Earth Planets Space, **xx**, 1–19, 2010

Evaluation of candidate geomagnetic field models for IGRF-11

C. C. Finlay[1], S. Maus[2], C. D. Beggan[3], M. Hamoudi[4], F. J. Lowes[5], N. Olsen[6], and E. Thébault[7].

[1] *Earth and Planetary Magnetism Group, Institut für Geophysik, Sonneggstrasse 5, ETH Zürich, CH-8092, Switzerland.*
[2] *NOAA/NGDC and CIRES, University of Colorado, U.S.A.*
[3] *British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, U.K.*
[4] *Helmholtz Centre Potsdam, GFZ German Research centre for Geosciences, Telegrafenberg, 14473, Germany.*
[5] *School of Chemistry, University of Newcastle, Newcastle Upon Tyne, NE1 7RU, U.K.*
[6] *DTU Space, Juliane Maries Vej 30, 2100 Copenhagen, Denmark*
[7] *Équipe de géomagnétisme, Institut de Physique du Globe de Paris, UMR 7154, CNRSINSU, Univ. Paris Diderot, Paris, France*

(Received xxxx xx, 2010; Revised xxxx xx, 2010; Accepted xxxx xx, 2010; Online published Xxxxx xx, 2010)

The eleventh revision of the International Geomagnetic Reference Field (IGRF) was agreed in December 2009 by a task force appointed by the International Association of Geomagnetism and Aeronomy (IAGA) Division V Working Group V-MOD. New spherical harmonic main field models for epochs 2005 (DGRF-2005) and 2010 (IGRF-2010) and predictive linear secular variation for the interval 2010-2015 (SV-2010-2015) were derived from weighted averages of candidate models submitted by teams led by DTU Space, Denmark (team A); NOAA/NGDC, U.S.A. (team B); BGS, U.K. (team C); IZMIRAN, Russia (team D); EOST, France (team E); IPGP, France (team F); GFZ, Germany (team G) and NASA-GSFC, U.S.A. (team H). Here, we report the evaluations of candidate models carried out by the IGRF-11 task force during October/November 2009 and describe the weightings used to derive the new IGRF-11 models. The evaluations include calculations of root mean square vector field differences between the candidates, comparisons of the power and degree correlations between the candidates and a mean model. Coefficient by coefficient analysis including determination of weight factors used in a robust estimation of mean coefficients is also reported. Maps of differences in the vertical field intensity at Earth's surface between the candidates and weighted mean models are presented. Candidates with anomalous aspects are identified and efforts made to pinpoint both troublesome coefficients and regions in physical space where large variations between candidates originate. A retrospective evaluation of IGRF-10 main field candidates for epoch 2005 and predictive secular variation candidates for 2005-2010 using the new IGRF-11 models as a reference is also reported. The high quality and consistency of main field models derived using vector satellite data is demonstrated; based on internal consistency DGRF-2005 has a formal root mean square vector error over Earth's surface of 1.0 nT. Difficulties nevertheless remain in accurately forecasting field evolution only five years into the future.

Key words: geomagnetism, field modelling, reference field, secular variation.

1. Introduction

The IGRF is an internationally agreed spherical harmonic reference model describing the largest scales of the internal part of the Earth's magnetic field. It is widely used by scientists studying local and regional crustal magnetic anomalies, by those studying space weather and solar-terrestrial magnetic interaction, and it is also sometimes used by individuals and commercial organizations for navigational purposes. Under normal circumstances the IGRF is updated every 5 years; for a history of IGRF and further background information consult Barton (1997) or Macmillan and Finlay (2010). An IGRF update involves collaboration between institutes collecting and disseminating geomagnetic measurements derived from satellites and ground-based observatories, and between teams of geomagnetic field modellers.

In May 2009 the task force for the eleventh-generation revision of IGRF, working under the auspice of IAGA, issued a call for main field (MF) candidate models for the Definitive Geomagnetic Reference Field for epoch 2005 (DGRF-2005), for the provisional IGRF for epoch 2010 (IGRF-2010) both to spherical harmonic degree 13, and for a prediction of the average secular variation (SV) over the upcoming five years (SV-2010-2015) to spherical harmonic degree 8. Seven MF candidate models were submitted for DGRF-2005 and IGRF-2010 while eight candidates were submitted for SV-2010-2015. Institutions leading the teams were based in Denmark, U.S.A., U.K., Russia, France and Germany making this a truly international enterprise.

The primary sources of data employed were from the German satellite CHAMP, the Danish satellite Ørsted and the Argentine-U.S.-Danish satellite SAC-C, along with data from the international network of geomagnetic observatories. The

teams adopted a variety of data selection and processing procedures. Furthermore, the required single epoch spherical harmonic model coefficients were derived from parent models that used a range of time durations (1 month to 10 years), temporal parameterizations (including Taylor series of degree 0 to 2, splines of order 1 to 6), and external field parameterizations of varying complexity. The parent models also used a number of alternative parameter estimation schemes (including least-squares, least absolute deviations, robust estimation based on Huber's distribution and natural orthogonal analysis). Further details concerning the techniques used to derive the individual models can be found in the papers appearing in this special issue (Olsen et al., 2010; Maus et al., 2010; Hamilton et al., 2010; Chambodut et al., 2010; Thébault et al., 2010; Lesur et al., 2010; Kuang et al., 2010). The candidate model coefficients and descriptions provided by the authors are also available from the web page http://www.ngdc.noaa.gov/IAGA/vmod/candidatemodels.html. The different strategies adopted naturally lead to differences in the submitted candidate models. The IGRF-11 task force therefore undertook testing and inter-comparison of the candidates to provide the information required for decisions on the weights to be used in the construction of the final IGRF-11 models.

The purpose of this article is to summarize the evaluations of the IGRF-11 candidates carried out by the task force. We follow closely the strategy adopted in previous evaluations (see, for example, Maus et al., 2005) focusing on statistical comparisons between the candidate models and various mean models, and utilizing well established diagnostic tools in both the spectral space and physical space. Model evaluations would ideally be based not only upon statistical analysis of candidates, but also on comparisons with independent data that accurately measured the relevant field (the internal magnetic field at Earth's surface) at the epochs of interest. Unfortunately such ideal evaluation data did not exist for the future epochs of 2010 and 2010-2015 at the time of the evaluations and it is even troublesome to obtain high quality independent data for the retrospective epoch 2005. Attempts to assess the candidate models using either observatory or satellite data are thus complicated by the necessity of propagating the models to suitable comparison epochs as well as with difficulties in separating internal and external field contributions in the observed data. Nonetheless, some workers have made interesting attempts at such comparisons, see for example the study by Chulliat and Thébault (2010) also in this issue.

As a mathematical preliminary, we begin in section 2 by providing the formulae defining the analysis tools employed. In section 3 MF candidates are studied, while section 4 presents evaluations of SV candidates. In section 3.1 we analyze the candidate models for DGRF-2005, then in section 3.2 a retrospective evaluation of the IGRF-10 candidates for epoch 2005 in comparison with the new DGRF-2005 model is carried out. In section 3.3 evaluations of the candidates for IGRF-2010 are presented followed in section 4.1 by a retrospective analysis of the predictive SV candidates for the epoch 2005-2010 from IGRF-10. Finally in section 4.2 the IGRF-11 predictive SV candidates for epoch 2010-2015 are analyzed. In each case global comparisons of root mean square (RMS) vector field differences are made first, then comparisons in the spectral domain, per degree and then coefficient by coefficient; finally maps of differences between candidate models and a weighted mean model in physical space are presented. Discussion of the evaluation results and a summary of the decision of the task force is provided for each IGRF-11 product. We conclude with an overall summary and some remarks on the implications of these evaluations for the future of IGRF.

2. Mathematical definitions and formulae used in evaluations

Formulae defining the diagnostic tools employed during the evaluations are first presented here to avoid ambiguity. The IGRF-11 candidate models take the form of Schmidt semi-normalized (sometimes also referred to as quasi-normalized) spherical harmonic coefficients (see, for example, Winch et al., 2004) with units of nT for MF models and nT/yr for SV models. In what follows g_n^m and h_n^m are used to denote the spherical harmonic coefficients associated with the $\cos m\phi$ and $\sin m\phi$ components respectively, where ϕ denotes geocentric longitude. As is conventional n denotes spherical harmonic degree while m denotes spherical harmonic order. Often we will be concerned with differences between a candidate model i whose coefficients we denote by ${}_ig_n^m$ and ${}_ih_n^m$ and some other reference model (labelled j) whose coefficients will be denoted by ${}_jg_n^m$ and ${}_jh_n^m$. It is also convenient at this point to define the difference between the coefficients of two such models as

$$ {}_{i,j}g_n^m = {}_ig_n^m - {}_jg_n^m \quad \text{and} \quad {}_{i,j}h_n^m = {}_ih_n^m - {}_jh_n^m. \tag{1} $$

Much use will be made below of the mean square vector field difference between models per spherical harmonic degree ${}_{i,j}R_n$ (see, for example, Lowes (1966) and Lowes (1974))

$$ {}_{i,j}R_n = (n+1)\left(\frac{a}{r}\right)^{(2n+4)} \sum_{m=0}^{n} \left[({}_{i,j}g_n^m)^2 + ({}_{i,j}h_n^m)^2\right] \tag{2} $$

where a is the magnetic reference spherical radius of 6371.2km which is close to the mean Earth radius, and r is the radius of the sphere of interest, which is taken as $r = a$ for comparisons at the Earth's surface and $r = 3480$ km for comparisons at the core-mantle boundary. Summing over degrees n from 1 to the truncation degree N and taking the square root yields the RMS difference between the models i and j averaged over the spherical surface

$$ {}_{i,j}R = \sqrt{\sum_{n=1}^{N} {}_{i,j}R_n}. \tag{3} $$

It is sometimes informative to calculate ${}_{i,j}R$ when the reference model j is a weighted mean of the K candidates models with each model allocated a certain weight ${}_iw$. The coefficients of the weighted mean model M_w are then

$$\widetilde{g_n^m} = \frac{\sum_{i=1}^{K} {}_iw \ {}_ig_n^m}{\sum_{i=1}^{K} {}_iw} \quad \text{and} \quad \widetilde{h_n^m} = \frac{\sum_{i=1}^{K} {}_iw \ {}_ih_n^m}{\sum_{i=1}^{K} {}_iw}. \tag{4}$$

The precise details of the various weightings used will be be discussed in detail below. In the special case when all ${}_iw = 1$ we obtain the simple arithmetic mean model (which we refer to below as model M) with coefficients

$$\overline{g_n^m} = \frac{1}{K}\sum_{i=1}^{K} {}_ig_n^m \quad \text{and} \quad \overline{h_n^m} = \frac{1}{K}\sum_{i=1}^{K} {}_ih_n^m. \tag{5}$$

In addition to calculating ${}_{i,j}R$ for individual models, it is also possible to compute the mean value of ${}_{i,j}R$ for the ith model compared to the $(K-1)$ other candidates labelled by j, such that

$${}_i\overline{R} = \frac{1}{(K-1)} \sum_{candidates \ j \neq i} {}_{i,j}R. \tag{6}$$

