Joint inversion of magnetotelluric and receiver function data

Ulf Max Stefan Moorkamp

A thesis submitted for the degree of Doctor of Philosophy
to the National University of Ireland

Research conducted at the
Dublin Institute for Advanced Studies, Ireland
and
Department of Earth and Ocean Sciences, National University
of Ireland, Galway

Supervisors: Prof. Alan G. Jones and Dr. Colin Brown

August 2007

Contents

Acknowledgements

Summary

This thesis falls into two separate parts that, although mathematically related, deal with very different problems in geophysical surveys. The first part is concerned with improving the quality of magnetotelluric sounding curves that are affected by high-amplitude, non-stationary noise from electric cow-fences. We present a number of different methods that have been proposed in the literature and that we have developed and the ability of both to deal with this difficult task. We find that LMS-adaptive filters have a low computational complexity and produce results that, at least for one polarization, have the properties of a regular sounding curve. For the other polarization, however, we do not achieve any improvement with any of the methods.

In the second, and main, part of the thesis we present a joint inversion algorithm for long-period magnetotelluric and receiver function data. Our inversion approach is based on a genetic algorithm for two reasons: The stochastic nature of the genetic algorithm avoids premature convergence to a local minimum, and from the inversion we obtain the trade-off curve that displays the extent to which the objective functions compete with each other. Our tests with synthetic datasets show that the structure of the trade-off curve can be used to assess the compatibility of the seismic and electric parts of the joint model. We explore different types of regularization and outline strategies to find a representative model of the subsurface. Finally, we apply our joint inversion algorithm to data from the Slave craton, where previous studies have postulated a correlation between seismic conversions and the location of a conductor in the upper mantle. Our joint inversion results indicate that we can model this upper-mantle structure as a correlated seismic and electric feature and that we obtain improved results for other more minor structures.

*The equation $a^n + b^n = c^n$ has solutions in positive integers a,b,c
and n only when $n = 2$ (and then there are infinitely many triplets
a,b,c which satisfy the equation); but there are no solutions for
$n > 2$. I have discovered a truly marvelous proof of this statement,
which, unfortunately, this margin is too small to contain.*

P. de Fermat

1

Introduction

Even with the great improvement of data quality and inversion methods in the last
20 years, the resolution of geophysical data due to noise and finite sampling limits
the level of detail at which we can image the subsurface of the Earth. Currently
there are two popular approaches to improve our understanding of the Earth's
crust and mantle. The first possibility is to increase the amount of measured data
of a certain type, both in terms of covered area and site spacing, and model these
data with three-dimensional codes. With this approach we can make sure that we
get the maximum amount of information from the data, and avoid artifacts from
spatial aliasing or simplifying assumptions. However, this does not circumvent
the fundamental resolution of the chosen method. The alternative approach is
to combine different datasets in the inversion process and create a joint model.
With a well-chosen combination of datasets we can hope to benefit from different
sensitivities of each set to improve the resulting model beyond a simple superpo-
sition of the individual models [Vozoff and Jupp, 1975]. One example of such an
approach is the joint inversion of surface waves and receiver functions [Julia et al.,
2000]. Surface waves are sensitive to bulk seismic velocity, but do not have good
sensitivity to the position of interfaces. Receiver functions, in contrast, primar-
ily sense seismic interfaces, but only sense velocities to a smaller degree. In this
example both datasets were sensitive to the same parameters, but we can even
go a step further and combine datasets that are sensitive to different parameters
[Meju et al., 2003, Linde et al., 2006, Dal Moro and Pipan, 2007], and this is what
we are attempting by combining magnetotelluric (MT) and receiver function (RF)
data.

1

At first the benefit of such an approach might not be clear. Magnetotelluric data senses the distribution of electrical conductivity, while receiver function data senses abrupt changes in seismic velocity. In the Earth's crust and mantle changes in these two parameters do not necessarily have to be spatially correlated. In fact a lot of the conductive anomalies, the main targets of MT surveys, are attributed to minor constituents of the rock matrix such as graphite, sulfide, or fluids, which we cannot expect to detect with seismic methods [Jones, 1992, Ji et al., 1996, Bahr et al., 2002]. On the other hand there are a number of physical parameters that affect both electrical conductivity and seismic velocity, such as temperature and rock composition. Furthermore we can expect that at major lithological boundaries both parameters change, and even where we have an electrical anomaly that is caused by graphite, for example, we would think that the extent of this graphitic body is determined by tectonic processes that, in turn, also govern the seismic structure of that area. Finally there have been observations of coincident changes in seismic velocity and electrical conductivity in the mantle below the Slave craton [Snyder et al., 2004] and this will also be the area where we apply our joint inversion approach. Furthermore the absence of correlation can also provide valuable information [Cook and Jones, 1995, Jones, 1998].

For these reasons we can expect some structural relationship between the seismic and electric parts of our joint model, but we have to allow for structures that are only present in one model. We will pursue this issue further when we discuss the coupling of the individual models. Also we have to be prepared to find two fundamentally different structures for the seismic and electrical models, in case our parametrisation does not permit sufficient flexibility to jointly model both datasets. This is one of the motivations for our genetic algorithm approach and we will discuss this issue in detail in our synthetic test examples. Once we have identified common interfaces and interfaces where the models do not allow coincident changes, we can use this additional information to characterise the structure of the Earth and obtain insight into the geological processes that formed these structures. Here our focus is on the methodology, though, and we will not discuss the implications of our results.

Although the focus of this thesis is on "Joint inversion of magnetotelluric and receiver function data", a considerable fraction of the presentation and the work we performed during the last three years is devoted to noise removal and signal processing of magnetotelluric time-series. The reason for this dichotomy lies in the poor quality of the data that we recorded at the beginning of this project in Ireland. The original idea of the magnetotelluric component of the Irish Lithospheric Experiment (ISLE) [Landes et al., 2004], from here on called ISLE-MT, was to augment each ISLE seismic station with an MT site, and therefore cre-

ate ideal conditions for the joint inversion experiment. It quickly became clear that at virtually all sites the high frequency part of the magnetotelluric data is contaminated by high amplitude interferences from electric cow-fences. At a few sites even the long-period data are disturbed by cultural noise. This motivates our attempts to remove these high amplitude disturbances that resist all modern processing algorithms.

We have two choices for the presentation of our results: From a mathematical point of view the two topics of adaptive time-series filtering and inversion are sub-disciplines of the broader field of optimisation. In both cases we seek to minimise the difference between a quantity calculated from the data and by some other means, e.g. a model, another time series or some other segment of the same time series. Therefore filtering and inversion should be presented together in order to understand the similarities and differences between the two. From a practical point of view filtering aims at providing reliable estimates of the impedance tensor, while inversion produces models from this impedance tensor, but without asking where the estimates come from. Usually signal processing is the first step after data acquisition, while modelling and model appraisal are the very last steps and a number of other steps lie in between. A presentation that follows the time-line of the researcher, from data acquisition to final model, better displays the motivation for dealing with these issues and gives answers to questions when they arise.

As a solution to this dilemma we will present the results in the order that they appear in practice. To satisfy partially those readers who seek to find the mathematical connections, we will try to draw as many parallels as possible between filtering and inversion, even if this means referring to a topic that will only be discussed later. We will therefore start with a discussion of the basic principles of electromagnetic induction and its application to magnetotelluric soundings. This leads directly to the question of how we calculate the impedance tensor elements and the problematic nature of the ISLE-MT recordings. After characterising the problems we will present our attempts to solve the interference issues with a number of different signal processing techniques. The first two chapters are self-contained, and most of the issues discussed there will not reappear when we present our joint inversion approach. Hence we will give a summary and discussion of the signal processing results at the end of Chapter 3.

Before we discuss our joint inversion method, we present the basic theory for the receiver function data. Apart from the methodological similarities between inversion and adaptive signal processing, this, and the following chapters, form another self-contained unit. The discussion of the receiver function method is followed by the presentation of the joint inversion algorithm. We will contrast traditional linearised methods with our genetic algorithm based approach and

demonstrate the advantages of genetic algorithms for joint inversion. Through the examination of some synthetic test problems we will highlight the capabilities and limitations of joint inversion for MT and receiver function data. The application of the algorithm to measured data from the Slave craton together with a discussion of the results and recommendations for further work concludes this part of the thesis.

2

The basic principles of magnetotelluric induction

In this chapter we will describe the steps required from recording the time-series in the field to obtaining the MT transfer function or impedance tensor that is subsequently used in analysis and modelling. We will start with a theoretical introduction that motivates the calculation of the MT impedances and shows how they are related to the properties of the Earth. This theoretical description will be limited to the one-dimensional case for two reasons: First, the basic concepts are more clearly understood without the complicated burden of a 2D or 3D treatment, and second, for most of this thesis we will only be concerned with the assumption of one-dimensional structure within the Earth. This is, of course, a gross oversimplification and we will refer to current state of the art methods were necessary.

In order to lay the foundation for the analysis of measured data in Chapter 7, we will finish this chapter with a discussion of different measures of electromagnetic dimensionality. These quantities can be calculated from the impedance tensor in order to assess whether a 1D, 2D, or even 3D description is most likely to be needed in order to explain the observations.

2.1 From Maxwell's equation to EM induction

The behaviour of any electromagnetic field is described by Maxwell's equations, viz.,

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \tag{2.1}$$

$$\nabla \times \mathbf{H} = \mathbf{j} + \frac{\partial \mathbf{D}}{\partial t}, \tag{2.2}$$

$$\nabla \cdot \mathbf{B} = 0, \tag{2.3}$$

$$\nabla \cdot \mathbf{D} = q. \tag{2.4}$$

Here $\mathbf{E}$ is the electric field in V/m, $\mathbf{B}$ the magnetic induction in T, $\mathbf{H}$ is the magnetic field strength in A/m, $\mathbf{j}$ is electric current density in A/m^2, $\mathbf{D}$ is the electric displacement in C/m^2 and q is the charge density in C/m^3.

To derive and solve the induction equation that forms the theoretical basis of magnetotellurics, we have to make a number of assumptions:

1. All electromagnetic sources are outside the Earth and produce uniform plane-polarized electromagnetic waves. For a one-dimensional conductivity distribution inside the Earth Dmitriev and Berdichevsky [1979] showed that a linearly varying field also fulfils the necessary conditions. These waves arrive at the Earth's surface at a near vertical angle. This assumption is a arguably the most important for magnetotellurics and certainly its most problematic. Since its publication by Cagniard [1953] and Tikhonov [1950] its validity has been discussed [Price, 1962, Madden and Nelson, 1963]. We will address this problem in more detail below.

2. The Earth acts as an ohmic conductor so that current density $\mathbf{j}$ and electric field $\mathbf{E}$ are linearly related , through Ohm's Law, by the scalar electrical conductivity σ,

$$\mathbf{j} = \sigma \mathbf{E}. \tag{2.5}$$

3. For Earth materials we can assume the conventional linear relationships between $\mathbf{D}$ and $\mathbf{E}$ and $\mathbf{B}$ and $\mathbf{H}$ and for magnetotellurics we neglect changes in electrical permitivities ε and magnetic permeabilities μ so that

$$\mathbf{B} = \mu_0 \mathbf{H} \text{ and } \mathbf{D} = \varepsilon_0 \mathbf{E}. \tag{2.6}$$

4. Displacement currents $\partial \mathbf{D}/\partial t$ are assumed to be negligible. Most magnetotelluric surveys use frequencies between 100,000 - 0.0001 Hz. Given typical resistivities of the Earth of $10^6 - 10^{-2}\,\Omega$m we can compare the terms in

equation 2.2,

$$\frac{\partial \mathbf{D}/\partial \mathbf{t}}{\mathbf{j}} = \frac{i\omega\varepsilon_0}{\sigma} \approx 0.1 - 10^{-19}. \tag{2.7}$$

For the highest frequencies in extremely resistive areas displacement currents are an order of magnitude smaller than the current density and we would start to see some effects from these currents. Typically both the measurement frequencies and the resistivities will be much smaller and we will not see any effect.

Under these assumptions the resulting equations are:

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial \mathbf{t}}, \tag{2.8}$$
$$\nabla \times \mathbf{B} = \mu_0\sigma\mathbf{E}, \tag{2.9}$$
$$\nabla \cdot \mathbf{B} = 0, \tag{2.10}$$
$$\nabla \cdot \mathbf{E} = \frac{q}{\varepsilon_0}. \tag{2.11}$$

We can transform these equations from the time domain into the frequency domain, and simplify the solution by replacing time derivatives with simple multiplications, i.e.,

$$-\frac{\partial \mathbf{B}(\mathbf{t})}{\partial \mathbf{t}} \xrightarrow{\mathcal{F}} -\imath\omega\mathbf{B}(\omega), \tag{2.12}$$

for a field at frequency ω.

We can use the fact that $\nabla \cdot (\nabla \times \mathbf{A}) = 0$ for any vector field $\mathbf{A}$. Applying the divergence to equation 2.9 yields

$$\mu_0\nabla \cdot (\sigma\mathbf{E}) = \mu_0 (\sigma\nabla \cdot \mathbf{E} + \mathbf{E}\nabla\sigma) = 0. \tag{2.13}$$

From equation 2.11 then follows that

$$\frac{q}{\varepsilon_0} = -\mathbf{E}\frac{\nabla\sigma}{\sigma} \Rightarrow \frac{q}{\varepsilon_0} = -\mathbf{E}\nabla\ln\sigma. \tag{2.14}$$

This results in modified Maxwell's equations:

$$\nabla \times \mathbf{E} = -\imath\omega\mathbf{B}, \tag{2.15}$$
$$\nabla \times \mathbf{B} = \mu_0\sigma\mathbf{E}, \tag{2.16}$$
$$\nabla \cdot \mathbf{B} = 0, \tag{2.17}$$
$$\nabla \cdot \mathbf{E} = -\mathbf{E}\nabla\ln\sigma. \tag{2.18}$$

These four equations can be combined into two equations by taking the curl of 2.15 and 2.16

$$\nabla \times \nabla \times \mathbf{B} = \nabla(\nabla \cdot \mathbf{B}) - \nabla^2\mathbf{B} = \mu_0(\sigma\nabla \times \mathbf{E} - \mathbf{E} \times \nabla\sigma), \quad (2.19)$$

$$\Rightarrow \nabla^2\mathbf{B} = \imath\omega\mu_0\sigma\mathbf{B} + \mu_0\mathbf{E} \times \nabla\sigma, \quad (2.20)$$

$$\nabla \times \nabla \times \mathbf{E} = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2\mathbf{E} = -\imath\omega\nabla \times \mathbf{B}, \quad (2.21)$$

$$\Rightarrow \nabla^2\mathbf{E} = \imath\omega\mu_0\sigma\mathbf{E} - \nabla(\mathbf{E}\nabla\ln\sigma). \quad (2.22)$$

Equations 2.20 and 2.22 describe the magnetic and electric fields under the assumptions above in their most general form. To explain the basic concepts of magnetotelluric we have to simplify the situation further by first assuming a homogeneous halfspace.

2.2 The homogeneous halfspace

For a homogeneous halfspace the conductivity is constant σ_0 inside the Earth and consequently $\nabla \cdot \sigma = 0$. Equations 2.20 and 2.22 then simplify to two uncoupled diffusion equations of the form,

$$\nabla^2\mathbf{F} = \mu_0\sigma\imath\omega\mathbf{F} \text{ with } \mathbf{F} = \mathbf{E}, \mathbf{B}, \quad (2.23)$$

with solutions,

$$\mathbf{F} = \mathbf{F}_1\exp(\imath\omega t - qz) + \mathbf{F}_2\exp(\imath\omega t + qz) \text{ and } q^2 = \mu_0\sigma\imath\omega. \quad (2.24)$$

As we assume that all sources are outside the Earth $\mathbf{F}_2 = 0$ for both the electric and magnetic field. For a uniform halfspace the electromagnetic field inside the Earth is described by

$$\mathbf{E} = \mathbf{E}_0\exp(\imath\omega t - \sqrt{\mu_0\sigma\imath\omega}z), \quad (2.25)$$

$$\mathbf{B} = \mathbf{B}_0\exp(\imath\omega t - \sqrt{\mu_0\sigma\imath\omega}z). \quad (2.26)$$

Inserting the two solutions into equation 2.15 yields

$$\nabla \times \mathbf{E} = \frac{\partial E_x}{\partial z} = -\sqrt{\mu_0\sigma\imath\omega}E_x, \quad (2.27)$$

$$= -\imath\omega B_y. \quad (2.28)$$

From this we can finally derive the basic equation for MT in a uniform halfspace

$$C = \frac{1}{q} = \frac{1}{\sqrt{\mu_0\sigma\imath\omega}} = \frac{E_x}{\imath\omega B_y} = -\frac{E_y}{\imath\omega B_x}. \quad (2.29)$$

The quantity C is know as the Schmucker-Weidelt transfer function [Weidelt, 1972, Schmucker, 1973]. It's direct connection to the differential equations makes it a usefull quantity in the theoretical examination of magnetotelluric problems. Equation 2.29 reveals that in this case the resistivity of the halfspace can be calculated from the ratio of the orthogonal components of the electric and magnetic fields at a given period. The realisation of this fact was the motivation for early magnetotelluric surveys,

$$\rho = \frac{1}{\sigma} = \mu_0 \omega |C|^2. \tag{2.30}$$

In practice instead of C the *magnetotelluric impedance* Z is used

$$Z = \frac{E_x}{H_y} = \frac{\mu_0 E_x}{B_y} = \iota \omega \mu_0 C \quad [\Omega]. \tag{2.31}$$

or alternatively as

$$Z = \frac{E_x}{B_y} = \iota \omega C \quad [m/s]. \tag{2.32}$$

The two definitions are known as "field-units" and "SI-units" conventions, respectively. Before we turn to slightly more realistic Earth models and the behaviour of the magnetotelluric impedance in those cases, we return to the solutions of the diffusion equations to look at the depth to which the fields penetrate. From 2.25 and 2.26 we see that the real part of the solution decays exponentially with depth. The depth where the field strength is reduced to $1/e$ is called the *skin-depth*, p, given by

$$p = \sqrt{\frac{2}{\mu_0 \sigma \omega}} \quad [m]. \tag{2.33}$$

Although this equation is only exact for a homogeneous half-space, it is often used to get a first idea of the penetration at a given frequency even in more complex environments. However, depending on the geometry of the structures, these estimates can be highly devious [Jones, 1983]

2.3 Layered half-space

Obviously the real Earth is far from being a homogeneous half-space. A first step toward a more realistic model is to allow for a number of layers of different thicknesses t_i and conductivities σ_i. To derive the expression for the magnetotelluric transfer function in this case we loosely follow Simpson and Bahr [2005]. Assuming N layers indexed $n = 1 \ldots N$ the solution of the diffusion equation 2.23 in each layer is still of the form 2.24, but now both terms do not vanish. For the electric field we have both upward and downward travelling waves, viz.,

$$E_x^n = E_1^n \exp(\iota \omega t - q_n z) + E_2^n \exp(\iota \omega t + q_n z), \qquad q_n = \sqrt{\mu_0 \sigma_n \iota \omega}, \tag{2.34}$$

except in the lowermost half-space, where the wave only travels downward.

We can combine the time varying exponential with the coefficients E_1^n and E_2^n, respectively, to a new set of coefficients a_n and b_n,

$$E_x^n = a_n \exp(-q_n z) + b_n \exp(+q_n z). \tag{2.35}$$

From equation 2.15 follows for the magnetic field

$$B_y^n = \frac{q_n}{\iota\omega} \left(a_n \exp(-q_n z) - b_n \exp(+q_n z) \right), \tag{2.36}$$

thus the transfer function C_n inside layer n is given by

$$C_n(z) = \frac{E_x^n}{\iota\omega B_y^n} = \frac{a_n \exp(-q_n z) + b_n \exp(+q_n z)}{q_n \left(a_n \exp(-q_n z) - b_n \exp(+q_n z) \right)}. \tag{2.37}$$

At the top of the layer $z = z_{n-1}$ we have

$$C_n(z_{n-1}) = \frac{E_x^n}{\iota\omega B_y^n} = \frac{a_n \exp(-q_n z_{n-1}) + b_n \exp(+q_n z_{n-1})}{q_n \left(a_n \exp(-q_n z_{n-1}) - b_n \exp(+q_n z_{n-1}) \right)}, \tag{2.38}$$

and similar at the bottom $z = z_n$

$$C_n(z_n) = \frac{E_x^n}{\iota\omega B_y^n} = \frac{a_n \exp(-q_n z_n) + b_n \exp(+q_n z_n)}{q_n \left(a_n \exp(-q_n z_n) - b_n \exp(+q_n z_n) \right)}. \tag{2.39}$$

We can rewrite 2.39 as

$$a_n = b_n \frac{1 + q_n C_n(z_n)}{q_n C_n(z_n) - 1} \exp(2 q_n z_n), \tag{2.40}$$

and insert the result into 2.38 to eliminate a_n and b_n,

$$
\begin{aligned}
C_n(z_{n-1}) &= \frac{\dfrac{1 + q_n C_n(z_n)}{q_n C_n(z_n) - 1} \exp(2 q_n z_n) \exp(-q_n z_{n-1}) + \exp(q_n z_{n-1})}{q_n \left(\dfrac{1 + q_n C_n(z_n)}{q_n C_n(z_n) - 1} \exp(2 q_n z_n) \exp(-q_n z_{n-1}) - \exp(q_n z_{n-1}) \right)}, \\[2ex]
&= \frac{\dfrac{1 + q_n C_n(z_n)}{q_n C_n(z_n) - 1} \exp(q_n(z_n - z_{n-1})) + \exp(-q_n(z_n - z_{n-1}))}{q_n \left(\dfrac{1 + q_n C_n(z_n)}{q_n C_n(z_n) - 1} \exp(q_n(z_n - z_{n-1})) - \exp(-q_n(z_n - z_{n-1})) \right)}, \\[2ex]
&= \frac{(1 + q_n C_n) \exp(q_n(z_n - z_{n-1})) + (q_n C_n - 1) \exp(-q_n(z_n - z_{n-1}))}{q_n \left((1 + q_n C_n) \exp(q_n(z_n - z_{n-1})) - (q_n C_n - 1) \exp(-q_n(z_n - z_{n-1})) \right)}, \\[2ex]
&= \frac{1}{q_n} \frac{\sinh(q_n(z_n - z_{n-1})) + q_n C_n \cosh(q_n(z_n - z_{n-1}))}{\cosh(q_n(z_n - z_{n-1})) + q_n C_n \sinh(q_n(z_n - z_{n-1}))}, \\[2ex]
&= \frac{1}{q_n} \frac{q_n C_n(z_n) + \tanh(q_n(z_n - z_{n-1}))}{1 + q_n C_n(z_n) \tanh(q_n(z_n - z_{n-1}))}. \tag{2.41}
\end{aligned}
$$

Both horizontal electric and magnetic field are continuous across the layer boundary, hence their ratio has to be continuous as well and consequently

$$C_n(z_n) = C_{n+1}(z_n). \tag{2.42}$$

Using the continuity requirement in 2.41 finally results in the *Wait recursion formula* [Wait, 1954]

$$C_n(z_{n-1}) = \frac{1}{q_n} \frac{q_n C_{n+1}(z_n) + \tanh\left(q_n(z_n - z_{n-1})\right)}{1 + q_n C_{n+1}(z_n) \tanh\left(q_n(z_n - z_{n-1})\right)}. \tag{2.43}$$

This recursion formula can be used to calculate the transfer function at the top of a N-layer model. The lowermost layer is a half-space with $C_N(z_N) = 1/q_N$. This solution is then propagated through the remaining layers from bottom to top using the recursion formula.

For a layered half-space no direct transformation between the transfer function or impedance and layer thickness and resistivities exists; this issue will be discussed in detail later. Equation 2.30, which is exact in the case of a homogeneous half-space, is now replaced by an *apparent resistivity*, the resistivity of an equivalent half-space, viz.,

$$\rho_a = \frac{1}{\mu_0 \omega} |Z(\omega)|^2 \text{ for Z in S.I. units.} \tag{2.44}$$

In addition the impedance *phase* is another useful quantity

$$\phi = \tan^{-1}\left(\frac{\Im Z}{\Re Z}\right). \tag{2.45}$$

In terms of the electromagnetic fields it represents the phase-difference between the magnetic and electric field. For a homogeneous half-space

$$Z = (1 + \imath)\sqrt{\frac{\omega \mu_0}{2\sigma}} \Rightarrow \phi = \tan^{-1} 1 = \frac{\pi}{4}. \tag{2.46}$$

Apparent resistivity and phase are an equivalent representation of the magnetotelluric impedance and are usually used to display and interpret magnetotelluric data because of their more intuitive physical relationship to the properties of the Earth. In a layered Earth and for the TM-mode in a two-dimensional environment, these two quantities are interrelated through a Hilbert transformation [Weidelt, 1972, Weidelt and Kaikkonen, 1994, Berdichevsky, 1999] and hence do not provide independent information about Earth structure. [Cavaliere and Jones, 1984] showed though that for real data the different sensitivities of apparent resistivity and phase help to constrain the conductivity in the subsurface.

2.4 Two-dimensional structures

If the conductivity varies in the vertical and one horizontal direction a single impedance that describes the relationship of the electric and magnetic field no longer exists. Instead two separate *modes* of the fields have to be considered. From equations 2.15 and 2.16 we can see that for structures aligned with the x-direction we get two uncoupled sets of differential equations

$$\frac{\partial E_x}{\partial y} = \iota\omega B_z, \tag{2.47}$$

$$\frac{\partial E_x}{\partial z} = \iota\omega B_y, \text{ and} \tag{2.48}$$

$$\frac{\partial B_z}{\partial y} - \frac{\partial B_y}{\partial z} = \mu_0\sigma E_x, \tag{2.49}$$

is called the *transverse electric* (TE) mode or *E-polarisation* and couples E_x with B_z and B_y. Conversely,

$$\frac{\partial B_x}{\partial y} = \mu_0\sigma E_y, \tag{2.50}$$

$$-\frac{\partial B_x}{\partial z} = \mu_0\sigma E_y, \text{ and} \tag{2.51}$$

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = \iota\omega B_x, \tag{2.52}$$

is called the *transverse magnetic* (TM) mode or *B-polarisation* and couples E_y and E_z with B_x. Analytic solutions for the TM-mode for 2D structures exist only in a few special cases such as two quarter-spaces or a dike. Hobbs [1975] gives an overview of analytical solutions for electromagnetic induction [see also Weaver et al., 1985]. For the TE-mode quasi-analytical solutions exist for some geometries [Weaver et al., 1986]. Despite the more complicated nature of the problem, the impedance in this case is defined as a straightforward extension of the one-dimensional case. We now have two impedance values,

$$Z_{xy} = \frac{\mu_0 E_x}{B_y} \text{ and } Z_{yx} = \frac{\mu_0 E_y}{B_x}, \tag{2.53}$$

with corresponding apparent resistivities and phases.

2.5 The 3D case

In the most general case, the electric and magnetic fields are described by equations 2.20 and 2.22. Particularly, the corresponding electric and magnetic fields are no

longer orthogonal two each other. Thus the concept of impedance is extended to the magnetotelluric *impedance tensor*, that describes the relation between the fields

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_x \\ H_y \end{pmatrix}. \tag{2.54}$$

Both the one-dimensional and two-dimensional cases can be regarded as special cases of the impedance tensor.

$$Z_{xx} = Z_{yy} = 0 \qquad Z_{xy} = -Z_{yx} \quad \text{in 1D} \tag{2.55}$$

$$Z_{xx} = Z_{yy} = 0 \qquad\qquad\qquad\quad \text{in 2D} \tag{2.56}$$

when the x-axis is directed along the strike of the 2D structure. For this reason all modern processing algorithms estimate the full impedance tensor. Only in the subsequent analysis different criteria are used to classify the electrical structure as 1D, 2D or 3D.

2.6 Dimensionality and invariants

We will now discuss some of the criteria used to classify the dimensionality of MT data. To date, most data from magnetotelluric surveys are interpreted in terms of two-dimensional profiles, assuming infinite uniformity in the direction perpendicular to the profile. In practice, the length scale of structures is often larger in one direction than the length scale of induction, but it is essential to quantify at which frequencies and sites this assumption holds, and where it is violated. Only then can we assess the reliability of models constructed this way. To this end a large number of classification schemes have been constructed to quantify the dimensionality, and extract the parts of the data appropriate for a chosen approximation [e.g. Groom and Bailey, 1989, Bahr, 1991, Groom and Bahr, 1992, Weaver et al., 2000, Caldwell et al., 2004, Martí et al., 2005]. We will concentrate here on indicators for one dimensional structures that we will need later when we model real data.

The most straightforward indication for 1D Earth structures near a measured site is directly given by Equation 2.55. If this equation holds within data error, a layered Earth model can be assumed to be appropriate. One thing to bear in mind with all classification schemes is that they all define necessary, but not sufficient, conditions. Therefore, the violation of a classification condition indicates that the chosen model is not appropriate, but the opposite does not prove its validity. This is a serious problem when modelling MT data, as one can never be completely sure whether the chosen approximation is correct. Without further information however, we do not have any better criterion to guide our decision to model the

data in a certain way. The generally accepted paradigm is to use the simplest model appropriate. This is also known as *Occam's razor*.

Even when the regional Earth structure can be adequately described by a plane layered model, we often observe violations of Equation 2.55. This violation is caused by *galvanic distortion* due to small inhomogeneities in the vicinity of the measurement site [Groom and Bailey, 1989, Bahr, 1991, Singer, 1992]. The effect of these inhomogeneities can be described by multiplication of the impedance tensor with a frequency independent, real-valued matrix

$$\mathbf{Z}_{\text{dist}} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \mathbf{Z}. \tag{2.57}$$

The complications that arise from this effect, and the fact that the entries of the *distortion matrix* $\mathbf{a}$ cannot be uniquely determined [Bibby et al., 2005] are a long standing problem for MT modelling and interpretation.

Fortunately the multiplication with a real valued matrix does not change the phase of the off-diagonal elements, if the impedance tensor is in the form of Equation 2.55. This enables us to use the information contained in the phase in situations where the magnitude of the impedance is affected by galvanic distortion. We define two measures of one-dimensionality:

$$\Sigma = \frac{D_1^2 + S_2^2}{|D_2|} \text{ with } D_1 = Z_{xx} - Z_{yy}, D_2 = Z_{xy} - Z_{yx} \text{ and } S_2 = Z_{xy} + Z_{yx} \tag{2.58}$$

is small when a layered Earth model without distortion is appropriate [Swift, 1967]. In the presence of galvanic distortion Σ can assume large values, as it is based on the magnitude of the impedance elements. When $\Sigma > 0.1$, we use a rotationally invariant measure of the phase difference,

$$\mu = \frac{\sqrt{|[D_1, S_2]| + |[S_1, D_2]|}}{|D_2|} \tag{2.59}$$

[Bahr, 1991]. Here $S_2 = Z_{xy} + Z_{yx}$, $D_1 = Z_{xx} - Z_{yy}$ and $[\cdot, \cdot]$ denotes the commutator. If both κ and μ are significantly different from zero, the data requires a 2D or 3D model approach to be explained fully. Caldwell et al. [2004] point out that when the enumerator of Equation 2.59 does not vanish, this dimensionality measure is also affected by galvanic distortion, as $|D_2|$ depends on the amplitude of the impedance tensor elements. When we determine the dimensionality of measured data, we therefore use a combination of indicators as suggested by Martí et al. [2005].

Even when the data requires a more complex model, it can be instructive to construct 1D models to get a first idea of the conductivity distribution. The Wait

algorithm only yields a single impedance value, as expected for a 1D Earth, so we have to condense the four elements of the observed impedance tensor into a single value that provides an appropriate approximation. Rotational invariants provide a convenient way to summarize the information in the impedance tensor without regard for the coordinate system of measurement. Two invariants are commonly used in modelling MT data: The arithmetic mean of the off-diagonal elements

$$Z_B = \frac{Z_{xy} - Z_{yx}}{2}, \qquad (2.60)$$

also known as the *Berdichevkiy invariant* [Berdichevskiy and Dmitriev], and the determinant

$$Z_D = Z_{xx}Z_{yy} - Z_{xy}Z_{yx}. \qquad (2.61)$$

The determinant has the advantage that it combines information from all 4 elements of the impedance tensor, although modelling studies show that the difference between the two for practical purposes is not significant in one-dimensional inversion [Park and Livelybrooks, 1989]. Pedersen and Engels [2005] advocate the use the determinant in two-dimensional inversion to incorporate information from the diagonal elements. In any case care has to be taken in the interpretation of results from modelling these quantities. Near 3D inhomogeneities, where the 1D approximation breaks down, models of these invariants can contain artificial resistors or conductors [Park and Livelybrooks, 1989]. Therefore a careful analysis of the data is needed before modelling it.

3

From magnetotelluric time series to transfer function

3.1 Processing magnetotelluric data, a short overview

As outlined in Section 2.1, in theory the transfer function between the horizontal electric and magnetic fields depends only on the conductivity distribution of the Earth. In an ideal situation the calculation of the transfer function from simultaneous recordings of the electric and magnetic field is simple and straightforward.

1. Fractionalize the recording in segments of equal length, the length of the segments depends on the longest period needed.

2. Multiply each segment with a suitable window function, e.g., the *Hamming window*, to avoid spectral leakage.

3. Fourier transform each segment.

4. Calculate the auto- and cross spectra for each segment.

5. Calculate the mean and the error from the various spectral estimates.

6. Calculate the resulting transfer function and its error.

To illustrate the basic principles, we will describe the most basic approach to impedance tensor estimation in some detail, before we discuss more advanced methods.

16

3.1.1 The anatomy of modern processing algorithms

As the impedance tensor is defined in the frequency domain, the first step is to obtain spectra for the recorded fields. A number of methods exist to estimate the spectrum of a time series. The most basic spectral estimate is the *periodogram*, the mean of the Fourier transform of different segments of the time-series,

$$S(\omega_i) = \sum_{j=0}^{N/L-1} \sum_{l=j \cdot L}^{(j+1) \cdot L-1} x_l \exp(-\imath \omega_i l) \quad i = 1 \ldots L-1. \tag{3.1}$$

It is easy to implement and computationally fast, but suffers from spectral bias problems because the segmentation of the time series is equivalent to multiplication with a boxcar function. The result in the frequency domain is convolution with a sinc function, and has numerous side-lobes that cause spectral leakage. This situation can be improved by applying a *window function* before calculating the Fourier transform,

$$S(\omega_i) = \sum_{j=0}^{N/L-1} \sum_{l=j \cdot L}^{(j+1) \cdot L-1} x_l w(l) \exp(-\imath \omega_i l) \quad i = 1 \ldots L-1. \tag{3.2}$$

The inductive nature of magnetotelluric soundings makes spectral resolution often a minor issue and the choice of spectral window is usually not critical. Popular choices are the *Hann window* and the *Hamming window*. When the time-series is contaminated with noise that is concentrated in spectral lines, the choice of spectral estimation method can make a noticeable difference. In these cases *multitaper* spectral analysis [Thomson, 1982] can produce more accurate results. The simple windowing functions are replaced by specially designed orthogonal taper functions that optimize spectral leakage and variance. Multitaper analysis is used in the bounded-influence remote reference (birrp) processing code by Chave and Thomson [2004].

In order to estimate the impedance tensor components we have to solve two bilinear regression problems. Each row of the impedance tensor relates two magnetic field components B_x and B_y to one electric field component through two unknown linear coefficients,

$$E_x = Z_{xx} H_x + Z_{xy} H_y, \text{ and} \tag{3.3}$$
$$E_y = Z_{yx} H_x + Z_{yy} H_y. \tag{3.4}$$

This seemingly straightforward problem has a lot of difficult and subtle problems to it, and the results depend on the chosen solution of these aspects. Obviously every physical measurement is affected by some sort of noise, so we have to estimate

errors for Z_{xx} and Z_{xy} and the plane described by them will not go exactly through all measured points. The first procedures to estimate the impedance were based on the idea of minimizing the distance of all points to the plane in the least-squares sense [Sims et al., 1971]. More modern algorithms based on *robust* methods will be described below. While least-squares estimation is mathematically uncomplicated, another problematic point remains. It is not possible to incorporate independent error estimates for the electric and magnetic field components, but only a single error estimate can be used. In most cases the magnetic field is assumed to be noise free and all disturbances concentrated in the electric field. This assumption is justified by the observation that in general electrical field recordings are much more noisy than magnetic fields.

Under these assumptions the impedance tensor estimates can be calculated from the mean cross and auto spectra of the electromagnetic fields. For Z_{xy} one possibility is

$$Z_{xy} = \frac{\overline{H_xH_x^\star}\,\overline{E_xH_y^\star} - \overline{H_xH_y^\star}\,\overline{E_xH_x^\star}}{\overline{H_xH_x^\star}\,\overline{H_yH_y^\star} - \overline{H_xH_y^\star}\,\overline{H_yH_x^\star}}, \tag{3.5}$$

others are given by Sims et al. [1971]. The expressions for the other elements are similar. This particular form is chosen because it is not biased by random noise on the electric channels, but only by random noise on the magnetic channels. As mentioned earlier, the magnetic channels are usually less affected by noise and the estimates from this formula are considered more reliable. When magnetic noise is present the estimator in equation 3.5 will result in a downward biased impedance. Other forms represent mixed bias or upward bias [Jones, 1980].

In practice a number of issues make this simple approach unsuitable for reliable estimation. A number of possible problems during the recording can strongly bias the output of such a simple processing scheme. These include

- Non-uniform natural fields that violate the plane-wave assumption, these occur mostly in high latitudes or during strong solar activity [Mareschal, 1986, Osipova et al., 1989, Jones and Spratt, 2002],

- Cultural noise from various electromagnetic sources such as pipelines, electric trains, power lines etc.,

- Temporary sensor problems or saturation of the AD-converter input, and

- Disturbances by humans or animals moving in the vicinity of the instrument.

The simplest way to remove the influence of problematic segments of the recording is manual editing of the time series. Segments that are considered unsuitable are excluded from further processing by visual inspection. However this approach is

time consuming and problematic segments can be impossible to identify. Furthermore preconceived ideas about the expected transfer function can influence the decision to include certain segments or not. A number of measures within the processing sequence itself aim at avoiding the arbitrariness of manual selection, while at the same time identifying problematic sections.

Coherence thresholding/ weighting: Under ideal circumstances the coherence between the magnetic and electric fields should be ≈ 1. A low coherence indicates problems in one of the channels and the corresponding segment is either completely excluded or its influence is reduced in the final calculation of the mean [Jones and Jödicke, 1984].

Remote reference: When several instruments are recording at the same time, the recorded magnetic field at another site can be used to improve the transfer function estimates. Under the plane wave assumption the magnetic field should be equal at any recording site in a purely layered Earth and vary coherently in the more general case. In practice the horizontal wavelength of the source is much larger than the induction scale length. For long-period data stations up to several hundred kilometres away are usually suitable as a remote reference site. If the noise in the magnetic field at the remote site is orthogonal to the noise at the local site, unbiased estimates can be obtained [Gamble et al., 1979a,b], in contrast to the downward biased estimates for single site processing.

Robust estimates: The mean as a statistical measure can be strongly influenced by a small number of events that fall far from the mean of all other estimates and thus violate the assumption of a Gaussian distribution of the residuals. These events, also known as *outliers*, then bias the final result. By down-weighting or excluding such outliers the estimation procedure can be made more *robust*, i.e. less influenced by spurious estimates [e.g. Egbert and Booker, 1986, Chave and Thomson, 1989].

Leverage points: Simple least-squares estimation methods can be strongly influenced by a single data-point with an amplitude that is much larger than the average. These *leverage points* can pose problems even for robust methods if they do not fall sufficiently far from the mean, but still bias the result. New *bounded-influence* methods try to identify and reduce the influence of such leverage points [Chave and Thomson, 2003].

Some or all of these methods are included in modern MT processing codes and they can greatly improve transfer function estimates compared to the simple processing scheme presented earlier. Still, some situations exist where even these
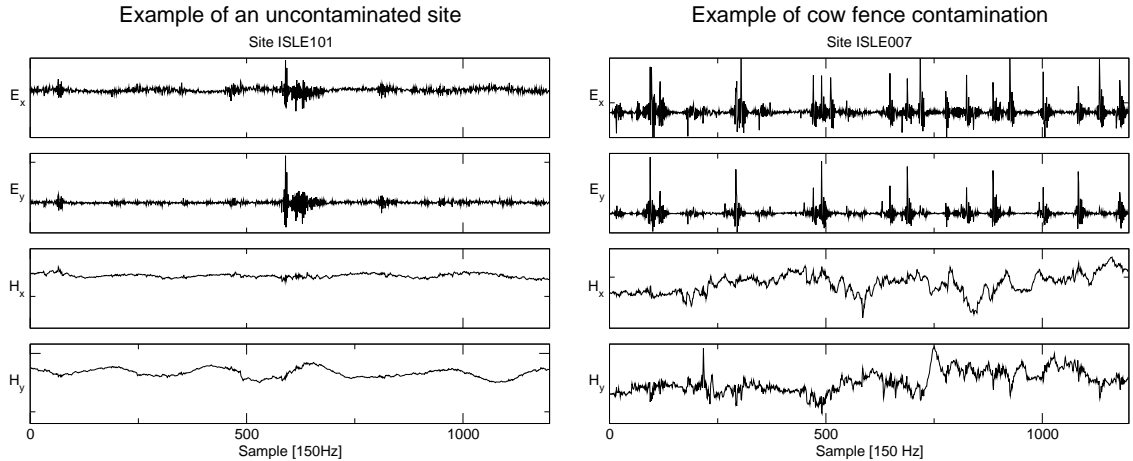
Figure 3.1: Recording of electric and magnetic channels from a undisturbed site (ISLE101) and a contaminated site (ISLE007). The high amplitude spikes in the electric field at site ISLE007 are caused by electric cow fences.

sophisticated methods fail to yield estimates of transfer functions representative of the conductivity distribution in the Earth. Coherence between the electric and magnetic fields is not always a reliable measure of data quality. Particularly cultural noise sources, such as DC railways or cow-fences, can produce signals that show high coherency, but violate the assumptions made in MT processing [Szarka, 1988, Qian and Pedersen, 1991, Padua et al., 2002]. In some cases these signals are so strong that they affect neighbouring sites making them unsuitable as remote reference sites. Also, when cultural noise is not just a temporary phenomenon but persists throughout the entire recording, robust methods break down. These methods rely on the fact that the majority of estimates is not strongly biased. If the majority of the estimates should be termed outliers, robust methods will use these as the "data" and discard the rarer reliable estimates.

3.1.2 The limit of current processing algorithms

The problems of available processing codes are dramatically illustrated by some of the high frequency data collected during the first phase of the ISLE-MT project. At most sites the high-frequency channels are strongly affected by noise from cow fences on the farms surrounding the recording site. The electromagnetic signal generated by the cow fences poses a particularly difficult problem for any processing algorithm. First, the noise usually continues throughout the whole time-series, making it difficult to distinguish "good" from "bad" segments. The noise is quasi-periodic, but both the shape of the signal as well as the frequency
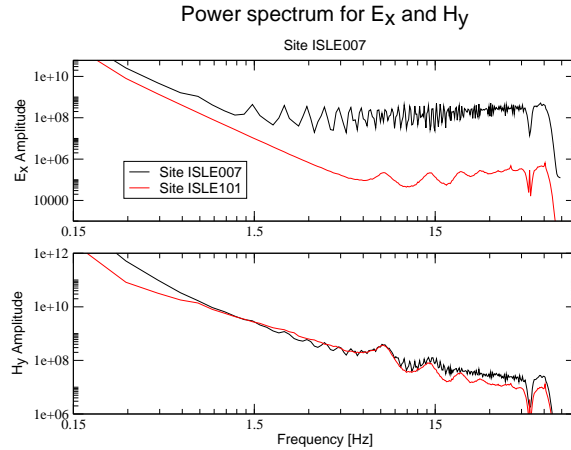
Figure 3.2: Spectra of E_x (top) and H_y (bottom) channels from sites ISLE007 (black) and ISLE101 (red). Site ISLE007 is severely contaminated by cow fence noise whereas ISLE101 is the least disturbed site in the survey area. The disturbed sites does not only show a high amplitude base frequency at 1.5 Hz and its harmonics, but is generally higher in amplitude.

changes with time. In some cases the power of the noise is also much higher than the power of the signal.

Figure 3.1 shows a comparison of a reasonably good site with such a problematic site. Both electrical components in the right hand panel show strong spiky signals. While there seems to be a certain repeatability, it can be clearly seen that the signal is far from being periodic. The magnetic channels seem to be unaffected by the cow fence noise, but close inspection shows that they also contain, to a much lesser degree, traces of the cow fence signal. In comparison the clean signal in the left hand panel does not contain such high amplitude spikes in the electric channels. In addition the magnetic signal seems much smoother, with a clear sinusoidal component that is typical for natural signals.

The impression from the time-series is emphasized by the corresponding spectra in Figure 3.2. The spectrum of the north-south electric channel at site ISLE007 is more than three orders of magnitude higher than at site ISLE101. It is theoretically possible that this difference is due to a much higher subsurface resistivity. Given that this would require a factor of 10^6 higher resistivities, makes this possibility highly unlikely though. The strong amplitude oscillations at 1.5 Hz and its harmonics reveal, that this effect is caused by the spikes observed in the time-series. Also the natural *Schumann resonances* [Schumann, 1954] at 7.8 Hz and 13.9 Hz can be clearly seen in the electric spectrum of site ISLE101, but are completely absent at site ISLE007.

21
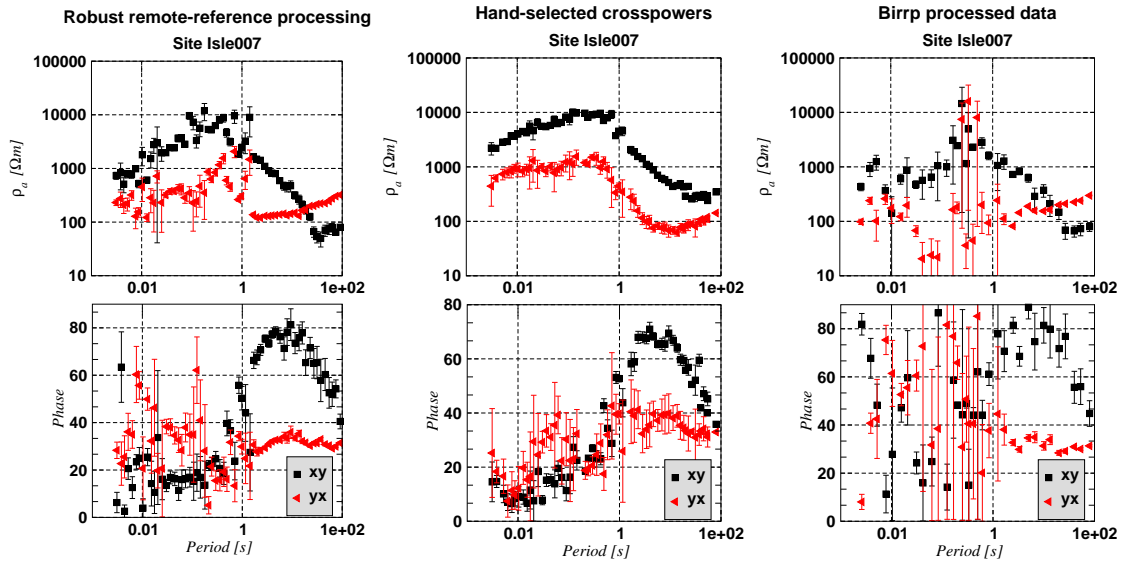
Figure 3.3: Transfer function estimates for site ISLE007 calculated with several popular processing codes. From left to right: Robust remote reference by Jones and Jödicke [1984], same code but with manually selected spectra, bounded-influence code by Chave and Thomson [2004].

The magnetic spectra are relatively similar at both sites. Close comparison shows though that there is an oscillatory component similar to the electric spectrum in the magnetic spectrum of site ISLE007. This indicates that the magnetic channels are affected by the cow-fence spikes as well.

Figure 3.3 compares the estimated off-diagonal impedance tensor elements for the contaminated site ISLE007 using three different popular approaches. The two left hand panels were both obtained using the same robust remote reference code [Jones and Jödicke, 1984, Jones et al., 1989, method 6]. The difference between the two plots is that for the middle panel individual impedance estimates were plotted as a function of time and estimates that were considered better were selected manually. The figure on the right was produced using birrp which includes protection against leverage points [Chave and Thomson, 2004]. The first thing to note is that all three methods produce different results at high frequencies. Although the general shape of the curves, particularly apparent resistivity, is similar, the absolute values at high frequencies differ up to an order of magnitude and much more than the error estimates suggest. The high apparent resistivities that sharply rise and then fall off abruptly, together with low phases in the left hand panel are indicative of a grounded dipole in the vicinity of the station [Qian and Pedersen, 1991]. There is also large scatter of values for adjacent frequencies. For an inductive method like MT, both apparent resistivity and phase should vary smoothly

with frequency. Also the theoretical relationships between apparent resistivity and phase can be used to check the data for internal consistency [Sutarno and Vozoff, 1991, Parker and Booker, 1996].

The smoothness of the apparent resistivity values as an indicator of data quality was used as a guideline in the manual selection process that produced the middle panel in Figure 3.3. Individual estimates have been selected to produce a smooth apparent resistivity curve. While the apparent resistivities consequently appear to be of better quality, the scatter between adjacent phase values reveals that this part of the sounding curve is still problematic. Finally for the birrp processed data, the apparent resistivities are much lower than for both other processing methods and appear to be generally more smooth than in the left hand panel. The phase estimates on the other hand scatter wildly between 0 and 90° and have large errors associated with them. While this visually appears to be the worst sounding curve, the large error bars at high frequencies at least signal that this part is problematic. To the untrained eye, a sounding curve like in the middle panel might seem of sufficient quality, but in essence the high frequency part of any of them cannot be used for interpretation or modelling. Only at periods >50 s, where the curves of all different processing schemes agree, the data appear to be reliable. These limitations of all processing codes available to date motivate the attempt to pre-process the data in the time domain before feeding it into the usual processing codes. We will present a number of signal processing methods and their application to the data in the following sections.

3.1.3 The ISLE-MT recordings

Before we can discuss signal processing methods and their application to MT data we have to look at some of the details of how the data were recorded during the ISLE-MT project. At each site we installed two different instruments that shared the electrodes to measure the electric fields, but used different magnetic sensors. LIMS instruments, borrowed from the Geological Survey of Canada, recorded with a sampling rate of 5 s and used a three-component ring-core fluxgate magnetometer to record the time varying magnetic field [Narod and Bennest, 1990]. These instruments recorded for 2 weeks to 2 months, depending on location, to obtain reliable estimates of the long period transfer functions. Due to their low sampling rate these recordings seem to be unaffected by the cow fences and data quality is generally high.

To obtain information about the shallow structure, Phoenix MTU-5A broadband instruments were also installed at each site for 2-3 days. They use magnetic coils to record the time derivative of the magnetic field and compute the high frequency transfer functions. Nearly all broad-band recordings are affected by cow

fence noise to some degree, and only at two sites could we obtain high-frequency impedance-tensor estimates that we regard as reliable. In contrast to the LIMS instruments, the broad-band instruments do not record with a single sampling rate, but in three different bands with different recording lengths for each band. The parameters for each band are shown in Table 3.1

Band	Sample rate [Hz]	Segment length [s]	Gap [s]	Total length [hrs]
3	2400	2	158	10
4	150	16	144	10
5	15	continuous		48-72

Table 3.1: Recording parameters for the different bands of the Phoenix MTU-5A broad-band MT recording unit.

These bands are then processed separately, and the results are combined into a single sounding curve. For the standard Phoenix processing software SSMT2000 the relationship between recording bands and frequency range of the impedance tensor are shown in Table 3.2.

Band	Frequency range [Hz]
3	352–35
4	35–5
5	5–0.009

Table 3.2: Standard processing frequencies for each band of the Phoenix MTU-5A broad-band MT recording unit.

In the following section we will mainly apply the methods to the time-series recorded in band 4, as the impedance estimates from this band seem to be most problematic. If we can successfully use the method in this band we will continue to apply it to the other band to test whether we can produce a reliable sounding curve for the whole frequency range.

3.2 Filter Theory

The term *filter* by itself is not well-defined. It is used in signal processing to denote a process that modifies some input signal and yields an output signal. In most cases, filters are associated with mathematical convolution as most filters

$$x(t) \longrightarrow \boxed{\quad f(t) \quad} \longrightarrow y(t)$$
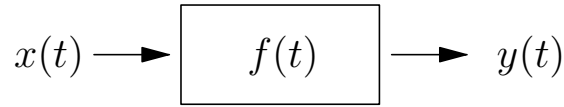
Figure 3.4: Block diagram of a general filter.

used in practical applications can be described as a convolution between the filter *impulse response* f(t) and the input x(t),

$$y(t) = \int_{-\infty}^{\infty} f(\tau)x(t-\tau)d\tau = f(t) * x(t), \tag{3.6}$$

or, for discrete and finite length data,

$$y(t_i) = \sum_{j=0}^{N} f(t_j)x(t_{i-j}). \tag{3.7}$$

It is often convenient and more intuitive to describe the filter in the frequency domain by its *transfer function*. This is because convolution in the time domain is replaced by element-wise multiplication in the frequency domain,

$$x(t) * y(t) \xleftrightarrow{\mathcal{F}} x(\omega) \cdot y(\omega). \tag{3.8}$$

The impedance tensor can be thought of as a filter that acts on the magnetic field to produce the electric field and that is why the term magnetotelluric transfer function is often used equivalently.

In signal processing applications a filter is often used to give the input signal certain properties, for example anti-alias filters that any digital recording instrument has to apply before digitizing to avoid artifacts from frequencies higher than the Nyquist frequency, given by

$$f_N = \frac{1}{2\Delta t}, \tag{3.9}$$

where Δt is the sampling rate. Anti-alias filters are analogue low-pass filters, but a similar step can occur digitally to down-sample the AD-converter output to the desired sampling rate. Filters can also be used to shape the input signal in the sense that the filter output should resemble some desired reference signal. We will discuss this later when we describe adaptive filters.

From Equation 3.8 it is clear that the application of any filter to a channel of an MT recording will change the spectrum of that channel. We have to distinguish two different cases of spectral distortion: 1) Removing the cow fence spike will obviously alter the spectrum and this change is desired, 2) In addition the filter

Method	Channels	Time dependence	Linearity
Delay	Single	static	
Template Substraction	Single	adaptive	
LMS	Multi	adaptive	linear
RLS	Multi	adaptive	linear
Neural Network	Multi	adaptive	non-linear

Table 3.3: Filters discussed in the following sections and their properties

results in some overall distortion of the spectrum that has nothing to do with the cow fence spikes but is an inherent property of the filter. Returning to equation 3.5 we see that the impedance tensor elements are calculated from ratios of auto- and cross-spectra. The overall spectral distortion therefore cancels out as long as it is identical for all channels of the time-series. For static filters, like the delay-filter that we will discuss below, this is easy to achieve. In contrast, for the more powerfull adaptive filters the filter transfer function depends on the data and this effect can become problematic.

The filters we will use to remove the cow fence noise can be classified in a number of ways. The simplest filters work on a single recorded channel at a time without using any information from other, simultaneously recorded, channels. In contrast multi-channel filters utilize expected relationships between different recording channels to distinguish noise from signal. The filters can also be static, i.e., their filter function remains constant for the whole time series, or adaptive, i.e., adjust to changing signal characteristics. Finally multi-channel filters can be categorized as linear or non-linear. Linear, multi-channel filters assume that the filter output is a linear combination of the input; we will discuss linear adaptive filters in particular. Non-linear filters, neural networks for example, can model more complicated relationships. Table 3.3 gives an overview over the filters we will discuss and their properties

3.3 Single channel methods

The simplest example of single-channel static filters are high-pass, low-pass or band-pass filters. High-pass and low-pass filters are unsuitable for noise cancellation in MT time series. They simply reduce the amplitude of all spectral estimates above or below a specified frequency, which is equivalent to discarding the impedance tensor estimates at those frequencies. This is sometimes necessary if all attempts to cancel the noise are unsuccessful but can be more easily and

more efficiently achieved by selecting estimates that are considered reliable after processing.

On the other hand, band-pass filtering can be useful under certain circumstances. If the noise is concentrated in a reasonably narrow frequency band, a band-pass filter can down-weight the spectra in that band and avoid leakage of the spectral estimates into neighbouring frequencies. Due to the inductive nature of magnetotelluric soundings, adjacent cross- and auto-spectra are not independent. Most processing codes use this to increase the number of estimates for a given output frequency by averaging over a window of frequencies around the central output frequency. If any of these frequencies are affected by noise, the final output will be so as well. Band-pass filtering can circumvent this problem, and is routinely undertaken to remove the influence of the 50 Hz electromagnetic signal associated with household electricity or the 16.3 Hz signal of AC powered railways. This is possible because, in most countries, the frequency is stable over long time and distance and the shape of the signal is sinusoidal, resulting in contamination at a very narrow range of frequencies. In contrast, the aforementioned cow-fence spikes contaminate the MT recording over a much broader range of frequencies and also at higher harmonics of the base frequency, as we already observed in Figure 3.2.

3.3.1 Delay Filter

One type of filter that cancels signals at a base frequency and its harmonics is the *Delay-Filter* [Schnegg and Fischer, 1980, Junge, 1996]. Its implementation is very simple: A copy of the time-series is time shifted and subtracted from the original. To preserve power we multiply the result with a normalizing factor,

$$y(t) = \frac{1}{\sqrt{2}} \left(x(t) - x(t - t_0) \right). \tag{3.10}$$

In the frequency domain the result is

$$\mathcal{F}\{y(t)\} = \frac{1}{\sqrt{2}} \mathcal{F}\{x(t) - x(t + t_0)\}, \tag{3.11}$$

$$= \frac{1}{\sqrt{2}} \left(\mathcal{F}\{x(t)\} - \exp\left(\iota\omega t_0\right) \mathcal{F}\{x(t)\} \right), \tag{3.12}$$

$$= \frac{1}{\sqrt{2}} \left(1 - \exp(\iota\omega t_0) \right) \mathcal{F}\{x(t)\}. \tag{3.13}$$

Hence the transfer function of the filter is

$$f(\omega) = \frac{1}{\sqrt{2}} \left(1 - \exp(\iota\omega t_0) \right), \tag{3.14}$$
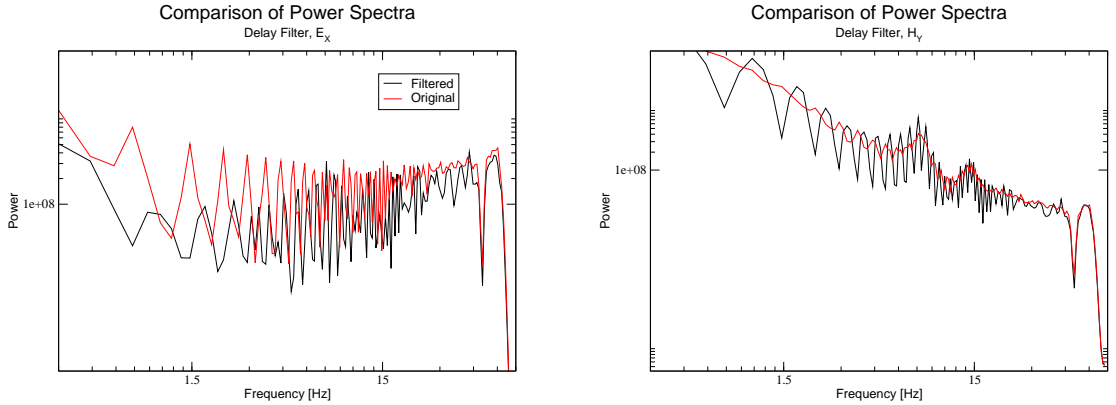
(a) Comparison of the E_x spectrum before (red) and after (black) delay filtering. The base frequency and the first couple of multiples have been successfully suppressed. Many of the higher harmonic frequency remain in the spectrum though.

(b) Comparison of the H_y spectrum before (red) and after (black) delay filtering. The filtered spectrum reveals the cosine shape of the filter transfer function.

Figure 3.5: The effect of a delay filter on the spectrum of E_x (left) and H_y (right).

and the resulting power spectrum is

$$\mathcal{P}\{y(t)\} = \mathcal{F}\{y(t)\}\mathcal{F}^*\{y(t)\} = (1 - \cos \omega t_0)\,\mathcal{P}\{x(t)\}. \qquad (3.15)$$

This means that the power at the base frequency $1/t_0$, and all its harmonics, will be suppressed by the vanishing filter transfer function. Note that the spectrum is also altered at all other frequencies, apart from $n/(2t_0)$ with n odd. As mentioned before it is not a problem for processing if we apply the same filter to all components of the time series as the resulting distortion of the spectra is identical for all components. The effect of such a delay filter on the spectrum of an MT recording is shown in figure 3.5. The time shift was determined by inspecting the time-series and measuring the prevalent distance between spikes. The corresponding shift of $n = 207$ resulted in better suppression of the cow-fence noise than a shift of $n = 216$ that was suggested by the lowest maximum in the electric spectrum. For our experiment we considered only integer valued shifts. In theory it would be possible to shift the time-series by a fraction of a sample by interpolation and resampling. However changing the shift by ± 1 does not vary the output of the filter significantly, so that this additional complexity seems unnecessary.

Figure 3.5 shows how the first maximum in the power-spectrum at 0.8 Hz and its harmonics are suppressed by the filter, but it also illustrates the problematic nature of cow fence noise. While the first few frequencies reduced by the filter are maxima in the original spectra, the high frequency filter minima coincide with
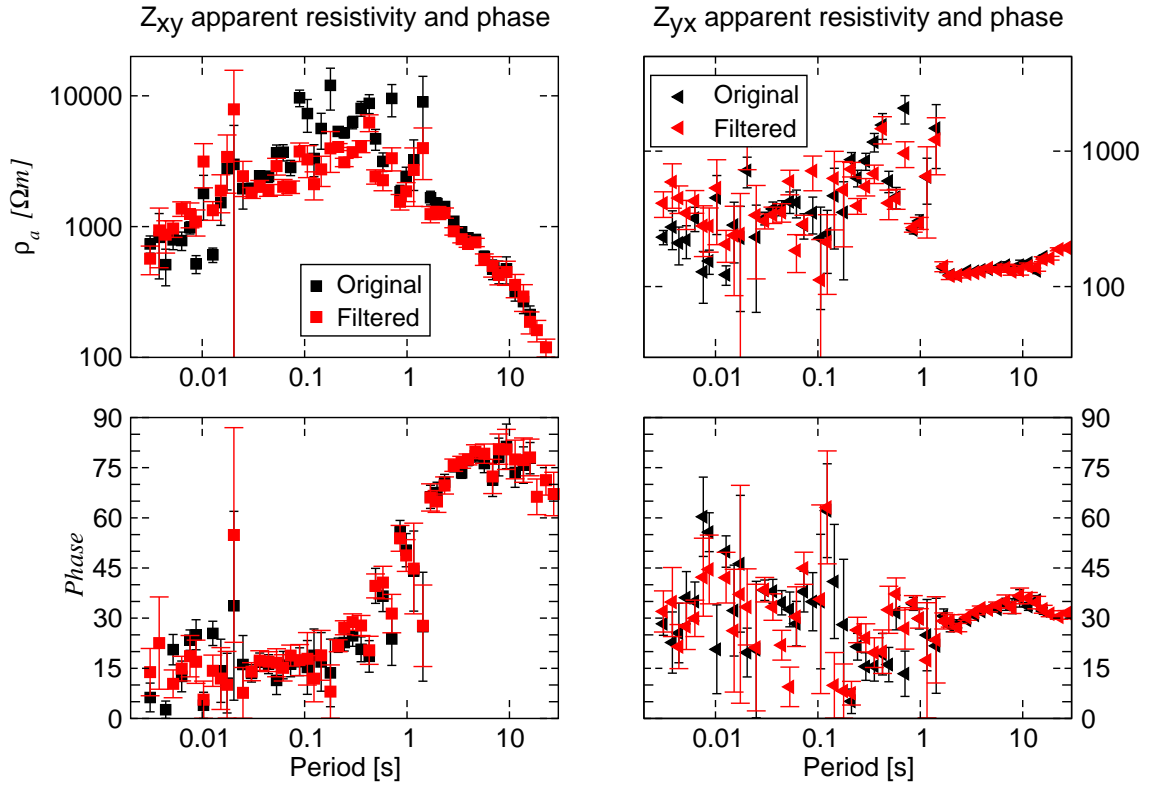
Figure 3.6: Off diagonal elements of the impedance tensor for site ISLE007. Black symbols show original estimates, red-symbols show estimates after applying a delay filter.

minima in the original spectrum, and some strong spikes in the spectrum are left undisturbed or are even amplified. Likely explanations are that the signal is not exactly repetitive and that the higher frequency signature is not an exact integer multiple of the base frequency. This can be due to the co-action of several fences with slightly different frequencies, or simply a variations in the mechanical mechanism that produces the spikes. Several attempts with different shifts t_0 for the filter, and a cascaded application with different shifts at each step did not result in any significant improvement.

