FULL PAPER

Open Access

Sequential modelling of the Earth's core magnetic field

Guillaume Ropp[1]* , Vincent Lesur[1] , Julien Baerenzung[2] and Matthias Holschneider[2]

Abstract

We describe a new, original approach to the modelling of the Earth's magnetic field. The overall objective of this study is to reliably render fast variations of the core field and its secular variation. This method combines a sequential modelling approach, a Kalman filter, and a correlation-based modelling step. Sources that most significantly contribute to the field measured at the surface of the Earth are modelled. Their separation is based on strong prior information on their spatial and temporal behaviours. We obtain a time series of model distributions which display behaviours similar to those of recent models based on more classic approaches, particularly at large temporal and spatial scales. Interesting new features and periodicities are visible in our models at smaller time and spatial scales. An important aspect of our method is to yield reliable error bars for all model parameters. These errors, however, are only as reliable as the description of the different sources and the prior information used are realistic. Finally, we used a slightly different version of our method to produce candidate models for the thirteenth edition of the International Geomagnetic Reference Field.

Keywords: Geomagnetic field, Secular variation, Kalman filter, IGRF

Introduction

The magnetic field surrounding the Earth is sustained by—and constantly evolving due to—the motions in the Earth's liquid outer core. It has been observed and studied for centuries, and recent evolution in both technological and mathematical methods has allowed us to understand more and more of its dynamics. Although studies of ancient variations of the geomagnetic field must rely on archeomagnetic and paleomagnetic data which are sparse and thus give only access to long time scales, the study of recent changes relies on a worldwide, dense data distribution, and efficient data acquisition platforms. With the significant increase of the number of magnetic observatories since the 1950s and the launch of several magnetic satellite missions in the last two decades, a continuous set of high quality data has allowed a deeper insight into the magnetic field evolution.

The field measured at the surface of the Earth results from the contributions of numerous sources, that must be separated in order to access their individual variations. We are particularly interested in those of the core field, i.e. the field generated in the liquid outer core of the Earth. One obstacle to progress in this domain is the presence of fields generated by phenomena such as the interactions between the core field and charged particles in the thermosphere, where ionospheric currents flow, and in the magnetosphere. These generate currents that in turn produce signals—the so-called external fields—that contribute to the measured magnetic field. They evolve on time scales ranging from seconds to years and these variations induce currents in the Earth's core, mantle, lithosphere and oceans that also generate magnetic signals. Separating all these contributions from the core field requires an adequate handling of data and has long been an important obstacle to the development of high-resolution core field models. In the past decades, however, the geomagnetic modelling community has been able to build better,

*Correspondence: ropp@ipgp.fr
[1] Université de Paris, Institut de physique du globe de Paris, CNRS, 75005 Paris, France
Full list of author information is available at the end of the article

Ropp *et al. Earth, Planets and Space*   (2020) 72:153

Page 2 of 15

more accurate models of the core field and its secular variation (SV). The separation of the different sources is now well controlled, with some remaining difficulties, especially at high latitudes. Modern models—e.g. the GRIMM (Lesur et al. 2015) or, more recently, the Chaos-6 model (Finlay et al. 2016), both of which use splines of order 6 for their time evolution—separate well contributions from the core and external sources. They are typically able to render short core field time scales, although with a precision depending on the spatial scales—e.g. small spherical harmonics degrees have a resolution of the order of 2 years. Besides using splines, sequential approaches have also been used. For example, the POMME model (Chulliat and Maus 2014) uses a 3-year sliding time window, but its time resolution remains of the same order.

In this paper, we describe an original method for modelling the Earth's magnetic field. Our aim is to build a high-resolution times series of field models over the satellite era by separating most external and internal sources at small time scales. In order to achieve this, all modelled sources must be described as reliably as possible. This implies that the models contain a large amount of parameters, leading to very large models and considerable computation times, as soon as the field is modelled for more than a few months. To overcome this problem, we use a Kalman filter process to generate a time series of snapshot models. Each snapshot model covers only a given time period, of about 3 months (or 12 months for the parent model of our IGRF-13 candidate). This reduces the amount of data analysed for a single inversion, so we need to add strong and reliable prior information on the spatial behaviour of the modelled sources, in order to further constrain our problem. The Kalman filter is a 3-step process. In the first step—the analysis, an a priori information on the field is updated through a correlation-based data assimilation method described in Holschneider et al. (2016). In this latter work, the spatial correlations of the different signals are described in the spherical harmonics (SH) domain and then used in the spatial domain to constrain the inversion process. Here, we have a similar approach, but we remain in the SH domain. Then, the information gained by the analysis at a given time step is used to predict an estimate of the field at the next time step, which serves as a new prior. When the whole time period is covered, a backward smoothing is applied. The evolution in time is not directly parameterised—e.g. through splines—but is a consequence of the analysis, prediction and smoothing steps. The Kalman filter has already been used, noticeably by Lesur et al. (2017), but no smoothing was applied there.

Gillet et al. (2015) and Beggan and Whaler (2009) also proposed an implementation of the Kalman filter for geomagnetic modelling. The Kalman filter has also been used to compute core field and core surface flow models in geomagnetic modelling. Barrois et al. (2018) describe an application of the Kalman filter to the estimation of the surface core flow. Baerenzung et al. (2018) use a full Ensemble Kalman filter to simulate the evolution of an ensemble of flow models, along with its statistical properties. None of these approaches have attempted to apply the Kalman filter to the modelling of all major sources as presented here.

The remaining of the paper is organised as follows: the first section describes the data sets used and the selection criteria applied to filter them. Then, the composition of the model and the parameterisation of each included source is detailed. Next, each of the three steps of the Kalman filter is described. In the Results section, we present our model time series, spanning 2000.0 to 2019.75. Via different representations, we assess the reliability of our method by comparing the resulting models to the Chaos-6 models, and discuss the features that characterise our models. We show how our approach can bring new insight on the core field secular variation. Before concluding, the application of this methodology in deriving candidates for the thirteenth edition of the International Geomagnetic Reference Field (IGRF) is detailed (see Thébault et al. (2015) for the preceding edition of the IGRF).

Data

The data set used for our modelling is compiled from ground observatory and satellite data, and is made up exclusively of vector magnetic data. It covers a period of 20 years, from 2000.0 to 2020.0. The observatory data set is built using hourly means from all available ground observatories from 2000.0 to 2020.0, reprocessed according to Macmillan and Olsen (2013), so that it covers this time period continuously. Satellite data are compiled from the Champ and Swarm missions data sets. The Champ data covers a time period ranging from Sept. 2000 to Aug. 2010. A gap of about 3 years in satellite data separates the Champ and Swarm missions. The Swarm data set spans Nov. 2013 to the end of 2020. It includes the latest available versions (0505, 0506 and 0507) of level-1b vectorial data files in December 2019 from the Swarm A (Alpha) and B (Bravo) satellites. Swarm C (Charlie) data were not used, as their information content on the core field is very similar to that of satellite A. We distinguish between "high latitude" (HL) data, with absolute magnetic latitudes above 55°, and "medium-to-low

latitude" (ML) data, of absolute magnetic latitudes below 55°. HL data are handled in the usual North, East, Centre (NEC) reference frame, whereas ML data are used in a Solar Magnetic (SM) reference frame, reducing this way the correlations between vector data component errors (Lesur et al. 2008). All data are originally taken in the NEC reference frame.

Data selection

We apply an overall light selection on the data set [see Thomson and Lesur (2007)]. The selection criteria for all types of data are detailed in Table 1. Different criteria apply for HL and ML data. ML data are taken only inside the 23:00–5:00 local time window, while HL data are selected for all local times. Different time samplings are set for ML and HL data, to compensate for the higher data density at high latitudes due to the nearly sunsynchronous satellite orbits (see Table 1). Data are also selected for a limited range of values of the D_{st} index, and for positive values of the z component of the interplanetary magnetic field (IMF). Finally, observatory and ML satellite data are selected only if they are located in a nonsunlit area.

We define N_v and N_d, the number of selected vector data, and the total number of selected data, respectively ($N_d = 3 \times N_v$).

Data weights

The variances attributed to each type of data are given in Table 2. X,Y and Z are the coordinates in the NEC reference frame for HL data, and the SM frame for the ML data. In this study, data errors are assumed to be uncorrelated, so the time and spatial covariance is set to zero between different data samples. Similarly, the errors of the different data vector components are assumed to be uncorrelated. The inverse of the variances are used to weight the data.

Model parametrisation

To parameterise the model in time, we introduced a grid over the time period 2000.0–2020.0, composed from $N_t + 1$ knots denoted t_k, such that $t_k = t_0 + k\Delta t$, for

Table 2 Data variances for each type and component in NEC coordinates (HL) or SM coordinates (ML)

Component/ data type	Satellite data		Observatory data	
	ML	HL	ML	HL
X	$v = 9\,\mathrm{nT}^2$	$v = 100\,\mathrm{nT}^2$	$v = 16\,\mathrm{nT}^2$	$v = 36\,\mathrm{nT}^2$
Y	$v = 9\,\mathrm{nT}^2$	$v = 81\,\mathrm{nT}^2$	$v = 16\,\mathrm{nT}^2$	$v = 25\,\mathrm{nT}^2$
Z	$v = 16\,\mathrm{nT}^2$	$v = 81\,\mathrm{nT}^2$	$v = 25\,\mathrm{nT}^2$	$v = 36\,\mathrm{nT}^2$

Data are initially weighted by the inverse of their variance, before undergoing an iterative Huber-weighting process. These values aim at including unknown signals as well as instrument noise. They correspond to typical level of data misfit in prior studies [see e.g. (Lesur et al. 2015)]

$k = 0\ldots, N_t$, with $\Delta t = 365.25/4$ days (i.e. roughly 3 months) and t_0 corresponding to decimal year 2000.0. It results that $N_t = 80$. A discrete time series of N_t snapshot models is computed over the whole time period.

