CARDIFF UNIVERSITY

PRIFYSGOL CAERDYDD

GEOPHYSICAL INVERSE THEORY

APPLIED TO RECONSTRUCTION

OF LARGE-SCALE HETEROGENEITIES

IN ELECTRICAL CONDUCTIVITY

OF EARTH'S MANTLE

by

Anna Kelbert

UMI Number: U211200

Dissertation Publishing

UMI U211200

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

In memory of Dmitry Khmelev

Abstract

This thesis addresses the non-linear ill-posed inverse problem of reconstructing the three-dimensional distribution of electrical conductivity in Earth's mantle. This problem has never previously been fully attacked. The major objective of this thesis is to develop a methodology allowing the resolution of any large-scale three-dimensional inhomogeneities in Earth's mantle based on a regularised inversion of electromagnetic field data.

We generalise the global three-dimensional forward problem of electromagnetic induction in the frequency domain for arbitrary sources, and solve it in a linear algebraic formulation. We develop data sensitivities analysis based on the generalised forward problem, and derive the analytic and numerical expressions for the Jacobian and the derivative of the regularised least squares penalty functional. This allows us to set up a non-linear conjugate gradient inverse solution. In doing so, we parametrize the model space by layered spherical harmonics. This inverse solution is tested on a series of checkerboard experiments; on this basis, we discuss spatial resolution of our analysis at different depths in the mantle.

This methodology is then applied to the most current low-frequency global observatory data set, and models are obtained satisfying the data statistically well. We discuss the features of these models and the implications of our experiments. We also plot and compare the corresponding magnetic fields and responses at the Earth's surface, and provide suggestions for future directions of research.

Acknowledgements

Contents

List of Notation

H-prisms	Primary prisms of the staggered-grid finite difference formulation
H-edges	Edges of *H-prisms* (pass through mid-points of *E-faces*)
H-faces	Faces of *H-prisms*
E-prisms	Staggered prisms of the staggered-grid finite difference formulation
E-edges	Edges of *E-prisms* (pass through mid-points of *H-faces*)
E-faces	Faces of *E-prisms*
l, l^E	Length elements defined on *H-edges* and *E-edges*, respectively
S, S^E	Surface area elements defined on *H-faces* and *E-faces*, respectively
$\mathbf{E}$	Vector space of complex components defined on *H-edges*
$\mathbf{E}_i \subset \mathbf{E}$	Subspace of complex components defined on interior *H-edges* of the domain
$\mathbf{E}_b \subset \mathbf{E}$	Subspace of complex components defined on boundary *H-edges*
$\mathbf{F}$	Vector space of complex components defined on *H-faces* (equivalently, on *E-edges*)
$\mathbf{G}$	Vector space of real components defined in *H-prisms*
i, b	Subscripts; denote restriction to either interior or boundary components
$\mathcal{D} \subset \mathbb{C}^M$	Admissible data space
$\mathcal{M} \subset \mathbb{R}^N$	Subspace of acceptable models (models, fitting the data)
K	Number of layers in a model
L	Degree of the spherical harmonic expansion

Vectors and non-linear mappings

$\mathbf{m} \in \mathbb{R}^N$	Vector of model parameters
$\mathbf{m}_0 \in \mathbb{R}^N$	Prior model
$\mathbf{d} \in \mathbb{C}^M$	Complex data vector
$\tilde{\mathbf{d}} \in \mathbb{C}^M$	Complex error-free data vector
$\boldsymbol{\epsilon} \in \mathbb{R}^M$	Real uniformly distributed errors
$\boldsymbol{\delta} \in \mathbb{R}^M$	Estimated errors
$\rho_{face} \in \text{Re}\,(\mathbf{F})$	Electrical resistivity mapped to the H-faces
$\rho \in \mathbf{G}$	Electrical resistivity on the grid
$\sigma \in \mathbf{G}$	Electrical conductivity on the grid, $1/\rho$
$\mathbf{h} \in \mathbf{E}$	Vector of magnetic field components on the grid
$\mathbf{h}_i \in \mathbf{E}_i$	Magnetic field components strictly inside the computational domain
$\mathbf{h}_b \in \mathbf{E}_b$	Magnetic field components on the boundary of the computational domain
$\mathbf{s} \in \mathbf{E}$	Vector comprising all sources
$\mathbf{f} \in \mathbf{E}_i$	The forcing on the interior of the computational domain
$\mathbf{b} \in \mathbf{E}_b$	The boundary conditions
$\tilde{\mathbf{s}} \in \mathbf{E}$	The source vector for the system $M_{\rho,\omega}\delta\mathbf{h} = \tilde{\mathbf{s}}$
$\mathbf{e} \in \mathbf{F}$	Vector of electric field components on the grid
α, β, μ	Regularisation parameters
$c_{lm}^{(k)}$	Spherical harmonic coefficients, $l = 1, \ldots L$, $m = -l, \ldots l$ and $k = 1, \ldots K$
$\mathbf{r}_\omega \in \mathbb{C}^{M_\omega}$	Vector of residuals with data, $\psi_\omega(\mathbf{m}) - \mathbf{d}$
$\tilde{\mathbf{r}}_\omega \in \mathbb{C}^{M_\omega}$	Vector of weighted residuals, $C_\epsilon^{-1}(\psi_\omega(\mathbf{m}) - \mathbf{d})$
$\mathbf{g}_j \in \mathbf{E}$	Linearised data functional for j'th observatory
$R(\mathbf{m}) \in \mathbb{R}$	The value of the penalty functional
$F : \mathbb{R}^N \to \mathbb{R}^M$	A general forward problem with real and exact data
$\psi_\omega : \mathbb{R}^N \to \mathbb{C}^{M_\omega}$	Non-linear data functionals: $\psi_j = (\psi_\omega(\mathbf{m}))_j$
$\hat{\psi}_\omega \in \mathbb{C}^{M_\omega}$	Linearised data functionals: $\hat{\psi}_j = \mathbf{g}_j^* \mathbf{h}$
$\lambda : \mathbf{E} \to \mathbb{C}^{M_\omega}$	Transfer functions: $\psi_j = \lambda_j(\mathbf{h})$
$\gamma_{\omega,s} : \mathbf{G} \to \mathbf{E}$	The forward mapping: $\mathbf{h} = \gamma_{\omega,s}(\rho)$
$\eta : \mathbb{R}^N \to \mathbf{G}$	Parametrization mapping: $\rho = \eta(\mathbf{m})$

Operators

D_{lE} : $\mathbf{E} \to \mathbf{E}$ Pre-multiplies by edge length elements on H-*edges*

D_{lF} : $\mathbf{F} \to \mathbf{F}$ Pre-multiplies by edge length elements on E-*edges*

D_{SF} : $\mathbf{F} \to \mathbf{F}$ Pre-multiplies by surface area elements on H-*faces*

D_{SE} : $\mathbf{E} \to \mathbf{E}$ Pre-multiplies by perpendicular surface area elements on H-*edges*

C : $\mathbf{E} \to \mathbf{F}$ Implements the curl operator $\nabla \times$ on H-*edges*

C^T : $\mathbf{F} \to \mathbf{E}$ Transpose of C. Implements the curl operator $\nabla \times$ on E-*edges*

L : $\mathbf{G} \to \mathrm{Re}\ (\mathbf{F})$ Averages real quantities from adjacent cells to their common face

$\tilde{L}$: $\mathbf{G} \to \mathrm{Re}\ (\mathbf{F})$ Defined as $D_{SF}^{-1} D_{lF} L$

$D_{L\rho}$: $\mathbf{F} \to \mathbf{F}$ Pre-multiplies by ρ on H-*faces*

$D_{\rho F}$: $\mathbf{F} \to \mathbf{F}$ Pre-multiplies by the weighted averages of ρ on H-*faces*

D_e : $\mathbf{F} \to \mathbf{F}$ Pre-multiplies by electric field components $\mathbf{e} = CD_{lE}\mathbf{h}$

I_{-b} : $\mathbf{E} \to \mathbf{E}$ Operator that nullifies boundary components on H-*edges*

I_{-i} : $\mathbf{E} \to \mathbf{E}$ Operator that nullifies interior components on H-*edges*

D_l : $\mathbf{E}_i \to \mathbf{E}_i$ Pre-multiplies by length elements. Acts on the interior components only. Shorthand notation for D_{lE_i}

D_S : $\mathbf{E}_i \to \mathbf{E}_i$ Pre-multiplies by perpendicular surface area elements. Shorthand notation for D_{SE_i}

B_ρ : $\mathbf{E}_b \to \mathbf{E}_i$ Implements the interior part of the operator $\nabla \times (\rho \nabla \times \cdot)$ acting on boundary components only

$A_{\rho,\omega}$: $\mathbf{E}_i \to \mathbf{E}_i$ The symmetric kernel of the forward solver

$M_{\rho,\omega}$: $\mathbf{E} \to \mathbf{E}$ The full generalised forward solver

E_h : $\mathbf{G} \to \mathbf{E}$ Defined as $D_{SE}^{-1} I_{-b} C^T D_e \tilde{L}$

J : $\mathbf{G} \to \mathbf{E}$ The Jacobian operator; computed as $-M_{\rho,\omega}^{-1} E_h$

L_j : $\mathbf{E} \to \mathbb{C}$ Sparse operator that maps the magnetic field vector to a point

G_ω : $\mathbb{C}^{M_\omega} \to \mathbf{E}$ Consists of $\mathbf{g}_j$ as columns

P : $\mathbb{R}^N \to \mathbf{G}$ The linearized parametrization mapping

$C_\mathbf{d}$: $\mathbb{C}^M \to \mathbb{C}^M$ The covariance operator in the data space (here taken to be D_{ϵ^2})

$C_\mathbf{m}$: $\mathbb{R}^N \to \mathbb{R}^N$ The covariance operator in the model space

C_H, C_V : $\mathbb{R}^N \to \mathbb{R}^N$ Horizontal and vertical "smoothing" operators

List of Figures

List of Software

Program	Description	Language	Copyright
earth3d	Modular global 3-D forward solver	Fortran 90	M. Uyeshima, A. Schultz, H. Toh and A. Kelbert, 1995 – 2006
inversion3d	Inverse solver for parallel and serial computations	C++	D. Khmelev and A. Kelbert, 2003 – 2006
GSD	Set of scripts to generate synthetic data sets, and more	Perl	A. Kelbert, 2004 – 2006
SHTOOLS	Spherical harmonics toolbox	Matlab	A. Kelbert, 2006
plot	Library of routines to present the output of *earth3d*	Matlab	A. Kelbert, 2006

Chapter 1

Introduction

Energy transmission through wave propagation and through diffusion is our only means at present to sense the structure of the Earth's deep interior. Both seismic waves and potential fields (gravitational and magnetic) provide constraints on deep interior. Although seismology provides us with the information about the density and elastic properties of Earth materials, electromagnetic (EM) induction and electrical resistance methods are the only means to study the electrical conductivity – a vitally important parameter that depends on the physical state of the Earth's interior. Conductivity is directly influenced by the composition and presence of melt, water and volatiles in the deep interior, and is more weakly dependent on the temperature and pressure. The aim of this Introduction is to give an overview of the current understanding of the Earth's interior as well as the Earth's electromagnetic fields and of the methodology developed for reconstructing the electrical conductivity distribution from the EM field observations taken on or above the surface of the Earth. We then give an overview of the contributed research and describe the format of this thesis.

1.1 General structure of Earth's interior

The primary internal structure of the Earth is defined by its main layers, and has been in the past determined on the global scale mainly from seismology. Any change in stress – either natural or man-made (such as an earthquake or an explosion), if sufficiently impulsive, leads to a wave that propagates from the point where the deformation occurs. These are known as *seismic waves*. There are different types of deformations that, depending on their nature either result in a wave propagation confined to the surface (Rayleigh waves and Love waves) or travel through the interior of the Earth (P-waves and S-waves). The latter are known as the *body waves*. The speed of the waves depends on various elastic properties of the material (elastic moduli) and on density. The arrival of seismic waves to various point on the surface can be detected, and travel times measured (alternatively, full seismic waveforms can be modelled). This information is used to determine seismic wave velocities in the interior. Hence, some constraints on density and elastic properties of materials can be set. It has also been shown, that seismic waves in the Earth experience some reflections and refractions, which has served to identify and locate major compositional boundaries inside the Earth (Figure 1.1).

Seismic sounding leads us to believe [e.g. *Dziewonski and Anderson*, 1981; *Masters and Shearer*, 1995] that the Earth has at its centre a solid inner core, and a dense liquid core surrounding it. The outer core is about 1/2 the radius of the Earth. It is widely believed that the inner core is made up of molten iron, perhaps mixed with nickel and sulfur [*Jeanloz*, 1990; *Allègre et al.*, 1995; *Stixrude et al.*, 1997]. The density seems appropriate, and iron, which among all elements has the most stable nucleus, is abundant in the universe; although the most convincing evidence is that coming from the composition of meteorites and from solar spectra [e.g. *Palme and O'Neill*, 2003]. The outer core is inferred to be liquid because it transmits compressional seismic waves, but not shear waves [*Davies*, 2001]. The vigorous convection in the outer core (Section 1.2.2) is dominated
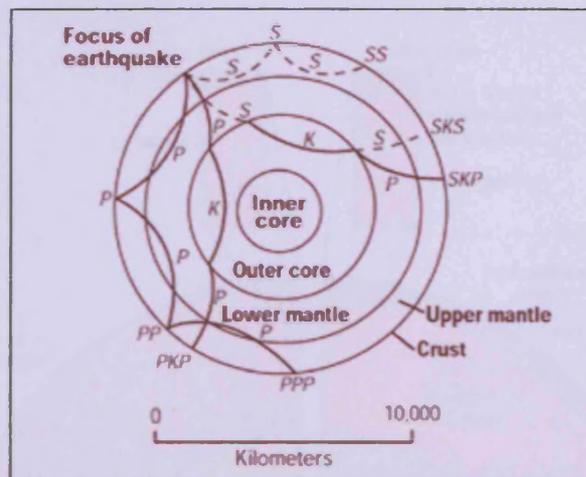
Figure 1.1: Cross section of the Earth, showing the complexity of paths of earthquake waves. The paths curve because the different rock types found at different depths change the speed at which the waves travel. Solid lines marked P are compressional waves; dashed lines marked S are shear waves. S waves do not travel through the liquid outer-core but may be converted to compressional waves (marked K) on entering the core (PKP, SKS). Waves may be reflected at the surface (PP, PPP, SS). *Reproduced from "The interior of the Earth" by Eugene C. Robertson*

by the rotation effects [e.g. *Olson and Hagee*, 1990], and sustained by radioactive heating and chemical differentiation.

The mantle is a broad layer of iron-magnesium-aluminosilicates surrounding the outer core. It extends to a depth of about 2900 km and accounts for around 82% of the Earth's volume. The mantle is mostly *plastic* – meaning a substance with the properties of a solid that however flows under stress, apart from the uppermost layer which is rigid, and determines the strength of the continental and oceanic plates.

The uppermost seismically defined layer of the Earth is the crust, which has been inferred to have a different composition from the substratum. The jump in seismic velocities at a depth of about 60km was originally identified by Mohorovičić in 1909, and later in many continental areas at a depth of 35-40km. This is known as *the Moho* boundary. The *lithosphere* is the rigid layer comprising the crust and the dense upper mantle. The *asthenosphere* refers to the plastic upper mantle.
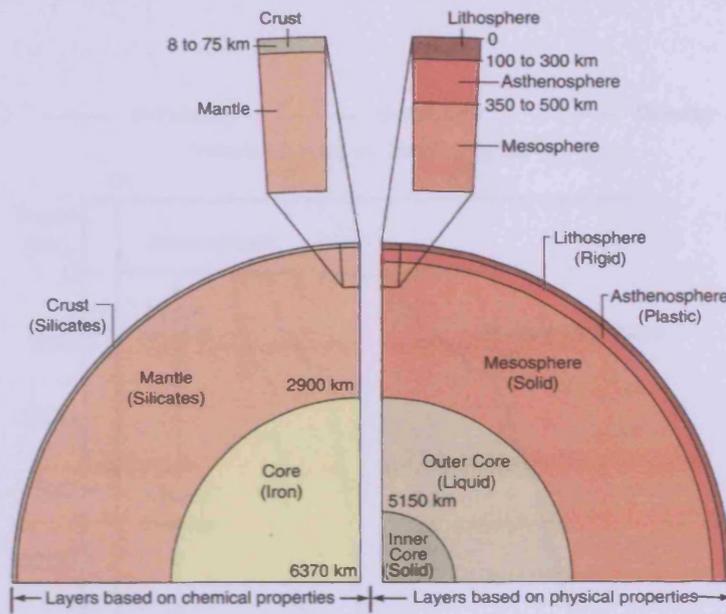
Figure 1.2: General chemical and physical structure of Earth's interior. *Taken from the Earth Sciences course by Prof Joe Licciardi, University of New Hampshire*

Under the assumption of a spherically symmetric Earth, one-dimensional profiles of seismic velocities have been developed. A classical seismic velocity profile is presented on Figure 1.3. It clearly indicates the main seismic discontinuities, that are now well-established compositional boundaries in the Earth's interior. The depth of the core-mantle boundary (CMB) is inferred to be 2,889km, and the inner core is defined by the jumps of seismic velocities near 5,150km depth [*Dziewonski and Anderson*, 1981; *Young and Lay*, 1987].

We often divide the mantle into two regions, upper and lower, based on the level of velocity heterogeneity. The region from near 410 to 670 km depth is called the *transition zone* and strongly affects body waves that "turn" at this depth and arrive about $20° - 30°$ distant from a shallow earthquake. In this depth range the minerals that make up the mantle silicate rocks are transformed by the increasing pressure. The atoms in these rocks rearrange themselves into compact structures that are stable at the high pressures and
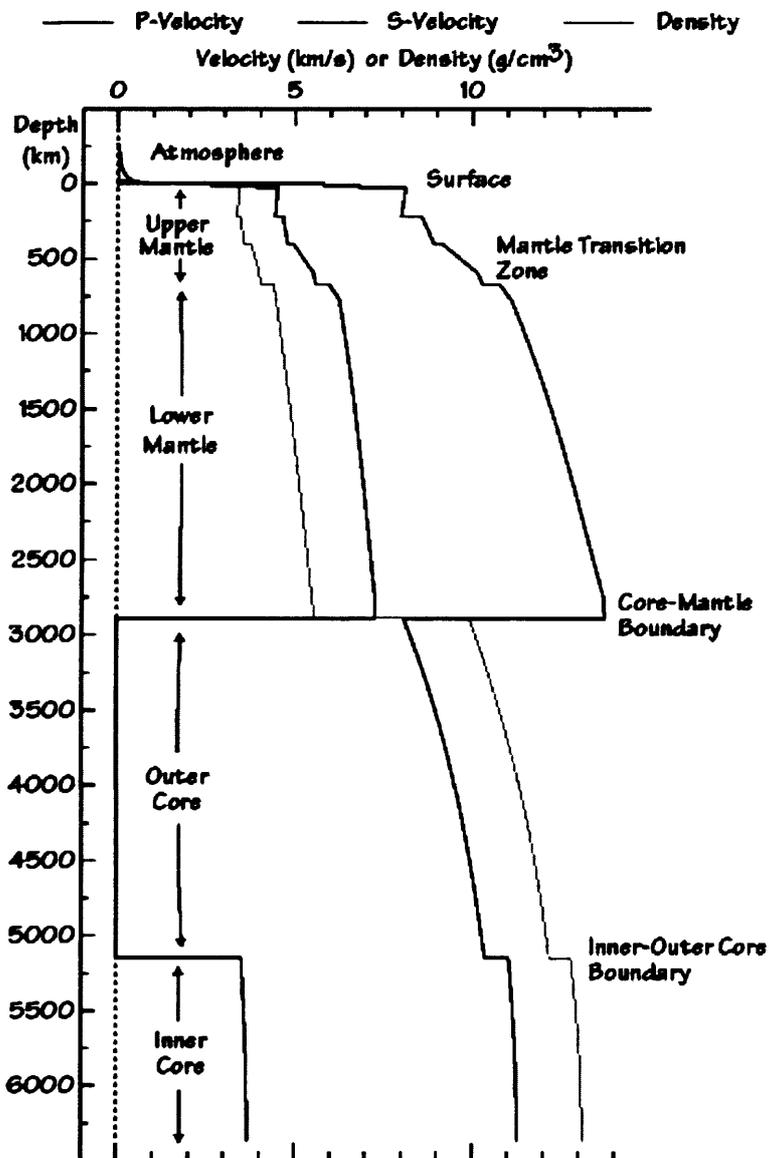
Figure 1.3: One-dimensional estimate of velocity and density variations within Earth based on seismic observations. The main regions of Earth and important boundaries are labeled. This model was developed in the early 1980's and is called PREM for Preliminary Earth Reference Model (*Dziewonski and Anderson* [1981])

the result of the rearrangement is an increase in density and elastic moduli, producing an overall increase in wave speed.

There is considerable structure in the mantle, that is indirectly resolved by seismic and geomagnetic sounding inverse techniques (Section 1.5). This is an active area of research at present, and no unique 3-D compositional model exists as yet. Please refer to section 1.4 for a more detailed discussion of the controversy related to the composition of the transition zone. Parameters such as density, electrical conductivity, temperature and pressure, presence of melts and volatiles, elastic moduli, elastic and electrical anisotropy, all form a single dynamic system in the mantle, and the aim of many studies is to infer the general structure and composition, based on what can be understood about each of the parameters by indirect measurements.

In this work, we concentrate on the geoelectromagnetic sounding techniques, that are concerned with estimating the electrical conductivity of the deep interior of the Earth using the electromagnetic (EM) fields measured at the surface.

1.2   Modelling Earth's magnetic field

The first known effort to study Earth's magnetic field was made by *Gilbert* [1600], who has shown that the Earth's magnetic field is similar to the field of a magnetic dipole (such as that of a bar magnet). It was later shown, that it is in fact similar to the magnetic field of a spherical magnet, with the principle axis inclined approximately 11° to the Earth's axis of rotation (Figure 1.4). In the nineteenth century, *Gauss* [1838] devised a method for measuring the strength of the field. He was the first to write a spherical harmonic expansion for the Earth's magnetic field, which immediately allowed him to obtain many fundamental results. He unambiguously decomposed the geomagnetic field into its internal and external parts, the sources of which lie inside and outside the body

of the Earth, respectively.

1.2.1 Spherical representation of the Earth's magnetic field

The Earth's atmosphere, crust and mantle may be treated as electrical insulators at very long periods of decades and longer which are of interest in mapping the main field. In this case there are no electric currents and the magnetic field **B** may be written as the gradient of a magnetic potential V (also known as the magnetic field strength),

$$\mathbf{B} = \nabla V. \tag{1.1}$$

In this case **B** is divergence free, and V satisfies Laplace's equation,

$$\nabla \cdot \mathbf{B} = \nabla^2 V = 0, \tag{1.2}$$

which can be solved everywhere in the Earth/air domain, provided reasonable boundary conditions (such as the value of V or the normal component of the magnetic field) at the top of the domain are specified.

A spherical harmonic decomposition is unambiguously defined for any solution of Laplace's equation in spherical curvilinear coordinates, and may be computed by separation of variables. Thus, V may be written as

$$V = a \sum_{l=1}^{\infty} \left[\left(\frac{r}{a}\right)^l S_l^e + \left(\frac{a}{r}\right)^{l+1} S_l^i \right], \tag{1.3}$$

or, equivalently, as

$$V = V_e + V_i,$$
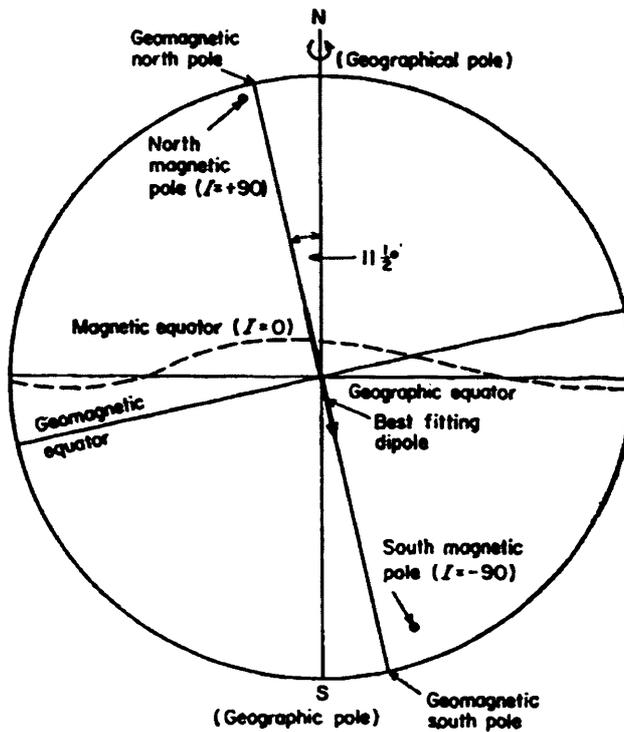
Figure 1.4: Graphical presentation of the magnetic, geomagnetic and geographic poles and equators, reproduced from *McElhinny* [1973]. The *magnetic poles* are defined as the points where the magnetic field is measured to be vertical. The *geomagnetic poles* (dipole poles) are the pole positions based on the first three terms of the International Geomagnetic Reference Field (IGRF), a model of the Earth's main magnetic field (Section 1.2.1). These positions are used to fix the geomagnetic coordinate system. Because of non-dipole component, magnetic poles do not coincide with geomagnetic poles.

where a is the radius of the Earth in metres, r is the distance from the centre of the Earth, which is also the origin in spherical coordinates, and S_l^e and S_l^i are two spherical harmonic expansions of order l. Expression 1.3 holds in any source-free region. The external V_e and internal V_i parts correspond to fields due to sources at larger and smaller radii, respectively, than the radius of the Earth. The part V_e, that consists of the terms in r^l, tends to zero as we move into the Earth, thus corresponding to an external current source. On the other hand, V_i, that consists of terms r^{-l}, tends to zero as we move away from the Earth's surface, thus corresponding to a source located inside the Earth. Expansions V_e and V_i are known as the external and internal source terms of the potential function.

It has been shown by spherical harmonic analysis of the observed magnetic field that at least 99% of fields at low frequencies (less than $\approx$ 10 Hz) originated inside the Earth (in Section 1.2.4 we describe the range of external sources responsible for the remaining 1% of the magnetic field). As we've learnt since, most of the internal magnetic field originates from the Earth's core, except for the background due to crustal magnetisation. The exact sources of this field are unknown; however, a plausible hypothesis exists, known as the *Earth's dynamo hypothesis* (Section 1.2.2). The possible sources of the external magnetic field are discussed in Section 1.2.4.

The internal geomagnetic field of the Earth, just like its gravitational field, can be represented in terms of its potential $V_i(r, \theta, \phi)$, decomposed onto spherical functions:

$$V_i(r, \theta, \phi) = a \sum_{l=1}^{\infty} \sum_{m=0}^{l} \left(\frac{a}{r}\right)^{l+1} [g_l^m \cos(m\phi) + h_l^m \sin(m\phi)] \, P_l^m(\cos \theta), \qquad (1.4)$$

where θ is the co-latitude (i.e. $90° -$ latitude), ϕ is the longitude to the East and P_l^m are the Schmidt semi-normalised associated Legendre functions of degree l and order m. Expansion 1.4 is valid outside the Earth. The real values g_l^m and h_l^m are the *geomag-*

netic coefficients, also known as Gauss coefficients. They have the units of the magnetic field strength, which is *Gauss* in the CGS unit system, however in geophysics a more convenient unit 1 nT = 10^{-5} Gauss is often used. Gauss coefficients (in nT's), and their secular variations $\dot{g}_l^m$ and $\dot{h}_l^m$ are recorded in the International Geomagnetic Reference Field (IGRF) model annually since 1965. The geomagnetic coefficients are functions of time and for the IGRF the change is assumed to be linear over five-year intervals. The number of known coefficients determines the accuracy of our knowledge about the Earth's main magnetic field. Currently, the main field expansion is recorded up to the degree and order 13^1. A new model of the Earth's magnetic field derived from magnetic satellite data has been recently offered by *Olsen et al.* [2006].

The first few coefficients of the Expansion 1.4 can be physically interpreted. Let us imagine the magnetic dipole **M** situated at the Earth's centre, and directed along the polar axis z. At an arbitrary point, the magnetic potential due to the vector **M** is $V = \mathbf{M} \cdot \mathbf{r}/r^3$.

Thus, the main magnetic field of the Earth may be uniquely defined as a model vector, which is simply the array of geomagnetic coefficients arranged in a suitable order:

$$\mathbf{m} = \{g_1^0, g_1^1, h_1^1, g_2^0, g_2^1, h_2^1, g_2^2, h_2^2 \ldots\}, \tag{1.5}$$

and the directional components of main Earth's magnetic field at any point on or above the Earth's surface are the partial derivatives of V with respect to r, θ or ϕ at $r = r_0 \geq a$ (where a is the Earth's radius) and the given co-latitudinal and longitudinal coordinates, respectively.

---

[1] In fact, the total internal field is estimated to a much higher degree of at least 60. However, separation between crustal and core contributions is not possible beyond about 12th or 13th degree. The core dominates the lower degrees, while the crust dominates highest modes, and they are generally interpreted as crustal magnetization. Degrees around 12–13 are ambiguous, and there are certainly longer wavelength crustal components which are poorly known, because they cannot at present be distinguished from the core contribution.

The internal field of the Earth is dominated by its dipole component. The Earth's multipole magnetic field is also significant, although the dipole field of the Earth exceeds the intensity of the multipole components by an order of magnitude.

Similarly, the external part of the Earth's magnetic field may be expressed as the gradient of the external vector potential

$$V_e(r, \theta, \phi) = a \sum_{l=1}^{\infty} \sum_{m=0}^{l} \left(\frac{r}{a}\right)^l \left[g_l^m \cos(m\phi) + h_l^m \sin(m\phi)\right] P_l^m(\cos\theta), \qquad (1.6)$$

for some other set of coefficients g_l^m and h_l^m, distinct from those in Equation 1.5.

1.2.2   Origin of the Earth's main magnetic field

The dynamic nature of the Earth's magnetic field suggests a source in the fluid, mobile outer core. It is now generally agreed, that the magnetic field is generated in the outer core by electric current loops, in turn powered by thermal convection of the molten material, forming a self-sustaining *geomagnetic dynamo*.

The idea of a hydromagnetic dynamo was first proposed by Sir Joseph Larmor in England in 1919 as an explanation of the creation of sunspot magnetic fields. It wasn't utilised in geophysics until Ya. I. Frenkel in the USSR and V. Elsasser in the USA proposed in the mid-1940's, that the convection of heat in the Earth's core may be the reason that sets in motion the dynamo in the Earth's core.

Attempts to model fluid dynamos were discouraged by the "anti-dynamo" theorem of Thomas G. Cowling (1906-1990), who proved that the fluid dynamo problem had no axially symmetric solutions. S. I. Braginsky in 1964 constructed the first kinematic models of the Earth's dynamo.

Dynamo equations usually include Navier-Stokes, with the Lorentz force contribution. Physically, the differentially rotating Earth's outer core may well be the consequence
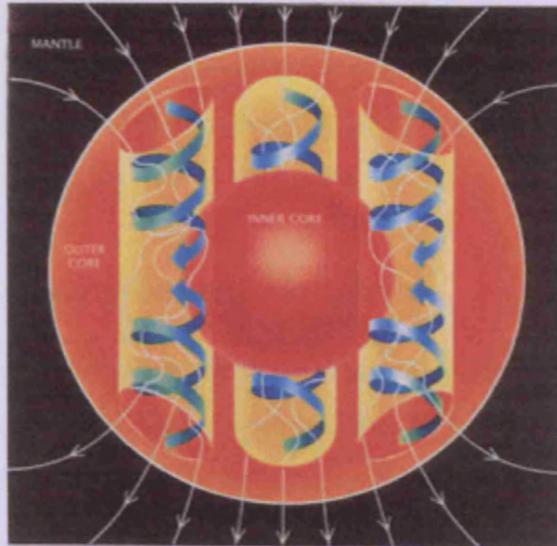
Figure 1.5: A schematic of the Earth's Dynamo. Taken from the website of the National Geomagnetism Program (USGS).

of convective flows, illustrated by Figure 1.5. Suppose that thermal and compositional convection occurs at the outer core. The moving particles keep their torque (turning moment) constant, which is why those particles that have moved up will slow down compared to the general particle flow they are in, while the denser material that moves down will also move faster relative to the general flow [*Zharkov*, 1986].

Measured variations in the Earth's length of day (LOD) on a decadal timescale are usually attributed to the exchange of angular momentum between the solid mantle and fluid core. One of several possible mechanisms for this exchange is *electromagnetic coupling* between the core and a (relatively) weakly conducting mantle. However, mantle conductivity, which is a determining parameter in this process, is only poorly known. This results is a controversy related to flows that may or may not exist that both explain the observed secular variation and generate the observed changes in length of day. There is a general agreement [e.g. *Buffett*, 1992; *Wicht and Jault*, 2000; *Mathews et al.*, 2002] that for the case of laterally varying mantle conductivity, just as in the simplest case of a

homogeneous mantle conductivity, a minimum average conductance of 10^8 S of the lower mantle is always required to make electromagnetic coupling efficient, possibly with a thin conducting layer located at the base of the mantle.

1.2.3 Time-varying geomagnetic field

Earth's magnetic field varies on time scales ranging from nanoseconds to many millions of years. This temporal variation may be separated into parts that are internal or external in origin (Section 1.2.1). Variations on short time scales are usually attributed to external sources, because the conductive mantle screens out high frequency variations arising in the core. The external field variations induce currents in Earth's mantle which can be related to the electrical conductivity profile (and thus more generally to structure and composition) within the Earth. Long term internal field variations are of interest because they give clues to the workings of the geodynamo in Earth's liquid outer core, as well as providing useful tools for magneto-stratigraphic, paleoclimate and tectonic problems. Observations of the field come from magnetic observatory and satellite data, which provide good global coverage over short time scales, from deployment of temporary instrument arrays, and from palaeomagnetic measurements and marine magnetic surveys for longer time scales [Backus et al., 1996].

After Gauss's pioneering work, the spherical harmonic analysis of the geomagnetic field has been repeated many times. The elements of the magnetic field have been represented on magnetic maps with isolines. Although the drift of Earth's magnetic field B has been known for centuries, a detailed comparison of such maps performed in 1950 has confirmed that the isolines have been shifting systematically to the West by about 0.2° a year in longitude.

Geomagnetic field reversals is an important very long time scale variation of the Earth's magnetic field. Reversal frequencies have varied significantly over mantle convection time
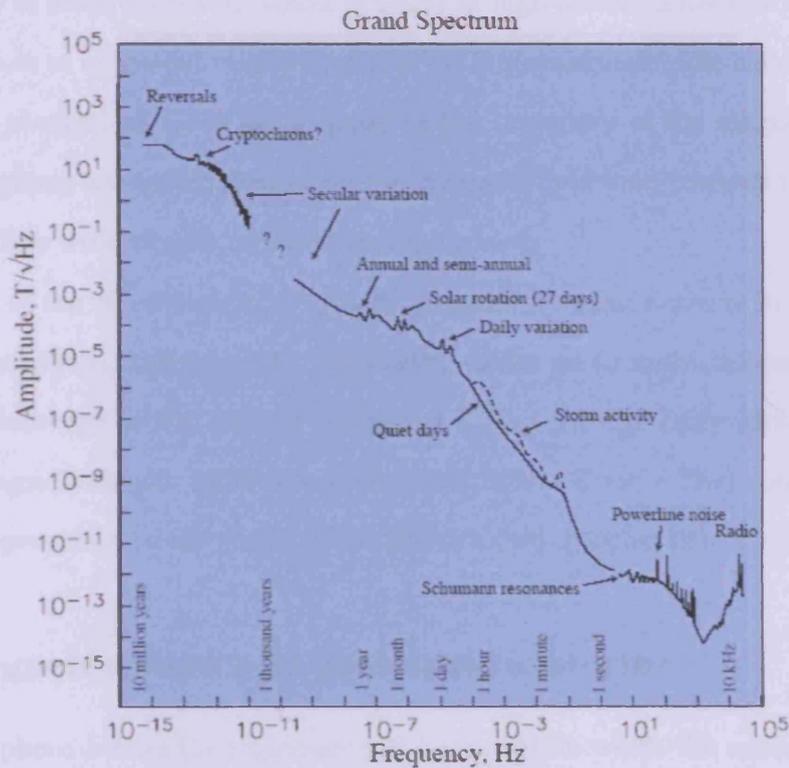
Figure 1.6: Time-varying magnetic field presented in the frequency domain (taken from S. Constable, "Global Electrical Conductivity and Magnetic Satellite Induction Studies", 2nd CHAMP Meeting, Sept 2003)

scales ($\approx$ 200 Ma); these variations are inferred from palaeomagnetism and (more importantly) from studies of seafloor magnetic anomalies.

The frequency-domain decomposition of the Earth's magnetic field variations (Figure 1.6) shows, apart from the irregular reversals, a number of regular secular, annual, solar rotation and daily variations, the latter sometimes disturbed by storm-time activity. The amplitude of magnetic disturbances is larger at high latitudes because of the presence of the oval bands of enhanced currents around each geomagnetic pole called auroral electrojets. Some charged particles get trapped at the boundary of the magnetosphere and, in the polar regions, are accelerated along the magnetic field lines towards the atmosphere and finally collide with oxygen and nitrogen molecules.

The study of the electromagnetic signals initiated by solar activity in the upper layers of the atmosphere, diffusing into the Earth, allows us to make inferences about the electrical conductivity of the interior (Section 1.3.3). Both the daily (Solar Quiet – Sq) and the geomagnetic storm variations (Disturbed Storm Time – Dst) contain energy at frequencies appropriate to the study of the Earth's deep geoelectrical structure.

1.2.4 Magnetosphere and geomagnetic storms

The magnetosphere defines the region around Earth within which the main magnetic field exerts an influence (e.g. *Dungey* [1961]; *Parkinson* [1983]; *Campbell* [2003]). Although the main magnetic field originates inside the Earth (Sections 1.2.1 and 1.2.2), the complex processes taking place in the magnetosphere create external source currents that are of major importance in global EM induction studies.

The plasma of the Sun's corona is so hot that gravity cannot contain it. Instead, the upper fringes flow away in all directions, in a constant stream of particles known as the *solar wind*. Moving at about 400 km/sec, the wind needs about 4-5 days to reach Earth.
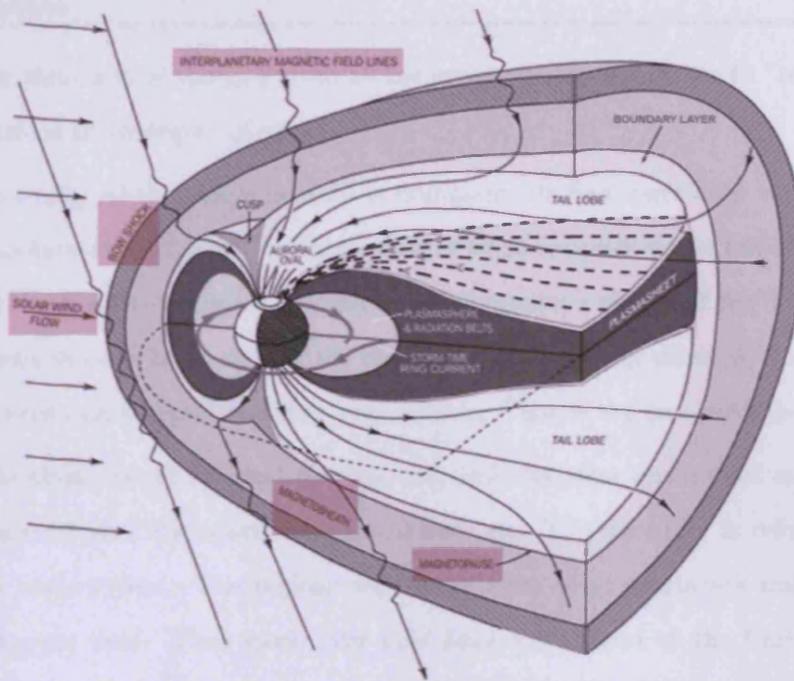
Figure 1.7: A wide-angle diagram of the magnetosphere showing the solar wind interaction with the Earth's magnetic fields and the various storm-time sources (ring current and auroral ovals). Figure taken from the IMAGE website of the Southwest Research Institute.

As the solar wind leaves the corona, it picks up the local magnetic field – contributed by sunspots and by the Sun's magnetic poles – and drags its field lines into space, forming the interplanetary magnetic field (IMF). Interacting with the Earth's magnetic field (Figure 1.7), it creates powerful magnetic disturbances in the Earth's environment that have been termed *geomagnetic storms*. The "sudden commencement" jumps of the magnetic intensity correspond to the arrival of fast plasma clouds, which plough through the ordinary solar wind and create shock fronts ahead of themselves [*Stern*, 1994].

Field lines in a plasma act like wires on which ions and electrons are strung. If the field is strong, these particles are forced to go to wherever the lines guide them. On the other hand, when the particles are numerous and energetic, as is the case with the solar wind, they can push the magnetic field around. When their flow is deflected, for instance, the lines will change shape, so as to always thread the same particles. Because of this

effect, the structure of the IMF even at the greatest distances tends to "remember" the Sun's rotation at its region of origin.

If the energy of the magnetic field is dominant, its field lines keep their shapes and particle motion must conform to them; that is what happens in the radiation belts. On the other hand, if the energy of the particles is dominant – that is, if the field is weak and the particles dense – the motion of the particles is only slightly affected, whereas the field lines are bent and dragged to follow that motion. That is the case with the solar wind.

By the equations of an ideal plasma, ions and electrons which start out sharing the same magnetic field line continue to do so later on. This property is referred to as the *field line preservation*. The regions where the solar wind starts are immersed in the Sun's magnetic field. Thus some solar field lines will extend to the Earth and further out, producing the interplanetary magnetic field (IMF). The IMF is quite weak – at the Earth's orbit, only 1/10,000 of the field at the Earth's surface – but still it exerts an extraordinary influence on the Earth's magnetosphere.

Sources produced in the magnetosphere include the ionospheric source effects in the plasma sheet [*Chappell et al.*, 1987] and other processes, e.g. field line resonances and lightning, which becomes a dominant source of signal for electromagnetic induction methods at frequencies above about 10 Hz. At significantly lower frequencies (Figure 1.6) the dominant external sources are created by electrons and positive ions trapped in the dipole-like field near the Earth. They tend to circulate around the magnetic axis of the dipole without gaining or losing energy. Viewed from above the northern magnetic pole, ions circulate clockwise, electrons counterclockwise, producing an electric current around the Earth's equator. Arthur Schuster in 1911 gave it the name we still use, the *ring current*. At mid-latitudes, the ring current is a good approximation to the source of the external field; however, at higher latitudes other (auroral) sources of similar nature come into play.

Geomagnetic storms significantly affect the strength of the ring current. The storm-

time growth of the ring current lasts from 3 to 12 hours and constitutes the "main phase" of a magnetic storm. Following this main phase, the ring current begins to decay, returning to its pre-storm state in two to three days. The growth and recovery of the ring current are indicated by changes in the Dst (disturbance storm time) index, the geomagnetic index that serves as the standard measure of ring current activity.

1.3 Overview of electromagnetic induction methods

The natural electromagnetic (EM) fields are distorted in the magnetosphere, as described in 1.2.4, then propagate through the atmosphere, the electrical conductivity of which is very low (the value of $\approx 1 \times 10^{-10} Sm^{-1}$ has been proposed by *Rokityansky* [1982]; *Mackie and Madden* [1993a]; *Uyeshima and Schultz* [2000] for computational convenience). This is the reason they lose almost none of their energy in the atmosphere (see Section 1.3.3 for a discussion of energy attenuation). They reach the Earth's surface as quasi-homogeneous waves, under the assumption that the frequency of interest is small. The larger part of the energy incident at the surface is reflected, while the rest diffuses deep into the Earth to a depth that can be approximated by the skin depth, also referred to as the *depth of penetration*, described further in Section 1.3.3.

This causes eddy currents in the conducting regions of the interior, by the principle of electromagnetic induction, that in turn result in a variation of the EM fields observed at the surface. This process is governed by Maxwell's equations, that involve the electrical conductivity and the magnetic permeability as parameters (Section 1.3.2). The latter doesn't vary much throughout the Earth, and for most practical purposes is assumed constant, while the conductivity varies a great deal and is an important constraint on the state of matter (Section 1.4).

Hence, the dependence of the EM field variations on the conductivity of the relevant

Earth's regions can be written out explicitly. Once the electromagnetic fields measured on the surface are separated into the internal and the external parts, this information can be used to draw inferences about the conductivity of the interior that is situated within the depth of penetration of the source. Thus, the Earth's magnetic field may be viewed as source of information about the structure of the deep interior. This approach is an example of a *geophysical inverse problem*, and forms the basis of the electromagnetic sounding techniques used in this thesis.

1.3.1 Forward and inverse problems of electromagnetic induction

The methodology of field separation is now well mastered, and employs the spherical harmonic expansion (Section 1.2.1). The data finally measured are functions of these separated fields, known as *response functions* (Section 1.3.4). They are recorded as a time series, but then are usually Fourier transformed into the frequency domain. Given that the source fields are known, these data, used in conjunction with Maxwell's equations, set up the foundation for the non-linear inverse modelling in the setup of deep geomagnetic exploration. The *forward problem* constitutes computing the response functions assuming the conductivity structure and the source fields are known. This can be attacked both numerically, and analytically under some restrictive assumptions. The corresponding *inverse problem* is to reconstruct the conductivity structure, given the response data. In trying to do so we're faced with numerous theoretical and practical complications, arising from the inevitable finiteness and noisiness of data measured, as opposed to the true continuous distribution we are trying to reconstruct. In this text the general setting of the ill-posed inverse problems and the application to the inversions that arise in a geophysics is considered in Section 4.1.

Geophysical inverse problems arise in the contexts of local, regional and global geo-

electromagnetic sounding. These are very different scales of looking at the same problem of reconstructing the conductivity distribution, and different methods are developed for each. For instance, local electromagnetic sounding is very widely applied throughout the exploration industry (mainly artificial source, e.g. TDEM), and is only concerned with relatively small areas and the depths of several km. For such problems Cartesian coordinates give a good enough approximation to the geometry of the survey area. Also, the source fields are often generated artificially, and hence the problem of their reconstruction does not arise. None of this is true for mantle-scale problems, either regional, with the data distributed on a segment of the Earth's sphere only (e.g. Europe), or global, with the data distributed world-wide. We focus our research on global-scale problems; hence we consider spherical geometries.

The natural source fields used in the global geomagnetic modelling, although not known explicitly, can be approximated analytically (Section 1.3.4). This sets up the necessity of additional preliminary data processing that accounts for the effects not taken into account by this approximation, such as that performed in *Fujii and Schultz* [2002], described in Chapter 2.

1.3.2 Governing equations

In this study, we are concerned with the period range 2 days $\leq T \leq 1$ year. This period range corresponds to penetration depths approximately $300 - 1500$ km for a realistic range of Earth's material conductivities. At these long periods, the energy transport mechanism within the Earth is almost purely a diffusion process and the *frequency domain* approximation to Maxwell's equations assuming harmonic $e^{+i\omega t}$ time dependence can be written as

$$\nabla \times \mathbf{H} = \sigma\mathbf{E} \tag{1.7}$$

$$\nabla \times \mathbf{E} = -i\omega\mu\mathbf{H} \tag{1.8}$$

where $\mathbf{H}$ is the *magnetic field strength*, $\mathbf{E}$ is the *electric field*, μ is the magnetic permeability and σ is the electrical conductivity of the medium. This system of equations is known as the *quasi-static approximation*, and is effectively equivalent to the vector Helmholtz equation on either the magnetic field strength, or the electric field.

We cover in Section 3.2 the derivation of this system of equations from the time domain formulation, and all the assumptions that need to be satisfied for this simplified system of equations to be valid. For now, we state that a variation of the above system of equations is applicable to the majority of deep electromagnetic sounding problems solved in the frequency domain. Forward problems that arise in this context can be solved analytically, numerically, and by the methods of physical modelling. Regardless of the problem setup, the solution is the EM field satisfying the following conditions:

1. Maxwell's equations in the quasi-static approximation 1.7 and 1.8,

2. conditions on the EM fields at the boundaries of computational domain or infinity.

1.3.3   Electromagnetic induction effect and skin depth

Consider a plane EM wave vertically incident to the boundary of a uniform halfspace of conductivity $\sigma = \text{const}$ [e.g. *Parker*, 1994; *Weaver*, 1994; *Egbert*, 2006]. At a given time t, the phase front of such a wave is a horizontal plane, with the fields $\mathbf{E}$ and $\mathbf{H}$ constant (*homogeneous*) in that plane. The quasi-static approximation to Maxwell's equations (Section 1.3.2) for a given angular frequency ω simplifies to

$$E_x''(z) - i\omega\mu_0\sigma E_x(z) = 0,$$ (1.9)

with orthogonal electric and magnetic fields related by

$$E_x'(z) = i\omega\mu_0 H_y(z).$$ (1.10)

Equation 1.9 is a second-order linear homogeneous ordinary differential equation (ODE) with constant coefficients. We write a characteristic equation $k^2 - i\omega\mu_0\sigma = 0$ and obtain the general solution

$$E_x(z) = \alpha e^{+kz} + \beta e^{-kz},$$ (1.11)

where $k^2 = i\omega\mu_0\sigma$. This is the definition of a *wavenumber* k, which is evidently a function of conductivity and the angular frequency of the wave diffusing into the medium. By Euler's formula, $\sqrt{i} = \pm(1+i)/\sqrt{2}$. Choosing Re $k > 0$, Im $k > 0$, $k = (1+i)\sqrt{\omega\mu_0\sigma/2}$.

We suppress the exponentially growing term to meet the zero boundary conditions at infinity, and obtain a particular solution

$$E_x(z) = \beta e^{-(1+i)z/z_0},$$ (1.12)

where

$$z_0 = \sqrt{\frac{2}{\omega\mu_0\sigma}}.$$ (1.13)

The value of z_0 has units of depth, and it is known as the *depth of penetration* of an EM wave into the medium, or *skin depth*. Notably, this is the depth at which the *amplitude* of the EM wave $E_x(z)$ decays to $1/e$ its surface amplitude. From Equation 1.10

we conclude that the horizontal magnetic fields decay in the halfspace in much the same way.

This frequency dependence aids the electromagnetic depth sounding approach in reconstructing the conductivity of the Earth's interior, via the MT, GDS, and other methodologies developed to study the geoelectric properties of materials, and the Earth in particular (Section 1.3.4). Intuitively, this can be understood as follows. The time varying EM signal induces time varying currents in the Earth. As they disseminate, they flow in the layers close to the surface, and the greater the frequency of the signal, the closer the currents keep to the surface. Therefore, the lower the frequency, the deeper layers may be sensed. This effect is known as the *skin-depth effect*. To calculate the frequency necessary to sense a certain depth within a uniform half-space, we rearrange the formula for skin depth (Equation 1.13) to get $f = (1/(\pi z_0^2 \mu_0 \sigma))$ Hz.

Problems may arise, however, if a horizontal layer of high conductivity is met, as this would shield the deeper layers from the EM fields. Encountering a good conductor, the field quickly dissipates its energy and dies out. This makes physical sense since external electrical current systems will induce internal eddy currents to flow only in the conductive regions. On the other hand, for the Earth model where the upper mantle is entirely resistive, hardly any currents will flow, meaning that hardly any dissipation of energy will take place. Thus, the technique of electromagnetic induction is almost completely insensitive to resistive regions, being particularly sensitive to the regions of enhanced conductance.

Take $\mu_0 = 4\pi \times 10^{-7}$ H m^{-1}. Further, define $\lambda = 2\pi z_0$, so that $\frac{2\pi}{\lambda} = \sqrt{\omega \mu_0 \sigma / 2}$. Just like the skin depth,

$$\lambda = \sqrt{\frac{10^7}{f\sigma}}$$

has the units of length and is known as the *wavelength* of an EM wave. Thus, the x (North) and y (East) components of the EM fields are harmonic waves that decay exponentially

with depth, with the spatial period defined by λ. The speed with which they decay is expressed by $\frac{2\pi}{\lambda}$, and the distance the phase front advances in one period is λ.

1.3.4 Response functions and source field assumptions

To derive a mathematical representation of the geophysical electromagnetic induction problem based on Maxwell's equations simplifying source assumptions are generally required. For most local induction studies, the source field is either imposed artificially or known to an approximate degree. Also, if natural EM fields are used in localised studies, the area under investigation is likely to be very small compared to the scale of the source field, in which case it may be impossible to determine the structure of the source given just the measured surface data [*Berdichevsky et al.*, 2000]. For example, a horizontal, plane wave magnetic field at the surface of a 1-D Earth could be generated by a horizontal current sheet located at any height above the surface, but in practice the origin of the source fields is sufficiently distant so that to make them horizontal by the time they reach the Earth, as argued originally by *Tikhonov* [1950] and *Cagniard* [1953]. Many workers have put together a solution to the EM induction problem in a uniform to multi-layered conducting half space under various source assumptions. A comprehensive review of the various methods is given in *Weaver* [1994] (see also Section 1.3.5). The problem of a good approximation to the source structure for a spherical Earth involves approximating the fields globally with an analytic expression. Often this is done with the aid of spherical harmonic expansion described in Section 1.2.1, cut down to the first few terms of the series. In practice, the elementary harmonic source field is assumed, to make the problem of 3-D EM induction tractable (e.g. Section 3.3.2). An alternative for a practical inversion is the approach proposed by *Fainberg et al.* [1990a], which is to invert not only with respect to the parameters modelling the structure of the Earth's interior, but also with respect to external field parameters, which can of course be coefficients of the spherical harmonic

expansion (see also *Mulhall* [2003], who has employed this approach in the time domain for a layered Cartesian Earth). The method can be used for laterally inhomogeneous as well as homogeneous Earth models. It does, however, aggravate the convergence problems by adding extra parameters to the model space, and can only be used with a certain type of (specifically written) forward solvers.

Simple source field assumptions include a magnetic dipole in free space or a conducting medium (approximating small current loops often used as artificial sources in EM sounding); infinite line current in free space (which is a first approximation to the ionospheric electrojets that flow in auroral and equatorial regions); and the infinite current sheet in free space (which is a reasonable assumption for natural systems of external currents for EM studies over a limited area of the Earth's surface, except in polar or equatorial regions where ionospheric electrojets may be present).

As we have previously discussed, natural EM fields external to the Earth induce the internal conductive currents in the Earth's interior, by the principle of EM induction. These in turn induce secondary fields on the surface. The resultant perturbations of the field on the surface can be measured. However, separation of the primary EM fields of external origin from the secondary EM fields of internal origin is not always possible. The alternative approach is to focus on data for which the sources have a simple form, and may be assumed to have a known spatial structure, and to cancel out the amplitude and phase of the source by taking local ratios of field components. There is a wide variety of approaches designed to cancel the unknown source parameters, each best suitable to its own range of problems.

The so called *response functions* (or transfer functions) are calculated from the values of electric and / or magnetic field values known at the locations of magnetic observatories on the Earth's surface. The response function can be a functional of either vertical or horizontal components of either the magnetic or the electric fields, or both. Two distinct

groups of methods used in geoelectromagnetic research are differentiated by the response functions they employ. These are the Magnetotelluric method (MT), that employs the impedance tensor originally given in *Tikhonov* [1950] and *Cagniard* [1953], which is in essence composed of the ratios of orthogonal horizontal electric & magnetic field components at the surface; the Geomagnetic Depth Sounding (GDS) method, that uses the ratios of vertical & horizontal components of the surface magnetic field as the response functions; and the Magnetovariational Sounding (MVS) method, that uses the horizontal (spatial) gradients of the magnetic fields. For many years it was disputable which of these response functions is preferable both in terms of the quality of their field measurements, and the resolving power. The origin of GDS goes back to *Schmucker* [1970]. In the last few decades the MT method has been generally receiving more attention than GDS, but the 1990's have seen the return to the GDS methodology. Different response functions can be efficiently combined in a practical EM inversion.

The main parameters of the MT method are the *apparent resistivity* and *phase* (which are transforms of the impedance tensor, that is itself determined from the relationships between the horizontal components of the electric and magnetic field on the surface of the Earth). Vertical field transfer functions are also commonly used, that relate the vertical component of the magnetic field to its horizontal components. They are referred to variously as the Tipper or the Weise-Parkinson vector, and are in essence parameters of GDS sounding. This study is mostly concerned with a variant of the GDS method, also known as *GDS on a point*. All the transfer functions discussed above derive from the assumption that external sources are spatially uniform. This is a reasonable assumption for many studies, because for small enough periods penetration depths are small compared to horizontal length scales of source variations. However, other sorts of assumptions are appropriate for global induction studies and large arrays. The GDS on a point Z/H transfer function [*Schmucker*, 1970; *Cox et al.*, 1970; *Larsen*, 1973; *Schultz and Larsen*, 1987] is

most commonly used for long period ($>$ 5 days) global studies of mantle conductivity.

At these periods (and especially for data at mid-latitudes) a zonal dipole is an appropriate source assumption because the magnetic field variations are almost totally dominated by the ring current (and the Earth rotates under these sources, guaranteeing zonal structure in an Earth fixed frame). The magnetic potential due to the magnetospheric ring current (Sections 1.2.4 and 1.2.3) is nearly azimuthally symmetric and is well approximated by the first associated Legendre function P_1^0 the cosine of the colatitude (*Banks and Ainsworth* [1992]). This is such a standard source field assumption in global geomagnetic studies, that we shall implicitly assume a P_1^0 source unless specified otherwise. GDS on a point, converted to c responses (Section 2.3) are appropriate transfer functions to cancel the effect of such an external source, leaving the response of the Earth.

1.3.5 One-dimensional forward and inverse problems

One-dimensional models include those with spherical symmetry, as well as stratified models in the rectangular cartesian coordinates. The problem of one-dimensional geoelectric modelling in the frequency domain has been solved completely. The 1-D forward and inverse problems in the time domain is a very different problem, both non-trivial and computationally demanding. The last few years have seen some progress in this respect [e.g. *Mulhall*, 2003].

For the 1-D frequency domain forward problem, a recurrence relationship can be written (e.g. *Rokityansky* [1982], *Weaver* [1994]), that allows the 1-D forward problem of geoelectrics to be solved numerically for any 1-D conductivity profile.

For the inverse 1-D problem, the uniqueness of solution [*Langer*, 1933; *Bailey*, 1970; *Weidelt*, 1972] and Fréchet differentiability [*Macbain*, 1986; *Parker*, 1986] has been proven for the case of ideal observations. However the inversion of actual – uncertain and incomplete – field data from a geoelectrical sounding experiment (e.g. magnetotelluric, dc

resistivity, or controlled-source electromagnetic) can never yield a unique solution. Thus, the inversions are stabilized artificially, by restricting the search to a simple class of models, and then controlling the number of parameters in models within that class. A classic 1-D example is the class of *simple layered models*, that is, models consisting of a small number of layers (sometimes less than five).

However, this approach produces solutions that are dependent upon the class of models chosen. For example, choosing a small number of layers imposes unjustified discontinuities between the layers, at the depths selected by the modeler. On the other hand, if more than four or five layers are specified, a least-squares inversion is unlikely to constrain such a highly parametrized model. Thus, for the solution to reflect the true structure of the Earth, both the parametrization and the starting model must be close to being correct. Below we discuss the few approaches, that manage to avoid the problem of non-uniqueness, and give new insights into the 1-D inversion.

Consider the frequency domain problem, where ω is the frequency, and z is the vertical coordinate (+ve upwards), of the 1-D profile. The conducting layer lies between $z = h$ at the surface and $z = 0$ at the bottom, where there is a perfect conductor.

D^+ inversion

The problem of finding the smallest achievable χ^2 misfit (Section 4.1) associated with an arbitrary 1-D profile for MT or dc resistivity data was completely solved in the early 1980's. This is done by choosing the best fitting conductivity models in a space restricted to delta functions. The full solution is presented in the series of articles *Parker* [1980]; *Parker and Whaler* [1981]; *Parker* [1984]. Let us here describe the development of the methodology, that has resulted in Parker's 1980 D^+ algorithm.

Effectively, D^+ comprises the set of Dirac delta functions in conductance - that is in a Cartesian coordinate system, D^+ models are a set of infinitesimally thin "layers"

each of which has finite conductance. The D^+ algorithm determines the number of delta functions required, the depths within a cartesian Earth at which these delta functions are located, and the conductance of each delta function. The algorithm is a very useful tool in many numerical geophysical inversions, because it provides a limit in terms of how well the data can be fit by a 1-D model. If any D^+ model exists that fits the data, then it is safe to say that the data are consistent with a 1-D interpretation; and D^+ uniquely finds the best possible fitting 1-D solution. In fact, it appears that the "roughest" space of δ-functions always contains at least one solution that is the best fitting of all possible solutions from all possible spaces of 1-D models, so that smoother models invariably have worse misfits.

The Weidelt's transformation can be used (*Weidelt* [1972]; Section 1.3.5) to project the models so computed to spherical coordinates, which allows one to obtain solutions for a spherically symmetric Earth.

Occam's inversion

To combat the ambiguity of finite and noisy data sets, that has been underscored by *Parker* [1984], a very different class of models has been proposed by *Constable et al.* [1987], as an alternative to the optimal but non-physical D^+. They propose finding the smoothest model in a special sense so that its features depart from the simplest case only as far as is necessary to fit the data. This algorithm has been developed in parallel with the *Smith and Booker* [1988] minimum structure inversion.

The idea behind smooth space inversions is that we do not want to be misled by features that appear in the model but are not essential in matching the observations. To attach any importance to the anomalous conductivity zones in a model with a general fine structure could be misleading, if another model can be found fitting the data equally well,

and not requiring this feature. In this case, although we cannot preclude the existence of the anomalous zones in reality, they are not demanded by the data. On the other hand, an advantage of inverting for maximally smooth models is that the model we obtain is unique (however, only within a narrow definition of objective function), in the sense that there is only one *smoothest* model satisfying any given data set, and should reflect any significant features that any other model fitting the data would include.

Constable et al. [1987] find that the balance between suppressing significant (too few parameters) and introducing spurious structure (too many parameters) is best attained by allowing the model to be as flexible, as possible, but to suppress complexity explicitly. This can be done by defining *roughness* (the converse of smoothness), as the integrated square of the first or second derivative with respect to depth:

$$R_1 = \int \left(\frac{dm}{dz} \right)^2 dz \text{ or } R_2 = \int \left(\frac{d^2m}{dz^2} \right)^2 dz, \tag{1.14}$$

where $m(z)$ could be the continuous profile of resistivity or log-resistivity. The original idea is that of regularisation of ill-posed problems [*Tikhonov*, 1943] (see Section 4.1), which can be reformulated as introducing the penalty for complexity, to combat the non-uniqueness generically arising from the instability of solutions. The concept is also widely used in the methods of data interpolation.

If we take our misfit to be the usual χ^2 (weighted least-squares criterion, see 4.1), the mathematical problem to be solved is the following: for given data $d_j, j = 1, \ldots M$, and the associated uncertainties, we must find the model $m(z)$ that makes R_1 or R_2 as small as possible, while χ^2 achieves an acceptable value. Because of the non-linearity of the problem, we may not be able to make χ^2 small enough, and we have to assume that the approximations of one-dimensionality, large-scale source fields etc. are all good enough for an acceptable fit to be possible. Fortunately, as already described, for the one-dimensional

problems, the lower limit of χ^2 can be found using Parker's D^+. The combination of these two numerical inversions solves the 1-D MT and GDS inverse problem, to the extent it can be solved.

Weidelt's transformation

If degree and order l of the spherical harmonic representation of the source field is independent of frequency (this is approximately true for signals originating in D_{st} but *not* for S_q), the 1-D inverse problem can be solved by deriving from spherical response functions a preliminary plane-Earth model $\hat{\rho}(\hat{z})$, using one the techniques described above. This model can be subsequently transformed into a spherical Earth model $\rho(r)$ with *Weidelt* [1972] transformation formula:

$$\rho(r) = f^4(r/a) \cdot \hat{\rho}(\hat{z}),\qquad(1.15)$$

where

$$f(r/a) = \frac{(l+1)\left(\frac{a}{r}\right)^l + l\left(\frac{r}{a}\right)^{l+1}}{2l+1},$$

and

$$\hat{z} = \frac{a\left[\left(\frac{a}{r}\right)^l - \left(\frac{r}{a}\right)^{l+1}\right]}{f(r/a) \cdot (2l+1)}.$$

Here a is the radius of the Earth and l is the degree and order of the source field assumption. If we only take the first harmonic of the source field (P_1^0) we set $l = 1$.

1.4 Composition and electrical conductivity of the mantle

Electrical conductivity is a measure of how well a material accommodates the transport of electric charge. Conductance is an electrical phenomenon where a material contains movable charges, which may be electrons, positive holes or ions. When a difference of electrical potential is placed across a conductor, its movable charges flow, and an electric current appears. Conductivity is defined as the ratio of the current density to the electric field strength.

In SI derived units (MKS) conductivity σ is measured in Siemens per metre, $S\ m^{-1}$, Siemens being 1/Ohm. Conductivity is the reciprocal of electrical resistivity ρ, which is measured in Ohm $\cdot$ m. For a uniform specimen of material of length l, cross-section area S and resistance R, $\rho = RS/l$. The two parameters are used interchangeably in the literature.

With the main Earth layers seismically determined (Section 1.1), composition is generally inferred from three main sources: field occurrences of rocks that are thought to come from the shallower parts of the Earth; igneous rocks thought to be formed from magma derived from melting the mantle; and laboratory measurements of the physical properties of minerals and rocks (e.g. *Davies* [2001]). In fact, there are several types of rocks that come directly from the upper mantle, and hence define very good constraints on the compositional models of the crust and the mantle. These are, most notably, *alpine peridotites* and *ophiolites*, that are magmatic rocks inferred to have been thrust to the surface by tectonic movements (that could have been extensive thrust faulting). Rocks carried to the surface by magmas are known as *xenoliths* (literally, "strange rocks") because typically their compositions are not directly related to those of the magmas that

carry them. There are two main magma types that bear mantle xenoliths: basalts and kimberlites. In particular, kimberlite pipes are notable for the extreme velocity of their eruptions, due to relatively high content of volatiles; so that magma erupting from kimberlite pipes seems to break off pieces of the mantle on its way up and carry them to the surface.

The upper mantle consists mainly of olivine dominated components, that are crystalline semi-conductors. It is predominantly composed of *peridotite*, with a less than 10% fraction of *eclogite*, hence the upper mantle is also known as *peridotite zone* in the context of its chemical and mineralogical composition. Peridotite's mineralogy is typically about 60% $(Mg,Fe)_2SiO_4$ olivine , 25% $(Mg,Fe)SiO_3$ orthopyroxene and the balance $(Ca,Mg)SiO_3$ clinopyroxene and garnet. The eclogites are composed of about 60% clinopyroxene, 30% garnet and variable amounts of other materials. Much less is known of the olivine to spinel transition at depths greater than 200 km. Olivine transforms to its high-pressure polymorph wadsleyite around 410 km depth, so that the upper mantle mineralogy at depths 410 − 520 km may be approximated by 60% wadsleyite and 40% clinopyroxene [*Xu et al.*, 2000]. Wadsleyite transforms to ringwoodite while clinopyroxene transforms to $(Mg,Fe)SiO_3$ ilmenite + garnet at approximately 520 km depth [*Xu et al.*, 2000]. Both upper mantle polymorphs of olivine, wadsleyite and ringwoodite have been found to have nearly two orders of magnitude higher conductivity than olivine [*Xu et al.*, 1998]. This may help reconcile the difference between laboratory measurements for dry single crystal olivine, and field observations that require higher upper mantle conductance [*Uyeshima and Schultz*, 2000].

Ringwoodite disproportionates to $(Mg,Fe)SiO_3$ perovskite + magnesiowüstite near 660 − 670 km [*Xu et al.*, 2000]. Although ilmenite and garnet start to transform to the perovskite structure at these depths, they can remain stable to $\sim$ 800 km depth [*Serghiou*

et al., 1998]. Garnet transforms completely to perovskite at $\sim$ 800 km. The lower mantle, from a depth of $\sim$ 670 km down to the coremantle boundary ($\sim$ 2,990 km), is mostly composed of (Mg,Fe)SiO$_3$ perovskite and (Mg,Fe)O magnesiowüstite [*Peyronneau and Poirier*, 1989], in proportions approximately 78% and 22% [*Xu et al.*, 2000].

Conductivity increases with increasing temperature and increasing pressure. It may also increase with iron content. All these properties of the mantle are known to vary with depth; thus, a strong radial dependence of the electrical conductivity is observed in Earth's mantle. However, other factors, such as enhanced hydrogen content, coexistence of multiple mineral phases, and oxygen fugacity also exert a significant influence on mantle conductivity. Recent convergence between laboratory studies of mineral conductivity and electromagnetic field data have provided some confidence in our estimates of mantle electrical properties (Figure 1.8).

The temperature of the mantle grows with depth and pressure, and we know that in insulators and in semi-conductors an increase in temperature implies a rapid increase in conductivity too. This has been used to infer relatively fast increase in conductivity with depth in the outer layers of the Earth. The increase in pressure also increases the melting temperature of materials, while the closer the temperature of a region is to the melting temperature of the material, the higher is the ionic conductivity. Hence at the regions where the ionic conductivity starts to dominate, the conductivity gradient is less steep.

Above the transition zone, the effect of thermal gradients and the presence of volatiles and melt may lead to several orders-of-magnitude variation in conductivity across a given convection cell. In contrast, lateral variations in electrical conductivity in the lower mantle may be less than one order of magnitude, which has been indeed inferred by both geophysical model and laboratory measurements.

Some step changes in conductivity have been observed at the seismic discontinuities.
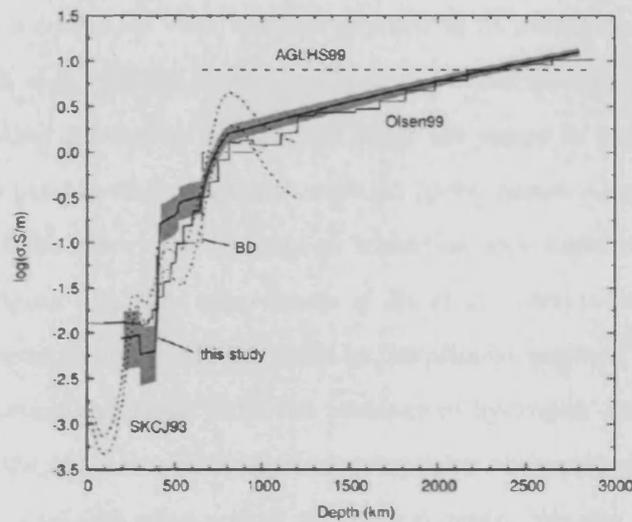
Figure 1.8: Laboratory-based conductivity-depth profile compared with geophysical models. Shaded areas illustrate the effect on the model of a ±100° C temperature variation. The laboratory-based profile is similar to BD if it is considered as a three-layer mantle and similar to Olsen99 if it is smoothed. Geophysical models shown are AGLHS99 [*Alexandrescu et al.*, 1999], SKCJ93 [*Schultz et al.*, 1993], BD [*Bahr and Duba*, 2000], and Olsen99 [*Olsen*, 1999a] (profile obtained beneath Europe). Figure taken from *Xu et al.* [2003].

For example, an order of magnitude change in electrical conductivity has been discovered at Moho depth below the highly resistive Slave craton [*Jones and Ferguson*, 2001]. A zone of enhanced conductivity near 416 − 456 km has been identified by *Schultz et al.* [1993] as a result of magnetotelluric inversion beneath the Canadian Shield, which implies a jump in electrical conductivity around the 410 km seismic discontinuity in this region. Other studies have shown that an up to two orders of magnitude jump in electrical conductivity around 410 km is consistent with, but not required to fit geophysical data (*Egbert et al.* [1992] and *Utada et al.* [2003]). In general, EM data can not normally resolve jumps: while there may be other reasons to assume that there are jumps in radial conductivity, the data are always fittable with a smooth transition (pers. comm. Gary Egbert). However, comparison with laboratory experiments on transition zone minerals argues in favour of such a jump (Figure 1.9). The experiments of *Xu et al.* [1998] indicate that the olivine-wadsleyite phase transition at 410 km could be the primary cause of this jump. Enhanced electrical conduction can result from the presence of hydrogen, and the contamination by volatiles of the impure semi-conductors comprising the upper mantle. *Utada et al.* [2003] also infer a smaller jump around the 520 km depth. We also present the electrical conductivity profile estimated by *Kuvshinov et al.* [2005] (Figure 1.10) in the same area, which takes the effects of the major distorting object (a non-uniform ocean) into account. We also note that the recent *Velimsky et al.* [2006] inversion of satellite data (Figure 1.12) favours a rather resistive upper mantle and transition zone. Another global 1-D model based on inversion of satellite data has been recently proposed by *Kuvshinov and Olsen* [2006], and is in general agreement with the previous inverse solutions of *Olsen* [1999a] and *Kuvshinov et al.* [2005].

Water content of the transition zone has long been a matter of dispute (e.g. *Constable* [1993], *Schultz et al.* [1993], *Lizarralde et al.* [1995]). It has been argued that if the rapid increase of conductivity at the 410 km discontinuity (proposed by *Xu et al.* [1998] and
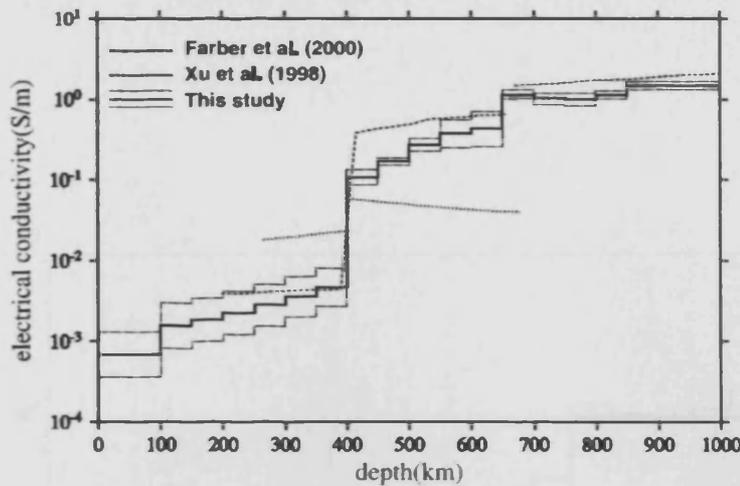
Figure 1.9: Comparison of the two-jump model of electrical conductivity beneath the north Pacific region and the results of laboratory experiments by *Farber et al.* [2000] and *Xu et al.* [1998]. Figure taken from *Utada et al.* [2003].

Utada et al. [2003]) exists, the increase in water content of about one order of magnitude across the discontinuity is required to explain this jump in electrical conductivity [*Bercovici and Karato*, 2003; *Huang et al.*, 2005, 2006]. The argument uses the ratio of oxygen fugacity of the mantle above and below 410 km, which is a better estimated quantity than the ratio of water content. This compositional model is known as the transition-zone water-filter, and assumes an anomalously high water content in the transition zone of about 1000 p.p.m. (compared to $50 - 200$ p.p.m. in the upper mantle). On the other hand, *Tarits et al.* [2004] claim (Figure 1.11) that a temperature of $350 - 450$ C less than normal in the transition zone explains their magnetotelluric conductivity profile beneath a subducted slab in the Western Alps. At $200 - 400$ km, their model favours a cold mantle with $1000 - 1500$ p.p.m. of water dissolved in olivine, suggesting that the subducted slab is dehydrated before reaching the transition zone. If this model is correct, this is strong evidence that there are regions of anomalously high water content in the upper mantle. Recently, a new model of olivine conductivity (SO3) has been proposed
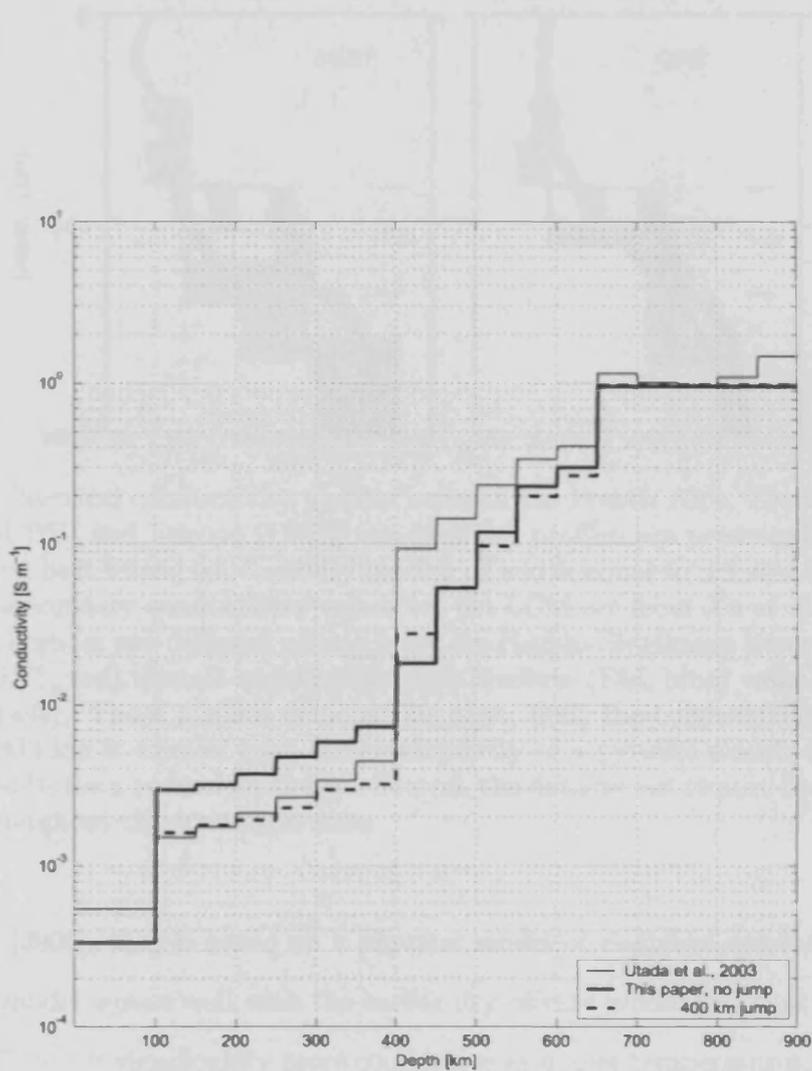
Figure 1.10: Two conductivity distributions obtained in *Kuvshinov et al.* [2005] by a quasi-1D inversion with an iterative correction for the ocean effect. Thick dashed and solid lines present the results with and without a hypothetical discontinuity at a depth of 400 km. The result conceptually confirms the findings of *Utada et al.* [2003].
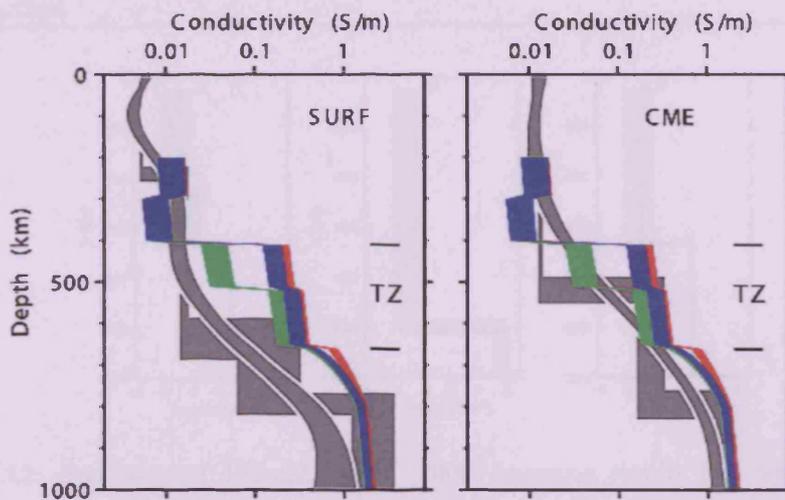
Figure 1.11: Electrical conductivity profiles beneath the French Alps, *Tarits et al.* [2004]. Sur-Frêtes (SURF) and Europe (CME) conductivity profiles are presented. The smooth and the layered best fitting conductivity models of width equal to ±1 standard deviation are in grey. Laboratory conductivity values for the LCM are from *Xu et al.* [2000]. They are presented here for two different mixing laws, the Hashin-Shtrikman lower (HS^-, green) and upper (HS^+, red) bounds and the Effective Medium (EM, blue) value (see *Xu et al.* [2000] for details). These profiles demonstrate that, first, the conductivity in the depth range $400 - 800$ km is smaller than the conductivity of a pyrolite mantle obtained from laboratory results for a normal geotherm. Second, the data do not require the conductivity to change throughout the transition zone.

by *Constable* [2006], that is based on a physical model of electrical conduction by point defects. The model agrees well with the earlier dry olivine model (SO2) at temperatures below 1200° C, but is significantly more conductive at higher temperatures, thus allowing higher conductivity at temperatures expected in the upper mantle.

