

IGRF-11 COMPARISON OF METHODS

This is an attempt to present the major aspects of the techniques used in preparing the various candidate models, in a form allowing easy comparison. I thank Team Leaders for their patience with my enquiries.

F J Lowes 5 Nov 2009

CONTENTS

p1	Data, weighting, and parent model
p3	Criteria used to define magnetically quiet conditions
p4	Tine variation, weighting, and parent model
p6	Correction for magnetospheric and other fields

DATA, WEIGHTING, AND PARENT MODEL

MODEL A *Based on parent model CHAOS-3*

Satellite data: Oersted, CHAMP, and SAC-C data March 1999-Aug 2009. Selected for quiet conditions (separately for polar- and non-polar regions).

For all latitudes, use data only if satellite non-sunlit. Also, at non-polar latitudes, for CHAMP only post-midnight LT.

Vector data used only in non-polar region; scalar data used in polar region, and to fill gaps elsewhere.

Simultaneous solution for magnetometer/star-imager Euler angles (two sets for Oersted, every 10 days for CHAMP).

Sampled every 60 s; $\sin \theta$ (equal-area) weighting.

Observatory data: Revised observatory monthly means 1997-2008. (These are obtained from the observatory hourly means by first subtracting the ionospheric field (external plus induced internal) predicted by CM4, and the magnetospheric field (external plus induced internal) predicted by CHAOS-2. Then these all the corrected hourly means for a given month are used to give a robust mean mean, using Huber weights. These revised monthly means are then used to give monthly values of a 12-month running mean of SV.

Data weighting

Anisotropic weighting for Oersted vector data, and CHAMP vector data if only one star camera.

Iterative re-weighting using Huber weights

MODEL B

Satellite data: Only CHAMP data. Selected for quiet conditions (separately for overlapping polar ($>50^\circ$) and low-latitude ($<60^\circ$) regions); diamagnetic correction calculable and small; both star cameras working.

Vector data used only in mid-latitude region (together with scalar data).

'Noisy' tracks rejected.

2200-0500 LT mid-latitude region. All LT used for polar region.

Sampled every 20 s (150 km)

Observatory data: None used

Data weighting

Weighting to allow for repeated data in the polar/mid-latitude overlap

Equal-area weighting

MODEL C2

Satellite data: Oersted and CHAMP (calibration level 51) data to 2009.6, selected for quiet conditions 2230-0500 LT.

Every 60th data (about 68 s).

Outliers with $|\Delta \mathbf{B}| > 100$ nT from a-priori model rejected.

Scalar data used only when no vector data.

Satellite data individually weighted according to along-track standard deviation of satellite data and activity at nearby-observatories, resulting in down-weighting of vector data at high latitudes, particularly in auroral zones.

Observatory data: 1999.0-2009.5 observatory hourly means, 0100-0200 LT, with manual correction for bad data. Jumps identified manually, and magnitude co-estimated along with observatory biases in parent model.

Satellite Data weighting

Weight used a combination of local noise (60 samples along track) and larger-scale noise from nearest observatory.

Weighted for equal spatial coverage using 1-degree equal-area tessera.

Observatory Data Weighting ?

MODEL D

Satellite data: Purely CHAMP vector data, Jan 2004 to Apr 2009, used for MF models. ALL data (at one second spacing) used (no selection).

Observatory data None used

Data weighting

All data given same weight.

No equal area weighting.

MODEL E

Three separate data sets, one for each model:

(i) Oersted & CHAMP vector and scalar data, selected for quiet conditions. About 12 month centred on 2005.0

Used only scalar data in polar regions (50°)

2200-0600 LT

Every tenth measurement.

Also decimated to maximum of one each month in $1^\circ \times 1^\circ$ equiangular bin; then $\sin \theta$ (equal-area) weighting.

Anisotropic weighting for Oersted vector data.

(ii) CHAMP vector and scalar data, selected for quiet conditions. About 1 month centred on 2009.485

Used only scalar data in polar regions (50°)

2200-0600 LT

Every tenth measurement.

Also decimated to maximum number in $3^\circ \times 3^\circ$ equiangular bin; then $\sin \theta$ (equal-area) weighting.

(iii) Observatory data – see under "Time Variation"

Data weighting

DGRF 2005 Weights corresponding to
CHAMP scalar/vector component 2 nT,
Oersted scalar 3 nt

Oersted vector component 3 nT, together with anisotropic weighting

IGRF 2010 (CHAMP) Scalar and vector component measurements given equal weight.

MODEL F

Satellite data: (Data set and selection for quiet conditions the same as used for Model B. It includes the correction for magnetometer Euler angles.)

