Geomagnetic field analysis — I. Stochastic inversion

David Gubbins Bullard Laboratories, Madingley Rise, Madingley Road,
Cambridge CB3 0EZ

Received 1982 November 16

Summary. Most of the Earth’s magnetic field and its secular change originate in the core. Provided the mantle can be treated as an electrical insulator, stochastic inversion enables surface observations to be analysed for the core field. A priori information about the variation of the field at the core boundary leads to very stringent conditions at the Earth’s surface. The field models are identical with those derived from the method of harmonic splines (Shure, Parker & Backus) provided the a priori information is specified appropriately.

The method is applied to secular variation data from 106 magnetic observatories. Model predictions for fields at the Earth’s surface have error estimates associated with them that appear realistic. For plausible choices of a priori information the error of the field at the core is unbounded, but integrals over patches of the core surface can have finite errors. The hypothesis that magnetic fields are frozen to the core fluid implies that certain integrals of the secular variation vanish. This idea is tested by computing the integrals and their standard and maximum errors. Most of the integrals are within one standard deviation of zero, but those over the large patches to the north and south of the magnetic equator are many times their standard error, because of the dominating influence of the decaying dipole. All integrals are well within their maximum error, indicating that it will be possible to construct core fields, consistent with frozen flux, that satisfy the observations.

1 Introduction

The magnetic field contains information about the behaviour and nature of the Earth’s core. In order to extract this information we have to analyse observations made at the Earth’s surface (and at satellite altitude), determine that part of the signal originating within the core, and downward continue the field to give a representation at the core—mantle boundary (CMB). The mantle will be regarded as an electrical insulator for the purposes of this paper; recent work suggests that this is a good approximation (Benton & Whaler 1983).

Maps of the magnetic field at the CMB are readily derived from a spherical harmonic
model of the potential (e.g. Kahle, Vestine & Ball 1967). These maps have inspired new insight into core processes, but quantitative testing of these ideas requires that error bounds be placed on the field values on these maps. Most spherical harmonic models include standard deviations of each of the coefficients, but they are all inadequate, for one reason or another. The IGRF (International Geomagnetic Reference Field) and DGRF (Definitive Geomagnetic Reference Field) series of models are a compromise between various competing national models, and it is impossible to assign scientifically defensible error estimates to their coefficients. Individual contributing models, such as Barraclough et al.'s (1978) definitive field, which was a constituent of the 1965 DGRF, are usually derived by conventional spherical harmonic analysis (SHA), a least squares fit to a truncated spherical harmonic expansion; standard deviations are quoted, but the full covariance matrix is not usually available. SHA is satisfactory for producing an interpolated field at the Earth's surface, but the errors at the CMB may be dominated by short wavelength contributions that are ignored by this method of analysis. Similar problems will arise when high quality satellite data (e.g. MAGSAT) are analysed by SHA.

Ideally one tests theoretical ideas by relating the relevant physical quantities (in this case properties of the magnetic field at the CMB, the velocity of the core flow, etc.), directly to the observations. Given an acceptable error for the measurement, one can test if the idea is consistent with the observations. Unfortunately the determination of these quantities is a complex and inherently unstable process, so that it may not always be possible to obtain a sufficiently simple relationship between desired model and observation to apply inverse theory. At best such an inversion would be a very laborious calculation, and would not make use of all of the available data.

Geomagnetic field analysis presents a dilemma: we can either produce maps which cannot be used in a quantitative manner, or progress very slowly by inversion of small subsets of the available data. What is required is some simple method of analysis, which can readily be converted to a map at the CMB, but which also contains the mapping of the observational errors. In this paper stochastic inversion (SI) is applied to the problem of finding the core field.

Stochastic inversion is a technique for dealing with ill-posed least squares problems (Franklin 1970). If some of the parameters are poorly determined by the data, a least squares procedure can give rise to spuriously large values of those parameters because noise has been mapped into the solution. SI damps out these spurious features by incorporating a priori information about the model itself. An extension of SI that makes use of the concept of 'a priori data' has been given by Jackson (1979), whose approach is followed here. We have good a priori information about the magnetic field because it originates in the core. The great distance between the CMB and the Earth's surface means that the short wavelength fields are strongly attenuated at the observation point, so that restrictions on the short wavelength fields at the CMB become very severe restrictions at the surface. In fact this is the usual justification for using a truncated spherical harmonic expansion for these analyses; SI puts the argument on a more explicit and precise footing.

