Geophys. J. Int. (2004) **157,** 1027–1044 doi: 10.1111/j.1365-246X.2004.02256.x

A more realistic estimate of the variances and systematic errors in spherical harmonic geomagnetic field models

F. J. Lowes[1] and N. Olsen[2]

[1]*Physics Department, University of Newcastle upon Tyne, Newcastle upon Tyne,* NE1 7RU, *UK. E-mail: f.j.lowes@ncl.ac.uk*
[2]*Danish Space Research Institute, Juliane Mariesvej 30,* DK-2100 *Copenhagen Ø, Denmark. E-mail: nio@dsri.dk*

Accepted 2004 January 21. Received 2004 January 21; in original form 2003 January 14

SUMMARY

Most modern spherical harmonic geomagnetic models based on satellite data include estimates of the variances of the spherical harmonic coefficients of the model; these estimates are based on the geometry of the data and the fitting functions, and on the magnitude of the residuals. However, the accuracy of these estimates depends on the correctness of various assumptions, some of which are in practice not valid. Using the Ørsted OSVM model as an example, we show that ignoring the serial correlation of the magnetospheric 'noise' in the data, and not solving separately for the ionospheric field, led to quite inaccurate variance estimates. We estimate correction factors which range from $\frac{1}{4}$ to 20, with the largest increases being for the zonal, $m = 0$, and sectorial, $m = n$, terms.

With no correction, the OSVM variances give a mean-square vector field error of prediction over the Earth's surface at sea level of 1.5 nT^2 (for the field of degrees $n = 1$–13); applying our correction increases this to 3.5 nT^2.

The leakage of the ionospheric field into the solution also gave a large systematic error field, of magnitude about 6 nT.

We discuss how our approach is applicable to other similar modelling approaches.

Key words: geomagnetic field, serial correlation, spherical harmonic analysis.

1 INTRODUCTION

Any numerical model of the geomagnetic field can only be an approximation to the actual field, and we would like to have a reasonable estimate of the magnitude of the errors involved. Such an estimate is particularly necessary if the field model is used in further analysis, e.g. to deduce the fluid velocity at the top of the core; (see e.g. Bloxham & Jackson 1991). Most modern spherical harmonic geomagnetic models do include estimates of the variances of the coefficients, but such estimates can be grossly wrong. For simplicity, the approach to variance estimation suggested in this paper will be discussed mostly in terms of its application to one particular spherical harmonic model, the Ørsted main field and secular variation model (OSVM) of Olsen (2002). However, the approach itself is applicable to most such models, and this more general situation is discussed at the end.

This OSVM model approximates the geomagnetic field in terms of a truncated spherical harmonic series representation of a magnetic scalar potential; for the internal terms it is of the form shown in eq. (1.1) with corresponding terms for a linear time variation (the secular variation), and similar terms (but with a more complicated time dependence) for the low-degree field of external origin:

$$V = a \sum_{n=1}^{N} \sum_{m=0}^{n} \left[P_n^m(\cos\theta) \times \left(g_n^m \cos m\lambda + h_n^m \sin m\lambda \right) (a/r)^{n+1} \right]. \quad (1.1)$$

In this expression a is a reference radius (taken to be the mean radius of the Earth), the $P_n^m(\cos\theta)$ are Schmidt semi-normalized associated Legendre functions, and (r, θ, λ) is the conventional geocentric spherical-polar coordinate system. The numerical Gauss coefficients g_n^m and h_n^m have the dimensions of magnetic field; conventionally they are expressed in units of nT. Estimates of these coefficients were produced by a weighted least-squares fit to measurements taken by the Ørsted satellite. The fitting process also produced an output covariance matrix as an estimate of the variances/covariances of the coefficient estimates. In the past such 'internal' estimates of uncertainty have often proved in retrospect to be significantly too small, and the aim of this paper is to produce a more realistic set of variances, and to show that the differences are due to the breakdown of various assumptions lying behind the least-squares process.

1.1 The least-squares process

A linear least-squares problem can be stated in the form

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{m} + \varepsilon, \tag{1.2}$$

where $\boldsymbol{y}$ is the vector of observations, $\mathbf{A}$ the known matrix which relates the observations to $\boldsymbol{m}$, the vector of coefficients to be estimated, and ε is the unknown vector of observation errors. For the time being we assume that the ε are purely random errors, i.e. having expectation zero. If all the observations are assumed to have the same variance (expectation of the mean-square random error), and if the observations have no covariances (no correlation of the errors from point to point), then the least-squares solution $\hat{\boldsymbol{m}}$ for $\boldsymbol{m}$ is obtained by minimizing the sum of squared residuals

$$S = (\boldsymbol{y} - \mathbf{A}\boldsymbol{m})^{\mathrm{T}}(\boldsymbol{y} - \mathbf{A}\boldsymbol{m}) \tag{1.3}$$

with respect to the set of coefficients $\boldsymbol{m}$, giving the estimate

$$\hat{\boldsymbol{m}} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\boldsymbol{y}. \tag{1.4}$$

This estimate will not be exact; if the variance of each observation is σ^2, then the covariance matrix for $\hat{\boldsymbol{m}}$ is

$$\mathbf{C} = \mathrm{Cov}[\hat{\boldsymbol{m}}] = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\sigma^2; \tag{1.5}$$

the diagonal terms of $\mathbf{C}$ are the variances of the coefficient estimates, and the off-diagonal terms are the appropriate covariances. The matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ incorporates all the relevant information about the nature of the data distribution and the fitting functions. For a given form of data distribution, the larger the number of observation points the larger will be $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, and hence the smaller the output (co)variances.

If, as is usually the case, we do not know σ^2 beforehand, we have to estimate it from the minimum value $S_{\min}$ of the sum of squared residuals, (1.3); an unbiased estimate is

$$\hat{\sigma}^2 = S_{\min}/(N - P), \tag{1.6}$$

where N is the number of observations and P the number of parameters estimated. Note that while the magnitude of $\mathbf{C}$ does depend on the actual values of the data through (1.6), its structure is given by the matrix $\mathbf{A}$, which depends only on the values of the fitting functions at the observation points, i.e. on the distribution of the data points, and not on the actual values of the data.

In our case the observations y_i are measurements of various components of magnetic field, the m_j are the Gauss coefficients of (1.1), and the elements A_{ij} are the appropriate partial derivatives of the spherical harmonics. If these fitting functions happen to be orthogonal over the data set, then $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, and hence cov[$\hat{\boldsymbol{m}}$], will be diagonal; each parameter can be estimated independently of the other parameters. But usually this is not the case; the value obtained for one parameter will depend on which other parameters are determined, and there will be covariances between the parameter estimates. In this case, if we wish to estimate the uncertainty of the field (or other function of the coefficients) as predicted at a given position by the model, we need to use the full covariance matrix. Fortunately, if we only wish to estimate the mean-square uncertainties of the vector field over a sphere about the origin, then, because the vector spherical harmonics are orthogonal over the sphere, we can simply add the mean-square field contributions given separately by the variance of each harmonic—the effects of the covariances cancel.

Throughout this paper, variances, or their equivalent, will usually be compared in terms of $(n + 1)\sigma^2$, their contribution to the mean-square vector-field uncertainty over the sphere. We will use the abbreviation ms for mean-square, and refer to the radius $r = a$ as sea level.

More usually, different (types of) data have different variances, and we need to weight the observations accordingly; there might also be correlation of the data errors from point to point. The data variances and covariances can be specified in terms of a data covariance matrix $\mathbf{V} = \mathrm{Expectation}[\varepsilon\varepsilon^{\mathrm{T}}]$; we can work either directly in terms of the weight matrix $\mathbf{V}^{-1}$, or indirectly in terms of a relative weight matrix $\mathbf{U}^{-1}$, where $\mathbf{V} = \mathbf{U}\sigma_0^2$ and σ_0^2 is the appropriate scaling factor. (The weight matrix $\mathbf{V}^{-1}$ (or $\mathbf{U}^{-1}$) might be imposed from the outset, or estimated iteratively; for the OSVM model both approaches were combined.) We now need to minimize

$$S = (\boldsymbol{y} - \mathbf{A}\boldsymbol{m})^{\mathrm{T}}\mathbf{U}^{-1}(\boldsymbol{y} - \mathbf{A}\boldsymbol{m}) \tag{1.7}$$

or, equivalently, the non-dimensional

$$S^* = (\boldsymbol{y} - \mathbf{A}\boldsymbol{m})^{\mathrm{T}}\mathbf{V}^{-1}(\boldsymbol{y} - \mathbf{A}\boldsymbol{m}), \tag{1.8}$$

for which the least-squares solution is

$$\hat{\boldsymbol{m}} = (\mathbf{A}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{U}^{-1}\boldsymbol{y} = (\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\boldsymbol{y}, \tag{1.9}$$

and the output covariance matrix is now

$$\mathbf{C} = \mathrm{Cov}[\hat{\boldsymbol{m}}] = (\mathbf{A}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{A})^{-1} = (\mathbf{A}^{\mathrm{T}}\mathbf{U}^{-1}\mathbf{A})^{-1}\sigma_0^2. \tag{1.10}$$

Again, we usually have to estimate σ_0^2 from the minimum value of (1.7), by

$$\hat{\sigma}_0^2 = S_{\min}/(N - P) = (\boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{m}})^{\mathrm{T}}\mathbf{U}^{-1}(\boldsymbol{y} - \mathbf{A}\hat{\boldsymbol{m}})/(N - P). \tag{1.11}$$

Because of the iterative re-weighting used in the OSVM modelling, the structure of $\mathbf{C}$ does now depend on the data values, but probably only to a small extent.

However, such 'internal' estimates of uncertainty (based on the residuals to the model) are subject to error. In fact the expressions (1.9) and (1.10) are only correct if three assumptions are valid:

(a) in the input relative covariance matrix $\mathbf{U}$, the relative variances and covariances are appropriate to the different types of data,

(b) the magnitudes of the data errors ε are properly estimated, i.e. σ_0^2 is estimated correctly, and

(c) the model is adequate, in that it solves for all significant non-random contributions to the measured quantities.

With regard to assumption (a), the use of iterative re-weighting in the Ørsted modelling should mean that the relative weighting of the different types of data is about right.

Because of the great computational cost of having to work with a non-diagonal error covariance matrix, it is usual to put to zero all the off-diagonal terms, i.e. all the data covariances. This is equivalent to assuming that there is no correlation between the errors at different points (or correlation between the errors of different components at one point); if this simplification is not justified then the solution $\hat{\boldsymbol{m}}$ we obtain will not be optimal, and our estimate of its uncertainty will also be wrong. Because of the anisotropic properties of the star camera used to determine the attitude of the Ørsted vector magnetometer, when the observed field vector is rotated from the satellite coordinate frame to a ground-based frame, correlated errors are in fact introduced into the magnetic field vector (X, Y, Z) triplet at each point. However, Holme & Bloxham (1996) have shown how the appropriate modification can be introduced into an otherwise diagonal $\mathbf{V}$ with little extra computation, and this was done in this analysis. But, as usual, for computational simplicity it was still assumed that there was no correlation between the data errors at successive locations.

The simple assumption made at the beginning of this section, that the ε_i of (1.2) are random and uncorrelated errors, is usually

valid for instrumental noise. But in practice there are other, often dominant, contributions to the observed residuals, and these can introduce significant covariances in the input data. For example, while at sea level the field from crustal magnetization mostly has correlation lengths much shorter than the spacing of the OSVM data, at satellite level there can be significant correlation over 5° to 10°.

Also, while we are trying to estimate the parameters of some model, this model is inevitably limited, so that our observations can include 'unfitted fields' from sources which are not being modelled. Depending on the geometry of the data set, some of the 'power' of these unfitted fields will 'leak', or be 'aliased', into the fitted parameters, so that the estimates obtained will be biased estimates of these parameters. The remainder of the power of the unfitted fields is left in the residuals, so that (1.11) might not be a good basis on which to estimate σ_0.

