Assimilation of geomagnetic data into dynamo models, an archeomagnetic study

Sabrina Sanchez

To cite this version:

Sabrina Sanchez. Assimilation of geomagnetic data into dynamo models, an archeomagnetic study. Geophysics [physics.geo-ph]. Institut de Physique du Globe de Paris (IPGP), France, 2016. English. NNT : . tel-01636599

HAL Id: tel-01636599

https://theses.hal.science/tel-01636599v1

Submitted on 16 Nov 2017

UNIVERSITÉ SORBONNE PARIS CITÉ

Thèse préparée
à l'INSTITUT DE PHYSIQUE DU GLOBE DE PARIS
École doctorale STEP'UP – ED N°560
IPGP - Equipe de Géomagnétisme

Assimilation of geomagnetic data into dynamo models, an archeomagnetic study

par
Sabrina Sanchez

présentée et soutenue publiquement le
29 Juin 2016

Thèse de doctorat de Sciences de la Terre et de l'environnement

dirigée par **Alexandre Fournier**
& **Julien Aubert**

devant un jury composé de :

Catherine Constable Professeur (Scripps Institution of Oceanography)	Rapporteur
Nicolas Gillet Chargé de recherche (ISTerre)	Rapporteur
Johannes Wicht Research Fellow (Max-Planck Institute)	Examinateur
Yves Gallet Directeur de recherche (IPGP)	Examinateur
Julien Aubert Directeur de recherche (IPGP)	Co-encadrant
Alexandre Fournier Professeur (IPGP)	Directeur de thèse

UNIVERSITÉ SORBONNE PARIS CITÉ

Assimilation of geomagnetic data into dynamo models, an archeomagnetic study

Sabrina Sanchez

Résumé :

Le champ magnétique mesuré à la surface de la Terre résulte de la superposition de plusieurs sources, dont la principale est celle du noyau. Le champ magnétique du noyau terrestre est généré par une dynamo naturelle qui évolue sur différentes échelles de temps. Pour extraire des informations sur les phénomènes se déroulant à des échelles de temps séculaires à millénaires, nous utilisons des données archéomagnetiques. L'inhomogénéité spatio-temporelle des données archéomagnetiques actuellement cataloguées fait pourtant obstacle à une inversion robuste du champ magnétique du noyau. Pour modéliser ce champ, des contraintes a priori sur la complexité du champ sont généralement imposées. Nous proposons une alternative en introduisant des informations a priori issues de simulations numériques de la dynamo terrestre. Cette étude est divisée en deux parties. Dans la première partie, les aspects statiques de cette inversion sont étudiés ainsi que leurs impacts sur la modélisation du champ archéomagnétique terrestre. L'information a priori, créée à partir d'un modèle de dynamo, est reliée aux données en surface d'orientation et d'intensité du champ magnétique par des opérateurs d'observation non-linéaires ou linéarisés. Le champ moyen est ainsi estimé (ou analysé) à l'issue de cette expérience. En combinant les résultats obtenus aux données archéomagnetiques des trois derniers millénaires, il est possible de quantifier leur capacité à reconstruire le champ magnétique. Les résultats ont démontré que le champ archéomagnétique est bien reconstruit jusqu'au degré 3 d'harmoniques sphériques pour le premier millénaire avant JC, jusqu'au degré 4 pour le premier millénaire après JC et jusqu'au degré 5 pour le dernier millénaire. La seconde partie de cette étude explore les apports de l'utilisation des méthodes séquentielles d'assimilation de données sur l'estimation du champ archéomagnétique. Dans ce contexte, l'inversion effectuée dans la première partie de la thèse est propagée dans le temps par un modèle numérique de la dynamo terrestre en cycles consécutifs de prévisions et analyses. L'utilisation de cette méthode a permis l'estimation non seulement des variables observables, mais aussi de celles typiquement inaccessibles, telles que le champ magnétique en profondeur, l'écoulement au noyau et la anomalie de densité. Cette assimilation fut testée dans le contexte d'expériences jumelles en utilisant des données synthétiques de type archéomagnétique, Les résultats obtenues se sont montrés performants en terme de précision sur l'état du noyau, même quand les observations n'étaient disponibles que dans un seul hémisphère. Ceux-ci devraient permettre l'utilisation de vraies observations archéomagnétiques et d'étudier ainsi les processus physiques ayant lieu dans le noyau sur des échelles temporelles séculaires.

Mot clés : archeomagnétisme, problème inverse, assimilation de donées, simulations numériques de la dynamo

Abstract:

The magnetic field measured at the Earth's surface is the superposition of a plurality of sources, the main component of which originates in the core. The core field is generated by a natural dynamo mechanism, which evolves on a variety of time scales. Its longer term dynamics are only accessible by indirect observations, the archeomagnetic data. The heterogeneous spatial and temporal character of the archeomagnetic data catalog, however, does not allow for a well-constrained inversion of the core field. Instead, the inverse problem is generally regularized by imposing prior constraints limiting the complexity of the field. Here we introduce the concept of using prior information derived from numerical models of the Earth's dynamo. This study is divided into two parts. The first part considers the static aspects of this inversion and its consequences for archeomagnetic field modeling. The prior information, built on a dynamo model, is connected to the surface data in terms of the directions and intensity of the field, by means of nonlinear and linearized observation operators. This yields an estimate, or analysis, of the core field given the available data. By means of these two pieces of information and the archeomagnetic dataset from the last three thousand years it is possible to quantify the archeomagnetic data resolution. Our results show that the archeomagnetic field is well-resolved up to spherical harmonic degree 3 for the first millennium BC, up to degree 4 for the first millennium AD and close to degree 5 for the past thousand years. The second part of the study explores how a sequential data assimilation framework can help improving the estimation of the field in the archeomagnetic context. In this case, the static inversion performed in the first part of the thesis is propagated in time by the numerical dynamo model in a sequence of forecast-and-analysis cycles. This methodology allows for the estimation of not only the observable, but also of the hidden variables of the dynamo system, the magnetic field in depth, the flow throughout the core and the density anomalies for instance. The assimilation, tested in the framework of closed-loop experiments for archeomagnetic-like synthetic observations, shows good performance in terms of accuracy and precision of the core state estimation. In particular, the assimilation is robust even in the case where observations are only available over one hemisphere. This thesis opens the possibility for the assimilation of real archeomagnetic observations and the subsequent estimation of the physical processes operating in the core on secular time scales.

Keywords: archeomagnetism, inverse problem, data assimilation, numerical dynamo simulations

Acknowledgements

This thesis is the product of many valuable interactions with different people I encountered before and during its development. I would like to thank them for helping me throughout this experience.

First of all, I would like to express my gratitude to my advisors Alexandre Fournier and Julien Aubert for having given me the opportunity to work with them at IPGP. Always very available, welcoming, encouraging, patient and at the same time very dynamic, Alex and Julien were role models both in terms of good scientific practice and of friendship. Through these three years and a half of thesis, they have guided me through my exploration of many subjects inside and outside geomagnetism, always remaining very open to ideas and discussions.

I am also very grateful to Cathy Constable and Nicolas Gillet for accepting to report this thesis. Their experience in diverse fields of science and their curiosity led to very important remarks and fruitful discussions. I am also grateful to Johannes Wicht for examining my thesis and having asked many pertinent questions during my defense. Many of the questions I received during my defense were open ones, and I am looking forward to answering them (or at least attempting to) in the near future, in our work together. Last but not least, I thank Yves Gallet for having accepted to preside the jury of my defense and lead the discussion. Yves' perception was crucial in the development of the many future perspectives in which I hope we will collaborate.

Throughout my thesis, I had the pleasure of meeting annually with my "Comité de suivi de thèse" composed by Sacha Brun, Emmanuel Cosme and Yves Gallet. Their different fields of expertise provided many different points-of-view to the thesis which enriched my scientific arsenal and allowed me to resolve the problems I encountered. I would also like to thank the dynamo team from the ISTerre at Grenoble, mostly Nicolas Gillet, Nathanael Schaeffer, Dominique Jault, Phillip Cardin, Claire Bouligrand and Olivier Barrois for the many discussions during the AVSgeomag meetings. I am grateful as well to the the planetology team from Nantes, Hagay Amit, Benoît Langlais and Erwan Thébault, for having invited me for a seminar at the LPG at the beginning of my thesis. I must single out and thank Erwan for accepting to develop with me a post-doctoral project during the end of my thesis.

Thanks also to the geomagnetism/paleomagnetism lab of IPGP for having received me for the period of my thesis, in particular to Gauthier, Olivier, Pierre, Boris, Lambert, Kevin, Ted, Virginie, Aude, Guillaume, France and Jean-Besse. For my early times in IPGP I am especially grateful to Fotini Vervelidou, my first friend in France. With her enormous generosity and goodwill, she presented me the lab, Paris and many interesting people. I am very grateful to Maylis as well for providing me with good advice at the beginning of my stay in Paris. A big thank you to Franck Houssein as well for the help with editing Parody and programming in general. At the end of my thesis I also received important insights from the new crew of the lab, Vincent Lesur, Thomas Gastine, Stasia, Tanya, Marie and Venkatesh, as well as visiting researchers at IPGP as Mathieu Dumberry and Matthias Morzfeld.

The geomagnetism lab has provided me with very special office mates. Boris and Lambert have been an amazing company throughout these years. Their good mood and jokes helped building up a very good ambiance in the office. Also, I want to thank them for teaching me so many French expressions, it helped me not to ‘craquer mon slip’. I thank also with all my heart Marina, Maria, Stasia and Tanya for the very good moments which helped me relax and focus during the time of writing of my manuscript and the preparation of my defense.

Beyond the lab, I am deeply grateful for the opportunity to get to know other amazing friends in the institute, in particular Claire, Virginia, Viviana and Andreina. What a source of inspiration for life you all have been. Sharing with you my experiences made me grow a lot along the thesis as a person, and made me feel better about myself. Outside the institute, I am very grateful as well to my friends Popi, Laurent, Jess, Leo, Mariana, Isaac, Damien, Kostas, Vasso and Bertrand, they made my life a lot easier in Paris.

I am forever indebted to Katia Pinheiro, Ricardo Trindade and Oscar Matsuura. Their encouragement and support while I was still in my masters in Brazil was decisive to give me strength to go in this adventure in Europe to make a thesis. I am also deeply grateful to Elmo and Igor, my fake son and husband, my bffs. Their continuous contact, care and humor helped me immensely in some sad days. My immense gratitude also to the Igor Fellowship Society (IFS), for both the moral and financial help to come to Paris and to go visit Rio. It was not easy to adapt to a foreign country, language and culture, and your tenderness was a remedy against despair. Thank you Igor, Elmo, Cicero, Tati, Djé, Detre, Yuzo, Eugenia, Day, Waffer, Lari, Carol and everybody else. I am very grateful also to my friends from the masters and undergrad that came to visit me during my thesis, Diego, Filllipe and Fernanda.

I would not have been able to make this thesis without the help and support of my family. The humor of my father, the tenderness of my mother, and the love of my brothers Fernando and Julian were pillars not only during this phase, but through my whole life. I am also very grateful to my dad for making many of the amazing figures present in this thesis manuscript. I would like to thank as well my parents-in-law Josiane and Théagenis and also their friends Renée, Jean-Pierre, Simone and Paul. Their generosity and welcoming arms have made me feel they were my adoptive family in Europe.

Last, but not least, I am grateful to Anastase, not only for your love, but also for our scientific discussions and for revising my thesis. I guess you cannot say that you don’t know anything about geomagnetism anymore! You have been the most important thing to me these last years. Thank you for everything.

Paris, 30/08/2016

Contents

Chapter 1

Introduction

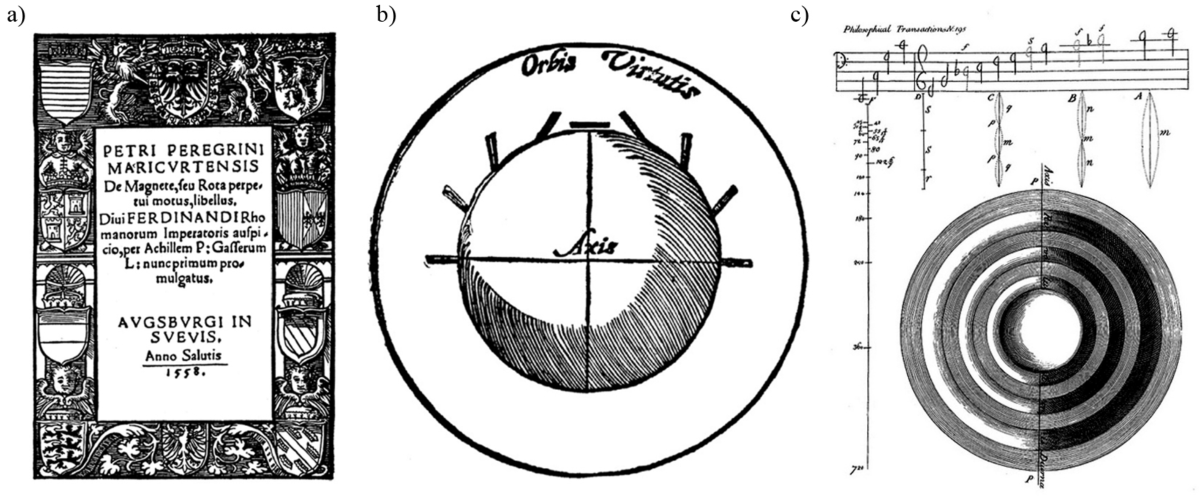
1.1 Early ideas on the Earth's magnetic field

One of the first documentations on the magnetic properties of certain types of rock, known as lodestones, is credited to the philosopher Thales de Miletus around 600 BC. It was said at the time, that the lodestones, relatively abundant in the region of Magnesia in Asia Minor, had a *magnetic* soul, through which they interacted with iron objects. In the following centuries, the lodestones were independently studied in China, where they were carved into thin objects, which led to the discovery of their close alignment with the geographic poles. The applicability of this discovery for geographical orientation, which generated the first compasses, was widely spread around 1000 AD. In Europe, compasses facilitated the exploration of the unknown world, triggering the Age of Discovery. In the following are listed three early studies that led to the development of the current theories on the Earth's magnetic field.

1.1.1 Pierre de Maricourt

Despite numerous subsequent records on the magnetic compass, the earliest scientific document related to magnetism is credited to Pierre de Maricourt's (also known as Petrus Peregrinus) 1269 AD letter *Epistola de Magnete* (Radelet-de Grave and Speiser, 1975). Differing from the general speculative texts of the epoch and displaying a similar structure as today's scientific reports, the *Epistola* (Figure 1.1a) presents a series of experiments on the attraction and orientation of the lodestone. Maricourt defined the notion of the magnetic poles of a spherical lodestone, and suggested that the lodestone poles are connected to the celestial sphere and its rotation around the Earth. Although not elaborating about the source of interaction between the celestial sphere and the lodestone, he does seem to imply a notion of magnetic filling of the space, resembling the modern idea of a magnetic field. Perhaps due to its assumed celestial origins, Maricourt asserts perfect symmetry to the magnetic interactions, which would be the cause of the coincidence between the magnetic and geographical meridians.

Breaking the symmetry ideal assumed by Maricourt, it was discovered around 1450 AD that the magnetic and geographic meridians are misaligned. The angle between the



PETRI PEREGRINI
MARICVRTENSIS
De Magnete, ſeu Rota perpetui motus, libellus,
Diui FERDINANDI Rhomanorum Imperatoris auſpicio, per Achillem P. Gaſſerum L. nunc primum promulgatus.
AVGSBVRGI IN SVEVIS.
Anno Salutis
1558.

Figure 1.1: Early documentation on the Earth's magnetic field and its origins. a) The *Epistola de Magnete* (1269 AD) of Pierre de Maricourt; b) Gilbert's *Terrella*, showing the inclination of a small magnetic needle as dependent on the lodestone latitude, as documented in his *De Magnete* (Gilbert, 1600); c) Halley's scheme of his theory of the concentric differentially rotating magnetized shells of the Earth's interior, responsible for the secular variation of the magnetic field (Halley, 1692).

geographic and the magnetic North was initially named the magnetic variation. Magnetic compasses of the epoch were commonly calibrated within a systematic variation, later known as declination, in order to be used for orientation. Almost a century after the British compass manufacturer Robert Norman discovered that the magnetic needle is not only misaligned with the geographic North, it is misaligned with respect to the horizon as well. The considerable dip of the needle discovered by Norman is known as the magnetic inclination, which was seen to be higher at high latitudes and smaller near the equator.

1.1.2 William Gilbert

After the change from Geo to Heliocentrism on the 16th century, William Gilbert presented his masterpiece *De Magnete* (Gilbert, 1600). Gilbert describes the use of a spherical lodestone as the prototype of a small Earth, the *Terrella*, and studies its interaction with a small iron needle assuming the role of a compass. Due to the pattern of inclination of the needle depending on its position over the *Terrella*, as well as the similarity with the latitude-dependent pattern of inclination measured around the Earth, he implies that the Earth itself is a magnet[1](Fig. 1.1b). Gilbert's theory was important for bringing the magnetic field from the perfect celestial sphere to the Earth's interior. This theory is, however, not the only one in Gilbert's *De Magnete*.

Gilbert also suggests that the close overlapping of the magnetic and geographical poles of the Earth is not a coincidence, that instead rotation and magnetism are closely connected in the universe. This idea, although not having enough contemporary repercussion, is presently acknowledged, as it will be further discussed. Furthermore, Gilbert

---

[1] Interestingly, the renowned statement "the terrestrial globe is itself a big magnet", generally attributed to Gilbert, is actually a citation of a paraphrase of Alexander von Humboldt (Sluijs, 2014).

also proposes that the declination of the compass is an effect of topological imperfections on the surface of the magnetic Earth, spoiling the perfect magnetic dipole. For Gilbert, the Earth's magnetic field was unchangeable in time.

In addition to the magnetism of the Earth, Gilbert documents in his book his experiments on the effects of temperature on the magnetization of iron-based objects. In synthesis, when the object is heaten up to very high temperatures it looses its previous magnetization, and when cooled down it acquires a new magnetization. Already mentioned by Maricourt, this notion anticipates the definition of the Curie temperature, a critical point linked to the physical properties of materials.

Contemporary to Gilbert, Guillaume de Nautonier came onto the same conclusions on the Earth as a magnet explored in *De Magnete*. He also incorporated numerous declination measurements around the globe to defend the idea of the magnetic poles being tilted relative to the Earth rotation axis (Courtillot and Le Mouël, 2007). The attention given to measurements by Nautonier is a sign of the change in paradigm of the 17th century, when diverse mathematical breakthroughs allowed a transition from purely qualitative to quantitative theoretical developments. Based on 50 years of declination measurements, Henri Gellibrand discovered in 1634 AD that the magnetic field presented temporal variability. This discovery, now known as the secular variation of the Earth's magnetic field, opposed Gilbert's ideal of an immutable magnetic field. The secular variation, shortly after confirmed in a worldwide sense, induced an awareness in the scientific community on the need of more measurements of the magnetic field. Subsequently, within a broad directional dataset, Peter Perkins introduced in 1680 AD the idea that the magnetic field would be better described in terms of a more complex system, formed by a superposition of two misaligned dipoles.

1.1.3 Edmond Halley

Based on Perkins' idea, Edmond Halley proposed in 1692 AD an ingenious theory of a magnetic Earth which accounted both for the multipolarity of the magnetic field and the temporal variations in declination (Halley, 1692). Disposing of a variety of declination data, he observed a general but complex westward tendency of the declination variation in Europe, Africa and the South Atlantic. From these observations he concluded that the magnetic field generated by the Earth bore a dynamic nature and proposed therefore a model of a dynamic Earth. In Halley's model, the magnetic field at the Earth's surface was the contribution of concentric magnetized spherical shells composing the internal layers of the Earth rotating at different rates (Fig. 1.1c). In his simple two-layer model, the external shell, which humanity inhabit, would have its North magnetic pole underneath North America. The inner layer would be rotating in a lower rate, with a North pole localized underneath Europe. The differential rotation of the spheres in Halley's model would under such configurations create a westward drift of the magnetic declination within a periodicity of approximately 700 years. Halley also suggested that additional data would perhaps reveal an increasingly complex stratification of the Earth's interior, and moreover, that the layers could comprise inner atmospheres. The diffusion of the atmospheres between layers would be in his view a possible origin for the creation of the Northern Lights, better known as Auroras. Even though an exaggerated extrapolation of his theory, it was the first scientific work linking the Auroras to magnetic phenomena.

Although Halley was the first one to link the Auroras with the Earth's magnetism, it was the clock-maker George Graham who presented clear evidences on the matter. In 1723 AD, he discovered by the use of an extremely precise compass that the declination presented daily perturbations at specific times of the day, around noon. This phenomenon suggested that the magnetic field was subject to a diurnal variation. Together with observers of the diurnal variation in Sweden, he noticed as well that the perturbations were latitude-dependent, and were stronger when Auroras were observed. These perturbations, often of very high amplitude and frequency, introduced the idea of geomagnetic storms.

1.2 Solar inspiration: the dynamo theory

In the days following September 1st of 1859, intense geomagnetic storms were recorded, generating very bright Auroras, able to be observed even in unusual low-latitudes locations, like the Caribbean, Hawaii and central Africa. The geomagnetic storm was so strong it affected telegraph networks all over Canada and Northern US. The amateur astronomer Richard Carrington had observed, the day before the beginning of the magnetic storms, a bright burst of light emitted in the vicinity of a sunspot group (Carrington, 1859); it was the first observation of a solar flare. The coincidence of the two phenomena suggested a solar-terrestrial connection of electric and magnetic nature, which was more evidenced in the following years by the joint monitoring of sunspot number and geomagnetic indices in magnetic observatories.[2] Evidence of the Sun being magnetic came almost five decades later, within the discovery of very strong magnetic field of sunspots (of the order of 10^3 G $= 10^5 \mu$ T, approximately 10^4 times bigger than the Earth's magnetic field)[3] by the Zeeman effect (Hale, 1908).

The discovery of the magnetism of the Sun puzzled the scientific community at the time, since the Earth's magnetic nature was still thought to originate from the magnetization of its rocky structure. Joseph Larmor proposed in 1919 a new theory on the source of the solar magnetism, suggesting that the dynamical magnetic field seen in sunspots had its origin in electromagnetic induction (Larmor, 1919). He suggested that the motion of the electrically conducting solar plasma under a background magnetic field would induce electric currents, which would in turn interact with the background magnetic field. Therefore, for a small seed magnetic field, such a system could generate a natural 'self-excited' dynamo. Larmor also said the same principle could be applied to the Earth, but that would require part of the Earth's interior to be fluid. In fact, magnetic fields generated by natural dynamos are known to be ubiquitous in the universe, not being restricted only to the Sun and the Earth. Within subsequent decades after Larmor's dynamo theory, the basics of magnetohydrodynamics (MHD) begun to be explored, and with it the study of the conditions for natural dynamo action. Although it is clear that natural dynamos are possible within fluidity and rotation (Moffatt, 1978), modeling dynamo systems proved to be extremely challenging.

In 1933 Thomas Cowling published what is now known as the first anti-dynamo theorem (Cowling, 1933). The theorem states that an axisymmetric flow cannot maintain a

---

[2] Even though many lines of evidence supported the solar-terrestrial electromagnetic link, Carrington himself was skeptical about it, having said on the matter "one swallow does not make a summer".

[3] 1 T $= 10^4$ G.

steady axisymmetric magnetic field by dynamo action. The very specific character of the theorem, focused on steadiness and axisymmetry, seemed to discredit the possibilities of dynamo action in the Sun and the Earth, given the nearly axisymmetrical character of the solar and terrestrial magnetic fields. The subsequent family of anti-dynamo theorems created a hiatus in solar and terrestrial magnetic field theory development. What was not accounted for at the time was the importance of the departure from symmetry for dynamo action, which would lead to the need of considering the three-dimensional structure of the problem.[4]

1.3 The terrestrial interior

The Earth's internal structure began to be truly unveiled within the work of Emil Wiechert, by means of comparisons of surface rocks densities with the mean density of the Earth. Wiechert created in 1897 the first quantitative model of the Earth, composed of a silicate mantle of rock and a deep and denser metallic core, similar in composition to iron meteorites. The core-mantle boundary (CMB) was then hypothesized as a discontinuity in density from the two main layers of the Earth. Wiechert speculated that such a region could be found by the propagation pattern of seismic waves originating from earthquakes, which inspired the beginning of seismologic imaging of the Earth. The existence of the core was confirmed by Richard Oldham in 1906 by a study of seismic P and S waves arrivals, and further estimation of the CMB depth (approximately 2,900 km, Figure 1.2) was made by Beno Gutemberg in 1912. In the next decade, the fluid nature of the core was revealed, and further, in 1936, Inge Lehman discovered yet another discontinuity, the existence of a solid inner core within the fluid outer core (Fig. 1.2).

After considerable developments in the analysis of seismological observations, comprising wave propagation and the normal modes of the Earth, as well as its mass and inertia moment, Dziewonski and Anderson (1981) built the Preliminary Reference Earth Model (PREM). Besides the seismic velocity estimates, the isotropic model specified the mean density variations with depth, shown in Figure 1.3. Despite the big density jump at the CMB and the smaller one at the inner core boundary (ICB), the density in the core does not vary much in the PREM. From density, the gravity and pressure throughout the core can be estimated as well (Fig. 1.3).

The discovery of the fluid outer core revived the ideas proposed by Larmor and paved the way for exploring the possibility of dynamo action within the Earth's interior. In 1946 Walter Elsasser proposed a first dynamo model, in which convection within the fluid outer core would sustain the magnetic field through electromagnetic induction compensating the Ohmic dissipation. He also introduced the different geometries of the core's magnetic field, poloidal and toroidal, which would interact generating the dynamic field we observe at the Earth' surface. Although Elsasser's work consisted of finding whether prescribed steady flows could sustain large-scale magnetic fields (what is known as the kinematic dynamo problem), it was the first step on the MHD modeling of the core.

---

[4] Despite Cowling's anti-dynamo theorem, nearly-axisymmetric 2D dynamo models are able to sustain dynamo action. Such models, generally in the mean-field context, manage to circumvent Cowling's theorem by the inclusion of a non-axisymmetric term based on the action of the Coriolis force, generically known as an α-effect.

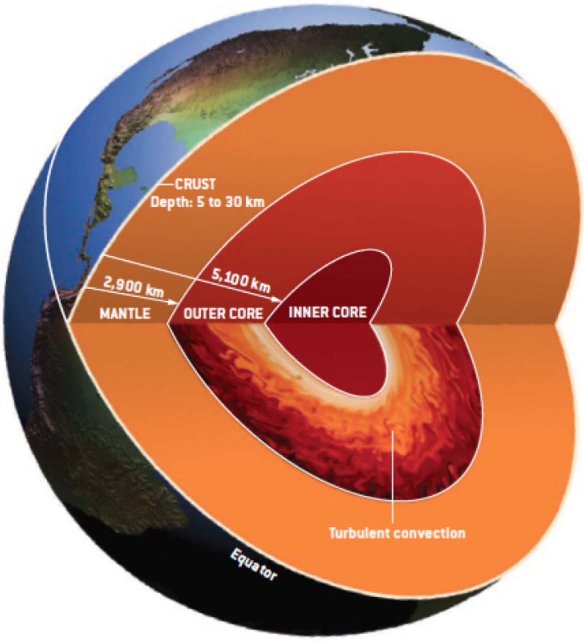


Figure 1.2: Scheme of the main layers composing the Earth's interior. From Glatzmaier and Olson (2005).

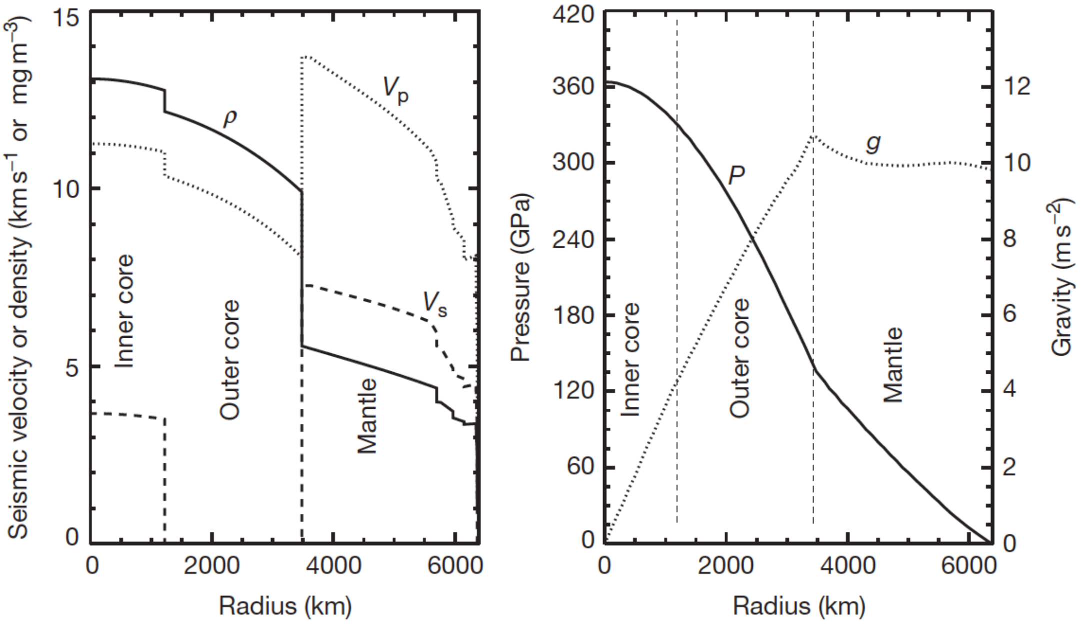


Figure 1.3: Seismic velocities through the Earth interior and its derived properties from the PREM model. The left figure shows the seismic P and S waves and density with depth, while the right figure shows the variability of pressure and gravity. From Olson (2015), based on Dziewonski and Anderson (1981).

1.3.1 Mantle processes

Since its formation, the Earth has been loosing its heat to the surrounding space. This process, known as secular cooling, is key to understanding the thermal state of the planet and therefore its dynamic behavior. The Earth's two major internal heat sources consist in radiogenic heat within the mantle and crustal rocks as well as the primordial heat trapped within the Earth since its formation from the planetary nebula. Independently from the origin, heat can be transported either by radiation, conduction or convection. Convection, in particular, plays a key role on the whole Earth dynamics. Plate tectonics is, for example, the surface manifestation of mantle convection. The mantle, although in solid state, is at sufficiently high temperatures for the rocks to present a ductile behavior over geological time scales. The thermal gradient between the colder top and its hotter base create slow convective motions, generating shear on the lower crust. This shear is responsible for many surface phenomena, like the spread of oceanic ridges and subduction of denser cold plates. Subduction of the cold plates results in a heterogeneous temperature and composition in the mantle, reinforcing the convective patterns (see Ricard, 2015, and references therein for a review on the physics of mantle convection).

Although the tracking of the sinking cold tectonic slabs is straightforward in the upper mantle, it is difficult to estimate the depth to which they survive within the mantle. Some studies reveal that such heterogeneities do make it to the lower mantle (Bercovici, 2015), and therefore determine a heterogeneous heat flux distribution at the CMB. Such lateral heterogeneities, in terms of temperature, phase and composition, reveal the complexity of the CMB. At the CMB, a certain degree of interchange in composition is supposed to happen, affecting the lowermost part of the mantle, the D” layer, and the upper part of the outer core, possibly creating a stratified layer (Buffett and Seagle, 2010; Gubbins and Davies, 2013; Souriau and Calvet, 2015).

1.3.2 The dynamic Earth’s core

The CMB controls the rate at which the core is losing heat, playing a key role on the dynamics of the core. As mentioned before, in contrast to the silicate rich rocky mantle, the core is mainly composed of molten iron. The viscosity of iron at the pressure and temperature corresponding to the outer core is counterintuitively small, comparable to the viscosity of water at the Earth’s surface (Poirier, 1988). The high fluidity of the outer core allows for thermal convection transporting heat from the hotter ICB to the cooler CMB.

Due to its relative abundance in chondritic iron meteorites, which are supposed to be residuals from the planetary core of protoplanets, there is strong evidence that the Earth’s core might consist mostly of iron (Birch, 1952), or most probably a Fe-Ni alloy. High pressure mineral physics experiments evidence that lighter elements must constitute an important part of the core composition, for observations show a core with lower density than the one formed by the pure Fe-Ni alloy (Alfè et al., 2003). Lighter elements like oxygen, silicon and sulfur are strong candidates for the core’s density deficit, but their proportion and even the possibility of the presence of carbon and hydrogen is still controversial (Hirose et al., 2013). The lighter elements are important not only for reconstructing the history of the Earth’s formation but also have an important impact on core convection.

Given the secular cooling of the Earth, the inner core is also losing its heat, which implies it is progressively growing with time as result of the crystallization of the lowermost part of the outer core. Since the heavier elements in the core have a lower freezing temperature, the crystallization of the core implies a solidification of the heavy elements and the release of liquid lighter elements into the outer core. This phase change is exothermic, releasing latent heat and low density, therefore buoyant, elements in the outer core. Such a phenomenon triggers what is known as thermochemical convection. Chemical convection by the crystallization of the inner core is though to be responsible for 80% of the whole convection in the outer core, while thermal convection would amount to the remaining 20%. However, a relatively new hypothesis for the core's convective mechanism considers the precipitation of magnesium (O'Rourke and Stevenson, 2016), a scenario which remains to be explored.

Still, independently of the driving mechanism, the convective pattern of the outer core is tightly constrained by the Earth's rotation. Theoretical fluid dynamics, confirmed by laboratory experiments and numerical simulations, state that homogeneous steady flows in a rotating spherical shell tend to be organized in columnar structures parallel to the rotation axis (the Taylor-Proudman theorem, Taylor, 1922). Such two-dimensional columnar flow structure is known as geostrophic. In presence of convection, the geostrophic columns display helicity, that is, an additional perpendicular component to vortical motions.

The flow inside the outer core is believed to be approximately geostrophic, in which case the region defined by the continuation of the inner core equatorial boundary through the rotation axis creates a region in the core known as the tangent cylinder (TC), illustrated in Figure 1.4. Although the columns outside the TC are continuous along the equatorial plane, the columns inside the North hemisphere TC do not necessarily imply a direct connection with their southern counterparts. The different flow regimes in the outer core are key to the Earth's dynamo mechanism.

1.4 The geodynamo

The electrical conductivity of the molten iron at outer core temperature and pressure conditions is currently estimated at 1.4×10^6 S/m (Pozzo et al., 2012), two times higher than previous estimates.[5] Such high conductivity favors the generation of electrical currents which interact with the embedding magnetic field. This phenomenon, the electromagnetic induction, has its basis in the laws of electrodynamics, rooted on Maxwell's equations. As shown in Fig. 1.4, the rotating flow under convection provides the generation of helical flows, closely aligned with the Earth's rotation axis. The relative movement of flow and seed magnetic field induces electrical currents, generating secondary magnetic field which in turn interacts with the flow. Given the helical flow configuration, the resulting magnetic field generated by the geodynamo is nearly dipolar.

The electromagnetic induction on a dynamo mechanism depends on two main phenomena, advection and diffusion of the magnetic field. If advection is ignored, the magnetic field variability is related to the diffusion of the magnetic field by the Joule effect. For the largest spatial scale of the magnetic field, the dipole, that would imply a free decay time of 50 ka. The fact that the magnetic field has been shown to be much older than this time interval, estimated through paleomagnetic observations to be as old as ~ 4.0 Ga (Tarduno et al., 2015), points to the important role of the inductive advection mechanism. On the contrary, if magnetic diffusivity is ignored, which is the case if the fluid is a perfect

---

[5] Iron at room temperature and pressure has 1×10^7 S/m conductivity.

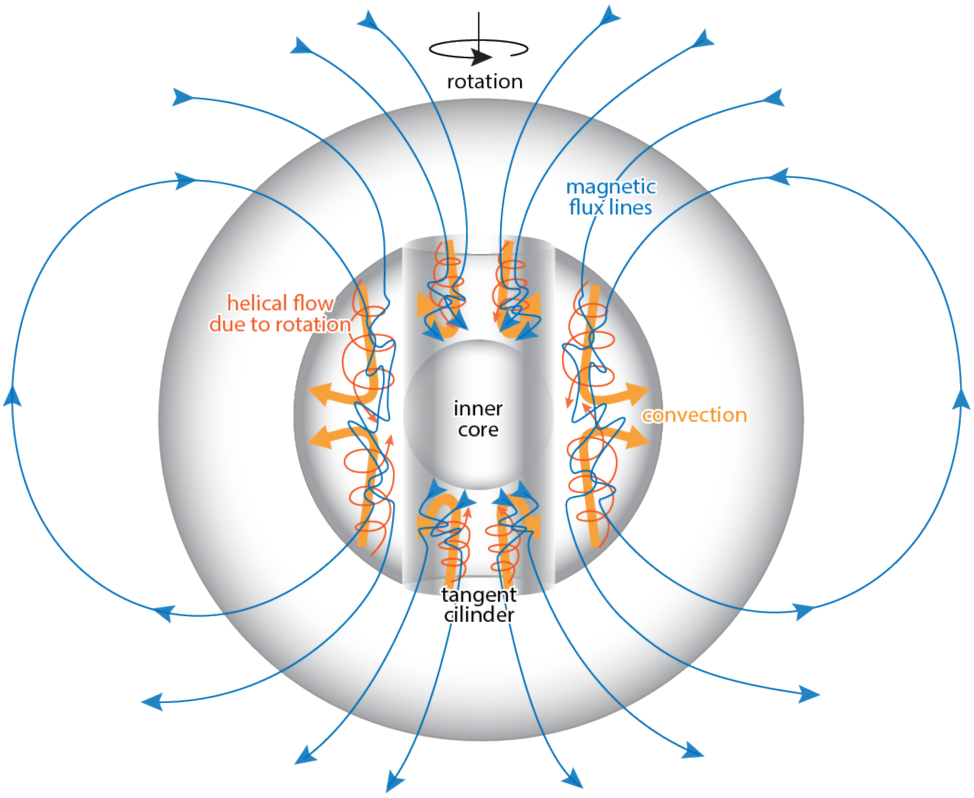


Figure 1.4: Basic features of the dynamo mechanism at work in the Earth's core. The secular cooling of the Earth provides the driving mechanism for core convection in the outer core, where the fluid tends to be organized in an helical flow due to the planet's rotation. Although different regimes are found for the interior and exterior of the tangent cylinder, the electrically conducting columnar flow tends to arrange the overall magnetic field in a dipolar configuration.

conductor, the magnetic field would be 'frozen' in the fluid and would then flow given its inertia. This condition is known as the 'frozen flux' hypothesis. Estimates of the flow velocities at the CMB using this hypothesis give a mean advection speed of 15 km/year (or 0.5 mm/s) (Holme and Olsen, 2006).

Another possibility is that the flow field in the outer core is close to geostrophic, embedded in a magnetic field frozen in the flow. A given rotational instability of the cylinders with respect to each other would generate a shear in the magnetic field lines perpendicular to the rotation axis. Since nature opposes changes in magnetic flux, as stated by Lenz's law, the magnetic field would impose a feedback reaction on the flow, the Lorentz force, generating the so called torsional oscillations. Such interactions would produce a wave-like pattern in the magnetic field lines reaching the Earth's surface. The interpretation of torsional oscillation time scales are still under debate. Gillet et al. (2010) support a periodicity of 6 years for the core torsional oscillations, while older estimates based on the root mean squared magnetic field at the core-mantle boundary suggests larger decadal time scales.

Many different features of the geodynamo are therefore still not fully understood. In order to investigate the diverse physical processes of the dynamo system, one has to resort to the MHD equations describing the laws of conservation of the different fields: the induction equation, the Navier-Stokes equations and the heat and compositional transport. Given the complexity of such a system of equations governing the main physics of the dynamo, together with the uncertainty related to its underlying parameter regimes, there is a great interest in tackling the problem numerically. The different aspects of numerical dynamo modeling will be explored in more detail in Chapter 2. However, before entering the discussion on the mechanisms responsible for driving the core magnetic field, it must be acknowledged that a considerable amount of information from the dynamo can be retrieved from Earth surface observations of the magnetic field.

1.5 The present geomagnetic field

The present magnetic field is currently monitored by a variety of networks involving different types of data. Local and regional monitoring are made by airborne, ship and field surveys and are very important in retrieving small spatial scales of the Earth's magnetic field. Such small scales are generally associated with the near surface signal of the magnetic crust, and therefore deliver little information about the magnetic field of the deep Earth. The global magnetic field can be directly monitored mostly by two different sources of data.

1.5.1 Geomagnetic data

Magnetic observatories

The first magnetic observatory was created in the German city of Göttingen by Carl Friedrch Gauss and Wilhelm Weber in 1832. It registered the intensity F of the magnetic field $\mathbf{B}$ and its horizontal component H. Since then, many new observatories have been created, measuring not only the intensity of the magnetic field but also its independent vector components, X (north), Y (east) and Z (downward). A scheme of the different components of the magnetic field vector is shown in Figure 1.5. The magnetic observatory network is nowadays organized by INTERMAGNET[6], where data are easily obtained.

Although worldwide, the geographical distribution of geomagnetic observatories is inevitably poorly distributed, as besides the obvious bias toward continental areas, the observatories are mainly concentrated in Europe for historical reasons. In order to compensate for this bias, many observatory-like measurements are performed in under sampled areas. Those measurement sites are called geomagnetic repeat stations.

Satellite missions

Since the dawn of the space age in 1957, scientists have been interested in obtaining satellite data of geophysical relevance. With the Russian satellite Sputnik 3 it became possible to contemplate the first satellite magnetic data (although the sampling was limited to Soviet Union territory), paving the way to new magnetic satellite missions. The missions

---

[6] http://www.intermagnet.org/

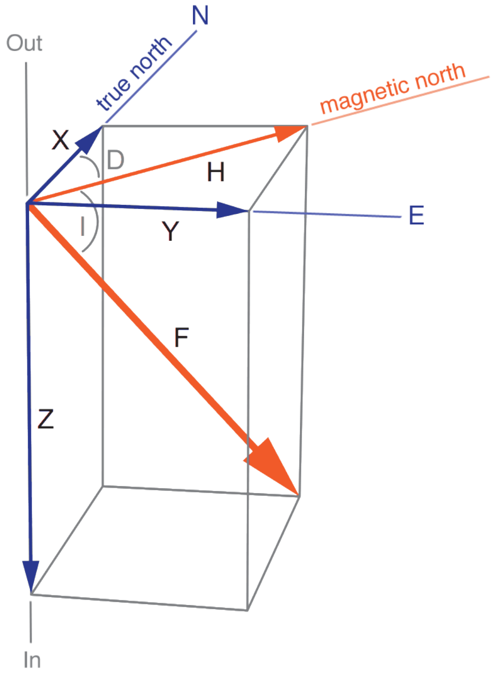


Figure 1.5: The different components of the magnetic field vector **B** and its additional elements, as measured at and beyond the Earth's surface. X, Y and Z are the northward, eastward and downward components of the magnetic field vector, respectively. Additionally, F represents the intensity of the magnetic field, D the declination (the angle between the horizontal projection H of the magnetic field vector and the true north) and I the inclination (the dip angle of the magnetic field vector).

of most relevance were POGO (1965-1971), Magsat (1979-1980), Ørsted (1999–present), CHAMP (2000-2010) and SAC-C (2001-2004) (see Friis-Christensen et al., 2009, for more details on the satellite missions for monitoring of the magnetic field). The continuity of the previous missions is now given by the Swarm satellite constellation. Launched in 2013, Swarm monitors the magnetic field in terms of vectorial and intensity data within so far unreached resolution.

The global coverage offered from the satellites tracks is by far the best amongst the various magnetic data types. However, since the satellites are constantly moving with respect to the Earth, there is an inherent difficulty in assessing the origins of the variations of the magnetic field registered by the data. Also, due to the satellite's altitude, placing them in between the ionospheric and magnetospheric typical layers, the separation of sources between the external magnetic field and the main field is difficult.

1.5.2 External and internal sources

The Earth's magnetic field is a superposition of many different sources. Beyond the core's magnetic field, there are magnetic sources of different natures in the mantle, lithosphere, oceans, atmosphere and beyond. A review of the different sources is given in the following.

Ionosphere

Above the Earth's surface, electrical currents can flow in the atmosphere, such as the ones related to electrical discharges in thunderstorms. Further on, at the upper layers of the atmosphere, particles are easily ionized by the UV solar input, generating a plasma region which is called the ionosphere (Kivelson and Russell, 1995). Due to the dependency on the solar irradiation, the ionosphere changes in accordance to the hour of the day, generating the diurnal variation mentioned in Section 1.1. Apart from the radiation, the solar activity can also produce additional features on the ionosphere, specially in its uppermost layers, like the Equatorial Electrojet (EEJ). However, the effects of the solar activity are mainly felt on the boundaries of the Earth's magnetic field, the magnetosphere. A simple description of the localization and geometry of such sources is given in Figure 1.6a.

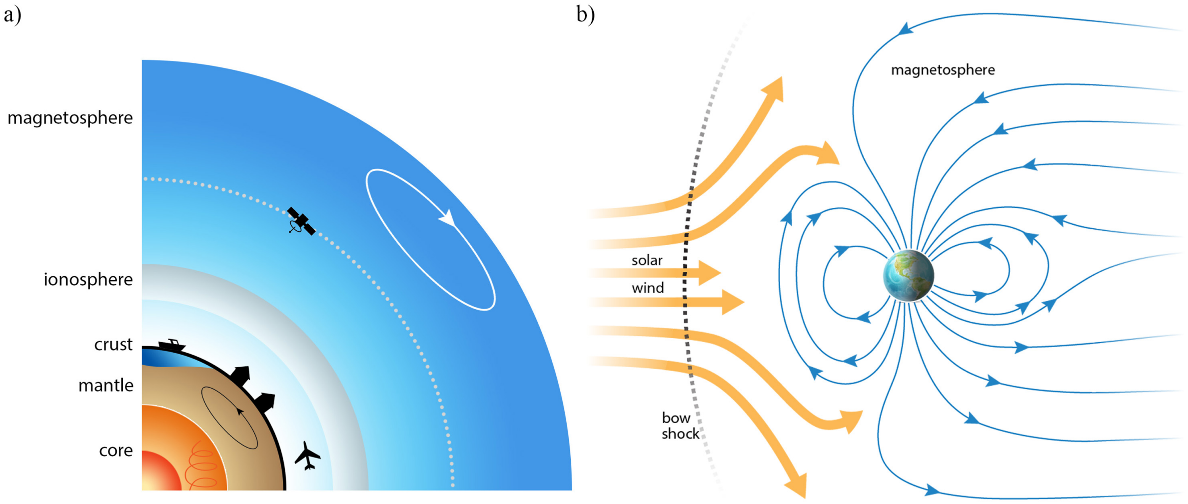


Figure 1.6: The different sources of the Earth's magnetic field and its variations. a) shows the illustrative stratification of the Earth from its deep internal layers to the higher parts of the atmosphere and beyond. The induced currents in the magnetosphere, crust, oceans and mantle, as well as the crustal magnetization, contribute to the magnetic signal measured by satellites, magnetic observatories, airborne and ship surveys. b) shows how the magnetic field of the Earth (Fig. 1.4) is distorted by the solar wind, shaping the magnetosphere.

Magnetosphere

The magnetosphere is the region enveloping the Earth which corresponds to the bow shock resulting from the solar wind deflection by the terrestrial magnetic field. It compares in shape to a distorted dipole of roughly 10 times the Earth's radius on the day side. The solar wind is a supersonic stream of plasma coming from the upper layers of the Sun, modulated by its magnetic activity. It is generally stronger when associated with Coronal Mass Ejections (CMEs), generally of higher frequency in times of high solar activity, like the Carrington event mentioned in Section 1.2. The solar wind is embedded in a magnetic field configuration resulting from its ejection from the Sun's magnetic field. The complex Sun-Earth interaction produces a variety of large-scale electric currents in very different configurations around the magnetospheric environment. Also, due to the Lorentz force, the ionized particles composing the solar wind are deflected from their initial path and follow the Earth's magnetic field lines (Fig. 1.6b). Some of those particles follow specific magnetic field lines which converge to the Earth's magnetic poles, and when interacting with the upper atmosphere, they produce the already mentioned polar Auroras.

Lithosphere and mantle

Going back to the solid Earth, it must be acknowledged that electric currents can also flow in the lithosphere and mantle, due to their finite, but low, electrical conductivity. Such induced currents are the feedback effect from changes in the embedding magnetic field, either from the core's secular variation or the external field variability. The combination of such feedback currents in the solid Earth and oceans is generically called the 'induced internal field' (Fig. 1.6a). In the oceans, the electric currents are generated not only by electric induction, but also due to the global oceanic circulation. This signal is, therefore, partially modulated by a seasonal signal. In addition to the induced magnetization, the uppermost part of the lithosphere, the crust, can also bear remanent magnetization.

The electrical conductivity of the lowermost part of the mantle is supposed to be quite a bit higher than the overall mantle (Ohta et al., 2010). This has important implications for the core field, for it would imply that the base of the mantle acts like a filter of the high frequency magnetic signal of the core due to the skin-effect. The electrical conductivity of the mantle is, however, poorly known, leaving the issue of high frequency core magnetic field variability as an open discussion.

1.5.3 Magnetic field models

Despite the difficulty in modeling the Earth's magnetic field due to the superposition of sources embedded in all geomagnetic data, there exist different techniques that can be applied to separate and construct global models of the magnetic field. In particular, the decomposition in spherical harmonic functions developed by Gauss in 1838, deals with the problem of dividing the external and internal components of a potential field. In the case of the magnetic field it is derived in the following.

As already mentioned, the magnetic field vector $\mathbf{B}$ at a radius a representing the mean radius of the Earth's surface is the superposition of different sources of the magnetic field. $\mathbf{B}$ can therefore be represented as the sum of sources external and internal to the surface defined by the radius a

$$\mathbf{B}(a,\theta,\phi,t) = \mathbf{B}_i(a,\theta,\phi,t) + \mathbf{B}_e(a,\theta,\phi,t), \tag{1.1}$$

where θ and ϕ are spherical geocentric angular coordinates, the colatitude and longitude respectively, and t the time.

At the Earth's surface, the intensity of the external field is small compared to the internal one. Additionally, the external field varies on very short time scales, from seconds to months, while the variations from the core measured at the Earth's surface range from decadal to secular time scales. The purpose of this study is to focus on the core's magnetic field and its secular variability, so the external sources of the magnetic field are going to be considered as negligible. In this case, the magnetic field vector can be expressed as the gradient of an internal magnetic potential V_i, which in spherical coordinates is given by

$$\mathbf{B}_i(a, \theta, \phi, t) \;=\; -\nabla V_i(a, \theta, \phi, t). \tag{1.2}$$

As the potential V_i satisfies Laplace's equation, the solution can be expressed in terms of the spherical harmonic (SH) basis

$$V_i(r, \theta, \phi, t) = a \sum_{\ell=1}^{\infty} \sum_{m=0}^{\ell} \left(\frac{a}{r}\right)^{\ell+1} [g_\ell^m(t) \cos m\phi + h_\ell^m(t) \sin m\phi] P_\ell^m(\cos\theta), \tag{1.3}$$

where ℓ and m are respectively the degree and order of the SH expansion, $g_\ell^m(t)$ and $h_\ell^m(t)$ the time-dependent Gauss coefficients and P_ℓ^m the Schmidt quasi-normalized associated Legendre functions.[7] Since it is not possible to actually compute a model with the infinite expansion, a given truncation L is chosen based on the expected resolution of the data. The largest scale coefficient groups are called the dipole (for which $\ell = 1$), the quadrupole ($\ell = 2$) and octupole ($\ell = 3$). The case $m = 0$ corresponds to zonal, or axisymmetric, modes, and the other extreme, for which $\ell = m$, corresponds to sectoral, or meridional, modes. If the lithosphere and mantle are considered as electrical insulators, the global magnetic field calculated from Equations 1.2 and 1.3 can be downward continued to the CMB, below which one must take into account the electric currents generated by the geodynamo.

The Gauss coefficients $g_\ell^m(t)$ and $h_\ell^m(t)$ can be retrieved given the set of available observations, that is, either X, Y, Z, H, D, I or F data (Fig. 1.5). Such retrieval consists in solving the geomagnetic inverse problem, extensively explored in the literature (e.g. Parker, 1994), which will be described in Chapter 4. Many models of the main geomagnetic field are available, differing basically in the dataset, time-span, inverse method and prior information. The main field model with a large projection outside the geomagnetic community is the International Geomagnetic Reference Field (IGRF). The IGRF consists of 5 year updates of the main field and of its secular variation. The models are derived from various candidate models from different teams around the globe, and the final estimate is a weighted mean. The IGRF model for the year 2015 (Thébault et al., 2015) can be seen in Figure 1.7.

The discrimination of the different internal sources of the magnetic field can be seen by means of the spatial power spectrum, also known as the Mauersberger-Lowes spectrum. This spectrum is calculated by

$$W_\ell^i = (\ell + 1) \left(\frac{a}{r}\right)^{2\ell+4} \sum_{m=0}^{\ell} [(g_\ell^m)^2 + (h_\ell^m)^2], \tag{1.4}$$

at any radius r in the current-free zone of the Earth's interior. Figure 1.8 shows an example of the power spectra as seen at the Earth's surface and at the top of the core. At the Earth's surface, the magnetic field is seen to be mainly dipolar, with an exponentially decreasing intensity towards the smaller scales. However, at around degrees 13 and 14,

---

[7] The coefficients g_ℓ^m and h_ℓ^m have the same units as the observed magnetic field vector, and are generally given in [nT] or [μT].

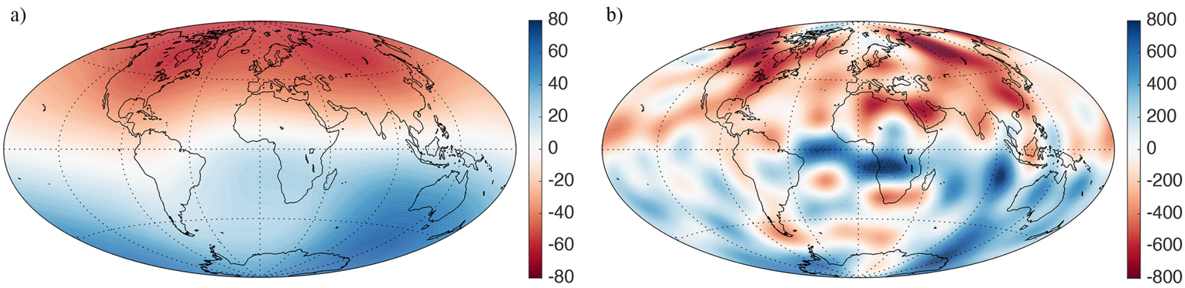


Figure 1.7: The IGRF model for the year 2015. The maps show the radial magnetic field a) at the Earth's surface and b) at the CMB (units displayed in μT).

this tendency is broken by a close-to-flat contribution. This change is better observed in the downward continued spectra, where it is seen to drastically and unrealistically increase for smaller scales. This change in regime corresponds to the contamination of the crustal magnetization in the internal magnetic field signal. The magnetic field of the core is therefore inevitably masked beyond SH degrees 13 or 14, what is known as the 'magnetic curtain' (Roberts and King, 2013).

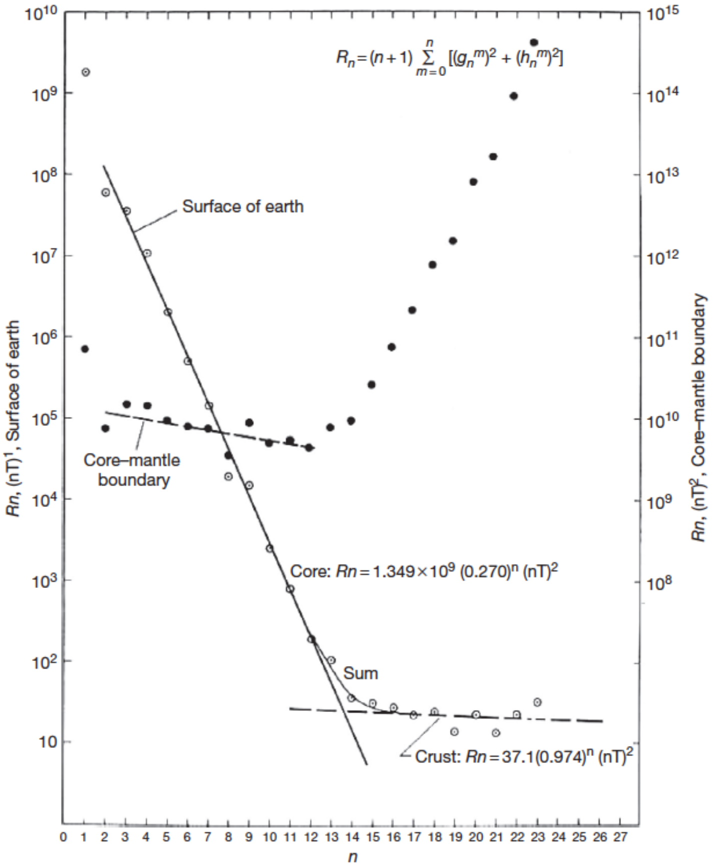


Figure 1.8: Magnetic field power spectra, or Lowes spectra, based on Magsat data. The power spectra is calculated both at the Earth's surface (open circles) and at the CMB (filled circles). The change in slope for both spectra shows the contribution of the crust magnetization to the core magnetic field, creating the 'magnetic curtain'. From Langel and Estes (1982).

Finally, as mentioned before, the magnetic field varies in a broad range of time scales. Although the external magnetic field, and also the induced internal magnetic field, are able to display decadal variability, due to the long-term modulation of the solar activity, the intensities changes are small compared to the secular variations of the magnetic field. The most dramatic example of such changes are the polarity reversals and excursions of the core magnetic field, with periods of thousand to million years (see Constable and Korte, 2015, and references therein for a review on secular to millennial variations of the magnetic field). On shorter, secular, time scales, the magnetic field displays variations originating from the advection of the field by the core flow.

1.6 Archeomagnetism

Magnetic field observations did not exactly begin with the establishment of magnetic observatories. Before that, historical measurements were registered, for instance, in navigational logbooks. This valuable record can extend the catalog of direct magnetic field observations into the past four centuries (e.g. Jackson et al., 2000). But in order to explore the longer-term behavior of the core's magnetic field and the dynamo, one has to face the problem of unavailability of direct observations of the magnetic field. However, indirect information from the magnetic field can be found in the locking of the magnetic signal in specific geological materials and some types of human artifacts, generally called archeomagnetic data.

1.6.1 Magnetic remanence

It has been long known that the magnetic field of certain rocks and artifacts depend on temperature. Bricks, for example, loose their magnetic signal when fired and acquire a magnetization when cooled. The fundamentals of this process are the following. When above a certain temperature known as the Curie point, the thermal energy of the given material is too high to allow for a memory of the ambient magnetic field. In this case, only induced magnetic fields are possible. After transitioning below the Curie temperature, however, the magnetic carriers embedded in the material, generally iron oxides as magnetite and hematite, remain roughly aligned with the ambient magnetic field directions. This process is known as thermal remanent magnetization (TRM), and is ubiquitous in igneous deposits such as volcanic rocks, as well as in fired clay-based archeological items (Figure 1.9a and 1.9b). In principle, the registered TRM of a volcanic rock or human artifact can be used to retrieve information from the past states of the magnetic field (Thellier and Thellier, 1959).

The directions of the fossil magnetic field vector can only be found in oriented samples, for example pieces composing ancient kilns and furnaces used to, amongst other things, harden clay in order to fabricate bricks and pottery. The retrieval of the full fossil magnetic vector information is not simple. In the archeological context, for instance, the directions of the original field cannot be accessed from portable objects like pottery. Those objects can therefore only give values of the field archeomagnetic intensity F. They are, however, important markers of the archeointensity variability on the secular time scale. Such information can be used to construct 'master curves' of the secular variation at a specific

Figure 1.9: Different types of archeomagnetic data. a) Lava flow (Morales et al., 2006), b) human artifacts (Genevey et al., 2009) and c) sedimentary cores.

location where archeomagnetic data are numerous (e.g. Gallet et al., 2002). In the case where the age of the magnetized object is not known, the master curves can serve as a tool for what is called archeomagnetic dating, frequently used for archeological reconstructions.

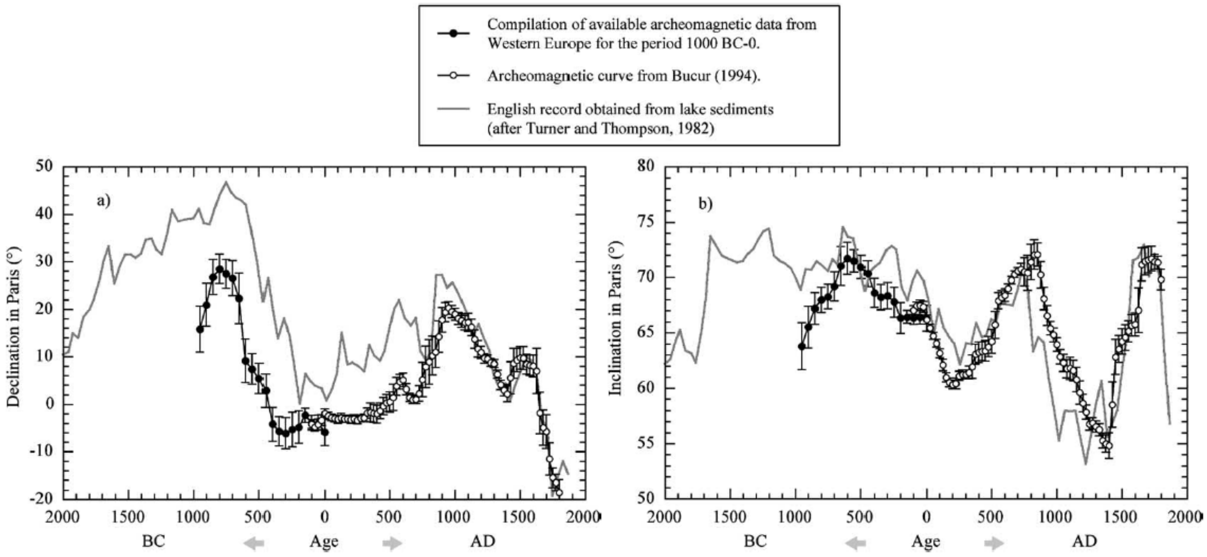


Figure 1.10: Archeosecular variation master curves for Paris from Gallet et al. (2002) for the magnetic a) declination and b) inclination. Also shown is the comparison with estimates from English lake sediments.

Furthermore, the information of the magnetic signal can be imprinted on materials by other processes, like for example depositional and detrital remanence (see Merrill et al., 1996, for more details). Detrital magnetic remanence (DRM) occurs in the context of lake or marine sediments (Fig. 1.9c), connected to the systematic alignment of magnetic carriers with the ambient magnetic field during sedimentary deposition. Due to orientation and measuring issues, magnetic declination and intensity from sediment cores are generally given in terms of relative values. Despite their relative character and other challenging characteristics, sediment data are, however, utterly valuable due to their continuous recording of the paleosecular variation (see for example Fig. 1.10).

Although simple, the primary magnetic register of the magnetized rock or artifact is hard to assess, for often the samples are seen to have more than one magnetic component, which could correspond to secondary processes leading to weaker remagnetizations or chemical alteration of the material. In order to retrieve the different components of a magnetized sample, the sample has to pass through careful laboratory protocols (Tauxe and Yamazaki, 2015, for a review on the measuring of archeointensity).

1.6.2 Chronology and uncertainties

The chronology of archeomagnetic data can be derived by various relative and absolute methods (Noller et al., 2000). In the case of human artifacts, archeological constraints based on either the stratigraphy of the archeological site or the specific style of the artifact are used to estimate the object age. Typical uncertainties for archeological dating generally are of the order of a few decades. In the case of lava flows, precise dating can be performed given the historical documentation of witnessed eruptions. Otherwise, the most frequently used source of paleomagnetic dating consists in radiometric methods, such as radiocarbon dating. Although not as frequent as archeologic dating, radiocarbon techniques are sometimes also applied for human artifacts. Radiogenic dating generally gives age uncertainties of the order of centuries. For sediments, varve counting can serve as a fine chronological tuning, improving the estimates from radiocarbon dating of certain stratigraphic layers with organic material. For non-varved sediments, however, age estimation can become difficult due to uncertainties on depositional rates.

A detailed description of paleomagnetic measurements can be found in Turner et al. (2015). The uncertainties in archeomagnetic data depend on both sampling and laboratory conditions and are generally given by a measure of the dispersion of a large number of specimens characterizing a given sample. In the case of directional measurements, the quality of observations is given by the α_{95} parameter, typically ranging from $1°$ to $2°$ (as it will be seen in Section 5.5). For archeointensity measures, the uncertainties are typically of a few μT. See Constable and Korte (2015) for a discussion on the many uncertainties affecting paleomagnetic studies.

1.6.3 Databases

There have been numerous archeo and paleomagnetic databases, differing in data type, data component, time span and regional or global character. Since the interest in this manuscript relies on global representations of the Earth's magnetic field, the focus is drawn to world-wide compilations. One of the first efforts in gathering globally distributed data is the PSVMOD1.0 (Constable et al., 2000), spanning the last 3 kyr and comprising archeomagnetic data as well as directional data from lake and marine sediments. Afterwards, the ArcheoInt compilation (Genevey et al., 2008), which as the name suggests focused on new archeointensities spanning the last 50 kyr, was incorporated to the contemporary database Geomagia50 (Donadini et al., 2006). The expanded version Geomagia50.V2 (Donadini et al., 2009), completed with available directional data, with complementary metadata, provides easy online access. A newer version, Geomagia50.V3 (Brown et al., 2015), extended the dataset to comprise sedimentary data as well. The different versions of the archeomagnetic dataset from Geomagia50 are shown in Figure 1.11.

