

# Geomagnetic core field models and secular variation forecasts for the 13th International Geomagnetic Reference Field (IGRF-13)

Ingo Wardinski (✉ [ingo.wardinski@unistra.fr](mailto:ingo.wardinski@unistra.fr))

Université de Strasbourg Ecole et Observatoire des Sciences de la Terre <https://orcid.org/0000-0001-8125-9476>

**Diana Saturnino**

Universite de Nantes

**Hagay Amit**

Universite de Nantes

**Aude Chambodut**

Universite de Strasbourg Ecole et Observatoire des Sciences de la Terre

**Benoit Langlais**

Universite de Nantes

**Mioara Manda**

CNES

**Erwan Thebault**

Universite de Nantes

---

## Full paper

**Keywords:** The geomagnetic field, Geomagnetic secular variation, Geomagnetic field models, Forecasts of the 14 geomagnetic field

**Posted Date:** July 10th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-26497/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published on October 22nd, 2020. See the published version at <https://doi.org/10.1186/s40623-020-01254-7>.

1 **Geomagnetic core field models and secular variation forecasts for**  
2 **the 13th International Geomagnetic Reference Field (IGRF-13)**

3 I. Wardinski, Institut de Physique du Globe de Strasbourg, Université de Strasbourg/EOST,  
CNRS, UMR 7516, Strasbourg, France, wardinski@unistra.fr

D. Saturnino, Laboratoire de Planétologie et Géodynamique, Université de Nantes, Univer-  
sité d'Angers , CNRS, UMR 6112, Nantes, France

H. Amit, Laboratoire de Planétologie et Géodynamique, Université de Nantes, Université  
d'Angers , CNRS, UMR 6112, Nantes, France

A. Chambodut, Institut de Physique du Globe de Strasbourg, Université de Stras-  
bourg/EOST, CNRS, UMR 7516, Strasbourg, France

B. Langlais, Laboratoire de Planétologie et Géodynamique, Université de Nantes, Université  
d'Angers , CNRS, UMR 6112, Nantes, France

M. Manda, Centre National d' Etudes Spatiales (CNES), Paris, France

E. Thébaud, Laboratoire Magmas et Volcans, Université de Clermont Auvergne, CNRS  
UMR 6524, Clermont Ferrand, France

5 **Abstract**

6 Observations of the geomagnetic field taken at Earth’s surface and at satellite altitude are combined to  
7 construct continuous models of the geomagnetic field and its secular variation from 1957 to 2020. From  
8 these parent models, we derive candidate main field models for the epochs 2015 and 2020 to the 13th  
9 generation of the International Geomagnetic Reference Field (IGRF). The secular variation candidate  
10 model for the period 2020 - 2025 is derived from a forecast of the secular variation in 2022.5, which  
11 results from a multi-variate singular spectrum analysis of the secular variation from 1957 to 2020.

12 **Keywords**

13 The geomagnetic field, Geomagnetic secular variation, Geomagnetic field models, Forecasts of the  
14 geomagnetic field

15 **Introduction**

16 The International Association of Geomagnetism and Aeronomy (IAGA) regularly releases the Interna-  
17 tional Geomagnetic Reference Field (IGRF), which is a mathematical description of Earth’s main mag-  
18 netic field and its secular variation. Previous versions (e.g. Finlay et al. 2010; Thébaud et al. 2015) are  
19 widely used in many disciplines of Earth sciences and **applied for navigational purposes (Jiménez et al.**  
20 **2012; Canciani & Raquet 2016) and in satellite orientation (Slavinskis et al. 2014).**~~applied sciences.~~

21 In this study, we combine geomagnetic field observations taken at Earth’s surface and at satellite altitude  
22 to construct continuous models of the geomagnetic core field and its secular variation between 1957 and  
23 2020. From these models, candidate models for the IGRF (Alken et al. 2020), i.e. main field models for  
24 the epochs 2015 and 2020 and a secular variation model for the period 2020 to 2025 centered at 2022.5  
25 are derived. We apply two modeling techniques to derive these models. First, a method that descends  
26 from the time-dependent modeling technique developed by Bloxham & Jackson (1992); we refer to this  
27 as the classical model. The second technique is based on a method for constructing core field models  
28 that satisfy the frozen-flux radial magnetic induction equation on the core-mantle boundary (CMB) by  
29 imposing the field evolution to be entirely due to advection of the magnetic field at the core surface  
30 (Lesur et al. 2010; Wardinski & Lesur 2012), which we refer to as the kinematic field model. The latter

31 method could be understood as a simple data assimilation approach, where the diffusion-less induction  
32 equation and assumptions about the dynamical regime of the core flow form the priors, and observations  
33 define their likelihood.

34 Methods of forecasting the future geomagnetic field evolution range from simple linear extrapolation to  
35 data assimilation into numerical dynamo simulations (Kuang et al. 2010; Aubert 2015; Fournier et al.  
36 2015). Here, we devise two strategies to forecast the geomagnetic secular variation. First, a direct forecast  
37 based on a multi-variate singular spectrum analysis (MSSA) (Broomhead & King 1986; Plaut & Vautard  
38 1994) of the magnetic field variability of past decades. Second, a kinematic forecast scheme is applied  
39 that is also based on the MSSA, but of the core flow variability of past decades. The reconstruction  
40 of the past flow variability and its forecast are used to predict the future geomagnetic field by forward-  
41 modeling the diffusion-less radial induction equation on the CMB. This approach is somewhat similar  
42 to geomagnetic field forecasts using steady and time invariant flows (Beggan & Whaler 2010; Hamilton  
43 et al. 2015; Whaler & Beggan 2015). However, such forecasts are expected to fail at the occurrence of  
44 geomagnetic jerks that are sudden changes in the secular variation. Such events occurred in the past  
45 decades (Mandea et al. 2010; Brown et al. 2013), most recently in 2014 (Torta et al. 2015). The cause  
46 of these events is not fully understood. Their occurrences have been related to different types of rapid  
47 wave motion within Earth’s liquid core (Bloxham et al. 2002; Aubert & Finlay 2019), temporal changes  
48 of the core flow (Wardinski et al. 2008) and Earth’s rotation variation (Holme & de Viron 2005; Holme  
49 & de Viron 2013).

50 This paper is organized as follows. Section 2 outlines the two techniques to derive the parent geomagnetic  
51 field model for the IGRF candidates. In the third section, we develop the methodology to predict future  
52 geomagnetic secular variation. Section 4 provides results of the geomagnetic field modeling, secular  
53 variation forecasts and the derivation of the candidate models. The last section discusses the results and  
54 concludes the study.

## 55 **Geomagnetic field modeling**

56 In this section, we summarize the derivation of a parent geomagnetic main field model from which we  
57 deduce an IGRF candidate model. The parent model, hereafter C<sup>3</sup>FM3, covers the period from 1957 to  
58 2020. The model derivation follows that of Wardinski & Lesur (2012) and consists of two branches, a  
59 classical model without the kinematic constraint applied (see section ), and a kinematic field model based

60 on the tangential geostrophic flow assumption (see section ). Like in the previous model, C<sup>3</sup>FM2, we use  
61 order 6 B-splines to parameterize field and flow coefficients in time. The spline knot spacing is set to be  
62 roughly 1.5 years. Both model branches are constrained to fit a main field model for the epoch 2015.  
63 This main field model is based on magnetic measurements taken by the Swarm satellite mission (Lesur,  
64 priv. comm.). We choose 2015, as it is the epoch of the last IGRF, with a good data coverage provided  
65 by geomagnetic observatories and the Swarm satellite mission. This data coverage decreases towards the  
66 model endpoint.

## 67 Data

68 In this work, we use two types of data, measurements taken at a network of ground-based geomagnetic  
69 observatories and satellite data taken at satellite virtual geomagnetic observatories (Mandea & Olsen  
70 2006). The idea of combining ground-based and virtual observatory data to perform a geomagnetic field  
71 modeling was already carried out by Barrois et al. (2018). However, here we derive secular variation  
72 estimates to avoid leakage of sub-annual external field variations into the description of the core field  
73 and to obtain a sufficient representation of the short-term secular variation, that may be used to forecast  
74 geomagnetic field changes.

75 **Ground-based geomagnetic observatories** A large portion of the data used in this study comes  
76 from ground-based observatories. Like in previous studies (Wardinski & Holme 2006; Wardinski & Lesur  
77 2012; Lesur et al. 2018), we derive estimates of secular variation by annual differences from observatory  
78 monthly means, where these monthly means are averages of observatory hourly means. Also, annual  
79 means are used for observatories for which hourly mean values are not available from the World Data  
80 Centre for Geomagnetism - Edinburgh (2019). These observatory annual means are part of a compilation  
81 that is provided by the British Geological Survey - Edinburgh (2020). Over the period 1957-2018 the  
82 number of geomagnetic observatories simultaneously in operation that have been providing vectorial  
83 hourly means of North, East and downward components ranges between 72 and 155. Data errors were  
84 removed when encountered and data gaps were not filled by interpolations. Figure 1 maps locations of  
85 the ground-based geomagnetic observatories used in this study.

86 **Satellite virtual geomagnetic observatories** We use vector magnetic field measurements from  
87 Swarm Level-1b data product, version 0505 (0506 for some data files). All three Swarm satellites  
88 are considered for the period between January 2014 to June 2019. Data is screened for quality flags  
89 defined in the Level-1b Product Definition Document (Tøffner-Clausen & Nielsen 2018). We keep only

100 measurements identified as nominal, and also Swarm C vector measurements after 4<sup>th</sup> November 2014.  
 101 We select only data where the Sun is below the horizon (Chambodut et al. 2002). Additionally, we retain  
 102 only data showing moderate magnetic activity. Sectorial magnetic activity index  $a\sigma$  (Chambodut et al.  
 103 2013), provided by the International Service of Geomagnetic indices (2020), were used and we only select  
 104 data corresponding to  $a\sigma < 25$  nT.

105 With the selected data we then construct a global mesh of virtual geomagnetic observatories (VO)  
 106 following Saturnino et al. (2018), with some small changes. An approximately equal area mesh is  
 107 obtained with the VO centers separated in latitude by  $12.8^\circ$  and defining 14 latitudinal bands, with  
 108  $\theta_{vo} = \pm 6.40^\circ, \pm 19.20^\circ, \pm 32.0^\circ, \dots, \pm 83.20^\circ$ . In each band, the longitude  $\phi_{vo}$  of each VO and the number  
 109 of longitudinal divisions,  $N_{\phi_{vo}}$  (rounded up to the nearest integer), are chosen so that:

$$N_{\theta_{vo}} = \frac{360}{12.8} \cos \theta_{vo}. \quad (1)$$

100 The resulting mesh contains 258 VOs. Figure 1 displays the locations of all VOs in the mesh. The data  
 101 set of each VO consists of selected data acquired inside a cylinder of  $3.0^\circ$  radius centered around each  
 102 VO and during a 30-day period, i.e., leading to nearly monthly values. The Equivalent Source Dipole  
 103 (ESD) technique is then used, following closely Saturnino et al. (2018). For each month (30-day period)  
 104 the ESD inversion is applied to each VO vector data for the equivalent magnetization of dipoles placed at  
 105 2900 km depth inside Earth’s interior, by a least squares fit in an iterative, conjugate gradient, inversion  
 106 scheme (Purucker et al. 1996). Then, the forward calculation is used to estimate a magnetic field value  
 107 at the VO center location and for a given time period. In this way, time series of magnetic field values  
 108 at the center of each VO and at a constant altitude of 500 km, are obtained.

109 ~~The distributions of ground-based and virtual observatories differ as can be seen in Figure 1. Overall,~~  
 110 ~~geomagnetic observatory locations cluster, which may lead to a higher spatial resolution of the model~~  
 111 ~~in some parts of the world, whereas in other parts the resolution may be lower than that of the virtual~~  
 112 ~~observatories. In addition to the VO’s, Figure 1 also maps locations of the ground-based geomagnetic~~  
 113 ~~observatories used in this study. The distribution of the latter observatory type is not similar, which may~~  
 114 ~~cause different spatial resolutions in the geomagnetic field model.~~

115 **Secular variation estimates** We derive secular variation estimates as input for the geomagnetic  
 116 field modeling. The technique is applied to monthly means of VOs and to ground-based geomagnetic

117 monthly means. The secular variation of the  $X$ -component at a given observatory is estimated as

$$dX/dt|_{\tau} = (X(\tau + 6) - X(\tau - 6))/dt \quad (2)$$

118 where  $\tau$  denotes a particular month. These are annual differences of observatory monthly means.

119 Likewise, observatory annual means are treated using

$$dX/dt|_{t+1/2} = (X(t + 1) - X(t))/dt \quad (3)$$

120 where  $t$  is in calendar years and  $dt$  is 1 year. Then, secular variation estimates derived by (2) and (3) are

121 given in nT/yr.

## 122 Classical modeling

123 Conventional geomagnetic field modeling approaches rely on the assumption that Earth's magnetic field

124  $\mathbf{B}(r, \theta, \phi, t)$  is a potential field without magnetic sources in the mantle and in the vicinity of satellite

125 virtual observatories. **Because of this, the geomagnetic field is determined as a gradient of a scalar**

126 **potential, i.e.**

$$\mathbf{B}(r, \theta, \phi, t) = -\nabla V(r, \theta, \phi, t). \quad (4)$$

127 Then in a spherical geometry, the scalar potential of the geomagnetic field can be represented as

$$V(r, \theta, \phi, t) = a \sum_{l=1}^{l_{\max}} \sum_{m=0}^l (g_l^m(t) \cos(m\phi) + h_l^m(t) \sin(m\phi)) \left(\frac{a}{r}\right)^{l+1} P_l^m(\cos \theta), \quad (5)$$

128 where  $a$  is Earth's radius (6371.2 km),  $(r, \theta, \phi)$  the geocentric spherical radial, co-latitude and longitude

129 coordinates and  $P_l^m(\cos \theta)$  are the Schmidt quasi-normalized associated Legendre functions, **with their**

130 **normalization defined by**

$$\int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} (P_l^m(\cos \theta) \cos(m\phi))^2 \sin \theta d\theta d\phi = \frac{4\pi}{2l+1}, \quad (6)$$

131 with degree  $l$  and order  $m$ . The maximum spherical harmonic degree in (5) is chosen to be  $l_{\max} = 14$ , to

132 minimize the contamination by the crustal field. The Gauss coefficients  $\{g_l^m, h_l^m\}$  are expanded in time

133 using order six B-splines  $M_n(t)$

$$g_l^m(t) = \sum_{n=1}^N g_l^{mn} M_n(t), \quad h_l^m(t) = \sum_{n=1}^N h_l^{mn} M_n(t). \quad (7)$$

134 The objective function  $\Phi(\mathbf{m})$  to be minimized in the inversion is:

$$\Phi(\mathbf{m}) = (\mathbf{y} - \mathbf{A}\mathbf{m})^T \mathbf{C}_e^{-1} (\mathbf{y} - \mathbf{A}\mathbf{m}) + \mathbf{C}_m^{-1} + \mathbf{C}_F^{-1}, \quad (8)$$

135 where  $\mathbf{A}$  is an operator which relates the model vector  $\mathbf{m}$  containing the Gauss coefficients to the data  
 136  $\mathbf{y}$ .  $\mathbf{y} - \mathbf{A}\mathbf{m}$  is the misfit between data and model, subject to the regularization.  $\mathbf{C}_e$  and  $\mathbf{C}_m$  are the error  
 137 and the prior model covariance matrix, respectively.  $\mathbf{C}_m$  is an expression of the model priors.

138 Here, we report solutions that are adjusted to minimize the integral of  $B_r^2$  over the core surface to obtain  
 139 a spatial smooth model

$$\int_{S(c)} B_r^2 d\Omega = \mathbf{m}^\top \mathbf{S}^{-1} \mathbf{m}. \quad (9)$$

140 The matrix  $\mathbf{S}$  has the diagonal elements

$$s_{ll}, t_{ll} = \frac{(l+1)^2}{2l+1} \left(\frac{a}{c}\right)^{(2l+4)} \quad \text{for } l = 1 \dots, l_{\max}. \quad (10)$$

141 Like in some previous studies (Lesur et al. 2010; Wardinski & Lesur 2012) we seek a reliable estimate  
 142 of the secular acceleration. Therefore, the temporal model constraint is to minimize the integral of the  
 143 third time derivative of the radial field component over the core surface and in the model period between  
 144  $t_S$  and  $t_E$

$$\frac{4\pi}{(t_E - t_S)} \int_{t_S}^{t_E} \int_{S(c)} \left( \frac{\partial^3 B_r}{\partial t^3} \right)^2 d\Omega dt = \mathbf{m}^\top \mathbf{T}^{-1} \mathbf{m}, \quad (11)$$

145 where  $S(c)$  is the spherical surface of the core at radius  $c = 3485$  km. The diagonal elements of the matrix  
 146  $\mathbf{T}$  are the same as for  $\mathbf{S}$ , and the time integral is computed using a Newton-Cotes formula of a closed  
 147 type, e. g. Bode's rule (Abramowitz & Stegun 1973). Minimization of the third time derivative requires  
 148 placing further conditions on the second time derivatives of the radial field at the model end-points; best  
 149 results are obtained when these are set to zero.

150 The model prior covariance matrix  $\mathbf{C}_m$  is then given by:

$$\mathbf{C}_m^{-1} = \lambda_S \mathbf{m}^\top \mathbf{S}^{-1} \mathbf{m} + \lambda_T \mathbf{m}^\top \mathbf{T}^{-1} \mathbf{m}. \quad (12)$$

151 The last term of (8) is the constraint to fit a given satellite field in 2015, which is

$$\mathbf{F} := \int_{r=a} (B - {}^0B)^2 dS|_{t=2015} = \sum_{l=1}^{l_{\max}} \sum_{m=0}^l (l+1) [(g_l^m - {}^0g_l^m)^2 + (h_l^m - {}^0h_l^m)^2]_{t=2015} \quad (13)$$

152 where  ${}^0g_l^m$  and  ${}^0h_l^m$  are the Gauss coefficients of the satellite geomagnetic main field model. This  
 153 constraint is necessary, as our model is based on secular variation data and it needs a main field model  
 154 at a given epoch in order to provide also description of the main field at all times. Traditionally, we use  
 155 a main field model that is derived in an independent study from satellite data. The constraint is then  
 156 written

$$\mathbf{C}_F^{-1} = \lambda_f \mathbf{m}^\top \mathbf{F}^{-1} \mathbf{m}. \quad (14)$$

157 with the damping parameter  $\lambda_f$ .