SI is reviewed and applied to geomagnetic fields in Section 2, where I also show how to choose a priori information in such a way that the models are identical with those derived from the method of harmonic splines (HS) of Shure, Parker & Backus (1982).

In Section 3 the method is applied to secular variation data and the results are used to test the frozen flux approximation. Roberts & Scott (1965) suggested that the core could be treated as a perfect conductor on time-scales of the secular variation, and that, with this assumption, measurements of secular variation could be used to deduce the core flow responsible for the temporal changes in the magnetic field. Backus (1968) emphasized that,
to be consistent with this hypothesis of frozen-in fields, the secular variation must satisfy the integral constraints

$$\int_{S_r} \vec{B}_r \, dS = 0$$  \hspace{1cm} (1)$$

where \( \{S_r\} \) are the patches of the CMB bounded by contours of zero radial field ('null flux curves'). Furthermore, Backus showed that, provided (1) is satisfied, it is possible to construct a large class of flows of the core fluid that could be responsible for the secular variation. The hypothesis is tested by computing the integrals in (1) from the models, and comparing the results with their variances computed from the covariance matrix of the coefficients.

The three primary methods of analysis (SI, SHA and HS) are compared by applying all three to the same dataset, and the results and conclusions are discussed in Section 4.

2 Method

2.1 STOCHASTIC INVERSION

The starting point for SI is the same as for conventional global analyses of magnetic data; details are to be found in Barraclough & Malin (1971). With the usual geomagnetic conventions, the radial component of magnetic field may be written as

$$B_r = \sum_{l=1}^{\infty} \sum_{m=0}^{l} \left( \frac{a}{r} \right)^{l+2} P_l^m(\cos \theta) (g_l^m \cos m\phi + h_l^m \sin m\phi)(l+1)$$  \hspace{1cm} (2)$$

where \( r = a \) is the Earth’s surface and the Schmidt quasi-normalized surface harmonics are such that their squares, integrated over the unit sphere, yield \( 4\pi(2l+1) \).

Setting \( r = a \) in (2) gives a relationship between the vertical component of magnetic field at the Earth’s surface and the geomagnetic coefficients: \( \{g_l^m ; h_l^m\} \). This equation, together with two similar equations for the horizontal components of field, are truncated at some degree, \( N \), to form the basis for a least squares analysis of observations for the coefficients. Arranging the observations into a data vector, \( \vec{y} \), and the coefficients in to a model vector, \( \vec{m} \), [specifically \( \vec{m} = (g_1^0, g_1^1, h_1^0, g_2^0, g_2^1, \ldots) \)], the truncated form of (2) is written as

$$\vec{y} = A \vec{m} + \vec{e}$$  \hspace{1cm} (3)$$

where \( \vec{e} \) is the error vector and the elements of \( A \) are terms like \( (l+1)P_l^m(\cos \theta) \cos m\phi \).

(3) are called the equations of condition.

Following Jackson (1979), we seek some estimate, \( \vec{m}_0 \), of the model vector, that is a linear combination of the data, in the form

$$\vec{m}_0 = H \vec{y}.$$  \hspace{1cm} (4)$$

Substituting into (3) gives

$$\vec{m}_0 = H A \vec{m} + H \vec{e}.$$  \hspace{1cm} (5)$$

The estimation error is defined as

$$\vec{m}_0 - \vec{m} = (H A - I) \vec{m} + H \vec{e}.$$  \hspace{1cm} (6)$$

which has covariance

$$C = (H A - I) C_m (H A - I)^T + H C_e H^T$$  \hspace{1cm} (7)$$

where \( C_m \) is the a priori covariance matrix of the model and \( C_e \) that of the data errors, \( \vec{e} \).
D. Gubbins

H is selected to minimize the sum of squares of the estimation errors, the trace of C, to give

$$ H = C_mE^T(A_mC_m^T + C_e)^{-1} = (A^T C_e^{-1} A + C_m^{-1})^{-1} A^T C_e^{-1}. $$

Substituting back into (4) gives the estimate

$$ \hat{m} = (A^T C_e^{-1} A + C_m^{-1})^{-1} A^T C_e^{-1} $$

(8)

(7) now gives

$$ C = (A^T C_e^{-1} A + C_m^{-1})^{-1}. $$

(9)

From equation (5), $\hat{m}$ is seen to be composed of a mapping of data errors (He) and a linear combination of model parameters (HAm). In an ideal case, e would be zero and HA would be the identity. HA is called the resolution matrix; it measures how well the individual model parameters have been separated by the procedure. The estimate error separates in the same way.

