Chapter 3

Time–dependent modelling of the geomagnetic field

A lack of information cannot be remedied by
any mathematical trickery.
C. Lanczoz

In this chapter I will outline the method used for the simultaneous time–dependent field modelling of the geomagnetic main field and its secular variation between 1980 and 2000. The idea of simultaneous inversion of main field and secular variation dates back to Cain et al. [1965] and has been further developed by Bloxham [1987]; Bloxham & Jackson [1989].

The new aspect of the methodology developed in this thesis is the usage of additional a priori information to constrain the time–dependent model in 1980 and 2000 by field models of high–quality satellite vector data. The use of the two satellite epochs as "bookends" will enable a higher resolution of secular variation modelling than has been possible to date.

I will begin this chapter with a description of the data and data processing. Then follows some basic discussions of inverse problems and their solution. In section 3.4 I will describe the new method used to develop time–dependent models of the geomagnetic main field and the secular variation. Also, I hope to make clear, why such a intricate apparatus is required to model the Earth's magnetic field. The results are presented in section 3.5.

3.1 Data

The data used in this work are magnetic observatory annual means, monthly means and repeat station measurements. The observatory annual means and the repeat survey data were kindly provided by S. Macmillan (British Geological Survey (BGS), Edinburgh). Monthly means, defined as being the average over all days of the month and all times of the day, were derived from hourly mean values downloaded from the World–Data–Center Copenhagen[1] and from the Intermagnet database[2]. Table (A.1) lists observatories and repeat

[1](http://web.dmi.dk/projects/wdcc1/master.html)
[2](http://www.intermagnet.org/myservlet/imotbl_e.jsp)

stations locations from which the data were used, the number of data for each field component and its uncertainty estimates. The computation of the errors for each component will be outlined in chapter 3. Figure (3.1) shows the global distribution of the observatories, where the different symbols correspond to observatories with monthly means (red circle), observatories with only annual means (black circle). A similar map (figure 3.2) is drawn for the repeat stations: a red triangle indicate multiply visited stations and black triangle station only visited once.

3.1.1 Geomagnetic observatory data

For the field modelling three components, northward (X), eastward (Y) and downward (Z), were compiled for each single observatory. Table (A.1) lists the number of available data for each component and observatory. The coding of the observatories is basically identical with the 3 character coding of IAGA (International Association of Geomagnetism and Aeronomy), but it differs for those sites, where location changes happened between 1980 and 2000, or an observatory with the same coding already existed at another location. These stations are marked with an additional digit after the IAGA-coding, and although a correction of these changes for most sites would be possible, it was not applied, e.g. Tucson as TUC1 and TUC2. Also sites, where for a longer period no data are available are marked with an extra digit, for example Addis Abbeba (Ethiopia), AAE1 before 1995.2 and AAE2 after 1997 [3].

Table (B.1) lists the known base line jumps, site changes or instrument changes during the period 1980 and 2000.

3.1.2 Repeat station data

Repeat stations are established to record the secular variation in a certain region. The spatial resolution should be high but is typically less than that for the ground measurements carried out to map magnetic anomalies. As the name implies, repeat stations mean a network of fixed sites, where the magnetic field is measured at a more or less regular interval. The occupation interval varies between one year and 10 years; in this study, it is restricted not to be longer than seven years. Any occupation period beyond that is believed not to be useful to uncover short term secular variation as analysed in this study.

The data of repeat surveys are commonly provided after they have been reduced to an epoch, such as annual mean. The reduction is made using

$$C_{rp}(t) = C_{rp}(t_0) + C_{obs}(t) - C_{obs}(t_0)\,, \tag{3.1}$$

where $C_{rp}(t)$ is the annual mean of the repeat station, $C_{rp}(t_0)$ the repeat station measurement, $C_{obs}(t)$ the observatory annual mean and $C_{obs}(t_0)$ the simultaneous measurement at the observatory. Relation (3.1) assumes, that the difference between the observation and annual mean is the same, i.e. that secular variation and external induced variations

[3] This coding corresponds to the coding preferred by the BGS

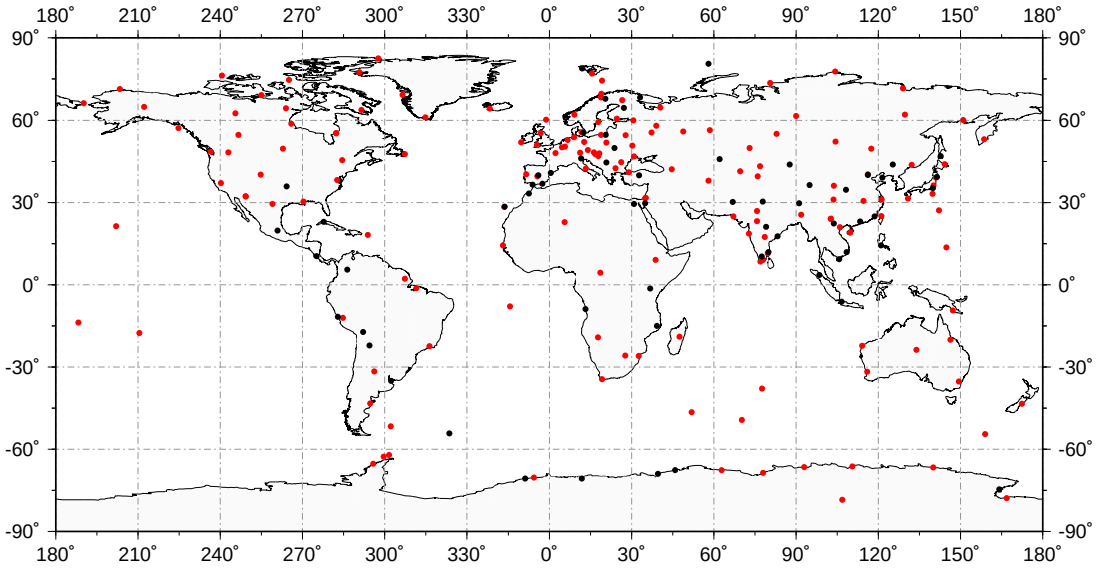

Figure 3.1: Global chart of permanent observatories, red circles shows places for which monthly means are available, black dots for sites where only annual means are available.

X	$\pm$ 5 nT – $\pm$ 20 nT	$\dot{X}$	$\pm$ 7 nT – $\pm$ 40 nT
Y	$\pm$ 4 nT – $\pm$ 13 nT	$\dot{Y}$	$\pm$ 3 nT – $\pm$ 26 nT
Z	$\pm$ 4 nT – $\pm$ 12 nT	$\dot{Z}$	$\pm$ 3 nT – $\pm$ 24 nT

Table 3.1: Error estimates of the magnetic field component X, Y and Z and for their computed secular variations $\dot{X}, \dot{Y}$ and $\dot{Z}$ for European repeat surveys

are identical at the nearby observatory and the repeat station. These idealizations lead to errors in the reduced data depending on the distance from the repeat station to the observatory. If there is no observatory within a radius of 1000 km, then the reduction is sometimes undertaken using a main field model instead of the observatory measurements, but even with main field models (e.g. DGRF–models) the reduction leads to errors, because of limited spatial resolution of these models. Korte [1999] gives the following error estimates of the measurements of the field components and the computed secular variation for European repeat surveys.

Ideally data for field modelling should not be reduced to observatory annual means, but rather to quiet night time. The data should be acquired as a mean of several measurements at different external conditions and separated by 24 hours in order to reduce the external contributions [Newitt et al., 1996].

The repeat survey data considered in this study are taken from a database maintained by the British Geological Survey (BGS)[4]. Only the data from repeat stations outside the

[4]http://www.geomag.bgs.ac.uk/gifs/surveydata.html

European mainland are incorporated. Certainly, the coverage and regular occupations of repeat stations across Europe are much better than in any other part of the world, but I would not expect any improvements in recovering the secular variation, as there already is a good coverage of observatories. Therefore, only repeat survey data are involved for regions, where the observatory density is sparse. The selection of repeat surveys will be listed below with a short description of the data.

Canada:

50 stations, of which nine were occupied more than 5 times, 23 visited four times.

United States including pacific islands:

61 stations, 5 more than 5 times, 7 more than 4 times

Columbia:

79 stations from which 13 were occupied 4 times between 1987 and 1995 and the rest at least twice in this period

Venezuela:

10 stations occupied twice in 1978 and years around 1983

Brazil:

78 stations, 19 of them were visited 4 times or more at an interval of about 5 years

Iceland, Norway, Finland, Portugal and Atlantic Islands:

3 Icelandic stations with 5 or more occupations

1 Norwegian station occupied once every year between 1991 and 1997

16 Finish stations visited up to six times in the period 1978.5 and 1986.5

7 Portuguese stations either on the Azores or on Madeira occupied up to three times during 1989 and 1991

2 stations in the south Atlantic (St. Helena and Ascension Island) which were visited three times.

Near East:

6 stations in Saudi Arabia visited twice, in 1980 and 1983

5 stations in Israel visited up to four times between 1992 and 1999

West and Central Africa:

26 stations were occupied in western and central Africa during 1986 and 1993

Southern Africa:

77 stations in southern Africa, 21 of them were visited five times or more during 1979 and 2000.

Japan and Indonesia:

3 Japanese station located not on the mainland

15 Indonesian sites.

Australia and New Zealand:

69 Australian stations, where 10 were occupied 5 or more times, and 4 stations were occupied four times.

45 stations in New Zealand, six four times and 30 five or more times.

Antarctic region:

14 Antarctic stations, from which seven were occupied five times or more and three visited four times.

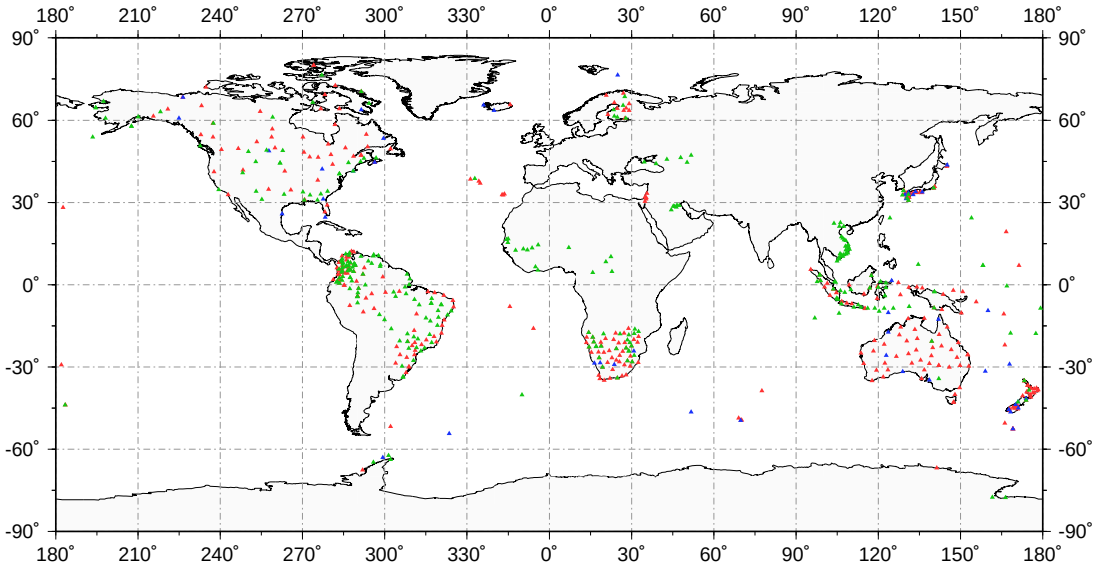

Figure 3.2: Global chart of repeat station, green triangles sites provided one secular variation estimate (SVE), red dots indicate sites of multiple visits, but less than 5 SVE, blue dots indicating sites with more than 4 SVE.

3.2 Data processing

In addition to internal field variation, ground based measurements are also subject to crustal fields and external field variations. Observatories and repeat station are fixed relative to crust and therefore the correlation in these data due to the crustal bias are in the temporal domain, as far as the spatial separation of observatories are comparable longer than the length scale of the crustal field.

