GEOACOUSTIC REFLECTIVITY INVERSION: A BAYESIAN APPROACH

by

Jan Dettmer
Diplom Geophysiker, University of Hamburg, 2002

A Thesis Submitted in Partial Fullfillment of the
Requirements for the Degree of
DOCTOR OF PHILOSOPHY
in the School of Earch and Ocean Sciences

Geoacoustic Reflectivity Inversion: A Bayesian Approach

by

Jan Dettmer
Diplom Geophysiker, University of Hamburg, 2002

Supervisory Committee:

Dr. Stan E. Dosso (School of Earth and Ocean Sciences, University of Victoria)
Supervisor

Dr. N. Ross Chapman (School of Earth and Ocean Sciences, University of Victoria)
Co-Supervisor

Dr. Michael J. Wilmut (School of Earth and Ocean Sciences, University of Victoria)
Departmental Member

Dr. Adam H. Monahan (School of Earth and Ocean Sciences, University of Victoria)
Departmental Member

Dr. Michel Lefebvre (Department of Physics and Astronomy, University of Victoria)
Outside Member

Dr. L. Neil Frazer (Department of Geology and Geophysics, University of Hawaii)
External Examiner

Supervisory Committee:

Supervisor: Dr. Stan E. Dosso (University of Victoria)

Co-Supervisor: Dr. N. Ross Chapman (University of Victoria)

Departmental Member: Dr. Michael J. Wilmut (University of Victoria)

Departmental Member: Dr. Adam H. Monahan (University of Victoria)

Outside Member: Dr. Michel Lefebvre (University of Victoria)

External Examiner: Dr. L. Neil Frazer (University of Hawaii)

Abstract

Propagation and reverberation of acoustic fields in shallow water depend strongly on the spatial variability of seabed geoacoustic parameters, and lack of knowledge of seabed variability is often a limiting factor in acoustic modelling applications. However, direct sampling (e.g., coring) of vertical and lateral variability is expensive and laborious, and matched-field and other long-range inversion methods fail to provide sufficient resolution.

This thesis develops a new joint time/frequency domain inversion for high-resolution single-bounce reflection data. The inversion approach has the potential to resolve fine-scale sediment profiles over small seafloor footprints ($\sim$100 m). The approach utilises sequential Bayesian inversion of time- and frequency-domain reflectivity data, employing ray-tracing inversion for reflection travel times and a layer-packet stripping method for spherical-wave reflection coefficient inversion. Rigorous uncertainty estimation is of key importance to yield high quality inversion results. Quantitative geoacoustic uncertainties are provided by a nonlinear Gibbs sampling approach together with full data error covariance estimation (including non-stationary effects). The small footprint of the measurement technique combined with the rigorous inversion of both time and frequency domain data provides a powerful new tool to examine seabed structure on finer scales than heretofore possible.

The Bayesian inversion is applied to two data sets collected on the Malta Plateau and the Strait of Sicily during the SCARAB98 experiment. The first application aims to recover multi-layered seabed structure and the second application recovers density and sound velocity gradient structure in the uppermost sediment layer.

An interesting new method of deriving reflectivity data from ambient noise measurements is briefly considered in simulation to examine the resolving power and limits of the approach.

Table of Contents

List of Tables

List of Figures

The most general formulation of Inverse Problems is obtained when using the language of probability calculus, and when using Bayesian interpretation of probability (Bayes, 1763). Inverse Problem Theory has to be developed from the consideration of uncertainties (either experimental, or in physical laws), and the right (well-posed) question to set is: given a certain amount of (*a priori*) information on some model parameters, and given an uncertain physical law relating some observable parameters to the model parameters, in which sense should I modify the *a priori* information, given the uncertain results of some experiments?

Albert Tarantola, 1986

Acknowledgements

I would like to thank Dr. Stan E. Dosso for support and being an excellent teacher. Further, I would like to thank Drs. N. Ross Chapman and Michael J. Wilmut for many helpful discussions regarding statistics and underwater acoustics. I thank Dr. Charles W. Holland for sparking many of the ideas that lead to this thesis, for helpful discussions, and for making the measured data available that is presented in this work. This work was financially supported through Dr. Dosso's research grants, the University of Victoria Fellowship, the Howard E. Petch Research Scholarship, the Maritime Awards Society of Canada Graduate Fellowship, the Donald Wagg Graduate Scholarship, and the Fessenden Student Prize in Underwater Acoustics.

1. Introduction

1.1. Background

The shallow water sound field is generally dominated by bottom-interacting paths and is therefore strongly influenced by seabed geoacoustic properties. Hence, the spatial variability of seabed properties at fine scales is of key interest to understand shallow-water sound propagation (lack of knowledge of physical seabed parameters often represents the limiting factor in ocean applications, Ferla and Jensen, 1987). However, direct sampling of vertical and lateral variability via coring can be prohibitively laborious and expensive. Geoacoustic inversion, which infers seabed properties from acoustic measurements in the water column, represents a very promising approach.

The majority of work to date in geoacoustic inversion has involved matched-field inversion of long-range acoustic fields measured on an array of sensors (e.g., Collins et al., 1992; Dosso et al., 1993; Lindsay and Chapman, 1993; Gerstoft, 1994; Fallat and Dosso, 1998; Gerstoft and Mecklenbräuker, 1998; Fallat and Dosso, 1999; Jaschke and Chapman, 1999; Mecklenbräuker and Gerstoft, 2000; Dosso and Nielsen, 2002; Dosso, 2002b, 2003; Dosso et al., 2006; Tollefsen et al., 2006; Huang et al., 2006). The long range nature of matched-field methods ensures many bottom interactions so that the acoustic field contains information about the seabed. This also results in an averaging effect over large areas since the bottom bounces cannot be separated. One of the main challenges in matched-field inversion is modelling the acoustic field with sufficient precision over large ranges, where the effects of environmental variability often result in large theory error. Standard matched-field and other long-range inversion methods lack fine-scale vertical and lateral resolution, suffer from large theory errors, and are often found to be insensitive to certain geoacoustic parameters of interest such as density.

To address these issues, single-bounce reflection methods have been applied to

more accurately resolve local sediment structure, and they are the subject of this thesis. Hastrup (1970) used a single bounce method to resolve seabottom reflection coefficients in deep water. Holland and Osler (2000) developed an experimental technique to collect high-resolution sea-bed reflectivity data in shallow water using a bottom-moored hydrophone and ship-towed impulsive source, with a seafloor experimental footprint of ~100 m. Because of this local scale, theory errors and the effects of spatial and temporal variability in the water column and seabed are greatly reduced compared to long-range acoustic measurements.

The reflectivity experiment yields seismo-acoustic data that can be analysed in the time domain (travel time data) and/or in the frequency domain (reflection coefficient data). Travel time data can be interpreted by normal move-out methods (Telford et al., 1995; Bryan, 1980) or ray tracing (Červený, 2001). Reflection coefficient data can be interpreted in a manner similar to seismic amplitude variation with offset (AVO) data (e.g., Riedel et al., 2003). Dosso and Holland (2006) applied a rigorous Bayesian inversion approach to single bounce reflection coefficient data, recovering the physical parameters of an elastic sediment halfspace with a plane wave reflection coefficient forward model, and showed that shear properties do not significantly affect the inversion. This thesis develops a joint time/frequency domain Bayesian inversion approach to recover complex fine-scale density and compressional-wave velocity structure from broadband reflectivity data.

Bayesian inversion formulates an inverse problem in terms of the posterior probability density (PPD) of the model parameters (Tarantola, 1987; Sambridge and Mosegaard, 2002; Brooks and Frazer, 2005), incorporating both data and prior information. The solution is quantified in terms of properties of the multi-dimensional PPD representing parameter estimates, uncertainties and inter-relationships. Optimal parameter estimates require nonlinear optimisation such as adaptive hybrid inversion (Dosso et al., 2001). Parameter uncertainties (e.g., marginal distributions, credibility intervals) can be computed using Markov-chain Monte Carlo methods such as fast Gibbs sampling (Dosso, 2002a; Dosso and Holland, 2006; Holland et al., 2005).

Rigorous uncertainty estimation for geoacoustic parameters is of key importance in order to distinguish spatial variability between inversion results for nearby measurement sites from the inherent inversion uncertainties. This requires not only a nonlinear inversion approach, but also rigorous estimation of the data error statistics. In particular, the standard approach of ignoring data error correlations (covariance) is

shown to lead to under-estimation of geoacoustic uncertainties and biased, misleading parameter estimates. The approach applied here is based on non-parametric estimation of the data error covariance matrix (i.e., error variances and covariances) from residual analysis, including effects of non-stationarity with source-receiver range, and on rigorous *a posteriori* statistical tests to validate these estimates and the underlying assumptions (Dettmer et al., 2004a, 2006; Holland et al., 2005; Dosso et al., 2006).

The Bayesian time/frequency reflectivity inversion developed here is based on a two-stage (sequential Bayesian) strategy. In the time domain, travel times are picked for multiple reflectors in the seismo-acoustic data, and ray-tracing inversion is carried out for layer thicknesses and sound velocities (the number of layers is equivalent to number of sub-bottom reflectors). For the frequency-domain inversion, the seismo-acoustic time series are processed to yield reflection coefficients as a function of angle and frequency. The reflection coefficients are inverted for layer thicknesses, sound velocities, densities and attenuations, with the PPD from the time-domain inversion applied as prior information. The frequency-domain inversion utilises a layer packet stripping approach in which the data are time-windowed into packets that contain the reflections of different sets of layers from the sea-floor to a sequence of increasing depths. Frequency-domain Bayesian inversion is applied stepping down through the packets, using the results (PPD) from one packet inversion as prior information to constrain subsequent packet inversions. The final (combined) PPD is considered the full solution to the reflectivity inverse problem. Modelled spherical wave reflection coefficient data used in the frequency-domain inversion are computed using a novel approach based on processing modelled full wave acoustic fields in the same manner as the measured data. This approach is relatively efficient and more general than plane wave decomposition methods for computing spherical-wave reflection coefficients (Brekhovskikh, 1980), allowing arbitrary water-column sound velocity profiles and sediment interface roughness. Finally, the inversion algorithms are implemented for massively parallel computers for further efficiency (Dettmer et al., 2006). Since computational power is poised to increase significantly, even on desktop computers, through multiple computing cores, the development of parallel inversion codes is timely and important for geoacoustic inversion in general.

The time- and frequency-domain inversions complement each other since the time-domain data provide information about the sediment sound-velocity profile through travel times while the frequency-domain data provide spectral amplitude informa-

tion on all parameters in terms of reflection coefficients. The sequential Bayesian approach includes the information from both data sets without resorting to more computationally-intensive simultaneous inversion. Further, the joint time/frequency domain approach utilises travel time and amplitude information without the very high computational costs associated with full-waveform inversion.

Bayesian time/frequency domain inversion yields much higher spatial resolution and smaller uncertainties than achieved with matched-field and other long-range methods (Holland, Dettmer, and Dosso, 2005; Dosso and Holland, 2006; Dettmer, Dosso, and Holland, 2006). These small uncertainties coupled with the small experimental footprint and relative simplicity of the measurements provide a powerful new tool to examine spatial variability on finer scales than heretofore possible.

An alternative to the single-bounce reflectivity measurements is to use ambient noise (generated by uniformly distributed breaking waves at the ocean surface) to measure the plane wave reflection coefficient locally. Harrison and Simons (2002) and Harrison and Baldacci (2002) showed that ambient noise measurements can be used to determine the plane wave reflection coefficient by beamforming vertical array data to separate the upward field from the downward field. The plane wave reflection coefficient as a function of angle and frequency follows from the ratio of the upward and downward fields, with some degradation due to angular smearing by the beamforming process. The simplicity of the method is striking: No ship or controlled source are needed. The reflection coefficient is measured within the area of one ray cycle around the array. Reflectivity inversion from ambient noise measurements is briefly considered in this thesis.

1.2. Overview of Work in this Thesis

The remainder of this thesis is organised as follows. Chapter 2 considers the theory of the reflection of plane and spherical waves in layered fluid media. These physical concepts are the basis of the forward models that are used to compute replica data in the inversion algorithms. Chapter 3 then develops the basic ideas of inverse theory including linear and linearised inverse problems and the maximum likelihood method.

Chapter 4 derives the Bayesian formulation of the inverse problem and develops a fully nonlinear approach to inversion. This includes optimisation of the PPD for optimal parameter estimates using hybrid inversion and integration of the PPD via

Metropolis Gibbs sampling to provide parameter uncertainty estimates. Finally, computationally efficient implementations of inversion algorithms, including massively parallel programming, are outlined.

The importance of formulating an appropriate likelihood function is of key importance to nonlinear Bayesian inversion. Chapter 5 derives data misfit functions from the likelihood and develops methods for treating data errors (including correlated and non-stationary errors). The importance of validating the statistical assumptions of the inversion is discussed, including rigorous parameter uncertainty estimates.

Chapter 6 develops a *joint time/frequency domain inversion* for single bounce reflectivity data. In the time domain, a detailed synthetic study is carried out to examine and account for the errors on picked travel times including non-stationary correlated and biased errors. The forward model consists of two point ray tracing. In the frequency domain, two different forward models are developed, one based on the plane wave assumption and the other on modelling the full acoustic pressure field (spherical wave reflection coefficients). The applicability of the models for reflectivity inversion is compared through a simulated experiment.

The following chapters apply the inverse methods developed here to experimental data. Chapter 7 applies the joint time/frequency inversion to data collected in the Strait of Sicily. This inversion aims at recovering the layered sediment velocity and density profiles for the upper several metres of the seabed, and the results are compared to a piston core taken at the experiment site. The time domain inversion is carried out using a ray tracing forward model and the frequency domain inversion is carried out using a full wave-field forward model. Chapter 8 applies the Bayesian inversion using a plane wave forward model to data collected at a different site in the Strait of Sicily. The importance of using appropriate statistical assumptions is examined in an inversion with simulated data with strongly correlated data errors. From the measured data, high resolution density and sound velocity gradients of the uppermost sediment layer (the transition layer) are recovered, and the results compared to cores taken at the experiment site.

The ability to recover geoacoustic information from ambient noise reflection coefficient data are examined in Chapter 9. First the technique to recover reflection coefficients from vertical line array data by beamforming is described. This is followed by a simulation study, examining the resolving power of ambient noise data.

Chapter 10 gives a brief summary and discussion of the work in this thesis.

2. Reflection of Waves in Layered Fluids

The basis of the work in this thesis is waves reflected off interfaces in the seabed sediments. The reflected energy is quantified in terms of reflection coefficients and the resulting data are inverted for seabed geoacoustic parameters. Depending on the geometry of the experiment, plane wave effects can be sufficient to describe the governing physical processes. However, in certain cases full wave effects become increasingly important. This chapter summarises the theory of acoustic waves in layered fluids. For a more detailed account, see Brekhovskikh (1980), Brekhovskikh and Lysanov (1991), Brekhovskikh and Godin (1990), and Brekhovskikh and Godin (1992). First, expressions for plane wave reflection coefficients will be derived for the case that the source can be assumed to be at infinite distance from the sediment interfaces and from the point of observation. Then the theory of reflection of spherical waves in layered media will be discussed using a plane wave decomposition of a spherical wave.

2.1. The Wave Equation

The wave equation for pressure p in a homogeneous fluid is given by

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = 0 \,, \tag{2.1}$$

where c is the sound velocity in the fluid. The wave equation can be derived from the equation of continuity and the equation of state. It describes the propagation of a perturbation with velocity c over time t, where the form of the perturbation does not change with time. A solution for a harmonic wave of angular frequency ω for Eq. 2.1

is given by

$$p(t) = p(\omega)\exp(-i\omega t)\,. \tag{2.2}$$

Superposition allows an arbitrary signal to be represented by harmonic waves and thus simplifies the analysis of complicated signals to the analysis of harmonic waves. The Fourier transformation (FT) pair for such a signal is

$$\begin{aligned} p(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} p(\omega)\exp(-i\omega t)d\omega \\ p(\omega) &= \int_{-\infty}^{\infty} p(t)\exp(i\omega t)dt\,. \end{aligned} \tag{2.3}$$

Substituting Eq. 2.3 into the wave equation 2.1 yields the Helmholtz equation

$$\nabla^2 p(\omega) + \frac{\omega^2}{c^2}p(\omega) = 0\,. \tag{2.4}$$

We can now consider a solution for the Helmholtz equation

$$p(\omega) = A(\omega)\exp(ik(n_x x + n_y y + n_z z)), \tag{2.5}$$

where k is the wavenumber, $\mathbf{k} = k\mathbf{n}$ is the wavevector, and $A(\omega)$ the amplitude. Substituting this solution into Eq. 2.4 yields

$$n_x^2 + n_y^2 + n_z^2 = 1\,. \tag{2.6}$$

The solution

$$\begin{aligned} p(t) &= p(\omega)\exp(-i\omega t) \\ &= A(\omega)\exp(i(\mathbf{k}\cdot\mathbf{r} - \omega t)) \end{aligned} \tag{2.7}$$

is a plane wave propagating with phase speed $c = \omega/k$, and the planes of constant phase $\mathbf{k}\cdot\mathbf{r} = \text{const.}$ are wavefronts.

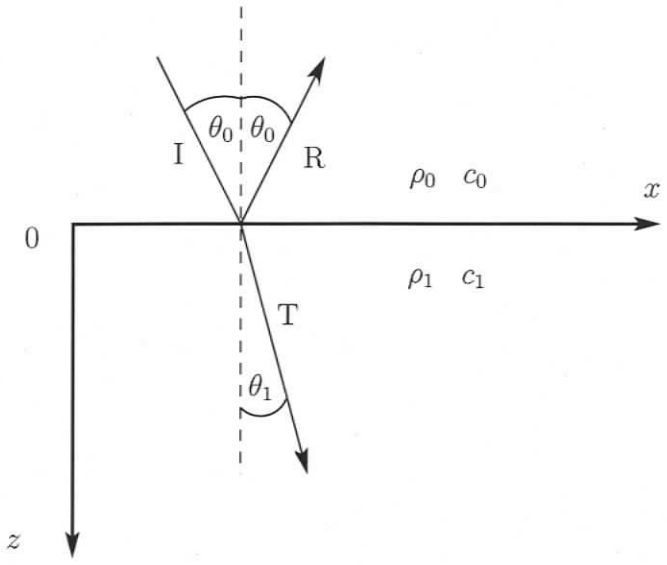

Fig. 2.1: Reflection of a plane wave from an interface.

2.2. Reflection of Plane Waves from Interfaces

Let the acoustic pressure $p(x, y, z, t)$ describe the sound field in a fluid. The fluid has an interface at $z = 0$ that separates two fluid half-spaces with different physical properties. The densities and sound velocities for the two half-spaces are denoted by ρ_0, c_0 and ρ_1, c_1. Let a plane wave with amplitude $A = 1$ be incident on the x-y-plane boundary at $z = 0$ (Fig. 2.1) with an angle of incidence θ_0. The wave vector for the plane wave is given by

$$k_x = k \sin\theta_0 \qquad k_y = 0 \qquad k_z = k \cos\theta_0 \,. \tag{2.8}$$

Let the reflection coefficient between the two media be denoted by V; the incident wave I and the reflected wave R can then be written

$$\begin{aligned} p_{\mathrm{I}} &= \exp(ik(x \sin\theta_0 - z \cos\theta_0)) \exp(-i\omega t) \\ p_{\mathrm{R}} &= \mathsf{V} \exp(ik(x \sin\theta_0 + z \cos\theta_0)) \exp(-i\omega t) \,. \end{aligned} \tag{2.9}$$

The transmitted wave T is given by

$$p_{\mathrm{T}} = \mathsf{W} \exp(ik_1(x \sin\theta_1 + z \cos\theta_1)) \tag{2.10}$$

with $k_1 = \omega/c_1$. θ_1 is the refracted angle and W is the transmission coefficient. The total field in the upper halfspace is given by the superposition of fields p_I and p_R. The reflective quality of the interface is determined from the boundary conditions at the interface.

1. Continuity of pressure p:

$$p(z = +0) = p(z = -0) \tag{2.11}$$

2. Continuity of normal particle velocity v_z:

$$v_z(z = +0) = v_z(z = -0) \tag{2.12}$$

It is common and useful to introduce the impedance Z, a ratio of field variables defined as

$$Z = -\frac{p}{v_z}, \tag{2.13}$$

and the velocity and density contrasts denoted by

$$n = \frac{c_0}{c_1} \qquad \text{and} \qquad m = \frac{\rho_1}{\rho_0}. \tag{2.14}$$

From Eqns. 2.11 and 2.12 it follows that Z also satisfies the boundary condition $Z(z = +0) = Z(z = -0)$. Applying the boundary condition for $z = 0$ to Eq. 2.9 and Eq. 2.10 yields

$$1 + \mathsf{V} = \mathsf{W} \exp(i(k_1 \sin\theta_1 - k \sin\theta_0)x)\,. \tag{2.15}$$

Since the left side of Eq. 2.15 is a constant, the term $\exp(i(k_1 \sin\theta_1 - k \sin\theta_0)x)$ must also be constant and hence

$$k_1 \sin\theta_1 \stackrel{!}{=} k \sin\theta_0\,. \tag{2.16}$$

This relation is known as Snell's law. Snell's law can also be written as $n = \sin\theta_1 / \sin\theta_0$ with $n = k_1/k_0 = c_0/c_1$ being the refractive index. Further, it follows that

$$1 + \mathsf{V} = \mathsf{W}\,. \tag{2.17}$$

For a harmonic wave in a homogeneous medium, the particle displacement can be written

$$\mathbf{v} = \nabla p/(i\omega\rho)\,. \tag{2.18}$$

Substituting this and Eq. 2.10 into Eq. 2.13 gives the impedance for the lower medium which is independent of z

$$Z_1 = \rho_1 c_1/\cos\theta_1\,. \tag{2.19}$$

Doing the same for the upper medium yields an impedance that is dependent on z. Equations (2.11) and (2.12) show that impedance is continuous at horizontal boundaries. Evaluating the boundary condition for the impedance at $z = 0$ yields

$$\begin{aligned}(1+\mathsf{V})(1-\mathsf{V}) &= (\cos\theta_0/(\rho_0 c_0))Z_1 \\ \mathsf{V} &= \frac{Z_1\cos\theta - \rho_0 c_0}{Z_1\cos\theta + \rho_0 c_0}\,.\end{aligned} \tag{2.20}$$

With $Z_0 = Z(z = 0)$ being the impedance for the upper medium, we can rewrite Eq. 2.20 to obtain the well known expression

$$\mathsf{V} = \frac{Z_1 - Z_0}{Z_1 + Z_0} \qquad \text{with} \qquad Z_0 = \frac{\rho_0 c_0}{\cos\theta_0} \tag{2.21}$$

for the reflection coefficient that is determined by the impedance contrast over the interface. Equation 2.21 gives the reflection coefficient for an arbitrary layered lower space as long as Z_1 is known. In the case of a homogeneous half-space, Z_1 is calculated in a straightforward manner by Eq. (2.19). For a layered system, an iterative approach (derived later) is needed to find the combined impedance.

There are a few special cases for the reflection coefficient that are worth discussing. For normal incidence, i.e., $\theta_0 = 0$, V becomes

$$\mathsf{V} = \frac{\rho_1 c_1 - \rho_0 c_0}{\rho_1 c_1 + \rho_0 c_0}\,. \tag{2.22}$$

At grazing incidence, i.e., $\theta_0 = \pi/2$, $\mathsf{V} = -1$. The products $\rho_0 c_0$ and $\rho_1 c_1$ are called the characteristic impedances. V is independent of θ for identical sound velocities in both layers and is given by the density contrast $(\rho_1 - \rho_0)/(\rho_1 + \rho_0)$.

Besides these trivial cases there are two more important cases that will be of special interest in this work. The *angle of intromission* is defined as the angle with perfect

transmission into the lower medium and no reflected energy off the interface. Setting $\mathsf{V} = 0$ gives the angle of intromission $\hat{\theta}$ as

$$\sin\hat{\theta} = \left(\frac{1 - \left(\frac{\rho_0 c_0}{\rho_1 c_1}\right)^2}{1 - \left(\frac{\rho_0}{\rho_1}\right)^2} \right)^{1/2} . \tag{2.23}$$

From this relation it follows that the argument of the square root has to be $0 \leq \arg \leq 1$. This means that

$$\begin{aligned} \rho_1/\rho_0 > 1 \quad &\Rightarrow \quad 1 < c_0/c_1 \leq \rho_0/\rho_1 \\ \rho_1/\rho_0 < 1 \quad &\Rightarrow \quad 1 > c_0/c_1 \geq \rho_0/\rho_1 \, . \end{aligned} \tag{2.24}$$

In underwater acoustics, the angle of intromission appears when the lower medium has lower sound velocity and higher density than the upper medium. Perfect intromission only exists in the case of ideal fluids without attenuation, however, the effect can be observed in fine grained sediments where highly water-saturated clays and silty-clays compose the uppermost sediment layers of the seabed. This type of sediment is referred to as a soft bottom.

Holland (2002) gives a simple method to extract sediment properties by measuring the angle of intromission and the normal incidence plane wave reflection coefficient. The angle of intromission is given by

$$\sin\delta = (1 - (\rho_w c_w/\rho_2 c_2)^2)^{1/2} (1 - (\rho_w/\rho_2)^2)^{-1/2} \, , \tag{2.25}$$

where δ is the angle of intromission and w indicates water properties. The sediment properties ρ_2 and c_2 can then be extracted as

$$\begin{aligned} \rho_2 &= \rho_w (1 - 4\bar{\mathsf{V}}/(\cos\delta(1+\bar{\mathsf{V}}))^2)^{-1/2} \\ c_2 &= \rho_w c_w/\rho_2 (1+\bar{\mathsf{V}})/(1-\bar{\mathsf{V}}) \, , \end{aligned} \tag{2.26}$$

where $\bar{\mathsf{V}}$ is the normal incidence plane wave reflection coefficient.

The *critical angle* is the angle of total reflection. For $c_1 > c_0$, $|\mathsf{V}| = 1$ for all incidence angles satisfying

$$\theta \geq \theta_c = \arcsin c_0/c_1 \, . \tag{2.27}$$

The change of phase upon reflection depends on the incident angle. At total reflection, the lower medium contains an inhomogeneous wave that decays exponentially with depth. The impedance of the lower medium becomes imaginary.

The plane wave reflection coefficient for an arbitrary number of homogeneous, horizontal layers is obtained as follows. Starting at the lower-most interface, the reflection coefficient for the sea-bed is recursively calculated by coherently adding up the reflection coefficients for each interface (Jensen et al., 1993). The reflection coefficient for one fluid layer (index 2) between two half spaces (indices 1 and 3) is given by

$$\mathsf{V} = \frac{\mathsf{V}_{12} + \mathsf{V}_{23}\exp(2i\phi_2)}{1 + \mathsf{V}_{12}\mathsf{V}_{23}\exp(2i\phi_2)}, \tag{2.28}$$

where V_{12} denotes the reflection coefficient for the interface in between layer 1 (the upper half space) and layer 2. The vertical phase delay is given by

$$\phi_2 = k_2 h_2 \sin\theta_2, \tag{2.29}$$

where k_2 is the wavenumber in the second layer, h_2 is the layer thickness, and θ_2 is the grazing angle in that layer. For a system of l fluid layers with 1 and l being half spaces, the reflection coefficient is found by starting with the three lower-most layers. Applying Eq. 2.28 to those layers and recursively going upwards yields:

$$\begin{aligned}
\mathsf{V}_{(l-2)l} &= \frac{\mathsf{V}_{(l-2)(l-1)} + \mathsf{V}_{(l-1)l}\exp(2i\phi_{l-1})}{1 + \mathsf{V}_{(l-2)(l-1)}\mathsf{V}_{(l-1)l}\exp(2i\phi_{l-1})} \\
\mathsf{V}_{(l-3)l} &= \frac{\mathsf{V}_{(l-3)(l-2)} + \mathsf{V}_{(l-2)l}\exp(2i\phi_{l-2})}{1 + \mathsf{V}_{(l-3)(l-2)}\mathsf{V}_{(l-2)l}\exp(2i\phi_{l-2})} \\
&\vdots \\
\mathsf{V}_{1l} &= \frac{\mathsf{V}_{12} + \mathsf{V}_{2l}\exp(2i\phi_2)}{1 + \mathsf{V}_{12}\mathsf{V}_{2l}\exp(2i\phi_2)}.
\end{aligned} \tag{2.30}$$

Absorption can be incorporated by introducing complex wave numbers. This results in complex refracted angles, impedances and reflection coefficients.

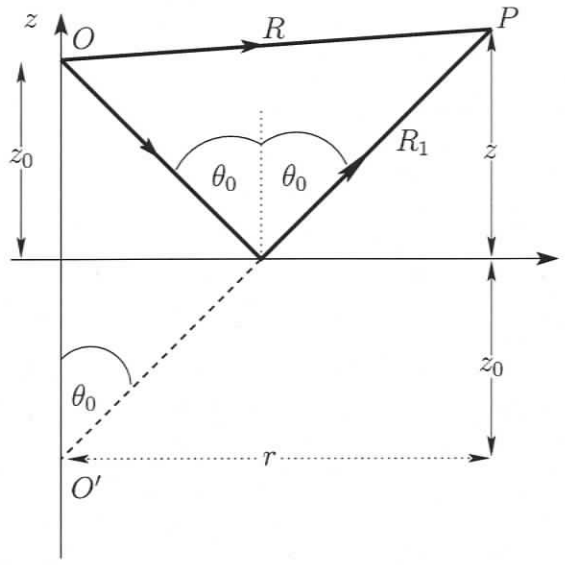

Fig. 2.2: Position of Source O, image source O' and point of observation P with respect to the interface for spherical wave reflection (after Brekhovskikh, 1980).

2.3. Reflection of Spherical Waves from Interfaces

For the case of finite distance between source and interfaces or receiver (i.e., for distances less than many wavelengths), spherical wave effects have to be taken into account (Stickler, 1977; Akal and Berkson, 1987). The classic work on reflection of spherical waves is by Sommerfeld (1949). In this section we mainly follow the derivations in Brekhovskikh (1980) and Brekhovskikh and Godin (1992). For the type of experiment geometry considered in this thesis, the difference of the plane interface and the curved wavefront are important in some cases and spherical wave effects must be used in calculating reflected energy.