When the elements of $C_m$ become large, (8) reduces to the normal equations for least squares analysis, and $HA = I$. The concept of resolution does not arise in conventional least squares parameter fitting. A very common application of SI takes $C_m$ to be some multiple of the identity matrix. This bland choice of $C_m$ does not prejudice the final estimate towards any particular model parameter. The procedure is equivalent to 'adding white noise' to stabilize Weiner filter design in signal processing.

2.2 APPLICATION TO GEOMAGNETISM

$C_m$ is specified by supposing that the a priori variance in the radial field on the CMB be independent of spherical harmonic degree. This builds in the physical requirement that the magnetic field originate in the core. The radial field is chosen because it is the only component of $B$ that is guaranteed to be continuous across the CMB. Independence of spherical harmonic degree is a bland hypothesis that will not prejudice the result for the core field, and any dependence on order would imply a departure from spherical symmetry, which is not expected.

The radial field at the CMB, taken to be at radius $c$, is obtained from (2) by setting $r = c$. Suppose that the a priori variance of a coefficient $g_l^n$ or $h_l^m$ is $v_l$. The variance of the radial field at the CMB will then be

$$ V = \sum_{l=1}^{\infty} v_l \left( \frac{a}{c} \right)^{2l+4} (l+1)^2 $$

(10)

where use has been made of the formula

$$ \sum_{m=0}^{l} \frac{m^2}{l^2} (\cos \theta) = 1. $$

The hypothesis requires that each term in the sum contribute equally to the total variance, which is achieved by the choice

$$ v_l^{-1} = \lambda (l+1)^2 \left( \frac{a}{c} \right)^{2l+4} $$

(11)

where $\lambda$ is a multiplicative factor called the damping constant. The hypothesis assigns very small variances to the high degree coefficients, which are heavily attenuated because of the great distance between the Earth's surface and the source of the field. The effect on the
normal equations (8) is to add very large numbers to the diagonal terms, particularly those corresponding to large degree.

In SHA, the truncation level, \( N \), is chosen from considerations of the available data. In this application of SI it will be chosen at a high value so as to achieve a good approximation to the limit \( N \to \infty \). The damping constant is varied so as to achieve a satisfactory fit to the data. Convergence is studied by computing and comparing models truncated at successively higher degree.

2.3 CORRESPONDENCE WITH THE METHOD OF HARMONIC SPLINES

The estimate given by (8) minimizes the quantity

\[
e^T C_e^{-1} e + \hat{m}^T C_m^{-1} \hat{m} = e^T C_e^{-1} e + \hat{m}^T \Lambda \hat{m}
\]

where \( \Lambda \) is the diagonal matrix with elements \( \lambda(l+1)^2(a/c)^{2l+4} \), from (11). To prove this, we regard \( \lambda \) as a Lagrange multiplier and minimize \( E = e^T C_e^{-1} e \), where \( E^{1/2} \) will be called the misfit, subject to the side constraint

\[
\hat{m}^T \Lambda \hat{m} = s.
\]

The quantity \( s \) defines a norm of \( \hat{m} \), and different norms may be defined for different choices of \( \Lambda \). This minimization gives the solution (8). Now consider a second estimate, \( \hat{m}_1 \), which satisfies \( \hat{m}_1^T \Lambda \hat{m}_1 \leq s_1 \), where \( s_1 > s \). \( \hat{m} \) satisfies this inequality as well, because of (13), and therefore the largest value that \( E(s_1) \) can take is \( E(s) \). Hence \( E(s_1) \leq E(s) \) if \( s_1 > s \), and \( E \) must be a monotonically decreasing function of \( s \). SI therefore yields the estimate that minimizes the misfit subject to the norm \( \hat{m}^T \Lambda \hat{m} \) being less than some prescribed value, \( s \). The smallest misfit is achieved when the norm is equal to \( s \), so that no other model fits the data as well with a smaller norm; the model has the ‘minimum norm’ property.

