

FULL PAPER

Open Access

Stochastic forecasting of the geomagnetic field from the COV-OBS.x1 geomagnetic field model, and candidate models for IGRF-12

Nicolas Gillet^{1*}, Olivier Barrois¹ and Christopher C Finlay²

Abstract

We present the geomagnetic field model COV-OBS.x1, covering 1840 to 2020, from which have been derived candidate models for the IGRF-12. Towards the most recent epochs, it is primarily constrained by first differences of observatory annual means and measurements from the Oersted, Champ, and Swarm satellite missions. Stochastic information derived from the temporal spectra of geomagnetic series is used to construct the *a priori* model covariance matrix that complements the constraint brought by the data. This approach makes it possible the use of a *posteriori* model errors, for instance, to measure the 'observations' uncertainties in data assimilation schemes for the study of the outer core dynamics.

We also present and illustrate a stochastic algorithm designed to forecast the geomagnetic field. The radial field at the outer core surface is advected by core motions governed by an auto-regressive process of order 1. This particular choice is motivated by the slope observed for the power spectral density of geomagnetic series. Accounting for time-correlated model errors (subgrid processes associated with the unresolved magnetic field) is made possible thanks to the use of an augmented state ensemble Kalman filter algorithm. We show that the envelope of forecasts includes the observed secular variation of the geomagnetic field over 5-year intervals, even in the case of rapid changes. In a purpose of testing hypotheses about the core dynamics, this prototype method could be implemented to build the 'state zero' of the ability to forecast the geomagnetic field, by measuring what can be predicted when no deterministic physics is incorporated into the dynamical model.

Keywords: Geomagnetic field; Secular variation; Stochastic equations; Ensemble Kalman filter

Background

The 12th generation of the IGRF has been computed from several candidates by different institutes, all submitted in September 2014 for evaluation by the IGRF working group (Thébault et al. 2015). We present here our proposed candidates for the main field (MF) DGRF model in 2010, IGRF in 2015, and the time average secular variation (SV) prediction over 2015 to 2020, together with their associated uncertainties. These are derived from the COV-OBS.x1 field model and its posterior model error covariance matrix. This model, that covers the observatory era since 1840, is primarily constrained at recent epochs by first differences of observatory annual

means (OAM) and by satellite observations, including data from the Swarm constellation (Olsen et al. 2013). It differs from regularized field models in the sense that the information contained into observations is complemented with a stochastic *a priori* information derived from temporal spectra of geomagnetic series, following Gillet et al. (2013). The *a priori* information integrated in the construction of COV-OBS-type models makes it possible to use the model errors, as estimated from the posterior covariance matrix, in data assimilation algorithms designed to re-analyze or forecast the outer core dynamics (Fournier et al. 2010). However, being derived over a long time-span, it cannot contain a detailed description of external magnetic signals (as in the comprehensive approach, see Sabaka et al. 2015). This leads to a compromise where external sources are handled in a simple way in comparison with models built

*Correspondence: nicolas.gillet@ujf-grenoble.fr

¹University Grenoble Alpes, ISTERre, CNRS, F-38041 Grenoble, France
Full list of author information is available at the end of the article

exclusively from the most recent observations (e.g., Lesur et al. 2010; Olsen et al. 2014).

We also present a test SV prediction (time average over 2015 to 2020) based on a stochastic model of the flow at the core-mantle boundary, using an augmented state ensemble Kalman filter (EnKF, see Evensen 2003). This prototype method is presented as a proof of concept study. It should be considered as an attempt to produce a 'state zero' of the ability to forecast the geomagnetic field, i.e., a measure of our ignorance when we have no deterministic knowledge about the core physics and only have access to statistical information. It may be used in the future for the purpose of validating geomagnetic data assimilation algorithms (Fournier et al. 2010). Possibilities for future improvements are listed in the 'Conclusions' section.

Methods

Derivation of the COV-OBS.x1 field model

Data selection

Our data selection process follows closely that used to construct the original COV-OBS field model. Full details can be found in Gillet et al. (2013); here, we only briefly describe the new or updated aspects of the datasets. Again, the most important data sources are satellite data, as measured by high-precision magnetometers on low Earth-orbiting missions (in particular, CHAMP, Oersted, and Swarm) and observatory annual means.

Regarding satellite data, we used data from ESA's Swarm satellite trio (Friis-Christensen et al. 2006) and also re-selected data from CHAMP, Oersted, and SAC-C, taking as a basis the dataset employed for a recent update of the CHAOS-4 field model (CHAOS-4plus_V4, Olsen et al. 2014). Most importantly for our IGRF-12 candidate models, the Swarm data used were ESA Level 1B operational data from 26th November 2013 to 17th August 2014 (baseline 0302 until 4th July 2014, baseline 0301 thereafter). Only quiet-time, night-side data were selected according to the CHAOS-4 criteria, and vector data were used only up to 55° degrees quasi-dipole latitude; at high latitudes, scalar intensity data were used. Following the procedure previously introduced for the COV-OBS model, the CHAOS-4plus_V4 dataset was then sub-sampled onto a grid of 72 cells in longitude by 36 cells in cosine latitude, with each cell refilled each year using random sampling from the data falling within that cell during the year. This provided a geographical and temporal coverage sufficient for studying the main field up to degree 14 and its secular variation up to degree 8, as required for IGRF, while avoiding problems due to along-track correlated errors. Vector data were selected where possible (preferentially CHAMP data when two star cameras were available) or later from Swarm-B (which is slightly further from the disturbing ionosphere than Swarm-A or Swarm-C); otherwise, scalar data was used. In particular,

during the gap from the end of the CHAMP mission in September 2010 and the start of the Swarm mission in November 2013, only Oersted scalar data were available.

All satellite vector data were rotated from the instrument frame to the North-East-Center frame using Euler angles determined during the derivation of the CHAOS-4plus_V4 model (see also Finlay et al. 2015). The CHAOS-4 estimate of the crustal field (for degrees 15 to 85) was removed from all data, but no correction was applied for the external field since, in COV-OBS type models, a simple external field model is co-estimated. In addition, following Finlay et al. (2015), all Swarm vector (VFM) data are scaled so that the magnetic field norm obtained from the vector data matches the intensity recorded by the absolute (ASM) scalar field instrument. Note that it is not the true vector field correction of the ASM-VFM difference, which was not available at the time of model construction. Data uncertainty estimates σ were allocated as described in Gillet et al. (2013): for the scalar and isotropic component of vector data, σ varies with geomagnetic latitude, taking its maximum value within 25° of the geomagnetic pole and its minimum value within 45° of the geomagnetic equator, with a cosine taper for the intervening 20°. The range for each dataset is given in Table 1. Swarm data uncertainties are allocated in the same manner as for the CHAMP data.

An update of the OAM dataset previously used to derive the COV-OBS model was also carried out. This was based on the worldwide OAM database from the World Data Centre for Geomagnetism, BGS Edinburgh (pers. comm. S. Macmillan, September 2014). It involved 13 new observatories compared to the previous dataset used by Gillet et al. (2013) and comprises annual means up to 2013.5 where available. First differences of OAM thus provide constraints on the secular variation up until 2013.0 that are not sensitive to the crustal field. Uncertainty estimates were re-calculated for each updated observatory time series using a generalized cross validation approach as in Bloxham and Jackson (1992), Gillet et al. (2013), and Jackson et al. (2000). No corrections for external field variations, beyond the averaging inherent in the annual means, were applied.

Parameterization and prior information for the COV-OBS.x1 model

We refer to Gillet et al. (2013) for the details concerning the construction of the COV-OBS field model. Here, we recall only the main points. We use a spherical harmonic expansion of the internal field up to degree $N = 14$. A single coefficient, the axial dipole in geomagnetic dipole coordinates, is used for the spherical harmonic expansion of the external field (with a 20-nT background value). All coefficients are projected onto order 4 cubic B-splines,

Table 1 Errors statistics on *F* data for various satellites

Satellite	Period	No	σ	\mathcal{M}^*	μ^*	\mathcal{M}	μ
Oersted	1999.2 to 2013.5	5580	5.5 to 8.5	0.95	0.03	5.58	0.10
CHAMP	2000.6 to 2010.7	4489	4 to 10	0.95	-0.05	8.39	-0.55
Sac-C	2001.1 to 2001.2	32	5.5 to 8.5	0.98	0.19	5.51	0.93
Swarm	2013.9 to 2014.6	574	4 to 10	0.65	-0.08	5.39	-0.88
Total	1999.2 to 2014.6	10675		0.93	-0.01	6.90	-0.23

Data number *No*, range for the prior error σ (nT), weighted L2 data misfit \mathcal{M}^* and bias μ^* , unweighted L2 data misfit \mathcal{M} and bias μ (in nT).

with knots every 2 years, spanning the period 1838 to 2022.