Interference from external field variations show temporal as well as spatial correlation. Distinct features of these variations are the semi-annual and the annual signal, which are caused by the movement of the ring–current, a magnetospheric current system in 3–5 Earth radii distance, and the change of heliographic latitude of Earth's orbit.
Using first differences of the measurements – secular variation estimates – should eleminate

the contributions due to crustal biases. In the next section a simple approach is outlined, which reduces the crustal and external biases.

3.2.1 Secular variation estimates

Filtering of time series is always a delicate issue: On one side the applied filter removes essential information inherent in the data and on the other side it introduces artificial signals to the time series. The right balance of both effects is decisive to any conclusion one will draw from this research.
A simple method for removing seasonality, or periodicity, in the data with a period p is differencing in the way, that, if p is a multiple of the sampling interval one could choose to examine

$$C_t - C_{t-p}.$$

In the case, that p and t are not commensurable then using

$$C_t - (1-\delta)C_{t-p} - \delta C_{t-n-1}\,,$$

where $n = |p|$ and $\delta = p - n$, the periodicity can be largely removed [Conradsen & Spliid, 1981]. This approach is known as the *p–step difference filter* [Box & Jenkins, 1976; Priestley, 1981] and it is appropriate to derive a secular variation estimate from observatory monthly means with reducing the contribution of annual and semi–annual variations, e. g. for the northward component

$$dX/dt = X(t+n/2) - X(t-n/2),\ n = 12 \tag{3.2}$$

Observatory annual means are treated using, e. g.

$$dX/dt = X(t) - X(t-1)\,. \tag{3.3}$$

This also reduces the crustal bias, because the crustal signal is assumed to be the same for both dates t and $t-1$ and therefore should canccel.

3.3 Some fundamental concepts of Inverse Problems

3.3.1 Statement of the forward problem and inverse problem

Any inverse problem is intertwined with its related forward problem which in general can be stated as

$$\mathsf{y} = F(\mathsf{x}) \tag{3.4}$$

where F is a integral operator, which is in general nonlinear, y is the observation, e.g. the observed magnetic field at Earth's surface, and x denotes the vector of the unknown magnetic field at the core-mantle boundary (CMB). The approach for the determination of the magnetic field at the CMB consists then in either:

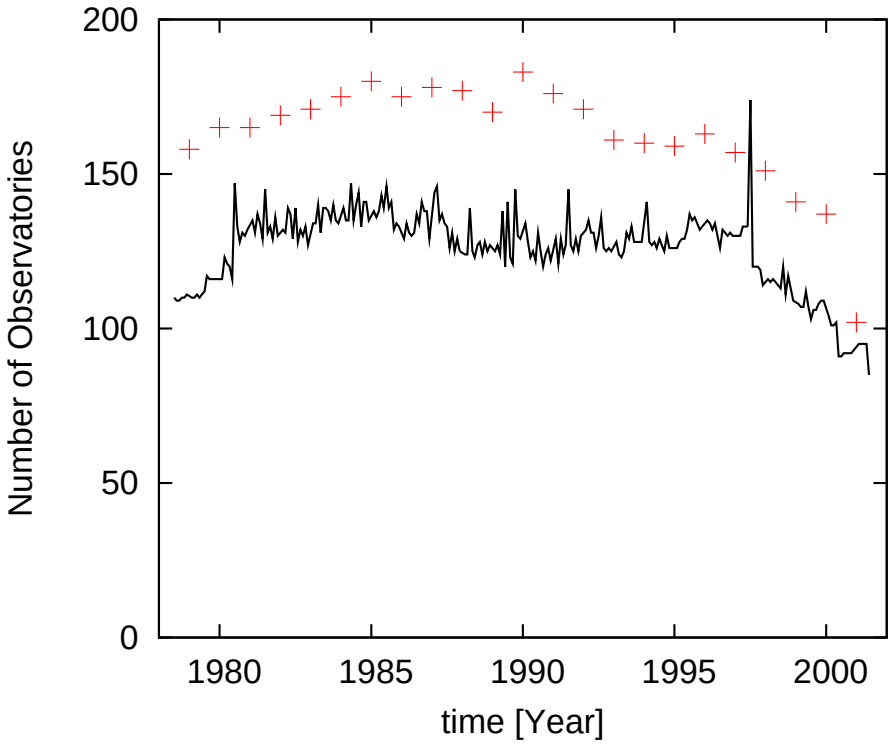

Figure 3.3: The number of available annual secular variation estimates shown as red crosses and the number of monthly secular variation estimates (either based on observatory or repeat station measurements) shown as solid black curve.

a) The repeated forward solving of (3.4) for a trial– and error model x until the observation y is satisfied.

b) The inversion of (3.4) to infer x directly from y.

While forward modelling provides one particular model which fits the observed data, inversion usually allows one to explore the whole model space and to delimit a range of possible models, which fit the data within its experimental and observational error interval equally well.

The inverse solution of (3.4) can be formally written as

$$x = F^{-1}(\mathsf{y})\,. \tag{3.5}$$

where F^{-1} is the inverse operator of F.

There are three issues related to the problem of finding the solution of (3.5):

1) **Existence of a solution** or of the inverse operator F^{-1}. This is a fundamental problem in the theoretical analysis of inverse problems. However, it is of minor

importance for the solution of geophysical inverse problems where an inverse operator can generally be computed by some discrete numerical technique.

2) **Uniqueness of the solution** holds if, and only if the operator F is injective, i.e.

$$\mathsf{y}_1 = \mathsf{y}_2 \Rightarrow x_1 = x_2\,. \tag{3.6}$$

As Backus & Gilbert [1967, 1968]; Backus [1970] pointed out, geophysical inverse problems are inherently non–unique as a consequence of the a priori infinite dimensionality of the model space. Therefore, with only finite data, the model cannot be uniquely determined.

3) **Stability of the solution** questioned how small fluctuation in the data y propagate into changes in the model x. The inverse solution is to be said stable if, for two sufficiently close data sets y, y' and two corresponding models x, x' it is

$$|\mathsf{y} - \mathsf{y}'| \leq \varepsilon \Rightarrow |x - x'| \leq d(\varepsilon). \tag{3.7}$$

An inverse problem which does not satisfy condition 2 and 3 is called an ill–posed inverse problem as opposed to a well–posed inverse problem. To gain a well–posed inverse problem several regularization techniques have been introduced and developed in the past decades, such as Tikhonov regularization, Least squares collocation, Bayes–estimation and Backus–Gilbert formalism [Parker, 1994; Backus & Gilbert, 1967, 1968; Backus, 1970].

3.4 Method

From now the inverse problem is considered to be linear: either linearized using Fermat's principle, or linear by nature, then the linear system of equation is given:

$$\mathsf{y} = \mathbf{A}\mathsf{m} + \mathsf{e}\,. \tag{3.8}$$

This is the equation of condition, where $\mathbf{A}$ is a design matrix mapping the model vector m to the data y, and e is the error vector.

3.4.1 The least–squares estimator

The classical approach to solve linear inverse problems such as (3.8) is based on measures of the distance of modeled or predicted conditions $\mathsf{y}^{mod} = \mathbf{A}\mathsf{m} + \mathsf{e}$ and data y as,

$$\mathsf{e} = \mathsf{y} - \mathsf{y}^{mod} = \mathsf{y} - \mathbf{A}\mathsf{m} \tag{3.9}$$

$$||\mathsf{e}||^2 = \mathsf{e}^{\mathsf{T}}\mathsf{e} = (\mathsf{y} - \mathbf{A}\mathsf{m})^{\mathsf{T}}(\mathsf{y} - \mathbf{A}\mathsf{m}) \tag{3.10}$$

The aim of utilizing this method is to gain a model where the predicted data are as close as possible to the observed data, so the misfit $||\mathsf{e}||^2$ is minimal. Assuming that the matrix

$\mathbf{A}$ has full column rank $p = m$ and that $m < n$, the condition for $||\mathsf{e}||^2$ to be a minimum follows:

$$\mathbf{A}^{\mathsf{T}}(\mathsf{y} - \mathbf{A}\mathsf{m}) = 0\,, \tag{3.11}$$

which is equivalent to the normal equations

$$\mathbf{A}^{\mathsf{T}}\mathbf{A}\mathsf{m} = \mathbf{A}^{\mathsf{T}}\mathsf{y}\,. \tag{3.12}$$

This gives the formal solution, which exists since $p = m$

$$\mathsf{m} = (\mathbf{A}^{\mathsf{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathsf{T}}\mathsf{y}\,. \tag{3.13}$$

This solution is the maximum likelihood; here, it is tacitly assumed the error in (3.8) are additive, independent and to be normal distributed with expectation $E(\mathsf{e}) = 0$ and standard deviation $\sigma^2 = 1$. If on the other hand the errors have a non-uniform variance, then (3.13) reads

$$\mathsf{m} = (\mathbf{A}^{\mathsf{T}}{\mathbf{C_e}}^{-1}\mathbf{A})^{-1}{\mathbf{C_e}}^{-1}\mathbf{A}^{\mathsf{T}}\mathsf{y}\,, \tag{3.14}$$

where ${\mathbf{C_e}}^{-1}$ is the error covariance matrix.

In case of an ill–posed inverse problem, evaluations of the least–squares solution (3.13) by any numerical technique, such as Gaussian elimination, Single Value– or Cholesky decomposition give solutions m and covariances of the model parameter $\mathbf{C_m}$, which are oscillating and large. In the next section I will discuss the origins and consequences of an ill–posed problem, and the general methodology of regularization.

3.4.2 Regularization of ill–posed inverse problem

If the third of the requirements in section 3.1.1 is not fulfilled, then the inverse problem is called ill–posed. As pointed out in the previous section, a solution of the inverse problem can generally be computed. Now the question is, how does the solution depend on the data? We need to quantify the criterion (3.7) for the stability of the solution and how variations of the data map into the model. To quantify the stability of solutions of a set of linear equations, we estimate the total error of the solutions

$$||\mathsf{m} - \mathsf{m}_\delta|| \leq ||\mathbf{A}^{-1}||\;||\mathsf{y} - \mathsf{y}_\delta||\,, \tag{3.15}$$

where m_δ and y_δ are estimates of the vectors m and y, respectively. Furthermore, $\mathsf{m}_\delta = \mathbf{A}^1\mathsf{y}_\delta$. Lawson & Hanson [1974] obtain the following conservative bound

$$\frac{||\mathsf{m} - \mathsf{m}_\delta||}{||\mathsf{m}||} < \kappa(\mathbf{A})\frac{||\mathsf{y} - \mathsf{y}_\delta||}{||\mathsf{y}||}\,, \tag{3.16}$$

where $\kappa(\mathbf{A})$ is the so called condition number of the matrix $\mathbf{A}$ defined as

$$\kappa(\mathbf{A}) = ||\mathbf{A}||\;||\mathbf{A}^{-1}|| = \frac{\max|s_i|}{\min|s_i|}\,, \tag{3.17}$$

s_i are the singular values of the matrix $\mathbf{A}$ computed using the singular value decomposition.

It follows that the solution will be unstable if $\kappa(\mathbf{A}) \gg 1$. Beside numerical round–off errors, the instability of the solutions are also caused by partly collinearity of columns and/or rows. This occurs when the data coverage is insufficient.