Langel *et al.* (1989) made some preliminary investigations to quantify some of these effects, by treating statistically various unfitted internal fields, such as the field from the crust, and the field from the ionosphere (which was 'internal' to the Magsat satellite). Simulations using synthetic (field + noise) 'data' on a small grid of about 150 vector data points showed that using the full covariance matrix did improve the fit to the original field, and gave more realistic, larger, estimates of coefficient variance. Similarly, for a sparse sample of Magsat data, 400 data points with a separation of about 10°, using the full covariance matrix gave more realistic, larger, estimates of coefficient variance.

However, when these authors re-analysed the GSFC(12/83) data set they were able to use only the diagonal terms of their 'inverse correlated weight matrix'. This gave coefficients which were not significantly different from the original coefficients, but coefficient variances which were considerably larger. But this increase was essentially all due to treating the coefficients as estimates of the core field, and therefore, in effect, adding to the output covariances their statistical estimates of the mean-square coefficients of the crustal and ionospheric fields.

Rygaard-Hjalsted *et al.* (1997) made a similar trial of the effect of crustal field, but using a model of crustal magnetization having a much shorter correlation length, about 5° at 400 km, than the tens of degrees of the model used by Langel *et al.* Their covariance matrix was therefore sparser, and they were able to use a larger subset of Magsat data, about 1600 B_r points at 5° spacing. Applying a regularized least-squares analysis, they found that using the full covariance matrix led to a smoother field model, with a reduction in the ms field from about $n = 10$ upwards; however, it is not clear if this would have happened without the regularization. They did not produce the corresponding coefficient covariance matrix. Note that in both simulations the whole of the crustal field, down to $n = 1$, was included when estimating the correlation length. If the crustal contribution were restricted to higher degree harmonics (above those solved for), the correlation length would be less.

In the present context note that in both simulations the unfitted fields were essentially time-independent, and the along-track correlation was due entirely to the spatial variation of the unfitted fields, which were specified only statistically.

It is the purpose of this paper to discuss two ways in which the errors in the assumptions (a) and (c) led to seriously wrong output coefficient variances in the OSVM modelling. Specifically, we find that

(i) significant along-track correlation of the residuals in time due to magnetospheric 'noise' can redistribute the effect of the noise among the output variances, and

(ii) the 'leakage' of unfitted ionospheric field gives a large systematic error; slightly different sampling of this systematic error leads to the increase in variance of some main-field coefficients.

1.2 The organization of this paper

In Section 2 we briefly describe the data used in the modelling, and explain how this data distribution accounts for the main properties of the formal output covariance matrix $\mathbf{C}$. In Section 3 we describe how other 'objective' estimates were made of the variances by comparing the results of analysing four independent subsets of the data. Comparing these objective estimates of the variance of individual coefficients with those of the formal output covariance matrix showed that major corrections were needed; in particular, while about half of the variances were overestimated by the formal matrix, this did significantly underestimate the variances of sectorial ($m = n$) and near-sectorial coefficients, and of zonal (m = 0) and near-zonal coefficients. In Section 4 we discuss the problem of assumption (a) above, and use Monte Carlo simulations to show how serial correlation of the data errors can lead to the observed increase of variance in the sectorial and near-sectorial coefficients, and to a reduction elsewhere. Then in Section 5 we discuss the way in which various types of field contributions, which are not fitted in the OSVM modelling, can affect certain coefficients, through failure of assumption (c); we find that unmodelled ionospheric field on average gives a large systematic error in near-zonal coefficients, and that slight differences in the averaging leads to the increased variances in these coefficients. Then in Section 6 we use these arguments to justify the use of a simple analytic form of a correction factor, to be applied to the formal output covariance matrix elements to give a much more realistic matrix. Section 7 summarizes our OSVM results, and Section 8 discusses the more general application of our results.

2 ØRSTED DATA AND ANALYSIS

The Ørsted satellite is in a near-polar orbit at an altitude of about 760 km (perigee 640, apogee 880 km). For night times on magnetically quiet days, data points were taken about 1 minute/sin θ (about 3.6°/sin θ) apart, so as to give a distribution roughly uniform over the sphere. Equatorward of 50° geomagnetic latitude, vector data $\boldsymbol{B}$ were used; scalar (field intensity) F data were used in the polar regions, and to fill in gaps in the coverage of the vector data. In total there were about 25 000 vector triplets and about 45 000 scalar values.

The analysis was to $n = 29$ for the field of internal origin (899 coefficients), together with linear secular variation to $n = 13$ (195 coefficients). External fields were approximated by an $n \leq 2$ model, allowing for a linear variation with (a modified version of) the *Dst* index, together with the appropriate internal induced field (11 coefficients); there were also independent annual and semi-annual variations of the zonal $n = 1, 2$ internal and external terms (16 coefficients). The data covariance matrix used was block diagonal, with covariances within each vector triplet to allow for the anisotropic accuracy of the attitude determination (Holme & Bloxham 1996). Different types of data were weighted differently, the weights being changed iteratively. This, together with the use of intensity data, meant that the problem was non-linear, but the final iteration was effectively of a linear model. The magnitude of the weight matrix $\mathbf{V}^{-1}$ (effectively the magnitude of σ_0^2) was estimated from the sum-square residual of the final iteration. No *a-priori* model, or other form of smoothing, was used.

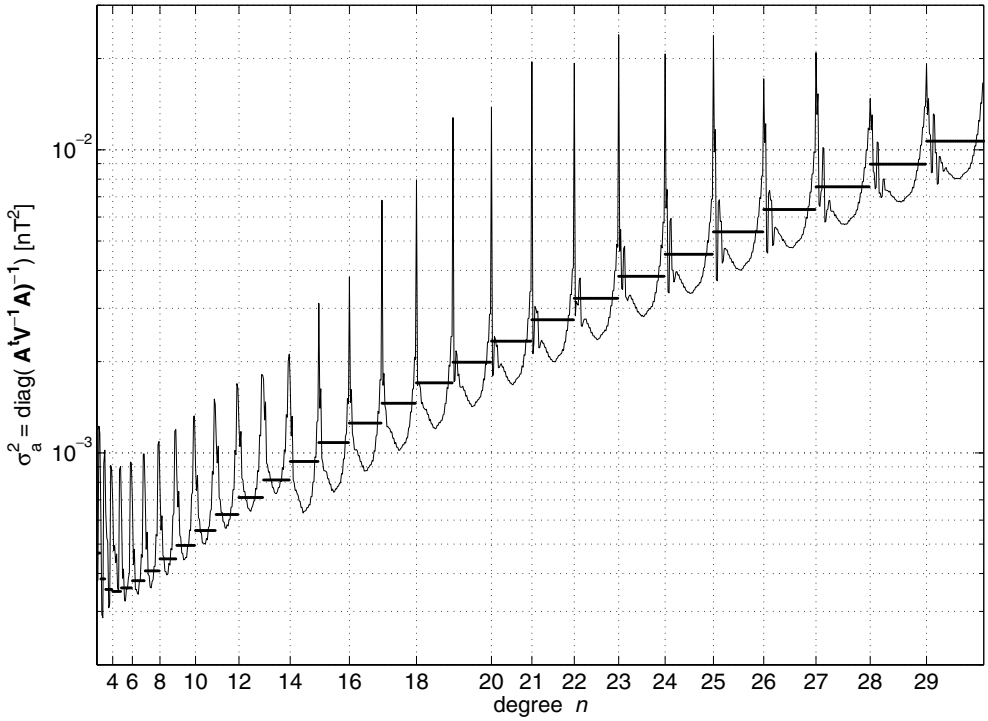

Figure 1. Individual main-field variances of the Ørsted model (09a/01). (The order is g_1^0, g_1^1, h_1^1, g_2^0, etc.) The horizontal bars are what would be expected for a uniformly distributed set of vector data at altitude 700 km.

For a full description of the data and the modelling see Olsen (2002). That paper presents the model known as OSVM. Some of the simulations described in the present paper used a slightly earlier model, Ørsted (09a/01), but there are only minor differences in the data used in the two models, so the results of the simulations are applied to the covariance matrix of the OSVM model.

The 899 formal variances for the main-field coefficients are shown as one continuous sequence in Fig. 1. For N vector triplets, uniformly distributed over a sphere of radius a, and having the same variance s^2 for each component, for a given degree n the resulting coefficient variances would be constant at $s^2/N(n+1)$ (see e.g. Langel 1987, p. 356). However, while the coefficients are expressed as at 'sea level', radius a, the Ørsted observations were at a larger radius r. The horizontal bars on Fig. 1 are $(r/a)^{2n+4}$ (3 nT)^2/18 000$(n+1)$, corresponding to uniform vector coverage at satellite altitude (the 3 nT and 18 000 observations are arbitrary figures), but clearly the actual variation is much more complicated. Within each degree n, the variances are large at $m = 0$, decrease to one or more minima, and then increase to another maximum at $m = n$.

The main features of this variation with order m can be explained in terms of the data distribution. For a uniform distribution of purely intensity data it is well known (see e.g. Lowes 1975) that the resultant variances increase monotonically with m for given n, with the sectorial, $m = n$, coefficients having the largest errors. (The sectorial harmonics have alternating signs in longitudinal sectors, with the latitudinal variation being more concentrated near the equator as n increases.) Even if there are also some vector data, we would expect a data set dominated by scalar observations to have this behaviour.

In contrast, a uniform distribution of purely vector data would give variances independent of m. But in the OSVM case there are no vector data in the polar regions; in effect the vector data are concentrated towards the equator. The least-squares process is such that it is the harmonics which are sampled most where their field is greatest which are best determined; for equatorial vector data these are precisely the sectorial harmonics which are poorly determined by scalar data. So a data set consisting of equatorially concentrated vector data will give coefficient variances which decrease monotonically with m for a given n; this has been confirmed by simulation. When the omitted polar cap is geomagnetic rather than geographic, there are also minor asymmetries between the g_n^m and h_n^m variances.

In the case of Ørsted, which has both polar concentration of scalar data and equatorial concentration of vector data, we can think of these two effects competing and so giving the observed pattern. This also has been confirmed by simulation.

In fact the Holme & Bloxham (1996) weighting of the vector data in effect gives the intensity F at the vector data points as accurately as at the scalar data points, together with two, relatively less accurate, angles defining its direction. So the OSVM data set can be thought of as a global F data set, which by itself would give variances increasing with m. The addition of equatorially concentrated directional data adds a secondary maximum at $m = 0$, and therefore a minimum at some intermediate value of m.

Note that this explanation of the form of Fig. 1 assumes, as did the least-squares analysis, that the measurements were subject only to random uncorrelated noise, and that the underlying model was adequate. We will see in Sections 4 and 5 that neither assumption was justified.

3 OBJECTIVE ESTIMATE OF VARIANCES

Ideally, the best method of validating a model is to compare it with good independent data not used in preparing the model. However, good independent data are scarce, and modellers like to use all available data, which leaves none for comparison. The 1-min sampling

rate used for the OSVM model was intended as a compromise between minimizing serial correlation and losing data.

Another approach is to model independent subsets of the data, and see how the resultant models differ. Cain *et al.* (1967) tried this with their 1900–1965 data set, but were able to produce only two subsets, which overlapped by 40 per cent. They did not give the least-squares estimates of coefficient variances, but stated that these were commensurate with the differences between the coefficients obtained from the two subsets.

We applied this approach to the data set used for Ørsted (09a/01). This was divided into four subsets, corresponding to every UT calendar first, second, third, and fourth day, and exactly the same analysis technique applied to each as was used for the model 09a. (In theory, the greater the number of subdivisions of the data, the more accurately can the output variance given by the whole data set be estimated. But smaller subsets are less likely to be uniformly distributed, and the differences in distribution could themselves lead to changes in the coefficients.)

To the extent that these four data subsets have independent errors, the differences between the resulting four models can be used to give unbiased estimates of the variances of the 09a model itself, which is based on the whole data set. We believe that these estimates will be more realistic than those of the 09a covariance matrix. However, the division cannot tell us anything about the presence of any systematic errors in the coefficients.