158 Solutions are sought iteratively in a very similar manner as for the previous model, C<sup>3</sup>FM2, by deriving  
159 an initial model to re-weight the observatory data by their residuals to this initial model. Then, the  
160 strength of external field variation is reduced by a noise-removal scheme (Wardinski & Holme 2011).  
161 From this data set the final model is derived.

## 162 **Kinematic field modeling**

163 In this section we describe our method to invert geomagnetic observations for field and flow at the core  
164 surface, which extends our previous study (Wardinski & Lesur 2012), where we imposed the core flow to  
165 be purely toroidal. The inversion of secular variation data for field and flow at the core surface formulates  
166 to a Bayesian inference. Assuming a Gaussian distribution, then this leads to an objective function similar  
167 to (8) that is minimal for the preferred solution  $\mathbf{m}$ .

168 In the following we provide details of the prior information (constraints) used to derive a preferred solution  
169 of the non-unique and non-linear inverse problem. The constraints are applied on the portion of the model  
170 vector that represents the core surface flow.

171 Here, we use the data set with the reduced external field noise obtained in section to jointly invert for the  
172 field and flow at the core surface. This eases the joint-inversion process, as it avoids the iterative solving  
173 scheme to re-weight the data for the non-linear problem. Different assumptions of the flow dynamic  
174 could be applied (Holme 2007). Among them: purely toroidal (Whaler 1980), tangential geostrophic  
175 (LeMouél 1984) and quasi-geostrophic flow (Pais & Jault 2008). However, we focus on the tangential  
176 geostrophic flow assumption, as it is more comprehensive than a purely toroidal flow, but less restrictive  
177 than a quasi-geostrophic flow. (Note that the term flow refers to its horizontal part only, as the radial  
178 part vanishes at the core surface.) The flow is decomposed into toroidal and poloidal components

$$\mathbf{u} = \mathbf{u}_{\text{tor}} + \mathbf{u}_{\text{pol}} = \nabla_h \times (\hat{\mathbf{r}}\mathcal{T}) + \nabla_h \mathcal{P}. \quad (15)$$

179  $\mathcal{T}$  and  $\mathcal{P}$  are scalars which are expanded in Schmidt-normalized real spherical harmonics in space and  
180 B-splines of order 6 in time, represented by  $t_l^m(n), s_l^m(n)$ .

181 Following Lesur et al. (2010), the objective function of the joint inversion for the field and the flow at the  
182 core surface reads

$$\Phi(\mathbf{m}) = (\mathbf{y} - \mathbf{A}\mathbf{m})^T \mathbf{C}_e^{-1} (\mathbf{y} - \mathbf{A}\mathbf{m}) + \lambda_1 \Phi_1(\mathbf{m}), \quad (16)$$

183 where the model vector  $\mathbf{m}$  now contains the sets of Gauss and flow coefficients. The functional  $\Phi_1$  is

184 related to the kinematic constraint and defined by

$$\Phi_1(\mathbf{m}) = \sum_t (\mathbf{A}_g(t) \cdot \mathbf{u} - \dot{\Psi}(t) \cdot g)^T \mathbf{C}^g (\mathbf{A}_g \cdot \mathbf{u} - \dot{\Psi}(t) \cdot g). \quad (17)$$

185  $\dot{\Psi}(t)$  is a design matrix based on the radial induction equation in the kinematic assumption, i.e.

$$\partial_t B_r = -\nabla_h \cdot (\mathbf{u} B_r). \quad (18)$$

186 According to (18) the secular variation of the radial field component at the core surface,  $\partial_t B_r$ , can be  
 187 expressed in terms of a core field,  $B_r$ , advected by the core fluid flow,  $\mathbf{u}$ , where  $\nabla_h$  is the horizontal  
 188 divergence. Then

$$\dot{\Psi}(t) \cdot g = \mathbf{A}_g \cdot \mathbf{u}, \quad (19)$$

189 where  $\mathbf{g}$  is a vector that contains the Gauss coefficients and  $\mathbf{u}$  contains the flow coefficients. The elements  
 190 of the diagonal weight matrix  $\mathbf{C}^g$  are defined as  $w_l^g = \frac{4\pi(l+1)^2}{(2l+1)}$ . Minimizing the mean square difference,  
 191 integrated over the core surface  $S(c)$  and time, between the observed secular variation and the secular  
 192 variation generated by the flow, then the functional  $\Phi_1$  is equivalent to the integral and we can write this  
 193 (similarly to (9) and (11)) as

$$\int_{\mathcal{T}} \int_{S(c)} (\partial_t B_r(t) + \nabla_h \cdot (\mathbf{u} B_r))^2 d\Omega dt = \mathbf{g}^T \mathbf{K}^{-1} \mathbf{g}, \quad (20)$$

194 where  $\mathbf{m}$  is now the model vector that holds the flow and Gauss coefficients. The diagonal elements of the  
 195 field part of  $\mathbf{K}$  are given by  $\mathbf{C}^g$ . The parameter  $\lambda_1$  controls the conformity of the model to the kinematic  
 196 constraint.

197 Because  $\mathbf{A}_g(t)$  in the functional  $\Phi_1$  depends on the Gauss coefficients  $g_l^m(t), h_l^m(t)$  and is multiplied by  $\mathbf{u}$ ,  
 198 this optimization problem (inversion) is clearly non-linear and has to be solved iteratively. However, the  
 199 iterative process is unlikely to converge unless some constraints are applied on the flow model, as finding  
 200 a flow model is an ill-posed inverse problem (Holme 2007). In order to obtain the optimal field model  
 201 and simultaneously reduce the null space for the flow, two types of constraints are considered. The flow  
 202 model is forced to have a convergent spectrum, i.e. to be large scale, and to minimize Bloxham's "strong  
 203 norm" (Bloxham 1988; Jackson 1997),

$$\int_{\mathcal{T}} \int_{S(c)} (\nabla_h(\nabla_h \cdot \mathbf{u}))^2 + (\nabla_h(\hat{r} \times \nabla_h \cdot \mathbf{u}))^2 d\Omega dt = \mathbf{u}^T \mathcal{N} \mathbf{u}. \quad (21)$$

204  $\mathcal{N}$  has the diagonal elements  $(l^3(l+1)^3)/(2l+1)$ . The damping parameter  $\lambda_S$  controls to what extent  
 205 the flow follows this constraint. Minimizing (21) constrains efficiently the secular variation. Secondly the

206 flow model is chosen such that it varies smoothly in time,

$$\int_{\mathcal{T}} \int_{S(c)} (\mathbf{u})^2 d\Omega dt = \mathbf{u}^\top \mathcal{V} \mathbf{u}, \quad (22)$$

207 with  $\lambda_T$  as the associated control parameter of the flow temporal evolution that efficiently regularize the  
 208 inverse problem. **The constraints (21) and (22) are similar to the temporal constraint of the classical**  
 209 **model, i.e. (11), as they involve a temporal integration to be minimized.** Finally, it is required that the  
 210 flow acceleration at starting and ending epochs is minimized by

$$\int_{S(c)} \left( \frac{\partial \mathbf{u}}{\partial t} \right)^2 d\Omega \Big|_{t_1, t_2} = \mathbf{u}^\top \mathcal{E} \mathbf{u}. \quad (23)$$

211  $t_1$  and  $t_2$  are the epochs 1957 and 2020.0, respectively. This becomes necessary, as, if the flow acceleration  
 212 is un-constrained at the endpoint, it may exceed realistic values. **The factor  $\lambda_E$  controls this constrain.**

213 These are the basic settings for the joint inversion for the magnetic core field and the core flow. We  
 214 impose a further constraint, in order to derive models that are based on different dynamical assumptions  
 215 of the flow. One possible assumption commonly used is a tangential geostrophic flow (Hills 1979; LeMouél  
 216 1984) which is established by minimizing

$$\int_{S(c)} (\nabla_h \cdot (\mathbf{u} \cos \theta))^2 d\Omega = \mathbf{u}^\top \mathcal{G} \mathbf{u}, \quad (24)$$

217 where  $\theta$  is the co-latitude. **Elements of  $\mathcal{G}$  are given by Pais et al. (2004). The constraint is controlled by**  
 218 **setting  $\lambda_{TG}$ .**

219 We apply two measures to discriminate models: the rms secular variation misfit is measured by differences  
 220 between model and the observed secular variation, i.e.

$$\mathbf{M} = \sqrt{\frac{1}{(N_{obs} - 1)} \sum_{i=1}^{N_{obs}} (\text{Obs} - \text{Mod})^2} \quad (25)$$

221 where  $N_{obs}$  is the number of observations. In addition, for the kinematic field models we derive the ratio

$$\mathbf{R}(t) = \frac{\int_{S(c)} \{ \partial_t B_r(t) + \nabla_h \cdot (\mathbf{u} B_r) \}^2 d\Omega}{\int_{S(c)} \{ \partial_t B_r(t) \}^2 d\Omega} \quad (26)$$

222 that specifies to what extent the frozen-flux radial induction equation is satisfied. Values similar to 1 and  
 223 larger mean that this constraint is not fulfilled by the model, and conversely values significantly smaller  
 224 than 1 indicate that the frozen-flux radial induction equation is approximately satisfied.

## 225 Forecasting schemes

226 In this study, we aim to obtain a description of a) the temporal dynamics of the secular variation and  
 227 b) the temporal variability of the advective motion within the liquid outer core. These should not differ

228 largely, but remaining external signals in the secular variation estimates may pose problems to identify  
 229 clearly signals from the core. To forecast states of the physical system that lead to the secular variation  
 230 it is necessary to analyze the observed time series. Our strategy relies on the derivation of multi-variate  
 231 time series models, where individual secular variation and flow coefficients are treated as time series.  
 232 Our development of time series models and forecasts is based on the multi-variate singular spectrum  
 233 analysis (MSSA) (Broomhead & King 1986; Plaut & Vautard 1994, and see appendix 1) of field and  
 234 core flow variability. Other methods could be chosen to find a sufficient description of the field and flow  
 235 variability, such as ARIMA-models, or vector auto-regressions models; however, these model types are  
 236 based on assumptions like stationarity of the time series and normality of the residuals (Box & Jenkins  
 237 1976; Brockwell & Davis 2002, and references therein), which are unlikely for the temporal variability of  
 238 the geomagnetic field (Hulot et al. 2010). Furthermore, the singular spectrum analysis does not rely on  
 239 assumptions about the linearity or non-linearity of the process that is generating the time series, whereas  
 240 auto-regression models implicitly assume a linear behavior of the observed data. If  $1, \dots, T$  serves as the  
 241 time range for a training set, then the model parameters are estimated from these observations. The  
 242 (out-of-sample) prediction is generated over the time range  $T + 1, \dots, T + h$  according to the generation  
 243 mechanism of the model. In this study we consider 10 years as the time range of the prediction. This  
 244 sets  $h = 120$  months. Prediction models are derived **using the first 22 eigenmodes obtained by the MSSA**  
 245 **of the temporal flow variability (see appendix 1).** ~~to truncation level  $k = 30$ .~~

#### 246 **Direct secular variation forecast**

247 The time series of the Gauss coefficients are taken from the classical model branch of C<sup>3</sup>FM3. The  
 248 analysis considers discrete series of secular variation coefficients

$$y_t := \dot{g}_t^m(n), n = 1, \dots, N \quad (27)$$

249 consisting in  $N$  observations regularly sampled in time. The sampling time  $\tau_s$  is chosen to be one month.  
 250 Time series models are derived using a multi-variate singular spectrum analysis, where the temporal  
 251 variability of the series is decomposed in Eigenmodes (Vautard et al. 1992; Golyandina et al. 2001; Ghil  
 252 et al. 2002). The time series models are then reconstructions based on superposition of these Eigenmodes,  
 253 and similarly forecasts. For the technical details of the model selection we refer the reader to appendix  
 254 1.

## Kinematic secular variation forecast

The kinematic forecasting scheme to predict future secular variation is motivated by results that the observed secular variation can be mostly explained by the advection of magnetic field (Roberts & Scott 1965; Wardinski & Lesur 2012). Additionally, results of a previous kinematic field model show that the spatial complexity imposed by a satellite magnetic field model of 2004 is maintained backward in time over decades (Wardinski & Lesur 2012). Here, we want to explore to what extent the approach can be used to forecast geomagnetic field changes.

The scheme consists of two steps. First, we derive time series models that give a sufficient accurate description of the temporal variability of the fluid motion in the liquid outer core. Therefore, we consider toroidal and poloidal flow terms, which are treated as discrete multi-variate time series from which we derive time series models for a given 'learning' phase. The learning phase is set to cover the interval 1957 to 2020.

The second step, the kinematic forecasting step, employs forward computation of the secular variation by using the diffusion-less induction equation and is initialized by

$$\partial_t B_r(t_1) = -\nabla_h \cdot (\tilde{\mathbf{u}}(t_0) B_r(t_0)) \quad (28)$$

where  $\tilde{\mathbf{u}}(t_0)$  represents the flow  $(t_l^m(n), p_l^m(n))$ , and  $B_r(t_0)$  the radial magnetic field at an initial epoch  $t_0$ . Here, we use the flow coefficients of the kinematic branch of C<sup>3</sup>FM3. The forecasts of  $\tilde{\mathbf{u}}(t_i)$  for  $i \neq 0$  are obtained by deriving forecasts of the flow coefficients from their time series models, similar to the operation in section . Generally, the forecasts of the secular variation are computed by

$$\begin{aligned} \partial_t B_r(t_n) &= -\nabla_h \cdot (\tilde{\mathbf{u}}(t_n) B_r(t_n)), \\ B_r(t_{n+1}) &= B_r(t_n) + \partial_t B_r(t_n) \delta t, \end{aligned} \quad (29)$$

where  $\delta t$  represents the sampling interval of one month, and  $\tilde{\mathbf{u}}(t_n)$  is the flow forecast vector that contains forecasts of the individual flow coefficients for the first month after 2020.0. Equation (28) is computed on a spherical grid, and then transformed to spherical harmonic secular variation coefficients  $(\dot{g}_l^m, \dot{h}_l^m)$  similar to the scheme of Lloyd & Gubbins (1990). The forecast period is 120 months, and ends in 2030.0.

## Assessing the accuracy of forecasts

Choosing an appropriate error measure in forecasting is problematic, because no single measure gives an unambiguous indication of forecasting performance, while the use of multiple measures makes comparisons between forecasting methods difficult and unwieldy (Mathews & Diamantopoulos 1994). We apply a

281 derivative of the widely used mean absolute percentage error, i.e. the symmetric absolute percentage  
 282 error (sAPE) to assess the difference between the observation, i.e. the actual field and flow coefficient  
 283  $y(t)$  and its forecast  $\hat{y}(t)$  as:

$$\text{sAPE}(t) = \frac{2|y(t) - \hat{y}(t)|}{|y(t)| + |\hat{y}(t)|}. \quad (30)$$

284 This follows the definition by Chen & Yang (2004), which differs from those of Armstrong (1985); Flores  
 285 (1986), as it does not become singular and has a maximum value of two when either  $y(t)$  or  $\hat{y}(t)$  is zero,  
 286 but is undefined when both are zero. The range of (30) is (0, 2), i.e. the maximum value corresponds to  
 287 200% percentage error. We define a prediction length, for which the sAPE( $t$ ) becomes larger than 10%.  
 288 This is a rather cautious definition and other limits could also be considered.

289 Another measure that is widely used (e.g. Aubert 2015; Whaler & Beggan 2015) is the rms-difference  
 290 between models

$$\sqrt{\text{dP}} = \sqrt{\sum_{l=1}^{l_{max}} \sum_{m=0}^l (l+1) [{}^A g_l^m - {}^B g_l^m]^2 + [{}^A h_l^m - {}^B h_l^m]^2}, \quad (31)$$

291 where  ${}^A g_l^m$ ,  ${}^A h_l^m$  and  ${}^B g_l^m$ ,  ${}^B h_l^m$  are Gauss coefficients of the compared field models A and B.  $\sqrt{\text{dP}}$   
 292 represents

$$\sqrt{\text{dP}} = \int B d\Omega \quad (32)$$

293 the total (difference) field integrated over the surface of the Earth, and similarly is the secular variation.  
 294 We compute it in order to allow comparisons to other studies which use this measures as primary  
 295 diagnostic. However, we must note, that  $\sqrt{\text{dP}}$  is biased towards large scale contributions, when computed  
 296 at Earth's surface where the spatial energy spectrum of the field is decreasing, and is biased towards small  
 297 scales at the core surface, where the spectrum is increasing. This is not the case for the sAPE( $t$ ), as it is  
 298 relative to the amplitude of the actual coefficient.

## 299 Results and Discussion

300 In this section we present results of the core field modeling and the secular variation forecasts. At the  
 301 end of this section we also present and discuss our candidates for the definitive geomagnetic reference  
 302 field model (DGRF) in 2015, and IGRF candidate model for 2020.

303 We focus on two models: Model 1 was derived with the classical modeling approach and Model 2 is a  
 304 kinematic field model based on the tangential geostrophic flow assumption. Table 1 compiles the set of  
 305 model parameters and related model characteristics. Both models provide almost equal fits to the data

306 in terms of their residual standard deviation.

307 We note that, by strongly imposing the tangential geostrophic constraint, the fit to the radial induction  
308 equation deteriorates, when compared to solutions with a weaker tangential geostrophic constraint, see  
309 appendix Table 3. The reason for this is not clear, but, perhaps, could be explained by the penalization  
310 of the poloidal  $p_n^0$  terms when the tangential geostrophic constraint is imposed.

