Earth Planets Space, **58**, 409–416, 2006

Deriving main field and secular variation models from synthetic Swarm satellite and observatory data

Vincent Lesur, Susan Macmillan, and Alan Thomson

British Geological Survey, Murchison House, West Mains Road, Edinburgh, EH9 3LA, UK

(Received November 18, 2004; Revised April 19, 2005; Accepted September 30, 2005; Online published April 14, 2006)

We investigate how well the core field component of the underlying model used in the Swarm End-to-End simulator can be recovered when relatively simple geomagnetic field models are fitted to synthetic Swarm satellite and observatory data. In particular we demonstrate the potential benefits of Swarm by deriving models without these data. From two years of observatory data, the underlying secular variation model is recovered up to about spherical harmonic degree 8. This maximum degree is higher than was expected but increases to 10 when satellite data are also used. The constant part of the geomagnetic field model is recovered to at least degree 17. These results improve when a better statistical treatment of the data errors is made, a slightly more complex field model is fitted, or when five years of data are used.

Key words: Geomagnetism, geomagnetic field, swarm.

1. Introduction

For the Swarm End-To-End Mission Simulator Study various synthetic data sets were derived, spanning five years from 1997, using the Comprehensive Model (CM), (Sabaka *et al*., 2004) and a model of the magnetospheric and associated induced fields constructed from observatory data (Olsen *et al*., 2006). In the simulator study, the underlying model of the main field, its variations with time, and the crustal field were all well recovered from selections of these synthetic data sets using comprehensive inversions (Sabaka and Olsen, 2006). Although these comprehensive inversions try to recover all sources of the underlying field, many widely available models, e.g. the International Geomagnetic Reference Field (IGRF) (IAGA, 2003), are simpler models. It is therefore of interest to study what can be achieved using relatively simple modelling techniques. In addition, as the forward modules and inverse modules in this work were not completely independent of one another, the results of other inversions were therefore of value in determining the final Swarm constellation (Olsen *et al*., 2004).

Our main objective here is to show the benefits of a Swarm mission by deriving models with and without the Swarm satellite synthetic data, rather than to demonstrate the benefits of one particular Swarm constellation over another. In order to compare like with like, only relatively simple geomagnetic field models are fitted to different synthetic data sets. In particular, we do not try to model ionospheric fields, or the crustal field for spherical harmonic degrees higher than nineteen. We derive three different basic models using data sets that include observatory data only, satellite data only and, finally, a data set combining observatory and satellite data.

In Sections 2 to 5 we describe the data selection, model parameterization, data weighting and model estimation process. In Section 6 the basic results are presented and discussed. In Section 7 we extend the analysis using a longer data set. We conclude with summary comments in Section 8.

2. Data Selection

2.1 Satellite data

We selected synthetic data from three of the Swarm satellites over two periods, 1997.0–1999.0 and for the whole 5-year time span. We used data from one high-altitude satellite and two low-altitude satellites (satellites A, B, C in Olsen *et al*., 2006). This constellation has been shown to be the best if only three satellites are selected (from four). The majority of the results presented use data for the 2-year time span. This was chosen as a compromise. On the one hand, if less than two years of data are used there is a risk of leakage of the internal field induced by the annual and semi-annual variations of the external field into the main field Gauss coefficients. Furthermore, the value of the Gauss coefficients from degree ten and above may not be very robust. On the other hand, if five years of data are selected, the complexity of the behaviour in time of the low degree main field Gauss coefficients may be difficult to model.

The Swarm data were treated as if they were a real data set, measured during varying external field conditions, traditionally quantified by magnetic indices and solar wind data. Thus we selected data using not only those indices that were input to the comprehensive model to generate the synthetic data, but also other indices commonly used in satellite magnetic data selection. The magnetic index and other data were obtained independently and were not provided with the Swarm database. For each period the data were filtered according to: local-time 23:00–06:00, minimizing the contribution from the ionosphere; $Kp < 1+$ (and $2-$

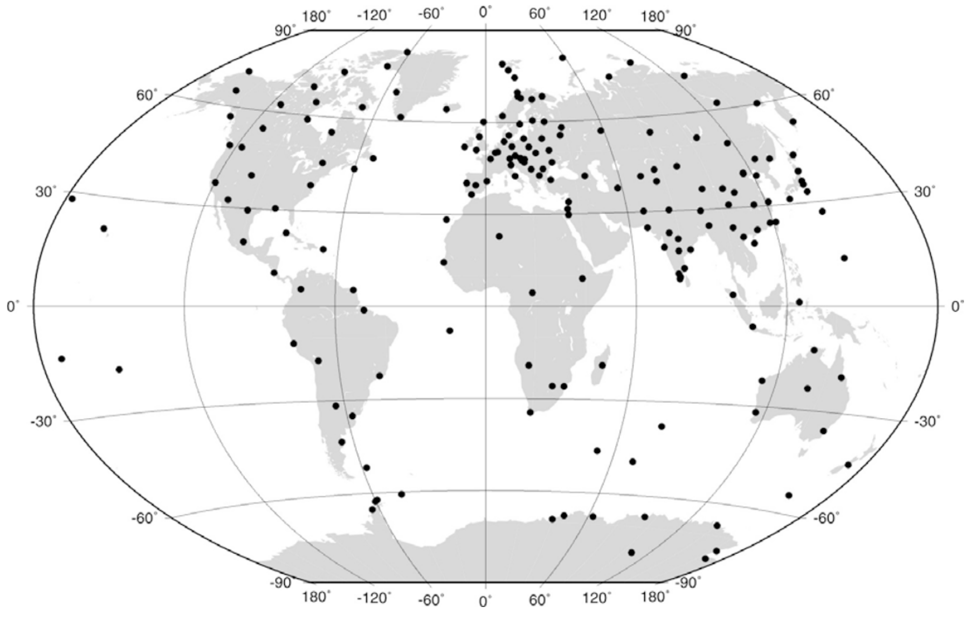

Fig. 1. Locations of observatories.