Taking the special case when the reference model is zero (2) reduces to the standard Lowes-Mauersberger geomagnetic power spectrum R_n for a given model

$$R_n = (n+1)\Big(\frac{a}{r}\Big)^{(2n+4)} \sum_{m=0}^{N} \left[(g_n^m)^2 + (h_n^m)^2\right]. \tag{7}$$

Analysis of spherical harmonic spectra is a powerful way to diagnose differences in amplitude between models but tells us little about how well they are correlated. The correlation per degree between two models again labelled by the indices i and j can be studied as a function of spherical harmonic degree using the quantity ${}_{i,j}\rho_n$ (see, for example p.81 of Langel and Hinze (1998))

$${}_{i,j}\rho_n = \frac{\sum_{m=0}^{n} ({}_ig_n^m \, {}_jg_n^m + {}_ih_n^m \, {}_jh_n^m)}{\sqrt{\left(\sum_{m=0}^{n} [({}_ig_n^m)^2 + ({}_ih_n^m)^2]\right)\left(\sum_{m=0}^{n} [({}_jg_n^m)^2 + ({}_jh_n^m)^2]\right)}}. \tag{8}$$

The degree correlation between a given model i and the arithmetic mean model M that is frequently considered below may then be defined as

$${}_i\rho_n = \frac{\sum_{m=0}^{n} ({}_ig_n^m \overline{g_n^m} + {}_ih_n^m \overline{h_n^m})}{\sqrt{\left(\sum_{m=0}^{n} [({}_ig_n^m)^2 + ({}_ih_n^m)^2]\right)\left(\sum_{m=0}^{n} \left[(\overline{g_n^m})^2 + (\overline{h_n^m})^2\right]\right)}}. \tag{9}$$

Assuming that the candidate models are independent, that they involve only random errors, and that these errors have a standard deviation at degree n common to all the K contributing models, then this common sample standard deviation can be estimated from the scatter about the mean. Expressed in terms of a per degree sample standard deviation s_n, the rms scatter of the resulting field, is given by

$$s_n = \sqrt{\frac{n-1}{K-1}\sum_{i=1}^{K}\sum_{m=0}^{n} ({}_ig_n^m - \overline{g_n^m})^2 + ({}_ih_n^m - \overline{h_n^m})^2}. \tag{10}$$

The corresponding standard error in the arithmetic mean determined from these K models is then

$$e_n = \frac{s_n}{\sqrt{K}}. \tag{11}$$

A final statistical tool of interest is the method of 'robust' estimation (see, for example, Huber, 1996; Hogg, 1979). This approach is known to be of value when error distributions are non-Gaussian, in particular if outliers are present. During the IGRF-11 evaluation process, in an investigation of the possible applicability of this method, the 'robust' weighted mean of each spherical harmonic coefficient was determined treating the set of values for each coefficient as an independent data

set. The weights entering this calculation were determined by an error distribution that is often referred to as the Huber distribution

$$H(\epsilon) = \frac{1}{N_c} \begin{cases} \exp(-\epsilon/2), & |\epsilon| < c \\ \exp(-c|\epsilon| + c/2), & |\epsilon| \geq c \end{cases} \tag{12}$$

where ϵ is the normalized departure from the mean, $c = 1.5$ is a parameter chosen for a compromise between a Laplacian distribution (obtained when $c = 0$) and a Gaussian distribution (obtained when $c \to \infty$), and $N_c = 2.6046$ is a normalization constant. This distribution treats large departures from the mean as coming from a Laplacian distribution, thus avoiding undue influence on the parameter estimate. Maximum likelihood estimates of a robust mean with the errors assumed distributed as in (12) can conveniently be determined by an iteratively-reweighted least squares (IRLS) procedure (Constable, 1988; Olsen, 2002). In this method for the qth iteration the weight for the ith model for a given spherical harmonic coefficient labelled by α i.e. $({}_i w_\alpha)_q$ are determined from the associated residuals from the current weighted mean ${}_i\epsilon_\alpha)_q$ such that

$$({}_i w_\alpha)_q = \min(c/|({}_i\epsilon_\alpha)_q|, 1.0). \tag{13}$$

Below we will plot the converged weights ${}_i w_\alpha$ for each spherical harmonic coefficient of each candidate model, i.e. for all ${}_i g_m^n$, ${}_i h_m^n$, in order to compare candidates. Coefficients allocated low weights are effectively identified as outliers under this scheme. However, note once again that this procedure treats each spherical harmonic coefficient α as independent, and the 'robust mean' coefficients neglect any prior information that may be gleaned from other coefficients. This may be particularly problematic if candidate models contain strongly correlated Gauss coefficients. Thus, we use the Huber weights only as a diagnosis tool and do not use them to determine the final weights given to the candidate models.

Having defined the tools used in the evaluations we proceed to present the results of the analysis, together with related discussion of the weightings allocated to candidates in the final IGRF models.

3. Evaluation of main field candidate models

3.1 Analysis of IGRF-11 DGRF-2005 candidates

Table 1 lists the seven candidates models for DGRF 2005 giving details of the teams, the major data sources used and very brief comments concerning the various modelling approaches adopted. Two candidates (C2 and E2) were resubmissions because the original candidates were withdrawn by their authors.

DGRF candidate models for main field epoch 2005				
Team	Model	Organization	Data	Comments (parent model etc.)
A	DGRF-2005-A	DTU Space / IPGP / GSFC-NASA	Ørsted, CHAMP, SAC-C revised obs mon. means	Based on CHAOS-3α in 2005.0 (6th order splines for parent)
B	DGRF-2005-B	NGDC-NOAA / GFZ	CHAMP 2003.5-2006.5	Based on POMME 6 2nd order Taylor series
C	DGRF-2005-C2	BGS	Ørsted, CHAMP and obs. hr. means for 01:30, 1999.0-2009.5	Revised submission: parent model linear splines (400day knots spacing)
D	DGRF-2005-D	IZMIRAN	CHAMP 2004.0-2006.0 no data selection	Natural Orthogonal Components Method with 5 terms
E	DGRF-2005-E2	EOST / LPGN / / LATMOS / IPGP	CHAMP & Ørsted 2004.5-2005.5	Revised submission: based on 12 month model with linear SV
F	DGRF-2005-F	IPGP / EOST / LPGN / LATMOS	CHAMP 2004.4-2005.7	2nd order (to n=5) Taylor series
G	DGRF-2005-G	GFZ	CHAMP 2001-2009.6 obs. hourly means	Based on GRIMM2x (6th order splines for parent) av. over 1 yr.

Table 1. Summary of DGRF-2005 candidate models submitted to IGRF-11.

3.1.1 RMS vector field differences for DGRF-2005 candidates

Rows of Table 2 present the RMS vector field differences ${}_{i,j}R$ in units of nT between a particular DGRF candidate model i and another candidate j. The final three columns document ${}_{i,j}R$ between a candidate model i and one of three mean possible mean models j. The mean models considered are the arithmetic mean model M, the model M_{noD} which is an arithmetic mean excluding candidate D, and the model M_{ABG} that is an arithmetic mean model derived only from candidates A, B and G. Note the symmetry about the diagonal entries in this table which is included as a check on the calculations. It is readily observed that model D is consistently furthest away from the other models in terms of ${}_{i,j}R$; furthermore the RMS vector field differences between the other candidates and the mean are reduced when D is removed from the calculation of the mean. On the other hand models A, B, G are found to be extremely similar displaying the smallest RMS vector field differences between each other. Besides candidate D, candidates C2 and E2 show the next largest ${}_{i,j}R$ followed by F.

The final three rows of table 2 involve the arithmetic means of the RMS vector field differences of ${}_{i,j}R$ of model i from the other models j. The third from last row is ${}_{i}\overline{R}$, the penultimate row is the same calculation excluding candidate D while the final row involves only ${}_{i,j}R$ from candidates A, B and G. Candidates A, B and G have the smallest ${}_{i}\overline{R}$ and the mean of the ${}_{i,j}R$ becomes smaller when only candidates A, B and G are retained.

${}_{i,j}R$ / nT	A	B	C2	D	E2	F	G	M	M_{noD}	M_{ABG}
A	0.0	2.3	4.3	14.9	5.4	4.6	2.9	3.1	2.0	1.6
B	2.3	0.0	4.8	14.5	5.2	3.8	2.2	2.6	1.7	1.2
C2	4.3	4.8	0.0	15.2	6.8	6.5	5.2	4.6	4.0	4.6
D	14.9	14.5	15.2	0.0	14.6	15.0	14.4	12.4	14.4	14.5
E2	5.4	5.2	6.8	14.6	0.0	5.6	5.6	4.5	4.2	5.2
F	4.6	3.8	6.5	15.0	5.6	0.0	4.4	4.1	3.4	4.0
G	2.9	2.2	5.2	14.4	5.6	4.4	0.0	3.0	2.4	1.6
Mean Diff	5.7	5.5	7.1	14.7	7.2	6.6	5.8	4.9	4.6	4.7
Mean Diff noD	3.9	3.7	5.5	17.7	5.7	5.0	4.1	3.7	3.0	3.0
Mean Diff ABG	2.6	2.2	4.8	14.6	5.4	4.2	2.6	2.9	2.0	1.4

Table 2. RMS vector field differences ${}_{i,j}R$ in units nT between DGRF-2005 candidate models and also between candidates and the arithmetic mean reference models M, M_{noD} and M_{ABG} shown in the rightmost columns. The bottom three rows are arithmetic means ${}_{i}\overline{R}$ of the ${}_{i,j}R$ where the means include respectively all candidates, exclude candidate D, and use only models A, B and G.

3.1.2 Analysis in spectral space of DGRF-2005 candidates

Figure 1 (left) presents the Lowes-Mauersberger spectra R_n (defined in (7)) of the DGRF-2005 candidate models as a function of spherical harmonic degree plotted at the Earth's core-mantle boundary ($r = 3480$km). The spectra of the candidate models are mostly very similar, almost completely overlapping for degrees less than 9. The most noticeable differences occur for candidate D at degree 11 (where it contains lower power than the other candidates) and for candidate E2 at degree 13 (where it contains higher power than the other candidates). Figure 1 (right) presents the degree correlation ${}_{i}\rho_n$ as defined in (9) between the DGRF-2005 candidate models and the arithmetic mean model M. Candidate D displays a low degree correlation to M above degree 9. The degree correlation of candidates C2, E2 and F to M above degree 10 is slightly lower than that of A, B and G which appear similar to each other and close to M.

