GEOPHYSICAL INTERPRETATION OF MAGNETIC FIELD MODELS

Estimation of the geothermal heat flux underneath the ice caps in Antarctica and Greenland from magnetic field models based on satellite magnetic data

Cathrine Fox Maule
PhD Thesis
Faculty of Science
University of Copenhagen
November 2005
Contents

Acknowledgments vii

1 Motivation and outline 1

I Earth’s magnetic field, satellite observations and field representations 5

2 Introduction to Earth’s magnetic field 7
  2.1 Earth’s magnetic environment 7
  2.2 Sources of the field 12
    2.2.1 Core field 13
    2.2.2 Crustal field 13
    2.2.3 Magnetospheric ring current 15
  2.3 Mathematical representation of the magnetic field 15
    2.3.1 Spherical harmonic expansions 16
  2.4 Geomagnetic coordinates 19
  2.5 Summary 20

3 Measuring the magnetic field from space 23
  3.1 Satellite magnetic data 23
  3.2 POGO and Magsat 26
  3.3 Ørsted 26
  3.4 CHAMP 28
  3.5 SAC-C 28
  3.6 Swarm 29
  3.7 Error sources in satellite data 30
  3.8 Summary 30

4 Spherical harmonic representations of the magnetic field 31
  4.1 Expectation, variance and covariance 31
  4.2 Estimation of Gauss coefficients 33
  4.3 The CM series 36
  4.4 The OSVM series 38
  4.5 The MF series 39
  4.6 Summary 41
CONTENTS

II Uncertainty estimates of field models – Variogram analysis of
field model residuals 43

5 Correlated errors, covariances and variograms 45
  5.1 Uncertainty estimates of field models 45
  5.2 The variogram 49
    5.2.1 Stationarity and detrending 51
    5.2.2 Experimental and theoretical variograms 52
    5.2.3 Nugget effect 53
    5.2.4 Finding the theoretical variogram 54
  5.3 Spatiotemporal covariance 54
  5.4 Summary 56

6 Covariance of field model residuals 57
  6.1 Residuals 57
    6.1.1 Removing trends 61
  6.2 Covariance of the trend 63
  6.3 Covariance of the corrected residuals 64
    6.3.1 Temporal variation 66
    6.3.2 Spatial variation 68
    6.3.3 The spatiotemporal covariance function 70
  6.4 The total along-track covariance 71
    6.4.1 The data covariance matrices 72
  6.5 Discussion 78
  6.6 Conclusions 81

III Application of field models – Magnetic crustal thickness and
geothermal heat flux 83

7 Magnetic crustal thickness 85
  7.1 The observed induced crustal field 86
  7.2 Equivalent source magnetic dipole method 89
    7.2.1 The magnetic field from a dipole 89
    7.2.2 Crustal magnetization and dipole moments 94
    7.2.3 Directions of the dipoles 95
    7.2.4 Modelling the induced crustal field 97
    7.2.5 Locations of observations and crustal dipoles 98
  7.3 Iterative forward modelling 99
    7.3.1 Annihilators 103
  7.4 Results for the magnetic crustal thickness 104
    7.4.1 Greenland 104
    7.4.2 Antarctica 105
    7.4.3 Error sources and uncertainty estimate 109
  7.5 Conclusions 110
# Contents

8 Geothermal heat flux 111
   8.1 Heat flux model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 111
   8.2 Heat flux underneath large ice sheets . . . . . . . . . . . . . . . . . . . 113
      8.2.1 Antarctica . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 114
      8.2.2 Greenland . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116
      8.2.3 Error sources and uncertainty estimate . . . . . . . . . . . . . . . 118
   8.3 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 120

9 Summary and conclusions 121

A Publications connected to this work 125

B Danish summary: Dansk resumé af afhandlingen 129

Bibliography 133
Acknowledgments

First and foremost I wish to thank my advisors Nils Olsen from the Danish National Space Center and Klaus Mosegaard from the Niels Bohr Institute for their patience and support and for all they have taught me not only about physics, but also about how the scientific community works.

This project was carried out at the Niels Bohr Institute, University of Copenhagen, in close cooperation with the Danish National Space Center. The project was funded by the Danish Natural Science Research Council through the ESA følgeforskning programme. During the project I spend three months at the Geodynamics Branch at Goddard Space Flight Center, NASA, working closely with Michael E. Purucker, to whom I am thankful of the fruitful cooperation and for the help and advice he provided particularly in the early stages of the project. I would also like to thank Herbert Frey, head of the Geodynamics Branch at GSFC, for making it possible for me to visit Goddard, and for making me feel very welcome. At Goddard I had the pleasure of meeting a large group of very nice and entertaining people, especially over lunch. In particular I met Terence Sabaka, whose friendship, advice and support has been of immense value to me since. I specifically wish to thank him for reviewing parts of this thesis.

I am also grateful to Tina Christensen, who also reviewed parts of this thesis and who temporarily took over some of my responsibilities as chairperson of Women in Physics in Denmark (KIF) when I was most busy writing the thesis. Another person that I wish to thank for taking over some of my responsibilities in KIF and in NorWiP is Anja C. Andersen, who I really admire and whose support and friendship means a lot to me.

There are many people at the Niels Bohr Institute it has been a pleasure to be around. Naming them all would be to extensive, but there is particularly one I wish to mention and that is Jens Martin Knudsen, who was truly a great person and who has been a major inspiration to me ever since I started studying physics. I am grateful that I have had the fortune of getting to known Jens Martin - he is greatly missed.

Finally I wish to thank my friends and my family for their support, and last but not least thanks to Kristoffer for being there and for always having energy to doing that little extra effort to help people around him - a quality I really admire.

Cathrine Fox Maule
Copenhagen, October 2005.
Chapter 1

Motivation and outline

The recent satellite missions dedicated to measuring the Earth’s magnetic field have increased the amount of magnetic field data considerably. With the new high quality data, the understanding of the Earth’s magnetic field is improving rapidly and new and better descriptions of the magnetic field by magnetic field models are continuously being developed. The resolution of magnetic field models of the crustal field is now approaching a level, where it becomes possible to use them for geologic and geophysical interpretation. However, an important aspect of using a field model as data for further interpretation is the need to have an uncertainty estimate of the field model, such that the uncertainty on quantities derived from the field model may be determined. Unfortunately uncertainty estimates of field models are not readily available, and so far there is no standard procedure to make reliable uncertainty estimates of field models in an easy and efficient way.

Two challenges have been pursued in the PhD work presented in this thesis. One is to use a field model to derive maps of the geothermal heat flux underneath the large ice sheets in Antarctica and Greenland. The other, which was brought about by the wish to determine the uncertainty of the derived heat flux, is to estimate the data covariance of field model residuals using variograms; with the data covariance matrix an uncertainty estimate of the field model can be made.

The geothermal heat flux can vary greatly over short distances e.g. due to changes in geology. This is also the case underneath the large ice sheets in Greenland and Antarctica, which is of interest because the geothermal heat flux influences the ice sheet dynamics. However, little is known about how the heat flux varies underneath the ice. Direct measurements are scarce because they are difficult and expensive to obtain, thus an indirect approach of determining the heat flux underneath the ice is an attractive alternative.

We have investigated whether magnetic field models, based on satellite magnetic data, can be used to infer the geothermal heat flux underneath the large ice sheets. To use a field model as data instead of the raw satellite data of the magnetic field is highly practical; the field measured by the satellites originates from several different sources: the core, the crust, the ionosphere and the magnetosphere, and it is only the field from one of the sources, the crust, that contains information about the heat flux. By deriving a field model from the raw data, the field from the various sources can be separated, and in particular the part of the field that contains information about the heat flux can be isolated.
Magnetic data can be used to infer the geothermal heat flux because the magnetic properties of crustal rocks are temperature dependent. At temperatures below their Curie temperature crustal rocks may sustain remanent or induced magnetism, but when heated above their Curie temperature the crustal rocks become essentially non-magnetic. From magnetic data it is possible to estimate the depth of this magnetic boundary, and thus the depth of the Curie isotherm. Aeromagnetic data have been widely used to determine the depth to the Curie isotherm, and it has in many of these studies been noted that the obtained depths to the Curie isotherm correlate with heat flow measurements. Several studies based on satellite magnetic data have also observed correlation between crustal magnetization models based on satellite magnetic data, the depth to the Curie isotherm and the surface heat flow. However, these earlier studies were based on satellite data of a much lower quality, than the data currently being obtained. Our study differs from the previous ones in that we use high quality satellite data from the current missions Ørsted, CHAMP and SAC-C, we estimate the depth to the Curie isotherm in the polar regions, and we use this estimate to infer the geothermal heat flux underneath the ice.

When using a field model for geophysical interpretation it is important to have an estimate of the uncertainty of the field model, such that the uncertainty on the quantity derived from the field model also can be estimated. However, valid uncertainty estimates of field models are in general not made, because there so far is no apparent way to obtain such estimates. A way to make a valid uncertainty estimate of a field model is to include the data covariance matrix in the field modelling procedure, but the data covariance matrix is in general not known. It is therefore common in field modelling to estimate the field model parameters without using a realistic data covariance matrix and the obtained field model parameters are generally accepted as being valid in spite of this simplification. The standard field modelling procedure does provide a method to determine the variances of the field model parameters, but this method does most likely not produce reasonable results unless a realistic data covariance matrix is used. To achieve this, the data covariance matrix can be estimated posterior to the field modelling by analysing the field model residuals. The residuals, which are that part of the observations that the field model in question does not explain, consist of instrument/measurement errors and more importantly of contributions from fields that are not modelled by the field model. Particularly the unmodelled field gives rise to correlation of the residuals, which means that non-zero off-diagonal elements are present in the data covariance matrix; it is these off-diagonal elements that usually are not included in the field modelling procedure, but which may significantly affect the estimate of the variances of the field model parameters. Thus, by analysing the residuals the data covariance matrix can be estimated in order to achieve reliable estimates of the variances of the field model parameters such that the overall level of uncertainty of the field model can be determined.

Error correlation of field model residuals and uncertainty estimates of field models have been the topics of many studies, which all concluded that correlation of field model residuals occurs and must be accounted for when making uncertainty estimates of field models. However, none has presented an efficient and reliable way to estimate the uncertainty of field models. We use a different method to estimate the data covariance matrix of a field model compared to what previously has been used. Variogram analysis of the field model residuals of the Ørsted(06s/05) field model is
used to estimate the along-track covariance function of the residuals, from which the
data covariance matrix of the Ørsted(06s/05) field model is found.

Outline of the thesis

This thesis falls in three parts.

In Part I a general introduction of topics, which are of relevance to the research conducted in this PhD project, namely the Earth’s magnetic field, magnetic satellite observations, and magnetic field representations, is given. This part is mainly aimed at the reader who has a background in geophysics or physics, but who is not working within geomagnetism. In chapter 2 a brief introduction to the Earth’s magnetic field and its sources is given, and spherical harmonic expansions of the magnetic field and field models are introduced. In chapter 3 a few past and all of the present satellite missions dedicated to measuring the magnetic field are briefly presented. In chapter 4 it is briefly outlined how magnetic field models are derived, and three series of field models are presented. With the information provided in these three introductory chapters, the reader should have the necessary background to fully comprehend the scientific work presented in the following chapters.

Part II is focussed on uncertainty estimates of field models, and presents the study of correlation of field model residuals. In chapter 5 a presentation of past studies of uncertainty estimates of field models and error correlation of data used to derive field models is given, and the variogram, which we have used to estimate the covariance of field model residuals, is introduced. In chapter 6 the actual analysis of field model residuals using variograms is presented.

Part III is focussed on the application of field models for geophysical interpretation, and the study concerning the estimation of the geothermal heat flux underneath the large ice caps in Greenland and Antarctica from a magnetic field model is presented. In chapter 7 the method we have used to estimate the thickness of the magnetic crust in Greenland and Antarctica from a magnetic field model is explained in detail, and the achieved results are presented. In chapter 8 the thermal model used to infer the geothermal heat flux from the obtained magnetic crustal thickness is presented and the results for the geothermal heat flux are discussed. With part II and part III the scientific results that comprise this PhD project have been presented.

In chapter 9 an overall summary and the conclusions are given. Appendix A contains a list of the publications that are a direct outcome of this PhD work. In appendix B a Danish summary of the thesis is given.

Units and Notation

SI units are used throughout the thesis. Regarding notation, overbars are used to denote vectors, while the norm/length of a vector is written with regular types, i.e.

\[ m = |\vec{m}|, \tag{1.1} \]

hat denotes unit vectors i.e.

\[ \hat{r} = \frac{\vec{r}}{r}, \tag{1.2} \]

and boldface capital letters are used to denote matrices.
Part I

Earth’s magnetic field, satellite observations and field representations
Chapter 2

Introduction to Earth’s magnetic field

Although invisible to the human eye, the Earth’s magnetic field affects our lives in many ways. The magnetic field shields us from cosmic radiation, it gives rise to the breath-taking natural phenomena of aurora, and it has been used for navigation through centuries. The compass can be traced more than 1000 years back (for a short review of the history of geomagnetism see e.g. Stern [2002]). The need for reliable navigation increased as trading over large distances became more and more common as civilization evolved. Thus there was a strong incentive to monitor and study the magnetic field, and systematic observations began centuries ago.

In this chapter a short introduction to some basic features of the Earth’s magnetic field is given. The most important sources of the field are introduced, and important mathematical tools are presented. More extensive treatments of these topics can be found in e.g. Campbell [1997] or Langel and Hinze [1998].

2.1 Earth’s magnetic environment

The idea that the Earth behaves like a great magnet was first proposed by Gilbert in 1600 in his *De Magnete* [Gilbert, 1958], and Gilbert was very close to the truth. To first approximation the Earth’s magnetic field is that of a dipole at the Earth’s centre, with the dipole axis tilted 11° with respect to the rotational axis. The magnetic south pole of the Earth’s field points towards the geographic north pole and vice versa. The magnetic field is predominantly generated in the liquid outer core, where convection of highly conducting material drives a self-sustaining dynamo. At the Earth’s surface, the field strength is about 60,000 nT at the poles and 30,000 nT at the equator. Figure 2.1 shows the total intensity and the radial component of the magnetic field at the surface.

The centre of the dipole is displaced a few hundred kilometres from the Earth’s centre, which gives rise to the South-Atlantic anomaly, also seen in figure 2.1. In this area the field is weaker and the protection against cosmic radiation lower. This can be directly observed by the failure of satellites flying over the South-Atlantic anomaly. Due to the safety of airline staff, aircrafts are routed around the anomaly.

The Earth and its nearest environment are completely embedded in the Earth’s
CHAPTER 2. INTRODUCTION TO EARTH’S MAGNETIC FIELD

Figure 2.1: The top panel shows the total intensity of the magnetic field (IGRF 2000, [Olsen et al., 2000b]) at the surface of the Earth. The South-Atlantic anomaly is clearly seen in South America. The lower panel shows the radial component of the field. The field is predominantly directed outwards in the southern hemisphere and inwards in the northern hemisphere. The white line, where the radial component is zero, is the dip equator.
2.1. EARTH’S MAGNETIC ENVIRONMENT

magnetic field. However, the dominating magnetic field in the solar system is the Sun’s: the solar wind, a plasma of primarily electrons and protons continuously emitted by the Sun, carries a frozen-in field known as the interplanetary magnetic field with it from the Sun towards the planets. The Earth’s magnetic field is an obstacle to the solar wind, which consequently flows around it. The (magnetic) boundary between the Earth’s magnetic field and the solar wind is called the magnetopause. Inside the magnetopause is the magnetosphere, which is dominated by the Earth’s magnetic field. Outside the magnetosphere the solar wind dominates the magnetic environment.

The interaction of the solar wind and the Earth’s magnetic field is highly dynamic and gives rise to many phenomena including the aurora. At distances less than a few Earth radii, the Earth’s field is close to dipolar, but beyond this the distortion by the solar wind is evident. Due to the dynamic pressure of the solar wind, the dayside of the magnetosphere is compressed towards the Earth, whereas it is extended hundreds of Earth radii on the nightside. This gives the magnetosphere a teardrop shape as illustrated in figure 2.2. Normally the magnetopause on the dayside is at a distance of about 12 Earth radii, but under severe solar storms, the increased pressure of the solar wind can move this boundary closer than 6–7 Earth radii, that is within the geostationary orbit.

A satellite orbiting the Earth at some hundred kilometres altitude measures a magnetic field comprised of contributions from several sources; figure 2.3 shows the source and data acquisition regions. The dominant contribution, about 94%, of the field observed by satellites is generated in the Earth’s core. A few percent (\sim 3\%) of

\footnote{Figure is from http://www-ssg.sr.unh.edu/tof/Smart/Students/lees/magball.html.}

Figure 2.2: Illustration of how the solar wind deforms the Earth’s magnetic field (not to scale)
CHAPTER 2. INTRODUCTION TO EARTH’S MAGNETIC FIELD

Figure 2.3: Sketch of the source regions of the magnetic field, and the regions where magnetic data are obtained (not to scale). Most of Earth’s magnetic field is generated in the outer core. A minor contribution comes from the crust, where both induced and remanent magnetism occurs. Another minor contribution comes from ionospheric and magnetospheric currents. The two current systems are coupled through field aligned currents. The currents generate magnetic fields observable both at ground-based observatories and by satellites. The acquisition of magnetic data for field models occurs primarily at two levels; at the surface by observatories, ships and aeroplanes, and at satellite altitudes of about 400–800 km. Figure provided by Nils Olsen.
the field originates from the crust. The core and crustal field together comprise the internal sources of the magnetic field, i.e. magnetic sources internal to the Earth. A few percent (≈ 3%) of the field detected by satellites comes from sources external to the Earth, primarily from currents flowing in the ionosphere and magnetosphere. The ionospheric and magnetospheric currents induce currents in the mantle providing yet another minor contribution to the field. In general the terms internal and external sources are referring to internal or external to the Earth. The satellites measuring the magnetic field are typically at altitudes of 400–800 km, which is above the primary ionospheric currents. Thus these external sources are internal to the satellite orbits.

The magnetic field varies on an enormous range of times-scales. The currents in the ionosphere and magnetosphere vary on several time-scales from seconds over minutes and hours to several years. The core field varies on time-scales from about half a year to several millions of years, and the crustal field changes on geologic time-scales.

The slow drift of the core field that occurs on a time-scale of a few years is known as the secular variation. Both the magnitude and the direction of the core field slowly changes; at the Earth’s surface the field change is typically 10–100 nT per year. The secular variation is not constant, and recently estimates of the secular acceleration are also made.

On longer time-scales the core field is known to change polarity from time to time. These reversals constitute the cornerstone of paleomagnetism. Reversals have occurred throughout the history of the magnetic field. They occur irregularly, without any predominant frequency. The time between reversals is typically hundred thousands of years; the longest period without reversals, the Cretaceous Quiet Interval, lasted more than 35 million years [Cande and Kent, 1995]. Paleomagnetism was crucial to the recognition and verification of seafloor spreading and plate tectonics.

The variation of the external fields occurs on much shorter time-scales, from years down to seconds. The external fields are influenced by the solar activity level and thus show a variation with the 11-year solar cycle. When the solar activity is low, the external currents give rise to a smoothly varying pattern, which is similar from day to day although with gradual seasonal changes. Days, where the field variation is dominated by quiet-level daily variations, are known as geomagnetic quiet days.

When the solar activity increases, the solar wind becomes denser and faster, and the constituent particles more energetic. This compresses and distorts the magnetosphere and large amounts of particles are injected into it. Many currents increase in strength and new short-lived currents can arise, resulting in disturbances from the quiet time conditions, known as active or disturbed times; very disturbed times are labelled geomagnetic storms. The number and intensity of geomagnetic storms vary with the 11-year solar cycle.

In general, the magnetic field is much more variable and disturbed in the polar regions than in the equatorial regions. Often there is activity in the polar regions even during periods of quiet times at most other latitudes. In the polar regions the field is close to vertical allowing charged particles flowing along field lines to enter the upper atmosphere. This particle precipitation occurs both day and night and at both quiet and disturbed times. The collision of the precipitating particles with the neutral atoms and molecules of the atmosphere has several effects. The collisions can ionize molecules thereby causing increased ionization in the polar regions,
which carries the many strong currents here. Another effect is that the collisions between the precipitating particles and the atmospheric constituents can excite the atmospheric molecules. When the excited molecules return to their ground state they emit photons, which constitute the aurora. The auroral zones are at about 64° to 70° geomagnetic latitude, but expand and shrink depending on the level of solar activity.

**Activity indices**

There exist a number of activity indices that describe the level of disturbance of the magnetic field. The various indices are based on different measurements and related to different ionospheric and magnetospheric processes. Here I will only introduce a couple of indices, which will be referred to later, namely the $D_{st}$ index and the $K_p$ index. An overview of the most common indices used in geomagnetism can be found in e.g. Campbell [1997] or Langel and Hinze [1998].

The $K_p$ (planetarische Kennziffer) index is based on worldwide data from 13 observatories. It gives a qualitative measure of the level of global disturbance. It is based on the amplitude range of the most disturbed horizontal component of the field after it has been corrected for quiet level daily variations. The largest amplitude range is determined at each observatory for every three-hour interval in Universal Time (UT). The scale of $K_p$ is quasi-logarithmic. The index is determined by converting the amplitude range at each station into a number on a scale from 0 to 9, with 0 being the most quiet and 9 the most disturbed time. The scale is subdivided into thirds giving $K_p$ values of 0°, 0+, 1−, 10, 1+ etc..

The $D_{st}$ (storm-time disturbance) index is based on magnetic field measurements made at four low-latitude observatories, evenly distributed in longitude. Quiet level daily variations are subtracted from the measurements, and the residual fields from all the stations are averaged to give the $D_{st}$ index on an hourly basis. Being at low latitudes the stations are particularly sensitive to the magnetospheric ring current (see below), and $D_{st}$ is widely used to model the temporal variation of the ring current strength.

**2.2 Sources of the field**

The sources of the magnetic field present at the surface of the Earth are divided into internal and external sources as mentioned earlier. The two source regions of the internal field are the core and the crust. Convective motions of hot iron alloy in the liquid outer core generate the strong core field. The much weaker crustal field is due to remanent and induced magnetism in the crustal rocks. Thus the core and crustal field are generated by fundamentally different processes.

The external field is generated from a number of dynamic currents in the ionosphere and magnetosphere including the equatorial electrojet, the polar electrojet, the magnetopause current, the magnetospheric ring current and the tail current. The currents are very variable and complex in many ways; some currents run regularly, some highly irregularly at different altitudes, latitudes and at different times. The ionospheric and magnetospheric currents are coupled by field aligned currents adding to the complexity. A description of the myriad of currents in the ionosphere and magnetosphere can be found in e.g. Kivelson and Russell [1995] or Christensen [2002],
below I will only describe the magnetospheric ring current, as this will be treated later (in chapter 5).

2.2. Core field

The core field is generated by currents in the outer core; since the temperature in the core is much higher than the Curie temperature of the core material remanent and induced magnetism is prevented here. The Curie temperature, which is very different from material to material, is the temperature above which a given material loses the ability to sustain a (remanent or induced) magnetic field. The outer core consists of a liquid iron alloy containing some nickel (∼8%) and about 6–10% lighter elements possibly oxygen or sulphur. The hot alloy is a very good conductor with respect to both electricity and heat. Motions of the conducting alloy within the Earth’s magnetic field (re)generate and reinforce the magnetic field; it is a self-sustaining dynamo. The dynamo is driven by convection in the core, but that there is convection in the outer core is no obvious matter. The thermal conductivity of the alloy is so high that thermal convection is not likely to occur in spite of the low viscosity and the present temperature gradient. However, due to the secular cooling of the Earth, the temperature at the inner-outer core boundary continuously falls below the melting point of iron at the given pressure. This causes an iron-enriched phase of the alloy to condense at the bottom of the outer core leaving behind a lighter residual (for a review of the inner core solidification process see Bergman [2003]). Being positively buoyant the residual rise towards the core-mantle boundary causing the (compositionally driven) convection of the outer core.

In situ measurements of the core is beyond technological reach, and so all our knowledge about the state of the core is based on measurements at or above the Earth’s surface, theoretical studies and small scale simulation experiments. Measurements of the field cannot show the fine details of the core flow, partly because short-wavelength features of the field are obscured by the crustal field, and partly because particularly the short-scale (spatial and temporal) features are attenuated between the core and the surface. Although the electrical conductivity of the mantle is low compared to the core, it is nevertheless high enough to attenuate magnetic fluctuations with periods shorter than about a year. The difficulties of obtaining good data from the core mean that there are still many unanswered questions about how the dynamo works and why the field behaves as it does.

2.2.2 Crustal field

In the outermost tens of kilometres of the Earth, the temperature is below the Curie temperature of the most common magnetic minerals. This allows for remanent and induced magnetism of the crustal rocks; a review of rock magnetism can be found in e.g. Dunlop [1995]. The magnetic minerals of importance in the crust are various iron oxides, of which magnetite is believed to be the primary cause of the crustal field [e.g. Schlinger, 1985, Frost and Shive, 1986, Clark and Emerson, 1991, Langel and Hinze, 1998]. Magnetite is fairly common, it is strongly magnetized and has a relatively high Curie temperature of about 580°C. Although the crustal field is much weaker than the core field, it is nevertheless strong enough to be observed by low
Remanent magnetism in the crust is created when a magma containing magnetic minerals cool below its Curie temperature in a magnetic environment. Large-scale remanent magnetization occurs along the oceanic spreading ridges. The remanent magnetization of the ocean floor contains a recording of the past magnetic field and how the field has changed.

Remanent magnetism also occurs in the continental crust, but to a lesser degree and with a much less systematic pattern. In the continental crust remagnetism induced by the core field is believed to dominate the remanent magnetism. The induced field depends on the strength of the core field, the magnetic susceptibility of the crustal rocks and thickness of the magnetic crust. The crustal field varies much less systematically in the continental crust compared to the oceanic. This is partly due to the much higher petrologic variability of the continental crust. Different types of rocks have different contents of magnetic minerals. Sedimentary rocks generally have a very low (negligible) magnetization, whereas igneous and metamorphic rocks can have a high magnetization. Thus large variations in the crustal field can occur over distances less than 1 km [Langel and Hinze, 1998], as the geologic properties of the crust vary on this scale.

The crustal rocks are chemically different from the mantle rocks, and it is the chemical boundary that defines the crust-mantle boundary. Normally the crust-mantle boundary is considered to coincide with the seismologically defined crust-mantle boundary, Moho, which can be determined relatively easy from seismic measurements. To what extent the source region of the crustal field extends beyond Moho has been somewhat debated, because the magnetic properties of mantle rocks are not so well known. However, based on a study of about 400 xenolith samples (mantle rock samples), Wasilewski and Mayhew [1992] concluded that the magnetization of mantle rocks in general is very low, and that the magnetic crust does not extend beyond Moho, if Moho in a region lies shallower than the Curie isotherm. Nevertheless the term lithospheric field is sometimes used for the magnetic field from the crust to signal the possible inclusion of mantle rocks. Here we use the term crustal field although we do not exclude the possibility that mantle rocks may contribute to the crustal field. In any case, the magnetization of the crust does not extend beyond the depth of the Curie isotherm. A central concept in this thesis is the magnetic crust; its upper boundary is the bedrock surface, its lower boundary is either the Curie isotherm or Moho, depending on which of the two that is shallower. The magnetic crust does not include sediments or ice sheets, but only crystalline rocks as illustrated in figure 2.4. In oceanic regions where the crust is relatively thin, about 7 km [White et al., 1992], the Curie isotherm is likely to be in the upper mantle [Langel and Hinze, 1998, Tanaka et al., 1999]. In the continental crust, which is tens of kilometres thick [Christensen and Mooney, 1995], the Curie isotherm is easily reached within the crust [Langel and Hinze, 1998, Tanaka et al., 1999]. A list of results obtained for the depth of the Curie isotherm for many different areas from several different studies by various authors is given in Tanaka et al. [1999].
2.3. Mathematical representation of the magnetic field

The magnetic field vector, \( \vec{B} \) (the magnetic flux density), can be expressed in several coordinate systems. The most commonly used coordinates when dealing with satellite measurements are the radial, colatitudinal and longitudinal components,

\[
\vec{B} = (B_r, B_\theta, B_\phi) .
\]  

(2.1)

These coordinates will be used throughout this thesis. Field lines, which at any point are tangent to the local magnetic field vector, are often used to illustrate the shape of the field.
CHAPTER 2. INTRODUCTION TO EARTH’S MAGNETIC FIELD

Before the satellite era, when magnetic measurements consisted of surface data from observatories, ships and aeroplanes, a topocentric system was used. In this, the field is resolved into a horizontal component, \( H \), and a vertical component, \( Z \), with \( Z \) being positive downwards. The horizontal component is resolved in two directions; the direction of the local meridian, \( X \), which is positive towards north, and the direction perpendicular to the meridian, \( Y \), which is positive towards east. These directions are used to define the inclination, \( I \), which is the angle between the total field and the horizontal plane

\[
I = \arctan \left( \frac{Z}{H} \right) = \arctan \left( \frac{Z}{\sqrt{X^2 + Y^2}} \right),
\]

and the declination, \( D \), which is the angle between magnetic and geographic north

\[
D = \arctan \left( \frac{Y}{X} \right).
\]

The declination is of great importance for navigational purposes.

If the Earth is approximated by a sphere, the relations between the spherical components and the topocentric system are

\[
\begin{align*}
X &= -B_\theta \\
Y &= B_\phi \\
Z &= -B_r.
\end{align*}
\]

2.3.1 Spherical harmonic expansions

All magnetic fields, including the Earth’s, obey some very fundamental laws of physics described by Maxwell’s equations, from which many interesting properties of the field can be deduced. A fundamental property of magnetic fields is that they are divergence-free, that is (Maxwell’s 2nd equation)

\[
\nabla \cdot \vec{B} = 0,
\]

where \( \vec{B} \) is the magnetic induction or flux density, measured in Tesla (T). That the fields are divergence-free is equivalent to saying that magnetic monopoles do not exist. Maxwell’s 4th equation state that in regions outside magnetic materials, where displacement currents can be neglected,

\[
\nabla \times \vec{B} = \mu_0 \vec{J},
\]

where \( \mu_0 \) is the vacuum permeability, which is \( 4\pi \cdot 10^{-7} \text{ N/A}^2 \), and \( \vec{J} \) is the current density. If the field is measured in a current-free region, \( \vec{J} = 0 \), \( \vec{B} \) is curl-free and can be described as the (negative) gradient of a scalar potential, \( V \),

\[
\vec{B} = -\nabla V,
\]

which due to equation 2.5 then satisfies the Laplace equation

\[
\nabla \cdot \nabla V = \nabla^2 V = 0.
\]
2.3. MATHEMATICAL REPRESENTATION OF THE MAGNETIC FIELD

Thus the Earth’s magnetic field can in current-free regions outside the Earth be described as the negative gradient of a scalar potential that satisfies the Laplace equation.

The solution for the magnetic potential in spherical coordinates is a spherical harmonic expansion

\[
V = a \sum_{n=1}^{\infty} \sum_{m=0}^{n} \left( \frac{a}{r} \right)^{n+1} \left[ g_n^m \cos (m\phi) + h_n^m \sin (m\phi) \right] P_n^m (\cos \theta) \\
+ a \sum_{n=1}^{\infty} \sum_{m=0}^{n} \left( \frac{r}{a} \right)^n \left[ q_n^m \cos (m\phi) + s_n^m \sin (m\phi) \right] P_n^m (\cos \theta),
\]

(2.9)

where \((r, \theta, \phi)\) is radius, colatitude, and longitude, respectively, in the conventional geocentric, geographical spherical-polar coordinate system, \(a\) is a reference radius, taken to be the mean radius of the Earth (6371.2 km), \(P_n^m\) are the Schmidt semi-normalized associated Legendre functions, and \(g_n^m, h_n^m, q_n^m\) and \(s_n^m\) are called Gauss coefficients and are usually given in nT, \(n\) is the degree and \(m\) the order of the individual terms. In the Schmidt normalization of the Legendre functions, which is the convention in geomagnetism, the tesseral and sectoral harmonics are normalized to have the same r.m.s. values over the spherical surface as the zonal harmonics of the same degree, whereas no normalization is applied to the zonal harmonics [Stacey, 1992].