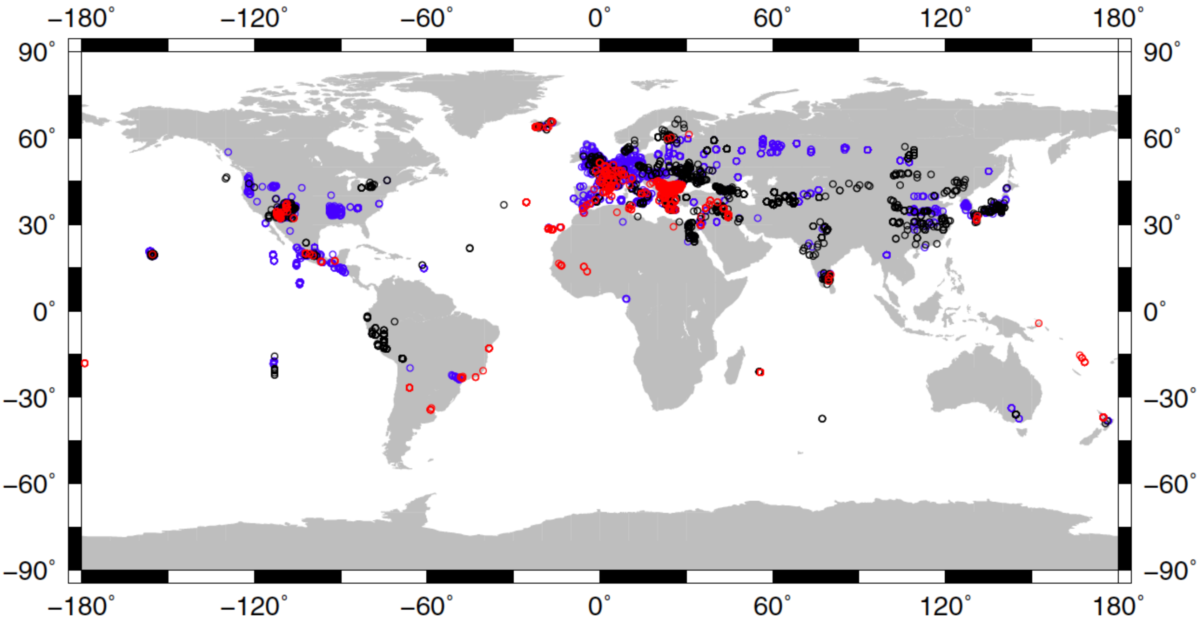


Figure 1.11: Global distribution of archeomagnetic data from the Geomagia50 compilation, corresponding to expansions of the dataset for different database versions: Geomagia50 (black), Geomagia50.V2 (blue) and Geomagia50.V3 (red). From Brown et al. (2015).

As can be seen in the figure, the archeomagnetic global database of Geomagia50 shows a clear hemispherical dichotomy. The data are concentrated mainly in Europe and western Asia, due to the availability of archeological artifacts, good historical documentation and proximity with more developed laboratories. The African continent, in contrast, is the least well sampled, despite its proximity with Europe. In general the Southern hemisphere is scarce in data, with a few intensity data, most of all in Brazil and Peru. In total, 95% of the data corresponds to the Northern hemisphere, possibly biasing estimates of the archeomagnetic field. However, the sampling of the archeomagnetic dataset is not as intuitive as one might think.

1.6.4 Sampling of the core

As mentioned in the previous section, archeomagnetic data are given by incomplete information of the magnetic vector by either its directions or its intensity. Those elements, the declination, inclination and intensity (Fig. 1.5), are given by

$$\begin{aligned} F &= \sqrt{X^2+Y^2+Z^2}, && (1.5)\\ I &= \arctan\frac{Z}{H}, \quad -\frac{\pi}{2} \le I \le \frac{\pi}{2} && (1.6)\\ D &= \arctan\frac{Y}{X}, \quad -\pi \le D \le \pi && (1.7) \end{aligned}$$

These data are nonlinear functionals of the vector components of the magnetic field at the Earth's surface. For the vector components, it is possible to represent them in terms of their corresponding Green functions. For example, following Gubbins and Roberts (1983), the downward component of the magnetic field at a given location of the Earth's surface $Z(\mathbf{r}_s)$ can be written in terms of the response from the source radial magnetic field at the CMB $B_r(\mathbf{r}_c)$, such that

$$Z(\mathbf{r}_s) = \int_{\Omega} G_Z(\mathbf{r}_s|\mathbf{r}_c) B_r(\mathbf{r}_c) d\Omega, \tag{1.8}$$

where $G_Z(\mathbf{r}_s|\mathbf{r}_c)$ is the Green function, or data kernel, over the CMB surface Ω which describes the mapping of $Z(\mathbf{r}_c)$ into $B_r(\mathbf{r}_s)$.[8] This assumption is of course valid assuming the mantle is an electric insulator.

The Green functions for each component of the magnetic field vector can be found in Constable et al. (1993) and are further discussed in Chapter 3. They describe the way each observation samples the top of the core, which is shown in Figure 1.12 for X, Y and Z observations at high latitudes. It is interesting to see how Z observations sample the top of the core directly underneath the observation site. X and Y observations, however, sample the top of the core within a certain angular distance, of approximately 23° (Gubbins and Roberts, 1983).

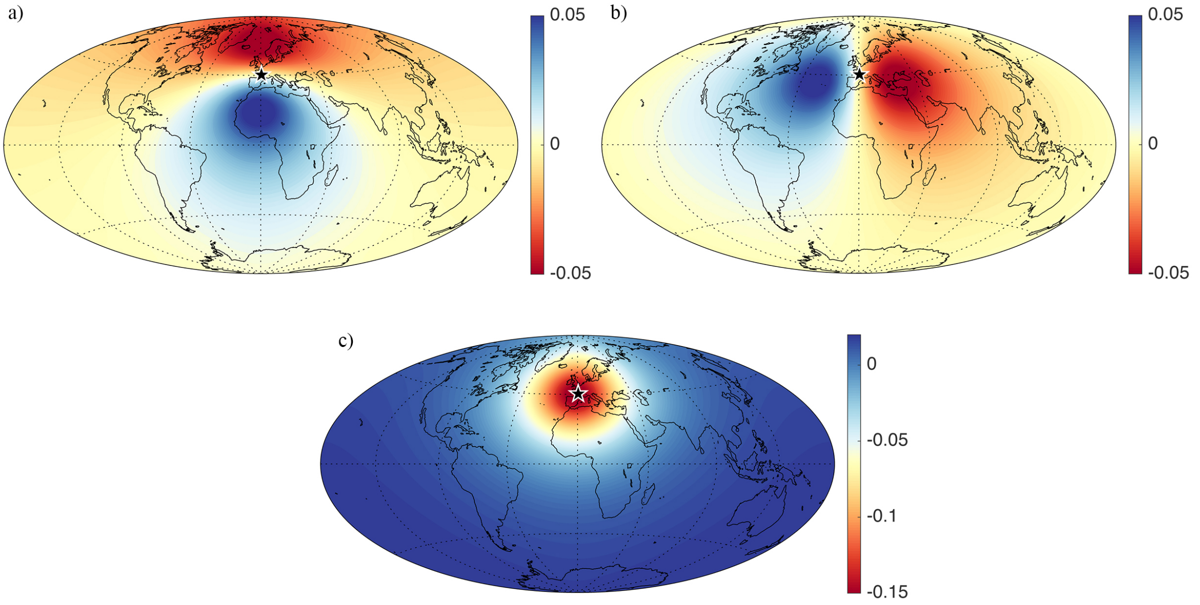


Figure 1.12: Sampling of the core by a) X, b) Y and c) Z geomagnetic surface observations at 45°N and the Greenwich meridian.

As the directions and intensity are nonlinear functions of the magnetic field vector components, they can not be described as a direct convolution of $B_r(\mathbf{r}_c)$ as in Eq. 1.8. Johnson and Constable (1997) proposed linearized sensitivity kernels with respect to changes in $B_r(\mathbf{r}_c)$ that approximate Green functions for declination and inclination. The same formalism can be easily extended to intensity. The corresponding linearized Green functions G_D, G_I and G_F are shown in Figure 1.13, in the case of changes with respect to an axial dipolar core magnetic field. As the archeomagnetic Green functions are linearized, and therefore depend on the underlying magnetic field configuration, they vary with the

---

[8] The radial component is sufficient to derive the magnetic potential within an additive constant (Gubbins and Roberts, 1983).

observation position, as it can be seen in Figure 1.14, for observations at lower latitudes. Such rich sampling of the CMB can allow for a certain compensation of the uneven sampling of the archeomagnetic data at the Earth's surface (Fig. 1.11) and the estimation of global models of the archeomagnetic field.

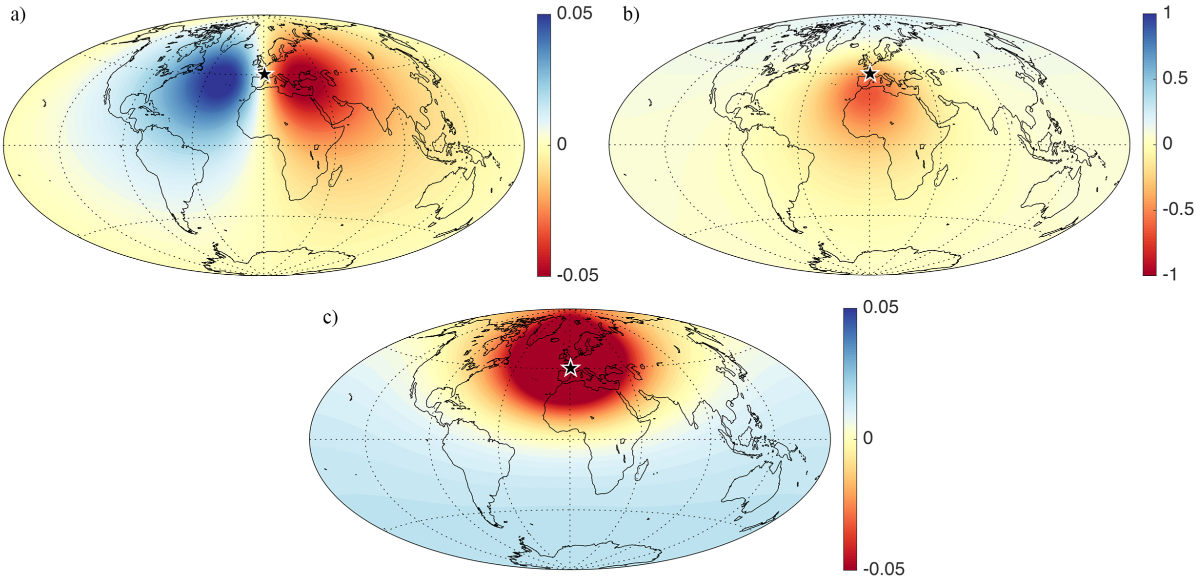


Figure 1.13: Same as Fig. 1.12, but for a) D, b) I and c) F observations.

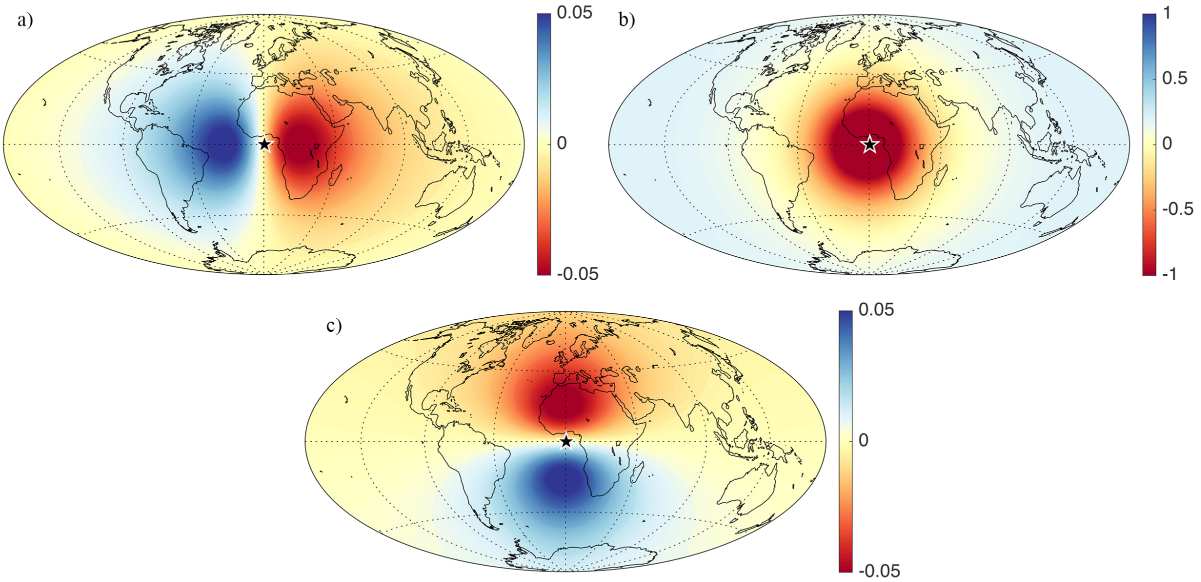


Figure 1.14: Same as Fig. 1.13, but for observations located at the equator.

1.6.5 Models of the archeomagnetic field

Given the general scarcity of archeomagnetic data, some models seek to reconstruct the more basic aspects of the archeomagnetic field. The virtual axial dipole moment (VADM), virtual dipole moment (VDM) and virtual geomagnetic pole (VGP) are examples of extrapolations of the magnetic field assuming the non-dipolar features can be neglected (e.g. Genevey et al., 2008; Valet et al., 2008; Nilsson et al., 2010).

Going further, global reconstructions of the archeomagnetic field are possible using the same methodology described in Subsection 1.5.3, based on the spherical harmonic description. Since the first low resolution model from Hongre et al. (1998), many higher resolution models have been published given the increase in paleo and archeomagnetic data during the last two decades, focusing on different details of the modeling strategy and data handling. One of the most notable global models of the magnetic field are the Continuous Archeomagnetic and Lake Sediment (CALS) series. The series comprises different models depending on duration, data type and year of publication. To cite, Korte et al. (2009) presents the archeomagnetic model ARCH3k.1, the sediment-based model SED3k.1 and archeomagnetic and sediment-based model CALS3k.1 for the last 3 millennia using Geomagia50.V2 database and a truncation of $L = 10$. In line with the CALS models, but within different strategies in data and modeling errors, Licht et al. (2013) produce an ensemble of archeomagnetic models with a lower truncation of $L = 5$. Panovska et al. (2012) and Nilsson et al. (2014, with the archeomagnetic field model pfm9k) have in turn addressed the issue of dating of sedimentary data and their respective uncertainties. The most recent models of the archeomagnetic field correspond to Pavón-Carrasco et al. (2014) SHA.DIF.14k and Panovska et al. (2015).

The different truncation of the archeomagnetic field models points to the uncertainty regarding the resolution allowed to the model due to the sparse and biased spatial distribution of the archeomagnetic dataset. In order to mitigate the uncertainty regarding the smaller and probably unresolved features of the archeomagnetic field, the model strategies generally use prior information from physical approximative constraints that acts as to damp the smaller scale features. However, one may argue that such regularization imposes too strong constraints in the archeomagnetic model solutions.

An alternative methodology can be used to provide additional information from the magnetic field, known generically as data assimilation. In this context, all the information available from the system, namely the geomagnetic data and numerical dynamo simulations, would be used in order to estimate the state of the core generating the observed magnetic field and understand its dynamics. Although data assimilation has already been introduced in geomagnetism (see Fournier et al., 2010, for a review of the state of the art), it is still in its first steps. In particular, little attention has been given to the use of actual geomagnetic observations in the geomagnetic assimilation. Instead, focus has been driven to the assimilation of parameterized models like the ones mentioned in the last paragraph.

1.7 Objectives and outline of the manuscript

The objective of this thesis is to introduce the use of physical information about the geodynamo from its numerical approximation in the archeomagnetic context. In other words, it aims at studying how to use the information from, on the one hand, the archeomagnetic observations and, on the other hand, numerical dynamo models, in a data assimilation framework. From the joint analysis of observations and models, it is possible to estimate not only the observable field, but also the hidden variables of the dynamo system.

This manuscript is organized as follows. In Chapter 2, the essential physics of the geodynamo and its numerical approximation are introduced and discussed. Chapter 3 introduces the main concepts of the data assimilation framework, followed by a discussion of its applicability to the archeomagnetic context. Chapter 4 presents a prelude to archeomagnetic data assimilation, by introducing the use of a dynamo-based prior in archeomagnetic inverse modeling. This study allows for the quantification of the resolution of the archeomagnetic dataset. Chapter 5 consists in a series of experiments on the assimilation of synthetic data closely related to the archeomagnetic dataset scenario. Finally, Chapter 6 summarizes the main conclusions of the present study, their relevance to the community and future perspectives.

Chapter 2

Numerical simulations of the geodynamo

The first efforts in numerical dynamo modeling considered the kinematic dynamo problem, in which different fluid flows were tested for their ability to sustain a magnetic field, aiming to overcome the generality of Cowling's anti-dynamo theorems (Section 1.2). Once the kinematic dynamo problem was shown to succeed for some prescribed flows (e.g. Backus, 1958), it set up a new aspiration, to understand the origin and form of the underlying outer core flow and its interactions with the magnetic field. The dynamo problem had therefore to increase in complexity from two-dimensional mean-field kinematic regimes in order to consider the full set of magnetohydrodynamical interactions.

Full three-dimensional MHD dynamo simulations became proeminent with the self-excited spontaneously reversing dynamo of Glatzmaier and Roberts (1995). Since then, geodynamo modeling has generated an assortment of different dynamo regimes, extensively studied and compared to the known properties of the observed core magnetic field (see for example Christensen and Wicht, 2015, for a review). Although most geodynamo models are based on the same magnetohydrodynamic principle, they may differ in terms of the underlying numerical method, the approach towards compressibility, heat sources, boundary conditions, couplings, choice of control parameters and resolution. In this chapter, the main ideas behind numerical geodynamo modeling are introduced. Attention is given to the discussion of the general morphology and dynamics of the core through simulations, and their implications for retrieving information from the state of the Earth's core.

2.1 Geometry of the core

In order to describe the core state, the following settings are considered in terms of the spherical coordinates (r, θ, ϕ), with r the radius, θ the colatitude and ϕ the longitude. The basic geometry for the geodynamo can be approximated by a spherical shell of inner radius r_i, outer radius r_o, with $D = r_o - r_i$ as the radial length of the outer core. As displayed in Figure 2.1, the shell is set to rotate with an angular velocity $\boldsymbol{\Omega} = \Omega\hat{\mathbf{z}}$, where $\hat{\mathbf{z}}$ is the unitary vector parallel to the rotation axis, assumed invariant through time.

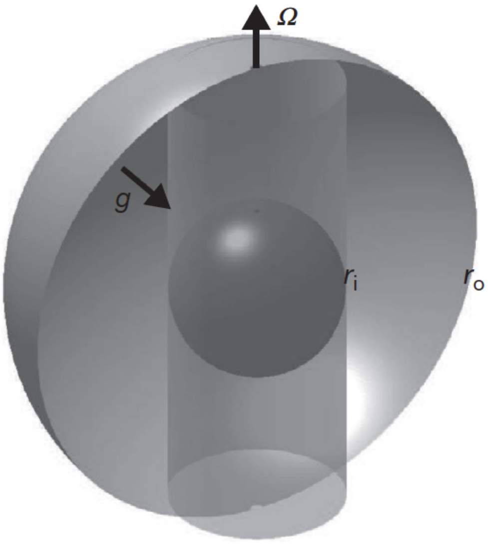


Figure 2.1: Spherical shell representing the basic aspects of a dynamo simulation in the outer core. Also shown is the geometry of the tangent cylinder associated with the existence of the inner core. From Christensen and Wicht (2015).

2.2 Flow in a rotating spherical shell

Let us consider a homogeneous fluid of density ρ filling the spherical shell. The conservation of mass of the core's fluid demands that

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \rho \mathbf{u} = 0, \tag{2.1}$$

where $\mathbf{u}$ is the flow velocity. If the fluid is incompressible, $\rho = \rho_0$ is constant through the shell, so Equation 2.1 imposes a certain restriction on the flow, namely

$$\nabla \cdot \mathbf{u} = 0. \tag{2.2}$$

When dealing with a continuum medium permeated by a velocity field $\mathbf{u}$, it is more practical to study the system's dynamics with respect to the moving flow. The rate of change of a certain property of the moving infinitesimal parcel of the fluid is given by the 'Lagrangian' or material derivative $D/Dt \equiv \{\partial/\partial t + \mathbf{u} \cdot \nabla\}$. The momentum conservation for an infinitesimal newtonian fluid parcel moving with the flow per unit volume is given by

$$\rho \frac{D\mathbf{u}}{Dt} = -\nabla p + \rho \nu \nabla^2 \mathbf{u} + \mathbf{f}_b, \tag{2.3}$$

where p is the pressure, ν the kinematic diffusivity[1], and $\mathbf{f}_b$ volumetric body forces acting on the infinitesimal fluid volume in question.

Under the system rotation with angular velocity $\mathbf{\Omega}$, it is interesting to choose a reference frame rotating along with the shell. In this case, 'fictitious' forces appear in the

---

[1] Counterintuitively, the kinematic viscosity of the outer core is relatively low, being equivalent to that of water at room conditions (Dormy, 1997).

preceding momentum equation from the fact that the reference frame is non-inertial. In consequence, for the fluid parcel at a position $\mathbf{r}$, the momentum equation will be

$$\rho\left[\frac{D\mathbf{u}}{Dt}+2\mathbf{\Omega}\times\mathbf{u}+\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r})\right]=-\nabla p+\rho\nu\nabla^2\mathbf{u}+\mathbf{f}_b. \tag{2.4}$$

The second term on the left-hand side is called the Coriolis force and the third is the centrifugal acceleration. The centrifugal term can be written as the gradient of a scalar and is supposed to be small in the core. The Coriolis force, on the other hand, is expected to have an important role on the geodynamo. Having established the inertial terms, the other relevant forces to the system must be defined. The body forces $\mathbf{f}_b$ include for instance the gravitational force, $\mathbf{f}_g=\rho\nabla\psi$, where ψ is the gravitational potential.

Equation 2.4, together with the corresponding mass conservation equation, are the Navier-Stokes equation for a rotating flow volume. The appropriate boundary conditions are discussed further. The equations do not yet take into account, however, the effects of convection supposed to power the dynamo.

2.2.1 Buoyancy force

Given a difference in temperature difference ΔT between the inner and outer boundaries of the shell, one must consider the variation in density imposed by the expansivity of the fluid and the further effect of buoyancy, by means of the Archimedes force. From PREM (see Section 1.3) the density of the outer core is supposed to vary only by 20% through the whole outer core. Under this condition, one can suppose that the flow can still be treated as an incompressible fluid in the sense that density variations are only important within the buoyancy force. This approximation is known as Boussinesq.

The volumetric buoyancy force is related to variations in density by $\delta\rho g\hat{\mathbf{r}}$, where $g=g(r)$ is the gravity acceleration, assumed to have only a negative radial component, and $\hat{\mathbf{r}}$ the radial unitary vector. Considering that the variations in density depend only on temperature changes, one can write the density fluctuation ρ' as the background density ρ_0 plus a variation in density $\delta\rho=\rho_0\alpha T'$, where T' is temperature fluctuation and α the thermal expansion coefficient. In other words, $\rho'=\rho_0(1-\alpha T')$. The volumetric buoyancy force can then be written as

$$\mathbf{f}_A=\rho_0\alpha g T'\hat{\mathbf{r}}. \tag{2.5}$$

Based on PREM, g is supposed to vary approximately linearly through the fluid outer core. This allows one to write $g\hat{\mathbf{r}}=g_0\mathbf{r}/r_o$, where g_0 is the gravity at the outer boundary radius r_o and $\mathbf{r}$ the radius vector.

2.2.2 Lorentz force

When an electrically conductive fluid is embedded in a magnetic field $\mathbf{B}$, relative motions of the flow and field will generate electric currents in order to restore the magnetic flux, as stated by Lenz' Law. Expressing the electric current density $\mathbf{j}$ in terms of the magnetic field by means of Ampères' law, for which $\mu_0 \mathbf{j} = \nabla \times \mathbf{B}$, the feedback of the magnetic field on the flow, known as Lorentz (or Laplace) force, can be written as

$$\mathbf{f}_L = \mathbf{j} \times \mathbf{B} = \frac{1}{\mu_0}(\nabla \times \mathbf{B}) \times \mathbf{B}, \tag{2.6}$$

where μ_0 is the magnetic permeability of the vacuum.

Rearranging Eq. 2.3 and inserting the buoyancy term in Eq. 2.5 and the Lorentz force in Eq. 2.6 in place of the body forces $\mathbf{f}_b$, we have that

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + 2\boldsymbol{\Omega} \times \mathbf{u} = -\frac{\nabla \pi}{\rho_0} + \alpha g T' \hat{\mathbf{r}} + \frac{1}{\rho_0 \mu_0}(\nabla \times \mathbf{B}) \times \mathbf{B} + \nu \nabla^2 \mathbf{u}, \tag{2.7}$$

where the terms have been rearranged in order to separate the inertial terms (on the left hand side) from the other forces (on the right hand side) and the centrifugal acceleration and the kinematic pressure have been regrouped into the scalar π. This equation represents the momentum conservation in the Boussinesq approximation for an electrically conducting fluid in a rotating spherical shell (a good description on the different geodynamo approximations can be found in Jacobs, 1987; Braginsky and Roberts, 1995).

2.2.3 Mechanical boundary conditions

If we are to consider there is no penetration of the fluid across the boundaries, the radial component of $\mathbf{u}$ must be zero at the inner and outer boundaries. In general, the natural boundary conditions for the flow is that of no-slip, $\mathbf{u}(r_i, \theta, \phi) = \mathbf{u}(r_o, \theta, \phi) = 0$, the fluid adheres to the boundary.

2.3 Convection in the core

The outer core is generally supposed to be under vigorous convection, which implies the fluid to be considered as 'well-mixed'. In this case, entropy is approximated as constant, and the state of the core can be defined as adiabatic (isentropic). The value of temperature, entropy S and pressure can be therefore be defined as fluctuations around the adiabatic state

$$T = T_a + T', \quad S = S_a + S', \quad \rho = \rho_a + \rho', \tag{2.8}$$

where the subscript a denotes adiabaticity. It is assumed that the fluctuations are small compared to the reference state. The adiabat represents the heat propagated by conduction, and fluctuations about this adiabat are responsible for driving the onset of convection.

Given the dependency of the momentum conservation equation on T, it is necessary to describe the essential thermodynamics of the system. The equation of energy can be written in many forms, for example involving the internal energy U and entropy S of the system, as $\mathrm{d}U + \mathrm{p}dV = T\mathrm{d}S$. The conservation of energy can then be written as

$$\rho T \frac{DS}{Dt} = -\nabla \cdot \mathbf{q} + s, \tag{2.9}$$

where $\mathbf{q}$ is the heat flux and s a given heat source (Jacobs, 1987). Using Fourier's law, the heat flux can be written as $\mathbf{q} = -k\nabla T$ for an isotropic fluid, where k is the thermal conductivity. Substituting the equations and neglecting lower order terms,

$$\rho_a T_a \frac{DS}{Dt} = \nabla \cdot (k \nabla T) + s. \tag{2.10}$$

Note that it was also implicitly assumed that the variability of thermal state of the core is negligible on short time scales and that the adiabatic temperature gradient is smooth. In such case where density variations are small, the entropy fluctuations can be approximated in terms of the temperature $S' \sim C_p T'/T_a$. Substituting S' in Eq. 2.10,

$$\frac{\partial T'}{\partial t} + \mathbf{u} \cdot \nabla T' = \kappa \nabla^2 T' + s', \tag{2.11}$$

where $\kappa = k/\rho C_p$ is the thermal diffusivity, and s' represents the heat sources, for example from radioactive decay. The aforementioned approximations are a simplification of the Boussinesq approximation introduced in the previous section. In synthesis, the Boussinesq approximation considers the situation where the convective fluid can be approximated as incompressible, except from the term of the buoyancy force in the momentum equation, and that the variations in density are only dependent on variations in temperature. For a more detailed derivation of the Boussinesq equations and a discussion on its relevance to the modeling of the geodynamo see Anufriev et al. (2005).

2.3.1 Thermal boundary conditions

Thermal boundary conditions are typically set as constant temperatures T_i and T_o at the inner and outer boundaries, respectively. However, in order to better represent the physical aspect of the secular cooling of the core, heat-flux based conditions should be applied. Moreover, seismic tomography of the Earth's mantle (e.g. Masters et al., 1996) has provided an idea of the heterogeneous distribution of heat flux at the top of the core, based on the distribution of higher and lower seismic velocities at the lower mantle. Olson and Christensen (2002) have studied the impact of heterogeneous CMB heat-flux conditions on dynamo simulations, suggesting the control of the mantle on high-latitude flux patches and secular variation localization. Compared to the CMB, however, the inner core boundary is much less constrained and understood.

2.3.2 Codensity

The inner core boundary corresponds to the conditions for solidification of the iron alloy in the core. Such solidification occurs by means of a differential cristalization, whereby the solidification is favored for heavy elements of the alloy, leaving place for a release of lighter elements as well as latent heat into the outer core (Sumita and Bergman, 2015). This process comprises an additional source for convection, in which the light elements enhance buoyancy, due to their lower density (as mentioned in Subsection 1.3.2). The resulting chemical convection can be modeled by a similar equation to Eq. 2.11, but for the fluctuations in light element mass fraction ξ'. Assuming the thermal and chemical diffusivities as equivalent, both temperature and composition fluctuations can be grouped into one single variable called the codensity, or density anomaly C (Braginsky and Roberts, 1995), given by

$$C = \alpha \rho T' + \Delta\rho \xi', \tag{2.12}$$

where $\Delta\rho$ is the density variation between the light and heavy elements in the alloy. The diffusivity is then assumed to be an 'effective' one, comprising the effect of turbulence. The equation for the thermochemical convection is then

$$\frac{\partial C}{\partial t} + \mathbf{u} \cdot \nabla C = \kappa \nabla^2 C + S_{T'/\xi'}. \tag{2.13}$$

The term $S_{T'/\xi'}$ represents the sources and sinks of the codensity in order to assure the mass conservation of the system (Aubert et al., 2009).

2.3.3 Compositional boundary conditions

The boundary conditions at the CMB become more complex for the codensity, for despite the heat being able to cross the boundary, the fluid is not allowed to penetrate the interfaces. This is generally described by prescribing a heat flux amplitude and pattern along the CMB. At the ICB, however, heat and light element flows are thought to be coupled due to the continuous crystallization of the inner core boundary.

Seismological evidence support anisotropies in the inner core properties (see Souriau and Calvet, 2015, for a recent review), which are supposed to be connected with a translational mode of the inner core, generating melting on the eastern inner core hemisphere and solidification on the western one (Alboussiere et al., 2010; Monnereau et al., 2010). As the inner core translation is thought to occur on geological time-scales, the codensity flux at the ICB is generally considered constant in numerical simulations.

2.4 Electromagnetic induction

The magnetic field is essentially solenoidal, and can be written as

$$\nabla \cdot \mathbf{B} = 0. \tag{2.14}$$

The magnetic field sources in the electrodynamics context are electric currents, as acknowledged by Ampère's law in the MHD approximation (the case when $u \ll c$, where c is the speed of light, neglecting the displacement currents), writing

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{j}, \tag{2.15}$$

already mentioned in Section 2.2.2. The electric currents, by Ohms law, can be written in terms of $\mathbf{B}$ and the electric field $\mathbf{E}$ as

$$\mathbf{j} = \sigma(\mathbf{E} + \mathbf{u} \times \mathbf{B}), \tag{2.16}$$

where σ is the electrical conductivity of the medium. The electric field can also be generated by variations in the magnetic field by means of the electromagnetic induction, described by Faraday's law

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{2.17}$$

Substituting 2.16 in Equation 2.15 and using the identity in 2.17, one is able to write the induction equation

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) - \nabla \times (\eta \nabla \times \mathbf{B}), \tag{2.18}$$

where $\eta = 1/\sigma\mu_0$ is the magnetic diffusivity. The electrical resistivity of the iron alloy is responsible for the diffusivity of the magnetic field, supposed to be many orders of magnitude higher than the viscous and thermal diffusivities (Dormy, 1997). Considering an homogeneous magnetic diffusivity, the induction equation can be simplified to

$$\frac{\partial \mathbf{B}}{\partial t} = \nabla \times (\mathbf{u} \times \mathbf{B}) + \eta \nabla^2 \mathbf{B}. \tag{2.19}$$

The induction equation relates the variability of the magnetic field to the phenomena of advection of the field by the flow (first term of the right hand side of the equation) and magnetic diffusion (second term on the right hand side), discussed in Section 1.4.

2.4.1 Magnetic boundary conditions

As mentioned before, for simplicity, the mantle is often considered as an electric insulator. Under such conditions, in order to match a self-excited dynamo, the magnetic field is generally set to meet the conditions of a potential field at the CMB. However, the lower mantle is thought to have a moderate conductive layer at its base. In such case, electrical currents would be able to trespass from the core to the lower mantle, generating large scale Lorentz forces which would results in important torques between the core and mantle. Such torques are suspected to be relevant to justify decadal changes in the length-of-day (LOD) of the Earth (Holme, 1998).

In a similar way, the inner core is sometimes presumed to be an insulator. Some studies, however, allow for viscous and magnetic torques between inner and outer core which generate a rotation of the inner core. Numerical models of the dynamo show that such magnetic and viscous couplings tend to entrain the inner core eastwards at different rates either steadily or in an oscillatory way (Aubert and Dumberry, 2011). The eastward rotation of the inner core generates the so-called thermochemical winds near the ICB, creating cyclonic movements inside the tangent cylinder. Evidence of inner core differential rotation (with respect to the mantle reference frame, also called super-rotation) from seismic observations supports a small but non-zero mean eastward rotation with greater decadal fluctuations around this mean (Tkalčić et al., 2013).

2.5 Geodynamo equations and parameters

The main physical ingredients of the geodynamo in the Boussinesq approximation are given by the equations for mass conservation 2.2, momentum conservation 2.7, codensity transport 2.13, solenoidal magnetic field 2.14 and magnetic induction 2.23. In order not to deal directly with each variable dimension involved in a simulation, it is practical to resort to a non-dimensional form of the equations. This is done by rescaling the dimensional variables in terms of the most relevant values in the problem. Taking, for example, the thickness of the spherical shell D as the typical length scale, Ω^{-1} as the typical time scale and $(\sigma\mu_0)^{1/2}\Omega D$ as the magnetic induction, the former equations can be rewritten as

$$\begin{aligned}
\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} + 2\hat{\mathbf{z}}\times\mathbf{u} &= -\nabla\pi + \mathrm{Ra}\frac{\mathbf{r}}{r_0}C + (\nabla\times\mathbf{B})\times\mathbf{B} + \mathrm{E}\nabla^2\mathbf{u}, && (2.20)\\
\nabla\cdot\mathbf{u} &= 0, && (2.21)\\
\frac{\partial C}{\partial t} + \mathbf{u}\cdot\nabla C &= \frac{\mathrm{E}}{\mathrm{Pr}}\nabla^2 C + S', && (2.22)\\
\frac{\partial \mathbf{B}}{\partial t} &= \nabla\times(\mathbf{u}\times\mathbf{B}) + \frac{\mathrm{E}}{\mathrm{Pm}}\nabla^2\mathbf{B}, && (2.23)\\
\nabla\cdot\mathbf{B} &= 0. && (2.24)
\end{aligned}$$

The dimensionality of the variables is replaced by four dimensionless parameters E, Ra, Pr and Pm which represent ratios between the different time scales of the system, which are summarized in Table 2.1. The Ekman number E

$$\mathrm{E} = \frac{\nu}{\Omega D^2} \qquad (2.25)$$

represents the ratio between the rotational and viscous time scales, or in an equivalent way, the ratio of the viscous to Coriolis forces. In the Earth's core, this number is very small, of the order of 10^{-15}, giving an idea of the striking difference in time scales involved on the geodynamo. It is important to note that although viscous effects are of minor importance on the bulk of the main flow, they play a key role near the boundaries, like in the Ekman layer. A very small E indicates very small Ekman layers ($\delta_\mathrm{E} \sim \mathrm{E}^{1/2}D$), which for the Earth's core would imply a thickness of around 10 cm. Considering the lower bound estimation by Dormy (1997) of 10 grid-points in the layer, the amount of resolution needed to account for a dynamo numerical simulation based on a realistic E is far from being computationally feasible (the Ekman number is typically 10^{-5} in high resolution numerical simulations, which either means a very slow rotation or very high viscosity compared to that of the core). Even by substituting the 'molecular' viscosity ν by an 'effective' viscosity, parameterizing the role of turbulence, the Ekman number would still be overestimated by a factor of 10^4 (Glatzmaier, 2002).

The mass anomaly flux Rayleigh number Ra (Aubert et al., 2009) is given by

$$\mathrm{Ra} = \frac{g_o f}{\rho\Omega^3 D^2}, \qquad (2.26)$$

where f is the mass anomaly flux at the ICB, represents the vigor of the thermochemical convection. The other two parameters, the Prandtl number

$$\mathrm{Pr} = \frac{\nu}{\kappa}, \qquad (2.27)$$

and the magnetic Prandtl number

$$\mathrm{Pm} = \frac{\nu}{\eta}, \qquad (2.28)$$

relate to the ratio between the thermal τ_κ and ohmic dissipation time scales τ_η and viscous diffusion time τ_ν, respectively. Due to the similarity between τ_κ and τ_ν, $\mathrm{Pr} \approx 1$. They are, however, significantly bigger than τ_η in the core, resulting in $\mathrm{Pm} \ll 1$ (Dormy and Le Mouël, 2008).

Table 2.1: Dimensionless numbers representing the control and diagnostic parameters relevant to the geodynamo equations, their range estimated in the core and their typical value in geodynamo simulations.

Symbol	Name	Definition	Core	Simulations
E	Ekman	$\nu/\Omega D^2$	10^{-15}	$10^{-3} - 10^{-7}$
Ra	Rayleigh	$g_o f/\rho\Omega^3 D^2$	10^{-12}	$10^{-4} - 10^{-6}$
Pr	Prandtl	ν/κ	10^{-1}	1
Pm	Magnetic Prandtl	ν/η	10^{-6}	0.1–1
Re	Reynolds	$u_0 D/\nu$	10^{9}	10^{2}
Rm	Magnetic Reynolds	$u_0 D/\eta$	10^{3}	$10^{2} - 10^{3}$

The dynamo models can also be characterized by their output parameters, representing a diagnostic of the solution regime. Taking the root-mean-squared flow field through the core, u_0, one can calculate the Reynolds number

$$\mathrm{Re} = \frac{u_0 D}{\nu} \tag{2.29}$$

and the magnetic Reynolds number

$$\mathrm{Rm} = \frac{u_0 D}{\eta} \tag{2.30}$$

which are diagnostics of the core flow regime. Re is estimated to be very high in the core, which gives an idea of the level of turbulence ($\mathrm{Re} \gg 100$). Rm is of particular importance for the analysis of the self-sustainable dynamo onset, generally corresponding to an $\mathrm{Rm} \approx 50$.

2.5.1 Poloidal-toroidal and spectral decomposition

The solenoidal form of both the magnetic and flow fields (Eqs. 2.21 and 2.24), allows for its description in terms of its poloidal and toroidal parts

$$\begin{aligned} \mathbf{B}(r,\theta,\phi) &= \nabla\times\nabla\times[B_p(r,\theta,\phi)\mathbf{r}] + \nabla\times[B_t(r,\theta,\phi)\mathbf{r}], & (2.31)\\ \mathbf{u}(r,\theta,\phi) &= \nabla\times\nabla\times[u_p(r,\theta,\phi)\mathbf{r}] + \nabla\times[u_t(r,\theta,\phi)\mathbf{r}], & (2.32) \end{aligned}$$

where the subscript p and t correspond to the poloidal and toroidal components of the fields (Dormy, 1997). Since the temperature/codensity field is scalar, there is no sense in applying such a decomposition. Due to the dimension reduction of the field representation, the poloidal-toroidal decomposition is regularly used in numerical simulations of the geodynamo. The field components, as well as the codensity field, can be further expanded in a spherical harmonic basis. For example, the poloidal and toroidal scalars of the magnetic field at a given radius r can be expressed as

$$B_p(r,\theta,\phi) = \sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell} \mathcal{P}_\ell^m(r) Y_\ell^m(\theta,\phi) \tag{2.33}$$

$$B_t(r,\theta,\phi) = \sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell} \mathcal{T}_\ell^m(r) Y_\ell^m(\theta,\phi) \tag{2.34}$$

where $Y_\ell^m = e^{im\phi} P_\ell^m(\cos\theta)$ are the fully normalized surface spherical harmonic functions, and P_ℓ^m are the associated Legendre functions. The same expansion holds for the flow fields and the codensity. The geodynamo Boussinesq equations 2.20–2.24 can be therefore be put in terms of the poloidal-toroidal spectral decomposition and solved numerically.

2.6 Numerical simulations

Given the complex nature of the geodynamo model described in the previous section, the system's evolution is simulated numerically. Different numerical methods can be used to simulate the dynamo system (see Christensen and Wicht, 2015, for a review), amongst which the spectral methods are the most popular. One example is the Parody code (Dormy et al., 1998; Aubert et al., 2008), used throughout this work.

Parody is based on a spherical harmonic decomposition of the flow, magnetic and temperature/codensity fields for the lateral directions and a second-order finite-difference scheme on the radial direction. The radial grid is non-uniform, being compressed near the boundaries to ensure an accurate resolution of the Ekman layers. The code is written in fortran and it is parallelized for the radial part within MPI and by Open-MP on the calculations involving the colatitude in physical space. Parody participated in the geodynamo benchmark from Christensen et al. (2001).

2.6.1 Boundary conditions

As mentioned above, the dynamo parameters in numerical simulations are not representative of that of the Earth's core. The smallness of the core's viscous diffusivity, for example, suggests an important role of the small spatial and temporal flow scales on the core dynamics. Numerical simulations are not capable of resolving such small-scale turbulent eddies, however, which are thought to be specially important in the domain of boundary layers. In particular, the too high viscosity from the simulations ends up producing very deep Ekman layers, which tend to generate unrealistically strong torques in between the boundaries. In order to avoid such strong torques, the Ekman layers are in some cases artificially suppressed by imposing the condition of vanishing shear in the boundaries (Kuang and Bloxham, 1997). This condition is called 'stress-free' and substitutes the physical mechanic 'no-slip' condition presented in Section 2.2.

2.6.2 Earth-likeness

Although geodynamo simulations lie far away from the true control parameters, many have been considered as being Earth-like with respect to the observable magnetic field. Christensen et al. (2010) introduced a way of quantifying such Earth compliance, by means of the parameter χ^2. This parameter is defined in terms of dipolarity, symmetry, zonality and flux concentration of the magnetic field at the CMB. By systematically diagnosing a large number of simulations, they found a specific wedge of compliance, bounded by $\mathrm{Em} < 10^{-4}$, where Em is the magnetic Ekman number $\mathrm{Em} = \mathrm{E}/\mathrm{Pm}$, and $\mathrm{Rm} > 10^2$. Simulations lying outside this wedge tend to be either too dipolar for lower Rm or too multipolar for higher Em. Such results suggests a compensation for the unrealistic Ekman number used in the numerical simulations by a high magnetic Prandtl number. One could assume the transition from the Earth-like simulations parameter range to the core conditions to be continuous, meaning that the underlying dynamo mechanism would be in line with that of the real geodynamo. This line of thought is however highly controversial, for either the transition of parameters might imply a disconnection in the dynamo regime or the Earth-likeness condition being shared by different possible underlying dynamo configurations.

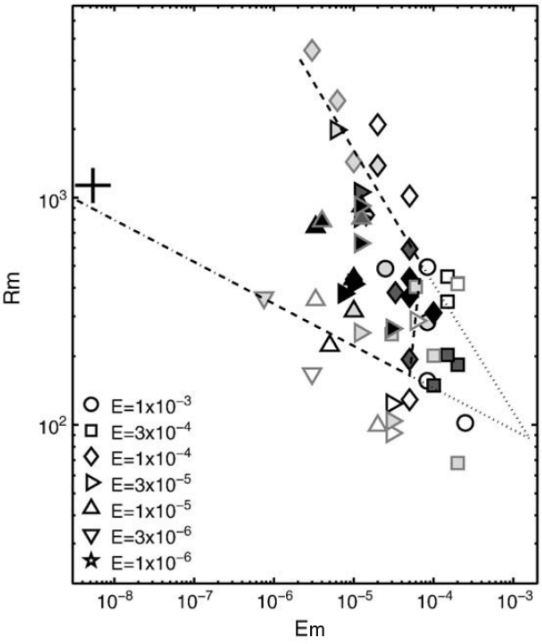


Figure 2.2: Diagram of Earth-like numerical simulations from Christensen et al. (2010) as function of the Reynolds and Ekman magnetic numbers and comparison with core conditions. Earth-like simulations are shown as shaded symbols and the + symbolizes the core situation in the parameter space.

Despite Christensen et al. (2010) having revealed the basic conditions for Earth-like dynamos based on the magnetic field average morphology, no systematic study has been made to analyze the Earth-likeness of the magnetic field secular variation. Additionally, a study of the Earth-likeness of reversing dynamos is still needed.

2.7 Time and magnetic field rescaling

Application of dynamo simulations in data assimilation require the consideration of the magnetic field variables with the proper dimensions to match that of the observed field. The formalism presented in Section 2.5 suggests a direct substitution of the scaling values in order to redimensionalize the dynamo system. However, as dynamo simulations are not generally holding to Earth-like control parameters, their rescaling for comparison with the observable magnetic field is not obvious, and somewhat arbitrary.

Lhuillier et al. (2011b) investigated the typical time scales of the magnetic power spectra in modern and historical models of the magnetic field at the CMB as well as those of representative dynamo simulations. They showed that a simple scaling law of

$$\tau_\ell = \frac{\tau_{SV}}{\ell} \tag{2.35}$$

satisfies well the correlation time scales of the observed magnetic spectrum.[2] For the Earth, the secular variation time scale is estimated from satellite observations to be around 415 years. The calibration of the time axis of any numerical geodynamo simulations can then easily be done by multiplying the simulation time by the ratio between the τ_{SV} from the core (expressed in years) to that estimated from the inverse linear fit to the dynamo model (non-dimensional). In synthesis, it would correspond to calibrating the non-dimensional time t_{nd} of the simulation as

$$t' = \frac{\tau_{SV}^{\oplus}}{\tau_{SV}^{sim}} \times t_{nd}, \tag{2.36}$$

where $\tau_{SV}^{\oplus}$ and τ_{SV}^{sim} are the secular variation characteristic time of the Earth's core and of the numerical simulation in question, respectively. Since $\tau_{SV} \propto \mathrm{Rm}^{-1}$, this also corresponds to the ratio between the magnetic Reynolds numbers, and therefore to matching the magnetic turbulence level of the simulation to that of the Earth (Lhuillier et al., 2011a; Davies and Constable, 2014).

The magnetic field rescaling can be done by using a scaling law connecting the magnetic field to the dynamo convective power (Christensen and Aubert, 2006). An example of this rescaling is given by Aubert (2013). Otherwise, an easier calibration of the magnetic field can be done by comparing the mean spatial spectrum of the core's magnetic field and that of numerical simulations, or even simpler, the axial dipole intensity.

2.8 Examples of dynamo models

A simple geodynamo model can be built by considering the thermochemical convection-based dynamo, following Equations 2.20–2.24, which will be here called STD (for 'standard') model throughout. The basic configuration for STD consists on 'no-slip' at the ICB and 'stress-free' at the CMB mechanical boundary conditions, a conductive inner core and an insulating lower mantle. A homogeneous heat flow is selected for the CMB, and a homogeneous codensity flow at the ICB. The model then allows for a magnetic

---

[2] The dipole is in general not considered in the scaling law.

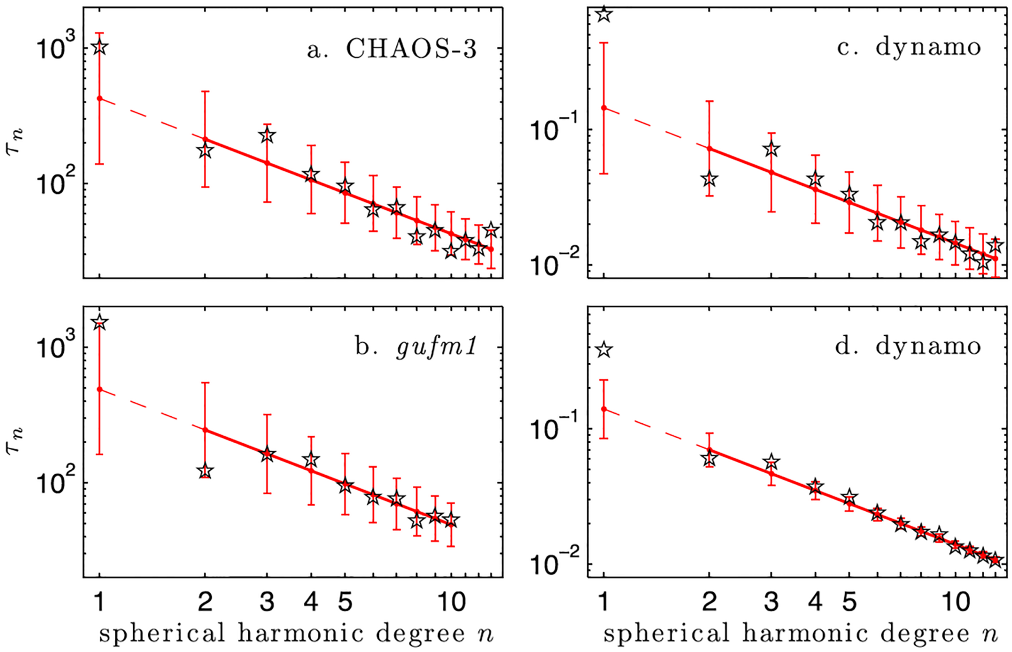


Figure 2.3: Secular variation time scales from the Earth and from numerical simulations, fitted by the inverse-linear law in Eq. 2.36, for a) the CHAOS-3 model (reference), b) the historical *gufm1* model (Jackson et al., 2000), c) the instantaneous spectrum for a dynamo simulation similar to STD, and d) to the mean spectrum over a long run of the simulation. From Lhuillier et al. (2011b).

coupling of the inner core with the outer core and viscous coupling at the CMB. The native truncation of the model is $L_{max} = 133$, which gives 480×200 horizontal grid resolution, with 160 levels on the radial grid. Rescaling of STD time is made by the secular variation time scale described in Eq. 2.36. The magnetic field is calibrated by means of the adjustment of the mean axial dipole from STD by the mean axial dipole from an archeomagnetic field model (as will be explained in Chapter 4). Further information on the STD model is displayed in Table 2.2.

Table 2.2: Dynamo simulations mentioned throughout this manuscript. The codes STD means 'standard' dynamo simulation, while CE is the 'Coupled Earth' model from Aubert et al. (2013). CEL is a 'lighter' version, in terms of resolution, of the CE model. The Prandtl number is not shown, for all models have an underlying Pr = 1. τ_c and β_c are respectively the time and magnetic field rescaling factors.

Code	E	Ra	Pm	Re	Rm	τ_c(yr)	β_c(μT)	χ^2	L_{max}	Δt
STD	3×10^{-5}	2.7×10^{-5}	2.5	370	926	42347	1033.0	2.8	133	1×10^{-7}
CE	3×10^{-5}	2.7×10^{-5}	2.5	377	943	56113	850.6	1.3	133	1×10^{-7}
CEL	1×10^{-4}	5×10^{-6}	10	55.5	555	9545	544.36	3.3	85	2×10^{-6}

Figure 2.4 shows a comparison between the radial magnetic field at the CMB from historical model *gufm1* (Jackson et al., 2000) for the year 1990 and the STD dynamo model at the *gufm1* truncation and the native model resolution. One can easily spot common features between *gufm1* and STD (Figs. 2.4a and 2.4b, respectively), like the mainly dipolar character of the magnetic field. There is also the presence of high latitude intense magnetic flux patches underneath what would be the intersection between the

tangent cylinder and the CMB, as well as the low magnetic field intensity around the poles. Those features correspond to the cyclonic dynamics of upwellings and downwellings on and around the tangent cylinders. Fig. 2.4c reveals the great amount of detail underneath the magnetic crustal curtain (recall Section 1.5).

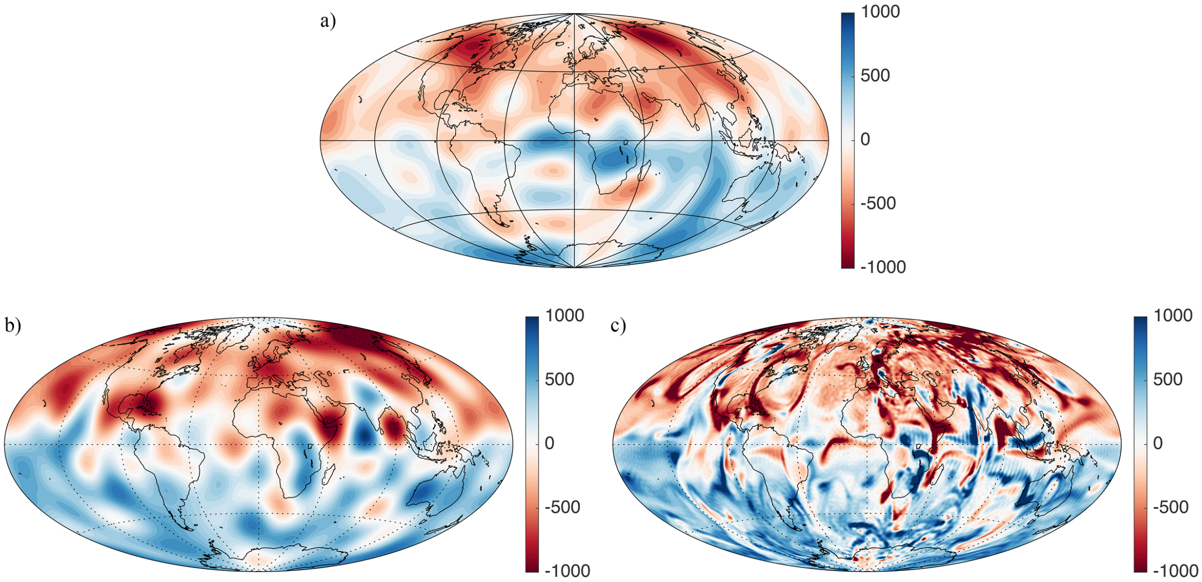


Figure 2.4: Radial magnetic field at the CMB given by a) *gufm1* (Jackson et al., 2000) for the year 1990, and the STD numerical simulation of the geodynamo b) limited to *gufm1* truncation ($L = 14$) and c) for the native truncation of degree 133. All quantities are shown in μT.

Although reasonably compliant relative to the morphological Earth-likeness parameter χ^2 of Christensen et al. (2010) (see Table 2.2), the STD model displays a not very Earth-like secular variation pattern. In general, geodynamo simulations have a tendency to display a greater secular variation at high latitudes, without a preferential direction (eastward or westward). Such a pattern strongly contrasts with the secular variation registered by *gufm1* (Jackson et al., 2000), which is concentrated at low latitudes preferentially on the Atlantic hemisphere (Finlay and Jackson, 2003). This disparity in secular variation can be seen in Figure 2.5.

The recent dynamo model from Aubert et al. (2013) manages to retrieve an Earth-like secular variation pattern by incorporating two complexities related to the inner core and lower mantle. The first one is a gravitational coupling in between the inner core and lower mantle, considering their heterogeneous mass anomaly distribution. Since the thermal winds tend to entrain the inner core to the east, the gravitational torque acts to align the mantle with the inner core. The inner core and mantle being entrained to the east, the conservation of angular momentum of the system consists in a westward rotation of the outer core at low latitudes near the CMB, creating a large-scale sheet-like gyre (Aubert et al., 2013). Secondly, it uses an asymmetric mass anomaly release at the top of the inner core, representing a differential inner core growth, as proposed by seismic studies (Monnereau et al., 2010, and references therein). The asymmetric buoyancy pattern generated by such inner boundary conditions acts to distort the gyre, creating higher secular variation underneath specific zones on the outer core (Aubert et al., 2013). The secular variation resulting for the coupling of those features match

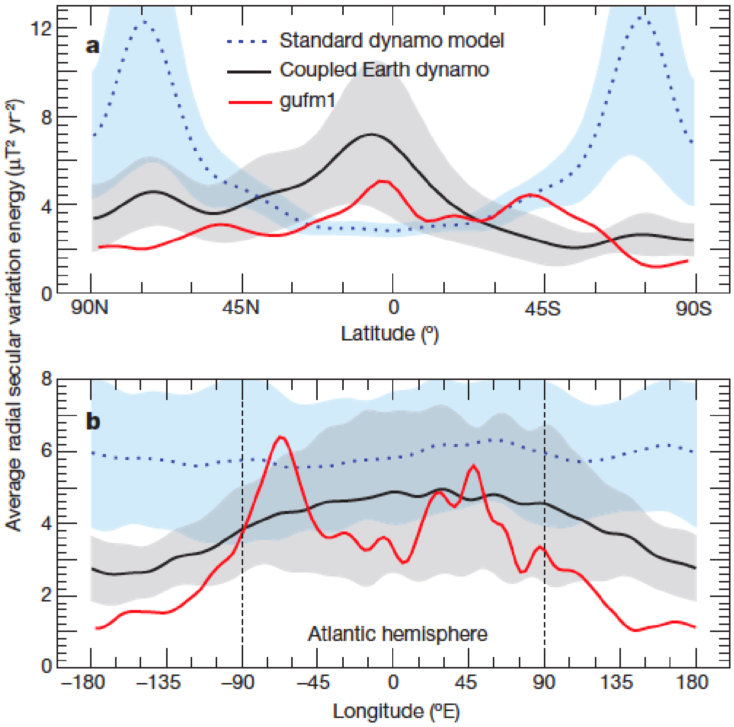


Figure 2.5: Secular variation localization in a) latitude and b) longitude for a similar model as STD, the CE dynamo model and the historical model *gufm1* (Jackson et al., 2000). From Aubert et al. (2013).

remarkably well that shown by *gufm1*, as can be seen in Fig. 2.5. The model, coined as 'Coupled Earth', will be abbreviated by CE throughout.

Envisaging the applicability of an Earth-like dynamo model in a data assimilation scheme, we are particularly interested in the use of a model with lower resolution, in order to speed-up the calculations. We introduce such model as the CEL, and its specifications are given in Table 2.2.

2.9 The e-folding time and predictability

Due to the important non-linearities in the governing dynamo equations and the expected turbulent level at its smaller spatial scales, the geodynamo is a chaotic system (Ryan and Sarson, 2008). A clear evidence of its chaotic nature relies on the non-periodic character of the dipole reversals. A chaotic system can be essentially described by its sensitivity to perturbations. That is, given an initially small perturbation, the system will evolve in such a way as to amplify the perturbation until a time when the perturbed and unperturbed solutions diverge. The characteristic time for the exponential amplification of an initial perturbation ε_0 at the time t_0 is given by the e-folding time τ_e, is connected to the normalized difference between the solutions Δ by

$$\Delta(t) = \varepsilon_0 e^{(t-t_0)/\tau_e}. \tag{2.37}$$

For the geodynamo, the e-folding time is dependent on the advective regime of the dynamo (Hulot et al., 2010), in the same way as τ_{SV}.Hulot et al. (2010) found an asymptotic limit of $\tau_e = 0.05\tau_{SV}$, with a corresponding e-folding time of $\tau_e \approx 30$ yr for the Earth's dynamo. From Equation 2.37, we have that at the time interval after which the perturbed solution diverges from the initial one ($\Delta = 1$), substituted by $\tau_f = t - t_0$, is given by

$$\tau_f = -\tau_e \log \varepsilon_0. \tag{2.38}$$

The characteristic time τ_f is often called the predictability limit of a system, and it is specially important when it comes to estimating the system's state, by means of data assimilation techniques, for example. In order to visualize the e-folding time of a dynamo model, Figure 2.6a and 2.6b show the axial dipole and equatorial dipole of the poloidal magnetic field at the top of the core from an ensemble of CEL solutions, departing from small perturbations in the initial axial dipole conditions. It is interesting to see how the perturbations, initially affecting only the axial dipole, are propagated to the equatorial dipole, leading to the divergence of the once very close solutions of the ensemble. After roughly 1000 years, the models are seen to diverge from each other, showing independent behaviors. Such divergence can be monitored through the spread of the ensemble, shown in Fig. 2.6c.

The predictability limit shows that even in a hypothetical situation where the underlying physics of the dynamo is perfectly known, important errors would arise by the discretization of the model, as well as the impossibility of perfectly observing the whole system. For the geodynamo, this is a specially difficult quantity to estimate, for the observable part of the dynamo system consist only of the largest scales of the magnetic field at the Earth's surface. It can be estimated, however, based on the e-folding time from Hulot et al. (2010). In an archeomagnetic context, taking into account solely the error level of 1 μT (Section 1.6) as the main uncertainty entering ε_0 based on a magnetic field of around 40 μT, this predictability limit would be of around 100 years. This limit would be considerably reduced if one considers the uneven spatial distribution of archeomagnetic data as well, which could amount to reducing the predictability limit at least by a half. Notwithstanding such rather pessimistic scenario, it can be mitigated by the data assimilation techniques, specifically by applying successive corrections from the observations at different times, bringing the dynamo system closer to its true state. A more elaborate discussion on data assimilation and its applicability to the dynamo problem is provided in the next chapter.

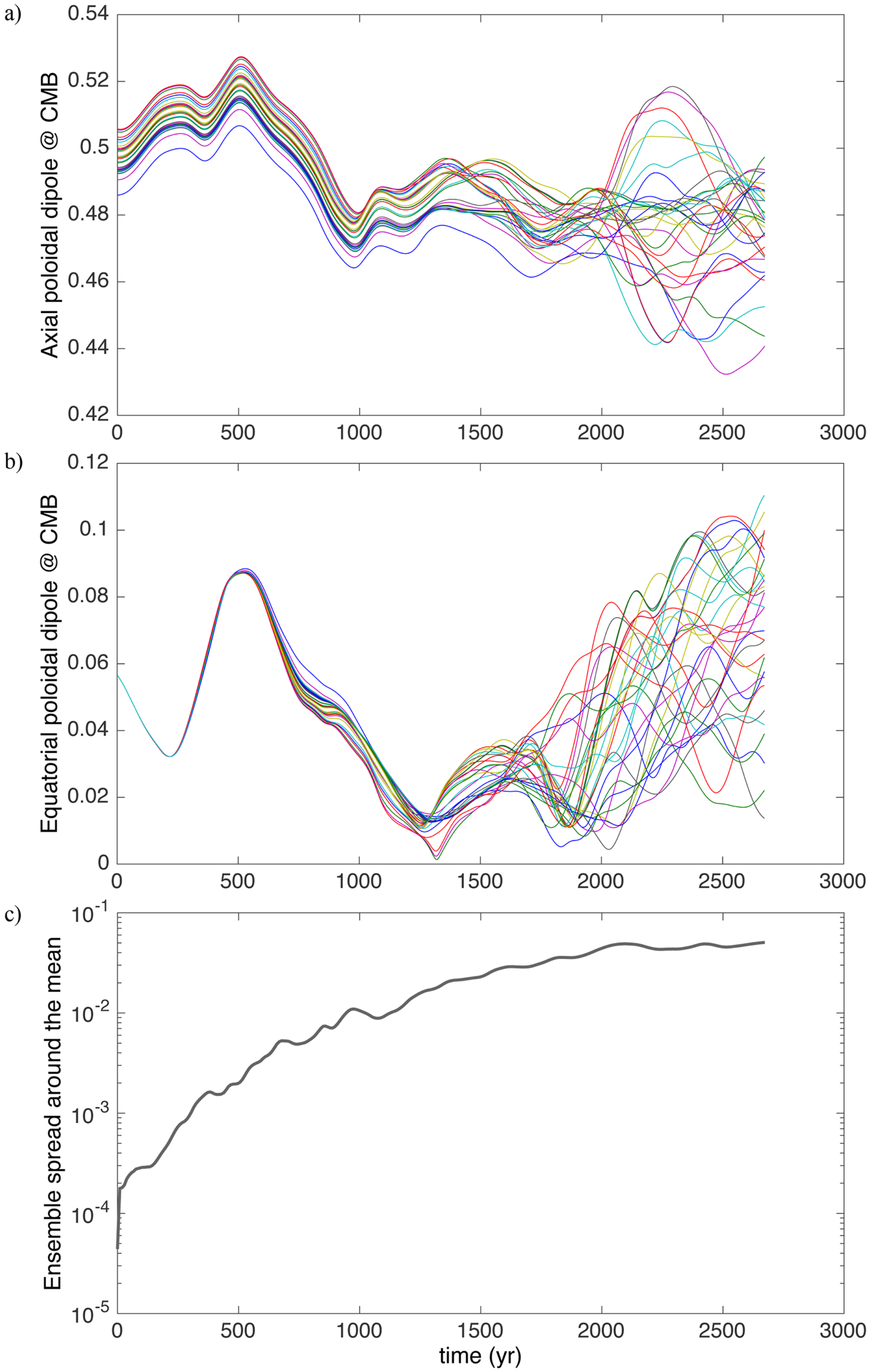


Figure 2.6: Ensemble of CEL models departing from small differences in the initial condition of the axial dipole of the poloidal magnetic field. The evolution in time of the ensemble is shown in terms of the a) axial and b) equatorial dipole of the poloidal magnetic field at the top of the core. Shown in c) is the ensemble spread of the full models around the ensemble mean.