Figure 3.6 shows, that the delay filter does not significantly improve the impedance estimates. Around the base frequency of 1 Hz the scatter in the xy-component apparent resistivities is slightly reduced, but for the most part the estimates remain the same as before. For the yx-component the estimates have changed as well, but they scatter in a similar manner as before, and it is not possible to decide which of the results could be termed better. The phases of both components have not changed significantly, apart from an outlier in the xy-component. This suggests that a delay filter is unsuitable to remove the cow-fence
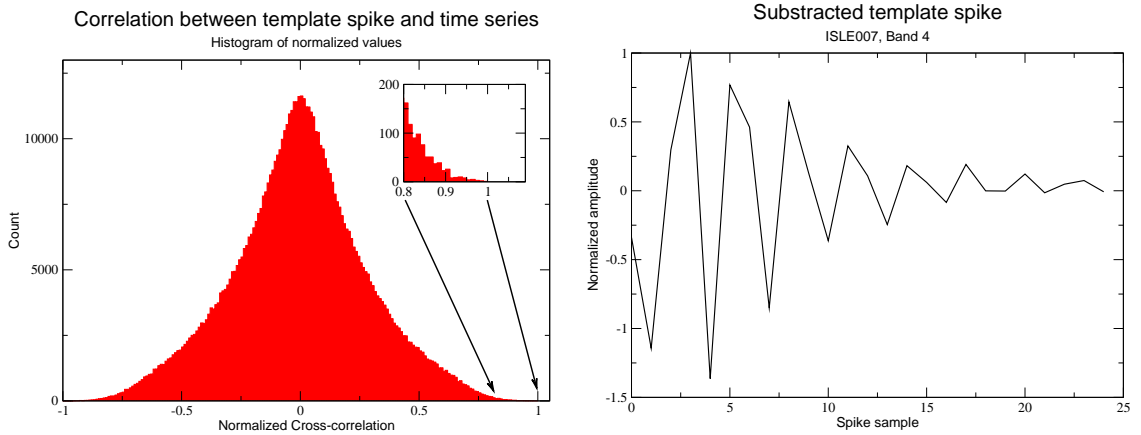
Figure 3.7: Histogram of normalized cross-correlation between a template spike and the time series (left) and the corresponding spike form (right) for a spike that is not representative of the cow fence spikes. The inset graph in the left hand figure shows a magnification of the $0.8\ldots1$ range.

noise from the data.

3.3.2 Direct Template Substraction

Apart from simple filtering operations with well controlled behaviour in the time and frequency domain, we will also try a rather ad-hoc method to substract the undesired signal from the time series. The first method that might come into mind is to identify a "typical spike signature" by visual inspection or some automated method, identify sections of the time-series where this template spike matches and substract it. If the shape of the spike remains approximately similar throughout the time series, perhaps with some amplitude variations, which are easy to correct for, this appears to be a reasonable approach. We will show the successes and problems of this approach in the case of cow fence noise in this section.

To identify segments of the time-series x_i that match our template spike s_i we use the normalized cross-correlation c_i given by

$$c_i = \frac{\sum_j s_j x_{j+i}}{\sqrt{\sum_j x_j^2 \sum_j s_j^2}}, \qquad (3.16)$$

between the two.

The possible values of c_i are in the range ± 1. An absolute value close to unity indicates a highly similar segment of the time-series, while values around 0 indicate dissimilar shapes in the time-series and the template spike. The histogram of cross-correlation values shown in Figure 3.7 illustrates the problem of this approach
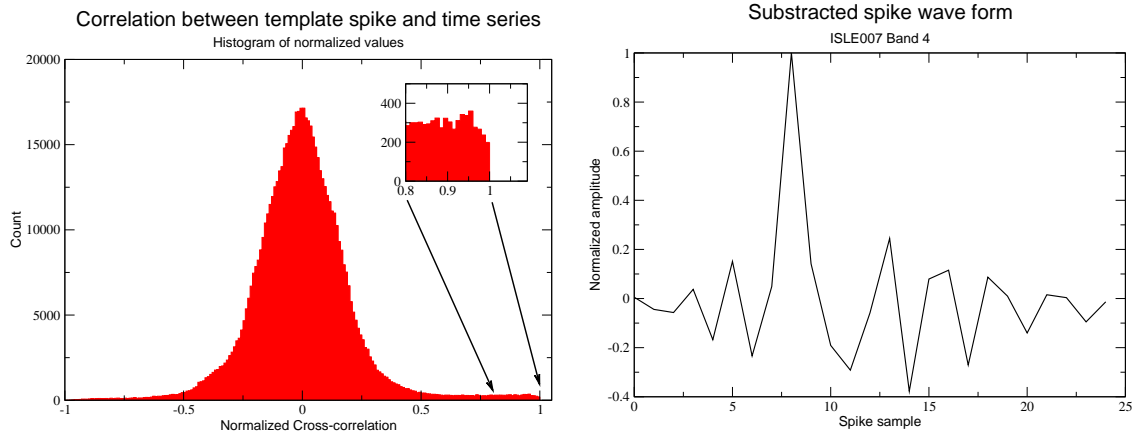
Figure 3.8: Histogram of normalized cross-correlation between a template spike and the time series (left) and the corresponding spike form (right) for a spike that represents some of the cow-fence disturbances in the time series. The inset graph in the left hand figure shows a magnification of the $0.8\ldots1$ range.

for a manually selected spike. As expected, the peak of the distribution is at 0 correlation, most of the time-series does not resemble the template spike at all. But if the template spike was as representative of the numerous spikes in the recording as we hope, there should be secondary peaks at ±1 or at least a significant number of points with a correlation > 0.9. What we observe instead is an extremely low count of high correlation points; the distribution drops rapidly off to zero at high correlation values. In the whole time series there are about 20 segments that have a correlation > 0.9 with the template spike. There are a number of possible reasons for the low agreement: The chosen spike might not be representative of the noise, but in fact just be a single disturbance, or the chosen segment of points around the spike might be too long and contain natural source data. If this is the case, the correlation will be lower than for an adequately chosen time window.

To avoid tedious trial and error searches for the right template spike, we use a simple automatic routine to identify possible cow fence spikes. The visual inspection of the time series shows that, at least in the electric fields, the spikes appear nearly as delta functions with a high amplitude. At the onset of the spike the first difference, the difference to the previous value, reaches values that are up to several magnitudes higher than for the rest of the data. We define a threshold value `diffthreshold` and if the first difference exceeds this value the segment is considered part of a spike.

Figure 3.8 shows the chosen spike and the corresponding histogram for a more successfull case that was identified automatically. We can now see from the his-

Parameter name	explanation
diffthreshold	Minimum first difference above which
	a point is considered a spike
trailpoints	Number of points before the identified spike
	included in the template
decaypoints	Number of points after the identified spike
	included in the template
minspikeavg	Minimum number of spikes needed for average so that
	the average is used for substraction
corrthreshold	Minimum correlation between template and other spikes
	for averaging and substraction
iterations	Number of iterations of the algorithm,
	at each iteration a new spike is selected

Table 3.4: Description of user-defined parameters for the direct template substraction algorithm

togram that large positive correlation values appear much more often than in the case displayed in Figure 3.7. Particularly the comparison with negative correlation values, where the distribution still falls off to zero quickly, confirms that there is a large number of similar waveforms within the time-series. On the basis of this observation we construct a simple algorithm to remove spikes from a time series shown in Algorithm 1. There are a number of user-defined parameters that determine which kind of shapes are considered spikes and how they are removed; these are summarized in Table 3.4. The same algorithm is applied with identical parameters to both components of the electric field for simplicity, although it would also be possible to use different parameters for each component. Applying the algorithm to the magnetic field components introduces strong artifacts into the time-series. As we can see from Figure 3.1 the characteristics of the cow-fence noise are very different for the magnetic channels. Particularly there are now high amplitude spikes that can be used to identify the disturbances and the signature is generally much weaker. We therefore leave these channels unchanged to avoid the introduction of artifacts.

As mentioned above, this algorithm is purely based on a phenomenological description of what is thought to be a cow fence spike, and the algorithm is the bare minimum necessary to identify spikes and similar shapes in the time series. Especially the minimum correlation threshold corrthreshold has to be set high enough to avoid averaging and subtracting spike shapes that have little similarity. This is also one of the differences to the similar iterative deconvolution algorithm

---

Algorithm 1 Algorithm to remove spikes from one component

1: Read parameters
2: **for** *iterations* **do**
3:   **repeat**
4:     $++i$
5:   **until** $x(i) - x(i-1) > \text{diffthreshold}$
6:   Copy $x(i - \text{trailpoints}) \ldots x(i + \text{decaypoints})$ into s
7:   Calculate cross-correlation c of s and x
8:   **for all** i **do**
9:     **if** $c(i) > \text{corrthreshold}$ **then**
10:       Stack $x(i - \text{trailpoints}) \ldots x(i + \text{decaypoints})$ with s
11:     **end if**
12:   **end for**
13:   **if** Stack consists of more than *minspikeavg* spikes **then**
14:     Normalize c, that maximum value is 1
15:     Calculate cross-correlation c of stacked s and x
16:     **for all** i **do**
17:       **if** $c(i) > \text{corrthreshold}$ **then**
18:         Scale s to maximum amplitude of current window
19:         Substract s from $x(i - \text{trailpoints}) \ldots x(i + \text{decaypoints})$
20:       **end if**
21:     **end for**
22:   **end if**
23: **end for**

[Ligorria and Ammon, 1999] that we will introduce in Chapter 4 to calculate receiver functions in the presence of noise. The other main differences are that we do not use the full time-series of another channel, but an averaged extract from the same channel, and that we are not interested in the transfer function, but in the modified time-series.

The performance of the algorithm on typical data from Ireland is shown in Figure 3.9. The figure shows that a number of spikes is significantly reduced, while some others remain unchanged. The reason for this is the high variability in spike shape. The correlation threshold in this case was set to 0.9; a lower threshold will remove more spikes from the time-series, but will introduce artifacts because of mismatches between the spike shape and the time-series.

The introduction of artificial features into the time-series is a risk that is common to all signal processing techniques, and is extremely difficult to assess because the noise-free time-series is unknown. We will examine the impact of the minimum
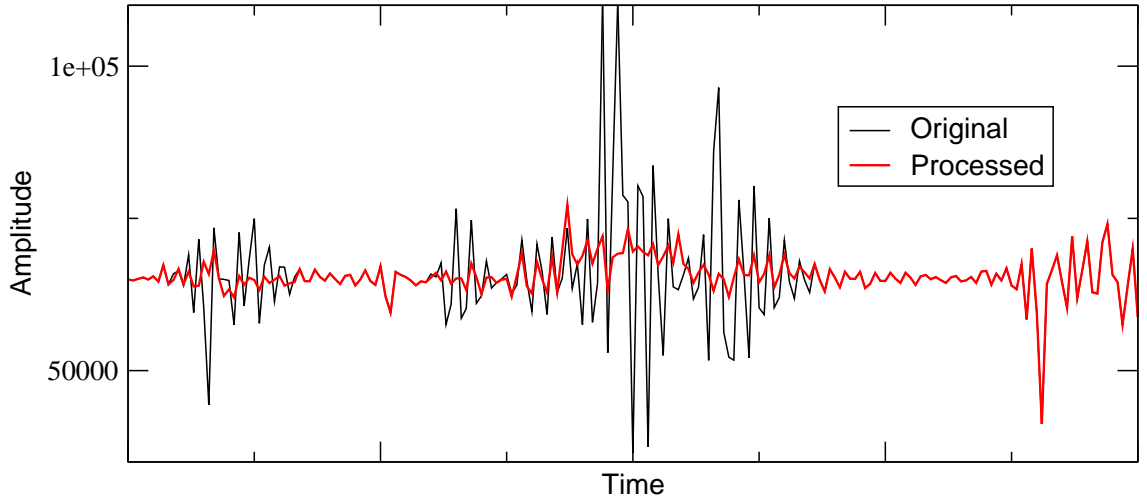
Figure 3.9: Comparison of original time series with time series after application of the template substraction algorithm. The correlation threshold was set to 0.9 in this case.

correlation parameter $\mathtt{corrthreshold}$ by systematically varying it and comparing the resulting time-series and spectra. The parameters for the various runs can be found in Table 3.5 and the resulting time-series in Figure 3.10.

	1	2	3
Spike height	10,000	10,000	10,000
Trail points	5	5	5
Decay points	20	20	20
Minimum correlation	0.95	0.9	0.8
Minimum samples	10	10	10
Iterations	50	50	50

Table 3.5: Parameters used to examine the performance of the template substraction algorithm

It is immediately obvious that the correlation threshold of 0.8 in the lowermost plot of Figure 3.10 is too low. A number of high amplitude spikes and downward shifts have been added to the time-series. Some other spikes have been reduced in amplitude and the remaining signal resembles natural variations, but the strong artifacts make it unusable for any further processing. Both other time-series, with correlation thresholds > 0.9, do not show any obvious artifacts. As observed before, some spikes remain unchanged while others seem to be removed. Carefull
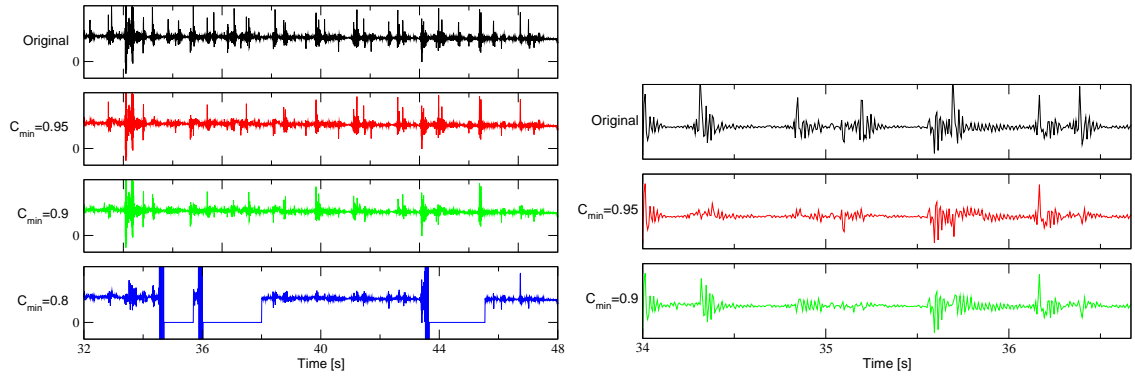
Figure 3.10: Comparison of original time series with time series after application of the template substraction algorithm with different correlation thresholds. The right graph is a magnified version of the upper three panels in the left graph to show the differences after spike removal.

comparison of the two time-series, with correlation thresholds of 0.9 and 0.95 respectively, reveals some subtle differences where spikes have been removed. The right panel of Figure 3.10 shows these differences for a few spikes. From the time-series plot alone it is impossible to decide which one is more representative of the natural signal, but it appears that with a higher correlation threshold, the artifacts are less severe.

The corresponding power-spectra provide an alternative view on the action of the noise removal algorithm. Figure 3.11 shows the power-spectra that correspond to the time-series in Figure 3.10. It is interesting to see that the maximum at the base-frequency of 0.8 Hz remains virtually unchanged by all different runs, only the high-frequency harmonics are reduced. The power spectra confirm the conclusion that a correlation threshold of 0.8 is too low and introduces strong artifacts. Throughout the whole frequency range the spectrum after spike removal is 4 orders of magnitude higher than the original spectrum, and is essentially flat. This indicates the presence of a number of high amplitude spikes whose frequency response is a straight line. Again there are some subtle differences between the other two spectra. In general the correlation threshold of 0.9 seems to remove more power from the time series, but this is, by no means, an indicator that the algorithm works better than with a higher correlation threshold.

Both the cleaned time-series and the corresponding spectra, for a correlation threshold of 0.9, indicate that the algorithm works and does not introduce major artifacts. We will therefore apply the algorithm to all frequency bands and process the resulting time series to compare the impedance estimates with the original. Due to the different sampling rates and the differing appearance of the cow-fence
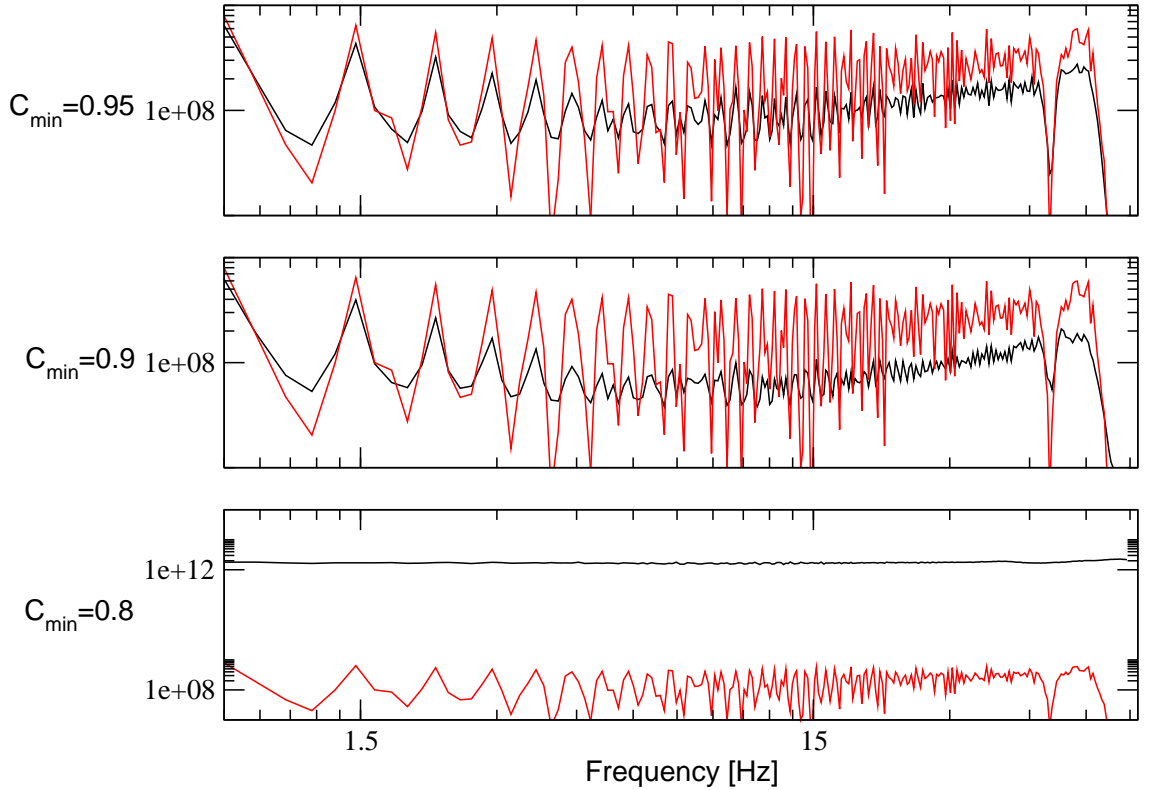
Figure 3.11: The spectra corresponding to the three processed time series in Figure 3.10. For comparison we plot the spectrum of the original time-series in red in each plot as well.

spikes we have to adjust some of the parameters for the other bands. These parameters are shown in Table 3.6

Finally we compare the impedance tensor estimates after applying the algorithm with the parameters given in Table 3.6 with the estimates from the unprocessed time-series. The off-diagonal apparent resistivity and phase are plotted in Figure 3.12. It appears that the apparent resistivities are generally smoother for both components, usually an indicator of better data quality, and in the critical band of 0.1-1 s also lower. This is the effect of removing the high energy spikes from the electric field recordings, which results in a smaller enumerator in equation 3.5 while the denominator remains constant. Still, there is considerable scatter in both the apparent resistivity and phase estimates. Especially the Z_{yx} estimates appear to be too noisy to be used for any kind of interpretation. The Z_{xy} estimates, on the other hand, appear to be of reasonable quality, but it is difficult to assess how much impact the cow-fence noise still has. At the end of this chapter we will

	Band 3	Band 4	Band 5
Spike height	100,000	10,000	5,000
Trail points	20	5	5
Decay points	200	20	5
Minimum correlation	0.9	0.9	0.9
Minimum samples	10	10	10
Iterations	50	50	50

Table 3.6: Parameters of the template substraction algorithm for the different recording bands

compare the results from all noise removal methods to get an idea of the quality of these estimates.

3.4 Linear adaptive filters

In contrast to static filters, *adaptive* filters can change their transfer function during the filtering process. Usually the adaption is performed in order to minimize a certain pre-defined criterion, for example the difference between the filter output and a reference signal. Adaptive filters are now an integral part of modern electronic devices, such as wireless network cards, mobile phones hearing aids and much more [Widrow et al., 1975, Widrow and Stearns, 1985]. A wide variety of adaptive filtering algorithms exists each of which has certain advantages and disadvantages depending on the requirements [Haykin, 2002]. We will only deal with fairly simple adaptive filtering algorithms which nevertheless have been proven to be most useful in practice. Also these filters are usually implemented digitally, so we will only be concerned with discretely sampled data.

3.4.1 The LMS adaptive filter

For a *linear adaptive combiner* the output of the filter y at time t_i is the sum of the input $\mathbf{x}$ weighted by the filter weights $\mathbf{w}$,

$$y(t_i) = \mathbf{w}^\mathsf{T}(t_i)\mathbf{x}(t_i) \text{ or shorter } y_i = \mathbf{w}_i^\mathsf{T}\mathbf{x}_i, \tag{3.17}$$

where $\mathbf{w}$ and $\mathbf{x}$ are both vector quantities. This is in contrast to non-linear filters, like neural networks, that we will discuss below, where the filter output is a non-linear function of this weighted sum. For $\mathbf{x}$ the vector property can be interpreted in different ways: Each component of the vector represents one sample at a given time from a different channel, each component represents a sample at different
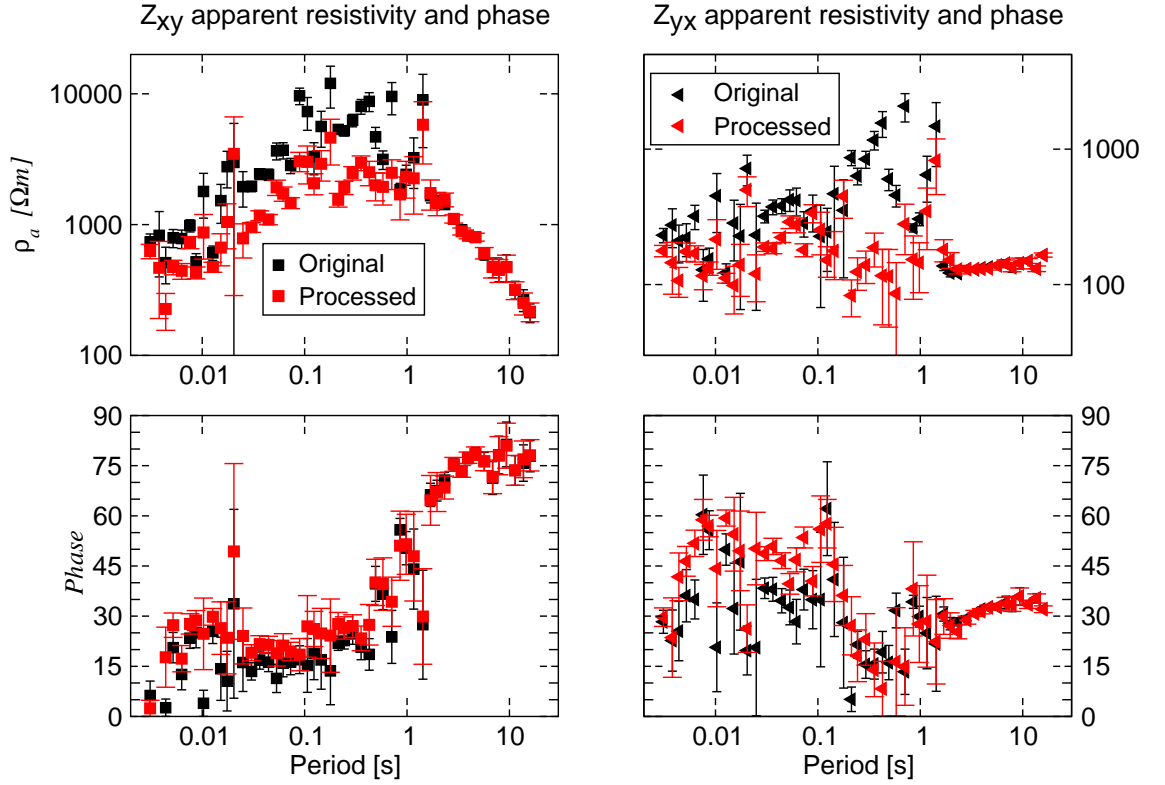
Figure 3.12: Off diagonal elements of the impedance tensor for site ISLE007. Black symbols show estimates without pre-processing, red-symbols show estimates after pre-processing with the template substraction algorithm.

times from the same channel, or a combination of both. A schematic overview of a general adaptive filter can be found in Figure 3.13. For the MT case we will combine a number of samples from two remote magnetic channels to predict one of the local electric or magnetic channels,

$$\mathbf{x_i} = \{H_x(t_i), \ldots, H_x(t_{i-L}), H_y(t_i), \ldots, H_y(t_{i-L})\}. \tag{3.18}$$

This approach was first proposed by Hattingh [1989] and proved to be the most successfull in the absence of direct information about the noise. Lezaeta et al. [2005] also used an LMS-adaptive filter to remove noise from sea-bottom magnetotelluric recordings. In their case though the noise source, the tilt of the instruments from wave motion, had been measured and could be used for removal.

In order to make the filter adaptive we have to define an error criterion $\epsilon(t_i)$ that we want to minimize. This is usually the difference between the filter output and a given reference signal $d(t_i)$,

$$\epsilon(t_i) = d(t_i) - y(t_i) = d_i - \mathbf{w_i^T x_i}, \tag{3.19}$$
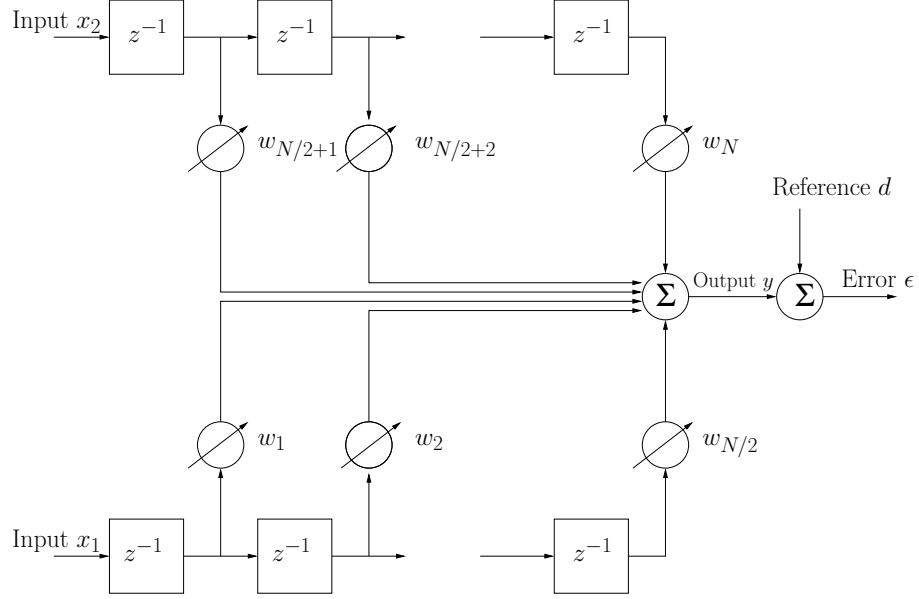
Figure 3.13: Schematic overview of a linear adaptive filter.

which in our case is a sample from the local channel that we want to match.

We want to minimize this error in the least-mean-square sense. If we assume, for now, that ϵ_i, d_i and $\mathbf{x}_i$ are all stationary, the expectation of the squared error over time is given by

$$E[\epsilon_i^2] = E[d_i^2] + \mathbf{w}_i^\mathsf{T} E[\mathbf{x}_i \mathbf{x}_i^\mathsf{T}] \mathbf{w} - 2E[d_i \mathbf{x}_i^\mathsf{T}] \mathbf{w}_i. \tag{3.20}$$

The square matrix $E[\mathbf{x}_i \mathbf{x}_i^\mathsf{T}]$ is the correlation matrix $\mathbf{R}$ of the input vector, and we can define the vector $\mathbf{P}$ as

$$\mathbf{P} = E[d_i \mathbf{x}_i^\mathsf{T}] = \{d_i \mathbf{x}_i, d_i \mathbf{x}_{i-1}, \ldots, d_i \mathbf{x}_{i-L}\}^\mathsf{T}, \tag{3.21}$$

i.e., the cross-correlation between the reference and the input signal. With these definitions we can rewrite the mean-square-error ξ in 3.20,

$$\xi = E[\epsilon_i^2] = E[d_i^2] + \mathbf{w}_i^\mathsf{T} \mathbf{R} \mathbf{w}_i - 2\mathbf{P} \mathbf{w}_i, \tag{3.22}$$

and we can see that for a stationary signal the mean-square-error is a quadratic function in $\mathbf{w}$ so it has a unique optimum solution $\mathbf{w}^\star$, the so called *Wiener solution*. We can find this solution by calculating the gradient of equation 3.22 with respect to $\mathbf{w}$,

$$\frac{\partial \xi}{\partial \mathbf{w}} = 2\mathbf{R}\mathbf{w} - 2\mathbf{P}. \tag{3.23}$$

At the optimum value $\mathbf{w}^\star$ the gradient is zero,

$$2\mathbf{R}\mathbf{w}^\star - 2\mathbf{P} = 0 \Rightarrow \mathbf{w}^\star = \mathbf{R}^{-1}\mathbf{P}. \tag{3.24}$$

This equation is the *Wiener-Hopf* equation. Notice that we assume here that the inverse of the auto-correlation matrix $\mathbf{R}^{-1}$ exists. This is not the case in general and even when it exists, the direct inverse can be numerically unstable.

All adaptive filters we will be concerned with try to minimize the mean-square-error; the difference lies in the way how the minimization is achieved. This is a direct analogy to the inversion algorithms we will discuss in Chapter 5 and, to some degree, the same algorithms are used. There are a number of important differences though, that influence the implementation details and choice of minimization method.

First, while in inversion we seek a model that explains the whole dataset, the analogue to the input vector, in adaptive filtering we operate only on a limited section of the data and the model (the weights of the filter) can vary with time. The filtering process thus consists of two stages: 1) Convergence from the starting weights to the optimum weights $\mathbf{w}^\star$, and 2) tracking of the optimum weights through time. Second, most filtering algorithms are intended for real time applications, which means that computational cost is a major concern and therefore filtering routines tend to use more simple minimization methods than commonly used in inversion.

The simplest possible adaptive filter is the *least mean square* or *LMS*-adaptive filter. It is similar to the method of steepest decent, but with some special adaptions for signal processing applications. For a steepest decent type iteration we follow the negative direction of gradient scaled by some stepsize μ,

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \mu(-\nabla\xi). \tag{3.25}$$

A straightforward implementation of this algorithm would use finite differences of short term averages of ϵ_i^2, an estimate for the expectation value, to obtain an estimate of the gradient. For the LMS algorithm we use ϵ_i^2 itself as an estimate of the expectation value ξ_i. Our estimated gradient is then

$$\nabla\xi \;\; = \;\; \left\{ \frac{\partial\epsilon_i^2}{\partial w_1}, \ldots, \frac{\partial\epsilon_i^2}{\partial w_L} \right\}, \tag{3.26}$$

$$= \;\; 2\epsilon_i \left\{ \frac{\partial\epsilon_i}{\partial w_1}, \ldots, \frac{\partial\epsilon_i}{\partial w_L} \right\}, \tag{3.27}$$

$$= \;\; -2\epsilon_i\mathbf{x}_i. \tag{3.28}$$

Equation 3.28 is equivalent to equation 3.23 as we can see from

$$2\mathbf{R}\mathbf{w} - 2\mathbf{P} = 2\left(\mathbf{x}\mathbf{x}^\mathsf{T}\mathbf{w} - \mathbf{d}\mathbf{x}\right) = 2\mathbf{x}\left(\mathbf{x}^\mathsf{T}\mathbf{w} - \mathbf{d}\right) = -2\mathbf{x}\epsilon. \tag{3.29}$$

Hence the filter weights are updated by

$$\mathbf{w}_{i+1} = \mathbf{w}_i + 2\mu\epsilon_i\mathbf{x}_i. \tag{3.30}$$
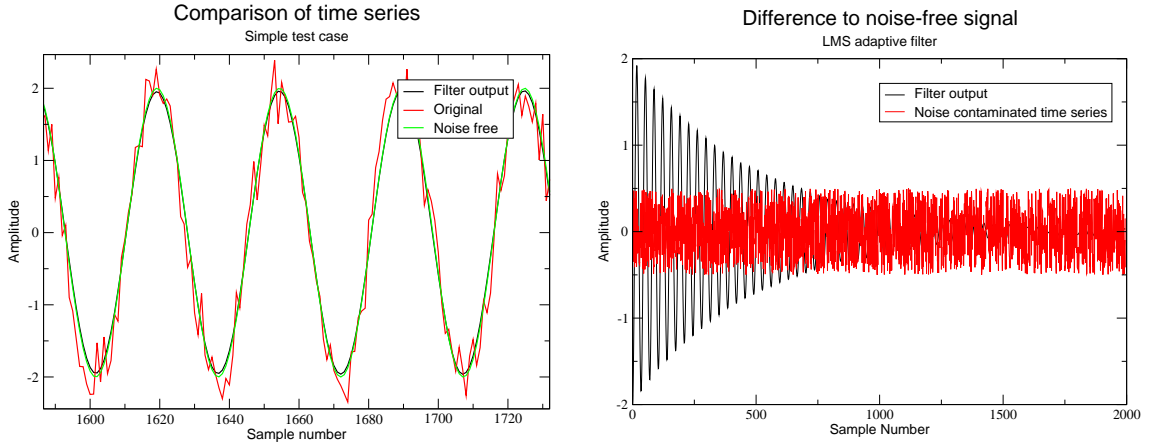
Figure 3.14: Simple performance test for the LMS adaptive filter. The left panel shows filter output (black), filter input (red) and the noise free signal (green). On the right the difference to the noise free signal is plotted for the filter output (black) and the original time-series (red). The filter length was N = 5 with a stepsize $\mu = 0.005$.

Usually all weights are initially set to zero. It would be possible to initialize them with values based on an a-priori estimate of the relationship between input and reference channels, but we will not consider these cases, but rely on the algorithm to find the appropriate weights. The stepsize μ has to be chosen by the user and is critical for the algorithm's performance. We will discuss the problem of choosing μ and a modified algorithm below.

We test the algorithm and its basic abilities with two simple harmonic signals for input and reference:

$$x_i = \sin(2\pi i/N), \quad d_i = 2\cos(2\pi i/N) + \mathrm{rand}(-0.5, 0.5). \qquad (3.31)$$

Here $\mathrm{rand}(-0.5, 0.5)$ denotes a random number in the interval $[-0.5, 0.5]$. The results can be seen in Figure 3.14. The left hand panel shows that the original noise contaminated reference time series (red) deviates from the noise free signal more than the filter output. This impression is confirmed by plotting the difference between the noise free signal and the reference time-series and filter output, respectively. After an initial adaptation stage, the difference falls below the original noise level and the filter output is a closer approximation to the noise free signal. This is, of course, because noise and signal are uncorrelated and a stable transfer function between input and noise free reference signal exists. As the adaptation process continues the filter weights approach the transfer function and we recover the noise free signal nearly perfectly.

Algorithm 2 The normalized LMS algorithm

1: $\mathbf{w}_0 = 0$
2: **for all** time samples i **do**
3: $\quad y_i = \mathbf{w}_i^\mathsf{T}\mathbf{x}_i$
4: $\quad \epsilon_i = d_i - y_i$
5: $\quad \mu_i = \frac{\tilde{\mu}}{\delta + \|x_i\|^2}$
6: $\quad w_{i+1} = w_i + 2\mu\epsilon_i x_i$
7: **end for**

When using the LMS adaptive filter, two choices have to be made that affect the performance of the filter. The number of filter weights, N, determines to which level of detail the filter can model the transfer function, or if thought of in the frequency domain, the spectral resolution of the filter. This parameter has to be chosen by trial and error and we will use long filters with $N > 100$ to deal with the spectral spikes in the contaminated electric field.

The second critical parameter is the choice of the stepsize, μ. If the step size is too small, the filter will not adapt fast enough and the filter output will not represent the desired signal. On the other hand, if the step size is too large the process will become unstable. In this case the corrections applied in each step of the algorithm are too large and subsequent attempts to correct this error result in even larger maladjustment. The optimum stepsize is difficult to chose a priori; it depends on the filter length, the input power and the error signal power. One solution to, at least, narrow the range of acceptable values for μ is to use the *Normalized LMS Algorithm* [Haykin, 2002]. The step size μ is replaced by the adaptation constant $\tilde{\mu}$ that relates to the step size through,

$$\mu = \frac{\tilde{\mu}}{\delta + \|x_i\|^2}.$$ (3.32)

The small constant δ avoids numerical problems with the divisor when $\|x_i\|^2 \ll 1$. In contrast to the step size, which has dimensions of inverse power, the adaptation constant is dimensionless, removing at least the need to estimate the signal power. Note also that now the step size changes during the adaptation process. During segments of high input power, a small step size is used to avoid divergence, whereas during segments of low input power the step size becomes larger to facilitate rapid convergence. The computational steps for the LMS adaptive filter are summarized in Algorithm 2

3.4.2 Applying a LMS filter to MT data

Before we can apply the LMS-adaptive filter to the MT time-series we have to per-form one step of additional preprocessing. As Figure 3.2 shows, the spectrum of the magnetic channel, in particular, has a considerable overall slope. This is prob-lematic because for a spectrum like this, the adaptation process will largely focus on reproducing this slope and not the variations we are interested in. The process of flattening the spectrum is known as *whitening*. In our case this can be achieved fairly easily by taking the first difference of the data. This an approximation for the first derivative and we can see from

$$\mathcal{F}\{x'(t)\} = \iota \omega \mathcal{F}\{x(t)\} \tag{3.33}$$

that this effectively acts as a high-pass filter subtracting a linear trend from the spectrum. At the same time it solves the problem that the individual recording segments in recording Bands 3 and 4 have a varying non-zero mean. When this pre-processing has been performed on all channels we can apply the adaptive filter to the data.

As mentioned above we filter a segment of the remote magnetic field to produce an improved version of one of the local channels, which we use as a reference. We shift the reference channel so that the current sample of the reference channel corresponds to a sample in the centre of the input segments. This allows the filter transfer function to model both positive and negative phase relationships between the input and reference channels. Finally, we will perform two filtering runs for each component. During the first run the filter can converge to the optimum value, a process that can take a considerable fraction of the recording length. After reaching the end of the recording we start again from the beginning with the filter weights from the previous run. We therefore avoid problems in the subsequent impedance estimation stage, where we would have to discard the segment of the time-series where the algorithm has not converged yet. With our setup we can utilize the complete time-series as the algorithm only has to track the minimum during the second iteration.

Figure 3.15 compares a segment from the contaminated E_x component before and after filtering. We can see that most of the high energy variations have been removed. During times that appear to be contaminated by cow fence noise the filter output is an order of magnitude lower than the original signal, but approaches the recorded time series during quieter segments that seem to be undisturbed.

The corresponding power-spectra in Figure 3.16 show that, in general, the adap-tive filter reduces signal-power significantly. This can be expected though because, regardless of the estimation method used, the high power cow-fence signal will pro-duce spectral leakage, and even the estimates that appear to be between the spikes
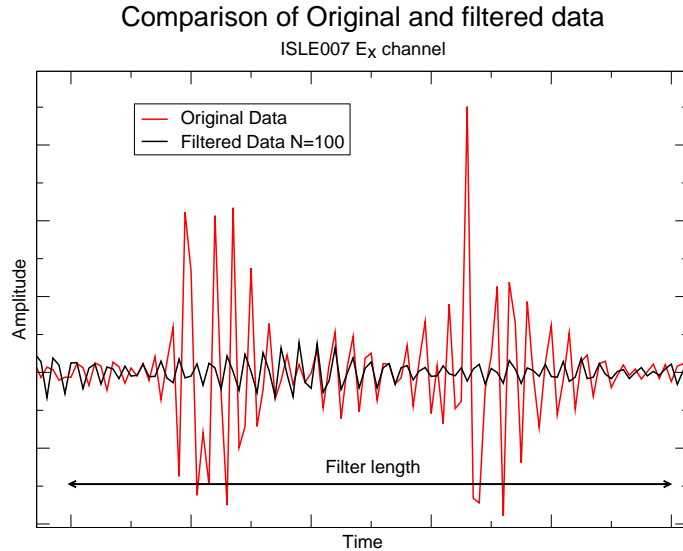
Figure 3.15: Comparison of the E_x component of site ISLE007 before (red) and after (black) application of a LMS adaptive filter with 100 filter coefficients.

are biased towards higher values. This is also the main reason that the cow-fence spikes produce higher apparent resistivities in the impedance estimates at high frequencies. In Figure 3.16 we can also observe the aforementioned spectral resolution characteristics of adaptive filters of differing length. The spectrum obtained after filtering with a filter of length N = 10 results in a smooth spectrum without large variations. With increasing filter length N the resulting spectra show more and more small scale variation.

It is reassuring that all power spectra of filtered time-series now reveal the Schumann resonances at 8 Hz and 14 Hz that are completely obscured in the original time series. Another encouraging observation is the overall agreement with the spectrum from site ISLE101. The exact shape of the electric field spectrum depends on the time of recording and the resistivity distribution in the sub-surface, so a direct comparison is not necessarily instructive. Nonetheless, for a site in a similar geologic setting and without unusual geomagnetic activity, we can expect at least a similar overall shape of the spectra. We observe this similar shape for the spectra after processing with LMS-filters of lengths N = 100 and N = 500, respectively.

We therefore continue and process the data, first with a very simple least-square type algorithm based on the principles outlined in chapter 3.1. Figure 3.17 shows individual estimates of the Z_{xy} impedance element for short segments of the original and filtered data. In addition the mean and the median, calculated separately for real and imaginary part, are plotted. These two plots show the potentially

Comparison of power spectra

LMS adaptive filter
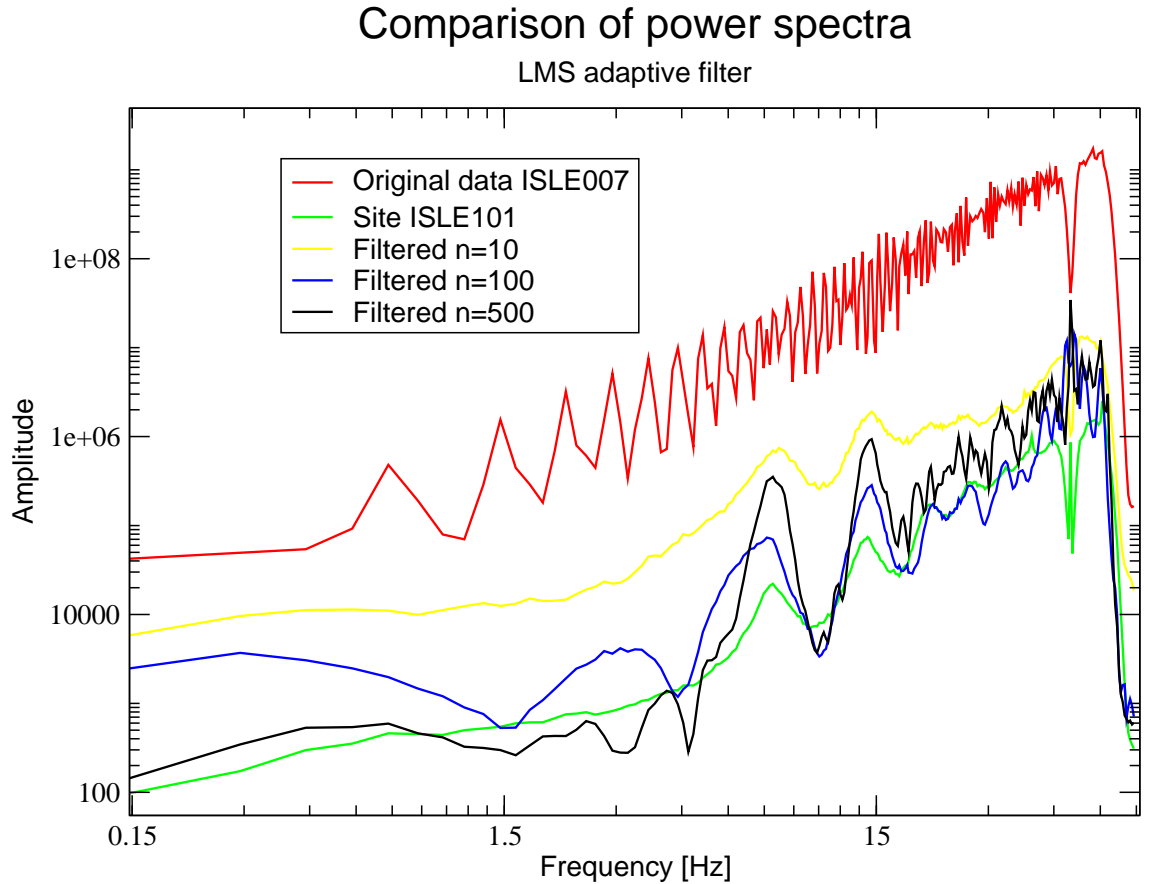
Figure 3.16: Comparison of the E_x component of site ISLE007 before (red) and after application of a LMS adaptive filter with a different number of filter coefficients. For comparison we also plot the spectrum for the undisturbed site ISLE101.

drastic impact of the filtering process on the estimates. Particularly for the estimate at 10 Hz the scatter of the individual estimates is strongly reduced and the filtered estimates concentrate in a small subregion of the original cloud of points. This fact is also reflected in the differing mean and median values at this frequency. At 2.5 Hz the action of the filter seems to be less strong. The scatter is also reduced and we see a considerable difference in the corresponding mean and median values, but the distribution of the estimates after filtering fills most of the area of the original distribution.

While initially the much more consistent impedance estimates at 10 Hz might be taken as an indicator of highly improved data quality, we have to consider the operation method of the adaptive filter. After the initial convergence stage the weights of the filter will model the transfer function between the input, in our case the magnetic channels at a remote site, and the reference signal, i.e., one
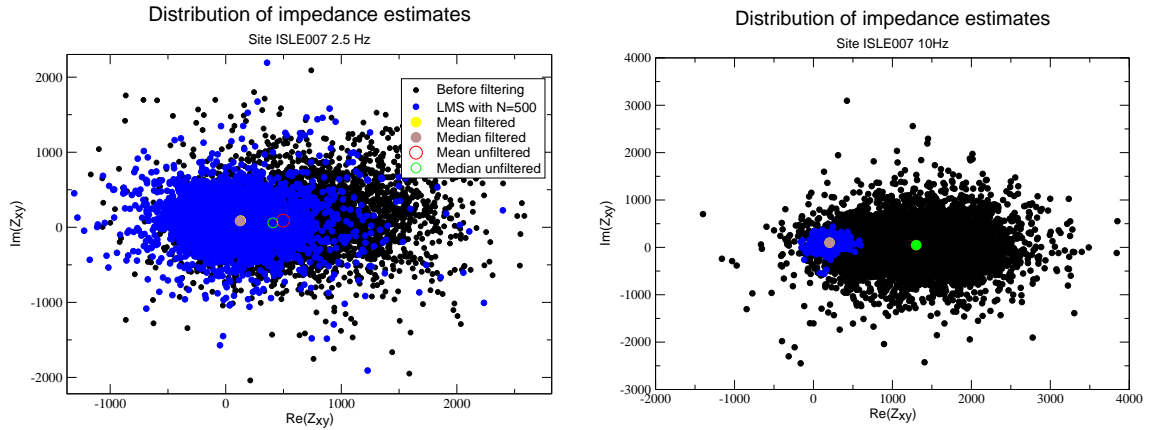
Figure 3.17: Comparison of the Z_{xy} impedance estimate for site ISLE007 before (black) and after (blue) application of a LMS adaptive filter with 500 filter coefficients. The left hand panel shows the estimates at a frequency of 2.5 Hz, the right hand panel at 10 Hz.

of the electric or magnetic channels. Through the adaption process this transfer function can change with time, but as long as the coherent parts of the signal can be described by this transfer function, the weights will not change. If the natural signal is contained in both remote and local recordings and can be described by a time-invariant source field, a stable transfer function between the local and remote channels exists. Thus the adaption process will attempt to keep the filter weights close to this transfer function and departures from this state are caused by high amplitude disturbances, but are usually fairly small. This applies to all local channels equally and consequently the impedance estimates will by design be relatively similar throughout the recording period. The relatively high scatter for the estimates at 2.5 Hz therefore indicates strong disturbance of the filter from the optimum state by the cow fence spikes. In contrast the concentrated cloud of points at 10 Hz only signals that the filtering algorithm stayed in a stable state throughout the filtering process, but does not directly reflect the quality of the data.

We now examine the sounding curves after processing with the robust remote-reference code by Jones and Jödicke [1984], as we did for the template substraction method. Figure 3.18 shows apparent resistivity and phase for the two off-diagonal elements in the critical frequency range of $0.01 - 10$ s for the filtered and unfiltered time-series. Similar to the template substraction method we observe a smoother apparent resistivity curve, particularly for Z_{xy}. Z_{yx} exhibits more scatter, particularly in the phase, despite very small error bars. The small error bars are a result of the time invariant impedance estimates discussed above and are certainly
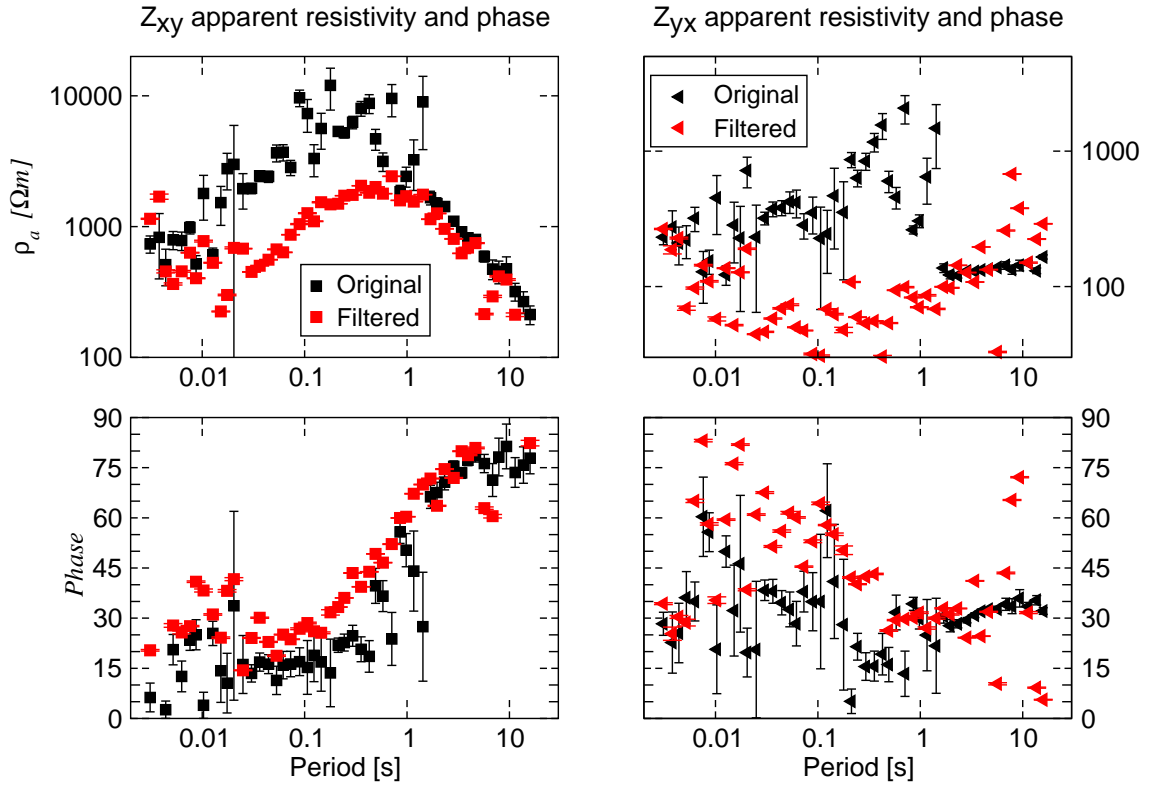
Figure 3.18: Off diagonal elements of the impedance tensor for site ISLE007 for the period range 0.003-10 s. Black symbols show estimates without pre-processing, red-symbols show estimates after pre-processing. The error estimates for the processed time-series are so small that they only appear as vertical bars.

grossly underestimated. In comparison with the template substraction results and the original estimates, the LMS-filtered results appear to be better than both for the Z_{xy} component and worse than both for Z_{yx}.

Finally we will look at the filter weights and their spectral characteristics. Figures 3.19 and 3.20 show the filter weights for the last output sample of the H_x and E_x components sampled at 15 Hz and 150 Hz, respectively. For the relatively unproblematic north-south magnetic component recorded at 15 Hz (Figure 3.19) we can see that the main filter weight is predominantly a single value at coefficient 150 for the H_x remote weights. This demonstrates the consistency of the magnetic field recordings. The location of the spike at coefficient 150 reflects the shift we introduced between the remote and local recordings in order to be able to model a wider range of filter transfer functions. If the two channels were exactly equal, this coefficient should be unity and all others zero. Instead the value is ≈ 0.6 with some small adjacent non-zero coefficients and a small contribution from the H_y coefficients. In comparison, the coefficients for the H_y component are two orders

47
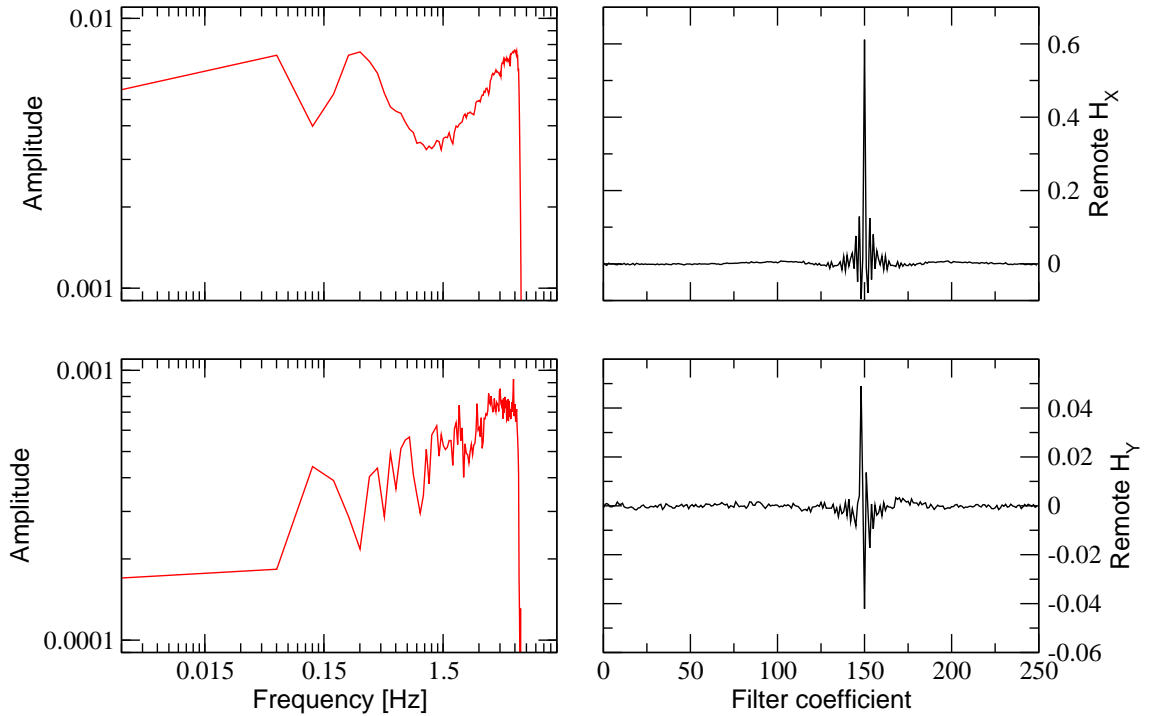
Figure 3.19: Final filter weights (black) of the LMS filter and corresponding spectra for the H_x channel of site ISLE007 sampled at 15 Hz (Band 5).

of magnitude smaller. They most likely represent a small amount of coherent signal due to imperfect alignment of the magnetic coils at the respective sites. The corresponding spectra show some variation with a minimum in the H_x spectrum around 1Hz where the cow-fence noise is concentrated. In general though both spectra do not show a large dynamic range.

The picture changes drastically when we look at the filter weights for the E_x component sampled at 150 Hz. We can still identify a maximum coefficient at a shift of 150 samples, now for the remote H_y component. In addition there are a large number of non zero coefficients with similar magnitude. The fact that the best correlation occurs between E_x and H_y can be expected, as we saw before from equation 2.29. The coefficient for H_x, however, is of similar magnitude and reflects the coherence between these two components, as expressed by the impedance tensor. In that sense the filter weights indicate that there is a natural signal in the recordings that obeys the laws of induction. In contrast, the broad range of non-zero values elsewhere reflects the strong disturbance by the cow-fence signal. The spectral signature of the filter weights is also very interesting: There is a large amplitude maximum at 50 Hz in the spectra, indicating that here we find the most coherent signal. This coincides with the notch filter applied during the recording
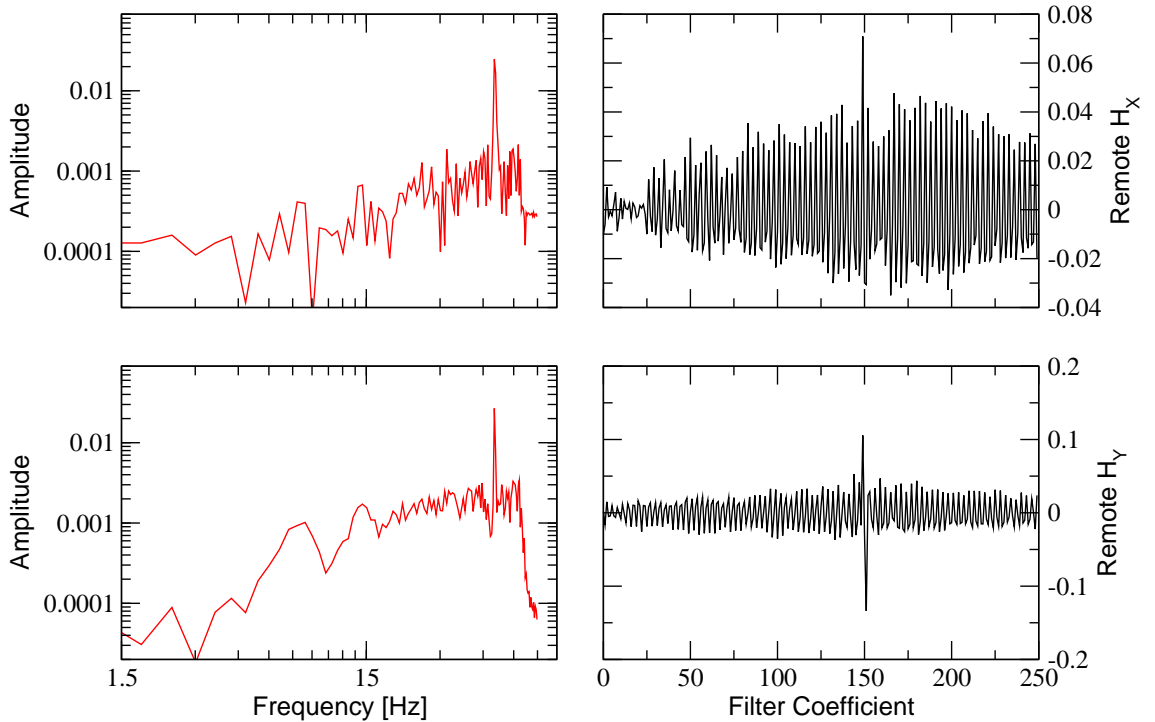
Figure 3.20: Final filter weights (black) of the LMS filter and corresponding spectra for the E_x channel of site ISLE007 sampled at 150Hz (Band 4).

to suppress the 50 Hz household electricity signal. The filter can let the signal at this frequency pass, simply because the amplitudes are extremely low. Two more maxima, particularly visible in the H_y section, around 8 Hz and 14 Hz correspond to the Schumann resonances mentioned above. A number of strong minima in the H_x section indicate strong disturbance at these frequencies, but it is difficult to associate them with a particular feature in the time-series spectrum.

3.4.3 The RLS adaptive filter

As shown in the previous section, the LMS adaptive filter works reasonably well on the contaminated data, but the quality of the sounding curve is still questionable. We will therefore try a more complicated linear filtering algorithm to see whether we can improve the results further. As mentioned before, the LMS adaptive filter is based on a steepest descent type optimization algorithm. In many situations, where the performance surface is not purely quadratic, the direction of steepest decent is not necessarily the best way to approach the minimum. Instead, the *Recursive Least Squares*-algorithm uses curvature information similar to the Newton method for optimization to improve the convergence rate.

Another issue that appears in analogy to inversion methods is *regularization*. We defer an extended discussion to Chapter 5 and only state here that we modify the minimization criterion to include the norm of the weight vector. The LMS error function only depends on the misfit at the current iteration. In contrast the RLS error function utilizes information from the previous iterations and includes a regularization term. The new error function $\mathcal{E}$ at time step i is then

$$\mathcal{E}_i = \sum_{n=1}^{i} \lambda^{i-n} |\epsilon_n^2| + \delta\,\lambda^i \|\mathbf{w}_i\|. \tag{3.34}$$

We can see that the error at the current iteration is the sum of errors from previous iterations weighted by the factor λ^{i-n} plus a weighted regularization term. λ lies in the interval $0 < \lambda \leq 1$ and controls how fast previous values of ϵ are forgotten and for how long the regularization term is applied. For $\lambda = 0$ the formula simplifies to the error function for the LMS filter and for $\lambda = 1$ the errors from all iterations contribute equally. In practice λ is usually chosen slightly less than unity, to utilize the information from previous times on the one hand, but emphasize the current filter error on the other hand. In term of the adaptive behaviour of the algorithm the *forgetting factor* λ plays a similar role to the step-size μ of the LMS-algorithm. δ, the other user defined parameter in the algorithm, is equivalent to the *Lagrangian Multiplier* in regularized inversion and is usually small for filtering. Note also that the regularization term decreases with each time-step i. The regularization is only applied during the initial phase, where it is most critical.

Algorithm 3 The recursive-least-squares (RLS) algorithm

1: $\mathbf{w}_0 = 0$
2: $\mathbf{P}_0 = \delta^{-1} \mathbf{I}$
3: **for all time samples i do**
4:      $\pi_i = \mathbf{P}_{i-1}\,\mathbf{x}_i$
5:      $\mathbf{k}_i = \frac{\pi_i}{\lambda + \mathbf{x}_i^\mathsf{T}\,\pi_i}$
6:      $\epsilon_i = d_i - \mathbf{w}_{i-1}^\mathsf{T}\,\mathbf{x}_i$
7:      $\mathbf{w}_i = \mathbf{w}_{i-1} + \mathbf{k}_i \epsilon_i$
8:      $\mathbf{P}_i = \lambda^{-1}\,\mathbf{P}_{i-1} - \lambda^{-1}\,\mathbf{k}_i\,\mathbf{x}_i^\mathsf{T}\,\mathbf{P}_{i-1}$
9: **end for**

A detailed derivation of the RLS-algorithm is given in Haykin [2002]. We will focus here on the practical aspects; the necessary computations are summarized in Algorithm 3.