in the previous three hours); $-10 < Dst < +10$ (± 15 for the previous hour); interplanetary magnetic field data with $-1 < Bz < +5$, $|By| < 3$, $|Bx| < 10$ nT and solar wind speed $Vsw < 450$ km/s. These criteria were used to select data during magnetically quiet night-time conditions. Vector data were selected in the geomagnetic latitude range $[-60°: 60°]$ and scalar data were selected at higher geomagnetic latitudes in an attempt to minimize the effect of noise from auroral and polar cap current systems.

An additional criterion that rejects satellite data over a sunlit ionosphere is occasionally reported in the literature. When this condition is used, the resulting spatial distribution, particularly at high latitudes, varies according to season. We have observed that this variation introduces a bias in the low internal degree Gauss coefficients of our models and thus we have not used it in the results reported here.

2.2 Observatory data

The Swarm observatory data set was originally synthesized using exclusively, for the magnetospheric field, the CM and most of the results we present here were obtained from this data set. Only a few results are presented that use data with the magnetospheric field and induced effects synthesized as in Olsen *et al.* (2006). In the latter higher degree external fields and a more realistic conductivity model are used. We used the same selection criteria in both cases.

Hourly mean vector data at 188 geomagnetic observatories were selected from 1997.0 to 1999.0. Their locations, the same as those of currently operating observatories, are shown in Fig. 1. The selection criteria were largely similar to those for satellite data: $Kp \leq 1+$, $0 \leq Bz$, $-15 < Dst < 0$ and local times between 23:00 and 05:00. Outside the latitude range $[-55°: 55°]$ the north and east component data are not used in order to minimize the effect of noise from auroral and polar cap current systems.

3. Model Parameterization

Away from its sources, the magnetic field B is a potential field and therefore can be written as the negative gradient of a potential $B(\theta, \varphi, r, t) = -\nabla V(\theta, \varphi, r, t)$. This potential can be expanded in terms of spherical harmonics:

$$\begin{aligned} V(\theta, \varphi, r, t) &= a\left\{\sum_{l=1}^{li}\sum_{m=0}^{l}(g_l^m(t)\cos(m\varphi) + h_l^m(t)\sin(m\varphi))\left(\frac{a}{r}\right)^{l+1}P_l^m(\cos\theta)\right\} \\ &+a\left\{\sum_{l=1}^{le}\sum_{m=0}^{l}(q_l^m(t)\cos(m\varphi) + s_l^m(t)\sin(m\varphi))\left(\frac{r}{a}\right)^{l}P_l^m(\cos\theta)\right\} \end{aligned} \quad (1)$$

where a (6371.2 km) is the Earth's reference radius, (θ, φ, r) are spherical coordinates in a geocentric reference frame, $P_l^m(\cos\theta)$ are the Schmidt semi-normalized Legendre functions, and $(g_l^m(t), h_l^m(t))$ and $(q_l^m(t), s_l^m(t))$ are the time-dependent Gauss coefficients describing internal and external sources respectively. The internal Gauss coefficients are assumed to have a polynomial dependence on time:

$$\begin{aligned} g_l^m(t) &= \sum_{it=itmi}^{itma} g_l^{m(it)}(t - t_0)^{it} \\ h_l^m(t) &= \sum_{it=itmi}^{itma} h_l^{m(it)}(t - t_0)^{it} \end{aligned} \quad (2)$$

where the time is given in decimal year and t_0 is the reference date of the model (1998.0). The minimum and maximum degrees of the polynomials are given by itmi and itma respectively. Thus if $itmi = 0$ and $itma = 1$, only constant and linear terms are solved for. The external Gauss coefficients have no polynomial dependence on time.

In the CM, seasonal variations are introduced in rather complex ionospheric and magnetospheric parts of the model. Here we introduce such variations only for external

Table 1. Mean and rms misfit of the three models to the input data (nT).

Component	N	*model_o*		*model_s*		*model_so*	
		mean	rms	mean	rms	mean	rms
X satellite	75957			0.09	7.92	0.28	8.37
Y satellite	75957			−0.97	10.05	−0.88	9.99
Z satellite	75957			0.03	3.80	0.03	3.63
F satellite	47299			0.80	9.09	0.24	7.02
X observatory	99922	0.24	1.79			0.62	4.24
Y observatory	99922	0.36	2.64			0.39	3.88
Z observatory	99922	0.03	1.47			0.13	1.73
Z_hl observatory	35510	0.14	3.95			0.25	4.27

The X, Y and Z components are oriented north, east and down respectively and N is the number of data values. The "Z_hl observatory" data are the vertical field components at the high latitude observatories.

