Geophysical Journal International

Geophys. J. Int. (2021) **225,** 1086–1109
Advance Access publication 2020 December 12
GJI Geomagnetism, Rock Magnetism and Palaeomagnetism

doi: 10.1093/gji/ggaa533

Modelling diurnal variation magnetic fields due to ionospheric currents

G.D. Egbert [1], P. Alken,[2] A. Maute[3] and H. Zhang[1]

[1] *College of Earth Ocean and Atmosphjeric Sciences, Oregon State University, Corvallis, OR,* 97331*, USA. E-mail: egbert@coas.oregonstate.edu,*
[2] *Cooperative Institute for Research in Environmental Scienes, University of Colourado, Boulder, CO,* 80309*, USA*
[3] *High Altitude Observatory, National Center for Atmospheric Research, Boulder, CO,* 80301*, USA*

Accepted 2020 November 5. Received 2020 October 8; in original form 2020 July 20

SUMMARY

Accurate models of the spatial structure of ionospheric magnetic fields in the diurnal variation (DV) band (periods of a few hours to a day) would enable use of magneto-variational methods for 3-D imaging of upper mantle and transition zone electrical conductivity. Constraints on conductivity at these depths, below what is typically possible with magnetotellurics, would in turn provide valuable constraints on mantle hydration and Earths deep water cycle. As a step towards this objective, we present here a novel approach to empirical modelling of global DV magnetic fields. First, we apply frequency domain (FD) principal components analysis (PCA) to ground-based geomagnetic data, to define the dominant spatial and temporal modes of source variability. Spatial modes are restricted to the available data sites, but corresponding temporal modes are effectively continuous in time. Secondly, we apply FD PCA to gridded surface magnetic fields derived from outputs of the physics-based Thermosphere–Ionosphere–Electrodynamics General Circulation Model (TIEGCM), to determine the dominant modes of spatial variability. The TIEGCM spatial modes are then used as basis functions, to fit (or interpolate) the sparsely sampled data spatial modes. Combining the two steps, we have a FD model of DV band global magnetic fields that is continuous in both space and time. We show that the FD model can easily be transformed back to the time domain (TD) to directly fit time-series data, allowing the use of satellite, as well as ground-based, data in the empirical modelling scheme. As an illustration of the methodology we construct global FD and TD models of DV band source fields for 1997–2018. So far, the model uses only ground-based data, from 127 geomagnetic observatories. We show that the model accurately reproduces surface magnetic fields in both active and quiet times, including those at sites not used for model construction. This empirical model, especially with future enhancements, will have many applications: improved imaging of electrical conductivity, ionospheric studies and improved external field corrections for core and crustal studies.

Key words: Geomagnetic induction; Magnetic field variations through time; Numerical modelling; Spatial analysis; Time-series analysis.

1 INTRODUCTION

The semi-regular diurnal variations (DV) of Earth's magnetic field result primarily from currents flowing in the ionospheric E-region but also the lower F-region, set up by a combination of the neutral-wind dynamo moving plasma across geomagnetic field lines, and high latitude field-aligned currents. The E-region current is strongly influenced by highly variable upward propagating atmospheric tides, which have less of an effect on F-region current flow. The ionospheric dynamo current also depends strongly on the ionospheric conductivity distribution, which is controlled primarily by solar ionizing radiation in the extreme ultraviolet band, and varies significantly with solar cycle and solar local time. Since the primary forcing of atmospheric tides, neutral winds and ionospheric conductivity are all strongly linked to the DVs in solar radiation the DV magnetic fields are, as the name suggests, quasi-periodic, with a fundamental period of one solar day. However, a number of harmonics (6 or more) are clearly evident as peaks in spectra computed from long magnetic field time-series (Egbert *et al.* 1992; Love & Rigler 2014; Wang *et al.* 2020). Furthermore, interactions with the highly variable solar wind and the underlying atmosphere result in substantial day-to-day variability, so these peaks are broad, not simple lines. We somewhat loosely use the term DV to refer to all of the surface magnetic field variability in

the frequency band $10^{-5} - 10^{-4}$ Hz, roughly 1−6 cycles per day (cpd).

Along with storm-time variations associated with large scale magnetospheric ring and tail currents, and sudden geomagnetic storm effects, for example, from penetrating electric fields, DVs dominate the external component of geomagnetic field temporal variations. An accurate representation of the spatial structure of the DV fields is thus a requirement of modern geomagnetic field models (e.g. Sabaka *et al.* 2015), which for many applications must account for significant external field variations. For example, these are required for increasingly demanding studies of core (Finlay *et al.* 2016; Alken *et al.* 2020) and crustal (Maus *et al.* 2008; Thébault *et al.* 2016; Olsen *et al.* 2017) magnetic fields with satellite magnetic field data, for aeromagnetic surveys, and for many applications where magnetic fields are used for navigation or orientation.

Imaging electrical conductivity of Earth's mantle also requires accurate models of DV fields. The magnetotelluric (MT) method can provide detailed images of conductivity of the lithosphere at near continental scales (e.g. Dong *et al.* 2013; Meqbel *et al.* 2014; Thiel *et al.* 2016), but because it is difficult to measure electric fields at periods much beyond a few times 10^4 s, MT penetration depths are typically limited to a few hundred kilometres. To image deeper into the mantle (e.g. Banks 1969; Olsen 1998; Olsen *et al.* 2003; Kuvshinov & Olsen 2006; Kelbert *et al.* 2009; Velímskỳ 2010; Püthe & Kuvshinov 2013) requires use of magnetovariational (MV) methods, which use magnetic field variation data (no surface electric fields) at periods of hours to days. Unlike MT, which is based on a local impedance (ratio of electric to magnetic fields) which is relatively insensitive to source wavelength, the MV approach requires detailed knowledge of source spatial structure. Thus, the MV approach has most commonly been applied to longer period ($T > 2$ d) D_{st} geomagnetic variations, where a zonal dipole source, representing a simplified distant ring-current, can be reasonably assumed, at least at mid-latitudes (but see Fujii & Schultz 2002; Kelbert *et al.* 2009; Semenov & Kuvshinov 2012). Filling the gap between the MT band (roughly $T < 2 \times 10^4$ s) and the D_{st} band ($T > 2 \times 10^5$ s) requires accurate models of the ionospheric sources in the DV band. As shown in Kelbert *et al.* (2008) this frequency band is critical for resolving conductivity of the upper mantle, from the asthenosphere into the transition zone.

Improved constraints on mantle electrical conductivity in this depth range are potentially of great importance. Laboratory data clearly show that significant quantities of water may be stored in Earths deep interior (e.g. Hirschmann & Kohlstedt 2012), but the actual volume and distribution remain highly uncertain (e.g. Karato 2011). Hydrogen would modify rheological properties (Karato & Jung 1998; Mei & Kohlstedt 2000; Karato & Jung 2003), and melting relationships (e.g. Hirschmann 2006), so reliable geophysical constraints on its concentration in the mantle would have important implications for the dynamics and geochemical evolution of the Earth. Electrical conductivity is uniquely sensitive to water content (Karato 2011), especially for upper mantle and transition zone minerals (Huang *et al.* 2005; Wang *et al.* 2006; Yoshino *et al.* 2008; Dai & Karato 2009), making EM induction methods an ideal probe of upper mantle water content. Improved capabilities for imaging upper mantle conductivity, from the asthenosphere to the transition zone, are a key motivation for the work reported here, which focuses on modelling the ionospheric source fields in the DV band.

To model DV fields for induction studies Schmucker (1999) restricted attention to variations on days of low geomagnetic activity, that is, S_q, with spatial structure defined in terms of a small number of spherical harmonics. Koch & Kuvshinov (2013) and Koyama *et al.* (2014) developed this modelling approach further, using a similar simplified parametrization of S_q, and then using the result to invert for 3-D conductivity in a continental-scale study of Australia. Guzavina *et al.* (2019) estimated source structure through a global analysis of mid-latitude sites, and used these source structures to define transfer functions for individual observatories, which were inverted for local 1-D mantle conductivity profiles. Here we follow a somewhat different approach, following the general scheme developed (and extensively justified) in Sun *et al.* (2015) to estimate long period ($T > 2$ d) sources, which were then used to invert for 3-D deep mantle conductivity.

In our approach, source estimation can be broken into two steps. Because the overall workflow is somewhat involved, we summarize key steps in Fig. 1. Readers may find it useful to refer back to this figure as details are filled in, and symbols are more precisely defined in subsequent sections.

(1) Frequency domain (FD) principal components analysis (PCA, Egbert & Booker 1989; Smirnov & Egbert 2012) is used to define the dominant spatial and temporal modes of variability in the external field field signal and its induced counterpart recorded by a large array of geomagnetic observatories (left two columns of Fig. 1), producing a low-dimensional approximation of the coherent signal. Following arguments given in Egbert & Booker (1989) each PCA spatial mode represents the total (external plus internal) fields that would be observed at the available sites for a fixed, but formally unknown, external source geometry. Temporal modes (which are effectively continuous in time (at the data sample rate) define how the mix of all modes varies (amplitude and phase) in time. In Sun *et al.* (2015) only a single spatial data mode was considered, corresponding primarily to a symmetric ring current source, but allowing for correlated auroral electrojets. Here we use many more modes (20 for each of 13 frequency bands) allowing us to model a large fraction of the variability in the DV band.

(2) A statistical model of the unknown source structure (represented as a covariance on the sphere) is used to define a set of basis functions which can be used to fit data spatial modes—that is, to interpolate the sparsely sampled observatory modes to a set of spatially continuous models of source spatial structure, one for each mode. Thus, these basis functions take the place of the selected spherical harmonics used by Schmucker (1999). The statistical model can also be used to define a regularization term for data fitting. In Sun *et al.* (2015) the basis functions were eigenmodes of a covariance matrix for an *ad hoc* statistical model, based on a superposition of current loops, in quasi-dipole (QD) coordinates (Richmond 1995), concentrated near the auroral zones and the magnetic equator, with random tilts and latitudes. The computed modes allowed modelling of complex source spatial structure in the auroral zone, as well as some deviation from a perfectly zonal dipole at mid latitudes. For the DV modelling we use outputs from the physics-based Thermosphere–Ionosphere–Electrodynamics General Circulation Model (TIEGCM; Richmond *et al.* 1992; Roble & Ridley 1994), to define the statistical model used to constrain source estimates. Basis functions used for source estimation are derived from FD PCA of model outputs, following the same approach applied to the data (rightmost column of Fig. 1). To the extent that the TIEGCM captures the dominant modes of spatial variability in the ionosphere, the dominant model modes should be efficient for modelling real source structure.

Our analysis approach is firmly rooted in the FD, consistent with our ultimate objective, imaging 3-D variations in mantle

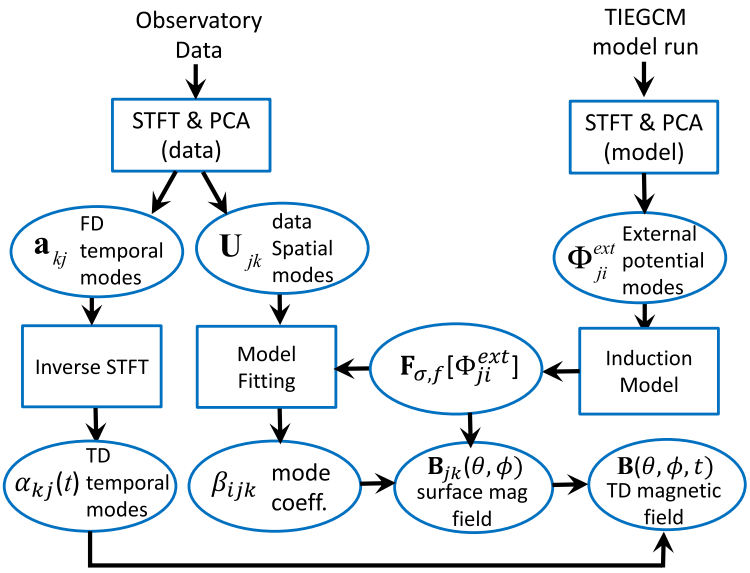

Figure 1. Overview of workflow used in this paper for modelling diurnal variation magnetic fields. Notation is summarized in Appendix A. Inputs include time-series for the relatively sparse set of geomagnetic observatories (left-hand side) and magnetic field time-series on a regular grid, derived from TIEGCM outputs (right-hand side). Outputs are FD and TD models of DV geomagnetic field variations for the time window where observatory data are available.

conductivity. Methods for 3-D EM inversion are by far best developed (and most efficient) in the FD. In fact, there are other potential advantages to a FD approach to modelling DV fields. The DV fields are generated by currents that remain, to first order, fixed relative to the sun. In the frame of the rotating Earth relevant to induction studies, magnetic variations have fundamental periods that are fractions of the solar day, that is, frequencies of 1, 2, 3, etc., cpd. Furthermore, spatial structure and frequency are closely linked—for example at m cpd the spherical harmonic Y_m^{m-1} is dominant (Schmucker 1999), so we can anticipate that the optimal set of spatial basis functions for fitting source fields will depend on frequency. Our FD approach naturally accounts for this.

However, as Fig. 1 (left-hand column) illustrates, the temporal modes can be represented in both frequency and time domains (TD). The FD model (in the figure this is represented by $\mathbf{a}_{kj}$ and $\mathbf{B}_{jk}(\theta,\phi)$) that is the most direct output of our analysis, can be easily converted to a TD model $\mathbf{B}(\theta,\phi,t)$. Note that temporal variations are derived from the data PCA temporal modes, and this data set will determine the time window of validity. TIEGCM is only used to define spatial structure. This TD model will have many other applications beyond induction modelling, including fundamental studies of the ionosphere and improving external field corrections for mapping crustal and core geomagnetic fields. Perhaps most importantly, the TD formulation yields an observation equation that will allow satellite data to be incorporated into our modelling scheme. This leads to a novel approach to merging ground and satellite magnetic data, and outputs of a physics based model such as TIEGCM, into a model of time varying magnetic fields. Our development is specifically focused on DV periods and external source fields associated with ionospheric currents, but the general scheme could be readily extended to treat a broader period range of external magnetic fields.

In the next section we review the PCA approach to data analysis, and present results of application to 21 yr (1997–2018) of hourly mean data from 127 geomagnetic observatories. In Section 3, we provide a brief description of the TIEGCM model and specific run parameters, and then discuss the computation of the spatial basis functions through PCA of model outputs. In Section 4, we combine data and model PCA results to estimate external source fields associated with each data spatial mode. In Section 5, we show how the FD model can be transformed to the TD, and discuss how, with this formulation, satellite data might be incorporated into the source estimation scheme. In Section 6, we present our DV source model, and compare model predictions to observatory time-series, including some not used in the modelling, and in the final section we discuss some extensions and areas for further work.

2 DATA PROCESSING

Our data processing approach builds on the FD PCA methods originally proposed for analysis of EM array data by Egbert & Booker (1989) and Egbert (1989), and developed further in Egbert (1997a, 2002) and Smirnov & Egbert (2012). Here we briefly review basic ideas, and set notation for subsequent developments. All notation is summarized in Appendix A. For an array of N_s three-component magnetic sites, the total number of data channels is $N_c = 3N_s$. We transform the N_c time-series to the FD with a windowed short-time Fourier transform (STFT), resulting in FD data X_{cfn}, where site/component is indexed by c, frequency by f, and time (i.e. sequence number for the FT analysis windows), n. The vector containing all N_c Fourier coefficients (FCs) for each frequency f and window n is denoted as $\mathbf{X}_{fn}$. In our work here, the FT is done in a way that allows seamless transition back to the TD (i.e. $\mathbf{X}_{fn} \rightarrow \mathbf{X}(t)$); details are provided in Appendix B.

The goal of PCA is a reduced dimension representation of the FD vectors $\mathbf{X}_{fn}$. The analysis is done independently for each of a series of J bands, consisting of one or more nearby frequencies f. To be explicit, we denote the set of frequencies f in each band $j = 1, \ldots, J$ by $f \in B_j$ with $P_j = |B_j|$, the number of frequencies in band j. Here we will assume that the bands are non-overlapping, and cover the full frequency range of interest. For each band B_j, PCA approximates the FD data vectors as a sum over products of spatial

($\mathbf{U}_{jk}$) and temporal (a_{kfn}) components, or modes, indexed by k,

$$\mathbf{X}_{fn} = \sum_{k=1}^{K_j} \mathbf{U}_{jk} a_{kfn} + \mathbf{e}_{fn}, \quad f \in B_j. \tag{1}$$

Here $\mathbf{e}_{fn}$ represents all sources of incoherent noise, as well as unmodelled signal due to truncation of the expansion to $K_j \ll N_c$ terms. As the notation indicates, the truncation level may depend on the band. We refer to the N_c-dimensional column vectors $\mathbf{U}_{jk}$, $j = 1, \ldots, J$, $k = 1, \ldots, K_j$ as the *spatial modes*. The components of these vectors correspond (exactly, in the idealized case of noise-free data) to total (external plus internal) magnetic fields that would be observed at the N_s observation sites for some fixed, but imprecisely known source geometry (Egbert 1989). In subsequent steps (described in Sections 3 and 4) we will estimate source geometry for each mode. Plots of the magnetic field vectors on a map of site locations are already informative about external field source geometry, but note that these also contain information about induced components of the field. In the source estimation procedures developed below we will account for these internal fields, at least approximately. The N-dimensional row vectors (a_{kf1}, ...a_{kfN}) define the variations in amplitude and phase of spatial mode k, frequency f for all time windows. Combining vectors for all P_j frequencies $f \in B_j$ over all time windows we obtain the NP_j-dimensional complex vectors $\mathbf{a}_{kj}$, which we will refer to as the *temporal modes*. Thus, for each band j there are K_j modes, each with spatial and temporal components. The spatial modes are constant across a band of nearby frequencies (i.e. depend only on j), but for the temporal mode vectors there are distinct components for each frequency in the band (i.e. these depend on $f \in B_j$), representing variations in amplitude and phase of different frequency components in the band for each segment.

The reduction in dimension (from N_c data channels, to K_j modes) offers some minor advantages in terms of computational efficiency, but there are other, more important benefits. First, PCA allows us to focus on a few statistically dominant source modes, which are *a priori* most likely to have relatively simple spatial structure, to exhibit global symmetries, and to thus be more easily modelled. The highest amplitude modes are also least effected by noise. Secondly, the factorization into products of temporal and spatial modes provides a framework for constructing a spatially and temporally continuous model, which can merge data from different eras, and as we shall discuss in Section 5, to incorporate satellite data. Note that the total dimension of the model is $\sum_j K_j$, so reducing the number of bands, if appropriate, can further simplify the description of spatial structure, and ultimately reduce the total number of source parameters to estimate.