Before we discuss more practical issues it is instructive to make the link to Newton-style optimization algorithms more explicit. Combining lines 4, 5 and 7
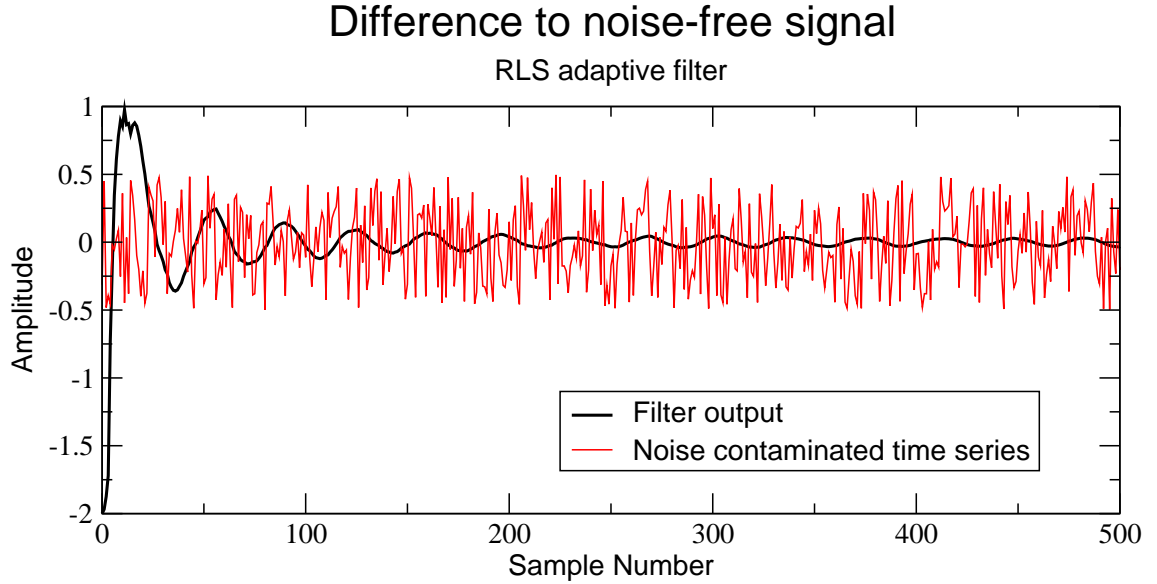
Figure 3.21: Analogue illustration to Figure 3.14 for the RLS-filter. We only plot the difference to the noise free signal (right hand panel in Figure 3.14) as it shows the difference to the LMS-filter more clearly.

of the algorithm we obtain

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mathbf{k}_i \xi_i = \mathbf{w}_{i-1} + \frac{1}{\lambda + \mathbf{x}_i^{\mathsf{T}} \pi_i} \mathbf{P}_{i-1} \mathbf{x}_i \epsilon_i. \tag{3.35}$$

The term $\mathbf{x}_i \epsilon_i$ is a scaled version of the gradient (see equation 3.23), so we can identify the matrix $\mathbf{P}_{i-1}$ with the inverse of the Hessian in a Newton-type optimization algorithm (see equation 5.15). Thus, we can expect similar advantages in conversion rate over a steepest descent type algorithm, like the LMS algorithm described above. This improved performance comes at the cost of much higher computational complexity. Whereas for the LMS algorithm we only had to update a vector of length n at each step, where n is the filter length, for the RLS algorithm we now need to update an $n \times n$ matrix. The computational complexity is therefore $\mathcal{O}(n^2)$ for the RLS algorithm compared with $\mathcal{O}(n)$ for the LMS algorithm. For example, for a filter length $n = 50$ applying an LMS-filter to a typical ISLE-MT recording of 500,000 samples takes about 2 minutes, compared with more than two hours for an RLS-filter.

We perform the same basic functionality test with the RLS algorithm as for the LMS algorithm, and apply it to a simple sinusoidal signal with added noise. The result can be seen in Figure 3.21. The difference between the filter output and the noise free signal approaches zero in only a few iterations. The prediction error of the LMS algorithm only reaches the noise level at about 700 iteration (see Figure
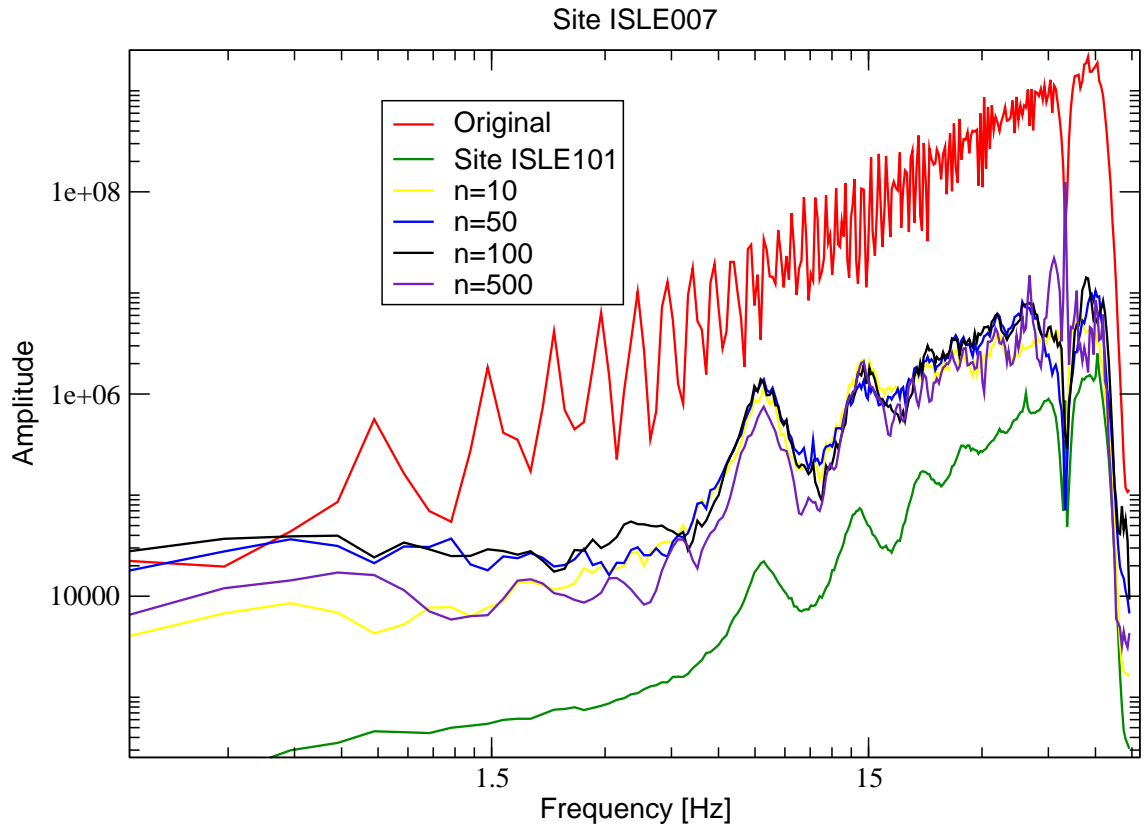
Figure 3.22: Comparison of the E_x component of site ISLE007 before (red) and after application of a RLS adaptive filter with a different number of filter coefficients. For comparison we also plot the spectrum for the undisturbed site ISLE101.

3.14), whereas the RLS algorithm achieves this in less than 50 iterations. This nicely illustrates the beneficial effects of incorporating curvature information in the optimization process.

We apply the RLS-filter with exactly the same setup as for the LMS-filter: The only difference between the two lies in the way the weights are updated, otherwise the philosophy remains the same. Tests with varying values of the regularization factor δ show that it has virtually no impact on our results for values between 10^6 and 10^{-6}. As before, we run the filter twice on each component to ensure that the initial convergence stage is completed. At that point the regularization term has already vanished and does not play any role in the further performance of the algorithm.

Figure 3.22 compares the power spectra for the E_x component before and after filtering with an RLS-filter in a similar manner as Figure 3.16 for the LMS-filter.

For a filterlength of $n = 500$ the high computational cost only allowed us to process the Band 4 time-series. A single step of the algorithm takes 0.2 s on a 1.6 Ghz Pentium4M using the highly optimized ATLAS linear algebra package [Whaley et al., 2001]. For the 500,000 samples of the Band 4 recordings the resulting run time is 1.2 days per component. We tried to run the computations for the other two bands on the DIAS owned LEDA cluster and a cluster of the SHARCNET consortium [SHARCNET]. In both cases interruptions in the operation of the cluster prevented completion of the filtering process a number of times and after one month we gave up.

In comparison to the LMS algorithm, the variation with filterlength is quite different. While the spectra of the filtered time-series showed substantial variation with power for the LMS-algorithm, the RLS-filtered results show very similar signal power, particularly at high frequencies. All spectral plots for the RLS-filter resemble the spectrum after filtering with an LMS-filter with $n = 10$ and we do not reach the spectral power of site ISLE101 as we did for the longer LMS-filters. There are several possible explanations for this phenomenon: The agreement of power-spectra for the RLS-filters with different lengths might indicate that this is the true power spectrum at site ISLE007 with the cow-fence noise removed. Alternatively the short LMS-filter and the RLS-filter might fail to completely remove the cow-fence noise, or the LMS filter results a higher overall reduction in power for all channels.

Figure 3.23 shows a comparison between the original and filtered E_x-component for a LMS-filter with $n = 500$ and a RLS-filter with $n = 50$. We also plot the magnetic input channels that were used to generate the filtered versions of the electric channel. The first thing to note is the strongly decreased amplitude for the electric component that we observed before and also saw in the corresponding spectra. Even though the original electric field is plotted with a scale of twice the range of the filtered counterparts, the time-series clips the plotting area for a substantial part of the plotted time. The overall appearance of the two filtered time-series is very similar, but, as also previously observed, the amplitude of the RLS-filtered time-series is generally higher than the LMS-filtered component. We can now also identify segments in the remote magnetic channels that have similar shape to the filtered electric channels. This can of course be expected, because the filtered electric channels were produced from the magnetic channels.

As before with the LMS-filters, we will now proceed and look at the results of processing the filtered time-series. Figure 3.24 shows the comparison of processed results between the original data and data after filtering with the RLS algorithm. In essence the results are very similar to the results with the LMS-adaptive filter. We obtain lower apparent resistivity values for both components and for the Z_{xy}
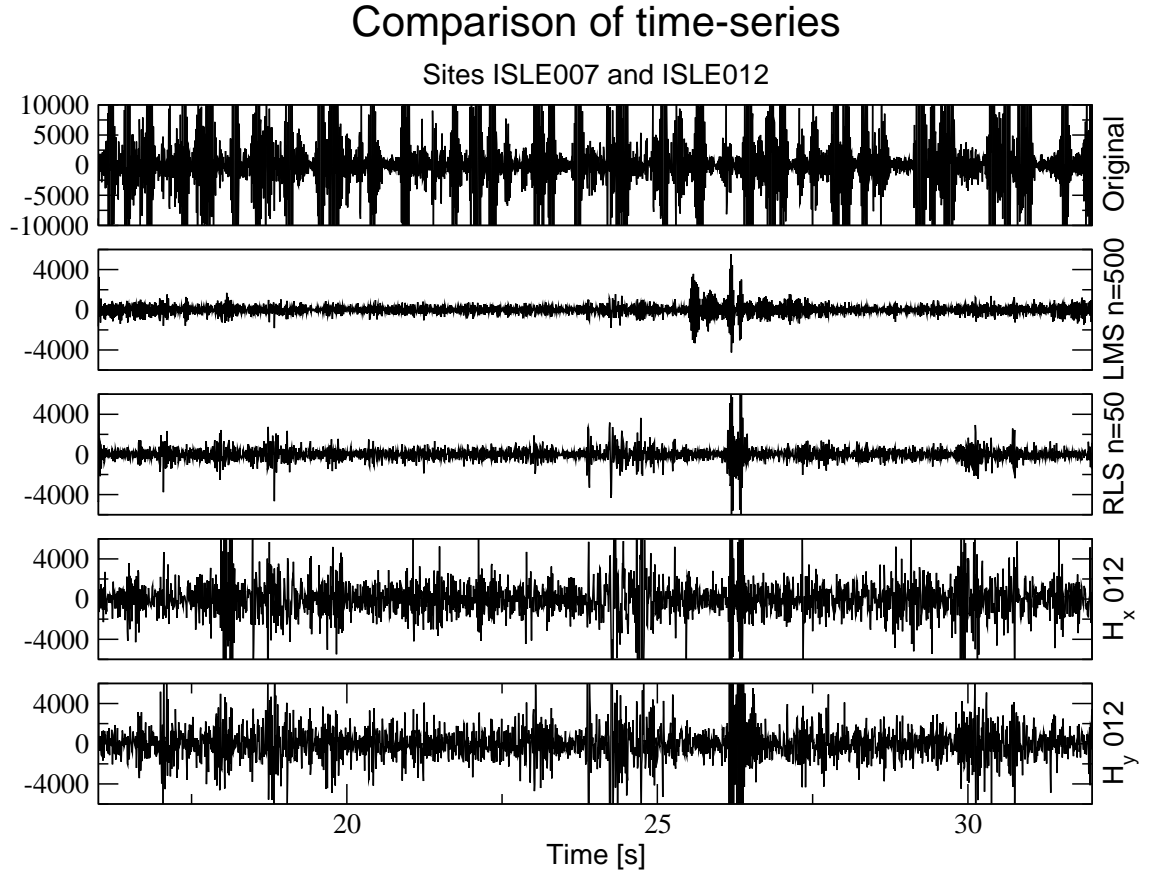
Figure 3.23: Comparison of the E_x component of site ISLE007 before (top) and after application of a LMS-filter and a RLS-filter. We also plot the two remote magnetic input channels from site ISLE012. Note the different scale for the original E_x component and the two filtered plots.

component the overall appearance is also smoother. In comparison with the LMS filtered results, the phase of both components exhibits more scatter between adjacent frequencies. It is unclear from this comparison whether this is due to the different algorithms or the variation in filter length from $n = 500$ for the LMS filter and $n = 50$ here, but comparison with results obtained from processing the data with a LMS filter with $n = 50$ (not shown), reveals that compared with those results, the RLS-filter performs slightly better. Another observation that is related to the shorter filterlength for the RLS-filter is the drop in apparent resistivity and the sharp rise in phase for long periods particularly visible in the yx-component.

As we mentioned before we only have a filtered time-series for a single band for the RLS filter with $n = 500$. Still, we want to present these results in order to assess whether it is worthwhile investing the considerable effort to apply long
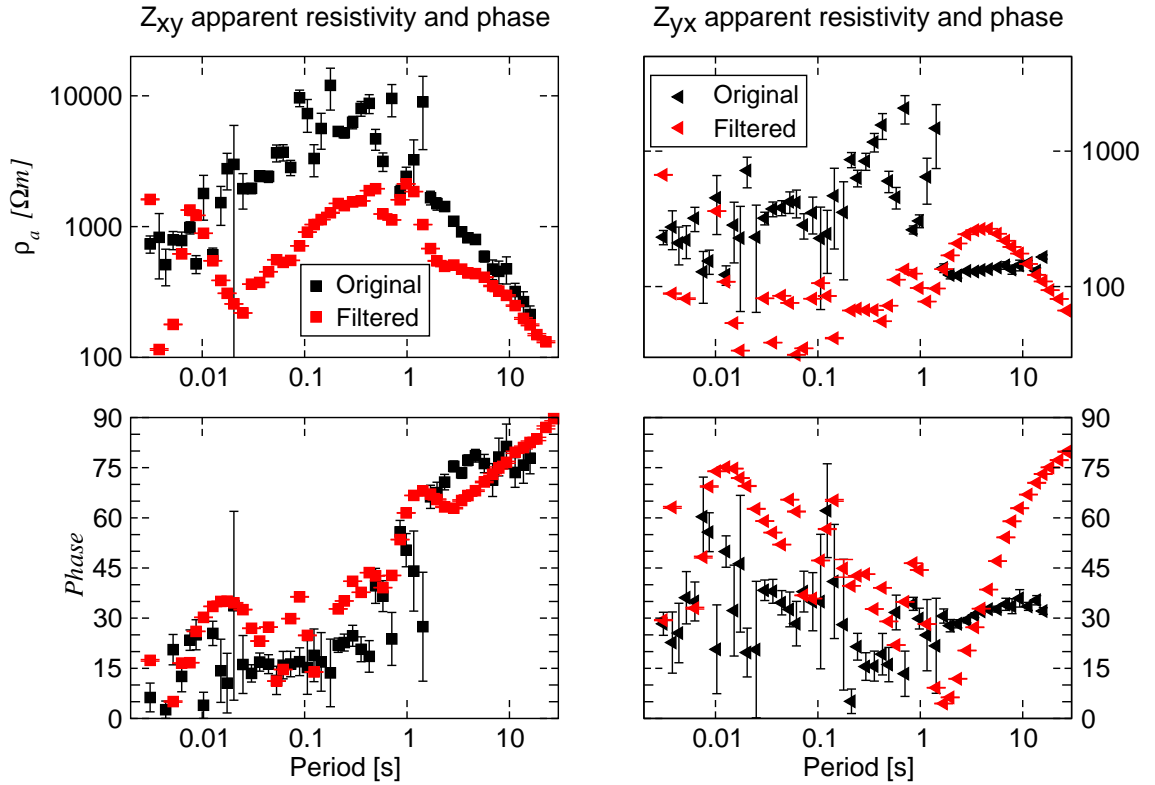
Figure 3.24: Off diagonal elements of the impedance tensor for site ISLE007. Black symbols show original estimates, red-symbols show estimates after application of a RLS-filter with $n = 50$.

RLS-type filters to the data. Figure 3.25 shows a comparison between the processed and original estimates as before. The impedance estimate after filtering are comparable with the estimates after applying a LMS-filter of the same length. For both polarizations the phase appears to be marginally more smooth, but the apparent resistivities, on this plotting scale, seem to be identical. Due to the limited range of estimates from the processed time-series, it is difficult to judge whether this method provides any improvement over the LMS-filters.

3.5 Neural networks for filtering

Linear adaptive filters assume that the filter output is a linear combination of the input values. As explained above this assumption is reasonable for magnetotelluric data where the filter weights then model the transfer functions between the different channels. Still, the question remains whether this kind of model and the corresponding algorithms provide optimum performance in the presence of high
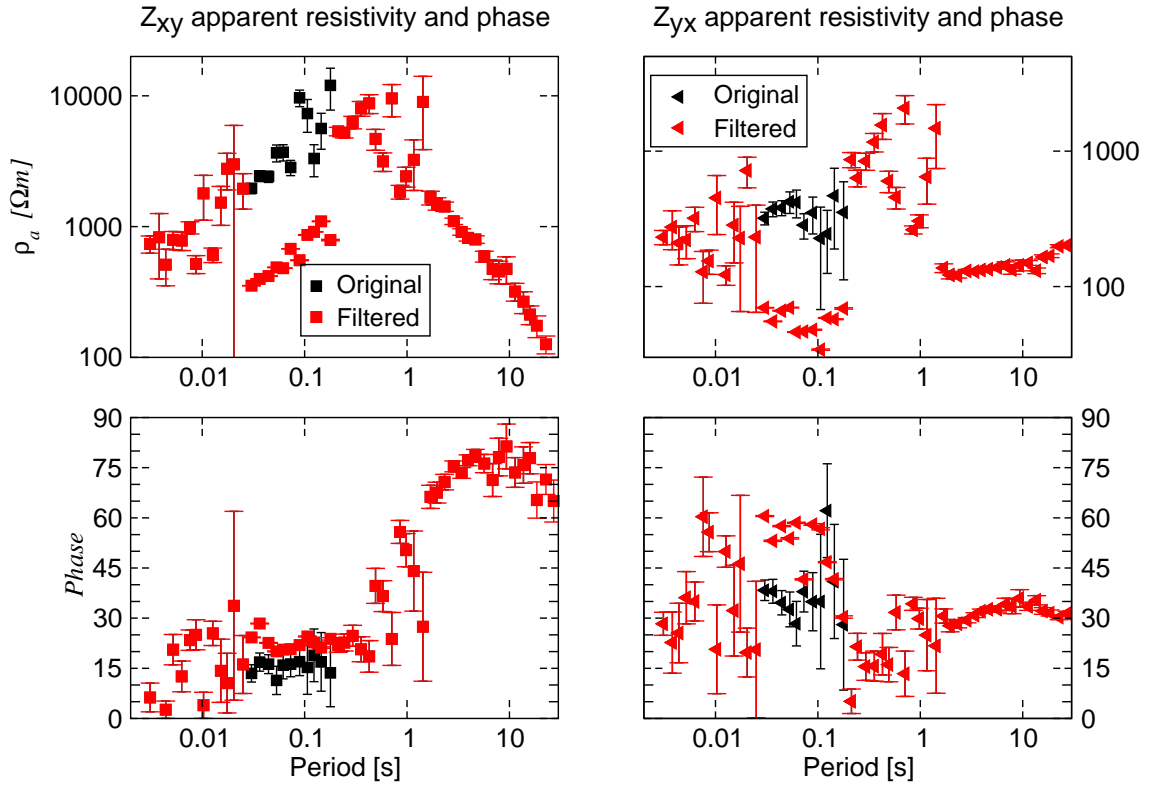
Figure 3.25: Off diagonal elements of the impedance tensor for site ISLE007. Black symbols show original estimates, red-symbols show estimates after application of a RLS-filter with $n = 500$. Note that only the Band 4 data has been filtered, therefore all estimates that are calculated from a different band are identical to the original, unprocessed estimates.

amplitude noise. We will therefore also explore non-linear filtering methods to get an idea of the limitations of the linear approach.

Neural networks have received a lot of attention particularly in the field of artificial intelligence and pattern recognition [Raiche, 1991], but also for inversion of geophysical data [Meier et al., 2007]. We explore the abilities of simple neural networks for filtering purposes following Haykin [2002]. Manoj and Nagarajan [2003] describe the use of neural networks for magnetotelluric processing. Their approach is based on the identification of suitable segments of the recordings that are not contaminated by noise. They report good success with this method for their data. Unfortunately it is not suitable for our purposes as there are no good segments of usable length in the contaminated records; the cow fence noise extends through the complete time-series.

The basic building blocks of neural networks are *neurons* which form an analogy to the biological neurons in the nervous system of animals. The common *sigmoidal*
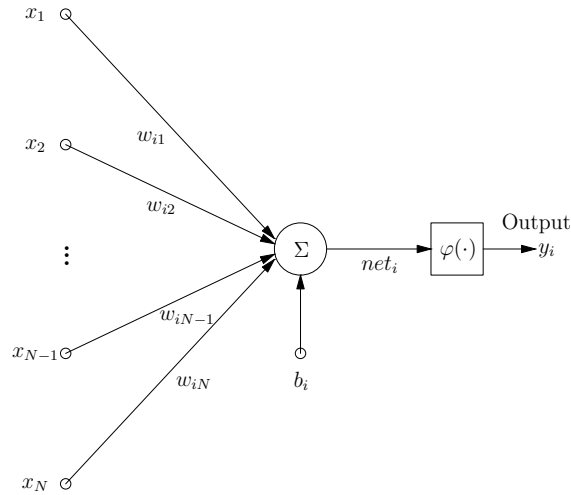
Figure 3.26: Schematic overview of a sigmoidal neuron i of a neural network.

model describes the neuron mathematically as a combination of a linear combiner with an additional bias weight b and a non-linear *activation function* $\varphi(\cdot)$. A schematic overview is given in Figure 3.26. The output net_i of the linear adaptive combiner

$$net_i = \sum_{j=1}^{N} w_{ij}x_i + b_i \tag{3.36}$$

is used as input into the scalar activation function. A number of choices exist for this function, the typical sigmoidal neuron uses the logistic function as an activation function

$$\varphi(net_i) = \frac{1}{1 + \exp(-net_i)}. \tag{3.37}$$

This function only assumes positive values, for all input values and is therefore called *unipolar*. We need a *bipolar* activation function, that can also yield negative values and choose

$$\varphi(net_i) = \tanh\left(\frac{1}{2}net_i\right) \tag{3.38}$$

for the filtering.

The individual neurons can now be combined into networks of different topology, depending on the application. We will use a setup known as *focused time-lagged feedforward network*, a special case of a *multilayer perceptron*, shown in Figure 3.27. The term *feedforward* refers to the fact that the signal only propagates in the forward direction, from left to right in Figure 3.27, without any feedback. It is called *focus time-lagged* because time is accounted for in the tapped delay line on
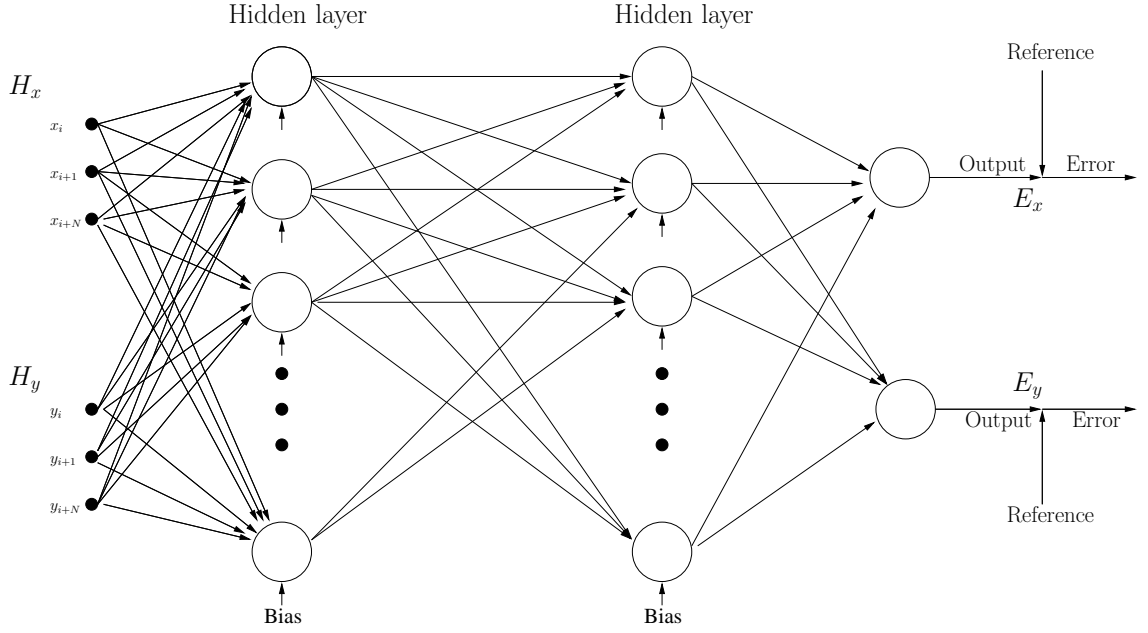
Figure 3.27: Setup of a focused time-lagged feedforward network with two hidden layers for predicting the local electric field from the magnetic field of a remote site. Each circle corresponds to a neuron as shown in Figure 3.26.

the input side [Haykin, 2002]. Other possibilities to include time in the learning process exist, but we will not consider them here.

The calculation of the output signal and adaptation of weights is very similar to the LMS-adaptive filter on which the neurons are based, only that we have to account for the non-linear activation function and the propagation through the hidden layers. For a network with M layers, the output of neuron i in layer l is given by

$$x_i^l = \varphi \left(\sum_{p=1}^{N} w_{ip}^l x_p^{l-1} + b_i^l \right) \text{ with } l = 1, \ldots, M, \tag{3.39}$$

where N is the number of neurons in the previous layer that this neuron is connected to.

Again we calculate the error signal, the difference between the network output and the reference signal, but we can now have a number of reference channels

$$e_i(n) = d_i(n) - y_i(n). \tag{3.40}$$

We use this error to adjust the weights and bias in a steepest descend fashion, now starting at the output layer and propagating the corrections backwards. This
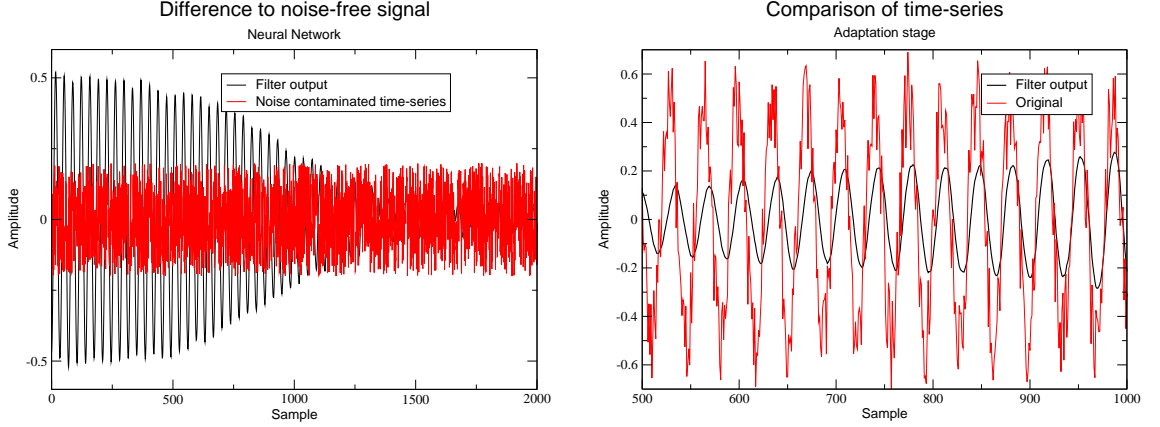
Figure 3.28: Simple performance test for a neural network with one hidden layer.

adaption algorithm is therefore known as the *backpropagation algorithm*,

$$w_{ij}^l(n+1) \;=\; w_{ij}^l(n) + \mu x_j(n)\delta_i^l(n), \tag{3.41}$$

$$b_i^l(n+1) \;=\; b_i^l(n) + \mu\delta_i^l(n), \tag{3.42}$$

and

$$\delta_i^l(n) = \begin{cases} \varphi'\left(net_i^l\right)\,[d_i(n) - y_i(n)] & l = M \\[1.5em] \varphi'\left(net_i^l\right)\,\sum_k w_{ki}\delta_k^{l+1}(n) & 1 \le l < M \end{cases}. \tag{3.43}$$

For the bipolar activation function that we use for our filtering

$$\varphi'\left(net_i^l\right) = \frac{1}{2\cosh\left(net_i^l\right)^2}. \tag{3.44}$$

These equations fully specify the evolution of weights and output and we can now perform our standard performance test that we used to test all adaptive algorithms. We have to make one modification though, as the output of the neural network lies in the range $(-1, 1)$. We therefore scale the problem and modify the reference signal to

$$d_i = 0.5\cos(2\pi i/N) + rand(-0.2, 0.2). \tag{3.45}$$

The results can be found in Figure 3.28.

We can see that it takes a little more than 1,000 samples before the filter output reaches the noise-level of the contaminated time-series, comparable with the adaptation time of the LMS-filter, as both are based on steepest descent type algorithms. The shape of the envelope of the filter error is very different though. The LMS-filter achieves the best improvement in the first few iterations (see Figure 3.14), while the maximum error for the neural network stays roughly constant
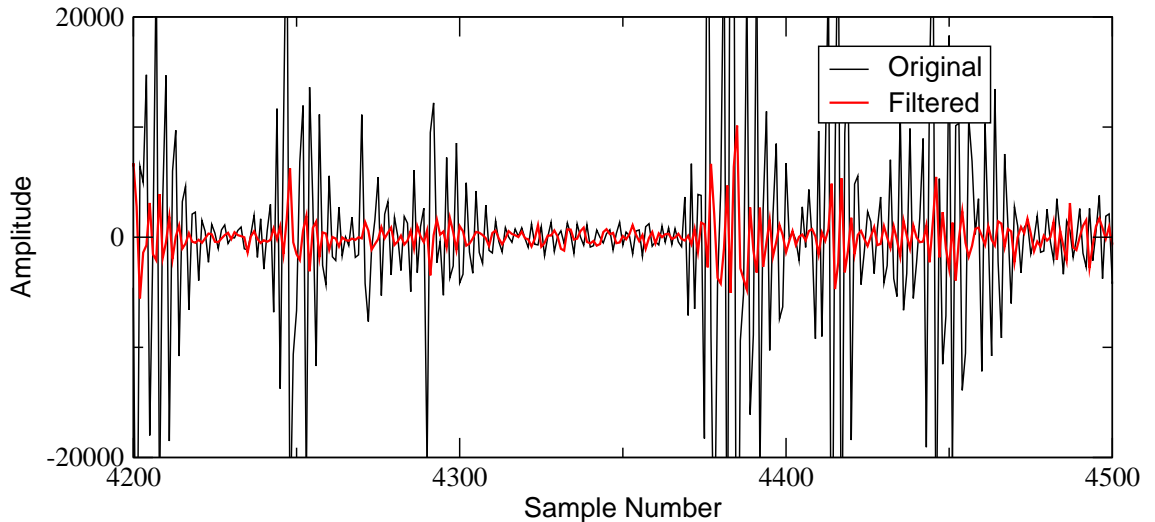
Figure 3.29: Comparison of the E_x time-series of site ISLE007 before (black) and after (red) application of a Neural Network filter with 100 filter coefficients.

during the first few hundred iterations and then rapidly falls below the original noise level. This behaviour demonstrates the non-linear nature of the algorithm. If we look in detail at the filter output during the rapid adaptation process, we can see that the filter first adapts to match the phase of the reference signal and then slowly increases the amplitude to the required level.

3.5.1 Filtering MT data

The layered design of the neural network obviously gives us many more possibilities to explore different settings for optimum filter results. Possible parameters to vary are the number of hidden layers, the number of neurons in each layer, the types of neurons, adaptation rate and the type of input and output. Unfortunately we cannot explore the potential benefits of these variations here. One general observation we made during our filtering experiments was that the adaptation rate greatly decreases with increasing number of hidden layers. To some degree this can be balanced by a higher adaptation constant, but for our purposes a single hidden layer appears to perform better. Therefore we only show results from Neural Networks with a single hidden layer. We also did not explore different type of non-linearities for the individual neurons due to time restrictions. The bipolar sigmoidal activation function, however, is regarded as a good general purpose function [Haykin, 2002].

Figure 3.29 shows a comparison between a segment of the original E_x-component at our test site and the same component predicted by the neural network. In this

Comparison of power spectra

Neural Network filter
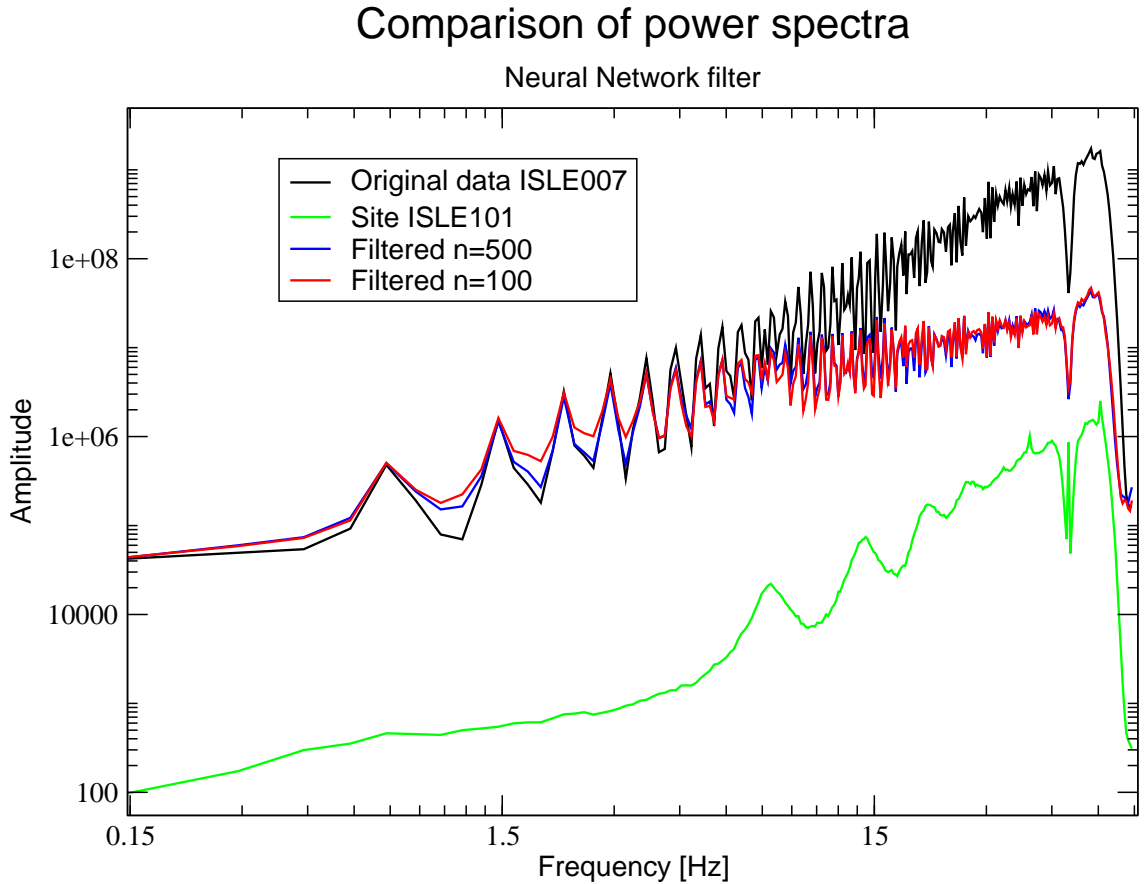
Figure 3.30: Comparison of the E_x component of site ISLE007 before (black) and after application of Neural Network filters with 100 (red) and 500 (blue) filter coefficients. For comparison we also plot the spectrum for the undisturbed site ISLE101.

case we used two remote magnetic channels as input, as for the linear adaptive filters, but now predicted all 4 channels of the local site simultaneously, similar to the setup shown in Figure 3.27 for only two local electric channels. The total filter length was 100 in this case. As before with the other methods, it is difficult to judge the filtered results from a comparison of the time-series. The high-energy segments that correspond to the cow-fence spikes have been reduced to an amplitude comparable to the overall signal strength. In the quieter segments we can see some similarities between the filtered time-series and the original, but at the same time there are some segments that are significantly different.

The plot of the power-spectra in Figure 3.30 gives some more insight into what happened in the filtering process. At frequencies <5 Hz the spectral peaks from the cow-fence disturbances are left unchanged after filtering. At frequencies >5 Hz
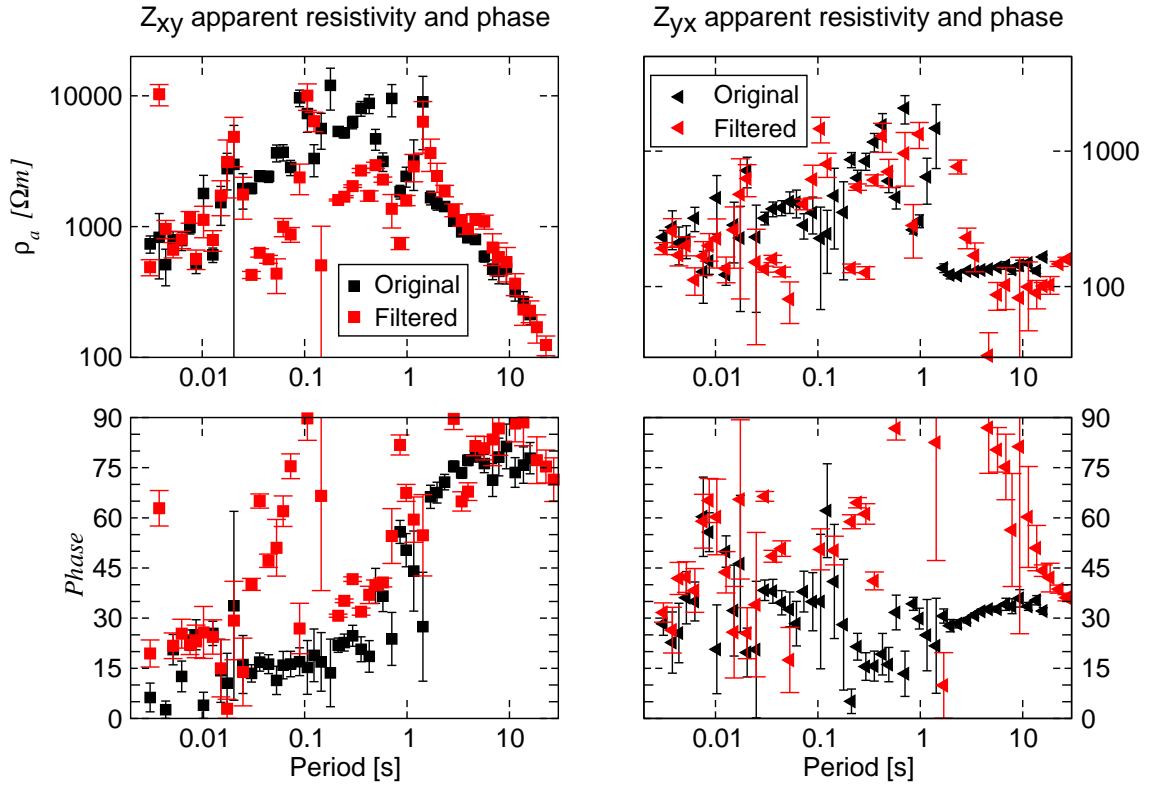
Figure 3.31: Off diagonal elements of the impedance tensor for site ISLE007. Black symbols show original estimates, red-symbols show estimates after application of a Neural Network filter with $n = 500$.

the filtered spectrum has the same small scale appearance as the original, but the overall shape is now flattened. It appears that the main impact of the neural network was an additional whitening of the spectrum without large modification of its local structure. Close comparison, however, reveals that also on a small scale the spectrum has been changed, but not to the extend of the previous methods. Our experience shows that we can only judge the success of a method by comparing the processed impedance elements.

Figure 3.31 shows the comparison between the impedance estimates before and after filtering. Of all filtering methods, the neural network shows the least improvement regardless of filter length. There are some frequencies where we observe similar apparent resistivity and phase values to the linear adaptive filter results, but the scatter between adjacent frequencies is generally higher. For the Z_{yx} component the scatter in phase values even seems to have increased compared with the original impedance estimates and at periods between 0.01 and 0.1 s the phases even leave the expected quadrant. However, for the Z_{xy} component between periods of $0.5 - 1$ s we observe a small number of estimates that seem to have improved
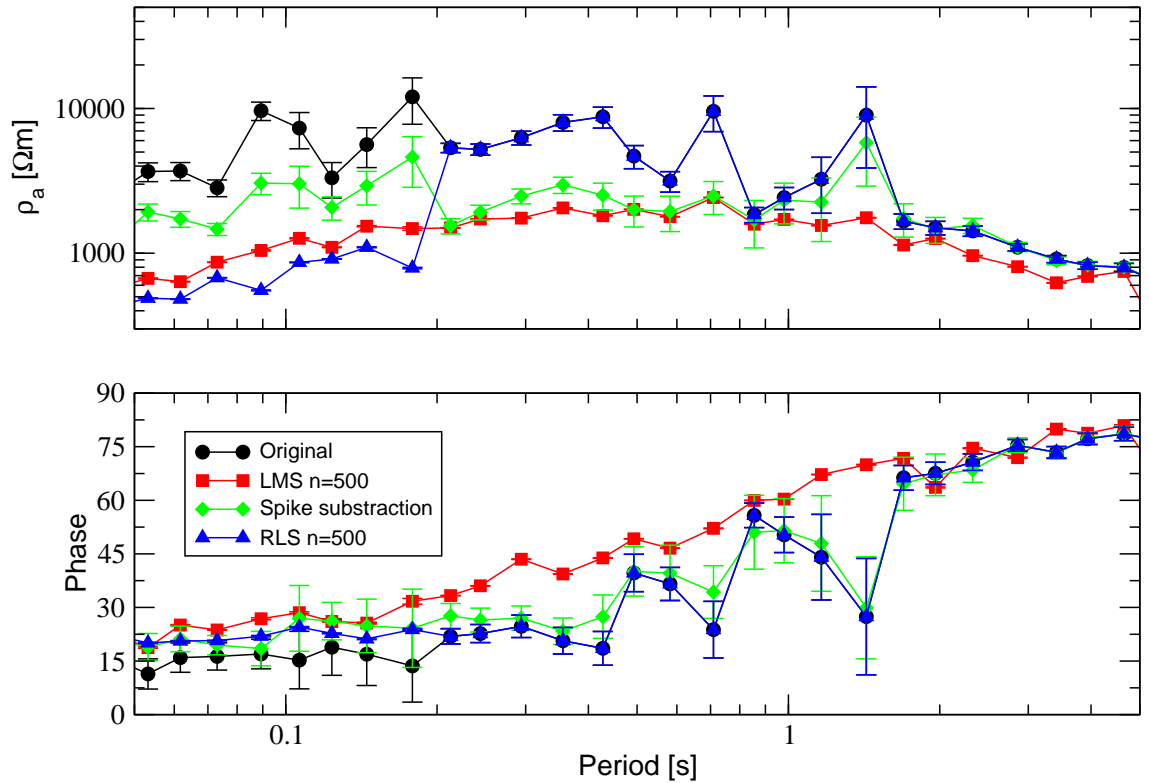
Figure 3.32: Comparison of the original apparent resistivity and phase estimates for the Z_{xy} component some of the filtering methods presented in this chapter.

in the sense that vary smoothly and fit in with the surrounding phase estimates.

3.6 Summary of results

All in all none of the methods produced results that immediately seem to be interpretable. Particularly the Z_{yx} component proved difficult for all signal processing methods and none of the filtered results for this component has sufficient quality to be used as input for modelling or inversion. For the Z_{xy} component we seem to achieve some improvement, but each method yields different impedance estimates.

Figure 3.32 shows a comparison of apparent resistivity and phase estimates for the Z_{xy} component in the critical period range of $0.05 - 0.5$ s. At short periods, between 0.05 s and 0.2 s, the apparent resistivity estimates from each method differ significantly from each other and from the original. The phase, in contrast, agrees within the estimated error. At periods > 0.2 s at least the apparent resistivity estimates from the LMS-filter and the template substraction algorithm show some general agreement. The agreement between the RLS-filter and the original has
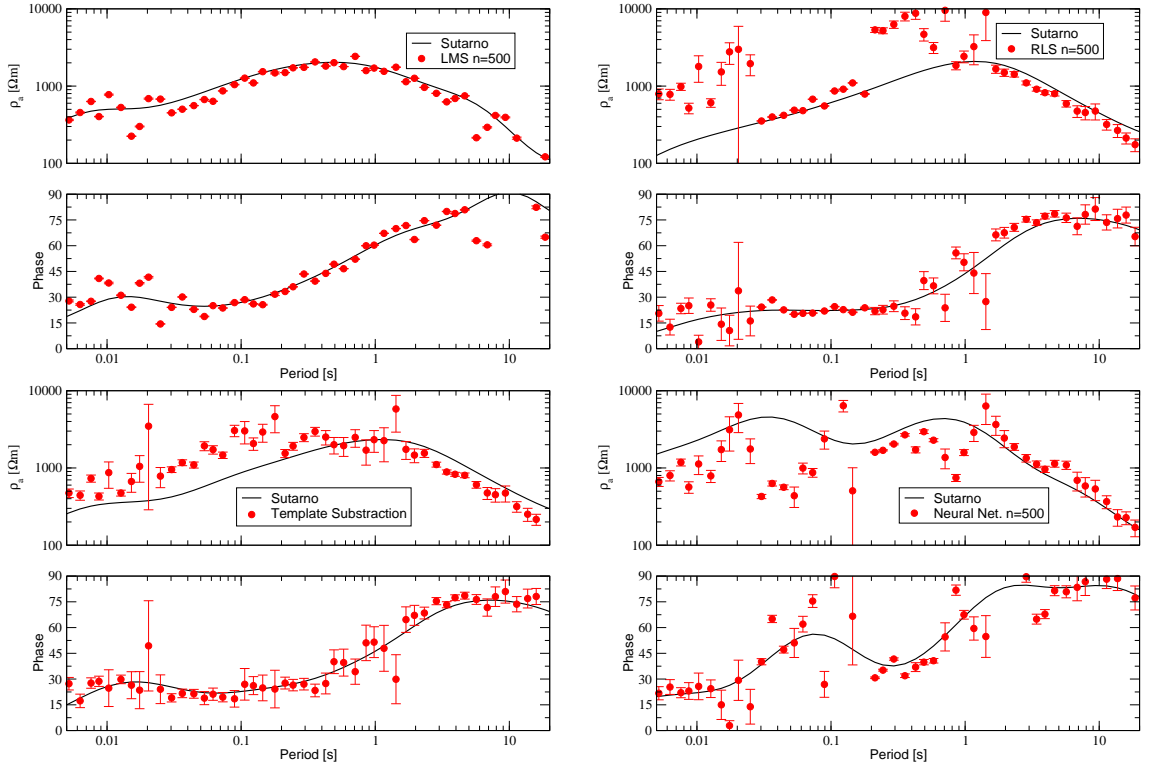
Figure 3.33: Comparison of the estimated apparent resistivities and phases for the Z_{xy} components with the values from the Sutarno phase consistent method.

no significance in this period ranges, as these estimates are calculated from the Band 5 recordings. In terms of phase, all but the LMS-filtered results follow the original phase estimates. The smoothness of the LMS phase estimates however suggests that these are reasonable quality estimates. At periods >2 s all estimates for both phase and apparent resistivity converge. We already speculated from the characteristics of the original data that at long periods the influence of the cowfences vanishes. The agreement of the filtered and unfiltered estimates confirms this observation.

As a final indicator of the quality of the estimates we compare the apparent resistivities with estimates from the *Sutarno phase consistent* method [Sutarno and Vozoff, 1991], as implemented in the WinGLink software package. This method uses the Hilbert transform relation between apparent resistivity and phase to calculate the expected apparent resistivity from the observed phase. Even though this relation is strictly speaking only valid in a 1D environment, or for the TM-mode in 2D, most high-quality data seems to obey this relationship.

Figure 3.33 shows the apparent resistivities and phases together with the Sutarno estimates for various filtering methods. Clearly the neural network based

filter does not create estimates that have apparent resistivities and phases that obey a Hilbert transform relationship. Together with the high scatter between adjacent values, and the out of quadrant phase, this demonstrates that in the form we applied the neural network it is not suitable to remove the noise. We only tested a small fraction of the possibilities of neural networks and it is highly likely that better configurations exist that achieve the same results as the other methods, or even better. Within our limited experimentation we were not able to find such a configuration.

The output from the template substraction algorithm also does not fulfil the Hilbert transform relation. At short periods the estimated apparent resistivities are higher than expected from the Sutarno prediction. The general shape, however, is similar. Both the LMS-filter and the RLS-filter produce consistent results. From the plot of the RLS results, where we have a mix of original and filtered data, we can see that in this case the Sutarno predictions bridge the gap between the reliable original results and the filtered results by following the overall shape of the phase. For the LMS-filter we see consistent results with only minor scatter for periods <5 s.

From this perspective the LMS and RLS results seem to be of similar quality. The difference between the apparent resistivity and phase estimates from the two methods is consistently larger than their estimated error. We remarked before that the error in both cases seem to be greatly underestimated. Using the difference between the two filtered results as a guideline a representative error is $\pm 5°$ in phase and ± 300 Ωm, or 20%, in apparent resistivity. These numbers are also consistent with the error estimates of the unprocessed impedances.

As we do not know the true impedance values for the contaminated site it is impossible to definitively say whether the filtered estimates are more representative of the subsurface. The quality indicators used, the smoothness of adjacent estimates and compliance with a Hilbert transform, indicate that for the Z_{xy} component they at least have the characteristics of regular MT data. For this component the LMS and RLS adaptive filters perform best and both equally well, so that the additional computational complexity of the RLS algorithm seems unjustified. The lack of improvement for the Z_{yx} component demonstrates that LMS-adaptive filters are not a panacea, and that the output has to be examined critically before using it to make inferences about the conductivity of the Earth.

The first principle is that you must not fool yourself and you are
the easiest person to fool

<div align="right">Richard P. Feynman</div>

4

Receiver functions

4.1 Wave propagation in a one dimensional Earth

For the same reasons as before with the theory of magnetotellurics we will only discuss one-dimensional wave-propagation and receiver function analysis. First, the underlying principles are more easily understood, and second, the inversion approach we will describe later is based on a one-dimensional approximation of Earth structure. Furthermore we will not derive all equations from first principles, as we did for the one-dimensional impedance for MT, but only briefly sketch the most important steps. More details can be found in any textbook on seismology [e.g., Lay and Wallace, 1995].

The basic source of energy for all teleseismic methods, such as receiver functions, are earthquakes that regularly occur in tectonically active regions around the world. Stress accumulates in collision zones between different plates, or in some cases even within one plate, and when this stress exceeds the strength of the material, a rupture initiates and propagates along the earthquake fault. The shock wave created by the rupture travels through the Earth and along the surface and the acceleration created by this wave is measured by seismometers around the world. In the process of travelling through materials with different propagation velocities and densities, the initial waveform is scattered and distorted, and the information contained in these distortions is then used to obtain information about the Earth's structure.

The general equation for a wave travelling through the Earth is

$$\nabla \left[(\lambda + 2\mu)\nabla^2\phi - \rho\frac{\partial^2\phi}{\partial t^2} \right] + \nabla \times \left[\mu\,\nabla^2\Psi - \rho\frac{\partial^2\Psi}{\partial t^2} \right] = 0 \qquad (4.1)$$

Where ϕ is a curl-free scalar potential, Ψ a divergence-free vector potential, μ is the *shear modulus*, λ Lamé's second constant and ρ density. Unfortunately two material parameters that we will be concerned with, electrical resistivity and density, are commonly represented by ρ in mathematical equations. Instead of introducing an arbitrary new variable for one of the parameters, we will simply use ρ for either of them where the context is clear and ρ_{el} for electrical resistivity where there is potential for confusion.

A possible solution to equation 4.1 is to set each term to zero independently, and we can simplify the expressions further by setting

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \qquad (4.2)$$

$$\beta = \sqrt{\frac{\mu}{\rho}}. \qquad (4.3)$$

With this we can write

$$\nabla^2\phi - \frac{1}{\alpha^2}\frac{\partial^2\phi}{\partial t} = 0, \qquad (4.4)$$

$$\nabla^2\Psi - \frac{1}{\beta^2}\frac{\partial^2\Psi}{\partial t} = 0. \qquad (4.5)$$

Equation 4.4 describes the propagation of compressional waves through the Earth, while Equation 4.5 describes the propagation of shear waves. α and β are the propagation velocities for the two wave types, respectively. Comparing Equations 4.2 and 4.3 we can see that compressional waves always travel faster than shear waves. For this reason, the common nomenclature is primary or P-waves for compressional waves and secondary or S-waves for shear waves.

A possible set of solutions to the above equations within a homogeneous region away from the source is given by

$$\phi(\mathbf{x}, t) = A\exp\left(\iota\omega t - \mathbf{k}_\alpha\mathbf{x}\right), \qquad (4.6)$$

$$\Psi(\mathbf{x}, t) = B\exp\left(\iota\omega t - \mathbf{k}_\beta\mathbf{x}\right). \qquad (4.7)$$

These equations describe harmonic waves that travel in the direction of the wave-vectors $\mathbf{k}_\alpha$ and $\mathbf{k}_\beta$, respectively, with $|\mathbf{k}_\alpha| = \omega/\alpha$ and $|\mathbf{k}_\beta| = \omega/\beta$. Note the structural similarity to the solutions for the electric and magnetic fields for magnetotellurics in Equations 2.25 and 2.26, respectively, but the different coefficients for the space variable in the exponential.

We can now consider the transmission of energy through an interface in a one-dimensional layered Earth. We will not derive the full expressions for the reflection and refraction coefficients, as the algebra is very tedious and not particularly instructive, but only demonstrate why for an incident P-wave, reflected and refracted SV-waves are generated at an interface. For a one-dimensional Earth we can set $\mathbf{x} = \{x_1, x_2, x_3\} = \{x_1, 0, x_3\}$. Consider a compressional wave travelling from below to an interface at $x_3 = 0$ where both β and α change. If we only consider P-waves we will have the incident and reflected wave solution ϕ_1 in the lower layer and the refracted P-Wave potential ϕ_2 in the upper layer,

$$
\begin{aligned}
\phi_1(\mathbf{x}, t) &= A_1 \exp\left(\iota\omega\left(px_1 + \eta_1 x_3 - t\right)\right) + A_2 \exp\left(\iota\omega\left(px_1 - \eta_1 x_3 - t\right)\right), &(4.8) \\
\phi_2(\mathbf{x}, t) &= A_3 \exp\left(\iota\omega\left(px_1 + \eta_2 x_3 - t\right)\right), &(4.9)
\end{aligned}
$$

where the *horizontal slowness* or *ray parameter* p and *vertical slowness* η are given by

$$
p = \frac{\sin i}{\alpha}, \tag{4.10}
$$

$$
\eta = \sqrt{\frac{1}{\alpha^2} - p^2} = \frac{\cos i}{\alpha}, \tag{4.11}
$$

and i is the angle of incidence of the wave.

Across the boundary we have to satisfy two conditions: 1) continuity of displacement and 2) continuity of stress. For our purposes it is sufficient to examine the continuity of displacement

$$
\left(\mathbf{u_1} = \nabla\phi_1 = \mathbf{u_2} = \nabla\phi_2\right)\big|_{\mathbf{x}=0}. \tag{4.12}
$$

If we evaluate this condition, we get two equations for the amplitudes A_1, A_2 and A_3

$$
\frac{\partial\phi_1}{\partial x_1} = \frac{\partial\phi_2}{\partial x_1}\bigg|_{x_1=0} \Rightarrow A_1 - A_2 = A_3, \tag{4.13}
$$

$$
\frac{\partial\phi_1}{\partial x_3} = \frac{\partial\phi_2}{\partial x_3}\bigg|_{x_3=0} \Rightarrow \eta_1\left(A_1 - A_2\right) = \eta_2 A_3. \tag{4.14}
$$

But this would required that $\eta_1 = \eta_2$, unless all amplitudes are zero, which, as we can see from Equation 4.11, is impossible for a velocity contrast between the layers. If we include the S-wave potentials in the calculations, we can find a solution for this condition, and including continuity of stress we can determine the reflection and refraction coefficients. A schematic version of the reflections and refractions for an incident P-wave is shown in Figure 4.1.
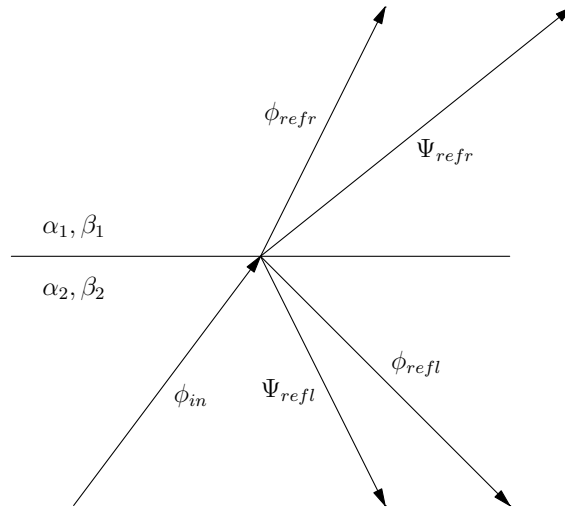
Figure 4.1: Reflections and refractions for an incidental P-Wave on a 1D velocity and density contrast.

The angles of reflection and refraction can be calculated from the ray parameter by *Snell's Law*

$$\frac{\sin i}{v} = p. \tag{4.15}$$

Here i is the angle between the ray and the normal to the interface and v is the velocity for the wave type in the respective layer. The transmission coefficients can be calculated from the boundary conditions and depend on the material properties, the angle of incidence and the type of incoming wave. Equations for these coefficients can be found in Lay and Wallace [1995], for example.

In a homogeneous halfspace without attenuation and a delta function source time function, we would expect to record only two delta function shaped spikes for the P-wave and S-wave arrival. The section between the two arrivals would be essentially flat, as shown in the left panel of Figure 4.2. What we observe instead, typically looks like the right panel in Figure 4.2. After the initial P-wave arrival, we observe a time segment with amplitudes only slightly lower than the initial amplitude before the high amplitude S-arrival. These are scattered and converted phases caused by changes of seismic properties within the Earth. We can even observe distinct arrivals of energy like, for example, just after the line labelled "pP" in the real seismogram in Figure 4.2.

Each arrival in a seismogram can be systematically labelled by the conversions and reflections it has undergone along the path between the source and the receiver. "pP", in the example above, denotes a phase that was reflected from the surface above the source and then continued as a P-wave for the rest of the path.

69
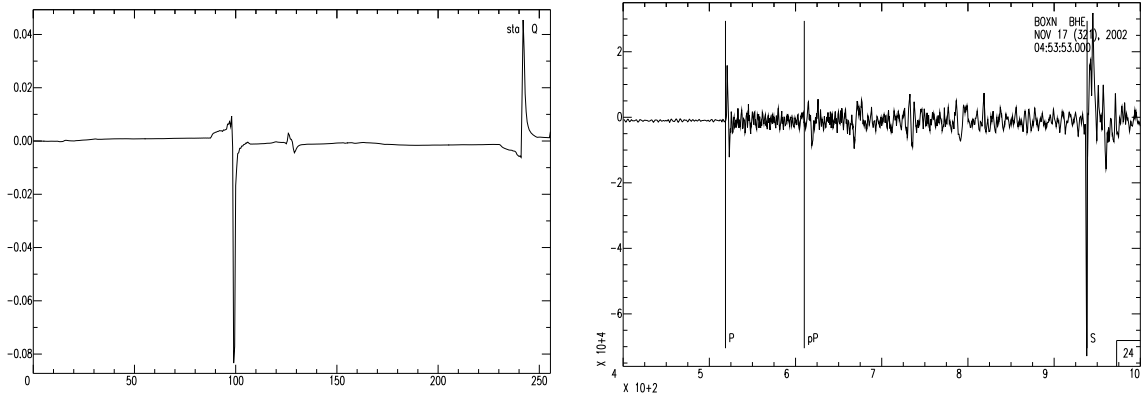
Figure 4.2: Comparison between a synthetic seismogram for a homogeneous half-space (left) and a recorded seismogram (right) for the P-wave and S-wave arrivals. Note the much higher amplitudes between the two main arrivals. The nearly flat line before the P-wave arrival demonstrates that this is not random noise, but scattered energy from structures inside the Earth.

This nomenclature for teleseismic waves is somewhat different from the nomenclature used to describe near-source reflection and refractions in receiver function studies. An overview of the labelling for local phases gives Table 4.1.

Position	Label	Explanation
First	P/S	initial wave at the source
any	p/s	up-going wave after refraction
	P/S	downgoing wave after reflection

Table 4.1: Labelling of seismic phases for local reflections and refractions.

All these reflections and refractions add up, and form the complicated waveform we observe in real seismograms. The total recorded time-series can be represented by a convolution of three different signals and a noise term $n(t)$, viz.,

$$u(t) = s(t) * g(t) * i(t) + n(t). \tag{4.16}$$

$s(t)$ is called the *source-time function* and describes the movement of the rupture at the earthquake source with time, $i(t)$ is a description of the response of the instrument to an incoming wave, usually this description is given as a frequency response and, as mentioned in the previous chapter, we can represent the convolution equivalently in the frequency domain, and $g(t)$ is the *Green's function* of the Earth and solely depends on its structure. This is what we try to calculate and

70
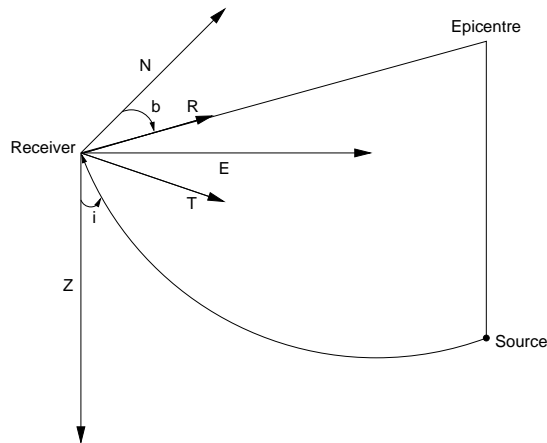
Figure 4.3: Schematic overview of a seismic wave coming into a seismic station and labelling of the corresponding coordinate system.For simplicity we do not show the curvature of the Earth.

model when we perform receiver function analyses. Before we can describe the steps necessary to isolate g we have to discuss the typical setup of a seismograph.

4.2 Recording seismic waves

The typical situation for teleseismic stations is shown in Figure 4.3. For clarity we do not display the curvature of the Earth in the illustration. Seismic energy travels from the source to the receiver along a curved path. This is due to the general increase of seismic velocity with depth. The point above the source location on the Earth's surface is called the *epicentre*. Usually seismic stations record three component of displacement, velocity or acceleration oriented in north-south, east-west and vertical direction. The angle b between the north-south component and the epicenter is called the *backazimuth* and the angle i between the vertical component and the incoming ray is the *incidence angle*. If we rotate the recorded components around the vertical axis, so that one component points towards the epicenter, parallel to the incoming wave, and one is orientated perpendicular to it, we can simplify the analysis of the incoming seismic energy. These two rotated components are called the *radial* and *transverse* components, respectively.

