Variability modes in core flows inverted from geomagnetic field models

M. A. Pais,[1,2] A. L. Morozova[1] and N. Schaeffer[3,4]

[1]*CGUC, Geophysical and Astronomical Observatory, University of Coimbra, Portugal. E-mail: pais@fis.uc.pt*
[2]*Department of Physics, University of Coimbra, P-3004-516 Coimbra, Portugal*
[3]*Univ. Grenoble Alpes, ISTerre, F-38041 Grenoble, France*
[4]*CNRS, ISTerre, F-38041 Grenoble, France*

SUMMARY

The flow of liquid metal inside the Earth's core produces the geomagnetic field and its time variations. Understanding the variability of those deep currents is crucial to improve the forecast of geomagnetic field variations and may provide relevant information on the core dynamics. The main goal of this study is to extract and characterize the leading variability modes of core flows over centennial periods, and to assess their statistical robustness. To this end, we use flows that we invert from two geomagnetic field models ('gufm1' and 'COV-OBS'), and apply principal component analysis and singular value decomposition of coupled fields. The quasi-geostrophic (QG) flows inverted from both geomagnetic field models show similar features. However, 'COV-OBS' flows have a less energetic mean and larger time variability. The statistical significance of flow components is tested from analyses performed on subareas of the whole domain. Bootstrapping methods are also used to extract significant flow features required by both 'gufm1' and 'COV-OBS'.

Three main empirical circulation modes emerge, simultaneously constrained by both geomagnetic field models and expected to be robust against the particular *a priori* used to build them (large-scale QG dynamics). Mode 1 exhibits three large vortices at medium/high latitudes, with opposite circulation under the Atlantic and the Pacific hemispheres. Mode 2 interestingly accounts for most of the variations of the Earth's core angular momentum. In this mode, the regions close to the tangent cylinder and to the equator are correlated, and oscillate with a period between 80 and 90 yr. Each of these two modes is energetic enough to alter the mean flow, sometimes reinforcing the eccentric gyre, and other times breaking it up into smaller circulations. The three main circulation modes added to the mean flow account for about 70 per cent of the flows variability, 90 per cent of the rms total velocities, and 95 per cent of the secular variation induced by the total flows.

Direct physical interpretation of the computed modes is not straightforward. Nonetheless, similarities found between the two first modes and time/spatial features identified in different studies of core dynamics, suggest that our approach can help to pinpoint the relevant physical processes inside the core on centennial timescales.

Key words: Inverse theory; Earth rotation variations; Dynamo: theories and simulations; Rapid time variations.

1 INTRODUCTION

Time variation of the main geomagnetic field on timescales smaller than about a couple of centuries (secular variation, SV) has been scrutinized for deeper understanding of the Earth's core dynamics. New insights have been gained both from numerical simulation of the Navier–Stokes, energy and induction equations inside the core (see e.g. Jones 2011) and inversion of magnetic field models fitting observatory and satellite observations (e.g. Holme 2007). Although the parameter regime characterizing the Earth's short timescale dy-

namics is not yet attainable in 3-D numerical simulations, different schemes have been used to extrapolate results. These include derivation of scaling laws and/or the use of simplifying approximations that allow the use of more Earth-like parameters.

The quasi-geostrophic (QG) approximation is such an example, since it reduces the system dimensionality from three to two (3-D to 2-D) making it possible to investigate a range of Ekman (ratio of viscous to Coriolis forces) and Lundquist (ratio of Alfvén wave to diffusion timescales) numbers that approach the Earth's core conditions. Convective QG rolls are known since the work

by Roberts (1968) and Busse (1970) and recent studies show that for timescales characteristic of SV and large-scale flows, elongated structures along the z-rotation axis will not be destroyed by Lorentz (magnetic) forces (Jault 2008; Gillet *et al.* 2011). One must however keep in mind the limitations of the QG approximation, mainly that it does not allow for antisymmetric flows (the ones that cross the equator), which could in reality be produced by any non-symmetric forcing.

From the point of view of flows computed from inversion of geomagnetic field models, the QG approximation is one example of constraint used to mitigate the intrinsic non-uniqueness characterizing the inverse problem. Other constraints relying on different dynamical assumptions or empirical reasons have also been used (e.g. purely toroidal flow, steady flow, tangential geostrophy, steady flow in a drifting reference frame, helical flow and tangential magnetostrophy). An account on these previously used constraints can be found in Holme (2007) and Finlay *et al.* (2010). In the following, we will restrict to QG flows due to their importance in SV dynamics (Gillet *et al.* 2011). In addition, although the flow is inverted on the core surface, it has a unique prolongation inside the core. QG flows inverted from different geomagnetic field models have been previously computed that explain the observed SV (Pais & Jault 2008; Gillet *et al.* 2009). In these studies, a large-scale spatial structure was identified and described as a large eccentric anticyclonic gyre. It interestingly gathers into one single structure, the well-known westward drift under the low-latitude Atlantic region, and the high latitude jet under the Bering Sea and close to the tangent cylinder (TC, cylinder coaxial with the Earth's rotation axis and tangent to the inner core surface) (see e.g. fig. 3 in Holme 2007). This structure seems to be more prominent in inversions derived from recent satellite geomagnetic field models, suggesting some time variability on relatively short timescales characteristic of SV. Other structures are also retrieved, a large-scale cyclonic vortex beneath the Pacific Hemisphere (PH) and smaller scale vortices dominantly cyclonic under the PH and anticyclonic under the Atlantic Hemisphere (AH). Gillet *et al.* (2009) pointed out that the smaller scale vortices may not in fact be well resolved, since they strongly depend on the induction effects that involve the small-scale main field (above harmonic degree 13–14), which are difficult to infer reliably from surface and satellite observations. In this respect, Finlay *et al.* (2012) stress the fact that the spectral slope above degree 12 depends significantly on the modelling regularizations in space. But there has not yet been a dedicated study to establish the significance of the flow structures. In this paper, we study the time variability of the inverted flow, and we assess their significance with statistical tools largely used in climate studies and oceanography.

Recently, a new geomagnetic field model has been computed, 'COV-OBS', which covers the period of historical geomagnetic field data provided by the international network of observatories (1840–1990), but also includes the recent period of satellite missions (1990–2010) with a dense coverage of high-quality data (Gillet *et al.* 2013). It avoids employing the regularized least-squares inversion approach used to compute the 'gufm1' model (Jackson *et al.* 2000), where roughness in the spatial and in the temporal domains is minimized when fitting the data to a truncated spherical harmonic (SH) expansion and using low-order splines in the time domain. Claiming that such regularizations are mainly *ad hoc* and do completely constrain the model above some harmonic degree (smaller at older epochs, larger at recent ones), Gillet *et al.* (2013) preferred to use *a priori* information on the statistical properties of the geomagnetic field, which they consider to be relatively well known, and follow a stochastic modelling approach (although their approach is

not completely free from some imposed temporal constraints, since the 'COV-OBS' model is expanded on a spline basis with a 2-yr knot spacing). The Gauss coefficients of the geomagnetic field are treated as resulting from a stationary process, their statistical properties being condensed into a large *prior* covariance matrix with variances and time correlations for all coefficients. Although the two models 'gufm1' and 'COV-OBS' fit the shared data at the same level for the common period (1840–1990), the different intrinsic characteristics are expected to distinguish flows inverted from one and the other.

In this study, we use QG flow models inverted from both 'gufm1' and from 'COV-OBS'. The computation of these flows, which is outlined in Section 2, is explained and discussed in Appendix A. Our main purpose is to test the possibility to describe the global circulation of the core flow in terms of a small number of patterns of variability, much like the climate variability can be described as a combination of patterns (e.g. Barry & Carleton 2001) such as the ENSO (El Niño–Southern Oscillation), PNA (Pacific–North America teleconnection pattern), NAO (North Atlantic Oscillation). We apply principal component analysis (PCA) tools, of standard use in the geophysical sciences of meteorology and oceanography, where we find the main references (e.g Preisendorfer 1988; Von Storch & Navarra 1995). These are non-parametric methods that do not assume any particular statistical distribution for the analysed data or any particular dynamical equation for the flow. In geomagnetic studies, PCA has been previously applied to space-time data grids in studies of geoelectromagnetic induction (Fujii & Schultz 2002; Balasis & Egbert 2006), and using ionospheric data (Matsuo *et al.* 2002; Zhang *et al.* 2013). The PCA approach has several limitations, as explained by Richman (1986). These may be due to domain shape dependence through the space orthogonality condition imposed, to sensitivity to the extent of spatial and time domains considered and to incomplete separation of consecutive modes. Also, there may be no correspondence between PCA-computed and physical modes. Nevertheless, PCA variability modes generally allow for a complexity reduction and, in certain cases, can provide information on the underlying dynamics (e.g North 1984). They have accordingly been extensively used in data analysis. Besides PCA, singular value decomposition (SVD) of coupled fields is also applied in this study to identify pairs of coupled spatial patterns explaining the covariance between 'gufm1' and 'COV-OBS' (see e.g. Bjornsson & Venegas 1997). At the same time, we propose to carry out significance tests that can decide which are the modes carrying relevant information on the system variability. The obtained results receive statistical support from analyses performed on subareas of the whole domain. Bootstrapping resampling methods are used to assess the statistical robustness of particular flow features. In the end, a physical interpretation of the most important/significant modes is sought.

The flow data to be analysed is described in Section 2. The basis of the PCA method and the more technical aspects that are used in this study are introduced in Section 3. Results are discussed in Section 4. They include results obtained when analysing flows inverted from 'gufm1' and from 'COV-OBS' separately, and results obtained by extracting correlated information contained in both of them. Section 5 contains a discussion on the identified structures and their possible dynamical interpretation. Finally, the last section summarizes the main conclusions in this study.

2 INVERTED QG FLOW MODELS

Two core surface flow models were computed, using regularized inversion, from the time-varying geomagnetic field models for the

observatory era 'gufm1' (1840–1990; Jackson *et al.* 2000) and COV-OBS (1840–2010; Gillet *et al.* 2013). These two geomagnetic field models employ cubic B-splines to model the time variation, with 2.5-yr and 2-yr knot spacing, respectively (but data are supplied twice the knot spacing). Our flow inversions generate a flow snapshot for each year in the geomagnetic field model period, and the whole set of flows is denoted flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$, respectively. The whole inversion procedure is explained in detail in Appendix A.

In the inversion, the flows were assumed to be columnar in the whole volume and incompressible. This implies that they can be completely retrieved from a single scalar pseudo-streamfunction ξ, symmetric with respect to the equatorial plane, which in turn can be inverted from geomagnetic field models. At the core surface, the relation between the flow **u** and ξ at a certain epoch is (Schaeffer & Cardin 2005; Pais & Jault 2008; Amit & Pais 2013):

$$\mathbf{u} = \frac{1}{\cos\theta}\nabla_H \wedge \xi(\theta,\phi)\hat{r} + \frac{\sin\theta}{R_c\cos^2\theta}\xi^{\text{NZ}}(\theta,\phi)\hat{\phi}, \quad (1)$$

where (r, θ, ϕ) are spherical coordinates, R_c is the core radius and ξ^{NZ} is the non-zonal (ϕ-dependent) part of ξ. The first term in (1) has exactly the same expression as tangential geostrophic (TG) flows computed at the CMB (e.g. Chulliat & Hulot 2000), where ξ would represent the geostrophic pressure to within a constant factor, $p_{\text{geo}} = -2\rho\Omega\xi$ (with ρ the core fluid density and Ω the Earth's rotation rate). For columnar flows touching the CMB, if they conserve mass they no longer follow isocontours of ξ because of the second term in (1). This term does not contribute to the zonal flow, is everywhere latitudinal and dependent on colatitude. It is required for fluid mass conservation, leading to equal contributions to upwellings and downwelling from the two terms (see e.g. Amit & Pais 2013). As also discussed in Amit & Pais (2013), the ξ field is uniquely determined from inversion of geomagnetic field models. Close to the equator, $\xi^{\text{NZ}}(\theta,\phi) \sim \cos^2\theta$ in order for the flow to remain finite. Canet *et al.* (2014) use other pseudo-streamfunction $\Psi(\theta,\phi)$ to study hydromagnetic QG modes in planetary interiors, the two scalar functions being related through $\Psi^{\text{NZ}} = -R_c\cos\theta\,\xi^{\text{NZ}}$ and $\partial_\theta\Psi^Z = -R_c\cos\theta\,\partial_\theta\xi^Z$ (where the Z superscript denotes the zonal component).

The assumption of equatorially symmetric flow due to columnar convection does not hold inside the TC. Different regimes can be envisioned there, depending on the dominant force balance believed to exist in that region. In particular, this can lead either to independent columnar convection in the North and South hemispheres (Pais & Jault 2008) or thermal wind convection (e.g Aurnou *et al.* 2003) that can change the sign of circulation close to the inner core and the outer core boundaries.

Because of this uncertainty, and for convenience, we use the same assumptions inside and outside the TC for computing **u**. Hence, to guarantee that flows computed for this small region (that usually tend to have high rms velocity u_{rms}) will not affect the PCA modes, the 20° caps around the North and South CMB poles were excluded from the analysis as explained in the following section.

3 PCA OF QG FLOWS

We follow the implementation of the method as explained in Bjornsson & Venegas (1997) and Hannachi *et al.* (2007). The field ξ, closely related to a geostrophic pressure especially at high latitudes, can be used to characterize the spatial flow features and its time variability over some time period, with a similar role to sea level pressure in studies of variability of atmospheric currents (e.g. Hannachi *et al.* 2007). In short, we will analyse the ξ field into a sum of

orthogonal functions in space, multiplied by coefficients which are uncorrelated functions of time, by applying a linear transformation to the data set that concentrates as much of the variance as possible into a small number of terms in the expansion.

3.1 Empirical orthogonal functions (EOFs) and principal components (PCs)

The spatio-temporal information retrieved from inverted flow models is gathered into matrices:

$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1N_p} \\ X_{21} & X_{22} & \cdots & X_{2N_p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{N_e1} & X_{N_e2} & \cdots & X_{N_eN_p} \end{bmatrix} \quad (2)$$

with

$$X_{ij} \equiv X(t_i, \vec{r}_j) = \sqrt{\sin\theta}\left(\xi(t_i,\theta_j,\phi_j) - \frac{1}{N_e}\sum_{i=1}^{N_e}\xi(t_i,\theta_j,\phi_j)\right). \quad (3)$$

In a few words, X_{ij} is the (scalar) value of the ξ pseudo-streamfunction at the core surface point $\vec{r}_j = \vec{r}(R_c,\theta_j,\phi_j)$ at time t_i, once the time-average has been subtracted. It is also weighted by a latitude-dependent factor $\sqrt{\sin\theta}$ to compensate for the fact that in a regular grid on a spherical surface the number of points to cover a given area increases with latitude. The index i takes values from 1 to the total number of epochs, N_e, and the index j takes values from 1 to the total number of gridpoints, N_p. For flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$, a set of $5° \times 5°$ latitude/longitude grids of the scalar function ξ were computed, one for each year, covering the CMB region under study defined by $20° < \theta < 160°$ for all values of longitude. We exclude the polar caps because, as referred, the computed QG flows are not reliable inside the TC.