The relations between the magnetic components and the potential are

\[
B_r = -\frac{\partial V}{\partial r} \\
B_\theta = \frac{-1}{a} \frac{\partial V}{\partial \theta} \\
B_\phi = \frac{-1}{a \sin \theta} \frac{\partial V}{\partial \phi}
\]

(2.10)

and the total (scalar) intensity is given by

\[
B = \sqrt{B_r^2 + B_\theta^2 + B_\phi^2}.
\]

(2.11)

Equation 2.9 may at first sight look complicated, but the spherical harmonic expansion provides a very powerful tool for handling and interpreting various features of the magnetic field. The first sum describes the field from internal sources; its amplitude decreases with increasing distance from the Earth. The second sum represents the field from external sources, its amplitude decreases when moving towards the Earth. Thus, if only the internal sources are considered the potential is simply

\[
V = a \sum_{n=1}^{\infty} \sum_{m=0}^{n} \left( \frac{a}{r} \right)^{n+1} \left[ g_n^m \cos (m\phi) + h_n^m \sin (m\phi) \right] P_n^m (\cos \theta).
\]

(2.12)

Each order of the expansion has a physical interpretation; \(n = 1\) gives the dipole terms of the field, \(n = 2\) the quadrupole terms, \(n = 3\) the octupole and so on. There is no \(n = 0\) term, which would correspond to a magnetic monopole. The Gauss coefficients of the IGRF 2000 magnetic field model [Olsen et al., 2000b] are listed in
<table>
<thead>
<tr>
<th>n</th>
<th>m</th>
<th>( g_n^m ) (nT)</th>
<th>( h_n^m ) (nT)</th>
<th>n</th>
<th>m</th>
<th>( g_n^m ) (nT)</th>
<th>( h_n^m ) (nT)</th>
<th>n</th>
<th>m</th>
<th>( g_n^m ) (nT)</th>
<th>( h_n^m ) (nT)</th>
<th>n</th>
<th>m</th>
<th>( g_n^m ) (nT)</th>
<th>( h_n^m ) (nT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>-29614.72</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>24.92</td>
<td>0</td>
<td>11</td>
<td>5</td>
<td>0.03</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>-1727.74</td>
<td>5186.27</td>
<td>8</td>
<td>1</td>
<td>6.3</td>
<td>11.78</td>
<td>11</td>
<td>6</td>
<td>-0.64</td>
<td>-0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>-2267.11</td>
<td>0</td>
<td>2</td>
<td>-9.06</td>
<td>-21.58</td>
<td>11</td>
<td>7</td>
<td>0.89</td>
<td>-2.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3071.78</td>
<td>-2477.69</td>
<td>8</td>
<td>3</td>
<td>-7.71</td>
<td>8.45</td>
<td>11</td>
<td>8</td>
<td>1.86</td>
<td>-0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1672.18</td>
<td>-457.36</td>
<td>8</td>
<td>4</td>
<td>-16.93</td>
<td>-21.26</td>
<td>11</td>
<td>9</td>
<td>0.05</td>
<td>-1.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1340.85</td>
<td>0</td>
<td>8</td>
<td>5</td>
<td>9.03</td>
<td>15.26</td>
<td>11</td>
<td>10</td>
<td>1.09</td>
<td>-2.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>-2289.53</td>
<td>-226.99</td>
<td>8</td>
<td>6</td>
<td>6.91</td>
<td>8.58</td>
<td>11</td>
<td>11</td>
<td>4.07</td>
<td>-0.83</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1252.52</td>
<td>296.26</td>
<td>8</td>
<td>7</td>
<td>-7.84</td>
<td>-15.5</td>
<td>12</td>
<td>0</td>
<td>-1.85</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>715.1</td>
<td>-492.28</td>
<td>8</td>
<td>8</td>
<td>-6.95</td>
<td>-2.7</td>
<td>12</td>
<td>1</td>
<td>-0.19</td>
<td>-0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>935.35</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>5.26</td>
<td>0</td>
<td>12</td>
<td>2</td>
<td>-0.15</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>786.71</td>
<td>271.92</td>
<td>9</td>
<td>1</td>
<td>9.06</td>
<td>-20</td>
<td>12</td>
<td>3</td>
<td>0.72</td>
<td>2.57</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>250.81</td>
<td>-231.63</td>
<td>9</td>
<td>2</td>
<td>2.61</td>
<td>13.4</td>
<td>12</td>
<td>4</td>
<td>-0.29</td>
<td>-2.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>-404.66</td>
<td>118.54</td>
<td>9</td>
<td>3</td>
<td>-8.49</td>
<td>12.4</td>
<td>12</td>
<td>5</td>
<td>0.91</td>
<td>0.64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>109.91</td>
<td>-303.88</td>
<td>9</td>
<td>4</td>
<td>6.44</td>
<td>-6.26</td>
<td>12</td>
<td>6</td>
<td>-0.51</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>-216.82</td>
<td>0</td>
<td>9</td>
<td>5</td>
<td>-8.77</td>
<td>-8.47</td>
<td>12</td>
<td>7</td>
<td>0.32</td>
<td>-0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>350.96</td>
<td>43.69</td>
<td>9</td>
<td>6</td>
<td>-1.55</td>
<td>8.59</td>
<td>12</td>
<td>8</td>
<td>-0.25</td>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>222.36</td>
<td>171.65</td>
<td>9</td>
<td>7</td>
<td>9.24</td>
<td>3.83</td>
<td>12</td>
<td>9</td>
<td>-0.34</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3</td>
<td>-130.72</td>
<td>-133.88</td>
<td>9</td>
<td>8</td>
<td>-3.98</td>
<td>-8.08</td>
<td>12</td>
<td>10</td>
<td>-0.02</td>
<td>-0.95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>-168.53</td>
<td>-40.26</td>
<td>9</td>
<td>9</td>
<td>-8.24</td>
<td>4.64</td>
<td>12</td>
<td>11</td>
<td>0.11</td>
<td>-0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>-11.83</td>
<td>106.79</td>
<td>10</td>
<td>0</td>
<td>-2.44</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>-0.55</td>
<td>0.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>72.34</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>-5.92</td>
<td>1.23</td>
<td>13</td>
<td>0</td>
<td>-0.04</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>67.76</td>
<td>-17.09</td>
<td>10</td>
<td>2</td>
<td>1.52</td>
<td>0.12</td>
<td>13</td>
<td>1</td>
<td>-1.11</td>
<td>-1.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>73.98</td>
<td>63.84</td>
<td>10</td>
<td>3</td>
<td>-3.15</td>
<td>3.97</td>
<td>13</td>
<td>2</td>
<td>-0.26</td>
<td>0.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>-161.06</td>
<td>65.19</td>
<td>10</td>
<td>4</td>
<td>-0.4</td>
<td>4.95</td>
<td>13</td>
<td>3</td>
<td>0.04</td>
<td>1.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>-5.13</td>
<td>-61.34</td>
<td>10</td>
<td>5</td>
<td>3.81</td>
<td>-5.79</td>
<td>13</td>
<td>4</td>
<td>-0.37</td>
<td>-0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>17.03</td>
<td>1.33</td>
<td>10</td>
<td>6</td>
<td>1.1</td>
<td>-1.25</td>
<td>13</td>
<td>5</td>
<td>1.39</td>
<td>-1.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>-90.53</td>
<td>44.11</td>
<td>10</td>
<td>7</td>
<td>2.03</td>
<td>-2.81</td>
<td>13</td>
<td>6</td>
<td>-0.39</td>
<td>-0.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>79.01</td>
<td>0</td>
<td>10</td>
<td>8</td>
<td>4.38</td>
<td>0.18</td>
<td>13</td>
<td>7</td>
<td>0.6</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>-74.49</td>
<td>-64.71</td>
<td>10</td>
<td>9</td>
<td>0.47</td>
<td>-2.02</td>
<td>13</td>
<td>8</td>
<td>-0.31</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>-0.24</td>
<td>-24.33</td>
<td>10</td>
<td>10</td>
<td>-0.77</td>
<td>-7.66</td>
<td>13</td>
<td>9</td>
<td>0.2</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>33.17</td>
<td>6.35</td>
<td>11</td>
<td>0</td>
<td>2.96</td>
<td>0</td>
<td>13</td>
<td>10</td>
<td>-0.16</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>9.38</td>
<td>23.59</td>
<td>11</td>
<td>1</td>
<td>-1.86</td>
<td>-0.45</td>
<td>13</td>
<td>11</td>
<td>0.5</td>
<td>-0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>6.72</td>
<td>15.07</td>
<td>11</td>
<td>2</td>
<td>-2.22</td>
<td>1.27</td>
<td>13</td>
<td>12</td>
<td>0.11</td>
<td>-0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>7.6</td>
<td>-25.47</td>
<td>11</td>
<td>3</td>
<td>1.54</td>
<td>-0.82</td>
<td>13</td>
<td>13</td>
<td>0.32</td>
<td>-0.73</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>-1.55</td>
<td>-5.55</td>
<td>11</td>
<td>4</td>
<td>-0.19</td>
<td>-2.61</td>
<td>13</td>
<td>14</td>
<td>0.16</td>
<td>-0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: The Gauss coefficients of the IGRF 2000 field model [Olsen et al., 2000b].
2.4. GEOMAGNETIC COORDINATES

Figure 2.5: The power spectrum of the magnetic field (CM4, [Sabaka et al., 2004]) has a clear bend around degree 15. The bend is the distinction between where the core field dominates \( (n < 14) \) and where the crustal field dominates \( (n > 16) \).

table 2.1; it is clearly seen that the dipole terms \( (g_1^0, g_1^1, \text{ and } h_1^1) \) have the largest values. Notice that \( h_0^0 \) always is zero; this occurs because \( \sin(0) \) is zero.

The Gauss coefficients of equation 2.12 represent the field from the internal sources, but it is not possible from the coefficients to separate the contributions from the crust and the core. For many practical purposes a crude separation is done by truncating the spherical harmonic expansion. This separation technique is inspired by the appearance of the fields power spectrum. The power in the \( n \)’th degree at the surface of the Earth is

\[
R_n = (n + 1) \sum_{m=1}^{n} \left[ (g_n^m)^2 + (h_n^m)^2 \right].
\]

(2.13)

The power spectrum of a spherical harmonic expansion of the field at the Earth’s surface has a distinct bend around degree 14–16, as shown in figure 2.5. The spectrum, when plotted as \( \log R_n \) vs. \( n \), has two distinct regimes. For degrees up to about 13, the power clearly decreases with increasing degree. For degrees higher than about 16 the power increases with increasing degree. The two regimes are interpreted as the field being dominated by different sources; for the low degrees the core field dominates the spectrum, for the high degrees the crustal field dominates. At the intermediate degrees, \( n = 13–15 \), both fields are significant and of roughly the same magnitude.

2.4 Geomagnetic coordinates

There exists a few different coordinate systems convenient to work with, when dealing with the Earth’s magnetic field. The normal coordinate system to work with in most geosciences is the geographic coordinate system based on the rotational axis. However, when working with the dipole field, which is tilted about 11° with respect to the
rotational axis, it is sometimes convenient to work with the geomagnetic coordinate
system.

The geomagnetic or dipole coordinate system is based on the dipole axis. The
ggeomagnetic, or dipole, poles are where the axis of the dipole cuts the Earth’s sur-
fäche. The line, where the vertical component ($B_r$) of the dipole is zero, is called the
geomagnetic equator.

The potential and the Gauss coefficients of the spherical harmonic expansion in
equation 2.12 are usually expressed in geographical coordinates. The geographic
colatitude and longitude of the dipole pole, $(\theta_0, \phi_0)$, are thus given by the Gauss
coefficients

$$\cos \theta_0 = \frac{g_0^0}{\sqrt{(g_1^0)^2 + (g_1^1)^2 + (h_1^1)^2}}, \quad \tan \phi_0 = \frac{h_1^1}{g_1^1}. \quad (2.14)$$

From the dipole poles, the geomagnetic or dipole colatitude, $\theta_d$, is defined as

$$\cos \theta_d = \cos \theta_0 \cos \theta + \sin \theta_0 \sin \theta \cos (\phi - \phi_0), \quad (2.15)$$

where $\theta$ and $\phi$ are geographic colatitude and longitude, respectively. Geomagnetic
longitude, $\phi_d$, is defined as

$$\tan \phi_d = \frac{\sin \theta \sin (\phi - \phi_0)}{\cos \theta_0 \sin \theta \cos (\phi - \phi_0) - \sin \theta_0 \cos \theta}. \quad (2.16)$$

A different coordinate system often encountered when working with the magnetic
field is the dip pole and the dip equator. The dip dipoles are defined as where the
(total) field, containing higher order degrees than just the dipolar is vertical, and the
dip equator where it is horizontal, i.e. the inclination is zero (see figure 2.1). Several
dip poles can occur, as there can be local maxima, minima or saddle points of the
potential. The poles, where the potential assumes the highest and the lowest values,
are called the principal poles. Note that in general the dipole poles and the dip poles
do not coincide, nor does the dipole and dip equator, although they may intersect.

In this work, the geographic and the geomagnetic coordinate systems will be used.

2.5 Summary

The Earth’s magnetic field is very dynamic, varying on a huge range of time- and
length-scales. The field experienced by a satellite orbiting the Earth, or measured by
observatories, ships and aeroplanes originates from several sources. The primary fields
are the core ($\sim 94\%$) and crustal fields ($\sim 3\%$). The crustal field is caused by both
remanent magnetism and magnetism induced by the core field in the magnetic crust.
The magnetic crust is the part of the crust able to sustain a magnetic field. Currents
in the ionosphere and magnetosphere give rise to external fields, one of the important
currents being the magnetospheric ring current. The currents vary regularly with
diurnal and seasonal variations during quiet times, and more irregularly and intensely
during disturbed and storm times. In general the field is more disturbed in the polar
regions than elsewhere.
When dealing with the magnetic field, a convenient mathematical framework is potential theory and spherical harmonic expansions. Besides providing a mathematical tool for handling the magnetic field(s), it also has many intuitively pleasant physical interpretations. With spherical harmonic expansions it is possible to separate the fields from internal and external sources. Although the fields from the core and the crust cannot be strictly separated, the dominating parts of each field can be extracted through truncation of the spherical harmonic expansion, with the limits of the truncation inspired by the power spectrum of the field. The core field dominates the low degrees ($n \leq 13$), whereas the crustal field dominates the higher degrees ($n > 16$) of the spherical harmonic expansion; for the intermediate degrees both sources contribute significantly.
Chapter 3

Measuring the magnetic field from space

Currently three satellites dedicated to measuring the magnetic field are gathering magnetic data in low Earth orbit: Ørsted, CHAMP, and SAC-C. These three satellites were all launched around the turn of the millennium and are still operating. A short presentation of each satellite is given in this chapter. Three other satellite missions are also mentioned, namely POGO, Magsat, and Swarm. POGO and Magsat, which flew some 40 and 25 years ago respectively, were the first satellites that measured the magnetic field. POGO only made scalar measurements of the field, whereas the short Magsat mission was the first satellite mission to make vector measurements of the magnetic field. These early magnetic satellite missions gave a glimpse of the vast scientific possibilities with satellite measurements of the magnetic field, which raised the interest for more extensive studies. In many ways POGO and Magsat were the inspiration for the current missions and their expected scientific outcome. To give a perspective on the future within satellite magnetics, a brief introduction to the Swarm mission is included.

As the focus of this thesis is on the use and interpretation of satellite magnetic data, I will only give basic information about the satellites, which are of the relevance to this work. Detailed information about the missions can be found in e.g. Olsen [2005] (Ørsted) and on the respective missions’ web sites.

3.1 Satellite magnetic data

Systematic observations of the magnetic field have been done at ground-based observatories for centuries providing long term data. However, as observatories require operating personnel and location on solid ground, the distribution of observatories is very uneven as seen in figure 3.1. Vast oceanic, polar, and other sparsely populated areas are not covered, and the number of observatories in developing countries is low. Satellites have the advantage that they can sample the field globally, as illustrated in figure 3.1, where the ground track covered by the Ørsted satellite during 24 hours is shown. Their ability to provide global data makes satellites invaluable for studies of the magnetic field and for generating magnetic field models. The current satellites measuring the magnetic field have provided systematic, global, dense and
homogeneous data sets of the field over the last 5–6 years.

How well a satellite covers the Earth’s surface depends on its orbit parameters. The range of geographic latitudes covered by a satellite depends on its inclination, which is the angle between the plane of the satellite’s orbit and the Earth’s equatorial plane. The highest geographical latitude reached by the satellite is equal to the value of its inclination, if this is less than 90°. If the inclination is above 90°, the highest geographical latitude reached by the satellites is 180° minus the true inclination. A satellite with an inclination of 90° would sample all geographic latitudes, but the perfect polar orbit has some disadvantages. For one thing it would take the satellite exactly half a year to sample all local times, making it impossible to separate daily and annual variations. Furthermore, to put a satellite into a polar orbit is technically difficult and very fuel consuming, and thus expensive. However, due to the tilt of the dipole axis with respect to the rotational axis of the Earth, all geomagnetic latitudes may be sampled, even if all geographic latitudes are not.

In order to get good global coverage, the satellite should also sample all longitudes. How well a satellite samples longitudes, depends on its orbital period compared to the Earth’s rotation period. If the satellite completes an integer number of orbits in 24 hours, the satellite will return to the same position after 24 hours. This is called one-day repeated track, as the satellite will have the same surface track day after day. Repeated track can occur after any given number of days. Repeated track means limited coverage in longitude and is in most cases an unwanted feature. Therefore much is done to prevent repeated track situations, but they cannot be avoided totally, as the orbital period of satellites changes with time. Due to atmospheric drag on the satellites, they continuously slow down and lose altitude; with decreasing altitude, the period decreases. From time to time the period will be in near resonance with the Earth’s rotation period, leading the satellite temporarily into a near repeated track state.

Good coverage also requires sampling the field at different local times. If only one local time is sampled, the contribution from ionospheric currents, which are highly local-time dependent, appears constant. This makes it difficult to isolate the ionospheric contribution from other contributions.

In order to get good data accuracy, it is important to know the exact time and location of each measurement, such accuracy is obtained with GPS. For vector magnetometers it is crucial to know the orientation in space, the attitude, which can be accurately measured by star imagers.

The satellites orbit at different altitudes, which have different advantages and disadvantages. For high resolution of small-scale features of the crustal field, a low altitude is preferred. But the atmospheric drag is higher at low altitude, causing a satellite at low altitude to lose altitude faster. This shortens the mission lifetime, thus the initial orbital altitude should not be too low. When choosing altitude, the effect of ionospheric currents should also be considered. It is preferable to orbit at an altitude, which has a minimum of currents and this also favours a not too low altitude.

Thus there are many different parameters to consider when designing the orbit parameters of a satellite mission dedicated to measuring the magnetic field. How to optimize the parameters depends on the primary scientific object of the specific mission.
Figure 3.1: In the top panel the locations of magnetic observatories are shown with red stars. The lower panel shows the surface track of the Ørsted satellite from a 24-hour period.


**3.2 POGO and Magsat**

The POGO (Polar Orbiting Geophysical Observatories) satellites, OGO-2, OGO-4, and OGO-6 made scalar measurements of the magnetic field from October 1965 to July 1971. Two of the satellites, OGO-4 and OGO-6, measured the field at all local times; due to a malfunction OGO-2 only made observations at local times of dawn and dusk.

In October 1979, about eight years after POGO, Magsat were launched into Earth orbit. Magsat was only in operation until mid-1980, but it nevertheless contributed significantly to the field of satellite magnetics. In addition to making scalar measurements as also the POGO satellites had done, Magsat carried a vector magnetometer and two star imagers to conduct vector measurements of the field. Magsat had an inclination of 97.15° [Langel and Hinze, 1998], thus sampling the field to about 83° geographic latitude. The orbit of Magsat was near-sun-synchronous in the dawn-dusk plane, so the span of local times sampled by the satellite was very narrow centred on the dawn-dusk plane.

Although POGO and Magsat by now have been out-performed by Ørsted, CHAMP, and SAC-C, it is important to recognize the importance of them as predecessors to the current missions. POGO and Magsat provided the basis for more than 20 years of studying satellite magnetic data and also gave the basis for wanting new and better satellite missions, in order to answer the questions the POGO and Magsat data had posed and left unanswered.

**3.3 Ørsted**

The Danish Ørsted satellite (figure 3.2), named after the Danish physicist H.C. Ørsted, was launched in February 1999. It is in a nearly polar orbit, with an inclination of 96.5°, sampling up to 83.5° geographic latitude. The slightly elliptic orbit had at the initial stages of the mission an average altitude of about 760 km, with 650 km at perigee and 860 km at apogee [Olsen, 2005]. The satellite orbits the Earth in about 100 minutes, with a speed of about 27,000 km/h or 7.5 km/s. During an orbit the Earth rotates about 25° around its own axis.

The satellite carries two magnetometers on an 8 m long deployable boom. The magnetometers are placed on a boom such that the magnetic field measurements will not be too heavily contaminated by magnetic noise from spacecraft fields. At the end of the boom there is a scalar magnetometer, which measures the total intensity of the magnetic field with a sampling rate of 1 Hz and an absolute accuracy better than 0.5 nT [Olsen, 2005]. The vector components of the field are measured with a fluxgate magnetometer situated two metres down the boom, 6 metres from the body of the satellite. It samples at 100 Hz at polar latitudes (burst mode), and at 20 Hz at lower latitudes (normal mode). Attitude is determined by a star camera placed next to the vector magnetometer. The absolute accuracy of the vector magnetometer is better than 0.1 nT [Olsen, 2005], but the accuracy of the star imager limits the accuracy of the vector data somewhat. However, the total accuracy of the data is believed to be better than 2 nT, for two of the three vector components.
Figure 3.2: The Ørsted satellite with its 8 m long boom. The magnetometers are mounted on the boom in order to reduce magnetic noise from the spacecraft. The scalar magnetometer is placed at the tip of the boom, the vector magnetometer is a few metres closer to the body of the satellite.
3.4 CHAMP

The German-built CHAMP satellite (CHAllenging Minisatellite Payload) depicted in figure 3.3 was launched in July 2000. It has a nearly circular orbit with an inclination of 87.3°. The altitude was initially 456 km, but it has decreased to about 360 km after five years in space [Maus et al., 2005b]. The orbital period of the satellite is 93.55 minutes, giving 15.4 revolutions per day. For conducting magnetic measurements the satellite is equipped with both a scalar magnetometer and a vector magnetometer, which as on Ørsted are placed on a boom for the exact same reason. The scalar magnetometer placed at the tip of the 4 m long boom, samples the field at 1 Hz with an accuracy better than 0.5 nT. The vector magnetometer placed halfway down the boom samples with 50 Hz at an accuracy better than 0.5 nT. Attitude is determined with two star cameras placed next to the vector magnetometer. More information about CHAMP can be found at the official CHAMP web site: http://www.gfz-potsdam.de/pb1/op/champ/.

3.5 SAC-C

The Argentinean satellite SAC-C was launched in November 2000. It is in a near-circular, sun-synchronous orbit sampling at 10.30 local-time. It has an altitude of about 712 km, an inclination of about 97°, and an orbital period of about 100 minutes.

The instrumentation of SAC-C is similar to that of Ørsted. It has a scalar and a vector magnetometer mounted on an 8 m long boom. For attitude determination SAC-C is equipped with a star camera. However, due to technical problems attitude data have never been available after launch, which severely limits the use of the vector data from SAC-C. There are also problems with the scalar magnetometer that unfortunately is not isotropic, which means the field value measured by the magnetometer depends on the direction of the magnetometer with respect to the ambient field. Part of this effect is corrected for, but still after calibration, the SAC-C data have an error of about 0.5–1.5 nT, which is higher than the corresponding values for Ørsted and CHAMP.
3.6 Swarm

The next generation of magnetic satellite missions is currently under development. The ESA Swarm mission, proposed by the Danish National Space Center, will consist of a ‘swarm’ of three identical satellites. By arranging the three satellites in a specific configuration it will be possible to extract much more information from the magnetic field measurements.

The Swarm satellites will have an inclination about 88°. Two of the Swarm satellites will be orbiting at an altitude of about 450 km practically next to each other, less than 150 km apart. This constellation enables us to obtain information about field gradients, leading to an unprecedented level of details of magnetic observations. The third Swarm satellite will be in a different orbit plane at about 530 km altitude to resolve some local time dependence of the external field; a sketch of the initial configuration of Swarm is shown in figure 3.4. Each satellite will have both a scalar and a vector magnetometer, and no less than three star imagers for attitude determination. Besides the advantages from the favourable constellation of the three satellites, improvements in the instrumentation means that each Swarm satellite is expected to be able to measure with an accuracy three to ten times better than that of the present missions. With the better data from Swarm, it is expected that the resolution of field models can be increased significantly.

If all goes well, the Swarm satellites will be launched late in 2009 and orbit the Earth for at least four years. Additional information about the Swarm mission can be found at e.g. http://spacecenter.dk/projects/swarm/index.html or http://www.esa.int/esaLP/LPswarm.html.
3.7 Error sources in satellite data

When measuring the magnetic field with satellites there are numerous potential error sources including instrument errors, position errors, time errors, attitude errors, digitization errors and errors due to spacecraft fields.

The absolute instrument error is determined on Earth before launch. For a good instrument the error should be time independent and uncorrelated between measurements. However, once in space the error of some of the instruments drift, which can cause correlation of the errors of consecutive measurements, i.e. serial correlation of the errors. For example, the error of one of the CHAMP magnetometers has a slight temperature dependence, and thus depends on the position of the satellite with respect to the Sun. This error is, however, believed to be less than 1 nT after correction.

Errors in position or time could potentially lead to large correlated errors, if e.g. the field values measured were systematically assigned to positions ahead of the true positions in the satellite orbit. However, with GPS the accuracy of position and time is so high that errors from this source can be considered negligible.

Attitude errors are, however, not negligible. Occasionally star imagers are not able to determine attitude, e.g. if they are blinded by the Moon or the Sun. This problem can be solved by having more than one star imager on the satellite pointing in different directions. Using more than one star camera also improves the general accuracy of the attitude. Attitude errors only affect the vector measurements and cause error correlation between the different vector components of a single measurement.

Errors from spacecraft fields are being reduced as better and better non-magnetic technology is developed. By placing the magnetic instruments on booms, the contribution from contaminating spacecraft fields are reduced even further.

3.8 Summary

Satellite measurements of the Earth’s magnetic field give dense, global, homogeneous data. Currently three satellites, Ørsted, CHAMP, and SAC-C are providing satellite magnetic data. From Ørsted and CHAMP both scalar and vector data are available, whereas only scalar data are available from SAC-C.
Chapter 4

Spherical harmonic representations of the magnetic field

The abundant magnetic satellite data available after the successful launches of Ørsted, CHAMP, and SAC-C gave rise to a new era of field modelling. Field models are of great importance in geomagnetism because they provide a useful way to describe the field in the three-dimensional space above Earth’s surface and because they allow for separation of the contributions to the magnetic field from different sources.

Although there are common characteristics among most field models, there are also some important differences between them. Different models are based on different data sets and have different model parameterizations. Some magnetic field models are based on satellite data alone, others use both satellite and observatory data, and yet others are based on observatory data alone. One of the reasons for choosing different data sets and model parameters is that the different models are developed for different purposes.

In this chapter I first introduce the central concepts of expectation, variance and covariance, as these enter the equations used to derive spherical harmonic representations of the magnetic field. Although these concepts will be used only briefly in this chapter, they will be used extensively in the two following chapters, chapters 5 and 6. In section 4.2 the concept behind how the coefficients of a spherical harmonic expansion of the magnetic field are determined from the magnetic field measurements is presented. In the remainder of this chapter I present three series of field models, which are of relevance to this work, the CM series of Sabaka et al., the OSVM series of Olsen et al., and the MF series of Maus et al..

4.1 Expectation, variance and covariance

The expectation, $E(X)$, of a random variable, $X$, is defined as the integral over the realizations, $x$, of $X$ weighted by a density function, $p(x)$, [Wackernagel, 2003]

$$E(X) = \int x \, p(x) \, dx.$$  \hspace{1cm} (4.1)
However, in most practical applications, the expectation may be approximated by
the mean of many realizations of the variable [Gubbins, 2004],

\[ E(X) = \frac{1}{N} \sum_{i=1}^{N} x_i, \]  

(4.2)

where \( x_i \) is the \( i \)'th measurement of \( X \) and \( N \) is the total number of measurements. The variance of the variable is

\[ \sigma^2(X) = E[(X - E(X))^2], \]  

(4.3)

where \( \sigma \) is the standard deviation. Given two random variables, \( X_1 \) and \( X_2 \), it is interesting to know how well they correlate. The covariance of \( X_1 \) and \( X_2 \) is defined as

\[ C(X_1, X_2) = E[(X_1 - E(X_1))(X_2 - E(X_2))]. \]  

(4.4)

Notice that the covariance of a variable with itself is the variance. If the covariance of two variables is zero, the variables are uncorrelated. A measure of the correlation of two variables is given by the correlation coefficient, \( \zeta \), which is the normalized covariance

\[ \zeta(X_1, X_2) = \frac{C(X_1, X_2)}{\sigma(X_1)\sigma(X_2)}. \]  

(4.5)

If the two variables are completely correlated, \( \zeta = 1 \), if they are completely anti-correlated, \( \zeta = -1 \), and if they are completely uncorrelated, \( \zeta = 0 \).

Given \( N \) random variables, the covariance is a \( N \times N \) matrix, whose \( ij \)'th entry is

\[ C_{ij} = C(X_i, X_j). \]  

(4.6)

If the variables are uncorrelated, then the covariance matrix will be diagonal with the variances on the diagonal, but if some of the variables depend on each other or simply correlate, the covariance matrix will have non-zero off-diagonal elements. Thus the number of non-zero elements in the covariance matrix is related to the correlation of the different variables.

If the variable in question can be considered the value of a random function, which is defined over the physical space or time, it is possible to define the autocovariance of the variable. The autocovariance function describes the covariance between the values of the random function in two difference places \( r_i \) and \( r_j \). If the variable is second-order stationary, that is if its expectation and variance do not change with position or time, then its autocovariance will depend only on the distance between the data points and not on their location. In this case the autocovariance can be written as

\[ C(r, r + h) = E[(X(r) - E(X))(X(r + h) - E(X))] = C(h). \]  

(4.7)

The autocovariance function, \( C(h) \), shows how the covariance varies with the separation or lag, \( h \). \( h \) can be a vector if the variable is distributed in more than one
dimension. For an isotropic covariance, the autocovariance function depends only on the length of $h$ and not on the direction, as opposed to anisotropic covariance, where the autocovariance function depends on both length and direction of $h$.

The autocovariance function is even and bounded by its value at the origin, the variance,

$$|C(h)| \leq C(0).$$  \hspace{1cm} (4.8)

The autocovariance function or simply the covariance function is by definition a positive definite function [Wackernagel, 2003] (see e.g. Wackernagel [2003] for an explanation of positive definiteness), thus only certain types of functions are valid covariance functions; examples can be found in Chilès and Delfiner [1999] pp. 80–90.

### 4.2 Estimation of Gauss coefficients

Magnetic field modelling is a science in its own right; a thorough treatment of this topic is beyond the scope of this thesis. To obtain more insight into field modelling the reader is referred to the many papers cited in this and the following sections. Here only some of the basic concepts behind field modelling will be presented to give the reader an idea of how the data covariance matrix appears in the field modelling procedure, and what significance it has on the model covariance matrix, i.e. the uncertainty of the Gauss coefficients and the field model.

To construct a spherical harmonic expansion of the field, the Gauss coefficients may be determined from the measured field vector components and total intensity, such that the spherical harmonic expansion fits the data. The relations between the magnetic field components and the Gauss coefficients are given in equations 2.9 and 2.10. There are different ways to determine the Gauss coefficients of a spherical harmonic expansion of the magnetic field from the data, but a method that has commonly been used is a standard least squares approach; here the linear version of this least squares problem is described.

