Geophys. J. Int. (2004) doi: 10.1111/j.1365-246X.2004.02421.x

Extending comprehensive models of the Earth's magnetic field with Ørsted and CHAMP data

Terence J. Sabaka,[1] Nils Olsen[2] and Michael E. Purucker[1]

[1]*Raytheon at Geodynamics Branch, NASA/Goddard Space Flight Center, Greenbelt,* MD 20771, *USA. E-mail: sabaka@geomag.gsfc.nasa.gov*
[2]*Danish Space Research Institute, Juliane Maries Vej* 30, DK - 2100 *Copenhagen Ø, Denmark*

Accepted 2004 July 7. Received 2004 July 7; in original form 2003 June 24

SUMMARY

A new model of the quiet-time, near-Earth magnetic field has been derived using a comprehensive approach, which includes not only POGO and Magsat satellite data, but also data from the Ørsted and CHAMP satellites. The resulting model shows great improvement over its predecessors in terms of completeness of sources, time span and noise reduction in parameters. With its well separated fields and extended time domain of 1960 to mid-2002, the model is able to detect the known sequence of geomagnetic jerks within this frame and gives evidence for an event of interest around 1997. Because all sources are coestimated in a comprehensive approach, intriguing north–south features typically filtered out with other methods are being discovered in the lithospheric representation of the model, such as the S Atlantic spreading ridge and Andean subduction zone lineations. In addition, this lithospheric field exhibits significantly less noise than previous models as a result of improved data selection. The *F*-region currents, through which the satellites pass, are now treated as lying within meridional planes, as opposed to being purely radial. Results are consistent with those found previously for Magsat, but an analysis at Ørsted altitude shows exciting evidence that the meridional currents associated with the equatorial electrojet likely close beneath the satellite. Besides the model, a new analysis technique has been developed to infer the portion of a model parameter state resolved by a particular data subset. This has proven very useful in diagnosing the cause of peculiar artefacts in the Magsat vector data, which seem to suggest the presence of a small misalignment bias in the vector magnetometer.

Key words: electromagnetic induction, geomagnetic secular variation, ionosphere, lithosphere, magnetic field of the Earth, magnetosphere.

1 INTRODUCTION

Within a few thousand kilometres of the surface of the Earth the magnetic field is rich in measurable contributions from several distinct current systems. However, the spatial and temporal scales of some of these constituent fields overlap, making it difficult to separate their effects when sampling the observed field. A successful approach known as comprehensive modelling has been developed to overcome this problem (see Langel *et al.* 1996; Sabaka *et al.* 2002). It entails the parametrization and coestimation of fields associated with the major current sources in the near-Earth regime from field measurements taken from ground-based observatories and satellite mapping missions, taking into account the crucial covariance between these fields. The result is a more proper partitioning of the amalgamous signal among the physical sources in a weighted least-squares sense. These parametrized fields include those of core and lithospheric origin, magnetospheric and ionospheric origin along with associated induced contributions, and toroidal magnetic fields produced by *in situ* poloidal currents that impinge the thin sampling shells of the satellites. Surface and satellite data together facilitate separation via their relative radial orientations with respect to the sources and their spatial and temporal sampling extents. Note that describing the complexities of the geomagnetic field is quite challenging even during magnetically quiet periods and so these studies have been restricted to those periods where the global index of geomagnetic activity, K_p, is roughly $\leq 2^0$.

The most recently published comprehensive model (CM) is CM3 (Sabaka *et al.* 2002), which spanned 1960–1985 and was derived from observatory data as well as data from the POGO and Magsat satellite missions. However, while the POGO data covered years 1965–1971, the data were only scalar intensity, and while Magsat delivered vector data, it was restricted to a 6-month sampling period from 1979 November to 1980 May at only two local times, dawn and dusk. Hence, global data coverage is severely limited in CM3. Finally, after an approximately 20-yr hiatus, the Danish Ørsted satellite was launched in 1999 February and in 2000 July the German CHAMP satellite was launched: both being high-precision scalar and vector magnetic mapping missions. Both satellites are in

© 2004 RAS

near-polar orbits and have already provided several years of quality data over all local times. Clearly, a CM augmented with these data will come much closer to its objective of properly representing the quiet-time, near-Earth magnetic field.

This paper reports on a natural extension of CM3, denoted CM4, in which scalar data from CHAMP and vector and scalar data from Ørsted have been incorporated, along with all available observatory data through 2000. Modifications have been made to the CM3 parametrization in order to accommodate these data and include: (i) an extension of the main field secular variation (SV) basis functions through mid-2002; (ii) *in situ* quasi-dipole (QD) meridional poloidal currents in the Magsat sampling shell; and (iii) *in situ* QD meridional poloidal currents in the Ørsted sampling shell, which are continuous in diurnal time. A comparison of CM3 and CM4 will be given and the above modifications will be further explained. A detailed description of the CM3 parametrization may be found in Sabaka *et al.* (2002).

2 DATA SELECTION

2.1 Observatory data

As CM4 is an extension of CM3, many of the data are common to both. CM3 incorporated observatory hourly means (OHMs) for the quietest day of the month, as determined by K_p, at two sampling rates: (i) the OHM values closest to 01:00 h local time for the entire 1960–1985 span of the model, denoted as OHM_1AM, and (ii) OHMs every 2 hr on those quiet days during the POGO and Magsat missions envelopes, denoted as OHM_MUL. This former rate allows for the determination of broad-scale main field SV, while the latter rate is sufficient for analysing diurnal periods as short as 6 hr. For CM4, the OHM values closest to 01:00 h local time on the quietest day per month were extended through 2000. The spatial and temporal distribution of these data are shown in Fig. 1. Post-1985 temporal distributions are consistent with earlier epochs with a conspicuous rise in reporting stations post-1995. In addition, station breaks (Langel *et al.* 1982) were introduced at times where baseline jumps have occurred. This leads to a total of 340 OHM_1AM and 216 OHM_MUL stations considered in this model.

2.2 Satellite data

Of the four satellite data sets used, the POGO and Magsat scalar data sets are identical to those used in CM3 and are described in Sabaka *et al.* (2002). The Magsat vector data, however, have been reselected for CM4 in order to achieve denser coverage for improved lithospheric field modelling. They have been selected for 20 arcsec accuracy per attitude flags during quiet conditions in which $K_p \leq 1^0$ for the time of observation and $K_p \leq 2^0$ for the previous 3-hr interval. The data have been decimated from the original 16 Hz to 1 min^{-1}. Ørsted and CHAMP data were selected for quiet conditions where $K_p \leq 1^+$ for the time of observation and $K_p \leq 2^0$ for the previous 3-hr interval. In addition, the D_{st} index was required to be within ± 20 nT. The Ørsted data span 1999 March to 2002 July, while the CHAMP data span 2000 August to 2002 July. During this period both satellite orbits precessed through all local times. A comparison of the local

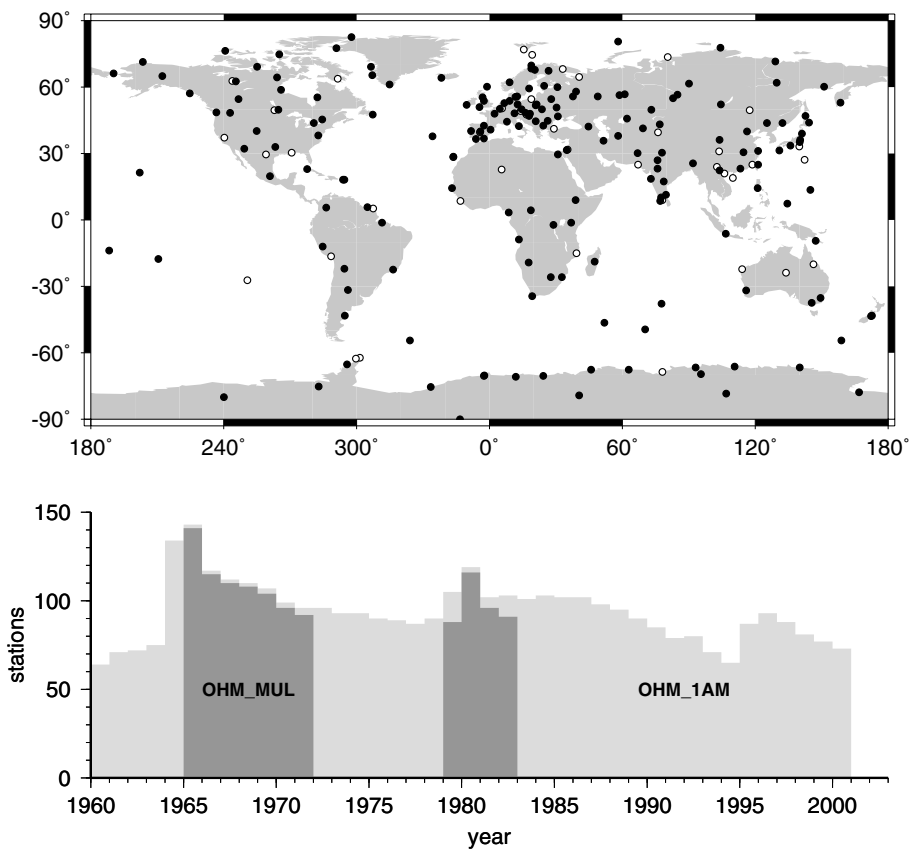

Figure 1. Spatial and temporal distributions of observatory data. The top panel shows observatory locations where only 01:00 h local time hourly means are used (open circles) or both 01:00 h and multiple hourly means are used (black circles). The bottom panel shows a histogram of the number of stations contributing either 01:00 h (OHM_1AM) or multiple hourly (OHM_MUL) means to the particular 1-yr bin.

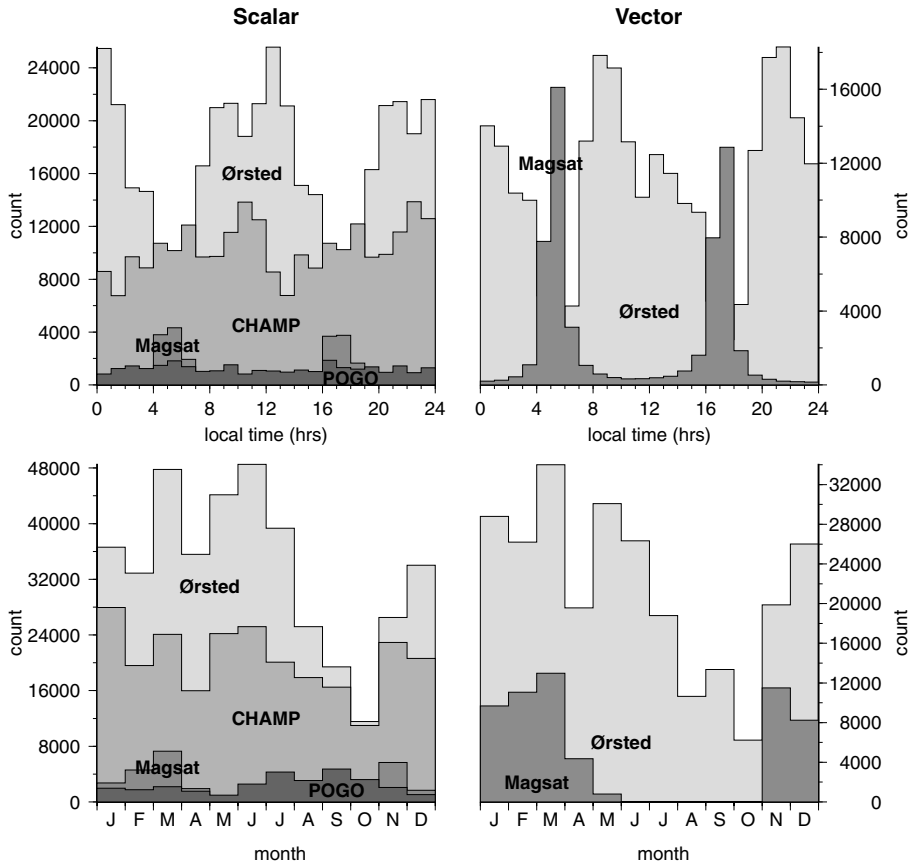

Figure 2. Local time (top) and seasonal (bottom) distributions of scalar (left) and vector (right) satellite data.

time distribution for satellite data is shown in the top two panels of Fig. 2, separately for scalar and vector data. The Ørsted dawn and dusk data are few, but are complemented by Magsat data. The bottom two panels show the seasonal distributions and indicate a general paucity of data during northern autumn. Because the local time of the equatorial crossing of the orbital plane of Ørsted precesses at a rate of approximately 6 hr yr^{-1}, there could be a coupling between seasonal and diurnal effects from these data. However, the data set as a whole probably does not suffer from this. Finally, vector and scalar data were used at all latitudes at a sampling rate of 1 min^{-1} and all satellite data were weighted proportional to $\sin\theta$ (where θ is geographic colatitude) to simulate an equal-area distribution.

To limit contamination from gross outliers, all satellite data were initially culled according to their residuals with respect to CM3. Elimination occurred if the magnitude of the X, Y, Z or F residual was greater than 100, 100, 50 or 40 nT, respectively, where X, Y, Z and F denote the north, east, down and scalar components of the magnetic field, respectively.

3 MODIFIED PARAMETRIZATION

The parametrization of field sources in CM4 closely follows that of CM3. The core and lithospheric fields are together expressed as the negative gradient of a potential function represented by a degree and order 65 internal spherical harmonic (SH) expansion in geographic coordinates, with SV represented by cubic B-splines through degree and order 13 (see Sabaka *et al.* 1997). The knot spacing was kept at 2.5 yr and extended through mid-2002. Thus, for $n \leq 13$ each Gauss coefficient is parametrized by 24 coefficients. Because OHMs are direct rather than derivative measurements of the field, and because of their close proximity to lithospheric and induced sources, a set of static vector biases is solved for each station, including its breaks. These biases have the effect of removing a weighted mean from the OHM residuals at each station (uniform weighting results in the removal of the usual arithmetic mean). If the external and induced portions of the model were perfect, then these biases would represent presumably small-scale lithospheric anomalies whose wave numbers are above the SH truncation level. However, variations exist that are not described well by the model, which contribute to the overall baselines of the residuals. Given that the OHM_1AM and OHM_MUL data sample not only different ranges in local time, but also different segments of the solar cycle, it is expected that their baselines could be very different at the same station. For this reason separate sets of biases are determined for each type of OHM data.

The currents responsible for the ionospheric field are considered to flow in a thin spherical shell at $h = 110$ km altitude. This field is thus expressed as the negative gradient of a potential function at surface and satellites altitudes and is constrained to have radial continuity across the current sheet. The ionospheric parametrization employs harmonic functions endowed with symmetry provided by a QD coordinate system (Richmond 1995), which is aligned with the ambient magnetic field. These functions are mainly sun-synchronous in time, but slightly slower and faster modes are also present with a minimum period of 6 hr, and are modulated with annual and semi-annual seasonal variability. Spatially, they have high QD latitudinal resolution in order to model the equatorial electrojet (EEJ). Induced contributions are accounted for by using an *a priori*

four layer, 1-D, radially varying conductivity model derived from Sq and D_{st} data at selected European observatories (Olsen 1998). The influence of solar activity is represented by an amplification factor, assumed to be equal for all harmonics, which is a function of a three-monthly moving mean of absolute $F_{10.7}$ solar radio flux values (Olsen 1993). This means that increasing solar flux inflates the whole ionospheric and associated induced current system without changing its shape.

The major sources of the magnetospheric field are currents that flow in the magnetotail, magnetopause and ring current complexes. Near the Earth, the field is cast as the negative gradient of a potential function represented by an external SH expansion in dipole coordinates, which has regular daily and seasonal periodicities. Ring current variability is modelled as a linear function of the D_{st} index for external dipole terms only. Because of a lag in availability of the final index, provisional D_{st} has been used for data during and after 2001. The induced contributions of the magnetosphere are treated in a similar manner as the ionosphere and are thus coupled with an internal SH expansion via the same *a priori* conductivity model.

3.1 *F*-region currents

Magsat and Ørsted sample the magnetic field in thin shells centered roughly at 400 and 750 km altitude, respectively. Consequently, these measurements contain contributions from toroidal magnetic fields as a result of poloidal F-region currents $\mathbf{J}$, which couple the ionospheric E region and magnetosphere. Hence, this field is not curl-free and cannot be expressed as the gradient of a potential. In CM3, which analysed Magsat vector data only, these coupling currents were assumed to be purely radial and were only sampled at two local times. However, Olsen (1997) found evidence of a strong J_θ component in Magsat and so for CM4 $\mathbf{J}$ is considered QD meridional. Because the coupling current morphology is also highly influenced by the ambient magnetic field, QD symmetric functions are again used here. Because Ørsted samples all local times, a continuous diurnal representation may now be attempted for toroidal $\mathbf{B}$ in the Ørsted shell. Radial continuity between the Magsat and Ørsted shells will be deferred for future work and so separate parametrizations are used for Magsat dawn and dusk and Ørsted.

The well-known (see Backus 1986; Olsen 1997) toroidal $\mathbf{B}$ and associated poloidal $\mathbf{J}$ may be expressed in spherical coordinates (r, θ, ϕ) as

$$\mathbf{B} = \begin{pmatrix} 0 \\ \frac{1}{\sin\theta}\frac{\partial}{\partial\phi}\Phi \\ -\frac{\partial}{\partial\theta}\Phi \end{pmatrix}, \tag{1}$$

$$\mathbf{J} = \begin{pmatrix} -\Delta_s(r\Phi) \\ \frac{1}{r}\frac{\partial}{\partial\theta}(r\Phi)' \\ \frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}(r\Phi)' \end{pmatrix}\frac{1}{\mu_0}, \tag{2}$$

where Δ_s is the angular part of the Laplacian, the prime indicates differentiation in r and Φ is the toroidal scalar function

$$\Phi(t, \mathbf{r}) = \Re\left\{\sum_{n,m,s,p} \phi^m_{nsp}(r) T^m_{nsp}(t, \theta_q, \phi_q)\right\}. \tag{3}$$

The $\Re\{\cdot\}$ operator used here takes the real part of the complex expression and

$$\begin{aligned} T^m_{nsp}(t, \theta_q, \phi_q) = {} & Y^m_n(\theta_q, \phi_q) \\ & \times \exp is\phi_s(t) \\ & \times \exp ip\phi_p(t_{mut}(t)) \end{aligned} \tag{4}$$

are the QD symmetric basis functions. $Y^m_n(\theta_q, \phi_q)$ is the Schmidt quasi-normalized surface SH function of degree n and order m evaluated at QD coordinates ($\theta_q(\theta, \phi)$, $\phi_q(\theta, \phi)$), but these may be expanded in terms of $Y^k_l(\theta, \phi)$ via spherical transforms. Thus, T^m_{nsp} is QD symmetric on a particular reference sphere and is simply a linear combination of the Y^k_l over k and l, modulated by periodic time functions in QD longitude ϕ_q. The arguments of these time functions include s and p, which are the seasonal and diurnal integer wave numbers, respectively, the seasonal time angle ϕ_s, which has a period of 1 yr and is a function of universal time (UT) t, and the diurnal time angle ϕ_p, which has a period of 24 hr and is a function of magnetic universal time (MUT) t_{mut}. The MUT of an observer is closely related to the observer's magnetic local time (MLT) t_{mlt} defined as

$$t_{mlt}(t) = (180^\circ + \phi_{d,o} - \phi_{d,s}(t))/15, \tag{5}$$

where, if the dipole longitude of the observer, $\phi_{d,o}$, and the sub-solar point, $\phi_{d,s}(t)$, are in degrees, then $t_{mlt}(t)$ is in hours (Langel 1987); note, however, that this reference contains a sign error, which is corrected in Sabaka *et al.* (2002). Thus, MUT is simply the MLT at the dipole prime meridian ($\phi_{d,o} = 0^\circ$), which runs roughly through central S America.

In CM3, the radial dependence of Φ was chosen to be $1/r$, which from eq. (2) leads to $J_\theta = J_\phi = 0$. To obtain QD meridional $\mathbf{J}$, one selects from two classes of admissible Φ: (i) those with a radial dependence of $1/r$ and (ii) those that are QD zonal, i.e. $m = 0$. Clearly, only this second class will contribute to the horizontal portion of the currents $\mathbf{J}_h = (J_\theta, J_\phi)^T$, where the superscript T denotes transposition. To show that $\mathbf{J}_h$ lies in QD meridional planes, let $f(t, r, \theta_q) = (r\Phi(t, r, \theta_q))'$, but from eq. (2)

$$\mathbf{J}_h = \frac{1}{\mu_0}\nabla_h f(t, r, \theta_q), \tag{6}$$

$$= \frac{1}{\mu_0}\frac{\partial f}{\partial\theta_q}\nabla_h\theta_q, \tag{7}$$

where ∇_h is the surface gradient. Therefore, $\mathbf{J}_h$ is a multiple of $\nabla_h\theta_q$ and thus lies in QD meridional planes. Let this component be denoted as J_{θ_q}. The radial dependence of these functions is obtained by following Backus (1986) and Olsen (1997) who consider a Taylor series expansion of ϕ^m_{nsp} around the shell mean radius, b, such that

$$\phi^m_{nsp}(r) = \left(\frac{R}{r}\right)\sum_j \phi^m_{nspj}\frac{\rho^j}{j!}, \tag{8}$$

where $\rho = (r - b)/R$ and R is the reference radius. Only the $j = 0, 1$ terms are retained for the QD zonal terms in the Magsat and Ørsted models and only the $j = 1$ terms contribute to $\mathbf{J}_h$.