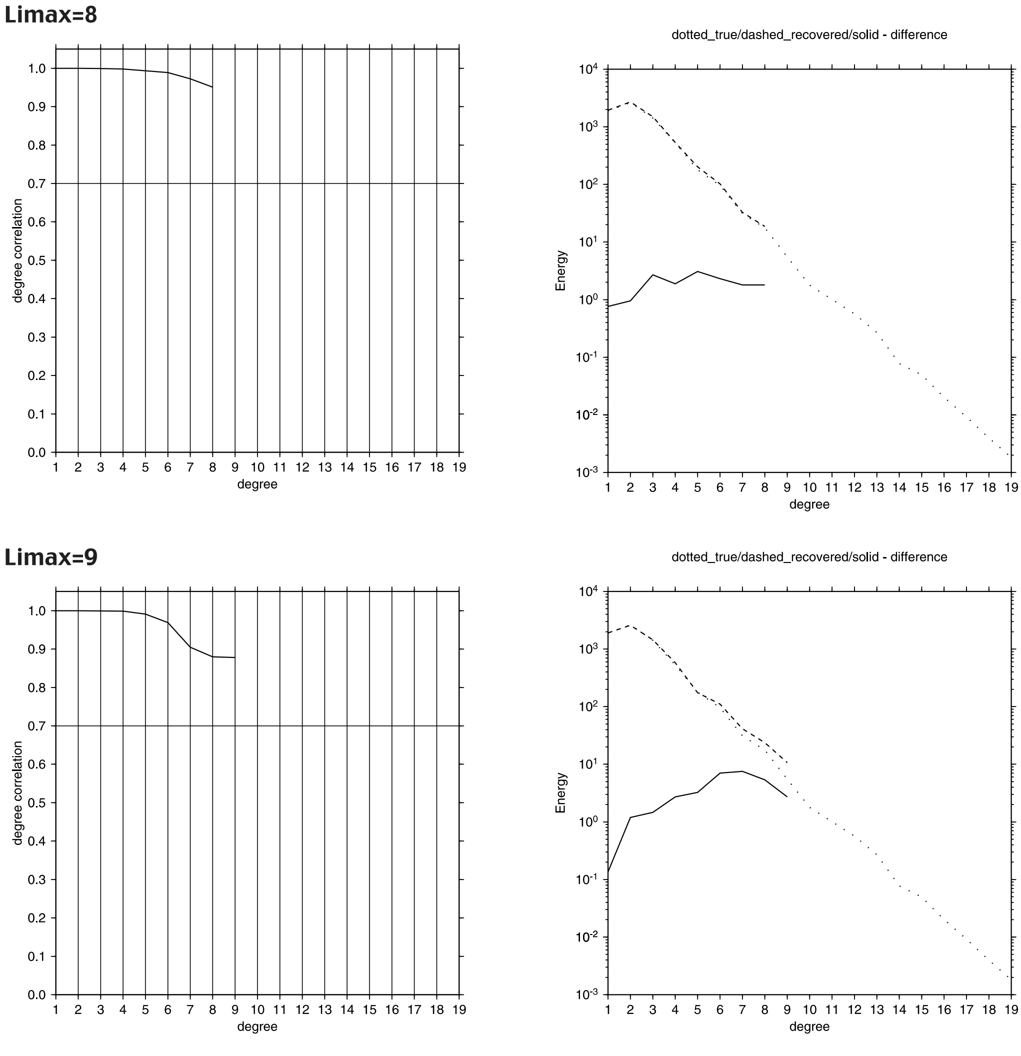

Fig. 2. Degree correlation and power spectrum plots for the secular variation (sv) model obtained from the observatory data set alone. Results are presented for an internal maximum spherical harmonic degree $li_{\max} = 8$ or 9.

and internal degree 1 and 2 Gauss coefficients. Let $\tilde{g}_1^m(t)$ be the parts of the $g_1^m(t)$ coefficients that account for these seasonal variations. Then:

$$\tilde{g}_1^m(t) = \tilde{g}_{1,1c}^m \cos(2\pi(t - t_0)) + \tilde{g}_{1,1s}^m \sin(2\pi(t - t_0)) + \tilde{g}_{1,2c}^m \cos(4\pi(t - t_0)) + \tilde{g}_{1,2s}^m \sin(4\pi(t - t_0)). \quad (3)$$

Similar representations are used for $h_1^m(t)$, $q_1^m(t)$, $s_1^m(t)$ and for degree 2 internal and external Gauss coefficients.

A Dst-dependence for the degree 1 Gauss coefficients is introduced to represent the variability of the magnetospheric ring current. We used a simple fixed ratio of 0.27 (Langel and Estes, 1985) between the external Dst dependence and the associated internal induced contribution.