The standard least squares approach can be used to obtain a maximum likelihood estimate of the solution of a set of linear equations, if the errors of the observations have a Gaussian distribution. Consider a set of linear equations

$$\bar{d} = Q\bar{p} + \bar{\epsilon},$$  \hspace{1cm} (4.9)

where $\bar{d}$ is the vector of observations, $\bar{p}$ is the vector of model parameters, $Q$ is the known matrix relating the observations and model parameters, and $\bar{\epsilon}$ is the unknown vector of errors of the observations. Assume that the errors, $\bar{\epsilon}$, are Gaussian distributed, random with zero mean and the same variance, and completely independent of each other. In addition assume that there are more observations than model parameters, but that all the model parameters are resolved by the data, such that equation 4.9 describes a purely overdetermined i.e. full-rank problem. In this case the best solution in a least squares sense to equation 4.9 is obtained by minimizing the norm of $(\bar{d} - Q\bar{p})$,

$$S = (\bar{d} - Q\bar{p})^T (\bar{d} - Q\bar{p}),$$  \hspace{1cm} (4.10)
with respect to the model parameters, $\bar{p}$. This yields the least squares estimate of $\bar{p}$,

$$\bar{p}_{lsq} = (Q^T Q)^{-1} Q^T \bar{d}. \quad (4.11)$$

If the variance of each observation is $\sigma^2$, then the covariance matrix of the model parameters is given by

$$C_p = (Q^T Q)^{-1} \sigma^2. \quad (4.12)$$

The diagonal elements of $C_p$ give the variances of the individual model parameters estimated, and the off-diagonal elements describe their covariance. Often $\sigma^2$ is not known in advance, and must be estimated from the observations. Provided that $Q$ is an adequate description of the underlying process, an unbiased estimate is

$$\tilde{\sigma}^2 = \frac{S_{min}}{N_d - N_p}, \quad (4.13)$$

where $N_d$ is the number of observations, $N_p$ is the number of model parameters, and $S_{min}$ is the minimum value of the sum of squared residuals given by equation 4.10 with $\bar{p}_{lsq}$ inserted for $\bar{p}$.

To achieve the solution above (equation 4.11), it was assumed that the errors were completely uncorrelated and all had the same variance, which corresponds to a diagonal data covariance matrix proportional to the identity matrix. However, in many realistic cases different (types of) data have different variances, and the errors are correlated such that the data covariance matrix,

$$C_d = E[\bar{\epsilon}\bar{\epsilon}^T], \quad (4.14)$$

contains non-zero off-diagonal elements. If the data covariance matrix is known, the least squares solution to equation 4.9, where data covariance is accounted for, is obtained by minimizing

$$S = (\bar{d} - Q\bar{p})^T C_d^{-1} (\bar{d} - Q\bar{p}). \quad (4.15)$$

The least squares solution to this problem is given by

$$\bar{p}_{lsq} = (Q^T C_d^{-1} Q)^{-1} Q^T C_d^{-1} \bar{d} \quad (4.16)$$

and the uncertainty of the model parameters is now given by the model covariance matrix,

$$C_p = (Q^T C_d^{-1} Q)^{-1}. \quad (4.17)$$

Note that the model covariance matrix given in equation 4.12 is just a special case of equation 4.17, where the data covariance matrix, $C_d$, is diagonal with the variance, $\sigma^2$, in the diagonal. Equation 4.16 gives the least squares solution of the model parameters from the set of linear equations (equation 4.9), and from equation 4.17 the uncertainty of the obtained model parameters can be determined if the data covariance matrix, $C_d$, is known. In most cases the data covariance matrix is, however, not known in advance. It is common practice in field modelling to estimate the model
parameters without using a realistic data covariance matrix, primarily because no procedure to obtain the data covariance matrix in advance is known. Although it appears that the estimates of the Gauss coefficients of at least the low degree harmonics (degrees less than about 14) do not depend significantly on the data covariance matrix [Langel et al., 1989], this is certainly not the case for the estimates of their variances. An estimate of the data covariance matrix is in some cases determined posterior to the estimation of the model parameters from the field model residuals, $(\bar{d} - Q\bar{p})$, in order to make an estimate of the variances of the model parameters.

The standard least squares approach has commonly been used in field modelling [Olsen et al., 2000a,b, Sabaka et al., 2002, Maus et al., 2004]. In this case $\bar{d}$ is the vector of magnetic field data, $\bar{p}$ is the vector of model parameters including the Gauss coefficients and the elements of $Q$ consist of the appropriate partial derivatives given by equations 2.9 and 2.10 relating the model parameters to the data. $\bar{\epsilon}$ is, however, in this case the vector of field model residuals, i.e. the difference between the data and the values predicted by the field model at the data locations. The residuals consist not only of the data errors but also of unmodelled field. The least squares solution is based on the assumption that the residuals, $\bar{\epsilon}$, are Gaussian distributed, but in most cases in field modelling, the distribution of the field model residuals has a much longer tail than a Gaussian distribution [e.g. Walker and Jackson, 2000, Olsen, 2002, Holme et al., 2003, Sabaka et al., 2004]. This has recently led field modellers to take a slightly different approach using an iteratively reweighted least squares method with Huber weights, see e.g. Olsen [2002], Holme et al. [2003] or Sabaka et al. [2004], which accounts for the non-Gaussian distribution of the residuals and which produce results that are less sensitive to data outliers than the standard least squares solution.

Some field models are derived from data from both satellites and observatories, and many include both scalar and vector data. The relation between the scalar data and the Gauss coefficients is non-linear, which means that when scalar data are used in field modelling, which most often is the case, a non-linear algorithm (as opposed to the linear least squares algorithm described above) is used. In this case the solution may e.g. be found by iteratively solving a linearized version of the set of non-linear equations.

It is common to include more than just the static field from the core and crust when deriving a field model. In many cases the secular variation, and in some cases the secular acceleration, is modelled. The low-degree external fields and their Earth-induced counterparts are also included in many models. This requires the inclusion of several other model parameters in the model parameter vector, $\bar{p}$, than just the static internal Gauss coefficients. To solve for all the model parameters from all the data simultaneously, the matrix $Q$ is expanded with the relevant relations combining the different data with the different model parameters, and the model parameters can then be co-estimated from all the data included in the data vector. Some examples of field models based on different data sets modelling different contributions from different sources of the magnetic field are presented below.

Theoretically, the spherical harmonic expansion (equation 2.9) of the magnetic field, relating the Gauss coefficients and the magnetic field potential, is only valid if the sum is taken to infinity, but in practice spherical harmonic expansions are always truncated at a finite level. At which degree the truncation is made, depends on what the specific field model is designed to model and the spacing of the available data.
CHAPTER 4. SPHERICAL HARMONIC REPRESENTATIONS

Core field models will typically be truncated at degree 10–13, whereas some crustal field models extend up to degree and order 90.

Global crustal field models can only be made with satellite data. At the surface of the Earth the strength of the crustal field can be several thousand nT, but at satellite altitude, the crustal field has reduced to a few tens of nT. The crustal field varies from broad, planetary sized length-scales to very small, local length-scales. The crustal field is expected to vary with the ocean-continent contrast, but this particular feature is difficult to observe, because it occurs on a length-scale where the core field is dominant (see e.g. Holme and Olsen [2002]). The smallest scales of the crustal field that can be resolved from satellite data, are of comparable order of magnitude as the altitude of the satellites, i.e. a few hundred km. Crustal field models derived from satellite data are in practice always band-pass filtered; it is impossible to extract the longest wavelengths of the crustal field because their signature is embedded in and completely dominated by the core field, while the smallest wavelengths are beyond the resolution that can be obtained with the available data. Inevitably, either some of the crustal signal of interest is eliminated or some of the unwanted core signal is retained when using truncation of a spherical harmonic expansion to model the crustal field, but at the moment there are no real alternatives to this method.

4.3 The CM series

The basic idea behind the Comprehensive Models (CM), primarily developed by Sabaka et al., is to co-estimate the major source fields using many different data sets. In most field models only some of the sources, e.g. the internal and magnetospheric fields, are modelled simultaneously, while fields from other sources, e.g. the ionosphere, are modelled separately. This approach may introduce errors in a field model, because the separation of the fields due to the various sources may be erroneous, and in this case the parameters of the field model are fitted to a field partly originating from a source, which is not parameterized by the field model.

The comprehensive models are based on data from both observatories and satellites. By including data sets from different altitudes, a separation of the various source fields is facilitated. The core, crustal and induced fields are internal and the magnetospheric field external to both satellite and observatory data (see figure 2.3). The ionospheric field is internal to the satellites, but external to the observatories. Thus observatory data can help separate the internal time-varying fields (induced fields and secular variation) from the fields due to ionospheric and magnetospheric sources, whereas the satellite data can separate the internal (core and crust), induced and ionospheric sources from the magnetospheric sources. Thus when analysing surface and satellite data jointly, it should at least theoretically be possible to separate magnetospheric, ionospheric, and internal+induced sources, but only if the model parameters are co-estimated. To separate the internal and induced fields, additional information is needed.

The first attempt of comprehensive modelling led to the model known as GSFC(12/93) [Sabaka and Baldwin, 1993], the second phase led to the GSFC(8/95-SqM) model [Langel et al., 1996]. Both of these models were based on POGO and Magsat satellite data as well as observatory data from quiet times. Both field models
include representations of the core and crustal field and the secular variation. Magnetospheric and ionospheric fields and their induced counterparts were modelled with local time variations, and the strength of the ring current was modelled through the $D_{st}$ index. In GSFC(8/95-SqM) seasonal variations of the ionospheric and magnetospheric fields were included.

The third model in the series, CM3 [Sabaka et al., 2002], was much more extensive than its predecessors. The description and parameterization of the sources already included in the two first comprehensive models were extended and refined, and additional sources were included in CM3. The new sources included field aligned currents and their seasonal variation, the influence of solar activity on the ionospheric field was included, the estimation of the crustal field was extended to a smaller scale (higher degree), and a more sophisticated mantle conductivity model for estimating the induced fields was applied. Only a few years after CM3 the latest comprehensive model, CM4 [Sabaka et al., 2004], was derived. The major difference between these two models is the inclusion of data from Ørsted and CHAMP in CM4; the parameterization of the field sources in the two models is basically the same.

CM3 was derived from quiet time POGO, Magsat, and observatory data, and spans the period 1960–1985. The field model had 16,594 parameters, which were estimated from the 591,432 data points via weighted least squares inversion [Sabaka et al., 2002]. For CM4 the data set was extended with observatory data through 2000, and with vector and scalar data from Ørsted and scalar data from CHAMP up to July 2002. CM4 thus spans from 1960 to mid-2002. The inclusion of Ørsted and CHAMP data in CM4 meant a great improvement from CM3. CM4 has 25,243 model parameters, which are estimated by an iteratively reweighted least squares method from the 2,156,832 data points [Sabaka et al., 2004].

Both CM3 and CM4 describe the static field up to degree and order 65. The secular variation, which obviously has to be taken into account in field models extending over several decades, is modelled up to degree and order 13. The models account for field aligned currents and the ionospheric and magnetospheric fields along with their induced counterparts and all their systematic and less systematic variations. Specifically, the strength of the ring current is assumed to vary proportionally to the $D_{st}$ index.

The $K_p$ and $D_{st}$ activity indices are used to select quiet time data. The selection criteria for the satellite data are $K_p \leq 1^+$ ($K_p \leq 1^-$ for Magsat) at the time of observation, and $K_p \leq 2^0$ in the previous three-hour interval, while the $D_{st}$ index had to be within $\pm 20$ nT. Both scalar and vector data were used for all latitudes and for all local times. The data were resampled with a one minute resampling rate (i.e. one data point per minute), and weighted proportionally to $\sin(\theta)$ (geographic colatitude) in order to simulate an equal-area distribution. Magsat and Ørsted vector data were additionally weighted to account for attitude error [Holme and Bloxham, 1996].

Due to the simultaneous description of most of the known field sources, the comprehensive models have extensive applications as e.g. reference models. In addition, due to the co-estimation of all the source fields, the comprehensive models allow for a division of the observed signal among the various sources. This makes the comprehensive models advantageous for studies where isolation of the field from one or several specific sources is desired.
4.4 The OSVM series

The OSVM (Ørsted Main and Secular Variation Models) series of Olsen et al. includes the Ørsted Initial Field Model (OIFM) [Olsen et al., 2000a], the Ørsted Main and Secular Variation model (OSVM) [Olsen, 2002], the Ørsted(09d/04) [Olsen et al., 2005b], and the Ørsted(06s/05) field models, all dedicated to model the magnetic field within the last 5–6 years.

Some common properties of the field models of the OSVM series are that they are based primarily on satellite data. They model the static field (core and crustal field) and the external magnetospheric fields including their induced counterparts. In all the models except OIFM, secular variation is estimated, and in some of the latest models (e.g. Ørsted(06s/05)) the secular acceleration is also modelled.

The earliest OSVM field models were based on Ørsted data alone (e.g. OIFM). Subsequently, CHAMP scalar data and later CHAMP vector data were included. In the most recent of the OSVM field models, SAC-C scalar data are also included. Some of the models include observatory data (e.g. OSVM), primarily to constrain the secular variation.

Similar data selection criteria have been used in the various OSVM field models, with some minor variations. All models are based on data from quiet times, using $K_p$ and $D_{st}$ indices for data selection. The restriction imposed by the $K_p$ index is that $K_p \leq 1^+$ at the time of observation, and $K_p \leq 2^0$ in the previous three-hour interval. For some models (e.g. OIFM and OSVM) it was required that the $D_{st}$ index was within $\pm 10$ nT, and that the temporal variation of the $D_{st}$ index was low, $|d(D_{st})/dt| < 3$ nT/h. However, the restriction on the $D_{st}$ index has later been relaxed. Olsen et al. [2005b] found that the baseline of the $D_{st}$ index appears to be varying in time, implying that absolute values of the $D_{st}$ index from different years not necessarily correspond to the same level of disturbance. Thus for some of the later field models (e.g. Ørsted(09d/04)), there is no restriction of the absolute value of the $D_{st}$ index in the data selection, only on the temporal variation of the $D_{st}$ index, which for Ørsted(09d/04) was that the $D_{st}$ index is not allowed to change more than 1 nT/h [Olsen et al., 2005b].

To minimize the effect from ionospheric currents at middle and low latitudes only data from dark regions, i.e. where the Sun is more than $5^\circ$ below the horizon, are used. In order to minimize the effect of ionospheric currents over the polar cap, data in the polar region are only included if the dawn-dusk component of the interplanetary magnetic field is less than 3 nT.

In general, vector data are used in the non-polar regions, whereas scalar data are used in the polar regions and whenever vector (i.e. attitude) data are not available. The limit for the use of vector data varies slightly between the different OSVM field models. In the OIFM and OSVM models, vector data are used for geomagnetic latitudes equatorward of $\pm 50^\circ$ and in the Ørsted(09d/04) model the limit is set at $\pm 60^\circ$ geomagnetic latitude. In the OIFM field model a resampling rate of 30 seconds was applied, while in all the later OSVM field models the data were resampled with a one-minute resampling rate. In all the models but OIFM the resampled data were weighted proportionally to sin($\theta$) (geographic colatitude) in order to simulate an equal-area distribution; in the OIFM model equal-area distribution was approximated by a slightly different method.
4.5. THE MF SERIES

The Gauss coefficients and other model parameters of OIFM were estimated by weighted least squares fit to the data, while in OSVM and all the later models an iteratively reweighted least squares approach with Huber weights were used. To which degree the static field is modelled varies somewhat; in the OIFM field model the static field is determined up to degree and order 19, in OSVM up to degree 29, in Ørsted(09d/04) up to degree 32, and in Ørsted(06s/05) up to degree 50. For most of the models, the Gauss coefficients with the highest degrees are not considered robust.

The approach taken to model the external magnetospheric field differs somewhat between the different OSVM models. Common for all the OSVM field models is, however, that the magnetospheric field is modelled up to degree and order two, the static magnetospheric field is modelled and its seasonal variability is included, and as in most other field models the time-varying strength of the magnetospheric ring current is estimated from the $D_{st}$ index. The induced counterpart of the ring current is also included in the models. There are, however, several difficulties with the use of the $D_{st}$ index when modelling the time-varying ring current strength as mentioned in the many papers describing OSVM field models. In the OSVM field model, the delay of the determination of the final $D_{st}$ index meant that the $D_{st}$ was not available for the latest data used in the model. Therefore a modified index based on worldwide observatory data was derived for this particular field model. This had the advantage that the index could be derived from observations in the same local-time sector as the selected field model data; the $D_{st}$ index is based on data from all local times, whereas the data used for the OSVM models are dark-side data only.

A second issue regarding the use of the $D_{st}$ index for modelling the ring current strength was raised in Olsen et al. [2005b], who found that the baseline level of the $D_{st}$ index has varied considerably over the past 40 years. This is the reason why the criterion on the absolute value of the $D_{st}$ index is not used for data selection for e.g. the Ørsted(06s/05) field model. To accommodate the problem of baseline instability, the static part of the external field was recalibrated for every month in the Ørsted(09c/04) and Ørsted(09d/04) field models. In Ørsted(06s/05) the static external field was recalibrated on a daily basis. Furthermore, Olsen et al. [2005b] suggested that as the induced counterpart of the ring current has a different temporal behaviour than the ring current due to the mantle conductivity, these two contributions should not be parameterized by the same index. Therefore they split the $D_{st}$ index into two parts, $E_{st}$ and $I_{st}$, parameterizing the external ring current and its induced counterpart, respectively. A similar decomposition was suggested independently by Maus and Weidelt [2004].

4.5 The MF series

The MF (Magnetic Field) models of Maus et al. are designed to isolate and represent the crustal (lithospheric) field only. The purpose is to determine the crustal field with sufficient precision such that it may be used for geological and geophysical interpretation, thus great effort is made to ensure that the high-degree coefficients are as accurate as possible.

The MF series consists of four field models: MF1 [Maus et al., 2002], MF2, MF3 [Maus et al., 2004], and MF4 [Maus et al., 2005b]. The model
coefficients as well as additional information can be found at http://www.gfz-potsdam.de/pb2/pb23/Models/model.html.

The MF field models are all predominantly based on CHAMP data. The differences of the models consist of some improvements in model parameterization, but the major difference is the amount of data included in the models, which increase as additional observations are made. MF1, which is expanded from spherical harmonic degree 15 to 80, was based on one year of CHAMP scalar data obtained during the first year of the mission. In MF2, which was expanded from degree 16 to 80, two years of scalar and one year of CHAMP vector data were included. MF3, was made from three years of scalar and two years of vector data, and expanded from degree 16 to 90. The most recent model in the MF series, MF4, is based on almost five years of scalar and vector data and is as MF3 expanded from degree 16 to degree and order 90.

To derive the crustal field models, the observations of the field are corrected for field contributions from all other major sources. The core field is removed by subtracting the values predicted by a core field model up to degree 15. The external fields are corrected for through an external field model to degree two. Induced fields and fields from tidal currents in the oceans are removed, and finally a track-by-track filtering to remove additional contributions from external fields is applied. The latter method is used to remove contributions from the polar electrojet and what additional magnetospheric ring current field that is left. Care must be exerted when applying the track-by-track filtering, as there is a great risk of also removing north-south trending crustal signals in the process.

One of the improvements of MF4 over MF3 is that a better correction for the polar electrojets for mid-latitude data is used in MF4. The core field model used in MF3 is the Ørsted(10b/03) field model, which is based on five years of Ørsted data, three years of CHAMP data and observatory measurements. In MF4 the POMME-2.5 [Maus et al., 2005a] field model is used. Both of these core field models include secular variation, secular acceleration and external and induced fields. These core field models will, however, also contain long-wavelength parts of the crustal field, which unfortunately also are subtracted from the data.

The CHAMP data used consist of quiet time scalar and vector data. Only vector data from latitudes equatorward of 65° geomagnetic latitude are used in MF4, while in MF3 the division was made at 50° geomagnetic latitude. Different selection criteria were used to find quiet time data for polar and non-polar data. For the non-polar scalar and vector data, measurements made at local times between midnight and 5.00 a.m. obtained in periods where \( K_p \leq 2 \), were used. The polar scalar data were also selected differently, but the criterion was in neither case based on the \( K_p \) index. In MF3, data from the tracks across the poles, which had the lowest root mean square deviation from the field predicted by MF2, were used [Maus et al., 2004]. In MF4, the selection was based on the behaviour of the interplanetary magnetic field [Maus et al., 2005b]. Once the quiet time data had been reduced to what is believed to originate from the crust only, they were weighted to give an equal cumulative weight to every unit area on the sphere, and subsequently the spherical harmonic coefficients representing the scalar potential of the crustal field were estimated by using a standard least squares method, with some regularization to dampen ill-determined coefficients.
4.6 SUMMARY

The resulting crustal field models, MF3 and MF4, provide good representations of the crustal field down to altitudes of 50–100 km [Maus et al., 2004, 2005b]. The noise on the models is somewhat larger in the polar regions than at geomagnetic mid and low latitudes as would be expected. In MF3, the noise appears larger in the southern polar region than in the northern polar region, in MF4 the opposite is the case.

4.6 Summary

To derive spherical harmonic representations of the magnetic field, the model parameters including the Gauss coefficients, are estimated by solving an overdetermined system of equations. There are two main approaches to this: a standard least squares approach has been widely used, but lately an iterative least squares method with Huber weights has been employed to account for the non-Gaussian distribution of the data errors.

Many different field models exist based on different types of data, selected from different criteria, and designed to model different features of the magnetic field. Three major series of field models were presented in this chapter: the comprehensive models, which are designed to co-estimate as many source fields as possible, the OSVM models, designed to model the present field, and the MF field models designed to model the crustal field. In chapter 6 I present the results of an analysis of the residuals of the Ørsted(06s/05) field model of the OSVM series. In chapters 7 and 8, we derive the magnetic crustal thickness and the geothermal heat flux in Greenland and Antarctica from MF3 and CM4.
Part II

Uncertainty estimates of field models – Variogram analysis of field model residuals
Chapter 5

Correlated errors, covariances and variograms

As a field model is a mathematical representation of the physical field, and as the magnetic field varies on a wide range of time-scales and length-scales, a field model can never become more than an approximation to the true field. Currently we are far from being able to predict or even explain all the variations that the magnetic field displays, and it is highly questionable whether we ever will be able to predict the behaviour of the field completely. Being only an approximation to the field, any field model will have some degree of uncertainty. It is important to have a feeling for this level of uncertainty as it is a measure of the validity of a given field model.

In this chapter I first present some earlier studies of error correlation and field model uncertainty estimates in order to motivate why it is important to determine the covariance of the field model errors. In the remainder of the chapter I will introduce the theory behind the approach we take to determine the correlation of field model residuals; the analysis itself is presented in the following chapter.

5.1 Uncertainty estimates of field models

An ideal way to make uncertainty estimates of field models is to first repeat the observations several times at different locations giving several independent but similar data sets, and then derive a field model from each data set. The distributions of the resulting Gauss coefficients would then give the uncertainty of all the derived field models. However, evaluating the Gauss coefficients of a field model, particularly for the ones including the crustal field, the external field, secular variation etc., requires extensive computations, thus calculating e.g. one hundred field models to make an average, is quite a formidable task, but more importantly, we only have few independent data sets.

For most field models developed nowadays, estimates of the variances of the Gauss coefficients are not published, since the estimates that are made, as pointed out by Lowes and Olsen [2004], can be very wrong, because they are based on assumptions which in many cases are not valid. These assumptions are that the measurement errors are uncorrelated and Gaussian distributed, that the errors assigned to the individual measurements are correct, and that the field model in question is adequate
and thus solves for all significant systematic variations.

If the field model does not take all fields into account, and we may at this point not even know all of them, it is likely that there will be systematic errors from unmodelled fields. Correlated errors due to unmodelled fields are observed in field modelling. Holme et al. [2003] found that the errors on the data used for the CO2 field model were highly correlated within a given orbit, but often also in between orbits; in this particular case the primary cause of the correlated errors was believed to be an unmodelled magnetospheric field. By assuming that the errors are uncorrelated when estimating the variances of the Gauss coefficients, the estimated level of uncertainty may be quite wrong [Lowes and Olsen, 2004], and thus we have no reliable uncertainty estimates of the field models.

Accounting for error correlation between the data points when estimating the Gauss coefficients and their variances, requires the inclusion of off-diagonal elements in the data covariance matrix. Even though the data covariance matrix may be sparse, the non-zero off-diagonal elements will nevertheless mean a large extra computational task; something that is very undesirable. A way to reduce the number of non-zero off-diagonal elements in the covariance matrix is to use a less dense data sampling alleviating serial correlation. But although this would give a better estimate of the variances of the Gauss coefficients, it would also reduce the resolution of the high-degree harmonics, as we would be throwing away valuable information, even though it is in a complicated form.

The deficiencies in the error estimates of field models have been known for a long time, and several studies have been made on how to make more realistic uncertainty estimates e.g. by Langel et al. [1989], Jackson [1990], Rygaard-Hjalsted et al. [1997] and Lowes and Olsen [2004], all using different approaches and focusing on different issues.

Langel et al. [1989] looked at the uncertainty of core field models and found that the uncertainty is underestimated when the modelling procedure does not take the field from truncated terms, crustal fields and external fields into account. They used statistical representations of the unknown fields from the ionosphere, the crust and the truncated terms. They used Magsat data and compared the results for core field models with and without error correlation and found that the inclusion of correlated errors increased the error estimate of the Gauss coefficients up to 7.5 times, whereas the values of the coefficients themselves remained much the same. Langel et al. [1989] concluded that not accounting for error correlation in field modelling does not significantly influence the results of the Gauss coefficients, only their uncertainty estimates. Their field models were, however, only derived to degree and order 13; it is possible that using a realistic data covariance matrix does affect the estimate of higher degree Gauss coefficients.

Jackson [1990] looked at the effect of the crustal field on core field models when satellite magnetic data are used for the field modelling. The crustal field influences a core field model from both low-degree crustal field that overlaps the core field and from truncated terms, which are dominated by the crustal field. Jackson [1990] found that if it is assumed that the crustal magnetization is completely uncorrelated, the crustal field will nevertheless be correlated out to 15° separation at an altitude of 400 km. Thus if the crustal magnetization is correlated, and geologic evidence suggests that it is, then the correlation length of the crustal field at satellite altitude will
be even larger. This observation underlines the necessity for taking error correlation due to e.g., the crustal field into account when making core field models.

Rygaard-Hjalsted et al. [1997] also looked at the effect of making erroneous assumptions about the contribution from the crustal field in core field models. They investigated the validity of the previously widely used assumption, that the contribution to the data error from the crustal field, in core field models, is white noise. Rygaard-Hjalsted et al. [1997] used a statistical representation of the crustal field, in which the crustal magnetization was given as a realization of a stationary, isotropic random process. They found that at a height of 400 km this field was correlated over ranges of 15–20°. They also found that the Gauss coefficients for models where correlation was accounted for compared to models where it was neglected, differed by as much as 27%, with the largest differences in the harmonic degrees 8–13. As Langel et al. [1989], they worked with Magsat data, which was the best satellite data available at the time, but contrary to Langel et al. [1989], Rygaard-Hjalsted et al. [1997] used a damped model, which probably is the reason for the different result.

Rygaard-Hjalsted et al. [1997] concluded that failure to include off-diagonal elements in the covariance matrix due to correlation of the crustal field in core field modelling, leads to biased and inconsistent core field models. They suggested that using a data set with large data spacing would be one way to reduce the problem of error correlation and avoid a non-diagonal covariance matrix. Otherwise, suitable covariances to account for unmodelled field should be included in the field modelling. Also they pointed out that for the specific task of generating better core field models, a better description of the crustal field was needed.

Today the crustal field is known better, and it is accounted for in field models derived from satellite magnetic data. Due to the arrival of new satellite data, especially data from the relatively low orbiting CHAMP satellite, the crustal field models have improved greatly in the last 6 years and are now robust up to about degree 60. Two of the approaches taken in core field modelling today are to either co-estimate the crustal field or to subtract a crustal field, determined from another field model, from the data before deriving the core field model, and in this way avoid the contribution from the crustal field.

Lowes and Olsen [2004] worked with a method of splitting the data into different subsets and run all data subsets through the modelling procedure, and then compare the results of the field models derived from the different subsets. The differences in the Gauss coefficients give estimates of the variances of the individual coefficients. This method was developed to mimic uncertainty estimates based on comparison of the field model with independent data not included in the field modelling. The reason for using the subset approach was to be able to include all data into the field model. Lowes and Olsen [2004] suggest that the method of using subsets is favourable: in this way the whole data set can be used to make the field model, but more realistic variances can be obtained than by using a simple diagonal covariance matrix, and the computational costs of including many off-diagonal elements can be avoided. However, care has to be exerted to ensure that the data are subdivided into independent subsets. Furthermore, if each subset is to have a significant amount of data, only a few non-overlapping subsets can be made.

Lowes and Olsen [2004] looked at the variances of the Gauss coefficients of the OSVM field model (see section 4.4) as an example, but showed that the distribution
CHAPTER 5. CORRELATED ERRORS

Figure 5.1: From Lowes and Olsen [2004]. The variances of the individual Gauss coefficients of the OSVM field model. The order is $g_1^0, g_1^1, h_1^1, g_2^0$, etc. It is seen that the variances of the Gauss coefficients follow a pattern depending on their degree and order. For a given degree, the variances of the zonal and near-zonal, and the sectoral and near-sectoral terms are higher than the intermediate terms. The horizontal lines are what Lowes and Olsen [2004] expected for a uniformly distributed set of vector data at an altitude of 700 km.

of the variances they obtained will be similar for other field models. In general, the variance will increase with degree due to downward continuation; the measurements are made at satellite altitude, whereas the Gauss coefficients are expressed at the surface. Additionally, high degree harmonics require better data coverage than low degree harmonics to be well resolved; thus in general, low degree harmonics are better constrained by the data than high degree harmonics. Most field models are like the OSVM model based on scalar data in polar regions and vector data in non-polar regions. Lowes and Olsen [2004] showed that this data distribution gives a characteristic variation of the variances, with the variance of the individual Gauss coefficient depending on its degree and order. For a given degree, $n$, the variance is higher for the zonal ($m = 0$) and near-zonal terms, and for the sectoral ($m = n$) and the near-sectoral terms, while the variance is lower for the intermediate values of $m$.

Figure 5.1 shows the variances they found.

Lowes and Olsen [2004] also found that systematic errors were present in this field model due to insufficiently modelled magnetospheric and ionospheric currents, and they showed that these errors enhance the variation with degree and order of the variances given by the data distribution. Magnetospheric currents are included into the field models with their time-varying contribution being modelled as a function of the $D_{st}$ index. However, in spite of the efforts to account for magnetospheric currents in the field models, there are still magnetospheric contributions which are unaccounted for. The OSVM field model was made of data with a 1 minute sampling rate, which was chosen to retain as much data as possible without getting too much
serial correlation. However, with a 1 minute sampling rate, the magnetospheric field introduce along-track correlation of the errors. Lowes and Olsen [2004] found that serial correlation tends to increase the variance of the sectoral and near-sectoral coefficients and decrease it for other harmonics, so that the variances of these latter coefficients become lower than they would have been if the error had been uncorrelated. In order to avoid contamination from ionospheric fields, the OSVM field model is based on night-time data. However, it turns out that this criterion is not sufficient to completely avoid ionospheric fields. Lowes and Olsen [2004] found from the CM3 model that there still are fields of a few nT at satellite altitude originating from ionospheric fields at night. It is not the ionospheric currents themselves that gives these contributions, but their induced counterparts, that are still present at night-time. These ionospheric fields make significant contributions to the variances of the zonal \((m = 0)\) and near-zonal coefficients, which are increased.

Based on their analysis Lowes and Olsen [2004] concluded that it would be realistic to use a simple overall uncertainty of 25% of the Gauss coefficients. They also advise that whatever method is used to determine the variances of the Gauss coefficients, it would be wise to compare the obtained variances with estimates based on their subset method.