### 311 **Comparison of modeled and observed secular variation**

312 In Figure 2 we show the rms differences (31) between the two models during the model period. The  
313 rms differences of the main field (Figure 2, left) and secular variation coefficients (Figure 2, right) are  
314 continuously decreasing during the period, except close to the model endpoint in 2020, where they  
315 increase. These increases can be explained by the constraint on the flow acceleration in Model 2. We  
316 argue that the decrease in differences during the model period is not directly related to the growth of  
317 the available ground-based geomagnetic observatory data, because this affects both models. Rather,  
318 the decrease in differences could be explained by the different modeling techniques. As it was found  
319 in a previous study (Wardinski & Lesur 2012), the kinematic field modeling seems to project the field  
320 morphology of a spatially high-resolved satellite field model imposed in 2010 backward in time. Therefore,  
321 the kinematic model is less subject to a varying data distribution, whereas the classical field model is.

322 Figure 3 compares the observed and modeled secular variation at different observatory sites. The model  
323 curves are derived from our preferred classical and kinematic field models. The observed secular variation  
324 series from ground-based geomagnetic observatories stop in 2017.5, as more recent data were not available  
325 at the time of this study.

326 The overall impression is that the classical and kinematic field models fit the observed secular variation  
327 equally well. An exception is, at the model endpoints where the two curves deviate. Most likely, this  
328 is due to the endpoint constraint of the kinematic field models. The known jerk occurrences (A - J)  
329 are all clearly visible in the Y-component of Chambon-la-Forêt, whereas in other observatory data only  
330 some jerks are detectable, e.g. the series of Sitka shows only events in 1969 and 2011. Data of the Sitka  
331 observatory show large changes of the secular variation of Y and Z components in the past decade, that  
332 may be related to the rapid movement of the magnetic North Pole.

333 The same large similarity between the models can be observed in Figure 4, which shows six secular  
334 variation coefficients with the largest amplitude of the two different models. These coefficients largely  
335 determine the morphology of the secular variation at the core surface. Particularly, the coefficient  $h_2^1$  is

336 closely related to the prominent patch of reverse magnetic flux in the southern hemisphere (Terra-Nova  
337 et al. 2017). Overall, differences between the model coefficients rarely exceed 2 nT/yr, except close to  
338 model endpoints, where models disperse largely.

339 Figure 4 also shows markers of known geomagnetic jerks, to allow their identification in the evolution of the  
340 secular variation coefficients. The coefficients  $\dot{g}_1^0$  and  $\dot{h}_2^1$  represent equatorial anti-symmetric contributions  
341 of the secular variation. These coefficients carry most of the known geomagnetic jerks, apart from the  
342 event in 1969 which is either not clearly visible in  $\dot{g}_1^0$ , or appears later in 1970. The identification of  
343 geomagnetic jerks in the equatorial symmetric parts of the secular variation described by  $\dot{g}_1^1$ ,  $\dot{h}_1^1$ ,  $\dot{g}_2^0$  is less  
344 clear.

345 The temporally averaged secular variation spectra of the two models are shown in Figure 5. On large  
346 and small scales, differences between spectra remain small at Earth’s surface. The spectra also indicate a  
347 very high temporal variability of the kinematic field secular variation at degrees 9 and 10, suggesting that  
348 at some epochs the power of the secular variation of these models is weaker by a few orders of magnitude  
349 than the secular variation of the classical field model.

350 At the core surface (Figure 5, right) the spectral power of the secular variation grows with spherical  
351 harmonic degree. The spectra of the two models start to deviate from degree 11, where the spectrum of  
352 **Model 2** flattens, whereas the spectrum of **Model 1** continues to increase. We note the same high temporal  
353 variability at spherical harmonic degree 9 of the kinematic secular variation model.

354 Maps of the radial component of the magnetic field and its secular variation derived from **Model 1** at the  
355 top of the core are shown in Figure 6. The radial magnetic field component shows a dipole-dominated  
356 morphology, but with considerable small scale features, such as reversed flux patches in both hemispheres.

357 The secular variation at the core surface tends to be dominated by small scales. Particular notable features  
358 are low latitude pairs of opposite polarity secular variation which are typical to advection (Amit 2014).

359 Also note larger patches of the radial secular variation in the vicinity of the magnetic North Pole, e.g.  
360 under Eastern Siberia.

361 Figure 7 shows maps based on **Model 2**. Maps of the radial component of the magnetic field at the  
362 core surface are widely in agreement with the respective maps of Figure 6. Perhaps most apparent, is  
363 the presence of the reversed flux patch underneath the East Pacific, in both models except at 1969 in  
364 Figure 6. Maps of the secular variation in 2020 differ substantially, which may be due to the endpoint  
365 constraint of **Model 2**. The secular variation map of **Model 2** in 2020 contains some large scale anomalously

366 intense structures below the geographic South Pole, indicating an intensification of the flux in this region.  
367 Based on the strong secular variation at high latitudes of the northern hemisphere, Livermore et al. (2017)  
368 inferred a zonal jet there. In the context of equatorial symmetry, they argued that such a zonal jet at high  
369 latitudes of the southern hemisphere would not produce detectable secular variation, because the field is  
370 oriented in the east-west direction there. Indeed in Figure 6 in 2020 the secular variation is very weak  
371 around the South Pole (and quite strong under the North Pole). All this makes the anomalously strong  
372 secular variation around the South Pole in 2020 as seen in Figure 7 quite suspicious. However, secular  
373 variation maps of other epochs (1969 and 2010) are very similar to maps of Figure 6, with dominant  
374 small scale features at mid- and low latitudes.

375 We conclude that classical and kinematic field models largely agree, both in spatial morphology as well as  
376 in the temporal evolution, but significantly deviate in their secular variation towards the endpoints. The  
377 cause for this deviation is the constraint of the kinematic field model onto the flow acceleration, while  
378 the classical model gets along without such constraint.

### 379 **Direct secular variation forecast**

380 Figure 8 displays observations and predictions of secular variation coefficients with the largest amplitude  
381 at the Earth surface, i.e.  $\dot{g}_1^0, \dot{g}_1^1, \dot{h}_1^1, \dot{g}_2^0, \dot{h}_1^2, \dot{h}_2^2$ , from 2000 to 2030. The most prominent feature of the  
382 forecast on Model 1 is the steep decrease of  $\dot{g}_1^0$  by almost 15nT/yr during a very short interval of about  
383 3 years. While this decrease is determined by the statistical properties of the time series model, it may  
384 be difficult to accept this when considering the past secular variation. In fact, this decrease would be  
385 caused by a large secular acceleration. Advective sources and sinks of  $\dot{g}_1^0$  exhibit a lot of cancellations  
386 and it is the rather small residual that gives the historical dipole decay (Olson & Amit 2006; Finlay et al.  
387 2016), i.e. subtle change in the core flow pattern and its interaction with the field may yield considerable  
388 changes in  $\dot{g}_1^0$ . However, apparently such a change did not happen since 1840. The other terms vary  
389 within ranges of previous oscillations (see Figure 4).

390 Most of the secular variation terms of Model 1 indicate possible occurrences of two future geomagnetic  
391 jerks in late 2020 - early 2021, and in early 2024. In Figure 8, at these dates the predicted secular  
392 variation shows notable changes in slope.

### 393 **Kinematic secular variation forecast**

394 The kinematic secular variation forecast is based on forecasts of individual flow coefficients derived from a  
395 multi-variate singular spectrum analysis (MSSA). Figure 9 shows the past and future temporal evolution  
396 of two toroidal flow coefficients,  $t_1^0, t_3^0$ . These coefficients are of a particular interest, as they relate to  
397 changes in Earth's rotation (Jault et al. 1988; Jackson et al. 1993), which coincide with geomagnetic jerks  
398 (Holme & de Viron 2005; Wardinski et al. 2008).

399 The MSSA of the temporal flow variability is truncated at the degree where past model and forecast of  
400 the flow coefficients becomes almost continuous, which is achieved for a truncation level  $k = 22$ , i.e. the  
401 **first 22 eigenmodes**. Both flow coefficients show spurious accelerations from 2017 on, which is caused the  
402 endpoint constraint of the flow acceleration, cf. (23). Therefore, we consider the kinematic field and flow  
403 coefficients to be flawed and start the kinematic forecast from 2017.

404 Figure 8 also provides a comparison between the direct and kinematic forecasts of secular variation  
405 coefficients. The kinematic forecast starts in 2017 to reduce the influence of the model's faulty temporal  
406 behavior close to its endpoint. These forecasts differ by their temporal evolution to those of the direct  
407 secular variation forecast. Most apparent are these differences in the forecast of  $\dot{g}_1^0$ , where both forecasts  
408 have opposite signs and differ by about 20 nT/yr. The kinematic forecast predicts a gently varying  $\dot{g}_1^0$ .

409 This forecast discrepancy is also clearly seen for  $\dot{g}_1^0, \dot{g}_1^1$  and  $\dot{h}_2^1$ . However, these forecasts show also common  
410 features related to future geomagnetic jerks. The cause for the forecast discrepancy is not understood,  
411 and it is not in relation to the anomalous flow acceleration at the kinematic model endpoint, as the  
412 kinematic forecast starts a few years prior to the model's endpoint.

### 413 **Predictability**

414 Previous studies (Aubert 2015; Whaler & Beggan 2015) used rms-based measures (31) to quantify  
415 differences between a reference model and forecasts. The reference model is a model for the epoch  
416 2017 of the classical and kinematic model branches and labeled as M1 2017 and M2 2017, respectively.

417 Table 2 lists the rms-differences between different epoch models of Model 1 and Model 2. Generally, the  
418 rms-differences of the direct secular variation forecasts (Model 1) is smaller than those of the kinematic  
419 forecasts. However, it is not clear whether it allows to conclude upon the performance of the forecasts.

420 Globally, it suggests that the main field and secular variation of Model 1 tend to be more similar over  
421 longer time intervals than for Model 2. This may agree with results based on numerical dynamos (Hulot  
422 et al. 2010) and data assimilation using numerical dynamos (Aubert 2015). Their studies used rms-based  
423 error estimates, and their results suggested a possible predictability of several decades to a century.

424 We perform several prediction experiments to measure the prediction length of the first 26 secular  
425 variation coefficients for different prediction setups. In these setups, the starting of the forecast is set to  
426 1995, 2005 and 2015, respectively. The forecast period is in all cases 10 years, for which we compute the  
427  $\text{sAPE}(t)$ , cf. (30).

428 First, it is found for the direct secular variation forecasts of the classical field model that the prediction  
429 length is largely independent of the number of eigenvalues considered in the MSSA forecast. Merely, this  
430 number has to be larger than the number of significant eigenvalues.

431 Results of the forecast experiments are shown in Figures 10 and 11. Generally, the prediction error  
432 ( $\text{sAPE}(t)$ ) of the zonal secular variation coefficients increases with time, except for the prediction of  $\dot{g}_4^0$   
433 of the 1995-experiment which improves with time. It is also noted that prediction error of the 2015-  
434 experiment is mostly smaller than for the other experiments (Figure 10). **We focus on the prediction of**  
435 **the large-scale secular variation represented by the first coefficients. For practical purposes, we arbitrarily**  
436 **examine the first 26 coefficients.** The mean prediction length for which the prediction error  $\text{sAPE}(t) < 10\%$   
437 of the first 26 secular variation coefficients is around 3 years (Figure 11). This is substantially shorter  
438 than results by Aubert (2015), who found a possible predictability of a century.

439 Similar experiments are performed for the flow forecasts of the kinematic field model (Model 2), with the  
440 resulting prediction lengths shown in Figure 12. The mean prediction length for all experiments is always  
441 shorter than 2.5 years, i.e. shorter than the mean prediction length of direct secular variation forecasts  
442 (Figure 11). However, we note that the prediction length of  $t_1^0$  for experiments started in 1995 and 2005  
443 is longer or equal to 5 years, and that only the forecast deteriorates only when the experiment is started  
444 closer to the model endpoint.

445 We avoided performing such experiments for the kinematic secular variation forecast, because of the  
446 faulty temporal behavior of Model 2 close to its endpoints. This behavior is caused by the constraint  
447 that controls the flow acceleration at the model endpoints by  $\lambda_E$ . Constraining the flow acceleration too  
448 strongly leads to a very small flow acceleration, which is as unrealistic as the counter case with large flow  
449 acceleration. To this end, conclusions about the flow acceleration at the endpoints are loose, and in the  
450 future we may constrain secular variation of kinematic field models to be similar to the secular variation  
451 of the classical model.

#### 452 **Derivation of candidate models**

453 Our candidate models for the DGRF 2015, IGRF 2020, and the secular variation for the period 2020 to  
454 2025 are derived from Model 1 and its respective forecast. This is justified by the large resemblance of  
455 the main field description of two models. The candidates for the DGRF 2015 and IGRF 2020 are given  
456 by the main field model in 2015.0 and 2020.0, respectively, truncated at spherical harmonic degree 13,  
457 in nT with two decimal places. The candidate of the secular variation model is given by the forecast for  
458 the epoch 2022.5 truncated to spherical harmonic degree 8.

#### 459 **Conclusion**

460 We derive field models for the period 1957 to 2020 from ground-based geomagnetic observatory data  
461 as well as satellite-based virtual observatory data. These models are constructed using two different  
462 techniques, a classical modeling (Model 1) that provides an optimal fit to observations, and a data  
463 assimilation to a dynamical assumption of Earth's core flow (Model 2). These strategies provide similar  
464 results for the core field and its temporal variations over the past decades.

465 In this study we set up two forecasting schemes that rely on analyses of multi-variate time series of  
466 secular variation coefficients. We derived time series models of the field variability, from which forecasts of  
467 individual secular variation coefficients are obtained. These serve as direct secular variation predictions.

468 Prediction experiments indicate a robust forecast of the secular variation of about three years. This  
469 might be a lower boundary and is determined by our cautious definition of the prediction length; the  
470 time interval when  $\text{sAPE}(t) > 10\%$ .

471 Forecasts of the flow coefficients are derived in the same manner and used in forward calculations of  
472 the secular variation due to advection in the core, where contributions from magnetic diffusion are  
473 neglected and a tangential geostrophy assumption couples the toroidal and poloidal flows (kinematic  
474 secular variation forecast). This approach extends beyond the approach of Beggan & Whaler (2010);  
475 Whaler & Beggan (2015) which used steady and constantly accelerated flow to predict future secular  
476 variation. However, our kinematic secular variation forecast suffers from faulty estimates of the core flow  
477 at the model endpoints due to a possibly inaccurate restriction of the flow acceleration. Therefore, the  
478 latter approach may not provide robust forecasts of the secular variation, **unless the endpoint constraint**  
479 **can be dropped**. Interestingly, both strategies consistently indicate the occurrence of future geomagnetic  
480 jerks in 2021 and 2024. **The uncertainty in the exact timing of these events is related to the original**  
481 **temporal resolution of our data, which is optimistically smaller than  $\pm 1$  year and realistically larger than**  
482  **$\pm 6$  months, as well as to the strength of the temporal constraint.**

483 Off course, our strategies of forecasting are limited by the lack of observations from within the (geo-  
484 dynamo) system (which are not available), but not by inferences made upon the geodynamo. Similar  
485 limitations may occur to approaches using a full description of the geodynamo in a data assimilation  
486 framework to forecast geomagnetic secular variation, like Aubert (2015); Fournier et al. (2015). We  
487 hypothesize that when longer time series are considered a longer behavior of the field can be modeled;  
488 however, this will not improve the predictability of the short term (decadal) field variations, as they may  
489 be chaotic.

490 To this end, results of the classical modeling (Model 1) are used to provide candidates for the definitive  
491 geomagnetic reference field model (DGRF) in 2015, and IGRF candidate model for 2020. The candidate  
492 of the secular variation model is given by the forecast for the epoch 2022.5.

493 **List of abbreviations**

494 CMB - core-mantle boundary

495 DGRF - Definitive Geomagnetic Reference Field

496 IAGA - International Association of Geomagnetism and Aeronomy

497 IGRF - International Geomagnetic Reference Field

498 MSSA - multi-variate singular spectrum analysis

499 sAPE - symmetric absolute prediction error

500 VO - virtual geomagnetic observatory

501 **Declarations**

502 **Availability of data and materials**

503 The datasets used and/or analysed during the current study are available from the corresponding author  
504 on reasonable request.

505 **Competing interests**

506 The authors declare that they have no competing interests.

507 **Funding**

508 IW was funded at an initial state of this study by the ESA project 3D-Earth and by the Centre National  
509 des Etudes Spatiales (CNES) within the context of the project “Exploitation des mesures de la mission  
510 Swarm”. DS was funded by CNES project “Exploitation des mesures de la mission Swarm”. HA  
511 acknowledges financial support from “Programme National de Planétologie” of “Institut national des  
512 sciences de l’Univers”. AC acknowledges financial support from CNES over the project “Indices d’Activité  
513 Magnétique : Météorologie de l’Espace et Gomagnetisme Spatial”. This work was supported by CNES.

514 **Authors’ contributions**

515 I. Wardinski organized the manuscript based on the analyses of the contributing co-authors. All authors  
516 contributed modeling results and/or detailed technical analyses and/or discussions for this study. All  
517 co-authors have read and approved the manuscript.

518 **Acknowledgments**

519 The authors would like to record their gratitude to scientists working at magnetic observatories, data  
 520 suppliers and involved national institutes for making geomagnetic observatory data available, to the  
 521 World Data Centre for Geomagnetism (Edinburgh) for compiling these data in a database and to the  
 522 International Service of Geomagnetic Indices for making data products (indices) available. The authors  
 523 acknowledge ESA for providing access to Swarm Level 1b products.”We also would like to thank V.  
 524 Lesur for sharing his satellite-based geomagnetic field model before its submission. This work was  
 525 supported by CNES. The critical reading and very valuable comments by two anonymous reviewers are  
 526 appreciated.

527 **Appendix A. Theory of singular spectrum analysis**

528 We briefly recall some aspects of the singular spectrum analysis (SSA). The SSA is a non-parametric  
 529 spectral estimation method that incorporates aspects of classical time series analysis, signal processing,  
 530 multivariate statistics and geometry (Vautard et al. 1992; Golyandina et al. 2001). We first formulate  
 531 the theoretical concepts for uni-variate time series, and then extend them to multi-variate time series.  
 532 For a deeper review and also its application to a variety of phenomena in meteorology, oceanography  
 533 and climatology we refer to (Ghil et al. 2002) and references therein.

