

Assimilation of ground and satellite magnetic measurements: inference of core surface magnetic and velocity field changes

O. Barrois,¹ M.D. Hammer,² C.C. Finlay,² Y. Martin¹ and N. Gillet¹

¹Univ. Grenoble Alpes, CNRS, ISTERRE, CS 40700, F-38058 Grenoble cedex 9, France. E-mail: oli.barrois@gmail.com

²Division of Geomagnetism, DTU Space, Technical University of Denmark, DK-2800 Lyngby, Denmark

Accepted 2018 July 19. Received 2018 July 16; in original form 2018 March 13

SUMMARY

We jointly invert for magnetic and velocity fields at the core surface over the period 1997–2017, directly using ground-based observatory time-series and measurements from the CHAMP and *Swarm* satellites. Satellite data are reduced to the form of virtual observatory time-series distributed on a regular grid in space. Such a sequential storage helps incorporate voluminous modern magnetic data into a stochastic Kalman filter, whereby spatial constraints are incorporated based on a norm derived from statistics of a numerical geodynamo model. Our algorithm produces consistent solutions both in terms of the misfit to the data and the estimated posterior model uncertainties. We retrieve core flow features previously documented from the analysis of spherical harmonic field models, such as the eccentric anticyclonic gyre. We find enhanced diffusion patterns under both Indonesia and Africa. In contrast to a steady flow that is strong under the Atlantic hemisphere but very weak below the Pacific, interannual motions appear evenly distributed over the two hemispheres. Recovered interannual to decadal flow changes are predominantly symmetrical with respect to the equator outside the tangent cylinder. In contrast, under the Northern Pacific we find an intensification of a high latitude jet, but see no evidence for a corresponding feature in the Southern hemisphere. The largest flow accelerations that we isolate over the studied era are associated with meanders, attached to the equatorward meridional branch of the planetary gyre in the Eastern hemisphere, that are linked to the appearance of an eastward equatorial jet below the Western Pacific.

Key words: Core; Magnetic field variations through time; Inverse theory; Probabilistic forecasting.

1 INTRODUCTION

Inferring information on the motions of the liquid outer core of the Earth requires properly separating the numerous sources of observed magnetic fields (geodynamo, crustal magnetization, ionospheric and magnetospheric currents and their Earth induced counterparts). To circumvent some of the leakage issues, magnetic field models are often built using regularizations, to ensure spectral convergence of the core field and its time variations. This prevents a proper assessment of *a posteriori* errors on model coefficients. When these are used as data in reconstructions of the core dynamics, it can lead to biased estimates. Furthermore, by proceeding in successive steps (to a field model and then on to the core flow), one loses information.

From the early 1990s alternative avenues of research arose, through which field models were built under topological constraints derived from physical insights. Constable *et al.* (1993) and O'Brien *et al.* (1997) proposed algorithms to apply, on single epoch pairs of models, magnetic flux conservation conditions at the core–mantle

boundary (CMB) that are appropriate assuming that magnetic diffusion is negligible. Along the same lines, Jackson *et al.* (2007) added a constraint on the radial vorticity. They showed that it was possible for a magnetic model to satisfy both these topological conditions, and the constraint from magnetic observations, from the late 19th century onwards.

Conversely, Chulliat & Olsen (2010) tested the validity of the frozen flux hypothesis using data from Magsat, Oersted and CHAMP satellite missions. They found an increase of the data misfit in some areas, potentially suggesting local failures of the constraint. Such studies motivated the coestimation, from magnetic observations, of both the field and the flow, imposing with a weak formalism the frozen flux radial induction equation at the CMB (Lesur *et al.* 2010; Wardinski & Lesur 2012). They concluded that the frozen flux constraint remained compatible with ground-based and satellite magnetic records. Pursuing an alternative approach, Beggan & Whaler (2009) and Whaler & Beggan (2015) obtained piecewise constant or linear flow models directly from magnetic data (see also Whaler *et al.* 2016).

One limitation though of such approaches is related to the uncertainties associated with the large-scale induction equation itself (and associated null-flux curves), assuming models truncated at spherical harmonic degree $n \simeq 13$ (Gillet *et al.* 2009). Subgrid-scale effects arising due to the nonlinear induction process (e.g. Eymin & Hulot 2005; Pais & Jault 2008; Gillet *et al.* 2009; Baerenzung *et al.* 2016) turn out to be the main source of uncertainty in the recovery of core surface flows from modern geomagnetic records. Barrois *et al.* (2017)—hereafter referred as BGA17—illustrate how ignoring their impact leads to severely biased flow models (see also Baerenzung *et al.* 2017, on the reliability of core flow reconstructions).

BGA17 furthermore show from the analysis of geodynamo simulations that magnetic diffusion at the core surface, enslaved to poloidal flow below the CMB, affects the recorded field changes at all timescales including rapid changes. This may seem at odds with the often used assumption of negligible magnetic diffusion that follows the argument of a high magnetic Reynolds number for large-scale motions in the core (see Holme 2015).

In the present work we invert, from magnetic field observations collected at and above the Earth's surface, for both the magnetic and velocity fields at the core surface, taking into account both magnetic diffusion and subgrid induction. We merge spatial information provided by numerical simulations, specifically from the Coupled Earth dynamo (CED) model (Aubert *et al.* 2013) and temporal constraints coming from a restriction of the field evolution to a chosen class of stochastic process. The sequential algorithm of BGA17, which considers as input data time-series of spherical harmonic coefficients of the main field, is extended to account for both virtual observatory (Mandea & Olsen 2006) and ground observatory time-series that cover the period 1997–2017. Our approach has similarities with the previous works of Gillet *et al.* (2015a) and Baerenzung *et al.* (2016), which favoured flat flow spatial spectra at the CMB, since the spatial dynamo norm employed here departs from the norms often employed to ensure spectral convergence. In addition, our stochastic framework allows us to discuss posterior model errors for both the flow and the magnetic field.

The paper is organized as follows. In Section 2 we describe the ground-based observatory data and satellite-based virtual observatory data, and the methodology we follow to recover magnetic and velocity fields at the CMB. In Section 3.1, we present our resulting geomagnetic model and its associated uncertainties, before we analyse in Section 3.2 our core flow solutions. Finally, implications for our understanding of the core dynamics and possible further improvements for the algorithm are given in Section 4.

2 METHODOLOGY

2.1 Ground-based and virtual observatory data

2.1.1 Ground observatory data

We use magnetic measurements made at 186 ground observatories (GOs) covering the period 1997–2017. Hourly mean values are taken from the BGS database¹, version 0111, using Intermagnet and WDC Edinburgh data as available in May 2017. The data have been checked and corrected for known baseline jumps (Macmillan & Olsen 2013). ‘Revised monthly means’ were then derived from these hourly means, following the procedure described by Olsen

et al. (2014). Briefly, predictions of the large-scale magnetospheric field (and the associated induced field) from the CHAOS-6 field model, as well as predictions for the ionospheric Sq field (and the associated induced field) from the CM4 model (Sabaka *et al.* 2004) are subtracted from the hourly mean values, and then robust (Huber-weighted) monthly mean values are computed using an iterative-reweighting procedure. Annual differences of such revised monthly means are routinely used in deriving the CHAOS series of field models and in order to study high resolution secular variation since, compared with simple monthly means, they are less contaminated by external field effects. Here, since we also wish to use the field itself for model construction, the median difference between each series and CHAOS-6 predictions was removed, in order to account in a simple way for the bias due to unmodelled crustal fields. In order to obtain the same sampling rate as that adopted for the virtual observatory series described below, the revised monthly mean series were finally averaged over 4 months windows to obtain the GO series used in our data assimilation scheme.

2.1.2 Virtual observatory data

In addition to GO data, we make use of satellite measurements from the CHAMP and *Swarm* missions covering respectively 2000–2010 and 2014–2017, through so-called virtual observatory (VO) data (Mandea & Olsen 2006; Olsen & Mandea 2007). These provide a regular spatial and temporal sampling of the global field, convenient for our Kalman filter algorithm (detailed in Section 2.2) and involve estimates from an easily manageable number of locations, which has computational advantages.

VO data were computed using measurements collected by the CHAMP vector field magnetometer between July 2000 and September 2010 and from the *Swarm* vector field magnetometers, onboard all three satellites (*Alpha*, *Bravo*, *Charlie*), between January 2014 and April 2017. Starting from the CHAMP MAG-L3 and *Swarm* Level 1b MAG-L, version 0501, data products, we subsampled at 15 s intervals the data in the vector field magnetometer (VFM) frame. Using the Euler rotation angles as given by the CHAOS-6-x3 model (which was based on *Swarm* and ground observation data up until April 2017²), we rotated the VFM data into an Earth-Centered Earth-Fixed (ECEF) coordinate frame.

Measurements from known problematic days were removed, for instance where satellite manoeuvres happened. Furthermore, gross data outliers with deviations more than 500 nT from CHAOS-6-x3 field model predictions were rejected. Based on previous studies of VO estimates (e.g. Beggan *et al.* 2009), we then employed data selection criteria retaining only data for which:

- (1) the sun was at maximum 10° above horizon;
- (2) geomagnetic activity index $K_p < 3^\circ$;
- (3) the RC disturbance index (Olsen *et al.* 2014) had $|dRC/dt| < 3 \text{ nT hr}^{-1}$;
- (4) merging electric field at the magnetopause $E_m \leq 0.8 \text{ mV m}^{-1}$, with $E_m = 0.33 v^{4/3} B_t^{2/3} \sin(|\Theta|/2)$. v is the solar wind speed, $\Theta = \arctan(B_y/B_z)$ and $B_t = \sqrt{B_y^2 + B_z^2}$. B_y and B_z are components of the interplanetary magnetic field (IMF) in the geocentric solar magnetospheric (GSM) coordinate system, calculated using 2

²<http://www.spacecenter.dk/files/magnetic-models/CHAOS-6/>

¹ftp://ftp.nerc-murchison.ac.uk/geomag/Swarm/AUX_OBS

hourly means of 1 min values of the IMF and solar wind extracted from the OMNI database³;

(5) IMF $B_z > 0$ nT and IMF $|B_y| < 10$ nT, again based on 2 hourly mean of 1 min values.

Following this data selection, estimates of the fields due to various unmodelled sources were next removed from the data:

(1) the magnetospheric and its induced fields as given by the CHAOS-6-x3 model;

(2) the ionospheric and its induced fields as given by the CM4 model (Sabaka *et al.* 2004);

(3) the static internal field for spherical harmonic degrees $n > 20$ given by the CHAOS-6-x3 model.

Although imperfect, in our opinion it is more consistent to remove such estimates rather to ignore known field sources.

Based on this data we then carried robust inversions for time-averaged point estimates (i.e. VOs) using data windows of 4 months width (60 d each side of an epoch t_j). In order to aid the robust inversion procedure in identifying and downweighting outliers, following Olsen & Manda (2007) as a pre-processing step, we also removed a time-dependent internal field, here taken from the CHAOS-6-x3 model (Finlay *et al.* 2016b), for spherical harmonic degrees 1 to 20, within each four month window. The CHAOS-6x-3 prediction at the target point and time was then added back at the end of the analysis. Note that this does not prevent our 4-monthly VO series, and the derived SV series from departing from CHAOS-6x-3; information about the time-dependence within each 4 month window is however lost.

We assume that the residual field $\tilde{\mathbf{B}}$, after the removal of the time-dependent internal field from the CHAOS-6-x3, can be represented as the gradient of a scalar potential V , that is,

$$\tilde{\mathbf{B}} = -\nabla V. \quad (1)$$

The residual field and associated positions are transformed into a local Cartesian coordinate system with origin at the VO points of interest, with x pointing towards geographic south, y pointing towards east and z pointing upwards. We use an expansion of the local potential up to cubic terms. Because the geomagnetic field is irrotational ($\nabla \times \tilde{\mathbf{B}} = 0$) and solenoidal ($\nabla \cdot \tilde{\mathbf{B}} = 0$), this local potential is entirely determined by 15 independent parameters:

$$\begin{aligned} V(x, y, z) = & v_x x + v_y y + v_z z + v_{xx} x^2 + v_{yy} y^2 - (v_{xx} + v_{yy}) z^2 \\ & + 2v_{xy} xy + 2v_{xz} xz + 2v_{yz} yz - (v_{xyy} + v_{xzz}) x^3 \\ & + 3v_{xxy} x^2 y + 3v_{xxz} x^2 z + 3v_{xyy} x y^2 + 3v_{xzz} x z^2 \\ & + 6v_{xyz} x y z - (v_{xxy} - v_{yzz}) y^3 + 3v_{yyz} y^2 z \\ & + 3v_{yzz} y z^2 - (v_{xxz} + v_{yyz}) z^3. \end{aligned} \quad (2)$$

For each VO position vector $\mathbf{r}_k = (\theta_k, \phi_k, r_k)$ and at epoch t_j , all data positioned within a cylinder of radius 850km ($\approx 7.5^\circ$) of the VO target \mathbf{r}_k , and within 60 days either side of t_j were used to build a local data vector $\mathbf{d}^{k,j}$. These data are then related to the 15 parameters defining the VO potential model $\mathbf{m}_{vo}^{k,j}$ at that site and epoch via $\mathbf{d}^{k,j} = \underline{\mathbf{g}}^{k,j} \mathbf{m}_{vo}^{k,j}$, where the elements of the matrix $\underline{\mathbf{g}}^{k,j}$ are determined from eqs (1) and (2).

