HAL open science

Testing IGRF-11 candidate models against CHAMP data and quasi-definitive observatory data

A. Chulliat, E. Thébault

To cite this version:

A. Chulliat, E. Thébault. Testing IGRF-11 candidate models against CHAMP data and quasi-definitive observatory data. Earth Planets and Space, 2010, 62, pp.805-814. 10.5047/eps.2010.06.004 . insu-03605281

HAL Id: insu-03605281

https://insu.hal.science/insu-03605281v1

Submitted on 11 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License

Earth Planets Space, **62**, 805–814, 2010

Testing IGRF-11 candidate models against CHAMP data and quasi-definitive observatory data

Arnaud Chulliat and Erwan Thébault

Equipe de Géomagnétisme, Institut de Physique du Globe de Paris, CNRS/INSU, Université Paris Diderot, 4 place Jussieu, 75005 Paris, France

(Received March 18, 2010; Revised June 8, 2010; Accepted June 11, 2010; Online published December 31, 2010)

As part of the evaluation of IGRF-11 candidate models, we compared candidate models and actual measurements. We first carried out a residual analysis between main field candidates and CHAMP data, which were pre-processed and corrected for the secular variation and the lithospheric, external and oceanic fields. For epoch 2005.0, one model (D) is abnormally far from the testing dataset, while four models (A, B, F, G) have the smallest data residuals. For 2010.0, three models (B, F, G) have smaller data residuals than other models. These results, although biased toward models relying on datasets close to the testing datasets (B, F), usefully complement the results of intercomparisons between models. We next tested secular variation candidate models for 2010–2015 against annual differences of (a) definitive monthly means in 2007 and 2008 at 86 observatories, and (b) quasi-definitive monthly means from January to October 2009 at nine observatories where this new type of data was produced. Quasi-definitive data are found to significantly improve the discriminating effect of the test, favoring models obtained at epochs close to the end of 2009 (B, F) and penalizing some extrapolated models (G). They also enable a truly independent validation of the candidate models.

Key words: IGRF, magnetic observatories, main field, geomagnetic secular variation.

1. Introduction

The *International Geomagnetic Reference Field* (IGRF) is a spherical harmonic model of the Earth's main magnetic field, prepared by an international task force of modellers under the umbrella of the International Association of Geomagnetism and Aeronomy (IAGA). It is widely used within the geomagnetism community, but also in other research areas, for example space physics, and by the industrial sector. The latest release, IGRF-11, was derived from an arithmetic mean of three candidate models for the main field in 2005.0, seven weighted candidate models for the main field in 2010.0 and eight weighted candidate models for the secular variation over the time interval 2010–2015. Candidate models were submitted by various groups of modellers, using different methodologies and datasets. They underwent a rigorous evaluation, led by the IAGA Working Group V-MOD, after which some models were rejected or downweighted. A general report on this evaluation, including details on candidate models, is provided by Finlay *et al.* (2010) also in this issue.

Most of the tests carried out during the evaluation were intercomparisons between models, in the real and spectral spaces (see Finlay *et al.*, 2010, for a detailed report). Other tests consisted in comparing candidate models with actual measurements. This second testing method, already used for previous versions of the IGRF (Cohen *et al.*, 1997; Lowes *et al.*, 2000), was criticized and even discarded during the preparation of the 10th generation of IGRF (Maus *et al.*, 2005), because attempts to use it were found inconclusive. The reasons invoked were the limited availability of independent data, the need to remove contributions from other sources than the core prior to comparing measured data with candidate models, and the absence of measurements during the time interval of the secular variation candidates. However, intercomparisons also have drawbacks. They rely on the implicit assumption that candidate models are unbiased noisy estimates of the true field; if this assumption fails, there is a risk that a good model appears as an outlier. Thus both testing methods have strengths and weaknesses and are expected to be complementary.

Here we report on comparisons between candidate models and observatory and satellite data that were carried out as part of the evaluation process of IGRF-11. We first present comparisons between candidate main field models in 2005.0 and 2010.0 and subsets of CHAMP data at these two epochs. We then present comparisons between candidate secular variation models for 2010–2015 and recent measurements at INTERMAGNET magnetic observatories. At nine of these observatories, a new type of observatory data, quasi-definitive data, was made available from January to October 2009 at the time of the evaluation (Peltier and Chulliat, 2010), thus providing a truly independent dataset, i.e., not used in any of the candidate models. This paper aims at documenting the results of the tests and showing how comparisons between models and actual measurements, particularly quasi-definitive data, can usefully contribute to the IGRF evaluation.

Copyright © The Society of Geomagnetism and Earth, Planetary and Space Sciences (SGEPSS); The Seismological Society of Japan; The Volcanological Society of Japan; The Geodetic Society of Japan; The Japanese Society for Planetary Sciences; TERRAPUB.

doi:10.5047/eps.2010.06.004

2. Testing Candidate Main Field Models against Satellite Data

2.1 Data selection and pre-processing

Candidate main field models for epochs 2005.0 and 2010.0 were tested against vector and scalar data from the CHAMP satellite. For epoch 2005.0, Ørsted data could have been a relatively independent testing dataset, as they were used in the preparation of only three out of seven models (see Finlay *et al.*, 2010, for model descriptions). However, CHAMP and Ørsted data do not seem to be fully compatible, as a systematic difference of about 1 nT was observed between Ørsted measurements and some models derived from CHAMP data after 2003 (Maus *et al.*, 2010; Thébault *et al.*, 2010, this issue). Likewise, models based exclusively on CHAMP data are expected to be slightly biased compared to other models derived from Ørsted data. For epoch 2010.0, there is no alternative satellite dataset as Ørsted vector measurements are no longer available after 2006. Ørsted data were thus discarded for this test.

The tests for epoch 2005.0 were performed using CHAMP data from 1 July 2004 to 30 June 2005. This 12-month time interval partly averages the seasonal external field variations. For epoch 2010.0, CHAMP data were selected from June 2009 to August 2009, in an attempt to be as close as possible to the epoch of the models. More recent data were actually available at the time of the evaluation but they were not corrected for the star camera misalignment. Using these uncorrected data would have introduced large-scale residuals of about 10 nT (Nils Olsen, personal communication).