Let the source for the acoustic spherical wave be a small pulsating sphere of radius r_0 at point O (Fig. 2.2). The acoustic pressure of a harmonic spherical wave at distance R from the source is given by

$$\Psi = \frac{r_0^2 v_0}{R} \exp\left(i(kR - \omega t)\right) , \tag{2.31}$$

where v_0 is the velocity of the surface of the source sphere. Pressure Ψ can be written as $R^{-1}\exp(ikR)$ by omitting the time dependence factor $\exp(-i\omega t)$ and the power of the source. The spherical wave can then be decomposed into plane waves. Using

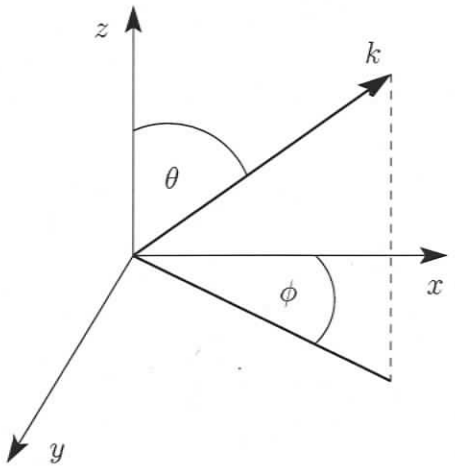

Fig. 2.3: Position of integration angles θ and ϕ for spherical wave reflection.

polar coordinates and a Fourier integral yields on the plane $z = 0$

$$\frac{\exp(ikR)}{R} = \frac{i}{2\pi}\int\int_{-\infty}^{\infty} \frac{\exp\left(i(k_x x + k_y y)\right)}{(k^2 - k_x^2 - k_y^2)^{1/2}} dk_x dk_y \,. \tag{2.32}$$

Each component of Eq. 2.32 represents a plane wave. This expression can be continued away from $z = 0$ since the integrand satisfies the wave equation (Eq. 2.1) and yields correct values for the acoustic field for $z = 0$. Adding $\pm ik_z z$ to the exponent of the integrand and substituting $k_z = (k^2 - k_x^2 - k_y^2)^{1/2}$ yields waves travelling in direction of positive z (plus sign) and in direction of negative z (minus sign):

$$\frac{\exp(ikR)}{R} = \frac{i}{2\pi}\int\int_{-\infty}^{\infty} \frac{\exp\left(i(k_x x + k_y y \pm k_z z)\right)}{k_z} dk_x dk_y \,. \tag{2.33}$$

This equation represents the decomposition of a spherical wave into plane waves with the direction of propagation given by the wave vector $\mathbf{k}$. The integral can also be written in terms of angles θ and ϕ (see Fig. 2.3) with

$$k_s = k\sin\theta\cos\phi\,, \qquad k_y = k\sin\theta\sin\phi\,, \qquad k_z = k\cos\theta\,. \tag{2.34}$$

Integration over θ cannot be restricted to the real axis, since $k_z \to i\infty$ for $k_x \to \pm\infty$

or $k_y \to \pm\infty$. Equation 2.33 then becomes

$$\frac{\exp(ikR)}{R} = \frac{ik}{2\pi} \int_0^{\pi/2-i\infty} \int_0^{2\pi} \exp(i(k_x x + k_y y \pm k_z z)) \sin\theta d\theta d\phi \,. \tag{2.35}$$

This expansion of a spherical wave contains body waves as well as inhomogeneous waves with complex angles θ. Inhomogeneous waves propagate with reduced wavelength and attenuate exponentially with depth.

The total acoustic field of a spherical wave that was reflected from an interface can be written as (omitting the lateral wave)

$$\Psi = \frac{\exp(ikR)}{R} + \Psi_{refl} \,, \tag{2.36}$$

where Ψ_{refl} is the reflected wave. Ψ_{refl} can be written as a superposition of reflected plane waves

$$\Psi_{refl} = \frac{ik}{2\pi} \int_0^{\pi/2-i\infty} \int_0^{2\pi} \exp(ik(x\sin\theta\cos\phi + y\sin\theta\sin\phi + (z+z_0)\cos\theta))\mathsf{V}(\theta)\sin\theta d\theta d\phi \,, \tag{2.37}$$

where $z + z_0$ is the source-receiver distance from the interface and $\mathsf{V}(\theta)$ is the plane wave reflection coefficient.

It can be shown that the integral over ϕ in Eq. 2.37 reduces to a Bessel function of zeroth order. By rearranging the integral limits in Eq. 2.37, the Bessel function can be expressed in terms of Hankel functions:

$$\Psi_{refl} = \frac{ik}{2} \int_{-\pi/2+i\infty}^{\pi/2-i\infty} \mathrm{H}_0^{(1)}(u) \mathrm{e}^{ik(z+z_0)\cos\theta} \mathsf{V}(\theta) \sin\theta d\theta \,, \tag{2.38}$$

where $u = kr\sin\theta$. Equation 2.38 is commonly referred to as the Sommerfeld integral (Sommerfeld, 1949) and can be used to calculate the up-going spherical wave and thus the up-going energy (including reflected and lateral waves) for an arbitrary set of layers, given the plane wave reflection coefficient $\mathsf{V}(\theta)$. The spherical reflection coefficient V_s can then be defined (Harrison and Nielsen, 2004a) as to include all

effects but spherical spreading

$$\begin{aligned}\Psi_{refl} &= \mathsf{V_s}\frac{\exp(ikR_1)}{R_1} \\ &= \frac{ik}{2}\int_{-\pi/2+i\infty}^{\pi/2-i\infty} \mathrm{H}_0^{(1)}(u)\exp(ik(z+z_0)\cos\theta)\mathsf{V}(\theta)\sin\theta d\theta\,.\end{aligned} \tag{2.39}$$

The integral can also be written in terms of grazing angles. It is interesting to note that V and $\mathsf{V_s}$ can be related through a Hankel transform pair (Harrison and Nielsen, 2004a,b). The integration is non-trivial since the integrand is highly oscillatory.

Under certain conditions, the integral in Eq. 2.39 can be solved with the method of steepest descent (Brekhovskikh, 1980). For distances large compared to the wavelength, the Hankel function $\mathrm{H}_0^{(1)}$ can be replaced by its asymptotic approximation

$$\mathrm{H}_0^{(1)}(u) \approx \left(\frac{2}{\pi u}\right)^{1/2} \mathrm{e}^{i(u-\pi/4)}(1+1/8iu+\cdots) \tag{2.40}$$

which can then be substituted into Eq. 2.38 to yield

$$\Psi_{refl} = \frac{k}{2\pi r}^{1/2} \mathrm{e}^{i\pi/4}\int_{-\pi/2+i\infty}^{\pi/2-i\infty} \mathrm{e}^{ikR_1\cos(\theta-\theta_0)}(1+1/8ikr\sin\theta)\mathsf{V}(\theta)(\sin\theta)^{1/2}d\theta\,. \tag{2.41}$$

Applying the method of steepest descent to Eq. 2.41 yields the final expression for the reflected wave

$$\Psi_{refl} = (\mathrm{e}^{ikR_1}/R_1)\left(\mathsf{V}(\theta_0) - \frac{iN}{kR_1}\right)\,, \tag{2.42}$$

where N is an expression that contains the first and second derivatives with respect to θ of the plane wave reflection coefficient V at angles θ_0

$$N = \frac{1}{2}\left(\mathsf{V}''(\theta_0) + \mathsf{V}'(\theta_0)\cot\theta_0\right)\,. \tag{2.43}$$

Since Eq. 2.42 contains the reflected wave and the incident wave, a measure for the total reflected energy $\mathsf{V_s}$ can be written as

$$\mathsf{V_s} = \Psi_{refl}/\left(\mathrm{e}^{ikR_1}/R_1\right) = \left(\mathsf{V}(\theta_0) - \frac{iN}{kR_1}\right)\,. \tag{2.44}$$

$\mathsf{V_s}$ can thus be calculated for an arbitrary layered sea-bed. However, this method is restricted to iso-velocity water-columns and contains a far field Hankel approximation. Further, this approximation does not include the lateral or head wave that becomes important past the critical angle (or below the critical angle in terms of grazing angles). See Brekhovskikh and Godin (1992) for a detailed discussion of lateral waves.

An alternative method to yield $\mathsf{V_s}$ (Eq. 2.39) is based on calculating the total acoustic field. The underwater acoustics community has developed several advanced models to compute the full acoustic field in the water-column with high precision and efficiency. These models can be used to calculate reflectivity, including spherical wave effects and without far field approximations or iso-velocity water-column assumptions. The numerical models can also handle shear effects and other complicated phenomena such as interface roughness, which can then also be included in the reflectivity computation to yield a very general forward model. Section 6.4.1 develops a spherical wave forward model to compute the reflected energy for horizontally stratified but otherwise arbitrary environmental models (water-column and seabed). The method uses a wavenumber integration model to calculate full acoustic fields, without Hankel transform approximations, and processes the field to yield reflected energy from spherical waves as a function of frequency and angle.

As mentioned earlier, spherical wave effects become important when the distances between source, receiver and interfaces are less than many wavelengths. This means that at infinite frequency or at infinite distance from the source, spherical wave effects vanish and V completely describes the reflection. For lower frequencies or finite distance, the reflection is not from a single point but rather from a finite volume (a Fresnel volume) around the specular point. In essence, a hydrophone registers energy from a range of angles for a reflector. Figure 2.4 shows a simulation of plane wave and spherical wave reflections for the case of a single interface between two halfspaces. The simulation was carried out for two different environmental models that differ only in the sound velocity of the lower halfspace. The source-receiver height was 178 m and the model was given by $c_1 = 1511$ m/s, $\rho_1 = 1.029$ g/cm^3, $\alpha_1 = 0$ dB/λ, $c_2 = 1700$ m/s, $\rho_2 = 1.4$ g/cm^3 and $\alpha_2 = 0.06$ dB/λ. For the second model, c_2 was set to 1670 m/s. It can be seen that $\mathsf{V_s}$ is consistently lower than V around the critical angle of the plane wave V. As Stickler (1977) has shown, the amount of displacement of the critical angle depends on the source receiver height above the interface as well as on the frequency. With increasing frequency or source receiver

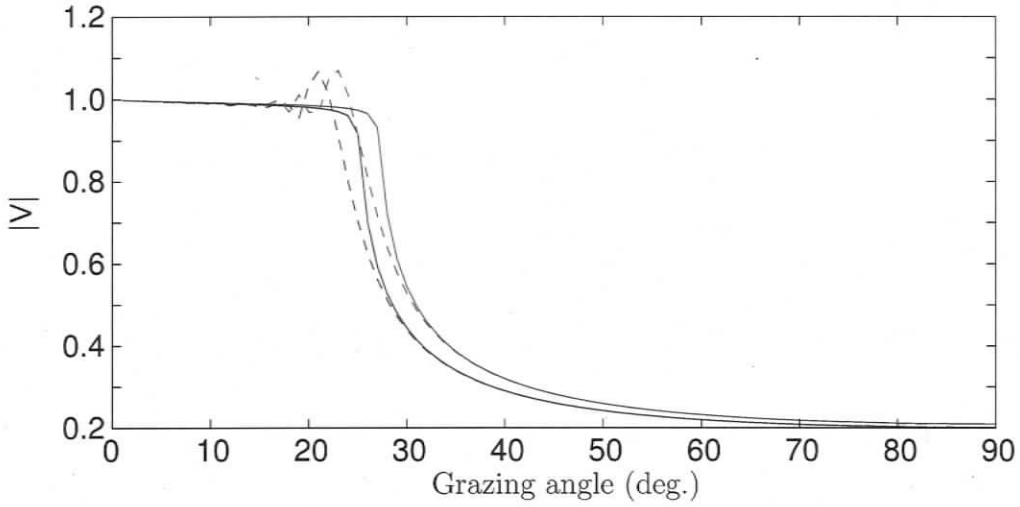

Fig. 2.4: Comparison of plane wave and spherical wave reflection for a single interface between two halfspaces at a frequency of 500 Hz. Dashed lines represent the spherical wave reflection, solid lines the plane wave reflection. Blue and red represent two different sets of parameters that only differ in the lower halfspace sound velocity. The lower sound velocity for the blue lines is 1670 m/s, for the red lines 1700 m/s.

height, the displacement of the critical angle decreases. Considering that the Fresnel zone is not symmetric around the specular point, this results in an effective critical angle that is shifted towards lower angles. The shift towards lower angles will result in a negative bias of sound velocities in inversion under the plane wave assumption. However, for a rigorous examination of the phenomenon purely intuitive effects are not sufficient and contributions from imaginary angles or inhomogeneous waves must be considered as well. The phenomenon of reflection coefficient values greater than one, as can be seen in Fig. 2.4, results from caustic effects (Stickler, 1977).

For a layered medium the effect of a point source can be understood from ray theory. In the plane wave case, at infinite distance from the source, every sediment interface contributes to the reflection coefficient at the same angle. On the other hand, at finite distance from the source, as can be seen in Fig. 2.5, the energy for each sediment interface originates at a different angle. This effect becomes more significant with larger depth. This means that spherical wave effects are more important for deeper layers in an environmental model. A plane wave solution is expected to work well for very shallow environmental models but will break down once increased depth is of interest.

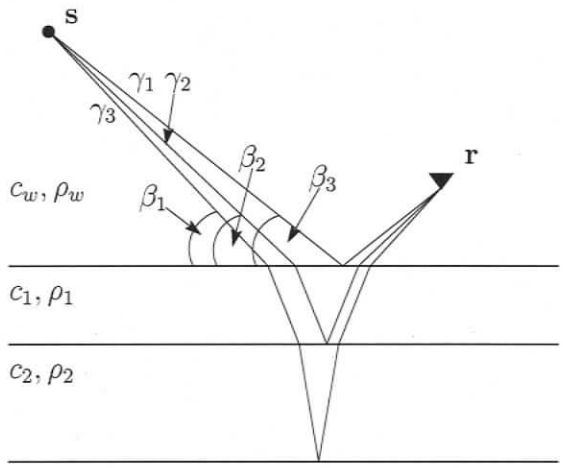

Fig. 2.5: Reflection of a spherical wave (represented by rays γ_1, γ_2, and γ_3) from multiple interfaces; the point **r** registers energy contributions from different angles β_i. The effect gets stronger with increasing depth of the interface.

3. Classical Inverse Theory

This chapter summarises the fundamentals of classical inverse theory, including linear and linearised inverse problems. Many of the ideas in this Chapter are useful for the nonlinear Bayesian inversion derived in Chapter 4. More complete treatments can be found in, e.g., Menke (1984), Tarantola (1987), Parker (1994) and Aster et al. (2005).

3.1. Fundamentals

Inverse theory allows for estimation of parameters of an assumed model for a physical system from observations (data) of a process (e.g., sound) that interacts with the system. The model is often a simplification of the physical system being studied and it must be considered whether or not the parameterisation is adequate (contributes to theory error). The goal is to find a parametrisation with a minimal number of model parameters that characterises the physical system with the desired precision.

The model can be represented by M model parameters in the model vector $\mathbf{m}$ that is a point in the M-dimensional model space $\mathcal{M} \subset \mathbb{R}^M$. The N data are represented by a vector $\mathbf{d} \in \mathbb{R}^N$. In some cases (e.g., matched field inversion) the data can also be complex. Model and data are related through the forward and inverse problems. The forward model predicts a data set $\mathbf{d}$ that would be observed in an experiment given the model $\mathbf{m}$

$$\mathbf{d} = F[\mathbf{m}]\,. \tag{3.1}$$

Formulating the forward model means developing the formalism that describes the physical laws in the studied system. Since forward problems are physical and exist in nature, they are unique, stable, and a solution exists. The same is generally true for the forward model.

The inverse model determines a set of model parameters $\mathbf{m}$ given a data set $\mathbf{d}$

$$\mathbf{m} = F^{-1}[\mathbf{d}]\,. \tag{3.2}$$

Inverse problems are mathematical problems, and are not guaranteed to have solutions. Hence, the inverse problem can only be solved once the forward problem is formulated. Also, the problem can be unstable, and is generally non-unique. The last two points cause significant difficulties in inverse theory.

Existence of a solution to an inverse problem is demonstrated if a model can be formulated that replicates data, i.e., fits the observations to within their uncertainties (assumes some knowledge about the errors of the observed data). However, non-existence is not proved if a forward model cannot be constructed. Stability of the forward problem means that small perturbations in the model cause only small perturbations in the data. However, for some problems, a small perturbation in the data can cause a large perturbation in the model.

Non-uniqueness is caused by inaccurate data (data errors) and incomplete data (finite number of observations). Every finite data set is inherently incomplete since it cannot capture all aspects of the physical system. Further, the physics of the forward model can result in non-uniqueness (e.g., gravity measurements). Data errors can be presented as a vector $\mathbf{n} \in \mathbb{R}^N$

$$\mathbf{d} = \mathbf{d}^{(t)} + \mathbf{n}\,, \tag{3.3}$$

where $\mathbf{d}^{(t)}$ is the *true* data (inherently unknown). The data errors $\mathbf{n}$ include *measurement* and *theory errors*. Measurement errors are associated with the experiment. They arise from instrument noise and other noise sources. It is sometimes possible to estimate measurement errors from knowledge about the experiment.

Theory error is due to simple parametrisation and approximations in the physics of the forward model. It is generally difficult to estimate theory error independently. Depending on the problem, one or the other error can dominate. In geoacoustic inversion, it is common for the theory error to dominate.

By assuming the statistics of $\mathbf{n}$ to be approximately known, information about the underlying model $\mathbf{m}$ can be recovered. It is common to assume that $\mathbf{n}$ is distributed according to a certain distribution, e.g., Gaussian, with standard deviation that is estimated from the data.

Assuming Gaussian distributed errors is supported by the *multivariate central limit theorem* (e.g., Rencher, 1998) which states that any sum of many independent distributed random variables is approximately Gaussian distributed. The assumption of Gaussian errors leads to the powerful tool of least-squares or L_2-norm inversion. The

Gaussian distribution is given by

$$P(\mathbf{d}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{d}-\langle\mathbf{d}\rangle)^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}(\mathbf{d}-\langle\mathbf{d}\rangle)\right), \tag{3.4}$$

where

$$\mathbf{C} = \langle(\mathbf{d}-\langle\mathbf{d}\rangle)(\mathbf{d}-\langle\mathbf{d}\rangle)^{\mathsf{T}}\rangle \tag{3.5}$$

is the data covariance matrix and $\langle\cdot\rangle$ denotes the expected value. For unbiased $\mathbf{d}$, it follows that $\mathbf{C} = \langle\mathbf{n}\mathbf{n}^{\mathsf{T}}\rangle$. In certain cases, it can be useful to assume other statistics for $\mathbf{n}$. For example, assuming a double sided exponential distribution results in L_1-norm inversion.

Inverse problems can be divided into two major categories: Linear and nonlinear problems. Linear problems can be represented by a set of linear equations and will only be discussed briefly. Most problems in underwater acoustics are nonlinear and the focus in this thesis will be on fully nonlinear numerical methods.

3.2. Linear Inverse Problems

Let the function $m(z)$ represent the model and d_k, $k = 1 \ldots N$, a discrete finite data set. The linear forward problem is then given by

$$d_k = \int_a^b g_k(z)m(z)dz\,, \tag{3.6}$$

where the $g_k(z)$ are the kernel functions. Then, the data d_k are the projections of the model onto the kernel functions.

Since there are only N kernel functions, the basis of the model space is $\leq N$. However, a finite basis cannot uniquely represent the model function of infinite dimension. Therefore, the model space $\mathcal{M}$ can be divided into two subspaces $\mathcal{M} = \mathcal{M}^{\|} \cup \mathcal{M}^{\perp}$. $\mathcal{M}^{\|}$ is spanned by the kernel functions and is $\leq N$ dimensional, depending on the amount of linear independent information in the kernel functions. The so called *null space* $\mathcal{M}^{\perp}$ has infinite dimension. Any model $m \in \mathcal{M}$ can be divided into the parts

$m^{\parallel} \in \mathcal{M}^{\parallel}$ and the annihilator part $m^{\perp} \in \mathcal{M}^{\perp}$ with

$$\begin{aligned} d_k &= \int_a^b g_k(z) m^{\parallel}(z) dz \\ &= \int_a^b g_k(z) \left(m^{\parallel}(z) + m^{\perp}(z) \right) dz \,. \end{aligned} \tag{3.7}$$

Discretisation of the model $m(z)$ then leads to a set of linear equations

$$\mathbf{d} = \mathbf{A}\mathbf{m} \,, \tag{3.8}$$

where the $N \times M$ matrix $\mathbf{A}$ is the kernel or sensitivity matrix. Depending on the rank and size of $\mathbf{A}$, we can distinguish between different situations. A system that is even-determined has an exact and unique solution. $\mathbf{A}$ is non-singular ($\text{rank}(\mathbf{A}) = M = N$), it has N linear independent rows and columns. A system with more linear independent rows than columns is called over-determined ($\text{rank}(\mathbf{A}) = M < N$, $\mathbf{A}$ is singular). An over-determined problem can not be solved exactly since there is no unique solution that fits the data exactly, resulting in the best fit model. The problem is called under-determined ($\text{rank}(\mathbf{A}) = N < M$, $\mathbf{A}$ is singular) when $\mathbf{A}$ does not contain enough information to determine all parameters individually. In some cases, a mixed-determined problem ($\text{rank}(\mathbf{A}) < N \leq M$, $\mathbf{A}$ is singular) has some parameters that are under-determined and some that are over-determined (e.g., seismic tomography). Further, problems can be ill-conditioned. This means that the data cannot resolve all individual parameters because the data are close to being linearly dependent (i.e., small singular values, see Sec. 3.2.3).

3.2.1. The Maximum-Likelihood Method

An experiment can be interpreted as measuring one realisation of data $\mathbf{d}$ from an infinite number of possible data sets that differ by the noise realisation. The problem can then be reversed and the question be asked what the probability of $\mathbf{d}$ being measured, given a particular model $\mathbf{m}$. If the model is good, the likelihood is high. Hence, maximising the likelihood finds the best fitting model.

This means that the maximum likelihood method is based on $P(\mathbf{d}|\mathbf{m})$, the conditional probability of $\mathbf{d}$ given $\mathbf{m}$. If $\mathbf{d}$ is observed, this can be interpreted as a function of $\mathbf{m}$, the likelihood function. Depending on the error statistics that are assumed for

the data, the maximum likelihood method yields different solutions. For example, the least squares solution follows for Gaussian errors and the least absolute value solution follows from double sided exponential errors. Both methods will be discussed in the following sections.

3.2.2. Least Squares Inversion

Let $\mathbf{d} = \mathbf{Am}$ be an over-determined problem with Gaussian errors. The distribution of the data errors is then given by the conditional probability of $\mathbf{d}$ given $\mathbf{m}$

$$P(\mathbf{d}|\mathbf{m}) = \frac{1}{(2\pi)^{N/2}|\mathbf{C}^{(\mathbf{d})}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{Am})^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}(\mathbf{d}-\mathbf{Am})\right). \tag{3.9}$$

If the data $\mathbf{d}$ are observed, $P(\mathbf{d}|\mathbf{m})$ can be interpreted as a function of $\mathbf{m}$, the likelihood function

$$\mathcal{L}(\mathbf{m}) \propto \exp\left(-\frac{1}{2}(\mathbf{d}-\mathbf{Am})^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}(\mathbf{d}-\mathbf{Am})\right). \tag{3.10}$$

Eq. 3.10 can be used to calculate the likelihood that $\mathbf{d}$ occured given a certain model $\mathbf{m}$. To find the most likely model $\mathbf{m}^{ML}$ for $\mathbf{d}$, $\mathcal{L}(\mathbf{m})$ has to be maximised, which is equivalent to minimising the negative of the exponent in Eq. 3.10, called the data misfit

$$E(\mathbf{m}) = (\mathbf{d}-\mathbf{Am})^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}(\mathbf{d}-\mathbf{Am}). \tag{3.11}$$

Solving $\partial E/\partial \mathbf{m} = 0$, the maximum likelihood (ML) solution is then given by

$$\mathbf{m}^{ML} = \left[\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{A}\right]^{-1}\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{d}, \tag{3.12}$$

provided that $\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{A}$ is non-singular. This is an important result that is equivalent to weighted least squares solution for the case of a diagonal covariance matrix $\mathbf{C}^{(\mathbf{d})}$. Hence, the ML method for Gaussian data errors is equivalent to the intuitive least squares solution.

It can be shown that the model covariance matrix $\mathbf{C}^{(\mathbf{m})}$ is given by

$$\mathbf{C}^{(\mathbf{m})} = \left[\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{A}\right]^{-1}, \tag{3.13}$$

which then leads to

$$\mathcal{L}(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\left(\mathbf{m}-\mathbf{m}^{ML}\right)^{\mathsf{T}}\left(\mathbf{C}^{(\mathbf{m})}\right)^{-1}\left(\mathbf{m}-\mathbf{m}^{ML}\right)\right). \tag{3.14}$$

This is a second important result that shows that for linear inverse problems with Gaussian distributed errors, the uncertainty for the model likelihood is also Gaussian distributed with mean $\mathbf{m}^{ML}$ and covariance $\mathbf{C}^{(\mathbf{m})}$.

If $\mathbf{A}$ is under-determined, no unique solution exists, and there is an infinite number of models that exactly fit the data. In this case, a modified least squares approach can be applied that introduces constraints to address the non-uniqueness of the problem. The constraints are introduced in terms of Lagrange multipliers that are generally used to find the minimum of a multivariate function. In this case, the smallest solution (the model vector with minimum length) is sought. The data misfit

$$E(\mathbf{m}) = \mathbf{m}^{\mathsf{T}}\mathbf{m} + 2\mathbf{l}(\mathbf{A}\mathbf{m}-\mathbf{d}) \tag{3.15}$$

is minimised with respect to $\mathbf{m}$ and $\mathbf{l}$, where $\mathbf{l}$ is a vector of Lagrange multipliers. The smallest model solution is then given by

$$\mathbf{m} = \mathbf{A}^{\mathsf{T}}\left[\mathbf{A}\mathbf{A}^{\mathsf{T}}\right]^{-1}\mathbf{d}, \tag{3.16}$$

provided that $\mathbf{A}\mathbf{A}^{\mathsf{T}}$ is non-singular. It should be noted that the smallest model is a mathematical solution to avoid the non-uniqueness, but might not be physically meaningful.

So far, the solutions only covered over- and under-determined systems. More general approaches exist to address any kind of problem and also guarantee stable solutions for ill-conditioned problems.

3.2.3. Stabilising Ill-Conditioned Inverse Problems

Singular value decomposition (SVD) and regularisation are two different approaches that can deal with ill-conditioned and singular systems of equations. SVD (Press et al., 1997) poses a mathematical way to stabilise problems, whereas regularisation yields solutions that are stabilised by including prior knowledge about the model and can be more physically meaningful.

The SVD theorem states that any $N \times M$ matrix $\mathbf{A}$ can be decomposed into matrices

$$\mathbf{A} = \mathbf{U}\mathbf{W}\mathbf{V}^{\mathsf{T}}, \tag{3.17}$$

where $\mathbf{U}$ is a $N \times M$ matrix containing the eigenvectors of $\mathbf{A}\mathbf{A}^{\mathsf{T}}$ in columns, $\mathbf{V}$ is an $M \times M$ matrix with the columns consisting of the eigenvectors of $\mathbf{A}^{\mathsf{T}}\mathbf{A}$, and $\mathbf{W}$ is a $M \times M$ diagonal matrix with the singular values w_i on the diagonal in descending order. The singular values w_i are the square roots of the eigenvalues of $\mathbf{A}\mathbf{A}^{\mathsf{T}}$ and $\mathbf{A}^{\mathsf{T}}\mathbf{A}$. In general, $\mathbf{W}$ has $p = \mathrm{Rank}(\mathbf{W}) \leq \min(N, M)$ non zero singular values which quantifies the amount of linearly independent information in the problem. For a singular $\mathbf{A}$, the first p columns of $\mathbf{U}$ are the basis for a subspace that can be reached by $\mathbf{A}$, the *range* of $\mathbf{A}$. The last $M - p$ columns of $\mathbf{V}$ are the basis for the *null space* $\mathcal{M}^{\perp}$. All annihilators $\mathbf{m}^{\perp} \in \mathcal{M}^{\perp}$ map to the null vector in the range of $\mathbf{A}$. The space that is spanned by the last $M - p$ column vectors in $\mathbf{V}$ cannot be reached by $\mathbf{A}$.

The inverse of $\mathbf{W}$ may not exist. However, a *pseudo inverse*

$$\mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^{\mathsf{T}} \tag{3.18}$$

can be defined where $\tilde{\mathbf{W}}^{-1}$ is a diagonal matrix with the i^{th} diagonal element set to

$$\tilde{w}_i^{-1} = \begin{cases} 1/w_i & \text{for i=1,...,p} \\ 0 & else \end{cases} \tag{3.19}$$

The SVD solution is then given by

$$\mathbf{m} = \mathbf{V}\tilde{\mathbf{W}}^{-1}\mathbf{U}^{\mathsf{T}}\mathbf{d}. \tag{3.20}$$

Further, it can be useful to set diagonal elements of $\tilde{\mathbf{W}}^{-1}$ to zero for small w_i to further stabilise the inversion, resulting in a trade off between parameter resolution and variance. It is often not obvious how many $1/w_i$ should be set to zero. Generally, too small a misfit would result in fitting noise on the data whereas too large a misfit means not representing the data well. For Gaussian data errors, the data misfit is chi-square (χ^2) distributed. The expected value of the χ^2 misfit for Gaussian data errors is

$$\langle \chi^2 \rangle = N. \tag{3.21}$$

The data should thus be fitted by removing small w_i, until $E(\mathbf{m}) \approx N$.

An alternative approach to stabilising inversion is regularisation. For the regularised solution, a function is minimised that includes data misfit and a prior misfit or regularisation term

$$\psi = (\mathbf{d} - \mathbf{Am})^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}(\mathbf{d} - \mathbf{Am}) + \mu(\mathbf{m} - \hat{\mathbf{m}})^{\mathsf{T}}\mathbf{H}(\mathbf{m} - \hat{\mathbf{m}}), \tag{3.22}$$

where $\mathbf{H}$ is the regularisation matrix (considered later) and μ is a trade-off parameter that controls how much importance is attributed to the data and prior information. The model $\hat{\mathbf{m}}$ represents the prior knowledge for the model. By minimising the function with respect to $\mathbf{m}$, the regularised solution

$$\mathbf{m} = \hat{\mathbf{m}} + \left[\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{A} + \mu\mathbf{H}\right]^{-1}\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}(\mathbf{d} - \mathbf{A}\hat{\mathbf{m}}) \tag{3.23}$$

is obtained. Here, μ and $\mathbf{H}$ are chosen in a way so that $\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{A} + \mu\mathbf{H}$ is non singular. Depending on the problem, μ and $\mathbf{H}$ can be chosen in different ways. A simple diagonal loading of the matrix can be achieved by using the identity matrix for $\mathbf{H}$ and $\hat{\mathbf{m}} = 0$. This results in a smallest model via smooth suppression of small singular values (in contrast to the singular value truncation SVD).

In the case of available prior knowledge about the model parameters, the prior parameter estimates can be applied in $\hat{\mathbf{m}}$ and weighted by their uncertainties by using the inverse of the model covariance matrix $\mathbf{C}_{\hat{\mathbf{m}}}^{-1}$ as a regularisation matrix. This is a powerful regularisation that assumes Gaussian distributed priors for the model parameters

Other complex options for regularisation involve higher order methods that include different order model derivatives. These are useful for problems with a natural ordering, e.g., functions of space or time, and can produce flattest (using first derivatives) or smoothest (using second derivatives) models.

The trade-off parameter μ is generally chosen so that $\langle\chi^2\rangle \approx N$, which results in the simplest model consistent with the data. The possibility to include prior knowledge to condition the inverse problem makes regularised inversion a very important and powerful tool for linear inverse problems. Applications in underwater acoustics can be found in e.g., Dosso et al. (1998) and Dosso and Collison (2002).