The main difference with the more commonly employed regularization procedures (penalizing second or third time derivatives plus a spatial norm involving damping parameters, see Finlay et al. 2012) is that we try to define an *a priori* covariance matrix on the model coefficients that is as realistic as possible, in order to be able to use the information contained in the *a posteriori* model errors' covariance matrix. For each Gauss coefficient g_n^m of degree n and order m (and similarly for h_n^m), we choose a time covariance function:

$$C_n^m(\tau) = E(g_n^m(t)g_n^m(t + \tau)) = \sigma_g(n)^2 \left(1 + \frac{|\tau|}{\tau_c(n)}\right) \exp\left(-\frac{|\tau|}{\tau_c(n)}\right). \quad (1)$$

τ is the time lag between two epochs, $E(\dots)$ means the statistical expectation. For each coefficient, the *a priori* distribution is centered on zero. We consider that coefficients of different orders or degrees are *a priori* independent (h_n^m and g_n^m are also independent). For the sake of simplicity, the C_n^m do not vary with the spherical harmonic order m . The main field variances $\sigma_g(n)^2$ are estimated from the variances of satellite MF model coefficients in 2005, averaged over all orders. Given Equation (1), the relevant time-scales are by definition $\tau_c(n) = \sqrt{3}\sigma_g(n)/\sigma_g'(n)$, with $\sigma_g'(n)$ the secular variation variances, estimated from the variances of satellite SV model coefficients in 2005, averaged over all orders (see Gillet et al. 2013).

Equation (1) corresponds to stationary, auto-regressive processes φ of order 2 governed by a stochastic differential equation of the form (e.g. Yaglom 1962)

$$\frac{d^2\varphi}{dt^2} + \frac{2}{\tau_c} \frac{d\varphi}{dt} + \frac{\varphi}{\tau_c^2} = \epsilon(t), \quad (2)$$

with ϵ a white noise process^a. This description is chosen since it is consistent with the slope close to -4 found for the power spectrum density (PSD) of observatory series at periods from 5 to 70 years (De Santis et al. 2003). The continuation of such a slope for the PSD of the internal field at periods shorter than a few years is unsure given the domination of the external signal towards high frequencies (see

Figure two in Ou et al. 2015), but there is no need either for a steeper slope given the current estimate of the mantle conductivity (Velínský and Finlay 2011).

Since *D*, *I*, and *F* data are involved, model estimation is a nonlinear optimization problem. We solve it with a Newton algorithm, using a L2 measure of the data misfit together with a rejection criterion. We first build, starting from COV-OBS, an intermediate model rejecting data for which the residual magnitude is above $10\sigma_e$ (σ_e is the *a priori* data error). Next, from this model we obtain within six iterations (typically a few iterations are enough for convergence) the COV-OBS.x1 field model together with its *a posteriori* model errors' covariance matrix using a $3\sigma_e$ rejection criterion.

Stochastic forecast of the geomagnetic field

We present below how we derive our test SV model for IGRF-12. It results from the integration of a stochastic core flow model into an augmented state ensemble Kalman filter (Evensen 2003).

Stochastic core flow modeling

In the frozen-flux approximation, core surface motions \mathbf{u} are related to changes in the radial geomagnetic main field B_r through the diffusive-less induction equation at the core-mantle boundary (e.g., Holme 2007):

$$\frac{\partial B_r}{\partial t} = -\nabla_H \cdot (\mathbf{u}B_r). \quad (3)$$

We consider the simplest stationary, stochastic model compatible with the -4 slope obtained for the temporal PSD of geomagnetic field series at ground-based observatories for periods $5 \leq T \leq 70$ years (De Santis et al. 2003). It consists of an order 1 stochastic differential equation, of the form^b

$$\frac{\partial \mathbf{u}}{\partial t} + \frac{\mathbf{u}}{\tau_u} = \mathbf{z}(t), \quad (4)$$

with \mathbf{z} a white noise process (see Gillet et al. 2015). Equation (4) defines the time correlation function for \mathbf{u} :

$$\rho_u(\tau) = \exp(-|\tau|/\tau_u). \quad (5)$$

For the sake of simplicity, a single value of $\tau_u = 100$ years is assumed for all the flow coefficients (see Gillet et al. 2015).

We consider here flow models calculated under the compressible quasi-geostrophic (QG) constraint, i.e., (Pais and Jault 2008):

$$\begin{aligned} u_\phi(\pi - \theta, \phi) &= u_\phi(\theta, \phi) \quad , \quad u_\theta(\pi - \theta, \phi) = -u_\theta(\theta, \phi) \quad , \\ \nabla_H \cdot (\mathbf{u} \cos \theta) &= 0. \end{aligned} \quad (6)$$

The flow model, decomposed into toroidal (t_n^m) and poloidal (s_n^m) components, is described up to spherical harmonic degree $N_u = 20$. Relatively large flow variances are assumed:

$$\sigma_u(n)^2 = E(t_n^{m2}) = \frac{10^2}{(n+1)(2n+1)} \mathcal{T}(n), \quad (7)$$

which is equivalent to assuming a flat CMB flow power spectrum with a taper $\mathcal{T}(n) = \min\{1, 10^{14-n}\}$ imposed to ensure the convergence of the scheme. Under the constraint (6), variances for the poloidal coefficients derive from those of the toroidal ones through the geostrophic chains (Le Mouél et al. 1985). Variances for \mathbf{z} are directly deduced from those for \mathbf{u} (e.g., Papoulis and Pillai 2002):

$$\sigma_z(n)^2 = \frac{2\sigma_u(n)^2}{\tau_u}. \quad (8)$$

From (8), we construct the covariance matrix \mathbf{C}_z for \mathbf{z} . The linear set of constraints (6) is transformed into matrix form, $\mathbf{Lz} = \mathbf{0}$ (if \mathbf{z} satisfies the constraints, then it is also the case for \mathbf{u}). We generate the random vectors \mathbf{z} by multiplying the Choleski decomposition of $\mu \mathbf{L}^T \mathbf{L} + \mathbf{C}_z$ (with μ a very large scalar, we use $\mu = 10^9$) to a random unit vector normally distributed. Then, flow models \mathbf{u} derived from \mathbf{z} satisfy the constraints (6) and present the variances of Equation (7).

Augmented state ensemble Kalman filter

The main source of errors, when imaging the flow from models of the secular variation $\partial_t B_r$ using Equation (3), comes from nonlinear interactions involving the unresolved parts of the flow and of the magnetic field (Baerenzung et al. 2014; Eymin and Hulot 2005; Pais and Jault 2008). The projection of Equation (3) onto large length-scales (denoted by overlines) gives

$$\frac{\partial \overline{B}_r}{\partial t} = -\overline{\nabla_H \cdot (\mathbf{u} \overline{B}_r)} + \mathcal{E}, \quad (9)$$

with $\mathcal{E} = -\overline{\nabla_H \cdot (\mathbf{u} \tilde{B}_r)}$ referred to as the SV model errors. The unresolved field \tilde{B}_r at small length-scales being correlated over decades, \mathcal{E} is also correlated in time (Gillet et al. 2015). We model it through the stochastic equation

$$\frac{\partial \mathcal{E}}{\partial t} + \frac{\mathcal{E}}{\tau_\mathcal{E}} = \mathcal{Z}(t), \quad (10)$$

with \mathcal{Z} a white noise process. This defines the correlation function for \mathcal{E} , $\rho_\mathcal{E}(\tau) = \exp(-|\tau|/\tau_\mathcal{E})$. This choice, with

$\tau_\mathcal{E} = 10$ years, builds on the investigations of Gillet et al. (2015) who constructed flow models from COV-OBS over 1940 to 2010 (see their Figure one). Variances for the SV model error coefficients, which depend only on the degree n , are also estimated from their results.

Equations (4), (9), and (10) are time-stepped to forecast the (augmented) model state:

$$\mathbf{x}^f = \left[\overline{B}_r, \mathbf{u}^f, \mathcal{E}^f \right]^T. \quad (11)$$

We use an Euler-Maruyama scheme (e.g., Kloeden and Platen 1992) to time-step the stochastic equations. We consider here as observations snapshots of the MF and SV coefficients of the COV-OBS.x1 field model, obtained from the continuous parameterization (through B-splines) of the field model. Observations fall on a subset of the discrete time-steps involved when integrating the forward problem, so there is no need for interpolating in time the stochastic model.