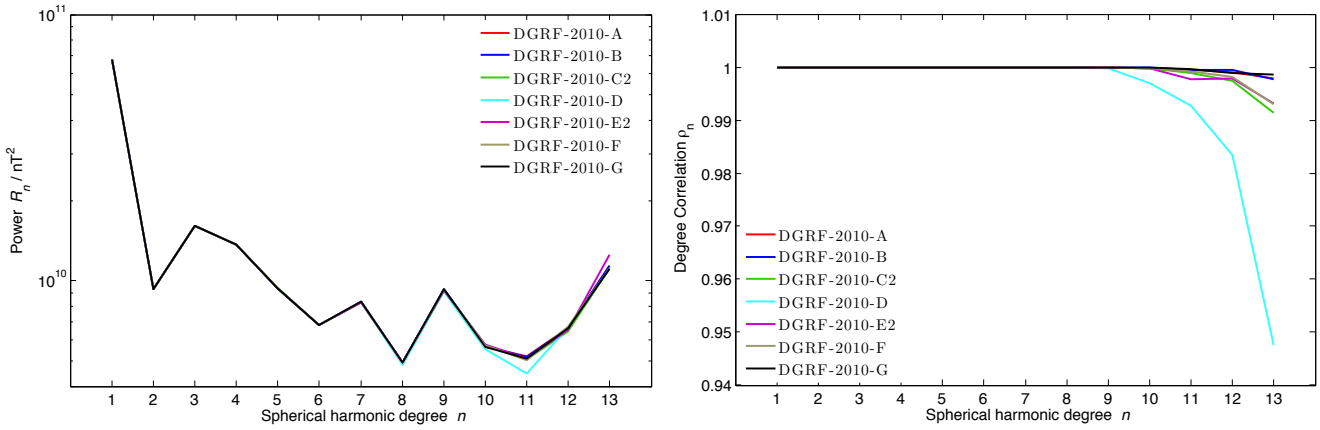

Fig. 1. Lowes-Mauersberger spectra R_n from (7) of DGRF-2005 candidate models at radius 3480km (core-mantle boundary) (left) and degree correlation ${}_{i}\rho_n$ from (9) between DGRF-2005 candidate models and their arithmetic mean model M (right).

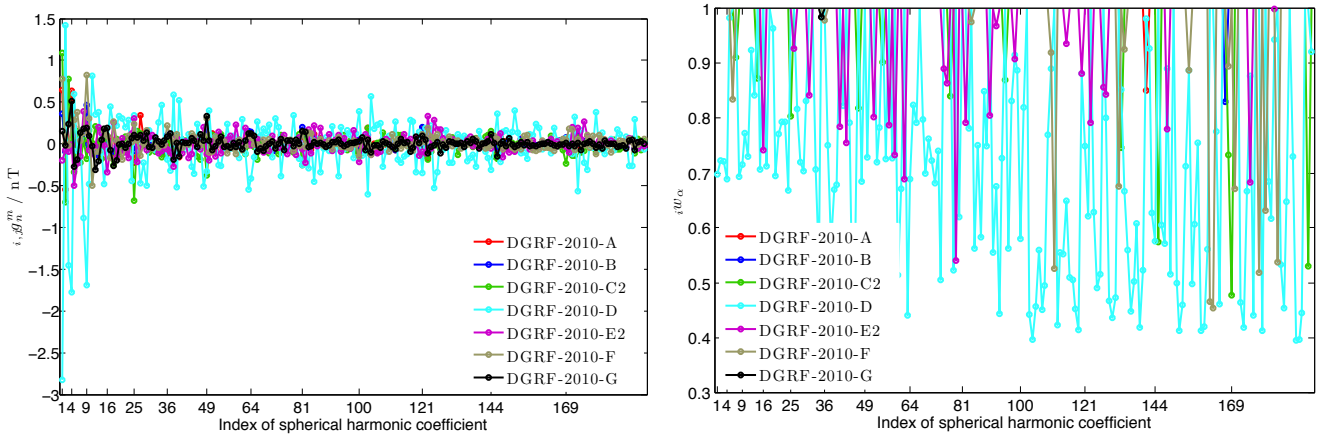

Fig. 2. Left plot shows differences ${}_{i,j}g_n^m$ as defined in (1) between DGRF-2005 candidate models and their mean model M as a function of the index of the spherical harmonic coefficient (running from $g_1^0, g_1^1, h_1^1, g_2^0, h_1^1$ etc indexed 1,2,3,4,5, etc). Right plot shows Huber robust weighting factor ${}_iw_\alpha$, where 1.0 indicates full weight 0.0 indicates zero weight, also as a function of the spherical harmonic coefficient.

In Figure 2 coefficient by coefficient analysis of the DGRF-2005 candidate models is presented. The plot on the left shows differences ${}_{i,j}g_n^m$ as defined in (1) between the candidate models and the arithmetic mean model M. The largest differences from M are found to occur for candidate D, with significant deviations also notable for candidates E2, C2 and F. The deviations associated with candidates A, B and G are smaller, so that the curves for candidates A and B are largely hidden behind those for the other candidates. The right hand plot shows the Huber weights calculated during the determination of robust mean coefficients. Notice that the coefficients of candidate D often receive the lowest weights, particularly for the coefficients associated with the highest harmonics which receive weights as low as 0.4. Candidates E2, C2 and F also receive low weights for certain coefficients; E2 also receives some fairly low weights for coefficients between n=6 and n=9. Almost all coefficients of candidates A, B and G receive full weights of 1.0 illustrating that they are consistently closer to the robust mean, so are apparently of consistently higher quality.

3.1.3 Analysis in physical space of DGRF-2005 candidates

An investigation of the DGRF-2005 candidate models in physical space is presented in Figure 3. This shows the differences between the vertical (Z) component of the candidates and model M_{ABG} at radius $r = a$. Model M_{ABG} was chosen as a suitable reference based on the earlier analyses presented in sections 3.1.1 and 3.1.2.

Studying differences between the candidate models and a reference model in physical space yields insight into the geographical locations where disparities in the candidates are located. Visual inspection of Figure 3 reveals that candidate D involves the most striking deviations from M_{ABG} that are locally as large as 50 nT. The differences are scattered over the globe and not confined to any particular geographical location, though the largest discrepancies occur in the polar regions and to the east of Australia. Candidates C2 and E2 display largest deviations from A, B and G in the polar regions (particular in the Arctic). Model E2 shows one localized anomalous region under equatorial Africa while model F shows rather minor differences at high latitudes and at mid-latitudes in the northern hemisphere. Candidates A, B and G only exhibit minor differences to the reference model M_{ABG} demonstrating once more that they are consistent with each other.

3.1.4 Choice of numerical precision for DGRF-2005

An important analysis for DGRF-2005 was to calculate (using (10) and (11)) the error per degree in the unweighted arithmetic means determined for sets of candidate models. Figure 4 shows the result of such a calculation using candidates A, B and G, on the assumption that all the candidates have the same per degree sample deviation s_n, which is estimated from their scatter about the mean. The dark blue line shows the resulting error in the mean per degree for model M_{ABG} which is typically around 0.3 nT. The red dashed line in Figure 4 shows the expected uncertainty due to rounding the model coefficients to 0.1 nT, given by the expression $0.1\sqrt{(2n+1)(n+1)/12}$ (see, for example, Lowes, 2000). It is observed that the error due to 0.1nT rounding dominates the error in the mean of candidates A, B and G above degree 7. Given the decision by the task force (see next section) to adopt model M_{ABG} for the DRGF-2005, this necessitates quoting DGRF-2005 to 0.01 nT rather than 0.1 nT to avoid introducing unnecessary rounding errors. Note that based on internal consistency, the total formal RMS error in the mean model M_{ABG} (which is DGRF-2005) is remarkably only 1.0 nT.

3.1.5 Discussion and summary for DGRF-2005

Based on the tests presented above, candidate D appears consistently different in both the spectral domain (with certain spherical harmonic coefficients apparently anomalous- see Figure 2) as well as in physical space where global problems are observed. In addition candidates E2, C2 and to lesser extend F were observed to have some problems, particularly at high degrees in the spectral domain and at high latitudes in physical space. In contrast candidates A, B and G were very similar despite being derived using different data selection criteria and using different modelling procedures. The task force therefore voted that DGRF-2005 be derived from a simple arithmetic mean of candidates A, B and G (i.e. model M_{ABG} as

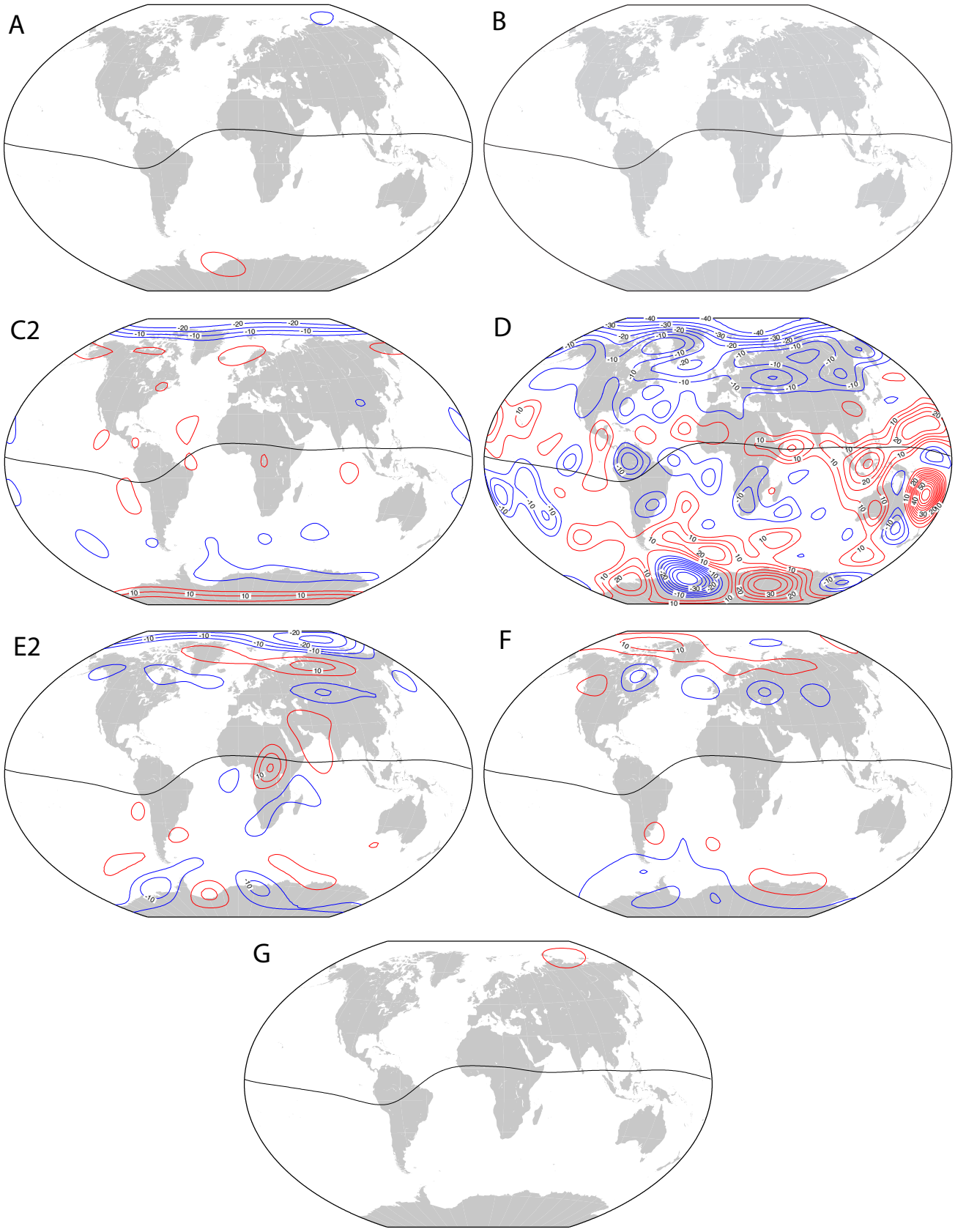

Fig. 3. Difference in the Z component of the magnetic field between each DGRF-2005 candidate model and the mean model M_{ABG} (i.e. the final DGRF-2005) plotted at Earth's reference radius in Winkel tripel projection projection. Contours are at intervals of 5 nT with labels every 10 nT when sufficiently large. Red: positive, Blue: negative. The dip equator is shown.