With four data sets there are only three independent differences, but a simple way to give equal weight to each of the four models is to estimate the variance of (each coefficient of) the 09a model by taking one-eighth of the mean-square of all six possible differences. This estimate will be called σ_k^2, while the value given by the 09a output covariance matrix will be called σ_a^2. This σ_a^2 is based on a large number of data and so has the detailed structure imposed by the data distribution, as shown in Fig. 1. On the other hand, our individual estimates σ_k^2 are based on only three degrees of freedom, and so are too scattered to show detailed structure. Despite the scatter, it is clear that the σ_k^2 show an even larger systematic variation with n and m than do the σ_a^2; for ease of comparison we will work with the variance ratios σ_k^2/σ_a^2. Our hope is to find some simple, overall, smooth relation between the two sets of variances which we can estimate, and then use to correct the 09a variances; in this way we can keep the detailed structure of the 09a variance vector, while scaling it to the more realistic magnitudes of our more objective estimates. We will then apply the same correction to the very similar OSVM variance vector.

3.1 Main field

For the $n = 1$–29 main-field coefficients, Fig. 2 shows the ratio σ_k^2/σ_a^2 as a function of n and m; to make the figure more legible, and to reduce the scatter, the mean of the corresponding g_n^m and h_n^m ratios has been used. Despite the scatter, we can see that many values are less than unity, and that for a given degree n the ratio tends to increase rapidly as m approaches 0 (zonal harmonics), and also to increase as m approaches n (sectorial harmonics). In both cases the increase is larger for smaller n.

We can think of no reason why our subdivision should introduce such a pattern, and believe that σ_k^2 is (apart from the inherent scatter) essentially a good estimate of the real variances of the model

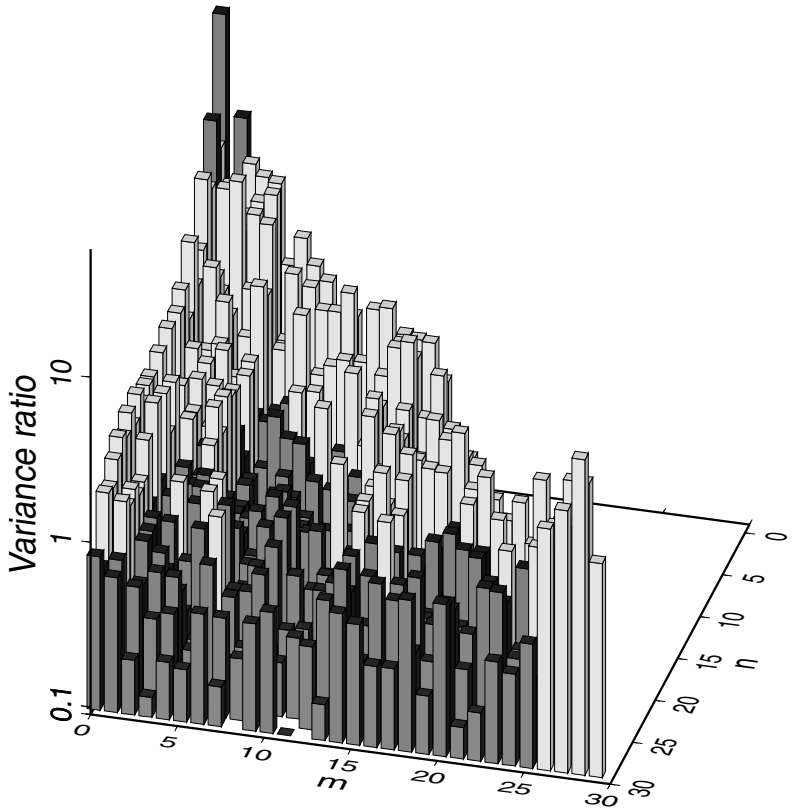

Figure 2. The ratio σ_k^2/σ_a^2 for the main field as a function of n and m; σ_k^2 is the estimate of the variance of the Ørsted model (09a/01) obtained by dividing the data into four subsets, and σ_a^2 is the value given by the 09a covariance matrix. To make the figure more legible, and to reduce the scatter, the mean of the corresponding g_n^m and h_n^m ratios has been used. Ratios greater than 10.0 or less than 1.0 are shown darker.

09a coefficients. Recall that while σ_a^2 itself has local maxima at $m = 0$ and $m = n$, our σ_k^2 show this effect even more markedly, particularly for the smaller degrees n. We conclude that we are seeing a real effect, which is not fully incorporated in the 09a formal variances.

We therefore try to understand why, for the ratio σ_k^2/σ_a^2,

(i) there is a rapid increase of the ratio as m approaches the value 0,

(ii) there is a (less rapid) increase of the ratio as m approaches the value n, and

(iii) for harmonics away from these limits, the variance is only about a third of that predicted by the 09a covariance matrix.

In Section 4 we suggest an explanation for (ii) and (iii), and in Section 5 for (i).

3.2 Secular variation

Fig. 3 shows the ratio σ_k^2/σ_a^2 as a function of n, m for the $n = 1$–13 secular variation coefficients. Again this is larger for the near-sectorial terms; however, there is no suggestion of any $m = 0$ enhancement.

4 THE EFFECT OF SERIAL CORRELATION IN THE DATA NOISE

In this and later sections we use the generic term 'noise' to represent the sum of instrumental noise and all sources of field other than those being modelled.

Even though the data used in the Ørsted (09a/01) and OSVM modelling come from magnetically quiet times, and an attempt has been made to allow for the large-scale magnetospheric contributions described by hourly mean values of the *Dst* index, it is likely that a significant part of the noise comes from other magnetospheric, sources. These fields often have timescales of minutes to hours, significantly longer than the about 1 min between data points. Therefore this contribution to the noise will not be independent from one data point to the next, and we would expect there to be serial correlation in the noise.

We checked this by looking at the scalar residuals (scalar data minus OSVM model values), selecting data segments of at least 30-min length (797 segments), and calculating the along-track autocorrelation. Fig. 4 shows the results, and confirms that there is in fact significant serial correlation, so that assumption (b) of Section 1 is not valid. (The 'damped sinusoidal' shape of the curve is puzzling; it is consistent with the presence of pseudo-periodic forcing, for example from applying random noise to a system which has a typical 'period' 10–15 min. At present we do not know the nature of the noise source in this 'quiet day' situation. However, it is clear that the effect is not produced by the satellite, as the same result was obtained for residuals from an analysis of scalar data from the quite different CHAMP satellite.)

In this section we try to characterize the likely effect of serial correlation, first by a qualitative discussion, and then by quantitative simulations using simple models of serial correlation.

4.1 Qualitative discussion of the effect of serial correlation

We argue that if there is serial correlation of the errors this will tend to increase the noise power in the sectorial and near-sectorial coefficients, and decrease it for other harmonics.

In the Appendix we show algebraically that, for Fourier analysis of measurements round a circle, for a given total mean-square noise

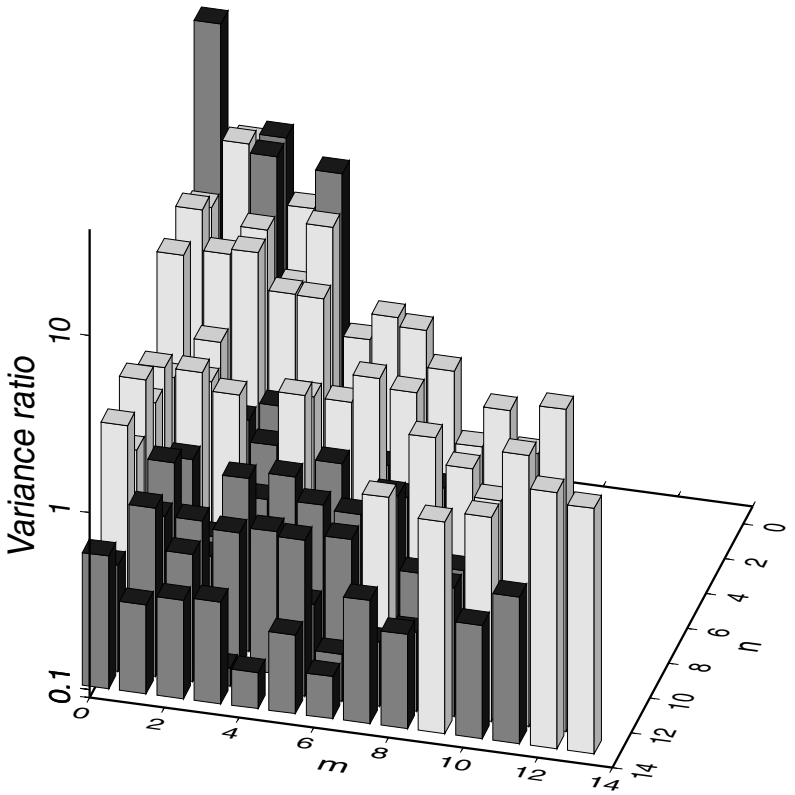

Figure 3. As for Fig. 2, but for the secular variation.

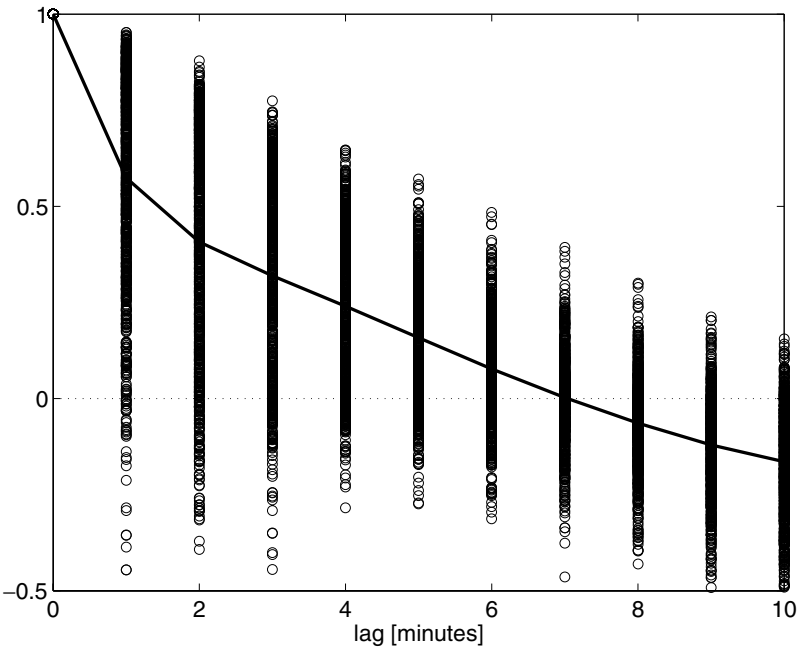

Figure 4. Autocorrelation of the F residuals. For each of 797 segments of F residuals (scalar data minus OSVM values), of length greater than 30 min, the autocorrelation was calculated. The line shows the average; the median is almost the same.

at each measurement point, the effect of increasing the serial correlation of the noise between points is to transfer output noise power from high frequencies to low frequencies. Applying this result to the near-polar orbit plane of our satellite, along-track corresponds to variation with latitude. With an orbit period of at least 100 min, for correlation times of 5–10 min there will be no correlation from orbit to orbit, i.e. in longitude. Therefore in terms of longitudinal analysis, this noise is essentially random, and will affect all longitudinal frequencies equally. The associated Legendre functions we are using in our surface harmonics are more complicated than simple great-circle sinusoids, but it seems clear that it will be the polynomials which on average have the slowest variation with latitude which will be most affected. $P_n^m(\cos\,\theta)$ has $(n-m)$ zeros between the two poles, so we can expect that it is the sectorial, $m = n$, harmonics which will be most affected. For large n the sectorial harmonics are large only near the equator, and have comparatively large gradients there; this probably explains the $1/n$ dependence we find later.