The EOF/PCA relies on the description of the space-time data in terms of decorrelated modes. They are the eigenvectors of the covariance (or variance–covariance) matrix $\mathbf{C}_X = \mathbf{X}^{\text{T}}\mathbf{X}$: each element $(\mathbf{C}_X)_{ij} = X_{li}X_{lj} = \sum_{l=1}^{N_e}X(t_l,\vec{r}_i)X(t_l,\vec{r}_j)$ is the discrete form of the (temporal) covariance between the data values at points $\vec{r}_i$ and $\vec{r}_j$. In search for decorrelated modes, we need to transform $\mathbf{C}_X$ in a diagonal matrix, for which the standard procedure is to transform the data matrix $\mathbf{X}$ into

$$\mathbf{Y} = \mathbf{XP}, \quad (4)$$

where $\mathbf{P}$ is the orthogonal matrix of eigenvectors of $\mathbf{C}_X$, each eigenvector $\mathbf{p}_i$ forming a column. The covariance matrix of $\mathbf{Y}$ is then

$$\mathbf{C}_Y = \mathbf{P}^{\text{T}}\mathbf{C}_X\mathbf{P} = \Lambda, \quad (5)$$

and is diagonal with non-zero elements λ_i, ordered from the highest to the lowest.

The transformed data-matrix $\mathbf{Y}$ is now expressed in terms of decorrelated modes, as was intended. From (4) the reconstructed data matrix yields

$$\mathbf{X} = \mathbf{YP}^{\text{T}}, \quad (6)$$

which can be written as

$$\mathbf{X} = \mathbf{y}_1\mathbf{p}_1^{\text{T}} + \mathbf{y}_2\mathbf{p}_2^{\text{T}} + \cdots + \mathbf{y}_{N_p}\mathbf{p}_{N_p}^{\text{T}} = \sum_{k=1}^{N_p}\mathbf{y}_k\mathbf{p}_k^{\text{T}}, \quad (7)$$

that is, as a linear combination of all eigenvectors of $\mathbf{C}_X$, the coefficients of the expansion being the different columns of matrix $\mathbf{Y}$.

Here, and according to Bjornsson & Venegas (1997), the following notation is used:

(i) each column $\mathbf{p}_k$ of $\mathbf{P}$ is an EOF and represents a certain spatial function;

(ii) each column $\mathbf{y}_k$ of $\mathbf{Y}$ is a PC and represents a time-series.

This decomposition presents the further advantage of data reduction, since there is no practical need to keep the whole set of N_p EOFs (the same number as gridpoints) in the analysis. The different eigenvalues λ_i of $\mathbf{C}_X$, determine which modes are to be kept in the expansion and in practice only the first few ones are important, those that explain the highest percentage of data variance as given by

$$f_i = \lambda_i \left(\sum_{k=1}^{N_p} \lambda_k \right)^{-1}. \tag{8}$$

In terms of describing the data covariance matrix, upon which the whole framework is based, the EOFs provide a simplified and (hopefully) compressed way to decompose this matrix. From (5) it follows

$$\mathbf{C}_X = \sum_{k=1}^{N_p} \lambda_k \mathbf{p}_k \mathbf{p}_k^{\mathrm{T}}, \tag{9}$$

with only the first modes required.

We note that some studies apply PCA using correlation instead of covariance matrices, by normalizing each time-series to unit standard deviation (see e.g. Bretherton *et al.* 1992). Here, such a choice could enhance small, non-meaningful signal and was not adopted.

3.2 SVD of coupled fields

SVD of coupled fields has been widely used in meteorology and oceanography, to study two combined data fields of different physical quantities such as for example, sea level pressure (SLP) and sea surface temperature (SST). In this study, it is used as a framework to isolate important coupled modes of variability between the two time-series of core flows, flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$, representing the same physical quantity but derived from geomagnetic field models computed using different *a priori* on the data and the model itself. We argue that by applying SVD to the correlation between the two flows, the most robust features can be recovered, those that do not depend specifically on the *a priori* used in one or the other geomagnetic field models.

We follow the implementation of the method as explained in Bretherton *et al.* (1992), Bjornsson & Venegas (1997) and Hannachi *et al.* (2007). Starting from two data matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ for flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$, respectively, built as explained, the temporal cross-variance matrix is constructed, $\mathbf{C}_{X_1 X_2} = \mathbf{X}_1^{\mathrm{T}} \mathbf{X}_2$. As will be made clear in the following, it makes no difference for the spatial patterns, the temporal functions or the percentage of cross-covariance explained, that this covariance matrix or its transpose, $\mathbf{C}_{X_2 X_1}$, be considered. The SVD general matrix operation is applied to $\mathbf{C}_{X_1 X_2}$:

$$\mathbf{C}_{X_1 X_2} = \mathbf{U} \Lambda \mathbf{V}^{\mathrm{T}}, \tag{10}$$

where $\mathbf{U}$ and $\mathbf{V}$ are both orthogonal matrices, the former having in columns the eigenvectors of $\mathbf{C}_{X_1 X_2} \mathbf{C}_{X_1 X_2}^{\mathrm{T}}$ and the later having the eigenvectors of $\mathbf{C}_{X_1 X_2}^{\mathrm{T}} \mathbf{C}_{X_1 X_2}$. As of Λ, it is a diagonal matrix with singular values λ_i in its diagonal, ordered from the highest to the lowest. These λ_i are the (positive) square roots of the eigenvalues of $\mathbf{C}_{X_1 X_2}^{\mathrm{T}} \mathbf{C}_{X_1 X_2}$, the same as for $\mathbf{C}_{X_1 X_2} \mathbf{C}_{X_1 X_2}^{\mathrm{T}}$.

Linear transformation of data matrices $\mathbf{X}_1$ and $\mathbf{X}_2$ with the orthogonal matrices $\mathbf{U}$ and $\mathbf{V}$, respectively, makes the new cross-correlation matrix diagonal, $\mathbf{C}_{Y_1 Y_2} = \Lambda$, where $\mathbf{Y}_1 = \mathbf{X}_1 \mathbf{U}$ and $\mathbf{Y}_2 = \mathbf{X}_2 \mathbf{V}$. The original data matrices are reconstructed as

$$\mathbf{X}_1 = \mathbf{Y}_1 \mathbf{U}^{\mathrm{T}}$$
$$\mathbf{X}_2 = \mathbf{Y}_2 \mathbf{V}^{\mathrm{T}}, \tag{11}$$

analogously to (6). It is nonetheless worth noting that $\mathbf{U}$ and $\mathbf{V}$ integrate information on both flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$, contrary to the transformation matrix $\mathbf{P}$ in Section 3.1. As a result, spatial patterns obtained from $\mathbf{U}$ for instance are different from EOFs obtained by applying PCA to $\mathbf{X}_1$. Also, the expansion coefficients time-series in (11), namely $\mathbf{Y}_1$ and $\mathbf{Y}_2$, are different from the PCs referred in (3.1).

In analogy with (9), the column vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ of the transformation matrices $\mathbf{U}$ and $\mathbf{V}$, respectively, can be used to decompose the matrix $\mathbf{C}_{X_1 X_2}$:

$$\mathbf{C}_{X_1 X_2} = \sum_{k=1}^{N_p} \lambda_k \mathbf{u}_k \mathbf{v}_k^{\mathrm{T}}. \tag{12}$$

This expression resorts the fact that each pair k of coupled modes describes a fraction of the cross-covariance between the two fields, the fraction being given by f_k (see 8).

3.3 Homogeneous and heterogeneous correlation maps

Besides direct representation of PCA or SVD modes in the form of maps of the EOFs and time plots of corresponding PCs, correlation maps are a means to highlight in the analysed spatial region, those regions that are evolving in time in a correlated manner. In each point of the space grid, a correlation coefficient is computed between a PC time function and the 'observed' time-evolving streamfunction in that same point. For homogeneous correlation maps, the concept is very similar to the one leading to different EOFs and correlating the different PCs with the original data gives rise to spatial structures that are the same as the corresponding EOFs. Nonetheless, it involves looking for strong correlation instead of strong covariance and, as such, it is normalized to local variance and the values in each gridpoint are between -1 and $+1$. In this way, regions of low amplitude variations that show up as faint regions in the charts of EOFs, will show up as strong as regions of high amplitude variations, if they belong to the same EOF structure. Heterogeneous correlation maps are meaningful in SVD analysis to show how PCs extracted from a certain flow model are correlated with the other flow model. If the identified SVD modes are to be given some physical meaning, it is required that a certain PC correlated with the two flow models gives similar structures.

From studies of climatic patterns analyses, we borrow standard tests used to assess the robustness of variability modes. The relevant quantities are introduced in Appendix B.

4 RESULTS OF STATISTICAL ANALYSES

The mean flows shown in Fig. 1 are computed by averaging the flow over the corresponding time period: 1840–1990 for flow$^{\text{gufm1}}$ and 1840–2010 for flow$^{\text{COV-OBS}}$. There are no significant differences if we consider the shorter 1840–1990 period for COV-OBS. The rms
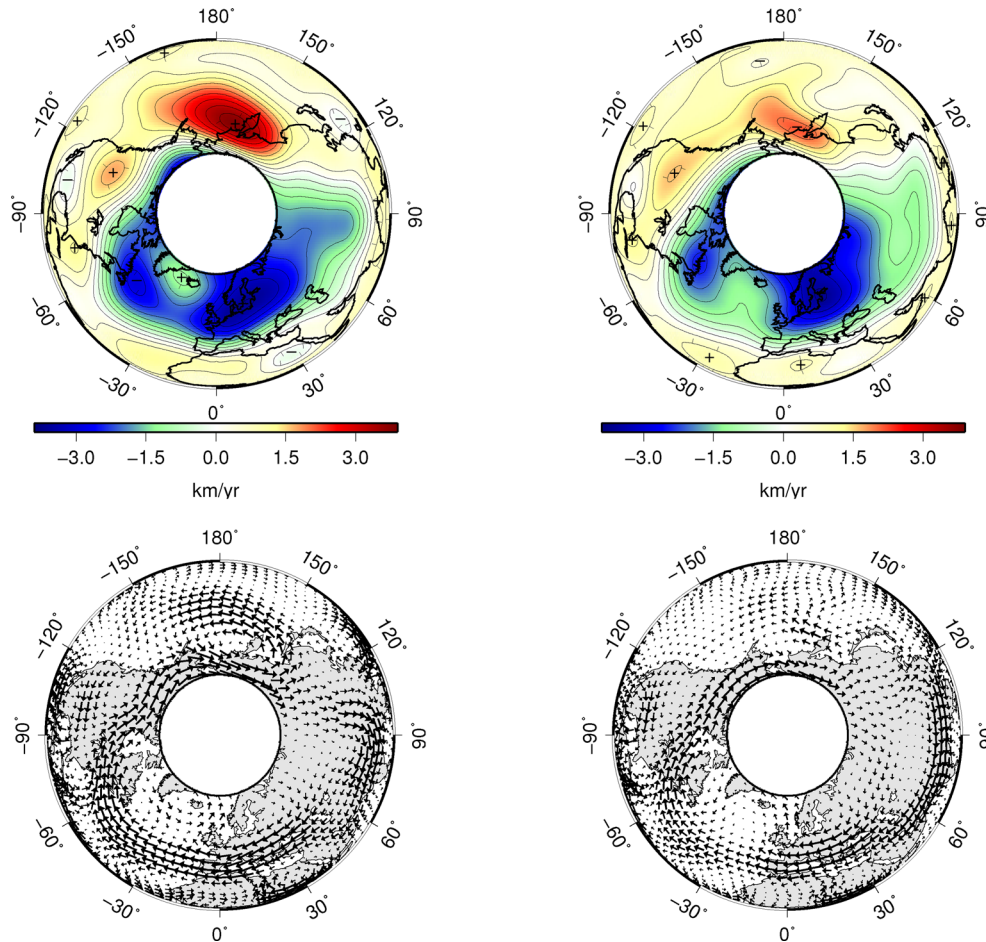
Figure 1. The mean QG flow on the equatorial plane over the time period 1840–1990, as seen from the North pole. For 'gufm1' (left-hand side) and 'COV-OBS' (right-hand side). At the top, the pseudo-streamfunction ξ/R_c in units of km yr^{-1}; at the bottom, arrows visualize the flow.

velocity for the surface mean flow, $\overline{u}_{rms}$, is stronger for flow^{gufm1} (12 km yr^{-1}) than for flow$^{COV\text{-}OBS}$ (10 km yr^{-1}), basically due to a stronger anticyclone under the PH and a stronger radial jet under the Eastern Asian continent, from high to low latitudes. The two different kind of charts in Fig. 1 illustrate the fact that the flow very closely follows contours of the ξ streamfunction at medium-high latitudes. The rule to keep in mind when interpreting ξ-contour charts, from eq. (1), is that the fluid circulates anticlockwise around centres of positive ξ and clockwise around centres of negative ξ.

The EOF/PCA tools can sort out the time variability associated to the main structures in the mean flow, and other structures which average out during the inspected period. In the following, we will be showing results for the analysis carried out over the computed flows, using tools described in Section 3 and Appendix B.

4.1 PCA applied to gufm1 and COV-OBS, separately

The first five PCA modes account for about 80 per cent of the flow total variability (see Table 1). The EOF patterns of the first three modes are shown in Fig. 2. Results for COV-OBS considering the two time periods are similar. Note however that the first three modes represent a larger fraction of the signal for the shorter period. Besides, the fourth mode may not be completely separated from the third when considering the longer time period. Comparing COV-OBS and gufm1, the first mode explains very similar variances.

Using North's criterion (eq. B2) to detect mode degeneracy, modes 4 and 5 in COV-OBS (1840–1990) and modes 3 and 4 in COV-OBS (1840–2010) are not completely separated (see Fig. 3), meaning that they may describe two aspects of a common structure. A simple illustration of this effect is found in the example of a propagating wave that can be decomposed into two spatial patterns space-shifted by fourth of a wavelength, multiplied by two sinusoidal functions time-shifted by fourth of a period. One unique structure (a propagating wave) would then appear in the PC analysis decomposed into two modes with exactly the same f value. In the case of the above-mentioned degenerate modes, they should be considered together.

As a test for subdomain stability, the whole CMB domain was subdivided into two longitudinal hemispheres, and EOF/PC modes were recalculated for each of them. The chosen meridian for the separation goes through 70°E and the two resulting hemispheres have longitudes 70°E to 250°E (PH) and −110°E to 70°E (AH). Results are shown in Tables 2 and 3. The first five modes are still non-degenerate when computed independently for each hemisphere and according to North's criterion (eq. B2 and see Table 2). While the spatial and temporal descriptions of modes 4 and 5 can be different depending on if a global or a hemispherical grid of data values is used, the first three modes are recovered under AH with very close characteristics as in the global grid. The first three variability modes are not so well recovered using only data under the PH, especially for mode 2 (see Table 3).

Table 1. Standard PCA applied to flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ separately. Percentage of variance explained by each mode ($100f_i$, using eq. 8). In parenthesis, the δf_i error due to $\delta\lambda_i$ as given by eq. (B1).

	Fraction of variance, f_i (per cent) (δf_i)				
	Mode-1	Mode-2	Mode-3	Mode-4	Mode-5
COV-OBS (1840–2010)	34.1 (4.1)	16.6 (2.0)	11.0 (1.3)	10.3 (1.2)	7.5 (0.9)
COV-OBS (1840–1990)	34.4 (4.5)	18.3 (2.4)	12.7 (1.7)	8.6 (1.1)	7.7 (1.0)
gufm1 (1840–1990)	33.3 (4.3)	22.9 (3.0)	16.2 (2.1)	7.9 (1.0)	6.6 (0.9)

Overall, when applying PCA to gufm1 and COV-OBS separately, some significant differences come out: (i) mode 2 is more important to explain the variability of flow$^{\text{gufm1}}$ than of flow$^{\text{COV-OBS}}$, possibly due to a less clear expression of mode 2 under the PH; (ii) also, differences between recovered modes in PH and AH hemispheres are more important for flow$^{\text{COV-OBS}}$ than for flow$^{\text{gufm1}}$; (iii) most importantly, corresponding modes in flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ exhibit significant spatial and temporal differences in general, as shown by relatively low values of congruence and correlation coefficients in square brackets in Table 4 (see also Fig 2).