Finally, since synthetic observatory data are used, we

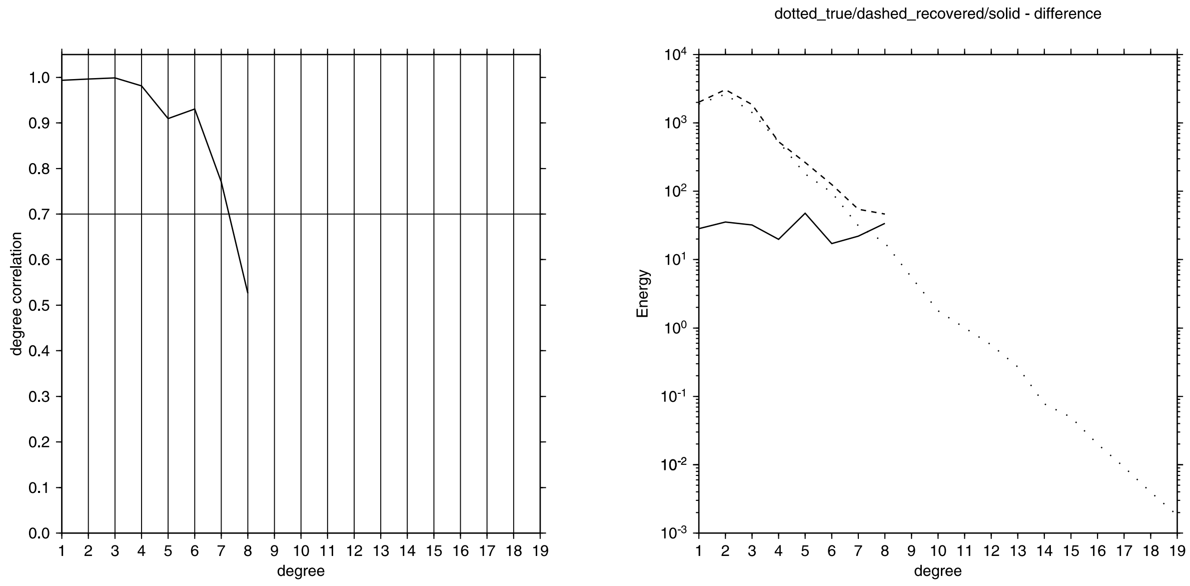

Fig. 3. Degree correlation and power spectrum plots for the sv model obtained from the observatory data set alone when the magnetospheric field was synthesized using the technique presented in Olsen *et al.* (2006). Results are presented for an internal maximum spherical harmonic degree $li_{\max} = 8$.

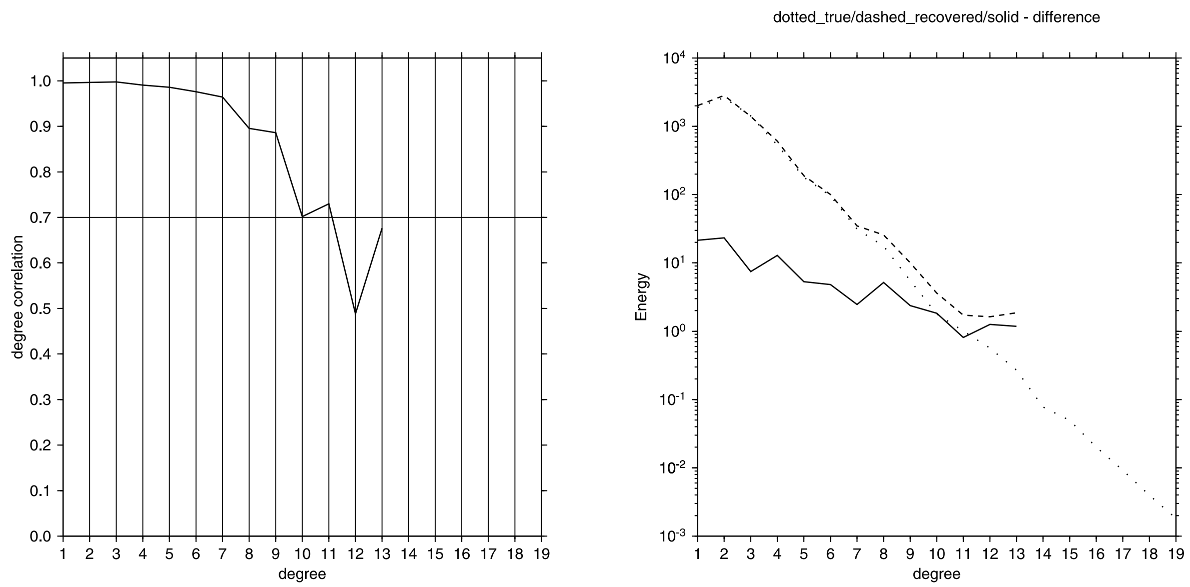

Fig. 4. Degree correlation and power spectrum plots for the sv model obtained from the satellite-only data set.

introduce offsets at each observatory to take into account the field, constant in time, which cannot be described by our models. At an observatory, the magnetic field B is:

$$B\,(\theta, \varphi, r, t) = -\nabla V\,(\theta, \varphi, r, t) + O(\theta, \varphi, r) \qquad (4)$$

where the offset vector $O(\theta, \varphi, r)$ is constant in time. There are therefore three new parameters per observatory in the geomagnetic latitude range -55° to 55°. Outside this range, we only use the vertical component of the magnetic field to maintain a linear relationship between the observatory offsets and the data, and, therefore, only one parameter is needed, per observatory, to account for the unmodelled field. The unmodelled field is usually mainly very short wavelength crustal field (i.e. the local field) but this was not included in the synthetic data. However, in the present case, the mean values of any unmodelled signal in the data will be included in the offsets. With such a parameterization, the observatory data only provide information about variation in time of the geomagnetic field, and not about its absolute level.

We derived three different models using the above parameterization. The first model ($model\_o$) was derived from observatory data only. In this model the maximum internal degree used in Eq. (1) was $li = 8$ and the time dependency defined in Eq. (2) was $itmi = 1$ to $itma = 3$. We also investigated the effect of increasing li and will therefore present results with $li = 9$. We did not increase the value of above 3 since the CM model uses only cubic B-splines in time with a knot separation of 2.5 years. The other models were derived either from satellite data alone ($model\_s$), or from a combination of satellite and observatory data ($model\_so$). Both these models had, in Eq. (1), a