The simplest approach to FD PCA (for a single frequency band j) is based on singular value decomposition (SVD) of the $N_c \times NP_j$ data matrices $\mathbf{X}_j$, with columns $\mathbf{X}_{fn}$, $f \in B_j$, $n = 1, \ldots, N$. It is also possible to formulate PCA in terms of the eigenvalues and eigenvectors of the spectral density matrix $(NP_j)^{-1}\mathbf{X}_j\mathbf{X}_j^{\dagger}$ (Egbert & Booker 1989), where the superscript † denotes complex conjugate transpose. However, these simple approaches only work with complete (or nearly complete) arrays of synoptic data. The geomagnetic observatory database is filled with gaps, with stations coming and going over the years (Figs 2 and 3). We thus adopt the missing data EM PCA (MsDEMPCA) scheme described in Smirnov & Egbert (2012) (see also Egbert 1997a), which incorporates robust statistical estimation methods, and allows for even large blocks of missing data. For completeness this approach to PCA is sketched in Appendix C. As discussed there, MsDMEPCA provides estimates of incoherent noise (or more precisely, residual) power levels for each channel, which we will use in the following.

Data analysed for this paper are hourly mean values from the compilation of observatory data of Macmillan & Olsen (2013), for years 1997–2018. Locations for all observatories with some data in this time window (182 total) are plotted in Fig. 2 (all symbols). The density of sites is of course very high in Europe, and in parts of Asia, and low in the Southern Hemisphere. Coverage over the oceans is particularly sparse. As the leading modes obtained from PCA by definition maximize total signal variance (summed over all sites), regions with high site density will tend to be overrepresented in the PCA modes. In an effort to obtain modes which are more spatially balanced, and more representative of the dominant magnetic variations over the whole globe, we use a more uniformly spaced subset of 127 sites (red circles in Fig. 2). To construct the thinned set of observatories, we first eliminated all Northern Hemisphere sites with fewer than 5 yr of data, then eliminated sites based on proximity to other sites, preferentially retaining sites with the highest quality (and most complete) data. No sites in the Southern Hemisphere, or in ocean basins, were eliminated, even though some of these had time-series shorter than 5 yr. Using all sites, but in a weighted fitting scheme (comparable to Shore *et al.* 2016) produced similar results. We adopt the thinning approach, as this leaves us some sites for model validation. Although the reduced array is more uniform, there remains some bias towards the Northern Hemisphere, and even more towards continents, especially when considering that a number of Southern Hemisphere and ocean sites have relatively short time-series. The latitudinal and temporal distribution of available data is shown for the set of $N_s = 127$ sites in Fig. 3. Clearly, allowing for missing data, as with the MsDEMPCA scheme, is required for multivariate array analysis of this data set.

Time-series for the results presented here were Fourier transformed using 8-d-long (192 samples) segments, multiplied by a modified Hanning window. Since our focus here is primarily on estimation of ionospheric source fields in the daily variation band, before transforming we subtracted from the magnetometer data the external and corresponding induced ring current component of the CHAOS-6 model (Olsen *et al.* 2014; Finlay *et al.* 2016), and then removed low frequency signals with a high-pass filter with 0.5 cpd cut-off. As described in Appendix B, the high pass filter is applied before windowing. The windows overlapped by 50 per cent (4 d) so there are roughly 90 time windows in 1 yr of data, resulting in $N = 1977$ segments in the full 21 yr period analysed. The nominal frequency resolution of the spectrum is 1/8 cpd (1.447×10^{-6} Hz), but given the relatively extreme tapering used, actual bandwidth is close to a factor of 2 larger.

First we do the PCA using single frequency bands (i.e. using $P_j = 1$) to obtain results with a nominal frequency resolution of 1/8 cpd. We restrict this analysis to frequencies from 0.5 cpd (period $T = 1.7 \times 10^5$ s) to 6 cpd ($T = 1.4 \times 10^4$ s), resulting in a total of $J = 48$ bands. The number of modes estimated [K_j in (1)] was set to 20 for all bands. In fact, the optimal number of modes to use is difficult to determine. Including more modes leads to a more complete description of the data, but as we shall see in Section 4, lower amplitude modes are noisier, and typically contain smaller scale features which are not so easily modelled, so additional modes might not improve overall reliability of the model. Furthermore, different numbers of modes could (and almost certainly should) be used for different bands. A careful examination of these issues is beyond the scope of this paper. The resulting PCA spectrum is shown in Fig. 4. Each curve represents the signal-to-noise ratio (SNR) for one mode (derived as described in Appendix C) for

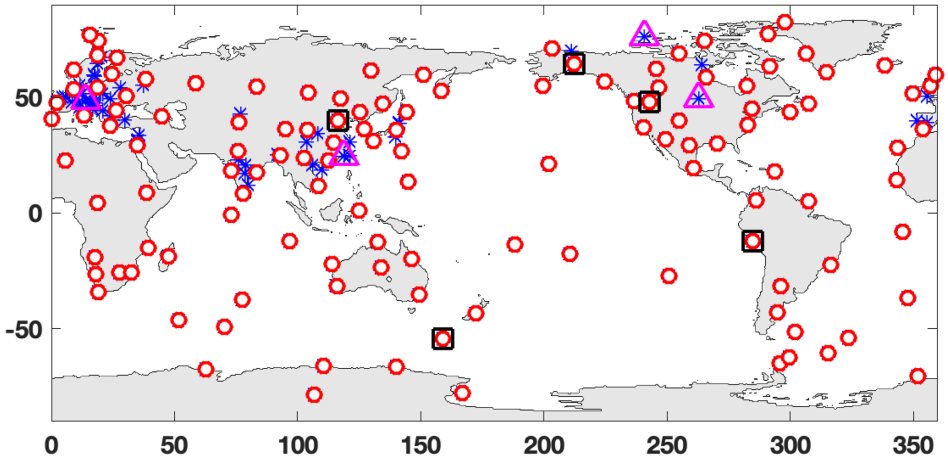

Figure 2. Geomagnetic observatories from the compilation of (Macmillan & Olsen 2013). Red symbols indicate the more evenly spaced subset of sites used for our analyses. Black squares and magenta triangles indicate sites used in Section 6 to illustrate and evaluate DV model.

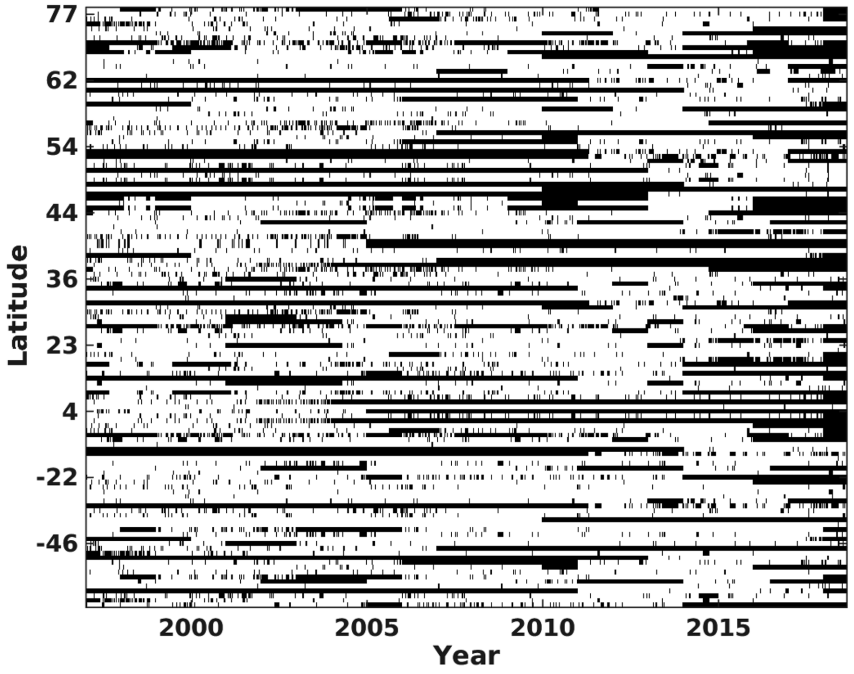

Figure 3. Temporal distribution of data available for observatories shown in Fig. 2. Black indicates missing data. Sites are organized by geographic latitude—but note that the *y*-axis scale is non-uniform, due to the strong bias towards northern latitudes.

all frequency bands. For example, a mode at 30 dB has power (summed over all channels) 1000 times background noise levels. As Fig. 4 shows, all 20 modes estimated are well above 0 dB. There are clear peaks at 1 cpd and the first four harmonics (to 5 cpd), with the highest (in terms of SNR) at 2 cpd. The peaks have a width at the base of roughly 0.5 cpd. However, given the spectral resolution (roughly 0.25 cpd) implied by the 8-d-Hanning window, actual peaks may be somewhat narrower. Wang *et al.* (2020) applied PCA to an array of 36 magnetic sites covering most of China, using longer (16 d) windows, with a modest taper (so spectral resolution was approximately 1/16 cpd in this case). Peaks at DV harmonics were indeed narrower, roughly 0.25 cpd, as further confirmed by tests with longer analysis windows. This corresponds to a typical modulation timescale for the daily variations of roughly 4 d.

Incoherent noise variances generally vary smoothly with period (albeit with subtle peaks at DV harmonics), so the enhancements of SNR at DV periods reflects an increase of signal power (relative to the background continuum) by roughly 10–15 dB (for 1–3 cpd). The cumulative fraction of power explained by 1, 2, 5, 10 and 20 modes is shown in Fig. 5(a). For the main DV peaks (1–3 cpd) roughly 90 per cent of the total variance is explained by the first five modes; 20 modes explain 95–98 per cent of the variance at these periods. At intermediate periods only 80–90 per cent of the signal variance is explained even with the full set of 20 modes. Thus, the global signal is simplest, and most readily modelled with a small number of modes, in the DV bands. However, since the power in these bands is much greater, the full PCA model (20 modes for all bands) explains close to 95 per cent of the total variance over the frequency band analysed.

The 48 single frequency bands allow us to define the DV peaks clearly standing above background signal levels (Figs 4 and 5a), but reduced resolution (with fewer total modes, and hence source

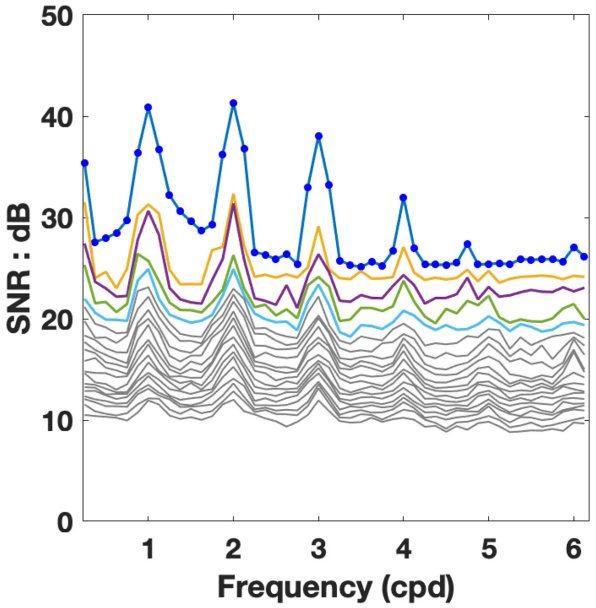

Figure 4. Variance as a function of frequency (cpd), in SNR units, of the 20 estimated modes. First five modes are plotted with different colours for emphasis.

parameters to estimate) will suffice for our purposes. For the remainder of this paper we merge these bands into a reduced set of $J = 13$ wider bands. This effectively increases the number of degrees of freedom in each frequency band, and increases the length of the temporal mode vectors. For example, for the bands centred at 1–6 cpd we include 2 adjacent frequencies, so $P_j = 3$, and the length of the temporal modes increases to $3N = 5931$. Bands with $P_j = 5$ are used for the intermediate (continuum) bands, and two long period band (beyond 1 cpd) uses 3 or 2 frequencies. These frequency bands are summarized in Table 1. The cumulative variance explained by up to 20 modes estimated for these wider bands are shown in Fig. 5(b). Although the picture is less detailed than with narrower bands, results are quite similar, particularly in terms of variance explained. Note that this implies that the same spatial structures can be used to model the data across a wider frequency band, justifying the increase in estimation bandwidth.

A different perspective on signal and noise characteristics is provided in Fig. 6(a). Here signal and noise variances are computed for the full band analysed (i.e. summing over all of the narrower bands of Table 1), but now broken down by magnetic channel (B_x, B_y, B_z in geographic coordinates), and QD latitude. For this computation 10° wide bins were used, combining Northern and Southern Hemispheres, and results were lightly smoothed over latitude. There are systematic variations of signal power with QD latitude, with much larger amplitudes at higher latitudes. There are also systematic variations between different channels, reflecting the main features of the ionospheric current systems, for example, the strong amplification of B_x due to the equatorial (0°) and auroral (65°) electrojets, and the relative enhancement of B_y at mid-latitudes. The dependence of noise variance on QD latitude is similar, but with less variation between channels. This demonstrates that a large fraction of the 'noise' actually represents source complexities not captured in the truncated PCA expansion. Additional modes would be required to capture all details, especially for high latitude current systems. These conclusions are supported further by Fig. 6(b) where signal and noise variances obtained when data are restricted to quiet times ($K_p \leq 1$) are plotted. Unsurprisingly, variances are reduced most significantly at high latitudes. Noise variances are reduced slightly at all latitudes, consistent with the idea that source structures are simpler and easier to model with a small number of PCA modes.

Examples of spatial modes $\mathbf{U}_{jk}$ are displayed in Fig. 7 by plotting the horizontal magnetic field vectors for each observatory at the site location. Real (in phase) and imaginary (quadrature) components of the FD vectors are plotted as blue and red vectors, respectively. At 1 cpd (Figs 7a and b) the signal due to the strong auroral electrojet currents dominate in both of the plotted modes. At mid-latitude the magnetic field vectors vary smoothly in space and reflect the signal of the Sq current; in areas with even moderate site density vectors point in the same direction, with slowly varying phase (or more precisely: polarization). Spatial patterns within mid- and high-latitude zones are quite distinct. At low latitude the signal in the first mode at 1 cpd is enhanced associated with the equatorial electrojet. For the two modes at 1 cpd, spatial patterns in the auroral zones are similar, but relative phases are different, almost reversed. The two modes together thus capture the main components of variability associated with the ionospheric dynamo, and local time effects in auroral electrojet current systems. As usual, distinct (but not necessarily independent) processes are mixed in the individual data modes. Dominant modes for frequency bands centred at the DV harmonics of 2 and 3 cpd are shown in Figs 7(c)–(d). Here the auroral zone is less dominant but still clearly visible. The mostly zonal high latitude current is mainly organized with respect to the magnetic coordinate system, and strong higher order modes exist when expressed in a geographic system. Spatial structure in current systems of SH order m of course map to temporal variations of m cpd. Careful examination reveals significantly more rapid variation of the phase with longitude as the period is reduced. To see this, compare the variations of in-phase (blue) and quadrature (red) vectors across North America or Asia between 1, 2, 3 cpd. At 1 cpd both components are almost constant across a continent, but by 3 cpd there are significant changes from east to west. This is consistent with expectations for DV fields, which are caused by rotation of the Earth under ionospheric current systems that are approximately fixed with respect to the sun. Variations at m cpd are thus expected to be associated with spatial variations in source currents of azimuthal order m. Note that is much easier to see phase shifts in the data modes after these have been interpolated by fitting TIEGCM model modes (see Section 4).

The temporal modes $\mathbf{a}_{kj}$ corresponding to the spatial modes of Fig. 7 are plotted in Fig. 8. For all of the frequency bands plotted (centred at 1, 2, 3 cpd) there are three complex coefficients (corresponding to slightly different frequencies) for each 8-d segment, which because of window overlap are spaced 4 d apart. To make the plot clearer, only the centre coefficient is plotted, as a function of time for the full 21 yr analysed. For the two modes at 1 cpd these variations are modulated by the 11-yr solar cycle associated with the stronger ionospheric current at middle and low latitudes during solar maximum when ionospheric conductivities are on average larger than at solar minimum. The more impulsive signals are consistent with the dominance of auroral current systems evident in the corresponding spatial modes. In contrast, the leading temporal modes for 2 and 3 cpd show very clear seasonal cycles which might be partially driven by the seasonal variation of upward propagating atmospheric tides. Similar to the 1 cpd mode, the strength of 2 and 3 cpd modes are modulated primarily by the solar cycle dependence of the ionospheric conductivities. Differences between temporal modes for 1 cpd, and the harmonics, agree with what we

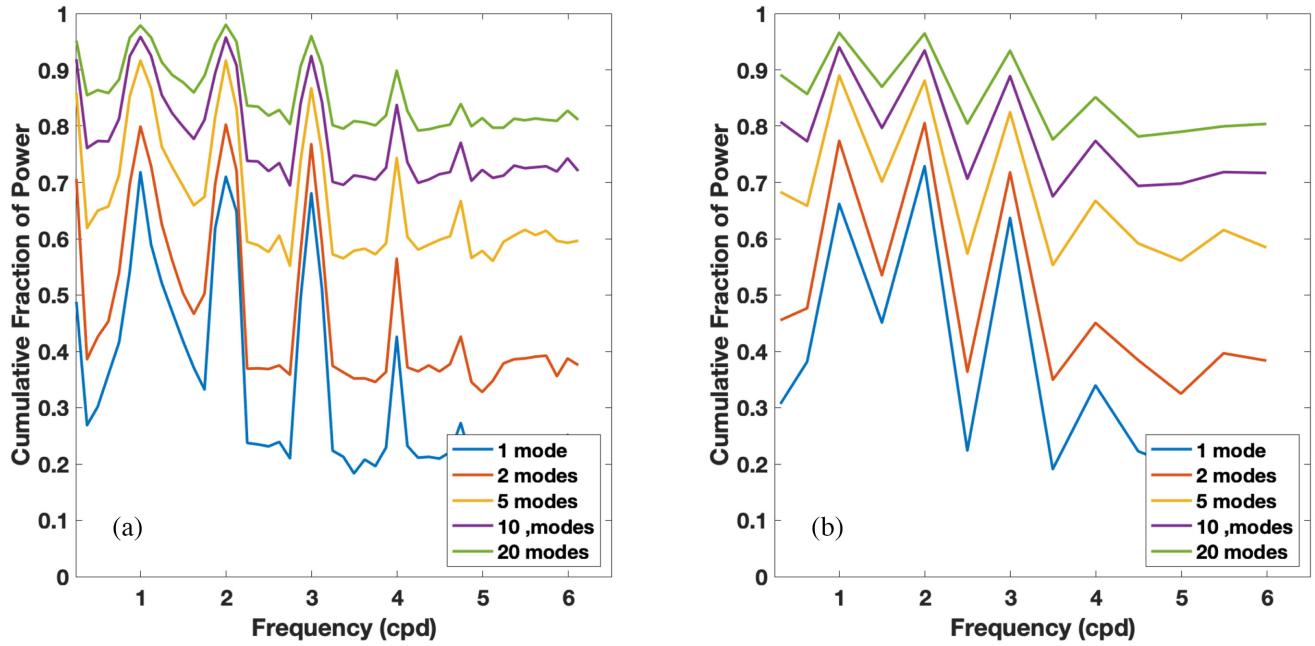

Figure 5. Cumulative fraction of total variance explained as a function of frequency for different values of PC truncation level K, for (a) PCA with 48 single frequency bands, and (b) the 13 bands defined in Table 1. In both plots variances are normalized by estimated noise level, as in Fig. 4.