We will look at error correlation of the data used to make the Ørsted(06s/05) field model. Our purpose is to determine the correlation of the field model residuals in order to find the data covariance matrix. Our approach is to make a direct statistical study of the field model residuals using variograms to investigate the behaviour of the residuals and find their covariance. With the direct approach it is neither necessary to make any assumptions about the error or variance of the individual measurements, nor to consider whether the field model is sufficiently good at resolving all non-random parts. Instead the answers to both of these considerations will emerge from the results of the analysis, which will be presented in detail in chapter 6. In the remainder of this chapter I will introduce the tools necessary to carry out the analysis.

5.2 The variogram

We want to analyse the behaviour of the satellite magnetic data left unexplained by a field model developed from the data: the field model residuals. We take a statistical approach to this using the geostatistical tool of the variogram to find the spatial and temporal variability of the residuals. Taking a statistical approach, we assume that the variable in question is a random variable that can be described stochastically. The residuals can be regarded as such; they are what remains of the data when the deterministic part of the data, the field model, has been subtracted. We can consider the residuals an estimate of the noise of the data, keeping in mind that they consist of both instrument error and unmodelled signal. Although it is not possible to predict the error on a given measurement in advance, it is possible to say something about how the error will behave if its statistical properties such as mean and variance are known. I will not go into the details of geostatistics here, but only introduce the geostatistical tool of the variogram, which we will use to estimate the covariance of the field model residuals. An extensive introduction to geostatistics can be found in e.g. Chilès and Delfiner [1999] or Wackernagel [2003].
Figure 5.2: Example of a variogram. For a phenomenon with finite variability, the variogram approaches a limit, the sill. The range is the distance where the sill is (almost) reached.

The variogram describes statistically how the values of two measurements differ with increasing distance between them. Given two measurements \((r_i, F_i)\) and \((r_j, F_j)\) the dissimilarity is defined as

\[
\gamma_{ij} = 0.5 (F_i - F_j)^2 .
\]  

(5.1)

In the variogram, the dissimilarity is plotted as a function of the distance between the data points,

\[
r_{ij} = \sqrt{r_i^2 - r_j^2}.
\]  

(5.2)

For a well-behaved physical quantity, it seems reasonable that if the value of the quantity is measured at two points located very close to each other, the two measurements will have almost the same value, and the dissimilarity will be small. If, however, the quantity is measured at two points far apart, the values obtained may be very different, and the dissimilarity will be high. Thus a typical variogram looks like the one shown in figure 5.2. The variogram can increase indefinitely with increasing distance, if the variability of the variable has no limit. But if the variability is limited, the variogram approaches a limit, the sill, and flattens. The distance where the sill is (almost) reached, the range, marks the distance beyond which the values of the variable are completely uncorrelated.

For a second-order stationary variable with limited variability, the variogram, \(\gamma (h)\), and the covariance function, \(C (h)\), are related through

\[
C (h) = \gamma (\infty) - \gamma (h) = C (0) - \gamma (h)
\]  

(5.3)
5.2. THE VARIOGRAM

Figure 5.3: An example of a non-stationary signal, the blue line is the measured signal, it obviously consists of a trend, the black line, and a high-frequency signal, which is the object of interest. If the trend is removed from the signal, the detrended signal (red line) can be subjected to variogram analysis. Axes are arbitrary.

where it is noticed that the sill, $\gamma(\infty)$, is equal to the variance, $C(0)$. The variogram is thus an even, nonnegative function, which has $\gamma(0) = 0$. Due to the intimate relation between the variogram and the autocovariance function the two terms will be used interchangeably.

For our work, presented in chapter 6, we determine the variogram from the data and use equation 5.3 to find the autocovariance from the variogram, thus using the variogram as the tool to find the autocovariance function. The reason for working with the variogram is that the variogram is a more general tool than the autocovariance. The variogram exists for variables that have an unlimited variability, whereas the autocovariance function only exists for variables with a limited variability. Furthermore, the variogram has the advantage of not requiring knowledge of the mean of the data, as the autocovariance does. If the autocovariance is determined directly from data, the mean has to be estimated also from the data, and this can introduce a bias [Chilès and Delfiner, 1999, and references therein]. The variogram also has the advantage of easily handling data with irregular data spacing (in both time and space). To some extent the variogram works better on irregular data grids than on regular, because the irregular grids get a larger variety of distances, whereas a regular sampling grid only has discrete distances between the data points.

5.2.1 Stationarity and detrending

The variogram is based on statistics of data pairs. The underlying assumption of the variogram is that all data pairs separated by the same distance vector are a measure
of the same thing, and that it makes sense to average properties of the data pairs. Thus some sort of homogeneity of the data is a necessary condition if the pooling of "two-points" used in the variogram, is to make sense. This homogeneity condition is that the variable must be second-order stationary on a length scale of the domain size. In the remainder of this thesis, stationarity refers to second-order stationarity.

In many practical applications within geophysics a high-frequency signal of interest is superposed on a low-frequency variation, as illustrated in figure 5.3. For example a daily varying signal measured over a period of some months superposed on an annual variation; in this case the annual variation will be seen as a trend in the data. Data with a trend are obviously non-stationary, as the mean depends on time (or location), and they should not be subjected to variogram analysis. However, with proper data pre-processing, variogram analysis may nevertheless be used in some cases. If it is possible to detrend the signal by removing the trend, leaving a stationary signal containing the higher frequency fluctuations around the trend, variogram analysis can be carried out on this remaining signal. The trend function must be chosen with great care. The residuals obtained from subtracting the trend function from the original data, will depend on the choice of the trend model. A wrong trend function will introduce cross-correlation into the system and create bias. Thus whenever possible, the choice of the trend function should be motivated by a physical explanation.

### 5.2.2 Experimental and theoretical variograms

The variogram comes in several versions including the variogram cloud, the experimental variogram, and the theoretical variogram.

The variogram cloud consists of the dissimilarities of all the possible data pairs plotted as a function of their distance, as shown in figure 5.4A. To display some of the interesting properties of the variogram cloud in a more manageable form, the experimental variogram is used: The data pairs are binned into different distance classes depending on their separation, and the value of the experimental variogram assigned to each distance class is the mean value of the dissimilarities within the given distance class. The experimental variogram is shown as the bars in figure 5.4B. The theoretical variogram is found by fitting a valid variogram function to the experimental variogram. From the theoretical variogram the autocovariance can be found, if the theoretical variogram is limited.

According to Chilès and Delfiner [1999] the variogram is usually not considered reliable if less than 50 data pairs have been used to make it. However, based on experiments with synthetic data, we find that this number is much too low. To ensure good statistics and a stable solution a large number of data points is necessary, and the data have to be fairly well behaved as well. If the data display regional differences, with some regions having higher average values and higher variability than others, this can corrupt the variogram. The variogram will in this case to a larger extent reflect the variation in the average value, rather than the actual spatial variability [Chilès and Delfiner, 1999], and in this situation regional variograms should be used. This is relevant for our analysis of field model residuals, which have higher values and larger variability in the polar regions than in the non-polar regions.

The variogram should not be analysed for distances larger than half of the width
5.2. THE VARIOGRAM

Figure 5.4: Panel A shows a variogram cloud. Panel B shows the same variogram cloud, including the experimental variogram (bars) and the theoretical variogram (solid line).

of the data domain for two reasons. First, the number of data pairs within a given distance class will decrease as the distance approaches the width of the data domain. Secondly, data pairs with inter-point distances more than half the width of the data domain do not sample the data domain properly, because the entire domain is not sampled.

5.2.3 Nugget effect

By definition the value of the variogram at the origin is zero, \( \gamma(0) = 0 \). However, in some cases the variogram has a discontinuity at the origin, an offset, and assumes a finite value

\[
\lim_{h \to 0} \gamma(h) = \eta > 0.
\]  \hspace{1cm} (5.4)

This is known as a nugget effect and occurs if part of the signal being analysed is uncorrelated between measurements. If the instrument errors are uncorrelated between measurements and independent of the measured variable, they will show up in the variogram as an offset that is equal to the variance of the instrument error [Chilès and Delfiner, 1999]. An offset of the variogram also occurs if the short range variability of the variable is shorter than the shortest distance between the data points.

In many cases both effects may be present and contribute to an offset. It is impossible from the variogram to tell how much of the offset that originates from which effect. It is, however, in most cases known if an instrument error is present such that an offset should be expected in the variogram, and in some cases the instrument error may even be known.
5.2.4 Finding the theoretical variogram

The experimental variogram can be calculated more or less automatically from the data, but there is no automatic way of fitting a theoretical variogram to the experimental. Several choices, which require some insight into the problem, have to be made. First the type of variogram function to be fitted must be chosen; several different functions with different properties are available (see e.g. Chilés and Delfiner [1999] pp. 80–90, for a variety of models). A second choice to be made is whether to allow for a nugget effect or force the chosen variogram function to pass through the origin. If a known instrument error is present, the variogram function can be forced to pass through the variance of the instrument error at zero distance. One can choose to use a single variogram or combine several functions and make complicated nested structures of the variogram.

These choices can significantly influence the results, so whenever possible the choices made should be based on knowledge of the physics of the variable being analysed. Once these critical choices have been made, there are systematic ways to optimize the fit of the theoretical variogram to the experimental, in much the same way that a function is fitted to a data set in other cases e.g. by ordinary least squares, generalized least squares or weighted least squares.

5.3 Spatiotemporal covariance

So far it has been implicitly assumed that the variable in question depends solely on either time or space. However, many geophysical quantities, including the magnetic field, vary in both space and time. Spatial and temporal variations are in most cases caused by fundamentally different physical processes, and a covariance function of a spatiotemporal variable needs to take both the temporal and the spatial variation into account. There are several different approaches to this; a review of different geostatistical space-time covariance models is given by Kyriakidis and Journel [1999], and several types of space-time covariance functions are mentioned and compared in De Cesare et al. [2001b].

We have chosen to work with the product-sum covariance function proposed by De Cesare et al. [2001b,a], where the spatiotemporal covariance function is given as

\[ C_{st}(h_s, h_t) = k_1 C_s(h_s) C_t(h_t) + k_2 C_s(h_s) + k_3 C_t(h_t), \]  

(5.5)

where \( C_t(h_t) \) is a purely temporal covariance function depending only on the temporal lag, \( h_t \), and \( C_s(h_s) \) is a purely spatial covariance function depending only on the spatial lag, \( h_s \), which in the general case can be a vector in three-dimensional space. \( C_{st} \) is a valid spatiotemporal covariance function if \( k_1 > 0, k_2 \geq 0 \) and \( k_3 \geq 0 \) [De Cesare et al., 2001b]. We have chosen to use the product-sum covariance model because the constants and the temporal and spatial covariance functions are easily determined from the spatiotemporal variogram of the data. It is positive definite for all choices of valid covariance functions of \( C_t \) and \( C_s \) (with the above constraint on the values of the constants), and has sufficient flexibility to model the data well. A comparison of the product-sum covariance function with other spatiotemporal covariance functions is given in De Cesare et al. [2001b].
5.3. SPATIOTEMPORAL COVARIANCE

The spatiotemporal variogram corresponding to the product-sum covariance function is

\[
\gamma_{st}(h_s, h_t) = [k_2 + k_1 C_t(0)] \gamma_s(h_s) + [k_3 + k_1 C_s(0)] \gamma_t(h_t) - k_1 \gamma_s(h_s) \gamma_t(h_t), \tag{5.6}
\]

where \(\gamma_t(h_t)\) and \(\gamma_s(h_s)\) are the variograms corresponding to \(C_t\) and \(C_s\), and \(C_t(0)\) and \(C_s(0)\) are the sill values of \(\gamma_t\) and \(\gamma_s\), respectively. By definition \(\gamma_{st}(0,0) = \gamma_t(0) = \gamma_s(0) = 0\). The global sill of the spatiotemporal variogram, \(C_{st}(0,0)\), is related to \(C_s(0)\) and \(C_t(0)\) through

\[
C_{st}(0,0) = k_1 C_s(0) C_t(0) + k_2 C_s(0) + k_3 C_t(0). \tag{5.7}
\]

The spatiotemporal covariance and variogram are two-dimensional, as the distance between two data points consists of both a temporal distance and a spatial distance. The experimental spatiotemporal variogram is found directly from the data, and from this the spatial and temporal covariance functions \(C_s\) and \(C_t\), as well as the constants \(k_1, k_2, \text{ and } k_3\), must be determined in order to find the theoretical spatiotemporal variogram. First, the spatial and temporal variograms \(\gamma_s\) and \(\gamma_t\) are estimated by looking at the marginal variograms, which are the intersections of the spatiotemporal variogram with the planes \(h_t = 0\) and \(h_s = 0\). From equation 5.6 it is seen that the spatial marginal variogram is

\[
\gamma_{st}(h_s, 0) = [k_2 + k_1 C_t(0)] \gamma_s(h_s) \tag{5.8}
\]

and the temporal marginal variogram is

\[
\gamma_{st}(0, h_t) = [k_3 + k_1 C_s(0)] \gamma_t(h_t). \tag{5.9}
\]

To estimate \(\gamma_s\) and \(\gamma_t\) from the marginal variograms, it is assumed that [De Cesare et al., 2001b,a]

\[
k_2 + k_1 C_t(0) = 1 \tag{5.10}
\]

\[
k_3 + k_1 C_s(0) = 1, \tag{5.11}
\]

so \(\gamma_s\) and \(\gamma_t\) are the best fitting valid variogram functions to the marginal variograms. The sill values, \(C_s(0)\) and \(C_t(0)\), are determined from \(\gamma_s\) and \(\gamma_t\), whereas the global sill, \(C_{st}(0,0)\), is estimated directly from the experimental two-dimensional spatiotemporal variogram. The constants \(k_1, k_2, \text{ and } k_3\) can subsequently be determined from the sill values; combining equations 5.7, 5.10 and 5.11 gives

\[
k_1 = \frac{[C_s(0) + C_t(0) - C_{st}(0,0)]}{C_s(0) C_t(0)} \tag{5.12}
\]

\[
k_2 = \frac{[C_{st}(0,0) - C_t(0)]}{C_s(0)} \tag{5.13}
\]

\[
k_3 = \frac{[C_{st}(0,0) - C_s(0)]}{C_t(0)}. \tag{5.14}
\]

As mentioned earlier the constants must satisfy \(k_1 > 0, \ k_2 \geq 0\) and \(k_3 \geq 0\) for equation 5.5 to be a valid covariance function. One consequence of this is that the
global sill value must be higher than either of the marginal sills, but less than their sum. The reason \( k_1 \) must be different from zero is that the sum of a temporal and a spatial covariance function is not always positive definite [Myers and Journel, 1990, Rouhani and Myers, 1990], and thus not always a valid covariance function.

We have chosen to work with the product-sum covariance function the way it was originally proposed by De Cesare et al. [2001a,b], because it has nice theoretical properties and is easy to implement on data. Later, however, the model, was generalized somewhat. De Iaco et al. [2001] showed that the assumptions made in equations 5.10 and 5.11 are not necessary for finding the covariance function, but they are nevertheless convenient to use in practice. The product-sum model is constructed for the case where all the variograms in play are limited and thus have covariances; Myers [2002] suggested a more generalized version of the product-sum model which included cases where either the spatial variogram or the temporal variogram or both are unlimited. By examining the marginal variograms \( \gamma_{st}(h_s, \infty) \) and \( \gamma_{st}(\infty, h_t) \) it is possible to find out whether it is necessary to include an unlimited component in the spatiotemporal variogram. From the product-sum covariance function De Iaco et al. [2002] derived a series of non-separable spatiotemporal covariance functions: the integrated product-sum covariance functions.

5.4 Summary

Although field models can be very extensive, they are only approximations of the field. There will always be a part of the field that a given field model does not account for. This unmodelled field may express itself as correlation of the field model residuals. The unmodelled field, along with the instrument error, constitute the errors of the data used to develop the field model. In field modelling, error correlation is only included to a limited extent; it is now common to include the correlation of the three components of each individual vector data point due to attitude uncertainty in the data covariance matrix, but additional error correlation is not included although obviously present. This is in part due to the fact that no standard procedure to determine the data covariance matrix prior to the estimation of the field model parameters exist at this point. Studies have, however, indicated that the inclusion of non-diagonal elements in the covariance matrix does not significantly effect the estimates of the Gauss coefficients at least for the low degree terms [Langel et al., 1989], but the estimate of the variances of the Gauss coefficients are. Thus, as the data covariance matrix in most cases is not used, realistic uncertainty estimates are not available for most field models. One method that is used to make more realistic uncertainty estimates of field models is to estimate the data covariance matrix posterior to the field modelling procedure, and then use this to find an estimate of the model covariance matrix. In order to estimate the data covariance matrix the autocovariance of the field model residuals must be determined; one way to find the autocovariance of a stationary variable with a limited variability is to use variograms.
Chapter 6

Covariance of field model residuals

The geomagnetic field varies on a variety of time-scales and length-scales, which are only rudimentarily considered in most present field models. Analysing the part of the observed field that is not explained by a given field model, the field model residuals, can provide insight into the deficiencies of the field model. The residuals may give an impression of the level of uncertainty of the field model and clues on how to improve the field model; the residuals can e.g. reveal the behaviour of the unmodelled field leading to recognition of the most important sources of the unmodelled field.

In this chapter I present the results we have obtained for the data covariance of the Ørsted(06s/05) field model when error correlation is taken into account. We estimate the data covariance directly from the field model residuals using variograms, and show that the data covariance matrix contains off-diagonal elements of the same order of magnitude as the diagonal elements.

When the data covariance matrix is known, it is possible to estimate the uncertainty of a field model and also the uncertainty on quantities deduced from the field model, e.g. the geothermal heat flux we derive in chapter 8. Furthermore, knowledge of how the errors correlate can be used to improve the field model and make an evaluation of the optimum data (re)sampling rate.

6.1 Residuals

The field model residuals consist of instrument error and unmodelled signal, and are found by subtracting values predicted by the field model from the data. The residuals are often considered an estimate of the data error, and are typically assumed to be uncorrelated and have a Gaussian distribution; these assumptions are, however, not justified. Unmodelled fields do, to a large extent cause serial correlation of the errors, and it has been shown [e.g. Walker and Jackson, 2000] that field model residuals in general do not have a Gaussian distribution. Ideally the instrument errors should be uncorrelated, but this is not always the case. Serial correlation of the errors introduces non-zero off-diagonal elements in the data covariance matrix, which are not accounted for if the errors are assumed uncorrelated. This is why the assumption of uncorrelated errors leads to erroneous estimates of the uncertainty of the field
model coefficients.

The Ørsted(06s/05) field model

We have chosen to work with the Ørsted(06s/05) field model developed by Olsen et al. (this model is a predecessor of the CHAOS field model [Olsen et al., 2006]). The Ørsted(06s/05) field model is based on data from Ørsted, CHAMP, and SAC-C from March 1999 to June 2005. Ørsted and CHAMP vector data are used equatorward of 60° geomagnetic latitude; scalar data are used poleward of this, and where vector (attitude) data were not available. For SAC-C only scalar data are used. Only data from quiet times are used, with the criteria that $K_p \leq 1^+$ at the time of observation and $K_p \leq 2^0$ in the previous three-hour interval. To minimize contributions from ionospheric and induced currents, only data from dark regions, that is with the Sun at least 5° below the horizon, are used. The satellite data have been resampled to 1 minute intervals and are weighted with $w \propto \sin \theta$ to simulate equal-area distribution.

The model describes the internal static field (core and crustal field) up to degree and order 50, the secular variation up to degree and order 18, and the secular acceleration up to degree and order 12. The model takes low-degree external fields, in particular the magnetospheric ring current, and induced fields into account. The ring current and its induced counterpart are modelled as a linear function of $D_{st}$, as is common in field modelling, but an additional ring current correction is made. To account for baseline instability of the $D_{st}$ index [Olsen et al., 2005b], the static part of the external field is determined on a daily basis. The model parameters are estimated from the data using an iterative reweighted least squares method with Huber weights to account for the non-Gaussian distribution of the residuals.

We work with the residuals of the total intensity, and estimate the correlation of the scalar data errors.

Data domain

A satellite data point has four space-time coordinates: altitude, (co)latitude, longitude and time. Although in principle it is possible to include all coordinates, we have chosen to simplify the problem. One simplification is that we in finding the covariance treat only one satellite at the time; we can therefore assume constant altitude. Although the orbit of Ørsted is slightly elliptic, this is a reasonable assumption. However, by treating the satellites individually we implicitly assume that the errors of the data from the different satellites are uncorrelated. This assumption is less likely to be good, as errors due to unmodelled fields probably are correlated between the satellites. Secondly we assume that there is no significant variation along geomagnetic longitude based on the general axi-symmetric shape of the main field. This leaves us with only two dimensions: geomagnetic colatitude and time, where we use Universal Time (UT) for the time coordinate. Or rather, the variation is primarily in these two directions: geomagnetic colatitude and time. The regularity of the satellite orbits means that it is difficult to separate temporal and spatial variations. A satellite in a circular orbit covers a certain (geographic) spherical angle within a certain time span; Ørsted with its period of 100 minutes, covers a spherical angle of 3.6° within one minute. With the orbit being nearly polar, the spherical angle is, in
Figure 6.1: The coordinates (geomagnetic colatitude vs. UT) of one day of Ørsted residuals (March 11th, 2001). The along-track variation is dominated by the variation along colatitude and short-scale temporal variation of the order of minutes. The dashed horizontal lines indicate ±50° geomagnetic latitude.

non-polar regions, dominated by the change in geographic latitude, as the change in longitude is small compared to the change in latitude. When the satellite observes an along-track variation of the field within 5 minutes, it can be viewed as a temporal variation over a range of 5 minutes, or a spatial variation over a range of 18°. In other words, the coordinates are entwined, which has to be kept in mind when interpreting the results.

Different viewpoints of the coordinates can be taken. We will mainly refer to them as geomagnetic colatitude vs. UT, and thus split the variation into spatial variation along the geomagnetic colatitude axis, and temporal variation along the UT axis. From a time perspective, the spatial variation is the short time-scale variation of the order of minutes (less than one orbit) and the temporal variation is the long term variation of the order of hours (more than one orbit). Taking a spatial perspective, the spatial variation is the geomagnetic latitudinal variation, while the temporal variation reflects the geomagnetic longitudinal variation. The most useful and coordinate independent interpretation is that the spatial variation is the along-track variation, whereas the temporal variation is the across-track variation. The along-track variation is dominated by the variation along colatitude and short-scale temporal variation (order of minutes).

In figure 6.1 the geomagnetic colatitude vs. UT distribution of one day of Ørsted residuals is shown. The residuals occur in segments as only dark-side data are used to derive the field model.
CHAPTER 6. COVARIANCE OF FIELD MODEL RESIDUALS

Figure 6.2: The Ørsted residuals plotted as a function of geomagnetic colatitude. The residuals have higher values in the polar regions as expected due to the more disturbed environment here.

Residual values

Figure 6.2 shows the Ørsted residuals as a function of geomagnetic colatitude (in the remainder of this chapter the term colatitude refers to geomagnetic colatitude unless otherwise stated). The residuals poleward of about 55° latitude clearly have much higher values and a much higher variability than the residuals equatorward of 55°, i.e. between 40° and 140° geomagnetic colatitude. This reflects the much more disturbed field in the polar regions and illustrates the difficulty of modelling the strong time-varying external field in the polar regions with field models. As mentioned in section 5.2.2, regional differences in values and variability of a variable can corrupt its variogram [Chilés and Delfiner, 1999]. We have chosen to focus on the variation of the non-polar field and determine the covariance of the residuals in the colatitude interval of 40° to 140°.

In a search for systematic trends in the residuals we have looked at the residual segments individually. The residuals as a function of colatitude from one day are shown in figure 6.3. Systematic long-wavelength features are clearly present, showing that the residuals are correlated. As stationarity is necessary for the variogram analysis, it is necessary to correct for the long-wavelength trend. The residuals are therefore split into two parts: a long-wavelength feature describing the trend, and the remaining signal, which should be stationary. The two parts of the residuals have different interpretations and properties, and will therefore be treated separately. The trend component is obviously non-stochastic, but can be modelled deterministically and has a physical explanation. The remaining part of the residuals cannot be corrected for or approximated with a deterministic model, and resembles true noise in
6.1. RESIDUALS

Figure 6.3: Ørsted residuals, segments by segment for one day (September 12th, 2004), the individual segments are displaced 5 nT for legibility. Systematic long-wavelength features are evident.

the sense that it cannot be predicted. In order to describe the properties and find the covariance of this part of the residuals, we take a statistical approach. To evaluate the covariance of the trend component, a deterministic approach is taken. From now on, the two parts of the residuals will be labelled as the trend and the corrected residuals, the latter being the residuals corrected for the trend.

6.1.1 Removing trends

We believe the trend is due to an unmodelled contribution from the magnetospheric ring current, and derive a deterministic model for this. The idea that the trend is due to the magnetospheric ring current is inspired by the minimum scatter of the residuals near $\pm 35^\circ$ geomagnetic latitude ($55^\circ$ and $125^\circ$ geomagnetic colatitude) and the maximum scatter at the geomagnetic equator, when looking only at the residuals in the interval from $40^\circ$ to $140^\circ$ geomagnetic colatitude, indicated by the dashed lines in figure 6.2. From the expression we will derive shortly (equation 6.5), it can be seen that a ring current flowing at the geomagnetic equator produces a magnetic field, which is perpendicular to the main field at $\pm 35^\circ$ geomagnetic latitude, and thus does not contribute to the scalar residuals at these geomagnetic latitudes. The magnetospheric ring current is accounted for in the Ørsted(06s/05) field model as mentioned previously, but apparently the applied ring current corrections are insufficient.

To first approximation the ring current can be considered a westward flowing current loop. Viewed from afar it has a dipolar field, but inside of the current loop the field is approximately a uniform axial field aligned along the normal of the plane
of the current. Due to the high altitude of the ring current compared to the satellites, we approximate the ring current field at the satellite orbits by an axial uniform field aligned along the dipole axis:

\[
\bar{B}_{rc} = -B_c \cos \theta_d \cdot \hat{r}_d + B_c \sin \theta_d \cdot \hat{\theta}_d,
\]  

(6.1)

where \( B_c \) is the magnitude of the ring current field and \( \theta_d \) is geomagnetic colatitude. The Earth’s dipole field expressed in geomagnetic coordinates is

\[
\bar{B}_{dipole} = B_p \left( \frac{a}{r} \right)^3 \cos \theta_d \cdot \hat{r}_d + \frac{1}{2} B_p \left( \frac{a}{r} \right)^3 \sin \theta_d \cdot \hat{\theta}_d,
\]  

(6.2)

where \( B_p \) is the magnitude of the dipole field at the dipole poles. The unit vector of the dipole field is

\[
\hat{B}_{dipole} = \frac{2 \cos \theta_d \cdot \hat{r}_d + \sin \theta_d \cdot \hat{\theta}_d}{\sqrt{1 + 3 \cos^2 \theta_d}}.
\]  

(6.3)

To find the contribution from the ring current field to the scalar residuals as a function of colatitude, we project the ring current field onto the dipole field:

\[
\bar{B}_{rc} \cdot \hat{B}_{dipole} = B_c \frac{1 - 3 \cos^2 \theta_d}{\sqrt{1 + 3 \cos^2 \theta_d}}.
\]  

(6.4)

Thus we attempt to fit the residuals with the function

\[
 f (\theta) = A \frac{1 - 3 \cos^2 \theta_d}{\sqrt{1 + 3 \cos^2 \theta_d}}
\]  

(6.5)

where \( A \) is a constant that is determined for each satellite segment by a least squares fit. The correction of the residuals for the ring current field by equation 6.5 is only done on segments longer than 25°, to ensure that there is a reasonable number of data points to fit the function to.

With this approximation of the trend it is assumed that it does not vary temporally within the duration of one segment (\(~ 30\) minutes), as equation 6.5 only varies spatially. But the amplitude, \( A \), is determined for each segment allowing temporal variation between the segments, which are about 90–100 minutes apart.

Figure 6.4 shows the fit of equation 6.5 to the residuals of a segment. The good fit supports the assumption that much of the trend is due to the magnetospheric ring current. Inherent in the assumptions made to develop the expression of equation 6.5, is the assumption that the ring current is symmetric around the geomagnetic equator, with highest amplitude here. But, as seen clearly in figure 6.3, the trend does not always peak at \( \theta_d = 90^\circ \). For many segments this leaves an asymmetrical signal in the corrected residuals as is evident in figure 6.4.

There are several possible causes for the displacement of the symmetry axis from the geomagnetic equator. It can e.g. occur if the ring current is displaced somewhat from the geomagnetic equator, if the amplitude of the ring current differs on either side of the geomagnetic equator, or if the current varies quickly with time when the satellite passes from one side of equator to the other. In geomagnetism the ring current is generally assumed to be symmetric around the geomagnetic equator [e.g.
Figure 6.4: The figure on the left shows a nice example of the ring current fit to a given satellite segment (Orsted, September 11th, 2004). The blue dots show the residuals, the green line the ring current fit, and the red dots the corrected residuals. The figure on the right shows an example where the peak of the residuals is displaced from the geomagnetic equator, which causes an asymmetric signal in the corrected residuals (Orsted, February 25th, 2004).

Jorgensen et al., 2004, Olsen et al., 2005b], although asymmetry has been suggested by Malin and Mete Isikara [1976], but what they suggested was, that the mean latitude of the ring current changes with season, and thus has an annual variation. We, however, observe variations within a day.

By subtracting the fitted trend from the residuals, we obtain the corrected residuals, which contain the high-frequency fluctuations around the trend in both space and time. The corrected residuals are assumed stationary, allowing us to calculate their spatiotemporal variograms to find their covariance. The covariance of the trend component is treated separately and is calculated from the deterministic expression of the trend.

6.2 Covariance of the trend

We take a numerical approach to determine the covariance of the trend using the expression in equation 6.5. We synthesize a set of satellite data for a satellite in a polar orbit with an orbital period of 90 minutes and a 1 minute sampling rate; a total of 1000 days, i.e. 16,000 orbits, are used. The field that the satellite observes is given by equation 6.5, with the amplitude changing value from orbit to orbit. The amplitude assigned to a given orbit is found as a random number from a normal distribution with zero mean, and a standard deviation equal to the observed standard deviation of the amplitudes of the segment-by-segment fits. The used standard deviations are from 2004 for Orsted and CHAMP and 2001 for SAC-C. A different year was used for SAC-C, because there are no residuals available in a wide belt (∼25–30°) around the geomagnetic equator for the 2003 and 2004 residuals from SAC-C. The lack of data in the equatorial region made the residuals from these years unsuited for fitting the ring current correction.

The covariance of the trend for each of the three satellites is shown in figure
6.5. As expected from the synthetic data the trend is correlated out to \(360^\circ\) and not beyond that, and it has a variance equal to the variance of the ring current fits. The covariance displays a clear cosine-shape modulated with a linearly decreasing amplitude. The standard deviation of the trend of Ørsted, CHAMP and SAC-C is 2.11 nT, 2.36 nT, and 2.63 nT, respectively (all are listed in table 6.2). We believe the higher standard deviation of SAC-C is due to the lower data quality.

### 6.3 Covariance of the corrected residuals

The second contribution to the total covariance comes from the corrected residuals. The covariance of the corrected residuals is determined by variogram analysis, again we use the residuals from 2004 for Ørsted and CHAMP and from 2001 for SAC-C. The experimental spatiotemporal variograms of the corrected residuals of Ørsted, CHAMP, and SAC-C are shown in the left column of figure 6.6. The spatiotemporal variograms display a clear structure of ridges and troughs parallel to the spatial axis. The periodicity with time lag is even more evident in the temporal marginal variogram shown in figure 6.8. A nugget effect is present in the variograms, which is evident from the spatial marginal variograms shown in figure 6.9. These findings will be discussed and interpreted below.