4.2 Simulation of the effect of serial correlation

To try to see what the actual effect was, we did a series of simulations. For simplicity we used global vector data, and made spherical harmonic analyses of a series of data sets in which each of the (X, Y, Z) 'data' were purely noise values of appropriate type, drawn from a serially correlated distribution having zero mean, and variance 1000^2 nT^2. The points were a constant 10.1° apart along an Ørsted-type orbit, but $\sin\theta$ weighting was used in the analysis to simulate the equal-area spacing of the selected Ørsted data. Each data set, consisting of 2000 points, was subject to least-squares spherical-harmonic analysis; the resulting coefficients were squared and multiplied by $(n+1)$ to give the mean-square vector field per harmonic. These mean-square fields were then averaged over the results from 10 or 20 data sets. Two noise types were used, as described in the Appendix:

(i) Overlapping noise with $p = 4$ (autocorrelation decreasing to zero at about 40°; Fourier power spectrum reducing to a half at 4 cycles/circle).

(ii) First-order autoregressive noise with $\beta = 0.6$, 0.7 and 0.9 (autocorrelation decreasing to 0.1 at 45°, 65° and 220°; Fourier power spectrum reducing to a half at 3, 2 or 0.6 cycles/circle).

A uniform distribution of vector 'data' which consisted of independent random noise would be expected to give, on average, coefficient noise corresponding to a constant mean-square field per harmonic. Our 2000, $\sin\theta$-weighted, data correspond to about 1300 data of unit weight, so an input noise of 1000^2 nT^2 per component would be expected to give a mean-square field per harmonic of about 750 nT^2. Even in the results for individual sets, for all types of correlated noise there was a strong suggestion that the mean-square field per coefficient had a base noise level significantly below this figure, to which was added a rapid rise for harmonics at and near $m = n$. This trend is clearly shown in Fig. 5, which shows the average result for 20 repetitions of the first-order autoregressive $\beta = 0.6$ case.

Averaging over all the relevant terms of Fig. 5 to give the mean-square field as a function of $(n-m)$ gives Fig. 6, which shows a base level, and a monotonic increase as $(n-m)$ decreases to the sectorial terms. (The increase in the simulation values for large $(n-m)$ is ignored as being due to the 10° spacing not being small enough to properly sample these short latitudinal wavelengths.) Various least-squares fits were tried, and it was found that the simple function

$$\text{ms field per harmonic} = a + b\exp[-(n-m)/d] \qquad (4.1)$$

gave a good fit to the $(n-m) \leq 25$ terms, as also shown on Fig. 6. However, we saw in Section 3 that the enhancement of the sectorial terms was larger for smaller degrees. Changing (4.1) to

$$\text{ms field per harmonic} = a + (b + c/n)\exp[-(n-m)/d] \qquad (4.2)$$

gave a better fit when applied to the, much more scattered, 675 individual ratios for the $(n-m) \leq 25$ g_n^m and h_n^m terms. The form (4.2) will be used in the scaling of Section 6.

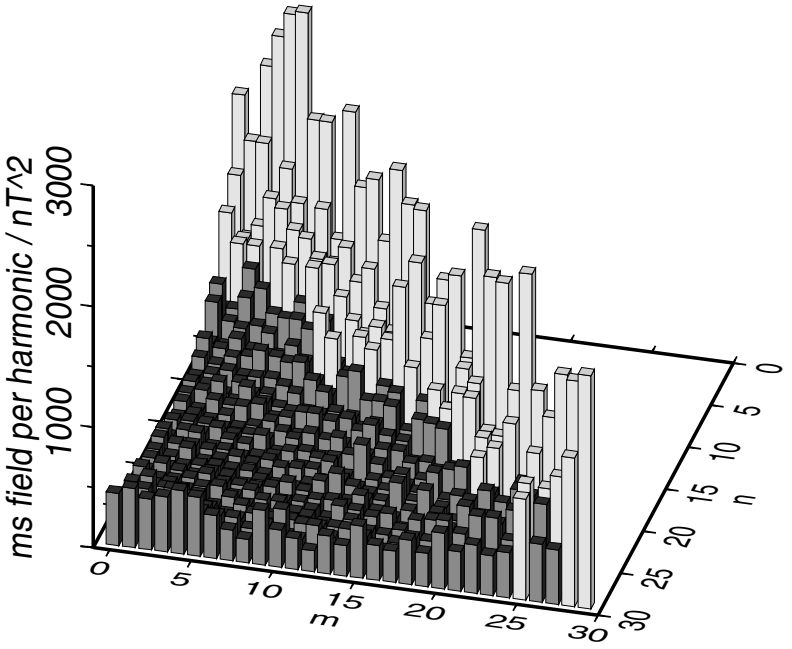

Figure 5. Linear plot of the mean-square (ms) vector fields corresponding to the variances produced by analysis of a satellite-type data set subject to serially correlated noise having rms value of 1000 nT, and an autocorrelation of 0.6 between successive data points. The figure shows the average of 20 repetitions of the analysis of 2000 points 10.1° apart in latitude. Values less than 750 nT^2, the expected value for uncorrelated noise, are shown darker.

We found a similar situation whichever model of serially correlated noise was used, with only the numerical values of the coefficients changing. A trial using a more realistic, combined vector and scalar, data set gave the same basic result.

Our simulations show that the presence of serial correlation can indeed produce the enhancement of the variances of the sectorial terms which is so obvious in the values of the σ_k^2/σ_a^2 ratio. They also show that for the other harmonics the variances can be less than they would be if the errors were independent. But these simulations did not show the enhancement of the zonal terms noted in Section 3; we suggest a reason for this zonal enhancement in the next section.

5 THE EFFECT OF UNMODELLED FIELDS

5.1 Introduction

So far we have discussed the effect of serial correlation of otherwise random fluctuations in the data, coming from magnetospheric 'noise'; this changes the variances of the resulting coefficients. However, there might also be (essentially) systematic contributions in the data 'noise'. These arise because the Ørsted OSVM model (and the very similar 09a model) is limited in its scope, so various known

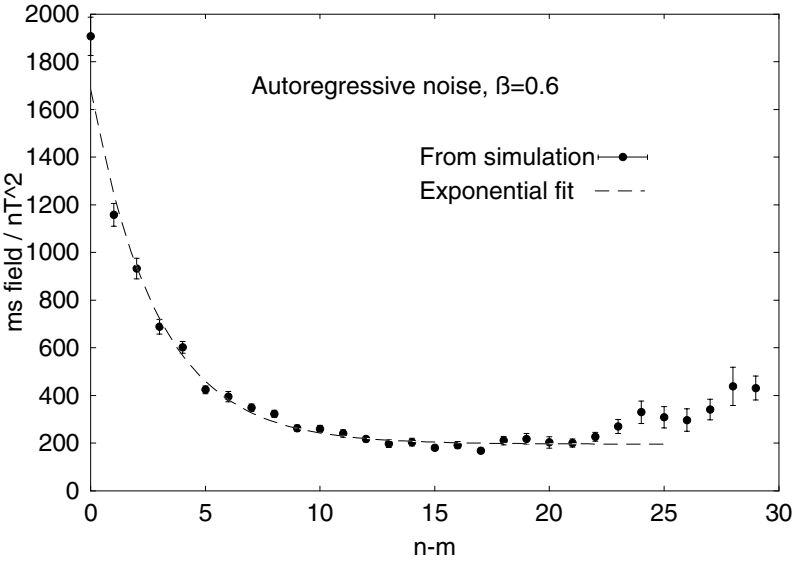

Figure 6. The values of Fig. 5 averaged as a function of $(n - m)$. (For $(n - m) = 29$ there is only one value; the standard deviation is a nominal value.) For $n \leq 25$ the (weighted) points are least-squares fitted by the exponential $a + b \exp[-(n - m)/d]$, shown by the dashed line.

regular contributions to the observed field are not modelled. We will call these unmodelled contributions 'confounding' fields.

If these unmodelled confounding fields were orthogonal to the fitted fields over the data points they would not change the values of the fitted coefficients. They would, however, increase the magnitude of the residuals, and hence our estimate of σ_0^2, even though they had had no effect on the accuracy of the fitted coefficients.

Even if these confounding fields were orthogonal over the surface to the fitted fields at any one time (UT), they are mostly Sun-based, and so vary with local time. But, while our data are distributed fairly uniformly in space, they deliberately sample only a very restricted range of local time. So in practice there is no orthogonality, and the unmodelled fields will change some of the fitted coefficients; the unmodelled field 'leaks' some of its 'power' mean-square field into some of the coefficients of the modelled field, making such coefficients more uncertain than is indicated by the OSVM covariance matrix.

Such leakage is essentially systematic, resulting in coefficient values which are biased away from the true value, though of course we might not know in what direction. If the unmodelled field is fixed, then for our particular Earth, strictly speaking, there is no effect on the variance of our estimate, which is a measure of the scatter about the mean. Sometimes, however, it is convenient to produce an overall estimate of uncertainty, random plus systematic, by adding to the variance the square of an estimate of the magnitude of the bias, to give a total ms deviation about the true value. This overall value is also sometimes called a variance, though one could argue that the term would only be valid if one were thinking of sampling a population of earths having different crustal fields etc. This is what Langel *et al.* (1989) did, using statistical estimates for various unmodelled fields, and showing how the 'variances' of the coefficients were increased.

In this paper we prefer to use 'variance' in its more restricted sense. Using a deterministic model of the ionosphere we estimate the resulting (quite large) average systematic bias of the coefficient estimates. As different data sets take slightly different averages, there is a also scatter about the average bias, leading to an increased variance of these coefficient estimates.

5.2 Estimation of the leakage from unmodelled fields

To estimate the magnitude of these effects, other simulations were done. At the positions and times of the OSVM data set, the appropriate field ($\boldsymbol{B}$ or F) of the internal part of the OSVM model was calculated. At the same positions and times we used the much more detailed CM3 model of Sabaka *et al.* (2002) to calculate realistic values of the contributions of each of four confounding fields, three of which had not been solved for in the OSVM model. For each of the four cases the confounding field was added to the OSVM synthetic data, and the OSVM solution technique was then applied to the sum field. The original OSVM internal coefficients were then subtracted from the resulting model coefficients. These differences, together with the ms values of the confounding field contributions and the residuals, let us estimate how much of each confounding field leaked through into the solution and how much remained unfitted in the residuals.

These four confounding fields were: crustal fields of higher degree ($n = 30$–45); ionospheric and magnetospheric fields, and their Earth-induced counterparts; and large-scale toroidal non-potential fields. The mean-square values of the confounding field, and of the residual after analysis, were determined for the actual data points, which covered a range of radii and times. The effect on the solution was estimated by subtracting the OSVM solution from the (OSVM + confounding field) solution, and using the resulting coefficient differences to give the ms field of the pseudo-internal difference field at the typical height of 760 km, as well as at sea level. The results are summarized in Table 1. Note that while the CM3 line gives the magnitudes of the actual confounding fields, which in one case has primary source external to the satellite, the last two lines give the magnitudes of the changes in the solved-for internal field produced by the 'leakage' of the confounding fields into the solution. In the case of the magnetospheric field, almost all of the CM3 contribution ended in the $n = 1$ external terms (giving a uniform field over the Earth), not shown in the table but having mean-square value 526 nT^2.

To put the 'residual' figures of the table into context, the unweighted mean-square vector residual in the OSVM analysis itself was about 57 nT^2, so, except for the toroidal field, the residuals in these simulations look reasonable; it is argued below that the CM3 toroidal field we used is much larger than would be relevant for the Ørsted measurement times. To put the last, sea level, line of the table into context, the total mean-square vector field corresponding to the OSVM output covariance matrix was about 100 nT^2; ignoring the toroidal field, the powers leaked by these confounding fields are comparable with, or smaller than, this total power of the OSVM covariance matrix. More important here is the way this leakage power is spread unevenly among the harmonic coefficients, and that it is effectively systematic, rather than random.