Table 1. Frequency bands used for PCA, for analysis of both data and TIEGCM model outputs.

Band (j)	Frequency numbers	Centre frequency (cpd)	Period (10^4 s)
1	48–50	6.0	1.44
2	43–47	5.5	1.57
3	40–42	5.0	1.73
4	35–39	4.5	1.92
5	32–34	4.0	2.16
6	27–31	3.5	2.47
7	24–26	3.0	2.88
8	19–23	2.5	3.46
9	16–18	2.0	4.32
10	11–15	1.5	5.76
11	8–10	1.0	8.64
12	5–7	0.625	13.82
13	3–4	0.3125	27.65

see in the spatial modes, reflecting the relative dominance of ionospheric dynamo currents in the 2 and 3 cpd modes, and auroral zone processes in the 1 cpd modes.

3 TIEGCM MODEL MODES

PCA of the geomagnetic array data represents DV band variations in terms of a relatively small number of time/space mode products. While the representation of temporal variation is essentially continuous, spatial structure is only sparsely sampled at the observatory locations. To construct a source model that is continuous in space (as well as time) interpolation of spatial modes is required. Generic sets of basis functions, such as spherical harmonics (SH) up to some fixed degree and order L, are not suitable for the task. If L were chosen large enough for accurate modelling of the spatially complex ionospheric current systems, the large number of free parameters would result in an unstable fitting problem. Thus, most previous efforts to model spatial structure of DV fields (e.g. Schmucker (1999), and references therein) have used a carefully chosen small subset of spherical harmonic basis functions. To account for the fact that DV current systems remain relatively fixed in local time, for variations at p cpd SH with order m near p are selected *a priori*, typically also with only a small range of degrees l. Regression variable selection procedures are then used to reduce the basis further. This is the approach used by Koch & Kuvshinov (2013) and Guzavina *et al.* (2019). Sabaka *et al.* (2002, 2004) used a variant on this approach for parametrizing a climatological model of DV fields, using similar limited subsets of spherical harmonics, but in QD coordinates, which should offer improved representation of small scale details in ionospheric current systems that are controlled by the background geomagnetic field. We use a different approach: basis functions for interpolation are derived from FD PCA of the TIEGCM simulated magnetic perturbations. The physics-based model captures the large scale ionospheric current systems so the dominant modes of model variability should provide an efficient basis for representing realistic ionospheric current systems. Note that TIEGCM uses a realistic model of the geomagnetic field, so possible advantages of using QD (or simply geomagnetic) coordinates are effectively already built in. Note also that the set of basis functions can be represented in the space domain, or through a SH expansion. We will adopt a SH representation for source currents, as this makes transitions between electric currents, magnetic fields, and potential functions straightforward, and in some cases at least (e.g. for a 1-D Earth conductivity), simplifies computation of induced fields.

3.1 TIEGCM

TIEGCM is a self-consistent model of the coupled thermosphere and ionosphere capturing variations from 97 km to approximately 450–600 km altitude depending on solar cycle conditions. The model physics include the dynamics, electrodynamics and chemical interactions of the thermosphere and ionosphere region, as well

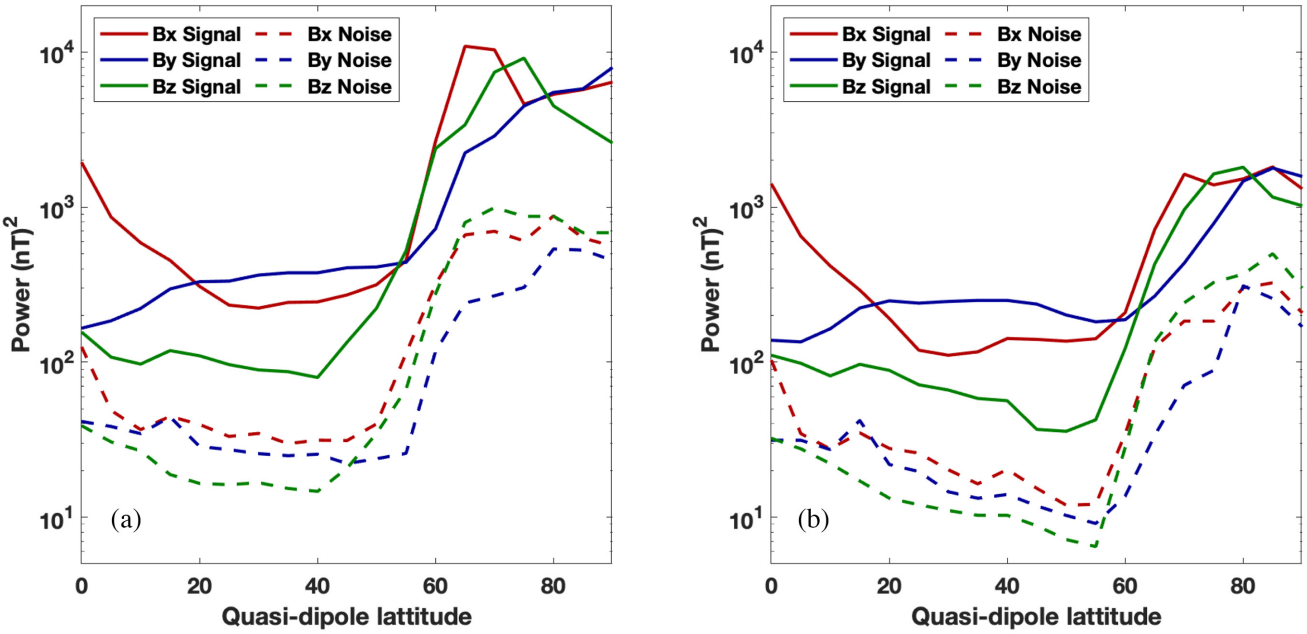

Figure 6. Signal and noise power (integral over the full frequency band considered) for each magnetic component separately, as a function of QD latitude. (a) all data; (b) quiet times, defined by $Kp \leq 1$; see Section 5 for details.

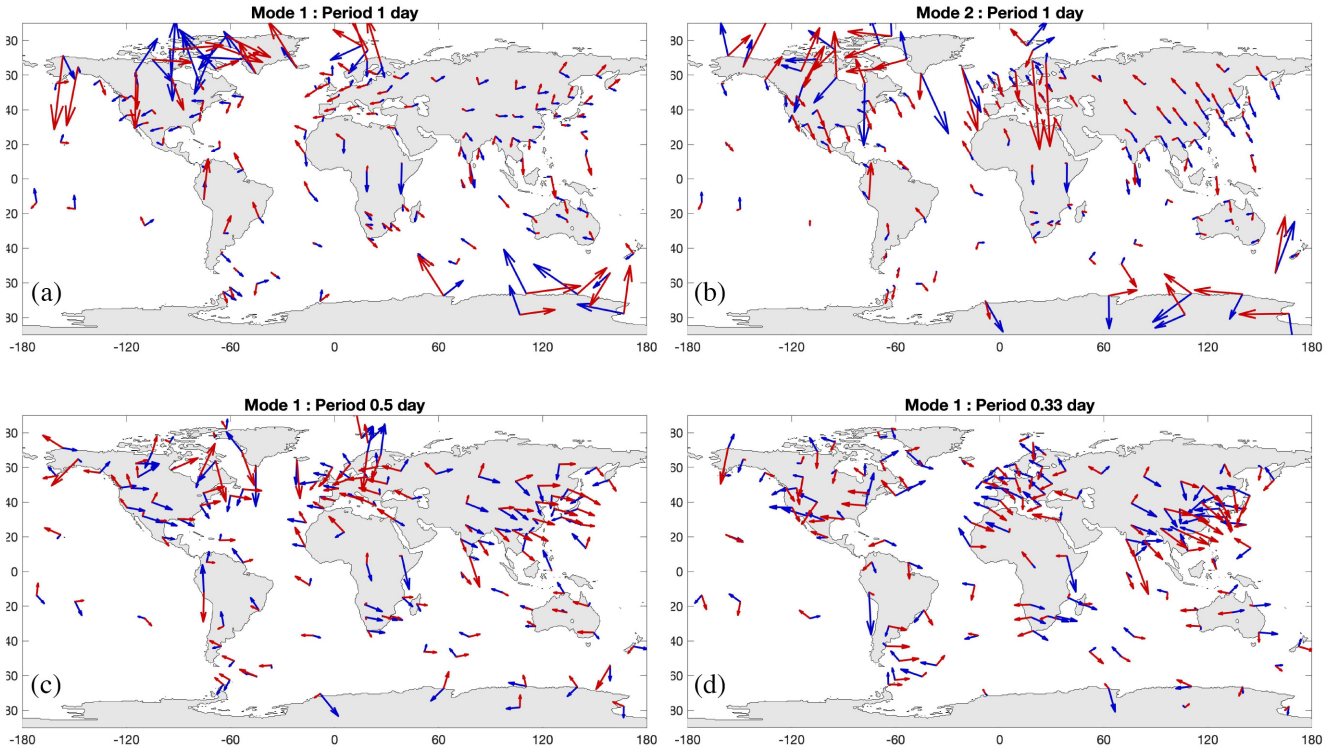

Figure 7. (a and b) First two spatial modes $\mathbf{U}_{jk}$, $k = 1, 2$ for the 1 cpd band. Horizontal magnetic field vectors are plotted on a map of station locations. As discussed in Appendix C, spatial mode components have units of nT/$\sqrt{\text{Hz}}$, but only relative amplitude (and phase) are meaningful. We thus omit a scale for this and subsequent figures based on spatial modes. Blue and red arrows are real and imaginary parts of the complex vectors. (c and d) First spatial mode for 2 and 3 cpd.

as its coupling to the mesosphere below and the magnetosphere above (Qian *et al.* 2014). The ionospheric electrodynamic solver in TIEGCM simulates the effect due to the neutral wind dynamo, gravity and plasma pressure-gradient currents (Maute & Richmond 2017), and high-latitude magnetospheric energy input (Richmond & Maute 2013). The TIEGCM set up for this study is described in Maute & Richmond (2017) and Alken *et al.* (2017). The TIEGCM resolution is 2.5° × 2.5° in geographic longitude and latitude, and 1/4 atmospheric pressure scale height (1.5–15 km, depending on temperature and composition) in the vertical direction.

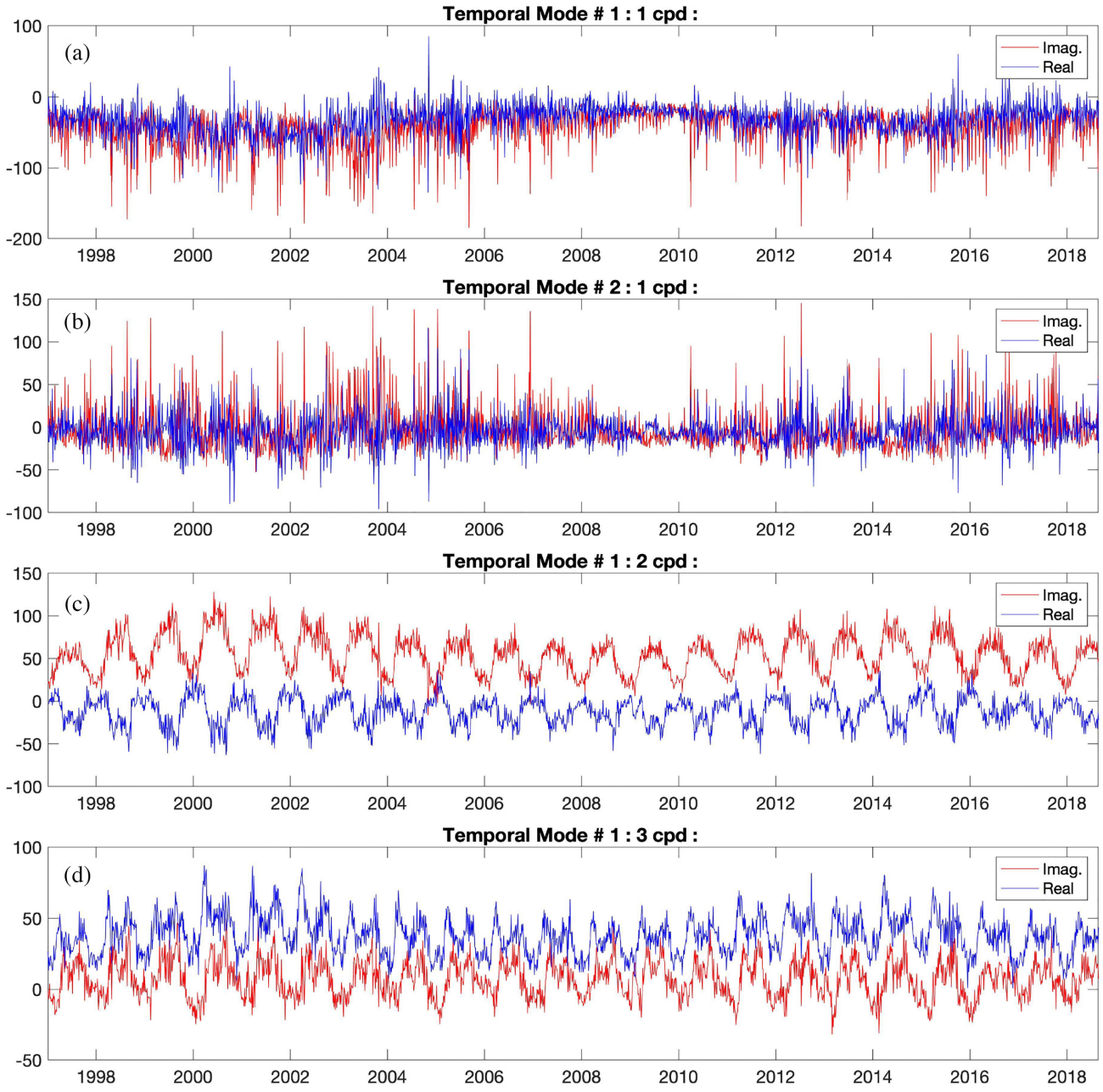

Figure 8. Temporal modes corresponding to spatial modes plotted in Fig. 7. (a) 1 cpd mode 1, (b) 1 cpd mode 2, (c) 2 cpd mode 1 and (d) 3 cpd mode 1. Temporal mode coefficients are complex; blue and red lines denote real and imaginary parts, respectively. Only coefficients for the centre frequency in the band are plotted. Note that the temporal modes are non-dimensional, but can be interpreted as SNR amplitude (comparable to Fig. 4).

At the lower boundary of the TIEGCM (approximately at 97 km) tidal and wave variability are prescribed, informed by 3-hourly Modern-Era Retrospective analysis for Research and Applications (MERRA) reanalysis data for the year 2009 (Häusler *et al.* 2015; Maute & Richmond 2017). The magnetosphere-ionosphere coupling at high latitudes is parametrized by the 3-hourly Kp index to determine the empirical ion convection (Heelis *et al.* 1982) and auroral particle precipitation patterns (Emery *et al.* 2012). TIEGCM solves for the global electrostatic potential by requiring that the total current is divergence-free. In a post-processing step using the TIEGCM neutral winds, ionospheric conductivities and high latitude ion convection, the full 3-D divergence-free ionospheric current can be determined using a stand-alone electrodynamo module (Maute & Richmond 2017). For this study hourly simulation output is used, mainly on the TIEGCM geographic grid. The resolution of the post-processing electrodynamo module is 4° in geomagnetic longitude and between 0.3° and 3.5° in geomagnetic latitude with higher resolution in regions of strong electric field and current gradients. Note that for calculating the magnetic perturbation the current is mapped to a regular geographic grid and geographic spherical coordinate system (Maute & Richmond 2017). We conducted two simulations, one during solar minimum conditions by using 2009 values of the solar radio flux $F_{10.7}$ and the Kp index, and one during solar maximum conditions using 2002 values for $F_{10.7}$ and the Kp index. Upward propagating tides and waves from the lower atmosphere do not strongly depend on the solar cycle and

therefore both simulations use the specification from 2009. This is a valid approximation since the focus of this study is on capturing the dominant modes and not the evolution of specific ionospheric current variations.

3.2 PCA of model outputs

The TIEGCM and its post-processor produce time-series of 3-D current density distributions throughout the ionospheric region. A much simpler description of the external sources suffices to define the forcing for Earth induction modelling, namely the magnetic fields on the surface of the Earth associated with these currents. For this study, where we focus on estimating source (and ultimately Earth conductivity) using ground-based data, the TIEGCM time-series can be reduced to a grid of magnetic fields on Earth surface, using SHA on the geographic grid (Maute & Richmond 2017). These (external) surface magnetic fields are then used, with a FD PCA approach to derive frequency dependent spatial modes for modelling of external sources. This can be viewed as a direct analogue of the data PCA, with the grid of modelled surface magnetic field components treated as a dense array of observatories. Time-series at each gridpoint would then be transformed to the FD with a windowed STFT, and PCA applied to the resulting sequence of complex FC vectors to derive the dominant spatial modes of variability for surface magnetic fields.

However, as already noted, there are advantages to working in the SH domain. In particular, the surface magnetic fields derived from TIEGCM can be represented as a scalar external potential (rather than a 3-component vector field), so this transformation results in some computational efficiency for subsequent processing. Furthermore, in the SH domain this potential is easily mapped to a stream function, or an equivalent sheet current, at some fixed altitude—for example, 110 km, a representative height for the E-layer—allowing source estimates to be presented in a physically meaningful form. Of course, transformation back to magnetic field vectors (at any altitude below the source region) is also straightforward. We thus first convert the TIEGCM surface magnetic fields $\mathbf{B}(\theta, \phi, t)$ to an external potential expanded in SH $\psi_l^m(t)$ (maximum degree $L = 60$, order $M = 30$; chosen to (approximately) match scales that are actually resolved in the TIEGCM), transform these time-series with a STFT, and then apply PCA to the resulting sequence of FCs. All processing parameters, such as window lengths and overlaps, and frequency bands used for PCA, are chosen to exactly match those used for the observatory data analysis. Of course, since there is no missing data, simpler approaches to PCA (i.e. based on SVD) can be used. Further details on the steps sketched here are given in Appendix D.

The model spatial modes which result from this FD PCA are sets of external potential functions Φ_{ji}, $i = 1, ..., I_j$, which best represent (in terms of variance) TIEGCM model spatial structure in each of the data analysis bands $j = 1, ...J$. Note that the PCA truncation level, I_j may vary between bands. Examples of these modes ($i = 1, 2, 3, 5, 10$), represented as a stream function (see Appendix D, eq. D3) are presented in Fig. 9 for the $j = 11$ band with centre frequency 1 cpd. Real and imaginary parts are shown in the left- and right-hand columns, respectively. The first mode exhibits multiple Sq-like vortices centred at different longitudes. The Sq structure is prominent at mid-latitudes in the TIEGCM current solution, and since we did not subtract means from the Fourier transformed model outputs before PCA, these patterns appear clearly in the first dominant principal component. For the higher modes, we see that the spatial patterns become more complex, particularly for the 5th and 10th mode. Since these stream function maps were generated from the columns of a unitary matrix (see Appendix D), the colour scale is dimensionless. In the source estimation problem, these modes will act as basis functions representing ionospheric sources, and physical units will come from model coefficients multiplying these modes.