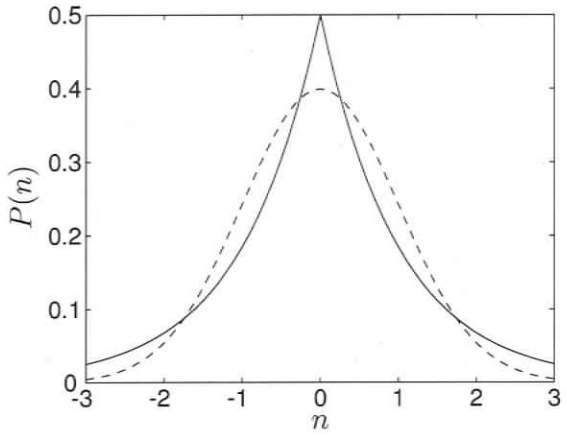

Fig. 3.1: Comparison of Gaussian (dashed) and double exponential (solid) error distributions. The double exponential distribution shows heavier tails and is more peaked than the Gaussian distribution.

3.2.4. Least Absolute-Value Inversion

L_2 norm or least squares inversion are based on the assumption of Gaussian error statistics. This assumption is ubiquitous in underwater acoustics, however other cases exist. If data residuals have heavier tails and/or are more peaked (Fig. 3.1), the assumption of double exponential error distributions (assuming uncorrelated errors)

$$P(n) = \frac{1}{2\sigma} \exp\left(-\frac{|n - \langle n \rangle|}{\sigma}\right) \tag{3.24}$$

can be more appropriate than simply interpreting heavy tails as data outliers.

Using double exponential error statistics for uncorrelated errors in a maximum likelihood approach leads to the L_1 error norm

$$\chi^1 = \sum_{i=1}^{N} \left| \frac{d_i - d_i(\mathbf{m})}{\sigma_i} \right| . \tag{3.25}$$

In linear inverse theory, there are a number of different approaches to use a L_1 norm in the inversion, however, in contrast to least squares methods, no analytic solution exists. Linear programming with the *simplex algorithm* (Kuenzi et al., 1971) and iteratively reweighted least squares inversion are common methods to use L_1 norms but will not be discussed here. It is worth noting that L_1 norms and any other norm are trivial to implement in the fully nonlinear numerical approach that will be discussed in Chapter 4 and is used throughout this thesis.

3.3. Linearised Inverse Problems

Many inverse problems of interest are in fact nonlinear. Nonlinear problems are often divided into groups, depending on the degree of nonlinearity. Weakly nonlinear problems can generally be solved with a linearised inversion and covariance and linearised resolution estimates are good approximations. Quasi nonlinear problems can be linearised and good solutions can be found after a few iterations. Strongly nonlinear problems cannot be solved satisfactorily with linearisation.

Newton's method of linearisation and iteration is based on a Taylor-series expansion of the forward model about an arbitrary starting point. In the Taylor expansion, nonlinear terms are neglected to find a linear solution. If the replica data for the first linear solution are not close enough to the data, further iterations can be used to improve the solution until convergence is satisfied. Newton's method's practicality is limited since it cannot escape from local minima in the parameter space. Hence, this and similar methods are often referred to as local methods.

The general problem is that it is impossible to tell the degree of nonlinearity without trying to invert the data with different methods. A fully nonlinear approach without any of the above restrictions can be formulated through inversion by forward modelling and Bayesian statistics. This nonlinear Bayesian inversion is discussed in the next chapter.

4. Nonlinear Bayesian Inversion

This chapter provides an overview of the nonlinear Bayesian inversion formulation developed and applied in this thesis to estimate geoacoustic model parameters; more complete treatments of Bayesian theory can be found in Tarantola (1987); Sambridge and Mosegaard (2002); Mosegaard and Sambridge (2002) and Sen and Stoffa (1995). The nonlinear formulation derived in this chapter uses efficient numerical optimisation and integration methods to extract the information content from the data. The approach is computationally efficient enough to be feasible for complex problems.

This approach is especially well suited for strongly nonlinear problems that have many local minima in the misfit function, where a linearisation does not characterise the parameter space sufficiently.

4.1. Bayes' Rule

Bayes' rule can be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \tag{4.1}$$

where $\mathbf{m} \in \mathbb{R}^M$ and $\mathbf{d} \in \mathbb{R}^N$ are random variables that represent the model parameters and data, respectively. $P(\mathbf{m}|\mathbf{d})$ is the conditional probability density function (PDF) of $\mathbf{m}$ given $\mathbf{d}$, $P(\mathbf{d}|\mathbf{m})$ is the conditional PDF of $\mathbf{d}$ given $\mathbf{m}$, and $P(\mathbf{m})$ is the conditional PDF of $\mathbf{m}$ independent of $\mathbf{d}$. $P(\mathbf{m})$ is also called the model prior distribution. $P(\mathbf{d})$ is the conditional PDF of $\mathbf{d}$ and represents a constant factor once the data are observed. In this case, $P(\mathbf{d}|\mathbf{m})$ is interpreted as a likelihood function $\mathcal{L}(\mathbf{d}|\mathbf{m})$. $\mathcal{L}(\mathbf{d}|\mathbf{m})$ quantifies the likelihood that the observed $\mathbf{d}$ occurred given a parameter set $\mathbf{m}$. For observed data $\mathbf{d}$, the likelihood function can be interpreted as a function of

$\mathbf{m}$. Equation 4.1 can then be written

$$P(\mathbf{m}|\mathbf{d}) = \frac{\mathcal{L}(\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}. \tag{4.2}$$

The posterior probability density (PPD) $P(\mathbf{m}|\mathbf{d})$ constitutes the state of information of $\mathbf{m}$ given the prior information $P(\mathbf{m})$ and the data $\mathbf{d}$. The likelihood function $\mathcal{L}(\mathbf{m})$ can generally be expressed as $\mathcal{L}(\mathbf{m}) \propto \exp(-E(\mathbf{m}))$ where $E(\mathbf{m})$ is an appropriate error function (considered later). Equation (4.2) then becomes

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp(-\phi(\mathbf{m}))}{\int\limits_{\mathcal{M}} \exp(-\phi(\mathbf{m}'))d\mathbf{m}'}, \tag{4.3}$$

where the integration is over the model space $\mathcal{M} \subset \mathbb{R}^M$ and $\phi(\mathbf{m})$ is the generalised misfit

$$\phi(\mathbf{m}) = E(\mathbf{m}) - \log_e P(\mathbf{m}). \tag{4.4}$$

The generalised misfit is not restricted to a certain prior distribution. For example, uniform prior distributions of the form

$$P(\mathbf{m}) = \begin{cases} \text{const.} & \forall\, \mathbf{m}^- \leq \mathbf{m} \leq \mathbf{m}^+ \\ 0 & \text{else} \end{cases}, \tag{4.5}$$

where $\mathbf{m}^+$ and $\mathbf{m}^-$ are the upper and lower parameter bounds, are commonly used but other model prior distributions are possible. Any other prior distributions (such as Gaussian) can also be used. In sequential Bayesian inversions, the PPD of one inversion can be used as the prior distribution for a subsequent inversion (see Chapter 6.6).

The PPD represents the full solution to the inverse problem in the Bayesian formulation. However, due to the PPD's multidimensional nature, interpretation is not trivial and properties such as the maximum a posteriori (MAP) estimate $\hat{\mathbf{m}}$, the mean model estimate $\mathbf{m}^{(mean)}$, the model covariance matrix $\mathbf{C}^{(\mathbf{m})}$, and marginal probability distributions $P(m_i|\mathbf{d})$, must be calculated to provide parameter estimates, uncertain-

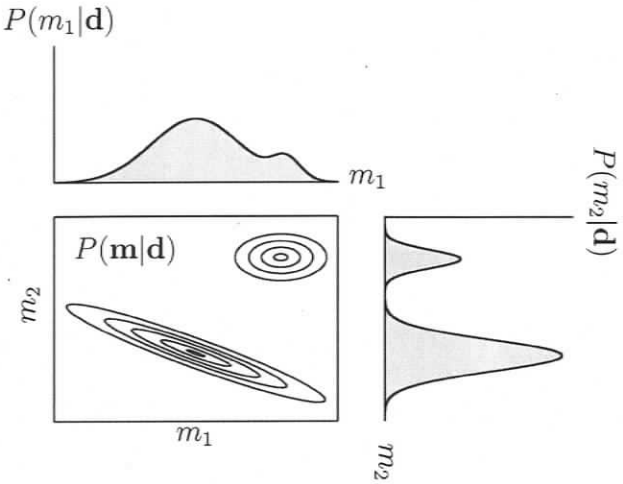

Fig. 4.1: One dimensional probability distributions integrate over all parameter dimensions of the PPD with the exception of the parameter of interest, as here illustrated for a two dimensional PPD.

ties and inter-relationships:

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max} P(\mathbf{m}|\mathbf{d}) \tag{4.6}$$

$$\mathbf{m}^{(mean)} = \int_{\mathcal{M}} \mathbf{m}' P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \tag{4.7}$$

$$\mathbf{C}^{(\mathbf{m})} = \int_{\mathcal{M}} (\mathbf{m}' - \mathbf{m}^{(mean)})(\mathbf{m}' - \mathbf{m}^{(mean)})^{\mathsf{T}} P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \tag{4.8}$$

$$P(m_i|\mathbf{d}) = \int_{\mathcal{M}} \delta(m_i' - m_i) P(\mathbf{m}'|\mathbf{d}) d\mathbf{m}' \,, \tag{4.9}$$

where δ denotes the Dirac delta function. As can be seen in Eq. 4.9, one dimensional marginal probability distributions are an integral property, where the integration is over all parameter dimensions with the exception of the parameter of interest (see Fig. 4.1) Higher-dimensional joint marginal distributions can be defined similar to Eq. (4.9). Uncertainties of parameter estimates can also be quantified in terms of highest probability density (HPD) credibility intervals. The β% HPD credibility interval is defined as the interval of minimum width that contains β% of the area of the marginal probability distribution. Interrelations of model parameters can be quantified by the model correlation matrix given by normalising the covariance matrix

$$R_{ij} = C_{ij}/(C_{ii}C_{jj})^{1/2} \,. \tag{4.10}$$

Thus $R_{ij} \in [-1, 1]$ with $R_{ij} = 1$ indicating perfect correlation and $R_{ij} = -1$ perfect

anti-correlation. The correlation between two model parameters can be visualised by joint marginal probability distributions.

While analytic solutions to Eq. 4.6–4.9 exist for linear inverse problems, solutions for nonlinear problems must be found using numerical optimisation and integration.

4.2. PPD Optimisation

The MAP model (Eq. 4.6) can be found by numerical optimisation of the PPD such as global searches and hybrid algorithms that minimise the data misfit which is equivalent to maximising the PPD. The following section gives a brief overview of optimisation algorithms and then describes adaptive simplex simulated annealing (ASSA) that will be used for PPD optimisation in this thesis.

4.2.1. Grid Search and Monte Carlo Search

Grid search is a brute force method where the model parameter space is divided into discrete values. The algorithm then tries all combinations of parameter values and the combination with the smallest misfit between measured and replica data is taken to be the best model. For M parameters and P discrete values for each parameter, P^M forward model computations have to be performed, which can quickly become impossible for today's computers. More sophisticated grid searches start searching on a coarse grid to identify regions of importance and then refine the search in those regions (Tolstoy, 1998).

Monte Carlo Search (Hammersley and Handscomb, 1964) is based on the idea of randomly choosing models and calculating the misfit for each model, which is more efficient than grid searches.

Both methods are inefficient for large numbers of parameters. However, they are simple to implement and can be run as parallel code where performance scales with the number of processors used.

4.2.2. Simulated Annealing

Simulated annealing (SA, Kirkpatrick et al., 1983) is a directed Monte Carlo method that combines basic ideas of combinatorial optimisation with the statistical mechanics formulation of annealing. Directed Monte Carlo searches converge much faster

than purely random searches since they use certain functions to only sample important parts of the parameter space. SA has been successfully applied to a variety of underwater acoustic applications, e.g., Collins et al. (1992), Lindsay and Chapman (1993), and Fallat and Dosso (1998). Annealing is the physical process of slowly cooling a hot liquid. During this process, crystals form that resemble a low state of energy. Similarly, the SA algorithm slowly cools the "*melted*" set of parameters until they are in the lowest energy state which is equivalent to the best global parameter estimate. If the inversion is cooled too fast, the algorithm can get stuck in a local misfit minimum which is analogous to amorphous material (e.g., glass) forming when quickly cooling a liquid.

In statistical mechanics, the probability that a system of many atoms in equilibrium is described by a configuration of $\mathbf{r}$ is given by the Boltzmann or Gibbs distribution

$$P(\mathbf{r}) = \frac{1}{Z(T)} \mathrm{e}^{-E(\mathbf{r})/k_B T}, \tag{4.11}$$

where

$$Z(T) = \sum_{\mathbf{r} \in S} \mathrm{e}^{E(\mathbf{r})/k_B T} \tag{4.12}$$

is the partition function that ensures normalisation, k_B is the Boltzmann constant, T the absolute temperature, E the energy of the system and S denotes the parameter space. For $T > 0$, the probability is greater than zero for all configurations. This means that even for low temperatures, there is a small probability that the system is in a high energy state. For fixed T, the system can undergo changes that increase or decrease E, i.e., the system can escape from local minima. The lower T, the less likely transitions are accepted that increase E. If T is lowered slowly enough for the system to remain in a Boltzmann equilibrium, the global minimum energy configuration will be reached. The Metropolis algorithm (Metropolis et al., 1953; Hastrup, 1970) was developed to numerically sample from Eq. 4.11 to simulate the thermal equilibrium of a solid body in a heat-bath:

In a system $\mathbf{r}$, perturb one parameter r_i by δ so that

$$r_i \rightarrow r_i' = r_i + \delta \tag{4.13}$$

$$\mathbf{r}' = [r_1, \ldots, r_i', \ldots, r_M]\,. \tag{4.14}$$

The energy for the perturbed system can then be written

$$E(\mathbf{r}') = E(\mathbf{r}) + \Delta E\,. \tag{4.15}$$

The probability of a change ΔE occurring is then given by

$$P(\Delta E) = \mathrm{e}^{-\Delta E/k_B T}\,. \tag{4.16}$$

The Metropolis algorithm accepts all perturbations that decrease energy ($\Delta E < 0$). If the energy increases, the probability $P(\Delta E)$ is compared to a uniform random number $\xi \in [0, 1]$ and the perturbation is accepted if $\xi \leq P(\Delta E)$. The probability of accepting perturbations that increase energy decreases with decreasing T and increasing ΔE.

Applying the Metropolis criterion over many perturbations at fixed temperature simulates a system in equilibrium. By slowly decreasing T, annealing is simulated while keeping the system in or close to equilibrium.

Simulated annealing is usually applied to optimisation problems where a function of many parameters is minimised. The main problem with SA algorithms is the slow rate of convergence. Different approaches exist to yield faster convergence. Quenching the algorithm involves setting the temperature to zero to force the SA algorithm to only move downhill. This will accelerate convergence once the algorithm is close to a solution. Fast simulated annealing (Szu and Hartley, 1987) uses a Cauchy distribution for parameter perturbations and decreases the distribution width with temperature. This speeds up convergence by reducing the number of large perturbations at low temperatures and hence avoiding high rejection rates that slow the algorithm down. Very fast simulated reannealing (Ingber, 1989) introduces individual cooling schedules for each model parameter to enhance convergence. Unfortunately, while these approaches can speed up convergence they require correct setting of control parameters which can be complicated. Finally, SA does not include any form of memory of good models. Due to the ability to move uphill out of minima, the algorithm can "forget" a particularly good model. Lindsay and Chapman (1993) developed an adaptive SA that included a memory about the topography of the search space to avoid this.

Other global optimisation schemes such as genetic algorithms (GA, see e.g., Goldberg, 1989) have been successfully applied to geoacoustic inverse problems (Gerstoft, 1994) but will not be discussed here.

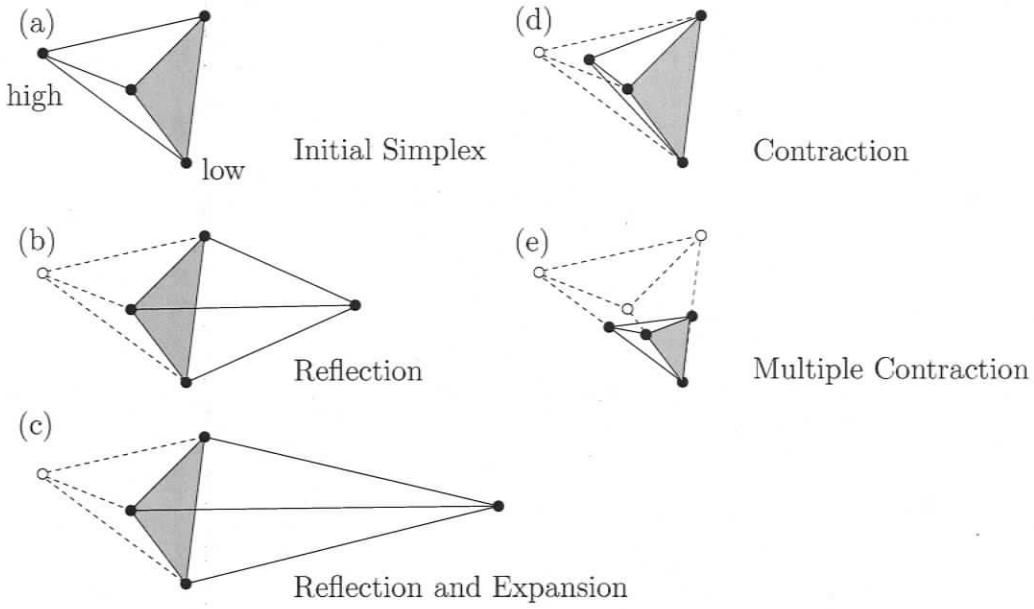

Fig. 4.2: The downhill simplex scheme in three dimensions after Press et al. (1997)

4.2.3. Hybrid Inversion: Adaptive Simplex Simulated Annealing

Global search methods, as described above, are effective at navigating parameter spaces with many local minima. Inefficiency arises when the algorithm is close to a good solution and needs to move downhill (following the local gradient in the misfit function). Close to a good solution, the misfit function can form long, oblique valleys due to parameter correlations. Global search methods are very inefficient in moving downhill in these parts of the parameter space. However, these valleys can be handled better by local gradient based searches that evaluate the local gradient of the misfit to move downhill (Fallat and Dosso, 1999).

There are many different approaches to combine the different local and global methods to search the complete space but at the same time move downhill efficiently. These are commonly referred to as hybrid algorithms. The following will discuss one of these methods. Adaptive simplex simulated annealing (ASSA, Dosso et al. (2001)) combines the global advantages of very fast simulated annealing with the local advantage of the downhill simplex method (DHS, Nelder and Mead (1965)).

DHS is a geometric scheme to move downhill without explicitly computing the misfit function gradient. Since the algorithm does not work with the exact gradient, it moves downhill less efficiently than other gradient methods (e.g., Newton's method).

DHS is not a purely local algorithm, but does get stuck in local minima since it only accepts downhill steps. For an M-dimensional parameter space, DHS is based on a simplex of $M+1$ models as illustrated in Fig. 4.2 (a). The data misfit for each of the $M+1$ models is computed and the models with highest and lowest misfit are identified. To improve the model with the highest misfit, a series of geometric operations are performed, as illustrated in Fig. 4.2 (b)-(e). First, the model with the highest misfit is reflected through the simplex and the new misfit is computed. If the reflection produces an energy lower than the second highest energy of the simplex, the move is accepted. If the reflection produced a misfit lower than the lowest energy of the simplex, an expansion by a factor of two in the same direction is attempted and accepted if it further reduces the energy. If the reflection does not decrease the energy below the second highest, the reflection is rejected and a contraction by a factor of two towards the lowest misfit model is tried and accepted if it lowers the energy. If none of the above steps is accepted, a multiple contraction is performed in which $M-1$ models are contracted halfway towards the model with the lowest energy. The above set of moves is repeated until convergence. Convergence is accomplished once the simplex has essentially contracted to a point, which can be tested by

$$\frac{E_{\text{high}} - E_{\text{low}}}{(E_{\text{high}} - E_{\text{low}})/2} < \epsilon\,, \tag{4.17}$$

where ϵ is a small arbitrary number. Typical values of ϵ are in $[10^{-6}, 10^{-4}]$.

The hybrid optimisation method of ASSA combines the DHS scheme adaptively with very fast SA. ASSA works on a simplex of models, where each DHS step is followed by a random perturbation to all model parameters. Acceptance of each step is decided by applying the Metropolis Algorithm (Eq. 4.16), i.e., uphill steps are sometimes accepted. The random perturbations C are drawn from a Cauchy distribution that is scaled adaptively for each parameter throughout the inversion. The Cauchy distribution (see Fig. 4.3) is given by

$$P(n; n_0, \gamma) = \frac{1}{\pi}\left[\frac{\gamma}{(n-n_0)^2 - \gamma^2}\right], \tag{4.18}$$

where n_0 is the location of the peak of the distribution and γ specifies the half-width at half-maximum. C can be drawn from the Cauchy distribution by

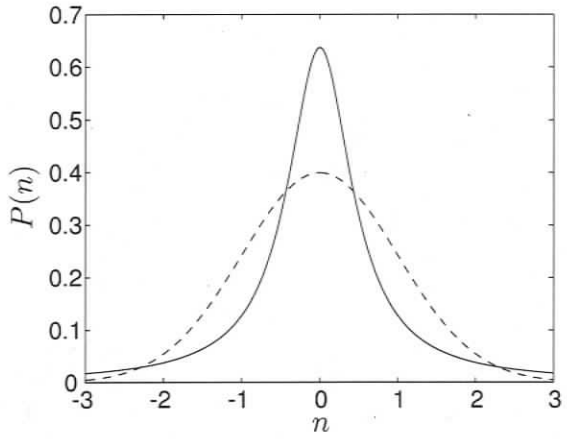

Fig. 4.3: Comparison of Gaussian (dashed) distribution and the Cauchy (solid) sampling distribution.

$$C = tan(\eta_2 \pi/2)\,, \tag{4.19}$$

where η_2 is a uniform random variable. The adaptive scaling of the distribution width for each parameter is based on an average of recently accepted perturbations. Perturbations are initiated with a large width that essentially results in a uniform distribution over the model parameter bounds $[\mathbf{m}^-, \mathbf{m}^+]$. As the inversion proceeds, the distribution width is scaled by $s > 1$ times the running mean magnitude of the last S accepted perturbations. Dosso et al. (2001) found that inversions are not very sensitive to these scaling parameters, and that $S = 30$ and $s = 3$ generally work well.

Cooling is applied after a fixed number of perturbations are accepted using

$$T_j = \beta^j T_0\,, \tag{4.20}$$

where $\beta < 1$ is a constant. The starting temperature T_0 should be chosen so that the problem is in a "melted" state, resulting in an acceptance rate of more than 90%. A safe value can be obtained by using the highest energy in the simplex as a starting temperature.

The ASSA algorithm is efficient at searching the whole parameter space while not getting stuck in local minima, and the simplex provides a form of memory, ensuring that no good models are lost in the process of escaping from local minima. ASSA is used extensively for optimisation throughout this thesis.

4.3. PPD Integration: Metropolis Gibbs Sampling

The integral properties of the PPD given in Eq. 4.7–4.9 of the general form

$$I = \int_{\mathcal{M}} f(\mathbf{m}')P(\mathbf{m}'|\mathbf{d})d\mathbf{m}', \tag{4.21}$$

are computationally challenging. I can be efficiently estimated by importance sampling (Press et al., 1997), which preferentially draws random samples from those regions of the model space that contribute most to the integral by using a sampling function.

Let $g(\mathbf{m})$ be a nonuniform sampling function with the normalisation

$$\int g(\mathbf{m}')d\mathbf{m}' = 1, \tag{4.22}$$

that is used to draw Q models. Substituting $g(\mathbf{m})$ into Eq. 4.21 gives

$$\begin{aligned} I &= \int \frac{f(\mathbf{m}')P(\mathbf{m}'|\mathbf{d})}{g(\mathbf{m}')} g(\mathbf{m}')d\mathbf{m}' \\ &\approx \frac{1}{Q}\sum_{i=1}^{Q} \frac{f(\mathbf{m}_i)P(\mathbf{m}_i|\mathbf{d})}{g(\mathbf{m}_i)}. \end{aligned} \tag{4.23}$$

The Gibbs distribution, which describes the state of energy of a system of many atoms in statistical mechanics, is used as a sampling function in Gibbs sampling and is given by

$$P_G(\mathbf{m}) = \frac{\exp(-\phi(\mathbf{m})/T)}{\int_{\mathcal{M}} \exp(-\phi(\mathbf{m}')/T)d\mathbf{m}'}, \tag{4.24}$$

where T is a factor that resembles the absolute temperature in statistical mechanics. For $T = 1$, Eq. (4.24) has the form of the PPD (Eq. (4.3)). Using importance sampling with $P_G(\mathbf{m})$ as sampling function at $T = 1$ approximates Eq. (4.21) as

$$I \approx \frac{1}{Q}\sum_{i=1}^{Q} f(\mathbf{m}_i), \tag{4.25}$$

providing an unbiased estimate of the integral based on a sample of Q models. To collect Q models, Metropolis GS applies the Metropolis algorithm, which accepts a

perturbation to the system, if

$$\xi \leq \exp(-\Delta\phi/T), \tag{4.26}$$

where $\xi \in [0, 1]$ is a uniform random number. Therefore, the algorithm accepts all perturbations that lower the energy of the system and some that increase it. The Metropolis algorithm samples from a normalised Gibbs distribution without summing over all possible states.

It can be shown, using Markov chain analysis, that the Metropolis algorithm (MA) converges to the Gibbs distribution for large Q and can provide an unbiased estimate of the PPD (Rothman, 1986; Geman and Geman, 1984; Sen and Stoffa, 1995).

Since standard Gibbs sampling can be rather slow, more efficient implementations have been developed (Gilks et al., 1996). For geoacoustic inversion, Dosso (2002a) developed the more efficient fast Gibbs sampling algorithm (FGS), which substantially enhances computational performance for these problems. To avoid sampling improbable models, sampling starts at a high temperature which is rapidly decreased during sampling. Once a reasonable model is found, it serves as a start for sampling at $T = 1$. Models only contribute to the sample if collected at $T = 1$.

FGS defines the maximum perturbation size of model parameters as a factor times the largest perturbation accepted to that point. The perturbation size cannot be larger than the parameter bounds.

Correlated parameters result in narrow, oblique valleys of the data misfit function. Such valleys are difficult to navigate, since standard GS perturbs along the parameter axes. To avoid inefficiencies due to correlated parameters, an orthogonal transformation is applied. This transformation rotates the parameter axes into the direction of the oblique valleys in order to carry out perturbations. To achieve this, an estimate of the (symmetric) model covariance matrix $\mathbf{C}^{(\mathbf{m})}$ is computed and the eigen-value decomposition

$$\mathbf{\Lambda} = \mathbf{U}^{\mathsf{T}}\mathbf{C}^{(\mathbf{m})}\mathbf{U} \tag{4.27}$$

is performed, where $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues and $\mathbf{U}$ is a matrix containing the orthonormal eigenvectors. The transformation-pair is then

given by

$$\begin{aligned} \mathbf{m}' &= \mathbf{U}^{\mathsf{T}}\mathbf{m} \quad \text{and} \\ \mathbf{m} &= \mathbf{U}\mathbf{m}', \end{aligned} \tag{4.28}$$

where $\mathbf{m}'$ is in the rotated space $\mathcal{M}' \subset \mathbb{R}^M$. While perturbations are performed in the rotated space, the forward modelling is done in physical parameter space. The estimate of $\mathbf{C}^{(\mathbf{m})}$ is found from initial sampling. Experience shows that $\mathbf{C}^{(\mathbf{m})}$ converges fairly quickly (Sen and Stoffa, 1995, 1996) and that an approximate estimate is good enough to substantially speed up convergence. To decide whether or not the estimate of the covariance matrix is good enough, correlation matrices for two independent samples are compared to each other since the widely varying magnitudes of the elements of the covariance matrix complicate convergence tests. If the maximum difference between the correlation matrices is less than a pre-defined threshold, the estimate is considered to be sufficient and the union of both samples is used to compute a final rotation matrix $\mathbf{U}$.

During this phase of the sampling, also called *burn-in*, no accepted models are used to estimate the PPD. Once the burn-in is completed, sampling continues with two independent samples in rotated parameter space until the samples converge. Convergence is satisfied when the maximum difference of the cumulative marginal distributions of all model parameters between both samples falls below a pre-defined threshold. Only the sampling in rotated space contributes to the final estimate of the PPD.

Once in rotated space, the sampling is much faster. To speed up the burn-in phase of the inversion, it is possible to apply a linear estimate of the model covariance matrix as an initial rotation matrix. Equation 3.13 gave the linear model covariance matrix for a L_2-norm as

$$\mathbf{C}^{(\mathbf{m})} = \left[\mathbf{A}^{\mathsf{T}}(\mathbf{C}^{(\mathbf{d})})^{-1}\mathbf{A}\right]^{-1}, \tag{4.29}$$

where $\mathbf{C}^{(\mathbf{d})}$ is the data covariance matrix. $\mathbf{A}$ is the kernel or Jacobian matrix

$$A_{ij} = \frac{\partial d_i(\mathbf{m}_0)}{\partial m_j} \tag{4.30}$$

and contains partial derivatives of the data for model $\mathbf{m}_0$ with respect to the model parameters. In a nonlinear problem, these partial derivatives will not be valid for the

complete parameter space but only in the vicinity of model $\mathbf{m}_0$. However, the linear estimate for the model covariance matrix can still be a worthwhile local estimate to approximately rotate the problem into principal component space and speed up the initial sampling in the area around a good starting model $\mathbf{m}_0$ (as determined, for example, by an optimisation algorithm). The partial derivatives in Eq. 4.30 are evaluated numerically as central differences at the starting model. Hence, some attention must be paid to the stability of the derivative estimates by considering a range of step sizes in the central difference. A size for which the derivatives are stable (Press et al., 1997) is then chosen to compute the matrix. If the model contains parameters of substantially different sensitivities, the process should be done for each parameter individually. To ensure numerical stability in Eq. 3.13, the inverse is replaced by the pseudo inverse (see Eq. 3.18). Another approach to speed up burn-in for particularly difficult cases is to iteratively update the rotation matrix, which results in faster sampling with every update.

4.4. Massively Parallel ASSA and FGS

ASSA and FGS have been optimised on scalar computers in the past to perform inversions efficiently. However, the speed of the algorithm strongly depends on the performance of the forward model. To conduct research using complicated forward models more efficiently, the ASSA and FGS codes were implemented for this thesis for massively parallel computers using the message-passing interface (Gropp et al., 1999). The performance gain for parallel implementations depends strongly on the granularity of the algorithm. Granularity describes how independently different parts of the algorithm can perform computations without the need of communication between the various processes. Since ASSA is largely a serial code, its granularity is small. The main performance gain for optimisation was obtained by parallelising the forward model. This implementation will not be further discussed here.