A conductivity-depth profile from 660 (or 670) to 2900 kilometers based on aluminum-bearing perovskite is thought to be consistent with most geophysical models [*Xu et al.*, 1998]. Extrapolation of the laboratory measurements to the temperature appropriate for a depth of 1100 km yields a conductivity of the lower mantle of the order of 1 S m^{-1}, varying by no more than about a factor of five across the entire lower mantle, reaching a maximum value of only 3-10 S m^{-1} [*Shankland et al.*, 1993]. Extrapolation to the temper-
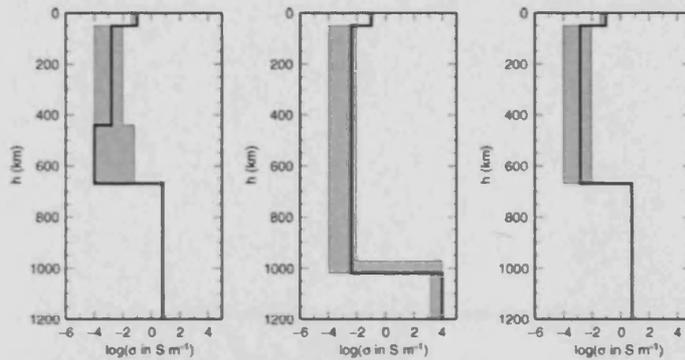
Figure 1.12: Summary of *Velimsky et al.* [2006] inversion results for, respectively, the four-layer model with fixed interface depths (left), the three-layer model with variable mantle interface depth and the three-layer model with all fixed interfaces (right). Solid lines show the best models, grey shadings show models with the misfit χ^2 within 0.2 per cent from the minimum for each particular parametrization.

ature and pressure of the core-mantle boundary yields values between 50 and 100 S m^{-1} [*Peyronneau and Poirier*, 1989]. However, a much higher conductivity value anywhere in the region $250 - 95000$ S m^{-1} has been also proposed for the lower-most mantle [*Semenov and Jozwiak*, 1999], based on a regional inversion of geophysical data. Strong lateral variations of seismic velocities and vigorous chemical activity have been detected at the boundary between Earth's liquid iron core and silicate mantle, with the total conductance of the lowermost 200 km of the mantle estimated to be 1.7×10^8 S [*Buffett et al.*, 2000]. Unfortunately, no data exist that constrain the electrical conductivity of the lowermost mantle well. The electrical conductivity of outer-core fluid, which is thought to be mainly molten iron, is relatively well-known from high-pressure experiments and inferred to be $6(\pm 3) \times 10^5$ S m^{-1} [*Jeanloz*, 1990].
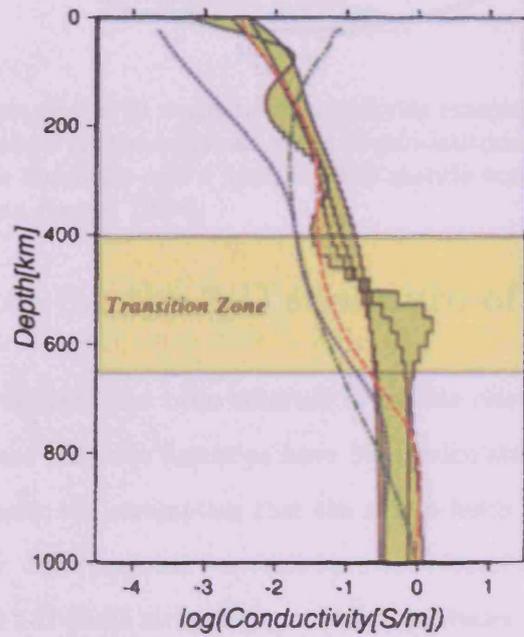
Figure 1.13: Comparison of conductivity profiles from the Changchun geomagnetic observatory (CHN) [*Ichiki et al.*, 2001] with those at Carty Lake (CLC, blue dotted line) in the Canadian Shield, Tucson (TUC, red broken line) in the southwestern United States and Honolulu (HON, green chain line) in the north central Pacific [*Neal et al.*, 2000]. The mantle transition zone is shown by yellow. Figure taken from *Ichiki et al.* [2001].
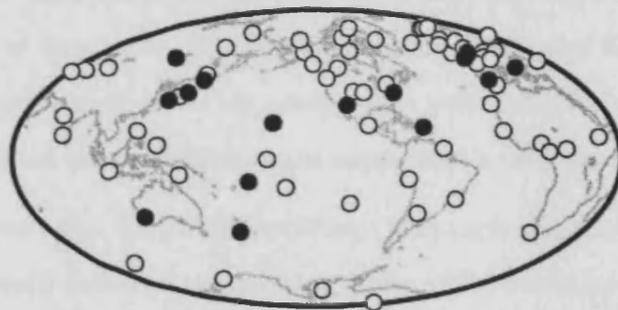
Figure 1.14: Locations of the 79 magnetic observatories examined in *Schultz and Larsen* [1987]. The dark symbols are the locations of the 15 mid-latitude observatories compatible with both P_1^0 source structure and a localized 1-D mantle conductivity profile. Figure taken from *Weiss and Everett* [1998].

1.5   Evidence for the 3-D structure of Earth's mantle

Strong lateral heterogeneity has been detected in mantle electrical conductivity. Geographically distributed response functions have been calculated in the period range of 2 days to 1 year, under the assumption that the source fields arise from the relaxation phase of D_{st} activity. Such response functions have been found to be simultaneously consistent with regional 1-D Earth structure, and at mid-latitudes with cosine co-latitudinal dependence of the source potential [*Schultz and Larsen*, 1987]. Figure 1.14 presents the distribution of the subset of observatories studied in that work whose responses are compatible with both of these assumptions. It has been established formally that no global one-dimensional model exists that fits this global magnetic observatory data set to within statistical bounds [*Schultz and Larsen*, 1990]. Furthermore, analysis of one-dimensional conductivity sounding profiles calculated from magnetic observatory data reveals significant regional variability [*Schultz*, 1990].

It has also been shown that the contrast in electrical conductivity between oceans and continents is not primarily responsible for the observed geographic variability of long-

period geomagnetic data [*Kuvshinov et al.*, 1990]. In fact, the crust has a negligible effect on the resolution of large-scale mid-mantle anomalies [*Weiss and Everett*, 1998]. It is however a significant disturbance that needs to be taken into account when trying to reconstruct any lateral heterogeneities in the upper mantle (Section 1.6).

In the last fifteen years, deep magnetotelluric (MT) experiments have been conducted that have substantially increased the resolving power of EM investigations of the upper to lower mantle [*Uyeshima and Schultz*, 2000]. These have included multi-year experiments on continents using long electric dipoles (*Egbert et al.* [1992]; *Egbert and Booker* [1992]; *Schultz et al.* [1993]; *Bahr et al.* [1993]), and also seafloor experiments employing abandoned submarine telecommunications cables (*Lizarralde et al.* [1995]; *Utada et al.* [2003]; *Koyama et al.* [2003]; *Kuvshinov et al.* [2005]). For example, the latter have shown that although responses of each station or cable in the north Pacific are individually consistent with a 1-D model, the inability to jointly invert all response functions indicates inconsistency with a 1-D model. A correspondence between patterns of electrical heterogeneity and tectonic regime (e.g. *Lizarralde et al.* [1995]; *Schultz and Semenov* [1993]) has been observed. See also Figure 1.13 that clearly indicates that the electrical conductivity of the transition zone beneath northeastern China is at least an order of magnitude greater than at a number of other locations. Such experiments confirm that there are substantial levels of lateral electrical heterogeneity in the upper mantle. Petrological evidence, as well as laboratory results applied to samples of olivine of varying composition (e.g. *Hirsch and Shankland* [1993b]; *Hirsch and Shankland* [1993a]) also sustains this view.

Additionally, seismic tomographic studies (*Jordan* [1975], *Su and Dziewonski* [1991], *Su and Dziewonski* [1992], *Woodward et al.* [1993] and many others) have shown large lateral changes in the seismic velocities in the lower mantle. There are several possible causes for such heterogeneities, that include changes in chemical composition and temperature variations due to cool descending material and hot rising plumes [*Boehler*, 1996],

as well as the introduction of water and volatiles into mantle by descending slabs [*Lay*, 1994]. Each of these factors is a strong reason to expect 3-D heterogeneities in mantle electrical conductivity structure.

1.6 The near-surface inhomogeneities

One of the well-known effects, making the problem of global EM induction in the Earth significantly three-dimensional, is the so-called *S-effect*, which stands for the galvanic distortions of the fields due to near-surface inhomogeneities. The electrical conductivity of the uppermost layer of the lithosphere (Section 1.1), also known as the thin shell, varies a great deal, from as low as $\approx 10^{-5} Sm^{-1}$ for the resistive continental rocks, to $\approx 4 Sm^{-1}$ for sea water.

The realistic S-distribution of conductance (conductance being the multiplicative product of conductivity and thickness) has been used by *Kuvshinov et al.* [1999, 2002] and other workers to analyze the effect of ocean screening on the MT and GDS responses. An updated S-map using more detailed data in the European region and further refined sea water information has been recently offered by *Vozár et al.* [2006].

Studies have been performed to estimate the effect of the near-surface distortion on the modelling of the D_{st} and S_q fields. For example, *Kuvshinov et al.* [1999] have demonstrated that the variations of electrical resistivity in the uppermost 100 km of the deep structure significantly change the vertical magnetic surface field Z near oceans. However, ocean effects in the horizontal magnetic surface fields X and Y have been found to be quite negligible on land for S_q.

Recently, *Chandrasekhar et al.* [2003] have investigated the role of oceans in the electromagnetic induction along the western Pacific coast. Their calculations have shown a considerable influence of the self-induction effect of the oceans on the induction response

at 24-hr period, significant enough that it can not be ignored in oceanic regions.

Although the effect of the uppermost thin shell conductivity distribution for very long-period ($\gg$ 1 day) EM sounding should not be too significant, it is not clear *a priori* whether it could be ignored in EM induction studies at periods $\approx$ 1−10 days. Thus, a case can be made that the effect of the near-surface inhomogeneities at periods corresponding to the D_{st} and S_q fields is not negligible, and should be accounted for in the modelling of the Earth's magnetic field in order to ensure accuracy of the computations. We shall discuss this issue further in Chapters 7 and 8.

1.7 Advances in global geoelectromagnetic induction studies

In recent years, the availability of computational resources that were not even expected several decades ago has stimulated an extensive research in three-dimensional electromagnetic modelling. A number of regional and global forward solutions in spherical geometry have been developed and implemented. Notably, these include a semi-analytic method based on the perturbation expansion of conductivity about the background 1-D model [*Zhang*, 1991], the finite element method (FEM) approach of *Everett and Schultz* [1996] (and then *Weiss and Everett* [1998], *Yoshimura and Oshiman* [2002]), spectral finite element [*Martinec*, 1999]), spectral decomposition [*Grammatica and Tarits*, 2002], a heterogeneous spherical thin-sheet method based on integral equation approach [*Avdeev et al.*, 1997; *Koyama et al.*, 2003; *Kuvshinov and Pankratov*, 1994; *Kuvshinov et al.*, 2005], a staggered-grid finite difference formulation [*Uyeshima and Schultz*, 2000], as well as time domain techniques based on spherical harmonic expansions [*Hamano*, 2002; *Velimsky et al.*, 2003; *Velimsky and Martinec*, 2005] for computing 3-D EM fields of a transient external source (2-D time domain spectral finite-element technique has also been devel-

oped by *Martinec et al.* [2003]). Integral equation and other quasi-analytic methods bear inherent restrictions on either the size, or the shape, or the magnitude of the 3-D inhomogeneities. Therefore, although they are invaluable for many specific purposes, we shall not discuss them here in depth, except to describe their usage in validation of the other, more general, formulations. A comprehensive overview of the state-of-the-art fully 3-D global forward and inverse modelling can be found in *Avdeev* [2005]. Here, we shall only describe in detail the publications directly relevant to this work.

Kuvshinov and Pankratov [1994] have developed a numerical solution for the global induction problem for the buried shell (BS) model using an integral equation formulation [*Fainberg et al.*, 1993]. The solution exploits the modified iterative-dissipative method by *Singer* [1995] to solve an integral equation. The method is based on a Neumann series expansion, which always converges, even for extremely large conductivity contrast. A buried shell model comprises an inhomogeneous thin spherical sheet embedded at a given depth within a radially symmetric medium.

Martinec [1998] developed a semi-analytic method to obtain a solution of the electromagnetic induction problem for an Earth comprising azimuthally-asymmetric eccentrically nested spheres. The method is an extension of *Everett and Schultz* [1995] solution for axisymmetric eccentrically nested spheres. The axisymmetric solution had been used previously to partially validate the *Everett and Schultz* [1996] finite element forward solver. A significant difference between the azimuthally-asymmetric class of models and those considered previously is that, even with axially-symmetric (zonal) source current excitation, eddy currents are forced across and around the electrical conductivity interface, producing strong galvanic effects.

A numerical global three-dimensional frequency domain forward solver has been devised in 1995 − 2000 (*Uyeshima and Schultz* [2000]; *Toh et al.* [2002]). The solver is based on 3-D mesh division geometry in spherical coordinates, and employs a staggered-grid

finite difference formulation, a spherical analogue of the cartesian formulation of *Mackie and Madden* [1993a] and *Mackie and Madden* [1993b]. The induced magnetic fields are found as a solution to the integral form of Maxwell's equations, while the system of linear equations is solved using stabilised biconjugate gradient methods. The solver has been carefully tested and cross-compared against the numerical solution of *Kuvshinov and Pankratov* [1994] and the quasi-analytic solution by *Martinec* [1998], and assessed to be accurate and efficient enough to be an attractive kernel of a practical inverse solution. Developing the methodology for such a solution has been an important objective of this thesis.

Also, a novel 3-D 'spherical' forward solution has been elaborated by *Kuvshinov et al.* [2005]. The solution combines the modified iterative-dissipative method with a conjugate gradient iteration and allows one to compute efficiently the electromagnetic fields in full 3-D spherical models with very high lateral contrasts of conductivity and for very dense grids. A quasi-1D inversion had been developed on the basis of the model employing an iterative correction for the ocean effect (Figure 1.10).

Quite recently, much expanded magnetic observatory (*Fujii and Schultz* [2002]) and satellite (*Constable and Constable* [2004] and *Balasis et al.* [2004]) data sets of EM responses have been collected. All these studies are significant advances towards making the inverse problem of 3-D geomagnetic depth sounding tractable. In order words, the major objective of these studies is to enhance our understanding of the three-dimensional distribution of electrical conductivity in the mantle, which is currently only marginally known (Section 1.5). However, due to the inherent difficulties of this ill-posed non-linear inverse problem (Section 4.1) and the extremely high computational requirements, as well as the sparseness and the uneven distribution of the available data, the three-dimensional inverse problem has not previously been fully attacked. Although *Avdeev et al.* [1997] have developed a very fast and stable multi-sheet inversion allowing to determine the conduc-

tance of a buried inhomogeneous layer masked by sedimentary cover, that was limited by the very restrictive model space. The first global scale mantle conductivity inversion was based on inversion of the augmented *Schultz and Larsen* [1987] data set [*Schultz and Pritchard*, 1999]. This model was limited in spatial resolving power, and accurate only for solutions containing mild levels of lateral heterogeneity, so that the dense European observatory data were found to be unfittable with those assumptions. Also, a regional mantle scale joint inversion of submarine cable data and near-shore observatory data has been recently performed beneath the North Pacific [*Koyama et al.*, 2003], and the 3-D perturbations around the prior 1-D model of *Utada et al.* [2003] have been obtained in this region.

1.8 Format of thesis

The objective of Chapter 1 was to familiarize the reader with the current understanding of the Earth's magnetosphere, geoelectromagnetic induction methods and the electrical conductivity of the mantle.

In Chapter 2, we define the c and d responses of the Earth and describe the *Fujii and Schultz* [2002] data set used in this study.

In Chapter 3, we provide a detailed review and discussion of the governing equations and the global 3-D frequency domain forward solver by *Uyeshima and Schultz* [2000]. A numerical scheme is developed, that is later employed in Chapter 5 to formulate a generalised forward solver.

In Chapter 4, we describe the methodology developed to attack the corresponding

inverse problem. We discuss the notion of regularisation and define the regularised least squares penalty functional. We also introduce the concept of parametrization, discuss the parametrization by layered spherical harmonics, and describe regularisation by preconditioning appropriate for this setup.

In Chapter 5, we define and use the generalised forward solver to set up a numerical formulation for the full Jacobian and the derivative evaluation for the spherical global 3-D EM induction problem using both direct and adjoint methods.

In Chapter 6 we discuss a variety of optimisation techniques, and specifically the non-linear conjugate gradient inversion. This, in conjunction with other techniques, has been implemented using the derivative formulation obtained in Chapter 5.

We then perform a set of synthetic inversions to validate both the Jacobian computations and the inverse solution, including a series of checkerboard inversions for various degree and order perturbations. These computational experiments are described in Chapter 7.

Finally, the data set of Chapter 2 is inverted using the methods developed in Chapters 3–6, and the resulting 3-D variations in electrical conductivity of the Earth are presented in Chapter 8. We also present and discuss the predicted magnetic fields and responses at the Earth's surface. Chapter 9 contains the conclusions drawn from this work; as well as suggestions for future research.

For the sake of clarity we also include a list of figures, a list of notation and a summary of the parallel and serial software developed and employed in this project.

Chapter 2

Compilation of global geomagnetic responses of the Earth in frequency domain

This dissertation is centred on the inversion of a catalogue of global geoelectromagnetic response functions that span the period range of 5 to 106.7 days and that was calculated by *Fujii and Schultz* [2002]. The sources of the time domain data analysed by *Fujii and Schultz* [2002], comprising magnetic observatory three component hourly mean values of the geomagnetic field, are the British Geological Survey, the Chambon la Forêt Observatory, the collections of Gupta and Winch, the Geological Survey of Canada, the Intermagnet programme, World Data Centers A (NGDC), B, C1 and C2. The data were pre-processed by a statistically robust procedure and converted into the frequency domain. This catalogue is improved over previously published sets of deep mantle response functions in terms of quality, quantity and geographical coverage of the data. The data have also been corrected for auroral effects, and the resulting data set is thought to be consistent with the P_1^0 source assumption at geomagnetic latitudes extending to at least
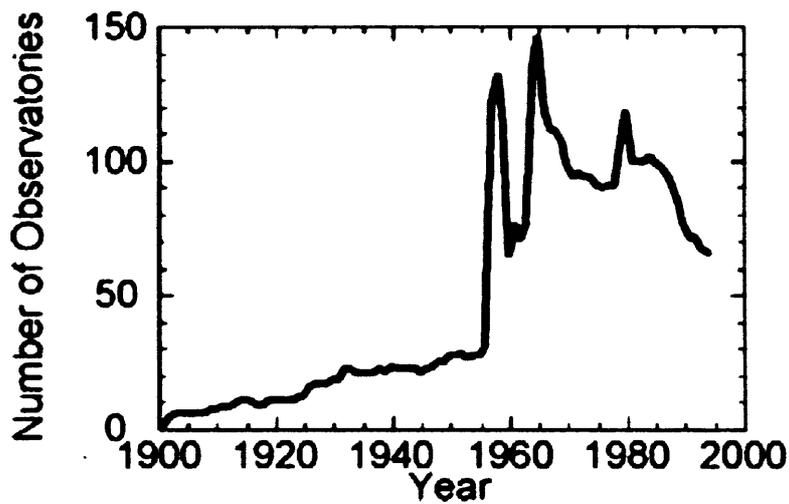
Figure 2.1: Distribution of geomagnetic observatories from 1900 to 1995 in time [*Fujii and Schultz*, 2002].

−60° South and 60° North. In this Section we describe the procedure that has resulted in these responses, give the definitions of the data functionals used and describe the space distribution of the resultant data set. Please refer to the *Fujii and Schultz* [2002] for any details regarding this study that are not covered in this brief overview.

2.1 Spatial distribution and analysis of the time series data

The hourly vector magnetic field values from 1957 to 1995 were used by *Fujii and Schultz* [2002] because of the availability of data from a particularly large number of observatories during this time period (Figure 2.1). The original time series were pre-processed and any gaps were filled using an iterative and statistically robust procedure of *Zhang and Schultz* [1990]. The solar quiet variations S_q and its harmonics, and 7 major oceanic tides, if detectable, were removed from the hourly data set by using a robust least-squares fitting method. The hourly data of horizontal force (H), declination (D) and downward

positive vertical component (Z) at 84 observatories were lowpass Hamming filtered with a non-recursive digital filter of 48 hour cut-off period and resampled into daily values. The annual and semiannual signals were removed from the daily values in the same manner. However, raw data quality in terms of continuity and length of embedded data gaps for six observatories was poor, prohibiting using these in the analysis by *Fujii and Schultz* [2002], leaving the total of 78 observatories in the data set. The locations of the remaining observatories are presented on Figure 2.2.

The North (B_x) and East (B_y) field components in geomagnetic coordinates were then computed from the H and D components in geographic coordinates. Locations of the geomagnetic pole for 1957 − 1995 were obtained from two dipole components of the IGRF model [*Barton*, 1997]. The secular motion of the pole is not relevant because this rate is much slower than the period range of interest in *Fujii and Schultz* [2002]. Finally, the mean and linear trend were removed from B_x, B_y and B_z (i.e. Z) from the entire time-series. The resulting data set was highpass filtered with Hamming filters of different cut-off periods and filter lengths and resampled into two data sets of 1 day and 5 day mean values, which were used for analysis at periods from 2 to 42.7 and from 10 to 106.7 days, respectively. After a statically robust[1] form of empirical orthogonal function (EOF) analysis had been performed [*Egbert*, 1997], the response functions were estimated from the resulting eigenvectors.

---

[1]It is important, in the analysis of a time series of responses, to use statistical procedures that are relatively insensitive to the presence of a moderate amount of bad data and poorly estimated errors, or to inadequacies of the model, and that react gradually rather than abruptly to perturbations of either [*Chave et al.*, 1987]. Such procedures are called *robust*, and have been developing since first introduced in 1953 by G. E. P. Box, presented in their present form by *Huber* [1981]. His definition is, that a procedure is deemed *robust* if its performance is insensitive to small deviations from the assumptions. More specifically, a statistical procedure is called *resistant* if the value of the estimate (or the test statistic) is insensitive to small changes in the underlying sample (either small changes in all, or large changes in a few of the values).
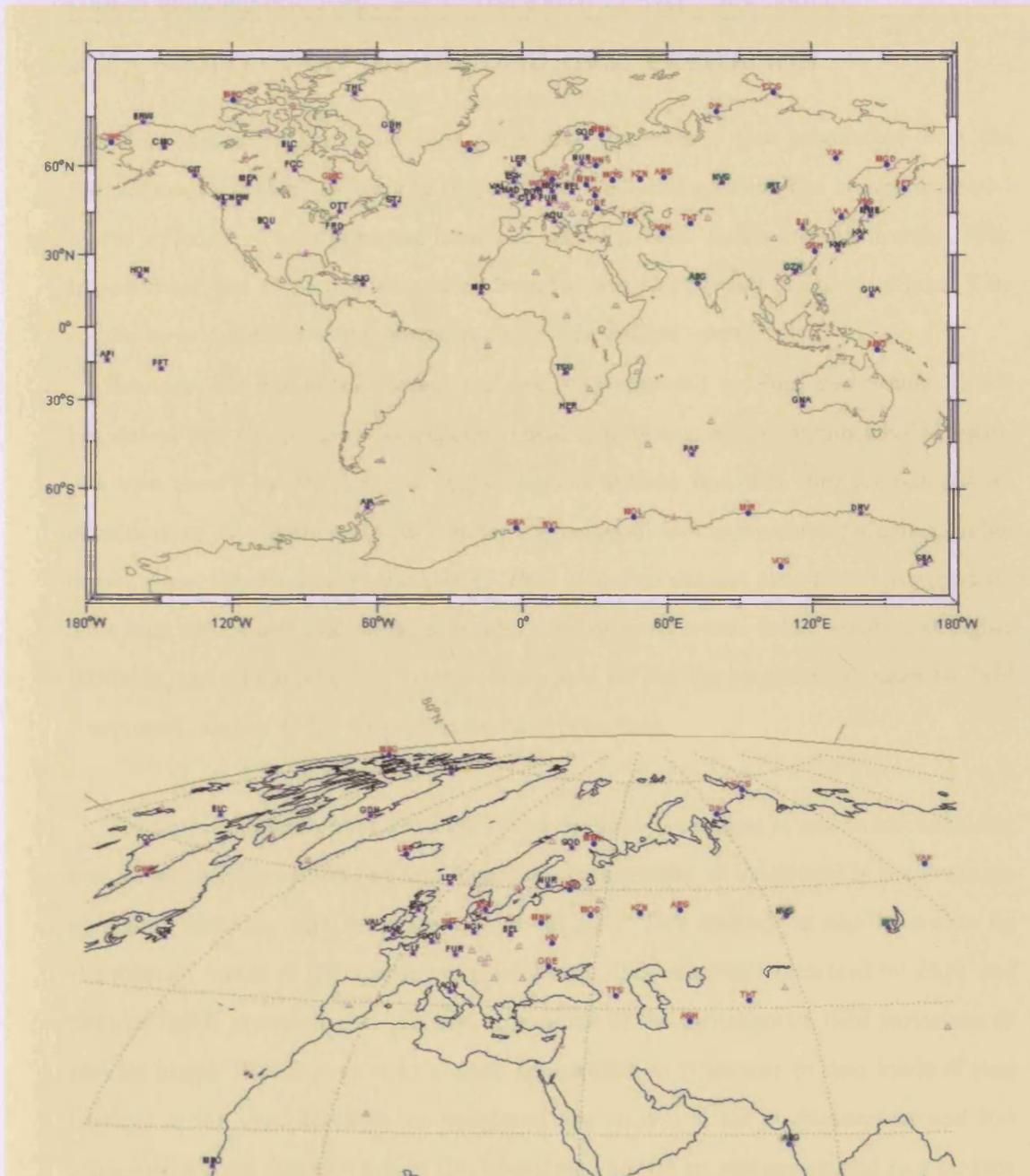
Figure 2.2: The 78 solid blue dots correspond to the *Fujii and Schultz* [2002] observatory locations; the 102 empty pink triangles correspond to IAGA (July 2006) observatory locations. Red letters denote observatories present in *Fujii and Schultz* [2002] but not in IAGA 2006 data set; green letters correspond to observatories whose geographic locations are shifted versus corresponding IAGA locations; black letters denote observatories that are common to both IAGA 2006 and *Fujii and Schultz* [2002]. We also provide a magnified view of the European observatories, on which the geomagnetic North pole is indicated with a yellow star.

2.2   Correction for auroral oval excitation

The first approximation to the external source structure is that originating from the Earth's magnetospheric equatorial ring current. This configuration can be approximated by the $P_1^0(\cos\theta) = \cos\theta$ spherical harmonic [*Schultz*, 1990; *Banks and Ainsworth*, 1992]. In particular, the P_1^0 source terms arise from the relaxation phase of D_{st} variations. This results in modulations of the magnetospheric ring current system.

However, the first mode eigenvector analysis performed by *Fujii and Schultz* [2002] has shown that the ϕ- and r- geomagnetic field components exhibit significant discrepancies with the P_1^0 model at higher latitudes. The authors find that they are not able to explain these discrepancies by an equatorially centered source function containing higher degree zonal spherical harmonic terms. They therefore suggest that these discrepancies arise from the influence of one or more additional external source fields, centered at higher latitudes, and whose influence extends down to $\sim 50°$ for the longitudinal magnetic field component, and to $\sim 40°$ for the vertical magnetic field.

The authors observe, that while the P_1^0 variation predominates at mid to low latitudes, a more localised variation of zonal form (i.e. independent of longitude) is predominant at higher latitudes, and is centered at about 67°. This latitude is also known to be the average centre of the statistical auroral oval. The analysis performed by *Fujii and Schultz* [2002] shows that 1) the first eigenmode of the geomagnetic field variations at periods longer than 5 days is to a good approximation generated by two kinds of ring currents in the same direction; an equatorial ring current in the magnetosphere and two conjugate auroral ring currents in the ionosphere, and 2) an average auroral ring current has stationary effects in the geomagnetic field variations at surprisingly long periods. This assumption has been verified by fitting curves to the data, and it has been heuristically found that the best fitting model is one setting the height h such that $a = R + h$ of
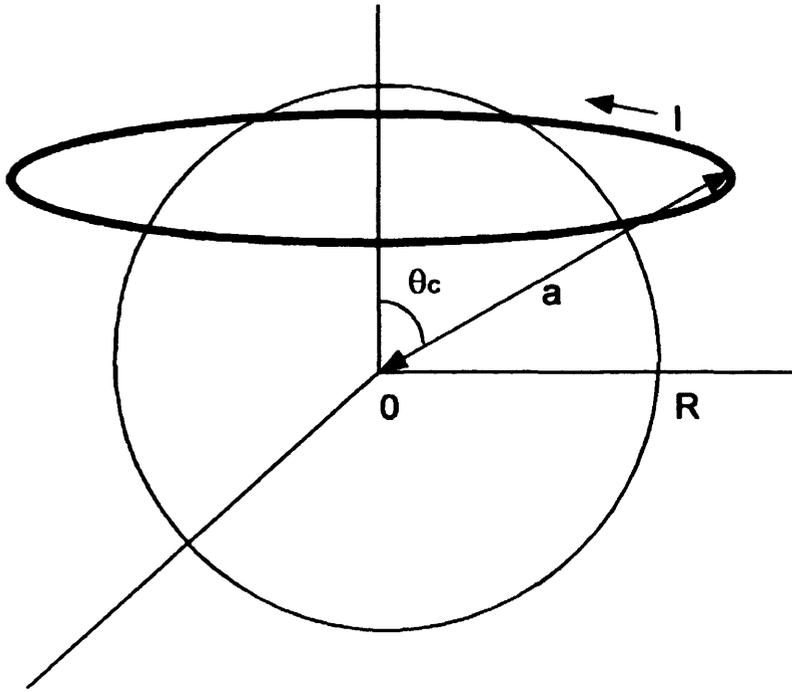
Figure 2.3: Schematic of a ring current I at colatitude of θ_c and distance from the origin a for a spherical Earth of radius R (as described by *Fujii and Schultz* [2002])

the auroral ring current from the surface of the Earth at approximately 300 km. In this model, the auroral ring current is offset a distance of $\theta_c = 23°$ along the rotational axis from the plane of the equator.

2.3 C and D response functions

According to the geomagnetic depth sounding (GDS) on a point method the scalar c response [*Banks*, 1969; *Schmucker*, 1970; *Schultz and Larsen*, 1987] at a single observatory and a single angular frequency ω (assuming an inducing source of D_{st}-type, so that the equatorial ring current is the primary source of the geomagnetic variations) may be defined as

$$c(\omega) = \frac{a\tan(\theta)}{2} \frac{B_r(\omega)}{B_\theta(\omega)} \tag{2.1}$$

where a is the radius of the Earth in metres, θ is co-latitude in radians, B_r is the vertical component of the magnetic field pointing downwards to the centre of the Earth, and B_θ is the co-latitudinal component pointing to magnetic South. Please note that this (spherical grid) notation relates to the standard geomagnetic field notation as $B_r \equiv -B_z$ (B_z is upward positive), $B_\theta \equiv -B_x$ (B_x points to magnetic North) and $B_\phi \equiv B_y$ (both pointing to the East). [2]

This definition of c response has been devised for a 1-D Earth, and is not very sensitive to 3-D deviations from this model. Since the term $\tan(\theta)$ compensates for the spatial structure of the source, the c response due to a zonal excitation should be the same everywhere on the surface of such a 1-D Earth. Such a c response has units of length and a positive real part, and provides a convenient means to estimate skin depth.

To increase the sensitivity to lateral inhomogeneities, a scalar response form complementary to c response has been proposed in *Fujii and Schultz* [2002]:

$$d(\omega) = \frac{a \sin(\theta)}{2} \frac{B_\phi(\omega)}{B_\theta(\omega)} \tag{2.2}$$

where a is the radius of the Earth in metres, θ is co-latitude in radians, B_ϕ is the longitudinal component of the magnetic field pointing to magnetic East, and B_θ is the co-latitudinal component pointing to magnetic South.

The d response is designed to extract information on 3-D inhomogeneities in conductivity whilst retaining the site independent properties of the local c response. The ϕ component of the geomagnetic field is zero if the mantle is 1-D and the source is zonal. Therefore, non-zero values of d response arise under such excitation only in the presence of lateral conductivity heterogeneities.[3]

---

[2]To express the c response in the (H,D,Z) coordinates, need to bear in mind that H points to the North while Z points down; hence the sign of the expression is negative.

[3]This work is only concerned with the inversion of the *Fujii and Schultz* [2002] c response values; d

2.4 Estimated errors in C responses

The smallest error ever encountered in the *Fujii and Schultz* [2002] data set is $\sim 3\%$ of

the absolute value of the respective measurement, while for some frequencies the smallest

measurement error is as large as $\sim 7\%$. On the other hand, the upper bound on the error

is quite flexible. In the *Fujii and Schultz* [2002] data set, the maximum error for various

frequencies ranges from 83% to as large as 1667% of the measurement, and the average

error in the data set constitutes 24.23% of the respective data values. It is notable that for

the mid-latitude data set ($-60° \le \theta \le 60°$, Section 2.5) the average error is 14.87%, with

the range being 3% to 285%, approximately. In total over all frequencies, 233 out of 2507

errors exceed 50%; corresponding to the ratio 43 out of 1767 for the mid-latitude data

set. All this suggests that the majority of the grossly inaccurate measurements have been

made close to the geomagnetic poles. In fact, from the graph on Figure 2.4 we can see

that apart from values at several observatories close to the equator and a single exception

– an inaccurate observatory at about 52° GM latitude, the majority of observations made

in the range $-75° \le \theta \le 75°$ are rather accurate.

It is often a good idea to use the hard rejection or bounded influence[4] least squares

functionals instead of the conventional form of least squares to downweight the influence

of too noisy data, as well as the data the error estimates of which are biased downward,

in the inversion. If the data errors are too optimistic, there is a danger of overfitting the

measurements at the expense of introducing unnecessary features into the model. On the

other hand, a handful of extremely noisy data may relax our convergence criterion too

much. Since we add the error weighted residuals together in the penalty functional, small

residuals may arise due to large measurement errors rather than to the goodness of fit. If

---

responses are not inverted. This remains a task for the future research. However, all the necessary work
has been completed to enable a joint inversion of these two data sets.

[4]The bounded influence least squares refers to a least squares functional that replaces all extreme
data errors (greater or smaller than a threshold) by the respective threshold values. The hard rejection
criterion ignores data values whose errors are greater or smaller than a threshold.
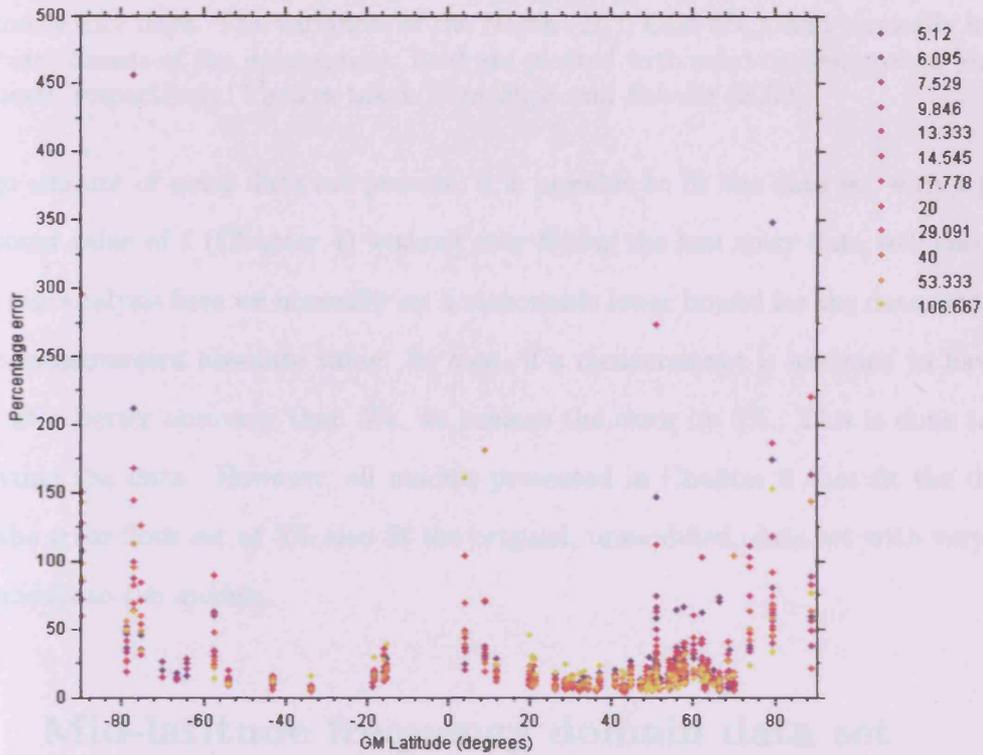
Figure 2.4: Percentage error in the *c* response *Fujii and Schultz* [2002] data set plotted against GM latitude, for every third frequency. The colour denotes the period (in days).
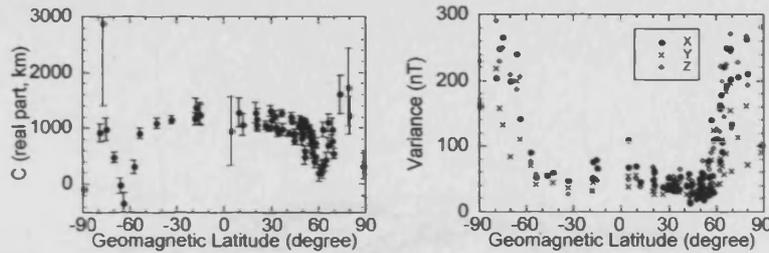
Figure 2.5: The real part of complex c response which are estimated from the first EOF mode (left) and estimates of the noise variance (right) computed for the daily data set at a period of 42.7 days. The variances of the North (B_x), East (B_y), and vertically upwards (B_z) components of the geomagnetic field are plotted with solid circles, crosses and open diamonds, respectively. Figures taken from *Fujii and Schultz* [2002].

a large amount of noisy data are present, it is possible to fit the data set with a penalty functional value of 1 (Chapter 4) without ever fitting the less noisy data well enough.

In our analysis here we normally set a reasonable lower bound for the data error of 5% of the measurement absolute value. So that, if a measurement is assumed to have been taken at a better accuracy than 5%, we replace the error by 5%. This is done to avoid overfitting the data. However, all models presented in Chapter 8 that fit the data set with the error floor set at 5% also fit the original, unmodified, data set with very minor refinements to the models.

2.5 Mid-latitude frequency domain data set

The resultant frequency domain data set comprises a set of c and d responses calculated at 78 observatories for 34 different periods ranging from 5 to 106.7 days. The assumption that the auroral ring currents are located at 300 km height has been found the most suitable [*Fujii and Schultz*, 2002]. The auroral effect calculated under this assumption has been removed from the resulting data set.

Since both the auroral and equatorial current systems are zonal, the B_x and B_z components of the first mode are affected by subtracting the reference field. Considering the
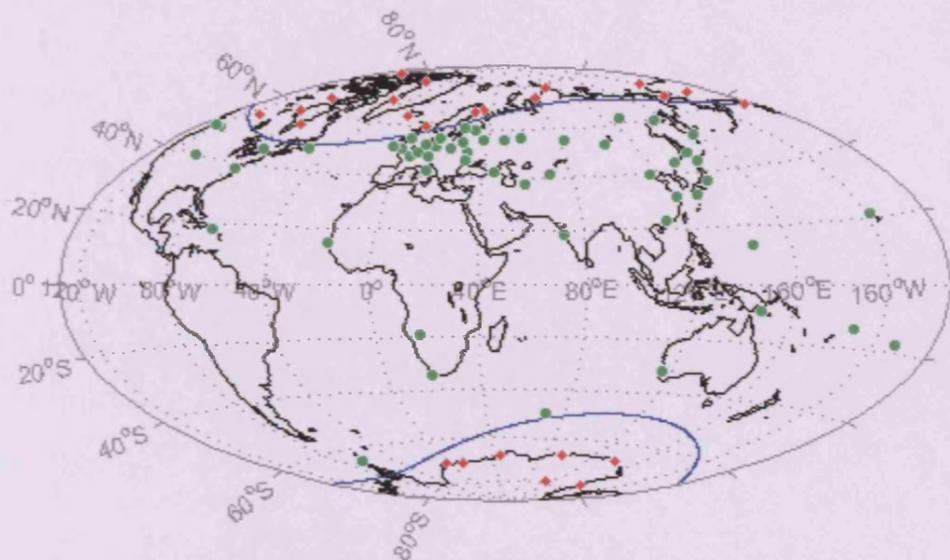
Figure 2.6: Shown are the locations of (78 in total) geomagnetic observatories from *Fujii and Schultz* [2002] data set. The blue lines correspond to 60° geomagnetic latitude. The 24 observatories located closer than that to geomagnetic poles are denoted with red diamonds. The other 54 observatories denoted by green solid circles are used in the analysis presented in this thesis.

large scatter in the first mode and poor data quality at high geomagnetic latitudes, as implied in Figure 2.5, in this study data were used only from observatories whose geomagnetic latitude was restricted to $-60° < \theta < 60°$ (Figure 2.6). Amplitudes of anomalous B_x and B_y are comparable within this latitude band. Attempts were made to expand on this band of latitudes. We found that the ability to fit the observatory data rapidly degenerated (Chapter 8).

Chapter 3

Forward problem of geoelectromagnetic induction in a heterogeneous sphere: setup and a 3-D staggered-grid finite difference numerical solution

In the Introduction, we have outlined the importance of being able to image the heterogeneities (and eventually the 3-D structure) of the electrical conductivity in the Earth's mantle. However, to make inferences about the interior structure, we require a comprehensive way of *mapping the interior to a quantity that can be measured* on the surface of the Earth. Such a mapping would be called a forward solution. We could then apply the methods of *inverse theory* (Chapter 4) to reconstruct some or all of the features of electrical conductivity distribution in the mantle. Due to the non-linearity arising in any forward solution that addresses this problem in three dimensions, this task is far from triv-

ial. In particular, the forward mapping is complicated enough that an analytic solution can not be written, and the problem has to be addressed numerically. Hence designing an accurate and computationally efficient forward solver is key to a successful inversion.

3.1   Overview of this chapter

In this chapter we aim to provide a comprehensive summary of the methods employed by *Uyeshima and Schultz* [2000] (henceforth referred to as U&S), that could be used as a reference; as well as to introduce some ideas that will be encountered further in this thesis. In particular, in Chapter 5 the numerical approach described here will be further generalised to compute the data sensitivities for an arbitrary forcing, and then employed as the core of a gradient-based global search algorithm described in Chapters 4 and 7.

In Section 3.2 of this Chapter we derive from first principles the system of equations commonly known as the *quasi-static approximation* of Maxwell's equations, that forms the basis of geo-electromagnetic exploration in the frequency domain. We would like to follow this argument here in order to clearly outline all the inherent assumptions of this approximation. Eventually we shall derive and describe the integral form of Maxwell's equations in the frequency domain, that constitutes our forward problem. References that cover in greater detail the physics behind the equations described in this section include *Jackson* [1975] and *Zhdanov* [1986].

In Section 3.3 we complete the setup of the problem. In particular, we outline the computational domain and the boundary conditions and describe the assumed external source configuration. We also provide reasons for choosing this particular setup.

In the following Section 3.4 we describe in detail the spherical staggered-grid finite

difference geometry of the computational domain. We write out expressions for the unit lengths and surface elements and give examples of typical calculations in the framework of this geometry. We also define some notation that is going to be actively employed in future chapters.

Then we provide a brief overview of the difference equations that approximate the system of integral equations given in Section 3.2. This is the system of equations that have been implemented by U&S. We also provide the linear algebraic formulation of these difference equations. This analysis constitutes Section 3.5. In the future we shall refer to the system of linear equations described here when we discuss the generalised forward solver in Chapter 5.

We conclude with an overview of the iterative method employed to solve this system of linear equation numerically. We describe the properties of this linear system in Section 3.6. A variant of the biconjugate gradient iterative relaxation method, called the Bi-Conjugate Gradient Stabilised (BiCGSTAB), has been employed. Better precision is achieved by repeatedly performing a correction of the magnetic field to ensure zero divergence. The algorithm has been developed and implemented by M. Uyeshima, A. Schultz and H. Toh, and then modified by A. Kelbert.

3.2 Quasi-static electromagnetic field model

The system of Maxwell's equations applicable to geomagnetic exploration (in the SI units) can be written as:

$$\text{Ampère's Law:} \qquad \nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t} + \mathbf{J} + \mathbf{J}^{ext} \qquad (3.1)$$

$$\text{Faraday's Law of Induction:} \qquad \nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad (3.2)$$

$$\text{Absence of magnetic monopoles:} \qquad \nabla \cdot \mathbf{B} = 0 \qquad (3.3)$$

$$\text{Gauss's Law:} \qquad \nabla \cdot \mathbf{D} = q + q^{ext} \qquad (3.4)$$

where $\mathbf{H}$ and $\mathbf{B}$ respectively denote *magnetic field strength* and *magnetic flux density* (also called *magnetic induction*), $\mathbf{E}$ and $\mathbf{D}$ are respectively the *electric field* and the *electric displacement field*, $\mathbf{J}$ is the *current density*, and q is the *electric charge*. The suffix "*ext*" relates to the external currents and charges, by which we mean currents and charges created in any way other than electromagnetically: e.g. those of mechanical or chemical nature. Such currents and charges affect the electromagnetic fields, being themselves maintained by other means. They should therefore be considered separately.

The following linear constitutive relationships are also assumed:

$$\mathbf{D} = \varepsilon \mathbf{E} \qquad (3.5)$$

$$\mathbf{B} = \mu \mathbf{H} \qquad (3.6)$$

$$\text{and} \quad \mathbf{J} = \sigma \mathbf{E} \qquad (3.7)$$

where ε is the *electrical permittivity* of the material, μ is the *magnetic permeability* and σ is the *electrical conductivity*. In non-dispersive, isotropic media, we can regard ε and μ as time-independent scalars. Thus in such media we can immediately assert that $\dfrac{\partial \mathbf{D}}{\partial t} = \varepsilon \dfrac{\partial \mathbf{E}}{\partial t}$ and that $\dfrac{\partial \mathbf{B}}{\partial t} = \mu \dfrac{\partial \mathbf{H}}{\partial t}$.

If we now assume *harmonic time dependence* of the form $e^{i\omega t}$, where ω is the angular frequency and t is time, then the equations 3.1 and 3.2 turn into the frequency domain

Maxwell's equations

$$\nabla \times \mathbf{H} = iw\varepsilon\mathbf{E} + \mathbf{J} + \mathbf{J}^{ext} \tag{3.8}$$

$$\nabla \times \mathbf{E} = -iw\mu\mathbf{H}. \tag{3.9}$$

It is also worth stating that equation 3.3 implies that $\nabla \cdot (\mu\mathbf{H}) = \mu\nabla \cdot \mathbf{H} + \mathbf{H} \cdot \nabla\mu = 0$. Therefore in the regions where $\nabla\mu = 0$ is a reasonable assumption, $\nabla \cdot \mathbf{H} = 0$. In this work we exploit the assumption that in the Earth/air region *no ferromagnetic materials are present* in large enough quantities to be significant, so that we set $\mu = \mu_0 = \text{const}$ throughout the region (in which case $\nabla \cdot \mathbf{H} = 0$ certainly holds). Under similar assumptions on the electrical permittivity, an expression for $\nabla \cdot \mathbf{E}$ can also be obtained. However, we will not require it here.

We also note that since we're solving the equations in the whole Earth/air domain (see 3.3 for the complete setup of the problem), we can also assume that there are *no external currents or charges*, so that $\mathbf{J}^{ext} = 0$ and $q^{ext} = 0$.

This allows us to obtain from 3.8, after the substitution of $\mathbf{J} = \sigma\mathbf{E}$,

$$\nabla \times \mathbf{H} = iw\varepsilon\mathbf{E} + \sigma\mathbf{E} \tag{3.10}$$

$$\nabla \times \mathbf{E} = -iw\mu_0\mathbf{H}. \tag{3.11}$$

Here the term $iw\varepsilon\mathbf{E}$ stands for the displacement currents, while $\mathbf{J} = \sigma\mathbf{E}$ represents the conduction currents in the system.

However, *for frequencies small enough*, $w\varepsilon\mathbf{E} \ll \sigma\mathbf{E}$, and the displacement currents

may be neglected. Thus we obtain Maxwell's equations in the *quasi-static* approximation,

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{3.12}$$

$$\nabla \times \mathbf{E} = -i\omega\mu_0\mathbf{H} \tag{3.13}$$

$$\mathbf{J} = \sigma\mathbf{E}. \tag{3.14}$$

The fields satisfying these equations are called *monochromatic*, meaning harmonic in time. Any time-dependent EM field can be represented as a superposition of monochromatic fields, since every function can be written in terms of its Fourier series or Fourier integral of its frequency components.[1] This allows us to work in the frequency domain, which simplifies the analysis of the fields and the computations in general, since the amount of data is greatly reduced.

It is also obvious from the equations 3.12-3.14, that the quasi-static EM fields are independent of the electric permittivity ε of the medium.

It is sometimes useful to separate the fields to solve for either the electric or the magnetic fields separately. In particular, this chapter describes a solution in terms of the magnetic fields. To obtain the equation on the magnetic fields only, we multiply Equation 3.12 by $\rho = 1/\sigma$, a quantity known as the *electrical resistivity*, and take the curl of both sides. After substituting the result into 3.13, we obtain the *vector Helmholtz equation* on $\mathbf{H}$:

$$\nabla \times (\rho\nabla \times \mathbf{H}) + i\omega\mu_0\mathbf{H} = 0. \tag{3.15}$$

Note also that, according to the arguments above, in this setting $\mathbf{H}$ is divergence-free.

---

[1] The reason one can separate time and space is that Maxwell's equations are time invariant (i.e. the spatial operator is *autonomous* – has no dependence on time). Hence one can Fourier transform the time dependence.

However, this is not very well enforced numerically when Equation 3.15 is being solved for a small frequency ω. Indeed, if an irrotational part $\mathbf{H}_i$ is added to the magnetic field (so that $\nabla \times \mathbf{H}_i = 0$ and $\nabla \cdot \mathbf{H}_i \neq 0$) it is annihilated by the first curl in the Helmholtz operator of Equation 3.15; if ω is small enough, the magnetic field still almost satisfies Equation 3.15. Enforcing the additional divergence free condition $\nabla \cdot \mathbf{H} = 0$ suppresses this computational mode, and thus can be used to improve the accuracy of an iterative numerical solver. This method has been introduced by U&S under the name of *divergence correction*.

A similar vector Helmholtz equation is derived for $\mathbf{E}$, applying curl operator to both sides of equation 3.13, and substituting 3.12 (here we are once again making use of the assumption that $\mu = \mu_0$ is constant in the region):

$$\nabla \times \nabla \times \mathbf{E} + i\omega\mu_0\sigma\mathbf{E} = 0. \tag{3.16}$$

The quasi-static Maxwell's equations can also be written in the integral form, using Stokes' theorem $\iint_S (\nabla \times \mathbf{A}) \cdot \mathbf{n}dS = \oint_C \mathbf{A} \cdot dl$, and defining $d\mathbf{S} = \mathbf{n}dS$:

$$\oint \mathbf{H} \cdot dl = \iint \mathbf{J} \cdot d\mathbf{S},$$

$$\oint \mathbf{E} \cdot dl = -\iint i\omega\mu\mathbf{H} \cdot d\mathbf{S}, \tag{3.17}$$

$$\mathbf{J} = \sigma\mathbf{E}.$$

It is easy to see that $\mathbf{H}$ is a non-linear mapping of parameter ρ - the electrical resistivity. Still, if we choose to solve this system of equations for $\mathbf{H}$, it makes more sense to parametrize with respect to ρ than with respect to $\sigma = 1/\rho$, since $\nabla \times (\rho\nabla \times \mathbf{H})$ is a linear

mapping of ρ, but not of σ. Let us concentrate further on the integral form of Maxwell's equations, since this is the formulation that has been discretized and implemented by U&S. In the following sections we cover the precise computational assumptions and the discretization for the staggered-grid finite difference mesh.

3.3 Setup of the forward problem

The computational domain and boundary conditions described here closely follow those used in the U&S paper, since the forward solver U&S have developed is used as the core of the inverse problem in this work. However, the assumptions on the source structure will later be reviewed (please refer to the generalised equations in Chapter 5) as we will find that these assumptions are too restrictive for our purposes while computing the data sensitivities. In fact, we will see from the theory developed alongside these generalizations, that only a few relatively simple operators need to be implemented to make inversion with respect to the boundary conditions, along with the resistivity distribution, feasible. In this chapter, however, we only cover the simple P_1^0 external source field assumption introduced by U&S. We are not employing more physically realistic source field models accounting for the auroral sources [*Fujii and Schultz*, 2002] since the data used in the inversion have already been corrected for the presence of auroral current systems (Section 2.2).

3.3.1 Computational domain and boundary conditions

We solve the Helmholtz equation 3.15 for the magnetic field **H** in the computational domain including the resistive air and conductive Earths' crust and mantle, the bottom of the domain being the core-mantle boundary (CMB). We complete our forward problem by assigning values to the tangential components of the magnetic field **H** at the upper and lower boundaries of the domain. This gives us a formulation for a Dirichlet boundary

value problem.

The magnetic permeability is assumed to be that of the vacuum, $\mu = \mu_0 = 4\pi \times 10^{-7}$ H m^{-1} throughout the domain.

Following *Mackie and Madden* [1993b], the authors of U&S paper find it convenient to set the electrical conductivity of the air to a moderately small non-zero value of 10^{-10} S m^{-1}. By so doing, the Helmholtz equation holds throughout the model domain, while the matrix system that approximates the equation numerically remains acceptably well-conditioned. Were the usual approach chosen, with $\sigma_{air} = 0$ (e.g. *Everett and Schultz* [1996]), then Laplace's equation would hold in the air layer, while Helmholtz equation would hold in the conducting Earth below. This would present the complication of having to match the Earth/air interface boundary conditions.

Numerically, the lower boundary can be set at or just below the CMB (~ 2890 km). In the core, the conductivity is set to be infinite, which has the effect of forcing **H** within the core's interior to vanish. We therefore set the tangential components of **H** to zero at the bottom of our computational domain. Although just above the core-mantle boundary **H** $\neq 0$, in practice the strong attenuation of externally induced fields in the conducting mantle leads all components of **H** in the vicinity of CMB to be approximately zero.

On the top of the air layers (the upper boundary of the domain) an appropriate source configuration is set by assigning values to the tangential components of the **H** field, and hence to their contour integrals. The air layer should be extended far enough above the Earth for all of the internally induced field perturbations to be damped out at the location of the source. In practice the top of the air layer is set at the 19113.1 km distance from the Earth's centre. The paper by U&S describes the location of the source field at

$r = 10R_e$ radius. The issue of choosing reasonable source fields and a reasonable height above the Earth's surface for the outer boundary is further addressed in section 3.3.2.

3.3.2 P_1^0 assumption on the source fields

As in the *Uyeshima and Schultz* [2000] paper, the equatorial ring current is approximated by a P_1^0 harmonic of unit strength placed at a radial distance from the Earth's surface of $r = 10R_e$, where R_e is the Earth's radius, i.e. at the upper boundary of the computational domain. At this distance, secondary magnetic fields induced by the presence of the conductive Earth are damped out to $< 10^{-3}$ of the external field intensity and may be considered negligible (U&S, Appendix).

3.4   Staggered-grid finite difference formulation

To solve numerically the 3-D spherical Dirichlet boundary value problem, described by the system of integral equations 3.17, a mesh division geometry needs to be set up on the computational domain. For the particular problem described here, it is in many ways advantageous to use a staggered-grid finite difference formulation (please see U&S for reference). Here we are planning to describe the spherical staggered-grid finite difference approach taken in U&S. We shall only review the basic facts; the reader can refer to U&S for any additional details. We will utilize this geometry as the foundation for the linear vector theory formulations in the Chapter 5 and for the specific operator implementations.

Let L, M and N be the user-specified grid dimensions in the longitudinal, latitudinal
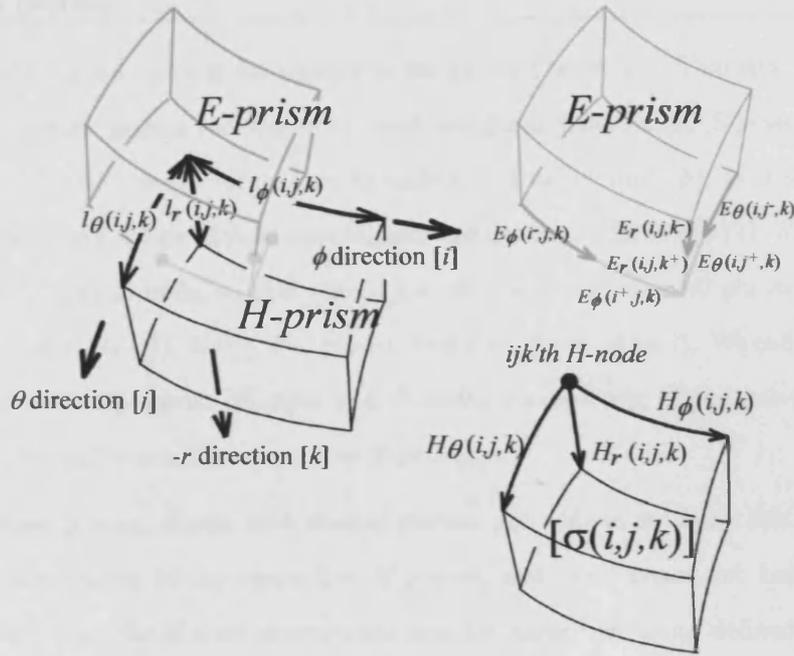
Figure 3.1: Staggered-grid finite difference formulation designed for the 3-D geoelectromagnetic induction problem in spherical coordinates [*Uyeshima and Schultz*, 2000]

and radial directions, respectively. Hereafter, let $\phi(i)$, $\theta(j)$ and $r(k)$ be the (user-specified) longitudes, colatitudes and distances from the Earth's surface of the primary grid nodes. So that, index i ranges from $[1, L]$, j from $[1, M + 1]$ and k from $[1, N + 1]$. Then the vectors ϕ, θ and r fully specify the grid. We choose the convention that index i is numbered from west to east, with zero meridian $\phi = 0$ associated with $i = 1$ and $i = L + 1$. Index j increases as we go from the North pole to the South along a meridian, with $j = 1$ corresponding to the North pole ($\theta = 0$) and $j = M + 1$ corresponding to the South pole ($\theta = \pi$). Index k is numbered from the top of the air layer, where $k = 1$, down, so that $k = N + 1$ corresponds to the lower domain boundary just below the CMB. This numbering convention is chosen to relate the fields to right-handed coordinates.

The whole computational domain including the resisting air and conducting Earth is

divided into curved rectangular prisms as shown on Figure 3.1. Two sets of prisms are defined: primary prisms (or *H-prisms*) and staggered (also called *E-prisms*). Primary prisms, drawn with solid curves, are bounded by longitudinal (ϕ), latitudinal (θ) and radial edges along which **H** field components are defined. The (i, j, k)'th *H-prism* builds up on the (i, j, k)'th node, so that the (i, j, k)'th prism embodies all points (ϕ, θ, r) such that $\phi(i) \leq \phi < \phi(i+1)$, $\theta(j) \leq \theta < \theta(j+1)$ and $r(k) \geq r > r(k+1)$. We call these prisms, edges and nodes *H-prisms*, *H-edges* and *H-nodes*, respectively. Conductivity values are defined to be uniform in the respective *H-prisms*.

Staggered prisms, drawn with shaded curves, are defined so that their nodes are located at the centres of the respective *H-prisms*, and their edges cut halfway through *H-surfaces*. Thus the **E** field components may be viewed as being defined on the edges of *E-prisms*. Alternatively, as can be seen on Figure 3.1, the **E** field components can be defined on a surface of an *H-prism*: they are defined on a vector perpendicular to that surface, and passing through the midpoint of the surface.

In the integral equations 3.17, $d\mathbf{l}$ is a line element and $d\mathbf{S} = \mathbf{n}dS$ is a surface element that is defined on a surface enclosed by the contour of $d\mathbf{l}$ and orthogonal to the element of the surface. The sign convention for the direction of **n** is taken so that **n** and $d\mathbf{l}$ satisfy the right-hand rule. We can see that the length and surface elements defined in this way may be approximated with the edges and the surfaces of the appropriate *H-prisms* (for computing discrete implementations of $\oint \mathbf{H} \cdot d\mathbf{l}$ and $\iint \mathbf{H} \cdot d\mathbf{S}$), or *E-prisms* (for $\oint \mathbf{E} \cdot d\mathbf{l}$ and $\iint \mathbf{J} \cdot d\mathbf{S}$, respectively).

The edge and surfaces of the primary *H-prisms* may be calculated using basic rules of spherical geometry, so that the length of the ijk'th *H-edges* in the respective directions are given by

$$l_\phi(i,j,k) = r(k)\sin\theta(j)[\phi(i+1) - \phi(i)], \tag{3.18}$$

$$l_\theta(i,j,k) = r(k)[\theta(j+1) - \theta(j)], \tag{3.19}$$

$$l_r(i,j,k) = r(k) - r(k+1), \tag{3.20}$$

and the *H-surfaces* normal to the respective directions are calculated by

$$S_\phi(i,j,k) = \frac{1}{2}[r(k)^2 - r(k+1)^2][\theta(j+1) - \theta(j)], \tag{3.21}$$

$$S_\theta(i,j,k) = \frac{1}{2}[r(k)^2 - r(k+1)^2]\sin\theta(j)[\phi(i+1) - \phi(i)], \tag{3.22}$$

$$S_r(i,j,k) = r(k)^2[\cos\theta(j) - \cos\theta(j+1)][\phi(i+1) - \phi(i)]. \tag{3.23}$$

For the calculations and notation related to *E-prisms*, U&S have relatively complicated notation with "+" and "-" signs, where e.g. the triple (i^+, j, k) represents the indices of the midpoint of the (i,j,k)'th *H-edge* in the ϕ-direction, and (i^+, j^-, k^-) is the index of the *E-node* of the prism centered at the $(i+1, j, k)$'th *H-node*.

Given this notation, the length of an *E-edge* is indexed by the edge's midpoint and direction (e.g. $l_\theta(i^+, j, k^+)$), and an *E-surface* is indexed by the respective *E-node* and perpendicular direction (e.g. $S_\phi(i^+, j^-, k^-)$).

U&S also point out, that it would be ambiguous to specify the values of electric field component on *E-edges*, since the electric field **E** is discontinuous across the boundary of *H-prisms* of different conductivity, whilst electric current **J** is continuous. They solve this problem by decomposing an electric field component along an *E-edge* onto two parts: the "-" component, meaning the part that is outside the *H-prism* of uniform conductivity $\sigma(i,j,k)$, and the "+" component, meaning the part inside that prism.

In this document we adopt a slightly different notational convention to that in U&S which, as we shall see, simplifies the linear algebraic formulation of the problem. We define the (i,j,k)'th *E-prism* to be the prism centered at the (i,j,k)'th *H-node*, so that we could refer to (i,j,k)'th *E-edges* and *E-surfaces* in much the same way as those of *H-prisms*.

Consider a single *H-loop*, which is the loop of four *H-edges* enclosing a surface. For this loop, we can compute an approximation to $\oint \mathbf{H} \cdot d\mathbf{l}$ by multiplying the elements of $\mathbf{H}$ by the respective length elements, and summing them up in the order defined by the sense of integration around contour in a right-hand coordinate system. The resultant vector is defined on an *E-edge* passing through the *H-surface*. We then obtain a local approximation of $\mathbf{J}$ on this *E-edge*, by dividing the resultant value by the surface area (thus implementing $\iint \mathbf{J} \cdot d\mathbf{S} / \iint d\mathbf{S}$ locally), according to the first of integral equations 3.17.

Now we use this value of $\mathbf{J}$ on the *E-edge* to obtain $\mathbf{E} = \rho\mathbf{J}$, where $\rho = 1/\sigma$ is interpolated to the *H-surface* on which this value of $\mathbf{J}$ is defined. Then $\mathbf{E}$ is also continuous across the surface, just as $\mathbf{J}$ is. In doing so, we have made use of the third equation in 3.17.

Having obtained the values of $\mathbf{E}$, we implement $\oint \mathbf{E} \cdot d\mathbf{l}$ by multiplying the elements of $\mathbf{E}$ by their respective *E-edge* lengths, and summing up the four elements belonging to the surfaces that share a single *H-edge*. Summing up the resultant values on *H-edges*, and equating the sum to the discrete implementation of $-i\omega\mu \iint \mathbf{H} \cdot d\mathbf{S}$ (second equation in 3.17), we obtain the full difference equation given in U&S (e.g. equation 18).

For details and implementation of each of the intermediate steps, please refer to Sections 3.5 and 5.4. Although we are not rewriting the full forward solver from this point of view, we have found that these intermediate steps need to be implemented if an efficient

view, we have found that these intermediate steps need to be implemented if an efficient data sensitivity analysis is to be developed.

Staggered-grid finite difference formulation has previously been applied to Cartesian problems in geophysics (e.g. *Mackie and Madden* [1993b]). Spherical staggered-grids have been used in some oceanic applications. In the global spherical staggered-grid construction some components are undefined (such as the ϕ-components as the poles, or the radial component at the lower boundary of the domain) and some are repetitious (such as the θ- and r-components defined at zero meridian). Thus, a number of exceptions and computational issues arise particularly in the *spherical* staggered-grid formulation.

Let us define vector spaces of *unique* components defined in the centres, on the edges and on the surfaces of *H-prisms*. It is convenient to define vectors whose components (ordered in a particular way) correspond to a node (i, j, k) and direction. The definition of these vectors depend on whether the components correspond to *H-edges* or *E-edges*.

Based on the notions described above, let us introduce the following vector spaces:

$\mathbb{E} \equiv$ vector space of complex components defined at the midpoints of each unique *H-edge* on the grid,

$\mathbb{F} \equiv$ vector space of complex components defined at the geometric centre points of each unique *H-face* on the grid (or, equivalently, at the midpoints of each unique *E-edge*, including those corresponding to the boundary surfaces),

$\mathbb{G} \equiv$ vector space of real components defined uniformly in the interior of each unique *H-prism*.

Then the quantities defined above can be described as $\mathbf{H} \in \mathbb{E}$, $\mathbf{E}, \mathbf{J} \in \mathbb{F}$ and $\rho \in \mathbb{G}$. To be specific, an H-component is defined at the centre point of the relevant *H-edge*, while
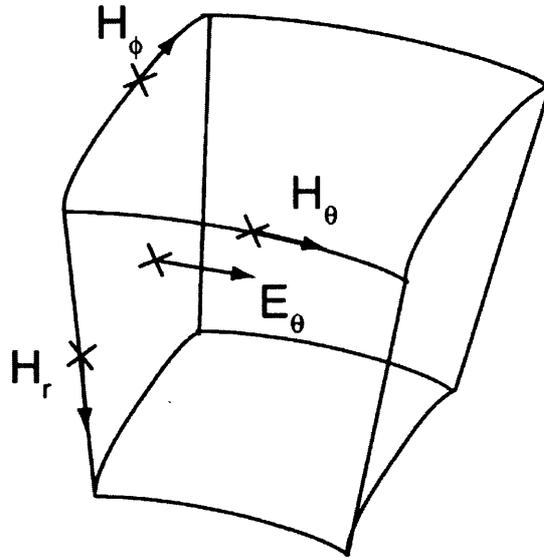
Figure 3.2: Spherical staggered-grid finite difference formulation: illustration of electric and magnetic field components in a single *H-cell*.

an E-component is defined at the centre point of the relevant *H-surface*. For a single grid cell, this construction is illustrated on Figure 3.2.

For example, the statement "**H** defined on edges" means that for each unique *H-edge* there exists the relevant directional component of **H**, either $H_\phi(i,j,k)$, or $H_\theta(i,j,k)$, or $H_r(i,j,k)$, respectively, defined on this edge.

These vector spaces will be used extensively in the analysis developed in Chapter 5.

3.5 Linear algebraic formulation

The Helmholtz equation in the context of the Dirichlet problem posed in U&S can be written as a set of second-order difference equations on the interior components of the magnetic field, while the boundary conditions are assumed to be specified, as described in Section 3.3.1. Let us therefore introduce new vector notation for the magnetic field to

stress this feature of the model:

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix}. \tag{3.24}$$

That is, we form an ordered vector of all the components of the magnetic field $\mathbf{H}$ corresponding to all the H-edges of the grid in such a manner, that all the components corresponding to the interior edges of the grid end up in vector $\mathbf{h}_i$ ($'i'$ for "interior"), while all the components corresponding to the upper and lower boundaries of the domain (i.e. $H_\phi(i,j,k)$ and $H_\theta(i,j,k)$ for $k = 1$ and $k = n + 1$, where n is the number of vertical division on the grid) constitute vector $\mathbf{h}_b$ ($'b'$ for "boundary").

Thus, if we introduce two complimentary subspaces of $\mathbb{E}$, such that $\mathbb{E}_b \subset \mathbb{E}$ and $\mathbb{E}_i \subset \mathbb{E}$ constitute the vector spaces of complex components defined on the *boundary* and *interior* H-edges, respectively, it is clear that $\mathbf{h}_b$ and $\mathbf{h}_i$ are elements of those subspaces, while $\mathbf{h} \in \mathbb{E}$.

3.5.1   Elementary operator definitions

The discrete representation of the vector Helmholtz equation on the magnetic field (Equation 3.15) constitutes a system of second-order difference equations carefully derived by U&S (Equations (8)-(19)). Starting from first principles it is also possible to decompose the discretized Helmholtz operator into a combination of elementary sparse matrices, acting consecutively on the field vector. This may be done by constructing step by step each of the operators required to solve numerically the system of integral equations 3.17, just like described in Section 3.4. This decomposition shall be immensely helpful in many ways. The benefits of this approach should become apparent later in this text. To summarise, the implementation of the iterative forward solution of Helmholtz equation is

simpler if the whole thing may be written in terms of elementary operators. Eventually, this approach would also allow to infer a direct relationship between the magnetic field and the parameter of interest - electrical resistivity, which would in turn be helpful for any kind of sensitivity analysis.

Notational convention

We are planning to introduce a set of operators that represent mappings between the spaces $\mathbb{G}$, $\mathbb{E}$ and $\mathbb{F}$, as defined in Section 3.4. For edges or faces, if just the interior ('i') or just the boundary ('b') components are intended to be included, this would be specified by the relevant subscript.

In general, the letter D for a matrix shall denote a square diagonal operator, with a subscript to indicate the nature of the values on the diagonal. Diagonal matrices are used extensively throughout this and the following chapter, and different diagonal operators are meant to act on elements of different spaces, such as $\mathbb{F}$, $\mathbb{E}$, $\mathbb{E}_i$ or $\mathbb{E}_b$. Therefore, to avoid clogging up the formulas with unnecessary notation, we have made the decision to allow the diagonal matrices encountered in formulas to assume the dimension of the element that it acts upon, unless the matrix dimension is specified explicitly. For example, operator D_l would always denote component-wise pre-multiplication of a vector defined on edges by the relevant edge length elements. However, depending on whether it acts on a vector of all edge components, or on the interior or boundary components only, it would have the dimensions, respectively, $|\mathbb{E}| \times |\mathbb{E}|$, $|\mathbb{E}_i| \times |\mathbb{E}_i|$ or $|\mathbb{E}_b| \times |\mathbb{E}_b|$. If encountered outside a well-defined context, the dimensions of such an operator will be clearly outlined.

Evaluating line integral $\oint \mathbf{H} \cdot d\mathbf{l}$

To compute the line integral of magnetic field $\mathbf{H}$ around the loop encircling an *H-surface*, we need to pre-multiply the relevant components of the field vector by the respective length elements, and then compute a signed sum. To perform these operations, we introduce the following elementary operators.

Operator D_{lE} : $\mathbb{E} \rightarrow \mathbb{E}$ is a diagonal operator pre-multiplying a vector defined on edges by the respective edge lengths (the lengths of *H-edges* may be computed using equations 3.18).

Operator C : $\mathbb{E} \rightarrow \mathbb{F}$ is a real sparse matrix whose non-zero elements are ± 1. It is constructed in such a way that if C is applied to a vector defined on *H-edges*, each of the elements of the resultant vector on *H-faces* (equivalently, on *E-edges*) is a signed sum of the four elements surrounding that face, in the sense of integration around contour (as shown in Figure 3.3). On the boundaries of the domain, C uses the elements on boundary edges to compute boundary radial components and the near-boundary horizontal field components on faces. In the most general sense, C may be viewed as the discrete implementation of the curl operator, except that no length scales are involved – so that C is, effectively, non-dimensional.

As we have seen, if $\mathbf{h}$ denotes the magnetic field vector, $CD_{lE}\mathbf{h}$ is a discrete implementation of $\oint \mathbf{H} \cdot d\mathbf{l}$. Denote this vector by letter $\mathbf{e}$. Thus, vector $\mathbf{e} \in \mathbb{F}$ is defined by $\mathbf{e} = CD_{lE}\mathbf{h}$.

$$+ \quad l_\theta H_\theta(i,j,k)$$

$$- \quad l_\theta H_\theta(i,j,k+1)$$
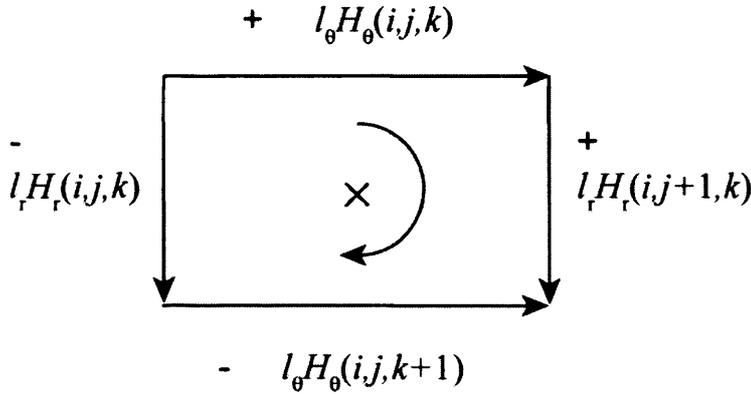
Figure 3.3: Operator C can be used to evaluate the discrete approximation of $\oint \mathbf{H} \cdot d\mathbf{l}$; value assigned to face is a signed sum as indicated, and points into the diagram in accordance with the sense of integration around contour.

Local estimate of electric current J

In the previous subsection, we have estimated the value of the line integral $\oint \mathbf{H} \cdot d\mathbf{l}$, which is a vector of complex values defined on *H-faces*. According to the first integral equation, $\iint \mathbf{J} \cdot d\mathbf{S}$ should equal to the same vector. This allows us to compute the local estimates of the electric current $\mathbf{J}$, flowing across each of the *H-faces* (equivalently, along the *E-edges*), by evaluating the expression $\iint \mathbf{J} \cdot d\mathbf{S} / \iint d\mathbf{S}$ for a single surface element. Hence, to compute the value of vector $\mathbf{J}$ we would need to divide $\mathbf{e} = CD_{lE}\mathbf{h}$ by surface area elements.

We introduce the diagonal operator $D_{SF} : \mathbb{F} \to \mathbb{F}$, that implements pre-multiplication of a vector on *H-faces* by surface area elements. Clearly, the inverse of this operator, $D_{SF}^{-1} : \mathbb{F} \to \mathbb{F}$ performs the opposite task and divides by respective surface area elements - which is the operator that we require. Thus, a local estimate of $\mathbf{J}$ is implemented by evaluating $D_{SF}^{-1}\mathbf{e} = D_{SF}^{-1}CD_{lE}\mathbf{h}$.

$$\ell^+ \quad \rho_{face} = \frac{\ell^+\rho^+}{\ell^++\ell^-} + \frac{\ell^-\rho^-}{\ell^++\ell^-} = \left(\frac{1}{\ell^++\ell^-}\right)\left(\frac{\ell^+}{\sigma^+} + \frac{\ell^-}{\sigma^-}\right)$$
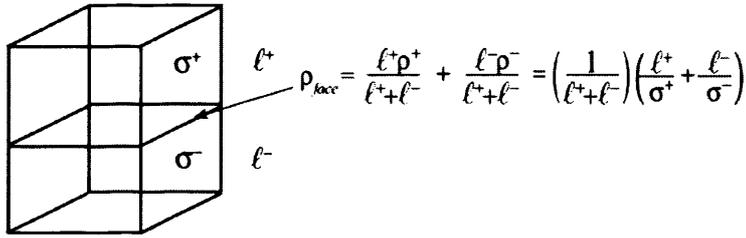
Figure 3.4: Operator L is an averaging operator used to compute a scalar quantity on a face of a cell, using the values in the centres of adjacent cells.

Mapping electrical resistivity from *H-prisms* to *E-prisms*

The values of electrical resistivity $\rho \in \mathbb{G}$ are discontinuous across *H-cell* boundaries. This makes the values of $\mathbf{E} = \rho \mathbf{J}$ also discontinuous across *H-surfaces*, which is an unwelcome feature for calculating the line integral of $\mathbf{E}$ around an *E-loop* (second integral equation). To make $\mathbf{E}$ continuous, we introduce new set of resistivity values, which we call $\rho_{face} \in \text{Re}\,(\mathbb{F})$, defined on *H-faces*. To do this, we construct an averaging operator from *H-cell* centres to *H-surfaces*.

Operator $L : \mathbb{G} \to \text{Re}\,(\mathbb{F})$ is a mapping from cell centres to faces, that averages quantities in adjacent cells to their common face, weighting by its perpendicular length elements. A common use of operator L would be to construct resistivity on faces. Consider a resistivity vector $\rho \in \mathbb{G}$. Once we choose an *H-face*, two entries of the vector would correspond to the resistivities of adjacent cells, call them ρ^+ and ρ^-. Then l^+ and l^- are, respectively, perpendicular length elements in these cells (see Figure 3.4).

Then, the entries of matrix L would be such that the resistivity on face

$$\rho_{face} = L\rho$$

would consist of averages of the entries ρ^+ and ρ^- of the original resistivity vector ρ, weighted by respective perpendicular length elements:

$$\rho_{face} = \frac{l^+}{l^+ + l^-}\rho^+ + \frac{l^-}{l^+ + l^-}\rho^-. \tag{3.25}$$

So that, for example, if we needed to compute the value of resistivity on the *H-face* corresponding to the area $S_\phi(i,j,k)$, this would be done by setting $\rho^+ = \rho(i,j,k)$, $\rho^- = \rho(i-1,j,k)$, $l^+ = l_\phi(i,j,k)$ and $l^- = l_\phi(i-1,j,k)$, and evaluating the above expression.

We also introduce a diagonal operator $D_{L\rho} : \mathbb{F} \to \mathbb{F}$, with the values of $\rho_{face} = L\rho$ aligned along the diagonal. Multiplication by this operator implements component-wise multiplication by ρ_{face} values.

Local estimate of electric field E

Having constructed the resistivity mapping in the previous subsection, it is now an easy task to implement $\mathbf{E} = \rho_{face}\mathbf{J}$. After computing $\mathbf{J} \in \mathbb{F}$ using $D_{SF}^{-1}CD_{lE}\mathbf{h}$, we now multiply this vector component-wise by the values of ρ_{face} to obtain an estimate of $\mathbf{E}$,

$$D_{L\rho}D_{SF}^{-1}CD_{lE}\mathbf{h}.$$

Evaluating line integral $\oint \mathbf{E} \cdot d\mathbf{l}$

To evaluate the line integral $\oint \mathbf{E} \cdot d\mathbf{l}$ along an *E-loop*, knowing $\mathbf{E}$, need to pre-multiply by the relevant length scales for *E-edges*, and calculate a signed sum of the elements defined on the four faces, perpendicular to the *E-edges* that form the loop (see Figure 3.5). Owing to the staggered-grid construction, electric field is mapped to the *H-edges* located at the
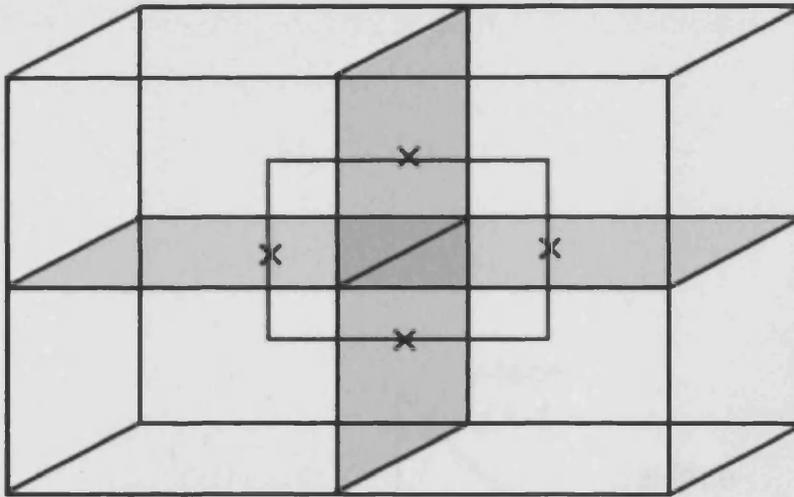
Figure 3.5: Operator C^T maps the values on the four adjacent faces onto the center edge of the "paddle wheel", in accordance with the sense of integration around contour. This operator can be used to evaluate the discrete approximation of $\oint \mathbf{E} \cdot d\mathbf{l}$.

centres of these "paddle wheels".

Thus, we define a diagonal operator $D_{lF} : \mathbb{F} \to \mathbb{F}$, that is intended to multiply by the appropriate length elements for *E-edges*. For each of the faces, using the same notation as that in Equation 3.25, these length elements may be computed as $(l^+ + l^-)/2$.

Operator $C^T : \mathbb{F} \to \mathbb{E}$ is a real sparse matrix whose non-zero elements are ± 1. It is constructed in such a way that if C^T is applied to a vector defined on *H-faces* (equivalently, on *E-edges*), each of the elements of the resultant vector on *H-edges* is a signed sum over four faces, that contain the *H-edge*, in the sense of integration around contour. On the boundaries of the domain, C^T uses the values on the near-boundary faces to compute the near-boundary radial components. The mapping to boundary horizontal field components is not well defined, however, it can still be constructed using the faces inside and on the boundary of the domain only. For the *spherical* staggered-grid, this operator is not as simple as for the cartesian case: complications arise due to the presence of the poles.
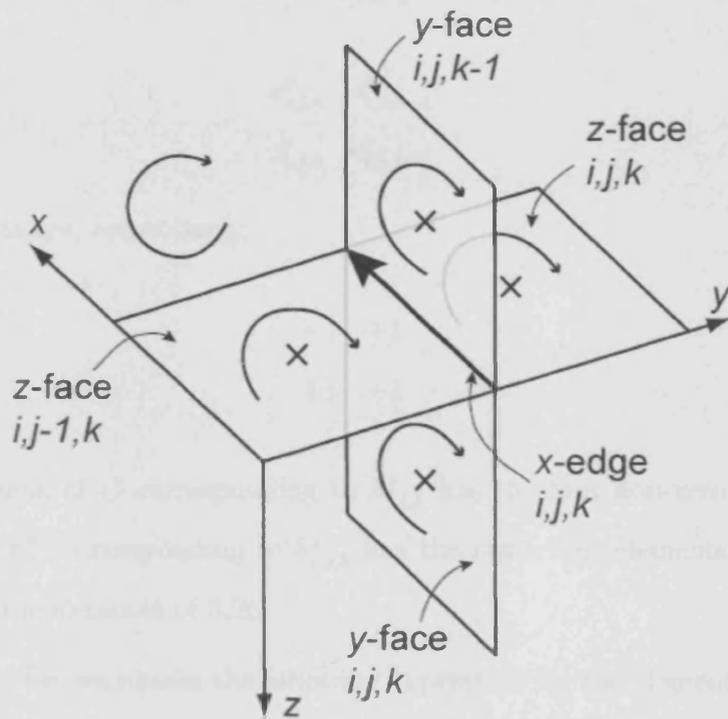
Figure 3.6: Illustration for the proof that the operator C^T is the transpose of C. The clockwise arrow corresponds to the sense of contour integral $\oint \mathbf{E} \cdot d\mathbf{l}$ around the (i, j, k)'th x-edge.