Fig. 10 presents the cumulative variance explained by the first 20 modes, plotted versus frequency. We see that the first mode explains more than 90 per cent of the variance in the simulated 2-D external ionospheric sheet current for the 1 cpd band. The first 20 modes explain more than 97 per cent of the variance in all frequency bands. Obviously the model exhibits less complexity (in a statistical sense) than the observations. However, we only require the model run to produce a rich enough set of physically realistic spatial basis functions to fit the data. As we shall see, in the next section, this appears to be the case.

4 SOURCE ESTIMATES FOR INDIVIDUAL FD DATA MODES

The FD spatial data modes $\mathbf{U}_{jk}$ define the total (external source, plus internal induced) magnetic fields sampled at the observatory locations. Our goal in this section is to estimate these external source components, and corresponding total surface magnetic fields ($\mathbf{B}_{jk}$ in Fig. 1). The steps described next are represented in Fig. 1 by the 'Induction Modelling' and 'Model Fitting' boxes. Note that throughout this section everything—data, external source potentials and magnetic fields are implicitly defined for a fixed frequency band B_j, with centre frequency f_j. As in Section 3 we represent the source (external potential) as a vector of SH coefficients (SHC), which we assume can be expanded in terms of the TIEGCM modes for the corresponding frequency band

$$\Psi_{jk}^{ext} = \sum_i \beta_{jki} \Phi_{ji}, \tag{2}$$

This reduces the estimation problem for data mode k, frequency band j, to finding the model-mode coefficients β_{jki}. In practice, we will use a subset of I_j model modes, corresponding to the largest singular values.

To compute total magnetic fields on Earth's surface corresponding to a given external source potential function Ψ^{ext} we must compute the corresponding induced internal component Ψ^{int} by solving an induction problem for some assumed Earth conductivity, evaluate gradients of the external and internal potentials to compute magnetic fields, and sum these components. We represent these steps as

$$\mathbf{B}(\theta, \phi) = \mathbf{F}_{f,\sigma}\left[\Psi^{ext}\right](\theta, \phi), \tag{3}$$

with the subscripts f, σ indicating dependence of the mapping on frequency and Earth conductivity. Thus, the result of applying $\mathbf{F}_{f,\sigma}$ to Ψ^{ext} is the total (internal plus external) surface magnetic fields, as a function of latitude θ and longitude ϕ, resulting from a source of frequency f with spatial structure defined by Ψ^{ext}.

Here we solve the induction problem using the global thin-sheet model of Sun & Egbert (2012). The simplified thin-sheet conductivity model comprises a radially symmetric Earth, overlain by a thin inhomogeneous surface layer representing the strong ocean–continent conductivity contrast. For results shown in this paper, the radial conductivity structure is taken from Püthe *et al.* (2015) and the heterogeneous thin sheet conductance is derived following the

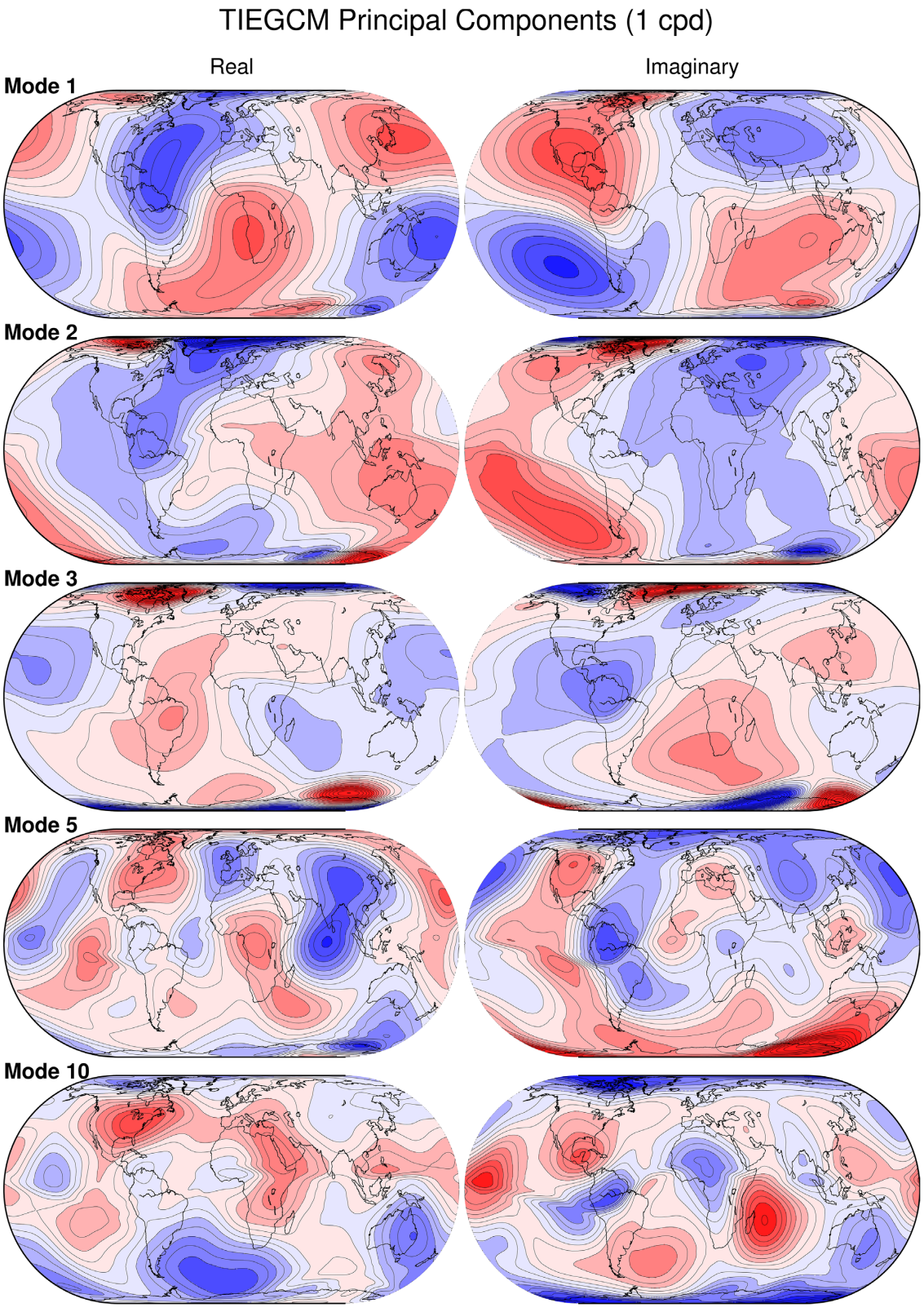

Figure 9. TIEGCM derived principal components for the 1 cpd band represented as sheet current stream functions at 110 km altitude. Modes 1, 2, 3, 5 and 10 are shown with real and imaginary parts (left- and right-hand columns, respectively) in Eckert projection. Colour scale is dimensionless.

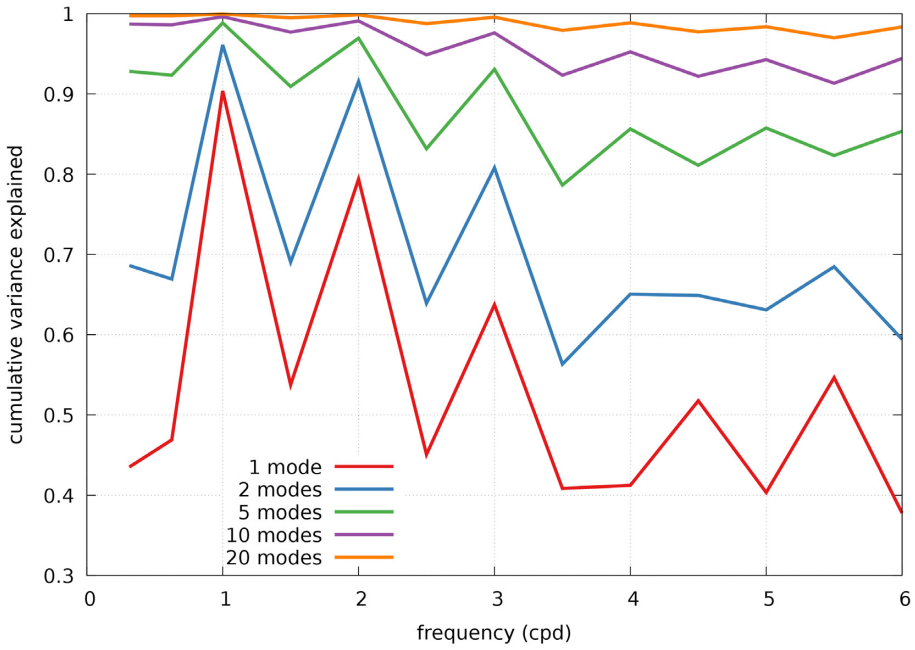

Figure 10. Cumulative fraction of total variance explained by the TIEGCM derived principal components as a function of frequency.

procedure described in Manoj *et al.* (2006) with seawater conductivity taken from Tyler *et al.* (2017), and the sediment contribution calculated following Everett *et al.* (2003). Certainly, this simplified Earth conductivity model should ultimately be refined—doing so is in fact a major motivation for our efforts to develop improved DV source models. However, the ocean–continent conductivity contrast, which is well constrained, is certainly the most important generator of internal anomalous fields which might impact source estimates. Other variations in upper mantle conductivity which might influence DV data are not now so well constrained, at least at a global scale. A fully 3-D spherical induction code, for example, as described by Kelbert *et al.* (2008), could easily replace the thin sheet model used here, if warranted by availability of sufficiently accurate global 3-D conductivity models. As outlined above, iterative refinement of models for source structure and Earth conductivity is perhaps the most promising path forward.

From eqs. (2) and (3) surface magnetic fields associated with data mode k in band j (with centre frequency f_j) are

$$\mathbf{B}_{jk} = \mathbf{F}_{f_j,\sigma}\left[\Psi^{ext}_{jk}\right] = \mathbf{F}_{f_j,\sigma}\left[\sum_{i=1}^{I_j}\beta_{jki}\Phi_{ji}\right] = \sum_{i=1}^{I_j}\beta_{jki}\mathbf{F}_{f_j,\sigma}\left[\Phi_{ji}\right], \quad (4)$$

the last step following from the linearity of the mapping $\mathbf{F}_{f,\sigma}$. We write $\mathbf{y}_{ji}$ for $\mathbf{F}_{f_j,\sigma}\left[\Phi_{ji}\right]$ evaluated at the N_s station locations (θ_s, ϕ_s), that is the $3N_s$-dimensional complex vector of magnetic field components associated with model mode ji restricted to the observatory locations. Since data mode $\mathbf{U}_{jk}$ is just $\mathbf{B}_{jk}$ sampled at the observatory locations we can write

$$\mathbf{U}_{jk} = \sum_{i=1}^{I_j}\mathbf{y}_{ji}\beta_{jki} + \mathbf{e}_{jk}. \quad (5)$$

Here we have added an error term to represent inadequacies in our model (e.g. incorrect Earth conductivity, spatial structure that is poorly represented in the TIEGCM, such as sharpness of auroral or equatorial electrojets), as well as noise in the data modes. To estimate the band j, mode k source model expansion coefficients $\mathbf{b}_{jk} = (\beta_{jk1}, \ldots \beta_{jkI_j})^T$, we thus must solve one small linear regression problem

$$\mathbf{U}_{jk} = \mathbf{Y}_j\mathbf{b}_{jk} + \mathbf{e}_{jk}, \quad (6)$$

with weights defined by our model for data errors. Note that the design matrix $\mathbf{Y}_j = \left[\mathbf{y}_{j1} \ldots \mathbf{y}_{jI_j}\right]$ is the same for all K_j modes in frequency band j.

To define approximate weights, we assume that the noise vector $\mathbf{e}_{jk}$ has diagonal covariance Σ_{jk}, with variances for each component derived as in Appendix C. This is certainly an oversimplification, both in terms of correlation structure, and error scales, which appear to be unreasonably small. However, as discussed in Appendix C, these diagonal variances probably provide reasonable estimates of the relative magnitude of errors for different sites/channels, and can thus provide reasonable weights for the linear regression model of eq. (6).

Errors in the Earth conductivity model—these are really errors in the prediction (right-hand side of eq. 5) not in the data (left-hand side)—are not accounted for in the formal spatial mode error estimates embedded in Σ_{jk}. In fact, the B_z components are much more strongly affected by details of Earth conductivity than the horizontal components (e.g. Koch & Kuvshinov 2013). This is illustrated in Fig. 11, where we plot forward modelling results obtained using two different Earth conductivity models, excited by the same source, defined by Ψ^{ext}_{jk} estimated as discussed below for the 3 cpd band, mode 1. In the left-hand column we show results obtained using a 1-D earth model, and in the right-hand column using the thin sheet earth model. Spatial patterns for the horizontal components are almost identical, with only subtle differences in amplitudes, and anomalies associated with induction in the thin sheet subdued. In contrast, for B_z differences obtained with the different models are quite clear, and the continent ocean boundaries are first order features when the thin sheet model is used. Thus, B_x, B_y contain most of the information about the source spatial structure, and are only weakly

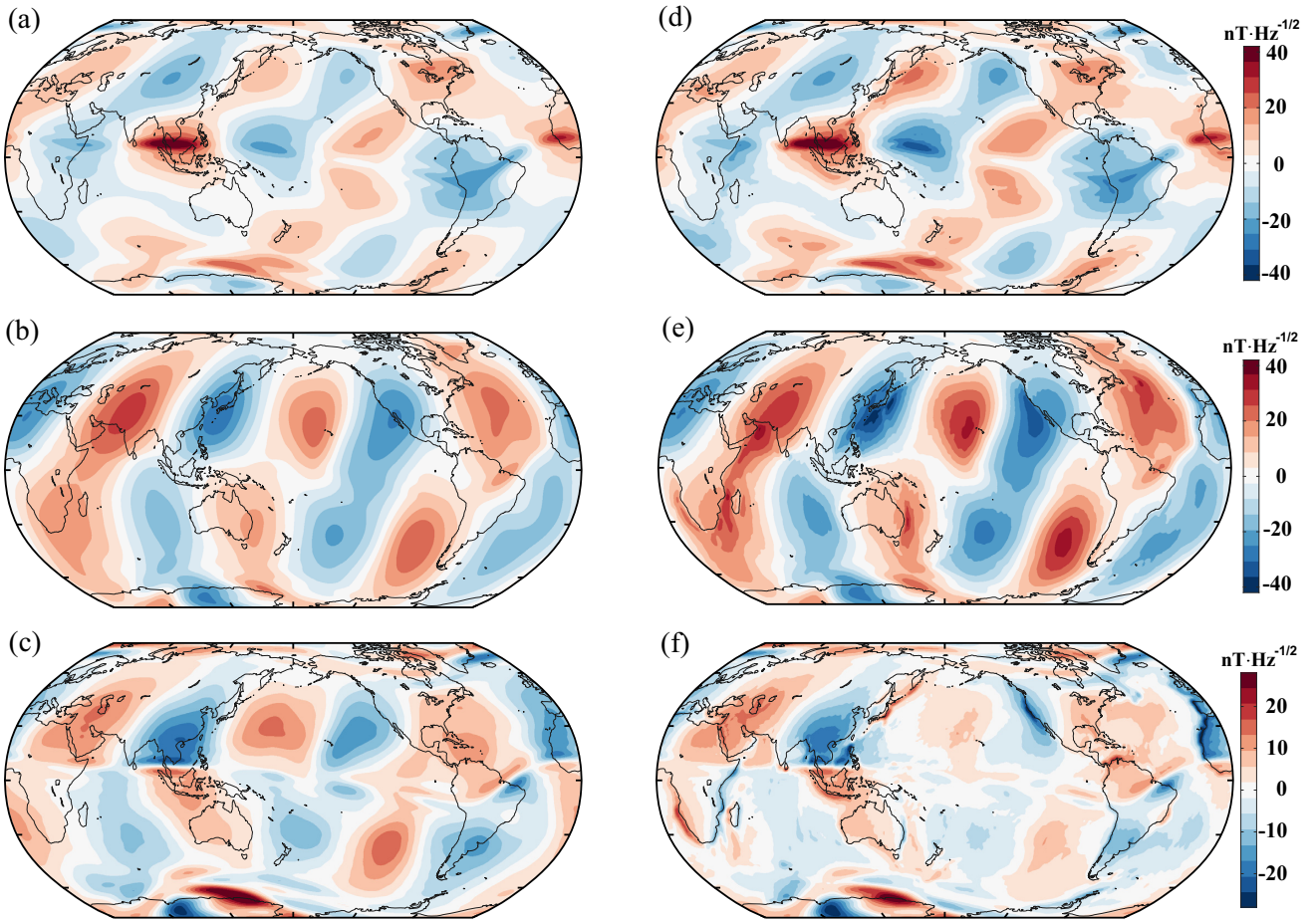

Figure 11. Estimated B_{jk} magnetic field components at Earth's surface for mode 1 at 3 cpd. Source fields are defined by the estimated source potential Ψ_{jk}^{ext} with internal induced fields computed using (a)–(c) a 1-D layered earth model, and (d)–(f) the thin-sheet model. The same 1-D profile (from Püthe *et al.* 2015) is used in both cases. Magnetic components are, top to bottom, B_x, B_y and B_z.

affected by errors in the assumed conductivity model. If vertical components are used in the source inversion, they should probably at least have error variances increased. We have chosen to omit B_z from the source estimates, and only fit the horizontal components of $\mathbf{U}_{jk}$.

We use regularized least squares to estimate the model expansion coefficients, minimizing the penalty functional

$$\phi(\mathbf{b}_{jk}) = (\mathbf{U}_{jk} - \mathbf{Y}_j\mathbf{b}_{jk})^T\mathbf{W}_d(\mathbf{U}_{jk} - \mathbf{Y}_j\mathbf{b}_{jk}) + \lambda\mathbf{b}_{jk}^T\mathbf{W}_s\mathbf{b}_{jk}, \qquad (7)$$

where λ is the trade-off parameter between data misfit and source magnitude. We assume for the data weight matrix $\mathbf{W}_d = \Sigma_{jk}^{-1}$, and for the model weight matrix $\mathbf{W}_s = \mathbf{S}_j^{-1}$ where $\mathbf{S}_j$ is the diagonal matrix of singular values from the TIEGCM PCA for frequency band j, as defined in Appendix D. As noted above, we only include horizontal data components in the fitting, and we use a total of 50 model modes for all bands, that is $I_j = 50$. To choose the regularization parameter λ, we use a leave-one-out cross validation (LOOCV) scheme (Hastie *et al.* 2009), similar to that used in Sun *et al.* (2015). For a fixed value of λ, the LOOCV score is the root-mean square prediction error. Predictions are calculated by minimizing the penalty function of eq. (7), with each data point omitted in succession, and using the resulting solution for the prediction at the omitted point. Fig. 12 illustrates the LOOCV score, as well as coefficient of determination R^2 between observed and predicted horizontal fields, calculated with different values of λ, for two modes. The minimum of the LOOCV score is then used to select λ for the final estimation of source coefficients. In our case, the optimal λ is generally around 10^3. In fact, the LOOCV score minimum is generally quite broad, and a wide range of regularization parameters produce very similar solutions. Thus, the source fitting procedure is rather stable.