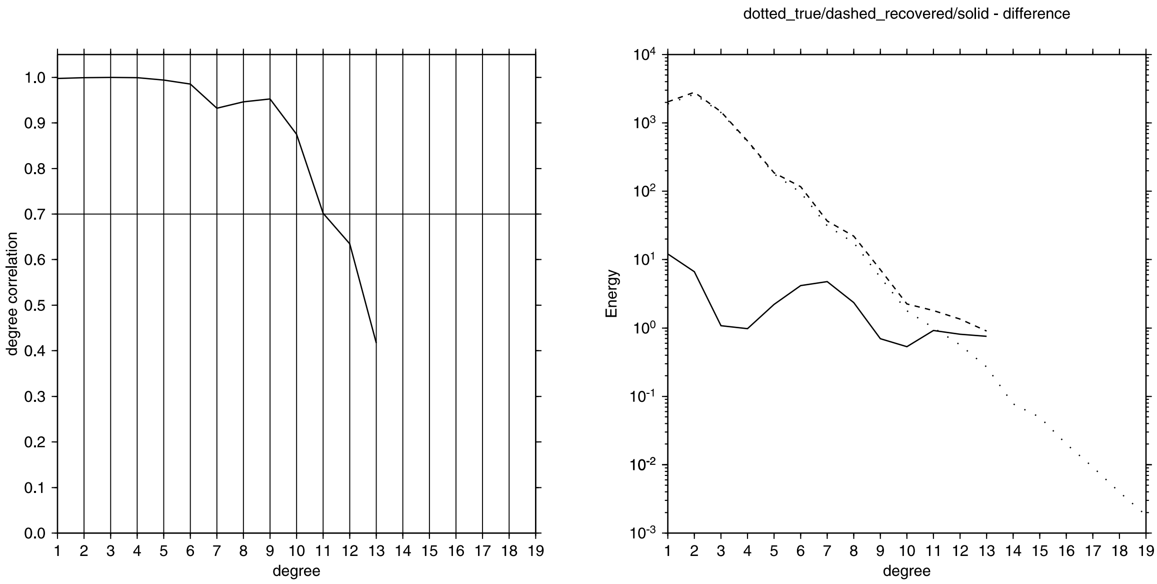

Fig. 5. Degree correlation and power spectrum plots for the sv model obtained from the combined satellite and observatory data set.

maximum internal degree $li = 19$ with $itmi = 0, itma = 3$ up to spherical harmonic degree 8, $itmi = 0, itma = 1$ for degrees 9 to 13, and $itma = 0$ for degrees higher than 13. These bounds were dictated by the maximum complexity of the model that we estimate the data set is able to resolve, rather than by any characteristic of the internal Gauss coefficients of the underlying model.

4. Data Weighting and Covariance Matrix

The variances associated with both satellite and observatory data were defined as:

$$\sigma^2 = (\sigma_0 + d_z(1 + \cos(za)))^2 \tag{5}$$

where za is the zenith angle of the sun, $d_z = 5$ nT and $\sigma_0 = 2$ nT for all scalar or vector data. The dependence on the zenith angle was introduced to account for the increased noise level due to the higher conductivity of the sunlit ionosphere.

No further selection or decimation of the data was introduced to deal with the high data density at high latitudes. Instead, the data values were multiplied by weights. The spherical surface was divided into roughly equal-area cells whose size at the equator was 5° in latitude and longitude. For a data point in a given cell, a weight was calculated as the ratio of the average number of data in all non-empty cells to the number of data in that cell. These weights were computed independently for observatory and satellite data. To introduce these weights into our parameter estimation scheme, their inverse values filled diagonal weight matrices that left and right multiplied the covariance matrix, whose diagonal elements are the variances given in Eq. (5). The resulting matrix C is then diagonal.

5. Model Estimation

We estimated the model parameters by fitting the data using a classic least-squares approach. For this application the iterative process can be written as:

$$\begin{aligned} d_i &= G(p_i) \\ \delta d_i &= d_{obs} - d_i \\ \delta p_i &= \left[G_i^t C^{-1} G_i\right]^{-1} \; G_i^t C^{-1} \, \delta d_i \\ p_{i+1} &= p_i + \delta p_i \end{aligned} \tag{6}$$

where the subscript $_i$ denotes the ith iteration, p is the model vector (i.e. the model parameters), $G(p)$ is the forward non-linear function used to calculate the predicted values d from a model p, d_{obs} is the data vector and G is the $n \times p$ matrix associated with the equations of condition (n: number of data values; p: number of model parameters):

$$G = \left[\frac{\partial G_i(p)}{\partial p_j}\right]_{\substack{i=1,n \\ j=1,p}} \tag{7}$$

The models were fitted to the data sets in four iterations as described in Eqs. (6). The inverse of the normal equation matrix was calculated using an eigenvalue/eigenvector decomposition. When solving the problem for *model_o*, eight null eigenvalues and their associated eigenvectors had to be removed from the inversion process. This was a result of the fact that, as for the internal field, the part of the external field that is constant in time cannot be separated from the observatory offsets using observatory data alone. The number of data, the mean and the rms misfits to the data for the resulting models are given in Table 1.