Stochastic inversion

I now discuss regularization methods to stabilize the solution of an ill–posed inverse problem. The technique treated here is the stochastic inversion. It is set up to reduce the covariances of the model parameters by invoking objective a priori information of the likeness of the sought solutions. The method was developed by Wiener [1949] and introduced to geophysics by Franklin [1970]. Now consider a linear system (3.8)

$$\mathsf{y} = \mathbf{A}\mathsf{m} + \mathsf{e}\,,$$

where m and e are stochastic processes with zero means. The auto– and cross–covariance matrices are given by

$$\mathbf{C_m} = E[\mathsf{m}\mathsf{m}^\mathsf{T}]\,, \tag{3.18}$$
$$\mathbf{C_y} = E[\mathsf{y}\mathsf{y}^\mathsf{T}]\,, \tag{3.18b}$$
$$\mathbf{C_e} = E[\mathsf{e}\mathsf{e}^\mathsf{T}]\,, \tag{3.18c}$$
$$\mathbf{C_{my}} = E[\mathsf{m}\mathsf{y}^\mathsf{T}]\,. \tag{3.18d}$$

The objective of stochastic inversion is then to find an inverse matrix $\mathbf{H}$ of $\mathbf{A}$, such that the covariance of the true error

$$E[\mathsf{e}_\mathsf{m}\mathsf{e}_\mathsf{m}^T] = E[(\mathsf{m} - \hat{\mathsf{m}})(\mathsf{m} - \hat{\mathsf{m}})^\mathsf{T}] \rightarrow \text{minimum} \tag{3.19}$$

with

$$\hat{\mathsf{m}} = \mathbf{H}\mathsf{y}\,, \tag{3.20}$$

where $\mathbf{H}$ is a $m \times n$ matrix such that the estimate $\hat{\mathsf{m}}$ is a stable average of the data with a meaningful physical interpretation [Jackson, 1979]. Combining (3.20) and (3.8) gives

$$\begin{aligned} \hat{\mathsf{m}} &= \mathbf{H}\mathbf{A}\mathsf{m} + \mathbf{H}\mathsf{e} \\ &= \mathbf{R}\mathsf{m} + \mathbf{H}\mathsf{e}\,. \end{aligned} \tag{3.21}$$

By definition it is $\mathbf{R} = \mathbf{H}\mathbf{A}$.

Following Jackson [1979] and Gubbins [1983] leads to an expression for the optimal $\mathbf{H}$

$$\begin{aligned} \mathbf{H}_{opt} &= \mathbf{C_{my}}\,\mathbf{C_y}^{-1} \\ &= (\mathbf{A}^\mathsf{T}\mathbf{C_e}^{-1}\mathbf{A} + \mathbf{C_m}^{-1})^{-1}\mathbf{A}^\mathsf{T}\mathbf{C_e}^{-1}. \end{aligned} \tag{3.22}$$

The final solution is given by

$$\mathsf{m} = (\mathbf{A}^\mathsf{T}\mathbf{A} + k\mathbf{I})^{-1}\mathbf{A}^\mathsf{T}\mathsf{y} \tag{3.23}$$

or more generally

$$\mathsf{m} = (\mathbf{A}^\mathsf{T}\mathbf{C_e}^{-1}\mathbf{A} + \mathbf{C_m})^{-1}\mathbf{A}^\mathsf{T}\mathbf{C_e}^{-1}\mathsf{y} \tag{3.24}$$

where $k = \sigma_{\mathsf{y}}^2/\sigma_{\mathsf{m}}^2$ is a constant and $\mathbf{I}$ the identity matrix. (3.24) follows directly from the assumption, that m and e are stochastic processes with zero means and variances $\mathbf{C_e} = \sigma_{\mathsf{y}}^2\mathbf{I}$ and $\mathbf{C_m} = \sigma_{\mathsf{m}}^2\mathbf{I}$. This method is exactly the damped least-squares method, here one adds white noise to the diagonal matrix elements of the normal matrix $\mathbf{A}^\mathsf{T}\mathbf{A}$ to stabilize the solution of the ill-posed inverse problem. The model vector m estimated by (3.24) minimizes the objective function [Bloxham, 1987]

$$\Theta(m) = (\mathsf{y} - \mathbf{A}\mathsf{m})^\mathsf{T}\mathbf{C_e}^{-1}(\mathsf{y} - \mathbf{A}\mathsf{m}) + \mathsf{m}^\mathsf{T}\mathbf{C_m}^{-1}\mathsf{m}. \tag{3.25}$$

3.4.3 Application to geomagnetic data

Now, after some theoretical discussions of solving ill-posed inverse problems, I want to apply these methods to develop a time–dependent model of the geomagnetic field and its secular variation between 1980 and 2000. But before, I would like to point out that the previous consideration are needed to model the geomagnetic field. The formalism, which is developed in this section, is based on an approach of Bloxham & Jackson [1992].
The Earth's magnetic field can be represented as a potential field as it is in an electrical insulator $\mathbf{B} = -\nabla V$, where V is a scalar potential field. The potential field has to satisfy the Laplace equation $\nabla^2 V = 0$, so the solution in spherical geometry reads:

$$\begin{aligned} V = a\sum_{l=1}^{\infty}\sum_{m=0}^{l}\Bigg\{ &(g_l^m\cos(m\phi) + h_l^m\sin(m\phi))\left(\frac{a}{r}\right)^{l+1} P_l^m(\cos\theta) \\ &+ (q_l^m\cos(m\phi) + s_l^m\sin(m\phi))\left(\frac{r}{a}\right)^{l} P_l^m(\cos\theta)\Bigg\}, \end{aligned} \tag{3.26}$$

where a is the Earth's radius (6371.2 km) and (r, θ, ϕ) the geocentric coordinates, θ the colatitude. The $P_l^m(\cos\theta)$ are the Schmidt normalized associated Legendre functions

$$\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}(P_l^m(\cos\theta)\cos(m\phi))^2\sin\theta\, d\theta\, d\phi = \frac{4\pi}{2l+1}, \tag{3.27}$$

where l are the degrees and m the order. The coefficients $\{g_l^m, h_l^m\}$ and $\{q_l^m, s_l^m\}$ are the Gauss coefficients.

$$\begin{aligned} V_{int} &= a\sum_{l=1}^{\infty}\sum_{m=0}^{l}\left(\frac{a}{r}\right)^{l+1}[g_l^m\cos(m\phi) + h_l^m\sin(m\phi)]P_l^m(\cos\theta) \\ V_{ext} &= a\sum_{l=1}^{\infty}\sum_{m=0}^{l}\left(\frac{r}{a}\right)^{l}[q_l^m\cos(m\phi) + s_l^m\sin(m\phi)]P_l^m(\cos\theta). \end{aligned} \tag{3.28}$$

The potential equation consists of two parts; terms in $(a/r)^{l+1}$ to describe the internal potential or field V_{int}, which vanish at infinity and terms in $(r/a)^l$ to describe fields with external origin with respect to the Earth's surface V_{ext}, which must remain finite within a volume of interest [Langel, 1987].

Basically, the method here relies on the evaluation of the geomagnetic potential at the core–mantle boundary, derived from quantities measured at the Earth's surface. The problem is ill-posed due to the fact that the geomagnetic potential is not known entirely at Earth's surface. In practice the expansion of (3.26) is truncated at a certain level, and therefore the geomagnetic potential is unknown from that level onward. Formally, there are infinite (truncated) description of the potential at the core–mantle boundary, which would describe the observed geomagnetic field at Earth's surface. One assumption which may help here, is to declare any Gauss coefficient greater than the truncation level as equal to zero. One other approach is developed in the following.

The Gauss coefficients are expanded in time as spline functions $S(\xi)$

$$g_l^m(t) = S(\xi) \tag{3.29}$$

and likewise for $h_l^m(t)$.

The (natural) spline functions $S(\xi)$ is a linear combination of B(asis)–spline functions

$$S(\xi) = \sum_{j=-1}^{n+1} a_j B_{j,k}(\xi)\,, \tag{3.30}$$

in which a_j are the spline coefficients (g_l^{mj}) and $B_{j,k}(\xi)$ are the piecewise polynomial functions of order k and degree $(k-1)$. For $k=4$ these polynomials are cubics. Unlike other polynomial functions (e.g. Legendre), which are defined on the whole interval $[t_1, t_j]$, a single B–spline function is defined very locally.

A set of non–decreasing real numbers $t_0 \le t_1 \le \ldots \le t_j$ are called the knots, the knot points and the half–open interval $[t_i, t_{i+1})$ is called the knot span. The knots can be considered as division points that subdivide the interval $[t_1, t_j]$ into $n+2$ knot spans. Generally, the i–th B–spline of degree k is defined recursively by the Cox–DeBoor formula

$$B_{i,1} = \begin{cases} 1 & t_i \le \xi < t_{i+1}, \\ 0 & \text{otherwise} \end{cases} \tag{3.31}$$

and for $k \ge 1$

$$B_{i,k}(\xi) = \frac{\xi - t_i}{t_{i+k-1} - t_i} B_{i,k-1}(\xi) + \frac{t_{i+k} - \xi}{t_{i+k} - t_{i+1}} B_{i+1,k-1}(\xi) \tag{3.32}$$

The B–splines are for $k=2$ piecewise linear, for $k=3$ piecewise quadratic, for $k=4$ piecewise cubic, and so on. In particular for cubic B–splines with equidistant knots it follows

$$B_{i,4}(\xi) = \begin{cases} 1/6z^3 & t_{i-2} < \xi < t_{i-1}, \\ 1/6[1+3(1+z(1-z))z] & t_{i-1} \le \xi < t_i, \\ 1/6[1+3(1+z(1-z))(1-z)] & t_i \le \xi < t_{i+1}, \\ 1/6(1-z)^3 & t_{i+1} \le \xi < t_{i+2}, \\ 0 & \text{otherwise}\,, \end{cases} \tag{3.33}$$

where d is the distance between the knots (e.g. $d = t_{i+1} - t_i$) and z the distance between the t_{i+k} and ξ divided by d (e.g. $z = (\xi - t_{i+2})/d$). Within the i-th knot span only k B–splines $B_{i-k+1}, \ldots, B_{i,k}$ are non–zero and normalized to 1. Also note, that the B–spline is symmetric about the knot around which it is centered.

The motivation of using splines is mainly because of three reasons:

1. basis functions are non–zero on a few adjacent subintervals, therefore ideal for recovering very localized features, i. e. temporal discontinuities,
2. the spline curve $S(\xi)$ has a minimum curvature in the interval $[t_1, t_j]$,
3. the banded form of the normal equation matrix makes large problems feasible for given computational power, what is the most important reason.

In matrix form (3.30) reads

$$S = \mathbf{A} \cdot \mathsf{x}\,, \tag{3.34}$$

where the vector x contains the spline coefficients a_j. $\mathbf{A}$ is symmetric and positive definite, and therefore nonsingular[5]. Cholesky decomposition is commonly used to solve linear systems equations with this matrix $\mathbf{A}$; it is a special case of the LU–decomposition. These decompositions are based on the following considerations: a symmetric and positive definite matrix $\mathbf{A}$ can be formulated as

$$\mathbf{A} = \mathbf{L}\mathbf{U}, \tag{3.35}$$

$\mathbf{L}$ and $\mathbf{U}$ are quadratic matrices, where for the upper matrix $\mathbf{U}$ all elements below the diagonal are equal to zero

$$\mathbf{U}_{ij} = 0 \text{ for } i > j \qquad \mathbf{U} = \begin{pmatrix} U_{11} & U_{12} & U_{13} & \ldots & U_{1n} \\ & U_{22} & U_{23} & \ldots & U_{2n} \\ & & U_{33} & \ldots & U_{3n} \\ & & & & \vdots \\ & & & & U_{nn} \end{pmatrix} \tag{3.36}$$

whereas for the lower matrix $\mathbf{L}$ all elements above the diagonal are equal to zero

$$\mathbf{L}_{ij} = 0 \text{ for } i < j \qquad \mathbf{L} = \begin{pmatrix} L_{11} & & & & \\ L_{21} & L_{22} & & & \\ L_{31} & L_{32} & L_{33} & & \\ \vdots & & & & \\ L_{n1} & L_{n2} & L_{n3} & \ldots & L_{nn} \end{pmatrix} \tag{3.37}$$

Equation (3.34) can then be written as

$$\mathbf{A} \cdot x = \mathbf{L} \cdot (\mathbf{U} \cdot x) = S\,, \tag{3.38}$$

[5] $\mathbf{A}$ becomes positive definite, when appropriate regularization is applied, which shifts zero eigenvalues by the value of damping

and solving for x takes two steps, first by solving the intermediate vector y

$$\mathbf{L} \cdot y = b \tag{3.39}$$

and then solving (backward substitution)

$$\mathbf{U} \cdot x = y\,. \tag{3.40}$$

The forward substitution of (3.39) reads

$$\begin{aligned} y_1 &= \frac{b_1}{L_{11}} \\ y_i &= \frac{b_i - \sum_{j=1}^{i-1} L_{ij} y_j}{L_i i} \qquad i = 2, \ldots, n \end{aligned} \tag{3.41}$$

and the backward substitution

$$\begin{aligned} x_n &= \frac{y_n}{U_{nn}} \\ x_i &= \frac{y_i - \sum_{j=i+1}^{n} U_{ij} x_j}{U_i i} \qquad i = n-1, n-2, \ldots, 1\,. \end{aligned} \tag{3.42}$$

The Cholesky decomposition takes advantage of the fact that if $\mathbf{A} = \mathbf{A}^{\mathsf{T}}$ then (3.35) is

$$\mathbf{A} = \mathbf{U}^{\mathsf{T}}\mathbf{U}\,, \tag{3.43}$$

i. e. upper and lower triangular matrices are the same and the forward and backward substitution read

$$\begin{aligned} y_1 &= \frac{b_1}{U_{11}^{\mathsf{T}}} \\ y_i &= \frac{b_i - \sum_{j=1}^{i-1} U_{ij}^{\mathsf{T}} y_j}{U_i^{\mathsf{T}} i} \qquad i = 2, \ldots, n \\ x_n &= \frac{y_n}{U_{nn}} \\ x_i &= \frac{y_i - \sum_{j=i+1}^{n} U_{ij} x_j}{U_i i} \qquad i = n-1, n-2, \ldots, 1\,. \end{aligned} \tag{3.44}$$

Utilizing the method of stochastic inversion to construct solutions to the simultaneous inverse problem requires minimizing (3.25). In fact it means minimizing of the misfit between model and observation and a model norm $\mathsf{m}^{\mathsf{T}}{\mathbf{C}_{\mathbf{m}}}^{-1}\mathsf{m}$. $\mathbf{C}_{\mathbf{m}}$ is chosen to reflects the spatial and temporal a priori variance in the field at the CMB.