In each case the proportion of the confounding field which leaks into the least-squares solution at satellite altitude is considerably more than would be expected from a degrees of freedom argument based on the assumption that the confounding field was high-frequency noise. So applying this degrees of freedom argument to the original OSVM residuals would have underestimated the σ_0^2 scaling factor to be applied to the OSVM relative output covariance matrix. The discussion below indicates that, particularly in the case of the ionospheric field, the increased uncertainty is not spread uniformly but is concentrated into particular harmonics, so a simple increase of σ_0^2 would be misleading.

Table 1. Leakage of confounding fields into a model OSVM type solution. Simulation of the effect of four separate confounding fields on the accuracy of the OSVM model. All values are mean-square vector fields, expressed in nT^2, averaged as shown in the last column. 'CM3' refers to the individual confounding fields, 'Residual' to the residuals left after that least-squares fit, and the Δg are the differences in the coefficients (both g_n^m and h_n^m) produced by the presence of that confounding field. The last two rows refer to the full $n = 29$ solution.

Confounding field	Crust	Ionosphere	Magnetosphere	Toroidal	
CM3 field	0.15	20	520	220	Average over all actual points
Residual	0.12	4	13	180	Average over vector points
$\Sigma\,(n+1)\Delta g^2$	0.05	14	1	13	At 760 km
$\Sigma\,(n+1)\Delta g^2$	13.43	94	33	273	At sea level

Satellite observations essentially give the spherical harmonic coefficients at satellite radius r. When, as is the convention, we express the coefficients as at 'sea level', radius a, there is an $(r/a)^{n+1}$ enhancement in the coefficients, or an $(r/a)^{2n+4}$ enhancement in the mean-square field, which can greatly increase the contribution of high harmonics. At sea level, except for the ionospheric field, about half of the leakage mean-square field is from the highest degrees, $n =$ 26–29; this is presumably due partly to the truncation and partly to these harmonics not being properly resolved. However, the absolute effect of these high-degree harmonics is trivial at satellite altitude.

The effect of each of the four fields will now be discussed separately, and summarized in Section 5.7. In each case the way the coefficient differences varied with n and m were analysed as in Section 4.2.

5.3 The effect of unmodelled crustal field

At satellite altitude, the CM3 short-wavelength crustal field ($n =$ 30–45) is very small. When integrated over a sphere, these high-degree fields are orthogonal to the $n \leq 29$ fields of the OSVM model. However, there is not exact orthogonality when the integration is replaced by summation over a discrete set of points, even for uniformly spaced vector data (see e.g. Jekeli 1996 eq. 24), so there is some leakage into the solution. At satellite altitude the effect is insignificant in magnitude, but when extrapolated down to sea level, the amplified higher harmonics do contribute significantly, giving a total leakage of 13 nT^2. (In the OSVM solution these higher harmonics were included specifically to avoid truncation errors; they were not claimed to be accurate.)

The vast majority of this sea-level power is in the zonal, $m = 0$, harmonics; the reason for this is not known, but compared with the effect of the ionosphere (see below) its effect is trivial.

Except at the highest degrees, the changes in the Gauss coefficients are much smaller than those obtained by Rygaard-Hjalsted *et al.* (1997). This is because our simulation used a deterministic model of the crustal field, over a limited range of spatial frequencies, for a large data set at 760 km, while their trial used a statistical model of the whole crustal spectrum, and a much smaller data set at 400 km.

5.4 The effect of ionospheric fields

By using only night-time measurements, it had been hoped that most of the effects of the ionosphere would be eliminated, so no ionospheric terms were included in the 09a and OSVM models. However, the CM3 model shows there are still fields of a few nanotesla at satellite level at these times; these are produced by the currents induced inside the Earth by the time-varying ionospheric field. (Even though the ionospheric currents as such are zero or almost zero at these local times, the field of the corresponding internal induced currents is not zero; see e.g. Ashour & Price 1965).

Our simulation shows that these night-time (ionospheric + induced) fields can leak significantly into the OSVM model. At satellite altitude the leakage is almost entirely into g_1^0, g_3^0 and h_3^1; however, at sea level, while g_3^0 dominates at 21 nT^2, higher harmonics also contribute, giving a total of 94 nT^2. This leakage is concentrated in the zonal and near-zonal terms; see Fig. 7.

The CM3 ionospheric currents are scaled by the daily mean of the F10.7 solar flux, giving a first-order approximation to the day-to-day variations. But the real day-to-day variation is more complicated; also, different data sets sample the ionospheric field differently in space and time. So different data sets will produce a somewhat different average ionospheric leakage, leading to a pseudo-random scatter of the coefficients, and this is what was shown by the near-zonal coefficients of our four subsets of data. The

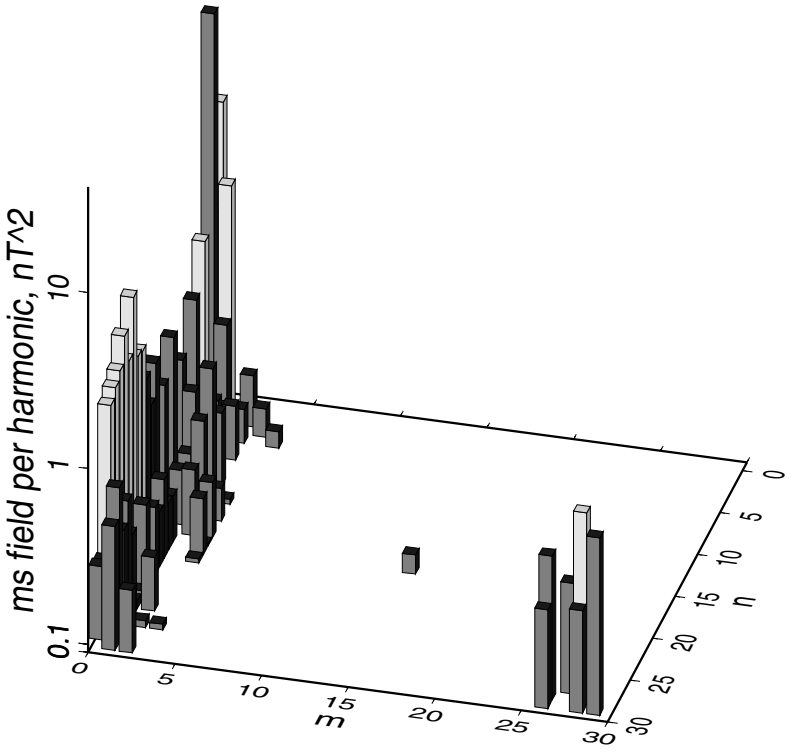

Figure 7. The mean-square field per harmonic leaking into an OSVM type solution from unmodelled ionospheric field; unit nT^2. Generally as for Fig. 2. Values greater than 10.0 or less than 1.0 are shown darker.

corresponding variances to be associated with the whole data set are only of the order of magnitude of 1 per cent of the squares of the systematic leakages, but are still many times larger than the formal variances.

By analogy with the fitting of the $m = n$ concentration found in the simulations of Section 4, we tried fitting the $m = 0$ concentration by the exponential

$$\text{ms field per harmonic} = a + e\exp(-m/g); \tag{5.1}$$

excluding the g_3^0 term, this fitted 35 per cent of the (ms field)^2 values. Allowing the coefficient of the exponent to increase with decreasing n, giving

$$\text{ms field per harmonic} = a + (e + f/n)\exp(-m/g), \tag{5.2}$$

fitted another 10 per cent. The form (5.2) will be used in the scaling of Section 6.

5.5 The effect of magnetospheric fields

The OSVM modelling did include magnetospheric terms, but this simulation was included for completeness.

As is to be expected, at satellite altitude the vast majority of the CM3 magnetospheric fields appears in the external $n = 1, 2$ terms, and there is only a trivial contribution to the main-field terms. At satellite altitude, most leakage into the main-field terms is to g_2^0, though with a small effect on h_2^1. At sea level there is relatively more power at the higher degrees, with a tendency to concentrate in the near-zonal and near-sectorial terms.

(Note that the CM3 magnetospheric model is based on hourly *Dst* values, modulated by seasonal and diurnal oscillations. It does not try to model the more rapid variations which we suggest are the reasons for the serial correlation we discussed in Section 4.)

5.6 The effect of toroidal fields

These fields come from large-scale ionospheric coupling currents, joining the northern and southern ionospheric current systems. They give a small leakage, mainly into low-degree terms, but continuing towards higher degrees mainly in the sectorial terms. At sea level, 222 of the total 273 nT^2 comes from $n = 28$ near-sectorial terms, which should probably be ignored, as being essentially a truncation problem. Unfortunately CM3 estimated these toroidal fields only for the dusk and dawn times of the Magsat observations. For this simulation we happened to use the larger dusk values, which will be many times too large for our night-time Ørsted data. It seems unlikely that there will be any significant leakage into the OSVM analysis.

5.7 Summary of the effect of unmodelled fields

These simulations show there is a large systematic leakage of the mean ionospheric field, mainly into the near-zonal coefficients, corresponding to a field of magnitude at least 6 nT. If the highest degree terms are ignored, the other three confounding fields probably only have a negligible effect, except possibly on g_2^0 by the magnetospheric field.

The enhancement of the variances of the near-zonal terms will come from the small variations of the leakage of unmodelled ionospheric field into the OSVM solution.

6 SCALING FACTOR FOR THE ØRSTED (09a/01) AND OSVM VARIANCES

We now want to estimate smooth scaling factors which are a reasonably good fit to the two sets of ratios, taking into account the results of Sections 4 and 5. Because the models 09a and OSVM are so similar, we will apply the scaling factors derived from the former to the variances of model OSVM.

6.1 Least-squares fitting of the ratio σ_k^2/σ_a^2 for the main field

6.1.1 n = 2–25 terms

A succession of least-squares fits were made to the 672 individual values of the ratio for $n = 2$–25; we excluded the $n = 1$, dipole, terms, which are discussed separately below. We excluded the $n > 25$ terms because they were noisy and not properly resolved. The results of successive fits are summarized in Table 2. Using the exponential expression (4.11) for the sectorial enhancement gave a fair fit, MF2, but using the modified expression (4.12) gave a substantial improvement, MF3. The exponential expression (5.1) for the enhancement of the zonal terms was added, MF4, and then its coefficient allowed to vary with n (eq. 5.2, MF5), giving additional small, but significant, improvements. Inspection of the residuals after this final fit showed no obvious trend.

The fit MF5 for the main field is

$$\sigma_k^2/\sigma_a^2 = 0.27 + (1.81 + 13.18/n)\exp[-(n-m)/4.49] + (1.62 + 9.83/n)\exp[-m/1.09]. \tag{6.1}$$

This simple seven-parameter fit accounts for two-thirds of the mean-square value of the ratio. Considering that the individual values are very scattered, this is a commendable result.

When (5.2) was fitted to the systematic field leaking from the ionosphere, the parameter g was 0.58, the same value was obtained when (5.2) was fitted to the residuals after MF2. This confirmed that slight variation of the systematic ionospheric leakage is the source of the enhancement of the zonal variances. (In the combined fit its value was modified to the 1.09 of eq. (6.1).) Such a comparison is not possible for the parameter d in the sectorial enhancement, as we do not know which, if any, of our models of serial correlation is relevant. We used the simulations only to guide us to a reasonable algebraic form for the enhancement. (In fact d was 2.55 for the $\beta = 0.6$ autoregressive noise, decreasing as β increased.) Also, the variation of the sectorial enhancement with n was larger than it was in the particular simulations of Section 4.

The values of MF5, eq. (6.1), are shown for integral n and m in Fig. 8. One way of looking at the different regions of the figure is in terms of the latitudinal and longitudinal spatial frequencies. The

Table 2. Least-squares fitting of the ratio σ_k^2/σ_a^2. The third column gives the percentage sum-square residual (SSR) after the fit. (Note that the fitting formula is schematic; the parameters have different numerical values at each stage.)