In theory the marginal variograms describe the variation of one parameter, while the other is kept fixed. This is, however, not possible to do in practice with the available satellite data. Since a satellite continuously moves, it is never at exactly the same place twice, fulfilling the necessary condition for a true temporal measure-
Figure 6.6: The experimental spatiotemporal variograms of the corrected residuals of the Ørsted data from 2004 (top), CHAMP data from 2004 (middle), and SAC-C data from 2001 (bottom) are in the left column. In the right column are the fitted theoretical spatiotemporal variograms.
ment, and it never measures at several locations simultaneously making true spatial measurements.

For the experimental variograms we bin the dissimilarities and take the marginal bins to represent the marginal variograms. The bin size has to be chosen with care. If the bins are too small, spurious features arise in the variograms due to the temporal periodicity of the data caused by the use of only nightside data. Figure 6.7 shows the number of data pairs of the Ørsted 2004 residuals with a given time differences; obviously some time differences do not occur. Thus, if the temporal bin width is chosen too small, empty bins will result. On the other hand, to get the best possible resolution, the bin size should not be too large either. The temporal marginal variogram is made of data pairs, which have almost the same latitude. It takes the satellite one period to reach the same latitude as only dark-side data is used. To get reasonable, well behaved statistics the temporal bin width has to be at least of the order of one period. Thus we use a temporal bin width of 100 minutes. For the spatial bin width we use 2°.

As already mentioned in section 6.1 the coordinates cannot be separated, and the variogram does not display true spatial vs. true temporal behaviour. The spatial marginal variogram primarily reflects the variation with latitude, but also includes short time-scale variations of order of minutes the combination of which is the along-track variation.

6.3.1 Temporal variation

The experimental temporal marginal variograms shown in figure 6.8 clearly display a 24-hour periodic behaviour for each satellite, which reveals correlation between measurements taken 24 hours apart. We believe this is due to an Earth-fixed feature
6.3. **COVARIANCE OF THE CORRECTED RESIDUALS**

Figure 6.8: The temporal marginal variograms of the spatiotemporal variograms shown in figure 6.6. The thin lines are the experimental variograms, the thick lines the fitted theoretical variograms of the Ørsted, CHAMP, and SAC-C residuals, respectively. We do not attempt to model the periodic behaviour of the experimental variograms.

that is unaccounted for in the model e.g. an unmodelled contribution of the crustal or core field, or a modulation of the magnetospheric ring current with geographic coordinates. The periodic correlation arises when the satellites are in repeated track orbits and fly over the same geographic location every 24 hours. Even for near repeated track orbits the problem arises as the unmodelled feature(s) has a finite extent. An unmodelled Earth-fixed feature means it is likely that there will be correlation between the residuals of the individual satellites.

The periodicity of the variogram is equally evident for all three satellites, with perhaps slightly higher amplitude for CHAMP compared to the other two. If the unmodelled field causing the 24-hour period is of internal (core or crustal) origin, the amplitude of the periodic behaviour of the CHAMP variogram should be larger than that of the Ørsted and SAC-C variograms, since CHAMP would be significantly closer to the source. If on the other hand the unmodelled field is due to an external, magnetospheric source that has a uniform field at the satellites orbits, there should be no significant difference in the amplitudes of the variograms. We believe that the observed difference in amplitude probably is not significant i.e. much smaller than it would be if the source is internal, which indicates that the unmodelled field causing the periodic behaviour probably is of external origin.
Fitting theoretical variograms

We have not attempted to model the periodic behaviour when fitting a theoretical variogram to the experimental, because our main interest is to estimate the correlation of the residuals within an orbit and possibly two successive orbits, that is, correlation on a time-scale of a few hours. The 24-hour periodic feature is on a much longer time-scale and corresponds to correlation between orbits 14–16 orbits apart. For the theoretical variogram, we have chosen an exponential function,

\[ \gamma_t(h_t) = \text{sill}_t \cdot \left(1 - \exp\left(-\frac{h_t}{c}\right)\right) \]  

(6.6)

which corresponds to a valid covariance function,

\[ C_t(h_t) = \text{sill}_t \cdot \exp\left(-\frac{h_t}{c}\right) \]  

(6.7)

when \( c > 0 \). For the exponential variogram the sill is reached asymptotically as \( h \to \infty \). Its practical range is about \( 3c \), when the value of the function reaches 95% of the sill value [Chilès and Delfiner, 1999].

The exponential function was chosen because of its simplicity. We did not expect significant correlation on a time-scale above one orbital period, and looking beyond the periodic behaviour, the variograms appear to take constant values. The best fits, shown by the thick lines in 6.8, were obtained through optimizing the values of the sill and width, \( c \), by ordinary least squares. The optimized values of sill and \( c \) are listed in table 6.1.

We expect the range of the temporal variograms to be small and no more than one orbital period, corresponding to one temporal bin width. We do, however, observe a finite width (\( c \sim 2.2 \) hours) of the temporal covariance functions. However, we believe this is due to the exponential function being forced to fit the curvature of the periodic variation for small temporal lags during least squares fitting. Thus, it is probably an artificial feature arising due to the periodic behaviour of the experimental variogram.

6.3.2 Spatial variation

The spatial marginal variograms of the corrected residuals are shown in figure 6.9. The variograms clearly approach non-zero values for zero spatial lag. However, the residuals are expected to have an error component which is uncorrelated. At least some of the instrument error will display this behaviour, and it is also likely that there is part of the unmodelled field that has a correlation length substantially shorter than the data sampling (1 minute) that will contribute to the nugget effect.

The experimental spatial marginal variograms only extend to a spatial lag of about 50°, because the data domain in the spatial direction only is 100° wide. The variograms have not completely reached sill at the maximum spatial lag of 50°. In order to investigate the possibility of extracting the behaviour of the experimental variogram beyond 50°, we conducted a series of tests with periodic boundary conditions of the residuals to see if this makes it possible to obtain reliable variogram values for distances beyond 50°. The tests were made in one dimension using synthetic data with a known covariance (the circular covariance model [Chilès and Delfiner, 1999, p.
6.3. COVARIANCE OF THE CORRECTED RESIDUALS

Figure 6.9: The spatial marginal variograms of the spatiotemporal variograms shown in figure 6.6. The thin lines are the experimental variograms, the thick lines the fitted theoretical variograms of the Ørsted, CHAMP, and SAC-C residuals, respectively.

82] was used). We found that the synthetic data were able to reproduce the correct range of the variogram, but the sill value was systematically estimated too low. However, it was possible to derive an empirical relation between the true sill value and the observed sill and range. The tests were very promising, but conducting a thorough study of this method, including various covariance functions, spatiotemporal variability, and deriving a theoretical basis, was beyond the scope of this work. So the idea was set aside for future investigation, and the interpretation of the variograms produced here is based only on spatial lags less than 50°.

Fitting theoretical variograms

To fit the spatial marginal variograms we have chosen a Gaussian covariance function, because it has a shape similar to that of the experimental variograms and thus gives good fits. The Gaussian variogram is given by

\[ \gamma_s(h_s) = \text{sill}_s \cdot \left(1 - \exp\left(-\frac{h_s^2}{b^2}\right)\right), \]

which corresponds to the valid covariance function

\[ C_s(h_s) = \text{sill}_s \cdot \exp\left(-\frac{h_s^2}{b^2}\right). \]

The Gaussian variogram function has a practical range of about 1.73b. Again, the fit is optimized for sill and width, b, by ordinary least squares: the values are listed in table 6.1, and the best fits are shown in figure 6.9. The experimental variograms have
CHAPTER 6. COVARIANCE OF FIELD MODEL RESIDUALS

Figure 6.10: The along track covariance of the corrected residuals.

not completely flattened at the maximum observed spatial lag, which obviously makes sill determination difficult. Ørsted and CHAMP have almost the same sill value of about 2.8 nT², SAC-C has a slightly higher value of 3.3 nT², and the ranges of the variograms are about 55°. The offset values are 0.55 nT², 1.1 nT², and 0.85 nT² for Ørsted, CHAMP, and SAC-C, respectively. The corresponding covariance functions are shown in figure 6.10. The offsets in the variograms correspond to spikes in the covariance functions at the origin.

It is relevant to compare the offset values with the accuracy of the measurements, as the uncorrelated part of the measurement errors should be less than the offset values of the variograms as mentioned in section 5.2.3. For Ørsted and CHAMP the accuracies of the total field measurements are about 0.5 nT corresponding to a variance of 0.25 nT², which is less than the offset values we find. The offset value of CHAMP is higher than that of Ørsted, but it seems reasonable that there could be a larger amount of uncorrelated signal in the CHAMP residuals than in the Ørsted residuals due to unmodelled short range variability of the field. CHAMP is in closer proximity to the sources in the crust and ionosphere, and to local plasma instabilities occurring in the post-sunset equatorial ionosphere that CHAMP flies through [Stolle et al., 2005]. The SAC-C data are less accurate than the CHAMP and Ørsted data (0.5–1.5 nT after calibration). The offset of 0.85 nT² we find for the SAC-C residuals corresponds to an uncorrelated error of about 0.9 nT.

6.3.3 The spatiotemporal covariance function

The theoretical variograms fitted to the experimental spatiotemporal variograms are found by the product-sum model, which was described in section 5.3. The spatiotem-
poral theoretical variograms, as functions of the marginal variograms, are given by

\[
\gamma_{st}(h_s, h_t) = [k_2 + k_1 C_t(0)] \gamma_s(h_s) + [k_3 + k_1 C_s(0)] \gamma_t(h_t) - k_1 \gamma_s(h_s) \gamma_t(h_t).
\]

(6.10)

where \(C_t(0) = sill_t\) and \(C_s(0) = sill_s\). The best fits to the experimental variograms are shown in the right column of figure 6.6. Besides the parameters obtained from the marginal variograms, the global sill value is necessary. It has been read directly from the experimental spatiotemporal variograms. All the necessary parameters are listed in table 6.1.

From the spatiotemporal covariance functions (not shown) corresponding to the spatiotemporal variograms it is possible to find the covariance between two arbitrary data points from their spatial and temporal distance. This is, however, only the covariance from the corrected residuals, not including the trend.

### 6.4 The total along-track covariance

Having determined the covariance of the trend and the covariance of the corrected residuals, it is possible to find the total along-track covariance of the residuals.

The total along-track covariance is found by adding the covariance of the trend and the spatial marginal covariance of the corrected residuals, which are the along-track covariances. By simply adding the covariances we implicitly assume that the two contributions are uncorrelated, such that no cross-correlation term exists. However, the two parts will only be uncorrelated if the trend function is perfect, otherwise the correction for the trend will introduce correlation between the trend and the corrected residuals. This is likely to occur, and it is reasonable to expect correlation between the corrected residuals and the trend. We do observe that cross-correlation is present, as we find that the variance of the total along-track covariance is larger than the variance of the residuals in the 40° to 140° interval (standard deviations are listed in table 6.2). In fact, we observe the variance of the trend alone is larger than the variance of the residuals, for all three satellites. This indicates that the trend correction we make is far from perfect. One of the major problems is that the peaks of the residual segments are displaced from the geomagnetic equator, which leaves long-wavelength features in the corrected residuals. It is, however, unclear to us how to estimate what effect the cross-correlation will have on the covariance function, and we therefore disregard it and find the total along-track covariance by simply adding the covariance of the trend and the corrected residuals.

<table>
<thead>
<tr>
<th>Satellite</th>
<th>(sill_s) (nT²)</th>
<th>offset (°)</th>
<th>(b) (nT²)</th>
<th>(c) (h)</th>
<th>(sill_t) (nT²)</th>
<th>(k_1) (nT⁻²)</th>
<th>(k_2)</th>
<th>(k_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ørsted</td>
<td>2.78</td>
<td>0.55</td>
<td>32.1</td>
<td>1.96</td>
<td>2.14</td>
<td>2.8</td>
<td>0.3560</td>
<td>0.3022</td>
</tr>
<tr>
<td>CHAMP</td>
<td>2.8</td>
<td>1.1</td>
<td>30.3</td>
<td>2.22</td>
<td>2.29</td>
<td>2.8</td>
<td>0.3571</td>
<td>0.2071</td>
</tr>
<tr>
<td>SAC-C</td>
<td>3.3</td>
<td>0.85</td>
<td>34.1</td>
<td>1.96</td>
<td>2.39</td>
<td>3.3</td>
<td>0.3030</td>
<td>0.4061</td>
</tr>
</tbody>
</table>

Table 6.1: \(b\), \(c\), and sill values, as well as the \(k_i\) values for the best fits of theoretical variograms.
The total along-track covariances for the three satellites, along with the along-track covariances of the corrected residuals and the trend, are shown in figure 6.11. The total along-track covariance is dominated by the trend. For Ørsted the covariance due to the trend has an amplitude about twice that of the covariance of the corrected residuals out to a spatial lag of about 40°. For CHAMP and SAC-C the amplitude of covariance of the trend is about three times that of the covariance of the corrected residuals. For spatial lags larger than about 50° the covariance of the corrected residuals is practically zero, and only the covariance of the trend contributes to the total covariance. The total covariance of each satellite has a spike at zero lag, originating from the covariance of the corrected residuals.

For better comparison of the total along-track covariance functions of the three satellites, they are plotted together in figure 6.12. SAC-C has the largest amplitude; the amplitude of CHAMP is slightly higher than that of Ørsted. The total standard deviations are 2.7 nT, 2.9 nT, and 3.2 nT for Ørsted, CHAMP, and SAC-C, respectively. These, as well as the standard deviations of the trend, the corrected residuals and the offsets are listed in table 6.2. We believe that the larger values of the covariance of CHAMP compared to Ørsted is due to CHAMP being in closer proximity to the internal and ionospheric sources. The larger values of the covariance of SAC-C is due to the lower data quality.

Serial correlation of both the corrected residuals and the trend contributes to the width of the total along-track covariance function. As seen in figure 6.11, the correlation length of the corrected residuals is about 40–50°, beyond 50° only the trend, which has a correlation length of 360°, contributes to the total covariance. However, we believe the correlation length of the trend is overestimated; the correlation length of 360° is an artificial property given to the synthetic data from which the covariance of the trend was estimated. If the correlation length is overestimated, then the true amplitude of the covariance decreases faster than the predicted. We do, however, believe that the predicted periodicity of the trend covariance is likely.

### 6.4.1 The data covariance matrices

Although we do observe a 24-hour periodic feature in the temporal marginal variograms, which shows that some (unwanted) across-track correlation does exist, we believe that across-track correlation apart from this is insignificant. Thus, we use the total along-track correlation to construct the data covariance matrices. The elements

<table>
<thead>
<tr>
<th>Satellite</th>
<th>(\sigma_{\text{trend}}) (nT)</th>
<th>(\sigma_{\text{corr. res.}}) (nT)</th>
<th>(\sigma_{\text{offset}}) (nT)</th>
<th>Total (\sigma) (nT)</th>
<th>(\sigma_{\text{non-polar res.}}) (nT)</th>
<th>(\sigma_{\text{residuals}}) (nT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ørsted</td>
<td>2.11</td>
<td>1.67</td>
<td>0.74</td>
<td>2.7</td>
<td>2.1</td>
<td>3.3</td>
</tr>
<tr>
<td>CHAMP</td>
<td>2.36</td>
<td>1.67</td>
<td>1.0</td>
<td>2.9</td>
<td>2.2</td>
<td>6.2</td>
</tr>
<tr>
<td>SAC-C</td>
<td>2.63</td>
<td>1.82</td>
<td>0.92</td>
<td>3.2</td>
<td>2.4</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Table 6.2: The standard deviations of the trend, the corrected residuals with the offset value included, the offset values separately, and the total covariance. The standard deviations of the residuals and the standard deviations of the non-polar residuals, which were used for the estimation of the covariance, are also listed. Values are for 2004 for Ørsted and CHAMP, and 2001 for SAC-C.
6.4. THE TOTAL ALONG-TRACK COVARIANCE

Figure 6.11: The total along-track covariances, the along-track covariances of the corrected residuals and the trend for Ørsted (top), CHAMP (middle), and SAC-C (bottom).
of the data covariance matrices are found from the total along-track covariance functions, by first calculating the along-track distance between the data pairs, and then assign the covariance as given by the total along-track covariance function for the given distance to the relevant element of the covariance matrix. Again we treat one satellite at the time, and find the data covariance matrix for each satellite separately. The total data covariance matrix of all the data can be obtained by simply combining the individual covariance matrices as

\[
C_{\text{all}} = \begin{pmatrix}
  C_\Theta & 0 & 0 \\
  0 & C_c & 0 \\
  0 & 0 & C_s
\end{pmatrix}
\]

where the zeros stand for matrices of zeros, and \(C_\Theta, C_c,\) and \(C_s\) are the data covariance matrices of the Ørsted, CHAMP, and SAC-C residuals for all the years, respectively. Recall that the residuals of the different satellites were assumed to be uncorrelated.

To give an impression of the structure of the data covariance matrices, small parts of each of them are shown in figures 6.13, 6.14, and 6.15 for Ørsted, CHAMP, and SAC-C, respectively. In each figure, the contributions to the covariance from the corrected residuals without the offset, from the offset itself and from the trend are shown. Again, it is clearly seen that the covariance of the corrected residuals has a shorter correlation length than the covariance of the trend in that the covariance matrix from the corrected residuals has fewer non-zero elements. The figures clearly demonstrate that the data covariance matrices have several non-zero off-diagonal elements, but are very sparse nevertheless. The number of non-zero elements and the densities of the covariance matrices are listed in table 6.3. Even though the densities of the covariance matrices are very low, of the order \(10^{-4}\), the large number of data
Figure 6.13: Part of the covariance matrix of the Ørsted residuals. Panel A shows
the contribution to the covariance matrix from the offset (ten times the value is shown); it only has elements on the diagonal. Panel B shows the contribution from the corrected residuals without the offset. The elements lie in a narrow band along the diagonal. Panel C shows the contribution from the trend. Again the elements lie in a band along the diagonal, which, however, is considerably wider than that from the corrected residuals. Panel D shows the total covariance.
Figure 6.14: As figure 6.13, but for the CHAMP residuals.
Figure 6.15: As figure 6.13, but for the SAC-C residuals.
Here all non-zero elements based on the along-track covariance functions have been included in the covariance matrices. However, for many practical purposes it would be natural to set values smaller than some limit to zero, for example when the (absolute) value of the covariance is below 10% or 20% of the maximum variance. In this way the number of non-zero elements in the covariance matrix is reduced. In figure 6.16 the fractions of retained non-zero elements as a function of the cutoff variance for the three covariance matrices are shown. A fraction of one corresponds to all non-zero elements being included, the scale of cutoff variances extends to but does not include zero, so all the exact zero elements of the covariance matrices are not included. The figure shows that if the cutoff variance is set to about 50% of the variance, then only about 25% of the non-zero elements are retained for all three satellites. For lower values of the cutoff variance, the fraction of retained non-zero elements increases rapidly. If the cutoff variance is set to 10% of the variances for each satellite, about 90% of the non-zero elements are retained.

The covariance of the trend dominates the covariance matrices as it did the covariance functions. As the correlation length of the trend most likely is overestimated, the off-diagonal elements of the covariance matrices are likely to have lower values than those predicted here. Thus the presented covariances are upper estimates of the true values, and so the numbers of non-zero elements listed in table 6.3 are also upper limits.

### 6.5 Discussion

We have taken a new approach to determine the error correlation of field model residuals by estimating the covariance directly from the residuals. As in previous studies of error correlation in field modelling discussed in section 5.1, we find that serial correlation is present and that off-diagonal elements should be included in the data covariance matrix if a valid uncertainty estimate of the field model is to be made. But as the previous attempts we fail in finding an easy, efficient way to estimate the uncertainty with low computational costs. There still is no standard procedure to make reliable uncertainty estimates of field models.

We have found that the data covariance matrix of the Ørsted(06s/05) field model has a density of $10^{-4}$. For this density it would be possible, with some effort and a very large computer to include the data covariance matrix in the estimation of the Gauss coefficients and their variance.

<table>
<thead>
<tr>
<th></th>
<th>Ørsted</th>
<th>CHAMP</th>
<th>SAC-C</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of data</td>
<td>127562</td>
<td>99620</td>
<td>81058</td>
<td>308240</td>
</tr>
<tr>
<td>Non-zero elements</td>
<td>$5.82 \cdot 10^6$</td>
<td>$4.04 \cdot 10^6$</td>
<td>$3.79 \cdot 10^6$</td>
<td>$13.65 \cdot 10^6$</td>
</tr>
<tr>
<td>Density</td>
<td>$3.52 \cdot 10^{-4}$</td>
<td>$4.07 \cdot 10^{-4}$</td>
<td>$5.77 \cdot 10^{-4}$</td>
<td>$1.43 \cdot 10^{-4}$</td>
</tr>
</tbody>
</table>

Table 6.3: The number of data points, the number of non-zero elements in the data covariance matrices and the densities of the covariance matrices for Ørsted, CHAMP, and SAC-C, as well as for the total covariance matrix.
6.5. DISCUSSION

Figure 6.16: The fraction of non-zero elements included in the three covariance matrices depending on the cutoff variance. A fraction of one corresponds to all non-zero elements are retained. If e.g. the cutoff variance of the Ørsted covariance is set to $3.5 \, \text{nT}^2$ about 30% of the non-zero elements are included in the covariance matrix. If the cutoff variance is set to $3 \, \text{nT}^2$, more than 45% of the elements are included.

It appears to us that most of the serial correlation of the field model residuals is due to an insufficient description of the field from the magnetospheric ring current. Thus a better model for the magnetospheric ring current is needed, as also suggested by Holme et al. [2003]. Although Jørgensen et al. [2004] concluded that the ring current strength is linearly related to the $D_{st}$ index, we find the ring current correction based on the $D_{st}$ index done in the field model is not adequate. There are several possible reasons for this.

We believe that the difficulties of making a proper ring current correction based on the $D_{st}$ index in part are caused by poor determination of the $D_{st}$ index. Olsen et al. [2005b] presented evidence that the baseline of the $D_{st}$ index is unstable and has varied significantly over the last 40 years. We did allow for this particular feature in this study, as the baseline level of the $D_{st}$ index was calibrated on a daily basis in the field model we analysed. Nevertheless long-wavelength trends are still evident in the field model residuals. We believe that improving the description of the ring current could, as suggested by Lowes and Olsen [2004], involve finding different or additional proxies to the $D_{st}$ index for modelling the amplitude and structure of the ring current. Possibly the determination of the $D_{st}$ index could be improved if data from more than four observatories were used.

Another explanation of the difficulties of modelling the magnetospheric field with the $D_{st}$ index is that other sources than the magnetospheric ring current could contribute to the $D_{st}$ index. The $D_{st}$ index is based on observatory data, and so it is possible that some of the sources contributing to the field component, from which the $D_{st}$ index is determined, are from the ionosphere below satellite altitude. In this
case the $D_{st}$ index does not only relate to the magnetospheric field, and it is wrong to assume so when developing field models.

Furthermore, the $D_{st}$ index is based on data from all local times, while the data included in the field model are nightside data only. The ring current is not symmetric in local time; during quiet times the peak of the ring current is in the afternoon sector [Jorgensen et al., 2004], during more disturbed times it shifts towards midnight. The asymmetry in local time of the ring current strength may also mean that the use of the $D_{st}$ index to describe the strength of the ring current on the nightside is not ideal.

We observe that in many of the residual segments the peak of the trend is displaced some $5–10^\circ$ from the geomagnetic equator, indicating a hemispherical asymmetry of the ring current field. Hemispherical asymmetry is allowed for in the external field of the field model, but again this correction appears inadequate as we do observe displacements of the peaks of the residual segments.

With an improved determination of the $D_{st}$ index and with a better understanding of what the $D_{st}$ index really represents, the contribution from the ring current could be better determined and accounted for in field modelling.

If a better correction for ring current had been included the Ørsted(06s/05) field model to begin with, we believe that not only the covariance of the trend would be significantly reduced, but the correlation length of the corrected residuals would also be less. This is the case because much of the long-wavelength signal that is present in the corrected residuals is due to the applied trend correction being too simple. Thus, getting a better hold on the magnetospheric contribution in general would decrease the serial correlation of the residuals, and the ultimate result is better field models.

In many of the studies of error correlation in field modelling that was mentioned in section 5.1, the problem was the poor determination of the crustal field, which influenced core field models. With the newly available satellite data, the core and crustal field are usually co-estimated, or the crustal field is accounted for in some other way, and thus should not contribute significantly to the error correlation. Although we did find a 24-hour periodic correlation of the corrected residuals, which may originate from an unmodelled crustal feature, we believe that the problem of error correlation in general has moved one step further, and no longer is that of the crustal field but now lies in the inadequate determination and modelling of the magnetospheric field. One may, however, question whether a better understanding of the magnetospheric contribution to the field, will solve all the error correlation problems. More likely, it is just the next step, beyond which a new “most important” error source will emerge.

Until a better description of the magnetospheric ring current becomes available various other measures can be taken to reduce serial error correlation. One way to alleviate error correlation in field model data is to use a lower resampling rate. For the Ørsted(06s/05) field model a resampling rate of 1 minute was used; decreasing the resampling rate to 6 minutes would decrease the variance by about 50%. Decreasing the resampling rate could, however, mean loss of information especially about short-scale structures. This method could be combined with using a non-diagonal covariance matrix, if e.g. a resampling rate of 3 minutes is used the correlation would be reduced, but the risk of loss of information would be less than for a lower resampling rate. With the reduction of correlation and number of data, the data covariance matrix would be smaller and have fewer non-zero elements, and could be included in
6.6. CONCLUSIONS

An alternative way is to detrend the data segment by segment, similar to the way the residuals was detrended in this study. This is e.g. used in crustal field modelling, where the data is high-pass filtered to remove long wavelengths. However, there is a great risk of also removing north-south aligned signal of interest in that process, and thus introduce new errors in the field model.

We find that the covariance of the Ørsted data in general is lower than of the CHAMP data which again is lower than the covariance of the SAC-C data. As mentioned earlier, it is likely that the CHAMP data has a higher variance than the Ørsted data due to its proximity to the crustal and ionospheric sources. The reason for the higher variance of the SAC-C data is the lower quality of the data.

To estimate the covariance function we used scalar intensity residuals, thus our covariance function applies to the scalar data. However, the along-track covariance function still gives an estimate of the error correlation between all (scalar and vector) data, it just does not account for the possible additional error correlation of the three components of the vector data. For a formalism on how to handle error correlation of the three components of the individual vector measurements see Holme and Bloxham [1995, 1996], Holme and Jackson [1997], Holme [2000]. Furthermore, our covariance estimate is based on non-polar data and thus applies to non-polar regions. As noted earlier the variability of the polar data is much higher than that of the non-polar data. So it is likely that our model overestimates the correlation length in the polar regions, but underestimates the variance. A way to accommodate for this in the data covariance matrix could be to assign a higher variance to the polar data; this was e.g. done in the derivation the OSVM field model [Olsen, 2002]. Once the data covariance matrix is known it is possible to estimate the uncertainty of the field model parameters.

Crustal field models are now considered to be accurate enough to be used for geologic and geophysical interpretation; in these situations it is desirable to know the uncertainty of the field model such that the uncertainty of quantities deduced from the field model can be estimated. In chapters 7 and 8 the results we have obtained for the magnetic crustal thickness and the geothermal heat flux derived from the MF3 field model are presented. However, as we do not know the uncertainty of MF3, we adopt the over all uncertainty estimate by Lowes and Olsen [2004] of 25% for this.

6.6 Conclusions

In field modelling, the field model residuals are often considered an estimate of the data uncertainty and are assumed to be uncorrelated, such that a diagonal data covariance matrix can be used when estimating the model parameters. We have shown that there is serial correlation of the Ørsted(06s/05) field model residuals and that there are off-diagonal elements of the data covariance matrix of the same order of magnitude as the diagonal elements. As the Ørsted(06s/05) field model is a fairly typical field model, it is reasonable to assume that serial correlation of field model residuals is quite common. To make valid error estimates of field models it is necessary to either account for this correlation by including non-zero off-diagonal elements into the data covariance matrix, parameterize and estimate the underlying
physical phenomenon, i.e. the co-estimation approach, or reduce the effect of serial correlation by e.g. using a low resampling rate (6–10 minutes between the individual data points).

We believe that it in particular is an inadequate description of the magnetospheric ring current that gives rise to the significant along-track correlation of the Ørsted(06s/05) field model residuals. Thus the correlation can be reduced and the field model improved if a better model for the magnetospheric ring current is developed.
Part III

Application of field models – Magnetic crustal thickness and geothermal heat flux
Chapter 7

Magnetic crustal thickness

With the recently obtained high precision magnetic field measurements by satellites it has become possible to derive models of the crustal field sufficiently detailed that they can be used for geophysical interpretation [Maus et al., 2005b, Hemant et al., 2005]. We use the equivalent source magnetic dipole method to determine the magnetic crustal thickness in Greenland and Antarctica from the magnetic field of the crust as given by the MF3 and CM4 field models (see chapter 4). As mentioned in section 2.2.2, the magnetic crust is that part of the crust that can sustain a magnetic field. Its upper boundary is the bedrock surface, thus it does not include sediments or ice sheets. The lower boundary of the magnetic crust is given by either Moho or the depth where the temperature reaches the Curie temperature of the crustal rocks [Wasilewski and Mayhew, 1992, Frost and Shive, 1986], depending on which of the two that is shallower (figure 2.4).

The reason for determining the magnetic crustal thickness in Greenland and Antarctica, is that we later (in chapter 8) wish to estimate the geothermal heat flux underneath the large ice sheets from the derived magnetic crustal thickness. That magnetic anomalies, the depth to the Curie isotherm, also referred to as the Curie depth, and the surface heat flow correlates is well known. Aeromagnetic data have been widely used to determine the depth to the Curie isotherm [e.g. Okubo et al., 1985, Tsokas et al., 1998, Tanaka et al., 1999, Stampolidis and Tsokas, 2002, and many others], and the obtained Curie depths have been seen to correlate with heat flow measurements. Magnetization models of the crust derived from satellite magnetic data have also been seen to correlate with the depth to the Curie isotherm and surface heat flow [e.g. Mayhew, 1982a,b, 1985, Hayling, 1991, Mayhew et al., 1991, Purucker et al., 1998, Hamoudi et al., 1998]; these earlier models were, however, based on POGO and Magsat data. With the present satellites, Ørsted, CHAMP, and SAC-C, data of a much higher quality have become available.

In this chapter it is first explained how the data necessary to determine the magnetic crustal thickness are calculated from the field models. Then the equivalent source magnetic dipole method is introduced, and it is explained how we use it to derive the magnetic crustal thickness from the data. Finally the results we have achieved for the magnetic crustal thickness are presented and discussed, and an estimate of the uncertainty of the obtained results is given.
Figure 7.1: The radial component of the observed crustal field at 300 km altitude as given by the MF3 crustal field model (degree and order 15 to 90).

7.1 The observed induced crustal field

To determine the magnetic crustal thickness we use data synthesized from a field model; this is an approach which previously has been used by e.g. Whaler [2003]. When working with the magnetic field it can be an advantage to use a field model instead of raw data, because the field model more easily allows for separation of the various sources contributions. We have worked primarily with the MF3 crustal field model (MF4 was not derived at the time of this work) for determination of the magnetic crustal thickness, but also with the CM4 field model, where we have applied a high-pass filter to remove the core field. Our preliminary studies were done with CM4, using spherical harmonics of degree 15–65 to represent the crustal field. When we later turned to use MF3, we wished to use the same range of spherical harmonic degrees, and we augmented MF3 with the degree 15 coefficients from the Ørsted(10a/03) field model.

Using a field model as data also has the advantage that the data can be synthesized at any data spacing at any altitude. MF3 provides a good representation of the field down to altitudes of 50 km over mid and low geomagnetic latitudes, and down to 100 km at geomagnetic polar latitudes [Maus et al., 2004], although the data from which the model is developed is obtained at 360–760 km altitude. Downward continuation of the field should, however, be done with care, as it amplifies short-wavelength (high degree terms) more than long-wavelength (low degree terms) features, and there is hence a real risk of amplifying noise more than signal. Thus in
7.1. THE OBSERVED INDUCED CRUSTAL FIELD

Figure 7.2: The radial component of the remanent magnetic field of the oceanic crust at 300 km altitude as given by the model of Dyment and Arkani-Hamed [1998b] and Purucker and Dyment [2000].

some of our initial studies particularly when using CM4, we synthesized the data at 400 km altitude. Later, when we started working with MF3, we used an altitude of 300 km. The radial component of MF3 at 300 km altitude is shown in figure 7.1.