For each epoch, the same satellite data selection scheme was applied as described in Thébault *et al.* (2010) and Maus *et al.* (2010). Several corrections were then applied to the selected data. The misalignment between the magnetometer reference system and the star tracker reference was estimated and corrected. 'Noisy' tracks were identified and removed by first sorting the tracks according to their longitudes, dates and latitude intervals (mid-latitude, north polar or south polar), and then looking at the root mean square of data residuals for each track with respect to the POMME-5 model (www.geomag.us/models/pomme5.html). Data were corrected for the lithospheric magnetic field using the POMME-5 model between spherical harmonic degrees 14 and 40, and for the magnetic signals of motional induction in the oceans using the model of Kuvshinov and Olsen (2004). Finally, the magnetospheric field was removed using two models for epochs 2005.0 and 2010.0 obtained by Thébault *et al.* (2010) while calculating their candidate main field models for those epochs. These models include a static field described by a degree 2 spherical harmonic expansion, and a time-varying field described by a dipole parameterized using the D_{st} index split into its external (E_{st}) and induced (I_{st}) parts (Maus and Weidelt, 2004; Olsen *et al.*, 2005).

Data were extrapolated at model epochs using: (1) for epoch 2005.0, the secular variation and acceleration up to degree 8 obtained by Thébault *et al.* (2010); (2) for epoch 2010.0, the F secular variation candidate (option #1) or the secular variation candidate associated with each main field candidate (option #2). We believe option #1 is preferable for epoch 2010.0 because secular variation candidates were prepared using very different methodologies and we want to test main field candidates alone, not main field and secular variation candidates together. For example, some secular variation candidates (models A, D, E, G, H) are extrapolations of core field time variations over 2010–2015, while others (models B, C, F) are not (Finlay *et al.*, 2010). For this reason we will show the results for option #1 and only briefly discuss those for option #2.

It is worth noting that the testing datasets cannot be considered independent as they have intersections with datasets used in the preparation of at least two candidates, B (Maus *et al.*, 2010) and F (Thébault *et al.*, 2010). These two models also relied on similar data selection and pre-processing as described above (with some differences; for example model B used data corrected for the diamagnetic effect and models B and F were based on different time intervals). Therefore the results of the tests presented here are expected to be biased toward these two models.

2.2 Results of the tests for epoch 2005.0

Figure 1 shows the geographical distribution of scalar residuals between the seven candidate models for 2005.0 and the selected and pre-processed CHAMP scalar data. We do not show the vector residuals, as they contain more external field contributions (especially from field aligned currents on the Y component). All 2005 maps in Fig. 1 contain some small scale residuals, which correspond to unmodelled crustal field (degrees larger than 40) or external field. Except for model D, the largest residuals are located at high latitudes and all models are in relatively good agreement with the data outside these regions. These results are confirmed by Fig. 2, which shows the histogram of scalar residuals for each model (without separation in polar and non polar regions). The histogram for D is clearly skewed and centered near -5 nT. The histograms for all other models exhibit bell-shaped curves centered near zero with similar widths.

Detailed residual statistics, including vector residuals and a separation between polar and non-polar regions, are provided in Table 1. As expected, the rms residuals for models B and F are among the lowest. Model D has its residuals biased for all vector components, confirming the results shown in Fig. 2 for the scalar field only. Models C and E have larger residuals in the polar regions compared to other models (also over the Southern and Northern Pacific Ocean, respectively; see Fig. 1).

From these results, we conclude that the candidate main field models for 2005.0 may be classified in three groups:

- models having the smallest data residuals everywhere (A, B, F, and G);
- models having slightly larger scalar residuals in the polar regions (C and E);
- models having larger residuals everywhere (D).

2.3 Results of the tests for epoch 2010.0

Like for epoch 2005.0, the largest scalar residuals for 2010.0 (using option #1, see Section 2.1) are found in the polar regions (Fig. 3; note that very little data are available in the Northern polar region because we applied a drastic data selection for day times). A large-scale, dipole-