534 **Uni-variate analysis**

535 The analysis is based on the embedding of a time series  $y_t$  in a vector space of dimension  $M$  that  
 536 determines the longest periodicity captured by the analysis. The spectral information on a time series  
 537 are obtained by diagonalizing the  $M \times M$  covariance matrix  $\mathbf{C}_y$  of  $y_t$ . The covariance matrix  $\mathbf{C}_y$  can be  
 538 estimated directly from the data, i.e., its entries  $c_{ij}$  depend only on the lag  $\delta = |i - j|$ :

$$c_{ij} = \frac{1}{N - \delta} \sum_{t=1}^{N-\delta} y_t \cdot y_{t+\delta}. \quad (\text{A.1})$$

539 Usually, the decomposition of the covariance matrix  $\mathbf{C}_y$  in  $q$  orthogonal eigenvectors  $\mathbf{E}_q$  that are called  
 540 temporal empirical orthogonal functions (EOF) is done by singular value decomposition. The  
 541 eigenvalues  $\epsilon_q$  of  $\mathbf{C}_y$  account for the partial variance in the direction  $\mathbf{E}_q$  and the sum of the eigenvalues,  
 542 i.e., the trace of  $\mathbf{C}_y$ , gives the total variance of the original time series  $y_t$ . By a projection of the time  
 543 series onto the EOF the temporal principal component  $\mathbf{A}_t^k$  can be obtained:

$$A_t^k = \sum_{j=1}^M X(t + j - 1)E_k(j). \quad (\text{A.2})$$

544 In fact, this method decomposes the time series in parts that correspond to trends, oscillatory modes or  
545 noise. Therefore, it allows the time series to be reconstructed and to be forecasted by using linear  
546 combinations of the temporal principal components and EOFs. The reconstructed components  $\mathbf{R}_t^k$  are  
547 given by

$$R_t^k = \frac{1}{N_t} \sum_{k \in K} \sum_{j=L_t}^{U_t} A^k(t-j+1)E^k(j), \quad (\text{A.3})$$

548 where  $K$  is the set of  $k$  EOFs and temporal principal components on which the reconstruction is based.

549 We refer to  $k$  as the truncation level of the reconstruction and forecast. This truncation level is  $k \leq q$ .

550 The values of the normalization factor  $N_t$ , as well as of the lower and upper bound of summation  $L_t$   
551 and  $U_t$ , differ between the central part of the time series and in the vicinity of its endpoints (Ghil et al.  
552 2002).

553 An oscillatory mode is characterized by a pair of nearly equal eigenvalues and by associated principal  
554 components that are in approximate phase quadrature (Ghil et al. 2002). Such a pair can represent a  
555 nonlinear, non-harmonic oscillation, because a single pair of eigenmodes are more sensitive to the basic  
556 periodicity of an oscillatory mode than methods with fixed basis functions, such as the sines and cosines  
557 used in the Fourier transform.

### 558 Multi-variate analysis

559 The multi-variate (or multi-channel) singular spectrum analysis (MSSA) is a generalization of the SSA  
560 from uni-variate to multi-variate time series, such as time series of individual Gauss coefficients. Its use  
561 was proposed theoretically in the context of nonlinear dynamics (Broomhead & King 1986) and  
562 numerous examples of successful application of this methods can be found, e.g. (Plaut & Vautard 1994).

563 The MSSA allows the identification of dynamically relevant temporal patterns that are coherent in series  
564 that form a multi-variate time series. These individual series are often called channels. By analogy to  
565 the SSA, each of  $L$ -channel data vectors  $y_{l,t} : l = 1, \dots, L, t = 1, \dots, N$  is expanded to a state vector

$$Y_l = \begin{pmatrix} y_{l,1} & y_{l,2} & \dots & y_{l,M} \\ y_{l,2} & y_{l,3} & \dots & y_{l,M+1} \\ \dots & \dots & \dots & \vdots \\ y_{l,N-M} & \dots & \dots & y_{l,N-1} \\ y_{l,N-M+1} & \dots & \dots & y_{l,N} \end{pmatrix} \quad (\text{A.4})$$

566 for each channel  $l = 1, \dots, L$ , and the window length  $M$ . Following the approach of (Broomhead &  
567 King 1986; Allen & Robertson 1996), then the MSSA relies on the construction a grand covariance

568 matrix  $C_x$  like

$$C_x = \frac{1}{N - M + 1} Y^T Y = \begin{pmatrix} C_{11} & C_{12} & C_{13} & \dots & C_{1L} \\ & C_{22} & C_{23} & \dots & C_{2L} \\ & & C_{33} & \dots & C_{3L} \\ & & & \ddots & \vdots \\ & & & & C_{LL} \end{pmatrix} \quad (\text{A.5})$$

569 where each block  $C_{ll'}$  is a covariance matrix between channels  $l$  and  $l'$ . The blocks  $C_{ll'}$  have the entries

$$(C_{ll'})_{j,j'} = \frac{1}{N - M + 1} \sum_{t=1}^{N-M+1} Y_{t+j-1}^l Y_{t+j'-1}^{l'}. \quad (\text{A.6})$$

570 The  $LM \times LM$  matrix  $C_x$  is symmetric and by diagonalizing,  $LM$  eigenvectors  $E^k : k = 1, \dots, LM$  can  
571 be obtained. The eigenvectors are composed of  $L$  consecutive segments with length  $M$ . Similarly to

572 (A.2) the associated principal components  $A^k$  can be computed by projecting  $Y$  in the directions of the  
573 eigenvectors (i.e. onto the EOFs):

$$A_t^k = \sum_{j=1}^M \sum_{l=1}^L Y_{t+j-1}^l E^{l,k}(j). \quad (\text{A.7})$$

574  $E^{l,k}(j)$  are the elements of the eigenvectors. The  $A^k(t)$  are single-channel time series and likewise to

575 (A.3) the  $k$ -th reconstruction of the signal of channel  $l$  can be obtained by:

$$R_t^{l,k} = \frac{1}{M_t} \sum_{j=L_t}^{U_t} A^k(t - j + 1) E^{l,k}(j), \quad (\text{A.8})$$

576 where  $L_T$  and  $U_t$  are the lower and upper bound of summation, respectively.

577 Figure 13 shows the first 12 eigenvectors of the toroidal flow decomposition. Clearly, the first two

578 eigenvectors capture the slow variation of flow coefficients, whereas other eigenvectors represent the

579 shorter temporal variations of the flow. The first few eigenvectors explain nearly the entire signal

580 variance, which is given on top of each single plot. However, higher indexed eigenvectors also show

581 non-zero partial variance, suggesting that these eigenvectors may carry relevant temporal information.

582 **(Figure A1 about here)**

583 An important aspect of the analysis is how well components of the time series are separated from each

584 other. The components generally group in two disjunctive parts, one corresponds to the signal, the

585 other to the noise. The signal could be composed of slowly varying, periodic and/or quasi periodic

586 components. A way to measure the separation between components, is to calculate the weighted

587 correlation or w-correlations, as given by Golyandina et al. (2001).

588 **(Figure A2 about here)**

589 In Figure 14 the, so-called, w-correlation matrix is displayed. It shows the weighted correlations for  
590 principal components,  $A^k$ , of the temporal flow variability. Accordingly, large values of w-correlation  
591 exist for diagonal elements of the matrix, where individual modes correlate with themselves (dark red  
592 color). Whereas, small values of the w-correlation (lighter colors) indicate orthogonality between  
593 components, as the scalar product between perpendicular vectors vanishes. Therefore, small correlation  
594 values indicate a good separation of components or noise. In particular, noise components do not  
595 correlate with each other and other components, and therefore, show zero correlation.

596 The w-correlation matrix (Figure 14) is found to be diagonal symmetric, **and we mainly identify two**  
597 **different regions. The first region relates to the first 11 eigenvalues and their modes, where correlations**  
598 **quickly diminish away from the central diagonal. In the second region, from the 11th eigenvector**  
599 **onward, off-diagonal correlations emerge. This suggests that principal components do not fully**  
600 **separate. However, these correlations are minor and those principal components may still carry**  
601 **important information. This is almost identical for toroidal and poloidal flow coefficients.**~~and we can~~  
602 ~~identify four blocks; first, the lower left block of the first 11 components, can be attributed to~~  
603 ~~orthogonal components, i.e. signal. Then, an intermediate block, made up of components 12 to 15~~  
604 ~~(between orange lines). Here, the w-correlation values are non-zero, an indication for a poorer~~  
605 ~~separation of these components. The separation for the components from 16 to 26 improves, and these~~  
606 ~~components carry a significant amount of signal. For  $k > 25$  the separation deteriorate significantly,~~  
607 ~~which we interpret as the leakage of noise into the signal.~~

608 The forecast of the flow variability requires to identify significant and non-noisy components. There are  
609 several possibilities to quantify the statistical significance of the temporal principal components. In the  
610 next section we develop a rule that provides a criterion for the selection of a sufficient correct time  
611 series model of the flow variability.

### 612 **Model selection criterion**

613 Generally, to what extend the variability of the time series can be explained strongly depends on  $k$ , the  
614 number of model parameter. Therefore, a selection criterion is applied which ascertains the statistical  
615 significance of the eigenvalues and their related EOF detected by the MSSA. The recovery of the  
616 temporal variability of the time series can be considered to be most complete, when the maximum  
617 number  $k$  of EOF's (truncation level) are used for the reconstruction of the time series. However, this is

618 for several reasons not desirable, for instance some modes may represent noise that is part of the time  
619 series (and we want a noise-free representation of the flow temporal variability). The goodness of the fit  
620 can be estimated by computing the sum of the squared residuals (RSS) as where  $\hat{y}_t$  is the model value  
621 and  $n$  is the length of the data series  $y_t$ . Here, we understand by model value ( $\hat{y}_t^k$ ) the reconstruction of  
622 the original data series ( $y_t$ ). In Figure 15 the RSS for different truncation levels (composed of  $k$  modes)  
623 are shown.

624 **(Figure A3, about here)**

625 The separation between signal and noise is indicated by the slope break in the eigenvalues spectrum vs.  
626  $k$ . The point  $k^* = S$  at which this break occurs should not be confused with a “dimension”  $D$  of the  
627 underlying deterministic dynamics (Vautard & Ghil 1989). Figure 15 shows the eigenvalue spectra of  
628 different analyses of the secular variation and the toroidal and poloidal flow variability. Broadly, the  
629 first 7 eigenvalues are  $\geq 1$ . A first break in the spectral slopes occurs at  $k \sim 10$ , **which is related to**  
630 **features of the w-correlation matrices shown in Figure 14**. A second break in slope occurs at  $k \sim 15$ .  
631 Thereafter, spectra of the kinematic secular variation and the flow variability continuously decrease.  
632 Based on these results a MSSA-model truncated at a level  $k > 15$  should provide a sufficient  
633 reconstruction of the temporal flow variability. However, in the kinematic forecast, we use  $k = 22$ , at  
634 this value past and predicted flow variability become continuous. This may indicate significant portion  
635 of the signal to be carried by eigenmodes for  $15 < k < 23$ .

636 The eigenvalue spectrum of the secular variation of Model 1 breaks at eigenvalue 11. Thereafter, it is  
637 flat, but eigenvalues remain an order of magnitude larger, than those of the kinematic secular variation  
638 and the flow coefficients.

## 639 **Appendix B. Kinematic field models**

640 Characteristics of kinematic field models for a range of damping parameters are given in Table 3. We  
641 explored purely toroidal (x0a) and tangential geostrophic (x3\*) flow assumptions, respectively. The  
642 parameter controlling the frozen-flux constraint is set to  $\lambda_1 = 1 \times 10^{-9}$  and  $\lambda_1 = 1 \times 10^{-7}$ . Different  
643 setups of the other constraints are also tested, as well as different temporal derivatives of the temporal  
644 norm. We tested

$$\begin{aligned} \mathbb{T} = 0 & : \int_{\mathcal{T}} \int_{S(c)} (\mathbf{u})^2 d\Omega dt \\ \mathbb{T} = 1 & : \int_{\mathcal{T}} \int_{S(c)} (\partial_t \mathbf{u})^2 d\Omega dt. \end{aligned} \tag{B.1}$$

645 **References**

- 646 Abramowitz, M. & Stegun, I. A., 1973. *Handbook of Mathematical Functions*, Dover (New York).
- 647 Alken, P., Thébault, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., Bertrand, F.,  
648 Bondar, T., Boness, A., Brocco, L., Canet, E., Chambodut, A., Chulliat, A., Coisson, P., Civet, F.,  
649 Du, A., Fournier, A., Fratter, I., Gillet, N., Hamilton, B., Hamoudi, M., Hulot, G., Jager, T., Korte,  
650 M., Kuang, W., Lalanne, X., Langlais, B., Léger, J.-M., Lesur, V., Lowes, F. J., Macmillan, S.,  
651 Manda, M., Manoj, C., Maus, S., Olsen, N., Petrov, V., Ridley, V., Rother, M., Sabaka, T. J.,  
652 Saturnino, D., Schachtschneider, R., Sirol, O., Tangborn, A., Thomson, A., Tøffner-Clausen, L.,  
653 Vigneron, P., Wardinski, I., & Zvereva, T., 2020. International Geomagnetic Reference Field: the  
654 13th generation, in prep.
- 655 Allen, M. R. & Robertson, A. W., 1996. Distinguishing modulated oscillations from coloured noise in  
656 multivariate datasets, *Climate Dynamics*, **12**, 775–784.
- 657 Amit, H., 2014. Can downwelling at the top of the Earth’s core be detected in the geomagnetic secular  
658 variation?, *Phys. Earth Planet. Inter.*, **229**, 110–121.
- 659 Armstrong, J. S., 1985. *From Crystal Ball to Computer*, Wiley, 2nd edn.
- 660 Aubert, J., 2015. Geomagnetic forecasts driven by thermal wind dynamics in the Earth’s core, *Geophys.*  
661 *J. Int.*, **203**(3), 1738–1751.
- 662 Aubert, J. & Finlay, C. C., 2019. Geomagnetic jerks and rapid hydromagnetic waves focusing at  
663 Earth’s core surface, *Nature Geoscience*, **12**(5), 393–398.
- 664 Barrois, O., Hammer, M. D., Finlay, C. C., Martin, Y., & Gillet, N., 2018. Assimilation of ground and  
665 satellite magnetic measurements: inference of core surface magnetic and velocity field changes,  
666 *Geophys. J. Int.*, **215**(1), 695–712.
- 667 Beggan, C. & Whaler, K., 2010. Forecasting secular variation using core flows, *Earth, Planets, and*  
668 *Space*, **62**(10), 821–828.
- 669 Bloxham, J., 1988. The dynamical regime of fluid flow at the core surface, *Geophys. Res. Lett.*, **15**,  
670 585–588.
- 671 Bloxham, J. & Jackson, A., 1992. Time-dependent mapping of the magnetic field at the core-mantle  
672 boundary, *J. geophys. Res.*, **97**, 19,537–19,563.
- 673 Bloxham, J., Zatman, S., & Dumberry, M., 2002. The origin of geomagnetic jerks, *Nature*, **420**(6911),  
674 65–68.

675 Box, G. E. P. & Jenkins, G. M., 1976. *Time Series Analysis, Forecasting and Control*, Holden Day, San  
676 Francisco.

677 British Geological Survey - Edinburgh, 2020. Worldwide observatory annual means,  
678 [http://www.geomag.bgs.ac.uk/data\\_service/data/annual\\_means.shtml](http://www.geomag.bgs.ac.uk/data_service/data/annual_means.shtml).

679 Brockwell, P. J. & Davis, R. A., 2002. *Introduction to Time Series and Forecasting*, Springer, 2nd edn.

680 Broomhead, D. S. & King, G. P., 1986. Extracting qualitative dynamics from experimental data,  
681 *Physica D Nonlinear Phenomena*, **20**, 217–236.

682 Brown, W. J., Mound, J. E., & Livermore, P. W., 2013. Jerks abound: An analysis of geomagnetic  
683 observatory data from 1957 to 2008, *Phys. Earth Planet. Inter.*, **223**, 62–76.

684 Canciani, A. & Raquet, J., 2016. Absolute Positioning Using the Earth’s Magnetic Anomaly Field,  
685 *NAVIGATION*, **63**(2), 111–126.

686 Chambodut, A., Schwarte, J., Langlais, B., Lühr, H., & M., M., 2002. The selection of data in field  
687 modeling, in *OIST- Proceedings. 4th Oersted International Science Team Conference*, pp. 201–213,  
688 eds Stauning, P., Lühr, H., Ultré-Guérard, P., LaBrecque, J., Purucker, M., Primdahl, F., Jørgensen,  
689 J., Christensen, F., Høeg, P., & Lauritsen, K., Danish Meteorological Institute, Copenhagen,  
690 Denmark.

691 Chambodut, A., Marchaudon, A., Menvielle, M., El-Lemdani Mazouz, F., & Lathuillere, C., 2013. The  
692 k-derived mlt sector geomagnetic indices, *Geophys. Res. Lett.*, **40**(18), 4808–4812.

693 Chen, Z. & Yang, Y., 2004. Assessing forecast accuracy measures, available on Researchgate.

694 Finlay, C. C., Maus, S., Beggan, C. D., Bondar, T. N., Chambodut, A., Chernova, T. A., Chulliat, A.,  
695 Golovkov, V. P., Hamilton, B., Hamoudi, M., Holme, R., Hulot, G., Kuang, W., Langlais, B., Lesur,  
696 V., Lowes, F. J., Lühr, H., MacMillan, S., Manda, M., McLean, S., Manoj, C., Menvielle, M.,  
697 Michaelis, I., Olsen, N., Rauberg, J., Rother, M., Sabaka, T. J., Tangborn, A., Tøffner-Clausen, L.,  
698 Thébault, E., Thomson, A. W. P., Wardinski, I., Wei, Z., & Zvereva, T. I., 2010. International  
699 Geomagnetic Reference Field: the eleventh generation, *Geophys. J. Int.*, **183**, 1216–1230.