For one-dimensional plane layered structures, there will not be any P-wave energy on the transverse component, only shear wave energy polarized in the plane parallel to horizontal, the so-called SV-wave, will be recorded. The P-wave and SH-wave energy on the other hand will be completely recorded by the radial and vertical component. In practice, real seismograms often show some energy on the

3-component data from site BOXN

Rotated to radial
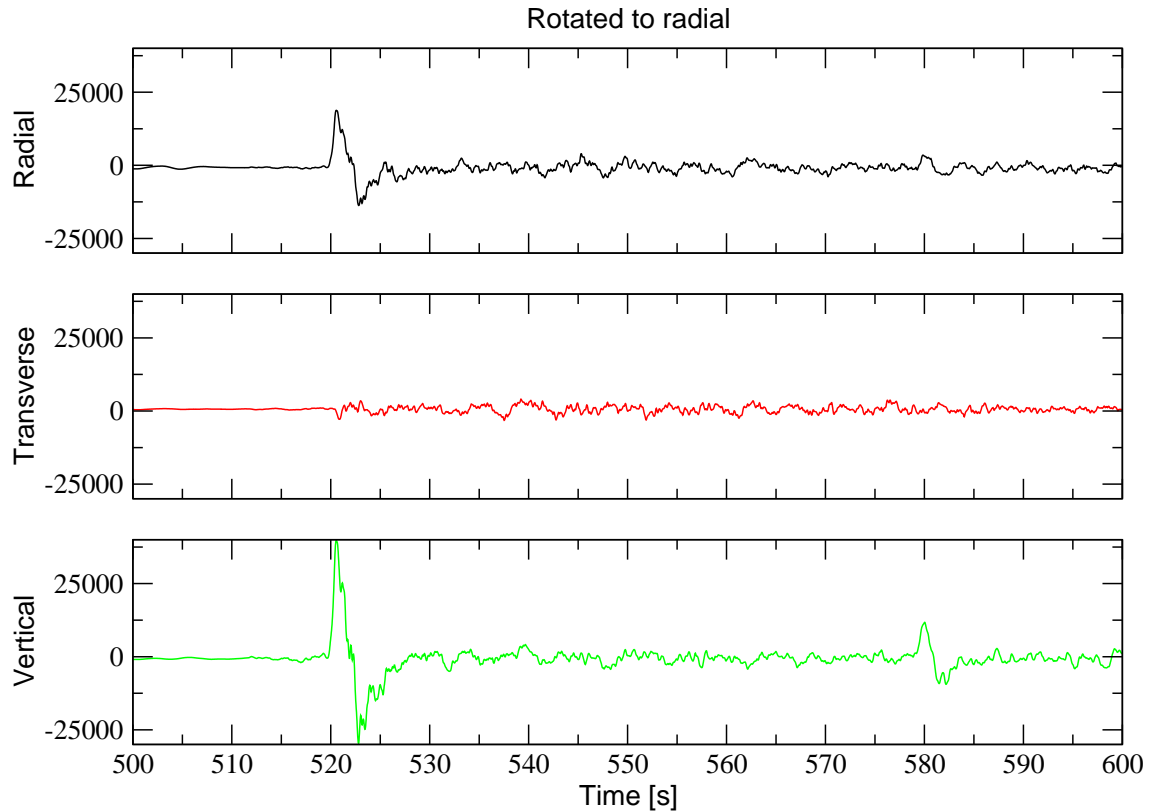
Figure 4.4: Example of seismic data after the initial P-Wave arrival after rotation in the ray coordinate system. The main P-Wave energy is concentrated on the vertical and radial components, but we can observe scattered energy on all three components.

transverse component during arrival of the P-wave. This can be due to scattering from 3D structures or anisotropy and we will discuss this issue again later when we look at receiver functions from the Slave craton.

Figure 4.4 illustrates what we just discussed. We plot the recorded data just before and after the P-wave arrival of a large teleseismic event with a moment magnitude $M_w = 7.3$. We can see a clear high amplitude arrival on the radial and transverse components about 520 s after the event. Because of the steep incidence angle of the P-wave, we observe the highest amplitude on the vertical component. In contrast the radial component only exhibits some low amplitude fluctuations. After the initial P-wave, the amplitude of the radial and vertical components drops quickly and reaches the level of the transverse component. These fluctuations are caused by reflections and refractions below the receiver and will be used for receiver function analysis. A sudden increase in amplitude on the vertical component at

580 s marks the arrival of another seismic phase, in this case PcP, a P-wave that has been reflected from the outer core.

4.3 Calculating receiver functions

We can now discuss how to extract information about Earth structure from the recorded P-wave and its conversions. As we saw in equation 4.16 the total measured seismic trace can be represented as a convolution of three contributors. The instrument response i(t) is known for every component and can be easily corrected for. The problem remains to separate the source time function contribution and distinguish effects from different structures along the wave path. This can be achieved by deconvolving the vertical component from the radial component [Vinnik, 1977, Berteussen, 1977, Langston, 1979, Kind and Vinnik, 1988], i.e., in the frequency domain

$$\mathrm{Rf}(\omega) = \frac{\mathrm{R}(\omega)\mathrm{Z}^{\star}(\omega)}{\mathrm{Z}(\omega)\mathrm{Z}^{\star}(\omega)}. \tag{4.17}$$

This operation removes the effect of the source time function, which is equal for both components. Transformation back into the time domain yields the actual receiver function. The representation in the time domain has the advantage that the features of the receiver function have a straightforward and direct connection to Earth structure. The time axis represent distance from the receiver and, because of the steep incidence of teleseismic waves, is a good proxy for depth. A significant non-zero amplitude marks the arrival of a P-to-S converted wave. As described above, the conversions take place at velocity and density contrasts and thus the signal of the receiver function can be interpreted in terms of velocity and density changes below the station. A positive amplitude marks a transition from higher to lower velocity, with respect to the direction of wave-propagation, while for a negative amplitude the situation is reversed. This direct interpretation of the receiver function is complicated though by the appearance of multiple reflections in the later sections of the receiver function. A synthetic receiver function and a part of the ray geometry are shown in Figure 4.5.

For the RTZ-coordinate system the highest amplitude of the receiver function is observed at time 0, corresponding to 0 lag between the two components, and is related to the initial P-wave arrival. We can identify a clear arrival at 4.5 s that is related to the P-to-S converted wave from the layer interface. The first multiple reflection from the surface has a comparable amplitude and identical polarity, while the polarity of the second multiple is reversed and its amplitude is slightly smaller. For a simple two-layer model the arrival times of the refracted wave and
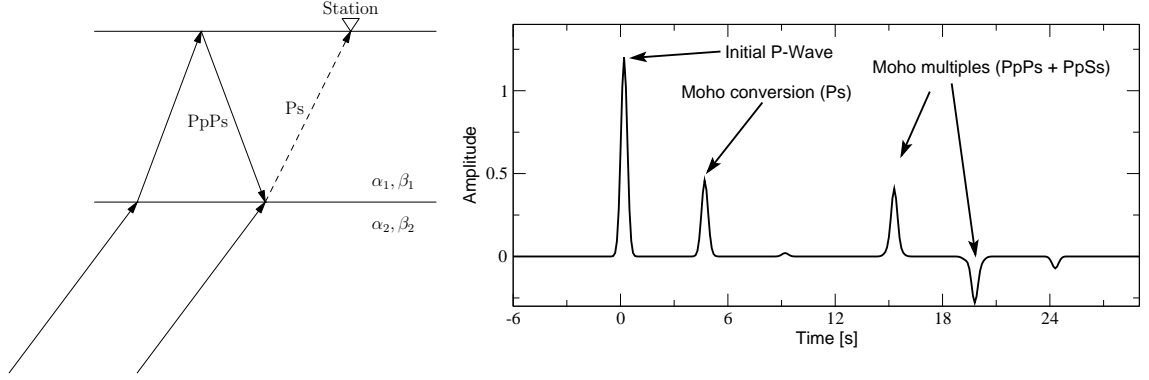
Figure 4.5: Synthetic receiver function (right) calculated for a two layer model (left). The main conversions are labelled in the receiver function plot, and two exemplary ray paths are shown on the left hand side.

the multiples can be easily calculated, viz.,

$$t_{Ps} = h \left(\sqrt{v_s^{-2} - p^2} - \sqrt{v_p^{-2} - p^2} \right), \tag{4.18}$$

$$t_{PpPs} = h \left(\sqrt{v_s^{-2} - p^2} + \sqrt{v_p^{-2} - p^2} \right), \tag{4.19}$$

$$t_{PpSs} = 2h\sqrt{v_s^{-2} - p^2}. \tag{4.20}$$

The ray parameter p depends on the distance from the earthquake and can be taken from an earthquake catalogue. Hence with these three equations we can uniquely determine the three parameters v_s, v_p and h.

Practical procedure

In practice Equation 4.17 does not work well to calculate receiver function directly from recorded seismograms. We have to perform some additional processing steps and modify the idealized equation to obtain usable results. The first step, is to apply a bandpass filter to the seismogram. The high frequency energy above 1 Hz is often related to scattering from small inhomogeneities, and does not fit into the theoretical formulation of a refracted wave that we assume for receiver functions. Below 0.05 Hz ambient noise from wind, tides and cultural sources can overwhelm the body wave signal which has dominant periods between 1-0.1 Hz. Therefore a simple band-pass filter can improve the quality of the receiver functions.

This step alone though does not alleviate another more fundamental problem with Equation 4.17. The division by the auto-spectrum of the vertical component becomes numerically unstable, if the spectrum has a high dynamic range. Real

seismograms often show pronounced minima in their power spectra, and the receiver function calculated from such a seismogram would be dominated by these spectral holes, which do not carry much information about earth structure. In magnetotelluric processing we solve the problem of unstable division by stacking several spectra and this can be done for receiver functions as well [Gurrola et al., 1995]. This approach however can be problematic, because as we see from Equation 4.18, the time of the conversions in the receiver function depends on the ray parameter, which varies from earthquake to earthquake. Careful selection of earthquakes with similar ray parameter or moveout correction (see below) can solve this problem, but in a lot of cases different solutions are preferred.

The easiest way to circumvent division by small numerical values, is to add a small but significant number to the denominator of equation 4.17 [Langston, 1979]. In addition a Gaussian filter is applied to the receiver function

$$\mathrm{Rf}(\omega) = \frac{R(\omega)Z^{\star}(\omega)}{\max\left(Z(\omega)Z^{\star}(\omega), l\right)} \exp\left(-\omega^2/4\sigma^2\right). \tag{4.21}$$

Usually the *water-level* parameter l is chosen with respect to the maximum amplitude of the vertical component

$$l = w \max\left(Z(\omega)Z^{\star}(\omega)\right), \tag{4.22}$$

and typical values for w range between 0.01 and 10^{-5}. The optimum value depends on the signal to noise ratio, but the final result is not overly sensitive to the chosen water-level. The advantage of this method is that it is easy to implement, and fast to calculate. One problem is though that because of the non-localized nature of the Fourier transform and the division operation, noise from all parts of the selected time-window influences the receiver function. This can often be observed for events with magnitudes less than 6. Despite an apparently high-quality seismogram, the resulting receiver function is often strongly distorted.

Some of the noise problems of the water-level deconvolution can be circumvented by using an iterative time-domain deconvolution method [Ligorria and Ammon, 1999]. At each iteration we calculate the cross-correlation between the vertical and radial component. We determine the receiver function value at the time corresponding to the maximum absolute correlation value by dividing the cross-correlation by the zero lag autocorrelation of the vertical component. After subtracting the predicted wave-form from the radial component the next iteration starts. The procedure stops when either the improvement between iterations becomes too small, or a chosen percentage of the radial component can be predicted from convolving the receiver function with the vertical component. The complete procedure is summarized in Algorithm 4.

Algorithm 4 Iterative deconvolution algorithm

1: Calculate zero lag auto-correlation of vertical a_Z
2: $R_i(t) = R(t)$
3: **repeat**
4:    Calculate cross-correlation $c(\tau)$ of current radial $R_i(t)$ and vertical $Z(t)$
5:    Find maximum of $c(\tau)$ and corresponding lag τ_{max}
6:    $Rf(\tau_{max}) = c(\tau_{max})/a_Z$
7:    $R_{i+1}(t) = R(t) - Rf(t) * Z(t)$
8: **until** $|R_{i+1} - R_i| < \delta_{min}$ or $|R_{i+1} - R| < e_{min}$ or $i = i_{max}$

Compared to the water-level deconvolution we often get better results with respect to the following criteria to assess the quality of the receiver function:

1. The zero lag amplitude should be the highest throughout the receiver function.

2. Even though the Moho depth is often not know a priori, a high amplitude conversion from the Moho can be expected at times between 3 and 5 seconds for Moho depths of $25 - 40$ km.

3. The Moho multiples should be visible at times given by equations 4.18 – 4.20.

4. In general the amplitude of conversions can be expected to decrease with time, strong oscillations throughout the receiver function indicate problems with noise.

5. The base-line should be flat, long-period background variations indicate noise contributions that often can be removed by high-pass filtering.

We will discuss these issues further when we look at receiver functions from the Slave craton in Chapter 7.

Figure 4.6 shows a comparison between a receiver function calculated with the waterlevel technique (top) and the iterative deconvolution technique (bottom) for the same event, station and filter parameters. The receiver function calculated with the waterlevel technique shows many high amplitude variations for the first 30 seconds. The Moho conversion at a lag time of about 5 seconds shows the highest amplitude after the zero lag pulse, but a high number of segments with similar amplitude make it difficult to identify the conversion reliably. This applies even more to the multiples and all other conversions. The iterative deconvolution technique, in contrast, yields a receiver function that is much easier to interpret.

Comparison of receiver function techniques

Event 2002/321
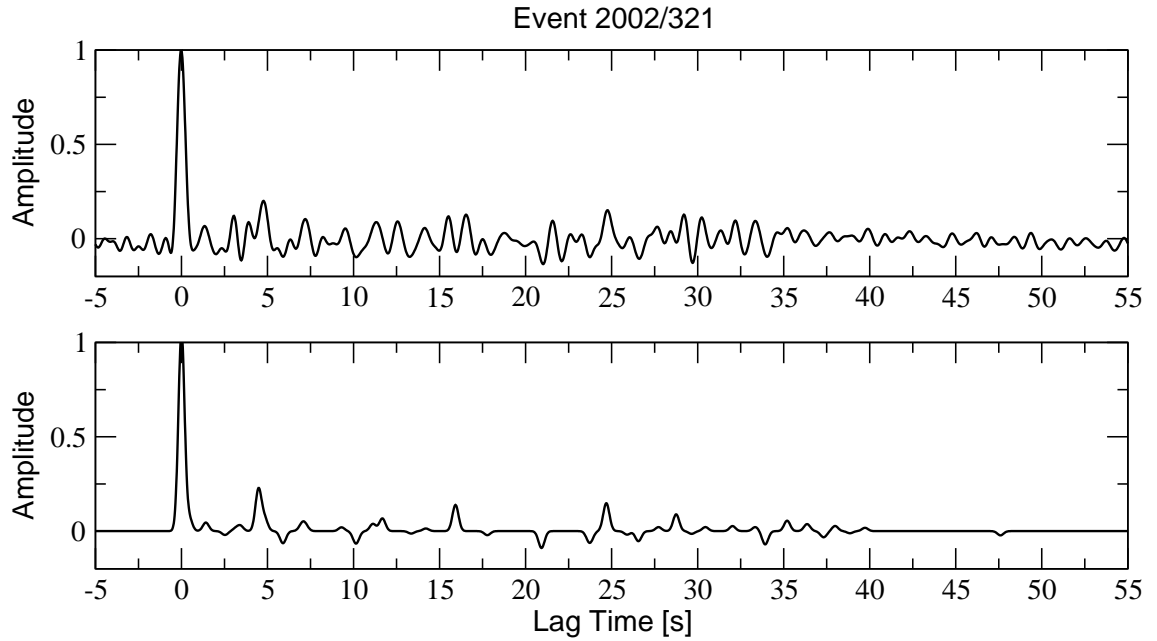
Figure 4.6: Comparison between receiver functions calculated with the waterlevel technique (top) and the iterative deconvolution technique (bottom).

The Moho conversions and its multiples are easy to identify. A number of smaller amplitude maxima and minima exist and a good correspondence with features of the waterlevel receiver function suggests that these are actual conversions within the Earth and not noise. Remarkably, convolving the vertical component with this receiver function retrieves 98.5% of the radial component's energy, illustrating the large influence of a reasonably low amount of noise on the waterlevel result.

4.4 Stacking receiver functions

Even after filtering and with the iterative deconvolution method, receiver functions can still be affected by noise which, if not identified correctly, can cause spurious structures in the subsequent inversion. The influence of random noise can be further reduced by stacking a number of high-quality receiver functions. A few things have to be considered before we can apply the stacking procedure. Equation 4.18 shows that the delay time of any conversion depends on the ray parameter, which in turn depends on the distance between the seismic station and the earthquake. If we stack receiver functions from different distances, we will not only suppress noise, but also smear out the signal we are interested in. Figure 4.7

Receiver function moveout
Distance Range 30 - 90
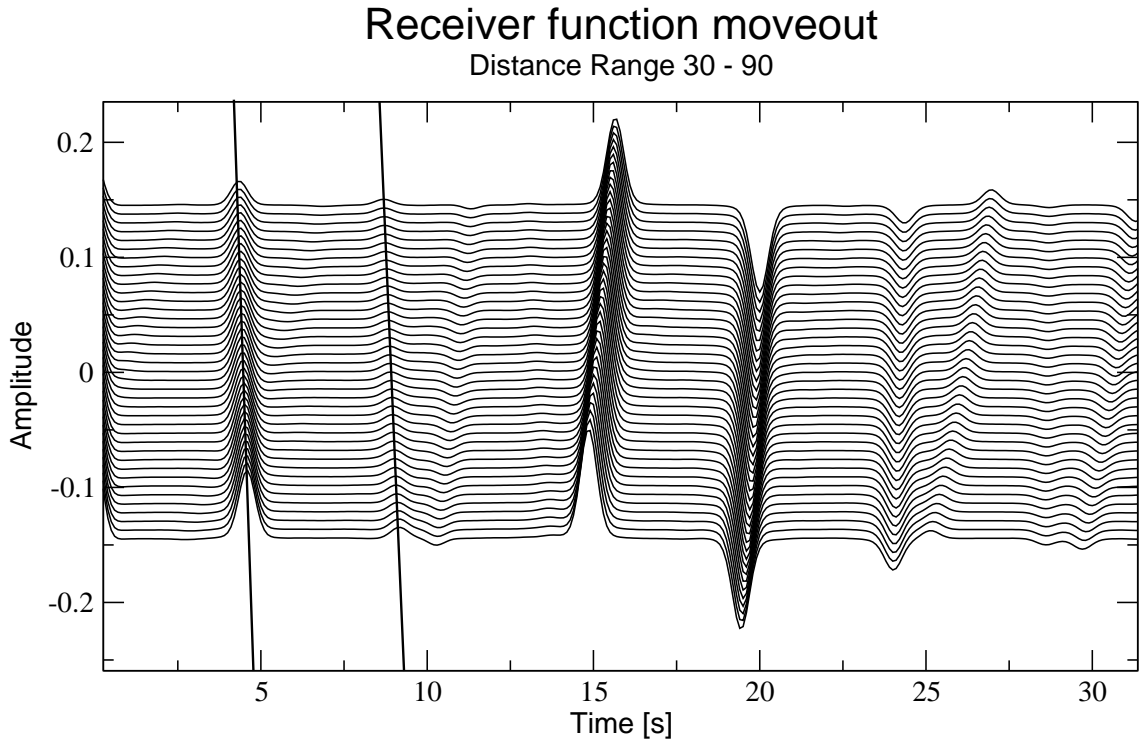
Figure 4.7: Synthetic receiver functions calculated for a number of distances be-
tween 30°and 90° (bottom to top). The two lines mark the maximum amplitude
of two primary conversions. We can identify multiples by their inverted moveout
relationship.

shows synthetic receiver functions calculated for a number of distances.

We can see that for the primary conversions the delay time decreases with
increasing distance, while for the multiples the delay time increases. For high
quality data this criterion can be used to distinguish the two, but in practice noise
can make it difficult to identify the weaker amplitudes reliably. For the Moho
primary conversion, the shift in time is less than 0.5 seconds across the whole
distance range, and therefore not significant for data that have been filtered to a
maximum frequency of 1 Hz. Already for the second conversion, marked with a
line in Figure 4.7, the moveout is significant and we can also see that the change
in time is not linear.

There are two ways to prevent problems with stacking events with different
moveout characteristics. The easiest possibility is to restrict the stacking to events
coming from the same distance range. A difference in distance of 10° is not critical
for delay times less than 25 seconds, which corresponds to the lithospheric depths
we are interested in. For deeper conversions, and if the number of earthquakes
from any distance range is not sufficient, an alternative approach is to correct for

the moveout. We can use equation 4.18 to convert the time of each sample to the corresponding time at a given reference distance. This, of course, requires assumptions about the seismic velocities, and usually a standard reference model such as PREM [Dziewonski and Anderson, 1981], IASP91 [Kennett and Engdahl, 1991] or AK135 [Kennett et al., 1995] is used for the correction. Another effect of the moveout correction is that all multiples are suppressed. Due to their inverted moveout characteristics with respect to the primary conversions, they are spread out even more after the correction and their amplitude is reduced in the following stacking process. This is an advantage when receiver functions are used as an imaging technique, and can bring out smaller conversions that are hidden by the multiples. When modelling receiver functions the advantage is not that clear though, as the multiples also contain information about Earth structure.

Another complication for stacking receiver functions can arise when the structures are not purely one-dimensional. As mentioned above, we will then observe energy on the receiver function for the transverse component, but also the radial receiver function then depends on the back-azimuth, the direction of the incoming wave. The only solution then is to restrict the stacking to a range of backazimuths where no variations are observed.

4.5 Modelling receiver function

In contrast to the 1D magnetotelluric case, there is no direct way to calculate the receiver function from a model. Instead, we calculate synthetic radial and vertical seismograms, and obtain the receiver functions with the same procedure as for observed data. A number of different methods exist to calculate synthetic seismograms. Ray-theoretical methods [Helmberger, 1974] are computationally fast, but require to specify the various rays that we want to model. This can be problematic, because the importance of the various conversions is not obvious a priori and the inclusion of a large number of reflected and refracted rays increases computation time and hence removes the speed advantage. Spectral methods [Fuchs and Muller, 1971] are exact in the 1D case, but are computationally expensive. Their main advantage is for modelling regional and local wave propagation, as they can include near-source effects.

For teleseismic events, matrix propagation methods [Haskell, 1962] provide a good compromise between precision and speed. Assuming a plane wave propagating into the region of interest, the effect of reflections and refractions at each interface can be calculated by a number of matrix operations. We use the program respktn described in Randall [1989].

*The ruler of the Universe thought for a long while whilst Zarni-
woop quivered with anger.*
*'You're very sure of your facts,' he said at last, 'I couldn't trust
the thinking of a man who takes the Universe – if there is one –
for granted.'*
Zarniwoop still quivered, but was silent.
'I only decide about my Universe,' continued the man quietly.
'My Universe is my eyes and my ears. Anything else is hearsay.'

Douglas Adams, The Restaurant at the End of the Universe

5

Optimization in the context of filtering and inversion

The mathematical fields of optimization and inverse theory, play an important role in geophysics. For all but the simplemost problems, the relationship between the observations and the properties of the Earth one is looking for is non-linear. In addition we usually only have observations at the Earth's surface and at a few locations, and there is no analytical formula to extract the information from the data. Usually we are faced with a situation, were we can only solve the forward problem

$$\mathbf{d}^{\mathrm{pred}} = \mathbf{f}(\mathbf{m}), \tag{5.1}$$

i.e., we can predict the expected data $\mathbf{d}$ for a given model $\mathbf{m}$ of the Earth. The entries of the vector $\mathbf{m}$ describe the distribution of one or more physical properties, in our case electrical resistivity or seismic velocities, and the function $\mathbf{f}$ captures the physics of the problem. What we need is the solution of the inverse problem, which formally can be written as

$$\mathbf{m} = \mathbf{f}^{-1}(\mathbf{d}^{\mathrm{obs}}), \tag{5.2}$$

but, as mentioned, the inverse of $\mathbf{f}$ is usually not accessible. What we have to do instead is to minimize the difference between the predicted and the observed data with respect to a certain norm,

$$\|\mathbf{d}^{\mathrm{obs}} - \mathbf{d}^{\mathrm{pred}}\| \rightarrow \min, \tag{5.3}$$

and for this we have to use optimization methods.

Minimization or maximization problems are encountered in various fields, and thus a large number of different methods exist, each geared towards special applications, and with certain advantages and disadvantages. We will discuss those methods that are relevant for geophysical applications, and in particular for joint inversion. We can separate optimization methods in two broad classes:

1. Direct search methods use a starting point in model space, i.e., a model of the Earth that the user regards as representative before the inversion, and look for directions that reduce the norm in equation 5.3. For non-linear problems this is an iterative process, and the minimization direction changes at each step. Often gradient and curvature information is exploited to improve the speed of convergence to the minimum. The main advantage of these methods is their rapid convergence, which makes them suitable for large scale problems. The major drawback is the risk of being trapped in a local minimum.

2. Stochastic search methods randomly sample model space, and preserve models that perform better with respect to the minimization criterion. As one would expect, the random nature of the process makes it necessary to consider a large number of models, which is the main disadvantage of these methods. Several variants try to increase the chance of finding suitable models, by introducing systematic elements into the random search, but still they are usually orders of magnitude slower than direct search methods. The great advantage of random methods, that justifies their usage, is the reduced influence of the starting point and the ability to escape local minima.

In the following sections we will first discuss some popular linearized methods and also describe various aspects of inverse theory that are usually applied in the context of these methods. We will then turn to genetic algorithms, a stochastic optimization method based on biological evolution, and describe their suitability for our joint inversion problem.

5.1 Linearized methods

Although we use a genetic algorithm for our optimization problem, it is instructive to discuss direct search and, particularly, linearized methods. A lot of the tools that we will use later to quantify the quality of our models were developed in the context of these methods, so we need to understand their mechanics. Also this discussion will enable us to understand some of the advantages and disadvantages of genetic algorithms.

Most linearized methods minimize equation 5.3 in the least-squares sense, i.e., with respect to the L_2 norm

$$E(\mathbf{m}) = \|\mathbf{d}^{obs} - \mathbf{d}^{pred}\|_2 = \sum_i \left(d_i^{obs} - d_i^{pred} \right)^2 \to \min. \tag{5.4}$$

This choice of norm has the advantage that the prediction error e is continuously differentiable, a property that is used by all linearized algorithms. Also the least-squares criterion provides an optimal solution in the presence of Gaussian errors. It is however possible to use other norms than the l_2 norm through an iterative reweighting scheme in linearized inversions [Farquharson and Oldenburg, 1998]. For genetic algorithm inversions we are not bound by the requirement of differentiability and we can choose arbitrary l_p norms a well as other measures of misfit such as correlation. These will, of course, influence the type of models we obtain.

For observed data that is affected by measurement errors we make a further modification to equation 5.4 to weight the contribution of each datum with respect to its error σ_i and a specified relative error floor ϵ

$$E(\mathbf{m}) = \sum_i \left(\frac{d_i^{obs} - d_i^{pred}}{\max\left(\sigma_i, \epsilon d_i^{obs}\right)} \right)^2 = \mathbf{e}^{\mathsf{T}}\mathbf{e}. \tag{5.5}$$

The division by the maximum of the data error and the absolute error floor for each datum avoids problems with unrealistically small error estimates for the data and at the same time balances numerical differences. Data points with high values or unrealistically small errors would otherwise dominate the misfit function, and all other data would be mostly ignored.

The gradient of E with respect to the model-parameters $\mathbf{m}$ can be calculated using the chain rule

$$\gamma = \nabla E(\mathbf{m}) = \mathbf{G}^{\mathsf{T}} \cdot \frac{d_i^{obs} - d_i^{pred}}{\max\left(\sigma_i, \epsilon d_i^{obs}\right)} = \mathbf{G}^{\mathsf{T}}\mathbf{e}, \tag{5.6}$$

where we assumed a linearized relationship between model and data

$$\mathbf{d} = \mathbf{G}\mathbf{m}. \tag{5.7}$$

The $n \times m$ matrix

$$\mathbf{G} = \frac{\partial \mathbf{f}}{\partial \mathbf{m}} \tag{5.8}$$

is called the *sensitivity matrix*, where n is the number of datapoints, and m is the number of model parameters. For a lot of problems the sensitivity matrix cannot be calculated analytically, but has to be approximated by a finite difference method

$$\mathbf{G}_{ij} \approx \frac{\mathbf{f}_j(\mathbf{m}_i + \Delta_i) - \mathbf{f}_j(\mathbf{m}_i)}{\Delta_i}. \tag{5.9}$$

or automatic differentiation [Sambridge et al., 2007].

As we can see, finite difference estimation requires $m + 1$ evaluations of the forward modelling function $\mathbf{f}$, which is the limiting factor in terms of computational time for large-scale problems. For certain types of problems, the calculation of the gradient can be achieved computationally faster using adjoint methods [Avdeeva and Avdeev, 2006, Plessix, 2006] which we will not discuss here.

The gradient points in the direction of steepest ascent, so we can construct a first simple optimization algorithm by performing a step in the opposite direction

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mu \nabla e. \tag{5.10}$$

This equation is exactly the same as Equation 3.25 for the LMS-adaptive filter, but now viewed in the more general context of minimization of an non-linear function, instead of adaptive signal processing.

The steepest descent algorithm is very simple, but not ideal with respect to convergence to the minimum, as we already saw in Chapter 3. For many problems the direction of steepest descent is not the shortest way towards the minimum. In situations where one forward model evaluation is computationally expensive, the simplicity of the optimization algorithm is outweighed by the fact that many iterations are needed to find the minimum.

The most commonly used method to improve the speed of convergence of linearized optimization algorithms is to incorporate curvature information by either calculating or approximating the *Hessian matrix* of second derivatives,

$$\mathbf{H} = \frac{\partial \mathbf{f}}{\partial m_i \partial m_j}. \tag{5.11}$$

For geophysical problems the Hessian matrix is rarely available analytically and computation by finite differences similar to the gradient is computationally prohibitive, even for rather simple problems such as one-dimensional magnetotelluric inversion. For large scale problems with a lot of model parameters, even the storage of the $m \times m$ matrix $\mathbf{H}$ can be problematic. For this reason, a number of different algorithms exist that approximate the Hessian matrix either directly or by its action on a vector [e.g. Nocedal, 1992, Tarantola, 2004]. We will discuss here the *variable metric* method, which uses information from gradients from previous iterations to build up an approximation of the Hessian matrix, as an example for a widely used state of the art method.

At the minimum $\mathbf{m}^\star$ the gradient vanishes

$$\mathbf{f}'(\mathbf{m}^\star) = \mathbf{0}. \tag{5.12}$$

From a Taylor expansion of $\mathbf{f}'$ around the current model $\mathbf{m}$ we get

$$\mathbf{f}'(\mathbf{m} + \mathbf{h}) = \mathbf{f}'(\mathbf{m}) + \mathbf{f}''(\mathbf{m})\mathbf{h} + \mathcal{O}(\mathbf{h}^2). \tag{5.13}$$

Assuming $|\mathbf{h}|$ is small we can discard second order terms, and we can find the next step $\mathbf{h}$ by solving

$$\mathbf{f}''(\mathbf{m})\mathbf{h} = \mathbf{H}\mathbf{h} = -\mathbf{f}'(\mathbf{m}). \tag{5.14}$$

Analytically the solution is, of course,

$$\mathbf{h} = -\mathbf{H}^{-1}\mathbf{f}'(\mathbf{m}) = -\mathbf{F}\mathbf{f}'(\mathbf{m}). \tag{5.15}$$

For this reason, some of the practical numerical algorithms do not approximate the Hessian, but its inverse. A detailed derivation of the approximations is beyond our scope here, and not of particular interest for the discussion, one important aspect of it is though, that for $\mathbf{h}$ being a descent direction a necessary requirement is that $\mathbf{H}$ is positive definite, i.e., for all vectors $\mathbf{u}$

$$\mathbf{u}^{\mathsf{T}}\mathbf{H}\mathbf{u} > 0. \tag{5.16}$$

The computational steps are summarized in Algorithm 5.

Algorithm 5 The variable metric optimization algorithm

1: Initialize $\mathbf{F}_0 = \mathbf{I}$
2: Choose starting model $\mathbf{m}_0$
3: **while** not converged **do**
4:     $\gamma_i = \mathbf{G}^{\mathsf{T}}\mathbf{e}$
5:     $\Phi_i = \mathbf{F}_i\gamma_i$
6:     $\mathbf{m}_{i+1} = \mathbf{m}_i + \mu\Phi_i$
7:     $\mathbf{F}_{i+1} = \mathbf{F}_i + \delta\mathbf{F}_i$
8: **end while**

Three important steps have been omitted in Algorithm 5, the adjustment of the stepsize parameter μ, the update $\delta\mathbf{F}_i$ of the inverse of the hessian matrix, and the definition of the convergence criterion.

In contrast to the algorithms we discussed in the context of filtering, the stepsize μ is usually not fixed, but adjusted through a *line-search* algorithm. The reason for this difference lies in the slightly different computational complexity criteria. For an optimization algorithm used for inversion of data, we want to achieve the minimization with the least amount of forward modellings. Even though the line-search algorithm also requires a few forward modelling calculations, the overall number within the whole optimization process is still lower than for a fixed stepsize. For filtering applications, the number of forward calculations is the number of samples in the time-series and consequently fixed. Adding a line search would increase this number even further and thus the stepsize is either kept fixed or estimated by a simple formula (see equation 3.32).

For the update of $\mathbf{F}_i$ a number of different possibilities exist, the most widely used is the BFGS formula [Fletcher, 1980]

$$\mathbf{F}_{i+1} = \left(\mathbf{I} - \frac{\delta\mathbf{m}\,\delta\gamma^\mathsf{T}}{\delta\gamma^\mathsf{T}\,\delta\mathbf{m}}\right)\mathbf{F}_i\left(\mathbf{I} - \frac{\delta\mathbf{m}\,\delta\gamma^\mathsf{T}}{\delta\gamma^\mathsf{T}\,\delta\mathbf{m}}\right) + \frac{\delta\mathbf{m}\,\delta\mathbf{m}^\mathsf{T}}{\delta\gamma^\mathsf{T}\,\delta\mathbf{m}}, \tag{5.17}$$

which preserves the positive definiteness of $\mathbf{F}$ in most situations.

The convergence criterion for all gradient-based optimization techniques is given by comparing the length of the gradient with the machine precision ϵ

$$\|\gamma\| \leq \epsilon. \tag{5.18}$$

This termination criterion, and the way the minimum is approached, has important implications for the performance of the algorithm compared with global optimization methods. Any point on the error surface, where the gradient becomes small enough, will be regarded as a valid optimum point, even though it might only be a local minimum, and not the global minimum we are seeking. As gradient based optimization algorithms only use local information from the current and, in some cases previous, iterations, the algorithm has no way to identify a minimum as local and will terminate as soon as the criterion in equation 5.18 is satisfied.

5.2 Genetic algorithms

Genetic algorithms (GA's) are one class of global search algorithms that avoid the localized search strategy of linearized methods. Using strategies mimicking biological evolution they provide a good compromise between computationally more expensive Monte-Carlo methods and linearized methods and offer particular advantages for multi-objective inversion. As the optimization method used in our joint inversion approach is based on a modern genetic algorithm, we will discuss them in some detail in the following sections. A good overview over the historical development and the basic properties of genetic algorithms can be found in Goldberg [1989]. Since the introduction of GA's in the field of optimization by Holland [1975], a large variety of genetic algorithms have been developed, and, because of the non-linear nature of geophysical inverse problems, applied to various types of datasets [e.g., Everett and Schultz, 1993, Jervis et al., 1993, Pérez-Flores and Schultz, 2002]. We will focus on the core properties common to all GA's and more specifically on the modern NSGA-II [Deb et al., 2002] which forms the core of our inversion algorithm. The general program flow for a genetic algorithm is shown in Figure 5.1. Each individual step can be performed in a variety of ways, depending on the problem at hand and the type of solution we are looking for.

The first design choice that influences most of the other steps of the genetic algorithm is the way in which the model parameters are represented within the GA. The earliest GA's explicitly encoded the parameters as a concatenated binary string, but in the meantime real valued genetic algorithms have gained popularity [Goldberg, 1989]. The binary representation is motivated by the analogy between DNA in an evolutionary process and the models in genetic algorithm optimization and makes the two modification operators crossover and mutation independent from model parameter storage.
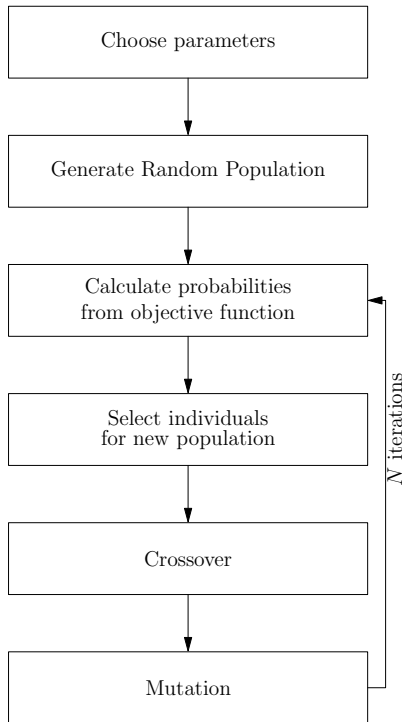
Figure 5.1: Program flow of a genetic algorithm

Given the model vector $\mathbf{m}$, as described in the previous section, we create the genetic string by translating each individual component m_i into a binary form and concatenating them into a single string. For this the user has to chose three transcription parameters for each model parameter m_i: The minimum possible parameter value m_i^{min}, the discretization step δ_i and the number of bits used to encode the parameter value n_i. The relationship between the actual model parameters for the inversion problem and the representation within the genetic algorithm is then given by

$$m_i = m_i^{min} + \delta_i \sum_{j=1}^{n_i} 2^{j-1} s_j, \qquad (5.19)$$

where s_j represents the part of the genetic string that corresponds to that model parameter. A simple example can illustrate the relationship between the model parameters and the genetic string.

Given a problem that depends on two parameters m_1 and m_2 where m_1 is expected to be in the range $-1 \ldots 1$ and m_2 between 0 and 3.1, we can adjust the encoding length of each parameter, so that the variations in the minimization function $f(m_1, m_2)$ are represented without aliasing. The Table 5.1 lists examples for transcription parameter values, corresponding example strings and model parameters values.

We can immediately see two major differences to the linearized algorithms discussed in the previous section. First of all the parameter values are constrained by the minimum value m_i^{min}, and the maximum that results from the other two parameters. The linearized algorithms we discussed do not have such limitations,

i	m_i^{min}	δ_i	n_i
1	-1	2/63	6
2	0	0.1	5

$\longrightarrow$ $\underbrace{101011}_{m_1}\underbrace{01011}_{m_2}$ $\longrightarrow$ $m_1 = 0.6825$ $m_2 = 2.6$

i	m_i^{min}	δ_i	n_i
1	-1	2/7	3
2	0	0.05	6

$\longrightarrow$ $\underbrace{011}_{m_1}\underbrace{101101}_{m_2}$ $\longrightarrow$ $m_1 = 0.7143$ $m_2 = 2.25$

Table 5.1: Example for transcription parameters of a genetic algorithm for a problem with two parameters

but modifications can be made to incorporate constraints into linearized methods if needed [e.g. Byrd et al., 1995]. The implicit constraints on the model parameters by the genetic algorithm are both and advantage and a disadvantage. Most geophysical problems have well known limits on "reasonable" values for physical parameters of the Earth. Therefore the limitedness of the inversion parameters can even be an advantage compared to linearized methods, where efforts have to be made to enforce the constraints. When the possible range of parameter values is large though, the user has to choose between a large number of bits for the respective model parameters and a large discretization step.

This leads directly to the next significant difference between linearized methods and genetic algorithms. The mathematical foundation of linearized methods assumes continuous model parameters, and the methods are continuous in as much as a computer can represent real numbers. For genetic algorithms, the discretization length is typically orders of magnitude above numerical precision, and therefore treat the model space as discrete. Again for geophysical problems this is usually not of concern, as we anyway assume that our objective function varies smoothly with the model parameters and typical variations are well known. Still the choices of the transcription parameters influence the outcome of the inversion and any user of genetic algorithms has to be aware of it.

With the information on the transcription parameters, the algorithm can then generate a number of random model strings, the so-called *population*. Each *member* of the population represents one individual model and the size N_p of the population is usually specified by the user. A larger N_p will result in more random starting models and thus sample a larger portion of model space. Of course, this comes at the cost of more forward modelling calls and a longer run-time of the algorithm.

Once the random starting population has been generated, each population member is transcribed back to the real-valued inversion parameters and its fitness is

calculated. For inversion applications this usually means calculating the misfit between the predicted data from the model parameters and the observed data in a manner similar to equation 5.5. However, because genetic algorithms do not require the misfit function to be differentiable, we can modify the way the misfit is calculated in a number of ways. We will return to this issue when we discuss multi-objective optimization and NSGA-II. The misfit determines the probability that a given member will be selected for the next generation. It is this probabilistic approach, that enables the genetic algorithm to escape local minima. Some of the models in the first generations might not have a good fit compared with the best model in that generation. Still, there is a possibility that these models are kept in further generations and through modification or exchange of information create a better model.

Once the models for the next generation have been selected, we have to make some modifications, otherwise there will be no new information and we could just stop after the first iteration. Two processes provide innovation to the original models. *Crossover* provides a mechanism to exchange information between different members of the population. With a probability p_c two random population members exchange part of their genetic string after a randomly chosen point. The concept behind this is that some of the models might perform well due to a certain segment within the genetic string, while the good performance of another model is caused by another part of the genetic string. Exchanging parts of these strings can then create an even better model, but in any case helps to distribute well performing parts through the population. Typically we choose $0.2 \leq p_c \leq 0.7$, so that a fair proportion of the population performs crossover.

The second mechanism to create new models is *mutation*. With a probability p_m a bit of the genetic string of a given member changes its value. This mechanism is undirected and creates completely new models. Especially in the later stages of the algorithm, when a small number of similar models dominate the population, this mechanism ensures that the algorithm does not converge prematurely. After the selection, crossover and mutation processes, the misfit for the new population is calculated and these steps are repeated for a fixed number of iterations. Due to the stochastic nature of the process and the lack of information about any gradients, there is no simple termination criterion. What we usually observe, is that the misfit of the best model and the average misfit decrease quickly in the first iterations until they reach a stable level. After this initial phase, the best misfit stays constant for a number of generations and then decreases again. With continuing evolution the stable phases become longer and longer, but there is always the possibility of another decrease.

5.3 NSGA-II

We will now turn our attention to the specific implementation of the genetic algorithm we use for our joint inversion approach, NSGA-II [Deb et al., 2002]. In this context we will also discuss some specific problems of multi-objective inversion and contrast the approach taken by NSGA-II with linearized methods and other genetic algorithms.

To represent the models we use the standard binary strings described above and consequently the initialization stage, as well as crossover and mutation work exactly as described for the general case. The selection process, and the way the fitness of each individual member is determined, is specifically geared towards multi-objective optimization. We will therefore look at this problem setting now. In contrast to single-objective optimization that we discussed at the beginning of the chapter, the optimization functional has now several components,

$$\mathbf{E}(\mathbf{m}) = \begin{pmatrix} e_1(\mathbf{m}) \\ \vdots \\ e_n(\mathbf{m}) \end{pmatrix}, \tag{5.20}$$

where each component $e_i(\mathbf{m})$ is defined similar to equation 5.5. In linearized approaches and some genetic algorithms this problem would be transformed into a single-objective problem by minimizing a weighted sum of the components

$$E_{so}(\mathbf{m}) = \sum_i \lambda_i e_i(\mathbf{m}). \tag{5.21}$$

The most prominent example for such an approach in geophysics is regularized inversion, which we will discuss in more detail later, where λ is called the *Lagrange multiplier* [Tikhonov and Arsenin, 1977], but it has also been applied to joint inversion problems [Julia et al., 2000, Gallardo and Meju, 2003]. The problem with this is that the optimum weighting factors λ_i are not know a priori, but have either to be guessed or systematically varied to see the influence on the resulting models [Hansen, 1992]. NSGA-II does not follow this approach but preserves the information from the various minimization functions by ranking the models according to *dominance*.

A model $\mathbf{a}$ is said to dominate another model $\mathbf{b}$, $\mathbf{a} \prec \mathbf{b}$, if and only if all objective function values are less or equal and at least one objective function value is less in $\mathbf{a}$ than in $\mathbf{b}$,

$$\mathbf{a} \prec \mathbf{b} \Leftrightarrow \forall i : e_i(\mathbf{a}) \leq e_i(\mathbf{b}) \wedge \exists i : e_i(\mathbf{a}) < e_i(\mathbf{b}). \tag{5.22}$$

We can see that with this definition, we only compare corresponding objective function values and avoid the mixing of Equation 5.21. Even though the definition is simple, it is quite difficult to envisage its implications and which models

Illustration of dominance
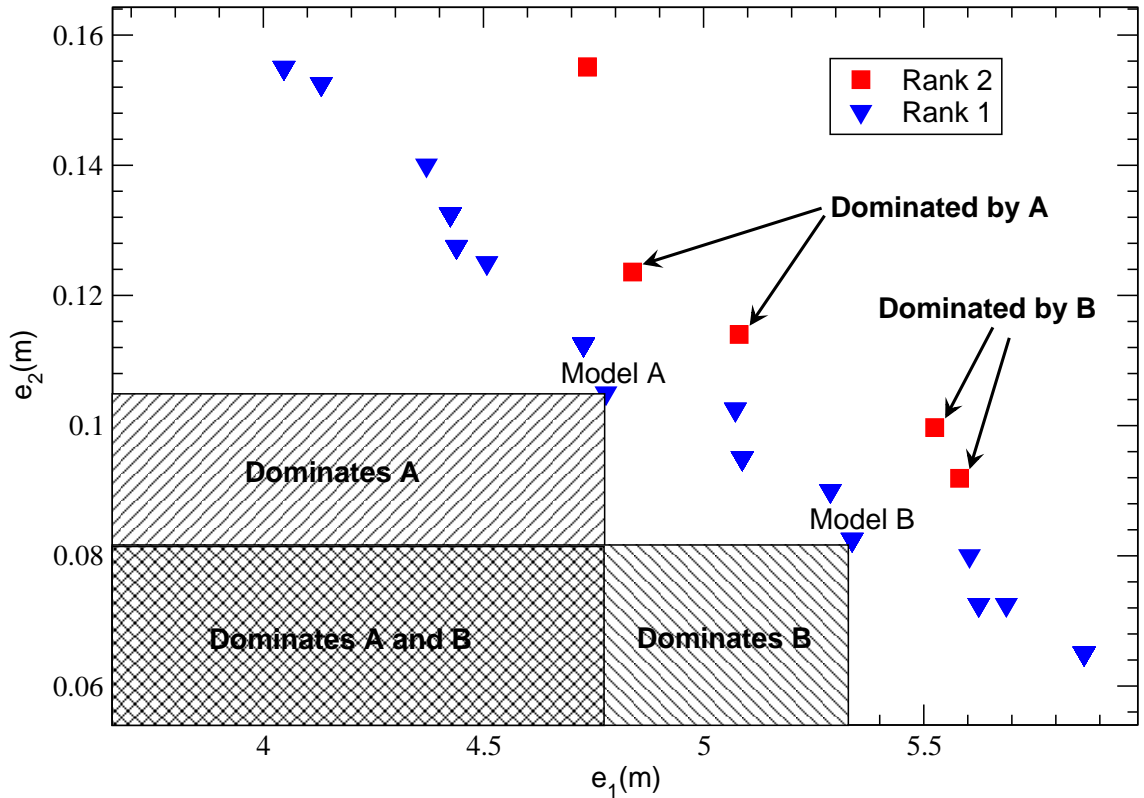
Figure 5.2: Graphical representation of the principle of dominance. The blue triangles are considered to be of equal quality with respect to the criterion of equation 5.22 and assigned rank 1. Every model plotted with a red square is dominated by one of the rank 1 models.

dominate others, a graphical representation in Figure 5.2 illustrates the situation.

For this example we assume a two component objective function $\mathbf{E}(\mathbf{m}) = (e_1(\mathbf{m}), e_2(\mathbf{m}))$. Each model then can be plotted as a point in a two-dimensional plane, for more components the same principles apply, but visualization becomes more difficult. The shaded regions in Figure 5.2 show the areas in objective function space that dominate the two models marked A and B. Basically the dominance criterion divides the objective function space into four quadrants with respect to the chosen model. Models located in the first quadrant are dominated by the chosen model, while it is dominated by models located in the third quadrant. Models in the second and fourth quadrants are neither dominated, nor do they dominate, the chosen model, and are considered to be of equal quality. This

is fundamentally different from the weighted sum in Equation 5.21 that describes ellipses around the origin where the ratio of the semi-major and semi-minor axes is given by the ratio of the weights λ_i.

Using the criterion of dominance, we can now rank the population. Members of the population that are not dominated by any other members are called *Pareto optimal* and assigned rank 1. Once all non-dominated members have been found, they are removed from the ranking scheme, and the process is repeated. Members of the population that are now non-dominated are assigned rank 2, and we continue until the complete population has been ranked.

Deb et al. [2002] describe a more complicated algorithm to reduce the computational complexity of the ranking operation at the cost of additional storage. In our case, this part of the inversion scheme only takes a small fraction of the overall computations, even for large populations, so there is no need for the more sophisticated implementation.

Another, secondary, criterion Deb et al. [2002] use to judge the quality of a model is the proximity to other models in objective function space. What we want to accomplish, by using the Pareto-optimality criterion, is to get an idea about the inherent trade-off between achieving the various minimization goals. To this end we want to obtain not only models that match both criteria reasonably well, but also find extremal cases where one of the goals is met, but the other one is not. To prevent the algorithm from converging only to the middling models, we have to implement special measures [Goldberg, 1989]. The approach taken by Deb et al. [2002] uses the *crowding distance* criterion without the need for any user specified parameters.

For each rank $\mathcal{I}$ we sort the members of the population in that rank successively by each objective function value, calculate the distance to the neighbouring members in objective function value space and normalize by the maximum distance within that rank. The average of those distances over all objective function components is the crowding distance. Models at the edge of each front that have only one neighbour are assigned a crowding distance of infinity. Again a graphical representation makes the concept much clearer (see Figure 5.3).

Once the rank and the crowding distance CD have been calculated, we can order the models with the crowded-comparison operator, $\prec_n$

$$\mathbf{a} \prec_n \mathbf{b} \Leftrightarrow \text{Rank}(\mathbf{a}) < \text{Rank}(\mathbf{b}) \vee (\text{Rank}(\mathbf{a}) = \text{Rank}(\mathbf{b}) \wedge \text{CD}(\mathbf{a}) > \text{CD}(\mathbf{b})) .$$
(5.23)

This crowded-comparison operator is used in a *binary-tournament* selection scheme to create candidates for the new population. In binary tournament selection two members of the population are chosen randomly and compared using the crowded-comparison operator. The smaller one of the two, with respect to the operator

Calculation of crowding distances
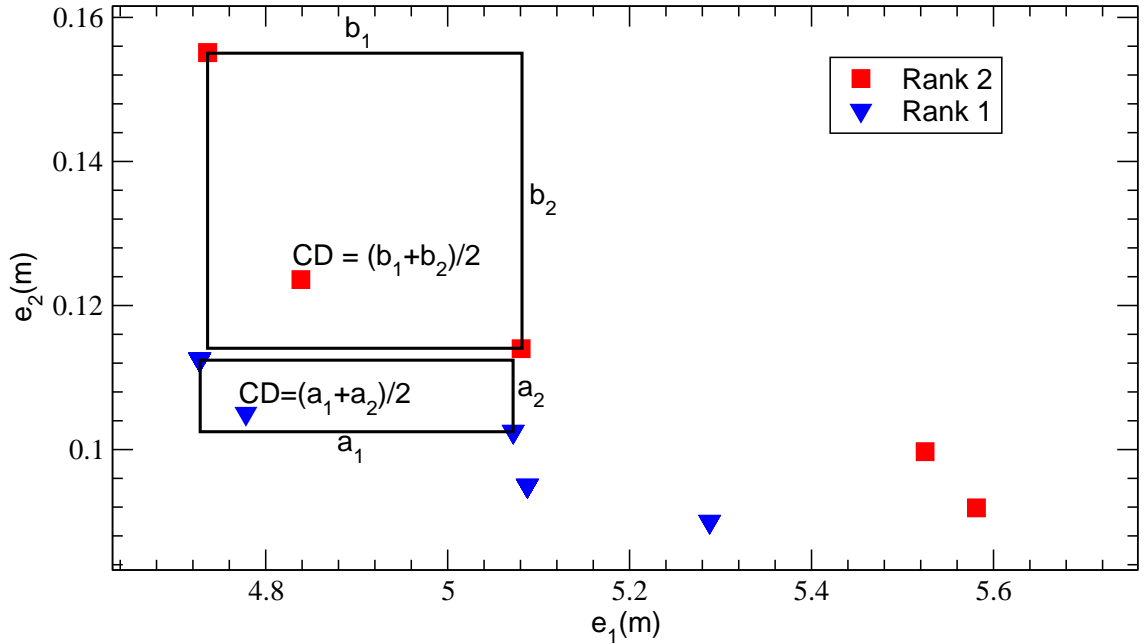
Figure 5.3: Graphical representation of crowding distance calculations. Within each rank the distance to the neighbouring models in objective function space is calculated and the average distance is the crowding distance of the respective models.

$\prec_n$, is selected for the next generation. This approach has two advantages over a fitness scaling [Goldberg, 1989] or annealing [Sambridge and Mosegaard, 2002] approach. First, it does not need a user selected parameter to determine the distribution of models with fitness, and second, it is robust in the sense that a single model with fitness much higher than average will not be replicated excessively in subsequent generations.

One drawback of binary tournament selection is that there is no guarantee that the best model will be present in the next generation. Even though it always wins the comparison stage, it might not be selected in the random drawing process. To avoid discarding good models, NSGA-II uses an *elitist* selection process. At each iteration, the new generation is combined with the last generation, and the combined population of size 2N is ranked. The best N members from this intermediate population form the final next generation.

These steps together with the diagram in Figure 5.1 completely describe NSGA-II. As in our case one calculation of the objective function value takes a much longer time than all other operations within the genetic algorithm, we implement an ad-

ditional feature. For each model we calculate a hash value, and this value and the corresponding objective function value are stored in an archive during one run of the algorithm. Before the objective function is evaluated, the algorithm checks whether this model has been calculated before. If this is the case, the objective function values are taken from the archive and not recalculated. Especially in the later stages of the inversion, where most new models are generated through mutation and comprise only a small fraction of the population, the speedup achieved by archiving is a factor of 2 or higher. Even in the earlier stages, the gain by avoiding recalculation of a few models is larger than the computational cost of hashing and retrieving the stored values.

5.4 Simple performance tests for multi-objective problems

Before we discuss inversion of magnetotelluric and receiver function data, we will demonstrate some of the abilities of the genetic algorithm on a simple, but well controlled, artificial minimization problem. This will highlight the benefits of using a genetic algorithm for multi-objective inversion.
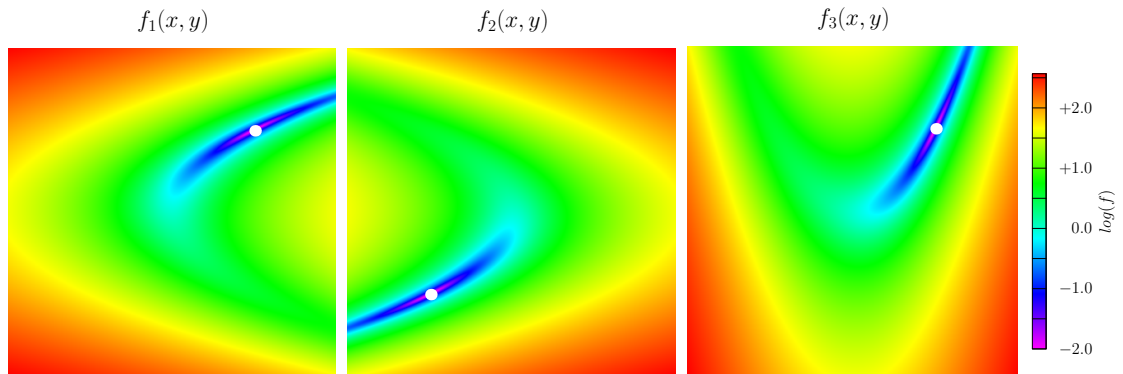
Figure 5.4: Graphical representation of the values of the three test functions f_1, f_2 and f_3 to demonstrate the abilities of the genetic algorithm. The minimum values are marked with a white dot.

Figure 5.4 shows the three functions we use to construct different multi-objective test problems for the genetic algorithm:

$$f_1(\mathbf{x}) = 10\left(x_1 - x_2^2\right)^2 + (1 - x_2)^2 \quad \min = (1, 1), \tag{5.24}$$

$$f_2(\mathbf{x}) = 10\left(x_1 + x_2^2\right)^2 + (1 + x_2)^2 \quad \min = (-1, -1), \tag{5.25}$$

$$f_3(\mathbf{x}) = 10\left(x_2 - x_1^2\right)^2 + (1 - x_1)^2 \quad \min = (1, 1). \tag{5.26}$$

$$f_1(x_1, x_2) + f_3(x_1, x_2)$$
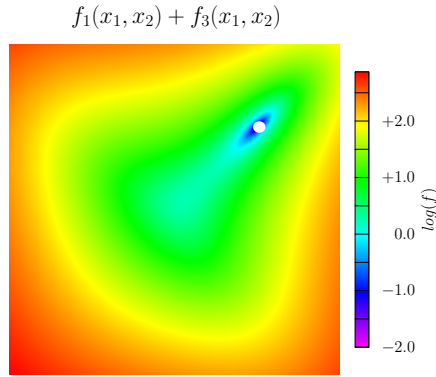
Figure 5.5: Equally weighted sum of functions f_1 and f_3 as an example for a multi-objective test problem where the individual minima coincide.

One problem in multi-objective optimization, and particularly in the joint inversion of different geophysical datasets, is the possibility of different minima in each objective function. In terms of Earth structure and our inversion problem, this means that magnetotellurics and seismics sense different structural features within the subsurface. We will discuss the arguments for and against similarity between the two datasets in Chapter 7, for now we only note that this problem exists, and that ideally a joint inversion scheme not only produces a model that fits the data, but also gives some indication whether the common model approach has any meaning for the given data.

From the three test functions we construct two multi-objective optimization problems,

$$\mathbf{O}_1(\mathbf{m}) = \begin{pmatrix} f_1(\mathbf{m}) \\ f_2(\mathbf{m}) \end{pmatrix} \text{ and } \mathbf{O}_2(\mathbf{m}) = \begin{pmatrix} f_1(\mathbf{m}) \\ f_3(\mathbf{m}) \end{pmatrix}. \tag{5.27}$$

f_1 and f_2 have different minima, so $\mathbf{O}_1$ simulates a case where the MT and seismic models are incompatible with each other, whereas the minima for f_1 and f_3 are identical, and thus $\mathbf{O}_2$ simulates compatible models. Before we analyse the performance of the genetic algorithm on these two problems, we will show why the linearized approach is problematic even in this very simple case.

As mentioned above, we have to construct a scalar objective function from Equations 5.27 by multiplication with a weight vector λ. For the minimization problem with identical minima, we see an example with identical weights in Figure 5.5. In this case the combination of the two objective function has made the minimization problem easier. The combined error surface does not show the elongated valley structure of the two original surfaces, but is more symmetric and closer to a quadratic function. We can expect any linearized method to perform well in this case. In addition, for any given relative weighting the absolute minimum remains
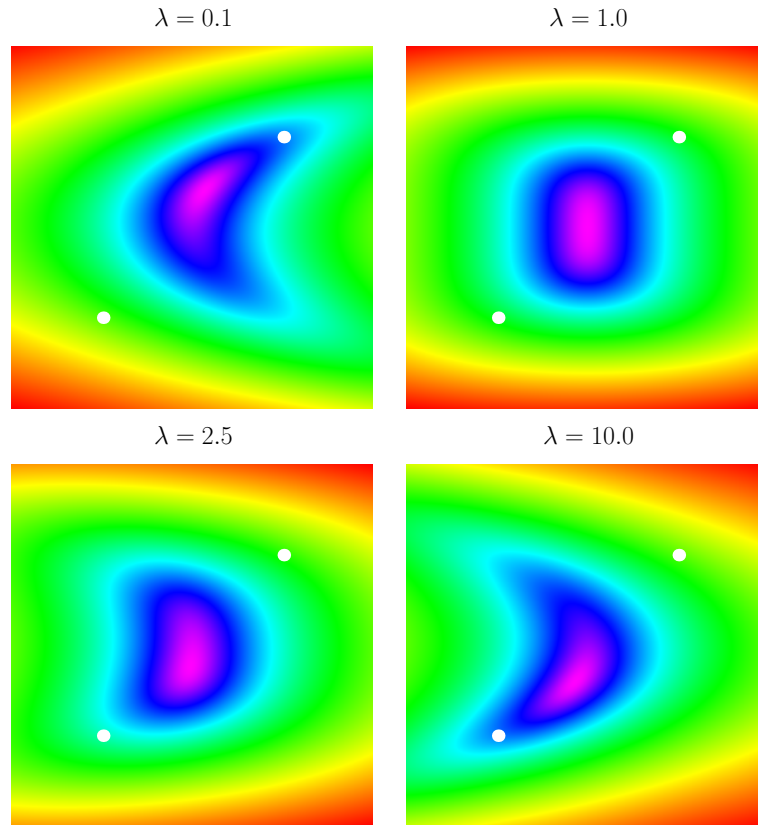
Figure 5.6: Weighted sums of the two incompatible test functions f_1 and f_2 for different relative weights λ. In each plot the minima of the individual test functions is marked by a white circle.

the same.

The situation changes when we examine weighted sums of $\mathbf{O}_1$. Figure 5.6 shows the weighted sums of $\mathbf{O}_1$ for 4 different relative weighting factors λ. The white dots mark the minima of the individual objective functions. We can see that none of the minima of the summed objective functions coincide with any of the minima of the individual objective functions for any weighting factor λ. As with the other test problem, the error surface for the summed objective functions is closer to a quadratic function, but in this case a linearized method would converge to a minimum that has no relation to the minima of the individual objective functions. The only way to realize this with a linearized method is to systematically vary the weighting factors and the starting models, and examine the impact on the results.

We now turn to the results from the genetic algorithm on these two test problems. To demonstrate the maximum potential of the method we use a large population size $N_p = 500$ and number of iterations $N_i = 50$ corresponding to 25,000

Figure 5.7: The set of Pareto optimum objective function values retrieved by the genetic algorithm for the multi-objective test functions O_2 (left) and O_1 (right).

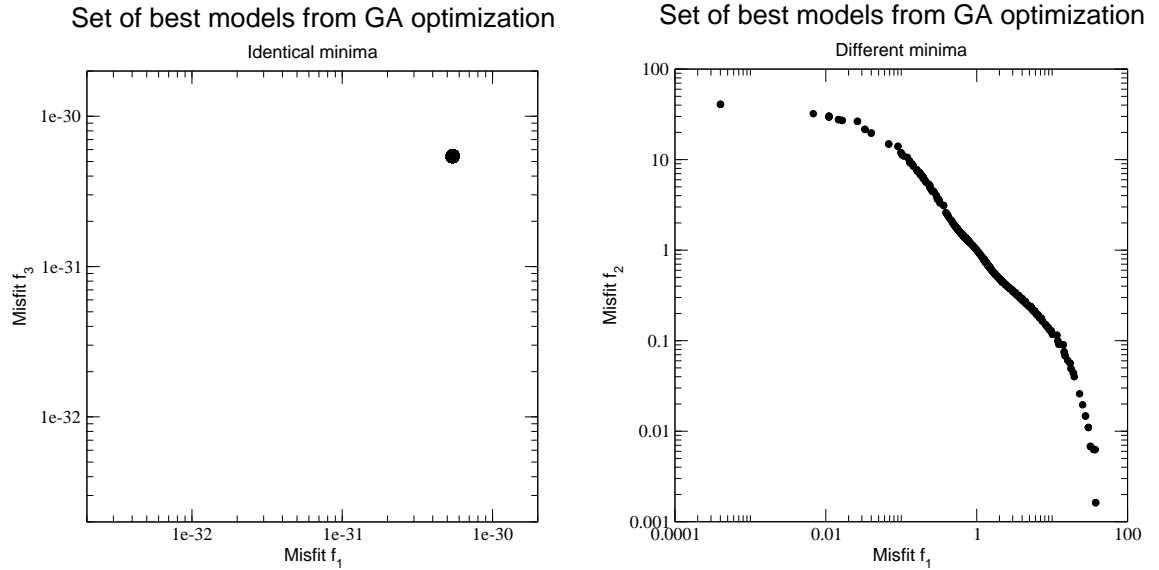model evaluations. The two model parameters are encoded with a 10-bit string, a minimum value of -10, and a stepsize of 0.02 each. The resulting search area is $-10 \ldots 10.48$ and if all models used in the search were different, we would cover about 10^{-4} % of the model space. This is unrealistically high, and later we show the results with more realistic settings. At the moment we are interested in what is theoretically possible.