Since FGS is a Monte Carlo method, it belongs to a group of algorithms that are commonly called *embarrassingly parallel* and have large granularity. FGS was implemented for parallel computers by distributing the sampling of the PPD over a large number of central processing units (CPUs). For the burn-in phase, each CPU samples for a certain number of forward steps and then passes the sample to the master CPU to carry out a convergence test of the various model covariance

estimates. Once convergence of the rotation matrix is obtained, the actual sampling of the PPD begins in parallel on all CPUs. It was found that after the burn-in, the performance scaled almost linearly with the number of CPUs used. The algorithm was tested with up to 80 CPUs with a speedup factor of close to 80.

5. Likelihood Function and Data Error Statistics

This chapter derives data misfit functions that are part of the generalised misfit (Eq. 4.4) from likelihood functions. Further, a framework to rigorously treat uncorrelated and correlated data errors is developed, including data covariance matrix estimates and quantitative statistical tests to validate statistical assumptions *a posteriori*.

5.1. Data Misfit Functions

Formulating the likelihood function $\mathcal{L}(\mathbf{m})$ requires specifying the data uncertainty distribution. Data uncertainties must include both measurement errors due to experimental processes, and theory errors due to model parameterisation and simplifications in the forward model. In general, the Bayesian inversion outlined above can be applied with arbitrary uncertainty distributions. In practice, however, the lack of information about the error distribution often suggests a mathematically simple distribution be assumed. In particular, Gaussian and double-sided exponential distributions are commonly considered, with their statistical parameters estimated from the data.

For N_i observed data $\mathbf{d}_i$ at each of F frequencies f_i with unbiased Gaussian-distributed random errors uncorrelated from frequency to frequency, the likelihood function $\mathcal{L}(\mathbf{m})$ is given by

$$\mathcal{L}(\mathbf{m}) \propto \prod_{i=1}^{F} \exp\left(-\frac{1}{2}\big(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})\big)^{\mathsf{T}} (\mathbf{C}_i^{(\mathbf{d})})^{-1} \big(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})\big)\right), \tag{5.1}$$

where $\mathbf{d}_i(\mathbf{m})$ are the replica data computed for model $\mathbf{m}$ and $\mathbf{C}_i^{(\mathbf{d})}$ are the F $N_i \times N_i$

data covariance matrices. The data misfit function is then given by the negative of the exponents in Eq. 5.1

$$E_2(\mathbf{m}) = \sum_{i=1}^{F} \frac{1}{2} \big(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})\big)^{\mathsf{T}} (\mathbf{C}_i^{(\mathbf{d})})^{-1} \big(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})\big) \,. \tag{5.2}$$

Hence, assuming Gaussian distributed errors leads to L_2-norm inversion. However, one of the strengths of the numerical approach chosen in this thesis is that data misfit functions for different error distributions can be applied. For instance, L_1-norm inversion corresponds to a double-sided exponential error distribution with an error function (Tarantola, 1987)

$$E_1(\mathbf{m}) = \sum_{i=1}^{F} \sum_{j=1}^{N_i} \left| \left(\mathbf{L}_i^{-1} \big(\mathbf{d}_i - \mathbf{d}_i(\mathbf{m})\big) \right)_j \right| , \tag{5.3}$$

where $\mathbf{L}$ is the lower triangular matrix of the Cholesky decomposition of the covariance matrix

$$\mathbf{C}_i^{(\mathbf{d})} = \mathbf{L}_i \mathbf{L}_i^{\mathsf{T}} \,. \tag{5.4}$$

In many cases, the covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ are not known. Depending on the data error statistics, different approaches can be applied to estimate the covariance matrices. If the data do not have significantly correlated errors, the covariance matrices can be approximated as diagonal $\mathbf{C}_i^{(\mathbf{d})} = \sigma_i^2 \mathbf{I}$ where $\mathbf{I}$ is the identity matrix and only the standard deviations σ_i have to be estimated. If error correlations cannot be neglected a covariance matrix must be estimated and applied in the inversion. Since not all elements of the covariance matrix can be estimated from N data, simplified covariance matrices are generally estimated (e.g., matrices of Toeplitz form). Alternative processes such as autoregressive moving average (ARMA) models can also be used to describe covariances with even less parameters (Rencher, 1998) but are not applied in this thesis. In this thesis, nonparametric approaches are used to estimate covariance matrices which do not assume a specific form of the matrix.

5.2. Inversion Assuming Uncorrelated Data residuals

This section will derive three different approaches to treat unknown standard deviations σ_i at each frequency in the inversion. Assuming uncorrelated errors, the data

covariance matrix at each frequency can be written as $\mathbf{C}_i^{(\mathbf{d})} = \sigma_i^2\mathbf{I}$. The likelihood function then becomes

$$\mathcal{L}(\mathbf{m}) = \prod_{i=1}^{F} \frac{1}{(2\pi)^{N_iF/2}} \frac{1}{\sigma_i^{N_i}} \exp\left[-\sum_{j=1}^{N_i} \left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2/(2\sigma_i^2)\right], \tag{5.5}$$

where $d_{i;j}$ represents the j^{th} datum at the i^{th} frequency. The first approach finds a global ML estimate for σ_i, by setting $\partial\mathcal{L}(\mathbf{m})/\partial\sigma_k \stackrel{!}{=} 0$, which leads to

$$\tilde{\sigma}_i = \left[\frac{1}{N_i} \left|\mathbf{d}_i - \mathbf{d}_i(\tilde{\mathbf{m}})\right|^2\right]^{1/2}. \tag{5.6}$$

Substituting Eq. 5.6 into Eq. 5.5 yields

$$\begin{aligned}\mathcal{L}(\mathbf{m}) = &\prod_{i=1}^{F} \frac{1}{(2\pi)^{N_iF/2}} N_i^{N_i/2} \exp\left[-\frac{N_i}{2}\right] \\ &\exp\left[\log_e\left[\prod_{i=1}^{F}\left(\sum_{j=1}^{N_i}\left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2\right)^{-N_i/2}\right]\right],\end{aligned} \tag{5.7}$$

which leads to the L_2-norm data misfit function

$$\begin{aligned}E_2(\mathbf{m}) &= -\log_e\left(\prod_{i=1}^{F}\left[\sum_{j=1}^{N_i}\left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2\right]^{-N_i/2}\right) \\ &= \sum_{i=1}^{F}\left(\frac{N_i}{2}\log_e\left[\sum_{j=1}^{N_i}\left(d_{ij} - d_{ij}(\mathbf{m})\right)^2\right]\right).\end{aligned} \tag{5.8}$$

Minimising this function (using numerical optimisation such as ASSA) provides the required ML estimate $\tilde{\mathbf{m}}$. To calculate $\tilde{\mathbf{m}}$, the function

$$E(\mathbf{m}) = \prod_{i=1}^{F}\sum_{j=1}^{N_i}\left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2 \tag{5.9}$$

can also be minimised using a global or hybrid search algorithm, e.g., ASSA. Equation 5.9 has the same global minimum as Eq. 5.8 (neglecting the logarithm does not change the location of the minimum) but is scaled differently due to the logarithm. Because of this, in some problems, it may be harder to find the global minimum with

Eq. 5.9, since the global minimum is scaled differently. Once $\tilde{\mathbf{m}}$ is obtained it is then substituted into Eq. 5.6 to calculate the ML estimates $\tilde{\sigma}_i$. These are then used in the Gibbs sampler that applies Eq. 5.2 which, in the case of uncorrelated data errors, simplifies to

$$E_2(\mathbf{m}) = \frac{1}{2}\sum_{i=1}^{F}\sum_{j=1}^{N_i}\frac{\left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2}{\tilde{\sigma}_i^2}\,. \tag{5.10}$$

The second approach is referred to as the full Bayesian approach. Here, the σ_i are treated explicitly as additional unknown parameters (nuisance parameters) in the inversion which results in the likelihood function

$$\mathcal{L}(\mathbf{m},\boldsymbol{\sigma}) = \prod_{i=1}^{F}\frac{1}{(2\pi)^{N_i F/2}}\exp\left[-\sum_{i=1}^{F}\sum_{j=1}^{N_i}\frac{\left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2}{2\sigma_i^2} + N_i\log_{\mathrm{e}}\sigma_i\right] \tag{5.11}$$

with the appropriate L_2 data misfit

$$E_2(\mathbf{m},\boldsymbol{\sigma}) = \sum_{i=1}^{F}\sum_{j=1}^{N_i}\frac{\left(d_{i;j} - d_{i;j}(\mathbf{m})\right)^2}{2\sigma_i^2} + N_i\log_{\mathrm{e}}\sigma_i\,. \tag{5.12}$$

The third approach considered here is an implicit approach that does not use a global ML estimate but rather a local ML estimate of σ. Further, the standard deviations are not explicitly included as additional parameters and therefore the approach uses less parameters compared to the full Bayesian approach, resulting in a faster algorithm. If no strong parameter correlations between standard deviations and other model parameters exist, the difference in performance is not expected to be large. The appropriate L_2 misfit function is already given in Eq. 5.8 but rewritten here

$$E_2(\mathbf{m}) = \sum_{i=1}^{F}\left(\frac{N_i}{2}\log_{\mathrm{e}}\left[\sum_{j=1}^{N_i}\left(d_{ij} - d_{ij}(\mathbf{m})\right)^2\right]\right). \tag{5.13}$$

The three different methods can be used to account for the standard deviations of the data. The idea behind the three different approaches is illustrated in a schematic diagram in Fig. 5.1. This figure shows the search space for one parameter x and one standard deviation σ of the data. A minimum in the misfit function is indicated by contour lines. A global maximum likelihood (ML) estimation would find the standard deviation that matches the minimum in the misfit function and FGS would

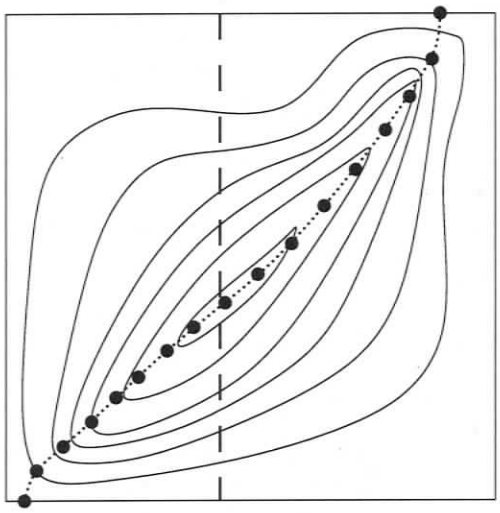

Fig. 5.1: Schematic comparison of three different methods to include standard deviation. x is a model parameter and σ the standard deviation of the data. The contour lines indicate a minimum in the misfit function. The dashed line illustrates sampling for the ML case and the dotted line for the implicit case. The full Bayesian approach samples the entire 2-D space.

then sample along a straight line. In multi-frequency inversion, a different standard deviation is estimated for each frequency.

The full Bayesian approach samples the whole x-σ space in Fig. 5.1. Finally, the implicit approach, samples along a curved line, making a local, optimal estimate of σ for each value of x.

Comparing inversion results for all three methods showed that resulting marginal probability distributions are similar for reflectivity inversion. The implicit approach has computational advantages over the full Bayesian approach while working without a global ML estimate for the standard deviations. Dosso and Wilmut (2005) showed that in matched field inversion, the implicit sampling gives similar results to explicitly including the standard deviations in the inversion. However, the implicit approach is computationally more efficient.

5.3. Data with correlated errors

To date, the simplifying assumption of uncorrelated data errors has almost always been made in geoacoustic inversion. In practice, however, strong error correlations are often present (Holland, Dettmer, and Dosso, 2005; Dosso, Nielsen, and Wilmut, 2006) due to theory errors. It is important to verify whether or not significant error correlations are present in Bayesian inversion, since neglecting these can result in unreasonably tight uncertainty bounds for the model parameters. One approach to

treat correlated errors is to down-sample the data below the correlation length of the errors, with an obvious loss of information. In this work, we apply a rigorous approach to analyse data residuals to test for randomness and incorporate error correlations into the inversion without loss of data information. A data covariance matrix is estimated from the data residuals using a nonparametric approach and is applied in the likelihood function of the Bayesian inversion.

The covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ can be estimated as follows. Initially assuming uncorrelated errors, the data covariance can be written as $\mathbf{C}_i^{(\mathbf{d})} = \sigma_i^2\mathbf{I}$. By minimising Eq. 5.9, $\tilde{\mathbf{m}}$ is found and used to calculate $\tilde{\sigma}_i$ via Eq. 5.6. With the estimates $\tilde{\sigma}_i$, another minimisation is carried out using Eq. (5.2) with $\mathbf{C}_i^{(\mathbf{d})} = \tilde{\sigma}_i^2\mathbf{I}$ to produce the MAP estimate $\hat{\mathbf{m}}^{(1)}$. Assuming a stationary process, the data covariance matrices at each frequency can then be estimated by calculating the autocovariance function of the data residuals $\mathbf{n}_i^{(1)} = \mathbf{d}_i - \mathbf{d}_i(\hat{\mathbf{m}}^{(1)})$. The autocovariance is defined as an ensemble average. However, the data provide only a single, realisation of the random process. For uniformly spaced data, the ensemble average can be replaced by an average over the N_i samples, assuming an ergodic process. The j^{th} element of the autocovariance function is then given by

$$c_{i;j} = \frac{1}{N_i} \sum_{k=0}^{N_i-j-1} (n_{i;j+k} - \bar{n}_i)(n_{i;k} - \bar{n}_i)\,, \tag{5.14}$$

where $\bar{n}_i$ is the mean of the residuals. Here, $1/N_i$ does not form a strict average but rather damps terms in Eq. 5.14 where $(N_i - j - 1)$ is small. These terms correspond to large lags with only few data pairs to constrain the covariance estimate. Further, it is unlikely to have strong correlations for widely spaced data. Hence the uncertainty is large and covariances should be small. Every term $c_{i;j}$ is then used to build the j^{th} diagonal of the initial data covariance matrix estimate $\hat{\mathbf{C}}_{(i,1)}^{(\mathbf{d})}$. This builds a Toeplitz or banded matrix so that every lag has a certain constant correlation parameter. This form is not as strict as, e.g., a first-order autoregressive model. The covariance matrix can then be applied in a subsequent inversion for a MAP estimate $\hat{\mathbf{m}}^{(2)}$. If the covariance matrix computed for $\hat{\mathbf{m}}^{(2)}$ using Eq. 5.14 is sufficiently close to $\hat{\mathbf{C}}_{(i,1)}^{(\mathbf{d})}$, the procedure can be considered to have converged. If not, further iterations can be performed. Synthetic data examples showed that the recovered estimate of the covariance matrix converged nicely after the second iteration. Figure 5.2 shows an example of a covariance matrix with the typical banded structure.

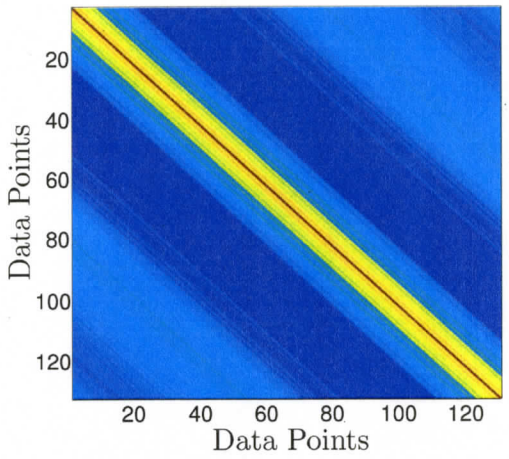

Fig. 5.2: Example of a covariance matrix for the case of uniform data spacing.

It is worth noting that the assumption of stationarity is not always justified. For certain data sets it is useful to take non-stationary errors into account as shown in Sec. 6.2.2.

The above procedure becomes somewhat more complicated for unevenly spaced data (e.g., reflection coefficient data as a function of seabed grazing angle β obtained for a source moving at constant velocity). Two approaches can be used to address this issue. First, the data can be resampled to be uniformly sampled. However, this can be disadvantageous in some cases (e.g., see Chapter 8) and a data covariance matrix for unevenly spaced data can be estimated as follows. First, the main diagonal of the covariance matrix (the variance) is computed as

$$C_{i;jj} = \frac{1}{N_i} \sum_{j=1}^{N_i} (n_{i;j} - \bar{n}_i)^2 \,, \tag{5.15}$$

where $n_{i;j}$ is the j^{th} residual at the i^{th} frequency. Then the product for every unique data pair $c_{i;j,k} = (n_{i;j} - \bar{n}_i)(n_{i;k} - \bar{n}_i)$ $(k \neq j,\ k > j)$ and their differences in angle $\Delta\theta_{i;j,k} = |\theta_{i;j} - \theta_{i;k}|$ are calculated and sorted according to increasing angle difference. So far, no assumptions were made about the stationarity of the data. The sorted $c_{i;l}$, where $l = 1 \ldots Q_i$, and $Q_i = (N_i(N_i + 1)/2) - N_i + 1$ is the number of pairs that equals the number of off-diagonal elements for the symmetric covariance matrix, are smoothed with a running cosine window that varies in width from 4° to 10° with

increasing $\Delta\theta$. This smoothing assumes stationarity of the data residuals over the length of the smoothing window. To ensure the covariance matrix estimate is positive definite, the smoothed autocovariance estimate is damped with an exponential factor

$$x_{i;l} = \exp\left(-\kappa_i \frac{l}{Q_i}\right) , \tag{5.16}$$

where κ_i is an empirical factor for each frequency, to obtain $\bar{c}_{i;l}$. In some cases, the matrix can also be diagonally loaded to ensure positive definiteness (Nocedal and Wright, 1999).

After smoothing and damping, the $\bar{c}_{i;l}$ are sorted back to the original order of the $\Delta\theta_{i;j,k}$ to obtain $\bar{c}_{i;j,k}$ that fill an initial symmetric data covariance matrix estimate $\hat{\mathbf{C}}^{(\mathbf{d})}_{(i,1)}$ according to $\hat{C}^{(\mathbf{d})}_{(i,1);jk} = \bar{c}_{i;j,k}$. This covariance matrix can then be applied in a subsequent inversion for a MAP estimate $\hat{\mathbf{m}}^{(2)}$. If the covariance matrix computed for $\hat{\mathbf{m}}^{(2)}$ using the above procedure is sufficiently close to $\hat{\mathbf{C}}^{(\mathbf{d})}_{(i,1)}$, the method can be considered to have converged. If not, further iterations can be performed. Measured data examples showed that the recovered estimate of the covariance matrix converged after the second iteration. The final estimate for the covariance can be incorporated into the FGS algorithm using Eq. (5.2). Figure 5.3 shows an example of the covariance matrix for nonuniform spacing for a measured data set. Figure 5.3 (a) shows that the covariance matrix is not banded; however, when the scale is stretched to the angle spacing of the data, Fig. 5.3 (b) shows an approximately banded structure is obtained (blocky structure results from data recording gaps, see Sec. 7.2).

5.4. Covariance Matrices for Non-stationary Residuals

As mentioned before, in some cases data residuals can be significantly non-stationary. This non-stationarity can be taken into account when estimating the data covariance matrix as follows. First, the data covariance matrix is estimated according to Sec.5.3. Second, to correct the non-stationarity, stationary residuals $\mathbf{n}$ were calculated from the data residuals $\mathbf{r}$ as described below.

The standard deviation for each reflector is calculated as a running RMS average over Q data

$$\sigma_i = \sqrt{\frac{1}{Q} \sum_{k=i-Q/2}^{i+Q/2} r_k^2} . \tag{5.17}$$

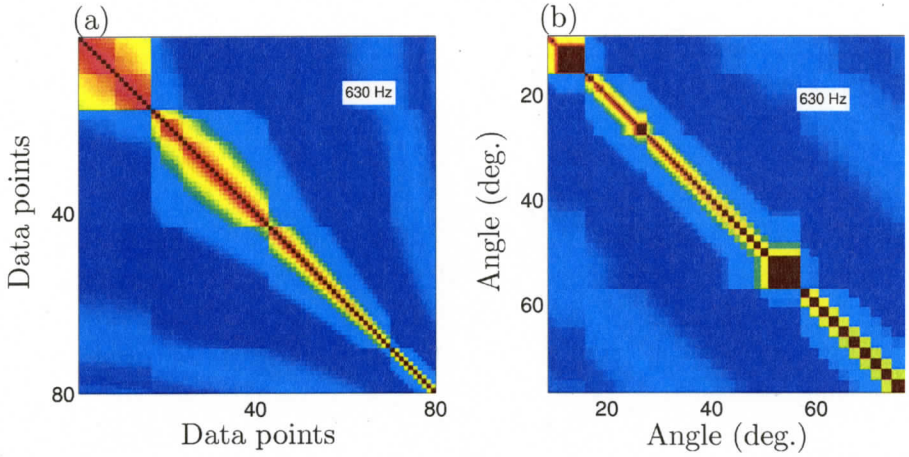

Fig. 5.3: Example of a covariance matrix for the case of nonuniform data spacing. (a) Estimated covariance matrix and (b) covariance matrix stretched to the nonuniform data spacing.

These standard deviations are then used to scale the data residuals so that

$$n_i = \frac{r_i}{\sigma_i} \,. \tag{5.18}$$

The n_i are then used to compute a Toeplitz data covariance matrix $\tilde{\mathbf{C}}^{\mathbf{d}}$ according to Eq. 5.14. To account for the non-stationary residuals, $\tilde{\mathbf{C}}^{\mathbf{d}}$ is scaled according to

$$C_{ij}^{\mathbf{d}} = \tilde{C}_{ij}\sigma_i\sigma_j \,, \tag{5.19}$$

resulting in a non-Toeplitz data covariance matrix $\mathbf{C}^{\mathbf{d}}$ that accounts for non-stationary data residuals (see Fig. 5.4).

5.5. Verifying Statistical Assumptions

Acoustic data always includes errors that are a combination of measurement errors (from the experiment) and theory errors (from approximations in the parametrisation and forward model). Formulating the likelihood function for Bayesian inversion requires making assumptions about the data error statistics (e.g., Gaussian or exponential distributed, with estimated covariance) since theses errors generally cannot be determined independently. Since the data misfit function for the inversion is based on the likelihood function, it is important for these assumptions to be reasonable to

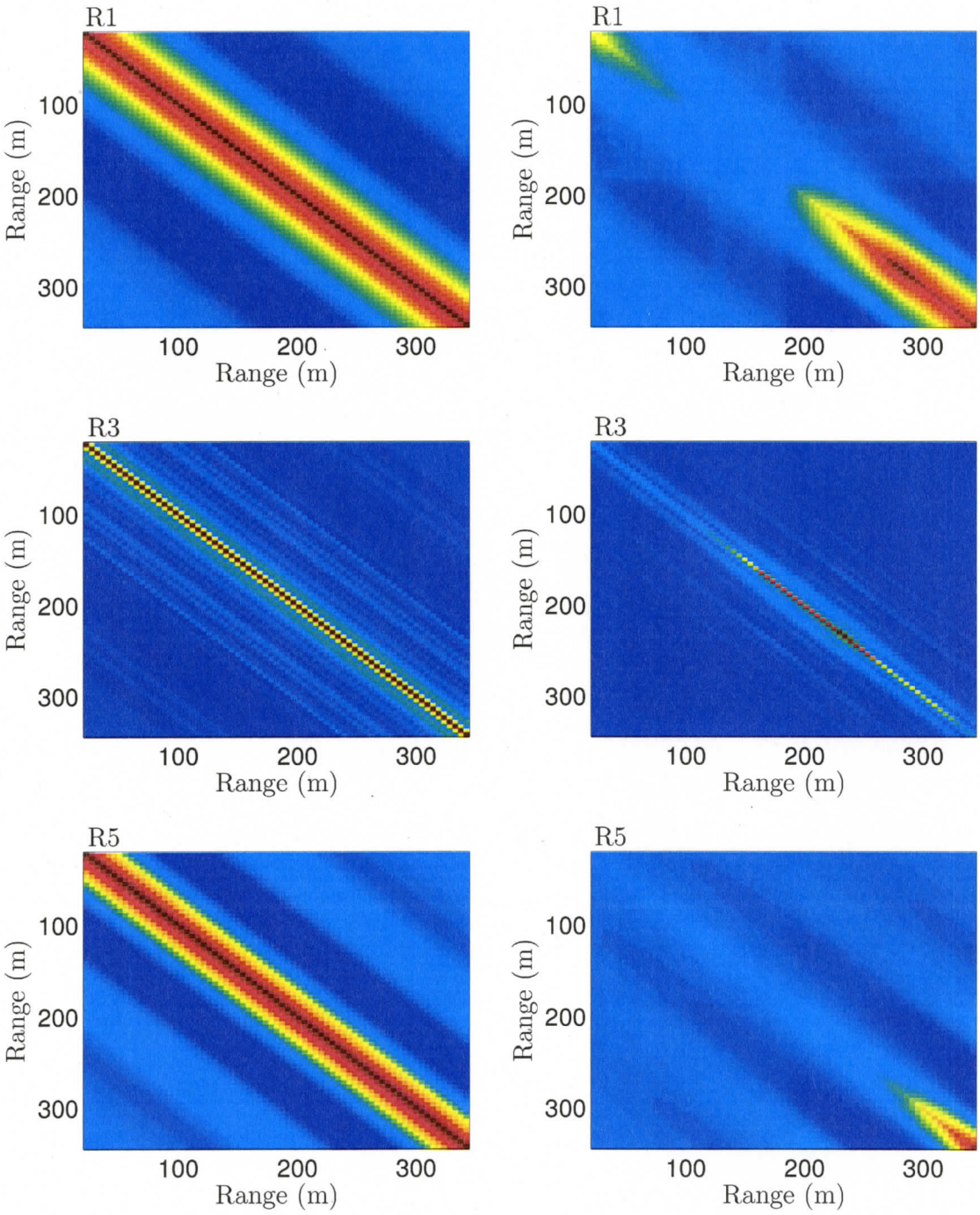

Fig. 5.4: Covariance matrix estimates for travel time inversion before (left column) and after correcting for non-stationarity (right column) for reflectors R1, R3 and R5.

produce meaningful results. To provide confidence in the solution of the inversion, the underlying statistical assumptions should be examined *a posteriori*.

The data residuals are considered a realisation of the data errors, and, hence, are required to satisfy the statistical error assumptions. The effect of taking covariances into account while inverting the data can be examined by standardising the data residuals at each frequency with the Cholesky decomposition $\mathbf{L}_i$ from Eq. (5.4) according to

$$\tilde{\mathbf{n}}_i = \mathbf{L}_i^{-1}\mathbf{n}_i \,. \tag{5.20}$$

To test whether the data residuals satisfy the initial assumptions, both qualitative (graphical) methods and quantitative statistical tests can be applied. The tests for standardised residuals can be compared to the same tests performed on raw residuals. Raw residuals are the residuals that are only scaled by their standard deviation.

Validating the assumptions essentially means testing the distribution form (e.g., Gaussian) and testing that the residuals are randomly distributed around zero. The form of the distribution can be qualitatively examined by looking at histograms of data residuals. A quantitative examination can be done with the KS test that is discussed below. Randomness of the mean removed data residuals is commonly examined by the runs test (discussed below).

5.5.1. Qualitative Tests

The autocorrelation function can be used to visualise the randomness of the residuals. Independent random errors should produce a narrow peak at zero lag. The wider the peak, the longer the correlation length of the errors. A qualitative test for the Gaussianity of the residuals can be done by plotting a histogram of the standardised residuals and comparing it to a theoretical Gaussian distribution $G(0, 1)$.

If there is strong evidence against the likelihood assumptions, the inversion should be repeated with different statistical assumptions that describe the fit of the best model to the data better, the model should be parametrised differently to decrease theory error, or physical assumptions in the forward model should be improved.

5.5.2. Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov (KS) test (Massey, 1951; Freund, 1967) is a quantitative statistical test for the hypothesis that a sample is drawn from a specific distribution by comparing the distribution of the sample to the specific distribution. This is referred to as a test of goodness of fit.

Let $X_1, \ldots, X_n$ be random variables forming a random sample with observed values $x_1, \ldots, x_n$. Let $F_n(x)$ be the sample distribution function such that $F_n(x)$ is the proportion of values in the sample with $x_i \leq x \;\forall\; i \in [1, n]$ and $x \in \mathbb{R}$. If the proportion is equal to k, $F_n(x) = k/n$. If $F(x)$ is the distribution function from which the random sample originated, $F_n(x)$ should converge to $F(x)$ for large n:

$$\lim_{n\to\infty} F_n(x) = F(x)\,. \tag{5.21}$$

This means that the maximum distance between $F_n(x)$ and $F(x)$

$$D_n = \sup |F_n(x) - F(x)| \tag{5.22}$$

will converge to zero for large n

$$\lim_{n\to\infty} D_n = 0\,. \tag{5.23}$$

This relation can be used to quantify how close a sample distribution function is to a theoretical distribution function $F(x)$. The hypotheses

$$\begin{aligned} &\mathrm{H}_0 : F_n(x) = F(x) \\ &\mathrm{H}_1 : \text{hypothesis } \mathrm{H}_0 \text{ is not true} \end{aligned} \tag{5.24}$$

can be tested by measuring D_n for a given sample of residuals and the distribution to be tested against (e.g., Gaussian or double-exponential). If D_n is small, H_0 is likely to be true. In this thesis, the p-value is used to quantify the rejection of the null hypothesis. The p-value is the probability of obtaining a result at least as extreme as that obtained, assuming the truth of the null hypothesis that the finding was the result of chance alone. Depending on the sample size n, tables with critical values for different levels of significance can be calculated. Generally, a level of significance of $p \geq 0.05$ is accepted as little or no evidence against H_0.

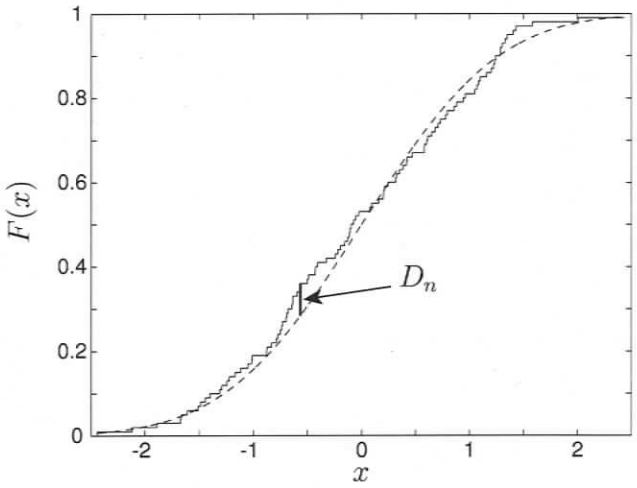

Fig. 5.5: KS test for a Gaussian distribution. The dashed line represents the theoretical distribution function $F(x)$ and the solid line the distribution function $F_n(x)$. The line D_n marks the maximum difference between both distribution functions.

Tables for critical values found in the literature are only valid for the case that the statistical parameters of a distribution are not estimated from the sample (Lilliefors, 1967). In the case that the statistical parameters (e.g., mean and variance) of the distribution have to estimated from the data, the usual critical values would present very conservative values. Lilliefors (1967) uses a Monte Carlo calculation to estimate the critical values for the case of estimating mean and variance from the sample and for different sample sizes n. For this work, the method was adapted by calculating critical values for fixed mean and variance estimated from the sample. The Monte Carlo calculation was performed for the appropriate sample sizes n and for 5000 realisations for both Gaussian and double-exponential distributed samples. These critical values are used throughout this work to calculate levels of significance.