Rather than working directly with $\mathbf{d}^{k,j}$ in deriving $\mathbf{m}_{vo}^{k,j}$ we make use of along-track (AT) and east–west (using *Swarm Alpha* and

Charlie only) sums and differences of the magnetic field components. An advantage of using field differences is that these have a reduced sensitivity to large-scale external signals, although data sums also need to be included in order to ensure sufficient information on the longer wavelengths core field. Using sums and differences has been found advantageous in a number of other field modelling efforts (Olsen *et al.* 2015; Sabaka *et al.* 2015). We calculate AT sums (Σ) and differences (Δ) as

$$\begin{cases} \Sigma d_i^{AT} = [\tilde{B}_i(\mathbf{r}, t) + \tilde{B}_i(\mathbf{r} + \delta\mathbf{r}, t + 15 \text{ s})]/2 \\ \Delta d_i^{AT} = [\tilde{B}_i(\mathbf{r}, t) - \tilde{B}_i(\mathbf{r} + \delta\mathbf{r}, t + 15 \text{ s})] \end{cases}. \quad (3)$$

$\tilde{B}_i = \mathbf{1}_i \cdot \tilde{\mathbf{B}}(\mathbf{r})$ are the residual magnetic field components in spherical polar coordinates (where $i = r, \theta$ or ϕ , and $\mathbf{1}_i$ are unit vectors). The east–west cross-track (CT) sums and differences between are calculated as

$$\begin{cases} \Sigma d_i^{CT} = [\tilde{B}_i^{\text{Alpha}}(\mathbf{r}_1, t_1) + \tilde{B}_i^{\text{Charlie}}(\mathbf{r}_2, t_2)]/2 \\ \Delta d_i^{CT} = [\tilde{B}_i^{\text{Alpha}}(\mathbf{r}_1, t_1) - \tilde{B}_i^{\text{Charlie}}(\mathbf{r}_2, t_2)] \end{cases}. \quad (4)$$

Here, for a given orbit of *Alpha* we select the corresponding *Charlie* data to be the one closest in latitude such that $|\delta t| = |t_1 - t_2| < 50$ s. Crucially, in order to relate these sums and differences to the VO model parameters, we also take sums and differences of the elements of the design matrices $\underline{\mathbf{g}}^{k,j}$ associated with the predictions of the VO model for the field components at the individual data locations. This results in a design matrix

$$\underline{\mathbf{G}}^{k,j} = \begin{bmatrix} \Sigma \underline{\mathbf{g}}^{k,j} \\ \Delta \underline{\mathbf{g}}^{k,j} \end{bmatrix} \quad (5)$$

associated with the data vector $\mathbf{D}^{k,j} = [\Sigma \mathbf{d}^{k,j} \quad \Delta \mathbf{d}^{k,j}]^T$. In this way we fully account for the change in the unit vectors associated with the two locations contributing to the sums and differences when deriving the parameters $\mathbf{m}_{vo}^{k,j}$. The inversion for each $\mathbf{m}_{vo}^{k,j}$ is carried out via a robust Huber weighted least-squares fit

$$\mathbf{m}_{vo}^{k,j} = \left[(\underline{\mathbf{G}}^{k,j})^T \mathbf{W} \underline{\mathbf{G}}^{k,j} \right]^{-1} (\underline{\mathbf{G}}^{k,j})^T \mathbf{D}^{k,j} \quad (6)$$

where \mathbf{W} is a diagonal vector of Huber weights that ensure a robust solution (Olsen 2002; Sabaka *et al.* 2004) and are iteratively updated until convergence. Once $\mathbf{m}_{vo}^{k,j}$ is determined, the three field components at the site and epoch of interest, $\tilde{\mathbf{B}}_k(\mathbf{r}_k, t_j) = -\nabla V_k(\mathbf{r}_k, t_j)$, are computed and added back on to the CHAOS-6-x3 prediction for the internal field (for degrees 1–14 only, to avoid as far as possible the lithospheric field) at the target location.

We constructed VO estimates at $P_{VO} = 200$ locations, with a spacing of about 1600 km ($\approx 14^\circ$, see dots in Fig. 1), located in an approximately equal area grid based on the spherical surface partition algorithm of Leopardi (2006). The altitude of the VOs are 300 and 500 km during the CHAMP and *Swarm* periods, respectively. Using predictions of the three components (B_r, B_θ, B_ϕ) of the magnetic field at P_{VO} locations, we finally obtain $3P_{VO}$ time-series (i.e. one point every 4 months during CHAMP and *Swarm* times, 48 epochs in all), stored in a vector $\mathbf{y}_{VO}(t)$. The SV was computed as annual differences of the 4 month time-series.

2.1.3 Uncertainty estimates for the GO and VO series

In order to obtain as much information as possible from the GO and VO data, while at the same time seeking to avoid overfitting them, it is important that appropriate uncertainty estimates are specified for each time-series. We define \mathbf{C}_{GO} and \mathbf{C}_{VO} to be the measurement error cross-covariance matrices for GO and VO data at each epoch,

³<http://omniweb.gsfc.nasa.gov>

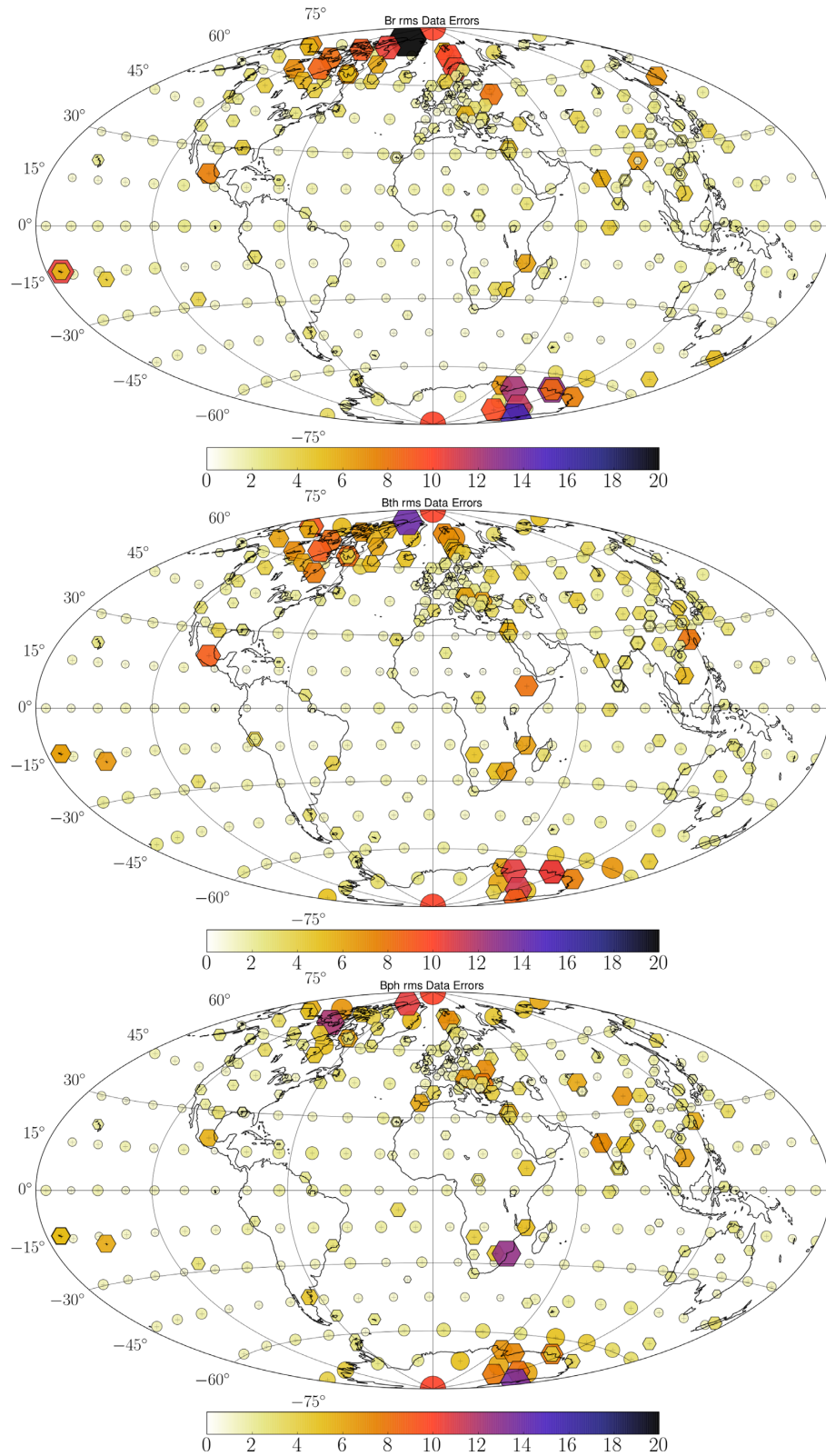


Figure 1. SV observation error estimates (colour scale in nT yr^{-1}) at all location where GOs (hexagons) and VOs (circles) are used in this study, for the three components \dot{B}_r , \dot{B}_θ and \dot{B}_ϕ (from top to bottom). The size of the markers is proportional to the magnitude of the *a priori* error estimates.

of sizes respectively $3P_{GO} \times 3P_{GO}$ and $3P_{VO} \times 3P_{VO}$. Data errors are supposed to be independent of time. Different data uncertainties are assigned for the VO's derived from CHAMP and *Swarm* respectively.

Regarding the GO time-series described above, we follow a similar approach to that used in CHAOS field model series (Olsen *et al.* 2014; Finlay *et al.* 2016b) and derive uncertainty estimates as follows. A 3×3 covariance matrix was computed for each observatory location from the time-series of the three components, after removing the predictions of the CHAOS-6 field model and detrending. The square roots of the diagonal elements of these covariance matrices were taken to be the uncertainty estimates for each component at each observatory. The same procedure was applied to both the MF and SV series.

For consistency, a very similar procedure was also applied to the VO series in order to obtain their uncertainty estimates. For each VO location, covariances were calculated between the time-series of the three components (after removing from each series the predictions of the CHAOS-6 model and de-trending), in order to obtain a 3×3 covariance matrix. A robust procedure for calculating the covariances (using the Minimum Covariance Determinant estimator, Verboven & Hubert 2005) was employed. However, only the square roots of the diagonal elements of the covariance matrices were taken to be the uncertainty estimates for each series, with similar procedures applied to both MF and SV series. To illustrate the range of the adopted uncertainty estimates, we show in Fig. 1 the r.m.s. SV uncertainty estimates for all locations where data (GO or VO) are used in this study.

Note that by using only the diagonal elements of \mathbf{C}_{GO} and \mathbf{C}_{VO} we effectively consider the errors on each GO and VO series to be uncorrelated with the errors on other series. In reality errors between components and between series will be correlated. This can be taken into account using full (i.e. dense) covariance matrices. It is however challenging to estimate cross-covariances for matrices of size larger than the length of the contributing times series (consisting of one sample every 4 months). We therefore postpone this step to future studies. Instead, by restricting to only 200 VO locations and ensuring that there was little overlap between the VO search radii we reduce as far as possible the correlations between distinct VO series.

Finally, we concatenate the above GO and VO main field data vectors for each epoch into $\mathbf{y}^o(t) = [\mathbf{y}_{VO}^T \mathbf{y}_{GO}^T]^T$. The associated observation errors covariance matrix $\mathbf{R}_{y,y}$, of rank $P = 3P_{VO} + 3P_{GO}$, is thus derived from the diagonals of \mathbf{C}_{VO} and \mathbf{C}_{GO} . In the next section we will consider both main field and secular variation data. SV data $\dot{\mathbf{y}}^o(t)$ are computed as annual differences of the four monthly (GO or VO) series. We follow the same approach as above to estimate the SV data errors variances (shown in Fig. 1) that are stored in a diagonal matrix $\mathbf{R}_{\dot{y},\dot{y}}$ of rank P .

2.2 Reanalysis of GO and VO data ground and satellite magnetic observations

The assimilation algorithm used in this study is essentially the one derived by BGA17 (see their table 2 for a summary). It is a sequential tool, consisting of a succession of forecast and analysis steps. The main modifications concern the direct integration of observations at and above the Earth's surface, while BGA17 considered data in the form of MF and SV spherical harmonic coefficients. We begin by recalling the main points of our stochastic forecast model, before we go on to describe the changes implemented in the present

study regarding the analysis step. These essentially concern the observation operator linking the state variables to the observations.