Figure 5.7 shows the Pareto optimum fronts that the genetic algorithms yields for the two test problems. For problem O_2 that consists of two components with the same minimum, the front consists of a single model with objective function values that are close to the numerical precision. The genetic algorithm identified the single minimum of the combined error surface. In contrast, the front for the problem with incompatible components consists of a large number of models that span a large range of objective function values for both components. Each point within this front represents a solution of the multi-objective problem with a different weighting factor between the two components. The full general solution to the problem is given by all the models in the theoretical Pareto-optimal front. The genetic algorithm retrieves a large part of this front, although the extreme values are not very well represented. We do not observe very small values for either of the components of the objective function, even though solutions with zero weight to one of the components are part of the theoretical Pareto front. Despite the mechanisms build into the algorithm to spread out the front, there is a tendency towards individuals that perform reasonably well for all criteria.
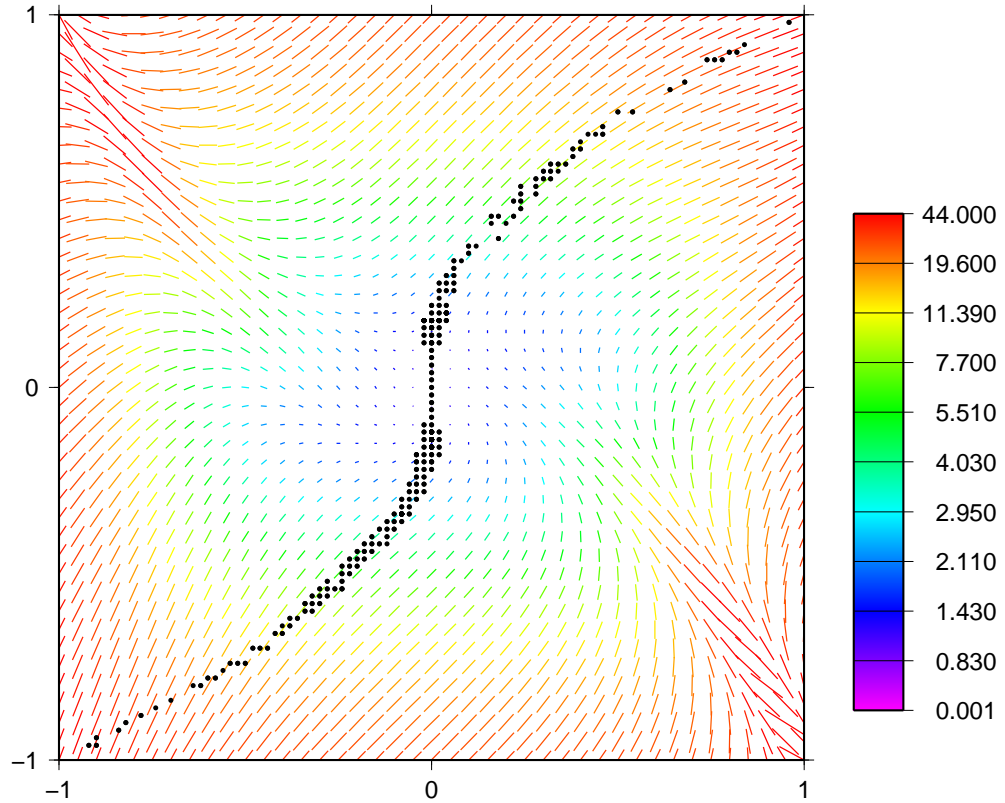
Figure 5.8: The set of Pareto optimum models plotted on top of a vectorial representation of $\mathbf{O}_1(\mathbf{m})$. The two components of the vector correspond to the logarithm of the two objective function components. In addition each vector is coloured according to its length.

Nonetheless, in this case the solutions to the two problems given by the genetic algorithm clearly distinguish the case of coincident minima from differing minima of the two components.

We now have a look at the distribution of solutions for the incompatible case. Figure 5.8 shows the solutions found by the genetic algorithm, together with the values of the two objective functions plotted as vectors. The x_1-component of each vector is given by the logarithm of the first objective function value, while the x_2 component is given by the logarithm of the second function value. The colour of the vector represents its overall length. We can see how the solutions trace the path of the minima with different weights that we already observed in Figure 5.6. We also observe the effect of finite discretization of the model parameters within the genetic algorithm. The solutions found by the GA fall onto a regular grid, and where the trace of the minima curves we can see the step-like fashion in which the GA solutions follow this trace.

Figure 5.9: Comparison of the results of different runs of the GA on the two test problems with a small population size and limited iterations. In general the Pareto fronts for the incompatible test problems contain more solutions and have higher misfit. However, some runs on the compatible problem exhibit relatively poor performance.

Finally we examine the effect of a small population size. By decreasing the population size by a factor of 20 to $N_p = 25$ and the number of iterations by a factor of 5 to $N_i = 10$, we decrease the overall function evaluations by a factor of 100. The new Pareto front from 20 different runs for each objective function is shown in Figure 5.9. Compared to the results shown in Figure 5.7, the difference between the incompatible and compatible objective functions is less clear. None of the runs for the compatible case retrieves the low value we achieved with the large population. More than half of those runs also result in two or more models within the Pareto front, suggesting a non-existent trade-off between fitting the two components of the objective function. In one case the performance of the models is even worse than that of most of the incompatible results. Still, the majority of solutions for the compatible problem has a lower final misfit and less solutions within the Pareto optimal front than the incompatible solutions. The distinction between the compatible and incompatible cases is not as clear as it was

for the large population. However, we still get and indication of the compatibility, particularly when we compare the results from several runs.

Another thing to bear in mind is that with the small population we used the statistical properties of the genetic algorithms are worse than for a large population, even if the search space for the large population is much bigger. The small number of iterations severely limits the interaction between the models, which in other cases would, to some degree, balance the fluctuations in starting population. One lesson to be learned from this experiment is that as large a population size as computationally feasible should be used and several runs of the algorithms should be performed to obtain reliable results. We will return to this issue when we test synthetic models of receiver function and magnetotelluric data.

5.5 Regularization and non-uniqueness

So far we have only discussed the minimization algorithm and how we measure the quality of our model in terms of data misfit, but neglected the issue of *regularization* [Tikhonov and Arsenin, 1977], which plays an important role in geophysical inversion. The objective functions we examined so far had one uniquely defined minimum, even though in the incompatible test case these minima were different for each function. When looking at the error surface of geophysical inverse problems, particularly if we consider measurement noise, we see that the minimum is not very well defined, but considerable trade-offs exist (c.f. Figure 6.1). All models that fit the measured data to the same level, can be considered to be of equal quality, so we have to use another criterion to select amongst these models.

The commonly used approach to overcome this problem, is based on a paradigm known as *Occam's razor*, which states that we should use the simplest possible model to explain the data [Constable et al., 1987]. Obviously the notion of simplicity is somewhat vague, but a few standard measures have been established and are used in most inversion codes. The choice of measure, however, depends on our preconception on the structure of the Earth and will influence what type of model we find.

The simplest possible type of regularization, even though it is often not regarded as regularization, is to limit the range of the parameter values and the number of model layers to specified values that are considered reasonable [Constable et al., 1987]. This type of regularization is an intrinsic part of our genetic algorithm, as we have to specify the parameter encoding before the inversion. The restrictions that result from the chosen encoding will keep the output models simple, as long as we keep the parameter range and number of layers small. This method is not without problem, because at the same time we want to allow the algorithm to have

enough flexibility to model the structure of the subsurface. By varying the number of layers and the model ranges, we can often identify appropriate choices in 1D inversion problems, for 2D or even 3D problems, this approach is not feasible, and we have to include a measure of model structure into the inversion. Even in 1D a properly chosen regularization functional can improve the inversion.

The most widespread measures to quantify the complexity of a model in 1D are

$$R_1 = \int_{z=0}^{\infty} |\mathbf{m} - \mathbf{m}_{\text{ref}}|^2 dz, \tag{5.28}$$

$$R_2 = \int_{z=0}^{\infty} \left| \frac{d}{dz} (\mathbf{m} - \mathbf{m}_{\text{ref}}) \right|^2 dz, \text{ and} \tag{5.29}$$

$$R_3 = \int_{z=0}^{\infty} \left| \frac{d^2}{dz^2} (\mathbf{m} - \mathbf{m}_{\text{ref}}) \right|^2 dz. \tag{5.30}$$

The optional reference model $\mathbf{m}_{\text{ref}}$ can be used to tie the inversion models to a known reference model such as PREM [Dziewonski and Anderson, 1981] for seismic data. If no sensible reference model exist, the corresponding vector can be set to zero. Note the similar structure of all these measures to the misfit function in Equation 5.4. In fact they are incorporated into the inversion as just another objective function.

R_1 limits the absolute value of the model parameters. If no reference model is used, it will chose the model with the lowest average parameter values, for example the least resistive model in the case of magnetotellurics. When a reference model is used, it limits the variation of the inversion model from that reference model.

R_2 reduces the variability of the gradient. Without a reference model, it limits the change of parameter values between adjacent layers. This represents the common notion of a smooth model. When we use a reference model with this measure, it results in small values when the changes in the inversion model are the same as in the reference model. The absolute values can differ by an arbitrary constant.

The use of the second derivatives in R_3, corresponds to minimizing the curvature of the models. Without a reference model this allows for inversion models with a uniform gradient, but suppresses strong oscillations. With a reference model the minimum value is given for models that have the same curvature as the reference model.

These three measures are not the only possible regularization functionals, but a large number of different functionals exists [e.g. Farquharson and Oldenburg, 1998, Zhdanov and Tolstaya, 2004]. For our purposes, and for 1D, the measures presented above are sufficient. It is important to note that in virtually all situations the regularized inversion problem provides an incompatible problem. For all but uniform models, the model with minimum data misfit will not be the model with lowest complexity. This trade-off, and the resulting difficulty in choosing

the correct weighting parameters, is well known in all inversion problems, and the plot of the trade-off between misfit and roughness is termed the *L-curve* [e.g. Hansen, 1992, Farquharson and Oldenburg, 1998, Farquharson and Oldenburg, 2004, Pedersen, 2004]. As we discussed above, NSGA-II provides an alternative way to incorporate this competing objective function into the minimization algorithm, and the first application of this algorithm in geophysical inversion was to automate the construction of the L-curve [Schwarzbach et al., 2005].

Apart from guiding the inversion process to a certain type of model, regularization has another role in linearized inversion methods. The equations 5.28 – 5.29 are all strictly quadratic. Adding such a quadratic functional to the possibly non-quadratic error functional, gives the objective function an approximately quadratic shape. This geometry makes it easier for the linearized algorithm to approach the minimum and reduces the influence of noise. For our genetic algorithm approach, this does not make any difference, as we do not use gradient information. We can incorporate a regularization functional as an additional objective function, however, this makes the trade-off plots as shown in Figures 5.9 and 5.7 three-dimensional.

5.6 Resolution and model covariance

One of the reasons we discussed linearized inversion in such detail, apart from contrasting it to GA based inversion, is that a lot of methods to appraise the inversion results have been developed in this context and we will need the concepts introduced above to understand them. Two characteristics of any model of the Earth are particularly useful to understand the quality of the model: *resolution* and *model covariance*. These two quantities were introduced into geophysical inversion by Backus and Gilbert [1968] and have since then been used to assess which parts of the model can be deemed reliable and where the data does not contain sufficient information on the structures in the subsurface.

Both of them are connected to the sensitivity matrix $\mathbf{G}$ and hence are only exact for linear inverse problems. Using them for non-linear problems will provide a first order approximation, which is useful to get an idea of the problem, but not always appropriate [Ledo et al., 2004]. The *model covariance matrix* simply expresses how errors in the data map into errors of the model [Menke, 1989]

$$\mathrm{cov}(\mathbf{m}) = \mathbf{G}^{-g}\mathrm{cov}(\mathbf{d})\mathbf{G}^{-g\mathsf{T}}. \tag{5.31}$$

Here the superscript $-g$ denotes the *generalized inverse* of the sensitivity matrix. In general the sensitivity matrix is not square, so we cannot use the ordinary matrix inverse. Even when it is square, the inverse might not exist, if the sensitivity

matrix is singular. Instead, we perform a *singular value decomposition* of the $n \times m$ matrix $\mathbf{G}$, i.e. we factorize it into three matrices

$$\mathbf{G} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathsf{T}}. \tag{5.32}$$

The $n \times n$ matrix $\mathbf{U}$ spans the data space, while the $m \times m$ matrix $\mathbf{V}$ spans the model space. The $n \times m$ matrix $\mathbf{\Lambda}$ contains the non-negative singular values on its diagonal. If $n < m$, i.e. we have more model parameters than data and are solving an underdetermined problem, we will have $n - m$ singular values that are exactly 0. But even when $n \geq m$, the magnitude of the eigenvalues can span several orders of magnitude and we can consider values below a certain threshold p as zero for numerical purposes. Setting these eigenvalues to zero gives the *truncated singular value decomposition*

$$\mathbf{G}_p = \mathbf{U}_p\,\mathbf{\Lambda}_p\,\mathbf{V}_p^{\mathsf{T}} \tag{5.33}$$

and the corresponding generalized inverse

$$\mathbf{G}_p^{-g} = \mathbf{V}_p\,\mathbf{\Lambda}_p^{-1}\,\mathbf{U}_p^{\mathsf{T}}. \tag{5.34}$$

Comparing equations 5.31 and 5.34, we see that with increasing threshold, i.e., using only the eigenvectors associated with large eigenvalues, the model covariance becomes smaller and smaller. This decrease in covariance comes at a price though. Recalling from equation 5.7 that

$$\mathbf{d}_{\mathrm{obs}} = \mathbf{G}\mathbf{m}_{\mathrm{true}} \tag{5.35}$$

and realizing that our estimated model $\mathbf{m}_{\mathrm{est}}$ is given by

$$\mathbf{m}_{\mathrm{est}} = \mathbf{G}^{-g}\mathbf{d}_{\mathrm{obs}}, \tag{5.36}$$

we see that in the truncated case

$$\mathbf{m}_{\mathrm{est}} = \mathbf{G}^{-g}\mathbf{d}_{\mathrm{obs}} = \mathbf{G_p}^{-g}\,\mathbf{G_p}\,\mathbf{m}_{\mathrm{true}} = \mathbf{R}\mathbf{m}_{\mathrm{true}}. \tag{5.37}$$

The matrix $\mathbf{R}$ is called the *resolution matrix*. If we do not have any vanishing singular values and do not truncate, we theoretically can recover the true model. This model will be strongly influenced by noise in the data though. Removing small singular values stabilizes the model, but $\mathbf{R}$ will no longer be the identity matrix and we recover a filtered version of the true model

$$\mathbf{R} = \mathbf{G_p}^{-g}\,\mathbf{G_p} = \mathbf{V}_p\mathbf{V}_p^{\mathsf{T}}. \tag{5.38}$$

Thus with increasing truncation level we reduce the model covariance, but the corresponding model becomes a strongly filtered version of the true model [e.g. Kalscheuer and Pedersen, 2007]. Conversely the *null-space projection matrix*

$$\mathbf{P} = \mathbf{I} - \mathbf{R} \tag{5.39}$$

indicates possible changes in the model that have no or only minor impact on the predicted data.

Muñoz and Rath [2006] use this fact to construct a model perturbation algorithm that illustrates the range of possible models that can explain the data

$$\mathbf{m_c} = \mathbf{m_0} + \mathbf{P}\Delta\mathbf{m}. \tag{5.40}$$

A random model perturbation is projected into the null-space and added to the inversion model to construct an equivalent model in terms of data misfit. We have to bear in mind though that this approach is based on a linear approximation around the inversion model and monitor the change in data misfit to avoid violating the approximation. Alternatively Kalscheuer and Pedersen [2007] employ a non-linear analysis which is computationally much more demanding, but improves the accuracy of the results.

We will use the simple perturbation algorithm described by Equation 5.40 to examine in how far we can modify our resulting models to violate the joint interface assumption of the joint inversion approach.

*If my answers frighten you Vincent, then you should cease asking
scary questions.*

Jules Winnfield

6

Is it real? Testing the algorithm and appraising the inversion results

Now that we have a good understanding of the inner workings of the inversion method, we can start to examine the datasets and their behaviour in an inversion procedure. First, we will define the exact objective functions we use for each of the datasets, and explain the motivation behind these choices. Then we will look at the dependence of the individual objective functions on the model parameters, to identify a suitable set of parameters for the inversion procedure. Finally, we look at the inversion results from synthetic data created from realistic Earth models, to see in how far the genetic algorithm can distinguish between compatible and incompatible datasets in the case of MT and receiver function inversion. In the process we introduce some additional tools to quantify the compatibility of the datasets.

6.1 Defining the individual objective functions

Although Equation 5.4 fully describes the misfit criterion of the inversion procedure, we have to make a number of choices for each dataset that will influence the results of the inversion. The two main decisions we have to make are, (1) which data to use and (2) how to specify the corresponding error. The first decision is mainly of concern for magnetotelluric data, while the second concerns mostly the

receiver function data.

The best, and most accurate, way to model our data would be to create a full three-dimensional model of the region of interest and model all aspects of both datasets. Even though 3D forward modelling codes for both MT and teleseismic data exist [Siripunvaraporn et al., 2002, Avdeev, 2005], and individual 3D inversion is slowly coming within reach [Weiland et al., 1995, Siripunvaraporn et al., 2005, Avdeeva and Avdeev, 2006], the forward modelling codes are computationally too demanding for our genetic algorithm approach. The more efficient linearized methods might be computationally feasible, but the issue of possible mismatches between seismic and electric structure would be completely ignored in such an approach as demonstrated with the simple example in Chapter 5. Without independent information, an indicator of the compatibility of the datasets is equally as important as the best fitting model. For this reason we have to constrain ourselves to a computationally feasible 1D approach.

For the MT data, we choose to model the real and imaginary parts of the Berdichevskiy invariant [Berdichevskiy and Dmitriev] of the observed and predicted data to calculate the misfit in Equation 5.5. For the synthetic input models that we examine here the choice of the invariant does not have any impact, as for the synthetic 1D data we invert the invariant is equal to the off diagonal elements. For recorded data, the invariant has two important effects: For purely one dimensional data that are affected by random noise, the averaging process reduces the impact of the noise. For data that are not strictly one dimensional, and where the the two polarizations differ, the Berdichevskiy invariant makes the inversion data independent of the coordinate system by combining the information from both polarizations to create an average one dimensional approximation.

For the receiver functions, we have to make a modification to Equation 5.5. We do not have individual error estimates for each datapoint, but only an overall relative error for the entire receiver function. Also, a number of datapoints will be close or equal to zero. For these datapoints the division by a relative error will be numerical unstable. For this reason we compute the squared difference between observed and modelled data, and divide the sum by a constant relative error estimate. Unfortunately this also removes the equalizing property from Equation 5.5, and therefore we have to exclude the initial correlation peak that is numerically much larger than the part of the receiver function we are interested in.

6.2 Choosing the inversion parameters

For the following synthetic tests and the inversions of real data we use a crossover probability $p_c = 0.6$ and a mutation probability of $p_m = 0.2$. Experiments with

different values have shown that these two values provide a good compromise between sufficient innovation and preservation of well performing models. The population size is usually set to $N_p = 1000$ and we perform 100 iterations. When these numbers differ we will indicate it in the text.

The model parameters we use in the inversion are the logarithm of electrical resistivity $\log(\rho_{el}^i)$, the thickness of each layer t^i, and the S-wave velocity v_s^i. In theory the shape of the receiver function also depends on P-wave velocity v_p^i and density ρ_i. Tests with different P-wave velocities and densities revealed, that this dependency is poorly resolved. We therefore predict these values from the S-wave velocities through the relations [Owens et al., 1984]

$$v_p = \sqrt{3}v_s, \text{ and} \tag{6.1}$$
$$\rho = 0.77 + 0.554v_s. \tag{6.2}$$

These relationships help to reduce the number of parameters and concentrate the search on the resolvable parts of the model, while at the same time keeping the unresolved parameters at realistic values.

One well known property of receiver functions is that they have a substantial trade-off between the thickness and the velocity of the surface layer [Ammon et al., 1990, Sambridge et al., 2006]. The characteristics of this trade-off are shown in Figure 6.1. We compute the misfit between a synthetic two-layer test model with 2% added noise and a number of test models with varying layer thickness and velocity in the top layer. We can see an elongated valley of low misfit values; within this valley the misfit changes only gradually. The position and shape of this valley can be easily understood by looking at Equation 4.18. An increase in v_s can be compensated for by a decreased layer thickness h, and result in the same lag time for the conversion. The transmission coefficient for the P-to-S conversion also depends on v_s and h, and consequently the amplitude of the conversion will be different for different parameter combinations. However this dependency is only weak, so within the valley we only have poor sensitivity to differences in structures. As mentioned in the discussion about regularization, this problem is widespread in geophysical inverse problems and not restricted to receiver function inversion. The issue of non-uniqueness will be of major concern when we examine our results.

It is also interesting to observe that there are a number of secondary local minima far away from the global minimum. Particularly at low velocities, we observe a number of ridge structures in the error surface. These ridges provide obstacles for linearized inversion schemes if the starting model is chosen in an area between two of them. This provides another compelling reason to use a non-linearized inversion instead of a linearized approach to invert receiver function
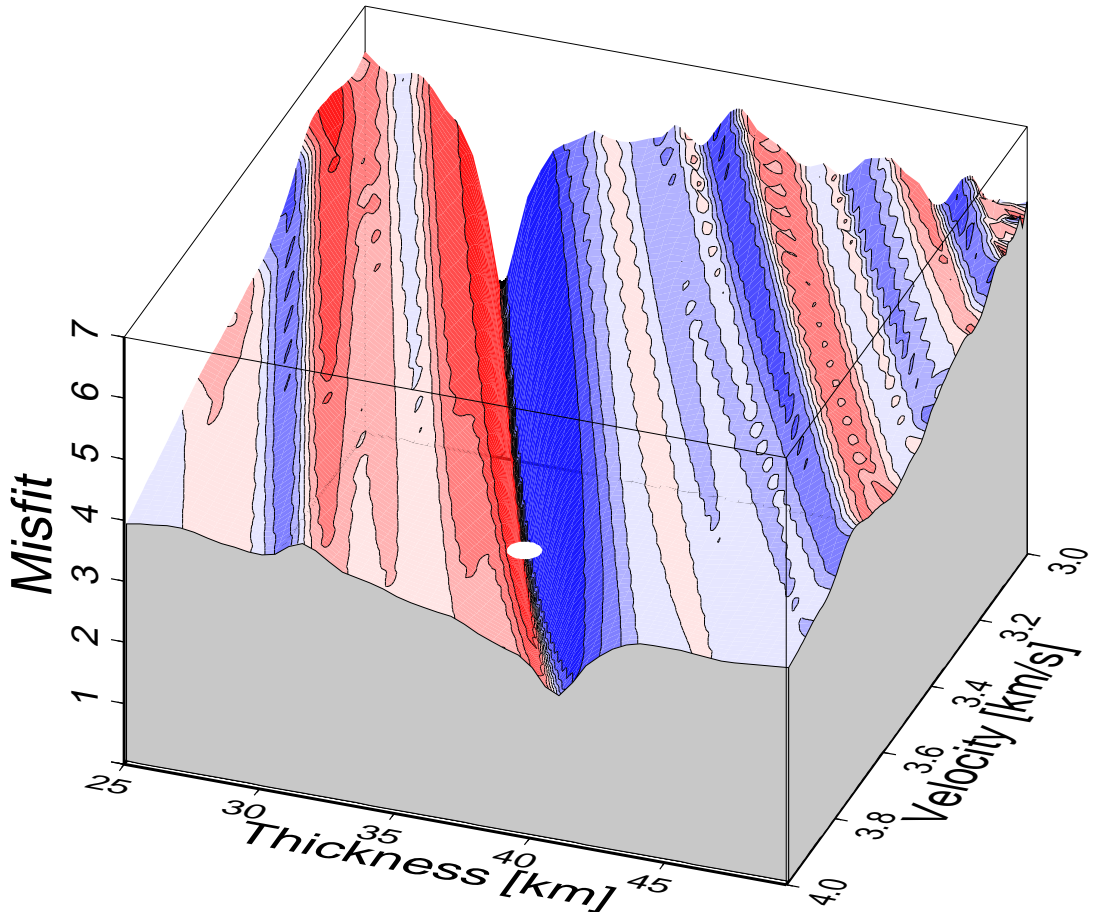
Figure 6.1: Receiver function misfit as a function of layer thickness and S-wave velocity of the surface layer. The true model parameters are marked by a white dot. The colour illustrates the length and direction of the gradient vector.

data [Shibutani et al., 1993]. Even when using a GA the structure of the error surface demonstrates the need to set the surface velocity to a fixed value. For the synthetic tests we present here, we set the velocity to the true value of the input models. For the real data inversion that we describe in Chapter 7, we will take values from regional and global crustal velocity models [Dziewonski and Anderson, 1981, Perry et al., 2002, Chen et al., 2007].

6.3 Connecting seismic velocities and electrical conductivities

Arguably the most important choice we have to make when we set up the inversion is how we link the magnetotelluric and seismic models to construct a joint model.

In our approach, the coupling between the seismic and magnetotelluric models is purely based on coincident layer boundaries. The model vector only contains one set of layer thickness values that is used for both types of forward models. The resistivities and velocities within each layer can vary independently, and there is no restriction on the gradient or second derivatives between the layers. This type of coupling provides only a loose connection between the models, and is the one-dimensional analogue to the cross-gradient approach [Gallardo and Meju, 2003]. The cross-gradient between two models is defined as

$$\mathbf{t(x)} = \nabla \mathbf{m_1(x)} \times \nabla \mathbf{m_2(x)}. \qquad (6.3)$$

It allows changes in the parameters of both models in the same direction, or a change in only one of the models while the other model stays constant. In 1D the cross-gradient is always zero, but our coincident layer approach provides a suitable approximation.

This is, of course, not the only possible approach. Saunders et al. [2005] use regularization based on the second derivative to couple electrical and seismic structure in a sedimentary environment. In their case the seismic model is fixed and only used to guide the electrical inversion. Another problem with their approach, as well as the similar one used by Haber and Oldenburg [1997], is that each time one parameter changes, the other has to change too.

Even more restrictive would be a functional coupling between the two parameters, where one parameter value is calculated from the other by some user specified function. For now, while we do not have a good understanding of the expected relationship between electrical conductivities and seismic velocities, we will use only the loose constraints of coincident layer boundaries. We will return to this discussion when we look at the inversion of real data.

6.4 Testing the inversion procedure with synthetic data: A simple problem

We test our joint inversion algorithm with synthetic input data that we calculated from a known model. We will start with noise free data from simple models, and gradually make the task more difficult by adding noise and making the models more complex. This will help to clarify a number of questions that we mentioned before:

- Can we retrieve the input models ?

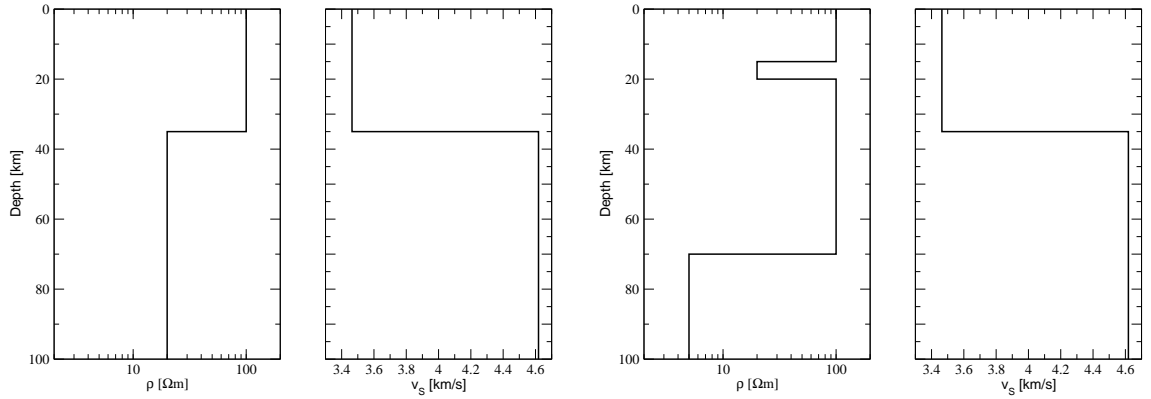- Can we distinguish compatible from incompatible models ?

Figure 6.2: The models used to create the compatible test problem (left) and the incompatible test problem (right).

- How robust is the genetic algorithm inversion ?

The first set of models we will test is shown in Figure 6.2. On the left hand side we see MT and seismic models with a single, coincident layer interface. This scenario is the simplest possible situation for the inversion algorithm. Both methods have good independent resolution for the thickness of the upper layer and the layer parameters. Therefore combining the two datasets should simplify the inversion and we can expect to retrieve both models.

In contrast, the two models on the right hand side in Figure 6.2 do not have any coincident layer boundaries. The seismic model remains the same, but the magnetotelluric model now consists of 4 layers. The first conductive layer at $15-20$ km depth models a widespread observation in MT measurements of a conductive layer in the lower crust [Jones, 1992]. Again this model does not pose any problems for an individual inversion, we have to bear in mind though that, in the presence of data errors, magnetotellurics does not resolve the conductivity and thickness of a conductive layer well, but mainly its product, the *conductance* [Weidelt, 1985]. This non-uniqueness problem in principle allows the joint inversion algorithm to adjust the parameters of this layer to move the boundary where it is required by the seismic data. The MT and receiver function data that we use as input for the inversion procedure is shown in Figure 6.3.

Noise free data

First, we perform a joint inversion on noise free synthetic data. We set the number of inversion layers to 4 in both cases and assume an error floor of 2%. This number of layers provides enough flexibility to fit all of the models individually, but does not allow modelling both structures in the incompatible case simultaneously.
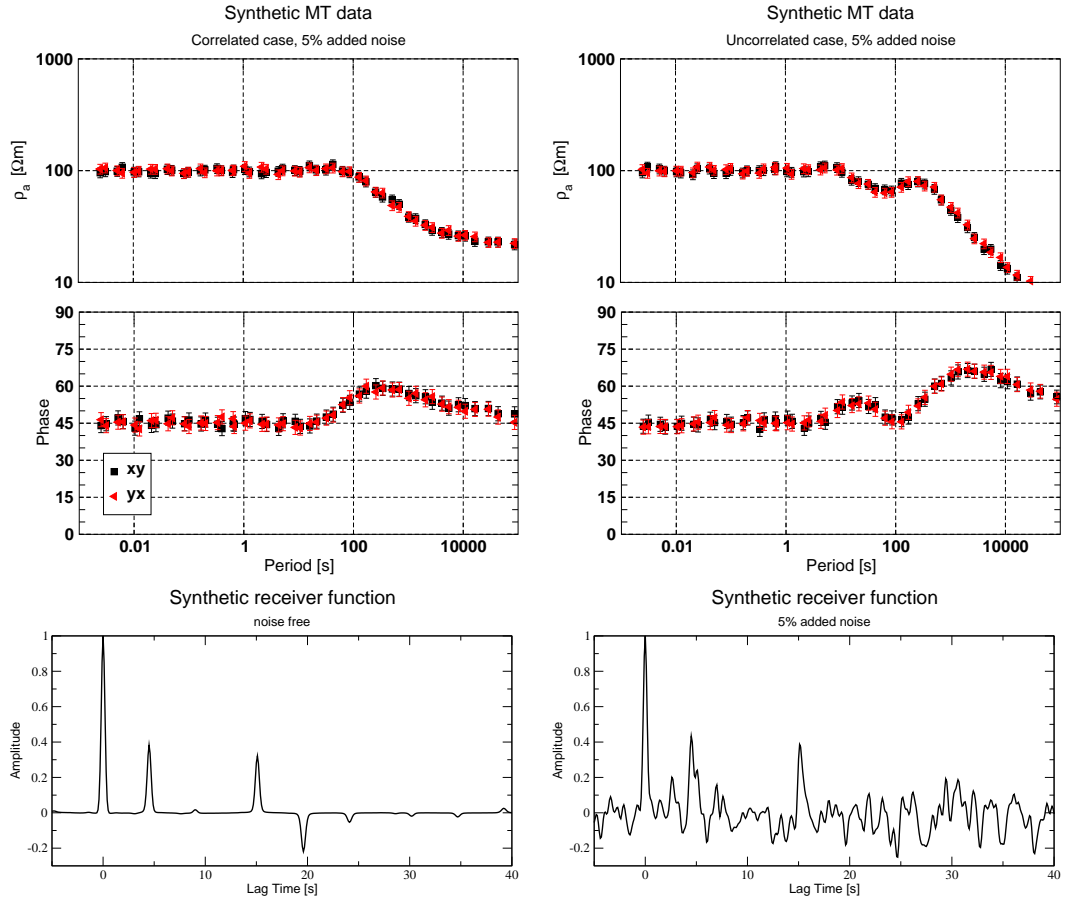
Figure 6.3: The synthetic MT data with 5% added noise for the compatible problem (upper left) and the incompatible problem (upper right). We do not show the noise free MT data, as it essentially is just a smooth version of the noisy data. The receiver functions for both problems are plotted in the lower row; the clean receiver function (lower left) and with 5% added noise (lower right).

Figure 6.4 shows the resulting fronts from 20 individual runs on each test problem. The picture we obtain is similar to the results of the test problem in Chapter 5 under ideal conditions. In all cases the inversion of the compatible models yields a single model with an RMS below 1. This indicates that we fit the data better than the estimated noise level. The misfit values for the receiver function data vary slightly between different runs of the inversion, but for the magnetotelluric data we reach the same misfit in all cases.

For the incompatible test problem the Pareto-front looks radically different. As expected, the algorithm cannot find a common model that fits both the MT and the receiver function data simultaneously. Instead we obtain a trade-off curve where only one or the other misfit value is sufficiently low. The lowest individual

Comparison of pareto fronts
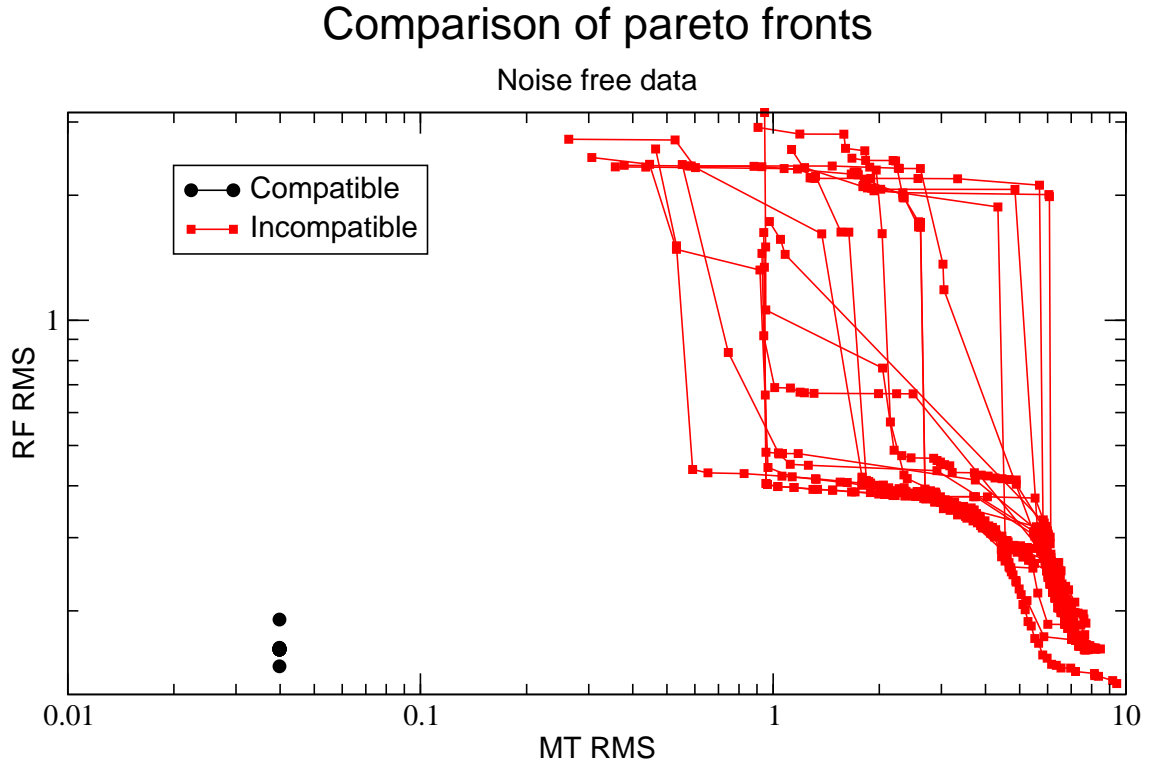
Noise free data

Figure 6.4: Comparison of the Pareto-fronts for 20 runs on the compatible and incompatible test problem. The synthetic input data were noise free.

misfit values are again below an RMS of 1. For this test scenario we only achieve misfits comparable to the compatible models for the receiver function data. The best misfit for the MT data is nearly an order of magnitude higher than for the compatible test case, although their RMS's would still classify them as suitable models. These results are encouraging because they show that, at least in this simple situation, the properties we observed for the artificial test problems also apply to our joint inversion problem.

Figure 6.5 shows the inversion output for the uncorrelated case in two different perspectives. On the left hand side we can see all models for one parameter plotted in the same graph. This gives us an impression of the spread of the models, but it is difficult to judge how many models have the same values in a certain region. For this reason we plot a 2D histogram of the model values on the left hand side. For a range of model parameters and depths we plot the number of models that have a given parameter value at that depth. In both cases we also plot the input models for comparison.

Most of the receiver function models reproduce the velocity discontinuity at 35 km depth correctly. This agrees with the observation in Figure 6.4 that the
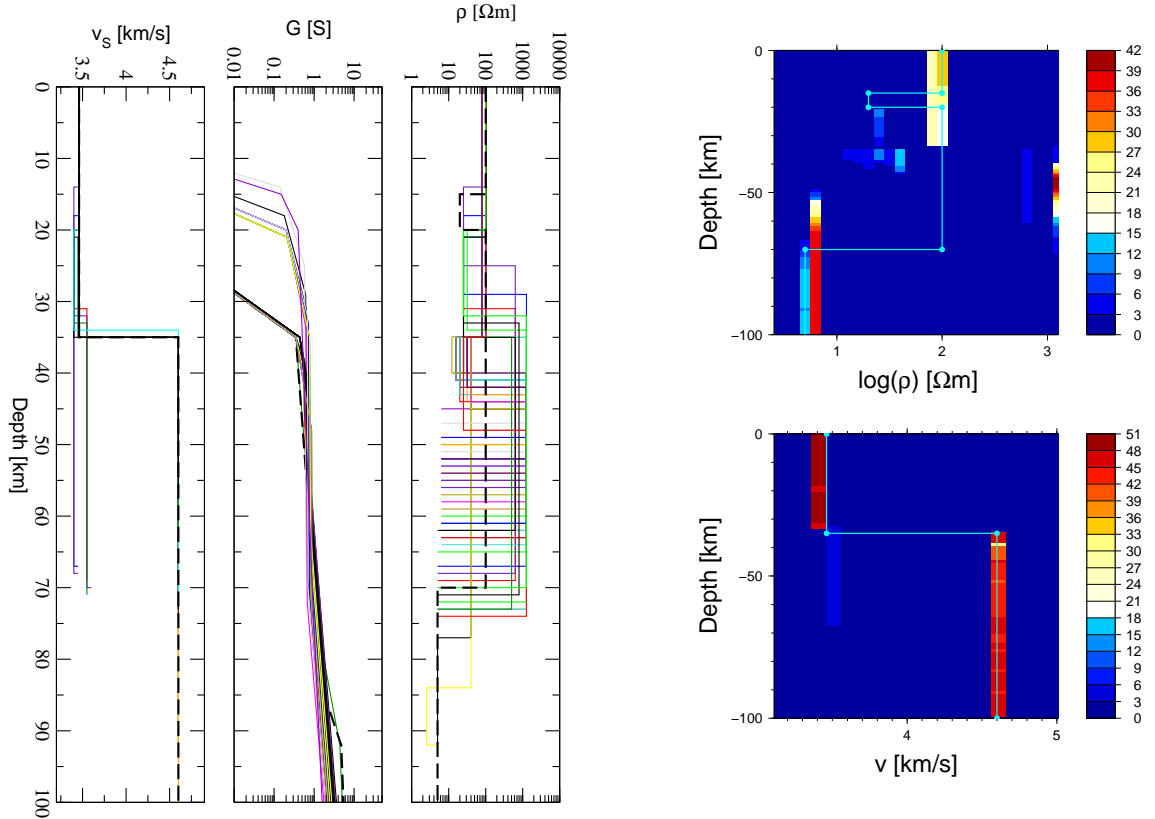
Figure 6.5: Resulting models from a single run for the uncorrelated test case. On the left hand side we plot all individual models for one parameter together in a single plot. The input model is shown as a dashed line. In addition to the inversion parameters v_s and ρ, we also plot the conductance G for the electrical parameters. On the right hand side we plot a 2D histogram that shows the number of models for a given parameter pair. The input models are shown as lines.

RMS for most models is below 1. There is a small number of models however that do not contain the discontinuity. These are the models that have their layer boundaries at the resistivity discontinuities. Due to the small number of layers, the algorithm has to choose between putting the layer interfaces at the seismic boundaries or electrical boundaries.

For the electrical models the picture is less clear. While the near surface and deep resistivity values are correctly reproduced, even without fixing any resistivities, we observe a large number of different boundaries at intermediate depths. Only a small number of models reproduces the first conductive layer correctly. As we speculated before, a number of models shows a smeared representation of that layer. Instead of a thin and highly conductive layer, we observe a much thicker and less conductive layer whose lower boundary coincides with the seismic

Comparison of pareto fronts
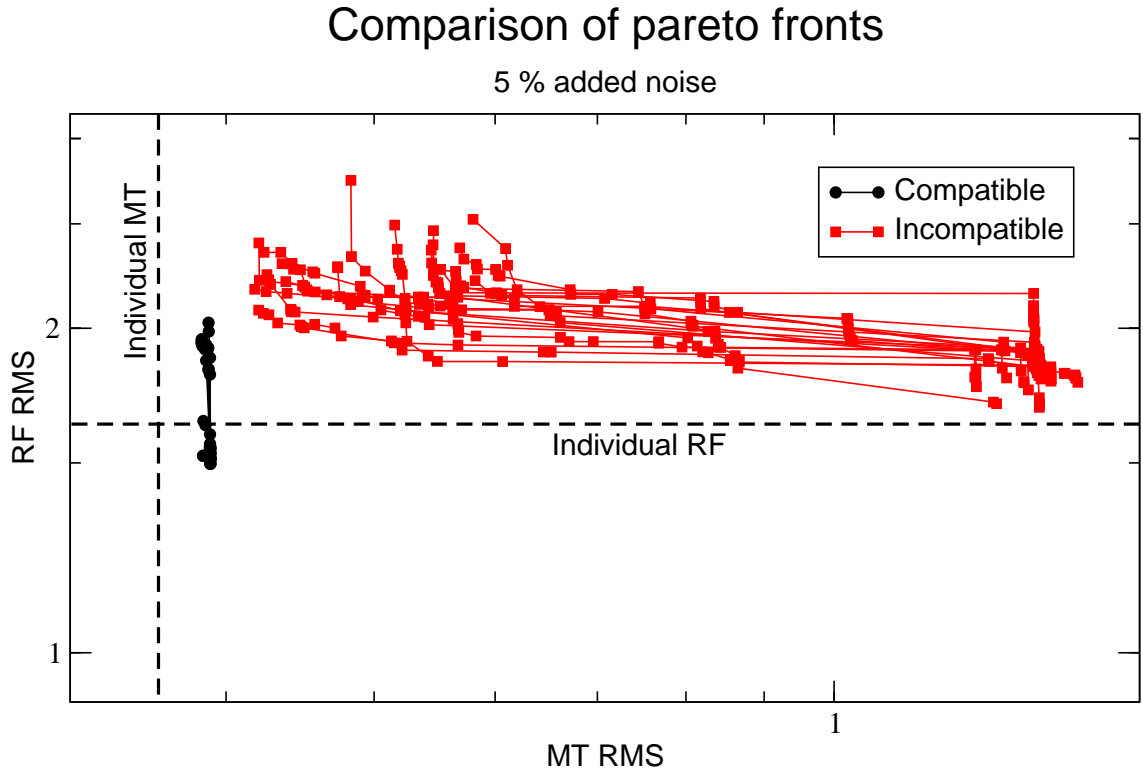
5 % added noise

Figure 6.6: Comparison of the Pareto-fronts for 20 runs on the compatible and incompatible test problem. Both MT and RF data had 5% random noise added.

discontinuity. This way the inversion algorithm can reasonably reproduce both datasets. In other cases this layer is not present in the inversion results. The expression of this layer in the data is small and the gain in terms of misfit is higher if the asymptotic behaviour is reproduced correctly. Virtually all models contain a highly resistive layer at depths of about 50 km. This is another non-uniqueness phenomenon with limited-bandwidth magnetotelluric data. Directly below a conductive layer we cannot resolve the resistivity of a more resistive layer. One way to avoid these artificial structures is to introduce regularization. We will show results from a regularized inversion later. From the plot of conductance with depth for all models we can see that the conductance, in contrast to the resistivities, is well constrained at depths >40 km [Weidelt, 1985].

Adding 5% noise

Observed data are never completely noise free, so we have to examine the impact of noise on our inversion scheme. We therefore add 5% random noise to the real and imaginary parts of the impedance tensor elements, and to the two components
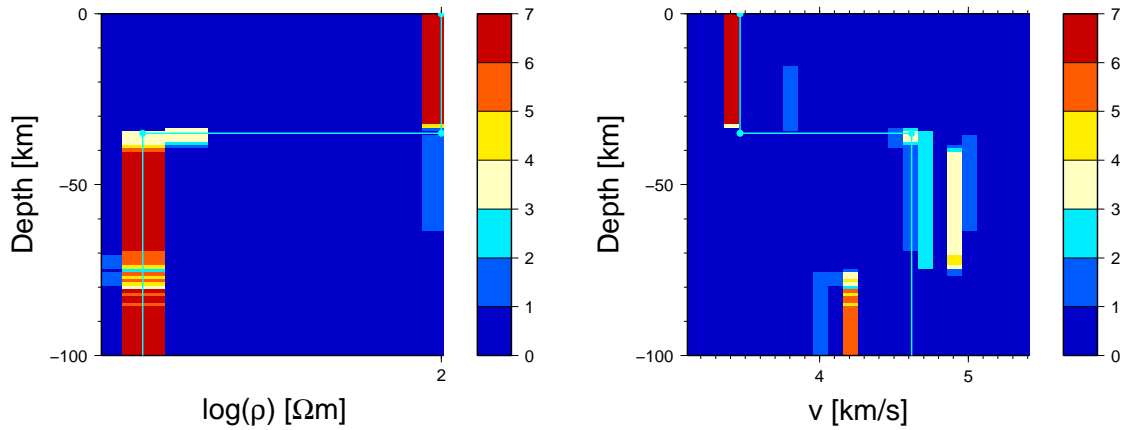
Figure 6.7: 2D histogram of receiver function (left) and MT (right) inversion models for the compatible test case with 5% added noise

of the synthetic seismograms, shown in Figure 6.3. In this case we have used the waterlevel deconvolution algorithm to calculate both receiver functions, to increase the impact of the noise. In the noisy example we can still clearly identify the three positive correlation spikes, but the negative multiple conversion is now buried in noise. We also observe a number of additional local minima and maxima that, at first sight, cannot be distinguished from the original conversions, and we will see how these can add artificial structures in the inversion. This appearance of a receiver function is representative for a fair quality real data example (compare Figure 4.6), so we can expect to encounter similar effects with real data.

Figure 6.6 shows the Pareto fronts for 20 runs of the genetic algorithm on both test problems. In addition we plot the misfit level we achieve when we invert the data individually with a linearized inversion. The general appearance of the graph is similar to Figure 6.4, but there are some important differences. For the compatible input models, we now get a number of output models with identical MT misfit, but varying receiver function misfit in each run. For the incompatible models, the range in both the MT and the receiver function misfit has decreased substantially. Before there was an order of magnitude difference in misfit for both data types, now it is a factor of 5 for the MT misfit, and less than a factor of 2 for the receiver function misfit. Also the separation between the compatible and incompatible results decreased. Still the two groups clearly separate; there is none of the overlap we observed in Figure 5.9, so we can distinguish the two cases.

Figure 6.7 shows the model histograms for the compatible test case. Compared with the noise free results, the receiver function models now contain a larger degree of artificial structure that is not present in the input models. The MT data, in contrast, seems to be much less affected by noise; the increase in scatter is less
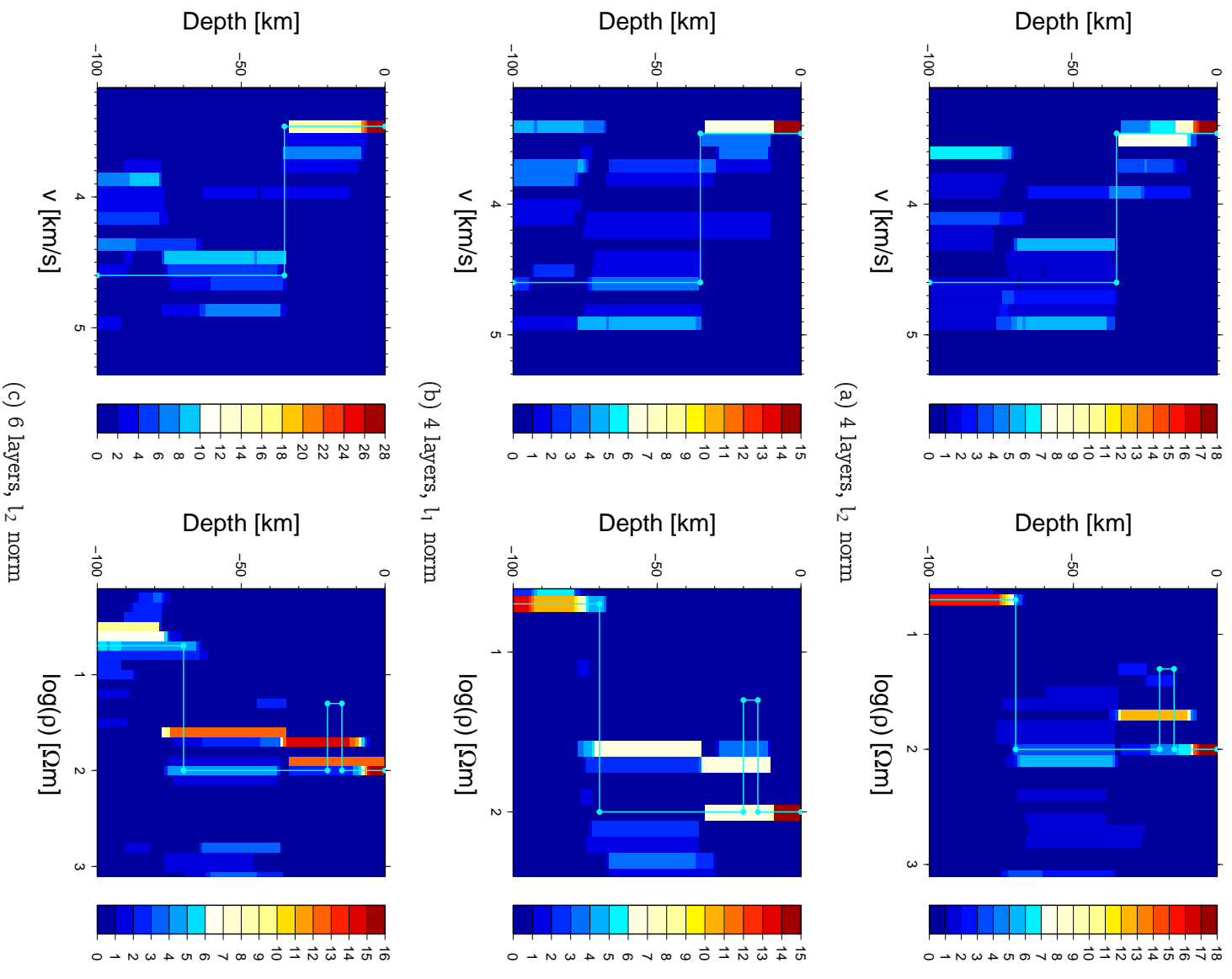
Figure 6.8: 2D histograms of the inversion results for the incompatible test problem with different inversion setting and 5% added noise.

severe. The explanation for this observation is simple. The amplitude changes in receiver functions are created by velocity contrasts at layer boundaries. As we remarked before, the appearance of the noise resembles the data. Therefore it is easy for the inversion algorithm to reproduce some of the noise by introducing additional layer boundaries with velocity contrasts. The noise for MT, on the other hand, is usually a high frequency phenomenon. Unless the data are systematically biased, only a small number of frequencies is shifted in one direction. The diffusive nature of the induction process distributes the effect of changing the conductivity or position of a layer over several neighbouring frequencies. If we only use a small number of layers, this effect prevents the introduction of strong artifacts due to noise.

Looking at the resulting models for the uncorrelated test case displayed in Figure 6.8 mostly confirms the impression we already gained with the compatible test problem. The shallow and deep structure of the MT model are well represented. Somewhat surprisingly we do not observe the high resistivity layer at intermediate depths. Ironically, this is most likely due to the noise in the receiver function data. As we saw with the correlated case, the receiver function models contain a considerable amount of artificial structure that results from fitting noise. This does not leave sufficient freedom for the algorithm to introduce unresolved structure in the MT models, but requires a concentration on the essential features. Although in this case the artifacts in the seismic model are certainly undesirable, this stabilizing effect is one of the objectives we want to achieve when we perform joint inversion.

A possible solution to suppress the influence of noise, is to use a more robust measure of misfit. Using the l_1-norm, instead of the commonly used l_2-norm, reduces the influence of high amplitude data and makes the inversion less prone to outliers [Farquharson and Oldenburg, 1998]. Our genetic algorithm allows complete flexibility in the kind of norm we employ, and we show the results using an l_1-norm measure of misfit for the receiver function data in Figure 6.8b. Otherwise all inversion parameters are kept the same. There are some discernable differences when comparing the distribution of models to the l_2-norm results, but overall the difference is only minor. We still observe artifact structures, although more models now correctly reproduce the velocity change at 35 km.

Finally we examine the effect of increasing the number of layers to 6. The overall effect is only minor, as we can see in Figure 6.8c. The trade-off curves (not shown) have a similar appearance to the ones shown in Figure 6.6 for 4 inversion layers. As can be expected, there is a larger amount of artificial structure in the receiver function models, as the algorithm now has more freedom to model noise contributions. In contrast, the MT model is not greatly affected by the change of inversion layers, its overall appearance is comparable to the 4 layer case.
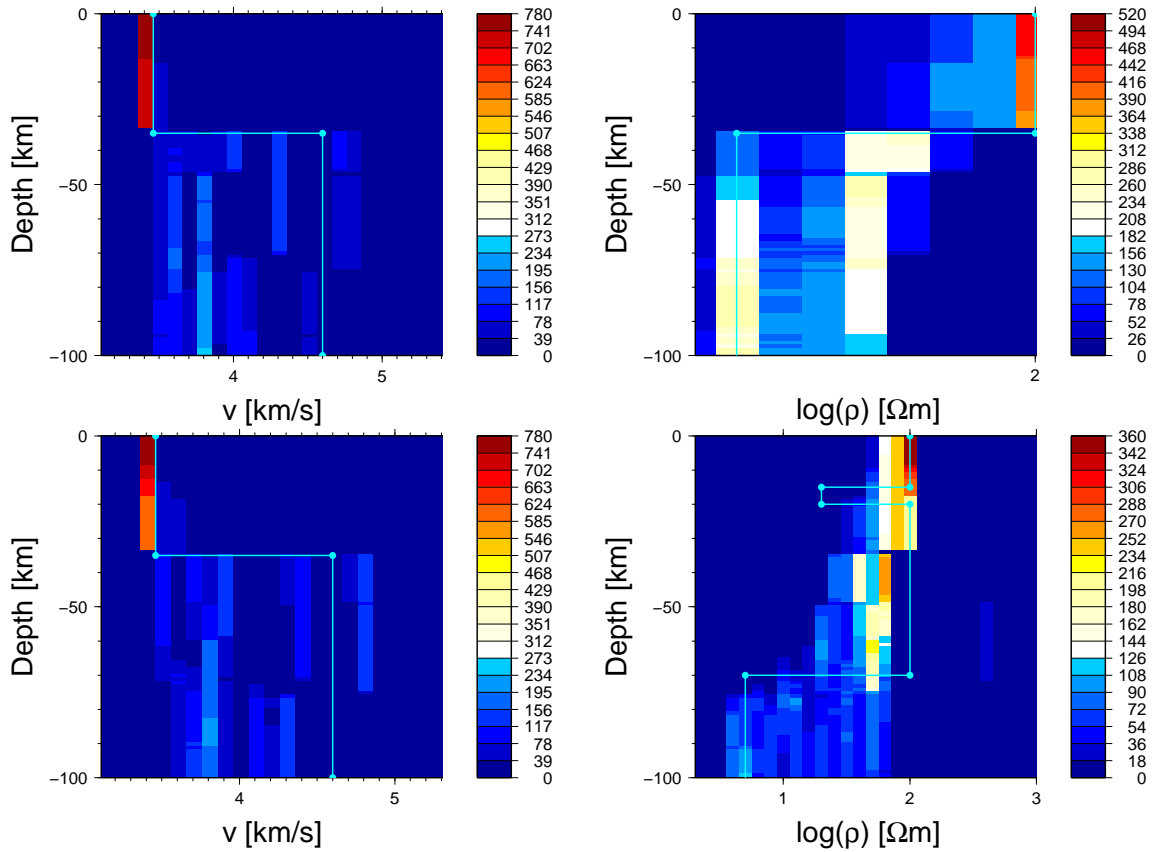
Figure 6.9: 2D histogram of receiver function (left) and MT (right) inversion models for the incompatible (top) and compatible (bottom) test cases with 5% added noise, and regularization based on the model gradient. The lines mark the input models.

Adding regularization

As we remarked before, adding regularization as a third objective function complicates the analysis of the Pareto-front, which, so far, has proven to be a reliable indicator of the compatibility of the seismic and electric models. The appearance of artificial structures in the seismic models, on the other hand, is a serious issue when we invert and interpret measured data, so the added complexity might be a small price to pay. We will test two types of regularization, with or without a reference model. Without a reference model we will minimize the model gradient, as given by Equation 5.29. With a reference model, we minimize the deviation from that model, given by Equation 5.28. The calculation of deviation from the reference model is only based on seismic velocities and layer thicknesses however. Typical variations in seismic velocities are on the order of a few percent and good global reference models exist that summarize the average structure of the Earth.
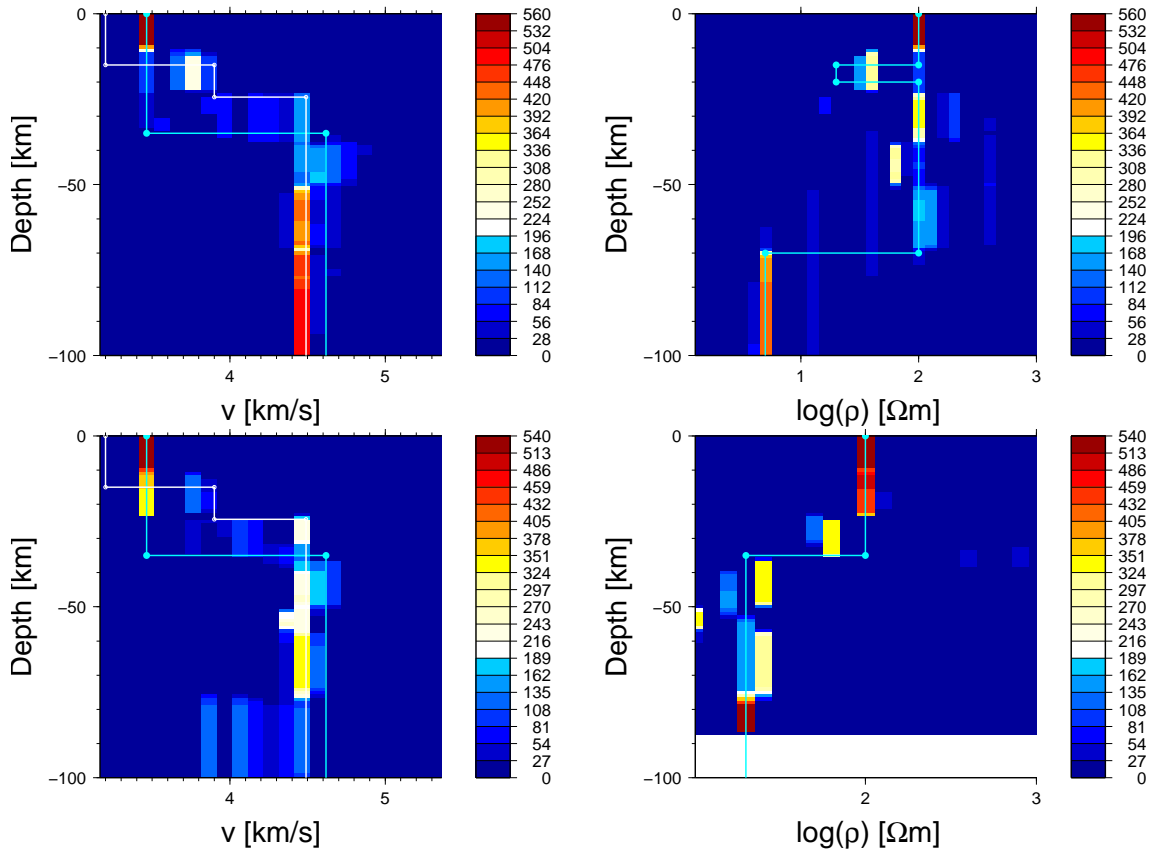
Figure 6.10: 2D histograms of receiver function (left) and MT (right) inversion models for the compatible (top) and incompatible (bottom) test cases with 5% added noise, and regularization based on difference to PREM. The turquoise line marks the input model, while the white line shows the reference model for the seismic parameters.

For MT, the variations are often several orders of magnitude, and no global reference models exist. Therefore, we do not have a good reference point which we can use for regularization of the electric models.

For model gradient based regularization, we obtain the models shown in Figure 6.9. For the MT models, the effect of regularization is exactly as we would expect. Instead of a sharp contrast, we get a number of models that have a gradual change of conductivity between the layers. Being an inductive method, MT only has limited resolution to the conductivity gradient. Therefore regularization facilitates the selection of models with gradual changes and the impact on the misfit is only minor. For the receiver function model the picture is similar, but we also observe now that a number of models shows no velocity contrast at all. This is certainly a case of over-regularization. We have to bear in mind though that these plots

summarize all models regardless of regularization level.

Comparing these histograms with the histograms in Figure 6.10, displaying the results for reference model regularization, we see a number of differences. The MT models appear to reflect the subsurface structure better than with the gradient based regularization. It is important to note though that these models have only been regularized indirectly. The difference to the reference model is only calculated based on layer thickness and seismic velocities, the electrical conductivities play no role in the regularization. The constraints on thicknesses on the other hand, limit possible model geometries and therefore, indirectly, limit the range of possible conductivity values.

The receiver function models also appear to resemble the input model more than with the previous approach. When comparing these models with the reference model used for regularization, we discover that the results are mostly influenced by this reference model. The fact that the seismic velocities only vary by a few percent and that depth to the Moho only varies by ± 15 km in most continental areas, makes it appear on this scale that we recovered the input model. Again these plots only summarize all models the inversion algorithm retrieved, but do not take into account the misfit and other characteristics of any of those models. We have to analyze the trade-off between the different criteria to identify the optimum model.

As mentioned above the trade-off is now represented by a complicated-shaped surface that can only be faithfully represented in three-dimensional plots. Experiments with different viewing angles and plotting styles showed that, while well visible on the computer screen, the printed versions do not provide much insight into the shape of the surface, as the sense for the third dimension gets lost. We therefore decided to represent the model roughness by colour, interpolated between the models and with added contours. While this plotting style has the disadvantage of suggesting values where we do not have any data (best seen in the upper right plot of Figure 6.11), it gives the best sense of the trade-off characteristics of the different criteria. To show in which region the colour is appropriate we also plot the position of the models as black circles.