4.2 SVD of coupled gufm1 and COV-OBS

In order to identify structures correlated in both flows, we apply SVD of coupled fields to flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ for the common period 1840–1990 (see Table 4).

In Fig. 4 we represent the contours of streamfunction ξ for the three first coupled modes. As all modes show important spatial structures at medium-high latitudes, where visualization of the flow through ξ contours is quite straightforward, the representation adopted uses these contours projected onto the equatorial plane as seen from the North pole. Each pattern EOF$_i$ contributes to the global flow with the flow circulation retrieved directly from Fig. 4 in epochs of positive PC$_i$ and opposite circulation in epochs of negative PC$_i$. We easily recognize common main structures in the pair of EOFs for the first two modes, namely three large rolls at medium-high latitudes for mode 1, two anticyclonic under the AH and one cyclonic under the PH, and an important zonal component around the TC in mode 2. Mode 2 also presents a strong zonal component at the equator, which is not apparent in Fig. 4 because of the chosen projection (see however Fig. 9 and the Discussion in Section 5). Special care must be taken in interpreting the EOF$_2$ pattern in Fig. 4. Indeed, the streamfunction ξ has an important negative zonal component at the equatorial region that contributes to the flow through the first term in the RHS of eq. (1), but not through the second term where only non-zonal components of ξ appear. Then as $\nabla_H\xi$ points radially inwards, the corresponding flow circulation is eastwards (in Fig. 4), the same direction as around the TC. The relative sign of the ξ streamfunction of mode 2 near the TC and at the equator appears in Fig. 7 and the common direction of the zonal longitudinal flow appears in Fig. 9. The two coupled spatial functions of mode 3 differ more than for modes 1 and 2 (Table 4). In particular, EOF$_3$ for flow$^{\text{gufm1}}$ (Fig. 4, bottom left-hand side) is mainly localized under the Atlantic, a feature not seen in the EOF$_3$ that reconstructs flow$^{\text{COV-OBS}}$ (Fig. 4, bottom right-hand side). There are nonetheless some corresponding features in these EOF$_3$ charts: a cyclone under West of Iberia and another under East of the Caribbean Sea, an anticyclone under Eastern Canada and another under the Bering Sea.

As to the expansion coefficient time-series (PCs), note that for a given mode, the flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ PCs for the SVD analysis are always closer to each other than the PC's of separated EOF/PC analysis of flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ (see Fig. 5). In addition, for mode 1 (respectively, mode 2), the expansion coefficients obtained by SVD are closer to the one obtained by PCA of flow$^{\text{COV-OBS}}$ (respectively, flow$^{\text{gufm1}}$). This ability of SVD to bring the two flow descriptions closer together is also observed for spatial structures: the EOFs for the SVD analysis shown in Fig. 4 are closer together than the corresponding EOFs for PCA analysis of individual flows (see Fig. 2). This result is also apparent in Table 4, from comparison of values inside and outside square brackets for congruence (spatial similarity) and correlation (temporal resemblance) coefficients.

A quasi-periodicity is seen in the temporal variation of mode 2. Given the importance of this mode in terms of data variability explained and significance of its spatial features close to the TC and at the equator, we further characterize this temporal behaviour. We use all PC$_2$'s obtained from PCA applied to flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ separately, both in a global grid and hemispherical subdomain grids, together with PC$_2$'s obtained from SVD of coupled flows. Each of these nine time-series was fitted with a single sinusoid where amplitude, frequency and initial phase were free parameters. An average and a standard deviation of the estimated period T for the ensemble PC$_2$'s were produced. When using the whole 1840–1990 data set we found $T = 90.5 \pm 4.6$ yr and, when using the more recent data set for 1900–1990 for which errors in geomagnetic field models are smaller, we found $T = 80.1 \pm 2.7$ yr.

Correlating PC$_i$ of flow$^{\text{gufm1}}$ (resp. flow$^{\text{COV-OBS}}$) with the data matrix corresponding to flow$^{\text{gufm1}}$ (resp. flow$^{\text{COV-OBS}}$) yields an homogeneous correlation map, whereas correlating it with the data matrix corresponding to flow$^{\text{COV-OBS}}$ (resp. flow$^{\text{gufm1}}$) yields an heterogeneous correlation map (see Fig. 6). Homogeneous correlation maps provide a representation of EOF$_i$, while heterogeneous correlation maps are of special interest, since they reveal the correlated structures present in the two flows. Strong blue and red patches show regions of flow$^{\text{gufm1}}$ or flow$^{\text{COV-OBS}}$ where the flow evolution is strongly correlated with a certain PC$_i$ (red for positive correlation, blue for anticorrelation). Note that the black solid lines, which separate regions of positive and negative correlation with PC$_i$, are also the zero iso-ξ lines of EOF$_i$.

The correlation maps of Fig. 6 highlight a spiraling structure in mode 1, better seen in flow$^{\text{gufm1}}$ than in flow$^{\text{COV-OBS}}$ and that was not obvious in Fig. 4 because of disparate flow amplitudes. It is brought out by the normalization underlying the computation of correlation coefficients.

If we denote the mean flow column vector as $\bar{\mathbf{x}}$, with the same dimension as $\mathbf{u}_i$ and $\mathbf{v}_i$ vectors (the EOFs to retrieve flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$, respectively), then we may write the mean flows as a combination of orthogonal SVD modes $\bar{\mathbf{x}}^{\text{gufm1,T}} = \alpha^{\text{T}}\mathbf{U}^{\text{T}}$ and $\bar{\mathbf{x}}^{\text{COV-OBS,T}} = \beta^{\text{T}}\mathbf{V}^{\text{T}}$. The column vectors α and β contain the expansion coefficients of the mean flows in the orthogonal basis of EOFs. Note that, because the mean flows $\bar{\mathbf{x}}^{\text{gufm1}}$ and $\bar{\mathbf{x}}^{\text{COV-OBS}}$ have been removed prior to PCA or SVD analysis applied to matrices $\mathbf{X}$,
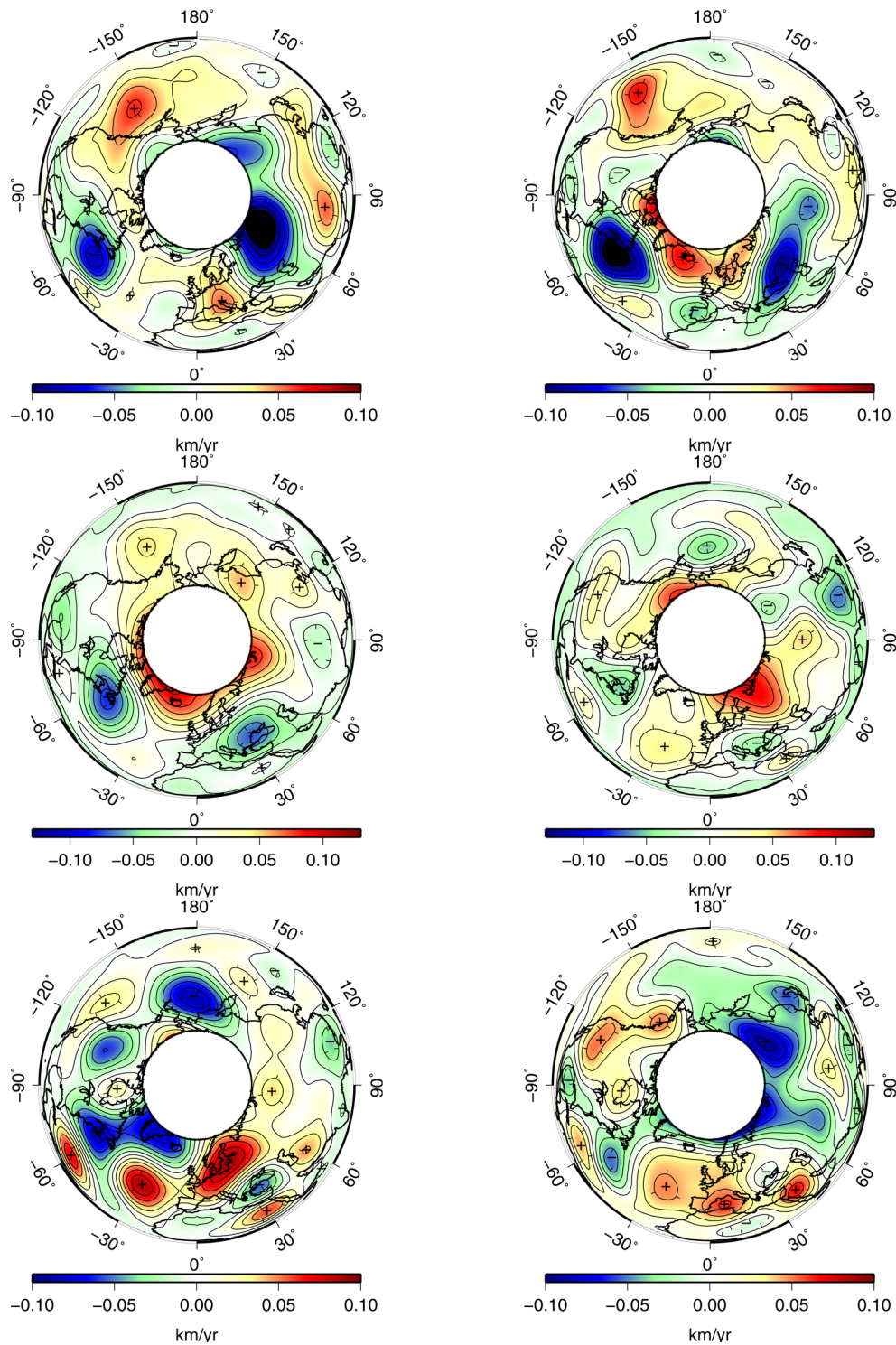
Figure 2. The three first EOFs resulting from PCA applied to flow[gufm1] (left-hand side) and flow[COV-OBS] (right-hand side) in the 1840–1990 period: mode 1 (top panel), mode 2 (middle panel) and mode 3 (bottom panel). Charted values are directly those from columns of matrices $\mathbf{P}$ (see Section 3.1) and are normalized such that $\mathbf{p}_i^T \mathbf{p}_j = \delta_{ij}$. Contours of ξ/R_c streamfunction are used to visualize the flow, projected onto the equatorial plane, as seen from the North pole.

there is no orthogonality imposed between them and the variability modes.

The cumulative importance of the three first SVD modes both in terms of u_{rms} and the energy in the SV that they account for is gath-

ered in Table 5. We can conclude that the compressed description of the flows using the mean and the three first variability modes, already accounts for about 90 per cent of the total u_{rms}. The percentage of SV explained by this simplified flow description is even
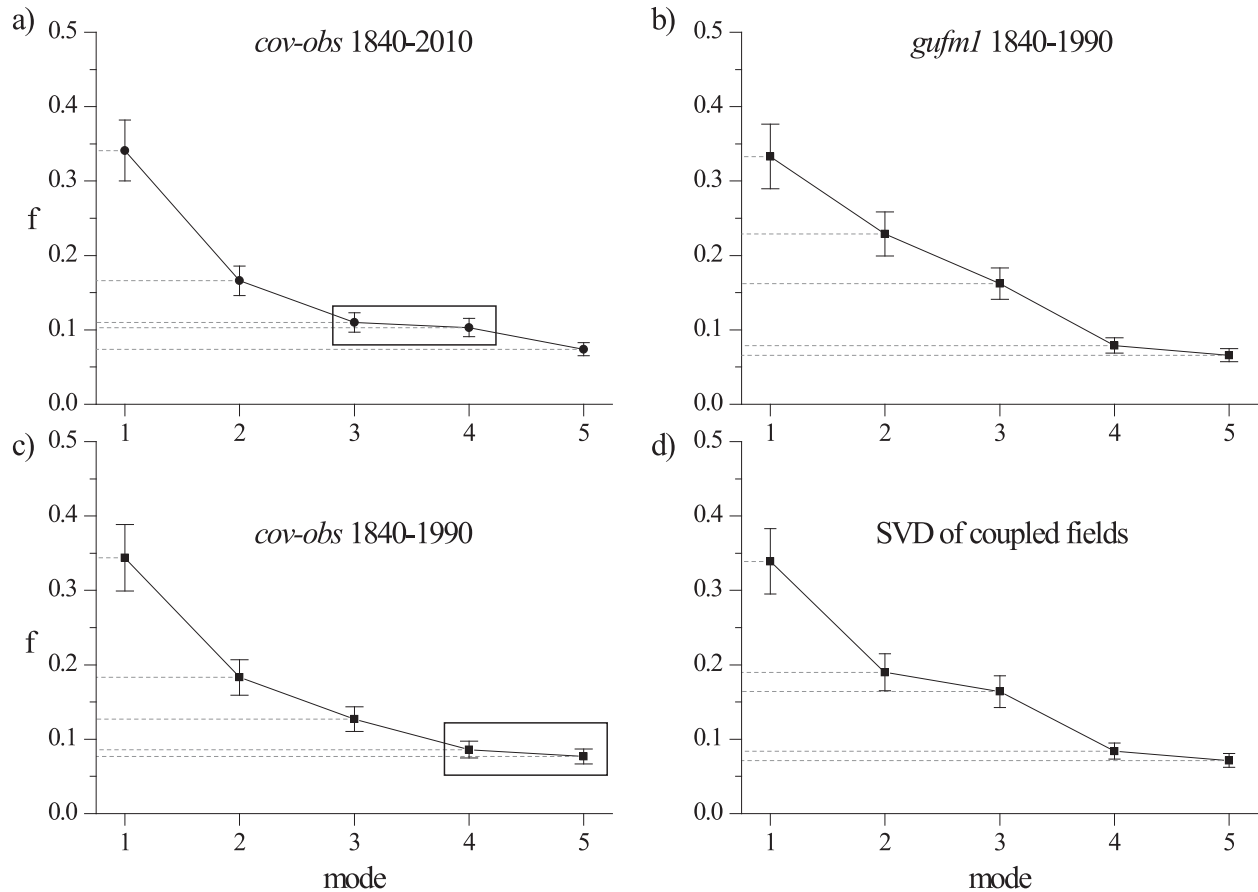
Figure 3. Fraction of explained variability f_i for PCA applied to different data sets and for SVD of coupled flow^{gufm1} and flow$^{COV-OBS}$ (on bottom, right-hand side), with error bars with total length $2\delta f_i$, computed from eq. (B1). Degenerate modes following North's criterion (eq. B2) are inside boxes.

Table 2. Subdomain stability of EOF/PCs for flow^{gufm1} and flow$^{COV-OBS}$, when considering separately the Pacific and the Atlantic hemispheres. Percentage of variance explained by each mode and, in parenthesis, estimate of corresponding standard error, calculated as in Table 1.