Shure et al. (1982) found the model which minimized a slightly different norm

\[
\int_{\text{CMB}} B_r^2 dS = 4\pi \sum_{l=1}^{\infty} \sum_{m=0}^{l} \frac{(l+1)^2}{(2l+1)} \left( \frac{a}{c} \right)^{2l+4} (g_l^m \hat{g}^m_{l+1} + h_l^m \hat{h}^m_{l+1})
\]

for which our matrix \( \Lambda \) has elements \( \lambda(2l+1)^{-1}(l+1)^2(a/c)^{2l+4} \). This choice of \( \Lambda \) differs from (11) by the factor \( (2l+1) \). Shure et al. (1982) show that any model with given norm and misfit will be unique. Therefore their solution will be equal to that derived by SI in the limit \( N \to \infty \). (11) corresponds to minimizing the integral

\[
\int_{\text{CMB}} \frac{1}{r} \frac{\partial}{\partial r} (r^2 B_r^2) dS.
\]

The extra factor of \( (2l+1) \) in (14) means that it will produce a field that is ‘rouglier’ on the CMB, but in practice it is the continuation factor \( (a/c)^{2l+4} \) which dominates the appearance of the resulting model, factors of \( l \) being relatively unimportant.

In the spherical harmonic series for the radial field at the CMB the a priori variance of a coefficient is constant for the uniform variance norm (11), but proportional to \( (2l+1) \) for the minimum energy norm (14). Calculations have been performed using both norms in this paper. Their differences become important when considering errors in the field at the CMB.

2.4 ERROR ANALYSIS

The spherical harmonic series is truncated when the field has converged; increasing the truncation point adds model parameters that are effectively zero. The minimum norm property
of the model means that it converges rapidly, and it will in fact converge faster than elements of the covariance matrix. At high degree the coefficients are not at all well determined by the observations and SI assigns to them the \textit{a priori} variance. The variance of these high degree coefficients will therefore be of the form \( k (l + 1)^{-2} x^{2l+4} \), where \( x = c/a \), for the uniform variance norm, and \( k' (l + 1)^{-2} (2l + 1) x^{2l+4} \) for the minimum energy norm. The constants \( k \) and \( k' \) can be found from high degree elements of the covariance matrix, and these asymptotic forms can be used to estimate the errors due to neglected coefficients in the series for the field.

Consider, as an example, the determination of the radial component of field at the Earth’s surface. From (2) with \( r = a \), we have

\[
\text{Var}(B_r) = a^T C a
\]

where \( a \) is a vector with components like \((l + 1) P_l^n (\cos \theta) \cos \phi\). The asymptotic form for the variances can be used to estimate the effect of increasing the truncation level in the calculation. For uniform variance the high degree terms will give

\[
\text{Remainder Var}(B_r) = k \sum_{l=N+1}^{\infty} x^{2l+4} = k \frac{x^{2N+6}}{1 - x^2}.
\]

The asymptotic form for the minimum energy norm is

\[
\text{Remainder Var}(B_r) = k' \sum_{l=N+1}^{\infty} (2l + 1) x^{2l+4} = k' \frac{(2N + 3) x^{2N+6} - (2N + 1) x^{2N+8}}{(1 - x^2)^2}.
\]

Point estimates of \( B_r \) have been found using the dataset described in Section 3. They have standard deviations comparable with the misfit, but with a geographical variation of a factor of 10, presumably due to the poor distribution of the observatories. Using only the diagonal terms of the covariance matrix gives very mild geographical variations. For one model of secular variation with a misfit of \( 2.47 \) yr\(^{-1}\), the standard deviation calculated for the point at \( 52^\circ \text{N}, 0^\circ \text{E} \) was \( 2.85 \) yr\(^{-1}\), of which 0.16 came from the remainder term (17); the overall geographical variation was from 0.4 to 3.1 yr\(^{-1}\).