	Fitting formula	SSR (per cent)
	0	100
MF1	a	65
MF2	$a + b\exp[-(n-m)/d]$	49
MF3	$a + (b + c/n)\exp[-(n-m)/d]$	38
MF4	$a + (b + c/n)\exp[-(n-m)/d] + e\exp(-m/g)$	35
MF5	$a + (b + c/n)\exp[-(n-m)/d] + (e + f/n)\exp(-m/g)$	34

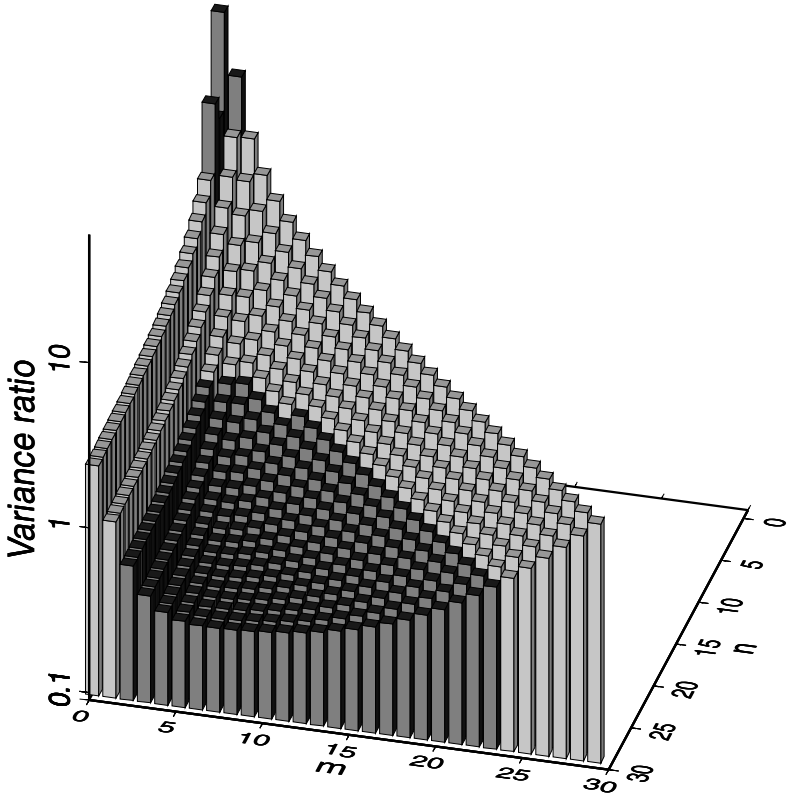

Figure 8. Overall correction factor as applied to the main-field variances of the formal Ørsted model OSVM covariance matrix. The values are those of eq. (6.1) except for the extra factors of 2 for g_1^0 and g_3^0. Factors greater than 10.0 or less than 1.0 are shown darker.

enhancement of the near-zonal variances, coming from variations in the ionospheric leakage, is for harmonics having low frequency in longitude and mainly high frequency in latitude. Conversely, the enhancement of the near-sectorial variances, coming from serial correlation of the magnetospheric 'noise', is for harmonics having low frequency in latitude and mainly high frequency in longitude. The reduction of variance is for harmonics having high frequency in both directions.

The formal standard deviations of the individual parameters are about 20 per cent for a, about 7 per cent for b, c and d, and about 16 per cent for e, f and g. This means that the formal standard deviation of the whole function is about 7 per cent for most values of (n, m), but increases to about 20 per cent at the baseline level of 0.27 (large n, intermediate m). But it will be argued in Section 7 that it would be more realistic to adopt an overall uncertainty of the whole expression of about 25 per cent.

6.1.2 The dipole terms

In the Ørsted (09a/01) model the $n = 1$, dipole, coefficients have some of the smallest 'internal' variances; in total they correspond to only about 6×10^{-3} nT^2 mean-square field at sea level. This means that the number and distribution of the data well define these coefficients provided that the assumptions given in Section 1 are valid. However, our 'objective' variance estimation of Section 3 showed that two of the three dipole terms had by far the largest variances of any coefficient, corresponding to about 120×10^{-3} nT^2. Hence the value of the ratio σ_k^2/σ_a^2 is 56 for g_1^0, and averages 12 for g_1^1 and h_1^1.

There are several reasons why these dipole terms may be subject to large error. In Section 5 we found by simulation that a realistic average ionospheric field leaked 5300×10^{-3} nT^2 zonal dipole field into a OSVM-type solution. Most of this is a systematic error, but it does confirm that the especially large 50×10^{-3} nT^2 random scatter corresponding to the σ_k^2 for g_1^0 is due to small variations in the leakage of the ionospheric field.

The $n = 1$ coefficients are probably also subject to extra uncertainty because of the difficulty in differentiating between the main part, coming from the core, and that part induced by time variation of the ring current. This is made worse by the arbitrary choice for the zero level of the *Dst* index.

Our empirical fit (6.1) to the $n = 2$–25 values of the ratio predicts large values for $n = 1$, as the sectorial enhancement coming from the serial correlation, and the zonal enhancement from the random part of the ionospheric leakage, are both at their largest here. In fact (6.1) gives about 20 for g_1^1 h_1^1, somewhat larger than the observed average, and about 24 for g_1^0, less than half the observed value.

Any *Dst* induced-current contributions should affect all the $n = 1$ terms equally, while the ionospheric leakage is much larger for the $m = 0$ term. To allow for the random part of this ionospheric leakage we propose, fairly arbitrarily, to increase the g_1^0 variance by a factor of two, in addition to the factor given by (6.1).

6.1.3 The g_3^0 term

We saw in Section 5.4 that at sea level the unmodelled mean ionospheric field gave about four times as much systematic leakage, 20 nT^2 mean-square field, into g_3^0 as it did to g_1^0. This suggests that

Table 3. Least-squares fitting of the ratio σ_k^2/σ_a^2 for the secular variation. The third column gives the percentage sum-square residual (SSR) after the fit. (Note that the fitting formula is schematic; the parameters have different numerical values at each stage.)

	Fitting formula	SSR (per cent)
	0	100
SV1	a	69
SV2	$a + b\exp[-(n-m)/d]$	47
SV3	$a + (b + c/n)\exp[-(n-m)/d]$	32

for the random part of the leakage (6.1) might again underestimate the real uncertainty of this coefficient by a factor of 2; it predicts a ratio of 8 against an observed 16. We therefore propose to use the same additional factor of 2 as for g_1^0.

6.1.4 The n = 26–29 terms

Although these terms were not used in the least-squares fit, it turns out that equation (6.1) is in fact a good fit to them. For completeness we will therefore apply (6.1) to the OSVM variances for these terms, but note that it was not claimed that these terms are meaningful.

6.2 Least-squares fitting of the ratio σ_k^2/σ_a^2 for the secular variation

The 09a variances for the secular variation show the same pattern of variation with n and m as do the main-field values. As for the main field, a series of least-squares fits was made for the ratio σ_k^2/σ_a^2 for the secular variation, with the results shown in Table 3. The four-parameter fit SV3 accounts for two-thirds of the mean-square value of the ratio. Adding an enhancement of the zonal terms made no significant improvement, so it appears that leakage of unmodelled ionospheric field did not significantly affect the secular variation within each of the four data subsets, even though it did affect the main field obtained from each subset. The fit for the secular variation is

$$\sigma_k^2/\sigma_a^2 = 0.22 + (0.17 + 23.32/n)\exp[-(n-m)/3.12], \qquad (6.2)$$

showing a more marked variation with n than did the main field, eq. (6.1). The appropriate values of eq. (6.2) are displayed in Fig. 9.

The lack of any effect from the ionospheric field is an example of how a secular variation analysis can be quite different from a main-field analysis. The value of the ratio for g_1^0 is about twice that given by (6.2), but on the other hand the other two $n = 1$ ratios are about one-half that given by (6.2). We decided, again fairly arbitrarily, that there was no clear need to treat the $n = 1$ terms any differently from the others, so will use (6.2) for all the terms for $n = 1$–13. The formal uncertainty of (6.2) ranges from about 7 per cent to 20 per cent, but again we will adopt an overall uncertainty of the expression (6.2) of 25 per cent.

7 DISCUSSION OF THE ØRSTED RESULTS

7.1 Variance estimates

The least-squares process that gave the set of spherical harmonic coefficients called the OSVM model (Olsen 2002), also produced an output covariance matrix, the diagonal terms of which estimate the variances of the coefficients. For the internal field we call these

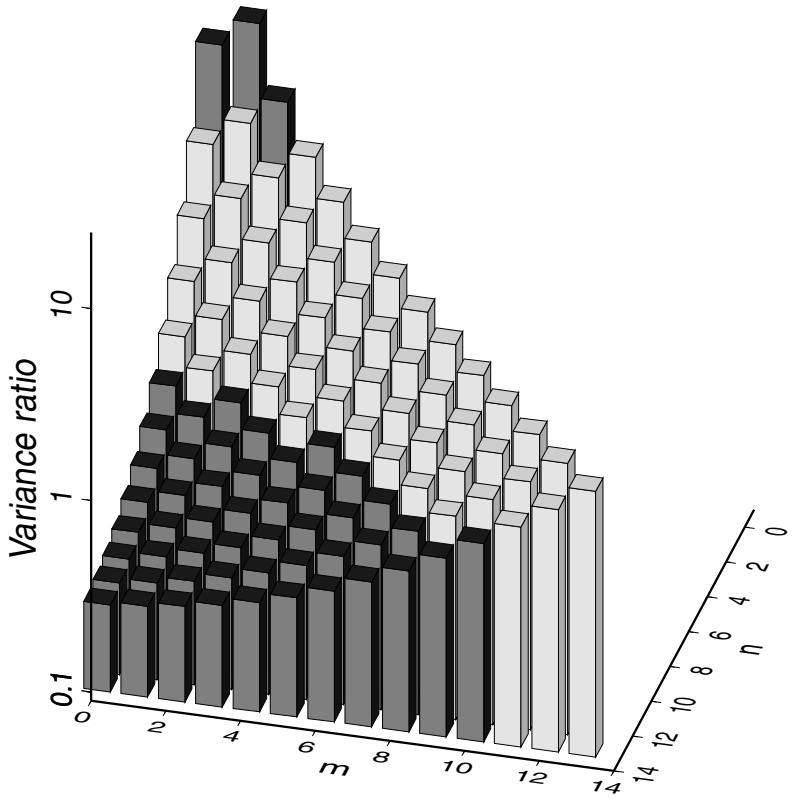

Figure 9. Overall correction factor as applied to the secular variation variances of the formal Ørsted model OSVM covariance matrix. The values are those of eq. (6.2). Factors greater than 10.0 or less than 1.0 are shown darker.

variances σ_a^2. (For simplicity, in this discussion we will ignore the small differences between the variances of the 09a model and those of the OSVM model.) Because the observations were at satellite altitude, while the coefficients are expressed as at sea level, σ_a^2 inevitably increases with degree n. Also, within each degree, because the data amount to equatorially concentrated vector measurements and polar-concentrated intensity measurements, the variances were largest at $m = 0$ and $=n$. Because the analysis was based on a large amount of data, the variation of σ_a^2 with n and m was piecewise smooth. Any similar analysis based on the same type of satellite data selection will give the same sort of pattern.

However, the accuracy of such 'internal' variance estimates depends on the validity of the assumptions (a) to (c) of Section 1.1, and in practice these assumptions are not all valid. We have produced objective 'external' estimates of the variances by performing OSVM-type analyses on four independent subsets of the data, and then using the scatter of the four resulting models to derive a corresponding set of variances σ_k^2. To ease comparison with the 'internal' estimates σ_a^2 we used the ratio σ_k^2/σ_a^2. Because each σ_k^2 was based on only three degrees of freedom, these ratios were very scattered, but various trends were apparent. Within each degree n we found that σ_a^2 underestimated the variances at and near $m = 0$ and $m = n$, but overestimated the variance for intermediate m at large n.