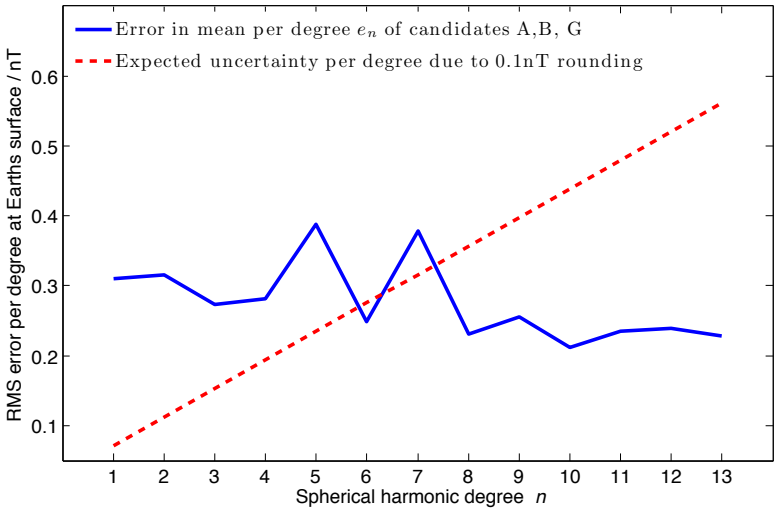

Fig. 4. Standard error e_n in the mean per degree defined by (10) and (11), calculated from the mean of DGRF-2005 candidates A, B and G is shown in blue. It was assumed that (within a given degree) all the candidates had the same standard deviation. The dashed red curve shows the expected uncertainty due to rounding to 0.1nT, given by the expression $0.1\sqrt{(2n+1)(n+1)/12}$. Note that above degree 7 the uncertainty due to rounding is greater than the error in the mean.

discussed above).

3.2 Retrospective analysis of IGRF-10 MF candidates for epoch 2005

Having established a new DGRF for epoch 2005 it is possible to carry out an assessment of the quality of the candidate models that contributed to the IGRF-10 provisional model for epoch 2005. Table 6 presents the RMS vector field differences ${}_{i,j}R$ between the various candidate models, the IGRF-2005 model (from IGRF-10) and the DGRF-2005 model (from IGRF-11). The naming convention for the candidates is that used by Maus et al. (2005). Candidate A1 agrees most closely with DGRF-2005 with a global RMS vector field difference of 9.9 nT followed closely by B3 which differs by 10.9 nT. Candidate D1 does a little worse with a difference 14.0 nT and candidate C1 is furthest from DGRF-2005 with global averaged vector field difference of 18.5 nT, almost twice that of candidate A1. The IGRF-2005 (which was the arthimetic mean of candidates A1, B3 and C1) differed from DGRF-2005 by 12.0 nT.

${}_{i,j}R$	IGRF-2005-A1	IGRF-2005-B3	IGRF-2005-C1	IGRF-2005-D1	IGRF-2005	DGRF-2005
IGRF-2005-A1	0.0	8.0	14.6	15.8	7.0	9.9
IGRF-2005-B3	8.0	0.0	11.4	15.6	4.6	10.9
IGRF-2005-C1	14.6	11.4	0.0	20.4	8.3	18.5
IGRF-2005-D1	15.7	15.6	20.4	0.0	16.1	14.0
IGRF-2005	7.0	4.6	8.3	16.1	0.0	12.0
DGRF-2005	9.9	10.9	18.5	14.0	12.0	0.0

Table 3. RMS vector field differences ${}_{i,j}R$ in units of nT between candidate models for IGRF-10 epoch 2005, the IGRF-2005 from IGRF-10 and the DGRF-2005 from IGRF-11. Note the symmetry about the diagonal, included as a check on the calculations.

In Figure 5 the difference in power per degree between the IGRF-10 candidates and DGRF-2005 (${}_{i,j}R_n$) are presented. It appears that the problems with candidate D1 are predominantly at high degree ($n > 7$); it is better than most other candidates at the lower degrees. Candidate C1 was further from DGRF-2005 than all the other candidates even at low degrees 1– 8 suggesting some systematic problem with this model. It is also noticeable that candidate A1 did better than the other candidates for the dipole ($n = 1$) terms while candidate B3 performed best at high degrees, especially $n = 12, 13$.

3.3 Analysis of IGRF-11 MF candidates for epoch 2010

Having completed the analysis of MF models for epoch 2005 we now move on to consider epoch 2010. Table 4 summarizes the candidate models submitted for IGRF-2010. Note that model C2 was a resubmission by BGS who withdrew their initial candidate. Further details are again given in the papers in this special issue focusing on the various candidate models and there descriptions are available online at http://www.ngdc.noaa.gov/IAGA/vmod/candidatemodels.html. Models for epoch 2010 were submitted in October 2009; teams therefore faced the additional challenge of how to propagate their estimates of the 2009 field forward to 2010; this was not an issue faced when deriving retrospective models for epoch 2005. A brief indication of the method used to propagate to epoch 2010 is provided in the final column of Table 4. Larger

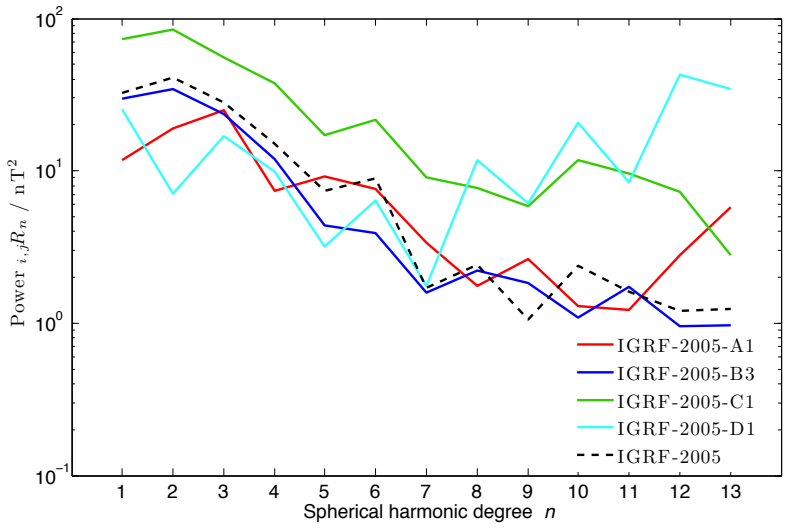

Fig. 5. Lowes-Mauersberger spectra of the vector field differences $_{i,j}R_n$ between DGRF-2005 and the candidate models considered for IGRF-10 epoch 2005 (Candidate A1 here in red was a model from DRSI/NASA/Newcastle, Candidate B3 here in blue was a model from NGDC/GFZ, Candidate C1 here in green was a model from BGS and Candidate D1 here in turquoise was a candidate from IZMIRAN. The vector field difference per degree between the final IGRF-2005 (the arithmetic mean of A1, B3 and C1) and DGRF-2005 is shown as the black dashed line.

differences in the candidate models are expected due to this additional complication and the IGRF-11 model for epoch 2010 is thus only provisional and will likely be updated to a DGRF in 2014 during the IGRF-12 process.

IGRF candidate models for main field epoch 2010				
Team	Model	Organization	Data	Comments (parent model, fwd propagation etc.)
A	IGRF-2010-A	DTU Space / IPGP / NASA-GSFC	Ørsted, CHAMP, SAC-C revised obs mon. means	Based on CHAOS-3α evaluated in 2010.0
B	IGRF-2010-B	NGDC-NOAA / GFZ	CHAMP 2006.5-2009.7	Based on POMME 6: 2nd order Taylor series SV & SA used for 2010.0 estimate
C	IGRF-2010-C2	BGS	Ørsted, CHAMP obs. hr. means for 0130 1999.0-2009.5	Revised sub: model evaluated 2009.0 MF and linear SV used to predict 2010 field.
D	IGRF-2010-D	IZMIRAN	CHAMP 2004.0-2009.2 no data selection	NOC method with extrapolation to 2010 using NOC1,2
E	IGRF-2010-E	EOST / LPGN / LATMOS / IPGP	CHAMP June/July 2009	Model at 2009.5 extrapol. to 2010 using SV models for 2009, 2010.
F	IGRF-2010-F	IPGP / EOST / / LPGN / LATMOS	CHAMP 2008.5-2009.6	2nd order Taylor series (to n=5 in quadratic) extrapolated to 2010
G	IGRF-2010-G	GFZ	CHAMP 2001-2009.6 obs. hourly means	Based on GRIMM2x MF and SV in 2009 extrapol. to 2010

Table 4. Summary of IGRF-2010 candidate models submitted for consideration in IGRF-11.

3.3.1 RMS vector field differences for IGRF-2010

Table 5 displays the RMS vector field differences $_{i,j}R$ between the IGRF-11 candidates for epoch 2010 and also between the candidates and the arithmetic mean model M and a weighted mean model M_w. M_w is considered here because it was important in the final voting process; it consists of candidates A, B, C2, F and G having weight 1.0 and candidates D, E having weight 0.25 (in addition coefficients g_1^0 and h_1^1 of candidate A were disregarded). The bottom row of Table 5 shows $_i\overline{R}$ the mean differences $_{i,j}R$ (excluding the zero value for the difference between candidates and themselves - see (6)).

As anticipated, the differences between the IGRF-2010 candidates is larger than between the DGRF-2005 candidates, with the mean of the $_{i,j}R$ between the candidates and model M being 7.3 nT here for epoch 2010 compared to 4.9 nT for epoch 2005. Candidates D and E display the largest differences from the other candidates and to the mean models M and M_w. Candidate B is most similar to M and it also agrees reasonably closely with candidates F and G (differences less than 5.5 nT) and slightly less well with candidates A and C2 (differences of less than 8.5 nT).

3.3.2 Analysis in spectral space of IGRF-2010 candidates

${}_{i,j}R$	A	B	C2	D	E	F	G	M	M_w
A	0.0	6.3	10.6	14.2	14.8	8.2	8.2	6.3	6.4
B	6.3	0.0	8.1	13.9	13.4	5.2	5.4	3.8	3.0
C2	10.6	8.1	0.0	16.9	11.8	10.0	8.9	7.1	6.8
D	14.2	13.9	16.9	0.0	19.4	15.0	14.2	12.3	13.4
E	14.8	13.4	11.8	19.4	0.0	14.0	12.4	10.9	12.0
F	8.2	5.2	10.0	15.0	14.0	0.0	6.6	5.8	5.3
G	8.2	5.4	8.9	14.2	12.4	6.6	0.0	4.6	4.4
Mean Diff	10.4	8.7	11.1	15.6	14.3	9.8	9.3	7.3	7.3

Table 5. RMS vector field differences ${}_{i,j}R$ from in units nT between IGRF-2010 candidates and also between them and the arithmetic mean of all candidates M and the weighted mean M_w (see text) . The bottom row displays the mean of the RMS vector field differences between each candidate model and all other candidate models ${}_i\overline{R}$ from (6) labelled 'Mean Diff'.