The crustal field as observed by the satellites and thus described by the field models is caused by both remanent and induced magnetization of the crustal rocks. The induced magnetization depends on the thickness of the magnetic crust, as well as on the susceptibility of the crustal rocks and the ambient field strength. Thus to determine the magnetic crustal thickness it is necessary to isolate the induced part of the crustal field. As it cannot be measured directly, we isolate the induced crustal field by subtracting a model for the remanent crustal field from the observed crustal field.

We use the model of remanent magnetization of the oceanic crust by Dyment and Arkani-Hamed [1998b] and Purucker and Dyment [2000], in which the remanent magnetization of the oceanic lithosphere is determined from ocean floor ages, plate motion, and polar wander. Unfortunately no global or regional models of the remanent magnetization of the continental crust exist, primarily because the remanent magnetization of the continental crust, contrary to that of the oceanic crust, does not display a systematic behaviour. It is, however, believed that induced magnetism in general dominate remanent magnetism in the continental crust, so we assume that remanent magnetism is negligible in the continental crust, and simply use the model by Dyment and Arkani-Hamed [1998b] and Purucker and Dyment [2000] to account
CHAPTER 7. MAGNETIC CRUSTAL THICKNESS

Figure 7.3: The radial component of the observed induced crustal field at 300 km altitude (spherical harmonic degrees 15–90). The observed induced crustal field is obtained by subtracting the remanent field (figure 7.2) from the observed crustal field (figure 7.1).

for remanent magnetism globally. The radial component of the remanent field from this model at 300 km altitude is shown in figure 7.2. Most of the magnetic anomalies associated with ocean ridge spreading do not show up at 300 km height, because the magnetic stripes in general are too narrow to be detected at this altitude. However, areas created during the Cretaceous Quiet Interval do appear on the map, as sufficiently large areas of crust with magnetization in the same direction were created during this long period of normal polarization.

In the model by Dyment and Arkani-Hamed [1998b] and Purucker and Dyment [2000] the remanent field is given on a $0.5^\circ \times 0.5^\circ$ map. To be able to apply it to our study we made a spherical harmonic expansion of the remanent field, and high-pass filtered it by setting all coefficients with degrees less than 15 to zero in order to get it on the same form as the observed crustal field. From the spherical harmonic expansion it is possible to synthesize the desired components of the remanent field at the wanted altitude.

By subtracting the remanent field from the observed crustal field, the observed induced crustal field is obtained, from which we determine the magnetic crustal thickness by the equivalent source magnetic dipole method [e.g. Dyment and Arkani-Hamed, 1998a, Langel and Hinze, 1998]. The radial component of the observed induced crustal field at 300 km altitude is shown in figure 7.3.
7.2 Equivalent source magnetic dipole method

In the equivalent source magnetic dipole (ESMD) method, the crustal magnetization is represented by a finite number of dipoles scattered over the area of interest, as illustrated in figure 7.4. The basic idea is to determine the magnetic moments of these crustal dipoles such that their total magnetic field at some given distance above the surface is similar to the observed magnetic field, within the uncertainty of the observed field. From the crustal magnetization it is possible to determine the thickness of the magnetic crust, if the magnetic susceptibility of the crustal rocks is known.

To avoid edge effects we apply the ESMD method globally and let the crustal magnetization of the Earth’s crust be represented by a large number (> 10,000) of dipoles scattered evenly across the surface of the Earth in an icosahedral grid (further details on the dipole locations will be given later). To find the magnetic moments of the crustal dipoles such that they together have a magnetic field equal to the observed field at the observation points, an expression relating the dipole moments of all the dipoles and their combined magnetic field is needed. The basic equation in the ESMD method, is that of the magnetic field of a dipole.

7.2.1 The magnetic field from a dipole

As we are interested in finding the crustal magnetization globally, and as the data (MF3, CM4) are given on a spherical shell surrounding the Earth, it comes naturally to work in spherical coordinates using the geographical coordinate system. Consider a dipole with magnetic moment \( \vec{m}_j \) positioned at \( \vec{r}_j = (r_j, \theta_j, \phi_j) \). We are interested in its (contribution to the) magnetic field at the observation point \( \vec{r}_i = (r_i, \theta_i, \phi_i) \), which can be found from the expression of the magnetic potential of a dipole. A schematic of the geometry of the problem is shown in figure 7.5. Let

\[
\vec{r}_{ij} = \vec{r}_i - \vec{r}_j,
\]  

(7.1)
and the cosine of the angle between $\vec{r}_i$ and $\vec{r}_j$, 

$$
\mu_{ij} = \frac{\vec{r}_i \cdot \vec{r}_j}{r_i r_j},
$$

(7.2)

which in terms of the angles involved is given by

$$
\mu_{ij} = \cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j \cos (\phi_i - \phi_j).
$$

(7.3)

By basic trigonometry

$$
(r_{ij})^2 = (r_i)^2 + (r_j)^2 - 2r_i r_j \mu_{ij}.
$$

(7.4)

In the spherical coordinate system the unit vectors $\hat{r}$, $\hat{\theta}$, and $\hat{\phi}$, change directions with position, $\vec{r}$. The transformations from the Cartesian unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}$ to the spherical unit vectors are

$$
\hat{r} = \sin \theta (\cos \phi \cdot \hat{x} + \sin \phi \cdot \hat{y}) + \cos \theta \cdot \hat{z}
$$

$$
\hat{\theta} = \cos \theta (\cos \phi \cdot \hat{x} + \sin \phi \cdot \hat{y}) - \sin \theta \cdot \hat{z}
$$

$$
\hat{\phi} = - \sin \phi \cdot \hat{x} + \cos \phi \cdot \hat{y},
$$

(7.5)

which clearly show that the unit vectors $\left(\hat{r}_i, \hat{\theta}_i, \hat{\phi}_i\right)$ at position $\vec{r}_i$ are not the same as the unit vectors $\left(\hat{r}_j, \hat{\theta}_j, \hat{\phi}_j\right)$ at position $\vec{r}_j$, unless $\vec{r}_i = \vec{r}_j$. To conduct simple algebra in the spherical coordinate system it is necessary to be able to express the same vector in both $j$-coordinates along the $j$ unit vectors and $i$-coordinates along the $i$ unit vectors. For example, the vector $\vec{r}_{ij}$ in equation 7.1 is

$$
\vec{r}_{ij} = \vec{r}_i - \vec{r}_j = r_i \hat{r}_i - r_j \hat{r}_j.
$$

(7.6)
7.2. EQUIVALENT SOURCE MAGNETIC DIPOLE METHOD

To find $\vec{r}_{ij}$ in either $i$-coordinates or $j$-coordinates, the projections of the each of the unit vectors in one coordinate system on the unit vectors of the other coordinate system are needed. These projections are found by taking the dot-products of all the unit vectors, these are easily calculated from the above equations (7.5) and are all related to spherical derivatives of $\mu_{ij}$. To simplify algebraic expressions we adopt a special notation for the dot-products and use $\delta^{ij}_{r\theta}$ to denote the projection of $\hat{r}_i$ upon $\hat{\theta}_j$. As the dot-product is commutative, $\delta^{ij}_{r\theta} = \delta^{ij}_{\theta r}$. All the dot-products, their notation and connection to the derivatives of $\mu_{ij}$ are listed in table 7.1. With the projections $\vec{r}_{ij}$ can be expressed in both $i$-coordinates

$$
\vec{r}_{ij} = r_i \hat{r}_i - r_j \hat{r}_j
= r_i \hat{r}_i - r_j \left( \delta^{ij}_{r\theta} \hat{r}_i + \delta^{ij}_{\theta r} \hat{\theta}_i + \delta^{ij}_{\phi r} \hat{\phi}_i \right)
= (r_i - r_j \delta^{ij}_{r\theta}) \hat{r}_i - r_j \delta^{ij}_{\theta r} \hat{\theta}_i - r_j \delta^{ij}_{\phi r} \hat{\phi}_i
$$

(7.7)

and in $j$-coordinates

$$
\vec{r}_{ij} = (r_i \delta^{ij}_{r\theta} - r_j) \hat{r}_j + r_i \delta^{ij}_{\theta r} \hat{\theta}_j + r_i \delta^{ij}_{\phi r} \hat{\phi}_j,
$$

(7.8)

which we will need later.

The magnetic potential at $\vec{r}$ due to a dipole with magnetic moment $\vec{m}_j$ positioned at $\vec{r}_j$ is given by

$$
V(\vec{r}_j) = \mu_0 \frac{\overline{m}_j}{4\pi} \cdot \nabla_j \left( \frac{1}{r_{ij}} \right),
$$

(7.9)

where $\mu_0$ is the vacuum permeability ($= 4\pi \cdot 10^{-7}$ N/A²), and $\nabla_j$ refers to that the nabla operator is operating on the $j$-coordinates. At position $\vec{r}_j$ it is

$$
\nabla_j = \hat{r}_j \frac{\partial}{\partial r_j} + \hat{\theta}_j \frac{1}{r_j} \frac{\partial}{\partial \theta_j} + \hat{\phi}_j \frac{1}{r_j \sin \theta_j} \frac{\partial}{\partial \phi_j}.
$$

(7.10)

<table>
<thead>
<tr>
<th>Unit vectors</th>
<th>$\delta$-notation</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{r}_i \cdot \hat{r}_j$</td>
<td>$\delta^{ij}_{rr}$</td>
<td>$\cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j \cos (\phi_i - \phi_j)$</td>
</tr>
<tr>
<td>$\hat{r}_i \cdot \hat{\theta}_j$</td>
<td>$\delta^{ij}_{r\theta}$</td>
<td>$-\cos \theta_i \sin \theta_j + \sin \theta_i \cos \theta_j \cos (\phi_i - \phi_j)$</td>
</tr>
<tr>
<td>$\hat{r}_i \cdot \hat{\phi}_j$</td>
<td>$\delta^{ij}_{r\phi}$</td>
<td>$\sin \theta_i \sin \phi_j$</td>
</tr>
<tr>
<td>$\hat{\theta}_i \cdot \hat{r}_j$</td>
<td>$\delta^{ij}_{\theta r}$</td>
<td>$-\sin \theta_i \cos \theta_j + \cos \theta_i \sin \theta_j \cos (\phi_i - \phi_j)$</td>
</tr>
<tr>
<td>$\hat{\theta}_i \cdot \hat{\theta}_j$</td>
<td>$\delta^{ij}_{\theta \theta}$</td>
<td>$\cos \theta_i \cos \theta_j + \sin \theta_i \sin \theta_j \cos (\phi_i - \phi_j)$</td>
</tr>
<tr>
<td>$\hat{\theta}_i \cdot \hat{\phi}_j$</td>
<td>$\delta^{ij}_{\theta \phi}$</td>
<td>$\cos \theta_i \sin (\phi_i - \phi_j)$</td>
</tr>
<tr>
<td>$\hat{\phi}_i \cdot \hat{r}_j$</td>
<td>$\delta^{ij}_{\phi r}$</td>
<td>$-\sin \theta_i \sin \phi_j$</td>
</tr>
<tr>
<td>$\hat{\phi}_i \cdot \hat{\theta}_j$</td>
<td>$\delta^{ij}_{\phi \theta}$</td>
<td>$-\cos \theta_i \sin (\phi_i - \phi_j)$</td>
</tr>
<tr>
<td>$\hat{\phi}_i \cdot \hat{\phi}_j$</td>
<td>$\delta^{ij}_{\phi \phi}$</td>
<td>$\cos (\phi_i - \phi_j)$</td>
</tr>
</tbody>
</table>

Table 7.1: The dot-products between the unit vectors, their full expression as functions of $\theta$ and $\phi$, and their relationship to $\mu_{ij}$. 
Applying the gradient operator on \((r_{ij})^{-1}\) gives

\[
\nabla_j \left( \frac{1}{r_{ij}} \right) = -\frac{\vec{r}_{ij}}{(r_{ij})^3}.
\]

(7.11)

Thus the magnetic potential due to the dipole can be expressed as

\[
V(\vec{r}_i) = -\frac{\mu_0}{4\pi} \frac{\vec{m}_j \cdot \vec{r}_{ij}}{(r_{ij})^3}.
\]

(7.12)

From this the magnetic field, \(\vec{A}\), due to the dipole is found by taking the negative gradient of the potential,

\[
\vec{A}(\vec{r}_i) = -\nabla V(\vec{r}_i)
= \frac{\mu_0}{4\pi} \frac{m_j}{(r_{ij})^3} [3(\vec{m}_j \cdot \vec{r}_{ij}) \vec{r}_{ij} - \vec{m}_j].
\]

(7.13)

This expression can be simplified to a matrix equation, which can be realized if the individual components of equation 7.13 are written in full length. First note that

\[
\vec{m}_j \cdot \vec{r}_{ij} = \frac{1}{m_j r_{ij}} m_i^j (r_i \delta_{ij}^r - r_j) + m_j^r r_i \delta_{ij}^r + m_j^\theta r_i \delta_{ij}^\theta + m_j^\phi r_i \delta_{ij}^\phi,
\]

(7.14)

where \((m_i^r, m_i^\theta, m_i^\phi)\) are the three components of the magnetic moment, \(\vec{m}_j\), of the dipole situated at \(\vec{r}_j\). The radial component of \(\vec{A}\) is then

\[
A_r(\vec{r}_i) = \frac{\mu_0}{4\pi} \frac{m_j}{(r_{ij})^3} \left[ \frac{3}{m_j (r_{ij})^2} \left( r_i - r_j \delta_{ij}^r \right) \right].
\]

(7.15)

\[
\left( m_i^r (r_i \delta_{ij}^r - r_j) + m_j^\theta r_i \delta_{ij}^\theta + m_j^\phi r_i \delta_{ij}^\phi \right)
- \frac{1}{m_j} \left( m_i^r \delta_{ij}^r + m_j^\theta \delta_{ij}^\theta + m_j^\phi \delta_{ij}^\phi \right)
\]

which can be rearranged to give

\[
A_r(\vec{r}_i) = \frac{\mu_0}{4\pi} \frac{1}{(r_{ij})^3} \left[ \left( \frac{3}{m_j (r_{ij})^2} \right) \left( r_i \delta_{ij}^r - r_j \right) \left( r_i - r_j \delta_{ij}^r \right) - \delta_{ij}^r \right] m_i^j
\]

\[
+ \left( \frac{3r_i \delta_{ij}^\theta}{m_j (r_{ij})^2} \right) \left( r_i - r_j \delta_{ij}^\theta \right) m_i^\theta
\]

\[
+ \left( \frac{3r_i \delta_{ij}^\phi}{m_j (r_{ij})^2} \right) \left( r_i - r_j \delta_{ij}^\phi \right) m_i^\phi.
\]

(7.16)

Similar expressions for \(\vec{A}_\theta\) and \(\vec{A}_\phi\) can be derived (not written), from which it can be seen that equation 7.13 can be written as

\[
\vec{A}(\vec{r}_i) = F_{ij} \vec{m}_j,
\]

(7.17)
where $F^{ij}$ is a $3 \times 3$-matrix relating the magnetic field at $\vec{r}_i$ to the dipole at $\vec{r}_j$ with dipole moment $\vec{m}_j$. The elements of the top row of $F^{ij}$ are given by the three brackets within the squared bracket in equation 7.16, note that they only depend on the geometry of the problem and not on the direction or magnitude of the dipole moment. The rest of the elements of $F^{ij}$ can be found from similar expressions for $A_\theta$ and $A_\phi$. All the elements of $F^{ij}$ are listed in table 7.2.

To calculate the magnetic field from a dipole, equation 7.13 can be used directly without specifying the $F^{ij}$-matrix. But in the inverse case of determining the dipole moment from the magnetic field it would be necessary to specify the $F^{ij}$-matrix explicitly.

The above expression was derived for one dipole; to represent the entire crustal field a large number of dipoles is necessary. Assume that $J$ dipoles are distributed at $J$ locations $\vec{r}_j$, with $j = 0, 1, \ldots, J$, then the total magnetic potential at $\vec{r}_i$ is simply

<table>
<thead>
<tr>
<th>Element</th>
<th>Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{11}^{ij}$</td>
<td>$\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3(r_i - r_j \mu_{ij}) (r_i \mu_{ij} - r_j)}{(r_{ij})^2} - \mu_{ij} \right)$</td>
</tr>
<tr>
<td>$F_{12}^{ij}$</td>
<td>$\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3(r_i - r_j \mu_{ij}) r_i (\hat{\theta}_j \cdot \hat{r}<em>i)}{(r</em>{ij})^2} - (\hat{\theta}_j \cdot \hat{r}_i) \right)$</td>
</tr>
<tr>
<td>$F_{13}^{ij}$</td>
<td>$\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3(r_i - r_j \mu_{ij}) r_i (\hat{\phi}_j \cdot \hat{r}<em>i)}{(r</em>{ij})^2} - (\hat{\phi}_j \cdot \hat{r}_i) \right)$</td>
</tr>
<tr>
<td>$F_{21}^{ij}$</td>
<td>$-\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3r_j (\hat{r}<em>j \cdot \hat{\theta}<em>i) (r_i \mu</em>{ij} - r_j)}{(r</em>{ij})^2} + (\hat{r}_j \cdot \hat{\theta}_i) \right)$</td>
</tr>
<tr>
<td>$F_{22}^{ij}$</td>
<td>$-\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3r_j (\hat{r}_j \cdot \hat{\theta}_i) r_i (\hat{\theta}_j \cdot \hat{r}<em>i)}{(r</em>{ij})^2} + (\hat{\theta}_j \cdot \hat{r}_i) \right)$</td>
</tr>
<tr>
<td>$F_{23}^{ij}$</td>
<td>$-\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3r_j (\hat{r}<em>j \cdot \hat{\phi}<em>i) (r_i \mu</em>{ij} - r_j)}{(r</em>{ij})^2} + (\hat{r}_j \cdot \hat{\phi}_i) \right)$</td>
</tr>
<tr>
<td>$F_{31}^{ij}$</td>
<td>$-\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3r_j (\hat{r}_j \cdot \hat{\phi}_i) r_i (\hat{\phi}_j \cdot \hat{r}<em>i)}{(r</em>{ij})^2} + (\hat{\phi}_j \cdot \hat{r}_i) \right)$</td>
</tr>
<tr>
<td>$F_{32}^{ij}$</td>
<td>$-\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3r_j (\hat{r}_j \cdot \hat{\phi}_i) r_i (\hat{\phi}_j \cdot \hat{r}<em>i)}{(r</em>{ij})^2} + (\hat{\phi}_j \cdot \hat{r}_i) \right)$</td>
</tr>
<tr>
<td>$F_{33}^{ij}$</td>
<td>$-\frac{\mu_0}{4\pi (r_{ij})^3} \left( \frac{3r_j (\hat{r}_j \cdot \hat{\phi}_i) r_i (\hat{\phi}_j \cdot \hat{r}<em>i)}{(r</em>{ij})^2} + (\hat{\phi}_j \cdot \hat{r}_i) \right)$</td>
</tr>
</tbody>
</table>

Table 7.2: The elements of the $F^{ij}$-matrix. Note that they only depend on the geometry of the problem.
the sum of the potentials from the individual dipoles

\[ V(\vec{r}_i) = \sum_j V_j(\vec{r}_i) = -\frac{\mu_0}{4\pi} \sum_j \vec{m}_j \cdot \vec{\nabla}_j \left( \frac{1}{r_{ij}} \right), \tag{7.18} \]

where \( V_j \) is the potential of the \( j \)'th dipole. The magnetic field is in this case given from equation 7.17 by

\[ \vec{A}(\vec{r}_i) = \sum_{j=1}^{J} \vec{F}_j \vec{m}_j. \tag{7.19} \]

With this equation we can find the magnetic field anywhere above the surface of the Earth from the dipole moments of the dipoles representing the crustal magnetization.

### 7.2.2 Crustal magnetization and dipole moments

The above equation (7.19) gives the relation between the observed crustal field, \( \vec{A} \), and the dipole moments, \( \vec{m}_j \), of the dipoles representing the crustal magnetization. The last step needed is to relate the dipole moments to the magnetic crustal thickness.

Each of the \( J \) dipoles represents the magnetization of a crustal block. If the \( j \)'th crustal block with volume \( v_j \) centred at \( \vec{r}_j \) has an average magnetization of \( \vec{M}_j \), then the dipole moment of this crustal block is

\[ \vec{m}_j(\vec{r}_j) = \vec{M}_j(\vec{r}_j) \; v_j. \tag{7.20} \]

The volume, \( v_j \), is simply the product of the surface area covered by the dipole, \( w_j \), and the thickness of the magnetized layer, \( h_j \),

\[ v_j = w_j \; h_j. \tag{7.21} \]

The induced magnetization in the crust is related to the inducing field, \( \vec{B} \), by

\[ \vec{M}(\vec{r}) = \frac{\kappa(\vec{r})}{\mu_0} \vec{B}(\vec{r}), \tag{7.22} \]

where \( \kappa \) is the magnetic susceptibility, which among other things is material dependent [Clark and Emerson, 1991] and thus varies with position. The susceptibility, which describes the relationship between the induced magnetization and the inducing magnetic field of a given material, is in general a second-order symmetric tensor. Most crustal rocks have slightly anisotropic magnetic susceptibilities [Clark and Emerson, 1991]. However, when considering the bulk properties of a large volume of rock the effective susceptibility will generally be close to isotropic, such that the susceptibility is proportional to the identity matrix and can be treated as a scalar. Thus, this is a common assumption to make in studies of the crustal field [e.g. Clark and Emerson, 1991, Langel and Hinze, 1998]. We adopt this assumption and also assume a scalar susceptibility; this has the implication that the induced field becomes aligned along the inducing field, which simplifies the problem of determining the dipole moments of the crustal dipoles considerably, as will be shown in the next section. The susceptibility also depends on temperature. In particular, the susceptibility tends towards
7.2. EQUIVALENT SOURCE MAGNETIC DIPOLE METHOD

Zero, when the temperature of the material approaches its Curie temperature. Here it is assumed that the susceptibility is constant as long as the temperature is below the Curie temperature and is zero at temperatures above the Curie temperature. This means the susceptibility is assumed constant with depth in the crust at depths shallower than the Curie depth and zero at depths larger than this.

Combining equation 7.20–7.22 the magnetic dipole moment of the \( j \)’th crustal block can be expressed as a function of the inducing field, \( \bar{B} \) at \( \bar{r}_j \), the susceptibility of the \( j \)’th crustal block, \( \kappa_j \), the surface area of the block, \( w_j \), and its thickness, \( h_j \),

\[
\bar{m}_j (\bar{r}_j) = \frac{\kappa_j}{\mu_0} \bar{B} (\bar{r}_j) \ w_j \ h_j. \tag{7.23}
\]

The inducing field, \( \bar{B} \), and the surface areas, \( w_j \), which depends on the chosen distribution of dipoles, are known and the dipole moments of the individual dipoles will be determined from the observed field. This leaves two unknowns: the magnetic crustal thickness, \( h_j \), and the susceptibility of the crustal rocks, \( \kappa_j \).

As mentioned above the susceptibility of the rocks varies with depth primarily due to the temperature dependence of the susceptibility. It is, however, generally believed that it is impossible to determine the vertical variation of \( \kappa \) from satellite data [e.g. Mayhew, 1982b, Mayhew et al., 1991, Langel and Hinze, 1998]. Instead the vertically integrated magnetization (the integral of \( \kappa \) over \( h \)) of the crust is determined, and we then use a model for the susceptibility to find the magnetic crustal thickness from the vertically integrated magnetization.

In general the crustal field is stronger over the continents than over the oceans [Maus et al., 2005b], and the continental crust is in general thicker than the oceanic crust (on average about 40 km [Christensen and Mooney, 1995] compared to about 7 km [White et al., 1992]). The range of susceptibility values of continental rocks and that of the oceanic rocks overlap to a large extent (see e.g. Clark and Emerson [1991]), although there is a tendency for the mafic, oceanic rocks to have a slightly higher susceptibility than the felsic, continental rocks, because the mafic rocks have a higher content of iron, magnesium and titanium [Clark and Emerson, 1991, Langel and Hinze, 1998]. We therefore find it reasonable to assume that the crustal thickness variations dominate the susceptibility variations. We use a simple bimodal model for the susceptibility with \( \kappa = 0.035 \) for the continental crust, and \( \kappa = 0.040 \) for the susceptibility of the oceanic crust [Schlinger, 1985, Purucker et al., 2002]; the division between oceanic and continental crust is taken to be where the depth to the sea floor exceeds 800 m in order to include the continental shelf as continental crust. With all the other parameters known, it is now possible to determine the magnetic crustal thickness from the dipole moments by equation 7.23.

7.2.3 Directions of the dipoles

As mentioned earlier the basic idea of the ESMD method is to represent the crustal magnetization by a large number of dipoles. This requires estimating both the direction and magnitude of the dipole moments of these dipoles such that they mimic the observed field at satellite altitude. However, we above made the assumption that the susceptibility is scalar, and that the induced field is aligned along the inducing (core) field. This simplifies the problem as it then only is necessary to determine the
magnitudes of the dipole moments, as their directions are given by the direction of the inducing field at the dipole locations.

In figure 7.6 the directions of the dipole moments as function of latitude are illustrated for the case where the inducing field is a perfect dipole at the Earth’s centre aligned with the rotation axis; in this case there is no longitudinal component. In the real world the multipole terms of the core field cause a more complex pattern of the dipole moment directions, but they are nevertheless straightforward to find. When the induced field is aligned along the main field, the three components of the dipole moment, $m_j$, of the individual dipole can be written as the magnitude of the dipole moment, $m_j$, multiplied by proper projections of the declination, $D$, and inclination, $I$, of the main field:

\[
\begin{align*}
    m^r_j &= -m_j \sin I_j \\
    m^\theta_j &= -m_j \cos I_j \cos D_j \\
    m^\phi_j &= m_j \cos I_j \sin D_j, 
\end{align*}
\]

where $I_j$ and $D_j$ are the inclination and declination, respectively, at $\mathbf{r}_j$, which can be found from the inducing field. The magnitude of each dipole moment is related to the magnetic crustal thickness of the $j$’th crustal block through

\[
m_j = \xi_j \, h_j, \quad \text{(7.25)}
\]

with

\[
\xi_j = \frac{\kappa_j}{\mu_0} B(\tau_j) \, w_j, \quad \text{(7.26)}
\]

which is known for all the crustal blocks.
By inserting the expressions of 7.24 into equation 7.19 it can be seen that e.g. the radial component of the magnetic field becomes

\[ A_r(\vec{r}_i) = \sum_{j=1}^{J} \left( -\mathbf{F}^{ij}_{11} \sin I_j - \mathbf{F}^{ij}_{12} \cos I_j \cos D_j + \mathbf{F}^{ij}_{13} \cos I_j \sin D_j \right) m_j, \]  

(7.27)

with similar equations for \( A_\theta \) and \( A_\phi \). So by assuming that the dipole moments are aligned along the inducing field, the equations have been simplified, and in particular the three elements of each row of the \( \mathbf{F}^{ij} \)-matrix have collapsed into one element. Equation 7.27 shows that the magnetic field at \( \vec{r}_i \) depends on the magnitude of the dipole moments, \( m_j \), the geometry between the dipole locations and the observation points (through \( \mathbf{F}^{ij} \)), and the inclination and declination of the inducing field. As \( A_r, \mathbf{F}^{ij}, I_j, \) and \( D_j \) are known, it is possible to solve equation 7.27 for the magnitudes of the dipole moments. From equation 7.25 the magnetic crustal thickness can be found, once the magnitudes of the dipole moments have been determined from the observed magnetic field.

### 7.2.4 Modelling the induced crustal field

Consider \( J \) dipoles scattered evenly across the surface of the Earth, let \( \vec{d} \) be a vector of the magnitudes of the dipole moments, \( m_j \), for \( j = 1, 2, \ldots, J \). Assume that the magnetic field are observed at \( N \) locations, and let \( \vec{b} \) be the vector of the values of \( A_r \) at location 1, 2, \ldots, \( N \). Let

\[ G_{ij} = -\mathbf{F}^{ij}_{11} \sin I_j - \mathbf{F}^{ij}_{12} \cos I_j \cos D_j + \mathbf{F}^{ij}_{13} \cos I_j \sin D_j, \]  

(7.28)

where the subscript \( ij \) similar to before means the contribution from the \( j \)'th dipole to the field at the \( i \)'th observation point. Then from equation 7.27

\[ b_i = \sum_j G_{ij} m_j, \]  

(7.29)

which can be written in matrix form

\[ \vec{b} = \mathbf{G} \vec{d}. \]  

(7.30)

It is thus possible to calculate the magnetic field from a large number of dipoles at many locations at once by simple matrix multiplication.

In the above example only the radial component of the magnetic field at the observation points was included in the \( \vec{b} \)-vector, but the system of equations can easily be extended to include all three components of the field at every observation point, just let

\[ \vec{b} = \left( A_r(1), A_r(2), \ldots, A_r(N), A_\theta(1), A_\theta(2), \ldots, A_\theta(N), A_\phi(1), A_\phi(2), \ldots, A_\phi(N) \right)^T \]  

(7.31)

and extend the \( \mathbf{G} \)-matrix with

\[ \mathbf{G} = \begin{bmatrix} \mathbf{G}^r \\ \mathbf{G}^\theta \\ \mathbf{G}^\phi \end{bmatrix} \]  

(7.32)
where the elements of $G_r$ are $G_{ij}$ from equation 7.28, and $G^\Theta$ and $G^\Phi$ are made similarly to $G_r$ by using the elements from the second and third row of the $F_{ij}$-matrix, respectively.

Since we only are interested in finding (or at least using) the solution of the magnetic crustal thickness in areas at high latitude, it suffices to use only the radial component of the magnetic field at the observation points to find a good solution for the magnitudes of the dipole moments, because the inducing field, and thus also the induced field, is close to vertical in the polar regions. If the area of interest was in equatorial regions, it would be sensible to also include the horizontal components of the observed field to better constrain the solution, as the inducing and induced field are close to horizontal here. In most of our studies, we solved for the magnitudes of the dipole moments using only the radial component of the observed field. We did, however, make some test where the north-south component ($\theta$-component) of the magnetic field observations was included to see if this had any significant effect on the results of the magnetic crustal thickness obtained for Antarctica and Greenland. We found that there was no significant difference between the two cases.

For the inducing field we used the Ørsted(05m/02) field model to degree and order 13, from this we determined the inclination, declination, and field strength at each dipole location. The choice of core field model is not crucial as the core field is fairly well determined and different core field models do not differ significantly.

### 7.2.5 Locations of observations and crustal dipoles

For the locations of the crustal dipoles we have chosen to use a spherical icosahedron grid [Covington, 1993]. The icosahedron grid has the advantage that the size of the area represented by each dipole is roughly the same and the inter-dipole distance between neighbouring dipoles is almost constant.

We have worked with two grid-sizes, a small grid of 11562 dipoles, and a large grid of 21162 dipoles. For the case of 11562 dipoles the distance between any two neighbouring dipoles are about 1.9° or about 211 km, in the 21162 dipole case the inter-dipole distance is about 1.4° or about 156 km. Thus our resolution is in both cases a few hundred of kilometres. For both the 11562 case and the 21162 case, the dipole locations in Antarctica and the surrounding area are shown in figure 7.7.