Figure 6.11 shows plots for the four setups we considered so far. For the correlated case with regularization based on the gradient, we observe a cloud of models with varying misfit values for MT and receiver functions, roughly distributed within a rectangular area. The coloured contours reveal that the roughness values associated with each model group in bands more or less parallel to the abscissa. We find the highest roughness values with the lowest MT and RF misfit values and vice versa, just as we would expect. Within each range of roughness values, i.e., each regularization level, we can find a small cluster of models with low misfit
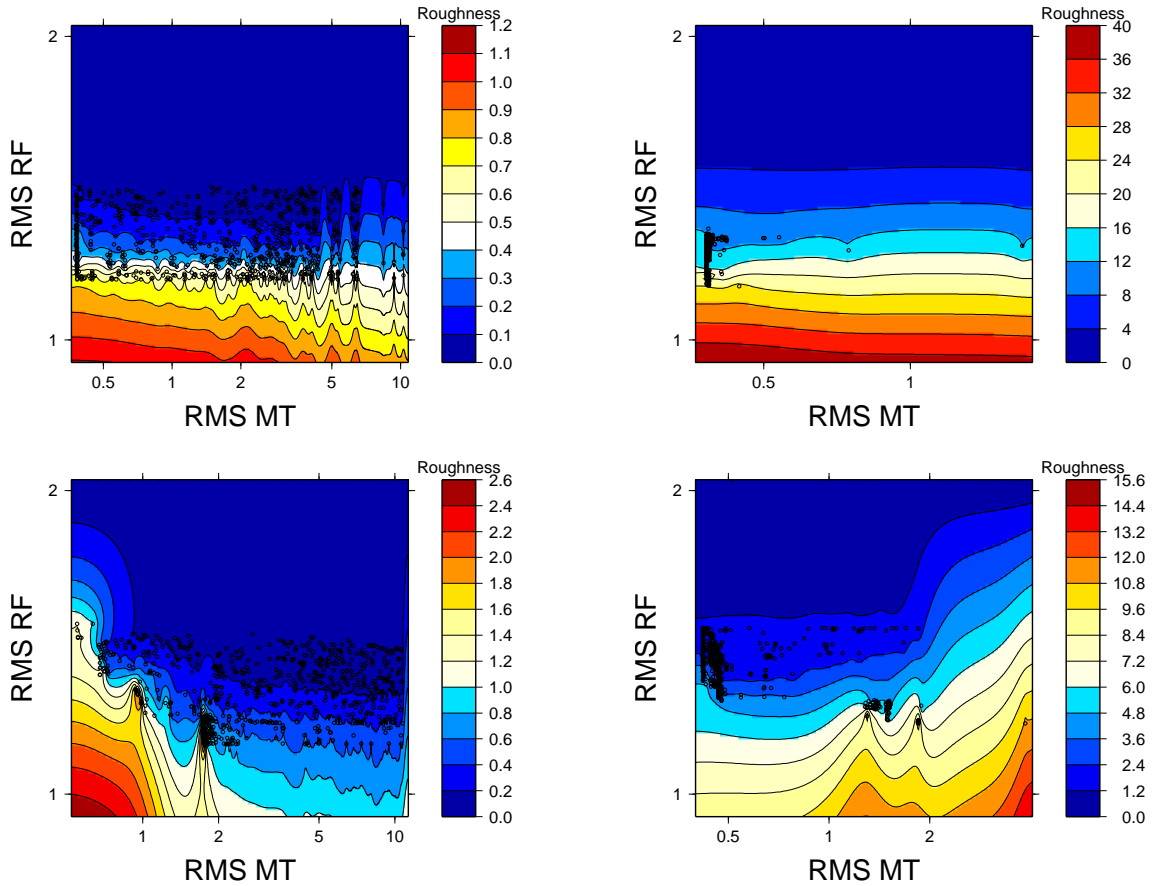
Figure 6.11: Trade-off plots for the two test scenarios with different types of regularization. The upper row shows the results for the compatible input data with gradient based regularization (left) and regularization with a reference model (right). The lower row shows the same for the incompatible input data.

values for both datasets. These are the optimum models for that regularization level.

When we change the type of regularization and use a reference model the picture changes in some sense but also retains some similar features. The largest change we observe, is that the cloud of models collapses to a line parallel to the y-axis. The receiver function misfit changes, while the MT misfit stays nearly constant. As mentioned before, the electrical model does not have a direct impact on the regularization value. Hence whenever the algorithm finds a model with lower MT misfit for a given RF misfit, the new model dominates the old one and replaces it. We observe a few cases where the models do not plot along the main line. These are most likely cases where the algorithm did not converge fully to the optimal solution. Again we observe bands of similar roughness values, although
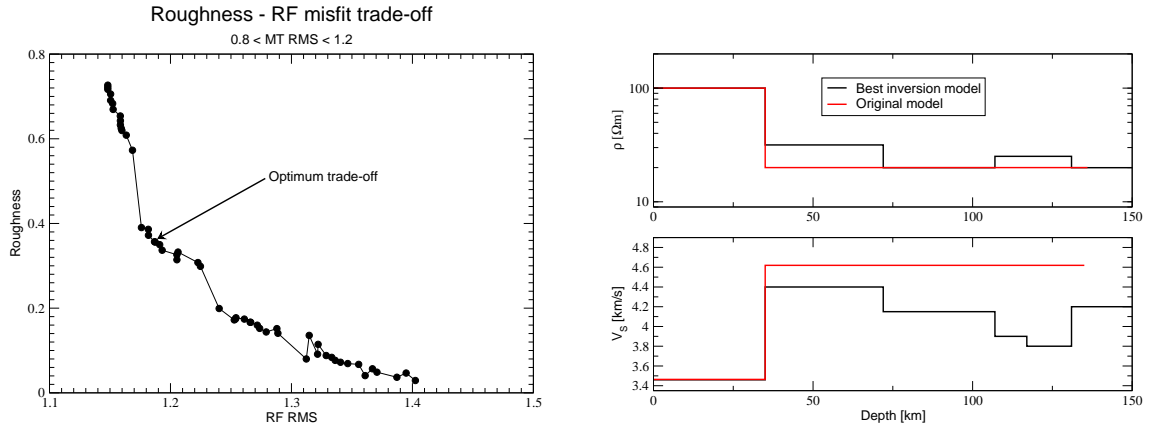
Figure 6.12: RF misfit – roughness trade-off curve for an MT misfit of $0.8-1.2$ (left) and the chosen best models based on the maximum curvature criterion (right).

this appearance is mostly due to the surface plotting algorithm. Still, within the main line of models, we can choose a regularization level and find an optimum model.

When we look at the incompatible test cases, we cannot find a single optimum model for a given regularization level any more. Particularly for high roughness values, i.e., low regularization, we get separate clusters of models with the same regularization value, but either a low MT misfit or a low RF misfit. This situation is analogous to the behaviour of the trade-off without any regularization. We do not obtain any models that minimize receiver function and MT data simultaneously. The rectangular area that we observed for the compatible case is now missing its lower left corner. The differences between the two types of regularization for the incompatible case are similar to the compatible case. When regularized with respect to a reference model, the models concentrate in a linear feature. Only now this line is offset to higher MT misfit values for low RF misfit values, as we would expect.

As an example we show how to find the best model for the correlated test case with gradient based regularization. From the full trade-off plot we select all models with an MT RMS between 0.8 and 1.2. By choosing an RMS of about 1 we avoid fitting noise contributions. For these models we can now plot the trade-off between receiver function misfit and roughness as shown in Figure 6.12. The plot resembles the trade-off plot for a single objective regularized problem, although, due to the projection of a range of MT misfit values, not all models are Pareto-optimal in the space spanned by the two objective functions. Still, we can pick an optimum model by choosing the point of maximum curvature [Hansen, 1992]. This point is marked in the trade-off plot and the corresponding models are shown on the right
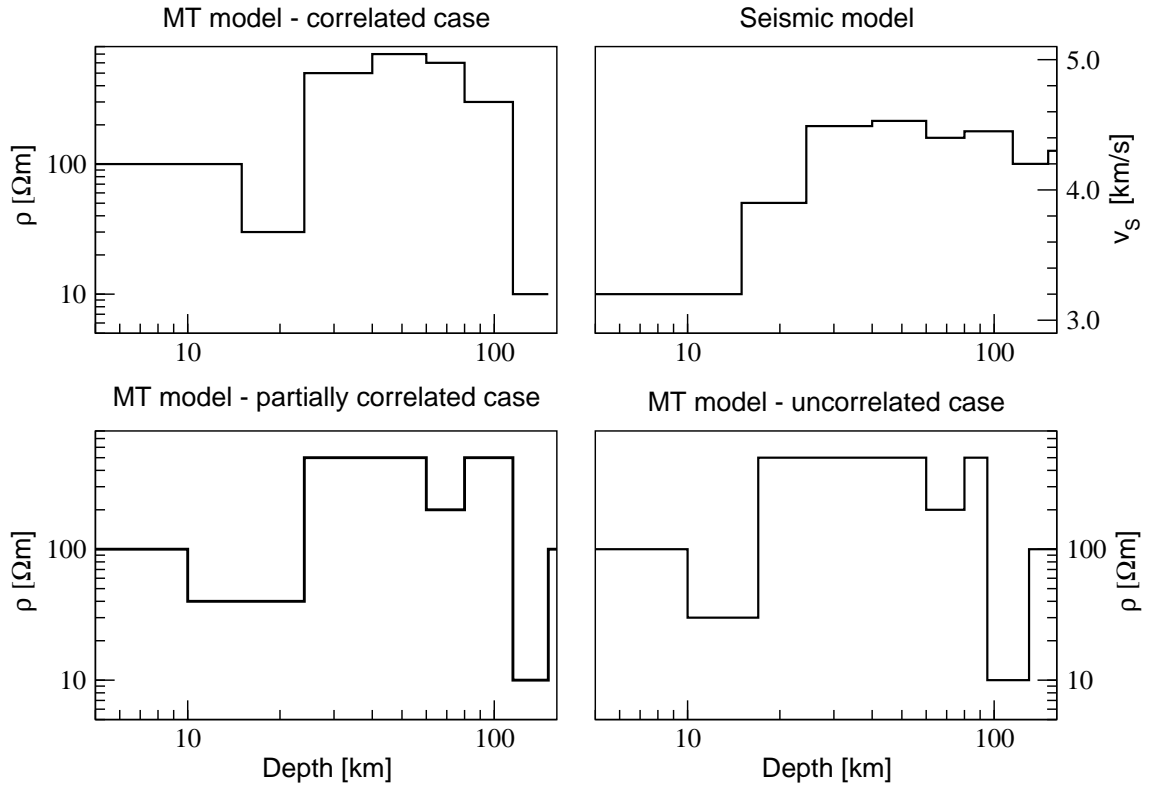
Figure 6.13: The input models used to create our three realistic test cases. In all cases the seismic model (upper right) remains constant. Only the MT changes to create a compatible (upper left), partially compatible (lower left) and incompatible (lower right) test case.

hand side of Figure 6.12.

We observe that the MT model agrees well with the input model. The layer is reproduced correctly and we observe only minor fluctuations around the true model in the lower part. This suggests that the chosen level of regularization is correct for the MT data. The receiver function model does not fit as well. Although the depth to the Moho is correctly recovered, the velocities in the lower part are systematically too low. In addition we observe a number of artificial low velocity layers with deviations that are substantial for seismic data. Comparison with the other seismic models along the trade-off curve shows that this is a consistent feature caused by the noise in the receiver function.
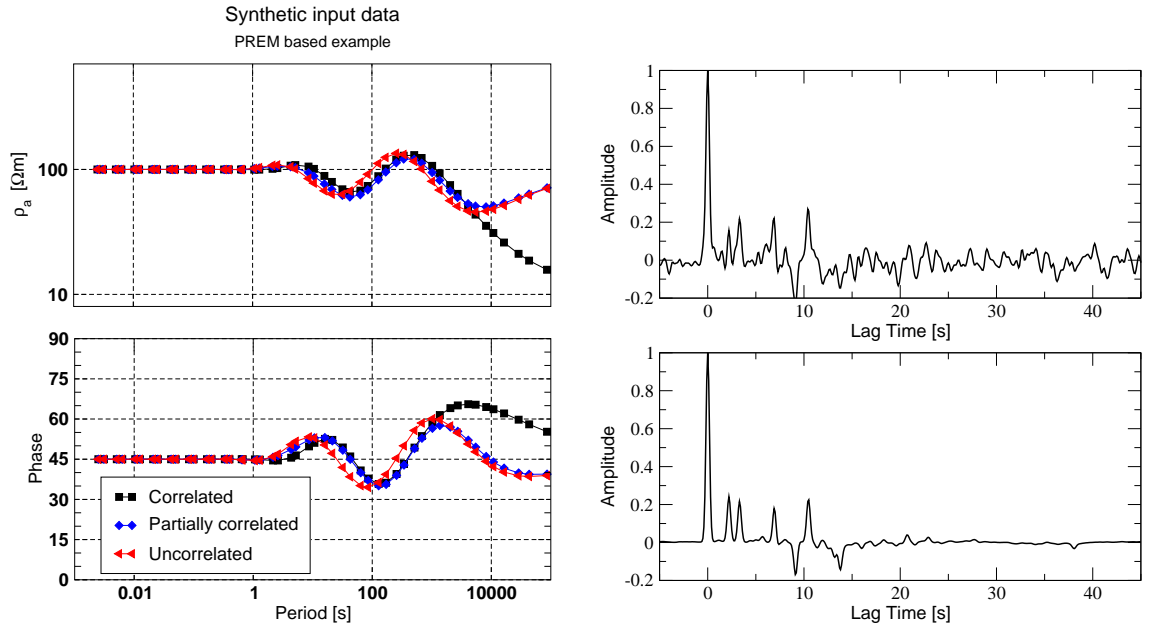
Figure 6.14: Synthetic input data generated from the models in Figure 6.13. We only display the noise free MT data (left), but show the receiver function data with 5% added noise (top right) and noise free (bottom right).

6.5 A more realistic model

We now look at a more representative test case. The two-layered model showed the potential of our approach, but the situation we can expect to find in the subsurface is certainly more complicated than that. Figure 6.13 shows the models we use to test the algorithm in this section. In addition to the two scenarios we tested in the last section, we add a partially correlated test case. In this case the Moho and lithosphere-asthenosphere interfaces correlate, but some other interfaces do not. The seismic model is based on PREM with some small velocity perturbations in the upper mantle, and is the same in all three cases. We vary the MT model to create varying degrees of correlation. Figure 6.14 shows the synthetic data that we create from these models.

There are some visible differences between the three MT datasets both in apparent resistivity and phase, but in the period range of $1 - 500$ s, these differences are as small as we can expect to detect with high quality data. Thus the three test scenarios will be much harder to distinguish than the simple examples before. Furthermore we cannot expect to resolve the small velocity and conductivity changes in the sub-Moho layers. We will examine these cases in the same way as we did in the previous section. First, we use noise free data to see what is
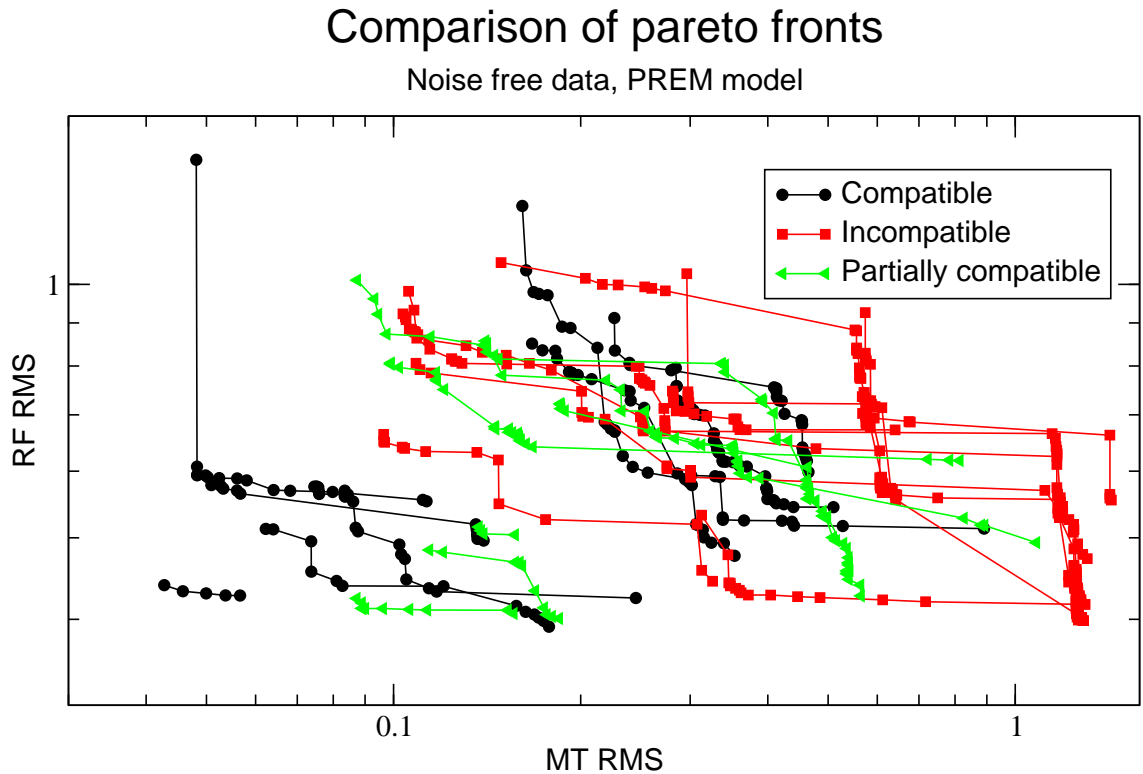
Figure 6.15: Comparison of the Pareto-fronts for 10 runs on the compatible, partially compatible and incompatible test problems.

theoretically possible, then we add random noise to simulate observed data.

Noise free data

The parameters we use for the inversion are nearly the same as for the two-layer test case. We only increase the number of inversion layers to 12, in order to give the algorithm more flexibility. The comparison of the Pareto-fronts from different runs of the genetic algorithm on the three test cases, shown in Figure 6.15, asserts the difficulty to distinguish those cases. In contrast to the two-layer test case, the fronts for the compatible and incompatible problems do not separate clearly. Some of the Pareto-front for the compatible test problem plot in the region of the incompatible results and their appearance is similar. Still, when looking at all the fronts, the compatible test problem gives lower misfits for both datasets and the fronts contain less models. As we would expect, the partially correlated test case plots between the two extremal cases and the appearance of the front is also mixed. Some fronts contain only a small number of models and achieve misfits comparable to the compatible problem, while others resemble the incompatible
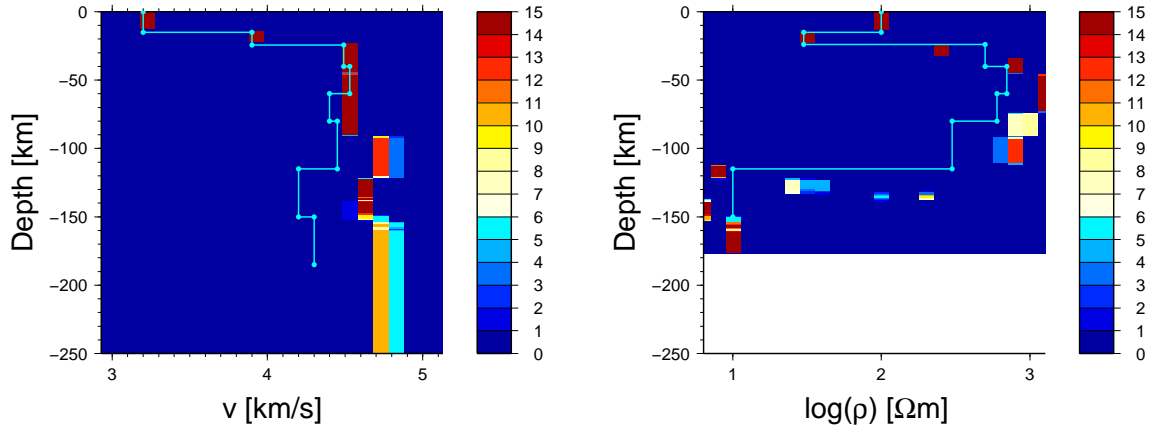
Figure 6.16: 2D histogram of receiver function (left) and MT (right) inversion models for the compatible test case for noise free data.

problem.

The model histograms for the compatible case in Figure 6.16 give an overview of the type of models we retrieve. For both seismic and electric models we obtain parameter values close to the true value for the upper layers. The resistive layers below the Moho and the small velocity variations in the same depth region are not well reproduced, as we discussed before. It is interesting to note though that most of the electric and seismic models reproduce the changes between the layers fairly well. For the deeper parts of the model, the electric inversion models approach the value of the half-space, while the seismic models overestimate the velocity at depth. The reason for this behaviour is that the receiver function data, as explained before, is mostly sensitive to changes in seismic velocities. If the velocity change is modelled incorrectly at one interface, subsequent interfaces will still reproduce the relative changes correctly,but the seismic velocities remain biased. Because of the velocity–thickness trade-off of receiver function data, this will also shift the position of the interfaces. This is exactly what we observe in this plot.

Adding 5% noise

In analogy in the two layer test case, we add 5% random noise to both datasets in order to simulate measured data. The resulting trade-off curve is shown in Figure 6.17. As we expect, the noise makes it even more difficult to distinguish the three different compatibility situations in the trade-off plot. Now all fronts comprise a number of models regardless of whether the input data was produced by compatible, incompatible or partially compatible models. Also the level of

Comparison of pareto fronts
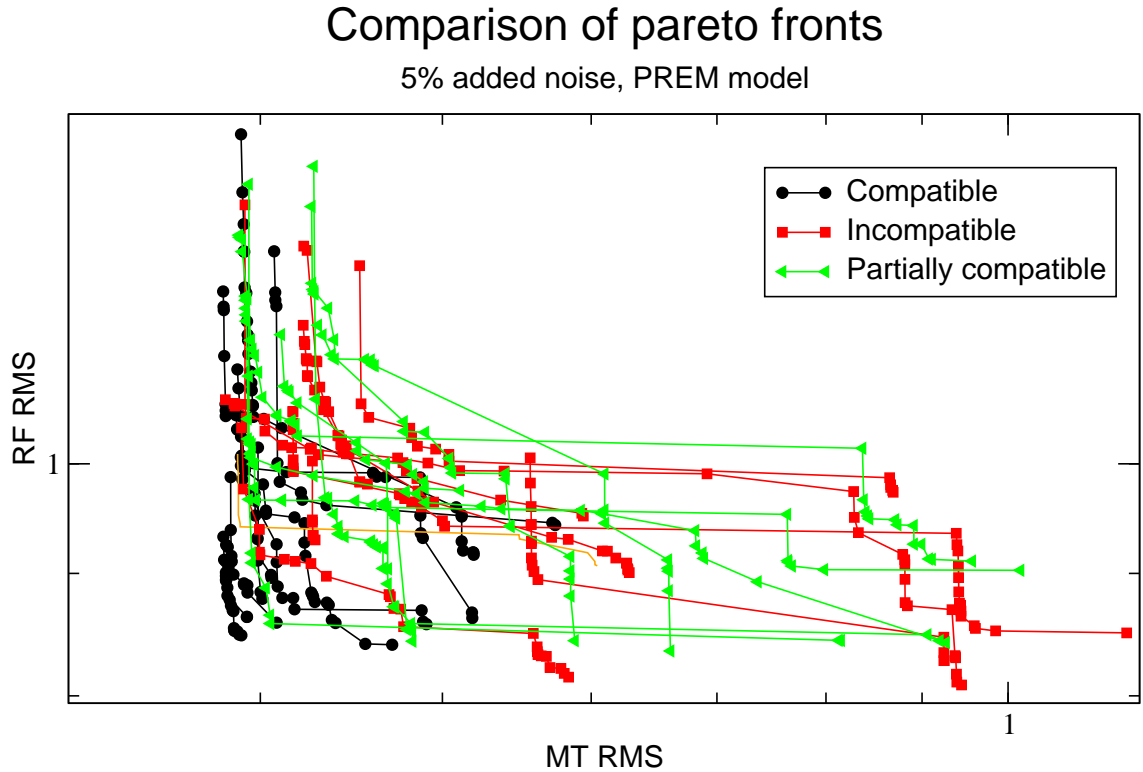
5% added noise, PREM model

Figure 6.17: Comparison of the Pareto-fronts for 10 runs on the compatible, partially compatible and incompatible test problems.

misfit that we achieve is similar for all three test cases. There is one visible difference between the compatible and incompatible cases, however. Apart from three runs, all compatible Pareto-fronts show a near vertical trade-off. Within this trade-off curve it is easy to pick the best model, because a substantial gain in receiver function misfit comes at a small decrease in MT misfit. In addition, all MT misfits are below an RMS of 1. For the incompatible case, most trade-off curves show a much larger decrease in MT misfit and are missing the lower left corner. There is one exceptional run though, where we simultaneously achieve low MT and receiver function misfit. Even when we increase the population size by a factor of 2 and the number of iterations to 150, these observations remain essentially the same.

The histograms in Figure 6.18 summarize the inversion results for a single run on each of the test cases. The general observations agree with what we already observed in the two layer case. Only now, due to the increased number of inversion layers and model complexity, the effect of noise is even more pronounced. Overall, the receiver function models are more strongly affected by noise. Where the MT models, at least in a general sense, reproduce the feature of the input model,

(a) Compatible test case

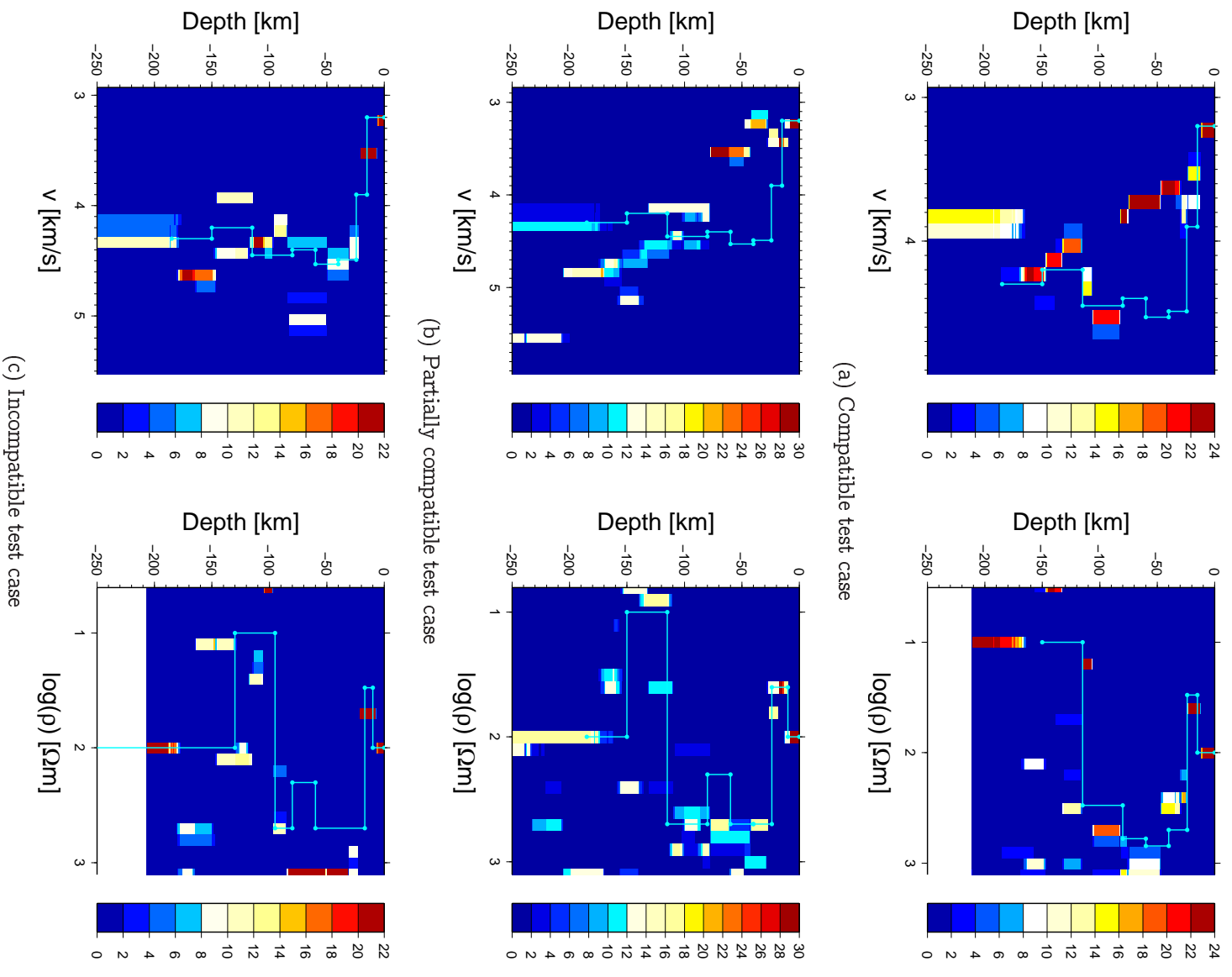(b) Partially compatible test case

(c) Incompatible test case

Figure 6.18: 2D histograms of the inversion results for the three different test problems with 5% added noise.
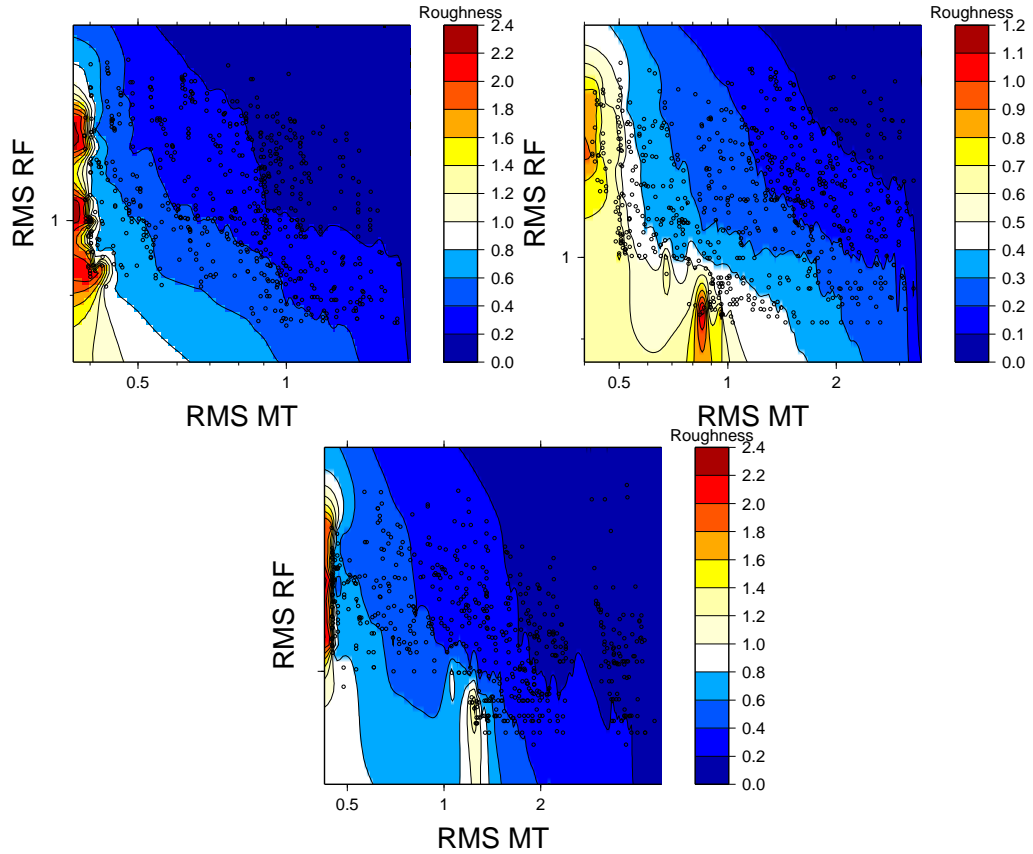
Figure 6.19: Trade-off plots for the three test scenarios with regularization based on the model gradient. The upper row shows the results for the compatible input data (left) and partially compatible data (right). The lower row shows the same for the incompatible input data.

the receiver function models show a high degree of artifacts in the upper part of the model. At depths between 100 and 170 km the majority of the models are reasonably accurate. It is puzzling though that the velocity of the lowermost layer is severely biased downwards for the compatible models. Even when considering the misfit of the receiver function data and selecting the best model (not shown), we observe the same characteristics as in the summary histogram. It is therefore important to make sure that we use high-quality receiver function data when we apply the inversion algorithm to real data. Also a good regional seismic model can be used as a reference model to suppress some of the artifacts as we showed for the two layer case before. We will now see in how far regularization helps in this case.
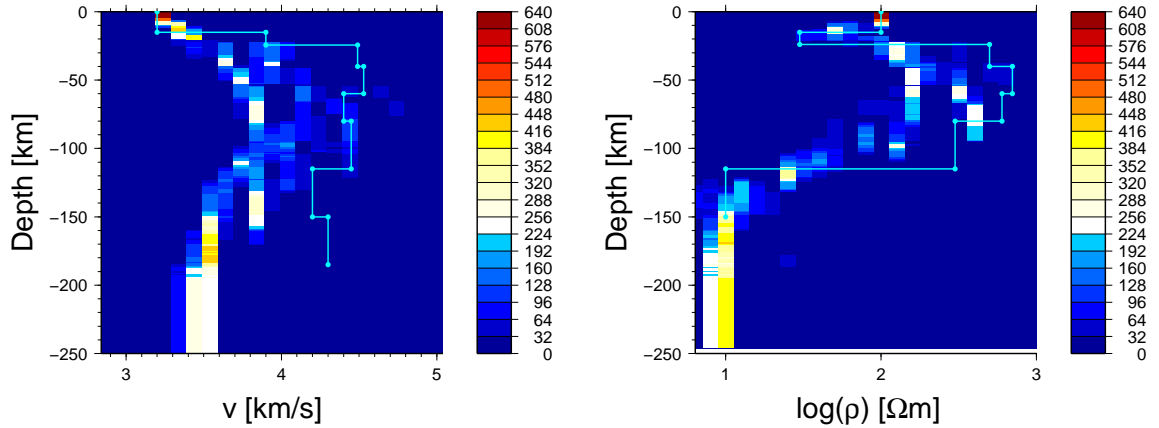
Figure 6.20: Histograms of the inversion results for the compatible test problem with gradient based regularization. The electric models (right) seem to represent the input better than the seismic models (left).

Adding regularization

The trade-off plots for the regularized runs in Figure 6.19 overall have a similar appearance as for the two-layer case. Only now the MT model seems to have the major impact on the model roughness measure. For the correlated case, each band of similar roughness plots roughly parallel to the ordinate and for high roughness values it is easy to pick the optimum model. For low roughness values, i.e., smooth models, this becomes increasingly difficult, mainly because these models are so smooth that they do not predict the data any more. For the partially correlated and uncorrelated case we observe a similar phenomenon as for the uncorrelated two-layer case. The vertical band of models with similar roughness values is offset to higher MT misfit values for low receiver function misfit values. As the two-dimensional plots before, the trade-off curve is missing its lower left corner. From these plots, however, it is not possible to distinguish the partially compatible from the incompatible case.

The model histograms in Figure 6.20 show the expected picture for the magnetotelluric models. The sharp boundaries are approximated by a succession of layers with small resistivity changes similar to the model generated by the OCCAM code [Constable et al., 1987]. The position of the transitions between resistive and conductive structures is reproduced correctly, although we do not recover the full amplitude of the resistive layer. This can be expected though, as the method does not have good resolution to this resistivity value. The results for the seismic models are somewhat disappointing. It is understandable that the strong change in velocity is now damped in the inversion models, it is unclear however why the
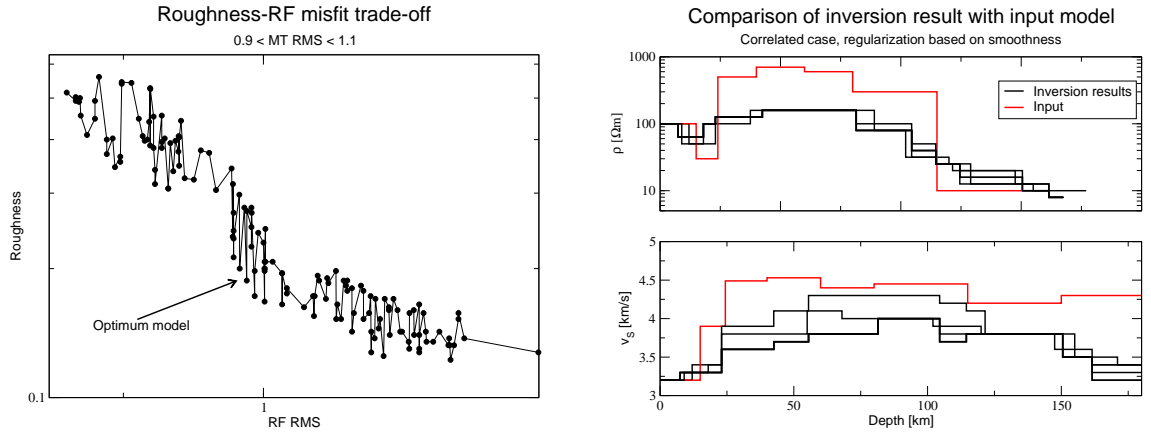
Figure 6.21: Receiver function misfit roughness trade-off for models with an MT misfit of ≈ 1 (left). We mark the optimum model based on the curvature of the trade-off curve and plot this model and the neighbouring models in the misfit plot together with the input model (right).

relatively small changes at depth between 50 and 200 km are replaced by a negative velocity gradient. One would expect that the regularization would prefer a uniform seismic velocity and the data misfit functional would cause the inversion to add interfaces where the data require it. This seems not to be the case for the majority of models here. We have to bear in mind again that these plots sum up all models regardless of data misfit and roughness. We will therefore follow the recipe laid out in the previous example to identify an optimum model and see how this model performs.

Figure 6.21 shows the trade-off between receiver function misfit and roughness for all models with an MT RMS of ≈ 1. In comparison with the two-layer case, we see that the trade-off curve now exhibits considerable scatter between adjacent models. Still, we can identify the general L-shape and we pick one of the models from the area of maximum curvature as our optimum model. We show the electric and seismic parameters on the right hand side of Figure 6.21 together with the input models. The inversion results mainly reflect the characteristics we already observed in the histogram. Still, the seismic velocity is not representative of the original structure and we observe the curious negative velocity gradient at large depths.

We perform a sensitivity study in order to identify the reason for this gradient. Figure 6.22 shows a comparison between the original data with noise, the data generated by our preferred model, data generated from the preferred model, but with a constant velocity at depth, and the input data without noise. We have marked the minima that result from the gradient with arrows. The comparison
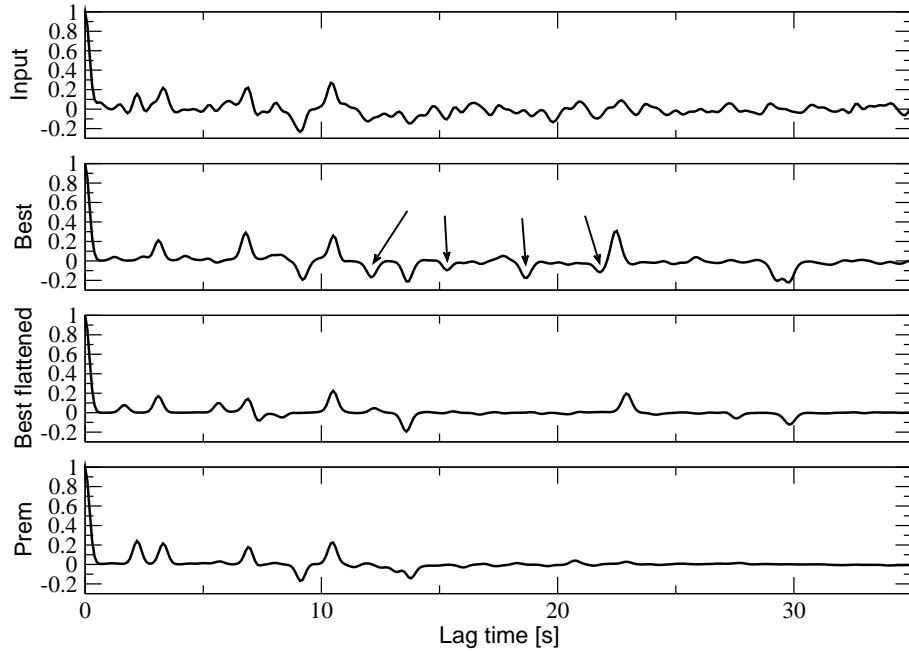
Figure 6.22: Comparison of receiver function data from different types of model. From top to bottom: Input model with added noise, inversion result, inversion result with constant velocity at depth, and input without added noise.

with the noisy input data reveals that most of the minima have a corresponding minimum in the inversion input. As the inversion procedure cannot distinguish noise from data, it adjusts the model structure to accommodate this feature. These results reinforce our previous statement that it is essential to use high-quality receiver function data in the inversion, and explains why so few receiver functions are modelled but mainly used as an imaging technique.

So far we have not discussed regularization with a reference model for this example. We will not show the trade-off plots for this type of regularization, as they resemble the two-layer case as much as the smooth regularization did. We will, however, see whether we can use this approach to find a better model in the correlated case.

The reference model we use is shown, together with the input model and a histogram of the inversion results, in Figure 6.23. Although it reproduces the basic shape of the input model, it does not have any of the small scale features, such as the velocity change in the crust or the variations in the mantle. Obviously this means that a model that resembles the reference model will reproduce some important features of the data, but this is by no means unrealistic, as for some parts of the Earth good regional models exist that can be used for regularization.

The histogram of the EM-models shows that these models reproduce the resis-
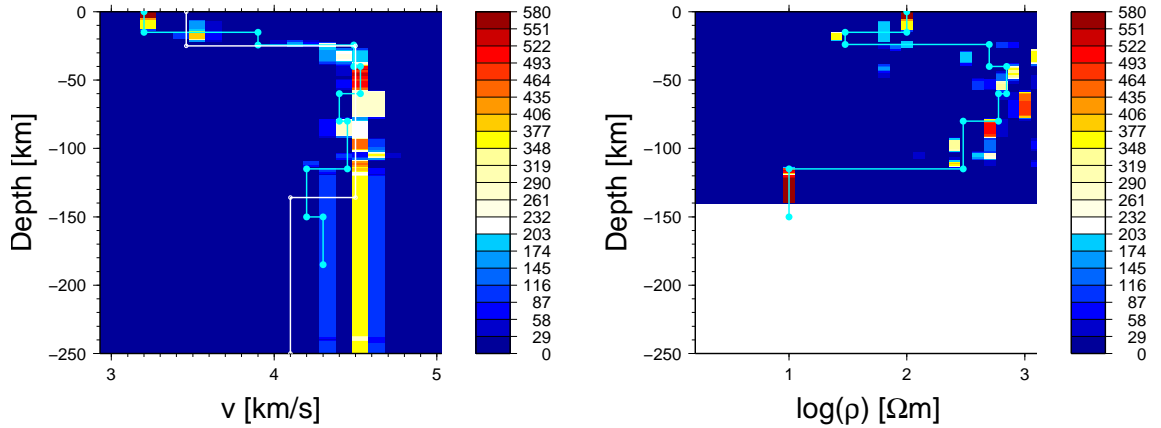
131

Figure 6.23: Histograms of the seismic (left) and electric (right) inversion models for a run with regularization with a reference model. We plot the original model (black) and the reference model (white) together with the seismic inversion results.

tivity structure well. We remarked before that this type of regularization does not have direct influence on this part of the joint model and consequently this plot resembles the results for the unregularized case. Also the seismic model, due to the proximity of the reference model to the input model, now resembles the input model more than before. We recover, although with reduced amplitude, the velocity contrast in the crust, that does not have a counterpart in the reference model. The position of the Moho is similar in reference and input models, and therefore well reproduced by the inversion results. In the upper mantle the situation is not clear, as input and reference do not differ substantially, but we seem to recover some structures there. There is even a small number of models that reproduce the main aspect of the velocity decrease at a depth of 120 km. Somewhat surprisingly the majority of models shows a constant velocity in this depth range, and neither follows the decrease of the input nor the reference model. One possible reason is that the limited number of layers in the inversion does not allow another change at this depth.

Following the same recipe as before to identify the optimum model, we construct the trade-off curve shown in Figure 6.24. As the MT data do not take part in the regularization, the achieved misfit is significantly below 1 for most of the models. We therefore pick a misfit range that represents the best 10% of the models in terms of MT misfit. For low misfit values the trade-off plots as a narrow band that branches out with increasing receiver function misfit. This is an effect of projecting the three-dimensional trade-off points onto a plane. The optimum model, where the curve bends most strongly, still lies in the region where we have only a single branch. The comparison of the resulting inversion model with the
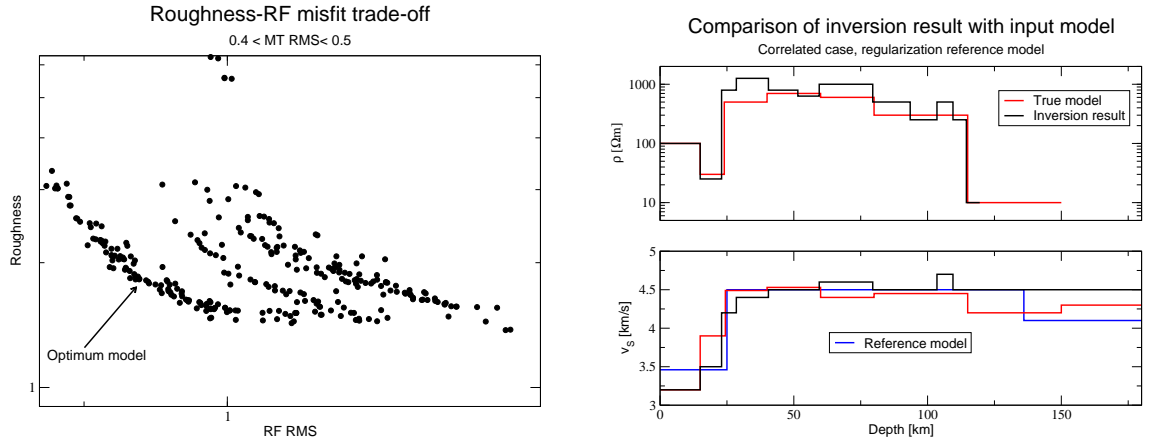
Figure 6.24: Trade-off between receiver function misfit and roughness for the best 10% MT models (left). The optimum model is marked in the trade-off plot and plotted together with the input models and the reference model (right).

input and reference models reveals the strong influence of the reference model on the seismic part of the inversion. Nevertheless we recover some of the features of the input model. It is also interesting to note that the position of the Moho in the inversion result is shallower than in both the reference and the input model. Another interesting effect can be observed at the interface of the layer between 60 and 80 km. The position of the interface is reproduce fairly accurately in both models, but the gradient across the boundary has the opposite sign in both cases. For both datasets this feature only has a minor impact on the data, so in general we cannot expect to find it in our inversion results. The transition to the low-velocity/low-resistivity layer at the bottom of the model is only reproduced by the MT model. The most likely explanation for the failure of the seismic model in this part is the fact that the reference model has a constant velocity in this part and only decreases at greater depths.

In summary these tests indicate that even with relatively complicated and realistic models we get some indicator on the compatibility of the seismic and electromagnetic data. The quality of this indicator, as well as the quality of the inversion models, strongly depends on the level of noise in the data, particularly for the receiver functions. This requires us to use only high-quality data in the inversion. In addition regularization with a good reference model can help to improve the inversion results. This of course carries the risk of enforcing spurious structures if the reference model is not appropriate. We therefore have to compare the resulting models with the output of runs with different reference models or different types of regularization.

One dimensional fool
In a three dimensional world

Ben Harper, Both sides of the gun

7

Application to data from the Slave craton

The synthetic test cases in the previous chapter outlined the strengths and limitations of our approach. We can use the insight we gained from those experiments to assess the quality of the results we obtain from measured data in this chapter. The Slave craton in north-western Canada is a well studied Archean craton and the observed coincidence of seismic interfaces with changes in conductivity [Snyder et al., 2004] was one of the motivating factors for this work. We can reasonably expect at least partial compatibility of the two datasets, and therefore apply our joint inversion method to two sites from this area.

7.1 An overview of the data and selecting suitable sites

Due to its diamoniferous kimberlites and thanks to the LITHOPROBE program the Slave craton has been studied extensively with different kinds of geophysical and geochemical methods [e.g., Griffin et al., 1999, Jones et al., 2003, Snyder et al., 2004, Clowes et al., 2005]. Figure 7.1 shows a selected number of sites located on the central Slave craton around the Ekati diamond mine. These were part of the studies by Jones et al. [2003] and Snyder et al. [2004], and therefore are good potential candidates for our joint inversion approach.

Here we apply the joint inversion algorithm to a small selection of stations, to demonstrate its potential, but do not attempt to construct a comprehensive
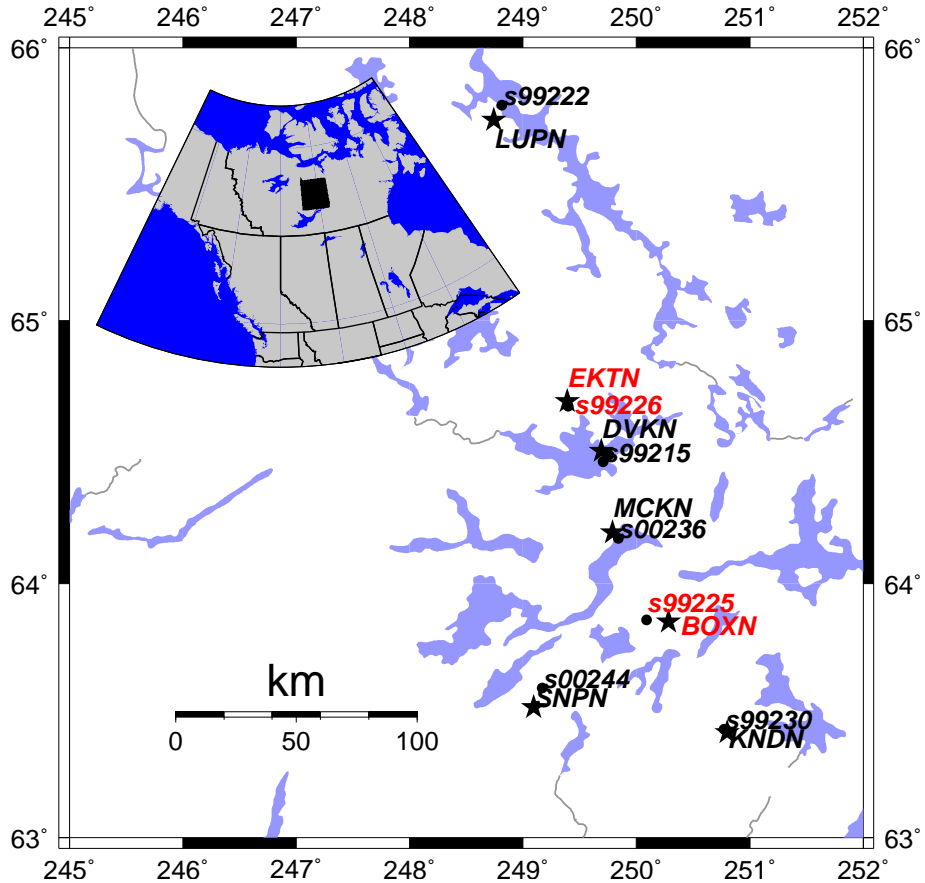
Figure 7.1: Location of some of the seismic stations (stars) and MT sites (dots) in the central Slave Craton. The sites we will discuss in detail are labelled in red. The location of the map area on the American continent is shown in the inset map.

model for the whole Slave craton. For this reason we concentrate on sites where we can expect few problems with the inversion approach. Figure 7.2 shows the magnetotelluric data for two of the sites in Figure 7.1 that appear to comply with our assumption of a layered Earth. In the measurement coordinate system there is no visible phase splitting at high frequencies, and less than 15° splitting at periods > 100 s. Figure 7.3 shows a number of different dimensionality measures as suggested by Martí et al. [2005] and the corresponding threshold values for one-dimensional structures.

The "classical" one-dimensional indicator Σ [Swift, 1967] has values larger than the proposed threshold for all frequencies and both sites. This suggests that a simple layered earth model is not adequate to describe the data. However, Σ is an amplitude based measure of dimensionality, and is very sensitive to near-surface
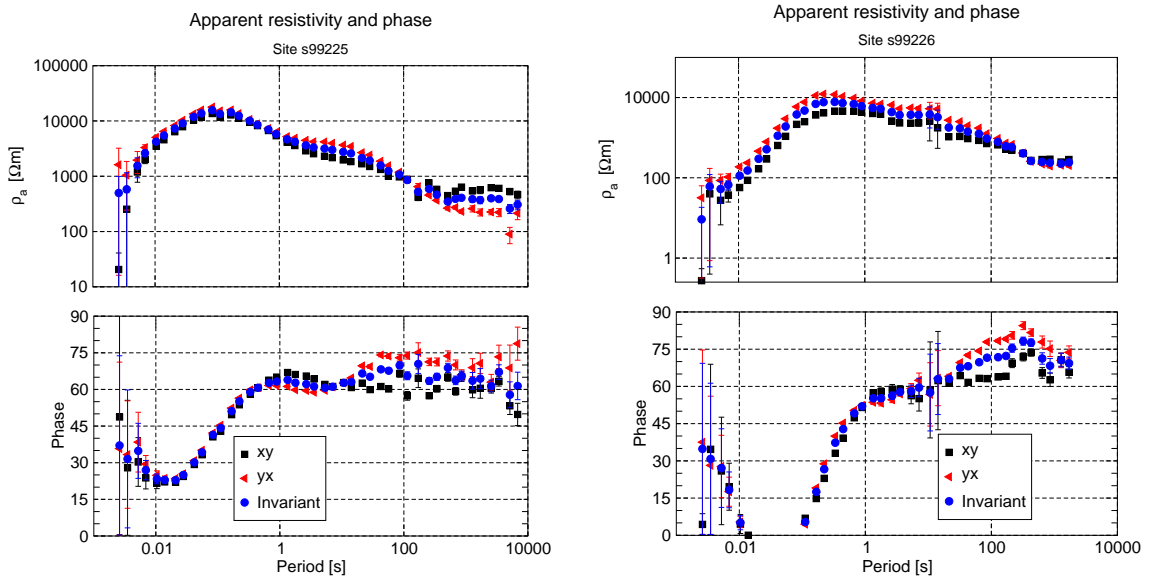
Figure 7.2: Apparent resistivity and phase for the off-diagonal elements and the Berdichevskyi invariant for two sites from the Slave craton.

distortion, as we already discussed in chapter 2. In contrast, the rotationally invariant phase difference μ, is below the threshold for most frequencies at both sites. This indicates a one-dimensional model with local 3D distortion, as proposed by Larsen [1975]. According to Martí et al. [2005] this would also require either η or Q to be below their respective threshold values. Considering the error bars, this is the case for most periods between $0.1 - 100$ s at site s99225 and all but the longest periods at site s99226. As mentioned in the theoretical discussion, we have to bear in mind that these are necessary, but not sufficient conditions. We can therefore conclude that the data admit the Larsen model for most of the period range, but cannot necessarily exclude a more complicated model. The high phase-sensitive skew values at long periods indicate the need for a three dimensional model, so at those periods our inversion results will only have limited validity. We can also see this from a comparison with the 3D model by Jones et al. [2003]. Site s99225 is located close to their site deGras (Figure 9 in Jones et al. [2003]) and we can see the three-dimensional character of the model at depths > 100 km. Still, these two sites show the best agreement with a one dimensional model hypothesis among all considered sites and unless the structure is very complicated, we can expect to retrieve the general characteristics.

For the receiver function analysis we initially inspected all recorded events with a moment magnitude >5 recorded between June 2002 and December 2005, resulting in a total of 598 events. It quickly became apparent that the receiver functions
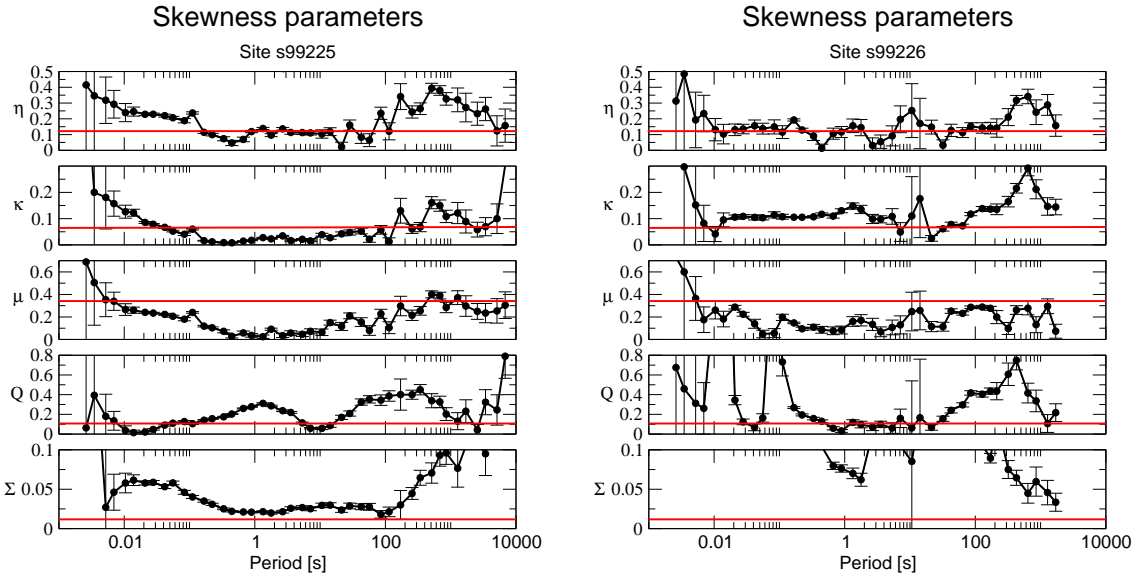
Figure 7.3: Frequency dependence of several dimensionality parameters for the two magnetotelluric sites. The red line in each plot shows the threshold value suggested by Martí et al. [2005].

from events with magnitude < 5.8 and shallow source depths did not satisfy the quality criteria we described on page 76 in chapter 4. Through manual selection we reduced the number to 59 reasonable to high quality events. The receiver functions calculated from these events are shown in Figure 7.4. For each site we show the receiver functions sorted once by backazimuth and once by distance.

In the plots each event occupies the same vertical space and we do not insert gaps to indicate missing backazimuthal or distance coverage. While this results in a highly non-uniform y-axis scale and distorts the moveout of coherent phases, it is the best way to give an overview of the data. The events are clustered in a small number of regions with similar backazimuth and distance, and consequently a plot that would take have a linear backazimuth or distance scale, would mostly be empty. Figure B.1 in the Appendix shows the location of the events in relation to the Slave craton.

Overall, we observe a good agreement of general features between most receiver functions. At both sites we only observe some receiver functions that either do not show the zero correlation peak, or the Moho. For most events we observe a consistent conversion at ≈ 5 s that is associated with the Moho. In addition a number of other significant features exist, with varying amplitude and varying degree of consistency between events with different backazimuth or distance.

We have marked the most important features for both sites in Figure 7.4. The
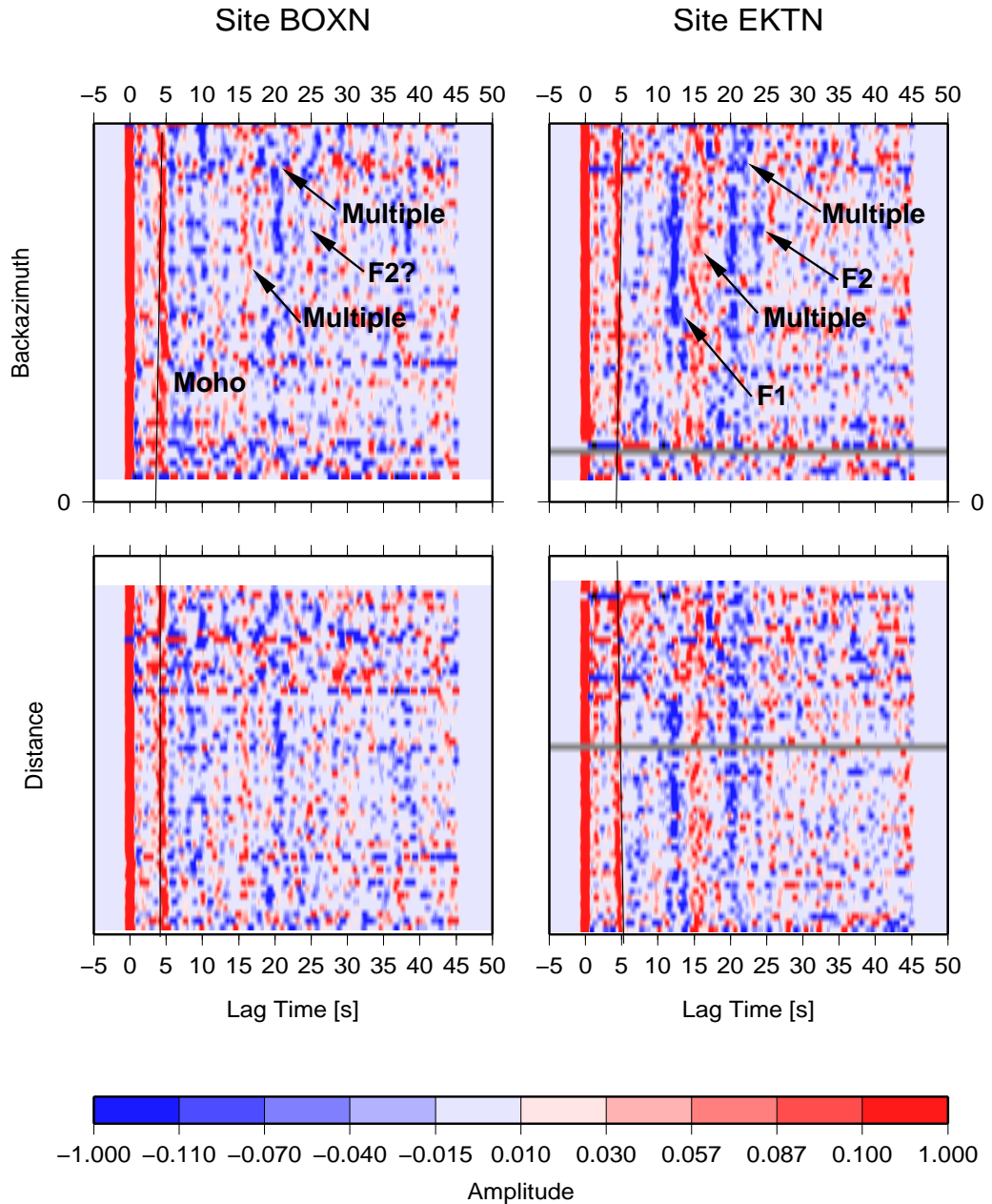
Figure 7.4: Receiver function data for 59 events with moment magnitudes between 5.8 and 8.3 recorded between June 2002 and December 2005. For each site we show the events sorted by backazimuth and epicentral distance. Note that while each receiver function is plotted with the same width, the difference in backazimuth or distance between adjacent events varies enormously and non-systematically along the axis. We have marked the Moho conversion, its multiples and two feature F1 and F2 that we will discuss in the text.

clearest and most consistent conversions at both sites are the primary Moho conversion and its multiples, although the primary conversion seems to be of higher amplitude at site EKTN. In addition we observe two more negative amplitude conversions labelled F1 and F2. F1 extends over several events at site EKTN, but we cannot identify it clearly at site BOXN. The negative amplitude suggests a conversion from high to low velocity and the rule of thumb conversion to depth, multiplying the lag time by 8, suggests a depth of ≈ 100 km. This agrees with results from Snyder et al. [2004], who make similar observations at the same site and identify a low velocity zone at a depth of $110 - 120$ km.

The expression of feature F2 is more subtle than F1 at site EKTN, and there is also some indication of it at site BOXN. It marks a negative amplitude conversion just after the second Moho multiple. Again using the rule of thumb depth conversion, we calculate a depth of 190 km. At this depth Snyder et al. [2004] identify a similar feature in the transverse receiver function. In a purely one dimensional environment, the transverse receiver function would only show noise. They therefore interpret it as a sign of anisotropic shear wave velocity, which is supported by other studies [Snyder and Bruneton, 2007], and anisotropic structures also have been identified in other depth intervals in the south-western Slave craton [Bostock, 1997]. Our observation on the radial receiver function suggests that there might be another low-velocity zone at this depth.

We do not model anisotropy of seismic velocity or electrical conductivity with the current inversion algorithm, as neither forward code has the ability to do so. As discussed above there is no sign of strong electrical anisotropy in the data, as this would require phase split and imply an elevated μ value, but the seismic component of our results will depend on the backazimuth of the events we chose for receiver function computation. Depending on whether the incoming wave is travelling along a fast direction of propagation, or slow axis of propagation, the retrieved velocities will be higher or lower, respectively. At this point we are more interested in the geometry of structures than the absolute velocities. Also, our synthetic tests showed that receiver functions do not have good resolution to absolute velocities anyway. Even in the presence of anisotropy we should be able to recover the location of velocity contrasts, as a change in velocity will produce a P-to-S conversion regardless whether the structure is anisotropic or not.