Separate minimization problems are solved for each of the $J = 13$ bands, and for all $K_j = 20$ spatial modes for each data band. One example of the estimated surface magnetic fields $\mathbf{B}_{jk}$ is provided in the right-hand column of Fig. 11. Fits are summarized, as coefficient of determination R^2 plotted as a function of frequency, for modes $k = 1, ..., 20$ in Fig. 13. As this shows, fits are best for the dominant modes, and for 1 cpd and harmonics up to the 4th. Fits degrade for higher modes, but the first 10 modes have $R^2 > .8$ for almost all bands. We will use all 20 of the fitted modes for all bands for the full DV source model presented here. The fits are all stable, and none of the estimates are obviously noisy. Note however, that the more poorly fit high modes carry little of the total signal. Omitting these would result in only small changes in the final model. Note also that from the perspective of EM induction, only the first few modes, which start from the highest data SNR, and are also generally simplest (and hence most easily fit with the TIEGCM modes), are likely to be useful. The relative misfit of the model (i.e. $1 - R^2$), broken down by magnetic field component, and QD latitude, are shown in Fig. 14. Solid lines are misfits for the PCA model with 20 modes, dashed lines are misfits to the spatially continuous model fit

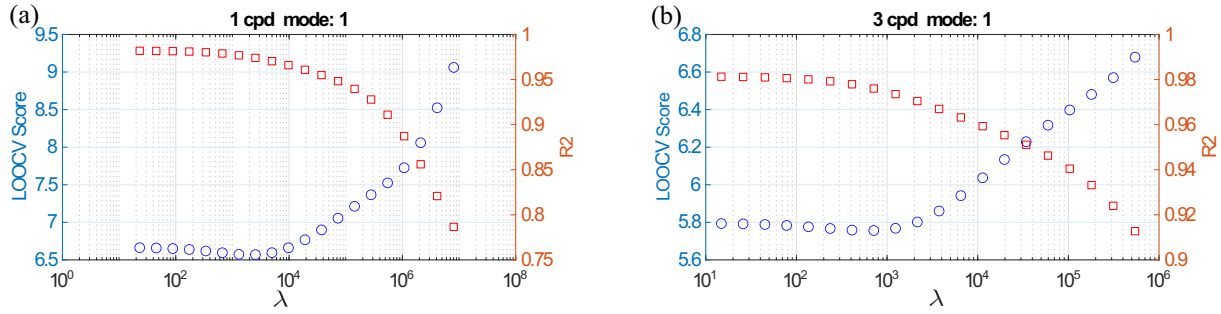

Figure 12. LOOCV score and R^2 as a function of λ for source inversion for mode 1 at (a) 1 cpd and (b) 3 cpd.

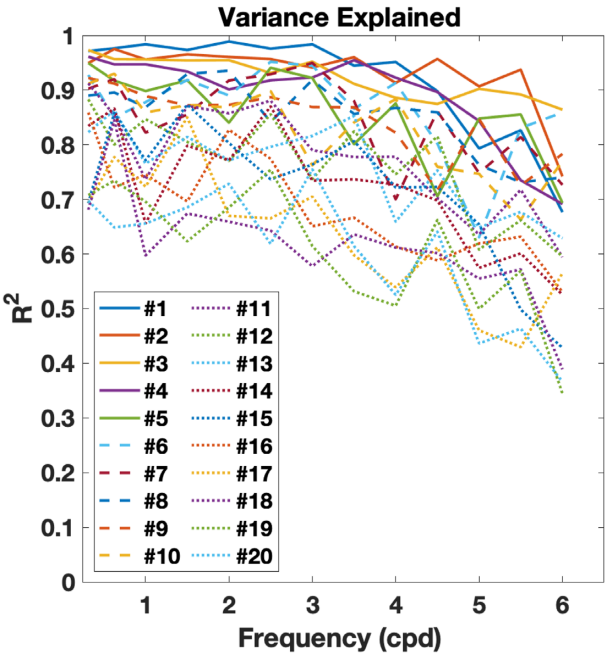

Figure 13. Fraction of variance explained by fitted source model, for all bands, 20 modes.

to TIEGCM modes. Relative misfits for the PCA model are greatest for B_z both because magnitudes for this component are smaller, and because more poorly resolved small scale source features have relatively large B_z. The relative misfit for B_z is even greater for the model, because of induced internal components and because we did not fit B_z. The relative misfit is only slightly reduced for quiet-time data.

Another perspective on the fitted fields is provided in Fig. 15, where we plot estimates of horizontal fields as vectors. The upper panels of the figure (a–b) are the estimated surface magnetic fields $\mathbf{B}_{jk}$ sampled at the observatory locations. These should be compared to the corresponding data modes $\mathbf{U}_{jk}$ displayed in Figs 7(a) and (d). In the lower panels (c–d) the estimated magnetic fields are plotted on a regular grid. Thus, our modelling procedure can be viewed as a physically based optimal interpolation scheme, which uses TIEGCM (and an Earth conductivity model) to constrain the interpolation between the sparsely sampled field vectors. Source currents associated with the magnetic fields of Fig. 15, represented as an equivalent sheet current at 110 km altitude, are shown in Fig. 16. These are derived from the source potential Ψ^{ext}_{jk}, following Appendix D. Note that these are the estimated external currents (reduced to a toroidal current sheet, no internal component), while Figs 11 and 15 show total magnetic fields, the sum of internal induced and external source components.

5 TD MODEL

We have developed a FD model for daily variation source fields, with spatial variations described by the potential functions Ψ^{ext}_{jk}, and temporal variations by the sequence of complex mode coefficients $\mathbf{a}_{kj}$. The spatial components of this model can be directly used for imaging deep Earth conductivity, refining the fit to eq. (5) by adjusting (or inverting for) conductivity parameters used to calculate total magnetic fields (i.e. σ in eq. 4), while leaving the source estimate fixed. This is the approach used in Sun *et al.* (2015), and, with different source estimation and data processing procedures, by Koch & Kuvshinov (2013) and Koyama *et al.* (2014). However, our analysis significantly extends the source modelling of these previous induction studies, both by retaining multiple modes of spatial structure for each frequency band, and by describing how the mix of source modes varies in time (i.e. the FD temporal modes $\mathbf{a}_{kj}$). As we now discuss this allows us to convert the FD source model to a (band limited) TD model of DV equivalent sheet current sources, and surface magnetic fields. These steps are represented in Fig. 1 along the bottom, and bottom part of the left-hand column. The model that results is global, and valid for 1997–2018.

As outlined in Appendix B, the STFT we use to transform the observatory data to the FD (i.e. for a single channel, $X_c(t) \rightarrow X_{cfn}$) can be easily inverted ($X_{cfn} \rightarrow X_c(t)$. The same inverse transformation can be applied to invert the sequence of mode-k temporal coefficients $a_{kfn} \rightarrow a_k(t)$. Considering a single frequency $f \in B_j$, eq. (1) can be transformed to the TD as

$$\mathbf{X}_f(t) = \mathrm{Re}\left[\sum_{k=1}^{K_j} \mathbf{U}_{jk} a_{kf}(t)\right] + \mathbf{e}_f(t). \tag{8}$$

$\mathbf{X}_f(t)$ is the original time-series, restricted to a very narrow frequency band—essentially a slowly modulated sine wave. The TD temporal mode function $a_{kf}(t)$ is complex, with real and imaginary parts (a Hilbert transform pair). Ignoring the error term for now, summing over all frequencies $f = 1, \ldots, N_f$, and noting that spatial variations are assumed to depend only on band j, the original (HP filtered) vector time-series can be approximated as

$$\begin{aligned}\mathbf{X}(t) &= \mathrm{Re}\left[\sum_{j=1}^{J}\sum_{k=1}^{K_l} \mathbf{U}_{jk} \sum_{f\in B_j} \alpha_{kf}(t)\right] \\ &= \mathrm{Re}\left[\sum_{j=1}^{J}\sum_{k=1}^{K_l} \mathbf{U}_{jk}\alpha_{kj}(t)\right].\end{aligned} \tag{9}$$

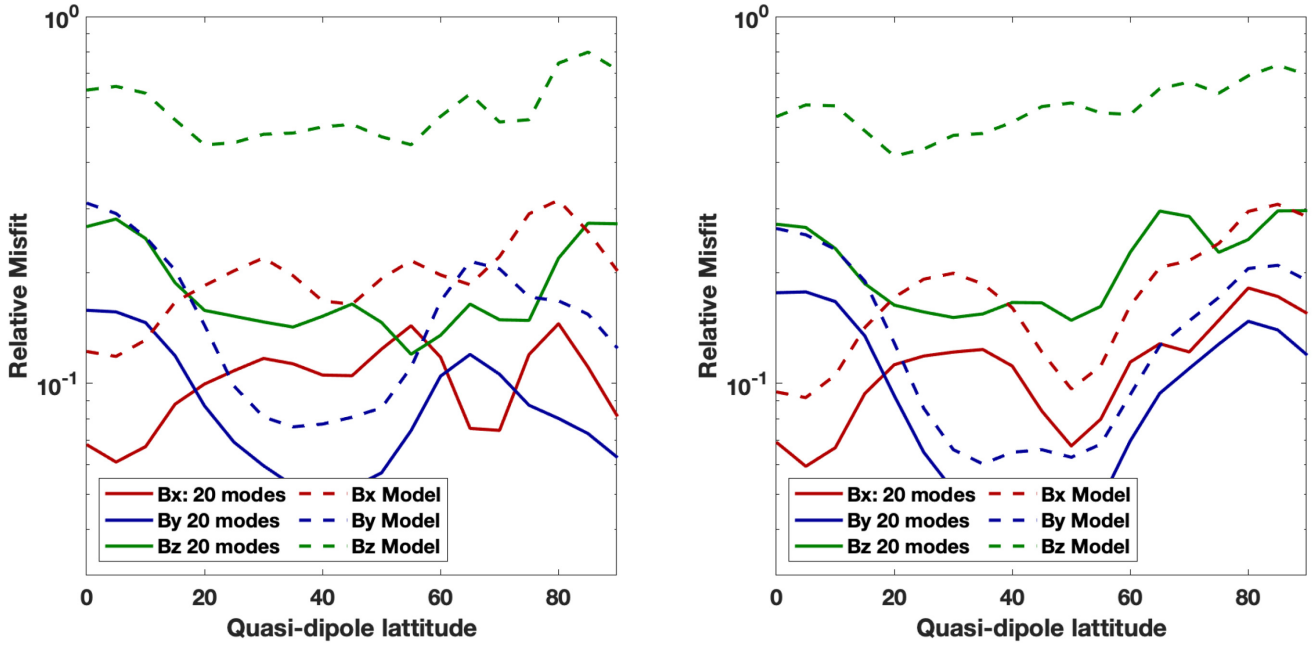

Figure 14. Fraction of variance unexplained (i.e. $1 - R^2$) for each channel, as a function of QD latitude. (a) all data for 1997–2009; (b) quiet time data $K_p < 1$.

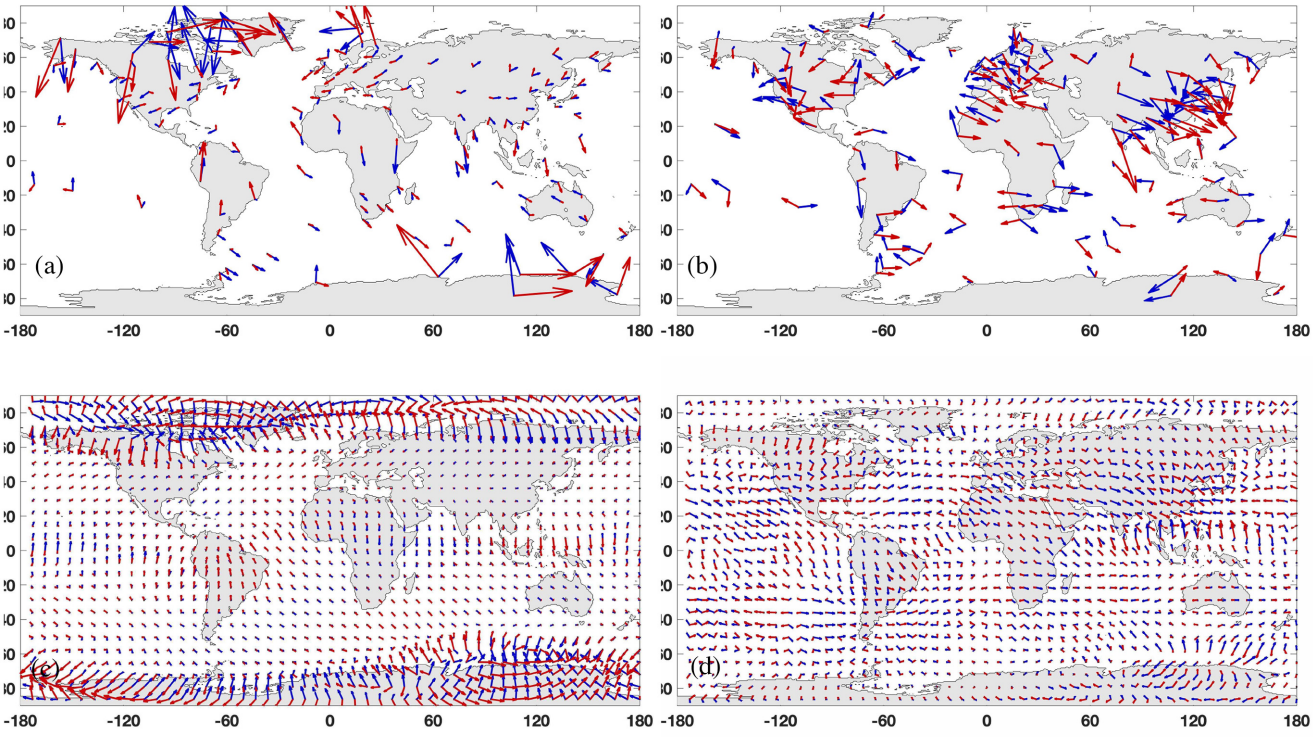

Figure 15. Magnetic fields at surface $\mathbf{B}_{jk}$ for mode $k = 1$. Left-hand column is for 1 cpd band ($j = 11$), right for 3 cpd ($j = 7$). Top panels (a–b) show horizontal field components at observatories; compare to corresponding spatial data modes shown in Fig. 7 panels (a) and (d), respectively. Bottom panels (c–d) show the same horizontal fields, but sampled on a regular grid. In all plots blue vectors are real parts, red imaginary. Units are nT/$Hz^{1/2}$, but only relative variations in vector magnitude are meaningful. Different scales are used for vectors in all panels.

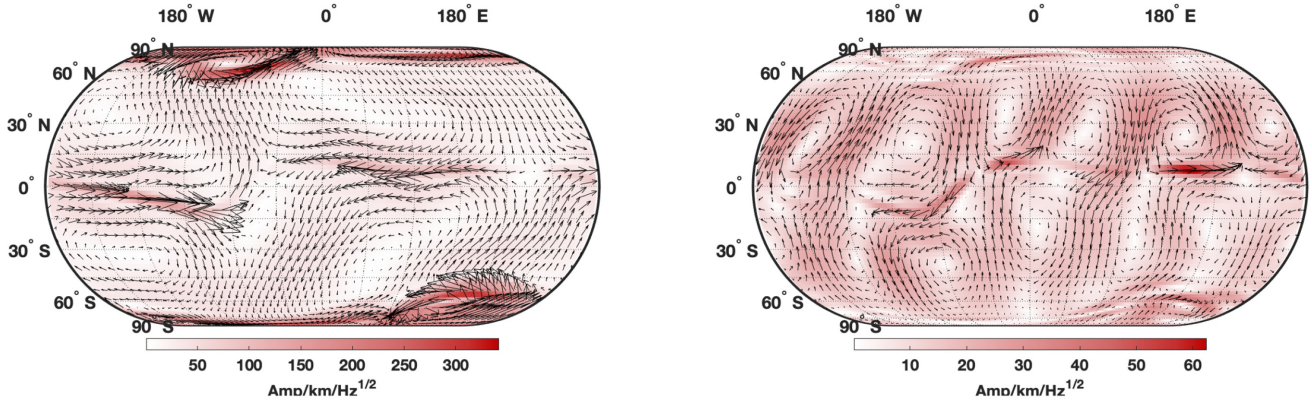

Figure 16. Real parts of FD equivalent sheet currents, computed from source stream function estimates Ψ_{jk}^{ext}, for mode $k = 1$, (a) 1 cpd ($j = 11$ in Table 1) and (b) 3 cpd ($j = 7$). Background colour is amplitude of vector field, with vectors overlain.

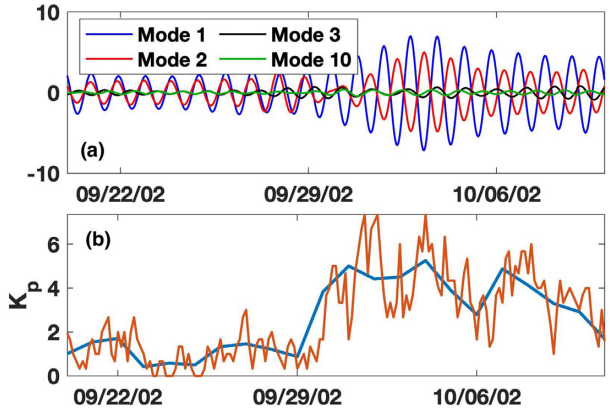

Figure 17. (a) Examples of narrow-band time-domain temporal modes $\alpha_{kj}(t)$ for the 1 cpd band. Only real parts are shown. Temporal modes are unit free; x-axis is date. (b) Kp index and daily mean Kp. Geomagnetic conditions are quiet in the first half of the interval, active in the second. Note the change in relative phase of modes, as well as the increase in amplitude, at the transition to active conditions. Note also that this transition is smoothed by the windowing (roughly 4-d timescale).

In this form there is one temporal mode $\alpha_{kj}(t)$ for each data mode $k = 1, \ldots, K_j$ for each frequency band j, which merges variations for all frequencies in the analysis band. Examples of the TD temporal mode functions for a 21-d time window are given in Fig. 17. Eq. (9) gives a TD representation of the FD PCA model.

Replacing the spatial mode $\mathbf{U}_{jk}$ with the corresponding interpolated fields $\mathbf{B}_{jk}(\theta, \phi)$ given in eq. (4) we obtain a representation of magnetic fields in the DV band that is continuous in both space and time

$$\mathbf{B}(\theta, \phi, t) = \mathrm{Re}\left[\sum_{j=1}^{J}\sum_{k=1}^{K_j}\alpha_{kj}(t)\mathbf{B}_{jk}\right] = \mathrm{Re}\left[\sum_{j=1}^{J}\sum_{k=1}^{K_j}\alpha_{kj}(t)\sum_{i=1}^{I_j}\beta_{jki}\mathbf{F}_{\sigma,f_j}(\Phi_{ji}(\theta,\phi))\right]. \quad (10)$$

In eq. (10) magnetic field variations are expressed as a sum over products of the temporal and spatial functions. The spatial functions $\mathbf{B}_{jk}(\theta, \phi)$ are derived from interpolating data spatial modes $\mathbf{U}_{jk}$ using basis functions derived from TIEGCM. The temporal functions $\alpha_{kj}(t)$ do not depend on the TIEGCM outputs in any way. Thus, the temporal window covered by the model depends only on the interval used for the data PCA.