	Fraction of variance, f_i (per cent) (δf_i)				
	Mode-1	Mode-2	Mode-3	Mode-4	Mode-5
COV-OBS (1840–1990), PH	29.7 (3.7)	19.3 (2.4)	16.2 (2.0)	9.5 (1.2)	7.7 (1.0)
COV-OBS (1840–1990), AH	42.8 (5.5)	15.7 (2.0)	13.0 (1.7)	8.3 (1.1)	5.6 (0.7)
gufm1 (1840–1990), PH	40.3 (5.0)	23.4 (3.0)	11.5 (1.5)	7.4 (0.9)	5.3 (0.7)
gufm1 (1840–1990), AH	33.5 (4.3)	27.0 (3.4)	20.3 (2.6)	5.6 (0.7)	4.4 (0.6)

Table 3. Subdomain stability of EOF/PCs for flow^{gufm1} and flow$^{COV-OBS}$, when considering separately the Pacific hemisphere (PH) and the Atlantic hemisphere (AH). Absolute values of congruence coefficients between corresponding EOFs computed using the global grid and the grid under only the PH or the AH ($|g(PH, global)|$ and $|g(AH, global)|$, respectively). Absolute values of correlation coefficients between corresponding PCs computed using the global grid and the grid under PH or AH ($|r(PH, global)|$ and $|r(AH, global)|$, respectively). Bold is used for g and r values above 0.8, indicating high similarity.

	$|g(PH, global)|$ / $|r(PH, global)|$		$|g(AH, global)|$ / $|r(AH, global)|$	
	gufm1	COV-OBS	gufm1	COV-OBS
Mode-1	**0.99** / **0.95**	**0.87** / 0.77	**0.94** / **0.95**	**0.98** / **0.98**
Mode-2	**0.88** / 0.77	0.62 / 0.65	**0.93** / **0.94**	**0.93** / **0.96**
Mode-3	**0.85** / **0.81**	**0.96** / **0.90**	**0.96** / **0.97**	**0.92** / **0.90**
Mode-4	0.25 / 0.30	0.68 / 0.79	0.72 / 0.70	0.39 / 0.34
Mode-5	0.25 / 0.32	**0.87** / **0.85**	0.73 / **0.84**	0.21 / 0.28

Table 4. SVD of coupled flows flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$. In the first column, values for the percentage of cross-covariance explained by each coupled mode and, in parenthesis, estimates of corresponding standard error according to eq. (B1). Second and third columns are for congruence (eq. B4) and correlation coefficients that compare corresponding EOFs and PCs, respectively, that reconstruct flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$. In square brackets, values comparing corresponding EOFs and PCs from independent PCA applied to flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$.

	f_i (per cent) (δf_i)	$|g(EOF_i^{\text{gufm1}}, EOF_i^{\text{COV-OBS}})|$	$|r(PC_i^{\text{gufm1}}, PC_i^{\text{COV-OBS}})|$
Mode-1	33.9 (4.4)	0.80 [0.56]	0.96 [0.75]
Mode-2	19.0 (2.5)	0.74 [0.45]	0.93 [0.40]
Mode-3	16.4 (2.1)	0.62 [0.24]	0.94 [0.38]
Mode-4	8.4 (1.1)	0.61 [0.63]	0.92 [0.75]
Mode-5	7.1 (0.9)	0.47 [0.31]	0.91 [0.70]

greater, amounting to 95 per cent or more. To see this, each mode added to the mean flow and previous variability modes, was made to interact with the corresponding geomagnetic field model for the whole 1840–1990 period and the relative SV energy was computed for each epoch. From Table 5 we can conclude that the variability in flow$^{\text{COV-OBS}}$ is higher than in flow$^{\text{gufm1}}$, since larger fractions of the whole flow u_{rms} and induced SV are explained by variability modes. Finally, the fraction of mean flow projected onto each one of the three modes is also shown in the last two columns. Mode 2, in particular, contributes significantly to both mean flows.

Estimates of significance of the correlation coefficients were computed as explained in Appendix B. To build one single map that could concentrate information from the SVD analysis, we computed for mode i, with $i = 1, 2, 3$, the mean of the two PCs, PC_i^{gufm1} and $PC_i^{\text{COV-OBS}}$ since they are very close (see Fig. 5). Then, the resulting time function $\overline{PC}_i$ was correlated with flow$^{\text{gufm1}}$ and with flow$^{\text{COV-OBS}}$, giving two correlation maps r_i^{gufm1} and $r_i^{\text{COV-OBS}}$, respectively. To condense information in one single map, a new parameter $(r_i^{\text{gufm1}} r_i^{\text{COV-OBS}})^{1/2} (r_i^{\text{gufm1}}/|r_i^{\text{gufm1}}|)$ is introduced and its geographical distribution over the Northern Hemisphere is shown in Fig. 7 using a colour code. For each of the correlation coefficients r_i^{gufm1} and $r_i^{\text{COV-OBS}}$, a corresponding p-value chart was computed (see Appendix B). The $p = 0.1$ iso-contours for correlation of $\overline{PC}_i(t)$ with flow$^{\text{gufm1}}$ (dashed lines) and for correlation of $\overline{PC}_i(t)$ with flow$^{\text{COV-OBS}}$ (solid lines) are plotted in Fig. 7. These contours enclose the regions with correlation coefficients r_i^{gufm1} or $r_i^{\text{COV-OBS}}$ statistically significant at level 90 per cent or higher (i.e. the probability that these correlations are obtained by chance is no more than 10 per cent), which means their variability is well explained by $\overline{PC}_i(t)$ in both flow models. In deep blue (or red) regions, both flows evolve in time, strongly correlated (or anticorrelated) with the corresponding $\overline{PC}_i(t)$. Furthermore, if such patches are inside $p = 0.1$ iso-contours of both r_i^{gufm1} and $r_i^{\text{COV-OBS}}$, we can conclude with a relatively high confidence (i.e. a 90 per cent probability of success) that they represent spatial features of variability modes common to both flows, their temporal variation being represented by $\overline{PC}_i(t)$. From Fig. 7, top panel, we conclude that not only the main three vortices are significant features of the first mode, but also several smaller ones. Similarly, the zonal equatorial flow characterizing mode 2 is also a relevant feature in both flows, unlike the zonal flow component next to the TC, which is significant in flow$^{\text{gufm1}}$ but not in flow$^{\text{COV-OBS}}$. As explained, the longitudinal flow component is in the same direction near the TC and in the equatorial region, in spite of different signs of ξ there (resulting in different signs of the correlation with $\overline{PC}_2(t)$). As for mode 3, the presence of correlated rolls under AH is significant, cyclonic around 1860 and 1970, anticyclonic around 1920 (see also Fig. 5). Their localization is however fuzzy as we can conclude from a relatively weak superposition of

corresponding p-level curves from correlation with flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$. Under the PH, this third mode shows more variability in flow$^{\text{COV-OBS}}$ than in flow$^{\text{gufm1}}$.

In order to visualize how and when these significant flow structures can be seen in the flow, Fig. 8 shows the flow reconstructed from the mean and the two first SVD modes for flow$^{\text{gufm1}}$, at epochs 1860, 1900, 1930, 1970 and 1980. Referring only to the most significant features simultaneously spotted in both flows (see Fig. 7), the jet breakup can occur either because of an intensification of the cyclone centred at $\sim\!-150°$E longitude (mode 1), either because of an intensification of the high-latitude cyclones cutting the TC under Asian continent, the most important centred at $\sim\!60°$E longitude (mode 2). Negative values of PC_1 and/or PC_2 tend to reinforce the jet, as can be seen in 1930.

5 DISCUSSION

5.1 Mean flow

The mean flows obtained from our inversions of gufm1 and COV-OBS, shown in Fig. 1 are not a surprise, both displaying an eccentric gyre already described by Pais & Jault (2008). There are however several differences between the two: flow$^{\text{COV-OBS}}$ needs a much weaker Pacific recirculation, and a lower overall rms value for the mean flow. As can be seen, the centennial timescale flow representing the mean shows strong latitudinal jets as main features, an anisotropy which is seen on different natural rotating flow systems, with or without magnetic fields. This fact was used by Schaeffer & Pais (2011) to propose an anisotropic regularization favouring zonation and leading to a mean flow not much different from that in Fig. 1. The dominance of $m = 1$ azimuthal wavenumber in the mean flow suggests that an $m = 1$ forcing strongly influences the core flow. With direct numerical simulations of the core, Aubert *et al.* (2013) advocate for a differential inner-core growth (Alboussière *et al.* 2010; Monnereau *et al.* 2010) to explain the jet eccentricity. In their model, the westward circulation is an effect of uppermost liquid core lagging behind the mantle which, due to gravitational coupling with the inner core is indirectly pushed eastwards by thermochemical winds inside the TC. However, the pattern observed here is strikingly close to the one found by Hori *et al.* (2014; compare their fig. 4c with our Fig. 1) with a geodynamo simulation with internal sources and an heterogeneous ($m = 1$) heat flux imposed by the mantle. Thus the mean flows inverted from both 'COV-OBS' and 'gufm1' could suggest a control of the convection by heterogeneous heat extraction from the mantle. They might indicate that more heat is extracted from the core in a hemisphere centred on the Western Atlantic, between $-60°$ and $-90°$ longitude, if we follow the findings of Hori *et al.* (2014).
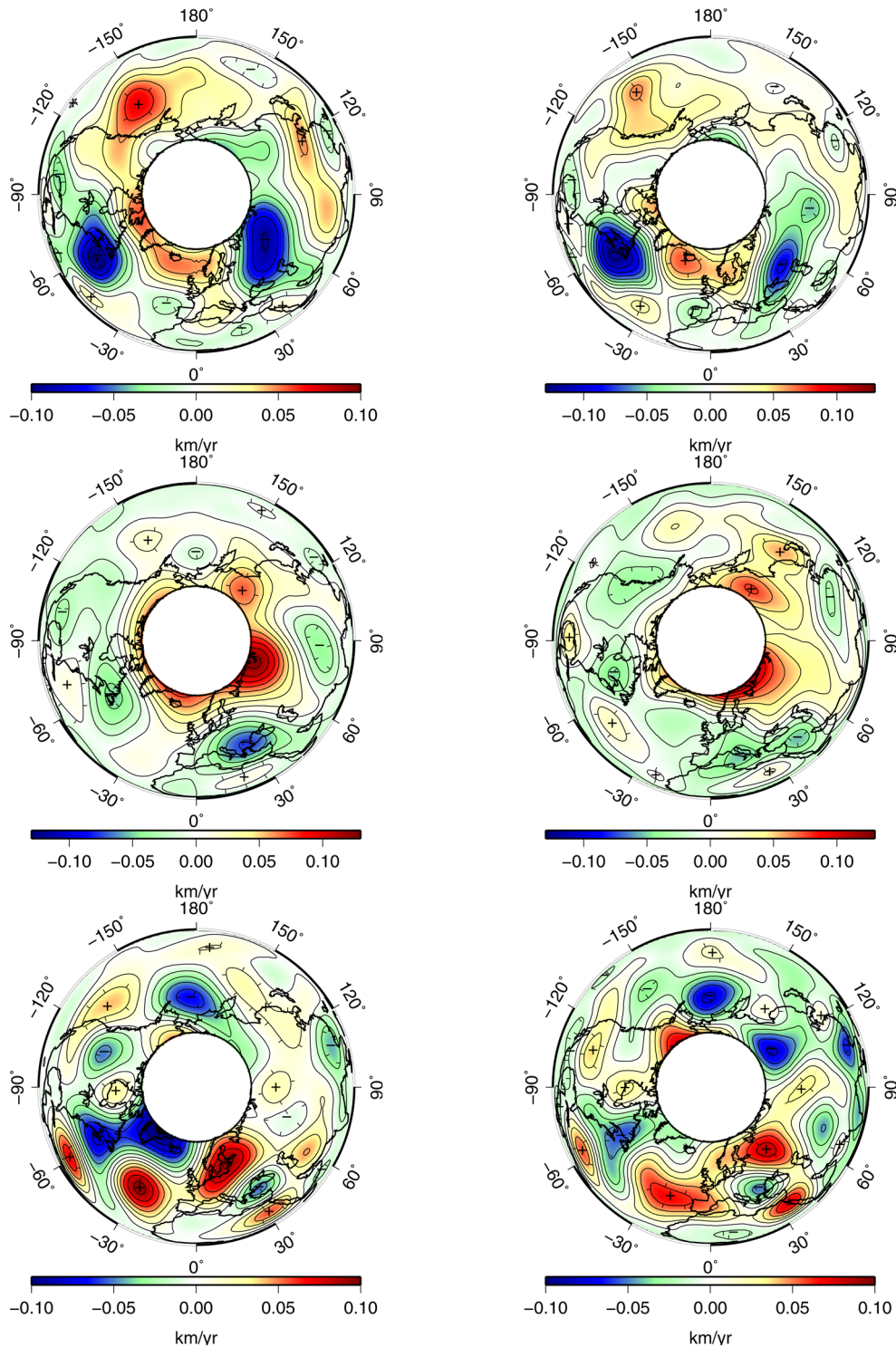
Figure 4. The three first pairs of EOFs resulting from SVD of coupled fields applied to flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$. For mode 1 (top panel), mode 2 (middle panel) and mode 3 (bottom panel), the EOFs allowing to reconstruct flow$^{\text{gufm1}}$ (left-hand side) and flow$^{\text{COV-OBS}}$ (right-hand side). Values are directly those from columns of matrices **U** and **V** (see Section 3.2) and are normalized such that $\mathbf{u}_i^{\text{T}}\mathbf{u}_j = \mathbf{v}_i^{\text{T}}\mathbf{v}_j = \delta_{ij}$. Contours of ξ/R_c streamfunction are used to visualize the flow, projected onto the equatorial plane, as seen from the North pole.