6. Results

In Fig. 2 we show, for *model_o*, the degree correlation, as defined in Olsen *et al.* (2006), and power spectrum of the secular variation of the reference (underlying) model together with the power spectrum of the differences between the recovered and the reference secular variation models. These results are presented for maximum degrees $li = 8$ and $li = 9$ in Eq. (1). Clearly, the best results are obtained with $li = 8$ where all 80 parameters of the secular variation model are robustly estimated as indicated by the degree correlation. A degree correlation greater than 0.7 is deemed "acceptable" (Olsen *et al.*, 2006). This in itself is

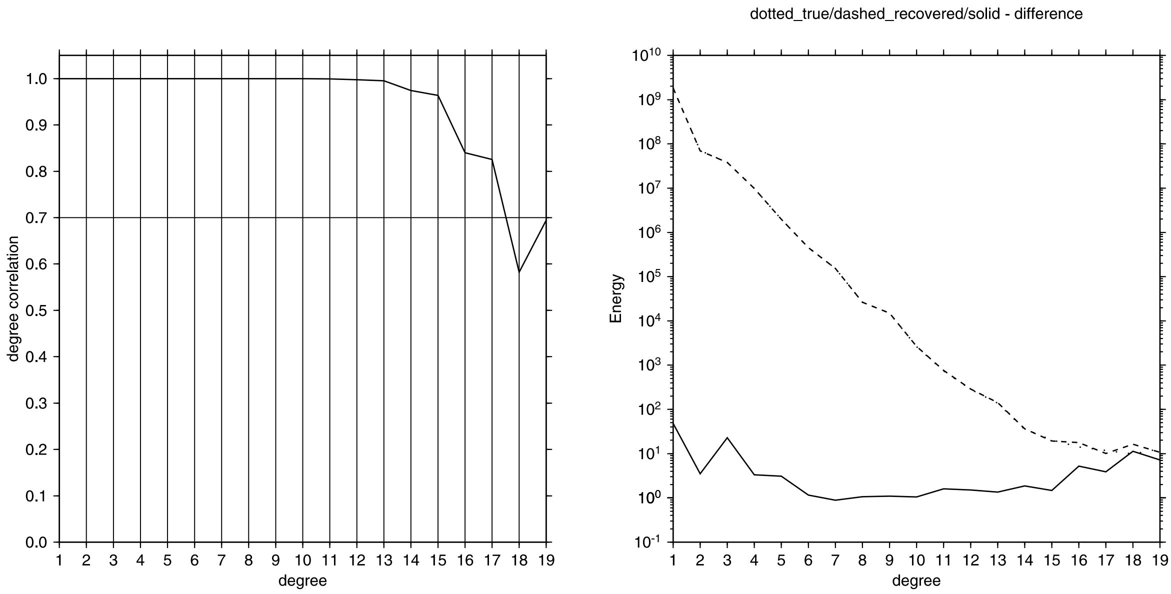

Fig. 6. Degree correlation and power spectrum plots for the main field model obtained from satellite data only.

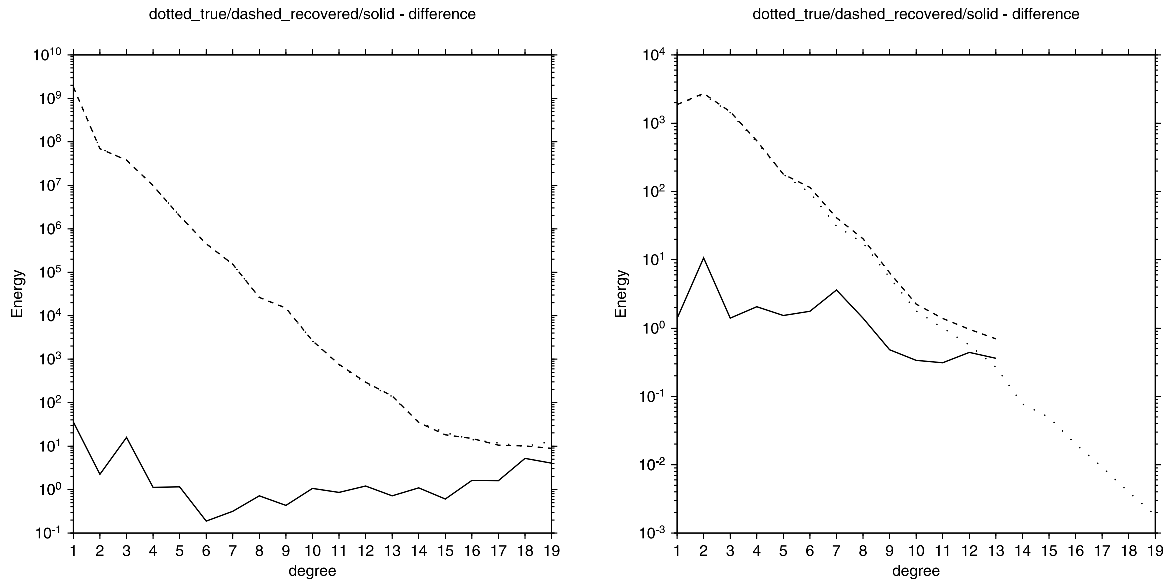

Fig. 7. Power spectrum plots for the "improved" main field (left) and sv (right) models obtained from the satellite-only data set.

a slightly surprising result as it was always thought that observatory data had to be complemented with other types of data, for example repeat station data, in order to model secular variation. It provides some justification for having the predictive secular variation part of the IGRF (IAGA, 2003) extending to degree 8 as observatory data are an essential component for making predictions for several years into the future. When the maximum degree is increased to $li = 9$ some of the $(g_l^{m(1)},\ h_l^{m(1)})$ estimates deviate from their reference values at degree five and above. If the maximum degree is increased further, the quality of the agreement with the underlying model collapses and some regularization (damping) is necessary.

Figure 3 shows the degree correlation and power spectrum plots for $li = 8$ when the model was built from the data set with the magnetospheric field synthesized as in Olsen *et al.* (2006). The effect of using this data set is to increase the power spectrum of the differences by one order of magnitude and this demonstrates the importance of using a good parameterization of the external field.