For the sake of simplicity, the MF and SV data error covariance matrices $\mathbf{R}_{MF}(t)$ and $\mathbf{R}_{SV}(t)$ are here assumed to be diagonal. Variances are obtained from the dispersion within 50 realizations of the COV-OBS.x1 model (an ensemble large enough to give a converged estimate of the diagonal elements), derived from the COV-OBS.x1 posterior model error covariance matrix (see Gillet et al. 2013). By ignoring cross-covariances, we avoid the issues related to rank-deficient matrices. Predictions for geomagnetic field observations \mathbf{y}^o are related to the model state \mathbf{x} through the forward equation:

$$\mathbf{y}^o = \mathbf{H}(\mathbf{x}) + \mathbf{e}^o \quad (12)$$

(\mathbf{e}^o the observation errors, \mathbf{H} the forward operator, see below). The state $\mathbf{x}(t)$ is obtained using an ensemble Kalman filter (Evensen 2003), composed of a succession of forecast and analysis steps:

$$\mathbf{x}^a = \mathbf{x}^f + \mathbf{P}^f \nabla \mathbf{H}^T \left[\nabla \mathbf{H} \mathbf{P}^f \nabla \mathbf{H}^T + \mathbf{R} \right]^{-1} \left(\mathbf{y}^o - \mathbf{H}(\mathbf{x}^f) \right), \quad (13)$$

with $\nabla \mathbf{H}$ the linear tangent of the forward operator and \mathbf{R} the observation error covariance matrix. The model covariance matrix \mathbf{P}^f is obtained from an ensemble of $K = 50$ realizations of the model as:

$$\mathbf{P}^f = \frac{1}{K-1} \sum_{k=1}^K \left(\mathbf{x}_k^f - \langle \mathbf{x}^f \rangle \right) \left(\mathbf{x}_k^f - \langle \mathbf{x}^f \rangle \right)^T, \quad (14)$$

with $\langle \mathbf{x}^f \rangle$ the ensemble average of the models. Given the small ensemble size compared to the model dimension, cross-covariances in \mathbf{P}^f are ignored to ensure a positive-definite matrix (one may consider reduced rank filters or localization techniques to address this issue, while conserving the information carried by off-diagonal elements).

All quantities are represented in the spherical harmonic domain.

We proceed to the analysis step every 5 years. Because the forecast field \bar{B}_r^f can deviate significantly from the exact large-length-scale field \bar{B}_r after 5 years, each analysis is proceeded in two steps (following Aubert 2014):

- we first perform an analysis of $\mathbf{x}_1 = \bar{B}_r$, from K noisy realizations of the COV-OBS.x1 field model, which involves $\mathbf{y}_1^o = \bar{B}_r^o$, $\mathbf{R} = \mathbf{R}_{MF}$, and $\mathbf{H}_1(\mathbf{x}) = \bar{B}_r$ (i.e., $\nabla \mathbf{H}_1 = \mathbf{I}_P$ the identity matrix of size $P = N(N + 2)$);
- next, we compute an analysis of $\mathbf{x}_2 = [\mathbf{u}, \mathcal{E}]$, with $\mathbf{y}_2^o = \partial_t \bar{B}_r^o$, $\mathbf{R} = \mathbf{R}_{SV}$ and:

$$\nabla \mathbf{H}_2 = \begin{bmatrix} \mathbf{A}(\bar{B}_r^a) & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_P \end{bmatrix},$$

where $\mathbf{A}(\bar{B}_r^a) \mathbf{u} = -\nabla_H \cdot (\mathbf{u} \bar{B}_r^a)$.

We perform an analysis of the observations every 5 years, starting from 1985 (except the last analysis, performed in 2014.5 instead of 2015). Several (at least two) analysis steps are required in order to eliminate the initial condition memory on the covariances that enter the EnKF.

Results and discussion

The COV-OBS.x1 field model

Statistics on the COV-OBS.x1 prediction errors

We now provide some statistics concerning the COV-OBS.x1 misfit to observatory and satellite observations, focusing on the period spanning 1990 to 2015. We present in Figure 1 the time evolution of the OAM normalized (i.e., weighted by the *a priori* data errors) data misfit and bias, together with the number of available (X, Y, Z) observations (standing for the northward, eastward, and downward components). We observe no particular change in the statistics with the advent of continuous satellite records in 1999, indicating a global compatibility between ground-based and satellite data.

The unweighted and weighted misfit and bias for various satellites are summarized in Table 1 for intensity data and in Table 2 for vector data. (B_b, B_\perp, B_3) stand for the magnetic field components rotated in the frame appropriate for describing anisotropic attitude errors (see Holme and Bloxham 1996; Olsen 2002). For each of the satellites, the weighted bias is reasonably close to zero when a large enough number of data is available. Normalized misfits are close to unity for all subsets of data, except for Swarm F data for which *a priori* error estimates seem to be over-estimated. We do not use any iterative scheme to re-weight the observation errors (e.g., Olsen 2002). Unweighted residuals to CHAMP F data are relatively large because with our selection process (see the ‘Data

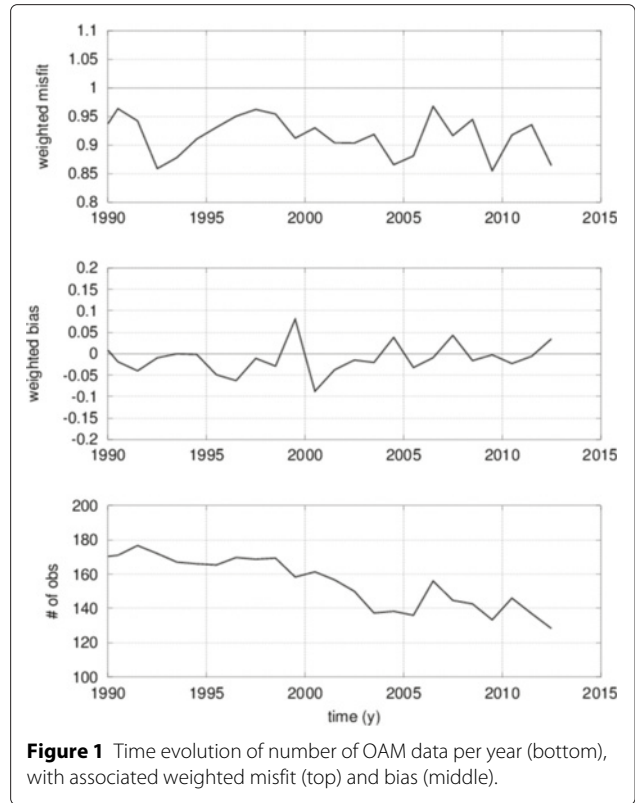


Figure 1 Time evolution of number of OAM data per year (bottom), with associated weighted misfit (top) and bias (middle).

selection’ section), all these data are located at high geomagnetic latitudes (see also Figure 2). Large misfit values obtained for Oersted B_\perp data are to be associated with the availability of a single star camera to determine Euler angles (see Olsen et al. 2014). We show in Figure 3 the time evolution of the satellite data weighted misfit and bias for all three components of the field in the frame aligned with the magnetic field vector. We see slightly larger residuals during the most active solar periods (2002 to 2005) and note some relatively large biases, for instance, on F data in 2002 and 2004.

The dependence of the unweighted residuals on the geomagnetic latitude is shown in Figure 2 for various satellites. As expected, larger prediction errors on F data are observed above 60° latitude. The envelope of the residuals shows a minimum around 35° latitude and a local maximum at the equator for both F and B_b data for all three satellites. This indicates some unmodelled contributions from the ring current (see the discussion of Figure two in Olsen 2002). It may at least partly be due to the inability of our model to fit high-frequency external sources due to the choice of a 2-year knot spacing also for the external field. As in Olsen (2002), B_p Oersted data and B_3 Oersted and CHAMP data are associated with lower residuals towards the geomagnetic equator. This trend is less obvious with Swarm data, which present on average lower residuals for both scalar and vector data.

Table 2 Errors statistics on vector data for various satellites

Satellite	No	\mathcal{M}^*			μ^*			\mathcal{M}			μ		
		B_b	B_{\perp}	B_3	B_b	B_{\perp}	B_3	B_b	B_{\perp}	B_3	B_b	B_{\perp}	B_3
Oersted	3971	1.01	1.12	0.86	0.02	-0.08	0.01	5.54	8.38	4.92	0.11	-0.61	0.08
CHAMP	18507	1.13	0.94	1.24	0.00	0.05	-0.04	4.51	4.07	5.43	0.00	0.21	-0.18
Swarm	3935	1.04	0.88	1.19	-0.01	0.01	-0.06	4.16	3.82	5.15	-0.03	0.05	-0.26
Total	26413	1.10	0.96	1.18	0.00	0.02	-0.04	4.63	4.94	5.31	0.01	0.07	-0.15

Data number No , weighted L2 data misfit \mathcal{M}^* and bias μ^* , unweighted L2 data misfit \mathcal{M} and bias μ (in nT).