2.2.1 Stochastic forecast model

We forecast the evolution of the radial magnetic field, B_r , at the CMB using the radial component of the induction equation, written as

$$\frac{\partial \overline{B}_r}{\partial t} = -\overline{\nabla_h \cdot (\mathbf{u}_H \overline{B}_r)} + e_r + d_r(\mathbf{u}_H, \overline{B}_r), \quad (7)$$

where overlines mean the projection onto large length-scales. e_r stands for the subgrid induction processes arising due to the unresolved magnetic field at small length-scales, \mathbf{u}_H is the horizontal flow, and d_r , enslaved to \overline{B}_r and \mathbf{u}_H , approximates the radial component of the diffusion operator (see below). The evolutions of e_r and \mathbf{u}_H are governed by order one autoregressive stochastic processes,

$$\frac{de_r}{dt} + \frac{e_r}{\tau_e} = \zeta_e, \quad (8)$$

$$\frac{d\mathbf{u}_H}{dt} + \frac{(\mathbf{u}_H - \hat{\mathbf{u}}_H)}{\tau_u} = \zeta_u, \quad (9)$$

with ζ_e and ζ_u white noise processes, and $\hat{\mathbf{u}}_H$ the background flow model (obtained as the time-averaged flow from the CED model). These processes come from the same family of process as employed by Baerenzung *et al.* (2017). For each process, an effective restoring force is implemented via single time scales that we respectively fix as $\tau_e = 10$ yr and $\tau_u = 30$ yr. Spatial cross-covariances of the two above fields are derived from statistics of a free run of the CED (Aubert *et al.* 2013).

The advected fields e_r , \mathbf{u}_H , \overline{B}_r and d_r are represented through spherical harmonics, whose coefficients are stored in vectors $\mathbf{e}(t)$, $\mathbf{u}(t)$, $\mathbf{b}(t)$ and $\mathbf{d}(t)$, respectively. Diffusion in eq. (7), and its dependence on e_r and \mathbf{u}_H , is also an expression of cross-covariances extracted from the CED (involving the radial magnetic field below the CMB). The projection onto large length-scales is processed in the spectral domain, restricting the induction equation (and thus the expansion of the fields e_r , \overline{B}_r and d_r) to spherical harmonic degrees $n \leq n_b = 14$, while the velocity field is truncated at $n_u = 18$. We write as $\mathbf{b}(t)$ the vector of SV spherical harmonic coefficients.

2.2.2 Integrating ground and satellite data in the assimilation tool

We write as \mathbf{M} the operator that links the vector $\mathbf{b}(t)$ to the three components main field observations $\mathbf{y}(t)$ in the spatial domain (e.g. Olsen *et al.* 2010):

$$\mathbf{y}(t) = \mathbf{M}\mathbf{b}(t). \quad (10)$$

At each epoch it is of size $n_o \times n_b(n_b + 2)$, with $n_o = 3(P_{VO} + P_{GO})$ the size of the observation vector. The matrix \mathbf{M} is composed of submatrices \mathbf{M}_r , \mathbf{M}_θ and \mathbf{M}_ϕ , depending on the considered component of the magnetic field. In practice, elements of the matrix are, for a column j corresponding to a coefficient $g_{n_j}^{m_j}$, and a line i to an observation at a coordinate $\mathbf{r}_i = (r_i, \theta_i, \phi_i)$,

$$\mathbf{M}_{r_i,j} = (n_j + 1) \left(\frac{a^\oplus}{r_i} \right)^{n_j+2} \mathcal{P}_n^m(\theta_i) \cos(m_j \phi_i), \quad (11)$$

$$\mathbf{M}_{\theta_i,j} = \left(\frac{a^\oplus}{r_i} \right)^{n_j+2} \frac{d\mathcal{P}_n^m(\theta_i)}{d\theta} \cos(m_j \phi_i), \quad (12)$$

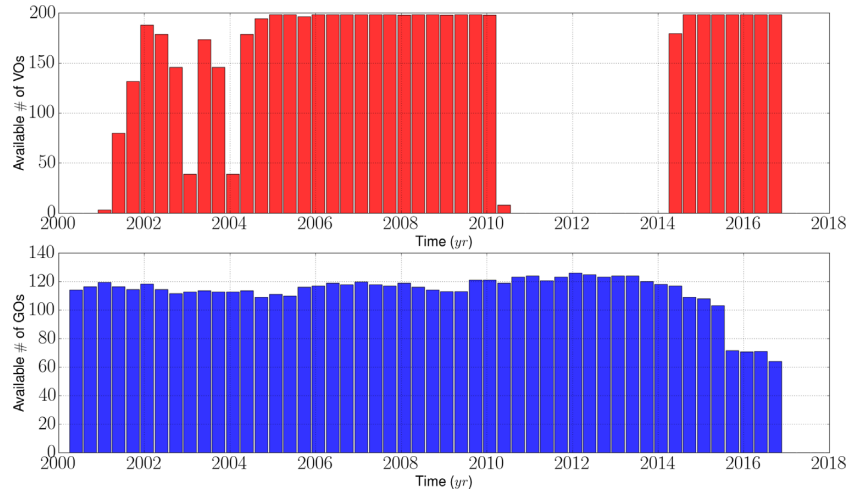


Figure 2. Time evolution of the number of SV data points (VOs in red, GOs in blue).

$$\mathbf{M}_{\phi_i, j} = \left(\frac{a^\oplus}{r_i} \right)^{n_j+2} \frac{m_j \mathcal{P}_n^m(\theta_i)}{\sin(\theta_i)} (-1) \sin(m_j \phi_i). \quad (13)$$

For a line j corresponding to a coefficient $h_{n_j}^{m_j}$, the function \sin replaces \cos in eqs (11) and (12), and \cos replaces $(-1)\sin$ in (13). $a^\oplus = 6371.2$ km is the Earth's spherical reference radius and \mathcal{P}_n^m are the Legendre polynomials.

The analysis in the Kalman filter algorithm employed by BGA17 consists of two steps: first an analysis of the vector \mathbf{b} containing MF spherical harmonic coefficients from MF spherical harmonic coefficients data, and second an analysis of the vector \mathbf{z} (that concatenates \mathbf{u} and \mathbf{e}) from SV spherical harmonic coefficients data. Writing as \mathbf{P}_{bb}^f the forecast model covariance matrix for \mathbf{b} , the first analysis (eq. 19 of BGA17) is replaced here by

$$\forall k \in [1, N_m], \mathbf{b}^{ka}(t_a) = \mathbf{b}^{kf}(t_a) + \mathbf{P}_{bb}^f \mathbf{M}^T \left[\mathbf{M} \mathbf{P}_{bb}^f \mathbf{M}^T + \mathbf{R}_{yy} \right]^{-1} (\mathbf{y}^{ko}(t_a) - \mathbf{M} \mathbf{b}^{kf}(t_a)), \quad (14)$$

with t_a the analysis epoch and the superscript k referring to the k th realization within an ensemble chosen to be of size $N_m = 50$. Writing as \mathbf{P}_{zz}^f the forecast model covariance matrix for \mathbf{z} , the second analysis (eq. 20 of BGA17) is replaced here by

$$\forall k \in [1, N_m], \mathbf{z}^{ka}(t_a) = \mathbf{z}^{kf}(t_a) + \mathbf{P}_{zz}^f \mathbf{G}^{kT} \left[\mathbf{G}^k \mathbf{P}_{zz}^f \mathbf{G}^{kT} + \mathbf{R}_{yy} \right]^{-1} (\delta \mathbf{y}^{ko}(t_a) - \mathbf{G}^k \mathbf{z}^{kf}(t_a)), \quad (15)$$

where the new observation operator is $\mathbf{G}^k = \mathbf{M} \mathbf{H}(\mathbf{b}^{ka})$, with \mathbf{H} as defined in BGA17. Here $\delta \mathbf{y}^{ko}(t_a) = \mathbf{y}^{ko}(t_a) - \mathbf{M} \mathbf{d}^{kf}(t_a)$ are the direct SV observations corrected by the forecast contribution from diffusion to the radial induction equation. This latter is sought iteratively at each analysis step, as in BGA17. Note that we consider an ensemble of observations \mathbf{y}^o and $\hat{\mathbf{y}}^o$, which are perturbed by random noise according to respectively \mathbf{R}_{yy} and $\mathbf{R}_{\hat{y}\hat{y}}$. We recall that we consider in eqs (14) and (15) forecast covariance matrices \mathbf{P}_{zz} and \mathbf{P}_{bb}^f that are frozen throughout the reanalysis period. These are derived directly from the CED cross-covariances on \mathbf{b} , \mathbf{u} and \mathbf{e} spherical harmonic coefficients, involving scaling pre-factors obtained analytically from the stochastic model presented in Section 2.2.1 (see BGA17 for details). For comparison, Baerenzung *et al.* (2017) employ a full implementation of the Ensemble Kalman filter (Evensen 2003), that is, they update the cross-covariances at each analysis step, requiring many more realizations to obtain well-conditioned matrices.

Finally, an extra complexity arises because the number of observation sites changes over time. Indeed, because of the selection criteria, the number of satellite data available may not always be sufficient to make a reliable VO estimate. Under these conditions the VO data point is considered to be absent: the associated elements of the data vector $\mathbf{y}^o(t)$ at a given time t are removed, together with the corresponding lines and columns of \mathbf{R}_{yy} , and the corresponding lines of the matrix \mathbf{M} (and thus \mathbf{G}). This procedure is performed during each analysis. Thus, the size P of the data vector changes through time, reflecting the changing number of available satellite observations through time (see Fig. 2).

To summarize, in this study we work with predictions made by spherical harmonic coefficients that are projected in physical space, where they are adjusted during the analysis step according to the observations and the covariance matrices. As such, our algorithm is still based almost entirely on the spectral domain; only the analysis steps are performed in physical space, in order to match the observed magnetic field data. Note that we corrected for two mistakes in the implementation of the algorithm by BGA17: a sign error in the background flow $\hat{\mathbf{u}}$, and off-diagonal elements of the covariance matrix for \mathbf{e} were non-intentionally ignored. Performing comparisons between reanalyses before and after correction, we found two consequences: a reduction of the dispersion within the ensemble of realizations, and a (almost stationary) shift in the analysed diffusion for some coefficients (including the axial dipole, see Section 3.1.2). This latter is almost entirely compensated by a shift in the analysed e_r , with minor impact on the recovered flow. Otherwise, the qualitative conclusions of BGA17 remain unaltered.

2.3 Posterior diagnostics

We now define several diagnostics used to evaluate the quality and the consistency of our results. We shall compare a quantity \mathbf{x} (MF, SV, subgrid error, diffusion... in the spatial or spectral domain) with observations \mathbf{x}^o (when available), or with the same quantities \mathbf{x}^c from the CHAOS-6 geomagnetic model (Finlay *et al.* 2016b). We define its time average

$$\hat{\mathbf{x}} = \frac{1}{t_f - t_i} \int_{t_i}^{t_f} \mathbf{x}(t) dt, \quad (16)$$