Figure 4 shows the same results as Fig. 3 but produced using $model\_s$. The satellite data provide an excellent distribution in space but the distribution in time is not necessarily optimal at all latitudes and longitudes. Consequently, if the external fields are well modelled, the satellite data can provide information about the secular variation at spherical harmonic degrees higher than eight, but the model must have a simple behaviour in time. This contrasts with observatory data where the maximum spherical harmonic degree used has to be relatively small, but where there is almost no limit to the complexity of the model's behaviour in time. Indeed, by using satellite data we are able to estimate the

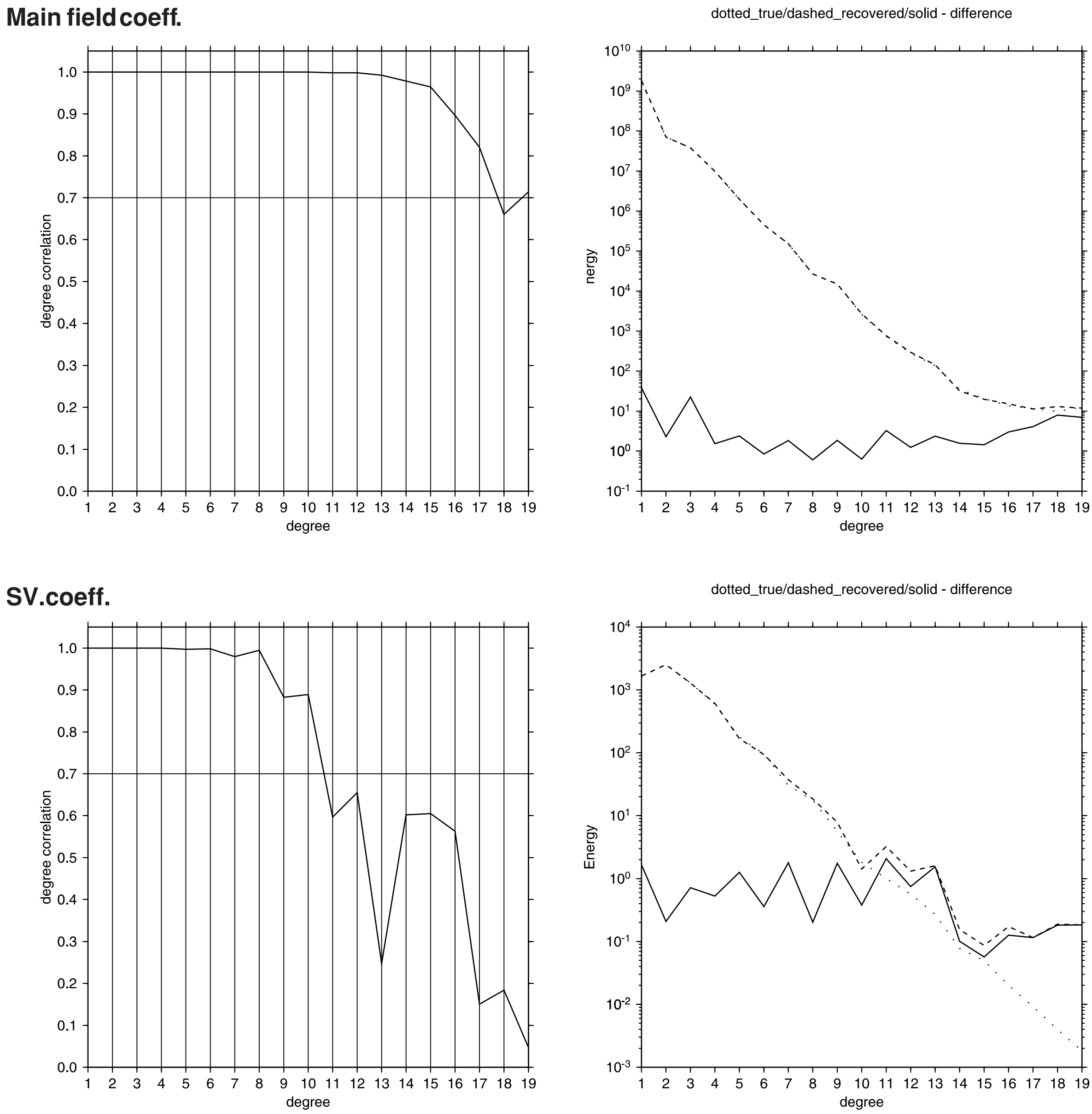

Fig. 8. Degree correlation and power spectrum plots obtained from five years of satellite data.

Gauss coefficients that are constant in time, i.e. estimate a full main field model. The recovery of the low degree secular variation Gauss coefficients is not as good as for the observatory data alone (Fig. 2). This is primarily because of the higher noise level in the residual, due to the complexity of the external field model and, in particular, because of the magnetic fields generated by field-aligned currents. By combining satellite and observatory data ($model\_so$: Fig. 5) we improve the recovery of the low degree secular variation Gauss coefficients.

In Fig. 6 we show the degree correlation and power spectrum plots for the constant Gauss coefficients of $model\_s$. Apart from the first three spherical harmonic degrees, and the two highest degrees, the underlying model is well recovered. The results for the first three degrees are due to the complexity of the external field model and its induced counterpart. These induced fields, including the fields induced by ionospheric variations, cannot be easily separated from the core-field Gauss coefficients.