Estimates of the field at the CMB are of greater interest to us. Neither of the asymptotic forms for the variances considered so far converge there, for any field component. It is therefore imperative to resist the temptation to estimate errors at the CMB from any analysis involving a truncated spherical harmonic expansion. Moreover, however large the dataset, there will always be a limiting degree beyond which the variances will be dominated by the \textit{a priori} information, so that the only way to limit the variance of estimates at the CMB is to assume them to be finite. SHA makes all variances zero beyond a certain degree, but this is unrealistic. Another possibility is to place limits on the allowed range of values that the field can take on the CMB, treating it as a \textit{a priori} data as described by Jackson (1979), but in this instance this amounts to assuming the desired result. Another possibility is to insist that the Ohmic heating be bounded. This restriction applies to magnetic field but not to secular change, and further discussion is postponed until a later paper of this series, when magnetic fields will be analysed.

It will never be possible to get good estimates of field values at points on the CMB. The
problem is less severe if we estimate an average over some patch of the CMB, because then
the effects of short wavelength fields are less severe. This is discussed in Section 3.2.

3 Applications

3.1 COMPARESOF DIFFERENT MODELS

The method has been tested on observatory annual mean data. Secular variation was
estimated for all three components of the magnetic field at those 106 observatories which
recorded continuously, or almost continuously, for 15 years from 1959 to 1974, by taking
the difference of 1974 and 1959 values. An error or weight was assigned to each measure-
ment, based on the scatter of the annual means from a smooth curve in time. Full details
are in Shure et al. (1983). The data are not corrected for height of the observatory above
mean sea-level, or ellipticity of the Earth's surface.

The models derived from this data are listed in Table 1. The defining parameters are the
truncation level, the damping constant and the definition of the norm (e.g. uniform
variance). Two important quantities derived from the models are the misfit, measured in
\( \gamma \text{ yr}^{-1} \), and the norm itself. It was shown in Section 2.3 that the misfit is a monotonically
decreasing function of the norm; a plot of \( E \) against \( s \) gives a trade-off curve. Model 6 (see
Table 1) was derived by HS (Shure et al. 1983) and is included only to demonstrate equality
with that derived by SI.

Convergence was evaluated by calculating solutions with \( \lambda \) fixed, and successively higher
truncation. Model 1 converged by degree 10. Model 3 is a better fit to the data and contains
more energy at high degree and order than model 1, and therefore converges more slowly.
It has converged by degree 12. The agreement between model 6, derived by HS, and the
equivalent model 5, which has a truncation at degree 14, is better than 0.005 \( \gamma \text{ yr}^{-1} \), demonstrat-
ing that the two methods are indeed equivalent.

Conventional analyses (SHA) correspond to low truncation and zero damping. They can
be compared with SI models provided the damping constant is chosen to give the same
misfit. The SI models minimize the norm, so that SHA must give a model with larger norm
(Table 1). Model 1 was chosen for comparison with model 2, a sixth-degree SHA model;
mmodel 3 is to be compared with model 4. Their differences are, for the most part, smaller
than their estimate errors. Exceptions to this rule are the high degree coefficients of the SHA
models, some of which are significantly different, particularly for the eighth-degree model 4.

Estimation errors on individual coefficients are typically 20 per cent smaller for the

Table 1. Details of the models used in this paper. The norms are measured in (\( \gamma \text{ yr}^{-1} \))^2, those
for \( E \) must be multiplied by \( c^2 \) to give an integral over the CMB.

<table>
<thead>
<tr>
<th>Model</th>
<th>Truncation</th>
<th>Misfit (( \gamma \text{ yr}^{-1} ))</th>
<th>Norm type</th>
<th>Norm</th>
<th>Damping constant (( \lambda ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>4.23</td>
<td>E</td>
<td>15.47</td>
<td>4.9 \times 10^{-4}</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>4.24</td>
<td></td>
<td>26.37</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>14</td>
<td>2.28</td>
<td>E</td>
<td>47.10</td>
<td>3.5 \times 10^{-4}</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>2.28</td>
<td></td>
<td>99.91</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>14</td>
<td>2.39</td>
<td>E</td>
<td>41.31</td>
<td>3.55 \times 10^{-5}</td>
</tr>
<tr>
<td>6</td>
<td>( \infty )</td>
<td>2.39</td>
<td>E</td>
<td>41.34</td>
<td>3.57 \times 10^{-5}</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>2.19*</td>
<td>E</td>
<td>41.33</td>
<td>3.53 \times 10^{-5}</td>
</tr>
<tr>
<td>8</td>
<td>14</td>
<td>2.18*</td>
<td>U</td>
<td>544.90</td>
<td>0.2 \times 10^{-5}</td>
</tr>
</tbody>
</table>

\( E \) denotes the minimum energy norm (14); \( U \) the uniform variance norm (11).