Only CHAMP data. Selected for quiet conditions (separately for overlapping polar ($>50^\circ$) and low-latitude ($<60^\circ$) regions); diamagnetic correction calculable and small; both star cameras working.

'Noisy' tracks rejected.

2200-0500 LT for non-polar region. All LT used for polar region.

Sampled every 20 s (150 km)

Used about 480 days centred on 2005.0, and about 480 days centred on 2009.0.

Decimated with respect to a quasi-regular grid.

Observatory data: None

Data weighting

Weighting to allow for repeated data in the polar/mid-latitude overlap

Equal-area weighting

MODEL G Based on parent model GRIMM-2x

Satellite data: Only CHAMP data (With Euler angles known)

Selected for quiet days

Decimated to 20 s.

For non-polar (55°) regions 2300-0500 LT, and only the X and Y components in the SM co-ordinate system are used.

For polar regions no LT selection, and the full vector is used.

Observatory data: Observatory hourly means. For all observatories, the same selection criteria are used as for non-polar satellite data. In the polar regions the full vector is used, but in the non-polar region only the X and Y components in the SM co-ordinate system are used.

Data weighting

Weight depends on polar/non-polar, field component, satellite/observatory

Also equal-area weighting

MODEL H (*Field input is from CHAOS model*)

No data as such. One particular model of core dynamics (involving various physical assumptions and approximations about the nature of fluid motion in the core) is intermittently constrained by imposing a particular external magnetic field. Forward integration is then used to predict how the field produced will change with time. For recent time this constraint was done yearly up to 1990, using the CHAOS model.

CRITERIA USED TO DEFINE MAGNETICALLY QUIET CONDITIONS

MODEL A

General data selection criteria:

Max D_{st} derivative 2 nT/hr

Sun 10° below horizon

Non-polar region (60°):

$K_p \leq 2_0$

CHAMP data are only used from local time past midnight

Polar regions specific also:

merging electric field at the magnetopause < 0.8 mV/m

MODEL B

General data selection criteria:

$|Dst| < 30$ nT

Diamagnetic effect < 5.0 nT

Maximum jump in diamagnetic effect: 2.0 nT

Low-latitude ($< 60^\circ$):

$2200 < \text{local time} < 0500$ LT

$|dDst/dt| < 2$ nT/h

Max am: 12

Max am 3 hours before: 15

Polar regions ($> 50^\circ$):

(All local time)

$|dDst/dt| < 5$ nT/h

Max IMF-By: ± 8 nT

Min IMF-Bz: -2 nT

Max IMF-Bz: 6 nT

Max merging electric field at the magnetopause: 0.8 mV/m

Max am: 27

Max am 3 hours before: 27

MODEL C

Satellite data

K_p and K_p for previous 3 hours ≤ 2 -

$|dDst/dt| \leq 5$ nT/hr

IE ≤ 30 nT; PC ≤ 0.2 mV/m

$0 \leq \text{IMF } B_z \leq +6$ nT; $-3 \leq \text{IMF } B_y \leq +3$ nT; $-10 \leq \text{IMF}$

$22:30 \leq \text{local time (hour:min)} \leq 05:00$,

$|\text{observed magnetic field value} - \text{value from } a \text{ priori model}| \leq 100$ nT;

$|\text{scalar } F \text{ from OVH} - \text{vector } F \text{ from CSC}| \leq 2$ nT

Observatory hourly means

$K_p \leq 2+$, $|dDst/dt| \leq 5$ nT/hour

IMF $B_z \geq 0$ nT

$01:00 \leq \text{local time} \leq 02:00$ LT **and** darkness test at 110 km altitude above observatory

MODEL D

NO selection for quiet conditions

MODEL E**DGRF 2005**

22:00 < local time < 06:00

Magnetic sectorial index alpha 30-minutes: $0 < \alpha_{30} < 4$ nT

$|Dst(t)| < 30$ nT $|dDst(t)/dt| < 10$ nT/h

IGRF 2010

$|Dst(t)| < 5$ nT $|dDst(t)/dt| < 3$ nT/h

$Kp(t) < 1+$ $Kp(t+/-3h) < 2-$

MODEL F (Same as for Model B)**General data selection criteria:**

$|Dst| < 30$ nT

Diamagnetic effect < 5.0 nT Maximum jump in diamagnetic effect: 2.0 nT

Low latitude:

2200 < local time < 0500 LT

$|dDst/dt| < 2$ nT/h

Max am: 12 Max am 3 hours before: 15

Polar regions:

(All local time)

$|dDst/dt| < 5$ nT/h

Max IMF-By: ± 8 nT Min IMF-Bz: -2 nT Max IMF-Bz: 6 nT

Max merging electric field at the magnetopause: 0.8 mV/m

Max am: 27 Max am 3 hours before: 27

MODEL G**General data selection criteria:**

IMF-B_z positive

Vector Magnetic Disturbance: VMD < 20 nT $|dVMD/dt| < 100$ nT/day

Non-polar region:

2300 < local time, 0500, sun below horizon at 100 km altitude.