One source of prior information about B_{r} at the CMB is the heat flux bound Q, derived from the Maxwell equations and Ohm's law. Gubbins [1975] analysis leads to the condition

$$F(\mathbf{B}_r) = \sum_{l=1}^{\infty} \frac{(l+1)(2l+1)(2l+3)}{l} \sum_{m=0}^{l} [(g_l^m)^2 + (h_l^m)^2] \leq Q \qquad \text{for } r = c\,, \tag{3.45}$$

where the heat flux bound is given by

$$Q = \Upsilon \frac{\kappa \mu_0^2}{4\pi c}, \tag{3.46}$$

with Υ the rate of the ohmic heat production of the core, κ the maximum electrical conductivity in the core, and μ_0 the magnetic permeability of vacuum. Taking $\kappa = 3 \times 10^5 \mathrm{Sm}^{-1}, c = 3.48 \times 10^6 \mathrm{m}, \mu_0 = 4\pi \times 10^{-7} \mathrm{NA}^{-2}$ and $\Upsilon = 3 \times 10^{13} \mathrm{W}$ (the observed heat flow out of the Earth's surface) gives

$$Q = 3 \times 10^{17} \mathrm{nT}^2 \,. \tag{3.47}$$

However this value of Q may vary between 3×10^{16} and $3 \times 10^{18} \mathrm{nT}^2$ depending on Υ[6], but it should give an upper limit.

The spatial regularization condition is then

$$S = \frac{4\pi}{(t_2 - t_1)} \int_{t_1}^{t_2} F(\mathbf{B}_r)\, dt = \mathsf{m}^\mathsf{T} N_S^{-1} \mathsf{m} \leq Q \,. \tag{3.48}$$

The condition to adjust the temporal behaviour of the solution, one seeks to minimize the norm

$$T = \frac{1}{(t_2 - t_1)} \int_{t_1}^{t_2} \oint_{CMB} (\partial \mathsf{B_r}/\partial t)^2 \, dS \, dt = \mathsf{m}^\mathsf{T} N_T^{-1} \mathsf{m} \,, \tag{3.49}$$

with

$$\oint_{CMB} (\partial \mathsf{B_r}/\partial t)^2 dS = \sum_{l=1}^{\infty} (l+1) \left(\frac{a}{c}\right)^{2l+4} \sum_{m=0}^{l} [(\dot{g}_l^m)^2 + (\dot{h}_l^m)^2] \tag{3.50}$$

Finally the regularization condition is given by

$$\mathbf{C_m}^{-1} = \lambda_s N_S^{-1} + \lambda_t N_T^{-1} \,, \tag{3.51}$$

whith the damping parameters λ_s and λ_t [c.f. Bloxham & Jackson, 1992]. Other possible regularization conditions are listed in table (3.3).

3.4.4 Extension of the Bloxham–Jackson formalism

The essential new aspect of this study is the invoking of a priori information for the geomagnetic field at the endpoints of the time interval 1980 and 2000.

$$\Theta(m) = (\mathsf{y} - \mathbf{A}\mathsf{m})^\mathsf{T} \mathbf{C_e}^{-1} (\mathsf{y} - \mathbf{A}\mathsf{m}) + \mathsf{m}^\mathsf{T} \mathbf{C_m}^{-1} \mathsf{m} + \mathbf{C_a}^{-1} \,. \tag{3.52}$$

[6]For simplicity, here is assumed that the heat flux out of the core is equal to that, which is observed at the Earth's surface. In fact, the largest portion of the heat flux comes from the crust. Labrosse [2002] gives estimates of the heat flux of the core ranging form 2×10^{12}W to 10×10^{12}W.

	integral to minimize	Norm function $f(r_0, l)$	Comments
(a)	$\oint \mathbf{B}^2 dS \vert r = r_0$	$(l+1)\left(\frac{a}{r_0}\right)^{(2l+4)}$	Mean square field intensity
(b)	$\oint \mathsf{B_r}^2 dS \vert r = r_0$	$\frac{(l+1)^2}{2l+1}\left(\frac{a}{r_0}\right)^{(2l+4)}$	Mean square radial field
(c)	$\oint (\nabla_h \mathsf{B_r})^2 dS \vert r = r_0$	$\frac{l(l+1)^3}{2l+1}\left(\frac{a}{r_0}\right)^{(2l+6)}$	Mean square of horizontal derivative of radial field

Table 3.2: Smoothing norms for magnetic field inversion after Holme & Bloxham [1996].

The magnetic field is constrained by models derived from high-quality satellite vector data m_0^{1980} and m_0^{2000}, respectively. The a priori model regularization condition is then

$$\begin{aligned} \oint \mathbf{B}^2 dS|_{t=1980} &= (\mathsf{m} - \mathsf{m}_0^{1980})^\mathsf{T} N_1^{-1} (\mathsf{m} - \mathsf{m}_0^{1980}) \\ \oint \mathbf{B}^2 dS|_{t=2000} &= (\mathsf{m} - \mathsf{m}_0^{2000})^\mathsf{T} N_2^{-1} (\mathsf{m} - \mathsf{m}_0^{2000}) \end{aligned} \tag{3.53}$$

giving

$$\mathbf{C_a}^{-1} = \lambda_1 N_1^{-1} + \lambda_2 N_2^{-1}. \tag{3.54}$$

The advantage of using satellite based models is their higher spatial resolution of the magnetic field. At the endpoints the time–dependent field model is constrained by a priori field models in 1980 and 2000 [Cain et al., 1989; Olsen, 2002]. These models are constructed by tapering single epoch models form satellite data. The taper is defined by the minimization conditions

$$(\mathbf{B} - \mathbf{B}_0)^2 \tag{3.55}$$

at the Earth's surface where $\mathbf{B}_0$ is the MAGSAT or ØRSTED model including crustal field, respectively. But also minimise

$$(\mathbf{B})^2 \tag{3.56}$$

at the CMB. Integration over the Earth's surface leaves to minimize

$$\sum_{l,m} (l+1)\left((g_l^m - {g_l^m}_0)^2 + (h_l^m - {h_l^m}_0)^2\right) + \lambda(l+1)\left(\frac{a}{c}\right)^{(2l+4)}\left((g_l^m)^2 + (h_l^m)^2\right), \tag{3.57}$$

where λ is a Lagrange multiplier. Differentiating with respect to the Gauss coefficients, setting each differential to zero, and rearranging, this gives

$$g_l^m = \frac{{g_l^m}_0}{1 + \lambda\left(\frac{a}{c}\right)^{(2l+4)}}. \tag{3.58}$$

The high-degree coefficients are damped much more than the low degree ones. The degree of fall-off is obviously controlled by λ.

3.4.5 Modelling and selection of damping parameter

The modelling is undertaken in an iteratively re-weighting scheme comprising four steps:

1. A first guess model is computed weighting all data with the same uncertainty (5 nT).
2. The second step is to derive the deviations of the data from the first guess model to compute new weights for the data. Data, which scatter extremely from the model are down-weighted.
3. Now a model is derived from the newly (re-) weighted data set.
4. In the final step, data are discarded which deviates more then 2 σ from the second model. From this reduced data set the final model is derived.

Models were derived for ranges of temporal damping parameters $3.5 \cdot 10^2 \leq \lambda_t \leq 3.5 \cdot 10^{-5}$. In order to choose an appropriate damping the objective function (3.52) might be transformed into

$$F(\kappa) = E + \kappa N \,, \tag{3.59}$$

where E is the misfit and N the solution norm. κ is a positive parameter which determines the relative importance of E and N. There are two extremal situations 1.) $\kappa = 0$, a priori information of the model are neglected 2.) $\kappa = \infty$, the data are ignored completely and the minimized norm gives $\mathsf{m} = 0$. Therefore minimizing $F(\kappa)$ is a compromise between these extrema. A satisfying solution can be read from a plot E versus N.

The selection of the spatial damping parameter is founded by an a priori knowledge of core physics, which is independent of the data set inverted. The condition (3.45) assigns to a sequence of Gauss coefficients in the model space the probability distribution which has a zero mean and the variance

$$\langle |g_l^m(c)|^2 \rangle = Ql(l+1)^{-1}(2l+1)^{-1}(2l+3)^{-1} \,. \tag{3.60}$$

This limits each Gauss coefficient only by requiring it not to produce by itself more ohmic heat in the core than is observed in the surface heat flow [Backus, 1988]. Following Backus [1988] gives a

$$\lambda_s = 8 \times 10^{-17} \mathrm{nT}^2$$

with $Q = 3 \times 10^{18} nT^2$. This Bayesian approach facilitates error estimates and to quantify the range of permissible models (see section 3.5.4). The solution for unchanged λ_s will be that where the trade–off curve has its knee. Figures (3.4) show the three trade–off curves of the temporal norm vs misfit for the three different values of departure damping. The damping controlling the departure from the a priori models is chosen in the way that the endpoint models for 1980.0 and 2000.0 are identical with the satellite main field models.

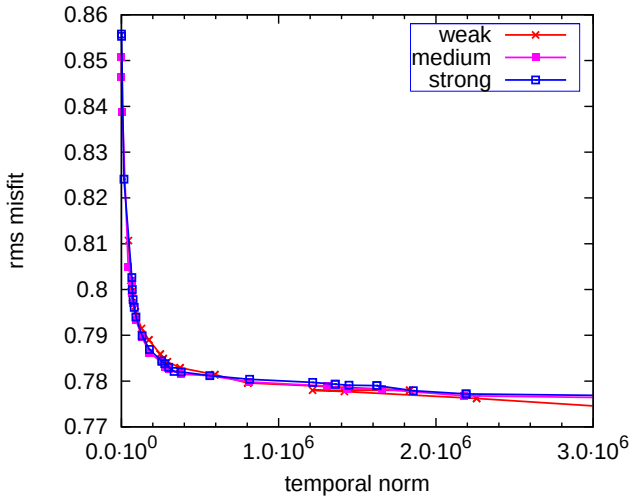

Figure 3.4: Trade–off curves of the temporal norm vs. misfit for different spatial damping weak: $\lambda_\Gamma = 4.0 \cdot 10^1$, medium: $\lambda_\Gamma = 4.0 \cdot 10^3$ and strong: $\lambda_\Gamma = 4.0 \cdot 10^6$.