5.5.3. Runs Test

The runs test (Freund, 1967) is a rigorous statistical test that quantifies randomness of a sample. Let $X_1, \ldots, X_n$ be a sample of variables with observed values $x_1, \ldots, x_n$, where the median has been removed from the sample. A *run* is defined as a succession of values with identical sign. For n_1 values with positive sign and n_2 values with negative sign, the total number of runs u is distributed with mean

$$\mu_u = \frac{2n_1n_2}{n_1 + n_2} + 1 \tag{5.25}$$

and variance

$$\sigma_u^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}\,. \tag{5.26}$$

The sampling distribution of u can be approximated with a normal distribution for $n_1 > 10$ and $n_2 > 10$, but for smaller sample sizes, the test has to be based on special tables (Owen, 1962). Hence, as long as the sample is large enough, the null hypothesis that the sample is random can be tested with

$$z_{\text{obs}} = \frac{u - \mu_u}{\sigma_u}\,. \tag{5.27}$$

The p-value is then given by

$$p = \frac{1}{\sqrt{2\pi}} \int_{|z_{\text{obs}}|}^{\infty} \exp(-x^2/2)dx\,. \tag{5.28}$$

6. Joint Time/Frequency Domain Inversion

6.1. Introduction

The spatial variability of seabed geoacoustic properties at fine scales and the uncertainties associated with geoacoustic inversion techniques are of key interest to understand shallow-water sound propagation and its applications. Standard matched-field and other long range-propagation methods lack fine-scale vertical and lateral resolution, have large theory error, and are often found to be insensitive to certain geoacoustic parameters such as density. To address these issues, single-bounce methods have been developed to more accurately resolve local sediment structure using plane wave approximations (Hastrup, 1970; Christensen et al., 1975; Chapman, 1983). Holland and Osler (2000) developed an experimental technique to measure high-resolution sea-bed reflectivity in shallow water. Because of the local scale, effects of spatial and temporal variability in the water column and seabed are greatly reduced compared to long-range acoustic measurements. Holland's experimental method collects seismo-acoustic data for a moving impulsive source (seismic boomer) and a seabottom moored hydrophone that can be interpreted in both the time and frequency domains. For impulsive sources, Stickler (1977) and Akal and Berkson (1987) showed that the plane wave approximation is often violated in deep water experiments. They proposed to take spherical wave effects into account when calculating reflection coefficients. Due to the geometry and the shallow water depth, the plane wave assumption may be violated in Hollands experiment in certain cases.

This chapter develops the methodology for a Bayesian joint time/frequency domain inversion, where the PPD resulting from a travel-time inversion is used as prior information for a frequency domain reflection coefficient inversion and examines certain important aspects of the inversion with synthetic studies. The inversion is then

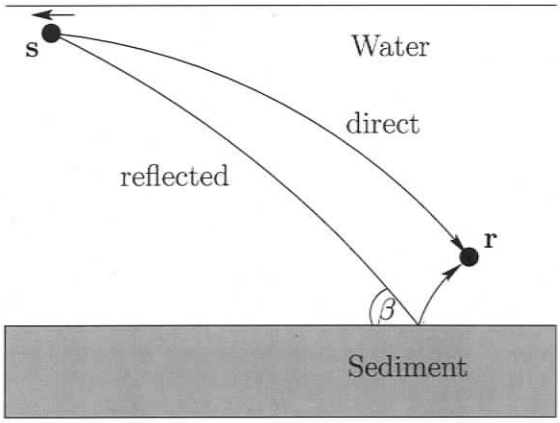

Fig. 6.1: Simplified experiment to measure single bounce seismo-acoustic data. The broad band source **s** is moving away from the single receiver hydrophone **r**, thus decreasing the grazing angle β.

applied to measured data in Chapter 7.

While the model parameters for the travel time inversion consist of layer thicknesses and sound velocities, the frequency domain inversion also includes densities and attenuation for each layer which are constrained by the reflection coefficients. For the frequency domain inversion, two forward models are developed to address both the plane wave approximation and full wave field effects, depending on the problem. This joint time/frequency domain approach utilises travel time and amplitude information but does not involve the extremely high computational costs of full waveform inversion. The sequential Bayesian formulation developed here applies the time domain information content within the frequency domain inversion.

The experiment (Fig. 6.1) consisted of a single, fixed hydrophone **r** and an impulsive broadband moving source **s**. The source is towed near the water surface at constant speed and transmits at uniform time intervals. Seismo-acoustic traces are recorded at a uniform range increment, as shown in Fig. 6.2. The time series show a direct arrival as well as numerous seabed reflections, including the water-sediment interface and deeper layers. At 0.130 s, the sediment-water interface reflection can be seen, followed by a number of sub-bottom reflections. The first water column multiple can be seen at approximately 0.270 s, which gives good separation between the main signal and the water multiple. The uniform spacing in range translates into a non-uniform spacing in grazing angle β, as shown in Fig. 6.3. Typical source-receiver offsets are from 20 to 1000 m. Gaps due to missing traces in Fig. 6.2 are caused by data being written to disk, which interrupts the recording. The ship usually moves towards the receiver, passes overhead, then moves away, resulting in two sets of recordings. The area where the rays bounce off the seabed defines the lateral footprint of the experiment. The

radius of the lateral footprint is typically of the order of 10^2 meters and the seabed is assumed to be laterally homogeneous over this area. Hence, the method averages over a seabed area in which no smaller lateral variations can be resolved. The method is also limited to areas with reasonably flat, horizontally stratified seabeds. To resolve spatial variability of seabed parameters, numerous recordings can be conducted with different hydrophone positions.

In the time domain inversion, travel time data are picked for multiple seabed reflectors and a ray tracing forward model is used to invert for layer thicknesses and sound velocities associated with the reflectors. For the frequency domain inversion, time domain data are windowed (see Sec. 6.3) into packets that contain the reflection effects of different sets of layers from the sea-floor to a sequence of increasing depths. The inverse problem is solved using a Bayesian layer-stripping approach that steps down through the packets, using the results for one packet inversion as prior information to constrain the subsequent packet inversions. The reflection coefficient $|V|$ is calculated for these various time windows by comparing the amplitudes of the direct and reflected paths. The value for $|V|$ is also corrected to eliminate effects of geometric spreading, refraction, absorption, and multi-path arrivals. A detailed description can be found in Sec. 6.3. An overview of the inversion methodology can be found in Fig. 6.4.

6.2. Single Bounce Travel Time Inversion

The seismo-acoustic travel time data (Fig. 6.2) include multiple sea-bottom reflections that are clearly defined, which suggest a high information content in the time domain. Holland and Osler (2000) used Bryan's method (Bryan, 1980, 1974) to estimate layer sound velocities and thicknesses from the seismo-acoustic data. To quantify the information content of the time domain traces, this chapter develops a fully non-linear travel time inversion that is based on picked travel times for multiple reflectors. This travel time inversion does not utilise the amplitude information of the acoustic data (in contrast to reflection coefficient inversion). Hence, the inversion results are not regarded as final environmental models, but rather prior information for a subsequent frequency domain inversion. The resulting PPD of the travel time inversion can be interpreted in terms of one dimensional marginal distributions, whose width can be used as prior knowledge to constrain certain model parameters in the final

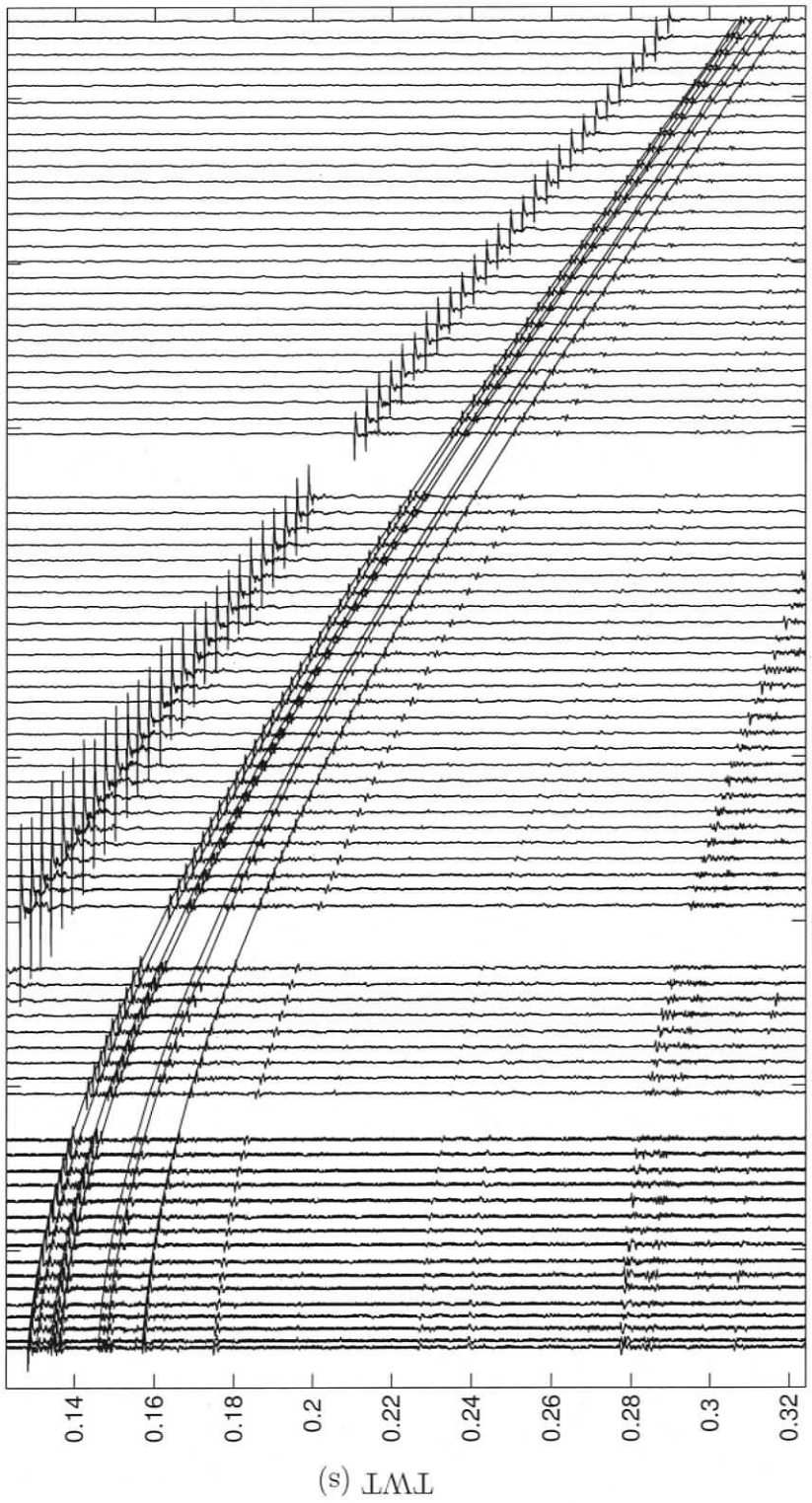

Fig. 6.2: Acoustic traces and picked travel times for six sediment layers in the Malta Plateau data. Missing traces are due to interruptions when data are written to disk.

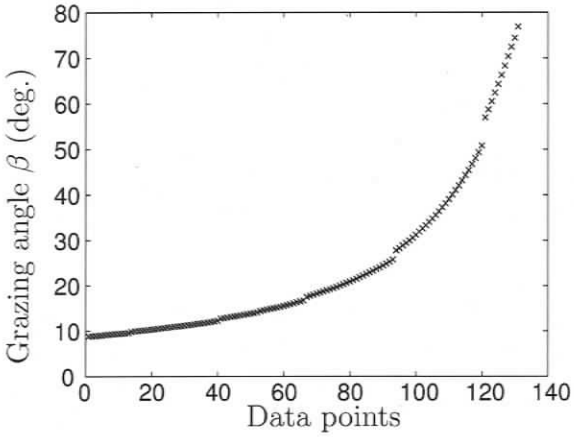

Fig. 6.3: Distribution of grazing angles β where data were recorded. The discontinuities are due to gaps in the data as shown in Fig. 6.2.

frequency domain inversion.

6.2.1. The Ray Forward Model

The forward model for the fully non-linear Bayesian approach uses ray tracing in horizontally stratified homogeneous layers, resulting in straight rays within layers with refraction at layer boundaries. Vertical incidence rays as well as turning rays are irrelevant for the experimental geometry and the expected environmental models and are neglected in the forward model. The ray tracing used here is based on the basic two point ray tracing problem of determining eigenrays that connect source and receiver (Aki and Richards, 1980; Shearer, 1999). The implementation using the method of images and Newton's method follows Dosso and Ebbeson (2005). For a brief summary of general asymptotic ray-series methods, see Appendix A.

Let a source point in a two dimensional, horizontally stratified fluid medium be given by $\mathbf{s} = (s_x, s_z)$. Let a receiver position be given by $\mathbf{r} = (r_x, r_z)$. The range R is then given by

$$R = \left((s_x - r_x)^2 + (s_z - r_z)^2\right)^{1/2} . \tag{6.1}$$

The travel-time t and range r of a non-turning ray follow from Snell's law and are given by

$$t = t_1 - t_0 = \int_{s_z}^{r_z} \frac{dz}{c(z)\left(1 - p^2c^2(z)\right)^{1/2}} , r = \int_{s_z}^{r_z} \frac{pc(z)dz}{\left(1 - p^2c^2(z)\right)^{1/2}} , \tag{6.2}$$

$$\tag{6.3}$$

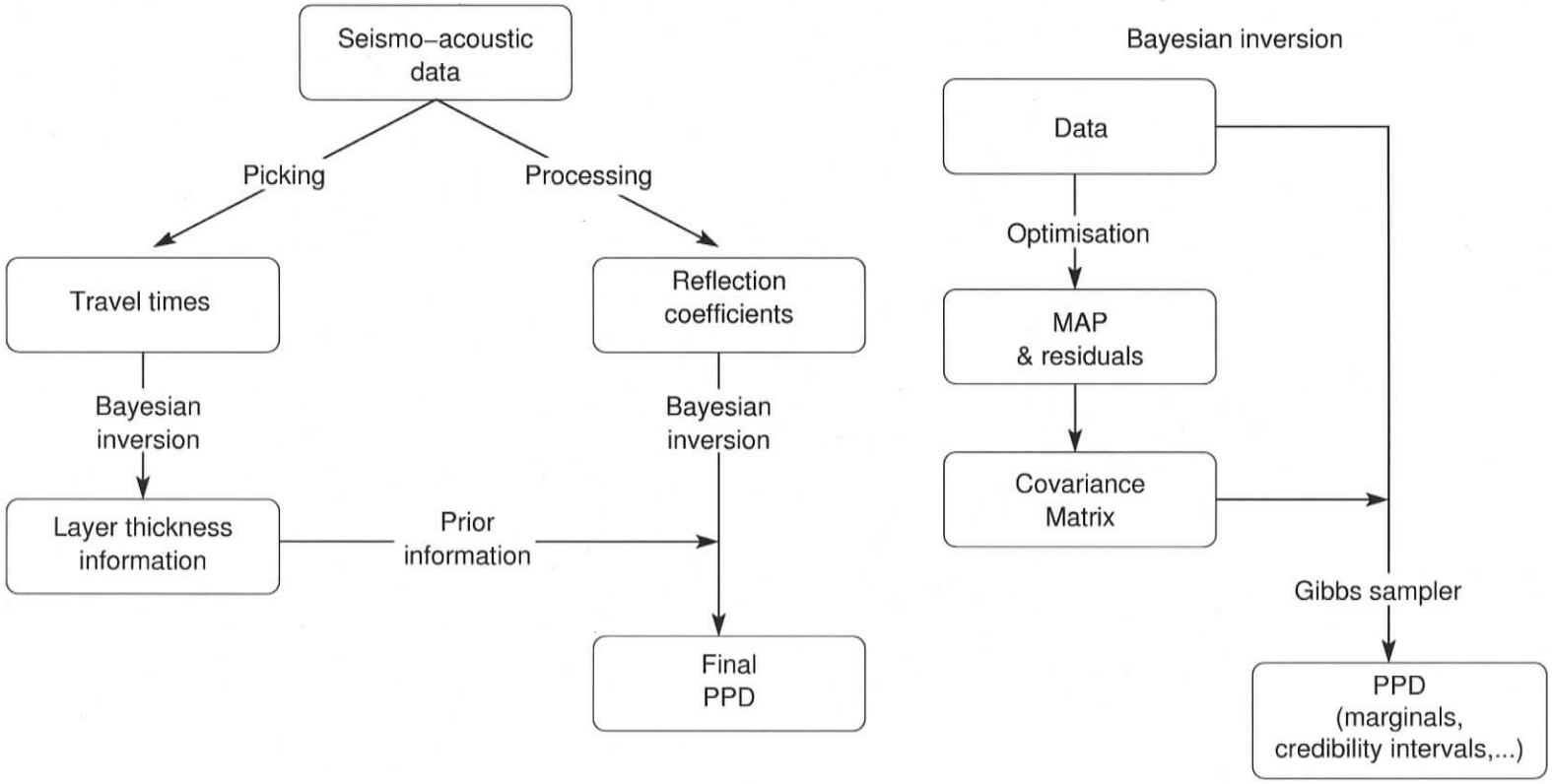

Fig. 6.4: Overview of the joint time/frequency domain inversion methodology.

where t_0 is the time of transmission of the source and t_1 is the arrival time at the receiver. The ray parameter or horizontal slowness $p = \cos(\beta(z))/c(z)$ is given in terms of grazing angle β as a function of depth and is constant along the ray (Snell's law). The two point ray tracing problem is to find the horizontal slowness that connects $\mathbf{r}$ and $\mathbf{s}$ for a given sound velocity profile $c(z)$ (the eigenray).

Hence, the search for the eigenray is a search over a range of take-off angles at the source. This search can be carried out efficiently with Newton's method. The search starts with an inital value p_0 that is based on a straight line between source and receiver where the layered medium is replaced by one layer with constant sound velocity c_H. The sound velocity c_H is chosen to be the harmonic mean of $c(z)$

$$c_H = (r_z - s_z) \Big/ \int_{s_z}^{r_z} \frac{dz}{c(z)} \,. \tag{6.4}$$

The initial value p_0 is then updated using Newton's method as follows. The range $r(p)$ is expanded in a Taylor series about p_0 up to the linear term

$$p_1 = p_0 + \left(\frac{\partial r(p_0)}{\partial p} \right)^{-1} (r(p) - r(p_0)) \,. \tag{6.5}$$

The derivative $\partial r/\partial p$ follows from Eq. 6.2

$$\frac{\partial r}{\partial p} = \int_{s_z}^{r_z} \frac{c(z)dz}{(1 - p^2 c^2(z))^{3/2}} \,. \tag{6.6}$$

Range r can then be computed and compared to the true range R; if the difference falls below a chosen convergence level, the method converges, otherwise p_1 replaces p_0 and another iteration is performed. Newton's method converges quadratically and generally only a few rays need (less than ten) to be computed to obtain convergence to a high level of precision (typically 1 cm in range). Once convergence is obtained, the travel time for the ray follows from Eq. 6.2.

The algorithm also uses the method of images (Brekhovskikh and Lysanov, 1991) to treat reflected rays as direct rays through an augmented sound velocity profile. Therefore, this simple ray tracing algorithm satisfies all requirements for the travel-time inversion in this section. The extension of the algorithm to layers with constant sound velocity gradient is straightforward.

6.2.2. Picking Travel Times

The inversion is based on picked travel time data from the seismo-acoustic traces as indicated in Fig. 6.2. For all reflectors, the first break was chosen for the picking. The picking is based on a matched filter

$$t_{\text{pick}} = \text{Arg}_{\text{max}} \left(A_{xy} \right) , \tag{6.7}$$

where A_{xy} is the cross correlation function of vectors x and y. Vector x consists of the acoustic trace for a certain offset and vector y represents the fixed-length wavelet that is matched. Initially, y is taken to be the arrival wavelet starting at a hand picked point on the first trace. The wavelet is automatically updated while proceeding through the traces (i.e., the wavelet beginning at the arrival time picked for trace i is used to correlate with trace $i + 1$). The matched filter picks are then edited by hand to ensure stable and consistent picks. The picking can be done iteratively by comparing replica data for a MAP model to the seismic traces and refining the picks in areas where the replica seems to fit better than the initial pick. Since the goal of this inversion is to quantify information content, the focus here is on treating the errors in picking rigorously rather than using the most sophisticated picking scheme. In particular, subjective criteria, such as picking arrivals along perfect hyperbolae are avoided. This assures that the uncertainty estimates are meaningful and actually relate to the true travel times (not an interpreted travel time model with unquantifiable theory error).

There are several challenges involved with picking travel times. First, the actual start of the signal is not always obvious due to overlap of different arrivals. This occurs at long offsets and due to complicated seabed structure, where waveforms of different events are superimposed. In general, it is easier to correlate events from trace to trace and consistently match the same signal feature across offset than it is to definitely pick the first break. This often results in a bias (offset) of the arrival time across all traces. Since even small errors in picking can have a significant effect on the recovered model, strong biases in the PPD can result from these picking errors. To address this in the inversion, additional nuisance parameters can be included to account for an unknown picking offset for each reflector. The data misfit for interface i, assuming unknown standard deviation of data errors, can then be written (see

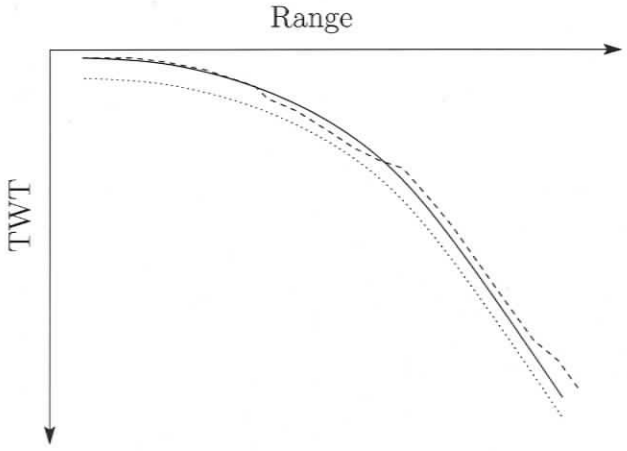

Fig. 6.5: Constant bias and correlated error in picking. The dotted line represents the true first break and the solid line a feature that appears to be the first break. The dashed line illustrates the type of picks that would likely be made by an interpreter. The picks show biased, strongly correlated errors.

Chapter 5) as

$$E_i(\mathbf{m}) = \frac{N_i}{2} \log_e \left(\sum_{j=1}^{N_i} (t_{i;j} + a_i - t_{i;j}(\mathbf{m}))^2 \right) , \tag{6.8}$$

where a_i is a unknown (constant) time offset parameter. Assuming independent errors from interface to interface, the total misfit is the sum over the misfits of the individual interfaces.

The picking of travel times typically becomes more difficult and less accurate with increasing offset, since for large offsets the arrivals of different interfaces are closer together and overlap more. Also, the measured data (Fig. 6.2) have missing traces (gaps) where data were written to disk. Correlating arrivals across these gaps is more difficult than for traces that are closer together. Hence, it is possible to consistently pick travel times for a number of traces and then lose a particular feature and begin picking on another, similar, feature, as illustrated in Fig. 6.5. The results are strongly correlated data errors which must be taken into account by estimating covariance matrices from the data residuals for each reflector. Further, as arrivals run into each other at large offsets, it becomes more difficult to identify the correct feature, resulting in increased picking uncertainty with offset. Note that the severity of these issues are different for different interfaces, and depend on the layer thickness. The thicker a layer is, the less likely the corresponding reflected arrivals are to overlap.

To address the above issues, data covariance matrix estimates were calculated according to Sec. 5.3. However, the increasing uncertainty with offset in picking causes the data residuals $\mathbf{r}$ to be non-stationary across range. To account for this corrected, data covariance matrices are estimated according to Sec. 5.4.

Accounting for both the covariance matrix estimate and the picking-offset nuisance parameters, the L_2-norm can then be written as

$$E(\mathbf{m}) = \sum_{i=1}^{L} \frac{1}{2} \left(\mathbf{t}_i + a_i - \mathbf{t}_i(\hat{\mathbf{m}})\right)^{\mathsf{T}} \left(\mathbf{C}_i^{(\mathbf{d})}\right)^{-1} \left(\mathbf{t}_i + a_i - \mathbf{t}_i(\hat{\mathbf{m}})\right) , \tag{6.9}$$

where L is the number of reflectors, $\mathbf{t}_i$ are vectors containing the travel times for the i^{th} reflector, and $\mathbf{C}_i^{(\mathbf{d})}$ are estimated data covariance matrices.

6.2.3. Data Simulated with Ray Tracing

To investigate the best possible resolution of the travel time inversion problem, where no biased and correlated picking errors exist, a synthetic travel time data set was generated by ray tracing and zero mean Gaussian errors were added. The true model consists of 6 sediment layers (i.e., 7 reflectors) and a constant sound velocity water column of 150 m depth. The source was placed at 0.35 m depth below the water surface and the receiver at 122 m depth. Numerical values for the true model parameters and for the MAP model are given in Table 6.1.

Layer	Thickness (m)			Sound velocity (m/s)		
	true	MAP	95% HPD	true	MAP	95% HPD
1	1.5	1.50	1.44–1.54	1540	1545	1500–1563
2	3.0	2.99	2.92–3.06	1520	1516	1500–1580
3	1.0	1.02	0.95–1.09	1560	1602	1507–1648
4	7.0	7.01	6.91–7.07	1600	1559	1501–1646
5	1.0	0.95	0.89–1.03	1700	1658	1585–1735
6	4.0	4.01	3.88–4.07	1600	1644	1553–1698

Table 6.1: Numerical values for true parameters and inversion results (MAP model and 95% HPD credibility intervals) for data simulated by ray tracing and no picking offsets.

The inversion results for unbiased data errors are compared to results for similar data but with realistic constant time offsets (biases) added to each reflector hyperbola. In the latter case, the picking offsets were taken into account by including seven unknown offset parameters a_i in the inversion. Figure 6.6 shows the resulting marginal distributions for both data sets. It can be seen that constant time offsets (which are easily introduced while picking) significantly widen the marginal distributions. While layer thicknesses are still resolved reasonably well, velocity resolution

is lost. Figure 6.7 shows the marginal distributions for the offset unknowns. The prior bounds for the last offset parameter a_7 were chosen to be narrower than the rest to simulate higher confidence in the travel time picks, which is a common feature for some reflectors in measured data. All offsets could be resolved reasonably well and the distributions agree with the true values. The numerical values for the MAP model and 95% HPD credibility intervals are given in Tables 6.1 and 6.2.

Layer	Thickness (m)			Sound velocity (m/s)		
	true	MAP	95% HPD	true	MAP	95% HPD
1	1.5	1.51	1.39–1.66	1540	1527	1500–1578
2	3.0	2.98	2.68–3.10	1520	1534	1500–1597
3	1.0	1.01	0.84–1.17	1560	1618	1503–1648
4	7.0	6.85	6.62–7.31	1600	1577	1501–1648
5	1.0	1.00	0.76–1.15	1700	1565	1550–1735
6	4.0	4.08	3.67–4.44	1600	1690	1550–1697

Table 6.2: Numerical values for true parameters and inversion results (MAP model and 95% HPD credibility intervals) for data simulated by ray tracing with added picking offsets.

The above examples illustrate the effect that picking biases have on the ability to recover physical information of the seabed structure. Not taking offsets into account can result in unreasonably tight credibility intervals that misrepresent the actual information content of the picked data. Further, as shown in Fig. 6.8, the geoacoustic parameter estimates can be strongly biased as a result of misrepresenting the data by not including offset nuisance parameters, and credibility intervals can be misleading.

6.2.4. Data Picked from Synthetic Seismo-Acoustic Traces

To consider more realistic correlated picking errors, the model in Table 6.2 was used to compute acoustic traces using the OASES program package. The resulting transfer functions (see Appendix B) were convolved with the source wavelet shown in Fig. 6.9. The wavelet was taken from experimental data by averaging four traces recorded at small offsets and then low pass filtered with a 3000 Hz cut-off to obtain realistic, simulated acoustic traces (Fig. 6.10). This results in a simulated data set including the limited resolution of a wavelet of finite length, causing realistic picking errors. Travel times were picked from these traces as outlined in Sec. 6.2.2 and used in the

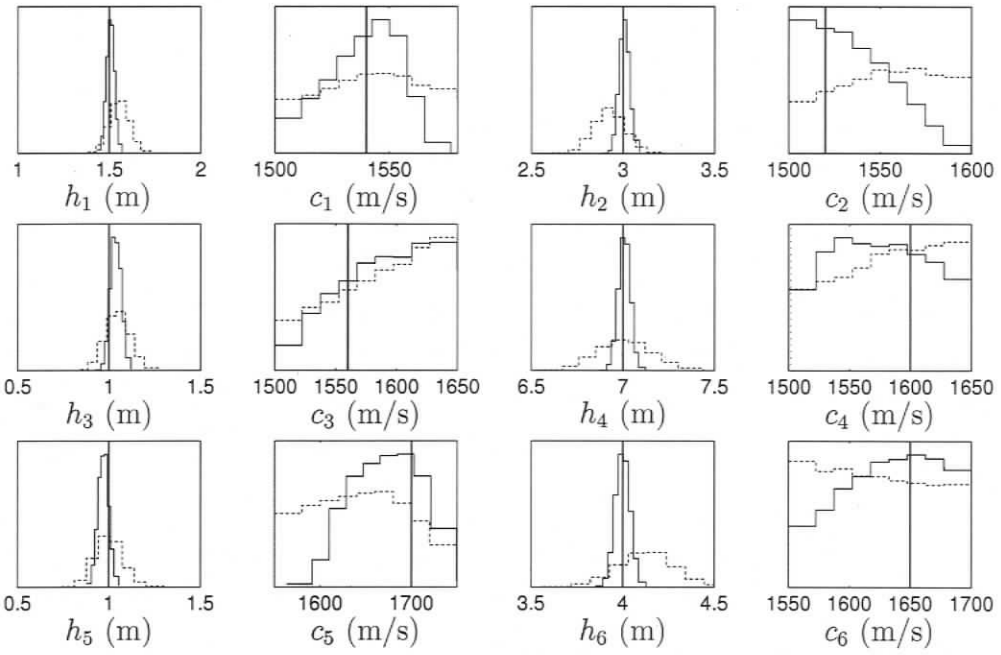

Fig. 6.6: Marginal distributions for data simulated by ray tracing and with additional travel time offsets for the different reflectors (dashed). True values as solid black lines.

inversion. Figure 6.10 also shows the true travel times as computed with ray tracing. For several reflectors, the picked travel times show an offset or bias as discussed before.

The picking bias is illustrated more clearly by plotting the differences between the true and picked data, which represents the picking error, as shown in Fig. 6.11. Note that the errors are strongly biased for some reflectors and show strong serial correlations (i.e., are not randomly distributed about zero). The resulting MAP estimates for the time offset parameters for each reflector (included in the inversion) fit the mean picking error reasonably well for most reflectors and are also shown in Fig. 6.11. The error correlations and biases are significantly reduced by including time offsets in the inversion. Although the time offsets do not account for all correlations, the error appears to be accounted for to first order. It can be seen that including the offsets provides a substantial advantage as it represents the median of the true errors reasonably well in most cases.