Star cameras flag: 3 (two cameras)

Polar region:

(All local times)

Star cameras flag: 3, 2, 1 (in order of preference)

Observatory data: Observatory hourly means selected as for non-polar CHAMP data.

TIME VARIATION TRUNCATION, AND REGULARISATION**MODEL A** *Parent model to n=60. Spline time variation to n=21*

Order 6 B-splines, with 6-month knots.

Values for 2005, 2010, taken from the spline (extrapolation).

SV 2010-15 taken to be the slope at 2010.

Constrains $|\partial^3 B / \partial t^3|^2$ averaged over CMB and in time. Also $|\partial^2 B / \partial t^2|^2$ averaged over CMB at the end-points 2007.0, 2010.0.

MODEL B *Two separate models. Parent models to n=40. SV and SA to n=20*

Data from 2003.5 to 2006.5 to give 2005 value.

Data from 2006.5 to 2009.67, to give 2010 value by extrapolation.

SV 2010-15 is taken to be the slope at 2009.67.

SV regularised from n=14, and SA regularised from n=9

MODEL C *Parent model to n=60. (Piecewise continuous) SV to n=13.*

Coefficients expressed as piecewise continuous straight lines of length 400 days, giving values at 2005.0 and 2009.0. Extrapolation to 2010 by extrapolating the straight line (2005.0 to 2009.0) to 2010.0.

The IGRF SV is the average of the four 400-day SVs.

{FJL *This is almost the same as the slope of the line used for extrapolation.*}

No regularisation, but very small eigenvalues rejected

MODEL D *Two separate models. Parent models (including NOC time variation) to n=13*

SHA done for 1 day of data, every 4 days.

For each coefficient, after subtracting time average performed NOC analysis in time. NOC0 is the value at the midpoint, NOC1 (which is almost linear) is identified as SV; NOC2 as SA (but numerically very small, so ignored), NOC3 as Dst variation, and NOC4 as annual variation.

2005 model: Used 24 months of data centred on 2005.0, so model is NOC0 of that analysis.

2010 model: (NOC0+NOC1) of the 2004-2009 model., extrapolated to 2010

SV 2010-15 model: is taken to be the NOC1 (straight-line) part of the 2004-2009 model.

{FJL I suspect NOC0 will correspond to an *average* value of Dst, rather than a *small* value of Dst. Also, depending on exactly how t=0 is defined, NOC0 could include some of the annual variation.}

MODEL E

Three separate models.

(i) *Parent model to n=15. SV to n=8*

DGRF 2005: linear in time model from Oersted and CHAMP data for about 12-months centred on 2005.0.

(ii) *Parent model to n=14.*

IGRF 2010 model – no time variation - from CHAMP data for about one month centred on 2009.485.

Extrapolated (for terms up to n=8) to 2010.0 using the appropriate section of their SV model.

(iii) *Parent SV model to n=8*

SV model 2010-15 from 96 sets of observatory monthly means (calculated from hourly means, data, manually checked, and with some linear interpolation to fill gaps), from (at least) 1997.0 to 2008.0. For each observatory series extrapolated to 2016.0 "using an exponential smoothing scheme, with an additional 12-month periodic term". Then used successive difference of annual means to give annual estimates of SV. The 2010-15 SV model is the mean of the SVs from 2009.5 to 2015.5. {FJL This is the same as one sixth of the difference between their MF models of 2009.5 and 2015.5.}

No regularisation

MODEL F

Two separate models, each to n=16, SV to n=10, SA to n=5.

DGRF 2005.0. For the 480-day period centred on 2005.0,

(Quadratic terms probably not needed, but had no adverse effect.)

IGRF 2010.0 For the 480-day period centred on 2009.0,

then quadratic extrapolation to 2010.0.

SV 2010-15 The linear part of the 2009.0 model.

(They suspect there might have been a jerk between 2007.5 and 2008.2, depending on the observatory.)

Probably no regularisation

MODEL G *Parent model to n=60. Spline time-variation to n=14*

Parent model (GRIMM-2x) uses order 6 B-splines with knots at (about) 400 day spacing.

DGRF 2005 obtained by averaging the parent model for one year around 2005.0.

{FJL comment: Think of the quadratic $y=a+bt+ct^2+dt^3$, with $t=0$ at 2005.0. Then the actual field at $t=0$ is $y=a$, but the average from $t=-T/2$ to $t=+T/2$ is $y=a+cT^2/12$. So in general there will be a small, but systematic, difference.}

IGRF 2010 obtained by extrapolating the parent model for 2009.0 to 2010.0, using dB/dt of the parent model at 2009.0.