We tried two further approaches to improve these results. Firstly it is possible to model other sources of the geomagnetic field. In particular, some parameters of the CM have a 24-hour periodicity. Our approach to model these fields is to represent them in a Geocentric Equatorial Inertial system of coordinates (Hapgood, 1992; Lesur *et al.*, 2005). Secondly, other distributions of residual errors can be used, rather than simply assuming a Gaussian distribution of error (Olsen, 2002; Lesur *et al.*, 2005). Using only satellite data, and assuming a Laplacian distribution of residual errors, we obtained the results shown in Fig. 7. Apart from the first three degrees, improvements can be observed in the recovery of both the main field and secular variation. The relatively large value at degree 2 of the secular variation power spectrum of differences is due to the erroneous estimate of the $g_2^{0(1)}$ term.

7. Other Data Selections and Analysis

The Swarm data set spans five years. We have therefore also produced several models over the full time span, using satellite data only. The main benefit of using several years of data should be the better recovery of the Gauss coefficients at high spherical harmonic degrees. In the following example, we tried to recover a model as close as possible to the underlying (reference) model. This model is the same as before with, in Eq. (1), a maximum internal degree $li = 19$, $itmi = 0$, $itma = 3$, up to spherical harmonic degree 13, $itma = 1$ and, for degrees higher than 13. The data set was obtained by combining data from satellites A, B, C and D (Olsen *et al.*, 2006). The selection process and the weights were otherwise the same as in Section 1 and 4.

In Fig. 8 we present the results for this model. This figure should be compared with the equivalent results, obtained from only two years of data, shown in Figs. 4 and 6. The introduction of the fourth satellite does not have a significant effect below degree 14. There is a clear improvement in the accuracy of the secular variation estimate, indicating that the error level in our original modelling was too high to see the complex behaviour in time of the low spherical harmonic degrees of the underlying model. This improvement in the secular variation model is due to the quantity of data used and also in the amplitude of the accumulated secular variation of the magnetic field over five years.

The main field coefficients are also well recovered up to spherical harmonic degree 17. Above this degree, we would not expect a good model recovery because of the likelihood of leakage of higher degree crustal field into our model. Again minor modifications to the model and a better handling of the noise in the data significantly improves the field model, particularly from degree 14 and above.

8. Conclusions

A relatively simple model of the geomagnetic field has been fitted to synthetic observatory and Swarm satellite constellation data. When only observatory data are used, the linear secular variation of the underlying model is recovered with an acceptable accuracy up to about spherical harmonic degree 8 depending on the magnetospheric model used to synthesize the data set. This maximum degree was higher than was expected and provides justification for the predictive secular variation part of the IGRF extending to degree 8. By introducing satellite data we first improve on the accuracy of the low degree secular variation Gauss coefficients, and second, acceptable estimates can now be derived up to degree 10. Satellite data also provide the opportunity to estimate the constant part of the geomagnetic field up to degree 17. Using a better statistical treatment of the data errors and a slightly more complex field model, we have achieved a significant improvement in the recovery of the Gauss coefficients. By using five years of data we also obtained better estimates of the Gauss coefficients, due to the quantity of data and the amplitude of the accumulated secular variation. These results, obtained from the Swarm synthetic data set, are believed to be realistic estimates of what can be expected with real data from the Swarm mission, due for launch in 2009. Clearly, an improvement in the modelling of the external fields and their induced counterpart is needed in order to take full advantage of the Swarm constellation data.

Acknowledgments. This paper is published with the permission of the Executive Director, British Geological Survey (NERC).

References

Hapgood, M. A., Space physics coordinate transformation: A user guide, *Planetary and Space Science*, **40**, 711–717, 1992.

International Association of Geomagnetism and Aeronomy (IAGA), Division V, Working Group 8. The 9th Generation International Geomagnetic Reference Field, *Earth Planets Space*, **55**, i–ii, 2003.

Langel, R. A. and R. H. Estes, Large-scale, Near-Earth magnetic fields from external sources and the corresponding induced internal field, *J. Geophys. Res.*, **90**, 2487–2494, 1985.

Lesur, V., S. Macmillan, and A. Thomson, Magnetic field model with daily variations of the magnetospheric field and its induced counterpart in 2001, *Geophys. J. Int.*, **160**, 79–88, 2005.

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

Olsen, N., R. Haagmans, T. J. Sabaka, A. Kuvshinov, S. Maus, M. E. Purucker, M. Rother, V. Lesur, and, M. Mandea, The *Swarm* End-to-End mission simulator study: A demonstration of separating the various contributions to Earth's magnetic field using synthetic data, *Earth Planets Space*, **58**, this issue, 359–370, 2006.

Olsen, N., E. Friis-Christensen, G. Hulot, M. Korte, A. V. Kuvshinov, V. Lesur, H. Lühr, S. Macmillan, M. Mandea, S. Maus, M. Purucker, C. Reigber, P. Ritter, M. Rother, T. Sabaka, P. Tarits, and A. Thomson, Swarm—End-to-End mission performance simulator study, Danish Space Research Institute Technical Report, 2004.

Sabaka, T. J. and N. Olsen, Enhancing comprehensive inversions using the *Swarm* constellation, *Earth Planets Space*, **58**, this issue, 371–395, 2006.

Sabaka, T. J., N. Olsen, and M. E. Purucker, Extending comprehensive models of the Earth's magnetic field with Ørsted and Champ data, *Geophys. J. Int.*, **159**, 521–547, 2004.

V. Lesur (e-mail: V.Lesur@bgs.ac.uk), S. Macmillan (e-mail: S.Macmillan@bgs.ac.uk), and A. Thomson (e-mail: A.Thomson@bgs.ac.uk)