Chapter 3

Overview of data assimilation

Data assimilation refers to a diverse group of methodologies applied to even more diverse fields. In its fundamentals, it involves using all the available information of a physical system in order to estimate its past, present or future state. Traditionally, such information comes on one hand from approximations of the physical system given by a numerical model, and on the other hand from sparse observations of such system. Data assimilation differs, therefore, from traditional inverse problems through its use of prior information strongly based on numerical models and the inclusion of physical based time-dependency.

Mainly known by its connection to meteorology in general and Numerical Weather Prediction (NWP) in particular (see Kalnay, 2003, for an extensive review), data assimilation has begun to be applied to many different geophysical areas in the last decade (glaciology, landscape formation, seismology, etc). Inside its general purpose of estimating the state of a system, there lies a variety of objectives in data assimilation: study its dynamical aspects, retrieve its hidden variables, estimate the underlying parameters of the system and acquire the ability to hindcast and forecast a model.

This chapter presents a brief overview of the basics of data assimilation, focusing on the formalism given by sequential algorithms. Special attention is given to its applicability in geophysics, and geomagnetism in particular.

3.1 State and observations of a system

Consider χ as the continuous state (in space and time) of the physical system we aim at estimating. The system is dynamic, and it evolves in accordance with the physical laws in the form of (non-linear) partial differential equations, here synthesized by the functional g

$$\frac{\partial \chi}{\partial t} = g(\chi). \tag{3.1}$$

Due to the impossibility in describing such continuous complex fields, one must resort to its temporal discretization in terms of $\mathpzc{g}$

$$\mathpzc{x}_{i+1} = \mathpzc{g}(\mathpzc{x}_i), \tag{3.2}$$

where the subscript i represents time, and spatial discretization, by projecting the continuous state into its discrete form $\mathbf{x}^t$ with the operator $\mathbf{\Pi}$ as

$$\mathbf{x}^t = \mathbf{\Pi}(\chi). \tag{3.3}$$

The superscript t implies the state is the discretized 'true' state. Since the model is dynamic, the discretized propagation of such state is described by

$$\mathbf{x}^t_{i+1} = \mathbf{\Pi}[g(\chi_i)]. \tag{3.4}$$

In the discrete case (Cohn, 1997), this would be equivalent to propagating the state by a numerical model $\mathbf{M}$ connecting the states at time i and $i+1$, following

$$\mathbf{x}^t_{i+1} = \mathbf{M}_{i,i+1}(\mathbf{x}^t_i) + \boldsymbol{\eta}_{i,i+1}. \tag{3.5}$$

Here $\boldsymbol{\eta}_{i,i+1}$ is the model error

$$\boldsymbol{\eta}_{i,i+1} = \mathbf{x}^t_{i+1} - \mathbf{M}_{i,i+1}(\mathbf{x}^t_i) \tag{3.6}$$

comprising different sources, like the error arising from the discretization of the model and state and from misrepresentations of the physical processes (e.g. approximations and unresolved scales, since g is not perfectly known). The model errors can be associated with an error covariance matrix $\mathbf{Q}_{i+1}$.

We can define the observations of the system χ as a mapping by the functional h

$$y = h(\chi). \tag{3.7}$$

Whenever the system is observed, however, measurement errors $\boldsymbol{\epsilon}^m$ contaminate the real observation $h(\chi)$

$$\mathbf{y}^o = h(\chi) + \boldsymbol{\epsilon}^m. \tag{3.8}$$

Also, errors in the formulation of how the system is observed can compromise the knowledge of h. An observation of the discretized state ignoring the measurement errors can be written as

$$\mathbf{y}^o = \mathbf{H}\mathbf{\Pi}(\chi), \tag{3.9}$$

where $\mathbf{H}$ is the discretized observation operator. The error regarding the representativeness of the observation can be described by

$$\boldsymbol{\epsilon}^r = h(\chi) - \mathbf{H}[\mathbf{\Pi}(\chi)]. \tag{3.10}$$

The combination of both sources of error are incorporated into the observation error $\boldsymbol{\epsilon}^o = \boldsymbol{\epsilon}^m + \boldsymbol{\epsilon}^r$, which allows us to redefine the observation vector as

$$\mathbf{y}^o = \mathbf{H}\mathbf{x}^t + \boldsymbol{\epsilon}^o. \tag{3.11}$$

The observation errors can be written in terms of the observation error covariance matrix $\mathbf{R}$.

The model and observation errors can be considered as stochastic perturbations (Cohn, 1997) in Eq. 3.5 and Eq. 3.11, respectively. Such supposition supports the interpretation of both model state and observations as random variables, allowing for a probabilistic description of the problem.

3.2 Stochastic estimation

Given the attributed stochasticity of the model state and observation, one can define their respective probabilities. The probability density function (pdf) p characterizing the multivariate model state is assumed to be Gaussian, given by

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}(\det\mathbf{P})^{1/2}} \exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{P}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right]\right\}, \tag{3.12}$$

where $\boldsymbol{\mu}$ is the expectation value, $\mathbf{P}$ is the covariance matrix related to $\mathbf{x}$ and det is the determinant. This is summarized by writting that the vector variables follow a normal distribution $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$. Disposing of a set of observations $\mathbf{y}$, we aim at estimating the system's state $\mathbf{x}$ by means of the conditional probability $p(\mathbf{x}|\mathbf{y})$ (the probability of the model state given its observations). For that, we can make use of the marginalization rule

$$p(\mathbf{y}) = \int p(\mathbf{x}, \mathbf{y})\mathrm{d}\mathbf{x} = \int p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\mathrm{d}\mathbf{x} \tag{3.13}$$

and Bayes' theorem

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{3.14}$$

where is $p(\mathbf{x}|\mathbf{y})$ the posterior distribution, $p(\mathbf{y}|\mathbf{x})$ is the likelihood of the set of observations given the model, and $p(\mathbf{x})$ is the prior (or background) distribution. The choice of the prior is connected to a given prior knowledge of the system state in question. Denoting $\mathbf{x}^b$ and $\mathbf{P}^b$ as the background associated mean and covariance, we have that $\mathbf{x} \sim \mathcal{N}(\mathbf{x}^b, \mathbf{P}^b)$, which gives

$$p(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\left[(\mathbf{x}-\mathbf{x}^b)^T{\mathbf{P}^b}^{-1}(\mathbf{x}-\mathbf{x}^b)\right]\right\}. \tag{3.15}$$

The observations follow $\mathbf{y}^o \sim \mathcal{N}(\mathbf{H}\mathbf{x}, \mathbf{R})$, for which the likelihood can be written as

$$p(\mathbf{y}^o|\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\left[(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x})\right]\right\}, \tag{3.16}$$

where $\mathbf{H}$ is considered to be linear. Solutions for $\mathbf{x}$ in light of $\mathbf{y}$ can be sought by constraining the posterior probability given by

$$p(\mathbf{x}|\mathbf{y}^o) \propto \exp\{-\mathcal{J}\}, \tag{3.17}$$

for which

$$\mathcal{J} = \frac{1}{2}\left[(\mathbf{y}^o-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}^o-\mathbf{H}\mathbf{x}) + (\mathbf{x}-\mathbf{x}^b)^T{\mathbf{P}^b}^{-1}(\mathbf{x}-\mathbf{x}^b)\right]. \tag{3.18}$$

The solution $\hat{\mathbf{x}}$ can be sought in many ways, generally pursuing optimality, which is somewhat subjective. One can define the optimal solution as the one that gives the minimum posterior variance, by assigning $\nabla_{\hat{\mathbf{x}}}\mathcal{J}(\hat{\mathbf{x}}) = \mathbf{0}$, for example. Other estimation methods are the maximum a posteriori estimation, and the maximum likelihood, for which the latter corresponds to ignoring the prior information.

3.2.1 The best linear unbiased estimate

We can instead focus on some basic assumptions on the form of the final estimate in order to estimate $\mathbf{x}^t$. Assuming that the background and observations are unbiased, for which the covariances are representative of the uncertainties, and the observation operator is linear, we have that

$$\begin{aligned}
\mathbf{x}^b &= \mathbf{x}^t + \boldsymbol{\epsilon}^b, && (3.19)\\
\langle \boldsymbol{\epsilon}^b \rangle &= \mathbf{0}, && (3.20)\\
\mathbf{P}^b &= \langle \boldsymbol{\epsilon}^b \boldsymbol{\epsilon}^{bT} \rangle, && (3.21)\\
\mathbf{y}^o &= \mathbf{H}\mathbf{x}^t + \boldsymbol{\epsilon}^o, && (3.22)\\
\langle \boldsymbol{\epsilon}^o \rangle &= \mathbf{0}, && (3.23)\\
\mathbf{R} &= \langle \boldsymbol{\epsilon}^o \boldsymbol{\epsilon}^{oT} \rangle, && (3.24)
\end{aligned}$$

where $\langle\rangle$ denotes the expectancy. We choose the estimate to be a linear combination of both background state and observations, writting

$$\mathbf{x}^a = \mathbf{A}\mathbf{x}^b + \mathbf{K}\mathbf{y}^o, \tag{3.25}$$

where the superscript a denotes the final estimates, here called the 'analysis'. Here $\mathbf{A}$ and $\mathbf{K}$ are linear and to be determined following the optimality criteria. Choosing the final estimate to be unbiased and of minimum variance (optimal accuracy), we assume

$$\begin{aligned}
\langle \boldsymbol{\epsilon}^a \rangle &= \mathbf{0}, && (3.26)\\
\mathrm{tr}(\mathbf{P}^a) &\rightarrow \text{minimum}, && (3.27)
\end{aligned}$$

where tr() is the trace and $\boldsymbol{\epsilon}^a = \mathbf{x}^t - \mathbf{x}^a$ and $\mathbf{P}^a$ are the errors and covariance of the analysis. Using the unbiased hypothesis,

$$\begin{aligned}
\mathbf{A} &= \mathbf{I} - \mathbf{K}\mathbf{H}, && (3.28)\\
\mathbf{K} &= \mathbf{P}^b\mathbf{H}^T(\mathbf{H}\mathbf{P}^b\mathbf{H}^T + \mathbf{R})^{-1}, && (3.29)
\end{aligned}$$

where $\mathbf{K}$ is generally known as the 'Kalman gain', the matrix which takes into account the information on the accuracies of both background and observations to compose the estimate (see Cohn, 1997, for the derivation of the Kalman gain matrix). The best linear unbiased estimator (BLUE) has therefore the form

$$\begin{aligned}
\mathbf{x}^a &= \mathbf{x}^b + \mathbf{K}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b), && (3.30)\\
\mathbf{P}^a &= (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}^b, && (3.31)
\end{aligned}$$

where the quantity $(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)$ is called the 'innovation', which characterizes the distance between the observations and background model in the observation space.

It is worth noticing that the BLUE corresponds to the minimum variance estimate mentioned in the previous section (in the case the pdf follows a Gaussian distribution)[1], and also similar to the generalized least-squares solution (Talagrand, 1997).

---

[1] The assumption of Gaussianity is widely used due to the fact that the first two moments are sufficient to completely represent the estimate.

3.2.2 The resolution matrix

The analysis error covariance in Equation 3.31 can be written as $\mathbf{P}^a = (\mathbf{I} - \mathbf{S})\mathbf{P}^b$, where

$$\mathbf{S} = \mathbf{KH} = \mathbf{P}^b\mathbf{H}^T(\mathbf{HP}^b\mathbf{H}^T + \mathbf{R})^{-1}\mathbf{H}, \tag{3.32}$$

is known as the model resolution matrix. It represents how much the estimate is sensitive to the background information. It also shows how the statistics of the background model are affected by the new information provided by the observations after the analysis. An illustrative example is given in the following.

For a model parameter i, the ith diagonal component of the resolution matrix can be approximately written as

$$S_{ii} \sim \frac{\sigma_i^b}{\sigma_i^b + \sigma_m^y}, \tag{3.33}$$

where σ_i^b is the uncertainty affecting the ith model variable and σ_m^y a representation of the data uncertainties in model space. In the case where $\sigma^b \ll \sigma^y$, $S \approx 0$ and $\mathbf{P}^a \approx \mathbf{P}^b$, therefore the update is negligible. Otherwise if $\sigma^b \sim \sigma^y$, $S \sim 1/2$ thus $\mathbf{P}^a \sim 0.5\mathbf{P}^b$. In the best case scenario, however, $\sigma^b \gg \sigma^y$, $S \sim 1$ and $\mathbf{P}^a \sim 0$, for which the observations fully constrain the model estimation.

It is worth noting that in Eq. 3.33, the information represented by σ_m^y does not only bear information on the observation uncertainties, but also of the observation type and distribution present on the observation operator $\mathbf{H}$ (Eq. 3.32). Finally, the trace of the model sensitivity matrix is a very useful diagnostic, as it can be interpreted as the number of resolved parameters (e.g. Backus, 1988).

Different to the model resolution matrix, the data resolution matrix $\mathbf{HK}$ represents the sensitivity of the final estimate to the observations in data space. Cardinali (2013) gives the details of the data resolution matrix, both for the least-squares estimation and the data assimilation formalism.

3.2.3 The Kalman filter algorithm

The Kalman filter algorithm (Kalman, 1960) consists in the application of the BLUE in a dynamical sense, and it can be described by the repetition of two successive steps, the forecast and the analysis. Here, the background state $\mathbf{x}^b$ will be given by the forecast of a dynamical model $\mathbf{x}^f$, say, a numerical simulation, where the superscript represents 'forecast'. The model state $\mathbf{x}^f$ has dimension $n_\mathbf{x}$, while the observations of the system $\mathbf{y}^o$ have dimension $n_\mathbf{y}$.

Forecast

The forecast consists in the propagation of the state (as described by Eq. 3.5) by the underlying model, here represented by the linear model operator $\mathbf{M}$

$$\mathbf{x}_{i+1}^f = \mathbf{M}_{i,i+1}\mathbf{x}_i^f. \tag{3.34}$$

Given the error covariance at time i as $\mathbf{P}_i^f$, the error propagation is written as

$$\mathbf{P}_{i+1}^f = \mathbf{M}_{i,i+1}\mathbf{P}_i^f\mathbf{M}_{i,i+1}^T + \mathbf{Q}_{i+1}, \tag{3.35}$$

for the error has two components, one from the propagation of the initial error and another $\mathbf{Q}$ from the model error itself discussed in Section 3.1.

Analysis

Whenever observations are available, at a given time $i+1$ for example, the BLUE equations (Eqs. 3.30 and 3.31) will be applied based on the information from the forecast $i+1$. This step will then consist on the computation of

$$\begin{aligned} \mathbf{K}_{i+1} &= \mathbf{P}_{i+1}^f\mathbf{H}_{i+1}^T(\mathbf{H}_{i+1}\mathbf{P}_{i+1}^f\mathbf{H}_{i+1}^T + \mathbf{R}_{i+1})^{-1}, & (3.36)\\ \mathbf{x}_{i+1}^a &= \mathbf{x}_{i+1}^f + \mathbf{K}_{i+1}(\mathbf{y}_{i+1}^o - \mathbf{H}_{i+1}\mathbf{x}_{i+1}^f), & (3.37)\\ \mathbf{P}_{i+1}^a &= (\mathbf{I} - \mathbf{K}_{i+1}\mathbf{H}_{i+1})\mathbf{P}_{i+1}^f. & (3.38) \end{aligned}$$

A good diagnostic of the the analysis is the 'residual', given by $\mathbf{r}_{i+1} = \mathbf{y}_{i+1}^o - \mathbf{H}_{i+1}\mathbf{x}_{i+1}^a$.

This two-step algorithm is then repeated for a sequence of analysis cycles, whenever observations of the system are available. This method is also called 'sequential data assimilation', for which a simple scheme is presented in Figure 3.1.

Further considerations

An important drawback of the sequential algorithm as it is described above is the computational infeasibility considering the typical dimensions of the problem (depending on the application). In the geomagnetic case $n_{\mathbf{x}} \sim 10^6$, which means there is a need of computing and possibly storing a $10^6 \times 10^6$ covariance matrix $\mathbf{P}_i^f$. Besides the storage problem, the propagation of the $\mathbf{P}_i^f$ matrix by Eq. 3.35 implies two applications of the forecasting linear operator, it is $2 \times n_{\mathbf{x}}$ computationally more expensive than the simple model forecast. Such drawbacks make the Kalman filter application unfeasible. The computational cost can be reduced by the use of a frozen covariance matrix, a method known as Optimal Interpolation (OI). Although easy to apply, the OI obviously suffers from the absence of dynamical information on the error statistics.

3.3 Dealing with non-linearities

The formulation of the BLUE and therefore of the Kalman filter considered the case where the model propagator and observation operator are linear, given by $\mathbf{M}$ and $\mathbf{H}$, respectively. However, in many situations they are known to be non-linearly related to the state vector. This situation spoils the optimality of the aforementioned estimators, for the Gaussian assumption is not likely to hold in such case. Another drawback is that in the nonlinear case, the model and/or observation operators cannot be written in simple matrix form, therefore not allowing for a description of the operator transpose. There are two methods derived from the Kalman filter that can handle different levels of nonlinearities.

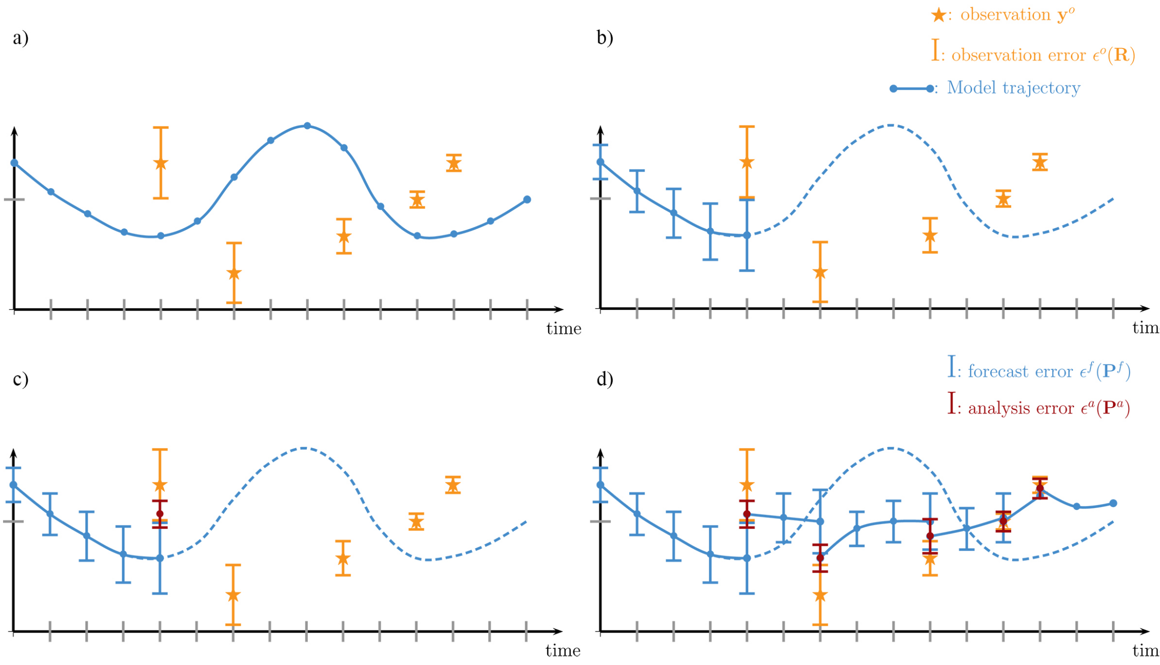


Figure 3.1: The main steps in a sequential data assimilation algorithm. a) An unconstrained model trajectory for the interval and a set of observations of the real system distributed over different times of the interval. b) The forecast step consists in the propagation of the unconstrained model up to a certain time where observations are available. c) Given the uncertainties in both the forecast and the observations and the innovation, an analysis is computed. d) The analysis is then propagated by the numerical model until new observations are available. From Fournier et al. (2010).

3.3.1 The Extended Kalman Filter

In the case the model propagator and observation operator, $\mathcal{M}$ and $\mathcal{H}$, are weakly non-linear, they can be well represented by their tangent linear approximation, $\mathbf{M}'$ and $\mathbf{H}'$ respectively, around a given reference state $\mathbf{x}_0$. They can be written as

$$\mathbf{M}' = \left.\frac{\partial \mathcal{M}}{\partial \mathbf{x}}\right|_{\mathbf{x}_0}, \tag{3.39}$$

$$\mathbf{H}' = \left.\frac{\partial \mathcal{H}}{\partial \mathbf{x}}\right|_{\mathbf{x}_0}. \tag{3.40}$$

The Kalman filter algorithm can be extended to the use of those operators, a formulation known as the Extended Kalman filter (EKF), explained in the following.

Forecast

The model state forecast is performed by the non-linear numerical simulation, while the covariance forecast is approximated by the use of the tangent linear operator

$$\mathbf{x}^f_{i+1} = \mathcal{M}_{i,i+1}(\mathbf{x}^a_i), \tag{3.41}$$

$$\mathbf{P}^f_{i+1} = \mathbf{M}'_{i,i+1}\mathbf{P}^a_i\mathbf{M}'^T_{i,i+1} + \mathbf{Q}_{i+1}. \tag{3.42}$$

Analysis

The calculation of the Kalman gain is done with the tangent linear observation operator, while the innovation is calculated by using the non-linear projector

$$\begin{aligned}
\mathbf{K}_{i+1} &= \mathbf{P}^f_{i+1}\mathbf{H}'^T_{i+1}(\mathbf{H}'_{i+1}\mathbf{P}^f_{i+1}\mathbf{H}'^T_{i+1} + \mathbf{R}_{i+1})^{-1}, && (3.43)\\
\mathbf{x}^a_{i+1} &= \mathbf{x}^f_{i+1} + \mathbf{K}_{i+1}[\mathbf{y}^o - \mathcal{H}_{i+1}(\mathbf{x}^f_{i+1})], && (3.44)\\
\mathbf{P}^a_{i+1} &= (\mathbf{I} - \mathbf{K}_{i+1}\mathbf{H}'_{i+1})\mathbf{P}^f_{i+1}. && (3.45)
\end{aligned}$$

Further considerations

The EKF can be expanded to the Iterated Extended Kalman filter (IEKF), in which, as the name suggests, the solution is sought iteratively. This small modification allows for better approximation of the linearization process. A description of the iterative solution for a non-linear observation operator can be found in Cohn (1997).

3.3.2 The Ensemble Kalman Filter

In the case of stronger non-linearities, the higher order terms on the expansion of Eqs. 3.39 and 3.40 play a significative role, compromising the EKF performance. Another possibility to account for the non-linearities of the problem is to conceive an ensemble of model states instead of a single state vector describing the system. This corresponds to a Monte-Carlo sampling of the model pdf, which is the idea behind the Ensemble Kalman filter (EnKF) proposed by Evensen (1994, 2009).

In synthesis, the EnKF consists in gathering at random an ensemble of models $\{\mathbf{x}^f_1, ..., \mathbf{x}^f_{N_e}\}$ of N_e ensemble members. We can define a mean background state $\langle\mathbf{x}^f_i\rangle$ and sample covariance matrix $\mathbf{P}^f$ given by

$$\begin{aligned}
\langle\mathbf{x}^f_i\rangle &= \frac{1}{N_e}\sum_{e=1}^{N_e}\mathbf{x}^f_{i,e}, && (3.46)\\
\mathbf{P}^f_i &= \frac{1}{N_e - 1}\sum_{e=1}^{N_e}(\mathbf{x}^f_{i,e} - \langle\mathbf{x}^f_{i,e}\rangle)(\mathbf{x}^f_{i,e} - \langle\mathbf{x}^f_{i,e}\rangle)^T, && (3.47)
\end{aligned}$$

where the subscript e refers to each ensemble member. Those two pieces of information relating to the first two moments of the background pdf are further incorporated on the Kalman filter algorithm. The observations $\mathbf{y}^o$ are set to match the random variables defining the model, therefore they should be as well defined as random variables (Burgers et al., 1998). For that reason, each ensemble member will assimilate the observation set at a time i assigned with different errors,

$$\mathbf{y}^o_{i,e} = \mathbf{y}^o_i + \boldsymbol{\epsilon}^o_{i,e}, \tag{3.48}$$

where $\boldsymbol{\epsilon_e}^o$ is the observation error added to a random perturbation with mean and standard deviation defined by the observation uncertainty $\boldsymbol{\epsilon^o}$.

Forecast

The forecast step is simple, consisting of a simple forwarding of the ensemble by the underlying model

$$\mathbf{x}_{i,e}^{f} = \mathcal{M}_{i,i-1}(\mathbf{x}_{i-1,e}^{a}). \tag{3.49}$$

Analysis

Equations 3.46 and 3.47 are then used to calculate the mean $\langle \mathbf{x}_i^f \rangle$ and the covariance $\mathbf{P}_i^f$. The Kalman gain matrix $\mathbf{K}_i$ can be calculated with the application of the direct nonlinear observation operator with the following steps

$$\begin{aligned}
\mathbf{H}_i\mathbf{P}_i^f &= \frac{1}{N_e-1}\sum_{e=1}^{N_e}\left[\mathcal{H}_i(\mathbf{x}_{i,e}^f) - \mathcal{H}_i(\langle \mathbf{x}_i^f \rangle)\right]\left[\mathbf{x}_{i,e}^f - \langle \mathbf{x}_i^f \rangle\right]^T, && (3.50)\\
\mathbf{H}_i\mathbf{P}_i^f\mathbf{H}^T &= \frac{1}{N_e-1}\sum_{e=1}^{N_e}\left[\mathcal{H}_i(\mathbf{x}_{i,e}^f) - \mathcal{H}_i(\langle \mathbf{x}_i^f \rangle)\right]\left[\mathcal{H}_i(\mathbf{x}_{i,e}^f) - \mathcal{H}_i(\langle \mathbf{x}_i^f \rangle)\right]^T, && (3.51)\\
\mathbf{K}_i &= \mathbf{P}_i^b\mathbf{H}_i^\dagger(\mathbf{H}_i\mathbf{P}_i^b\mathbf{H}_i^T + \mathbf{R})^{-1}, && (3.52)
\end{aligned}$$

(Nerger, 2004). The analysis is then performed for each ensemble member

$$\mathbf{x}_{i,e}^{a} = \mathbf{x}_{i,e}^{f} + \mathbf{K}_i(\mathbf{y}_{i,e}^{o} - \mathcal{H}_i\mathbf{x}_{i,e}^{f}). \tag{3.53}$$

The analysis estimate is then given by the mean analysis $\langle \mathbf{x}_i^a \rangle$ and the sample analysis covariance $\mathbf{P}_i^a$ is updated implicitly by means of the ensemble

$$\begin{aligned}
\langle \mathbf{x}_i^a \rangle &= \frac{1}{N_e}\sum_{e=1}^{N_e}\mathbf{x}_{i,e}^a, && (3.54)\\
\mathbf{P}_i^a &= \frac{1}{N_e-1}\sum_{e=1}^{N_e}(\mathbf{x}_{i,e}^a - \langle \mathbf{x}_{i,e}^a \rangle)(\mathbf{x}_{i,e}^a - \langle \mathbf{x}_{i,e}^a \rangle)^T. && (3.55)
\end{aligned}$$

Further considerations

One of the practical aspects of the EnKF is the way the covariances are handled. Since they depend only of the N_e model states, they do not need to be stored or even explicitly computed. Also, model uncertainty estimation is a rather difficult task, and the ensemble spread is easy to compute and thought to be a good first order proxy. It is important to note that for a sufficiently large ensemble, the analysis sample covariance converges to that of the classic Kalman filter (Burgers et al., 1998).

The number of ensemble members needed for the EnKF corresponds to $\mathcal{O}(10^2)$, since the error in sampling decreases as $N^{-1/2}$ (Evensen, 1994). This somehow negatively impacts the ease of dealing with the forecast covariance matrix. Such problem can be minimized, however, by a parallelization of the ensemble forecast itself (Nerger, 2004).

An alternative approach for dealing with a nonlinear observation operator consists in artificially augmenting the state vector. In such a case, the state vector is extended in order to admit the observation vector. As a result, the effective observation operator is linear and affects only the extended part of the state vector. This approach leads to the same performance as the EnKF.

3.4 Comparison with the variational approach

Data assimilation is essentially divided between two different methodologies, the sequential Kalman filter-based algorithm and variational methods. While sequential assimilation is based on estimation theory, the variational methods are based on optimal control theory. The variational methods (3D-Var or 4D-Var, depending if the problem is time-dependent), aim at minimizing the functional in Eq. 3.18, generally called the 'cost', or 'objective', function. If the observations are time-dependent, the adjustment of the model given the observations is performed over the whole interval, meaning that the model is influenced by all the available observations at once.

An important control parameter for a numerical model is its initial condition $\mathbf{x}_0$. In the variational approach, it is possible to determine it from all subsequent data considered over a given time interval. A scheme of the variational data assimilation is showed in Figure 3.2a. As for the minimum variance estimate, the estimation is done by seeking the inflection point of the cost function, by means of $\nabla_{\mathbf{x}_0}\mathcal{J}(\mathbf{x}_0) = \mathbf{0}$. For a situation where the state vector is large, the brute force calculation of the gradient is ruled out, since the model would have to be forwarded $n_{\mathbf{x}}$ times. Instead, most variational algorithms are based on the adjoint method (for a good and simple description of the adjoint method, see Talagrand, 1997). In this situation, the backwards calculation of the adjoint field is needed, based on the adjoint tangent linear model operator $\mathbf{M}^T$. The derivation of the adjoint model by hand is a rather difficult task, possibly becoming unfeasible depending on the complexity of the forward model. Instead, one can resort to automated differentiation algorithms in order to calculate the adjoint model from a specific code (e.g. Sambridge et al., 2007).

The variational approach can easily be extended in order to adjust not only the model initial conditions, but also its underlying control parameters and boundary conditions. However, there are two strong drawbacks with respect to the sequential algorithms. One is that the variational formalism generally assumes the model as perfect (that is, $\eta = 0$). The other is the absence of analysis error statistics, although recent applications (Egbert et al., 1994) impose additional constraints on the cost function in order to estimate the a posteriori uncertainties. It is interesting to note, however, that both sequential and variational algorithms are convergent in terms of their estimates at the end of the assimilation period, in the case of linear operators, perfect model and Gaussian statistics (Talagrand, 1997). Moreover, despite the considerable computational cost of both variational and sequential assimilation algorithms, 4D-Var and EnKF for example, their operational performance in terms of forecast are seen to be equivalent for NWP (Buehner et al., 2008).

Also, a big advantage of the variational approach with respect to the sequential one is that in the former, the information from the observations is propagated backwards as well

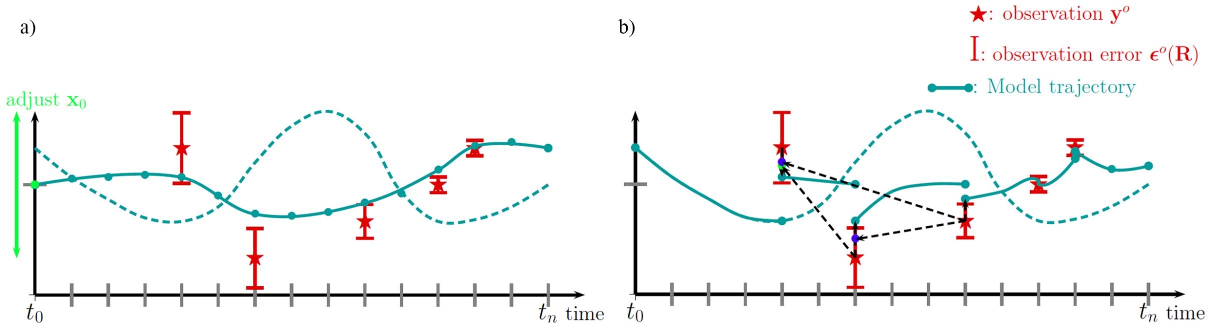


Figure 3.2: Variational data assimilation and sequential smoother. In the variational approach a) the background is adjusted within all set of observations spanning the whole time interval by an iterative scheme in order to better fit the observations and through that find the model initial condition $\mathbf{x}_0$. The sequential smoother b) is performed on top of the already performed sequential analysis, retropropagating the observation information in time. From Fournier et al. (2010).

as forward (compare Fig. 3.2a and Fig. 3.1). This, however, can be compensated by the use of the Kalman *smoother* (KS) in sequential methods. In the KS, once the sequential forward assimilation is performed, a second backwards sequential assimilation is applied, retropropagating the information of the observations into the past (Cohn et al., 1994). A scheme of the KS algorithm is shown in Figure 3.2b. In fact, filtering is one of the many aspects of the sequential estimation problem, in which observations up to a given time k are used to estimate the state of the system at all times until and including k. Smoothing is the case for which observations before and after k are used to constrain the state at the time k. Prediction, finally, is when observations at the time k are used to estimate the state after k.

3.5 Applications

Although data assimilation has flourished in the atmospheric science context, it has begun inspiring diverse scientific fields when it comes to understanding the evolution of a given physical system and predicting its behavior. Widely known from NWP (Kalnay, 2003), where the main objective is predicting the short-range dynamics of the atmosphere, data assimilation has been extended for longer time scales of the atmosphere dynamics, in the field of climatology (Dee et al., 2011). Climatology research is important in face of the observed current climate changes, and has lead to the creation of the IPCC (Intergovernmental panel on climate change) whose reports on the estimated future evolution of climate due to human activity and carbon gas emissions are strongly based on data-assimilation and ensemble methods (Knutti and Sedláček, 2013).

However, climate itself is strongly dependent on the state of the ocean, and the increasing interest in unraveling the ocean dynamics motivated progresses in oceanographic data assimilation (Ghil and Malanotte-Rizzoli, 1991). Given the big technical improvements for the acquisition of indirect data which recorded the history of the states of the atmosphere and the oceans, the use of reanalysis and smoothing has been of particular interest to oceanographic data assimilation (e.g. Carton and Giese, 2008; Cosme et al., 2010).

While the big scales of modeling of the Earth's climate can be found through coupling ocean and atmospheric models, the accurate prediction of its subsystems requires researchers to further specialize in fields such as hydrology, sedimentology, ecology, continental surfaces studies and glaciology. All these fields are currently using data assimilation either for reanalysis or parameter estimations (Rodell et al., 2004). Additionally, the applications of data assimilation on surface and near-surface processes also cross into monitoring natural resources (Nævdal et al., 2005) with potential applications both for the industrial sector and for environmental impact studies.

The study of fields of geoscience that are less dependent on human activity also have benefited from using data assimilation techniques. For instance, mantle convection models have been combined with tectonic data (like surface kinematics and seafloor age distribution) with promising results (Bunge et al., 2003; Bocher et al., 2016). The study of the ionosphere, has been using data assimilation to incorporate variations of the GPS signal intensities to better estimate its evolution (Bust et al., 2004).

A decade ago, data assimilation was also introduced in astrophysics, in particular on solar dynamo research by Brun (2007). Solar dynamo data assimilation would have particular importance in space weather, due to the strong link between the terrestrial magnetic environment and the solar activity. The first efforts in solar data assimilation were made by Kitiashvili and Kosovichev (2008), followed by Dikpati et al. (2014), in a sequential approach, and by Jouve et al. (2011) in a variational one. Going beyond prediction of the solar magnetic field, Hung et al. (2015) enhanced the latter variational framework in order to constrain the underlying flow in the solar convection zone.

The dynamo models used in the aforementioned examples are mean-field models (e.g. Sanchez et al., 2014b), mentioned in Section 1.2. Their relative simplicity due to the reduced dimension imposed by axisymmetry is the reason for their use in solar data assimilation. Given their relevance to solar activity prediction, it is important to acknowledge the existence of a predictability limit (recall Section 2.9) of such mean-field based solar data assimilation. Such study was performed by the author at the beginning of the thesis project described in this manuscript. The results, shown in Sanchez et al. (2014a), pointed to a forecast horizon of ~ 3 solar cycles. This article is presented in Appendix A.

At about the same time data assimilation was introduced in the solar dynamo, it begun to draw attention to the geomagnetic and geodynamo context, as detailed in the following.

3.6 Geomagnetic assimilation: state of the art

As stated in Fournier et al. (2010), geomagnetic data assimilation is still in its infancy. Preliminary studies of simple one-dimensional versions of the MHD dynamo equations mimicking the geomagnetic data assimilation case have been introduced in a variational form by Fournier et al. (2007) and EnKF context by Sun et al. (2007). Furthermore, a more complex geodynamo model (like the ones described in the previous chapter) has been used in the specific sequential context of OI by Liu et al. (2007); Kuang et al. (2008) and EnKF (Fournier et al., 2013), confirming the feasibility of sequential geomagnetic assimilation. The variational route by the adjoint method has also been further explored in terms of a quasi-geostrophic model by Canet et al. (2009) and by a simplified 3D MHD dynamo introduced by Li et al. (2011, 2014). All such studies are closed-loop experiments (also known as twin or fraternal experiments, or even OSSE's, observing system simulation experiments), where synthetic data is generated from a reference numerical model (characterized as the true state) so the assimilation performance can be directly assessed.

Given the relative success of geomagnetic data assimilation in twin scenarios, different methodologies have been applied to 'real' geomagnetic data (Kuang et al., 2009; Tangborn and Kuang, 2015), sometimes with the objective to contribute to new IGRF models and its short term predictions (Kuang et al., 2010; Gillet et al., 2015; Fournier et al., 2015). Longer term predictions of the Earth's magnetic field have also been performed, showing a probable continuation of the decrease in the axial dipole moment intensity and westward drift of the South Atlantic magnetic anomaly (Aubert, 2015, where the secular variation is also considered as an observation).

However, all studies aforementioned do not directly assimilate proper geomagnetic data, that is, the components of the magnetic field vector at the Earth's surface or in altitude. Instead, they consider Gauss coefficients from inverse models of the magnetic field at the CMB (introduced in Section 1.5) as proxy to geomagnetic observations. Such 'level-2' observations (in the space agency community jargon) are privileged for two main reasons. One is that since most dynamo codes are written in spectral form, the observation operator in order to assimilate 'level-2' observations consists in a simple diagonal matrix (see for example Kuang et al., 2009). The second reason is that such observations are already pre-processed in order to isolate the core's signal from all other magnetic field sources, like the magnetosphere, ionosphere and induced currents in the mantle. However, the use of 'level-2' observations should be performed with caution in the assimilation exercise. Since such observations are the result of inverse modeling of the magnetic field of the CMB, they are subject to the influence of regularization, generally acting to damp the smaller scales of the observable core's magnetic field. Also, the assessment of observational errors is not evident for the 'level-2' observations, although recent models have drawn specific attention to the matter (Gillet et al., 2013). All in all, following the reasoning from Talagrand (1997), direct (raw) observations should always be privileged in entering the data assimilation exercise, an idea which should be extended to geomagnetism.

3.7 Assimilating real geomagnetic observations

As seen in Section 1.5, the assumption of an electrically insulating mantle is appropriate at first order. In this case, a vector component α of the magnetic field at or above the Earth's surface at a radius $\mathbf{r}_o$ is connected to the magnetic field at the CMB by Eq. 1.8. The equation can be rewritten as

$$B_\alpha(r_o,\theta_o,\phi_o) = \int G_\alpha(r_o,\theta_o,\phi_o|r_c,\theta,\phi) B_r(r_c,\theta,\phi) r_c^2 \sin\theta \mathrm{d}\theta \mathrm{d}\phi, \tag{3.56}$$

where $\alpha = \{X, Y, Z\}$ is the vector component, (r_o,θ_o,ϕ_o) are the coordinates at which the observation is made and r_c, the radius of the CMB, and G_α is the corresponding data kernel, or Green function, corresponding to the α observation. The Green functions for the vector components of the magnetic field for a given a location at the Earth's surface

$$\begin{aligned} G_Z &= \frac{1}{4\pi}\left[\rho^2 - \frac{\rho^2(1-\rho^2)}{f^3}\right] && (3.57)\\ G_X &= \frac{1}{4\pi}\left[\frac{\rho^3(1+2f-\rho^2)}{f^3(1+f-\rho\mu)}\right]\hat{r}_c\cdot\hat{\theta}_o && (3.58)\\ G_Y &= -\frac{1}{4\pi}\left[\frac{\rho^3(1+2f-\rho^2)}{f^3(1+f-\rho\mu)}\right]\hat{r}_c\cdot\hat{\phi}_o. && (3.59)\end{aligned}$$

where $f=\sqrt{1-2\mu\rho+\rho^2}$, $\rho = r_c/r_o$ and $\mu = \hat{r}_c\cdot\hat{r}_o$ (see Constable et al., 1993, for the derivation of the Green functions, and Fig. 1.12 for their graphical representations).

The information from Eq. 3.56 indicates the basic dependency to enter the observation operator for raw direct geomagnetic observations. Its form in a data assimilation scheme, however, will depend on the form of the state vector $\mathbf{x}$. Details on the derivation of the geomagnetic observation operators $\mathbf{H}_X, \mathbf{H}_Y$ and $\mathbf{H}_Z$ for the outputs from a Parody code numerical simulation (Section 2.6) state vector are given in the next chapter.

As mentioned in Section 3.2.3, the typical size of a geodynamo simulation-based state vector is $n_{\mathbf{x}} \sim 10^6$. Figure 3.3 shows a histogram of the number of observations for different types of geomagnetic data through the last thousand years. Although existing in a narrow time window, direct observations of the magnetic field, from magnetic observatories and satellites, are relatively abundant. If we consider, for example, one year of satellite observation, we would in principle consider an observation vector dimension of $n_{\mathbf{y}} \sim 10^5$, closely matching the size of the state vector. One may think that the tendency of data quantity to increase in time might lead the system to be finally overdetermined. This perception is, however, misleading due to the magnetic curtain (Section 1.5). In other words, no matter how copious the magnetic data, the crust magnetization inevitably masks the core magnetic field beyond spherical harmonic degree 13. Considering that approximately 200 Gauss coefficients are needed to determine the field up until degree 13, the full state vector $\mathbf{x}$ would be sampled up to 0.02% in the best case scenario (Fournier et al., 2010).

As previously stated, one of the main difficulties in assimilating raw geomagnetic observations is the separating the magnetic field sources. This is specially important for very precise satellite and observatory data, for which $\sigma \sim 10$ nT. However, for the archeomagnetic dataset, the observational uncertainties are such that the contamination of the core signal from different magnetic field sources is negligible in the estimation problem.

3.8 The case for archeomagnetic observations

As mentioned in Section 1.6, the archeomagnetic catalog consists of indirect observations of the magnetic field vector. Since it is often given as incomplete information from either its directions or its intensity, a full vector reconstruction from the database is not possible. This situation corresponds to the problem of a nonlinear observation operator presented in Section 3.3. The possibilities in hand are to resort to a linearized observation operator

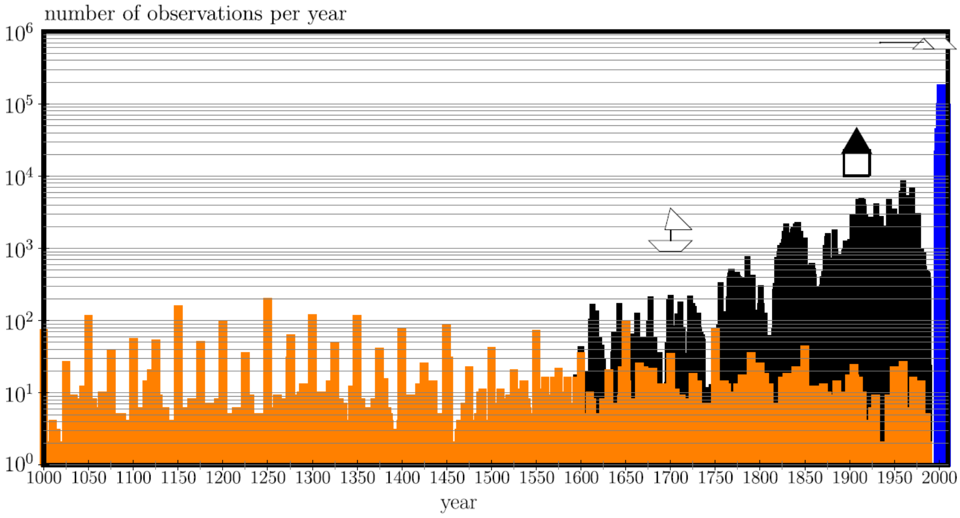


Figure 3.3: Histogram of observations from the magnetic field per year through the last millennium in logarithmic scale. The colors represent the different data catalogs: the archeomagnetic dataset in orange, the historical dataset (data from navigation logbooks and magnetic observatories) in black, and the satellite dataset in blue. From Fournier et al. (2010).

(as in the EKF, Section 3.3.1) or to an ensemble assimilation strategy (as in the EnKF, Section 3.3.2). A hybrid of those two methodologies is applied for the time-independent problem of archeomagnetic field modeling in the following part, Chapter 4. This study makes possible an analysis of the resolution power of the archeomagnetic field observations, preparing the ground for a sequential assimilation of the archeomagnetic data in Chapter 5.

Chapter 4

Dynamo-based archeomagnetic field modeling

Our knowledge of the core magnetic field in the global scales is based on diverse types of observations, direct and indirect, at the Earth's surface. Due to both the uneven distribution of the magnetic observations in space and time and the contamination of the magnetic signal by a plurality of sources close to the observation sites, the inverse problem for modeling the magnetic field of the core is of special complexity. There exist numerous inverse core field models (as described in Section 1.6.5) differing mainly in resolution, data catalogs and data processing. The underlying inverse framework remains the same.

This chapter highlights the possibility of changing a key ingredient in such methodology, the prior spatial information entering the core field inverse problem. Instead of an adjustable spatial norm, dynamo-based correlations of the magnetic field at the top of the core are used to generate the spatial prior. This approach serves as an introduction to the use of data assimilation in geomagnetism, preparing the ground for a sequential assimilation algorithm as the ones described in Chapter 3. Since the interest of this thesis is to understand the long-term dynamics of the magnetic field, the data entering the inverse problem consists of archeomagnetic observations. In particular, this chapter focuses on archeomagnetic field modeling through the time when archeomagnetic data are more abundant, between 1200 BC and 2000 AD.

4.1 The general inverse problem

As seen in Section 1.5, the magnetic field is generally represented in a spherical harmonic basis. A model of the core magnetic field at a given time can be defined as the grouping of Gauss coefficients $\mathbf{x} = [..., g_\ell^m, h_\ell^m, ...]$ up to a given truncation. Given a set of data $\mathbf{y}^o$ with error covariance $\mathbf{R}$, connected to the model by H, the traditional inverse problem involves minimizing the functional

$$\mathcal{J}(\mathbf{x}) = [\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x})], \tag{4.1}$$

which is equivalent to the cost function introduced by Eq. 3.18, without the prior constraint. The solution to the minimization of Equation 4.1 corresponds to the weighted least-squares one. Complications arise, however, due to the general non-unique character of the problem and the instability of the solutions (see Parker, 1994; Gubbins, 2004, for extensive reviews on inverse theory and applications in geomagnetism). To mitigate those issues an additional constraint, or regularization, is added to the cost function based on some prior information of the model. A specific model norm is then added to Eq. 4.1 to stabilize the solutions, as in

$$\mathcal{J}(\mathbf{x}) = [\mathbf{y}^o - H(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y}^o - H(\mathbf{x})] + \mathbf{x}^T \lambda \mathbf{C}^{-1} \mathbf{x}, \tag{4.2}$$

where **C** is a matrix defining the model complexity and λ is called the damping parameter. In the case where λ is big, the solution will favor the information on the model norm, and conversely, if λ is small, the solution will prioritize the fit to the data.

The damping parameter being adjustable, one generally seeks a compromise solution between model complexity and data fit. This trade-off between the fit and the norm is illustrated in Figure 4.1. The matrix **C** is chosen based on plausible assumptions about the basic features of the model. In the case of the geomagnetic field, the 'dissipation norm' is often used, involving an approximated expression for a lower bound for the Ohmic dissipation at the CMB (Gubbins and Bloxham, 1985). Such norm scan differ, however, from model to model, mostly in order to mitigate the damping effect on the dipole (e.g. Korte et al., 2009). Also, an additional model norm can be introduced for treating the time-dependency of the problem, which for simplicity will be set aside in this chapter.

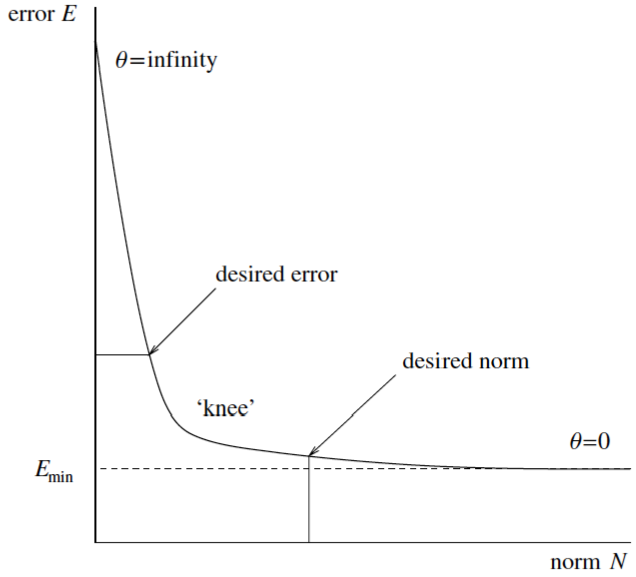


Figure 4.1: Trade-off curve showing different solutions of the inverse problem depending on the damping parameter. The error E represents the misfit between the data and model projection, the first term on the RHS of Eq. 4.2, the norm N the second term on the RHS of the equation and θ the damping parameter. From Gubbins (2004) .

Since the model norm should represent the basic information from the actual magnetic field at the core, one may wish to benefit from the prior information given by dynamo simulations. Recalling the discussions in Chapter 2, many numerical dynamo models are categorized as morphologically Earth-like, and can therefore be used to bring statistical information from the magnetic field at the CMB to build **C**. In this case, the prior

’dissipation norm’ will be replaced by what is introduced in this chapter as a ’dynamo norm’.

4.2 Prior information from dynamo models

The prior dynamo information can be given by the mean dynamo state and the covariance between the different variables composing the dynamo system, in a similar way as introduced in the EnKF formalism in Section 3.3.2. In fact, the mean and covariance can be calculated from a given ensemble of models originating from a dynamo simulation, like for example the CE model, following the EnKF approach. Since we are interested in the time-independent estimation problem, only the observable part of the dynamo system, that is, the poloidal component of the magnetic field (Eq. 2.33), is important to compose the model state vector $\mathbf{x}$. Such difference in input information from the ’dissipation’ and ’dynamo’ norm is likely to impose very different spatial correlations, affecting the way information is propagated from the observations to the model.

4.3 Nonlinear observation operator

The archeomagnetic observations, declination, inclination and intensity, are nonlinearly related to the radial component of the magnetic field, and therefore to the state vector. In such case, the two options are either to rely on the use of a linearized observation operator (as in the EKF approach, Section 3.3.1) or on an ensemble formalism (as in the EnKF, Section 3.3.2) to mitigate the effects of spoiling the assumed Gaussian statistics by non-linearities. It is also possible to combine both methodologies, in order to improve the solution.

As mentioned in Section 1.6, the archeomagnetic dataset can comprise observations originating from the magnetic remanence of volcanic rocks, archeological artifacts and sedimentary deposits. In comparison with the first two data, the sedimentary ones are hard to integrate in terms of observation operator, since often they are given in terms of relative components. In the following, for simplicity, they will be set aside. The archeomagnetic dataset mentioned further will imply therefore data only from volcanic rocks and archeologic artifacts.

4.4 Archeomagnetic resolution

The significant heterogeneity in archeomagnetic data distribution (Fig. 1.11) leaves open the question to what extent such observations can image the core magnetic field (Donadini et al., 2009). The resolution of the archeomagnetic data is not only compromised by the biased sampling of the Earth’s surface but also by the high uncertainties affecting the indirect data, much higher than the ones from direct observations from the historical and contemporary record. Additionally, the uncertainties in age attribution for both radiochronology and archeologic dating can span decades and even centuries sometimes, as mentioned in Section 1.6.

Inverse models of the main core field for the archeomagnetic period based on the adjustable dissipation norm do not directly tackle the quantification of resolution of the archeomagnetic dataset. Initial archeomagnetic field models supported the impossibility of going beyond the dipole and quadrupole, truncating the model to low spherical harmonic degree and order (Hongre et al., 1998). However, this approach allows for features of higher complexity of the magnetic field to artificially contribute to lower spatial scales, compromising the low resolution model. As a consequence, more recent models consider the archeomagnetic field truncation up to degree 5 (Licht et al., 2013) or further up to degree 10 (Korte et al., 2009). From comparison of the power spectra of archeomagnetic models and the ones based on historical and satellite data, Korte and Constable (2008) suggest the former are resolved spatially up to degree 4. However, this criterion does not take into account the fact that the regularization by the 'dissipation' norm implies a damping of the model coefficients, especially the higher degrees. Although initial models focused on providing snapshots or time averages over some specific intervals (Constable et al., 2000), current ones apply a temporal regularization imposing a magnetic field with smooth variations, minimizing the second time derivative of the field at the CMB (Korte and Constable, 2003). To that matter, Korte and Constable (2008) also suggests a limited temporal resolution for archeomagnetic data, of the order of 100 years.

In the previous two sections, emphasis was placed on the possibility of describing a well-defined prior based on Earth-like dynamo simulations and a linearized observation operator. These two pieces of information allow then for a straightforward calculation of the model resolution matrix, introduced in Section 3.2.2. The prior being fixed, the quantification of the resolution depends solely on the data quantity, their type (declination, inclination or intensity), geographic distribution and uncertainty. In order to understand the weight of this information on archeomagnetic record, we propose an analysis of the evolution of the archeomagnetic resolution in time.

4.5 A prelude to archeomagnetic data assimilation

The temporal treatment of core field inverse modeling is generally handled by the development of each model variable in cubic B-splines functions (e.g. Korte and Constable, 2003), creating a continuous model. The spline knot spacing can vary from model to model, depending on the assumption in data temporal resolution, discussed in the previous section. This temporal dependency is then regularized by the assumption that the field must evolve smoothly in time, by means of the minimization of the second time derivative of the field. Alternatively, the 'smooth' temporal norm can be substituted by defining the time variability of the model coefficients as stochastic processes and calculating their respective temporal correlation functions, as in Gillet et al. (2013).

From the data assimilation point-of-view, the temporal dependency would enter the estimation problem by means of the underlying numerical model representing the physics of the system. In particular for the sequential assimilation (Section 3.2.3), the numerical model would provide the prior information at analysis time and also a means of propagating the information gained from the data at the forecast stage. Further exploration of the temporal evolution would be provided by the exploration of the temporal covariances, smoothing out the final archeomagnetic model trajectory.

In order to prepare the ground for such a task, the performance of the analysis algorithm should be tested and the spatial resolution of the archeomagnetic observations should be explored. As an introductory study for archeomagnetic assimilation, the next section focuses on the time-independent estimation problem. A matlab algorithm based on the EnKF analysis step was developed by the author in order to estimate the magnetic field under the dynamo constraints imposed by outputs from a dynamo model. A constant prior information is given by an ensemble of model states from the CE simulation and used for estimating the magnetic field at the CMB at different time windows of the recent archeomagnetic period. Being time-independent, such an estimation represents therefore the average field for each time window interval.

4.6 Article: Modeling the archeomagnetic field under spatial constraints from dynamo simulations: a resolution analysis

The following article, currently in minor revision for publication in the Geophysical Journal International, comprises an introductory approach to archeomagnetic data sequential assimilation. It can be regarded as an inverse modeling exercise with the innovation of being based on dynamo spatial covariances as prior information. Such a physically based prior allows for the possibility of quantification of the archeomagnetic data resolution of the magnetic field at the CMB. The quantification of resolution of the archeomagnetic data is very important to any estimation problem, for it makes possible the assessment of the trustworthiness of the different archeomagnetic core models variables. The algorithm is used to construct a dynamo-based archeomagnetic field model, the AmR model. In order to validate the results from the archeomagnetic resolution matrix, AmR is compared to other archeomagnetic field models, ARCH3k.1 from Korte et al. (2009) and AFM from (Licht et al., 2013). The differences in the models enable the easy distinction between data-driven and prior-driven periods for the Gauss coefficients, in line with the resolution of the archeomagnetic data.

Geophys. J. Int. (0000) **000**, ??–??

Modeling the archeomagnetic field under spatial constraints from dynamo simulations: a resolution analysis

S. Sanchez[1], A. Fournier[1], J. Aubert[1], E. Cosme[2], Y. Gallet[1]
[1] *Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot, CNRS (UMR 7154), Paris, France.*
E-mail: `ssanchez@ipgp.fr`
[2] *Université Grenoble Alpes-CNRS, Grenoble, France*

March 3, 2016

SUMMARY

Archeomagnetic observations are key to recovering the behavior of the geomagnetic field over the past few millennia. The corresponding dataset presents a highly heterogeneous distribution in both space and time. Furthermore, the data are affected by substantial age and experimental uncertainties. In order to mitigate these detrimental properties, time-dependent global archeomagnetic field models are usually constructed under spatial and temporal regularization constraints, with the use of bootstrap techniques to account for data uncertainties. The models so obtained are the product of an adjustable tradeoff between goodness-of-fit and model complexity. The spatial complexity is penalized by means of a norm reflecting the minimization of Ohmic dissipation within the core. We propose in this study to resort to alternative spatial constraints relying on the statistics of a numerical dynamo simulation with Earth-like features. To that end, we introduce a *dynamo norm* in an ensemble least-squares iterative framework, the goal of which is to produce single-epoch models of the archeomagnetic field. We first validate this approach using synthetic data. We next construct a redistributed archeomagnetic dataset between 1200 BC and 2000 AD by binning the data in windows of 40-year width. Since the dynamo norm is not adjustable, we can legitimately calculate a resolution matrix to quantify the resolving power of the available archeomagnetic dataset. Gauss coefficients are resolved up to spherical harmonic degree 3 for the first thousand years of the interval, to degree 4 for the next thousand years and to degree 5 during the last millennium. These conclusions are based on the distribution and uncertainties that characterize the dataset, and do not take into account the possible presence of outliers. Comparison between our model, called AmR, and previously published archeomagnetic field models confirms the archeomagnetic resolution analysis: it highlights the dichotomy between data-driven coefficients for which model predictions coincide (within their respective uncertainties), and prior-driven coefficients. This study opens the way to physics-based models of the archeomagnetic field; future work will be devoted to integrating the framework here introduced into a time-dependent ensemble assimilation scheme.

Key words: Archeomagnetism – Dynamo: theories and simulations – Inverse theory – Magnetic field variations through time

1 INTRODUCTION

The main source of the Earth's magnetic field is located in the Earth's core. Composed mainly of iron, the core is in liquid phase in its external part and solid in its interior. The temperature gradient between the inner and outer boundaries, together with the release of lighter buoyant elements from the crystallization of the inner core, induces convection inside the outer core. Such convective motions of the highly conductive iron flow enable electromagnetic induction, generating thus a magnetic field via a dynamo process. This process, affected by the Coriolis force due to the Earth's rotation, gives place to a mainly dipolar structure of the global magnetic field, roughly aligned with the rotation axis. There are, however, important departures from dipolarity, like the presently observed South Atlantic Magnetic Anomaly. The westward drift and spread of this anomaly, as well as the intensity decrease of the axial dipole throughout the past hundred years, illustrate the late Secular Variation (SV) of the Earth's magnetic field (Hartmann & Pacca, 2009; Finlay et al., 2010; Jackson et al., 2000).

Although current direct measurements of the magnetic field offer global coverage in terms of very precise satellite, magnetic observatory and survey data, the magnetic field and its SV beyond the past few centuries remain far less constrained. The key to unravel it relies on the indirect magnetic field information provided by the remanent magnetization acquired by geological deposits (mainly lava flows) and archeological artifacts at the time of their formation, emplacement, fabrication or use. These observations are known as archeomagnetic data, which are compiled in global databases, such as GEOMAGIA (Donadini et al., 2006), for directional and intensity data covering the past 50 millennia, and ArcheoInt (Genevey et al., 2008) for intensity data over the past 10 millennia. Unfortunately, these datasets are clustered around some specific areas of the globe, like Europe and Asia, leaving the South Hemisphere practically undocumented. They are also sparse before the last two thousand years. Despite the challenging character of archeomagnetic compilations, they have allowed the development of several inverse models of the global magnetic field spanning the past few millennia (Korte & Constable, 2005; Korte et al., 2011; Licht et al., 2013; Pavón-Carrasco et al., 2014; Nilsson et al., 2014).

Inverse magnetic field models are the result of a minimization of the misfit between data and model, generally regularized in space and time by a certain norm of the model to restrict the solution of such an ill-posed inverse problem (Gubbins & Bloxham, 1985). The regularization is based on prior information built on the hypothesis that the core magnetic field varies smoothly in time and that the dissipation of magnetic energy at the core-mantle boundary (CMB) is minimum (Gubbins, 1975; Bloxham & Jackson, 1992). The shorter temporal and spatial scales of inverse geomagnetic field models are consequently damped, thereby concealing a possibly more complex behaviour of the geomagnetic field. As an alternative, different constraints can be sought in numerical simulations of the Earth's dynamo, which make it possible to understand the main dependencies between the field variables of the dynamo system and also to construct, for instance, the statistics describing the variability of the magnetic field at the CMB. Such statistics can be used to define a "dynamo norm" to constrain the archeomagnetic inverse problem, in place of the cannonical regularizations discussed above. We shall explore this possibility in this study.

Different kinds of dynamo models can be built depending on the control parameters and coupling mechanisms within the different regions of the deep Earth. Some standard dynamo models, considering thermochemical convection as the driving mechanism of the system, succeed in displaying an Earth-like morphology of the magnetic field relative to reference inverse models (Christensen et al., 2010), but fail to present an Earth-like SV pattern (Aubert et al., 2013). However, the recent model designed by Aubert et al. (2013) is able to reproduce the localized westward drift of low latitude magnetic anomalies, like the South Atlantic Anomaly. The model, coined "Coupled Earth", considers gravitational coupling mechanisms and heterogeneous crystallization of the inner core to produce an Earth-like SV. Given its Earth-likeness, the statistical information from Coupled Earth can be used as prior information in the problem of estimating the magnetic field at the CMB (e.g. Aubert, 2014).

Estimating the state of a system given a set of noised observations together with a prior model state is a filtering problem (Jazwinski, 1970), for which Kalman filter-based algorithms are largely used. However, while errors are usually prescribed for observations, uncertainties affecting the prior information are difficult to assess. An option is to work with an ensemble of numerical simulations, a stragegy at the heart of the Ensemble Kalman filter (EnKF) (Evensen, 2003). In this case, the prior estimate is given by the mean of an ensemble of model states derived from the numerical integration of the system in question, while prior uncertainties are represented by the sample ensemble covariance. Assuming the ensemble covariance is a good proxy for model error, the EnKF will update the prior information in light of the observations and produce a posterior estimate (also called the analysis), characterized as well by its mean and covariance.

The EnKF is one example of the many techniques used in the data assimilation framework, the backbone of which relies on a dynamical model describing the physics of a system to be updated with information from its observations. Although data assimilation has been widely used in meteorology and oceanography (Kalnay, 2003), it has begun to be explored in geophysics only recently. In particular, an introduction to data assimilation applied to the specific problematic of the Earth's magnetic field can be found in Fournier et al. (2010). So far, the ongoing work on geomagnetic data assimilation has focused on updating a numerical model of the geodynamo with Gauss coefficients of the magnetic field as input data (e.g. Fournier et al., 2013; Li et al., 2014; Tangborn & Kuang, 2015). Although the results are promising, in terms of propagation of information from the CMB throughout the core (see also Kuang et al., 2009; Aubert & Fournier, 2011, for earlier attempts), in an ideal data assimilation scenario raw pointwise observations of the magnetic field should be assimilated in place of field model coefficients.

We present in this paper a preliminary attempt at assimilating such pointwise observations, whose goal is to model the archeomagnetic field at a given time, under the spatial constraint specified by the Coupled Earth dynamo norm. For simplicity, we ignore at this stage the possibility of performing a temporal sequence of analyses (as permitted by a full EnKF setup), and focus instead on the impact of the dynamo norm on single-epoch field models. In order to ease comparison of our temporally unregularized approach with previously published archeomagnetic field models for the period 1200 BC to 2000 AD, we resort to the Geomagia50.V2 database of Donadini et al. (2006, 2009) that was used to produce these models (despite the recent update of GEOMAGIA by Brown et al., 2015).

This paper can be summarized as follows. We begin by presenting the estimation method and the dynamo model in

Section 2, including details on the archeomagnetic observation operators, the ensemble approach and a synthetic validation test. Next, in Section 3 we apply the methodology to an archeomagnetic context, focusing on the resolving power of archeomagnetic data. Section 4 describes a dynamo-based archeomagnetic field model relying on a redistributed archeomagnetic dataset, which we compare with previously published models. A summary and some perspectives are finally given in Section 5.

2 METHOD AND MODEL

2.1 Generalized least-squares iterative framework

At any given epoch, we wish to minimize a functional $\mathcal{J}$ of the form

$$\mathcal{J}(\mathbf{x}) = [\mathbf{y} - \mathcal{H}(\mathbf{x})]^T \mathbf{R}^{-1} [\mathbf{y} - \mathcal{H}(\mathbf{x})] + \left(\mathbf{x} - \mathbf{x}^b\right)^T \mathbf{P}^{b^{-1}} \left(\mathbf{x} - \mathbf{x}^b\right)$$

in which the observations $\mathbf{y}$, whose error covariance matrix is denoted by $\mathbf{R}$, are nonlinearly connected with the system state $\mathbf{x}$, by means of a nonlinear observation operator $\mathcal{H}$. The initial guess on the state, $\mathbf{x}^b$, also called the background, is characterized by the background error covariance matrix $\mathbf{P}^b$. Unless otherwise noted, all vectors are column vectors, and superscript T denotes transpostion.