5.2 First mode

The PCA of both flow models show similar empirical orthogonal modes. The SVD of the coupled flows further pinpoints common features. The method is particularly convenient in making corresponding time functions (PCs) and spatial patterns (EOFs) that

reconstruct flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$ more similar (Table 4). The empirical mode that carries the largest variability of the flow consists mainly of three big vortices (see Fig. 4): two anticyclones located at mid-latitudes around $-60°$ and $+45°$ longitude, respectively; a cyclone located under the Eastern Pacific ocean around $-150°$
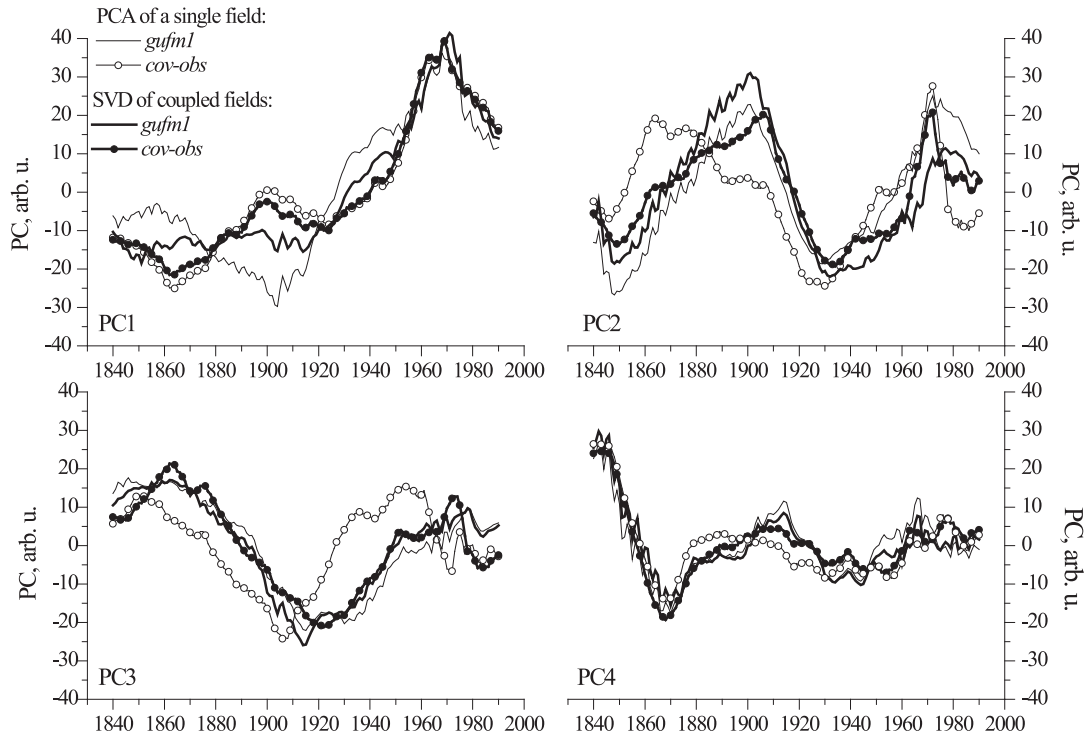
Figure 5. Temporal functions representing: the expansion coefficients of PCA modes that reconstruct flow^gufm1 (thin solid line) and flow^COV-OBS (thin line with empty circles); the expansion coefficients of SVD coupled modes that reconstruct flow^gufm1 (thick solid line) and flow^COV-OBS (thick line with filled circles).
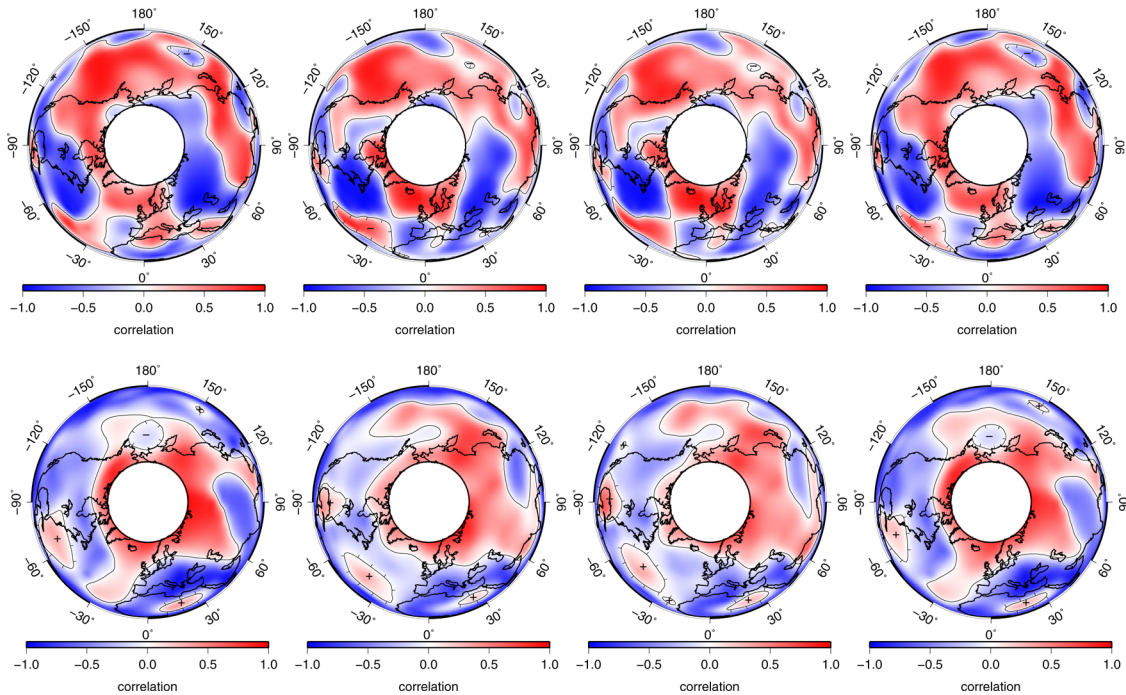

Figure 6. Homogeneous (two left columns, first for PC_i^{gufm1} correlated with flow^gufm1, second for $PC_i^{COV-OBS}$ correlated with flow^COV-OBS) and heterogeneous (two right columns, first for PC_i^{gufm1} correlated with flow^COV-OBS, second for $PC_i^{COV-OBS}$ correlated with flow^gufm1) correlation maps for SVD modes 1 (top row) and 2 (bottom row). Results are projected onto the equatorial plane and seen from the North pole.

longitude, but which appears weaker on the flow inverted from 'COV-OBS'. The significance analysis carried out on this mode further indicates that all the main three rolls (as well as a few weaker ones) are meaningful, the centres of the most significant regions being located at medium latitudes between 40° and 50° (see Fig. 7).

Note also the cyclonic features close to the TC under the North Atlantic ocean. The main effect of this empirical mode when added to the mean flow is to slightly reinforce the mean eccentric gyre before 1940, and later break it by making the three vortices largely dominant after 1960, and especially around 1970 (see Fig. 8).

Table 5. First and second columns show cumulative values of $u_{\mathrm{rms,i}} = \langle \mathbf{u}_i \rangle = \left((1/4\pi R_c^2) \int \mathbf{u}_i \cdot \mathbf{u}_i \, dS \right)^{1/2}$, starting from the mean flow over the 1840–1990 period; average and standard deviation (std) values are displayed. Third and fourth columns show the cumulative fraction of SV predicted, $\langle \dot{\mathbf{B}}_i \rangle / \langle \dot{\mathbf{B}} \rangle$ where $\langle \dot{\mathbf{B}}_i \rangle = \left((1/4\pi R_c^2) \int \dot{\mathbf{B}}_i \cdot \dot{\mathbf{B}}_i \, dS \right)^{1/2}$, when comparing to the SV signal induced by the total flows, $\langle \dot{\mathbf{B}} \rangle$; average and std values are displayed. Fifth and sixth columns show the normalized contribution of each orthogonal mode in the mean flow expansion, that is, $\left(\alpha_i^2 / \sum_{k=1}^{N_p} \alpha_k^2 \right)^{1/2}$ and $\left(\beta_i^2 / \sum_{k=1}^{N_p} \beta_k^2 \right)^{1/2}$, respectively (see text). The last row shows values of u_{rms} mean and standard deviation for the whole flows flow$^{\mathrm{gufm1}}$ and flow$^{\mathrm{COV\text{-}OBS}}$ (first and second column). Note that they induce an SV signal that does not fit exactly the SV models, mainly because of the parametrization error (see Appendix A). The percentage of SV rms explained by both global flows is on average 87 per cent.

	u_{rms} (km yr^{-1})		$\langle \dot{\mathbf{B}}_i \rangle / \langle \dot{\mathbf{B}} \rangle$		Mean flow projection	
	gufm1	COV-OBS	gufm1	COV-OBS	gufm1	COV-OBS
Mean flow	12.2	10.0	0.78 ± 0.16	0.75 ± 0.11	–	–
Mode-1	14.0 ± 3.2	12.2 ± 3.6	0.88 ± 0.19	0.86 ± 0.15	0.17	0.06
Mode-2	14.8 ± 3.1	13.1 ± 3.2	0.94 ± 0.20	0.91 ± 0.13	0.35	0.42
Mode-3	15.9 ± 4.0	14.6 ± 2.9	0.98 ± 0.17	0.95 ± 0.08	0.22	0.32
Total flow	17.1 ± 5.1	16.7 ± 3.1	1.00	1.00	–	–

5.3 Second mode

The second mode coming out of our analysis has the interesting property of carrying most of the angular momentum of our core flows, as shown in Fig. 9 (right-hand side) where the estimations for length-of-day (LOD) variations δT have been computed for our core flows and compared to LOD observations. The flow contributing to these variations is concentrated near the equator and the TC (see Fig. 9, left-hand side). In addition to the zonal part, mode 2 also shows several vortices and especially two co-rotating ones located close to the TC under Eastern and Western Russia. There are also vortices with opposite circulation, the most significant one being located under north of the Mediterranean basin, at mid-latitudes (see Fig. 7, region with $10°\mathrm{E} < \mathrm{long.} < 20°\mathrm{E}$ and $30° < \mathrm{lat.} < 50°\mathrm{N}$). During the period 1840–1990, this mode shows two main oscillations basically reinforcing the eccentric gyre around 1850 and 1930, and weakening it around 1900 and 1980, with vortices and an opposite circulation around the TC. The fact that most of the variation in zonal motion is correlated to these vortices might indicate either an excitation of a (zonal) eigenmode by the convecting flow, or the presence of slave secondary vortices excited by the interaction of a zonal eigenmode with the background magnetic field.

In this study, we show that these oscillations, have a mean period between 80 and 90 yr (see Fig. 5). These rather long periods cannot be explained by torsional oscillations which are thought to have much shorter periods of about 6 yr (Gillet *et al.* 2010), but could, in principle, be a match for (a combination of) MAC oscillations (Braginsky 1993), as shown recently by Buffett (2014). However, these MAC oscillations are restricted to a thin stratified layer at the top of the core. Fig. 9 (right-hand side) shows that the simple restriction of our QG core flow to a thin layer largely underestimates the LOD variations. Hence, with the hypothesis of MAC oscillations, the core flow cannot explain the LOD variations directly (at least not for the velocity rms values of 10–20 km yr^{-1} constrained by the SV), while a QG model extending in the whole core nicely explain these (e.g. Gillet *et al.* 2010; Schaeffer & Pais 2011). It might be worth mentioning that the solar activity exhibits a period of 80–100 yr, the Gleissberg cycle (e.g. Hathaway 2010), which might provide an excitation for our mode 2. Periods of 96 yr and 76 yr associated to damped waves with parameters fitted to core surface flows have been reported by Zatman & Bloxham (1997, 1998). A core flow mode with a similar period has also been identified by Dickey & de

Viron (2009, 85 yr) and Buffett *et al.* (2009, 86.3 yr) using different methods. However, all those studies focused on the zonal longitudinal flow component inverted from 'gufm1' or a prior version, 'ufm1' (Bloxham & Jackson 1992). In some cases, a shorter time span (1900–1990) was analysed (e.g. Zatman & Bloxham 1997, 1998). The standing oscillation reported in our study also involves non-zonal features, mainly close to the TC, and is detectable during the whole 1840–1990 period. Furthermore, it combines information retrieved from two geomagnetic field models, 'gufm1' and 'COV-OBS'. It must be emphasized that the time-dependence of mode 2 is not a nice sine wave, and that several frequencies are needed to reconstruct the signal. We obtain a dominant period of 80–90 yr for the time-span 1840–1990 as well as for SVD applied to the shorter interval 1900–1990. However, if we restrict the SVD analysis from 1920 to 1990 the period decreases to $\sim$70 yr and a 60 yr sinusoid fitted to mode 2 gives an almost equally high *R*-squared value, in agreement with other studies (Roberts *et al.* 2007; Buffett 2014).

5.4 Flow reconstruction and other modes

The flow reconstructed from the first two modes and the mean flow is shown in Fig. 8 and basically exhibits a very strong eccentric gyre around 1860 and 1930, which is destroyed by the appearance of several vortices around 1900 (especially two counter-rotation vortices under Asia) and by three very strong and large vortices around 1970, basically splitting the unique global gyre into two circulation systems, a cyclonic one located under the Pacific ocean, and an anticyclonic one under Europe and the Atlantic ocean.

The third mode is less important in terms of the pseudo-streamfunction ξ variability explained, still it accounts for more than 10 per cent either using PCA or SVD (Fig. 3). It's spatial pattern is also more complex showing a distribution of smaller scale vortices, with often no clear correspondence between flow$^{\mathrm{gufm1}}$ and flow$^{\mathrm{COV\text{-}OBS}}$. The correlated vortices clustered under the Atlantic region show nonetheless some correspondence in both flows, significant only in reduced regions of Fig. 7.

We also checked higher order modes, which concentrate smaller fractions of the data variability. We looked in particular for a 6-yr periodicity, which has been reported by Gillet *et al.* (2010). Such periodicities can be glimpsed at modes 8 and 10 of PCA applied
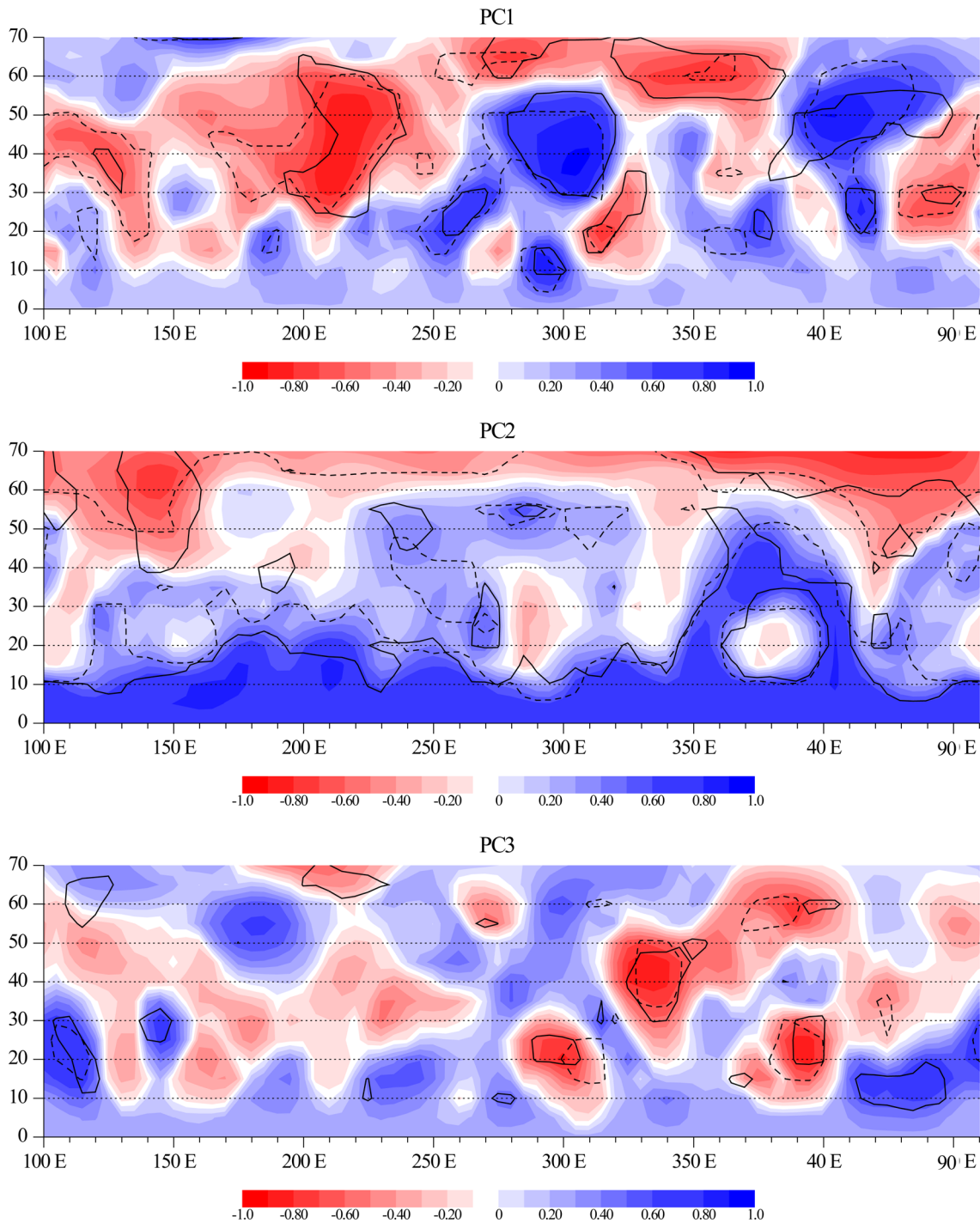
Figure 7. Maps for the parameters $(r_i^{\mathrm{gufm1}}\, r_i^{\mathrm{COV\text{-}OBS}})^{1/2}\,(r_i^{\mathrm{gufm1}}/|r_i^{\mathrm{gufm1}}|)$ (coloured code), overlaid with p-level lines of 0.1 for the correlation coefficients r_i^{gufm1} (dashed) and $r_i^{\mathrm{COV\text{-}OBS}}$ (solid). Top panel: $i = 1$, middle panel: $i = 2$, bottom panel: $i = 3$.

to COV-OBS for the time period 1960–2010. However, they never stand out above other relatively small periods that are also present, as is the case of a period of 9 yr. Indeed, to isolate this component, Gillet *et al.* (2010) need to extract the zonal flow of the ensemble average, which is then bandpass filtered around 6 yr, resulting in a signal several order of magnitudes smaller than the whole flow. Other tools should be employed to look at small signals of specific periods, as the EOF/PC analysis is not the adequate tool to isolate such a relatively low amplitude variability.