Treating this constraint strongly obliges the degrees 10 to 14 of the main field for an epoch in between (i. e. 1991) to have significant information. Figure (3.5) shows power spectra of main fields for the epochs 1980, 1991 and 2000 and also for the main field models derived from MAGSAT or ØRSTED satellite data, respectively. Here the departure constraint is relaxed, the power spectrum for the 1991 main field model falls off at order 10. What means that the spectrum from this degree onward is comparable with those of a white noise process, where the Gauss coefficients behave erratically. Figure (3.6) shows when the departure constraint is applied strongly and the spectrum of the main field model for 1991 tends to fall off around degree 14.

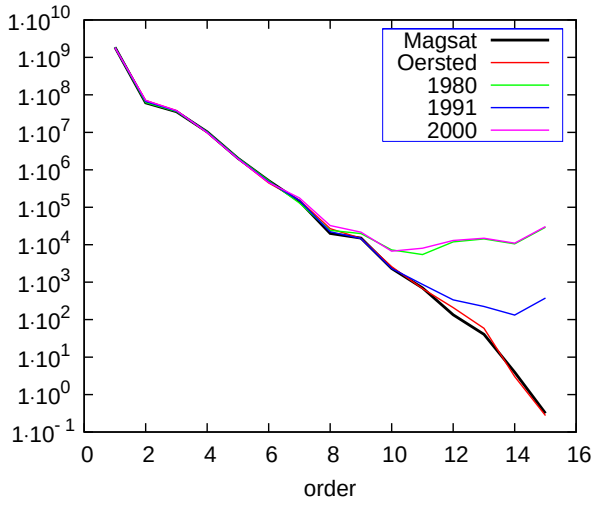

Figure 3.5: Spectra of main field of a time–dependent model for three different times 1980, 1991 and 2000 in comparison with the satellite main field models from MAGSAT and ØRSTED. The departure damping is chosen to be weak ($\lambda_\Gamma = 6.0 \cdot 10^{-1}$).

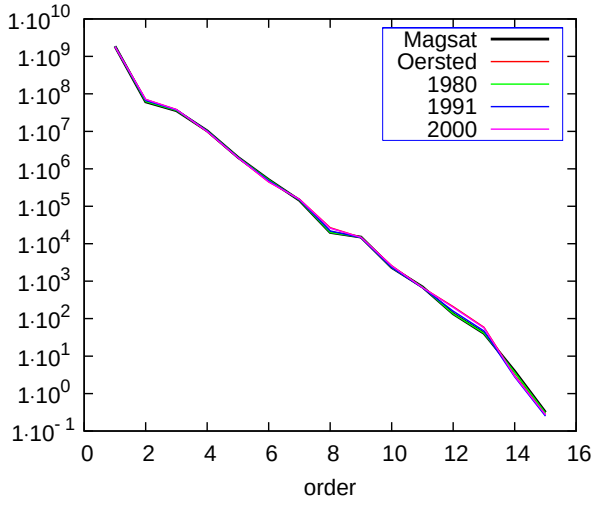

Figure 3.6: Spectra of main field of the preferred time–dependent model for three different times (1980, 1990 and 2000), where the departure damping is chosen to be strong ($\lambda_\Gamma = 6.0 \cdot 10^{6}$) in comparison with the satellite main field models from MAGSAT and ØRSTED.

Statistics	
Number of data	70939
Number of Splines	13
Damping	
spatial λ_s	0.3×10^{-17}
temporal λ_t	0.3×10^{-3}
departure λ_Γ 1980.0	0.60×10^{3}
departure λ_Γ 2000.0	0.60×10^{3}
Model norms	
spatial	3.56×10^{8}
temporal	3.69×10^{5}
departure	1.93×10^{9}
RMS Residual	0.782

Table 3.3: Statistics and parameters of the time–dependent modelling

3.5 Results

In the previous section I developed a time–dependent model of the main field and its secular variation for the period 1980 - 2000. The results of the time–depending modelling have been achieved under the assumption that the mantle is an insulator, except for possibly at its base, where a 200 km thick layer exists, known as D″. This layer is likely to be a boundary layer chemically and physically distinct from the rest of mantle. If this layer is conducting, then it would have only a little effect on the results of the downward continuation. In fact Voorhies [2000] could show that the core radius inferred from seismic methods differ by only 62 km from that of geomagnetic estimates. However, a conductivity jump across the CMB is expected, what causes that the horizontal components are discontinuous there. The radial component remains continuous. Therefore, in the next sections only results of the preferred model solution for the radial component of the magnetic field and secular variation are discussed.

3.5.1 Morphology of the magnetic field at the core-mantle boundary

In figure (3.7) the radial component of the magnetic field is shown for 5 epochs 1980.0, 1985.0, 1990.0, 1995.0 and 2000.0. The maps show basically a dipolar structure which is riddled with localized concentrations of high flux separated by larger areas of weaker flux. All features are apparently equally present in all maps.

In order to characterize the features a similar nomenclature as in Bloxham et al. [1989] is applied. Regions of high field are marked as N (normal) and R (reversed). The term reversed means here that these patches have a reverse flux with respect to their vicinity and are bounded by a null–flux curve. See for locations of these features figure (3.8). Bloxham et al. [1989] found in their analysis that patches have drifted by as much as 60° in longitude over a period of 265 years. This rate might be detected for some of the R patches. Table 3.4 lists the positions, size and the flux through the R patches.

The most prominent such feature is the patch R2 beneath the southern Atlantic. This dog–bone shaped patch with reversed flux (with respect to its surroundings) has grown continuously since it appeared around 1970 [Jackson et al., 2000]. In 1984 the patch merged with a minor patch R1 underneath Antarctica and finally united with the major positive flux pattern of the northern hemisphere in 1997. The size, position and flux of the other reversed flux patches of the southern hemisphere remained almost unchanged during this period. There are also reversed flux patches in the north polar region. Initially a single patch R8 existed, which split into two individual patches R8A and R8B in 1986.

It should be mentioned that some of the high flux patches remained steady over the last 20 years namely N1 and N2 in the southern and N9 and N10 in northern hemisphere. Certainly the period is too short to infer anything definitively, but as pointed out by Bloxham et al. [1989] these patches remained steady over the last 300 years and were unaffected by the westward drift, which should have shifted these by about 60°. This fact strongly suggests that these features are linked to a steady dynamo process and are

Patch	Year	Long.	Lat.	Flux [MWb]	$\langle B_r \rangle$ [nT]	Area [km^2]
R1	1980	53.042	-78.746	45.8	81650.4	1.8×10^{12}
	2000	–	–	–	–	–
R2	1980	345.940	-46.246	603.3	137814.3	14.6×10^{12}
	2000	–	–	–	–	–
R3	1980	344.833	-19.600	77.5	80684.6	3.2×10^{12}
	2000	341.263	-20.136	119.4	110936.9	3.5×10^{12}
R4	1980	241.259	-18.106	20.7	34081.6	2.0×10^{12}
	2000	240.895	-17.555	19.3	35121.4	1.8×10^{12}
R5	1980	165.165	19.727	46.1	60198.2	2.5×10^{12}
	2000	165.076	20.109	40.2	56636.3	2.3×10^{12}
R6	1980	316.803	25.524	10.9	34746.5	1.0×10^{12}
	2000	313.089	25.090	6.2	24390.6	0.8×10^{12}
R7	1980	188.700	44.285	16.1	28328.0	1.8×10^{12}
	2000	190.656	45.029	30.8	42026.2	2.4×10^{12}
R8	1980	317.882	80.531	34.9	46351.6	2.5×10^{12}
R8A	2000	31.141	71.053	13.0	32394.2	1.3×10^{12}
R8B	2000	258.085	83.815	65.6	71798.9	3.0×10^{12}

Table 3.4: Listing of individual reverse flux patches for the epochs 1980 and 2000

most likely convection columns along 120° longitude west and 120° longitude east. These columns originally postulated by Busse [1970], are aligned parallel with the rotation axis of the Earth and touch the inner core at the equator. There are substantiated arguments that a well–ordered arrangement of columnar vortices with helical velocity distribution exists in a highly conducting liquid core. This idea has been proven convincingly by model experiments of Carrigan & Busse [1983]. However, there are also arguments against this flow pattern, which basically rely on the fact that these rolls collapse, when the Lorentz force act on [Fearn, 1998]. That might be a reason, why a third column, which was expected on grounds of geometrical reasons [Gubbins & Bloxham, 1987] is not traceable.

Testing the frozen flux hypothesis

The hypothesis that the secular variation on short time scales is entirely given by the advection term of the induction equation is known as the frozen flux hypothesis [Alfvén, 1942; Roberts & Scott, 1965]. For reasons which will be discussed in more detail in section 4.1, the radial field at the core–mantle boundary has to satisfy certain conditions, such as

$$F_i = \int_S \partial_t B_r dS = 0\,, \tag{3.61}$$

the flux through a patch S on the core surface bounded by a contour of zero radial field must be constantly. A valid test of this hypothesis is to compare the changes of the flux

through individual flux patches. The dog bone patch R2 shows in 17 years (1980 – 1997) a change of its flux of about -20 MWb, in good agreement with Bloxham [1988], who found -25 MWb in 11.5 years. Another patch which shows drastic increase of flux is the St. Helena patch R3: within 20 years it changed its flux by 40 MWb, also in agreement with Bloxham [1988]. In the northern hemisphere the North Pole patches show considerable changes in size and flux. Beside the analysis of the flux through individual flux patches, there is one further possibility to test the consistency of the model, and consequently of the data, with the frozen flux hypothesis. A necessary, but not sufficient condition for the frozen flux to apply is that the quantity of the unsigned flux integral

$$\oint_{CMB} |B_r| dS = \text{const.}\,, \tag{3.62}$$

integrated over the surface of the core-mantle boundary should not change with time. Therefore a computation of this integral for each year is carried out and the differences for successive years are averaged of the whole period. In figure (3.9) the averaged differences, in fact time derivatives of eq.(3.62), is shown. The given error-bars are the sample variance of the integral for a specific truncation degree. For truncation degrees less than 9 a conservation of unsigned flux is not achieved, for higher degree this seems to be achieved within error margins, but it should be mentioned that the higher degrees of the model more and more reflects the a priori beliefs, i.e. spatial and temporal damping.

Holme & Olsen [2005] computed the unsigned flux integral as here, but using models of the secular variation and main field based on satellite data. Their results for unsigned flux integral concur with those here discussed up to truncation degree 9 or 10. Beyond that the integrals differ due to different modelling approaches. The conclusion on this might be that diffusion is more effective on larger length scales than on smaller ones.

This all gives rise to the presumptive evidence that the model can resolve the diffusion and further indicates a violation of the frozen flux hypothesis even on sub-decadal time scales [Bloxham & Gubbins, 1986], but it does not inevitably mean that the frozen flux assumption has to be abandoned, rather that the assumption is inadequate to explain the resolved secular variation entirely. Thus the observed secular variation in the southern hemisphere may be partly due to diffusion. It is most likely that diffusion might be enhanced by expulsion of toroidal flux by upwelling [Bloxham, 1986]. This mechanism also operates on the Sun, where regions of toroidal flux expulsion are seen as Sun spots. I will return to this topic at the end of the section 4.2.3, where I discuss flux expulsion due to fluid upwelling.