As the data are given by a field model, the grid of observation points must be specified by the user. We use the same icosahedron grid for the data points as for the dipole locations. The icosahedron grid has the advantage that it weights the observations evenly with geographic latitude; a regular latitude-longitude grid has a much closer data spacing in the polar regions, which gives the polar data a higher weight than the equatorial data. Thus, we use the same geographical latitudes and longitudes for both dipole locations and observation points. The crustal dipoles are assumed to lie on a spherical shell with a radius of 6371.2 km, thus we do not take the Earth’s ellipticity into account, as it is a minor effect. We have used two different altitudes of the observations; 300 km and 400 km. The results presented in this thesis are for the case of 21162 dipoles, with the observations being synthesized at an altitude of 300 km from MF3, with the remanent field subtracted. Some results for the magnetic crustal thickness obtained using 11562 dipoles, with the observations being synthesized from CM4 (degrees 15–65) can be seen in Fox Maule et al. [2005a].
7.3. Iterative forward modelling

Equations 7.25 and 7.30 describe how the magnetic field from the crustal dipoles can be calculated if the magnetic crustal thickness is known. To calculate the magnitude of the dipole moments of the crustal dipoles from the observed field, as we want, is the inverse problem. In the general case the number of observations and the number of dipoles are not necessarily the same, and the inverse of $G$ cannot be found directly. Our $G$-matrix is by incident square, as we have chosen the same number of dipole locations and observation points and $G$ does in fact have full rank and so the inverse of $G$ does exist, but as $G$ has a fairly high condition number, the inverse of $G$ is unlikely to yield a reasonable solution to the problem, because noise will be blown up in such a manner, that it is likely to drown the signal. Therefore and due to the large size of $G$ we solve for the dipole moments from the observations using an iterative forward modelling procedure. The $G$-matrix is non-symmetric, as the contribution to the magnetic field at the $p$'th observation point from the $q$'th dipole in general is not the same as the contribution to the field at the $q$'th observation point from the $p$'th dipole. The condition number of the small $G$-matrix ($11562 \times 11562$) is about $5 \cdot 10^6$, and $G$ has a very small determinant.

I will first give a very brief description of the iterative procedure to provide an overview of the process, then a more detailed description of the individual steps will be given. A schematic of the iterative procedure is shown in figure 7.8 (from Fox Maule et al. [2005b]). First the induced field of an initial model of the crustal thickness is calculated using equations 7.25 and 7.30. This is then compared to the observed induced field. If the difference between the two is higher than the uncertainty of the observations, the crustal thickness model is improved by calculating a correction to the crustal thickness model. The correction is found by inverting the difference
between the magnetic fields with a least square conjugated gradient algorithm on a sparse version of the \( G \)-matrix. The correction is then added to the initial crustal thickness model, to obtain an improved crustal thickness model. Then the induced field of the improved crustal thickness model is calculated and compared to the observed induced field etc. until the crustal thickness model has an induced field, which is sufficiently close to the observed induced field. This crustal thickness model is then accepted as the solution.

The very first step of the procedure is to “create” the observations, the observed induced field, this was already described in section 7.1. The observed crustal field is synthesized from a field model (MF3 or CM4) and high-pass filtered by removing all spherical harmonic coefficients for degrees less than 15. The contribution to the crustal field from remanent magnetism in the crust is taken from the model by Dyment and Arkani-Hamed [1998b] and Purucker and Dyment [2000], which is also high-pass filtered. By subtracting the high-pass filtered remanent field, \( \vec{b}_{\text{rem, hp}} \), from the high-pass filtered observed field, \( \vec{b}_{\text{obs, hp}} \), the high-pass filtered observed induced
7.3. **ITERATIVE FORWARD MODELLING**

Field,
\[ \vec{b}_{\text{obs,ind,hp}} = \vec{b}_{\text{obs,hp}} - \vec{b}_{\text{rem,hp}}, \]  
(7.33)
is obtained. This is shown for MF3 at 300 km altitude in figure 7.3. It is this field, the field from the crustal dipoles shall reproduce.

To start the iterative procedure, an initial model for the magnetic crustal thickness is needed. We use the 3SMAC crustal thickness model by Nataf and Ricard [1996] (3SMAC: Three-dimensional seismological model a priori constrained). The 3SMAC model is designed to provide a good crustal model to be used in global seismology. It is based on seismic data, as well as other geophysical, geological, and geochemical data. In the 3SMAC model the crust is partitioned into layers of water, ice, sediments and igneous rock on a \(2^\circ \times 2^\circ\) grid; it is the igneous crustal thickness from the 3SMAC model, shown in figure 7.9, that we use as initial model for the magnetic crustal thickness. The thickness of the individual layers specified in the 3SMAC model is compiled from many different studies, using a variety of observation techniques, however, the locations of crustal thickness data are very unevenly distributed, and the amount of data is particularly small in polar regions. Thus it is reasonable to improve the model in these areas; besides the igneous crustal thickness is in general not equal to the magnetic crustal thickness, but it does provide a good starting point. Apart from providing a reasonable initial crustal thickness model, 3SMAC also delivers the long-wavelength part of the solution for the magnetic crustal thickness. Since it is necessary to high-pass filter the magnetic field observations to remove the crustal field, the observations cannot resolve the long-wavelength variations of the magnetic crustal thickness. So we use the long-wavelength part of the crustal thickness from 3SMAC to provide the long-wavelength part of the solution of the magnetic crustal thickness and iteratively solve for the short-wavelength part (degrees > 14) of the magnetic crustal thickness from the observations. As we directly adopt the long-wavelength part of our initial model for our end model, it is naturally very important to select a good initial model, which we believe 3SMAC is.

The first step of the iterations is to calculate the dipole moments, \(\vec{d}_j^{(0)}\), at the \(J\) dipole locations from the initial crustal thicknesses at the dipole locations, \(h_j^{(0)}\), as provided by 3SMAC
\[ \vec{d}_j^{(0)} = \xi_j \cdot h_j^{(0)}, \]  
(7.34)
where the superscript denotes the iteration number and \(\xi_j\) is given by equation 7.26. The close relation between the magnetic crustal thickness and the dipole moments means that they will be referred to interchangeably in the description of the modelling procedure. The induced magnetic field as described by these dipole moments of the starting model is then calculated by using
\[ \vec{b}_{\text{model,ind}}^{(0)} = \mathbf{G}\vec{d}^{(0)}, \]  
(7.35)
although we, due to its size, do not actually form the \(\mathbf{G}\)-matrix in the process, but simply calculate the magnetic field due to each dipole at each data point and sum the results continuously. In order to compare the induced field of the crustal thickness model with the observed induced field, the modelled induced field is high-pass filtered,
The difference between the modelled induced high-pass filtered field, \( \tilde{b}_{\text{model,ind,hp}} \), and the observed induced high-pass filtered field, \( \tilde{b}_{\text{obs,ind,hp}} \), is obtained by subtraction

\[
\Delta \tilde{b}^{(0)}_{\text{ind,hp}} = \tilde{b}_{\text{obs,ind,hp}} - \tilde{b}_{\text{model,ind,hp}}^{(0)}.
\] (7.36)

If the difference between the induced field of the model and the observed induced field is larger than some threshold value, the crustal thickness model should be improved, such that its induced field comes closer to the observed induced field. A correction to the initial model of the dipole moments, \( \Delta \tilde{d}_{hp}^{(0)} \), is estimated by inversion of

\[
\Delta \tilde{b}_{\text{ind,hp}}^{(0)} = G^s \Delta \tilde{d}_{hp}^{(0)},
\] (7.37)

using a least squares conjugate gradients (LSQR) method [Paige and Saunders, 1982] solving the normal equations of the linear system. To reduce the number of computations the inversion is done with a sparse version of the \( G \)-matrix, \( G^s \). In \( G^s \) only dipoles lying within a distance of 2500 km from a given observation point, contribute to the field at this point, thereby most of the elements of \( G \) are set to zero; in \( G^s \) only about 3–4% of the elements of are non-zero.

An improved estimate of the magnitudes of the dipole moments is then obtained by

\[
\tilde{d}^{(1)} = \tilde{d}^{(0)} + \Delta \tilde{d}_{hp}^{(0)}.
\] (7.38)
From this the induced field of the improved estimate of the magnitudes of the dipole moments is calculated using equation 7.35, which then is compared to the observed induced field etc. until a model for the dipole moments is found that gives a satisfying fit to the observed induced field.

The iterations should in principle be stopped once the difference between the induced field of the crustal model and the observed induced field, becomes less than the level of uncertainty of the observed induced field. However, as described in detail in chapters 5 and 6, the level of uncertainty of most field models, including MF3 and CM4, is not very well known, which makes it difficult to specify a level of halt. Instead we stop iterating when the magnetic crustal thickness model has stabilized, and we note that the solution obtained appears reasonable in that it neither is too smooth nor too noisy.

### 7.3.1 Annihilators

Solutions to the magnetic inverse problem of finding the crustal magnetization from the observed field are in general non-unique. It has been shown by Runcorn [1975] and Maus and Haak [2003] that there exist nontrivial crustal magnetizations, annihilators, that have a vanishing magnetic potential and thus no magnetic field outside the Earth’s surface. The simplest of the annihilators is a homogeneous shell of constant thickness and susceptibility magnetized by interior sources, which does not have a magnetic field external to the shell [Runcorn, 1975]. But there also exist much more complicated annihilators, which, however, all are centred on the dip equator. According to Maus and Haak [2003], any vertically integrated magnetization distribution along the dip equator can for a given dipole dominating inducing field be extended north-south to form an annihilator [Maus and Haak, 2003, Nolte and Siebert, 1987]. As we have focused on areas in the polar regions, we expect that annihilators have little effect on our results, but we have nevertheless investigated if annihilators are a problem for our solution.

To see how much of the solution of the inverse problem that is unconstrained by the data, the solution should be projected onto the solution null space, which consists of all the annihilators. The part of the solution that lies in the solution null space is unconstrained by the data. The analytically determined annihilators discovered by Maus and Haak [2003] and Runcorn [1975] were found for the continuous solution of the magnetic inverse problem working with spherical harmonic representations. Discretization of a continuous problem is, however, likely to affect the null space. In a finite discretization the solution cannot wriggle as much as a continuous solution can, thus the discretization decreases the solution space to only contain the solutions that can be spanned by the base vectors of the discrete space, which in our case are the dipoles. This subspace of the continuous solution space may not contain the full original null space. It seems unlikely that the dipoles would be able to produce a magnetic field that is zero everywhere outside the surface of the Earth, but it could be imagined that a particular distribution of dipole moments could produce a field which is vanishing at a particular altitude (e.g. 300 km); this would constitute an annihilator in our formulation of the problem.

As the magnetic field from the dipoles is found by multiplying the full $G$-matrix with the vector of dipole moments, the annihilators to the discretized version of
the problem is given by the null space of the $G$-matrix. Any null vector, as well as any linear combination of null vectors of $G$, constitutes an annihilator. The $G$-matrix depends on geometry between the data points and dipole locations, and on the direction of the inducing field (inclination and declination), as can be seen in equation 7.28, thus so will the annihilators as expected. The strength of the inducing field does not enter the equations, thus for a particular geometry, a specific magnetization becomes an annihilator no matter the strength of the inducing field.

To determine whether our discretization had conserved (part of) the null space, we determined the rank of the $G$-matrix, as a matrix with full rank has no null space [Messer, 1994]. We found that the small $G$-matrix ($11562 \times 11562$) has full rank, and thus no null space. It was not possible for us to calculate the rank of the large $G$-matrix, but since the geometry of the discretizations for the two cases are similar, it seems reasonable that the large $G$-matrix also has full rank. However, if the condition number of $G$ was very high, solutions lying very close to the null space could occur, which in practice would act as null vectors, but the condition number of the $G$-matrix is not large enough for this to be a serious problem in our case. Thus with the used discretization, no significant part of the solution null space existing for the continuous problem is contained in the discrete solution subspace, which means that the solution we have obtained for the magnetic crustal thickness is unique in the solution subspace spanned by our discretization.

7.4 Results for the magnetic crustal thickness

As mentioned previously we have solved for the magnetic crustal thickness from MF3 and CM4 globally to avoid edge-effects. However, our main objective is to use the obtained magnetic crustal thickness to estimate the geothermal heat flux underneath the large ice sheets of Antarctica and Greenland, thus we only present and interpret the results there. A discussion of error sources and an uncertainty estimate is given subsequently to the presentation of the results.

7.4.1 Greenland

The results we have obtained for the magnetic crust of Greenland, using 21162 dipoles to represent the crustal field and with the observed field given by MF3 at 300 km altitude is shown in figure 7.10, along with the used initial model for the igneous crustal thickness by 3SMAC.

We find that the magnetic crust is thickest in the northern part of Greenland, particular in north-west Greenland, where the magnetic crust is more than 50 km thick. Along the east coast of Greenland, in and between King Christian X Land and King Christian IX Land, we find that the magnetic crust is very thin. Intermediate magnetic crustal thickness is found in the southern part of Greenland, with somewhat lower values extending as a band across Greenland at about 67.5°N.

In a recent seismic study, the depth to Moho (the thickness of the geologic crust) was determined at 19 locations in Greenland by Dahl-Jensen et al. [2003b]. We have compared these results with our results for the magnetic crustal thickness in figure 7.11, to see if the magnetic crustal thickness that we find is less than the thickness of the geologic crust. At almost all the locations where the depth to Moho
7.4. RESULTS FOR THE MAGNETIC CRUSTAL THICKNESS

Figure 7.10: The left panel shows the initial model of the crustal thickness from 3SMAC over Greenland. In the right panel the result for the magnetic crustal thickness in Greenland is shown.

was determined, shown in figure 7.11, the geologic crust is thicker than the magnetic crust. Thus, it appears the Curie isotherm lies within the geologic crust in Greenland. It should, however, be mentioned, that Dahl-Jensen et al. [2003b] unfortunately did not determine the depth to Moho in the area in north-western Greenland, where we find a very thick magnetic crust. In particular, we note that the magnetic crust is thinner than the geologic crust in the eastern part of Greenland, where the data of Dahl-Jensen et al. [2003b] can be supplemented by the results of a local seismic study that was made in central east Greenland by Mandler and Jokat [1998]. They found the crust to be between 22 and 48 km thick in the area of their study (70–72°N, 21–29°W) with the crustal thickness decreasing towards the coast.

Results obtained for the magnetic crustal thickness in Greenland using 11562 dipoles, with data synthesized from CM4 (degrees 15–65) are shown and published in Fox Maule et al. [2005a]. These results are very similar to the results presented here.

7.4.2 Antarctica

The result we have obtained for the magnetic crustal thickness in Antarctica, using 21162 dipoles and with the observed field given by the MF3 crustal field model synthesized at 300 km altitude is shown in figure 7.13, along with the initial crustal thickness from the 3SMAC model. These results are published in Fox Maule et al.
The geologic crust is in general thicker in East Antarctica than in West Antarctic, a fact which is supported by both seismic and gravity studies (see e.g. von Frese et al. [1999], Ritzwoller et al. [2001], Llubes et al. [2003] and references therein). We find that also the magnetic crust is thicker in East Antarctica than in West Antarctica. In the central part of East Antarctica, our model predicts a magnetic crustal thickness in excess of 50 km. The area around Victoria Land, Oates Land and George V Land in East Antarctica has a thin magnetic crust. We also find that the magnetic crust is thin in a belt along the East-West Antarctica boundary and along the Siple Coast. The coastal part of West Antarctica is characterized by intermediate values of the magnetic crustal thickness.

The thin magnetic crust we find in West Antarctica do at least to some extent overlap with the West Antarctic Rift System [e.g. Behrendt, 1999, Winberry and Anandakrishnan, 2003, Bennet, 2003] (see figure 7.12). It is reasonable to expect a thin magnetic crust in a rift zone due to the extensional crustal thinning which result
7.4. RESULTS FOR THE MAGNETIC CRUSTAL THICKNESS

Figure 7.12: Map of Antarctica. WARS is the West Antarctic Rift System, its boundaries (after Winberry and Anandakrishnan [2003]) are illustrated with the blue lines. Antarctica is, with its area of about $1.4 \times 10^6 \text{ km}^2$, about six times larger than Greenland, which has an area of about $2.2 \times 10^6 \text{ km}^2$.

in elevated heat flow.

Moho depth determinations from local seismic experiments in Antarctica are scarce (see e.g. Ikami et al. [1985], Guterch et al. [1985]), but seismic and gravity data have been used to estimate the crustal thickness over the Antarctic continent. Ritzwoller et al. [2001] used data from seismic surface waves to determine the structure of the crust and the upper mantle of Antarctica. To constrain their solution of the crustal thickness they used another seismic crustal model CRUST5.1 [Mooney et al., 1998] as an initial model in their inversion for the crustal and upper mantle structure. The resulting crustal thickness they found to be consistent with the seismic data is close to that of CRUST5.1. They found that the crustal thickness in West Antarctica on average is about 27 km, and in East Antarctica the average crustal thickness is about 40 km, with the maximum thickness approaching 45 km in the central parts of East Antarctica.

Llubes et al. [2003] used CHAMP gravity data to infer the crustal thickness in Antarctica. They also used an initial model based on seismic data for their inversion for the crustal thickness, as us Llubes et al. [2003] used the 3SMAC model. The resulting crustal thickness they found varies with 34.1 km from 8.5 km to 42.6 km, with thicker crust in East Antarctica than in West Antarctica. This is somewhat less than the crustal thickness variation of 41.9 km of the 3SMAC model (from 7.3 km to 49.2 km). For their inversion they assumed constant density of the crustal rocks.

Although there obviously are made simplifying assumptions in both of the above described attempts to infer the crustal thickness of Antarctica, we also make (other) simplifying assumptions, so we find it instructive to compare the results of the three
Figure 7.13: The left panel shows the initial model of the crustal thickness from 3SMAC over Antarctica. In the right panel the result for the magnetic crustal thickness in Antarctica is shown.

The magnetic crust in most of West Antarctica appears to be thinner than the geologic crust. The thick magnetic crust we find in the central part of East Antarctica, however, appears to be thicker than the predicted thickness of the geologic crust from seismic and gravity studies. All of the estimates of the (magnetic) crustal thickness (the seismic study, the gravity study and our magnetic study) are based on several simplifications, which may explain the discrepancy. The seismic study is for example limited by the low number of seismic stations in Antarctica. The gravity study is among other things limited by the lack of knowledge of the variation of density of the crustal rocks. We assumed constant magnetic susceptibility of the crustal rock in the entire continent; if the susceptibility of the rocks is underestimated, our magnetic crustal thickness may be overestimated. Furthermore the presence of unknown remanent magnetism may cause us to overestimate the thickness of the magnetic crust. Finally, although unlikely [Wasilewski and Mayhew, 1992], we cannot rule out the possibility of magnetic rocks in the mantle below Moho [Hayling, 1991, Hamoudi et al., 1998] in the central part of East Antarctica. Summarizing, we find it likely that the Curie isotherm in general lies shallower than Moho in West Antarctica, and around Victoria Land, Oates Land and George V Land, whereas this is more doubtful in central East Antarctica.
7.4. RESULTS FOR THE MAGNETIC CRUSTAL THICKNESS

7.4.3 Error sources and uncertainty estimate

There are several different sources of error that may influence the results we have obtained for the magnetic crustal thickness, these include the uncertainty of the field model, the separation of the induced and remanent crustal magnetism, the susceptibility model and the initial crustal thickness model, which also provides the long-wavelength part of the solution.

The level of uncertainty of magnetic field models was the main topic of chapters 5 and 6, and were discussed extensively there. As mentioned in chapters 5 and 6, there are usually not given valid uncertainty estimates of field models, this includes MF3 and CM4. The uncertainty of the magnetic field model is probably of the order of 25% [Lowes and Olsen, 2004], and is thus the dominating error source on the estimate of the magnetic crustal thickness.

Another potential source of error is the model for remanent magnetization we applied to isolate the induced part of the crustal field. In particular the model of remanent magnetization does not account for remanent magnetism in the continental crust, although this does occur and can give rise to significant magnetic anomalies. There does not exist detailed maps of remanent magnetization for neither Greenland nor Antarctica, but no large remanent anomalies are obvious from the observed crustal field. Furthermore our method does not give unrealistic (either too high or negative) magnetic crustal thickness values, hence a removal of remanent magnetization is not required to explain our results, thus we conclude that neither Antarctica nor Greenland have any ‘super’ anomalies that require hard remanent magnetization to be explained. Purucker and Whaler [2003] found that remanent magnetism appears to be negligible in the continental crust in Greenland, so we consider this source of error minor in Greenland, but the uncertainty due to unconsidered remanent magnetization may be important in some regions of the Antarctic. Local magnetic anomalies detected in aeromagnetic studies in West Antarctica have been interpreted as at least in part being due to remanent magnetization in the direction of the current field [Behrendt et al., 1994]. However, the seismic starting model we have used predicts the location and approximate magnitude of 8 of the 10 major regional magnetic anomalies associated with crustal thickness changes, and it predicts all of the major features in the vicinity of the thermal anomalies. Thus we believe that induced magnetization can explain at least 80% of the observed magnetic anomalies. We therefore estimate that the error from remanent magnetization is at most 20%.

It is not really clear to us how to quantify the uncertainty of the starting model, and thus on the long-wavelength part of the resulting magnetic crustal thickness. It is, however, obvious that as the long-wavelength part of the 3SMAC model is directly adopted in the resulting model, an error in the 3SMAC model, particularly in the long-wavelength part of the model, will be directly transferred to the magnetic crustal thickness result. However, the long-wavelength part of the 3SMAC model is likely to be that part of the model that is best constrained by the data used to develop the 3SMAC model. We believe the 3SMAC model at the moment is the best available crustal thickness model to be used as initial model in our study, and that any errors introduced by the 3SMAC model are negligible compare to the errors introduced from the field model. As the starting model predicts 8 of the 10 major regional magnetic anomalies associated with crustal thickness changes as previously
discussed, we consider this a minor error. We estimate that the error from this source is less than about 7%.

We applied a crude model for the magnetic susceptibility of the crustal rocks, which do not account for vertical variation, or lateral variations other than the continent-ocean crust difference of the susceptibility. There does not exist global maps of the magnetic susceptibility of the crustal rocks, that are not based on satellite magnetic data of the crustal field. However as already mentioned, the crustal field is in general stronger over the continents than over the ocean, and the crustal thickness difference is larger than the susceptibility difference. Thus we consider the lack of accounting for lateral variation of the susceptibility to be a minor source of error; we expect this error to be less than about 7%. We do, however, acknowledge that e.g. a large magnetite-ore deposit as the one in Kiruna in Sweden is a potential source of error. The large susceptibility and the possible remanent magnetization of such an ore body would cause us to make an erroneous estimate of the magnetic crustal thickness in the area surrounding such an ore body if present in either Greenland or Antarctica.

In total the level of uncertainty of the derived magnetic crustal thickness due to these independent error sources is about 29%.

7.5 Conclusions

We have estimated the thickness of the magnetic crust in Greenland and Antarctica from the crustal field given by the MF3 and CM4 field models using the equivalent magnetic source dipole method. The magnetic crustal thickness was determined with an iterative forward modelling procedure.

We found that the thickness of the magnetic crust is very varying in both Greenland and Antarctica. In Greenland we find that the magnetic crust is thick in north-west Greenland, and thin in east Greenland. In Antarctica we find a thick magnetic crust in central East Antarctica, and a thin magnetic crust in a band along the boundary between East and West Antarctica, and in Victoria Land, Oates Land and George V Land. Comparing these results with results obtained for the depth to Moho from seismic and gravity studies, we believe that the Curie isotherm lies shallower than Moho in most of Greenland, West Antarctica, and in Victoria Land, Oates Land and George V Land in East Antarctica. It is, however, possible that the Curie isotherm lies deeper than Moho in central East Antarctica, and possibly also in north-west Greenland.
Chapter 8

Geothermal heat flux

Heat from the hot interior of the Earth continuously flows through the (solid) surface of the planet to the surrounding oceans, ice sheets and atmosphere. The geothermal heat flux varies with location, it is e.g. high along ocean spreading ridges and in other volcanic areas. In general, the geothermal heat flux is higher in the oceans than over the continents [Pollack et al., 1993], but the variability of the heat flux is large within both the oceans and the continents. One of the places where the geothermal heat flux is of interest is underneath the large ice sheets. The geothermal heat flux is an important factor in the dynamics of ice sheets, the occurrence of subglacial lakes and onset of ice streams, and it may affect the mass balance of the ice sheets. However, direct heat flux measurements in ice covered regions are difficult for obvious reasons. Thus we have developed a method using satellite magnetic data to estimate the heat flux underneath the ice sheets in Greenland and Antarctica.

In this chapter I first present the thermal model we have derived for the crust to determine the geothermal heat flux. Then the results we have obtained for the geothermal heat flux underneath the Antarctic and Greenlandic ice sheets are presented and discussed.

8.1 Heat flux model

To estimate the geothermal heat flux underneath the ice sheets we use that the depth to the Curie isotherm can be determined from the induced crustal field, which may be inferred from satellite magnetic measurements. With the temperature known at a specific depth in the crust, it is possible to determine the temperature profile in the crust and from this the geothermal heat flux.

Both the crust and a part of the upper mantle belong to the thermal lithosphere, which is defined as that outer region of the Earth where heat conduction is the dominant heat transfer mechanism. The transition between the thermal lithosphere and the underlying mantle is gradual, but for many purposes the lower boundary of the thermal lithosphere is taken to be around 1350°C [e.g. Sahagian and Holland, 1993]. Thus the magnetic crust bound by the Curie isotherm of 580°C is always well within the thermal lithosphere and the heat conduction regime. Although the dominant heat transfer mechanism of the crustal rocks is heat conduction, heat advection by transport of hot magma or by water in aquifers may locally be an important heat transfer mechanism.
transfer mechanism in the crust. If the temperature at some depth within the thermal lithosphere is known, it is possible to estimate the temperature gradient and the surface heat flux from the heat conduction equation, if the conductivity of the rocks, the bedrock surface temperature and the heat production within the crust/lithosphere is known.

To estimate the geothermal heat flux we have derived a thermal model for the continental crust. The vertical temperature gradient in the continental crust is much larger than the lateral gradients, thus we use a one-dimensional approximation for the thermal regime of the crust. This has the implication that the thermal model is not valid for the oceanic crust as it for the oceanic crust in general is necessary to account for horizontal heat advection due to the lateral motion of the ocean floor caused by plate tectonics. We assume steady state, this is likely to be reasonable for the stable part of the continental crust, whereas the crust in tectonic active areas may not be in thermal equilibrium and the temperature profile may be changing with time. To account for crustal heat production by decay of radioactive elements, we have applied a widely used model of exponentially decreasing heat production with depth [e.g. Sandiford and McLaren, 2002],

\[ H(z) = H_0 \exp\left(\frac{-z}{\delta}\right), \]  

(8.1)

where \( H_0 \) is the heat production at the surface in W/m³ and \( \delta \) is the scale depth at which the heat production is decreased to \( 1/e \) of \( H_0 \). The heat production in general varies with age and type of the rock (see e.g. Sass et al. [1972]), but as very little is known about the geology underneath the ice sheets, we assume the same heat production profile everywhere. Thus the heat production only varies laterally due to changes in the thickness of the magnetic crust.

Under these assumptions the heat conduction equation is

\[ \frac{\partial^2 T(z)}{\partial z^2} = -\frac{H_0}{k} \exp\left(\frac{-z}{\delta}\right), \]  

(8.2)

where \( k \) is the thermal conductivity of the crustal rocks. In the general case, the thermal conductivity is a second order tensor, but for many rocks, particularly volcanic and plutonic rocks, the thermal conductivity is isotropic and may be treated as a scalar [Clauser and Huenesch, 1995]. The thermal conductivity is material dependent and thus varies with geology, it is also temperature and pressure dependent and thus varies with depth [Clauser and Huenesch, 1995]. However since we know very little of the crustal geology underneath the ice, we use a constant thermal conductivity of the rocks.

To solve the second order differential equation (8.2) for the temperature profile of the crust, it is necessary to specify two boundary conditions. We use that the temperature at the Curie depth, \( z_c \), is the Curie temperature, \( T_c \), and that the temperature is \( T_0 \) at the bedrock surface, \( z = 0 \). The solution for the temperature profile of the crust, \( T(z) \), to this boundary value problem is

\[ T(z) = T_0 + \frac{H_0 \delta^2}{k} \left( 1 - \exp\left(\frac{-z}{\delta}\right) \right) + \frac{z}{z_c} \left( T_c - T_0 - \frac{H_0 \delta^2}{k} \left( 1 - \exp\left(\frac{-z_c}{\delta}\right) \right) \right). \]  

(8.3)
The heat flux is related to the temperature through
\[ q(z) = -k \frac{\partial T}{\partial z}. \] (8.4)

To get the surface heat flux \( q \) is evaluated at \( z = 0 \), which gives
\[ q(z = 0) = \frac{k(T_c - T_0)}{z_c} - H_0 \delta + \frac{H_0 \delta^2}{z_c} \left( 1 - \exp \left( -\frac{z_c}{\delta} \right) \right). \] (8.5)

From this equation the heat flux underneath the ice sheet is found.

We take the radiogenic heat production at the surface to be \( H_0 = 2.5 \times 10^{-6} \text{ W/m}^3 \), and use a scale depth of \( \delta = 8 \text{ km} \) [Sandiford and McLaren, 2002, and references therein]. The temperature at bedrock surface varies slightly underneath the ice, but it is in many places close to the melting point temperature of the ice at the given pressure. Since only the difference between the Curie temperature and the bedrock surface temperature appear in equation 8.5, it is not necessary to specify either boundary temperature but only their difference. We take \( T_c - T_0 = 580 \text{ K} \), and assume that the Curie depth is given by the magnetic crustal thickness, and use the values we obtained in chapter 7. For the thermal conductivity we use \( k = 2.8 \text{ W/mK} \) for Antarctica, and \( k = 2.4 \text{ W/mK} \) for Greenland; both of these values are within the range of thermal conductivity typically observed for volcanic and plutonic rocks of 1.5–3.5 W/mK [Clauser and Huenges, 1995], but with these specific choices of the thermal conductivity the average heat flux in both Antarctica and Greenland becomes 65 mW/m², which is the average continental heat flux [Pollack et al., 1993].

## 8.2 Heat flux underneath large ice sheets

The geothermal heat flux can vary significantly over distances of only tens of kilometres due to local geologic settings [e.g. Majorowicz and Embry, 1998, Dahl-Jensen et al., 2003a, Näslund et al., 2005]. It is very difficult to determine the geothermal heat flux underneath the large ice caps. Direct measurements can be made at the bottom of boreholes as it has been done by e.g. Dahl-Jensen et al. [1998] in Greenland and Engelhardt [2004] in Antarctica. However, making these measurements is time consuming and costly, and therefore only very few have been made. Furthermore, as mentioned by Engelhardt [2004], reliable measurements of the heat flux can only be done at the bottom of boreholes where there is little or no ice flow and the ice is frozen solid at bedrock to avoid heat sources from shear heating of the basal ice, heat advection, basal friction and latent heat from melting of basal ice. All these possible additional heat sources and sinks at the bottom of the ice makes it difficult to isolate the contribution from the geothermal heat flux when the heat flux at the bottom of the ice is measured [Engelhardt, 2004].

Thus it is highly desirable to infer the geothermal heat flux underneath the ice from indirect methods; for example Shapiro and Ritzwoller [2004] derived a heat flux map of Antarctica using seismic information to interpolate worldwide heat flux measurements, Dahl-Jensen et al. [2003a] used airborne radio-echo sounder observations combined with information from the NGRIP drill site to locally infer the geothermal heat flux at and around the NGRIP drill site in central Greenland, and
Fahnestock et al. [2001] used airborne radar soundings to estimate the basal melting and the geothermal heat flux along an extensive flight track across Greenland. We suggest that using satellite magnetic data is a useful supplement or alternative. We have determined the heat flux underneath the ice sheets in Greenland and Antarctica from the thermal model described above using the values obtained for the magnetic crustal thickness, i.e. the depth to the Curie isotherm, in the previous chapter. In the following first the achieved results for Antarctica and Greenland are presented and compared with other heat flux estimates, then a discussion on error sources and an uncertainty estimate is given.

8.2.1 Antarctica

The geothermal heat flux underneath the Antarctic ice sheet is shown in figure 8.1 (from Fox Maule et al. [2005b]). We find that the heat flux map reflects the magnetic crustal thickness as expected, with high heat flux in regions where the magnetic crust is thin and vice versa (see figure 7.13). In the central part of East Antarctica we find heat flux values of 50–60 mW/m\(^2\), which is below the continental average of 65 mW/m\(^2\) [Pollack et al., 1993]. High heat flux is found on the East-West Antarctica boundary and in the area of Victoria Land, Oates Land and George V Land. Heat flux values above average are also found in the Siple Coast area.