One snapshot model is constructed for time t_k using data from the time interval $[t_k; t_{k+1}]$. In this time interval, the model includes multiple internal and external sources, nearly all of which are parameterised through spherical harmonics (SH). Internal sources include the static core field and its secular variation (up to SH degree 18) and the lithospheric field (for SH degrees 15 to 30). They also include internal fields induced by magnetospheric currents that evolve on time scales up to a few months, along with their variations (up to SH degree 6), and the internal part of the D_{st} indexed field, which is denoted I_{st} (up to SH degree 3). Including this dependance to I_{st} allows to track as fast as hourly variations of the induced fields. A known lithospheric field model (Lesur et al. 2013), computed from SH coefficients of degrees 30 to 120, is subtracted from the data. External sources include the outer magnetospheric field in Geocentric Solar Magnetic (GSM) coordinates, the inner magnetospheric field in Solar Magnetic (SM) coordinates, a time varying field indexed on the E_{st}—the external part of the D_{st} (in SM coordinates)—and another one indexed on hourly mean values of the Y component of the interplanetary magnetic field (IMF), in SM coordinates. All these sources are modelled for

Table 1 Selection criteria for satellite (Champ and Swarm) and ground observatory data

Mag. latitude	|mag. lat.| ≥ 55°			|mag. lat.| ≤ 55°		
Data type	Swarm	Champ	Obs.	Swarm	Champ	Obs.
Time sampling	60 s	60 s	1 h	30 s	20 s	1 h
Local time	–	–	–	23:00–5:00	23:00–5:00	23:00–5:00
D_{st}	$[-30 : 30]\,\mathrm{nT}$	$[-30 : 30]\,\mathrm{nT}$	$[-30 : 10]\,\mathrm{nT}$	$[-30 : 30]\,\mathrm{nT}$	$[-30 : 30]\,\mathrm{nT}$	$[-30 : 10]\,\mathrm{nT}$
IMF	$B_z^{IMF} > 0$	$B_z^{IMF} > 0$	$B_z^{IMF} > 0$	$B_z^{IMF} > 0$	$B_z^{IMF} > 0$	$B_z^{IMF} > 0$
Sunlit area excl.	No	No	Yes	Yes	Yes	Yes
N_v	1,104,672	637,561	906,513	822,246	1,004,634	1,646,564

Ropp *et al. Earth, Planets and Space* (2020) 72:153

Page 4 of 15

SH degrees 1 to 3. For each ground observatory, local static contributions are modelled by 3 constant values, one for each vector direction. All modelled sources, along with their characteristics, are listed in Table 3. Note that the separation of the D_{st} in I_{st} and E_{st}, its internal and external components, respectively, is part of the data preprocessing [see Maus and Weidelt (2004)].

The equation linking the model parameters to the value of the field **B** at a point (r, θ, φ, t) in spherical coordinates (i.e. radius, colatitude and longitude), and time $t = t_k + \delta t$, with $\delta t < \Delta t$, is given in Eq. (1),

where the various g_ℓ^m and q_ℓ^m are the model parameters for internal and external sources, respectively. The sources and coefficients are listed in Table 3. $\mathbf{O}_i = \left(O_x^i, O_y^i, O_z^i \right)$ is the crustal offset at the location of observatory i, with $(r_i, \theta_i, \varphi_i)$ the respective observatory spherical coordinates. The $(\theta_{SM}, \varphi_{SM})$ (resp. $(\theta_{GSM}, \varphi_{GSM})$) are the coordinates in the SM (resp. GSM) system of coordinates. The reference radius for the SH development, denoted a, is set to the usual Earth's surface radius, $a = 6371.2$ km. The symbol $\sum_{\ell, m}$ stands for the double sum $\sum_{\ell=1}^{L} \sum_{m=-\ell}^{\ell}$, with L the maximal SH degree considered. The $\hat{\mathbf{Y}}_{\ell,\ell+1}^m$ and $\hat{\mathbf{Y}}_{\ell,\ell-1}^m$ are the vector spherical harmonics defined by:

$$
\begin{aligned}
\mathbf{B}(r, \theta, \varphi, t) = &\sum_{\ell, m} \left(\frac{a}{r}\right)^{\ell+2} \left[{}^C g_{\ell,k}^m + {}^L g_{\ell,k}^m + \delta t\, {}^C \dot{g}_{\ell,k}^m \right] \hat{\mathbf{Y}}_{\ell,\ell+1}^m(\theta, \varphi) \\
&+ \sum_{\ell, m} \left(\frac{a}{r}\right)^{\ell+2} \left[{}^I g_{\ell,k}^m + \delta t\, {}^I \dot{g}_{\ell,k}^m + {}^{I_{st}} g_{\ell,k}^m\, I_{st}(t) \right] \hat{\mathbf{Y}}_{\ell,\ell+1}^m(\theta, \varphi) \\
&+ \sum_{\ell, m} \left(\frac{r}{a}\right)^{\ell-1} {}^G q_{\ell,k}^m\, \hat{\mathbf{Y}}_{\ell,\ell-1}^m(\theta_{GSM}, \varphi_{GSM}) \\
&+ \sum_{\ell, m} \left(\frac{r}{a}\right)^{\ell-1} \left[{}^S q_{\ell,k}^m + {}^I q_{\ell,k}^m\, B_y^{IMF}(t) + {}^{E_{st}} q_{\ell,k}^m\, E_{st}(t) \right] \hat{\mathbf{Y}}_{\ell,\ell-1}^m(\theta_{SM}, \varphi_{SM}) \\
&+ \sum_{i \in \mathcal{I}_{obs}} \delta(r - r_i, \theta - \theta_i, \varphi - \varphi_i)\, \mathbf{O}_{i,k},
\end{aligned}
\tag{1}
$$

Table 3 Summary of all modelled sources with relevant parameters and prior information values

Source	L_i, L_a	Time scale	Prior	Nb. of par.	Coef.
Static core field	1, 18	$\tilde{\tau}_\ell^c = 415/\ell$ year	$R = 3000$ km $S = 2 \cdot 10^9$ nT^2	360	${}^C g_\ell^m$
Core SV	1, 18	$\tilde{\tau}_\ell^{sv} = 11$ year	$R = 3500$ km $S = 1.5 \cdot 10^4$ (nT/year)^2	360	${}^C \dot{g}_\ell^m$
Lithospheric field	15, 30	$\tau^{lt} = 10^6$ year	$R = 6280$ km $S = 2.7 \cdot 10^{-2}$ nT^2	736	${}^L g_\ell^m$
Induced field	1, 6	$\tau^I < \Delta t$	–	48	${}^I g_\ell^m$
Induced field variation	1, 6	$\tau^{\partial_t I} < \Delta t$	–	48	${}^I \dot{g}_\ell^m$
GSM ext. field	1, 3	$\tau^{GSM} < \Delta t$	$R = 1.6 \cdot 10^4$ km $S = 5.4 \cdot 10^3$ nT^2	15	${}^G q_\ell^m$
SM ext. field	1, 3	$\tau^{SM} < \Delta t$	$R = 6900$ km $S = 3.56$ nT^2	15	${}^S q_\ell^m$
I_{st} indexed field	1, 3	$\tau^{Ist} < \Delta t$	$R = 2537$ km $S = 1.0$	15	${}^{Ist} g_\ell^m$
E_{st} indexed field	1, 3	$\tau^{Est} < \Delta t$	$R = 1.6 \cdot 10^4$ km $S = 5.4$	15	${}^{Est} q_\ell^m$
B_y^{IMF} indexed field	1, 3	$\tau^{IMF} < \Delta t$	$R = 6900$ km $S = 0.1$	15	${}^I q_\ell^m$
Obs. offset	Non SH	$\tau^{lt} = 10^6$ year	–	3×200	$\mathbf{O}_i$

From left to right: name of the source, minimum (L_i) and maximum (L_a) harmonic degrees (when relevant), time scale, radius and scale for the HS prior (see "Analysis step" section), total number of parameters, and coefficient notation used in Eq. (1). Δt is set to 3 months in this study. No R and S values are given for the induced field and its variation. See "Analysis step" section for the construction of the statistics for these latter contributions

Ropp *et al. Earth, Planets and Space* (2020) 72:153

Page 5 of 15

$$\hat{\mathbf{Y}}_{\ell,\ell+1}^{m}(\theta,\varphi) = -a\left(\frac{r}{a}\right)^{\ell+2}\nabla\left(\left(\frac{a}{r}\right)^{\ell+1}Y_{\ell}^{m}(\theta,\varphi)\right), \tag{2}$$

$$\hat{\mathbf{Y}}_{\ell,\ell-1}^{m}(\theta,\varphi) = -a\left(\frac{a}{r}\right)^{\ell-1}\nabla\left(\left(\frac{r}{a}\right)^{\ell}\ Y_{\ell}^{m}(\theta,\varphi)\right), \tag{3}$$

where ∇ is the gradient operator. The $Y_{\ell}^{m}(\theta,\varphi)$ are the Schmidt semi-normalised real spherical harmonics usually employed in geomagnetism. Positive orders ($m \geq 0$) are associated with $\cos(m\theta)$ terms, and negative orders ($m < 0$) are associated with $\sin(|m|\theta)$ terms. The vector harmonics $\hat{\mathbf{Y}}_{\ell,\ell-1}^{m}(\theta_{\mathrm{SM}},\varphi_{\mathrm{SM}})$ (resp. $\hat{\mathbf{Y}}_{\ell,\ell-1}^{m}(\theta_{\mathrm{GSM}},\varphi_{\mathrm{GSM}})$) are vectors in the SM (resp. GSM) system of coordinates.

Modelling method

The models we compute are model distributions that we assume to be Gaussian, a property which is required for the Kalman filter framework. Each distribution is described by the normal probability density function (*pdf*) $\mathcal{N}(\mathbf{m}_k, \mathbf{C}_k)$, where k is the index for the time interval $[t_k : t_{k+1}]$, $\mathbf{m}_k$ is the mean model, and $\mathbf{C}_k$ is the covariance of the model distribution. The mean model $\mathbf{m}_k$ is a N_m-sized vector containing the mean values of the model parameters ($N_m = 2227$, see Table 3) and $\mathbf{C}_k$ is a $N_m \times N_m$ matrix. The Gaussian nature of the distribution is preserved through the whole process, since all operations applied are linear.