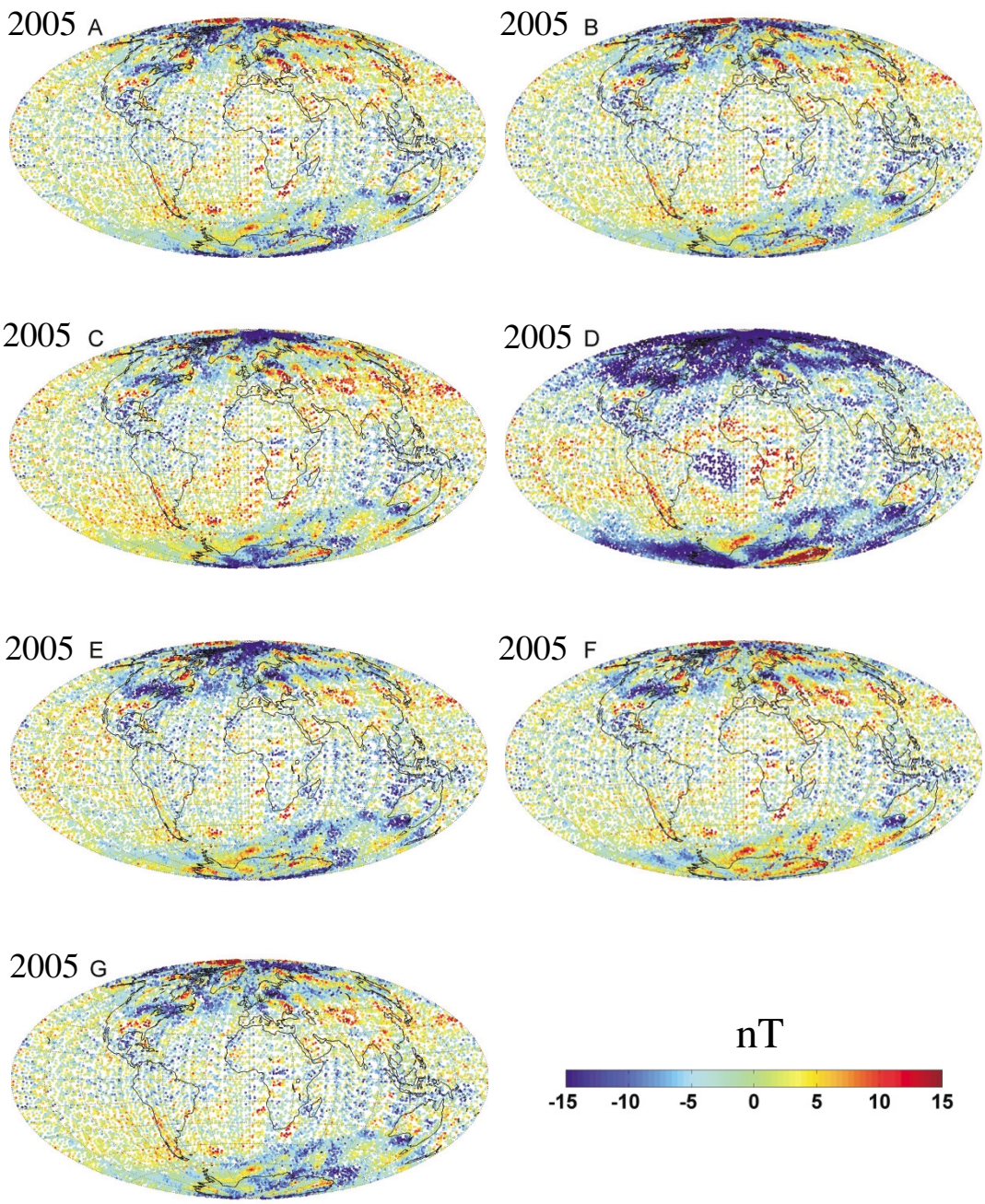

Fig. 1. Scalar residuals between the testing dataset and the seven main field candidate models for epoch 2005.0.

like structure is noticeable in the residuals of models A and D, suggesting either a contamination by some unmodelled magnetospheric field or a wrong estimate of the internal dipole field. Model C has large negative residuals over Antarctica and model E has large-scale negative residuals in the Southern Pacific Ocean not seen in other models. Models B, F and G have similar residual maps, without noticeable large scale structures except for model F in the southern Atlantic region. The histograms shown in Fig. 4 partly confirm this visual inspection. The curves for models A, B, E, F, and G are reasonably well centered on zero and roughly symmetrical, while model D is biased towards negative values. The histogram of model C has a negative skewness even though the distribution is well centered on zero.

Table 2 provides some more insights. The misfits are larger for models C and E in polar areas and larger for model D everywhere. Model A has relatively large mean residuals for the field intensity and for the B_θ component in nonpolar regions. The negative patches in residual maps of models C and E are explained by the underestimation of the radial component.

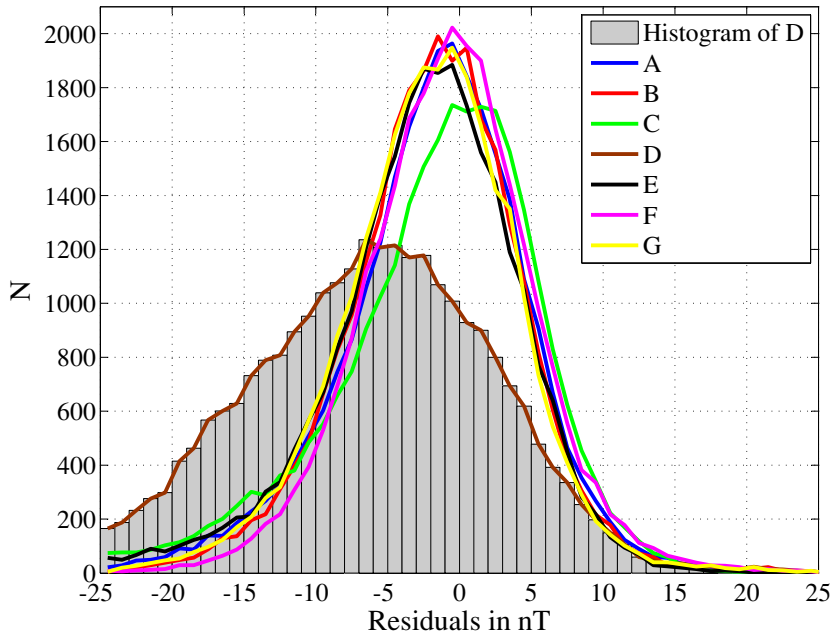

Fig. 2. Histograms of the scalar residuals between the testing dataset and the seven main field candidate models for epoch 2005.0. For all models except D, only the curve going through the top of histogram bars is shown.

Table 1. Mean and rms residuals (in nT) of the main field candidate models for epoch 2005.0 with respect to the testing dataset. Residuals are calculated for the field intensity (dF) and each component of the vector field (dB_r, dB_θ, dB_ϕ), in nonpolar ($-60°$ to $60°$ magnetic latitudes) and/or polar regions.

	A	B	C	D	E	F	G
mean dF polar	−3.40	−2.87	−4.56	−11.99	−4.06	−0.94	−3.49
rms dF polar	8.03	7.36	10.07	16.20	8.73	6.73	7.77
mean dF nonpolar	−0.05	−0.54	0.61	−2.58	−1.07	−0.61	−0.55
rms dF nonpolar	5.04	5.12	5.35	7.03	5.29	5.15	4.98
mean dB_r nonpolar	−0.43	−0.53	−0.58	−0.87	−0.75	−0.66	−0.41
rms dB_r nonpolar	5.30	5.25	5.73	7.67	5.66	5.52	5.31
mean dB_θ nonpolar	−0.24	0.05	−0.54	2.69	0.51	−0.13	0.23
rms dB_θ nonpolar	5.86	5.87	5.72	7.46	6.17	6.02	5.82
mean dB_ϕ nonpolar	0.02	0.08	0.08	0.11	0.13	0.02	0.12
rms dB_ϕ nonpolar	4.57	4.49	4.77	4.88	4.47	4.46	4.48

The same conclusions are reached if data are corrected using the secular variation candidate associated with each main field candidate (option #2, see Section 2.1), except for one model: for model A, the bias of scalar residuals towards positive values disappears. We note that the A main field and secular variation models for 2010.0 are both derived from the same parent model (Olsen *et al.*, 2010, this issue). We thus venture that this peculiarity could be explained by a possible bias in the dipole term of the secular variation model that would counterbalance the bias in the 2010 main field model.

We conclude that models B, F and G have smaller data residuals than models A, D, C and E. The case of model A is special as smaller data residuals are obtained by correcting data with the associated secular variation. (However, an independent spectral analysis showed that the dipole terms of both the main field and secular variation were anomalous; see Finlay *et al.*, 2010).

3. Testing Candidate Secular Variation Models against Observatory Data

3.1 Data

For the purpose of evaluating the candidate secular variation models, we selected 87 INTERMAGNET magnetic observatories; see Fig. 5. Other observatories were not used due to problematic data or data gaps after 2007. For all selected observatories, definitive data were available until the end of 2008.