Of course, the picking errors are not available for measured data since the true (error-free) data are not known. However, the data residuals (between MAP replica

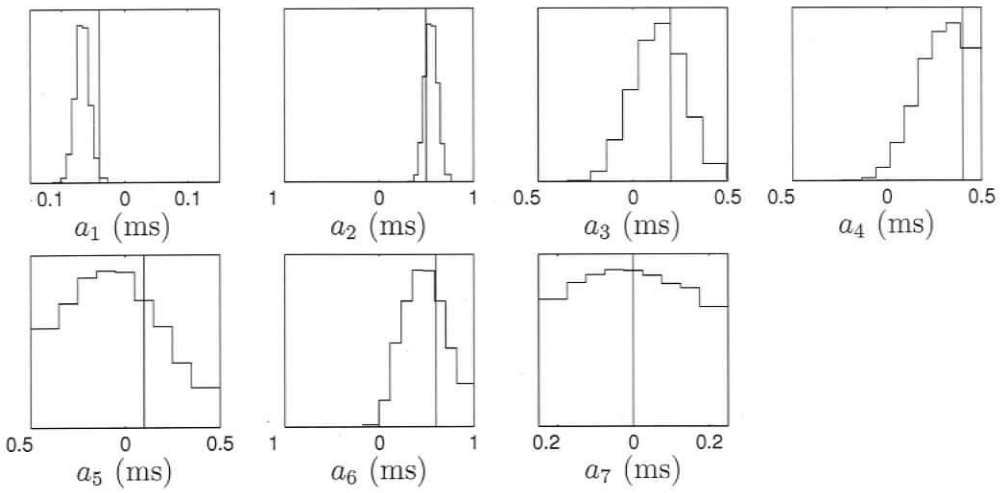

Fig. 6.7: Marginal distributions of the recovered picking offsets that were added to the simulated ray-tracing data. True values are given as solid black lines.

and picked data) are available and are shown in Fig. 6.12. It should be noted that data biases can be difficult to detect in measured data. The misfit of an inversion accounting for biases can be very similar to the misfit of an inversion neglecting biases. However, the biased data will cause misleading inversion results. Therefore, in this case it is important to closely examine data residuals and inversion results in realistic synthetic studies where the simulated data are computed using a different forward model that is independent from the inversion. Possible data biases can then be detected and corrected for in the misfit function of the forward model. It is important to realise that the effect of including offset parameters may not be understood by only examining the residuals and validating the statistical assumptions, as the data may be fit well. Neglecting offsets results in a substantial theory error that biases the resulting PPD as shown in Sec. 6.2.3. To quantify the effect of including offsets, two different misfits can be calculated for the j^{th} reflector

$$E_i = \sum_j \left(\frac{t_{i;j}^{(pick)} - t_{i;j}^{(true)}}{\sigma_{i;j}} \right)^2 \quad \text{and} \quad \bar{E}_i = \sum_j \left(\frac{t_{i;j}^{(pick)} - t_{i;j}^{(true)} - a_j}{\sigma_{i;j}} \right)^2, \quad (6.10)$$

where σ_i is estimated as described in Sec. 5.4. Here, $\bar{E}$ is the misfit accounting for constant time offsets and E is the misfit neglecting the offsets. Further, it is useful

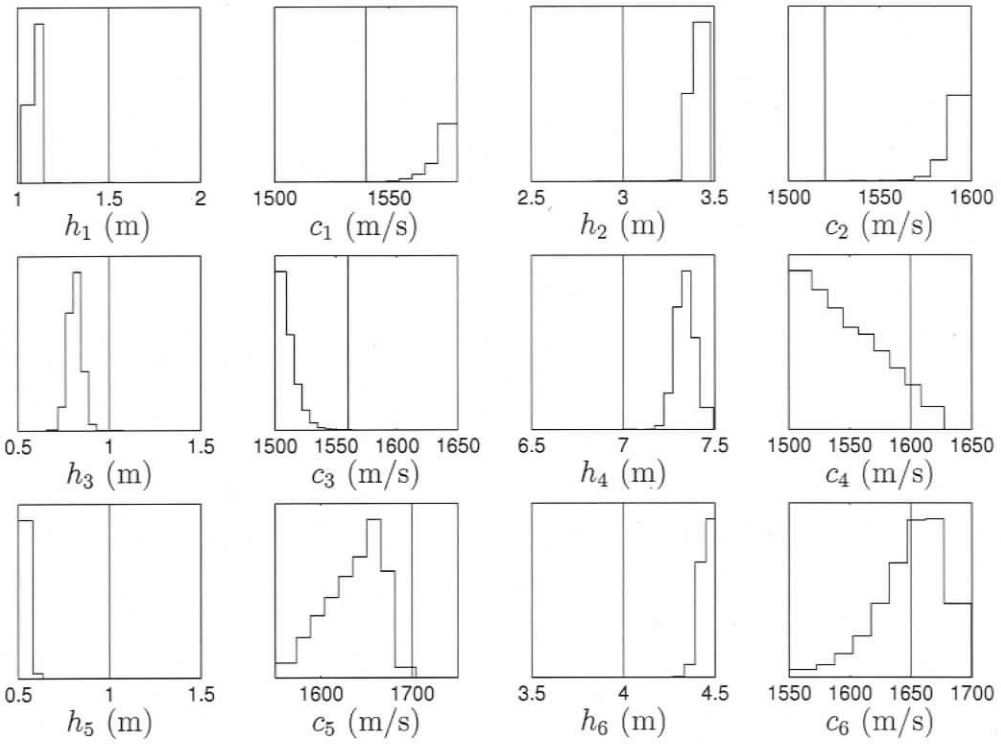

Fig. 6.8: Marginal distributions for data simulated by ray tracing with additional travel time offsets for the different reflectors that were neglected in the inversion. True values as solid black lines. Strong biases in the parameter estimates and misleading credibility intervals result from neglecting picking offsets.

to look at the arithmetic mean values which represent the mean offset

$$\bar{t}_{1i} = \text{MEAN}\left(\mathbf{t}_i^{(pick)} - \mathbf{t}_i^{(true)}\right) \qquad \text{and} \qquad \bar{t}_{2i} = \text{MEAN}\left(\mathbf{t}_i^{(pick)} - \mathbf{t}_i^{(true)} - a_i\right), \tag{6.11}$$

where the travel time vectors represent the travel times across range for the j^{th} reflector. Similarly to E_i and $\bar{E}_i$, $\bar{t}_{1i}$ and $\bar{t}_{2i}$ give a measure of how well the travel time data are fit. The closer $|\bar{t}_{lj}|$ is to zero, the better the fit. The average values of the absolute values of all reflectors are $E = 68.336$, $\bar{E} = 30.207$, $\bar{t}_1 = 0.1798$ ms, and $\bar{t}_2 = 0.0793$ ms. Table 6.3 gives numerical values for these misfits and mean values for the individual reflectors. With the exception of R1, the results for all reflectors improve significantly.

Even with constant offset parameters included in the inversion, the data residuals (Fig. 6.12) indicate strongly correlated data errors. Further, the residuals are non-stationary, in that the magnitude of the standard deviation changes significantly

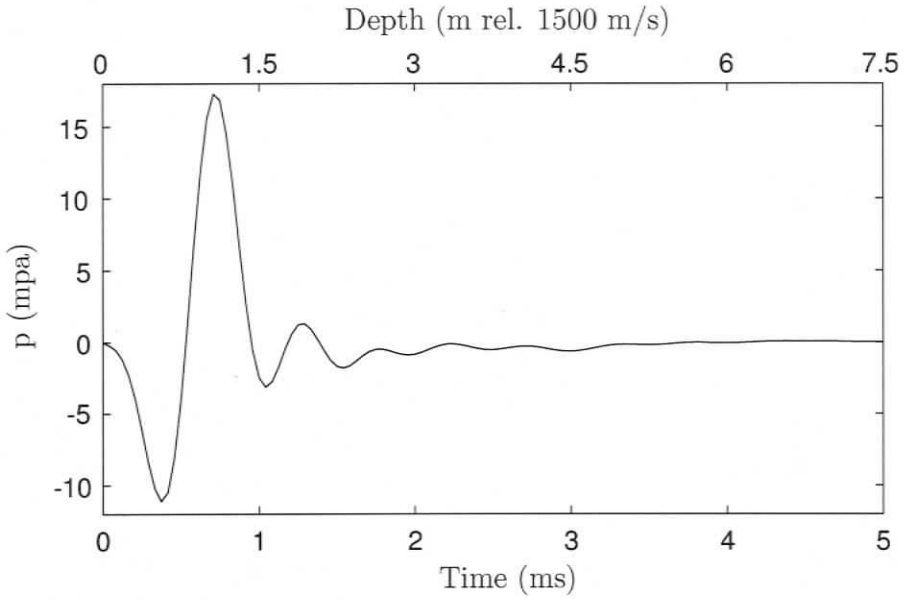

Fig. 6.9: Source wavelet for the simulated experiment.

Parameter	R_1	R_2	R_3	R_4	R_5	R_6	R_7
E	70.513	66.687	70.296	66.774	70.241	69.059	64.784
$\bar{E}$	93.156	16.997	31.354	21.188	7.516	9.107	32.132
$\bar{t}_1$ (ms)	-0.0147	-0.3034	-0.1418	-0.2478	0.2509	-0.1897	0.0018
$\bar{t}_2$ (ms)	-0.0536	-0.1472	0.0828	-0.1325	-0.0461	-0.0112	0.0307

Table 6.3: Misfit and mean residuals for reflectors R_1 to R_7 both, neglecting offsets (E, $\bar{t}_1$) and taking offsets into account ($\bar{E}$, $\bar{t}_2$)

across range, as illustrated by a running RMS average (Fig. 6.13). To meaningfully estimate credibility intervals under the assumption of Gaussian distributed random errors, these data covariances must be estimated and included in the inversion. To do so, Toeplitz covariance matrices are estimated from the data residuals and then scaled according to Eq. 5.19 to account for the non-stationary data residuals, resulting in non-Toeplitz covariance matrices (Fig. 6.14).

After applying Gibbs sampling inversion using the covariance matrix estimates (prior bounds given in Table 6.4), the information recovered from the PPD is given as marginal distributions for the physical parameters in Fig. 6.15 and for the nuisance parameters in Fig. 6.16. Numerical values for the MAP parameter estimates and the

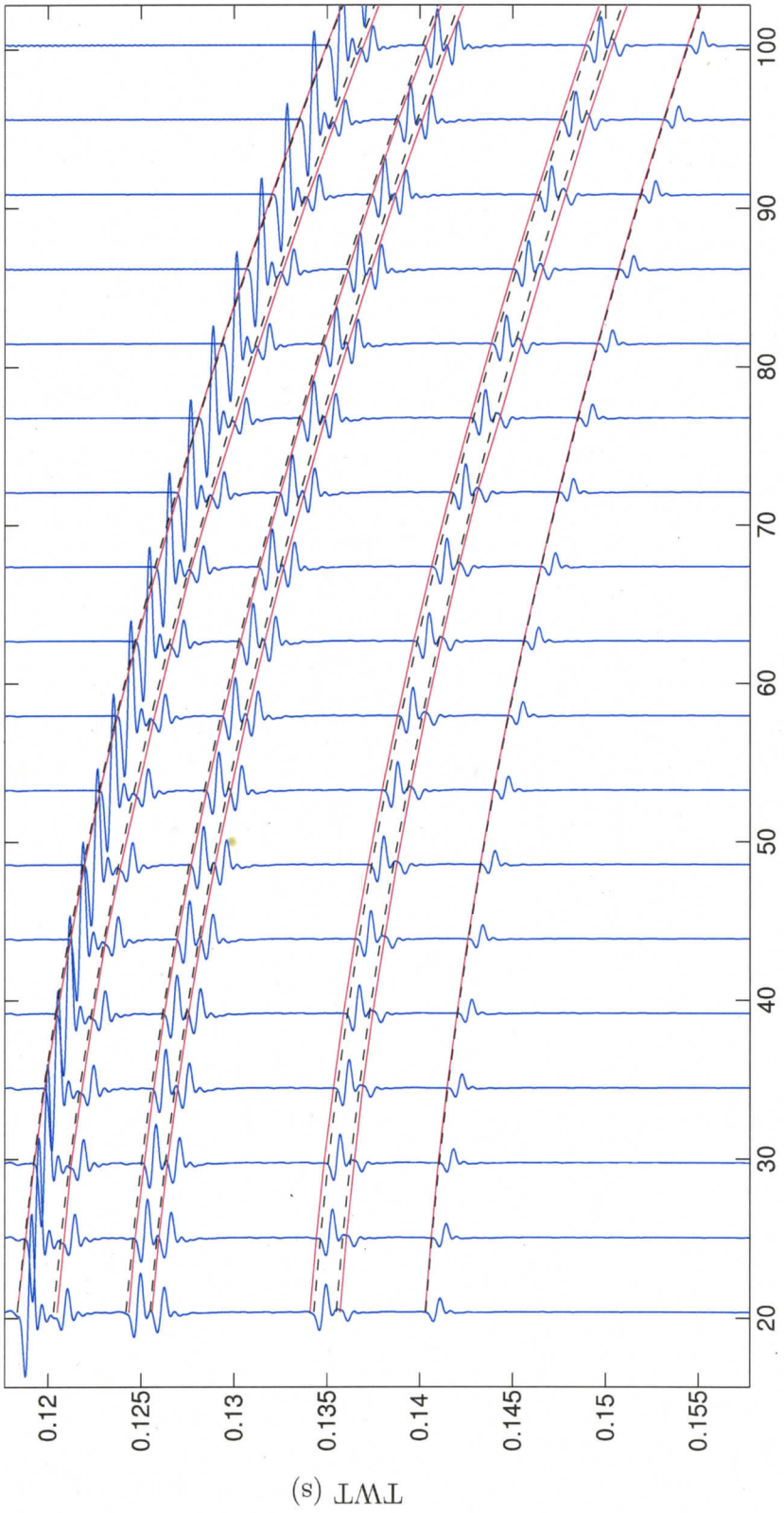

Fig. 6.10: Simulated acoustic traces and picked travel times for six sediment layers. The dashed lines indicate the hyperbola for the true model as computed by ray tracing. Note that in some cases there is an offset between the picked travel times and true solution. Only a crop of ranges from 20 to 100 m is shown for clarity.

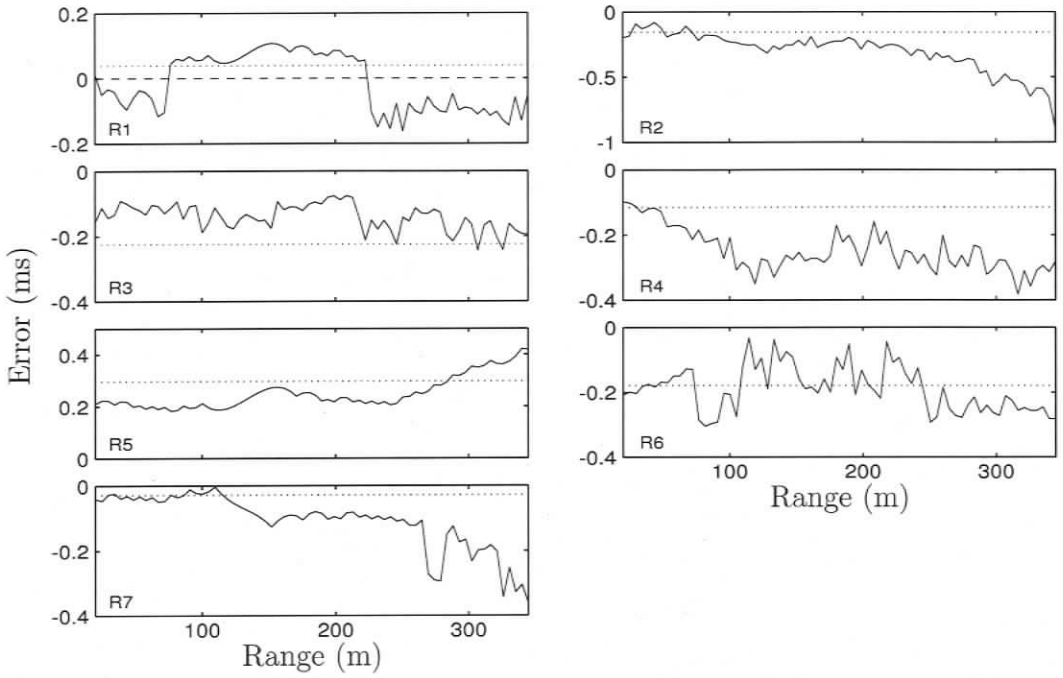

Fig. 6.11: Picking error for the seven reflectors. Dotted lines indicate the time offsets recovered by the inversion.

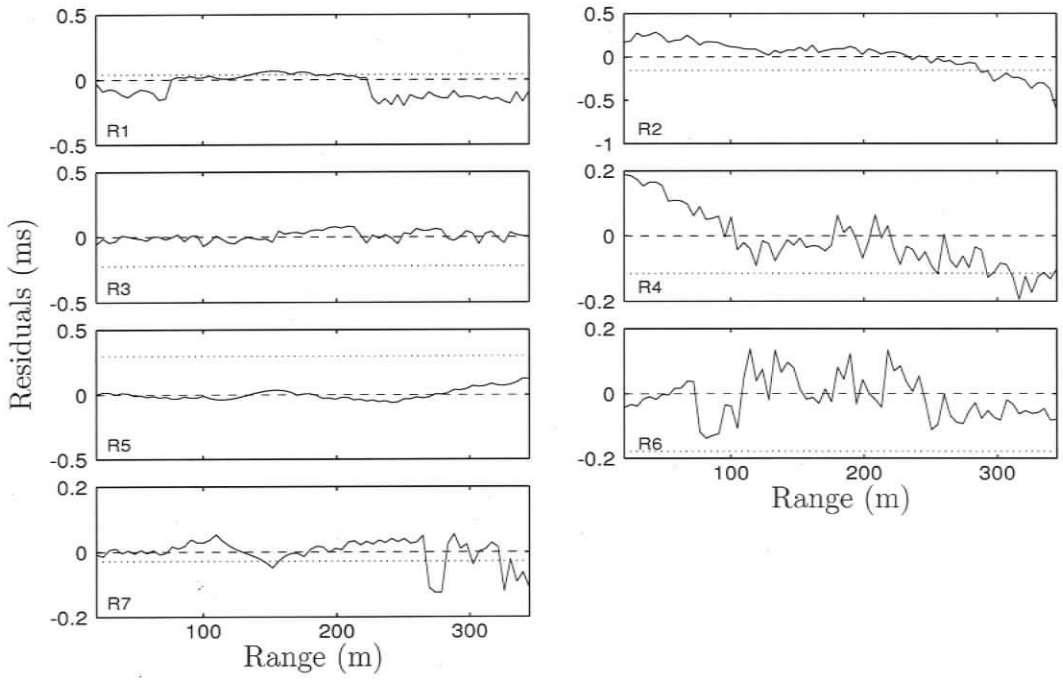

Fig. 6.12: The data residuals for picked and MAP replica data for 7 reflectors. The dotted line indicates the recovered time offset parameters.

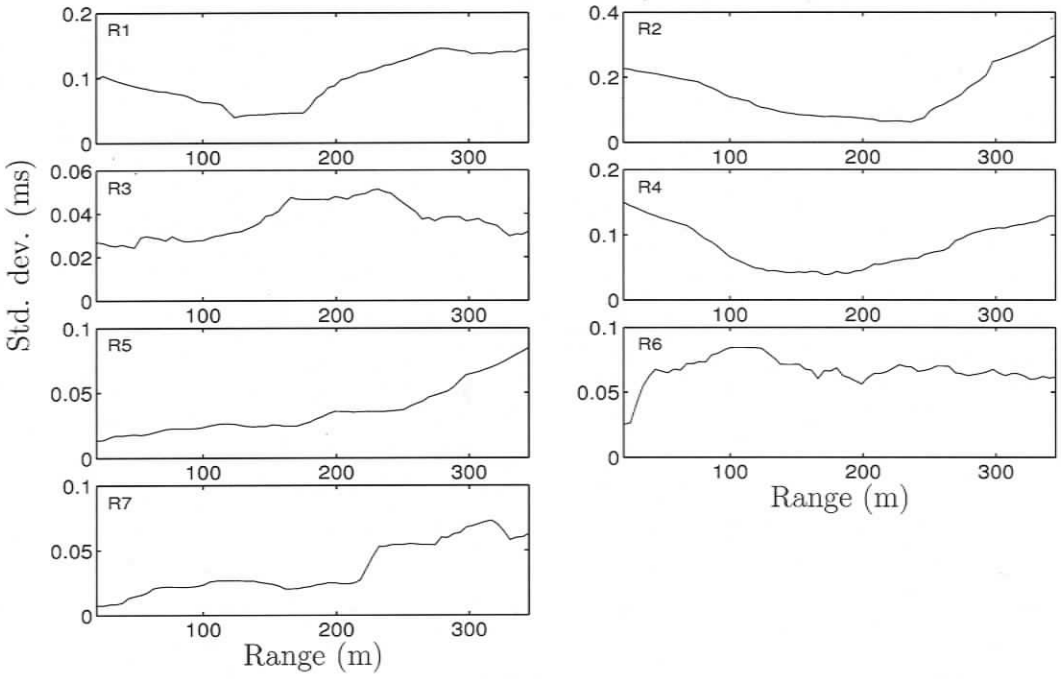

Fig. 6.13: Standard deviation estimates averaged with 20 point running RMS filter. Significant changes in standard deviation across range indicate non-stationarity of the residuals.

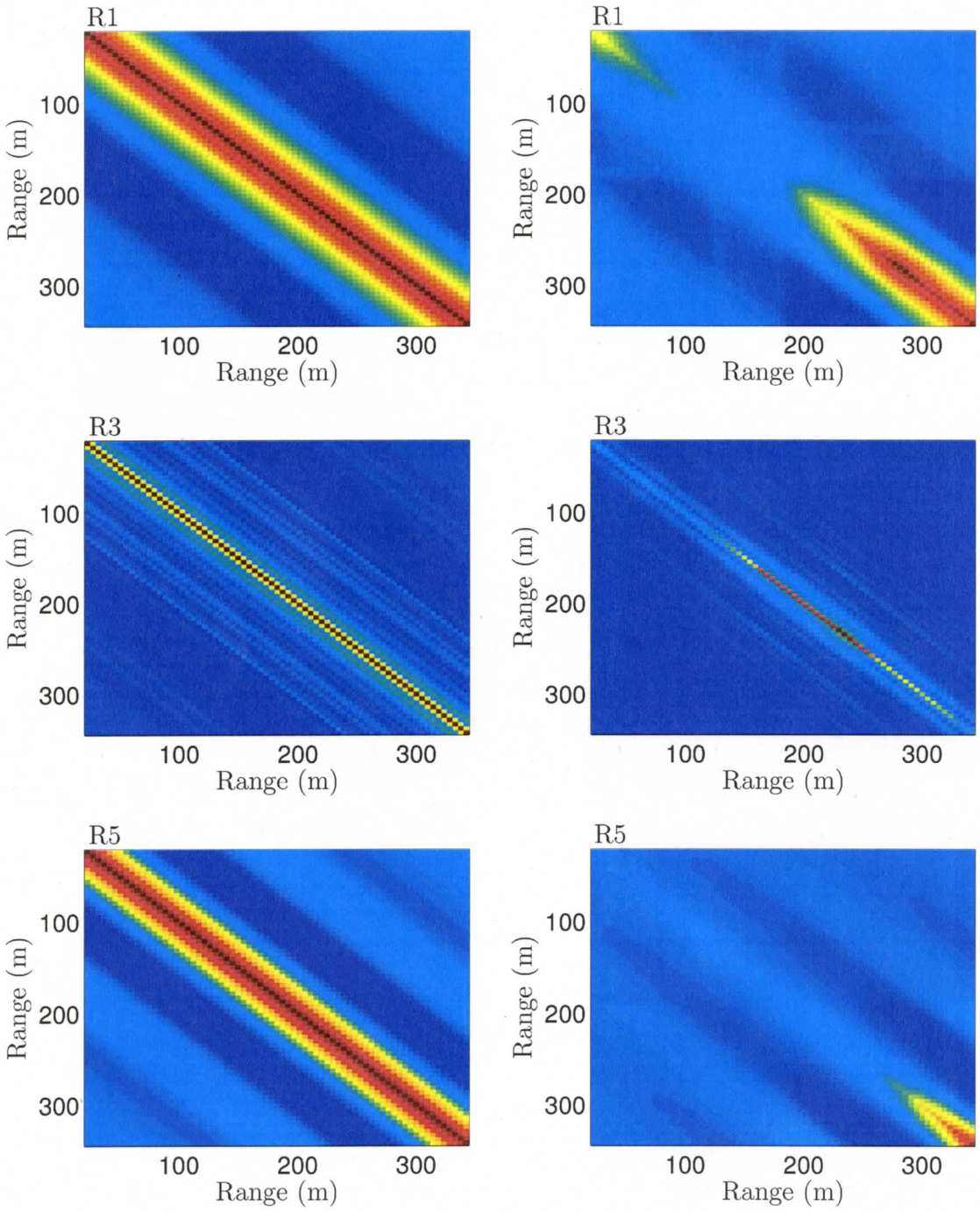

Fig. 6.14: Covariance matrix estimates for travel time inversion before (left column) and after correcting for non-stationarity (right column) for reflectors R1, R3 and R5.

Layer	h (m)			c (m/s)		
	Prior	MAP	95% HPD	Prior	MAP	95% HPD
1	0.10–8.00	1.79	1.63–1.97	1450–1600	1530	1457–1595
2	0.10–8.00	2.80	2.51–3.05	1500–1600	1508	1500–1596
3	0.10–8.00	1.66	1.45–1.87	1500–1700	1581	1537–1662
4	0.10–8.00	7.65	7.56–7.67	1500–1700	1637	1537–1679
5	0.10–8.00	1.61	1.47–1.82	1500–1650	1556	1500–1628
6	0.10–8.00	7.06	6.68–7.28	1500–1700	1569	1500–1664

Table 6.4: Numerical values of prior information and inversion results (MAP model and 95% HPD credibility intervals) for picked travel time data.

credibility intervals for the 12 physical parameters are given in Table 6.5. It can be seen that layer thicknesses are well resolved within their prior bounds. Since the prior bounds for the sound velocities were already closely constrained relative to their numerical values, the PPD does not provide substantial new information about the sound velocities. The total width of the marginal distributions is given in Table 6.6. The 100% credibility intervals for the layer thicknesses will be used as prior information in the frequency domain inversion in the following section. For sound speeds, the prior bounds applied in the frequency domain inversion are the same as in the time domain inversion. Constraining layer thicknesses in the reflection coefficient inversion can be very useful as layer thicknesses often have multiple solutions due to interference effects. The travel time inversion thus complements the reflection coefficient inversion well.

Layer	Thickness (m)			Sound velocity (m/s)		
	true	MAP	95% HPD	true	MAP	95% HPD
1	1.5	1.34	1.24–1.73	1540	1501	1500–1558
2	3.0	3.19	2.67–3.24	1520	1516	1500–1575
3	1.0	1.10	1.00–1.39	1560	1613	1542–1648
4	7.0	6.84	6.74–7.01	1600	1608	1516–1645
5	1.0	0.85	0.79–1.04	1700	1614	1550–1684
6	4.0	4.19	3.90–4.19	1600	1569	1554–1684

Table 6.5: Numerical values for true parameters and inversion results (MAP model and 95% HPD credibility intervals) for picked travel time data. Picking was based on simulated seismo-acoustic traces.

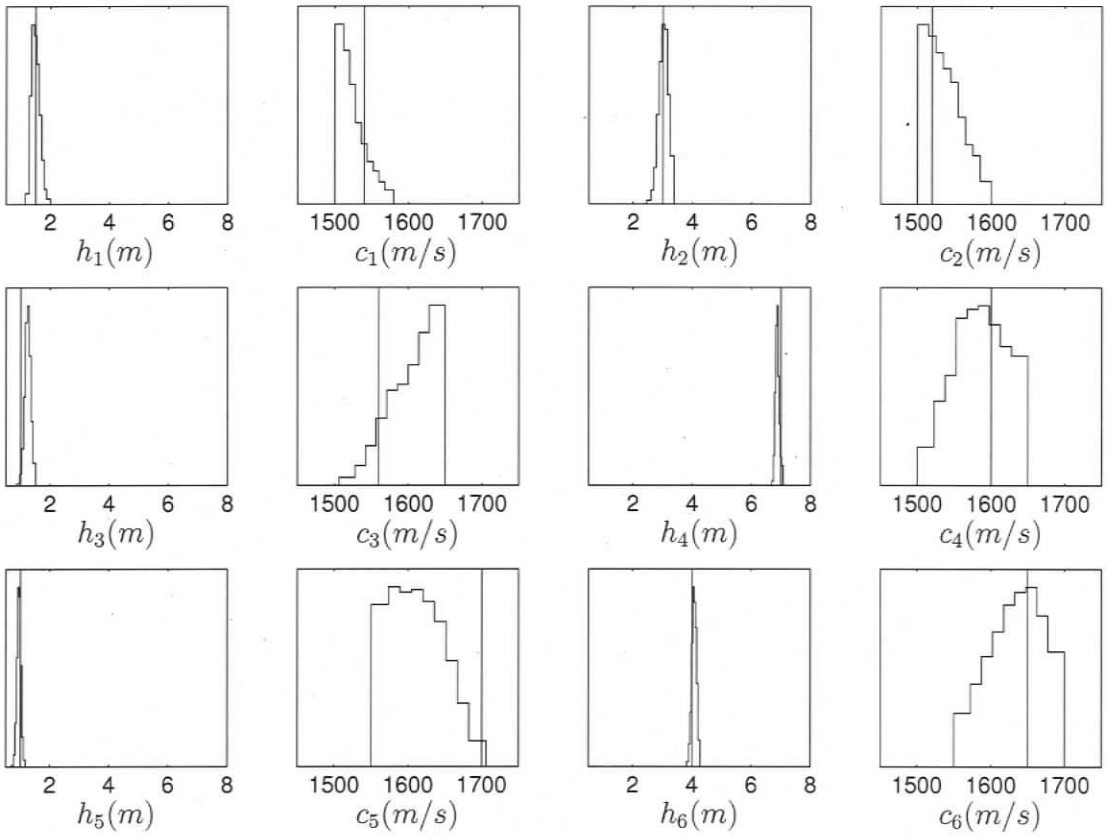

Fig. 6.15: Recovered marginal probability distributions of physical model parameters for the inversion of picked travel time data from simulated seismo-acoustic traces.