\*Indicates model derived from weighted data; misfit is from weighted data.
stochastic models 1 and 3 than for 2 and 4. Consider the error in a model prediction of the field at the Earth’s surface. The variance in the field value will contain contributions from coefficients of degrees 1–6 (for model 2; 1–8 for model 4) plus, for the stochastic models, an additional contribution from the coefficients of degrees 7–14 (for model 1; 9–14 for model 3) plus a ‘remainder’ term that would be incurred if we were to increase the truncation point beyond degree 14. The variance out to degree 14 is computed from the covariance matrix using equation (15), and the remainder from (17). The remainder term is $0.1 - 0.2 \gamma \text{yr}^{-1}$ for both stochastic models; the total variance varies by about a factor of 10 depending on geographical location but is typically $4 \gamma \text{yr}^{-1}$ (models 1 and 2) and $2 \gamma \text{yr}^{-1}$ (models 3 and 4), as would be expected. Thus the stochastic models do not give a larger variance for the vertical field at the Earth’s surface, despite having more coefficients than the SHA models.

Resolution is very high for the low degree coefficients. The diagonal elements of the resolution matrix fall below 0.9 at degree 6 for model 3, and at lower degree for model 1, which is more heavily damped. Resolution at degree 8 is about 0.5, and it falls off rapidly thereafter.

Weighting was applied by estimating the covariance matrix of the data errors as

$$C_e = K C_0$$

where $C_0$ is a matrix of weights, determined from the scatter in the annual means, and $K$ is the square of the overall misfit to the weighted data. In all cases the data were assumed independent, so that all off-diagonal elements of $C_0$ are taken to be zero.

Models 5 and 7 were derived from weighted and unweighted data respectively, and have virtually the same norm. If the weights derived by Shure et al. (1983) are good estimates of the errors, then model 7 would have misfit of unity. With this norm we are only able to achieve a misfit of 2.18 $\gamma \text{yr}^{-1}$. The problem of being unable to fit the observations as well as expected was discussed by Shure et al. (1983). These calculations have shown that the difficulty cannot be resolved by weighting the data. However, models 5 and 7 are significantly different (i.e. differ by more than the estimate errors): estimate errors are smaller for model 7, and yet resolution is larger, demonstrating that it is easier to fit core fields to the weighted data.

Uniform variance is illustrated by model 8, which has the same misfit to the weighted dataset as model 7. These two models are very similar even when downward continued to the CMB. Model 8 is smoother and converges faster, but the differences between individual coefficients is much less than the corresponding estimate error. The estimate errors for model 8 are somewhat larger than those for 7 at low degree, but smaller at higher degree; resolution is better at low degree for model 8, but worse at high degree. This reflects the relatively heavier damping at high degree in model 8.

### 3.2 Testing the ‘Frozen Flux’ Hypothesis

Testing whether the core is behaving as a perfect conductor involves calculating the integrals of the secular variation in (1), and showing them to be zero. The integrals are over patches of the CMB that are bounded by null flux curves. The definitive model for the magnetic field at epoch 1965 (Barraclough et al. 1978) has six such patches, yielding five independent integrals. These integrals have been calculated for the secular variation model 8 and are displayed in Table 2. We can ask the following two questions:

1. Are the observations consistent with measurements of a core field frozen to the core fluid, plus some random errors?
Table 2. Integrals of secular variation, in the radial component, over patches of the CMB bounded by null flux curves. A map of the patches is to be found in Shure et al. (1983); the left-hand column gives the geographical location of the patch. The numbers in the table have to be multiplied by $c^2 \times 10^{-9}$ to convert them to Wb yr$^{-1}$.