700 Finlay, C. C., Aubert, J., & Gillet, N., 2016. Gyre-driven decay of the Earth’s magnetic dipole, *Nature*  
701 *Communications*, **7**, 10422.

702 Flores, B. E., 1986. A pragmatic view of accuracy measurement in forecasting, *Omega*, **14**(2), 93–98.

703 Fournier, A., Aubert, J., & Thébault, E., 2015. A candidate secular variation model for IGRF-12 based  
704 on Swarm data and inverse geodynamo modelling, *Earth, Planets, and Space*, **67**, 81.

705 Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W.,  
706 Saunders, A., Tian, Y., Varadi, F., & Yiou, P., 2002. Advanced spectral methods for climatic time  
707 series, *Rev. Geophys.*, **40**, 1–41.

708 Golyandina, N., Nekrutkin, V., & Zhigljavsky, A., 2001. *Analysis of Time Series Structure: SSA and*  
709 *Related Techniques*, Chapman & Hall.

710 Hamilton, B., Ridley, V. A., Beggan, C. D., & Macmillan, S., 2015. The BGS magnetic field candidate  
711 models for the 12th generation IGRF, *Earth, Planets, and Space*, **67**, 69.

712 Hills, R. G., 1979. *Convection in the Earth's mantle due to viscous shear at the core-mantle interface*  
713 *and due to large-scale buoyancy*, Ph.D. thesis, New Mexico State University, Las Cruces.

714 Holme, R., 2007. Large-scale flow in the core, in *Core Dynamics*, vol. 8 of **Treatise on Geophysics**,  
715 chap. 4, pp. 107–130, ed. Olson, P., Elsevier.

716 Holme, R. & de Viron, O., 2005. Geomagnetic jerks and a high-resolution length-of-day profile for core  
717 studies, *Geophys. J. Int.*, **160**, 435–440.

718 Holme, R. & de Viron, O., 2013. Characterization and implications of intradecadal variations in length  
719 of day, *Nature*, **499**(7457), 202–204.

720 Hulot, G., Lhuillier, F., & Aubert, J., 2010. Earth's dynamo limit of predictability, *Geophys. Res. Lett.*,  
721 **37**, 6305.

722 International Service of Geomagnetic indices, 2020. , <http://isgi.unistra.fr>.

723 Jackson, A., 1997. Time-dependency of tangentially geostrophic core surface motions, *Phys. Earth*  
724 *Planet. Inter.*, **103**, 293–311.

725 Jackson, A., Bloxham, J., & Gubbins, D., 1993. Time-dependent flow at the core surface and  
726 conservation of angular momentum in the coupled core-mantle system, in *Dynamics of the Earth's*  
727 *deep interior and Earth rotation*, pp. 97–107, eds LeMouël, J.-L., Smylie, D. E., & Herring, T.,  
728 AGU/IUGG.

729 Jault, D., Gire, C., & LeMouël, J. L., 1988. Westward drift, core motions and exchanges of angular  
730 momentum between core and mantle, *Nature*, **333**, 353–356.

731 Jiménez, A. R., Seco, F., Zampella, F., Prieto, J. C., & Guevara, J., 2012. Improved heuristic drift  
732 elimination with magnetically-aided dominant directions (MiHDE) for pedestrian navigation in  
733 complex buildings, *Journal of Location Based Services*, **6**(3), 186–210.

734 Kuang, W., Wei, Z., Holme, R., & Tangborn, A., 2010. Prediction of geomagnetic field with data

735 assimilation: a candidate secular variation model for IGRF-11, *Earth, Planets, and Space*, **62**(10),  
736 775–785.

737 LeMouél, J.-L., 1984. Outer core geostrophic flow and secular variation of Earth’s magnetic field,  
738 *Nature*, **311**, 734–735.

739 Lesur, V., Wardinski, I., Asari, S., Minchev, B., & Manda, M., 2010. Modelling the Earth’s core  
740 magnetic field under flow constraints, *Earth, Planets, and Space*, **62**, 503–516.

741 Lesur, V., Wardinski, I., Baerenzung, J., & Holschneider, M., 2018. On the frequency spectra of the  
742 core magnetic field Gauss coefficients, *Phys. Earth Planet. Inter.*, **276**, 145–158.

743 Livermore, P. W., Hollerbach, R., & Finlay, C. C., 2017. An accelerating high-latitude jet in Earth’s  
744 core, *Nature Geoscience*, **10**(1), 62–68.

745 Lloyd, D. & Gubbins, D., 1990. Toroidal fluid motion at the top of the Earth’s core, *Geophys. J. Int.*,  
746 **100**, 455–467.

747 Manda, M. & Olsen, N., 2006. A new approach to directly determine the secular variation from  
748 magnetic satellite observations, *Geophys. Res. Lett.*, **33**, 15306.

749 Manda, M., Holme, R., Pais, A., Pinheiro, K., Jackson, A., & Verbanac, G., 2010. Geomagnetic Jerks:  
750 Rapid Core Field Variations and Core Dynamics, *Space Sci. Rev.*, **155**(1-4), 147–175.

751 Mathews, B. P. & Diamantopoulos, H., 1994. Towards a taxonomy of forecast error measures: a  
752 factor-comparative investigation of forecast error dimensions, *Journal of Forecasting*, **13**, 409–416.

753 Olson, P. & Amit, H., 2006. Changes in earth’s dipole, *Naturwissenschaften*, **93**, 519–542.

754 Pais, M. A. & Jault, D., 2008. Quasi-geostrophic flows responsible for the secular variation of the  
755 Earth’s magnetic field, *Geophys. J. Int.*, **173**, 421–443.

756 Pais, M. A., Oliveira, O., & Nogueira, F., 2004. Nonuniqueness of inverted core-mantle boundary flows  
757 and deviations from tangential geostrophy, *J. geophys. Res.*, **109**, 8105.

758 Plaut, G. & Vautard, R., 1994. Spells of Low-Frequency Oscillations and Weather Regimes in the  
759 Northern Hemisphere., *Journal of Atmospheric Sciences*, **51**, 210–236.

760 Purucker, M. E., Sabaka, T. J., & Langel, R. A., 1996. Conjugate gradient analysis: A new tool for  
761 studying satellite magnetic data sets, *Geophys. Res. Lett.*, **23**, 507–510.

762 Roberts, P. H. & Scott, S., 1965. On the analysis of secular variation, 1, A hydromagnetic constraint:  
763 Theory, *J. Geomag. Geoelectr.*, **17**, 137–151.

764 Saturnino, D., Langlais, B., Amit, H., Civet, F., Manda, M., & Beucler, E., 2018. Combining virtual

765 observatory and equivalent source dipole approaches to describe the geomagnetic field with swarm  
766 measurements, *Phys. Earth Planet. Inter.*, **276**, 118–133.

767 Slavinskis, A., Kvell, U., Kulu, E., Sünter, I., Kuuste, H., S., L., Voormansik, K., & Noorma, M., 2014.  
768 High spin rate magnetic controller for nanosatellites, *Acta Astronautica*, **95**, 218–226.

769 Terra-Nova, F., Amit, H., Hartmann, G. A., Trindade, R. I. F., & Pinheiro, K. J., 2017. Relating the  
770 South Atlantic Anomaly and geomagnetic flux patches, *Phys. Earth Planet. Inter.*, **266**, 39–53.

771 Thébaud, E., Finlay, C. C., Beggan, C. D., Alken, P., Aubert, J., Barrois, O., Bertrand, F., Bondar, T.,  
772 Boness, A., Brocco, L., Canet, E., Chambodut, A., Chulliat, A., Coisson, P., Civet, F., Du, A.,  
773 Fournier, A., Fratter, I., Gillet, N., Hamilton, B., Hamoudi, M., Hulot, G., Jager, T., Korte, M.,  
774 Kuang, W., Lalanne, X., Langlais, B., Léger, J.-M., Lesur, V., Lowes, F. J., Macmillan, S., Mandea,  
775 M., Manoj, C., Maus, S., Olsen, N., Petrov, V., Ridley, V., Rother, M., Sabaka, T. J., Saturnino, D.,  
776 Schachtschneider, R., Sirol, O., Tangborn, A., Thomson, A., Tøffner-Clausen, L., Vigneron, P.,  
777 Wardinski, I., & Zvereva, T., 2015. International Geomagnetic Reference Field: the 12th generation,  
778 *Earth, Planets, and Space*, **67**, 79.

779 Tøffner-Clausen, L. & Nielsen, J. B., 2018. Swarm Level 1b Product Definition, SW-RS-DSC-SY-0007,  
780 Issue 5.22, <https://earth.esa.int/documents/10174/1514862/Swarm.L1b.Product.Definition>.

781 Torta, J. M., Pavón-Carrasco, F. J., Marsal, S., & Finlay, C. C., 2015. Evidence for a new geomagnetic  
782 jerk in 2014, *Geophys. Res. Lett.*, **42**(19), 7933–7940.

783 Vautard, R. & Ghil, M., 1989. Singular spectrum analysis in nonlinear dynamics, with applications to  
784 paleoclimatic time series, *Physica D Nonlinear Phenomena*, **35**, 395–424.

785 Vautard, R., Yiou, P., & Ghil, M., 1992. Singular-spectrum analysis: A toolkit for short, noisy chaotic  
786 signals, *Physica D Nonlinear Phenomena*, **58**, 95–126.

787 Wardinski, I. & Holme, R., 2006. A time-dependent model of the earth’s magnetic field and its secular  
788 variation for the period 1980 to 2000, *J. geophys. Res. (Solid Earth)*, **111**, 12101.

789 Wardinski, I. & Holme, R., 2011. Signal from noise in geomagnetic field modelling: denoising data for  
790 secular variation studies, *Geophys. J. Int.*, **185**, 653–662.

791 Wardinski, I. & Lesur, V., 2012. An extended version of the C<sup>3</sup>FM geomagnetic field model:  
792 application of a continuous frozen-flux constraint, *Geophys. J. Int.*, **189**, 1409–1429.

793 Wardinski, I., Holme, R., Asari, S., & Mandea, M., 2008. The 2003 geomagnetic jerk and its relation to  
794 the core surface flows, *Earth and Planetary Science Letters*, **267**, 468–481.

Figure 1. The distribution of ground-based and satellite virtual geomagnetic observatories. The ground-based observatory data (blue circles) cover the period 1957 to 2018, whereas the VO data (red stars) are derived from Swarm and cover the period 2015 to 2019.5.

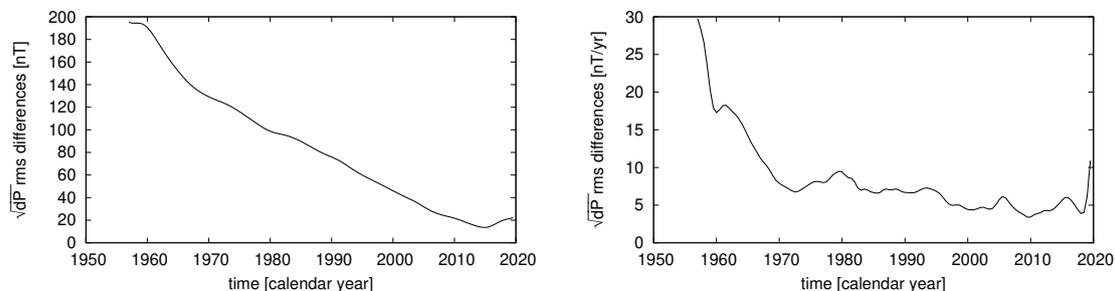


Figure 2. Temporal evolution of the rms difference ( $\sqrt{dP}$ ) of the main field (left) and secular variation (right) between models 1 and 2.

Figure 3. Observed and modeled secular variation at some observatory sites. From top to bottom: Sitka (Alaska), Chambon-la-Forêt (France), Kakioka (Japan), Hermanus (South Africa), Gwangara (Australia). The gray dots represent the observed monthly secular variation in X, Y and Z (from left to right). Solid curves display the modeled secular variation of Model 1 (black) and Model 2 (red). Vertical lines mark occurrences of geomagnetic jerks.

Figure 4. Comparison of six modeled secular variation coefficients with the largest amplitude. Same colors are applied as in Figure 3.

795 Whaler, K. A., 1980. Does the whole of the Earth's core convect?, *Nature*, **287**, 528–530.

796 Whaler, K. A. & Beggan, C. D., 2015. Derivation and use of core surface flows for forecasting secular  
797 variation, *J. geophys. Res. (Solid Earth)*, **120**(3), 1400–1414.

798 World Data Centre for Geomagnetism - Edinburgh, 2019. Geomagnetic Data Master Catalogue,  
799 <http://www.wdc.bgs.ac.uk/catalog/master.html>.

Figure 5. The temporally averaged secular variation spectra of Models 1 and 2 at Earth's surface (left) and at the core surface (right). Same line styles are applied as in Figure 3. The vertical bars indicate the range, i.e. the temporal variability of the individual spectra.

Figure 6. Radial component of the magnetic field at the core surface (left) and its secular variation (right) derived from Model 1 for epochs 1969, 2010 and 2020.

Figure 7. As in Figure 6 for Model 2.

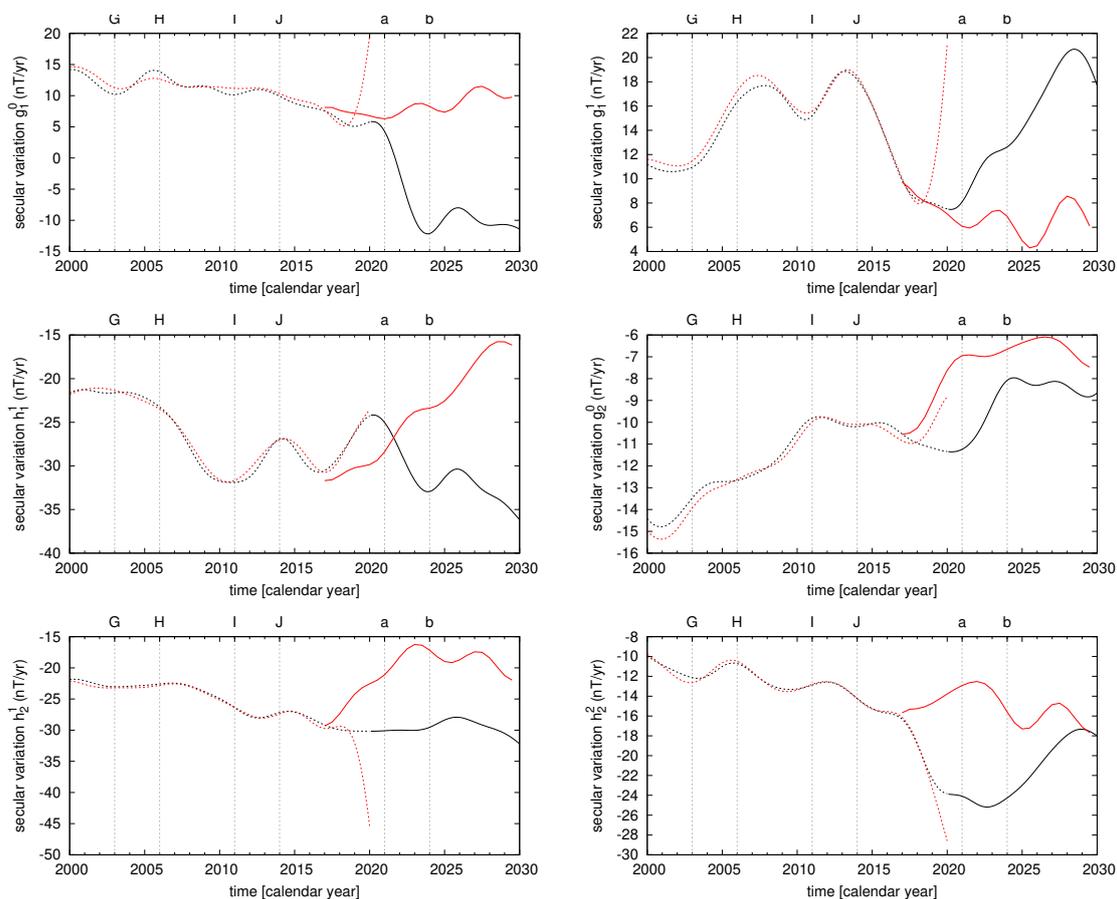


Figure 8. Comparison of six observed and predicted secular variation coefficients with the largest amplitude at Earth's surface. Model 1 (dashed black line), Model 2 (dashed red line) and their forecasts (solid lines). Vertical lines labeled with a and b mark two possible geomagnetic jerks. The other labels mark occurrences of known jerks.

Figure 9. Evolution of the toroidal flow coefficients  $t_1^0$  (left) and  $t_3^0$  (right); their past variations (black) and their forecasts (red) based on the MSSA. The vertical line labeled with FC marks the start of the forecast.

Figure 10. Evolution of the sAPE(t) of zonal secular variation coefficients based on Model 1 of the three forecasts experiments started in 1995 (top panel), in 2005 (middle panel) and in 2015 (bottom panel), respectively.

Figure 11. Prediction lengths defined by the time interval when sAPE(t) > 10% of different forecast experiments base on Model 1 with forecasts started in 1995 (top panel), in 2005 (middle panel) and in 2015 (bottom panel), respectively. The solid horizontal red lines represent mean prediction lengths in years derived from the first 26 secular variation coefficients, i.e  $g_1^0 \dots g_5^1$ . Top axes indicate degrees of the spherical harmonic expansion.

Figure 12. Prediction lengths for the first 26 toroidal flow coefficients estimated from different forecast experiments with forecasts started in 1995 (top panel), in 2005 (middle panel) and in 2015 (bottom panel), respectively. The solid horizontal red lines represent mean prediction lengths in years derived from the first 26 toroidal flow coefficients. Top axes indicate degrees of the spherical harmonic expansion.