In Figure 6 (left) we plot the Lowes-Mauersberger spectra R_n from (7) of the IGRF-2010 candidates at the core-mantle boundary. Candidates E and D have noticeably higher power in degrees 11 and 13 suggesting that they may have difficulties with noise being mapped into the model coefficients at high degree.

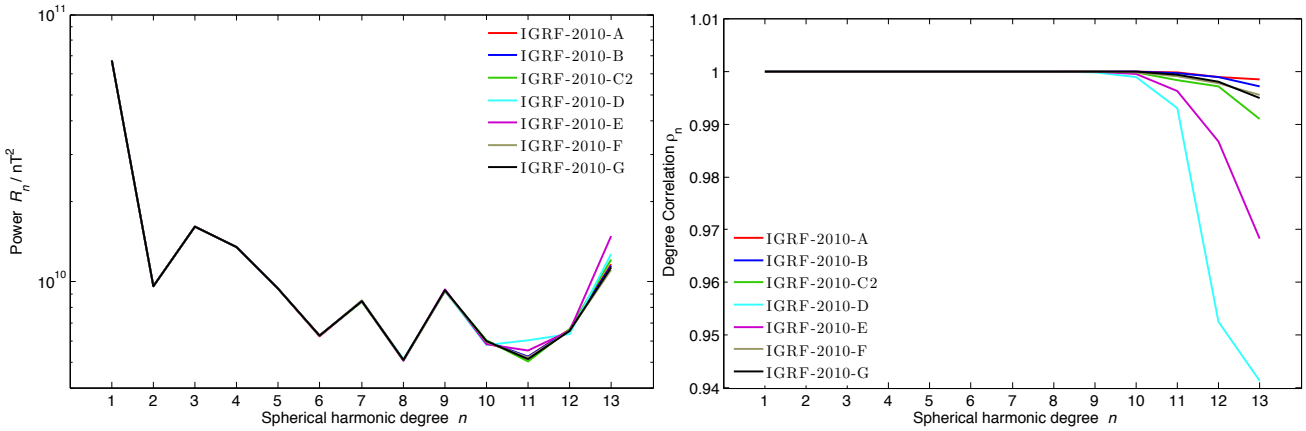

Fig. 6. Lowes-Mauersberger power spectra R_n from (7) of IGRF-2010 candidate models at radius 3480km (core-mantle boundary) (left) and degree correlation ${}_i\rho_n$ (9) of IGRF-2010 candidate models with the arithmetic mean model M (right).

Figure 6 (right) shows the degree correlation per degree ${}_i\rho_n$ from (9) between the candidates and the arithmetic mean model M. Candidates E and especially D show the largest differences above degree 10; candidates C2, F and G show smaller deviations from M while candidates A and B are closest to M.

In Figure 7 the left hand plot presents the coefficient by coefficient differences ${}_{i,j}g_n^m$ as defined in (1) between the IGRF-2010 candidates and the mean model M. It is apparent that there are some systematic problems. Candidate A possesses particularly large differences from M in coefficients g_1^0 and h_1^1. Candidate D displays many remarkable differences from M in the h_n^n sectoral harmonics while candidate E shows anomalous h_n^1 coefficients, particularly at degrees $n = 11 - 13$. Candidate C2 shows differences from M predominantly in the g_n^0 terms, most noticeably in degrees $n = 3 - 9$. The right hand plot in Figure 7 displays the Huber weights as a function of the index of the spherical harmonic coefficient. It shows how the robust weighting scheme would in this circumstance strongly down-weight many (but not all) of the coefficients of candidate D at $n > 10$, as well as many of the h_n^1 coefficients of candidate E. The lowest Huber weight for the important g_1^0 axial dipole coefficient is allocated to candidate A. Aside from this exception candidates A, B, C2, F and G receive Huber weighting factors close to 1 for the majority of their coefficients.

3.3.3 Analysis in physical space of IGRF-2010 candidates

In Figure 8 we plot at Earth's surface the differences between the Z component of the IGRF-2010 candidate models and the weighted mean model M_w in which candidates D and E are weighted by a factor 0.25 and the g_1^0 and h_1^1 coefficients of candidate A are discarded. The largest discrepancies are observed for candidates D and E. Candidate D displays major differences from M_w along the dip equator, and in the high latitude Arctic region where differences as large as 50 nT are evident. Candidate E also displays prominent deviations from M_w in the Arctic region, but of the opposite sign to those of candidate D; in addition it possesses low latitude anomalies linked to its anomalous sectoral harmonics. For both candidates E and D the deviations are globally distributed rather than localized. Candidate C2 has its largest differences from the other models in the polar regions. Candidates A, B, F, and G show more minor deviations from M_w, the differences being largest

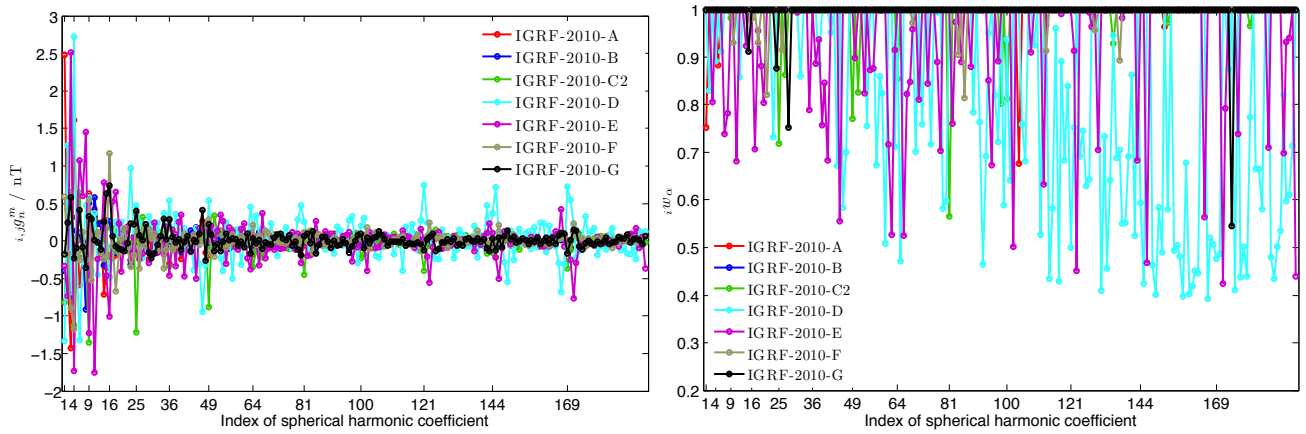

Fig. 7. Left plot shows differences ${}_{i,j}g_n^m$ as defined in (1) between IGRF-2010 candidate models and the arithmetic mean Model M as a function of the index of the spherical harmonic coefficient (running from $g_1^0, g_1^1, h_1^1, g_2^0, h_1^1$ etc indexed 1,2,3,4,5, etc). Right plot shows Huber robust weighting factors ${}_iw_\alpha$ for the candidate models for IGRF-2010 (1.0 full weight, 0.0 zero weight) also as a function of the index of the spherical harmonic coefficient.

in the polar regions in all cases. The analysis of the IGRF-2010 candidate models in physical space highlights that the most serious differences in the candidate models occur in the polar regions and to a lesser extent along the dip equator. Future efforts towards improved field models will require better models of external and induced fields in these regions.

3.3.4 Discussion and summary for IGRF-2010

The evaluations of the IGRF-2010 candidates presented above suggest that candidates D and E have some problems, particularly at spherical harmonic degree greater than 10. Consequently the task force voted to allocate these candidates weight 0.25 while candidates A, B, C2, F, G were allocated weight 1.0 in the determination of the new IGRF-11 model for epoch 2010. In addition the task force voted to disregard coefficients g_1^0 and h_1^1 from candidate A since these were thought to be suspect. Subsequent analysis has shown that a model that includes more recent data but is otherwise similar to the parent model for candidate A results in values of g_1^0 and h_1^1 that are much better agreement with model M (Olsen et al., 2010). The final IGRF-2010 was therefore fixed to be the model discussed above as M_w.

4. Evaluation of predictive SV candidate models

4.1 Retrospective analysis of IGRF-10 SV-2005-2010 candidates.

With the evaluations of the main field candidates for 2005 and 2010 complete we now move on to consider evaluations of predictive SV models. First we present a retrospective analysis of the predictive average SV-2005-2010 candidates (with central epoch 2007.5) used in IGRF-10. We treat as a reference SV model IGRF-2010 minus DGRF-2005 divided by 5 years - this provides the required coefficients in nT/yr centered on epoch 2007.5. We refer to this model in the following discussion as SV-2007.5-G11. The predictive SV from IGRF-10 (a weighted mean of IGRF-10 candidates A3, C1 and D1 with weight 1.0, and B1 and B2 with weight 0.5) is referred to in the following as SV-2007.5-G10. For further details on the IGRF-10 candidate models readers should consult Maus et al. (2005).

In Table 6 we report the RMS vector field differences ${}_{i,j}R$ between the IGRF-10 SV candidate models, their weighted mean SV-2007.5-G10, and the model derived from IGRF-11, SV-2007.5-G11.

${}_{i,j}R$	SV-2007.5-A3	SV-2007.5-B1	SV-2007.5-B2	SV-2007.5-C1	SV-2007.5-D1	SV-2007.5-G10	SV-2007.5-G11
SV-2007.5-A3	0.0	11.1	6.7	11.8	16.9	6.0	21.9
SV-2007.5-B1	11.1	0.0	12.2	17.4	19.5	11.3	21.3
SV-2007.5-B2	6.7	12.2	0.0	10.3	16.6	5.9	23.8
SV-2007.5-C1	11.8	17.4	10.3	0.0	19.1	9.4	28.4
SV-2007.5-D1	16.9	19.5	16.6	19.1	0.0	12.6	20.3
SV-2007.5-G10	6.0	11.3	5.9	9.4	12.6	0.0	21.5
SV-2007.5-G11	21.9	21.3	23.8	28.4	20.3	21.5	0.0

Table 6. RMS vector field differences ${}_{i,j}R$ in units of nT/yr between SV candidate models from IGRF-10 for epoch 2007.5, their weighted mean SV-2007.5-G10 and the mean SV between 2005 and 2010 as determined from IGRF-11, using DGRF-2005 and IGRF-2010, SV-2007.5-G11. Note the symmetry about the diagonal, again included as a check on the calculations.