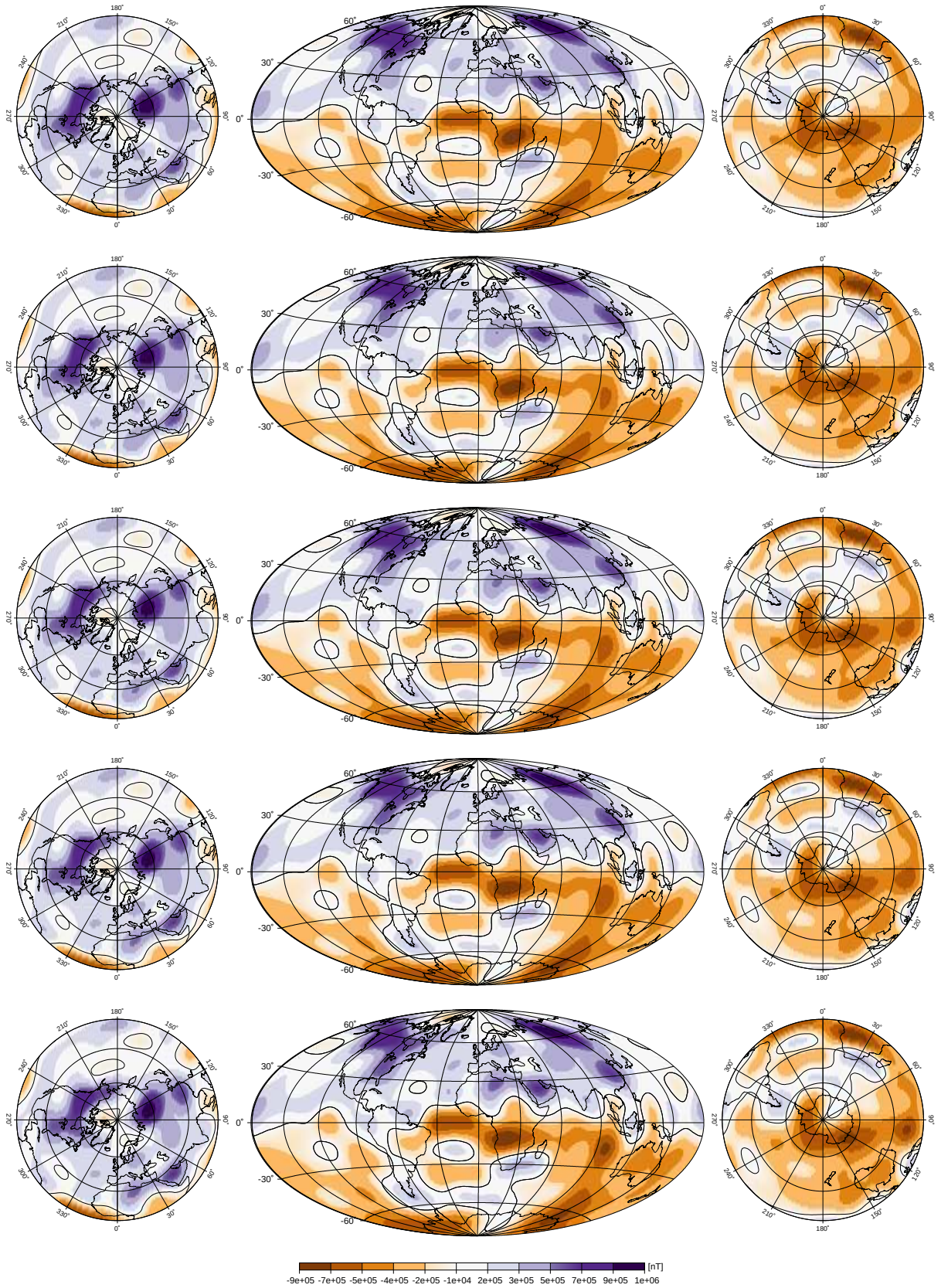

Figure 3.7: Radial component of the geomagnetic field at 1980, 1985, 1990, 1995 and 2000 from top to bottom (Mollweide equal–area projection).

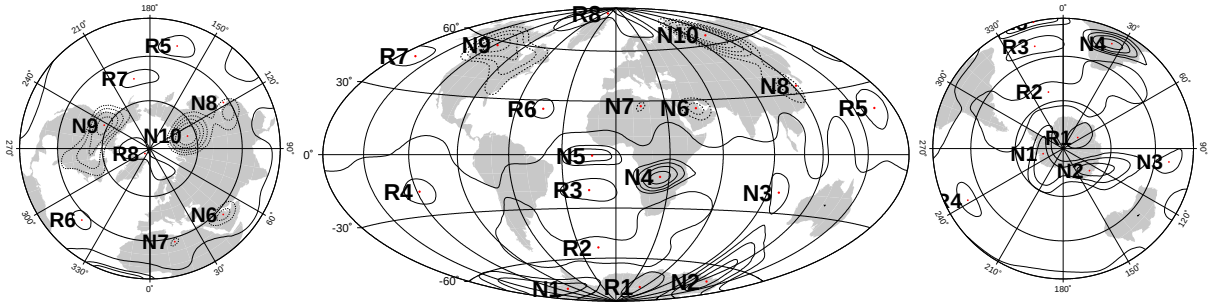

Figure 3.8: Radial component of the geomagnetic field, with key features of the field: reverse flux patches are labeled R1 – R8, normal flux patches as N1 – N10 (Mollweide equal–area projection).

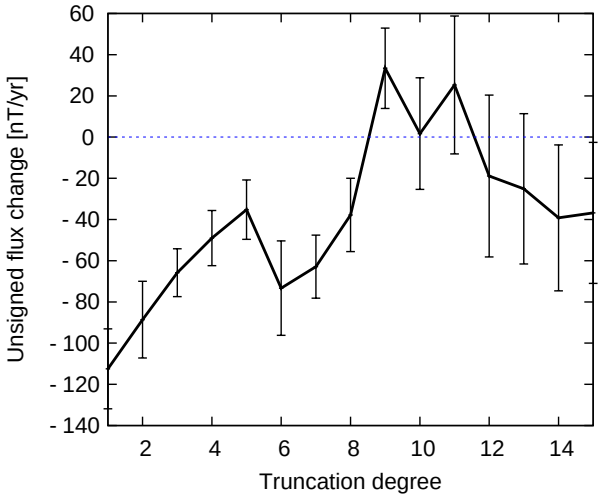

Figure 3.9: The mean change of the unsigned flux integral for 1980 – 2000.

3.5.2 The secular variation and geomagnetic jerks during this period

I now discuss the details of the modelled secular variation using the comparison between the model and the secular variation estimates of monthly means at a number of permanent observatories. In figure 3.10 this comparison is shown. The observatories are chosen to represent different latitudinal regions, polar as well as mid and low latitudinal regions.

The time–dependent model renders the observed secular variation well and even makes it possible to detect data errors such as unreported base line jumps, see for instance the graphs of Y and Z in Hermanus around 1997.

The model reveals a short term variability on sub–decadal scale for all components of the presented observatories in mid and low latitudes. What is also appearing, the secular variation of observatories at nearly the same latitude, but different longitude, show different monotonic behaviour, for example Niemegk and Newport or MBour and Pamatai. Only the monotonic behaviour of Hermanus and Eyrwell are comparable. These periods of increasing and decreasing secular variation are delineated by sudden events, where the slope of the secular variation changes its sign – geomagnetic jerks. These jerks are mainly identified on ground of the secular variation of the Y–component, because this component is less influenced by external field variation. I will discuss these events first on ground of the Y–component of the secular variation. The first of the geomagnetic jerks, which occurred during this period is visible around 1983 in the Y component of Hermanus [Dowson et al., 1987; Kotzé, 2003]. It was originally thought, that this event was only discernible in the recordings of observatories in the southern hemisphere, but the extent of this event seems to reach even mid latitudinal observatories as MBour. In 1990 a jerk was reported [Macmillan, 1996], mainly visible in East components of European observatories such as Niemegk. The third jerk during this period took place around 1999, also visible in European observatories [Mandea et al., 2000].

So far a discussion of the secular variation recorded in polar observatories has been left out, certainly because of the jammer of external field variation, which makes a discussion of secular variation at those sites not easy. However, the secular variation modelled by the time–dependent field model seems to be unaffected by this external noise. The variance of the observed secular variation is by a factor of 10 larger than the modelled secular variation. The 1983 and 1999 jerks are clearly seen in the Y–components of Resolute Bay and Scott Base, whereas the 1990 event is missed out in both observatories.

Beside the known jerks, there are further sudden events present in the secular variation. Clearly around 1986 and 1995, jerks happened in the southern hemisphere, whose amplitude fades away at more northly observatories.

Also, the other components show distinct variation in this 20 years time span. My belief is that the secular variation in Y should not be discussed independently of those in X and Z, but caution should be paid to induction effects which might be present in the Z–component. However, the short term variations (shorter then 5 years) caused by external field variation seem to be not present in the model. This will be much clearer when considering figure 4.15. In these figures, without the large amplitude variation of the

secular variation estimates, the modelled secular variation (black line) is easier to view. Whereas the secular variation in X and Z of Resolute Bay and Scott Base is more or less featureless, for the others it is rich on details. All non–polar observatories show a distinct feature in Z around 1982 which also seen in dX/dt of Mbour and Pamatai. This feature is chronologically before the 1983 jerk. Also for the 1991 jerk, there are sharp changes in the secular variation of X and Z for most of the observatories visible, which leading or trailing the 1991 event by about one year.

In figure 3.11 the secular variation of the radial component of the magnetic field at the core-mantle boundary is shown. The structure is very heterogeneous. The largest amplitudes of the secular variation appear within a band $\pm$ 30° about the equator. These patches remain steady within some margins during the 20 years.

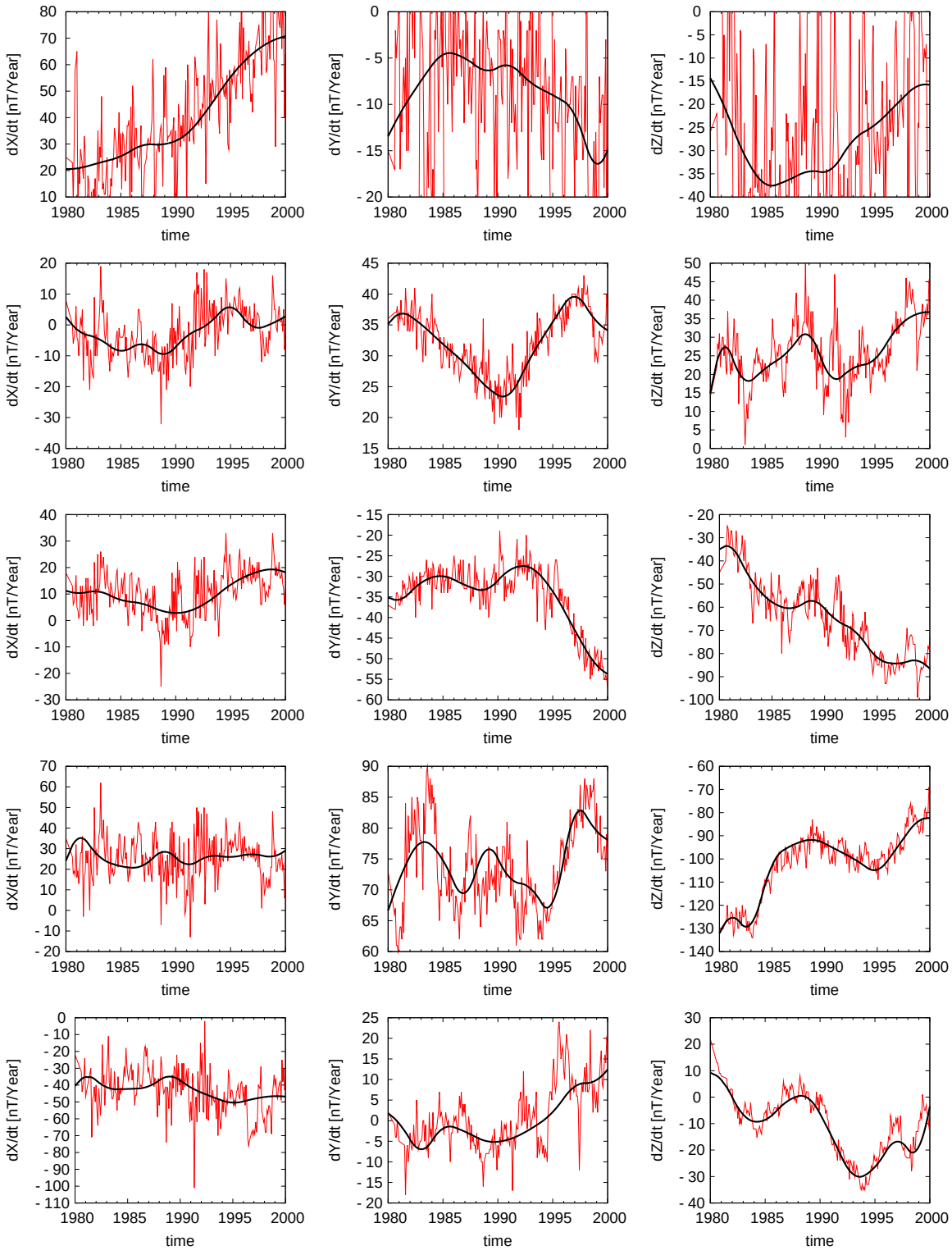

Figure 3.10: Comparison of modeled secular variation (black line) with the secular variation estimates at selected permanent observatories (red line). From left to right: $dX/dt, dY/dt, dZ/dt$. From top to bottom: Resolute Bay (Canada), Niemegk (Germany), Newport (USA), MBour (Senegal), Pamatai (French Polynesian).