Our modelling approach relies on a Kalman filter (Kalman 1960), a 3-step process where data assimilation relies on a correlation-based technique (Holschneider et al. 2016). The first step of the process is to model the magnetic field from a subset of data spanning the time interval $[t_k; t_{k+1}]$, through a re-weighted least square process (hereinafter the analysis step). The model thus obtained is used to predict the model for the next time interval, through an extrapolation of its mean and covariance (hereinafter the prediction step). When the full time series is built, a backward smoothing is applied, in order to constrain each model but the last with information from posterior time intervals (hereinafter the smoothing step). Each one of these three steps is described below.

Analysis step

The data available in the time interval $[t_k, t_{k+1}]$ are gathered in a vector $\mathbf{d}_k$. The model parameters and the data are linked by a linear operator $\mathbf{A}_k$, and the uncertainty on the data is accounted for by an error vector $\mathbf{e}_k$. The elements of $\mathbf{A}_k$ are directly derived from Eq. (1).

To lighten notations, the index k is dropped for $\mathbf{d}$, $\mathbf{A}$ and $\mathbf{e}$ as it does not impede the understanding of the equations. Therefore, the relation between the data and mean model is

$$\mathbf{d} = \mathbf{A}\mathbf{m}_k + \mathbf{e}. \tag{4}$$

The error $\mathbf{e}$ is a zero mean multivariate random variable with a distribution described by a *pdf* $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. $\mathbf{\Sigma}$ is a diagonal matrix, its elements being the data variances given in Table 2.

Equation (4) is solved for the mean model $\mathbf{m}_k$ and covariance $\mathbf{C}_k$ via a re-weighted least-square (RLS) process using Huber weights (Huber 1981). The solution is

$$\mathbf{m}_k = \tilde{\mathbf{m}}_k + \left(\mathbf{A}^t\mathbf{W}_j\mathbf{A} + \tilde{\mathbf{C}}_k^{-1}\right)^{-1}\mathbf{A}^t\mathbf{W}_j(\mathbf{d} - \mathbf{A}\tilde{\mathbf{m}}_k), \tag{5}$$

$$\mathbf{C}_k = \left(\mathbf{A}^t\mathbf{W}_j\mathbf{A} + \tilde{\mathbf{C}}_k^{-1}\right)^{-1}, \tag{6}$$

where the weight matrix is $\mathbf{W}_j = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{U}_j\mathbf{\Sigma}^{-\frac{1}{2}}$, and j denotes the index of the iterations. $\mathbf{W}_j$ is updated at each of the 3 iterative steps of the RLS process. $\mathbf{U}_0 = \mathbf{I_d}$, and $\mathbf{U}_j$ for $j > 0$ is the diagonal matrix for Huber weights. The super-script t denotes the transpose.

In Eqs. (5) and (6), the prior mean model $\tilde{\mathbf{m}}_k$ and its prior covariance $\tilde{\mathbf{C}}_k$ are updated according to the information extracted from the data, to give the posterior mean model $\mathbf{m}_k$ and covariance $\mathbf{C}_k$. This prior model distribution $\mathcal{N}(\tilde{\mathbf{m}}_k, \tilde{\mathbf{C}}_k)$ describes what we know of the model before assimilation of the data. It is, for all time steps but the first, derived from the previous time step posterior model distribution through the prediction process that is described below. However, the prior distribution $\mathcal{N}(\tilde{\mathbf{m}}_0, \tilde{\mathbf{C}}_0)$ for time interval $[t_0 : t_1]$ needs to be defined.

The initial mean prior model is null: $\tilde{\mathbf{m}}_0 = 0$. This means we allow for each parameter to vary around zero, in a range characterised by the variances and covariances given in the matrix $\tilde{\mathbf{C}}_0$.

Regarding its structure, $\tilde{\mathbf{C}}_0$ is a block diagonal matrix, with one block for the core field and SV, and one block for each of the other sources.

Regarding the construction of the different matrix blocks (designated as covariance blocks in the following), two possibilities are investigated. The first possibility is to use the Holschneider et al. (2016) type of prior that consists in information on the energy spectrum of each modelled source. For each contribution, this spectrum is defined by a scaling S, and a radius R. The value of S is set either empirically, or by optimisation, or alternatively by using the actual spectrum of the source if it is known. R is the radius at which the spectrum of the source is flat. The values of R and S are given for all modelled sources in Table 3. The resulting covariance block is diagonal, containing the respective a priori variances of each model parameter. The variances $v_{\ell}^{n_s}$, where n_s refers to the source type (i.e. core,

Ropp *et al. Earth, Planets and Space* (2020) 72:153

Page 6 of 15

SV, etc.) and ℓ is the spherical harmonics degree, are defined as

$$v_\ell^{n_s,int} = S^{n_s}\left(\frac{a}{R^{n_s}}\right)^{2\ell+4}, \tag{7}$$

$$v_\ell^{n_s,ext} = S^{n_s}\left(\frac{R^{n_s}}{a}\right)^{2\ell}, \tag{8}$$

for internal and external sources, respectively.

Note that for a given SH degree ℓ, variances do not depend on the SH order m. The initial variance for observatory offsets is set to $1000\,\mathrm{nT}^2$ for all observatories.

The second possibility is to derive covariance information from a range of parameter samples. We used this approach for the core and induced fields. For the core field and SV, Gauss coefficient samples were obtained from numerical dynamo runs, of the *Coupled Earth* model described by Aubert (2013). Nine thousand samples were taken for each Gauss coefficient, and variance and covariances derived from them. This yields a block containing variances and covariances for both the core field and SV, as well as the cross-covariances between the two sources. For the induced field, we built the prior through an empirical approach. We first built a model where all ${}^{I}g_\ell^m$ and ${}^{I}\dot{g}_\ell^m$ were imposed to take zero values. This resulted in very noisy core and SV time series of Gauss coefficients. These time series were smoothed using a 2-years averaging window, sliding with time. The residuals of this smoothing process were used to derive variances for the ${}^{I}g_\ell^m$ and ${}^{I}\dot{g}_\ell^m$ Gauss coefficients. Given the simplicity of this ad hoc process, no covariance terms were calculated. The corresponding covariance block is therefore diagonal. It should be noted that these matrices concern only the spatial correlations of the core and induced fields, and do not affect their time dependance, which is handled separately (via the prediction and smoothing).

In this work, two series of model distributions were derived. Both series are parameterised in the same way and are using the same data and the same prior, except for the core field and SV. The HS series uses the Holschneider et al. (2016) type of prior information for the core field and its secular variation, whereas the CE series use the statistics derived from the *Coupled Earth* model for these two contributions.

Prediction step

The prediction step defines the prior model distribution $\mathcal{N}(\tilde{\mathbf{m}}_{k+1}, \tilde{\mathbf{C}}_{k+1})$ at time t_{k+1} from the posterior model

distribution $\mathcal{N}(\mathbf{m}_k, \mathbf{C}_k)$ at time t_k. The prediction is based on the assumption that the model is a multivariate random variable with stationary second order statistics. It evolves according to

$$\tilde{\mathbf{m}}_{k+1} = \mathbf{P}\mathbf{m}_k + \mathbf{w}, \tag{9}$$

$$\tilde{\mathbf{C}}_{k+1} = \mathbf{P}\mathbf{C}_k\mathbf{P}^t + \mathbf{C_w}, \tag{10}$$

where $\mathbf{P}$ is the prediction operator. Since this prediction step is not exact, an error term is introduced. This term is also a multivariate random variable normally distributed: $\mathcal{N}(\mathbf{w}, \mathbf{C_w})$. The mean of the error is null: $\mathbf{w} = 0$, but its covariance matrix $\mathbf{C_w}$ depends, as the operator $\mathbf{P}$, on the physics governing the evolution of the different sources. In particular, for all the sources, but the core field and its secular variation, it is assumed that the parameters evolve in time as auto-regressive processes of order 1 (AR1). For such processes a single parameter g with a timescale τ evolves following:

$$g_{k+1} = \alpha g_k + \omega, \tag{11}$$

where α is defined by:

$$\alpha = \begin{cases} e^{-\frac{\Delta t}{\tau}} & \text{for } \tau > \Delta t \\ 0 & \text{for } \tau < \Delta t \end{cases} \tag{12}$$

The timescale τ for the different sources is specified in Table 3. The prediction error ω is a random variable with zero mean. The requirement that the statistics of the parameter g are stationary over time defines the variance v_ω of ω. Assuming the parameter g has a variance v_g constant in time, i.e. v_g is the corresponding diagonal element on $\tilde{\mathbf{C}}_0$, then the variance of ω is $v_\omega = v_g(1 - \alpha^2)$. For external sources, the timescales are smaller than the time step Δt. Therefore, $\alpha = 0$ and the predicted parameter has, as ω, a zero mean and a variance $v_\omega = v_g$. On the contrary, for the lithospheric field, the prediction is strongly dependent on the precedent state, $\alpha \simeq 1$ and the variance of ω is $v_\omega \simeq 0$.

For the core field and secular variation, the evolution of a given coefficient is given through the coupled set of equations:

$$g_{\ell,k+1}^m = \alpha_\ell\left(g_{\ell,k}^m + \Delta t\,\dot{g}_{\ell,k}^m\right) + \omega_c, \tag{13}$$

$$\dot{g}_{\ell,k+1}^m = \dot{\alpha}_\ell\,\dot{g}_{\ell,k}^m + \dot{\omega}_c, \tag{14}$$

where ω_c and $\dot{\omega}_c$ are the errors. These errors have, to the first order, amplitudes proportional to the coefficient accelerations $\ddot{g}_\ell^m$. As in the case of AR1 processes, we assume that the errors have zero means, and we set $\dot{\omega}_c$ (resp. ω_c) variances to $v_{\dot{\omega}_c} = v_{\dot{g}_\ell^m}(\frac{\Delta t}{\tau_{sa}})^2$ (resp.