<table>
<thead>
<tr>
<th>Region</th>
<th>Integral</th>
<th>Standard deviation</th>
<th>Maximum error</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Africa</td>
<td>-159</td>
<td>73</td>
<td>813</td>
</tr>
<tr>
<td>South America</td>
<td>39</td>
<td>73</td>
<td>676</td>
</tr>
<tr>
<td>North Pole</td>
<td>22</td>
<td>32</td>
<td>215</td>
</tr>
<tr>
<td>North hemisphere</td>
<td>752</td>
<td>140</td>
<td>1481</td>
</tr>
<tr>
<td>South hemisphere</td>
<td>-649</td>
<td>162</td>
<td>1695</td>
</tr>
<tr>
<td>Pacific</td>
<td>-15</td>
<td>60</td>
<td>672</td>
</tr>
</tbody>
</table>

(2) Will it be possible to construct a model which satisfies these constraints and is still a satisfactory fit to the data?

The answer to (1) will be yes if the values of the integrals are within one or two standard deviations of zero. The standard deviations in Table 2 were calculated from the covariance matrix of the coefficients from the formula $\hat{a}_i^{T} C \hat{a}_i$, where $\hat{a}_i$ has components like

$$(a/c)^{2l+4}(l+1) \int_{S_i} P_{l}^{m}(\cos \theta) \cos m\phi dS.$$ 

It does not include any contribution beyond degree 14. (2) will be true provided the integral does not exceed the maximum error, which is defined to be

$$\sum_{i} a_i |\hat{a}_i|,$$

where $\sigma$ is the vector of estimate errors, because then it will be possible to adjust the model coefficients by small amounts ($< a_i$), so that the data fit is unchanged, until the integral is zero. These adjustments need not be a random departure from the original model, which is why the requirement is less stringent than for (1). Maximum errors are displayed in Table 2. These have been calculated for degrees 1 to 14 and do not include a remainder term.

All the integrals lie within one standard deviation of zero except the patch over South Africa, which is two standard deviations away, and those over the large regions to north and south of the magnetic equator, denoted north and south hemisphere in the table. These latter two are far in excess of their standard deviations.

Before seeking reasons for these large integrals, we must first establish that the errors in Table 2 are good estimates of the true uncertainty. There are two reasons for the errors being too small; the first is that we have not taken account of uncertainty in the location of the null flux curves themselves. If the integrals have been computed over the wrong patch, then further errors will be incurred. The integrals were recomputed using null flux curves derived from the IGRF for 1965 (Zmuda 1971) with results that were within 10 per cent of Table 2. The IGRF was derived from observations made prior to 1965, and was in some sense predictive, whereas the definitive model was based on data which straddled the epoch 1965. The two sets of null flux curves coincide very well, showing that magnetic field is much better determined than is secular variation. The same may not be true at other epochs. A second reason is the remainder term. The remainder can be investigated analytically for the full northern hemisphere, $0 \leq \theta \leq \pi/2$, using the asymptotic forms for the standard deviations of the coefficients. The contribution to the variance, for the uniform variance...
norm, is
\[ V_R = k \sum_{l=N+1}^{\infty} Y_l^2 \]  
(19)

where
\[ \bar{Y}_l = 2\pi \int_0^1 P_l(x) \, dx \]

\[ \bar{Y}_{2n+1} = 2\pi (-1)^n \frac{(2n-1)!}{4^n (n+1)! (n-1)!}; \quad \bar{Y}_{2n} = 0; \quad n > 0 \]  
(20)

(from, e.g. Gradsteyn & Ryzhik 1965, 7.111 and 8.832), (19) becomes

\[ V_R = 4k\pi^2 k \sum_{n=M}^{\infty} \left\{ \frac{(2n-1)(2n-3) \ldots 3}{(2n+2)(2n) \ldots 2} \right\}^2 \]  
(21)

where \( M \) is the smallest integer such that \( 2M > N \).

Comparison with the series whose general term is \( n^{-1+\alpha}; \alpha > 0 \), shows this series to converge at least as fast as \( n^{-3} \). When the highest term omitted from the sum is \( (2M+1) \) the integral test can be used to show that the remainder satisfies, approximately

\[ V_R \leq \frac{0.16 TC^2 k}{M^2} \]

For model 8 the covariance matrix gives a standard deviation of 118 for the integral, and the remainder 100, with the units of Table 2. The remainder is very large for this slowly convergent series, but not large enough to give the five-fold increase needed to account for the integral in Table 2. The remainder term for the maximum error is even larger, and for the minimum energy norm it does not even converge; it can be shown to converge for the uniform variance norm. However, this does not concern us now because all the maximum errors in Table 2 are well in excess of the corresponding integrals, and it is clearly possible to construct 14th degree models which satisfy both data and flux constraints. This point will be pursued in paper II.