We also used quasi-definitive data provided by the Bureau Central de Magnétisme Terrestre (BCMT, www.bcmt.fr) from January to October 2009 at nine of these observatories: AAE, BOX, CLF, KOU, LZH, MBO, PHU, PPT, TAM. Quasi-definitive data are data corrected using temporary baselines shortly after their acquisition and very near to being the final data of an observatory. For the observatories listed above, the means and standard deviations of the differences between quasi-definitive and definitive data are estimated to be less than 0.3 nT (Peltier and Chulliat, 2010). Quasi-definitive data

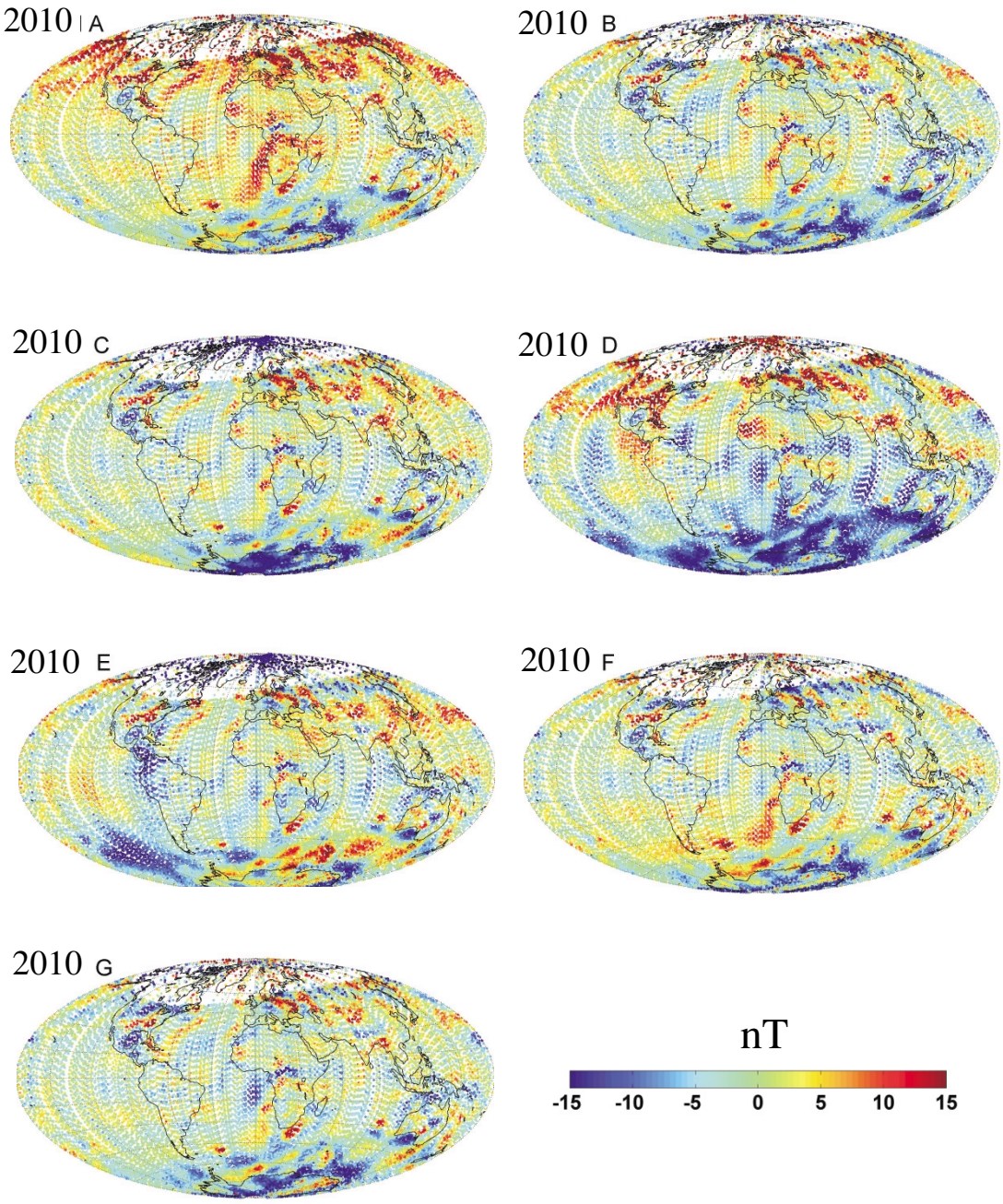

Fig. 3. Scalar residuals between the testing dataset and the seven main field candidate models for epoch 2010.0.

have recently emerged as a new observatory data product and are currently produced by only a few observatories, but others are expected to join this effort soon. These data make it possible to use observatory data for internal field studies within a few days or weeks after their acquisition, without having to wait about one year for the release of definitive data.

In all calculations, the secular variation at a given observatory was obtained by taking differences between monthly means at $t + 6$ months and $t - 6$ months, in order to remove at least part of the seasonal variations of external origin (see, e.g., Olsen and Mandea, 2007). Figure 6 shows the obtained secular variation at two observatories, MBour in Senegal (IAGA code MBO) and Kourou in French Guyana (IAGA code KOU).

3.2 Testing method

The IGRF secular variation is set to be constant over the upcoming five year time interval, i.e., the secular acceleration and higher time derivatives are set to zero over this time interval. This parameterization reflects our currently limited forecasting capability. Despite recent progress in data assimilation and other forecasting techniques applied

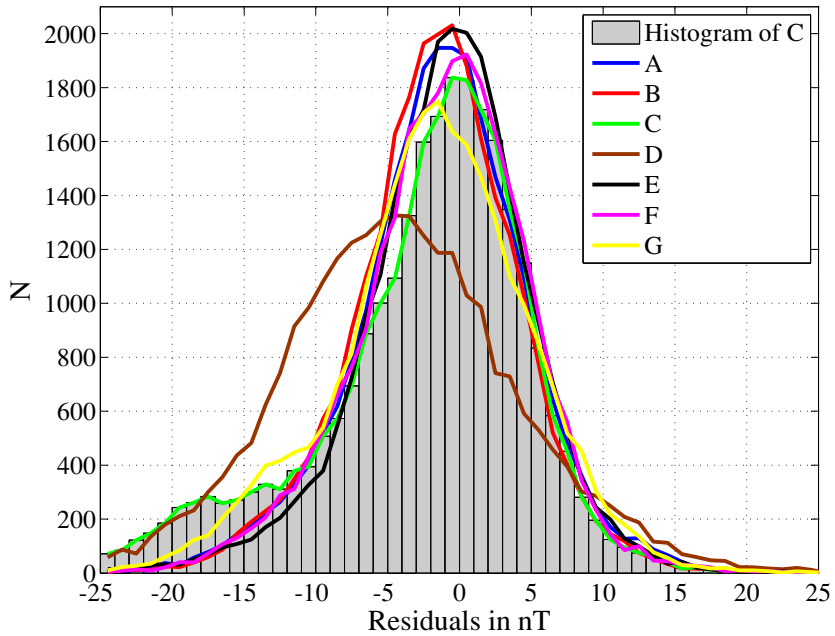

Fig. 4. Histograms of the scalar residuals between the testing dataset and the seven main field candidate models for epoch 2010.0. For all models except C, only the curve going through the top of histogram bars is shown.