The solution to this problem is sought iteratively, by constructing a sequence of $\mathbf{x}^k$ such that (Tarantola & Valette, 1982, Eq. 23)

$$\mathbf{x}^{k+1} = \mathbf{x}^b + \mathbf{P}^b\mathbf{H}^{k^T} \left(\mathbf{H}^k\mathbf{P}^b\mathbf{H}^{k^T} + \mathbf{R}\right)^{-1} \left[\mathbf{y} - \mathcal{H}\left(\mathbf{x}^k\right) + \mathbf{H}^k\left(\mathbf{x}^k - \mathbf{x}^b\right)\right] \tag{1}$$

or, equivalently (Tarantola & Valette, 1982, Eq. 24),

$$\mathbf{x}^{k+1} = \mathbf{x}^b + \left(\mathbf{H}^{k^T}\mathbf{R}^{-1}\mathbf{H}^k + \mathbf{P}^{b^{-1}}\right)^{-1} \mathbf{H}^{k^T}\mathbf{R}^{-1} \left[\mathbf{y} - \mathcal{H}\left(\mathbf{x}^k\right) + \mathbf{H}^k\left(\mathbf{x}^k - \mathbf{x}^b\right)\right], \tag{2}$$

starting in both cases with the initial guess $\mathbf{x}^0 = \mathbf{x}^b$. In these equations, $\mathbf{H}^k$ refers to the observation operator linearized about the current estimate,

$$\mathbf{H}^k = \left.\frac{\partial\mathcal{H}}{\partial\mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}^k}. \tag{3}$$

These two formulations of the iterative algorithm assume that data and background errors are independent. They are equivalent mathematically, but their effective cost differs, depending on the respective size of the data and state vectors, $n_\mathbf{y}$ and $n_\mathbf{x}$, respectively. Equation 2 should be preferred in situations where $n_\mathbf{y} \gg n_\mathbf{x}$, while Eq. 1 is more expedient if $n_\mathbf{y} \ll n_\mathbf{x}$. As will be clear form Sections 2.3 and 3.1, the state and observation vectors dimension considered here correspond to the latter situation. We therefore choose the algorithm defined by Eq. 1, for which, once convergence is reached, the uncertainty affecting the final estimate $\hat{\mathbf{x}}$ is given by the posterior covariance matrix

$$\hat{\mathbf{P}} = \left[\mathbf{I} - \mathbf{P}^b\hat{\mathbf{H}}^T \left(\hat{\mathbf{H}}\mathbf{P}^b\hat{\mathbf{H}}^T + \mathbf{R}\right)^{-1} \hat{\mathbf{H}}\right] \mathbf{P}^b, \tag{4}$$

where $\mathbf{I}$ is the identity matrix, and we understand that $\hat{\mathbf{H}}$ denotes the observation operator linearized about the final estimate of the state, $\hat{\mathbf{x}}$. The resolution matrix $\hat{\mathbf{S}}$, which relates to how well the final estimate $\hat{\mathbf{x}}$ resolves the true model state $\mathbf{x}^t$ given the observations, is defined as

$$\hat{\mathbf{S}} = \mathbf{P}^b\hat{\mathbf{H}}^T \left(\hat{\mathbf{H}}\mathbf{P}^b\hat{\mathbf{H}}^T + \mathbf{R}\right)^{-1} \hat{\mathbf{H}}. \tag{5}$$

In this nonlinear context, the interpretation of $\hat{\mathbf{S}}$ is subject to caution, as high-order derivatives of $\mathcal{H}$ are neglected in the analysis, in the hope that the final estimate is close to the true state. We will check on the validity of this assumption when dealing with synthetic data, in Section 2.6 below.

In standard geomagnetic field modeling, the prior information places regularization constraints on the sought solution, by setting

$$\mathbf{P}^b = \lambda^{-1}\mathbf{C}, \tag{6}$$

where λ is an adjustable 'damping' parameter and $\mathbf{C}$ a suitably chosen matrix, see e.g Bloxham et al. (1989), Eqs. 3.8–3.11. The novelty of our approach stands in the prior information we supply, which rests on the statistics of a numerical dynamo simulation, a dynamo norm, and in the ensemble approach that we adopt in order to assess the uncertainties of our estimate of the archeomagnetic field. We dedicate the next paragraphs to these aspects. We begin by the description of the numerical simulation, the choice of which in turn dictates the nature of the state vector $\mathbf{x}$.

2.2 A geodynamo model as source of prior information

The essential physics of the dynamo can be explained by the interaction between the fluid flow, magnetic field, temperature and composition within the core. Such interactions are governed by the magnetohydrodynamic (MHD) equations, which for the case of the Earth's core are often considered in the Boussinesq approximation due to the supposedly weak density stratification of the outer core (e.g. Braginsky & Roberts, 1995). Given the complexity of the dynamo system, the MHD dynamo equations are solved by a numerical approximation in a rotating spherical shell (consult Christensen & Wicht, 2015, for a recent review on numerical dynamo simulations). The parameter space reached by numerical simulations is, however, restricted by computational feasibility, resting still far away from Earth-like values (Glatzmaier, 2002). Despite this distance in parameters, many numerical simulations have been claimed to be Earth-like. To address the issue of semblance between simulations and the magnetic field of the Earth, Christensen et al. (2010) defined different morphological criteria to characterize the Earth-likeness of the output of numerical dynamo simulations. However, such criteria aim at describing only the static, morphological semblance with the Earth's magnetic field, leaving aside its dynamical aspect.

2.2.1 The Coupled Earth dynamo model

Although archeomagnetic field features have been shown to drift both to the west and the east at high latitudes, a systematic westward drift is observed around the equator, especially in the Atlantic hemisphere, most notably in historical data (Dumberry & Finlay, 2007; Finlay & Jackson, 2003). Such geographical concentration of the SV strongly suggests a persistent heterogeneous control from the upper and/or lower boundaries of the core. Furthermore, the evidence for seismic anisotropy at the top of the inner core (e.g. Souriau & Calvet, 2015, for a review) points to a hemispherical dichotomy in terms of inner core growth and therefore of buoyancy release at the inner core boundary (Monnereau et al., 2010; Alboussière et al., 2010). In light of these studies, Aubert et al. (2013) proposed a bottom-up control of the dynamo by the inner core. The model considers an heterogeneous buoyancy control at the inner core boundary (ICB), which, combined with the heterogeneous heat-flux control at the CMB, is responsible for a higher SV underneath the Atlantic hemisphere. Moreover, the model also includes the gravitational interaction of the lower mantle mass anomalies with the inner core, which results in large scale westward flow (and concurrent advection of field structures) beneath the CMB at low latitudes. This so-called 'Coupled-Earth' model is referred to as CE henceforth.

In this study we aim at investigating the influence of the statistical properties of a dynamo simulation, the CE model, on the estimation of the magnetic field at the CMB. To that end, we need to find the correspondence between the magnetic field strength and its time scales from the numerical simulation (which operates in a non-dimensional world) and those observed. As in Aubert et al. (2013), the time scaling of CE is done by calculating the model non-dimensional secular time-scale τ_m and normalizing it to the Earth's secular variation $\tau_\oplus \approx 415$ years (Lhuillier et al., 2011). The rescaling ratio $\tau_\oplus/\tau_m$ is found to be 56113 years. The magnetic field, on the other hand, is calibrated here by adjusting the root-mean-squared axial dipole of the simulation to that of the archeomagnetic model A_FM-0 of Licht et al. (2013), giving a magnetic field rescaling ratio of 850.6 μT. It is important to note that once the calibration is so defined, it leaves no room for an additional adjustable parameter factorizing the dynamo norm entering the estimation problem.

2.2.2 Statistical properties of the Coupled Earth dynamo model

After rescaling the CE dynamo simulation, we define two basic statistical characteristics of the numerical simulation, namely the mean and spread (covariance) of the radial magnetic field B_r at the CMB. Instead of considering the whole simulation interval, we define an ensemble of 820 states of the radial magnetic field at the CMB by randomly picking snapshots on a long sequence of the dynamo simulation, over an interval of about 100,000 years (the value of 820 ensures statistical convergence). The mean and spread of this ensemble projected in physical space are shown in Figure 1 for two spherical harmonic truncations (5 and 133, the latter being that of the native CE model). The average CE model shows an eccentric axial dipole, concentrated underneath northern Asia. Such feature is in agreement with recent and past configurations of the magnetic field; it results from the heterogeneous forcing from the inner core, stronger in the eastern hemisphere (Olson & Deguen, 2012). The mean also shows an invariance in scale, as seen by the similarity between Figs. 1a and 1c despite the different truncations. In contrast, the spread varies with the truncation, in morphology and intensity. As shown by Figs. 1b and 1d, it is mostly concentrated underneath the Earth's North-East quadrant for the larger scales, and around low latitudes in between 90°W and 90°E for the smaller scales. The biased distribution of the spread illustrates the effect of the combination of heterogeneous thermal couplings at the ICB and CMB. These two pieces of information, the mean and covariance of the radial magnetic field at the CMB, define the prior information that fuels the iterative procedure outlined in Section 2.1.

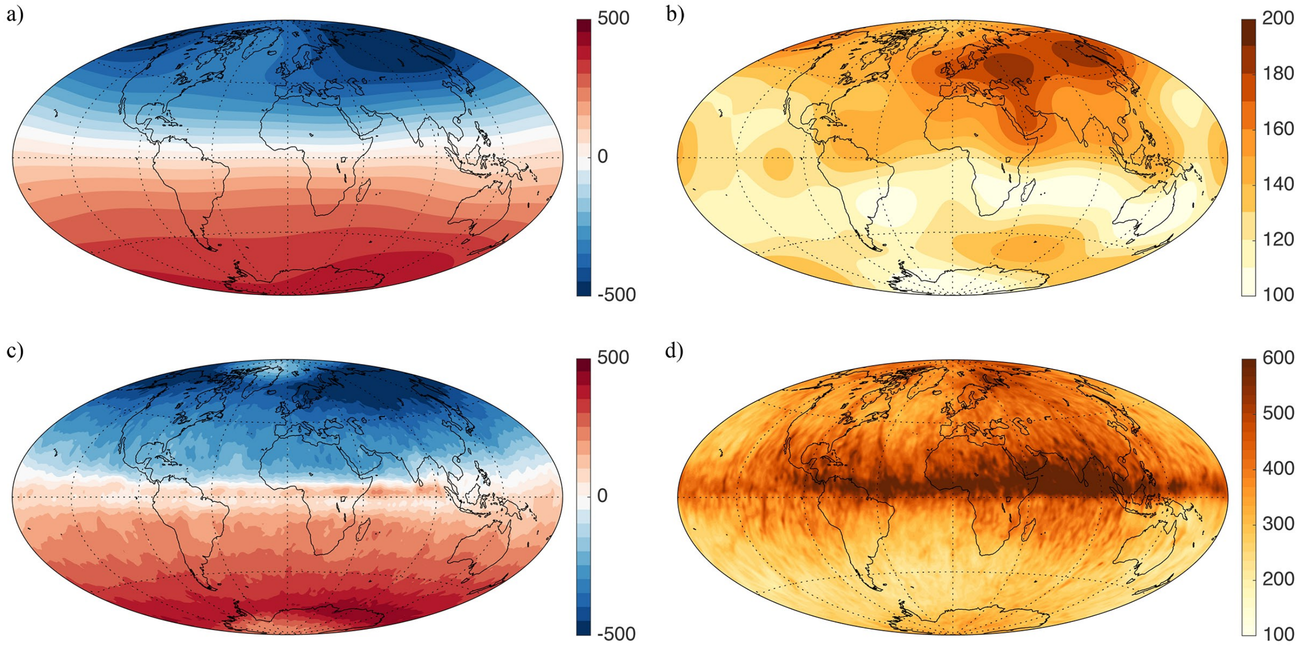


Figure 1. Mean and spread of the radial magnetic field at the CMB of the Coupled Earth (CE) dynamo model. a) and b) are respectively the mean and spread truncated at spherical harmonic degree $L = 5$, with contours every 50 μT; c) and d) are the mean and spread at the native model truncation, $L = 133$, with contours every 10 μT. All values in μT.

2.3 The state vector $\mathbf{x}$

Due to its solenoidal nature, the magnetic field vector $\mathbf{B}$ can be described in terms of its poloidal and toroidal components $\mathbf{B}_{\mathcal{P}}$ and $\mathbf{B}_{\mathcal{T}}$,

$$\mathbf{B}(r,\theta,\phi) = \mathbf{B}_{\mathcal{P}}(r,\theta,\phi) + \mathbf{B}_{\mathcal{T}}(r,\theta,\phi) = \nabla \times \nabla \times [\mathcal{P}(r,\theta,\phi)\mathbf{r}] + \nabla \times [\mathcal{T}(r,\theta,\phi)\mathbf{r}]\,, \tag{7}$$

where $\mathcal{P}$ and $\mathcal{T}$ are the poloidal and toroidal scalars and $\mathbf{r}$ is the radius vector (e.g. Dormy, 1997). This decomposition is used in the CE model, in which the poloidal and toroidal scalars are further projected onto a spherical harmonic (SH) basis to account for their horizontal dependency. The poloidal scalar at a given radius r is expressed as

$$\mathcal{P}(r,\theta,\phi) = \sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell} \mathcal{P}_\ell^m(r) Y_\ell^m(\theta,\phi), \tag{8}$$

where $Y_\ell^m = e^{im\phi} P_\ell^m(\cos\theta)$ is a complex-valued, fully normalized spherical harmonic of degree ℓ and order m, and L is the truncation of the expansion. As the toroidal field does not have a radial component, it is confined within the core (under the assumption that the mantle is an electrical insulator). For that reason, the only observable part of the core field at the Earth's surface is the poloidal field. We consequently define the (column) state vector of our system by

$$\mathbf{x} \equiv [\ldots, \mathcal{P}_\ell^m(r_c), \ldots]^T\,, \tag{9}$$

that is, the poloidal scalar in spectral space at the CMB (radius $r_c = 3485$ km), which is a native component of the state vector characterizing the full CE simulation, for all admissible values of degree ℓ and order m below the truncation of degree and order $L = 133$ ($n_{\mathbf{x}} = 9045$). We use the full CE truncation bearing in mind the future application of the present methodology to sequential archeomagnetic assimilation, a situation in which the full set of dynamo state variables define the model state vector.

As stated above, our state vector rests on a fully normalized, complex-valued spherical harmonic basis. Consequently, in order to connect its components with the standard Gauss coefficients g and h (whose definition rests on the use of real-valued spherical harmonics subject to the Schmidt convention for normalization), one must resort to the following relationships

$$g_\ell^m = \left(\frac{r_c}{r_a}\right)^{\ell+2} \frac{\ell\sqrt{2\ell+1}}{r_c} \mathrm{Re}\,[\mathcal{P}_\ell^m(r_c)], \tag{10}$$

$$h_\ell^m = -\left(\frac{r_c}{r_a}\right)^{\ell+2} \frac{\ell\sqrt{2\ell+1}}{r_c} \mathrm{Im}\,[\mathcal{P}_\ell^m(r_c)], \tag{11}$$

in which Re [] and Im [] denote the real and imaginary parts, and r_a is the mean radius of the Earth ($r_a = 6371.2$ km). If $\mathbf{g}$ represents the vector of Gauss coefficients, one can introduce a matrix $\mathbf{Q}$ to represent the above formula

$$\mathbf{g} = \mathbf{Q}\mathbf{x}. \tag{12}$$

2.4 The observation operator $\mathcal{H}$

As introduced above, the observation operator $\mathcal{H}$ connects the state $\mathbf{x}$ we just defined mathematically with the observations of the magnetic field, $\mathbf{y}$.

2.4.1 Linear functionals of B_r at the core surface

The full magnetic field vector observed at the surface of the Earth is described today by its X (north), Y (east) and Z (downward) components. Under the assumption of an insulating mantle, each component can be expressed as a linear functional of the magnetic field B_r at the CMB, by means of the convolution with the adequate Green function for the exterior Laplace problem (Gubbins & Roberts, 1983). For example, a measurement of Z at a location (r_0, θ_0, ϕ_0) at Earth's surface writes

$$Z(r_0, \theta_0, \phi_0) = \int_S G_Z(r_0, \theta_0, \phi_0 | r_c, \theta, \phi) B_r(r_c, \theta, \phi) \mathrm{d}S, \tag{13}$$

where G_Z is the Green function (or data kernel) for a Z datum (e.g. Bloxham et al., 1989; Constable et al., 1993) and S is the core surface of radius r_c. Using spherical harmonics, B_r at the core surface is expressed as

$$B_r(r_c, \theta, \phi) = \sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell} {B_r}_\ell^m(r_c) Y_\ell^m(\theta, \phi), \tag{14}$$

noting that we have already truncated the expansion according to the native resolution of the CE dynamo model. We can now connect each coefficient of this expansion with the poloidal scalar introduced above

$${B_r}_\ell^m(r_c) = \frac{\ell(\ell+1)}{r_c} \mathcal{P}_\ell^m(r_c), \tag{15}$$

which then allows us to write formally the link between $\mathcal{P}$ and the prediction for the datum Z,

$$Z(r_0, \theta_0, \phi_0) = \frac{1}{2r_c} \sum_{\ell=1}^{L} \sum_{m=0}^{\ell} \ell(\ell+1) \left[\mathcal{P}_\ell^m(r_c) {\mathcal{G}_Z}_\ell^{m\dagger} + \mathcal{P}_\ell^{m\dagger}(r_c) {\mathcal{G}_Z}_\ell^m \right], \tag{16}$$

where we have now restricted the sum to positive values of m only, and consequently introduced the † for conjugation (Z is real-valued). In this expression, ${\mathcal{G}_Z}_\ell^m$ is the coefficient of degree ℓ and order m of the spherical harmonic expansion of $G_Z(r_0, \theta_0, \phi_0 | r_c, \theta, \phi)$. The linear observation operator for a Z measurement, $\mathbf{H}_Z$, can then be defined as

$$\mathbf{H}_Z = \frac{1}{2r_c} \left[\ldots, \ell(\ell+1) {\mathcal{G}_Z}_\ell^{m\dagger}, \ldots, \ell(\ell+1) {\mathcal{G}_Z}_\ell^m, \ldots \right], \tag{17}$$

which is applied to the expanded state vector $\mathbf{x} = [\ldots, \mathcal{P}_\ell^m, \ldots, \mathcal{P}_\ell^{m\dagger}, \ldots]^T$, where m can only take positive values. The same formalism can be applied in order to define the operators $\mathbf{H}_X$ and $\mathbf{H}_Y$ for the North and East measurements of the magnetic field, respectively.

2.4.2 Nonlinear functionals of B_r at the core surface

An important amount of data from the past magnetic field consists of indirect and often incomplete information about the magnetic field vector. Such information is given in terms of the directions, declination D and inclination I of the vector, or its intensity F. The directions and intensity are nonlinear combinations of the magnetic field vector components (X, Y, Z),

$$D = \arctan \frac{Y}{X}, \tag{18}$$

$$I = \arctan \frac{Z}{\sqrt{X^2 + Y^2}}, \tag{19}$$

$$F = \sqrt{X^2 + Y^2 + Z^2}. \tag{20}$$

To calculate the linearized observation operator given by Equation 3, we have to bear in mind that in practice, this operator achieves a mapping from the complex-valued $\mathbf{x}$ to a real-valued measurement. Real functions of complex variables are not holomorphic, and their gradients cannot be consequently defined in standard complex analysis (Hjorungnes & Gesbert, 2007). However, a real-valued function f of a complex variable z can be written in terms of z and its complex conjugate $z^\dagger$, in order to get rid of a possible imaginary part of f. If we take z and $z^\dagger$ to represent independent variables, the differential of a functional $f(z, z^\dagger)$ is given by the Wirtinger derivatives

$$\mathrm{d}f = \frac{\partial f}{\partial z}\mathrm{d}z + \frac{\partial f}{\partial z^\dagger}\mathrm{d}z^\dagger, \tag{21}$$

where $\partial f/\partial z$ ($\partial f/\partial z^\dagger$) is the derivative of f with respect to a reference point for a fixed $z^\dagger$ (z).

Rewriting the expression for the declination in Equation 18, we have

$$D\left(\dots, \mathcal{P}_\ell^m, \dots, \mathcal{P}_\ell^{m\dagger}, \dots\right) = \arctan\left[\frac{\sum_{\ell,m}\frac{1}{2r_c}\ell(\ell+1)\left(\mathcal{P}_\ell^m {\mathcal{G}_Y}_\ell^{m\dagger} + \mathcal{P}_\ell^{m\dagger}{\mathcal{G}_Y}_\ell^m\right)}{\sum_{\ell,m}\frac{1}{2r_c}\ell(\ell+1)\left(\mathcal{P}_\ell^m {\mathcal{G}_X}_\ell^{m\dagger} + \mathcal{P}_\ell^{m\dagger}{\mathcal{G}_X}_\ell^m\right)}\right] \tag{22}$$

where $\sum_{\ell,m}$ is a simplified notation for the previous double summation over ℓ and m. Differentiating the previous equation with respect to $\mathcal{P}_{\ell'}^{m'}$, around $\mathcal{P}_{\ell',0}^{m'}$, we have that

$$\frac{\partial D}{\partial \mathcal{P}_\ell^m} = \frac{1}{1+(Y_0/X_0)^2}\left[\sum_{\ell',m'}\frac{\ell'(\ell'+1)}{2r_c}\delta_{\ell\ell'}\delta_{mm'}\left({\mathcal{G}_Y}_{\ell'}^{m'\dagger}X_0 - {\mathcal{G}_X}_{\ell'}^{m'\dagger}Y_0\right)\right]\frac{1}{X_0^2}, \tag{23}$$

where δ is the Kronecker symbol and X_0, Y_0 and Z_0 are the magnetic vector components obtained from the reference state $\mathbf{x}_0 = \left[\dots, \mathcal{P}_{\ell,0}^m, \dots, \mathcal{P}_{\ell,0}^{m\,\dagger}, \dots\right]^T$. Upon simplification, we see that

$$\frac{\partial D}{\partial \mathcal{P}_\ell^m} = \frac{1}{H_0^2}\frac{\ell(\ell+1)}{2r_c}\left(X_0{\mathcal{G}_Y}_\ell^{m\dagger} - Y_0{\mathcal{G}_X}_\ell^{m\dagger}\right), \tag{24}$$

in which $H_0 = \sqrt{X_0^2 + Y_0^2}$. Differentiating now with respect to $\mathcal{P}_{\ell'}^{m'\dagger}$

$$\frac{\partial D}{\partial \mathcal{P}_\ell^{m\dagger}} = \frac{1}{H_0^2}\frac{\ell(\ell+1)}{2r_c}(X_0{\mathcal{G}_Y}_\ell^m - Y_0{\mathcal{G}_X}_\ell^m). \tag{25}$$

The linearization of the nonlinear declination operator, $\mathcal{H}_D(\mathbf{x})$ around $\mathbf{x}_0$, following Equation 3, which we define formally as

$$\left.\frac{\partial \mathcal{H}_D}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_0} \equiv \mathbf{H}_{D,0},$$

can be written in terms of its Wirtinger derivatives as

$$\mathbf{H}_{D,0} = \frac{1}{2r_cH_0^2}\left[\dots, \ell(\ell+1)\left(X_0{\mathcal{G}_Y}_\ell^{m\dagger} - Y_0{\mathcal{G}_X}_\ell^{m\dagger}\right), \dots, \ell(\ell+1)\left(X_0{\mathcal{G}_Y}_\ell^m - Y_0{\mathcal{G}_X}_\ell^m\right), \dots\right]. \tag{26}$$

In the course of the iterative scheme (Eq. 1), this operator is to be applied to a generic incremental state vector of the form

$$\mathrm{d}\mathbf{x} = \left[\dots, \mathrm{d}\mathcal{P}_\ell^m, \dots, \mathrm{d}\mathcal{P}_\ell^{m\dagger}, \dots\right]^T.$$

Using the same reasoning, we find the following expressions for the linearized inclination and intensity operators, $\mathbf{H}_{I,0}$ and $\mathbf{H}_{F,0}$:

$$\begin{aligned}\mathbf{H}_{I,0} = \frac{1}{2r_cF_0^2}\Big[\quad & \dots\quad, \ell(\ell+1)\left(H_0{\mathcal{G}_Z}_\ell^{m\dagger} + \frac{Z_0}{H_0}(X_0{\mathcal{G}_X}_\ell^{m\dagger} + Y_0{\mathcal{G}_Y}_\ell^{m\dagger})\right), \\ & \dots\quad, \ell(\ell+1)\left(H_0{\mathcal{G}_Z}_\ell^m + \frac{Z_0}{H_0}(X_0{\mathcal{G}_X}_\ell^m + Y_0{\mathcal{G}_Y}_\ell^m)\right), \dots\Big],\end{aligned} \tag{27}$$

$$\begin{aligned}\mathbf{H}_{F,0} = \frac{1}{2r_cF_0}\Big[\quad & \dots\quad, \ell(\ell+1)\left(X_0{\mathcal{G}_X}_\ell^{m\dagger} + Y_0{\mathcal{G}_Y}_\ell^{m\dagger} + Z_0{\mathcal{G}_Z}_\ell^{m\dagger}\right), \\ & \dots\quad, \ell(\ell+1)\left(X_0{\mathcal{G}_X}_\ell^m + Y_0{\mathcal{G}_Y}_\ell^m + Z_0{\mathcal{G}_Z}_\ell^m\right), \dots\Big].\end{aligned} \tag{28}$$

2.5 An ensemble approach

We finish this methodological preamble by explaining how we define in practice the background state $\mathbf{x}^b$ and the background error covariance matrix $\mathbf{P}^b$ which must be supplied to the iterative algorithm. We resort to an ensemble of N_e state vectors $\mathbf{x}_e$ taken randomly from a $100,000$-yr long integration of the coupled Earth dynamo model. Our estimator is the ensemble mean, and the background state is therefore the average of these N_e states,

$$\mathbf{x}^b = \frac{1}{N_e}\sum_{e=1}^{N_e}\mathbf{x}_e. \tag{29}$$

As already stated above, we take $N_e = 820$ such states in order to get statistical convergence. Let us recall that $\mathbf{x}^b$ is a

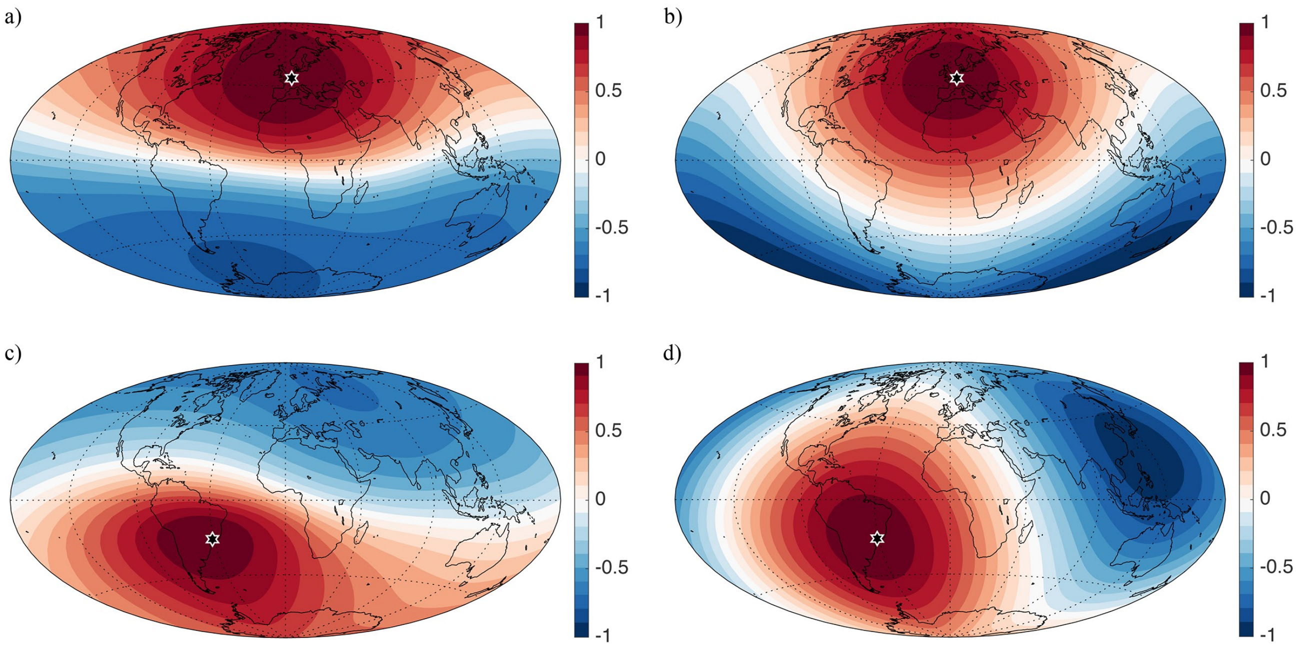


Figure 2. Correlations implied by the prior information between the radial magnetic induction at different core surface points. Top row: correlations underneath Paris (shown with the star) and any point on the core surface for a) prior information supplied by the Coupled Earth dynamo and b) Prior information supplied by the dissipation norm. Bottom row: same, save that correlations are expressed with respect to a point underneath São Paulo for c) CE dynamo norm prior and d) dissipation norm prior.

complex-valued field containing the spherical harmonic representation of the poloidal scalar at the core surface. We therefore define the background covariance matrix as

$$\mathbf{P}^b = \frac{1}{N_e - 1} \sum_{e=1}^{N_e} \left(\mathbf{x}_e - \mathbf{x}^b\right) \left(\mathbf{x}_e - \mathbf{x}^b\right)^{\dagger T}, \tag{30}$$

noting that $\mathbf{P}^b$ is a complex-valued, Hermitian matrix.

Implicit is the assumption that these two moments suffice to characterize the prior information. This is only true, however, when dealing with Gaussian statistics. Previous studies indicate that the large-scale magnetic field produced by a buoyancy-driven dynamo is indeed close to possess Gaussian statistics (Fournier et al., 2011, their fig. 4). In order to gain insight into the type of prior information supplied by $\mathbf{P}^b$, we introduce the linear operator $\mathbf{M}$ which maps $\mathbf{x}$ to a vector $\mathbf{v}_B$ containing the values of B_r on a grid approximating the core surface,

$$\mathbf{v}_B = \mathbf{M}\mathbf{x}. \tag{31}$$

The matrix $\mathbf{W} \equiv \mathbf{M}\mathbf{P}^b\mathbf{M}^{\dagger T}$ then provides us with the covariances contained in $\mathbf{P}^b$ expressed in physical space, and we can express the correlation ρ_{ij} between B_r at the i-th grid point and B_r at the j-th grid point as follows

$$\rho_{ij} = \frac{W_{ij}}{\sqrt{W_{ii}W_{jj}}}. \tag{32}$$

The first column of Figure 2 shows mappings of the correlations given by ρ between core surface grid-points at specific locations (Paris and São Paulo) and every grid-point of the CMB. For comparison, the figure also shows in its second column these correlations when the dynamo-based $\mathbf{P}^b$ is replaced by the minimum dissipation matrix $\mathbf{C}$ of Bloxham et al. (1989), recall Eq. 6 above. The prior information supplied by the Coupled Earth dynamo favors antisymmetry with respect to the equator, whereas the dissipation norm implies correlations whose geometry reflects that of an inclined dipole piercing through Paris or São Paulo.

The posterior covariance is calculated by solving the iterative algorithm for each member of the ensemble, as would be be required in a sequential data assimilation framework based on the EnKF (e.g. Evensen, 2003). If $\hat{\mathbf{x}}_e$ denotes the estimate for each ensemble member e, then one can get $\hat{\mathbf{P}}$ by computing

$$\hat{\mathbf{P}} = \frac{1}{N_e - 1} \sum_{e=1}^{N_e} \left(\hat{\mathbf{x}}_e - \hat{\mathbf{x}}\right) \left(\hat{\mathbf{x}}_e - \hat{\mathbf{x}}\right)^{\dagger T}, \tag{33}$$

where $\hat{\mathbf{x}}$ is the estimator (the mean). For that expression to give a $\hat{\mathbf{P}}$ equal to the one provided by Equation 4, the iterative algorithm to apply to each initial $\mathbf{x}_e$ must write

$$\mathbf{x}_e^{k+1} = \mathbf{x}_e + \mathbf{P}^b\mathbf{H}^{k\dagger T}\left(\mathbf{H}^k\mathbf{P}^b\mathbf{H}^{k\dagger T} + \mathbf{R}\right)^{-1}\left[\mathbf{y} + \boldsymbol{\epsilon}_e - \mathcal{H}\left(\mathbf{x}^k\right) + \mathbf{H}^k\left(\mathbf{x}^k - \mathbf{x}_e\right)\right], \tag{34}$$

where $\mathbf{x}^k$ is the ensemble mean at the k-th step of the algorithm, and $\mathbf{H}^k$ is the observation operator linearized about this ensemble mean. Note that for each ensemble member, the data is noised by an amount $\boldsymbol{\epsilon}_e$ drawn from a normal distribution of zero mean and covariance $\mathbf{R}$ (Burgers et al., 1998).

Finally, let us note that instead of opting for a linearized approach, one could simply compute each state estimate $\hat{\mathbf{x}}_e$ using the nonlinear observation operator, according to

$$\hat{\mathbf{x}}_e = \mathbf{x}_e + \left(\mathbf{HP}^b\right)^{\dagger T}\left(\mathbf{HP}^b\mathbf{H}^{\dagger T} + \mathbf{R}\right)^{-1}\left[\mathbf{y} + \boldsymbol{\epsilon}_e - \mathcal{H}\left(\mathbf{x}_e\right)\right], \tag{35}$$

with

$$\mathbf{HP}^b = \frac{1}{N_e - 1}\sum_{e=1}^{N_e}\left[\mathcal{H}(\mathbf{x}_e) - \mathcal{H}\left(\mathbf{x}^b\right)\right]\left[\mathbf{x}_e - \mathbf{x}^b\right]^{\dagger T}, \text{ and} \tag{36}$$

$$\mathbf{HP}^b\mathbf{H}^{\dagger T} = \frac{1}{N_e - 1}\sum_{e=1}^{N_e}\left[\mathcal{H}\left(\mathbf{x}_e\right) - \mathcal{H}\left(\mathbf{x}^b\right)\right]\left[\mathcal{H}\left(\mathbf{x}_e\right) - \mathcal{H}(\mathbf{x}^b)\right]^T \tag{37}$$

computed based on the ensemble of states. This strategy has the advantage of expediency, since it only requires one iteration. In the context of archeomagnetism, it appears to be slightlty less accurate than the iterative one (see Section 2.6 below).

2.6 Validation

With archeomagnetic applications in mind, we proceed with a simple test considering the retrieval of the large scales of the magnetic field at the CMB from synthetic data. The synthetic dataset consists of magnetic data at 35 locations spread in a close-to-uniform fashion at Earth's surface, a nearly ideal (and admittedly unrealistic) scenario for the distribution of archeomagnetic data, shown in Figure 3a. The number of data locations corresponds to the amount of full vector observations necessary to recover the magnetic field up to SH degree $L = 5$ (35 Gauss coefficients). A reference model state $\mathbf{x}^t$ is used to produce synthetic vector observations at the given locations, to which a 1 μT value is introduced for the data uncertainty. The (X, Y, Z) observations are transformed into (D, I, F) and the uncertainties are propagated and noised with a random component drawn from a normal distribution (whose standard deviation is obtained from the uncertainty propagation). The observation error covariance matrix $\mathbf{R}$ is diagonal, with each diagonal element equal to the variance of each synthetic observation.

In order to quantify the efficacy of the scheme we use three diagnostics for the estimates: the normalized data misfit

$$M(\mathbf{x}) = \sqrt{\frac{1}{n_\mathbf{y}}\left[\mathbf{y} - \mathcal{H}(\mathbf{x})\right]^T\mathbf{R}^{-1}\left[\mathbf{y} - \mathcal{H}(\mathbf{x})\right]}, \tag{38}$$

the normalized spread (i.e. the standard deviation) of the ensemble

$$S\left(\mathbf{x}_e\right) = \sqrt{\frac{1}{n_\mathbf{x}}\frac{1}{N_e - 1}\sum_{e=1}^{N_e}\left(\mathbf{x}_e - \langle\mathbf{x}_e\rangle\right)^{\dagger T}\mathbf{P}_d^{b\,-1}\left(\mathbf{x}_e - \langle\mathbf{x}_e\rangle\right)}, \tag{39}$$

and the normalized error with respect to the known true state $\mathbf{x}^t$

$$E(\mathbf{x}) = \sqrt{\frac{1}{n_\mathbf{x}}\left(\mathbf{x} - \mathbf{x}^t\right)^{\dagger T}\mathbf{P}_d^{b\,-1}\left(\mathbf{x} - \mathbf{x}^t\right)}, \tag{40}$$

where $\mathbf{P}_d^b$ is the diagonal matrix assembled from the diagonal elements of $\mathbf{P}^b$. It is important to note that the spread and error are evaluated throughout this paper up to spherical harmonic degree 5, i.e. over that part of the state vector which is effectively constrained by observations (as will be clear in Section 3.3). The evolution of these quantities with iteration count can be seen in Figure 4, which shows that the algorithm converges after three iterations. The misfit M decreases from 5 to around 0.8 with the iterative scheme and 0.95 with the nonlinear one described at the end of Section 2.5, which points to the better fit to the dataset of the iterative solution. The iterative solution is also slightly superior in terms of spread S and error E, which represent the uncertainty level of the model and its distance with respect to the truth, respectively. It is interesting to note the agreement of 20% for the final estimates of S and E, suggesting that the ensemble spread is a good proxy for the posterior model error. Additionally, the fast convergence of the iterative scheme indicates the weak nonlinear character of the problem at hand, which gives us confidence in the ability of the algorithm to deal with the true archeomagnetic dataset and validates the interpretation of the resolution matrix defined in Equation 5.

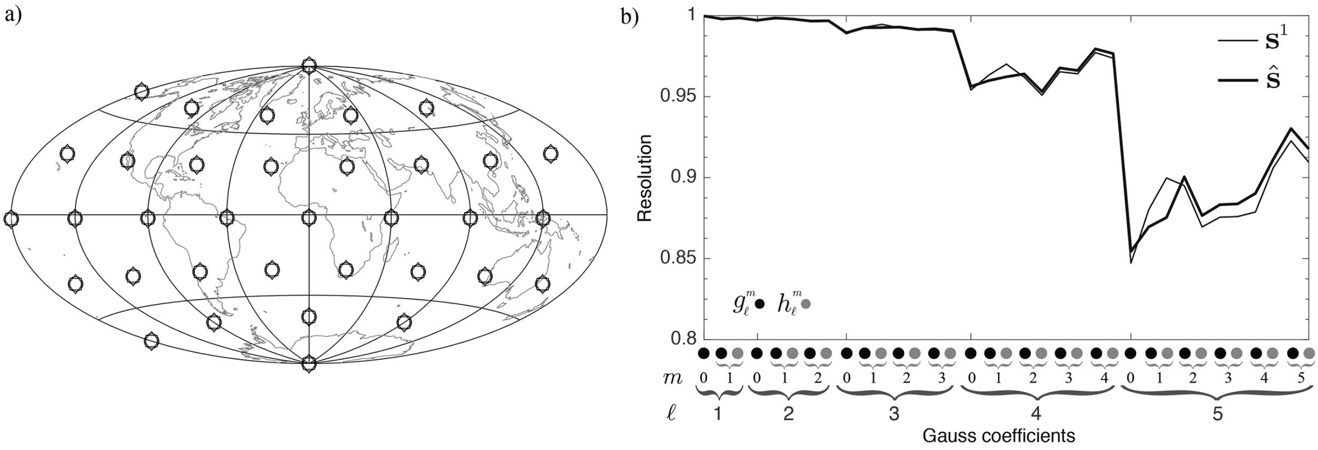


Figure 3. Synthetic directional and intensity dataset and its resolution in model space. a) Distribution of synthetic observations at Earth's surface; b) resolution of Gauss coefficients up to degree 5. See text for details.

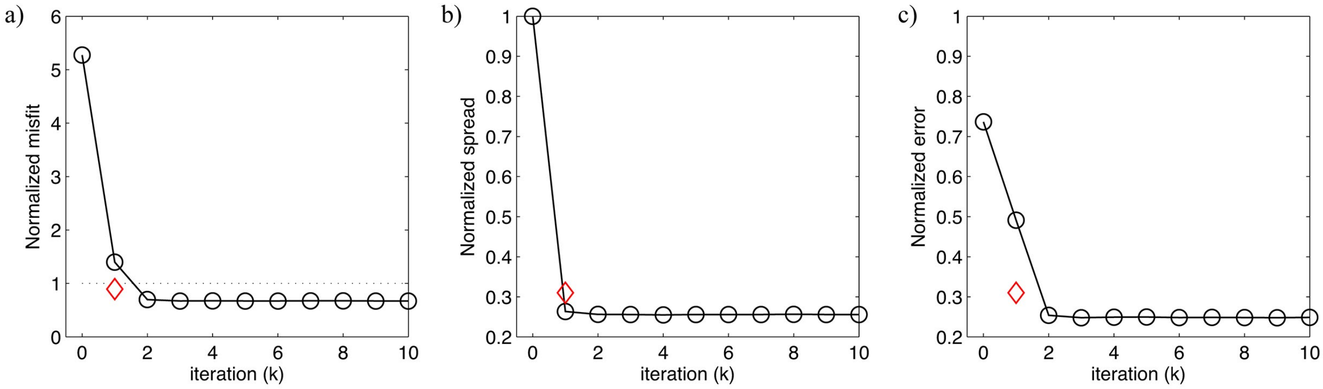


Figure 4. Convergence of the iterative solutions for the validation test in terms of normalized a) misfit, b) spread of the ensemble and c) error of the mean estimate. The black curve shows the iterative estimates by means of the linearized observation operator approach, while the red diamond corresponds to the non-linear observation operator-based analysis. See text for details.

The resolution matrix for the dataset displayed in Fig. 3a is close to diagonal, which is why we will only discuss its information content in terms of its diagonal elements. Since the resolution matrix of Eq. 5 represents the ratio between the covariances in the background model and data ($\hat{S}_i \approx {\sigma_i^b}^2/({\sigma_i^b}^2 + {\sigma_y}^2)$ for a model state variable i), we analyze our results in terms of its square-root $\hat{\mathbf{S}}^{1/2}$, thereby dealing with standard deviations instead of variances. Projecting the diagonal elements of the square-root of the resolution matrix into Gauss coefficients through the $\mathbf{Q}$ operator introduced in Section 2.3, we can introduce a resolution vector $\hat{\mathbf{s}}$ constructed from the diagonal elements of $\mathbf{Q}\hat{\mathbf{S}}^{1/2}\mathbf{Q}^{\dagger T}$. We show this resolution vector up to degree 5 in Figure 3b. Also plotted is the resolution vector at the first iteration $\mathbf{s}^1$, which closely resembles the resolution vector of the final estimate, $\hat{\mathbf{s}}$. The resolution vector coefficients are of order unity up to degree 3, meaning that the data constrain the corresponding Gauss coefficients almost to 100%. Degree 4 and 5 are constrained up to 97% and 90%, respectively. The relative lack of resolution of the $m = 0$ zonal coefficients (such as g_4^0 and g_5^0), reflects the small amount of data underneath polar regions, precisely where associated Legendre functions of order 0 reach their maximum amplitudes. Since the resolution vector is seen not to change considerably from the first to the last iteration, we will consider the resolution vector at the first iteration, $\mathbf{s} = \mathbf{s}^1$, as representative of the model resolution throughout this paper. The interest in this choice is that $\mathbf{s}^1$ depends only on the data and prior information, allowing us to appreciate the resolution before performing the iterative estimation per se.

The sum of the elements of the vector $\hat{\mathbf{s}}$, corresponding to the trace of the square-root of the resolution matrix, is 33. This figure corresponds to the number of resolved degrees of freedom of the model, given the dataset in question. Note that an independent validation of the results discussed in this paragraph was carried out by a brute force calculation of the tangent observation operator, for a state vector made directly of Gauss coefficients. The weakly nonlinear character of the problem at hand as well as the satisfactory behavior of the scheme in this academic test prompt us to exercise it further in the more realistic and challenging context of archeomagnetic observations.

3 ARCHEOMAGNETIC RESOLUTION ANALYSIS

3.1 Archeomagnetic database

The archeomagnetic dataset used in this proof-of-concept paper relies on the Geomagia50.V2 database (Donadini et al., 2009) for the period from 1200 BC to 2000 AD. Geomagia50.V2 consists of directional and intensity data spanning the past 50 millennia that derive from the analysis of the remanent magnetization carried by volcanic deposits and archeological artifacts, like bricks, pottery, ovens, etc. This version of the database is not the newest one, but is sufficiently mature to help illustrate the fundamental aspects of the method we designed. Using it allows us to compare our findings with models previously published relying on practically the same dataset.

The dataset clearly shows that the available archeomagnetic data present a very uneven temporal and geographical distribution. Spatially, most of the data are located in Western Eurasia. In fact, almost all observations are located in the Northern Hemisphere, with just 2.5% of the data in the Southern one (Donadini et al., 2009; Licht et al., 2013). At any given epoch, we therefore rely on the prior information supplied by the dynamo to recover the field behavior in the Southern hemisphere, based on the spatial correlations of the prior (recall Figs. 2a and 2c). Regarding the data, we follow a hybrid strategy based on Korte et al. (2009) and Licht et al. (2013) in order to better compare the different models, attributing lower bounds for the data uncertainties and asserting uncertainties to data without such information. We arbitrarily choose a lower bound of 2 μT for intensity data uncertainties and 3.4° for the α_{95} parameter uncertainties, following the modeling errors assigned by Licht et al. (2013) as a consequence of the degree 5 truncation of their modeling strategy. This implies a lower bound for the standard deviation of $\sigma_I = (81/140) \times 3.4°$ for the inclination I and $\sigma_D = (81/140 \cos I) \times 3.4°$ for the declination D. Data with no uncertainty information are assigned an uncertainty of 8.25 μT for intensity and 4.5° for α_{95}, which amounts to 150% of the average of known uncertainties, as in Licht et al. (2013).

With regard to ages, results are dated either by radiogenic techniques, generally the case for volcanic data, or by means of archeological constraints, in the case of archeological artifacts. These different dating techniques result in different probability density functions for the age, Gaussian for radiogenic dating or uniform for archeological dating (e.g. Ramsey, 2009). Age uncertainties have been treated in different ways in archeomagnetic field modeling. The most recent global models of Korte & Constable (2011), Licht et al. (2013) and Pavón-Carrasco et al. (2014) use different flavors of bootstrap sampling to account for errors in age estimations (jointly with measurement errors), while, in a regional context, Hellio et al. (2014) resort to a Monte Carlo Markov Chain to sample the space of possible ages of a regional dataset.

In the single-epoch approach outlined in this study, we do not treat the archeomagnetic field model temporal dependency. Instead, we group the data within bins of 40-year width, assuming that the field is constant over that period. The width of 40 years is a compromise between the shortest time possible (for the assumption of steadiness to hold) and the longest one, which will favor a larger number of data (and should therefore improve the accuracy of the model). Over 40 years, one can expect that the harmonic components of the field of degree $\ell \leq 5$ will not undergo significant changes, since their typical time scales τ_ℓ are thought to vary like $\tau_\oplus/\ell$, with a master secular variation coefficient $\tau_\oplus$ of order 400 years (Lhuillier et al., 2011).

3.2 Data redistribution based on age uncertainties

A first dataset can be constructed by taking the mean age of each datum, ignoring the age uncertainties, and rejecting these data for which the age uncertainty is either not provided or is strictly larger than ±100 years (total 200 years). Such selected and reworked Geomagia50.V2 data is labeled here as the N (native) dataset. Figure 5a shows the histogram of the N dataset comprising 2528 declination, 3757 inclination and 2230 intensity data distributed within the 40-year bins through the time interval from 1200 BC to 2000 AD, with thus a total of 8515 data.

Within our snapshot framework, a possibility to deal with age uncertainties would be to draw an ensemble of datasets based on the age pdf of each datum, and to apply our framework to each draw. The final archeomagnetic field model (for each time bin) would then consist of a weighted average of the ensemble of models, following the procedure detailed by Lanos (2004) for the construction of regional archeomagnetic curves. Although we are planning to use that strategy in the future, we opt here for a more expedient (and less accurate) strategy of redistribution, whereby each datum is distributed in the bins where its age pdf is not zero, with a weight given by the pdf integrated over each bin. For a given datum, the reworked uncertainty inside the j-th bin b_j, defined by the time window $[t_i(b_j), t_f(b_j)]$, is described by

$$\sigma_{b_j} = \sigma_o \times w_{b_j}^{-1/2}, \tag{41}$$

where σ_o is the original uncertainty of the data and the weight w_{b_j} is defined by

$$w_{b_j} = \int_{t_i(b_j)}^{t_f(b_j)} f(\sigma_o, t_o, t)\mathrm{d}t, \tag{42}$$

where $f(\sigma_o, t_o, t)$ is a function representing the pdf associated with the datum's age uncertainty (Gaussian or boxcar), t_o is the expected age and t is time. If the redistributed data are counted with respect to their new weights in each bin, the result

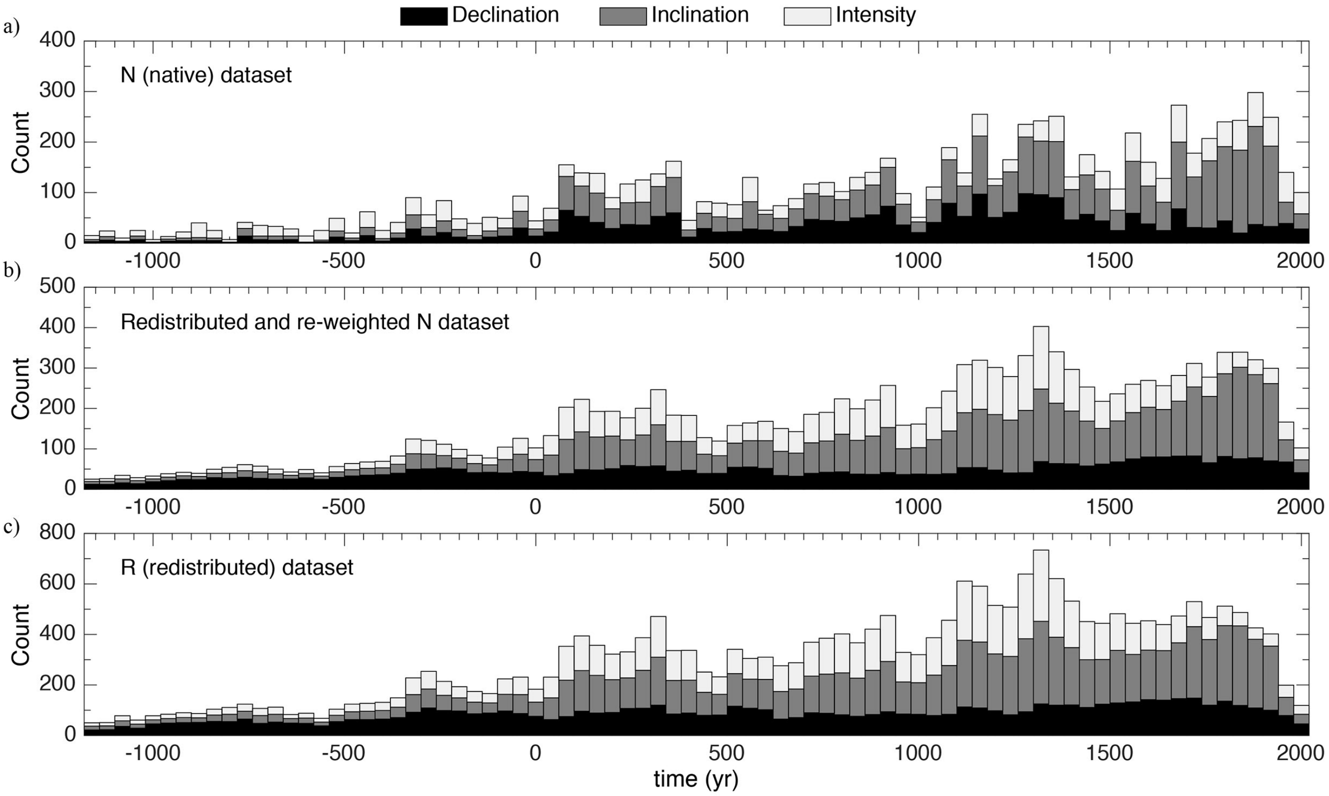


Figure 5. Histograms of the archaeomagnetic datasets used in this study. a) The native (N) dataset constructed from Geomagia50.V2 (Donadini et al., 2009); b) Redistribution of the N dataset given the re-weighting based on the pdf of the age uncertainty of each datum, counted by their weights; c) The redistributed (R) dataset, which assumes that the redistributed data are independent, with a recalibrated measurement uncertainty. See text for details.

is a smoother distribution of the N dataset, as can be seen in Figure 5b. The data redistribution and uncertainty re-weighting of the native N dataset leads to a new data group, called R (for redistributed) henceforth. If each redistributed datum in each bin is considered as a discrete independent datum, the R dataset comprises 7432 declination, 10055 inclination and 6904 intensity data, for a total of 24391 data, whose histogram is shown in Figure 5c.

Although the present archeomagnetic field estimation method does not include a regularization of the model in time, it is important to remark that the grouping of data in 40-year bins, the assignment of lower bounds to the uncertainties and the further redistribution of the dataset (meaning the transformation of N into R), act effectively as a temporal regularization.

3.3 Resolution of the archeomagnetic dataset

Before applying our iterative scheme to the N and R archeomagnetic datasets, let us analyze the impact of the data under consideration on our knowledge of the magnetic field at the core surface. We have seen in Section 2.1 that the resolution matrix given by Equation 5 quantifies the influence of the properties of the dataset on the model. For simplicity, we resort to the resolution vector **s** after the first iteration (recall the end of Section 2.6) to synthesize the information contained in the matrix and to represent the resolving power of the archeomagnetic dataset in each 40-year interval from 1200 BC to 2000 AD.

The resulting sequence of resolution vectors, which we will call the archeomagnetic resolution matrices throughout, based both on the N and R datasets, represents the evolution of the archeomagnetic resolution in time, and is shown in Figure 6. The similarity between both matrices in Fig. 6a and 6b in terms of amplitude and behavior shows that the increase of the uncertainties by the data re-weighting of dataset R is mitigated by the increase in the number of data. The smoother character of the R-based resolution matrix is a consequence of the better distribution of the data given their age uncertainties information, and therefore, of its effective temporal regularization.

The key aspect of Fig. 6 is that it demonstrates the inability of the present archeomagnetic dataset to resolve the magnetic field at the core surface for spatial scales beyond spherical harmonic degree 5. Due to the smoother character of the R dataset-based archeomagnetic resolution matrix, the following description will focus on Fig. 6b. For the whole 3,000 years interval, only the dipole and quadrupole are continuously well resolved. The axial dipole g_1^0 is the best constrained coefficient, resolved around 90% from 1200 BC to 0 AD and up to almost 100% from 0 AD to 2000 AD. The other components of the dipole,

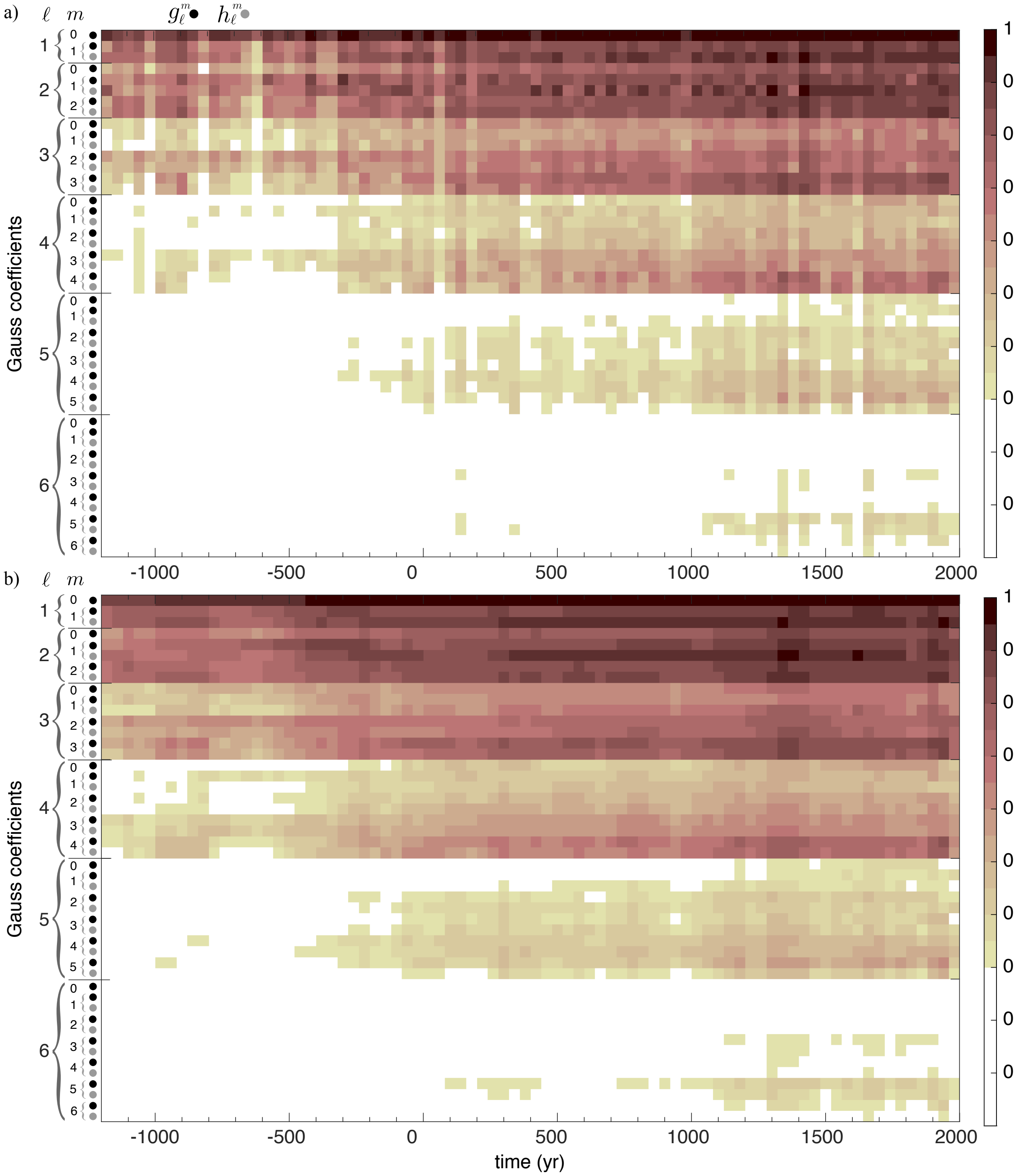


Figure 6. Archeomagnetic resolution matrices, given the Coupled Earth dynamo prior and a) the N or b) the R dataset. The color scale represents the normalized resolving power of the data on the model state variables, cast here in terms of Gauss coefficients, up to spherical harmonic degree 6.

together with the quadrupole terms, evolve from roughly 70% up to 85% after 0 AD. The octupole ($\ell = 3$) shows a resolution of 50% before 0 AD, while after 0 AD the resolution increases up to around 75% at present-day. The degree 4 starts presenting resolution lows, meaning that at times the resolution of some coefficients is lower than an arbitrary value of 30%. The period after 0 AD presents a reasonable resolution of 50% for degree 4 and after 1000 AD a resolution around 40% for degree 5. Beyond degree 5 the model variables display little to no sensitivity to the datasets. For all coefficients of degree higher than the quadrupole, the general rule is that the zonal modes ($m = 0$), are not well constrained within the corresponding degree, due to the lack of observations in polar regions. In contrast, sectoral modes ($m = \ell$) are best resolved within a given harmonic degree family, due to their better sampling of lower latitudes, which coincides with the data spatial distribution.

In synthesis, we observe a good resolution of the archeomagnetic field up to degree 3 for the period between 1200 BC

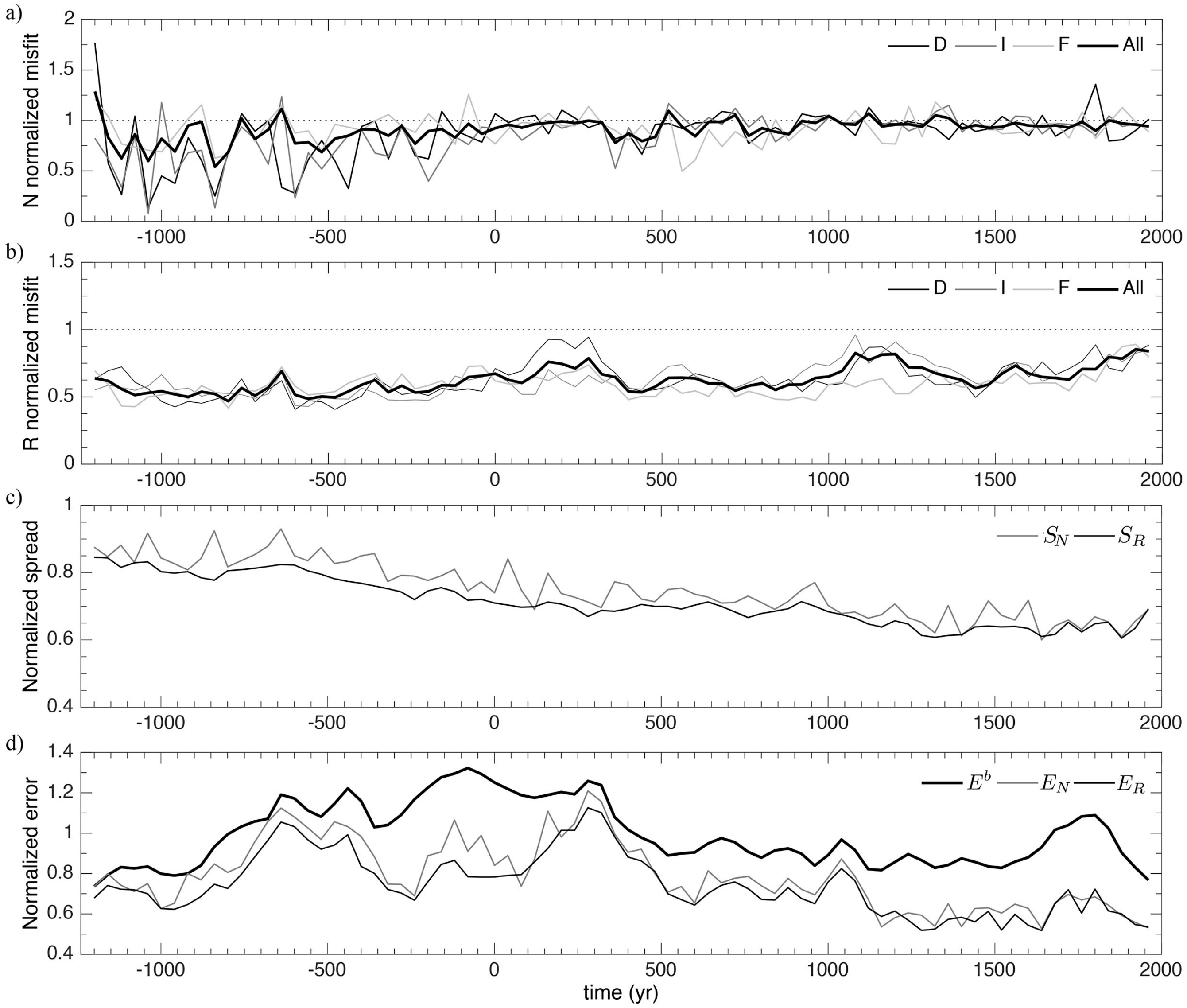


Figure 7. Diagnostics of the twin experiments in terms of a) the normalized misfit M of the final estimation for the synthetic N dataset b) normalized misfit M of the final estimate for the synthetic R dataset, c) normalized total spread S for both estimates and d) the normalized total error E of the background and final estimate for the synthetic N and R datasets. Note that S and E are calculated considering a degree 5 truncation.

and 0 AD, degree 4 from 0 AD to 1000 AD, and degree 5 from 1000 AD onwards. The trend corresponding to the increase in resolving power in time contained in both matrices is related mainly to the increase of data availability in the interval, as can be seen by comparing Figures 5a and 6a and Figures 5c and 6b. It is important to note that the archeomagnetic resolution matrices shown in Fig. 6 represent an idealization of the model resolution by the data, in the sense that they do not depend on data values themselves, and therefore, do not consider the existence and impact of incoherent data, which are likely to be present in the real archeomagnetic dataset (e.g. Licht et al., 2013). In order to test the robustness of the conclusions drawn from the analysis of the archeomagnetic resolution matrices and to prepare the application of the scheme to real datasets, we next design archeomagnetic twin experiments, whereby we can directly assess error on the state estimates based on synthetic analogs of the N and R datasets.

3.4 Twin experiments

In this twin experiment, like in the validation test of Section 2.6, synthetic data are generated from a reference model simulation, the "true" state, which can be directly compared with the estimates of the iterative scheme. The true state evolves in time, since it is the product of the integration of the coupled Earth dynamo, and for each 40-year interval it is calculated as the average over 40 years of this integration. In this case, the two archeomagnetic datasets N and R are tested to assess the validity of the data regularization by the redistribution and re-weighting based on the ages uncertainties. The synthetic observations are based on the original mean age, geographical distribution and age and data uncertainties from the two real datasets. The observations are further noised based on a random draw from the data uncertainty distribution. The noised synthetic observations are then used to estimate the true state given the background model.

We proceed with the iterative estimation by applying an outlier rejection scheme, whereby data with misfits larger than 3 standard deviations after the first estimation are discarded. A second estimation is then produced with the remaining dataset. Following this strategy, 22 data were excluded from the N dataset (12 declination, 4 inclination and 6 intensity data) while 12 data were rejected from the R one (8 declinations and 4 inclinations).