Following the steps outlined above, we have created a model for the 21-yr period (1997–2018). Temporal resolution is nominally 1 hr, and the longest periods represented are roughly 2–3 d. The model provides estimates of three components of DV band magnetic field variations at any point on Earth's surface. Spatial resolution is of course limited, both by the density of observations, and the resolution of the TIEGCM model runs used to define spatial structure. Our approach extends previous efforts to model DV fields (Sabaka *et al.* 2004), as we seek to represent variations at all latitudes, and in active as well as quiet geomagnetic conditions.

The TD model of eq. (10) can be viewed as an observation equation that relates magnetic fields at any point in space and for any time in the data analysis period to the coefficients β_{jki}, treated as unknowns. Thus, the model can be easily adapted to directly fitting satellite magnetic data, such as from the CHAMP or Swarm missions. The only significant modification required for this extension is that spatial basis functions $\mathbf{F}_{\sigma,f}\left[\Phi_{ji}\right]$ must be defined from the surface of the Earth to satellite altitudes—that is as functions of θ, ϕ, r. Since the 3-D current based on the TIEGCM model can be determined, this is relatively straightforward. Now PCA of the model outputs must be applied to 3-D model outputs (allowing for both poloidal and toroidal modes), rather than to the 2-D potential functions used here, and the induced fields must be upward continued into the ionospheric domain. Again, these extensions are straightforward. Thus, we may use the modelling approach developed here to merge ground-based and space observations to constrain spatial structure of DV source fields. We emphasize that with this approach the ground data, which samples the variations continuously in time, would still be used to define the temporal mode functions $\alpha_{kj}(t)$. Further details on this important extension, and initial applications to incorporating satellite data into a DV model will be presented in a future publication.

6 RESULTS: AN EMPIRICAL MODEL FOR 21 YR OF DV-BAND MAGNETIC FIELDS

In this section we explore how well our initial model reproduces ground observations, both from the observatory database used for fitting the model coefficients, and from independent magnetic observations.

In Fig. 18 we plot magnetic field time-series at five representative sites for the same 21-d period (9/20/2002–10/10/2002) used for Fig. 17, near the fall equinox. For roughly the first half of this interval geomagnetic conditions were quiet (mean $K_p \approx 1$), transitioning to active conditions ($K_p > 4$) in the second half, with a significant storm ($K_p = 7$) on 10/01/2002. For each site we show the original observatory time-series ((high-pass filtered, as above; blue lines), the TD PCA approximation of eq. (8), using 20 modes for all bands (black lines), and the estimated DV model of eq. (10) (red lines). The observatories selected (black squares in Fig. 2) include mid-latitude (BMT, QD latitude 34.1°), moderately high latitude (NEW, QD latitude 54.6°), auroral zone (CMO, QD latitude 64.7°; MCQ QD latitude –64.0°), and equatorial electrojet (HUA, QD latitude 0.7°). Only horizontal components are shown; fits to B_z are somewhat poorer, in part because this field component was not directly fit. Perhaps more importantly, larger misfits in B_z reflect errors in the internal fields, due to an incorrect Earth conductivity model. Except for HUA, where modelled amplitudes for B_x are systematically too small, the three curves overlie each other almost perfectly during quiet conditions. We believe that the relatively poorer performance at HUA reflects the limited resolution of our modelling effort; the equatorial electrojet is almost certainly artificially broadened and smoothed, if not in the resolution of TIEGCM model runs, then in subsequent post-processing to compute surface magnetic fields, and the model PC modes. Higher resolution modelling and/or processing might be expected to improve the agreement. Not surprisingly, fits are somewhat poorer during the active interval, where again modelled amplitudes are reduced relative to observations. Still, the model tracks the complex evolution of the disturbed magnetic fields quite well, demonstrating that our model has some validity for all times.

Snapshots of the global source model for a few times during the interval displayed in Fig. 18 are shown in Fig. 19, represented as stream functions for the equivalent sheet current (defined in eq. D3). Each column represents the source current system for one day, plotted, top to bottom, at 0:30, 6:30, 12:30 and 18:30 UT. The first column is for 9/24/2002, during the interval of quiet geomagnetic conditions, so the plotted stream functions essentially represent S_q, propagating from east to west. The second column is for 10/3/2002, during the active period, when mid-latitude fields are highly disturbed (and do not particularly resemble S_q), and current systems are much stronger in the auroral zones. We emphasize that plots of this sort could be made for any time during the 21-yr period (1997–2018) used for the data PCA. The full model, represented in eq. (10) as a sum over products of temporal and spatial modes is relatively compact, compared to the full set of spatio-temporal fields.

To further validate our model of surface magnetic fields, we compare predictions at some observatories not used to construct the model, selected from the blue symbols in Fig. 2. Since these sites were not included in the original data PCA, we cannot project these data onto the PCA spatial modes. Comparison between the measured magnetic fields and the model predictions is shown in Fig. 20, for the four sites marked by magenta triangles in Fig. 2. The first observatory used (BDV; QD latitude 44.7) is in Europe, and was omitted from our analysis because of the high density of sites in this area. Given the proximity to sites that were used to construct the model, it is perhaps not too surprisingly that the agreement between model predictions and data is excellent, comparable to fit at the mid-latitude observatory BMT shown in Fig. 18. For BDV we use the same time window as for the comparisons shown in Figs 18 and 19. At the other 3 sites no data were recorded during this time interval. QZH (QD latitude 17.9) is on the Chinese coast, and also relatively close to sites used in the model construction. The fit is excellent for B_y, slightly poorer for B_x. Possibly this reflects the relatively low latitude, or incomplete modelling of internal fields. During the interval plotted K_p varied between 0 and 4. The other two observatories are not near sites used in the modelling—these sites were omitted because there was little available data—in both cases just a few months, in early 1997, when K_p was mostly 2–4, but reached 6 on 2/9/1997. Rather surprisingly, the fit for the polar observatory MBC (QD latitude 80.7) is quite good. The fit for GLN which is close to the auroral zone (QD latitude 59.5) is a bit poorer. It is at least possible that this partly reflect some problem with these data, as the available data at this site was very limited and had many gaps.

Overall, the result of the validation exercise are encouraging, and suggest that the model should have predictive value, for both quiet and active times, at least near areas where some data constraints are available. This includes most of the continents. The present model is almost certainly less accurate over the ocean basins, and in the Southern Hemisphere.

7 DISCUSSION AND CONCLUSIONS

In this paper we have developed a novel approach to modelling the spatial and temporal structure of external ionospheric magnetic field variations and applied this to empirical modelling of DV source fields. The original motivation for this effort was to enable use of MV methods for 3-D imaging of upper mantle and transition zone electrical conductivity. However, both the DV source model, and the methodologies developed here, have considerably broader applicability. Accurate models of external magnetic fields are required to provide corrections for space-based studies of core and crustal components of the geomagnetic field, and also for more practical applications, such as aeromagnetic surveys. Our DV model represents a step towards developing models for this component of the geomagnetic field that are useful at all latitudes, and for all geomagnetic conditions. In terms of methodology, our approach can clearly be extended, to include lower frequencies for deeper imaging of global Earth conductivity, or higher frequencies, where detailed models of source spatial structure may be required for work with geomagnetically induced currents (GIC).

The DV model presented here should be considered preliminary. Many improvements are possible, for example using additional ground-based magnetic data available from the THEMIS (Russell *et al.* 2008) or SuperMag (Gjerloev 2012) compilations, and extending the time period using more recent, as well as older, observatory data. We can also extend and improve the set of spatial basis functions, using additional runs of TIEGCM, for example explicitly including more intense storms, or modifying model forcings. Higher resolution could be used, both for modelling and post-processing, to better capture narrow features like the equatorial electrojet. We could also incorporate magnetospheric models, ultimately coupled to the ionosphere, to further improve completeness and realism of the spatial basis functions, and we could treat tidal magnetic fields more carefully (Schnepf *et al.* 2018). Perhaps most importantly, we can include satellite data, using eq. (10) as a TD observation equation. The methodology described explicitly here only needs minor extension, replacing the 2-D equivalent sheet current (or external potential) by a 3-D representation of currents in the definition of the source function Ψ^{ext}, and modifying the mapping $\mathbf{F}_{\sigma,f}$ to include upward continuation of the internal induced component, and

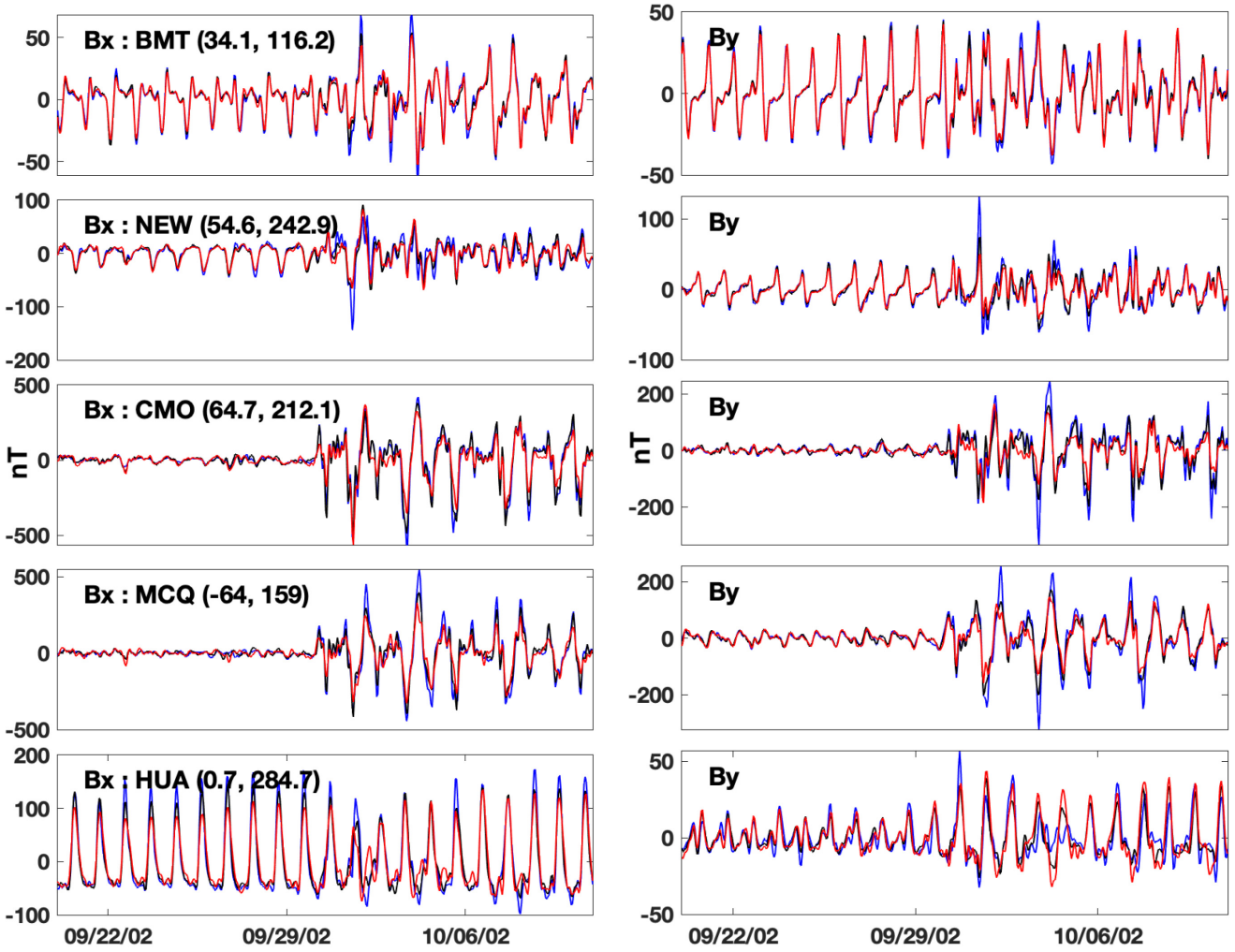

Figure 18. Horizontal components of magnetic fields for 20 d at five representative sites (black squares in Fig. 2), comparing high-pass filtered observatory data (blue lines), projection onto 20 data modes (black lines), and the global time domain model (red lines). Note that scales (all in nT) differ between plots. The period selected (around the fall equinox in 2002) includes both relatively quiet and active intervals; see Fig. 17(b) for plot of K_p. Observatories plotted include mid-latitude (BMT, NEW), auroral (CMO, MCQ) and equatorial electrojet (HUA). QD latitude, and geographic longitude are displayed in B_x plots in left-hand column. In general the model fit is best during quiet times, and at mid-latitudes. The model tracks data well at all latitudes, even during disturbed times, but model amplitudes are often reduced relative to data.

evaluation of magnetic fields as a function of θ, ϕ, r. All of these modifications are straightforward, and in fact have already been implemented and used for preliminary modelling of Swarm data (Egbert *et al.* 2019).

Our empirical modelling approach can be viewed as a somewhat simplified ensemble data assimilation scheme, which merges information from observations (ground-based so far, but ultimately also including satellite data) with model simulations. Merging data with modelling outputs through data assimilation provides a physically consistent way to interpolate sparse observations, as illustrated by Fig. 15. Data assimilation also provides a framework for exploring consistency between data and model inputs and assumptions. These dual aspects of data assimilation are discussed in (Egbert 1997b). Empirical models of ionospheric currents that can be derived with our approach may thus also be useful for ionospheric studies, particularly when extended to a fully 3-D representation of source currents.

Our modelling approach straddles the frequency and TDs. Spatial structure is represented as the sequence of source functions Ψ_{jk}^{ext}, one for each band/mode. Temporal structure can be represented in the FD through the sequence of complex temporal coefficients $\mathbf{a}_{kj}$, or in the TD through $\alpha_{kj}(t)$. The FD representation is convenient for induction studies, as modelling and inversion methods for conductivity are by far best developed and most efficient in the FD. The temporal representation is convenient for incorporating satellite data, which because of the moving observation platform, cannot be easily transformed to the FD. We have also argued that an FD approach is justified for magnetic DV, where optimal spatial basis functions are expected to depend on frequency. However, this will not always be true. In some cases, spatial structure may be decoupled from frequency, with a single set of spatial basis functions appropriate for representing sources for a broad range of frequencies. This is readily accommodated in our modelling framework by using broader frequency bands. Possibly narrower bands would then be required for calculating induced fields (and for using these source models for induction studies). Generalization to allow for different bands for modelling source, and induction, is possible, but beyond the scope of this paper. Careful tuning of frequency bands,

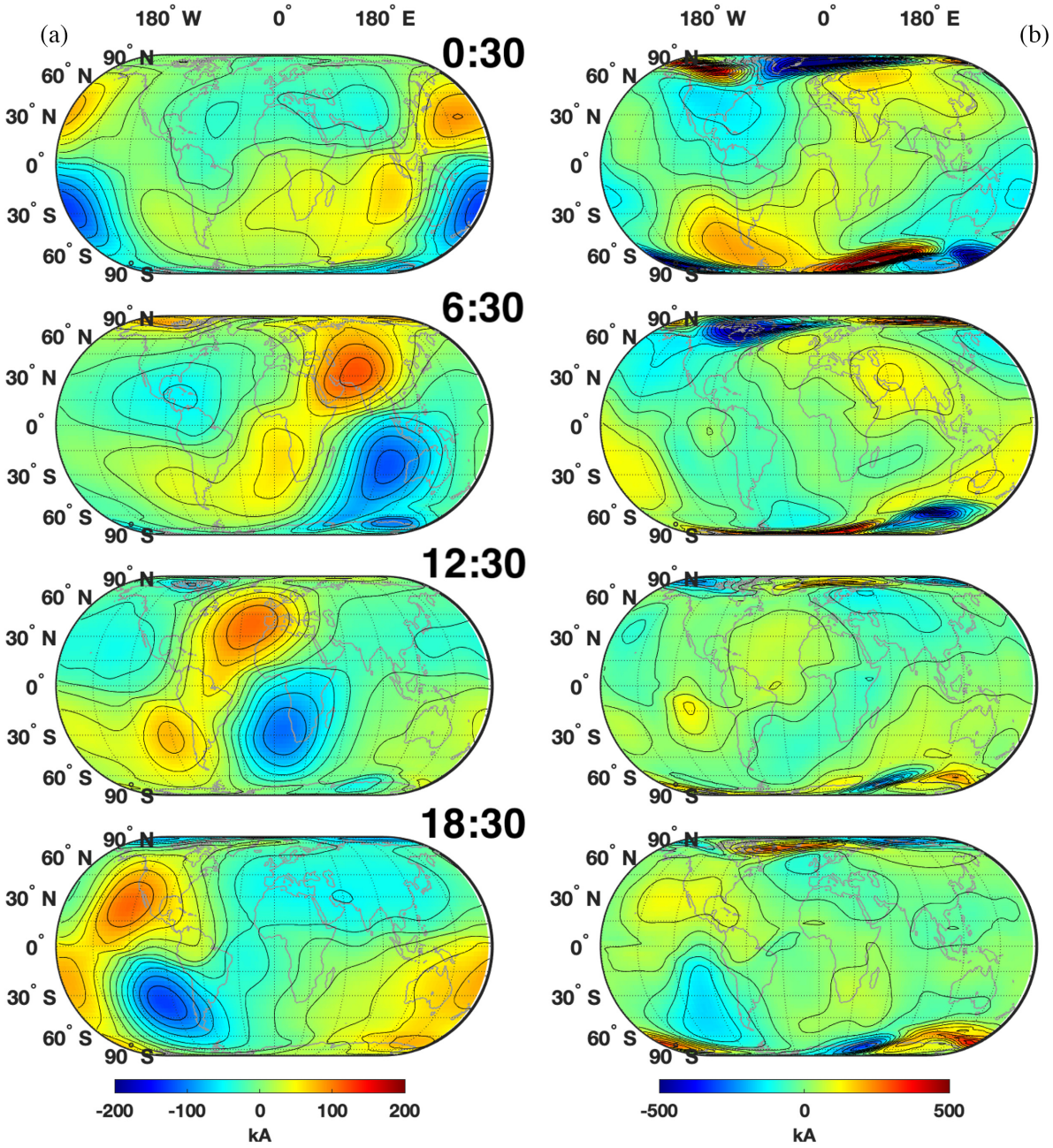

Figure 19. Estimated equivalent sheet current stream functions, sampled at 6 hr intervals, for two days from the time-series plotted in Fig. 18. (a) 9/24/2002, a quiet day. (b) 10/02/2002, from near the start of the active interval. Stream function units are kAmp; note that different colour scales are used for the two days. From top to bottom times are 0:30, 6:30, 12:30, 18:30 UT.

and the number of modes used in each band, would in fact be a major refinement of the DV model, leading to further reduction in the total number of spatial and temporal modes used to represent the full model [e.g. eq. (10) with variable K_j]. This in turn will result in fewer coefficients β_{jki}, and more stable estimation of these parameters with satellite data. These are refinements to be implemented in future work.