We have shown that the residuals of the OSVM least-squares fit have significant serial correlation. Qualitative arguments show this would be expected to lead to the OSVM covariance matrix overestimating the variances of many of the coefficients, but underestimating those for sectorial ($n = m$) and near-sectorial coefficients. Our simulations confirmed the existence of this effect, and suggested an algebraic function of n and m which fits the effect quite well.

We also used simulations to investigate the leakage of various types of field not included in the OSVM model into the solution. It turns out that, even though the satellite data were all taken at night time, unmodelled ionospheric fields, and especially their Earth-induced counterparts, make a significant contribution to the zonal ($m = 0$) and near-zonal coefficients, some of which is pseudo-random between data sets. A similar algebraic function fitted this effect.

We then used the forms of these functions to make an overall least-squares fit to the set of ratios σ_k^2/σ_a^2. The resulting correction factors, equation (6.1) for the main field (together with additional factors of two for g_1^0 and g_3^0) and (6.2) for the secular variation, are applied to the OSVM variances. These factors vary from about 1/3 (for the large n, intermediate m, variances), to 2–20 (for the zonal and sectorial variances), and up to 50 for g_1^0, so their effect is significant. The resultant scaled variances for the main field are shown in Fig. 10; all the variances are available at www.dsri.dk/Oersted/Field_models. We believe that they are much more realistic estimates of the variances of the OSVM harmonic coefficients.

With no correction, for the field of degrees $n = 1$–13, approximating the main field of mainly core origin, at sea level the formal OSVM variances give a mean-square vector field error of prediction over the Earth's surface of 1.5 nT^2; applying (6.1) increases this to 3.5 nT^2. At higher degrees there are more terms for which the ratio is less than 1, so the overall effect is less; for $n = 1$–25 the corresponding values are 31.5 and 38.5 nT^2. But it must be remembered that these are global averages; within the overall increase in noise power there is a major transfer of power from terms having large n, intermediate m, to terms at and near $m = 0$ and $m = n$, so we expect the uncertainty to vary significantly over the surface. For the secular variation, applying (6.2) increases the $n = 1$–13 mean-square uncertainty from 1.4 to 2.1 $(nT\ yr^{-1})^2$, with the same proviso about redistribution.

Of course our estimates of the variances are still just estimates, and will not be exact; the formal standard deviations of (6.1) and (6.2) range from about 7 per cent to 20 per cent. As in any least-squares process, these formal values are based not only on the lack

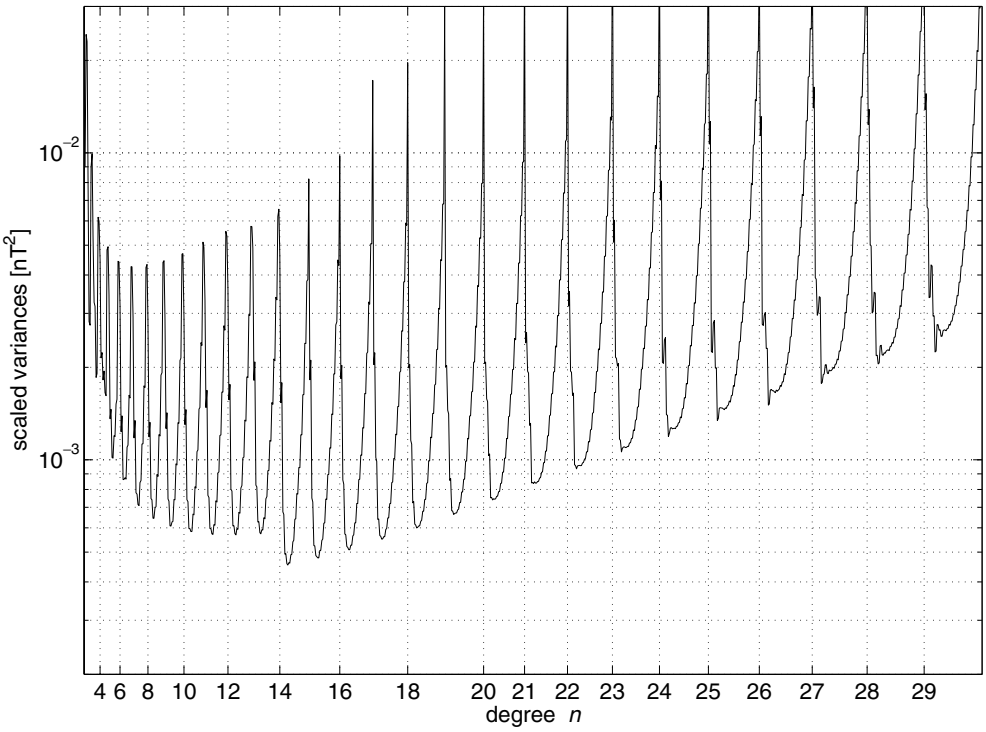

Figure 10. Individual main-field variances of the Ørsted model OSVM after scaling. (The order is g_1^0, g_1^1, h_1^1, g_2^0, etc.)

of fit to the individual values, but also on the assumption that the fitting function is appropriate. We suggest that it would be more realistic to adopt a simple overall uncertainty of 25 per cent. It must be remembered that these variances estimate the pseudo-random scatter about the mean; as discussed below there are also larger systematic errors in some of the coefficients. Also, if there are other defects in the modelling not yet considered, these might well give other contributions to uncertainty!

In the OSVM the various external and induced internal terms were co-estimated essentially to better model the internal field and its secular variation. We have therefore not tried to investigate the inaccuracy of the corresponding variances/covariances.

Because of the very large numbers of covariances, only the variances were investigated directly. To extend our results to the covariances we have assumed that the OSVM output correlation matrix is correct; we have therefore scaled the OSVM covariances by the geometric mean of the two corresponding variance ratios. This scaled matrix is also available at www.dsri.dk/Oersted/Field_models/OSVM.

7.2 Systematic errors

We have shown that there is significant systematic leakage of the average quiet-day ionospheric field into the OSVM solution, and ideally we need to correct for this as far as possible. Table 1 shows that, in our simulation using the CM3 model of an average ionosphere, a total of 94 nT^2 was leaked into the model, but much of this was in the noisy high harmonics. At this stage we do not know how accurate our estimation of the leakage is. However, omitting the g_3^0 contribution, the sea-level power spectrum of the leaked field tends to decrease with degree to about $n = 8$, and then increase; we take this to indicate that above about $n = 8$ the observed 'leakage' is probably mainly noise. It would therefore seem sensible to truncate any correction at $n = 8$. (But note that such a truncation does not mean that there is no significant correction at higher degrees.) This truncated estimate of ionospheric leakage is shown for sea level in Fig. 11; it has a ms field of 39 nT^2. Of the 80 coefficients, 48 have leakage greater than the scaled standard deviation corresponding to eq. (6.1); 32 have leakage more than twice the scaled standard deviation.

Applying such a correction will reduce the systematic errors of the OSVM coefficients, but will increase their random errors, as the numbers we use will not be exact. More work using different ionospheric models would be needed to give some indication of the uncertainty of the correction. We therefore chose NOT to make this correction. The listing at www.dsri.dk/Oersted/Field_models gives the original OSVM coefficients, together with the revised variances, which are our best estimates of the effect of random error in the observations. It lists separately the values we estimate for the correction (to $n = 8$) for the ionospheric leakage (to be subtracted from the OSVM values); the user can decide whether or not to apply the correction.

Our approach uses a deterministic model of the ionosphere to estimate the systematic leakage. This in contrast to that of Langel *et al.* (1989), who used a statistical model of the ionosphere, and increased their coefficient 'variances' by the square of the magnitude of the resulting leakage; for a given degree n, their model spread the ms ionospheric power equally between all the harmonics.

8 GENERAL DISCUSSION

It is clear that in any simple modelling of satellite magnetometer data, the resultant output covariance matrix can be a very poor estimate of the accuracy of the resulting coefficients. And if incorrect weighting has been used in the least-squares process, the coefficients themselves might not be optimum.

It has perhaps not been realized sufficiently just how large is the variation of coefficient of variance with n and m when using a typical satellite (vector + scalar) data set, even if the noise were uncorrelated. And the effect of correlated noise is to significantly

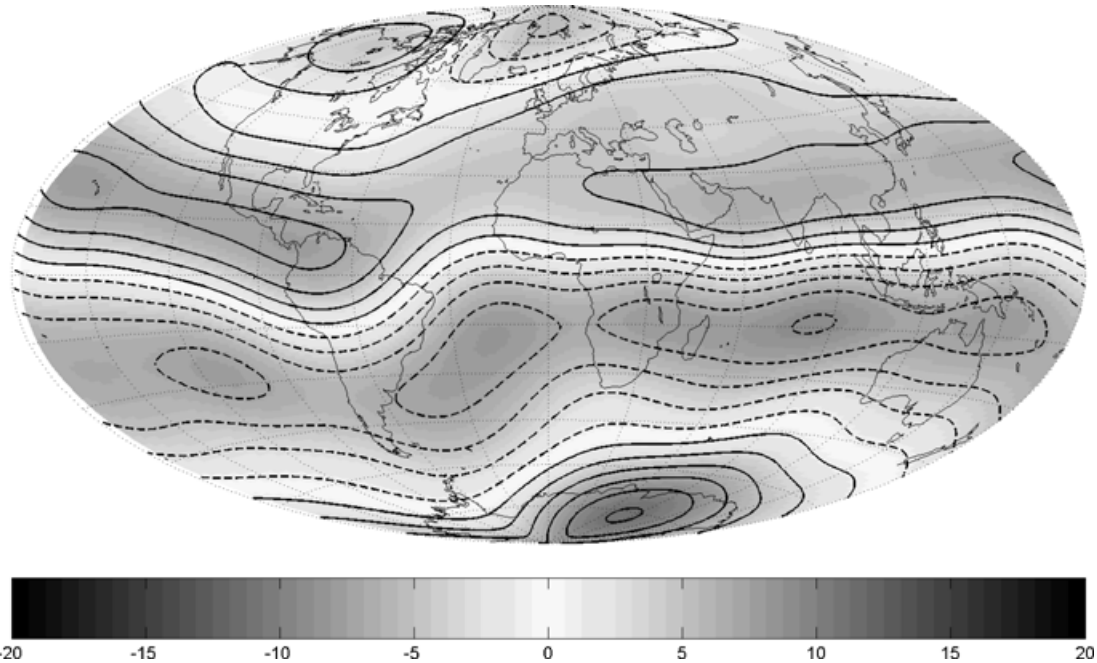

Figure 11. An estimate (truncated to $n = 8$) of the spurious field 'leaked' into the OSVM solution from the average ionosphere. The figure shows the radial component at ground level. Contour interval 2 nT; Hammer projection.

increase this variation. More work is needed to see how this variation affects the various users of the models.

Although the numerical details of this paper are appropriate only to the OSVM model, the general approach is applicable to any model. Ideally any modelling should be such that all the three assumptions of Section 1 are valid (or, if they do not hold, are properly accounted for in the modelling approach), so that we can trust the output coefficients and covariance matrix to be valid estimates, but in practice this is probably never fully possible. The modeller will have to compromise between the relative merits of increasing the accuracy of the output coefficients, having a proper knowledge of what this accuracy actually is, and the amount of computation involved.

Given sufficient data, the simple approach of this paper is one possible compromise; use the coefficient set obtained by a simple least-squares fit of the whole data set, but use the differences between the results from independent subsets of that data to estimate more realistic variances. For a data set having similar separation and distribution to that of the OSVM modelling, the scaling factors obtained in this paper could be used instead.

Because the least-squares process used in the OSVM modelling is not quite linear, averaging the subset coefficients does not give quite the same values as analysing the whole data set. However, the differences are small. For example, for our four data subsets, the mean coefficients rarely differed from the coefficients derived from the total data set by more than the scaled standard deviation. Because it will have more uniform coverage, the total data set probably gives a slightly better result, but without doing a proper least-squares fit incorporating the serial correlation we cannot be sure.