Figure 13. The first 12 eigenvectors of the toroidal flow temporal variability. The x-axes of each subplot represent the model interval (1957-2020). The percentages of the partial variance of each eigenvector are given in the top bar of the individual plots.

Figure 14. The weighted correlation matrix. Red colors represent stronger correlation between eigenvectors and therefore a weaker separation. The color scale corresponds to absolute values of correlations from 0 (white) to 1 (dark red).

Figure 15. The eigenvalue spectra of the classical, kinematic secular variation, toroidal and poloidal flow variability, respectively. The horizontal line marks eigenvalues equal 1.

	standard deviation [nT/yr]			damping	
	X	Y	Z	$\lambda_S$	$\lambda_T$
Model 1	13.69	11.07	14.47	3.0d-10	3.0d-2
Model 2	12.42	10.89	13.89	see Model x3d Table 3	

Table 1. Model diagnostics: residuals standard deviation in [nT/yr] and damping parameters.

	main field		secular variation	
	M1 2017	M2 2017	M1 2017	M2 2017
2020	271.41	863.49	21.00	22.11
2025	306.11	1511.43	34.70	40.79
2030	306.54	359.97	35.13	90.02

Table 2. The rms differences between models of different epochs and Model 1 and Model 2. Differences of main field models are given in [nT] and differences of secular variation models are in [nT/yr].

	$\lambda_s$	$\lambda_T$	$\lambda_{TG}$	$\lambda_E$	S/T	R	M
x0a	$1 \times 10^{-3}$	$1 \times 10^{-4}$	-	$1 \times 10^2$	0/1	$0.957 \times 10^{-3}$	2.1184
x3a	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^3$	$1 \times 10^2$	0/1	$0.791 \times 10^{-3}$	2.1196
x3b	$1 \times 10^{-3}$	$1 \times 10^{-5}$	$1 \times 10^3$	$1 \times 10^2$	0/1	$0.941 \times 10^{-3}$	2.1185
x3c	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^4$	$1 \times 10^2$	0/1	$0.123 \times 10^{-2}$	2.1233
x3d	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^5$	$1 \times 10^2$	0/1	$0.102 \times 10^{-2}$	2.1237
x3e	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^6$	$1 \times 10^2$	0/1	$0.166 \times 10^{-3}$	2.1431
x3f	$1 \times 10^{-3}$	$1 \times 10^{-3}$	$1 \times 10^4$	$1 \times 10^2$	0/1	$0.774 \times 10^{-3}$	2.1322
x3g	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^{10}$	$1 \times 10^2$	0/1	$0.136 \times 10^{-2}$	2.1627
x3g1 <sup>†</sup>	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^{10}$	$1 \times 10^2$	0/1	$0.431 \times 10^{-3}$	2.1856
x3h	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^{12}$	$1 \times 10^2$	0/1	$0.106 \times 10^{-2}$	2.1632
x3j	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^8$	$1 \times 10^2$	0/0	$0.311 \times 10^{-2}$	2.1695
x3k	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^7$	$1 \times 10^2$	0/0	$0.696 \times 10^{-2}$	2.1634
x3k1 <sup>†</sup>	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^7$	$1 \times 10^2$	0/0	$0.332 \times 10^{-2}$	2.2221
x3l	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^6$	$1 \times 10^2$	0/0	$0.129 \times 10^{-1}$	2.1525
x3m	$1 \times 10^{-3}$	$1 \times 10^{-4}$	$1 \times 10^4$	$1 \times 10^2$	0/0	$0.524 \times 10^{-3}$	2.1396

Table 3. Specification of damping parameters of the frozen-flux field model based on the tangential geostrophic flow assumption, except model x0a, which is based on a purely toroidal flow assumption. The parameter controlling the frozen-flux constraint is kept fixed, i.e.  $\lambda_1 = 1 \times 10^{-9}$  for most of the inversions, except those marked with <sup>†</sup>, there we set  $\lambda_1 = 1 \times 10^{-7}$ . In all inversions, the parameter controlling the flow acceleration at model end points is set to  $\lambda_4 = 1 \times 10^2$ . S/T identifies which temporal derivative of the spatial and temporal norm is applied, respectively (see text). R and M are defined by (26) and (25), respectively. M is given in nT/yr. The model x0a represents a field model with purely toroidal flow.

# Figures

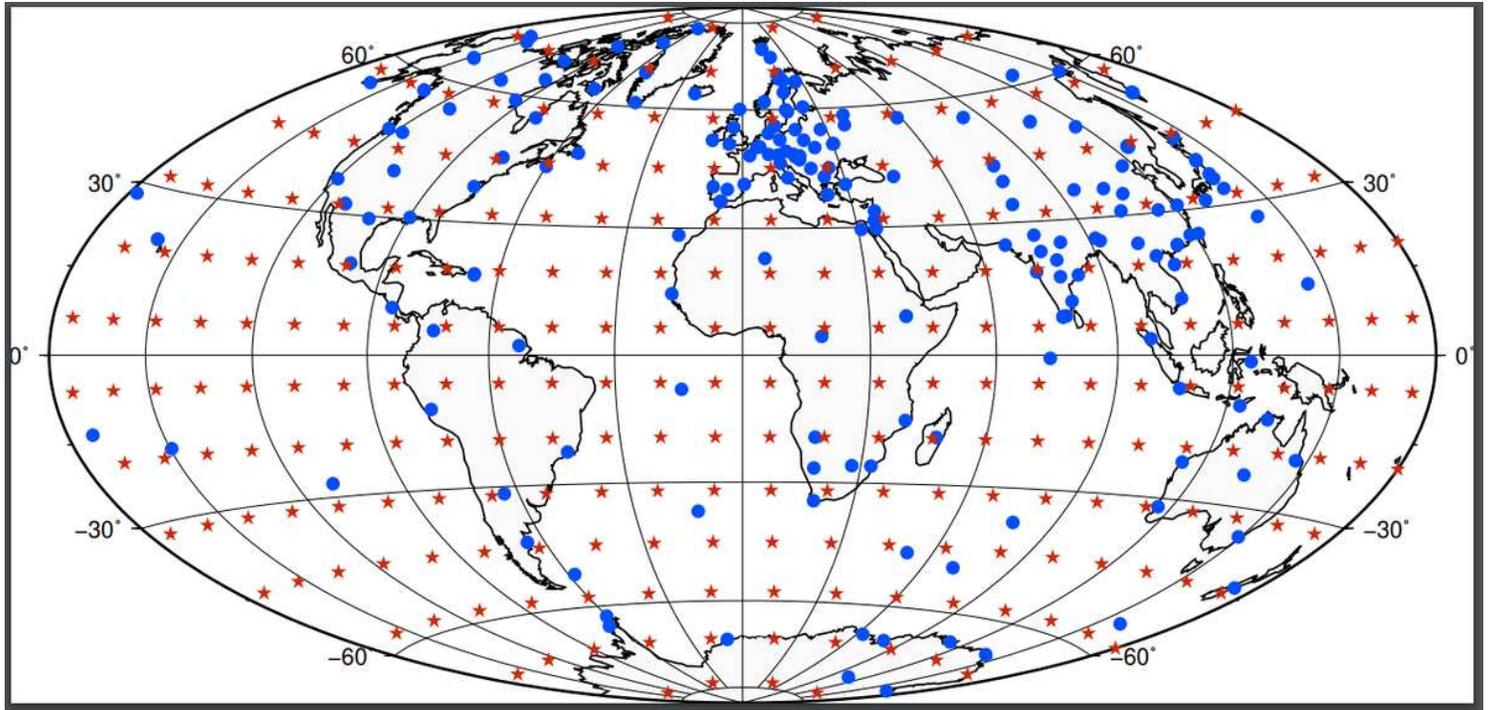


Figure 1

The distribution of ground-based and satellite virtual geomagnetic observatories. The ground-based observatory data (blue circles) cover the period 1957 to 2018, whereas the VO data (red stars) are derived from Swarm and cover the period 2015 to 2019.5.

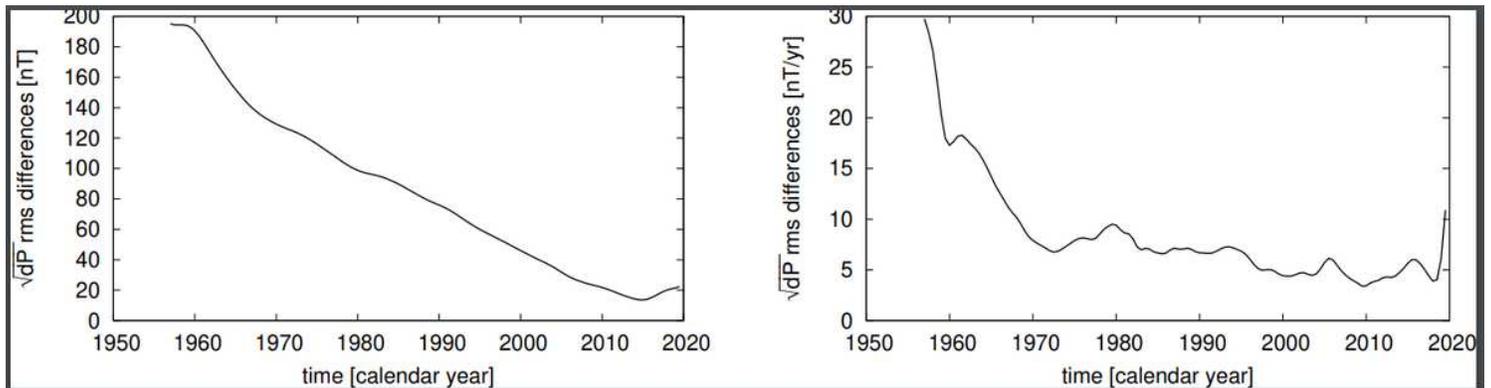
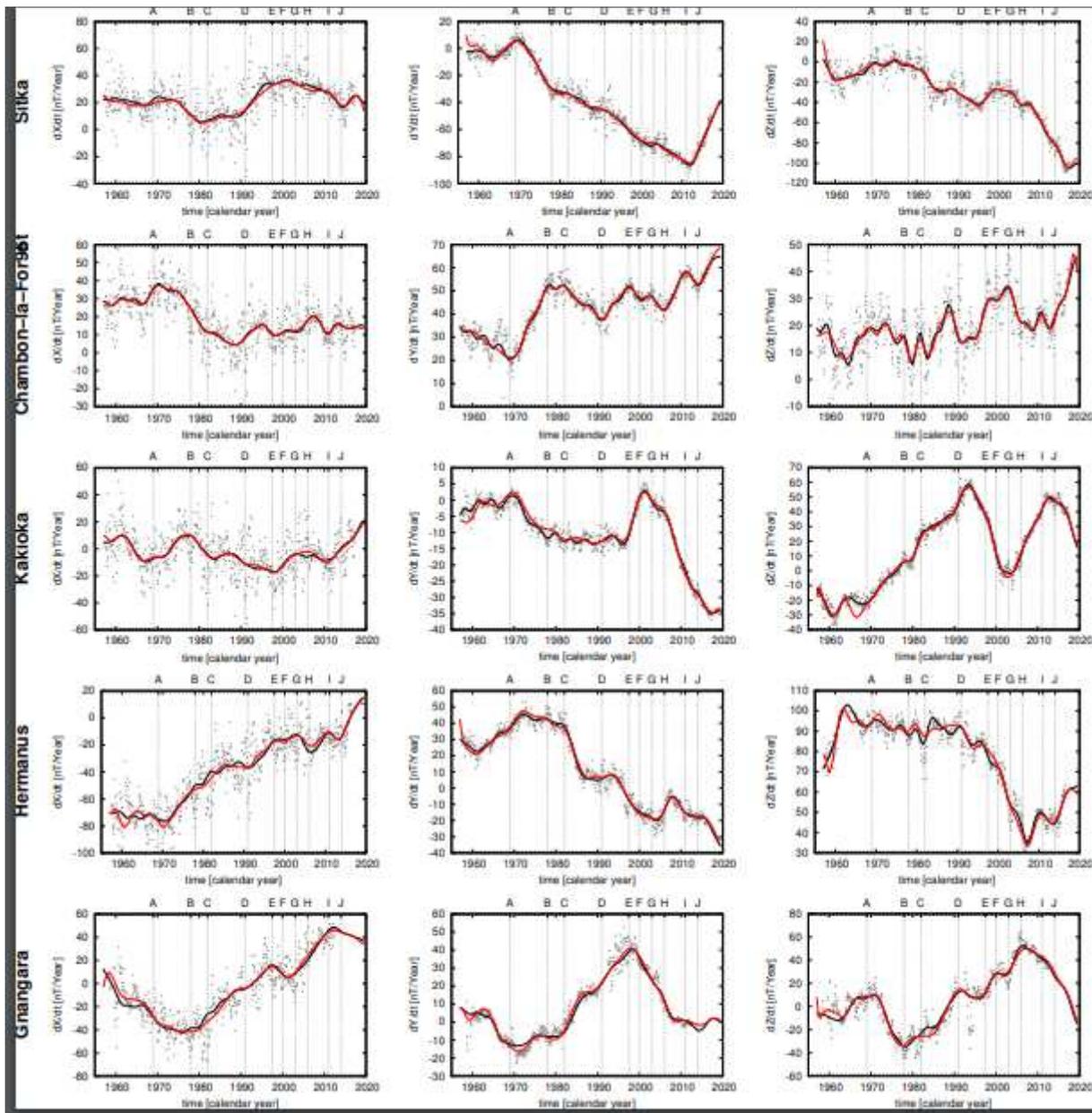


Figure 2

Temporal evolution of the rms difference ( $\sqrt{dP}$ ) of the main field (left) and secular variation (right) between models 1 and 2.



**Figure 3**

Observed and modeled secular variation at some observatory sites. From top to bottom: Sitka (Alaska), Chambon-la-Forêt (France), Kakioka (Japan), Hermanus (South Africa), Ghangara (Australia). The gray dots represent the observed monthly secular variation in X, Y and Z (from left to right). Solid curves display the modeled secular variation of Model 1 (black) and Model 2 (red). Vertical lines mark occurrences of geomagnetic jerks

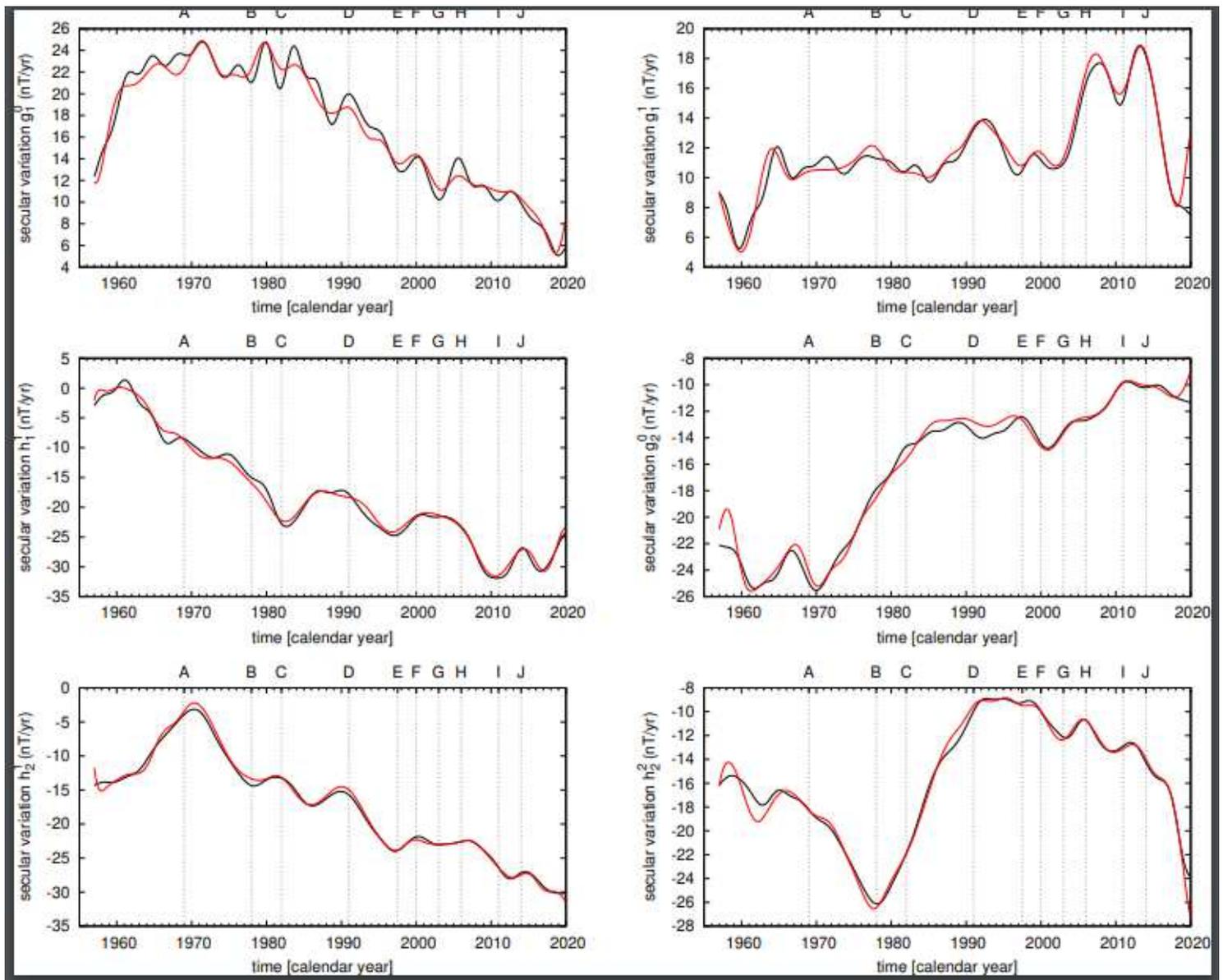


Figure 4

Comparison of six modeled secular variation coefficients with the largest amplitude. Same colors are applied as in Figure 3.