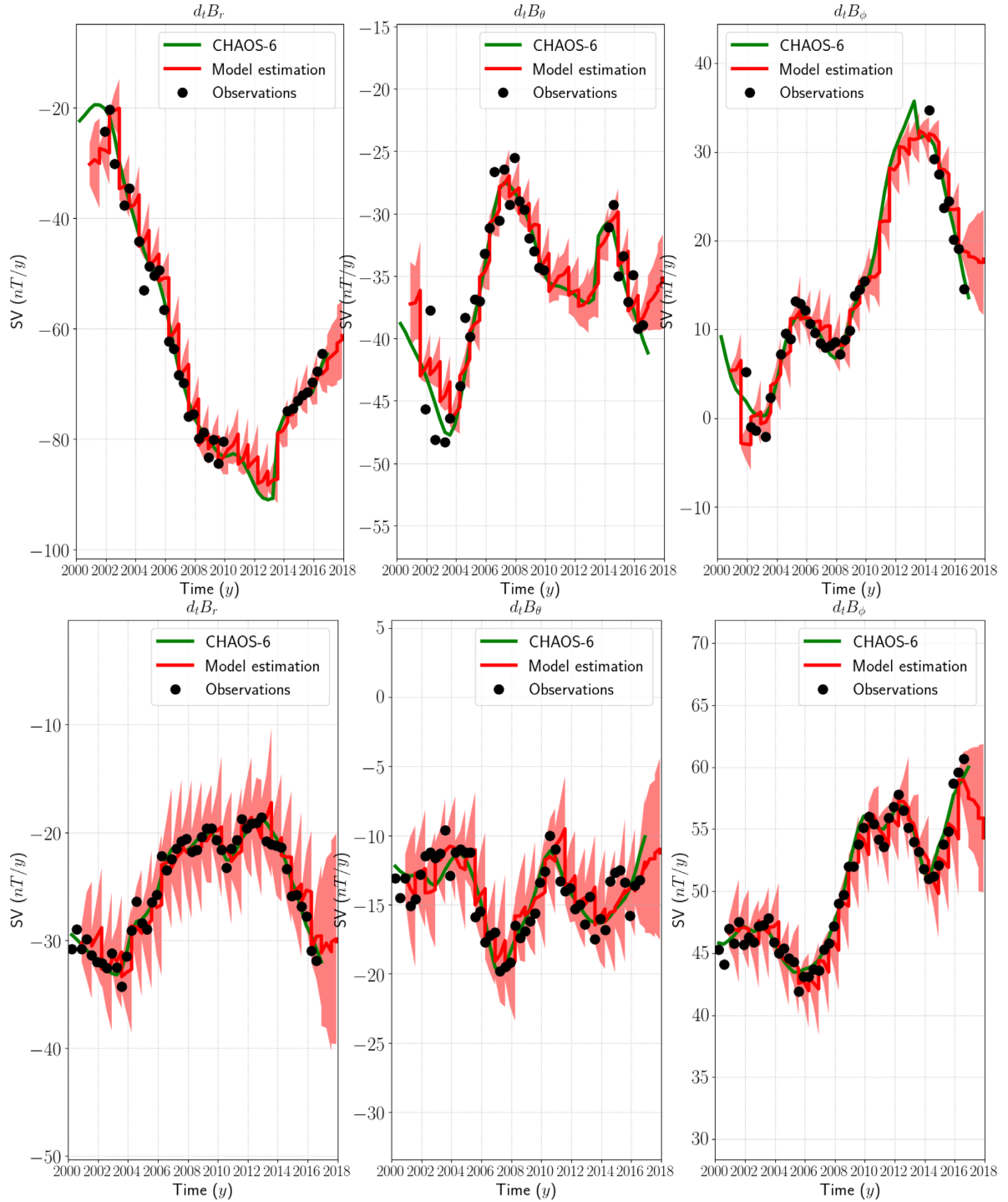


Figure 3. SV time-series for the three components (dB_r/dt , dB_θ/dt , dB_ϕ/dt), at one VO location $\{r = 6671 \text{ km}, \theta = 90^\circ, \phi = 88, 8^\circ\}$ (top), and at Chambon-la-forêt $\{r = 6366 \text{ km}, \theta = 42^\circ, \phi = 2^\circ\}$ (bottom). SV observations are shown in black, CHAOS-6 predictions in green and predictions from our analysis in red. The shaded area correspond to $\pm\sigma_j$, see eq. (18).

with t_i and t_f the initial and final epochs, its ensemble mean

$$\langle \mathbf{x}(t) \rangle = \frac{1}{N_m} \sum_{k=1}^{N_m} \mathbf{x}^k(t), \quad (17)$$

the dispersion within the ensemble

$$\sigma_x(t) = \sqrt{\frac{1}{N_m - 1} \sum_{k=1}^{N_m} (\mathbf{x}^k(t) - \langle \mathbf{x}(t) \rangle)^2}, \quad (18)$$

and finally the bias between our ensemble mean model and the reference \mathbf{x}^c ,

$$\delta_x(t) = \mathbf{x}^c - \langle \mathbf{x}(t) \rangle. \quad (19)$$

We also define spatial power spectra of any magnetic trajectory $\mathbf{b}(t)$ as

$$\mathcal{R}_b(n, t) = (n + 1) \left(\frac{a^\oplus}{c} \right)^{2n+4} \sum_{m=0}^n [g_n^m(t)^2 + h_n^m(t)^2], \quad (20)$$

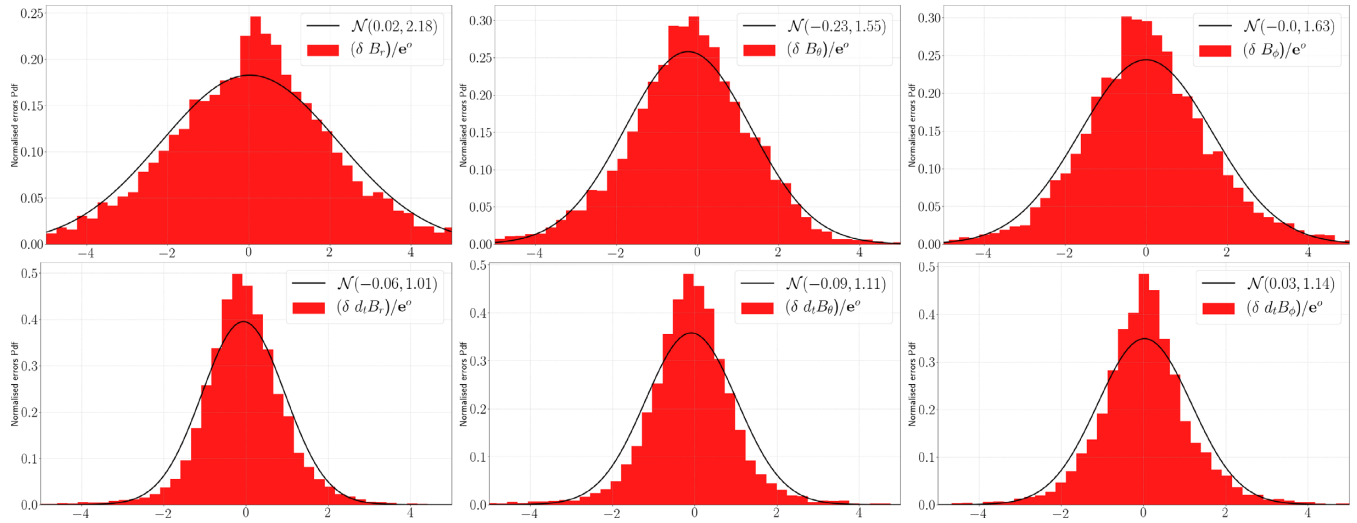


Figure 4. Top: histograms of MF prediction errors δ_{MF} (eq. 19), accumulated over all analysis epochs, normalized to the observation errors, for the components B_r (left), B_θ (middle) and B_ϕ (right). Superimposed in black are the Gaussian distribution fits obtained with the mean μ and the variance σ^2 for each of the three distributions. Bottom: same histograms for the SV prediction errors δ_{SV} .

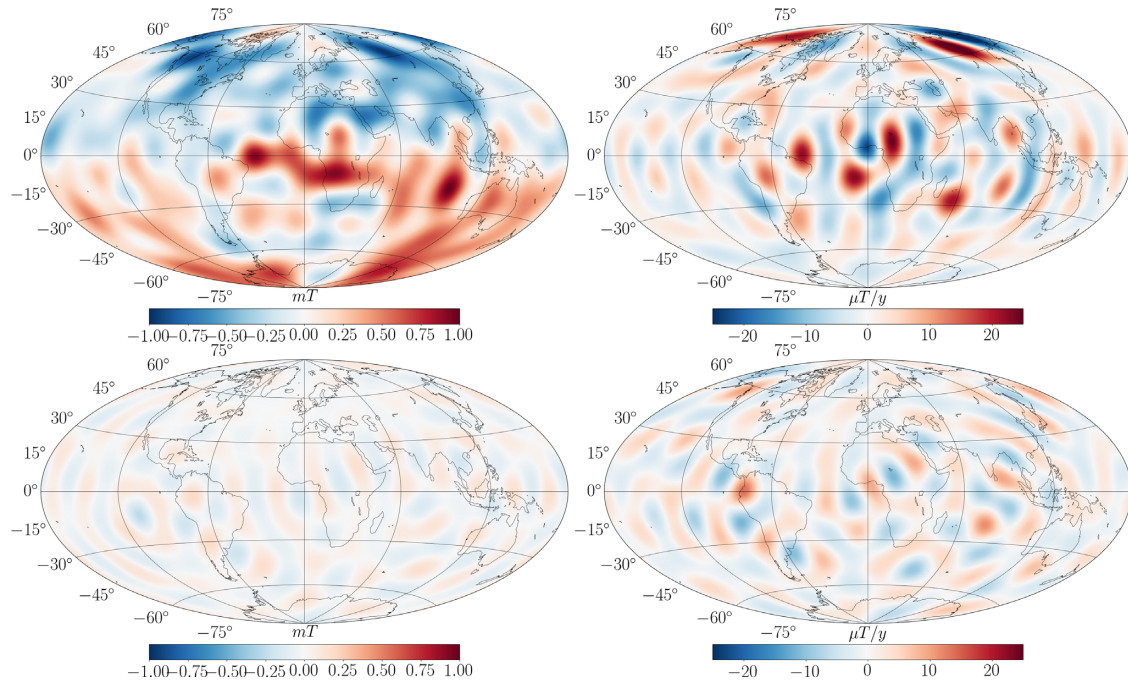


Figure 5. Top: CMB maps of the ensemble average radial magnetic field (\mathbf{b}) (eq. 17) in 2017 (left: MF in mT; right: SV in $\mu\text{T yr}^{-1}$), as estimated with our algorithm. Bottom: MF (left) and SV (right) maps of the difference of our ensemble average field with CHAOS-6 (truncated at degree 14) at the CMB (with the same colour scales).

with similar notations for $\mathbf{b}(t)$, $\mathbf{d}(t)$ and $\mathbf{e}(t)$. $c = 3485$ km is the Earth's core radius, and g_n^m and h_n^m are Schmidt semi-normalized spherical harmonic coefficients for the magnetic field at the Earth's surface. Finally, the spatial power spectrum for core flow trajectories \mathbf{u} writes

$$\mathcal{S}(n, t) = \frac{n(n+1)}{2n+1} \sum_{m=0}^n [t_{c_n}^m(t)^2 + t_{s_n}^m(t)^2 + s_{c_n}^m(t)^2 + s_{s_n}^m(t)^2], \quad (21)$$

with t_{c,s_n}^m and s_{c,s_n}^m Schmidt semi-normalized spherical harmonic coefficients for the toroidal and poloidal components of the flow.

We also define the flow norm

$$\mathcal{N} = \sum_{n=1}^{n_u} \frac{n(n+1)}{2n+1} \sum_{m=0}^n [t_{c_n}^{m^2} + t_{s_n}^{m^2} + s_{c_n}^{m^2} + s_{s_n}^{m^2}]. \quad (22)$$

The above power spectra can be considered for the ensemble mean or the dispersion within the ensemble, in which case they are respectively noted $\mathcal{R}_{\langle x \rangle}(n, t)$ and $\mathcal{R}_{\delta x}(n, t)$. Additionally, all those quantities may be averaged in time and/or computed only at analysis periods. For example, the time-averaged spatial power spectrum of the dispersion of magnetic field solutions at analysis epochs is $\overline{\mathcal{R}}_{\delta b}^a(n)$. The same convention as above holds for core flow spectra.

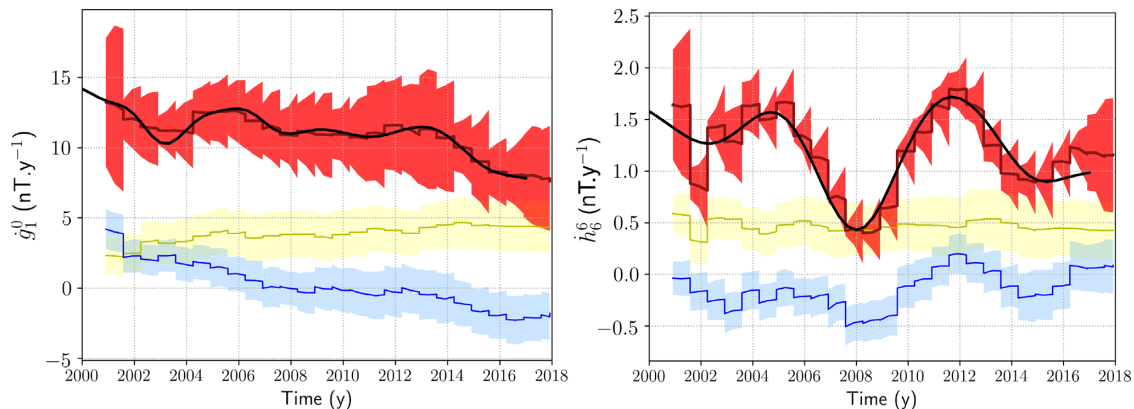


Figure 6. SV spherical harmonic coefficient time-series for g_1^0 (left) and h_6^6 (right). Predictions from our ensemble average model are shown in dark red ($\pm 2\sigma_b$ in red) and CHAOS-6 in black. Contributions from subgrid errors and diffusion extracted from our ensemble of realizations are superimposed in respectively blue and yellow (with dispersions $\pm 1\sigma_{\text{diff}}$ and $\pm 1\sigma_{\text{er}}$ in the corresponding colours).

3 RESULTS

We apply our algorithm to VO and GO magnetic field observations over a period spanning from $t_i = 1996.92$ to $t_f = 2016.92$. We recall that since we use satellite measurements from CHAMP and *Swarm* missions, VOs are available only over the periods 2000–2010 and 2014–2017, whereas GOs are available over the whole time span. Analysis are performed every $\Delta t^a = 4$ months. The sequences of analyses and forecasts between 1997 and 2001 are used to warm up the filter (see fig. 7 in BGA17), avoiding an increase in the ensemble spread over the first years of the targeted satellite era. This warm-up period is not considered below when interpreting the ensemble of inverted magnetic field and flow. We first describe predictions from our reanalysis for observations in the physical domain (Section 3.1.1), before we present the resulting magnetic model (Section 3.1.2), and insights on core flows over various timescales (Section 3.2).