6 CONCLUSIONS

We use PCA tools applied to 'data' consisting of the pseudo-streamfunction ξ evaluated on a regular spatial grid at the core surface outside the TC. Our main goal is to identify some underlying structure in the flow and reconstruct the data from a linear combination of a small number of spatial patterns multiplied by time-varying coefficients. In this attempt for a simplification of observations, we first extracted the mean flow over the observatory
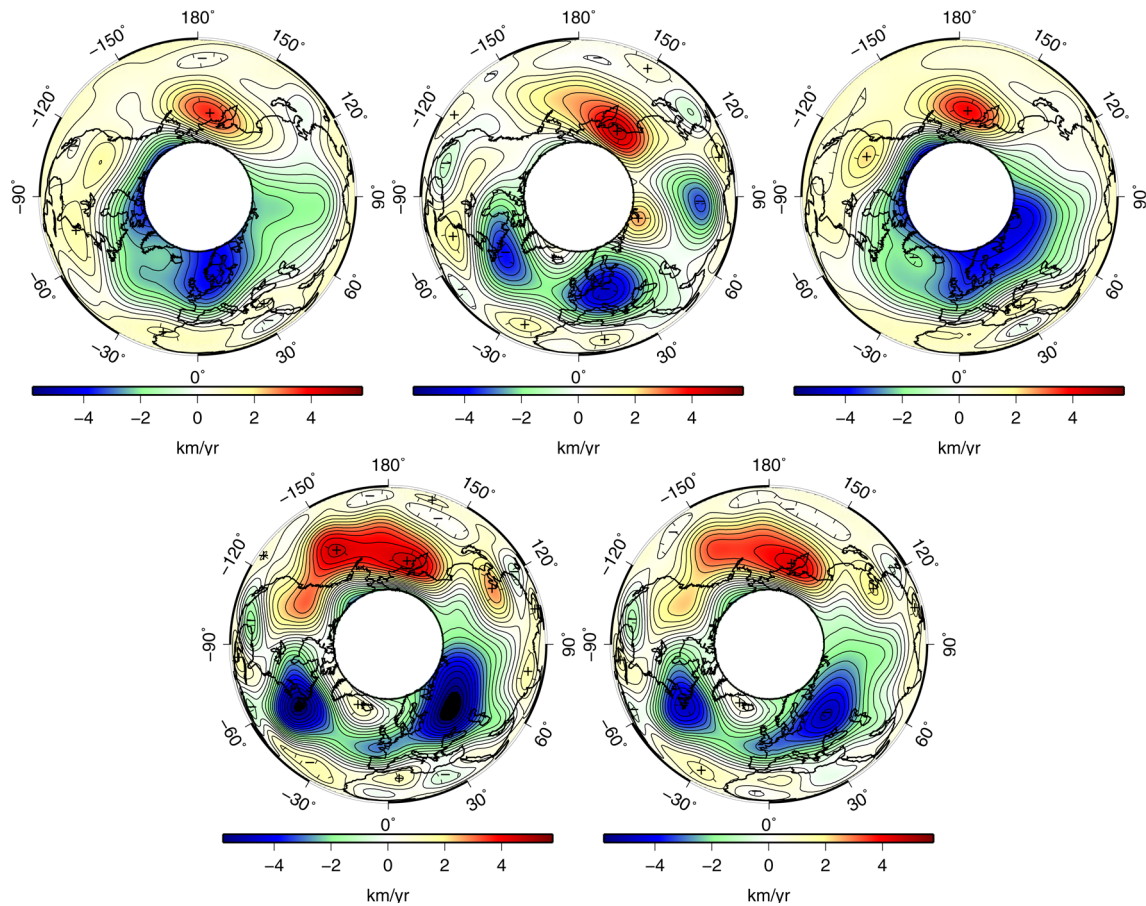
Figure 8. Contours of pseudo streamfunction ξ/R_c in units of km yr^{-1}, for a reconstruction of flow^{gufm1} using the mean flow and the two first SVD modes. Shown epochs are, from top left to bottom right: 1860 (both PC$_1$ and PC$_2$ are weak, negative), 1900 (PC$_2$ strong, positive), 1930 (PC$_2$ strong, negative), 1970 (PC$_1$ very strong, positive) and 1980 (PC$_1$ and PC$_2$ strong, positive). In epochs of high positive values of PC$_1$ and/or PC$_2$, the flow features in these two modes act to destroy the large eccentric jet, their amplitude being of the same order of magnitude as the mean flow. See also the movies (Pais *et al.* 2014).
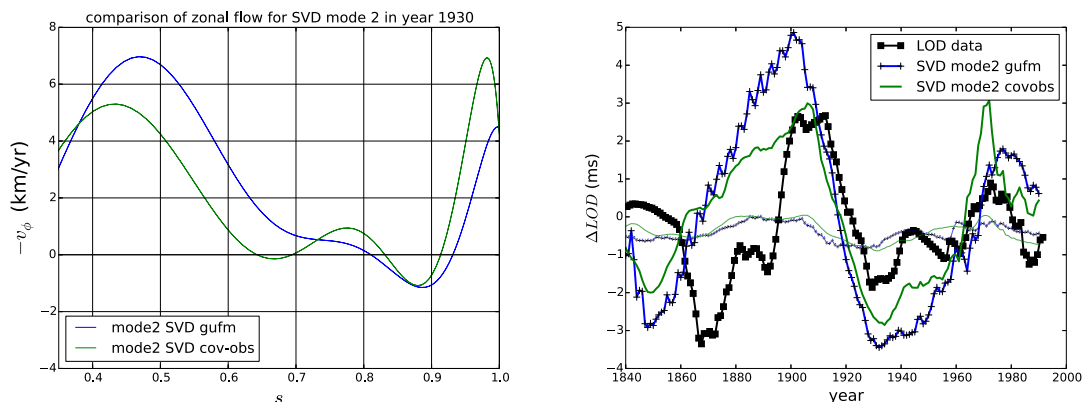
Figure 9. On the left-hand side, the zonal component of u_ϕ for SVD mode 2, in km yr^{-1}, as a function of normalized distance to the rotation axis. On the right-hand side, the observed LOD variation (thick black line with squares), estimates from our SVD mode 2 assuming a columnar flow (thick blue and green lines, see Jault & Finlay in press, for the formula) or assuming the flow is restrained to a stratified layer at the top of the core of thickness 140 km (thin lines).

era period (1840 to present): a large eccentric jet, flowing at lower latitudes under the AH and at higher latitudes under the PH and a large cyclone under the PH, centred at medium-latitudes. The clear dichotomy revealed in this mean flow can be due to heterogeneous thermal forcing at the core–mantle boundary or at the solid core boundary (Aubert *et al.* 2007; Aubert *et al.* 2013; Hori *et al.* 2014).

We further identified three main circulation modes that account together for about 70 per cent of the observed variability, 90 per cent

of the total flow u_{rms} and 95 per cent of the SV predicted by the total flow. Mode 1, consists of three large QG vortices indicative of large-scale convection. It has an aperiodic variability during the inspected period, with a main boost around the 1970 epoch. At this time, the amplitude of mode 1 largely dominates the mean flow u_{rms} and the three-roll flow breaks the large eccentric jet. The same dichotomy as in the mean flow reveals the dominance of anticyclonic circulation under the AH and cyclonic one under the PH.

A second mode was identified, which concentrates variations of core angular momentum, with a quasi-periodicity of 80–90 yr. Periods close to 80 or 90 yr have been found in magnetic field observatory data (Jackson & Mound 2010), geomagnetic field models (Kang *et al.* 2008) and inverted flows (Zatman & Bloxham 1997, 1998; Buffett *et al.* 2009; Dickey & de Viron 2009). Never before, however, have they been put forward without focusing the analysis on the zonal component of the flow and resorting to flows inverted from different geomagnetic field models. Here, this period emerges associated with an empirical orthogonal mode of the whole system, naturally associated to a large zonal flow but also, interestingly, to some significant small-scale circulations mainly next to the TC. These vortices evolve correlated with the equatorial flow and can also contribute to destroy or reinforce the large jet. The spatial structure of this mode has an important projection on the mean flow suggesting, if we believe on the presence of this flow component inside the core, that it can give some insight into the morphology of the sources responsible for accelerating the large eccentric jet. The physical existence of this mode needs further testing, however. The presence of a mode of 80–90 yr period both in historical observatory data series and field models spanning centennial time periods seems confirmed. However, its absence from decade LOD historical data where a period of ∼60 yr is seen instead can sound suspicious, if we accept that decade LOD variations have their origin in angular momentum exchanges between the mantle and the core. A possible explanation would be that this flow mode results from a leakage of an external magnetic field signal (associated to the Gleissberg Sun-cycle) into internal geomagnetic field models and inverted core flows. Unfortunately, we are not in a position to deny (or confirm) this hypothesis based only on the results from this study.

Mode 3 shows the presence of a few correlated rolls under AH but their localization is fuzzy due to the fact that they are retrieved differently in flow$^{\text{gufm1}}$ and flow$^{\text{COV-OBS}}$. This precludes the emergence of significant spatial features.

Significance tests were carried out with results that support the spatial stability of modes 1, 2 and 3 in general, and of specific spatial features present in these modes in particular. The PCA tools used in this study look for the directions of maximum variance in data space. The modes corresponding to projection along these directions will be of special interest in the case that higher variability modes correspond to interesting dynamics and lower ones correspond to noise. We cannot guarantee from our results that this condition is fulfilled. However, we point out the similarities we found between certain identified modes (modes 1 and 2), significant both in terms of spatial features described and amount of variance explained, and other time/spatial features identified in previous studies of core dynamics.

This study gives important insight into the variability of the core flows on centennial timescales. We have shown that the mean flow and three empirical modes can account for most of the SV of the geomagnetic field. Direct physical interpretation of the observed variability modes is not straightforward, although we have been able to link the first one to a dominant vortex pattern, and the second one to observed LOD variations. These modes may also guide future analytical and numerical studies on QG modes (e.g. Canet *et al.* 2014) and maybe help to constrain geometry of the internal magnetic field.

Finally, decomposition of QG flows inverted from geomagnetic field models into a few number of spatial structures and time-dependent coefficients is of practical interest to serve as input for dynamical studies of the Earth's core interior where conditions closer to the real Earth are sought. An interesting question that this decomposition in empirical modes could help to answer is whether and how the core flow that we can reconstruct from magnetic models contributes to the generation of the Earth's magnetic field. We plan to address this question in a future study.

REFERENCES

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

Amit, H. & Pais, M.A., 2013. Differences between tangential geostrophy and columnar flow, *Geophys. J. Int.,* **194,** 145–157.

Aubert, J., 2013. Flow throughout the Earth's core inverted from geomagnetic observations and numerical dynamo models, *Geophys. J. Int.,* **192,** 537–556.

Aubert, J., 2014. Earth's core internal dynamics 1840–2010 imaged by inverse geodynamo modelling, *Geophys. J. Int.,* **198,** 1683–1683.

Aubert, J., Amit, H. & Hulot, G., 2007. Detecting thermal boundary control in surface flows from numerical dynamos, *Phys. Earth planet. Inter.,* **160,** 143–156.

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

Aurnou, J., Andreadis, S., Zhu, L. & Olson, P., 2003. Experiments on convection in Earth's core tangent cylinder, *Earth planet. Sci. Lett.,* **212,** 119–134.

Balasis, G. & Egbert, G., 2006. Empirical orthogonal function analysis of magnetic observatory data: further evidence for non-axisymmetric magnetospheric sources for satellite induction studies, *Geophys. Res. Lett.,* **33**(11), doi:10.1029/2006GL025721.

Barry, R.G. & Carleton, A.M., 2001. *Synoptic and Dynamic Climatology,* Routledge.

Bjornsson, H. & Venegas, A., 1997. A manual for EOF and SVD analyses of climatic data. Tech Rep No. 97-1, Department of Atmospheric and Oceanic Sciences and Centre for Climate and Global Change Research.

Bloxham, J. & Jackson, A., 1992. Time-dependent mapping of the magnetic field at the core–mantle boundary, *J. geophys. Res.,* **97,** 19 537–19 563.

Braginsky, S.I., 1993. MAC-oscillations of the hidden ocean of the core, *J. Geomagn. Geoelectr.,* **45,** 1517–1538.

Bretherton, C.S., Smith, C. & Wallace, J.M., 1992. An intercomparison of methods for finding coupled patterns in climate data, *J. Clim.,* **5,** 541–560.

Buffett, B., 2014. Geomagnetic fluctuations reveal stable stratification at the top of the Earths core, *Nature,* **507,** 484–487.

Buffett, B.A., Mound, J. & Jackson, A., 2009. Inversion of torsional oscillations for the structure and dynamics of Earth's core, *Geophys. J. Int.,* **177,** 878–890.

Busse, F.H., 1970. Thermal instabilities in rapidly rotating systems, *J. Fluid Mech.,* **44,** 441–460.

Canet, E., Finlay, C.C. & Fournier, A., 2014. Hydromagnetic quasi-geostrophic modes in rapidly rotating planetary cores, *Phys. Earth planet. Inter.,* **229,** 1–15.

Chulliat, A. & Hulot, G., 2000. Local computation of the geostrophic pressure at the top of the core, *Phys. Earth planet. Inter.,* **117,** 309–328.

Dickey, J.O. & de Viron, O., 2009. Leading modes of torsional oscillations within the Earth's core, *Geophys. Res. Lett.,* **36,** doi:10.1029/2009GL038386.

Eymin, C. & Hulot, G., 2005. On core surface flows inferred from satellite magnetic data, *Phys. Earth planet. Inter.,* **152,** 200–220.

Finlay, C.C., Dumberry, M., Chulliat, A. & Pais, M.A., 2010. Short timescale core dynamics: theory and observations, *Space Sci. Rev.,* **155,** 177–218.

Finlay, C.C., Jackson, A., Gillet, N. & Olsen, N., 2012. Core surface magnetic field evolution 2000-2010., *Geophys. J. Int.,* **189,** 761–781.

Fujii, I. & Schultz, A., 2002. The 3D electromagnetic response of the Earth to ring current and auroral oval excitation, *Geophys. J. Int.,* **151,** 689–709.

Gillet, N., Pais, M.A. & Jault, D., 2009. Ensemble inversion of time-dependent core flow models, *Geochem. Geophys. Geosyst.,* **10,** Q06004, doi:10.1029/2008GC002290.