Description of COV-OBS.x1 and its uncertainties

We now provide details concerning the characteristics of the COV-OBS.x1 field model. We present in Figure 4 MF and SV spectra at the Earth's surface for epochs 2010, 2015, and 2020, for both the average model and

the estimated model errors (derived from the posterior covariance matrix). Uncertainties are relatively larger at higher degrees. Starting from epoch 2010 to 2015 then 2020, uncertainties gradually increase, illustrating the weakening of the constraint from the data (and thus the

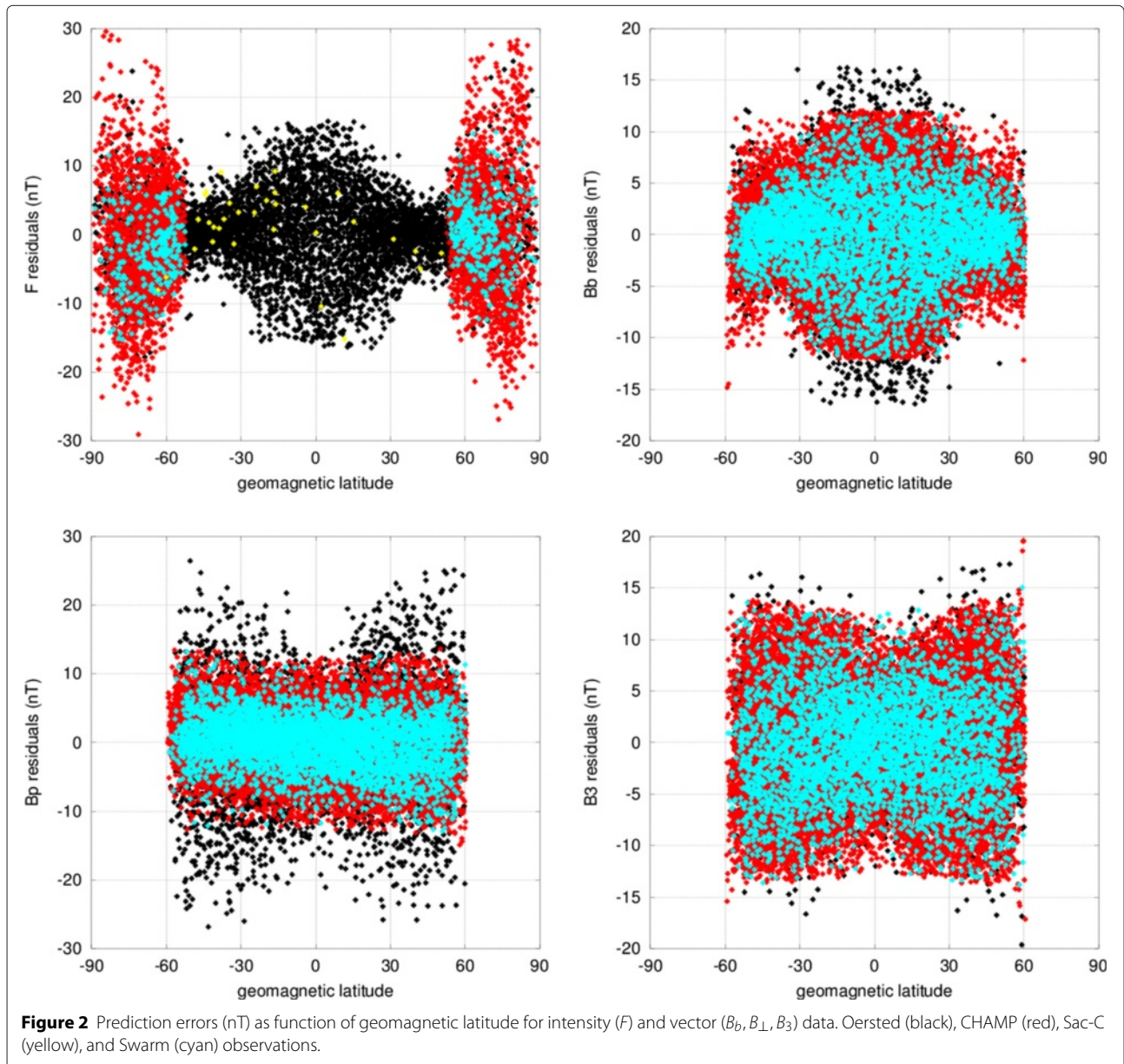


Figure 2 Prediction errors (nT) as function of geomagnetic latitude for intensity (F) and vector (B_b, B_{\perp}, B_3) data. Oersted (black), CHAMP (red), Sac-C (yellow), and Swarm (cyan) observations.

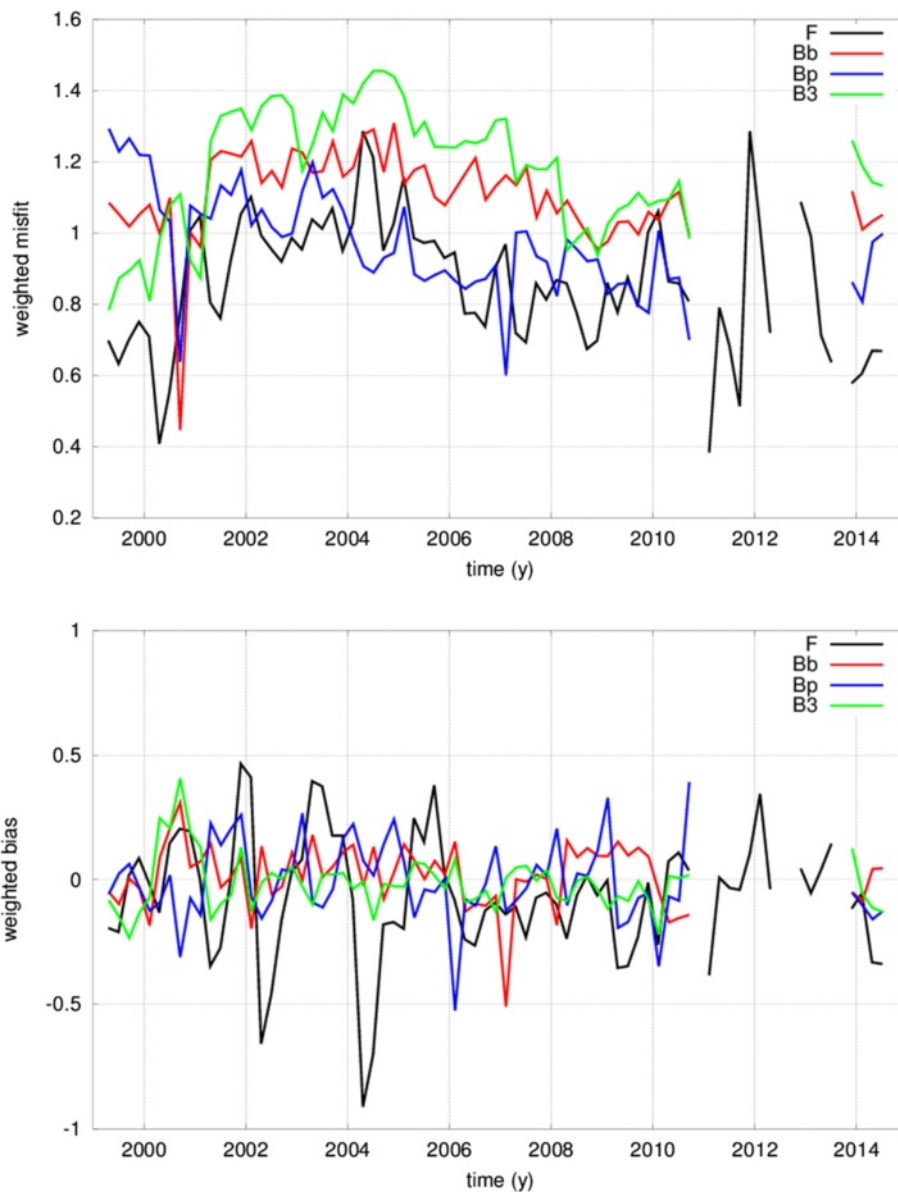
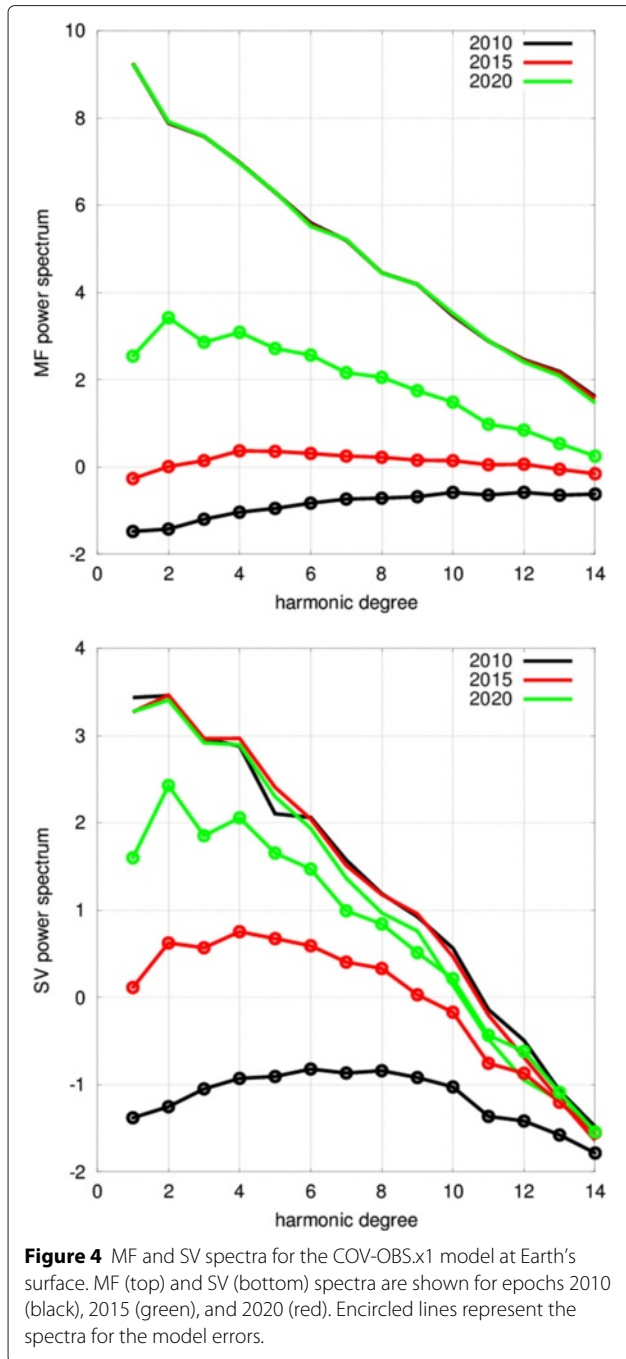


Figure 3 Satellite data weighted misfit (top) and bias (bottom) as a function of time.

increasing importance of the *a priori* covariances) towards the future. We also notice a significant decrease of the SV power in 2020 compared to 2010 (see below).