SV 2010-15 Plot SV of parent model against time, make a straight-line fit from 2001.0 to 2009.5, and extrapolate the line to 2012.5.

(They note that SA of h_1^1 increased rapidly after about 2008)

Probably no regularisation

MODEL H

The time variation is not constrained to a particular algebraic form.

The calculation predicts the ratio g_n^m/g_1^0 . So g_1^0 for 2010 is obtained from CHAOS, and its time variation to 2015 assumed to be linear, with the same gradient as from 2001 to 2009.

No regularisation as such, but there might be some arising from the assumed physics.

CORRECTION FOR MAGNETOSPHERIC AND OTHER FIELDS

MODEL A

Magnetospheric and ring-current field modelled and subtracted using a quiet-time model (based on GSM and SM co-ordinates) derived separately, and including internal induced field. Also a SM $n=1$ slow time variation estimated for each 12-hours.

Observatory data had a model ionospheric field subtracted.

MODEL B

Magnetospheric and ring-current field modelled and subtracted using an 18-parameter quiet-time model (based on GSM and SM co-ordinates) derived separately, and including internal induced field).

Also a SM $n=1$ field, strength estimated each 24 hrs.

Ocean tide effects estimated and removed.

Also, ionospheric "diamagnetic" correction made

(Data set probably had a model crustal field removed)

MODEL C2

Co-estimate degree 1 external field. Time dependence using periodic functions for the annual and semi-annual signals. Also a dependence on the 20-minute Vector Magnetic Disturbance index. Signals induced by these external fields are also parameterised.

MODEL D

Except for annual variation, no explicit allowance for external field sources. They argue that their NOC time-series approach adequately separates the more rapidly varying external fields from the (purely) internal NOC_{0,1,2}.

MODEL E

For 2005 co-estimate constant degree 1,2 external field, and degree 1 external field having Dst dependence.

For 2010 co-estimate constant degree 1,2 external field.

For SV co-estimate constant degree 1 external field each year.

MODEL F

Co-estimate degree 1,2 external/induced-internal field up to $n=2$ for Ist, Est time dependency

Co-estimate degree 1 external/induced-internal pseudo-static field

Ocean tide effects estimated and removed.

(NO "diamagnetic" correction)

(Data set probably had a model crustal field removed)

MODEL G

The $n=1$ magnetospheric field is co-estimated, as a piece-wise continuous linear fit with a node every 3 months (corresponding induced field assumed zero). For mid and low latitude data only, a dependence on the VMD index, and its corresponding induced internal field.

In the polar region, the field of the ionospheric and field-aligned currents is co-estimated, using localised functions.

MODEL H

(Not relevant)

NUMERICAL COMPARISON OF IGRF-11 CANDIDATE MODELS

F.J. Lowes

5 Nov 2009

1. INTRODUCTION

All my numerical comparisons are specified in terms of the mean-square vector (difference in) field averaged over the reference sphere, i.e. in terms of $\sum_{n,m} (n+1)(\Delta\text{coeff})^2$,

where $\Delta\text{coeff} = (\text{coeff} - \text{average coeff})$,

using a variety of "average coeff".

(To possibly help comparison with other assessments, I also give the corresponding rms value. To avoid possible confusion, in the text I will spell out "mean-square" in full, but use the abbreviation "rms".)

I have chosen this presentation, not only because it gives the total (root) mean-square vector field difference, but also because for the SHA of vector data, of uniform variance, and uniformly spaced over the sphere, the variance of the resulting degree-n coefficients will be proportional to $1/(n+1)$. So the combination above would then give figures which will be scattered, but with no significant variation with degree n. This is what I find for "noisy" models (those most deviating from the average. But for "good" models there is a distinct trend for the values to increase with decreasing n; there must be some other factor increasing the scatter towards lower degree terms (for example magnetospheric and ionospheric noise).

At BGS, Beggan has used a somewhat different approach; for a given model, say model i, he takes Δcoeff to be the difference between that c_i for model i and that c_j for (the different) model j, and averages over all the possible differences ($c_j - c_i$) for j not equal to i. Then (for simplicity ignoring the sum over n,m) if the errors in the coefficients are random, and independent between models, and have standard deviations σ_i , it can be shown that if $i=1,2,3 \dots N$, we get for the respective parameters

$$\text{Lowes} \quad L = (c_i - m)^2 = \quad [(N-2)/N]\sigma_i^2 + [1/N^2]\sum_{j=1}^N \sigma_j^2$$

and

$$\text{Beggan} \quad B = \sum_{j \text{ not } i} (c_i - c_j)^2 / (N-1) = [(N-2)/(N-1)]\sigma_i^2 + [1/(N-1)]\sum_{j=1}^N \sigma_j^2$$

In both cases this is a biased estimate of some known fraction of σ_i^2 , the bias being the same for all i. The difference is that in the Beggan approach the bias is roughly constant at about the average variance), while in the Lowes approach the bias is the average variance, divided by N. In both approaches the real variation of the variances between models is diluted by the bias, but the dilution is much more in the Beggan approach.