The Φ parametrizations used for Magsat dawn and dusk are

$$\begin{aligned} \Phi(t, \mathbf{r}) = \Re\Bigg\{ & \sum_{s=0}^{2}\sum_{n=1}^{40}\sum_{m=0}^{\min(n,4)} \phi^m_{ns00} T^m_{ns0}(t, \theta, \phi) \\ & + \sum_{s=0}^{2}\sum_{n=1}^{40} \rho\phi^0_{ns01} T^0_{ns0}(t, \theta, \phi)\Bigg\}\left(\frac{R}{r}\right), \end{aligned} \tag{9}$$

where $b = 6801.2$ km for dawn and $b = 6786.2$ km for dusk and $R = 6371.2$ km for both. As with CM3, the seasonal phase angle is fixed as a result of the limited seasonal coverage of Magsat. This results in retaining only terms in $\cos s\phi_s(t)$, giving a total of 1164 coefficients in each expansion.

For Ørsted, the Φ parametrization used is

$$\Phi(t, \mathbf{r}) = \Re \left\{ \sum_{s=-2}^{2} \sum_{p=0}^{4} \sum_{m=p-1}^{p+1} \sum_{n=\max(1,|n|)}^{|n|+40} \phi^m_{nsp0} T^m_{nsp}(t, \theta, \phi) + \sum_{s=-2}^{2} \sum_{p=0}^{4} \sum_{n=1}^{40} \rho \phi^0_{nsp1} T^0_{nsp}(t, \theta, \phi) \right\} \left(\frac{R}{r} \right), \quad (10)$$

where now $b = 7121.2$ km and as before $R = 6371.2$ km. Note that the second term, describing $\mathbf{J}_\mathrm{h}$, only contains terms with $m = 0$, whereas the first terms allows for more variability with m and also includes, for instance, terms with $m = p$ (local time terms). The total number of coefficients in the expansion is 6120. These ranges on s, p, n and m were chosen to match those of the ionospheric E-region, which it couples, because Ørsted vector data give full local time coverage.

Because the bulk of toroidal $\mathbf{B}$ is usually perpendicular to the ambient magnetic field, it is assumed that scalar measurements will not be sensitive to it. Thus, only Magsat and Ørsted vector measurements are considered functions of the ϕ^m_{nspj}.

4 ESTIMATION PROCEDURE

4.1 Iteratively reweighted least squares

Let $\mathbf{x}$ be the vector of model parameters collected from each field source and let $\mathbf{a}(\mathbf{x})$ be the model prediction corresponding to the vector of observed field measurements, $\mathbf{d}$. For CM4, $\dim \mathbf{x} = 25\,243$ and $\dim \mathbf{d} = 2\,156\,832$. The number of parameters in each source is given in Table 1. Clearly, the estimation of $\mathbf{x}$ from $\mathbf{d}$ is an overdetermined inverse problem. If the model residuals, $\mathbf{e} = \mathbf{d} - \mathbf{a}(\mathbf{x})$, are Gaussian distributed, then a weighted least-squares estimate, which minimizes the ℓ_2 norm of a vector, would provide the maximum-likelihood estimate. However, Gaussian distributed residuals are surely not the case resulting from instrument errors and unmodelled sources (Walker & Jackson 2000). Instead, the iteratively reweighted least-squares (IRLS) approach with Huber weights is employed here (see Huber 1964; Constable 1988). It is a hybrid method whose cost function contains two terms: $\|\mathbf{e}_G\|_2$ and $\|\mathbf{e}_L\|_1$, where $\mathbf{e}_G$ and $\mathbf{e}_L$ are the subvectors of $\mathbf{e}$ whose distributions are considered Gaussian and Laplacian, respectively. Large residuals are treated as Laplacian to avoid unduly influencing the parameter estimate.

Operationally, ILRS may be cast in the framework of the usual least-squares method by employing a special data weight (inverse covariance) matrix, W, which varies at each iteration in response to changes in $\mathbf{e}_G$ and $\mathbf{e}_L$. This matrix reflects a Huber distribution, which has a Gaussian core ($|e| \le c\sigma$, where e is the random deviate and σ^2 is its variance) and Laplacian tails. Assuming uncorrelated errors (a diagonal W), IRLS assigns Huber weights to the kth data point at the nth iteration as a function of its standard deviation σ_k and current residual value $e_{k,n}$ as

$$w_{k,n} = \frac{1}{\sigma_k^2} \min \left(\frac{c\sigma_k}{|e_{k,n}|}, 1 \right), \quad (11)$$

with $c = 1.5$ (the $w_{k,n}$ for satellite data are multiplied by $\sin\theta$ as stated earlier).

In this study, *a priori* information is to be injected for purposes of physical plausibility and regularization and takes the form of K additional quadratic terms in the cost function. Formally, these ℓ_2 norms describe $\mathbf{x}$ as the realization of K Gaussian processes having covariances of $(\lambda_i\ \Lambda_i)^{-1}$, $i = 1, \ldots, K$. Thus, the damping parameters, λ_i, and Λ_i remain static throughout IRLS and are exempt from modifications as in eq. (11).

The non-linear IRLS cost function

$$\mathcal{L}(\mathbf{x}) = (\mathbf{d} - \mathbf{a}(\mathbf{x}))^T \mathrm{W}(\mathbf{x}) (\mathbf{d} - \mathbf{a}(\mathbf{x})) + \sum_{i=1}^{K} \lambda_i \mathbf{x}^T \Lambda_i \mathbf{x}, \quad (12)$$

is minimized to obtain an estimate of $\mathbf{x}$. At the nth step, the Gauss method (Sorenson 1980) is used to update the current parameter estimate $\mathbf{x}_n$ as

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \left(\mathrm{A}_n^T \mathrm{W}_n \mathrm{A}_n + \sum_{i=1}^{K} \lambda_i \Lambda_i \right)^{-1} \left(\mathrm{A}_n^T \mathrm{W}_n (\mathbf{d} - \mathbf{a}(\mathbf{x}_n)) - \sum_{i=1}^{K} \lambda_i \Lambda_i \mathbf{x}_n \right), \quad (13)$$

where W_n and A_n are the Huber weight matrix and the Jacobian of $\mathbf{a}(\mathbf{x})$ at $\mathbf{x}_n$, respectively. Three iterations were taken in this study and the starting model was CM3 or zero for new parameters.

4.2 Error estimates

It remains to assign the standard deviations σ_k in eq. (11) under the expectation of uncorrelated residuals. For the OHMs, the errors were simply assumed to be uncorrelated and were assigned according to the *a posteriori* residuals from CM3. Thus, for stations poleward of $\pm 50^\circ$ dipole latitude, $\sigma_X = \sigma_Y = 16$ nT and $\sigma_Z = 19$ nT and, for stations equatorward, $\sigma_X = \sigma_Y = \sigma_Z = 9$ nT. For satellite scalar data, $\sigma_F = 4$ nT was used. However, for satellite vector data, the errors in measured X, Y and Z are correlated as a result of attitude inaccuracies.

The correlated covariance matrix of Holme & Bloxham (1996) was employed to account for the anisotropy in the attitude accuracy of the Ørsted vector data (see Holme 2000; Olsen *et al.* 2000; Olsen 2002). It can be shown that the principal components of such a matrix lie in the directions of $\mathbf{B}$, $\hat{\mathbf{n}} \times \mathbf{B}$ and $\mathbf{B} \times (\hat{\mathbf{n}} \times \mathbf{B})$, with an associated coordinate system denoted as $(B, B_\perp, B_3)$, where $\hat{\mathbf{n}}$ is the unit vector in the direction of the boresight of the star imager (SIM) and $\mathbf{B}$ is the observed magnetic field vector. The principal variances are then $(\sigma_B^2, \sigma_\perp^2, \sigma_3^2)$, where $\sigma_\perp^2 = \sigma_B^2 + |\hat{\mathbf{n}} \times \mathbf{B}|^2 \chi^2 + (\hat{\mathbf{n}} \cdot \mathbf{B})^2 \psi^2$ and $\sigma_3^2 = \sigma_B^2 + B^2 \psi^2$. ψ, χ, and σ_B are the standard deviations of the boresight direction, angle about the boresight and scalar intensity, respectively. In this study, $\psi = 10$ arcsec, $\chi = 60$ arcsec before and 40 arcsec after 2000 January 22 reflecting improvements in attitude accuracy and $\sigma_B = 4$ nT consistent with σ_F. For IRLS, the residual vector in the $(B, B_\perp, B_3)$ system and the principal standard deviations are used in eq. (11) and the resulting weight submatrix is rotated back into the (X, Y, Z) system.

Table 1. Number of parameters in each field source.

Field source	Number of parameters
Observatory biases	1635
Core/lithosphere	8840
Ionosphere	5520
Magnetosphere	800
Magsat coupling currents	2328
Ørsted coupling currents	6120
Total	25 243

© 2004 RAS, *GJI*

The same basic scheme is used for Magsat vector data, except that any anisotropic treatment is made difficult by the unavailability of directional information from the attitude determination instruments. Therefore, an isotropic treatment is pursued in this study, i.e. let $\psi = \chi$. Attitude errors for Magsat are approximately 20 arcsec (Langel *et al.* 1981) and so let $\psi = 20$ arcsec; the scalar error is again chosen to be $\sigma_B = 4$ nT. This leads to $\sigma_\perp^2 = \sigma_3^2 = \sigma_B^2 + B^2\psi^2$ such that the principal variances are of the form $(\sigma_B^2, \sigma_\perp^2, \sigma_\perp^2)$. The principal direction of the first component is still **B**, but the last two components exist in a linear subspace perpendicular to **B**. For IRLS, an arbitrary assignment of a coordinate system fixed in the plane perpendicular to **B** will lead to an arbitrary weighting of components in that plane because this weighting is now also a function of azimuth, which has no physical meaning with respect to attitude error. To eliminate this dependency on azimuth, let the residual vector $\boldsymbol{\delta}\mathbf{B}$ be decomposed into components δB_B and δB_3 along and in the plane perpendicular to **B**, respectively. Choose $\hat{\mathbf{n}}$ to be the unit vector in the direction corresponding to δB_3 so that $\delta\mathbf{B} = (\delta B_B, 0, \delta B_3)$ in the $(B, B_\perp, B_3)$ system. Note that the coordinate system is not fixed in the plane perpendicular to **B**, but changes with δB_3. From eq. (11), the $B_\perp$ component will always be weighted as Gaussian, but the B_3 may not, depending upon the value of $|\delta B_3|$. Thus, IRLS may impart some anisotropy in the plane perpendicular to **B**, but this anisotropy will be a function of the magnitude of δB_3 only.

4.3 Regularization and *apriori* information

In addition to magnetic field observations, information has been introduced either to restrict the set of admissible parameter estimates as a result of insufficient data (regularization) or, for external fields, to impart some physical knowledge to the problem that is otherwise not supplied by data or theory. As stated earlier, this information is in the form of ℓ_2-type norms on **x** and, with the exception of F-region currents, is the same as in CM3. Main field SV is smoothed by two norms: (i) the mean-square magnitude of $\ddot{B}_\mathrm{r}$ over the core–mantle boundary (CMB) over the span of the model, denoted $\mathcal{Q}_{|\ddot{B}_\mathrm{r}|}$, and (ii) the mean-square magnitude of the surface Laplacian of $\dot{B}_\mathrm{r}$ over the CMB and over the span of the model, denoted $\mathcal{Q}_{|\nabla_\mathrm{h}^2\dot{B}_\mathrm{r}|}$. Night-side ionospheric E-region currents are minimized by a norm, denoted $\mathcal{Q}_{\|\mathbf{J}_\mathrm{eq}\|}$, which measures the mean-square magnitude of the E-region equivalent currents, $\mathbf{J}_\mathrm{eq}$, flowing at 110 km altitude over the night-time sector through the year. In addition, these currents are further smoothed by a norm, denoted $\mathcal{Q}_{\|\nabla_\mathrm{s}^2\mathbf{J}_{\mathrm{eq},p>0}\|}$, which measures the mean-square magnitude of the surface Laplacian of the diurnally varying portion of $\mathbf{J}_\mathrm{eq}$ at mid-latitudes at all local times. In the magnetosphere, the mean-squared magnitude of deviations from a dipole field in MLT is damped at Magsat altitude (450 km), independent of D_st.

For CM3, the mean-square magnitude of the radial F-region currents were minimized at Magsat altitude at dawn and dusk. For CM4, the mean-square magnitude of the surface Laplacian of J_r (denoted $\mathcal{Q}_{|\nabla_\mathrm{h}^2 J_\mathrm{r}|}$) and $\mathbf{J}_\mathrm{h}$ (denoted $\mathcal{Q}_{\|\nabla_\mathrm{h}^2\mathbf{J}_\mathrm{h}\|}$) were damped separately. The damping was on spheres at 430 and 415 km for Magsat dawn and dusk, respectively, and at 750 km for Ørsted at all local times. In addition, because Ørsted samples continuous diurnal variations, the weighted mean-square magnitude of J_r was damped over the same night-time sector used in $\mathcal{Q}_{\|\mathbf{J}_\mathrm{eq}\|}$ in order to stablize meridional coupling currents associated with the EEJ and to allow interhemispheric coupling currents to still flow via $\mathbf{J}_\mathrm{h}$. The function $\sin^8\theta_\mathrm{d}$, where θ_d is the dipole colatitude, was used to weight the norm more heavily at low and mid-dipole latitudes. It is denoted as $\mathcal{Q}_{|J_\mathrm{r}|}$. The values used for the damping parameters associated with these norms are listed in Table 2.

Table 2. Damping parameter values.

Norm	Damping parameter (λ)	
$\mathcal{Q}_{\|\ddot{B}_\mathrm{r}\|}$	6.2×10^0	$(\mathrm{nT\ yr^{-2}})^{-2}$
$\mathcal{Q}_{\|\nabla_\mathrm{s}^2\dot{B}_\mathrm{r}\|}$	6.2×10^{-8}	$(\mathrm{nT\ yr^{-1}km^{-2}})^{-2}$
$\mathcal{Q}_{\|\mathbf{J}_\mathrm{eq}\|}$	8.4×10^2	$(\mathrm{A\ km^{-1}})^{-2}$
$\mathcal{Q}_{\|\nabla_\mathrm{s}^2\mathbf{J}_{\mathrm{eq},p>0}\|}$	3.8×10^{-2}	$(\mathrm{A\ km^{-3}})^{-2}$
$\mathcal{Q}_{\|\Delta\mathbf{B}_\mathrm{ltd}\|}$	6.3×10^4	$(\mathrm{nT})^{-2}$
Magsat dawn		
$\mathcal{Q}_{\|\nabla_\mathrm{h}^2 J_\mathrm{r}\|}$	1.3×10^{-15}	$(\mathrm{nA\ m^{-4}})^{-2}$
$\mathcal{Q}_{\|\nabla_\mathrm{h}^2\mathbf{J}_\mathrm{h}\|}$	1.3×10^{-2}	$(\mathrm{nA\ m^{-4}})^{-2}$
Magsat dusk		
$\mathcal{Q}_{\|\nabla_\mathrm{h}^2 J_\mathrm{r}\|}$	1.3×10^{-15}	$(\mathrm{nA\ m^{-4}})^{-2}$
$\mathcal{Q}_{\|\nabla_\mathrm{h}^2\mathbf{J}_\mathrm{h}\|}$	1.3×10^{-2}	$(\mathrm{nA\ m^{-4}})^{-2}$
Ørsted		
$\mathcal{Q}_{\|\nabla_\mathrm{h}^2 J_\mathrm{r}\|}$	1.3×10^{-16}	$(\mathrm{nA\ m^{-4}})^{-2}$
$\mathcal{Q}_{\|\nabla_\mathrm{h}^2\mathbf{J}_\mathrm{h}\|}$	1.3×10^{-2}	$(\mathrm{nA\ m^{-4}})^{-2}$
$\mathcal{Q}_{\|J_\mathrm{r}\|}$	3.4×10^{-2}	$(\mathrm{nA\ m^{-2}})^{-2}$

5 RESULTS AND DISCUSSION

5.1 Residuals

The weighted mean and rms of the residual components of the various data sources are given in Table 3 with respect to CM4. If e_k is the kth residual for a particular component, then these quantities are defined as

$$\mathrm{mean} = \Sigma_k w_k e_k / \Sigma_k w_k, \tag{14}$$

$$(\mathrm{rms})^2 = \Sigma_k w_k e_k^2 \big/ \Sigma_k w_k, \tag{15}$$

where $w_k = w_{k,n}$, evaluated at the final model. Also listed in the table are the measurement counts, N, and the percentage of residuals residing in the Laplacian tails of the Huber distribution, denoted L%. Residuals are provided according to two classifications: (i) poleward (polar) and equatorward (non-polar) of $\pm 50^\circ$ dipole latitude and (ii) local time from 06:00 to 18:00 (day) or from 18:00 to 06:00 (night). Magsat dawn and dusk are listed under night and day, respectively. The (X, Y, Z) components are oriented (north, east, down) and the $(B_B, B_\perp, B_3)$ components are described in Section 4.2 for the particular satellite. Because of correlations resulting from attitude uncertainties, the (X, Y, Z) components of the satellite vector residuals are not independent; their statistics are unweighted and are provided for completeness.

Given that robust estimation was not employed in CM3, the OHM fits are now superior, and the POGO and Magsat fits are commensurate with those of CM3. The rms values are $\leq$16.23 nT for all OHM components listed, including polar and day-side. As a side note, the small magnitude of the means for the OHM_MUL polar and OHM_1AM polar and non-polar groups indicates that the model has nearly converged, because one would expect vanishing

Table 3. Weighted residual statistics, where N is the number of data points, L% is the percentage of residuals residing in the Laplacian tails of the Huber distribution, and the mean and rms are in units of nT. Polar denotes data poleward of ± 50° dipole latitude. Day denotes a local time from 06:00 to 18:00 h, otherwise night. Statistics for the (X, Y, Z) components of the satellite vector residuals are unweighted and are provided for completeness.

Component	Polar				Non-polar							
					Day				Night			
	N	L%	mean	rms	N	L%	mean	rms	N	L%	mean	rms
OHM_1AM												
X	22 744	9	0.01	13.20	—	—	—	—	19 826	10	−0.02	8.12
Y	22 766	6	−0.00	11.29	—	—	—	—	19 796	5	0.00	6.10
Z	22 771	10	0.01	16.23	—	—	—	—	20 300	8	−0.00	7.29
OHM_MUL												
X	66 476	10	0.01	13.55	28 517	21	−0.38	10.12	28 555	6	0.34	7.02
Y	66 544	8	−0.01	12.64	28 539	17	0.47	9.34	28 575	3	−0.45	5.63
Z	66 390	9	0.02	15.53	29 357	12	0.13	8.27	29 396	4	−0.14	5.89
POGO												
F	11 857	15	−0.04	3.97	8 789	20	−0.34	4.40	8 793	11	−0.01	3.51
CHAMP												
F	105 224	34	−0.17	5.36	70 439	24	−0.21	4.80	70 251	7	−0.01	3.29
Magsat dawn												
$F + B_B$	20 560	17	−0.28	3.76	—	—	—	—	23 368	5	0.28	2.93
B_3	11 834	69	(10.44)	15.14	—	—	—	—	19 653	35	(6.12)	7.06
X	11 834	—	1.01	24.65	—	—	—	—	19 653	—	0.72	4.79
Y	11 834	—	0.81	30.05	—	—	—	—	19 653	—	−0.86	6.12
Z	11 834	—	−0.90	6.95	—	—	—	—	19 653	—	−2.16	4.01
Magsat dusk												
$F + B_B$	19 127	20	0.40	4.20	19 404	12	0.06	3.75	—	—	—	—
B_3	10 871	67	(10.12)	15.02	16 257	42	(6.76)	7.76	—	—	—	—
X	10 871	—	−2.81	27.68	16 257	—	−0.24	5.43	—	—	—	—
Y	10 871	—	2.09	30.87	16 257	—	0.11	6.81	—	—	—	—
Z	10 871	—	0.01	7.29	16 257	—	0.36	4.13	—	—	—	—
Ørsted												
$F + B_B$	280 937	21	0.05	4.20	187 974	14	0.02	3.89	192 513	5	0.02	2.80
$B_\perp$	108 595	43	−0.06	12.58	73 547	33	−0.18	8.50	77 765	15	−0.00	6.06
B_3	108 595	51	−0.28	9.14	73 547	42	−0.03	6.98	77 765	6	−0.11	3.16
X	108 595	—	0.48	21.50	73 547	—	0.20	7.39	77 765	—	−0.18	5.59
Y	108 595	—	0.06	23.70	73 547	—	0.06	10.53	77 765	—	0.20	5.22
Z	108 595	—	0.20	7.25	73 547	—	−0.39	6.74	77 765	—	0.21	4.36

means from a converged solution as a result of the estimation of biases. However, the differences between CM4 and a truly converged model are considered negligible. For Magsat, the vector rms value is slightly larger, particularly in the Y component, but $F + B_B$ appears to be approximately the same or smaller with respect to CM3. This is consistent with attitude error considerations, which place a higher weight on the measurement magnitude at the expense of directional information; all satellites were in near-polar orbits in a strong dipolar field. In addition, Magsat vector data is poorly fit at high latitudes in components perpendicular to the ambient field, i.e. the B_3, X and Y components; clearly unmodelled signal from polar field-aligned currents (FACs). At first glance, one may be surprised to see such large positive mean values for the Magsat B_3 residual components (shown in parentheses in Table 3). Recall, however, that this component is always taken in the direction of the residual in the plane perpendicular to **B**, rendering δB_3 non-negative. The δB_3 means are equal to their mean absolute values and therefore appear anomalously large. However, the expected value of $|x|$ for a normally distributed x with mean μ and standard deviation σ is shown in Appendix A to be

$$\mathrm{E}\,[|x|] = \mu \cdot \mathrm{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \sqrt{\frac{2}{\pi}}\sigma e^{-\mu^2/(2\sigma^2)}, \tag{16}$$

where E[·] is the expectation operator and erf(·) is the error function defined as (Beyer 1981)

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt. \tag{17}$$

In the simple case that $\mu = 0$, then rms $= \sigma$ and eq. (16) reduces to $\mathrm{E}[|x|] = \sqrt{2/\pi} \cdot (\mathrm{rms}) \approx 0.8 \cdot (\mathrm{rms})$. Assuming an underlying Gaussian distribution, one can now see that the δB_3 mean values are not unreasonable. Again, any assignment of a preferred direction that might mitigate the large means would be totally arbitrary.