There are subtleties in the implementation of the operators around the poles and at zero longitude. They are not included in this text, but they are not necessary for conceptual understanding of this and the following chapters.

It can be shown, that C^T is indeed the transpose of operator C. In the mapping $C : \mathbb{E} \rightarrow \mathbb{F}$ we use the x-edge element $h^x_{i,j,k}$ (see Figure 3.6) in computing four face elements:

$$
\begin{matrix}
e^y_{i,j,k} & e^y_{i,j,k-1} \\
e^z_{i,j,k} & e^z_{i,j,k-1}
\end{matrix}
\tag{3.26}
$$

and the coefficients are, respectively,

$$
\begin{matrix}
-1 & +1 \\
+1 & -1
\end{matrix}
\tag{3.27}
$$

Thus, the *column* of C corresponding to $h^x_{i,j,k}$ has the four non-zero entries in 3.27. Then, the *row* of C^T corresponding to $h^x_{i,j,k}$ has the same four elements in the columns corresponding to the elements of 3.26.

Hence, if $\mathbf{h} = C^T \mathbf{e}$, we obtain the following expression for the element $h^x_{i,j,k}$:

$$
-e^y_{i,j,k} + e^y_{i,j,k-1} + e^z_{i,j,k} - e^z_{i,j,k-1},
\tag{3.28}
$$

which is numerical approximation to $\oint \mathbf{E} \cdot d\mathbf{l}$ around the (i,j,k)'th x-edge with the correct sense of integration around contour. By symmetry, we can extend this argument to all the other edges of the grid.

A combination of these operators, acting on the electric field vector, implements $\oint \mathbf{E} \cdot d\mathbf{l}$.

Computation of $\iint \mathbf{H} \cdot d\mathbf{S}$

Thus, we have devised the necessary machinery to compute the LHS of the second integral equation (Equation 3.17). To compute the RHS, which is $- \iint i\omega\mu\mathbf{H} \cdot d\mathbf{S}$, we need one more elementary operator.

Diagonal operator D_{S^E} : $\mathbb{E} \rightarrow \mathbb{E}$ implements pre-multiplication of the values on *H-edges* by the perpendicular surface elements, passing through the mid-points of *H-edges*. On the staggered-grid, these are also *E-faces*. Then the operator $i\omega\mu_0 D_{S^E}$: $\mathbb{E} \rightarrow \mathbb{E}$ is a diagonal operator that multiplies a vector defined on edges by $i\omega\mu_0 S^E$, where S^E is the *E-face* area normal to the respective edge.

The surfaces S_b^E, perpendicular to the horizontal boundary edges of the domain, are no longer fully inside the domain, and multiplication by these elements is not well-defined in the context of the staggered-grid formulation. We shall see that since we're only interested in the interior components of this surface integral, this shall not create a problem.

A careful decomposition of the operator into the interior and boundary parts would show the following structure:

$$
D_{S^E} = \left(
\begin{array}{cc}
D_{S_i^E} & 0 \\
0 & D_{S_b^E}
\end{array}
\right)
\begin{array}{l}
\left.\right\} |\mathbb{E}_i| \\
\left.\right\} |\mathbb{E}_b|
\end{array}
\tag{3.29}
$$

We can now rewrite the Helmholtz equation 3.15 in linear algebraic notation using a composition of elementary linear operators defined in this section.

3.5.2 Decomposition of Helmholtz operator

Following carefully the steps described in the previous section, we conclude that an equivalent linear algebraic expression to the general vector Helmholtz equation 3.15 with zero RHS on a staggered-grid is

$$(C^T D_{lF} D_{L\rho} D_{SF}^{-1} C D_{lE} + i\omega\mu_0 D_{SE})\mathbf{h} = 0. \tag{3.30}$$

Let us first simplify the notation by introducing a new diagonal operator, $D_{\rho F} : \mathbb{F} \to \mathbb{F}$, such that

$$D_{\rho F} = D_{lF} D_{L\rho} D_{SF}^{-1}. \tag{3.31}$$

Since operator $D_{\rho F}$ is a concatenation of three real diagonal operators, it is itself real (although it may of course act on complex components to produce something complex). A typical diagonal element of $D_{\rho F}$ is computed using the expression $\frac{1}{2SF}(l^+\rho^+ + l^-\rho^-)$, which we recognize as the multiplication factor used in Equation (18) of U&S.

An examination of Equations (18)-(19) in U&S and Equation 3.30 shows, that these expressions are indeed consistent - in fact, the two systems are identical. However, in the context of the Dirichlet problem being solved, this equation may not be considered valid on the boundaries of the domain. In fact, the interior components of the field satisfy Equation 3.30, while the boundary components are specified via an external source, and henceforth kept invariable. We would like to complete the derivation by making the system mathematically valid everywhere inside and on the boundaries of the domain. We shall later see, that the computation of the adjoint of Helmholtz operator will make this level of mathematical precision necessary.

3.5.3 Further decomposition of Helmholtz operator onto interior and boundary parts

To proceed, we rewrite the system in Equation 3.30 using Equation 3.31 and bearing in mind that the field vector $\mathbf{h}$ may be decomposed into two parts constituting the interior and the boundary components respectively, just like already described by Equation 3.24. We should also bear in mind, that only the interior part of Helmholtz operator matters. We obtain

$$C^T D_{\rho^F} C D_{lE} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} + i\omega\mu_0 D_{SE} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} = 0. \tag{3.32}$$

We aim to bring this system into a set of linear algebraic equations, acting on the *variable* interior components only, while the boundary conditions are used to force the system on the RHS. In order to do this, let us introduce, additionally, the following useful operators.

Let $I_{-b} : \mathbb{E} \to \mathbb{E}$ be a diagonal operator such that

$$I_{-b} = \begin{pmatrix} I_i & 0 \\ 0 & 0 \end{pmatrix} \begin{matrix} \}|\mathbb{E}_i| \\ \}|\mathbb{E}_b| \end{matrix} \tag{3.33}$$

where $I_i : \mathbb{E}_i \to \mathbb{E}_i$ is the identity matrix. That is, I_{-b} serves the purpose of replacing the boundary conditions on edges with zeros.

Similarly, introduce the diagonal mapping $I_{-i} : \mathbb{E} \to \mathbb{E}$ such that

$$I_{-i} = \begin{pmatrix} 0 & 0 \\ 0 & I_b \end{pmatrix} \begin{matrix} \}|\mathbb{E}_i| \\ \}|\mathbb{E}_b| \end{matrix} \tag{3.34}$$

so that multiplication by I_{-i} only leaves the boundary components non-zero.

Further, we notice that we can now mathematically restrict the above equation to the interior of the domain by pre-multiplying by I_{-b}:

$$I_{-b}C^T D_{\rho^F} C D_{lE} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} + i\omega\mu_0 I_{-b} D_{S^E} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} = 0. \qquad (3.35)$$

Now, since D_{S^E} is diagonal,

$$I_{-b} D_{S^E} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} \equiv D_{S_i^E} \mathbf{h}_i.$$

Let operator $C_i = CI_{-b}$ be the operator C restricted to act on the interior edges only, i.e. with columns corresponding to the boundary edges deleted. Thus, $C_i : \mathbb{E}_i \to \mathbb{F}$, and

$$C_i \mathbf{h}_i \equiv CI_{-b} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} \equiv C \begin{bmatrix} \mathbf{h}_i \\ 0 \end{bmatrix}.$$

Similarly, operator C_b is the operator C restricted to the boundary edges, with columns corresponding to the interior edges deleted. Operator $C_b : \mathbb{E}_b \to \mathbb{F}$ is such that

$$C_b \mathbf{h}_b \equiv C \begin{bmatrix} 0 \\ \mathbf{h}_b \end{bmatrix}.$$

Thus,

$$C \equiv \left(\begin{array}{c|c} C_i & C_b \end{array} \right). \qquad (3.36)$$

It is clear that $I_{-b}C^T = (CI_{-b})^T = C_i^T$. Equation 3.35 is equivalent to

$$C_i^T D_{\rho^F} CD_{lE} \begin{bmatrix} \mathbf{h}_i \\ \\ \mathbf{h}_b \end{bmatrix} + i\omega\mu_0 D_{S_i^E}\mathbf{h}_i = 0, \tag{3.37}$$

and can be further rewritten by taking the boundary components to the RHS as follows:

$$C_i^T D_{\rho^F} CD_{lE} \begin{bmatrix} \mathbf{h}_i \\ \\ 0 \end{bmatrix} + i\omega\mu_0 D_{S_i^E}\mathbf{h}_i = -C_i^T D_{\rho^F} CD_{lE} \begin{bmatrix} 0 \\ \\ \mathbf{h}_b \end{bmatrix}. \tag{3.38}$$

We can now write the linear equations on the interior components, keeping the boundary conditions on the RHS:

$$(C_i^T D_{\rho^F} C_i D_{l_i^E} + i\omega\mu_0 D_{S_i^E})\mathbf{h}_i = -C_i^T D_{\rho^F} C_b D_{l_b^E}\mathbf{h}_b, \tag{3.39}$$

and complete the problem setup by setting the boundary components $\mathbf{h}_b = \mathbf{b}$, where $\mathbf{b}$ is specified.

3.5.4   System of linear equations on H

U&S find it convenient to express the linear relationship between the variable interior components of the magnetic field and the fixed boundary components through an operator that we shall denote $A_{\rho,\omega}$. This operator depends on ρ, ω and μ only (however, since we take $\mu = \mu_0$ to be constant, we omit the explicit dependence on this parameter) and it has some very welcome properties. In fact, it is symmetric and real everywhere except along the diagonal. The combination of these features makes it possible to solve the linear system of equations efficiently (see Section 3.6). Symmetric operator $A_{\rho,\omega}$ acts on the interior components of the magnetic field, pre-multiplied by the respective edge lengths. The RHS of the equation constitutes a mapping from the boundary *H-edges* of the grid

to the interior *H-edges*, effectively producing a forcing. We shall denote this mapping B_ρ, since it only depends on resistivity.

To summarize, the set of second-order difference equations (Equations (18)-(19) of U&S) can be rewritten in terms of the operators $A_{\rho,\omega}$ and B_ρ and the elementary operator $D_l \equiv D_{l_i^E}$, as

$$A_{\rho,\omega} D_l \mathbf{h}_i = -B_\rho \mathbf{h}_b. \qquad (3.40)$$

By comparison with the analysis developed in Section 3.5.1 (see specifically Equation 3.39), we identify these operators to be

$$A_{\rho,\omega} = C_i^T D_{\rho^F} C_i + i\omega\mu_0 D_{S_i^E} D_{l_i^E}^{-1}, \qquad (3.41)$$

$$B_\rho = C_i^T D_{\rho^F} C_b D_{l_b^E}. \qquad (3.42)$$

As we see from this decomposition into elementary operators, operator $A_{\rho,\omega}$ indeed possesses the helpful properties claimed in the beginning of this section.

It has to be stressed, that $A_{\rho,\omega} : \mathbb{E}_i \to \mathbb{E}_i$, and Equation 3.40 (or equivalently Equation (20) in U&S) is solved only for the interior edges, keeping the magnetic field at the boundary edges fixed. This equation is equivalent to the system of equations on both the interior and the boundary components, as follows:

$$A_{\rho,\omega} D_l \mathbf{h}_i = -B_\rho \mathbf{b} \qquad (3.43)$$

$$\mathbf{h}_b = \mathbf{b}$$

where $\mathbf{b} \in \mathbb{E}_b$ denotes the fixed vector of boundary conditions.

We shall return to this system of linear equations when we proceed to discuss the generalized forward solver in Chapter 5.

3.6   Iterative numerical solution

Equation 3.40, as already discussed, constitutes a large sparse complex symmetric linear system. Given electrical resistivity ρ everywhere on the grid, and the boundary conditions $\mathbf{b}$, it can then be solved for the magnetic field $\mathbf{h}_i$ in the interior of the domain by the standard methods of numerical analysis. Since the coefficient matrix $A_{\rho,\omega}$ has complex diagonal elements (although all of the off-diagonal elements are real-valued) the matrix is non-Hermitian; while the convergence properties of most otherwise appropriate iterative relaxation methods (such as the Minimum Residual Accelerated (MRA) method used by *Mackie and Madden* [1993b] and most other variants of the conjugate gradient method) have only been established for real and symmetric (thus Hermitian), positive-definite matrices. An additional complication arises from the fact that $A_{\rho,\omega}$ is also indefinite and very ill-conditioned. The large condition number mainly stems from the large conductivity contrast between the insulating air layers and the conducting Earth.

In order to recover a Hermitian system of linear equations, *Toh et al.* [2002] have reposed the problem so that a variant of the biconjugate gradient method, valid for complex symmetric matrices [*Smith*, 1996], could be employed to solve this linear system.

In the numerical implementation written by H. Toh, A. Schultz and M. Uyeshima, the system of Equation 3.40 is solved for the magnetic field values by a variant of the biconjugate gradient method, Bi-Conjugate Gradient Stabilised (BiCGSTAB). Every few iterations, the values of the magnetic field are corrected to satisfy the zero divergence

condition. This method has made the code accurate and efficient enough to be employed as a core of 3-D inverse computations to reconstruct global mantle conductivity.

Chapter 4

Issues in geophysical inverse theory

The spherical staggered-grid forward formulation for the global EM Induction problem, described in Chapter 3, allows us to obtain an accurate estimate of the magnetic fields at the Earth's surface given a detailed electrical conductivity grid and an approximation of the sources. Our aim is to develop the inverse computational solution that would that would make it possible to recover as much information as we can about the distribution of electrical conductivity in the Earth's mantle, based on the measurements of EM fields, or their ratios, on the Earth's surface. In this Chapter we consider the difficulties inherent to this non-linear ill-posed inverse problem. We also describe tools necessary to make this task computationally feasible.

4.1 An introduction to inverse theory

The theory of inverse problems has a long and fruitful history and is carefully outlined from different viewpoints in such fundamental texts as *Tikhonov and Arsenin* [1977], *Parker* [1984], *Tarantola* [1987] and *Tarantola* [2005], *Kirsch* [1996], *Hansen* [1998], *Engl et al.* [2000] and others. In this section, we shall outline the concept of an inverse problem,

and provide an illustration using the linear forward and inverse formulations. We shall also discuss the issues arising from ill-posedness and non-linearity of the inverse problems commonly encountered real-world situations.

4.1.1 General setting

Suppose we are given a set of geophysical measurements, d_1, d_2, ..., d_M, that can be represented by the data vector $\mathbf{d} \in \mathbb{R}^M$, whose values depend on a *physical parameter with unknown spatial distribution*. We represent this unknown parameter distribution as an element $\mathbf{m} \in V$, where V is a suitable linear vector space. Under the assumption that the data are error free, the solution to the general *forward problem* can always be expressed as $d_j = F_j[\mathbf{m}], j = 1 \ldots M$, where F_j is a (usually non-linear) forward functional associated with the jth datum. In other words, the solution $F_j, j = 1 \ldots M$ to the forward problem supplies means of calculating the values of the observations if the function describing the parameter distribution is specified.

The corresponding *inverse problem* is to learn as much as possible about the model $\mathbf{m}$ from the data $\mathbf{d}$, once the solution to the forward problem is at hand.

4.1.2 Ill-posedness and regularisation

Hadamard's definition of *well-posedness* of an inverse problem states that a problem is well-posed if

1. For all admissible data, a solution exists (*existence*).

2. For all admissible data, the solution is unique (*uniqueness*).

3. The solution depends continuously on the data (*stability*).

To make this precise, we need to specify the notion of a solution, which data are considered admissible and which topology is used for measuring continuity. This has been

Model space

Data space

$$d = F(m)$$

$$\tilde{m} \bullet$$

$$\|d - \tilde{d}\| \leq \varepsilon$$

$$M \subset R^N$$

$$F^{-1}(\tilde{d})$$

$$D \subset R^M$$
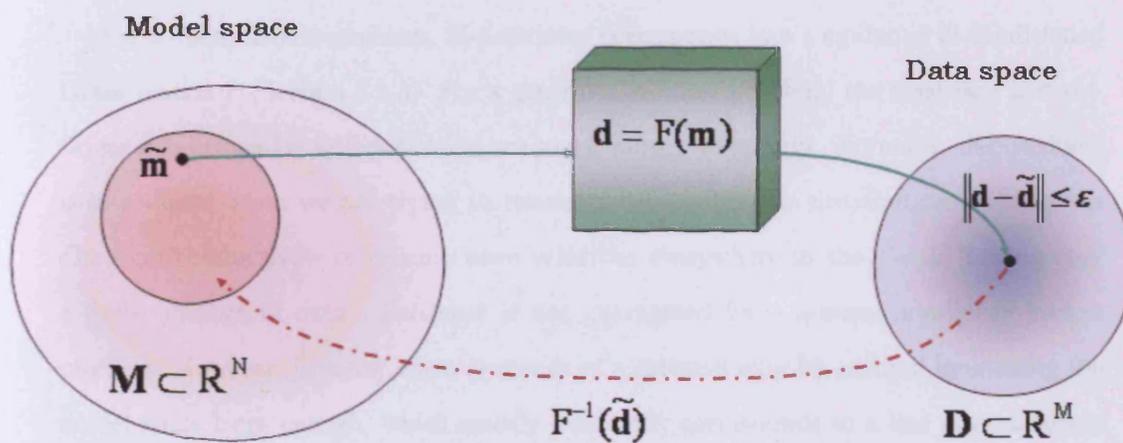
Figure 4.1: Schematic of an ill-posed inverse problem. The (green) cube and the solid arrows represent the forward problem, while the (brown) dashed arrow represents the inverse solution. For a well-posed inverse problem, a one-to-one correspondence exists between model and data. For an ill-posed problem, the inverse mapping $m = F^{-1}(\tilde{d})$ is not necessarily uniquely defined. In this case, there could be a subspace $\mathcal{M}$ of models that fit the data. In a practical inverse problem, we search for the subspace that fits a data point $d \in \mathcal{D}$, since we do not normally possess error-free data. This subspace may or may not include the model $\tilde{m}$.

done by many workers (e.g. *Kirsch* [1996], *Engl et al.* [2000]) and we are not going to discuss the formalism of a well-posed problem in detail here. Conceptually, *ill-posedness* is introduced when at least one of the above conditions is not satisfied. Most inverse problems encountered in real life situations are *ill-posed*. Let us discuss in greater detail the complication arising from ill-posedness in the inverse problem theory.

For a linear inverse problem, ill-posedness corresponds to a singular or ill-conditioned Gram matrix Γ (Section 4.1.3). For a general non-linear problem, the necessary and sufficient conditions for well-posedness are much harder to specify. Normally, ill-posedness is introduced when we are trying to reconstruct a continuous distribution (such as the electrical conductivity or seismic wave velocities everywhere in the Earth) having only a finite number of data. *Existence* is not guaranteed for a general non-linear inverse problem. However, in many cases existence of a solution may be ensured by making the model space large enough, which usually essentially corresponds to a fine discretization. However, the *uniqueness* condition is necessarily violated in this class of problems since they are by definition under-determined. Even if the *uniqueness* condition is fulfilled in the continuous problem when the data are present at uncountably many points, non-uniqueness is usually introduced by the need for discretization. Figure 4.1 illustrates the concept of non-uniqueness of a discrete ill-posed inverse problem.

In problems where violation of *stability* occurs, traditional numerical methods become unstable. For these problems, a small variation (or error) in the data may correspond to an arbitrarily large departure from the original solution. Problems of this kind are very difficult to solve computationally. Moreover, most geophysical inverse problems are unstable in that the solution that best fits the data (or fits the data exactly for a linear problem with no constraints) will change dramatically for even a tiny change in the data;

obviously in the presence of data errors this is undesirable.

Techniques that have been developed to tackle ill-posed inverse problems are known as *regularisation* techniques. Conceptually, these are techniques for construction of an approximation of an ill-posed problem by a family of neighbouring well-posed problems (*Engl et al.* [2000]). Thus, the notion of regularisation is central to restoration of stability and solution of an ill-posed inverse problem.

Regularisation techniques originate from *Tikhonov* [1943], who has proposed that linear ill-posed problems could be stabilised by adding a diagonal matrix to the ill-conditioned Gram matrix, thus getting away from the region of instability and ensuring that the eigenvalues are far enough from zero. Since then, many regularisation techniques have been developed for both linear and non-linear ill-posed inverse problems; classical references include *Marquardt* [1963]; *Tikhonov* [1963]; *Turchin et al.* [1971]; *Tikhonov et al.* [1987]; *Turchin* [1985]. Thus, regularization techniques have been in essence developed for discrete under-parametrized systems. The notion of regularisation also refers to the techniques that combat the non-uniqueness of an inverse problem by penalising the model for any features that are not necessary from the point of view of fitting the data, thus restricting the search domain to a class of "smoother" or in some ways more desirable models. Suppose that we have some prior knowledge about the model. For example, we're given a prior model $m_0 \in \mathbb{R}^n$ and we would like to search for model parameters that do not deviate too far from the prior model parameters. This prior information could be incorporated into the problem by adding a second term to the penalty functional that would penalise the departures from the prior model. We shall discuss these techniques further in Section 4.3.

Perturbation analysis and *resolution analysis* of non-linear inversions are two other important techniques developed to complement the construction of a solution for a non-linear ill-posed inverse problem. Perturbation analysis sets out to quantify the effect of a small perturbation in the solution on the data, and vice versa. This is an invaluable technique that allows to assess the stability of an inverse problem and the sensitivity of the data to small model parameter variations. Works centered on the perturbation analysis of non-linear inverse problems include *Snieder* [1990, 1998]. Resolution analysis, originally developed for a linear inverse problem by *Backus and Gilbert* [1967], and adapted since then to non-linear problems [e.g. *Snieder*, 1991] allows one to locate the features of the solution that may be trusted, and others that are not so well recovered (*resolved*). It also allows to estimate the *resolving power* of the data - that is, assists in quantifying the information that any particular data set contains. This is a very powerful technique, but it is not used very widely due to the related analytical and computational difficulties.

4.1.3   Linear inverse theory

Consider a Hilbert space[1] $\mathcal{X}$ endowed with an inner product weighted by the inverse model norm C_m^{-1}, so that $(\mathbf{m}_1, \mathbf{m}_2) = \mathbf{m}_1^T C_m^{-1} \mathbf{m}_2$; and take all the functionals F_j to be *linear and bounded.*

Then, by Rieszs theorem[2], we may rewrite the general form above as the inner product form

$$d_j = F_j[\mathbf{m}] = (g_j, \mathbf{m}), j = 1 \ldots M.$$

---

[1] A Hilbert space is a complete inner product space. An *inner product space* is a linear vector space (over some field, e.g. $\mathbb{R}$) with a function that takes an inner product of every two members of itself into the field it's defined upon (equivalently, a vector space endowed with an inner product). A metric space $\mathcal{X}$ is said to be *complete* if every Cauchy sequence of points in $\mathcal{X}$ has a limit that is also in $\mathcal{X}$. A *metric space* is a space endowed with the notion of distance between elements. Every inner product space is also a metric space with the metric induced by the inner product.

[2] Rieszs representation theorem for Hilbert spaces states that every bounded linear functional $L[f]$ on a Hilbert space may be written as an inner product, and the representer is uniquely determined by L (e.g. *Parker* [1984, p.32]).

Here $g_j \in \mathcal{X}$ are called *representers*, also known as *data kernels* in geophysics. Let us define the matrix G such that g_j is the j'th column of G. Then, G is the matrix of discrete representer elements and it has dimension $N \times M$, where M is the number of data and N is the model dimension.

Then, for the weighted scalar product space $\mathcal{X}$, the model data mapping could be written as $\mathbf{d} = G^T C_m^{-1} \mathbf{m}$. Let us define also the Gram matrix Γ [*Backus and Gilbert*, 1967] such that $\Gamma_{ij} = (g_i, g_j)$. For this special case, $\Gamma = G^T C_m^{-1} G$ (an $M \times M$ matrix).

Then, the model that minimises the norm defined by the inner product is a linear combination of representers, i.e. the columns of G [*Parker*, 1994]. Thus, the inverse solution that minimises the model norm among all models that fit the data exactly is $\tilde{\mathbf{m}} = G^T \mathbf{b}$; the coefficients $\mathbf{b}$ satisfy $\Gamma \mathbf{b} = \mathbf{d}$.

Hence, the error-free linear inverse problem is reduced to solving a (possibly large) system of simultaneous linear equations.

4.1.4 Non-linear inverse problems

If the forward problem F (Section 4.1.1) is non-linear and can not be linearised, then instead of solving a system of linear equations (a linear inverse problem) we are faced with a non-linear system that may very rarely be solved analytically. Typically, this is a system of differential or integral equations that needs to be discretized and solved numerically. Due to the discretization, many practical difficulties are encountered.

For a non-linear inverse problem, the exact fit error free procedure will generally be unstable (small changes in input to inversion, i.e. the data, will result in large changes in the inverse solution).[3] Thus, one introduces some sort of data misfit statistic and,

---

[3]For linear overparametrized problems (without constraints) one can always fit the data exactly, and any misfit statistic would be reduced to zero.

instead of fitting the data exactly, reduces this statistic to a reasonable value. Stability of the inverse problem is attained by minimising not the misfit statistic, but some penalty functional that includes a regularising term; or by minimising a model norm subject to fitting the data adequately, as measured by some data misfit statistic.

Additional difficulties related to non-linearity include:

- an evaluation of a single non-linear forward operator F is typically very computationally expensive. It is crucial that an efficient search algorithm is developed, in terms of the number of evaluations of the forward solver (and possibly also its gradient), to use when searching for the model that corresponds to the physical reality. Devising an algorithm to exploit parallel computing architecture may be beneficial;

- the need to linearise the forward operator F to compute gradients of the penalty functional required for a gradient-based search;

- non-uniqueness and resolution is harder to assess in non-linear problems. There is no satisfactory theory such as *Backus and Gilbert* [1967] for non-linear problems;

- local minima in a penalty functional may complicate the search; for practical linearised schemes the inverse solution may depend on where the search is started.

4.1.5 Practical inverse problems in geophysics

Most practical problems in geophysics are non-linear ill-posed inverse problem. A non-linear inverse solution requires an iterative search in the model space, in order to find models consistent with the data, and possibly with prior assumptions or constraints. This is done by solving the forward problem for a set of models, and comparing the result to any measurements we have of the real world, then updating the model iteratively in an intelligent way (Chapter 6). Once the domain of the forward problem is discretized,

a model constitutes a set of values in each of the cells of the computational mesh. The finer the discretization, the larger the model space. For fine enough discretizations to adequately model the real world, this model space has such a high dimension that searching in this space for the true model becomes unpractical. Thus, we try to compress the information about the model into a few parameters, and introduce a new mapping that maps these parameters to the computational mesh (thus parametrizing the model domain). Unfortunately, if the number of model parameters is too small, the parametrization is too restrictive to include a reasonably large set of models. Thus, even a parametrized problem still normally has a high-dimensional model space.

A practical inverse problem is non-trivial for the following reasons:

- the dimension of the model space may be too large for the algorithm to converge, let alone be stable. Thus, regularisation is usually necessary in a practical inversion;

- which function to minimise? It is important to define a good measure of distance between the data and the computed responses of the forward solver, to enable us to judge the goodness of model;

- having found a solution that satisfies the data, is it unique? If not, how close is it to the physical reality? Many non-linear inverse problems are ill-posed in the sense of (non-)uniqueness;

- noisiness of real data. In real life, measurements are rarely error free. Thus we are in fact inverting the data that depart slightly (or more than slightly) from the values that would have been observed if they were measured exactly. The modelling is also rarely exact. These errors may create a huge problem if the problem is ill-posed in the sense of stability.

4.2 Full setup of a geophysical inverse problem

Let us consider general non-linear forward and inverse problems in geophysics with a reference to the specific example of EM induction setting that we are concerned with in this text. Let us first define the mappings that constitute the forward solver and express them analytically.

Consider the relationship between the magnetic field $\mathbf{H}$ and the parameter of interest, electrical resistivity ρ. The Helmholtz operator (see e.g. Equation 3.15) is clearly linear in both ρ and $\mathbf{H}$. However, $\mathbf{H}$ and ρ themselves are related through the Helmholtz equation by a complex non-linear mapping, say, γ. Thus, if we go back to the finite difference approximation described earlier in Chapter 3 and think of the values in question as of vectors defined on the staggered-grid edges and cells centres (see Section 3.4 for a thorough explanation of this statement), we can denote the magnetic field by $\mathbf{h} \in \mathbb{E}$ and resistivity by $\rho \in \mathbb{G}$, and write that

$$\mathbf{h} = \gamma(\rho). \tag{4.1}$$

Then, evaluating the value of $\mathbf{h}$ on the grid given the electrical resistivity ρ constitutes a non-linear forward problem. The corresponding non-linear inverse problem would be to reconstruct the resistivity ρ, given the values of the magnetic field $\mathbf{h}$.

However, in practice the setup of a forward (and the corresponding inverse) problem is usually more involved than just the non-linear mapping $\gamma(\rho)$. If we were to reconstruct the resistivity ρ by solving the inverse problem, we would like to be able to assess whether our guess of ρ is good by comparing the resultant magnetic field $\mathbf{h} = \gamma(\rho)$ to something that can be measured. In practise we don't measure the magnetic field within the Earth's interior. Instead, magnetic field is measured on the surface of the Earth, or somewhere

above the Earth's surface. If the points where we can measure the magnetic field are still inside the domain, such as in the computational setup described in Section 3.3, then the values of the magnetic field that we know something about constitute at best a small number of entries of vector **h**. Furthermore, each magnetic field value has three complex components (radial, longitudinal and latitudinal), and each of these components can be measured with a different degree of accuracy. It is also quite possible that the locations at which the data can be taken do not directly correspond to the mid-points of the edges of the computational grid, so that a further error is introduced when trying to compare the computed magnetic field values to those that can be measured. Thus is has become customary to employ a data functional for such comparison purposes, so that the final data can be expressed as

$$\psi_j = \lambda_j(\mathbf{h}), \tag{4.2}$$

where the index j corresponds to the location of the data point, and ψ_j is the value computed from the measurements taken in some neighbourhood of this location. Such data functionals are called *transfer functions* (we have discussed EM transfer functions, also known as response functions, at length in Section 1.3.4). In our problem, a transfer function λ_j may use any of the values of magnetic field vector. For some problems it may also depend on components other than the magnetic field. For example, in magnetotellurics (MT) transfer functions, known as impedance tensors, would also require components of electric field solutions. We are not going to discuss such transfer functions in this document - although the generalization of the analysis developed here to MT problems could be readily derived. For our purposes, transfer functions would only be slightly non-linear and shall only depend on the magnetic fields in the immediate neighbourhood of the location of the observation. Thus, for an inverse problem we would not regard the values of the magnetic field as data, but instead use some (possibly non-linear) functionals of the magnetic field vector.

To solve the practical inverse problem, one more functional needs to be introduced. We notice, that for fine enough computational grid vector $\rho \in \mathbb{G}$ is high-dimensional. Thus, while performing our search for a model that satisfies the data well enough we find ourselves searching in the domain of the same dimension as the number of cells in the grid, $|\mathbb{G}|$. For a 3-D problem, this can be a very large number. In fact, for a fine 3-D grid $|\mathbb{G}|$ is so large, that it quickly becomes computationally unfeasible to search in this domain. A practise common to the researchers who wish to solve an inverse problem in a high-dimensional domain is to introduce a (possibly non-linear) mapping to the domain from a much smaller set of parameters. This mapping is called a *parametrization*. A parametrization allows to perform the search using just a few parameters. These parameters are then mapped to the grid and the forward solution is computed. Normally this mapping would use an orthogonal or orthonormal set of vectors to ensure that the search parameters are independent. Clearly, if the search domain is too small, the mapping would be too restrictive and therefore unlikely to allow a realistic model. Choosing a reasonable parametrization is therefore crucial to a successful inversion.

To construct a parametrization mapping, we would first need to define the domain. Suppose $\mathbf{m} \in \mathbb{R}^n$ is a real vector of model parameters, where n is defined as the dimension of this space. In general, n would be much smaller than $|\mathbb{G}|$, the total dimension of the grid. For our problem, $\mathbf{m}$ would be mapped to a vector $\rho \in \mathbb{G}$, which would then be employed to compute the magnetic field, and further the transfer functions. We could describe the most general parametrization as

$$\rho = \eta(\mathbf{m}). \tag{4.3}$$

The mappings appropriate to parametrize the 3-D spherical Earth require a separate discussion and shall be considered in detail in Section 4.4.

Thus, the full setup of the forward problem would require an expression that relates the response functions to the model parametrization coefficients:

$$\psi_j = \psi_j(\mathbf{m}) = \lambda_j(\gamma_{\omega,s}(\eta(\mathbf{m}))). \tag{4.4}$$

Thus, we have put together an analytic expression for the model response of a general non-linear forward problem.

To solve the corresponding inverse problem, such model responses need to be repeatedly computed and compared to the field measurements in an intelligent way. We assume, that we have some measured *data*, taken at different frequencies and at different locations in the computational domain. In the error-free case, those are the values of the data functionals $\psi_j(\tilde{\mathbf{m}})$ for the model $\tilde{\mathbf{m}}$ that corresponds to the physical reality - "*the truth*". The model is unknown. The data are in general complex, and usually contain *errors*. Errors may be real, or they could be regarded complex, if the real and imaginary components of the data are estimated to have been obtained with different accuracy.

In a geophysical inverse problem, a *data misfit* is a measure of the distance between the observed data, and the computed response of the model, where the data are regarded as a compilation of ideal measurements and noise[4] (the statistical distribution of the noise

---

[4]The assumption is that each data entry is drawn from a distribution, say, Gaussian, that has the unknown "ideally observed" true response value as the mean, and the corresponding recorded error as a random deviation from this mean. That means, that if the same measurement could have been taken repeatedly, with all the background conditions except for the amount of noise, being constant, such data, when recorded, would follow the specified distributional curve. This p.d.f. assumption on the data errors is critical in establishing the goodness of fit, both in the forward and inverse modelling. In particular, it serves for the calculation of misfit between the observed data and the response calculated

is assumed to be known). We assume that the ideal measurements are precisely those yielded by *the truth*. However, the relationship between the measurements (even ideally accurate), and the model is not necessarily one-to-one, and is never such for a non-unique inverse problem. It is important to define a misfit function that realistically reflects the proximity between the model response and the data, and is statistically robust, in that the result wouldn't change much if any of the underlying assumptions on the data (e.g. error distribution) were slightly wrong. It is also useful to have some understanding of the statistical properties of the misfit itself, for instance, the value that would be expected if the model was to fit the imprecise data ideally, and how great a deviation from that value could be allowed due to the data imprecision only.

A *penalty functional* normally comprises the data misfit and a term penalising the model norm, which is necessary to regularise the inverse problem (see discussion in Section 4.1.2). If the forward modelling results in a small enough misfit with the data, we say that the model we have used in the procedure belongs to the set of acceptable models, or *fits the data*. We cannot, however, conclude, that we have found the true conductivity distribution (while the reverse inference can be made: if we find that the misfit is unacceptably large, the corresponding model can be safely rejected). Then, the problem may be posed as follows:

1. Find at least one model that fits the data well;

2. Characterize in some way the set of models that fit the data well;

---

by the forward modelling, which is important in order to distinguish between the realistic and non-realistic models. However, this assumption normally deviates significantly from the observed reality. Some statistical outliers are to be expected, due to such side effects as geomagnetic storms, auroral and equatorial electrojet effect, and at some frequencies also cultural urban noise, as well as that from DC electric rail systems. It is not always possible to take them into account in the process of assignment of the *a priori* error estimates, which may invalidate the assumptions, underlying any further inverse calculations. Thus, careful statistical analysis of the data is required, and even then the use of robust measures of misfit is preferable.

3. Try to distinguish between the models using other sources of information.

While the ultimate task of the first stage of this procedure is to reach a target (small enough) data misfit, in practice this is usually not possible unless regularisation is performed. Thus, the inverse problem is solved by minimising the penalty functional comprising both the data and the model norms. Any L_p-norm can be employed for this purpose, and while there are extensions to L_2 for both data and model norms, in this work we have only used a variant of the L_2-norm squared, known as the *least-squares penalty functional*. Let us discuss this specific problem in greater detail.

4.3 Regularised least-squares inverse problem

Least-squares inversion involves minimising a linear combination of the data and the model L_2-norms squared (quadratic norms or quadratic semi-norms). Least-squares geophysical inversion techniques are widely used, and have been described in the classical references *Marquardt* [1963], *Lawson and Hanson* [1974], *Tarantola and Valette* [1982], *Lines and Treitel* [1984], *Tarantola* [1987] and others.

4.3.1 Least squares penalty functional

We could write the general regularised least squares penalty functional as

$$R(\mathbf{m}) = (\psi(\mathbf{m}) - \mathbf{d})^{\mathrm{H}} \, C_{\mathbf{d}}^{-1} \, (\psi(\mathbf{m}) - \mathbf{d}) + \mu(\mathbf{m} - \mathbf{m}_0)^{\mathrm{T}} C_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}_0), \qquad (4.5)$$

Here $C_{\mathbf{d}}$ is a data covariance operator that performs weighting of some sort by data errors and correlations, and the letter H denotes the Hermitian operator, meaning a combination of transposition and conjugation. We shall also denote the Hermitian operator by a star when convenient. Also, $C_{\mathbf{m}}$ is the model covariance, so that $C_{\mathbf{m}}^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ is a weighting

operator designed to penalise large deviations from the starting model parameters, and other features that are not desired, such as unwelcome large jumps between the parameters in neighbouring layers and higher order terms that correspond to finer details in the model. The model covariance operator allows us to ensure that any non-trivial features present in the model found by minimizing this functional should really be there in order to satisfy the data. Thus, larger entries in C_m^{-1} correspond to those of the parameters (or parameter variations) that we would not like to be there unless they are absolutely required by the data and that we therefore choose to penalise.

There are two ways, almost equivalent, to implement regularisation for a least-squares problem in practice, and both essentially involve the penalty functional in Equation 4.5. The first is to minimise this penalty functional; then the extra term involving the model norm is a regularising term in the traditional sense of *Tikhonov* [1963]. The second is Parker's 1984 approach: minimise the model norm subject to finding a model which fits the data to within tolerance. Minimisers of the penalty functional (4.5) in fact minimise the model norm subject to the data misfit achieved. Minimising with a range of μ results in a range of data misfits; choosing the one which achieves the appropriate misfit solves the problem for Parker's approach. In either case, solution of the inverse problem involves to some extent minimising the least-squares penalty functional, which is done by standard methods of global optimisation discussed in Chapter 6. In Chapter 5 we describe how the derivative of the least squares penalty functional may be computed, which is a necessary step to enable usage of any of the gradient-based optimisation techniques.

Another angle of looking at the regularisation needs to be mentioned: choosing model parametrizations to reduce the number of degrees of freedom also acts to regularise.

4.3.2   Regularisation by pre-conditioning

The problem with directly inverting the formulation of Section 4.3.1 is that the model norm part of the penalty functional (4.5) has an unbounded eigenvalue spectrum, making the overall penalty functional very poorly conditioned, and potentially difficult to minimise. The inverse of the model covariance operator could be considered a roughening operator, in the same sense as that discussed in Section 1.3.5: a roughening operator has the same sort of unbounded spectrum as the inverse covariance.

On the other hand, the model covariance operator $C_{\mathbf{m}} : \mathbb{R}^n \to \mathbb{R}^n$ could be considered a smoothing operator. This is an operator with relatively small eigenvalues that we could safely apply without being concerned that the solution is going to blow up. Thus, we would like to rewrite the problem in terms of $C_{\mathbf{m}}$, if possible. We achieve this by transforming the model parameter space to make the quadratic form defining the linearised penalty functional better conditioned; thus preconditioning the penalty functional.

Let $C_{\mathbf{m}}$ be a symmetric positive-semidefinite $n \times n$ matrix, and suppose we can find an invertible linear operator $C_{\mathbf{m}}^{1/2}$ such that $C_{\mathbf{m}} = C_{\mathbf{m}}^{1/2} C_{\mathbf{m}}^{1/2}$. Let the inverse of $C_{\mathbf{m}}^{1/2}$ be $C_{\mathbf{m}}^{-1/2}$. If, in addition, $C_{\mathbf{m}}^{-1/2}$ is symmetric, we can rewrite Equation 4.5 as

$$R(\mathbf{m}) = (\psi(\mathbf{m}) - \mathbf{d})^{\mathrm{H}}\, C_{\mathbf{d}}^{-1}\, (\psi(\mathbf{m}) - \mathbf{d}) + \mu [C_{\mathbf{m}}^{-1/2}(\mathbf{m} - \mathbf{m}_0)]^{\mathrm{T}} C_{\mathbf{m}}^{-1/2}(\mathbf{m} - \mathbf{m}_0).$$

Let us define a new model vector $\hat{\mathbf{m}} \in \mathbb{R}^N$ such that

$$\hat{\mathbf{m}} = C_{\mathbf{m}}^{-1/2}(\mathbf{m} - \mathbf{m}_0). \tag{4.6}$$

Then,

$$\mathbf{m} = C_{\mathbf{m}}^{1/2}\hat{\mathbf{m}} + \mathbf{m}_0, \tag{4.7}$$

and the expression (4.5) for the least-squares penalty functional may be rewritten in terms

of $\hat{\mathbf{m}}$:

$$R(\hat{\mathbf{m}}) = (\psi(C_{\mathbf{m}}^{1/2}\hat{\mathbf{m}} + \mathbf{m}_0) - \mathbf{d})^{\mathrm{H}} \, C_{\mathbf{d}}^{-1} \, (\psi(C_{\mathbf{m}}^{1/2}\hat{\mathbf{m}} + \mathbf{m}_0) - \mathbf{d}) + \mu \, \hat{\mathbf{m}}^{\mathrm{T}}\hat{\mathbf{m}}, \qquad (4.8)$$

An advantage of this scheme is that we would only directly use $\hat{\mathbf{m}}$, and if $\mathbf{m}$ is ever needed it could be evaluated using Equation 4.7. Thus, assuming the operator $C_{\mathbf{m}}^{1/2}$ is available, there is no need to invert any operators at all. Additionally, $C_{\mathbf{m}}^{1/2}$ acts as a "smoothing" operator, thus allowing us to penalise unwelcome features of the model without getting unboundedly large values of the misfit. When we define a covariance that acts as a smoother, the pre-conditioning used here is simple to implement; implementing the inverse of the smoother (which appears in the penalty functional, but is never needed for the preconditioned algorithm) would be hard. Variants of this technique have been encountered in literature [e.g. *Rodi and Mackie*, 2001; *Haber et al.*, 2004], except that in these cases the roughening operator has been easy to write down, while the smoothing term would have been difficult to write down and implement.

In terms of the transformed model parameters the model norm term is just the identity (multiplied by a scalar), and hence adding this term regularises (as in *Tikhonov* [1963]) instead of making the penalty functional more poorly conditioned.

4.3.3 The data misfit statistic

Let us consider in greater detail the data norm term of the least-squares penalty functional,

$$R_{\mathbf{d}}(\mathbf{m}) = (\psi(\mathbf{m}) - \mathbf{d})^{\mathrm{H}} \, C_{\mathbf{d}}^{-1} \, (\psi(\mathbf{m}) - \mathbf{d}). \qquad (4.9)$$

This formulation of the data misfit has helpful statistical properties. Operator $C_{\mathbf{d}}$ is a Hermitian covariance operator of a complex random vector. In this analysis we shall

only consider the case of diagonal C_d.

Suppose that we possess M data points, d_j, $j = 1, \ldots M$, where complex data values have been assumed. Suppose also, that each of these values has been measured with an ϵ_j (complex) error, and we denote the real and imaginary components of ϵ_j by $\epsilon_j^{(1)}$ and $\epsilon_j^{(2)}$, respectively. The data inaccuracies are such that

$$\text{Re } (d_j) = \text{Re } (\tilde{d}_j) + \epsilon_j^{(1)}, \tag{4.10}$$

$$\text{Im } (d_j) = \text{Im } (\tilde{d}_j) + \epsilon_j^{(2)}, \tag{4.11}$$

where $\tilde{d}_j$ is the "true" value that would have been observed if no measurement error were added. In practice, the measurement errors ϵ_j, $j = 1, \ldots M$ are never exactly known. Instead, they are *estimated* by the values δ_j. A common assumption in science is that measurement errors in the distinct data points may be viewed as independent random variables distributed as

$$\epsilon_j^{(1,2)} \sim N(0, \delta_j^2) \tag{4.12}$$

where $\delta_j \in \mathbb{R}$ have been somehow determined. In general, if we believe that the j's observatory is accurate and the background noise contaminating the measurement is small in amplitude, δ_j can be set small. If, however, we expect that a large amount of noise and inaccuracies contaminates the observation, we set δ_j to a larger value. Careful preliminary analysis of the data is usually performed to ensure that this assumption holds true to a reasonable extent. Please see Sections 2.4 and 7.1.1 for a discussion of errors δ_j in field and synthetic data, respectively.

Under this assumption, data values may be viewed as instances of random variables

distributed according to $N(\tilde{d}_j, \delta_j^2)$. Then,

$$\frac{\text{Re } (d_j - \tilde{d}_j)}{\delta_j} \sim N(0,1),$$

$$\frac{\text{Im } (d_j - \tilde{d}_j)}{\delta_j} \sim N(0,1).$$

The functionals $\psi_j(\tilde{\mathbf{m}}) = \tilde{d}_j$ computed for the true model are assumed to be error-free. Thus for the values of $\mathbf{m}$ close to the *truth* $\tilde{\mathbf{m}}$, the entities $(\psi_j(\mathbf{m}) - d_j)/\delta_j$ may also be considered normally distributed random variables with zero mean and unit variance, $N(0,1)$.

If we now revisit Equation 4.9 for this specific problem, we obtain a new expression for the least squares penalty functional,

$$R_\mathbf{d}(\mathbf{m}) = \sum_{j=1}^{M} \left\{ \left(\frac{\text{Re } (\psi_j(\mathbf{m}) - d_j)}{\delta_j} \right)^2 + \left(\frac{\text{Im } (\psi_j(\mathbf{m}) - d_j)}{\delta_j} \right)^2 \right\}. \tag{4.13}$$

In this example the data covariance operator $C_\mathbf{d}$ is in fact a diagonal operator with the values of δ_j^2 on the diagonal.

We see, that for the values of the model parameters $\mathbf{m} \approx \tilde{\mathbf{m}}$ expression (4.13) is an instance of a $2\chi_M^2$ random variable[5], which has the mean of $2M$. If we now divide the expression by the number of degrees of freedom $(2M)$ then the expectation $\mathbf{E}\left(R(\mathbf{m})/2M\right) \approx 1$ for values of $\mathbf{m} \approx \tilde{\mathbf{m}}$.

Thus, the least squares penalty functional has a pleasant feature that if we approach

---

[5]If the real and imaginary parts of the data values each had a distinct and independent error estimate, they would also be pairwise independent, and the expression (4.13) would correspond to a χ_{2M}^2 random variable. This would not affect the mean, which would still be $2M$.

the true model in our search, the expression, once it is scaled by the appropriate number of degrees of freedom, will approach the value of 1. For this data misfit, we can say that a model m *fits the data* if the condition $R_d(m)/2M \approx 1$ is achieved. This happens when the misfit is close to its expected value in the proximity of the true model.

The p.d.f. assumption on the data errors allows us to postulate, that once the model, used in the forward solution, gets sufficiently close to the *truth*, any deviates of the misfit are due to the noisiness of the observations. Thus, we cannot distinguish between the models that give a good enough fit. Trying to do so would require a higher degree of resolving power than that provided by the data. This is just another view at the ill-posedness of the geophysical inverse problem: there are never enough constraints on the model due to the finiteness and the noisiness of the data set. Additionally, errors are introduced through a discretization of the model domain, so that the computed responses are not precise either. While we have ignored modelling errors here, in general they require a separate analysis.

4.3.4 Extension to multiple frequencies

In practice, we are often faced with a least-squares problem that needs to be solved for several different frequencies, or other control parameters, in which case we would want to sum up the misfit terms for different parts of the problem. In this case, let us denote our expression in Equation 4.13 by R_ω to specify dependence on frequency. If we assume that the total number of data values for all parts of the problem summed up together is M, and for one frequency it is M_ω (this number may be different for different frequencies), we obtain

$$R_d(m) = \frac{1}{2M} \sum_\omega R_\omega(m), \quad M = \sum_\omega M_\omega. \tag{4.14}$$

Obtaining $R_\omega(\mathbf{m})$ as separate quantities for different values of ω also opens other possibilities for the total penalty functional, such as computing a weighted sum of R_ω. This would allow to downweight the influence of some frequencies with respect to the others. In general, this approach allows flexibility in the choice of the penalty functional, any linear combinations of R_ω certainly being acceptable.

4.4 Parametrization by surface spherical harmonics

Recall that the term *parametrization* in the context of geophysical inverse problems refers to the method of replacing the full distribution of the parameter of interest on the computational grid by a smaller number of parameters and a mapping that takes these parameters to the full distribution on the grid. Parametrization may also be viewed as a method of compressing the search domain to a smaller space, to make the search in that space computationally tractable.

Whatever the specific details of the parametrization, the idea is that we introduce a small set of parameters $\mathbf{m} \in \mathbb{R}^n$ to fully describe the staggered-grid resistivity distribution $\rho \in \mathbb{G}$, and map from the smaller set to the larger set using a (possibly non-linear) mapping $\eta(\mathbf{m})$. Clearly, not all distributions on the staggered-grid could be described using this mapping, due to the reduction in the number of degrees of freedom. However, in practice it is usually possible to choose a parametrization that would describe most physically realizable models (within the resolving power of our technique).

Various parametrizations may be used to describe the 3-D electrical resistivity in the Earth's mantle, and most parametrizations use a truncated (to some degree and order)

expansion in terms of some orthogonal (or orthonormal) set of basis functions. We have chosen the orthogonal set of surface spherical harmonics for this purpose, and we are parametrizing in the vertical with layers. This choice of basis functions has certain advantages that include linearity and ease of implementation. We are considering employing different base functions eventually, but since this work uses parametrization by spherical harmonics only we shall not discuss other possibilities in detail in this chapter.

4.4.1 Associated Legendre functions and spherical harmonics

Associated Legendre functions $P_l^m(x)$ are solutions of the differential equation on $[-1, 1]$

$$(1 - x^2)y'' - 2xy' + \left[l(l+1) - \frac{m^2}{1 - x^2}\right] y = 0. \tag{4.15}$$

These solutions are non-singular if and only if $l \in \mathbb{Z}_+^1$ and $0 \le |m| \le l$ [e.g. *Arfken and Weber*, 2001]. In this case the associate Legendre functions are polynomials, and for $0 < m \le l$ may be defined by

$$P_l^m(x) = (-1)^m (1 - x^2)^{m/2} \frac{d^m}{dx^m} P_l(x), \tag{4.16}$$

where

$$P_l(x) = \frac{1}{2^l l!} \left[\frac{d^l}{dx^l}(x^2 - 1)^l\right] \tag{4.17}$$

is the Legendre polynomial of degree l. The first two Legendre polynomials are $P_0(x) = 1$ and $P_1(x) = x$.

Spherical harmonic expansion in geophysics is normally based on the Schmidt semi-normalised associated Legendre functions $S_l^m(x)$. These are related to the non-normalised

associated Legendre functions $P_l^m(x)$ by

$$S_l^m(x) = P_l^0(x) \text{ for } m = 0;$$

$$S_l^m(x) = (-1)^{|m|}\sqrt{\frac{2(l - |m|)!}{(l + |m|)!}} P_l^{|m|}(x) \text{ for } 0 < |m| \leq l. \qquad (4.18)$$

Many alternative definitions of surface spherical harmonics exist, which essentially vary in the weighting factor by the $e^{im\phi} \cdot P_l^m(\cos\theta)$ term. Here, we define surface spherical harmonics $Y_l^m(\theta, \phi)$ in their complex form as

$$Y_l^m(\theta, \phi) = e^{im\phi} \cdot S_l^m(\cos\theta), \qquad (4.19)$$

where $S_l^m(\cdot)$ are the Schmidt semi-normalised associated Legendre functions (Equation 4.18).

Spherical harmonics form a complete set of orthogonal functions (that may also be normalised by an appropriate weighting if required) and thus form a vector space analogue to unit basis vectors. Any (square-integrable) complex function on a sphere $f(\theta, \phi)$ can be expanded as a linear combination of spherical harmonics,

$$f(\theta, \phi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm} Y_l^m(\phi, \theta). \qquad (4.20)$$

in the limit of $L \to \infty$. Here, Y_l^m are as defined by (4.19) and c_{lm} are the complex coefficients.

The expansion coefficients can be obtained by multiplying the above equation by the complex conjugate of spherical harmonics and integrating:

$$c_{lm} = w_{lm} \int_0^\pi \int_0^{2\pi} f(\theta, \phi)[Y_l^m(\theta, \phi)]^* \sin\theta \, d\phi \, d\theta, \qquad (4.21)$$

SPHERICAL HARMONICS
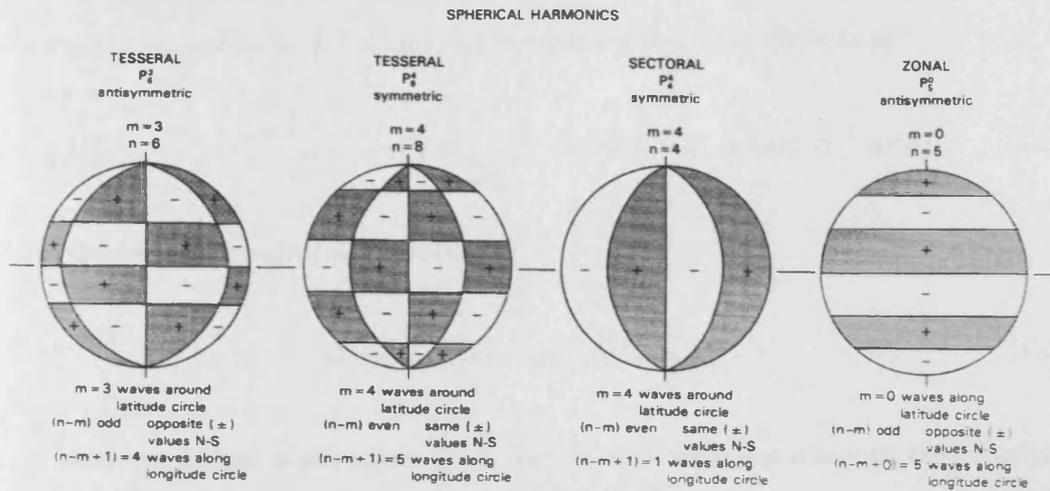
Figure 4.2: Tesseral, sectoral, and zonal forms of Legendre polynomial P_n^m (note change of notation $l \mapsto n$) used in construction of a spherical harmonic expansion, with value indicated by positive (shaded) and negative (clear) areas. The harmonics are symmetric or antisymmetric with respect to the equatorial plane. The four spheres show selections of n and m to determine the number of waves around latitude and longitude circles. Figure adapted from *Campbell* [2003].

where w_{lm} is the appropriate weighting factor. From the definition of spherical harmonics given by Equation 4.19,

$$
\begin{aligned}
w_{lm} &= 1 \text{ for } m = 0; \\
w_{lm} &= \frac{2l + 1}{8\pi} \text{ for } m \neq 0.
\end{aligned}
\tag{4.22}
$$

Various specific types of spherical harmonics are distinguished for convenience of scientific analysis. The *zonal harmonics* are defined to be those of the form $Y_l^0(\cos\theta) = P_l(\cos\theta)$. The *tesseral harmonics* are those of the form Y_l^m for $l \neq m$. The *sectorial harmonics* are of the form Y_m^m. See Figure 4.2 for an illustration.

We can ensure that the expression (4.20) is real by demanding that $c_l^{-m} = (c_l^m)^*$.

Then, the expansion for $f(\theta, \phi)$ may be rewritten with real coefficients as

$$f(\theta, \phi) = a_0^0 + \sum_{l=1}^{L} \left[a_l^0 P_l(\cos\theta) + \sum_{m=1}^{l} [a_l^m \cos(m\phi) + b_l^m \sin(m\phi)] S_l^m(\cos\theta) \right], \quad (4.23)$$

and the vector of coefficients is written as

$$\mathbf{m} = \{a_0^0, a_1^0, a_1^1, b_1^1, a_2^0, a_2^1, b_2^1, a_2^2, b_2^2 \ldots\}. \quad (4.24)$$

A convenient feature of this expression is that the first coefficient is exactly the underlying uniform distribution of the parameter of interest.

Equivalently, we may allow the index m to specify the coefficient uniquely by allowing m to become negative. Then, define

$$f(\theta, \phi) = a_0^0 + \sum_{l=1}^{L} \left[a_l^0 P_l(\cos\theta) + \left[\sum_{m=1}^{l} a_l^m \cos(m\phi) + \sum_{m=-1}^{-l} a_l^m \sin(m\phi) \right] S_l^m(\cos\theta) \right],$$
$$(4.25)$$

and (setting $a_l^{-m} = -b_l^m$)

$$\mathbf{m} = \{a_0^0, a_1^0, a_1^1, a_1^{-1}, a_2^0, a_2^1, a_2^{-1}, a_2^2, a_2^{-2} \ldots\}. \quad (4.26)$$

This is the convention we have used in our implementation. We define $f_i(\theta, \phi)$ as the function that corresponds to the i'th coefficient in this sequence, multiplied by the respective real spherical harmonic. Here, real spherical harmonics are defined by

$$Y_l^m(\theta, \phi) = \begin{bmatrix} \cos(m\phi) S_l^m(\cos\theta) & \text{for} \quad 0 \leq m \leq l; \\ \sin(m\phi) S_l^m(\cos\theta) & \text{for} \quad -l \leq m < 0. \end{bmatrix} \quad (4.27)$$

We also assume that $\mathbf{m}$ is finite-dimensional and belongs to $\mathbb{R}^N$, so that $i = 1, \ldots N$.

Surface spherical harmonics are a complete orthonormal set for L_2-functions on the unit sphere; hence any real function on a spherical surface can be represented (in the limit) with an expansion in these basis functions. Letting the expansion coefficients to be a function of radius allows any function in 3-D to be represented. Layers, or polynomials in radius are natural ways to represent radial dependence.

4.4.2 Parametrizing the log resistivity

As discussed earlier in this Section, in order to solve the practical inverse problem we would need a more compact way to store the model, which is otherwise described as the distribution of resistivity ρ (or conductivity $\sigma = 1/\rho$) everywhere on the staggered-grid cells - a space far too large to search in.

The duality between electrical conductivity σ and resistivity ρ gives us the choice of the property to parametrize. We shall see in Chapter 5 that an inversion with respect to electrical resistivity is closer to the linear problem, than the problem in which the electrical conductivity is used to describe the computational domain. However, linear parametrization of electrical resistivity (or conductivity) with respect to spherical harmonics creates a positivity constraint issue in the inversion. The physical reality requires $\rho > 0$ at every point in the domain. This condition does not necessarily hold if a spherical harmonic expansion with arbitrary coefficients is evaluated. This introduces an additional constraint into the global search and complicates the optimisation (Chapter 6). A common solution to this problem is the so-called *logarithmic parametrization*, in which we parametrize the $\log \rho$ rather than ρ. We employ $\log_{10} \rho$ for this purpose.

Thus, instead of parametrizing

$$\rho(\phi, \theta, r) = \eta_{\phi,\theta,r}(\mathbf{m}) = \sum_{i=0}^{N} f_i(\phi, \theta, r) \tag{4.28}$$

we use the $\log_{10}$ parametrization

$$\log_{10}(\rho(\phi, \theta, r)) = \eta_{\phi,\theta,r}(\mathbf{m}) = \sum_{i=0}^{N} f_i(\phi, \theta, r). \tag{4.29}$$

In general, the orthogonal functions $f_i(\phi, \theta, r)$ could be continuous in r, but we substitute surface spherical harmonics on a set of K layers.

Define the values $r_{min}^{(k)}$ and $r_{max}^{(k)}$ to denote the lower and upper boundaries of the k'th layer, the total number of layers being K. Then we would compute the value of ρ at a point using the expression

$$\log_{10}(\rho(\phi, \theta, r)) = \sum_{k=1}^{K} \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^{(k)} Y_l^m(\phi, \theta) \, \mathrm{I}(r_{min}^{(k)} < r \le r_{max}^{(k)}). \tag{4.30}$$

Here I is the identity function

$$\mathrm{I}(X) = \begin{cases} 1 \text{ if X is true,} \\ 0 \text{ otherwise.} \end{cases}$$

We could regard the spherical harmonic expansion coefficients $c_{lm}^{(k)}$, $l = 1, \ldots L$, $m = -l, \ldots l$, $k = 1, \ldots K$ to be the full set of our parametrization coefficients, equivalent to the vector $\mathbf{m} \in \mathbb{R}^N$ in our previous notation. Thus, the total number of spherical harmonic coefficients is $N = K[1 + \sum_{l=0}^{L}(2l + 1)]$, where L is the degree of the expansion, and K is the number of layers. The number of parameters for a single layer is $n = 1 + \sum_{l=0}^{L}(2l + 1)$.

Note that in the logarithmic parametrization ρ is no longer linear with respect to $c_{lm}^{(k)}$.

The electrical resistivity ρ in layer number k is independent of r and may be computed using the expression

$$\rho(\phi, \theta) = 10^{\sum_{l=0}^{L} \sum_{m=-l}^{l} c_{lm}^{(k)} Y_l^m(\phi,\theta)}. \tag{4.31}$$

No smooth transitions between layers are implied by the parametrization. These are enforced, where necessary, by the regularising model covariance operator $C_{\mathbf{m}}$ while minimising the penalty functional. The specific smoothing operators that could be used for the spherical harmonic parametrization are proposed in Section 4.4.3.

4.4.3 Smoothing operators for layered spherical harmonic parametrization

Consider a spherical harmonic parametrization $c_{lm}(r)$, $l = 1, \ldots L$, $m = -l, \ldots l$, as described in Section 4.4.2.

We would like to regularise this parametrization by adding a term to the penalty functional that penalises for large jumps in adjacent parameters - which will hopefully also penalise for spatial jumps in electrical resistivity/conductivity. Effectively, we would like to add to the penalty functional some sort of discretization to

$$w_1 \sum_{l,m} \left(\int \frac{\partial c_{lm}(r)}{\partial r} dr \right)^2 + w_2 \sum_{l,m} \left(\int l^2 (c_{lm}(r))^2 dr \right). \tag{4.32}$$

The first term of the sum would penalise for vertical discontinuities in the model, thus allowing the inversion to prefer a smoother model to one that has larger jumps that are not physically justified while the second term would penalise higher order terms of

spherical harmonics, that could otherwise introduce the unjustified amount of detail into the model structure. The multipliers w_1 and w_2 are arbitrary weights.

In practice, we would normally discretize the model further by choosing a set of layers over which these parameters are defined. For the layered spherical harmonic parametrization, our parameters would take form $c_{lm}^{(k)}$, $l = 1, \ldots L$, $m = -l, \ldots l$ and $k = 1, \ldots n$, where n is the number of layers in the model. Then the first term in Equation 4.32 would in fact correspond to adding large weights to the penalty functional if there are unjustified jumps in parameter values between layers.

We have already seen in Section 4.3.2 that in practice we would rather construct a combination of *smoothing* operators that would serve as pre-conditioners, so that $C_\mathbf{m}$ would be some combination of these smoothing operators, possibly applied more than once. Then, the operator that would regularise by assigning large weights to the unwelcome features of the model, if ever needed, could be constructed by inverting the operator $C_\mathbf{m}$. For this purpose, let us define two basic operators, from which we would construct $C_\mathbf{m}$.

Horizontal smoothing:

$C_H \equiv$ multiply by $c_l = l^{-\alpha/2}$. To downweight the effect of higher order spherical harmonic terms, take $\alpha \geq 0$. A reasonable weighting is achieved by setting $\alpha \approx 1$.

Vertical smoothing:

$C_V \equiv$ smooth $c_{lm}^{(k)}$ over the index k, via a variant of a mean filter. We would act by C_V on the vector of parameters for l, m harmonic, c_{lm}, across the layers. Then the smoothed vector of parameters could be $\hat{c}_{lm} = C_V c_{lm}$.

We could write an appropriate operator as

$$C_V = \begin{pmatrix} \beta & \frac{1-\beta}{2} & & & & \\ \frac{1-\beta}{2} & \beta & \frac{1-\beta}{2} & & 0 & \\ & \frac{1-\beta}{2} & \beta & & & \\ & & & \ddots & & \\ 0 & & \frac{1-\beta}{2} & \beta & \frac{1-\beta}{2} \\ & & & \frac{1-\beta}{2} & \beta \end{pmatrix}. \tag{4.33}$$

In general, $0 \le \beta \le 1$. For $\beta = 1/3$, this is exactly the mean filtering kernel: the smoothed parameter value is the mean of the original parameter value and the values of the two "adjacent" parameters. The closer β is to 1, the less smoothing occurs. In fact, if $\beta = 1$, C_V is the identity matrix, and no vertical smoothing occurs whatsoever. In general, it is a good practice to take $\beta > 1/2$, to avoid overweighing the parameters for the neighbouring layers with respect to the parameter of interest.

Then the smoothing operator could be written as, e.g.

$$C_{\mathbf{m}} = C_V C_H C_V C_V C_H C_V,$$

in which case the symmetric "square root" operator is

$$C_{\mathbf{m}}^{1/2} = C_V C_H C_V.$$

Chapter 5

Formal derivation of data

sensitivities

In this chapter we generalise the forward problem described in Chapter 3, to enable it to be solved with an arbitrary forcing. We then derive numerical expressions for the Jacobian and the penalty functional derivative of the generalised forward problem. These expressions can be readily implemented as numerical routines, which would allow us to run an inversion based on a variant of gradient search.

The analysis in this chapter is developed for a single frequency, but can readily be applied to multiple frequencies. Some of the operators do not depend on the frequency and may only be computed once in this case. However, most of the computations are frequency-dependent. Any dependence on frequency shall be clearly outlined.

5.1   Motivation

We would like to employ the analysis developed in Chapter 3 to study the dependence of the magnetic field **H** on the parameter of interest, electrical resistivity ρ. As we have seen in Section 4.2, the non-linear dependence of the magnetic field **H** on the electrical resistivity ρ through the Helmholtz equation (3.15) can be expressed in its finite difference approximation as a non-linear mapping (Equation 4.1)

$$\mathbf{h} = \gamma(\rho).$$

Here, $\mathbf{h} \in \mathbb{E}$ and $\rho \in \mathbb{G}$.

If we now introduce a small perturbation in ρ, we would like to know how this affects **h**. To perform this estimate, some kind of linearisation of the mapping γ would be required. If we introduce a small perturbation $\delta\rho$ around $\rho = \rho_0$, we could use Taylor's expansion to write an approximate expression for $\mathbf{h}_0 + \delta\mathbf{h} = \gamma(\rho_0 + \delta\rho)$ as follows:

$$\gamma(\rho_0 + \delta\rho) = \gamma(\rho_0) + \frac{\partial\gamma}{\partial\rho}\big|_{\rho_0}\delta\rho + O(\delta\rho^2). \qquad (5.1)$$

Capital O denotes higher order terms. If we keep in mind that both ρ and $\gamma(\rho)$ are vectors, then $\partial\gamma/\partial\rho$ is the Jacobian matrix for this non-linear mapping. Thus, to develop some kind of sensitivity analysis we would require an expression for $\partial\gamma/\partial\rho$.

Furthermore, we would like to consider the inverse to the problem we have just described: having observed a perturbation $\delta\mathbf{h}$ in the magnetic field, can we deduce anything about $\delta\rho$? Or, in fact, the more general problem: having observed the magnetic field **h**, can we deduce anything about the resistivity ρ on the grid? This question is known as the *inverse* problem. Since the mapping between these values is non-linear this problem

is non-trivial, and can only be solved numerically. It is well-known, that many gradient-based numerical solutions to such inverse problems (such as conjugate gradients) perform much more efficiently than those that do not use any information about gradients. Indeed, any numerical inverse solver for our non-linear problem would be based on the following simple principle: we choose a model (a particular value of vector ρ), then apply the non-linear function (γ) and obtain the resultant forward solution (in our case, h). Then we compare the result to the value (or values) we're interested in, such as the known entries of vector h, and try again, until our forward solution gets sufficiently close to the known value of magnetic field. Thus, we would have found a resistivity vector ρ (which is not necessarily unique) such that $h = \gamma(\rho)$. The available information about the function clearly plays a crucial role in the choice of the model during the search. We would normally like to go as far away as possible from purely random search, and use a strategy of some sort for choosing the model. Knowledge of the derivative (or, in our case, the Jacobian, since the value of interest h is a vector) of the function is invaluable for the construction of the inverse solution.

5.2 Analytic expression for the Jacobian $\psi'(\mathbf{m})$

The core of the forward solver, evoked in Section 5.1 as the operator γ, would explicitly depend on the frequency ω and on any forcing introduced into the system. This dependence should be clearly outlined, so let us write $\gamma_{\omega,s}(\rho)$ to denote the forward solver mapping $\gamma(\rho)$ that corresponds to a specific frequency ω and is forced by specific sources s, normally arising from the boundary conditions chosen in accordance with the setup described in Section 3.3.2. In terms of the three problem components $\eta(\mathbf{m})$, $\gamma_{\omega,s}(\rho)$ and $\lambda_j(\mathbf{h})$, introduced in Section 4.2, the complete non-linear parameter to data mapping can

now be expressed as (Equation 4.4)

$$\psi_j = \psi_j(\mathbf{m}) = \lambda_j(\gamma_{\omega,s}(\eta(\mathbf{m}))). \tag{5.2}$$

The complex data functional ψ_j depends on the location (ϕ, θ, r) for which this value is computed through index j, that each corresponds to a different location. The mapping also depends on frequency ω and the sources s through the core of the forward solver γ. This dependence, once outlined, shall be omitted in the future for clarity of notation, since the problem described here is only being solved for a single frequency and a single forcing.

The derivative of this expression may be written using the chain rule:

$$\psi_j'(\mathbf{m}) = \lambda_j'(\mathbf{h}) \, \gamma_{\omega,s}'(\rho) \, \eta'(\mathbf{m}). \tag{5.3}$$

or, equivalently,

$$\frac{\partial \psi_j}{\partial \mathbf{m}} = \frac{\partial \psi_j}{\partial \mathbf{h}} \frac{\partial \mathbf{h}}{\partial \rho} \frac{\partial \rho}{\partial \mathbf{m}}. \tag{5.4}$$

We may view the elements $\mathbf{m}$, ρ and $\mathbf{h}$ as vectors, while the expressions $\partial \mathbf{h}/\partial \rho$ and $\partial \rho/\partial \mathbf{m}$ are Jacobian matrices for the problems $\mathbf{h} = \gamma_{\omega,s}(\rho)$ and $\rho = \eta(\mathbf{m})$, respectively. Also, $\partial \psi_j/\partial \mathbf{h}$ is a row vector, that relates a variation in a single data functional to perturbations in each of the elements of vector $\mathbf{h}$. Clearly, the total derivative $\partial \psi_j/\partial \mathbf{m}$ is also a row vector, that relates a variation in the data functional to small perturbations in each of the model parameters m_i, $i = 1, \dots n$. The full Jacobian $\psi'(\mathbf{m})$, computed for a single frequency ω, is a matrix, composed of these row vectors.

5.3   Derivative of the data misfit

We could apply the above analysis to the problem outlined in Sections 5.1 and 4.2 and in Chapter 3. For example, ψ_j can be the computed complex C- or D-responses, discussed in the Introduction (see also *Fujii and Schultz* [2002]), while d_j would be the corresponding measurements, which are for our purposes from now on considered constant.

Let us introduce the vector of residuals $\mathbf{r} \in \mathbb{C}^m$ such that

$$r_j = \psi_j(\mathbf{m}) - \mathbf{d}_j, \tag{5.5}$$

where ψ_j and d_j are as defined in Sections 4.3 and 5.2.

Then Equation 4.9 may be rewritten as

$$\begin{aligned} R(\mathbf{m}) &= \mathbf{r}^* C_\mathbf{d}^{-1} \mathbf{r} & (5.6) \\ &= \sum_{i,j} c_{ij} \bar{r}_i r_j, & (5.7) \end{aligned}$$

where c_{ij} are the relevant entries of matrix $C_\mathbf{d}^{-1}$.

While it would be easy to compute the derivative of this expression for $\mathbf{r} \in \mathbb{R}^m$, for complex vector $\mathbf{r}$ the derivative is not so trivial, as we shall now see.[1]

We are currently only interested in the case when $C_\mathbf{d}$ is diagonal, with $c_{jj} = \delta_j^2$, or, equivalently, in the specific penalty functional described by Equation 4.13. Thus, in our case the data covariance operator is

$$C_\mathbf{d} = diag(\delta_j^2). \tag{5.8}$$

---

[1]In fact, it is easy to show that for real vector $g(\mathbf{x})$ and a symmetric matrix A such that $f(\mathbf{x}) = g(\mathbf{x})^T A \, g(\mathbf{x})$, partial derivative $\dfrac{\partial f}{\partial x_i} = 2 \dfrac{\partial g}{\partial x_i}^T A \, g(\mathbf{x})$.

Then

$$R_{\mathbf{d}}(\mathbf{m}) = \sum_{j=1}^{m} \frac{1}{\delta_j^2} \bar{r}_j r_j,$$ (5.9)

and

$$\frac{\partial R_{\mathbf{d}}}{\partial \mathbf{m}} = \sum_{j=1}^{m} \frac{1}{(\delta_j)^2} \frac{\partial}{\partial \mathbf{m}} (\bar{r}_j r_j) = \sum_{j=1}^{m} \frac{1}{(\delta_j)^2} \frac{\partial}{\partial \rho} (\bar{r}_j r_j) \frac{\partial \rho}{\partial \mathbf{m}}.$$ (5.10)

Unfortunately, complex mapping $r_j \mapsto \bar{r}_j r_j$ is not analytic, therefore the standard chain rule that is valid for real and complex analytic functions cannot be applied to decompose $\frac{\partial}{\partial \rho}(\bar{r}_j r_j)$ (Equation 5.10) into simpler partial derivatives that could be numerically implemented. To do this, we will need a short *Lemma*.

Lemma: Consider mappings $f : \mathbb{C} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{C}$, such that $f(\xi) = |\xi|^2$, and g is analytic. Let $\rho \in \mathbb{R}$. Then

$$\frac{\partial}{\partial \rho} \{ f(g(\rho)) \} = 2 \operatorname{Re} \left(\bar{g} \frac{\partial g}{\partial \rho} \right).$$ (5.11)

Proof: The mapping $f(g) = |g|^2 = g\bar{g}$ is not analytic. Following *Ahlfors* [1987], Chapter IA,

$$\frac{\partial f}{\partial \rho} = \frac{\partial f}{\partial g} \frac{\partial g}{\partial \rho} + \frac{\partial f}{\partial \bar{g}} \frac{\partial \bar{g}}{\partial \rho}.$$ (5.12)

But $\frac{\partial f}{\partial g} = \bar{g}$ and $\frac{\partial f}{\partial \bar{g}} = g$. Therefore,

$$\frac{\partial f}{\partial \rho} = \bar{g} \frac{\partial g}{\partial \rho} + g \frac{\partial \bar{g}}{\partial \rho}.$$ (5.13)

We note that since ρ is real, $\frac{\partial \bar{g}}{\partial \rho} = \left(\frac{\bar{\partial g}}{\partial \rho} \right).$

Hence

$$
\begin{aligned}
\frac{\partial f}{\partial \rho} &= \bar{g}\frac{\partial g}{\partial \rho} + g\left(\overline{\frac{\partial g}{\partial \rho}}\right) \\
&= \bar{g}\frac{\partial g}{\partial \rho} + \left(\overline{\bar{g}\frac{\partial g}{\partial \rho}}\right) \\
&= 2\,\mathrm{Re}\,\left(\bar{g}\frac{\partial g}{\partial \rho}\right).
\end{aligned}
$$

$\square$

Using the above *Lemma* and Equation 5.10, we find that

$$
\frac{\partial R_{\mathbf{d}}}{\partial \mathbf{m}} = \sum_{j=1}^{m} 2\,\mathrm{Re}\,\left(\frac{\bar{r}_j}{\delta_j^2}\frac{\partial r_j}{\partial \rho}\frac{\partial \rho}{\partial \mathbf{m}}\right) \tag{5.14}
$$

$$
= 2\,\mathrm{Re}\,\left(\sum_{j=1}^{m}\frac{\bar{r}_j}{\delta_j^2}\frac{\partial r_j}{\partial \mathbf{h}}\frac{\partial \mathbf{h}}{\partial \rho}\frac{\partial \rho}{\partial \mathbf{m}}\right) \tag{5.15}
$$

$$
= 2\,\mathrm{Re}\,\left(\mathbf{r}^{*}C_{\mathbf{d}}^{-1}\frac{\partial \psi}{\partial \mathbf{h}}\frac{\partial \mathbf{h}}{\partial \rho}\frac{\partial \rho}{\partial \mathbf{m}}\right). \tag{5.16}
$$

For convenience, let us introduce new vector $\tilde{\mathbf{r}} = C_{\mathbf{d}}^{-1}\mathbf{r}$, the vector of weighted residuals. Then

$$
\frac{\partial R_{\mathbf{d}}}{\partial \mathbf{m}} = 2\,\mathrm{Re}\,\left(\tilde{\mathbf{r}}^{*}\frac{\partial \psi}{\partial \mathbf{h}}\frac{\partial \mathbf{h}}{\partial \rho}\frac{\partial \rho}{\partial \mathbf{m}}\right). \tag{5.17}
$$

Both $\tilde{\mathbf{r}}^{*}$ and $\partial \psi/\partial \mathbf{h}$ are frequency-dependent quantities, so that Equation 5.14 only holds for a single frequency. If we're interested in more than one frequency, the above expression could be denoted $\partial R_{\omega}/\partial \mathbf{m}$. The total derivative for the multiple frequency problem (Equation 4.14) would of course be

$$
\frac{\partial R_{\mathbf{d}}}{\partial \mathbf{m}} = \frac{1}{2M}\sum_{\omega}\frac{\partial R_{\omega}}{\partial \mathbf{m}}. \tag{5.18}
$$

5.4 Generalised forward solver with arbitrary forcing

5.4.1 Motivation for generalised forward solver

Suppose we introduce a small perturbation into the Helmholtz equation 3.15, just like described in Section 5.1. The Helmholtz equation could now be written as

$$\nabla \times (\rho + \delta\rho)\nabla \times (\mathbf{H} + \delta\mathbf{H}) + i\omega\mu_0(\mathbf{H} + \delta\mathbf{H}) = 0. \tag{5.19}$$

Since the Helmholtz operator is linear in $\mathbf{H}$, and $\nabla \times \rho(\nabla \times \mathbf{H})$ is linear in ρ we can rewrite Equation 5.19 as

$$(\nabla \times \rho(\nabla \times \mathbf{H}) + i\omega\mu_0\mathbf{H}) + (\nabla \times \rho(\nabla \times \delta\mathbf{H}) + i\omega\mu_0\delta\mathbf{H}) +$$
$$\nabla \times \delta\rho(\nabla \times \mathbf{H}) + \nabla \times \delta\rho(\nabla \times \delta\mathbf{H}) = 0.$$

The first bracket cancels since $\nabla \times \rho(\nabla \times \mathbf{H}) + i\omega\mu_0\mathbf{H} = 0$ according to the unperturbed Helmholtz equation. Dropping the second order variation term $\nabla \times \delta\rho(\nabla \times \delta\mathbf{H})$, we are left with a new Helmholtz equation:

$$\nabla \times \rho(\nabla \times \delta\mathbf{H}) + i\omega\mu_0\delta\mathbf{H} = -\nabla \times \delta\rho(\nabla \times \mathbf{H}). \tag{5.20}$$

Thus calculation of sensitivities requires solving the same system of equations as the forward problem, but with different forcing and boundary conditions.[2] To solve the equation for the perturbation $\delta\mathbf{H}$ in the magnetic field we would need to propagate the perturbation $\delta\rho$ by forcing the equation with a general interior source, that depends on the unperturbed magnetic field $\mathbf{H}$ and on $\delta\rho$.

---

[2]Perturbed solution of (5.19) has the same boundary conditions as the unperturbed. Thus, the boundary conditions for (5.20) are zero ($\mathbf{H} + \delta\mathbf{H} \equiv \mathbf{H}$ on boundary $\Rightarrow \delta\mathbf{H} = 0$ on boundary).

Thus, to perform any kind of sensitivity analysis we would require a forward solver that can be forced with a more sophisticated source than that arising from the boundary condition only. This urges us to reconsider our analysis from Chapter 3 for a generalized Helmholtz equation with an arbitrary non-zero RHS, as follows:

$$\nabla \times \rho(\nabla \times \mathbf{H}) + i\omega\mu_0\mathbf{H} = f. \tag{5.21}$$

5.4.2 Linear algebraic formulation for Helmholtz equation with arbitrary forcing

The correction described by Equation 5.21 is equivalent to the quasi-static approximation to Maxwell's equations (equations 3.12–3.14), modified as follows:

$$\nabla \times \mathbf{H} = \mathbf{J} \tag{5.22}$$

$$\nabla \times \mathbf{E} = -i\omega\mu_0\mathbf{H} + f \tag{5.23}$$

$$\mathbf{J} = \sigma\mathbf{E} \tag{5.24}$$

Indeed, if we now take the curl of the first equation (divided by σ), and substitute second and third equations in, we would end up with Equation 5.21.