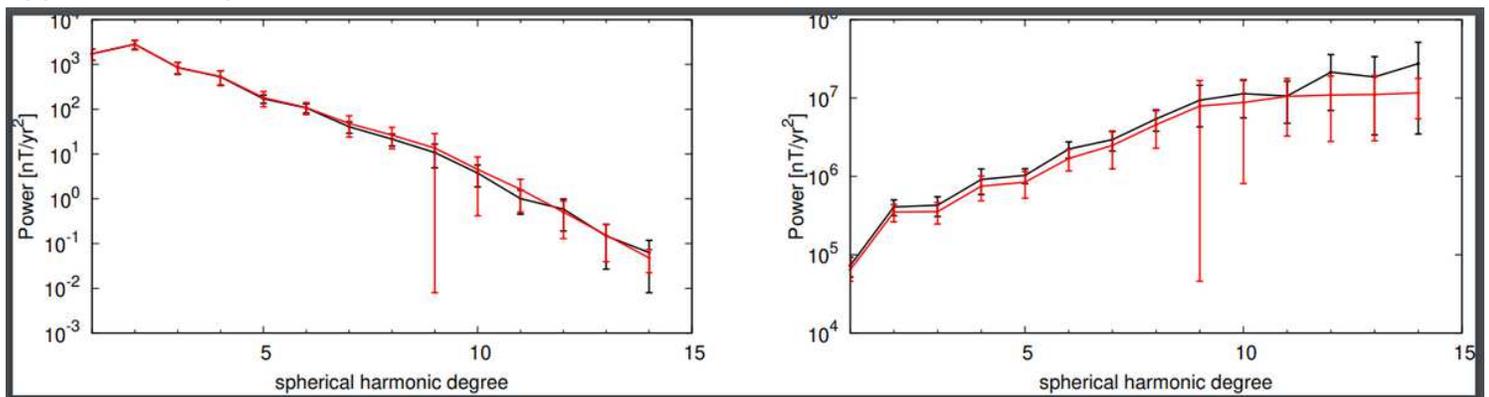
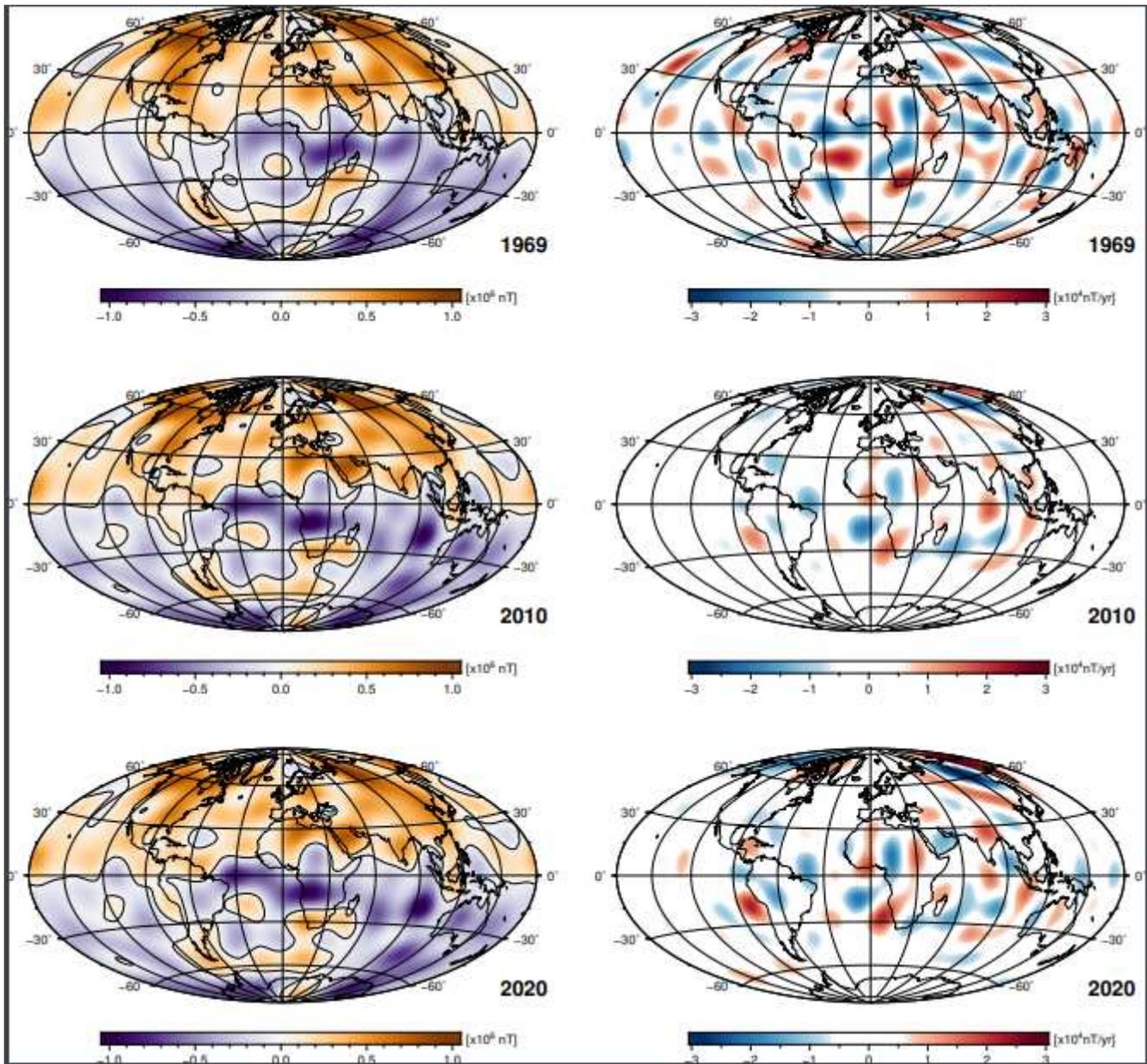


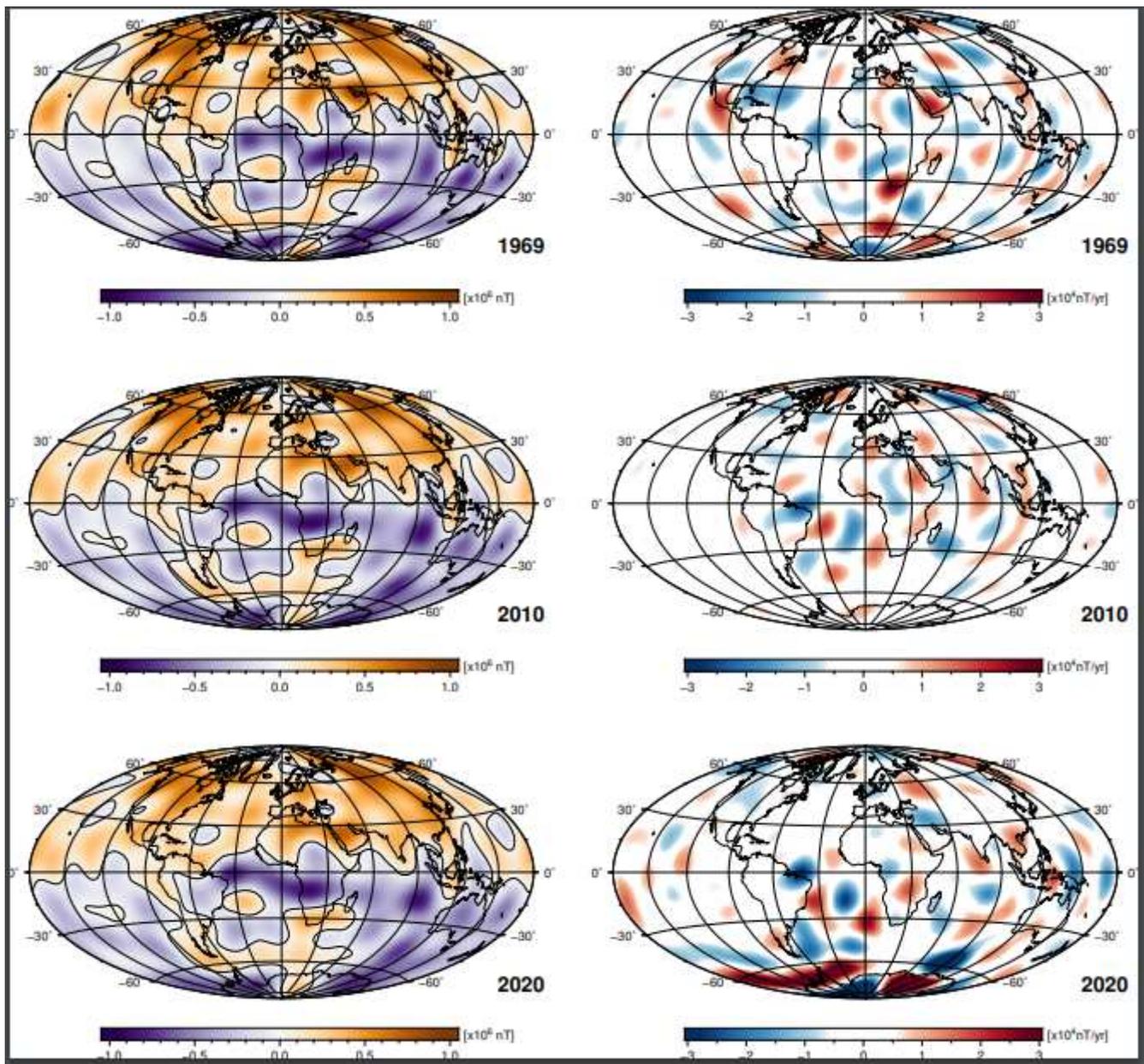
Figure 5

The temporally averaged secular variation spectra of Models 1 and 2 at Earth's surface (left) and at the core surface (right). Same line styles are applied as in Figure 3. The vertical bars indicate the range, i.e. the temporal variability of the individual spectra.



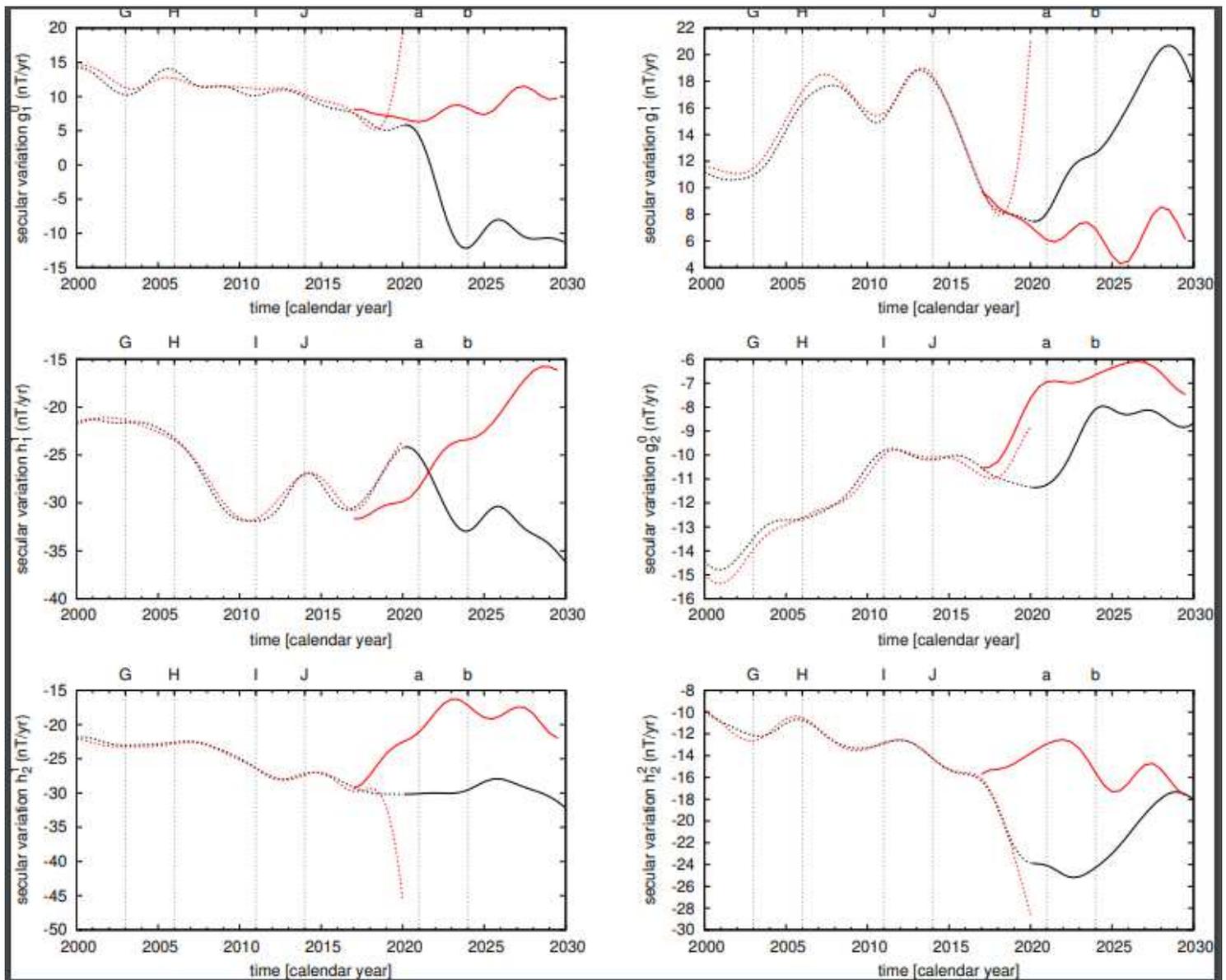
**Figure 6**

Radial component of the magnetic field at the core surface (left) and its secular variation (right) derived from Model 1 for epochs 1969, 2010 and 2020.



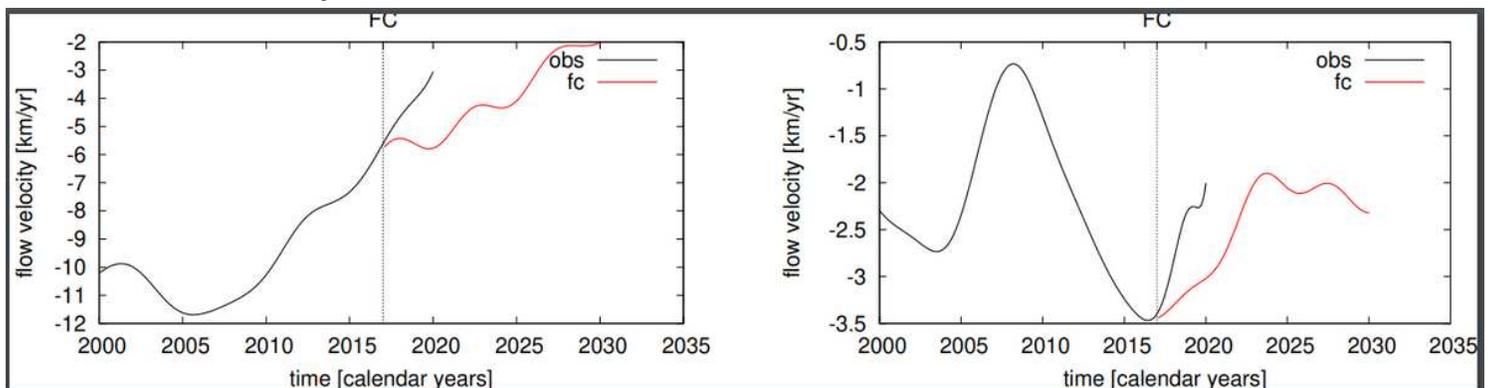
**Figure 7**

As in Figure 6 for Model 2.



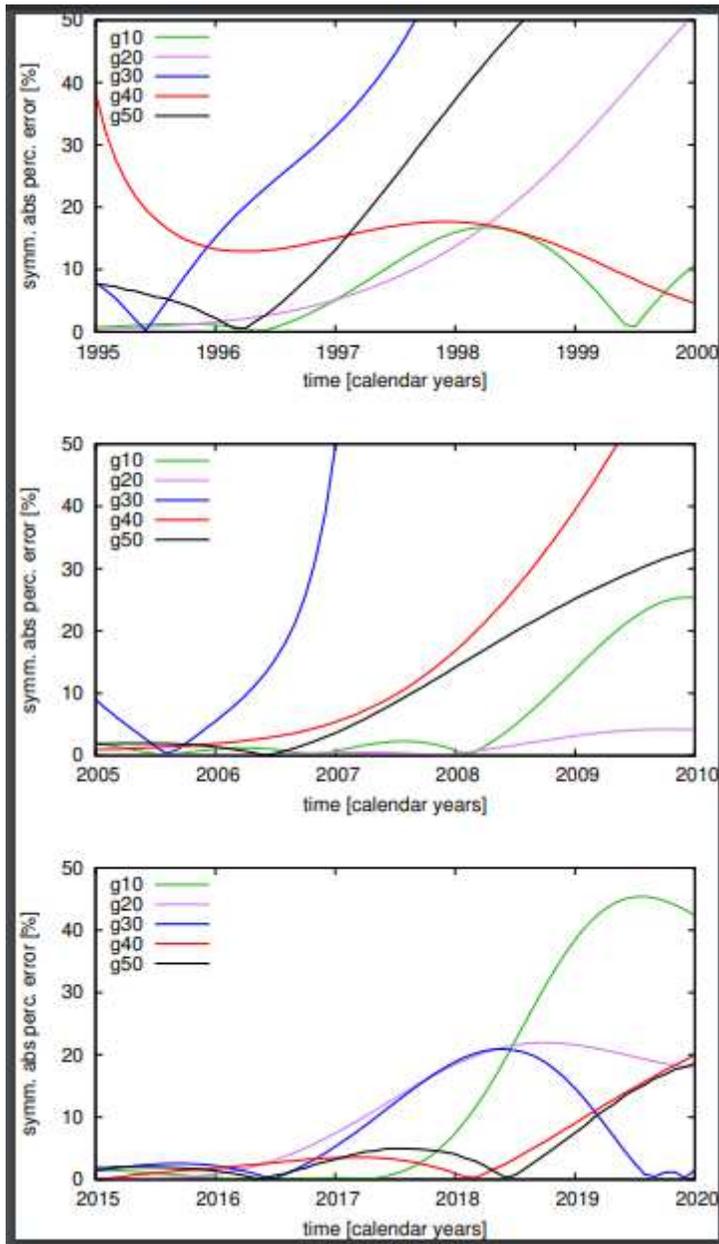
**Figure 8**

Comparison of six observed and predicted secular variation coefficients with the largest amplitude at Earth's surface. Model 1 (dashed black line), Model 2 (dashed red line) and their forecasts (solid lines). Vertical lines labeled with a and b mark two possible geomagnetic jerks. The other labels mark occurrences of known jerks.