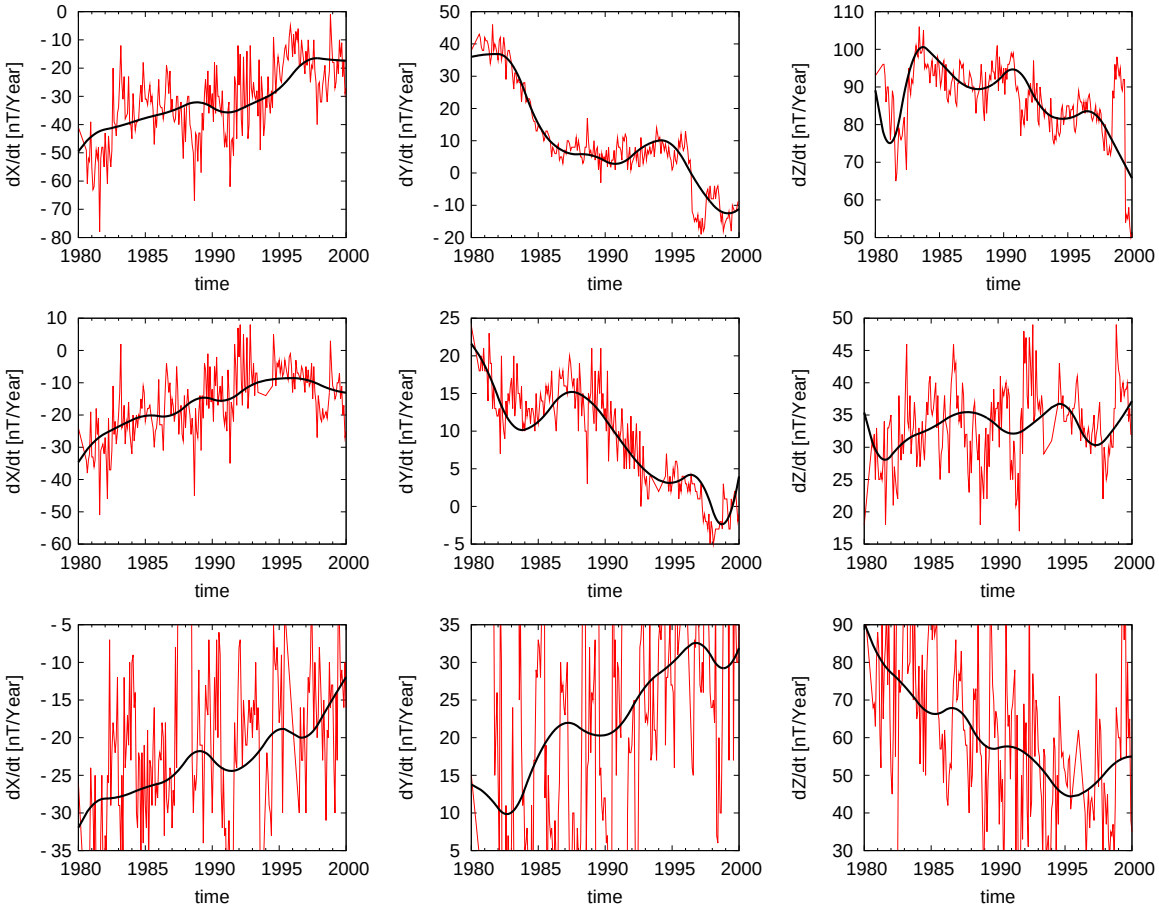

Figure 3.10: (*continued from previous page*) From top to bottom: Hermanus (South Africa), Eyrwell (New Zealand) and Scott Base (Antarctica).

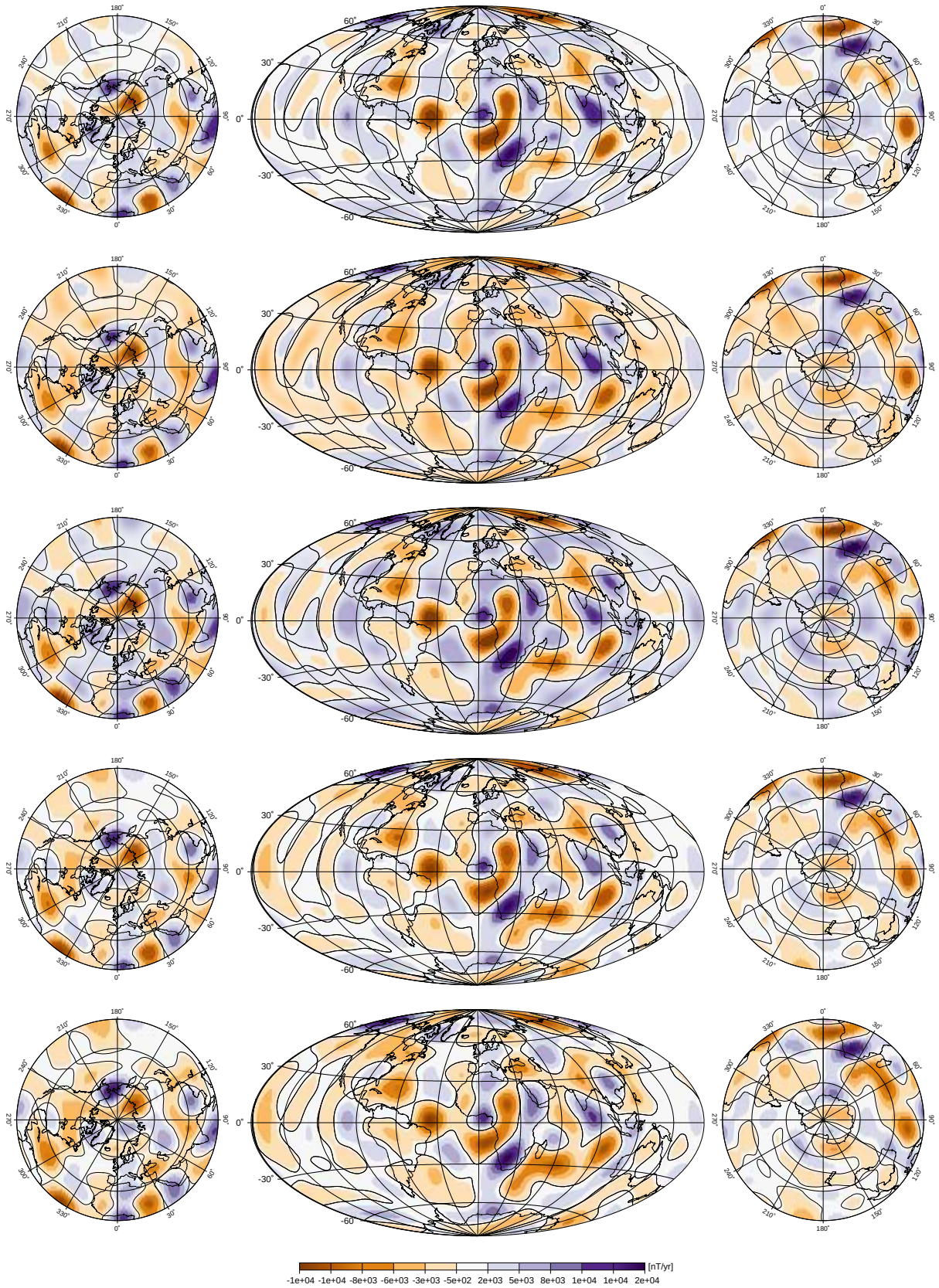

Figure 3.11: Radial component of the secular variation at 1980, 1985, 1990, 1995 and 2000 from top to bottom (Mollweide equal–area projection).

3.5.3 Analysis of the residuals

Figure 3.12 shows the residuals between model and data (left column). The highest residuals appear in both polar observatories Resolute and Scott Base. The variance of the residuals of the Y component in Resolute is smaller than for the other components. The same is true for the observatories in Niemegk, Guam and Hermanus. Attention should be paid to the base line jumps in Hermanus in Y around 1997 and in Z around 1998. The curves in the middle column depict the autocorrelation function of the residuals in each component. The autocorrelation function is given by

$$A(\tau) = \frac{1}{N-r} \sum_{t=1}^{N-\tau} (x_t - \langle x_t \rangle)\ (x_{t+\tau} - \langle x_{t+\tau} \rangle) / (x_t - \langle x_t \rangle)^2 \,, \tag{3.63}$$

where x_t is time series of the residuals and τ is the shift. Variations with periods multiple of τ will have a local maximum in $A(\tau)$. Apparently the dip at $\tau = 12$ in all curves is due to the processing of the data, to derive the secular variation estimates (cf. eq.3.2). The residuals show basically the same periodicities, with external origins as already discussed in chapter 1 for single observatories. The cross correlation function of the residuals against each other is given in the right column of the figure and is defined as

$$C(\tau) = \frac{1}{N-r} \frac{\sum_{t=1}^{N-\tau} (x_t - \langle x_t \rangle)(y_{t+\tau} - \langle y_{t+\tau} \rangle)}{\sqrt{(\sum_t (x_t - \langle x_t \rangle)^2)} \sqrt{(\sum_t (y_{t+\tau} - \langle y_t \rangle)^2)}} \,. \tag{3.64}$$

It shows the common correlations of two independent time series x_t and y_t.

The maxima at a lag of zero, which are evident in all figures, indicate that the variation of the residual components are correlated or anti–correlated, respectively. This could mean that the effect of variations, i.e. the semi–annual variation, causes a increase of the residuals lets say in X and Y, what gives a correlation, an increase in the residuals of X and a decrease of the residuals of Y signify an anti–correlation. In detail; Y and Z are anti–correlated in RES and HER and correlated in NGK and SBA, X and Y are correlated in RES and GUA and anti–correlated in NGK and HER, X and Z are anti–correlated in RES and NGK and correlated in HER. This behaviour is mainly due to the different geometries of the ring current at different location and due to different interaction with the distinct magnetization of the crust at each observatory site.