Note that $\nabla \cdot \mathbf{H}$ is no longer zero in this setting. In fact, the divergence of a curl is zero, therefore taking $\nabla \cdot$ of both sides of Equation 5.21 gives:

$$\nabla \cdot (i\omega\mu_0\mathbf{H}) = \nabla \cdot f,$$

or, equivalently,

$$\nabla \cdot \mathbf{H} = -i\frac{\nabla \cdot f}{\omega\mu_0}, \tag{5.25}$$

for the interior components of the magnetic field, using the fact that divergence is linear. This is an important point, since the identity $\nabla \cdot \mathbf{H} = 0$ valid for magnetic field satisfying Equation 3.15 with zero interior forcing has been used extensively to enhance convergence of iterative numerical computations (see also U&S, Section 3). As we conclude from Equation 5.25, this identity no longer holds for the generalised problem, unless the divergence of the interior forcing is also zero. Instead, the new identity 5.25 may still be used to enhance convergence.[3]

The integral equation formulation of the problem with an added interior forcing now looks like:

$$\oint \mathbf{H} \cdot d\mathbf{l} = \iint \mathbf{J} \cdot d\mathbf{S},$$

$$\oint \mathbf{E} \cdot d\mathbf{l} = -\iint i\omega\mu\mathbf{H} \cdot d\mathbf{S} + \iint f \cdot d\mathbf{S}, \qquad (5.26)$$

$$\mathbf{J} = \sigma\mathbf{E}.$$

Let us briefly review the staggered-grid discretization of the original forward solver, fully described in Chapter 3. Consider an *H-cell* bounded by *H-faces* (see Figure 3.1). The magnetic field components are defined on *H-edges* (edge centers). Electric fields are defined on *H-faces* (which correspond to the centers of *E-edges*). Length elements l_r, l_ϕ and l_θ are defined on *H-edges*. We can also define elementary areas S_r, S_ϕ and S_θ on both the *H-faces* and on the *H-edges* (which correspond to the *E-faces*). The latter are the elementary surface areas of a surface perpendicular to an *H-edge*, passing through the edge center. Let us use the notation $\mathbb{E} = \mathbb{E}_i \cup \mathbb{E}_b$ and $\mathbb{F}$ to denote the vector spaces of components defined on the *H-edges* and on the *H-faces*, respectively, as specified in Section 3.4.

---

[3]There is a technical subtlety with this approach that arises when the adjoint solution is implemented; here we omit the details.

The RHS of the second order difference equation (18) in U&Swould now be written as

$$-i\omega\mu_0 S_\phi(i^+,j^-,k^-)H_\phi(i,j,k) + S_\phi(i^+,j^-,k^-)f_\phi(i,j,k). \tag{5.27}$$

The equivalent linear algebraic formulation would be to add the term $D_{S_i^E}\mathbf{f}$ to the RHS of the Helmholtz equation 3.39, where $\mathbf{f} \in \mathbb{E}_i$ is the vector that contains the interior forcing components.

Thus, we can now write the generalised system of linear equations for both the interior and boundary components of the magnetic field (c.f. Equation 3.43). Let $\mathbf{b} \in \mathbb{E}_b$ be the specified boundary conditions and $\mathbf{f} \in \mathbb{E}_i$ be the specified interior forcing. Then the magnetic field components

$$\mathbf{h} = \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix}$$

satisfy the system of equations

$$
\begin{aligned}
A_{\rho,\omega} D_l \mathbf{h}_i &= -B_\rho \mathbf{b} + D_S \mathbf{f} \\
\mathbf{h}_b &= \mathbf{b}
\end{aligned}
\tag{5.28}
$$

where D_l is the diagonal operator pre-multiplying the interior components on edges by edge lengths; D_S is the diagonal operator pre-multiplying the interior components on edges by perpendicular surface elements; and $A_{\rho,\omega}$ and B_ρ are specified by Equations 3.41 and 3.42, respectively.

5.4.3   Generalised Helmholtz operator

We complete the preliminary analysis developed above by defining a new linear algebraic operator, that corresponds to the generalised Helmholtz operator, acting on the full set of finite difference magnetic field components (both on the interior and on the boundary edges of the staggered-grid).

Consider vector $\mathbf{s} \in \mathbb{E}$ that we define as

$$\mathbf{s} = \begin{bmatrix} \mathbf{f} \\ \mathbf{b} \end{bmatrix} \tag{5.29}$$

that could be thought of, collectively, as the source terms of the system. The two vector components will normally have different units.

Let us also remind ourselves that $I_b : \mathbb{E}_b \to \mathbb{E}_b$ has been defined as the identity matrix acting on boundary edges of the grid (which should not be confused with the operator I_{-b}, that has a completely different definition - see Section 3.5.1).

Operator $M_{\rho,\omega} : \mathbb{E} \to \mathbb{E}$ is such that

$$M_{\rho,\omega} \begin{bmatrix} \mathbf{h}_i \\ \mathbf{h}_b \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{b} \end{bmatrix}. \tag{5.30}$$

Equivalently, $M_{\rho,\omega}\mathbf{h} = \mathbf{s}$. For

$$M_{\rho,\omega} = \begin{pmatrix} D_S^{-1} A_{\rho,\omega} D_l & D_S^{-1} B_\rho \\ 0 & I_b \end{pmatrix} \tag{5.31}$$

equations 5.28 and 5.30 are identical.

5.4.4 Numerical expression for $\partial h/\partial\rho$

We can now use the numerical operator scheme we have devised to perform sensitivity analysis of the forward problem. Suppose that the linearised relationship between a perturbation $\delta\rho$ in the electrical resistivity of the model and the respective perturbation δh in the magnetic field can be described by an operator $J : \mathbb{G} \to \mathbb{E}$, so that $\delta h = J \delta\rho$. Clearly, J is an $|\mathbb{E}| \times |\mathbb{G}|$ matrix. Let us write J in terms of the elementary operators that are already at our disposal.

Consider a small perturbation of equation 5.30,

$$M_{\rho+\delta\rho,\omega}(\mathbf{h} + \delta\mathbf{h}) = \mathbf{s}. \tag{5.32}$$

We note, that at $\omega = 0$ vector Helmholtz operator is linear in ρ. The following important identities can be stated, using the elementary decompositions 3.41 and 3.42:

$$A_{\rho+\delta\rho,\omega} = A_{\rho,\omega} + A_{\delta\rho,0} \tag{5.33}$$

and

$$B_{\rho+\delta\rho} = B_\rho + B_{\delta\rho}. \tag{5.34}$$

By the definitions of operators $M_{\rho,\omega}$ (Equation 5.31) and I_{-b} (Equation 3.33) we obtain

$$M_{\rho+\delta\rho,\omega} = M_{\rho,\omega} + I_{-b}M_{\delta\rho,0}. \tag{5.35}$$

It is easy to visualise $I_{-b}M_{\delta\rho,0}$ by noticing that post-multiplication of a matrix by I_{-b} has the effect of deleting all rows corresponding to the boundary edges. Thus,

$$I_{-b}M_{\delta\rho,0} = \begin{pmatrix} D_S^{-1}A_{\delta\rho,0}D_l & D_S^{-1}B_{\delta\rho} \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} D_S^{-1}C_i^T D_{\delta\rho^F}CD_{l^E} \\ 0 \end{pmatrix} \qquad (5.36)$$

is just the real part of the relevant finite difference Helmholtz operator, restricted to the interior edges of the domain (c.f. Equation 3.35). Using identity 5.35, we can now state Equation 5.32 as

$$M_{\rho,\omega}h + M_{\rho,\omega}\delta h + I_{-b}M_{\delta\rho,0}h + I_{-b}M_{\delta\rho,0}\delta h = s.$$

Both $M_{\rho,\omega}h$ and s cancel. Ignoring second order terms, get the equation we would solve to obtain a perturbation in the magnetic field,

$$M_{\rho,\omega}\delta h = \tilde{s}, \qquad (5.37)$$

where $\tilde{s} = -I_{-b}M_{\delta\rho,0}h$.

We recognize this equation as the generalised forward problem with a purely interior forcing, just like predicted by the preliminary analysis in Section 5.4.1. Zero forcing from boundary conditions in Equation 5.37 makes perfect physical sense for invariant boundary conditions: we would not expect them to change for the perturbed system.

Note that $\tilde{s} = -I_{-b}M_{\delta\rho,0}h$ is in fact linearly dependent on $\delta\rho$. Let us make this dependence explicit by rearranging the elementary operators in a clever way. Full decomposition

of the expression for $\tilde{s}$ based on Equation 5.36 yields

$$\tilde{s} = -I_{-b}D_{SE}^{-1}C^T D_{\delta \rho^F} C D_{lE}\mathbf{h}. \tag{5.38}$$

Let us remind ourselves that the diagonal operator $D_{\delta \rho^F} : \mathbb{F} \to \mathbb{F}$ is a diagonal operator acting on the *H-faces* of the staggered-grid, with the entries of the averaged resistivity on faces aligned along the diagonal. Using Equation 3.31, $D_{\delta \rho^F} = D_{lF}D_{L\delta \rho}D_{SF}^{-1}$ (see Section 3.5.1 for the detailed derivation and operator definitions).

Diagonal operators may of course be freely permuted. Let us define operator $\tilde{L} = D_{SF}^{-1}D_{lF}L$, so that $\tilde{L} : \mathbb{G} \to \text{Re}\,(\mathbb{F})$. Then vector $\delta \rho^F$ may be computed using the expression

$$\delta \rho^F = D_{SF}^{-1}D_{lF}L\delta \rho = \tilde{L}\delta \rho. \tag{5.39}$$

On the other hand, we could write vector $\mathbf{e} \in \mathbb{F}$,

$$\mathbf{e} = C D_{lE}\mathbf{h}, \tag{5.40}$$

as a diagonal operator by arranging its elements along the diagonal of a matrix $D_e : \mathbb{F} \to \mathbb{F}$,

$$D_e = diag(\mathbf{e}). \tag{5.41}$$

Then

$$D_{\delta \rho^F} C D_{lE}\mathbf{h} = D_e\delta \rho^F = D_e\tilde{L}\delta \rho, \tag{5.42}$$

and the setup may be completed by introducing the operator $E_h : \mathbb{G} \to \mathbb{E}$ so that

$$E_h = D_{SE}^{-1}I_{-b}C^T D_e\tilde{L}. \tag{5.43}$$

The notation has been chosen to remind the reader that the range of the operator is the space of the edges, and to stress the dependence on the magnetic field $\mathbf{h}$.

We can finally rewrite the system 5.32 as follows:

$$M_{\rho,\omega}\delta\mathbf{h} = \tilde{\mathbf{s}}, \tag{5.44}$$

where $\tilde{\mathbf{s}} = -E_h\delta\rho$.

Hence (assuming that the Helmholtz operator is invertible) we can write the dependence between the perturbation $\delta\rho$ and the respective variation $\delta\mathbf{h}$ in the magnetic field as

$$\delta\mathbf{h} = -M_{\rho,\omega}^{-1}E_h\delta\rho, \tag{5.45}$$

which implies that the Jacobian matrix for the generalised forward solver is

$$J = -M_{\rho,\omega}^{-1}E_h = -M_{\rho,\omega}^{-1}D_{S^E}^{-1}I_{-b}C^TD_e\tilde{L}. \tag{5.46}$$

In practice to compute the variation in the magnetic field that results from a small perturbation $\delta\rho$ in the resistivity model, for a given frequency, we would perform the following steps:

1. Run the forward solver $M_{\rho,\omega}\mathbf{h} = \mathbf{s}$ to obtain $\mathbf{h}$.

2. Compute $\mathbf{e_1} \in \mathbb{F}$ such that $\mathbf{e_1} = CD_{l^E}\mathbf{h}$.

3. Compute $\mathbf{e_2} \in \text{Re}\,(\mathbb{F})$ such that $\mathbf{e_2} = \tilde{L}\delta\rho$.

4. Compute $\mathbf{e} \in \mathbb{F}$ such that $\mathbf{e}$ is the component-wise product of the two vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ without any complex conjugation.

5. Compute $\tilde{\mathbf{h}} \in \mathbb{E}$ such that $\tilde{\mathbf{h}} = C\mathbf{e}$, and zero any resulting boundary values.

6. Compute the forcing $\tilde{\mathbf{s}} = -D_{S\dot{E}}^{-1}\tilde{\mathbf{h}}$.

7. Run the forward solver $M_{\rho,\omega}\delta\mathbf{h} = \tilde{\mathbf{s}}$ to obtain $\delta\mathbf{h}$.

5.5 Numerical scheme: direct and adjoint formulations

We now possess a numerical expression for the Jacobian for the core of the forward problem, $\partial\mathbf{h}/\partial\rho$. We would like to complete the setup with the linear algebraic expressions for Jacobians of the other two parts of the problem, the parametrization and the data functionals, namely, $\partial\rho/\partial\mathbf{m}$ and $\partial\psi/\partial\mathbf{h}$.

5.5.1 Linearised data functionals

Consider a general non-linear data functional such as described by $\psi_j = \lambda_j(\mathbf{h})$ (Section 4.2). We are interested in a numerical approximation of its partial derivative, $\partial\psi_j/\partial\mathbf{h}$ with respect to the magnetic field vector. Thus, we would like to derive a linearisation of this data functional with respect to the magnetic field, such that $\delta\psi_j/\delta\mathbf{h}$ may be approximated by a vector for small $\delta\mathbf{h}$.

Thus, we write a linearisation

$$\hat{\psi}_j = \mathbf{g}_j^*\mathbf{h}, \quad j = 1,\ldots m. \tag{5.47}$$

for some complex vector $\mathbf{g}_j$. Data functional ψ_j depends on the location of the observatory (ϕ, θ, r) through index j, but there is also implicit dependence on frequency, which

we are going to imply for analytic expressions. We shall however make this dependence very clear when we derive the linear algebraic formulations since it would be critical for any numerical computations.

Data functionals of interest all have the form

$$\psi_j(\mathbf{h}) = \psi(\mathbf{h}(x_j)) \tag{5.48}$$

where ψ is some (possibly non-linear) functional, and $\mathbf{h}(x_j)$ is a triplet of complex components (h_ϕ, h_θ, h_r) at the chosen location $x_j = (\phi, \theta, r)$ in the problem domain. So that, in general, we may write

$$\mathbf{h}(x_j) = L_j^T \mathbf{h}, \tag{5.49}$$

and L_j would be a general real sparse matrix, that would allocate the weights to the magnetic field components in the neighbourhood of the chosen location. Then we may write the data functional of the type that we're interested in as $\psi \equiv \psi(h_\phi, h_\theta, h_r)$, the components of the magnetic field all computed at a single location. If we decompose the sparse matrix L_j as

$$L_j = (L_j^\phi, L_j^\theta, L_j^r), \tag{5.50}$$

then the derivative of the mapping ψ_j with respect to $\mathbf{h}$ may be written as

$$\mathbf{g}_j^* = \frac{\partial \psi}{\partial h_\phi}(L_j^\phi)^* + \frac{\partial \psi}{\partial h_\theta}(L_j^\theta)^* + \frac{\partial \psi}{\partial h_r}(L_j^r)^*. \tag{5.51}$$

Clearly, unless ψ is linear in $\mathbf{h}$, the derivatives in the above expression would depend on $\mathbf{h}$, and therefore on all the information that is required to compute $\mathbf{h}$ (including frequency ω, the forcing and the model ρ).

Thus, we have defined a general complex mapping g_j^* : $\mathbb{E} \rightarrow \mathbb{C}$, or equivalently g_j : $\mathbb{C} \rightarrow \mathbb{E}$ (which is equivalent to saying that the column vector $g_j \in \mathbb{E}$). Multiplication of a vector defined on the edges of the staggered-grid by operator g_j^* would convert this vector into a single complex component, defined at a particular location anywhere in the domain. Implementation of this operator would require knowledge of the analytic derivatives of ψ with respect to each directional component of the magnetic field, and the implementation of a sparse operator L_j.

In our particular implementation, L_j is evaluated once per observatory location and stored as a sparse matrix. The non-zero components correspond to the nearest grid edges to the observatory, and are determined by the distances to their mid-points. Multiplication by L_j implements bi-linear interpolation. For the specific case of c response transfer functions (Section 2.3)

$$\psi_j(h_\phi, h_\theta, h_r) = \frac{R \tan \theta}{2} \frac{h_r}{h_\theta} \tag{5.52}$$

for a single frequency, where (h_ϕ, h_θ, h_r) are the magnetic field components at the observatory location. Then,

$$\frac{\partial \psi}{\partial h_\phi} = 0, \quad \frac{\partial \psi}{\partial h_\theta} = -\frac{R \tan \theta}{2} \frac{h_r}{h_\theta^2}, \quad \frac{\partial \psi}{\partial h_r} = \frac{R \tan \theta}{2} \frac{1}{h_\theta} \tag{5.53}$$

are used to calculate g_j^* using Equation 5.51. The magnetic field (h_ϕ, h_θ, h_r) is calculated using Equation 5.49.

We also define the operator G_ω^* : $\mathbb{E} \rightarrow \mathbb{C}^{M_\omega}$, together with its adjoint G_ω : $\mathbb{C}^{M_\omega} \rightarrow \mathbb{E}$ that we shall soon require. If we recall our definition of vector g_j, we shall see that G_ω is a complex matrix such that its j'th column is g_j. Let us also recall that accord-

ing to the linearisation (5.47), $\delta\psi_j = \mathbf{g}_j^* \delta\mathbf{h}$. Thus the complex mapping $\psi_\omega(\mathbf{m})$ (vector of components $\psi_j(\mathbf{m})$ for all locations, grouped by frequency) satisfies $\delta\psi_\omega(\mathbf{m}) = G_\omega^* \delta\mathbf{h}$, for a perturbation $\delta\mathbf{h}$ in the magnetic field, caused in turn by a perturbation in the model.

5.5.2 Sensitivity with respect to parametrization coefficients

The non-linear parametrization mapping is defined as $\rho = \eta(\mathbf{m})$ (Section 4.2). To describe the Jacobian of this mapping in detail, we would need to specify the parametrization, while we would like to keep the analysis general at this point. Thus, let us consider the linear relation $\delta\rho = P \, \delta\mathbf{m}$, so that $P : \mathbb{R}^n \to G$, $P^T : G \to \mathbb{R}^n$. We see that in this definition operator P may be regarded as the linear operator that approximates the matrix of partial derivatives. In practice, we would normally compute entries of operator P using the analytic derivatives $\partial\rho_j/\partial m_i = \partial\rho(\phi, \theta, r)/\partial c_{lm}$. For a linear parametrization, these values correspond to the respective orthonormal functions, computed at the chosen point (ϕ, θ, r), that corresponds to the centre of the j'th cell of the staggered-grid . In some cases of non-linear parametrizations, such as the logarithmic parametrization, the entries of operator P would depend on ρ.

To see how this may be true, consider the example of the $\log_{10}$ parametrization by layered surface spherical harmonics degree and order L (Section 4.4). Then (Equation 4.30)

$$\log_{10}(\rho(\phi, \theta, r)) = \sum_{k=1}^{K} \sum_{l=1}^{L} \sum_{m=-l}^{l} c_{lm}^{(k)} Y_{lm}(\phi, \theta) \; \mathrm{I}(r_{min}^{(k)} < r \leq r_{max}^{(k)}).$$

The electrical resistivity ρ in layer number k may be computed using the expression (Equation 4.31)

$$\rho(\phi, \theta) = 10^{\sum_{l=1}^{L} \sum_{m=-l}^{l} c_{lm}^{(k)} Y_{lm}(\phi, \theta)}.$$

Then for a single layer k

$$\frac{\partial}{\partial c_{lm}^{(k)}}(\log_{10}(\rho(\phi, \theta))) = \frac{1}{\rho(\phi, \theta)\ln(10)}\frac{\partial\rho(\phi, \theta)}{\partial c_{lm}^{(k)}} = Y_{lm}(\phi, \theta),$$

hence

$$\frac{\partial\rho(\phi, \theta)}{\partial c_{lm}^{(k)}} = \rho(\phi, \theta)Y_{lm}(\phi, \theta)\ln(10). \tag{5.54}$$

5.5.3 Direct versus adjoint approaches

We have now fully described the numerical definitions required to write down the linear expression that we would use for data sensitivity analysis. We have seen that the partial derivatives encountered in Equations 5.4 and 5.14 may be approximated by concatenations of linear operators, as follows,

$$
\begin{array}{rcll}
\partial\psi_\omega/\partial\mathbf{h} & = & G_\omega^* & \text{(Section 5.5.1)} \\[1mm]
\partial\mathbf{h}/\partial\rho & = & -M_{\rho,\omega}^{-1}E_h & \text{(Section 5.4.4)} \\[1mm]
\partial\rho/\partial\mathbf{m} & = & P & \text{(Section 5.5.2)}
\end{array}
$$

while the partial derivative with respect to $\mathbf{h}$ of a single data functional $\psi_j \equiv \psi_\omega^j$ can be written as

$$\partial\psi_\omega^j/\partial\mathbf{h} = \mathbf{g}_j^*,$$

where vector $\mathbf{g}_j$ is the j'th column of operator G_ω.

We can now state the linear algebraic expression for the derivative of a single data

functional (Equation 5.4),

$$\frac{\partial \psi_\omega^j}{\partial \mathbf{m}} = -\mathbf{g}_j^* M_{\rho,\omega}^{-1} E_h P. \tag{5.55}$$

The value of this expression is a complex row vector $1 \times N$. To compute it, we would need to apply the sequence of operators to every column of P; i.e. run the whole computation N times for each frequency, looping over the model parameters. For a large model space, if the parametrization is detailed enough to include a large number of parameters N, evaluation of this expression would be computationally expensive. To compute the value of a single data functional (or values for a set of data functionals for a given frequency) using Equation 5.55, $N + 1$ runs of the generalised forward solver would be required. We would need to run it once to compute the underlying magnetic field $\mathbf{h}$, and then once for each of the parametrization coefficients.

In some cases, when the number of the data functionals for a single frequency is much less than the number of parameters in the model, the so-called *adjoint* approach is preferable, meaning that we compute the adjoint of the data functional derivative described by Equation 5.55. Since the parameter vector $\mathbf{m}$ is real, this is equivalent to computing the derivative of $\bar{\psi}_\omega^j$ with respect to $\mathbf{m}$.

The expression we would utilize in this case is

$$\frac{\partial \bar{\psi}_\omega^j}{\partial \mathbf{m}} = -P^T E_h^* (M_{\rho,\omega}^{-1})^* \mathbf{g}_j. \tag{5.56}$$

Now, $\frac{\partial \bar{\psi}_\omega^j}{\partial \mathbf{a}}$ is a complex column vector $N \times 1$, which is computed by applying the operator $P^T E_h^* (M_{\rho,\omega}^{-1})^*$ to the column vector $\mathbf{g}_j \in \mathbb{E}$. Hence, in this setting we must run the computation once (for every frequency) to obtain the full derivative of a single data functional. To compute these values, we would need to run the adjoint of the generalised

forward solver once to compute **h**, and then once for each data functional, $M + 1$ times altogether. More precisely, to obtain the Jacobian at frequency ω we would apply the sequence of operators $M_\omega + 1$ times, where M_ω is the number of observations for this particular frequency.

In many cases, such as when a large number of parameters is required to fully describe the model, this approach is preferable. To utilize it, we would need to derive and implement the expression for all the adjoint operators. Once computed, we would conjugate $\partial \bar{\psi}^j_\omega / \partial \mathbf{m}$ again to obtain $\partial \psi^j_\omega / \partial \mathbf{m}$. Since in our problem the global data are very sparse, the data space domain is relatively low-dimensional. Therefore, the adjoint approach would allow us to consider far more complex parametrizations than we could have afforded otherwise.

Similarly, the derivative of the least squares penalty functional (Equation 5.17) may be calculated using

$$\frac{\partial R_\omega}{\partial \mathbf{m}} = -2 \operatorname{Re} \left(\tilde{\mathbf{r}}^*_\omega G^*_\omega M^{-1}_{\rho,\omega} E_h P \right) \tag{5.57}$$

(with the dependence on ω explicitly specified).

The adjoint of this expression is

$$\left(\frac{\partial R_\omega}{\partial \mathbf{m}} \right)^{\mathrm{T}} = -2 \operatorname{Re} \left(P^T E^*_h (M^{-1}_{\rho,\omega})^* G_\omega \tilde{\mathbf{r}}_\omega \right). \tag{5.58}$$

This has been obtained by transposing the composition of operators in Equation 5.57. Since we take the real part of that composition, the result stays the same. Then $G_\omega \mathbf{r}_\omega \in \mathbb{E}$ is a column vector, and the multiplication by $P^T E^*_h (M^{-1}_{\rho,\omega})^*$ has to be performed once for every frequency.

As we see, if we run the inverse solver based on the values of the penalty functional and its derivative, the adjoint approach is going to be far more efficient. In fact, as opposed to running the forward solver $n + 1$ times to obtain this expression in the model space (Equation 5.57), if we do so in the data space we only require a single additional run to compute the derivative after the penalty functional has been evaluated. So that, to compute $\partial R_\omega / \partial m$ using the adjoint expression (Equation 5.58) we would need to perform the computations twice in total for every frequency: once we would run the forward solver to evaluate h and the penalty functional, and the second time we would run the adjoint to evaluate the derivative. As we shall now see, running the adjoint computations require solving a different system of equations, but it is not significantly more complicated than the original system. In essence, the data space approach allows us to evaluate the derivative for the same cost as the value of the functional itself.

While $J = -M_{\rho,\omega}^{-1} E_h$ is a numerical representation of $\partial h / \partial \rho$, $J^* = -E_h^*(M_{\rho,\omega}^{-1})^*$ is a numerical representation of its complex conjugate $(\partial h / \partial \rho)^*$, or, equivalently, $\partial \bar{h} / \partial \rho$, since ρ is real. To compute the adjoint operator $J^* : \mathbb{E} \to \mathbb{G}$ we need some means to evaluate E_h^* and $(M_{\rho,\omega}^{-1})^*$.

First we note, that the adjoint of an inverse is the inverse of an adjoint (true for every bounded linear transformation defined on a finite dimensional or Hilbert space), so

$$(M_{\rho,\omega}^{-1})^* = (M_{\rho,\omega}^*)^{-1}.$$

For E_h^*, we remind ourselves that (see Section 5.4.3)

$$E_h = D_{SE}^{-1} I_{-b} C^T D_e \tilde{L}.$$

We note that D_e is a complex diagonal operator, therefore $D_e^* = D_{\bar{e}}$. Also, the adjoint of operator $C^T : \mathbb{F} \to \mathbb{E}$ is quite clearly operator $C : \mathbb{E} \to \mathbb{F}$ (see definitions of operators C and C^T in Section 3.5.1). Hence $E_h^* : \mathbb{E} \to \mathbb{G}$ can be expressed as

$$E_h^* = \tilde{L}^T D_{\bar{e}} C I_{-b} D_{SE}^{-1}. \tag{5.59}$$

We end up with the following expression for the adjoint of J:

$$J^* = \tilde{L}^T D_{\bar{e}} C I_{-b} D_{SE}^{-1} (M_{\rho,\omega}^*)^{-1}. \tag{5.60}$$

Thus, to evaluate J^* we only need an expression for $M_{\rho,\omega}^*$.

5.5.4 Adjoint of the generalised Helmholtz operator

Recall that (see Equation 5.31)

$$M_{\rho,\omega} = \begin{pmatrix} D_S^{-1} A_{\rho,\omega} D_l & D_S^{-1} B_\rho \\ 0 & I_b \end{pmatrix}.$$

We see that the adjoint of this operator is

$$M_{\rho,\omega}^* = \begin{pmatrix} D_l A_{\rho,\omega}^* D_S^{-1} & 0 \\ B_\rho^* D_S^{-1} & I_b \end{pmatrix}. \tag{5.61}$$

We can now employ our knowledge of operator $A_{\rho,\omega}$ to find its complex conjugate. We know that $A_{\rho,\omega}$ is complex symmetric, real everywhere except the diagonal, where its imaginary part comes from the $i\omega\mu D_{SE_i}$ term only. Therefore, its hermitian conjugate

(adjoint) satisfies

$$A_{\rho,\omega}^{*} = A_{\rho,-\omega}. \tag{5.62}$$

We remind ourselves that the only complex component of $M_{\rho,\omega}$ is due to the $i\omega\mu D_{S^{E_i}}$ term on the diagonal. Hence

$$M_{\rho,\omega}^{*} = M_{\rho,-\omega}^{T}, \tag{5.63}$$

and

$$M_{\rho,\omega}^{*} = \begin{pmatrix} D_l A_{\rho,-\omega} D_S^{-1} & 0 \\ B_\rho^T D_S^{-1} & I_b \end{pmatrix}. \tag{5.64}$$

5.6 The Jacobian matrix

We define the full Jacobian, separately for every frequency, as $\dfrac{\partial\psi_\omega}{\partial\mathbf{m}}$, an $(M_\omega \times N)$-dimensional complex matrix of partial derivatives of each data functional with respect to each model parameter. This quantity may be implemented using the numerical scheme devised in Sections 5.4 and 5.5.

We have already seen (Section 5.4) that a perturbation $\delta\mathbf{h}$ in the magnetic field can be viewed as being linearly related to the respective perturbation $\delta\rho$ in the electrical resistivity of the model. They are related by an operator $J : \mathbb{G} \to \mathbb{E}$, so that $\delta\mathbf{h} = J\,\delta\rho$. Operator J has explicit dependence on ω, ρ and $\mathbf{h}_0$ - the point at which the gradient is computed. Then $J^* : \mathbb{E} \to \mathbb{G}$.

As we have seen,

$$J = -M_{\rho,\omega}^{-1} D_{SE}^{-1} I_{-b} C^T D_e \tilde{L}$$

and

$$J^* = \tilde{L}^T D_{\tilde{e}} C I_{-b} D_{SE}^{-1} (M_{\rho,-\omega}^{T})^{-1}. \tag{5.65}$$

Diagonal operator D_e is such that $\mathbf{e} = CD_l\mathbf{h}$.

Then the *direct* expression for the Jacobian entries is

$$\frac{\partial \psi_\omega^j}{\partial \mathbf{m}} = -\mathbf{g}_j^* JP, \tag{5.66}$$

and *adjoint* is

$$\frac{\partial \bar{\psi}_\omega^j}{\partial \mathbf{m}} = -P^T J^* \mathbf{g}_j. \tag{5.67}$$

We see that $J^*\mathbf{g}_j \in \mathbb{G}$ is in general complex, and so is the vector $P^T J^*\mathbf{g}_j$. The results of the direct and adjoint expressions are hermitian conjugates of each other. If both expressions are implemented, this fact may be used for testing the self-consistency of the numerical algorithm. The expressions (5.57) and (5.58) can also be used for this purpose.

5.7 Derivative of regularised least-squares problem

Recall the definition of the regularised least squares penalty functional (Equation 4.5)

$$R(\mathbf{m}) = (\psi(\mathbf{m}) - \mathbf{d})^H C_\mathbf{d}^{-1} (\psi(\mathbf{m}) - \mathbf{d}) + \mu(\mathbf{m} - \mathbf{m}_0)^T C_\mathbf{m}^{-1}(\mathbf{m} - \mathbf{m}_0).$$

There are two equivalent formulations for the derivative of this penalty functional, based on the direct (5.57) and the adjoint (5.58) expressions. Either of the two may be preferable, depending on whether the data space or the model space is higher-dimensional, so we state both.

According to the direct formulation (Equation 5.57), for $C_\mathbf{m}^{-1}$ symmetric, the derivative

of this expression with respect to **m** is

$$\frac{\partial R_\omega}{\partial \mathbf{m}} = -2\,\mathrm{Re}\ (\tilde{\mathbf{r}}_\omega^* G_\omega^* JP) + 2\mu\,(\mathbf{m} - \mathbf{m}_0)^{\mathrm{T}} C_{\mathbf{m}}^{-1}, \tag{5.68}$$

where $\tilde{\mathbf{r}}_\omega = C_\delta^{-1}(\psi_\omega(\mathbf{m}) - \mathbf{d})$. Expression (5.68) is a real row vector $1 \times n$. In the adjoint formulation (Equation 5.58),

$$\left(\frac{\partial R_\omega}{\partial \mathbf{m}}\right)^{\mathrm{T}} = -2\,\mathrm{Re}\ \left(P^T J^* G_\omega \tilde{\mathbf{r}}_\omega\right) + 2\mu\,C_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}_0), \tag{5.69}$$

However, we have seen in Section 4.3.2 that it is possible to rewrite the expression for the least squares penalty functional as a function of $\hat{\mathbf{m}}$ (Equation 4.8)

$$R(\hat{\mathbf{m}}) = (\psi(C_{\mathbf{m}}^{1/2}\hat{\mathbf{m}} + \mathbf{m}_0) - \mathbf{d})^{\mathrm{H}}\ C_{\mathbf{d}}^{-1}\ (\psi(C_{\mathbf{m}}^{1/2}\hat{\mathbf{m}} + \mathbf{m}_0) - \mathbf{d}) + \mu\,\hat{\mathbf{m}}^{\mathrm{T}}\hat{\mathbf{m}},$$

Then we may compute the (adjoint) derivative as

$$\left(\frac{\partial R_\omega}{\partial \hat{\mathbf{m}}}\right)^{\mathrm{T}} = -2\,\mathrm{Re}\ \left(C_{\mathbf{m}}^{1/2} P^T J^* G_\omega \tilde{\mathbf{r}}_\omega\right) + 2\mu\,\hat{\mathbf{m}}. \tag{5.70}$$

Thus, effectively, instead of penalising the second term of the sum we're introducing a pre-conditioner on the first term, which should have the same effect, while resulting in greater numerical stability of the computations.

Gradient computation for a single frequency in the context of the inverse problem could now be summarised as follows.

1. Start with providing $\hat{\mathbf{m}}$ and possibly $\mathbf{m}_0$ (if $\mathbf{m}_0$ is not specified, this is equivalent to using a zero prior - i.e. a uniform Earth). These models may be provided by the inverse solver, and they have not been pre-conditioned in any way.

2. Compute the "smoothed" model $\mathbf{m} = C_{\mathbf{m}}^{1/2}\hat{\mathbf{m}} + \mathbf{m}_0$.

3. Use $\mathbf{m}$ to compute $\mathbf{r}_\omega = \psi_\omega(\mathbf{m}) - \mathbf{d}$, $\tilde{\mathbf{r}}_\omega = C_{\mathbf{d}}^{-1}\mathbf{r}_\omega$ and $P^T J^* G_\omega \tilde{\mathbf{r}}_\omega$.

4. Evaluate

$$
\begin{aligned}
R(\hat{\mathbf{m}}) &= \mathbf{r}_\omega^H C_{\mathbf{d}}^{-1} \mathbf{r}_\omega + \mu \, \hat{\mathbf{m}}^T \hat{\mathbf{m}}, \\
\frac{\partial R_\omega}{\partial \hat{\mathbf{m}}} &= -2 \, \mathrm{Re} \, \left(C_{\mathbf{m}}^{1/2} P^T J^* G_\omega \tilde{\mathbf{r}}_\omega \right) + 2\mu \, \hat{\mathbf{m}}.
\end{aligned}
$$

These are the values we would use for the inversion. Thus, although $\mathbf{m}$ corresponds to the physical reality, the inversion would run in the space of $\hat{\mathbf{m}}$, and the model $\mathbf{m}$ could be easily computed (Equation 4.7) when needed.

5.8 Generalisation for other penalty functionals

In the most general case, define the misfit (for a single frequency that we omit from the notation for the sake of clarity)

$$
R = \sum_{j=1}^{M} f(r_j), \tag{5.71}
$$

where $r_j = \psi_j(\mathbf{m}) - \mathbf{d}_j$ are data residuals.

Then

$$
\begin{aligned}
\frac{\partial R}{\partial \mathbf{m}} &= \sum_{j=1}^{M} f'(r_j) \frac{\partial \psi_j}{\partial \mathbf{m}} \tag{5.72} \\
&= \sum_{j=1}^{M} w_j r_j \frac{\partial \psi_j}{\partial \mathbf{m}}, \tag{5.73}
\end{aligned}
$$

and the weights are

$$
w_j = \frac{f'(r_j)}{r_j}.
$$

For instance, if the misfit were a function of residuals squared, the weights would be $w_j = 2$. For the least-squares misfit (Equation 4.13), the weights are $w_j = \bar{r}_j/r_j$, as we have seen from Equation 5.11.

Therefore, with a slight modification to the derivative calculations, other robust schemes based on non-L_2 penalty functionals (e.g. L_1 norm) can also be considered in this framework.

5.9   A note on the modularized numerical scheme

As we have seen in Sections 5.5.3 and Sections 5.6–5.7, both the Jacobian operator and the gradient of the penalty functional may be expressed as a concatenation of linear operators acting on some input vector or a set of vectors. It is possible to implement each of these operators as an abstract procedure, acting on a complex or real vector, the result of the procedure being another vector belonging to the relevant vector space on the staggered-grid. Then, any concatenation of these operators could be evaluated by calling each procedure in sequence. Whilst forming all the matrices and storing them as number arrays would be computationally expensive, computing the action of a sequence of operators on a given vector would be quite practical in general.

The formal division of the sensitivity calculations into steps as described here opens the possibility to develop the code in a modular way. This includes the development of a set of functional blocks, acting on vectors to produce other vectors. So that, for example, a numerical implementation of each of the linear operators defined in Section 3.5.1 involves writing abstract subroutines that implement multiplication and/or division by these operators. We think of the input vectors to each of these operators as of particular instances of abstract data types, designed to store the information corresponding to a

particular vector space on the staggered-grid in a convenient way.

This degree of abstraction and modularization allows implementing additional specific data functionals (e.g. d response; long period MT; ocean cables), as well as different model parametrizations, different source models, and other modular extensions. Initial steps in this direction have been taken, but a complete treatment of this problem is a task for the future.

To summarise, it is important to note that any linearised operator we discuss here is not being thought of as a matrix that makes sense by itself, but rather as a function that both has a range and a domain in the staggered-grid formulation.

Chapter 6

Methods of global optimisation and the inverse solver

Any practical inversion algorithm could be broken into two stages: searching for a reasonable initial model satisfying the data, which involves finding at least one optimal point in the parameter space, and then going on to find as many optimal points as possible. For an ill-posed inverse problem such as the one considered in this thesis, there may be an uncountable set of optimal models, satisfying the data to an acceptable level. However, even finding one such model may be a non-trivial task, if the prior knowledge is insignificant and hence the initial search space is large. This is the stage of the inversion we would like to concentrate on in this chapter.

The fast derivative computation based on the adjoint approach (Chapter 5) allows us to develop an inverse solver using one of the available non-linear gradient-based global optimisation techniques. These methods are not new in geophysics. Previously, non-linear conjugate gradient techniques (NLCG) have been successfully applied to local and regional magnetotelluric problems. A preconditioned NLCG algorithm for 2-D MT prob-

lems has been developed by *Rodi and Mackie* [2001]. An efficient model-space approach, which does not require computation or storage of full sensitivity matrix, has been developed for the 3-D MT problems by *Newman and Alumbaugh* [2000]. Variants of NLCG such as quasi-Newton methods also exist [*Haber*, 2005] and are widely applied. There are also related approaches, including the Gauss-Newton method (based on calculation of all or some of the Jacobian) both for 2-D and 3-D MT problems (see, for example, the data space approach of *Siripunvaraporn and Egbert* [2000]; *Siripunvaraporn et al.* [2005]).

Here, we discuss several well-known optimisation techniques that we have implemented for our problem, some of them modified from their traditional versions. We have confirmed that those non-gradient based techniques that we have implemented all require a much greater number of iterations to converge to an acceptable solution, than those using the information about the gradients. Although we briefly describe several other methods for completeness, we stress that the inverse solutions described in the following chapters have all been performed using the non-linear conjugate gradient algorithm (Section 6.1.2). We also briefly describe the related parallelization issues.

6.1 Multi-dimensional non-linear optimisation techniques

A great number of sophisticated general purpose global non-linear optimisation techniques have been developed. Among many others, these methods include the neighbourhood sampling algorithm [*Sambridge*, 1999a,b], a multistart simplex technique [*Törn and Zilinskas*, 1989], a genetic algorithm-based method [*Goldberg*, 1989], a shuffled complex evolution method [*Duan et al.*, 1993], the MultiDE differential evolution algorithm [*Storn and Price*, 1997].

Different methods are preferable for solving different problems, depending mostly on the amount of information we can afford to obtain about the cost function. Sometimes it is possible to evaluate the Hessian, thus allowing new possibilities for efficient optimisation, while in other cases evaluating the derivative of the cost function may be impractical.

Following *Zhigljavsky* [1991], we cite a set of guidelines that any global optimisation algorithm should adhere to. This is most applicable to the algorithms that do not require the evaluation of gradients, but rather aim to minimise a function using a number of function evaluations at a set of points in the domain. A global optimisation algorithm should have the following components:

- global – a means of distributing the initial points over the whole parameter space,

- local – an implementation of local descents,

- identifying prospective regions – the region in the parameter space which is covered by the global algorithm should be cut down every N iterations, for some N,

- control over the precision – the closer we get to the optimum point, the greater precision is required in calculating the function.

Let us consider our forward solver schematically as a non-linear functional f defined on a subspace of $\mathbb{R}^N$, where N is the number of parametrization coefficients. Then if $\mathbf{x} \in \mathbb{R}^N$ is a given vector of parametrization coefficients, $f(\mathbf{x}) \in \mathbb{R}$ is the penalty functional that we would like to minimize. The specifics of our problem are such that evaluating the value of this functional at a particular point in the domain is computationally expensive. Therefore, we're interested in designing an efficient multi-dimensional optimisation technique, that would require as few evaluations of f and its gradients as possible. After $f(\mathbf{x})$

has been computed, the additional evaluation of $f'(\mathbf{x})$ is possible and is comparable to the evaluation of $f(\mathbf{x})$ in terms of the related computational costs.

This scheme suggests that utilizing a parallel computing environment would be advantageous. Indeed, the necessity of a reasonable exchange of information between processors would be a negligible drawback compared to the opportunity to compute f for different values of $\mathbf{x}$ simultaneously (Section 6.2).

Here, we describe a generalised simplex (Nelder-Mead) algorithm as an example of a parallelizable non-gradient based global optimisation method, as well as two methods that require evaluating the gradients of the cost function: the conjugate gradient algorithm (CG), and the regularised steepest descent. The methods described here have been implemented and run on a parallel cluster.

6.1.1   Generalised Nelder-Mead algorithm

The *global distribution* of the algorithm is attained by running a Monte-Carlo simulation, generating the quasi-random Sobol' points in an N-dimensional rectangular region $U \in \mathbb{R}^N$ specified by the user, where N is the number of parameters being varied during the inversion.[1] At each of these points, a simplex in multi-dimensions *local descent* algorithm is initialised. The points are fed to the parallel processors for function computations, and the results are then gathered back at the main node, as described in Section 6.2. We define a number, call it L, which is being compared to the value of the function $f(\mathbf{x}_j)$, as soon as it is passed back to the main node. Here $\mathbf{x}_j$ is the j'th quasi-random point in U.

---

[1]Quasi-random points provide a higher degree of uniformity than random points, leading to a better low-dimensional performance, while exhibiting a milder dimensional dependence than random points. However, the performance of all quasi-random sequences deteriorates in higher dimensions. For large N (such as $N = 200$) a random sequence would be preferable to the quasi-random Sobol' sequence.

If at any time $f(\mathbf{x}_j) > L$, the local descent is terminated and a new quasi-random point $\mathbf{x}_{j+1}$ is passed to that processor for function computation. Otherwise, $f(\mathbf{x}_j) \leq L$, and the point is deemed perspective.

We define a *cycle* as a set of local descents run from quasi-random starting points in a chosen region. Once a cycle is complete, the search region is refined. Thus, the iterative search may be summarised in general terms as the following algorithm (starting from $i = 0$):

1) A search space U_i is chosen, comprising an N-dimensional rectangular region;

2) A critical misfit value L_i is chosen; also the maximum number of local minima K_i in the i'th cycle and some local descent parameters;

3) A point $\mathbf{x}$ is chosen at random (pseudo-random) in the N-dimensional rectangular region, and the value $f(\mathbf{x})$ computed;

4) If the value is greater than the critical misfit in 2), go to 3). If not, continue from point $\mathbf{x}$, using it as an initial point of a local descent algorithm (e.g. Nelder-Mead simplex). Once the local descent completes, check whether the required number K_i of local minima in the cycle has been achieved. If so, go to 5); eise, go back to 3).

5) Once the cycle is complete, define the new search domain, typically a subspace $U_{i+1} \subset U_i$ of the previous N-dimensional rectangular region, and the new critical misfit L_{i+1}, based on the "best" points found: a) the minimum values found, b) the best local minima found. Then go to 3).

Each local descent algorithm stops if and only if it meets one of the four stopping criteria:

1) *tolerance*: the user-defined accuracy is achieved, so that several consecutive iterations return close enough values that we can assert that we have found a local minimum;

2) *number of iterations*: the user-defined maximum number of local descent iterations is achieved;

3) $f(\mathbf{x}) > L_i$: a large value of the misfit is encountered, as described above;

4) the cycle is being restarted, and all processes stopped.

When it quits for either of these reasons, a new quasi-random point is sent to the processor for computations. The algorithm keeps track of the local minima obtained by each simplex algorithm, and the minimum value found is constantly updated. The algorithm runs until we are satisfied with the best local minima found.

Unfortunately, this algorithm becomes impractical if the dimension and size of the initial search space are too large. Our experience has shown, that ~ 20000 function evaluations are generally required to reduce the search domain by approximately a factor of ten (depending on the function and the initial parameters) in a 12-dimensional space. However, there have been improvements/bug fixes which would likely improve the performance of non-gradient search, if these tests were repeated.

6.1.2 Non-linear conjugate gradient algorithm

The conjugate gradient algorithm in multi-dimensions is probably the most widely used gradient-based global optimisation technique. Based on a standard set of principles, it however has a great number of variants. Here, we only describe the particular variant of the non-linear conjugate gradient algorithm that has been used in this work, which may be specified as the non-linear Polak-Ribière conjugate gradients with the line search based on a Secant method. Our analysis here is pretty standard and follows mostly *Press et al.* [1992] and *Shewchuk* [1994], although the particular algorithm described here is not encountered in its entirety in these references.

General strategy for non-linear conjugate gradients

Consider the task of minimising the function $f : \mathbb{R}^N \to \mathbb{R}$. Choose a starting solution $x_0 \in \mathbb{R}^N$, and the initial search direction $h_0 = -\nabla f(x_0)$. Also, set $g_0 = -\nabla f(x_0)$. Starting from $i = 0$, perform the following steps:

1) Proceed from x_i along the direction h_i to a *local minimum* $x_{i+1} = x_i + \alpha_i h_i$, for some α_i so that at the local minimum $h_i \cdot \nabla f(x_{i+1}) = 0$ (i.e. gradient is perpendicular to the descent direction).

2) Compute $g_{i+1} = -\nabla f(x_{i+1})$.

3) Compute either of the expressions

$$
\begin{cases}
\beta_i = \dfrac{g_{i+1} \cdot g_{i+1}}{g_i \cdot g_i}, & \text{Fletcher-Reeves;} \\[3ex]
\beta_i = \dfrac{(g_{i+1} - g_i) \cdot g_{i+1}}{g_i \cdot g_i}, & \text{Polak-Ribière.}
\end{cases}
$$

4) Compute $h_{i+1} = g_{i+1} + \beta_i h_i$.

Repeat this procedure N times, where N is the number of dimensions of the domain of function f. If f were a perfect quadratic form, N iterations would be enough to successfully minimise it. However, for functions encountered in real life, this procedure usually needs to be restarted and repeated several times.

Possible stopping criteria for CG include, among many others,

$$
\begin{cases}
\|f'(\mathbf{x}_i)\| \le \varepsilon \|f'(\mathbf{x}_0)\|, & ([\textit{Shewchuk}, 1994] \\
2\,|f(\mathbf{x}_i) - f(\mathbf{x}_{i-1})| \le \varepsilon(|f(\mathbf{x}_i)| + |f(\mathbf{x}_{i-1})| + \delta), & [\textit{Press et al.}, 1992]
\end{cases}
$$

where both ε and δ are small and positive.

Line search using Secant method

The process of minimising the function $f(\mathbf{x}_i + \alpha\mathbf{d}_i)$ along a given direction $\mathbf{d}_i$ is known as *line search*. As a general rule, the value of α_i such that

$$
\alpha_i = \arg\min_\alpha f(\mathbf{x}_i + \alpha\mathbf{d}_i) \tag{6.1}
$$

is found by ensuring that the gradient at the local minimum $\mathbf{x}_i + \alpha_i\mathbf{d}_i$ is orthogonal to the search direction $\mathbf{d}_i$. This can be easily seen:

$$
\frac{\partial}{\partial\alpha}f(\mathbf{x}_i + \alpha\mathbf{d}_i) = [f'(\mathbf{x}_i + \alpha\mathbf{d}_i)]^T\mathbf{d}_i. \tag{6.2}
$$

Thus, searching for the extrema of this expression with respect to α corresponds to finding zeros of the expression $[f'(\mathbf{x}_i + \alpha\mathbf{d}_i)]^T\mathbf{d}_i$.

The two most commonly used methods for performing line search are the Newton-Raphson method and the Secant method. Here, we describe these methods as they appear

in *Shewchuk* [1994]. Both of these iterative methods rely on the truncated Taylor series expansion of $f(\mathbf{x}_i + \alpha \mathbf{d}_i)$ around $\alpha = 0$:

$$f(\mathbf{x}_i + \alpha \mathbf{d}_i) \approx f(\mathbf{x}_i) + \alpha \left[\frac{d}{d\alpha} f(\mathbf{x}_i + \alpha \mathbf{d}_i) \right]_{\alpha=0} + \frac{\alpha^2}{2} \left[\frac{d^2}{d\alpha^2} f(\mathbf{x}_i + \alpha \mathbf{d}_i) \right]_{\alpha=0}. \tag{6.3}$$

Both methods also require that f be twice continuously differentiable. Differentiating 6.2 further with respect to α gives

$$\frac{\partial^2}{\partial \alpha^2} f(\mathbf{x}_i + \alpha \mathbf{d}_i) = \mathbf{d}_i^T [f''(\mathbf{x}_i + \alpha \mathbf{d}_i)] \mathbf{d}_i. \tag{6.4}$$

Here, f'' is now a matrix. Setting $\alpha = 0$ in both expressions 6.2 and 6.4 and differentiating 6.3 with respect to α yields:

$$\frac{\partial}{\partial \alpha} f(\mathbf{x}_i + \alpha \mathbf{d}_i) \approx [f'(\mathbf{x}_i)]^T \mathbf{d}_i + \alpha \mathbf{d}_i^T [f''(\mathbf{x}_i)] \mathbf{d}_i. \tag{6.5}$$

The Newton-Raphson method finds zeros of this expression by setting

$$\alpha = -\frac{[f'(\mathbf{x}_i)]^T \mathbf{d}_i}{\mathbf{d}_i^T [f''(\mathbf{x}_i)] \mathbf{d}_i}.$$

To perform an exact line search using Newton-Raphson method, this expression needs to be evaluated and $\mathbf{x}_i \Leftarrow \mathbf{x}_i + \alpha \mathbf{d}_i$ needs to be iteratively updated until α is sufficiently close to zero. Thus, we see that the method requires a repeated evaluation (or approximation) of the Hessian f''.

Therefore, although the Newton-Raphson method has a better convergence rate, in cases where a repeated evaluation of f'' is to be avoided due to high computational costs, the Secant method is preferable.

Instead of computing or approximating f'', the Secant method approximates the second derivative of $f(\mathbf{x}_i + \alpha \mathbf{d}_i)$ by evaluating the first derivative at two distinct points $\alpha = 0$ and $\alpha = \sigma$, where σ is an arbitrary small non-zero number:

$$\frac{\partial^2}{\partial \alpha^2} f(\mathbf{x}_i + \alpha \mathbf{d}_i) \approx \frac{[f'(\mathbf{x}_i + \sigma \mathbf{d}_i)]^T \mathbf{d}_i - [f'(\mathbf{x}_i)]^T \mathbf{d}_i}{\sigma}. \tag{6.6}$$

Substituting this approximation into 6.3 yields a new expression for the derivative with respect to α:

$$\frac{\partial}{\partial \alpha} f(\mathbf{x}_i + \alpha \mathbf{d}_i) \approx [f'(\mathbf{x}_i)]^T \mathbf{d}_i + \frac{\alpha}{\sigma} \left([f'(\mathbf{x}_i + \sigma \mathbf{d}_i)]^T \mathbf{d}_i - [f'(\mathbf{x}_i)]^T \mathbf{d}_i \right). \tag{6.7}$$

Setting

$$\alpha = -\sigma \frac{[f'(\mathbf{x}_i)]^T \mathbf{d}_i}{[f'(\mathbf{x}_i + \sigma \mathbf{d}_i)]^T \mathbf{d}_i - [f'(\mathbf{x}_i)]^T \mathbf{d}_i} \tag{6.8}$$

minimises $f(\mathbf{x}_i + \alpha \mathbf{d}_i)$. Typically, we will choose an arbitrary σ on the first Secant method iteration; on subsequent iterations we will choose $\mathbf{x}_i + \alpha \mathbf{d}_i$ to be the value of $\mathbf{x}_i$ from the previous Secant method iteration. Equivalently, set $\sigma_{j+1} = -\alpha_j$.

Unless the Hessian of the function can be easily implemented and efficiently evaluated, the Secant method is in practice preferable to the Newton-Raphson method; however, convergence of Secant method depends a great deal on the choice of $\sigma = \sigma_0$.

Non-linear conjugate gradients with Secant and Polak-Ribière

Given a function f, a means to evaluate f', a starting value $x_0 \in \mathbb{R}^N$, a maximum number of CG iterations i_{max}, a CG error tolerance $\varepsilon < 1$, a maximum number of Secant method iterations j_{max}, a Secant method error tolerance $\epsilon < 1$, and an initial Secant method step parameter σ_0, we propose the following algorithm to minimise f.

$i \Leftarrow 0$

$k \Leftarrow 0$

$x \Leftarrow x_0$

Evaluate $f'(x)$

$r \Leftarrow -f'(x)$

$s \Leftarrow r$

$d \Leftarrow r$

$\delta_{new} \Leftarrow r^T d$

$\delta_0 \Leftarrow \delta_{new}$

While $i < i_{max}$ and $\delta_{new} > \varepsilon^2 \delta_0$ do

    $j \Leftarrow 0$

    $\sigma_d \Leftarrow d^T d$

    $\alpha \Leftarrow -\sigma_0$

    $\eta \Leftarrow [f'(x)]^T d$

    Evaluate $f'(x + \sigma_0 d)$

    $\eta_{prev} \Leftarrow [f'(x + \sigma_0 d)]^T d$

    Do

        $\alpha \Leftarrow \alpha \dfrac{\eta}{\eta_{prev} - \eta}$

        $\eta_{prev} \Leftarrow \eta$

        $x \Leftarrow x + \alpha d$

        Evaluate $f'(x)$

        $\eta \Leftarrow [f'(x)]^T d$

        $j \Leftarrow j + 1$

    while $j < j_{max}$ and $\alpha^2 \delta_d > \epsilon^2$

    $r \Leftarrow -f'(x)$

    $\delta_{old} \Leftarrow \delta_{new}$

    $\delta_{mid} \Leftarrow r^T s$

    $\delta_{new} \Leftarrow r^T r$

    $\beta \Leftarrow \dfrac{\delta_{new} - \delta_{mid}}{\delta_{old}}$

    $k \Leftarrow k + 1$

    If $k = N$ or $\beta \leq 0$

        $d \Leftarrow r$

        $k \Leftarrow 0$

    else

        $d \Leftarrow r + \beta d$

    $s \Leftarrow r$

    $i \Leftarrow i + 1$

Just before evaluating α in the inner loop of CG method, in the beginning of a Secant method iteration, we propose inserting additional checks that have proved useful in practice to deal with extremely small and extremely large values of α. For $\eta_{tol} < 1$ and $\eta_{frac} < 1$, insert the following lines:

> If $|\eta - \eta_{prev}| < \eta_{tol}$
>   Exit CG
> If $|\eta/\eta_{prev}| < \eta_{frac}$
> $\sigma_0 \Leftarrow k\sigma_0$
>   Restart CG

We see that the maximum of $j_{max} + 1$ gradient evaluations is required per iteration of the CG algorithm. Under normal circumstances (unless the conditions above are satisfied) the algorithm terminates when the maximum number of iterations i_{max} is exceeded, or when $||f'(\mathbf{x}_i)|| \le \varepsilon||f'(\mathbf{x}_0)||$. A fast inexact line search may be accomplished by using a small j_{max}. The algorithm is very sensitive to the Secant step starting parameter σ_0, and it is often necessary to adjust that parameter manually or during the algorithm iterations to achieve convergence. The outer CG iterations are effectively restarted (by setting $d \Leftarrow r$) whenever the Polak-Ribière parameter β is negative. It is all restarted every N iterations, where N is the dimension of the domain of function f. This is required to improve convergence for small N.

6.1.3 Steepest descent with relaxation

The basic steepest descent algorithm has the form

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \alpha_i \nabla f(\mathbf{x}_i), \tag{6.9}$$

where ∇f is the gradient of f and

$$\alpha_i = \arg\min_\alpha f(\mathbf{x}_i - \alpha \nabla f(\mathbf{x}_i)). \tag{6.10}$$

Here, we follow the algorithm developed by *Pronzato et al.* [2000] and *Pronzato et al.* [2001] as part of their research on applications of dynamical systems in optimisation. They have shown that an introduction of a relaxation coefficient γ, with $0 < \gamma < 1$, in the steepest descent algorithm totally changes its behaviour. The algorithm then becomes

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \gamma \alpha_i \nabla f(\mathbf{x}_i), \tag{6.11}$$

with α_i given by the expression (6.10).

This method has been given the name of *Renormalised Steepest Descent*, or *Steepest Descent with Relaxation*. It has been shown by the authors that although the process exhibits a chaotic behaviour in the N-dimensional space, the asymptotic rate of convergence for large values of the relaxation coefficient $\gamma < 1$ is typically significantly better than the worst-case rate of the steepest descent algorithm.

Then, the algorithm could be described as follows. Choose a starting solution $\mathbf{x}_0$, and the initial search direction $\mathbf{h}_0 = -\nabla f(\mathbf{x}_0)$. Set $\gamma = 1 - \delta$, $\delta > 0$ small. A good choice is the value of $\gamma \approx 0.99$. Starting from $i = 0$, perform the following steps:

1) Proceed from $\mathbf{x}_i$ along the direction $\mathbf{h}_i$ to a *local minimum* $\mathbf{x}_{i+1} = \mathbf{x}_i + \gamma \alpha_i \mathbf{h}_i$, for some α_i such that $\mathbf{h}_i \cdot \nabla f(\mathbf{x}_i + \alpha_i \mathbf{h}_i) = 0$.

2) Set $\mathbf{h}_{i+1} = -\nabla f(\mathbf{x}_{i+1})$. Repeat step 1).

6.2 Parallelization and the inverse solver

As we have already discussed, the inverse problem that we are interested in is computationally expensive. There are two conceptual ways to take advantage of the efficient parallel clusters available in the modern world. One is to parallelize the forward solver over frequencies and to employ it by a sequential inverse solver. The other is to parallelize the inverse solver itself, by implementing a parallel optimisation routine. Both options are being implemented in the inverse solver code that was developed for this project; both have their advantages and disadvantages. An important consideration is that parallelization over frequencies ruins the speedup that is otherwise obtained by using the solution computed for each of the previous frequencies as a starting point for the next frequency iterations. While parallelization over frequencies may be a useful tool if a few forward solutions need to be fast evaluated, parallelization of the inverse solver, if implemented efficiently, is the optimal computational solution for the inverse search. However, such a scheme requires careful synchronization of the computations to minimise the waiting times and hence the processor-time expenses. This is not always easy to accomplish.

The current implementation of the inverse solver may be run both in serial and in parallel. The parallelization has been implemented in such a way that any independent computations may be run in parallel. This has been accomplished as follows. First, n initial points are specified, for which the function F needs to be computed. This is done so that the number of points is just exceeding the number of available processors. These points form a queue, and the head point in the queue is sent for computation as soon as a processor becomes available. After a certain number of independently computed function values is received, the queue status is set to "locked", so that no further points are sent for computation until all processors are freed. Then, the function values are sorted and a certain number of minimal values $F(x_1), \ldots F(x_n)$ is chosen. The new search domain
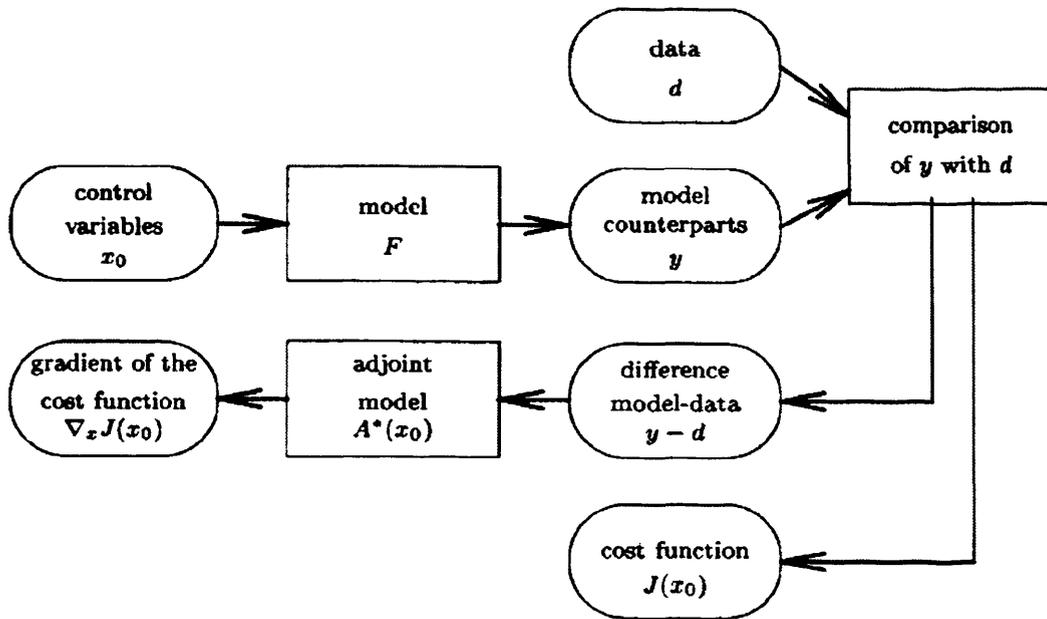
Figure 6.1: Schematic of a single run of the adjoint forward solver. Figure taken from *Giering and Kaminski* [1998].

is defined as the minimum subspace of the initial domain, that contains all the optimal points $x_1, \ldots x_n$. The procedure is repeated until the search domain is sufficiently reduced or until the optimal function value is satisfactorily small. This is essentially a variant of the domain decomposition approach [e.g. *Meyer*, 1990].

This parallel scheme is very flexible, however a gradient-based search can only be efficiently parallelized in terms of running several such descents with different starting parameters simultaneously on different processes. This may be beneficial, but it does not improve the convergence time of a single descent algorithm. An efficient parallelization does, however, improve convergence times of non-gradient based algorithms (e.g. Section 6.1.1) almost linearly with the number of processors used.

The implementation of the adjoint forward solver (Section 5.5) has allowed us to efficiently compute the derivative of the penalty functional (Section 4.2), also known as the *cost function*, thus making it practical to use in conjunction with a gradient-based inverse solver. A single computation of the cost function and its gradient by the adjoint forward solver closely follows the scheme on Figure 6.1 [*Giering and Kaminski*, 1998]. Once a vector of *control variables* (or model parameters) x_0 is proposed by the inverse solver, it is passed on to the forward solver (*model F*) to compute the *model counterparts*, or, equivalently, the responses $y = F(x_0)$. They are compared to the data via a cost function that is passed back to the inverse solver and is used to decide on the next step of the search (an update to x_0). Simultaneously, the vector of residuals between the data and the responses may be used in the adjoint gradient computations, at the same computational expense as the computation of the cost functional itself. In a gradient-based search, the gradient of the cost function is needed along with the cost function itself to decide on the new value of x_0; and the results of the computation may not be used until the whole cycle is complete. Additionally, each step of a conjugate gradient algorithm depends on the values of the cost function and its gradient at a previous point. Therefore, the search can only be performed sequentially.

To conclude, a gradient-based algorithm such as the non-linear conjugate gradients is only parallelizable over the starting points. This is not a huge benefit, since we have seen in practice that the end result of the search does not significantly depend on the starting point, but rather on the regularisation parameters, and especially on the prior model (since regularisation essentially penalises any departures from the prior). In fact, we have established a strong dependence of the inverse result on the prior model, so that depending on the prior model we always get slightly (or, in fact, significantly) different models that fit the data, given that parametrization is flexible enough. Please refer to

Chapters 7 and 8 for further discussions and illustrations of this feature of the inverse search. We believe that the parallel inverse solver could be readily modified to parallelize over prior models rather than starting (initial) models, which could be a highly valuable feature (Section 9.2).

Chapter 7

Non-linear global 3-D Inversion – I. Computational experiments

The aim of this chapter is to demonstrate the convergence and resolution of the inverse method on a series of simple checkerboard and other synthetic examples. We also illustrate the issues arising in this ill-posed inverse problem due to sparseness, noisiness and irregular spatial distribution of the data as well as uncertainties in the prior model. We also discuss the way the large magnitude inhomogeneities due to the highly conducting oceans affect the magnetic fields at the surface of the Earth. We conclude with a set of guidelines for inverting field data.

7.1   Design of the computational experiments

In order to test the performance of the inverse solver, we perform a set of computational experiments with models known a priori. To do this, we generate synthetic data for a set of idealized conductivity distributions.

The response of the forward solver is uniquely determined by the electrical conductivity distribution on the grid, under the source field assumptions described in Chapter 2 and the

frequency domain assumptions required to reduce Maxwell's equations to the quasi-static approximation (Chapter 3). Running the forward solver with a specified input model and a chosen computational grid for each of the frequencies of interest we generate the exact[1] set of responses of the Earth, that would have been observed at geomagnetic observatories if

1) the specified model was infinitesimally close to the true structure of the Earth,

2) the source field assumptions were perfectly satisfied, and

3) the measurements were error free.

These values are referred to as the computed responses of the Earth. In order to achieve resemblance to the field data, random errors are introduced and added to these values according to the procedure described in Section 7.1.1. Thus, we obtain a synthetic data set with approximate responses to a particular electrical conductivity structure in the Earth. We can now employ an inverse solution (Chapter 4) to see if and how well we can reconstruct this model by inverting the synthetic data. This may or may not be possible, depending on the complexity of the model, the distribution of the synthetic observatory locations and the errors, and on the degree of ill-posedness and resolution in the inverse problem. Often the same (noisy) data set can be fit by a number of different models apart from the one used to construct the data in the first place. This issue illustrates the inherent ill-posedness of the problem in the sense of uniqueness. It is also possible, especially if the introduced data errors are small, that the inversion turns out unstable and never converges to the original model, nor in fact to any model that fits the data. This illustrates the ill-posedness in the sense of stability.

---

[1]Clearly, these values are only "exact" in the sense that they contain no measurement errors. In fact, any computational solution contains both truncation and discretization errors. Here we do not discuss in detail the accuracy of our forward solver, since it has already been attested by the analysis performed in the U&S paper, although we shall briefly describe the accuracy tests performed for the Jacobian. However, a discussion that explains our choice of the computational grid is provided.

In this Section we discuss the methodology developed by the author to combat or at least to quantify these issues in the inversion of EM field data. We also describe the full setup of the various synthetic inversion tests covered in this Chapter.

7.1.1 The structure of synthetic data

Having generated the computed responses, we introduce synthetic errors into the data set following pretty much the same procedure as that described in Section 4.3.1. Let us denote the computed responses by $\tilde{d}_j$ (in this case they represent the "true" values observed for the model of interest).

For this data set, we introduce synthetic errors that constitute a percentage deviation about the calculated response, weighted by the modulus of the response. That is, we estimate the errors by evaluating the expression

$$\delta_j = \kappa \; |\tilde{d}_j|, \tag{7.1}$$

where κ is a constant. For example, if we wanted to generate a data set with 5% errors, we would set $\kappa = 0.05$. Then, δ_j is the estimated error for the measurement taken at the j'th observatory at a given frequency. We generate the synthetic data according to Equation 4.10,

$$\mathrm{Re} \; (d_j) \;\; = \;\; \mathrm{Re} \; (\tilde{d}_j) + \epsilon_j^{(1)},$$
$$\mathrm{Im} \; (d_j) \;\; = \;\; \mathrm{Im} \; (\tilde{d}_j) + \epsilon_j^{(2)},$$

where $\epsilon_j^{(1,2)} \sim N(0, \delta_j^2)$. This distribution is achieved by drawing pseudo-randomly from $N(0,1)$ and multiplying by δ_j. The final data set used in synthetic inversions contains d_j and δ_j, $j = 1, \ldots M_\omega$ for every frequency ω.

In this text, we shall not systematically study the rate of convergence for different values of κ. However, the general rule is that if very small errors $< 3\%$ are used a gradient-based inversion is very unstable. Since we divide the residuals by the errors in the penalty functional (e.g. Equation 4.13), a model that achieves a reasonably small misfit if data errors are large would result in unreasonably large values of both the misfit and the derivative if the same data set had only a little bit of noise. Gradients become uncontrollably large and a lot of manual adjustment of the initial parameters is required to achieve convergence. In the field data, the measurement errors are normally $\gg 5\%$ (see Section 2.4). In this text, all synthetic experiments have been performed with 5% estimated errors, unless a different value is explicitly specified.

7.1.2   The choice of the computational grid

As we have already discussed earlier in this Section, discretization errors also exist and may be significant. In general, a coarser discretization introduces greater errors. That is, the EM fields computed using the true electrical resistivity distribution discretized to the grid of $2° \times 2°$ should in general be closer to the truth than those computed on the $20° \times 20°$ grid (assuming the number of vertical layers is the same in both cases). However, a single forward computation on a $2° \times 2°$ grid requires a large amount of computational resources. We have found that with the currently available computing power a practical grid spacing to use in the inversion is $5 - 10°$ in both longitude and latitude. Thus, the grid mesh used in the inversions in this work varies between $[L, M, N] = [36, 18, 43]$ and $[L, M, N] = [72, 36, 63]$, where $[L, M, N]$ indicates the number of longitudinal, latitudinal and radial grid divisions, respectively (including the air layers). We normally use the grid $[36, 18, 43]$ for the majority of computations, and then refine the obtained models using a finer grid, if necessary. The grids we have been using are all regular in longitude and

latitude, although this is by no means a requirement of the forward solver; rather the specifics of this work only.

In this Chapter, the [36, 18, 43] grid (that we shall in the future refer to as $10° \times 10°$) has been employed both to generate the synthetic data and for the inversions, unless specified otherwise. There are both advantages and disadvantages to inverting with the same grid as that used to generate the synthetic data. This way we eliminate the effect of the discretization errors from the synthetic inversions. This is an advantage since we do not currently possess an accurate way to quantify these errors, and thus it would be difficult to separate the effect of the discretization from the performance of the inversion itself, that we are primarily interested in. On the other hand, a fair test of the inversion of field data would have been to generate synthetic data with an extremely fine computational grid (since the true parameter distribution in the Earth is best modelled on a finely discretized computational domain), while possibly performing the inversion using a coarser grid. However, in order to use this technique to test the performance of the inversion a detailed analysis of discretization errors needs to be performed (see Section 9.2). This has not been the central focus of this work, although for simple interior structure it has been shown by the comparison with 1-D layered Weidelt corrected (Section 1.3.5) and other computational solutions [*Uyeshima and Schultz*, 2000] that grids of $5 − 10°$ in both longitude and latitude allow an accurate forward solution. Any discretization errors introduced by the coarseness of these grids would be significantly smaller than the noise in the field data, and we propose that they may be neglected. We leave to future work a detailed analysis of the effect of discretization on the error in the response of the Earth.

7.1.3   Accounting for the near-surface inhomogeneities

For the synthetic inversions described in Sections 7.2 and 7.3, and for the inversion of real data described in Chapter 8, we consider both of the following approaches to the near-surface conductance. We either ignore the existence of the S-distribution in the inversion by not including it in the model (thus effectively inverting for it in the case of real data), or we include it in the model as a constant layer extending down to 12.65 km at the same resolution as the grid. In Figure 7.1 we show the S-distribution for a set of discretizations. We see that for a $10° \times 10°$ grid this distribution, although very approximate, still resembles the true oceans versus land distribution.

The effect of the near-surface conductance distribution on the responses of the Earth, and in particular on the magnetic fields at coastal observatories has been studied and carefully described in a number of works including *Fainberg et al.* [1990b], *Fainberg et al.* [1990c], *Weiss and Everett* [1998], *Kuvshinov et al.* [1999], *Kuvshinov et al.* [2002], *Olsen and Kuvshinov* [2004] and others. Please refer to Section 1.6 for an overview. Therefore, we shall not analyse it here in depth. We shall however describe this effect in relation to our models (see for example Section 7.2.5). Further considerations regarding the effect of the near-surface conductance on the inverse solutions can be found later in the text of this and the following chapters.

7.1.4   Non-linear conjugate gradient optimisation

The inversions in this work have been performed using the Polak-Ribière non-linear conjugate gradient optimisation with Secant method used for line search. We have describe this scheme in Section 6.1.2. Here we only note, that this method requires computations of both the penalty functional (cost function) and the derivative at each step.

Parameters of this algorithm include:
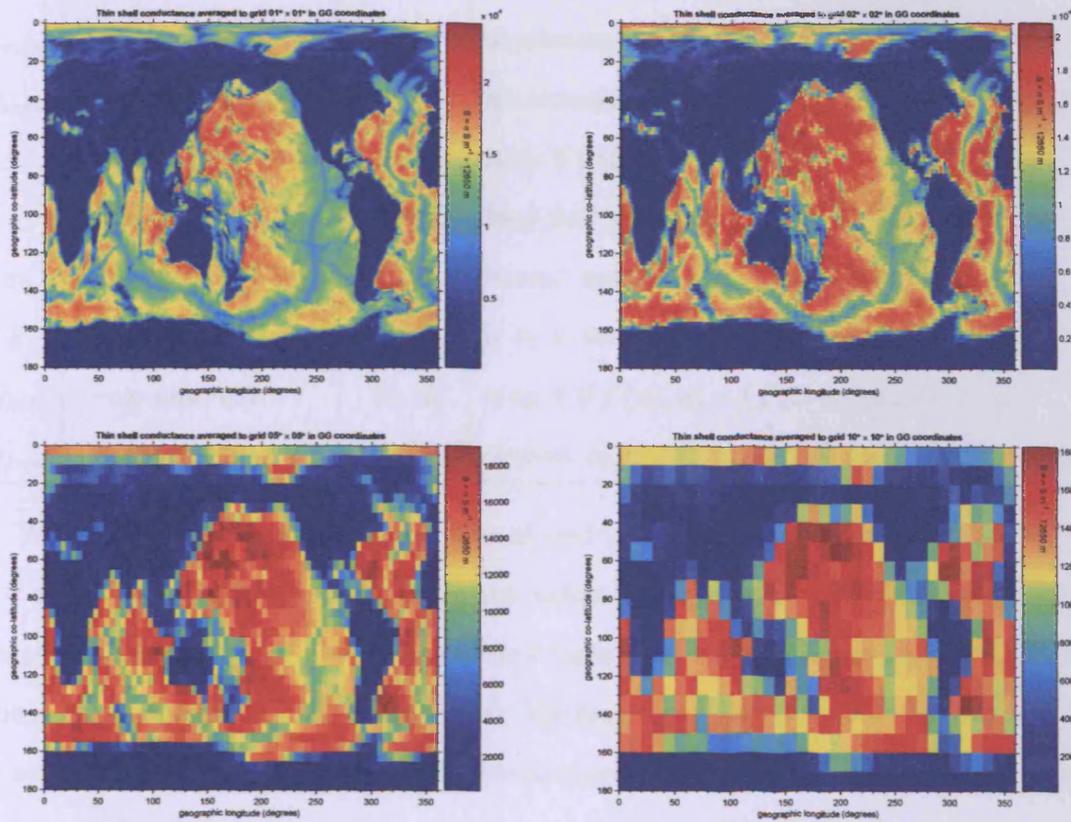
Figure 7.1: Near-surface conductance of the uppermost 12.65 km for fine to coarser discretizations in geographic coordinates. Figure created from thin shell conductance data obtained courtesy to Alexei Kuvshinov.

	Option	Default	Purpose
x_0	-G2 --fpts	random	starting value
$f(x)$	--prog,--test-f	none	function to minimise
	--script	none	parametrization script (if f is a program)
	--config	none	configuration script (if f is a program)
i_{max}	--cg-itermax	80	maximum number of CG iterations
j_{max}	--cg-sec-itermax	3	maximum number of Secant method iterations
ε	--cg-tol	1e-2	stop if $\|r(i)\| < \varepsilon^2 \|r(0)\|$
ϵ	--cg-sec-tol	1e-5	stop line search if $\|\alpha d\| < \epsilon$
σ_0	--cg-sigma0	0.01	Secant method step parameter
k	--cg-adjust	1e-2	if σ_0 is too large, we correct it by setting $\sigma_0 \Leftarrow k\sigma_0$
η_{tol}	--cg-eta-tol	1e-10	stop if $|(f'(x), d) - (f'(x + \sigma_0 d), d)| < \eta_{tol}$
η_{frac}	--cg-eta-frac	1e-8	correct σ_0 if $|(f'(x), d)/(f'(x + \sigma_0 d), d)| < \eta_{frac}$

We can also specify the total number of evaluations of the function $f(\cdot)$. In our case, $f(\cdot)$ is the forward solver, and $f(x)$ is the value of the least squares penalty functional, where x is the vector of model parameters (spherical harmonic expansion coefficients). Then, x_0 corresponds to the prior model. We have used the parameter setting $\varepsilon = 10^{-2}$ as a stopping criterion allowing to test whether performance is still acceptable or a restart with an updated damping parameter is required (Section 7.1.6). For the last stages of an optimisation routine this parameter has to be set to a smaller value to ensure accuracy of the solution. Another important parameter is j_{max}, one of the two stopping criteria for the line search (the other is ϵ). Normally, only a few iterations of line search are required. We have observed that for our practical inversions setting $j_{max} = 1$ allowed faster convergence in terms of the number of function and derivative evaluations than the larger values of j_{max} in most cases, which suggests that may be a different algorithm (such as a Relaxed steepest descent algorithm, Section 6.1.3) could be more efficient for our purposes.

Stability of the algorithm is governed by σ_0, the parameter of the first step of the line search. Effectively, σ_0 depends on the second derivative of the cost function, $f''(x_0)$ which we do not know. We do not currently have the capability to compute the matrix of second derivatives (Hessian) explicitly; thus we instead approximate it in the line search as described in Section 6.1.2. Therefore, an ad-hoc criterion needs to be employed for choosing the value of σ_0. In general, if we expect the second derivative to be large, it is a good idea to choose a small value of σ_0 to avoid encountering an unpredictable value of $f'(x_0 + \sigma_0 d)$, where $d = -f'(x_0)$. A large σ_0 will in many cases result in an unstable algorithm. In other cases, setting σ_0 to a larger value actually improves convergence by preventing the search from wandering around local minima.

7.1.5 Achieving stability of the inversion

Stability is ensured to a great extent by a good choice of σ_0 (Section 7.1.4).

Another important stabiliser is the regularisation. This includes, first of all, the model norm term in the penalty functional with the appropriate damping parameter μ. Updating the damping parameter once in a while as described in Section 7.1.6 allows convergence while stabilising the search. Apart from the damping parameter, two other parameters also constitute regularisation as they determine the model covariance matrix C_m. These are: α, that penalises higher degree and order terms, and β, which prevents large conductivity jumps from layer to layer (see Section 4.4.3 for details). The default values of these parameters are $\alpha = 0$ and $\beta = 1$, and these values do not have an effect on the inversion. Any departure from these values ($\alpha > 0$, $0 < \beta < 1$) has an additional stabilising effect on the inversion by "smoothing" the model $\hat{m}$.

7.1.6   Dependence of convergence on the damping parameter μ

We minimise the regularised least squares penalty functional $R(\hat{m})$ (Equation 4.8) with respect to the model $\hat{m}$. This penalty functional has two terms corresponding to the data misfit and the model norm, respectively. They could be summarised as follows:

$$R(\hat{m}) = R_{d,m_0}(\hat{m}) + \mu R_{\hat{m}}, \tag{7.2}$$

where R_{d,m_0} depends on the data, data and model covariance matrices and the prior model, and $R_{\hat{m}}$ is the model norm. By minimising the penalty functional $R(\hat{m})$, we are trying to reach a target (small enough) misfit $R_{d,m_0}(\hat{m})$, but at the same time to find the model that minimises the model norm subject to this misfit.

In this setting, μ is called the *damping parameter*, and regulates the reciprocal weighting of the two terms. If we minimise (7.2) with μ fixed, we will not generally achieve the target misfit. Normally, given a particular value of μ, the inversion would increase the model norm gradually while it is significantly decreasing the data misfit, so that the total still decreases. It stops making progress when an increase in the model norm no longer results in a total improvement; the misfit achieved at this point can not be further decreased unless the value of μ is corrected. For larger values of μ the achievable misfit is larger; but convergence for a larger μ is generally faster and more stable. Thus we pursue a strategy of starting with a large μ, the reducing μ (by a factor of 10) when the misfit stops decreasing, until we either achieve the target (generally, 1 for the normalised misfit) or stop making progress. If the procedure converges to a misfit higher than the target, it usually means that it is not possible to find a model that fits the data well enough, either because the data does not satisfy the distributional or source assumptions well, or because of the inherent restrictions of model parametrization. It can also diverge, if the

value of μ at some stage of the procedure is too small to successfully regularise.

7.1.7 Choosing the prior model

The inversion penalises for any departures from the prior model m_0 (see Sections 4.3 and 7.1.6). In general, a prior model should include as few features as possible, unless the aim of the inversion is to specifically test the compatibility of a particular feature with the data. In that case, the feature of interest would be included in the prior model to see, whether it is possible to fit the data assuming that specific feature is present in the model.