Gillet, N., Jault, D., Canet, E. & Fournier, A., 2010. Fast torsional waves and strong magnetic field within the Earths core, *Nature,* **465,** 74–77.

Gillet, N., Schaeffer, N. & Jault, D., 2011. Rationale and geophysical evidence for quasi-geostrophic rapid dynamics within the earth's outer core, *Phys. Earth planet. Int.,* **187,** 380–390.

Gillet, N., Jault, D., Finlay, C.C. & Olsen, N., 2013. Stochastic modeling of the earth's magnetic field: inversion for covariances over the observatory era, *Geochem. Geophys. Geosyst.,* **14,** 766–786.

Gubbins, D., 1983. Geomagnetic field analysis: I. Stochastic inversion., *Geophys. J. Int.,* **73,** 641–652.

Hannachi, A., Jolliffe, I.T. & Stephenson, D.B., 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review, *Int. J. Climatol.,* **27,** 1119–1152.

Harman, H.H., 1976. *Modern Factor Analysis,* Univ. of Chicago Press.

Hathaway, D.H., 2010. The solar cycle, *Living Rev. Solar Phys.,* **7,** doi:10.12942/lrsp-2010-1.

Holme, R., 2007. Large-scale flow in the core, in *Treatise on Geophysics,* Vol. **8,** pp. 107–130, ed. Olson, P., Elsevier Science.

Hori, K., Wicht, J. & Dietrich, W., 2014. Ancient dynamos of terrestrial planets more sensitive to core-mantle boundary heat flows, *Planet. Space Sci.,* **98,** 30–40.

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

Jackson, L.P. & Mound, J.E., 2010. Geomagnetic variation on decadal time scales: what can we learn from empirical mode decomposition? *Geophys. Res. Lett.,* **37,** L14307, doi:10.1029/2010GL043455.

Jackson, A., Jonkers, A. & Walker, M., 2000. Four centuries of geomagnetic secular variation from historical records, *Phil. Trans. R. Soc. Lond. A,* **358,** 957–990.

Jault, D., 2008. Axial invariance of rapidly varying diffusionless motions in the earth's core interior, *Phys. Earth planet. Inter.,* **166,** 67–76.

Jault, D. & Finlay, C.C., in press. Waves in the core and mechanical core-mantle interactions, in *Treatise on Geophysics,* 2nd edn, eds Schubert, G. & Olson, P., Elsevier.

Jones, C.A., 2011. Planetary magnetic fields and fluid dynamos, *Annu. Rev. Fluid Mech.,* **43,** 583–614.

Kang, G.-F., Bai, C.-H. & Gao, G.-M., 2008. Periodical characteristics of the geomagnetic secular variation and decadal length-of-day variation, *Chin. J. Geophys.,* **51,** 240–247.

Kunsch, H.R., 1998. The jackknife and the bootstrap for general stationary observations, *Ann. Statist.,* **17,** 1217–1241.

Liu, R.Y. & Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence, in *Exploring the Limits of the Bootstrap,* pp. 225–248, eds Lepage, R. & Billard, L., Wiley.

Matsuo, T., Richmond, A.D. & Nychka, D., 2002. Modes of high-latitude electric field variability derived from DE-2 measurements: empirical orthogonal function (EOF) analysis, *Geophys. Res. Lett.,* **29,** 1107, doi:10.1029/2001GL014077.

McLeod, M.G., 1996. Spatial and temporal power spectra of the geomagnetic field., *J. geophys. Res.,* **101,** 2745–2763.

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

North, G.R., 1984. Empirical orthogonal functions and normal modes, *J. Atmos. Sci.,* **41,** 879–887.

North, G.R., Bell, T.L., Cahalan, R.F. & Moeng, F.J., 1982. Sampling errors in the estimation of empirical orthogonal functions, *Mon. Weather Rev.,* **110,** 699–706.

Pais, A. & Hulot, G., 2000. Length of day decade variations, torsional oscillations and inner core superrotation: evidence from recovered core surface zonal flows, *Phys. Earth planet. Inter.,* **118,** 291–316.

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

Pais, M.A., Oliveira, O. & Nogueira, F., 2004. Nonuniqueness of inverted core-mantle boundary flows and deviations from tangential geostrophy, *J. geophys. Res. (Solid Earth),* **109,** 8105–8200.

Pais, A., Morozova, A. & Schaeffer, N., 2014. COV-OBS (1840-2010): pseudo-streamfunction—world projection, *figshare,* doi:10.6084/m9.figshare.1061462.

Peng, S. & Fyfe, J., 1996. The coupled patterns between sea level pressure and sea surface temperature in the midlatitude North Atlantic., *J. Clim.,* **9,** 1824–1839.

Preisendorfer, R.W., 1988. *Principal Component Analyses in Meteorology and Oceanography,* Elsevier.

Richman, M.B., 1986. Rotation of principal components, *Int. J. Clim.,* **6,** 293–335.

Richman, M.B. & Lamb, P.J., 1985. Climatic pattern analysis of three- and seven-day summer rainfall in the central United States: some methodological considerations and a regionalization, *J. appl. Meteorol.,* **24,** 1325–1343.

Roberts, P.H., 1968. On the thermal instability of a self-gravitating fluid sphere containing heat sources, *Phil. Trans. R. Soc. Lond.* **A263,** 93–117.

Roberts, P.H., Yu, Z.J. & Russell, C.T., 2007. On the 60-year signal from the core, *Geophys. Astrophys. Fluid Dyn.,* **101**(1), 11–35.

Schaeffer, N. & Cardin, P., 2005. Quasi-geostrophic model of the instabilities of the Stewartson layer in flat and depth varying containers., *Phys. Fluids,* **17,** 104–111.

Schaeffer, N. & Pais, M.A., 2011. On symmetry and anisotropy of Earth-core flows, *Geophys. Res. Lett.,* **38,** L10309, doi:10.1029/2011GL046888.

Venegas, S.A., Mysak, L.A. & Straub, D.N., 1997. Atmosphere-ocean coupled variability in the South Atlantic, *J. Clim.,* **10,** 2904–2920.

Von Storch, H. & Navarra, A., 1995. *Analyses of Climate Variability: Applications of Statistical Techniques,* Springer-Verlag.

Wallace, J.M., Smith, C. & Bretherton, C.S., 1992. Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies, *J. Clim.,* **5,** 561–576.

Whaler, K.A., 1986. Geomagnetic evidence for fluid upwelling at the core-mantle boundary, *Geophys. J. Int.,* **86,** 563–588.

Zatman, S. & Bloxham, J., 1997. Torsional oscillations and the magnetic field within the Earth's core, *Nature,* **388,** 760–763.

Zatman, S. & Bloxham, J., 1998. A one-dimensional map of bs from torsional oscillations of the earths core, in *The Core-Mantle Boundary Region,* pp. 183–196, eds Gurnis, M., Wysession, M.E., Knittle, E. & Buffett, B.A., American Geophysical Union.

Zhang, S.-R., Chen, Z., Coster, A.J., Erickson, P.J. & Foster, J.C., 2013. Ionospheric symmetry caused by geomagnetic declination over North America, *Geophys. Res. Lett.,* **40,** 5350–5354.

APPENDIX A: ITERATIVE INVERSION OF gufm1 AND COV-OBS

In this appendix we outline the procedure to invert the geomagnetic field models 'gufm1' in the time period 1840–1990 and 'COV-OBS' in 1840–2010, for a QG core flow with surface core expression **u**. The method is regularized weighted least-squares inversion, whereby the objective function $\Phi(\mathbf{m})$ that expresses a linear combination of the discrepancy between observables and theoretical predictions and a term with quadratic forms on the flow model **u** is minimized (e.g Gubbins [1983]):

$$\Phi(\mathbf{m}) = \left(\mathbf{A}(\mathbf{b})\,\mathbf{m} - \dot{\mathbf{b}}\right)^{\mathrm{T}} \mathbf{C}_{\mathbf{e}}^{-1} \left(\mathbf{A}(\mathbf{b})\,\mathbf{m} - \dot{\mathbf{b}}\right) + \lambda\,\mathbf{m}^{\mathrm{T}}\mathbf{C}_{\mathbf{m}}^{-1}\mathbf{m}. \quad (A1)$$

m is the vector of poloidal ($s_\ell^{m(c,s)}$) and toroidal ($t_\ell^{m(c,s)}$) coefficients of the SH expansion of the scalars $\mathcal{S}$ and $\mathcal{T}$ that

determine $\mathbf{u} = R_c \nabla_H \mathcal{S} - R_c \hat{r} \wedge \nabla_H \mathcal{T}$; $\nabla_H = \nabla - \hat{r} \partial / \partial r$ is the horizontal nabla operator; $\mathbf{b}$ is the vector of SH time-dependent coefficients of the scalar potential V that determines the internal component of the geomagnetic field at the Earth's surface $\mathbf{B} = -\nabla V$ and $\dot{\mathbf{b}}$ is the vector of the corresponding first time derivatives; $\mathbf{A}$ is the interaction matrix or matrix of equations of condition, with elements that depend on the geomagnetic field model and on the Elsasser and Adams-Gaunt integrals (e.g. Whaler 1986):

$$\dot{\mathbf{b}} = \mathbf{A}(\mathbf{b})\,\mathbf{m} + \mathbf{e}, \tag{A2}$$

where $\mathbf{e}$ is the error.

The solution to this minimization problem is of the form

$$\hat{\mathbf{m}} = \left(\mathbf{A}^T \mathbf{C}_e^{-1} \mathbf{A} + \lambda \mathbf{C}_m^{-1}\right) \mathbf{A}^T \mathbf{C}_e^{-1}\,\dot{\mathbf{b}}. \tag{A3}$$

The formalism above is standard and can be found in different studies (see e.g. Pais & Hulot 2000). We now concentrate on specific features of our inversions.

A1 The error covariance matrix, $\mathbf{C}_e$

In the inversions for this study, coefficients up to degree and order $\ell_{\mathbf{b}} = \ell_{\dot{\mathbf{b}}} = 13$ from model *mod* are used, for the deterministic part of the main field and its SV, where *mod* stands for 'gufm1' or 'COV-OBS'. As to the dimension of vector $\mathbf{m}$, it is that required to solve for flow coefficients up to degree and order $\ell_{\mathbf{m}} = 26$, the maximum degree that can be constrained (even if only slightly) by the first 13 SH degrees of the MF and of the SV.

The error $\mathbf{e}$ is treated as a Gaussian random variable with mean zero and covariance matrix $\mathbf{C}_e$, which can be condensed using the notation $\mathbf{e} \sim \mathcal{N}(0, \mathbf{C}_e)$. We consider two contributions for $\mathbf{C}_e$: (i) $\mathbf{C}_e^{mod}$ is the *a posteriori* covariance matrix for the SV coefficients coming from the geomagnetic field model 'calculation'; (ii) $\mathbf{C}_e^r$ is the covariance matrix characterizing the spatial resolution or modelling errors. This second term is dependent on the flow solution $\mathbf{m}$:

$$\mathbf{C}_e = \mathbf{C}_e^{mod} + \mathbf{C}_e^r(\mathbf{m}). \tag{A4}$$

For $\mathbf{C}_e^{mod}$, we will be using the diagonal $\mathbf{C}_e^{gufm}$ matrix from Jackson (1997) and the dense $\mathbf{C}_e^{COV\text{-}OBS}$ matrix provided by Gillet *et al.* (2013). Previous studies have identified the modelling errors arising from an incomplete knowledge of the main field ($\mathbf{C}_e^r(\mathbf{m})$) as having a higher contribution to the large scales of the SV than do observational errors (e.g. Eymin & Hulot 2005; Pais & Jault 2008; Gillet *et al.* 2009). Inversions reported in recent papers do already take into account these errors, though different approaches have been followed, in particular the iterative approach in Pais & Jault (2008) and the stochastic approach in Gillet *et al.* (2009). More recently another method has been proposed, called inverse geodynamo modelling, whereby the estimation of the effects of underparametrization is based on a statistical study of a numerical dynamo used as a prior model (Aubert 2013, 2014).

A law is required to prolongate the spectrum of the geomagnetic field to length scales that are unperceivable at the Earth's surface and above. We used the form $R(\ell, R_c) = W/(2\ell + 1)$ for the Lowes–Mauersberger spectrum at the CMB, from McLeod (1996). This function was fitted to the geomagnetic spectrum computed from gufm-sat-E3 (Finlay *et al.* 2012) for the 2005 epoch and using only $3 \le \ell \le 13$. The value $W = 1.11 \times 10^{11}\,\text{nT}^2\,\text{yr}^{-2}$ was found. The small-scale magnetic field ($14 \le \ell \le 40$) was then treated as a Gaussian random variable of zero mean and variance–covariance matrix $\mathbf{C}_{\mathbf{b}^S}$, that is,

$$\mathbf{b}^S \sim \mathcal{N}(0, \mathbf{C}_{\mathbf{b}^S}), \tag{A5}$$

where $\mathbf{C}_{\mathbf{b}^S}$ is a diagonal matrix with elements given by

$$\sigma_{\mathbf{b}^S}^2(\ell) = \frac{W}{(\ell+1)(2\ell+1)^2}\left(\frac{R_c}{R_E}\right)^{2\ell+4}, \tag{A6}$$

R_E being the mean Earth radius.

To compute $\mathbf{C}_e$ for a certain flow solution $\mathbf{m}$, an ensemble of K modelling error vectors $\mathbf{e}^{r,k}(\mathbf{m}) = \mathbf{A}(\mathbf{b}^{S,k})\,\mathbf{m}$ are calculated, from an ensemble of K small-scale magnetic fields $\mathbf{b}^{S,k}$ with statistical properties given by (A5) and (A6). From this sampling, estimates of the mean and covariance matrix for the modelling (or spatial resolution) error are computed according to standard formulae, the (non-diagonal) covariance matrix being denoted $\mathbf{C}_e^r(\mathbf{m})$. The updated error covariance matrix is then computed from eq. (A4). In all our computations we used $K = 80$.

A2 The flow regularization

The second term on the RHS of (A1), also referred to as model regularization, incorporates the assumptions on the flow; its relative weight in the linear combination is controlled by the magnitude of the regularization parameter λ. Two important assumptions are considered in this study: (1) that most of the observed SV is due to a large-scale flow interacting with the magnetic field and (2) that this flow is the surface expression of incompressible columnar convection inside the core. To impose condition (1) an ℓ^3 norm that minimizes flow gradients was used (Gillet *et al.* 2009). As to condition (2), it requires the flow at the CMB to be equatorially symmetric and verifying $\nabla_H \cdot (\mathbf{u}\cos^2\theta) = 0$ (e.g. Pais & Jault 2008; Schaeffer & Pais 2011; Amit & Pais 2013). This second condition is imposed using a very high λ multiplier, such that deviations are zero to the machine precision. The regularizing parameter multiplying the ℓ^3 norm, denoted λ_R, is allowed to change from one epoch to the other, though, in order for the normalized misfit to be nearly 1 at each epoch. For each inversion the following condition is used to select an acceptable λ_R (Pais *et al.* 2004):

$$0.8 < \sqrt{\frac{\chi^2}{N}} < 1.2, \tag{A7}$$

where

$$\chi^2 = \left(\mathbf{A}(\mathbf{b})\,\mathbf{m} - \dot{\mathbf{b}}\right)^T \mathbf{C}_e^{-1} \left(\mathbf{A}(\mathbf{b})\,\mathbf{m} - \dot{\mathbf{b}}\right) \tag{A8}$$

is the normalized misfit and N is the total number of SV coefficients.