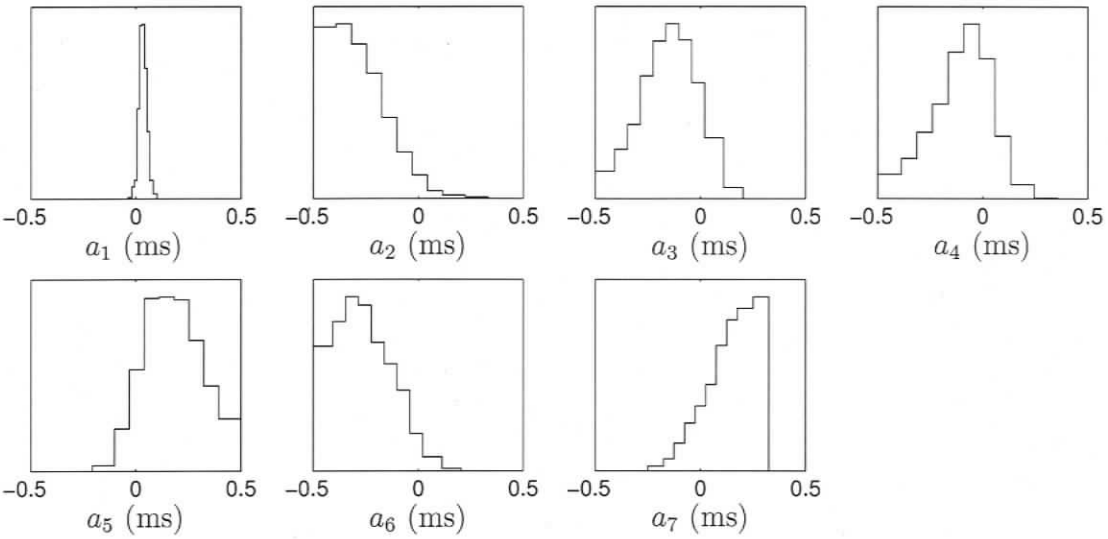

Fig. 6.16: Recovered marginal probability distributions of picking offsets of 7 reflectors for the inversion of picked travel time data from simulated seismo-acoustic traces.

Layer	Thickness (m)	Sound velocity (m/s)
1	1.15–1.99	1500–1580
2	2.45–3.37	1500–1599
3	0.87–1.51	1506–1650
4	6.68–7.07	1500–1649
5	0.69–1.17	1550–1706
6	3.82–4.27	1550–1700

Table 6.6: Information recovered from travel time inversion that can be used as prior knowledge in the following frequency domain inversion.

6.3. From Time Domain Traces to Reflection Coefficients

The frequency domain inversion takes reflection amplitude information into account in terms of reflection coefficients at various frequencies. The magnitude of the pressure reflection coefficient is estimated from the experiment as described in Sec. 6.1. The raw data for this experiment are time series (seismic traces) for shots at different source offsets. The seismo-acoustic recordings are provided by Charles W. Holland, ARL, Pennsylvania State University. To derive the magnitude of the pressure reflection coefficient from the time domain data, Holland developed a high resolution processing technique (Holland, 2003). This section will elaborate on the processing

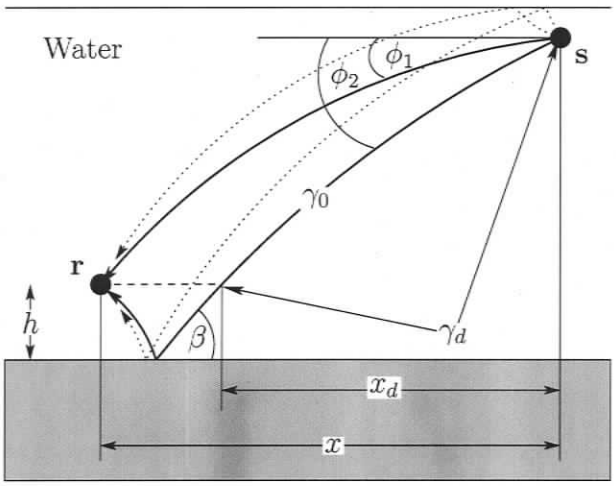

Fig. 6.17: Experiment geometry to measure reflection loss versus angle. The direct and the bottom bounce path are shown as solid lines, the surface reflected paths as dashed lines. Source **s** moves away from receiver **r** which is mounted at height h above the seabed. (after Holland and Osler (2000))

implementation developed for this thesis which closely follows Holland's approach.

The experiment geometry is illustrated in Fig. 6.17 for a shot at position **s**. As the source moves, data are recorded for each shot position as a time series of acoustic pressure. This results in a common receiver gather for a single hydrophone as shown in Fig. 6.18. Reflection coefficient data are then computed as a function of angle and frequency by time windowing the direct and bottom bounce paths and comparing their energy (correcting for geometric spreading and source directivity) as follows.

The magnitude of the pressure reflection coefficient can be written as

$$|\mathsf{V}(\beta, f)| = \left| \frac{p_r(x, f)}{p_0(x, f)} \right| , \tag{6.12}$$

where p_r is the received pressure for the bottom reflected path at range x and p_0 is the received pressure for a bottom reflected path with a perfectly reflecting halfspace which can be written as

$$p_0(x, f) = |p_s(\phi_2, f)|\, \gamma_0 \,. \tag{6.13}$$

In Eq. 6.13, p_s is the source pressure amplitude at 1 m from the source and γ_0 constitutes a transmission factor along the bottom bounce path that accounts for spreading, refraction, absorption, and multi-path effects. Due to non-uniform directionality of the source, we cannot obtain p_s from the direct path but must use the offset x_d for the correct take-off angle ϕ_2:

$$|p_s(\phi_2, f)| = q_d(x_d, f)\gamma_d^{-1} \,. \tag{6.14}$$

Here, q_d is the amplitude for the path with take-off angle ϕ_2 at receiver height h and γ_d is the transmission factor along the same path to range x_d and height h. Thus, q_d is not measured directly. Equation 6.12 then becomes

$$|\mathsf{V}(\beta, f)| = \frac{|p_r(x, f)|}{q_d(x_d, f)} \frac{\gamma_d}{\gamma_0} . \tag{6.15}$$

To obtain p_r, the seismic traces are time windowed around the energy of the bottom reflected path. The time windowed traces are then transformed into the frequency domain using a fast Fourier transform (FFT) and band averaged into appropriate frequency bands. As indicated, above q_d cannot be extracted from the data directly. Instead, the seismic traces are time windowed around the energy of the direct arrival. An FFT and frequency average are applied as before. The field is then interpolated (or extrapolated, depending on geometry) onto ranges x_d using a polynomial (of order 3–5, depending on the data). These ranges are found by ray-tracing for the particular geometry.

The transmission factors γ_d and γ_0 are also found from ray-tracing in the water column. Using the path lengths of the rays and an approximate attenuation coefficient for seawater (Jensen et al., 1993), absorption and geometrical spreading can be accounted for. The experimental set up results in only two significant surface multiples. These two multiples (see Fig. 6.17) can be taken into account by calculating transmission factors for these paths as well and then averaging over the direct path and the surface-bounce path factors.

It is also possible to correct the reflection coefficient estimate for source level variability. This can be important if the source amplitude fluctuates for other reasons than changes in the beam pattern (e.g., time dependent roll or pitch of the frame on which the source is mounted). In this case, the source pressure p_s can be corrected by the ratio of interpolated pressure $q_d(x_d, f)$ and fitted pressure $q_d(x, f)$:

$$|p_s(\beta, f)| = |p_d(x, f)| \frac{q_d(x_d, f)}{q_d(x, f)} \gamma_d^{-1} . \tag{6.16}$$

To eliminate bad traces from the data, two criteria are used. First, when interpolating q_d, outliers of more than two standard deviations are excluded. Several iterations of polynomial fitting and screening for outliers are performed for each frequency band. Second, the signal to noise ratio (SNR) is calculated for each trace

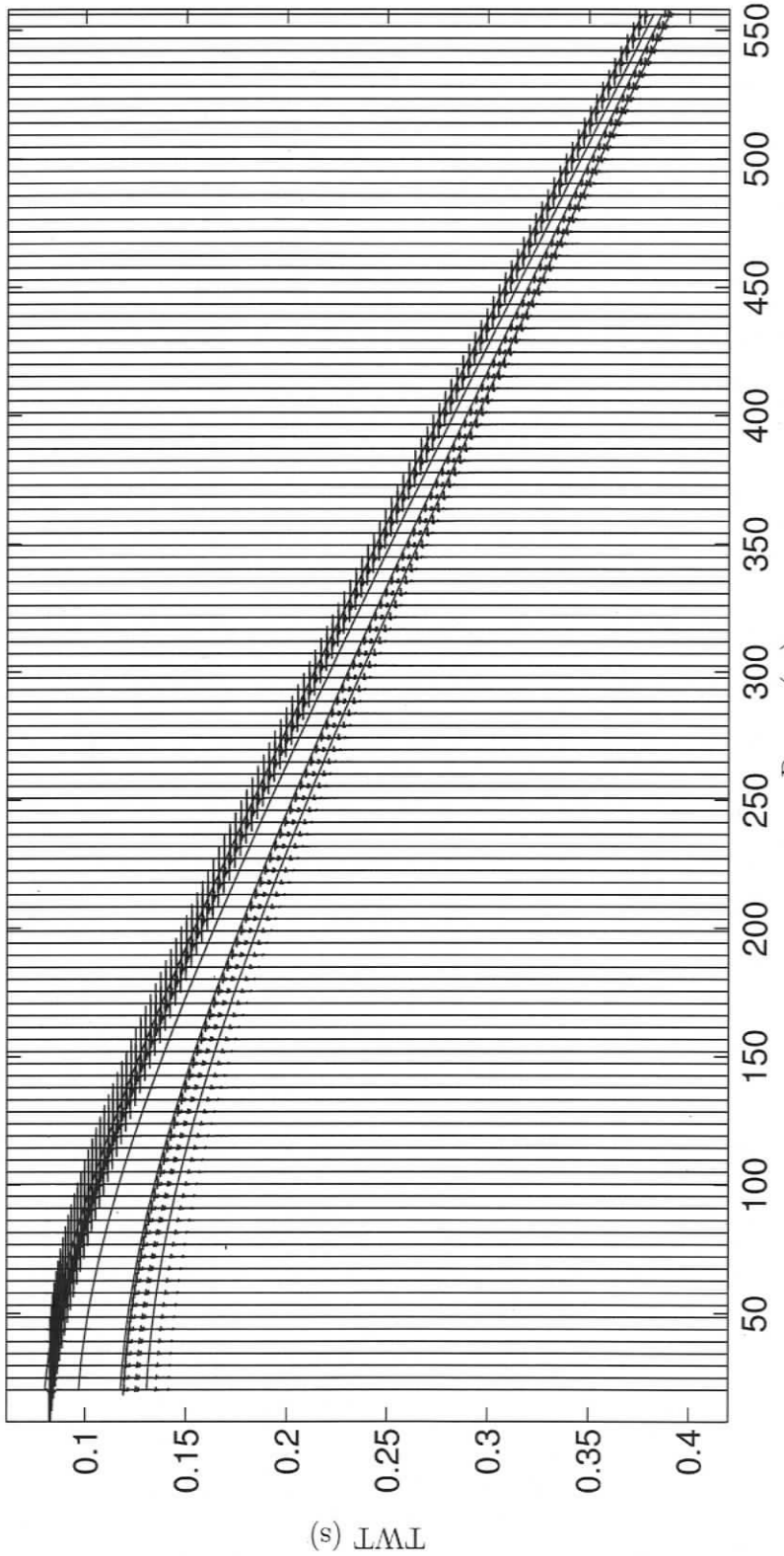

Fig. 6.18: The acoustic traces for a simulated data set consisting of 6 fluid layers over a halfspace. The time window hyperbolas for two windows are shown as black lines.

and frequency band (the noise level is calculated from time zero to the first acoustic arrival at each range). A sliding window is used to smooth the SNR over traces. All traces with SNR < 6 dB are excluded from the data set.

6.4. Frequency Domain Inversion

This section develops the second part of the joint time/frequency domain inversion that is based on the reflection coefficient data as derived from the seismo-acoustic traces (see Sec. 6.3). Different forward models for plane and spherical waves are developed, respectively. A layer packet stripping approach is introduced to recover complicated seabed structures from the frequency domain data while using the results of the time domain inversion as prior knowledge. Finally, two computer simulated experiments are used to illustrate the importance of accounting for spherical wave effects in certain environments.

6.4.1. Single Bounce Reflection Coefficient Forward Models

The numerical inversion approach performs iterations on a forward model (inversion by forward modelling). Hence, the formulation of the forward model is a critical part of the inversion process. The forward model used here approximates the seabed as a layered lossy fluid. This is justified, since shear velocities in fine grained sediments are low (Richardson, 1997) and earlier inversion studies (Dosso and Holland, 2006) that treated the seabed as a halfspace have shown that the reflection coefficient is relatively insensitive to shear properties and that ignored shear properties do not result in a bias for the physical properties of the fluid.

As indicated in Fig. 6.1, the experiment uses a point source generating spherical waves. Therefore, the measured reflectivity contains spherical wave effects. Depending on the geometry of the experiment and environment at the measurement site, these effects may be negligible or may be important and need to be taken into account to recover meaningful estimates of the geoacoustic parameters. In the following, both a plane wave reflectivity model and a spherical wave reflectivity model are developed to address this issue.

The parameterisation of the sea-bed used in the forward model depends on the characteristics of the experimental site. It is important to examine the time domain

seismo-acoustic traces to decide what parameterisation would be effective. For example, the time traces can help determine the number of layers. The reflection data can have prominent features (such as critical angles or an angle of intromission) that determine which parameterisation is appropriate. Layers can be parameterised as homogeneous or contain variably shaped gradients in sound velocity, density, and attenuation. However, it is advantageous to keep the parameterisation as simple as possible.

6.4.2. Plane Wave Reflectivity Model

For shallow water (100–200 m water depth), low sound velocities (e.g., in fine grained sediments), small layer thicknesses (10^{-1}–10^{0} m), and a frequency range of 300–1600 Hz, a plane wave assumption is reasonable (see Chapter 2). In this case, the forward model consist of a recursive calculation as explained in detail in Chapter 2, Eq. 2.30:

$$\begin{aligned} \mathsf{V}_{(m-2)m} &= \frac{\mathsf{V}_{(m-2)(m-1)} + \mathsf{V}_{(m-1)m}\exp(2i\phi_{m-1})}{1 + \mathsf{V}_{(m-2)(m-1)}\mathsf{V}_{(m-1)m}\exp(2i\phi_{m-1})} \\ \mathsf{V}_{(m-3)m} &= \frac{\mathsf{V}_{(m-3)(m-2)} + \mathsf{V}_{(m-2)m}\exp(2i\phi_{m-2})}{1 + \mathsf{V}_{(m-3)(m-2)}\mathsf{V}_{(m-2)m}\exp(2i\phi_{m-2})} \\ &\vdots \\ \mathsf{V}_{1m} &= \frac{\mathsf{V}_{12} + \mathsf{V}_{2m}\exp(2i\phi_{2})}{1 + \mathsf{V}_{12}\mathsf{V}_{2m}\exp(2i\phi_{2})}. \end{aligned} \tag{6.17}$$

6.4.3. Spherical Wave Reflectivity Model

For reflection coefficients rapidly changing with angle and complex environments to more than a few wavelengths depth below the seafloor, spherical wave effects are often significant. Harrison and Nielsen (2004a,b) outline an interesting approach to derive plane wave reflection coefficients from point source measurements, and showed that plane wave and spherical wave reflection coefficients can be significantly different even for large source-receiver distance (210 m source-receiver separation at 500 Hz). In this thesis, a forward model is developed that derives a reflection coefficient, including spherical wave effects, from complex acoustic fields. The full acoustic fields are computed by wavenumber integration techniques (OASES, Schmidt, 1988) and then processed using ray tracing to resemble reflection coefficients that take full spherical wave effects into account. Several numerical optimisations are applied to run the

computationally intensive forward model very efficiently.

The full acoustic field calculation, which is computationally intensive, is based on the transmission loss module of OASES. Let the full acoustic field be given by Ψ. The field can then be split into two parts, a direct wave and a seabed response term

$$\Psi = \frac{e^{i\int_{r_1} k(u)du}}{r_1} + \mathsf{V_s}\frac{e^{i\int_{r_2} k(u)du}}{r_2}, \tag{6.18}$$

where k is the wavenumber and r_1 and r_2 are the path lengths of the direct and bottom bounce paths, respectively. $\mathsf{V_s}$ is a measure for the seabed response which can be recovered by rearranging Eq. 6.18

$$\mathsf{V_s} = \frac{\Psi - e^{i\int_{r_1} k(u)du}/r_1}{e^{i\int_{r_2} k(u)du}/r_2}. \tag{6.19}$$

The path lengths r_1 and r_2 and integrals along the rays are found by ray-tracing in the water-column using the measured sound velocity profile. Hence, $\mathsf{V_s}$ quantifies the seabed response by correcting the field Ψ for spreading loss of direct and bottom reflected paths (ignoring ray tube effects). All spherical wave effects (including possible lateral and inhomogeneous waves) are contained in $\mathsf{V_s}$ and will be accounted for in the inversion. It should be noted that Eq. 6.18 is a convenient way to process the acoustic field replica data to resemble the processing of the measured data and is therefore suitable for the inversion. For convenience, $\mathsf{V_s}$ is referred to as the spherical reflection coefficient, even though more than just reflection effects enter the coefficient.

Since reflection coefficients show a strong frequency dependence, frequency averaging must be applied. Here, the frequency average is replaced by an equivalent range average as outlined in Sec. 6.5. With the shortest range being in the order of 10^1 m, the wavenumber integration model is run with full Hankel transforms eliminating any far field approximation. This ensures precise modelling of the acoustic field even at small ranges at the price of higher computational effort.

This forward model is powerful and general and can calculate the spherical wave reflection coefficient for arbitrary sound velocity profiles in the water and arbitrary layering in the seabed. It can also account for interface roughness, and shear and gradient layers can be built into the model parametrisation.

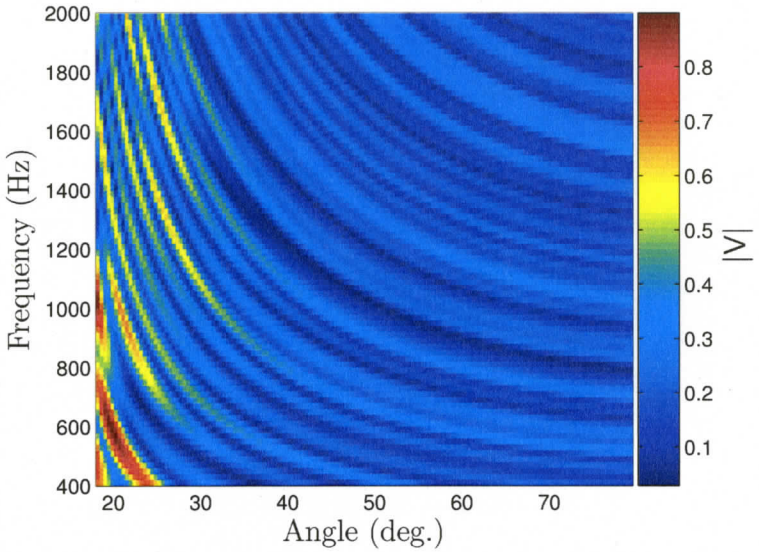

Fig. 6.19: Simulated reflection coefficient over angle and frequency for a typical seabed case with 6 sediment layers over a half-space.

6.5. Frequency and Range Averaging

Some of the forward models that are used in this work are very computationally expensive. This, combined with the number forward computations needed for a typical ASSA or FGS run (in the order of 100,000 models), can make it slow to calculate reflection coefficients at many frequencies, even on sophisticated supercomputers. In particular, the measured data need to be frequency averaged to improve the signal to noise ratio. Due to the strong frequency dependence of the reflection coefficient, the forward model must also include frequency averaging. Depending on the bandwidth and the particular environmental model, the number of frequencies to average over varies. In most cases, 8 or 9 frequencies have proven to provide sufficient averaging.

Figure 6.19 shows the reflection coefficient over angle (which translates into range as shown in Sec. 6.3) and frequency for a typical sediment model consisting of six layers over a semi-infinite half-space. The advantage of averaging over range rather than frequency is that the wavenumber integration in the forward model used to calculate the acoustic fields provides the field at a fairly dense range spacing. Thus, spacial averaging can be carried out at almost no extra computational cost. The computational cost of frequency averaging, as indicated above, scales linearly with the number of frequencies used in the average (the majority of computational time

is spent computing the data at each frequency). In this work, range averaging was found to be an order of magnitude faster than frequency averaging.

Harrison and Harrison (1995) illustrated a simple relationship between frequency and range averaging for broadband sonar. Consider a Gaussian frequency average as

$$I_f = \frac{\int \Psi(f, r_0) \exp(-(f - f_0)^2/(\kappa f_0)^2) df}{\int \exp(-(f - f_0)^2/(\kappa f_0)^2) df}, \tag{6.20}$$

where f_0 denotes the centre frequency of the average, κ the fractional bandwidth, and Ψ is the acoustic field which is explicitly given as a function of frequency and range. A range average can also be defined as

$$I_r = \frac{\int \Psi(f_0, r) \exp(-(r - r_0)^2/(\kappa r_0)^2) dr}{\int \exp(-(r - r_0)^2/(\kappa r_0)^2) dr}, \tag{6.21}$$

where r_0 is the centre range and κ is the same fraction as above. Equation 6.21 is a sliding window of width proportional to the centre range. Harrison and Harrison (1995) showed for Lloyd's mirror (i.e., a point source in a semi-infinite half-space with perfectly reflecting interface) that the above averages are identical. In other examples they illustrated that the two are very similar. Depending on the choice of κ, the Gaussian window length can vary from monochromatic to broadband.

Figure 6.20 compares frequency and range averages for a simulated experiment. The seabed consists of three sediment layers over a sediment halfspace; data were generated as time series (synthetic seismo-acoustic traces) and then processed to resemble reflection coefficients. The agreement is excellent over the whole frequency and angle range. Other simulations showed that the agreement is not as good for frequencies much below those in Fig. 6.20. However, for the frequency range of interest in this work, the range average reduces computation time by an order of magnitude at no significant loss in accuracy.

6.6. Bayesian Layer Packet Stripping Inversion

This section develops a layer packet stripping inversion algorithm for reflection coefficient data that resolves complex multi-layered sediment environments and provides rigorous uncertainty estimates. The approach is based on windowing acoustic time-series to isolate reflections from sequences of sediment layers (packets) containing

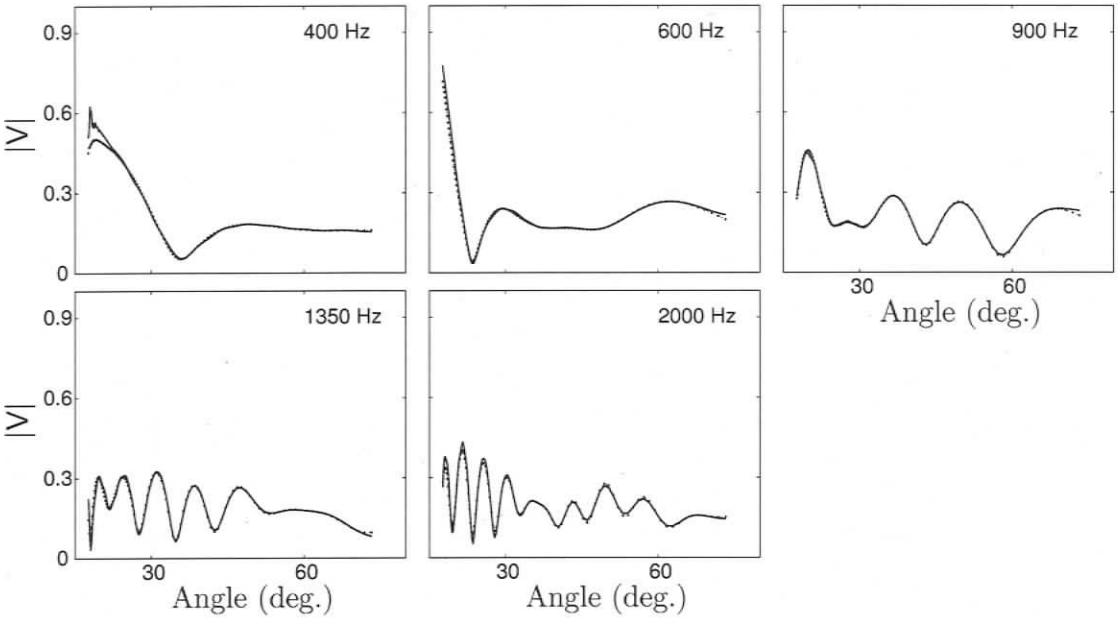

Fig. 6.20: Comparison of frequency (solid dotted line) and range (solid line) averages for simulated data derived from synthetic seismic traces for three sediment layers over a halfspace.

information from the sea-floor to increasing depths, i.e., every packet samples the layers of the previous packet and additional, deeper layers. The acoustic data are processed to represent reflected energy across frequency and grazing angle.

For the layer-packet stripping inversion, the L_2-norm is used in the generalised misfit function (Eq. 4.4). The first packet is inverted with wide uniform prior information for all parameters (unless the parameters are constrained through a prior time domain inversion). Subsequent inversions for packets that include deeper layers use the information from previous inversions as prior knowledge. The packet that is inverted thus has a number of layers that have uniform priors and a number of layers that have prior knowledge from earlier inversions. Once an inversion for a certain packet is completed, marginal distributions are calculated from the PPD. These marginals provide parameter bounds in the next inversion as follows. The prior probability $P(\mathbf{m})$ is computed from the probabilities p_i of the marginal distributions for layer parameters that were included in earlier inversions

$$P(\mathbf{m}) = \prod_{i=1}^{M} p_i(m_i)\,. \tag{6.22}$$

Using marginal distributions for the calculation of the prior probability is an approximation that does not account for correlations between parameters. In theory, the full PPD can be used to compute prior probabilities, however, due to the large number of parameters in many problems, this is not feasible (the "curse of dimensionality"). To partly include parameter correlations, a principal component rotation is performed. The rotation matrix is obtained from the eigenvector decomposition of the model covariance matrix determined from the PPD. This matrix is then used to rotate the PPD into its principal components and marginal distributions are computed for the rotated parameters. The rotation is determined from the eigenvector decomposition of the model covariance matrix $\mathbf{C}^{(\mathbf{m})}$ (see Eq. 4.8 and Eq. 4.27 for details)

$$\mathbf{C}^{(\mathbf{m})} = \mathbf{U}_p \mathbf{\Lambda} \mathbf{U}_p^{\mathsf{T}} . \tag{6.23}$$

The prior information then consists of the rotated marginals (Fig. 6.21) and the prior rotation matrix $\mathbf{U}_p$. When evaluating the misfit, the prior probability $P(\mathbf{m})$ is computed for the rotated model $\mathbf{m}'$. The transformation pair between rotated and physical space is

$$\begin{aligned} \mathbf{m}' &= \mathbf{U}_p^{\mathsf{T}} \mathbf{m} \\ \mathbf{m} &= \mathbf{U}_p \mathbf{m}' . \end{aligned} \tag{6.24}$$

The following sections apply the Bayesian layer stripping approach in two simulations. The first simulation uses the plane wave forward model that, while fitting the data well, was found to be insufficient to resolve the parameters of complex multilayered structures correctly, especially at greater depths. Therefore, in the second simulation, the spherical wave forward model is used and provides superior results.

6.7. Plane Wave Inversion Simulation

For the plane wave inversion, the time domain experiment was closely simulated by calculating synthetic seismic traces with the OASES program package. The underlying geoacoustic model consists of 6 sediment layers over a sediment half-space. Each layer is defined by layer thickness h, sound velocity c, density ρ, and attenuation coefficient α. A single receiver was placed at a depth of 122 m in a 150 m water-

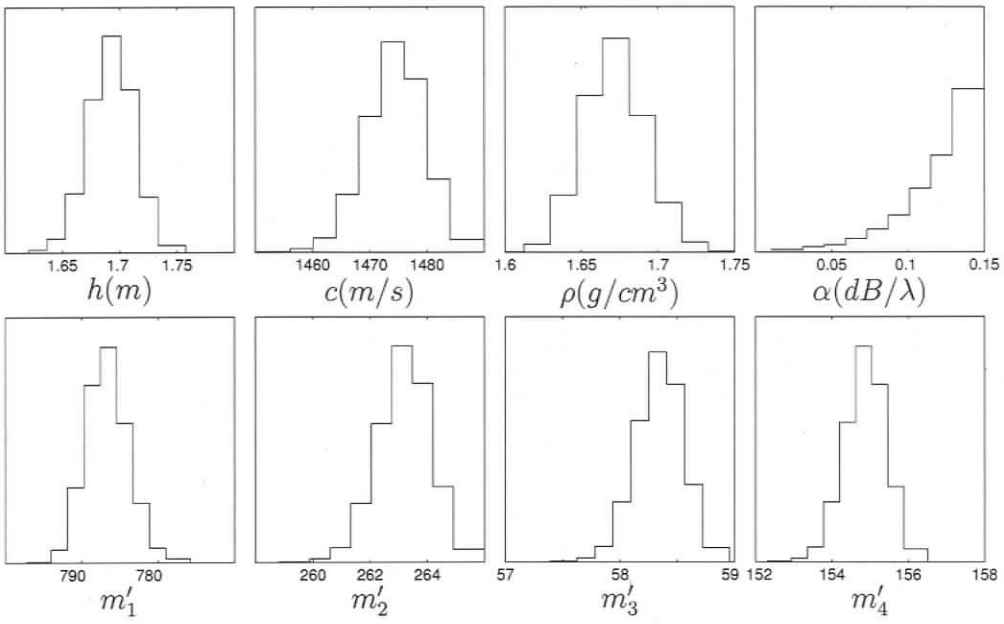

Fig. 6.21: The upper row shows an example of unrotated prior information for one layer with four parameters. The lower row shows the rotated prior for the same PPD.

column. A source was simulated by the source wavelet shown in Fig. 6.9 at ranges up to 560 m (see Fig. 6.17) with a range spacing of 4.7 m. The reflectivity data, which includes spherical wave effects, were extracted from the time series and processed in nine frequency bands from 315–2000 Hz. Each frequency represents a 1/3 octave band frequency average. Gaussian random errors (1 dB standard deviation) were added to the reflection coefficient data. The replica data were computed using the plane wave forward model. Frequency averaging was performed for each band using nine frequencies which were logarithmically spaced through the band. The data were processed into three layer packets. The first packet contained the effects to below the first layer only, the second packet contains the effects to below the third layer, and the last packet contained the reflected energy of all six sediment layers.