In this work, we use 1-D profiles for the prior models, to avoid introducing laterally non-uniform structure not required by the data. These prior models are based on the 1-D global and regional profiles found in the literature (please refer to Section 1.4 for a literature review of the 1-D models of the mantle).

We tried variants of these prior models to assess sensitivity of our inversions to the prior. We have found that both the rate of convergence and the inverse solution depend a great deal on the choice of the prior model m_0. In general, data can constrain the 1-D structure, making the inversion reasonably insensitive to the prior assumptions. However, we use only fairly long-period data, so information about typical upper mantle conductivity, which can be obtained from interpretation of shorter period data, is not in the data set. We can take advantage of what have been learnt from analysis of long period MT, daily variations, etc. to help set a more reasonable 1-D prior for the upper mantle.

7.1.8   Choosing the initial model

We start our computations from the prior model m_0. We see from (Equation 4.7)

$m = C_m^{1/2}\hat{m} + m_0$ that this corresponds to setting the perturbation to the prior $\hat{m} = 0$ in

the log-resistivity $\hat{m}$–space (see Section 4.3.2, specifically Equation 4.6 for the definition

of $\hat{m}$). Then, the initial data misfit is calculated as $R_0 = R_{d,m_0}(0)$ and the initial model

norm (Equation 4.8) is zero.

Choosing $\hat{m} = 0$ as the initial model in general makes the search more efficient, since

the prior is often (though not necessarily) the best-fitting 1-D model. However, we have

seen in practice that the choice of the initial model does not significantly affect the inverse

solution. This argument is justified by the computational experiments we have performed,

which have shown that for all synthetic models considered and for a number of reason-

able initial models, setting a large damping parameter μ (of the order of 10 or greater)

ensures convergence to the 1-D prior in the first place. Therefore, starting from a 1-D

parametrization that we believe to be reasonable eliminates the steps that would have

been taken by the inverse solver to converge to this 1-D profile if a different initial model

were used. Having said that, in some cases starting from some model other than the prior

with a small damping parameter may in fact ensure faster convergence.

7.1.9   Notational and plotting conventions

In the notation of the plots shown below, H_ϕ, H_θ and H_r are magnetic field values on

the sphere, pointing to magnetic East, magnetic South and downwards to the centre of

the Earth, respectively. This notation has already been defined and used in Chapters 3

and 5, and we employ it here for consistency.

All inverse problems discussed in this text have been parametrized by writing the surface Schmidt semi-normalised spherical harmonics expansions for $\log_{10} \rho$, where ρ is the electrical resistivity, just like described in Section 4.4. Thus the respective plots included in this text are logarithmic. However, instead of plotting the value of $\log_{10} \rho$, we normally plot the logarithmic electrical conductivity $\log_{10} \sigma = -\log_{10} \rho$. So that in general, warm colours (e.g. red) in the conductivity plots represent conductive anomalies, while cold (e.g. blue) colours correspond to resistive anomalies. Scales are provided for every plot.

Note also, that all the resistivity models described in this work have been parametrized in geomagnetic coordinates, while the plots presented here have been converted to geographic coordinates; hence some asymmetry appears in the plots even for the latitudinally symmetric synthetic models. Note also, that although the responses for different periods are plotted on different axes, the range is kept invariable so that the magnitudes of inhomogeneities may be compared between periods as well as between models. Similarly, when the inhomogeneities (rather than perturbation against a common background) in the model are plotted for different layers, the range is kept invariable between layers so that the size of the inhomogeneities may be compared at different depths in the Earth. Although all the inversions in this work have been performed using purely c response data sets (real and imaginary parts), we also plot the d responses for comparison. We shall discuss the possibilities of usage of d responses in a practical inversion further in this Chapter, as well as in Chapter 9.

For each of the examples considered below, we specify whether the surface conductance layer has been included in the model (Section 7.1.3). If so, we parametrize the Earth from the depth of 12.65 km (the lowermost boundary of the near-surface layer) down to the CMB (the exact depth of which varies in our models and is always specified explicitly).

Otherwise, we parametrize all of the mantle. Apart from the domain that is parametrized and varies during the inversion, the computational domain also includes 6–10 constant air layers of electrical conductivity 10^{-10} S m^{-1} (Section 3.3.1) and the infinitely conductive core. These assumptions are employed throughout this work.

7.2   Synthetic checkerboard inversion degree and order 3

We present in Figure 7.2 the synthetic model that has been used for this series of inversions. This has been generated by setting the two P_2^2 terms of the harmonic expansion in layer 2 of the three-layer model, corresponding to the depths $450 - 670$ km, to 0.6. These depths roughly define the transition zone in the true mantle. The synthetic 1-D model that has been perturbed by the degree 2 checkerboard in layer 2 is also shown (Figure 7.2).

For the purpose of the inversion, the $\log_{10} \rho$ (ρ standing for electrical resistivity) on the layer corresponding to the depth range $450 - 670$ km has been parametrized by the degree and order 3 surface spherical harmonic expansion, thus allowing a finer structure than that required by the true model. This corresponds to a total of 16 free parameters in layer 2. Additionally, there were two more free parameters representing the uniform background resistivities of layers 1 (down to 450 km) and 3 (670 $-$ 3000 km depths), respectively. These resistivity values were set to 10^3 and 10^{-1}Ohm $\cdot$ m in the synthetic model. The resistivity of the lowermost model layer $3000-3500$ km was set to 10^{-5}Ohm$\cdot$m to achieve approximate resemblance to the true Earth.

We should also note that this particular series of test inversions has been performed

450 - 670 km

Electrical conductivity anomaly, $\log_{10}(\sigma/\sigma_{prior})$
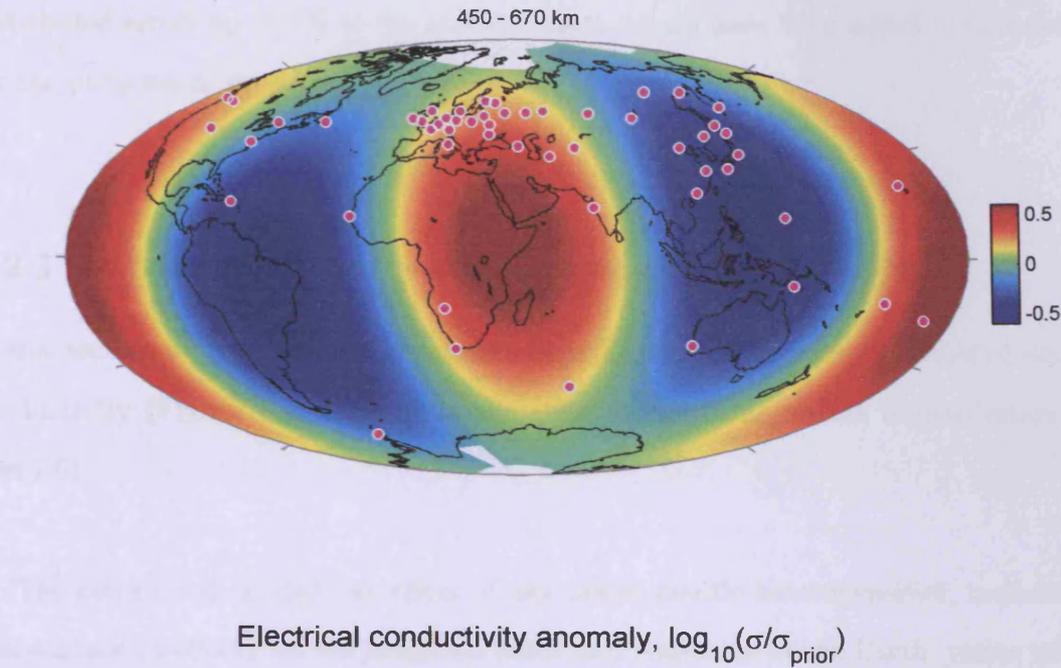
Figure 7.2: The synthetic model generated for the checkerboard degree and order 3 synthetic inversion tests. The model has been plotted in terms of the logarithmic perturbation around the prior uniform electrical conductivity $\sigma_{prior} = 1/\rho_{prior}$, which for this layer is 10^{-1} (this can be seen from the 1-D plot). Thus, the anomaly that is shown for the depths $450 - 670$ km corresponds to the values of electrical conductivity ranging from $10^{-1.5}$ to $10^{-0.5}$ S m^{-1}. The pink dots represent the locations of the *Fujii and Schultz* [2002] mid-latitude observatories.

using data for 5 periods only. These are: 80, 40, 20, 8 and 5.12 days. Random, normally distributed errors up to 3% of the absolute data values have been added to this data set for the purposes of inversion (see Section 7.1.1).

7.2.1   Analysis of the synthetic responses of the Earth

In this section we consider the responses of the Earth from the simple model of electrical conductivity (Figure 7.2), as well as the effect of the near-surface S-distribution (Section 1.6).

The general rule is that the effect of any upper-mantle heterogeneities, including the near-surface (S-effect), on the magnetic fields and responses of the Earth, varies with the frequency at which these responses are computed: for longer periods, this effect is less significant. However, at large periods ~ 80 days the S-effect is still visible: responses for the 1-D Earth when the S-map is part of the model still differ noticeably from the 1-D Earth responses without the near-surface. This effect is certainly above the 3% confidence level imposed for this experiment in both the magnetic fields and responses, for our period range. Here, we shall only discuss a set of model responses for a sample period of 8 days. We shall further discuss the effect of near-surface inhomogeneities at a few sample periods in Section 8.6.

On Figures 7.3, 7.4 and 7.5 we have plotted the synthetic H_ϕ, H_θ and H_r components of the magnetic fields at the Earth's surface produced as follows. Parts (a) of each figure correspond to the real and imaginary parts of the respective magnetic field components for a 1-D three-layer conductivity structure in the mantle, based on the prior on Figure 7.2. We see that for this 1-D model the real and imaginary parts of H_ϕ are pretty

much uniform (except for noise), while the real and imaginary parts of H_θ and H_r are laterally varying. Parts (b) of each figure show the magnetic fields obtained with the same 1-D mantle structure, with a map of near-surface conductance imposed on top. Both real and imaginary parts of H_ϕ are significantly affected. So are the imaginary parts of H_θ and H_r; they quite obviously resemble the sea water versus land distribution (sea water being in general more conductive). Parts (c) have been obtained by perturbing the 1-D structure by the simple checkerboard pattern (Figure 7.2) at the depths 450 − 670 km. Again, we see that the real parts of H_θ and H_r are totally or almost totally insensitive to the 3-D structure. Then, parts (d) have been obtained by combining the 1-D prior perturbed by the 3-D checkerboard with the near-surface conductance distribution on top of the mantle. In comparison with (b) and (c), this gives us an idea of the scale of the perturbation that we are considering here: this is a relatively mild heterogeneity in terms of the effect seen at the surface, compared to the effect of the surface conductance distribution, at the frequencies considered in this work.

Thus, if the near-surface distribution is included into the model, we are effectively inverting a totally different set of field ratios that no longer has a clear pattern of the underlying mantle structure even in this simple case. This is illustrate by the plots of the modulus and phase of c responses and d responses (as defined in Section 2.3) for the same set of models (please see Figures 7.6 and 7.7). We see that the phase of c responses does not contain any information about the heterogeneities (for this model at least), while the phase of d responses clearly tracks the boundaries of each conductive anomaly. These figures are included to demonstrate that there may be substantial additional information in d responses, if they are used along with the c responses in a practical inversion. However, **no d responses have been inverted in this study**. Whenever they are included in the figures, this is done for illustrative purposes only, rather than to imply that they have

1. Real
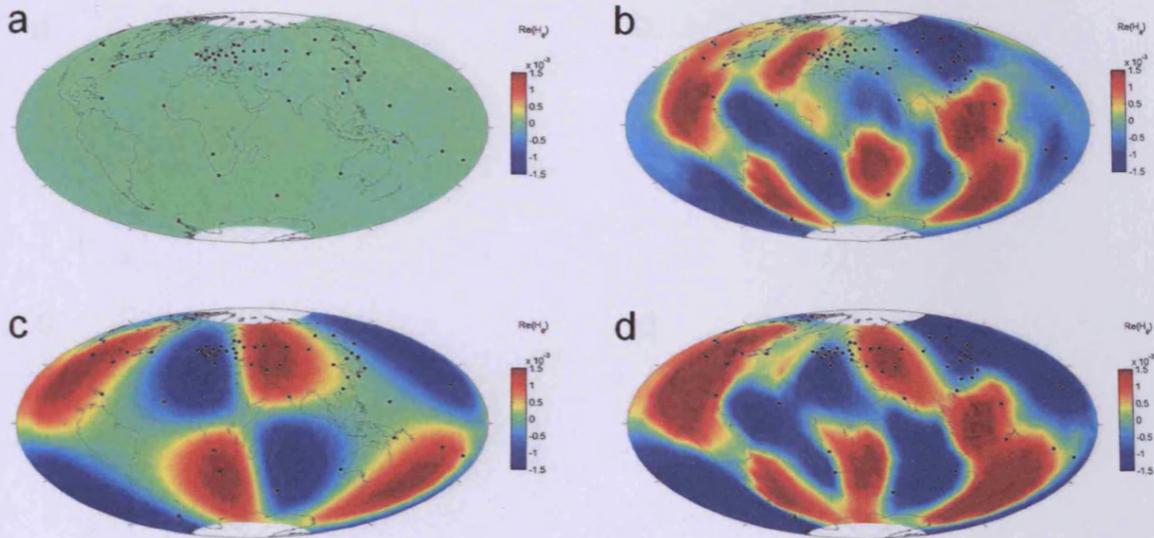
2. Imag
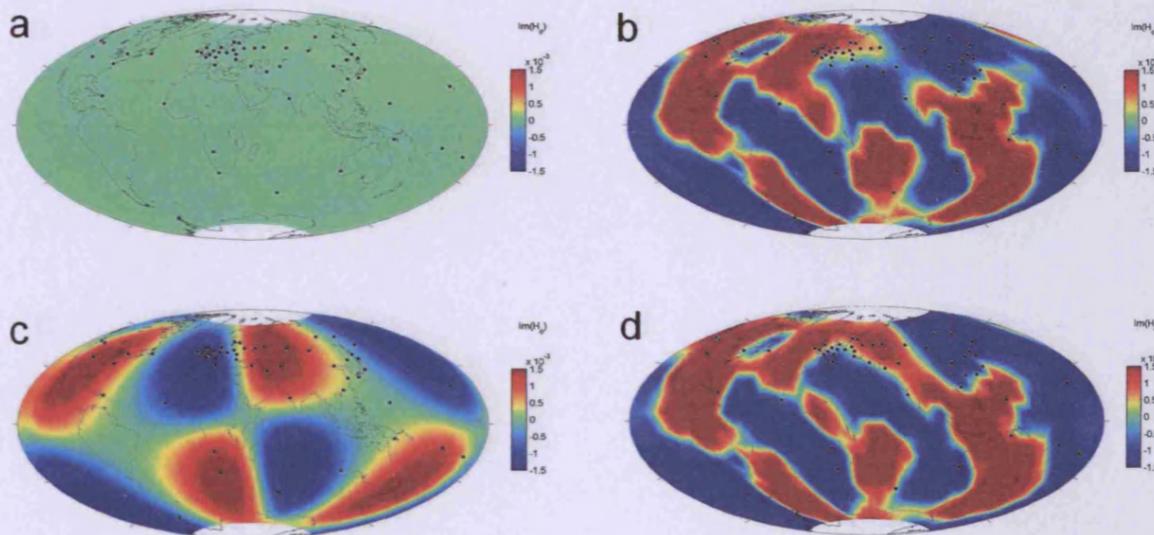
Figure 7.3: A set of real and imaginary parts of surface H_ϕ, corresponding to the synthetic checkerboard degree and order 3 model (see Figure 7.2) for a sample period of 8 days. Figure (**a**) is the response of the 1-D layered Earth with no 3-D heterogeneities. Figure (**b**) is the response of the same 1-D Earth with the near-surface conductance distribution imposed on top. Figure (**c**) is the response of the checkerboard perturbation at depths 450–670 km for an Earth without the near-surface heterogeneities. Figure (**d**) is the response of the Earth with both the checkerboard perturbation at depths 450–670 km and the near-surface heterogeneities.

1. Real
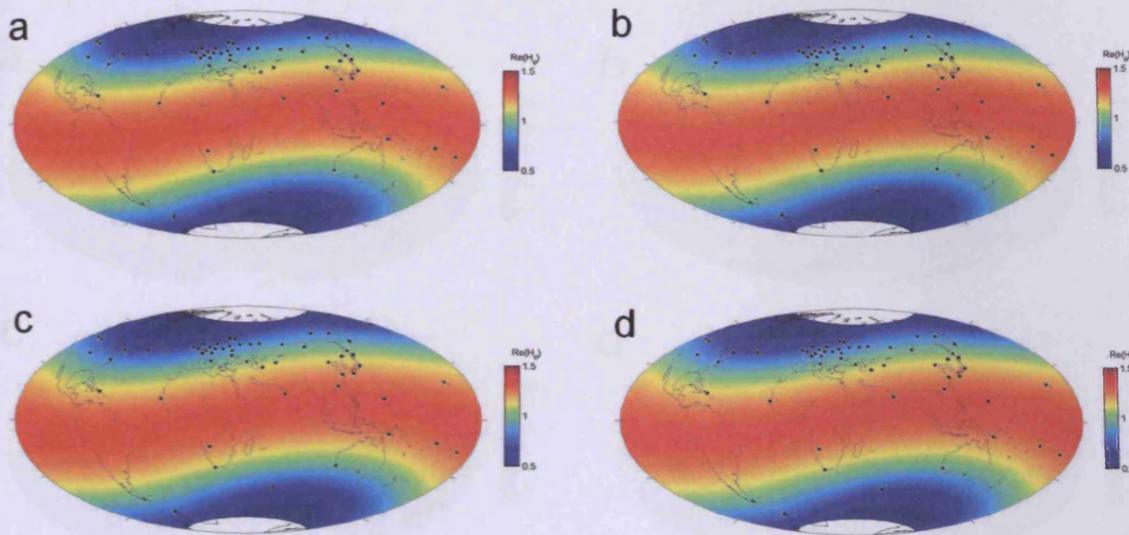
2. Imag
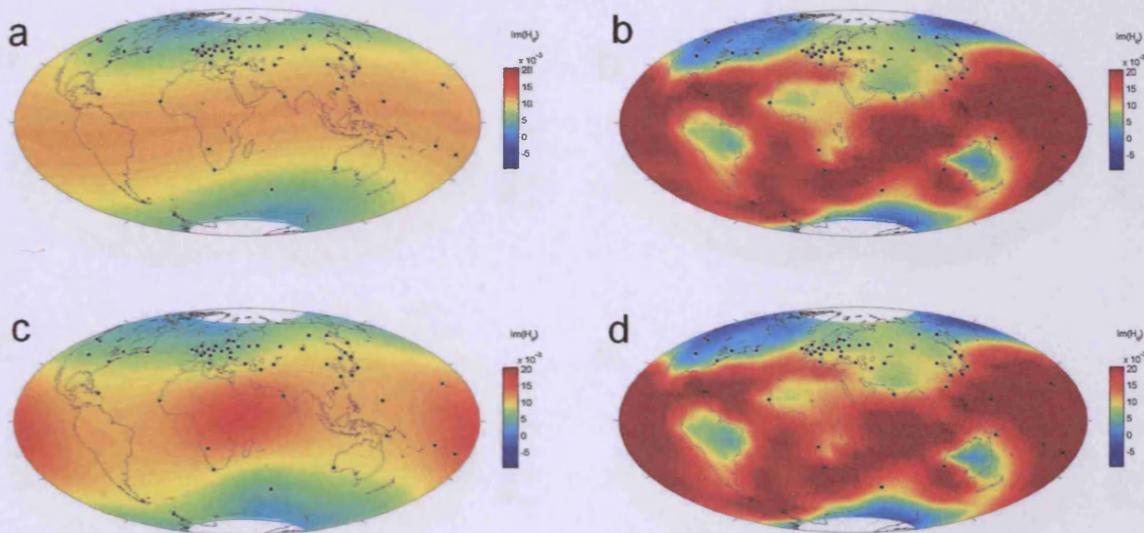
Figure 7.4: A set of real and imaginary parts of surface H_θ, corresponding to the synthetic checkerboard degree and order 3 model (see Figure 7.2) for a sample period of 8 days. Figure (a) is the response of the 1-D layered Earth with no 3-D heterogeneities. Figure (b) is the response of the same 1-D Earth with the near-surface conductance distribution imposed on top. Figure (c) is the response of the checkerboard perturbation at depths 450–670 km for an Earth without the near-surface heterogeneities. Figure (d) is the response of the Earth with both the checkerboard perturbation at depths 450–670 km and the near-surface heterogeneities.

1. Real
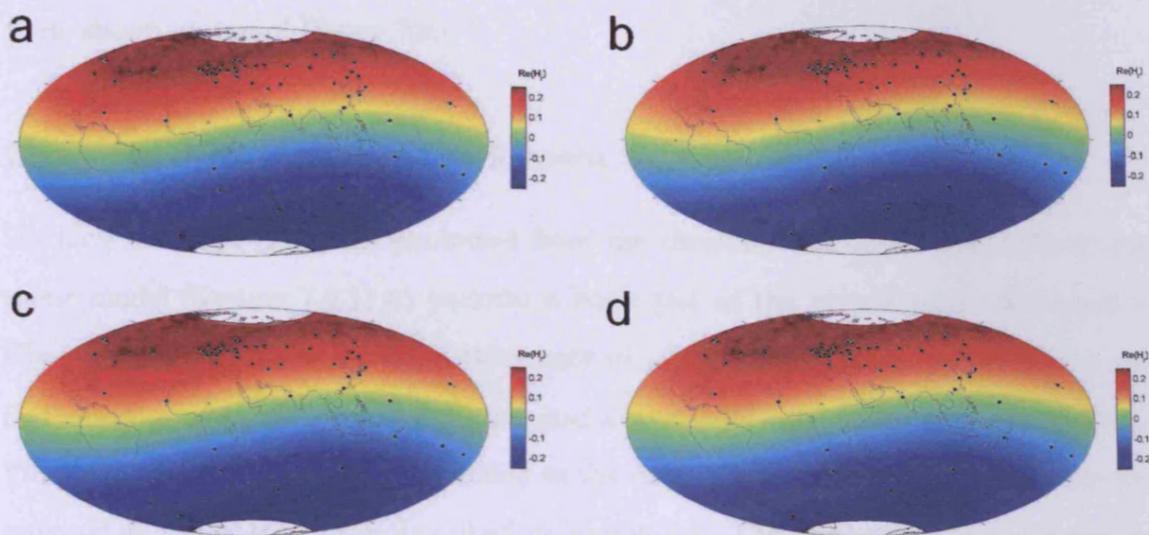
2. Imag
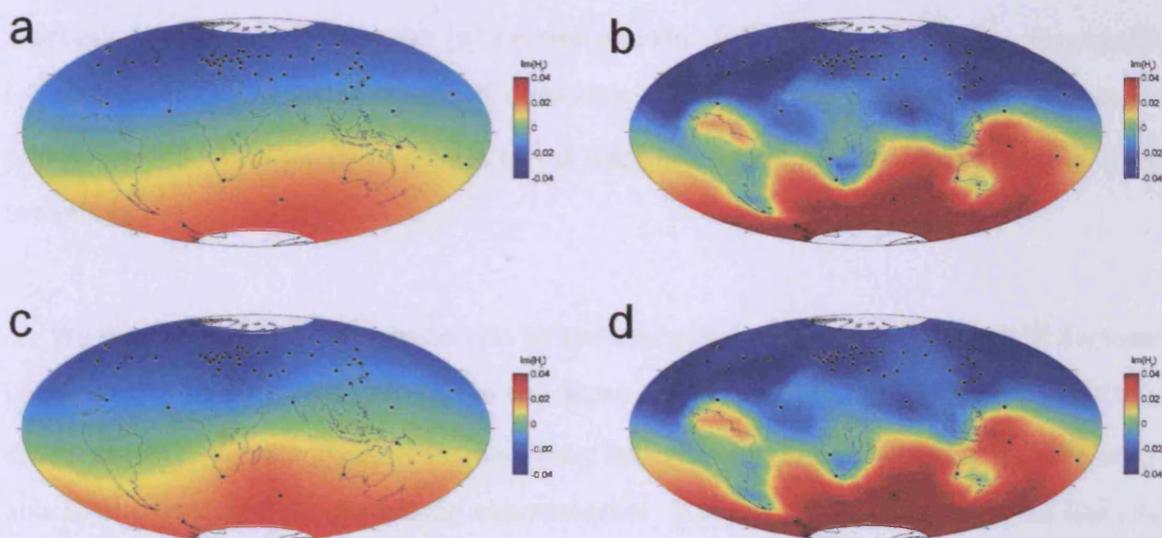
Figure 7.5: A set of real and imaginary parts of surface H_r, corresponding to the synthetic checkerboard degree and order 3 model (see Figure 7.2) for a sample period of 8 days. Figure (**a**) is the response of the 1-D layered Earth with no 3-D heterogeneities. Figure (**b**) is the response of the same 1-D Earth with the near-surface conductance distribution imposed on top. Figure (**c**) is the response of the checkerboard perturbation at depths 450–670 km for an Earth without the near-surface heterogeneities. Figure (**d**) is the response of the Earth with both the checkerboard perturbation at depths 450–670 km and the near-surface heterogeneities.

been used in the inversion. Joint inversions of c and d responses are a task for the future. Additionally, modulus and phase have also been presented for illustrative purpose: for the practical inversions we have used the real and imaginary parts of c responses, such as those shown on top of Figure 7.9.

7.2.2   Inversions with the known 1-D prior

We have used the responses generated from the checkerboard degree and order 3 synthetic model (Section 7.2.1) to perform a basic test of the inverse solver developed in Chapters 5–6. We have inverted three sets of synthetic responses, including the real and imaginary parts of c responses generated at three different observatory distributions. Pseudo-random errors of 3% were added to the data. For this simple test, no S-map was imposed on top of the model described by Figure 7.2. The prior model used for this set of inverse runs was exactly the true synthetic 1-D model. We present on Figure 7.9 a set of inverse models for these experiments. Column 1 (**b,c,d**) contains the heterogeneous layer in the three inverse models with the corresponding observatory distributions. The synthetic model is plotted on top (**a**) for comparison. Column 2 contains the corresponding real and imaginary components of c responses, compared to the synthetic c responses plotted on top. Column 3 illustrates the d responses, that have not been used in these inversions.

We see that the synthetic model can be reconstructed quite well, with a slight decrease in the quality of the model going from the dense regular distribution of observatories to a sparse regular set, with some skewness being introduced when information is only available in the irregular set of realistic observatories. Note, however, that this test has not been so trivial since in the parametrization all the coefficients up to the degree and order 3 had been allowed to vary freely, while the synthetic model has been originally generated

1. Modulus
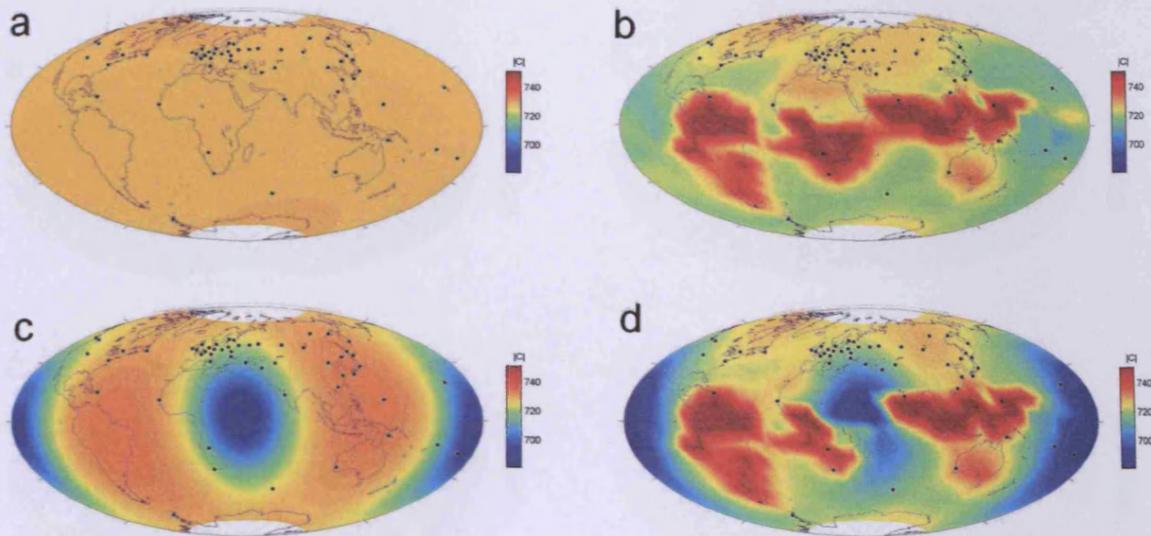
2. Phase
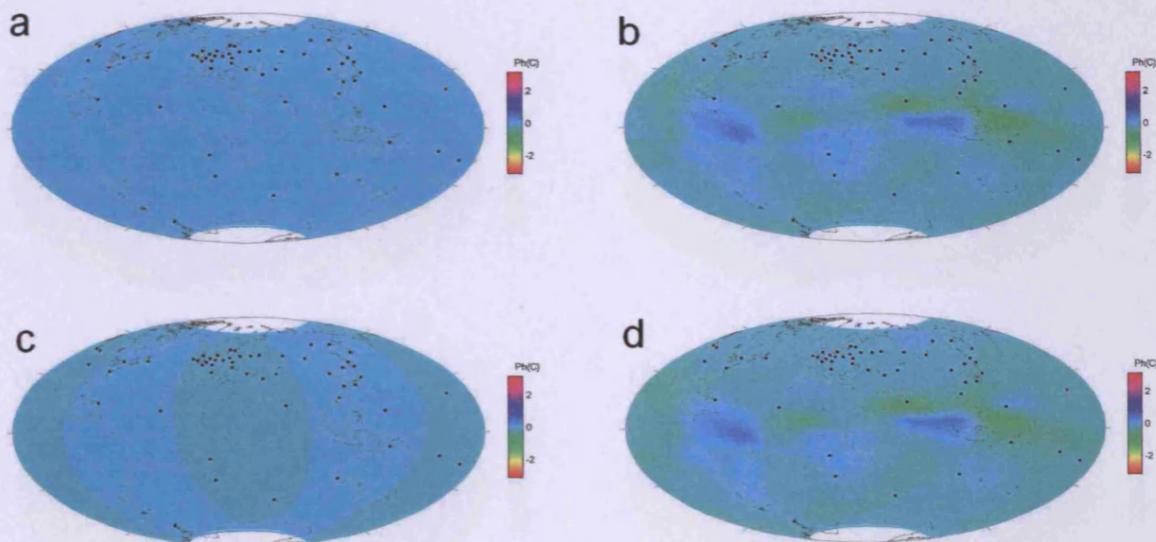
Figure 7.6: A set of absolute values and phases of the complex c responses, corresponding to the synthetic checkerboard degree and order 3 model (see Figure 7.2) for a sample period of 8 days. Figure (**a**) is the response of the 1-D layered Earth with no 3-D heterogeneities. Figure (**b**) is the response of the same 1-D Earth with the near-surface conductance distribution imposed on top. Figure (**c**) is the response of the checkerboard perturbation at depths 450–670 km for an Earth without the near-surface heterogeneities. Figure (**d**) is the response of the Earth with both the checkerboard perturbation at depths 450–670 km and the near-surface heterogeneities.

1. Modulus

a

b

c

d

2. Phase

a

b

c

d
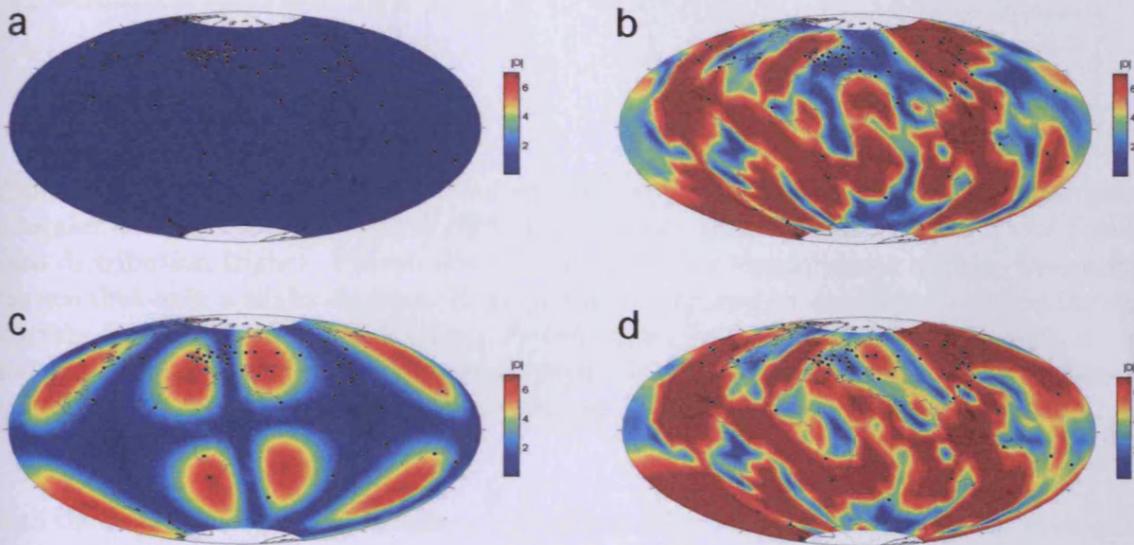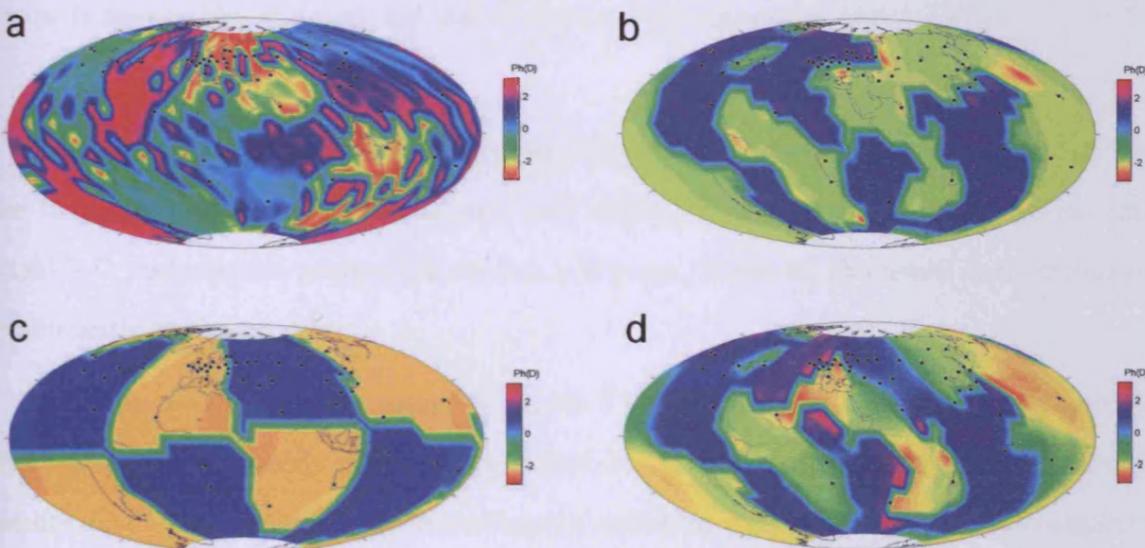
Figure 7.7: A set of absolute values and phases of the complex d responses, corresponding to the synthetic checkerboard degree and order 3 model (see Figure 7.2) for a sample period of 8 days. Figure (**a**) is the response of the 1-D layered Earth with no 3-D heterogeneities. Figure (**b**) is the response of the same 1-D Earth with the near-surface conductance distribution imposed on top. Figure (**c**) is the response of the checkerboard perturbation at depths 450–670 km for an Earth without the near-surface heterogeneities. Figure (**d**) is the response of the Earth with both the checkerboard perturbation at depths 450–670 km and the near-surface heterogeneities.
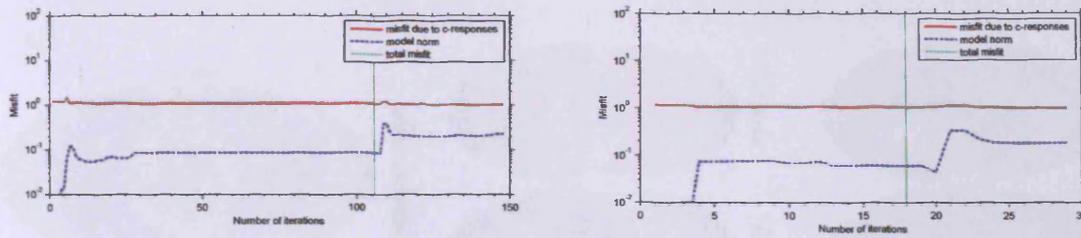
Figure 7.8: We present two examples of converging inversions with a true prior: one using a regular dense set of observatories (left) and another using the irregular *Fujii and Schultz* data distribution (right). Please refer to Figure 7.9 for the solutions of these inversions. We see that only a slight decrease in the misfit is required to converge to these solutions starting from a true 1-D conductivity distribution. In both cases, we start with $\mu = 1$ and gradually decrease the damping parameter to $\mu = 0.01$. The vertical lines (green) correspond to the points at which the damping parameter was updated.

with the degree and order 2 non-zero coefficients. No horizontal smoothing has been used; thus, higher degree structure could have been introduced than that required by the model. We do not see any evidence of this happening. In fact, given a large number of perfect data (Figure 7.9,b), the size of the conductive anomalies is well reconstructed and the shape is reasonable, although the size of the resistive anomalies is not perfect.

We also include two example plots of convergence with the true prior, Figure 7.8. We see that the data misfit decreases only very slightly to enable the inversion to see this slight 3-D perturbation around the correct 1-D prior. However, the model norm increases significantly in the process.

The equatorial band appearing on Figure 7.9 and on other c response plots later in this text is most probably noise, perhaps introduced by the noise in the data. This noise results in subtle noise in the inverse estimates, and then very subtle noise in the magnetic fields. It does not show up when the respective magnetic fields are plotted; nor is it seen in the d responses. However, since the fraction H_r/H_θ is multiplied by a large constant around the equator for the predicted c responses, any small variations due to noise are
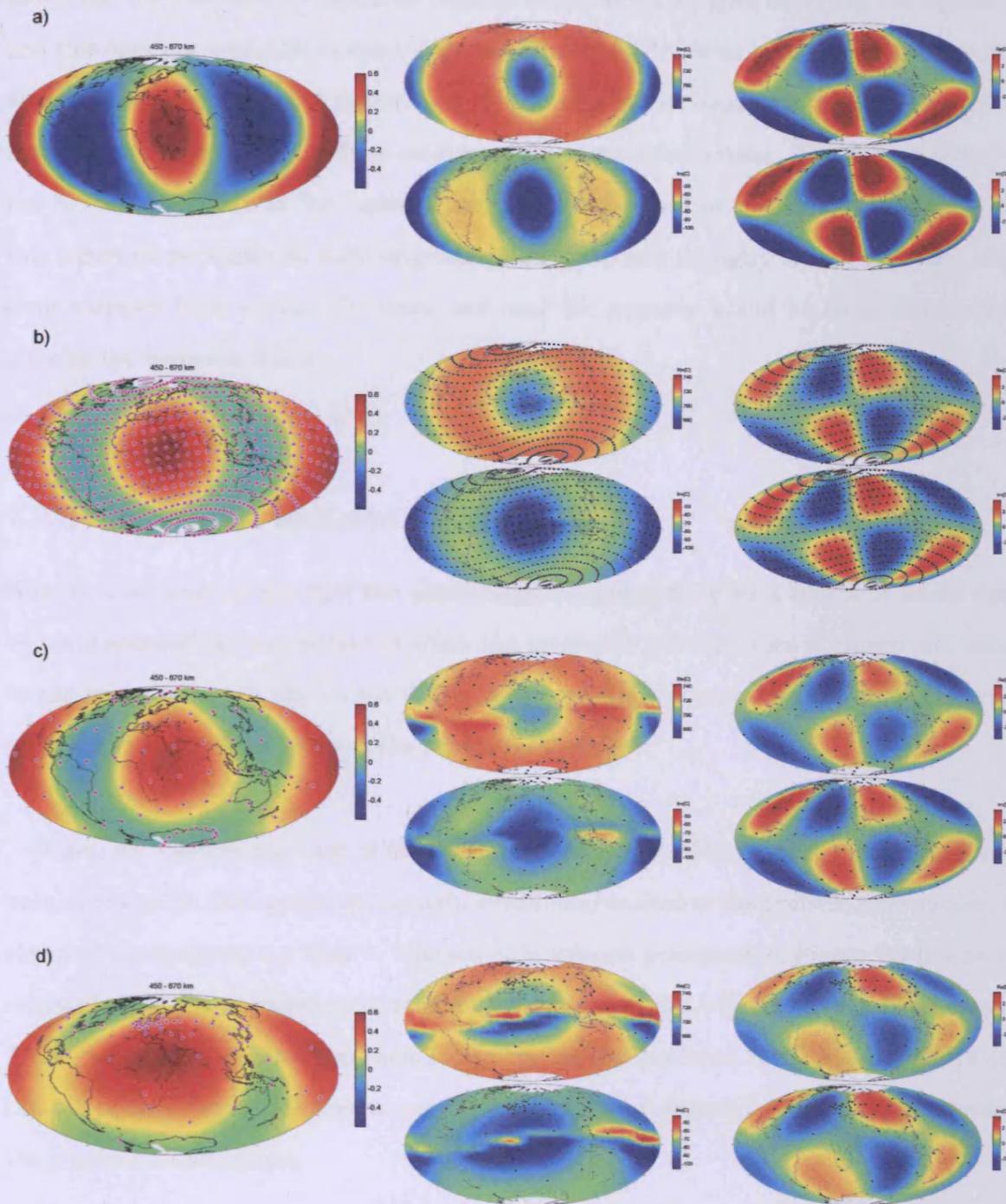
Figure 7.9: A set of inverse solutions for the checkerboard degree and order 3 synthetic inversion. Figure (**a**) shows the synthetic model (identical to Figure 7.2) and the corresponding real and imaginary parts of c and d responses for a sample period of 8 days. Figures (**b**), (**c**) and (**d**) show the inversion results with the data present at 648 regularly distributed observatories, at 72 regularly distributed observatories and at 54 *Fujii and Schultz* [2002] observatories, respectively.

blown up. The calculated c response remains finite only if H_r goes to zero at the equator, and this does not generally happen if the model is not 1-D (or at least exactly symmetric about the equator). Fitted c responses will look noisy at the equator if the model deviates from equatorial symmetry, unless we have data there to constrain. The d responses do not have a singularity at the equator, and hence they are free of the magnified noise in this region (d responses do have singularities at the poles, though). In any real data, the error variance both around the poles and near the equator would be large, due to the noise in the magnetic fields.

7.2.3   Averaged uniform Earth prior

Now, that we have shown that the checkerboard degree and order 3 synthetic model can be quite successfully reconstructed when the synthetic 1-D structure is known and used as the prior, we would like to see if this is still possible when we do not know the true prior, or when our knowledge of the prior is not perfect.

First, we consider the case of the uniform prior model. The synthetic 1-D model has been averaged (in the logarithmic space), which has resulted in the uniform prior model of electrical conductivity 0.1 S m^{-1}. The set of inversions presented on Figure 7.9 has been repeated with this averaged prior model used instead of the 1-D underlying distribution. Thus, the a priori information included in the search has been much less significant in this set of inversions. The inverse results are plotted on Figure 7.10 (a,b,c), along with the graphs of convergence.

We see that not only the resultant 3-D perturbation is still acceptably well recovered, but also the magnitude of the perturbation is much closer to that of the synthetic model.

The uniform prior had also been used as the starting model, and the initial data misfit had been significantly greater (~ 1000 instead of ~ 2). From the convergence graphs it can be seen, that the relative weighting of the data misfit and the model norm had been therefore different from our previous test, in favour of the data misfit (even though the damping parameter strategies had been identical). Thus, it was no longer possible to fit the data with a small model norm, which had allowed the search to continue enhancing the magnitude of the conductivity anomalies until the data could be fit. The anomalies had also become more localised. They are still deformed in the same way as seen in the previous example (Figure 7.9) where there isn't data to constrain them.

7.2.4 An example of an inversion with a skewed prior

We have also run a set of synthetic inversions with the prior intentionally skewed against the synthetic 1-D model. Each of the 1-D model parameters had been distorted in the logarithmic (base 10) space by 0.5, making the log conductivity values -2.5, -1.5 and 1.5, respectively, for the three mantle layers. Not all of these inversions converged. Divergence is likely if the prior model is significantly skewed from the truth, so that the model norm penalises for the departures from a wrong model, thus preventing the search from approaching the true solution. This problem may be solved by decreasing the damping parameter, but not if the errors in the data are too small to allow that much flexibility in the model: in that case the search becomes unstable. If we do not have a good estimate on the prior model it is generally wise to start with a small damping parameter; in that case, regularisation has to be performed by other means (see e.g. Section 4.3). Please refer to Figure 7.9, part (d) for an example of the converging (regular sparse observatory distribution) inversion and to Figure 7.11 for two examples of diverging inversions.
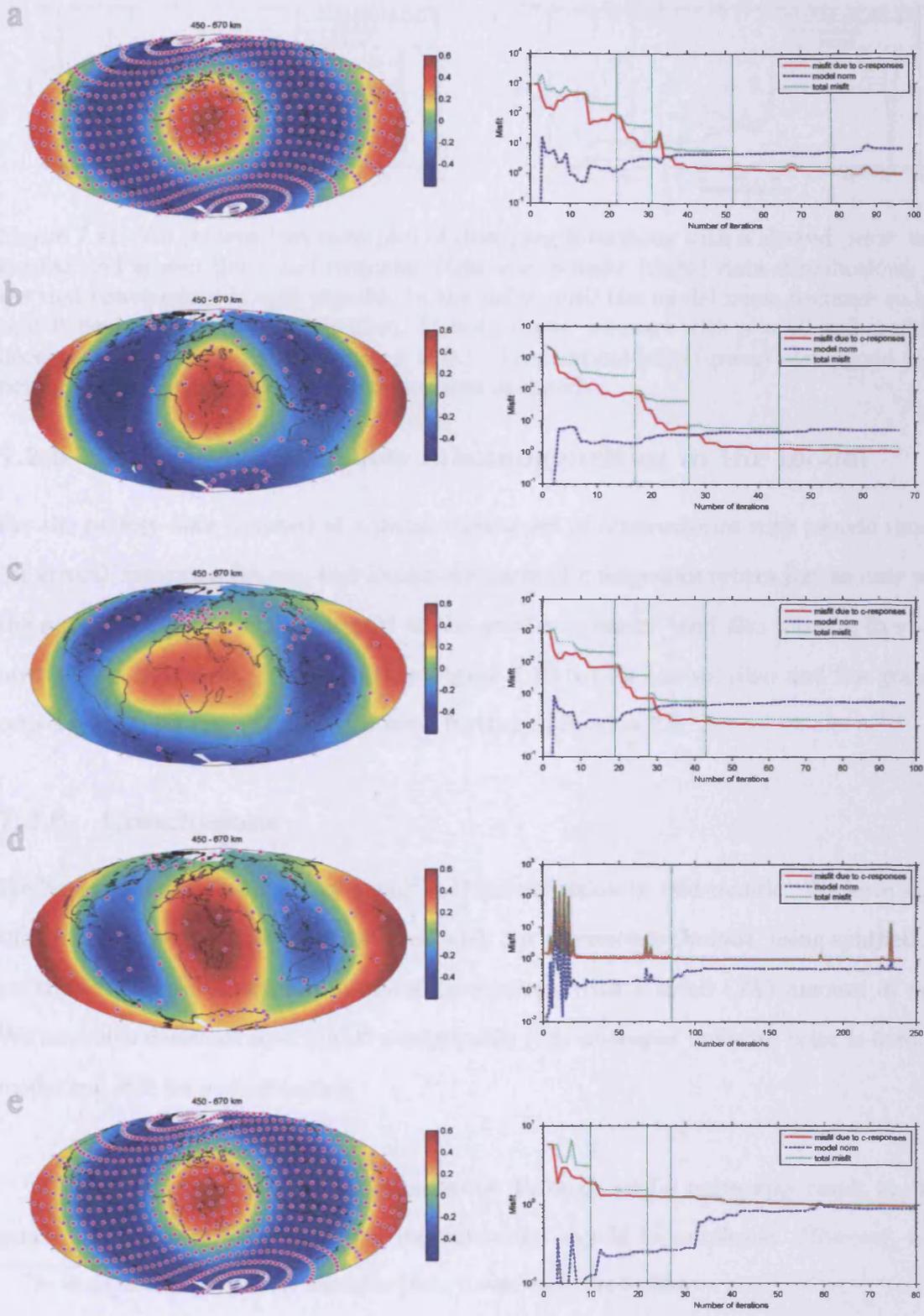
Figure 7.10: A set of five inverse solutions of the checkerboard degree and order 3 experiment: (a), (b) and (c) have been obtained with a uniform averaged prior model, using synthetic data at regular dense, regular sparse and irregular [*Fujii and Schultz*, 2002] sets of locations, respectively; (d) – synthetic 1-D prior with the S-map imposed on top both in the synthetic data and in the inversion (sparse data); (e) – the dense data set, no S-map distribution, with a skewed prior. We also present the corresponding rates of convergence.
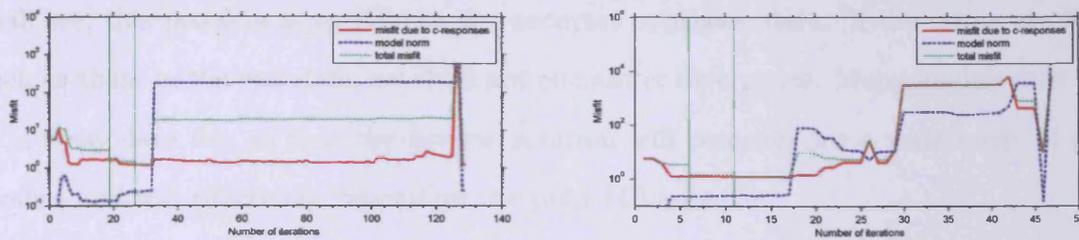
Figure 7.11: We present two examples of diverging inversions with a skewed prior: using regular and sparse (left) and irregular *Fujii and Schultz* (right) data distributions. We see that convergence is only possible to the point until the model norm becomes so large that it begins to drive the algorithm. In both cases, we start with $\mu = 10$ and gradually decrease the damping parameter to $\mu = 0.1$. The vertical lines (green) correspond to the points at which the damping parameter was updated.

7.2.5 Adding near-surface inhomogeneities to the model

For the perfect data (located at a dense regular set of observatories with pseudo-random 3% errors), inverting the real and imaginary parts of c responses proves just as easy when the near-surface conductance is part of the synthetic model (and also forms a fixed thin layer in the inversion). Please refer to Figure 7.10 (**e**) for the solution and the graph of convergence. We shall explore this issue further in Section 7.3.

7.2.6 Conclusions

We have considered a relatively small[2] 3-D perturbation in mid-mantle. We have shown that it can be successfully reconstructed with our inversion technique, using synthetically generated real and imaginary parts of c responses with a small (3%) amount of noise. We have also demonstrated that if a reasonable (e.g. averaged uniform) prior is used, the model can still be reconstructed.

If the prior is wildly wrong, regularisation through model norm may result in divergence. In that case, other types of regularisation should be employed. However, as we

---

[2]In terms of the effect on the magnetic fields, compared to the S-effect

shall see, this problem is specific to the accurate synthetic data. If the errors are large, such as those in the real data, we shall not encounter divergence. Many models exist that fit a noisy data set, so that the inverse solution will converge for a wide range of prior models and will effectively depend on the prior (Chapter 8).

Equatorial bands seen in estimated c response plots are errors magnified by the large factor near the equator. Note that in a real data set error variance near the equator would be large due to the noise in magnetic field estimates (H_r). Synthetic c response data are unrealistic in the sense that a constant error percentage is used, so the synthetic data in the equatorial region is unrealistically accurate.

7.3 Resolution testing using degree and order 8 checkerboard inversion

Here we describe a set of synthetic experiments designed to test the resolving power of our data at different depths in the mantle. We have divided the Earth into eight layers, and defined two different 1-D conductivity profiles on these layers: (a) to model the Earth with a relatively conductive upper mantle and (b) to model the resistive upper mantle (Figure 7.12). We have generated a checkerboard perturbation by setting the two coefficients corresponding to the degree 8 and order 6 spherical harmonics to 0.6 on each of the layers. Thus, identical perturbations have been imposed on each of the eight layers, for both of the two models (a) and (b). We have generated synthetic data for both models at a subset of 11 *Fujii and Schultz* frequencies, and pseudo-random 5% errors have been added to the synthetic c responses.

The synthetic data has been generated at several hypothetical observatory distribu-
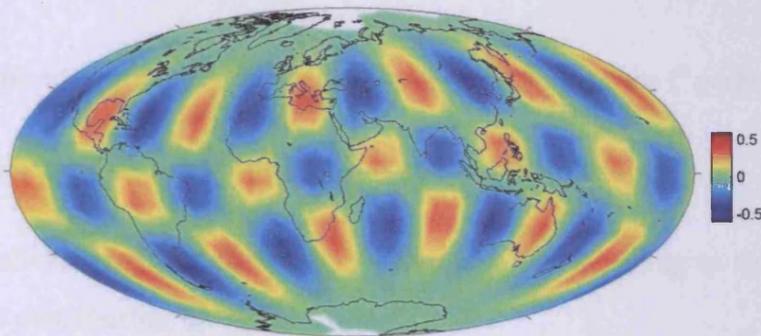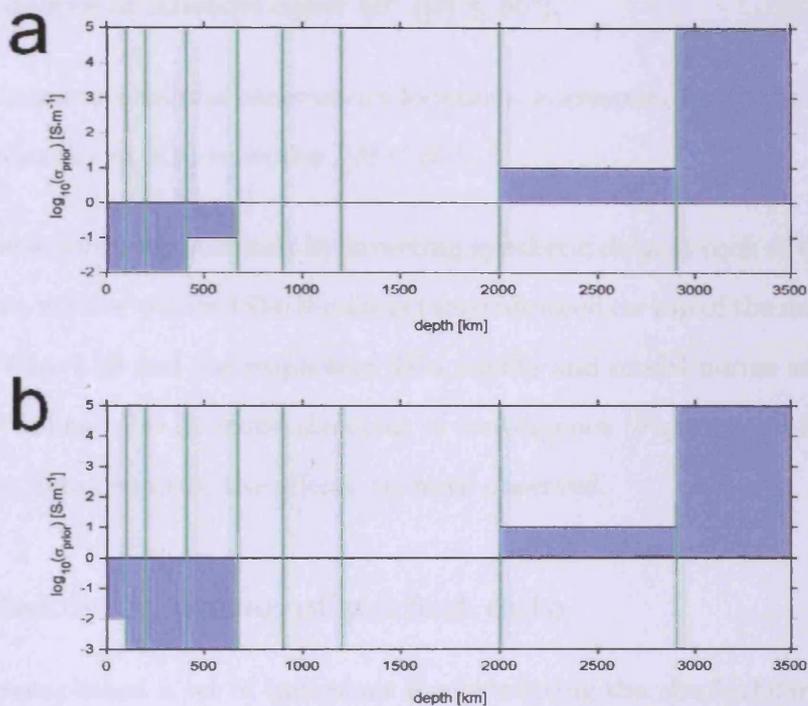
Figure 7.12: The synthetic checkerboard degree and order 8 model, generated by setting the non-zero $l = 8$ $m = 6$ terms on all layers to 0.6. The top figure shows the log conductivity perturbation around the prior model on all layers; (a) the prior model used for the "conductive upper mantle" experiment; (b) the prior model used for the "resistive upper mantle" experiment. The log conductivity on every layer is represented by the shaded area; and the layer boundaries by vertical lines. The layer boundaries for the two models are: 200 km, 410 km, 670 km, 900 km, 1200 km, 2000 km and 2900 km depths. The lowermost layer is uniform and fixed; it is added to enhance similarity with a realistic Earth. Note that the two prior models only differ in the depth range $200 - 670$ km.

tions:

dense: 648 regularly distributed locations at the Earth's surface ("perfect" data);

sparse: 72 regularly distributed locations at the Earth's surface;

iaga90: 103 realistic irregular observatory locations corresponding to the INTERMAG-NET 2006 distribution of observatories;

real65: 63 realistic irregular observatory locations, corresponding to the *Fujii and Schultz* observatories at latitudes below 65° ($|\theta| \leq 65°$);

real60: 54 realistic irregular observatory locations, corresponding to the *Fujii and Schultz* observatories at mid-latitudes ($|\theta| \leq 60°$).

The inverse solutions obtained by inverting synthetic data at each of these observatory distributions, with or without the S-conductance imposed on top of the mantle, are plotted on Figures 7.14–7.30 and the respective data misfits and model norms are summarised in Table 7.1. We have also included the rates of convergence (Figure 7.13) for reference. Let us discuss in detail some of the effects we have observed.

7.3.1   Resolving power of perfect data

We have accomplished a set of inversions reconstructing the checkerboard degree and order 8 synthetic model from the perfect (densely and regularly distributed) synthetic data. This has served as a more sophisticated test of our inverse procedure than the example described in Section 7.2. Additionally, this series of tests has been designed to test the resolution of the c response data at the frequency range of the real *Fujii and Schultz* [2002] data set.

Experiment	Prior	Obs. distr.	S-cond.	α	Initial data misfit	Initial model norm	$\mu = 10$ data misfit	$\mu = 10$ model norm	$\mu = 1$ data misfit	$\mu = 1$ model norm	$\mu = 0.1$ data misfit	$\mu = 0.1$ model norm	$\mu = 0.01$ data misfit	$\mu = 0.01$ model norm
chessboard8a	a	dense	no	0	35.03	0.00	4.52	0.36	1.39	1.20	1.04	2.10	–	–
chessboard8a	a	sparse	no	0	13.77	0.00	2.78	0.20	1.11	0.67	0.97	0.97	–	–
chessboard8a	a	real60	no	0	7.48	0.00	1.69	0.06	1.14	0.23	0.98	0.72	–	–
chessboard8a	a	real65	no	0	6.46	0.00	1.62	0.06	1.12	0.21	0.98	0.64	–	–
chessboard8a	a	iaga90	no	0	4.40	0.00	1.64	0.04	1.10	0.22	1.03	0.41	–	–
chessboard8a	a	dense	yes	0	26.21	0.00	4.26	0.31	1.40	1.12	1.05	2.00	–	–
chessboard8a	a	real60	yes	0	7.35	0.00	1.65	0.07	1.06	0.24	0.96	0.49	–	–
chessboard8	b	dense	no	0	24.07	0.00	–	–	1.40	0.72	0.01	1.48	–	–
chessboard8	b	dense	yes	0	19.86	0.00	–	–	1.35	0.77	1.05	1.47	–	–
chessboard8	b	sparse	no	0	11.04	0.00	2.17	0.17	1.05	0.47	0.90	0.91	–	–
chessboard8	b	sparse	yes	0	10.58	0.00	2.25	0.16	1.12	0.47	0.96	0.93	–	–
chessboard8	b	real60	no	0	5.76	0.00	1.33	0.05	0.99	0.15	0.88	0.52	–	–
chessboard8	b	real60	yes	0	5.75	0.00	1.42	0.05	1.04	0.17	0.91	0.57	–	–
chessboard8	b	real60	yes	1	5.75	0.00	2.24	0.09	1.33	0.38	1.01	1.40	–	–
chessboard8	b	real60	yes	2	5.75	0.00	3.90	0.07	1.95	0.74	1.23	3.10	0.98	11.18

Table 7.1: Summary of degree and order 8 checkerboard experiments with 8 layers performed over 11 periods, ranging from 5.12 to 106.667 days. The values corresponding to a particular value of μ have been obtained by minimising the penalty functional $R = R_d + \mu R_m$, where R_d is the data misfit and R_m is the model norm. Each new optimisation has been performed starting from the best fitting model of the previous iterations.
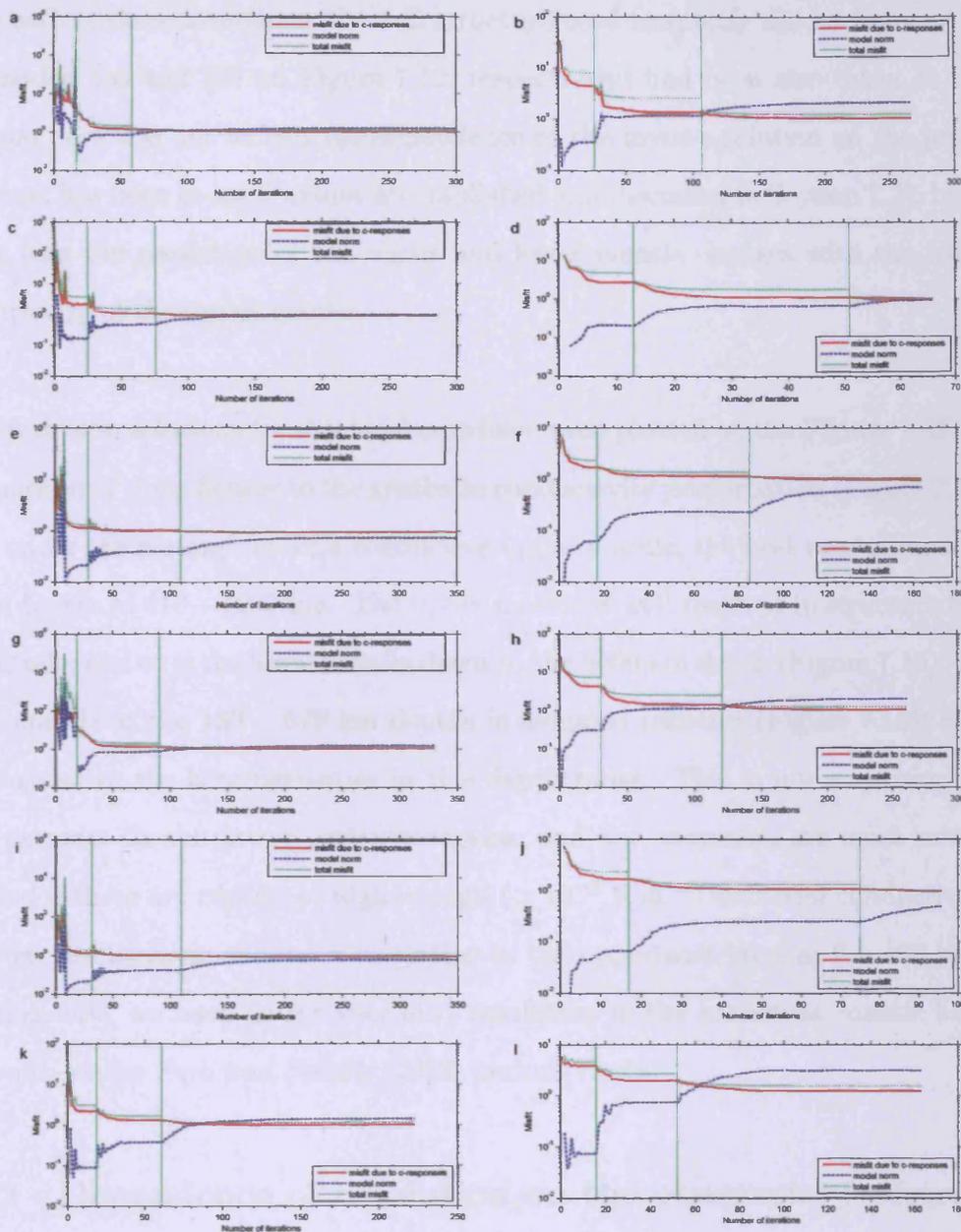
Figure 7.13: Convergence rates for the checkerboard 8 experiment. Solid (red) lines represent the data misfit for c responses, $R_{\mathbf{d}}$. The dotted (blue) lines represent the model norm $R_{\mathbf{m}}$. The dashed (green) lines represent the total $R_{\mathbf{d}} + \mu R_{\mathbf{m}}$, that is being minimised. The vertical lines (green) correspond to the points at which the damping parameter μ has been updated. Shown are the experiments with (**a**) resistive upper mantle, dense observatory distribution; (**b**) conductive upper mantle, dense observatory distribution; (**c**) resistive upper mantle, sparse observatory distribution; (**d**) conductive upper mantle, sparse observatory distribution; (**e**) resistive upper mantle, $-60° \leq \theta \leq 60°$ *Fujii and Schultz* observatory distribution; (**f**) conductive upper mantle, $-60° \leq \theta \leq 60°$ *Fujii and Schultz* observatory distribution; (**g**) resistive upper mantle, dense observatory distribution, near-surface thin layer; (**h**) conductive upper mantle, dense observatory distribution, near-surface thin layer; (**i**) resistive upper mantle, $-60° \leq \theta \leq 60°$ *Fujii and Schultz* observatory distribution, near-surface thin layer; (**j**) conductive upper mantle, $-60° \leq \theta \leq 60°$ *Fujii and Schultz* observatory distribution, near-surface thin layer; (**k**) regularised with $\alpha = 1$, resistive upper mantle and the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* observatory distribution and the near-surface thin layer; (**l**) regularised with $\alpha = 2$, resistive upper mantle and the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* observatory distribution and the near-surface thin layer.

In each of these inversions, the 1-D structure used to specify the background 1-D profile (models (a) and (b) on Figure 7.12, respectively) had been also taken as the prior. Our aim here was not to test the dependence of the inverse solution on the prior model (this task has been to some extent accomplished and discussed in Section 7.2), but instead to see how the resolution of the upper and lower mantle changes with the background conductivity of the upper mantle.

The inverse solutions for this ideal case have been plotted on the Figures 7.15 and 7.14. Comparison of these figures to the synthetic conductivity perturbation (Figure 7.12) shows that, under the assumption of a conductive upper mantle, the best resolution is achieved at the depths $\approx 410 - 1200$ km. The upper mantle is well resolved in structure but not in magnitude, and so is the lower mantle down to the 2000 km depth (Figure 7.15). However, if the mantle at the $100 - 670$ km depths is assumed resistive (Figure 7.15), we are not able to resolve the heterogeneities in this depth range. This is not surprising since the eddy currents do not flow in resistive regions, and any anomalies are much more readily resolved if there are regions of high enough ($> 10^{-2}$ S m^{-1}) electrical conductivity. This, however, results in an enhanced resolution in the uppermost layer at $0 - 100$ km depths. In either case, we have no (or very low) resolution in the lowermost mantle below 2000 km depth at the *Fujii and Schultz* [2002] periods range.

7.3.2 Dependence of resolution on the observatory distribution

Comparison of the Figures 7.15, 7.16 and 7.21 allows us to see how the resolved model changes when the data distribution goes from the perfectly regular and dense, to regular and sparse, to the irregular distribution of realistic observatories. We see how the quality of the resolved model diminished when a sparse synthetic data distribution is used, and how the picture gets totally distorted when a small number of irregular observatories is

used.

Figure 7.22 shows the predicted c responses and d responses for these models (Figures 7.16 and 7.21) compares to the synthetic c responses and d responses. We see how the quality of the response distribution deteriorates when a smaller number of irregularly distributed data is used for the inversion. We also see the equatorial distortions in c responses, as discussed in Section 7.2.

Additionally, we have performed a set of experiments with all five data sets defined earlier in this section, under the assumptions of the conductive upper mantle (prior model (a) on Figure 7.12) and no near-surface complications. The results are plotted on the Figures 7.14, 7.17, 7.20, 7.19 and 7.18. We see that the data at a sparse and regular set of synthetic observatories (Figure 7.17) resolve the mid-mantle structure significantly better than even a larger data set with values distributed irregularly (Figure 7.18). Irregularly distributed data best resolve specific longitudes and latitudes, that depend of the distribution of observatories (compare Figures 7.18, 7.19, 7.20). From these figures it can be seen that an inversion of an irregular data set can deform the anomalies in the the regions of low data coverage. For example, we see that the inversion reconstructs a single zone of enhanced conductivity on Figure 7.20 (the diagonal anomaly across the Eurasian plate), where there should in fact be two separate, smaller, anomalies. We see similar features in other inversions of irregular data. Apart from these distortions no other features are reconstructed that are not in the synthetic model. Even so, we see how distorted our picture of the "checkerboard" mantle becomes when irregularly distributed data are inverted.

For the real inversions (Chapter 8) we have mostly used the *Fujii and Schultz* [2002] $|\theta| \leq 60°$ observatory distribution. We see that with this data distribution and for these

synthetic models (Figures 7.20 and 7.21) most anomalies at the European latitudes (excluding the mid-Atlantic, North American and mid-Pacific anomalies) are reasonably well reconstructed. Additionally, we see most anomalies at the longitudes of Japan, Australia and West Pacific; some anomalies under and around Africa and South America can also be seen. We have very poor resolution in the regions such as the Atlantic, mid- to East Pacific and Indian oceans and, surprisingly, North America.[3]

7.3.3 Effect of the near-surface inhomogeneities on resolution

We have also performed a set of experiments that should better model the real Earth, by including a fixed thin layer of the near-surface conductance (S-conductance) in the synthetic model for the generation of synthetic responses, and then in the respective inversions. We have seen in Section 7.2, how significantly the addition of the near-surface distribution can affect the modelled responses of the Earth. The inverse solutions of these experiments are plotted on the Figures 7.23, 7.26, 7.24, 7.25 and 7.27. Comparison of these results to the respective inverse solutions without this added complication (Figures 7.15, 7.14, 7.16, 7.21 and 7.20) shows no significant deterioration in the quality of solutions. Comparison of the plots of responses for some of these models (Figure 7.28) to the respective responses without the near-surface conductance included in the model (Figure 7.22) shows exactly how different the response distributions are. Here, the synthetic model no longer generated symmetric responses, and the imaginary parts of c and d responses are very different. Note that the c responses only have been inverted, while d responses are plotted for illustrative purposes.

---

[3]Judging from the results of this synthetic experiment.