2. DGRF 2005

There are 7 candidate models; A, B, C2 (the revised BGS model), D, E (the revised EOST model), F, and G.

Call the (unweighted) arithmetic mean of all 7 models mean7. The resulting rms differences from mean7 are given in the second row of Table 1a, and the corresponding mean-square differences in Table 1b.

TABLE 1a rms vector difference (nT) between models and various means

Model	A	B	C2	D	E2	F	G	Models used in mean
rms from mean7	3.1	2.6	4.6	12.3	4.5	4.1	3.0	A+B+C+D+E+F+G
rms from mean6	2.0	1.7	4.1	<i>14.4</i>	4.2	3.4	2.3	A+B+C +E+F+G
rms from mean3g	1.6	1.1	<i>4.6</i>	<i>14.5</i>	<i>5.3</i>	<i>4.0</i>	1.5	A+B +G
rms from mean3p	<i>3.2</i>	<i>2.9</i>	4.0	<i>14.5</i>	3.5	3.4	<i>3.6</i>	C+E+F

TABLE 1b mean-square vector difference (nT)² between models and various means

Model	A	B	C2	D	E2	F	G	Models used in mean
ms from mean7	9.8	6.9	21.4	152.	20.4	16.4	8.8	A+B+C+D+E+F+G
ms from mean6	4.1	2.8	16.4	<i>208</i>	17.9	11.6	5.5	A+B+C +E+F+G
ms from mean3g	2.5	1.3	<i>21.0</i>	<i>210</i>	<i>27.6</i>	<i>16.0</i>	2.4	A+B +G
ms from mean3p	<i>9.9</i>	<i>8.4</i>	16.0	<i>209</i>	12.4	11.2	<i>12.8</i>	C+E+F

Note: In the Tables, figures in italics are for models NOT used in that mean.

There is a range of a factor of 17 between the largest and smallest mean-square values. Model D has by far the largest value, about 7 times larger than the next largest. Many of its individual coefficient ms differences are up to 50 times larger than for the other models; looking at the 195 coefficients, in 152 cases it has the largest deviation from the mean. I am sure that something has gone wrong, and that it should be completely omitted from further consideration. (The D model for 2005 was derived independently from the 2010 and SV models.)

Omitting model D from the average gives mean6, and the resulting rms/ms differences are given on the next line of the tables. The deviations from the mean have reduced significantly for all the six included models; this suggests that D is systematically different from the others, and confirms that including D will almost certainly make the DGRF poorer.

There are now two groups of models (see Table 3), a 'good' group (A,B,G) having mean-square deviations from the mean in the range 2.8 to 5.5 and a 'poor' group (C,E,F), having values 11.6 to 17.9. I have tried comparing the models with mean3g, the mean of (A,B,G), and with mean3p, the mean of (C,E,F). For mean3g, the values for the contributing models are all reduced significantly compared to their mean6 values (the mean squares are reduced by a factor of two), while the values for the others are increased, but by much smaller factors. Comparing the mean3p values with the mean6 values, all of C,E,F are reduced, (but only the value for F significantly), and those for A,B,G are increased by a factor of 2 to 3. (This suggests a small systematic difference between the 3g and 4p groups of models, though the actual differences between the means is consistent with the errors being purely random.) In fact, given only the values of this line of the table, it would be very difficult to decide which three models had been used to obtain the mean!

At first sight, the fact that for both mean3g and mean3p the respective values either all increase or all decrease suggests that there could be a small systematic difference between the two groups, but it is shown below that the differences between the means are consistent with purely random errors.

The differences between the various means are given in Table 2., in which a blank line indicates a change in the way the models are divided into groups.

TABLE 2a rms vector differences (nT) between different means

	0		mean6	mean3g	mean3p
Mean7	43907.3		2.1	2.5	2.5
mean6	43906.7			1.4	1.4
mean3g	43906.8				2.9
mean3p	43906.5				

(The second column gives the total rms field of that model, i.e. the rms intensity of that model. The other columns give the rms intensity of the vector difference between the respective models.)