Ropp *et al. Earth, Planets and Space*   (2020) 72:153

Page 7 of 15

$v_{\omega_c} = v_{\dot{g}_\ell^m}(\frac{\Delta t}{\tilde{\tau}_{sa}})^2(\frac{\Delta t}{2})^2)$, where $\tilde{\tau}_{sa}$ is the timescale for acceleration. The value of $\tilde{\tau}_{sa}$ is known to be around 11 years up to spherical harmonics degree 13 (Christensen et al. 2012), and we used this same value up to degree 18. $v_{\dot{g}_\ell^m}$ is the known variance of the secular variation Gauss coefficient $\dot{g}_\ell^m$—i.e. a diagonal element of the $\tilde{\mathbf{C}}_0$ matrix. From there, the stationary hypothesis defines the α_ℓ and $\dot{\alpha}_\ell$:

$$\alpha_\ell = \sqrt{\frac{v_{g_\ell^m} - v_{\dot{g}_\ell^m}(\frac{\Delta t}{\tilde{\tau}_{sa}})^2(\frac{\Delta t}{2})^2}{v_{g_\ell^m} + v_{\dot{g}_\ell^m}\Delta t^2}} \simeq \sqrt{\frac{1 - \frac{1}{4}(\frac{\Delta t}{\tilde{\tau}_{sa}})^2(\frac{\Delta t}{\tilde{\tau}_{sv}})^2}{1 + (\frac{\Delta t}{\tilde{\tau}_{sv}})^2}}$$

$$\dot{\alpha}_\ell = \sqrt{1 - (\frac{\Delta t}{\tilde{\tau}_{sa}})^2}, \qquad (15)$$

where $v_{g_\ell^m}$ is the variance of the core field Gauss coefficient g_ℓ^m as estimated in $\tilde{\mathbf{C}}_0$. $\tilde{\tau}_{sv}$ is the timescale for the secular variation. Note that we used in this work the definition of α_ℓ that does not involve $\tilde{\tau}_{sv}$. As an indication, $\tilde{\tau}_{sv} \simeq 415/\ell$ is in the range of acceptable values for Chaos [see Christensen et al. (2012)]. In Additional file 1: Appendix S1, more details are given on the way the α, α_ℓ and $\dot{\alpha}_\ell$ values are derived. The exact construction of the matrices $\mathbf{P}$ and $\mathbf{C_w}$ is also detailed.

Smoothing

The result of the process presented above is a time series of model distributions $\mathcal{N}(\mathbf{m}_k, \mathbf{C}_k)$ for $k = 0, \ldots, N_t$. The final model distributions are smoothed versions of this time series. The smoothing consists in re-computing the model at a time step k by using the information provided by the data analysed at every time step $k' > k$. The smoothed model distributions are identified with a upper script ^{s}: $\mathcal{N}(\mathbf{m}_k^s, \mathbf{C}_k^s)$. This smoothing is achieved through the following equations (Anderson and Moore (1979), Rauch et al. (1965)) for the mean and covariance of the smoothed model at time t_k, respectively:

$$\mathbf{m}_k^s = \mathbf{m}_k + \mathbf{G}_k\left(\mathbf{m}_{k+1}^s - \mathbf{Pm}_k\right), \qquad (16)$$

$$\mathbf{C}_k^s = \mathbf{C}_k - \mathbf{G}_k\left(\mathbf{PC}_k\mathbf{P}^t + \mathbf{C_{w}}_k - \mathbf{C}_{k+1}^s\right)\mathbf{G}_k^t, \qquad (17)$$

where the matrix $\mathbf{G}_k$ is defined by

$$\mathbf{G}_k = \mathbf{C}_k\mathbf{P}^t\left(\mathbf{PC}_k\mathbf{P}^t + \mathbf{C_w}\right)^{-1} \qquad (18)$$

and $\mathbf{P}$ and $\mathbf{C_w}$ are, respectively, the prediction operator and covariance matrix introduced in the previous section. These equations are similar to Eqs. (5) and (6), as they give the solution $\mathbf{m}_k^s$ of the inverse problem set by Eq. (9), where $\mathbf{m}_{k+1}^s$ defines the data.
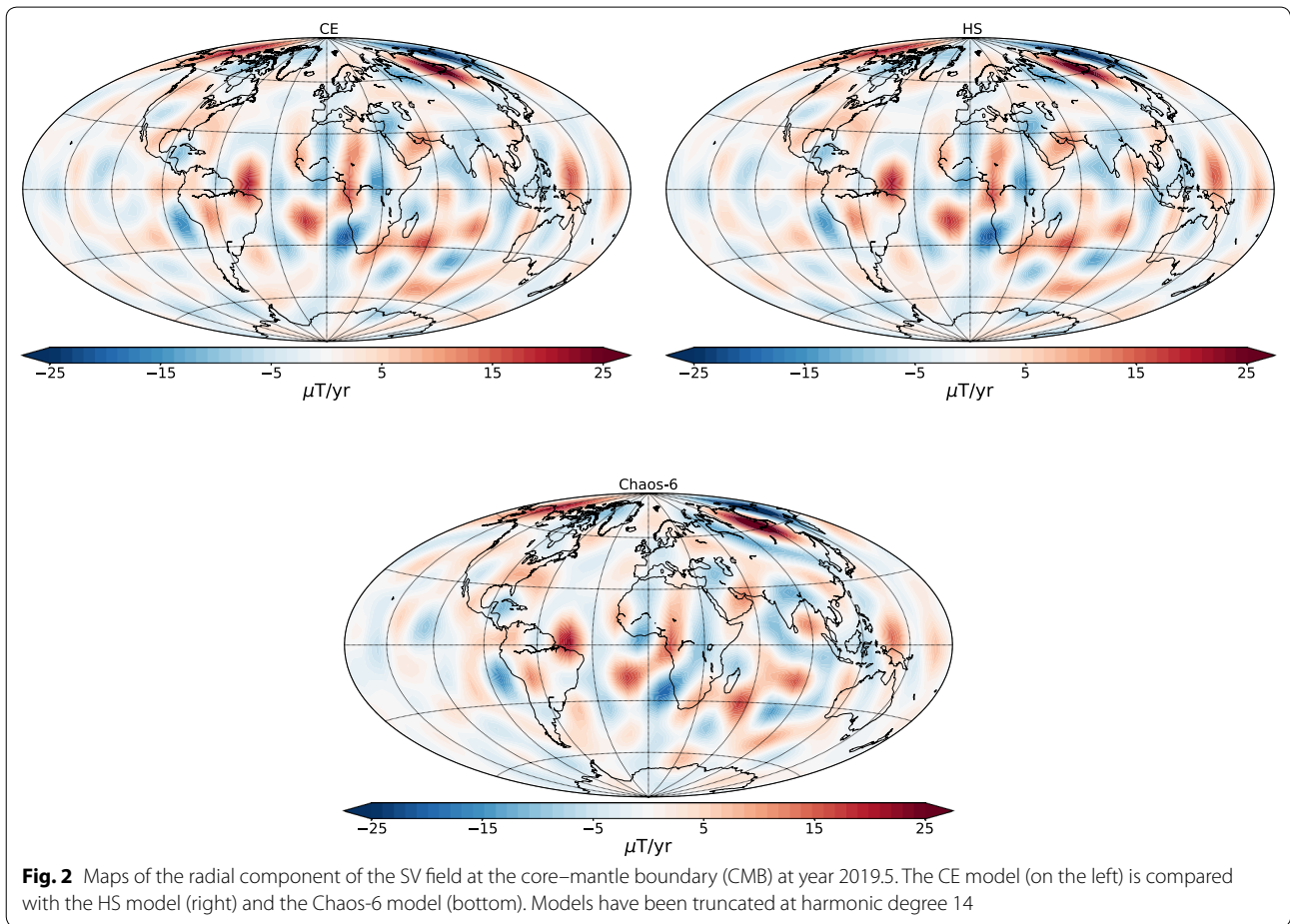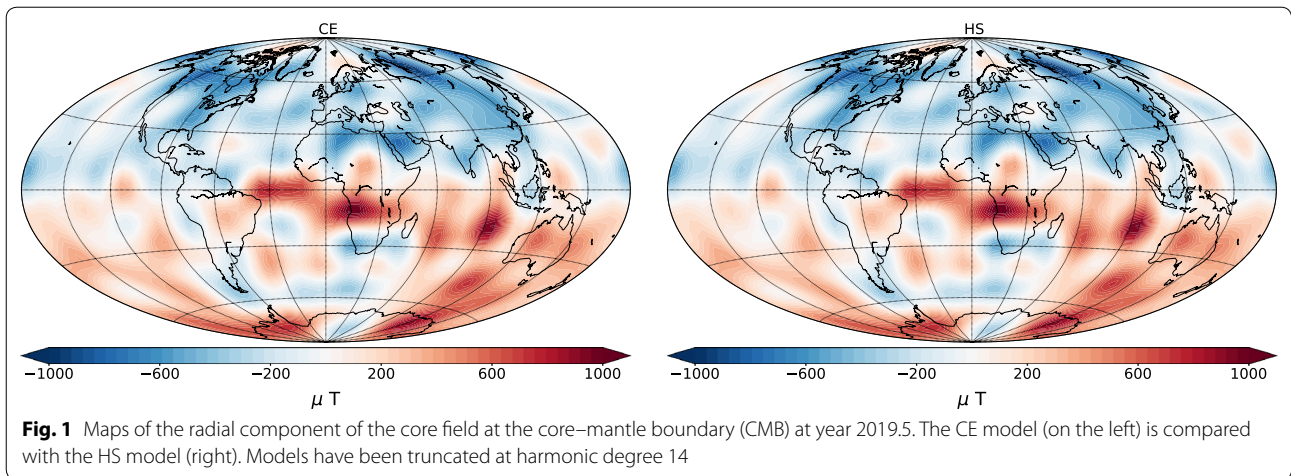
Results

We present the results of the process described above applied to our data set. We recall that the full output is described by a series of mean models and covariances, each pair defining a normal distribution of models. Strictly speaking, a single model would have to be randomly drawn from this ensemble. Here, we systematically use the series of mean models for our representations. When specified, the mean models are presented with a $\pm 2\sigma$ wide error range. We recall also that two series of model distributions have been derived, the HS and CE series, that differ only by the prior information used for the core field and secular variation statistics. Through this section, our results are compared with the Chaos-6 model (Finlay et al. 2016) in its version Chaos-6-×9.