Table 2. Mean and rms residuals (in nT) of the main field candidate models for epoch 2010.0 with respect to the testing dataset. Residuals are calculated for the field intensity (dF) and each component of the vector field (dB_r, dB_θ, dB_ϕ), in nonpolar ($-60°$ to $60°$ magnetic latitudes) and/or polar regions.

	A	B	C	D	E	F	G
mean dF polar	−1.72	−4.72	−7.59	−7.67	−2.24	−3.04	−4.69
rms dF polar	7.30	7.87	12.21	11.54	8.43	7.24	8.24
mean dF nonpolar	2.57	−0.58	−0.07	−1.77	−0.83	0.20	−0.56
rms dF nonpolar	6.29	5.19	5.14	7.27	5.90	5.11	5.06
mean dB_r nonpolar	−0.58	−0.86	−1.66	−0.61	−2.39	−0.69	−0.96
rms dB_r nonpolar	6.01	5.23	5.38	7.88	7.72	5.25	5.54
mean dB_θ nonpolar	−1.76	0.46	0.64	1.21	0.59	−0.15	0.45
rms dB_θ nonpolar	6.23	5.52	5.25	7.22	6.12	5.60	5.45
mean dB_ϕ nonpolar	0.02	−0.06	−0.01	−0.00	−0.08	−0.05	−0.04
rms dB_ϕ nonpolar	4.13	4.06	4.09	5.22	5.21	3.96	4.29

to geomagnetism (e.g., Fournier *et al.*, 2010), we cannot accurately forecast the secular variation over more than a few months to a year. This can be seen on Fig. 6, where two geomagnetic jerks, defined as jumps in the second derivative of the main field recorded at magnetic observatories (Courtillot *et al.*, 1978), are conspicuous in 2003 and 2007. The 2007 jerk, recently studied by Chulliat *et al.* (2010), could not have been forecasted five years ago, while preparing the IGRF-10 secular variation for 2005–2010. Similarly, we don't know if another jerk will be occurring before 2015, or if the secular variation trend will remain unchanged.

Facing this uncertainty about the future secular variation and in the absence of a specific recommandation from the IAGA Working Group, some modelers (B and F) chose to calculate a model of the secular variation as close as possible to 2010.0, others (C) to average the secular variation over a time interval prior to 2010.0, and yet others (A, D, E, G and H) to extrapolate the secular variation on or after 2010.0 (see Finlay *et al.*, 2010, for details). Among these three types of models, only the first two types were rigorously testable against data before 2010, since no data was available after 2010.0. However, it seems reasonable to expect from all candidate models, even the extrapolated ones, that they are not too far from the observed secular variation by the end of 2009. This can be tested by comparing, at each selected observatory, the secular variation predicted by each candidate model and the observed secular variation. As can be seen in Fig. 6, some candidate models are closer to the secular variation observed at KOU and MBO than others.

Two tests were performed on the eight candidate secular variation models: (1) a comparison between models and annual differences from the nine BCMT observatories having produced definitive data from November 2007 to December 2008 and quasi-definitive data from January 2009 to October 2009 at the time of the evaluation; (2) a comparison between models and annual differences from 86 observatories having produced definitive data from January 2007 to December 2008 at the time of the evaluation. (The second test could not be performed using data from all 87 selected

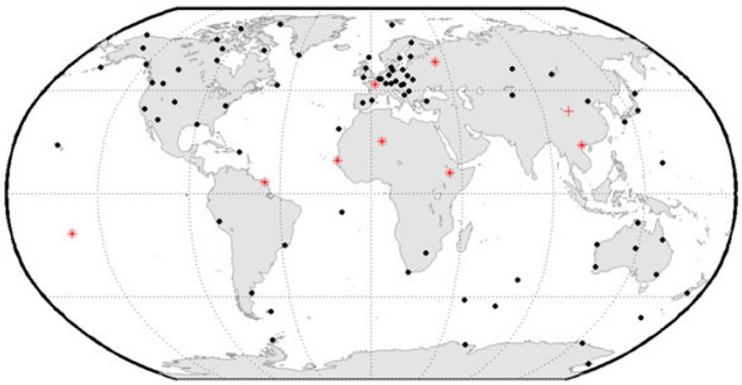

Fig. 5. Map of the 87 observatories used for test (1) only (red cross), test (2) only (black dots) or both tests (red stars): AAA, AAE, AIA, AMS, ASC, ASP, BDV, BEL, BFE, BFO, BLC, BMT, BOU, BOX, BRW, BSL, CBB, CLF, CMO, CNB, CTA, CZT, DOU, DRV, EBR, ESK, EYR, FCC, FRD, FRN, FUR, GCK, GNA, GUA, GUI, HAD, HBK, HER, HLP, HON, HRB, HRN, HUA, IQA, IRT, IZN, KAK, KDU, KNY, KOU, LER, LRM, LVV, LZH, MAB, MAW, MBO, MCQ, MEA, MMB, NAQ, NCK, NEW, NGK, NUR, NVS, PAF, PHU, PPT, PST, RES, SBA, SHU, SIT, SJG, SOD, SPT, STJ, TAM, THL, THY, TRW, TUC, UPS, VAL, VIC, VSS (see www.intermagnet.org for full names and positions).