Another approach would be to use a sparser data set, so that the data errors were no longer correlated. Then, apart from the complications caused by the ionospheric leakage, the diagonally weighted least-squares approach would indeed give valid estimates of the coefficients and their variance, but it might reduce the resolution of high-degree harmonics. It would also throw away useful information, even if this was in a complicated form. To see this, suppose that we used only every second point, and that this was sufficient to avoid serial correlation. Then by raising the curve of Fig. 1 by a factor of 2, we would have valid estimates of the resulting variances. We see that, for large n, the peaks of Fig. 10 have roughly the same value as the peaks of the displaced Fig. 1; the effect of the extra information contained in the larger, if more complicated, data set giving Fig. 10 has been to reduce the variances everywhere except at these $m = n$ peaks. (This can be explained by appropriately modifying the arguments of Section 4.) Presumably, by using even more intermediate data (of which there is a vast amount), we could reduce the non-sectorial uncertainties even more, though with rapidly diminishing returns. At this stage we do not know what the optimum data spacing is. Nor do we know how users of the models would be affected by the increased variation of variance with n and m.

In theory the nature of the average serial correlation of the data could be obtained empirically, and, at least for the quiet-time data normally used, would be applicable to other data sets, as there is probably not much variation with altitude in the range used by magnetometer satellites intended for measurements of the main field. So, if the relevant autocorrelations can be obtained (and again ignoring the effects of ionospheric leakage), a modeller prepared to work with the complications of a very large, but sparse, near-diagonal, input covariance matrix could use the appropriate non-zero input covariances, and obtain a better set of coefficient estimates, together with a better estimate of their variances and covariances; clearly this is the ideal approach.

Unfortunately there are several complications in this approach; while the autocorrelation function is easily obtained for the intensity measurements—see for example Fig. 4; it is not clear at this stage how best to obtain it for the vector measurements. There will be a fairly simple autocorrelation function when the measurements are expressed in a fixed geocentric coordinate system, but the actual observations are of field components in the rotating spacecraft reference frame. Therefore the effective correlation will depend on how these rotate in space, and so will be different from spacecraft to spacecraft. And in either coordinate system there will now also be (additional) cross-correlations between the components!

Of course, if eventually we better understood the (probably mainly) magnetospheric processes which led to this correlated 'noise', and were able to find some other proxy measurement related to it, we could apply a correction in the modelling; then the remaining residuals would have much less serial correlation.

Could we gain by using several data sets overlapping in time, each with data far enough apart in time that there is no significant error correlation within each subset? Each subset will give a set of coefficients having well estimated, if comparatively large, variances, and we can certainly average the resulting coefficients over all the subsets. But the individual data of each set are highly correlated with the corresponding ones of the other sets, and so the resulting coefficient errors will also be correlated; therefore averaging the coefficients will not decrease the variances as much as might be expected. Also, because of the correlation, we cannot use the coefficient differences to give an independent estimate of the real variances. (We found another problem. In a simulation allocating alternate data points to two subsets, because the data points were now twice as far apart in time the pseudo-random part of the ionospheric leakage was different; it was comparable to that of our main simulations up to $n = 8$, but tended to be larger, and to move to $m = 1$ or 2, for higher degrees.)

Whatever the approach, it is probably still desirable to check the variance estimates by comparing the coefficients obtained by analysing appropriate subsets of the data.

So far this discussion has concerned random errors. The effect of these can be reduced (if inefficiently) by increasing the number of days of data. Unfortunately it turns out that if the field due to the average night-time ionosphere is not modelled then there is also a very significant systematic error. The exact nature of this will depend on the details of the data distribution, but its magnitude is not changed simply by increasing the number of data points. For the OSVM modelling this systematic ionospheric leakage corresponded to a root-mean-square vector field at the surface of about 6 nT up to $n = 13$; this is larger than the estimated random error of about 2 nT for these main-field terms. In fact, as this systematic leakage is concentrated in the zonal and near-zonal coefficients, for these terms its magnitude is much larger than that of the random errors. We have made an estimate of this leakage, and hence could approximately correct for it, but at this stage we do not know how accurate the estimate is. If its relative uncertainty is in fact less than about 10 per cent, then for the OSVM model applying such a correction would not significantly increase the overall random error of these coefficients. For a longer data set, or for a data set using data from several satellites, the variances could be less, and this simplification might no longer be valid. More work is needed to determine the extent to which the leakage (and hence correction) depends on the nature of the data distribution.

The ideal would be to include terms representing a realistic ionospheric contribution in the least-squares solution; then the problem of 'leakage' of the ionospheric field into the main-field solution

would be very much reduced, if not eliminated. (Though the simulations of Section 5 suggest that the remaining unmodelled fields might give similar, if smaller effects.) This approach is taken in the 'Comprehensive Model' of Sabaka *et al.* (2002) and Sabaka *et al.* (2004).

ACKNOWLEDGMENTS

We are grateful to Richard Holme for helpful discussion, and to David Barraclough and Coerte Voorhies for helping to improve the presentation of the paper. Some of the diagrams in this paper were prepared using the Generic Mapping Tools produced by Wessel & Smith (1998).

REFERENCES

Ashour, A.A. & Price, A.T., 1965. Night-time earth currents associated with the daily magnetic variations, *Geophys. J. R. astr. Soc.,* **10**, 1–15.

Bloxham, J. & Jackson, A., 1991. Fluid flow near the surface of Earth's outer core, *Rev. Geophys.,* **29**, 97–120.

Cain, J.C., Hendricks, S.J., Langel, R.A. & Hudson, W.V., 1967. A proposed model for the International Geomagnetic Reference Field-1965*, *J. Geomag. Geoelectr.,* **19**, 335–355.

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

Jekeli, C., 1996. Spherical harmonic analysis, aliasing, and filtering, *J. Geodesy,* **70**, 214–223.

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

Langel, R.A., Estes, R.H. & Sabaka, T.J., 1989. Uncertainty estimates in geomagnetic field modeling, *J. geophys. Res.,* **94**, 12 281–12 299.

Lowes, F.J., 1975. Vector errors in spherical harmonic analysis of scalar data, *Geophys. J. R. astr. Soc.,* **42**, 637–651.

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

Rygaard-Hjalsted, C., Constable, C.G. & Parker, R.L., 1997. The influence of correlated crustal signals in modelling the main field, *Geophys. J. Int.,* **130**, 717–726.

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

Sabaka, T.J., Olsen, N. & Purucker, M., 2004. Extending comprehensive models of the Earth's magnetic field with Ørsted and CHAMP data, accepted by, *Geophys. J. Int.*

Wessel, P. & Smith, W.H.F., 1998. New, improved version of Generic Mapping Tools released, *EOS, Trans. Am. geophys. Un.*, **79**, 579.

APPENDIX : THE EFFECT OF SERIAL CORRELATION ON FOURIER ANALYSIS

Consider the 1-D case of N observations equally spaced by $2\pi/N$ around a circle. If the noise at each point is independent random noise, of mean-square value σ^2, then a Fourier analysis along the circle will give a noise power (mean-square value over the circle) for each of the two harmonics at ν cycles/circle of

$$P(\nu) = 2\sigma^2/N. \tag{A1}$$

This is a 'white' noise spectrum; the input noise is spread equally among all the harmonics (except for the $\nu = 0$ average value).

Now consider what happens when the noise is serially correlated. One simple type of correlation is where a new independent noise value is added in at each point, and remains constant for a total of p points before vanishing; there are then (a progressing selection of) p independent noise contributions at each point. Such noise amounts to overlapping rectangular signals, each of width $\Delta\theta = p2\pi/N$, but of random amplitude A_i. First consider the Fourier analysis of the particular individual rectangular noise contribution which has amplitude A_0 and length $\Delta\theta$ symmetrical about $\theta = 0$. Treating this as a signal $y(\theta)$, it is represented by the Fourier series

$$y(\theta) = \frac{A_0\Delta\theta}{\pi}\sum_{\nu=0}^{\infty}\frac{\sin(\nu\Delta\theta/2)}{\nu\Delta\theta/2}\cos(\nu\Delta\theta) \tag{A2}$$

and its power spectrum is

$$P(\nu) = \frac{1}{2}A_0^2(\Delta\theta/\pi)^2[\sin(\nu\Delta\theta/2)/(\nu\Delta\theta/2)]^2. \tag{A3}$$

Now add in the other, independent, contributions, each displaced by $2\pi/N$. Each error contribution will produce a similar spectrum, but with the appropriate phase shift in the $\cos(\nu\Delta\theta)$ term, and the appropriate random amplitude A_i. When these are summed, the result squared, and the expectation taken, we get

$$P(\nu) = \frac{1}{2}NA^2(\Delta\theta/\pi)^2[\sin(\nu\Delta\theta/2)/(\nu\Delta\theta/2)]^2, \tag{A4}$$

where A^2 is the expectation of A_i^2. Comparing (A.4) with (A.3) we see that, as would be expected, having N independent noise sources increases the output noise power by a factor of N. In our case we want the total mean-square noise at each point to be σ^2; therefore for p overlapping contributions, the expectation of A_i^2 must be σ^2/p. Also $\Delta\theta = p2\pi/N$, so finally we have

$$P(\nu) = (2p\sigma^2/N)[\sin(p\nu\pi/N)/(p\nu\pi/N)]^2. \tag{A5}$$

For low frequencies, $\nu \leq N/\pi\ p$, (long wavelengths) the factor $[\ldots]^2$ is nearly unity, and we see by comparing (4.5) with (4.1) that the effect of the serial correlation is to increase the output noise power by a factor of p. (For the $\nu = 0$, average, term we can see that this is because the overlapping of the noise means there is less cancellation of the random contributions, and the effective number of degrees of freedom is reduced by the factor p.)

On the other hand, for the higher frequencies (short wavelengths) the effect of the serially correlated noise rapidly becomes trivial; the effect is only significant for wavelengths comparable with, or larger than, the noise 'correlation length' $p\Delta\theta = p2\pi/N$. The reason for this is that in the Fourier analysis the 'signal' is multiplied by the appropriate sine wave and then integrated; for high frequencies/short wavelengths the correlated noise looks like a constant 'signal' over each period, and so does not contribute much to the integral.

For a given total mean-square noise at each point, the overall effect is that increasing the correlation length of the noise transfers output noise power from high frequencies to low frequencies.

Another form of serial correlation is autoregression. For a first-order autoregressive process, a fraction β of the noise at one data point is still present at the next data point. In this case, if the expected mean-square value of any one error is σ^2, then the noise y_i at point i is given by

$$y_i = \beta y_{i-1} + \gamma e_i, \tag{A6}$$

where $\gamma = (1-\beta^2)^{\frac{1}{2}}$, and e_i is a random sample of noise of variance σ^2. Again the noise consists of the addition of independent components, but in this case each component decreases by a factor of β between adjacent points. This corresponds to the exponential decay

$$y(\theta) = A_i\exp[-\alpha(\theta-\theta_i)], \tag{A7}$$

where $A_i = \gamma\, e_i$, and $\alpha = (N/2\pi)\ln(1/\beta)$. Using the approximation that α is large enough (β small enough), so that $y(\theta)$ is negligible for $\theta \geq 2\pi$, the power spectrum of the decay starting at $\theta = 0$ is

$$P(\nu) = A_0^2/(\alpha^2 + \nu^2). \tag{A8}$$

By the same argument as above, on average the power spectrum of the N uncorrelated random-amplitude signals starting at equally spaced points around the circle is just N times that of one signal:

$$P(\nu) = NA^2/(\alpha^2 + \nu^2). \tag{A9}$$

Again the effect of the serial correlation is to transfer noise power from high frequencies to low frequencies.

In effect, discrete Fourier analysis over the N points approximates the input signal $y(\theta)$ by a series of trapezoids; the above analysis is based on exact integration, and so will become invalid for rectangular pulses which are too short, or exponential decay which is too rapid. However, the analysis can be expected to reproduce the main effects for the situations we are considering.

We see that, whatever the exact nature of the serial correlation, its effect is to transfer noise power from high spatial frequencies to low frequencies.