7.3.4   Effect of penalising higher order terms of parametrization

Two of our inverse model runs employed regularisation by horizontal smoothing, thus penalising higher order terms. This has been done to see the effect of regularisation on an inverse model that actually requires higher-order terms to be successfully reconstructed. Two tests have been performed, with a milder regularisation by setting $\alpha = 1$ and a stronger regularisation, setting $\alpha = 2$. The results are presented in Figures 7.29 and 7.30, and should be compared to Figure 7.25 to see the effect of regularisation. Although we have still been able to fit the data, we were unable to reconstruct most features of the synthetic model, since the checkerboard has been created with degree 8 and order 6 coefficients, penalised by the regularised inversions. Fewer features are reconstructed, and the upper mantle has been made unnecessarily resistive to somehow "compensate" for the lack of sensitivity of these inversions to the small-scale mid-mantle features. Thus, regularisation by smoothing reconstructs larger scale features at the cost of suppressing smaller scale features, which might also be present in the true model. This is an undesirable effect if used while reconstructing a model in which small-scale features are dominant.
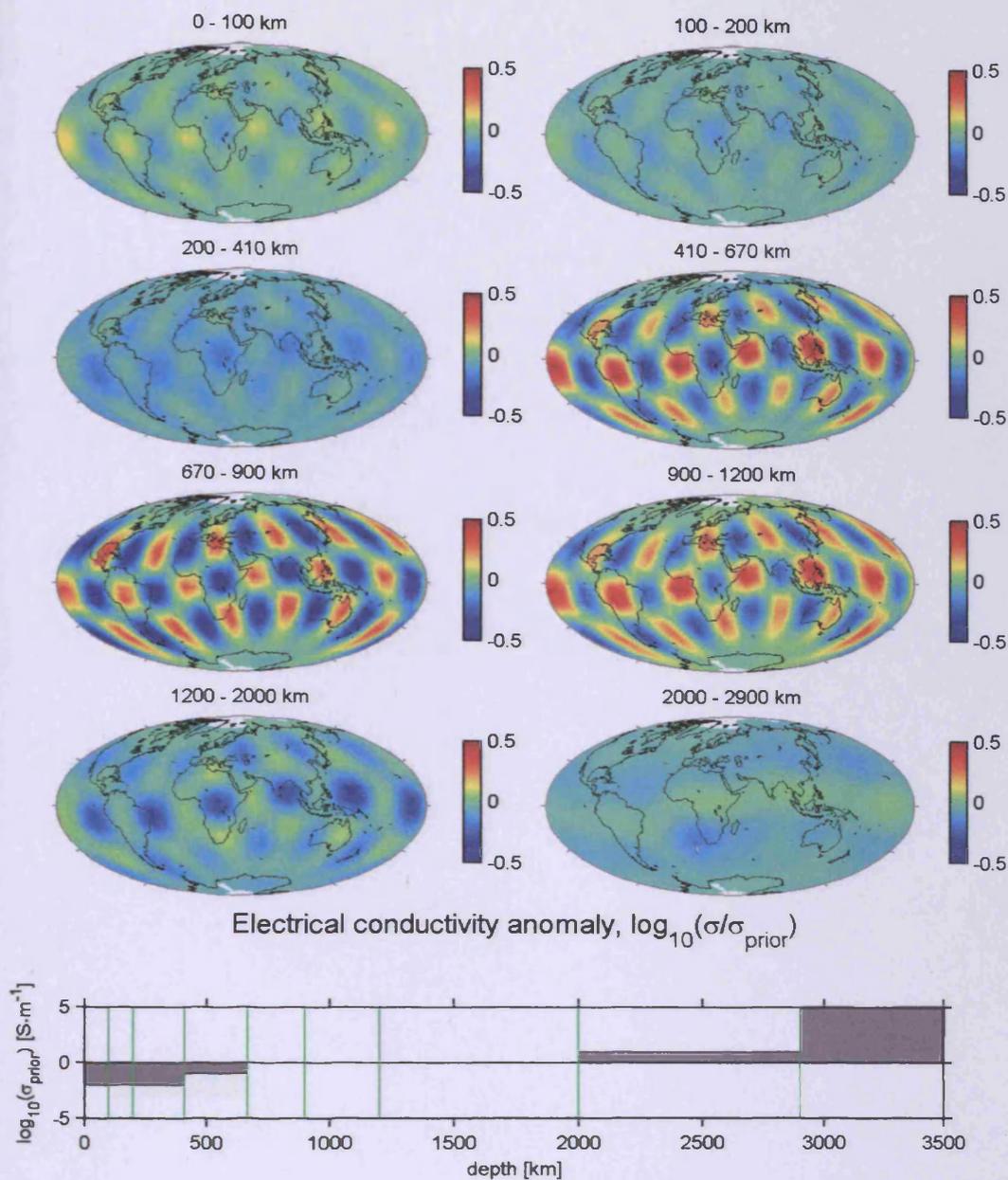
Figure 7.14: Inverse solution of the conductive upper mantle experiment **without** the near-surface conductance, with a regular and dense distribution of observatories. The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
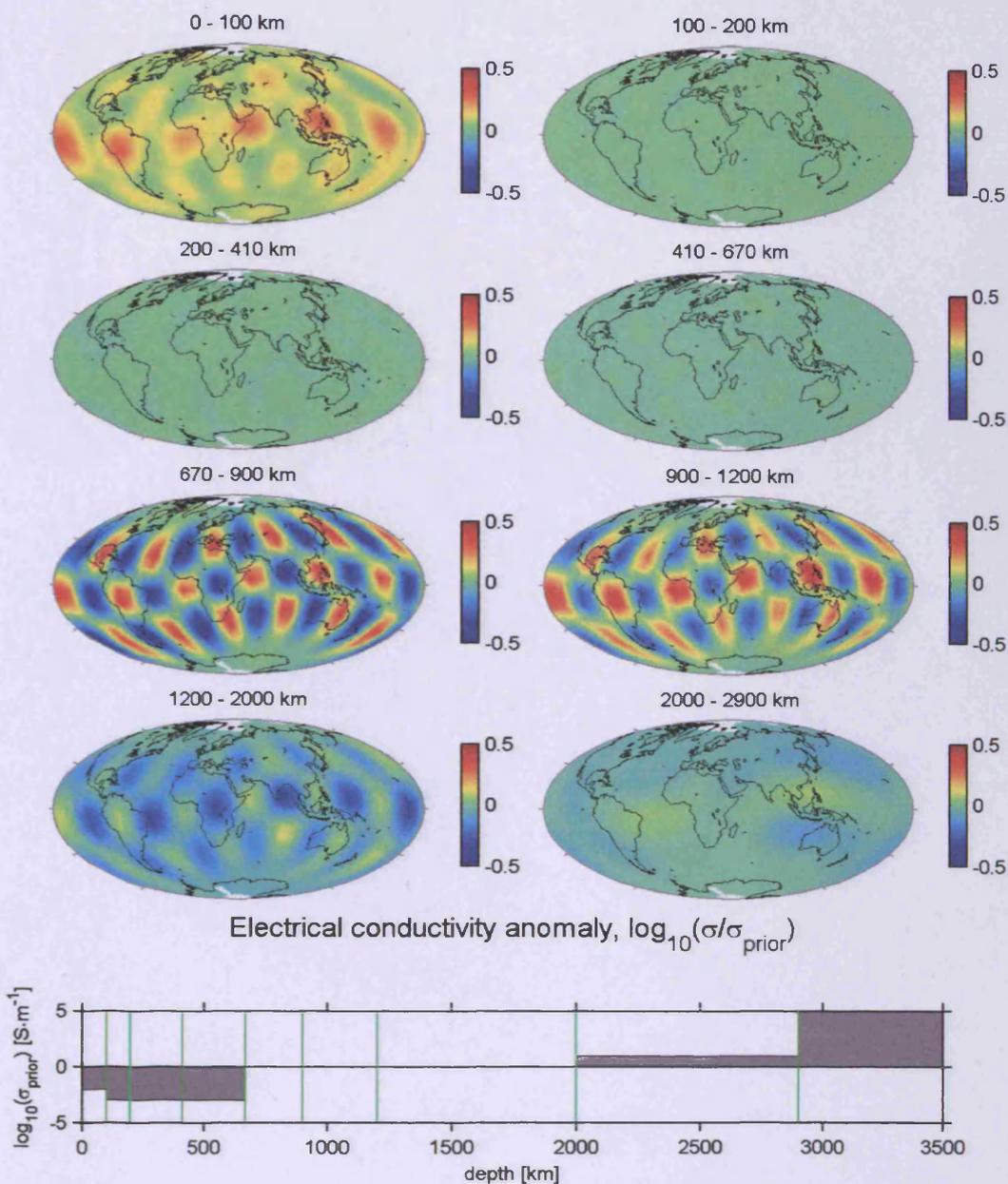
Figure 7.15: Inverse solution of the resistive upper mantle experiment **without** the near-surface conductance, with a regular and dense distribution of observatories. The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
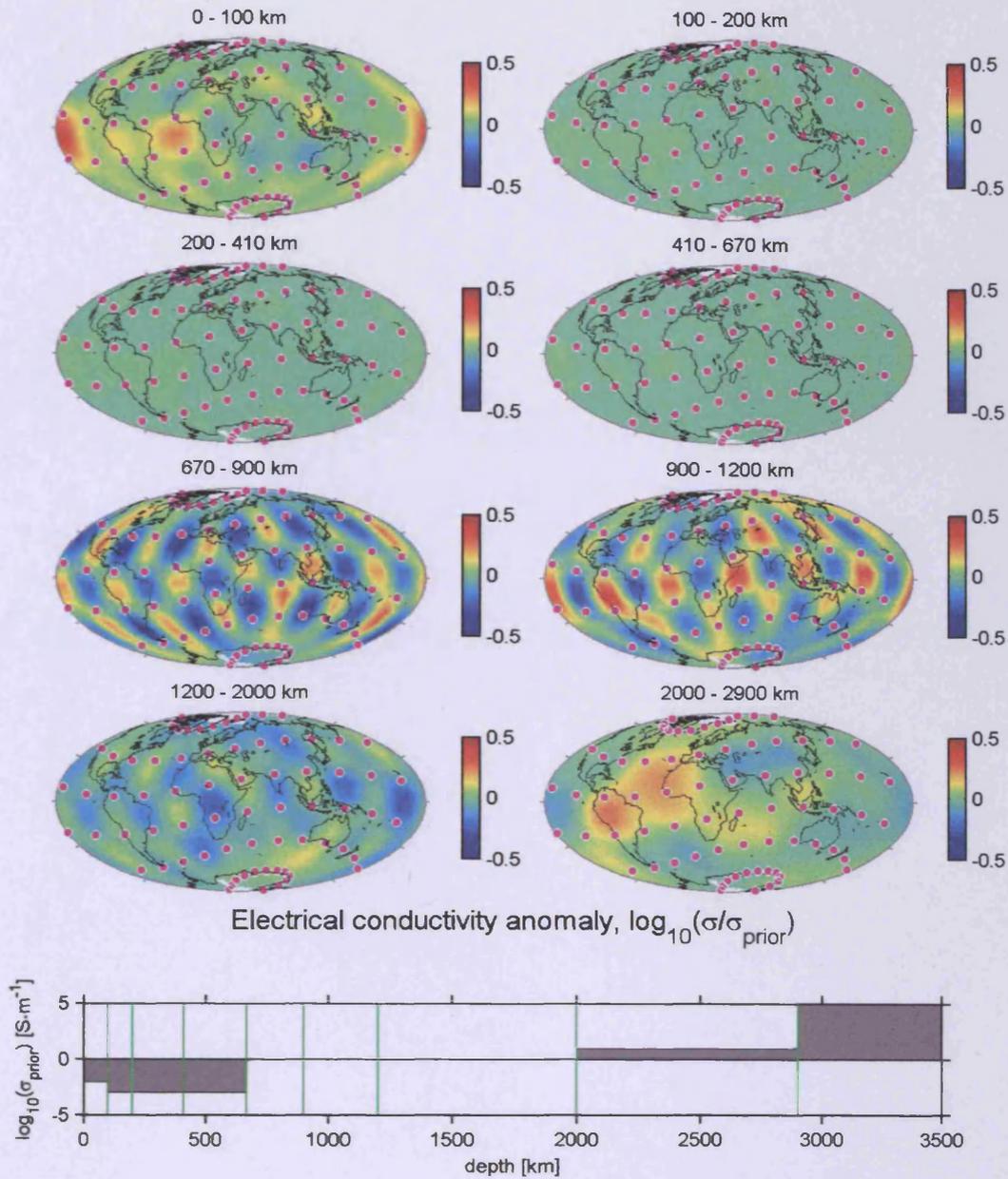
Figure 7.16: Inverse solution of the resistive upper mantle experiment **without** the near-surface conductance, with a regular and sparse distribution of observatories (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
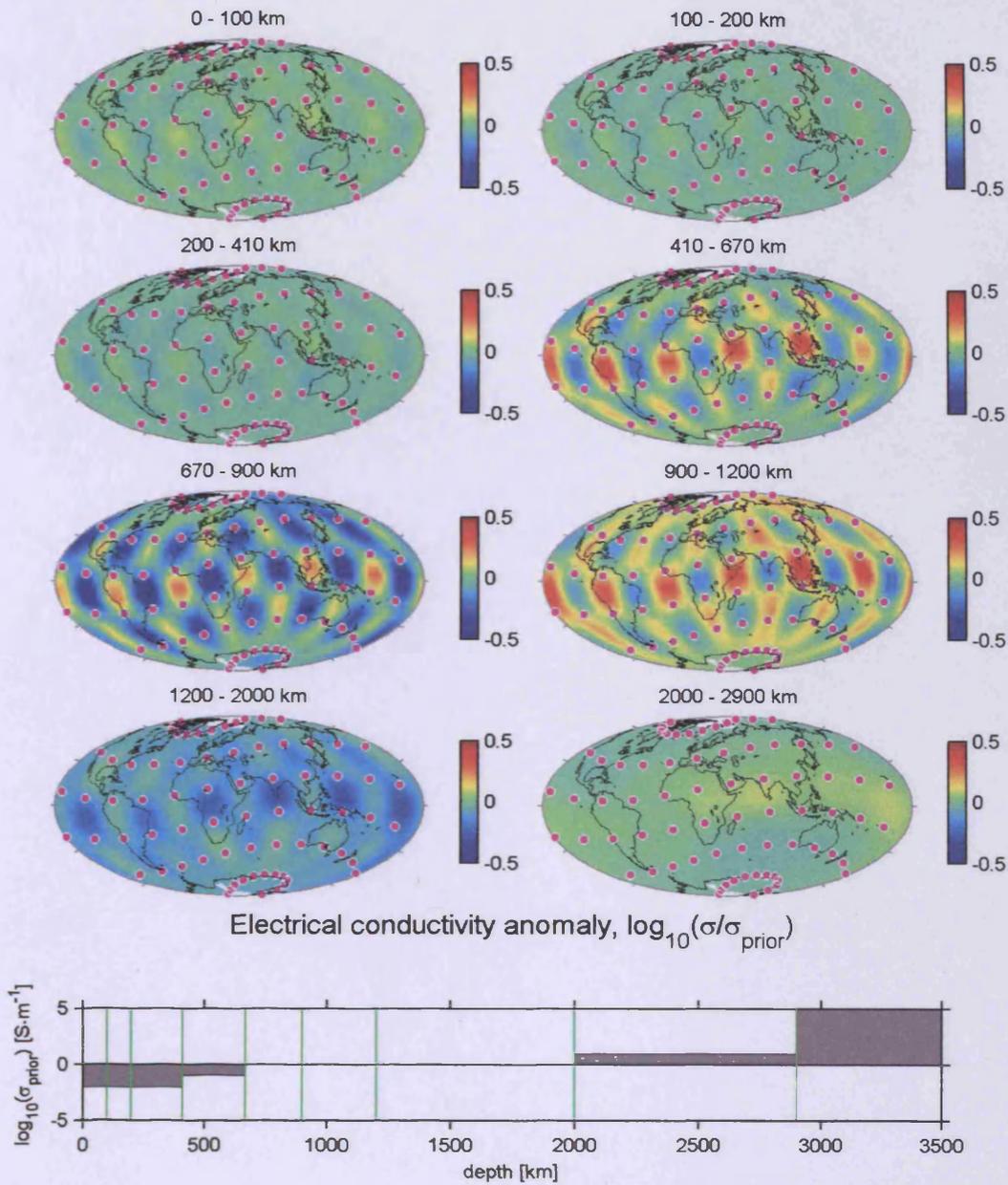
Figure 7.17: Inverse solution of the conductive upper mantle experiment **without** the near-surface conductance, with a regular and sparse distribution of observatories (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
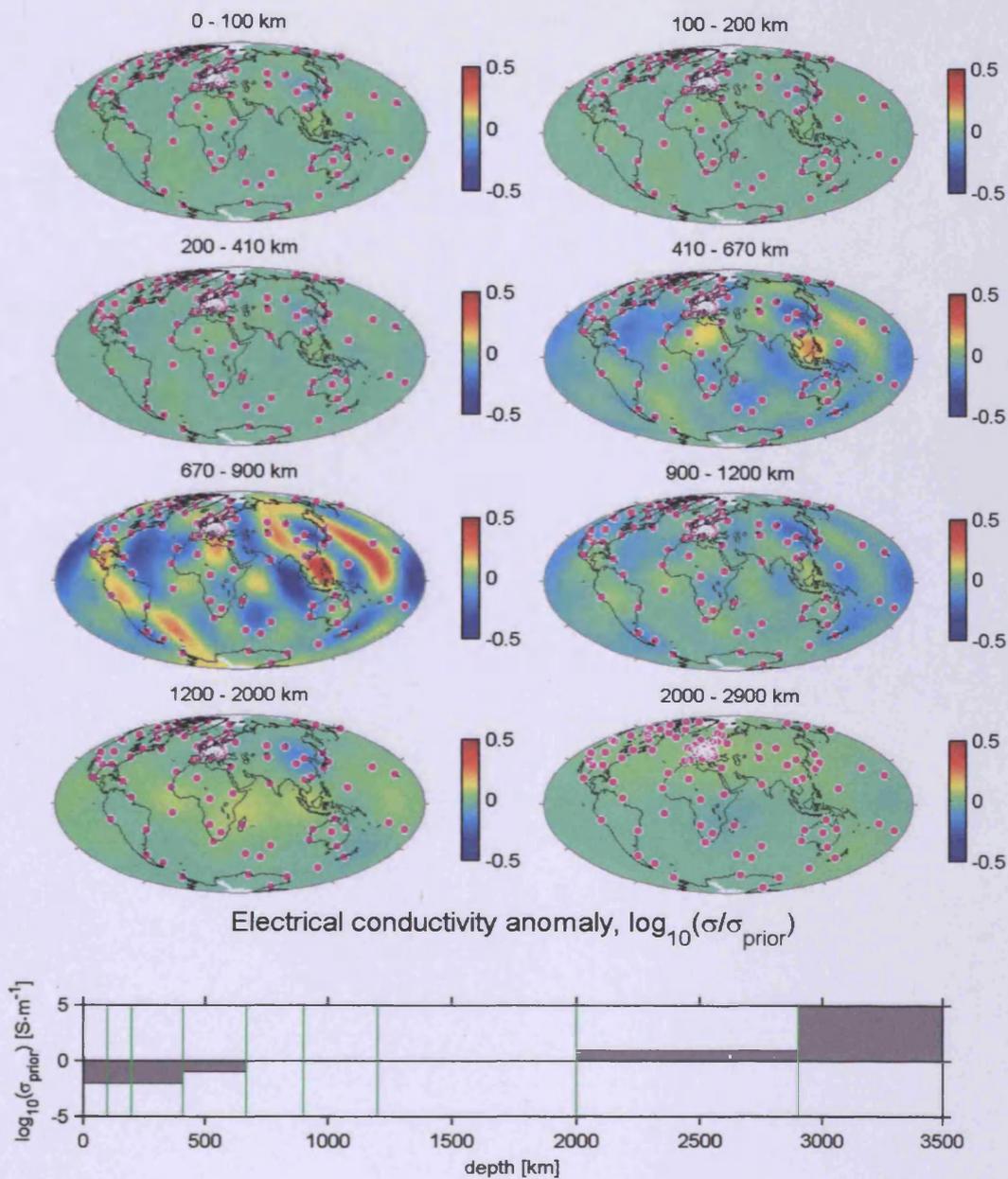
Figure 7.18: Inverse solution of the conductive upper mantle experiment **without** the near-surface conductance, with an irregular INTERMAGNET 2006 distribution of observatories (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
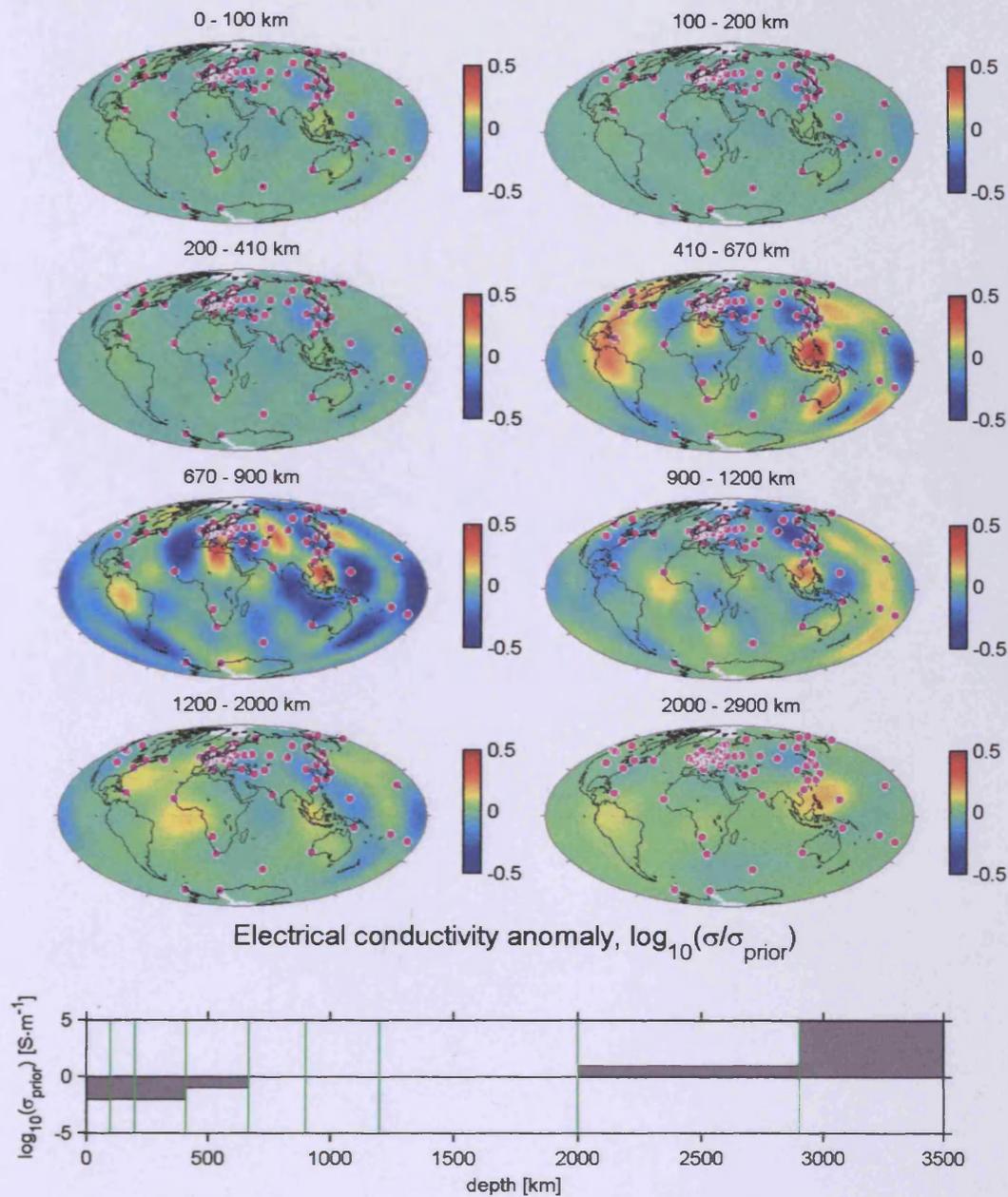
Figure 7.19: Inverse solution of the conductive upper mantle experiment **without** the near-surface conductance, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-65° \leq \theta \leq 65°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
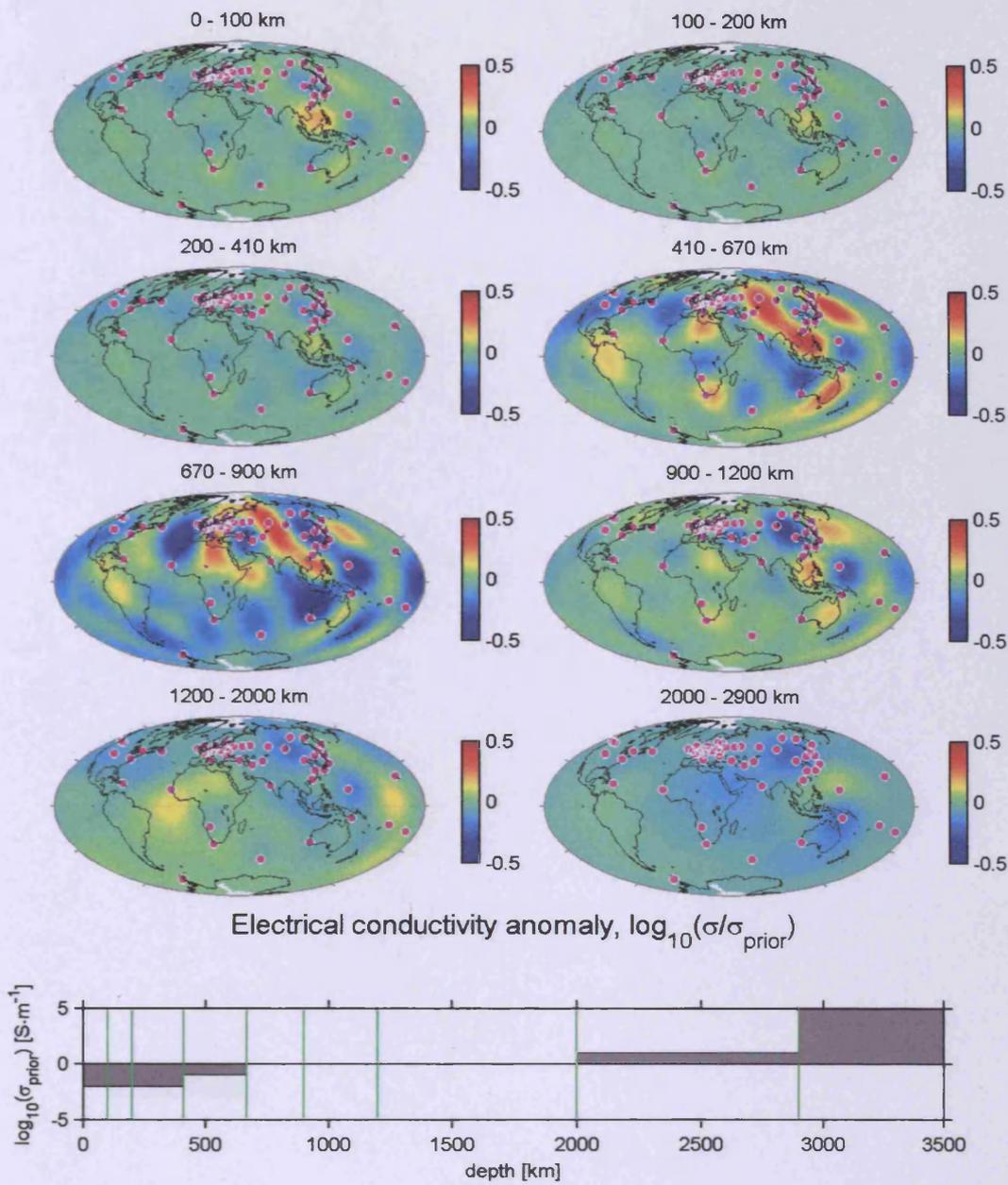
Figure 7.20: Inverse solution of the conductive upper mantle experiment **without** the near-surface conductance, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-60° \leq \theta \leq 60°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
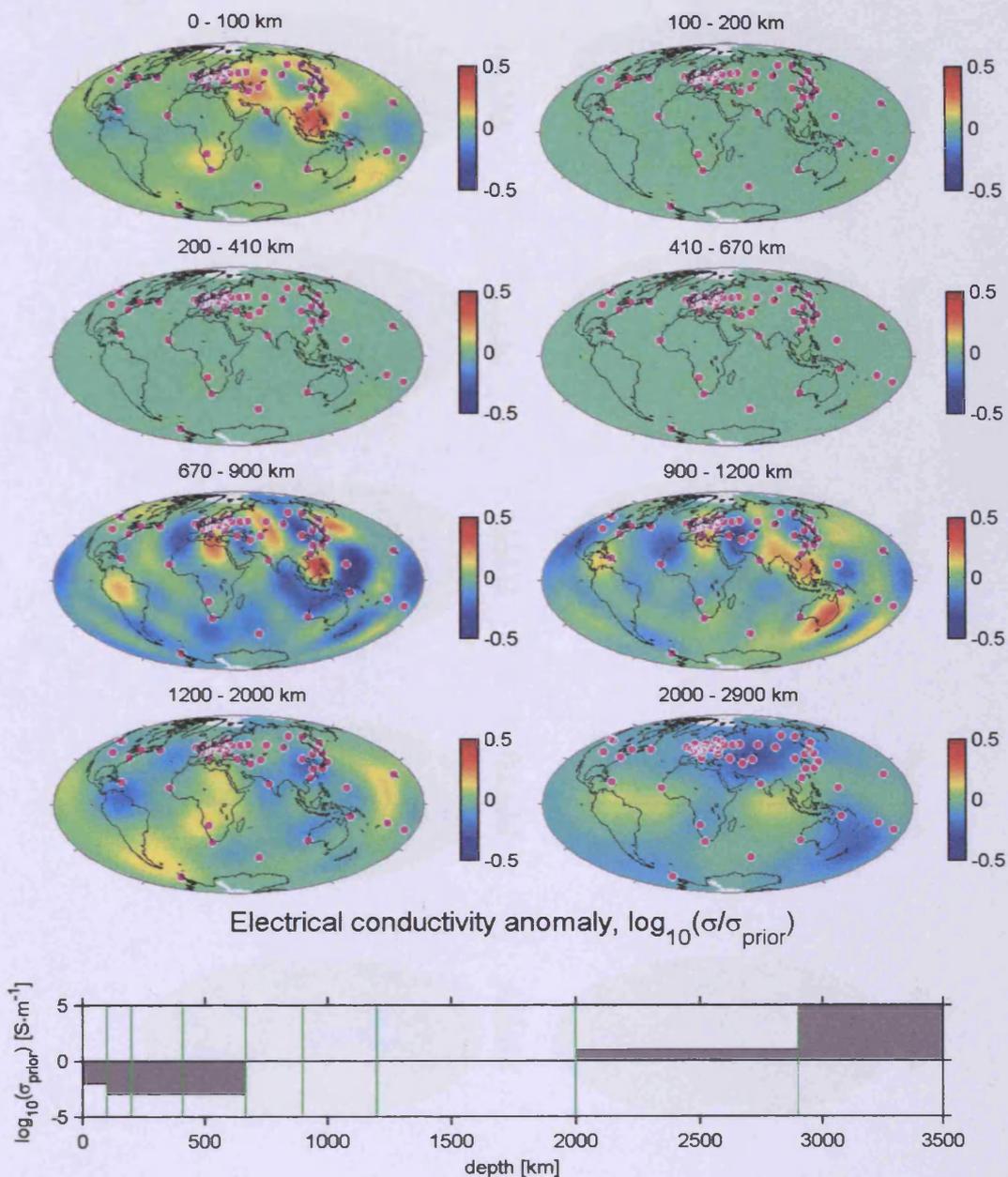
Figure 7.21: Inverse solution of the resistive upper mantle experiment **without** the near-surface conductance, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-60° \leq \theta \leq 60°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
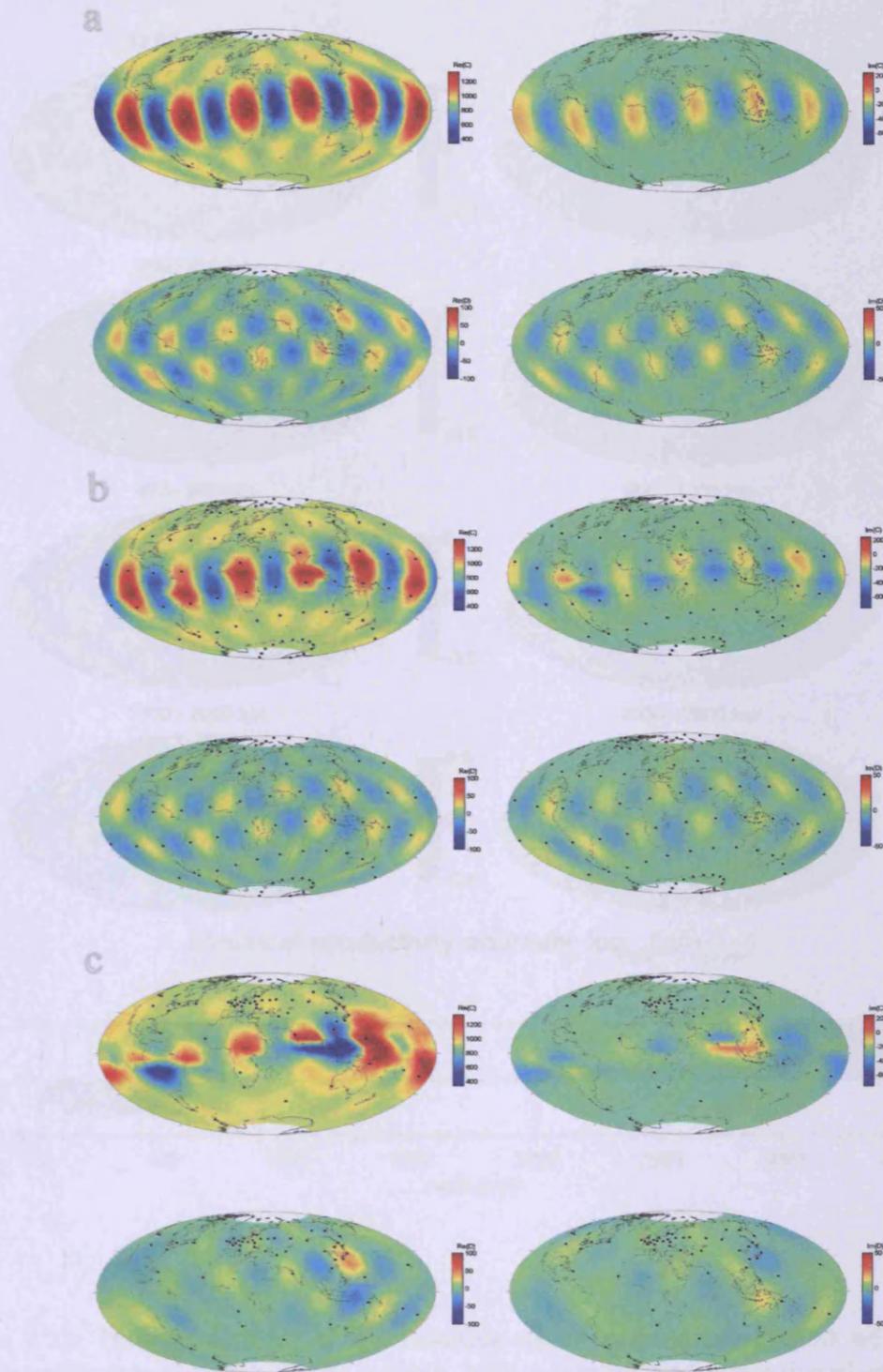
Figure 7.22: Real and imaginary parts of c responses and d responses at the Earth's surface for a sample period of 11.636 days. No near-surface conductance map was used in these models. Plotted are (**a**) responses for the synthetic "resistive upper mantle" checkerboard 8 model; (**b**) responses for the respective inverse solution that utilized data at a regular sparse set of observatories; (**c**) responses for the inverse solution that utilized data from the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of observatories. The two inverse models fit the synthetic data well (least squares misfit ≈ 0.99).
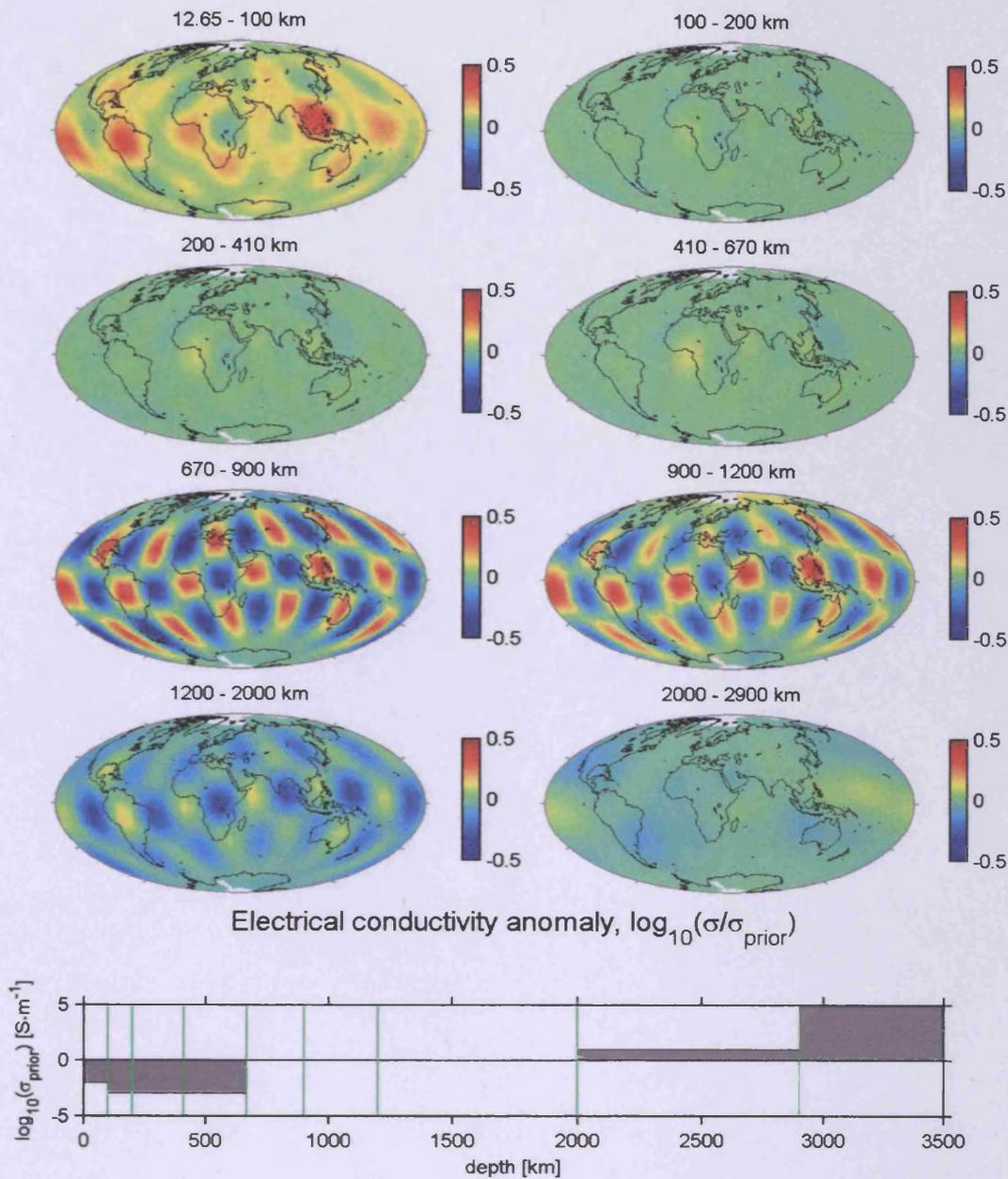
Figure 7.23: Inverse solution of the resistive upper mantle experiment **with** the near-surface conductance on top of the mantle, with a regular and dense distribution of observatories. The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
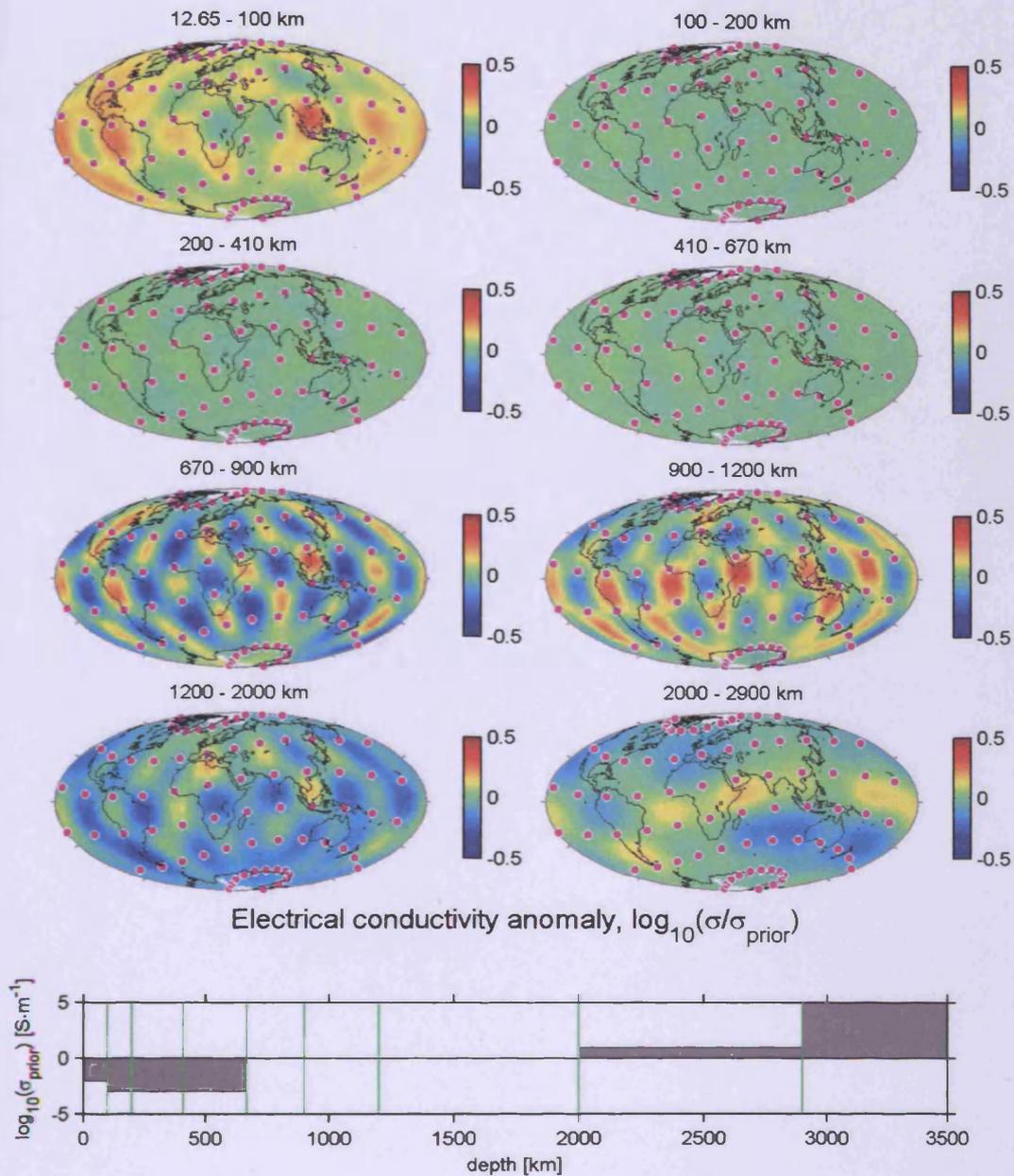
Figure 7.24: Inverse solution of the resistive upper mantle experiment **with** the near-surface conductance on top of the mantle, with a regular and sparse distribution of observatories (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
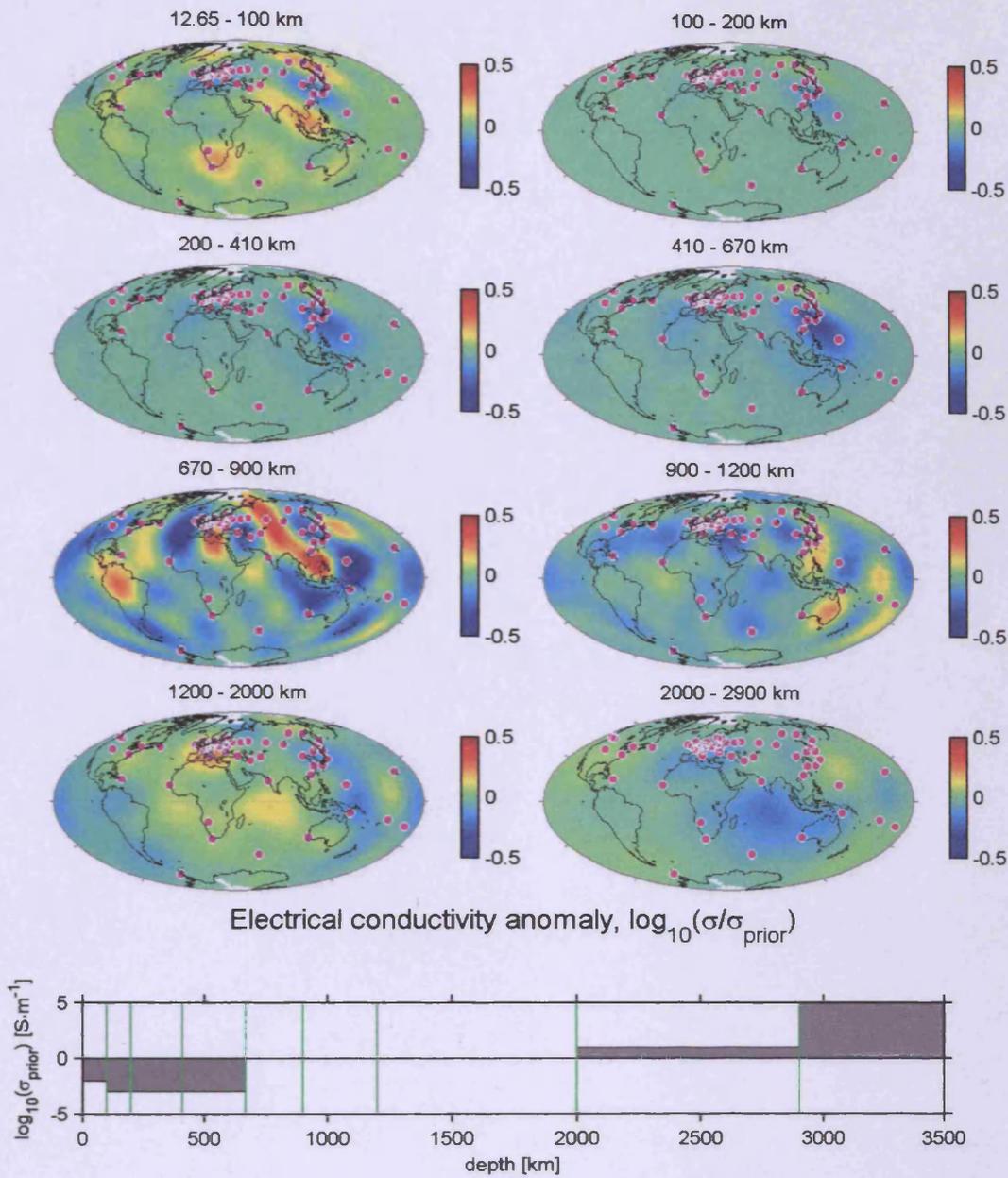
Figure 7.25: Inverse solution of the resistive upper mantle experiment **with** the near-surface conductance on top of the mantle, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-60° \leq \theta \leq 60°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
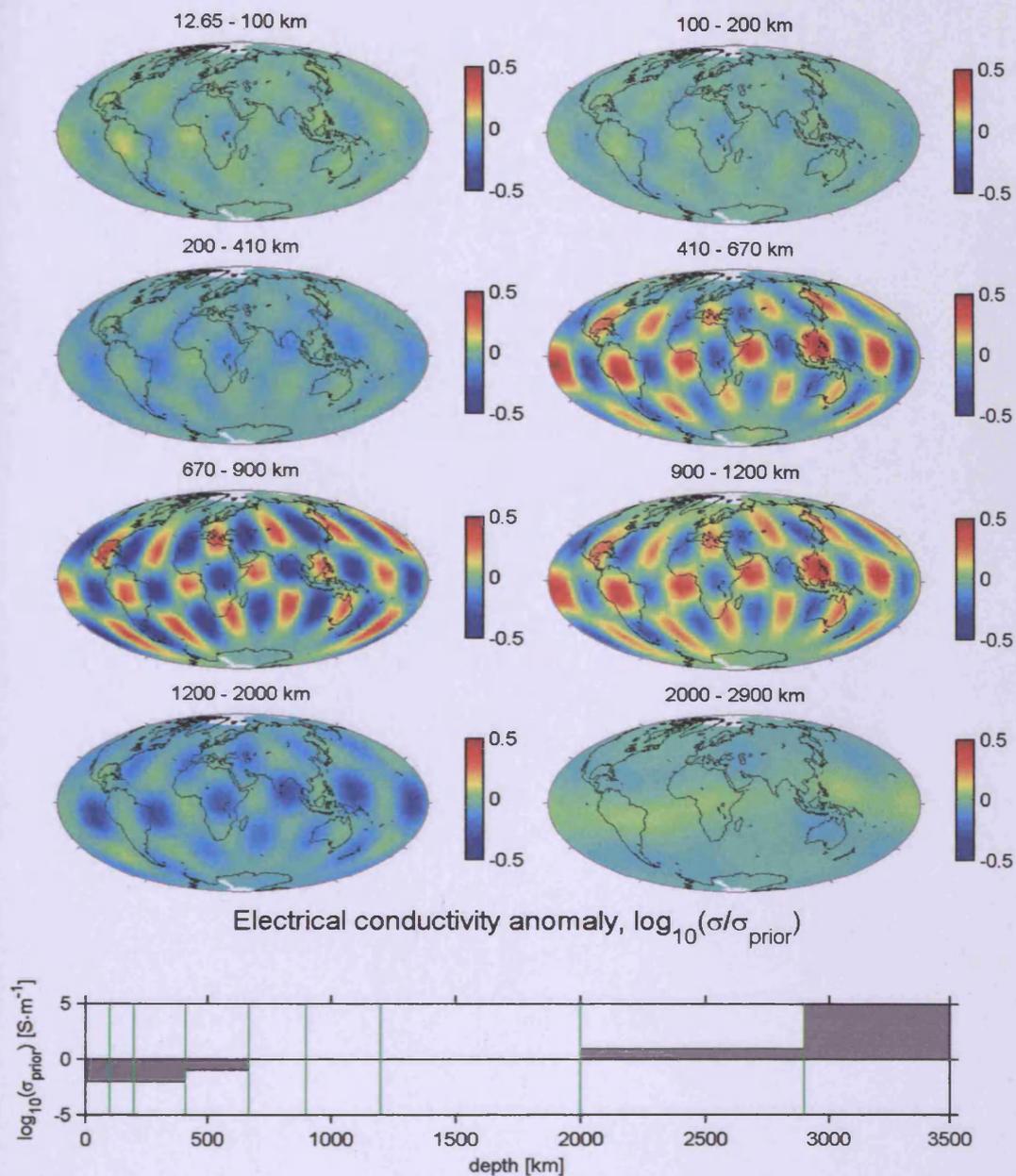
Figure 7.26: Inverse solution of the conductive upper mantle experiment **with** the near-surface conductance on top of the mantle, with a regular and dense distribution of observatories. The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
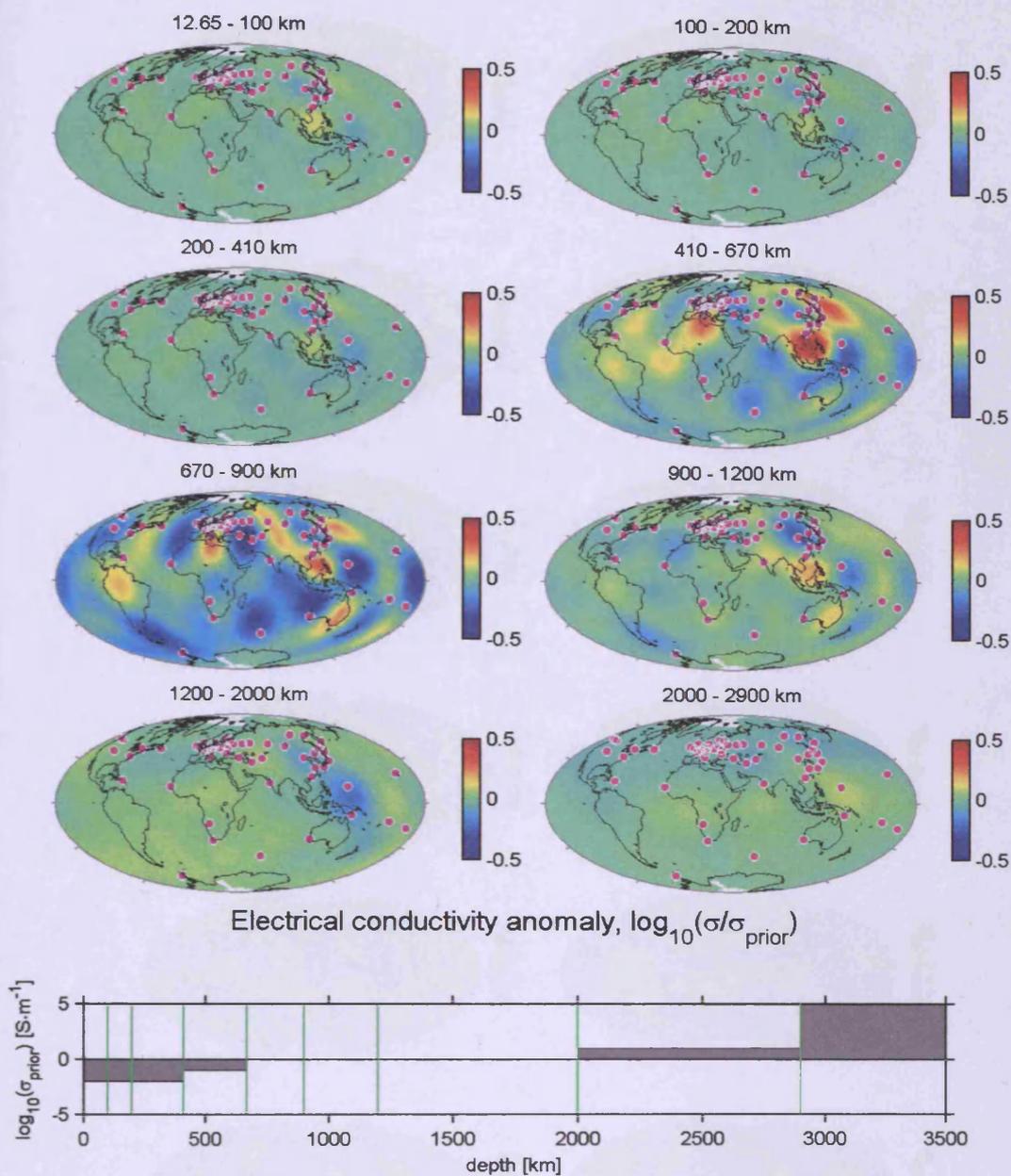
Figure 7.27: Inverse solution of the conductive upper mantle experiment **with** the near-surface conductance on top of the mantle, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-60° \leq \theta \leq 60°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure.
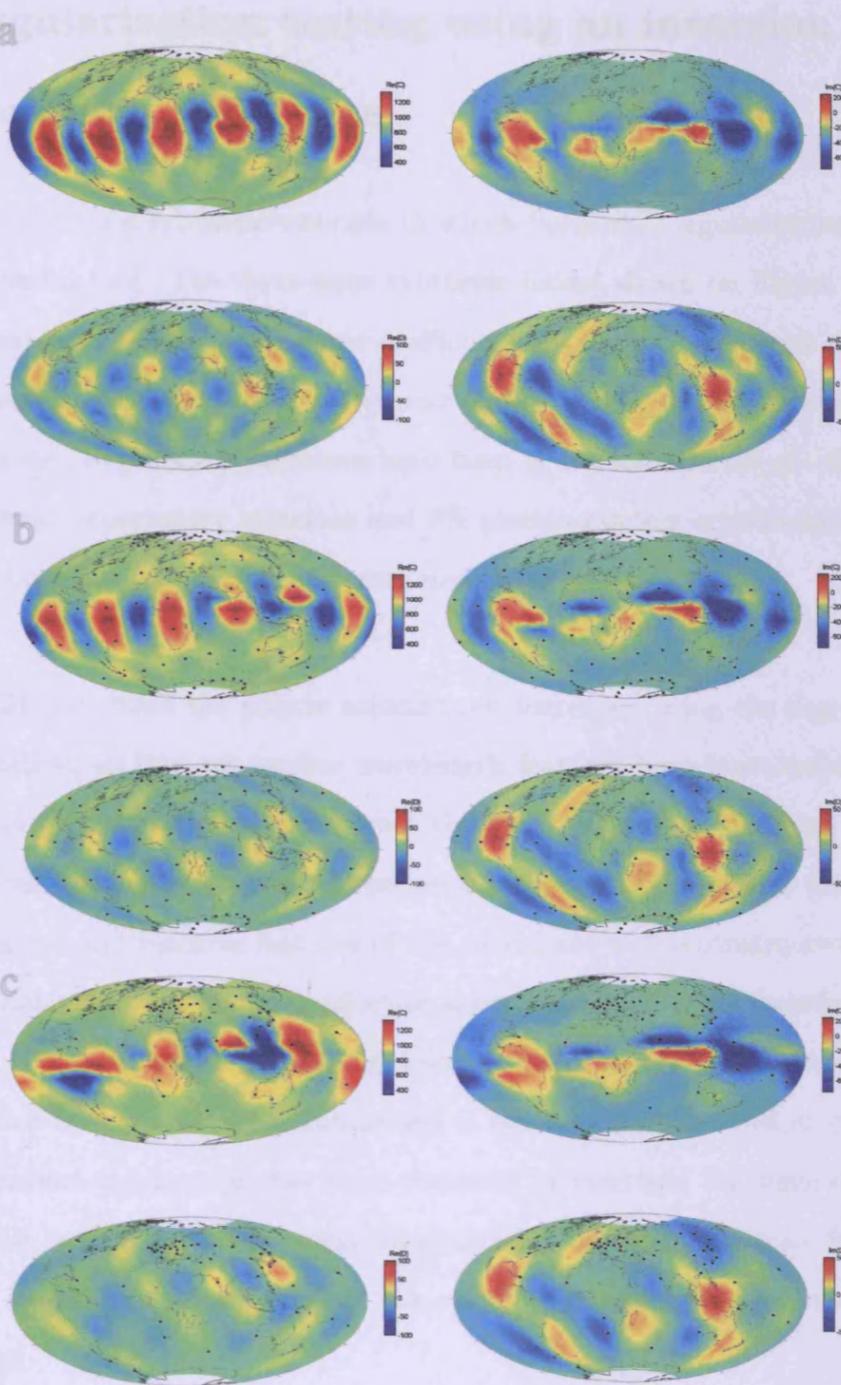
Figure 7.28: Real and imaginary parts of c responses and d responses at the Earth's surface for a sample period of 11.636 days. Near-surface conductance map was included in these models. Plotted are (**a**) responses for the synthetic "resistive upper mantle" checkerboard 8 model; (**b**) responses for the respective inverse solution that utilized data at a regular sparse set of observatories; (**c**) responses for the inverse solution that utilized data from the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of observatories. The two inverse models fit the synthetic data well (least squares misfit ≈ 0.99).

7.4 Regularisation testing using an inversion of large wavelength features

Let us now resort to a synthetic example in which horizontal regularisation would be a particularly useful tool. The three-layer synthetic model shown on Figure 7.31 (a) has been generated with spherical harmonic coefficients up to the degree and order 4, to include large wavelength features only. The near-surface conductance has been imposed on top of this model. Synthetic c responses have been generated at a set of $-60° \le \theta \le 60°$ *Fujii and Schultz* observatory locations and 5% pseudo-random errors added. This data set has then been inverted by four different means, as follows.

Figure 7.31 (b) shows the inverse solution reconstructed using the degree and order 4 parametrization, so that no smaller wavelength features have been included into the inversion than that required to reconstruct the synthetic model. Keeping in mind the poor spatial resolution of this data set, we see a rather satisfying result: pretty much all of the conductive and resistive features of the model are well reconstructed, apart from the resistive anomalies and one of conductive anomalies in the upper mantle. As we have already seen, the upper mantle, extending down to the seismic transition at 670 km depth, is most affected by low spatial resolution and is not very well resolved at our frequency range. The inverse solution (c) has been obtained by inverting the same data set with the degree and order 8 parametrization, thus allowing much smaller scale features to be seen (if they are in the synthetic model). We see that unnecessarily small scale anomalies are introduced.

It could be misleading to see such small-scale features in a real inversion, if they are not necessarily there. A lower degree and order parametrization ensures that they do not

12.65 - 100 km

100 - 200 km

200 - 410 km

410 - 670 km

670 - 900 km

900 - 1200 km

1200 - 2000 km

2000 - 2900 km

Electrical conductivity anomaly, $\log_{10}(\sigma/\sigma_{prior})$
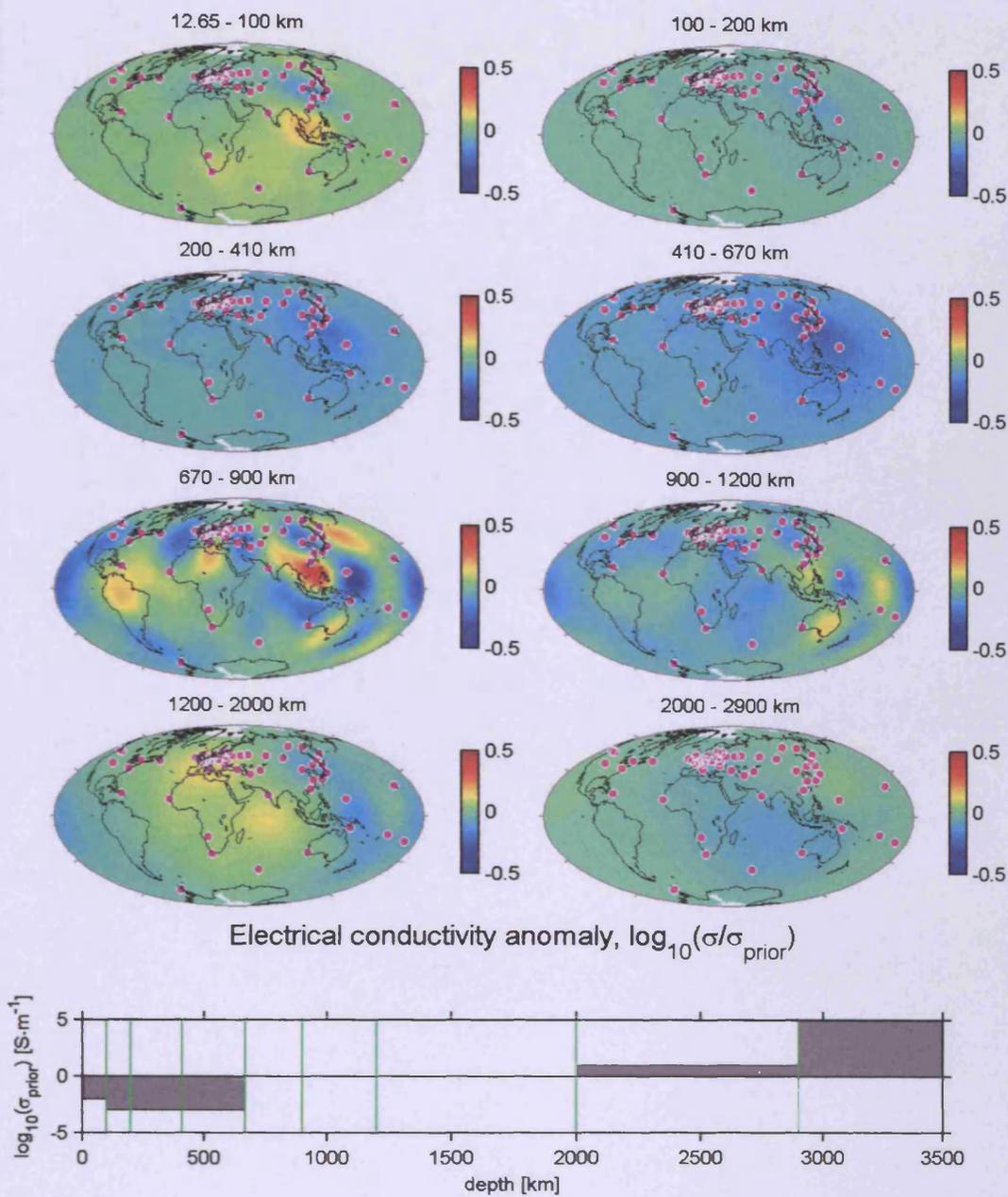
Figure 7.29: Inverse solution of the resistive upper mantle experiment **with** the near-surface conductance on top of the mantle, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-60° \leq \theta \leq 60°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure. Additional regularisation has been achieved by setting the parameter $\alpha = 1$.
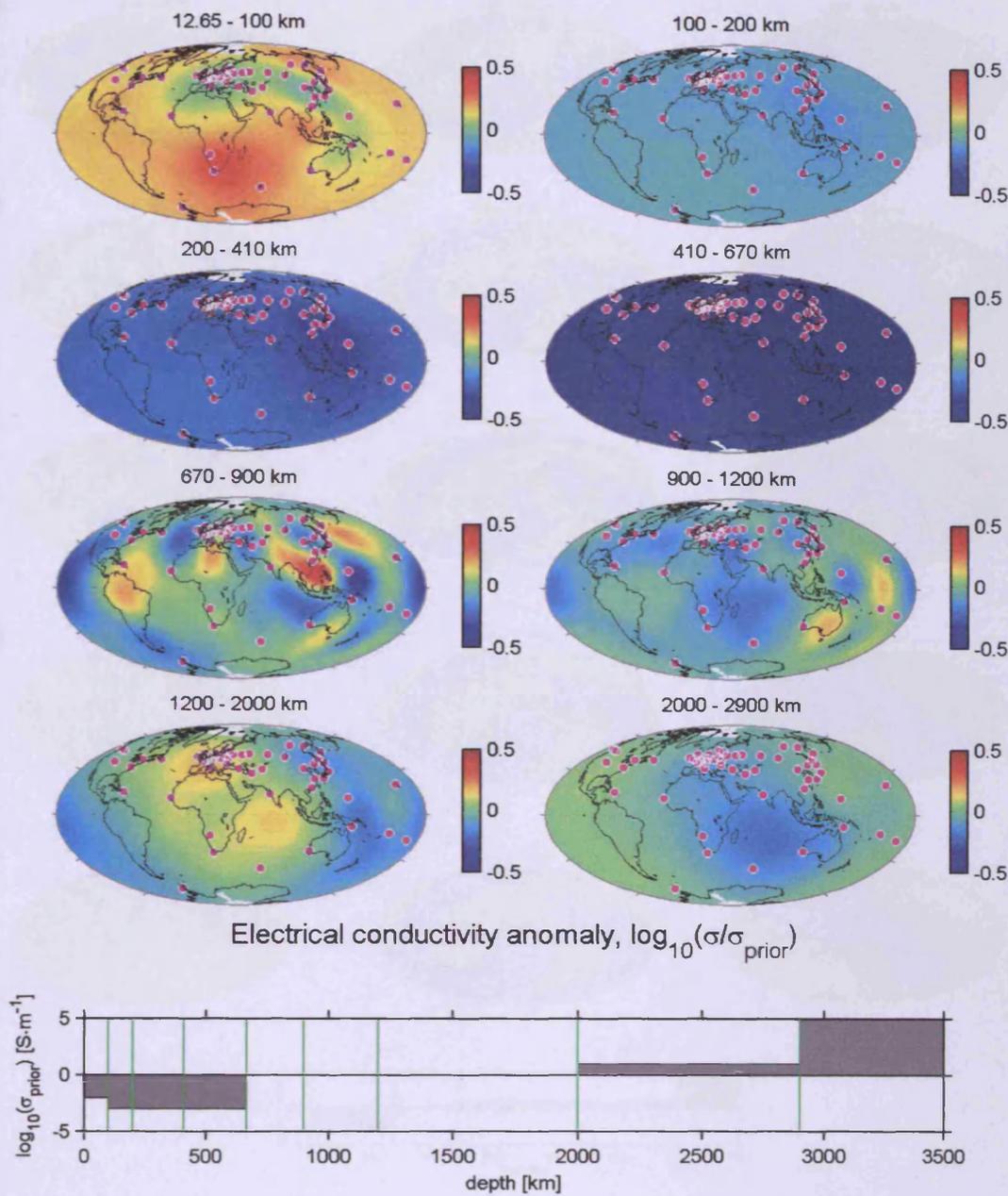
Figure 7.30: Inverse solution of the resistive upper mantle experiment **with** the near-surface conductance on top of the mantle, with an irregular *Fujii and Schultz* mid-latitude distribution of observatories at $-60° \leq \theta \leq 60°$ (shown by pink dots). The solution is presented as a logarithmic perturbation around the prior model, which is specified by the shaded area on the bottom figure. Additional regularisation has been achieved by setting the parameter $\alpha = 2$.
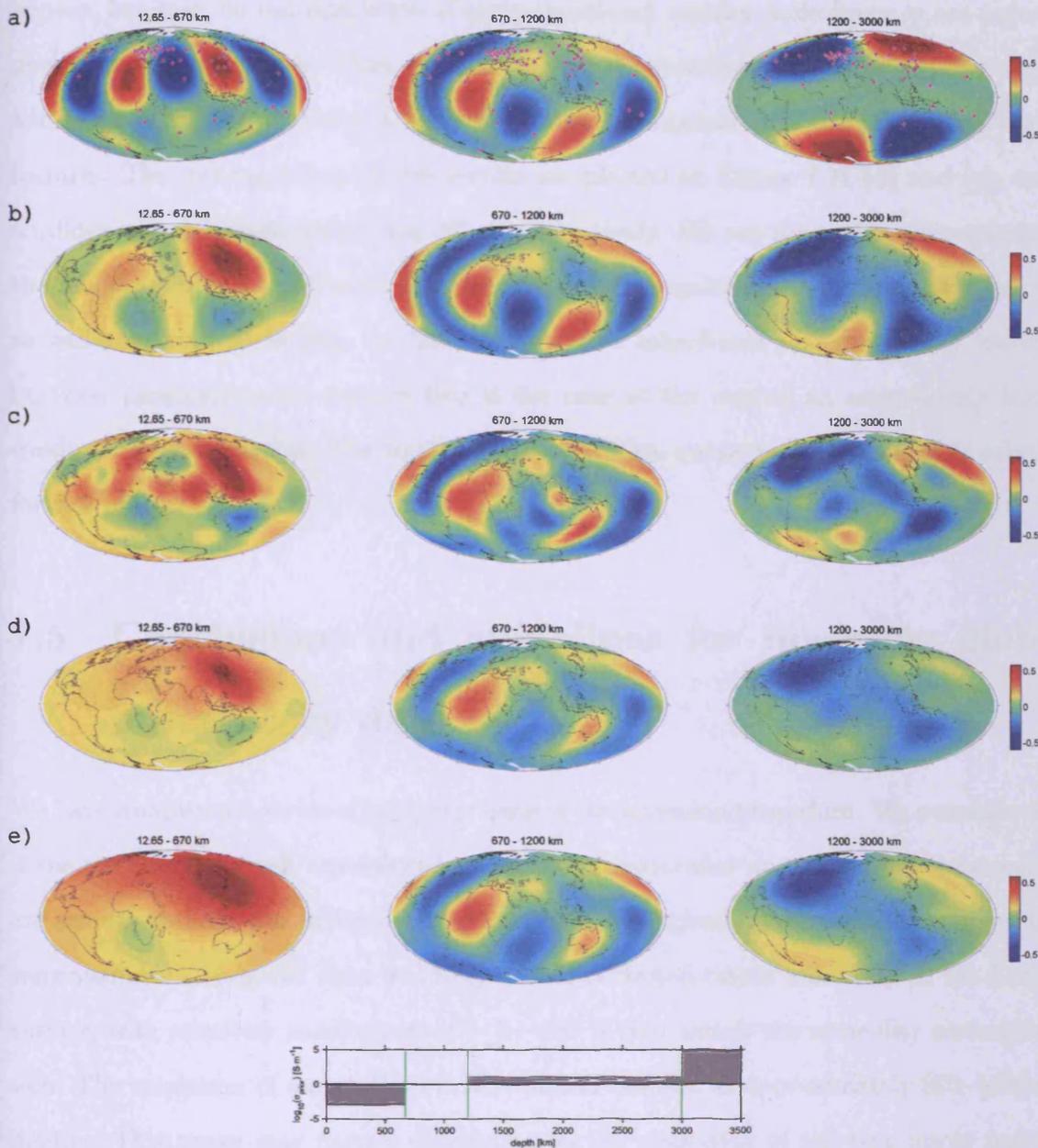
Figure 7.31: Inversion results for the large wavelength (degree and order 4) synthetic experiment designed to test the regularisation. Figure (**a**) is the three-layer synthetic model, on top of which the near-surface conductance distribution has been imposed. Figure (**b**) shows the inversion result obtained by searching in the domain of degree and order 4 models. Figure (**c**) shows the inversion result obtained by searching in the domain of degree and order 8 models. Figures (**d**) and (**e**) are inversion results obtained in the class of degree and order 8 models, with additional regularisation: $\alpha = 1$ and $\alpha = 2$, respectively.

appear, but may be too restrictive if some significant smaller scale features are actually present in the true Earth. Thus, instead we can parametrize by penalising higher degree and order terms, while using a flexible enough parametrization to include small scale features. The corresponding inverse results are plotted on Figure 7.31 (**d**) and (**e**), using a milder and a stronger smoothing effect, respectively. We see that we again reconstruct the model quite well, and although the shape and magnitude of the anomalies are not as well recovered as in (**b**), we have nevertheless introduced no unnecessary features. Stronger parametrization ensures this is the case at the cost of an anomalously highly conductive upper mantle. The milder parametrization seems to be a reasonable solution for this problem.

7.5 Conclusions and guidelines for inverting global observatory data

We have completed a series of synthetic tests of the inversion procedure. We conclude that if the true Earth is well represented by spherical harmonics and all the inhomogeneities are smooth enough, our inversion would perform well, given a large amount of good data were available. By 'good' data we imply data distributed rather uniformly on the Earth's surface, with relatively small errors of $\sim 5 - 100$ % that satisfy the normality assumptions well. The resolution of our method is best in mid-mantle, at approximately 500–1600 km depths. This range may narrow depending on the resistivity of the true upper mantle, which affects both the skin depth and the sensitivity to the upper mantle structures. It could also widen, if additional data at periods < 5 days and > 107 days were employed.

We have shown that the quality of synthetic inverse solutions is not affected by the S-distribution. We have also studied how the resolution in the mantle depends on the

number and spatial distribution of the data. We have shown that the solution quality deteriorates significantly going from the regular to an irregular data distribution. We have also specified the spatial regions in which the resolution of $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data is best, and those in which the resolution is poor. This knowledge could be of help in our analysis of field data inversions in Chapter 8, enhancing our confidence in the resolved conductive anomalies in these regions. Additionally, we have studied the effects of regularisation by horizontal smoothing. We have shown that mild regularisation could be quite helpful, assuming the true model is such that the large wavelength features dominate. However, if small scale features are dominant in the true model, regularisation creates undesirable effects.

We should also note that these rather satisfying results have all been obtained for the examples in which both the synthetic perturbations and the inverse solutions have been restricted to the space of harmonic functions on layers. If the true Earth is best represented by some other space, such as, for example, a space of discontinuous 'blocky' models, or by a much larger number of inhomogeneous layers (i.e. radially continuous functions), we are unlikely to reconstruct anything like the true structure using spherical harmonic decomposition on 3–8 layers. For the analysis of Chapter 8 we assume that the Earth could be well represented with this parametrization.

Chapter 8

Non-linear global 3-D Inversion – II.

Results and discussion

We describe a set of global 3-D models of the electrical conductivity distribution in the Earth, obtained by inverting the *Fujii and Schultz* [2002] data set. We discuss the features present in our models and compare them to the three-dimensional features of global and regional electromagnetic models obtained previously. We also describe and compare the corresponding predicted magnetic fields and responses of the Earth.

8.1   Choosing an adequate parametrization

By the choice of the parametrization we mean the choice of the number and the depths of the heterogeneous layers in the mantle, as well as the degree of spherical harmonic perturbation imposed on each of these layers.

The resolving power of our data is such that we would only be able to reconstruct a small number of independently parametrized heterogeneous layers.[1] Therefore, for this

---

[1] Judging from Figures 8.6 and 8.7, a larger number of layers may be required to reconstruct the vertical

study we restrict our parametrization in terms of the number of layers, never exceeding 8 heterogeneous layers in the mantle.

By setting fixed values for the depths of layer boundaries, we effectively allow abrupt jumps of conductivity at these depth (unless vertical regularisation is performed). Thus, we generally base our choice of layer boundaries on the seismic evidence for compositional discontinuities in the mantle (Section 1.1). In particular, we always set two of the layer boundaries to ~ 410 km and ~ 670 km depths to mark the top and the bottom of the transition zone. Other layer boundaries vary between the models.

The choice of the degree of spherical harmonic expansion depends on the spatial scale (or wavelength) of inhomogeneities expected in each of the layers. Higher degree and order expansions allow the reconstruction of smaller scale features. On the other hand, we are restricted in this choice by the resolution of the sparse and noisy data: features smaller than some threshold will not be resolved; thus there is no point including them in the model.

Analysis of travel-time anomalies of long-period shear waves [e.g. *Su and Dziewonski*, 1991] suggests, that the heterogeneity in the Earth's mantle, which probably reflects the patterns of mantle convection, is dominated by long-wavelength features. The analysis of the power of SS residuals[2] as a function of harmonic degree, performed by *Su and Dziewonski* [1992], shows a distinct difference in the rate of decrease of the power below

---

variation in the electrical conductivity from the uppermost mantle down. The adjoint approach that we have developed in Chapter 5 allows us to model the mantle as a large number of thin heterogeneous layers, regularised by vertical smoothing, without increasing the cost of forward solver and derivative computations. However, this parametrization would require an extremely large model space comprising up to a hundred thousand of parameters - a number comparable to the size of the grid. Such an increase in the model space dimension would significantly complicate the inverse search; possibly making it totally infeasible to search in this domain.

[2]SS waves are shear waves that reflect off the Earth's surface, Figure 1.1
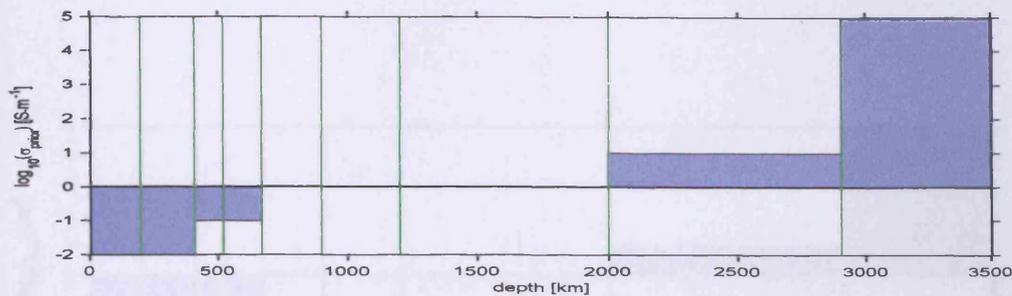
Figure 8.1: Prior model (**a**), used in the experiments **earth4**, **earth8**, **uppermantle8** (see Table 8.1). Log conductivity is shown by the shaded area, the vertical lines denote layer boundaries.

and above degree 8. The rate is first slower than L_1, and then it exceeds L_2. This indicates that there is a preference for features with dimensions larger than 2500–3500 km (when projected onto a two-dimensional map at the Earth's surface). Assuming the scale of the heterogeneities in the electrical conductivity distribution is comparable to the scale of mantle convection patterns, we conclude that a degree and order 8 spherical harmonic parametrization would be appropriate for this study.

Thus, for most inverse runs discussed in Sections 8.3 and 8.4, we parametrize the Earth with degree and order 8 spherical harmonics on all layers. We have also tested a lower degree and order parametrization to see whether the *Fujii and Schultz* [2002] is fittable with long-wavelength structures only. Additionally, we have attempted to penalise higher degree and order features by regularisation. Please refer to Table 8.1 and to the rest of Section 8.3 for a discussion.
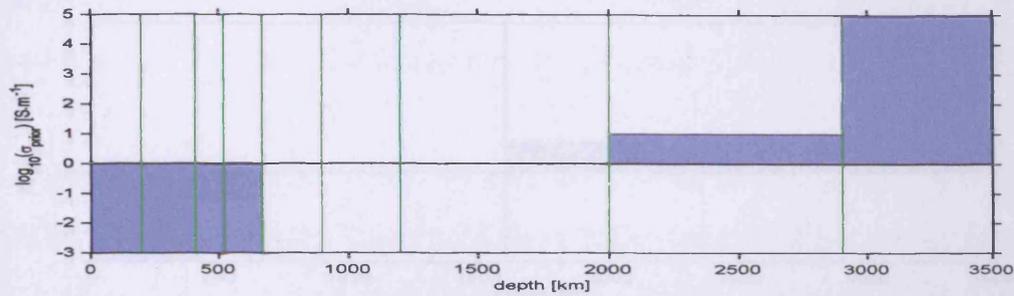
Figure 8.2: Prior model (**b**), used in the experiments earth4a, earth8a, uppermantle8a (see Table 8.1). Log conductivity is shown by the shaded area, the vertical lines denote layer boundaries.
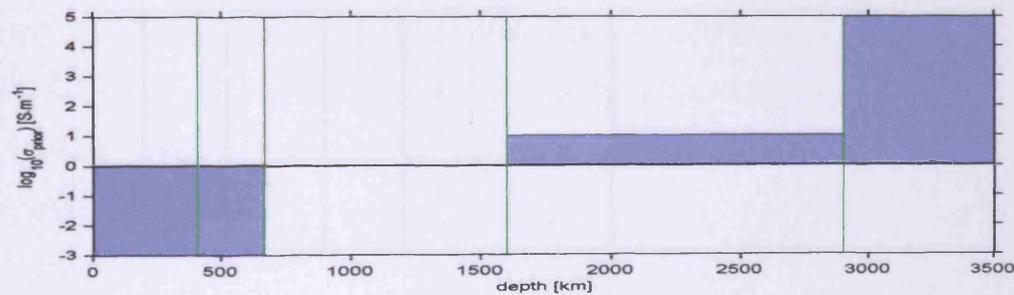
Figure 8.3: Prior model (**c**), used in the experiment fourlayers8 (see Table 8.1). Log conductivity is shown by the shaded area, the vertical lines denote layer boundaries.
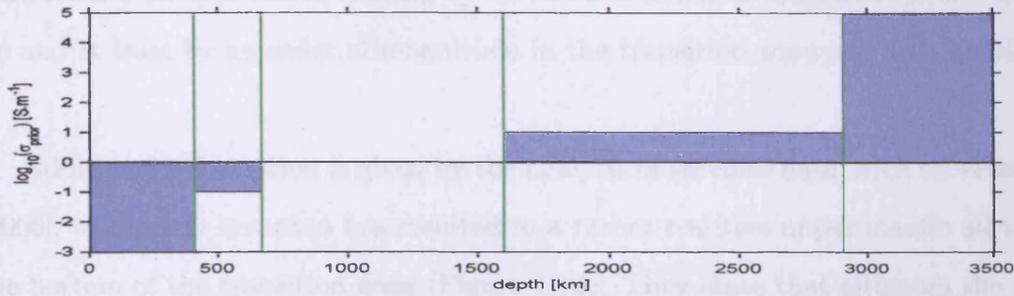
Figure 8.4: Prior model (**d**), used in the experiment fourlayers8a (see Table 8.1). Log conductivity is shown by the shaded area, the vertical lines denote layer boundaries.
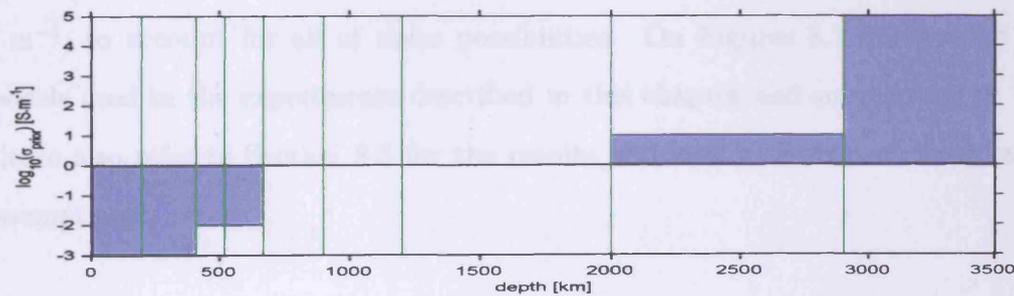
Figure 8.5: Prior model (**e**), used in the experiment transition8 (see Table 8.1). Log conductivity is shown by the shaded area, the vertical lines denote layer boundaries.

8.2   Considerations for the choice of the prior model

We have included in Section 1.4 a range of global 1-D profiles found in the literature. Many of them are summarised in Figure 8.6. We see that a wide diversity of electrical conductivity profiles exists, varying by up to three orders of magnitude in the upper mantle and at least by an order of magnitude in the transition zone and mid-mantle regions.

Additional information is given by the analysis of satellite data, such as *Velimsky et al.* [2006], whose 1-D inversion has resulted in a rather resistive upper mantle all the way to the bottom of the transition zone (Figure 1.12). They state that although the resolution of their method in the resistive upper mantle sandwiched between conductive crust and lower mantle is poor, an upper bound of 0.01 S m^{-1} is suggested by the data. On the other hand, the satellite inversion of *Kuvshinov and Olsen* [2006], corrected for the induction in the oceans, results in the values that are quite close to those previously obtained by inverting observatory data (Figure 8.7).

Here, we obtain and describe a set of inverse results for several prior electrical conductivity values in the upper mantle and the transition zone, ranging from 10^{-3} to 10^{-1} S m^{-1}, to account for all of these possibilities. On Figures 8.1–8.5 we plot the prior models used in the experiments described in this chapter and summarised in Table 8.1. Please also refer to Section 8.3 for the results obtained with each of these prior model assumptions.
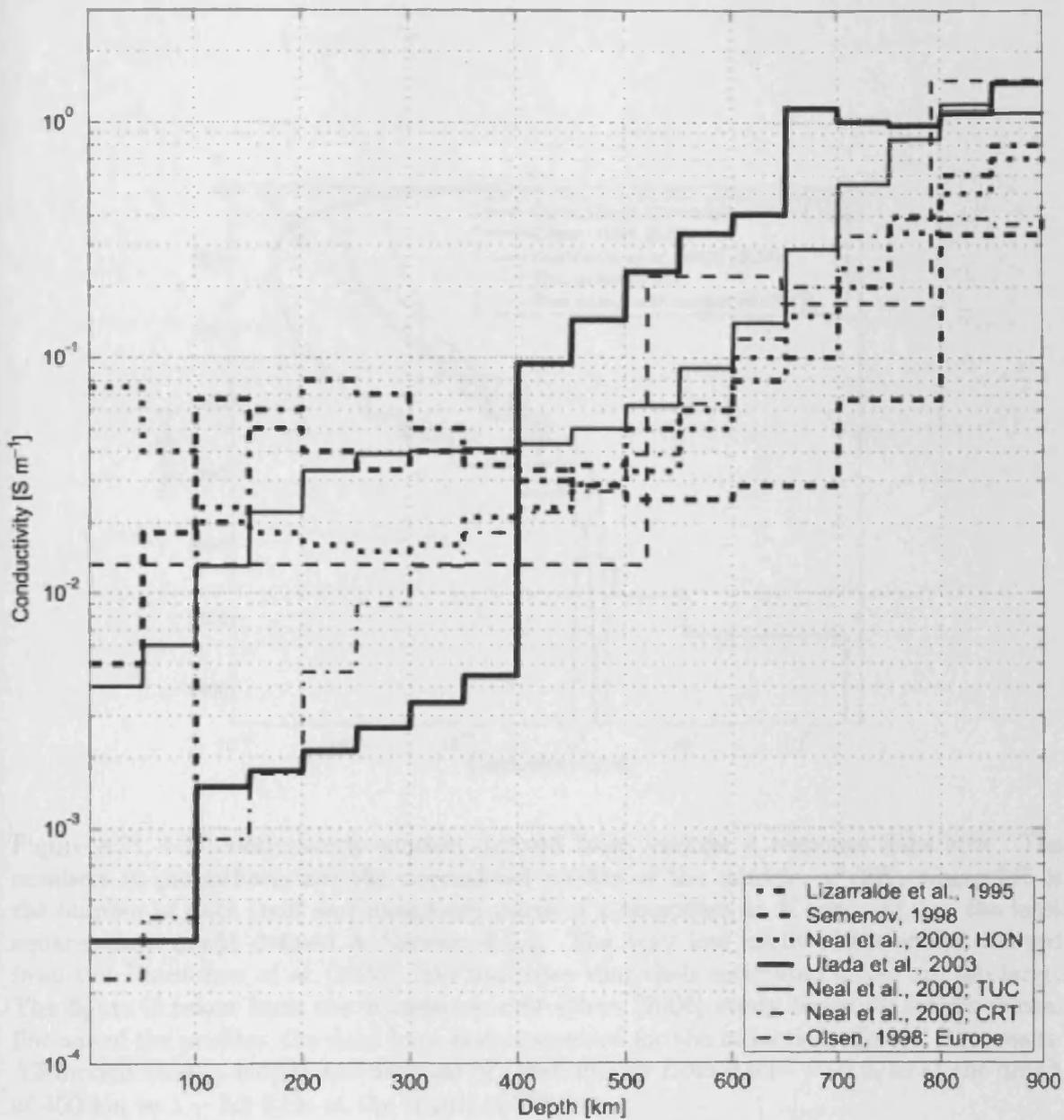
Figure 8.6: 1-D mantle conductivity profiles obtained for the North Pacific Ocean region. The mean 1-D section for Europe [*Olsen*, 1998] and the 1-D sections at Carty Lake (CRT) and Tucson (TUC) from [*Neal et al.*, 2000] are presented for comparison. Figure taken from *Kuvshinov et al.* [2005].
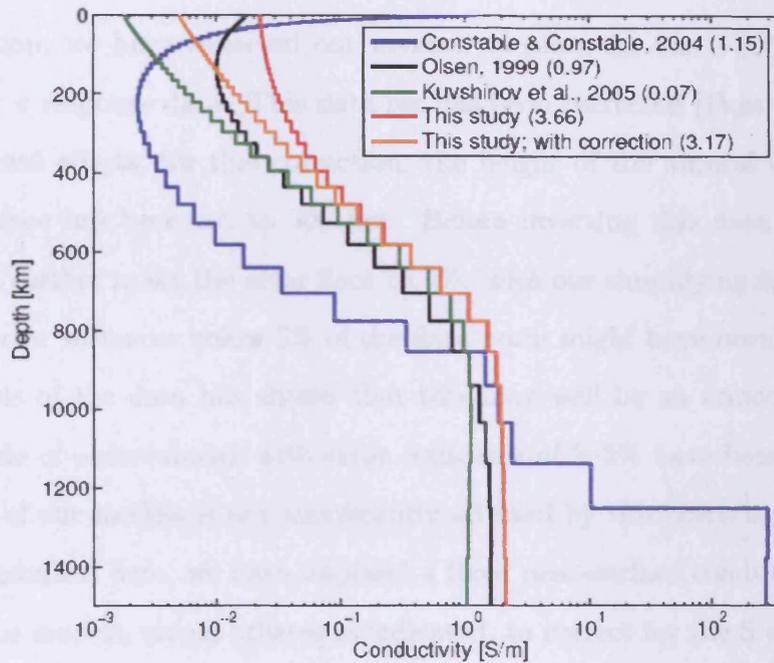
Figure 8.7: 1-D conductivity models derived from various c response data sets. The numbers in parenthesis are the normalized misfits of the models, $\chi^2/2K$, where $2K$ is the number of data (real and imaginary parts of c responses at K periods), i.e. the least squares data misfit defined in Section 4.3.1. The very low misfit of the model derived from the *Kuvshinov et al.* [2005] data indicates that their estimated errors are too large. The figure is taken from the *Kuvshinov and Olsen* [2006] study based on satellite data. For one of the profiles, the data have been corrected for the induction effect in the oceans. All models show a monotonic increase of conductivity from $0.03 - 0.09$ S/m at the depth of 400 km to $1 - 2.5$ S/m at the depth of 900 km.

8.3   Reconstruction of the 3-D electrical conductivity distribution using the mid-latitude data from the *Fujii and Schultz* [2002] data set

In this section, we have collected our inverse solutions for the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* c response data. This data set has been corrected [*Fujii and Schultz*, 2002] for the auroral effects; for this correction, the height of the auroral currents above the Earth's surface has been set to 300 km. Before inverting this data set, we have pre-processed it further to set the error floor to 5%: with our simplifying source assumptions, estimated error variances below 5% of the data norm might have been too optimistic. In fact, analysis of the data has shown that this may well be an unnecessary precaution: only a couple of observatories with error estimates of 3–5% have been affected, and the data misfit of our models is not significantly affected by this correction. For the inverse solutions described here, we have imposed a fixed near-surface conductance distribution on top of the models, unless otherwise indicated, to correct for the S-effect.

8.3.1   Experiments with four layers

We have performed a set of experiments with four-layer inhomogeneous degree and order 8 mantle, assuming either a resistive or a relatively conductive transition zone. We have been able to fit the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data set of c responses under both assumptions. However, the regularised experiment penalising higher order terms ($\alpha = 1$) has not resulted in a model which fits the data, which suggests that in this experiment higher order terms are just as significant.