For each time interval $[t_k, t_{k+1}]$ the normalised misfit to the data is defined by:

$$\mathcal{R}_k = \sqrt{\frac{1}{N_d}(\mathbf{d} - \mathbf{Am}_k)^t\mathbf{W}_i(\mathbf{d} - \mathbf{Am}_k)}, \qquad (19)$$
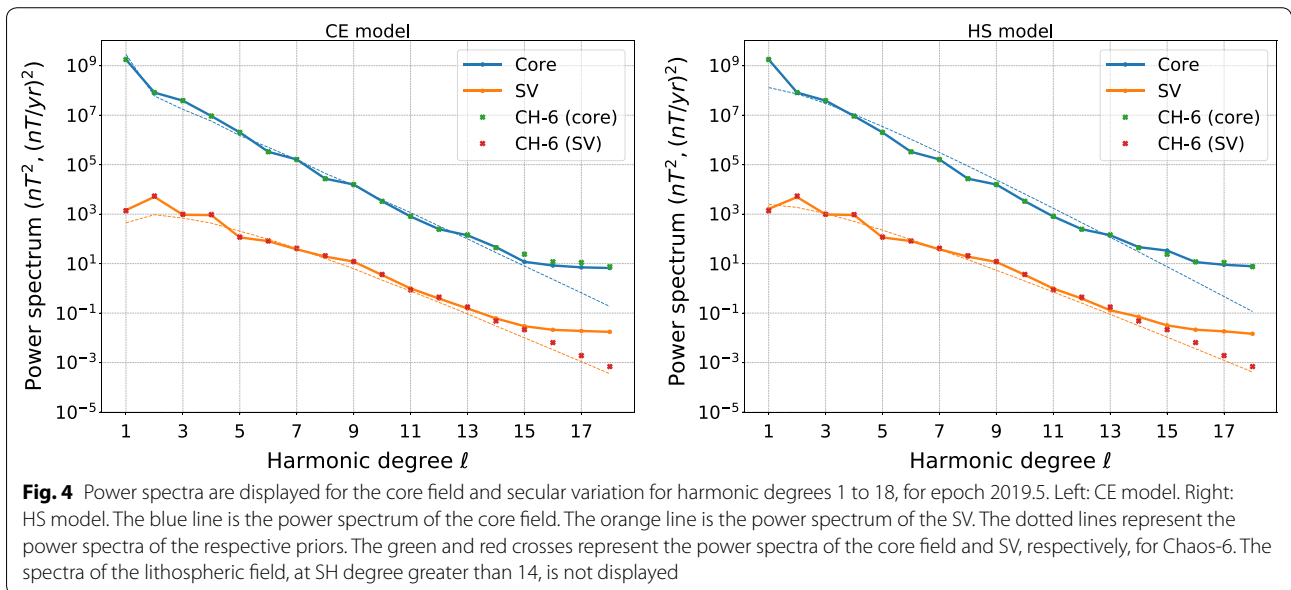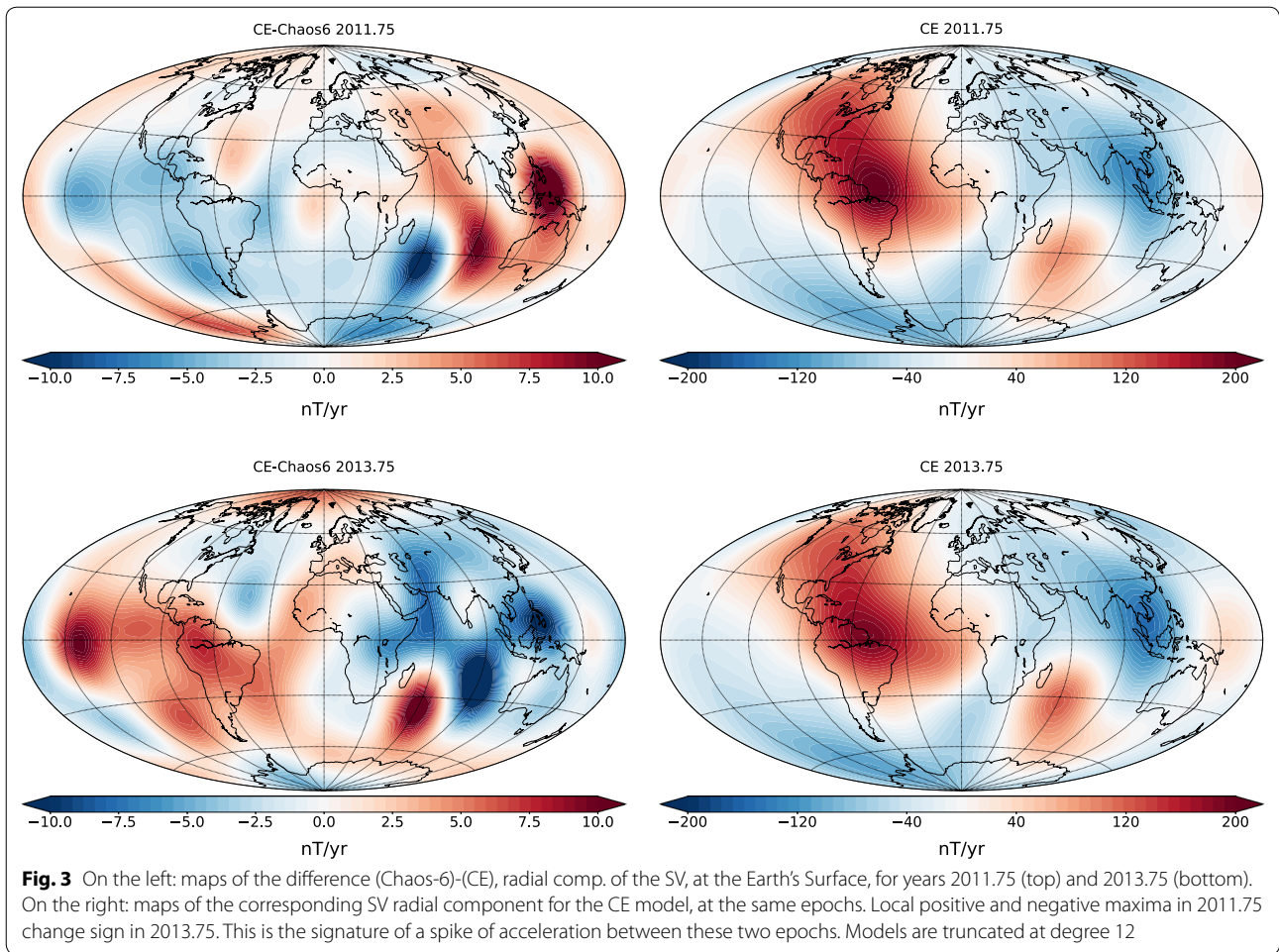
where $\mathbf{W}_i$ is defined in "Analysis step" section. The values of $\mathcal{R}_k$ are in the range $[0.98, 1.92]$ nT for both CE and HS models, depending essentially on the type of data available during this time interval. Over the whole time period, the mean of the misfits for both models is 1.44 nT. It shows that our models achieve a fairly good fit to the data. These values are estimated after the analysis process, but before the smoothing steps. The misfit after smoothing is not reported here as its estimation requires adjusting other contributions to the geomagnetic field such as external and induced fields.

In Fig. 1, the CE and HS model core field radial components are plotted at the core–mantle boundary (CMB) for year 2019.5. The field models are truncated at degree 14. Figure 2 displays the radial components of the SV of the same two models at the CMB. The Chaos-6 SV model is also displayed. All three are truncated at degree 14. Both CE and HS models are quite consistent with Chaos-6, although some differences appear in small-scale features, over the Pacific or Indian oceans, for example. Some noticeable and interesting differences can be highlighted at other time periods. Concerning the core field, the largest differences occur at the poles, where they rarely exceed ± 10nT at the Earth's surface, when satellite data are available. Secular variation differences occur mostly during years 2010 to 2014. They are characterised by localised maxima in the Indian Ocean, as shown in Fig. 3. These maxima change signs between 2011.75 and 2013.75 suggesting a spike of acceleration in the HS/CE models during this period of time. Otherwise, SV radial

Ropp *et al. Earth, Planets and Space*   (2020) 72:153

Page 8 of 15

Fig. 1 Maps of the radial component of the core field at the core–mantle boundary (CMB) at year 2019.5. The CE model (on the left) is compared with the HS model (right). Models have been truncated at harmonic degree 14

Fig. 2 Maps of the radial component of the SV field at the core–mantle boundary (CMB) at year 2019.5. The CE model (on the left) is compared with the HS model (right) and the Chaos-6 model (bottom). Models have been truncated at harmonic degree 14

component differences are mostly contained inside a ±5nT/year interval. The maps shown in Fig. 3 do not include the HS model, for which the results described above are also valid. All models are truncated at degree 12.

Figure 4 displays the spherical harmonics power spectra of the calculated core field and SV (HS on the left, CE on the right) for year 2019.5. The crosses picture the Chaos-6 power spectra, and the spectra of the prior are displayed in dotted lines. These spectra show the

Fig. 3 On the left: maps of the difference (Chaos-6)-(CE), radial comp. of the SV, at the Earth's Surface, for years 2011.75 (top) and 2013.75 (bottom). On the right: maps of the corresponding SV radial component for the CE model, at the same epochs. Local positive and negative maxima in 2011.75 change sign in 2013.75. This is the signature of a spike of acceleration between these two epochs. Models are truncated at degree 12

Fig. 4 Power spectra are displayed for the core field and secular variation for harmonic degrees 1 to 18, for epoch 2019.5. Left: CE model. Right: HS model. The blue line is the power spectrum of the core field. The orange line is the power spectrum of the SV. The dotted lines represent the power spectra of the respective priors. The green and red crosses represent the power spectra of the core field and SV, respectively, for Chaos-6. The spectra of the lithospheric field, at SH degree greater than 14, is not displayed

Ropp *et al. Earth, Planets and Space* (2020) 72:153

Page 10 of 15

general agreement of both HS and CE model core field with those obtained by other modellers, but they have too much energy at small scales compared to the prior (from degree 14 for the SV, degree 15 for the core field). As a simple, first analysis, this suggests either that our priors are poor—e.g. too small for some Gauss coefficients of the lithospheric field, or that there are some un-modelled signals in the data at these wavelengths. This excess of energy probably results from a combination of the two hypotheses. To discriminate between them would require a more thorough analysis of the results, focusing on the output distributions of the core field Gauss coefficients, their variances and possible covariances with other contributions. The conclusions of this analysis are not yet completely clear. However, it suggests that sources (e.g. small-scale induced signals and high latitude ionospheric contributions) are missing in our description of the observed signal. In favour of the first hypothesis, the CE distribution of models is closer to the prior at earlier epochs (not shown) and keeps closer to the prior at later epochs. This highlights the importance of cross-correlations, which are absent from the HS prior information.