A comparison of Ørsted and CHAMP fits cannot be made with CM3. However, Olsen (2002) has derived a model based upon night-side Ørsted scalar data at all latitudes and vector data equatorward of ±50° dipole latitude. The resulting weighted rms values are 2.89, 6.40 and 3.25 nT for the $F + B_B$, $B_\perp$ and B_3 components, respectively, which are quite close to the non-polar, night-side values for CM4. In fact, Olsen (2002) may have been more stringent in the

tolerance of outliers than in this study. As with Magsat, the vector data are fit poorly in horizontal components at high latitudes. Although CHAMP data was not fit by Olsen (2002), weighted residual rms values with respect to that model were quoted to be 3.4 nT for non-polar and 5.4 nT for polar F. These agree quite well with the CM4 non-polar, night-side and polar rms values.

In order to get a better picture of the nature of these residuals, samplings (every tenth point) were plotted in Fig. 3 in the (B_B, $B_\perp$, B_3) system for Ørsted, along with CHAMP F, versus both dipole latitude and UT rendered in modified Julian days (MJD). Clearly, there is much dispersion at high latitudes in all components of both satellites. One can also see a general thickening of the $\delta B_\perp$ residual band over the others at low latitudes, which is expected. δB_3 shows the existence of some possible systematic outliers, which should be removed. As for behavior in UT, no major differences are seen before or after 2001 January 1 (denoted by the vertical dotted line), the dividing point before (after) which the final (provisional) D_{st} index was used, although $\delta B_\perp$ and δB_3 may be slightly more dispersed after this date.

With $c = 1.5$, the Laplacian tails of the Huber distribution account for approximately 17 per cent of the population. If the residuals are N samples from such a distribution, one expects that approximately 17 per cent of them would reside in these tails as $N \to \infty$. These percentages are shown under the column labelled L% in Table 3 for various data subsets. In general, these percentages are higher for the polar and non-polar day-side samples: this is attributed to unmodelled fields, especially rapidly varying FAC fields. This implies that the data weights are probably too heavy for polar and non-polar day-side data. The OHM residuals are more confined to the Gaussian core of the distribution while satellite vector components normal to the direction of **B** cluster in the Laplacian tails. The B_3 component of Ørsted has more samples in the tails than the $B_\perp$

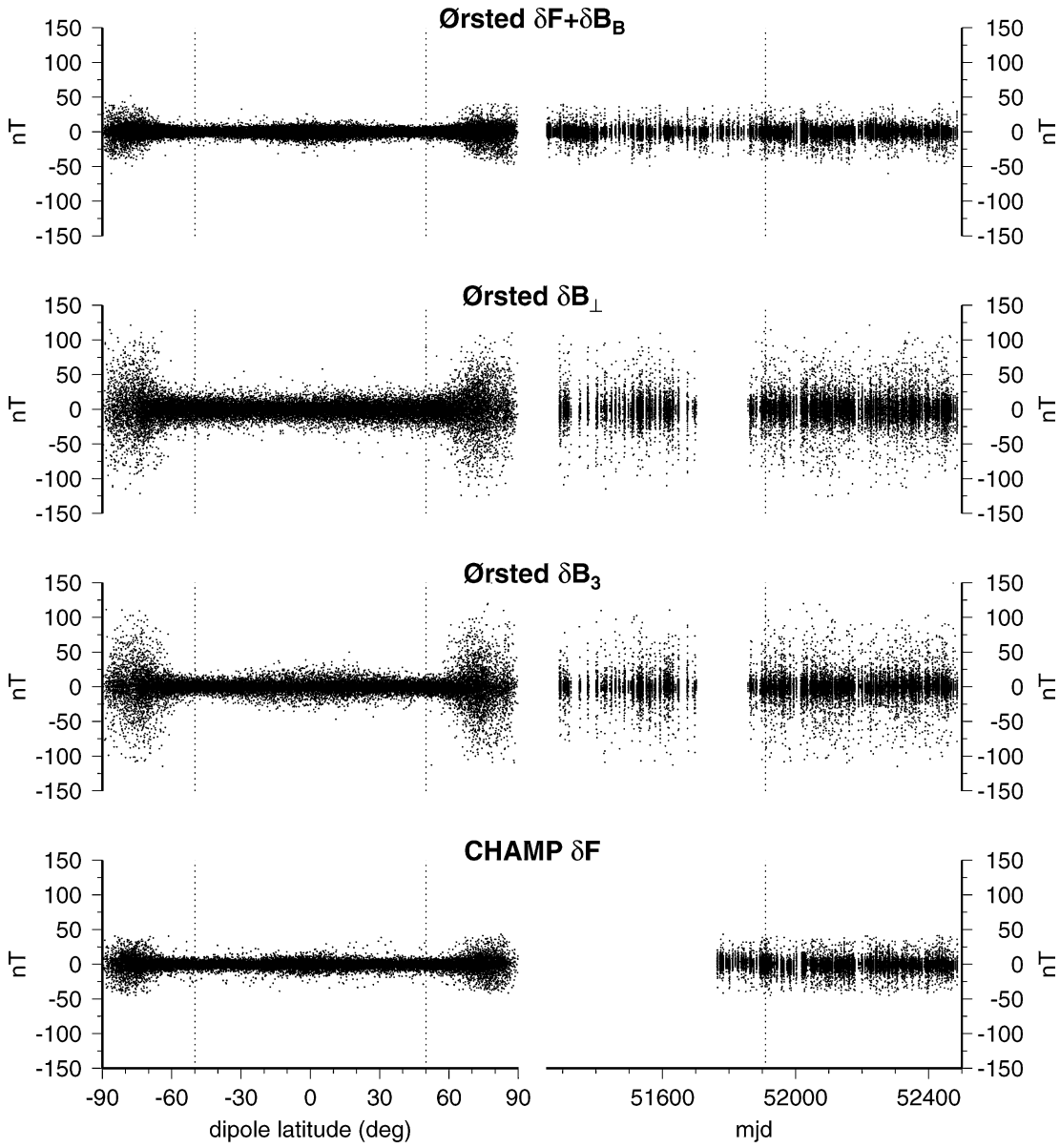

Figure 3. A sampling of the residual distributions for Ørsted and CHAMP as a function of dipole latitude (left) and universal time (right) rendered as modified Julian days (MJD). Every 10th point is plotted here. Vertical dotted lines in the left panel denote $\pm 50^\circ$ dipole latitude and in the right denotes 2001 January 1, which divides the earlier (later) days in which the final (provisional) D_{st} index was used.

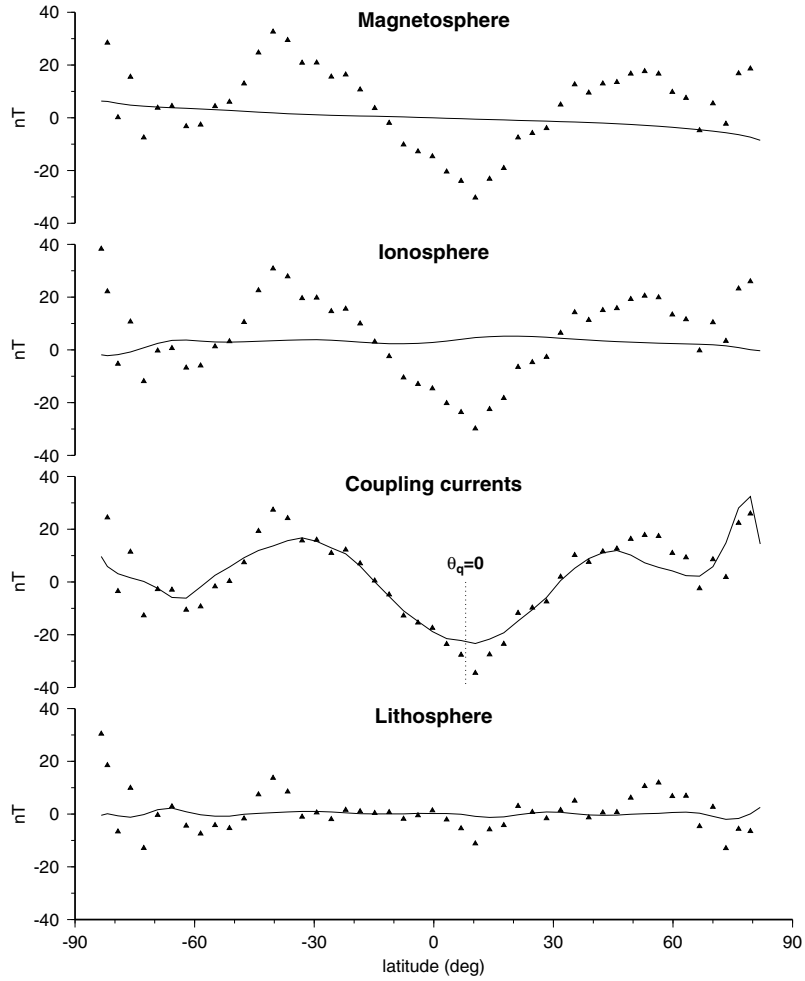

Figure 4. Residual progression for the Y component of an Ørsted descending pass on 2001 August 20 beginning at 07:00 h UT and crossing the equator at 80°E and 12:30 h MLT versus geographic latitude. For a given panel, the symbols represent residuals with respect to the main field (up to degree 13) plus all fields labelled in the panels above; the line is the prediction from the field labelled in the current panel. The vertical dotted line indicates the geographic latitude where the pass crosses the dip equator (approximately 8°N).

component in polar and day-side regions, even though the latter is more susceptible to rotation error about the SIM boresight. This may be the result of systematic errors in these regions not accounted for in the formal error estimates. If one assumes that residual samples are drawn from the same population across sources, e.g. POGO, Magsat, etc., or across the polar and non-polar day- and night-side partitions, then the same 17 per cent outlier property would still hold. This was done for the F and B_B satellite components resulting in L% values of 15, 23, 11, 16 and 14 for POGO, CHAMP, Magsat dawn, Magsat dusk and Ørsted, respectively. Thus, treating the entire residual samplings of a satellite as coming from a common population is much more in agreement with the $c = 1.5$ assumption. The L% values of 24, 17 and 6 were found for polar and non-polar day-side and night-side, respectively; the first two groups appear to be consistent with the statistical assumptions of this paper.

One strength of the CM approach is the ability to properly divide the signal among the sources, so it is interesting to look at a residual progression that highlights the new capability of modelling Ørsted F-region fields. Fig. 4 shows a progression generated for the Y component of an Ørsted descending pass on 2001 August 20 beginning at 07:00 h UT and crossing the equator at 12:30 h MLT versus geographic latitude. For a given panel, the symbols represent residuals with respect to the main field (up to degree 13) plus all fields labelled in the panels above; the line is the prediction from the field labelled in the current panel. Clearly, the toroidal Y prediction from the F-region coupling currents accounts for most of the variance in δY after removal of the main field. However, a typical Magsat δY signature would appear asymmetric with respect to the dip equator along a polar pass, with a crest (trough) to the south (north) of the dip equator. This is because the meridional currents (Maeda *et al.* 1982) upwell at the dip equator and then downwell within 10° to 20° on either side of it. This produces eastward (westward) magnetic fields south (north) of the dip equator. Yet at Ørsted altitude the toroidal Y is closer to being symmetric with respect to the dip equator (shown as a vertical dotted line in Fig. 4), indicating little vertical current flow directly at the dip equator. Furthermore, this trough is much broader than expected from the meridonal currents. This appears to be the signature of the interhemispheric connecting currents, which evidently flow above the meridional currents at Ørsted altitude. These currents and their implications will be discussed more in a later section.

© 2004 RAS, *GJI*

5.2 Prediction

In the previous section it was shown that CM4 fits very well the data used in the analysis, at least in regions where the major fields have been parametrized. While this is necessary for a good model, it does not test its predictive capabilities, which may be of considerable interest. In regards to true forecasting, any significant extrapolation of future behavior by the full model will be precluded by a failure to predict the core field SV accurately as a result of the absence of core dynamics in the model; the dynamo process is chaotic. However, because of the regularity of the ionospheric field and the tracking of the ring current and ionospheric amplification by D_{st} and $F_{10.7}$, respectively, there is a possibility to predict daily variations and variations with magnetic activity in external and induced fields during data gaps within the time span of the model. To test this, five observatories were chosen over a range of magnetic latitudes during 1990 April: GDH, KAK, TRD, CTA and SBA. This particular month and year are of interest because 1990 lies midway between the Magsat and Ørsted/CHAMP missions when only OHM_1AM data were used and April contains a major magnetic storm that commenced on the 10th day. Comparisons during this month will test the predictive capacity of the model over both daily variation and variation with magnetic activity.

Fig. 5 shows the measured X, Y and Z hourly-mean values (black symbols) and those predicted by CM4 (light blue lines) during this month at the five observatories. Biases were adjusted to remove the residual means. This is justified because OHM_MUL data were not analysed during 1990 and because magnetic activity is much higher on average during this month than reflected by the quietest day; neither OHM_MUL nor OHM_1AM biases apply. Both the northern and southern mid-latitude stations, KAK and CTA, respectively, are fit very well during this active month, even across the storm. The rms fits for (X, Y, Z) are (14.1, 10.3, 8.4) nT and (17.7, 12.6, 5.5) nT for KAK and CTA, respectively. Note how well the variations in X are tracked by D_{st}. This indicates that at these latitudes the storm response is linear in D_{st} to a good approximation to perhaps −200 or −300 nT. One can also see how well the modulations in the daily $F_{10.7}$ value (used for synthesis here instead of the three-monthly means) match the daily variability in Y. The equatorial station TRD is also fit well in X, considering its magnitude,

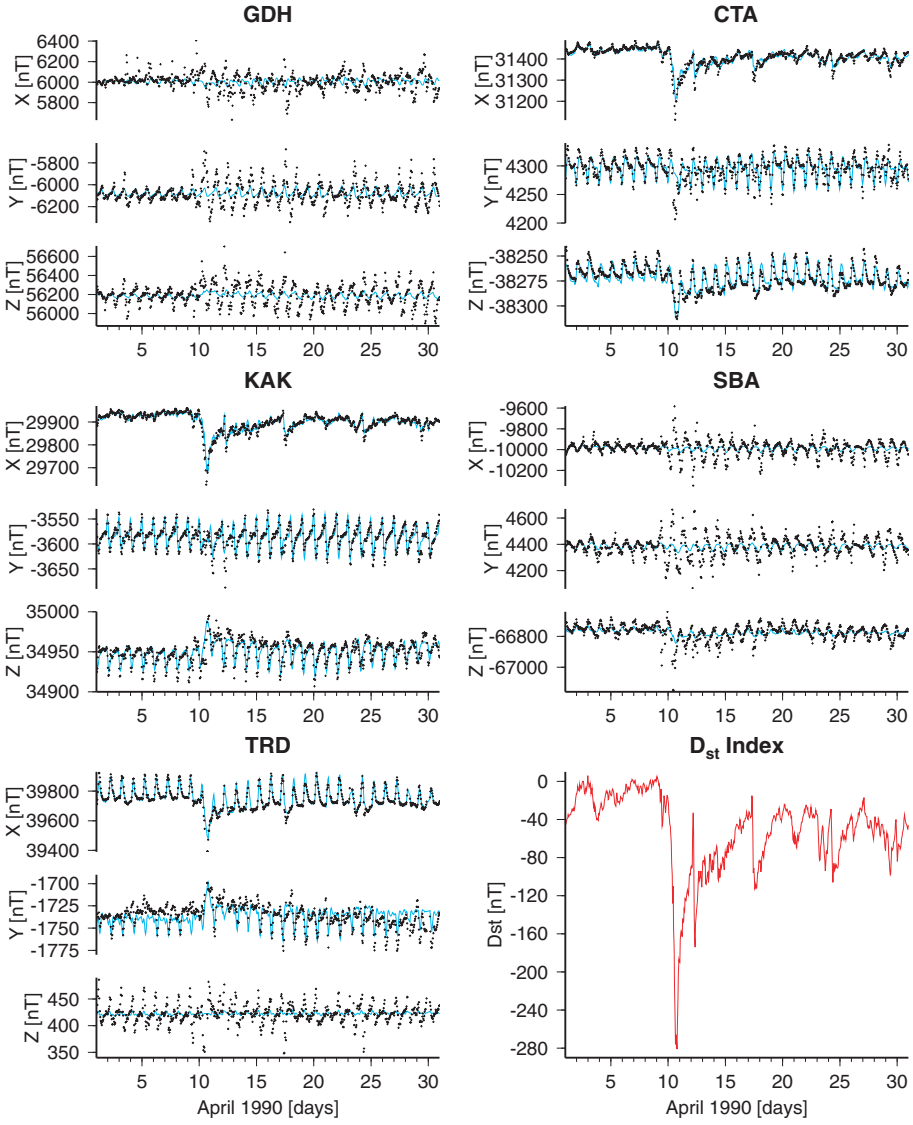

Figure 5. Comparison of measured X, Y and Z hourly-mean values (black symbols) and those predicted by CM4 (light blue lines) during 1990 April at five observatories. Biases were adjusted to remove the residual means. The D_{st} index is shown as a red line in the lower right panel during the same period. Note the commencement of the magnetic storm on the 10th day.

but somewhat less in Y and Z (rms fits are (20.5, 10.6, 17.0) nT). Most of the unmodelled variation in Z is probably the result of coastal induction effects (the station is at the southern tip of India), which are not modelled in CM4. Finally, the phase of the signal at the northern and southern polar stations, GDH and SBA, respectively, is described reasonably well throughout the month. However, the daily modulation in amplitude is fit much better before the storm commencement; they are not tracked well by either $D_{\rm st}$ or daily $F_{10.7}$. The rms fits are (85.0, 83.6, 96.1) nT and (63.0, 64.5, 42.4) nT for GDH and SBA, respectively. Evidently, the storm energy is dissipating rather quickly at mid- and low latitudes, but is more persistent near the poles. It appears that CM4 does indeed predict well the daily variations of its parametrized fields under quiet to moderate magnetic activity conditions within the time span of the model. The predictions understandably fail for fields that are not considered in the model, e.g. coastal induction, polar electrojets etc.

5.3 Resolution

It is important to understand how information in the data is used to construct the model parameter estimates. This can reveal both limitations in the model and possible avenues of design improvement. First, let

$$\mathcal{L}_{{\rm r}_i} = (\mathbf{d}_i - \mathbf{a}_i(\tilde{\mathbf{x}}))^T \, {\rm W}_i \, (\mathbf{d}_i - \mathbf{a}_i(\tilde{\mathbf{x}})) \,, \tag{18}$$

$$\mathcal{L}_{{\rm e}_i} = \lambda_i \tilde{\mathbf{x}}^T \Lambda_i \tilde{\mathbf{x}}, \tag{19}$$

$$\mathrm{C} = \left(\mathrm{A}^T \mathrm{W} \mathrm{A} + \sum_{i=1}^{K} \lambda_i \Lambda_i \right)^{-1}, \tag{20}$$

$$\mathrm{R}_{{\rm d}_i} = \mathrm{C} \mathrm{A}_i^T \mathrm{W}_i \mathrm{A}_i \,, \tag{21}$$

$$\mathrm{R}_{{\rm a}_i} = \lambda_i \mathrm{C} \Lambda_i \,, \tag{22}$$

where $\tilde{\mathbf{x}}$ is the final estimate of the model parameter vector, $\mathrm{W} = \mathrm{W}_n$ and $\mathrm{A} = \mathrm{A}_n$, evaluated at $\tilde{\mathbf{x}}$, and $\mathbf{d}_i$, $\mathbf{a}_i$, A_i and W_i are the ith subvector or submatrix of $\mathbf{d}$, $\mathbf{a}$, A and W, respectively. Thus, $\mathcal{L}_{{\rm r}_i}$, $\mathcal{L}_{{\rm e}_i}$, C, $\mathrm{R}_{{\rm d}_i}$ and $\mathrm{R}_{{\rm a}_i}$ are the weighted residual and prior error variances, the error-covariance matrix, the ith data subset resolution matrix and the ith norm resolution matrix, respectively, all evaluated at $\tilde{\mathbf{x}}$. Table 4 lists $\mathcal{L}_{{\rm r}_i}$, $N_i = \dim \mathbf{d}_i$ and tr[$\mathrm{R}_{{\rm d}_i}$] for the data subsets and $\mathcal{L}_{{\rm e}_i}$, $M_i = \mathrm{rank}[\Lambda_i]$ and tr[$\mathrm{R}_{{\rm a}_i}$] for the norms, where tr[$\cdot$] is the trace operator.