We show in Figure 5 examples of the MF and SV COV-OBS.x1 model coefficients, in comparison with those from the CHAOS-4plus_v4 model (Olsen et al. 2014). A reasonable agreement is found between the two models, COV-OBS.x1 presenting sharper SV variations at high degrees, due to the strong damping used when building the CHAOS model. Indeed, regularized field models tend to provide a smoothed estimate (i.e., weighted time average) of the SV coefficient series towards high degrees. An almost constant bias of 4 nT between the two models is

found for coefficient g_1^0 , COV-OBS.x1 showing more negative values. This may be associated with several factors, among which (see Gillet et al. 2013) a possible aliasing in time due to the projection of q_1^0 onto splines with a 2-year knot spacing, the use of a 20-nT background value for the parameterization of q_1^0 , or the simple description of the induced response to external fields (COV-OBS.x1 being designed to model field changes on interannual periods and longer, we simply assume an infinitely conducting outer core, with no consideration of induced currents in the mantle). Some rapid SV fluctuations observed on several low-degree CHAOS-4plus_v4 coefficients are also absent in COV-OBS.x1 (see, for instance, g_2^1). These are



possibly due to the filtering of rapid SV changes by the projection onto splines with 2-year knot spacing and/or to some leakage of high-frequency external sources into the CHAOS internal model.

We obtain relatively larger error bars at higher degrees. In comparison with the period prior to 1999, where the SV is mainly constrained by OAM, we find smaller uncertainties during the satellite era. These increase after 2014.6, when the control imposed by *a priori* covariances

gradually takes over from that of the observations. The ensemble average SV prediction after 2014.6 becomes almost flat. This behavior is in agreement with our assumed prior information. Indeed, the correlation function in Equation (1) is that of an auto-regressive process of order 2 for the MF, i.e., approximately a random walk for the SV on short periods, characterized by an envelope of possible solutions $\propto \sqrt{t}$.

We present in Figure 6 the prediction of an ensemble of 20 realizations of the COV-OBS.x1 (internal plus external) field model at several ground-based observatories. These reflect the behavior described above on the field coefficients. Note that because both internal and external coefficients are projected onto splines, fluctuations at periods much shorter than the (2-year) knot spacing are not modeled, also after 2014.6. This is particularly obvious on the *X* series, because (i) *Y* data are less affected by auroral electrojets and the equatorial ring current and (ii) the magnitude of decadal (mainly from internal origin) SV changes is larger on *Z* than on *X* for the two sites presented here.

Results of the MF and SV stochastic forecast

We now present predictions of the geomagnetic field obtained with the stochastic core flow modeling and the augmented state EnKF method. We show in Figure 7 SV and MF spectra for the forecast and analysis errors in 2014.5. These provide a measure of the forecast uncertainties for the upcoming 5-year intervals. The analysis errors being below the observation error, our algorithm produces solutions able to fit the MF and SV data within the error bars provided by COV-OBS.x1.

We present in Figure 8 some examples of MF and SV coefficient predictions for several cycles of analysis and forecast steps, spanning 1990 to 2020. Please note that the stochastic forecast is discrete (contrary to the continuous spline model COV-OBS.x1), with a time-step of 0.5 years and that it has not been filtered in time. The dispersion within the ensemble of stochastic SV forecasts is comparable to that obtained directly from the COV-OBS.x1 posterior covariance matrix (the latter being sometimes a bit less, as for instance with \dot{g}_1^0).

For most coefficients, observations are accounted for by one standard deviation σ^f estimated from our ensemble of forecasts. This indicates that the statistics of our stochastic model are fairly realistic, although this diagnostic is imperfect: COV-OBS.x1 coefficients at two different epochs are not independent. Thus, the forecast at a given epoch is not independent from the observation at this epoch. To obtain an independent diagnostic, forecasts produced from COV-OBS-type field models constructed from datasets without data from the forecast era would be required. For some coefficients, there exists a systematic bias (see, for instance, the positive trend taken by

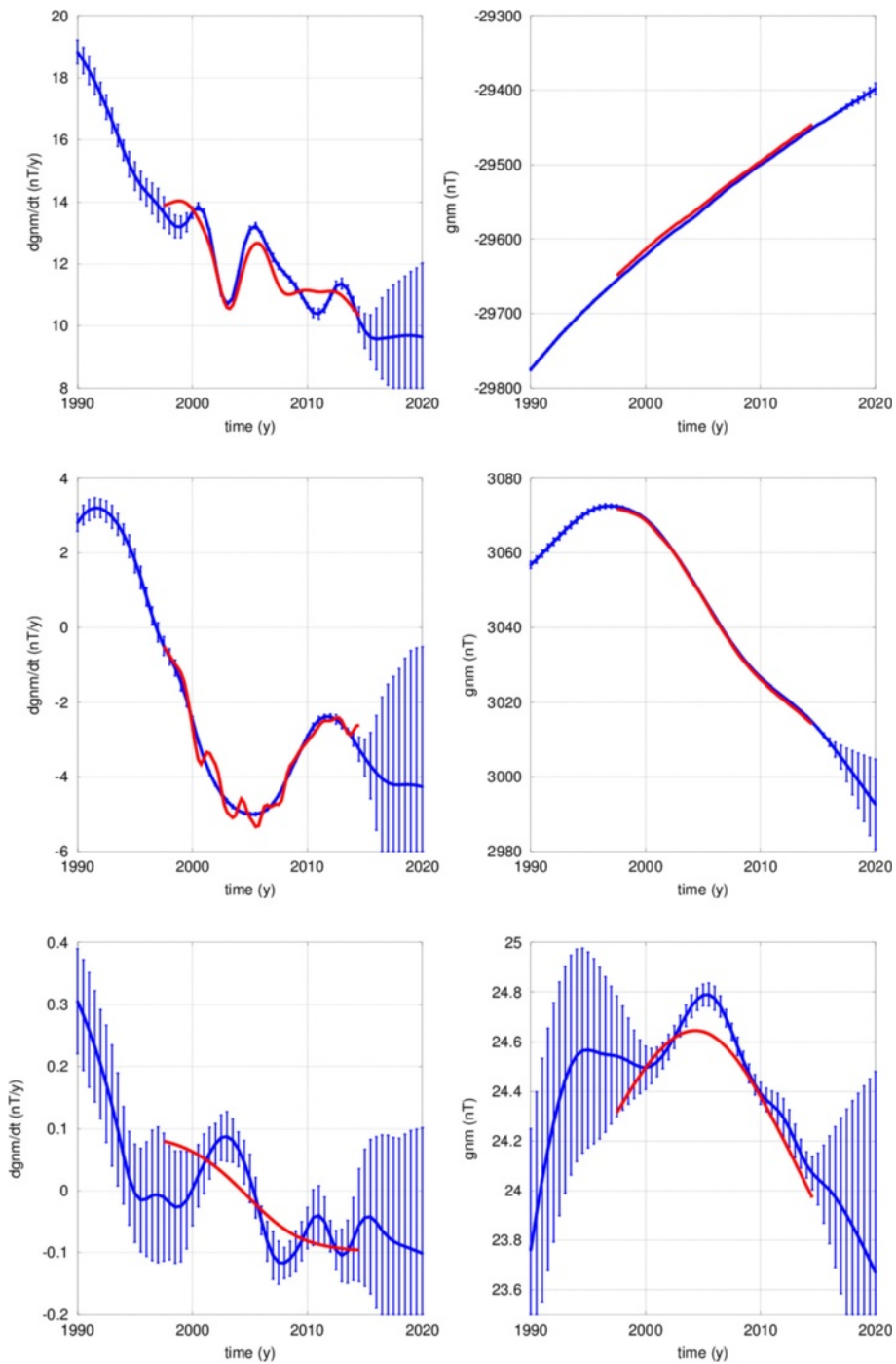
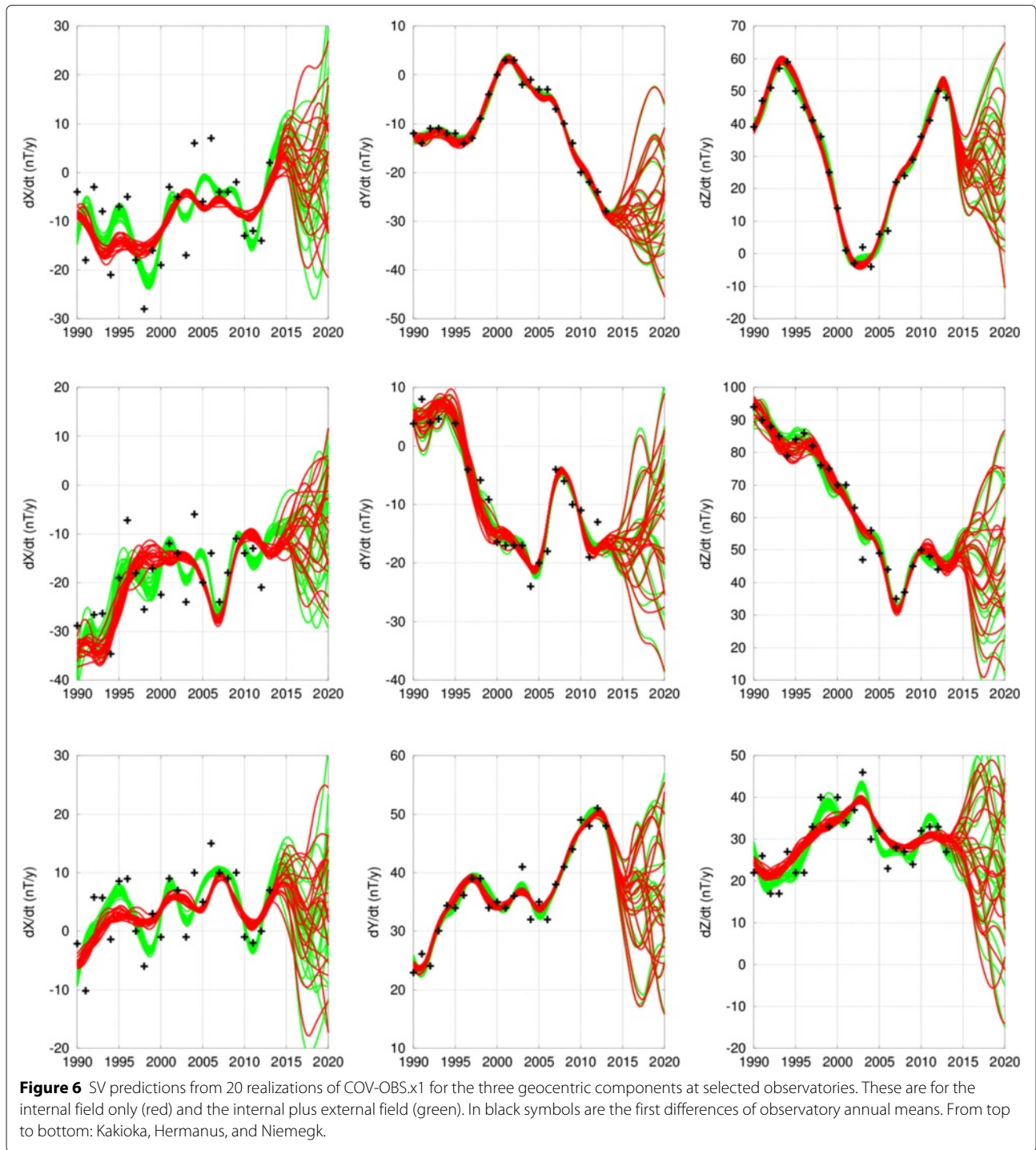