In Fig. 5, we compare the time series for several SH coefficients of the secular variation models before and after the Kalman smoothing step (see the process description in "Modelling method" section). Results are shown for both CE and HS secular variation models. The non-smoothed time series present a time lag, as compared to the smoothed and Chaos-6 time series, which agree on the general trend of the variation for most large-scale coefficients. This lag results from the difficulty to resolve the secular variation within a single time step, using mainly information from the past—thanks to the prediction step, but very little information from the future, as only 3 months worth of data are used. It leads to an underestimation of the secular variation when the acceleration is positive, and an overestimation when the acceleration is negative. The smoothing step clearly alleviates this time lag.

Figure 6 displays the time series of CE and HS secular variations compared with Chaos-6 for some low SH degrees. The variations are overall compatible with Chaos-6. Major differences can appear during the time period 2010–2014, because of the lack of satellite data combined with radically different ways of handling ground observatory data. The bottom part of the figure displays the Fourier Transform (FT) spectra of the time series to which a linear trend has been first removed to avoid spectral leakage.
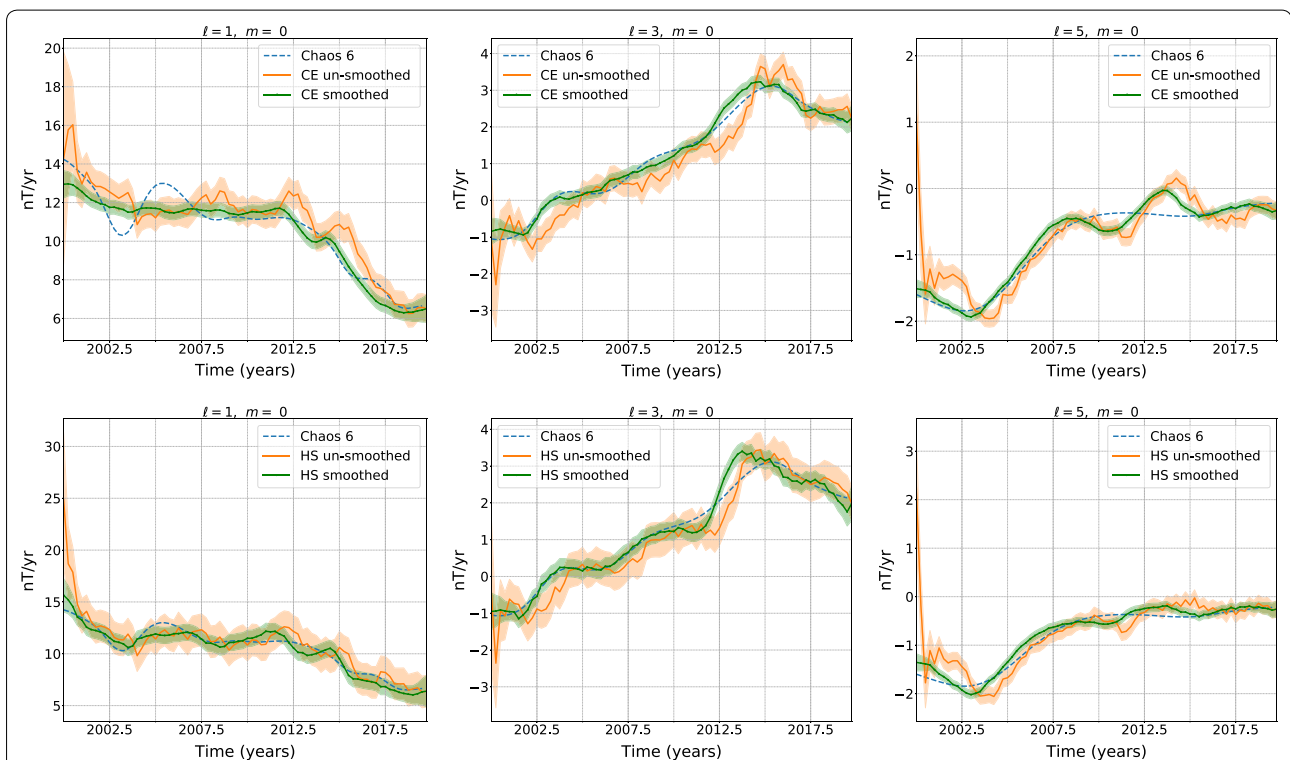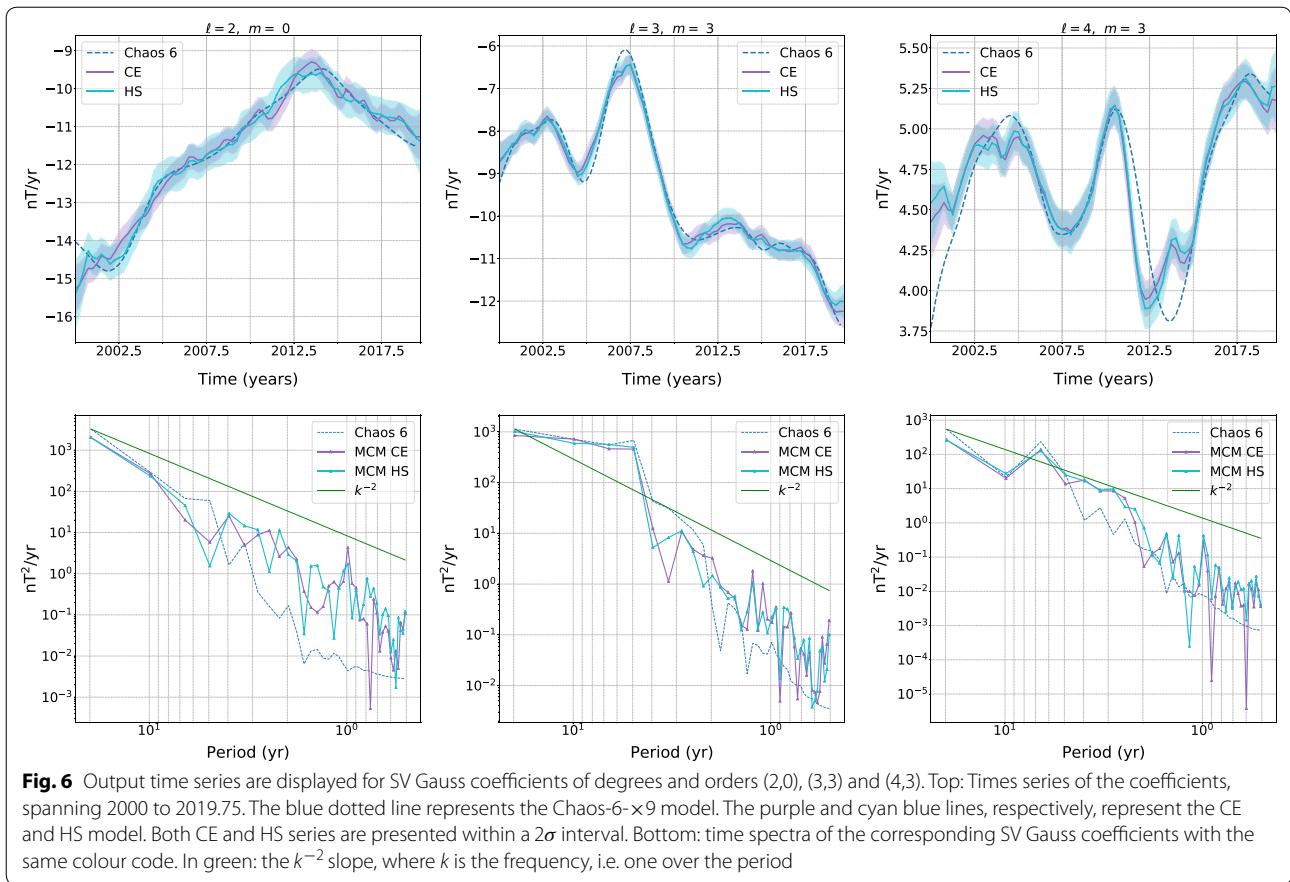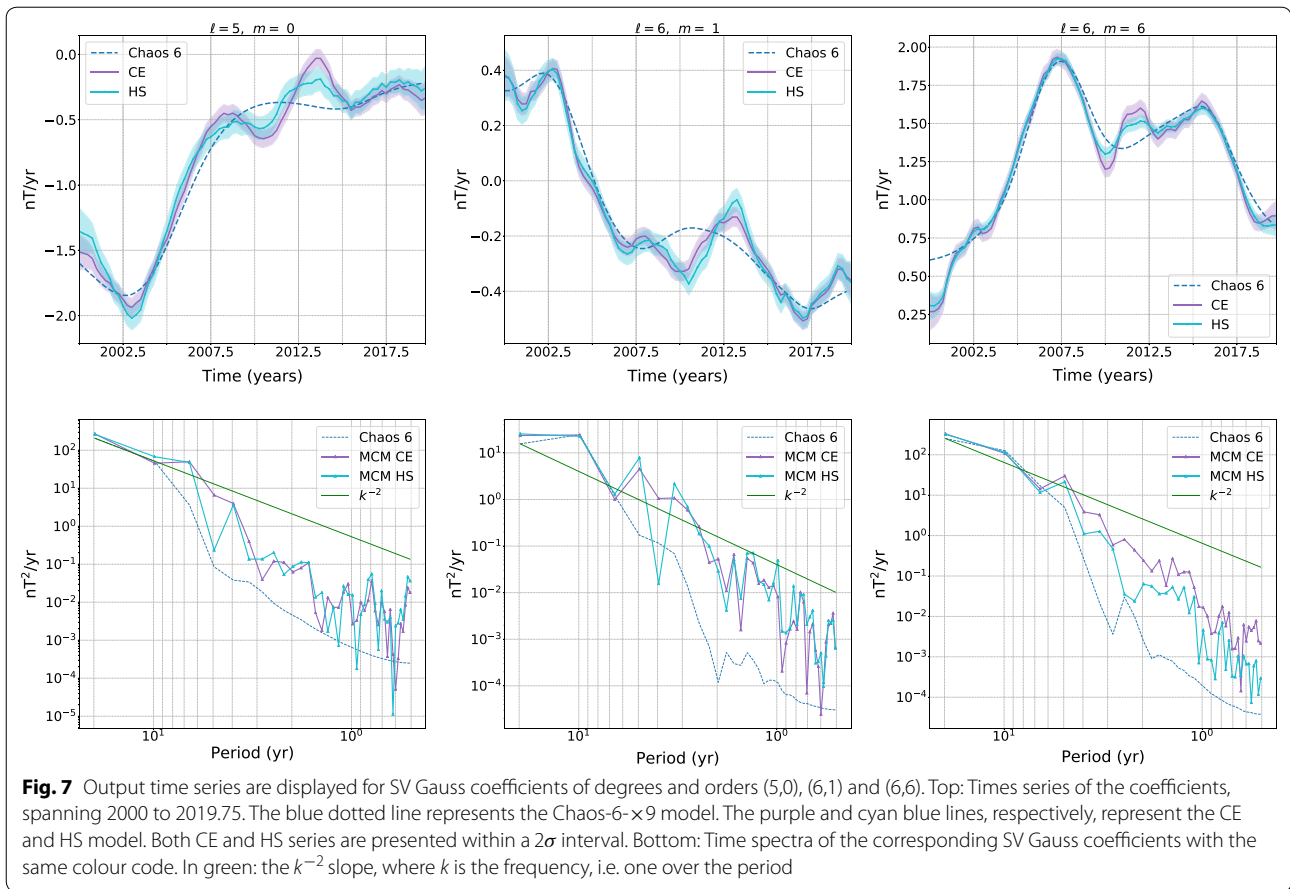