3.1 Geomagnetic field models

3.1.1 Predictions for GO and VO series

We compare in Fig. 3 our series of SV forecasts and analysis with two examples of observation series (one VO and one GO), and with the predictions from CHAOS-6. The large spread of the SV forecasts is to be expected given the large uncertainties associated with subgrid errors and the large-scale flow (see BGA17). At both sites, the dispersion within the ensemble of SV trajectories encompasses most of the time the observations. Moreover the predictions from CHAOS-6 and from our ensemble of SV models are generally consistent. Our algorithm thus seems able to provide a coherent estimate of the SV probability density function (PDF) at the Earth’s surface and at satellite altitude. In addition, we highlight that even during the period 2010–2014 where no VO data are available, the trajectory of SV model, controlled by the stochastic prior and GO data only, remains reasonable, with a slight increase in the ensemble spread that always contains CHAOS-6. Note that our algorithm tends to drive the system toward low SV values (see the saw-tooth patterns in Fig. 3). This feature is to be expected given our choice of the stochastic models for \mathbf{u}_H and e_r , which control the evolution of the SV. In the absence of data constraints, the process will drift back the ensemble average trajectories for \mathbf{u}_H and e_r towards the average dynamo state, which by construction is responsible for a weak

SV. This is not a major drawback as soon as we analyse frequently enough, though it does limit the prediction capabilities of our tool (as discussed in BGA17).

We check in Fig. 4 the accuracy with which our model fits MF and SV observations, with the histograms of the prediction errors (over all analyses) normalized to the observation errors, for the three components of the magnetic field. Concerning the MF, prediction errors are only weakly biased, excepted for B_θ (normalized biases on the three components are $\mu_r = -0.02$, $\mu_\theta = -0.23$ and $\mu_\phi = 0.0$). The histograms of prediction errors are reasonably close to Gaussian for the three components with observation errors that appear to be under-estimated on average, in particular on B_r (normalized r.m.s. errors on the three components are $\sigma_r = 2.18$, $\sigma_\theta = 1.55$ and $\sigma_\phi = 1.63$). The SV predictions errors are remarkably consistent with the a priori errors with small biases and standard deviation close to unity for the three components ($\mu_r = -0.06$, $\sigma_r = 1.01$; $\mu_\theta = -0.09$, $\sigma_\theta = 1.11$ and $\mu_\phi = 0.03$, $\sigma_\phi = 1.14$), even though the distributions appears more peaked than a Gaussian. The Kalman filter employed here implicitly assumes Gaussian distributed data errors. However, the above remark suggests that alternative treatments of data residuals may be worth considering in future studies (e.g. L1 or Huber norms, see Constable 1988; Farquharson & Oldenburg 1998).

3.1.2 Field models, and contributions to the SV

We now describe in more detail our MF and SV models. We present in Fig. 5 MF and SV maps for our ensemble average model at the CMB truncated at spherical harmonic degree $n = 14$. Comparing it to a more traditional field model CHAOS-6, which is temporally regularized, the overall agreement is very good, indicating that our tool is indeed capable of producing reasonable field models. MF discrepancies to CHAOS-6 are relatively small, with peak to peak values less than 10% of the total amplitude for a field truncated at degree 14. They are dominated by isotropically distributed, small length-scale patterns. As well as being dominated by small length-scales, the disagreements are larger for the SV, with peak to peak differences about 30% of the total amplitude, which is to be expected given the blue SV spectrum at the CMB, meaning that small length scales dominate. Interestingly, the largest differences are localized under South America and the Indian Ocean, where the planetary gyre respectively detaches from and joins the equatorial belt (Pais

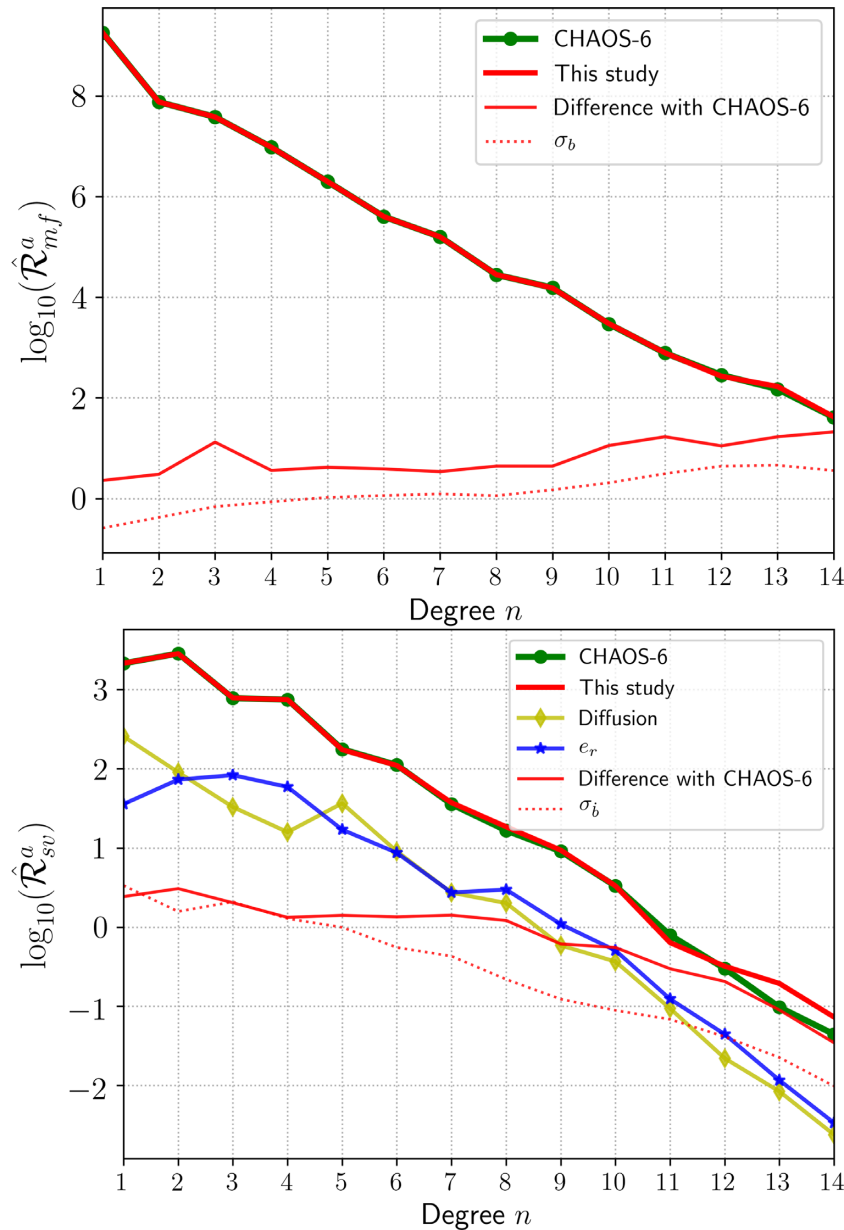


Figure 7. Top: time averaged spatial power spectra at the Earth surface of the magnetic field of CHAOS-6 ($\hat{\mathcal{R}}_{mf}^a$, eq. 20, in green), our estimate ($\hat{\mathcal{R}}_{(b)}^a$, in red), the difference between the two ($\hat{\mathcal{R}}_{(b)}^a - \hat{\mathcal{R}}_{mf}^a$, red thin line) and the dispersion within our ensemble of analyses ($\hat{\mathcal{R}}_{\sigma_b}^a$, dotted line). Bottom: *idem* for the SV, superimposed with the spectra of the contributions from subgrid errors (blue) and from diffusion (yellow).

& Jault 2008) and where rapid time-dependence is observed (Finlay et al. 2016a).

In Fig. 6 we show the various contributions in our model to two SV spherical harmonic coefficient series. The dispersion within the ensemble of models is large enough to include time changes as estimated by CHAOS-6, with some exceptions during the high solar activity era, for example, in 2002 for h_0^6 , and at the very end of the CHAOS-6 era (this latter possibly in link with the damping of SA towards end-points in the regularized field model). We note a larger spread of the analysis for the axial dipole than for non-zonal coefficients of intermediate length-scale such as h_0^6 . This may be a consequence of the weaker constraint on zonal coefficients from surface observations (e.g. Kotsiaros & Olsen 2012), although we only note such behaviour for g_0^1 . An enhancement of the dispersion is notable between 2010 and 2014, displaying in the spectral domain

the impact of the decreasing number of data during this era when no vector satellite data were available. Over 2001–2006, the ensemble average h_0^6 trajectory shows distinctive square shaped variations, probably partly related to variations in the number of data satisfying selection criteria during this interval of enhanced solar activity when only CHAMP data were available.

Spatial spectra shown in Fig. 7 summarize the characteristics of our model in the spectral domain. We find excellent agreement with CHAOS-6 for the main field and its secular variation, except at the small length scales of the SV ($n > 10$), which are more likely to be affected by the different data set chosen and by the different temporal kernel used (short time windows in our case against whole time span for CHAOS-6). The ensemble spread gives a good approximation of the characteristic distance between our model and CHAOS-6. Diffusion and subgrid errors in the SV have

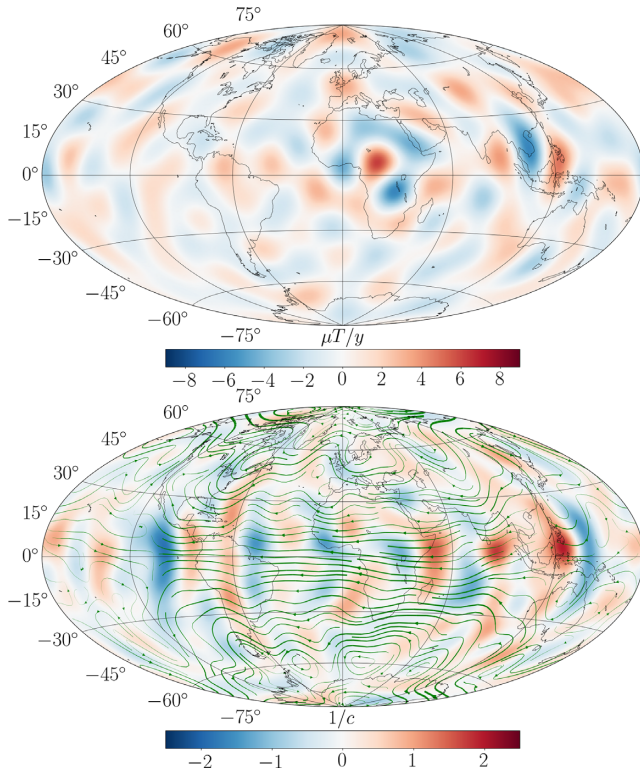


Figure 8. Magnetic diffusion at the CMB (top, colour scale in $\mu\text{T y}^{-1}$), and horizontal divergence $\nabla_h \cdot \mathbf{u}_h$ (bottom, colour scale in 10^{-3} yr^{-1}) superimposed with passive tracers trajectories (black, tracer size scale in km yr^{-1}), for the ensemble average model in 2017. Core flow visualizations are performed using the tools provided at <https://geodyn.univ-grenoble-alpes.fr/>. The size of the tracers is proportional to the velocity field (see the legend). The initial positions of the tracers is random; each trajectory is advected by the velocity field for a fixed time; along each trajectory, the late (early) positions are darker (lighter).

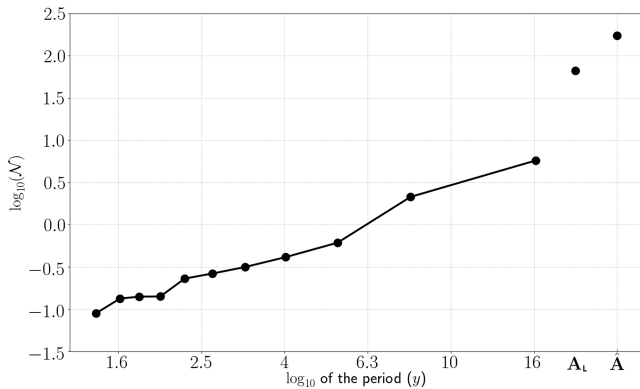


Figure 9. Core flow norm \mathcal{N} for all flow constituents that enter eq. (23). The norm \mathcal{N} for the linear flow acceleration is obtained by integrating the linear trend over the 16 yr.

approximately the same amplitude except for the dipole. The power stored in these two SV sources represents about 10 to 20% of the total SV energy at all scales.