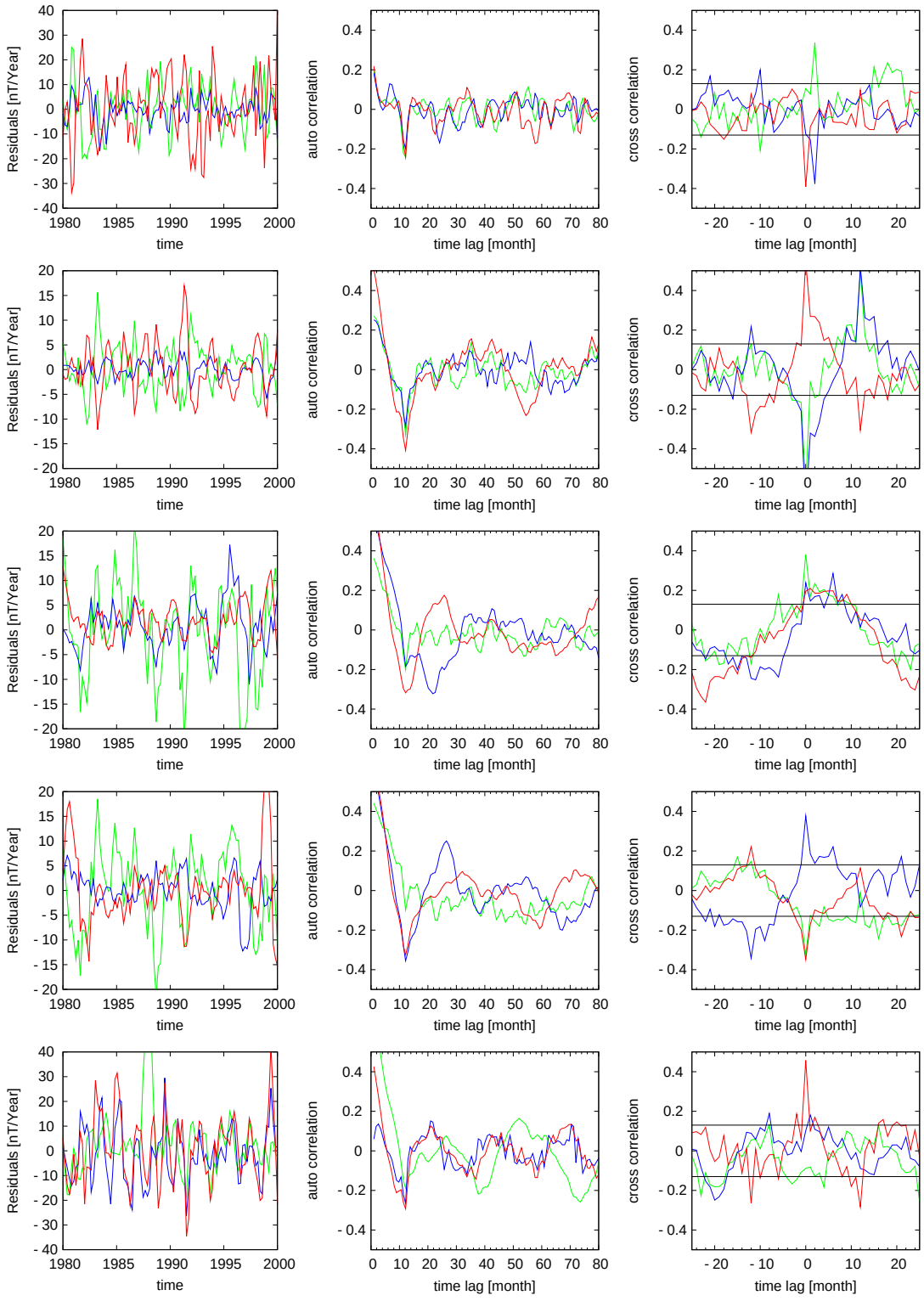

Figure 3.12: In the left column, Residuals between model and the secular variation estimates for Resolute Bay, Niemegk, Pamatai, Hermanus and Scott Base (from top to bottom). Residuals for X in green, Y in blue and Z in red. Autocorrelation function is hown with the same color scheme as for the residuals (middle column). Cross-correlation functions are presented, where the pink line is the cross-correlation function of X and Y, blue for X and Z and red for Y and Z. The horizontal black lines represent the 95 % significance level.

3.5.4 Covariance and Resolution Analysis

Up to now nothing has been stated about the confidence and errors of the gained solution m. Therefore in this section the estimation of the uncertainties in the model is developed and discussed.

The significance of the model parameters are measured by the resolution and covariance matrices. The first gives the extent to which an individual parameter is determined by the data. The second quantifies the uncertainties in the model estimates due to inaccuracies in the measurements. Low resolution and inability to satisfy the data entirely are both sources of uncertainty in the model estimates. The resolution matrix is given by

$$\mathbf{R} = (\mathbf{A}^{\mathsf{T}}{\mathbf{C_e}}^{-1}\mathbf{A} + {\mathbf{C_m}}^{-1} + {\mathbf{C_a}}^{-1})^{-1}\mathbf{A}^{\mathsf{T}}{\mathbf{C_e}}^{-1}\mathbf{A}\,, \tag{3.65}$$

where $\mathbf{C_a}$ and $\mathbf{C_m}$ are defined in (3.54) and (3.51). Ideally, this matrix would be an identity matrix. Due to inadequacy of the data, a regularization scheme must be applied in the solving process to obtain a stable solution this is reflected in the form of the resolution matrix. A resolution near 1.0 signifies that a model parameter is wholly determined by the data, whereas a low resolution means that the model is mostly constrained by the a priori information. Figure 3.13 shows the estimates of the resolution of all 255 model parameters $(g_1^0, \ldots, g_{15}^{15})$ for six different times of the interval. The model is less resolved by the data for 1980 and 2000, rather being constrained by the satellite models. This follows from (3.30) and (3.33), two of the four B–splines are constrained dominantly by the a priori model, whereas the two other are confined by the data. This is not an evidence of a contradiction in the method, it is rather a competition between the different qualities of satellite models and models derived from observatory data. The step–like appearance of these curves indicates that the resolution of the coefficients within a degree is nearly even. Only the coefficients g_8^m, h_8^m (coefficient numbers 64–80) show a slight tendency to a higher resolution with increased order. Gauss coefficients of order and degree greater than 12 (coefficient number > 168) are not resolved at all, only the resolution of the coefficients for 1980 (solid black line) seems not to vanish. This is certainly, because of the high spatial resolution of the MAGSAT satellite model, which originally was derived to degree and order 66.

Following the same considerations as in section 3.4.2, the covariance matrix of the solution is given by

$$\mathbf{C} = \hat{\sigma}^2(\mathbf{A}^{\mathsf{T}}{\mathbf{C_e}}^{-1}\mathbf{A} + \mathbf{C_m} + \mathbf{C_a})^{-1} \tag{3.66}$$

where $\hat{\sigma}^2$ is the misfit of the model

$$\hat{\sigma}^2 = \frac{\mathsf{e}^{\mathsf{T}}{\mathbf{C_e}}^{-1}\mathsf{e}}{N - Tr(\mathbf{R})} \tag{3.67}$$

and N the number of data and $Tr(\mathbf{R})$ the trace of the resolution matrix is equivalent to the degree of freedom. On ground of these considerations one would expect this matrix to be purely diagonal, but in fact the non–diagonal elements are not zero, which means that there is a dependency between coefficients of the same degree but different order.

This dependency is caused by an uneven distribution of observatories. Figure 3.14 shows the covariance estimates of the Gauss coefficients for six different times of the interval, separated by four years. The highest variance is shown by the dipole coefficients for the model in 1980. The variances for the Gauss coefficient of order and degree > 10 (coefficient numbers > 120) are almost zero. Higher damping parameters would force these curves to be equal zero for lower degrees. A zero variance of a model parameter means that this parameter does not deviate from the a priori beliefs. Hence, the high covariance of the dipole term is caused by a deviation of the MAGSAT model from the time–dependent model for 1980. Here again the different quality of satellite and observatory data are noticed.

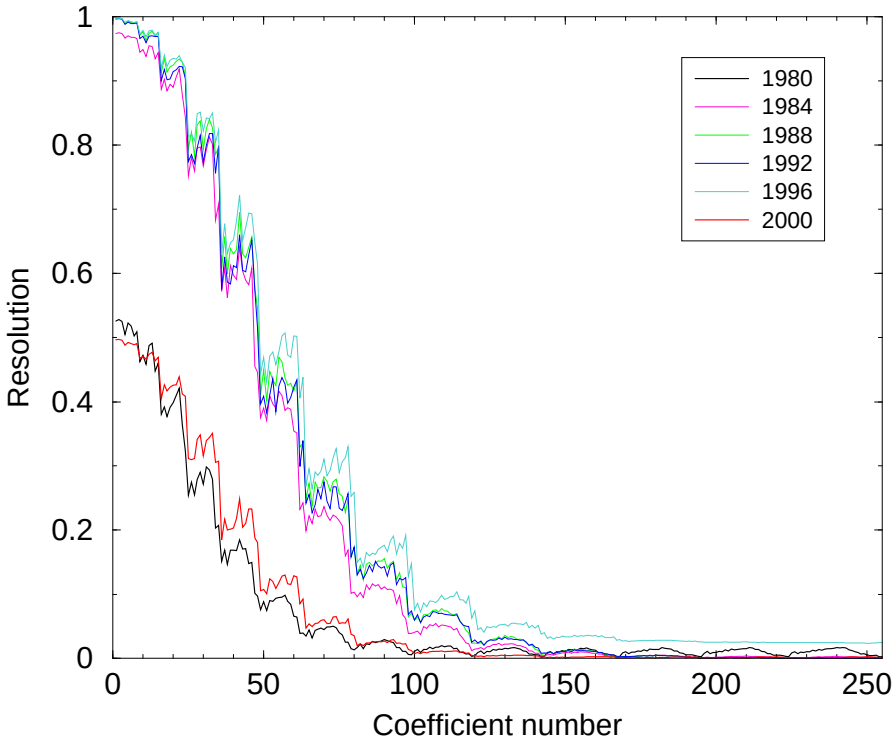

Figure 3.13: Resolution estimates of the Gauss coefficients at different times. Gauss coefficients are ordered with increasing l and m, i. e. $g_1^0, g_1^1, h_1^1, \ldots, h_{15}^{15}$. The resolution of the coefficients with an order greater than 12 are nearly zero this indicates a nearly complete control of these coefficients by the a priori beliefs.

3.5.5 Conclusion

In this chapter, a time–dependent model of the secular variation up to degree and order 15 has been developed for the period 1980 to 2000. The model is constrained at the end points by satellite models derived from MAGSAT, CHAMP and ØRSTED data. It is unique in the (linear square) sense that it minimizes the model norms for the chosen

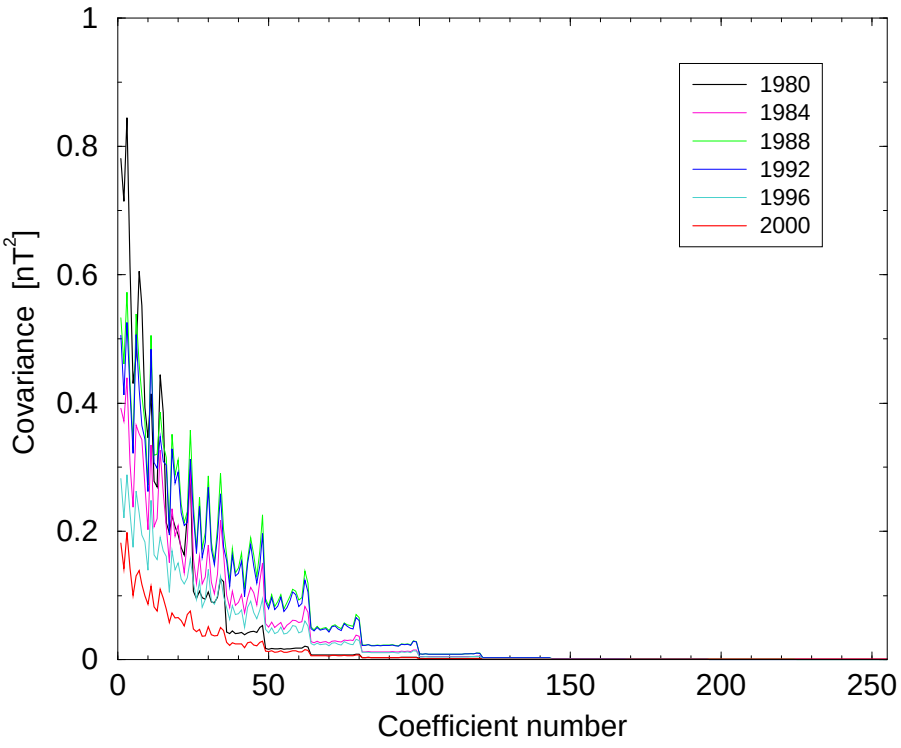

Figure 3.14: Covariance estimates of the Gauss coefficients at six different times. Gauss coefficients are ordered with increasing degree l and order m, i. e. $g_1^0, g_1^1, h_1^1, \ldots, h_{15}^{15}$.

damping parameters. The results of the resolution analysis suggest that the constructed model is resolved by the data at least up to degree 11, hence, a good representation of the real geomagnetic field and its temporal behaviour. The model is a sensitive test to recover the geomagnetic jerks occurred in this period, and even resolve the changes of flux through the core–mantle boundary. Further, it gives evidence for a violation of the frozen flux hypothesis and therefore facilitates a valuable test of this hypothesis, which is essential in the computation of core surface motion.