Fig. 5 Times series for SV Gauss coefficients of degrees 1, 3, 5 and order 0. The times series span 2000 to 2019.75. The dotted blue lines represent the Chaos-6-×9 model. The orange line represents the MCM model before the smoothing operation. The green line represents the final model, after smoothing. Both series are presented within a 2σ interval. Top: CE series. Bottom: HS series

Ropp *et al. Earth, Planets and Space*   (2020) 72:153

Page 11 of 15

Fig. 6 Output time series are displayed for SV Gauss coefficients of degrees and orders (2,0), (3,3) and (4,3). Top: Times series of the coefficients, spanning 2000 to 2019.75. The blue dotted line represents the Chaos-6-×9 model. The purple and cyan blue lines, respectively, represent the CE and HS model. Both CE and HS series are presented within a 2σ interval. Bottom: time spectra of the corresponding SV Gauss coefficients with the same colour code. In green: the k^{-2} slope, where k is the frequency, i.e. one over the period

Coefficients time series (and their Fourier spectra) are displayed for higher SH degrees (i.e. lower spatial scales), in Figs. 7 and 8. At these spatial scales, the SV models sometimes present very different behaviours. The Fourier spectra of several Gauss coefficients show significant periodicities of 4 to 6 years (e.g. for $\dot{g}_6^1$ and $\dot{g}_6^6$ at 5 years). Higher degree Gauss coefficients show various peaks in their Fourier spectra, such as $\dot{g}_7^0$, which presents very strong variations between 4 and 10 years, or $\dot{g}_8^1$ with peaks at 3 and 4 years. Chaos-6 is systematically and strongly smoothed, with little energy for periods under 5 years. This most certainly results from the temporal smoothing used in the process leading to this latter model. It contrasts with the approach used here where the time scales controlling the temporal behaviour of the core field are set to realistic values. However, although our HS and CE models present more energy at periods under 5 years, they are not free of anomalous features. The high variability of the Fourier spectra at small time scales is probably a signature of noise (or unidentified contributions) affecting the models. For example, the periodicity observed in some coefficients (e.g. $\dot{g}_7^0$ in Fig. 8) might be explained by an incorrect separation of some ionospheric fields with the core field. This

is a significant risk, considering that satellites see ionospheric fields as internal sources. These ionospheric fields could be co-estimated in the future, using a suitable parametrisation and proper prior information. Furthermore the ad hoc way in which we handled the induced fields has smoothed out short time variations of the core field Gauss coefficients, but did not allow a clear separation of core and induced field signals for periodicity longer than 2 years. To progress in this matter, it should be assessed whether the amplitude of the estimated induced fields is physically plausible, considering the mantle conductivity and external fields amplitudes. In particular, harmonics of the 11 years solar cycle (e.g. 5 years periodicity) are often more pronounced in our models than, e.g. in CHAOS-6. This is visible, for example, in the ${}^C\dot{g}_1^0$ time series of the HS model, but not in the CE model or the Chaos-6 model.

To test the coherency of our models, we compared the modelled SV time series (HS and CE SV) to the variations of the modelled core field time series. For this purpose, we computed a SV series by finite differences of the mean core field model series (referred to as the FD SV). The coherency between both the CE and HS SV series and the FD SV series was verified by evaluating the power

Ropp *et al. Earth, Planets and Space* (2020) 72:153

Page 12 of 15

Fig. 7 Output time series are displayed for SV Gauss coefficients of degrees and orders (5,0), (6,1) and (6,6). Top: Times series of the coefficients, spanning 2000 to 2019.75. The blue dotted line represents the Chaos-6-×9 model. The purple and cyan blue lines, respectively, represent the CE and HS model. Both CE and HS series are presented within a 2σ interval. Bottom: Time spectra of the corresponding SV Gauss coefficients with the same colour code. In green: the k^{-2} slope, where k is the frequency, i.e. one over the period

spectra of their respective differences. The residuals of this comparison have a total energy of less than 0.03% of the total energy of the SV spectrum, for both CE and HS models. The difference between the FD SV and the CE or HS SV is most important where the secular variation presents steep slopes.

Finally, we would like to point out that, because there is no smooth parametrisation in time—such as splines would provide—in our model, the core field secular acceleration cannot be derived in a robust way directly from our mean SV Gauss coefficient series. If, for further analysis, a core field acceleration model is required, it is recommended to pick a SV Gauss coefficient series in the model distribution such that the derived acceleration varies smoothly in time, rather than using the mean SV.

Discussion

The results presented in the previous section show that the core field and secular variation general features are well recovered through the process presented. Some difficulties have been identified, e.g. extracting information on the secular acceleration requires further processing of the derived series of snapshots. There are nonetheless technical advantages in using a sequential modelling approach over more classical methods, such as, for example, the relatively light computing power requirements. However, the main motivation for applying this method is first to extract from the data better models through the information provided by the prior, and second to have, as an output, reliable estimates of the model uncertainties. We discuss in the following these two points.

The spatial prior information for a time interval $[t_k, t_{k+1}]$ is provided to the model through the covariance matrix $\tilde{\mathbf{C}}_k$ in Eq. (5). This matrix cannot be singular because it would imply that some parts of the model are perfectly known, an hypothesis that we reject by imposing a minimum variance to all eigenvectors of the matrix. However, in the same equation the matrix $\mathbf{A}^t \mathbf{W}_j \mathbf{A}$ can be singular when the data set cannot resolve the model and the spatial prior information has an important role in this case. The prior information on the temporal behaviour of the model components intervene in the prediction part of the Kalman filter through the definition of the α_ℓ and $\dot{\alpha}_\ell$ in Eq. (9). They influence the output of the model as soon as they differ from unity. It follows that in our core field model, the prior information plays an important

Fig. 8 Output time series are displayed for SV Gauss coefficients of degrees and orders (7,0), (8,1) and (8,8). Top: Times series of the coefficients, spanning 2000 to 2019.75. The blue dotted line represents the Chaos-6-×9 model. The purple and cyan blue lines, respectively, represent the CE and HS model. Both CE and HS series are presented within a 2σ interval. Bottom: time spectra of the corresponding SV Gauss coefficients with the same colour code. In green: the k^{-2} slope, where k is the frequency, i.e. one over the period

role in the definition of the small scales of the secular variation, and in the separation of core and induced field contributions.

We point out that the HS and CE models differ only by the spatial prior information. It is therefore the spatial prior information that leads to different temporal behaviour for the HS and CE Gauss coefficients of the secular variation, for intermediate-to-high SH degrees as shown in Figs. 7 and 8. In contrast, both the spatial and temporal prior information influence the deviations of the Chaos-6 SV model from our model series. At these spatial scales, and for a given SH degree, the HS and CE prior variances are very similar. The observed differences are therefore mainly due to the covariance between coefficients of different degree and order. These covariances are present only in the initial prior of the CE model. In particular, covariances between well resolved large scales and poorly defined small scales can be responsible for observed differences. As long as the information carried by the *Coupled Earth* dynamo model outputs is correct, there is no reason to believe less in the variations of the SV Gauss coefficient distributions of our CE models than in those of our HS or Chaos-6 models. Of course, the spread of

the model distributions, characterised by the variance values, plays an important role, and it is worth studying what influences these values. The example of the induced fields is here particularly instructive.

The first point to notice is that if the contribution of the induced field is neglected, the output variances of the core field distributions are particularly small. In that case, of course, the induced field signals are partly described by the core field model, the noise, and possibly spread over other components of the model. The prior information we can provide to separate them from the core field is extremely limited, particularly regarding their low frequency components. We used here the fact that they are potential fields of internal origins, with small amplitudes. This information is not sufficient to separate them from the core contributions, and when co-estimated, the output variances of the latter increase considerably. However, we also assumed that the induced fields over a 3-months period are uncorrelated to the induced field of the next time period (whereas the core field components are strongly correlated in time). It is this characteristic, although not entirely valid, that allows an acceptable separation of the two contributions

Ropp *et al. Earth, Planets and Space*   (2020) 72:153

Page 14 of 15

in our model. In principle, it should be possible to use our knowledge of the deep mantle conductivity and the external fields behaviours to improve the separation.