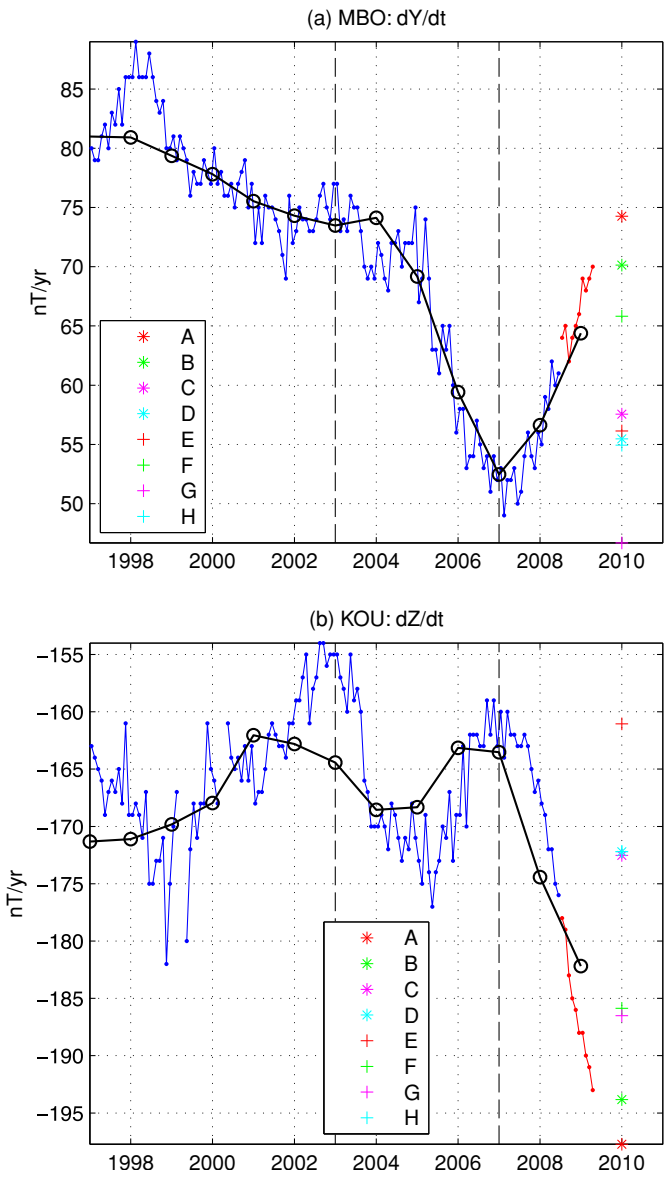

Fig. 6. Annual differences of the Y component at MBO (a) and of the Z component at KOU (b), from definitive monthly means (blue dots) and quasi-definitive monthly means (red dots). The secular variation calculated from models at these two observatories is shown as black lines and circles (CHAOS-2 model, Olsen *et al*., 2009) and colored marks (IGRF-11 candidates, see the legend).

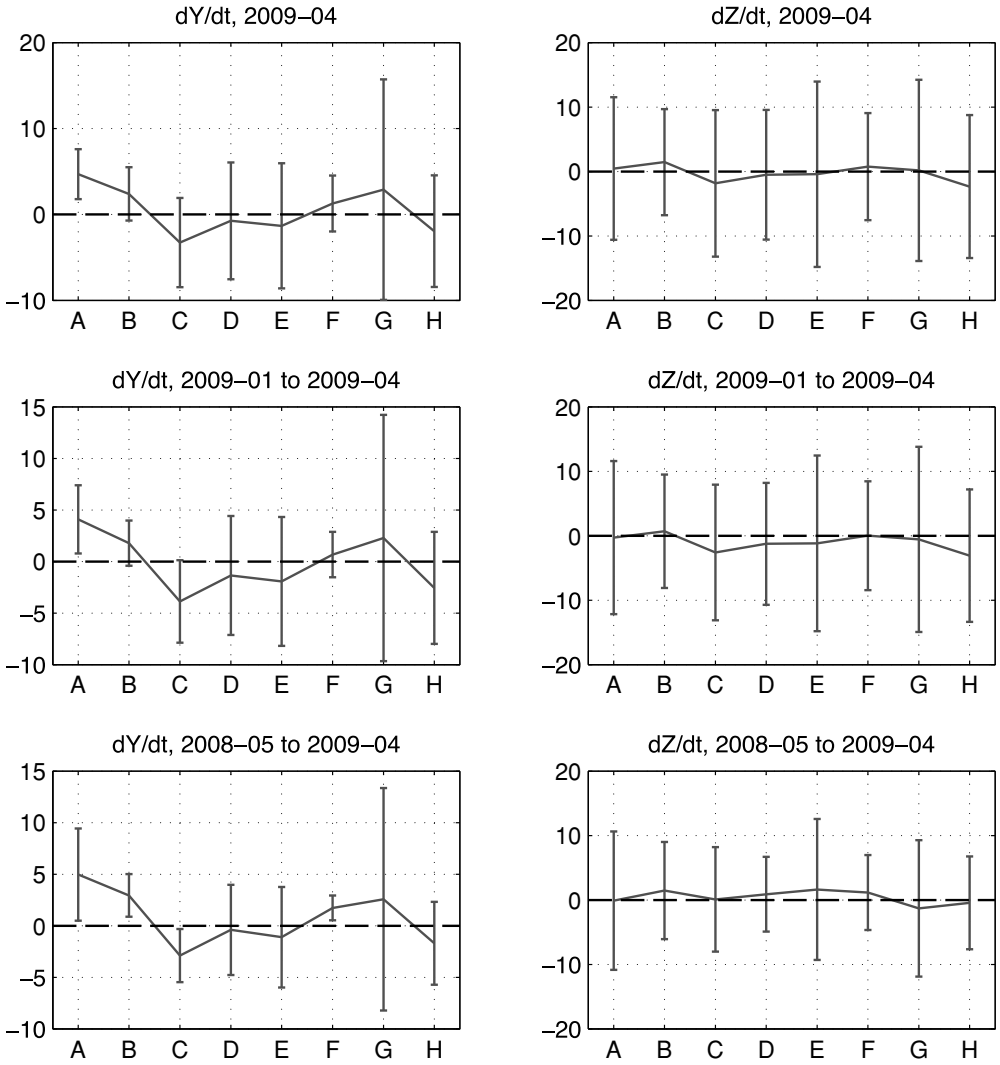

Fig. 7. Means and standard deviations (represented as error bars) of the differences (in nT/yr) between the observed and predicted secular variations at the nine selected observatories producing quasi-definitive data (Test (1)), calculated for the Y (left) and Z (right) components. At each observatory, the observed secular variation is obtained by averaging annual differences over three different time intervals: April 2009 (no averaging; top), January to April 2009 (middle), and May 2008 to April 2009 (down).

observatories as one of them, LZH, did have a data gap in the second half of 2007.)