According to information theory, the trace of the resolution matrix gives the expected number of parameters resolved by that particular data subset or norm (Tarantola 1987). For CM4, the data are on average resolving approximately 63 per cent of the model parameters, with roughly equal amounts for the OHM, Magsat and CHAMP data sets. Ørsted is resolving almost twice as many parameters as CHAMP for example, but is also a much larger data set. POGO resolves many fewer parameters, but is also the smallest data set. In addition, the $\mathcal{Q}_{|\ddot{B}_{\rm r}|}$ and $\mathcal{Q}_{\|\nabla_{\rm h}^2 \mathbf{J}_{\rm h}\|}$ norms are resolving high percentages of their associated parameters. If the residuals were from a Gaussian distribution, then $\mathcal{L}_{{\rm r}_i}$ would be from a χ^2 distribution with an expected value of $N_{\rm i} - \mathrm{tr}[\mathrm{R}_{{\rm d}_i}]$. Similarly, if the model parameters were from a Gaussian distribution, then $\mathcal{L}_{{\rm e}_i}$ would be from a χ^2 distribution with an expected value of $M_i - \mathrm{tr}[\mathrm{R}_{{\rm a}_i}]$. The ratio of observed to expected values gives an indication as to the relative influence of that data subset or norm (Sabaka *et al.* 2002). Because not all residuals have been treated as realizations of Gaussian distributions (this is the reason for employing robust estimation) and because an underlying Gaussian distribution for the model parameters is dubious, these ratios have not been computed. However, there is one ratio of particular interest defined as

$$s^2 = \frac{(\mathbf{d} - \mathbf{a}(\tilde{\mathbf{x}}))^T \, \mathrm{W} \, (\mathbf{d} - \mathbf{a}(\tilde{\mathbf{x}}))}{N - \mathrm{tr}\,[\mathrm{C}\mathrm{A}^T \mathrm{W} \mathrm{A}]}, \tag{23}$$

where $N = \dim \mathbf{d}$. Because s^2 is the ratio of the observed weighted residual variance from all the data to its expected value (Bloxham *et al.* 1989), it measures how well the model fits the weighted data per degree of freedom (DOF). It should be approximately unity if the weighting is correct. One may calibrate C by simply multiplying by s^2. Such a calibrated C reflects how well the observations can actually be fit. For the CM4 model, $s^2 = 0.74$, indicating that the initial overall uncertainties are a bit too large.

While the trace of the resolution matrix can provide the number of parameters expected to be resolved by a particular data subset, it cannot discriminate what portion of a parameter state is resolved by that subset. To address this, consider a linear model whose parameters can be successfully estimated from each of several groups of data, i.e. the parameters are observable by each group. This situation might describe a set of static main field coefficients at 2001 estimated independently by both Ørsted and CHAMP

Table 4. CM4 resolution and calibration information. $\mathcal{L}_{{\rm r}_i}$ and $\mathcal{L}_{{\rm e}_i}$ are the weighted residual and prior error variances, respectively; $\mathrm{R}_{{\rm d}_i}$ and $\mathrm{R}_{{\rm a}_i}$ are the data and norm resolution matrices, respectively; $N_i = \dim \mathbf{d}_i$ and $M_i = \mathrm{rank}\,[\Lambda_i]$; and tr[$\cdot$] is the trace operator.

Data subset	$\mathcal{L}_{{\rm r}_i}$	N_i	tr$\left[\mathrm{R}_{{\rm d}_i}\right]$
OHM	336 799.73	500 552	3429.08
POGO	18 368.00	29 439	216.64
Magsat	139 632.45	199 689	3348.84
CHAMP	181 765.50	245 914	3 253.65
Ørsted	915 319.60	1 181 238	5649.97
Subtotal	1 591 885.28	2 156 832	15 898.18
Norm	$\mathcal{L}_{{\rm e}_i}$	M_i	tr[$\mathrm{R}_{{\rm a}_i}$]
$\mathcal{Q}_{\|\ddot{B}_{\rm r}\|}$	24 495.92	4485	3453.08
$\mathcal{Q}_{\|\nabla_{\rm s}^2 \dot{B}_{\rm r}\|}$	1353.76	4485	9.31
$\mathcal{Q}_{\|\mathbf{J}_{\rm eq}\|}$	20 103.88	5520	979.10
$\mathcal{Q}_{\|\nabla_{\rm s}^2 \mathbf{J}_{{\rm eq},p>0}\|}$	1058.97	4910	1103.39
$\mathcal{Q}_{\|\Delta \mathbf{B}_{\rm ltd}\|}$	19 779.89	740	256.32
Magsat dawn			
$\mathcal{Q}_{\|\nabla_{\rm h}^2 J_{\rm r}\|}$	346.41	1044	634.35
$\mathcal{Q}_{\|\nabla_{\rm h}^2 \mathbf{J}_{\rm h}\|}$	149.32	120	109.02
Magsat dusk			
$\mathcal{Q}_{\|\nabla_{\rm h}^2 J_{\rm r}\|}$	521.22	1044	648.90
$\mathcal{Q}_{\|\nabla_{\rm h}^2 \mathbf{J}_{\rm h}\|}$	187.21	120	109.96
Ørsted			
$\mathcal{Q}_{\|\nabla_{\rm h}^2 J_{\rm r}\|}$	2402.67	5520	1156.71
$\mathcal{Q}_{\|\nabla_{\rm h}^2 \mathbf{J}_{\rm h}\|}$	197.10	600	440.50
$\mathcal{Q}_{\|J_{\rm r}\|}$	557.84	6120	444.18
Subtotal	71 154.19	34 708	9344.82
Grandtotal	1 663 039.47	2 191 540	25 243.00

vector data. If one now considers a joint solution from both data sets, then intuitively it seems that this solution should be some type of weighted average of the two independent solutions. Each term would then represent that portion of the joint parameter state resolved by the corresponding data subset. Furthermore, if one is interested only in the resolution of a particular subset of parameters, the target parameters, then these terms can be collectively smoothed such that their sum has minimal structure, but still resolves the target parameters. This smoothing essentially eliminates extraneous effects that average to zero over the data subsets and are the result of parameters outside of the target subset. It is shown in Appendix B that these weights are normalized and correspond to the resolution matrices of the data subsets and smoothing norms. It is also shown that under fairly general conditions the length of the resolved contribution to the joint parameter state is less than that of the independent state for a given data subset. This means that the resolution matrices generally downweight the various independent estimates that form the joint estimate.

A complete mathematical development of the general resolution analysis case, i.e. one in which non-linear F data are considered and in which the data subsets are assumed insufficient for determining independent parameter estimates, is given in Appendix B. In this case, the resolved contribution from the ith data subset may be interpreted as the solution resulting from an analysis in which all but the ith subset of residuals with respect to non-target parameters are set to zero, i.e. how much the ith residual subset requires the model to deviate from zero when all other data are forcing it to zero. Thus, the target parameters are cast as a sum of perturbations as a result of the influence of the target fields present in the data subsets.

The resolution analysis may now be applied to fields where data interactions are crucial to proper modelling, such as the lithosphere. Therefore, let the target subset be the vector of lithospheric field parameters for $n = 16$–65. Fig. 6 shows the B_r component of this field at 400 km altitude from CM4 as predicted by the contributions resolved from CHAMP, Magsat, Ørsted, POGO and all the data. At this altitude the rms magnitudes of the fields resolved by CHAMP, Magsat, Ørsted and POGO are 2.1, 1.4, 0.9 and 0.2 nT, respectively, as compared with 4.1 nT from all the data. The level of contribution appears to be controlled by two dominant effects: the weight of the data subset (heavier weights and/or more data lead to larger relative magnitudes) and satellite altitude (higher altitudes lead to broader, less intense features). Evidently, the lithospheric signal in POGO is used very little in resolving the lithospheric parameters: this contribution is greatly downweighted compared with independent estimates derived exclusively from POGO (see Langel 1990).

Although there is much commonality between the CHAMP, Ørsted and Magsat maps, there are conspicuous, along-track streaks present near the dip equator in the Magsat map at low–mid-latitudes. Further analysis indicates that these features exist at both dawn and dusk in the B_r and B_ϕ components, but not F, at approximately the 1-nT-amplitude level, and that they appear over track lines rather than regions absent of data. The B_r-B_ϕ morphology seems to be correlated in local time as opposed to geographic position, that is, there is a tendency towards a change in sign in B_ϕ between the dawn and dusk portions of an orbit (B_ϕ is either sunward or antisunward over the orbit). The associated B_r structure suggests weak meridional currents in one direction at dawn and the opposite direction at dusk, with both polarities present. Depending upon the seasonal structure of the polarities, the stripes could be manifestations of the mid-latitude FACs envisioned by Fukushima (1994). Being that these currents are field-aligned, they would not be expected to have a signature in F. However, these current lines would have to close below the Magsat sampling shell and well below that of Ørsted (because Ørsted does not detect these features), otherwise the resulting magnetic fields would be non-potential and would be difficult to map into the lithospheric field. Alternatively, a misalignment of the vector magnetometer would also result in features that were transparent to F, but would be manifested in the components perpendicular to the direction of the ambient field at the dip equator, mainly B_r and B_ϕ. The sign change in B_ϕ between dawn and dusk portions of the orbit would be expected for a fixed alignment bias in the yaw angle as a result of a single rotation of Magsat per orbit. This attitude bias would be of the order of 7 arcsec, given the magnitude of the features in a predominately B_θ-directed 30 000-nT field, which is well within the 20 arcsec nominal accuracy of the Magsat vector data. One might expect to see similar signs in the B_ϕ features for the ascending or descending passes if these were truly a result of misalignments. However, these maps are the result of a least-squares fit to all of the measurements and cannot be expected to treat alignment biases, which in some instances appear to be lithospheric signal, in a systematic way.

One goal of the new generation of magnetic mapping missions, such as Ørsted and CHAMP, is a continuous sampling of the field through time, which will offer a detailed look into the behavior of the core field SV. It is therefore interesting to see how the satellite data are being used to resolve the SV of particular harmonics of this field. Let the target subset be the vector of coefficients describing $\dot{g}_1^0$. The upper panel of Fig. 7 shows g_1^0 from CM4 over the time span of the model (solid line), along with g_1^0 values predicted by integrating the $\dot{g}_1^0$ resolved by the OHMs (dashed line), POGO plus Magsat (dotted line) and Ørsted plus CHAMP (dot-dashed line) from the 1980 value of g_1^0 (long-dashed line). Satellite mission envelopes are indicated by the shaded regions. Ørsted and CHAMP contributions account for nearly all of the signal during their missions while POGO clearly dominates during its mission. The OHM contribution is strongest during the gap between the Magsat and Ørsted missions when no other data are available. This behavior makes good sense, but the influence of the satellites outside of their respective mission envelopes may seem counter intuitive. However, this can be explained by the fact that the B-splines are correlated in time, that is, the support of adjacent splines are shifted by one knot interval, which leads to overlap. Hence, an event at one end of the domain can influence the other end via propagation through the sequence of splines. This influence is enhanced between the mission envelopes and the core SV epoch (1980) by integration. Indeed, the Ørsted and CHAMP contributions ramp up after 1980 while POGO and Magsat contributions slowly dissipate after 1980.

Because the g_1^0 term is somewhat special, the analysis was carried out on a less dominant harmonic to see if the resolution patterns change signficantly. The lower panel of Fig. 7 shows a similar plot for h_4^4. In this case, the OHMs appear to have resolved a large percentage of the signal at all times except during the Ørsted and CHAMP missions. The same gradational behavior in POGO plus Magsat and Ørsted plus CHAMP contributions is seen here, although the former is greatly reduced. If this term is indicative of the general resolution levels amongst the low degree harmonics, then it confirms the utility of observatory measurements in determining the SV, even in the presence of satellite data.

5.4 Geomagnetic jerks

The time span of CM4 has been extended from 1960 to mid-2002 and, as a result, traverses several instances of a phenomenon that

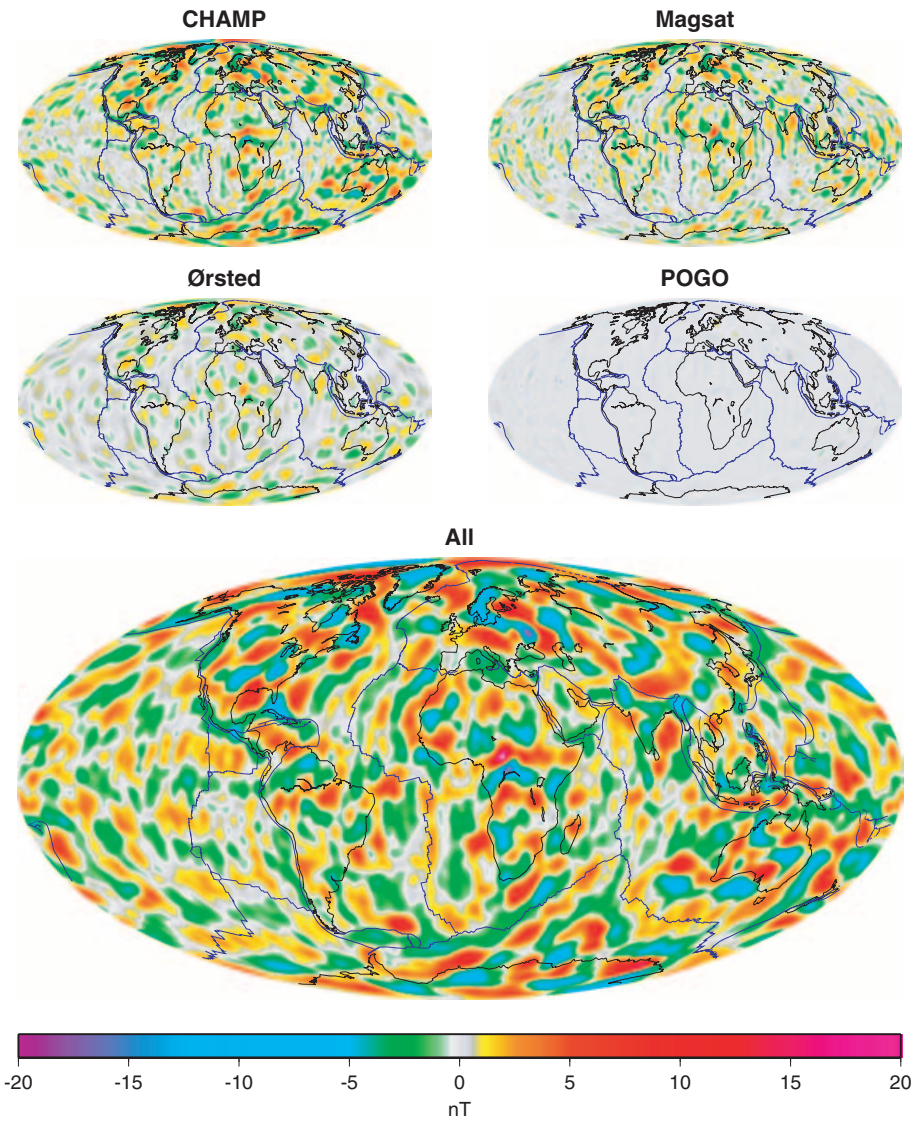

Figure 6. The portion of the B_r component of the lithospheric field ($n = 16 - 65$) at 400 km altitude from CM4 resolved by CHAMP, Magsat, Ørsted, POGO and all the data (Mollweide projections). Major tectonic boundaries are drawn in blue.

has come to be known as a geomagnetic jerk (see Courtillot & Le Mouël 1976; Malin & Hodder 1982; Courtillot & Le Mouël 1984). These are generally accepted as worldwide events of internal origin, which have been theorized by some to be jumps in acceleration of the fluid motion at the CMB (Le Huy *et al.* 2000) or to be created by torsional oscillations of the core (Bloxham *et al.* 2002). In simple terms, they are manifested in the magnetic record as the sudden change in slope of a V-shaped segment of SV, i.e. they are an impulse or a Delta function in the third time derivative of the field. In reality, the duration of a jerk may be subannual to several years. In addition, certain jerks may not be detected over the entire globe and may even be influenced by external signals (Alldredge 1975, 1984). These last points are intriguing: CM4 provides not only a global description of the evolution of the field through time, but also attempts to separate the internal from the external signals, even the induced from the core, during quiet-time conditions. The model could be used as a tool to test, or at least elucidate, these kinds of claims.

Alexandrescu *et al.* (1995, 1996) have put forth a sophisticated wavelet analysis for detecting and characterizing geomagnetic jerks from horizontal field components measured from a worldwide distribution of observatories. The data used are actually monthly means defined as the average over all days of the month and all times of the day. Citing edge-effect problems in the wavelet analysis, Mandea *et al.* (2000) investigated smoothed first differences of these monthly means, particularly $\dot{Y}$, to infer a possible jerk near 1999. Neither study attempted to separate external from internal nor core from induced effects. It is therefore interesting to see what CM4 says about the contributions of these individual sources to the observatory data being analysed. Consider then a time-series of monthly-mean first differences at the stations NGK and HER from 1960 to 2002 that have been smoothed in accordance with Mandea *et al.* (2000) (smoothing centred first differences with a 12-month running average). This process eliminates most of the annual variation and leaves a time-series in which candidate jerks are fairly recognizable. Fig. 8 shows the $\dot{X}$, $\dot{Y}$ and $\dot{Z}$ components smoothed from actual monthly means and those produced from a progression of sources from CM4. For each component, there is a comparison of observed (black symbols) versus core (red line), core plus induced

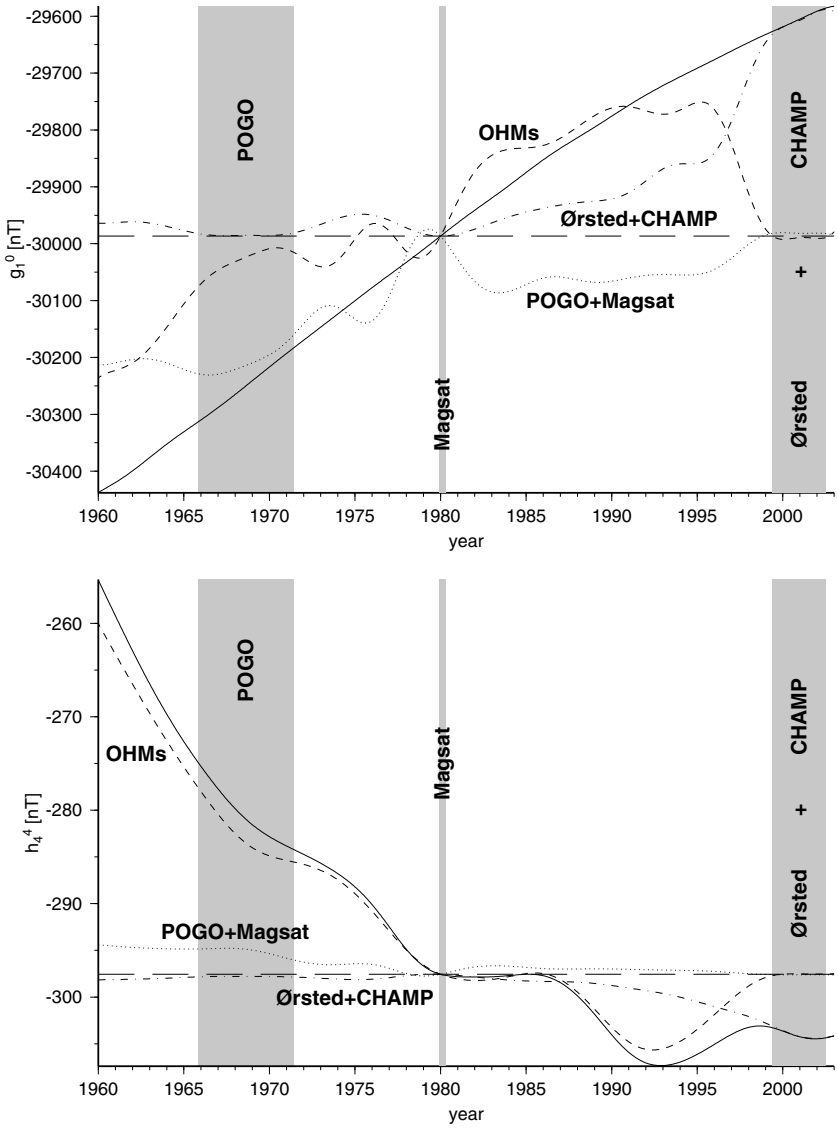

Figure 7. The top panel shows the portion of g_1^0 from CM4 predicted by integrating the $\dot{g}_1^0$ resolved by the OHMs (dashed line), POGO plus Magsat (dotted line), Ørsted plus CHAMP (dot-dashed line) and all the data (solid line) from the 1980 value of g_1^0 (long-dashed line). The bottom panel shows a similar figure for h_4^4.