Figure 5 Example of COV-OBS.x1 MF (right) and SV (left) coefficient series (blue) and comparison with CHAOS4plus_v4 (red). From top to bottom: g_1^0 , g_2^1 , and g_8^1 . Error bars indicate \pm one standard deviation of the COV-OBS.x1 model errors.

the average h_2^1 forecast): even though the observed series usually lies within the ensemble of forecast realizations, it sometimes lies in an area of lower probability, as measured by the pdf of the forecast. Initial investigations of this problem suggest it is not related to our choice of topological constraint (6) but probably to our use of stochastic

models for \mathbf{u} and \mathcal{E} that are centered around zero (we use no background model).

We also make a comparison of the EnKF forecast with the IGRF-11 predictions for 2010 to 2015 (Finlay et al. 2010). We find that the IGRF-11 SV predictions are accounted for by one standard deviation σ^f estimated



from our ensemble of forecasts. The results presented in this section should be considered as a proof of concept study; possible future applications or modifications are presented in the ‘Conclusions’ section.

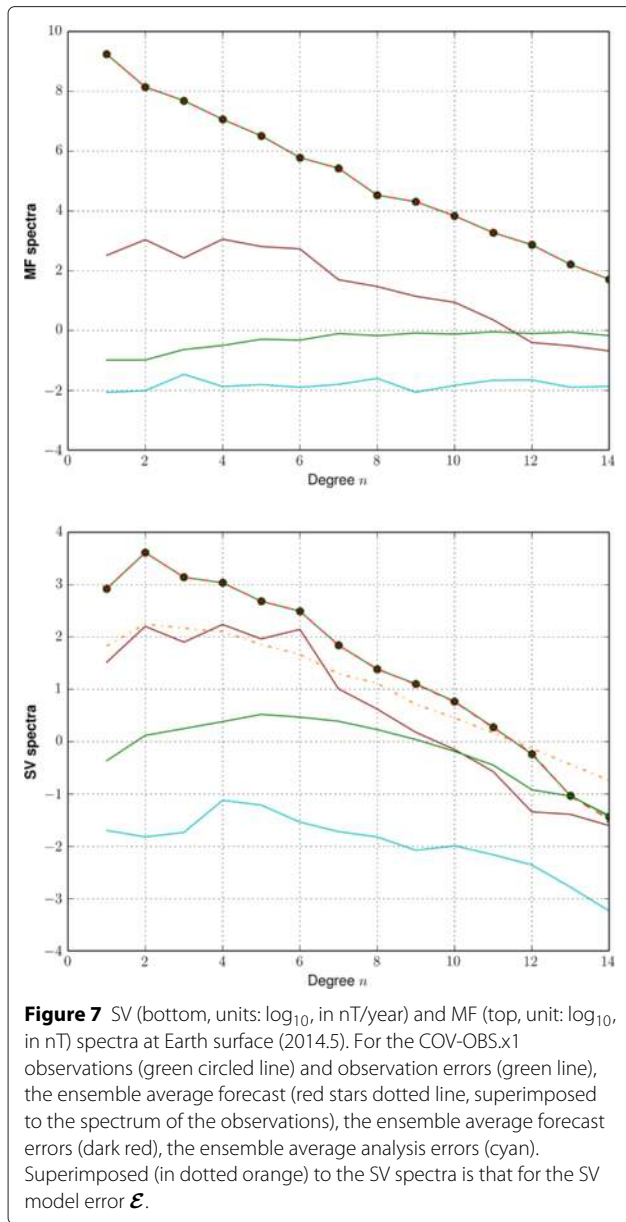
Our IGRF-12 candidate models

Candidate models for IGRF-12 have been derived from an ensemble of $K = 50$ realizations of the COV-OBS.x1 MF

and SV field coefficients, $\{\mathbf{g}_k(t), \dot{\mathbf{g}}_k(t)\}_{k \in [1, K]}$. Our DGRF MF candidate is the COV-OBS.x1 ensemble average MF model evaluated in 2010:

$$\mathbf{g}_{DGRF} = \langle \mathbf{g}(2010) \rangle . \tag{15}$$

The associated error bars are estimated as the dispersion relative to the ensemble average within the K realizations of the MF model:



$$\delta \mathbf{g}_{\text{DGRF}} = \sqrt{\frac{1}{K} \sum_{k=1}^K [\mathbf{g}_k(2010) - \langle \mathbf{g}(2010) \rangle]^2}. \quad (16)$$

Similarly, we estimate the IGRF MF candidate, and its associated error bars, from the ensemble average of the COV-OBS.x1 realizations, and the dispersion within the ensemble, evaluated in 2015. Our IGRF SV candidate is the COV-OBS.x1 ensemble average SV model, averaged in time over 2015 to 2020:

$$\dot{\mathbf{g}}_{\text{IGRF}} = \frac{1}{2020 - 2015} \int_{2015}^{2020} \langle \dot{\mathbf{g}} \rangle (t) dt. \quad (17)$$

The associated error bars are estimated as the dispersion relative to the ensemble average within the K realizations of the SV model:

$$\delta \dot{\mathbf{g}}_{\text{IGRF}} = \sqrt{\frac{1}{K} \sum_{k=1}^K \left[\frac{1}{2020 - 2015} \int_{2015}^{2020} (\dot{\mathbf{g}}_k(t) - \langle \dot{\mathbf{g}} \rangle (t) dt \right]^2}. \quad (18)$$

Additionally, we presented a test IGRF SV candidate as the ensemble average of SV forecast resulting from the stochastic EnKF, averaged in time over 2015 to 2020 (its associated error bars are estimated from the dispersion relative to the ensemble average within the ensemble of SV forecast realizations).