Compared to the SV derived retrospectively from the IGRF-11 MF models (SV-2007.5-G11), the IGRF-10 candidate model D1 was found to perform best with an RMS difference of 20.3 nT/yr. Candidates A3 and B1 did almost as well

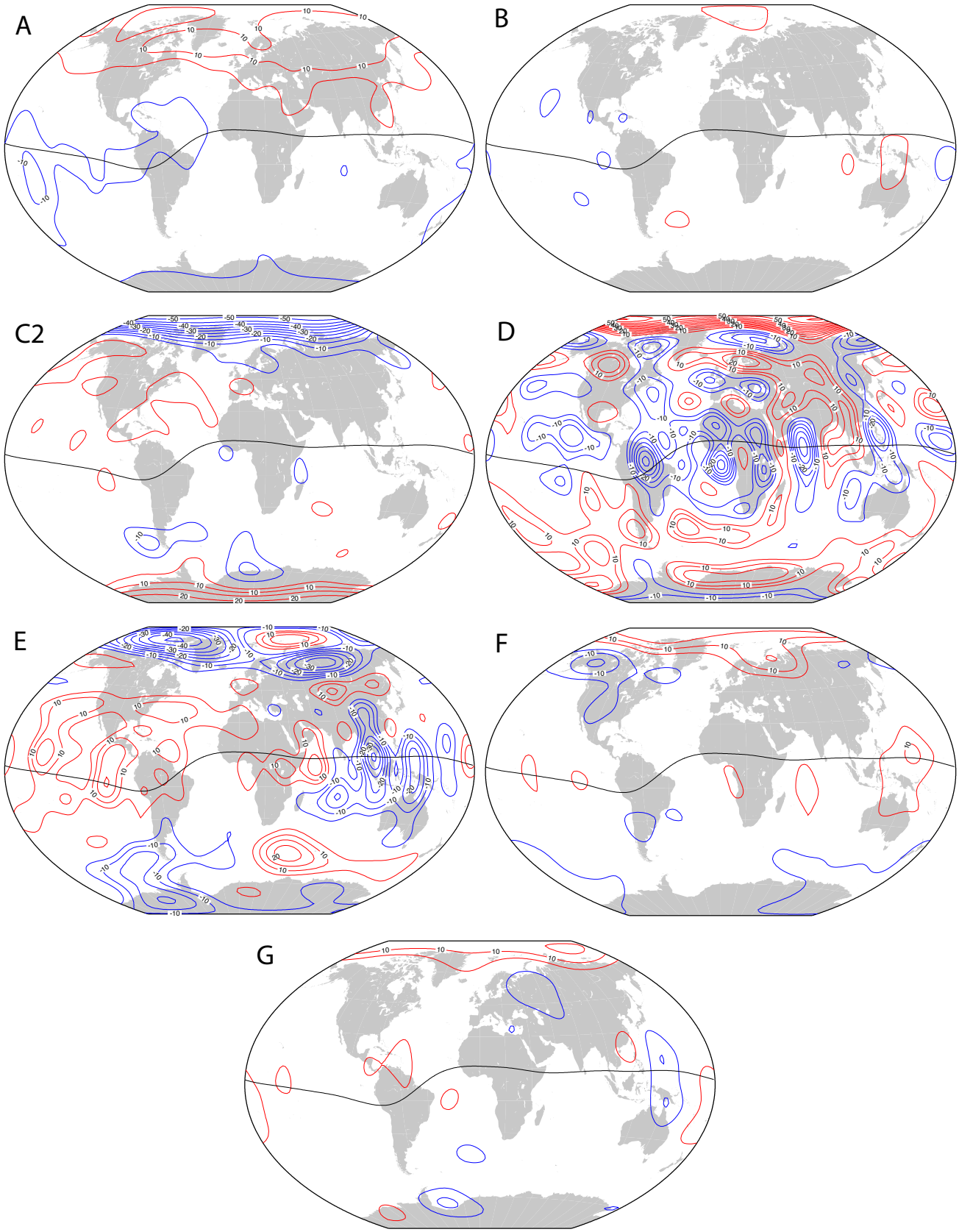

Fig. 8. Difference between the Z component of the magnetic field of IGRF-2010 candidate models and the weighted mean model M_w plotted at Earth's reference spherical surface in Winkel tripel projection projection. Contours are at intervals of 5 nT with labels every 10 nT when sufficiently large. Red: positive, Blue: negative. The dip equator is shown.

with differences of 21- 22 nT/yr. Candidate B2 performed slightly less well with a difference of almost 24 nT/yr while candidate C1 performed worst with a difference of 28.4 nT/yr. In comparison the IGRF-10 prediction differed from the IGRF-11 model by 21.5 nT/yr. Interestingly Beggan and Whaler (2010), also in this issue, demonstrate that using a steady, tangentially geostrophic, core flow they are able to derive a predictive SV model that performs slightly better than any of the candidate models for IGRF-10, with a RMS vector field difference of $\sim$ 17 nT/yr.

In Figure 9 the power spectra of the RMS vector field differences per degree between the candidates and also SV-2007.5-G10 compared to SV-2007.5-G11 are presented. Candidate C1 is found to have the largest differences at all degrees less than 6 while candidate B2 performs most poorly at degrees 7 and 8. Candidates A3, B1 and B2 involved extrapolation via quadratic terms out to 2007.5 and consequently had higher power at high degrees; the simpler linear models C1 and D1 are found in this case to perform better at high degree suggesting that extrapolation using quadratic terms was not beneficial.

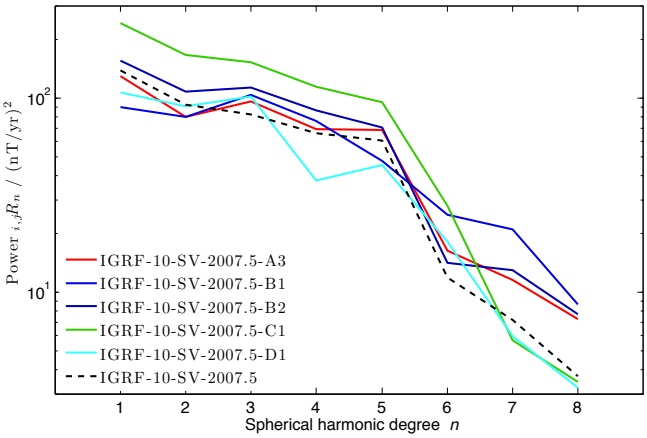

Fig. 9. Lowes-Mauersberger spectra of the vector field differences $_{i,j}R_n$ between the IGRF-11 average SV over the interval 2005 to 2010 (IGRF-2010 minus DGRF-2005 divided by 5, referred to as SV-2007.5-G11 in the text) and the SV candidate models considered for IGRF-10 epoch at 2007.5 (Candidate A1 here in red was a from DRSI/NASA/Newcastle, candidates B1, B2 here are in light and dark blue respectively were models from NGDC/GFZ, candidate C1 here in green was a model from BGS and candidate D1 here in turquoise was produced by IZMIRAN). The difference per degree between the final IGRF-10 SV for 2007.5 (a weighted mean of A1, B1, B2, C1 and D1 referred to as SV-2007.5-G10 in the text) and SV-2007-G11 is shown as the black dashed line.

4.2 Analysis of IGRF-11 SV-2010-2015 candidates

The final evaluation carried out was that of candidates submitted for the IGRF-11 average predictive SV for the interval 2010-2015. SV candidates were sought only to degree 8 although test models to higher degree were also submitted by some teams and are available for future evaluations from http://www.ngdc.noaa.gov/IAGA/vmod/candidatemodels.html. These test models should help to determine whether it is worthwhile to extend the truncation level for predictive SV in IGRF-12 (for further discussion of this point see Silva et al. (2010), this issue). Eight teams submitted SV-2010-2015 candidates, the same teams that submitted candidates for DGRF-2005 and IGRF-2010 and in addition a team, referred to as team H, led by Weijia Kuang at NASA-GSFC (Kuang et al., 2010). The latter team for the first time in the history of IGRF used assimilation of a retrospective field model (CHAOS-2s) into a geodynamo simulation based on an approximation of core dynamics in order to obtain a SV forecast for the upcoming 5 years. Details of the teams submitting SV candidates are collected for reference in Table 7.

4.2.1 RMS vector field differences for SV-2010-2015

Analysis begins as before with a compilation of RMS vector field differences $_{i,j}R$ in Table 8 between the candidates and an arithmetic mean model M and a weighted mean model M_w (the latter in this case consists of candidates B, C, D, F, H with weight 1.0 and candidates A, E and G with weight 0.5 with coefficients g_1^0 and h_1^1 of A discarded). In comparison to the earlier analyses of the MF models there is much more spread in the predictions of the candidate models with the mean of $_{i,j}R$ between candidates being 14.4 nT/yr . Candidates D and F are now the closest to the mean model M, followed by candidates H and B, then candidates A and E, with candidate G most different from M.

4.2.2 Analysis in spectral space of SV-2010-2015 candidates

Given the spread in the candidate models it is instructive to consider the Lowes-Mauersberger spectra (7) of the SV candidates at Earth's surface in Figure 10 (left), rather than at the core-mantle boundary as was done for the MF models. There appears to be no obvious way to choose between the candidates; they are widely spread at all degrees rather than form a consistent with a few anomalous outliers. Candidate G contains noticeably more power in degree 6 while candidate E has a very different spectral slope for degrees 5 to 8 (with degree 7 appearing anomalously high). The degree correlation $_i\rho_n$ from (9) between the candidate models and mean model M as shown in Figure 10 (right) illustrates that candidate E possesses a lower correlation to the mean model above degree 5 while candidate G is also noticeably different in degrees 3, 5 and 8. Candidate C2 also has marginally lower degree correlations to M than the remaining models, but it is less obviously

Predictive SV candidate models for epoch 2010-2015				
Team	Model	Organization	Data	Comments (parent model etc.)
A	SV-2010-2015-A	DTU Space / IPGP / NASA-GSFC	Ørsted, CHAMP, SAC-C revised obs monthly means	Based on CHAOS-3α SV at 2010.0
B	SV-2010-2105-B	NGDC-NOAA / GFZ	CHAMP 2006.5-2009.7	Based on POMME 6: 2nd order Taylor SV at 2009.7 used.
C	SV-2010-2015-C2	BGS	Ørsted, CHAMP and obs. hourly means	Revised sub: Av. SV 2005.0-2009.0 from parent model used
D	SV-2010-2105-D	IZMIRAN	CHAMP 2004.0-2009.25	Based on linear NOC extrapolated
E	SV-2010-2015-E2	EOST / LPGN / LATMOS / IPGP	Obs. hourly mean used to derive mon. means 1980 – 1998	Extrap. gives 1st diff of ann. means 1981-2015: SV models is av. over last 6 yrs
F	SV-2010-2015-F	IPGP / EOST / / LPGN / LATMOS	CHAMP 2008.5-2009.6	2nd order Taylor series (to $n = 5$): used slope at 2009.0.
G	SV-2010-2015-G	GFZ	CHAMP 2001-2009.6 obs. hourly means	Based on GRIMM2x: linear fit. SV 2001.0-2009.5, extrap. to 2012.5.
H	SV-2010-2015-H	NASA GSFC / UMBC / Univ. Liverpool		Geodynamo Sim: assim. from CALS7K.2, gufm1, CM4, CHAOS-2s

Table 7. Summary of SV-2010-2015 candidate models submitted to IGRF-11.