For each test and each observatory, the following secular variation values and averages were used for the statistics: (a) the latest annual difference available (April 2009 for Test (1), June 2008 for Test (2)); (b) the mean of the latest four annual differences available (January to April 2009 for Test (1), March to June 2008 for Test (2)); (c) the mean of the latest twelve annual differences available (May 2008 to April 2009 for Test (1), July 2007 to June 2008 for Test (2)). These three sets of annual differences were used in order to assess the robustness of the results.

3.3 Results of the tests using quasi-definitive data

As can be seen in Fig. 7, the means and standard deviations of the model-data differences for the Y component are quite stable when changing the number of annual differences used (one, four or twelve, see above), with a few exceptions (for example, model A). This is also the case for the Z component when using one and four annual differences, and to a lesser extent when using the latest twelve values. This provides some confidence in the consistency of the results.

The Z component does not appear to be very discriminant and we do not further comment on it. (Even less discriminant results were obtained for X, and for this reason we do not show them here.) Four categories of models may be distinguished when comparing the results of the test for the Y component:

- B and F have the smallest mean and rms differences, whatever the number of annual differences considered.
- A and C have small rms differences for one and twelve annual differences, respectively, but large mean differences relatively to other models.
- D, E and H have intermediate mean and rms differences.
- G has a much larger rms difference (>10 nT/yr) than other models, whatever the number of annual differences considered.

These results partly reflect the various modeling strate-

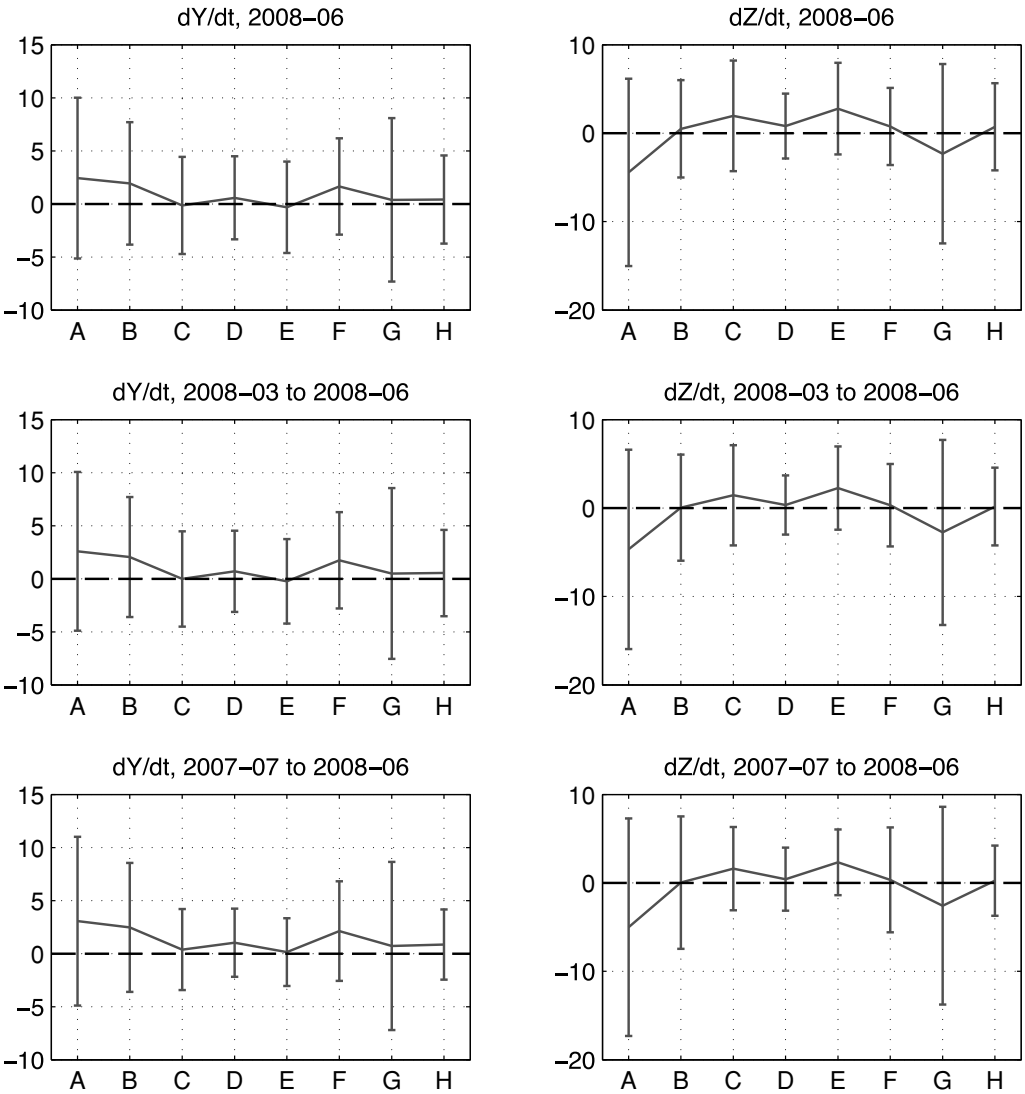

Fig. 8. Means and standard deviations (represented as error bars) of the differences (in nT/yr) between the observed and predicted secular variations at the 86 selected observatories producing definitive data (Test (2)), calculated for the Y (left) and Z (right) components. At each observatory, the observed secular variation is obtained by averaging annual differences over three different time intervals: June 2008 (no averaging; top), March to June 2008 (middle), and July 2007 to June 2008 (down).

gies used for constructing the candidate models. It is not surprising that models B and F are closer than other models to all testing datasets, as they were calculated at epochs 2009.67 and 2009.0, respectively. As expected, model A, calculated at epoch 2010.0, is closer to the last annual differences, while model C, calculated by taking the average of a parent model over 2005–2009, is closer to the average of the last twelve annual differences. Extrapolated models are more remote from the testing datasets, especially model G. The rms difference of this latter model is abnormally large compared to other extrapolated models, suggesting a more risky extrapolation scheme.

3.4 Results of the tests using definitive data

The statistics of the model-data differences for the second test are provided in Fig. 8. Again, the results are quite consistent (even more than in the first test) when going from one to four and then twelve annual differences. This time the Z component is as discriminant as the Y component and results for both components are in good agreement. The models can be categorized as follows:

- A and G have the largest rms differences (>6 nT/yr on Y, about 10 nT/yr on Z); A also has the largest average differences (>2 nT/yr on Y, 4 to 5 nT/yr on Z).
- B, C, D, E, F and H have smaller rms differences than the other two models, with some variations from one component to the other and one dataset to the other.