(green line), and core plus induced plus external (light blue line) predictions from CM4. One can immediately see a striking agreement in the horizontal components, particularly $\dot{X}$, when all sources are considered. Some deviation exists at HER from 1994–1998 in $\dot{Y}$ and $\dot{Z}$, but recall that CM4 is derived during quiet times only, whereas the actual first differences include disturbed times (comparison with smoothed monthly means from 01:00 h local time data shows a closer agreement with CM4). There are unmatched excursions in $\dot{Z}$, but the model attempts to track most of these, which suggests that they also emanate from more disturbed conditions. One can also see that the induced and external contributions are in phase for the horizontal components, but out of phase for $\dot{Z}$. This makes sense when considering that the Sq current system consists of two large current vortices of opposing flow in the Northern and Southern Hemispheres.

With induced and external contamination removed, the core contribution from CM4 indeed shows the well-known 1969 and 1978 internal events in the $\dot{Y}$ components at NGK and HER and even the 1–2 yr lag of the Southern behind the Northern Hemisphere noted by Alexandrescu *et al.* (1996). The 1991 event (Macmillan 1996; De Michelis *et al.* 1998) can also be clearly seen along with an apparently similar lag. The event proposed by Mandea *et al.* (2000) near 2000 is present, but it appears that another candidate event might predate this near 1997 (more will be said about this later). What is interesting here is that many of these events are also manifested in the core contributions of other components, e.g. the 2000 event in NGK $\dot{X}$ and $\dot{Z}$. This broadens the amount of information that can be used in determining these events. Perhaps even more importantly, the observed monthly means will be contaminated by induced and external fields, particularly during active periods, which may influence the determination of the event locations. For instance, local extrema exist in the data record of $\dot{Y}$ at NGK near the times of each of the mentioned events. Separation of the fields in CM4 reveals that these extrema can be shifted by perhaps as much as 1–2 yr from the underlying extrema in the core contribution by external field effects. Similar problems are likely from induction effects even if internal

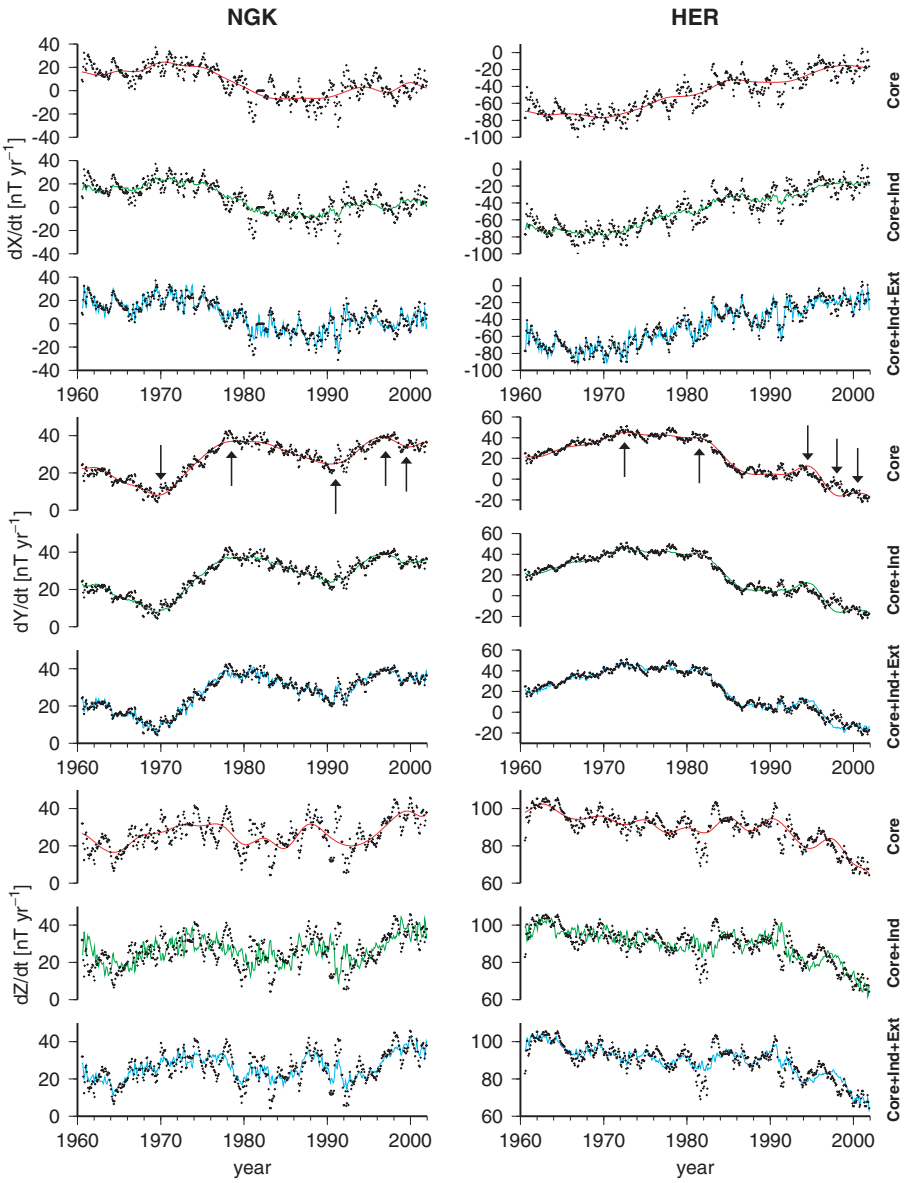

Figure 8. Comparison of smoothed monthly-mean first differences, $\dot{X}$, $\dot{Y}$ and $\dot{Z}$, from observations (black symbols) and from predictions using a progression of sources from CM4 (colored lines) at NGK and HER from 1960 to 2002. Red, green and blue lines reflect core, core plus induced, and core plus induced plus external sources, respectively, from CM4. Arrows indicate approximate jerk locations from visual inspection of the core contribution to $\dot{Y}$.

and external signals are properly separated. Although it is beyond the scope of this paper to quantify the effects of non-core signals in characterizing jerks via wavelet analyses, it is clear that they need to be properly considered.

As previously mentioned, inspection of $\dot{Y}$ at NGK, as well as other European observatories, indicates that two events might have occurred between 1996 and 2000 as suggested by two changes in the sign of the slope. Alternating signs in slope have been seen in jerk sequences and have prompted an interest in characterizing successive jerks. Le Huy *et al.* (1998) have studied the 1969, 1978 and 1991 jerk sequence and found anticorrelation in global models of secular acceleration, which is accentuated in the implied core flow models, between successive jerk events. Their methods are applied here to see how well CM4 reproduces this behavior and to test whether the two events between 1996 and 2000 conform to the successive jerk scenario and whether they are local or global in extent. In their study, first differences of smoothed annual means for each component at each of 160 observatories were best fit by piecewise linear functions of four sections having common but adjustable vertices. The final vertice locations were 1969, 1979 and 1992 and represent the jerk locations. The differences in slope between two successive segments for each component at each observatory, $\delta\ddot{X}$, $\delta\ddot{Y}$ and $\delta\ddot{Z}$, were then fit with internal and external SH expansions to degree 4. In this study, first differences of smoothed annual means of core field Gauss coefficients from CM4 were fit by piecewise linear functions of six sections having vertices 1969, 1979, 1992, 1997 and 2000, to match the previous studies and to test the proposed new position at 1997. Note that the annual means were computed from 20 samples per year, centred on the half year from 1960.5 to 2002.5. A simple difference in Gauss coefficients between successive segments, $\delta\ddot{g}$, then provides a point of comparison to the previous study. Fig. 9 shows the north ($\delta\ddot{X}$), east ($\delta\ddot{Y}$) and down ($\delta\ddot{Z}$) components of the

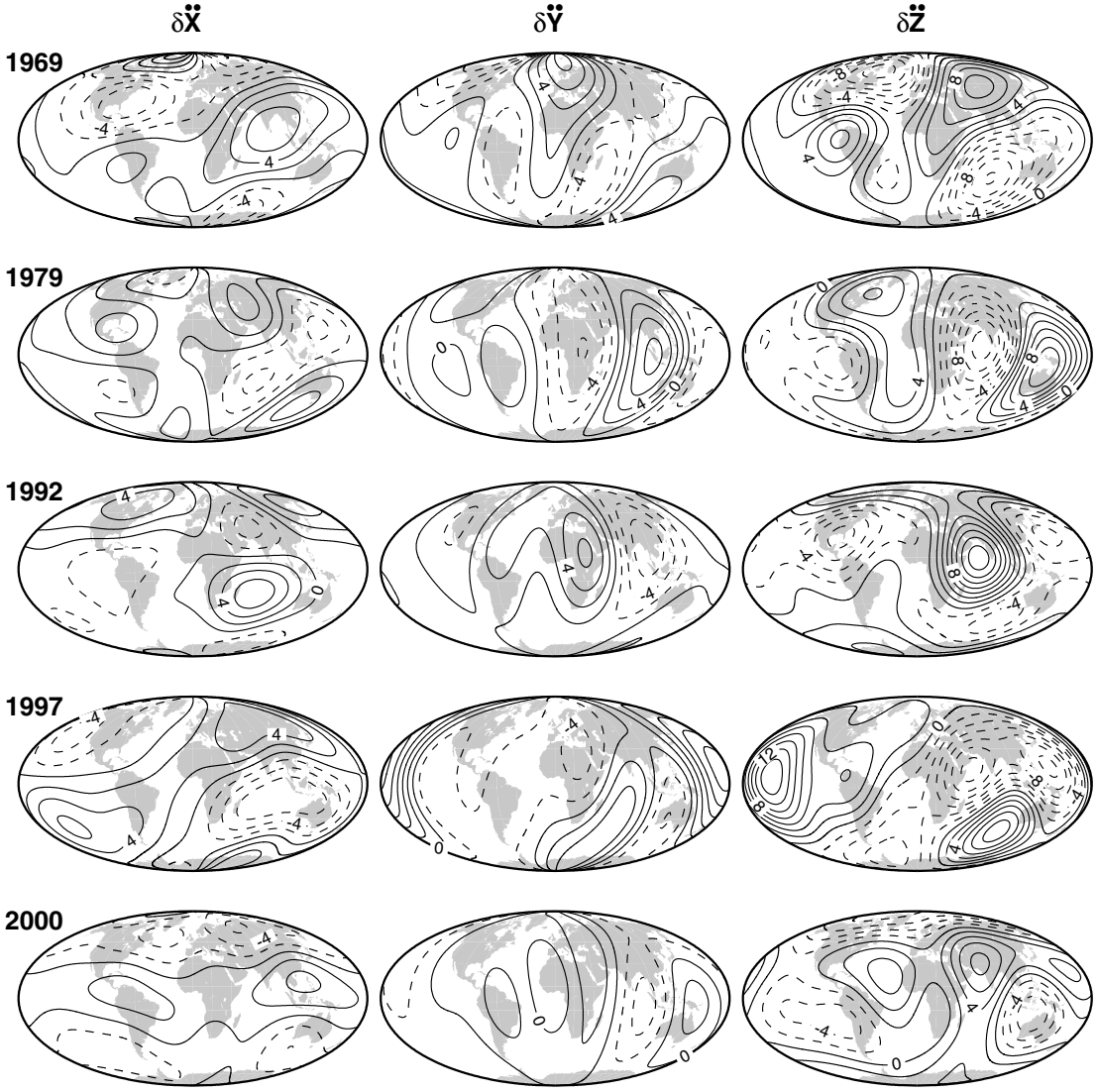

Figure 9. North ($\delta\ddot{X}$), east ($\delta\ddot{Y}$) and down ($\delta\ddot{Z}$) components of the 1969, 1979, 1992 and 2000 jerks and the 1997 point of interest. The contour interval is 2 nT/yr^2.

1969, 1979, 1992 and 2000 jerks and the 1997 point of interest computed from the sets of $\delta\ddot{g}$. The 1969 and 1979 patterns and amplitudes look very similar to those of Le Huy *et al.* (1998) and the 1992 agreement is not bad. Going to 1997 and 2000, the patterns appear to be generally anticorrelated between each event, especially in the Eastern Hemisphere. This could be an indication that the 1997 event does not have a global extent. To quantify these successive patterns, Le Huy *et al.* (1998) employ the global correlation coefficient of McLeod (1985),

$$c_{12} = \frac{\int_\Omega \mathbf{B}_1 \cdot \mathbf{B}_2 \, ds}{\sqrt{\int_\Omega |\mathbf{B}_1|^2 \, ds \int_\Omega |\mathbf{B}_2|^2 \, ds}}, \tag{24}$$

where Ω is the surface of the Earth. These coefficients have been computed between all events and are listed in Table 5. One immediately sees a checkerboard pattern in the signs of c_{12} indicating anticorrelation between alternating events. The strongest anticorrelations occur between successive events, although the 1979–2000 coefficient is also relatively large in magnitude. Le Huy *et al.* (1998) quote coefficients of −0.61 between 1969 and 1979, −0.43 between 1979 and 1992, and 0.31 between 1969 and 1992. The anticorrelation between 1992 and 1997 is relatively strong at −0.47, but is slightly weaker between 1997 and 2000 at −0.36. Some caution must be taken when looking at the 2000 event because it occurs near the edge of the CM4 time span. However, given these caveats, it appears that CM4 indeed confirms the hypothesis that successive jerks do change their signs. There is also evidence that at least a local jerk may have occurred near 1997 in addition to the one at 2000, but this will require a more rigorous analysis. If the 1997 event turns out to be a jerk of global extent, then the 1992, 1997 and 2000 sequence will exhibit the closest occurrences of jerks detected so far.

Table 5. Global correlation coefficients, c_{12}, between jerk events computed from CM4.

	1979	1992	1997	2000
1969	−0.49	0.18	−0.21	0.26
1979	—	−0.57	0.04	−0.40
1992	—	—	−0.47	0.17
1997	—	—	—	−0.36

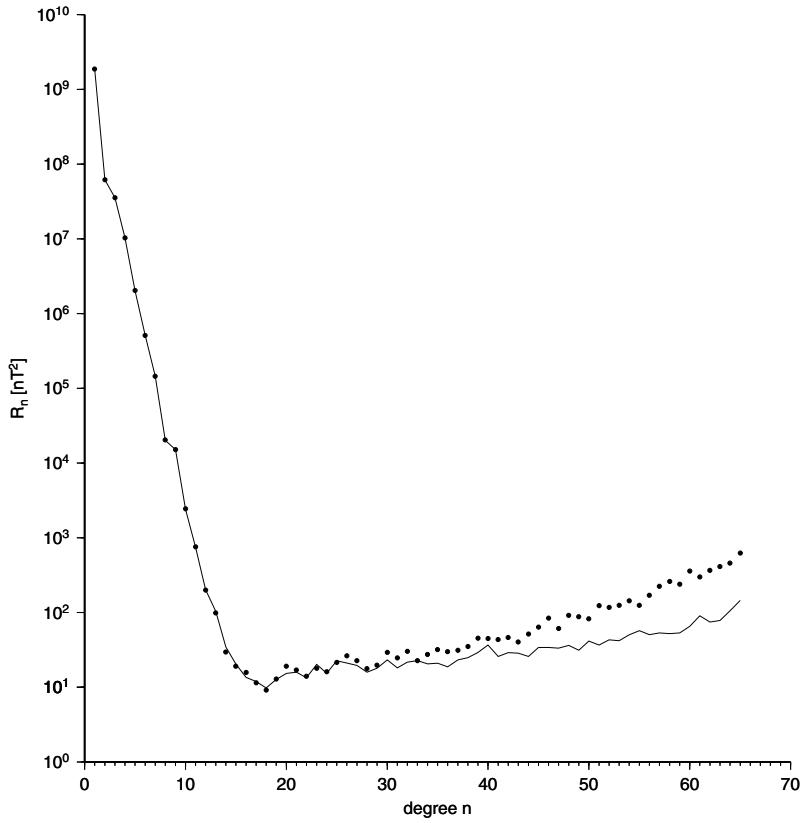

Figure 10. Comparison of the Lowes–Mauersberger (R_n) spectra for CM4 (line) and CM3 (symbols) at the surface of the Earth. R_n is the mean-square magnitude of the magnetic field over a sphere produced by harmonics of degree n.

5.5 Lithospheric fields

Besides the extension of main field SV to mid-2002, the lithospheric field, here taken to be the internal SH expansion corresponding to $n \geq 16$, is one of two constituent fields that have changed most noticeably over their CM3 counterparts. Its B_r component is shown in the bottom map of Fig. 6 at 400 km altitude, along with some of the more major tectonic boundaries. The CM4 map has fewer spurious, small-scale oscillations at low and mid-latitudes, especially along the dip equator. This is shown perhaps more convincingly by a comparison of the Lowes–Mauersberger, R_n, spectra at the surface of the Earth in Fig. 10. R_n is the mean-square magnitude of the magnetic field over a sphere produced by harmonics of degree n and at $n = 65$ the CM4 value is less than one quarter of that of CM3. The departure begins arguably around $n = 26$ after which the CM4 spectrum falls increasingly below that of CM3. This is likely the result of the addition of more high-precision data, such as CHAMP, which would tend to drive down the variance in the field, and the use of a Magsat data set with much denser coverage, which would fill data holes known to exist in the CM3 data set.

It is also fruitful to compare with the lithospheric fields of other models that are not part of the CM series. For this, the MF3 lithospheric field model of Maus *et al.* (2004) was chosen. It is derived from quiet, night-side scalar and vector CHAMP data, which have been corrected for main and low-degree external fields from a pre-existing Ørsted-CHAMP based model and for tidal effects. The remaining magnetospheric ring current and associated induced contributions are removed on a track-by-track basis, including corrections for auroral electrojets (AEJs). The lithospheric field is solved in the range $n = 16$–90 with regularization employed above $n = 60$ in order to extract clusters of lithospheric coefficients that are best resolved by the data. Global difference maps in B_r, B_θ and B_ϕ for lithospheric portion $n = 16$–65 between MF3 and CM4 at 400 km altitude are shown in Fig. 11. The largest deviations are the sectoral groupings of north–south streaks in δB_r and δB_ϕ, which appear at all latitudes, and the zonal bands along the dip equator and at polar latitudes in δB_θ, each of which can reach several nT in amplitude. The former are reminiscent of the features isolated in the Magsat lithospheric contribution discussed in Section 5.3. Indeed, some of these features are common to both maps, but there are additional features in the difference maps that may reflect differences in the along-track processing of the data. The latter exhibit a morphology consistent with an intense westward current flow along the dip equator in the mid-Atlantic between its points of highest curvature: perhaps an induction effect related to the EEJ.