Through this example, it is clear that a contribution which is not described in the model "leaks" inside other modelled contributions. These model components may then present spurious behaviours and small variances. As another example, the lithospheric field contribution at SH degrees lower than 15 cannot be distinguished from the static core field, and trying to model it would again increase the variance of the core field model. In contrast, the tidal signals separate well from the other components because of their well defined periodicities in time. When un-modelled they are likely to remain in the noise, whereas trying to co-estimate them should not increase the variances of the core field model. Overall, the core field model we obtained is probably not completely free of induced field contributions, it also certainly describes the small SH degree lithospheric field, some ionospheric fields, and possibly unexpected other minor, more or less static contributions. Nonetheless, while improvements are possible, our field model provides a good description of the main field, together with an estimation of its variance, that is probably slightly under-estimated.

Candidate models to the IGRF

The models presented here have contributed to set two independent contributions to the IGRF-13: the IPGP candidate to the IGRF main field 2020, and the Japanese team candidate to the IGRF predictive secular variation (Minami et al. 2020). In both cases, the data selection criteria were those described in the first section, but Swarm-B satellite data were not used.

Regarding the contribution of the Japanese candidate, the model is nearly the same as the HS model presented here with a data set including data only up to 2019.5. The scaling and reference radius used for the different covariance matrices were those of Table 3. The obtained core field and secular variation components of the model were then used as input data to the En4dVar assimilation process that led ultimately to a prediction of the mean SV over 2020–2025—see Minami et al. (2020) for details on this assimilation process.

For the IPGP main field model, the general scheme for deriving the candidate is the same as the CE model presented here, but with a data set reduced to 2013.5–2019.5 and a Kalman time step set to a year. In total only 6 time steps had to be done. This length of the Kalman step insures robust estimations of the secular variation and reduces significantly the contribution of the induced field components that required to be modelled. The scaling and radius parameter were the same as in Table 3 except for the $^\mathrm{I}g_\ell^m$ and $^\mathrm{I}\dot{g}_\ell^m$. Their radii R were set to 2200 km and

their S values to $1 \cdot 10^{-3}$. From the main field model and its SV derived this way for 2019.0, the main field was linearly extrapolated to 2020.0 giving a main field candidate not much different from other candidates, although with generally slightly lesser energy.

Conclusion

In this paper, we have presented a sequential approach to core field modelling based on a combination of a Kalman filter and a correlation-based modelling method. We aim at modelling separately most major sources contributing to the observed Geomagnetic field. The separation of these different contributions relies on a strong prior information on their spatial and temporal behaviours. We built a sequence of snapshot models constituting a time series that spans 2000.0 to 2019.75, using data from 2000.0 to 2020.0.

Our model time series present mean values that are generally in agreement with recent, reliable models such as the Chaos-6 model (Finlay et al. 2016).

Nonetheless, the results suggest that more temporal variability exists in the small spatial scales of the core field compared to what is shown by classic modelling techniques. In particular, several Gauss coefficient time series present significant periodicities at time scales ranging from 3 to 10 years that are absent in the Chaos-6 time series. This technique offers the further advantage to give reasonably good estimates of the Gauss coefficients variances, provided that the separation of the different sources is appropriately handled. Our main field and SV models, as it is, probably yield slightly under-estimated variances for their Gauss coefficients.

The idea behind the method presented in this paper is that, starting from a prior information on the behaviour of the different sources contributing to the magnetic field, we seek to improve this information through the analysis of data. This knowledge comes in the posterior model distribution, via the Gauss coefficients mean values and variances, that are reduced through the process. This can come, however, at the cost of increasing covariances between model coefficients, due to the incomplete separation of sources. The parametrisation of the modelled sources, and the tuning of their associated prior information is therefore the backbone of this technique. As further work is achieved for this purpose, the produced models should provide more precise and reliable information on the dynamics of the core field. The setup described in this paper, used with a different time step, has allowed the production of a candidate model for the IGRF-13.

Ropp *et al. Earth, Planets and Space*   (2020) 72:153

Page 15 of 15

Supplementary information

Additional file 1: Appendix S1. This appendix gives details on the construction of the projection operator and of the error covariance matrix used in the Kalman filter. These are defined in section Prediction Step.

Abbreviations

SV: secular variation; IGRF: International Geomagnetic Reference Field; IGRF-13: thirteenth edition of the IGRF; ML: medium-to-low latitudes; HL: high latitudes; IMF: interplanetary magnetic field; NEC: North East Centre (reference frame); SM: solar magnetic (reference frame); SH: spherical harmonics; GSM: geocentric solar magnetic (reference frame); HS: series of models using the Holschneider et al. (2016) type of prior for the core and SV fields; CE: series of models using the prior derived from the *Coupled Earth* model; AR1: auto-regressive (process) of order 1; CMB: core–mantle boundary; FD SV: secular variation computed from the core field series by finite differences; IPGP: Institut de Physique du Globe de Paris.

Authors' contributions

GR and VL wrote the manuscript and produced the models presented here. VL and MH coordinated the study and designed the original method used. JB contributed to the theoretical development and implementation of the Kalman filter. All authors participated in the discussion and commented on the manuscript. All authors read and approved the final manuscript.

Data availability statement

Observatory data are available at ftp://ftp.nerc-murchison.ac.uk/geomag/Swarm/AUX_OBS/. Champ data are available on the GFZ data center at https://isdc.gfz-potsdam.de/champ-isdc/access-to-the-champ-data/. Swarm data are available on the ESA data center at https://earth.esa.int/web/guest/swarm/data-access.
The coefficients of the models presented are available upon request.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] Université de Paris, Institut de physique du globe de Paris, CNRS, 75005 Paris, France. [2] Institute for Mathematics, University of Potsdam, Potsdam , Germany.

References

Anderson BD, Moore JB (1979) Optimal filtering. Information and system sciences series. Prentice-Hall, Englewood Cliffs

Aubert J (2013) Flow thoughout the earth's core inverted from geomagnetic observations and numerical dynamo models. Geophys Int. https://doi.org/10.1093/gji/ggs051

Baerenzung J, Holschneider M, Wicht J, Sanchez S, Lesur V (2018) Modeling and predicting the short term evolution of the geomagnetic field. J Geophys Res Solid Earth. https://doi.org/10.1029/2017JB015115

Barrois O, Hammer M, Finlay C, Martin Y, Gillet N (2018) Assimilation of ground and satellite magnetic measurements: inference of core surface magnetic and velocity field changes. Geophys J Int 215(1):695–712

Beggan CD, Whaler KA (2009) Forecasting change of the magnetic field using core surface flows and ensemble kalman filtering. Geophys. Res. Lett. 36(18). https://doi.org/10.1029/2009GL039927

Christensen U, Wardinski I, Lesur V (2012) Time scales of geomagnetic secular acceleration in satellite field models and geodynamo models. Geophys J Int 190:243–254. https://doi.org/10.1111/j.1365-246X.2012.05508.x

Chulliat A, Maus S (2014) Geomagnetic secular acceleration, jerks, and localized standing wave at the core surface from, 2000 to 2010. J Geophys Res (Solid Earth). https://doi.org/10.1002/2013JB010604

Finlay CC, Olsen N, Kotsiaros S, Gillet N, Tøffner-Clausen L (2016) Recent geomagnetic secular variation from swarm and ground observatories as estimated in the CHAOS-6 geomagnetic field model. Earth Planets Space 68:112. https://doi.org/10.1186/s40623-016-0486-1

Gillet N, Barrois O, Finlay CC (2015) Stochastic forecasting of the geomagnetic field from the cov-obs.x1 geomagnetic field model, and candidate models for igrf-12. Earth Planets Space 67(1):1–14

Holschneider M, Lesur V, Mauerberger S, Baerenzung J (2016) Correlation based modelling and separation of geomagnetic field components. J Geophys Res Solid Earth 121:3142–3160. https://doi.org/10.1002/2015JB012629

Huber PJ (1981) Robust statistics. Wiley, New York

Kalman RE (1960) A new approach to linear filtering and prediction problems. J Basic Eng 82:35–45. https://doi.org/10.1115/1.3662552

Lesur V, Wardinski I, Rother M, Mandea M (2008) GRIMM—The GFZ reference internal magnetic model based on vector satellite and observatory data. Geophys J Int 173(2):382–394. https://doi.org/10.1111/j.1365-246X.2008.03724.x

Lesur V, Rother M, Vervelidou F, Hamoudi M, Thébault E (2013) Post-processing scheme for modeling the lithospheric magnetic field. Solid Earth 4:105–118. https://doi.org/10.5194/sed-4-105-2013

Lesur V, Whaler K, Wardinski I (2015) Are geomagnetic data consistent with stably stratified flow at the core-mantle boundary? Geophys J Int 201(2):929–946. https://doi.org/10.1093/gji/ggv031

Lesur V, Wardinski I, Baerenzung J, Holschneider M (2017) On the frequency spectra of the core magnetic field Gauss coefficients. Phys Earth Plan Int. https://doi.org/10.1016/j.pepi.2017.05.017

Lesur V, Rother M, Wardinski I, Schachtschneider R, Hamoudi M, Chambodut A (June 2015) Parent magnetic field models for the igrf-12 gfz-candidates. *Earth, Planets and Space*, 67(87), https://doi.org/10.1186/s40623-015-0239-6

Macmillan S, Olsen N (2013) Observatory data and the swarm mission. Earth Planets Space 65(11):1355–1362. https://doi.org/10.5047/eps.2013.07.011

Maus S, Weidelt P (2004) Separating the magnetospheric disturbance magnetic field into external and transient internal contributions using a 1D conductivity model of the Earth. Geophys Res Lett 31:L12614. https://doi.org/10.1029/2004GL020232

Minami T, Nakano S, Lesur V, Takahashi F, Matsushima M, Shimizu H, Nakashima R, Taniguchi H, Toh H (2020) A candidate secular variation model for igrf-13 based on mhd dynamo simulation and data assimilation, en4dvar. Earth Planets Space. https://doi.org/10.1186/s40623-020-01253-8

Rauch HE, Tung F, Striebel CT (1965) Maximum likelihood estimates of linear dynamic systems. AIAA J 3(8):1445–1450. https://doi.org/10.2514/3.3166

Thébault et al (2015) International geomagnetic reference field: the twelfth generation. Earth Planets Space 67:79. https://doi.org/10.1186/s40623-015-0228-9

Thomson AWP, Lesur V (2007) An improved geomagnetic data selection algorithm for global geomagnetic field modelling. Geophys J Int 169:951–963. https://doi.org/10.1111/j.1365-246X.2007.03354.x