The quality assessment of our estimates is done by using the three diagnostics introduced in Section 2.6, namely the normalized data misfit M, the normalized spread of the ensemble S and the normalized error E. Those are shown in Figure 7. Fig. 7a reveals a mean misfit of 0.9 for the N-based estimate. Considering that the background mean misfit is close to 3 (based on the average field of figure 1a), the decrease amounts to 70%. Further inspection of Fig. 7a indicates a stronger variability of the misfit during the first millennium BC than that prevailing since 0 AD. The R-based misfit shown in Fig. 7b reveals a mean of 0.6, which, compared to the background mean misfit of 1.75, reflects a decrease of 63%. The smaller misfit values, and weaker variability of the R-based estimates compared with the N-based ones result from the increase in data uncertainties implied by the redistribution scheme.

The normalized spread S is shown in Fig. 7c. The spread level is similar for both N and R-based estimates, around 85% at the beginning of the interval, with an almost linear decrease down to 65% at 2000 AD. This decreasing trend reflects the increase in data quantity with time (recall the histograms of Figs. 5a and 5c). The spread of the R-based estimate displays a smoother variability than the N-based one.

Variations of the normalized error E are shown in Fig. 7d. The background error E^b fluctuates since the true state dynamically evolves over the time interval, while the background state remains the same in our single-epoch analysis approach. E^b exhibits long-term millennial scale fluctuations, between 0.8 and 1.2. Although the normalized error for both N and R based estimations (E_N and E_R) presents an irregular behavior through the interval, the mean error value decreases from 1 for E^b to approximately the same value of 0.8 for the N and R based estimations (interestingly, these values correspond to the mean values of the normalized spread we just discussed). The detailed temporal behavior reflects the combination of the time-dependency of the true state on the one hand and of the varying distribution and quality of the data on the other hand. Sometimes, the initial guess (the background) is rather close to the truth, and the data does not strongly correct the estimate. In contrast, like around 0 AD and around 1800 AD, the initial guess is at times far from being optimal, and a substantial benefit is then drawn from the data.

4 STANDARD VERSUS DYNAMO-BASED ARCHEOMAGNETIC FIELD MODELS

We now move to the application of our iterative approach to the real N and R archeomagnetic datasets. Using the 3σ-rejection scheme discussed above, we discard 634 data from the N dataset (114 declinations, 302 inclinations and 218 intensities) and 826 data from the R dataset (136 declinations, 369 inclinations and 321 intensities). Figure 8 shows the normalized misfit M for the final estimations based on the N and R datasets. We see in Fig. 8a that the N misfit does not differ substantially from that computed for the twin experiments, showing again a larger degree of variability during the first millennium BC (especially during its first half). The decrease from background to final estimation in misfit is of 57% in this case, lower than that of the twin experiment. This lower performance in the misfit decrease indicates that the real archeomagnetic dataset is penalized by a number of data whose “real” values lie outside their error bars. The R dataset normalized misfit in Fig. 8b shows an almost constant value over the whole interval, as seen in the corresponding twin experiment in Fig. 7b, albeit at an higher level (close to 0.9, to contrast with 0.6), and a decrease of 44% in misfit. In synthesis, the differences between Fig. 8 and Fig. 7 point to the presence of a certain amount of outliers in the archeomagnetic dataset (see also Licht et al., 2013).

As the error of the models cannot be directly assessed, we choose the estimate based on the smoother R dataset as our reference model in order to compare it with inverse models built using a similar database (Geomagia50.V2). For convenience, we use the tag AmR to refer to our archeomagnetic model based on the dataset R and the CE dynamo background, and the tag Bm to refer to our background model. Figures 9 and 10 present groups of Gauss coefficients from AmR compared to models ARCH3k.1 from Korte et al. (2009) and A_FM-0 and A_FM from Licht et al. (2013). A brief description of the models is provided in the following:

ARCH3k.1: The ARCH3k.1 model consists of an inverse model built under adjustable temporal and spatial regularization truncated to degree 10 using lower bounds for data uncertainties (i.e. $\sigma = 5$ μT for intensities and $\alpha_{95} = 4.3°$ for directions) and an outlier rejection scheme. The model is derived from an inversion beginning at 2000 BC, the product of which is not considered between 2000 BC and 1000 BC in order to avoid spurious spline-end effects. The opposite end of the model is drawn towards the *gufm1* model (Jackson et al., 2000) after 1650 AD for all non-axial dipole coefficients and after 1840 AD for

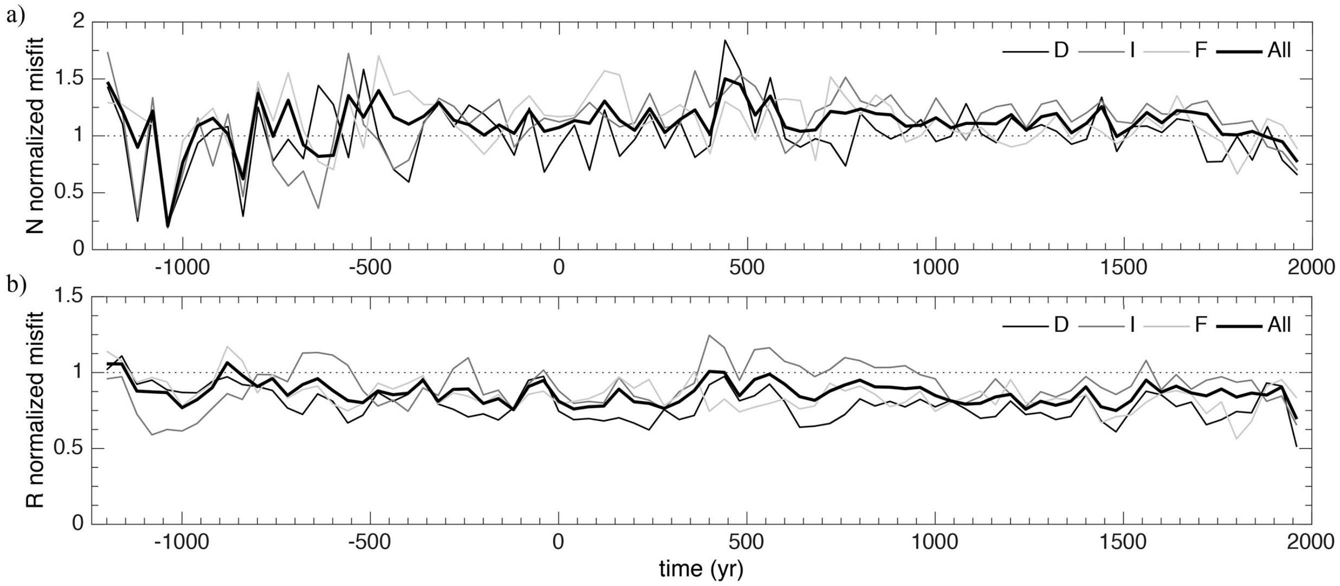


Figure 8. Normalized misfit M for a) the N and b) the R dataset analysis.

the axial dipole coefficient. Since *gufm1* is based on historical, observatory and satellite data, the last 500 years of ARCH3k.1 result in a more precise estimate in comparison to the rest of the time interval. The model rests on a bootstrap to account for the different dataset errors in order to provide a posteriori errors.

A_FM: The A_FM models are inverse models under the same general regularization approach as in ARCH3k.1. Acknowledging the fact that the available database does not allow to resolve field characteristics beyond degree 5, the field models are truncated to degree 5. Instead of applying lower bounds to data errors (which implies penalizing data of good quality), modeling errors due to the lower degree truncation are introduced in the data errors. Following a philosophy of using all information provided by the data, the model applies an outlier re-weighting instead of a rejection scheme. While A_FM-0 is a direct inversion of the A_dat dataset (restricted to volcanic and archeological data), the A_FM model consists of an ensemble of models generated by a similar bootstrap technique as in Korte et al. (2009).

AmR: The AmR model is the result of an inverse scheme regularized by a non-adjustable spatial dynamo norm provided by the CE dynamo, using the full truncation of the latter. With respect to ARCH3k.1 and A_FM, it considers a hybrid approach to the handling of lower bounds/modeling errors relative to the data uncertainties (see Subsection 3.1). The redistributed R dataset, from which the model is derived, is described in Section 3.2.

The differences between ARCH3k.1, A_FM and AmR are illustrated in Fig. 9 for the dipole and quadrupole terms and Fig. 10 for coefficients of degree 3 and above. AmR exhibits frequencies higher than those displayed by ARCH3k.1, A_FM, and A_FM-0, pointing to a weaker effective temporal regularization. As can be seen in Fig. 9a, the four models concur on the intensity decrease of the axial dipole g_1^0 over the past millennium. The axial dipole shows a considerable decrease in the standard deviation from the background model Bm to the posterior model AmR, which could be expected from the archeomagnetic resolution matrix (recall the first line in Fig. 6b). The general agreement of all models for g_1^0 reflects its control by the dataset. Such a statement can also be made for the equatorial dipole coefficients, g_1^1 and h_1^1 in Figs. 9b and 9c. The uncertainty levels of the (g_1^0, g_1^1, h_1^1) triplet of the AmR model are in line with those predicted by A_FM and ARCH3k.1, except over 1600 AD-2000 AD when the latter is constrained by *gufm1*. Again, this agreement highlights the robust characterization of model uncertainties, strongly controlled by the characteristics of the common dataset. The quadrupole coefficients (shown in Fig. 9d-h) present more variability between models, in particular during the first millennium BC. In this case, the disagreement underlines the larger influence of the prior information supplied to the inversion scheme.

Higher order coefficients are exhibited in Figure 10. They essentially show that before 0 AD, AmR is not properly updated from the background model Bm, both in terms of its mean value and respective uncertainty level. This indicates a poor sensitivity of the model to the data; see for instance the zonal coefficient g_4^0 in Fig.10c. Here, as in many other instances in Fig. 10, the estimate is strongly driven by the CE dynamo model. Periods of model-driven versus data-driven coefficient estimates are easy to assess, and they follow the information conveyed by the archeomagnetic resolution matrix (recall Figure 6b). In contrast, such distinction is less straightforward to make for ARCH3k.1 and A_FM. For instance, they do not show any significant increase of uncertainty estimates at times of paucity of data (see for instance the first 1000 years of g_4^0 in Fig. 10c). Since these models prioritize fitting the data, they are prone to displaying a variability relatively higher than

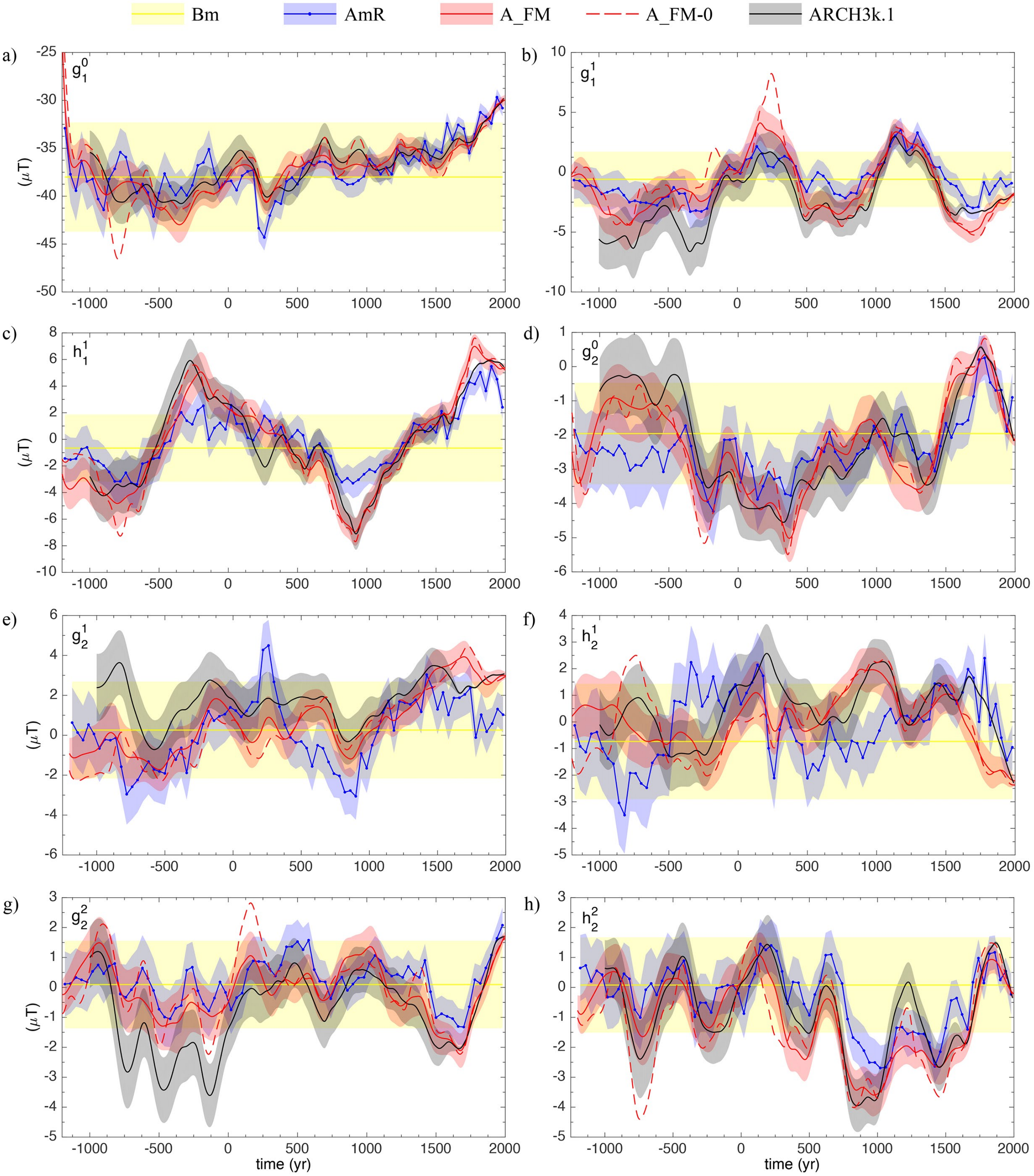


Figure 9. Gauss coefficients of the dipole and quadrupole magnetic field at the CMB. The background model Bm (yellow) is shown together with model AmR (blue). Both estimations are shown in terms of their mean and standard deviation calculated from the ensemble spread. Also shown are the AFM-0 model (red dashed curve), AFM-M model (red curve and shading) of Licht et al. (2013) and ARCH3k.1 model (black curve and shading) of Korte et al. (2009).

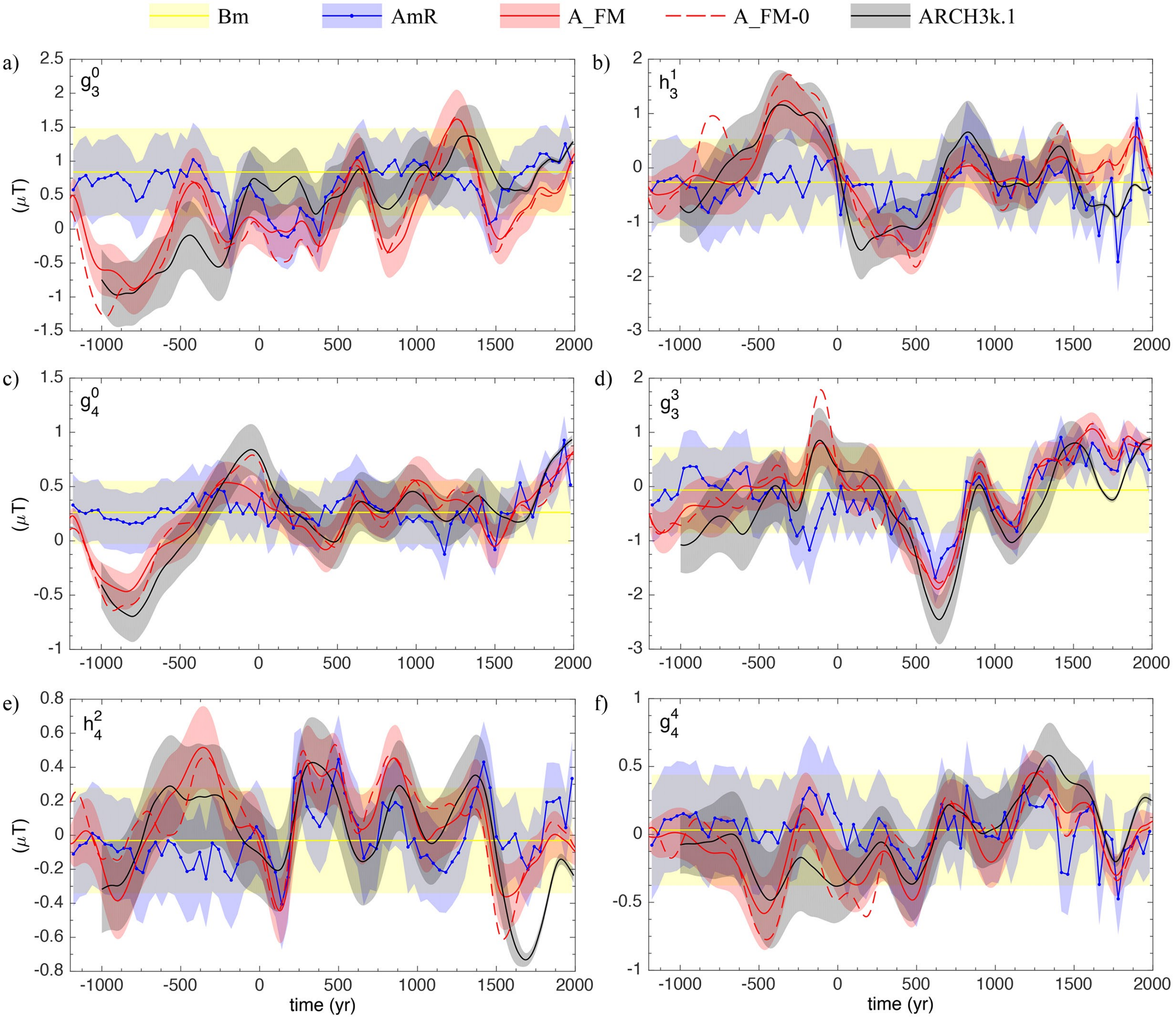


Figure 10. Continuation of Fig. 9 for higher order Gauss coefficients.

that of our estimate, which on the contrary favors geomagnetic quiescence (within the background uncertainty) at periods poorly documented by observations.

In order to further discuss the accuracy of the models in light of the available data, we show in Figure 11 model predictions at Paris and São Paulo. In Fig. 11, left column, we report the direction and intensity data lying inside a circular region of 300 km radius around Paris. All models display a rather good agreement with the data, and a similar behavior after 0 AD. The most significant difference among models is observed for intensity during the first millennium BC. This reflects a period poorly constrained by data, which puts consequently more weight to the prior chosen to construct any given model. Posterior model uncertainties vary considerably among models, regardless of the epoch. As stressed above, this is a consequence of different strategies to handle posterior model uncertainties: while the error on AmR is estimated from the standard deviation of the ensemble spread, ARCH3k.1 provides the error by propagating the coefficient uncertainties, and A_FM directly adds modeling errors to the ensemble standard deviation.

A different situation is found in Fig. 11, right column, where we show the model projections at São Paulo. Each model provides different evolutions, clearly in relation with the lack of data in the Southern hemisphere. This is particularly evident for the declination: both ARCH3k.1 and A_FM give large amplitude fluctuations on centennial to millennial time scales, while AmR predictions favor geomagnetic quiescence, namely weak fluctuations around the background state, within the prior uncertainties. This contrasting behavior must arise from the regional propagation of information by means of the prior correlations previously discussed in Section 2.5 and shown in Fig. 2.

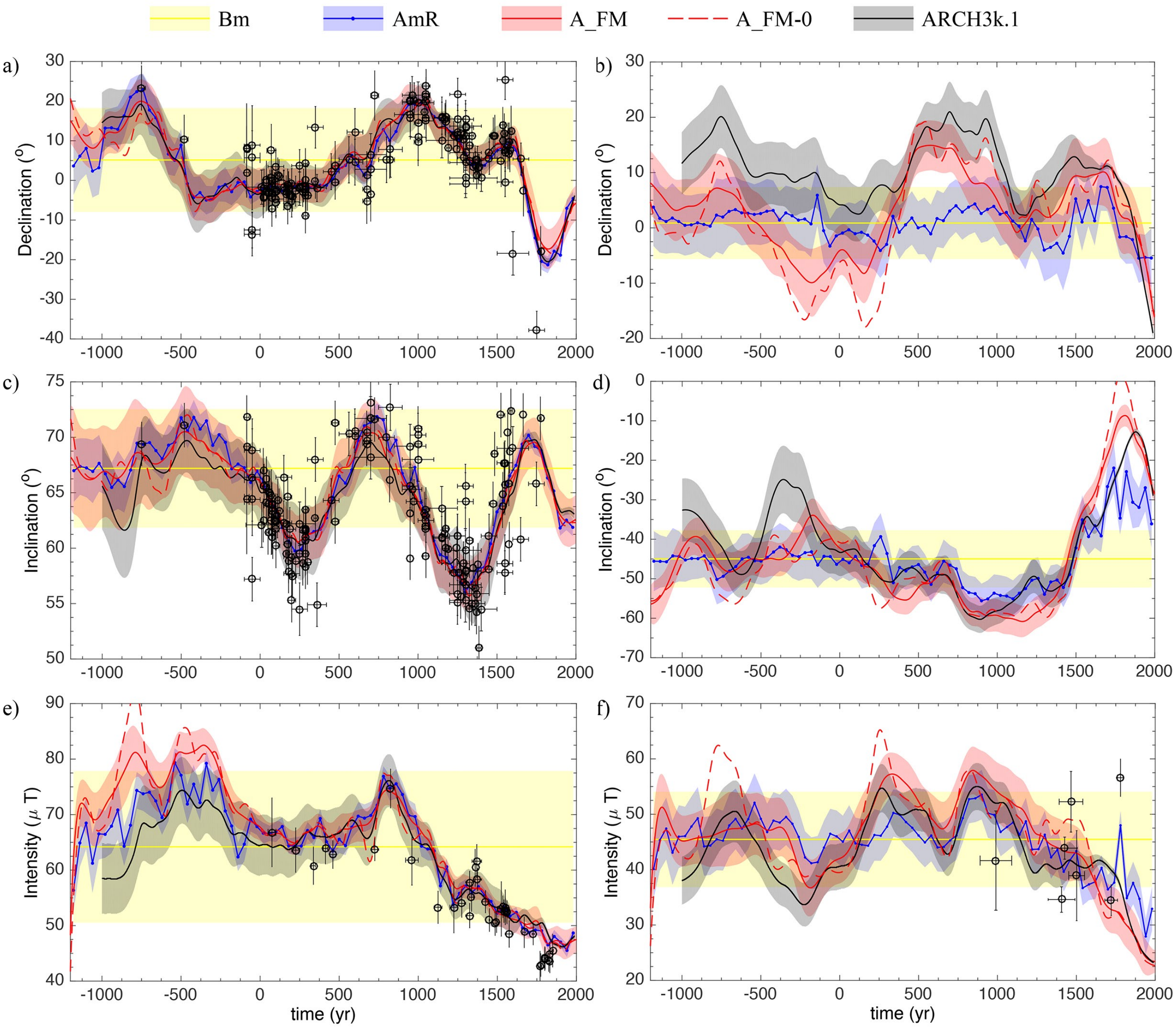


Figure 11. Projection of the ARCH3k.1 (black line with gray shadings), A_FM-M (red line with shadings), AFM-0 (red dashed line), AmR (blue line with shading) and Bm (yellow line with shading) models in PAris and São Paulo, and comparison with the available archeomagnetic data. a) Declination, c) inclination and e) intensity at Paris (location 48.85°N 2.35°E), for which data was collected over a disk of radius 200 km centered on Paris. b) Declination, d) inclination and f) intensity at São Paulo (location 23.54°S 46.63°W), for which data was collected over a disk of radius 300 km centered on São Paulo.

Finally, a comparison of the radial induction B_r mapped at the CMB averaged over a time window of width 40 years centered on 1700 AD between A_FM, ARCH3k.1 and AmR is shown in Figure 12. We use that specific time interval in order to compare the situation in which one of the models, ARCH3k.1 (Fig. 12a), is regularized by *gufm1*, while A_FM and AmR are built exclusively with archeomagnetic data. The three models show northern high latitude flux lobes underneath North America and Asia. Yet, the intensity and the location of the lobes predicted by AmR differ from those given by the other two models. In particular, the eastward offset of the American flux lobe in AmR might be interpreted as the influence of an eastward pull by the prior (see Fig. 1a), for which an intense flux patch is localized underneath North-East Asia. While the intensity of AmR seems reduced in comparison with the others, the general morphology of the field remains the same in all models. Still, the main differences between A_FM and ARCH3k.1 on the one hand, and AmR on the other hand occur mostly in the Southern hemisphere, as shown in Figs. 12b and 12d. Differences with ARCH3k.1 are more pronounced around Southernmost America and the South Pacific (but also on the Northernmost and central west Pacific), while the largest differences with A_FM occur in South Africa and the southern Indian ocean. The magnitude of these regional discordances are within the characteristic posterior uncertainties of the AmR model, whose standard deviation is mapped in Fig. 12f.

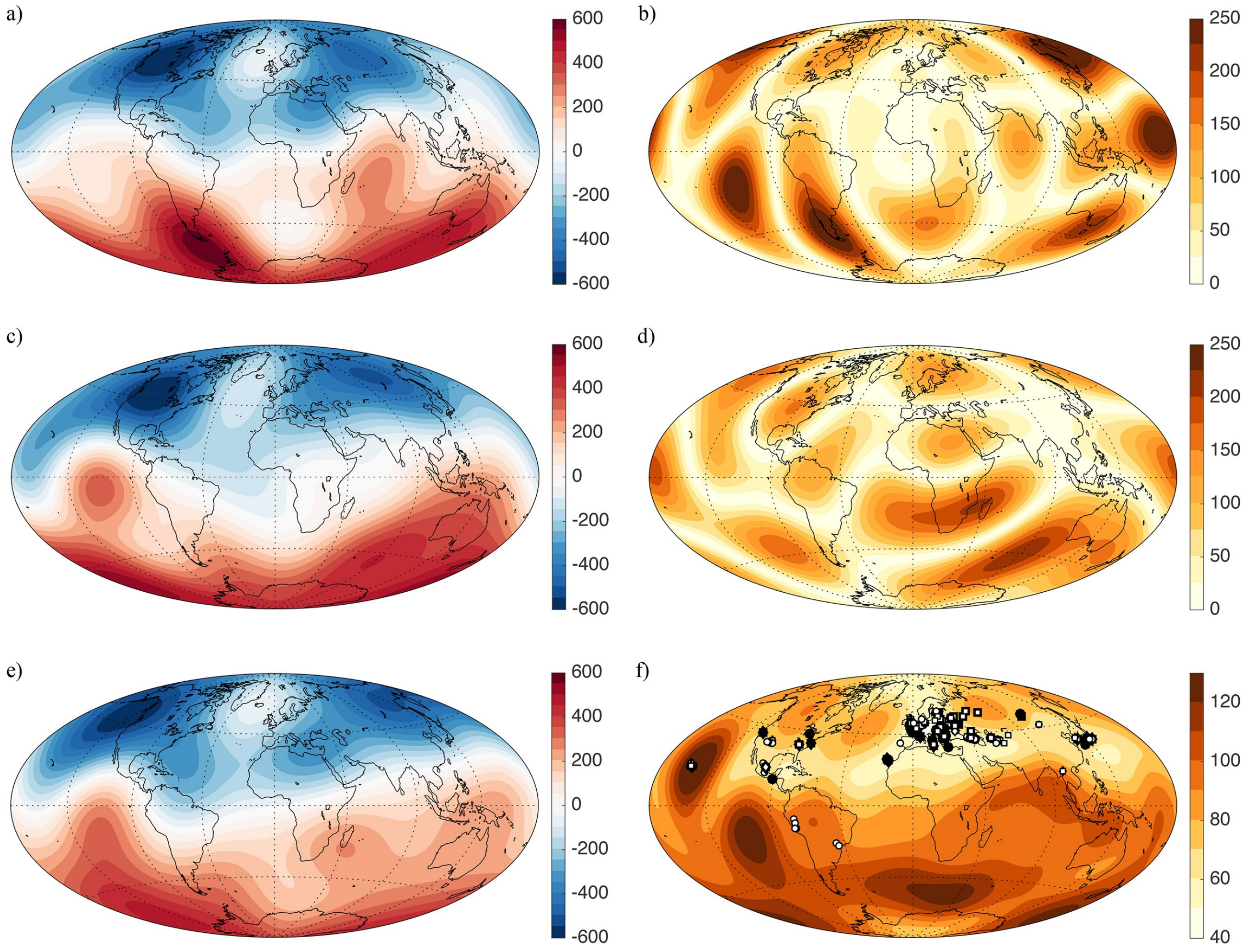


Figure 12. Different models of the magnetic field at the CMB for the time interval between 1680 AD and 1719 AD and its differences and uncertainties. a), c) and e) show the average field for the 40 year interval from the inverse models ARCH3k.1, AFM-M and our AmR model respectively, with contours every 50 μT; b) shows the absolute difference $||$ARCH3k.1$-$AmR$||$ while d) shows the absolute difference $||$AFM-M$-$AmR$||$, with contours every 25 μT; f) the standard deviation associated to the ensemble spread of AmR, with contours every 10 μT, overlapped by the R dataset showing redistributed data, with high uncertainties, (in white) and non redistributed data, with small uncertainties (in black). All values in μT.

5 SUMMARY AND PERSPECTIVES

The main objective of this study is to present a framework for modeling the archeomagnetic field at single epochs, in which a set of directional and intensity data is complemented by the prior information from a geodynamo simulation. Here, the background information is represented by the mean and covariance of an ensemble of magnetic field states extracted from the long integration of a dynamo model, in this instance the Coupled Earth model (Aubert et al., 2013). These statistics are used to define a spatial norm that enters the inverse archeomagnetic problem. The dynamo norm considered in this work contrasts with the regularization norm based on minimum Ohmic dissipation used in standard inverse archeomagnetic models. As shown in Fig. 2a and 2c, the dynamo norm places strong antisymmetrical constraints on the sought solution, as a consequence of the dominant axial dipolarity of the geodynamo. On the contrary, standard regularization norms lead to the spatial correlations implied by a dipole whose axis passes through the observation site. Besides, once the simulation outputs are rescaled, the physics-based dynamo norm needs no further adjustment of the trade-off between goodness-of-fit and model complexity.

One of the most important aspects of using a dynamo norm in the inverse archeomagnetic problem is that it enables a straightforward computation of the resolution matrix associated with archeomagnetic data. In this study we use the GEOMAGIA50.V2 database (Donadini et al., 2009) to ease comparison with previously published archeomagnetic field models. Since these field models are constructed using some form of temporal regularization, we perform a redistribution of the archeomagnetic dataset in a discretized time interval from 1200 BC to 2000 AD, based on age uncertainties, thereby creating the R dataset. The resulting archeomagnetic resolution matrix, shown in Fig. 6, allows us to distinguish, for each coefficient, data-driven from prior-driven periods. The dipole, quadrupole, and to some extent the octupole, appear well resolved over the whole period. From 0 AD to 2000 AD, the degree 4 is resolved as well, whereas from 1000 AD onward, resolution is achieved up to degree 5. In both cases, an enhanced resolution is observed for sectoral coefficients. Beyond degree 5, the archeomagnetic dataset has almost no influence on coefficient estimates. It is noteworthy that the recent update of the database (GEOMAGIA50.V3 Brown et al., 2015) does not alter significantly these conclusions (results not shown).

We choose the R dataset to build up a preliminary model, referred to as AmR. Comparison of Gauss coefficients predicted by AmR and ARCH3k.1 (Korte et al., 2009), AFM and AFM-0 (Licht et al., 2013) is in line with what is to be expected from the archeomagnetic resolution matrix, despite the presence of outliers in the dataset. In general, when coefficients are resolvable from the data, similar estimates are obtained, regardless of the model and methodology. Most differences are observed for those coefficients which are prior-driven. In the same way, pointwise predictions are in rough agreement in those regions where data are abundant, whereas stark differences are found in poorly documented areas.

In this study, attention was drawn to a sequence of independent snapshot analyses of the magnetic field through the last 3,000 years, assuming a constant background state. The next step will be to consider the temporal aspect of this scheme, and therefore the estimation of the state via a sequence of analysis cycles. More specifically, in a data assimilation framework such as the EnKF, an ensemble of states will be forecasted using the underlying numerical model, and this ensemble of forecasts will be used in conjunction with data (whenever they are available) to produce an ensemble of analyses, whose mean will hopefully be closer to the true state of Earth's core. In parallel, this novel approach will be strengthened by our ability to mitigate the impact of the strongly heterogeneous properties of the catalog of data at our disposal.

ACKNOWLEDGEMENTS

We are grateful to Sacha Brun, Andy Jackson and Erwan Thébault for fruitful exchanges and to Olivier Sirol for his assistance in the design of the calculations. This work has been partially supported by the French “Agence Nationale de la Recherche” under the grant ANR-11-BS56-011. Numerical computations were performed at S-CAPAD, IPGP, France and using HPC resources from GENCI-IDRIS (Grants 2014-042122 and 2015-042122). This is IPGP contribution xxxx.

References

Alboussière, T., Deguen, R., & Melzani, M., 2010. Melting-induced stratification above the Earth's inner core due to convective translation, *Nature*, **466**(7307), 744–747.

Aubert, J., 2014. Earth's core internal dynamics 1840–2010 imaged by inverse geodynamo modelling, *Geophysical Journal International*, p. ggu064.

Aubert, J. & Fournier, A., 2011. Inferring internal properties of Earth's core dynamics and their evolution from surface observations and a numerical geodynamo model, *Nonlinear Processes In Geophysics*, **18**(5), 657–674.

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

Bloxham, J. & Jackson, A., 1992. Time-dependent mapping of the magnetic field at the core-mantle boundary, *Journal of Geophysical Research: Solid Earth (1978–2012)*, **97**(B13), 19537–19563.

Bloxham, J., Gubbins, D., & Jackson, A., 1989. Geomagnetic secular variation, *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pp. 415–502.

Braginsky, S. I. & Roberts, P. H., 1995. Equations governing convection in Earth's core and the geodynamo, *Geophysical & Astrophysical Fluid Dynamics*, **79**(1), 1–97.

Brown, M. C., Donadini, F., Korte, M., Nilsson, A., Korhonen, K., Lodge, A., Lengyel, S. N., & Constable, C. G., 2015. Geomagia50.v3: 1. General structure and modifications to the archeological and volcanic database, *Earth, Planets and Space*, **67**(1), 1–31.

Burgers, G., van Leeuwen, P. J., & Evensen, G., 1998. Analysis scheme in the ensemble Kalman filter, *Monthly weather review*, **126**(6), 1719–1724.

Christensen, U. & Wicht, J., 2015. 8.10 - numerical dynamo simulations, in *Treatise on Geophysics (Second Edition)*, pp. 245 – 277, ed. Schubert, G., Elsevier, Oxford, second edition edn.

Christensen, U. R., Aubert, J., & Hulot, G., 2010. Conditions for Earth-like geodynamo models, *Earth and Planetary Science Letters*, **296**(3), 487–496.

Constable, C. G., Parker, R. L., & Stark, P. B., 1993. Geomagnetic field models incorporating frozen-flux constraints, *Geophysical Journal International*, **113**(2), 419–433.

Donadini, F., Korhonen, K., Riisager, P., & Pesonen, L. J., 2006. Database for Holocene geomagnetic intensity information, *Eos, Transactions American Geophysical Union*, **87**(14), 137–143.

Donadini, F., Korte, M., & Constable, C., 2009. Geomagnetic field for 0–3 ka: 1. New data sets for global modeling, *Geochemistry, Geophysics, Geosystems*, **10**(6).

Dormy, E., 1997. *Modélisation numérique de la dynamo terrestre*, Ph.D. thesis.

Dumberry, M. & Finlay, C. C., 2007. Eastward and westward drift of the Earth's magnetic field for the last three millennia, *Earth and Planetary Science Letters*, **254**(1), 146–157.

Evensen, G., 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean dynamics*, **53**(4), 343–367.

Finlay, C., Maus, S., Beggan, C., Bondar, T., Chambodut, A., Chernova, T., Chulliat, A., Golovkov, V., Hamilton, B., Hamoudi, M., et al., 2010. International geomagnetic reference field: the eleventh generation, *Geophysical Journal International*, **183**(3), 1216–1230.

Finlay, C. C. & Jackson, A., 2003. Equatorially dominated magnetic field change at the surface of Earth's core, *Science*, **300**(5628), 2084–2086.

Fournier, A., Hulot, G., Jault, D., Kuang, W., Tangborn, A., Gillet, N., Canet, E., Aubert, J., & Lhuillier, F., 2010. An introduction to data assimilation and predictability in geomagnetism, *Space science reviews*, **155**(1-4), 247–291.

Fournier, A., Nerger, L., & Aubert, J., 2013. An ensemble Kalman filter for the time-dependent analysis of the geomagnetic field, *Geochemistry, Geophysics, Geosystems*, **14**(10), 4035–4043.

Fournier, A., Aubert, J., & Thébault, E., 2011. Inference on core surface flow from observations and 3-D dynamo modelling, *Geophysical Journal International*, **186**(1), 118–136.

Genevey, A., Gallet, Y., Constable, C., Korte, M., & Hulot, G., 2008. Archeoint: An upgraded compilation of geomagnetic field intensity data for the past ten millennia and its application to the recovery of the past dipole moment, *Geochemistry, Geophysics, Geosystems*, **9**(4).

Glatzmaier, G. A., 2002. Geodynamo simulations-how realistic are they?, *Annual Review of Earth and Planetary Sciences*, **30**(1), 237–257.

Gubbins, D., 1975. Can the Earth's magnetic field be sustained by core oscillations?, *Geophysical Research Letters*, **2**(9), 409–412.

Gubbins, D. & Bloxham, J., 1985. Geomagnetic field analysis—III. Magnetic fields on the core-mantle boundary, *Geophysical Journal International*, **80**(3), 695–713.

Gubbins, D. & Roberts, N., 1983. Use of the frozen flux approximation in the interpretation of archaeomagnetic and palaeomagnetic data, *Geophysical Journal International*, **73**(3), 675–687.

Hartmann, G. A. & Pacca, I. G., 2009. Time evolution of the South Atlantic magnetic anomaly, *Anais da Academia Brasileira de Ciências*, **81**(2), 243–255.

Hellio, G., Gillet, N., Bouligand, C., & Jault, D., 2014. Stochastic modelling of regional archaeomagnetic series, *Geophysical Journal International*, **199**(2), 931–943.

Hjorungnes, A. & Gesbert, D., 2007. Complex-valued matrix differentiation: Techniques and key results, *IEEE Transactions on Signal Processing*, **55**(6), 2740–2746.

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

Jazwinski, A. H., 1970. *Stochastic processes and filtering theory*, Academic Press, New York.

Kalnay, E., 2003. *Atmospheric modeling, data assimilation, and predictability*, Cambridge University Press.

Korte, M. & Constable, C., 2005. The geomagnetic dipole moment over the last 7000 years–new results from a global model, *Earth and Planetary Science Letters*, **236**(1), 348–358.

Korte, M. & Constable, C., 2011. Improving geomagnetic field reconstructions for 0–3ka, *Physics of the Earth and Planetary Interiors*, **188**(3), 247–259.

Korte, M., Donadini, F., & Constable, C., 2009. Geomagnetic field for 0–3 ka: 2. A new series of time-varying global models, *Geochemistry, Geophysics, Geosystems*, **10**(6).

Korte, M., Constable, C., Donadini, F., & Holme, R., 2011. Reconstructing the Holocene geomagnetic field, *Earth and Planetary Science Letters*, **312**(3), 497–505.

Kuang, W., Tangborn, A., Wei, Z., & Sabaka, T., 2009. Constraining a numerical geodynamo model with 100 years of surface observations, *Geophysical Journal International*, **179**(3), 1458–1468.

Lanos, P., 2004. Bayesian inference of calibration curves: application to archaeomagnetism, in *Tools for Constructing Chronologies*, pp. 43–82, Springer.

Lhuillier, F., Fournier, A., Hulot, G., & Aubert, J., 2011. The geomagnetic secular-variation timescale in observations and numerical dynamo models, *Geophysical Research Letters*, **38**(9).

Li, K., Jackson, A., & Livermore, P. W., 2014. Variational data assimilation for a forced, inertia-free magnetohydrodynamic dynamo model, *Geophysical Journal International*, **199**(3), 1662–1676.

Licht, A., Hulot, G., Gallet, Y., & Thébault, E., 2013. Ensembles of low degree archeomagnetic field models for the past three millennia, *Physics of the Earth and Planetary Interiors*, **224**, 38–67.

Monnereau, M., Calvet, M., Margerin, L., & Souriau, A., 2010. Lopsided growth of Earth's inner core, *Science*, **328**(5981), 1014–1017.

Nilsson, A., Holme, R., Korte, M., Suttie, N., & Hill, M., 2014. Reconstructing Holocene geomagnetic field variation: new methods, models and implications, *Geophysical Journal International*, p. ggu120.

Olson, P. & Deguen, R., 2012. Eccentricity of the geomagnetic dipole caused by lopsided inner core growth, *Nature Geoscience*, **5**(8), 565–569.

Pavón-Carrasco, F. J., Osete, M. L., Torta, J. M., & De Santis, A., 2014. A geomagnetic field model for the holocene based on archaeomagnetic and lava flow data, *Earth and Planetary Science Letters*, **388**, 98–109.

Ramsey, C. B., 2009. Bayesian analysis of radiocarbon dates, *Radiocarbon*, **51**(1), 337–360.

Souriau, A. & Calvet, M., 2015. The Earth's cores, in *Treatise on Geophysics (Second Edition)*, pp. 725 – 757, ed. Schubert, G., Elsevier, Oxford, second edition edn.

Tangborn, A. & Kuang, W., 2015. Geodynamo model and error parameter estimation using geomagnetic data assimilation, *Geophysical Journal International*, **200**(1), 664–675.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Reviews of Geophysics*, **20**(2), 219–232.

4.7 Further considerations

The objective of the article presented in the previous section is twofold: to quantify the resolution power of the archeomagnetic data in providing information of the magnetic field at the CMB, and pave the way for the basic analysis algorithm which will enter the archeomagnetic sequential assimilation framework. Both objectives are closely connected to the exploitation of prior information from dynamo simulations entering the estimation problem.

4.7.1 Robustness of the dynamo prior

The prior information based on dynamo models has the advantage of relying on the known basic physics of the core magnetic field. The criterion of Earth-likeness is essential to the choice of an appropriate background model to provide reliable statistic information to enter the estimation problem. Given the broad range of simulations capable of displaying an Earth-like morphology (Fig. 2.2), many models can be used as reliable first guess of the state of the magnetic field at the CMB.[1] An Earth-like criterion for the secular variation, on the contrary, has not been explored so far. However, the Coupled Earth model (Aubert et al., 2013) succeeds at displaying a secular variation pattern close to that observed for the last 400 years (Jackson et al., 2000). The different secular variation pattern displayed by the CE model reflects richer dynamics, and therefore different spatial interdependencies amongst the model variables in comparison with other models. The mean and covariance matrix of a long run of CE is then assumed a robust source of prior information, mainly for the calculation of the data resolution.

Another argument for the robustness of the dynamo norm is that after proper rescaling of the non-dimensional variables, it is non-adjustable. Discussion might arise, however, on the rescaling itself, in particular for the magnetic field amplitude. As the resolution analysis is carried out as balance between the magnetic field prior and data uncertainties, the prior scaling directly affects the outcome. Since the rescaling the background model corresponds well to that of the Earth, the rescaling allows for a good comparison of the power spectra.

The question of the dependency on the background model might arise, since it is natural to think different dynamos models would give different background information. Although this argument is fair, there is increasing evidence that the dynamo models in the Earth-like parameter range show similar large-scale spatial structures (Davidson, 2013). Since the larger scales, up to degree 5, are the only ones resolved by the archeomagnetic dataset, an invariance of the spatial correlations would imply retrieving the same archeomagnetic resolution matrix. In order to test this hypothesis, the same resolution analysis performed in Section 3.3 of the article was made using the STD model (recall Chapter 2.8) instead. The comparison of the resolution matrices from CE and STD in terms of the trace of the archeomagnetic resolution matrix, tr($\mathbf{s}$), is shown in Figure 4.2, evidencing the convergence of the estimates and the robustness of the resolution analysis.

---

[1] One should take into account, however, that such Earth-like criteria is based on the morphological analysis of inverse models of the magnetic field, which are themselves submitted to regularization and limitation in resolution by the available data.

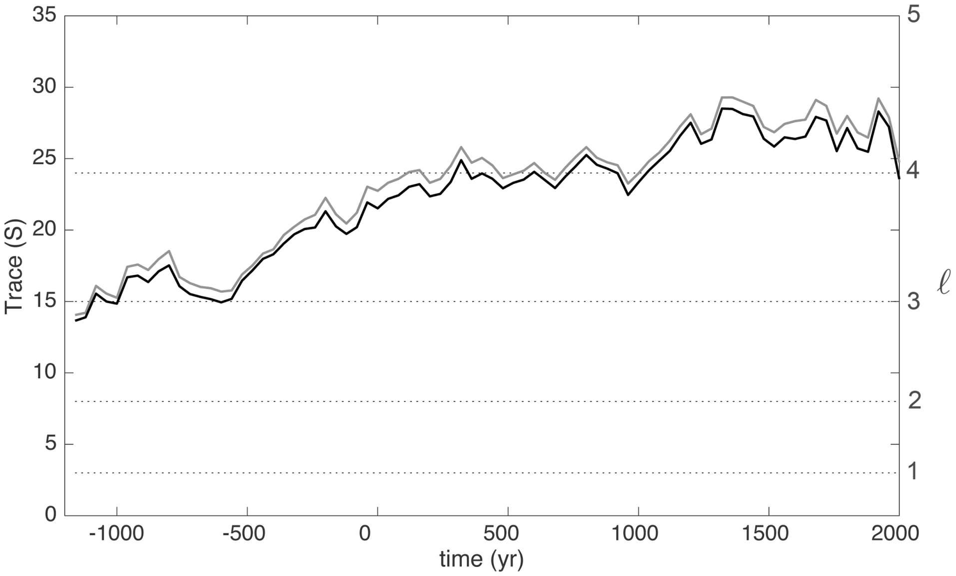


Figure 4.2: Trace of the archeomagnetic resolution matrix based in different priors. The black curve represents the CE dynamo-based resolution and the grey curve the STD dynamo-based resolution. Dashed lines represent the corresponding SH degree ℓ associated with the respective trace.

4.7.2 Archeomagnetic resolution

The simple inspection of the highly heterogeneous geographical and temporal distribution of the archeomagnetic dataset at the Earth's surface, shown in Figure 1.11, anticipates its limited influence in imaging the magnetic field at the top of the core. The objective quantification of this influence is given by the resolution matrix, connecting the information from the data distribution and uncertainty by means of a proper observation operator to that information from the prior uncertainty. The archeomagnetic resolution matrix for both the native N and redistributed R datasets derived from Geomagia50.V2 (Donadini et al., 2009), shown in Figure 6 of the article, corresponds well to the initial suspicion from Korte and Constable (2008) of an archeomagnetic resolution reduced to spherical harmonic degree 4. What is seen in more detail is the quantification of resolution as well as its distribution amongst the different model coefficients and its evolution in time. The trace of the archeomagnetic resolution matrices for the N and R datasets is shown in Figure 4.3. The figure confirms the previous conclusions from the resolution matrices, indicating mainly the increasing ladder-like character of the resolution in time. It is also worth noticing that the trace from the R dataset is higher than the one based on the N dataset. This suggests that the information of age uncertainties, which is taken into account on the R dataset, is highly valuable for the resolution. Although the jump in resolution around 0 AD is evident, the one around 1000 AD is somewhat less clear and seems to show a smoother tendency.

It is important to acknowledge the fact that lower bounds were asserted to the data uncertainties of the original Geomagia50.V2 dataset to construct the N and R datasets discussed in the article. Such uncertainty reworking was applied in order to provide a equivalent database as the ones used from the reference models ARCH3k.1 (Korte et al., 2009) and AFM (Licht et al., 2013) for the sake of model comparison under similar

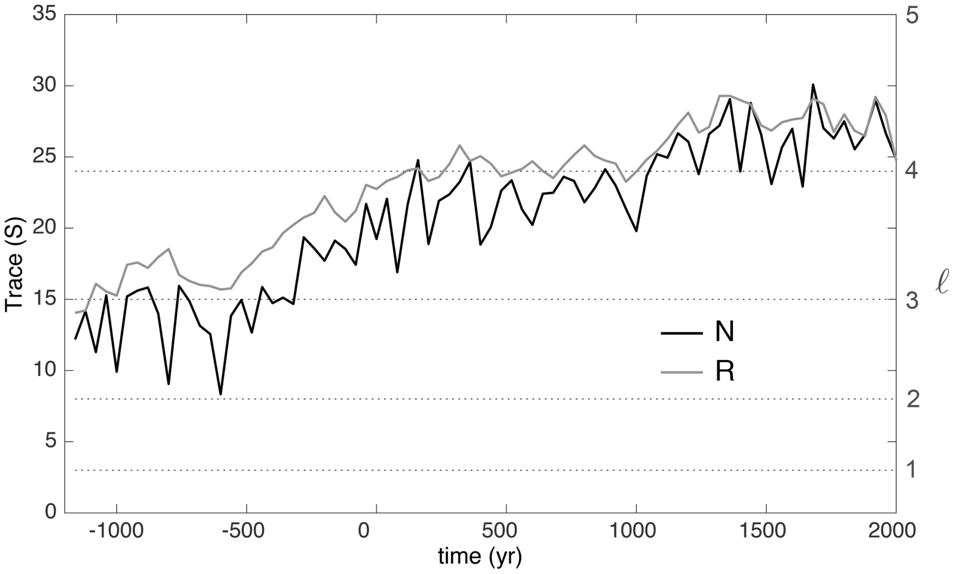


Figure 4.3: Trace of the resolution matrices based on the N (black curve) and R (grey curve) datasets with the CE prior, indicating the approximative degree up to which the field can be resolved for different times. As in Fig. 4.2, the dashed lines represent the corresponding SH degree ℓ associated with the respective trace.

data constraints. The uncertainty lower bounds are generally used due to the suspicion of overestimated confidence in very precise data. Such lower bounds in uncertainties act therefore to penalize what is thought to be high quality data. Although discussion can arise wether precision indeed indicates really high quality, this penalization strongly impacts the resolution power of the dataset. As seen in Figure 4.4, the original version of the Geomagia50.V2 dataset, non penalized by uncertainty lower bounds, resolves the model considerably better than the N dataset.[2]

Also shown in Fig. 4.4 is the native dataset without lower bounds of the newest version of the Geomagia database, Geomagia50.V3 (Brown et al., 2015). Although the additional data does improve the archeomagnetic resolution for the time window of the last three thousand years, the gain is relatively small. The small changes despite the high amount of data incorporated in the last version might reflect the fact that data is still biased to denser data regions and scarce on the Southern hemisphere (around 5% of the database).

Although the results from the archeomagnetic resolution matrices are objective with respect to the influence on observations on the model, it is important to keep in mind that they represent an upper bound for the real archeomagnetic resolution. Such optimistic case reflects the situation in which the observation errors would be coherent with their uncertainty bounds, either for the data themselves and their age. The archeomagnetic dataset is known to have an important amount of such incoherent data, characterized in inverse modeling as outliers. Licht et al. (2013) estimates around 10% of outliers in the construction of the ensemble of archeomagnetic models.

---

[2] The dataset is again restricted to data possessing age uncertainties, as in the N dataset.

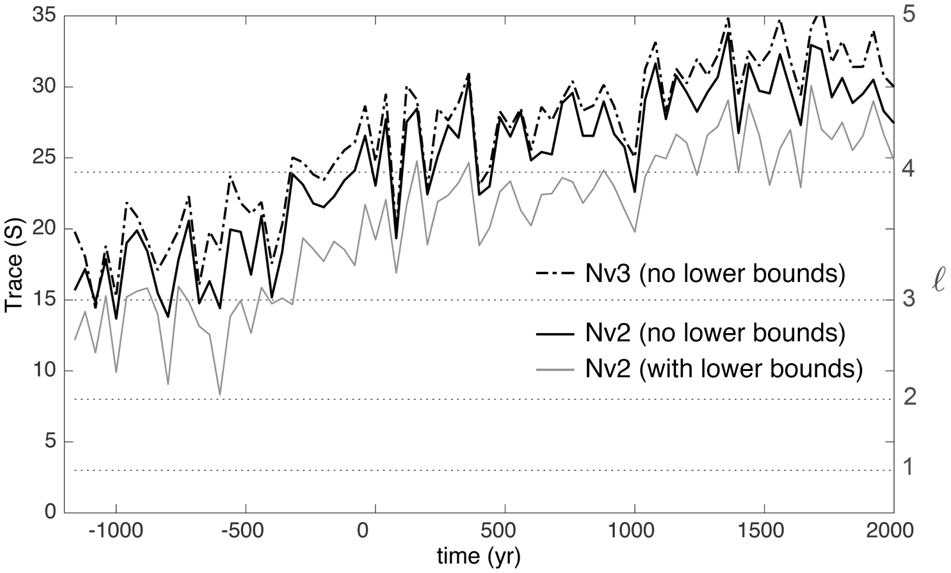


Figure 4.4: Trace of the resolution matrix for the N dataset, without uncertainty lower bounds, based on Geomagia50.V2 (Donadini et al., 2009) (black curve) and Geomagia50.V3 (Brown et al., 2015) (black dashed curve). The resolution estimation for the R-based Geomagia50.V2 datase with uncertainty lower bounds in also shown for comparison (grey curve).

4.7.3 Connection with sequential archeomagnetic assimilation

Even though the archeomagnetic resolution matrix provides an upper bound to the amount of detail that can be imaged by the archeomagnetic data using an inverse technique, it does not necessarily apply to a data assimilation estimation. In particular for the sequential assimilation, where observations are fed frequently to the dynamically evolving model, information can be propagated between spatial scales. This information cascade can possibly allow for an improvement in data resolution, and therefore an increment to our knowledge of the archeomagnetic field.

Chapter 5

Sequential geomagnetic data assimilation

As introduced in Chapter 3, data assimilation has been explored so far in the context of geomagnetism by means of the incorporation in inverse models, in the form of spectral coefficients of the magnetic field at the top of the core, as observations. Although the different assimilation schemes show promising results, little attention has been given to the possibility of assimilating actual observations of the geomagnetic field instead of parameterized models. The favoring of the assimilation of raw observations instead of inverse models is also strongly supported in other fields of research the (e.g. in meteorology, Talagrand, 1997). The main reason behind this favoring is that substitution of raw observations by observation-derived models, like inverse magnetic field maps at the CMB, could suffer from serious biasing of the initial signal as well as unreliable uncertainties, depending on the definition of the prior information (discussed in Chapter 4). One could argue that such biasing could be small in the case where observations resolve well the model, like in the modern satellite and observatory era. However, this is not the case for the archeomagnetic period.

As seen in the previous chapter, the archeomagnetic dataset has the power to resolve the magnetic field at the CMB up to spherical harmonic degree 3 or 5 depending on the time period. This lack of resolution, together with its heterogeneous character amongst the different harmonic degrees, leaves place for a biasing by the inverse method prior information. In the raw, point-wise, geomagnetic data assimilation case, the observations are connected to the underlying physical model by a proper observation operator consisting on the Green functions associated to the measurement type and position (Subsection 1.6.4 and Section 3.7). This observation operator bypasses the need for using inverse models as observations and can therefore open the way for direct assimilation of geomagnetic data.

This chapter presents a sensitivity study of the assimilation of point-wise geomagnetic observations and its comparison with the assimilation of spectral observations. Although consisting of closed-loop experiments, this study focuses in case scenarios closely representative of what would be the assimilation of archeomagnetic data through the last three millennia.

5.1 The EnKF and Parody-PDAF

As previously described, data assimilation comprises a variety of methods used to handle different characteristics of the time-dependent estimation problem. For the geomagnetic case, two of the main challenging components are the size of the problem (the dimension of the state vector) and the non-linear character of the dynamo model. Focusing on the sequential methods, the EKF or EnKF would be best suited to deal with the non-linear aspect. The EKF (Subsection 3.3.1) would imply an impractical calculation of the forecast covariance matrix, due to the linearization of the dynamo model. Moreover, the EKF is supposed to perform well in the case of weak non-linearities only. Differently, the EnKF (Subsection 3.3.2) deals with the non-linearity of the model by means of working with an ensemble of models. Although it is supposed to handle reasonably well stronger non-linear behavior, at a first glance the ensemble of models considerably enlarges the problem size. The model error propagation is, however, far less expensive than in the EKF. In this study, we choose the EnKF as the sequential data assimilation method to be applied to the point-wise geomagnetic data assimilation.

Instead of forecasting the model statistics by means of the linearized model operator, in the EnKF the background statistics are provided by the model ensemble. The analysis step (presented in Section 4.6), is performed for each ensemble member, and the analysis error given by the new ensemble covariance. An illustrative example of the EnKF algorithm is shown in Figure 5.1. In order for the EnKF to be of practical implementation for large-size assimilation problems, the algorithm must profit from parallel computing in the calculation of the ensemble integration.

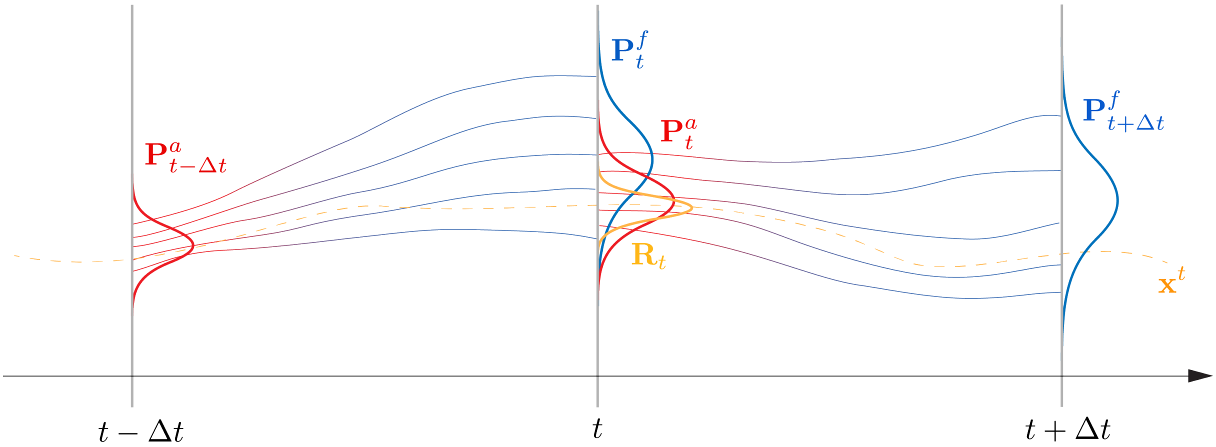


Figure 5.1: Scheme of the forecast and analysis step of the EnKF. An initial ensemble of models at time $t-\Delta t$ with covariance matrix $\mathbf{P}^a_{t-\Delta t}$ is forecasted up to the time step t, when the covariance matrix has evolved to $\mathbf{P}^f_t$. Given the set of observations at time t with a error covariance $\mathbf{R}_t$, the analysis step is performed and the ensemble state is updated, within a new covariance $\mathbf{P}^a_t$. The analysis is then forecasted up to subsequent times continuing the EnKF analysis cycle.

Given the broad applicability of the diverse data assimilation strategies, Nerger and Hiller (2013) developed the Parallel Data Assimilation Framework (PDAF). This suite of routines parallelized in MPI can be attached to the code simulating the physical system in question, usually also parallelized. Although initially put in place for oceanographic problems, it has been been further applied to the Parody code (Dormy et al., 1998; Aubert et al., 2008) by Fournier et al. (2013). The EnKF is one of the many different Kalman filter-based algorithms proposed by PDAF (Nerger and Hiller, 2013). The resulting EnKF Parody-PDAF combines the forecast of the ensemble of models based on the Parody code (the forecast phase is parallelized in a second level by PDAF), to the subsequent analysis performed on the go by PDAF. An important adaptation of PDAF to the geomagnetic context is the transformation from transpose operations to transpose conjugate, since Parody, as many spectral dynamo methods, rely on complex-valued variables. More details can be found in Fournier et al. (2013).

5.2 Model and state vector

In order to enter the assimilation problem, the numerical model must reflect the basic aspects of the known characteristics of the observed magnetic field. Those are the morphological criteria by Christensen et al. (2010) and the localization of the secular variation (Aubert et al., 2013). The obvious candidate would be the CE model (Section 2.8), but its high resolution implies a model dimension of approximately 10^7. In order to decrease the computational cost of an ensemble calculation with an Earth-like numerical model, preference is given to a lower resolution version of CE. The characteristics of the lower model version, CEL, are listed in Table 2.2.

Differently from the formalism presented in Chapter 4, where the state vector comprised the poloidal magnetic field at the CMB, the present state vector now takes into account all the variables of the dynamo model. Recalling the introduction of Section 2.6, Parody, as many spectral codes, work with the spectral decomposition of its field variables, decomposed in poloidal and toroidal components. Here, the poloidal and toroidal spectral coefficients of the magnetic field are defined by $\mathcal{P}$ and $\mathcal{T}$, the poloidal and toroidal spectral coefficients of the velocity field by $\mathcal{V}$ and $\mathcal{U}$, and the spectral coefficients of the codensity field by $\mathcal{C}$. The full state vector of the dynamo system is then written as

$$\mathbf{x}(t) = \left[..., \mathcal{P}^m_{\ell,k}(t), ..., \mathcal{T}^m_{\ell,k}(t), ..., \mathcal{V}^m_{\ell,k}(t), ..., \mathcal{U}^m_{\ell,k}(t), ..., \mathcal{C}^m_{\ell,k}(t), ...\right]^T , \qquad (5.1)$$

for all ℓ, m and k, which represent respectively the degree, order and radial level. Following the resolution of the CEL model, the state vector has a size close to 2.6×10^6.

5.3 True state and initial ensemble

As we are interested in assessing the performance of the assimilation by twin experiments, a reference trajectory of the CEL has to be defined. In order to compute this reference trajectory, a long integration corresponding to 20,000 years (using the time rescaling described in Section 2.7) of CEL has been performed. This discretized sequence of states, called the truth, or true state $\mathbf{x}^t$, is used to generate synthetic observations.

Together with the true state, a initial condition should be fed to the ensemble of models to enter the EnKF assimilation. The initial ensemble $\{..., \mathbf{x}^0_e, ...\}$, where e is the ensemble index, is chosen to consist in a broad variety of states, representing a large initial spread. In order to test the performance of the algorithm in a case other than the ideal, the initial ensemble is chosen in a biased way. The bias is chosen based on the axial dipole coefficients. From the long run, only model states which axial dipole were higher than the axial dipole of the initial true state $\mathbf{x}_0^t$ were allowed to compose a pre-selection of the ensemble of initial states. Finally, a random draw of the pre-selected states was made to create the initial ensemble of $N_e = 512$ states.

The number of ensemble states N_e was selected based on the sensitivity study of Parody-PDAF by Fournier et al. (2013). The authors show that for a simpler geodynamo model, the assimilation system reaches statistical convergence for ensemble sizes greater than 480, as can be seen in Figure 5.2. We choose therefore N_e based on the 480 lower bound and the characteristics of the computing resources available for the calculations. Also shown in the figure is the characteristic time for the convergence of the assimilation, in terms of the field errors. This time-scale is known as the spin-up time of the system. For the case studied in Fournier et al. (2013), corresponding to the assimilation of spectral coefficients up to degree $L = 13$ and analysis cycles spaced by 25 years, the spin-up time corresponds to roughly 1,000 years (Fig. 5.2). This estimate, however, should not be regarded as universal in the geomagnetic assimilation. It depends both on the underlying model and the assimilation parameters like the observation vector spatial resolution and frequency. The dynamo model used by Fournier et al. (2013) is also more laminar (less turbulent) than CEL. The transfer of information, dictated by the rate of propagation between the different spatial scales, is supposed to be greater in CEL, therefore yielding a shorter spin-up time. This subject will be further explored in Section 5.6.

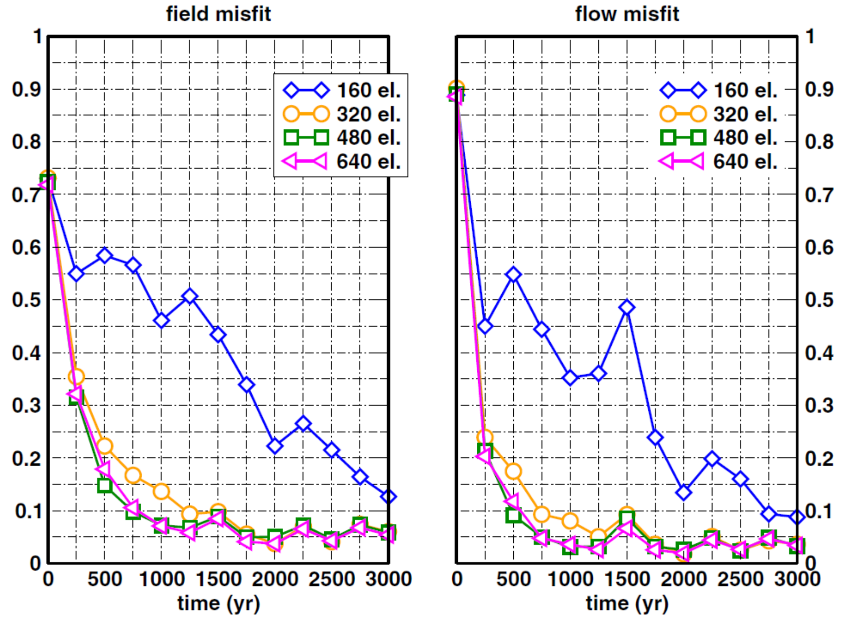


Figure 5.2: Spin-up time of a closed-loop assimilation experiment from Fournier et al. (2013). The left and right figures show the normalized error of the magnetic field and flow field, respectively, of the assimilation of spectral coefficients up to degree $L = 13$ in time windows of 25 years.