Finally, we return to our original motivation, with a few general comments on using the estimated source fields for induction studies. Our DV source model includes a large number of modes in each band, as we seek to represent a large fraction of the signal for both quiet and active conditions. However, the first few modes correspond to the largest signals and will be less contaminated by noise. They are also typically the simplest, and most easily modelled using TIEGCM basis functions (Fig. 13). Furthermore, as demonstrated by the comparison of data and model in Fig. 18, sources are simplest, and most readily modelled using TIEGCM basis functions during quiet conditions. Thus, while all of the sources (represented by Ψ_{jk}^{ext}) contain some information about Earth conductivity, a subset (or more properly, a subspace) of these modes are likely to have the most accurately modelled sources, and be most useful. Thus, a first step in using the DV model for induction studies may be to extract the dominant quiet time source subspace. This is easily done within our general framework: for each band j linear combinations of the modes $k = 1$, K_j ($\mathbf{U}_{jk}$, $\mathbf{a}_{kj}$, with source estimates Ψ_{jk}) that have the greatest power during quiet conditions are sought. These can be

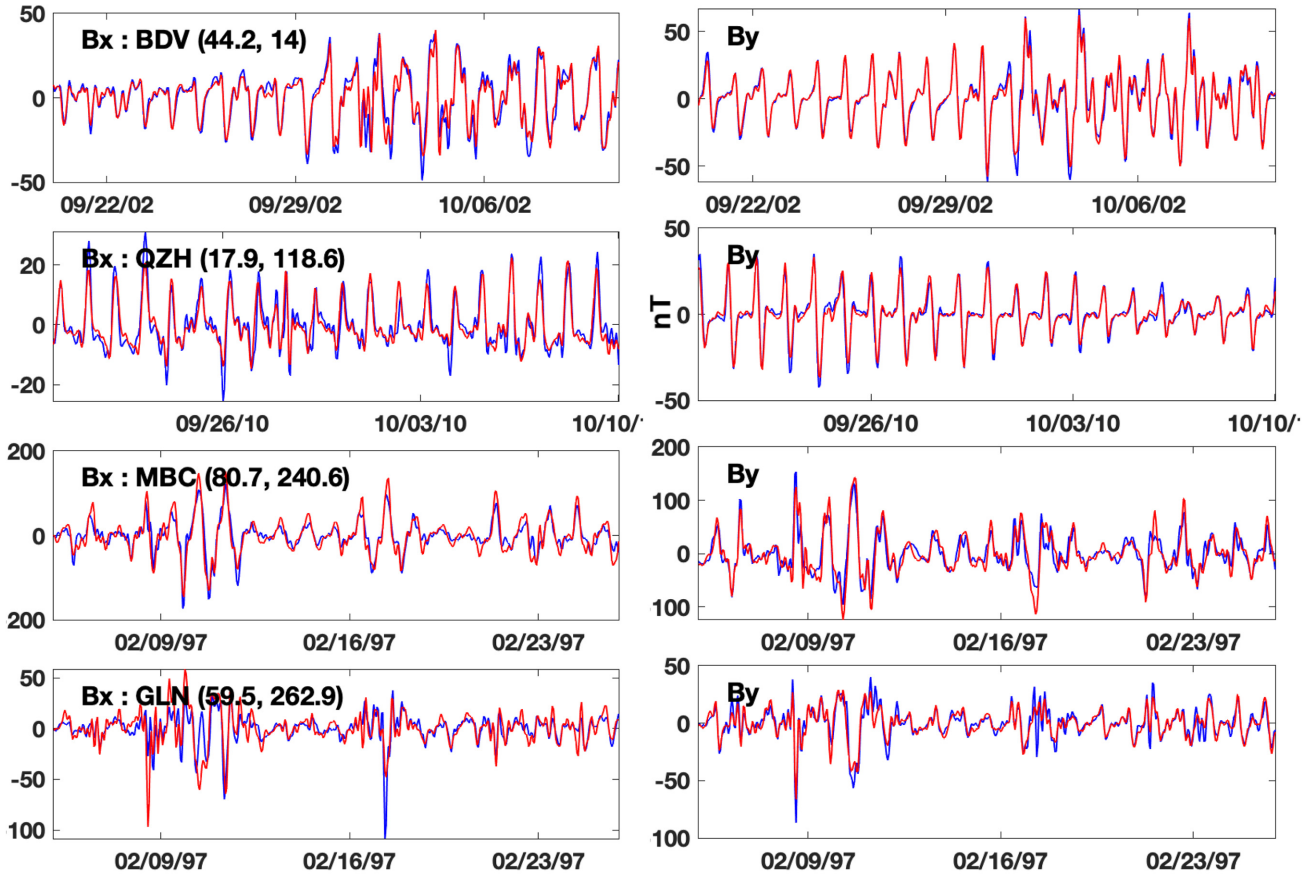

Figure 20. Horizontal components of magnetic fields (in nT) for 20 d at some observatories that were not used to construct the model (magenta triangles in Fig. 2), comparing raw observatory data (blue lines), and the global time domain model (red lines). With the exception of the European observatory BDV, the observatory data were not available for much of the 21 yr window, so different time windows are used for different sites. QD latitude and geographic longitude are shown in left-hand panels.

derived from the temporal coefficients $\mathbf{a}_{kj}$ (which carry amplitude and phase of each mode), restricted to segments with, for example, $K_p < 1$. Details will be given elsewhere, where we use our DV model for mantle induction studies.

In summary, we have developed new methods for empirical modelling of external source magnetic fields, and applied these methods to construct a preliminary model of DVs (periods of 4 hr to 2 d) of surface magnetic fields. The model is global, has a nominal sample rate of 1 hr, and extends over the time period 1997–2018. Efforts to use this model for induction and ionospheric studies, and to incorporate satellite data into a fully 3-D model of ionospheric magnetic fields are ongoing, and will be reported separately.

ACKNOWLEDGEMENTS

This work was partially supported by NSF Grants EAR-1447109 (GDE), EAR-1447036 (PA), EAR-1446856 (AM) and NASA grant NNX16AK88G (GDE, PA, AM). Part of this material is based upon work supported by the National Center for Atmospheric Research, a major facility supported by NSF under Cooperative Agreement No. 1852977. We acknowledge high-performance computing support from Cheyenne (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by NSF. We thank the national institutes that support magnetic observatories and INTERMAGNET for promoting high standards of magnetic observatory practice (http://www.intermagnet.org). We especially thank Art Richmond for useful discussions in the early phases of this work, and for a careful review of this paper. Reviews from Alexander Grayver and an anonymous reviewer have helped us improve clarity of this paper.

DATA AVAILABILITY

The simulated ground magnetic perturbations of this study are archived at the NCAR Digital Asset Service Hub (https://doi.org/10.5065/akxe-2a56) where they are available for download. Observatory time-series used were obtained from publicly available sources, as cited.

REFERENCES

Alken, P., Maute, A., Richmond, A., Vanhamäki, H. & Egbert, G., 2017. An application of principal component analysis to the interpretation of ionospheric current systems, *J. geophys. Res.,* **122**(5), 5687–5708.

Alken, P., Olsen, N. & Finlay, C.C., 2020. Co-estimation of geomagnetic field and in-orbit fluxgate magnetometer calibration parameters, *Earth, Planets Space,* **72,** 1–32.

Banks, R., 1969. Geomagnetic variations and the electrical conductivity of the upper mantle, *J. geophys. Int.,* **17**(5), 457–487.

Dai, L. & Karato, S.-I., 2009. Electrical conductivity of wadsleyite at high temperatures and high pressures, *Earth planet. Sci. Lett.,* **287**(1–2), 277–283.

Dong, S.-W., *et al.*, 2013. Progress in deep lithospheric exploration of the continental China: a review of the Sinoprobe, *Tectonophysics,* **606,** 1–13.
Egbert, G., Alken, P., Maute, A. & Richmond, A., 2019. Modeling magnetic field variations in the daily variation band with ground-based and satellite data, in *Proceedings of the AGU Fall Meeting 2019,* AGU.
Egbert, G.D., 1989. Multivariate analysis of geomagnetic array data: 2. Random source models, *J. geophys. Res.,* **94**(B10), 14 249–14 265.
Egbert, G.D., 1997a. Robust multiple-station magnetotelluric data processing, *J. geophys. Int.,* **130**(2), 475–496.
Egbert, G.D., 1997b. Tidal data inversion: interpolation and inference, *Prog. Oceanogr.,* **40**(1–4), 53–80.
Egbert, G.D., 2002. Processing and interpretation of electromagnetic induction array data, *Surv. Geophys.,* **23**(2–3), 207–249.
Egbert, G.D. & Booker, J.R., 1986. Robust estimation of geomagnetic transfer functions, *J. geophys. Int.,* **87**(1), 173–194.
Egbert, G.D. & Booker, J.R., 1989. Multivariate analysis of geomagnetic array data: 1. The response space, *J. geophys. Res.,* **94**(B10), 14 227–14 247.
Egbert, G.D., Booker, J.R. & Schultz, A., 1992. Very long period magnetotellurics at Tucson observatory: estimation of impedances, *J. geophys. Res.,* **97**(B11), 15 113–15 128.
Emery, B., *et al.*, 2012, Parameterization of the ion convection and the auroral oval in the NCAR Thermospheric General Circulation Models, Tech. rep., National Center for Atmospheric Research, Boulder, CO, USA.
Everett, M.E., Constable, S. & Constable, C.G., 2003. Effects of near-surface conductance on global satellite induction responses, *J. geophys. Int.,* **153**(1), 277–286.
Finlay, C.C., Olsen, N., Kotsiaros, S., Gillet, N. & Tøffner-Clausen, L., 2016. Recent geomagnetic secular variation from swarm and ground observatories as estimated in the chaos-6 geomagnetic field model, *Earth, Planets Space,* **68**(1), 112.
Fujii, I. & Schultz, A., 2002. The 3d electromagnetic response of the earth to ring current and auroral oval excitation, *J. geophys. Int.,* **151**(3), 689–709.
Gabriel, K.R. & Zamir, S., 1979. Lower rank approximation of matrices by least squares with any choice of weights, *Technometrics,* **21**(4), 489–498.
Gjerloev, J.W., 2012. The supermag data processing technique, *J. geophys. Res.,* **117,** doi:10.1002/2017JA024543.
Guzavina, M., Grayver, A. & Kuvshinov, A., 2019. Probing upper mantle electrical conductivity with daily magnetic variations using global-to-local transfer functions, *J. geophys. Int.,* **219**(3), 2125–2147.
Haines, G. & Torta, J., 1994. Determination of equivalent current sources from spherical cap harmonic models of geomagnetic field variations, *J. geophys. Int.,* **118**(3), 499–514.
Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction,* Springer Science & Business Media.
Häusler, K., Hagan, M.E., Forbes, J.M., Zhang, X., Doornbos, E., Bruinsma, S. & Lu, G., 2015. Intraannual variability of tides in the thermosphere from model simulations and in situ satellite observations, *J. geophys. Res.,* **120**(1), 751–765, 2014JA020579.
Heelis, R.A., Lowell, J.K. & Spiro, R.W., 1982. A model of the high-latitude ionospheric convection pattern, *J. geophys. Res.,* **87**(A8), 6339–6345.
Hirschmann, M. & Kohlstedt, D., 2012. Water in earths, *Phys. Today,* **65**(3), 40.
Hirschmann, M.M., 2006. Water, melting, and the deep earth H_2O cycle, *Annu. Rev. Earth planet. Sci.,* **34,** 629–653.
Huang, X., Xu, Y. & Karato, S.-I., 2005. Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite, *Nature,* **434**(7034), 746–749.
Karato, S.-I., 2011. Water distribution across the mantle transition zone and its implications for global material circulation, *Earth planet. Sci. Lett.,* **301**(3–4), 413–423.
Karato, S.-I. & Jung, H., 1998. Water, partial melting and the origin of the seismic low velocity and high attenuation zone in the upper mantle, *Earth planet. Sci. Lett.,* **157**(3–4), 193–207.
Karato, S.-I. & Jung, H., 2003. Effects of pressure on high-temperature dislocation creep in olivine, *Philos. Mag.,* **83**(3), 401–414.
Kelbert, A., Egbert, G.D. & Schultz, A., 2008. Non-linear conjugate gradient inversion for global em induction: resolution studies, *J. geophys. Int.,* **173**(2), 365–381.
Kelbert, A., Schultz, A. & Egbert, G., 2009. Global electromagnetic induction constraints on transition-zone water content variations, *Nature,* **460**(7258), 1003–1006.
Koch, S. & Kuvshinov, A., 2013. Global 3-d em inversion of sq variations based on simultaneous source and conductivity determination: concept validation and resolution studies, *J. geophys. Int.,* **195**(1), 98–116.
Koyama, T., Khan, A. & Kuvshinov, A., 2014. Three-dimensional electrical conductivity structure beneath australia from inversion of geomagnetic observatory data: evidence for lateral variations in transition-zone temperature, water content and melt, *J. geophys. Int.,* **196**(3), 1330–1350.
Kuvshinov, A. & Olsen, N., 2006. A global model of mantle conductivity derived from 5 years of CHAMP, Ørsted, and SAC-C magnetic data, *Geophys. Res. Lett.,* **33**(18), doi:10.1029/2006GL027083.
Love, J.J. & Rigler, E.J., 2014. The magnetic tides of Honolulu, *J. geophys. Int.,* **197**(3), 1335–1353.
Macmillan, S. & Olsen, N., 2013. Observatory data and the Swarm mission, *Earth, Planets Space,* **65**(11), 15.
Manoj, C., Kuvshinov, A., Maus, S. & Lühr, H., 2006. Ocean circulation generated magnetic signals, *Earth, Planets Space,* **58**(4), 429–437.
Maus, S., *et al.*, 2008. Resolution of direction of oceanic magnetic lineations by the sixth-generation lithospheric magnetic field model from CHAMP satellite magnetic measurements, *Geochem. Geophys. Geosyst.,* **9**(7), doi:10.1029/2008GC001949.
Maute, A. & Richmond, A.D., 2017. Examining the magnetic signal due to gravity and plasma pressure gradient current with the TIE-GCM, *J. geophys. Res.,* **122**(12), 12 486–12 504.
Maute, A. & Richmond, A.D., 2017. F-region dynamo simulations at low and mid-latitude, *Space Sci. Rev.,* **206,** 471–493.
Mei, S. & Kohlstedt, D.L., 2000. Influence of water on plastic deformation of olivine aggregates: 1. Diffusion creep regime, *J. geophys. Res.,* **105**(B9), 21 457–21 469.
Meqbel, N.M., Egbert, G.D., Wannamaker, P.E., Kelbert, A. & Schultz, A., 2014. Deep electrical resistivity structure of the northwestern us derived from 3-D inversion of USarray magnetotelluric data, *Earth planet. Sci. Lett.,* **402,** 290–304.
Olsen, N., 1998. The electrical conductivity of the mantle beneath Europe derived from c-responses from 3 to 720 hr, *J. geophys. Int.,* **133**(2), 298–308.
Olsen, N., Vennerstrøm, S. & Friis-Christensen, E., 2003. Monitoring magnetospheric contributions using ground-based and satellite magnetic data, in *First CHAMP Mission Results for Gravity, Magnetic and Atmospheric Studies,* pp. 245–250, Springer.
Olsen, N., Lühr, H., Finlay, C.C., Sabaka, T.J., Michaelis, I., Rauberg, J. & Tøffner-Clausen, L., 2014. The chaos-4 geomagnetic field model, *J. geophys. Int.,* **197**(2), 815–827.
Olsen, N., Ravat, D., Finlay, C.C. & Kother, L.K., 2017. LCS-1: A high-resolution global model of the lithospheric magnetic field derived from CHAMP and Swarm satellite observations, *J. geophys. Int.,* **211**(3), 1461–1477.
Portnoff, M., 1980. Time-frequency representation of digital signals and systems based on short-time Fourier analysis, *IEEE Trans. Acoust., Speech, Signal Process.,* **28**(1), 55–69.
Püthe, C. & Kuvshinov, A., 2013. Determination of the 1-D distribution of electrical conductivity in earths mantle from swarm satellite data, *Earth, Planets Space,* **65**(11), 4.
Püthe, C., Kuvshinov, A., Khan, A. & Olsen, N., 2015. A new model of earth's radial conductivity structure derived from over 10 yr of satellite and observatory magnetic data, *Geophys. Suppl. Mon. Not. R. Astron. Soc.,* **203**(3), 1864–1872.
Qian, L., *et al.*, 2014. The NCAR TIE-GCM: A community model of the coupled thermosphere/ionosphere system, in *Modeling the Ionosphere-Thermosphere System,* Geophys. Monogr. Ser., pp. 73–84, eds Huba, J., Schunk, R. & Khazanov, G., John Wiley & Sons.
Richmond, A., 1995. Ionospheric electrodynamics using magnetic apex coordinates, *J. Geomag. Geoelectr.,* **47**(2), 191–212.

Richmond, A., Maute, A. & Khazanov, G., 2014, Ionospheric electrodynamics modeling, in *Modeling the Ionosphere-Thermosphere System,* pp. 57–73, eds Huba, J. & Schunk, R., AGU Geophysical Monograph Series.

Richmond, A., Ridley, E. & Roble, R., 1992. A thermosphere/ionosphere general circulation model with coupled electrodynamics, *Geophys. Res. Lett.,* **19**(6), 601–604.

Roble, R. & Ridley, E., 1994. A thermosphere-ionosphere-mesosphere-electrodynamics general circulation model (time-gcm): equinox solar cycle minimum simulations (30–500 km), *Geophys. Res. Lett.,* **21**(6), 417–420.

Russell, C.T., Chi, P.J., Dearborn, D.J., Kuo-Tiong, B., Means, J.D., Pierce, D.R., Rowe, K.M. & Snare, R.C., 2008. Themis ground-based magnetometers, *Space Sci. Rev.,* **141,** 389–412.

Sabaka, T.J., Olsen, N. & Langel, R.A., 2002. A comprehensive model of the quiet-time, near-earth magnetic field: phase 3, *J. geophys. Int.,* **151**(1), 32–68.

Sabaka, T.J., Olsen, N. & Purucker, M.E., 2004. Extending comprehensive models of the Earth's magnetic field with Ørsted and CHAMP data, *J. geophys. Int.,* **159**(2), 521–547.

Sabaka, T.J., Olsen, N., Tyler, R.H. & Kuvshinov, A., 2015. CM5, a pre-Swarm comprehensive geomagnetic field model derived from over 12 yr of CHAMP, Ørsted, SAC-C and observatory data, *J. geophys. Int.,* **200**(3), 1596–1626.

Schmucker, U., 1999. A spherical harmonic analysis of solar daily variations in the years 1964–1965: response estimates and source fields for global induction II. Results, *J. geophys. Int.,* **136**(2), 455–476.