It appears, therefore, that this secular variation model cannot be regarded as an estimate of a frozen-in core field plus random errors; either the frozen flux assumption is invalid, or there is some agent causing a systematic difference between the true field and its estimate. Booker (1969) has pointed out that the decay of the dipole field makes a substantial contribution to the flux integrals, and it is this effect we are seeing in the two integrals north and south hemisphere. The available data constrain the dipole well but do not have sufficient resolution to determine smaller scale features that may compensate for the change in flux. All analysis procedures set poorly determined features to zero, so that this would appear as a systematic bias away from a frozen-in field. The next step is to produce models that are consistent with frozen flux, and to test the hypothesis further by incorporating more data from satellites and earlier epochs.

4 Discussion

The main theme of this paper has been to present stochastic inversion as a good routine method for analysing magnetic observations for core fields. It yields models that are smooth on the CMB, and useful error estimates, so that results can be used for core studies: it is also a viable alternative for producing maps of fields at the Earth's surface. The two alternative methods of analysis are the conventional method (SHA) and harmonic splines. The relative merits of all three techniques are now compared.
Booker (1968) and Whaler & Gubbins (1981) have argued that a truncated spherical harmonic expansion will suffer from the same problems of ‘ringing’ as does a truncated Fourier expansion. Attenuation of the high degree coefficients alleviates this problem, because at the Earth’s surface there is little energy in the degree range 6–12 where the cut-off is likely to be taken. However, at the CMB there is no such hole in the spectrum, and ringing is likely to be a problem. SI forces a smooth fall-off with spherical harmonic degree, which reduces this problem.

In practical applications it is an advantage to be able to adjust the damping constant by small amounts, rather than choose the truncation point of a spherical harmonic series. Cutting off the series to include only some of the coefficients of a certain degree (e.g. Fougere 1965) is undesirable because it introduces a geographical bias, for no good reason. SHA is therefore restricted by truncation at specific degrees, in the case of the dataset in this paper the choice is at 6, 7 or 8. This choice is rather a coarse one, and it will sometimes be an advantage to specify the misfit rather more precisely than this.

The most important feature of SI is that it enables a proper error estimate to be made. Given some assumption about the a priori variation of the field on the CMB, it is possible to assess the error due to the (essentially unknown) high degree coefficients. With conventional analyses, the high degree coefficients are assumed to be identically zero, and the overall error estimates are optimistic. The difference between the two approaches is critical when considering the fields at the CMB, where the error is dominated by small-scale features that are not constrained by the data. The main disadvantage of SI is that the higher truncation level requires more computation — degree 12 or 14 for this dataset rather than degree 7.

The method of harmonic splines, as described by Shure et al. (1982), gives the same models as SI, and it could even be developed to give the same error estimates. Some workers may have strong opinions on the differences between the underlying rationales of the two approaches (see the discussion in Jackson 1979 and Sabatier 1979), but when two methods give virtually identical results, practical considerations may dictate our choice. SI requires less computation than HS when the number of data exceeds the number of coefficients retained (318 against 224 for degree 14), but HS will be faster if the optimal basis set is used (Parker & Shure 1983). SI has the distinct advantage that it only requires a few lines of alteration to a computer program already running for conventional analyses!

The principal advantage of SI is greater flexibility. Changing norms is trivial to do, whereas HS involves the calculation of new sets of infinite sums for each norm (Shure et al. 1982). It is not always possible to do these sums analytically, and much of the advantage of HS is lost once the sums need to be evaluated numerically. There are several extensions to the present method that will require a major extension of the present version of HS, but which will be straightforward by SI; these include analysis of inclination and declination data, total field data, and building in the constraints of frozen-in fields. This last application will be dealt with in paper II of this series.

Acknowledgment

This work was partially supported by NERC Grant GR3/3475.

References