${}_{i,j}R$	A	B	C2	D	E	F	G	H	M	M_w
A	0.0	10.0	20.2	15.9	22.2	11.4	21.0	18.1	12.8	13.8
B	10.0	0.0	15.4	10.2	18.3	5.1	18.1	12.5	7.4	7.8
C2	20.2	15.4	0.0	8.0	11.0	11.4	24.2	6.5	9.7	8.6
D	15.9	10.2	8.0	0.0	12.7	6.7	18.2	4.7	4.1	3.5
E	22.2	18.3	11.0	12.7	0.0	15.3	26.3	11.6	12.9	12.6
F	11.4	5.1	11.4	6.7	15.3	0.0	18.1	9.0	4.1	4.3
G	21.0	18.1	24.2	18.2	26.3	18.1	0.0	20.7	16.9	17.8
H	18.1	12.5	6.5	4.7	11.6	9.0	20.7	0.0	6.6	5.7
Mean Diff	17.0	12.8	13.8	10.9	16.8	11.0	21.0	11.9	9.3	9.3

Table 8. RMS vector field differences ${}_{i,j}R$ in units nT/yr between SV-2010-2015 candidate models and also between these and the mean model M and the weighted mean model M_w in the columns. The final row labelled 'Mean Diff' is the mean ${}_i\overline{R}$ of the ${}_{i,j}R$ for each candidate or mean model.

different than candidates E and G.

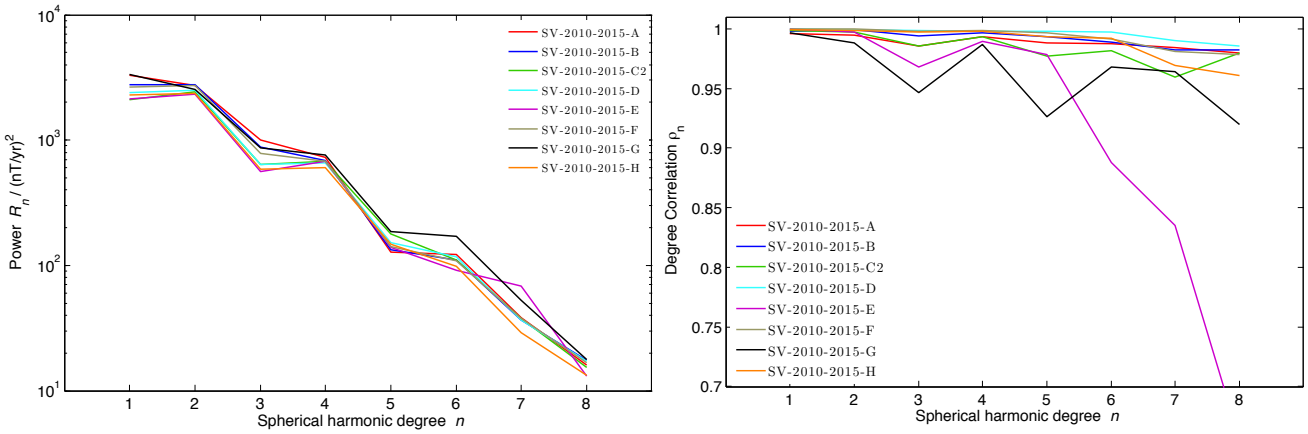

Fig. 10. Lowes-Mauersberger spectra R_n from (7) of SV-2010-2015 candidates at Earth's surface (left) and the degree correlation ${}_i\rho_n$ from (9) between the candidate models and a mean model M (right).

In Figure 11 the left hand plot presents the coefficient by coefficient differences defined in (1), between the candidate

models and model M while the right hand plot presents the Huber weights allocated by the robust weighting procedure. Candidate E is consistently allocated Huber weights as low as 0.4 for many coefficients above degree 6 while candidate G possesses some noticeably anomalous coefficients even at low degree (this is also apparent in the plot of ${}_i\rho_n$ in Figure 10).

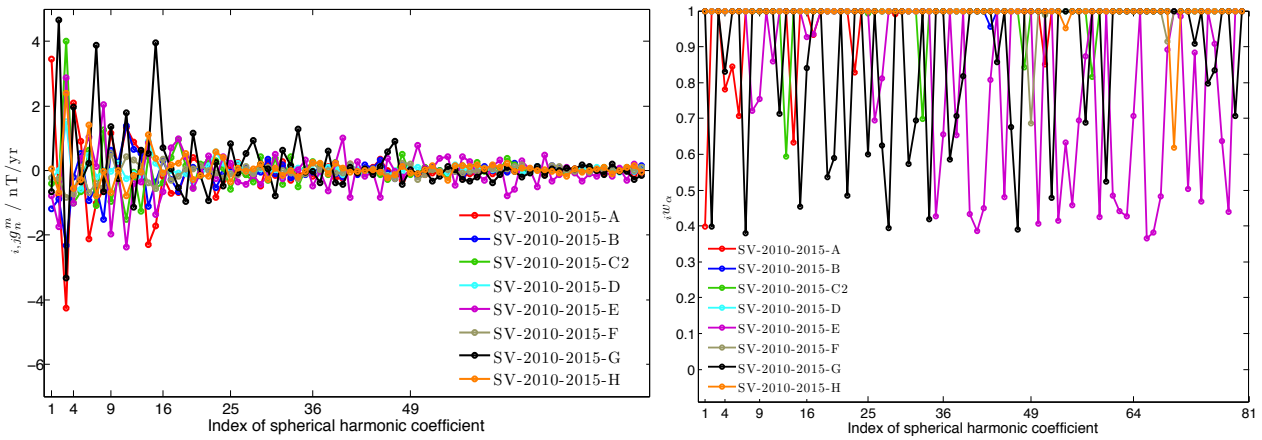

Fig. 11. Left plot shows differences ${}_{i,j}g_n^m$ defined in (1) between SV 2010-2015 candidate models and the mean model M as a function of the index of the spherical harmonic coefficient (running from $g_1^0, g_1^1, h_1^1, g_2^0, h_1^1$ etc indexed 1,2,3,4,5, etc). Right plot shows Huber robust weighting factor ${}_iw_\alpha$ (1.0 full weight, 0.0 zero weight) also as a function of the index of the spherical harmonic coefficient.

4.2.3 Analysis in physical space of SV-2010-2015 candidates

In Figure 12 the differences in the vertical (Z) component of SV at the Earth's surface between M_w and the SV-2010-2012.5 candidates is presented. The largest difference from the weighted mean is observed for candidates G and E. Candidate G predicts a large negative change in the Z to the west of South America that is not present in the other candidates; this feature has a maximum amplitude that is more than 55 nT/yr different from M_w. The majority of the differences for candidate G occur at low latitudes. Candidate E has a number of positive and negative anomalies of amplitude >20 nT/yr that are distributed over all latitudes. Candidates B, D, F and H show only minor differences from M_w. Candidate A shows global differences at Earth's surface that appear to be mainly due to a differences in its axial dipole term, but also a significant difference in its equatorial dipole contribution evident at equatorial latitudes. Candidate C2 possess some noticeable differences at low latitudes and also at northern polar latitudes.

We remark that the differences between SV candidates are often most striking at low latitudes; this becomes even more obvious when the models are analyzed at the core-mantle boundary. These differences amount to different predictions concerning the evolution (especially westward drift) of high amplitude flux features that are found at low latitudes at the core-mantle boundary and are responsible for a large amount of the present secular variation. Accurate determination of the evolution these low latitude features is crucial for accurate SV predictions - it will be of great interest in the upcoming five years to see whether any of the candidates (including H which based on an approximation of core physics) performs better in this regard than the weighted mean of the candidates M_w - it is unfortunately not currently possible to make a prior judgment on this matter.

4.2.4 Discussion and summary for SV-2010-2015

The decision on how best to weight the SV candidates to produce the final SV-2010-2015 model for IGRF-11 is much more challenging than in the MF scenarios because the candidates agree less well and because there is no well established technique for accurately predicting the future evolution of the core magnetic field. The difficulties associated with field prediction are illustrated in Figure 13. This shows in black dg_1^0/dt predictions from the GRIMM-2x time-dependent field model (constrained by CHAMP satellite and observatory data between 2001 and 2009.5) along with their formal error bars, including an extrapolation back to 2000.2 and forward to 2010.8. Also shown as a red line is a maximum entropy fit (Burg, 1967; Lacoss, 1971; Ulrych, 1972) to the GRIMM-2x model extrapolated back to 1997.25 and forward 2013.25. An assessment of the future prediction cannot yet be carried out because no observations are available. The quality of the prediction back in time can however be assessed. This is illustrated here by comparison with a 50 year field model (Lesur and Wardinski, 2009) that spans the interval 1997-2003.5 and was derived from observatory monthly means shown by the yellow stars. The 50 year model shows that $dg_1^0(t)/dt$ prior to 2001 lies close at the limits of the formal error bars on the extrapolation of GRIMM-2x, and the maximum entropy prediction is observed to be rather poor. This is a consequence of the fundamental nonlinearity of the variations in $dg_1^0(t)/dt$ due to MHD processes in the Earth's core. Looking forward to the interval 2010-2015 one can see the spread in the predictions of the SV candidate models on the right hand side of the plot at time 2012.5. The prediction of candidate A for dg_1^0/dt is clearly anomalous. For this reason the task force voted to disregard the prediction for g_1^0 (and also for h_1^1) of candidate A. Aside from this outlier the difficulty in choosing between the dg_1^0/dt predictions of the candidates is apparent.

The analyses presented earlier in this section, in both physical space and spectral space, suggested that candidate E (which