We have several possibilities to deal with the issue of seismic anisotropy: We can select a receiver function from a single event that we regard as representative. Obviously, we do not have any quantitative measure of noise in this case, but through careful comparison with other events we can get at least a qualitative idea. Alternatively we can select receiver functions from a narrow range of backazimuths that show interesting conversions, or we can average over all backazimuths to average
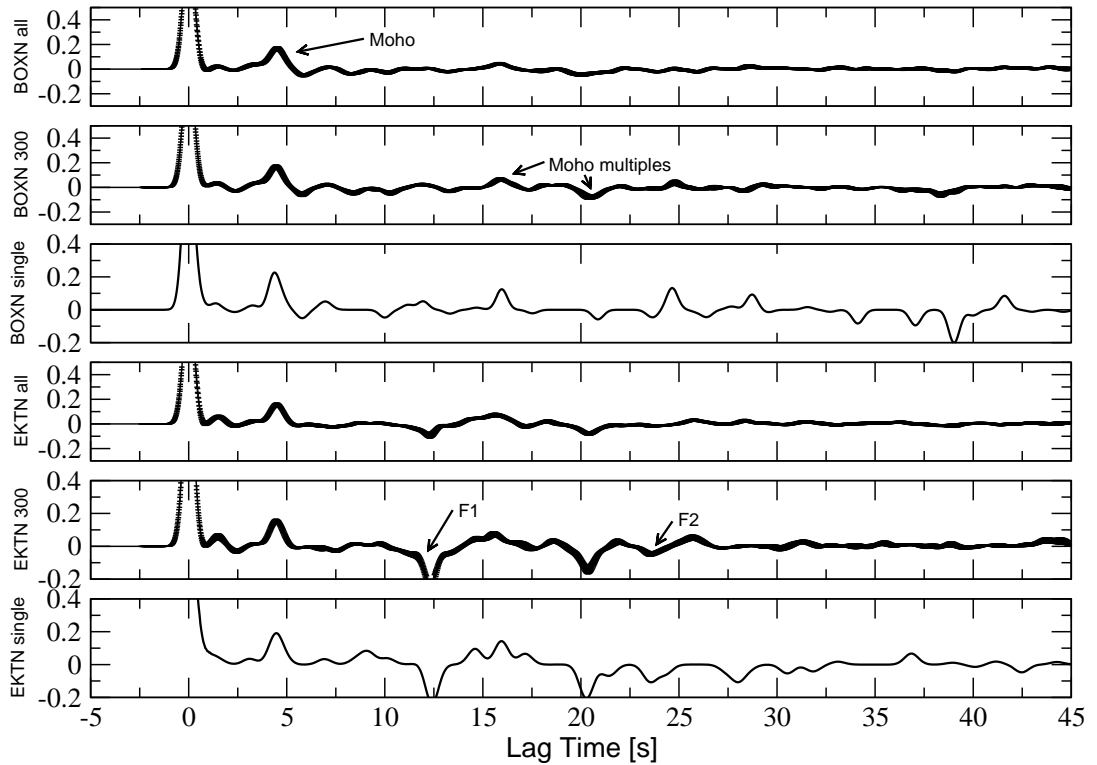
Figure 7.5: Comparison of receiver functions for the two Slave craton sites with different types of events used for the average and a single event receiver function. For each site we compute the average of all high quality events (labelled all), of events with a backazimuth between 299° and 306° (labelled 300), and also show a single high-quality receiver function (labelled single). The single event RF is calculated from event 4 in Table B.1

out the anisotropy. The first approach has the advantage that we will get strong signatures from features that only appear in a limited range of backazimuths, such as F1 and F2. This also implies though that we are looking at a certain direction from which the wave is travelling to the station, and the feature we are observing might be only present in this direction. The MT data in contrast will be sensitive to a volume around the measurement site, and the seismic feature might only occupy a small fraction of this volume. This would suggest that we average over all backazimuths to construct a receiver function for inversion. The problem with this approach is that we do not have uniform coverage over all backazimuths and that we also might average conversions from major interfaces in the process.

Figure 7.5 shows two types of averages for both Slave craton sites plus a receiver function that has been calculated from a single deep event with magnitude 7.3 (number 4 in Table B.1). The two plots labelled "all" show averages over all

backazimuths and distances (a histogram of the distribution can be found in Figure B.2), while the plots with label "300" are only over a backazimuthal range between 299°– 306°. This backazimuthal range corresponds to the region in Figure 7.4 where we can observe features F1 and F2 most clearly and a lot of the high quality events are located in this backazimuth range. The thicknesses of the individual receiver function lines display the error based on the standard deviation. There are some significant differences between the two types of averages. In general the constrained averages show more pronounced features. This is even true for the Moho multiples, but particularly for both F1 and F2, as we expect. We do not employ any moveout correction, which might account for some of the amplitude reduction. However, most of the events come from a similar distance range, so this effect should not be very strong.

This plot also reveals the different expression of F1 and F2 at both sites. While F1 is barely visible at site BOXN, it is, together with the Moho conversion, the most significant feature at site EKTN. F2 is more subdued at both sites, but still more pronounced at site EKTN. The single event receiver function shows the same features as the limited backazimuth average with some additional conversion amplitudes. It is difficult to quantify which of these are noise, and which are actual conversions. In general, this receiver function seems to be of high quality and we will also perform some inversions with just this one receiver function. For these two sites, it appears that a simple stacking of receiver functions from all backazimuths strongly reduces the amplitude of all conversions but the primary Moho conversion. We will therefore not consider it for joint inversion.

7.2 Results from separate inversions

We will invert the data with a number of different parameter settings to investigate the effect of different numbers of layers, regularization and different types of data. First we will invert each dataset separately to get an idea of features required by the individual datasets. Then we will combine the datasets to see in how far this compares to the results of the individual inversions. From here on we will label both the MT and the RF data with the names of the seismic stations to make it easier to relate the results.

Inverting the MT data

Figure 7.6 shows the results when we invert the MT data for each site separately. We use 20 layers to parametrize the model and regularize its smoothness in an Occam-like approach [Constable et al., 1987]. For site BOXN we observe the typ-
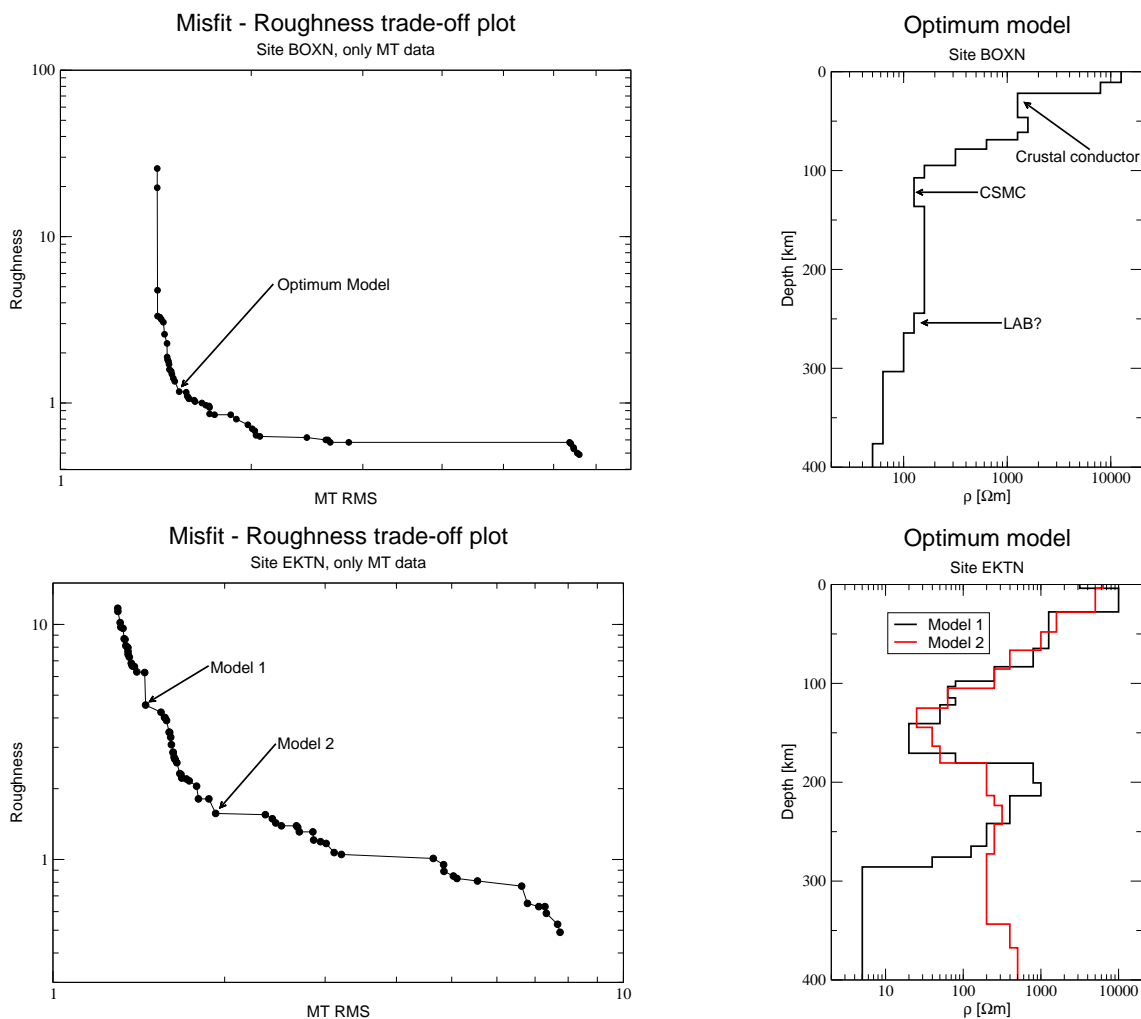
Figure 7.6: Roughness-misfit trade-off curves and optimum models for inversion of the MT data alone. Each model was parametrized with 20 layers. We also mark the most relevant features in the model plot.

ical L-shaped trade-off curve for regularized inversion when we plot misfit against roughness, and we pick a model from the region of maximum curvature as the optimum model. There is a general decrease of resistivity with depth, but we can identify three main features of the model. A conductor at a depth of ≈ 20 km is the main feature in the crust. An increase of conductivity in the middle to lower crust is a widespread feature that is observed in various regions around the world and attributed to either salinous fluids or interconnected graphite [e.g. Jones, 1992, Bahr et al., 2002]. The second conductor, between 90 and 130 km depth, has been termed the Central Slave Mantle Conductor (CSMC) in earlier studies [Jones et al., 2001, 2003]. We also see another decrease of conductivity

at a depth of 240 km. This coincides roughly with the reported depth of the lithosphere-asthenosphere boundary (LAB) in this region. Jones et al. [2003] report a depth of 260 km based on averaged MT data, and 200 km for their Lac de Gras site in the vicinity of our study area, while receiver function results suggest 190 km [Snyder et al., 2004], and surface wave studies 210 ± 65 km [Chen et al., 2007]. In our models this decrease is a minor feature though and not well resolved, we will see whether joint inversion can improve this situation.

The trade-off curve for site EKTN has a slightly different appearance than at site BOXN. We still observe the overall L-shape, but we observe two regions of high curvature. We therefore plot two models, one from each area. The two models are generally very similar and only differ significantly at large depths. In comparison with the model for site BOXN, they both show larger resistivity variations, but the main features remain the same. We observe a similar crustal conductor, and also the low resistivity zone at 100 km depth. Only now it appears to be wider and bounded by higher resistivity below. The resistivity of this resistor, however, is not well resolved, as we can see by comparison with Model 2. Again we see an expression of the LAB in form of another resistivity decrease. As before this is not a well resolved feature and this also is the area where the two models begin to differ significantly.

Inverting the RF data

For the receiver function data we have to take a different approach for the individual inversion. The non-uniqueness of the models that we demonstrated in Figure 6.1 requires to choose a velocity for the crust. Perry et al. [2002] suggest $v_p = 5.9 \pm 0.1$ km/s for the south-western Slave craton, based on the wide-angle reflection and refraction study by Fernandez Viejo and Clowes [2003]. Assuming a Poisson solid this translates into $v_s = 3.4 \pm 0.06$ km/s. Chen et al. [2007] obtain $v_s = 3.546 \pm 0.037$ km/s from Rayleigh wave inversion of array data from the Slave craton. We will test both of these values and examine the impact on the inversion results.

To determine the basic crustal structure, we start with a simple two-layer inversion to determine the depth to the Moho. Figure 7.7 shows a comparison between observed and predicted data from the best fitting model for two different kinds of input data and different crustal velocities at site BOXN. As before the receiver function labelled "BAZ 300" has been calculated from averages restricted to a back-azimuth around 300°, while "BAZ all" denotes an average over all backazimuths. We also noted before that the multiples are less pronounced in the general average, particularly in comparison to the predicted data. For all models the amplitude of the multiple conversions is significantly smaller than predicted by the synthetic
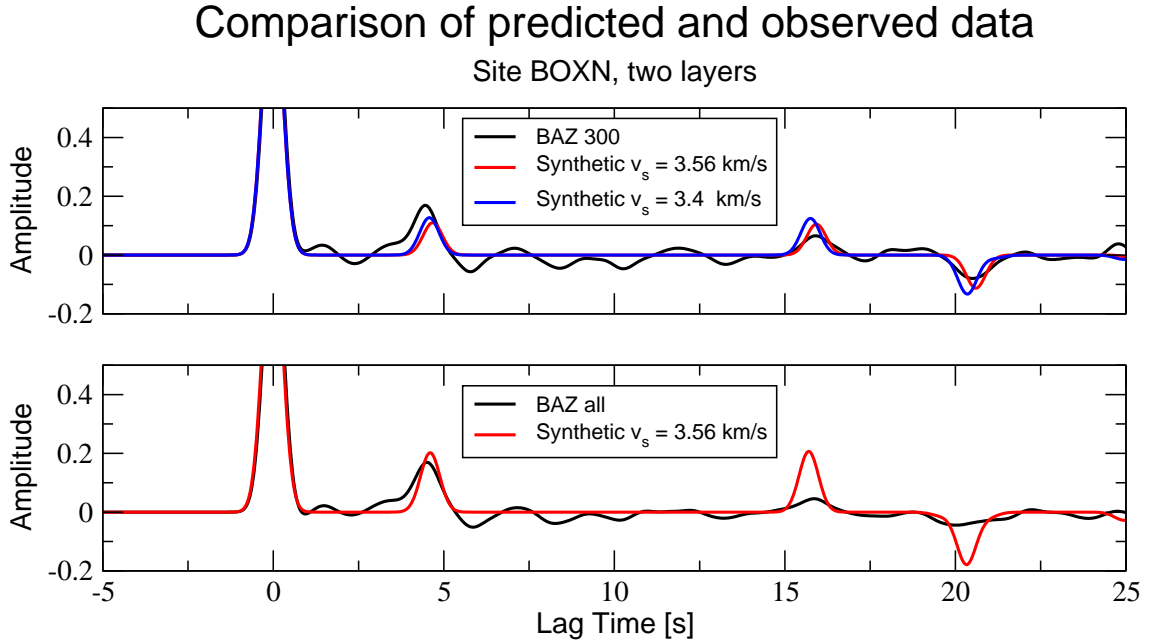
Figure 7.7: Comparison of observed receiver functions and best fitting synthetics from a two-layer model for site BOXN. For the events from a backazimuth of $299 - 306°$ (labelled "BAZ 300") we plot the results with two different crustal velocities. For the average over all backazimuths (labelled "BAZ all") we only plot the results for one crustal velocity, as there is not significant difference.

models. In contrast the Moho conversion has approximately the correct amplitude. We also see that there is only a minor difference between the models with different crustal velocities and this is also reflected in a nearly identical numerical misfit value for both of them. The depth to the Moho varies between 35 km and 37.5 km for the three models. Assuming that the difference between the two crustal velocities reflects the uncertainty, it means that we can model the depth to the Moho within ± 1.25 km. These values are smaller than those given by Bank et al. [2000], who obtain estimates between 37 and 39 km in the same region. They assume a similar crustal velocity, but consistently observe the primary Moho conversion at greater lag times.

Figure 7.8 shows a similar comparison for site EKTN, but this time we only consider a crustal velocity of 3.4 km/s, as the difference is only minor. The resulting Moho depth for both datasets is 35.5 km. For this site we can see that the discrepancy in amplitude for the Moho multiples is much smaller. For the receiver functions with backazimuth 300°, the amplitude of the second multiple matches exactly at the plotting scale and there is only a minor difference for the first multiple. Even when we average all data, the difference increases less than

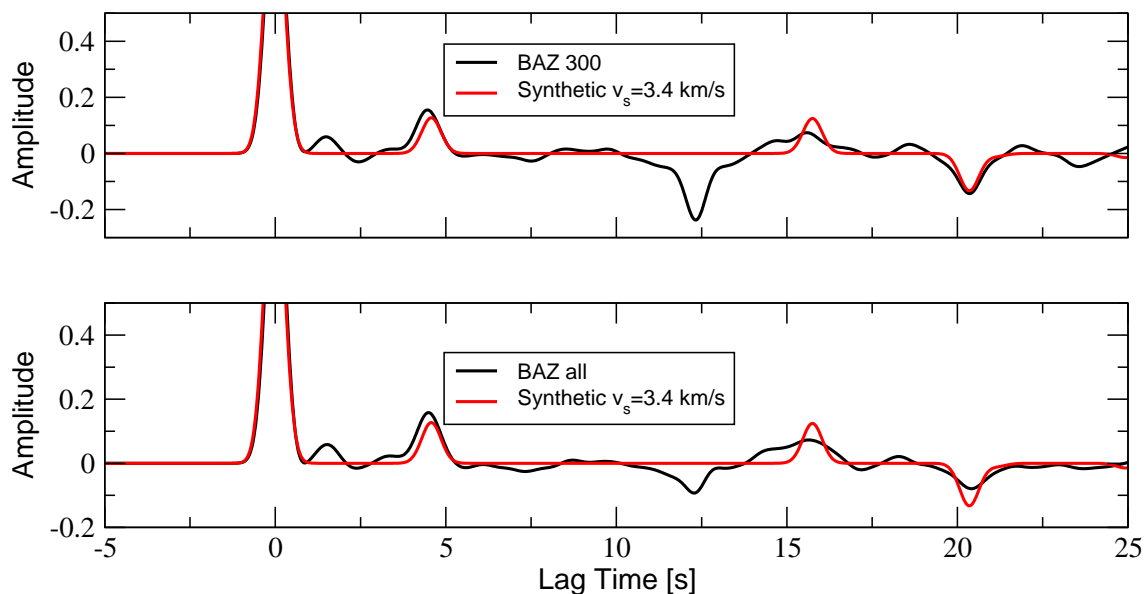## Comparison of predicted and observed data

Site EKTN, two layers

Figure 7.8: Observed data and best fitting two-layer prediction for the two types of averaged receiver functions at site EKTN.

for site BOXN. One possible explanation for this observation is that site BOXN shows a greater departure from a simple one-dimensional structure, either in the form of anisotropy or due to higher-dimensional structure. Another possibility is that there is some interference with conversions from other depth ranges that decrease the amplitude of the multiples at site BOXN.

We have to incorporate the information about crustal thickness into subsequent runs with more degrees of freedom. We performed several experiments where we only fixed the velocity of the uppermost layer, in the hope that the inversion would automatically infer crustal and mantle structure. Disappointingly this was not the case in any of the runs. While the results for the MT data were robust between different inversion runs, the seismic structure varied greatly depending on the number of layers and the permitted range of velocities in each layer. Once we constrain the structure of the crust for the seismic parameters, the results become stable. Figure 7.9 shows a comparison of optimum models for the RF data alone, obtained from runs with different types of regularization. In each run, the crust was parametrized with 3 layers, a thin surface layer with velocities in the range between $2.5-4.0$ km/s, the main crustal layer with a minimum thickness of 25 km and a fixed S-velocity of 3.4 km/s, and another thin layer with S-velocities in the range between $3.8-5.3$ km/s. The complete model comprised 20 layers.
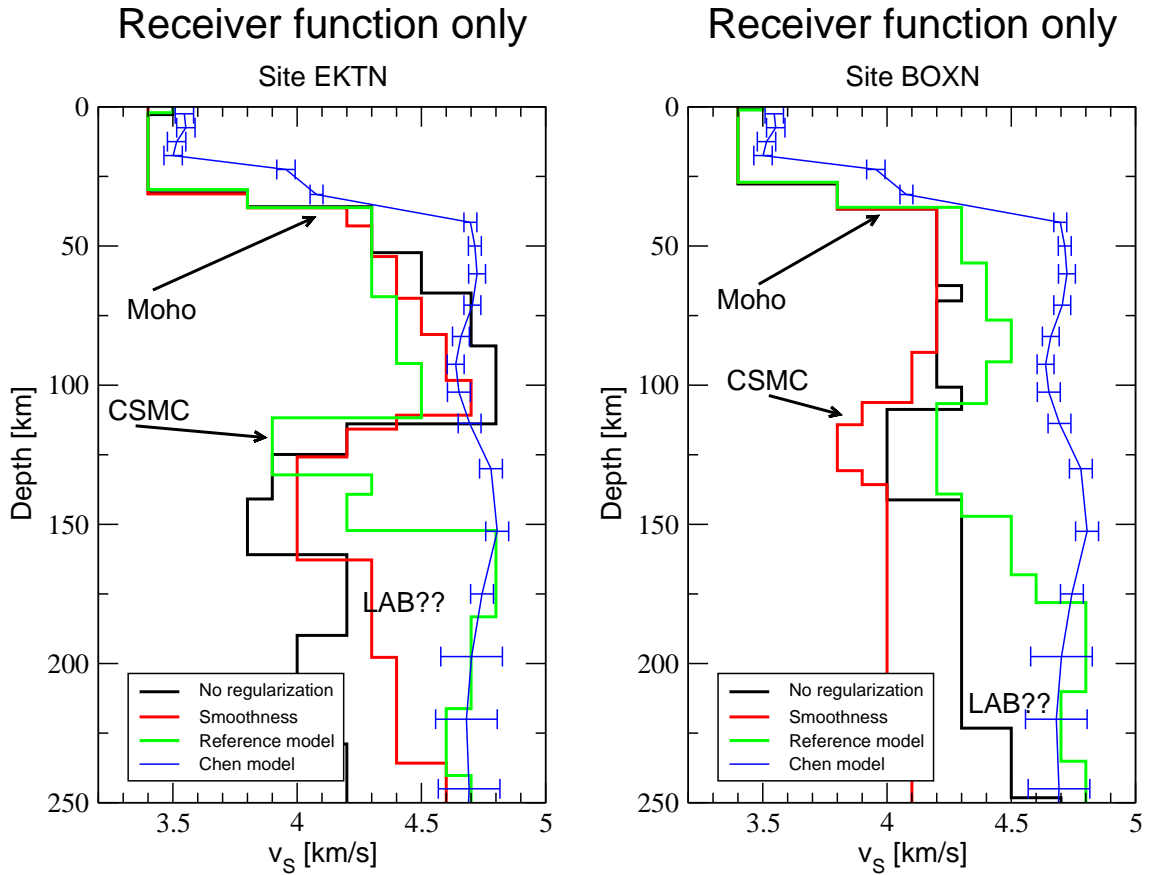
145

Figure 7.9: Comparison of individual receiver function models with different types of inversion settings for site EKTN (left) and site BOXN (right). We also show the surface wave model by Chen et al. [2007] that we use as a reference model in the regularization.

Regardless of the type of regularization we obtain a number of similar features in each model. The Moho is always located at a depth of 35.5 km. This is not surprising, because by enforcing the bulk of the crust to have a velocity of 3.4 km/s we guided the model to resemble the two-layered results. In each model we also observe a similar velocity change across the Moho. This was not specified in the inversion and can therefore be seen as an indicator of the reliability of this estimate. In the depth range between $40 - 110$ km, all models show a gradual velocity increase, although the magnitude of this increase differs depending on the type of regularization. At a depth of $110 - 115$ km the models for site EKTN show an abrupt decrease in velocity. Again, the magnitude of this decrease depends on the type of regularization, but its general depth range between 115 and 150 km agrees between all models. This is the depth at which we observed the CSMC

in the independent inversion of the MT data from this site, and this correlation has been observed before [Snyder et al., 2004]. Below 150 km the models differ significantly, indicating poor resolution. However, two models show a decrease in velocity at a depth of 190 km, which could be interpreted as an expression of the LAB. Yuan et al. [2006] argue that in most cases for P-receiver functions the conversion from the LAB is buried in the conversions from more shallow structures, and advocate the use of S-receiver functions to identify the LAB. On the other hand, Rychert et al. [2005] interpret negative Ps conversions as an expression of a sharp transition into the asthenosphere. From our limited results so far, this interpretation remains highly speculative for our data.

The structure of the mantle at site BOXN is similar to site EKTN. We also observe a low-velocity zone and obtain a similar depth for the Moho. In contrast to the models for site EKTN, the low velocity zone appears to start at shallower depth. This agrees with the shallower position of the low-resistivity zone in the MT models. The similar behaviour makes it likely that the low velocity and low resistivity have a common cause. We cannot identify the LAB at site BOXN. Only the model that has been regularized with a reference model shows a velocity decrease at 200 km depth. This is also a feature of the reference model and therefore might be an artifact of regularization.

7.3 Joint inversion without regularization

The observed coincidence between the low-resistivity CSMC and the low-velocity zone in the independent inversion suggests that we can expect at least a partial compatibility of the two datasets in the joint inversion. Ideally the joint model will also clarify the position of the LAB, which is a poorly resolved feature in both independent inversions.

For the individual inversions, we chose a large number of inversion layers that exceeds the expected number of layers and chose the appropriate model based on regularization. For the joint inversion we have to modify this approach. For a very large number of layers the seismic and electric model effectively decouple, as the coupling is established through coincident layer interfaces. If the thickness of layers is smaller than the resolution of the data, the electrical conductivity and resistivity changes can occur anywhere within a broad region and we will not be able to attribute the changes to a common structure. If we go to the other extreme and use too small a number of layers, we enforce artificial correlations and will not be able to explain the data adequately. We therefore have to determine the optimum number of layers.

Figure 7.10 shows the dependence of the minimum RMS on the number of inver-
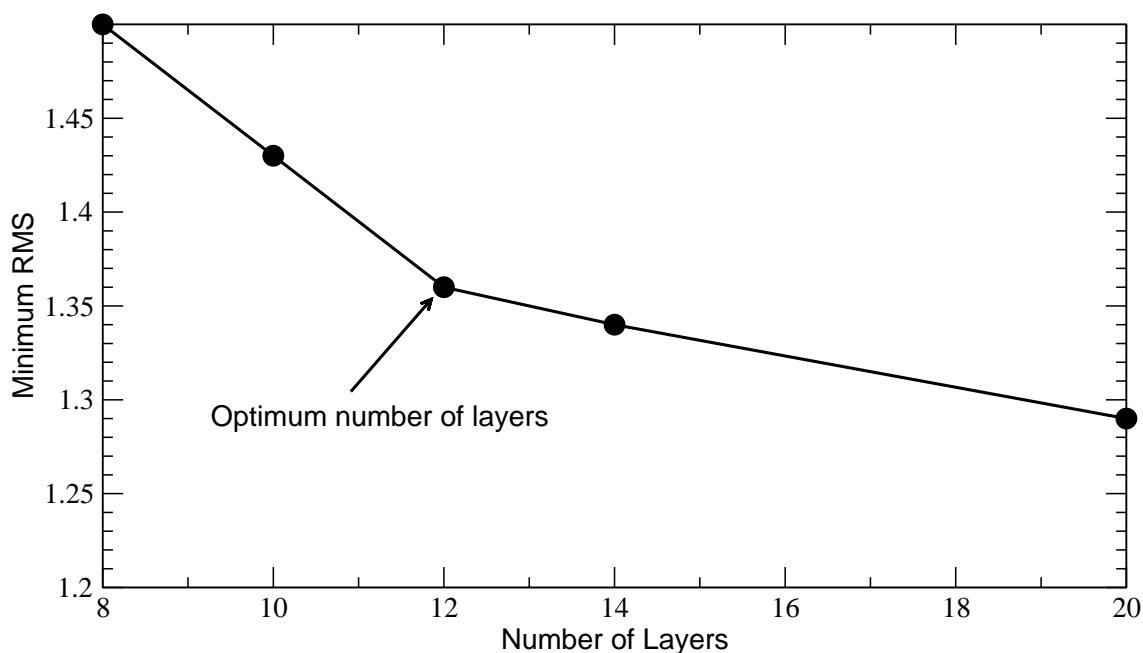
Figure 7.10: Minimum RMS as a function of number of layers for the MT data at site EKTN. For more than 12 layers the misfit does not decrease significantly with increasing number of layers.

sion layers for the MT data from site EKTN. Between 8 and 12 layers we observe a significant decrease in misfit, while for more than 12 layers the improvement is less with increasing number of layers. From this we conclude that 12 layers are sufficient to model the MT data at this site.

For site BOXN the picture is similar, although 10 layers appear to be sufficient for that site. As the difference between 10 layers and 12 layers is not very large in both cases, we use 11 layers for both sites as a compromise. For the receiver function data the situation is more complicated due to the need to fix certain structures, but an inspection of Figure 7.9 shows that 12 layers are located in the depth interval of interest between 0 and 250 km depth, and 10 of these in the well resolved region above 150 km depth, suggesting that 11 layers are also appropriate for the seismic data.

Table 7.1 summarizes the parameter search ranges we use for each layer in the following joint inversion runs. Apart from the fixed crustal velocities, all other parameter ranges should be large enough not to have an effect on the inversion procedure. We show the resulting trade-off curve and some representative models for both sites in Figure 7.11.

For site EKTN we obtain a trade-off curve in the form of a strongly bent L, and the difference in misfit between the worst and best fitting model is only about 15%

Layer no .	1	2	3	4	5	6	7	8	9	10	11
Min. Thick. t_{min} [km]	1	20	10	10	10	10	10	10	10	10	∞
Δt	1	1	1	1	1	1	1	1	1	1	0
Max. Thick. t_{max}	16	35	25	25	25	25	41	41	41	41	∞
Min. Res. ρ_{min} [Ωm]	10	10	10	10	10	10	10	10	10	10	10
$\log(\Delta\rho)$	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Max. Res. ρ_{max}	$19 \cdot 10^6$				$15 \cdot 10^3$						
Min. v_s [km/s]	2.5	3.4	3.4	3.7	3.7	3.7	3.7	4.0	4.0	4.0	4.0
Δv_s	0.1	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Max. v_s	4.0	3.4	4.5	5.2	5.2	5.2	5.2	5.5	5.5	5.5	5.5

Table 7.1: Parameter search range for the unregularized joint inversion runs.

for both datasets. This resembles the results for the compatible test problem with added noise, where we also obtained a cluster of models, but with low variance in misfit. For site BOXN we plot two trade-off curves, one with the setup as given in Table 7.1, labelled "conductor in the crust", and one where we enforce a velocity change at the second interface, effectively allowing only for two crustal layers, labelled "conductor in the mantle". We can see how this change in setup drastically affects the trade-off curve. While the "conductor in the crust" trade-off curve resembles the trade-off curve for site EKTN, and plots as a cluster of points on the scale of this plot, the "conductor in the mantle" curve clearly resembles the trade-off for an incompatible problem. There is not any model with both a low MT and RF misfit when we use this type of setup. The reason for this experiment is the observation of a lower resistivity just below the Moho at site EKTN. This is an unusual result, but a similar observation has been made in the south-western Slave craton [Jones and Ferguson, 1997], and therefore is possibly a real feature. As we discussed before though, the high-frequency MT data at this site is problematic and we can therefore only include periods >0.8 s. As a consequence, this gives us only limited resolution in this depth range. At site BOXN we can clearly exclude a mantle origin of the conductor, as we can only explain both datasets when its top resides in the crust. Regardless of the position of the first conductor, we obtain essentially identical results for the mantle structures.

For site EKTN we examine the nature of the trade-off curve by comparing two models within that curve. Model A is located at the position of highest curvature and has both a low receiver function and MT misfit. In comparison, Model B has a significantly higher MT misfit, but only marginally smaller receiver function misfit. As we saw in the synthetic examples, this might be an indicator of a partial incompatibility between the datasets. Comparing the two models we see a number
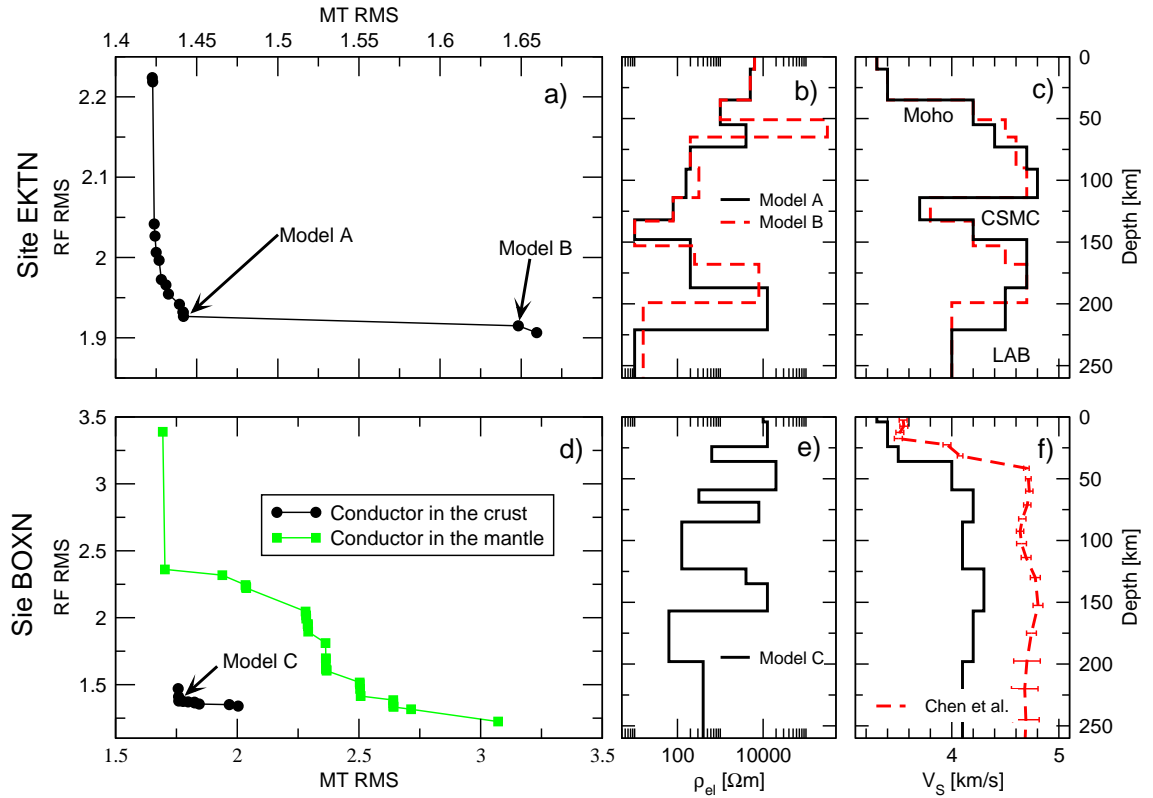
Figure 7.11: Trade-off curves (left) and some representative models from an un-regularized run of the joint inversion algorithm. For site EKTN we examine the nature of the trade-off by comparing two models within the trade-off curve. For site BOXN we examine the impact of the position of the first conductor on the trade-off curve. The seismic model for site BOXN also shows the results obtained by Chen et al. [2007].

of differences, although some of them are mostly an indicator of poor resolution. For example the greatly increased resistivity below the first conductor in Model B can certainly not be resolved by the data. The most significant difference between Model A and Model B is the different position of the lowermost conductive low-velocity zone. The receiver function data seem to favour a velocity decrease at 190 km, while the MT data favour 220 km for the top of the conductor. Figure 7.12 shows a comparison between the measured data and the predicted data from Model A. We also indicate the feature that corresponds to the low velocity zone in Model B by an arrow. Certainly the feature modelled by Model A is more pronounced, and this is what we labelled F2 in our discussion of the receiver function data. Considering the differences in misfit and our assessment of the quality of the features, we regard Model A as the representative model. It appears
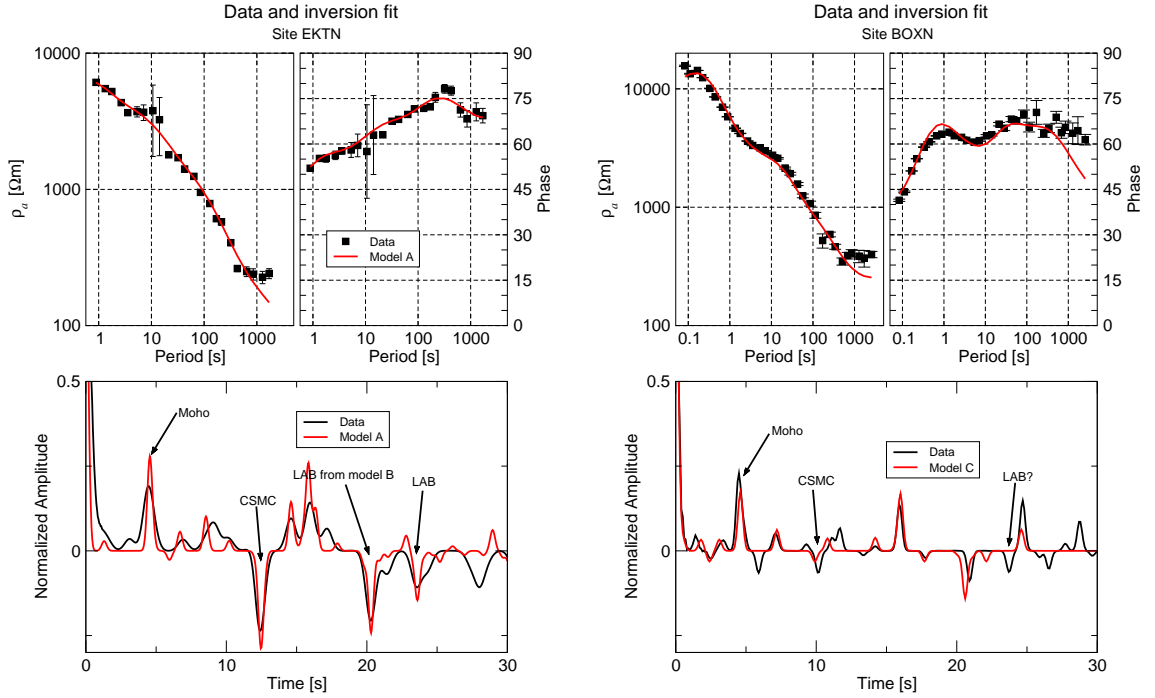
Figure 7.12: Measured (black) and predicted (red) RF and MT data for the optimum model at each of the sites. We label the main features in the receiver function plot.

that in this case we have benefited from the joint inversion approach, as it guided the inversion method to a more robust feature. Furthermore within the resolution of the data we can explain both datasets by a joint model.

In general the joint model reproduces the structures we already observed in the individual inversions. Particularly the CSMC and the corresponding low velocity and low resistivity zone have been reproduced in all models. In addition we can now say that we can model this feature by a jointly compatible seismic and electric structure. Before the joint analysis there was some indication based on the similar depth range in which these features occur in the individual models, now the joint model allows us to join these two separate observations into one common feature. Still, there are limits to the extend to which we can identify these structures. For example, the bottom of the CSMC cannot be resolved by the MT data. The increase in velocity, in contrast, is related to significant features in the receiver function data. We therefore can model the bottom of the conductor coincident with the velocity increase, but this is not strictly required by the data. We can assess the possible variations in the models by using the nullspace perturbation approach [Muñoz and Rath, 2006] that we described in Chapter 5. To ensure that we only perturb the models within the nullspace, we recalculate the misfit after
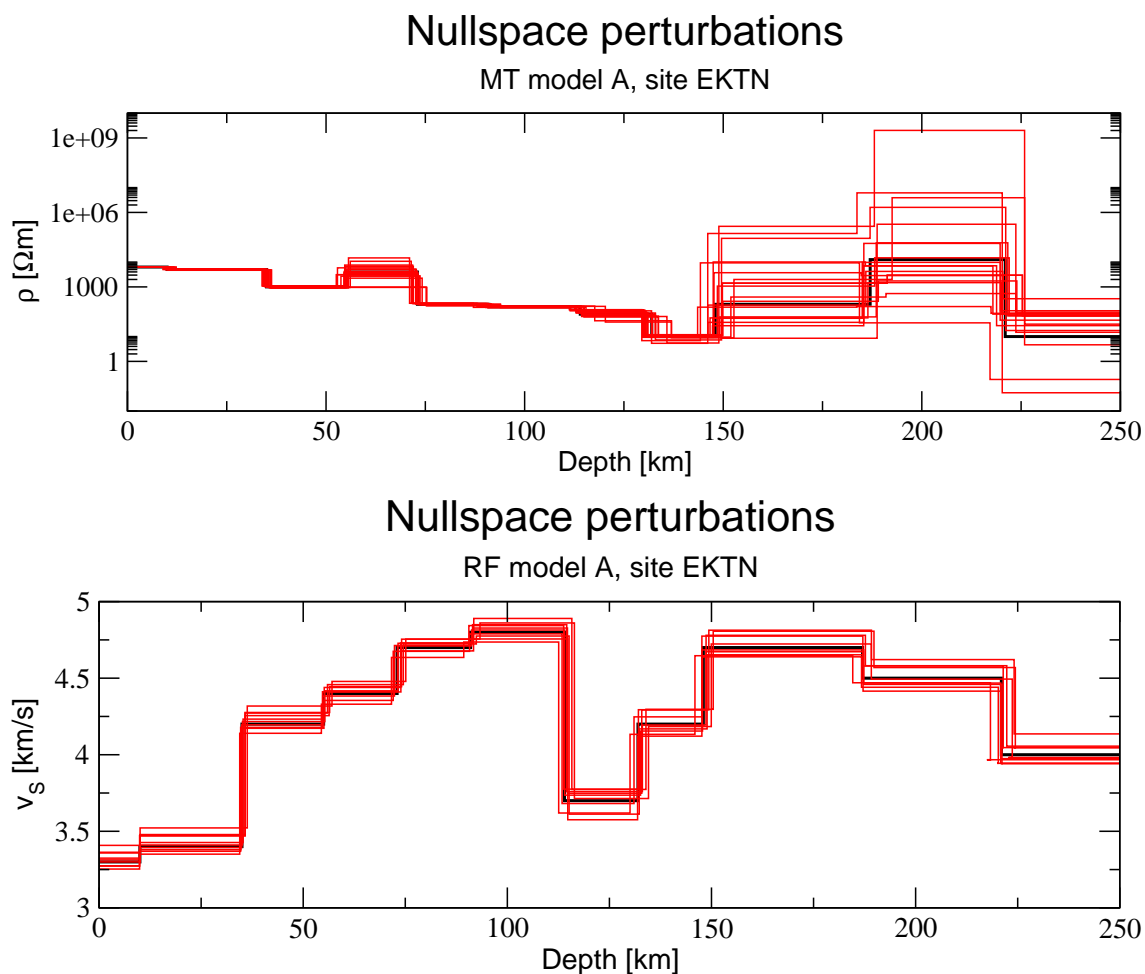
Figure 7.13: Optimum MT model for site EKTN (black) and 20 perturbed models within the nullspace. The variation between the models illustrates the resolution to the corresponding features.

each perturbation and stop the algorithm when the difference in misfit exceeds 5%. For the following examples the relative cut-off for the eigenvalues was set to 10^{-2}. While this choice is somewhat arbitrary, it typically allowed one or more iterations of the perturbation algorithm before we hit the prescribed threshold. This suggests that the cut-off is appropriate.

Figure 7.13 shows Model A from site EKTN plus 10 models that have been generated by random perturbations with the nullspace projection approach. We can see that the resistivity of the resistive lower layers is virtually unconstrained. In contrast the top of the conductive layers and their resistivities are much more tightly bound. The top of the CSMC varies by 7 km within all these models and while the bottom of this conductor varies by the same amount in most models,

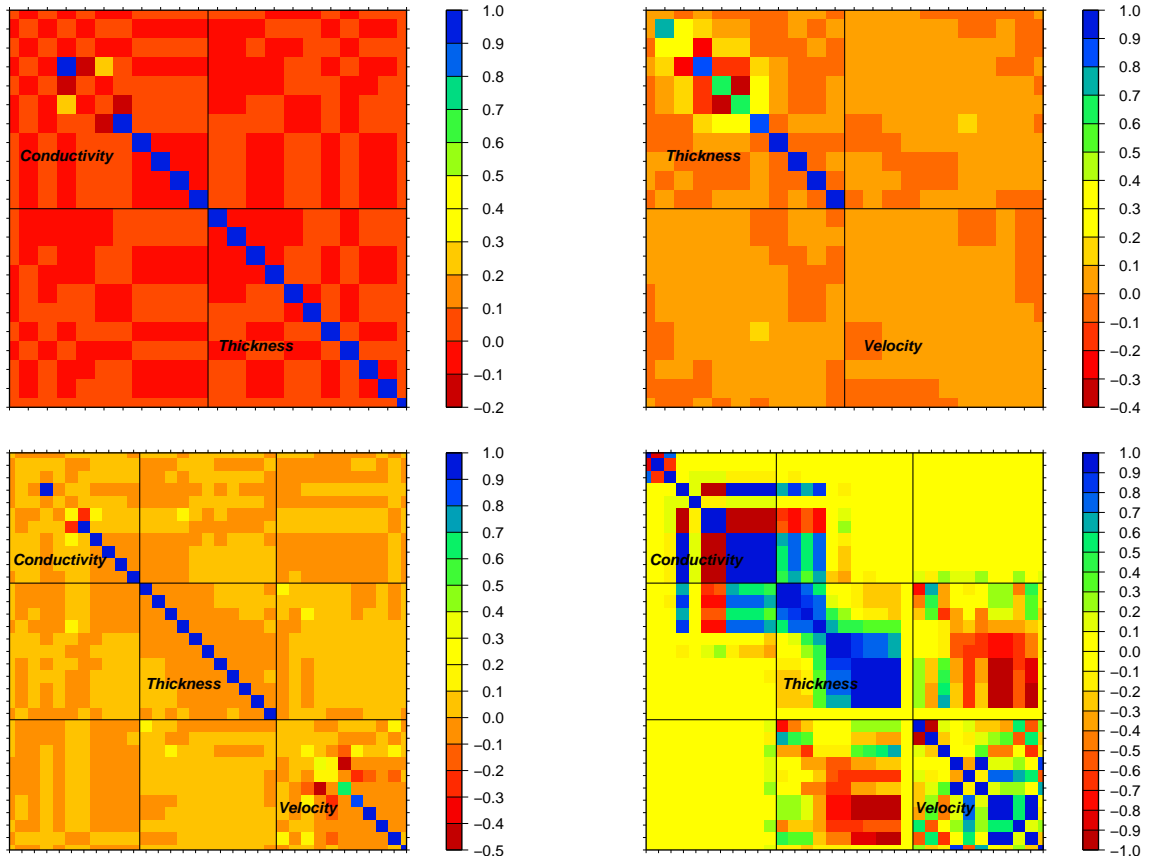Figure 7.14: The individual nullspace projection matrices for the MT model (top left) and the RF model (top right) at site EKTN. We also plot the joint nullspace matrix (lower left) and the joint correlation matrix (lower right).

we obtain one model, where it extends to much greater depth.

The receiver function models do not show such strong differences between well and poorly resolved parameters. Both seismic velocities and the position of the interfaces vary by similar amounts throughout the model; only the position of the Moho seems to be better defined. Typical variations are ± 0.1 km/s for the velocities and ± 5 km for the interface positions.

The graphical display of the nullspace projection matrices in Figure 7.14 visualizes these observations in a different form. The values close to unity for the diagonal elements in the lower right quadrant for the MT matrix suggest that we have hardly any resolution for the thickness of most of the layers. While the perturbed models show some variation in thickness in each of the layers, it seems that the nullspace matrix paints an overly pessimistic picture. The high values for the entries corresponding to the resistivity of the last three layers, in contrast, is clearly reflected by the variations in the perturbed models. The single high value

in the conductivity quadrant corresponds to the resistive layer between 50 and 70 km depth. Here we also find significant off-diagonal elements that display the interaction between the inversion parameters.

For the receiver function model parameters we only observe high values of the nullspace matrix associated with the thickness parameters, particularly for the lower layers. In addition, there is a larger number of significant off-diagonal elements, suggesting considerable interaction between the layer thickness values. This confirms the observation from our experiments and is due to the non-uniqueness problem of receiver functions we discussed before. In this case the low values for the velocities give a more optimistic view on the situation, as the application of this matrix in the model perturbation algorithm demonstrates. From the matrix we would expect to see only small velocity variations, but they are on the same relative order as the thickness variations, as the off-diagonal elements project these variations on the velocities of the model.

These observations provide us with another indication of the compatibility of the datasets, as they demonstrate how far we can perturb each model and still maintain a similar misfit. For example, the small variation in the top of the CSMC makes a common interface plausible, while the larger variations of its bottom in the MT data do not exclude a common interface, but neither prove its existence.

Using averaged receiver functions

As discussed before the use of a receiver function calculated from a single event bears the risk of inverting noise contributions for structure and creating artifacts in the models. We will now examine the results when we use the averaged receiver functions.

Figure 7.15 shows the results for the averaged receiver function from a back-azimuth of 300° for site BOXN with otherwise identical settings to the previous inversions. We achieve a similar MT RMS, but the receiver function RMS is nearly halved and now even reaches value below unity. As the comparison between the data and the synthetic RF reveals though, this is not so much due to a greatly improved fit, but more to the overall lesser amplitude of the receiver function. We do achieve a good fit to the Moho conversion and its multiples and some additional features, but there are a number of unexplained amplitudes, particularly in the time window between 5 and 15 s.

The two models we show for comparison do not differ significantly for both the MT and the receiver function parts. Both MT models are similar to the model we presented before, which is not surprising as they are derived from the same data. The receiver function models, in turn, are much simpler than before. Especially Model B shows only the low velocity zone associated with the CSMC,
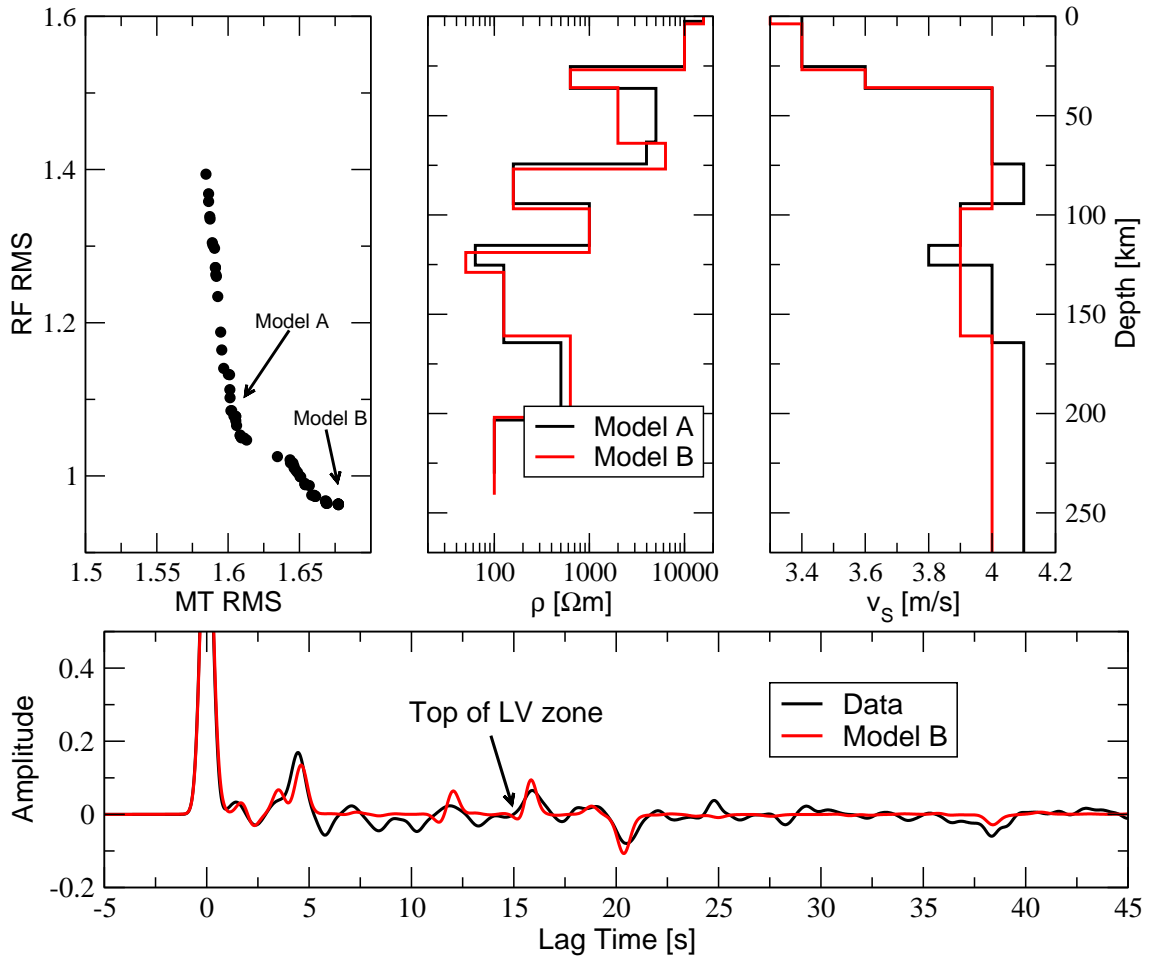
Figure 7.15: The trade-off curve (top left), two representative electrical (top middle) and seismic (top right) models, and the predicted receiver function data from Model B compared with the receiver function average with a backazimuth of 300° at site BOXN.

but an otherwise constant velocity in the mantle. The depth region of the low velocity zone agrees with the previous inversion results. Comparing the predicted data with the averaged receiver function, we see that it is based on a feature in the data, although the fit to could certainly be improved, and was better for the single event receiver function. It appears that this model represents the absolute minimum in terms of structure that we need to explain the most important aspects of the data.

As a consequence of the homogeneity of the receiver function model, we see first signs of decoupling between the electric and the seismic parts of the joint model. The resistivity model shows much more structure throughout the mantle, and we therefore obtain a number of interfaces where the resistivity changes and
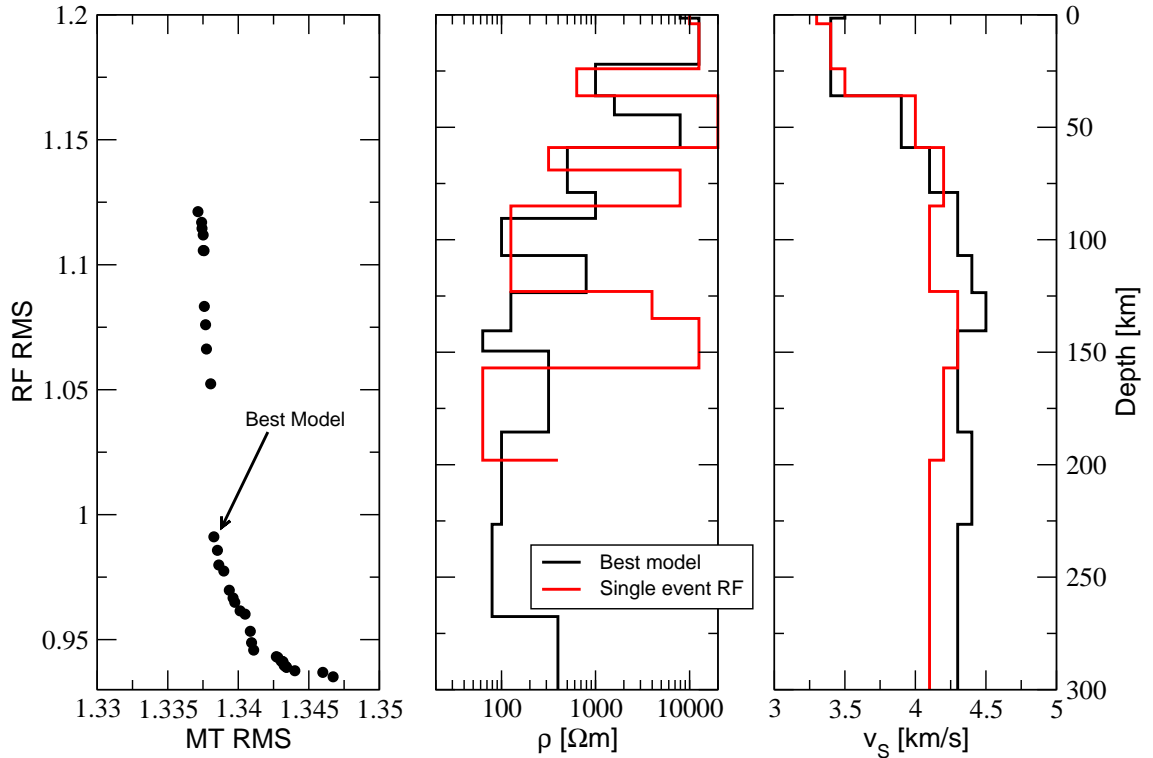
155

Figure 7.16: The trade-off curve (left), and electric (middle) and seismic (right) parameters for the best joint model with 20 layers and without regularization for site BOXN. In addition we also plot the best model from the single event results.

the velocity remains constant. One of the results is that we no longer have a one-to-one correlation between low resistivity and low velocity for the CSMC, as we had in the previous model.

We can take the decoupling even further by increasing the number of layers in the joint model to 20. Figure 7.16 shows the resulting trade-off curve and our chosen best model for site BOXN. The variation in MT misfit is ≈ 1% and therefore negligible. Again we reach and RF misfit of less than unity. Therefore, and because the difference to the best fitting model is not large, we chose the model with a receiver function RMS closest to 1.

Surprisingly, the low-velocity zone has now shifted to depths below 140 km, deeper than the previous estimates of its bottom and this is a consistent features of all models from this run. What happened is that the inversion now picked a negative conversion just after the first Moho multiple and modelled it as a low velocity zone. If we maintain our definition of the CSMC as a low resistivity and low velocity zone, this model suggests it to be at a depth of 140 km and only about 10 km wide. Another surprising feature of this model is a correlated change
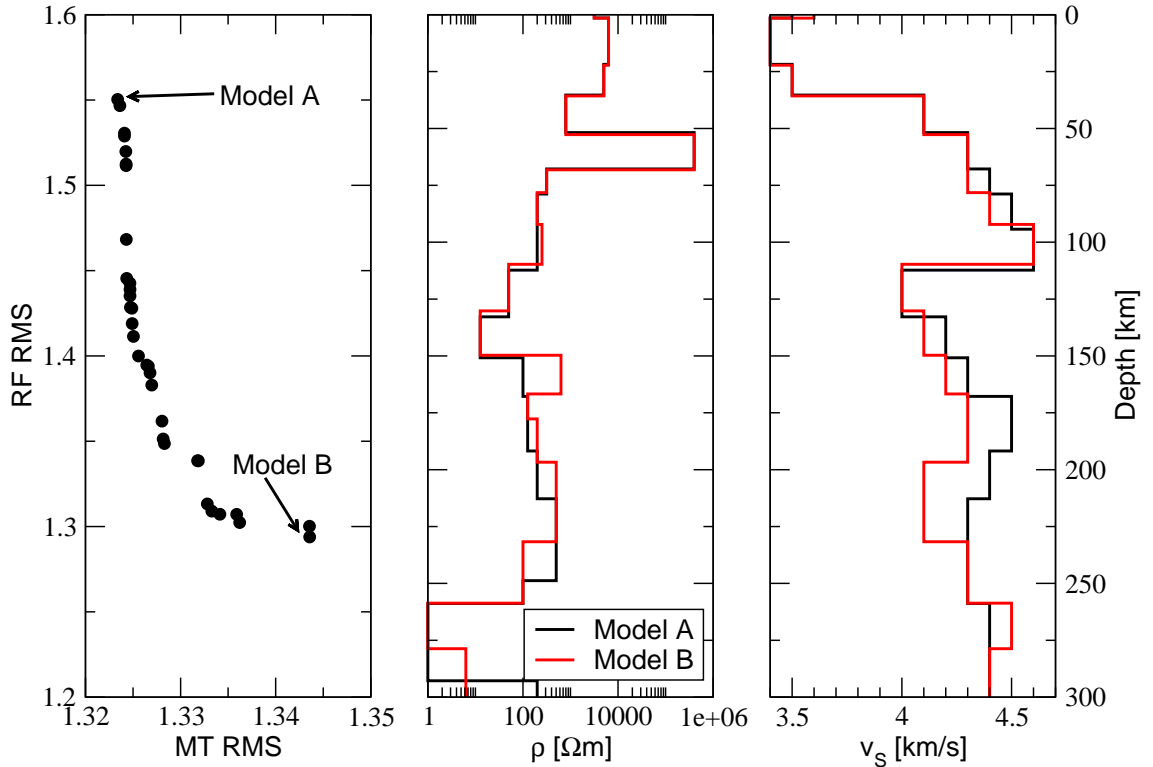
Figure 7.17: The trade-off curve (left), and electric (middle) and seismic (right) parameters for two joint models with 20 layers and no regularization for site EKTN.

to lower velocities and resistivities at a depth of 220 km, suggesting the transition into the asthenosphere. We did not find this feature in previous models and even here its expression is only weak.

We observed before that the difference between the averaged and single event receiver function is less pronounced at site EKTN. Consequently we expect the models based on the averaged RF to be resemble the single event models more closely than at site BOXN. This expectation is confirmed by the inversion results shown in Figure 7.17. As before with site BOXN we observe a reduction in receiver function misfit, but this time not as drastic as for site BOXN. Also, the range in MT misfit for all models is less than 5%, but the difference in RF misfit is considerable. We plot the two extremal models in order to see what the reason for the difference might be. At depths < 170 km the models generally agree. There are some minor variations in velocities, but the general appearance is the same. In particular we retrieve the low resistivity and low velocity zone at depths between 110 and 150 km that we observed in the previous model for this site.

At depths > 170 km the velocities for the two models begin to differ, whereas the resistivities, at least qualitatively, remain the same. In both resistivity models
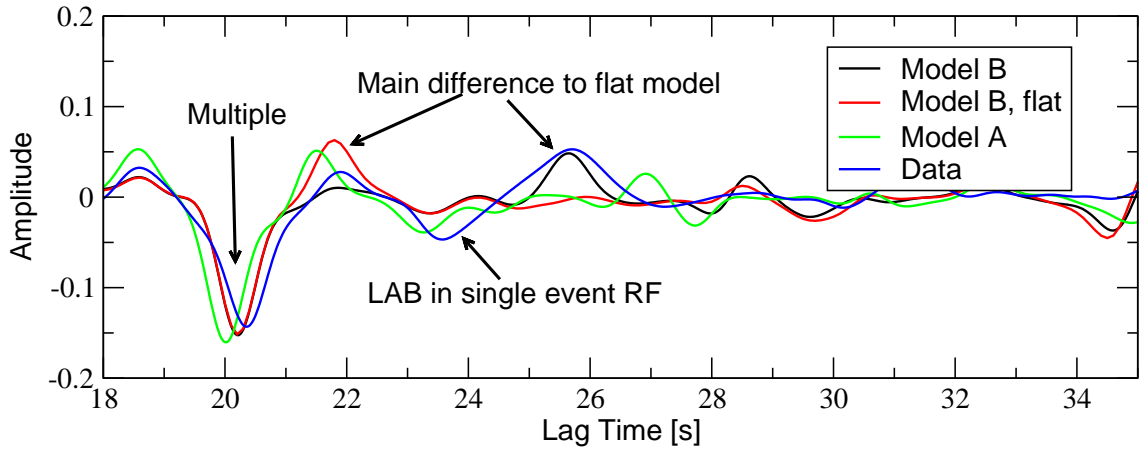
Figure 7.18: Comparison between measured receiver function and various model predictions in the time window sensitive to the depth region of the LAB.

we observe a decrease in resistivity at depths between 220 and 240 km that we might interpret as an expression of the LAB. The receiver function models, in contrast, show a first decrease in velocity at 190 km depth. We discussed this issue before with the single event receiver function and concluded that the feature in the recorded data that corresponds to a velocity decrease at 220 km is more reliable.

We can examine this situation for the averaged receiver function by comparing the expression of different models in the predicted data, as shown in Figure 7.18. In addition to the predicted data from the two models in Figure 7.17, we also plot the predicted data from a model derived from Model B, but with constant velocity between 170 and 250 km depth. We can see that none of the models reproduces the measured data exactly, but they approximate the shape of some of the features. The comparison between Model B and its flat counterpart shows two regions of major difference. The amplitude after the Moho multiple is decreased in Model B and increased at a lag time of 26 s. Both of these features have a corresponding expression in the measured data, although Model B now underestimates the amplitude after the multiple and performs only marginally better there than the flat model which overpredicts it. Model A, which has a higher misfit, also overestimates the amplitude after the multiple and, in general, appears to be shifted with respect to the other two models and the data. If we shifted the predicted data from Model A by 0.5 s it seems that we would get an improved fit, and even reproduce the negative amplitude that we previously associated with the LAB better than in the best fitting Model B. This could be achieved by either deeper interfaces or reduced velocity in one of the layers above.

Figure 7.19: Regularized inversion results for site EKTN. We plot the trade-off curves between MT misfit and roughness (upper left) and RF misfit and roughness (upper right) together with the electrical (middle) and seismic (right) parameters for two different selection orders.

It is interesting that the MT component of Model A does not show a resistivity decrease in this depth region, but only at about 250 km depth. There are two possible explanations for this observation: Either the LAB is sensed differently by electrical and seismic data, which would not be very surprising given the ambiguous definition of the LAB itself, or the higher number of layers compared to the single event inversion decouples the two models and poor resolution allows for different positions. We cannot give a definitive answer to this questions, but in the light of the analysis from the single event receiver function we favour a common origin for the LAB, as identified from the joint model with fewer number of layers.

	MT RMS	RF RMS	MT Rough.	RF Rough.
EKTN MT first	1.69	1.56	0.86	11.05
EKTN RF first	1.66	1.69	4.64	4.73
EKTN RF ind.	0	1.55	0	17.89
EKTN MT ind.	1.62	0	2.86	0
BOXN MT first	1.78	1.60	1.56	6.74
BOXN RF first	2.33	1.55	2.04	3.77
BOXN RF ind.	0	1.11	0	11.659
BOXN MT ind.	1.45	0	2.43	0

Table 7.2: Misfit and roughness values for the two Slave craton sites depending on the order of trade-off selection. We also show the misfit and roughness values for the individual inversions for each site.