TABLE 2b mean-square vector differences (nT)² between different means

			mean6	mean3g	mean3p
Mean7			4.2	6.4	6.2
mean6				2.1	2.1
mean3g					8.3
mean3p					

Dropping model D changes the mean by 2.1 nT rms, and also reduces the rms intensity of the field ; that removing only one from 7 models should change the mean by so much shows just how different model D is. Going to either the 'good' or 'poor' mean3 changes the mean by 1.4 nT rms. (That the two values are both 1.4 follows from mean6 being the arithmetic mean of mean3g and mean3p. The actual value of 1.4 is consistent with the mean-square values of line mean6 representing the variances of random errors.) Similarly, the rms difference of (mean3g-mean3p) would be expected to be twice as large, consistent with the actual value of 2.9.

Proposed weighting for DGRF

I propose that we should use a weighted mean EXCLUDING model D.

Putting the models in order of the mean-square values for mean6 we get Table 3, where there is a not too arbitrary division into two groups (plus model D):

TABLE 3 Ordering in terms of mean-square deviation from mean6

Model	B	A	G		F	C2	E2		D
ms from mean6	2.8	4.1	5.5		11.6	16.4	17.9		208
suggested weight	1	1	1		1/2	1/2	1/2		0

If these ms values are taken to indicate the magnitude of the respective ms errors, this suggests that typically the models (C,E,F) have about 4 times the variance of the models (A,B,G). For purely random errors this would indicate using relative weights of 1/4 for the 'poor' models. But as well as random errors, there are almost certainly systematic errors in the models, and to reduce the effect of these it would be good to keep a variety of models. So as a compromise I suggest that we use a weight of 1/2.

So for DGRF 2005 I propose the following weights:

A, B, G given weight 1
C, E, F " " 1/2
D " " zero.

3. IGRF 2010

There are 7 candidate models: A, B, C2, D, E2, F, G.

Call the (unweighted) arithmetic mean of all 7 models mean7. The resulting rms/ms differences from mean7 are given in the second row of the Tables 4a,b.

TABLE 4a rms vector difference (nT) between models and various means

Model	A	B	C2	D	E	F	G	Models used in mean
rms from mean7	6.3	3.8	7.1	12.3	10.9	5.8	4.6	A+B+C+D+E+F+G
rms from mean5g	5.6	2.7	6.8	<i>14.0</i>	<i>12.3</i>	4.7	4.3	A+B+C +F+G
rms from mean2p	<i>10.8</i>	<i>9.5</i>	<i>10.9</i>	9.7	9.7	<i>10.8</i>	<i>9.2</i>	D+E

TABLE 4b mean-square vector difference (nT)² between models and various means

Model	A	B	C2	D	E	F	G	Models used in mean
ms from mean7	40	15	50	151	120	33	21	A+B+C+D+E+F+G
ms from mean5g	32	7	46	<i>197</i>	<i>152</i>	22	18	A+B+C +F+G
ms from mean2p	<i>116</i>	<i>90.8</i>	<i>118</i>	94	94	<i>117</i>	<i>84</i>	D+E

Note: In the Tables, figures in italics are for models NOT used in that mean.

There is a range of a factor of 10 in the mean-square differences. If the two 'poor' models D and F (see Table 6) are omitted we get mean5g, giving the mean5g lines of Tables 4a,b. The values for all the 'good' models have been slightly reduced, and those for all the 'poor' models slightly increased, by about the same factor.

If the two 'poor' models are averaged to give mean2p, this gives the mean2p lines of the Tables 4a,b. Compared with the values for mean7, the values for all the 'good' models have been much increased, and those for all the 'poor' models somewhat reduced, though by a much smaller factor. Again this suggests a small systematic difference between the 5g and 2p groups of models, though the actual differences between the means is consistent with the errors being purely random. Also again, given only the mean 2p line, it would be very difficult to decide which models had been used to obtain the mean!

The differences between the various means are given in Tables 5

TABLE 5a rms vector differences (nT) between different means

	0		mean5g	mean2p
mean7	43819.2		2.6	6.4
mean5g	43818.7			9.0
mean2p	43820.3			

(The second column gives the total rms field of that model, i.e. the rms intensity of that model. The other columns give the rms intensity of the vector difference between the respective models.)
(That 9.0 appears to be the sum of 2.6 and 6.4 is due purely to chance and rounding!)

TABLE 5b mean-square vector differences (nT)² between different means

			mean5g	mean2p
mean7			6.5	40.9
mean5g				80.1
mean2p				

As would be expected, these differences are quite a bit larger than for the DGRF. As for the DGRF, the three values are consistent with the mean7 values of Table4 representing random errors.