Schnepf, N.R., Nair, M., Maute, A., Pedatella, N.M., Kuvshinov, A. & Richmond, A.D., 2018. A comparison of model-based ionospheric and ocean tidal magnetic signals with observatory data, *Geophys. Res. Lett.,* **45**(15), 7257–7267.

Semenov, A. & Kuvshinov, A., 2012. Global 3-D imaging of mantle conductivity based on inversion of observatory c-responses. II. Data analysis and results, *J. geophys. Int.,* **191**(3), 965–992.

Shore, R., Whaler, K., Macmillan, S., Beggan, C., Velímskỳ, J. & Olsen, N., 2016. Decadal period external magnetic field variations determined via eigenanalysis, *J. geophys. Res.,* **121**(6), 5172–5184.

Smirnov, M.Y. & Egbert, G.D., 2012. Robust principal component analysis of electromagnetic arrays with missing data, *J. geophys. Int.,* **190**(3), 1423–1438.

Sun, J. & Egbert, G.D., 2012. A thin-sheet model for global electromagnetic induction, *J. geophys. Int.,* **189**(1), 343–356.

Sun, J., Kelbert, A. & Egbert, G.D., 2015. Ionospheric current source modeling and global geomagnetic induction using ground geomagnetic observatory data, *J. geophys. Res.,* **120**(10), 6771–6796.

Thébault, E., Vigneron, P., Langlais, B. & Hulot, G., 2016. A swarm lithospheric magnetic field model to sh degree 80, *Earth, Planets Space,* **68**(1), 1–13.

Thiel, S., Reid, A., Heinson, G. & Robertson, K., 2016. Insights into lithospheric architecture, fertilisation and fluid pathways from Auslamp Mt, *ASEG Extended Abstracts,* **2016**(1), 1–6.

Tyler, R.H., Boyer, T.P., Minami, T., Zweng, M.M. & Reagan, J.R., 2017. Electrical conductivity of the global ocean, *Earth, Planets Space,* **69**(1), 1–10.

Velímskỳ, J., 2010. Electrical conductivity in the lower mantle: constraints from champ satellite data by time-domain em induction modelling, *Phys. Earth planet. Inter.,* **180**(3–4), 111–117.

Wang, D., Mookherjee, M., Xu, Y. & Karato, S.-I., 2006. The effect of water on the electrical conductivity of olivine, *Nature,* **443**(7114), 977–980.

Wang, H., Egbert, G., Yao, Y. & Cheng, J., 2020. Array analysis of magnetic and electric field observatories in china: estimation of magnetotelluric impedances at very long periods, *J. geophys. Int.,* **222**(1), 305–326.

Yoshino, T., Manthilake, G., Matsuzaki, T. & Katsura, T., 2008. Dry mantle transition zone inferred from the conductivity of wadsleyite and ringwoodite, *Nature,* **451**(7176), 326–329.

APPENDIX A: NOTATION

APPENDIX B: SHORT TIME FOURIER TRANSFORM

To transform from the TD to FD we use a windowed short-time Fourier transform (STFT, e.g. Portnoff 1980), which is readily inverted. For completeness we give key details of our approach here. In the first processing step, vector time-series $\mathbf{X}(t)$ are high-pass (HP) filtered using an autoregressive-moving average (ARMA) filter, with a cut-off frequency of $4\Delta t N_{\rm win}$ (where $\Delta t = 3600$ s is the sampling rate, and $N_{\rm win} = 192$ is the window length) to obtain $\mathbf{X}^1(t)$. The filter is run forward, then backward, in time, so that there is no phase shift. $\mathbf{X}^1(t)$ is multiplied by a series of overlapping taper functions $W_n(t), \quad n = 1, ..., N$, each a standard taper of length $N_{\rm win}$, shifted by $N_{\rm win}/2 = 4$ d. The sequence of tapered data windows $\mathbf{X}^1(t)W_n(t)$ are then transformed to the FD data vectors $\mathbf{X}_{fn}$ with a standard FFT. If more than two data points are missing in a segment, the entire segment is treated as missing. Clearly, by inverse FT we can recover each tapered data window, and hence

$$\sum_{n=1}^{N} \mathbf{X}^1(t) W_n(t) = \mathbf{X}^1(t) \sum_{n=1}^{N} W_n(t). \tag{B1}$$

Provided $\sum_{n=1}^{N} W_n(t)$ does not vanish we can recover $\mathbf{X}^1(t)$ by dividing this out. For a Hanning window with 50 per cent overlap the sum is actually close to 1 everywhere already, except at the beginning and end of the time-series. To simplify the back-transform we have slightly modified the Hanning window so that the sum is identically 1, except at the start and end of the time-series. Note however, that any window/overlap can be used, provided the sum over windows remains non-zero. This very simple scheme can be viewed as a discrete variant on the inverse of the continuous time STFT.

Note that the residual $\mathbf{R}^1 = \mathbf{X} - \mathbf{X}^1$ can be processed in the same way as the original time-series, HP filtering $\mathbf{R}^1$ to construct $\mathbf{X}^2$ with the cut-off frequency reduced further, and then using longer data tapers for the FT—for example if the low-frequency cut-off is reduced by 4, the taper length would be increased by 4. Another set of FCs, extending to lower frequencies, would be obtained by FT of these longer data segments, which could be inverted to recover $\mathbf{X}^2$. The original time-series could then be recovered as $\mathbf{X}(t) = \mathbf{X}^1(t) + \mathbf{X}^2(t) + \mathbf{R}^2(t)$. This obviously could be extended by repeating this process for multiple levels, limited only by the length of the original time-series. For this study, we used only a single HP filter step, as with hourly mean data and $N_{\rm win} = 192$, $\mathbf{X}^1(t)$ contains all frequencies above 0.5 cpd, covering the full DV band of interest. However, for modelling magnetic variations over a broader period range, the multilevel scheme sketched here will be useful.

APPENDIX C: MSDEMPCA: PCA WITH MISSING DATA

Egbert (1997a) and Smirnov & Egbert (2012) describe an approach to PCA that incorporates robust statistical estimation methods, and allows for even large blocks of missing data. This approach to PCA, which is sometimes referred to as criss-cross regression (Gabriel & Zamir 1979), is based on the observation that if one set of model parameters in the basic PCA equation (eq. (1), repeated here for

Table A1. Summary of notation.

Indices and limits		
Index	Limit	
s	N_s	Site number
c	N_c	Component number (channel)
n	N	Segment (window) number
f	$N_{win}/2$	Frequency for DFT; window of length N_win
j	J	Frequency band
B_j	P_j	Frequency indices (f) in band j
k	K_j	Data PCA mode, number of modes used for band j
l	L	SH degree (representation of model modes)
m	M	SH order (representation of model modes)
i	I_j	Model PCA mode, number of modes used for band j
Vectors		
Vector	Dimension	
$\mathbf{X}_{fn}$	N_c	FD data vector (complex)
$\mathbf{U}_{jk}$	N_cD	Data spatial mode (complex)
$\mathbf{a}_{kj}$	NP_j	Data temporal mode (complex)
Φ_{ji}	$M(M+2)+(L-M)(2M+1)$	Model modes (SH potential coefficients)
Ψ^{ext}_{jk}	$M(M+2)+(L-M)(2M+1)$	External potential corresponding to $\mathbf{U}_{jk}$
Ψ^{int}_{jk}	$M(M+2)+(L-M)(2M+1)$	Internal potential corresponding to $\mathbf{U}_{jk}$
$\mathbf{b}_{jk}=(\beta_{jk1},\ldots\beta_{jkI_j})$	I_j	Model mode coefficients for band j, mode k
$\mathbf{y}_{ji}$	N_c	$\mathbf{F}_{f,\sigma}\left[\Phi^{ext}_{ji}\right]$ evaluated at N_s observatory locations
Matrices		
matrix	Dimension	
$\mathbf{X}_j$	$N_c \times NP_j$	FD data matrix, band j
$\mathbf{Y}_j$	$N_c \times I_j$	$\left[\mathbf{y}_{j1}\ldots\mathbf{y}_{jI_j}\right]$
Operators		
Operator	Input	Output
$\mathbf{F}_{f,\sigma}$	Ψ^{ext}	$\mathbf{B}(\theta,\phi)$ = total surface magnetic field

convenience)

$$\mathbf{X}_{fn}=\sum_{k=1}^{K_j}\mathbf{U}_{jk}a_{kfn}+\mathbf{e}_{fn},\quad f\in B_j, \tag{C1}$$

is known (e.g. the spatial modes $\mathbf{U}_{jk}$) the model is linear in the other set (the temporal mode coefficients $\mathbf{a}_{kj}$). Once initialized, we thus can alternately update estimates of temporal and spatial mode parameters using a robust estimator for the linear model, such as the regression M-estimate (e.g. Egbert & Booker 1986). Clearly, these linear estimation steps can be applied even if there are significant amounts of missing data. Following Smirnov & Egbert (2012) we refer to our implementation of this scheme as 'MsDEMPCA', short for missing data EM PCA. To initialize the iterative scheme we start from an estimate of the spatial modes $\mathbf{U}_{kj}$ obtained by applying a more conventional PCA approach (e.g. SVD) to a core array, with fewer sites and or time segments, but little or no missing data. MsDEMPCA then builds up the array, successively extending to more segments and more sites until all data are incorporated. Further details on MsDEMPCA can be found in Smirnov & Egbert (2012). An example application to a large array of electrical and magnetic sensors covering much of China can be found in Wang *et al*. (2020).

The MsDEMPCA scheme also computes estimates of the incoherent noise covariance for each band

$$\mathbf{Cov}(\mathbf{e}_{fn})=\mathbf{diag}(\sigma^2_{cj},c=1,N_c), \tag{C2}$$

which is assumed diagonal (noise is uncorrelated between channels) for simplicity. FCs for band j for each channel ($c=1,N_c$) are separately fit using FCs from all other sites, and residual variances are used [with some correction; see Egbert (1997a) for details] to estimate the component noise variances σ^2_{cj}. Because the error vector in eq. (C1) includes signal omitted due to PCA truncation as well as actual noise, the assumption of a diagonal error covariance is certainly an oversimplification. However, the incoherent noise estimates do provide a measure of how well the model fits each data channel, as illustrated in Fig. 6. Noise standard deviations are used to normalize data channels so that noise levels are uniform across the array, and to provide optimal weights for estimation of the temporal mode coefficients. With this normalization components of the data matrix $\mathbf{X}$ are non-dimensional, as are components of both spatial and temporal modes. Then with the spatial mode vectors all chosen to have unit amplitude, relative signal amplitude (variations over time, and between modes) is carried in the temporal coefficient vectors $\mathbf{a}_{kj}$, which represent a signal-to-noise ratio (SNR)—that is the total signal power, relative to the average background noise level for each mode in each time window. The variance of $\mathbf{a}_{kj}$ defines the average SNR (over all time segments and all channels) of mode k for band j, as displayed in Fig. 4. To make the relative amplitudes of components of the spatial modes physically meaningful we must restore the actual physical units, by multiplying each component of the estimated spatial mode unit vector by the corresponding noise SD, that is multiply channel c by σ_{cj}, for all modes in band j. Thus, in the model of eq. (1) the spatial modes carry units [which in some cases, such as in the MT example of Wang *et al.* (2020), can be different for different channels], and the temporal modes define variations of SNR, both with time, and between modes. However, note that while the spatial modes have units (nT/Hz$^{1/2}$ here) the overall amplitudes are still controlled in part by the initial normalization, and so really only relative amplitudes (and directions and phases) are meaningful. Thus, for example, all spatial modes have similar overall amplitude.

Finally note that the estimates of noise variances σ_{cj}^2 can also be used to provide very rough error estimates for the spatial mode components. With temporal mode coefficients taken as known, the linear problems for each component ($c = 1, N_c$) of $\mathbf{U}_{jk}$ are decoupled

$$X_{fn}^c = \sum_{k=1}^{K_j} U_{jk}^c a_{kfn} + e_{fn}^c, \quad f \in B_j. \tag{C3}$$

The design matrix for the linear problem (call this $\mathbf{A}_{cj}$) is constructed from the temporal mode coefficients, and is nominally of dimension $N_c P_j \times K_j$. If some time segments (index n) are missing, the actual number of rows in the matrix will be smaller than $N_c P_j$. Following standard theory, the error covariance for estimates of the regression parameters (here, $U_{jk}^c, k = 1, K_j$) is $\sigma_{cj}^2 \left[\mathbf{A}_{cj}^\dagger \mathbf{A}_{cj}\right]^{-1}$. We use the diagonal elements of these covariance estimates to define approximate error bars for the spatial mode components, which we use in the fitting procedure described in Section 4. While this represents only a rough approximation to the true error (for example, it ignores the fact that the design matrix elements are also estimated), we believe that this approach does provide reasonable estimates of the relative magnitude of errors for different sites/channels. In particular, the error estimates do account for how well the PCA model fits the data for each channel/site—which is quite variable and depends systematically on a site's latitude and the magnetic component (Fig. 6). The estimated errors also properly account for the distribution of missing data. Thus, components at observatories with only a few years of data will have larger variances.

APPENDIX D: PCA OF MODEL OUTPUTS

Below the ionosphere, the magnetic field produced by ionospheric sources can be represented as the gradient of an external potential function, $\mathbf{B} = -\nabla V^{ext}$, with

$$V^{ext}(r, \theta, \phi, t) = a \sum_{l=1}^{L} \sum_{m=-M_l}^{M_l} \psi_l^m(t) \left(\frac{r}{a}\right)^l Y_l^m(\theta, \phi), \tag{D1}$$

where r, θ, ϕ are the usual radius, colatitude and longitude in a geocentric spherical coordinate system, $a = 6371.2$ km is the mean Earth radius, $Y_l^m(\theta, \phi)$ are complex spherical harmonics of degree l and order m, $\psi_l^m(t)$ are time-dependent spherical harmonic coefficients, L is the maximum spherical harmonic degree used in the expansion, $M_l = \min(l, M)$, and M the maximum spherical harmonic order. We chose maximal spherical harmonic degree $L = 60$ and order $M = 30$ to represent the ionospheric sources. The magnetic field below the ionosphere is given by

$$\begin{pmatrix} B_r \\ B_\theta \\ B_\phi \end{pmatrix} = -\sum_{l=1}^{L} \sum_{m=-M_l}^{M_l} \psi_l^m(t) \left(\frac{r}{a}\right)^{l-1} \begin{pmatrix} l Y_l^m \\ \partial_\theta Y_l^m \\ \frac{im}{\sin\theta} Y_l^m \end{pmatrix} \tag{D2}$$

We invert the surface magnetic fields produced from TIEGCM for the $\psi_l^m(t)$ time-series with a least squares approach. Since the surface magnetic fields are real valued, the complex SH coefficients will satisfy the symmetry condition $\psi_l^{-m}(t) = (-1)^m \psi_l^m(t)^*$. While the magnetic field perturbations on Earth's surface are generated by a complicated 3-D ionospheric current system, they can be represented by a simpler 2-D equivalent sheet current flowing on a spherical shell of radius b. As detailed in Haines & Torta (1994, eqs 23–25), the equivalent 2-D sheet current density is defined as $\mathbf{K}(\theta, \phi, t) = \nabla \times [\hat{r}\chi(\theta, \phi, t)]$, with the *current stream function* given by

$$\chi(\theta, \phi, t) = -\frac{a}{\mu_0} \sum_{l=1}^{L} \sum_{m=-M_l}^{M_l} \psi_l^m(t) \left(\frac{2l+1}{l+1}\right) \times \left(\frac{b}{a}\right)^l Y_l^m(\theta, \phi). \tag{D3}$$

The set of all SH coefficients (SHC) $\psi_l^m(t)$ appearing in eqs (D1–D3) will be referred to collectively as $\Psi^{ext}(t)$, and used subsequently to define the 2-D external source. With the chosen values of $L = 60$, $M = 30$, the dimension of this vector is 2790, however due to the symmetry condition $\psi_l^{-m}(t) = (-1)^m \psi_l^m(t)^*$, there are 1425 independent (complex valued) parameters in this vector.

To derive the spatial basis functions that we will use for modelling DV fields, we apply PCA to the vector time-series $\Psi^{ext}(t)$ (with components $\psi_l^m(t)$) transformed to the FD. The approach closely follows that used for the ground station analysis. We apply windowed STFT to each component $\psi_l^m(t)$, using exactly the same modified Hanning window length (8 d), and shift (4 d). Since the simulated ionospheric current has a temporal resolution of 1-hr, we have $N_{\text{win}} = 192$, exactly as for the data. The TIEGCM simulations each span 1 yr (2009. solar minimum, and 2002, solar maximum). These were processed separately, giving a total of $N = 85$ windows for each simulation. In terms of span (i.e. the subspace of fields that can be represented as a linear combination of the modes), the two runs were very similar. We thus just used the model modes from the solar maximum simulation, though it would also be simple to do PCA on the combined 2-yr time-series. We note that the previous symmetry condition on the $\psi_l^m(t)$ does not apply to the STFT coefficients, as these are complex. Thus for each window n and frequency f, a full set of 2790 parameters are needed to specify the Fourier coefficients, which we denote $Q_{l,n}^m(f)$. Fixing l, m, f and allowing n to vary, we may consider $Q_{l,n}^m(f)$ as a time-series representing the changing amplitude and phase of $\psi_l^m(t)$ at the frequency f and PCA can be applied to these Fourier coefficients. To do this, for each frequency f we define the matrix

$$\mathbf{Q}(f) = \left(\mathbf{Q}_1(f)\ \mathbf{Q}_2(f) \ldots \mathbf{Q}_N(f),\right) \tag{D4}$$

where the column vectors $\mathbf{Q}_n(f)$ contain the 2790 coefficients $Q_{l,n}^m(f)$. As discussed in Section 2, we merge individual frequency bins into $J = 13$ larger bands, as presented in Table 1. This is done by defining, for $j = 1, \ldots, J$, a set of $2790 \times NP_j$ matrices,

$$\mathbf{X}_j = \left(\mathbf{Q}(f_1) \cdots \mathbf{Q}(f_{P_j})\right), \quad f_1, \ldots, f_{P_j} \in B_j \tag{D5}$$

SVD is applied to each of these matrices $\mathbf{X}_j = \mathbf{U}_j \mathbf{S}_j \mathbf{V}_j^\dagger$; note that in this appendix only these symbols refer to the complex data matrix and singular vectors from TIEGCM outputs; in the main text these symbols represent analogous quantities derived from observatory data. The left singular vectors $\mathbf{U}_j$ are the spatial modes which are the linear combinations of the $\mathbf{Q}_n(f)$ which maximize variance explained in the TIEGCM model run. We denote the columns of $\mathbf{U}_j$ as $\mathbf{\Phi}_{ji}$, with i representing column, or mode number. We will use the leading I_j modes to define spatial basis functions for frequency band j. The elements of these basis functions can be written as $(\Phi_{ji})_l^m$, allowing a direct substitution for $\psi_l^m(t)$ in eqs (D1)–(D3) to determine the magnetic potential, vector magnetic field and stream function corresponding to spatial mode i of frequency band j. Note that in contrast to the data PCA, the right singular vectors and values (which together define temporal variations of amplitude and phase of each $\mathbf{\Phi}_{ji}$ in the model run) are not used.