5.4 Observation operators

Although the mantle has finite conductivity by which it impacts the temporal magnetic signal of the core (Jault, 2015), this contribution is generally assumed as negligible when it comes to modeling the magnetic field of the core. Given the approximation of negligible influence of the mantle, the magnetic field signal at the Earth's surface can be directly connected to the observable part of the dynamo system, poloidal magnetic field at the top of the core. The connection between observations and the state vector must therefore depend only of $\mathcal{P}_\ell^m(r_c)$.

5.4.1 Gauss coefficients

In the assimilation studies mentioned in Section 3.6, spectral coefficients of the magnetic field at the CMB are used as observations of the magnetic field. In such a case, the observation operator is linear and diagonal, only non-zero in the part affecting the radial level $k = r_c$ of the poloidal magnetic field. Next, one must acknowledge that the resolution of the observable core magnetic field is limited either by the crustal magnetic curtain (Subection 1.5.3) in the case of recent observations, or due to data scarcity in the case of historical and paleo and archeomagnetic periods. Given such limitation, the 'spectral data' must be truncated up to a certain maximum observable degree L_o. The final spectral observation operator $\mathbf{H}_s$ to be applied to the full state vector in Eq. 5.1 will have the form

$$\mathbf{H}_s = \begin{cases} 1, & \text{for } x = \mathcal{P}_{\ell,k}^m, \text{ where } \ell \leq L_o, m \leq \ell, k = r_c \\ 0, & \text{all other variables} \end{cases} \tag{5.2}$$

Note that in this case, the assimilated spectral coefficients do not correspond to the real-valued Gauss coefficients g_ℓ^m and h_ℓ^m introduced in Subection 1.5.3. Instead, they correspond to the Parody code state vector, composed of complex variables. Therefore, in the case of assimilation of spectral observations, $\mathbf{H}_s$ performs a direct mapping $\mathbb{C} \to \mathbb{C}$.

5.4.2 Point-wise data

As presented in the article in Section 4.6, the actual observations of the dynamo system consist of point-wise data, in the form of the vector components of the magnetic field at the Earth's surface. Since such data are real-valued, the observation operators in question must connect real-valued variables with the complex-valued Parody code variables.

XYZ observations

In synthesis, the linear X, Y and Z observations consist of the real part of the convolution between the radial component of the poloidal field at the top of the core and the Green function corresponding to the data type and location. The details on the form of the observation operator for the components of the magnetic field vector, $\mathbf{H}_X, \mathbf{H}_Y$ and $\mathbf{H}_Z$ are given in the appendix of the article in Section 4.6.

Although the assimilation algorithm involves the use of well defined matrices, as described in the analysis step in Section 3.2.3, some matrix products can be written in a much more practical way. As noted in Subsection 3.3.2 In the EnKF, for example, the matrix products $\mathbf{HP}^f$ and $\mathbf{HP}^f\mathbf{H}$ can be calculated directly, which apart from avoiding the calculation and storage of the forecast covariance matrix also allows an easy implementation of the analysis process in the underlying code.

DIF observations

Since the angular and intensity data are non-linear combinations of the magnetic field vector components, attention must be given on how they enter the assimilation scheme. In the article presented in Section 4.6, two approaches were compared in the framework of the single-epoch EnKF, the iterative linearized approach and the direct application of the nonlinear observation operators. Due to the $\mathbb{C} \to \mathbb{R}$ form of the non-linear observation operators H_D, H_I and H_F, in order to linearize them one needs to calculate the Wirtinger derivatives of the operator and apply it to an expanded state vector (Subsection 2.4.2 of article in Section 4.6 for more details).

Recalling Figure 4 of Section 4.6, we see that the iterative framework based on the linearized observation operators shows a superior performance to the application of the nonlinear operators. Such improvement in accuracy is however small, suggesting the weak non-linear character of the observation operators. Combined with this is the fact that the calculations of the linearized observation operator based on the Wirtinger derivatives in an iterative framework would considerably slow the analysis step in the assimilation.

That being said, the direct application of the non-linear observation operators H_D, H_I and H_F is more suited for the assimilation of point-wise archeomagnetic-like observations through the EnKF. The analysis step is therefore also performed by the algorithm shown in Section 3.3.2. The next Sections are dedicated to the description of the synthetic observations and the characterization of their respective uncertainties entering the EnKF twin experiments.

5.5 Synthetic observations

In order to test the sensitivity of the assimilation scheme to the nature of the different geomagnetic data, three types of synthetic observations were derived from the true state (Section 5.3). The first type consists of the spectral coefficients of the radial poloidal magnetic field at the CMB truncated up to a certain degree L_o. The second type consists in point-wise data linearly related to the magnetic field of the core. They correspond to the X, Y and Z components of the magnetic field vector at a certain location beyond the Earth's surface. The third type consists in point-wise data non-linearly related to the magnetic field of the core. They are the directional and intensity information of the magnetic field vector also at a certain site above the Earth's surface.

5.5.1 Geographical distribution

The core magnetic field has been modeled up to different degrees depending on the epoch and therefore on the amount of available observations. In the recent times of satellite and observatory measurements, this resolution is thought to reach the magnetic curtain, that is, up to spherical harmonic degree $L = 13$ or 14. For the archeomagnetic period, however, this resolution is limited by the much lower amount of data, its bad spatial distribution and higher uncertainties, reaching at maximum resolution of degree $L = 5$ (recall Chapter 4). Since the objective of this chapter is to test the performance of the geomagnetic EnKF to different types of data, it is interesting to define a simple link between the point-wise data resolution and the spherical harmonic one.

In a well-posed inverse problem, the number of variables of the model must correspond to the number of data. In terms of Gauss coefficients up to a certain degree L_{max}, this equivalence corresponds to

$$N_d = (L_{max} + 1)^2 - 1, \tag{5.3}$$

where N_d is the number of data available. Although objective, the relation in Equation 5.3 does not take into account important information such as the spatial distribution of observations and their corresponding weight, or uncertainties. However, it is evident that data distribution over the Earth's surface close to a uniform one will favor a better resolution of the magnetic field. A close-to-uniform distribution algorithm (as the tessellation used in Constable et al., 1993) is used to provide a number of locations at the Earth's surface corresponding to the data needed to model the magnetic field up to degree L_{max}. Figure 5.3 represents a close-to-uniform data distribution for three situations, one with 35 data, in which the magnetic field is resolved up to degree $L = 5$, another with 195 data, in which the magnetic field resolution reaches up to degree $L = 13$, and finally for 22 data, corresponding roughly to $L = 4$ but with a biased distribution towards one hemisphere.

It is important to remark that each data point corresponds to full vector information on the magnetic field. That is, in the case of a model resolution of $L = 5$, for example, the 35 locations displayed in Fig. 5.3 correspond each to X, Y and Z or D, I and F measurements. The grids are then used to construct a dataset of synthetic observations at different times spaced by a given time window Δt_a.

5.5.2 Uncertainties

The uncertainties affecting the data are, together with the spatial distribution, a crucial information for the assimilation into the physical model. Since in this study the interest is in assessing the feasibility of archeomagnetic data assimilation, it is important to know beforehand the information from the typical uncertainties of the archeomagnetic dataset. Figures 5.4a, 5.4b and 5.4c show the histogram of the uncertainties affecting the declination, inclination and uncertainty data from the last three millennia from the Geomagia50.V2 (Donadini et al., 2009) database. The corresponding modes for the declination, inclination and intensity data are respectively 1°, 2.5° and 1μ T.

A first approximation of synthetic data uncertainties can be done by means of the standard deviation σ^o of the time variability of a point-wise observation at a given grid

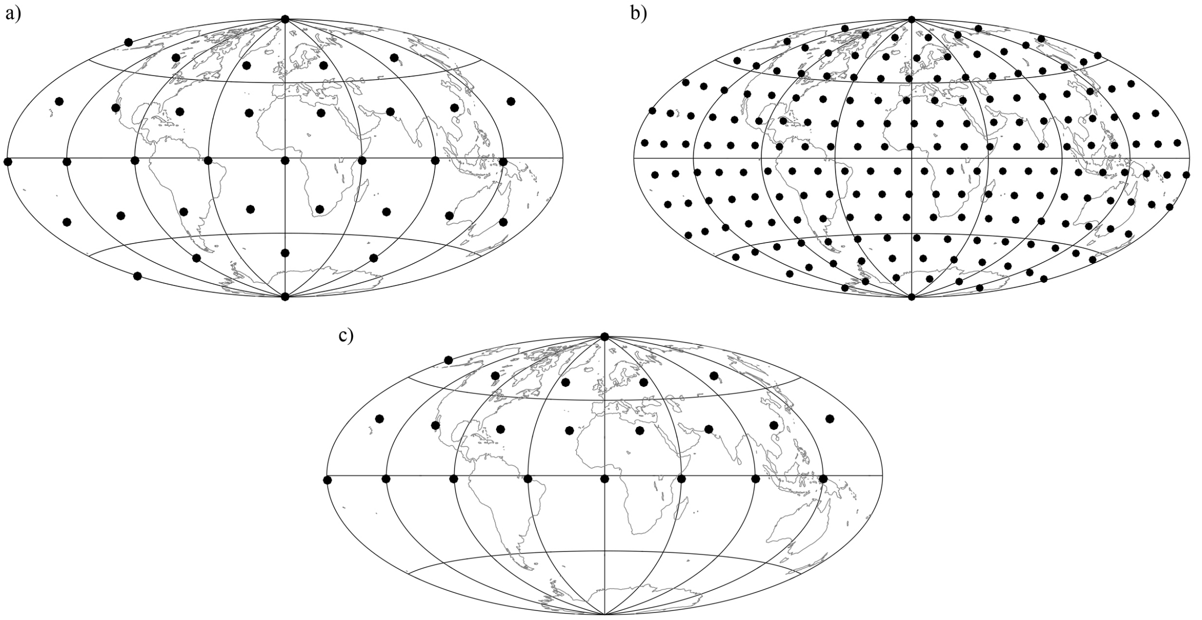


Figure 5.3: Close-to-uniform global coverage of the synthetic point-wise observations for different resolutions at the Earth's surface. In a) the 35 locations are chosen to be representative in resolution to the spherical harmonic degree $L = 5$. In b) the same, but for 195 locations associated to degree $L = 13$. In c) 22 observations are distributed only over the North Hemisphere, corresponding to $L = 4$ according to Eq. 5.3.

point. For long intervals, like 10,000 years, the distributions are convergent, Gaussian and present high standard deviation. These values represent the characteristic variability of the observations in the context of the dynamo model. They are, however, high compared to the archeomagnetic typical uncertainties. Choosing shorter time windows for the uncertainty calculation yields generally smaller σ^o, for the variability of the larger scales of the magnetic field will be reduced (Fournier et al., 2013). The time window of 500 years of the CEL simulation was chosen amongst others to derive the σ^o in view of the resemblance with that of the archeomagnetic case. The histograms of the σ^o from the synthetic observations for the grid of Fig. 5.3a are shown in Figure 5.4a, 5.4b and 5.4c. They show a mode of 1°, 0.7° and 1.3 μT for the declination, inclination and intensity, respectively. For simplicity, the uncertainties of the synthetic observations are considered to be constant for each grid location.

5.5.3 Datasets

Having prepared the synthetic point-wise observations and their respective uncertainties, it is important to recall that real data are not perfect. Every observation of a physical system is supposed to bear errors, either from the action of measuring or the representativeness of the connection to its source (recall the ideas introduced in Section 3.1). In order to add observation errors to the synthetic dataset, the observations are contaminated with an error associated to its uncertainty. Specifically, the noise comes in the form of a random draw from a normal distribution whose standard deviation corresponds to the datum uncertainty estimate σ^o.

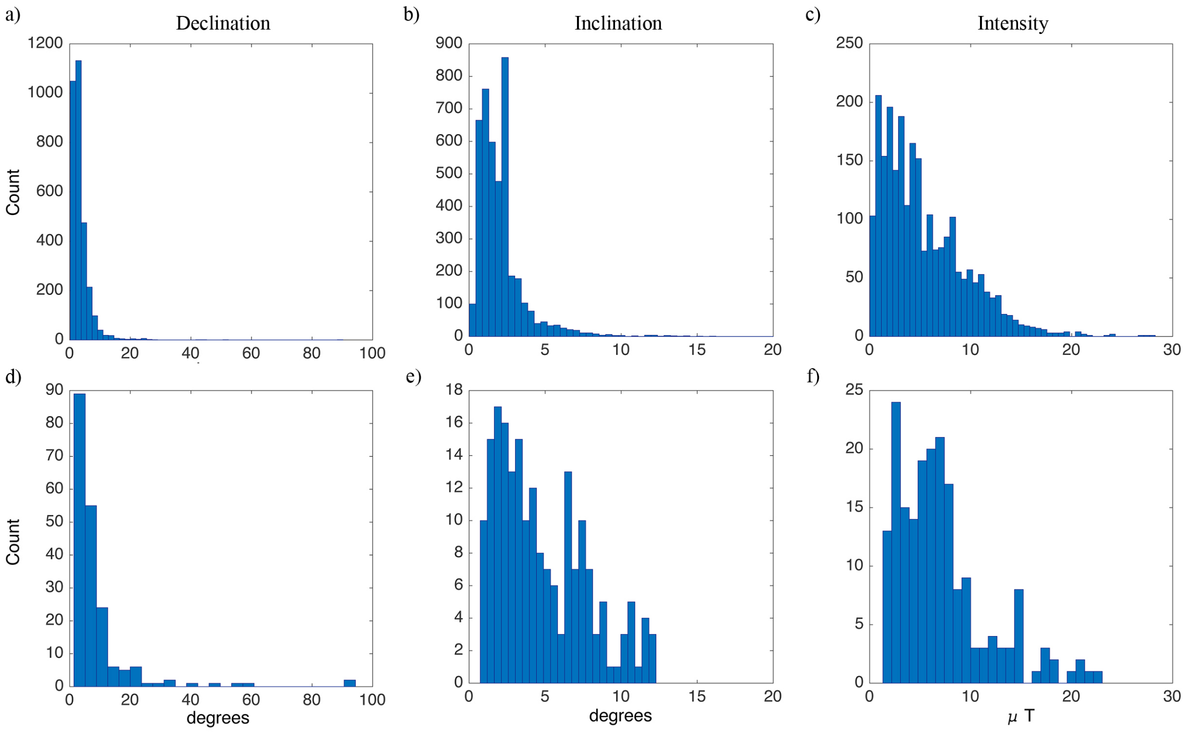


Figure 5.4: Archeomagnetic and synthetic data uncertainties σ^o histograms for declination, inclination and intensity measurements. The top row shows the distributions corresponding to the GeomagiaV2 database, while the bottom row corresponds to the synthetic data discussed in the text.

Although explained only for the synthetic point-wise DIF dataset, the same grid specification, uncertainty definition and noising were also applied to prepare the synthetic point-wise XYZ dataset. In the case of spectral observations, the synthetic dataset, here coined SPC, is associated with uncertainties corresponding to the standard deviation of each spectral coefficient of the poloidal magnetic field at the top of the core over a 500 year CEL simulation interval and noised. These three types of data, XYZ, DIF and SPC are further fed into the EnKF assimilation algorithm supplied by the Parody-PDAF in order to test the sensitivity of the framework.

5.6 Assimilation results

In order to test the performance of different assimilation configurations, in terms of data type, assimilation window, resolution and observation bias, a reference case is chosen. The monitoring of the assimilation scheme can be done by different diagnostics. As an ensemble framework, the EnKF formalism describes the precision in the model by means of the standard deviation of the ensemble, assuming the corresponding pdf is Gaussian. Therefore, as in Chapter 4, the spread of the ensemble plays a key role on the description of the model's uncertainty. However, since here the underlying background model varies in time, the spread will not be normalized by the background uncertainties as in Eq. 39 of the article in Section 4.6. Instead, the non-normalized standard deviation S of the state vector is used, as in

$$S = \sqrt{\frac{1}{n_{\mathbf{x}}}\frac{1}{N_e - 1}\sum_{e=1}^{N_e}(\mathbf{x}_e - \langle\mathbf{x}\rangle)^{\dagger}\mathbf{I}(\mathbf{x}_e - \langle\mathbf{x}\rangle)}. \tag{5.4}$$

Note here that by including the identity matrix **I**, only the diagonal terms, that is, the model variables variances, are used to calculate the spread. The spread S can be computed at any time based on the EnKF estimate, the ensemble mean $\langle\mathbf{x}\rangle$, either in its forecast or analysis stages. However, attention is generally given to the spread at the analysis steps $\langle\mathbf{x}^a\rangle$.

Further on, since the following tests are closed-loop experiments, it is possible to calculate the total error of the scheme, meaning the difference between the assimilation estimate and the true state. The normalized error E is written as

$$E = \sqrt{\frac{(\langle\mathbf{x}\rangle - \mathbf{x}^t)^{\dagger}(\langle\mathbf{x}\rangle - \mathbf{x}^t)}{\mathbf{x}^{t\dagger}\mathbf{x}^t}}. \tag{5.5}$$

As for the spread, if otherwise not mentioned, the error is always computed for the analysis estimate. The error E and spread S are therefore diagnostics of the accuracy and precision of the assimilation scheme, respectively.

It is important to keep in mind that since the true state is actually what one wants to retrieve in the estimation problem, the error E is only accessible in closed-loop experiments. It is therefore essential to use as well assimilation diagnostics based on the errors connected to the observable part of the system. The observable error can be given by the difference between the system's observation $\mathbf{y}^o$ and the projected observation by the forecast estimate. This forecast difference, called the innovation vector **d**, is written as

$$\mathbf{d} = \mathbf{y}^o - \mathcal{H}(\langle\mathbf{x}^f\rangle), \tag{5.6}$$

and its norm given by $d = \|\mathbf{d}\|$. In case the assimilation scheme is not biased to a certain state regardless of the observation input, the innovation vector should converge to zero after many analysis cycles (Talagrand, 2003). One way to monitor this convergence is by means of the cumulative innovation d_c, given by

$$d_c = \left\|\frac{1}{N_c}\sum_{i=1}^{N_c}\mathbf{d}_i\right\|, \tag{5.7}$$

where N_c is the number of analysis cycles up to the present analysis.

The quality of the analysis can be measured similarly to the innovation, by the observable analysis error. This quantity, called the residual **r**, is written as

$$\mathbf{r} = \mathbf{y}^o - \mathcal{H}(\langle\mathbf{x}^a\rangle), \tag{5.8}$$

and the residual norm as $r = \|\mathbf{r}\|$. In a well-behaved assimilation scheme, one expects $r < d$.

In the following, the aforementioned assimilation diagnostics are evaluated for different closed-loop EnKF experiments, summarized in Table 5.1. The different experiments

Table 5.1: List of closed-loop EnKF assimilation experiments. The experiments vary in terms of data type, interval between observation availability and resolution.

Code	Data	Δt_a	Resolution
dif10yl5	DIF	10 yr	35 sites
dif10yl5b	DIF	10 yr	22 sites
dif10yl13	DIF	10 yr	195 sites
dif40yl5	DIF	40 yr	35 sites
xyz10yl5	XYZ	10 yr	35 sites
spc10yl5	SPC	10 yr	$L_o = 5$

allow for an introductory sensitivity analysis of the assimilation scheme relevant to the geomagnetic case. As a side note, given the synthetic character of the experiments there is no specific need for rescaling the variables strength, like the magnetic field, velocity, codensity, etc. The time, however, is rescaled in order for the reader to appreciate the time-scales involved and its consequences for the real assimilation of geomagnetic data. The typical assimilation length used here is the same as in the previous chapter, that is 3,000 years.

5.6.1 Reference assimilation case

Since the intention of this chapter is to assess the feasibility of the assimilation in the context of archeomagnetism, a synthetic case of a DIF assimilation scheme is taken as a reference case and closely inspected. The reference assimilation experiment is labeled here as dif10yl5 (Table 5.1), which corresponds to the assimilation of DIF data over an analysis cycle of $\Delta t_a = 10$ yr and 35 observation sites corresponding to a $L = 5$ sampling (see Fig. 5.3a).

For the reference case, therefore, the initial ensemble described in Section 5.3 is confronted with the first set of surface DIF noisy observations after a forecast phase of 10 yr. The observations being connected with the poloidal magnetic field at the CMB are expected to directly influence the part of the state vector corresponding to the radial magnetic field at the top of the core. This is verified in Figure 5.5a–d by means of the comparison between the mean forecast, the mean analysis, the difference between both and the true state. The mean forecast is seen to be close to an axial dipole configuration, contrasting with the true state where many non-dipolar features of the magnetic field can be found over different spatial scales. The retrieval of such features is not perfect, since the data are finite and bear considerable uncertainties, as seen by the comparison between the analysis and the true state. However, the large scale features of the magnetic field like the intense flux patch beneath southern South America are well captured.

Still, since the assimilation scheme connects the different variables of the state vector by means of the covariance matrix $\mathbf{P}^f$, the analysis can propagate information into the hidden variables of the system. This is seen, for example, for the azimuthal velocity field at the top of the core $\mathbf{u}_\phi(r_c)$ in Fig. 5.5e–h. While the mean estimate of the system consists of a closely axisymmetric flow composed by a westward flow spread along a latitudinal band between 45°N and 45°S, the true estimate presents a rich signature of the different columnar flows underneath the core surface. The true state, as well as the forecast state,

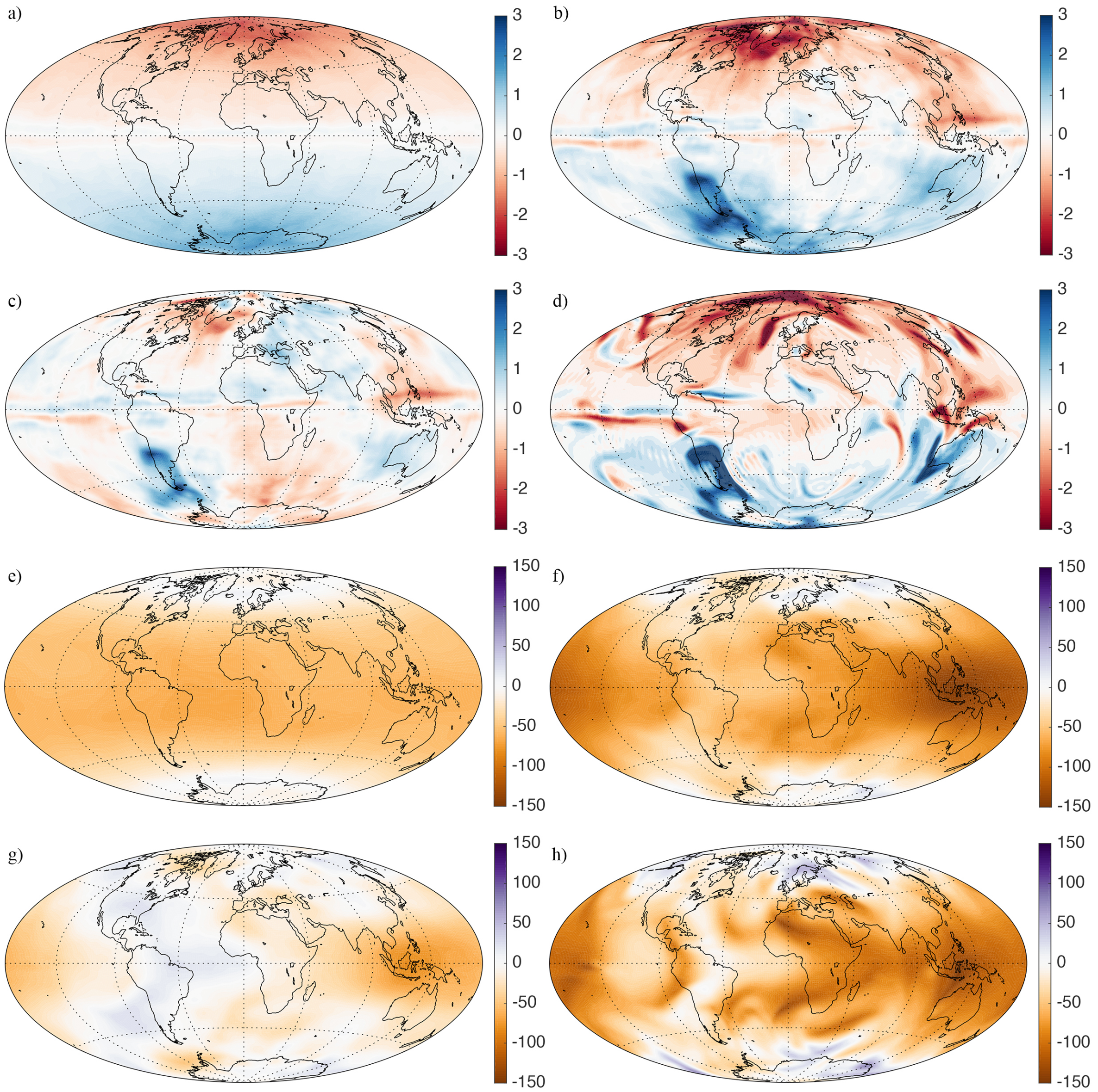


Figure 5.5: Comparison between mean forecast, mean analysis and true state for the first assimilation step of the dif10yl5 assimilation scheme at the top of the core. The a) mean forecast, b) mean analysis c) difference between forecast and analysis and d) true state is shown for the radial magnetic field at the top of the core. The same is shown in e), f), g) and h), but for the azimuthal flow at the top of the core. Since this is a synthetic experiment, continent contours are shown only in order to compare spatial scales and refer to features of the fields in the text.

shows high equatorial symmetry, that is, an apparent high degree of equatorial symmetry. Interestingly, the main features of the true azimuthal flow at the CMB are also well captured in the analysis, like the stronger westward drift of the flow underneath the western Pacific.

In fact, the analysis update not only consists of a propagation of information through the hidden system's variables, but it also extends to the different depths. This is observed in Figure 5.6, where the mean forecast, analysis and true state are again shown, but for the radial magnetic field at mid-depth in the core. Departing from weak dipolar magnetic field in the mean forecast, the analysis consists of miscellaneous flux patches. The amplitudes of the mean analysis are small, however, compared to those of the true state. This indicates that the information injected into the system is propagated in a limited way in depth.

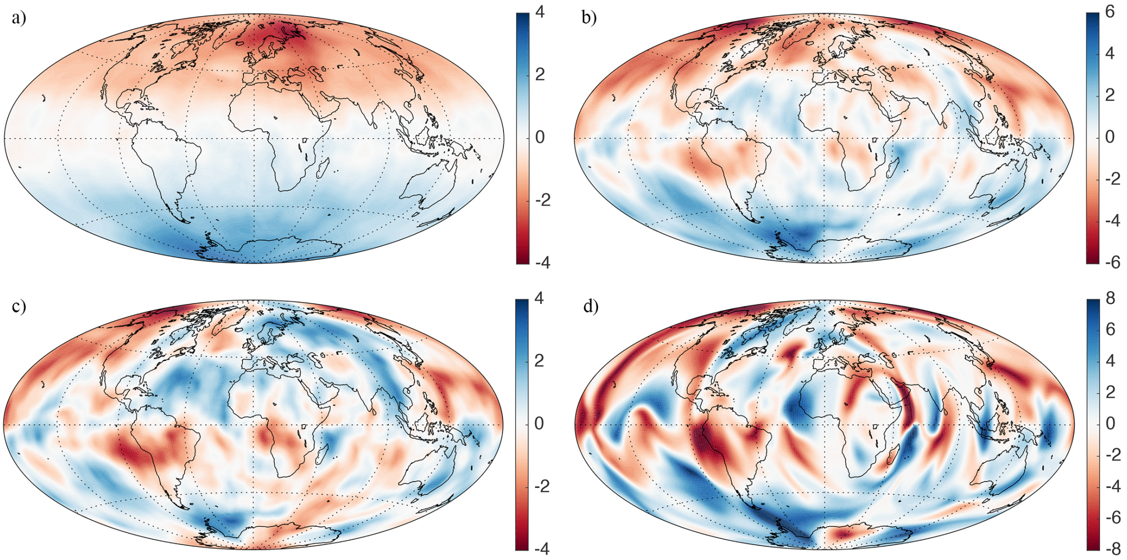


Figure 5.6: Comparison between mean forecast, mean analysis and true state for the first assimilation step of the dif10yl5 assimilation scheme under the core surface. The radial magnetic field at mid-depth in the core is shown in terms of the a) mean forecast, b) mean analysis, c) difference between forecast and analysis and d) the true state.

The effects of the analysis throughout the core are better seen in an equatorial view of the simulation. Figure 5.7 shows the azimuthal flow, radial flow and density anomaly in the equatorial plane. The azimuthal velocity in Fig. 5.7a–b is in line with the CMB estimates presented in Fig. 5.5. In particular, one can see in Fig. 5.7a the extent of the smooth and nearly axisymmetric westward drift of the flow. The analysis update in Fig. 5.7b, shows a considerably larger amount of corrections in the flow with depth, including a heterogeneous eastward drift close to the ICB, similar to the true state one.

The radial velocity flow $\mathbf{u}_r$, shown in Fig. 5.7d–f, is very weak in the mean forecast, indicating the randomness in the flow downwellings and upwellings. The mean analysis, although also weak, is generally coherent with the structure found in the true core flow. The codensity C, shown in Fig. 5.7d–f, corresponds to the density anomalies in the core. The mean forecast, displaying a simple structure concentrated along one of the hemispheres of the inner core, contrasting with the existence of structures coherent with the downwellings and upwellings seen in the radial flow. The mean analysis manages to capture only a reduced amount of such features on the codensity. It is important to recall that the comparisons showed in Fig. 5.5, Fig. 5.6 and Fig. 5.7 correspond to the first analysis of the assimilating scheme, when the model trajectory is supposedly far from the true state. The consecutive analysis cycles act to constantly update the ensemble of models, leading to an improving estimation of the model subsequent states.

An example of the temporal evolution of the assimilating closed-loop system is shown in Figure 5.8, by means of the monitoring of the total magnetic and kinetic energy of the

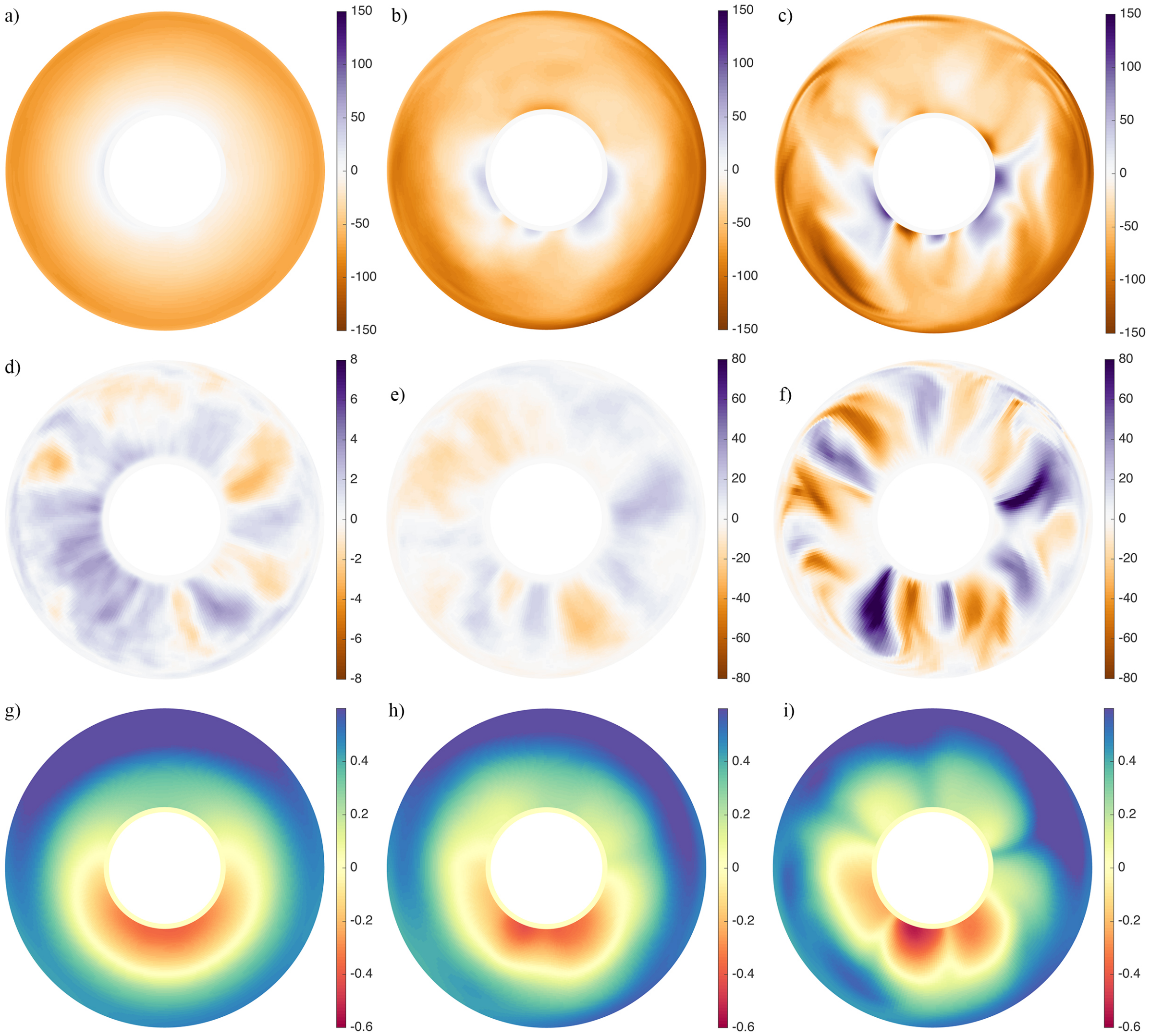


Figure 5.7: Comparison of the mean forecast, mean analysis and true state for different fields on the equatorial plane for the dif10yl5 assimilation. a), b) and c) show the mean forecast, mean analysis and true state of the azimuthal flow. d), e) and f) show the same, but for the radial velocity field. g), h), i) show the same, but for the density anomalies.

models. The start of the assimilation has the interesting effect of decreasing the total magnetic energy and increasing the kinetic energy of the ensemble mean. After some assimilation cycles, however, the ensemble begins to converge into the reference state from which the synthetic observations were created. Although the mean is not exactly coincident with the model trajectory, the latter is generally within the bounds of the ensemble. The energy profiles shown in Fig. 5.8, includes all the spatial scales composing the dynamo model, and it is natural that the incomplete and noisy observations do not affect the small scales of the system. If the monitoring is done in terms of the large scales of the fields, the estimation is found to be much more precise and accurate. Figure 5.9 shows the axial and equatorial dipole coefficients of the poloidal magnetic field at the CMB, evidencing that the assimilation succeeds in estimating the main magnetic field characteristics.

Having shown different aspects of the trajectory of the assimilating system, the qual-

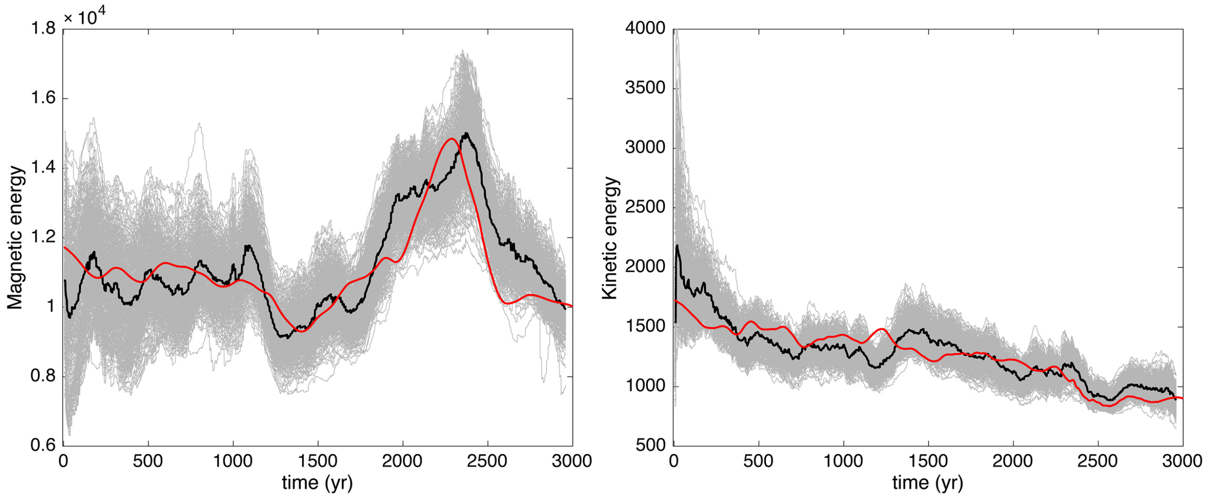


Figure 5.8: Total a) magnetic and b) kinetic energy of the dif10yl5 assimilation experiment. The ensemble of models of the EnKF is shown by the grey curves, the mean of the ensemble by the black curve, and the true trajectory, from which the synthetic data are generated, by the red curve.

ity of the scheme is quantified by means of the different diagnostics mentioned at the beginning of the present section. Figure 5.10 presents the total normalized error (Eq. 5.5) and spread (Eq. 5.4) of the previously discussed dif10yl5 assimilation test. As seen in the figure, the initial error of the ensemble starts at relatively small normalized error of 0.27. During the assimilation, this error is decreased and generally maintained under the 0.2 accuracy level, with small sharp variations around this mean. These sharp variations derive from the relatively high errors embedded in the synthetic observations, compromising the coherence between the observation set at each assimilation step. In order to assess the performance of the assimilation, the error from a free-run (that is, an unconstrained ensemble propagation) is also shown in Fig. 5.10.

In contrast to the error, the spread of the ensemble is a smoothly varying quantity, for it is not directly affected by the errors in the dataset. It depends only on the forecast uncertainties and the observation set uncertainties and spatial distribution. Fig. 5.10b shows that after a time span of approximately 500 or 700 years the assimilation system begins to reach statistical convergence in terms of precision, characterizing the assimilation spin-up time described in Section 5.3. After this time-window, the ensemble spread stays at a level defined by the initial spread divided by a factor of two.

Turning now to the diagnostics related to the observable part of the system, the innovation and residuals are presented in Figure 5.11. It is easy to see in Fig. 5.11a that the condition $r < d$ holds for the dif10yl5 experiment, confirming the well-behaved character of the assimilation. The cumulative innovation d_c shown in Fig. 5.11b decreases considerably, by two orders of magnitude, reinforcing the good performance of the assimilation. The behavior of d_c shows an apparent stagnation of towards the end of the assimilation cycle, implying a possible small bias on the model forecasts. However, a longer assimilation run is needed to confirm this tendency.

Given the improvement in estimation towards the end of the assimilation experiment shown in Figs. 5.10 and 5.11, it would be interesting to have a glance on the different field

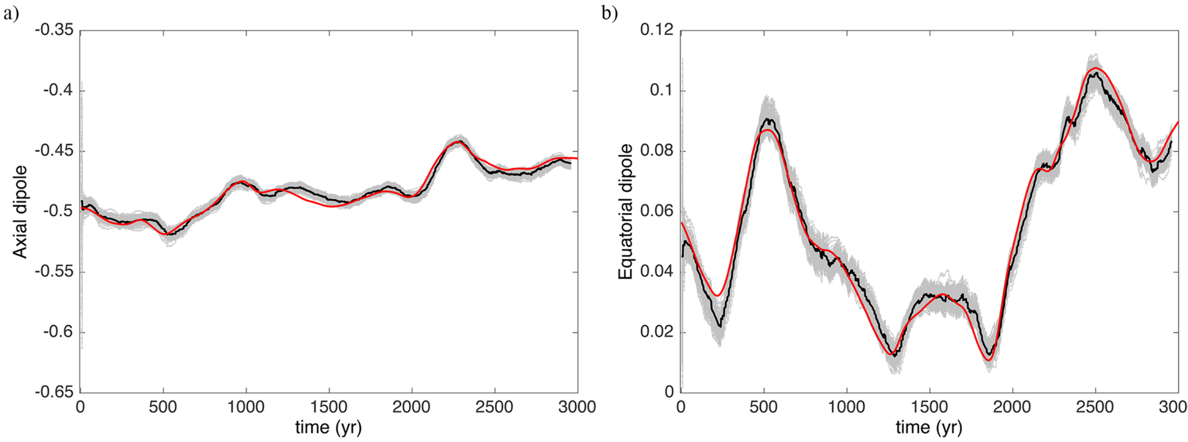


Figure 5.9: a) Axial and b) equatorial dipole of the poloidal magnetic field at the top of the core for the dif10yl5 assimilation experiment. The red curve shows the true state, the grey curves show the ensemble of models and the black curve, the ensemble mean.

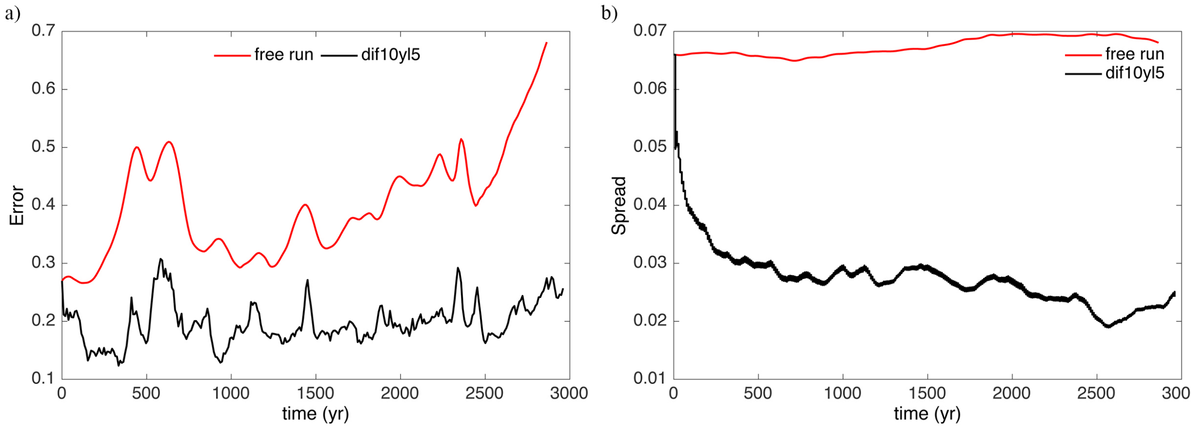


Figure 5.10: Total a) normalized error E and b) spread S of the dif10yl5 assimilation experiment, representing the accuracy and the precision of the estimates. The red curve represents a free-run of the ensemble, that is, the evolution of the ensemble members in case they are not constrained by observations.

variables at the last analysis cycle. Figures 5.12 and 5.13 show a comparison between the mean analysis and the true state at the end of the assimilation experiment (after 300 assimilation cycles, corresponding to approximately 3,000 years). The comparison shows the fields are much closer to the true state than in the first analysis, an improvement originating from the constant feeding of observations into the system and propagation by the dynamical model. All fields show a considerable degree of detail and, in particular for the radial field shown in Fig. 5.13, comparable magnitudes. Given the good results displayed by the assimilation run dif10yl5, other assimilations with different characteristic are compared to it in order to assess the sensitivity of the method. It is important to note that this is a work in progress. Some of the assimilation experiments shown in the following are still under calculation.

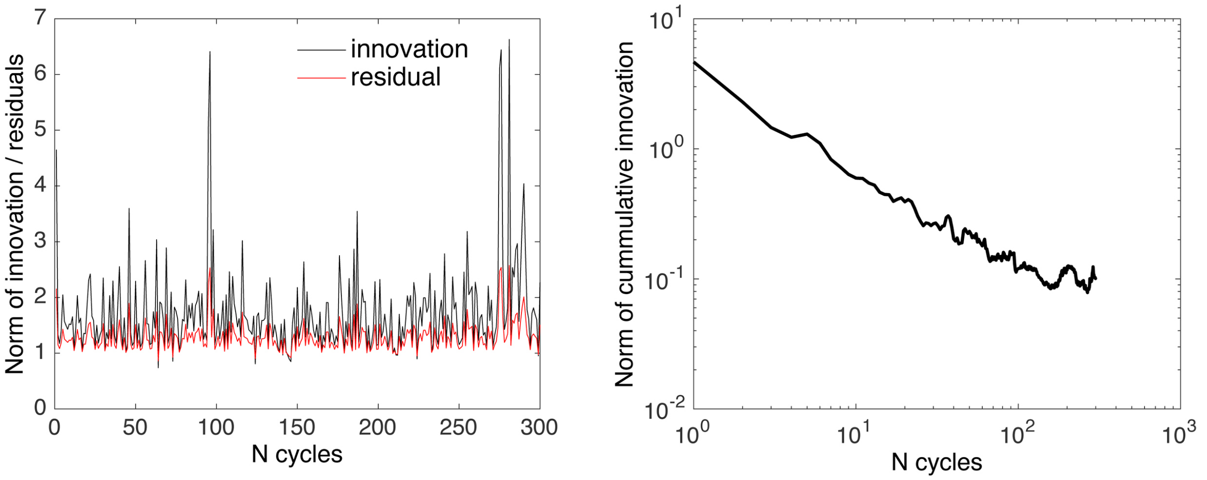


Figure 5.11: Comparison of the a) innovation and residuals and b) cumulative innovation of the dif10yl5 assimilation.

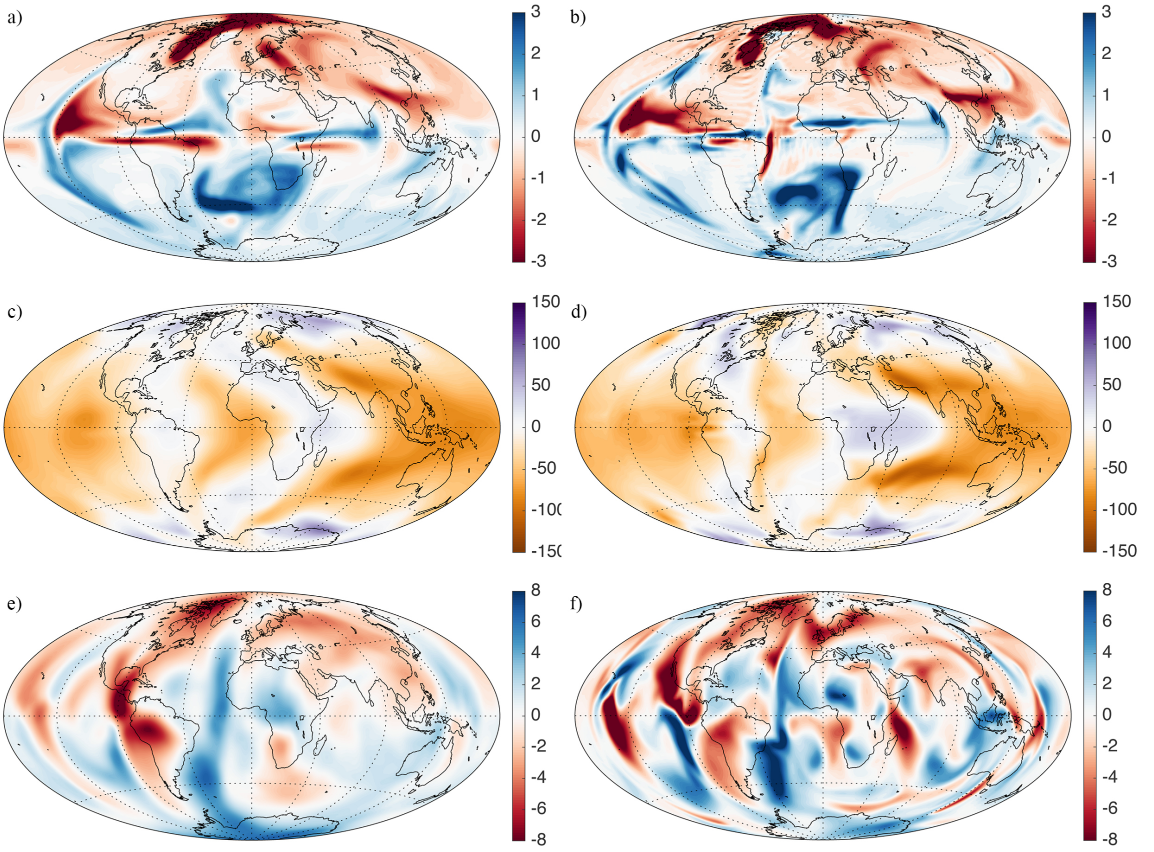


Figure 5.12: Final analysis of the dif10yl5 assimilation, showing the comparison between the the analysis (left column) and true state (right column) of different fields. a) and b) show the radial magnetic field at the top of the core, c) and d) the azimuthal flow also at the top of the core, and e) and f) the radial magnetic field at mid-depth in the core.

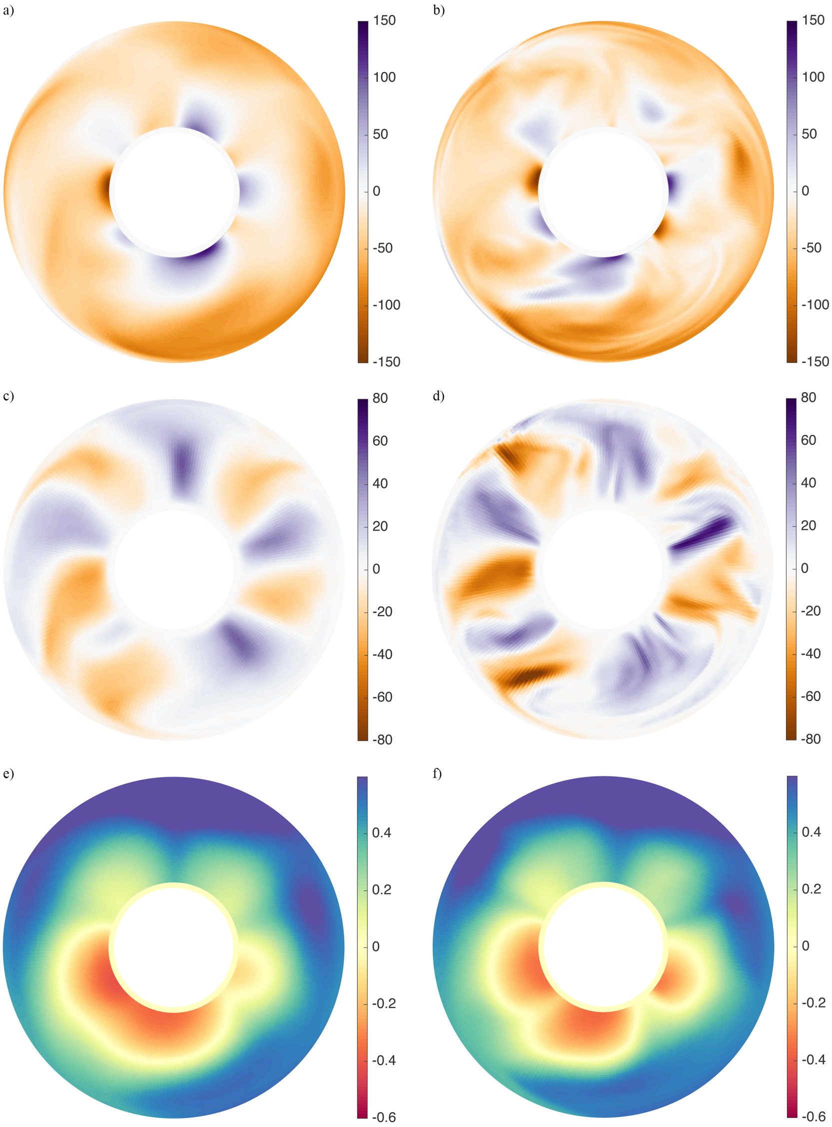


Figure 5.13: Equatorial views of the final analysis estimate of dif10yl5 in terms of the a) and b) azimuthal flow, c) and d) radial flow, e) and f) the density anomaly. The analysis is shown on the left column, while the true state is shown on the right column.

5.6.2 Sensitivity to the data type

Figure 5.14 shows the sensitivity of the assimilation to the data type, XYZ, DIF and SPC. Although for the XYZ-based assimilation the error starts quite smaller than the DIF-based one, they seem to be equivalent in the long term. The similarity between both solutions is even higher in terms of spread, pointing to the equivalence in the assimilation of the different geomagnetic data. Indeed, the invariance of the assimilation in terms of data type shows in particular the weak non-linear character of the indirect observation operators H_D, H_I and H_F, already discussed in the article in Section 4.6. The differences might be due to the fact that the observation errors are not originated by error propagation following the spectral, XYZ and DIF sequence. Instead, they are computed for each dataset independently considering the same method discussed in Section 5.5.3.

The assimilation of spectral coefficients, however, does not corresponds as well as the

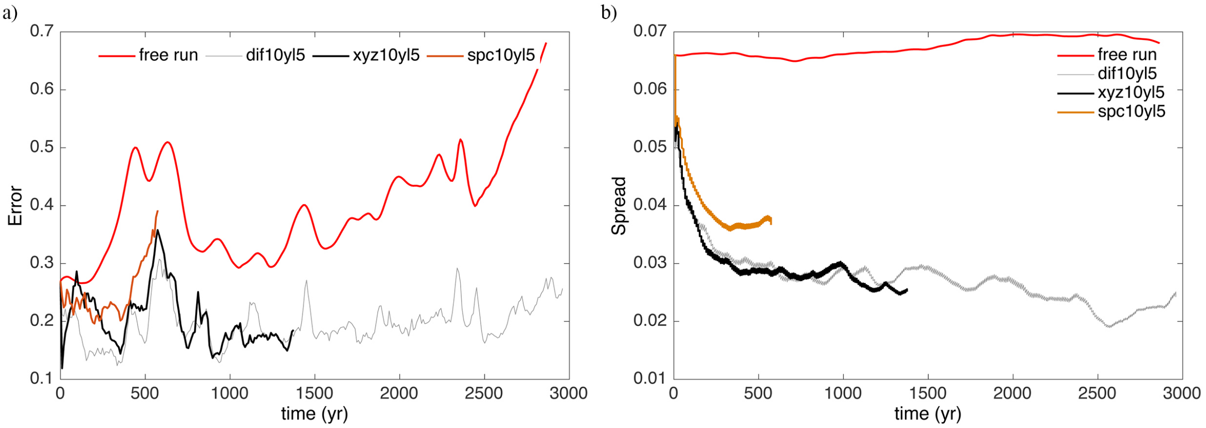


Figure 5.14: Comparison of the total a) normalized error E and b) spread S of the assimilation experiments differing on the data type.

point-wise assimilation experiments. Although the spread reveals a less precise estimation than that of DIF and XYZ, the accuracy stays around the same level. The length of the assimilation is, however, too short to allow for any conclusion.

5.6.3 Dependency on the assimilation window

The sensitivity of the assimilation to different time windows is shown in Figure 5.15 for the DIF dataset. In the figure, the DIF reference assimilation with an analysis window of $\Delta t_a = 10$ yr is compared with the DIF assimilation with $\Delta t_a = 40$ yr. Although the spread of the dif40yl5 assimilation is similar to the dif10yl5 one, the error is considerably higher. In fact, at some intervals the assimilation errors of dif40yl5 reach close to those of the free-run simulation, despite the good departure of the first analysis. The decrease in performance of the $\Delta t_a = 40$ yr assimilation points to the impact of the predictability limit of the dynamo model. The impact of analysis time intervals shorter than 40 years should therefore be carried out in the near future in order to assess the practical predictability limit of the archeomagnetic-like assimilations studied here.

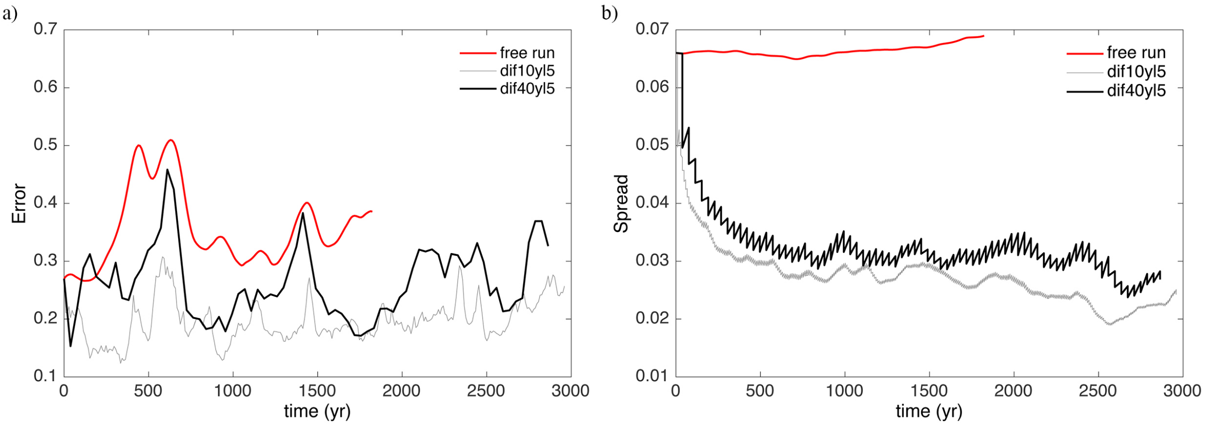


Figure 5.15: The same ad Fig. 5.14, but comparing the influence of the analysis cycle interval.

5.6.4 Impact of data distribution

Intuitively, one might think that an increase in the data quantity would increase the quality of the model. Figure 5.16 shows that this is not exactly the case. The assimilation of 195 × 3 observations corresponding to a field coverage of SH degree $L = 13$, dif10yl13, displays the same accuracy level of the assimilation of data from 35 × 3 observations corresponding to the SH degree $L = 5$, dif10yl5 (Fig. 5.3). The precision, however, is slightly better in the former case. This result suggests that for the error uncertainty level used in the present closed-loop assimilation (see Fig. 5.4), there is a substantial limit in resolution. Improving the data availability does seem to improve the precision, but does not improve the accuracy level.

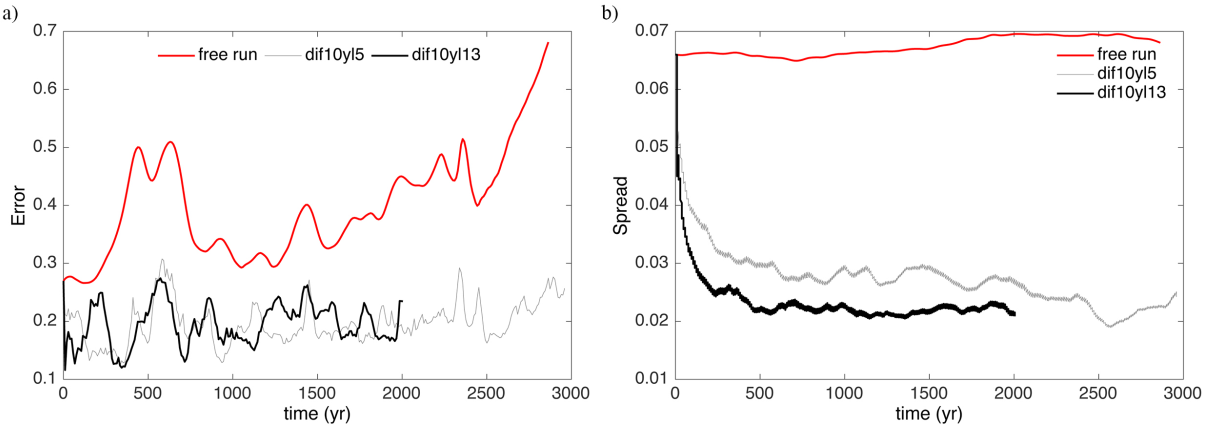


Figure 5.16: The same ad Fig. 5.14, but comparing the influence on the data quantity.

Finally, the assimilation strategy is tested for the case of a biased distribution of the surface data. The biased distribution is the one shown in Fig 5.3, depleted of data over the South Hemisphere. Such a biased sampling represents a first approximation of the archeomagnetic dataset configuration, strongly biased towards the North Hemisphere (Fig. 1.11). It can be seen in Figure 5.17 that although dramatic, the hemispherical bias does not compromise the general behavior of the assimilation. The precision of the assimilation is slightly decreased in dif10yl5b due to the lower amount of data compared to dif10yl5 (see Table 5.1), but there is no great change in the spin-up time of the assimilation.

5.7 Summary and discussion

The state of the art of geomagnetic assimilation, presented in Section 3.6, is restricted to the assimilation of spectral coefficients of the magnetic field. The present study aims at introducing the idea of the feasibility of assimilating direct and indirect magnetic observations in a context close to the archeomagnetic one. The algorithm used here is an adaptation of the Parody-PDAF code (Fournier et al., 2013). This newer version consists in the development of observation operators corresponding to the point-wise direct and indirect observations of the magnetic field at the surface. The EnKF framework is chosen in order to deal with the nonlinearities of both model and observation operator (in the case of indirect magnetic data). Closed-loop, or twin experiments, were carried out in order to assess the performance of the assimilation experiments. For that, synthetic noisy

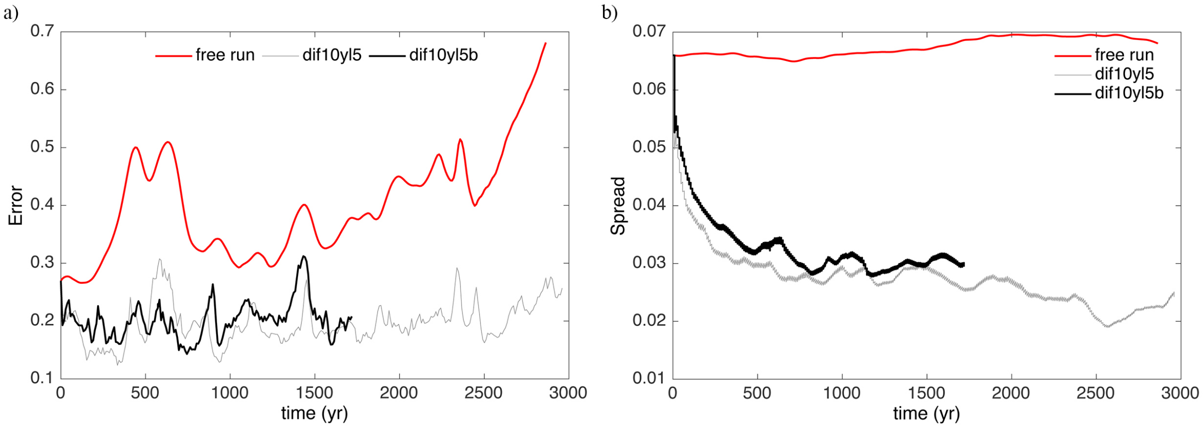


Figure 5.17: The same as Fig. 5.14, but comparing the influence on the geographic distribution of the dataset.

observations were created from a reference model trajectory of the CEL dynamo introduced in Section 2.8. The uncertainty affecting the data was chosen to be approximately representative of the archeomagnetic dataset.

The assimilation results of the closed-loop experiments show a remarkably good performance, both in terms of accuracy, precision and data fit. Comparison of the different fields composing the state vector reveals that the update of information given by the synthetic data set not only constrains well the observed part of the state vector, the poloidal magnetic field at the top of the core, but also the hidden variables of the system, like the magnetic field at depth, the flow field and the density anomaly. The good propagation of information between the different fields is possible due to two main aspects of the assimilation algorithm. One is the use of the correlations between the different variables embedded in the forecast covariance matrix in the analysis. The other is the propagation of information from the dynamical model.

The statistical convergence of the assimilation in terms of the precision proxy, the spread, suggests an assimilation spin-up time of the order 500 to 700 years (for the reference run dif10yl5, Fig. 5.10). After this time window, the accuracy of the assimilation reaches an error stagnation around a 0.2 level, and the precision decreases from the initial spread by a factor of two. The cumulative innovation points to the good forecast quality of the assimilation, but suggests a small bias $\sim 10^{-1}$ after 1,000 years. This issue remains to be investigated for a longer run of the assimilation experiment.

Given the good results of the reference assimilation test, the algorithm was tested for its sensitivity to the dataset characteristics and assimilation parameters. It was seen, for instance, that the assimilation is not sensitive to the data type, whether it consists of XYZ or DIF observations, at least for the case in which the point-wise observations uniformly sample the Earth's surface. The spectral assimilation showed different results, but is still too short in order to give any conclusions. Moreover, the analysis window, has an important impact on the assimilation. The assimilation with analysis cycle of $\Delta t_a = 40$ yr shows a considerable decrease in accuracy, despite maintaining a similar, but slightly lower, precision than that of the reference case. This results shows that shorter intervals for the analysis window are best suited for the archeomagnetic-like dataset. The sensitivity test of the analysis cycle time window should be further expanded to shorter times, in order to assess the practical predictability limit of the archeomagnetic assimilation.

Finally, the increase in data availability for the same archeomagnetic-like uncertainties is seen to not considerably influence the accuracy of the assimilation, although its precision is slightly increased. This conclusion points to the limit in resolution affected by the data uncertainties. It is likely that increasing the data quantity even more would not considerably improve the accuracy, while a general decrease in the data uncertainties most probably would. Still in terms of data availability, it was seen that a biased hemispherical observation distribution did not strongly affect the assimilation accuracy. This particular result has very important implications for real archeomagnetic data assimilation. It shows that the absence of observations over one of the hemispheres (Fig. 1.11 and Fig. 5.3), does not weaken the assimilation. Instead, the propagation of observation seems to profit from the high level of equatorial symmetry displayed by the model (see Fig.5.5 and Fig. 5.12). The extension of this conclusion to the real archeomagnetic assimilation is tempting. However, the degree of geostrophy on the core must be taken into account.