A3 The inversion algorithm

In this study, the iterative approach of Pais & Jault (2008) was used, whereby the effects of the small-scale magnetic field are computed for each step i of the flow calculation and the result is assimilated as a modelling error into the covariance matrix $\mathbf{C}_e^i$. The procedure is repeated iteratively until a convergence is met, when either the modelling error or the flow solution no longer suffer significant changes. At this point, the found solution is consistent with the error covariance matrix $\mathbf{C}_e$ used to invert it. Another iterative search is made with iteration index j, within each step i, in order to find the regularization parameter λ_R that provides a flow solution $\mathbf{m}$ satisfying (A7).

The whole procedure can be described in the following way, for each epoch. Each step i of the cycle involves three intermediate flows $\mathbf{m}^{i,0}$, $\mathbf{m}^{i,1}$ and $\mathbf{m}^{i,2}$ in four intermediate steps within step i, denoted (A) to (D):

(A) $\mathbf{m}^{i,0}$ is used to compute a first estimate $\mathbf{C}_{\mathbf{e}}^{i,0}(\mathbf{m}^{i,0})$ of the error covariance matrix, as explained in Section A1;

(B) $\mathbf{C}_{\mathbf{e}}^{i,0}$ is used in eq. (A3) to invert for $\mathbf{m}^{i,1}$ and from this a new error covariance matrix $\mathbf{C}_{\mathbf{e}}^{i,1}(\mathbf{m}^{i,1})$ is computed as in previous item;

(C) $\mathbf{C}_{\mathbf{e}}^{i,1}$ is used to invert for $\mathbf{m}^{i,2}$ and $\mathbf{m}^{i,2}$ is used to compute $\mathbf{C}_{\mathbf{e}}^{i,2}(\mathbf{m}^{i,2})$;

(D) The flow solution resulting from step i is $\mathbf{m}^{i} = (\mathbf{m}^{i,1} + \mathbf{m}^{i,2})/2$, which is used as $\mathbf{m}^{i+1,0}$ in the next step.

To monitor the convergence of flow solutions $\mathbf{m}^{i,1}$ and $\mathbf{m}^{i,2}$ and of covariance matrices $\mathbf{C}_{\mathbf{e}}^{i,1}(\mathbf{m}^{i,1})$ and $\mathbf{C}_{\mathbf{e}}^{i,2}(\mathbf{m}^{i,2})$ as i increases, they are compared using as diagnostic parameters the L^2 norm of the vector difference $\mathbf{m}^{i,2} - \mathbf{m}^{i,1}$ and the Frobenius norm of the matrix difference $\mathbf{C}_{\mathbf{e}}^{i,2}(\mathbf{m}^{i,2}) - \mathbf{C}_{\mathbf{e}}^{i,1}(\mathbf{m}^{i,1})$.

Each of the two flow inversions made at step i is done according to a second (internal) iterative cycle with index j, whereby the found solution must verify condition (A7):

$j = 0, 1$: Two trial solutions are first computed from (A3), using $\lambda_{R,0}$ and $\lambda_{R,1} = \lambda_{R,0} + \Delta\lambda_0$. Computing χ^2 for each of these two inversions, makes possible to estimate the derivative of χ^2 with respect to λ. Then a Newton–Raphson-type algorithm is applied to choose $\Delta\lambda_1$ such that $\lambda_{R,2} = \lambda_{R,1} + \Delta\lambda_1$ makes χ^2 approach the due condition (A7).

$1 < j < J$: For each new value $\lambda_{R,j} = \lambda_{R,j-1} + \Delta\lambda_{j-1}$ a new inversion using (A3) is made and the corresponding χ^2 computed. A Newton–Raphson-type algorithm is applied to choose $\Delta\lambda_j$.

$j = J$: The cycle ends at iteration J, when $\lambda_{R,J}$ verifies condition (A7). This final estimate gives λ_R of the inversion.

In the first of all j-cycles, we used in our calculations $\lambda_{R,0} = 1.0 \times 10^6$ and $\Delta\lambda_0 = \lambda_{R,0}/4$, from previous tests. In all other cycles, the converged λ_R from the previous cycle is used.

Finally, the two particular i-cycle steps are:

$i = 0$: The initial flow $\mathbf{m}^{0,0}$ is computed from (A3), using as error covariance matrix $\mathbf{C}_{\mathbf{e}} = \mathbf{C}_{\mathbf{e}}^{mod} + \mathbf{C}_{\mathbf{e}}^{r}$, where only diagonal elements are considered for $\mathbf{C}_{\mathbf{e}}^{r}$, given by $\sigma_{\mathbf{e}}^{r}(\ell)^2 = 36\exp(-\ell)$ as in, for example, Schaeffer & Pais (2011).

$i = I$: Convergence is met for $i = I$ if $\lambda_R^{I,1} = \lambda_R^{I,2}$ AND $\chi^{I,1}(\mathbf{m}^{I,1}) = \chi^{I,2}(\mathbf{m}^{I,2})$ (see eq. A8), within a certain precision imposed by A7. These two conditions ensure that the inversion is well converged, as confirmed by very small values of root squares of the L^2 norm of flow differences and of the Frobenius norm of error covariance matrices differences (10^{-7} to 10^{-8} and 10^{-3} to 10^{-5}, respectively.)

Fig. A1 shows the Lowes–Mauersberger power spectrum at the Earth's surface of the SV of model 'COV-OBS', together with predictions from our inverted flows and different errors referred above: misfit errors, SV prior data errors from $\mathbf{C}_{\mathbf{e}}^{COV\text{-}OBS}$ and modelling errors from $\mathbf{C}_{\mathbf{e}}^{r}(\mathbf{m})$. For the two last cases, although the covariance matrices are dense, only diagonal elements are used. Besides, rotationally invariant errors are calculated by averaging the diagonal elements (variances) within each degree ℓ. Along the time interval from 1840 to 1990 the prior data errors decrease as the quality and amount of data improve, but the modelling errors are always of the same order of magnitude with very similar spectra. It is worth noting that the prediction error for epoch 1850 is very similar to the corresponding error obtained by Aubert (2014, see his fig. 1), in spite of a more involved approach whereby the error covariance matrix to use is obtained from an ensemble of direct numerical simulations of the geodynamo.
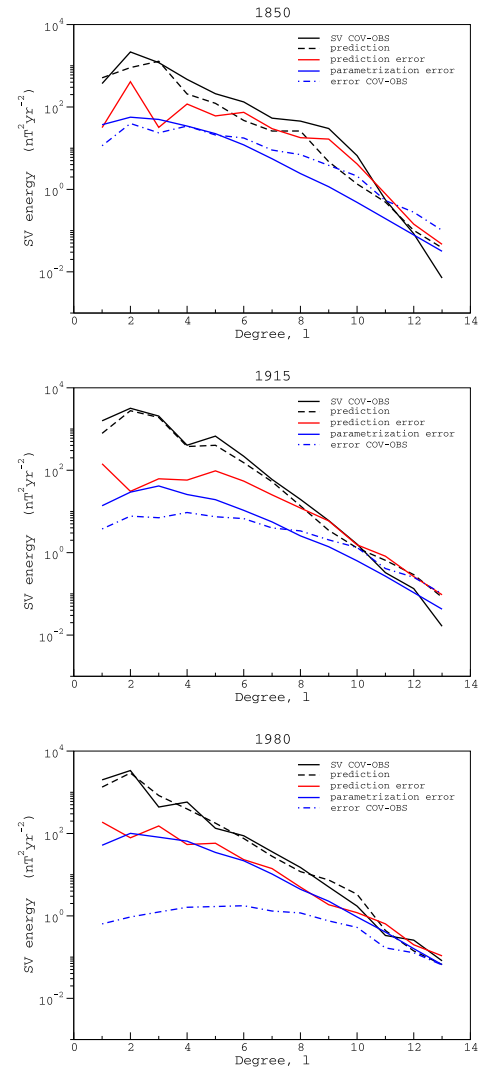
Figure A1. The Lowes–Mauersberger power-spectrum at the Earth's surface of: the SV of model 'COV-OBS' (black solid line), flow predictions (black dashed line), misfit errors (red line), SV prior data errors from $\mathbf{C}_{\mathbf{e}}^{COV\text{-}OBS}$ (blue dashed line) and modelling errors from $\mathbf{C}_{\mathbf{e}}^{r}(\mathbf{m})$ (blue solid line).

When dealing with the two different geomagnetic field models 'gufm1' and 'COV-OBS', the main differences come from the fact that we used the diagonal $\mathbf{C}_{\mathbf{e}}^{gufm}$ matrix from Jackson (1997) and the dense $\mathbf{C}_{\mathbf{e}}^{COV\text{-}OBS}$ matrix provided by Gillet *et al.* (2013). The total matrices $\mathbf{C}_{\mathbf{e}}$ are always non-diagonal, contrary to most previous inversions, thus allowing to take into account the fact that errors in the SV coefficients are correlated.

For 'COV-OBS', this correlation comes in part from the geomagnetic field model computation, but the most important contribution comes from the parametrization errors iteratively computed during the inversion as explained in Section A1.

APPENDIX B: ESTIMATION OF SIGNIFICANCES FOR IDENTIFIED MODES

A test on the covariance matrix spectrum allows to assess the uncertainty of each eigenvalue of the covariance matrix $\mathbf{C}_X$ (Section 3.1) or each singular value of $\mathbf{C}_{X_1 X_2}$ (Section 3.2), that reflects the fact

that the sampling EOF's and PC's differ from the asymptotic ones (when $N_e \to \infty$). Following North *et al.* (1982), the sampling error in the eigenvalue λ_i of $\mathbf{C}_X$ is, to first order,

$$\delta\lambda_i \sim \lambda_i \sqrt{\frac{2}{N}}, \tag{B1}$$

where N is the number of independent realizations or the number of degrees of freedom, which we take equal to N_e. Then, two consecutive eigenvalues λ_i and λ_j are considered degenerate if their difference $\Delta\lambda = \lambda_i - \lambda_j$ is such that

$$\Delta\lambda \leq \lambda_i \sqrt{\frac{2}{N_e}}. \tag{B2}$$

This is known as North's rule of thumb. When two (or more) eigenvalues of $\mathbf{C}_X$ are degenerate it means they cannot be clearly distinguished and the corresponding EOFs patterns cannot be considered independent (Hannachi *et al.* 2007). The same rule of thumb can be used to decide on the separation of consecutive singular values of the cross-covariance matrix when applying SVD (e.g. Venegas *et al.* 1997).

Temporal functions are compared using the correlation coefficient r (e.g. Von Storch & Navarra 1995), which measures the extent to which there is a linear relationship between the two functions. It always takes values in the interval $[-1, 1]$, the two extreme values meaning there exists a perfect linear relationship between the two compared functions. For $\mathbf{t}_1$ a column vector representing a certain time-series and $\mathbf{t}_2$ a column vector representing another time-series, with the same dimension,

$$r(\mathbf{t}_1, \mathbf{t}_2) = \frac{\left(\mathbf{t}_1 - \overline{\mathbf{t}_1}\right)^{\mathrm{T}} \left(\mathbf{t}_2 - \overline{\mathbf{t}_2}\right)}{\sqrt{\left(\mathbf{t}_1 - \overline{\mathbf{t}_1}\right)^{\mathrm{T}} \left(\mathbf{t}_1 - \overline{\mathbf{t}_1}\right)} \sqrt{\left(\mathbf{t}_2 - \overline{\mathbf{t}_2}\right)^{\mathrm{T}} \left(\mathbf{t}_2 - \overline{\mathbf{t}_2}\right)}}, \tag{B3}$$

where $\overline{\mathbf{t}_1}$ and $\overline{\mathbf{t}_2}$ are the average values of the two series over the time period considered.

To estimate the level of similarity between two different spatial structures, we use congruence coefficients, g (e.g. Harman 1976). If $\mathbf{s}_1$ is a column vector representing a given spatial structure and $\mathbf{s}_2$ a column vector for some other spatial structure, both with the same dimension, then

$$g(\mathbf{s}_1, \mathbf{s}_2) = \frac{\mathbf{s}_1^{\mathrm{T}} \mathbf{s}_2}{\sqrt{\mathbf{s}_1^{\mathrm{T}} \mathbf{s}_1} \sqrt{\mathbf{s}_2^{\mathrm{T}} \mathbf{s}_2}}. \tag{B4}$$

Congruence coefficients have the same range as correlation coefficients, that is from -1 (negative similarity) to $+1$ (positive similarity). However, as noted by Richman & Lamb (1985) who uses this parameter in the context of PCA analysis, the congruence coefficient measures not only the pattern similarity but the magnitude similarity as well. This is because, unlike the correlation coefficient, it does not subtract the mean value of the spatial patterns to compare.

Finally, a Monte Carlo (MC) approach is used to estimate the significance of the correlation coefficients r for each gridpoint shown in correlation maps (see Fig. 7). To this end, the two original time-series to be correlated are shuffled, one at a time, using a technique called bootstrapping with moving blocks. The theoretical discussion of the method can be found in Kunsch (1998) and Liu & Singh (1992). Here, we use it in a similar context as in several studies on coupled atmosphere–ocean variability (e.g. Wallace *et al.* 1992; Peng & Fyfe 1996; Venegas *et al.* 1997). The method is used for resampling and is applicable to weakly dependent stationary data, that is, data which are nearly independent if far apart in time. There is nonetheless a correlation time to consider and that should be kept in resampled as in original data, that is, the autocorrelation structure of the original series should be preserved. The reason is that shuffling the data in time randomly would increase the temporal degrees of freedom, giving in the end lower p-values for the statistical tests and improving artificially the significance of estimated parameters. So, in order to avoid this effect, the shuffling in time is applied not to every single data point randomly, but to a number of blocks containing the original data. The length of the blocks (b) depends on the autocorrelation structure of the original data. For a data set x_i of total time length N_e, a number of blocks $N_e - b + 1$ are created so that the data from x_1 to x_b will be in block 1, the data from x_2 to x_{b+1} will be in block 2, etc. After creating all the blocks, N_e/b of these blocks are randomly selected to create a new pseudo-time-series x_i', from which a certain quantity to test may be recalculated (if N_e is not an integer multiple of b, x_i' is truncated to have the same length as x_i). The length b of the blocks was chosen after inspection of the flow data autocorrelation function in each point of the spatial grid. Having verified that the time lag at which the autocorrelation function becomes zero or very closely zero does not exceed 55 yr, this time interval was therefore used as the time-length of the moving blocks. For the significance tests on correlation maps produced from the SVD of coupled flow[gufm1] and flow[COV-OBS], each of the MC runs included two subruns. During the first subrun the first of the series to correlate (e.g. the $\overline{PC}_i$ time-series for SVD mode i, see Section 3.2) was shuffled and the resulting artificial series was correlated in each gridpoint with the second series which is the observed time-series from flow[gufm1] or from flow[COV-OBS]. During the second subrun the shuffling in time procedure was applied to the flow data matrix (flow[gufm1] or flow[COV-OBS], respectively), and the resulting resampled data series for each gridpoint were then correlated with $\overline{PC}_i$. The main goal is always to break the chronological order of one field relative to the other. For each MC run, the mean of the two artificial correlation coefficients for each gridpoint obtained from the two subruns was calculated. After a large number of MC runs (2×1000), the number of the correlation coefficients of magnitude higher than the original r divided by the total number of the runs is the p-value of this specific r. This p-value gives the probability to obtain this specific r just by chance. For instance, a p-value $= 0.1$ associated to a certain correlation coefficient, means that the probability to obtain such correlation coefficient for this specific pair of correlated time-series just by chance is only 10 per cent.