Even though the dispersion within the model predictions is large enough to encompass most of the MF and SV observations, the dispersion within the ensemble of realizations is lower, by a factor about 2.5, than the distance between the ensemble average model and CHAOS-6 for both the MF (at all length-scales) and the SV

(towards small length-scales only). A complete account of SV errors from all subgrid interactions (see Baerenzung *et al.* 2017) may help reduce the above under-estimation. Our current estimate is nevertheless larger than that obtained for the COV-OBS.x1 model Gillet *et al.* (see fig. 4 in 2015a, the error spectrum in 2010). We suspect that the accumulation of data (assumed independent) during the construction of this latter field model involved too strong a decrease of the posterior error within the COV-OBS framework. The more consistent approach to error propagation developed here and presented in Fig. 7 favours larger uncertainties on spherical harmonic coefficients during the satellite era.

Overall, we are generally able to retrieve earlier well-established results. For instance the contribution from advection dominates (over diffusion) the axial dipole decay (Finlay *et al.* 2016b; Barrois *et al.* 2017) and its fluctuations—even though our estimate for the contribution from diffusion to dg_0^0/dt , shifted upward by a couple of nT yr^{-1} in comparison with the results of BGA17 (see Section 2.2.2), amounts to a relatively larger fraction over the latest years where the dipole decay tends to be weaker. The ensemble average SV originating from diffusion is presented in Fig. 8 for 2017: the most significant contributions appear below Africa and Indonesia. The strongest diffusion appears linked to intense patches of up-/downwellings in the equatorial belt at the CMB (see Fig. 8) and/or where strong gradient of \mathbf{B} occur. This is a direct consequence of our estimation of diffusion through cross-covariances involving core surface velocity and magnetic fields (see BGA17 and Amit & Christensen 2008). In the framework of our modelling, such diffusion patterns seem to be required by magnetic observations rather by the imposed prior cross-covariances (or if it is the case, it does not show up in the background state).

3.2 Core flow solutions

Next, we study with more details the temporal information contained in our core flow solutions. The idea is to extract an average signal and a linear acceleration, together with the flow at different periods, to check if we witness any preferential frequency, or if the characteristics of the flow change with the period. To do so, we apply a least-squares regression to our core flow solution with a function of the form

$$\mathbf{u}(t) = \hat{\mathbf{A}} + \mathbf{A}_L(t - t_0) + \sum_{k=1}^{11} \left[\mathbf{A}_k^s \sin\left(2\pi(t - t_0)\frac{k}{T}\right) + \mathbf{A}_k^c \cos\left(2\pi(t - t_0)\frac{k}{T}\right) \right], \quad (23)$$

with $t_0 = (t_i + t_f)/2 = 2008.92$ and $T = t_f - t_i = 16 \text{ yr}$. Vectors $\hat{\mathbf{A}}$, \mathbf{A}_L , \mathbf{A}_k^c and \mathbf{A}_k^s store respectively the spherical harmonic coefficients of the time average velocity, time average flow acceleration, and cosines and sines from periods 16 yr (for $k = 1$) to 1.45 yr (for $k = 11$) – of course the longer periods are not well constrained given the short time span considered here.

We show in Fig. 9 the norm (22) of all flow constituents for the ensemble average solution. The flow is dominated by long periods, translating onto core surface motions the red SV temporal spectrum (see Gillet *et al.* 2015a; Lesur *et al.* 2017). In comparison with a r.m.s. time average flow of 11.1 km yr^{-1} , the linear acceleration \mathbf{A}_L corresponds, integrated over 16 yr, to a r.m.s. flow increment of 6.6 km yr^{-1} .

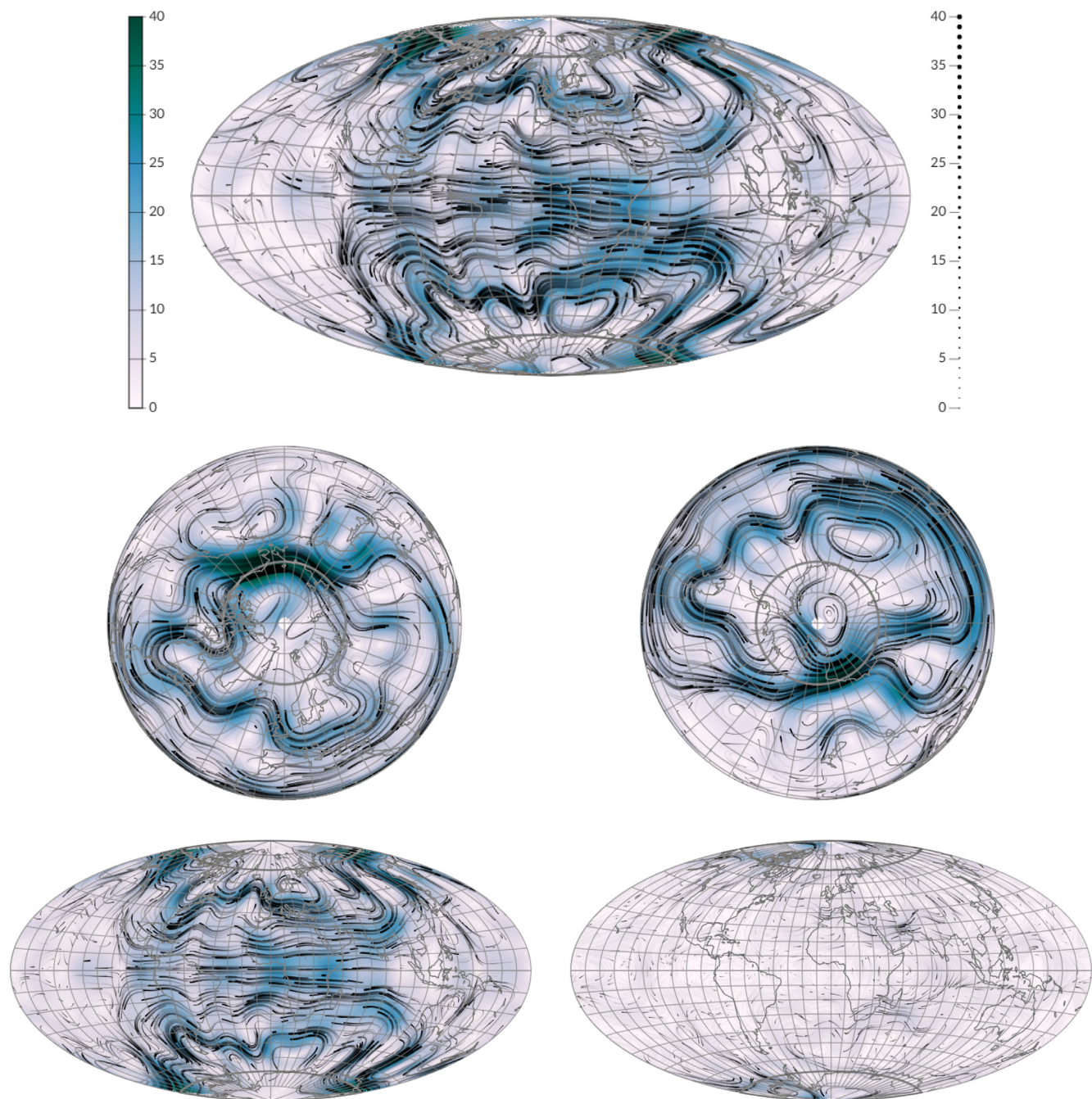


Figure 10. Intensity maps at the CMB of the flow constituents \hat{A} (in km yr^{-1}) for the ensemble average flow solution, superimposed with passive tracers trajectories (black). Top: Aitoff projection. Middle: north (right) and south (left) polar projections. Bottom: Aitoff projection for equatorially symmetric (left) and antisymmetric (right) components. The colour scale and tracer size scale are the same for all subfigures.

3.2.1 Stationary motions, and flow model uncertainties

We show in Fig. 10 core surface maps of the flow intensity and tracers trajectories for the ensemble average flow constituents \hat{A} . We retrieve on the map for the time average flow classical features, such as the westward gyre offset towards the Atlantic Ocean found in many studies (e.g. Pais & Jault 2008; Aubert 2014; Gillet *et al.* 2015b; Baerenzung *et al.* 2017), with a Pacific hemisphere that is on average much less energetic. The most energetic flow features are associated with (i) azimuthal motions in the equatorial belt below Africa, (ii) high latitudes azimuthal jets in the Pacific

hemisphere and (iii) meridional circulations, poleward (resp. equatorward) around 90°W (resp. 90°E).

Our solution is dominated by equatorially symmetric features (see Fig. 10, bottom), as expected outside the tangent cylinder (or TC, the cylinder tangent to the inner core, whose axis coincides with the rotation axis) when rotation forces dominates the momentum balance (e.g. Pais & Jault 2008). Nevertheless, the symmetry may be locally broken. The most striking examples of this are anticyclonic circulations within the TC, retrieved in both the Northern and Southern hemispheres (Fig. 10, middle). In contrast with polar

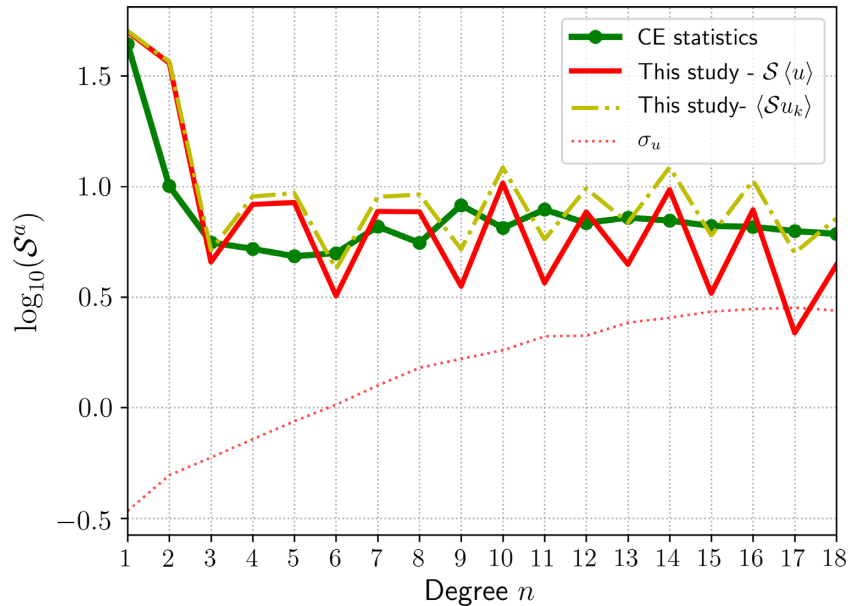


Figure 11. Time averaged spatial power spectra for ensemble average core flows (\hat{S}_{u}^a , red thick line) and the spectra for the ensemble average of each realization ($\langle \hat{S}_{u_k}^a \rangle$, yellow thick dotted line, eq.21), obtained from the reanalysis of VO and GO data. Spectra for the corresponding dispersion within the ensembles of models are displayed in dotted lines. In green is shown the averaged spectrum for the prior CED.

vortices previously inferred from geomagnetic observations (Olson & Aurnou 1999; Amit & Olson 2006), features we isolate here are offset to one side of the polar caps (i.e. they contain an important $m = 1$ contribution). This is a common configuration for polar vortices found in the most up to date numerical simulations (Schaeffer *et al.* 2017), which show much variability through epochs.

We show in Fig. 11 the time-average spatial power spectra for the ensemble average solution and for the dispersion within the ensemble of models. The former is comparable with the spectrum of the prior CED. The latter indicates that uncertainties, as measured by the ensemble spread, constitute a large fraction of the flow magnitude for degrees $n \geq 10$. The oscillation in the power seen between odd and even degrees might be magnified by possibly too low subgrid error budget (see Section 3.1.2).

3.2.2 On transient core surface motions

We now explore transient flow motions. We particularly focus on the amount of equatorial symmetry of our solutions inside and outside the TC, in order to detect if our model is sensitive to the specific geometry of the Earth's core (does it hold a signature of the TC?). As for the time-average flow, the linear acceleration over the past 16 yr is primarily symmetric with respect to the equator (see Fig. 12). The largest contributions consist of accelerating circulations around the meridional, Eastern branch of the gyre. Associated with these time-changing eddies around the equatorward branch of the planetary gyre, an Eastward equatorial jet intensifies under the Western Pacific. This suggests an underlying dynamics more complex than a simple longitudinal shift of the planetary gyre.

Interestingly, our average solution does not show a major intensification of equatorially symmetric azimuthal jets at high latitudes in the Pacific hemisphere, as inferred by Livermore *et al.* (2017). Indeed, we see an increase of the Northern jet only, by about 67%

in average (the one σ dispersion within the ensemble of flow realizations allowing for an increase up to 100%). Although still an appreciable acceleration, it is significantly less than the factor of 3 found by Livermore *et al.* The disagreement is likely due to our global inversion (in opposition to their local model). The difference seems to be related with antisymmetric circulations within the TC. One should keep in mind that in these high and low latitude areas, gradients of B_r are much larger in the Northern Hemisphere, meaning that the signature of any motions near the TC below the Southern Pacific are significantly weaker. As for the stationary constituent, the equatorial symmetry is not perfectly respected, and we retrieve the largest antisymmetrical features within the TC, associated with polar jets.