While Fig. 11 indicates that differences exist between the models, it cannot attribute the features to a particular model. To address this, consider an azimuthal spectrum of the power in the field, that is, consider the mean-square magnitude of the field over a sphere produced by harmonics having the same m/n ratio, or azimuthal number, denoted as R_a (Maus, private communication, 2003). If the azimuthal numbers are defined to be non-negative for the g_n^m and negative for the h_n^m, then zonal terms occur at $m/n = 0$ and sectoral terms occur at $m/n = \pm 1$. Fig. 12 shows the R_a spectra for CM4 (solid black) and MF3 (light blue) at 400 km altitude. It is apparent that most of the $\delta\mathbf{B}$ field resides in CM4, having approximately 20 per cent more power for zonal terms and almost an order of magnitude more power at $m/n = 1$ compared with MF3. Inspection of the resolved

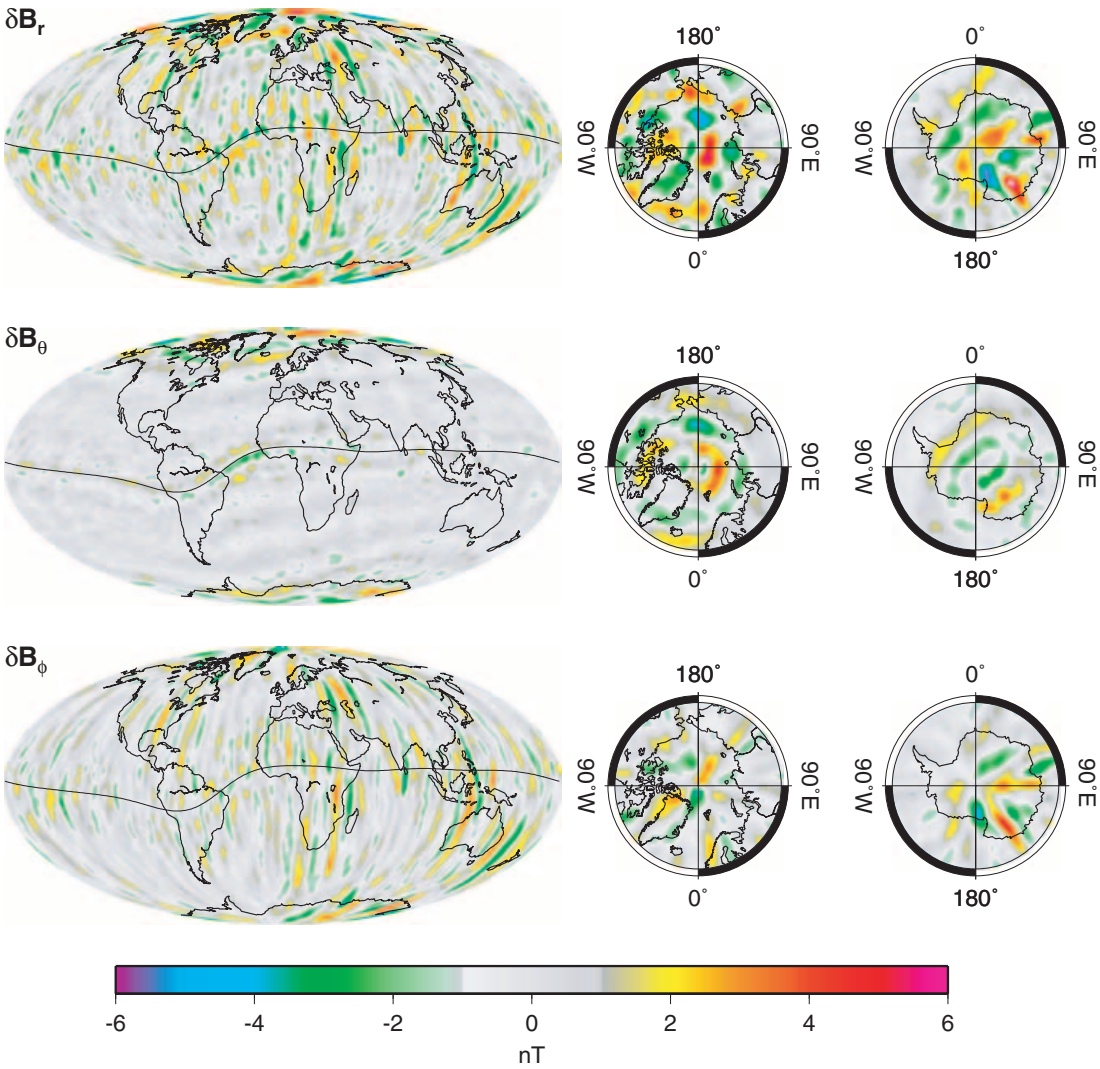

Figure 11. Global difference maps in B_r, B_θ and B_ϕ for lithospheric portion $n = 16$–65 between MF3 and CM4 at 400 km altitude (Mollweide and polar projections). The dip equator is included with the continental outlines.

contributions from each satellite reveals that CHAMP and Magsat are responsible for most of the R_a signal and so included are the spectra for the CHAMP squared terms (red), Magsat squared terms (dark blue), CHAMP and Magsat crossed terms (green, denoted as CHAMP $\otimes$ Magsat), and the sum of the CHAMP and Magsat squared and crossed terms (dashed black, denoted as CHAMP $\oplus$ Magsat). These plots clearly show that Magsat is responsible for a high percentage of the sectoral structure in the CM4 lithosphere and presumably $\delta\mathbf{B}$, which has previously been attributed to non-lithospheric processes. It should be mentioned, however, that part of the large discrepancy in R_a between CM4 and MF3, particularly in the sectoral terms, is likely the result of first correcting for ring-current effects on a track-by-track basis and then estimating the lithospheric field in MF3. Corrections along polar orbital arcs have significant influence on sectoral terms. This type of serial estimation can be shown to result in a reduction in power of the lithospheric model. The coestimation of all magnetic sources in the CM approach is designed to exactly mitigate such problems. In addition, the regularization of lithospheric terms above $n = 60$ in MF3 is actually a damping, which drives the coefficients towards zero in accordance to a prescribed power law. Conversely, CM4 uses no direct damping on its lithospheric coefficients and these coefficients generally increase in power with n. The true model state probably lies somewhere between the two models.

While the CM approach allows for a consistent analysis of along-track features, the presence of perpendicular errors or external field contamination in the Magsat vector data casts suspicion on some of them, particularly in the neighbourhood of the dip equator. However, there are also an abundance of north–south lineaments that appear to be legitimate. Focusing again on Fig. 6, many features align well with tectonic boundaries such as the Aleutian trench, the Izu-Bonin trench, the Himalayan plateau, the mid-Atlantic ridge and the Andes subduction zone; several of these strike north–south. A particularly intriguing area is the S American/S Atlantic region shown in Fig. 13. The S Atlantic spreading centre (A) and the subduction zone defining the western boundary of S America (B) are features that are either obfuscated or are totally absent in the CM3 and MF3 maps. The S Atlantic feature has been predicted from a model of induced and remanent magnetization of the lithosphere (Purucker *et al.* 2002) and shows up remarkably well in CM4 as a continuous anomaly stretching from the Romanche fracture zone to the Bouvet triple junction. A spreading zone signature is expected because of the enhanced magnetization associated with those zones (Dyment & Arkani-Hamed 1998). A thickened crust, coupled with normal to

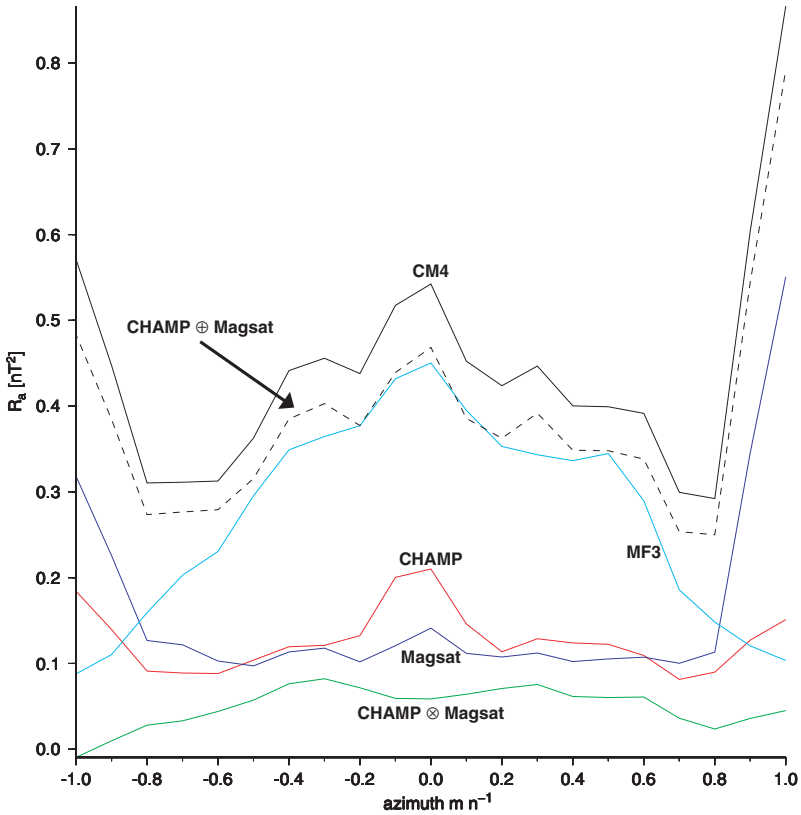

Figure 12. Comparison of the azimuthal (R_a) spectra for CM4 (solid black) and MF3 (light blue) at 400 km altitude. Also shown are those contributions resolved by CHAMP and Magsat, which include CHAMP squared terms (red), Magsat squared terms (dark blue), CHAMP and Magsat crossed terms (green, denoted as CHAMP ⊗ Magsat), and the sum of the CHAMP and Magsat squared and crossed terms (dashed black, denoted as CHAMP ⊕ Magsat). R_a is the mean-square magnitude of the magnetic field over a sphere produced by harmonics having the same m/n ratio, or azimuthal number. By definition, the g_n^m have non-negative and the h_n^m have negative azimuthal numbers. In addition, the spectra have been smoothed in order to accentuate the broader trends.

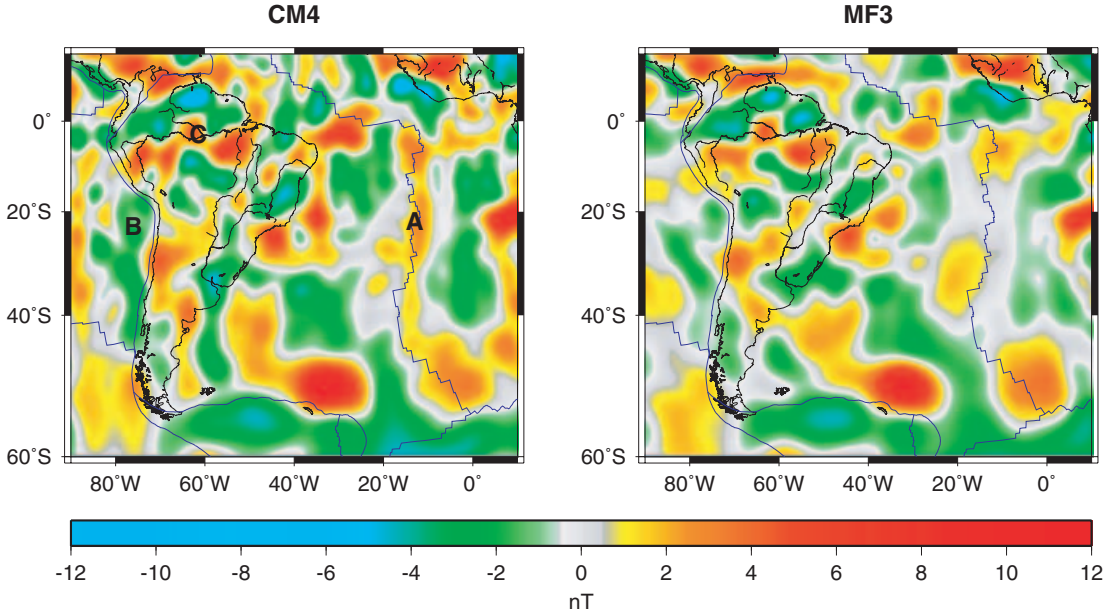

Figure 13. Exploded view of the B_r component of the lithospheric field ($n = 16$–65) at 400 km altitude over the S American and S Atlantic region from CM4 and the MF3 model of Maus *et al.* (2004; Mercator projections). Letters identify particular anomalies referenced in the discussion.

slightly elevated heat flows, would explain the magnetic feature observed along the western boundary of S America. An alternative, or additional, explanation would invoke induced magnetizations associated with the Andean subducting slab. Again, the CM approach is well suited for analysing along-track (in this case north–south) features, but Fig. 6 shows that feature B comes almost exclusively from Magsat vector data. Further analysis will thus be required before a definitive interpretation can be given. One other feature is worth noting: the Amazon river loosely defines an east–west striking magnetic boundary (C) both near its mouth and in its upper reaches, with a more complicated pattern in the middle reaches. This pattern is seen in both the CM4 and MF3 maps and an east–west boundary was predicted by the model of Purucker *et al.* (2002) on the basis of a thinner magnetic crust in this region. The more complicated magnetic pattern seen in the middle reaches of the river is consistent with a northward extension of the thinner magnetic crust as seen on the Geological Map of the World (Bouysse 2002).

A second area of interest is mainland Australia over which CM4 and MF3 differ in two primary areas with respect to the lithospheric total anomaly, T, corresponding to degrees $n = 16$–65. Recall that T is the component of the lithospheric field in the direction of the ambient field. In the southeastern portion of Fig. 14, CM4 shows a paired north–south trending high and low (A2), which is largely absent from MF3. In the west, CM4 shows a north–south trending saddle between two highs (B and D) in northwestern and northeastern Australia. In MF3, the saddle is replaced by an east–west trending low (C).

Australia is a good location to compare the longest aeromagnetic wavelengths because of several recent high-altitude aeromagnetic surveys that took the form of two concentric rings covering mainland Australia. These surveys have allowed a dewarping (Tarlowski *et al.* 1996) of the aeromagnetic map of T for Australia. However, a comparison of the satellite and aeromagnetic maps is still difficult because of the need to compare at a common altitude. On the one hand, upward continuation of the aeromagnetic map systematically distorts the anomalies of interest near the edges of mainland Australia. On the other hand, a downward continuation of CM4 is possible only to intermediate altitudes, say 200 km, but continuation to the surface reveals high-frequency artefacts. So a different approach is taken in which a single level (the 5250 nT level) is contoured in the aeromagnetic T data. This level is near the top of the data range and, because the survey is so dense (a 5 km grid was contoured here), the contour effectively shows in red the lateral extent of the highest amplitude aeromagnetic anomalies.

In southeastern Australia, the aeromagnetic compilation delineates an area devoid of magnetic highs (A2) that extends north from 144°W, 36°S to 27°S before assuming a westerly trend (A1). The northerly trending low is shown most clearly in CM4 and is absent from MF3. However, both maps show the connecting west-trending feature. These lows coincide with a known area of high heat flow (Cull 1991) and an elevated geotherm (O'Reilly & Griffin 1985). So the north–south trending feature seen in CM4 is more consistent with the aeromagnetic compilations, and heat flow and geotherm data sets, than is MF3.

In western Australia, the aeromagnetic compilation delineates an area of high magnetic fields coincident with the Archean Pilbara craton centred at 118°W, 23°S (B). Both CM4 and MF3 also see this feature as a high in the total intensity field of northwestern Australia. In southwestern Australia, the aeromagnetic compilation also delineates an area of high magnetic fields coincident with the Archean Yilgarn craton centred at 118°W, 33°S (D). Likewise,

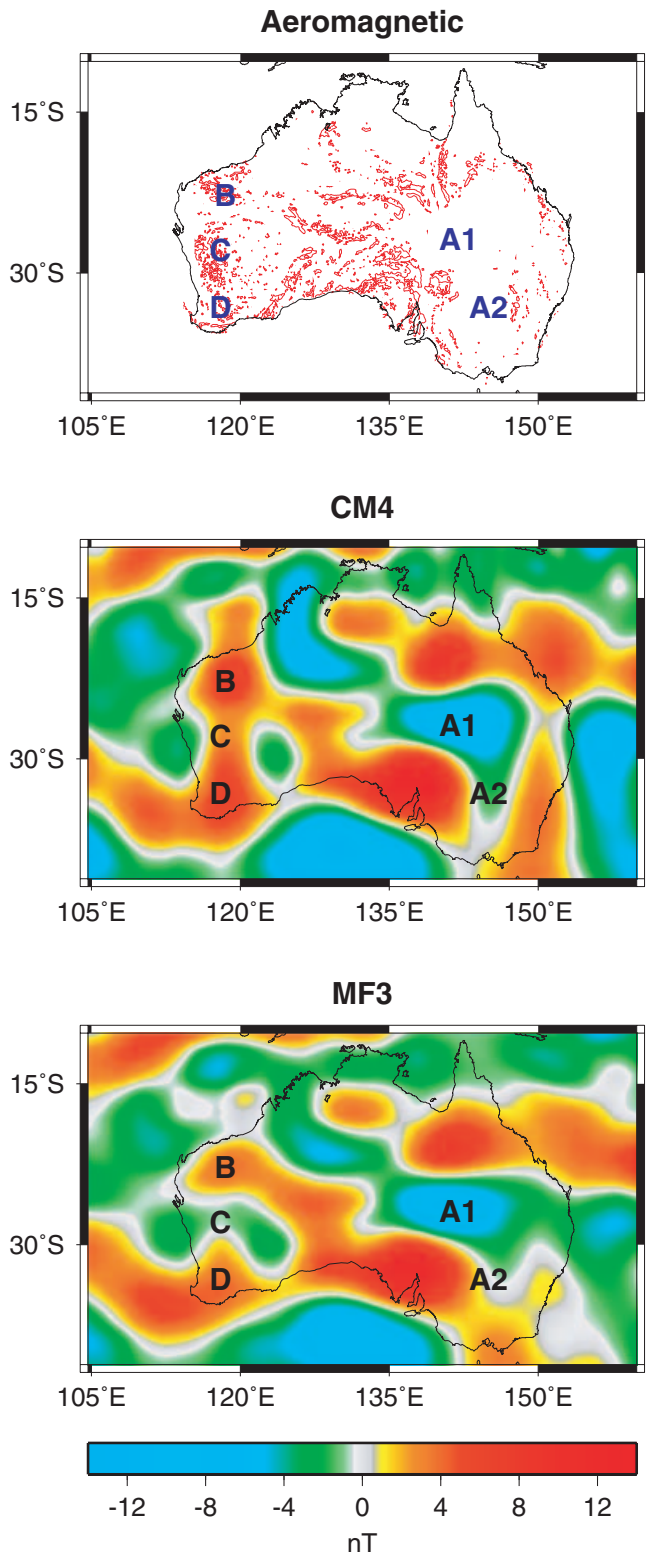

Figure 14. The 5250 nT contour level of the lithospheric total anomaly, T, of the Australian aeromagnetic map of Tarlowski *et al.* (1996; top) and exploded views of T for degrees $n = 16$–65 at 400 km altitude over the Australian region from CM4 (middle) and the MF3 model of Maus *et al.* (2004; bottom; Mercator projections). Letters identify particular anomalies referenced in the discussion and are in the same positions on all maps.

both CM4 and MF3 resolve this feature. The aeromagnetic compilation shows a relative low between the two features, coincident with an area of regional metamorphism and extensive plutonism (the Capricorn orogen). A recent high-fidelity broad-band seismic survey (Reading & Kennett 2003), extending north–south from the Pilbara to the Yilgarn cratons, images the Moho in the three regions and finds that the Moho is distinct, and at approximately 30 km under the Pilbara and 40 km under the Yilgarn, and diffuse under the Capricorn. So the east–west trending feature (C) seen in MF3 would be more consistent with the aeromagnetic and seismic compilations than is the north–south trending saddle seen in CM4. However, its location does not appear to be coincident with the location of the Capricorn orogen nor the aeromagnetic low, but is rather to the south.

5.6 *F*-region currents

Perhaps the largest difference between CM3 and CM4 is the resolution of toroidal fields generated by *in situ* F-region currents in the satellite sampling shells. Fig. 15 shows global maps of the J_{r} and $J_{\theta_{\mathrm{q}}}$ components of these fields from CM4 within these shells. The top two pairs are for fixed dawn and dusk MLTs on December 21 at 430 and 415 km altitudes, respectively. Both components appear to be in general agreement with predictions from simulations of the ionospheric dynamo (Richmond & Roble 1987): asymmetric flow across the dip equator such that a net flow exists from the winter (Northern) to summer (Southern) Hemisphere at dusk and the opposite at dawn. J_{r} shows the well-known meridional currents associated with the EEJ upwelling along the dip equator and downwelling along side lobes at dusk, but no such feature at dawn. $J_{\theta_{\mathrm{q}}}$ also shows flow away from the dip equator at low QD latitudes during dusk associated with flying above the meridional currents, which is superimposed on the background interhemispheric flow. Interestingly, J_{r} at dawn shows weak upward flow in the north and downward flow in the south. This might appear to be in contradiction with the associated $J_{\theta_{\mathrm{q}}}$ component; however, the currents need not arch upward over the satellite, but could instead be dipping beneath the satellite at dawn. Olsen (1997) also notes complicated behavior in this component at dawn in his analysis; moreover, $J_{\theta_{\mathrm{q}}}$ intensities are much greater than J_{r} at low to mid-latitudes. This is attributed to current continuity in which the horizontal component of a solenoidal **J** flowing in a thin horizontal layer must compensate the radial component, which flows through a much thicker vertical layer.

The bottom two panel pairs in Fig. 15 are for fixed noon MLT and MUT on August 20 at 750 km, a typical Ørsted altitude. As stated earlier, because the precessional period of the local time where the Ørsted orbit crosses the equator is close to an integer year number, there could be correlations between seasonal and diurnal variations in the Ørsted F-region currents, so such plots must be interpreted with caution. There is better agreement in the noon MLT plots with the predictions of Richmond & Roble (1987): net flow from the winter (Southern) to summer (Northern) Hemisphere at noon, although $J_{\theta_{\mathrm{q}}}$ indicates that some variation exists with longitude. J_{r} shows that this interhemispheric flow is symmetric across the dip equator, and also shows that the meridional currents are either weak or absent at this altitude. The plausibility of this flow structure is supported, at least in the Indian sector, by the toroidal Y component predicted along the pass that crossed the dip equator at approximately $81^{\circ}E$ in Fig. 4 (the track is shown in the fixed noon MLT plot as a dashed line). This component would be produced by just such a pattern in J_{r} and $J_{\theta_{\mathrm{q}}}$. The continuous local time dependency of J_{r} is also in good agreement with model predictions, as illustrated in the fixed noon MUT plot; Richmond & Roble (1987) found strong currents in the opposite direction in the morning sector (the eastern Pacific ocean). Here, however, this feature is taken with caution, because there is a paucity of Ørsted vector data at dawn and dusk. The weaker current density during night-time is probably the result of the $\mathcal{Q}_{|J_{\mathrm{r}}|}$ damping.