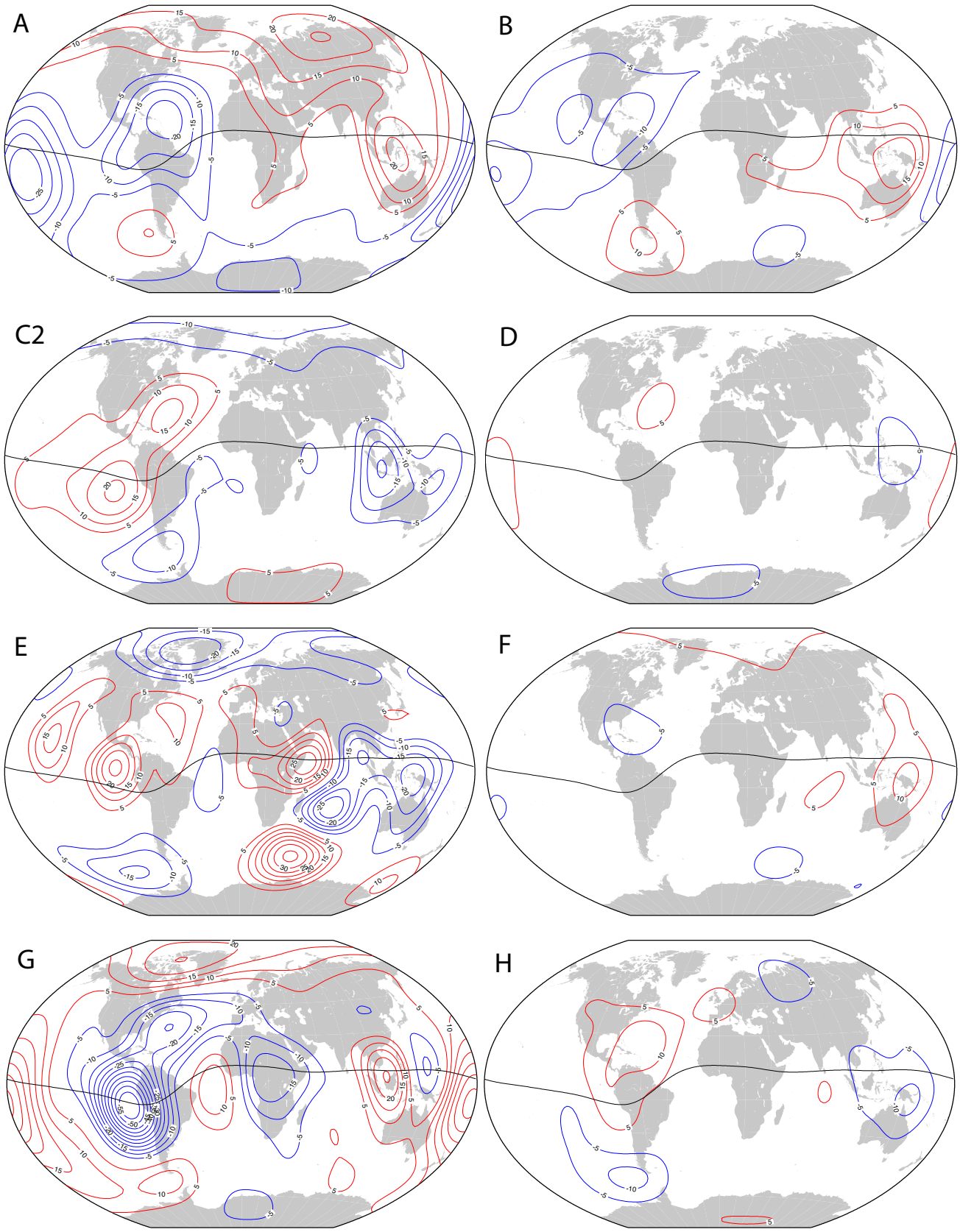

Fig. 12. Difference in the Z component of the magnetic field between SV-2010-2015 candidate models and the weighted mean model M_w plotted at Earth's reference surface in Winkel tripel projection projection. Contours are at intervals of 5nT/yr; all are labelled. Red: positive, Blue: negative. The dip equator is shown.

may have problems at degrees greater than 5), candidate G (which made predictions for the sectoral harmonic different from other candidates) and candidate A (which possessed anomalous dipole terms) were consistently different from the other candidates. The task force therefore voted to allocated weights of 0.5 to A, E and G with the remaining candidates B, C, D, F, H allocated weights of 1.0 for the construction of final IGRF-11 SV-2010-2015 model. The SV-2010-2015 model for IGRF-11 is thus identical to the the model M_w discussed above. We emphasize that in the case of SV models it is much more difficult to be certain that a particular candidate is in error simply because it differs from a mean model, because there are non-random difficulties in field forecasting, and because it is not obvious that a mean model is more likely to be correct. Further study of how best to propagate forward information from accurate MF and SV models at the current epoch is urgently needed.

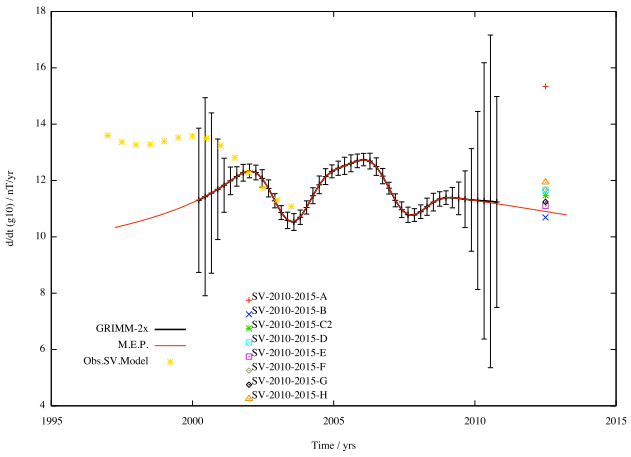

Fig. 13. Rate of change of axial dipole (in units of nT/yr) from model GRIMM-2x (black bars showing formal error estimates), along with a maximum entropy fit and extrapolation (red). A 50 year model derived from observatory monthly means (labelled Obs. SV. Model) constructed by Lesur and Wardinski (2009) is shown in yellow stars to illustrate the rate of change prior to the validity of GRIMM-2x. The symbols show the predictions of the IGRF-11 SV candidates plotted for their central epoch of 2012.5.

5. Conclusion

In this article we have described some of the statistical tests carried out by the IGRF-11 task force in order to evaluate candidate models for DGRF-2005, IGRF-2010 and SV-2010-2015. As a result of these tests, the task force voted in December 2009 that DGRF-2005 be composed of an unweighted combination of candidates A, B and G; that IGRF-2010 be composed of candidates A, B, C2, F, G with weight 1.0 and candidates D, E with weight 0.25 with g_1^0 and h_1^1 of candidate A discarded; and that SV-2010-2015 be composed from candidates B, C2, D, F, H with weight 1.0 and candidates A, E and G with weight 0.5, with dg_1^0/dt and dh_1^1/dt of candidate A discarded.

The retrospective main field models submitted for DGRF-2005 were found to largely be in good agreement. Candidates A, B and G, based on parent models from the established series of MF models CHAOS, POMME and GRIMM, were found to agree particularly well, with the formal RMS vector field error in their mean being only 1.0 nT. This close agreement is a consequence of the advances in main field modelling that have occurred in the past decade, in particular thanks to the availability of high quality satellite data from the CHAMP mission. Differences in the MF models to degree 13 are now primarily due to differences in the data selection and pre-processing strategies employed by the various teams, as well as in their choice of parameterization of external field variations. We note however that it remains possible that minor systematic errors (common to all or many candidates) could remain for example due to limitations in common techniques used to account for the external field variations. Improved knowledge of external current systems (particularly those originating) can be anticipated from ESA's multi-satellite constellation mission *Swarm* (Friis-Christensen et al., 2006) that is expected to be underway before the next IGRF revision in 2015.

Concerning the provisional IGRF model for epoch 2010, differences in how teams forward propagated their estimates from mid-2009 to 2010, depending on the nature of the time-dependence of their parent models, was an additional source of variation between the candidates. It is also clear (see Figure 13) that accurate determination of predictive SV remains the major challenge in the IGRF process; a noticeable scatter in the submitted candidate models was again present in the IGRF-11 SV candidates and it was not possible to clearly identify one group of candidates that were demonstrably of superior quality. These difficulties were further underlined by retrospective analysis of IGRF-10 SV candidates centered in 2007.5 which differed by 20-30 nT/yr from the retrospective IGRF-11 estimate for the same interval. It will be of considerable interest over the next 5 years to discover whether data assimilation methods utilizing approximations of core physics to forward propagate information (see Kuang et al., 2010; Beggan and Whaler, 2010, this issue) are yet at the stage where they can provide better forecasts than the traditional extrapolation strategies.

Acknowledgements

We thank the institutes responsible for supporting the CHAMP, Ørsted and SAC-C missions for operating the satellites and making the data available. We also thank the national institutes that support ground magnetic observatory and INTERMAGNET for promoting high standards of practice. Vincent Lesur is thanked for his help in producing Figure 13.

References

Barton, C. E., 1997. International Geomagnetic Reference Field: The Seventh Generation. J. Geomag. Geoelect. 49, 123–148.

Beggan, C. D., Whaler, K. A., 2010. Forecasting secular variation using core flows. Earth, Planets, Space, (submitted).

Burg, J., 1967. Maximum entropy spectral analysis. Proc. 37th Meet. Soc. Exploration Geophysicists, 1967; reprinted in Modern Spectrum Analysis (D. G. Childers, d.), IEEE Press, New York, 1978 94, 34–39.

Chambodut, A., Langlais, B., Menvielle, M., Thébault, E., Chulliat, A., Hulot, G., 2010. Candidate models for the IGRF - 11th generation. Earth, Planets, Space, (submitted).

Chulliat, A., Thébault, E., 2010. Testing IGRF-11 candidate models against CHAMP data and quasi - definitive observatory data. Earth, Planets, Space, (submitted).

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

Friis-Christensen, E., Lühr, H., Hulot, G., 2006. *Swarm*: A constellation to study the Earth's magnetic field. Earth, Planets, Space. 58, 351–358.

Hamilton, B., Macmillan, S., Thomson, A., 2010. The BGS magnetic field candidate models for the 11th generation IGRF. Earth, Planets, Space, (submitted).

Hogg, R. V., 1979. Statistical robustness: One view of its use in applications today. The American Statistician 33, 717–730.

Huber, P. J., 1996. Robust Statistical Procedures. SIAM.

Kuang, W., Wei, Z., Holme, R., Tangborn, A., 2010. Prediction of Geomagnetic Field with Data Assimilation: a Candidate Secular Variation Model for IGRF-11. Earth, Planets, Space, (submitted).

Lacoss, R. T., 1971. Data adaptive spectral analysis methods. Geophysics 36, 661–675.

Langel, R. A., Hinze, W. J., 1998. The Magnetic Field of the Earth's Lithosphere: The Satellite Perspective. Cambridge University Press.

Lesur, V., Wardinski, I., 2009. A 50 years core magnetic field model under frozen-flux constraints. Geophysical Research Abstracts 11, EGU2009–5352.

Lesur, V., Wardinski, I., Hamoudi, M., Rother, M., 2010. The second generation of the GFZ Reference Internal magnetic Model: GRIMM-2. Earth, Planets, Space, (submitted).

Lowes, F. J., 1966. Mean-square values on the sphere of spherical harmonic vector fields. J. Geophys. Res. 71, 2179.

Lowes, F. J., 1974. Spatial power spectrum of the main geomagnetic field. Geophys. J. R. Astron. Soc. 36, 717–730.

Lowes, F. J., 2000. An estimate of the errors of the IGRF/DGRF field 1945-2000. Earth, Planets, Space 52, 1207–1211.

Macmillan, S., Finlay, C. C., 2010. The International Geomagnetic Reference Field. IAGA Sopron Book Series, (submitted).

Maus, S., Macmillan, S., Lowes, F. J., Bondar, T., 2005. Evaluation of candidate geomagnetic field models for the 10th generation of IGRF. Earth. Planet. Space 57, 1173–1181.

Maus, S., Manoj, C., Rauberg, J., Michaelis, I., Lühr, H., 2010. NOAA/NGDC candidate models for the 11th generation International Geomagnetic Reference Field and the concurrent release of the 6th generation POMME magnetic model. Earth, Planets, Space, (submitted).

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., Mandea, M., Sabaka, T. J., Tøffner-Clausen, L., 2010. The CHAOS-3 Geomagnetic Field Model and Candidates for the 11th Generation IGRF. Earth, Planets, Space, (submitted).

Silva, L., Maus, S., Hulot, G., Thébault, E., 2010. On the possibility of extending the IGRF predictive secular variation model to a higher spherical harmonic degree. Earth, Planets, Space, (submitted).

Thébault, E., Chulliat, A., Maus, S., Hulot, G., Langlais, B., Chambodut, A., Menvielle, M., 2010. IGRF candidate models at times of rapid changes in core field acceleration. Earth, Planets, Space, (submitted).

Ulrych, T., 1972. Maximum entropy power spectrum of long period geomagnetic reversals. Nature 235, 218–219.

Winch, D. E., Ivers, D. J., Turner, J. P. R., Stening, R. J., 2004. Geomagnetism and Schmidt quasi-normalization. Geophys. J. Int. 149, 487–504.

Corresponding Author: C. C. Finlay (e-mail: cfinlay@erdw.ethz.ch)