Conclusions

Beyond the ‘state zero’ forecast of the geomagnetic field

Forecasting the geomagnetic field requires, upon accurate observations at or above the Earth’s surface, understanding the physics occurring within the Earth’s outer core over millennial to interannual time-scales. The knowledge of the field behavior on long time-scales can be inferred from field models derived from archeomagnetic and lake sediment data bases (e.g., Korte et al. 2011; Licht et al. 2013; Nilsson et al. 2014; Pavón-Carrasco et al. 2014). Since the turn-over time is of the order of a few hundreds of years, the re-analysis of such models may help define a background flow above which a flow perturbation could be described (equivalent of a climatic mean in oceanic studies). This may reduce systematic biases observed in the SV predictions. One may think, for instance, of the coupled-Earth dynamo scenario by Aubert et al. (2013), where a slowly evolving westward drift naturally arises from the coupling between the inner core and the mantle.

In this study, the evolution of the geomagnetic field is constrained by statistical properties inferred from the temporal PSD recovered for magnetic series. There exists dynamical interpretations associated with the observed slopes of the spectra (Tanriverdi and Tilgner 2011) (see also Buffett and Matsui 2015; Buffett et al. 2014; Olson et al. 2012, for studies focused on the analysis of the dipole moment in geodynamo simulations). Nevertheless, the method derived in the ‘Stochastic forecast of the geomagnetic field’ section does not contain any deterministic equation for advecting the magnetic field and the flow inside the outer core. As a consequence, it provides an envelope of possible states, and its average prediction can only be marginally better than those obtained from a stationary flow hypothesis (Beggan and Whaler 2009; Waddington et al. 1995). To go further, still accounting for unresolved processes, one may introduce random forcings such as those of Equations (4) and (10) into prognostic models of the core dynamics used for the re-analysis of the core state – for instance, into geodynamo (Fournier et al. 2011) or quasi-geostrophic (Canet et al. 2009) equations.

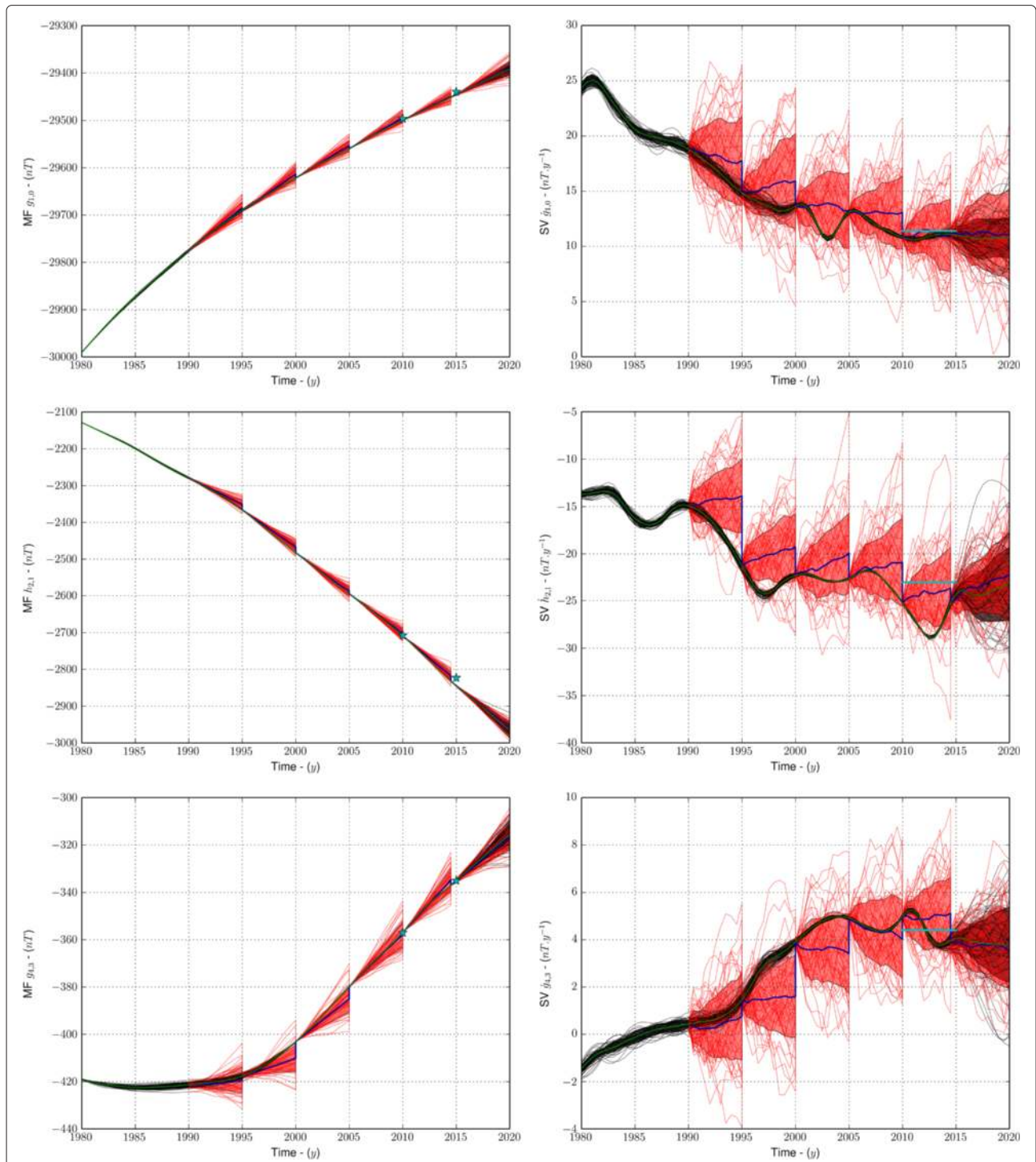


Figure 8 Example of MF and SV Gauss coefficients forecast obtained with 50 realizations of the stochastic core flow model. From top to bottom: g_0^0 , h_2^1 , and g_4^3 . In black are the ensemble of COV-OBS.x1 observations (one σ uncertainties are indicated by the grey shaded area), in red the stochastic forecast realizations (one σ uncertainties are indicated by the red shaded area), in green the ensemble average COV-OBS.x1 model, and in blue the ensemble average stochastic forecast. Superimposed on the MF series (cyan stars) are the IGRF-11 (Finlay et al. 2010) MF model in 2010 and its prediction for 2015 (i.e., the $MF_{2010} + 5 \times SV_{2010to2015}$). Superimposed on the SV series (cyan line) is the IGRF-11 average SV prediction for 2010 to 2015. These are taken from http://www.ngdc.noaa.gov/IGRF/vmod/igrf_old_models.html.

With this in mind, the development of smoother algorithms may also be envisioned (see Evensen and Van Leeuwen 2000).

The model presented in this study aims to produce, as far as possible, an unbiased estimate of the core state, considering SV model error covariances via a stochastic equation. To go further, we could reduce the importance of SV model errors by co-estimating the unresolved field at degrees $n > 14$ (and its associated uncertainties) together with the field at large length-scales, following Aubert (2014) who performed such a co-estimation using the cross-covariances between Gauss coefficients obtained from the forward integration of a geodynamo simulation. Implementing more sophisticated stochastic models for \mathcal{E} can also be envisioned, given that the Laplace correlation function associated with the AR-1 Equation (10) is only an approximation of that found by Gillet et al. (2015). Finally, we used here the compressible QG constraint on core surface flows. Any possible topological constraint may be considered, and the algorithm presented above could be used to test the ability of each hypothesis to predict the observed SV within re-analysis studies. With the prototype algorithm illustrated above, one could, for instance, explore the hypothesis of a stratified layer at the top of the outer core, through its implications on the structure in space and time of core surface flow (Buffett 2014).

Endnotes

^a Note that Equation (2) differs from Equation (twelve) in Gillet et al. (2013). Both define processes with the same auto-covariance function (1), but their Equation (twelve), contrary to our Equation (2), corresponds to a non-stationary process. This does not affect their results that have been computed from the covariance function and not the stochastic equation.

^b By replacing in Equation (4) the scalar quantity $1/\tau_u$ by a matrix homogeneous to a frequency, one could derive more complex stochastic models still compatible with the -4 slope for the PSD of observatory series.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NG adapted and ran the code to produce the COV-OBS.x1 field model. CF performed the selection of the satellite and ground observatory geomagnetic data. NG and CF performed the analysis of the model statistics. NG and OB derived the augmented-state Kalman filter algorithm used to produce the SV candidate model. This tool has been implemented, validated, and run by OB. NG lead the writing of the manuscript. All authors discussed the results and commented on the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank the national institutes that support ground magnetic observatories and INTERMAGNET for promoting high standards of practice and making the data available. The support of the CHAMP mission by the German Aerospace

Center (DLR) and the Federal Ministry of Education and Research is gratefully acknowledged. The Oersted Project was made possible by extensive support from the Danish Government, NASA, ESA, CNES, DARA, and the Thomas B. Thriges Foundation. We thank T. Sabaka and A. De Santis for their review of our manuscript. This work was supported in part by the French Centre National d'Etudes Spatiales (CNES) for the preparation of the Swarm mission of ESA. This work has been partially supported by the French 'Agence Nationale de la Recherche' under the grant ANR-11-BS56-011. We thank the International Space Science Institute for providing support for the workshops of international team no. 241.