To clarify the meridional structure of the F-region currents predicted by CM4, profiles in QD latitude are shown in Fig. 16 for the J_{r} (dotted lines) and $J_{\theta_{\mathrm{q}}}$ (solid lines) components, integrated along ϕ_{q}. The resulting profiles represent the average QD meridional structure of the currents for dawn MLT at 430 km and dusk MLT at 415 km during vernal equinox and northern winter solstice from Magsat and for dawn, noon, dusk and midnight MLTs at 750 km during the solstices from Ørsted. These seasons were chosen in order to illustrate the annual extremes in structure experienced by each satellite. Again, one can see the meridional currents associated with the EEJ in the Magsat dusk MLT profiles; interestingly, J_{r} is more symmetric during vernal equinox, but $J_{\theta_{\mathrm{q}}}$ is more symmetric during northern winter, resulting in more flow from north to south during vernal equinox. Horizontal flow at dawn is clearly opposite that at dusk during vernal equinox, but less clear during northern winter. At Ørsted altitude and low QD latitudes, $J_{\theta_{\mathrm{q}}}$ indicates that bulk flow is generally from south to north during northern summer, with the strongest flow concentrations moving from north to south across the dip equator with increasing MLT. The opposite is seen during northern winter, but is less pronounced. However, what is clear in all these profiles is a weak J_{r} component suggesting broad, horizontal current movement at Ørsted altitude.

Inspection of the fixed noon MUT map of $J_{\theta_{\mathrm{q}}}$ exposes a weakness in the present parametrization: there is no variation along lines of constant θ_{q} with t_{mlt} as would be expected from the oppositely flowing interhemispheric currents detected by Magast in the morning and the evening (top two pairs). There is variation in the map for fixed noon MLT. To see this, the $\theta_{\mathrm{q}} = \pm 60^{\circ}$ lines have been plotted on the $J_{\theta_{\mathrm{q}}}$ maps along with the dip equator. To see why this is so, recall from eq. (7) that $J_{\theta_{\mathrm{q}}}$ is the component in the direction of $\nabla_{\mathrm{h}}\theta_{\mathrm{q}}$. Assuming $\|\nabla_{\mathrm{h}}\theta_{\mathrm{q}}\|$ is constant along lines of constant θ_{q}, then only $\partial f/\partial\theta_{\mathrm{q}}$ changes along these lines. For a fixed season, this change comes from multipliers of the form $\exp ip\,\phi_{\mathrm{p}}(t_{\mathrm{mut}}(t))$. Thus, $J_{\theta_{\mathrm{q}}}$ is constant along lines of constant θ_{q} when: (i) t, and therefore $t_{\mathrm{mut}}(t)$, is constant, as in the fixed noon MUT map; or (ii) $p = 0$, as in the fixed dawn and dusk MLT maps. Only fixed MLT maps produced by terms involving $p \neq 0$ will show variation along lines of constant θ_{q} because $t_{\mathrm{mut}}(\phi_{d,o}) = t_{\mathrm{mlt}} - \phi_{d,o}/15$ varies across the map. However, despite this weakness the results indicate an average current flow from the winter (Southern) to the summer (Northern) Hemisphere, as expected.

Finally, it should be pointed out that unlike potential magnetic fields, whose angular and radial dependence is completely specified, the toroidal magnetic fields associated with the F-region currents are non-potential such that the radial dependence of the $\phi^{m}_{nsp}(r)$ terms in eq. (8) is rather arbitrary. This limits the extent to which these fields can be accurately extrapolated in the radial direction by the model. Hence, the models in this section should be applied only within their appropriate sampling shells, which are approximately between 350–500 km altitude for Magsat and 675–850 km altitude for Ørsted.

6 CONCLUSIONS

The inclusion of Ørsted vector and scalar and CHAMP scalar data in CM4 has resulted in a great improvement over its predecessor

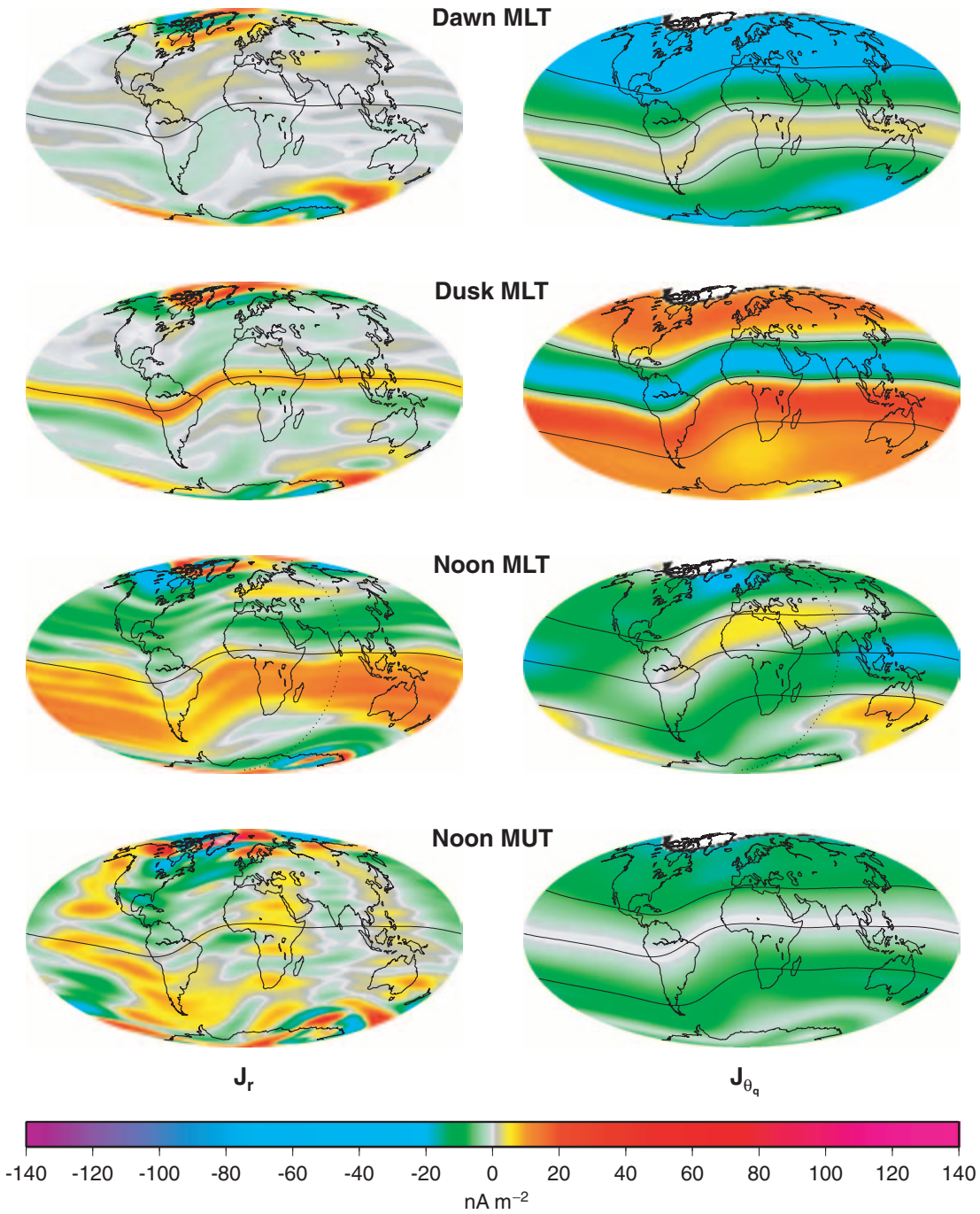

Figure 15. J_r (left) and J_{θ_q} (right) components of the F-region current density. From the top, the component pairs are for dawn MLT at 430 km and dusk MLT at 415 km on December 21 then noon MLT and MUT at 750 km on August 20 (Mollweide projections). Note that noon MLT on August 20 is very close to the time of the Ørsted pass of Fig. 4, whose track is indicated on the noon MLT plots here. The dip equator is plotted on all maps and the $\theta_q = \pm 60^\circ$ curves are plotted on the J_{θ_q} maps as well.

CM3. This model is offering new insights into complicated processes such as geomagnetic jerks by virtue of its ability to separate various field sources. Because of its coestimation approach to modelling, intriguing north–south features are being discovered in its lithospheric representation, such as the long extension of the magnetic signatures of the S Atlantic spreading ridge and possibly the Andean subduction zone. The analysis of the meridional component of the F-region currents, through which both Magsat and Ørsted fly, has yielded exciting new information showing the likely closure below Ørsted altitude of the meridional currents associated with the EEJ. In addition to the model itself, new resolution analysis techniques have been developed that have greatly aided in diagnosing problems; most notably, the possible presence of a misalignment bias in the Magsat vector magnetometer. Such information will expedite future improvements to the model. To better serve the geomagnetics community, the CM4 model and its forward code are available from the authors by request. Additional material on the CM series may be found at `http://core2.gsfc.nasa.gov/CM/`.

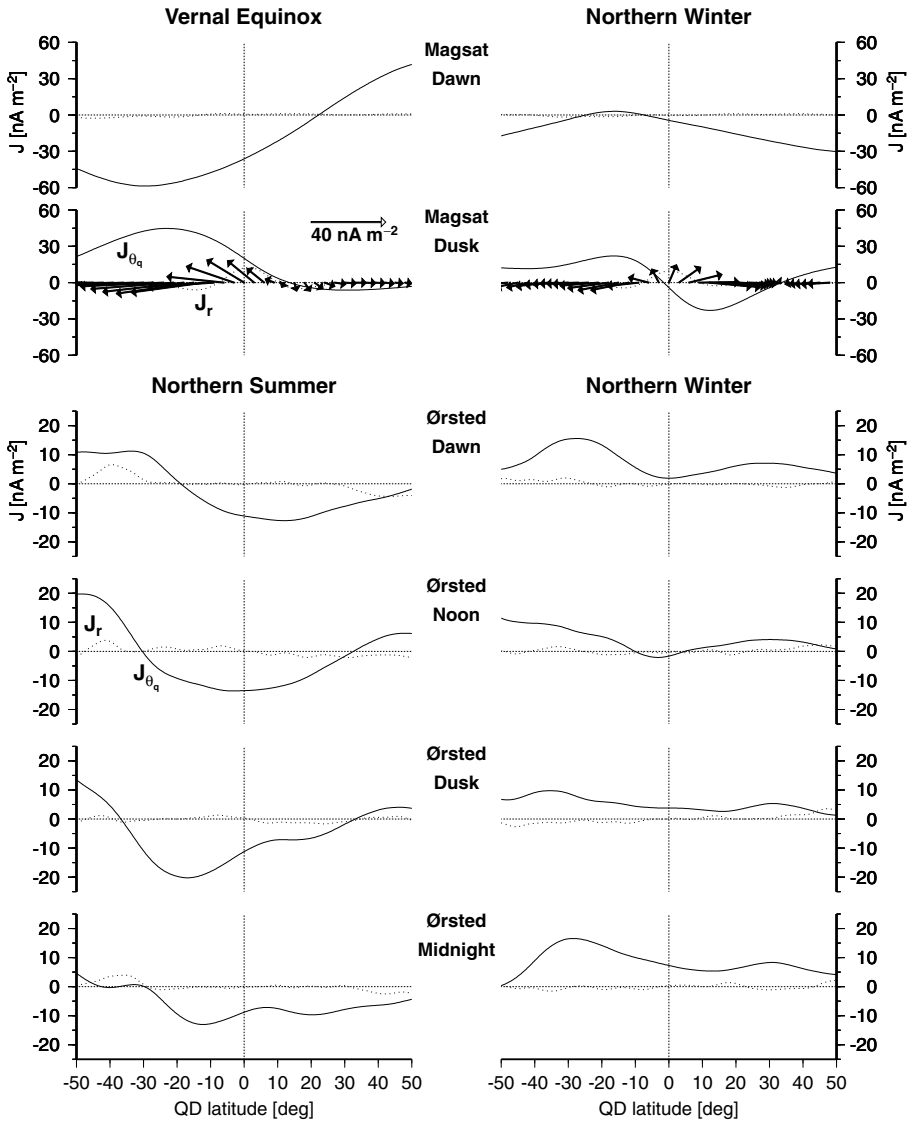

Figure 16. QD latitude profiles of the J_r (dotted lines) and J_{θ_q} (solid lines) components of the F-region current density, **J**, integrated along ϕ_q. From the top, the profiles are for dawn MLT at 430 km and dusk MLT at 415 km during vernal equinox and northern winter solstice from Magsat followed by dawn, noon, dusk and midnight MLTs at 750 km during the northern summer and winter solstices from Ørsted. The **J** vectors are shown as arrows in the Magsat dusk MLT profiles.

ACKNOWLEDGMENTS

We thank Richard Holme and Stefan Maus for fruitful reviews and Gauthier Hulot and Coerte Voorhies for useful comments. The NASA Center for Computational Sciences at Goddard Space Flight Center provided computer support. All figures were produced with GMT (Wessel & Smith 1991). TJS and MEP were supported under NASA Contract NAS5-00181.

REFERENCES

Alexandrescu, M., Gibert, D., Hulot, G., Le Mouël, J.-L. & Saracco, G., 1995. Detection of geomagnetic jerks using wavelet analysis, *J. geophys. Res.,* **100,** 12 557–12 572.

Alexandrescu, M., Gibert, D., Hulot, G., Le Mouël, J.-L. & Saracco, G., 1996. Worldwide wavelet analysis of geomagnetic jerks, *J. geophys. Res.,* **101,** 21 975–21 994.

Alldredge, L.R., 1975. A hypothesis for the source of impulses in geomagnetic secular variations, *J. geophys. Res.,* **80,** 1571–1578.

Alldredge, L.R., 1984. A discussion of impulses and jerks in the geomagnetic field, *J. geophys. Res.,* **89,** 4403–4412.

Backus, G.E., 1986. Poloidal and toroidal fields in geomagnetic field modeling, *Rev. Geophys.,* **24,** 75–109.

Backus, G.E. & Gilbert, F., 1968. The resolving power of gross Earth data, *Geophys. J. R. astr. Soc.,* **16,** 169–205.

Backus, G.E. & Gilbert, F., 1970. Uniqueness in the inversion of gross Earth data, *Phil. Trans. R. Soc. Lond.,* A., **266,** 123–192.

Beyer, W.H., 1981. *CRC standard mathematical tables, 27th edition,* CRC Press, Boca Raton, FL.

Bierman, G.J., 1977. *Factorization methods for discrete sequential estimation,* Academic Press, London.

Bloxham, J., Gubbins, D. & Jackson, A., 1989. Geomagnetic secular variation, *Phil. Trans. R. Soc. Lond.,* A., **329,** 415–502.

Bloxham, J., Zatman, S. & Dumberry, M., 2002. The origin of geomagnetic jerks, *Nature,* **420,** 65–68.

Bouysse, P., 2002. *Geological Map of the World,* Scale 1:50,000,000, ed. Bouysse, P., CGMW/CCGM & Bureau de Recherches Geologiques et Minieries, France.

Constable, C.G., 1988. Parameter estimation in non-Gaussian noise, *Geophys. J.,* **94,** 131–142.

Courtillot, V. & Le Mouël, J.-L., 1976. On the long-period variations of the Earth's magnetic field from 2 months to 20 years, *J. geophys. Res.,* **81,** 2941–2950.

Courtillot, V. & Le Mouël, J.-L., 1984. Geomagnetic secular variation impulses, *Nature,* **311,** 709–716.

Cull, J.P., 1991. Geothermal gradients in Australia, in *The Australian Lithosphere,* Geol. Soc. Australia Spec. Pub. 17, pp. 147–156, ed. Drummond, B., Geol. Soc. Australia, Sidney, Australia.

De Michelis, P., Cafarella, L. & Meloni, A., 1998. Worldwide character of the 1991 jerk, *Geophys. Res. Lett.,* **25,** 377–380.

Demmel, J.W., 1997. *Applied numerical linear algebra,* SIAM, Philadelphia.

Dyment, J. & Arkani-Hamed, J., 1998. Contribution of lithospheric remanent magnetization to satellite magnetic anomalies over the world's oceans, *J. geophys. Res.,* **103,** 15 423–15 441.

Fukushima, N., 1994. Some topics and historical episodes in geomagnetism and aeronomy, *J. geophys. Res.,* **99,** 19 113–19 142.

Holme, R., 2000. Modelling of attitude error in vector magnetic data: Application to Ørsted data, *Earth, Planets and Space,* **52,** 1187–1197.

Holme, R. & Bloxham, J., 1996. The treatment of attitude errors in satellite geomagnetic data, *Phys. Earth. planet. Int.,* **98,** 221–233.

Huber, P.J., 1964. Robust estimation of a location parameter, *Ann. Math. Statist.,* **35,** 73–101.

Langel, R.A., 1987. The main geomagnetic field, in, *Geomagnetism,* Vol. 1, pp. 249–512, ed. Jacobs, J.A., Academic Press, London.

Langel, R.A., 1990. Global magnetic anomaly maps derived from POGO spacecraft data, *Phys. Earth. planet. Int.,* **62,** 208–230.

Langel, R.A., Berbert, J., Jennings, T. & Horner, R., 1981. Magsat data processing: A report for investigators, *NASA Tech. Mem.,* **82160,** 1–178.

Langel, R.A., Estes, R.H. & Mead, G.D., 1982. Some new methods in geomagnetic field modeling applied to the 1960–1980 epoch, *J. Geomag. Geoelectr.,* **34,** 327–349.

Langel, R.A., Sabaka, T.J., Baldwin, R.T. & Conrad, J.A., 1996. The near-Earth magnetic field from magnetospheric and quiet-day ionospheric sources and how it is modeled, *Phys. Earth. planet. Int.,* **98,** 235–267.

Le Huy, M., Alexandrescu, M., Hulot, G. & Le Mouël, J.-L., 1998. On the characteristics of successive geomagnetic jerks, *Earth Planets Space,* **50,** 723–732.

Le Huy, M., Mandea, M., Le Mouël, J.-L. & Pais, A., 2000. Time evolution of the fluid flow at the top of the core. Geomagnetic jerks, *Earth Planets Space,* **52,** 163–173.

McLeod, M.G., 1985. On the geomagnetic jerk of 1969, *J. geophys. Res.,* **90,** 4597–4610.

Macmillan, S., 1996. A geomagnetic jerk for the early 1990's, *Earth planet. Sci. Lett.,* **137,** 189–192.

Maeda, H., Iyemori, T., Araki, T. & Kamei, T., 1982. New evidence of a meridional current system in the equatorial ionosphere, *Geophys. Res. Lett.,* **9,** 337–340.

Malin, S.R.C. & Hodder, B.M., 1982. Was the 1970 geomagnetic jerk of internal or external origin?, *Nature,* **296,** 726–728.

Mandea, M., Bellanger, E. & Le Mouël, J.-L., 2000. A geomagnetic jerk for the end of the 20th century?, *Earth planet. Sci. Lett.,* **183,** 369–373.

Maus, S., Rother, M., Hemant, K., Lühr, H., Kushinov, A. & Olsen, N., 2004. Earth's crustal magnetic field determined to spherical harmonic degree 90 from CHAMP satellite measurements, *Geophys. J. Int.,* submitted.

Olsen, N., 1993. The solar cycle variability of lunar and solar daily geomagnetic variations, *Ann. Geophys.,* **11,** 254–262.

Olsen, N., 1997. Ionospheric *F*-region currents at middle and low latitudes estimated from Magsat data, *J. geophys. Res.,* **102,** 4563–4576.

Olsen, N., 1998. The electrical conductivity of the mantle beneath Europe derived from C-Responses from 3 h to 720 h, *Geophys. J.,* **133,** 298–308.

Olsen, N., 2002. A model of the geomagnetic field and its secular variation for epoch 2000 estimated from Ørsted data, *Geophys. J. Int.,* **149,** 454–462.

Olsen, N. *et al.*, 2000. Ørsted initial field model, *Geophys. Res. Lett.,* **27,** 3607–3610.

O'Reilly, S.Y. & Griffin, W.L., 1985. A xenolith-derived geotherm for southeastern Australia and its geophysical implications, *Tectonophysics,* **111,** 41–63.

Purucker, M., Langlais, B., Olsen, N., Hulot, G. & Mandea, M., 2002. The southern edge of cratonic North America: Evidence from new satellite magnetometer observations, *Geophys. Res. Lett.,* **29,** 10.1029/2001GL013645.

Reading, A.M. & Kennett, B.L.N., 2003. Lithospheric structure of the Pilbara Craton, Capricorn Orogen, and northern Yilgarn Craton, Western Australia, from teleseismic receiver functions, *Australian J. Earth Sci.,* **50,** 439–445.

Richmond, A.D., 1995. Ionospheric electrodynamics using magnetic apex coordinates, *J. Geomag. Geoelectr.,* **47,** 191–212.

Richmond, A.D. & Roble, R.G., 1987. Electrodynamic effects of thermospheric winds from the NCAR thermospheric general circulation model, *J. geophys. Res.,* **92,** 12 365–12 376.