Here, we present two four-layer models fitting the data (Figures 8.8 and 8.9), cor-

Experiment	Degree	N. of layers	N. of 3-D layers	α	GM lat. (degrees)	S-cond.	Prior	Initial	$\mu = 1$	$\mu = 0.1$	$\mu = 0.01$	$\mu = 0.001$	$\mu = 0$
earth4	4	8	8	0	60	yes	b	2.68 (11)	1.91 (11)	1.54 (11)	1.32 (11)	1.21 (11)	1.32 (13)
earth8	8	8	8	1	60	yes	b	2.68 (11)	2.04 (11)	1.56 (11)	1.24 (11)	1.23 (34)	1.29 (13)
earth8	8	8	8	0	60	yes	b	2.68 (11)	1.71 (11)	1.32 (11)	1.02 (11)	–	1.00 (34)
earth8	8	8	8	1	60	no	b	2.77 (11)	2.18 (11)	1.69 (11)	1.10 (11)	–	0.97 (34)
uppermantle8	8	8	4	1	60	yes	b	2.68 (11)	2.42 (11)	2.33 (11)	2.32 (11)	2.28 (11)	–
uppermantle8	8	8	4	0	60	yes	b	2.68 (11)	2.41 (11)	2.29 (11)	2.22 (11)	2.25 (11)	–
earth8	8	8	8	0	75	yes	b	8.14 (11)	6.65 (11)	6.07 (11)	3.42 (11)	–	–
earth8	8	8	8	1	75	yes	b	8.14 (11)	7.21 (11)	6.43 (11)	5.90 (11)	–	–
earth8	8	8	8	0	65	yes	b	3.37 (11)	2.33 (11)	1.88 (11)	1.61 (11)	1.51 (11)	1.59 (34)
fourlayers8	8	4	4	0	60	yes	c	2.64 (11)	2.02 (11)	1.59 (11)	1.07 (11)	–	1.04 (34)
earth8a	8	8	8	1	60	yes	a	3.29 (11)	2.37 (11)	1.67 (11)	1.10 (11)	1.11 (13)	1.05 (34)
earth8a	8	8	8	0	65	yes	a	3.88 (11)	2.57 (11)	1.81 (11)	1.47 (11)	1.62 (13)	1.50 (34)
earth8a	8	8	8	0	60	yes	a	3.29 (11)	1.94 (11)	1.18 (11)	0.92 (11)	–	1.00 (34)
fourlayers8a	8	4	4	0	60	yes	d	2.94 (11)	1.94 (11)	1.24 (11)	1.00 (11)	–	1.04 (34)
earth4a	4	8	8	0	60	yes	a	3.29 (13)	2.52 (13)	1.84 (13)	1.36 (13)	1.26 (13)	1.23 (13)
uppermantle8a	8	8	4	0	60	yes	a	4.01 (13)	2.70 (13)	1.65 (13)	1.17 (13)	1.16 (13)	1.10 (13)
fourlayers8a	8	8	4	1	60	yes	d	2.94 (13)	2.76 (13)	1.96 (13)	1.47 (13)	1.30 (13)	–
earth8a	8	8	8	1	60	no	a	3.70 (13)	2.62 (13)	1.88 (13)	1.04 (13)	–	–
transition8	8	8	6	0	60	yes	e	3.29 (13)	1.98 (13)	1.51 (13)	1.11 (13)	–	–

Table 8.1: This table summarizes the computational experiments performed with the *Fujii and Schultz c* response data and described in this Chapter. The values corresponding to a particular value of μ are the minimum least squares data misfits obtained subject to minimising the penalty functional with the model norm weighted by μ. Each new optimisation has been performed starting from the best fitting model of the previous iterations. The numbers in brackets indicate the number of frequencies for which the optimisation has been performed.

responding to the damping parameter $\mu = 0.01$, as well as a regularised solution that has the misfit of 1.47 with the c responses (Figure 8.10). Note that the only difference between the prior information of the experiments in Figures 8.8 and 8.9 has been the electrical conductivity of the transition zone layer, $410 - 670$ km depths.

We see many similarities between the two models, mostly in the lower mantle. The upper mantle, however, is conceptually different between the two models, demonstrating that both a highly inhomogeneous and conductive transition zone with a relatively resistive uppermost mantle layer, and a totally resistive transition zone with inhomogeneous uppermost mantle both fit the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data set. Having said that, we do see several common features in the upper mantle between these models. There is a large resistor beneath the South-West Africa, as well as relatively small conductive anomalies beneath South-East parts of South America and Australia, and beneath Iceland, present is all these models. The presence of the conductors is not so clear in Figure 8.10, since features of that size have been strongly penalised in this inversion.

8.3.2   Experiments with 8 inhomogeneous layers degree and order 4

Two experiments have been performed with the prior models (a), Figure 8.1, and (b), Figure 8.2, allowing inhomogeneities up to the degree and order 4. We have found that we have not been able to quite fit the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data set of c responses with this degree and order, with either of the priors. Here, we present the two solutions obtained with $\mu = 0.001$, fitting the c response data set with the least squares misfit of 1.26 and 1.23, respectively (Figures 8.11 and 8.12). Even relaxing the model norm restriction to $\mu = 0.001$ from the usual $\mu = 0.01$ has not allowed us to fit the data statistically well enough.
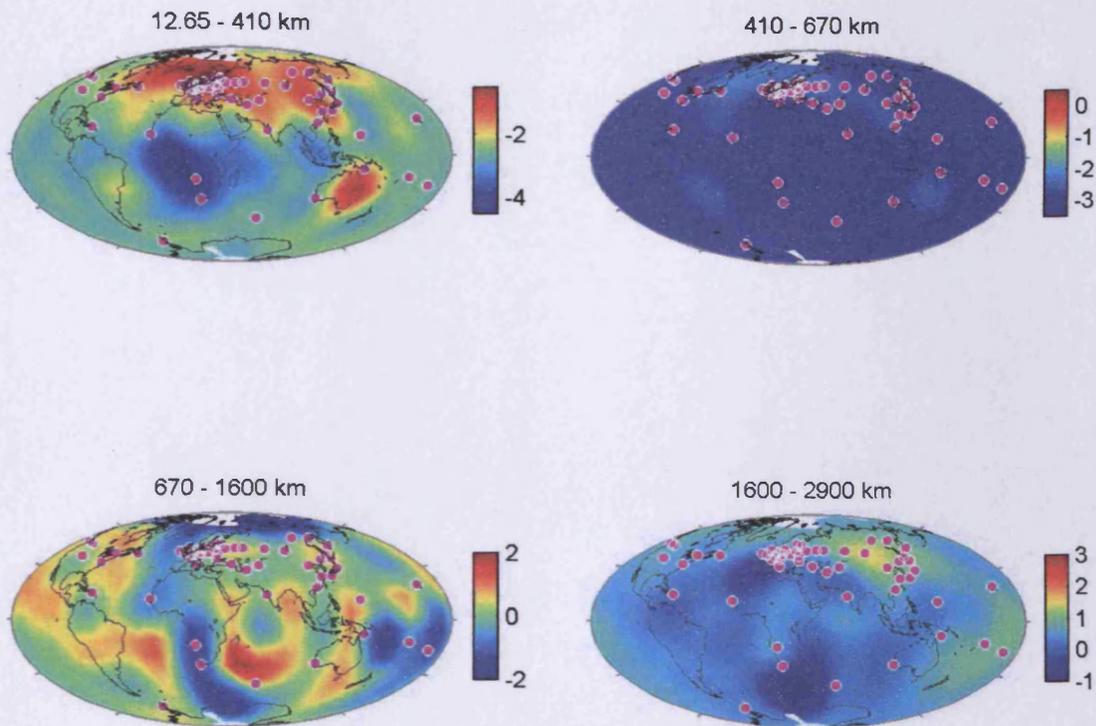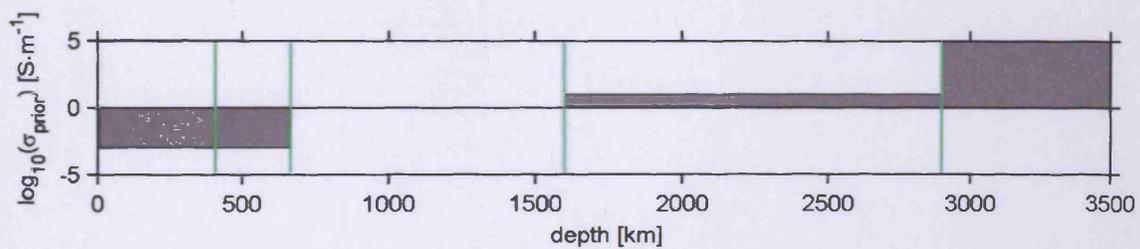
Figure 8.8: Four-layer inhomogeneous mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of resistive transition zone. Least squares misfit with the data: 1.07.
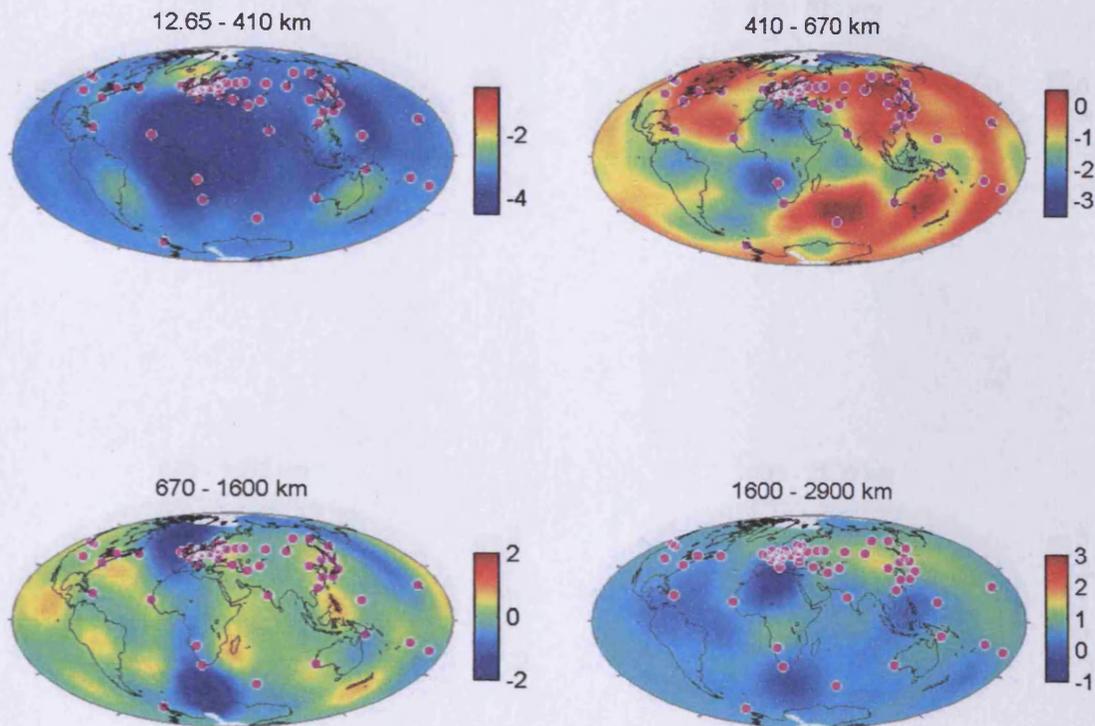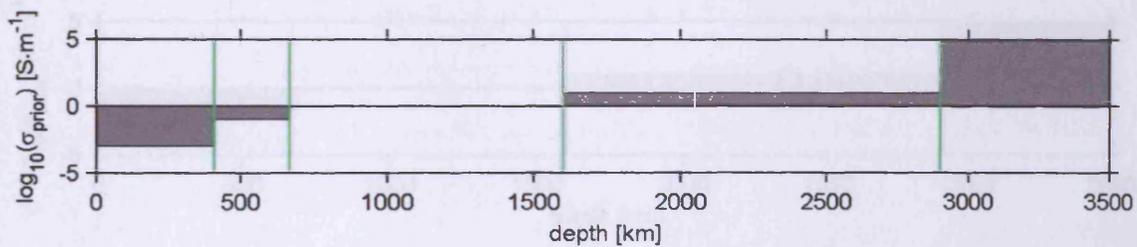
Figure 8.9: Four-layer inhomogeneous mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of conductive transition zone. Least squares misfit with the data: 1.00.
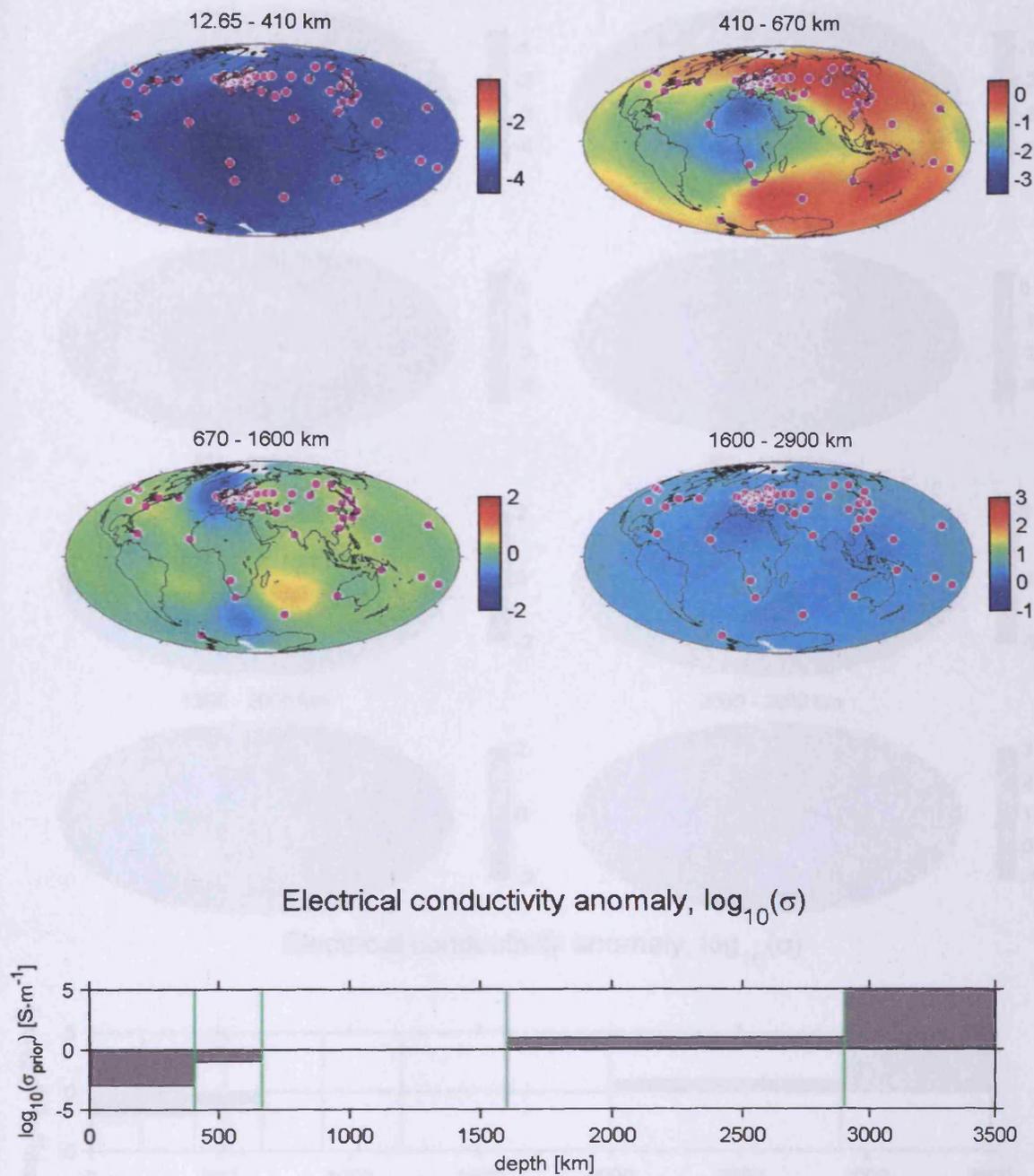
Figure 8.10: Four-layer inhomogeneous mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of conductive transition zone, penalising higher order terms by setting $\alpha = 1$. Least squares misfit with the data: 1.47.
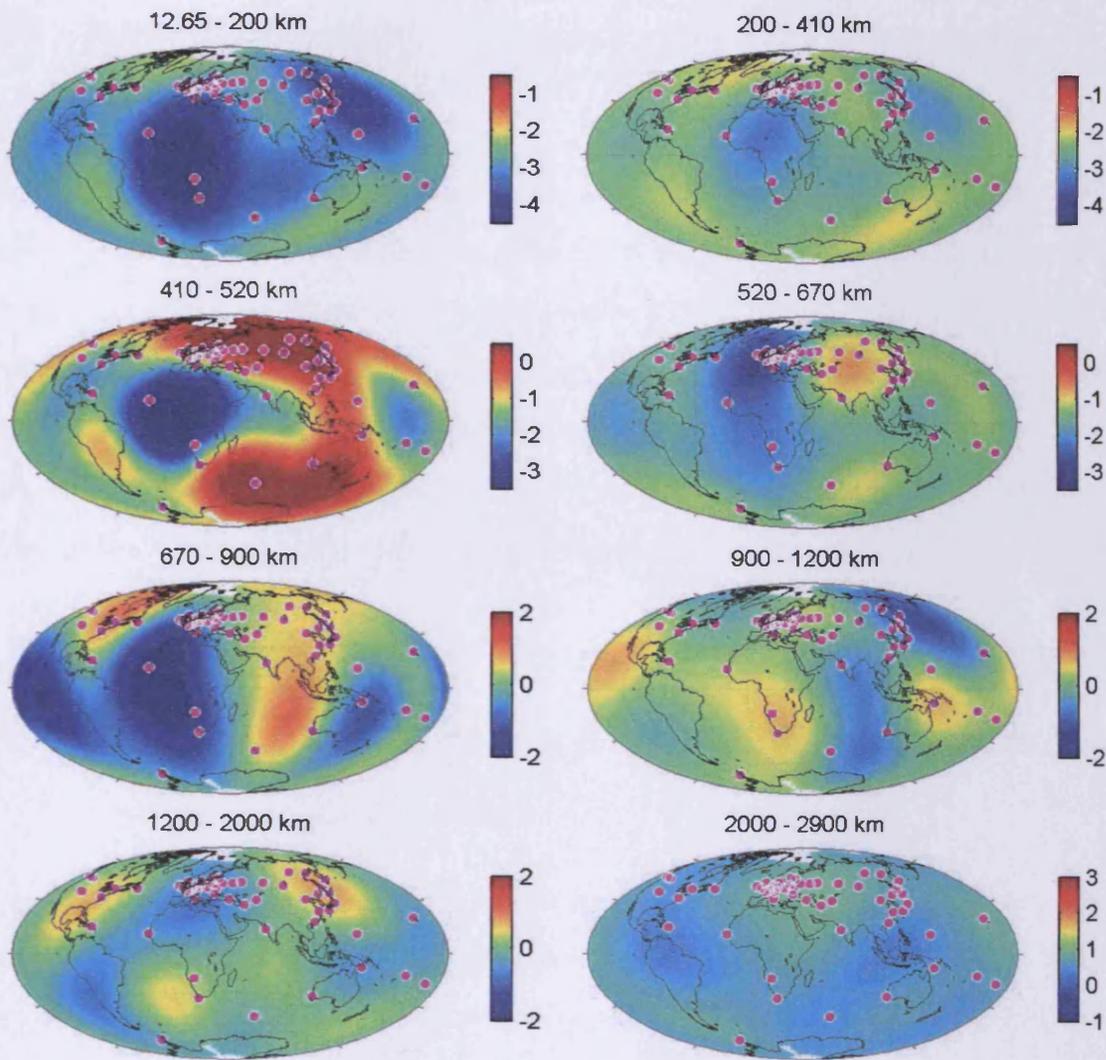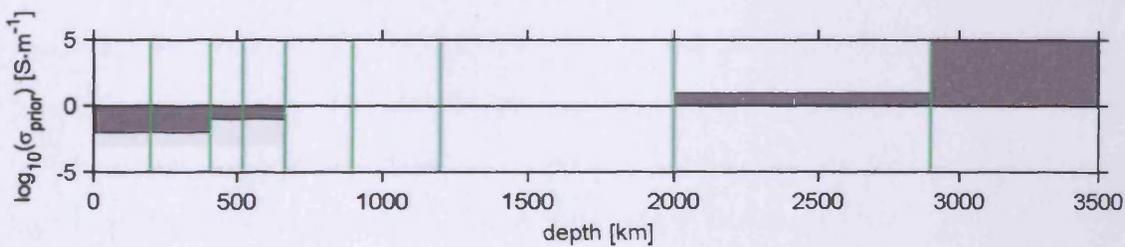
Figure 8.11: Eight-layer inhomogeneous mantle of degree and order 4, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.001$ under the assumption of conductive upper mantle. Least squares misfit with the data: 1.26.
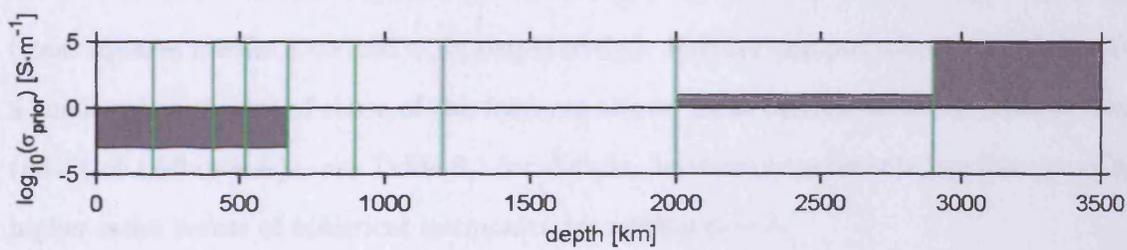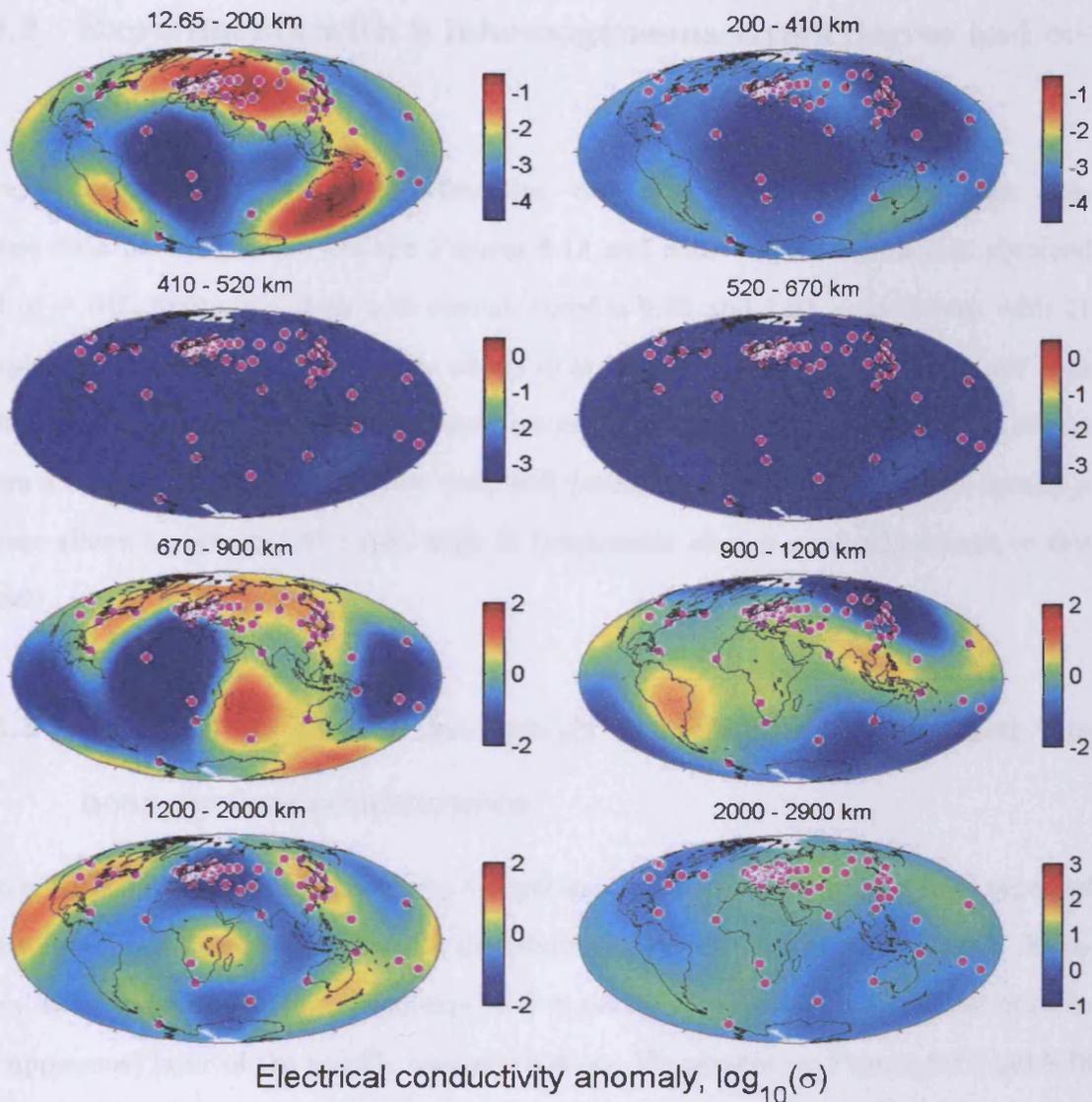
Figure 8.12: Eight-layer inhomogeneous mantle of degree and order 4, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.001$ under the assumption of resistive upper mantle. Least squares misfit with the data: 1.23.

8.3.3 Experiments with 8 inhomogeneous layers degree and order 8

Here, we present a set of models fitting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz c* response data perfectly well. On the Figures 8.13 and 8.15 we plot the models obtained with $\mu = 0.01$ fitting the data well enough (misfits 0.92 and 1.02, respectively, with 11 sample frequencies). Setting μ to zero allows to obtain perfect fit with the data (1.00 with 34 frequencies) after a small enhancement in some of the model features. Figure 8.16 also shows a regularised model fitting the data well (misfit 1.10 with 11 frequencies; setting μ to zero allows to obtain 1.05 misfit with 34 frequencies after a small adjustment to this model).

8.3.4 Eight-layer degree and order 8 experiments without the near-surface conductance

Two of the eight-layer degree and order 8 experiments described above have been repeated without the near-surface conductance distribution imposed on top of the model. Effectively, these experiments have been expected to partly reconstruct oceans distribution in the uppermost layer of the mantle parametrization. We present on Figures 8.17 and 8.18 the two solution models obtained for these experiments, that fit the data well enough (least squares misfits 1.04 and 1.10, respectively). Further computations have shown that a small enhancement of some of the features allows these models to fit the data perfectly (misfit of 1.00 or less) – see Table 8.1 for details. In these experiments, we have penalised higher order terms of spherical harmonics by setting $\alpha = 1$.

Notably, while we have not been able to quite fit the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data set with the prior model b when we regularised with $\alpha = 1$ in Section 8.3.3
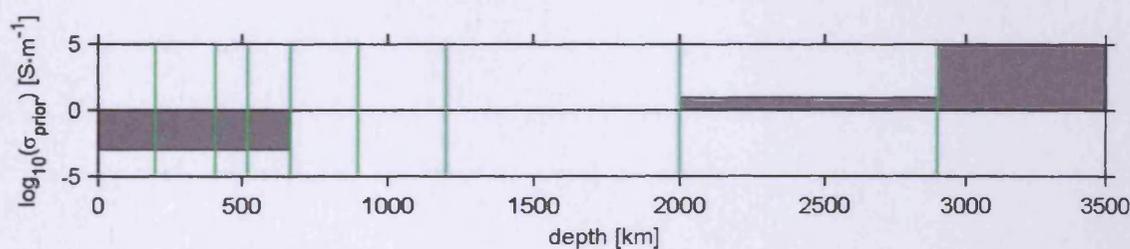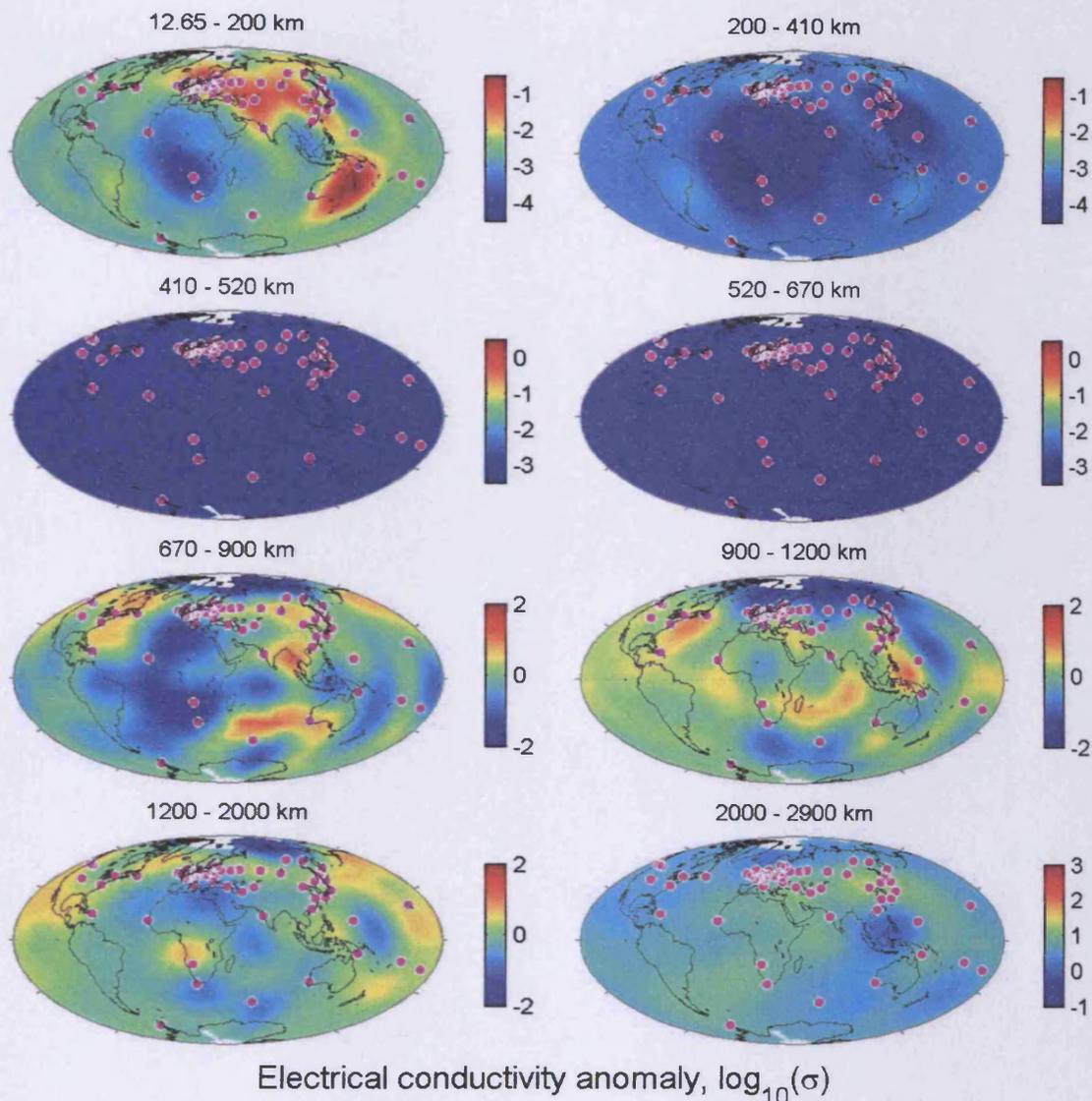
Figure 8.13: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of a resistive upper mantle. Least squares misfit with the data: 1.02.
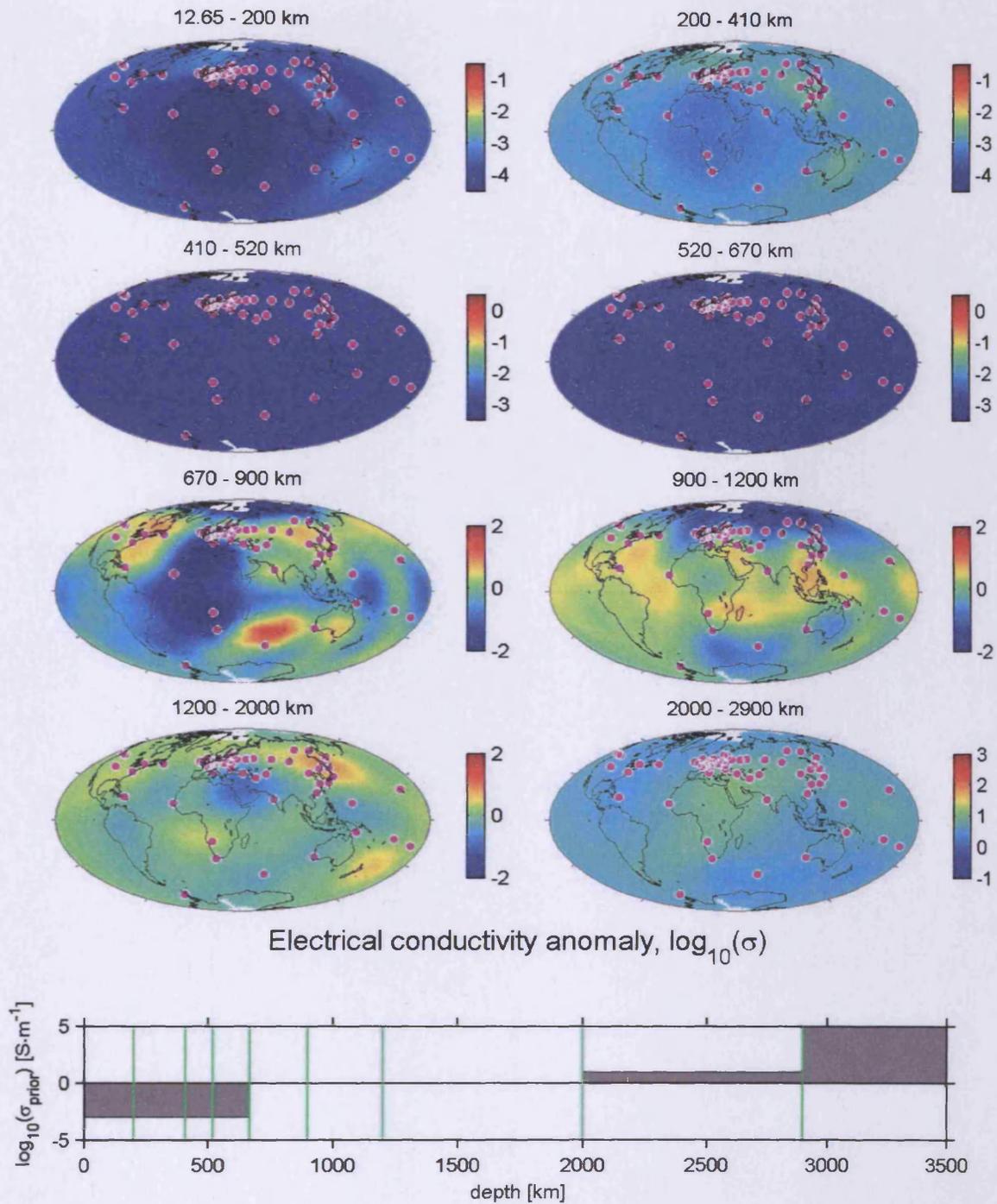
Figure 8.14: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of a resistive upper mantle, penalising higher order terms by setting $\alpha = 1$. This model has then been further improved by setting $\mu = 0.001$. Least squares misfit with the data: 1.23.
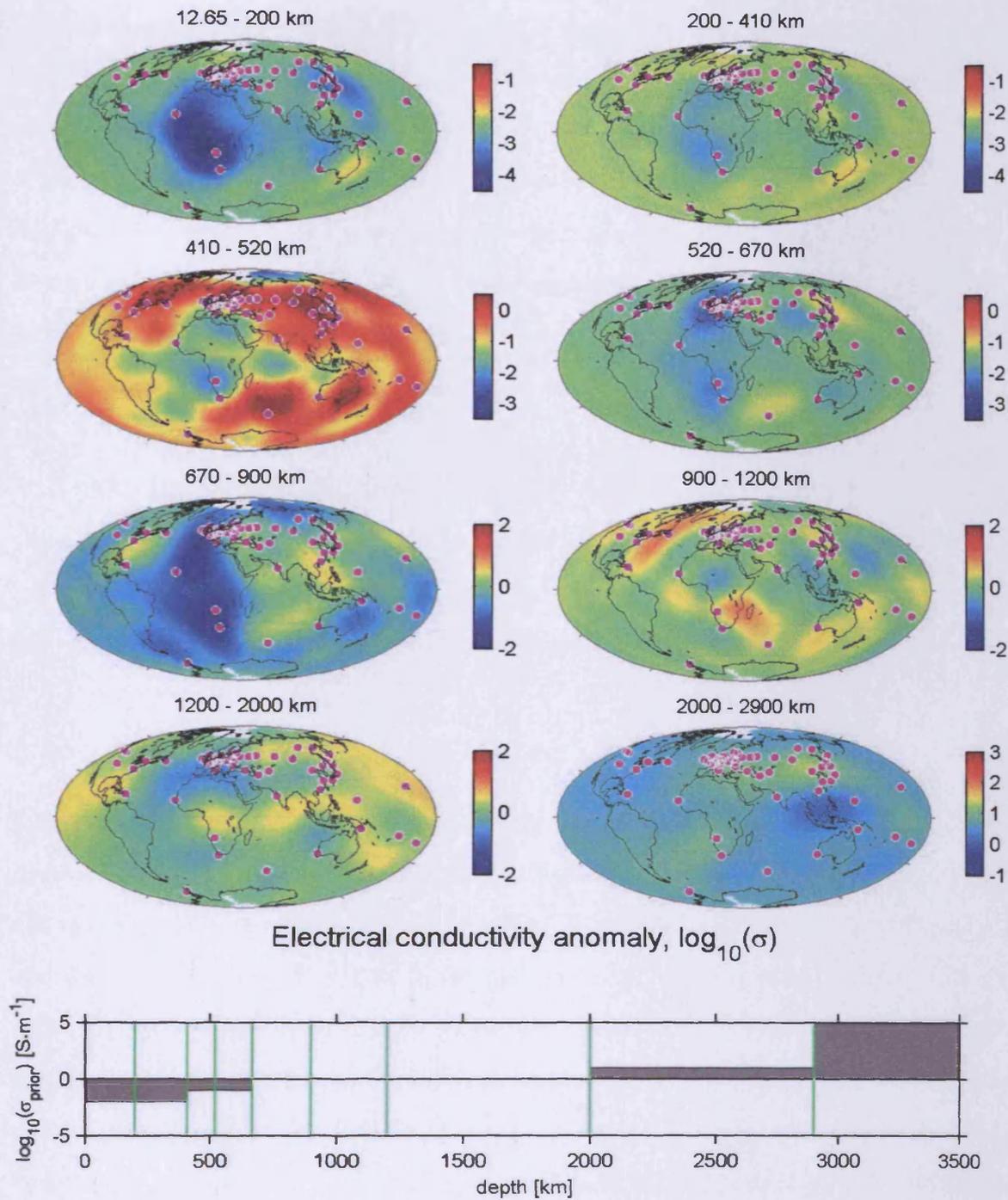
Figure 8.15: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of a conductive upper mantle. Least squares misfit with the data: 0.92.
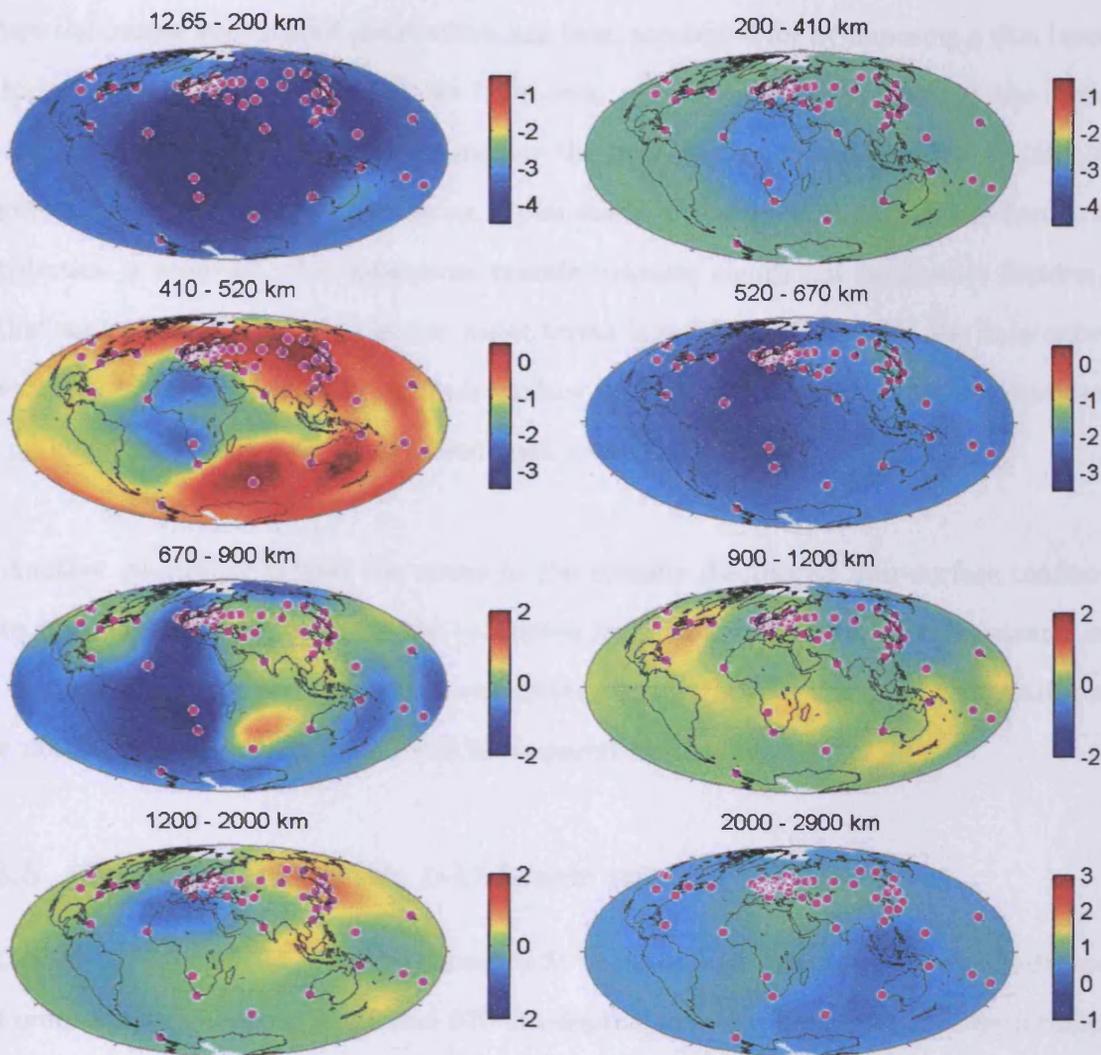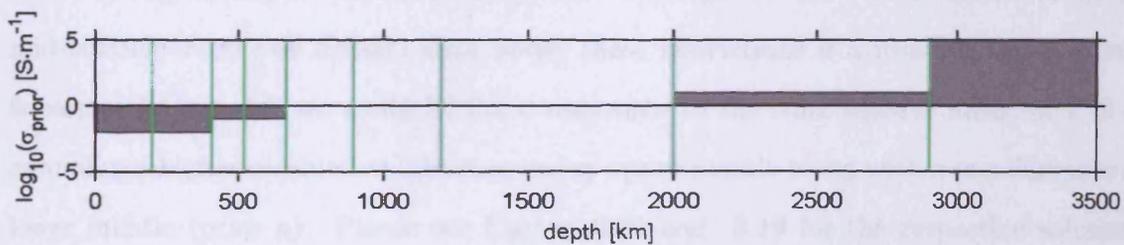
Figure 8.16: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of a conductive upper mantle, penalising higher order terms by setting $\alpha = 1$. This model has then been further improved by setting $\mu = 0$. Least squares misfit with the data: 1.05.

(where the oceans versus land distribution has been accounted for by imposing a thin layer on top of the model, Figure 8.14), we have been able to successfully perform the same regularised inversion without accounting for the near-surface inhomogeneities explicitly (Figure 8.18). We suggest the following explanation. Once the near-surface conductance distribution is removed, the uppermost mantle contains significant small-scale features, so that an inversion penalising higher order terms is not being able to fit the data quite as well. However, if we include the near-surface into the uppermost layer, these features are no longer dominant and do not need such a careful modelling.

Another possibility is that the errors in the roughly discretized near-surface conductance introduce, in turn, errors in the computed responses; so that small-scale features in the uppermost mantle are required to extenuate these errors. In this case, these features may not be there, but they would still be required by the data.

8.3.5 Experiments with 1-D lower mantle

In this set of experiments, we were trying to fit the data with four inhomogeneous degree and order 8 layers between 12.65 and 670 km depths, and four homogeneous layers below 670 km. Two prior models have been used for this purpose: (a) and (b) (see Figures 8.1 and 8.2, respectively). We have found that although we have been unable to fit the mid-latitude *Fujii and Schultz* data under these restrictions if a resistive upper mantle is assumed (prior b), we could fit the c responses to the least squares misfit of 1.10 by assuming a highly conductive inhomogeneous upper mantle along with a one-dimensional lower mantle (prior a). Please see Figures 8.20 and 8.19 for the respective solutions, obtained with $\mu = 0.01$ and fitting the mid-latitude data with the least squares misfits 1.17 and 2.22, respectively. We can see that in spite of all the differences between the two models, the large resistive anomaly beneath the South-West Africa and two conductive
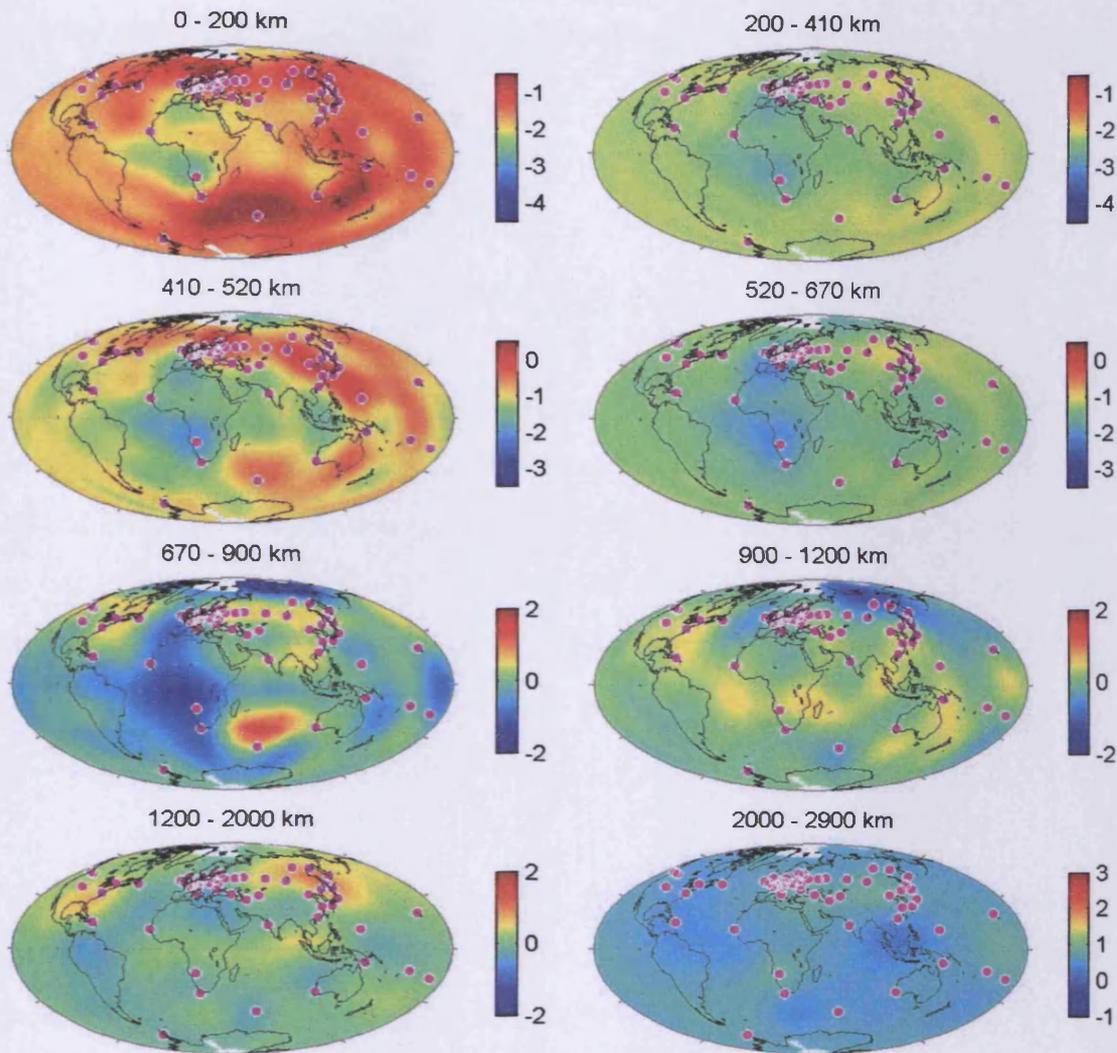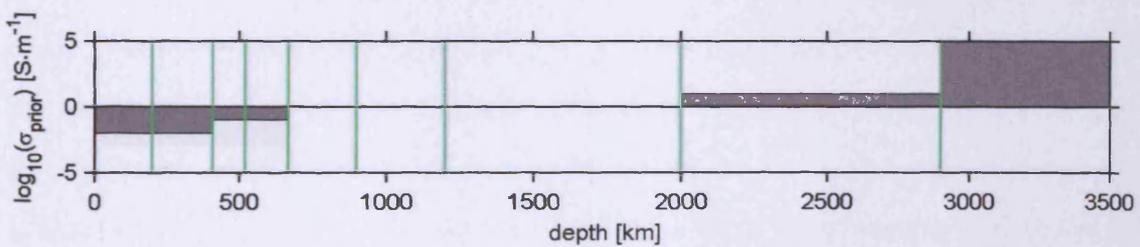
Figure 8.17: Eight-layer inhomogeneous mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of conductive upper mantle, penalising higher order terms by setting $\alpha = 1$. No near-surface conductance has been imposed on top of this model, hence the uppermost layer compensates for the ocean versus land distribution. Least squares misfit with the data: 1.04.
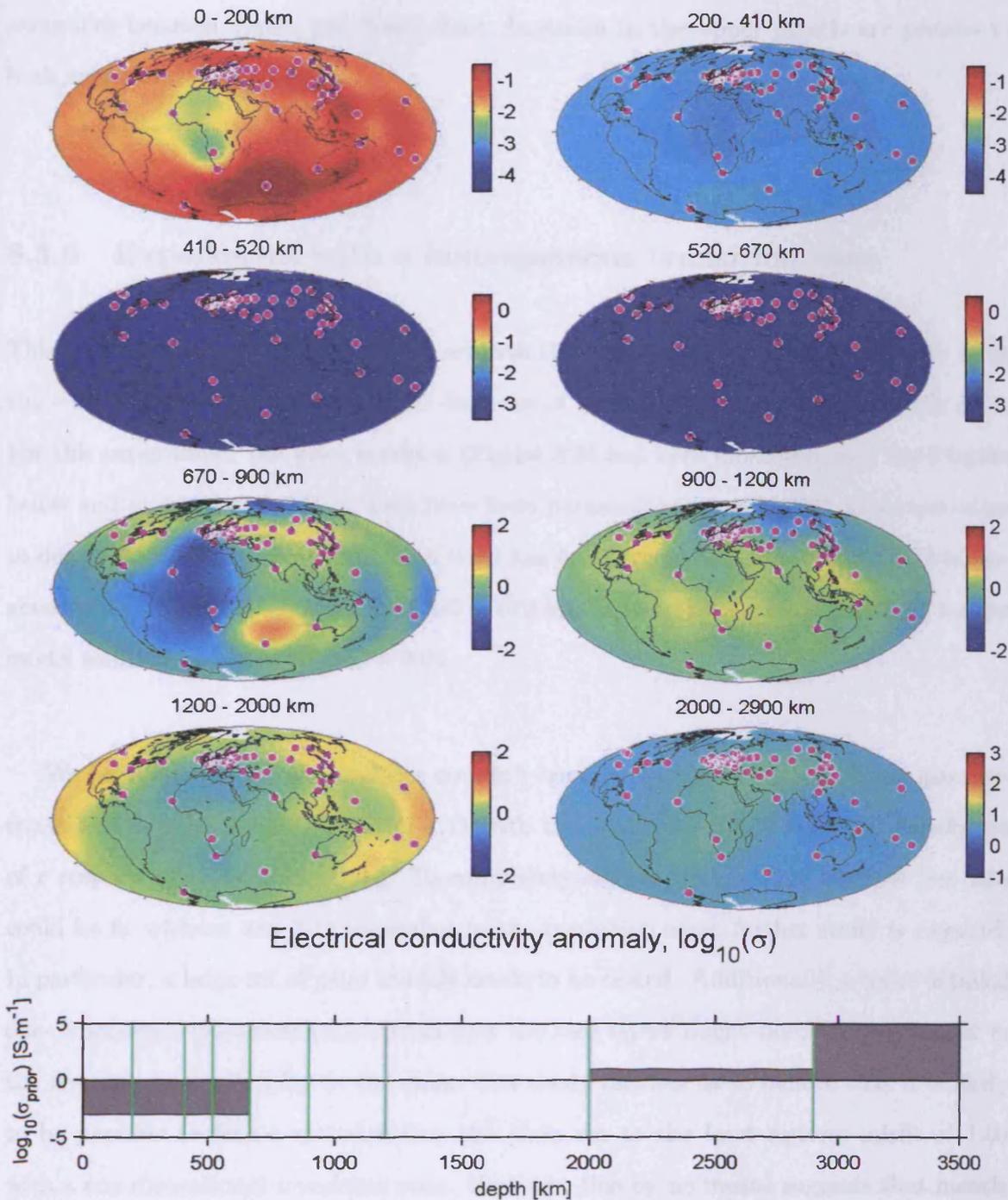
Figure 8.18: Eight-layer inhomogeneous mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of resistive upper mantle, penalising higher order terms by setting $\alpha = 1$. No near-surface conductance has been imposed on top of this model, hence the uppermost layer compensates for the ocean versus land distribution. Least squares misfit with the data: 1.10.

anomalies beneath Japan and South-East Australia in the upper mantle are present in both solutions.

8.3.6 Experiment with a homogeneous transition zone

This experiment has been designed to address the question of whether it is possible to fit the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data set of c responses with a 1-D transition zone. For this experiment, the prior model **e** (Figure 8.5) has been employed, and the 6 layers below and above the transition zone have been parametrized to allow inhomogeneities up to degree and order 8. The transition zone has been parametrized to consist of 2 homogeneous layers at $410 - 520$ km and $520 - 670$ km depths. Please see Figure 8.21 for the model solution obtained with $\mu = 0.01$.

We have found that although we couldn't completely fit the data with this parametrization, the least squares misfit of 1.11 with the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses has been obtained. To completely answer the question whether the data could be fit without any 3-D anomalies in the transition zone, further study is required. In particular, a large set of prior models needs to be tested. Additionally, a more detailed one-dimensional parametrization than just the two layers might need to be imposed in the transition zone to fully fit the data. Our study inclines us to believe that it is likely to be possible to find a model fitting this data set to the least squares misfit of 1.00 with a one-dimensional transition zone. However, this by no means suggests that mantle transition zone is not heterogeneous. In order to establish that, better vertical resolution is required than that provided by this data set. This can only be achieved if additional data are obtained.
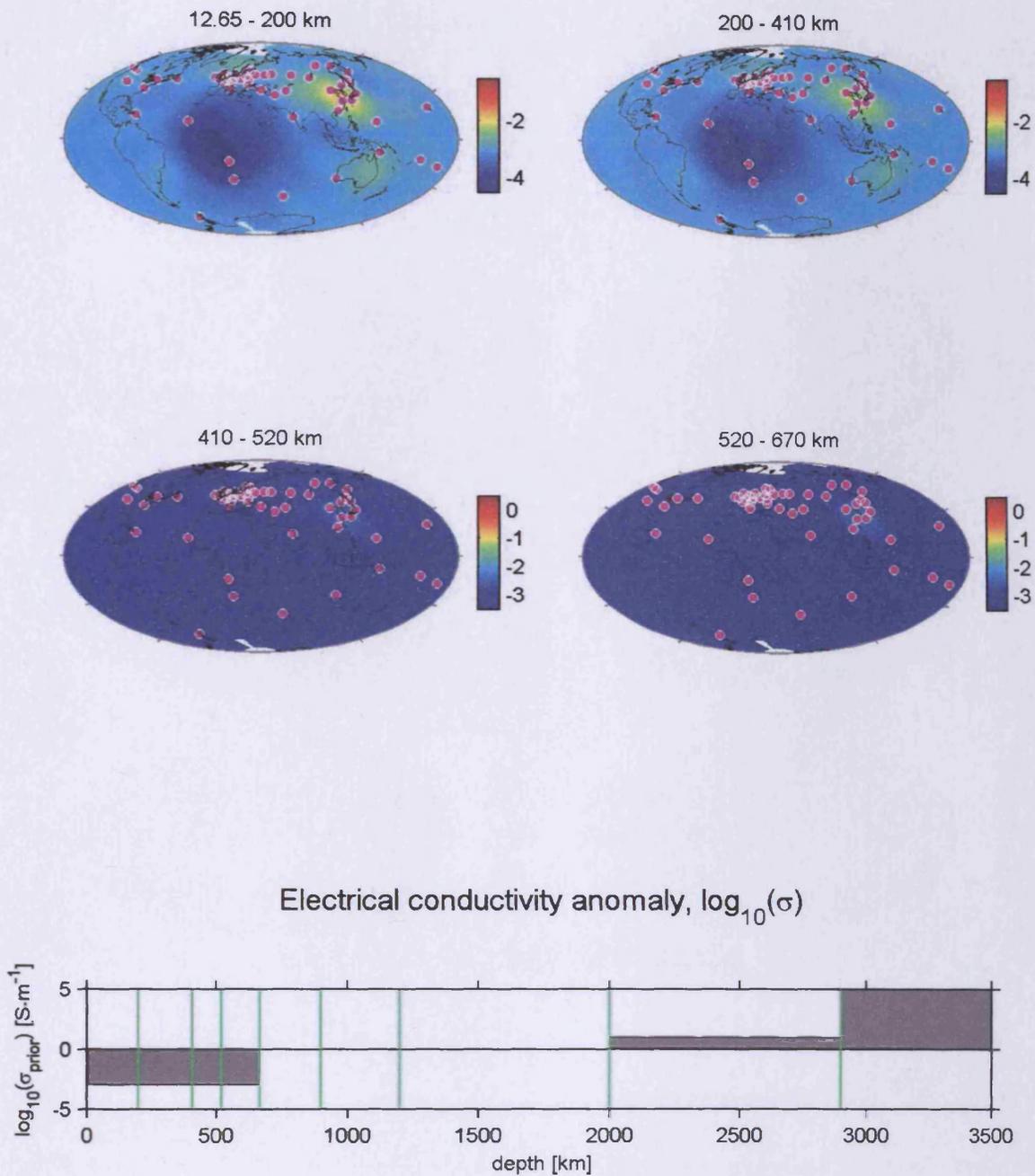
Figure 8.19: Inhomogeneous upper mantle above a one-dimensional lower mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of resistive upper mantle. Least squares misfit with the data: 2.22. The respective perturbations around the prior $\log_{10} \sigma$ in the four homogeneous layers below 670 km are: -0.44, 0.35, -0.15, -0.07.
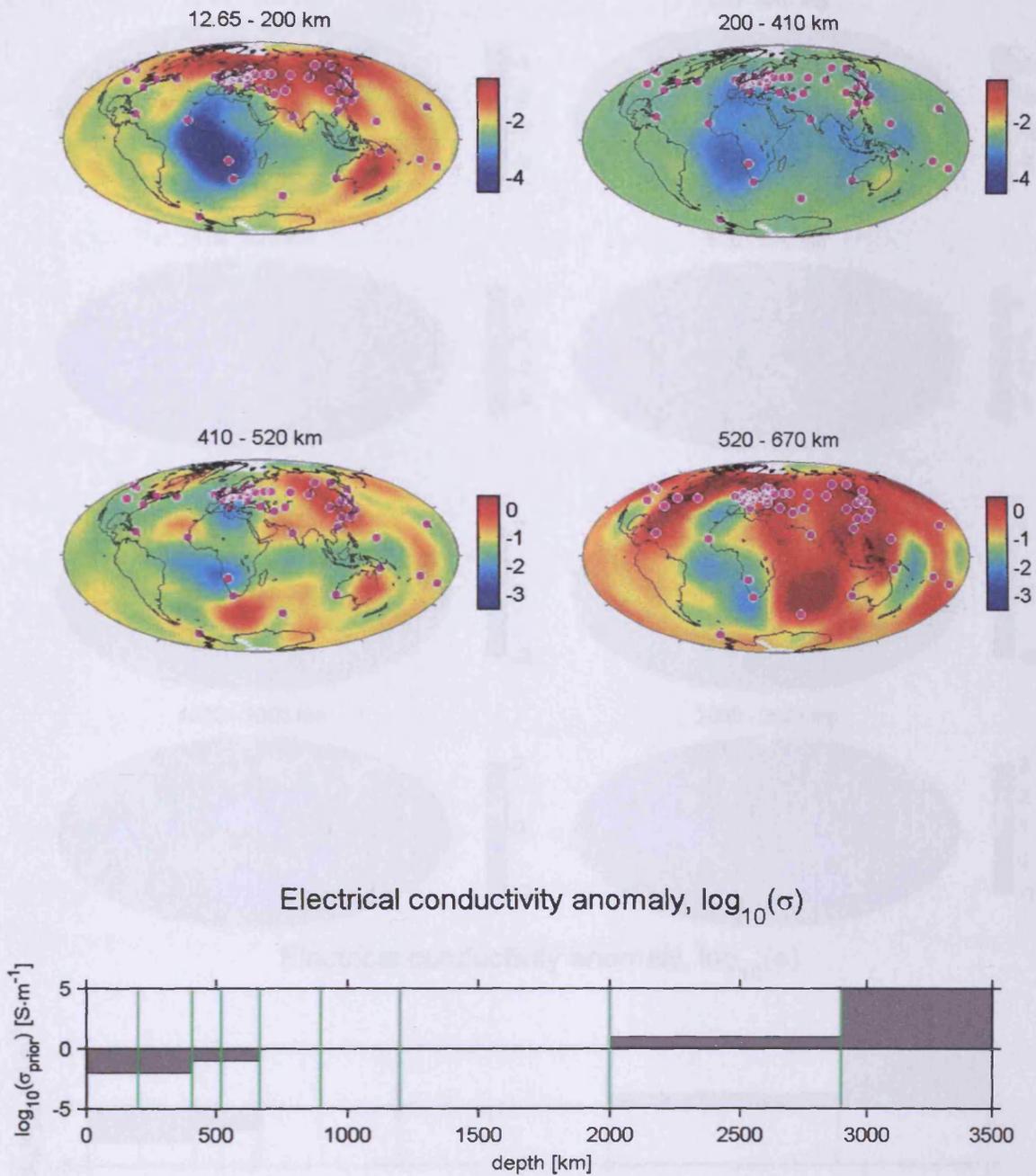
Figure 8.20: Inhomogeneous upper mantle above a one-dimensional lower mantle, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of conductive upper mantle. Least squares misfit with the data: 1.17. The respective perturbations around the prior $\log_{10} \sigma$ in the four homogeneous layers below 670 km are: -0.93, 0.35, -0.32, -0.13.
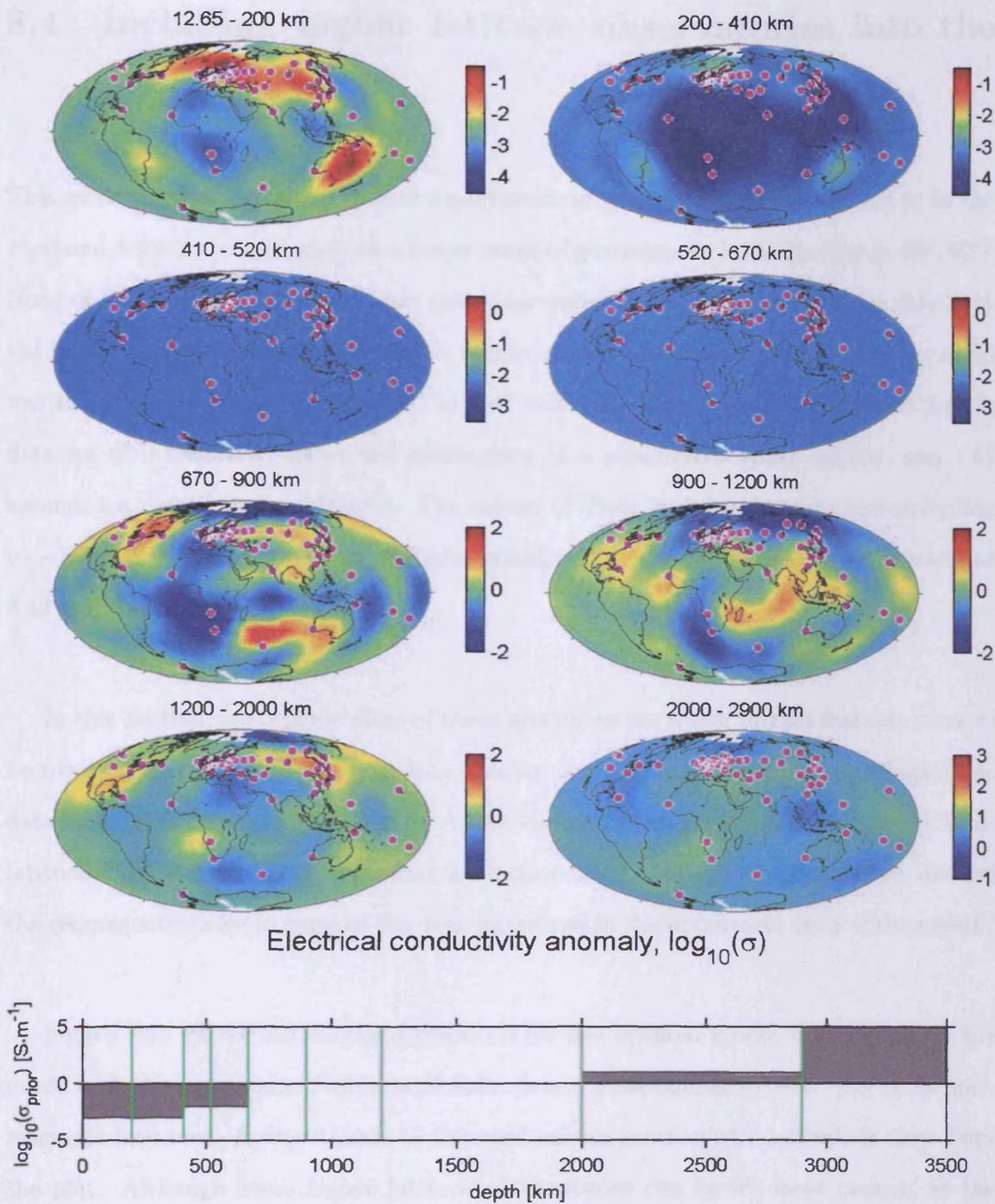
Figure 8.21: An eight-layer degree and order 8 model with a 1-D transition zone, obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$. Least squares misfit with the data: 1.11. The respective perturbations around the prior $\log_{10} \sigma$ in the two homogeneous layers at $410 - 510$ km and $520 - 670$ km depths are: 0.26 and 0.23.

8.4   Including higher latitude observatories into the analysis

This section describes several inverse experiments in which we have been trying to fit the *Fujii and Schultz* c responses from a larger range of geomagnetic latitudes than $[-60°, 60°]$. None of these experiments have been quite successful in that we have not been able to fit the higher latitude observatory data to within statistical bounds. The best we could do was to obtain the least squares misfit of 1.47 with the $-65° \leq \theta \leq 65°$ *Fujii and Schultz* data set of c responses under the assumption of a conductive upper mantle, and 1.61 assuming a resistive upper mantle. The subset of *Fujii and Schultz* data corresponding to $-75° \leq \theta \leq 75°$ geomagnetic latitudes could only be fit to the least squares misfit of 3.42 (see Table 8.1).

In this Section, we provide plots of these models to see which further features seem to be required by the higher latitude data compared to the models fitting the mid-latitude data well (Figures 8.23, 8.24, 8.22). As we can see from these plots, additional high-latitude data seems to require further anomalies, both conductive and resistive, around the geomagnetic poles in some of the deep layers and in the uppermost layer of the mantle.

Figure 8.25 shows the averaged residuals for the optimal model that minimises the misfit with the c response *Fujii and Schultz* data corresponding to $-75° \leq \theta \leq 75°$ geomagnetic latitudes. A dependence of the residuals on geomagnetic latitude is clear from the plot. Although some higher latitude observatories can be fit, most cannot; at the same time, the model satisfies the data from all the observatories between the latitudes of $-60°$ and $60°$ reasonably well. This suggests an explanation for not being able to fit the expanded *Fujii and Schultz* data set: quite possibly some of the higher latitude data
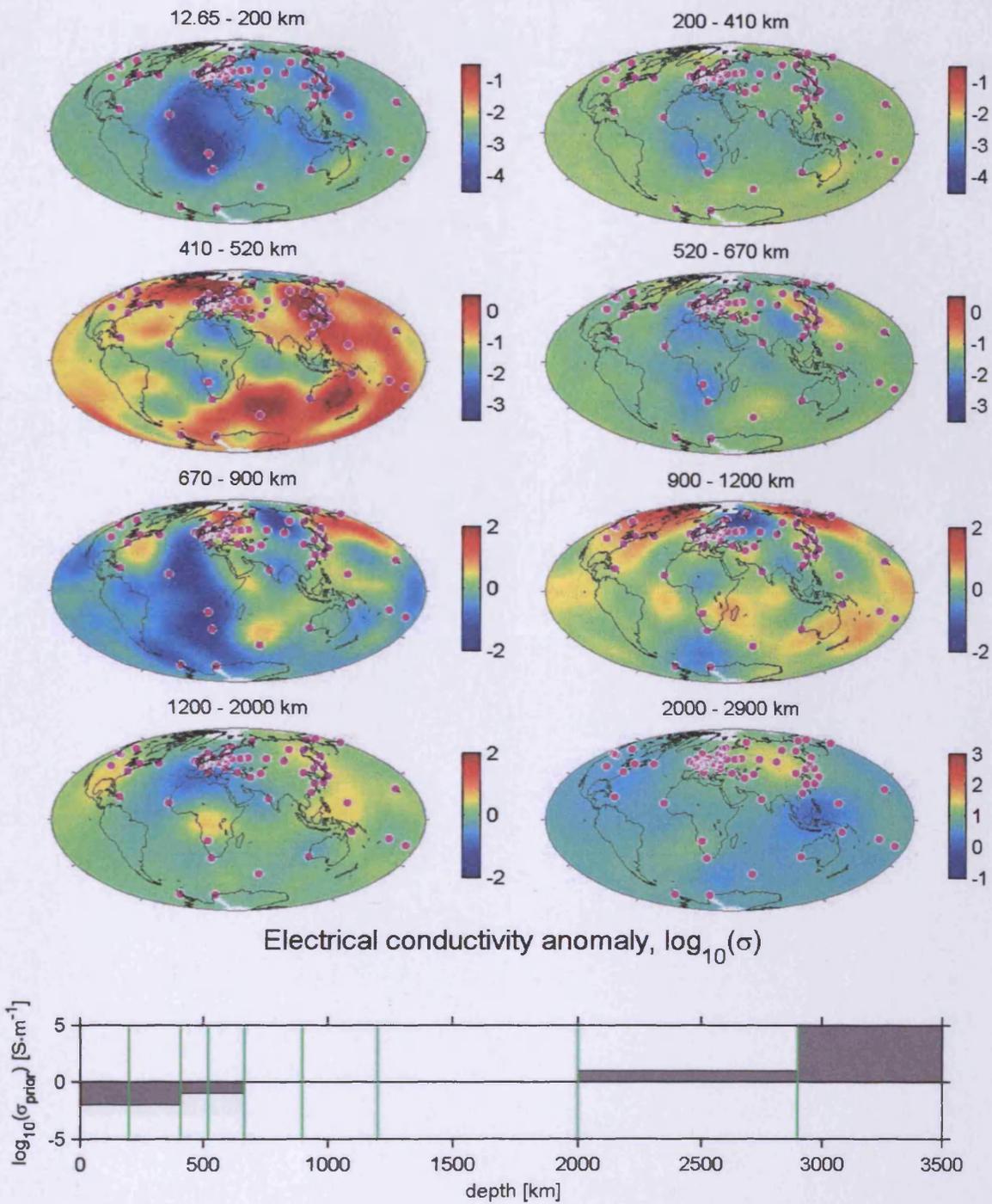
Figure 8.22: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-65° \leq \theta \leq 65°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of a conductive upper mantle. Least squares misfit with the data: 1.47.
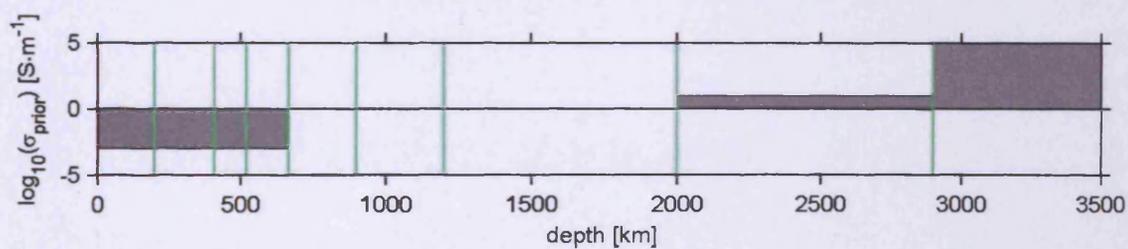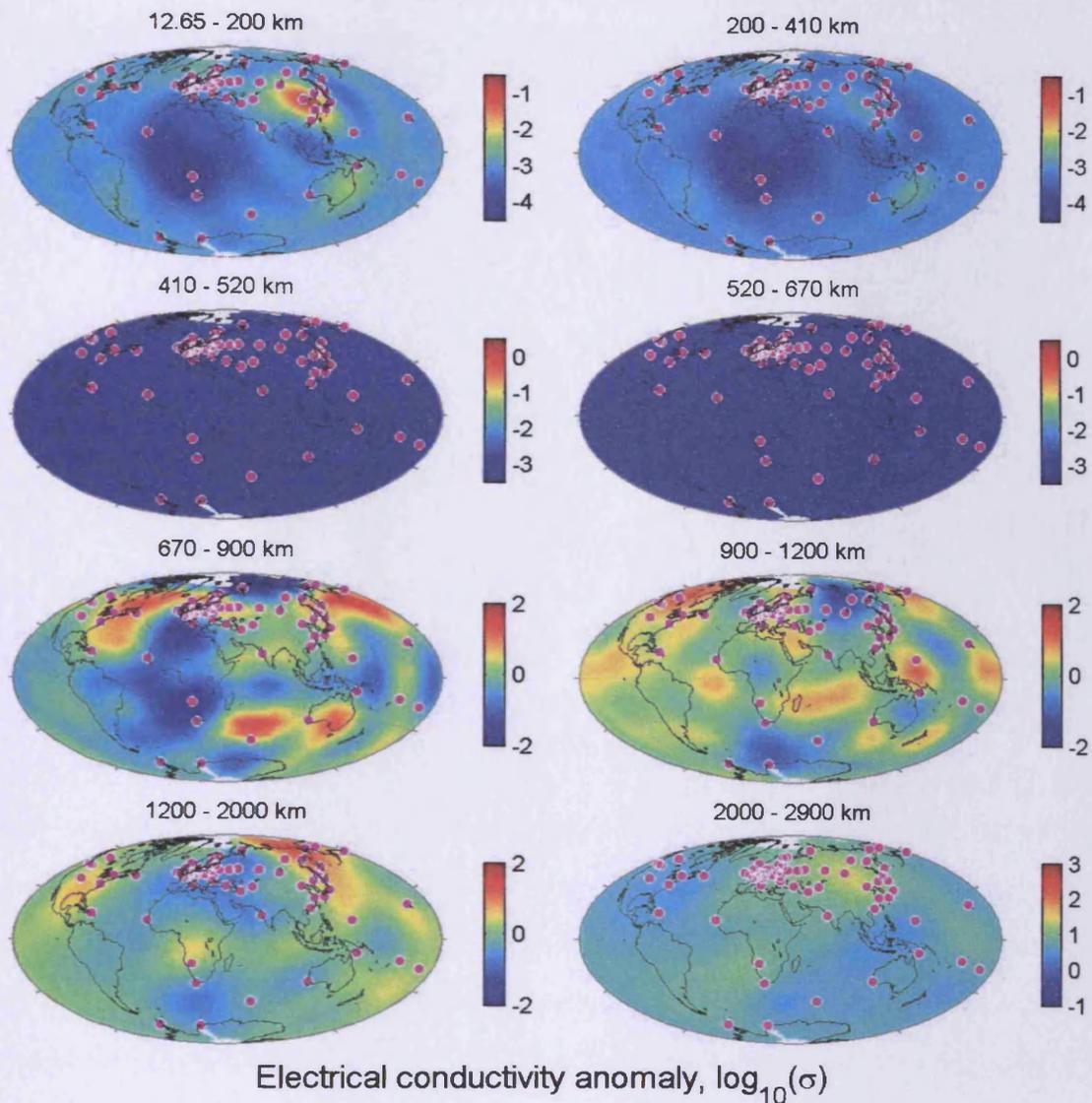
Figure 8.23: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-65° \leq \theta \leq 65°$ *Fujii and Schultz* set of *c* responses with the damping parameter $\mu = 0.01$ under the assumption of a resistive upper mantle. Least squares misfit with the data: 1.61.
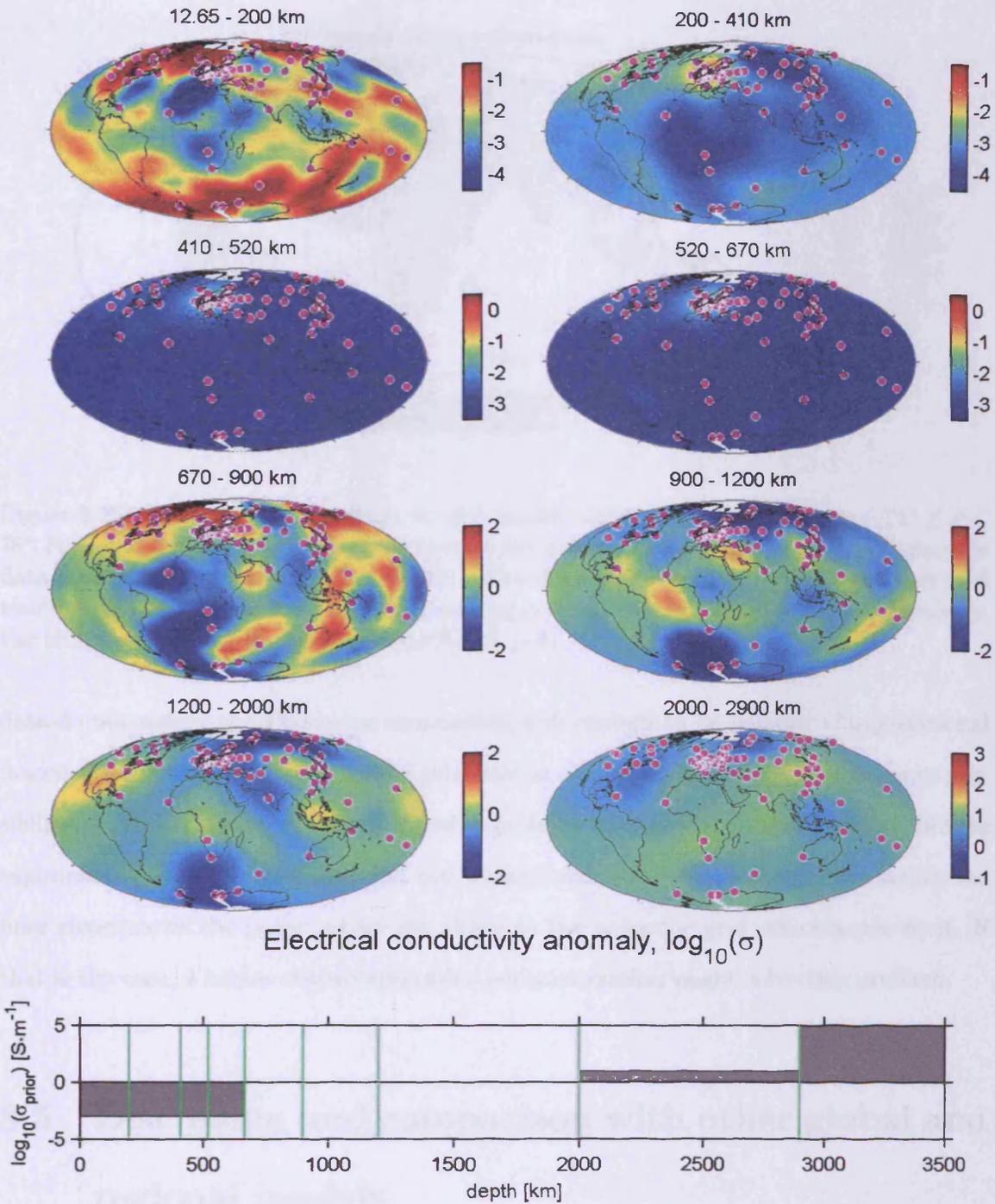
Figure 8.24: Eight-layer inhomogeneous mantle degree and order 8, obtained by inverting the $-75° \leq \theta \leq 75°$ *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ under the assumption of a resistive upper mantle. Least squares misfit with the data: 3.42.