7.4 Regularized joint inversion

As the last examples showed, increasing the number of layers decouples the two datasets and reduces the stabilizing effect of the joint inversion procedure. The synthetic tests demonstrated that regularization is one way to increase the stability of the results, particularly if we have a suitable reference model. We will apply different strategies to identify the best models and see how the results differ.

We already remarked during the discussion of the synthetic tests that a smoothness-based regularization is not appropriate for the receiver function data, whereas a reference model approach is problematic for MT. We therefore try a combination of both approaches by increasing the number of objective functions in the inversion to four. The two data misfit functionals remain identical to the ones we used before, as well as the difference to the reference model for the seismic parameters. We now add a smoothness-based functional defined purely on the electrical parameters. The resulting four-dimensional hypersurfaces cannot be directly displayed in any form, but we can look at projections onto different planes to visualize the different trade-offs. In the following examples we use the model by Chen et al. [2007] from inversion of surface wave data as a reference model.

The first strategy we use is to plot the trade-off between MT misfit and roughness for the MT parameters for all models, as shown in Figure 7.19 (upper left). From this cluster of points we can identify the optimum trade-off for the MT part of the inversion by selecting models from the left border of the cloud that fulfil the maximum curvature criterion. Due to the projection from four to two dimensions we usually cannot identify a single model, but a number of models with similar misfit and roughness. With these models that are all optimal for the MT, we can
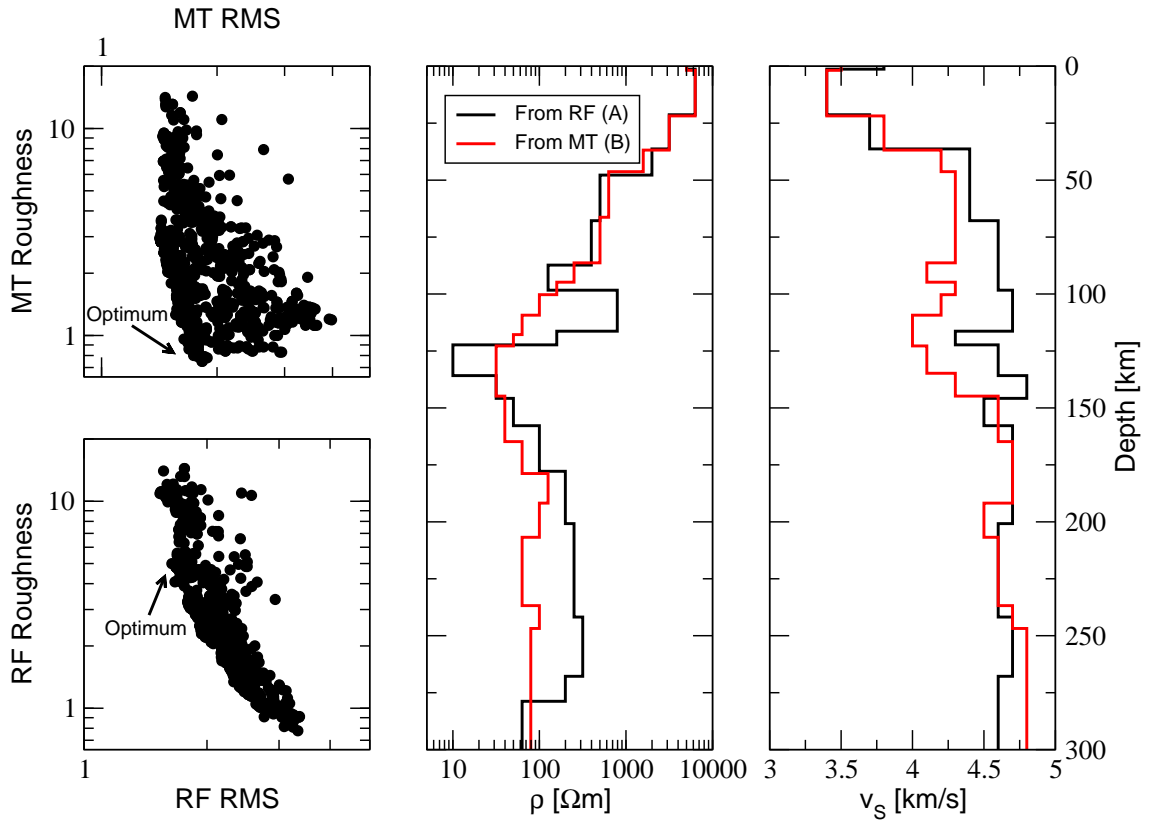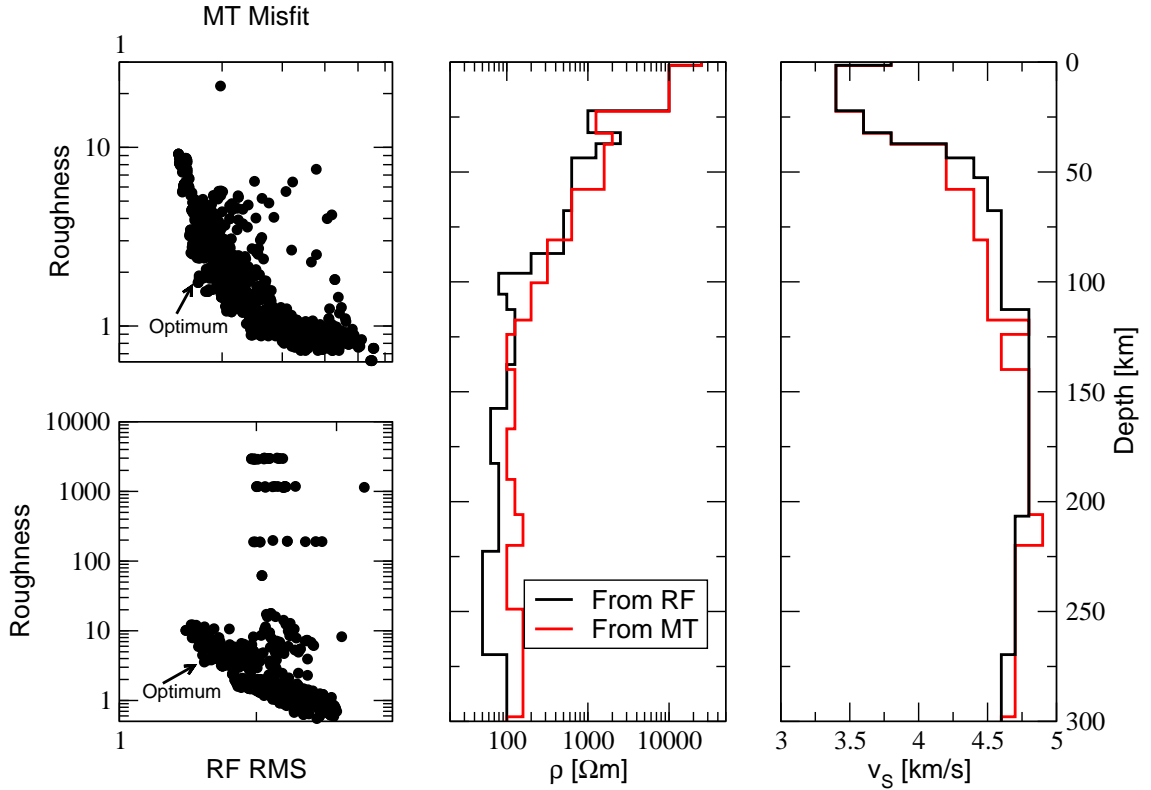
Figure 7.20: Regularized inversion results for site BOXN. We plot the trade-off curves between MT misfit and roughness (upper left) and RF misfit and roughness (upper right) together with the electrical (middle) and seismic (right) parameters for two different selection orders.

repeat the procedure for the seismic part of the model. Obviously the order of this selection process can be reversed and we can first plot the receiver function roughness trade-off (also plotted in Figure 7.19, lower left) and then select MT models from the resulting subset.

Surprisingly, the models obtained with these two methods have similar misfit values for both MT and receiver functions. When we base the selection on the MT trade-off first, we even only obtain a single model with a RF misfit of 1.56, comparable to the best receiver function we obtain in the run. When we reverse the order, we obtain a number of models with similar RF misfit, but varying MT misfit. From this group we pick the model with lowest MT misfit, in this case 1.66. A comparison of all the misfit parameters for the two selection orders and both sites is shown in Table 7.2

The two models we obtain in this way for site EKTN have the same overall appearance, but we can clearly see the influence of the selection order on the

characteristics of the model. When we select a model based on optimum trade-off between receiver function misfit and seismic parameter roughness first, we do not consider the smoothness of the electrical parameters. Model A in Figure 7.19 shows two distinctive conductive zones at depths of 80 km and 120 km with a resistive zone in between, whereas Model B shows a broad conductor which starts at the same depth and encompasses both conductors of Model A. In the shallower part the electrical parameters are virtually identical, while in the deeper part we again observe the elusive LAB in one of the models, but not in the other. The results are similar for the seismic parameters. Where Model A appears more focused, with smaller variations and only a small low-velocity zone at 120 km depth, Model B shows a much broader low velocity zone that correlates with the low-resistivities of the MT. At depth both models converge, partially because we do not have sufficient resolution and the regularization dominates this part, but we observe differences in the depth range of the LAB. Unfortunately, as in the example before, we do not have a clear coincident velocity and resistivity decrease that would clearly mark the position of the LAB.

We repeat the same analysis with both selection orders for site BOXN and show the results in Figure 7.20. The main difference to the previous results is that the MT misfit is considerably higher, if we select the optimum receiver function model first. Furthermore this model does not show any low velocity zone that could be interpreted as an expression of the CSMC. Similarly to site EKTN, when we select the optimum MT model first we see smoother variations in resistivity and, at this site, a small low velocity zone between 120 and 140 km depth. This results agrees with the 20 layer, unregularized model, but not with the 11 layer model which suggests that the low velocity zone is more shallow.

*Hofstadter's Law: It always takes longer than you expect, even
when you take into account Hofstadter's Law.*

Douglas Hofstadter, Gödel, Escher, Bach

8

Discussion

Are seismic and MT data an "essential combination" [Jones, 1987] or are correlations between seismic velocity changes and electrical conductivities just "A case of Holmes's curious dog" [Cook and Jones, 1995]? Joint inversion methods cannot give an answer to this question, but they can provide indications in how far certain structures have compatible expressions in both datasets. We saw in the synthetic test cases that we can distinguish the situation of coincident interfaces from differing interfaces through the shape of the misfit trade-off curve, as long as the data quality and resolution are high enough. To a certain degree the question of compatibility is a matter of parametrization. For very large numbers of layers we effectively decouple the seismic and electric datasets and can always find a model that explains both datasets. Therefore one of the challenges of joint inversion is to find parametrizations that provide enough freedom to model all essential features, but still couples both parts of the model.

One way to achieve this is to regularize the inversion by choosing the minimum number of layers appropriate for the problem. This requires trial and error inversion runs with different numbers of layers, but seems to provide the strongest coupling and stabilization. We also demonstrated how we can use different types of regularization with a large number of layers to obtain a similar effect. This removes the necessity to test different parametrization, but adds the complication of finding appropriate levels of regularization, and, for the seismic part, finding a suitable reference model. Once we have found a suitable parametrization we achieve misfit levels for the joint models that are comparable to the misfit of individual inversions, at least for the two sites on the Slave craton. This suggests

that we can reconcile these two different datasets in terms of a joint model. From this point of view our joint inversion approach with a genetic algorithm is a great success. We can assess the degree of compatibility between the two datasets if they are compatible, we can reconcile the two models and find a joint model that explains the data equally well.

Unfortunately, the non-unique nature of seismic receiver functions requires considerable efforts to extract stable and meaningfull models, be it in individual or joint inversion. Even then we observe considerable variations in seismic velocities and position of interfaces in our models. In comparison, the electric part of the model is much more stable and requires less effort in the choice of parameters. For this reason the stabilization effect is less than we hoped for and it is still difficult to answer relatively simple questions such as: Is the CSMC shallower at site BOXN than at site EKTN? At what depth is the LAB below the two sites? Particularly at site BOXN the expression of those two features is only minor in the receiver function data. This causes shifts in the depth to the CSMC and varying expressions of the LAB in the models.

For site EKTN it seems that at least the CSMC is a consistent feature, both in the receiver function and in the MT data and its position does not vary greatly with differing inversion settings. All joint models show a significant reduction in seismic velocities and resistivities at a depth of 113 ± 3 km. Figure 8.1 shows a comparison of the individual and joint inversion results for site EKTN. We can see that in the individual inversions the decrease in velocity does not coincide exactly with any resistivity interface, but the two differ by 10 km. In the joint model these two separate interfaces have been joined together which suggests that there is a common cause for the decrease in velocity and resistivity. Obviously, the joint inversion result cannot prove that this is the case, and it is still possible that this apparent coincidence is caused by different processes [Cook and Jones, 1995], without indication for the contrary, Occam's razor justifies an interpretation of a common source.

So far we have not discussed what such a process could be that causes a simultaneous decrease in both velocity and resistivity. The reason is that a full discussion of this topic would require an extended review of conductivity mechanisms and the relation between velocities, conductivities and geological structure. This is easily a topic for a separate thesis and therefore we will not even attempt to do so.

The results presented here provide the basis for a wide range of improved joint inversion methods. One of the advantages of the genetic algorithm inversion is that it can easily be extended to include further datasets or different types of regularization. The first logical extension is to include seismic surface wave dispersion data
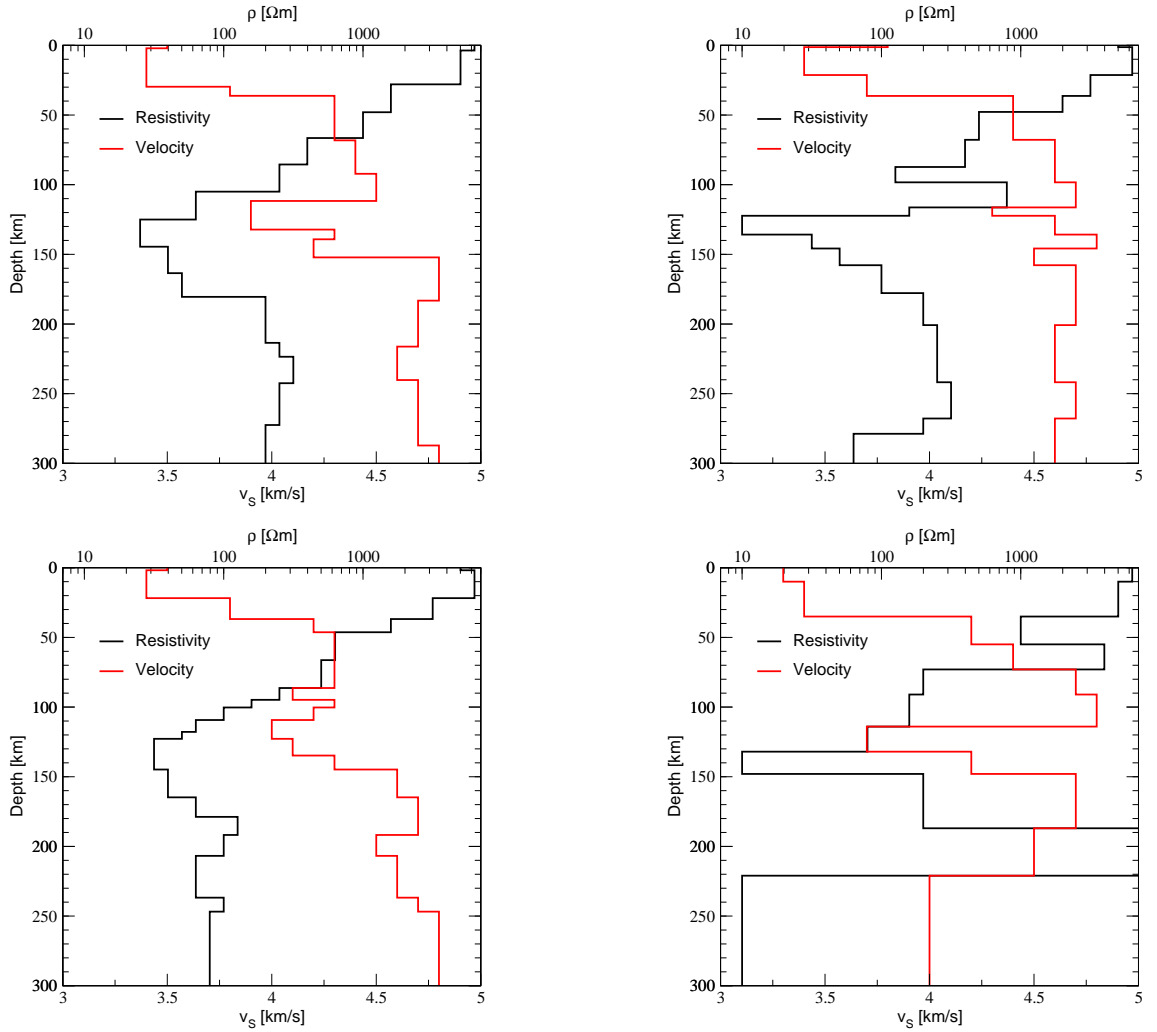
Figure 8.1: Electrical and seismic parameters of the optimum models plotted together for different inversion settings: Individual inversion (top left), regularized inversion with selection of optimum RF model first (top right), regularized inversion with selection of optimum MT model first (lower left), and unregularized inversion with a minimum number of layers (lower right).

in the inversion. We discussed the deficiencies of receiver functions to resolve seismic velocities and we mentioned the joint inversion method by Julia et al. [2000] that combines surface waves and receiver functions to circumvent this problem. This can easily be added to our approach as well, all that is needed is a forward modelling code for surface waves, which is part of the Computer programs in seismology package (see Appendix), for example, and another misfit functional. With a more stable seismic component in the joint model we might be able to obtain information about the LAB in the examples presented above.

We did not explore the possibilities of different types of coupling between seismic velocities and resistivities, mainly because we did not want to restrict the algorithm too much. Now that we have some first results, we can start to experiment with different types of coupling, either by putting more restrictions on the interfaces, e.g. requiring both parameters to change, to change in the same direction, to have the same curvature, or by inverting for a set of meta-parameters and calculate velocities and resistivities from these parameters. The simplest way to achieve this is to use a simple linear or logarithmic function to calculate the model parameters, but this is most likely not appropriate. More interesting and potentially more rewarding is the approach to use mineral composition and calculate velocities and resistivities from it, as has been recently done for electromagnetic data [Kopylova et al., 2004, Khan et al., 2006, Bagdassarov et al., 2007].

The third obvious extension is to increase the model dimensions. The one-dimensional approach we presented here is limited by the fact that most data requires more complete models. During the inspection of the Slave craton data we had to discard the majority of sites because they showed clear signs of two-dimensional structures in the magnetotelluric data. Even the two sites we presented, contained segments were a one-dimension model was not fully appropriate. This extension is the most difficult of the three. Not only does the time to calculate a single forward model increase substantially when going from one to two dimensions, but additional issues such as model gridding and geometry for both datasets have to be considered. With increasing computational power and large parallel clusters the increase in complexity will likely be a nuisance rather than a serious problem, and two-dimensional inversion of MT data with a genetic algorithm has already been performed [Everett and Schultz, 1993]. Finding appropriate grids and providing appropriate regularization and coupling between the models is much more challenging. A first step in this direction that circumvents at least the gridding problem would be to include anisotropy in the one-dimensional models. We already discussed the issue of seismic anisotropy for the receiver function data, and electrical anisotropy is also a well know, albeit controversial, issue [Wannamaker, 2005]. This would require good azimuthal coverage for the receiver function data and would raise the question in how far an interpretation of single station MT data in terms of anisotropy is appropriate.

Bibliography

C. J. Ammon, G. E. Randall, and G. Zandt. On the nonuniqueness of receiver function inversions. *Journal of Geophysical Research*, 95:15303–15318, 1990.

D. B. Avdeev. Three-dimensional electromagnetic modelling and inversion from theory to application. *Surveys in Geophysics*, 26(6):767–799, 2005.

A. Avdeeva and D. Avdeev. A limited-memory quasi-newton inversion for 1d magnetotellurics. *Geophysics*, 71(5):G191–G196, 2006.

G. E. Backus and J. F. Gilbert. The Resolving Power of Gross Earth Data. *Geophysical Journal*, 16:169–205, 1968.

N. S. Bagdassarov, M. G. Kopylova, and S. Eichert. Laboratory derived constraints on electrical conductivity beneath Slave craton. *Physics of the Earth and Planetary Interiors*, 161:126–133, Apr. 2007. doi: 10.1016/j.pepi.2007.01.001.

K. Bahr. Geological noise in magnetotelluric data: a classification of distortion types. *Physics of the Earth and Planetary Interiors*, 66:24–38, 1991.

K. Bahr, M. Smirnov, E. Steveling, and BEAR Working Group. A gelation analogy of crustal formation derived from fractal conductivity structures. *Journal of Geophysical Research*, 107(B11):18, 2002.

C. G. Bank, M. G. Bostock, R. M. Ellis, and J. F. Cassidy. A reconnaissance teleseismic study of the upper mantle and transition zone beneath the archean slave craton in nw canada. *Tectonophysics*, 319(3):151–166, 2000.

M. N. Berdichevskiy and V. I. Dmitriev. Basic principles of interpretation of magnetotelluric sounding curves. In A. Adam, editor, *Geoelectric and geothermal studies*, pages 165–221. Budapest, Akademai Kiado.

M. N. Berdichevsky. Marginal Notes On Magnetotellurics. *Surveys in Geophysics*, 20:341–375, 1999.

Bibliography

K. A. Berteussen. Moho depth determinations based on spectral-ratio analysis of NORSAR long-period P waves. *Physics of the Earth and Planetary Interiors*, 15:13–27, 1977.

H. M. Bibby, T. G. Caldwell, and C. Brown. Determinable and non-determinable parameters of galvanic distortion in magnetotellurics. *Geophysical Journal International*, 163:915–930, 2005.

M. G. Bostock. Anisotropic upper-mantle stratigraphy and architecture of the slave craton. *Nature*, 390:392–395, 1997.

R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu. A limited-memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing*, 16: 1190–1208, 1995.

L. Cagniard. Basic theory of the magneto-telluric method of geophysical prospecting. *Geophysics*, 18:605–635, 1953.

T. G. Caldwell, H. M. Bibby, and C. Brown. The magnetotelluric phase tensor. *Geophysical Journal International*, 158:457–469, 2004.

T. Cavaliere and A. G. Jones. On the identification of a transition zone in electrical conductivity between the lithosphere and asthenosphere: a plea for more precise phase data. *Journal of Geophysics*, 55:23–30, 1984.

A. D. Chave and D. J. Thomson. Some comments on magnetotelluric response function estimation. *Journal of Geophysical Research (Solid Earth)*, 94(B10): 14215–14225, 1989.

A. D. Chave and D. J. Thomson. A bounded influence regression estimator based on the statistics of the hat matrix. *Applied Statistics*, 52(3):307–322, 2003.

A. D. Chave and D. J. Thomson. Bounded influence magnetotelluric response function estimation. *Geophysical Journal International*, 157:988–1006, 2004.

C. W. Chen, S. Rondenay, D. S. Weeraratne, and D. B. Snyder. New constraints on the upper mantle structure of the Slave craton from Rayleigh wave inversion. *Geophysical Research Letters*, 34:L10301, 2007.

R. M. Clowes, T. C. Hammer, G. Fernandez-Viejo, and J. K. Welford. Lithospheric structure in northwestern canada from lithoprobe seismic refraction and related studies: a synthesis. *Canadian Journal of Earth Sciences*, 42:1277–1293, 2005.

S. C. Constable, R. L. Parker, and C. G. Constable. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data. *Geophysics*, 52(3):289–300, 1987.

F. A. Cook and A. G. Jones. Seismic reflections and electrical conductivity: A case of Holmes's curious dog. *Geology*, 23:141–144, 1995.

G. Dal Moro and M. Pipan. Joint inversion of surface wave dispersion curves and reflection travel times via multi-objective evolutionary algorithms. *Journal of Applied Geophysics*, 61:56–81, 2007.

K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, 2002.

V. I. Dmitriev and M. N. Berdichevsky. The fundamental model of magnetotelluric sounding. *Proceedings of the IEEE*, 67(7):1034–1044, 1979.

A. M. Dziewonski and D. L. Anderson. Preliminary reference earth model. *Physics of the Earth and Planetary Interiors*, 25(4):297–356, 1981.

G. D. Egbert and J. R. Booker. Robust estimation of geomagnetic transfer functions. *Geophysical Journal of the Royal Astronomical Society*, 87(2):173–194, 1986.

M. E. Everett and A. Schultz. Two-dimensional nonlinear magnetotelluric inversion using a genetic algorithm. *J. Geomag. Geoelectr.*, 45:1013–1026, 1993.

C. G. Farquharson and D. W. Oldenburg. Non-linear inversion using general measures of data misfit and model structure. *Geophysical Journal International*, 134:213–227, 1998.

C. G. Farquharson and D. W. Oldenburg. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems. *Geophysical Journal International*, 156(1):411–425, 2004.

G. Fernandez Viejo and R. M. Clowes. Lithospheric structure beneath the archaean slave province and proterozoic wopmay orogen, northwestern canada, from a lithoprobe refraction/wide-angle reflection survey. *Geophysical Journal International*, 153(1):1–19, 2003.

R. Fletcher. *Practical methods of optimization*, volume 1. Wiley, New York, 1980.

K. Fuchs and G. Muller. Computation of synthetic seismograms with the reflectivity method and comparisons with observations. *Geophysical Journal of the Royal Astronomical Society*, 23:713–433, 1971.

L. A. Gallardo and M. A. Meju. Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data. *Geophysical Research Letters*, 30(13):1658, 2003.

T. D. Gamble, W. M. Goubau, and J. Clarke. Magnetotellurics with a remote magnetic reference. *Geophyiscs*, 44(1):53–68, 1979a.

T. D. Gamble, W. M. Goubau, and J. Clarke. Error analysis for remote reference magnetotellurics. *Geophysics*, 44(5):959–968, 1979b.

D. E. Goldberg. *Genetic algorithms in search, optimization & machine learning*. Addison Wesley, 1989.

W. L. Griffin, B. J. Doyle, C. G. Ryan, N. J. Pearson, S. Y. O'Reilly, R. Davies, K. Kivi, E. van Achterbergh, and L. M. Natapov. Layered mantle lithosphere in the lac de gras area, slave craton: Composition, structure and origin. *Journal of Petrology*, 40(5):705–727, 1999.

R. W. Groom and K. Bahr. Corrections for near surface effects: Decomposition of the magnetotelluric impedance tensor and scaling corrections for regional resistivities: A tutorial. *Surveys in Geophysics*, 13(4-5):341–379, 1992.

R. W. Groom and R. C. Bailey. Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion. *Journal of Geophysical Research (Solid Earth)*, 94(B2):1913–1925, 1989.

H. Gurrola, G. E. Baker, and J. B. Minster. Simultaneous time-domain deconvolution with application to the computation of receiver functions. *Geophysical Journal International*, 120:537–543, 1995.

E. Haber and D. W. Oldenburg. Joint inversion: A structural approach. *Inverse Problems*, 13:63–77, 1997.

P. C. Hansen. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Rev.*, 34(4):561–580, 1992. ISSN 0036-1445.

N. A. Haskell. Crustal reflections of plane p and sv waves. *Jour. Geophys. Res.*, 67(12):4751–4767, 1962.

M. Hattingh. The use of data-adaptive filtering for noise removal on magnetotelluric data. *Physics of the Earth and Planetary Interiors*, 53:239–254, 1989.

S. Haykin. *Adaptive Filter Theory*. Pearson Education, 2002.

D. V. Helmberger. Generalized ray theory for shear dislocations. *Bull. Seism. Soc. Am.*, 64:45–64, 1974.

B. A. Hobbs. Analytical solutions to global and local problems of electromagnetic induction in the earth. *Physics of the Earth and Planetary Interiors*, 10: 250–261, 1975.

J. H. Holland. *Adaptation in natural and artificial systems*. The University of Michigan Press, Ann Arbor, 1975.

M. Jervis, P. L. Stoffa, and M. K. Sen. 2-D migration velocity estimation using a genetic algorithm. *Geophysical Research Letters*, 20:1495–1498, 1993.

S. Ji, S. Rondenay, M. Mareschal, and G. Senechal. Obliquity between seismic and electrical anisotropies as a potential indicator of movement sense for ductile shear zones in the upper mantle. *Geology*, 24(11):1033–1036, 1996.

A. G. Jones. Geomagnetic induction studies in scandinavia: I. determination of the inductive response function from the magnetometer array data. *J. Geophys.*, 48:181–194, 1980.

A. G. Jones. The problem of current channelling: A critical review. *Surveys in Geophysics*, 6:079–122, 1983.

A. G. Jones. Mt and seismic reflection: an essential combination. *Geophysical Journal of the Royal Astronomical Society*, 89:7–18, 1987.

A. G. Jones. Electrical properties of the lower continental crust. In D. M. Fountain, editor, *Continental Lower Crust*, pages 81–131. Elsevier, Amsterdam, 1992.

A. G. Jones. Waves of future: Superior inferences from collocated seismic and electromagnetic experiments. *Tectonophysics*, 286:273–298, 1998.

A. G. Jones and I. J. Ferguson. The electric Moho. *Nature*, 390:331–333, 1997.

A. G. Jones and H. Jödicke. Magnetotelluric transfer function estimation improvement by a coherence-based rejection technique. 54th Annual International Meeting, Soc. of Expl. Geophys, Expanded Abstracts, 1984.

A. G. Jones and J. Spratt. A simple method for deriving the uniform field MT responses in auroral zones. *Earth Planets Space*, 54:443–450, 2002.

A. G. Jones, A. D. Chave, G. D. Egbert, D. Auld, and K. Bahr. A comparison of techniques for magnetotelluric response function estimation. *Journal of Geophysical Research (Solid Earth)*, 94(B10):14201–14213, 1989.

A. G. Jones, I. J. Ferguson, A. D. Chave, R. L. Evans, and G. W. McNeice. Electric lithosphere of the Slave craton. *Geology*, 29(5):423–426, 2001.

A. G. Jones, W. J. Davis, W. Bleeker, H. Grutter, P. Lezaeta, A. D. Chave, R. L. Evans, and I. J. Ferguson. Lithosphere development of the Slave Craton: A linked crustal and mantle perspective. *EGS - AGU - EUG Joint Assembly, Abstracts from the meeting held in Nice, France, 6 - 11 April 2003, abstract*, page 12444, 2003.

A. G. Jones, P. Lezaeta, I. j. Ferguson, A. D. Chave, R. L. Evans, X. Garcia, and J. Spratt. The electrical structure of the Slave craton. *Lithos*, 71:505–527, 2003.

J. Julia, C. J. Ammon, R. B. Herrmann, and A. M. Correig. Joint inversion of receiver function and surface wave dispersion observations. *Geophysical Journal International*, 143(1):99–112, 2000.

A. Junge. Characterisation of and correction for cultural noise. *Surveys in Geophysics*, 17:361–391, 1996.

T. Kalscheuer and L. B. Pedersen. A non-linear truncated svd variance and resolution analysis of two-dimensional magnetotelluric models. *Geophysical Journal International*, 169(2):435–447, 2007.

B. L. N. Kennett and E. R. Engdahl. Traveltimes for global earthquake locations and phase identification. *Geophysical Journal International*, 105(1):429–465, 1991.

B. L. N. Kennett, E. R. Engdahl, and R. Buland. Constraints on seismic velocities in the earth from traveltimes. *Geophysical Journal International*, 122(1): 108–124, 1995. doi: 10.1111/j.1365-246X.1995.tb03540.x.

A. Khan, J. A. D. Connolly, and N. Olsen. Constraining the composition and thermal state of the mantle beneath Europe from inversion of long-period electromagnetic sounding data. *Journal of Geophysical Research (Solid Earth)*, 111(B10):B10102, 2006.

R. Kind and L. P. Vinnik. The upper-mantle discontinuities underneath the grf array from p-to-s converted waves. *Journal of Geophysics*, 62:138–147, 1988.

Bibliography

M. G. Kopylova, J. Lo, and N. I. Christensen. Petrological constraints on seismic properties of the Slave upper mantle (Northern Canada). *Lithos*, 77:493–510, 2004.

M. Landes, J. R. R. Ritter, V. C. Do, P. W. Readman, and B. M. O'Reilly. Passive Teleseismic Experiment Explores the Deep Subsurface of Southern Ireland. *EOS Transactions*, 85:337–340, 2004.

C. A. Langston. Structure under Mount Rainier, Washington, inferred from teleseismic body waves. *Journal of Geophysical Research*, 84:4749–4762, 1979.

J. C. Larsen. Low frequency (0.1 - 6.0 cpd) electromagnetic study of deep mantle electrical conductivity beneath the hawaiian islands. *Geophysical J.R. Astron. Soc.*, 43:17–46, 1975.

T. Lay and T. C. Wallace. *Modern Global Seismology*. Academic Press, 1995.

J. Ledo, A. G. Jones, I. J. Ferguson, and L. Wolynec. Lithospheric structure of the Yukon, northern Canadian Cordillera, obtained from magnetotelluric data. *Journal of Geophysical Research (Solid Earth)*, 109(B18):B04410, 2004.

P. F. Lezaeta, A. D. Chave, and R. L. Evans. Correction of shallow-water electromagnetic data for noise induced by instrument motion. *Geophysics*, 70(5): G127–G133, 2005.

J. P. Ligorria and C. J. Ammon. Iterative deconvolution and receiver-function estimation. *Bulletin of the Seismological Society of America*, 89(5):1395–1400, 1999.

N. Linde, A. Binley, A. Tryggvason, L. B. Pedersen, and A. Revil. Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data. *Water Resources Research*, 42:12404, 2006.

T. Madden and P. Nelson. A defense of Cagniard's magnetotelluric method. Report for Office of Naval Research. Published in Vozoff (1966), pages 89 - 95, 1963.

C. Manoj and N. Nagarajan. The application of artificial neural networks to magnetotelluric time-series analysis. *Geophysical Journal International*, 153: 409–423, 2003.

M. Mareschal. Modelling of natural sources of magnetospheric origin in the interpretation of regional induction studies: a review. *Surveys in Geophysics*, 8: 261–300, 1986.

A. Martí, P. Queralt, A. G. Jones, and J. Ledo. Research note: Improving bahr's invariant parameters using the wal approach. *Geophysical Journal International*, 163:38–41, 2005.

U. Meier, A. Curtis, and J. Trampert. Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, 169(2): 706–722, 2007.

M. A. Meju, L. A. Gallardo, and A. K. Mohamed. Evidence for correlation of electrical resistivity and seismic velocity in heterogeneous near-surface materials. *Geophysical Research Letters*, 30:26–1, 2003.

W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory*, volume 45 of *International Geophysics Series*. Academic Press, 1989.

G. Muñoz and V. Rath. Beyond smooth inversion: the use of nullspace projection for the exploration of non-uniqueness in mt. *Geophysical Journal International*, 164:301–311, 2006.

B. B. Narod and J. R. Bennest. Ring-core fluxgate magnetometers for use as observatory variometers. *Physics of the Earth and Planetary Interiors*, 59: 23–28, 1990.

J. J. Nocedal. Theory of algorithms for unconstrained optimization. *Acta Numerica*, 1:199–242, 1992.

I. L. Osipova, S. E. Hjelt, and L. L. Vanyan. Source field problems in northern parts of the Baltic Shield. *Physics of the Earth and Planetary Interiors*, 53: 337–342, 1989.

T. J. Owens, G. Zandt, and S. R. Taylor. Seismic evidence for an ancient rift beneath the cumberland plateau, Tennessee: A detailed analysis of broadband teleseismic P waveforms. *Journal of Geophysical Research*, 89:7783–7796, 1984.

M. B. Padua, A. L. Padilha, and I. Vitorello. Disturbances on magnetotelluric data due to dc electrified railway: A case study from southeastern brazil. *Earth Planets Space*, 54:591–596, 2002.

S. K. Park and D. W. Livelybrooks. Quantitative interpretation of rotationally invariant parameters in magnetotellurics. *Geophysics*, 54(11):1483–1490, 1989.

R. L. Parker and J. R. Booker. Optimal one-dimensional inversion and bounding of magnetotelluric apparent resistivity and phase measurements. *Physics of the Earth and Planetary Interiors*, 98:269–282, 1996.

L. B. Pedersen. Determination of the regularization level of truncated singular-value decomposition inversion: The case of 1d inversion of mt data. *Geophysical Prospecting*, 52(4):261–270, 2004.

L. B. Pedersen and M. Engels. Routine 2D inversion of magentotelluric data using the determinant of the impedance tensor. *Geophysics*, 70(2):G33–G41, 2005.

M. A. Pérez-Flores and A. Schultz. Application of 2-D inversion with genetic algorithms to magnetotelluric data from geothermal areas. *Earth, Planets, and Space*, 54:607–616, 2002.

H. K. C. Perry, D. W. Eaton, and A. M. Forte. LITH5.0: a revised crustal model for Canada based on Lithoprobe results. *Geophysical Journal International*, 150:285–294, 2002.

R. E. Plessix. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503, 2006.

A. T. Price. The theory of magnetotelluric methods when the source field is considered. *Journal of Geophysical Research*, 67(5):1907–1918, 1962.

W. Qian and L. B. Pedersen. Industrial interference magnetotellurics: An example from the tangshan area, china. *Geophysics*, 56(2):265–273, 1991.

A. Raiche. A pattern recognition approach to geophysical inversion using neural networks. *Geophysical Journal International*, 105:629–648, 1991.

G. E. Randall. Efficient calculation of differential seismograms for lithospheric receiver functions. *Geophysical Journal International*, 99:469–481, 1989.

C. A. Rychert, K. M. Fischer, and S. Rondenay. A sharp lithosphere-asthenosphere boundary imaged beneath eastern north america. *Nature*, 436:542–545, 2005. ISSN 0028-0836.

M. Sambridge and K. Mosegaard. Monte carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40:3–1, 2002.

M. Sambridge, C. Beghein, F. J. Simons, and R. Snieder. How do we understand and visualize uncertainty? *The Leading Edge*, 25(5):542–546, 2006.

M. Sambridge, P. Rickwood, N. Rawlinson, and S. Sommacal. Automatic differentiation in geophysical inverse problems. *Geophysical Journal International*, 170:1–8, 2007.

J. H. Saunders, J. V. Herwanger, C. C. Pain, M. H. Worthington, and C. R. E. de Oliveira. Constrained resistivity inversion using seismic data. *Geophysical Journal International*, 160(3):785–796, 2005.

U. Schmucker. Regional induction studies: a review of methods and results. *Physics of the Earth and Planetary Interiors*, 7:365–378, 1973.

P. A. Schnegg and G. Fischer. On-line determination of apparent resistivity in magnetotelluric soundings. In J. Homilius, editor, *Kolloquium Elektromagnetische Tiefenforschung*, pages 173–184, Berlin-Lichtenrade, 1980.

W. O. Schumann. Luftelektrische Störungen (atmospherics) und elektrische langwellenausbreitung um die Erde. *Kleinheubacher Berichte*, 1:139, 1954.

C. Schwarzbach, R. U. Börner, and K. Spitzer. Two-dimensional inversion of direct current resistivity data using a parallel, multi-objective genetic algorithm. *Geophysical Journal International*, 162:685–695, 2005.

SHARCNET. Sharcnet homepage. http://www.sharcnet.ca.

T. Shibutani, M. Sambridge, and B. Kennett. Genetic algorithm inversion for reciever functions with application to crust and uppermost mantle structure beneath eastern australia. *Geophysical Research Letters*, 20(24):1829–1832, 1993.

F. Simpson and K. Bahr. *Practical Magnetotellurics*. Cambridge University Press, 2005.

W. E. Sims, F. X. Bostick, and H. W. Smith. The estimation of magnetotelluric impedance tensor elements from measured data. *Geophysics*, 36(5):938–942, 1971.

B. S. Singer. Correction for distortions of magnetotelluric fields: Limits of validity of the static approach. *Surveys in Geophysics*, 13(4-5):309–340, 1992.

W. Siripunvaraporn, G. Egbert, and Y. Lenbury. Numerical accuracy of magnetotelluric modeling: A comparison of finite difference approximations. *Earth, Planets, and Space*, 54:721–725, 2002.

W. Siripunvaraporn, G. Egbert, Y. Lenbury, and M. Uyeshima. Three-dimensional magnetotelluric inversion: data-space method. *Physics of the Earth and Planetary Interiors*, 150:3–14, 2005.

D. B. Snyder and M. Bruneton. Seismic anisotropy of the Slave craton, NW Canada, from joint interpretation of SKS and Raleigh waves. *Geophysical Journal International*, 169:170–188, 2007.

D. B. Snyder, S. Rondenay, M. G. Bostock, and G. D. Lockhart. Mapping the mantle lithosphere for diamond potential using teleseismic methods. *Lithos*, 77: 859–872, 2004.

D. Sutarno and K. Vozoff. Phase-smoothed robust m-estimation of magnetotelluric impedance functions. *Geophysics*, 56(12):1999–2007, 1991.

C. M. Swift. *A magnetotelluric investigation of an electrical conductivity anomaly in the Southwestern United States*. PhD thesis, MIT, Cambridge, MA, 1967.

L. Szarka. Geophysical aspects of man-made electromagnetic noise in the earth - a review. *Surveys in Geophysics*, 9:287–318, 1988.

A. Tarantola. *Inverse Problem Theory*. SIAM, 1 edition, 2004. ISBN 0-89871-572-5.

D. J. Thomson. Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, 70(9):1055–1096, 1982.

A. N. Tikhonov. The determination of electrical properties of the deep layers of the earth's crust. *Dokl. Acad. Nauk. SSR*, 73:295–297, 1950. In russian.

A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. Wiley-Interscience, New York, 1977.

L. P. Vinnik. Detection of waves converted from P to SV in the mantle. *Physics of the Earth and Planetary Interiors*, 15:39–45, 1977.

K. Vozoff and D. L. B. Jupp. Joint Inversion of Geophysical Data. *Geophysical Journal*, 42:977–991, 1975.

J. R. Wait. On the relation between telluric currents and the earth's magnetic field. *Geophysics*, 19:281–289, 1954.

P. E. Wannamaker. Anisotropy Versus Heterogeneity in Continental Solid Earth Electromagnetic Studies: Fundamental Response Characteristics and Implications for Physicochemical State. *Surveys in Geophysics*, 26:733–765, 2005.

J. T. Weaver, B. V. Le Quang, and G. Fischer. A comparison of analytic and numerical results for a two-dimensional control model in electromagnetic induction - i. b-polarization calculations. *Geophysical J.R. Astron. Soc.*, 82: 263–277, 1985.

J. T. Weaver, B. V. Le Quang, and G. Fischer. A comparison of analytic and numerical results for a 2-d control model in electromagnetic induction - ii. e-polarization calculations. *Geophysical J.R. Astron. Soc.*, 87:917–948, 1986.

J. T. Weaver, A. K. Agarwal, and F. E. M. Lilley. Characterization of the magnetotelluric tensor in terms of its invariants. *Geophysical Journal International*, 141:321–336, 2000.

P. Weidelt. The inverse problem of geomagnetic induction. *Zeitschrift für Geophysik*, 8(2):257–290, 1972.

P. Weidelt. Construction of conductance bounds from magnetotelluric impedances. *Journal of Geophysics*, 57:191–206, 1985.

P. Weidelt and P. Kaikkonen. Local 1-d interpretation of magnetotelluric b-polarization impedance. *Geophysical Journal International*, 117:733–748, 1994.

C. M. Weiland, L. K. Steck, P. B. Dawson, and V. A. Korneev. Nonlinear teleseismic tomography at Long Valley caldera, using three-dimensional minimum travel time ray tracing. *Journal of Gephysical Research*, 100:20379–20390, 1995.

C. R. Whaley, A. Petitet, and J. J. Dongarra. Automated empirical optimization of software and the ATLAS project. *Parallel Computing*, 27(1–2):3–35, 2001.

B. Widrow and S. D. Stearns. *Adaptive Signal Processing*. Pearson Education, 1985.

B. Widrow, J. R. Glover, J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, and R. C. Goodlin. Adaptive noise cancelling: Principles and applications. *Procedings of the IEEE*, 63(12):1692–1716, 1975.

X. Yuan, R. Kind, X. Li, and R. Wang. The s receiver functions: synthetics and data example. *Geophysical Journal International*, 165(1):555–564, 2006.

M. Zhdanov and E. Tolstaya. Minimum support nonlinear parametrization in the solution of a 3d magnetotelluric inverse problem. *Inverse Problems*, 20(3): 937–952, 2004.

A

Software

This work could not have been done without the use of a variety of different software. Most of it is freely available and I would like to thank the authors for generously sharing their work with others, and letting me benefit from their efforts. As my code becomes more stable and as I have time, I will release it so that others can benefit from my work. The latest information can be found at `http://gplib.sourceforge.net`.

The following list of programs I used is probably not complete, but I have tried to give credit to everything I have used more than just occassionally.

antlr: This freely available parser generator was used to solve the tricky problem of reading in EDI files and the somewhat less tricky problem to read J-Format files. It also ensures a flexible and robust way of reading in the configuration files for various programs. Look at `http://www.antlr.org` for more information.

asymptote: A few of the figures in this thesis were prepared using this open-source illustration language. See `http://asymptote.sourceforge.net`.

atlas: Changing the matrix operations in my RLS-adaptive filter to the atlas library, improved the speed of the program by a factor of 20. These linear algebra routines are not only automatically, but truly highly tuned. With the bindings for boost, it is also just another function call. You can get the library from `http://math-atlas.sourceforge.net`.

179

beamer: If you think of presentations and think of Powerpoint, think again. This LaTeXpackage does not only produce much nicer looking presentations, they are also smaller and do not have the annoying font problems Powerpoint has. `http://latex-beamer.sourceforge.net`.

birrp: Some of the MT data, particularly the sites with high noise, were processed using this state of the art processing code. It is available from `achave@whoi.edu`.

Boost: The high-quality, peer reviewed, and open source C++ libraries from `http://www.boost.org` are used in virtually every program that was written during this thesis. Starting from the ublas matrix and vector package, the date and time functionality and the filesystem library to name just a few, this work would not have been possible in this form without these libraries.

cfmake: This is actually my own creation, but it is released separately, as it has potential use for people that have nothing to do with geophysics. It automatically creates C++ classes that read in configuration files. If you find yourself continuously programming something like that, have a look at `http://cfmake.sourceforge.net`.

dot: The dot language (`http://www.graphviz.org`) was used to automatically create figures that show the topology of neural networks. It is also used in doxygen.

doxygen: The automatically generated documentation from doxygen (`http://www.doxygen.org`) is probably more useful to other users of the program than me, but they greatly helped to keep an overview of the ever growing files and classes.

Eclipse: All programming was done within the Eclipse IDE. Together with the subclipse plugin and the CDT this provided everything I could ask for in terms of development environment and more.
Go to `http://www.eclipse.org`.

fftw: All fourier transforms were done using the *Fastest Fourier Transform in the West* library, available from `http://www.fftw.org`.

GMT: No practical geophysics thesis without the Generic Mapping Tools. In addition to the standard map plots these were also used for plotting netcdf data, as seen in Figure 6.1, for example. `http://gmt.soest.hawaii.edu` has the source code as well as excellent documentation and examples.

gsac: This program is part of the Computer Programs in Seismology package (`http://www.eas.slu.edu/People/RBHermann/ComputerPrograms.html`), and is considered an alternative and updated version to the popular SAC package (see below).

GSL: The `Gnu Scientific library` was also used in various programs; most importantly for resampling and numerical precision checking.
`http://www.gnu.org/software/gsl/`

kdissert: Thomas Nagy's minmapping software helped to organize spontaneous ideas. If you use it right, it will even write your thesis (`http://freehackers.org/~tnagy/kdissert`).

kile: This integrated environment for writing LaTeX documents provided me with a nice and easy to use tool that makes writing a joy. It might not be as hard-core as using *vi* or *ed*, but it makes life a lot easier. It is part of KDE, so it can be obtained from `http://www.kde.org`.

Lapack: Without *lapack* calculating singular value compositions would have taken much more time, both to execute the code, but also to develop without major bugs. This library is the reason, that all matrices are column major format. Long live Fortran ! (`http://www.netlib.org/lapack`).

latex: Obvious to the trained eye, this thesis was written using LaTeX 2_ε and the concrete package plus a few more. Thanks to Leslie Lamport and Donald Knuth for saving me from using W#@%. Probably the best place to get yourself informed about LaTeX and TeX is `http://www.tug.org`.

levmar: Even though I implemented my own linearized inversion routines, there is a huge difference between the core optimization method and a full-blown optimization library with all bells and whistles that handles all types of possible cases correctly. Therefore I performed the practical work with linearized algorithms using the levmar library available at
`http://www.ics.forth.gr/~lourakis/levmar/`.

netcdf: *netcdf* is not only a library, but more importantly a standardized data format for scientific data. Using the provided C-functions writing netcdf files is very easy and these can then be used with GMT, ncview, visit or other visualization tools. More information can be found at
`http://www.unidata.ucar.edu/software/netcdf`.

NSGA-II: I didn't use this software for my inversion algorithm, but wrote my own implementation. Having the possibility to look at the sourcecode and

see how "they" have done it, was certainly very helpfull and saved me some bug hunting. You can download the original NSGA-II from `http://www.iitk.ac.in/kangal/soft.htm`.

respktn: I used this program to compute all the synthetic seismograms in this thesis, apart from the ones shown in Figure 4.2, which were calculated with a spectral code. This is part of the Computer Programs in Seismology (see gsac).

sac: The Seismic Analysis Code helps to perform all standard and some not so standard operations `http://www.llnl.gov/sac/`.

scons I never got the hang of make files, so I decided to give this build tool a go and I achieved what I tried for weeks before within half an hour. `http://www.scons.org` has more if you need to build libraries that are more than just a couple of files.

Subversion: A version management system might seem overkill for a single developer, but I try to compensate my lack of organisation by using this immensely useful development tool. Trace the evolution of your files and find out where things finally went completely wrong. `http://subversion.tigris.org`

taup: This collection of Java programs calculates traveltimes from different global models, displays ray paths and a lot of other usefull seismological parameters. Without the help of this valuable little tool teleseismic analyses would be much more difficult.
You can find it at `http://www.seis.sc.edu/software/TauP`.

Ubuntu: My favorite Linux-distribution. Easy to set up, easy to use and fairly complete for all scientific computing needs. Thanks a lot to all contributors. The latest Packages are available at `http://www.ubuntu.com`.

valgrind Memory leaks and access violations can be the most annoying and difficult to find bugs in any program with dynamic memory allocation. With the help of this little tool you can not only find those fairly easily, but can also get all kinds of different diagnostics on the performance of your programs. `http://valgrind.org`.

wikindx: This web-based reference management tool not only helps you keeping track of all your bibliographic references, but also allows you to attach the corresponding paper and share the information between different users. You can even write papers with it. `http://wikindx.sourceforge.net`

xfig: It might look like it comes straight from the eighties (which it does), but underneath the shabby surface you can find an excellent drawing program. Less trouble with importing ps, pdf and other files and drawing on them, then corel draw, illustrator and all that "modern" stuff. There is not a lot of recent improvement, but it doesn't need any, it's perfect. `http://www.xfig.org`

xmgrace: Most of the plots in this thesis were done using xmgrace. If you need to plot xy data of any kind, with or without error bars, I cannot imagine a better plotting tool. It can even be invoked through C-library calls, to give your programs their own plot window. However I prefer the old style version 5, but that's just me getting old.
Look at `http://plasma-gate.weizmann.ac.il/Grace/` for the latest version.

B
Event parameters

No.	Date	Time	Lat	Long	Depth [km]	M_w
1	2002/06/22	02:58:21.30	35.6260	49.0470	10.00	6.50
2	2002/08/02	23:11:39.13	29.2800	138.9700	426.10	6.30
3	2002/10/12	20:09:11.46	-8.2950	-71.7380	534.30	6.90
4	2002/11/17	04:53:53.54	47.8240	146.2090	459.10	7.30
5	2003/01/21	02:46:47.74	13.6260	-90.7740	24.00	6.50
6	2003/01/22	02:06:34.61	18.7700	-104.1040	24.00	7.60
7	2003/03/17	16:36:17.31	51.2720	177.9780	33.00	7.10
8	2003/04/17	00:48:38.58	37.5290	96.4760	14.00	6.40
9	2003/04/27	22:57:44.84	-8.1950	-71.5920	559.90	6.00
10	2003/05/14	06:03:35.86	18.2660	-58.6330	41.50	6.70
11	2003/05/19	16:27:10.20	17.5460	-105.4730	10.00	6.10
12	2003/05/21	18:44:20.10	36.9640	3.6340	12.00	6.80
13	2003/05/26	09:24:33.40	38.8490	141.5680	68.00	7.00
14	2003/06/20	06:19:38.91	-7.6060	-71.7220	558.10	7.10
15	2003/06/23	12:12:34.47	51.4390	176.7830	20.00	6.90
16	2003/07/25	22:13:29.97	38.4150	140.9960	6.00	6.10
17	2003/07/27	06:25:31.95	47.1510	139.2480	470.30	6.80
18	2003/08/25	23:24:59.13	18.5410	-106.6950	10.00	5.80
19	2003/09/25	19:50:06.36	41.8150	143.9100	27.00	8.30
20	2003/09/29	02:36:53.14	42.4500	144.3800	25.00	6.50
21	2003/10/31	01:06:28.28	37.8120	142.6190	10.00	7.00
					Continued on next page	

No	Date	Time	Lat	Long	Depth	M_w
					Continued from previous page	
22	2003/11/12	08:26:43.74	33.1710	137.0720	384.90	6.40
23	2003/11/17	06:43:06.80	51.1460	178.6500	33.00	7.80
24	2003/12/05	21:26:09.48	55.5380	165.7800	10.00	6.70
25	2004/02/24	02:27:46.23	35.1420	-3.9970	0.00	6.40
26	2004/03/12	22:45:19.00	36.3970	70.7740	218.00	5.80
27	2004/03/27	18:47:29.20	33.9540	89.1790	8.00	6.00
28	2004/04/14	23:07:39.94	71.0670	-7.7470	12.20	6.00
29	2004/04/19	08:14:11.42	3.6140	-32.2340	10.00	5.60
30	2004/05/10	23:27:25.49	37.4850	96.6040	10.00	5.60
31	2004/05/29	20:56:09.60	34.2510	141.4060	16.00	6.50
32	2004/06/10	15:19:57.75	55.6820	160.0030	188.60	6.90
33	2004/06/14	22:54:21.32	16.3370	-97.8450	10.00	5.90
34	2004/07/22	09:45:14.90	26.4890	128.8940	20.90	6.10
35	2004/08/07	09:30:16.94	51.7530	-166.3130	8.00	6.00
36	2004/08/10	01:47:32.81	36.4440	70.7960	207.00	6.00
37	2004/08/11	15:48:26.82	38.3770	39.2610	7.40	5.70
38	2004/09/05	10:07:07.82	33.0700	136.6180	14.00	7.20
39	2004/09/05	14:57:18.61	33.1840	137.0710	10.00	7.40
40	2004/09/06	23:29:35.09	33.2050	137.2270	10.00	6.60
41	2004/10/09	21:26:53.69	11.4220	-86.6650	35.00	7.00
42	2004/10/15	04:08:50.24	24.5300	122.6940	94.00	6.70
43	2004/10/23	08:56:00.86	37.2260	138.7790	16.00	6.60
44	2004/11/15	09:06:56.56	4.6950	-77.5080	15.00	7.20
45	2004/11/28	18:32:14.13	43.0060	145.1190	39.00	7.00
46	2005/02/05	03:34:25.73	16.0110	145.8670	142.70	6.60
47	2005/06/13	22:44:33.90	-19.9870	-69.1970	115.60	7.80
48	2005/06/14	17:10:12.28	51.2390	179.3140	17.00	6.80
49	2005/07/25	16:02:07.57	71.1110	-7.4320	10.00	5.50
50	2005/07/26	12:17:14.27	52.8710	160.1050	27.60	5.80
51	2005/08/16	02:46:28.40	38.2760	142.0390	36.00	7.20
52	2005/09/21	02:25:08.11	43.8920	146.1450	103.00	6.10
53	2005/09/26	01:55:37.67	-5.6780	-76.3980	115.00	7.50
54	2005/10/08	03:50:40.80	34.5390	73.5880	26.00	7.60
55	2005/10/08	10:46:28.79	34.7330	73.1000	8.00	6.40
56	2005/10/15	15:51:07.21	25.3210	123.3560	183.40	6.50
57	2005/11/14	21:38:51.42	38.1070	144.8960	11.00	7.00
					Continued on next page	

					Continued from previous page	
No	Date	Time	Lat	Long	Depth	M_w
58	2005/12/02	13:13:09.52	38.0890	142.1220	29.00	6.50
59	2005/12/12	21:47:46.07	36.3570	71.0930	224.60	6.50

Table B.1: List of events used in the receiver function analysis for sites BOXN and EKTN.
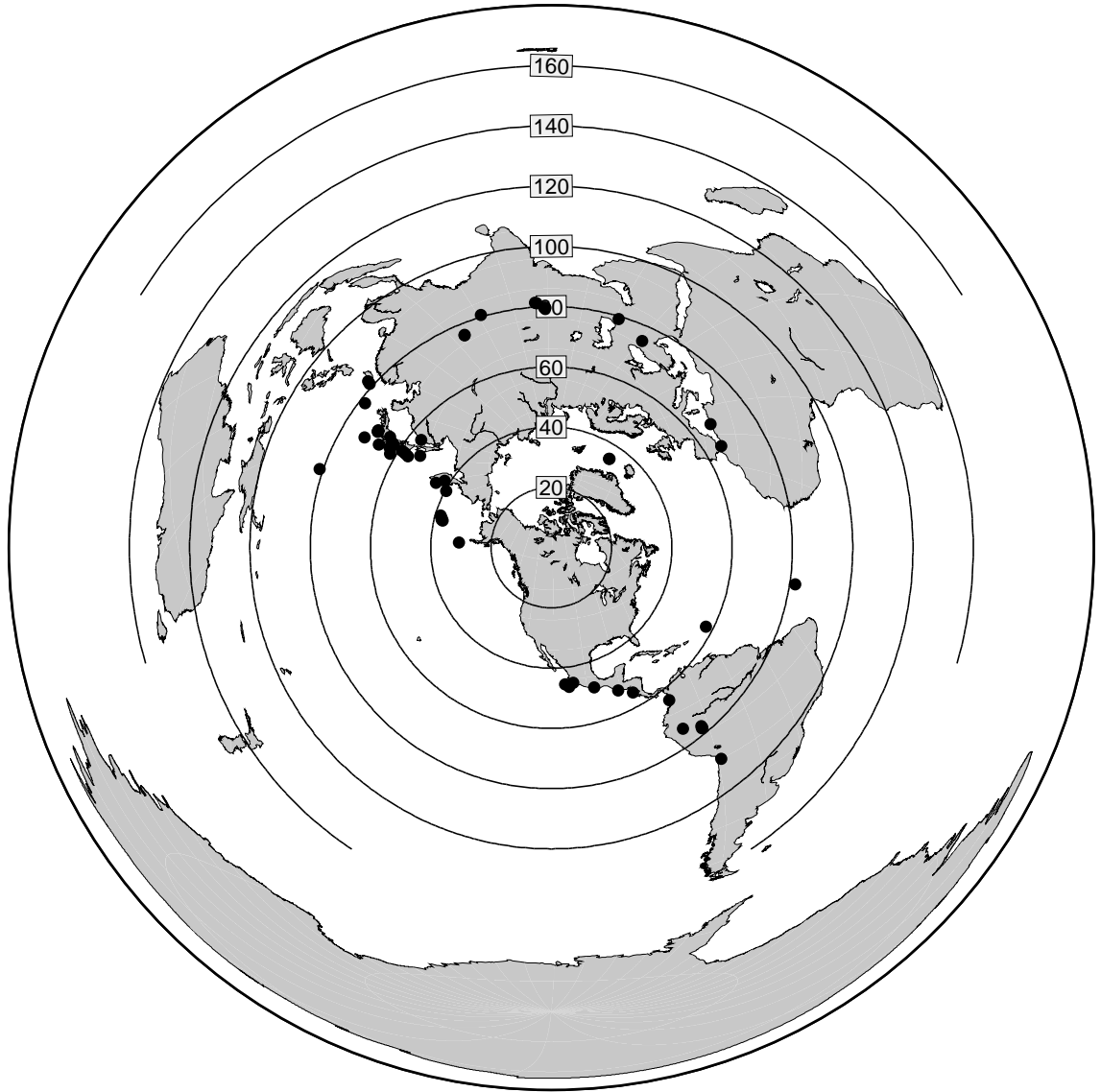
Figure B.1: Map of the events used in our receiver function analysis. The circles show the approximate epicentral distance from the two sites in degree.
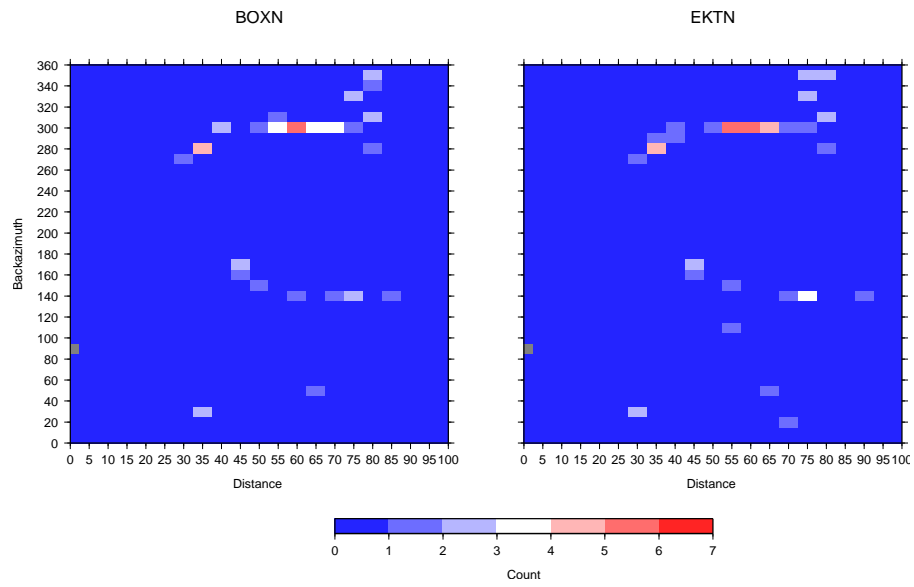
Figure B.2: Histogram of the backazimuth and distance distribution of events used in the receiver function analysis of the two Slave craton sites.

C
Publications

Peer reviewed

- M. Moorkamp, A.G. Jones and D.W. Eaton, Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: Are seismic velocities and electrical conductivities compatible ?, Geophys. Res. Let., 34, L16311, doi:10.1029/2007GL030519

- M. Moorkamp, Comment on "The magnetotelluric phase tensor", accepted for publication in GJI, doi:10.1111/j.1365-246X.2007.03490.x

- CK Rao, A.G. Jones and M. Moorkamp, The geometry of the Iapetus suture zone in central Ireland deduced from a magnetotelluric study, Physics of the Earth and Planetary Interiors, 2007, 161, 131-141

Conference abstracts

- M. Moorkamp, CK. Rao and A.G. Jones, Examination and removal of repetitive noise on MT time series, EM Induction Workshop 2004, Hyderabad

- M. Moorkamp, CK. Rao and A.G. Jones, Noise cancellation in magnetotelluric time series using adaptive methods, EGU05-A-00713, EGU General Assembly 2005, Vienna

- M. Moorkamp, CK. Rao and A.G. Jones, Joint inversion of magnetotelluric and receiver function data using a genetic algorithm, BGA 2005, Galway, talk

- M. Moorkamp, CK. Rao and A.G. Jones, Joint inversion of magnetotelluric and receiver function data EMTF-Workshop 2005, Königstein, poster

- M. Moorkamp, CK. Rao and A.G. Jones, Inverting receiver-function and MT-data with a genetic algorithm IGRM 2006, Cork, talk

- M. Moorkamp, CK. Rao and A.G. Jones, Joint inversion of MT and seismic receiver function data using a genetic algorithm BGA 2006, Edinburgh, talk

- M. Moorkamp, CK. Rao and A.G. Jones, Processing magnetotelluric time-series with adaptive filters and neural networks EM Induction Workshop 2006, Barcelona, poster

- M. Moorkamp, CK. Rao and A.G. Jones, Joint inversion of MT and seismic receiver function data using a genetic algorithm EM Induction Workshop 2006, Barcelona, poster

- A.G. Jones and M. Moorkamp, Are seismic velocities and electrical conductivities reconcilable?, S44B-01, AGU Fall meeting 2006, San Francisco, talk

- M. Moorkamp, A.G. Jones and D. W. Eaton, A lithosphere-scale relationship between electrical conductivity and seismic velocity in the Slave Craton ?, EGU2007-A-08277, EGU General Assembly 2007, Vienna, talk