Model A has larger model-data differences in this test than in Test (1), which can be explained by the fact that Test (1) relied on secular variation values at epochs closer to 2010.0 (up to 2009.3 for Test (1), up to 2008.5 for Test (2)). Like in Test (1), model G is the furthest from the data. Other models are not clearly discriminated by Test (2). This could be explained by the larger number of observatories used in Test (2). However, it is worth noting that the nine observatories used in Test (1) are globally distributed, with three observatories in Africa (AAE, MBO and TAM), two in Eastern Asia (LZH and PHU), one in South America

(KOU), one in Russia (BOX), one in Europe (CLF) and one in the Pacific Ocean (PPT). This geographical distribution might even help avoiding the usual bias toward Europe of the global distribution of INTERMAGNET observatories.

4. Concluding Remarks

Despite the bias introduced by the use of a non-independent testing datasets for some models (B and F), the results of the comparison with CHAMP data are in relatively good agreement with the results of intercomparisons between models (see Finlay *et al.*, 2010). For epoch 2005.0 (resp. 2010.0), the same models A, B and G (resp. B and G) were identified as being the closest from the testing datasets and from the mean model, with the exception of model F. Similarly, the same models were found to perform less well (D for epoch 2005.0; D and E for epoch 2010.0). Differences between both evaluation methods were found for other models, hence providing a complementary view on these models and helping with the selection of weights.

When it comes to testing the predictive secular variation, comparisons between models and observatory data appear to be more discriminant than intercomparisons between models. One reason for this could be the underlying assumption that the secular variation over 2010–2015 should be correct by the end of 2009, as it cannot be reliably forecasted after 2010 anyway. This approach certainly penalized models obtained by extrapolating the secular variation after 2010.0, or averaging the secular variation over the last five years, versus those obtained at epochs close to those of the testing datasets (from July 2007 to April 2009, depending on the test). However, it brought some useful insights regarding the levels of risk associated with the various modeling strategies used to derive candidate models. Also, the use of quasi-definitive data provided an independent validation of the various candidate models, including the ones calculated without extrapolation.

Acknowledgments. The results presented in this paper rely on data collected at magnetic observatories. We thank the national institutes that support them and INTERMAGNET for promoting high standards of magnetic observatory practice (www.intermagnet.org). We also thank the institutes responsible for supporting the CHAMP satellite mission for making their data available. The research reported here was financially supported by CNES. We thank B. Langlais and H. Shimizu for their contructive reviews. This is IPGP contribution number 3028.

References

Chulliat, A., E. Thébault, and G. Hulot, Core field acceleration pulse as a common cause for the 2003 and 2007 geomagnetic jerks, *Geophys. Res. Lett.*, **37**, L07301, doi:10.1029/2009GL042019, 2010.

Cohen, Y., M. Alexandrescu, G. Hulot, and J.-L. Le Mouël, Candidate models for the 1995 revision of IGRF, a worldwide evaluation based on observatory monthly means, *J. Geomag. Geoelectr.*, **49**, 279–290, 1997.

Courtillot, V., J. Ducruix, and J.-L. Le Mouël, Sur une accélération récente de la variation séculaire du champ magnétique terrestre, *C. R. Acad. Sci. D*, **287**, 1095–1098, 1978.

Finlay, C. C., S. Maus, C. D. Beggan, M. Hamoudi, F. J. Lowes, N. Olsen, and E. Thébault, Evaluation of candidate geomagnetic field models for IGRF-11, *Earth Planets Space*, **62**, this issue, 787–804, 2010.

Fournier, A., G. Hulot, D. Jault, W. Kuang, A. Tangborn, N. Gillet, E. Canet, J. Aubert, and F. Lhuillier, An introduction to data assimilation and predictability in geomagnetism, *Space Sci. Rev.*, doi:10.1007/s11214-010-9669-4, 2010 (in press).

Kuvshinov, A. and N. Olsen, 3-D modelling of the magnetic fields due to ocean tidal flow, in *Earth Observation with CHAMP, Results from Three Years in Orbit*, edited by C. Reigber, H. Lühr, P. Schwintzer, and J. Wickert, 628 pp, Springer, Berlin, 2004.

Lowes, F. J., T. Bondar, V. P. Golovkov, B. Langlais, S. Macmillan, and M. Mandea, Evaluation of the candidate Main Field model for IGRF 2000 derived from preliminary Ørsted data, *Earth Planets Space*, **52**, 1183–1186, 2000.

Maus, S. and P. Weidelt, Separating the magnetospheric disturbance magnetic field into external and transient internal contributions using a 1D conductivity model of the Earth, *Geophys. Res. Lett.*, **31**, L12614, doi:10.1029/2004GL020232, 2004.

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

Maus, S., C. Manoj, J. Rauberg, I. Michaelis, and H. Lühr, 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*, **62**, this issue, 729–735, 2010.

Olsen, N. and M. Mandea, Investigation of a secular variation impulse using satellite data: The 2003 geomagnetic jerk, *Earth Planet. Sci. Lett.*, **255**, 94–105, 2007.

Olsen, N., T. J. Sabaka, and F. Lowes, New parameterization of external and induced fields in geomagnetic field modeling, and a candidate model for IGRF 2005, *Earth Planets Space*, **57**, 1141–1149, 2005.

Olsen, N., M. Mandea, T. J. Sabaka, and L. Tøffner-Clausen, CHAOS-2—A geomagnetic field model derived from one decade of continuous satellite data, *Geophys. J. Int.*, **179**, 1477–1487, 2009.

Olsen, N., M. Mandea, T. J. Sabaka, and L. Tøffner-Clausen, The CHAOS-3 geomagnetic field model and candidates for the 11th generation IGRF, *Earth Planets Space*, **62**, this issue, 719–727, 2010.

Peltier, A. and A. Chulliat, On the feasibility of promptly producing quasi-definitive magnetic observatory data, *Earth Planets Space*, **62**, e5–e8, 2010.

Thébault, E., A. Chulliat, S. Maus, G. Hulot, B. Langlais, A. Chambodut, and M. Menvielle, IGRF candidate models at times of rapid changes in core field acceleration, *Earth Planets Space*, **62**, this issue, 753–763, 2010.

A. Chulliat (e-mail: chulliat@ipgp.fr) and E. Thébault