Residuals, averaged over 34 frequencies
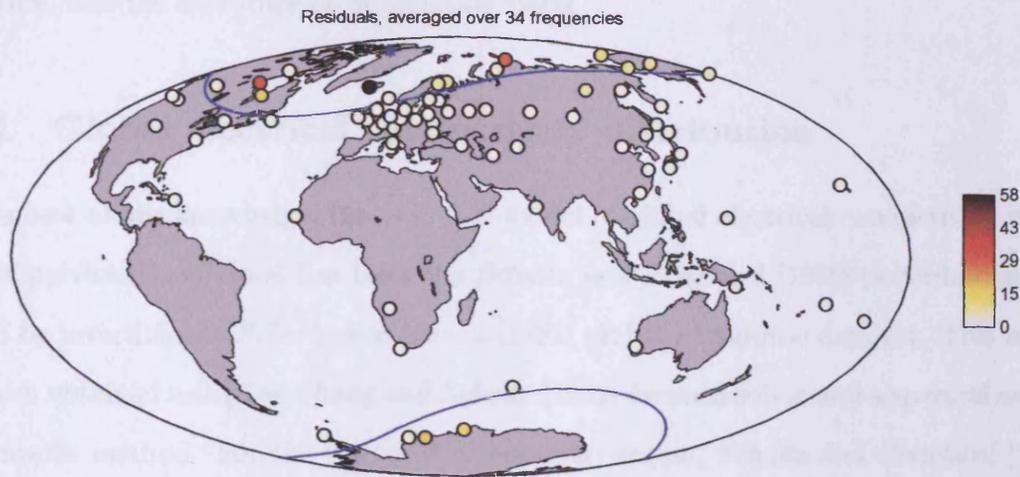
Figure 8.25: The averaged residuals for the model obtained by inverting the $-75° \leq \theta \leq$ 75° *Fujii and Schultz* set of c responses with the damping parameter $\mu = 0.01$ (fitting the data with a least squares misfit of 3.42). The colour denotes the value of the averaged residual; the two blue lines denote the geomagnetic latitudes $-60°$ and $60°$, respectively; the blue star denotes the geomagnetic North pole.

data do not satisfy the P_1^0 source assumption well enough to be fittable. Computational inaccuracies gathering around the singularities at the geomagnetic poles is one more possibility; however, we have not witnessed a problem of this sort in the synthetic inverse experiments. It is also possible that our parametrization is too restrictive to satisfy the finer structure at the poles: as we get closer to the poles the grid cells become finer. If that is the case, a higher degree and order parametrization might solve this problem.

8.5 Discussion and comparison with other global and regional models

Very few 3-D models of electrical conductivity in the mantle exist to date. In this section, we compare them to our models. We also include a discussion of the relevant seismic

literature, and the literature on subducting slabs.

8.5.1 Global electrical conductivity distribution

To the best of our knowledge, the only 3-D model of global electrical conductivity in the mantle previously obtained has been the *Schultz and Pritchard* [1999] perturbation, obtained by inverting the *Schultz and Larsen* [1987] global c response data set. This model has been obtained using the *Zhang and Schultz* [1992] forward solver and a spectral expansion inverse method. For the purposes of parametrization, *Schultz and Pritchard* [1999] have used the assumption that the spatial variations of conductivity within the Earth are smooth; and the inverse model has been obtained as a perturbation expansion around the prior. This 4-layer inverse model is presented on Figure 8.26 (left) as a percentage perturbation about the prior electrical conductivity values (Figure 8.26, bottom left). On the contrary, our inverse models for the experiments fourlayers8a with $\alpha = 1$ and $\alpha = 0$, respectively (Section 8.3.1) are shown on Figure 8.26 as $\log_{10}$ perturbations, so that the heterogeneities are much stronger. Compensation for the S-effect has been attempted by including a fixed near-surface layer down to the 12.65 km depth in the Earth. The least squares misfit of the *Schultz and Pritchard* [1999] model with the *Schultz and Larsen* [1987] global data set has been 4.41. The least squares misfits of our models with the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* data set are 1.47 and 1.00, respectively.

The comparison of the two models is only possible qualitatively, rather than quantitatively, since the *Schultz and Pritchard* [1999] perturbation has been restricted, as mentioned above, to the class of smooth models. As far as we are aware, the authors of *Schultz and Pritchard* [1999] have not corrected for the S-effect, thus the reconstructed perturbation in the uppermost mantle is not expected to consist of the same features as the two fourlayers8a models. However, the regions of enhanced conductivity in the lower

Figure 8.26: We compare two of our 4-layer models to the perturbation given in *Schultz and Pritchard* [1999]. The *Schultz and Pritchard* [1999] model is presented as percentage perturbations about the prior electrical conductivity values. Our models are shown as $\log_{10}$ perturbations. The respective least squares misfits with the data are: 4.41 [*Schultz and Larsen*, 1987], 1.47 and 1.00 [*Fujii and Schultz*, 2002].
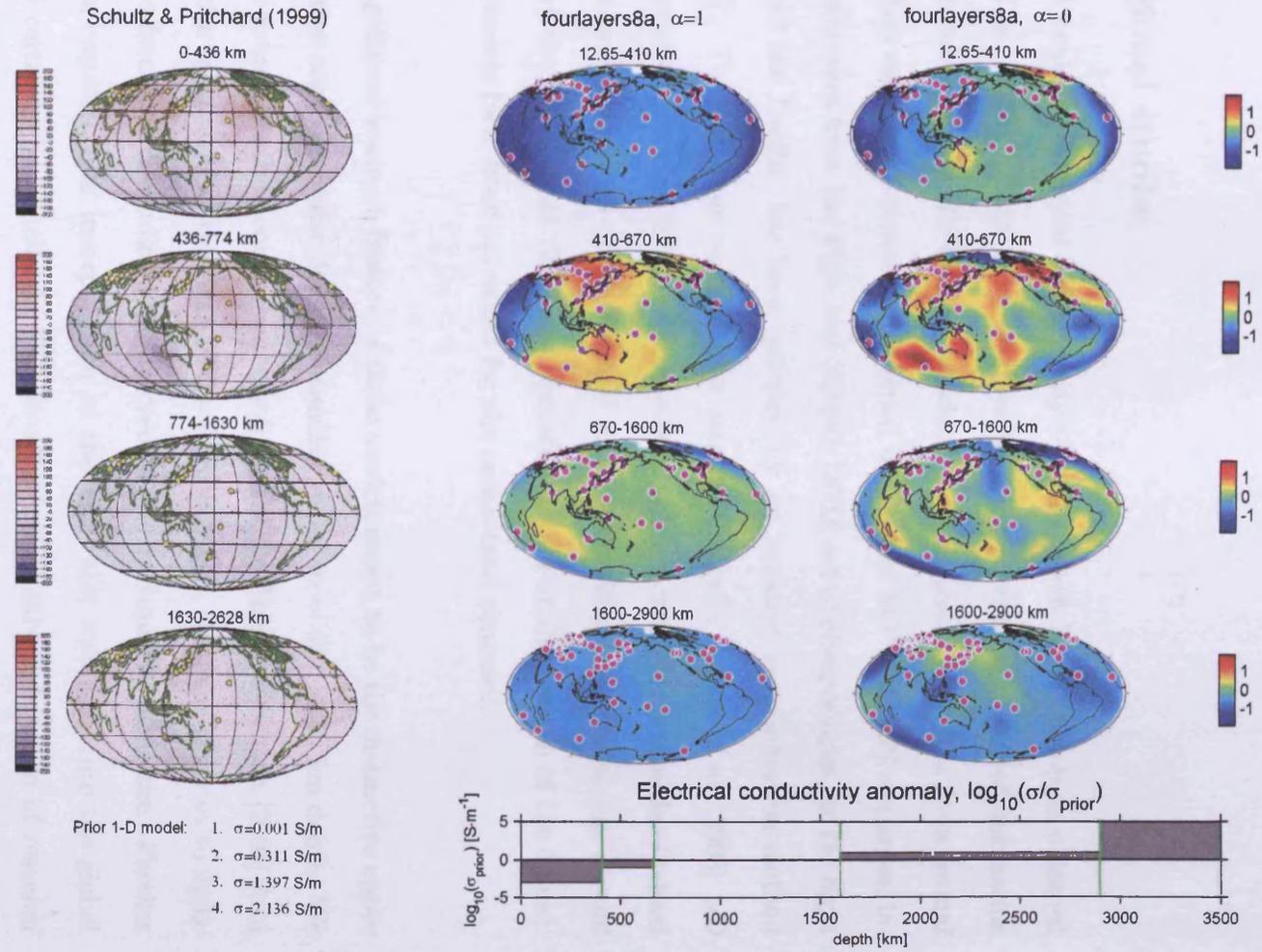
three layers, and especially layers 2 and 4, seem to very roughly correlate between these three models.

8.5.2 Regional studies

A regional 3-D model of electrical conductivity beneath North Pacific has been obtained by *Koyama et al.* [2003]. They have used voltage data from eight retired submarine communication cables across the Pacific, in addition to magnetic field data from several geomagnetic observatories on shore. A combined data set of MT and GDS responses, including the c responses from the *Fujii and Schultz* [2002] set of observatories, at the eight locations around the Pacific, has been inverted by an iterative approximation method (quasi-Newton). The resultant perturbation around the prior *Utada et al.* [2003] 1-D model is presented on Figure 8.27. It is compared to one of our 8-layer models obtained with similar prior assumptions (prior **a**), and also presented as a perturbation around (approximately) the *Utada et al.* [2003] 1-D profile (plotted at the bottom of the figure). Both of these models have been corrected for the ocean-land contrasts.

The only significant common feature of these models seems to be the distinctive upper mantle conductive anomaly under Japan, extending down to $\sim 450 - 550$ km depth. We should perhaps stress that the poor vertical resolution of the *Fujii and Schultz* [2002] data set, as well as the sparseness of the data in the Pacific region, does not allow us to make any conclusions from the dissimilarities observed in the mid-mantle structure. Further study would be required, and incorporation of the MT cable responses into the global data set would certainly be beneficial. This is an intended future direction of research (Section 9.2).

A 2-D regional model has been obtained for the upper mantle beneath the East Pacific
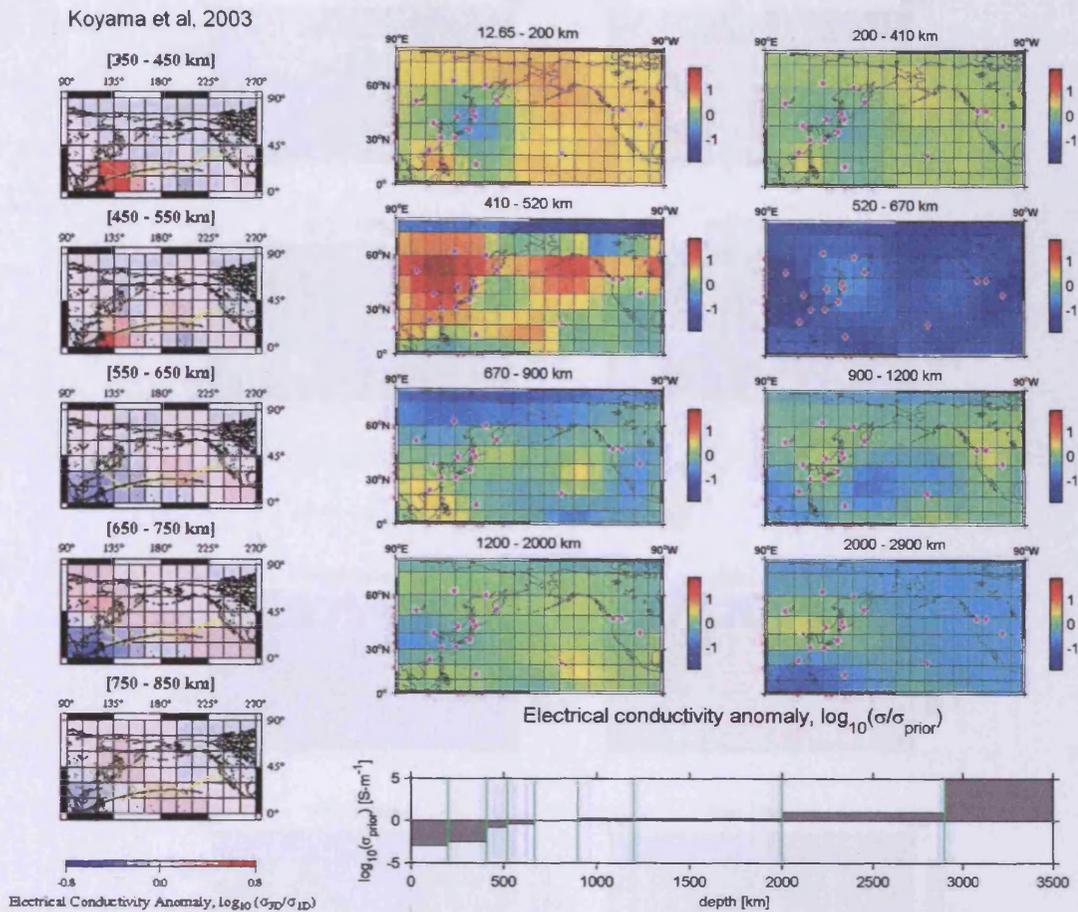
Figure 8.27: Here, we compare our results to the regional model of *Koyama et al.* [2003] in the North Pacific region. Both models are plotted as a 3-D logarithmic perturbation around the 1-D prior electrical conductivity profile computed by *Utada et al.* [2003] (Figure 1.9). Our best estimate for this profile, discretized for an 8-layer model, is plotted at the bottom of this figure. Our 8-layer model has been obtained by inverting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* c response data with the prior profile (**a**), the S-map, $\alpha = 0$ and $\mu = 0.01$. The global view of this model is provided on Figure 8.15. Our 8-layer model fits the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* c response data with the least squares misfit of 0.92.
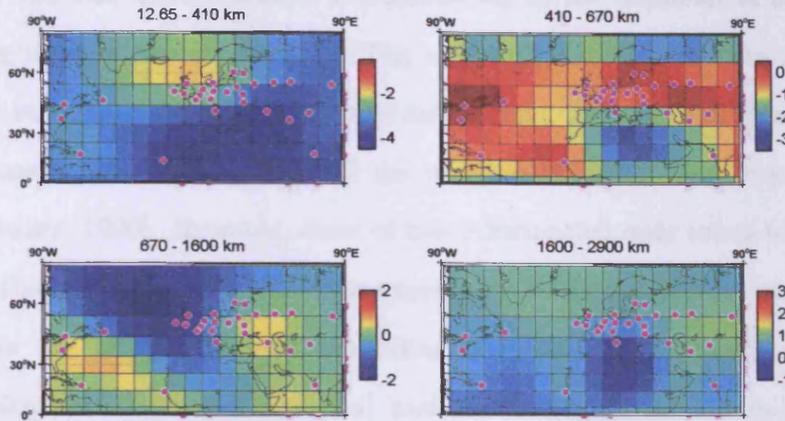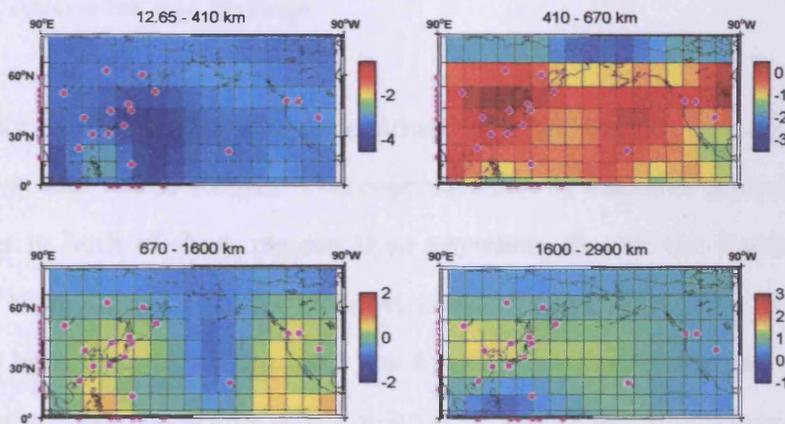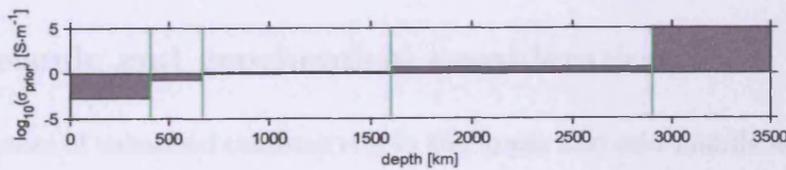
Figure 8.28: The 4-layer model degree and order 8 model with $\alpha = 0$ (Figure 8.9) is plotted for the two regions of interest: Europe and North Pacific. The model fits the $-60° \leq \theta \leq 60°$ *Fujii and Schultz c* response data with the least squares misfit of 1.00.

Rise [*Evans et al.*, 1999]. Unfortunately, the length of the array used in this experiment has been only 500 km, approximately corresponding to the diameter of a single grid cell at the spacing used in our experiment. The resulting a model consists of too fine-scale features to be informatively compared with our finding. Information about the electrical conductivity anomalies in the crust and the upper mantle has been summarised in the review by *Schwarz* [1990]. However, most of this information only refers to the uppermost 4 – 15 km of the sediment layer. The few exceptions include a discussion of a conductive anomaly under the Tibetan plateau, extending down to 40 – 50 km depths, and in the central Rio Grande rift (North America) and the so-called tin belt in Southern Peru, Bolivia and northern Argentina (both anomalies extending down to 20 km depth). It is not clear whether these features are significant enough to show in the global models; they clearly do not contradict our findings.

Additionally, we plot the electrical conductivity distribution for the two regions of interest: Europe and North Pacific. The concentration of available geomagnetic observatories is higher in both of these regions than anywhere else on the Earth's surface; this makes both of them a checkpoint for global studies. Both of the regional models presented in Figure 8.28 have been extracted from the 4-layer global inverse model obtained under the assumption of the conductive upper mantle (Figure 8.9). Unfortunately, we are not aware of any other deep EM studies in the European region.

8.5.3 Seismic and geochemical considerations

In general, zones of enhanced conductivity in the upper and mid-mantle would be expected to correlate with the known subducting slabs and the boundaries of tectonic plates, where the water had been introduced into the mantle. Seismic imaging has located a number

of subducting slabs in the mantle, including the Tethyan subducted slabs, imaged seismi-
cally beneath India and South-East Asia at the depths $\sim$ 1100 – 1500 km [e.g. *van der
Voo et al.*, 1999]; subducted trailing fragments of the Farallon plate in the upper mantle
beneath the western margin of North America [e.g. *van der Lee and Nolet*, 1997]; the sub-
ducted slab associated with the Australian-Antarctic Discordance, presently being drawn
up by the Southeast Indian Ridge [e.g. *Forsyth et al.*, 1987; *Gurnis et al.*, 1998]. Other
subducted slabs have been discussed by *Widiyantoro et al.* [1999]; *Zhao* [2004].

On the other hand, the old and stable parts of the continental crust, known as *cratons*,
could be a cause of anomalously resistive zones in the upper mantle, since they are known
to be cold (seismically fast), possibly dry and compositionally light (thus probably low in
iron content) [e.g. *Deschamps et al.*, 2002].

Strong conductive anomalies of 2 to 4 orders of magnitude that are, apparently, re-
quired by the data suggest an explanation in terms of the presence of water that causes
partial melt [*Thompson*, 1992], or the presence of positive charge carriers. For example,
Li and Jeanloz [1991] have shown that the inclusion of small amounts of water in the
silicates can enhance the electrical conductivity of the lower-mantle assemblage by more
than three orders of magnitude at these temperatures and pressures. Presence of water
could have been caused by subducting slabs.

Fukao et al. [1992] have suggested that descending slabs of the lithosphere in the
Western Pacific tend to be stagnant in the transition zone under the subtle control of the
670-km discontinuity under the Japan arc. Although stagnant slab materials eventually
descend into the lower mantle, they no longer maintain their original configuration below
the 670-km discontinuity. This could explain the strong conductive anomaly beneath the

Western Pacific reconstructed in most of our models (in fact, in all except those for which the penalty for conductive regions in the transition zone is too large). This anomaly continues into the mid-mantle just below the 670-km discontinuity in all of our models, although the anomaly is smaller in magnitude at greater depths.

Ichiki et al. [2001] suggest that the mantle transition zone beneath northeastern China is more conductive than those of other tectonic settings by almost one order of magnitude in the depth range of 400–600km. According to *Ichiki et al.* [2001], this feature may correspond to the presence of a stagnant slab beneath this region. This argument favours the strong conductive heterogeneity in this region reconstructed by those of our models that allow heterogeneity in the transition zone.

Unfortunately, direct comparison of our results with seismic wave models, such as *Dziewonski and Woodhouse* [1987], *Su et al.* [1994]; *Su and Dziewonski* [1997], *Romanowicz* [1995, 2001]; *Romanowicz and Gung* [2002], and others, has not been too illuminating. Although zones of low seismic velocities correspond to the high temperature anomalies, other factors, such as partial melt, influence the electrical conductivity to a greater extent. Quite clearly, it is not possible to extract anomalous electrical conductivity information from seismic models as such, unless we also possess additional information on e.g. the water content of the region (like we do in the subducted slabs studies). This is not surprising: ideally, EM induction studies should complement the corresponding seismic models by creating additional compositional constraints.

8.6 Comparative analysis of predicted magnetic fields and responses

We have plotted on Figures 8.29–8.49 the predicted magnetic field components and c and d responses at the Earth's surface for a range of our inverse models. The black dots denote the locations of the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* geomagnetic observatories. For the inversions, only the real and imaginary components of the c responses at the locations of the black dots have been exploited (Figures 8.38–8.40).

On Figures 8.29, 8.32, 8.35 we have plotted the real and imaginary parts of the longitudinal, latitudinal and vertical predicted magnetic fields, respectively, for the following four models: 1) fourlayers8 resistive upper mantle, $\alpha = 0$; 2) fourlayers8a conductive upper mantle, $\alpha = 0$; 3) earth8a conductive upper mantle with the S-map, $\alpha = 1$; 4) earth8a conductive upper mantle without the S-map, $\alpha = 1$. All of these models fit the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* c response data well. We see a number of smaller scale features in the maps of magnetic fields for the first two model, compared to the third and the fourth, which were regularised. We have also plotted the real and imaginary components of c and d responses on Figures 8.38 and 8.44, and the same values as magnitude and phase on Figures 8.41 and 8.47. All of these maps have been produced for a sample period of 8 days.

Similarly, we have shown the corresponding magnetic fields (Figures 8.30, 8.33, 8.36) and responses (Figures 8.39, 8.45, 8.42, 8.48) for the following four models: 1) earth4 resistive upper mantle, $\alpha = 0$; 2) earth8 resistive upper mantle, $\alpha = 0$; 3) earth8 resistive upper mantle without the S-map, $\alpha = 1$ and 4) the S-effect due to the S-map on top of the 1-D resistive upper mantle prior model. This information has been plotted for the

same sample period of 8 days; additionally, we plot it for an order of magnitude longer periods of 80 days (Figures 8.31, 8.34, 8.37, 8.40, 8.46, 8.43, 8.49).

We find it illuminating to see how remarkably similar the magnetic fields at the surface are for this range of rather different inverse models (displayed in Section 8.3). Some of the regions in these maps coincide even through there are no data to constrain them; some are different. The latter contain additional information about the model. Let us discuss in detail some of the significant features we see in these plots.

First of all, we note that the S-effect, plotted on the same scale as the magnetic field maps of the 3-D mantle models fitting the data, is still rather significant at the period of 8 days, showing especially in the imaginary parts of H_ϕ, H_θ and H_r and in the imaginary parts and phases of both c and d responses. The effect of the S-map on the d responses is much stronger than that on the c responses, compared to the respective perturbations due to the 3-D inverse models. The S-effect on the fields and c responses at a period of 80 days is essentially negligible. Some minor latitudinal distortion is seen in the H_θ component, which is definitely not significant enough to affect the c responses. It does however affect the phase of the d responses, which is visually seen as distorted even at a period of 80 days. We have also confirmed that the inversions in which no S-map has been included have somehow compensated for it, so that the magnetic fields at the surface do not look systematically different from those that have been obtained with the S-map distribution.

In general, we see that the imaginary components of H_θ clearly increase in size as the period gets longer, and so do both the real and imaginary parts of H_r. On the contrary, the real parts of H_θ decrease in size with longer periods. The average magnitude of H_ϕ does not depend on the period, but the distribution of the field does.

The appearance of the equatorial band in c responses has been already discussed in Section 7.2. The vertical magnetic field component H_r does not in general go to zero at the equator, if the model is latitudinally asymmetric, while the latitudinal components H_θ stay small; so that the ratio H_r/H_θ multiplied by $\tan(\theta)$ is unbounded at the equator. We do not have data close enough to the geomagnetic equator to constrain this feature well. Thus, we get these thin equatorial strikes in c responses seen in the Figures 8.38–8.40, even through these features do not show up in the magnetic field values. We also see, that the d responses are free from this problem. Surprisingly, even though they have not been used to obtain these inverse models, the c responses seem to constrain the models well enough that the maps of the d responses also turn out rather similar. Apart from the similarities, they certainly contain a great deal of additional information about the models even at some of the currently available observatory locations. This information could be used to further distinguish between the models by evaluating the data misfit with the d response values, along with the c response data.
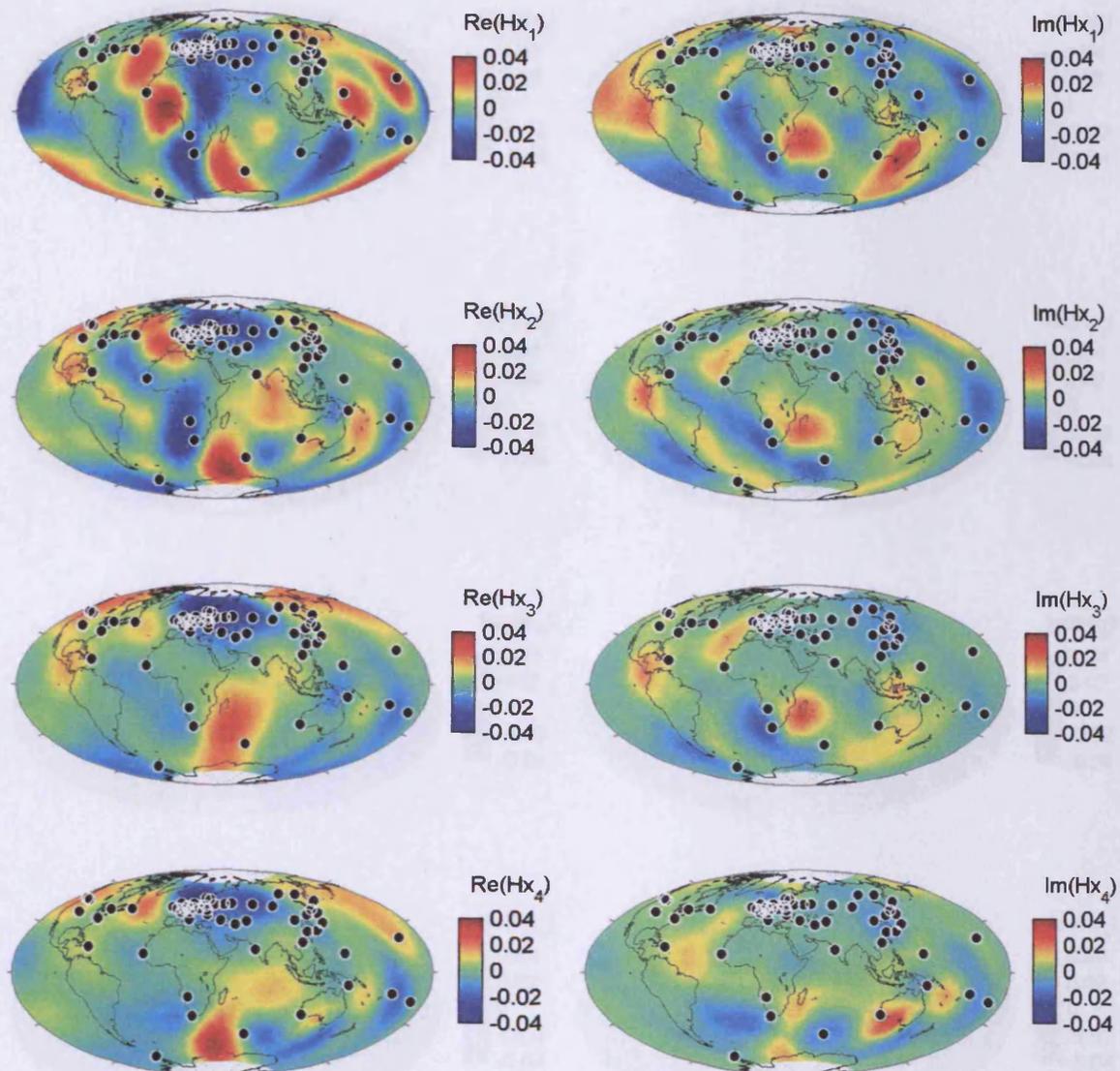
Figure 8.29: The plot shows the real and imaginary parts of the predicted longitudinal magnetic field H_ϕ at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (**c**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (**d**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (**a**), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (**a**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.30: The plot shows the real and imaginary parts of the predicted longitudinal magnetic field H_ϕ at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.31: The plot shows the real and imaginary parts of the predicted longitudinal magnetic field H_ϕ at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.32: The plot shows the real and imaginary parts of the predicted latitudinal magnetic field H_θ at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (c), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (d), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (a), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (a) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.33: The plot shows the real and imaginary parts of the predicted latitudinal magnetic field H_θ at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (b), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (b), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (b) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (b). The black dots denote geomagnetic observatories.

Figure 8.34: The plot shows the real and imaginary parts of the predicted latitudinal magnetic field H_θ at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.35: The plot shows the real and imaginary parts of the predicted vertical magnetic field H_r at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (**c**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (**d**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (**a**), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (**a**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.36: The plot shows the real and imaginary parts of the predicted vertical magnetic field H_r at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**).

Figure 8.37: The plot shows the real and imaginary parts of the predicted vertical magnetic field H_r at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.38: The plot shows the real and imaginary parts of the predicted c responses at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (**c**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (**d**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (**a**), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (**a**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.39: The plot shows the real and imaginary parts of the predicted c responses at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.40: The plot shows the real and imaginary parts of the predicted c responses at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.41: The plot shows the modulus and phase of the predicted c responses at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (c), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (d), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (a), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (a) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.42: The plot shows the modulus and phase of the predicted c responses at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.43: The plot shows the modulus and phase of the predicted c responses at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.44: The plot shows the real and imaginary parts of the predicted d responses at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (**c**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (**d**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (**a**), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (**a**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.45: The plot shows the real and imaginary parts of the predicted d responses at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.46: The plot shows the real and imaginary parts of the predicted d responses at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

8.7 Conclusions

We have performed a series of inversions of a frequency domain c response data set compiled by *Fujii and Schultz* [2002]. We have obtained a range of 3-D models of electrical conductivity distributions in the mantle of the Earth, fitting the $-60° \leq \theta \leq 60°$ *Fujii and Schultz* portion of the data set statistically well (with a scaled least squares misfit ~ 1). The variety of models fitting this data set has demonstrated poor vertical resolution of the *Fujii and Schultz* [2002] data in the upper mantle. Both additional observatory data and a larger range of frequencies would improve the resolution of the upper mantle and allow to further restrict our models. Although some features of the reconstructed models are likely to be robust, geodynamic interpretation is premature until inversions of d values and additional higher frequency data are accomplished.
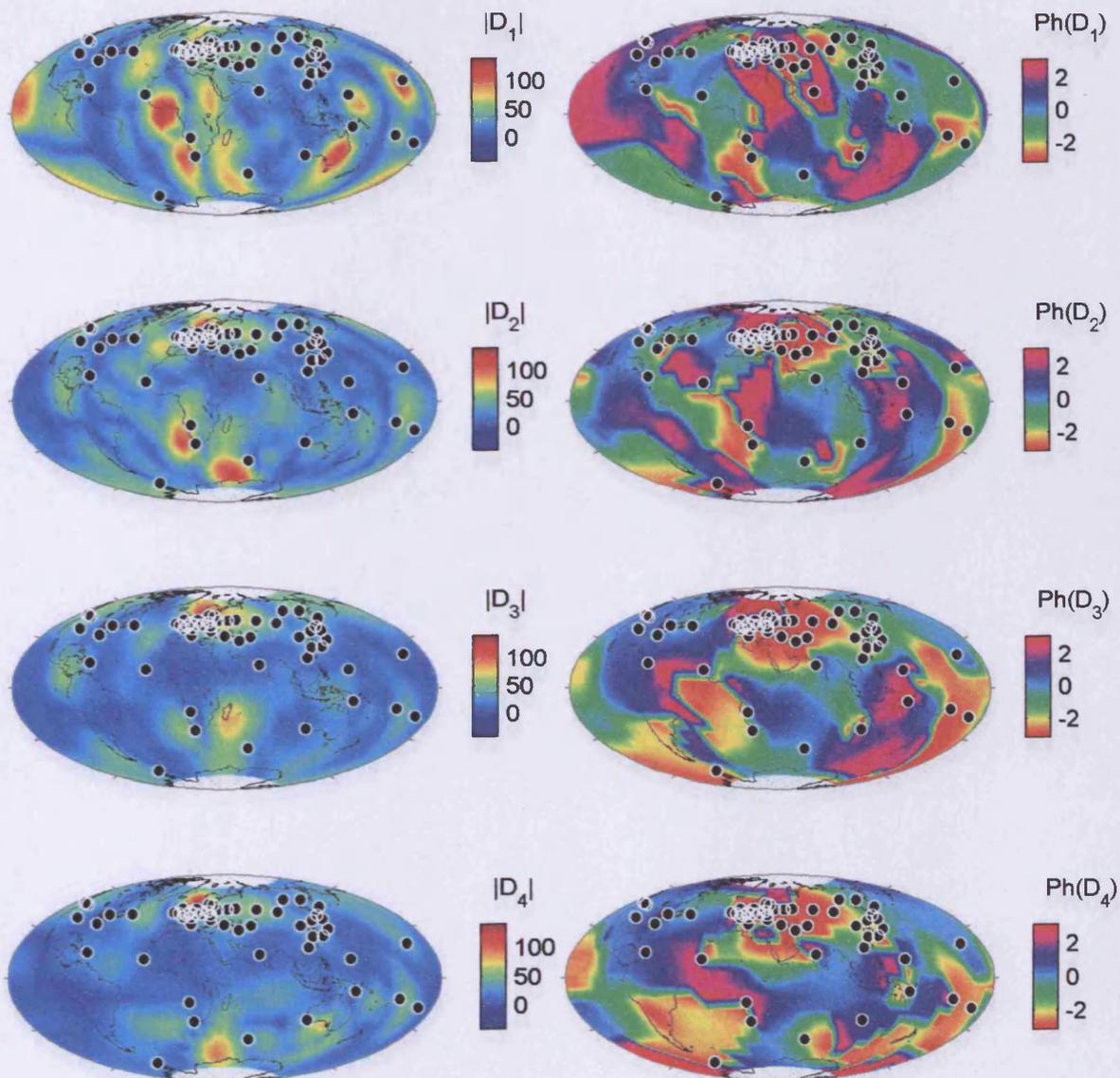
Figure 8.47: The plot shows the modulus and phase of the predicted d responses at the Earth's surface for a sample period of 8 days and a set of four models plotted on Figures 8.8, 8.9, 8.16 and 8.17, respectively. The models, from the top down, correspond to the 1) four layers degree and order 8 experiment with the prior (c), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.07; 2) four layers degree and order 8 experiment with the prior (d), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.00; 3) eight layers degree and order 8 experiment with the prior (a), the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10; 4) eight layers degree and order 8 experiment with the prior (a) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.04. The black dots denote geomagnetic observatories.

Figure 8.48: The plot shows the modulus and phase of the predicted d responses at the Earth's surface for a sample period of 8 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Figure 8.49: The plot shows the modulus and phase of the predicted d responses at the Earth's surface for a sample period of 80 days and a set of four Earth models. Models 1–3 are plotted on Figures 8.12, 8.13, 8.18, respectively. They correspond to the 1) eight layers degree and order 4 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.32; 2) eight layers degree and order 8 experiment with the prior (**b**), the $|\theta| \leq 60°$ data and $\alpha = 0$, data misfit 1.02; 3) eight layers degree and order 8 experiment with the prior (**b**) without the S-map, with the $|\theta| \leq 60°$ data and $\alpha = 1$, data misfit 1.10. The last figure shows the effect of the S-map at this frequency, imposed on top of the 1-D prior (**b**). The black dots denote geomagnetic observatories.

Chapter 9

Conclusions

9.1   Conclusions and summary of results

In this thesis, we have developed a least squares inverse solution for reconstructing large-scale inhomogeneities in the electrical conductivity distribution of Earth's mantle. We have also made an initial application of this method. We conclude that it is possible to fit the most current field data set consistent with the P_1^0 source structure with a model that can be described by a 4 to 8 layer, degree and order 8 surface spherical harmonic parametrization. However, many such models exist, and in order to distinguish between them to determine which one (if any) corresponds to the physical reality in the mantle more information is required. Our analysis suggests that a few strong anomalies at specific locations in both the upper and mid-mantle are probably required by the field data, but that the geodynamic analysis is premature until additional data are available. We have also shown that the data requires some inhomogeneity in both the transition zone and in the lower mantle.

9.2   Future directions

First of all, we stress the importance of obtaining more data. We have demonstrated the huge deterioration of the inverse model resolution if the sparse and irregularly distributed field data are used instead of a denser and regular distribution (Section 7.3). We note that this is true even in the case of randomly generated and perfectly normally distributed 5% errors. The irregular distribution of the data affects the latitudinal resolution so such an extent, that only a few inhomogeneity areas best covered by the data can be resolved. However, an even greater problem, as we have seen, is the poor vertical resolution of the present data set. While in all cases the inversion reconstructs pretty much the same pattern of irregularities, it is not at all clear where to place it vertically. The vertical resolution of the upper mantle would be greatly improved if higher frequency data at periods $\sim 1 - 5$ days were obtained and inverted jointly with the present data set. The d response data from the *Fujii and Schultz* [2002] should also be inverted. There are reasons to believe that the d responses are more sensitive to the inhomogeneities in the Earth's mantle than are the c responses at a point, and should therefore provide a better constraint, as well as add more information (see Sections 2.3 and 7.2 for discussions). In fact, it has been shown that d-responses contain the bulk of the 3-D information about the Earth's interior (pers. comm. Adam Schultz). Ultimately, inversion of the EOF's is also planned. Additionally, satellite EM data (e.g. *Olsen* [1999b]), global geomagnetic observatory data from the currently operating network (INTERMAGNET) and cable responses (e.g. *Utada et al.* [2003]) should all be jointly inverted to significantly improve the resolution. This is where we propose to concentrate our efforts in the future work.

An interesting test would be to invert various subsets of observatories, and to see how that affects the solution. Alternatively, could try deleting observatories from the data set one by one and inverting the rest. It would be interesting to see and quantify how the solutions to these problems would differ.

9.2.1   Accuracy issues

Another important issue would be to study the effect of discretization and truncation errors on the forward solution, and consequently on the final result of the inversion. While for several grid divisions is has been shown that the solution of the forward solver is very accurate compared to both quasi-analytic and other numerical solutions (in U&S, grid divisions were [2, 36, 136] and [50, 31, 53] for the two examples, respectively) the effect of grid spacing on the calculated response of the Earth has never been systematically studied. Although in general the accuracy of solution grows as the computational grid is refined, we have encountered convergence problems at the poles as the grid spacing gets very fine ($1° - 3°$). This effect needs to be carefully studied and, if possible, eliminated, before truly fine computational grids may be used for the inversion. This is a necessary step in increasing the accuracy of both the forward and the inverse solutions. The effect of the precision used in calculations could also be studied to see if higher precision significantly improves the accuracy of the solver, but this is a minor issue compared to ensuring convergence at the poles.

Another accuracy issue concerns the current implementation of the thin shell conductance distribution. We include the distribution of the oceans directly into the computational grid just before running the forward computations. Thus, if there were any inaccuracies in this distribution the inversion would have no way of correcting them: these values are taken constant just like the conductivity of the air layers. We have already seen (Sections 7.2 and 7.3) the significant effect that the S-distribution has on the Earth's responses. Thus, inaccuracies in the S-distribution could introduce a large error into the solution, due to the ill-posedness of the inverse problem. We note, that as we map the

S-distribution on the relatively coarse $5 - 10°$ grid that we are using, inaccuracies are inevitably introduced. However, allowing the inversion to treat the thin shell as a variable layer results in trying to reconstruct this fine structure with relatively low degree and order spherical harmonics, which doesn't allow for good precision either. A better way to account for the near-surface inhomogeneities is required. This is an important issue but a solution is mainly a matter of a clever implementation.

9.2.2 Inversion with respect to the sources

We have seen that the data at very high geomagnetic latitudes ($\ll -60°$ and $\gg 60°$) can not be fit by a degree and order 8 spherical harmonic parametrization, which the mid-latitude data is perfectly fittable with the same parametrization (Section 8.4). The most likely cause for this phenomenon is that the source assumptions described in Chapter 2 are not a good enough approximation to the true source at high latitudes. The generalised forward solution developed in Chapter 3 allows inverting the data with respect to sources along with the conductivity model. This approach is not totally novel in geomagnetic sounding (see *Mulhall* [2003]), however, it has never been performed for the global 3-D problem. We feel it would be beneficial to develop this kind of joint inversion.

9.2.3 Quantifying the dependence on the prior

We also suggest the following procedures that would in our view complement obtaining and inverting more data to let us achieve a better understanding of the Earth's interior electrical conductivity structure. In our view, the observed dependence of the solution on the prior model, i.e. our prior understanding of the Earth's one-dimensional electrical conductivity structure, opens possibilities for novel research in this area.

We propose to study more rigorously the dependence of the computed electrical conductivity distribution on the prior 1-D model. This can be done by varying smoothly each of the prior model parameters throughout its physically realisable domain. For each of these prior models, a solution may exist that fits our current data set perfectly well. Quantifying the discrepancies between this series of solutions would give us an idea as to which features of solutions are most affected by varying the prior conductivity of each model layer. Superimposed on the knowledge of the 1-D structure of the Earth that is constantly updated and refined by Earth sciences, this would serve as a useful addition to our resolution analysis.

Another study that we propose is rather ambitious in its computational cost, although we believe it could prove very useful in the long run. Apart from several regional profiles that exist, our current knowledge regarding the 1-D approximation of the Earth's electrical conductivity structure could be summarized as inequality constraints both from below and from above: $\sigma_{min}^{(n)} \leq \sigma_{1D}^{(n)} \leq \sigma_{max}^{(n)}$, where the index n corresponds to a layer in the Earth, or, in the continuous case, to depth. We could make use of this knowledge in a rigorous way to generate a pseudo-random set of prior models in this domain, in much the same way as *Press* [1970] did. Inverting field data using each of these prior models, even if this inversion is only approximately performed on a sparse computational grid, would give us a database of all acceptable models fitting the data. Now, each of these models gives rise to a unique distribution of magnetic fields at the Earth's surface. Quantifying the differences between these magnetic field maps by some sort of ϕ- and θ-dependent function (a measure of information) would allow us to plot the regions on the Earth's surface that contain most information about the interior. In particular, if the discrepancies in the magnetic field distribution arising from all of our models are large is a particular region on the Earth's surface, we say that this region contains a lot of information. Thus, if an observatory was set up in that area that measured the true magnetic field values, these values would allow

us to further constrain our knowledge of the Earth's interior by discriminating between the models in our database. Also, any additional data that we obtain could be used to further distinguish between the models that are or are not acceptable. This procedure, if quantified in a mathematically rigorous way, could complement inverting larger data sets, since it is always easier to search in a discrete than in a continuous space. However, in generating the models that would constitute such a database one should be very careful not to introduce any features that are not necessarily required by the data. Models that are favoured by additional field measurements could always be refined by inverting the updated EM field data set using the favoured model as the starting model for the search or the prior, or both.

9.2.4 Resolution analysis

The Jacobian computed via the scheme devised in Chapter 5 allows us to study the resolution of our inverse method quantitatively. Formal resolution analysis could be performed, based on a non-linear extension of Backus-Gilbert theory such as the methodology described by *Snieder* [1990, 1991, 1998]. Resolving kernels could be constructed to give us a better idea whether the features seen in the inverse models are truly required by the data.

Let us define a mapping $M : \mathbb{R}^N \to \mathbb{G}$ which corresponds to the parametrization mapping from the spherical harmonic coefficients to the grid cells. Then, using the notation of Chapters 3 and 5, define

$$J_{\mathbf{m}} = -G^* J M \qquad (9.1)$$

and

$$J_c = -G^* J. \qquad (9.2)$$

Then

$$R = C_{\mathrm{m}} J_{\mathrm{m}}^{\mathrm{T}} (J_{\mathrm{m}} C_{\mathrm{m}} J_{\mathrm{m}}^{\mathrm{T}} + C_{\mathrm{d}})^{-1} J_{\mathrm{m}} \tag{9.3}$$

tells how spherical harmonic parameters are resolved, and

$$R_c = M C_{\mathrm{m}} J_{\mathrm{m}}^{\mathrm{T}} (J_{\mathrm{m}} C_{\mathrm{m}} J_{\mathrm{m}}^{\mathrm{T}} + C_{\mathrm{d}})^{-1} J_c \tag{9.4}$$

tells how cell conductivities get blurred through the inversion (pers. comm. Gary Egbert).

Then, we could generate checkerboards using spherical harmonics of a specific degree (up to the highest degree used in the parametrization), e.g. $l = 8$ $m = 6$, or just the Y_m^m tesseral harmonics. For each degree m look at the column of R corresponding to Y_m^m (or, if several coefficients are perturbed, at the linear combination of these columns). A single column of R corresponds to generating a unit perturbation in a single spherical harmonic coefficient and evaluating R δm.

This can be plotted, by mapping spherical harmonic coefficients to cells (i.e. multiply the column of R by M), and looking at the plots at various depths and for various m. We could define the notion of *semblance* in much the same way as *Zelt* [1998], and use it to roughly quantify which degree is resolved at selected depths and locations (e.g. beneath Europe, North America, the middle of Pacific, etc).

A more complicated and computationally demanding analysis would involve R_c. For instance, we could evaluate $R_c \delta\rho$, where $\delta\rho$ is a perturbation in resistivity in selected locations. This would involve evaluating spherical harmonic parameters $C_{\mathrm{m}} J_{\mathrm{m}}^{\mathrm{T}} (J_{\mathrm{m}} C_{\mathrm{m}} J_{\mathrm{m}}^{\mathrm{T}} + C_{\mathrm{d}})^{-1} J_c \delta\rho$, and mapping them back to the grid with M. This would result in a distribution on the grid, that could be plotted for specific depths, showing spherical projections at the cell depth and meridional/zonal cross sections. We could also try to define some length scales through some sort of ad hoc procedure; this would allow us to display some resolution information for more points simultaneously. Setting up a specific procedure to

accomplish this could be a challenge.

Bibliography

Ahlfors, L. V. (1987), *Lectures on quasiconformal mappings*, The Wadsworth & Brooks/Cole Mathematics Series, viii+146 pp., Wadsworth & Brooks/Cole Advanced Books & Software, Monterey, CA.

Alexandrescu, M. M., D. Gibert, J.-L. Le Mouël, G. Hulot, and G. Saracco (1999), An estimate of average lower mantle conductivity by wavelet analysis of geomagnetic jerks, *Journal of Geophysical Research*, *104*, 17,735–17,746, doi:10.1029/1999JB900135.

Allègre, C. J., J.-P. Poirier, E. Humler, and A. W. Hofmann (1995), The chemical composition of the Earth, *Earth and Planetary Science Letters*, *134*(3-4), 515–526.

Arfken, G., and H. Weber (2001), *Mathematical methods for physicists*, Academic Press.

Avdeev, D. (2005), Three-dimensional electromagnetic modelling and inversion from theory to application, *Surveys in Geophysics*, *26*(6), 767–799.

Avdeev, D. B., A. V. Kuvshinov, O. V. Pankratov, and G. A. Newman (1997), High-performance three-dimensional electromagnetic modelling using modified Neumann series. Wide-band numerical solution and examples, *Journal of Geomagnetism and Geoelectricity*, *49*(11-12), 1519–1539.

Backus, G., R. Parker, and C. Constable (1996), *Foundations of Geomagnetism*, Cambridge University Press.

Backus, G. E., and F. Gilbert (1967), Numerical application of a formalism for geophysical inverse problems, *Geophys. J. Roy. Astron. Soc.*, *13*, 247–276.

Bahr, K., and A. Duba (2000), Is the asthenosphere electrically anisotropic?, *Earth Planet. Sci. Lett.*, *178*.

Bahr, K., N. Olsen, and T. J. Shankland (1993), On the combination of the magneto-telluric and the geomagnetic depthsounding method for resolving an electrical conductivity increase at 400 km depth, *Geophysical Research Letters*, *20*, 2937–2940.

Bailey, R. C. (1970), Inversion of the Geomagnetic Induction Problem, *Royal Society of London Proceedings Series A*, *315*, 185–194.

Balasis, G., G. D. Egbert, and S. Maus (2004), Local time effects in satellite estimates of electromagnetic induction transfer functions, *Geophysical Research Letters*, *31*(16), L16,610.

Banks, R. (1969), Geomagnetic variations and the electrical conductivity of the upper mantle, *Geophys. J. R. astr. Soc.*, *17*, 457–487.

Banks, R. J., and J. N. Ainsworth (1992), Global induction and the spatial structure of mid-latitude geomagnetic variations, *Geophysical Journal International*, *110*, 251–266.

Barton, C. E. (1997), International geomagnetic reference field: the seventh generation, *J. Geomag. Geoelectr.*, *49*, 123148.

Bercovici, D., and S.-i. Karato (2003), Whole-mantle convection and the transition-zone water filter, *Nature*, *425*, 39–44.

Berdichevsky, M. N., V. I. Dmitriev, and N. A. Mertzhikova (2000), *Ob obratnoi zadache zondirovaniya s ispol'zovaniem magnetotelluricheskih i magnetovaryatsionnyh dannyh*

(On the inverse problem of sounding using magnetotelluric and magnetovariational data), MAKS Press, Moscow, in Russian.

Boehler, R. (1996), Melting temperature of the Earth's mantle and core: Earth's thermal structure, *Annu. Rev. Earth Planet. Sci.*, *24*(7), 15–40.

Buffett, B. A. (1992), Constraints on magnetic energy and mantle conductivity from the forced nutations of the earth, *Journal of Geophysical Research*, *97*, 19,581–+.

Buffett, B. A., E. J. Garnero, and R. Jeanloz (2000), Sediments at the Top of Earth's Core, *Science*, *290*(5495), 1338–1342.

Cagniard, L. (1953), Basic theory of the magnetotelluric method of geophysical prospecting, *Geophysics*, *18*, 605–635.

Campbell, W. H. (2003), *Introduction to Geomagnetic Fields*, 2nd ed., Cambridge University Press.

Chandrasekhar, E., N. Oshiman, and K. Yumoto (2003), On the role of oceans in the geomagnetic induction by S_q along the 210 degrees magnetic meridian region, *Earth Planets and Space*, *55*(6), 315–326.

Chappell, C. R., T. E. Moore, and J. H. Waite, Jr. (1987), The ionosphere as a fully adequate source of plasma for the earth's magnetosphere, *Journal of Geophysical Research*, *92*, 5896–5910.

Chave, A. D., D. J. Thomson, and M. E. Ander (1987), On the robust estimation of power spectra, coherences, and tranfer functions, *Journal of Geophysical Research*, *92*(B1), 633–648.

Constable, S. (1993), Constraints on mantle electrical-conductivity from field and laboratory measurements, *J. Geomagn. Geoelectr.*, *45*(9), 707–728.

Constable, S. (2006), SEO3: A new model of olivine electrical conductivity, *Geophysical Journal International*, *166*(1), 435–437, doi:10.1111/j.1365-246X.2006.03041.x.

Constable, S., and C. Constable (2004), Observing geomagnetic induction in magnetic satellite measurements and associated implications for mantle conductivity, *Geochemistry Geophysics Geosystems*, *5*, Q01,006.

Constable, S. C., R. L. Parker, and C. G. Constable (1987), Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, *52*(3), 289–300, http://mahi.ucsd.edu/Steve/Occam/home.html.

Cox, C., J. H. Filloux, and J. Larsen (1970), *The Sea*, vol. 4, chap. Electromagnetic studies of ocean currents and electrical conductivity below the ocean floor, pp. 637–693, Wiley, New York.

Davies, G. F. (2001), *Dynamic Earth : Plates, Plumes and Mantle Convection*, Cambridge University Press.

Deschamps, F., J. Trampert, and R. Snieder (2002), Anomalies of temperature and iron in the uppermost mantle inferred from gravity data and tomographic models, *Physics of The Earth and Planetary Interiors*, *129*(3-4), 245–264.

Duan, Q. Y., V. K. Gupta, and S. Sorooshian (1993), Shuffled complex evolution approach for effective and efficient global minimization, *J. Optim. Theory Appl.*, *76*(3), 501–521.

Dungey, J. W. (1961), Interplanetary magnetic field and the auroral zones, *Phys. Rev. Lett.*, *6*(2), 47–48, doi:10.1103/PhysRevLett.6.47.

Dziewonski, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Physics of the Earth and Planetary Interiors*, *25*, 297–356, doi:10.1016/0031-9201(81) 90046-7.

Dziewonski, A. M., and J. H. Woodhouse (1987), Global Images of the Earth's Interior, *Science*, *236*, 37–48.

Egbert, G. (2006), *Encyclopedia of Geomagnetism and Paleomagnetism*, chap. Inverse Modeling of EM; Robust Electromagnetic Tranfer Functions Estimates, Edited by D. Gubbins and E. Herrero-Bervera. Berlin: Springer, 2006.

Egbert, G. D. (1997), Robust multiple-station magnetotelluric data processing, *Geophysical Journal International*, *130*(2), 475–496.

Egbert, G. D., and J. R. Booker (1992), Very long period magnetotellurics at Tucson observatory - Implications for mantle conductivity, *Journal of Geophysical Research*, *97*, 15,099–+.

Egbert, G. D., J. R. Booker, and A. Schultz (1992), Very long period magnetotellurics at Tucson-observatory – Estimation of impedances, *J. Geophys. Res.-Solid Earth*, *97*(B11), 15,113–15,128.

Engl, H., M. Hanke, and A. Neubauer (2000), *Regularisation of Inverse Problems*, Kluwer Academic Publishers.

Evans, R. L., et al. (1999), Asymmetric Electrical Structure in the Mantle Beneath the East Pacific Rise at 17 degrees S, *Science*, *286*(5440), 752–756.

Everett, M. E., and A. Schultz (1995), Geomagnetic induction in eccentrically nested spheres, *Phys. Earth Planet. Inter.*, *92*(3-4), 189–198.

Everett, M. E., and A. Schultz (1996), Geomagnetic induction in a heterogenous sphere: Azimuthally symmetric test computations and the response of an undulating 660-km discontinuity, *J. Geophys. Res.-Solid Earth*, *101*(B2), 2765–2783.

Fainberg, E. B., A. V. Kuvshinov, L. P. Mishina, and B. S. Singer (1990a), The new approach to global deep sounding, *Pure Appl. Geophys.*, *134*(4), 527–531.

Fainberg, E. B., A. V. Kuvshinov, and B. S. Singer (1990b), Electromagnetic induction in a spherical Earth with nonuniform oceans and continents in electric contact with the underlying medium: 1. Theory, method and example, *Geophys. J. Int.*, *102*(2), 273–281.

Fainberg, E. B., A. V. Kuvshinov, and B. S. Singer (1990c), Electromagnetic induction in a spherical Earth with nonuniform oceans and continents in electric contact with the underlying medium: 2. Bimodal global geomagnetic sounding of the lithosphere, *Geophys. J. Int.*, *102*(2), 283–286.

Fainberg, E. B., O. V. Pankratov, and B. S. Singer (1993), Thin sheet modelling of subsurface and deep inhomogeneities, *Geophysical Journal International*, *113*(1), 144.

Farber, D. L., Q. Williams, and F. J. Ryerson (2000), Divalent cation diffusion in Mg_2SiO_4 spinel (ringwoodite), b phase (wadsleyite), and olivine: Implications for the electrical conductivity of the mantle, *J. Geophys. Res.*, *105*, 513 529.

Forsyth, D. W., R. L. Ehrenbard, and S. Chapin (1987), Anomalous upper mantle beneath the Australian-Antarctic discordance, *Earth and Planetary Science Letters*, *84*(4), 471–478.

Fujii, I., and A. Schultz (2002), The 3D electromagnetic response of the Earth to ring current and auroral oval excitation, *Geophys. J. Int.*, *151*(3), 689–709.

Fukao, Y., M. Obayashi, H. Inoue, and M. Nenbai (1992), Subducting slabs stagnant in the mantle transition zone, *Journal of Geophysical Research*, *97*, 4809–4822.

Gauss, C. F. (1838), *Allgemeine Theorie des Erdmagnetismus*, Resultate aus den Beobachtungen des Magnetischen Vereins im Jahre 1838, english translation by E. Sabine and R. Taylor, 1848 in Scientific Memoirs Selected from the Transactions of Foreign Academies and Learned Societies and from Foreign Journals, pp. 184-251, J. and R. E. Taylor Pub., London.

Giering, R., and T. Kaminski (1998), Recipes for adjoint code construction, *ACM Trans. Math. Softw.*, *24*(4), 437–474, doi:http://doi.acm.org/10.1145/293686.293695.

Gilbert, W. (1600), *De Magnete (On the Magnet)*, Peter Short, London, translation by Silvanus Thompson, 1900 commemorative issue reproduced in facsimile by Basic Books, Inc., NY 1956 translation by Paul Fleury Mottelay (1893), reprinted by Dover Books, 1958.

Goldberg, D. E. (1989), *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Pub. Co.

Grammatica, N., and P. Tarits (2002), Contribution at satellite altitude of electromagnetically induced anomalies arising from a three-dimensional heterogeneously conducting Earth, using Sq as an inducing source field, *Geophysical Journal International*, *151*, 913–923, doi:10.1046/j.1365-246X.2002.01817.x.

Gurnis, M., R. D. Müller, and L. Moresi (1998), Cretaceous Vertical Motion of Australia and the Australian–Antarctic Discordance, *Science*, *279*(5356), 1499–1504, doi: 10.1126/science.279.5356.1499.

Haber, E. (2005), Quasi-Newton methods for large-scale electromagnetic inverse problems, *Inverse Problems*, *21*(1), 305–323.

Haber, E., U. M. Ascher, and D. W. Oldenburg (2004), Inversion of 3D electromagnetic

data in frequency and time domain using an inexact all-at-once approach, *Geophysics*, *69*(5), 1216–1228.

Hamano, Y. (2002), A new time-domain approach for the electromagnetic induction problem in a three-dimensional heterogeneous Earth, *Geophys. J. Int.*, *150*(3), 753–769.

Hansen, P. C. (1998), *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

Hirsch, L., and A. Shankland, T.J.and Duba (1993a), Electrical-conduction and polaron mobility in Fe-bearing olivine, *Geophys. J. Int.*, *114*, 3644.

Hirsch, L., and T. Shankland (1993b), Quantitative olivine-defect role of Fe content, *Geophys. J. Int.*, *114*, 2135.

Huang, X., Y. Xu, and S.-i. Karato (2005), Water content in the transition zone from electrical conductivity of wadsleyite and ringwoodite, *Nature*, *434*, 746–749, doi:10.1038/nature03426.

Huang, X., Y. Xu, and S.-I. Karato (2006), Earth science: A wet mantle conductor? (Reply), *Nature*, *439*, 3–4, doi:10.1038/nature04529.

Huber, P. J. (1981), *Robust Statistics*, John Wiley & Sons, New York.

Ichiki, M., M. Uyeshima, H. Utada, Z. Guoze, T. Ji, and M. Mingzhi (2001), Upper mantle conductivity structure of the back-arc region beneath northeastern China, *Geophysical Research Letters*, *28*, 3773–3776, doi:10.1029/2001GL012983.

Jackson, J. D. (1975), *Classical Electrodynamics*, second ed., John Wiley & Sons.

Jeanloz, R. (1990), The nature of the Earth's core, *Annual Review of Earth and Planetary Sciences*, *18*(1), 357–386.

Jones, A. G., and I. J. Ferguson (2001), The electric Moho, *Nature*, *409*, 331–333.

Jordan, T. H. (1975), Lateral heterogeneity and mantle dynamics, *Nature*, *257*(5529), 745–750.

Kirsch, A. (1996), *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag New York, Inc.

Koyama, T., H. Shimizu, and H. Utada (2003), Three-dimensional electrical conductivity structure beneath North Pacific by using a submarine cable network, Scientific Use of Submarine Cables and Related Technologies, 2003. The 3rd International Workshop.

Kuvshinov, A., and N. Olsen (2006), A global model of mantle conductivity derived from 5 years of CHAMP, Ørsted, and SAC-C magnetic data, *Geophys. Res. Lett.*, *in press*.

Kuvshinov, A., H. Utada, D. Avdeev, and T. Koyama (2005), 3-D modelling and analysis of $D_s t$ C-responses in the North Pacific Ocean region, revisited, *Geophysical Journal International*, *160*, 505–526, doi:10.1111/j.1365-246X.2005.02477.x.

Kuvshinov, A. V., and O. V. Pankratov (1994), Electromagnetic induction in a spherical Earth with inhomogeneous conducting mantle: thin sheet forward problem, in *Proc. 12th Workshop on the Earth, Brest, France*, p. 83.

Kuvshinov, A. V., O. V. Pankratov, and B. S. Singer (1990), The effect of the oceans and sedimentary cover on global magnetovariational field distribution, *Pure and Applied Geophysics*, *134*(4), 533–540.

Kuvshinov, A. V., D. B. Avdeev, and O. V. Pankratov (1999), Global induction by S_q and D_{st} sources in the presence of oceans: Bimodal solutions for non-uniform spherical surface shells above radially symmetric Earth models in comparison to observations, *Geophys. J. Int.*, *137*(3), 630–650.

Kuvshinov, A. V., N. Olsen, D. B. Avdeev, and O. V. Pankratov (2002), Electromagnetic induction in the oceans and the anomalous behaviour of coastal c-responses for periods up to 20 days, *Geophys. Res. Lett.*, *29*(12), art. no–1595.

Langer, R. E. (1933), An inverse problem in differential equations, *Bull., Am. Math. Soc.*, ser. *2*, *29*, 814–820.

Larsen, J. C. (1973), An introduction to electromagnetic induction in the ocean, *Physics of the Earth and Planetary Interiors*, *7*, 389–398, doi:10.1016/0031-9201(73)90063-0.

Lawson, C. L., and R. J. Hanson (1974), *Solving Least Squares Problems*, Prentice-Hall Series in Automatic Computation, Prentice-Hall, Inc., Eaglewood Cliffs, New Jersey.

Lay, T. (1994), The Fate of Descending Slabs, *Annual Review of Earth and Planetary Sciences*, *22*, 33–61, doi:10.1146/annurev.ea.22.050194.000341.

Li, X., and R. Jeanloz (1991), Phases and electrical conductivity of a hydrous silicate assemblage at lower-mantle conditions, *Nature*, *350*(6316), 332–334.

Lines, L. R., and S. Treitel (1984), Tutorial - a review of least-squares inversion and its application to geophysical problems, *Geophysical Prospecting*, *32*(2), 159–186.

Lizarralde, D., A. Chave, G. Hirth, and A. Schultz (1995), Northeastern Pacific mantle conductivity profile from long-period magnetotelluric sounding using Hawaii-to-California submarine cable data, *J. Geophys. Res.-Solid Earth*, *100*(B9), 17,837–17,854.

Macbain, J. (1986), On the Frechét differentiability of the one-dimensional magnetotellurics problem, *Geophys. J. Int.*, *86*(2), 669–672.

Mackie, R. L., and T. R. Madden (1993a), 3-Dimensional magnetotelluric inversion using conjugate gradients, *Geophys. J. Int.*, *115*(1), 215–229.

Mackie, R. L., and T. R. Madden (1993b), Conjugate direction relaxation solutions for 3-D magnetotelluric modeling, *Geophysics*, *58*(7), 1052–1057.

Marquardt, D. W. (1963), An algorithm for least-squares estimation of non-linear parameters, *J. Soc. Ind. Appl. Math.*, *11*, 431–441.

Martinec, Z. (1998), Geomagnetic induction in multiple eccentrically nested spheres, *Geophysical Journal International*, *132*, 96–110.

Martinec, Z. (1999), Spectral-finite element approach to three-dimensional electromagnetic induction in a spherical Earth, *Geophysical Journal International*, *136*, 229–250.

Martinec, Z., M. E. Everett, and J. Velimsky (2003), Time-domain, spectral finite-element approach to transient 2-D geomagnetic induction in a spherical heterogeneous Earth, *Geophysical Journal International*, *155*(1), 33–43.

Masters, T. G., and P. M. Shearer (1995), Seismic Models of the Earth: Elastic and Anelastic, in *Global Earth Physics: A Handbook of Physical Constants*, edited by T. J. Ahrens, pp. 88–+.

Mathews, P. M., T. A. Herring, and B. A. Buffett (2002), Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior, *Journal of Geophysical Research (Solid Earth)*, *107*, 3–1, doi:10.1029/2001JB000390.

McElhinny, M. W. (1973), Mantle plumes, palaeomagnetism and polar wandering, *Nature*, *241*(5391), 523–524.

Meyer, A. (1990), A parallel preconditioned conjugate gradient method using domain decomposition and inexact solvers on each subdomain, *Computing*, *45*(3), 217–234.

Mulhall, A. (2003), 1-D Time Domain Inversion, Ph.D. thesis, Cambridge University.

Neal, S. L., R. L. Mackie, J. C. Larsen, and A. Schultz (2000), Variations in the electrical conductivity of the upper mantle beneath North America and the Pacific ocean, *J. Geophys. Res.-Solid Earth*, *105*(B4), 8229–8242.

Newman, G. A., and D. L. Alumbaugh (2000), Three-dimensional magnetotelluric inversion using non-linear conjugate gradients, *Geophysical Journal International*, *140*, 410–424.

Olsen, N. (1998), The electrical conductivity of the mantle beneath Europe derived from c-responses from 3 to 720 hr, *Geophys. J. Int.*, *133*(2), 298–308.

Olsen, N. (1999a), Long-period (30 days-1 year) electromagnetic sounding and the electrical conductivity of the lower mantle beneath Europe, *Geophys. J. Int.*, *138*(1), 179–187.

Olsen, N. (1999b), Induction studies with satellite data, *Surveys in Geophysics*, *20*(3 - 4), 309–340.

Olsen, N., and A. Kuvshinov (2004), Modeling the ocean effect of geomagnetic storms, *Earth, Planets, and Space*, *56*, 525–530.

Olsen, N., H. Luhr, T. J. Sabaka, M. Mandea, M. Rother, L. Toffner-Clausen, and S. Choi (2006), CHAOS – a model of the Earth's magnetic field derived from CHAMP, ørsted, and SAC-C magnetic satellite data, *Geophysical Journal International*, *166*(1), 67–75, doi:10.1111/j.1365-246X.2006.02959.x.

Olson, P., and V. L. Hagee (1990), Geomagnetic polarity reversals, transition field structure, and convection in the outer core, *Journal of Geophysical Research*, *95*, 4609–4620.

Palme, H., and H. S. C. O'Neill (2003), Cosmochemical estimates of mantle composition, in *Treatise on Geochemistry*, edited by H. D. Holland and K. K. Turekian, pp. 1–38, Pergamon, Oxford.

Parker, R. L. (1980), The inverse problem of electromagnetic induction: existence and construction of solutions based on incomplete data, *J. Geophys. Res.*, *85*, 4421–4428.

Parker, R. L. (1984), The inverse problem of resistivity sounding, *Geophysics*, *49*, 2143–2158.

Parker, R. L. (1986), On the Frechét differentiability of the one-dimensional magnetotellurics problem - comment, *Geophys. J. Int.*, *86*(2), 673–673.

Parker, R. L. (1994), *Geophysical Inverse Theory*, Princeton University Press, Princeton, New Jersey.

Parker, R. L., and K. A. Whaler (1981), Numerical methods for establishing solutions to the inverse problem of electromagnetic induction, *J. Geophys. Res.*, *86*, 9574–9584.

Parkinson, W. (1983), *Introduction to Geomagnetism*, Scottish Academic Press, Edinburgh.

Peyronneau, J., and J. P. Poirier (1989), Electrical-conductivity of the Earth's lower mantle, *Nature*, *342*(6249), 537–539.

Press, F. (1970), Earth models consistent with geophysical data, *Physics of the Earth and Planetary Interiors*, *3*, 3–22, doi:10.1016/0031-9201(70)90039-7.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992), *Numerical Recipes in Fortran 77 – The Art of Scientific Computing*, second ed., Cambridge University Press.

Pronzato, L., H. P. Wynn, and A. A. Zhigljavsky (2000), *Dynamical Search*, CRC Press.

Pronzato, L., H. Wynn, and A. Zhigljavsky (2001), Renormalised steepest descent in Hilbert space converges to a two-point attractor, *Acta Applicandae Mathematicae: An*

International Survey Journal on Applying Mathematics and Mathematical Applications, *67*(1), 1–18.

Rodi, W., and R. L. Mackie (2001), Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, *66*(1), 174–187.

Rokityansky, I. I. (1982), *Geoelectromagnetic Investigation of the Earth's Crust and Mantle*, Springer-Verlag, Berlin, Heidelberg, New York, 381pp.

Romanowicz, B. (1995), A global tomographic model of shear attenuation in the upper mantle, *Journal of Geophysical Research*, *100*, 12,375–12,394, doi:10.1029/95JB00957.

Romanowicz, B. (2001), Can we resolve 3D density heterogeneity in the lower mantle?, *Geophysical Research Letters*, *28*, 1107–1110, doi:10.1029/2000GL012278.

Romanowicz, B., and Y. Gung (2002), Superplumes from the Core-Mantle Boundary to the Lithosphere: Implications for Heat Flux, *Science*, *296*, 513–516, doi:10.1126/science.1069404.

Sambridge, M. (1999a), Geophysical inversion with a neighbourhood algorithm - I. Searching a parameter space, *Geophys. J. Int.*, *138*(2), 479–494.

Sambridge, M. (1999b), Geophysical inversion with a neighbourhood algorithm - II. Appraising the ensemble, *Geophys. J. Int.*, *138*(3), 727–746.

Schmucker, U. (1970), *Anomalies of geomagnetic variations in the southwestern United States, Bull. Scripps. Inst. Oceanogr.*, vol. 13, University of California Press, 165pp.

Schultz, A. (1990), On the vertical gradient and associated heterogeneity in mantle electrical-conductivity, *Phys. Earth Planet. Inter.*, *64*(1), 68–86.

Schultz, A., and J. Larsen (1990), On the electrical-conductivity of the mid-mantle. 2. Delineation of heterogeneity by application of extremal inverse solutions, *Geophys. J. Int.*, *101*(3), 565–580.

Schultz, A., and J. C. Larsen (1987), On the electrical-conductivity of the mid-mantle. 1. Calculation of equivalent scalar magnetotelluric response functions, *Geophys. J. Int.*, *88*(3), 733–761.

Schultz, A., and G. Pritchard (1999), A three-dimensional inversion for large-scale structure in a spherical domain, in *Three Dimensional Electromagnetics, Geophysical Developments Series*, vol. 7, edited by B. Spies and M. Oristaglio, pp. 451–476, Society of Exploration Geophysicists.

Schultz, A., and V. Y. Semenov (1993), Modeling the electrical-conductivity structure of the Earths mantle, *Fiz. Zemli*, *29*(10), 39–43.

Schultz, A., R. D. Kurtz, A. D. Chave, and A. G. Jones (1993), Conductivity discontinuities in the upper-mantle beneath a stable craton, *Geophys. Res. Lett.*, *20*(24), 2941–2944.

Schwarz, G. (1990), Electrical conductivity of the earth's crust and upper mantle, *Surveys in Geophysics*, *11*(2 - 3), 133–161.

Semenov, V. Y., and W. Jozwiak (1999), Model of the geoelectrical structure of the mid- and lower mantle in the Europe-Asia region, *Geophys. J. Int.*, *138*(2), 549–552.

Serghiou, G., A. Zerr, A. Chopelas, and R. Boehler (1998), The transition of pyrope to perovskite, *Physics and Chemistry of Minerals*, *25*(3), 193–196.

Shankland, T. J., J. Peyronneau, and J. P. Poirier (1993), Electrical-conductivity of the Earth's lower mantle, *Nature*, *366*(6454), 453–455.

Shewchuk, J. R. (1994), An introduction to the conjugate gradient method without the agonizing pain, technical report, School of Computer Science, Carnegie Mellon University.

Singer, B. S. (1995), Method for solution of Maxwell's equations in non-uniform media, *Geophysical Journal International*, *120*(3), 590.

Siripunvaraporn, W., and G. Egbert (2000), An efficient data-subspace inversion method for 2-D magnetotelluric data, *Geophysics*, *65*(3), 791–803.

Siripunvaraporn, W., G. Egbert, Y. Lenbury, and M. Uyeshima (2005), Three-dimensional magnetotelluric inversion: data-space method, *Physics of the Earth and Planetary Interiors*, *150*(1-3), 3–14.

Smith, J. T. (1996), Conservative modeling of 3-D electromagnetic fields: 2. Biconjugate gradient solution and an accelerator, *Geophysics*, *61*(5), 1319–1324.

Smith, J. T., and J. R. Booker (1988), Magnetotelluric inversion for minimum structure, *Geophysics*, *53*(12), 1565–1576.

Snieder, R. (1990), A perturbative analysis of nonlinear inversion, *Geophys. J. Int.*, *101*(3), 545–556.

Snieder, R. (1991), An extension of Backus-Gilbert theory to nonlinear inverse problems, *Inverse Problems*, *7*(3), 409–433.

Snieder, R. (1998), The role of nonlinearity in inverse problems, *Inverse Problems*, *14*(3), 387–404.

Stern, D. P. (1994), The art of mapping the magnetosphere, *Journal of Geophysical Research*, *99*, 17,169–+.

Stixrude, L., E. Wasserman, and R. E. Cohen (1997), Composition and temperature of Earth's inner core, *Journal of Geophysical Research*, *102*, 24,729–24,740, doi:10.1029/97JB02125.

Storn, R., and K. Price (1997), Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, *Journal of Global Optimization*, *11*(4), 341–359.

Su, W.-J., and A. M. Dziewonski (1991), Predominance of long-wavelength heterogeneity in the mantle, *Nature*, *352*, 121–126.

Su, W.-J., and A. M. Dziewonski (1992), On the scale of mantle heterogeneity, *Physics of the Earth and Planetary Interiors*, *74*, 29–54, doi:10.1016/0031-9201(92)90066-5.

Su, W.-J., and A. M. Dziewonski (1997), Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Physics of the Earth and Planetary Interiors*, *100*, 135–156.

Su, W.-J., R. L. Woodward, and A. M. Dziewonski (1994), Degree 12 model of shear velocity heterogeneity in the mantle, *Journal of Geophysical Research*, *99*, 6945–6980.

Tarantola, A. (1987), *Inverse Problem Theory*, Elsevier.

Tarantola, A. (2005), *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.

Tarantola, A., and B. Valette (1982), Generalised nonlinear inverse problems solved using the least squares criterion, *Reviews of Geophysics and Space Physics*, *20*(2), 219–232.

Tarits, P., S. Hautot, and F. Perrier (2004), Water in the mantle: Results from electrical conductivity beneath the French Alps, *Geophysical Research Letters*, *31*, 6612–+, doi: 10.1029/2003GL019277.

Thompson, A. B. (1992), Water in the Earth's upper mantle, *Nature*, *358*(6384), 295–302.

Tikhonov, A. N. (1943), Ob ustoichivosti obratnyh zadach (on the stability of inverse problems), *Dokl. Akad. Nauk SSSR*, *39*(5), 195–198, in Russian.

Tikhonov, A. N. (1950), Ob opredelenii electricheskih kharakteristic glubokih sloev zemnoi kory (Determination of the electrical characteristics of the deep strata of the Earth's crust), *Dokl. Akad. Nauk SSSR*, *73*(2), 295–297, in Russian.

Tikhonov, A. N. (1963), Solution of incorrectly formulated problems and the regularisation method, *Sov. Math. Dokl.*, *4*, 1035–1038.

Tikhonov, A. N., and V. Y. Arsenin (1977), *Solutions of Ill-posed Problems*, Winston and Sons, Washington, D.C., 258pp.

Tikhonov, A. N., V. I. Dmitriev, and B. S. Enenstein (1987), Perspectives of development of deep soundings by transient electromagnetic-field, *Vestn. Mosk. Univ. Seriya 1 Mat. Mekhanika*, *295*(5), 1089–1092.

Toh, H., A. Schultz, M. Uyeshima, and I. Fujii (2002), Three-dimensional modelling of the Earth's mantle, *Proceedings of the Ocean Hemisphere Project 2, Tokyo*.

Törn, A., and A. Zilinskas (1989), *Global optimization*, Lecture Notes in Computer Science, Berlin: Springer, 1989.

Turchin, V. F. (1985), Statistical regularisation, *Lecture Notes in Physics*, *236*, 33–49.

Turchin, V. F., V. P. Kozlov, and M. S. Malkevich (1971), The use of mathematical-statistics methods in the solution of incorrectly posed problems, *Soviet Physics Uspekhi*, *13*(6), 681–703.

Utada, H., T. Koyama, H. Shimizu, and A. D. Chave (2003), A semi-global reference model for electrical conductivity in the mid-mantle beneath the North Pacific region, *Geophys. Res. Lett.*, *30*(4), art. no–1194.

Uyeshima, M., and A. Schultz (2000), Geomagnetic induction in a heterogeneous sphere: a new three-dimensional forward solver using a conservative staggered-grid finite difference method, *Geophys. J. Int.*, *140*(3), 636–650.

van der Lee, S., and G. Nolet (1997), Seismic image of the subducted trailing fragments of the Farallon plate, *Nature*, *386*(6622), 266–269.

van der Voo, R., W. Spakman, and H. Bijwaard (1999), Tethyan subducted slabs under India, *Earth and Planetary Science Letters*, *171*, 7–20.

Velimsky, J., and Z. Martinec (2005), Time-domain, spherical harmonic-finite element approach to transient three-dimensional geomagnetic induction in a spherical heterogeneous Earth, *Geophysical Journal International*, *161*(1), 81–101.

Velimsky, J., M. E. Everett, and Z. Martinec (2003), The transient D_{st} electromagnetic induction signal at satellite altitudes for a realistic 3-D electrical conductivity in the crust and mantle, *Geophys. Res. Lett.*, *30*(7), art. no–1355.

Velimsky, J., Z. Martinec, and M. E. Everett (2006), Electrical conductivity in the Earth's mantle inferred from CHAMP satellite measurements; I. data processing and 1-D inversion, *Geophysical Journal International*, *166*(2), 529–542, doi:10.1111/j.1365-246X. 2006.03013.x.

Vozár, J., V. Y. Semenov, A. V. Kuvshinov, and C.-S. Manoj (2006), Updating the map of Earth's surface conductance, *Eos*, *87*(83).

Weaver, J. T. (1994), *Mathematical Methods for Geo-Electromagnetic Induction, Applied and engineering mathematics series*, vol. 7, Wiley.

Weidelt, P. (1972), The inverse problem of geomagnetic induction, *Z. Geophys., 38*, 257–289.

Weiss, C. J., and M. E. Everett (1998), Geomagnetic induction in a heterogeneous sphere: fully three-dimensional test computations and the response of a realistic distribution of oceans and continents, *Geophysical Journal International, 135*(2), 650–662.

Wicht, J., and D. Jault (2000), Electromagnetic core-mantle coupling for laterally varying mantle conductivity, *J. Geophys. Res.-Solid Earth, 105*(B10), 23,569–23,578.

Widiyantoro, S., B. L. N. Kennett, and R. D. van der Hilst (1999), Seismic tomography with P and S data reveals lateral variations in the rigidity of deep slabs, *Earth and Planetary Science Letters, 173*, 91–100.

Woodward, R. L., A. M. Forte, W.-J. Su, and A. M. Dziewonski (1993), *Constraints on the Large-Scale Structure of the Earth's Mantle*, pp. 89–+, Evolution of the Earth and Planets.

Xu, Y. S., B. T. Poe, T. J. Shankland, and D. C. Rubie (1998), Electrical conductivity of olivine, wadsleyite, and ringwoodite under upper-mantle conditions, *Science, 280*(5368), 1415–1418.

Xu, Y. S., T. J. Shankland, and B. T. Poe (2000), Laboratory-based electrical conductivity in the Earth's mantle, *J. Geophys. Res.-Solid Earth, 105*(B12), 27,865–27,875.

Xu, Y. S., T. J. Shankland, and B. T. Poe (2003), Laboratory-based electrical conductivity in the Earth's mantle (vol 108, pg 2314, 2003), *J. Geophys. Res.-Solid Earth, 108*(B6), art. no-2314.

Yoshimura, R., and N. Oshiman (2002), Edge-based finite element approach to the simulation of geoelectromagnetic induction in a 3-D sphere, *Geophysical Research Letters*, *29*, 9–1.

Young, C. J., and T. Lay (1987), The Core-Mantle Boundary, *Annual Review of Earth and Planetary Sciences*, *15*, 25–+, doi:10.1146/annurev.ea.15.050187.000325.

Zelt, C. A. (1998), Lateral velocity resolution from three-dimensional seismic refraction data, *Geophysical Journal International*, *135*(3), 1101–1112, doi:10.1046/j.1365-246X.1998.00695.x.

Zhang, T. s. (1991), Three-dimensional electromagnetic induction in a spherical Earth, Ph.D. thesis, Geophysics Program, University of Washington.

Zhang, T. S., and A. Schultz (1990), Exorcise - an algorithm for detection of spurious values and prediction of missing data, *Comput. Geosci.*, *16*(8), 1027–1065.

Zhang, T. S., and A. Schultz (1992), A 3-D perturbation solution for the EM induction problem in a spherical Earth - the forward problem, *Geophys. J. Int.*, *111*(2), 319–334.

Zhao, D. (2004), Global tomographic images of mantle plumes and subducting slabs: insight into deep Earth dynamics, *Physics of the Earth and Planetary Interiors*, *146*, 3–34, doi:10.1016/j.pepi.2003.07.032.

Zharkov, V. N. (1986), *Interior Structure of the Earth and Planets*, Harwood Academy Press, London.

Zhdanov, M. S. (1986), *Electorazvedka (Geoelectrics)*, Nedra, Moscow.

Zhigljavsky, A. A. (1991), *Theory of Global Random Search, Mathematics and Its Applications (Soviet Series)*, vol. 65, Kluwer Academic Publishers.