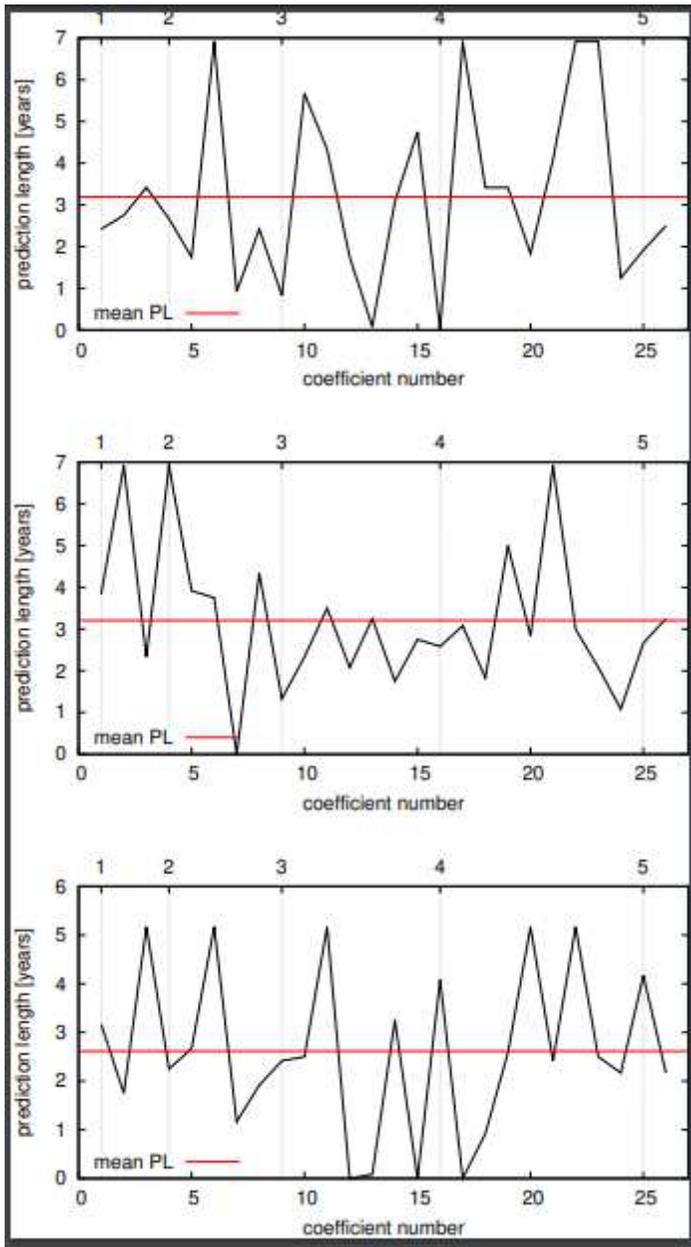
**Figure 9**

Evolution of the toroidal flow coefficients  $t_{0\_1}$  (left) and  $t_{0\_3}$  (right); their past variations (black) and their forecasts (red) based on the MSSA. The vertical line labeled with FC marks the start of the forecast.



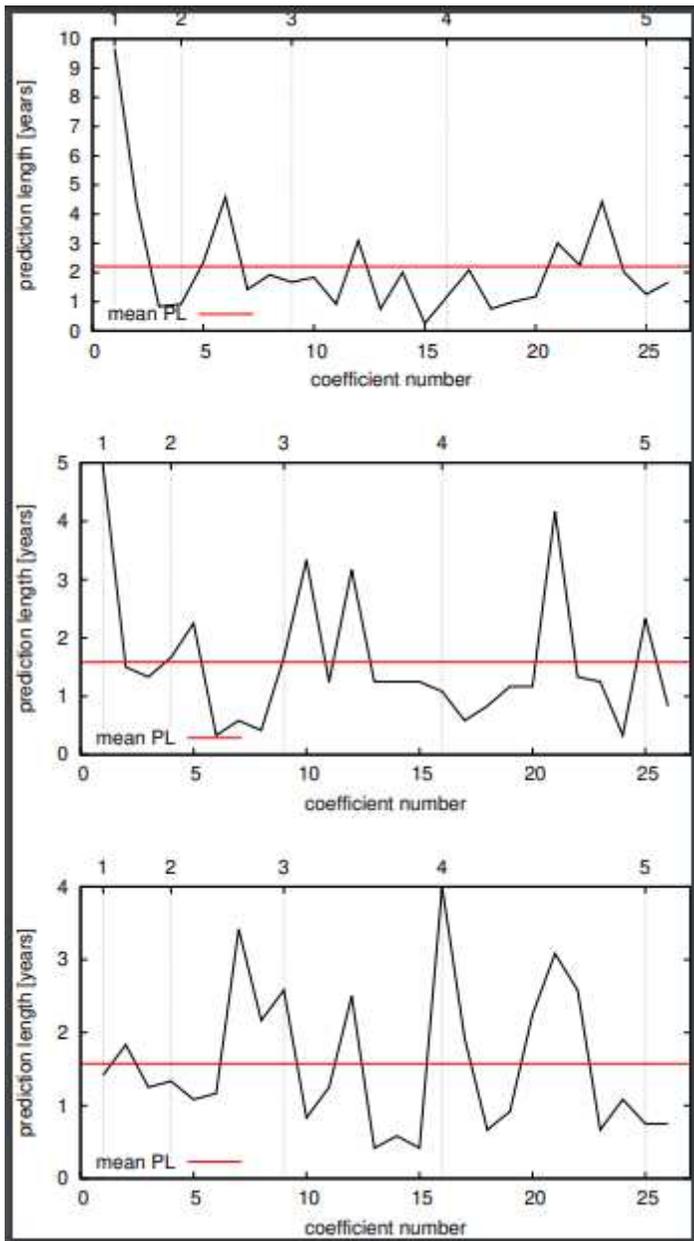
**Figure 10**

Evolution of the SAPE(t) of zonal secular variation coefficients based on Model 1 of the three forecasts experiments started in 1995 (top panel), in 2005 (middle panel) and in 2015 (bottom panel), respectively.



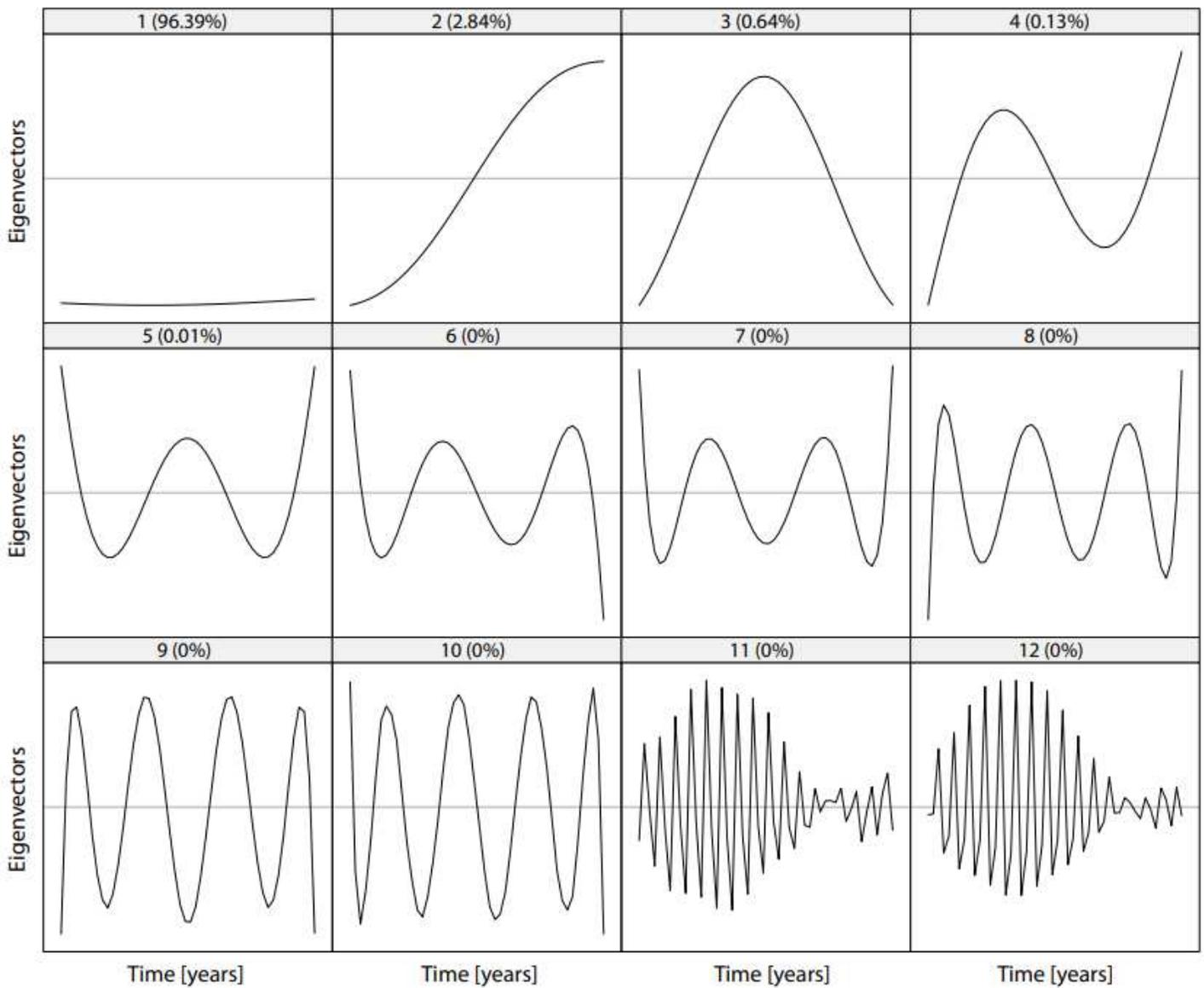
**Figure 11**

Prediction lengths defined by the time interval when  $sAPE(t) > 10\%$  of different forecast experiments based on Model 1 with forecasts started in 1995 (top panel), in 2005 (middle panel) and in 2015 (bottom panel), respectively. The solid horizontal red lines represent mean prediction lengths in years derived from the first 26 secular variation coefficients, i.e.  $g_0 \dots g_1$ . Top axes indicate degrees of the spherical harmonic expansion.



**Figure 12**

Prediction lengths for the first 26 toroidal flow coefficients estimated from different forecast experiments with forecasts started in 1995 (top panel), in 2005 (middle panel) and in 2015 (bottom panel), respectively. The solid horizontal red lines represent mean prediction lengths in years derived from the first 26 toroidal flow coefficients. Top axes indicate degrees of the spherical harmonic expansion.



**Figure 13**

The first 12 eigenvectors of the toroidal flow temporal variability. The x-axes of each subplot represent the model interval (1957-2020). The percentages of the partial variance of each eigenvector are given in the top bar of the individual plots.

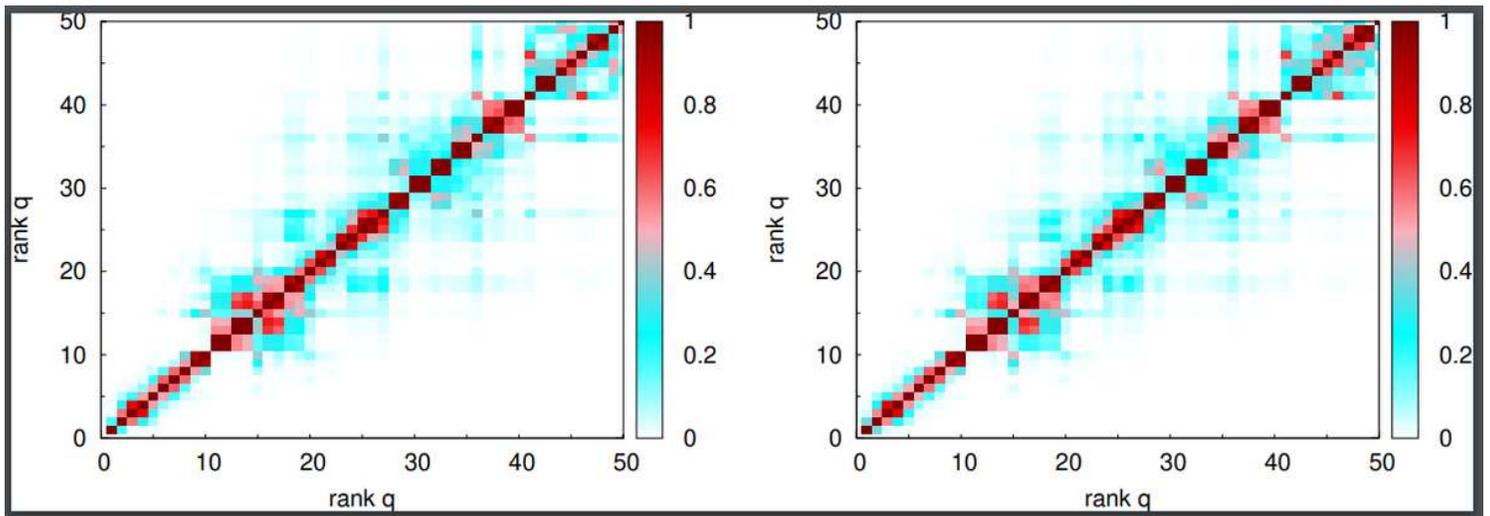


Figure 14

The weighted correlation matrix. Red colors represent stronger correlation between eigenvectors and therefore a weaker separation. The color scale corresponds to absolute values of correlations from 0 (white) to 1 (dark red).

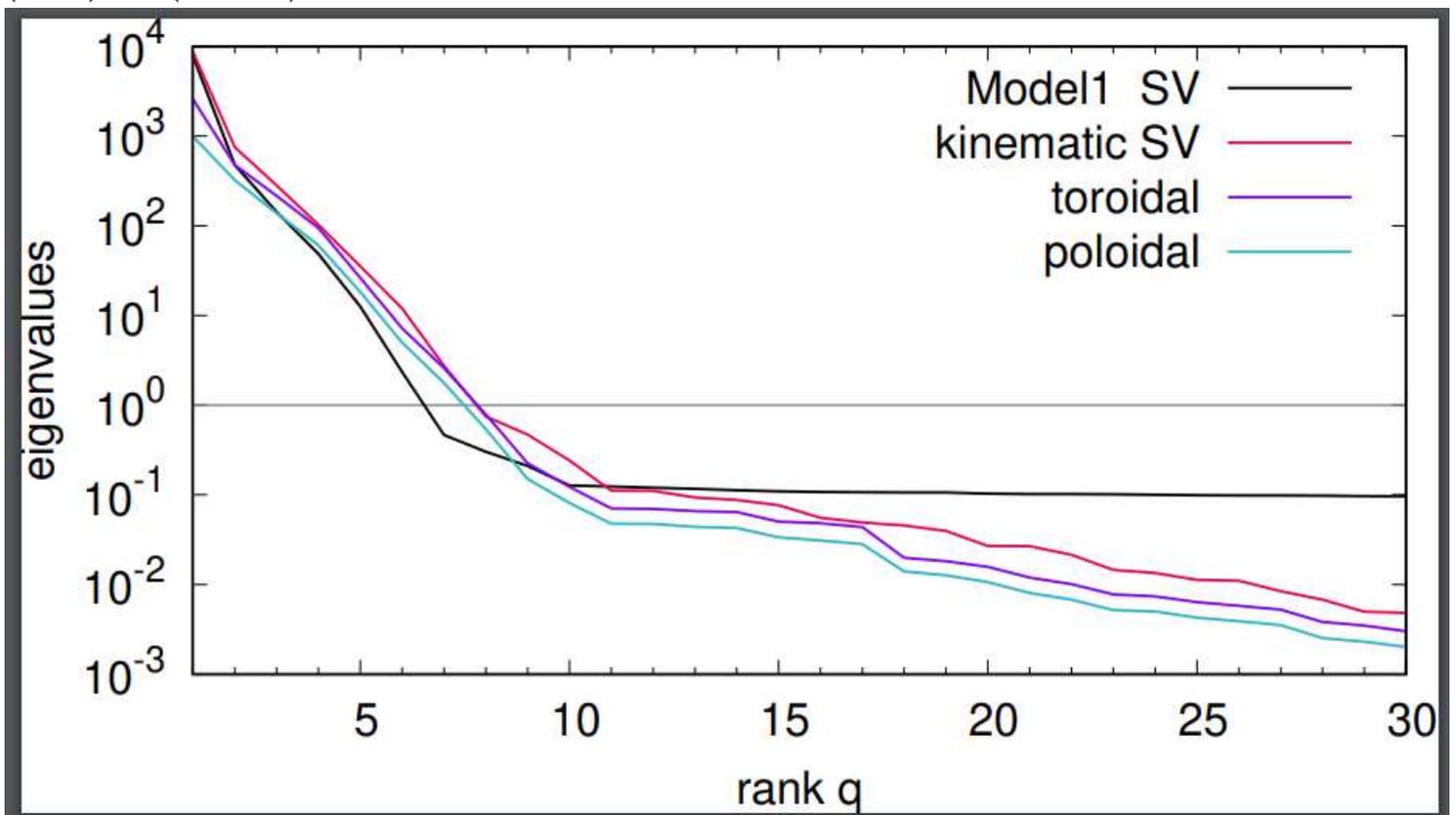


Figure 15

The eigenvalue spectra of the classical, kinematic secular variation, toroidal and poloidal flow variability, respectively. The horizontal line marks eigenvalues equal 1.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [graphicalabs.jpg](#)