We give in Fig. 13 an example of one interannual flow constituent at the CMB for a period of 5.3 yr. In this case, the most energetic flows are concentrated into non-axisymmetric azimuthal jets near the equator (already highlighted by Gillet *et al.* 2015b; Finlay *et al.* 2016b), and into localized circulations at mid and high latitudes. These are not confined to the Atlantic hemisphere: despite being less energetic on average, the Pacific hemisphere shows interesting interannual flow variations. At these sub-decadal periods, we have not detected any obvious propagation of non-zonal flow patterns, which might be interpreted as the signature of azimuthally propagating waves (as advocated for by Chulliat & Maus 2014; Chulliat *et al.* 2015). The other periods display globally the same kind of features and no particular behaviour is found at any period. At these timescales also show up less intense antisymmetric features; the most significant shows up in the equatorial area (for instance under the Atlantic ocean and the Western Pacific), and towards high latitudes on the edge of the TC.

Fig. 14 summarizes the amount of equatorial symmetry found in regions inside and outside the TC, for our core flow solutions at all periods. It appears almost independent of the considered period: outside the TC, it is within 90 to 95% of the surface energy

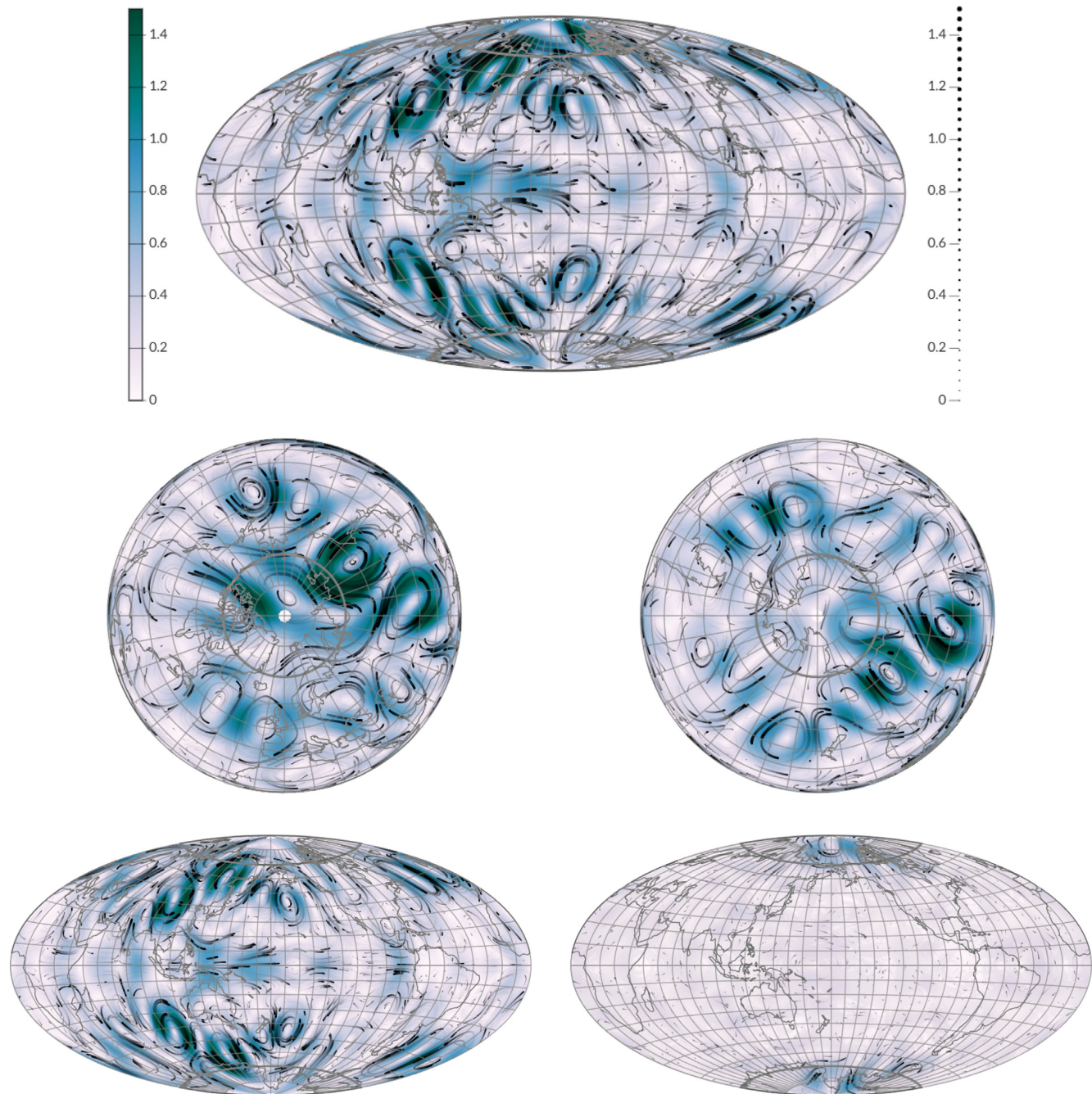


Figure 12. Same as Fig. 10 for the flow constituent A_L (in km yr^{-2}).

for all flow constituents of eq. (23). The partition of energy between symmetric and antisymmetric flow components is more balanced inside the TC where, depending on the considered timescale, $\approx 55 \pm 15\%$ of the energy is contained in equatorially symmetric flows. This latter observation could be expected because the presence of the inner core is intended to partially break the equatorial symmetry. However, it is remarkable that the algorithm appears accurate enough to detect a specific behaviour within the tiny areas covered by polar caps. Moreover, although our ensemble average model and the CED show very similar amounts of equatorial symmetry outside the TC (the value for the CED model is 95% of symmetrical flows inside and outside TC), they differ

significantly inside the TC (it is much less in the inverted flows). As a consequence, the larger proportion of equatorial antisymmetry inside the TC is driven by observations (against the prior information).

4 SUMMARY AND DISCUSSION

Following earlier strategies for geomagnetic field model reconstruction (e.g. Jackson *et al.* 2007; Lesur *et al.* 2010), and moving towards geomagnetic data assimilation (Aubert 2015; Gillet *et al.* 2015a; Baerenzung *et al.* 2017), we continue the work initiated in BGA17. We retain their idea of combining spatial information from

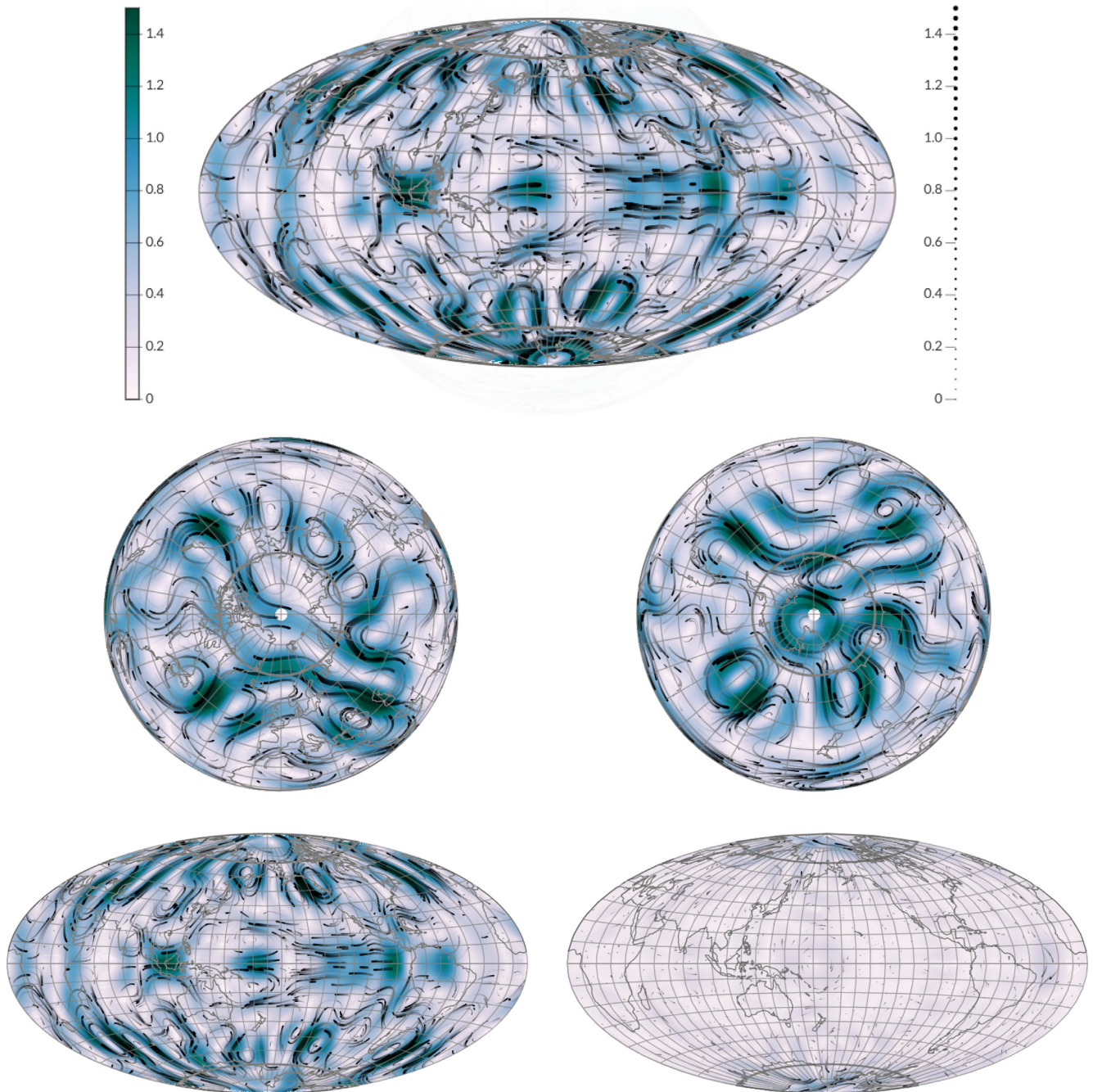


Figure 13. Same as Fig. 10 for the flow constituent A_{ξ}^{ξ} (in km yr^{-1}).

numerical simulations of the geodynamo with temporal information implemented through stochastic equations, chosen to replicate the frequency spectrum of ground-based geomagnetic series. However, instead of considering spherical harmonic coefficients of the main field as data, here we have inverted observations (GOs and VOs) directly, at and above the Earth's surface. In this respect we follow the studies by Beggan & Whaler (2009) and Whaler & Beggan (2015), although we account for subgrid processes (of great importance, as shown by BGA17 or Baerenzung *et al.* 2016) and for surface magnetic diffusion. This avenue allows us to propose PDFs for the main field and its secular variation, as well as for the recovered core motions.

4.1 Geophysical insights

The MF models presented here are consistent both with observations and with the imposed dynamical prior. The model uncertainties, as suggested by the ensemble spread, are slightly less than the distance of the average model to CHAOS-6. We recover in our core flow solutions a westward gyre that circulates around the TC at high latitudes in the Pacific hemisphere, and flows closer to the equator in the Atlantic hemisphere. The largest contributions from magnetic diffusion are associated with up-/downwellings where the gyre meets the equatorial region (under Indonesia) and in the equatorial region below Africa. At all timescales, the flow is predominantly symmetric with respect to the equator, except inside the TC where the

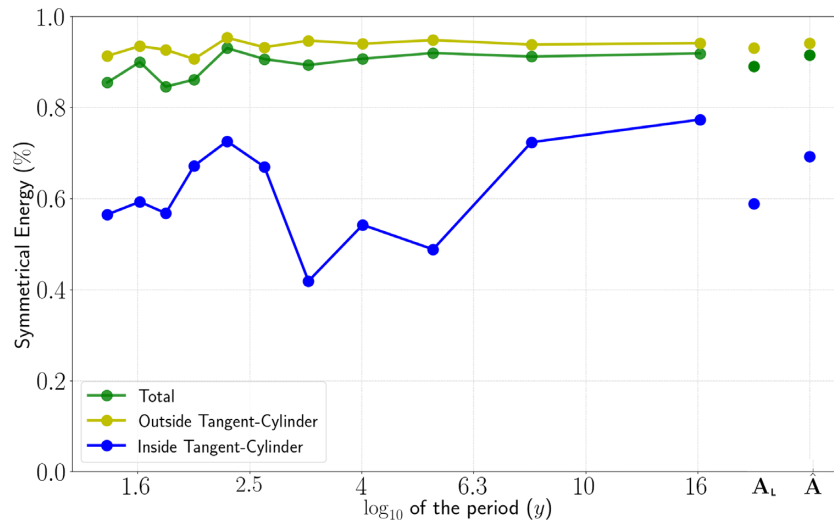


Figure 14. Fraction of energy contained into the equatorial symmetric part of the flow, inside (blue line) and outside (yellow line) of the tangent cylinder (TC), for each of the flow constituent that enters eq. (23). The total symmetric part of the flow is also displayed in green. The value for the CED dynamo used as a prior is 0.95 both inside and outside the TC.

situation is more balanced (contrary to our dynamo prior that is mostly symmetric everywhere).