Chapter 6

Conclusions and perspectives

Inverse modeling of the archeomagnetic field is an important but difficult task which can allow for a deepening in our understanding of the processes responsible for the dynamics of the core. This thesis focuses on the inverse problem framework known as data assimilation to propose new ways of modeling the archeomagnetic field. It can be separated into two main parts, for which the main difference lies in the temporal aspect. The first part, detailed in Chapter 4, describes the static inverse problem by using prior information from numerical dynamo models. In this part, attention is drawn to the influence of the spatial correlations imposed by the prior and how it allows a robust analysis of the resolution of a given dataset, making possible the quantification of the archeomagnetic data resolution. The second part, presented in Chapter 5, introduces the time-dependent inverse problem in terms of a sequential assimilation framework. In this part, closed-loop experiments are used to assess the performance of the sequential assimilation of archeomagnetic-like synthetic observations. Both parts are summarized and discussed in the following.

6.1 Static modeling of the archeomagnetic field

It is evident from the heterogeneous distribution of the archeomagnetic data catalog, that it is not possible to reliably constrain the archeomagnetic field without the complement of additional information. Such information is classically based on an approximate condition for the minimization of Ohmic dissipation at the CMB, entering the inverse problem as an adjustable spatial model norm. In this thesis, that norm is substituted by dynamo-based statistical information. The resulting 'dynamo norm' consists of well-defined information about the correlations of the different spatial scales of the magnetic field, in this case, based on the Coupled Earth dynamo model (Aubert et al., 2013). Since the large spatial structures of the magnetic field are suspected to be invariant for models bearing the same magnetic turbulence level (Davidson, 2013), the dynamo norm is considered to be robust.

The dynamo norm being non-adjustable once the magnetic field rescaling is defined, it allows for the calculation of the resolution matrix, which represents the extent to which the data can resolve the model. It is found that the archeomagnetic data based on the GeomagiaV2 dataset (Donadini et al., 2009) resolves the magnetic field at different degrees given the studied time period of the last 3,000 years. For the first millennium BC, the field is constrained up to spherical harmonic degree 3. This resolution increases to degree 4 for the first millennium and closely reaches degree 5 for the second millennium AD. GeomagiaV3 (Brown et al., 2015) is seen to improve this resolution estimate, but within a narrow contribution. The archeomagnetic resolution matrix provides the ability to differentiate data-driven from prior-driven conditions for the model variables with time. The prior-driven coefficients therefore differ for models derived from different inverse modeling strategies. The comparison between the dynamo-based archeomagnetic field model AmR and well-known inverse archeomagnetic models (ARCH3k.1 and AFM) confirms the dichotomy between data-driven and prior-driven estimates.

The framework used in this time-independent study provided the possibility of validating different aspects of inverse modeling, for instance the observation operators corresponding to the archeomagnetic data. Such observation operators, projecting the field model variables into declination, inclination and intensity of the magnetic field at the Earth's surface, are nonlinearly related to the core magnetic field. Tested in the context of a closed-loop experiment, it was seen that the observation operators are weakly nonlinear. In an ensemble framework, the results from the direct application of the nonlinear observation operator were very similar to the ones calculated with an iterative scheme based on the linearization of the observation operator.

Given the robustness of the time-independent approach of the dynamo-based archeomagnetic field modeling, the next direction to go was that of the implementation of the temporal dependency of the model.

6.2 Sequential archeomagnetic assimilation

The temporal dependency in the data assimilation framework comes from the physical model in question itself, in its numerical approximation. In the geomagnetic case, this model is provided by numerical dynamo simulations. A lower resolution version of the Coupled Earth model, CEL, was chosen to represent the dynamo model used to serve as the dynamic prior in the assimilation. In order to handle the nonlinearities of model and observation operator, an ensemble framework, the Ensemble Kalman Filter, was chosen as the assimilating algorithm. In synthesis, the EnKF assimilates, in a sequence of analysis cycles, the observations by means of an ensemble of models describing the model statistics. The adaptation of the EnKF for the geomagnetic case was provided by the machinery behind the Parody-PDAF code (Fournier et al., 2013).

Since the state of the art of geomagnetic data assimilation rests on the assimilation of inverse field models, the risk is that the assimilated models display important bias towards its prior-driven parameters. That being said, actual point-wise measurements of the magnetic field at the surface should be assimilated instead. This problematic lead to the development of a modified version of Parody-PDAF which allows the assimilation of point-wise observations. In order to validate the algorithm in an archeomagnetic context, a series of closed-loop assimilation experiments was performed.

The assimilation of noisy synthetic declination, inclination and intensity data with a global surface close-to-uniform coverage corresponding to 35 observation sites every 10 years has shown a good performance in retrieving the hidden variables of the dynamo model. The assimilation spin-up time of the system was seen to range from 500 to 700 years, after which the accuracy of the assimilation stays at a mean value of approximately 20% and the precision increases by a factor of two with respect to its initial value. It is worth mentioning that keeping the same archeomagnetic-like uncertainties of the synthetic dataset but increasing data quantity was seen to not significantly improve the accuracy of the assimilation.

The main conclusion of the closed-loop experiments lies in the good response of the assimilation in the case of a strong bias in the geographic distribution of the synthetic data. Considering the data localized only on one hemisphere, the assimilation algorithm manages to show a level of accuracy and precision close to that when data from both hemispheres are used. The reason behind this good result is probably related to the high degree of equatorial symmetry displayed by the CEL model. In this case, information updated from observations over the North Hemisphere would be propagated to the South Hemisphere by means of the closely cylindric columns of the flow field. This conclusion opens important perspectives for the assimilation of real archeomagnetic data. Its feasibility, however, should be confronted with the uncertainties regarding the degree of geostrophy displayed by the core (Baerenzung et al., 2014).

6.3 Perspectives

This study paves the way for a series of perspectives associated with the use of the resolution of the archeomagnetic dataset and the reconstruction of the state of the core for the archeomagnetic period.

The archeomagnetic resolution matrix represents the degree of influence of the archeomagnetic observations on the magnetic field variables at the top of the core. The formalism used to construct such a matrix could be in principle used in the inverse problem with the purpose of searching for the best way to expand the archeomagnetic dataset. Given a set of constraints imposing the feasible locations for the archeomagnetic sampling and uncertainty of observations, the conditions embedded on the observation operator could be estimated by maximizing the resolution matrix. Such a framework could provide guidance for future archeomagnetic missions and help optimizing the expansion of the archeomagnetic databases.

From a methodological point-of-view, the sequential EnKF assimilation explored in the present study could be further improved, for example, by means of algorithms relying on the retrospective analysis of the model state. Given that the archeomagnetic databases (e.g. GeomagiaV3) are constantly being expanded, new data can be used in the reanalysis and smoothing of the previously analyzed archeomagnetic model trajectory.

The closed-loop assimilation experiments performed in this study serve as an introduction on the feasibility of sequential assimilation of archeomagnetic-like observations. Further testing of the framework should be extended to assess the sensitivity of the assimilation to the observation uncertainties, to even more biased scenarios for the data distribution and more important, to the model itself. In order to reveal the sensitivity of the data assimilation to the underlying model, the scheme should be tested in a fraternal experiment. In such case, one type of dynamo model is used to be the reference, or true, solution, while another type of dynamo is used to provide the ensemble forecast information. The monitoring of the assimilation in the fraternal context would give an idea of the expected performance of the assimilation in the case of real archeomagnetic data assimilation. In the positive case, the real archeomagnetic assimilation could help to unveil the hidden processes in the core responsible for the observed secular variation of the Earth's magnetic field.

Finally, the geomagnetic sequential data assimilation framework can also be extended to other types of observation beyond the archeomagnetic data. Historical and modern observatory and satellite data are able to register, for instance, shorter time scale variations of the magnetic field than the archeomagnetic one, like decadal variations. Decadal time scales are particularly interesting for the unveiling of the core processes, since their origins bear a certain degree of uncertainty, either thought to be the contribution of dynamo waves or from advective motions in the core. The investigation of such question through geomagnetic data assimilation could help increase our understanding of the core dynamics over shorter time scales.

Appendix A

The predictability of advection-dominated flux-transport solar dynamo models

The Astrophysical Journal, 781:8 (15pp), 2014 January 20 doi:10.1088/0004-637X/781/1/8

THE PREDICTABILITY OF ADVECTION-DOMINATED FLUX-TRANSPORT SOLAR DYNAMO MODELS

Sabrina Sanchez, Alexandre Fournier, and Julien Aubert
Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Université Paris Diderot UMR 7154 CNRS, F-75005 Paris, France; ssanchez@ipgp.fr

Received 2013 May 30; accepted 2013 December 2; published 2013 December 24

ABSTRACT

Space weather is a matter of practical importance in our modern society. Predictions of forecoming solar cycles mean amplitude and duration are currently being made based on flux-transport numerical models of the solar dynamo. Interested in the forecast horizon of such studies, we quantify the predictability window of a representative, advection-dominated, flux-transport dynamo model by investigating its sensitivity to initial conditions and control parameters through a perturbation analysis. We measure the rate associated with the exponential growth of an initial perturbation of the model trajectory, which yields a characteristic timescale known as the *e*-folding time τ_e. The *e*-folding time is shown to decrease with the strength of the α-effect, and to increase with the magnitude of the imposed meridional circulation. Comparing the *e*-folding time with the solar cycle periodicity, we obtain an average estimate for τ_e equal to 2.76 solar cycle durations. From a practical point of view, the perturbations analyzed in this work can be interpreted as uncertainties affecting either the observations or the physical model itself. After reviewing these, we discuss their implications for solar cycle prediction.

Key words: chaos – dynamo – Sun: activity

Online-only material: color figures

1. INTRODUCTION

The Sun is a magnetic active star, which undergoes successive phases of high and low magnetic activity with a quasi-periodicity of approximately 11 yr, powered by a natural dynamo mechanism (Moffatt 1978). This magnetic activity encompasses the recurrent manifestation of dynamical phenomena at the solar surface and in its atmosphere, such as sunspots, flares, and coronal mass ejections (Priest 1982). In addition to its remarkable regularity, solar activity exhibits longer term (decadal to centennial) fluctuations (Hathaway 2009), and occasional periods of long-lasting near-quiescence, such as the Maunder Minimum. Since the solar cycle affects the energy radiated by the Sun, its understanding is key in elucidating the potential control of solar activity on the long-term variability of the Earth's climate (Haigh 2003).

Solar activity influences the terrestrial environment in other important aspects, connected with the operation of satellites (Baker 2000), and the occurrence of geomagnetic storms, which can damage electric power grids and interfere with radars and radio communications. These important issues highlight the strong need for an accurate prediction of solar magnetic phenomena, which is one of the main goals of space weather (Pulkkinen 2007). Until recently, such forecasting exercises were mostly conducted within an entirely data-driven framework, based, for instance, on geomagnetic precursors methods (Hathaway 2009; Wang & Sheeley 2009). It is sensible to believe, though, that more accurate and effective predictions could be obtained by combining these data with physical models of the Sun, using data assimilation (e.g., Talagrand 1997). The most salient illustration of the application of data assimilation emanates every day from numerical weather prediction (NWP) centers, in the form of weather forecasts (consult, e.g., Kalnay 2003 for a historical perspective on NWP). Application of data assimilation in geoscience also include oceanography (e.g., Brasseur 2006), the study of air quality (e.g., Elbern et al. 2010), and land surfaces (e.g., Houser et al. 2010). In a context similar to that of the solar dynamo, data assimilation has also recently come to the fore for the study of the Earth's dynamo, a surge motivated by our increased ability to observe and simulate the geomagnetic field (e.g., Fournier et al. 2007, 2010, 2013; Aubert & Fournier 2011). Over the past 15 yr, the study of the solar dynamo has witnessed an even more spectacular increase in its observational and modeling capabilities. The question of the feasibility of applying data assimilation techniques to the solar dynamo was asked a few years ago (Brun 2007), and was followed by a series of studies bearing promises (Kitiashvili & Kosovichev 2008; Jouve et al. 2011; Dikpati & Anderson 2012).

The physical model of the solar dynamo that should enter this inverse problem machinery remains to be defined. Forward modeling of the solar dynamo has shed light on the main physical processes believed to be responsible for the solar cycle (see Charbonneau 2005 for a review). Kinematic dynamo theory stresses that these processes are connected with the continuous transformation of poloidal magnetic energy into toroidal magnetic energy (the $P \rightarrow T$ conversion), and vice versa (the $T \rightarrow P$ conversion, necessary to close the dynamo loop). There is now little doubt that the Ω-effect, which denotes the shearing action of the differential rotation of the plasma flow, is responsible for the $P \rightarrow T$ conversion. Through the advent of helioseismology, the large-scale, interior, differential rotation was mapped in detail (Tomczyk et al. 1995), which made it possible to infer that the most likely location of the Ω-effect is the base the convection zone, a region known as the tachocline (Howe et al. 2000). There is less consensus regarding the processes at work behind the $T \rightarrow P$ conversion. The mean-field α-effect (Parker 1955), and the Babcock-Leighton (BL) mechanism (Babcock 1961; Leighton 1969) are two commonly envisioned possibilities. The former rests on the large scale effect of small scale turbulent motions whose twisting action can transform a toroidal field line into a poloidal field line. The latter relies on empirical evidences of the process of diffusion and reconfiguration of the magnetic field of sunspots. The three aforementioned processes (α-effect, Ω-effect, and BL mechanism) are illustrated in Figure 1.

The ambiguity between the α-effect scenario and the BL mechanism would disappear, should one be in a position to carry out the full three-dimensional numerical integration

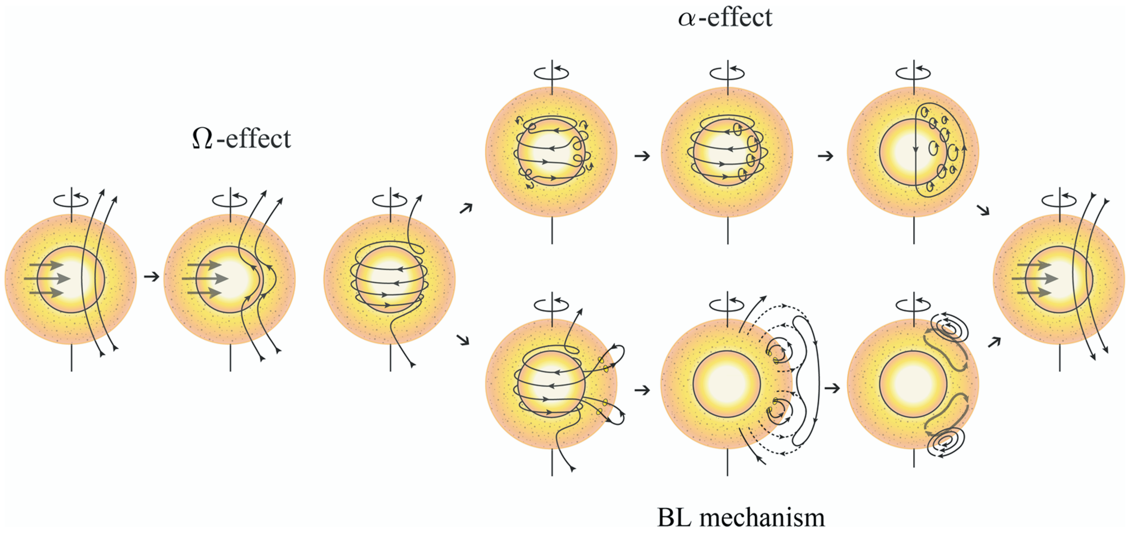


Figure 1. Illustration of the main processes at work in our solar dynamo model. The Ω-effect (left) depicts the transformation of a primary poloidal field into a toroidal field by means of the differential rotation. The poloidal field regeneration is next accomplished either by the α-effect (top) and/or by the Babcock-Leighton mechanism (bottom). In the α-effect case, the toroidal field at the base of the convection zone is subject to cyclonic turbulence. Secondary small-scale poloidal fields are thereby created, and produce on average a new, large-scale, poloidal field. In the Babcock-Leighton mechanism, the primary process for poloidal field regeneration is the formation of sunspots at the solar surface from the rise of buoyant toroidal magnetic flux tubes from the base of the convection zone. The magnetic fields of those sunspots nearest to the equator in each hemisphere diffuse and reconnect, while the field due to those sunspots closer to the poles has a polarity opposite to the current one, which initiates a polarity reversal. The newly formed polar magnetic flux is transported by the meridional flow to the deeper layers of the convection zone, thereby creating a new large-scale poloidal field.

(A color version of this figure is available in the online journal.)

of those equations governing the solar dynamo. Despite the monotonic and dramatic increase in compute power which already led to substantial achievements (e.g., Brun et al. 2004; Charbonneau & Smolarkiewicz 2013), such a comprehensive integration remains out of reach due to the wide range of temporal and spatial scales induced by the high level of turbulence expected inside the solar convection zone. On the other hand, and from a more practical perspective, a large body of work has shown that axisymmetric mean-field solar dynamo models were able to reproduce many of the observed features of solar activity (Charbonneau 2005). The most recent and representative illustrations of this strand rely on the advection of magnetic flux by a meridional flow (following in general the BL mechanism). These models, called "flux-transport" models, are in particular successful in accounting for the equatorward migration of the solar toroidal field and the observed phase-locking of the solar cycle (Dikpati & Charbonneau 1999; Charbonneau & Dikpati 2000).

Such flux-transport models may make it possible to predict the amplitude and duration of the upcoming solar cycles. The first studies addressing this possibility (Dikpati et al. 2006; Choudhuri et al. 2007) considered direct incorporation of data into models, essentially by imposing (in a strong sense) surface boundary values inherited from the data onto the model, whereas an assimilation scheme would require this to happen in a weak sense, through some flavor of the so-called best linear unbiased estimator, whose goal is to combine in an optimal fashion the data and the model, considering the uncertainties affecting both. Independently of the data assimilation scheme one may resort to, and as good as it may be, there exists an intrinsic limit to its predictive power. Bushby & Tobias (2007) point out that this limit arises either from the stochastic nature of the BL and α-effects, or from nonlinear deterministic processes. They stress, in addition, that the lack of constraints on the exact nature of the key physical mechanisms which sustain these models and govern their time-dependency, such as the α-effect, make their ability to capture the essentials of the solar dynamo process questionable. They conclude that under the best circumstances of a near-perfect model, the shape of the solar cycle could only be predicted one or two cycles ahead. As this best case scenario is out of reach, they argue that a reliable forecasting exercise is untractable.

The same critic was made regarding weather prediction during its early years. The seminal work by Lorenz (1963) showed the extreme sensitivity of a deterministic system governed by a simple set of nonlinear coupled differential equations to its initial conditions. In a subsequent study, Lorenz (1965) estimated the timescale of divergence τ of two initially very close dynamical trajectories (called *twin* trajectories in the following) to be of a few days (Lorenz's simple model aimed at representing atmospheric convection). More realistic models of the atmosphere have now established that τ is equal to two weeks. This value has to be confronted with the current forecast horizon of NWP, which is (depending on the center) between seven and nine days. The combined progress of observation, models, and data assimilation algorithms over the past 30 yr has resulted roughly in a gain of one day per decade, bringing the operational limit closer and closer to the theoretical limit.

One may wonder to which extent the progress made by the atmospheric community could be expected within the solar community. Doing so, one immediately realizes that these two dynamical systems (the atmosphere and the Sun) are dramatically different. Whereas the Earth's atmosphere is a thin and directly observable layer, the solar convection zone is an almost entirely concealed thick shell. Moreover, the physics of the atmosphere is much better constrained than that at work behind the solar dynamo (consult Vallis 2006 for a review of atmospheric processes). Bearing these substantial differences in mind, and assuming that the basic physics involved in the solar dynamo is faithfully captured by mean-field models, one may

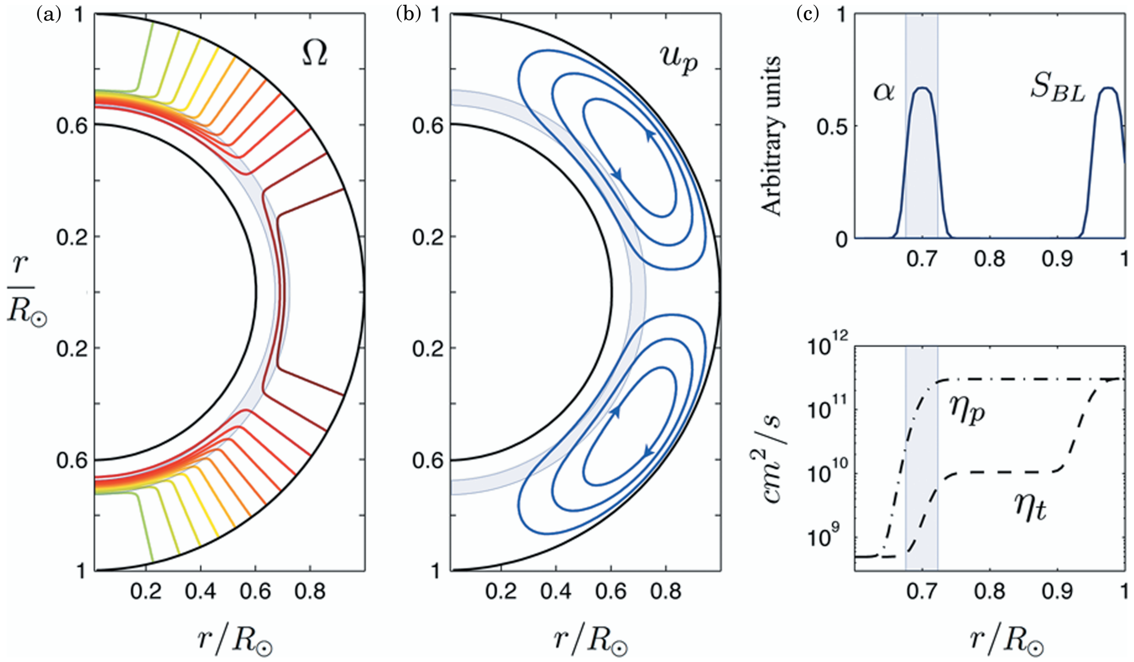


Figure 2. Components defining the class of solar dynamo models used in this study. (a) Isocontours of the angular velocity Ω; (b) Meridional circulation streamlines; (c) radial profiles of the α-effect and Babcock-Leighton poloidal source terms; (d) radial profiles of the magnetic diffusivities. In each panel, the shaded regions symbolize the tachocline. Aside from the differential rotation, whose amplitude remains fixed in this study, we vary the magnitude of these various components, whose relative contributions are described by a suite of non-dimensional numbers (see the text for details).

(A color version of this figure is available in the online journal.)

still hope that the short-term prediction of at least some of the features of the solar cycle (e.g., duration and mean amplitude) is possible.

Knowledge of the modulations and mean intensity of the upcoming solar cycles from mean-field models may serve as an important input for more specific space weather considerations. In this study, we therefore wish to adopt an operational perspective. Assuming that mean-field models will be effectively used to forecast solar activity, our goal here is to quantify their intrinsic limit of predictability τ (the equivalent of the two weeks discussed above for the atmosphere), following the methodology proposed recently by Hulot et al. (2010) and Lhuillier et al. (2011) in order to estimate τ for the Earth's dynamo.

This paper is organized as follows. In Section 2, we describe our working mean-field model and detail its numerical implementation. We next inspect the sensitivity of this model to its control parameters in Section 3. Section 4 presents the systematic study of the error growth between twin trajectories. This allows us to evaluate τ, and to assess its sensitivity to its control parameters. Finally, we discuss in Section 5 the influence of modeling and observational errors on the practical limit of predictability of the model.

2. THE MODEL AND ITS NUMERICAL IMPLEMENTATION

Our flux-transport model is the one presented by Sanchez et al. (2014); it includes both the α and BL scenarios for the $T \rightarrow P$ conversion. The first reason for adding an α-effect to a standard BL flux-transport model is that a dynamo running on a BL mechanism alone cannot recover from a quiescent phase devoid of sunspots. As reported by Sanchez et al. (2014), the model setup enables the appearance of a long-term variability (succession of active and quiet phases), which can then be interpreted as the result of the competition between the α-effect operating at the tachocline and a BL mechanism operating at the solar surface. In addition, a deep location of the α-effect is known to favor the sought antisymmetrical evolution of the magnetic field in the Northern and Southern hemispheres (Dikpati & Gilman 2001; Bonanno et al. 2002).

Let us now write accordingly the modified mean-field induction equation (Moffatt 1978) for the large-scale magnetic field **B**

$$\frac{\partial \mathbf{B}}{\partial t} = \boldsymbol{\nabla} \times [\mathbf{U} \times \mathbf{B} - \eta \boldsymbol{\nabla} \times \mathbf{B} + \alpha \mathbf{B} + S_{\mathrm{BL}} B_\varphi \hat{\mathbf{e}}_\varphi], \qquad (1)$$

where **U** is the prescribed flow, η is the turbulent diffusivity, α is the turbulent magnetic helicity, and $S_{\mathrm{BL}} B_\varphi \hat{\mathbf{e}}_\varphi$ is the BL source term ($\hat{\mathbf{e}}_\varphi$ is the unit vector in the direction of longitude). We will specify the profiles of these various physical fields in the following. The definitions that we will need are summarized in Table 1 and the profiles shown in Figure 2.

Under the assumption of axisymmetry, the magnetic and flow fields are further expressed in terms of their poloidal and toroidal components in spherical coordinates (r, θ, φ) as

$$\mathbf{B}(\mathbf{r}, \mathbf{t}) = \boldsymbol{\nabla} \times [A_\varphi(\mathbf{r}, t)\hat{\mathbf{e}}_\varphi] + B_\varphi(\mathbf{r}, t)\hat{\mathbf{e}}_\varphi, \qquad (2)$$

$$\mathbf{U}(\mathbf{r}) = \mathbf{u}_p(\mathbf{r}) + r \sin\theta\, \Omega(\mathbf{r})\hat{\mathbf{e}}_\varphi, \qquad (3)$$

in which A_φ is the poloidal potential and B_φ is the toroidal field. The prescribed time-independent flow is defined by the angular velocity Ω and the meridional circulation $\mathbf{u}_p$, shown in Figures 2(a) and (b), respectively. Helioseismic data provide strong constraints on Ω, which will thus remain fixed in the remainder of this work, and approximated using the analytic formula of Dikpati & Charbonneau (1999). On the

Table 1

Summary of the Mathematical Symbols Used in the Model, Their Values and a Brief Explanation of Their Meaning

Symbol	Value	Interpretation
r_{tc}	$0.7\, R_\odot$	Radial location of the center of the tachocline
δr	$0.05\, R_\odot$	Thickness of the tachocline
$\Omega_{\rm eq}$	$2\pi \times 460.7$ nHz	Rotation rate at the equator
α_0	0.34–1.03 m s^{-1}	Strength of the α-effect
$S_{\rm BL_0}$	0.02–0.06 m s^{-1}	Strength of the Babcock-Leighton mechanism
u_0	13.27–17.68 m s^{-1}	Velocity of the superficial meridional flow at mid-latitude
η_r	5×10^8 cm^2 s^{-1}	Effective diffusivity near the radiative zone
η_{cz}	1×10^{10} cm^2 s^{-1}	Effective diffusivity at the bottom of the convection zone
η_s	3×10^{11} cm^2 s^{-1}	Effective diffusivity at the solar surface

contrary, the large-scale meridional circulation $\mathbf{u}_p$ remains poorly constrained. For the sake of simplicity, we will follow the one-cell per hemisphere description of Dikpati & Charbonneau (1999).

The poloidal–toroidal decomposition of the magnetic and flow fields prompts us to define poloidal and toroidal components for the turbulent diffusivity η, denoted by η_p and η_t, respectively. This distinction rests on the analysis made by Chatterjee et al. (2004), who pointed out that the toroidal field strength is expected to be much larger than the poloidal field strength throughout the convection zone. This should decrease notably the efficiency of toroidal turbulent diffusion compared with its poloidal counterpart. With this distinction at hand, injection of Equations (2) and (3) into Equation (1) gives rise to a set of two coupled partial differential equations for A_φ and B_φ

$$\begin{aligned}\frac{\partial A_\varphi}{\partial t} + \frac{\mathbf{u}_p}{r\sin\theta}\cdot\boldsymbol{\nabla}(r\sin\theta A_\varphi) = {} & \eta_p\left(\nabla^2 - \frac{1}{r^2\sin^2\theta}\right)A_\varphi \\ & + \alpha(r,\theta;B_\varphi)B_\varphi \\ & + S_{\rm BL}\big(r,\theta;B_\varphi^{tc}\big)B_\varphi^{tc}, \qquad (4)\end{aligned}$$

$$\begin{aligned}\frac{\partial B_\varphi}{\partial t} + r\sin\theta\boldsymbol{\nabla}\cdot\left(\frac{\mathbf{u}_pB_\varphi}{r\sin\theta}\right) = {} & \eta_t\left(\nabla^2 - \frac{1}{r^2\,\sin^2\theta}\right)B_\varphi \\ & + \frac{1}{r}\frac{\partial\eta_t}{\partial r}\frac{\partial(rB_\varphi)}{\partial r} \\ & + r\,\sin\theta(\boldsymbol{\nabla}\times A_\varphi\hat{\mathbf{e}}_\varphi)\cdot(\boldsymbol{\nabla}\Omega), \qquad (5)\end{aligned}$$

where $B_\varphi^{tc} = B_\varphi(r = r_{tc},\theta,t)$ is the toroidal field at the tachocline, defined in this work as the spherical shell of mean radius $r_{tc} = 0.7\,R_\odot$, with a thickness $\delta r = 0.05\,R_\odot$. The dependency of the $S_{\rm BL}$ term in Equation (4) expresses the non-local character of the BL source term. Even if it is active within the surface layers, the BL regeneration process is thought to originate from processes occurring in the vicinity of the tachocline—numerical models indeed indicate that the formation of tilted bipolar regions at the surface is mostly controlled by the strength of toroidal flux tubes prior to their buoyant instability (D'Silva & Choudhuri 1993). Their finite rise time should induce a time lag between the onset of the instability and the formation of the bipolar regions, on the order of some days to a few weeks (Jouve et al. 2010). We will neglect this delay on the account of it being small compared to the timescales of interest here.

Turning now our attention to the α-effect, we use the standard formula of α-quenching, which is written as

$$\alpha(\mathbf{r};B_\varphi) = \frac{\alpha_0}{1+\left(\frac{B_\varphi}{B_{\rm eq}}\right)^2}f_\alpha(\mathbf{r}), \qquad (6)$$

in which α_0 is a typical magnitude, $B_{\rm eq} = 10^4$ G (Fan 2009) and $f_\alpha(\mathbf{r})$ restricts the α-effect to the mid-latitudes of the tachocline, according to

$$f_\alpha(\mathbf{r}) = \frac{1}{4}\left[1+\mathrm{erf}\left(\frac{r-r_1}{d_1}\right)\right]\left[1-\mathrm{erf}\left(\frac{r-r_2}{d_2}\right)\right]\cos\theta\sin\theta, \qquad (7)$$

where $r_1 = r_{tc} - \delta r/2$, $r_2 = r_{tc} + \delta r/2$, and $d_1 = d_2 = 0.01\,R_\odot$. The radial variations of f_α are shown in Figure 2(c).

The BL $S_{\rm BL}$ source term operates within bounds of the magnetic field strength (D'Silva & Choudhuri 1993), specifically between $B_{\varphi,\min}^{tc} = 10^4$ G and $B_{\varphi,\max}^{tc} = 10^5$ G. Denoting the magnitude of this source term by $S_{\rm BL_0}$, we write accordingly

$$\begin{aligned}S_{\rm BL}\big(\mathbf{r};B_\varphi^{tc}\big) = {} & \frac{S_{\rm BL_0}}{4}\left[1+\mathrm{erf}\left(B_\varphi^{tc\,2} - B_{\varphi,\min}^{tc\,2}\right)\right] \\ & \times\left[1-\mathrm{erf}\left(B_\varphi^{tc\,2} - B_{\varphi,\max}^{tc\,2}\right)\right]f_{\rm BL}(\mathbf{r}). \qquad (8)\end{aligned}$$

The radial and latitudinal distribution $f_{\rm BL}(\mathbf{r})$ is in turn given by

$$\begin{aligned}f_{\rm BL}(\mathbf{r}) = {} & \frac{1}{4}\left[1+\mathrm{erf}\left(\frac{r-r_3}{d_3}\right)\right] \\ & \times\left[1-\mathrm{erf}\left(\frac{r-r_4}{d_4}\right)\right]\cos\theta\sin\theta, \qquad (9)\end{aligned}$$

where $r_3 = 0.95\,R_\odot$, $r_4 = R_\odot$ and $d_3 = d_4 = 0.01\,R_\odot$. The radial distribution of $f_{\rm BL}$ is shown in Figure 2(c).

The poloidal and toroidal diffusivities in Equations (4) and (5), are written as

$$\eta_p(r) = \eta_r + \eta_s\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{r-r_5}{d_5}\right)\right], \qquad (10)$$

$$\begin{aligned}\eta_t(r) = {} & \eta_r + \eta_{cz}\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{r-r_6}{d_6}\right)\right] \\ & + \eta_s\frac{1}{2}\left[1+\mathrm{erf}\left(\frac{r-r_7}{d_7}\right)\right], \qquad (11)\end{aligned}$$

in which $r_5 = 0.7\,R_\odot$, $r_6 = 0.72\,R_\odot$, $r_7 = 0.95\,R_\odot$, $d_5 = d_6 = d_7 = 0.025\,R_\odot$, η_r is the diffusivity at the boundary with the radiative zone, η_{cz} is the diffusivity in the turbulent convection zone, and η_s is the diffusivity in the surface layers (which applies to the poloidal field over the entire convection zone). The radial profiles of the diffusivities are shown in Figure 2(d). This model pertains to the generic class of advection-dominated models: owing to the low values of the diffusivities throughout the convection zone, the coupling between the regions where the poloidal and toroidal fields are generated is ensured by the meridional circulation. In diffusion-dominated models (e.g., Chatterjee et al. 2004), this coupling is on the contrary accomplished by turbulent diffusion.

In order to express the dynamo equations in their nondimensional form, we choose the solar radius $R_\odot$ as the length scale and the magnetic diffusion time $R_\odot^2/\eta_s$ as the timescale (roughly equal to 500 yr). This yields

$$\frac{\partial A_\varphi}{\partial t} + \frac{\mathrm{Rm}}{r\sin\theta}\tilde{\mathbf{u}}_p \cdot \boldsymbol{\nabla}(r\sin\theta A_\varphi) = \tilde{\eta}_p\left(\nabla^2 - \frac{1}{r^2\sin^2\theta}\right)A_\varphi + C_\alpha\tilde{\alpha}\,B_\varphi + C_{\mathrm{BL}}\tilde{S}_{\mathrm{BL}}\,B_\varphi^{tc}, \tag{12}$$

$$\frac{\partial B_\varphi}{\partial t} + \mathrm{Rm}\,r\sin\theta\boldsymbol{\nabla}\cdot\left(\frac{\tilde{\mathbf{u}}_p B_\varphi}{r\sin\theta}\right) = \tilde{\eta}_t\left(\nabla^2 - \frac{1}{r^2\,\sin^2\theta}\right)B_\varphi + \frac{1}{r}\frac{\partial\tilde{\eta}_t}{\partial r}\frac{\partial(rB_\varphi)}{\partial r} + C_\Omega\,r\,\sin\theta \times(\boldsymbol{\nabla}\times A_\varphi\hat{\mathbf{e}}_\varphi)\cdot(\boldsymbol{\nabla}\tilde{\Omega}). \tag{13}$$

Equations (12) and (13) contain six nondimensional numbers characterizing the relative importance of each term in the equations

$$\mathrm{Rm} = u_o R_\odot/\eta_s, \tag{14}$$

$$C_\Omega = \Omega_{\mathrm{eq}} R_\odot^2/\eta_s, \tag{15}$$

$$C_\alpha = \alpha_0 R_\odot/\eta_s, \tag{16}$$

$$C_{\mathrm{BL}} = S_{\mathrm{BL}_0} R_\odot/\eta_s, \tag{17}$$

$$\text{the ratio } \eta_r/\eta_s, \tag{18}$$

$$\text{and the ratio } \eta_{cz}/\eta_s. \tag{19}$$

The magnetic Reynolds number Rm is associated with the amplitude of the large-scale meridional flow, u_0. The three following coefficients C_Ω, C_α, C_{BL}, respectively, express the ratio of the equatorial rotation, turbulent and BL timescales to the diffusive timescale. In these expressions, Ω_{eq} is the equatorial rotation rate, and α_0 and S_{BL_0} are the amplitudes of the α and BL terms seen above. The remaining two terms η_r/η_s and η_{cz}/η_s are magnetic diffusivity ratios entering the nondimensional forms of Equations (10) and (11). The $\sim$ in Equations (12) and (13) denotes normalization with respect to those quantities. Note that a suitable rescaling of A_φ can decrease the number of control parameters by one, as it can scale either C_α or C_{BL} out of the problem (it is the ratio of these two that would remain). Albeit more elegant, we did not consider this possibility. We shall therefore analyze the α and BL effects independently in the remainder of this study.

Finally, our formulation has to be complemented with boundary and initial conditions. The inner boundary condition is that of a perfect conductor. An approximation of this condition is that

$$A_\varphi = B_\varphi = 0 \text{ at the inner radius } r = 0.6\,R_\odot \text{ (Chatterjee et al. 2004)}. \tag{20}$$

The outer boundary condition corresponds to the interface with an insulating medium, and requires matching of the internal solar field with a potential field (Dikpati & Charbonneau 1999).

As an initial condition, we choose a dipolar field confined inside the convection zone. In this case,

$$A_\varphi(\mathbf{r}, t=0) = \sin\theta/r^2 \text{ for } 0.7\,R_\odot \leqslant r \leqslant R_\odot, \tag{21}$$

$$A_\varphi(\mathbf{r}, t=0) = 0 \text{ elsewhere}, \tag{22}$$

$$B_\varphi(\mathbf{r}, t=0) = 0 \text{ everywhere}. \tag{23}$$

The numerical approximation of the problem at hand is based on the Parody code, which was originally designed for three-dimensional geodynamo simulations (Dormy et al. 1998; Aubert et al. 2008), and successfully passed the dynamo benchmark of Christensen et al. (2001). The magnetic field is expanded according to the three-dimensional poloidal-toroidal decomposition

$$\mathbf{B} = \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times(\mathcal{P}\mathbf{r}) + \boldsymbol{\nabla}\times(\mathcal{T}\mathbf{r}), \tag{24}$$

where the poloidal and toroidal scalar potentials $\mathcal{P}$ and $\mathcal{T}$ are further expanded upon an axisymmetric spherical harmonic basis $Y_n^0(\theta)$, according to

$$(\mathcal{P}, \mathcal{T})(r, \theta, t) = \sum_{n=1}^{N}(\mathcal{P}_n, \mathcal{T}_n)(r, t)\;Y_n^0(\theta), \tag{25}$$

and truncated at spherical harmonic degree N. The discretization is completed by applying a second-order finite differencing in radius and second order time integration, comprising a Crank-Nicolson scheme for the diffusive terms and a second order Adams-Bashforth scheme for the nonlinear terms. The resulting code was then successfully tested against the reference solutions of Jouve et al. (2008). Details of this benchmark are provided in the Appendix. The results presented in what follows were obtained using $N = 65$, and $N_r = 65$ uniform radial levels in $[0.6\,R_\odot, R_\odot]$, and a constant non-dimensional time step size $\Delta t = 5\times10^{-6}$. A typical run comprised 10^7 time steps, which corresponds roughly to 25,000 yr.

3. FORWARD MODELING: MODEL PROPERTIES AND VARIABILITY

With our operational purpose in mind, a representative solution of the model should match some of the basic solar cycle features (Charbonneau 2005): cyclic polarity reversals with approximately 11 yr periodicity; strong toroidal fields at the base of the convection zone migrating from mid-latitudes toward the equator; poleward migration of a weaker high-latitude magnetic field; phase lag of $\pi/2$ between the toroidal field at mid-latitudes and polar field at the poles; antisymmetry of the magnetic field between the northern and southern hemispheres; and long-term variability of the solar cycle.

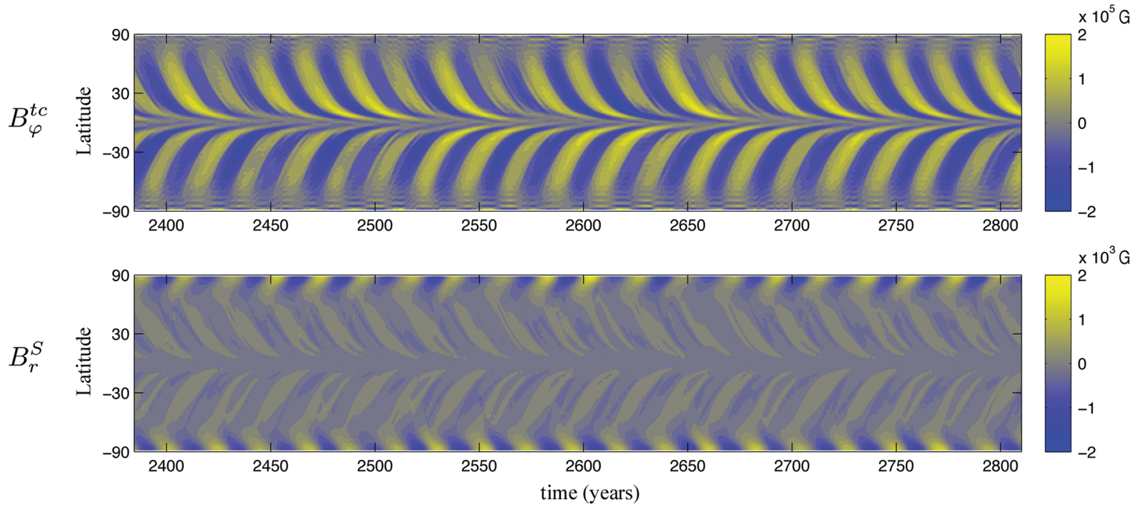


Figure 3. Time-latitude (butterfly) diagrams of the reference solution S, corresponding to $\mathrm{Rm} = 318$, $C_\alpha = 8$ and $C_{\mathrm{BL}} = 1$. Top: toroidal magnetic field at the tachocline; bottom: radial magnetic field at the solar surface.

(A color version of this figure is available in the online journal.)

In the following, we will impose the fixity of some of those non-dimensional numbers appearing in Equations (14)–(19). As helioseismological data give $\Omega_{\mathrm{eq}} \sim 2\pi \times 460.7$ nHz, we set accordingly $C_\Omega = 4.710^4$. In addition, the turbulent diffusivity in the solar interior is not well constrained (Ossendrijver 2003), and we consequently hold for simplicity the ratios η_r/η_s and η_{cz}/η_s fixed to values (see Table 1) previously shown to yield a satisfactory degree of solar semblance (e.g., Dikpati & Charbonneau 1999). Variations of the remaining free parameters C_α, C_{BL}, and Rm allow for a broad range of solutions. Rm represents the strength of the meridional circulation and controls the periodicity of the solar cycle, a well-known characteristic of flux transport dynamos (Dikpati & Charbonneau 1999). Consequently, and because of the strong observational constraint to obtain a period close to 11 yr, our family of models works in relatively narrow range of Rm. As the meridional flow measured at the solar surface at mid-latitudes has an average magnitude u_0 of 15 m s^{-1} (Hathaway 1996), we vary Rm between 308 and 378 ($u_0 \approx 13$ and 17 m s^{-1}, respectively). Within this range, getting a self-sustained reversing dynamo requires $C_\alpha \gtrsim 2$ and $C_{\mathrm{BL}} \gtrsim 0.5$.

We pick a reference (standard) solution (labeled S in the following) which has $\mathrm{Rm} = 318$, $C_\alpha = 8$ and $C_{\mathrm{BL}} = 1$; it generates quasi-periodic reversals, separated by approximately 10.95 yr. Figure 3 represents the simulated evolutions of the toroidal field at the tachocline, B_ϕ^{tc}, and of the radial field at the surface, B_r^S. It illustrates that the criteria for solar semblance which we listed are essentially met. This does not include the equatorial antisymmetric field configuration, a known recurring issue with BL models (Chatterjee et al. 2004; Charbonneau 2005). In this respect, Dikpati & Gilman (2001) and Bonanno et al. (2002) previously showed that the addition of an α-effect in a thin layer above the tachocline (as done here, recall Figure 2(c)) helps in obtaining antisymmetric solutions. However, and even if the portion of the dynamical trajectory represented in Figure 3 does display an antisymmetric magnetic field configuration, let us stress that there does not seem to exist a clear preferred mode of operation for the magnetic field: periods of symmetric, antisymmetric, and out-of-phase modes alternate over the dynamical trajectory followed by the standard model.

Long-term variability of the solar cycle is also present in this reference solution. Charbonneau et al. (2005) point out that chaotic modulation is a characteristic of BL models in which the BL term includes a lower operational threshold, as is the case in our model. Short periods of weaker than average activity level, lasting for approximately three cycles, are frequently found in our simulations, over a vast range of input parameters. In addition to this short-term variability, some of the solutions we obtain (including the reference solution S) display as well long periods of grand minima, lasting for several centuries, during which the cycle is not fully developed, but persists with a residual activity (see Sanchez et al. 2014 for more details). The occurrence of long periods of minimum activity is rare in our simulations; we chose accordingly to focus on their regular, quasi-cyclic behavior to carry out the predictability analysis exposed below.

The quantities C_α and C_{BL} (recall their definition in Equations (16) and (17)) are less tightly constrained by observations than Rm, and they will constitute the effective degrees of freedom of our class of models when we investigate its horizon of predictability in the following section. Variations in C_α and C_{BL} affect the overall morphology of the solar cycle in different ways. While an increase in C_α tends to excite higher frequencies during the solar cycle, it does not result in strong alterations of the magnetic field strength and cycle periodicity. On the other hand, the intensity of the magnetic field is strongly and irregularly sensitive to variations of C_{BL}—the overall trend is that it grows with C_{BL}. Increasing C_{BL} also usually results in the appearance of a feature respecting the Gnevyshev-Ohl rule, which is the persistent pattern of alternating high and low amplitudes of the solar cycles (Hathaway 2010). A too large an increase, though, gives rise to intermittent, non-solar-like, solutions. This forces us to define an upper bound of 2 for any admissible C_{BL}. On the other hand, as the main role of Rm is that of setting the pace of the solar cycle, increasing its value leads to a shortening of the simulated periodicity (note that the first columns of Table 3, which we will discuss further below, document in detail this variability).

4. PREDICTABILITY ANALYSIS

4.1. Methodology

Our mean-field solar dynamo model is a dynamical system, characterized by a limited range of predictability, owing to its chaotic nature (Lorenz 1963). As stated in the Introduction, two initially very close, *twin*, dynamical trajectories are bound to diverge in a finite time τ. The analysis of the divergence between these twin trajectories forms the backbone of our methodology; it is based on the work carried out by Hulot et al. (2010) and Lhuillier et al. (2011) to study the limit of predictability of the geodynamo.

We create a twin from a reference trajectory by perturbing a field variable (or control parameter) ξ at a given instant t_p in the following way

$$\xi(t_p) \longmapsto \widetilde{\xi}(t_p) = \xi(t_p)(1+\varepsilon), \tag{26}$$

where $\widetilde{\xi}$ and ε are the perturbed quantity and the amplitude of the perturbation, respectively.

Of importance for the assessment of the predictability is the evolution of the distance between the two trajectories over time. In order to monitor this distance, we resort to two pointwise measures, which are related to the toroidal field B_φ at a point $\mathbf{r}_{tc} \equiv (r = r_{tc}, \theta = 70^{\circ})$ on the tachocline, and to the radial field B_r at a point $\mathbf{r}_S \equiv (r = R_\odot, \theta \sim 2^{\circ})$ at the solar surface. These measures write

$$\Delta B_\varphi(\mathbf{r}_{tc}, t) \equiv \frac{|B_\varphi(\mathbf{r}_{tc}, t) - \widetilde{B}_\varphi(\mathbf{r}_{tc}, t)|}{\sqrt{\langle B_\varphi^2(\mathbf{r}_{tc})\rangle}} \tag{27}$$

and

$$\Delta B_r(\mathbf{r}_S, t) \equiv \frac{|B_r(\mathbf{r}_S, t) - \widetilde{B}_r(\mathbf{r}_S, t)|}{\sqrt{\langle B_r^2(\mathbf{r}_S)\rangle}}, \tag{28}$$

respectively. In these two definitions, notice that the distance is normalized since the brackets $\langle\cdot\rangle$ represent time averaging (which we perform over a period of about 1000 yr after $t = t_p$). In the following, we will use Δ as a shorthand for ΔB_φ or ΔB_r, when the distinction need not be made, and we will refer to the evolution of Δ as the *error growth*: in a forecasting perspective, the perturbation which we insert can indeed be interpreted as the uncertainty affecting the initial condition (or the control parameters) of the model. In this sense the distance we measure is analogous to the growth of the forecast error of interest for the data assimilation practitioner.

Figure 4 shows the typical evolution of the error growth (measured here in terms of ΔB_φ) in our numerical experiments. It corresponds to a $\varepsilon = 10^{-6}$ perturbation applied to the spectral poloidal coefficient $\mathcal{P}_1$. The evolution of Δ comprises three distinct phases. First, both trajectories remain fraternal, as their distance remain similar to ε (phase I in Figure 4). This is called the mobilization phase by Lhuillier et al. (2011). Next, the error enters a phase of exponential growth (phase II), until it reaches saturation (phase III). From then on, the reference and perturbed solutions evolve in an uncorrelated way.

Among these three phases, the phase of exponential growth is the most meaningful to constrain the limit of predictability. Considering that this phase starts at $t = t_\varepsilon$ with an initial value ε, the distance evolves according to

$$\Delta(t) = \varepsilon\, e^{\lambda(t-t_\varepsilon)}, \tag{29}$$

where λ denotes the exponential growth rate. Its inverse λ^{-1} is the so-called e-folding time τ_e, namely the divergence time τ we discussed above. We set out to estimate λ (or τ_e) as accurately as possible for the class of mean-field models considered in this work. Visual inspection of the time series of Δ allows us to pick the phase of exponential growth; we next perform a least-squares analysis to estimate λ (this procedure yields the purple line in Figure 4).

That estimate may depend on the type and amplitude of the perturbation, though, which calls for a systematic approach to evaluating λ. In the next subsection, we use the standard model S presented in Section 3 to vary extensively the type and amplitudes of perturbations. Within this single-model context, we find that the characteristics of the error growth are robust. Therefore, in order to push the analysis further, we shall consider in Section 4.3 how λ may be influenced by the values of the triplet (Rm, C_{BL}, C_α).

4.2. Error Growth in the Standard Model

4.2.1. Magnetic Perturbations

As explained above, we focus here on the standard model S and begin by examining its response to perturbations of the magnetic field. We study different scenarios. The perturbation can affect either the poloidal scalar $\mathcal{P}$ or the toroidal scalar $\mathcal{T}$. It can be either large-scale (restricted to the $n = 1$ harmonic degree), in which case it writes

$$\mathcal{P}_1(r, t_p) \longmapsto \widetilde{\mathcal{P}}_1(r, t_p) = \mathcal{P}_1(r, t_p)(1+\varepsilon), \tag{30}$$

(and the same for $\mathcal{T}_1$), or distributed randomly over the entire spectrum, according to

$$\mathcal{P}_n \longmapsto \widetilde{\mathcal{P}_n}(r, t_p) = \mathcal{P}_n(r, t_p)(1+\gamma_n\varepsilon),\ 1 \leqslant n \leqslant N, \tag{31}$$

(and the same for $\mathcal{T}_n$), in which the γ_n are random numbers from 0 to 1 distributed over all the harmonic degrees. In the remainder of this subsection, the amplitude of the perturbation ε is set to 10^{-6}.

The pink curves in Figure 5 show that large-scale perturbations of the poloidal or toroidal scalars defined by Equation (30) yield the same well-defined three phases for the evolution of ΔB_r and ΔB_φ. In addition, each panel of Figure 5 comprises five gray curves obtained from five random realizations of the small-scale perturbations defined by Equation (31). Despite some scatter, visual inspection indicates a common error growth behavior. In particular, if we were to estimate λ from this catalog of curves, we would probably get a robust value. This is rather encouraging, but before proceeding with the actual calculation of λ, let us now inspect in more detail its sensitivity to a broader range of perturbations.

4.2.2. Systematic Perturbations

We thus investigate now the error growth induced by perturbations of different origins, varying amplitudes ε and different times of insertion t_p on the standard model S. The origin of the perturbation ξ in Equation (26) can be one of the following: $\mathcal{P}_1$, $\mathcal{T}_1$, $\mathcal{P}_n$, $\mathcal{T}_n$ (as in the previous section), Ω or u_p (the flow), α or S_{BL} (the poloidal source terms), or η_p (the poloidal diffusivity

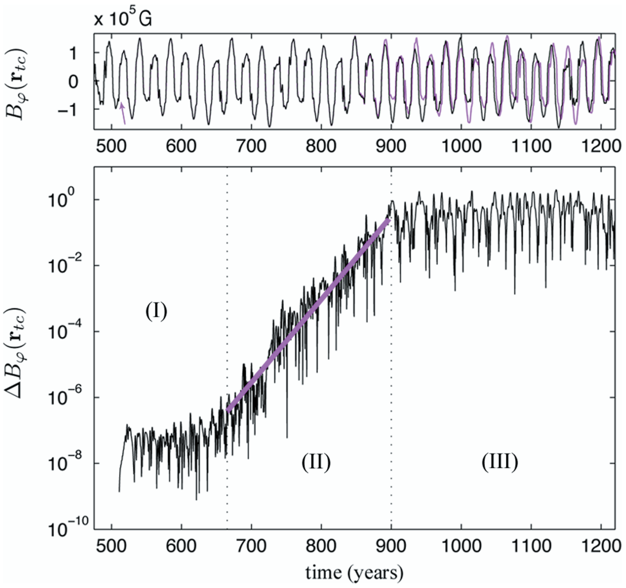


Figure 4. Error growth behavior for the reference model S (see Table 3). Top: toroidal field at 20^0 latitude on the tachocline. The reference solution (shown in black) is perturbed at a given instant, shown by the pink arrow, by a relative amount $\varepsilon = 10^{-6}$. This generates the perturbed solution (purple) which progressively diverges from the reference one. Bottom: the difference ΔB_φ between both solutions on a logarithmic scale. The error growth can be separated in three well-defined stages: (I) a mobilization phase, (II) an exponential growth phase and (III) a saturated phase. In phase (II), we perform a least-squares regression (purple line) in order to estimate the error growth rate λ. See the text for details.

profile). For each of these nine possibilities, we consider perturbations of amplitudes 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8} or 10^{-10}. Finally, we perturb the reference dynamical trajectory at three different times, $t_p = t_1, t_2$ or t_3. We therefore consider 9 (origins) × 5 (amplitudes) × 3 (instants) = 135 ways of perturbing the standard trajectory. Since both ΔB_φ and ΔB_r are used to monitor the error growth, this allows us to construct a database of 270 estimates of λ. The database is completely described in Table 2.

Figure 6 illustrates the variability within the database of model S, and shows that regardless of this variability, the error growth displays a fair amount of dynamical similarity in the 270 scenarios we envisioned. Figure 6(a) shows the evolution of ΔB_φ, for different origins, times of perturbation insertion and different perturbation amplitudes. We see that the error growth is weakly sensitive to the origin of the perturbations. Still, the mobilization phase seems to vary depending on the way the perturbations were inserted. For perturbations corresponding to $\xi = \alpha$ or $\xi = u_p$, the mobilization phase lasts longer (several centuries), and there is a mild dependency of the duration of that phase on t_p. The mobilization phase has a duration which decreases with ε as well. However, this variability on the mobilization phase does not strongly affect the estimate of λ. On another note, it can also be seen that the error growth due to smaller perturbations can experience secondary mobilization phases, and resume its exponential growth after some time.

Figure 6(b) presents the distribution of the error growth rates (one histogram per value of ε, which integrates all other dependencies) of model S. The exponential growth is steeper for smaller levels of perturbations (noticing that $\varepsilon = 10^{-8}$ and $\varepsilon = 10^{-10}$ yield essentially the same behavior, though), that is, large perturbations lead to smaller values of λ. There is also a general tendency for the growth to slacken as the error reaches macroscopic values.

Still, focusing on small to extremely small values of ε (10^{-6} and less), the estimated λ does not vary by more than 20%. This robustness suggests the fact that λ is an intrinsic property of our standard model S: regardless of the perturbation time and origin, and as long as it is small, the exponential growth of the

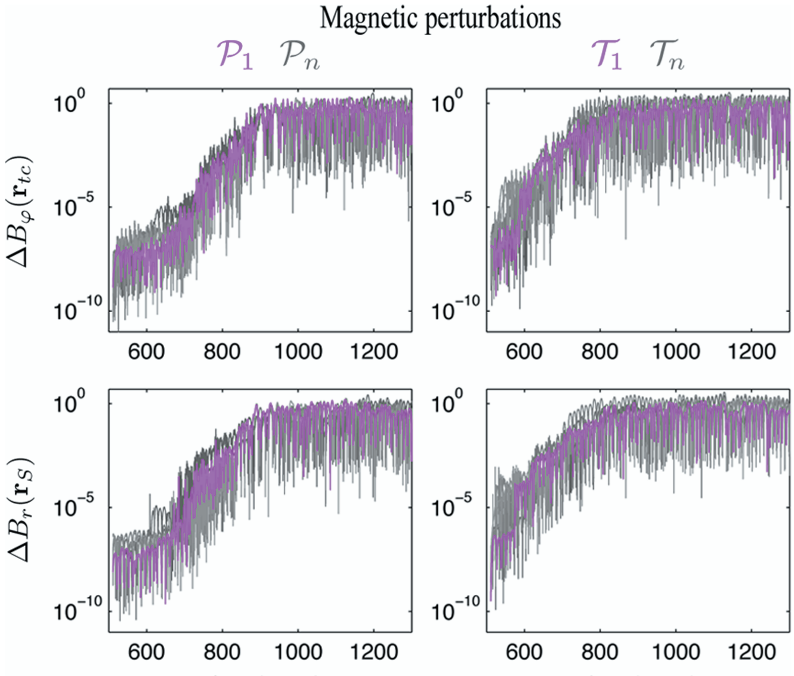


Figure 5. Time series of ΔB_φ and ΔB_r, following the application of magnetic perturbations of relative amplitude $\varepsilon = 10^{-6}$ on standard model S. The perturbations are inserted either on the poloidal (left column) or toroidal (right column) component of the magnetic field, and they affect either the first harmonic degree ($\mathcal{P}_1/\mathcal{T}_1$ pink curves) or all the harmonic degrees randomly ($\mathcal{P}_n/\mathcal{T}_n$), in which case five curves with different shades of gray are shown in each graph.

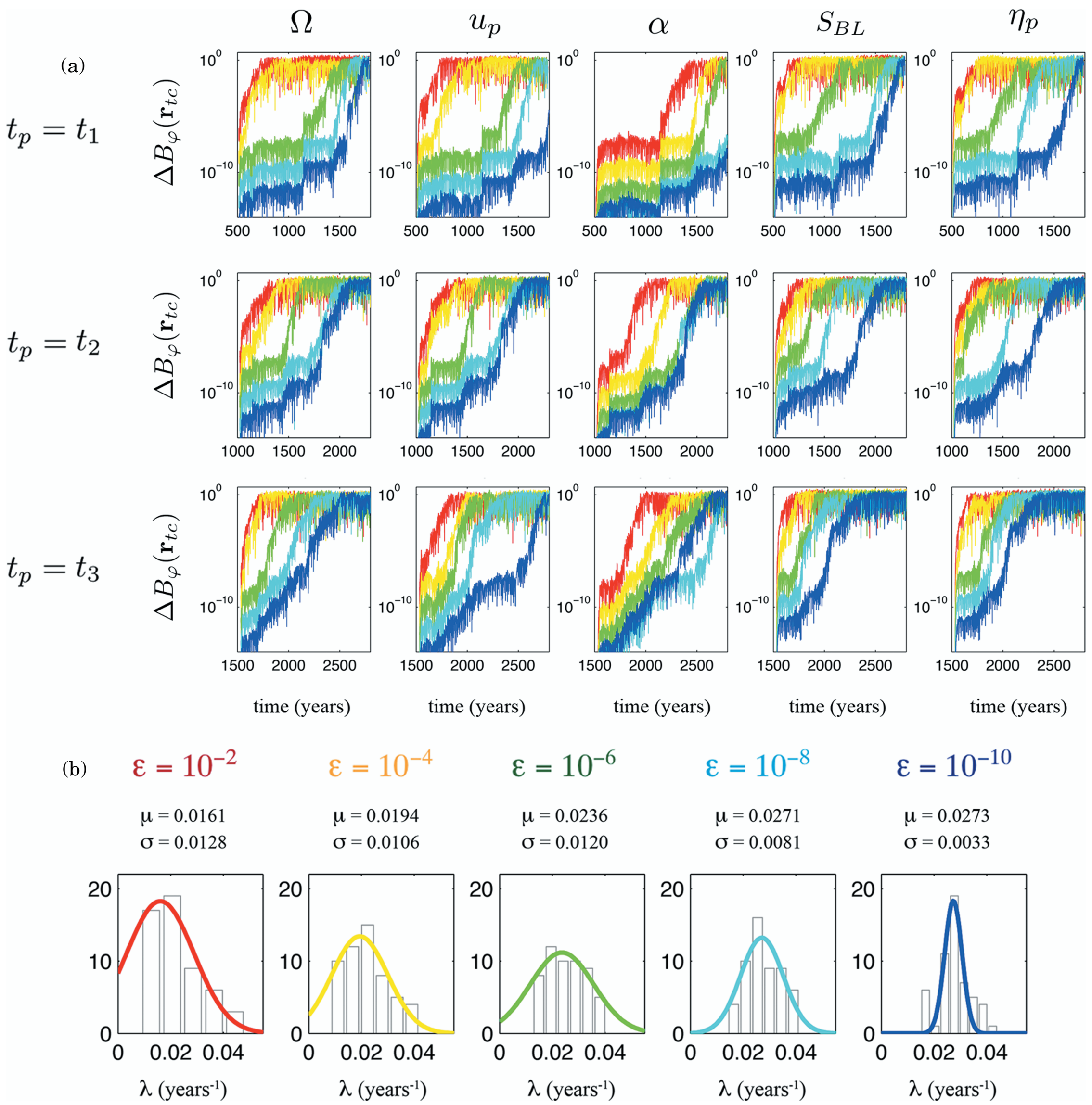


Figure 6. Sensitivity of the error growth rate against perturbation types for model S: (a) Error growth considering different origins ($\xi = \Omega, u_p, \alpha, S_{\mathrm{BL}}$, and η_p), times ($t_p = t_1, t_2$, and t_3) and amplitudes ($\varepsilon = 10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}$, and 10^{-10}); (b) histograms of the exponential growth rates λ from the set of perturbations displayed in Table 2 ordered by perturbation amplitude. The histograms are modeled by Gaussian curves with mean μ and standard deviation σ.

error is likely to occur on a timescale τ of roughly 40 yr, that is over slightly more than three simulated cycles. More precisely, if T_c denotes the period of the simulated cycle and considering a least-squares analysis of the $\varepsilon = 10^{-10}$ histogram, we find that $\tau_e = (3.34 \pm 0.40)\, T_c$.

4.3. Sensitivity of λ to the Control Parameters

We now explore the more general dependency of τ_e to the control parameters of our class of mean-field models. Since the simulated T_c varies with these parameters as well, and since we wish to express τ_e in units of T_c, we investigate the joint dependency of these two quantities on the triplet (Rm, C_α, C_{BL}).

First, we increase the α-effect coefficient from $C_\alpha = 8$ to $C_\alpha = 16$ and consider the same 270 possibilities as the ones used for the standard model (this new model is labeled T in the following). Figure 7 illustrates the corresponding database of model T, and highlights consistent differences when compared with the standard case S shown in Figure 6. Most notably, the mobilization phase is in every instance much shorter (not lasting more than a few decades), while the exponential growth phase is in all cases much steeper, two effects pointing toward an increased influence of turbulence as the value of C_α increases, leading to larger estimates for λ. We still retrieve the tendency for λ to decrease with increasing ε, while its uncertainties decrease with ε. Accordingly, we find that for $\varepsilon = 10^{-10}$, $\tau_e = (2.45 \pm 0.42)\, T_c$ (here, $T_c = 10.15$ yr).