Proposed weighting for the IGRF

Putting the models in order of the ms values in line mean7 we get Table 6:

TABLE 6 Ordering in terms of mean-square deviation from mean7

Model	B	G	F	A	C		E	D
ms from mean7	15	21	33	40	51		120	151
suggested weight	1	1	1	1	1		1/3	1/3

In Table 6 the first five models have a range of about 3 in mean-square values, but it would be difficult to divide this first group into sub groups. There is then a jump of a factor of 2.5 to 3 to the remaining two models. I therefore suggest taking the five 'good' models (A,B,C,F,G) as one group, and the two 'poor' models (D,E) as another group. If the mean7 line is used to indicate the magnitude of the respective errors, this suggests that on average the 'poor' models have about 7 times the variance of the 'good' models. For purely random errors this would indicate relative weights of 1/7 for the 'poor' models. But there are almost certainly systematic errors, as well as random errors, in the models, and to reduce the effect of these it would be good to keep a variety of models. So as a compromise I suggest that we use a weight of 1/3.

So for DGRF 2005 I propose the following weights:

A, B, C, F, G	given weight	1
D, E	" "	1/3

4. SV 2010-2015

There are eight candidate models: A, B, C2, D, E, F, G, H

Call the (unweighted) arithmetic mean of all 8 models mean8. The resulting rms/ms differences from mean8 are given in the mean8 rows of Tables 7a,b.

TABLE 7a rms vector difference (nT/yr) between models and various means

	A	B	C2	D	E	F	G	H	Models used in mean
Mean8	12.8	7.4	9.7	4.1	12.9	4.1	16.9	6.6	A+B+C+D+E+F+G+H
Mean5g	<i>14.4</i>	8.2	7.6	3.2	<i>12.7</i>	4.5	<i>19.1</i>	4.9	B+C+D +F +H
Mean3p	11.4	<i>8.6</i>	<i>13.9</i>	<i>8.3</i>	14.6	<i>7.2</i>	14.0	<i>10.8</i>	A +E +G
Mean2g	<i>13.5</i>	<i>7.3</i>	<i>9.3</i>	3.4	<i>13.7</i>	3.4	<i>17.8</i>	<i>6.3</i>	D +F
Mean6p	12.7	7.6	10.0	<i>4.7</i>	12.7	<i>4.6</i>	16.7	6.9	

TABLE 7b mean-square vector difference (nT/yr)² between models and various means

	A	B	C2	D	E	F	G	H	Models used in mean
Mean8	164	55	93	17	166	17	287	44	A+B+C+D+E+F+G+H
Mean5g	<i>208</i>	68	58	11	<i>162</i>	20	<i>365</i>	24	B+C+D +F +H
Mean3p	130	<i>74</i>	<i>102</i>	<i>68</i>	24	<i>51</i>	97	<i>117</i>	A +E +G
Mean2g	<i>182</i>	<i>54</i>	<i>86</i>	11	<i>187</i>	11	<i>318</i>	<i>40</i>	D +F
Mean6p	160	58	98	<i>22</i>	162	<i>22</i>	278	47	A+B+C +E +G+H

Note: In the Tables, figures in italics are for models NOT used in that mean.

(For mean8, the apparent equality for the D and F values is due to rounding.) There is a range of a factor of 17 between the largest and smaller mean-square values, which is about the same range as for the DGRF candidates. However for the SV models the values are much more evenly spread; the ratio between the largest value and the next two largest is only 1.7, compared with 7 for the DGRF.

If we ignore the three models having the largest values we get mean5g, and the values in the mean5g lines of Tables 5. Now it is only on average that for the five models used in mean5 their mean-square decreases; while for (C,D,H) the table values decrease, for the other two (B,F) there is a (relatively smaller) increase. Similarly, on average the mean square value increases for the three models excluded from mean5, but while the values for (A,G) increase significantly, that for E has a marginal decrease.

Conversely, if one takes only the three models having the largest values in line mean8, we get mean3p. Compared with lines 2, for the three contributing models there are reductions in the values (large for E and G, much smaller for A), and for the five excluded models there are factors of 3 to 4 increases for D, F, H, and much smaller increases for B, C. Again, given only line 4, it would be very difficult to decide which models had been used to obtain the mean!

If we use only the two models having the smallest values, we get mean2g. Of the other 6 models, compared with the results for mean8, three were increased and three were decreased. Using the other 6 models gives mean6p. Not surprisingly, compared with the results for mean8, the values for the excluded D and F are increased, and for the six included models there are small decreases for four and small increases for two.

Table 8a. rms vector differences (nT/yr) between means.

	0		mean5g	mean3p	mean2g	mean6p
mean8	82.6		3.0	5.0	2.4	0.8
mean5g	81.5			8.0	2.1	3.6
mean3p	84.5				7.0	4.5
mean2g	82.6					3.2
mean6p	82.6					

(The second column gives the total rms field of that model, i.e. the rms intensity of that model. The other columns give the rms intensity of the vector difference between the respective models.)