Sabaka, T.J., Langel, R.A., Baldwin, R.T. & Conrad, J.A., 1997. The geomagnetic field 1900–1995, including the large-scale field from magnetospheric sources, and the NASA candidate models for the 1995 revision of the IGRF, *J. Geomag. Geoelectr.,* **49,** 157–206.

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

Sorenson, H.W., 1980. *Parameter estimation: Principles and problems,* Dekker, New York.

Tarantola, A., 1987. *Inverse problem theory: Methods for data fitting and model parameter estimation,* Elsevier, New York.

Tarlowski, C., McEwin, A.J., Reeves, C.V., & Barton, C.E., 1996. Dewarping the composite aeromagnetic anomaly map of Australia using control traverses and base stations, *Geophysics,* **61,** 696–705.

Walker, M.R. & Jackson, A., 2000. Robust modelling of the Earth's magnetic field, *Geophys. J. Int.,* **143,** 799–808.

Wessel, P. & Smith, W.H.F., 1991. Free software helps map and display data, *EOS, Trans. Am. geophys. Un.,* **72,** 441.

APPENDIX A: EXPECTED VALUE OF $|x|$ WHEN x IS NORMALLY DISTRIBUTED

The probability density function (PDF) for a normally distributed variable x with mean μ and standard deviation σ is given by (Tarantola 1987)

$$\mathrm{P}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}. \tag{A1}$$

The expected value of $|x|$ is then given by

$$\mathrm{E}\left[|x|\right] = \int_{-\infty}^{\infty} |x| \mathrm{P}(x)\, dx, \tag{A2}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} |x| e^{-(x-\mu)^2/(2\sigma^2)}\, dx, \tag{A3}$$

$$= \frac{1}{\sqrt{2\pi}\sigma} \int_{0}^{\infty} x e^{-(x-\mu)^2/(2\sigma^2)}\, dx + \frac{1}{\sqrt{2\pi}\sigma} \int_{0}^{\infty} x e^{-(x+\mu)^2/(2\sigma^2)}\, dx, \tag{A4}$$

where E[·] is the expectation operator. If the following changes of variables are made

$$y = \frac{x-\mu}{\sqrt{2}\sigma}, dy = \frac{dx}{\sqrt{2}\sigma}, \tag{A5}$$

$$z = \frac{x+\mu}{\sqrt{2}\sigma}, dz = \frac{dx}{\sqrt{2}\sigma}, \tag{A6}$$

then eq. (A4) becomes

$$\mathrm{E}[|x|] = \frac{1}{\sqrt{\pi}}\int_{-\mu/(\sqrt{2}\sigma)}^{\infty}(\sqrt{2}\sigma y+\mu)e^{-y^2}\,dy + \frac{1}{\sqrt{\pi}}\int_{\mu/(\sqrt{2}\sigma)}^{\infty}(\sqrt{2}\sigma z-\mu)e^{-z^2}\,dz, \tag{A7}$$

$$= \frac{\mu}{\sqrt{\pi}}\int_{-\mu/(\sqrt{2}\sigma)}^{\mu/(\sqrt{2}\sigma)} e^{-y^2}\,dy + \frac{2\sqrt{2}\sigma}{\sqrt{\pi}}\int_{\mu/(\sqrt{2}\sigma)}^{\infty} ze^{-z^2}\,dz, \tag{A8}$$

$$= \frac{2\mu}{\sqrt{\pi}}\int_{0}^{\mu/(\sqrt{2}\sigma)} e^{-y^2}\,dy + \sqrt{\frac{2}{\pi}}\sigma e^{-\mu^2/(2\sigma^2)}, \tag{A9}$$

$$= \mu\cdot\mathrm{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right) + \sqrt{\frac{2}{\pi}}\sigma e^{-\mu^2/(2\sigma^2)}, \tag{A10}$$

where erf(·) is the error function defined as (Beyer 1981)

$$\mathrm{erf}(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt. \tag{A11}$$

APPENDIX B: GENERAL RESOLUTION ANALYSIS

In this Appendix, a resolution analysis will be developed that is generalized for non-linear data from subsets considered insufficient for determining independent parameter estimates. Unless otherwise stated, all evaluations with respect to the model parameter state are assumed to be at the point of convergence, $\tilde{\mathbf{x}}$. As in Section 5.3, let

$$\mathrm{C} = \left(\mathrm{A}^T\mathrm{W}\mathrm{A} + \sum_{i=1}^{K}\lambda_i\Lambda_i\right)^{-1}, \tag{B1}$$

$$\mathrm{R}_{\mathrm{d}_i} = \mathrm{C}\mathrm{A}_i^T\mathrm{W}_i\mathrm{A}_i, \tag{B2}$$

$$\mathrm{R}_{\mathrm{a}_i} = \lambda_i\mathrm{C}\Lambda_i, \tag{B3}$$

where $\mathbf{d}$ is the vector of measurements, $\mathbf{a}$ is the vector of model predictions with Jacobian A, W is the Huber weight matrix, and $\mathbf{d}_i$, $\mathbf{a}_i$, A_i and W_i are the respective subvectors or submatrices corresponding to the ith data subset. Recall that CM4 is derived not only from vector measurements, but also from satellite scalar F measurements, which are non-linear functions of $\mathbf{x}$. However, F has the special property that if $\mathbf{x} \neq \mathbf{0}$, then

$$F(\mathbf{x}) = \frac{\partial F}{\partial \mathbf{x}}\cdot\mathbf{x}, \tag{B4}$$

$$= \left(\frac{\partial F}{\partial X}\frac{\partial X}{\partial \mathbf{x}} + \frac{\partial F}{\partial Y}\frac{\partial Y}{\partial \mathbf{x}} + \frac{\partial F}{\partial Z}\frac{\partial Z}{\partial \mathbf{x}}\right)\cdot\mathbf{x}, \tag{B5}$$

$$= \frac{1}{F}(X^2+Y^2+Z^2) \tag{B6}$$

and so $\mathbf{a}(\mathbf{x}) = \mathrm{A}\mathbf{x}$ for all data used to derive CM4. Thus, at $\tilde{\mathbf{x}}$, eq. (13) becomes

$$\tilde{\mathbf{x}} = \mathrm{C}\mathrm{A}^T\mathrm{W}\mathbf{d}. \tag{B7}$$

Because $\partial^2 F/\partial\mathbf{x}^2$ is proportional to F^{-1}, it is assumed that A is constant in the neighbourhood of a terrestrial $\tilde{\mathbf{x}}$, where F is large as a result of the large dipole core field. This means that the converged $\tilde{\mathbf{x}}$ is a linear combination of all data types and that eq. (B7) is the sum of all $\tilde{\mathbf{x}}_i$ of the form

$$\tilde{\mathbf{x}}_i = \mathrm{K}_i\mathbf{d}_i, \tag{B8}$$

where

$$\mathrm{K}_i = \mathrm{C}\mathrm{A}_i^T\mathrm{W}_i. \tag{B9}$$

It is clear that a combined solution from all the data subsets is just a weighted average of the subset measurements. In fact, the ith weight, K_i, is the Kalman gain for the ith data subset (Bierman 1977). These terms would then represent that a portion of the combined parameter state resolved by the corresponding data subset. Mathematically, $\tilde{\mathbf{x}}_i$ may be interpreted as the solution resulting from an analysis in which all but the ith subset of data measurements are set to zero, i.e. how much the ith data subset requires the model to deviate from zero when all other data are forcing it to zero. Indeed, this data influence is conveyed through K_i, which is statistically related to $\mathrm{R}_{\mathrm{d}_i}$. According to the Backus–Gilbert theory for linear models (see Backus 1968, 1970), a resolution matrix is effectively a data-imposed filter through which the true model parameter vector, $\mathbf{x}$, passes. The connection of $\tilde{\mathbf{x}}_i$ with resolution may now be seen by taking the expected value of $\tilde{\mathbf{x}}_i$ such that

$$\mathrm{E}[\tilde{\mathbf{x}}_i] = \mathrm{K}_i\mathrm{E}[\mathbf{d}_i], \tag{B10}$$

$$= \mathrm{K}_i\mathrm{A}_i\mathbf{x}, \tag{B11}$$

$$= \mathrm{R}_{\mathrm{d}_i}\mathbf{x}, \tag{B12}$$

where E[·] is the expectation operator. Clearly, the entire resolution matrix plays a role in determining what portion of the parameters is resolved by which data.

If one is interested only in the resolution of a particular subset of parameters, the target parameters, then one can consider a modification to $\tilde{\mathbf{x}}$ in eq. (B7) such that it has minimal structure (length), but retains the target parameters. This can be seen by inspection to be a vector $\tilde{\mathbf{x}}_t$ whose target parameters match those of $\tilde{\mathbf{x}}$, but is zero otherwise. This smoothing essentially eliminates extraneous effects that average to zero over the data subsets and that are the result of parameters outside of the target subset. It is natural to distribute this smoothing over the right side of eq. (B7) by removing the extraneous effects, i.e. a bias, directly from the measurements. This is commensurate to subtracting the vector of extraneous parameters, $\tilde{\mathbf{x}}_{t'}$ (where $\tilde{\mathbf{x}}_t = \tilde{\mathbf{x}} - \tilde{\mathbf{x}}_{t'}$), from both sides of eq. (B7). From eq. (B1), this becomes

$$\tilde{\mathbf{x}}_t = \mathrm{C}\left(\mathrm{A}^T\mathrm{W}\mathbf{r}_t - \sum_{i=1}^{K}\lambda_i\Lambda_i\mathbf{x}_{t'}\right), \tag{B13}$$

where

$$\mathbf{r}_t = \mathbf{d} - \mathrm{A}\tilde{\mathbf{x}}_{t'}. \tag{B14}$$

It follows that the contribution of the ith data subset to $\tilde{\mathbf{x}}_t$ is

$$\tilde{\mathbf{x}}_{t,i} = \mathrm{K}_i \mathbf{r}_{t,i}, \tag{B15}$$

where $\mathbf{r}_{t,i}$ is the subvector of $\mathbf{r}_t$ corresponding to the ith data subset. Although there were no direct contributions of smoothing norms ($\mathbf{d}_i = \mathbf{0}$) to $\tilde{\mathbf{x}}$, there are now contributions to $\tilde{\mathbf{x}}_t$ of the form

$$\tilde{\mathbf{x}}_{t,i} = -\mathrm{R}_{\mathrm{a}_i} \tilde{\mathbf{x}}_{t'}. \tag{B16}$$

In the case of linear data, $\mathbf{r}_{t,i}$ is the vector of residuals with respect to the extraneous fields predicted by the best model estimate for the ith data subset. If the model design is adequate, then one would expect that only the target field signal and a random noise component would remain in $\mathbf{r}_{t,i}$. There is a slight caveat to this interpretation when considering F residuals that stems from the fact that in eq. (B4) one is evaluating $\partial F/\partial \mathbf{x}$ at $\tilde{\mathbf{x}}$ but forming the inner product with $\tilde{\mathbf{x}}_{t'}$. However, the discrepancy in the resulting approximation to $F(\tilde{\mathbf{x}}_{t'})$ is negligible if the field associated with the target parameters is relatively weak with respect to the field from all parameters. With this in mind, the target parameters of $\tilde{\mathbf{x}}_{t,i}$ may be interpreted as that portion of the target parameters of $\tilde{\mathbf{x}}$ resolved by the nominal target field remaining in $\mathbf{r}_{t,i}$. For example, the estimated lithospheric field parameters would be a linear combination of the vector lithospheric field in (X, Y, Z) data and the total anomaly field in F data; the removal of the bias has eliminated the direct effects of the extraneous fields present in the data. This is intuitively reasonable because it takes advantage of the field separation achieved during the estimation by expressing the target parameters in terms of the target fields present in the data.

It is interesting to derive expressions for the expected values and covariances of the $\tilde{\mathbf{x}}_{t,i}$. In the following discussion, let α and β designate target parameter subsets, i and j designate data subsets or particular smoothing norms, and R_i represent a generic resolution matrix such as $\mathrm{R}_{\mathrm{d}_i}$ or $\mathrm{R}_{\mathrm{a}_i}$. For example, $\tilde{\mathbf{x}}_{\alpha,i}$ is the contribution of the ith data subset or norm to target parameter subset α, $\tilde{\mathbf{x}}_{\beta,j}$ is the contribution of the jth data subset or norm to target parameter subset β, while $\mathrm{C}_{\alpha\beta,\mathrm{ij}}$ is the covariance between these two contributions. After some algebraic manipulation, it can be shown that

$$\mathrm{E}[\tilde{\mathbf{x}}_{\alpha,i}] = \mathrm{R}_i \mathbf{x}_\alpha, \tag{B17}$$

$$\mathrm{C}_{\alpha\beta,ij} = \delta_{ij}\mathrm{R}_i\mathrm{C} - \mathrm{R}_i\mathrm{C}_{\alpha\cap\beta}\mathrm{R}_j^T, \tag{B18}$$

where δ_{ij} is the Kronecker delta and $\mathrm{C}_{\alpha\cap\beta}$ is the C matrix whose elements are zero in the intersection of the rows of α and columns of β. Given that

$$\mathrm{I} = \sum_i \mathrm{R}_i, \tag{B19}$$

where the summation is over all data subsets and smoothing norms, it is straight forward to show that

$$\mathrm{C}_{\alpha\beta} = \sum_i \sum_j \mathrm{C}_{\alpha\beta,ij}, \tag{B20}$$

$$= \mathrm{C} - \mathrm{C}_{\alpha\cap\beta}. \tag{B21}$$

The (α, β) block of this matrix is equal to that of C, but otherwise is zero. Clearly, C may be partitioned into individual contributions as was $\tilde{\mathbf{x}}$.

In the previous discussion, a general resolution analysis was developed that made no assumptions about the observability of the model parameters by a data subset, i.e. the ability of a particular data subset to successfully estimate the parameters. Mathematically, the statement is equivalent to the condition of the ith data information matrix $\mathrm{E}_i = \mathrm{A}_i^T\,\mathrm{W}_i\mathrm{A}_i$; if it is well conditioned, then the ith subset of data provide an independent estimate, $\tilde{\mathbf{x}}_{[i]}$, of the parameters. Note that *a priori* information introduced as a quadratic form supplies a preferred model regardless of the condition of the norm matrix Λ_i and this preferred model is zero for smoothing norms. Assuming that E_i is invertible, $\tilde{\mathbf{x}}_{t,i}$ for the ith data subset may now be expressed as

$$\tilde{\mathbf{x}}_{t,i} = \mathrm{C}\mathrm{E}_i\mathrm{E}_i^{-1}\mathrm{A}_i^T\mathrm{W}_i\mathbf{r}_{t,i}, \tag{B22}$$

$$= \mathrm{R}_{\mathrm{d}_i}\left(\tilde{\mathbf{x}}_{[i]} - \tilde{\mathbf{x}}_{t'}\right). \tag{B23}$$

The contribution from the ith norm is the same as in eq. (B16). If E_i is invertible for all i, then $\tilde{\mathbf{x}}_t$ may be written as

$$\tilde{\mathbf{x}}_t = \sum_i \mathrm{R}_i\left(\tilde{\mathbf{x}}_{[i]} - \tilde{\mathbf{x}}_{t'}\right), \tag{B24}$$

where the summation is over all data subsets and smoothing norms and R_i is generic. Thus, from eq. (B19), one can see that $\tilde{\mathbf{x}}_t$ is a normalized weighted average of the independent estimates (adjusted by the $\tilde{\mathbf{x}}_{t'}$ baseline). Finally, if R_i has full-rank partitions, then the length of $\tilde{\mathbf{x}}_{t,i}$ is bounded as (see Appendix C)

$$|\tilde{\mathbf{x}}_{t,i}| \le \sqrt{\kappa\left(\mathrm{E}_i\right)}\left(1 + \frac{1}{|\mathrm{E}_i|\,\left|\mathrm{F}_i^{-1}\right|}\right)^{-1}\left|\tilde{\mathbf{x}}_{[i]} - \tilde{\mathbf{x}}_{t'}\right|, \tag{B25}$$

where $\kappa(\mathrm{E}_i)$ is the condition number of E_i and

$$\mathrm{F}_i = \sum_k^{k\neq i} \mathrm{E}_k. \tag{B26}$$

If E_i and F_i are both well conditioned, but the information in the latter dominates the former, then the length of $\tilde{\mathbf{x}}_{t,i}$ will generally be less than that of $\tilde{\mathbf{x}}_{[i]} - \tilde{\mathbf{x}}_{t'}$ (see Appendix C, eq. (C19)), reflecting the normalized weighting in eq. (B24).

APPENDIX C: UPPER BOUND ON THE ℓ_2 NORM OF A RESOLUTION MATRIX WITH FULL-RANK PARTITIONS

The ℓ_2 norm of a general matrix A is defined as (Demmel 1997)

$$|\mathrm{A}| \equiv \sqrt{\lambda_{\max}\left(\mathrm{A}^*\mathrm{A}\right)}, \tag{C1}$$

where $\lambda_{\max}(\cdot)$ denotes the largest eigenvalue and $*$ denotes the conjugate transpose. Three key properties of the ℓ_2 norm will be needed: If S and A are symmetric and general matrices, respectively, then

$$|\mathrm{S}| \equiv \lambda_{\max}\left(\mathrm{S}\right), \tag{C2}$$

$$|\mathrm{S}^{-1}| = \left(\lambda_{\min}\left(\mathrm{S}\right)\right)^{-1}, \tag{C3}$$

$$|\mathrm{A}| = |\mathrm{A}^T|, \tag{C4}$$

where $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue.

Let R_k be the resolution matrix for the kth data subset such that

$$\mathrm{R}_k = \left(\mathrm{E}_k + \mathrm{F}_k\right)^{-1}\mathrm{E}_k, \tag{C5}$$

where E_k is the normal matrix for the kth data subset and F_k is the sum of the remaining normal matrices. These two matrices will be termed the partitions of R_k. These partitions are both Gram matrices, i.e. can be expressed in the form $\mathrm{E}_k = \mathrm{A}^T\mathrm{A}$ and $\mathrm{F}_k = \mathrm{B}^T\mathrm{B}$ for some matrices A and B. Therefore, they are both symmetric positive semi-definite (SPSD) and so have a full set of real non-negative eigenvalues. However, it is assumed here that the partitions are also invertible, which means their eigenvalues are

positive. Therefore, there exists a Cholesky factorization for E_k of the form

$$E_k = L_k L_k^T, \tag{C6}$$

where L_k is lower triangular and invertible. Note that the invertibility of E_k assures that R_k is full-rank and that the following useful factorization exists

$$R_k = \left(L_k L_k^T + F_k\right)^{-1} L_k L_k^T, \tag{C7}$$

$$= L_k^{-T}\left(I + L_k^{-1} F_k L_k^{-T}\right)^{-1} L_k^{-1} L_k L_k^T, \tag{C8}$$

$$= L_k^{-T}\left(I + L_k^{-1} F_k L_k^{-T}\right)^{-1} L_k^T. \tag{C9}$$

First, the upper bound on $\left|(I + L_k^{-1}\, F_k\, L_k^{-T})^{-1}\right|$ is derived as

$$\left|\left(I + L_k^{-1} F_k L_k^{-T}\right)^{-1}\right| = \frac{1}{\lambda_{\min}\left(I + L_k^{-1} F_k L_k^{-T}\right)}, \tag{C10}$$

$$= \frac{1}{1 + \lambda_{\min}\left(L_k^{-1} F_k L_k^{-T}\right)}, \tag{C11}$$

$$= \left(1 + \frac{1}{\left|L_k^T F_k^{-1} L_k\right|}\right)^{-1}, \tag{C12}$$

$$\leq \left(1 + \frac{1}{|L_k|^2 \left|F_k^{-1}\right|}\right)^{-1}, \tag{C13}$$

$$= \left(1 + \frac{1}{|E_k| \left|F_k^{-1}\right|}\right)^{-1}. \tag{C14}$$

The upper bound on $|R_k|$ is now straight forward to derive

$$|R_k| = \left|L_k^{-T}\left(I + L_k^{-1} F_k L_k^{-T}\right)^{-1} L_k^T\right|, \tag{C15}$$

$$\leq \left|L_k^{-T}\right| \left|L_k^T\right| \left|\left(I + L_k^{-1} F_k L_k^{-T}\right)^{-1}\right|, \tag{C16}$$

$$\leq \left|L_k^{-T}\right| \left|L_k^T\right| \left(1 + \frac{1}{|E_k| \left|F_k^{-1}\right|}\right)^{-1}, \tag{C17}$$

$$= \sqrt{\kappa\,(E_k)} \left(1 + \frac{1}{|E_k| \left|F_k^{-1}\right|}\right)^{-1}, \tag{C18}$$

where $\kappa(E_k)$ is the condition number of E_k, i.e. the ratio of the largest to smallest eigenvalues of E_k. It can be shown, after some algebra, that $|\,R_k| \leq 1$ is guaranteed whenever

$$\lambda_{\min}(F_k) \geq \lambda_{\max}(E_k)\left(\sqrt{\kappa(E_k)} - 1\right). \tag{C19}$$

© 2004 RAS, *GJI*