Next, we carry out a similar sensitivity analysis with different triplets (Rm, C_α, C_{BL}). More specifically, we consider the following possibilities

1. $308 \leqslant \mathrm{Rm} \leqslant 378$,
2. $0.5 \leqslant C_{\mathrm{BL}} \leqslant 2$,
3. $8 \leqslant C_\alpha \leqslant 32$,

providing a total of 48 different models (including models S and T). For each model we calculate λ restraining the amplitude of the perturbation ε to what we consider its most reliable level,

Table 2
Mean Values of λ and Its Uncertainties σ (in Units of 10^{-2} yr^{-1}) from the Systematic Perturbation Analysis of the Standard Model S

		ΔB_φ										ΔB_r									
	ε	10^{-10}		10^{-8}		10^{-6}		10^{-4}		10^{-2}		10^{-10}		10^{-8}		10^{-6}		10^{-4}		10^{-2}	
t_p	ξ	λ	σ	λ	σ	λ	σ	λ	σ	λ	σ	λ	σ	λ	σ	λ	σ	λ	σ	λ	σ
t_1	$\mathcal{P}_1$	3.66	0.07	2.49	0.05	2.52	0.06	1.68	0.09	3.57	0.21	3.69	0.09	2.33	0.06	2.58	0.06	1.44	0.09	4.23	0.35
	$\mathcal{T}_1$	3.62	0.07	3.20	0.13	1.52	0.05	1.33	0.07	3.82	0.35	3.70	0.09	3.51	0.11	1.51	0.05	1.60	0.09	1.19	0.22
	$\mathcal{P}_n$	3.48	0.06	1.49	0.06	1.87	0.08	1.21	0.08	2.76	0.27	3.69	0.08	1.51	0.05	2.05	0.06	1.33	0.07	1.21	0.21
	$\mathcal{T}_n$	2.89	0.06	2.12	0.08	1.86	0.11	2.43	0.21	1.61	0.15	3.04	0.06	2.11	0.07	1.89	0.10	2.29	0.14	1.52	0.13
	Ω	3.16	0.05	3.82	0.09	2.30	0.06	1.40	0.05	2.11	0.09	3.98	0.10	3.52	0.11	2.90	0.08	2.22	0.06	2.20	0.14
	u_p	1.74	0.03	3.52	0.04	2.90	0.08	3.98	0.06	2.20	0.14	2.57	0.05	3.73	0.07	2.77	0.07	2.43	0.06	2.15	0.13
	α	3.12	0.05	2.32	0.04	3.03	0.05	3.98	0.12	2.01	0.05	3.18	0.05	2.28	0.03	3.39	0.05	4.19	0.15	1.62	0.06
	S_{BL}	2.73	0.05	3.39	0.08	1.92	0.04	1.91	0.13	4.83	0.25	2.91	0.05	3.58	0.09	1.84	0.05	1.67	0.09	2.18	0.14
	η_p	2.66	0.05	2.53	0.05	1.83	0.04	2.22	0.11	1.56	0.28	2.90	0.04	2.31	0.06	1.85	0.05	1.12	0.08	1.12	0.12
t_2	$\mathcal{P}_1$	2.48	0.04	2.57	0.04	1.37	0.06	1.46	0.04	1.37	0.07	2.50	0.04	2.77	0.05	1.26	0.05	1.15	0.05	1.61	0.14
	$\mathcal{T}_1$	2.62	0.04	3.69	0.06	3.15	0.10	1.36	0.04	1.39	0.09	2.76	0.04	3.97	0.07	3.91	0.12	1.28	0.07	0.87	0.16
	$\mathcal{P}_n$	3.06	0.06	3.92	0.07	1.97	0.08	1.58	0.06	1.81	0.15	2.91	0.10	4.12	0.10	1.84	0.06	1.67	0.05	1.31	0.09
	$\mathcal{T}_n$	3.44	0.06	2.87	0.08	2.07	0.08	1.53	0.08	2.10	0.23	3.61	0.08	2.76	0.08	2.48	0.11	1.37	0.06	1.79	0.22
	Ω	2.53	0.04	2.71	0.05	3.46	0.07	2.13	0.05	1.43	0.05	2.50	0.04	2.79	0.05	3.93	0.10	2.07	0.05	1.40	0.04
	u_p	2.68	0.04	2.80	0.05	3.47	0.05	2.16	0.11	1.96	0.09	2.67	0.04	2.54	0.04	3.60	0.06	1.32	0.05	2.16	0.10
	α	3.11	0.05	2.81	0.05	4.03	0.04	2.65	0.05	3.35	0.09	3.19	0.05	2.83	0.06	2.31	0.04	3.07	0.05	2.57	0.11
	S_{BL}	2.65	0.05	3.65	0.06	4.03	0.12	2.34	0.08	1.87	0.17	2.49	0.05	1.68	0.08	3.53	0.12	1.42	0.07	1.30	0.08
	η_p	2.47	0.04	3.15	0.04	1.98	0.12	2.22	0.08	1.78	0.08	2.57	0.06	1.42	0.05	1.38	0.04	1.42	0.09	1.98	0.14
t_3	$\mathcal{P}_1$	2.37	0.04	2.32	0.03	2.85	0.05	2.05	0.12	2.53	0.21	2.75	0.05	2.26	0.03	2.76	0.05	2.94	0.20	3.06	0.35
	$\mathcal{T}_1$	2.59	0.05	2.57	0.07	3.19	0.07	2.45	0.16	1.37	0.32	2.69	0.06	2.07	0.06	2.61	0.08	3.65	0.25	1.84	0.22
	$\mathcal{P}_n$	4.40	0.12	2.48	0.05	2.73	0.14	0.82	0.33	2.00	0.29	2.73	0.06	2.47	0.05	2.45	0.12	2.52	0.20	2.06	0.25
	$\mathcal{T}_n$	1.64	0.02	2.81	0.06	2.81	0.10	3.07	0.19	2.72	0.32	1.96	0.04	2.59	0.05	2.48	0.10	3.16	0.24	3.65	0.34
	Ω	1.50	0.01	1.87	0.03	3.07	0.08	2.74	0.18	1.98	0.14	1.54	0.01	1.91	0.03	3.29	0.07	2.65	0.13	2.82	0.22
	u_p	2.77	0.06	2.77	0.06	3.53	0.08	3.17	0.06	3.11	0.05	2.70	0.06	2.61	0.06	3.43	0.08	2.96	0.06	2.02	0.05
	α	1.76	0.03	3.56	0.09	1.33	0.04	2.10	0.04	2.95	0.05	1.67	0.03	2.66	0.06	1.60	0.02	2.15	0.04	2.57	0.06
	S_{BL}	2.42	0.08	2.44	0.07	2.80	0.05	2.84	0.18	4.31	0.35	2.64	0.06	2.03	0.06	2.60	0.07	2.11	0.09	1.51	0.20
	η_p	2.74	0.06	3.36	0.06	2.21	0.08	2.75	0.17	3.59	0.21	2.39	0.03	3.26	0.07	1.55	0.05	3.24	0.19	3.46	0.34

Table 3
Summary of the Values of Solar Cycle Periodicity T_c, e-folding Time τ_e and Its Uncertainty δ, and the Ratio τ_e/T_c for a Large Number of Configurations of the Triplet (Rm, C_α, C_{BL})

	C_{BL}	C_α	Rm	T_c	τ_e	δ	τ_e/T_c	C_{BL}	C_α	Rm	T_c	τ_e	δ	τ_e/T_c
	0.50	16	308	9.02	24.48	3.50	2.71	1.00	16	368	9.40	42.15	5.50	4.48
	0.50	16	318	12.78	23.25	5.94	1.81	1.00	16	378	9.28	46.00	17.59	4.95
	0.50	16	328	12.52	23.33	6.79	1.86	1.00	20	318	9.73	23.74	4.77	2.43
	0.50	16	338	12.13	25.40	5.61	2.09	1.00	24	308	9.48	20.98	5.80	2.21
	0.50	16	348	11.61	26.37	5.74	2.29	1.00	24	318	9.47	19.00	2.72	2.00
	0.50	16	358	11.49	29.06	6.61	2.52	1.00	24	328	9.26	19.15	3.67	2.06
	0.50	16	368	11.26	26.60	3.29	2.36	1.00	24	338	9.94	20.85	4.02	2.09
	0.50	16	378	11.04	42.66	2.52	3.86	1.00	24	348	8.96	21.75	1.67	2.42
	0.75	16	318	10.64	28.05	4.62	2.63	1.00	24	358	8.84	19.74	2.81	2.23
	1.00	8	308	11.13	32.21	5.67	2.89	1.00	24	368	8.61	25.07	4.39	2.91
S	**1.00**	**8**	**318**	**10.95**	**36.64**	**4.41**	**3.34**	1.00	24	378	8.52	21.06	2.48	2.47
	1.00	8	328	10.77	62.13	14.35	5.76	1.00	28	318	9.32	18.54	2.00	1.99
	1.00	8	338	10.50	52.30	9.98	4.98	1.00	32	318	9.28	16.75	2.46	1.80
	1.00	8	348	10.30	47.97	7.21	4.65	1.25	16	318	10.19	33.69	10.99	3.30
	1.00	8	358	10.17	53.44	19.78	5.25	1.50	16	308	10.13	25.21	2.01	2.48
	1.00	8	368	10.08	63.59	3.03	6.30	1.50	16	318	10.30	24.12	2.57	2.34
	1.00	8	378	10.07	54.52	3.27	5.41	1.50	16	328	10.14	25.50	3.35	2.51
	1.00	12	318	10.53	32.81	3.36	3.11	1.50	16	338	9.45	22.46	7.07	2.37
	1.00	16	308	10.18	27.35	2.36	2.68	1.50	16	348	9.56	25.77	4.00	2.69
T	**1.00**	**16**	**318**	**10.15**	**24.94**	**4.33**	**2.45**	1.50	16	358	9.44	30.21	3.04	3.20
	1.00	16	328	9.99	27.95	4.33	2.79	1.50	16	368	9.23	26.78	4.89	2.90
	1.00	16	338	9.85	25.44	2.55	2.58	1.50	16	378	9.19	26.05	2.74	2.83
	1.00	16	348	9.65	31.90	4.66	3.30	1.75	16	318	10.37	24.91	2.26	2.40
	1.00	16	358	9.58	31.63	5.21	3.30	2.00	16	318	10.89	26.21	4.37	2.40

Notes. The letters S and T make reference to the main two models discussed in the bulk of the paper. All the timescales are expressed in years. Boldface highlights the values of the timescales for the two specific cases (S and T) discussed throughout the text.

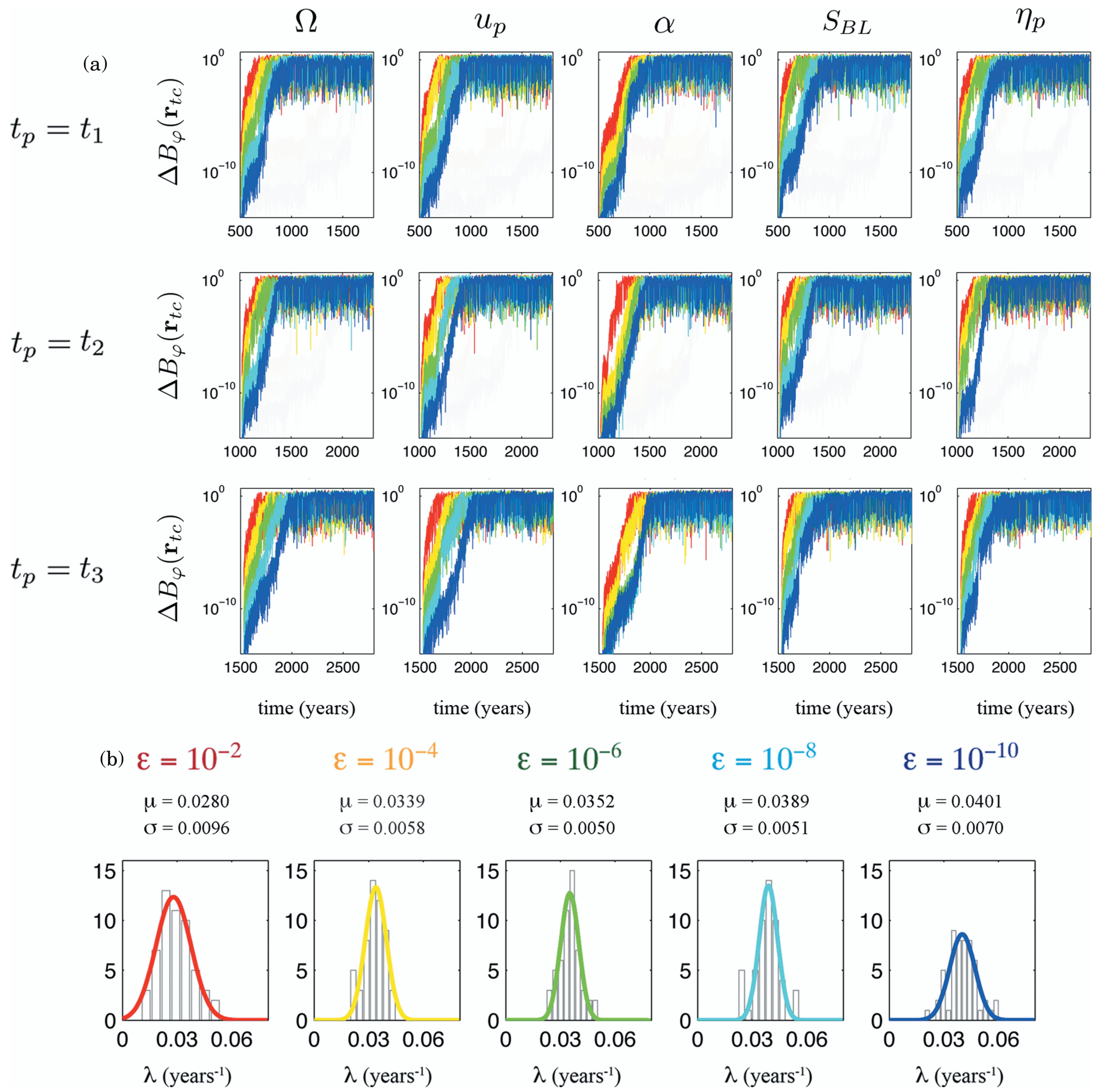


Figure 7. Same as Figure 6, for model T which has a stronger α-effect than model S ($C_\alpha = 16$ instead of $C_\alpha = 8$), all other control parameters being the same.

namely 10^{-10}. This survey is summarized in Table 3, and the results (expressed in terms of the corresponding timescales T_c and τ_e) are shown in Figures 8 and 9.

When Rm is fixed to its standard value (model S, Rm = 318), our results show that τ_e and T_c are mostly sensitive to C_α, and less affected by variations in C_{BL}. Regarding the former dependency, it can be seen in Figure 8(a) that both τ_e and T_c decrease with C_α. The decrease in τ_e is a consequence of the stronger destabilizing effect of turbulence. We also note (Figure 8(a), right) that the ratio τ_e/T_c decreases with increasing C_α. In the parameter region which we explored, τ_e is thus more sensitive to variations in C_α than T_c.

According to Figure 8(b), the cycle period T_c displays a non-monotonic behavior with respect to changes in C_{BL}, which measures the intensity of the non-local coupling in the governing Equation (12). It is worth mentioning here that for the lower value of C_{BL}, the system undergoes a transition to an α-dominated dynamo, characterized by a longer (and less solar-like) periodicity of about 13 yr. As indicated by Figure 8(b), the e-folding time τ_e does not vary substantially with C_{BL} over our narrow interval of investigation (recall Section 3). Overall, we find that the ratio τ_e/T_c remains approximately constant (equal to 2.5) over this interval.

Turning our attention to the dependency of T_c and τ_e on Rm, we see (Figure 9) that the former decreases with increasing Rm. The cycle duration scales indeed approximately in inverse proportion to Rm, as shown in the left panel of Figure 9. On the other hand, the dependency of τ_e on Rm is less clear. There seems to be a mild trend in the cases of low to intermediate values of C_α (orange and red points in the middle plot of Figure 9), with τ_e slightly increasing with increasing Rm. This behavior can be interpreted as a regulatory effect of the meridional circulation: as u_0 gets larger the meridional circulation tends to make the system more stable against perturbations. This is no longer true for a large C_α (dark red points in the middle plot of Figure 9), which indicates that τ_e is then controlled by the α-effect. It is worth mentioning that some realizations of τ_e are affected by large uncertainties, mostly in cases with low values of C_α and C_{BL}, and large values of Rm. These cases are the less chaotic ones, and the introduction of a perturbation can sometimes lead to a mobilization phase lasting for more than 1000 yr.

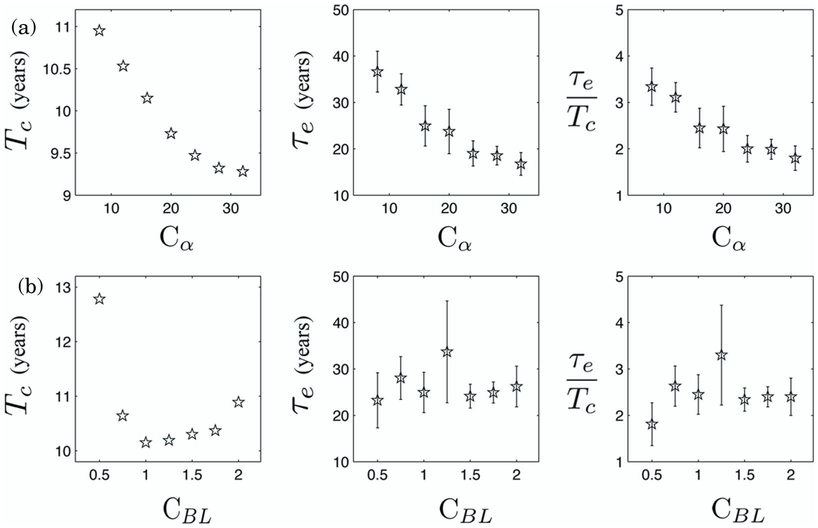


Figure 8. Solar cycle periodicity T_c, e-folding time τ_e, and their ratio for (a) (top row) Rm = 318, $C_{\mathrm{BL}} = 1.0$ and a varying C_α and (b) (bottom row) Rm = 318, $C_\alpha = 16$, and a varying $\mathrm{C_{BL}}$.

To conclude this analysis, let us stress (as shown in Figure 9, right) that the τ_e/T_c ratio is mainly concentrated around two values, 2.5 and 5, with a larger concentration of points around the former. Using all the available data at our disposal (as summarized in Table 3), we can finally calculate a weighted average for the ratio τ_e/T_c, and find

$$\frac{\tau_e}{T_c} = 2.76 \pm 0.05. \tag{32}$$

5. SUMMARY AND DISCUSSION

Our extensive analysis of the e-folding time τ_e for our preferred (in the sense of solar semblance) standard model S led us to conclude that if the control parameters (Rm, C_α, C_{BL}) are fixed, then τ_e can be regarded as an intrinsic property of the model, regardless of the source of the error, with a small dependence on its initial magnitude (Section 4.2.2).

In view of using that standard model (or a close version) for operational forecasting, we extended the analysis to a series of models, and investigated the sensitivity of τ_e to (Rm, C_α, C_{BL}) in detail. Our results reveal three salient properties.

1. A decrease of τ_e with increasing C_α. This reflects the influence of the non-linear nature of the quenched α-effect on the amplification of errors, leading to a more chaotic (and less predictable) dynamo.
2. An apparent independence of τ_e on C_{BL}, indicating the secondary role played by this non-local forcing term on the error growth. However, let us stress that this may be caused by the narrow range of possible C_{BL} we explored, a consequence of the extreme sensitivity of the solar semblance of the flux-transport model to this parameter.
3. A slight tendency for τ_e to increase with Rm for those models with low to intermediate strength of the α-effect, pointing to a stabilizing role of the meridional circulation on the system under these conditions.

In addition, the moderate variability of the ratio of τ_e to the simulated cycle period T_c in our database of simulations (which comprises approximately 50 members) prompts us to propose the master value $\overline{\tau}_e = 2.76\, T_c$ for the class of mean-field models we considered, should they be used for operational forecasting (and keeping in mind that we focused our analysis on the regular working of those models, not considering extreme events such as grand minima).

From a practical point of view, the perturbations artificially inserted into the model in Section 4 can be interpreted as uncertainties in the measurements or in the model itself, which are the causes of errors any data assimilation scheme needs to deal with. These uncertainties are ultimately responsible for the limited horizon of predictability of the chaotic system we are interested in. If ε denotes the relative level of these uncertainties, we derive from Equation (29) that the forecast horizon τ_f is given by

$$\tau_f = -\overline{\tau}_e \ln \varepsilon. \tag{33}$$

Let us begin by estimating the level of uncertainties on the measurement side. It is likely that an operational data assimilation scheme will assimilate observations connected with large-scale maps of B_r at the solar surface, B_r^S. Such magnetograms are contaminated by errors, due to limited resolution, asynchronous sampling and sparse polar measurements. A way to quantify those errors is to analyze the spherical harmonic decomposition of B_r^S. Theory demands the monopole term in this expansion (g_0^0) to be zero; a non-zero g_0^0 can consequently be used as a means to quantify the uncertainty $\varepsilon(B_r^S)$ we are after. Figure 10 shows the time series of the monopole and axial dipole coefficient (g_1^0) derived from the database of magnetograms of the Wilcox Solar Observatory (WSO).[1] The figure shows that g_1^0 evolves in phase with the global poloidal magnetic field—it changes sign at the time of maximum activity, and is correlated with the polar flux (DeRosa et al. 2012). The monopole coefficient g_0^0 constantly oscillates around zero. We can therefore use the ratio of the root-mean-squared (rms) value of g_0^0, $\langle g_0^0 \rangle$, to the rms value of g_1^0, $\langle g_1^0 \rangle$, to estimate $\varepsilon(B_r^S)$. This yields

$$\varepsilon\left(B_r^S\right) \approx \frac{\langle g_0^0 \rangle}{\langle g_1^0 \rangle} = \frac{0.1535\ \mathrm{G}}{1.2550\ \mathrm{G}} \approx 12\%. \tag{34}$$

On the model side now, one of the most obvious sources of errors lies in the large-scale kinematic approximation on which

---

[1] http://wso.standford.edu/Harmonic.rad/ghlist.html

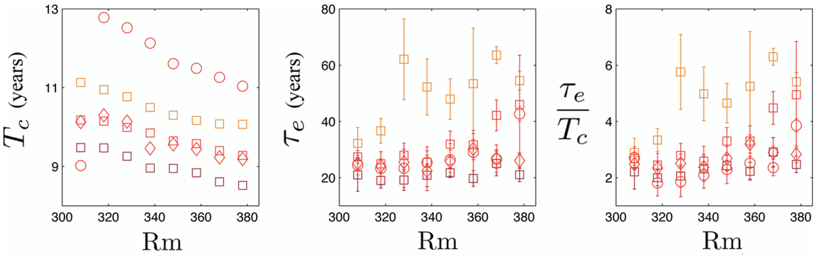


Figure 9. Solar cycle periodicity T_c, e-folding time τ_e and their ratio for different values of C_α, C_{BL} and Rm. The magnitude of C_α is color-coded: $C_\alpha = 8$, orange; $C_\alpha = 16$, red; $C_\alpha = 24$, dark red. Symbols indicate different C_{BL}: $C_{\mathrm{BL}} = 0.5$, circles; $C_{\mathrm{BL}} = 1.0$, squares; $C_{\mathrm{BL}} = 1.5$, diamonds.

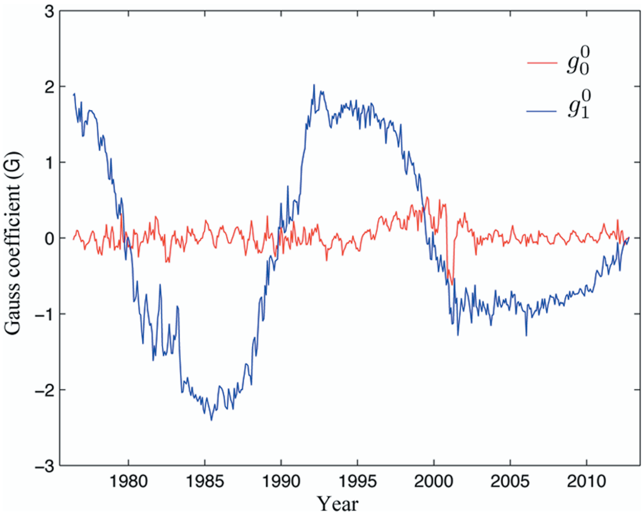


Figure 10. Time series of the monopole g_0^0 and axial dipole g_1^0 Gauss coefficients derived from the magnetic charts of the Wilcox Solar Observatory (http://wso.standford.edu/Harmonic.rad/ghlist.html).

our modeling rests. In particular, observations indicate that it may be inappropriate to assume that the large-scale flow driving the dynamo is steady. As a consequence, there are errors arising from the variability of both the patterns of differential rotation and meridional circulation. Howe et al. (2000) discovered a persistent pattern of low-amplitude time variation of Ω, $\delta\Omega$, of about 6 nHz, due to solar torsional oscillations. Consequently, we get

$$\varepsilon(\Omega) = \frac{\delta\Omega}{\Omega_{\mathrm{eq}}} \approx 1\%. \tag{35}$$

This small figure must be contrasted with the one owing to those uncertainties impacting $\mathbf{u}_p$. The long-term variability of the meridional circulation $\delta\mathbf{u}_p$ has an amplitude δu_0 close to 5 m s^{-1} (Hernández et al. 2006), which yields

$$\varepsilon(\mathbf{u}_p) = \frac{\delta u_0}{u_0} \approx 33\%, \tag{36}$$

if computed based on the mean value of the surface meridional flow at mid-latitudes, $u_0 \approx 15$ m s^{-1}. Injecting $\varepsilon(\mathbf{u}_p)$ in Equation (33) yields

$$\tau_f \approx 3\, T_c. \tag{37}$$

In addition to these fluctuations in amplitude, there exists considerable uncertainties on the large-scale structure of the meridional circulation itself. The depth at which the equatorward return flow occurs (Hathaway 2011) and the possible multi-cell pattern of meridional flow (Zhao et al. 2013) are two examples illustrating the current lack of robust observational constraints on $\mathbf{u}_p$. These cannot be readily incorporated in the current analysis, for they would require different families of simulations to be integrated, and their region of solar semblance be identified in parameter space (in the same way we mapped it for the ensemble of single-cell, tachocline equatorward return flow simulations considered here).

Further uncertainties affect the turbulent diffusivity, $\eta(\mathbf{r})$. As explained in Section 2, we resort in this study to spherically symmetric $\eta_p(r)$ and $\eta_t(r)$, of relatively low values. Both reach an amplitude of 3×10^{11} cm s^{-2} at $r = R_\odot$ (recall Figure 2(d)). This value has to be contrasted with the value of 10^{12} cm s^{-2} used by Choudhuri et al. (2007) for their prediction of cycle 24, using a so-called diffusion-dominated flux-transport model. More recently, Miesch et al. (2012) put forward theoretical and observational arguments in favor of the same figure, 10^{12} cm s^{-2}, as a lower bound of this turbulent transport coefficient. The exact nature of $\eta(\mathbf{r})$ in the solar interior remains strongly debated, and no consensus has been reached.

In this study, we opted for an advection-dominated forward model, on the account of its first order dynamical semblance with the solar dynamo. If one were to choose instead a diffusion-dominated model for data assimilation purposes, one would have to carry out a sensitivity analysis similar to the one pursued here, in order to compute the e-folding time and estimate the forecast horizon for that different family of models. In this respect, note that Karak & Nandy (2012) resorted to a simpler, correlation-based, analysis in order to study the "memory" of such a diffusion-dominated model (which includes in their case turbulent pumping and a stochastic component to the poloidal source term). Their analysis demonstrates that turbulent diffusion shortens the memory of the system to less than one cycle. This preliminary work should be complemented by the proper derivation of the e-folding time characterizing that class of models, along the methodological lines presented in this paper.

Regardless of the model ultimately chosen, one should keep in mind that data assimilation remains in any case the only sensible way of testing the compatibility of a given physical model of the solar dynamo with observations of its dynamical activity. By enabling on-the-fly parameter adjustments (in addition to state estimation), data assimilation offers in principle the possibility of correcting the radial profiles of diffusion coefficients (and those of poloidal source terms). Even if the advection-dominated model we studied has an optimistic theoretical predictability limit of three solar cycles, we must bear in mind that any data assimilation scheme aiming to forecast solar activity will be unperfect, and its effective forecast

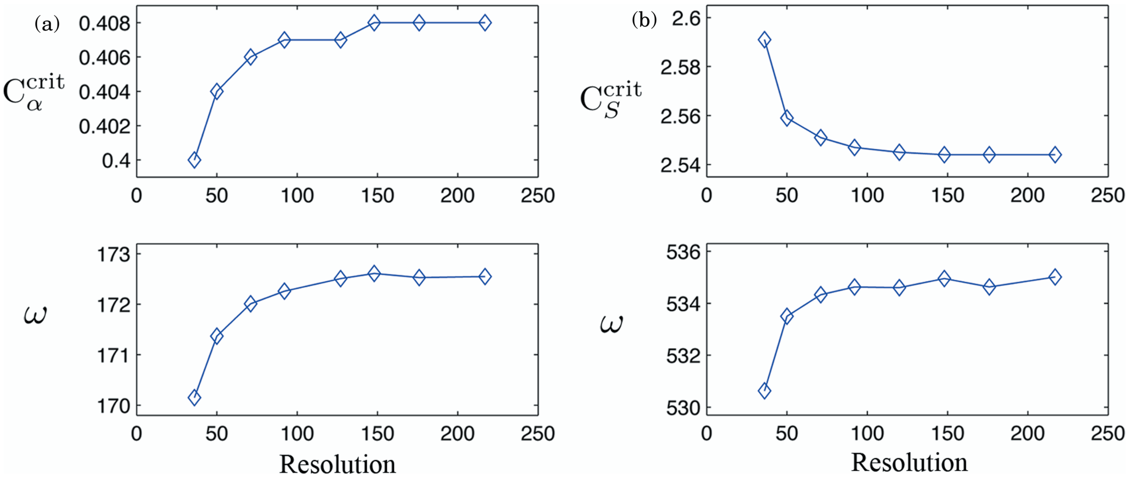


Figure 11. Convergence tests: critical dynamo numbers C_α^{crit} and C_S^{crit} and solar cycle periodicity ω for (a) case B and (b) case C, as defined by Jouve et al. (2008). The resolution is defined by $\sqrt{N_r\, N}$, where N_r is the number of radial levels and N is the truncation of the spherical harmonic expansion.
(A color version of this figure is available in the online journal.)

Table 4
Comparison of the Critical Dynamo Numbers $C_{\alpha,S}^{\text{crit}}$ and Frequency of the Solar Cycle ω in the Benchmark Cases A, B and C from Jouve et al. (2008)

Case	Results				Reference	
	Resolution	Δt	$C_{\alpha,s}^{\text{crit}}$	ω	$C_{\alpha,s}^{\text{crit}}$	ω
A	71 × 71	5×10^{-5}	0.385	158.00	0.387 ± 0.002	158.1 ± 1.472
B	71 × 71	5×10^{-5}	0.406	172.01	0.408 ± 0.003	172.0 ± 0.632
C	120 × 120	1×10^{-6}	2.545	534.6	2.489 ± 0.075	536.6 ± 8.295

Notes. The spacial and temporal resolutions are given in terms of radial points and harmonic degree ($N_r \times N$) and time-step Δt.

horizon will consequently decrease. Taking this into account, one can hope, though, that if such an advection-dominated model were to be chosen for operational forecasting, its practical limit of predictability could reach (and perhaps exceed) one solar cycle.

The authors thank the referee for helpful and constructive review, and Allan Sacha Brun, Emmanuel Dormy and Martin Schrinner for enlightening discussions. Sabrina Sanchez also thanks Oscar Matsuura and Katia Pinheiro for the fruitful contributions to the beginning of this project, the Observatório Nacional of Brazil for the initial support, and the Space Physics and Aeronomy group of the American Geophysical Union for the student grant award at the 2012 AGU Meeting. Numerical calculations were performed on IPGP's S-CAPAD computing facility. This is IPGP contribution 3459.

APPENDIX

PARODY CODE—MEAN FIELD BENCHMARKING

The Parody code used in this work was originally proposed for full 3D MHD dynamo simulations (ACD code, benchmarked in Christensen et al. 2001; see Dormy et al. 1998 and Aubert et al. 2008). In order to perform an analysis of the predictability of standard mean-field solar dynamos, it was necessary to ensure the compatibility of the model with the ones used within the solar dynamo community. For such reason, we modified and compared outputs from our 3D MHD code with a mean-field solar dynamo benchmark.

The full spherical harmonic expansion of the code writes

$$(\mathcal{P}, \mathcal{T})(r, \theta, \varphi, t) = \sum_{n=1}^{N} \sum_{m=1}^{M} \left(\mathcal{P}_n^m, \mathcal{T}_n^m\right)(r, t)\; Y_n^m(\theta, \varphi), \quad \text{(A1)}$$

truncated at spherical harmonic degree and order N and M, respectively. As most mean-field models assume axisymmetry, we set $M = 0$ throughout.

The original inner boundary conditions of Parody considered the inner core as an insulating or electrically conducting medium of finite conductivity (Christensen et al. 2001); in contrast, in the solar context, the radiative zone is modeled as a perfect conductor. This last condition requires to impose

$$\mathcal{P} = 0, \text{ and} \quad \text{(A2)}$$

$$\partial(r\,\mathcal{T})/\partial r = 0 \text{ at the inner boundary.} \quad \text{(A3)}$$

Further modifications of the code included the incorporation of the α and BL source terms in the poloidal induction Equation (12), and the prescription of the flow fields, Ω and $\mathbf{u}_p$, and depth-dependent turbulent diffusivities $\eta(r)$.

The resulting code was tested against published reference solutions of the mean-field community benchmark effort described by Jouve et al. (2008). The benchmarking consists of computing the critical dynamo numbers C^{crit}, and solar activity cycle frequency ω, for three case studies. The three cases include two $\alpha\Omega$ mean-field dynamos (cases A and B, differing only by the prescribed $\eta(r)$) and a BL dynamo (case C). Table 4 displays the values obtained from our code and the Jouve et al. (2008) benchmark ones for each case, while convergence tests

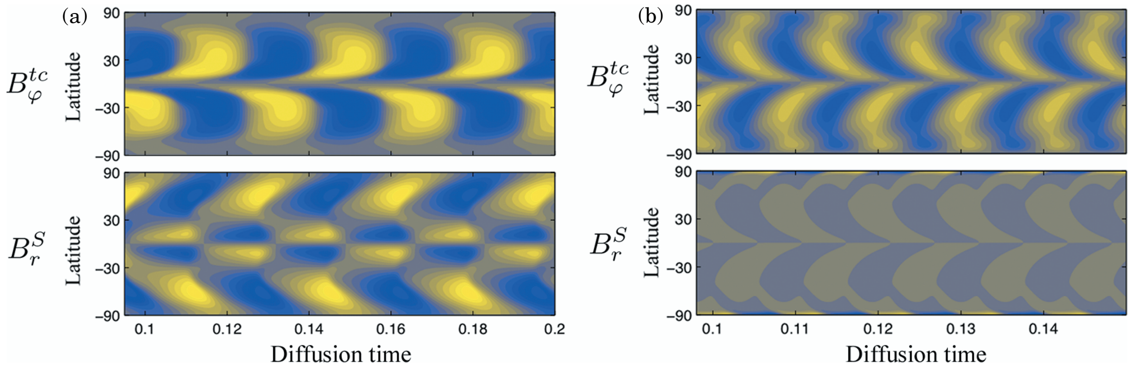


Figure 12. Butterfly diagrams summarizing two different benchmark cases from Jouve et al. (2008): (a) $\alpha\Omega$ dynamo from the supercritical case SB and (b) a Babcock-Leighton dynamo from the supercritical case SC. For each case, the upper panel displays the toroidal field at the tachocline and the lower one the radial field at the surface.

(A color version of this figure is available in the online journal.)

of the critical numbers of cases B and C are shown in Figure 11. In addition, butterfly diagrams for the supercritical cases SB and SC (the supercritical cases include α and S_{BL} quenching) are displayed in Figure 12.

Note that in the present study, and compared with the benchmark cases, we use slightly different inner boundary conditions, namely

$$\mathcal{P} = 0, \text{ and} \tag{A4}$$

$$\mathcal{T} = 0 \text{ at the inner boundary,} \tag{A5}$$

as is common in mean-field solar dynamo simulations (e.g., Dikpati & Charbonneau 1999). Inspection of results obtained with both types of inner boundary conditions shows that they are virtually the same, in agreement with Chatterjee et al. (2004). An interpretation of this is that the low diffusivity of the radiative zone and the absence of a deeply penetrating meridional flow inhibit the penetration of the strong tachocline magnetic field to the deepermost layers.

REFERENCES

Aubert, J., Aurnou, J., & Wicht, J. 2008, GeoJI, 172, 945
Aubert, J., & Fournier, A. 2011, NPGeo, 18, 657
Babcock, H. W. 1961, ApJ, 133, 572
Baker, D. N. 2000, ITPS, 28, 2007
Bonanno, A., Elstner, D., Rüdiger, G., & Belvedere, G. 2002, A&A, 390, 673
Brasseur, P. 2006, in Ocean Weather Forecasting: An Integrated View of Oceanography, ed. E. Chassignet & J. Verron (Dordrecht, The Netherlands: Springer), 271
Brun, A. S. 2007, AN, 328, 329
Brun, A. S. M. S., & Toomre, J. 2004, ApJ, 614, 1073
Bushby, P. J., & Tobias, S. M. 2007, ApJ, 661, 1289
Charbonneau, P. 2005, LRSP, 2, 2
Charbonneau, P., & Dikpati, M. 2000, ApJ, 543, 1027
Charbonneau, P., & Smolarkiewicz, P. K. 2013, Sci, 340, 42
Charbonneau, P., St-Jean, C., & Zacharias, P. 2005, ApJ, 619, 613
Chatterjee, P., Nandy, D., & Choudhuri, A. R. 2004, A&A, 427, 1019
Choudhuri, A. R., Chatterjee, P., & Jiang, J. 2007, PhRvL, 98, 131103
Christensen, U. R., Aubert, J., Cardin, P., et al. 2001, PEPI, 128, 25
DeRosa, M. L., Brun, A. S., & Hoeksema, J. T. 2012, ApJ, 757, 96
Dikpati, M., & Anderson, J. L. 2012, ApJ, 756, 20
Dikpati, M., & Charbonneau, P. 1999, ApJ, 518, 508
Dikpati, M., De Toma, G., & Gilman, P. A. 2006, GeoRL, 33, L05102
Dikpati, M., & Gilman, P. A. 2001, ApJ, 559, 428
Dormy, E., Cardin, P., & Jault, D. 1998, E&PSL, 160, 15
D'Silva, S., & Choudhuri, A. R. 1993, A&A, 272, 621
Elbern, H., Strunk, A., & Nieradzik, L. 2010, in Data Assimilation, ed. W. Lahoz, B. Khattatov, & R. Menard (Berlin: Springer), 491
Fan, Y. 2009, LRSP, 6, 4
Fournier, A., Eymin, C., & Alboussière, T. 2007, NPGeo, 14, 163
Fournier, A., Hulot, G., Jault, D., et al. 2010, SSRv, 155, 247
Fournier, A., Nerger, L., & Aubert, J. 2013, GGG, 14, 4035
Haigh, J. D. 2003, RSPTA, 361, 95
Hathaway, D. H. 1996, ApJ, 460, 1027
Hathaway, D. H. 2009, SSRv, 144, 401
Hathaway, D. H. 2010, LRSP, 7, 1
Hathaway, D. H. 2011, arXiv:1103.1561
Hernández, I. G., Komm, R., Hill, F., et al. 2006, ApJ, 638, 576
Houser, P. R., De Lannoy, G. J., & Walker, J. P. 2010, in Data Assimilation, ed. W. Lahoz, B. Khattatov, & R. Menard (Berlin: Springer), 549
Howe, R., Christensen-Dalsgaard, J., Hill, F., et al. 2000, Sci, 287, 2456
Hulot, G., Lhuillier, F., & Aubert, J. 2010, GeoRL, 37, 6
Jouve, L., Brun, A. S., Arlt, R., et al. 2008, A&A, 483, 949
Jouve, L., Brun, A. S., & Talagrand, O. 2011, ApJ, 735, 31
Jouve, L., Proctor, M. R. E., & Lesur, G. 2010, A&A, 519, 13
Kalnay, E. 2003, Atmospheric Modeling, Data Assimilation and Predictability (Cambridge: Cambridge Univ. Press)
Karak, B. B., & Nandy, D. 2012, ApJL, 761, L13
Kitiashvili, I., & Kosovichev, A. G. 2008, ApJL, 688, L49
Leighton, R. B. 1969, ApJ, 156, 1
Lhuillier, F., Aubert, J., & Hulot, G. 2011, GeoJI, 186, 492
Lorenz, E. N. 1963, JAtS, 20, 130
Lorenz, E. N. 1965, Tell, 17, 321
Miesch, M. S., Featherstone, N. A., Rempel, M., & Trampedach, R. 2012, ApJ, 757, 128
Moffatt, H. K. 1978, Field Generation in Electrically Conducting Fluids (Cambridge: Cambridge Univ. Press)
Ossendrijver, M. 2003, A&ARv, 11, 287
Parker, E. N. 1955, ApJ, 122, 293
Priest, E. R. 1982, Solar Magnetohydrodynamics (Dordrecht: Reidel)
Pulkkinen, T. 2007, LRSP, 4, 1
Sanchez, S. M., Fournier, A., Pinheiro, K. J. R., & Aubert, J. 2014, An. Acad. Bras. Cienc., 86, in press (arXiv:1307.4115)
Talagrand, O. 1997, MeSJJ, 75, 81
Tomczyk, S., Schou, J., & Thompson, M. J. 1995, ApJL, 448, L57
Vallis, G. K. 2006, Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-scale Circulation (Cambridge: Cambridge Univ. Press)
Wang, Y. M., & Sheeley, N. R. 2009, ApJL, 694, L11
Zhao, J., Bogart, R., Kosovichev, A., Duvall, T., Jr., & Hartlep, T. 2013, ApJL, 774, L29

Bibliography

Alboussiere, T., Deguen, R., and Melzani, M. (2010). Melting-induced stratification above the Earth's inner core due to convective translation. *Nature*, 466(7307):744–747.

Alfè, D., Gillan, M., and Price, G. (2003). Thermodynamics from first principles: temperature and composition of the Earth's core. *Mineralogical Magazine*, 67(1):113–123.

Anufriev, A., Jones, C., and Soward, A. (2005). The Boussinesq and anelastic liquid approximations for convection in the Earth's core. *Physics of the Earth and Planetary Interiors*, 152(3):163–190.

Aubert, J. (2013). Flow throughout the Earth's core inverted from geomagnetic observations and numerical dynamo models. *Geophysical Journal International*, 192(2):537–556.

Aubert, J. (2015). Geomagnetic forecasts driven by thermal wind dynamics in the Earth's core. *Geophysical Journal International*, 203(3):1738–1751.

Aubert, J., Aurnou, J., and Wicht, J. (2008). The magnetic structure of convection-driven numerical dynamos. *Geophysical Journal International*, 172(3):945–956.

Aubert, J. and Dumberry, M. (2011). Steady and fluctuating inner core rotation in numerical geodynamo models. *Geophysical Journal International*, 184(1):162–170.

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

Aubert, J., Labrosse, S., and Poitou, C. (2009). Modelling the palaeo-evolution of the geodynamo. *Geophysical Journal International*, 179(3):1414–1428.

Backus, G. (1958). A class of self-sustaining dissipative spherical dynamos. *Annals of Physics*, 4(4):372–447.

Backus, G. E. (1988). Bayesian inference in geomagnetism. *Geophysical Journal International*, 92(1):125–142.

Baerenzung, J., Holschneider, M., and Lesur, V. (2014). Bayesian inversion for the filtered flow at the Earth's core-mantle boundary. *Journal of Geophysical Research: Solid Earth*, 119(4):2695–2720.

Bercovici, D. (2015). Mantle dynamics: An introduction and overview. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 1 – 22. Elsevier, Oxford, second edition edition.

Birch, F. (1952). Elasticity and constitution of the Earth's interior. *Journal of Geophysical Research*, 57(2):227–286.

Bocher, M., Coltice, N., Fournier, A., and Tackley, P. (2016). A sequential data assimilation approach for the joint reconstruction of mantle convection and surface tectonics. *Geophysical Journal International*, 204(1):200–214.

Braginsky, S. I. and Roberts, P. H. (1995). Equations governing convection in Earth's core and the geodynamo. *Geophysical & Astrophysical Fluid Dynamics*, 79(1-4):1–97.

Brown, M. C., Donadini, F., Korte, M., Nilsson, A., Korhonen, K., Lodge, A., Lengyel, S. N., and Constable, C. G. (2015). Geomagia50. v3: 1. general structure and modifications to the archeological and volcanic database. *Earth, Planets and Space*, 67(1):1–31.

Brun, A. S. (2007). Towards using modern data assimilation and weather forecasting methods in solar physics. *Astronomische Nachrichten*, 328(3-4):329.

Buehner, M., Charente, C., He, B., et al. (2008). Intercomparison of 4-D Var and EnKF systems for operational deterministic NWP.

Buffett, B. A. and Seagle, C. T. (2010). Stratification of the top of the core due to chemical interactions with the mantle. *Journal of Geophysical Research: Solid Earth*, 115(B4).

Bunge, H.-P., Hagelberg, C., and Travis, B. (2003). Mantle circulation models with variational data assimilation: inferring past mantle flow and structure from plate motion histories and seismic tomography. *Geophysical Journal International*, 152(2):280–301.

Burgers, G., Jan van Leeuwen, P., and Evensen, G. (1998). Analysis scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 126(6):1719–1724.

Bust, G., Garner, T., and Gaussiran, T. (2004). Ionospheric data assimilation three-dimensional (IDA3D): A global, multisensor, electron density specification algorithm. *Journal of Geophysical Research: Space Physics*, 109(A11).

Canet, E., Fournier, A., and Jault, D. (2009). Forward and adjoint quasi-geostrophic models of the geomagnetic secular variation. *Journal of Geophysical Research: Solid Earth*, 114(B11).

Cardinali, C. (2013). Observation influence diagnostic of a data assimilation system. In *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications (Vol. II)*, pages 89–110. Springer.

Carrington, R. C. (1859). Description of a singular appearance seen in the Sun on September 1, 1859. *Monthly Notices of the Royal Astronomical Society*, 20:13–15.

Carton, J. A. and Giese, B. S. (2008). A reanalysis of ocean climate using Simple Ocean Data (SODA). *Monthly Weather Review*, 136(8):2999–3017.

Christensen, U. and Aubert, J. (2006). Scaling properties of convection-driven dynamos in rotating spherical shells and application to planetary magnetic fields. *Geophysical Journal International*, 166(1):97–114.

Christensen, U., Aubert, J., Cardin, P., Dormy, E., Gibbons, S., Glatzmaier, G., Grote, E., Honkura, Y., Jones, C., Kono, M., et al. (2001). A numerical dynamo benchmark. *Physics of the Earth and Planetary Interiors*, 128(1):25–34.

Christensen, U. and Wicht, J. (2015). Numerical dynamo simulations. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 245 – 277. Elsevier, Oxford, second edition edition.

Christensen, U. R., Aubert, J., and Hulot, G. (2010). Conditions for Earth-like geodynamo models. *Earth and Planetary Science Letters*, 296(3):487–496.

Cohn, S. E. (1997). An introduction to estimation theory. *Meteorological Society of Japan Series 2*, 75:147–178.

Cohn, S. E., Sivakumaran, N., and Todling, R. (1994). A fixed-lag Kalman smoother for retrospective data assimilation. *Monthly Weather Review*, 122(12):2838–2867.

Constable, C. and Korte, M. (2015). Centennial- to millennial-scale geomagnetic field variations. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 309 – 341. Elsevier, Oxford, second edition edition.

Constable, C. G., Johnson, C. L., and Lund, S. P. (2000). Global geomagnetic field models for the past 3000 years: transient or permanent flux lobes? *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 358(1768):991–1008.

Constable, C. G., Parker, R. L., and Stark, P. B. (1993). Geomagnetic field models incorporating frozen-flux constraints. *Geophysical Journal International*, 113(2):419–433.

Cosme, E., Brankart, J.-M., Verron, J., Brasseur, P., and Krysta, M. (2010). Implementation of a reduced rank square-root smoother for high resolution ocean data assimilation. *Ocean Modelling*, 33(1):87–100.

Courtillot, V. and Le Mouël, J.-L. (2007). The study of Earth's magnetism (1269–1950): A foundation by Peregrinus and subsequent development of geomagnetism and paleomagnetism. *Reviews of Geophysics*, 45(3).

Cowling, T. G. (1933). The magnetic field of sunspots. *Monthly Notices of the Royal Astronomical Society*, 94:39–48.

Davidson, P. (2013). Scaling laws for planetary dynamos. *Geophysical Journal International*, 195(1):67–74.

Davies, C. J. and Constable, C. G. (2014). Insights from geodynamo simulations into long-term geomagnetic field behaviour. *Earth and Planetary Science Letters*, 404:238–249.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597.

Dikpati, M., Anderson, J. L., and Mitra, D. (2014). Ensemble Kalman filter data assimilation in a Babcock-Leighton solar dynamo model: An observation system simulation experiment for reconstructing meridional flow speed. *Geophysical Research Letters*, 41(15):5361–5369.

Donadini, F., Korhonen, K., Riisager, P., and Pesonen, L. (2006). Database for holocene geomagnetic intensity information. *Eos*, 87(14):137.

Donadini, F., Korte, M., and Constable, C. (2009). Geomagnetic field for 0–3 ka: 1. New data sets for global modeling. *Geochemistry, Geophysics, Geosystems*, 10(6).

Dormy, E. (1997). *Modélisation numérique de la dynamo terrestre*. PhD thesis.

Dormy, E., Cardin, P., and Jault, D. (1998). MHD flow in a slightly differentially rotating spherical shell, with conducting inner core, in a dipolar magnetic field. *Earth and Planetary Science Letters*, 160(1):15–30.

Dormy, E. and Le Mouël, J.-L. (2008). Geomagnetism and the dynamo: where do we stand? *Comptes Rendus Physique*, 9(7):711–720.

Dziewonski, A. M. and Anderson, D. L. (1981). Preliminary reference earth model. *Physics of the earth and planetary interiors*, 25(4):297–356.

Egbert, G. D., Bennett, A. F., and Foreman, M. G. (1994). TOPEX/POSEIDON tides estimated using a global inverse model.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162.

Evensen, G. (2009). *Data assimilation: the ensemble Kalman filter*. Springer Science & Business Media.

Finlay, C. C. and Jackson, A. (2003). Equatorially dominated magnetic field change at the surface of earth's core. *Science*, 300(5628):2084–2086.

Fournier, A., Aubert, J., and Thébault, E. (2015). A candidate secular variation model for IGRF-12 based on Swarm data and inverse geodynamo modelling. *Earth, Planets and Space*, 67(1):81.

Fournier, A., Eymin, C., and Alboussiere, T. (2007). A case for variational geomagnetic data assimilation: insights from a one-dimensional, nonlinear, and sparsely observed mhd system. *Nonlinear Processes in Geophysics*, 14(2):163–180.

Fournier, A., Hulot, G., Jault, D., Kuang, W., Tangborn, A., Gillet, N., Canet, E., Aubert, J., and Lhuillier, F. (2010). An introduction to data assimilation and predictability in geomagnetism. *Space science reviews*, 155(1-4):247–291.

Fournier, A., Nerger, L., and Aubert, J. (2013). An ensemble Kalman filter for the time-dependent analysis of the geomagnetic field. *Geochemistry, Geophysics, Geosystems*, 14(10):4035–4043.

Friis-Christensen, E., Lühr, H., Hulot, G., Haagmans, R., and Purucker, M. (2009). Geomagnetic research from space. *Eos*, 90(25):213–214.

Gallet, Y., Genevey, A., and Le Goff, M. (2002). Three millennia of directional variation of the Earth's magnetic field in western Europe as revealed by archeological artefacts. *Physics of the Earth and Planetary Interiors*, 131(1):81–89.

Genevey, A., Gallet, Y., Constable, C., Korte, M., and Hulot, G. (2008). Archeoint: An upgraded compilation of geomagnetic field intensity data for the past ten millennia and its application to the recovery of the past dipole moment. *Geochemistry, Geophysics, Geosystems*, 9(4).

Genevey, A., Gallet, Y., Rosen, J., and Le Goff, M. (2009). Evidence for rapid geomagnetic field intensity variations in western europe over the past 800 years from new french archeointensity data. *Earth and Planetary Science Letters*, 284(1):132–143.

Ghil, M. and Malanotte-Rizzoli, P. (1991). Data assimilation in meteorology and oceanography. *Adv. Geophys*, 33:141–266.

Gilbert, W. (1600). De Magnete.

Gillet, N., Barrois, O., and Finlay, C. C. (2015). Stochastic forecasting of the geomagnetic field from the COV-OBS. x1 geomagnetic field model, and candidate models for IGRF-12. *Earth, Planets and Space*, 67(1):1–14.

Gillet, N., Jault, D., Canet, E., and Fournier, A. (2010). Fast torsional waves and strong magnetic field within the Earth's core. *Nature*, 465(7294):74–77.

Gillet, N., Jault, D., Finlay, C., and Olsen, N. (2013). Stochastic modeling of the Earth's magnetic field: inversion for covariances over the observatory era. *Geochemistry, Geophysics, Geosystems*, 14(4):766–786.

Glatzmaier, G. A. (2002). Geodynamo simulations-how realistic are they? *Annual Review of Earth and Planetary Sciences*, 30(1):237–257.

Glatzmaier, G. A. and Olson, P. (2005). Probing the geodynamo. *Scientific American*, 292(4):50–57.

Glatzmaier, G. A. and Roberts, P. H. (1995). A three-dimensional convective dynamo solution with rotating and finitely conducting inner core and mantle. *Physics of the Earth and Planetary Interiors*, 91(1):63–75.

Gubbins, D. (2004). *Time series analysis and inverse theory for geophysicists*. Cambridge University Press.

Gubbins, D. and Bloxham, J. (1985). Geomagnetic field analysis III. Magnetic fields on the core-mantle boundary. *Geophysical Journal International*, 80(3):695–713.

Gubbins, D. and Davies, C. (2013). The stratified layer at the core–mantle boundary caused by barodiffusion of oxygen, sulphur and silicon. *Physics of the Earth and Planetary Interiors*, 215:21–28.

Gubbins, D. and Roberts, N. (1983). Use of the frozen flux approximation in the interpretation of archaeomagnetic and palaeomagnetic data. *Geophysical Journal International*, 73(3):675–687.

Hale, G. E. (1908). On the probable existence of a magnetic field in sun-spots. *The astrophysical journal*, 28:315.

Halley, E. (1692). An account of the cause of the change of the variation of the magnetical needle; with an hypothesis of the structure of the internal parts of the Earth: As it was proposed to the royal society in one of their late meetings. by edm. halley. *Philosophical transactions*, 16(179-191):563–578.

Hirose, K., Labrosse, S., and Hernlund, J. (2013). Composition and state of the core. *Annual Review of Earth and Planetary Sciences*, 41:657–691.

Holme, R. (1998). Electromagnetic core-mantle coupling I. explaining decadal changes in the length of day. *Geophysical journal international*, 132(1):167–180.

Holme, R. and Olsen, N. (2006). Core surface flow modelling from high-resolution secular variation. *Geophysical Journal International*, 166(2):518–528.

Hongre, L., Hulot, G., and Khokhlov, A. (1998). An analysis of the geomagnetic field over the past 2000 years. *Physics of the Earth and Planetary Interiors*, 106(3):311–335.

Hulot, G., Lhuillier, F., and Aubert, J. (2010). Earth's dynamo limit of predictability. *Geophysical Research Letters*, 37(6).

Hung, C. P., Jouve, L., Brun, A. S., Fournier, A., and Talagrand, O. (2015). Estimating the deep solar meridional circulation using magnetic observations and a dynamo model: a variational approach. *The Astrophysical Journal*, 814(2):151.

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

Jacobs, J. A. (1987). *Geomagnetism*, volume 2. Academic Press.

Jault, D. (2015). Illuminating the electrical conductivity of the lowermost mantle from below. *Geophysical Journal International*, 202(1):482–496.

Johnson, C. L. and Constable, C. G. (1997). The time-averaged geomagnetic field: global and regional biases for 0–5 Ma. *Geophysical Journal International*, 131(3):643–666.

Jouve, L., Brun, A. S., and Talagrand, O. (2011). Assimilating data into an $\alpha\omega$ dynamo model of the sun: a variational approach. *The Astrophysical Journal*, 735(1):31.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability.* Cambridge university press.

Kitiashvili, I. and Kosovichev, A. (2008). Application of data assimilation method for predicting solar cycles. *The Astrophysical Journal Letters*, 688(1):L49.

Kivelson, M. G. and Russell, C. T. (1995). *Introduction to space physics.* Cambridge university press.

Knutti, R. and Sedláček, J. (2013). Robustness and uncertainties in the new CMIP5 climate model projections. *Nature Climate Change*, 3(4):369–373.

Korte, M. and Constable, C. (2003). Continuous global geomagnetic field models for the past 3000 years. *Physics of the Earth and Planetary Interiors*, 140(1):73–89.

Korte, M. and Constable, C. (2008). Spatial and temporal resolution of millennial scale geomagnetic field models. *Advances in Space Research*, 41(1):57–69.

Korte, M., Donadini, F., and Constable, C. (2009). Geomagnetic field for 0–3 ka: 2. A new series of time-varying global models. *Geochemistry, Geophysics, Geosystems*, 10(6).

Kuang, W. and Bloxham, J. (1997). An Earth-like numerical dynamo model. *Nature*, 389(6649):371–374.

Kuang, W., Tangborn, A., Jiang, W., Liu, D., Sun, Z., Bloxham, J., and Wei, Z. (2008). MoSST-DAS: the first generation geomagnetic data assimilation framework. *Commun. Comput. Phys*, 3(1):85–108.

Kuang, W., Tangborn, A., Wei, Z., and Sabaka, T. (2009). Constraining a numerical geodynamo model with 100 years of surface observations. *Geophysical Journal International*, 179(3):1458–1468.

Kuang, W., Wei, Z., Holme, R., and Tangborn, A. (2010). Prediction of geomagnetic field with data assimilation: a candidate secular variation model for igrf-11. *Earth, planets and space*, 62(10):775–785.

Langel, R. and Estes, R. (1982). A geomagnetic field spectrum. *Geophysical Research Letters*, 9(4):250–253.

Larmor, J. (1919). How could a rotating body such as the Sun become a magnet?, concrete mine timbers. *Scientific American*, 88:287–287.

Lhuillier, F., Aubert, J., and Hulot, G. (2011a). Earth's dynamo limit of predictability controlled by magnetic dissipation. *Geophysical Journal International*, 186(2):492–508.

Lhuillier, F., Fournier, A., Hulot, G., and Aubert, J. (2011b). The geomagnetic secular-variation timescale in observations and numerical dynamo models. *Geophysical Research Letters*, 38(9).

Li, K., Jackson, A., and Livermore, P. W. (2011). Variational data assimilation for the initial-value dynamo problem. *Physical Review E*, 84(5):056321.

Li, K., Jackson, A., and Livermore, P. W. (2014). Variational data assimilation for a forced, inertia-free magnetohydrodynamic dynamo model. *Geophysical Journal International*, 199(3):1662–1676.

Licht, A., Hulot, G., Gallet, Y., and Thébault, E. (2013). Ensembles of low degree archeomagnetic field models for the past three millennia. *Physics of the Earth and Planetary Interiors*, 224:38–67.

Liu, D., Tangborn, A., and Kuang, W. (2007). Observing system simulation experiments in geomagnetic data assimilation. *Journal of Geophysical Research: Solid Earth*, 112(B8).

Masters, G., Johnson, S., Laske, G., Bolton, H., and Davies, J. (1996). A shear-velocity model of the mantle [and discussion]. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 354(1711):1385–1411.

Merrill, R. T., McElhinny, M., and Mcfadden, P. L. (1996). The magnetic field of the Earth. *International Geophysics Series*, 63.

Moffatt, H. K. (1978). *Field Generation in Electrically Conducting Fluids.* Cambridge University Press, Cambridge, London, New York, Melbourne.

Monnereau, M., Calvet, M., Margerin, L., and Souriau, A. (2010). Lopsided growth of Earth's inner core. *Science*, 328(5981):1014–1017.

Morales, J., Alva-Valdivia, L. M., Goguitchaichvili, A., and Urrutia-Fucugauchi, J. (2006). Cooling rate corrected paleointensities from the Xitle lava flow: Evaluation of within-site scatter for single spot-reading cooling units. *Earth, planets and space*, 58(10):1341–1347.

Nævdal, G., Johnsen, L. M., Aanonsen, S. I., Vefring, E. H., et al. (2005). Reservoir monitoring and continuous model updating using ensemble Kalman filter. *SPE journal*, 10(01):66–74.

Nerger, L. (2004). *Parallel filter algorithms for data assimilation in oceanography.* PhD thesis, Universität Bremen.

Nerger, L. and Hiller, W. (2013). Software for ensemble-based data assimilation systems. implementation strategies and scalability. *Computers & Geosciences*, 55:110–118.

Nilsson, A., Holme, R., Korte, M., Suttie, N., and Hill, M. (2014). Reconstructing holocene geomagnetic field variation: new methods, models and implications. *Geophysical Journal International*, page ggu120.

Nilsson, A., Snowball, I., Muscheler, R., and Uvo, C. B. (2010). Holocene geocentric dipole tilt model constrained by sedimentary paleomagnetic data. *Geochemistry, Geophysics, Geosystems*, 11(8).

Noller, J. S., Sowers, J. M., and Lettis, W. R. (2000). *Quaternary geochronology: Methods and applications*, volume 4. American Geophysical Union.

Ohta, K., Hirose, K., Ichiki, M., Shimizu, K., Sata, N., and Ohishi, Y. (2010). Electrical conductivities of pyrolitic mantle and morb materials up to the lowermost mantle conditions. *Earth and Planetary Science Letters*, 289(3):497–502.

Olson, P. (2015). Core dynamics: An introduction and overview. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 1 – 25. Elsevier, Oxford, second edition edition.

Olson, P. and Christensen, U. (2002). The time-averaged magnetic field in numerical dynamos with non-uniform boundary heat flow. *Geophysical Journal International*, 151(3):809–823.

O‘Rourke, J. G. and Stevenson, D. J. (2016). Powering Earth’s dynamo with magnesium precipitation from the core. *Nature*, 529(7586):387–389.

Panovska, S., Finlay, C., Donadini, F., and Hirt, A. (2012). Spline analysis of holocene sediment magnetic records: Uncertainty estimates for field modeling. *Journal of Geophysical Research: Solid Earth*, 117(B2).

Panovska, S., Korte, M., Finlay, C., and Constable, C. (2015). Limitations in paleomagnetic data and modelling techniques and their impact on Holocene geomagnetic field models. *Geophysical Journal International*, 202(1):402–418.

Parker, R. L. (1994). *Geophysical inverse theory*. Princeton university press.

Pavón-Carrasco, F. J., Osete, M. L., Torta, J. M., and De Santis, A. (2014). A geomagnetic field model for the holocene based on archaeomagnetic and lava flow data. *Earth and Planetary Science Letters*, 388:98–109.

Poirier, J. (1988). Transport properties of liquid metals and viscosity of the Earth’s core. *Geophysical Journal International*, 92(1):99–105.

Pozzo, M., Davies, C., Gubbins, D., and Alfè, D. (2012). Thermal and electrical conductivity of iron at Earth’s core conditions. *Nature*, 485(7398):355–358.

Radelet-de Grave, P. and Speiser, D. (1975). Le ”De Magnete” de Pierre de Maricourt: Traduction et commentaire. *Revue d’histoire des sciences*, pages 193–234.

Ricard, Y. (2015). Physics of mantle convection. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 23 – 71. Elsevier, Oxford, second edition edition.

Roberts, P. H. and King, E. M. (2013). On the genesis of the Earth’s magnetism. *Reports on Progress in Physics*, 76(9):096801.

Rodell, M., Houser, P., Jambor, U. e. a., Gottschalck, J., Mitchell, K., Meng, C., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., et al. (2004). The global land data assimilation system. *Bulletin of the American Meteorological Society*, 85(3):381–394.

Ryan, D. and Sarson, G. (2008). The geodynamo as a low-dimensional deterministic system at the edge of chaos. *EPL (Europhysics Letters)*, 83(4):49001.

Sambridge, M., Rickwood, P., Rawlinson, N., and Sommacal, S. (2007). Automatic differentiation in geophysical inverse problems. *Geophysical Journal International*, 170(1):1–8.

Sanchez, S., Fournier, A., and Aubert, J. (2014a). The predictability of advection-dominated flux-transport solar dynamo models. *The Astrophysical Journal*, 781(1):8.

Sanchez, S., Fournier, A., Pinheiro, K. J., and Aubert, J. (2014b). A mean-field Babcock-Leighton solar dynamo model with long-term variability. *Anais da Academia Brasileira de Ciências*, 86(1):11–26.

Sluijs, M. A. (2014). A common miscitation of William Gilbert. *Eos, Transactions American Geophysical Union*, 95(16):137–137.

Souriau, A. and Calvet, M. (2015). Deep earth structure: The earths cores. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 725 – 757. Elsevier, Oxford, second edition edition.

Sumita, I. and Bergman, M. (2015). Inner core dynamics. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 297 – 316. Elsevier, Oxford, second edition edition.

Sun, Z., Tangborn, A., and Kuang, W. (2007). Data assimilation in a sparsely observed one-dimensional modeled MHD system. *Nonlinear Processes in Geophysics*, 14(2):181–192.

Talagrand, O. (1997). Assimilation of observations, an introduction. *Journal-Meteorological Society of Japan Series 2*, 75:81–99.

Talagrand, O. (2003). *Data Assimilation for the Earth System*, chapter A Posteriori Validation of Assimilation Algorithms, pages 85–95. Springer Netherlands, Dordrecht.

Tangborn, A. and Kuang, W. (2015). Geodynamo model and error parameter estimation using geomagnetic data assimilation. *Geophysical Journal International*, 200(1):664–675.

Tarduno, J. A., Cottrell, R. D., Davis, W. J., Nimmo, F., and Bono, R. K. (2015). A Hadean to paleoarchean geodynamo recorded by single zircon crystals. *Science*, 349(6247):521–524.

Tauxe, L. and Yamazaki, T. (2015). Paleointensities. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 461 – 509. Elsevier, Oxford, second edition edition.

Taylor, G. I. (1922). The motion of a sphere in a rotating liquid. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 102(715):180–189.

Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., Bertrand, F., Bondar, T., Boness, A., Brocco, L., et al. (2015). International geomagnetic reference field: the 12th generation. *Earth, Planets and Space*, 67(1):1–19.

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

Tkalčić, H., Young, M., Bodin, T., Ngo, S., and Sambridge, M. (2013). The shuffling rotation of the Earth's inner core revealed by earthquake doublets. *Nature Geoscience*, 6(6):497–502.

Turner, G., Rasson, J., and Reeves, C. (2015). Observation and measurement techniques. In Schubert, G., editor, *Treatise on Geophysics (Second Edition)*, pages 91 – 135. Elsevier, Oxford, second edition edition.

Valet, J.-P., Herrero-Bervera, E., LeMouël, J.-L., and Plenier, G. (2008). Secular variation of the geomagnetic dipole during the past 2000 years. *Geochemistry, Geophysics, Geosystems*, 9(1).