The most intense time-average flow acceleration over the past 16 yr is linked with evolving meanders around the equatorward branch of the gyre in the Eastern hemisphere, also associated with the appearance of an Eastward equatorial jet under the Western Pacific. We do find a decadal intensification of jets near the TC, although the magnitude of the acceleration we infer is lower than that estimated by Livermore *et al.* (2017) with their reduced model. In our study, it is furthermore confined to the Northern Hemisphere. This equatorial asymmetry may be interpreted as the signature of an ageostrophic acceleration, keeping in mind that main field gradients are weak in the Southern Pacific, implying a weaker constraint on flow motions there (see fig. 7 in Baerenzung *et al.* 2016). However, because our prior does not show any particular bias in those areas, it is likely that those features are mostly driven by the data. On interannual periods, we find relatively energetic flow changes in both the Atlantic and the Pacific hemispheres, with both non-zonal equatorial jets and time-dependent mid-to-high latitudes eddies evident.

4.2 Future work

We currently lack a physical understanding for the features described above, whether it be through quasi-geostrophic flows (e.g. Labbé *et al.* 2015), motions within a stratified layer (e.g. Buffett & Knezek 2017), or any other interpretation through a reduced model. We also lack suitable long coverage by high quality satellite records to perform spectral analyses with a refined sampling in the frequency domain, which would allow us to isolate possible waves at interannual periods. Development of such reduced models, and their coupling with stochastic processes for modelling unresolved processes, will be an important next step in our ability to understand and predict geomagnetic field changes.

Meanwhile, our stochastic model itself could be improved; in particular it is desirable to avoid driving back the average trajectory towards an average dynamo simulation. This is indeed an unlikely state for the current era (say over decadal to centennial timescales), which might be better represented by a reanalysis of for instance

centennial motions from historical records (Jonkers *et al.* 2003). Furthermore, because of the short time span covered today by satellite data, we found it challenging to derive well-conditioned matrices for VO uncertainties. This is a key-point for such data assimilation studies, which calls for further developments, for example, through projections onto reduced basis in the data space. Alternatively, we may wish to coestimate, together with the core state, time-dependent external fields. Although possible, this calls for a severe re-encoding of both the forecast and analysis steps, in order to integrate satellite measurements along the tracks.

The general philosophy of our work is to retrieve information on the state of the Earth's core, and to provide realistic uncertainties on all state variables in a simple way. The encouraging magnetic models obtained with this approach render our algorithm suitable for deriving candidates to the International Geomagnetic Reference Field (Thébault *et al.* 2015). Remaining in a stochastic framework, modifications of the forward model parametrization—such as accounting for a background state closer to the flow responsible for the magnetic field over the past decades—may extend the prediction capability of our algorithm. However, targeting accurate field predictions one will have to resort to deterministic (i.e. dynamically based) equations for the core state.

ACKNOWLEDGEMENTS

We thank Julien Aubert for providing the Coupled-Earth dynamo series used to build the core state statistics, and Loic Huder for spotting two errors in the code at the origin of the results of BGA17. We also thank Julien Baerenzung and an anonymous referee for their useful comments that helped improve the quality of our manuscript. We would like to thank as well GFZ German Research Centre for Geoscience for providing access to the CHAMP MAG-L3 data and to ESA for providing access to the Swarm L1b MAG-L data. We also like to thank the staff of the geomagnetic observatories and INTERMAGNET for supplying high-quality observatory data. NG and OB were partially supported by the French Centre National d'Etudes Spatiales (CNES) for the study of Earth's core dynamics in the context of the *Swarm* mission of ESA. ISTERre is part of Labex OSUG@2020 (ANR10 LABX56), which with the

CNES also finance the PhD grant of OB. Numerical computations were performed at the Froggy platform of the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>) supported by the Rhône-Alpes region (GRANT CPER07 13 CIRA), the OSUG@2020 Labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01). MH and CF were supported by the Danish Council for Independent Research - Natural Sciences, Grant DFF-4002-00366.

REFERENCES

- Amit, H. & Christensen, U.R., 2008. Accounting for magnetic diffusion in core flow inversions from geomagnetic secular variation, *Geophys. J. Int.*, **175**(3), 913–924.
- Amit, H. & Olson, P., 2006. Time-average and time-dependent parts of core flow, *Phys. Earth planet. Inter.*, **155**, 120–139.
- Aubert, J., 2014. Earth's core internal dynamics 1840–2010 imaged by inverse geodynamo modelling, *Geophys. J. Int.*, **197**, 1321–1334.
- Aubert, J., 2015. Geomagnetic forecasts driven by thermal wind dynamics in the Earth's core, *Geophys. J. Int.*, **203**(3), 1738–1751.
- Aubert, J., Finlay, C.C. & Fournier, A., 2013. Bottom-up control of geomagnetic secular variation by the Earth's inner core, *Nature*, **502**(7470), 219–223.
- Baerenzung, J., Holschneider, M. & Lesur, V., 2016. The flow at the Earth's core-mantle boundary under weak prior constraints, *J. geophys. Res.*, **121**(3), 1343–1364.
- Baerenzung, J., Holschneider, M., Wicht, J., Sanchez, S. & Lesur, V., 2017. Modeling and predicting the short term evolution of the Geomagnetic field, *J. geophys. Res.*, **123**, 4539–4560.
- Barrois, O., Gillet, N. & Aubert, J., 2017. Contributions to the geomagnetic secular variation from a reanalysis of core surface dynamics, *Geophys. J. Int.*, **211**(1), 50–68.
- Beggan, C. & Whaler, K., 2009. Forecasting change of the magnetic field using core surface flows and ensemble Kalman filtering, *Geophys. Res. Lett.*, **36**(18), 1–5.
- Beggan, C., Whaler, K. & Macmillan, S., 2009. Biased residuals of core flow models from satellite-derived virtual observatories, *J. geophys. Int.*, **177**(2), 463–475.
- Buffett, B.A. & Knezek, N., 2017. Stochastic generation of MAC waves and implications for convection in Earth's core, *Geophys. J. Int.*, **212**, 1523–1535.
- Chulliat, A. & Maus, S., 2014. Geomagnetic secular acceleration, jerks, and a localized standing wave at the core surface from 2000 to 2010, *J. geophys. Res.*, **119**(3), 1531–1543.
- Chulliat, A. & Olsen, N., 2010. Observation of magnetic diffusion in the Earth's outer core from Magsat, Ørsted, and CHAMP data, *J. geophys. Res.*, **115**(B5), 1–13.
- Chulliat, A., Alken, P. & Maus, S., 2015. Fast equatorial waves propagating at the top of the Earth's core, *Geophys. Res. Lett.*, **42**(9), 3321–3329.
- Constable, C., 1988. Parameter estimation in non-Gaussian noise, *Geophys. J. Int.*, **94**(1), 131–142.
- Constable, C.G., Parker, R.L. & Stark, P.B., 1993. Geomagnetic field models incorporating frozen-flux constraints, *Geophys. J. Int.*, **113**(2), 419–433.
- Evensen, G., 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dyn.*, **53**(4), 343–367.
- Eymic, C. & Hulot, G., 2005. On core surface flows inferred from satellite magnetic data, *Phys. Earth planet. Inter.*, **152**(3), 200–220.
- Farquharson, C.G. & Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.*, **134**(1), 213–227.
- Finlay, C.C., Aubert, J. & Gillet, N., 2016a. Gyre-driven decay of the Earth's magnetic dipole, *Nat. Communi.*, **7**, 1–8.
- Finlay, C.C., Olsen, N., Kotsiaros, S., Gillet, N. & Toffner-Clausen, L., 2016b. Recent geomagnetic secular variation from Swarm, *Earth Planets Space*, **68**(1), 1–18.
- Gillet, N., Pais, M. & Jault, D., 2009. Ensemble inversion of time-dependent core flow models, *Geochem. Geophys. Geosyst.*, **10**(6), 1–20.
- Gillet, N., Barrois, O. & Finlay, C.C., 2015a. 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.
- Gillet, N., Jault, D. & Finlay, C., 2015b. Planetary gyre, time-dependent eddies, torsional waves, and equatorial jets at the Earth's core surface, *J. geophys. Res.*, **120**(6), 3991–4013.
- Holme, R., 2015. Large scale flow in the core, in *Treatise in Geophysics, Core Dynamics*, Vol. **8**, chap. 4, pp. 91–113, eds Olson, P. & Schubert, G., Elsevier.
- Jackson, A., Constable, C., Walker, M. & Parker, R., 2007. Models of Earth's main magnetic field incorporating flux and radial vorticity constraints, *Geophys. J. Int.*, **171**(1), 133–144.
- Jonkers, A.R., Jackson, A. & Murray, A., 2003. Four centuries of geomagnetic data from historical records, *Rev. Geophys.*, **41**(2), 1–37.
- Kotsiaros, S. & Olsen, N., 2012. The geomagnetic field gradient tensor, properties and parametrization in terms of spherical harmonics, *Int. J. Geomath.*, **3**, 297–314.
- Labbé, F., Jault, D. & Gillet, N., 2015. On magnetostrophic inertia-less waves in quasi-geostrophic models of planetary cores, *Geophys. Astrophys. Fluid Dyn.*, **109**(6), 587–610.
- Leopardi, P., 2006. A partition of the unit sphere into regions of equal area and small diameter, *Electron. Trans. Numer. Anal.*, **25**(12), 309–327.
- Lesur, V., Wardinski, I., Asari, S., Minchev, B. & Mandea, M., 2010. Modelling the Earth's core magnetic field under flow constraints, *Earth Planets Space*, **62**(6), 503–516.
- Lesur, V., Wardinski, I., Baerenzung, J. & Holschneider, M., 2017. On the frequency spectra of the core magnetic field Gauss coefficients, *Phys. Earth planet. Inter.*, **276**, 145–158.
- Livermore, P.W., Hollerbach, R. & Finlay, C.C., 2017. An accelerating high-latitude jet in Earth's core, *Nat. Geosci.*, **10**(1), 62–68.
- Macmillan, S. & Olsen, N., 2013. Observatory data and the Swarm mission, *Earth Planets Space*, **65**(11), 1189–1200.
- Mandea, M. & Olsen, N., 2006. A new approach to directly determine the secular variation from magnetic satellite observations, *Geophys. Res. Lett.*, **33**(15), 1–5.
- O'Brien, M.S., Constable, C.G. & Parker, R.L., 1997. Frozen-flux modelling for epochs 1915 and 1980, *Geophys. J. Int.*, **128**(2), 434–450.
- 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. & Mandea, M., 2007. Investigation of a secular variation impulse using satellite data: The 2003 geomagnetic jerk, *Earth planet. Sci. Lett.*, **255**(1), 94–105.
- Olsen, N., Glassmeier, K.-H. & Jia, X., 2010. Separation of the magnetic field into external and internal parts, *Space Sci. Rev.*, **152**(1–4), 135–157.
- Olsen, N., Lühr, H., Finlay, C.C., Sabaka, T.J., Michaelis, I., Rauberg, J. & Toffner-Clausen, L., 2014. The CHAOS-4 geomagnetic field model, *Geophys. J. Int.*, **197**(2), 815–827.
- Olsen, N. *et al.*, 2015. The swarm initial field model for the 2014 geomagnetic field, *Geophys. Res. Lett.*, **42**(4), 1092–1098.
- Olson, P. & Aurnou, J., 1999. A polar vortex in the Earth's core, *Nature*, **402**(6758), 170–173.
- Pais, M. & Jault, D., 2008. Quasi-geostrophic flows responsible for the secular variation of the Earth's magnetic field, *Geophys. J. Int.*, **173**(2), 421–443.
- Sabaka, T.J., Olsen, N. & Purucker, M.E., 2004. Extending comprehensive models of the Earth's magnetic field with Ørsted and CHAMP data, *Geophys. J. Int.*, **159**(2), 521–547.
- Sabaka, T.J., Olsen, N., Tyler, R.H. & 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.
- Schaeffer, N., Jault, D., Nataf, H.-C. & Fournier, A., 2017. Turbulent geodynamo simulations: a leap towards Earth's core, *Geophys. J. Int.*, **211**(1), 1–29.
- Thébault, E. *et al.*, 2015. International geomagnetic reference field: the 12th generation, *Earth Planets Space*, **67**(1), 1–19.

- Verboven, S. & Hubert, M., 2005. LIBRA: a MATLAB library for robust analysis, *Chemometr. Intell. Lab. Syst.*, **75**(2), 127–136.
- Wardinski, I. & Lesur, V., 2012. An extended version of the C3FM geomagnetic field model: application of a continuous frozen-flux constraint, *Geophys. J. Int.*, **189**(3), 1409–1429.
- Whaler, K. & Beggan, C., 2015. Derivation and use of core surface flows for forecasting secular variation, *J. geophys. Res.*, **120**(3), 1400–1414.
- Whaler, K., Olsen, N. & Finlay, C., 2016. Decadal variability in core surface flows deduced from geomagnetic observatory monthly means, *J. geophys. Int.*, **207**(1), 228–243.