Table 8b. mean-square vector differences (nT/yr)² between means.

			mean5g	mean3p	mean2g	mean6p
mean8			9.0	25.1	5.8	0.6
mean5g				64.4	4.2	13.2
mean3p					48.6	19.9
mean2g						10.3
mean6p						

(Much of the differences in overall rms field comes from large differences in the three dipole coefficients.)

For the group (mean8, mean5g, and mean3p) the values of Table are probably consistent with the mean8 values of Table 7 being purely random errors. But for the group (mean8, mean2g, and mean6p) the values are a factor three less than I would have expected; I have no explanation for this.

Proposed weighting for the SV

Putting the models in order of the ms values in line mean8 we get Table 9

TABLE 9 Ordering in terms of mean-square deviation from mean8

Model	F	D	H	B	C	A	E	G
ms from line 2	17	17	44	55	93	164	166	287
possible weight	1	1	1/2	1/2	1/2	1/3	1/3	1/4

There is a not too arbitrary division into four groups. Using the arguments I have used above it would seem appropriate to give the first three groups weights of 1, 1/2, and 1/3. Although model G has the largest value, in this SV case I do not think there is evidence that it is so different as to be rejected, so I suggest including it with a weight of 1/4.

So for SV 20010-15 I propose the following weights:

F, D	given weight	1
B, C, H	" "	1/2
A, E		1/3
G	" "	1/4

BEHAVIOUR OF PARTICULAR GROUPS OF HARMONICS

A reminder:

For models A, C, G	DGRF, IGRF, SV are all derived from one parent model
For models B, D, F	the DGRF uses a different parent model from IGRF/SV
For model E	each of DGRF, IGRF, SV are derived separately
For model H	there is only SV

In the discussion of this section, the differences are with mean6 for DGRF, mean7 for IGRF, and mean8

BACKUS EFFECT (Larger scatter for $m=n$ coefficients. It occurs when there is a significant proportion of intensity data; this can reduce the error *except* for near $m=n$.)

For the DGRF, looking at the deviations from mean6 suggests this is probably absent for model B, and small or very small for A,C,E,F,G.

For the IGRF, the deviations from mean7 suggest it is small (though still present) for models B,F,G, moderate for E, and large for A,C,D.

For the SV, the deviations from mean8, suggest that models B,E probably show no effect, D,E,H show a very slight effect, and A,C,G show significant effect. (Model H does not use intensity data directly. That it shows a small effect is probably due to the CHAOS model used to constrain the dynamics; the corresponding model A IGRF and SV models do show significant effects.)

POLAR DATA GAP

(A data gap near the geographic poles leads to increased scatter at $m=0,1$)

For the DGRF: Small for model A, moderate for F,G, and large for B,C (though mainly for n odd) and E.

For the IGRF: Small for models B,E, moderate for G, and large for A,C,D,E (but only in the $m=1$ coefficients) and F.

For the SV: Probably none for model B, small for A,C,H, and moderate for D,E,F,G.

DIPOLE TERMS. PARTICULARLY h_1^1

(There is the problem of whether the dipole term does, or should, contain an average induced component. This can lead to the dipole coefficients having the largest absolute uncertainties. If the induced field were the problem, we would expect the deviations to be in the same ratio as for the static dipole.)

For the DGRF: No effect. For all of the 6 models the largest m s deviation from mean6 is not a dipole term; the largest deviations are scattered up to $n=7$.

For the IGRF: For model A the largest deviation is for g_1^0 , and for E it is h_1^1 . But for the other four models the largest deviations are scattered up to $n=4$ (one Backus term, three $m=0$ terms).

For the SV: For 6 of the 8 models, the largest is for h_1^1 . For the seventh model the h_1^1 value is only slightly less than g_2^0 , and for the eighth model the h_1^1 value is exceeded only by g_2^2 , and g_1^1 . Only for model A is the g_1^0 value comparable in magnitude.

It looks as though there is no problem with the main field, but there is certainly a problem with the SV dipole, though probably from another origin. **The model G team noted explicitly that SA of h_1^1 increased rapidly after about 2008, and warn of a possible jerk! Similarly the model F team suspect there might have been a jerk between 2007.5 and 2008.2, depending on the observatory!**

NIGHT-TIME IONOSPHERIC INDUCED FIELD

(Lowe & Olson (2004) showed that assuming there is no ionospheric (induced) field at night led to systematic errors, particularly in the g_3^0 and h_3^1 coefficients.)

Only Model A corrected for this, and only with observatory data. However with the present sets of models it is not possible to separate any such effect from that of the polar gap.