We find that our results of the heat flux in Antarctica are supported by geologic evidence of high heat flux as volcanoes, subice lakes and ice streams, which are corridors of fast flowing ice within the ice sheets. The pink dots in figure 8.1 mark the locations of known volcanoes. It is seen that the volcanic areas, e.g. in Victoria Land and coastal areas of West Antarctica, lie in areas were we find high heat flux as would be expected. Based on aeromagnetic, seismic and radar ice sounding data it is believed that there is subice volcanism in connection with the West Antarctic Rift System in West Antarctica [Blankenship et al., 1993, Behrendt et al., 1994, Behrendt, 1999]. Blankenship et al. [1993] found evidence of recent active volcanic activity in the West Antarctic Rift System and estimated the local heat flux value around the specific volcano to be 10–25 W/m\(^2\), two orders of magnitude larger than the highest heat flux we find. This high heat flux is, however, likely only to be over an area of about 50 km\(^2\), much smaller than our resolution. Satellite altitude limits our resolution to at least a few hundred kilometres [Mayhew, 1982a], this causes a smoothing which should be kept in mind when comparing the heat flux map derived from satellite magnetic data with direct measurements or localized estimates of the heat flux.

The Siple Coast area, where we find elevated heat flux, is known to have several ice streams; a review of the Siple Coast ice streams is given by Bennet [2003]. It has been suggested that high heat flux could be one of the trigger mechanisms for ice stream formation [Blankenship et al., 1993, Fahnestock et al., 2001, Engelhardt, 2004], again supporting results produced by our model. Direct measurements of the heat flux at the bottom of boreholes around Siple Coast by Engelhardt [2004] gave a value of 69 mW/m\(^2\), we find a value of 80 mW/m\(^2\) in this area, but the discrepancy is within our level of uncertainty of 37% (see below). However, a modelling study of the energy balance of one of the Siple Coast ice streams [Raymond, 2000] concluded that the geothermal heat flux underneath this particular ice stream must exceed 80 mW/m\(^2\).
8.2. HEAT FLUX UNDERNEATH LARGE ICE SHEETS

Figure 8.1: The geothermal heat flux in Antarctica, the pink dots mark the locations of volcanoes. Elevated heat flux is found in a band along the East-West Antarctica boundary, along Siple Coast and in the area around Victoria Land, Oates Land and George V Land. From Fox Maule et al. [2005b].

to maintain basal melting. The difference between the direct measurement at Siple Dome and the estimated value underneath one of the ice streams is possibly due to local variations, which is beyond the resolution of our model.

Several subice lakes have been discovered at different locations in Antarctica; a review of subice lakes in Antarctica is given by Siegert [2000]. Siegert and Dowdeswell [1996] made theoretical studies of the thermal regimes of the lakes, among other things determining the necessary basal heat flux to maintain liquid water at the sites of lakes. Whether the temperature at the bottom of the ice is above freezing depends among other things on parameters as ice thickness, snow accumulation rate and ice movement, both vertical and horizontal. Ice is a good thermal insulator, thus the basal temperature tends to be higher under thick ice than under thin. A large vertical velocity will advect cold surface ice quickly to the bottom and thus cool the base of the ice, a large horizontal flow is however likely to heat the ice at the bottom due to deformational heating. Taking all these considerations into their models, Siegert and Dowdeswell [1996] calculated the basal heat flux necessary to maintain subice lakes at different locations on Antarctica. The basal heat consists of both heat supplied by the geothermal heat flux and heating caused by internal ice deformation and basal sliding. They found e.g. that a basal heat flux of 43 mW/m²
is necessary to sustain Lake Vostok, we find the geothermal heat flux in this area to be 55 mW/m$^2$, thus sufficient to maintain the lake. In general they find that a basal heat flux of 55 mW/m$^2$ is required to maintain subglacial lakes in the central parts of East Antarctica, whereas substantially higher basal heat fluxes are necessary to maintain lakes in e.g. George V Land and Oates Land. We find that the geothermal heat flux is about 55 mW/m$^2$ in the central parts of East Antarctica, and about 70–100 mW/m$^2$ in George V Land and Oates Land, which can explain much of the necessary basal heat flux needed to maintain subice lakes here.

Shapiro and Ritzwoller [2004] constructed a map of the geothermal heat flux in Antarctica by using seismic information of the crustal structure to extrapolate worldwide heat flux measurements. From a global seismic three-dimensional shear velocity model of the crust and upper mantle and global heat flux measurements they determined typical heat flux values for certain crustal structures from the places where heat flux measurements are available, also taking the large variability of heat flux for a given crustal structure into account. Then they assigned heat flux values for the given crustal structures to areas where no heat flux measurements are available. They found that the geothermal heat flux is almost three times larger and much more variable in West Antarctica than in East Antarctica. In East Antarctica they find the heat flux to vary between approximately 40–55 mW/m$^2$, with about 50 mW/m$^2$ in most of East Antarctica and somewhat higher values in Victoria Land of about 80 mW/m$^2$; their uncertainty is about 20 mW/m$^2$ in most of East Antarctica and about 50 mW/m$^2$ in Victoria Land. In West Antarctica they find values from 80 mW/m$^2$ to more than 125 mW/m$^2$, with the highest values in the central part of West Antarctica; the uncertainty on their heat flux in West Antarctica is from about 50 to 85 mW/m$^2$ with the highest uncertainty where the heat flux is highest. Our results and the results of Shapiro and Ritzwoller [2004] for the heat flux agree very well in East Antarctica, where we find similar values. In West Antarctica, our heat flux values have a slightly larger span than the ones of Shapiro and Ritzwoller [2004], and the area of highest heat flux is in our map somewhat further to the south (closer to the pole), but accounting for the uncertainty in both our and their heat flux results, our heat flux map is in agreement with the one by Shapiro and Ritzwoller [2004] also in West Antarctica.

### 8.2.2 Greenland

A map of the geothermal heat flux in Greenland is shown in figure 8.2. We find the geothermal heat flux to be low in the north-western part of Greenland, whereas we find very high values for the heat flux in east Greenland in and around King Christian IX Land and King Christian X Land.

There have been a few direct measurements of the geothermal heat flux in Greenland at coastal areas, and the geothermal heat flux underneath the ice has been estimated by various methods at a few more locations. Sass et al. [1972] measured the geothermal heat flux, as well as the thermal conductivity and the radioactive heat production of the surface rocks at two coastal sites in south Greenland less than 100 km apart. At one site, Ivigtut, they found the surface heat flux to be 42 mW/m$^2$, the thermal conductivity to be 2.8 W/mK, and the surface heat production to be $2.3 \times 10^{-6}$ W/m$^3$. At the other site, Ilímaussaq, they found a heat flux of
8.2. HEAT FLUX UNDERNEATH LARGE ICE SHEETS

Figure 8.2: The geothermal heat flux in Greenland. The pink stars mark the locations of known volcanoes. 1) the locations of the heat flux measurements by Sass et al. [1972], where they found 42 mW/m$^2$ at the northern location and 37 mW/m$^2$ at the southern. 2) the location of the GRIP ice core site, where Dahl-Jensen et al. [1998] found the heat flux to be 51 mW/m$^2$. 3) the location of the NGRIP site, where Dahl-Jensen et al. [2003a] estimated the heat flux to be 120 mW/m$^2$. 4) the site where Fahnestock et al. [2001] find large basal melting, which they attribute to locally high heat flux.
37 mW/m$^2$, a thermal conductivity of 2.3 W/mK, and a very high heat production rate of $9.7 \times 10^{-6}$ W/m$^3$ due to a highly radioactive, but thin intrusion present at the surface at this site. The study by Sass et al. [1972] shows how variable the radioactive surface heat production rate can be. The values they find for the geothermal heat flux in this area appears to be lower than the values we find here. However, the sites measured by Sass et al. [1972] (location 1 in figure 8.2) are on the border of our heat flux map, where our model is less likely to be accurate, in particular, our boundary value used for the bedrock surface temperature is not likely to be valid here as the sites are not covered by ice.

Dahl-Jensen et al. [1998] determined the geothermal heat flux at the site of the GRIP ice core (site 2 in figure 8.2) to be 51 mW/m$^2$ by inversion of the measured temperature profile through the ice cap. We find a slightly higher heat flux here of about 55 mW/m$^2$, but the discrepancy is within our level of uncertainty.

Only 324 km from the GRIP ice core site lies the NGRIP ice core drill site (location 3 in figure 8.2). From the knowledge that the basal ice is melting at the NGRIP drill site, Grinsted and Dahl-Jensen [2002] and Dahl-Jensen et al. [2003a] concluded that the geothermal heat flux must exceed 55 mW/m$^2$ here; based on the estimated amount of melting Grinsted and Dahl-Jensen [2002] estimated the geothermal heat flux here to be 98 mW/m$^2$, Dahl-Jensen et al. [2003a] later revised the estimate to 120 mW/m$^2$. By using radio-echo sounder observations along with information obtained from the ice cores at the GRIP and NGRIP drill sites Dahl-Jensen et al. [2003a] furthermore estimated that the heat flux along the ice flowline leading to the NGRIP drill site varies between 50 mW/m$^2$ and 200 mW/m$^2$ 100 km upstream of the NGRIP site. We find a heat flux of about 65 mW/m$^2$ in this area, which is less than the heat flux estimated by Dahl-Jensen et al. [2003a], however the variation of the heat flux they find is on a length-scale much shorter than our resolution.

Fahnestock et al. [2001] used airborne radar soundings to track internal layers in the Greenlandic ice sheet. From the internal layering they found the melt rate of the basal ice to vary greatly, and that the geothermal heat flux locally is in excess of 15 times the continental average. Particularly around $(74^\circ N, 40^\circ W)$ (site 4 in figure 8.2) they find a very high basal melt rate. They speculate, supported by gravity and bedrock topography data, that a volcano could be present under the ice at this location. We do not find a particularly high heat flux in this area, but again the feature reported by Fahnestock et al. [2001] is smaller than our resolution.

### 8.2.3 Error sources and uncertainty estimate

There are several sources of error that may influence the results for the geothermal heat flux. An important source of error is the uncertainty of the depth to the Curie isotherm, which we assumed to be the thickness of the magnetic crust. The uncertainty on the magnetic crustal thickness was discussed in the last chapter and estimated to be at the 29% level. Besides the error sources mentioned in chapter 7 (uncertainty of the field model, lateral variation of the magnetic susceptibility and separation of induced and remanent magnetization), uncertainties associated with the temperature boundary conditions, and lateral variations in the thermal conductivity contribute to the uncertainty of the geothermal heat flux.

The presence of the ice caps means that the bedrock surface temperature prob-
ably has not varied significantly with time in the last tens of thousands of years. The temperature at bedrock surface varies geographically from about 0°C to about −30°C underneath the ice in Antarctica [Huybrecths, 1991], the lowest values are only present locally. It is likely that the range of the bedrock surface temperature in Greenland is similar. We did not account for variation of the bedrock surface temperature, which we assumed to be 0°C everywhere. The maximum difference of 30 K corresponds to an error of about 5%, as the temperature difference used to derive the heat flux is $T_c - T_0 = 580$ K. The lowest basal ice temperatures in Antarctica are found beneath the Transantarctic Mountains, in areas where our model predicts high heat flux. A bedrock surface temperature lower than assumed would increase the resulting heat flux in these areas.

The uncertainties associated with the lower temperature boundary are comparable to that associated with the upper temperature boundary. The Curie temperature of the crustal rocks is dependent on the specific composition of the rock, particularly on the titanium content of the magnetites, and on pressure. Schlinger [1985] found the Curie temperature of rocks from the lower crust to vary between 550° and 580°. Thus, we estimate the uncertainty of the temperature of the lower boundary to be at the 5% level as well. Summarizing, we find the uncertainty due to the two assumed boundary temperatures, which are independent, to be 7% (of 65 mW/m²) or 5 mW/m².

In chapter 7 we concluded that the Curie isotherm may lie below Moho in the central parts of East Antarctica and possibly in north-west Greenland. If the mantle rocks are non-magnetic [Wasilewski and Mayhew, 1992, Frost and Shive, 1986], the lower magnetic boundary, and therefore the thickness of the magnetic crust, is determined by Moho in these areas. Thus the assumption that the magnetic crustal thickness gives the depth to the Curie isotherm is not correct in these areas. If the Curie isotherm lies below Moho, the temperature at the bottom of the magnetic crust will be lower than the Curie temperature. In this case the heat flux will be lower than our estimate, thus the heat flux we have determined in these areas may be considered an estimate of the maximum heat flux. However, we note that particularly in the central part of East Antarctica our heat flux estimate is in agreement with other estimates obtained for the heat flux in this region [Shapiro and Ritzwoller, 2004, Siegert and Dowdeswell, 1996].

It was assumed that it only was necessary to account for heat conduction in the vertical direction. Thus we are not taking possible lateral heat flow into account. Although lateral heat flow occurs due to e.g. variations in topography, we consider the amount of lateral heat flow to be negligible compared to the vertical heat flow.

When solving for the temperature profile of the crust we assumed steady state. It is reasonable to assume steady state or nearly steady state in stable cratonic regions, whereas areas where tectonic activity has recently occurred, may not be in thermal steady state. If not in steady state the temperature profile of the crust will be different than what we find, and so will the geothermal heat flux. As mentioned previously tectonic activity has occurred in recent geologic time in the West Antarctic rift zone, however several studies have concluded that the extension in the late Cenozoic has been very slow, and that the rift system currently is dormant [Behrendt, 1999, Ritzwoller et al., 2001, Winberry and Anandakrishnan, 2003], thus we consider the assumption of steady state to be reasonable and possible errors due
to this assumption to be negligible.

The thermal conductivity depends largely on temperature, pressure, quartz content, porosity, and fluid content of the rocks, and may in general be very variable in the upper crust. The conductivity decreases with increasing temperature, whereas it increases with increasing pressure. At depth, for temperatures above 300°C and pressures above 20 MPa (typically below 10 km depth), the variability of the thermal conductivity decreases significantly [Clauser and Huenges, 1995]. Significant lateral changes in the thermal conductivity might be expected if there are significant lateral changes in the quartz content of the entire crust. However, quartz is preferentially found in sedimentary sections in the uppermost crust, and secondarily, as quartzite in metamorphic rocks. Lower and mid-crustal rocks are typically mafic in composition, with relatively minor amounts of free quartz. If we use an amphibolite from the KTB drill hole [Clauser and Huenges, 1995] as representative of rocks of this type, 84 measurements of thermal conductivity gave a mean of 2.6 W/mK, with a standard deviation of 15%. Using this standard deviation as representative of the lateral variation of thermal conductivity suggests that variations at the 15% level or about 10 mW/m² in the heat flux might be expected as a consequence.

The combined error on the geothermal heat flux estimate from these independent, uncorrelated error sources is 37% or 24 mW/m². However, we believe, based on the comparison with other heat flux estimates given in section 8.2.1 and 8.2.2, that the uncertainty on the heat flux will be slightly lower in stable cratonic regions as e.g. the central part of East Antarctica, and slightly higher in regions that have experienced recent tectonic activity as e.g. West Antarctica.

8.3 Conclusions

We have developed a method to estimate the geothermal heat flux underneath large ice caps from field models of the crustal field based on satellite magnetic data. The result for the heat flux is shown in figure 8.1 and 8.2 for Antarctica and Greenland, respectively. We find the heat flux to vary between about 40 mW/m² and 185 mW/m²; the uncertainty is estimated to be about ±24 mW/m². We have compared the heat flux map with direct heat flux measurements, geologic evidence of heat flux anomalies, and heat flux estimates based on alternative data and methods. Our results are in coherence with the available direct measurements of the heat flux, within the level of uncertainty. We also find a good correlation of the geologic evidence of high heat flux with areas of high heat flux found in our model. All this indicates that this way of extracting the geothermal heat flux from satellite magnetic data is possible and useful.

We believe this method of deriving heat flux from satellite magnetic data will become even more favourable in the future, as the resolution of field modelling increases. Better field models can be expected due to more and better data, which e.g. will be obtained with the Swarm mission. An increased understanding of the structure and dynamics of the magnetic field will also lead to better field models that can explain more of the observed field, than what currently is the case. Finally, perhaps a better method of extracting the crustal field from the measured total field will be developed.
Chapter 9

Summary and conclusions

Two challenges were pursued in this PhD work. One was to look at uncertainty estimates of field models, in particular to apply variogram analysis to field model residuals to derive the along-track covariance function and from that an estimate of the data covariance matrix. The other was to use a field model to infer the geothermal heat flux underneath the large ice sheets in Greenland and Antarctica.

We have estimated the data covariance of a recent field model (Ørsted(06s/05)) directly from the scalar field model residuals using variograms. The field model is based on five years of quiet time, dark-side scalar and vector data from Ørsted, CHAMP, and SAC-C; a resampling rate of one minute was used. The analysis was focussed on the residuals in the geomagnetic colatitude interval from 40° to 140° in order to avoid corruption of the variogram by the difference in the variability of the polar and non-polar residuals. A long-wavelength trend was clearly visible in the individual satellite segments, which we interpret as originating from the magnetospheric ring current field. Although great effort was made in the particular field model to account for the field from the magnetospheric ring current, the applied corrections appear to be inadequate. As variogram analysis requires stationarity of the variable analysed, it was necessary to separate the residuals into two components: the trend component describing the long-wavelength variability of the residuals, and the corrected residuals, consisting of the remaining part of the residuals, which were assumed to be stationary. The covariances of the two parts of the residuals were analysed separately. The covariance of the trend was found from the expression derived for the trend, whereas the covariance of the corrected residuals was found from their spatiotemporal variograms. By adding the covariance of the trend to the along-track covariance of the corrected residuals the total along-track covariance was determined. From this the data covariance matrices of the Ørsted, CHAMP, and SAC-C residuals were made.

We found that serial correlation is present in the field model residuals, and that the data covariance matrices contain off-diagonal elements of the same order of magnitude as the diagonal elements. From this we conclude that in order to make a valid uncertainty estimate of field models it is necessary to include non-zero off-diagonal elements in the data covariance matrix, when calculating the variances and covariances of the Gauss coefficients. Alternatively the effect of serial correlation can be reduced significantly if a lower resampling rate of e.g. six minutes is used instead of the one minute resampling rate used in the analysed field model. We believe that most of the
serial correlation of the residuals is due to an insufficient description of the magnetospheric ring current, thus if a better description for the magnetospheric ring current is derived, serial correlation of field model residuals can be reduced significantly.

In spite of the difficulties that still are present in field modelling, field models have in general improved significantly over the past few years. Crustal field models are now becoming sufficiently detailed that they can be used for geophysical interpretation. We have used the MF3 crustal field model (and CM4) to estimate the geothermal heat flux underneath the large ice sheets in Greenland and Antarctica. The observed induced crustal field was obtained by subtracting a model of the remanent crustal field from the observed crustal field as given by MF3. By using the equivalent source magnetic dipole method, the thickness of the magnetic crust was determined from the observed induced field by iterative forward modelling. An initial model of the magnetic crustal thickness was iteratively improved until the induced field of the improved magnetic crustal thickness model matched the observed induced field. Since the observed crustal field is high-pass filtered to remove the core field it can only constrain the short-wavelength part of the magnetic crustal thickness variation. To provide the long-wavelength part of the solution of the magnetic crustal thickness, we used the igneous crustal thickness from the 3SMAC model. We found that the thickness of the magnetic crust in Greenland and Antarctica varies from only a few kilometres to more than 50 km. By comparing with seismic determined depths to Moho in Greenland, we concluded that the Curie isotherm lies shallower than Moho in most of Greenland. For Antarctica we compared the magnetic crustal thickness with crustal thickness estimates based on seismic and gravity studies and concluded that it is likely that the Curie isotherm lies shallower than Moho in West Antarctica and in the area of Victoria Land, Oates Land and George V Land, whereas it is possible that the Curie isotherm is deeper than Moho in the central part of East Antarctica.

To determine the heat flux underneath the large ice sheets a thermal model for the crust was derived. The one-dimensional heat conduction equation was used to determine the temperature profile of the crust, and from this the geothermal heat flux was determined. The boundary conditions used to solve the heat conduction equation are that the temperature at the bedrock surface is close to the melting point temperature of ice, and that the temperature at the bottom of the magnetic crust is the Curie temperature. Figure 9.1 provides a short overview of the method used to derive the geothermal heat flux from the crustal field.

We found that the geothermal heat flux varies between 40 mW/m² and 185 mW/m² in both Greenland and Antarctica. In Greenland we find high heat flux in the central part of east Greenland in and between King Christian IX Land and King Christian X Land. In Antarctica high heat flux is found in a band along the East-West Antarctica boundary, at Siple Coast, and in Victoria Land, Oates Land and George V Land. The resulting heat flux maps were compared with direct measurements of the heat flux and with indirect heat flux estimates inferred from other techniques and types of data. Comparison was also made with geologic evidence of high heat flux as e.g. the presence of volcanoes. In general we found good agreement of the results obtained for the heat flux in this study with both the direct observations of heat flux and the other heat flux proxy.

Thus we believe that this method of inferring the geothermal heat flux underneath
Figure 9.1: The observed induced magnetic field over Antarctica as obtained by subtracting a model of the remanent crustal field from the crustal field of the MF3 field model is shown in panel A. From this the thickness of the magnetic crust is found by iteratively improving a magnetic crustal thickness model; the initial model of the crustal thickness given by 3SMAC is shown in panel B. The resulting magnetic crustal thickness shown in panel D is then used to derive the geothermal heat flux shown in panel D. Figure from Fox Maule et al. [2005b].
the large ice caps from magnetic field models based on satellite magnetic data is possible, and that this method will become more useful in the future when better field models will be developed. Better field models can be expected when better satellite data are obtained from e.g. the Swarm mission. An increased understanding of the magnetic field in general, and in particular of the magnetospheric ring current field, can also improve field models, even before Swarm data become available.

All in all we conclude from this study that the potential of using magnetic field models for geophysical interpretation is increasing as field models improve. However, at the same time the need for a method to make reliable uncertainty estimates of field models grows.
Appendix A

Publications connected to this work

The following publications, posters and talks are a direct outcome of this PhD project. All talks mentioned below were given by the author.

Peer-reviewed papers

Peer-reviewed conference proceedings

Manuscripts in preparation
Fox Maule, C., N. Olsen & K. Mosegaard, Estimation of the covariance of magnetic field model residuals using variograms (working title), to be submitted to Geophysical Journal International.

Fox Maule, C., M. E. Purucker, N. Olsen & K. Mosegaard, The geothermal heat flux underneath the Greenland ice sheet estimated from satellite magnetic data (working title), to be submitted to Geophysical Research Letters.

Fox Maule, C., & K. Mosegaard, Using periodic boundary conditions of data to aid range estimation of variograms (working title), to be submitted to Mathematical Geology.

Posters presented at conferences
Fox Maule, C., K. Mosegaard, & N. Olsen, Satellite magnetic residuals investigated with geostatistical methods, EGU General Assembly, Vienna, April, 2005. This
poster was also presented at the IUPAP Second International Conference on Women in Physics, Rio de Janeiro, May, 2005, and at the Women in Physics in Denmark’s Annual Meeting, Nyborg, June, 2005.


Fox Maule, C., M. E. Purucker, N. Olsen, & K. Mosegaard, Magnetic crustal thickness in Greenland from CHAMP and Ørsted data, 2nd CHAMP Science Meeting, Potsdam, September, 2003.

**Talks presented at conferences**

Fox Maule, C., M. E. Purucker, N. Olsen, & K. Mosegaard, Geothermal heat flux underneath ice sheets estimated from magnetic satellite data, talk to be given at AGU Fall meeting, San Francisco, December, 2005.

Fox Maule, C., M. E. Purucker, N. Olsen, & K. Mosegaard, Heat flux in Antarctica revealed from satellite magnetic data, talk at 1st Workshop of Nordic Network for Women in Physics, Bergen, August, 2005.

Fox Maule, C., K. Mosegaard, & N. Olsen, Geostatistical methods applied to field model residuals, talk at the 10th Scientific Assembly of the International Association of Geomagnetism and Aeronomy, Toulouse, July, 2005.


Fox Maule, C., M. E. Purucker, & N. Olsen, Magnetic crustal thickness in Greenland from satellite magnetic data, talk at Geofysikdag, Copenhagen, October, 2003.

**Public outreach talks**

Ørstedsatelliten og varmefluxen i Antarktis (The Ørsted satellite and the heat flux in Antarctica), Senior Universitetet, to be given December 1st, Værløse Bibliotek, 2005.

Jordens magnetfelt, Ørstedsatelliten og varmefluxen i Antarktis (Earth’s magnetic field, the Ørsted satellite and the heat flux in Antarctica), to be given November 28th, Experimentarium, Hellerup, 2005.

The North Pole, the dipole and the multipole expansion, Danish Geophysical Society, December 9th, 2004. (Talk was given jointly by C. Fox Maule and N. Olsen)

Jordens magnetfelt (Earth’s magnetic field), Sankt Annæ Gymnasium, October 4th, 2004.

Det overordnede tema for dette ph.d.-projekt er brugen af magnetfeltsmodeller til geofysisk fortolkning; to centrale områder indenfor dette emne er behandlet nærmere: Formålet med det ene delprojekt er at bestemme datakovariansen af feltmodelsresidualer ved brug af variogrammer. Kendes datakovariansmatricen kan usikkerheden på en feltmodel beregnes. Formålet med det andet delprojekt er at bestemme den geoterme varmeflux under iskapperne i Grønland og Antarktis. Direkte målinger af varmefluxen i disse isdækkede områder er både besværlige og kostbare; derfor er indirekte metoder til at bestemme varmefluxen særlig fordelagtige.

Vi har ved brug af variogrammer estimeret datakovariansen af Ørsted(06s/05)-feltmodellen direkte fra feltmodellens skalarresidualer, som er forskellen mellem de af feltmodellen forudsagte værdier og de faktiske data. For at undgå korruption af variogrammet på grund af forskellig variabilitet af polare og ikke-polare data blev analysen foretaget på data liggende mellem 50°S og 50°N geomagnetisk bredde. Studeres de enkelte datasegmenter nærmere, ses tydeligt en langbølget variation af residualerne. For at kunne analysere en parameter ved hjælp af et variogram, er det nødvendigt at parameteren er stationær. Det var derfor nødvendigt at opdele residualerne i to dele: en trendkomponent og de korrigerede residualer, der findes ved at trække trenden fra residualerne. Trendkomponenten beskriver den langbølgede variation af residualerne, mens de korrigerede residualer beskriver den højfrekvente fluktuation af residualerne omkring trenden. De korrigerede residualer antages at være stationære. For at bestemme kovariansen af residualerne var det nødvendigt
at behandle de to dele hver for sig. Kovariansen af trenden blev estimeret fra et
deterministiske udtryk for trenden, mens variogrammet blev brugt til at estimere
kovariansen af de korrigerede residualer. Summen af de to kovarianser gav den to-
tale along-track kovariansfunktion, hvorfra datakovariansmatricerne for hhv. Ørsted-,
CHAMP- og SAC-C-residualerne blev bestemt.

Vores studie viser, at der er seriel korrelation af residualerne, og at datakovari-
ansmatricerne indeholder ikke-diagonal elementer af samme størrelsesorden som di-
agonalelementerne. Deraf konkluderer vi, at det er nødvendigt at inkludere ikke-
diagonalelementer forskellige fra nul i datakovariansmatricen for at opnå et pålideligt
usikkerhedsestimat af feltmodellen. Alternativt kan man mindske den serielle kor-
relation betydeligt ved at mindske resampling-raten fra hvert til f.eks. hvert sjette
minut. Vi tror, at langt størstedelen af den serielle korrelation af residualerne skyldes,
at der er et betydeligt bidrag fra den magnetosfæriske ringstrøm, som feltmodellen
ikke beskriver. Vi mener derfor, at den serielle korrelation kan reducieres ganske be-
tydeligt, hvis en bedre forståelse for og beskrivelse af den magnetosfæriske ringstrøm
udvikles og inkluderes i magnetfeltmodeller.

På trods af de problemer der stadig findes indenfor feltmodellering, er kvaliteten
af magnetfeltmodeller forøget betydeligt i de senere år. Detaljéringsgraden af skor-
pelfeltsmodeller er nu ved at nå et niveau, hvor de kan bruges til geofysisk fortolkning.
Vi har brugt skorpelfeltsmodellen MF3 til at bestemme varmefluxen under iskapperne
på Antarktis og i Grønland. Dette har vi gjort ved først at bestemme tykkelsen af den
magnetiske skorpe, der er den del af skorpen, der ligger mellem grundfjeldets over-
flade og enten Curie isolteren eller Moho, afhængig af hvilken af de to, der ligger
tættest på overfladen. Først blev en model for det remanente skorpelfelt trukket fra
det observerede skorpelfelt givet ved MF3 for at finde det observerede inducerede skor-
pelfelt. Ved brug af den ækvivalente dipolmetode blev den magnetiske skorpetykkelse
derefter bestemt udfra det observerede inducerede skorpelfelt ved iterativt at forbedre
en begyndelsesmodel for den magnetiske skorpetykkelse indtil det inducerede skorpe-
felt af den resulterende skorpetykkelsesmodel svarede til det observerede. Fordi det
observerede skorpelfelt (MF3) er højpasfilteret, kan det kun bruges til at bestemme
den kortbølgede del af variationen af den magnetiske skorpetykkelse. Den langbølgede
del af skorpetykkelsesvariationen er taget fra en seismisk bestemt skorpetykkelses-
model, 3SMAC. Vi finder, at den magnetiske skorpetykkelse både i Grønland og på
Antarktis varierer fra nogle få kilometer til over 50 km. Ved sammenligning med
seismisk bestemte Mohodybder i Grønland, konkluderer vi at Curie-isolteren ligger
i lavere dybde end Moho i størstedelen af Grønland. I Antarktis har vi sammenlignet
de opnåede magnetiske skorpetykker med skorpetykker bestemt fra seismiske
studier og tyngdestudier. Fra dette konkluderer vi, at Curie-isolteren ligger i en
lavere dybde end Moho i Vestantarktis, mens dette muligvis ikke er tilfældet i store
dele af Østantarktis.

Til at bestemme den geoterme varmeflux blev en termisk model for skorpen op-
stillet. Den én-dimensionale varmeledningsligning blev brugt til at bestemme temper-
aturprofilen af den magnetiske skorpe hvorfra varmefluxen blev bestemt. De grænse-
betingelser, der er brugt til at løse varmeledningsligningen, er, at temperaturen ved
grundfjeldets overflade er smeltepunktstemperaturen af is, og at temperaturen ved
bunden af den magnetiske skorpe er Curie-temperaturen af magnetit.

Vi finder, at den geoterme varmeflux varierer mellem 40 mW/m² og 185 mW/m²

Alt i alt mener vi, at den anvendte metode til bestemmelse af den geoterme varmeflux under de store iskapper udfra en magnetfeltsmodel baseret på magnetiske satellitdata er anvendelig og giver fornuftige resultater. Vi er af den overbevisning, at metoden vil blive endnu mere fordelagtig i fremtiden, når magnetfeltsmodeller bliver bedre. Dette kan bl.a. forventes at ske, når Swarm missionen bliver sendt op i 2009, idet vi da kan forvente at få endnu bedre satellitdata til rådighed. Feltmodellerne kan dog forbedres allerede inden nye data bliver tilgængelige, i takt med at forståelsen for magnetfeltet generelt og i særlighed feltet fra den magnetosfæriske ringstrøm forøges.

Samlet set konkluderer vi på baggrund af dette projekt, at potentialet for at bruge magnetfeltsmodeller til geofysisk forståelse udvikles og at det forøges i takt med at feltmodellerne bliver bedre, men samtidig bliver behovet for en metode til at opnå pålidelige usikkerhedsestimater af feltmodeller større.
Bibliography


M. Purucker and K. Whaler. Merging satellite and aeromagnetic data over Europe, the north Atlantic, and the arctic. Poster presented at the second CHAMP meeting, Potsdam, Germany, 2003.