Author details

¹University Grenoble Alpes, ISTerre, CNRS, F-38041 Grenoble, France. ²DTU Space, Technical University of Denmark, Elektrovej 328, 2800 Kgs. Lyngby, Denmark.

Received: 15 December 2014 Accepted: 4 April 2015

Published online: 14 May 2015

References

- Aubert J (2014) Earth's core internal dynamics 1840–2010 imaged by inverse geodynamo modelling. *Geophys J Int* 197(3):1321–1334
- Aubert J, Finlay CC, Fournier A (2013) Bottom-up control of geomagnetic secular variation by the Earth's inner core. *Nature* 502:219–223
- Baerenzung J, Holschneider M, Lesur V (2014) Bayesian inversion for the filtered flow at the Earth's core mantle boundary. *J Geophys Res Solid Earth* 119:2695–2720. DOI 10.1002/2013JB010358
- Beggan C, Whaler K (2009) Forecasting change of the magnetic field using core surface flows and ensemble Kalman filtering. *Geophys Res Lett* 36:L18303. DOI 10.1029/2009GL039927
- Bloxham J, Jackson A (1992) Time-dependent mapping of the magnetic field at the core-mantle boundary. *J Geophys Res* 97:19,537–19,563
- Buffett B (2014) Geomagnetic fluctuations reveal stable stratification at the top of the Earth's core. *Nature* 507:484–487
- Buffett B, Matsui H (2015) A power spectrum for the geomagnetic dipole moment. *Earth Planetary Sci Lett* 411:20–26
- Buffett BA, King EM, Matsui H (2014) A physical interpretation of stochastic models for fluctuations in the earth's dipole field. *Geophys J Int* 198(1):597–608
- Canet E, Fournier A, Jault D (2009) Forward and adjoint quasi-geostrophic models of the geomagnetic secular variation. *J Geophys Res* 114:B11,101. doi:10.1029/2008JB006189
- De Santis A, Barraclough D, Tozzi R (2003) Spatial and temporal spectra of the geomagnetic field and their scaling properties. *Phys Earth Planet Int* 135(2):125–134
- Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn* 53(4):343–367
- Evensen G, Van Leeuwen PJ (2000) An ensemble Kalman smoother for nonlinear dynamics. *Mon Weather Rev* 128(6):1852–1867
- Eymin C, Hulot G (2005) On core surface flows inferred from satellite magnetic data. *Phys Earth Planet Inter* 152:200–220
- Finlay CC, Maus S, Beggan CD, Bondar TN, Chambodut A, Chernova TA, Chulliat A, Golovkov VP, Hamilton B, Hamoudi M, Holme R, Hulot G, Kuang W, Langlais B, Lesur V, Lowes FJ, Lühr H, Macmillan S, Mandea M, McLean S, Manoj C, Menvielle M, Michaelis I, Olsen N, Rauberg J, Rother M, Sabaka TJ, Tangborn A, Tøffner-Clausen L, Thébault E, et al. (2010) International geomagnetic reference field: the eleventh generation. *Geophys J Int* 183(3):1216–1230
- Finlay CC, Jackson A, Gillet N, Olsen N (2012) Core surface magnetic field evolution 2000–2010. *Geophys J Int* 189:761–781. DOI 10.1111/j.1365-246X.2012.05395.x
- Finlay CC, Olsen N, Tøffner-Clausen L (2015) DTU candidate field models for IGRF-12 and the CHAOS-5 geomagnetic field model. *Earth, Planets and Space*, in press
- 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 Sci Rev* 155:247–291. <http://dx.doi.org/10.1007/s11214-010-9669-4>
- Fournier A, Aubert J, Thébault E (2011) Inference on core surface flow from observations and 3-D dynamo modelling. *Geophys J Int* 186:118–136

- Friis-Christensen E, Lühr H, Hulot G (2006) Swarm a constellation to study the Earth's magnetic field. *Earth Planet Space* 58:351–358
- Gillet N, Jault D, Finlay CC, Olsen N (2013) Stochastic modeling of the earth's magnetic field: inversion for covariances over the observatory era. *Geochem Geophys Geosyst* 14(4):766–786
- Gillet N, Jault D, Finlay CC (2015) Planetary gyre, time-dependent eddies, torsional waves and equatorial jets at the Earth's core surface. *J. Geophys Res*, in press
- Holme R (2007) Large scale flow in the core. In: Olson P, Schubert G (eds). *Treatise in geophysics, core dynamics*. Elsevier, chap 4 Vol. 8. pp 107–129
- Holme R, Bloxham J (1996) The treatment of attitude errors in satellite geomagnetic data. *Phys Earth Planet Inter* 98:221–223
- Jackson A, Jonkers ART, Walker MR (2000) Four centuries of geomagnetic secular variation from historical records. *Phil Trans R Soc Lond A* 358:957–990
- Kloeden PE, Platen E (1992) Numerical solution of stochastic differential equations, Vol. 23. Springer-Verlag, Berlin Heidelberg
- Korte M, Constable C, Donadini F, Holme R (2011) Reconstructing the holocene geomagnetic field. *Earth Planet Sci Lett* 312:497–505
- Le Mouél JL, Gire C, Madden T (1985) Motions of the core surface in the geostrophic approximation. *Phys Earth Planet Inter* 39:270–287
- Lesur V, Wardinski I, Hamoudi M, Rother M (2010) The second generation of the GFZ reference internal magnetic model: GRIMM-2. *Earth Planets Space* 62:765–773
- Licht A, Hulot G, Gallet Y, Thébault E (2013) Ensembles of low degree archeomagnetic field models for the past three millennia. *Phys Earth Planetary Inter* 224:38–67
- Nilsson A, Holme R, Korte M, Suttie N, Hill M (2014) Reconstructing holocene geomagnetic field variation: new methods, models and implications. *Geophys J Int* 198(1):229–248
- Olsen N (2002) A model of the geomagnetic field and its secular variation for epoch 2000 estimated from ørsted data. *Geophys J Int* 149(2):454–462
- Olsen N, Friis-Christensen E, Floberghagen R, Alken P, Beggan CD, Chulliat A, Doornbos E, da Encarnação JT, Hamilton B, Hulot G, et al (2013) The swarm satellite constellation application and research facility (scarf) and Swarm data products. *Earth Planets Space* 65(11):1189–1200
- Olsen N, Luhr H, Finlay CC, Sabaka TJ, Michaelis I, Rauberg J, Tøffner-Clausen L (2014) The CHAOS-4 geomagnetic field model. *Geophys J Int* 197:815–827
- Olson PL, Christensen UR, Driscoll PE (2012) From superchrons to secular variation: a broadband dynamo frequency spectrum for the geomagnetic dipole. *Earth Planet Sci Lett* 319–320:75–82. <http://www.sciencedirect.com/science/article/pii/S0012821X11007229>
- Ou J, Gillet N, Du A (2015) Geomagnetic observatory monthly series, 1930 to 2010: empirical analysis and unmodeled signal estimation. *Earth Planets Space* 67(1):13. DOI 10.1186/s40623-014-0173-z
- Pais MA, Jault D (2008) Quasi-geostrophic flows responsible for the secular variation of the Earth's magnetic field. *Geophys J Int* 173:421–443
- Papoulis A, Pillai SU (2002) Probability, random variable and stochastic process. 4th ed., McGraw-Hill, New York
- Pavón-Carrasco FJ, Osete ML, Torta JM, De Santis A (2014) A geomagnetic field model for the holocene based on archaeomagnetic and lava flow data. *Earth Planetary Sci Lett* 388:98–109
- Sabaka TJ, Olsen N, Tyler RH, Kuvshinov A (2015) Cm5, a pre-swarm comprehensive geomagnetic field model derived from over 12 yr of champ, ørsted, sac-c and observatory data. *Geophys J Int* 200(3):1596–1626
- Tanriverdi V, Tilgner A (2011) Global fluctuations in magnetohydrodynamic dynamos. *New J Phys* 13(3):033,019
- Thébault E, Finlay CC, Beggan C, Alken P, Aubert J, Barrois O, Bertrand F, Bondar T, Boness A, Brocco L, Canet E, Chambodut A, Chulliat A, Coisson P, Civet F, Du A, Fournier A, Fratter I, Gillet N, Hamilton B, Hamoudi M, Hulot G, Jager J, Korte M, Kuang W, Lalanne X, Langlais B, Léger JM, Lesur V, Lowes FJ, et al. (2015) International geomagnetic reference field: the twelfth generation. *Earth, Planets and Space*, accepted
- Velínský J, Finlay CC, Effect of a metallic core on transient geomagnetic induction (2011). *Geochem Geophys Geosyst* 12(5):Q05,011. DOI 10.1029/2011GC003557
- Waddington R, Gubbins D, Barber N (1995) Geomagnetic field analysis –V. Determining steady core surface flows directly from geomagnetic observations. *Geophys J Int* 122:326–350
- Yaglom AM (1962) An introduction to the theory of stationary random functions, Prentice-Hall, Englewood Cliffs NJ

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