After applying the layer stripping inversion (prior bounds given in Table 6.7), the inversion results of the final PPD are interpreted in terms of MAP models and 95% credibility intervals for the sound velocity and density profiles. Figure 6.22 shows that the data fit of the MAP model (last packet) is reasonably good across all frequencies. The figure also shows data that were generated with the plane wave model using the true geoacoustic parameters. There is a noticeable difference between the spherical

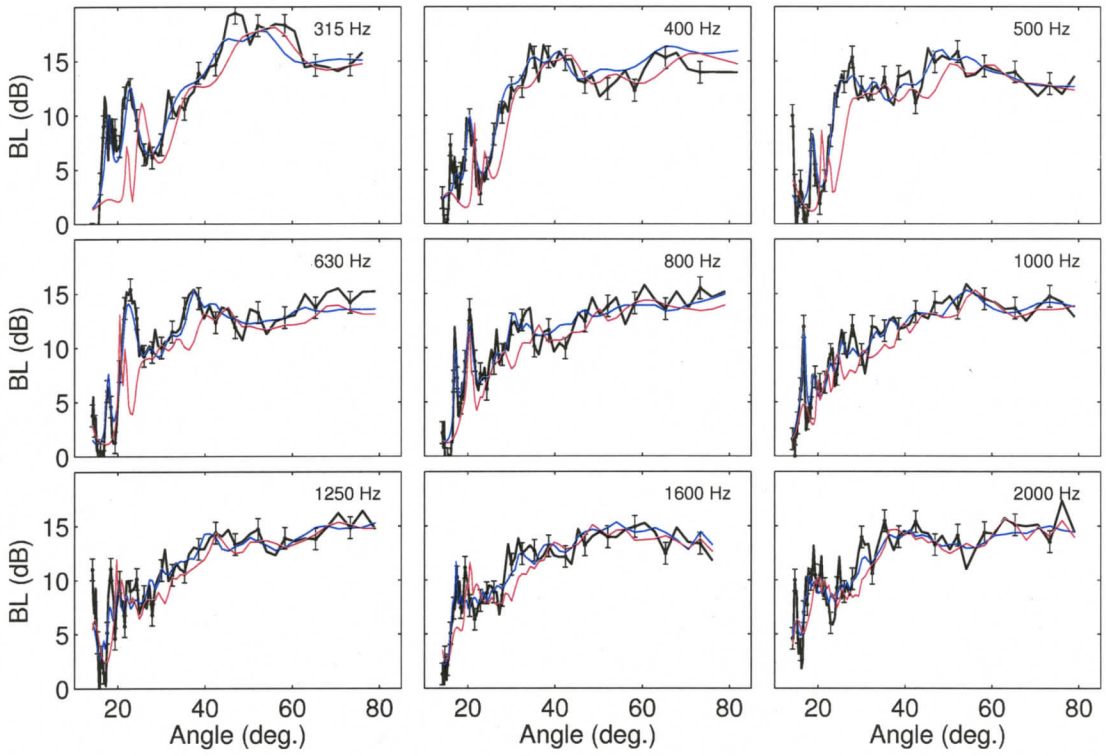

Fig. 6.22: Fit of the MAP replica data (blue) to the simulated data (black dotted line). One standard deviation error bars are given for every third datum. The red line shows data generated with the plane wave forward model for the true parameters.

and plane wave data for the true model, especially at low frequencies.

Figure 6.23 shows the inversion results extracted from the final PPD. It can be seen that while velocities and densities of the upper three layers are reasonably well determined, deeper layers cannot be recovered with the plane wave forward model. The velocity profile shows strong negative biases at depth and does not represent the true profile well. Strong biases do not occur for the density profile; however, the 95% HPD credibility intervals indicate that the inversion cannot resolve density well below the third layer. Further, layer thicknesses show strong discrepancies from the true model. Numerical values for the inversion results are given in Table 6.8.

To some degree, the negative bias observed in the sound velocity can be understood by the following simplified example. Consider two models of a single fluid layer between two fluid half spaces. Let the physical parameters for the three sediments of the first model be given by $c_1 = 1511$ m/s, $\rho_1 = 1.029$ g/cm^3, $\alpha_1 = 0$ dB/λ, $h_2 = 1.5$ m, $c_2 = 1500$ m/s, $\rho_2 = 1.4$ g/cm^3, $\alpha_2 = 0.03$ dB/λ, $c_3 = 1700$ m/s, $\rho_3 = 1.4$ g/cm^3

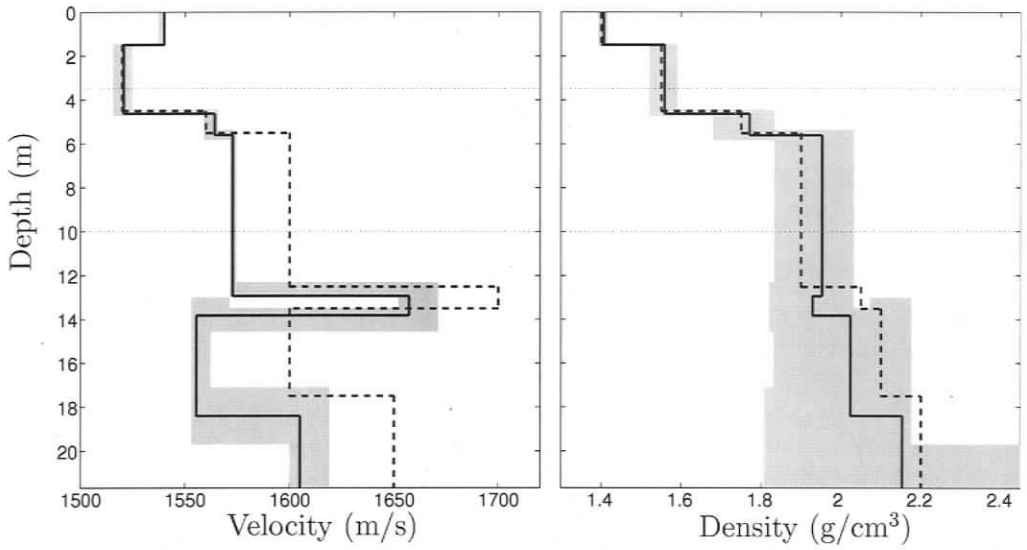

Fig. 6.23: Bayesian layer-stripping inversion results of spherical-wave reflection coefficient data using the plane wave forward model. The solid line indicates the *a posteriori* mean model; shaded areas are the 95% HPD credibility intervals. The dashed line represents the true model and dotted horizontal lines are the lower limits of the packets.

and $\alpha_3 = 0.06$ dB/λ. In a second model, c_3 is replaced by 1670 m/s. Figure 6.24 shows a comparison between spherical and plane wave reflection coefficients for the two models. The critical angle for the spherical reflection coefficient is lower than for the plane wave case. This is compared to a plane wave reflection coefficient computed for a slightly lower sound velocity in the basement, with all other parameters held constant. It can be seen that a negative bias in sound velocity shifts the critical angle to lower values, which is consistent with the bias to lower velocities observed in the inversion.

Layer	h (m)	Sound velocity (m/s)	Density (g/cm^3)	Attenuation (dB/λ)
1	1.15–1.99	1500–1600	1.2–1.6	0.01–0.10
2	2.45–3.37	1500–1600	1.3–1.8	0.05–0.15
3	0.87–1.51	1500–1600	1.3–1.8	0.05–0.15
4	6.68–7.07	1500–1600	1.4–2.0	0.10–0.20
5	0.69–1.17	1550–1650	1.6–2.2	0.05–0.15
6	3.82–4.27	1550–1650	1.8–2.2	0.01–0.15
7		1600–1700	1.8–2.4	0.00–0.05

Table 6.7: Prior bounds for all layers.

Thickness (m)		Sound velocity (m/s)		Density (g/cm^3)		Attenuation (dB/λ)	
true	95% HPD	true	95% HPD	true	95% HPD	true	95% HPD
1.5	1.46–1.53	1540	1537–1541	1.40	1.39–1.41	0.05	0.13–0.18
3.0	2.99–3.21	1520	1515–1524	1.55	1.52–1.59	0.10	0.04–0.08
1.0	0.90–1.09	1560	1559–1566	1.75	1.68–1.83	0.15	0.05–0.11
7.0	6.96–7.64	1600	1571–1574	1.90	1.83–2.03	0.10	0.05–0.11
1.0	0.69–1.08	1700	1652–1671	2.05	1.82–2.07	0.10	0.03–0.47
4.0	4.09–5.14	1600	1553–1562	2.10	1.83–2.17	0.10	0.13–0.47
		1650	1600–1619	2.20	1.81–2.49	0.01	0.03–0.50

Table 6.8: Numerical values for true parameters and inversion results (95% HPD credibility intervals) for the plane wave inversion.

The inversion results in Fig. 6.23 show that the plane wave forward model is not sufficient to recover the complex structure in this case. Even though the data can be fit well, the good fit is only obtained by introducing a bias. Not only is the data fit misleading about the quality of the results but the HPD credibility intervals also provide false confidence in the results. It is thus important to pay attention to the simplifications in the forward model and check their validity. In this case, the plane wave model is insufficient and a spherical wave model must be considered.

6.8. Spherical Wave Inversion Simulation

For the spherical wave inversion, the time domain experiment was simulated as described in Sec. 6.7. The resulting simulated time traces were windowed into three packets. The first packet contained energy to below the first layer, the second packet contained reflected energy to below the third layer and the final packet included all layers. The time domain data were then processed to represent reflected energy across frequency (400–2000 Hz) and grazing angle (20–75°) according to Holland (2003). The five frequencies represent the centre of frequency bands of fractional bandwidth 1/20. Finally, Gaussian noise (standard deviation 0.1) was added to the reflection data (Fig. 6.25).

Several numerical optimisations are applied to run the computationally intensive forward model as efficiently as possible. Both ASSA and the Gibbs sampler were implemented and run on a massively parallel computer using message passing, resulting in a saving of several orders of magnitude in computational time. Further, range

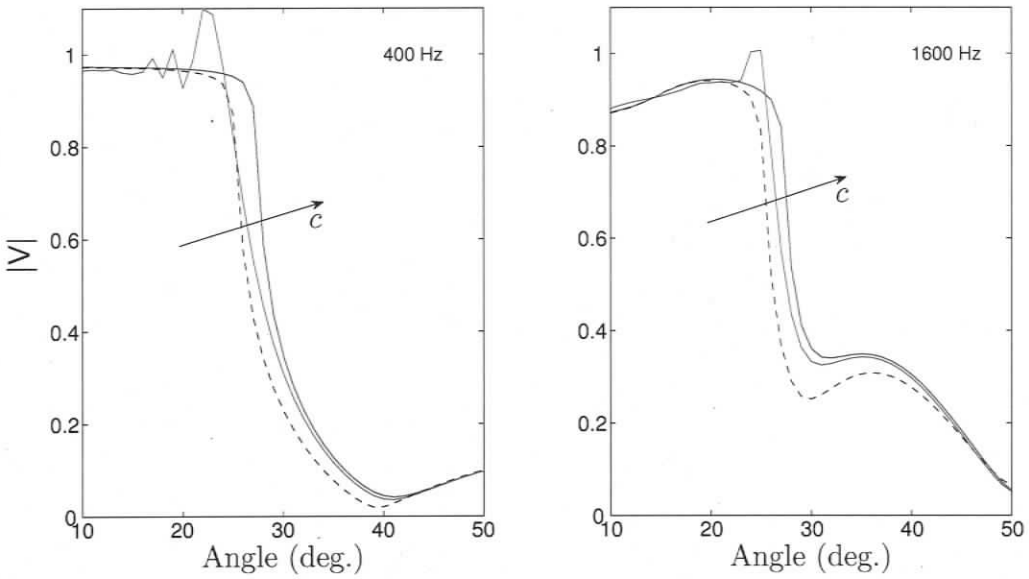

Fig. 6.24: Spherical (red) and plane wave (blue) reflection coefficients at two discrete frequencies for a single layer over a halfspace (both fluid). The critical angle for the spherical reflection coefficient is shifted to lower angles compared to the plane wave reflection coefficient. This shift translates into a lower sound velocity contrast (black dashed line) for plane wave reflection coefficient. The arrow indicates the shift of the critical angle for increasing basement sound velocity.

averaging is applied to replace the frequency average in each band (see Sec. 6.5).

Each packet was inverted using ASSA under the assumption of unknown error magnitudes at each frequency. The resulting MAP model was then used to find a maximum likelihood estimate of the error standard deviation at each frequency. These standard deviations were then used in a fast Gibbs sampler to sample the PPD for each packet. The PPD for the final packet (27 parameters) is regarded as the full solution to the inverse problem. Figure 6.26 shows marginal probability distributions extracted from this PPD. The *a posteriori* mean model is very close to the true model parameters for both the velocity and density profiles. Layer thicknesses are also matched well. The velocity tends to be better resolved in thick layers. In general, the uncertainty in the parameter estimate grows with depth, particularly for the density profile. Attenuation was included as a parameter in the inversion but in most cases is not resolved well. The best resolution for attenuation was obtained in the thickest layers (see Table 6.9).

The inversion resolved velocity and density profiles and their associated uncertainties for six sediment layers and a basement (27 parameters) from simulated single

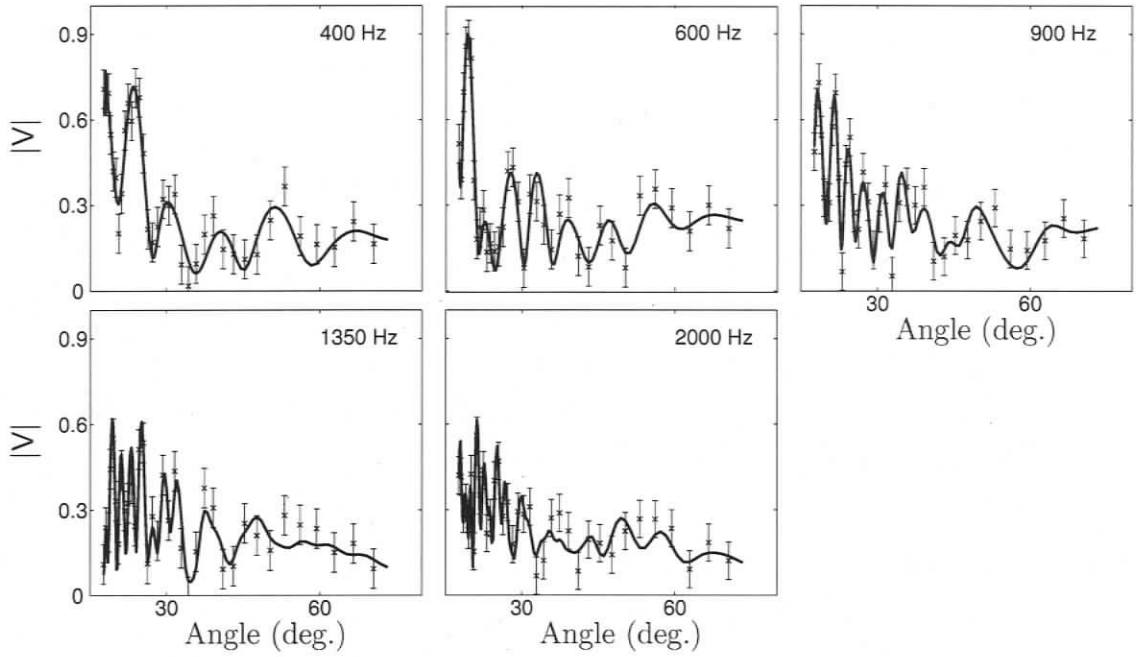

Fig. 6.25: Simulated reflection data for the final packet. Data are shown with one standard deviation error bars. For clarity, only every 3rd datum is shown. The solid lines are the replica data generated from the MAP model for the final packet.

bounce seabed reflectivity data. The final results quantify the information in the data correctly and show substantial improvements over the plane wave inversion.

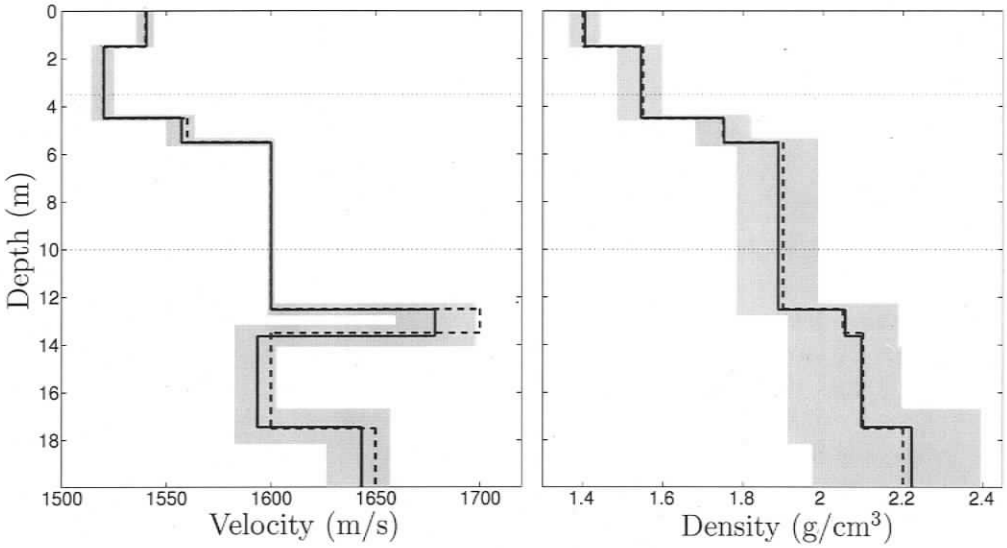

Fig. 6.26: Bayesian layer-stripping inversion results for the spherical wave forward model. The solid line indicates the *a posteriori* mean model; shaded areas are the 95% HPD intervals. The dashed line represents the true model that was used to generate the data and dotted horizontal lines are the lower limits of the packets.

Thickness (m)		Sound velocity (m/s)		Density (g/cm^3)		Attenuation (dB/λ)	
true	95% HPD	true	95% HPD	true	95% HPD	true	95% HPD
1.5	1.46–1.51	1540	1539–1543	1.40	1.35–1.39	0.05	0.02–0.10
3.0	2.98–3.05	1520	1515–1520	1.55	1.51–1.57	0.10	0.05–0.12
1.0	0.95–1.02	1560	1552–1559	1.75	1.68–1.78	0.15	0.10–0.20
7.0	6.91–7.07	1600	1597–1599	1.90	1.79–1.96	0.10	0.08–0.13
1.0	0.86–1.12	1700	1684–1716	2.05	1.92–2.17	0.10	0.05–0.15
4.0	3.80–4.17	1600	1589–1602	2.10	1.94–2.19	0.10	0.05–0.15
		1650	1635–1655	2.20	2.01–2.39	0.01	0.00–0.05

Table 6.9: Numerical values for true parameters and inversion results (95% HPD credibility intervals) for the spherical wave inversion.

7. Joint Time/Frequency Domain Inversion: Malta Plateau Data

7.1. Introduction

This chapter applies the joint time/frequency domain inversion to acoustic data collected during the SCARAB98 experiment on April 22, 1998, on the Malta Plateau (Site 2), Mediterranean Sea (see Fig. 7.1). The goal is to extract high-resolution, multi-layer sound velocity and density structure for the upper parts of the seabed, including rigorous uncertainty estimates. The data processing to obtain travel time and reflection coefficient data are described in Chapter 6 and the Bayesian inversion methodology in Chapters 4 and 5.

The experiment geometry is shown in Fig. 6.1. In this particular case, the source was an electro-mechanical impulsive source (EG&G model 265 Uniboom referred to as a "boomer") with a short pulse length (< 1 ms) and a broad bandwidth. Data were recorded at a single receiver that was part of a vertical line array with 16 Benthos AQ-4 hydrophones. The hydrophone used in this data set was at 122 m depth and the water depth was 153 m. The sound velocity profile is given in Fig. 7.2 and shows a fairly constant velocity structure with less than 5 m/s variation over the whole water column.

At the experimental site, seismo-acoustic data were collected for two perpendicular ship tracks, both crossing the coordinates of the receiver. This results in four individual data sets, two with the ship approaching the receiver and two with the ship departing from the receiver. All four data sets were considered for this site, but the results were very similar and only one set is shown here. The seismo-acoustic data are inverted for an environmental model consisting of multiple flat lying layers without shear properties. The full set of seismo-acoustic traces is given in Fig. 7.3. The direct arrival or water wave can be seen at 0.075 s zero offset (range zero) time. This is

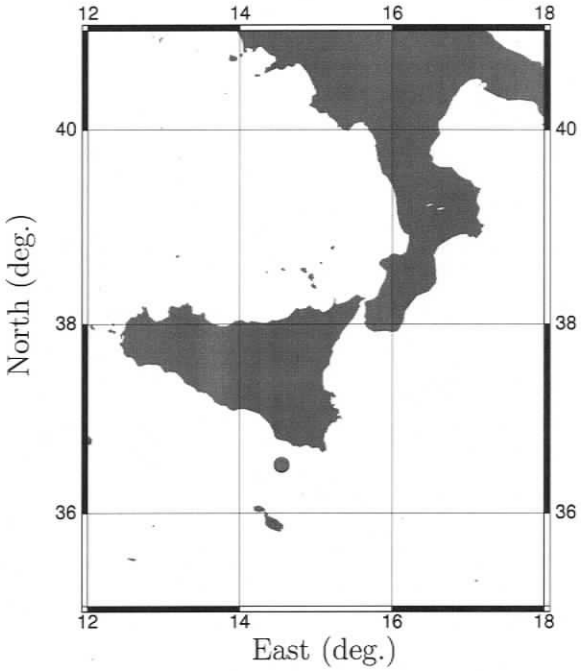

Fig. 7.1: Malta Plateau 1998 experiment location.

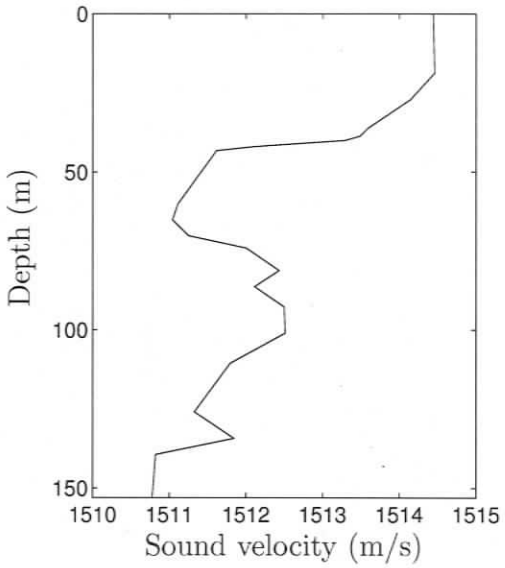

Fig. 7.2: Malta Plateau sound velocity profile.

followed by the sediment-water interface reflection (0.13 s zero offset time), followed by a number of sub-bottom reflections. The largest zero offset time of interest is 0.165 s. The first water column multiple can be seen at approximately 0.270 s, which gives good separation between the single bounce arrivals and water multiples. The data show exceptionally low noise levels and multiple reflections are clearly visible. The time domain inversion will be used to recover layer thickness probability distributions that are then used as prior information in the frequency domain reflection coefficient inversion. Chapter 6 derived plane wave and spherical wave forward models and showed the limitations of the plane wave approximation. Due to rapid changes of the reflection coefficient with angle and a maximum depth of greater than 25 m below the water-sediment interface in the Malta Plateau data, the spherical wave forward model is used in this chapter.

The remainder of this chapter describes the inversion and results. The final PPD is interpreted in terms of marginal distributions, optimal (MAP) parameter estimates and HPD credibility intervals. Recovered density and velocity profiles and credibility intervals are compared to cores taken at the experiment site. Finally, the MAP model is used to generate synthetic seismo-acoustic traces that are qualitatively compared

to the measured data.

It is worth noting that this data set turned out to be particularly challenging for the inversion process. A major problem with the environment was identified only after the inversion was completed and several different approaches to improve the forward model. Complicated three-dimensional geological structure was found on a seismic section that was shot on a perpendicular track to the experiment (shown later). This is an unfortunate and serious violation of the assumption of a horizontally stratified environment. The receiving hydrophone in the experiment was in fact directly on top of this complicated structure. Nonetheless, this chapter will present the complete inversion conducted for this data set. For deeper layers, problems with the parameterisation are evident and appear in the inversion results for deeper layers, particularly sediment densities. However, the results for the upper layers are not affected by this problem and show good agreement with cores.

7.2. Travel Time Inversion

The travel time data were picked (first breaks) from the seismo-acoustic traces as discussed in Sec. 6.2.2. Two iterations of the picking procedure were used to ensure high quality data. Particular attention was paid to pick the travel times objectively without altering the picks by interpreting the seismo-acoustic data (as is often done). A crop of the resulting picks is shown in Fig. 7.4. Seven reflectors including the sediment water interface were picked. The gaps in the data are due to the recording system writing to hard disk during the experiment. Travel times were interpolated over the gaps to simplify the processing, particularly the calculation of covariance matrices (Eq. 5.14) and the statistical validation of assumptions which were based on uniformly spaced residuals. Since the travel time hyperbolas can be assumed to be well behaved in these gaps, which were at most three traces wide, linear interpolation was found to be sufficient. This resulted in a travel time data set consisting of 83 points per reflector for offsets between 30 and 420 m with a range spacing of ~ 4.7 m. The maximum offset was limited to just over 400 m, because travel time differences (due to relatively thin layers) became too small to pick confidently at larger offsets.

The data residuals between the arrival times predicted for the MAP model and the picked data are shown in Fig. 7.5 and indicate correlated errors. Figure 7.5 also shows recovered time offset parameters for all reflectors (as introduced in Eq. 6.8).

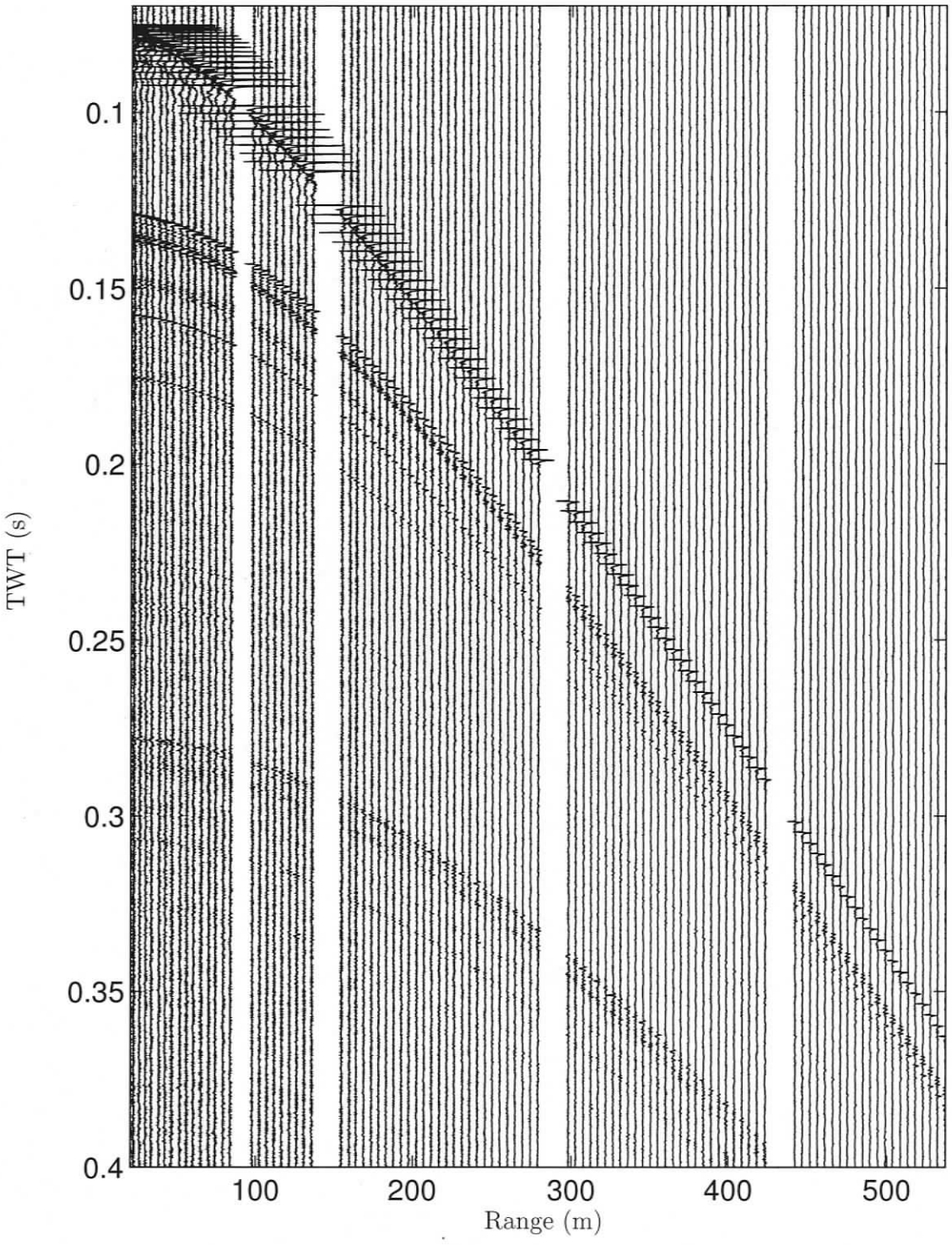

Fig. 7.3: Malta Plateau seismo-acoustic data. The reflections of interest end at 0.165 s zero offset time. The first water column multiple starts at approximately 0.270 s.

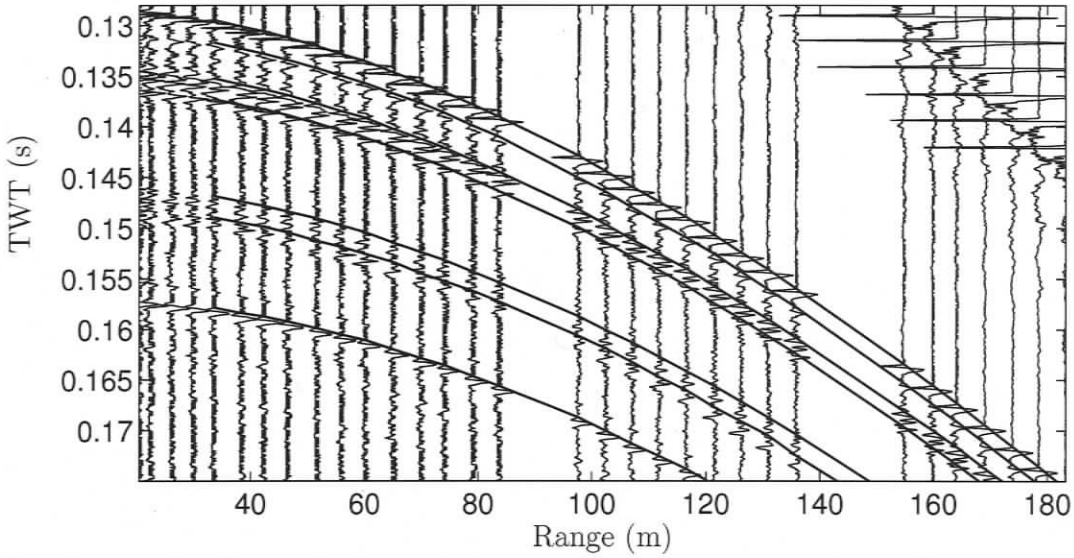

Fig. 7.4: Crop of picked travel times.

The recovered offset parameters illustrate typical limitations in picking. For example, reflectors two and four show large offsets, and both these reflectors were difficult to pick since they are located in a complicated part of the signal. These complications can be due to single thin layers where the events of two different reflectors overlap, or can be due to multiple thin layers below the resolution limit of the source pulse. Figure 7.6 shows a 20 point RMS filter average of the estimated standard deviation of the data residuals. The variable standard deviation level across range indicates non-stationarity of the data residuals. Due to the correlated and non-stationary errors, a non-Toeplitz data covariance matrix was estimated and included in the likelihood function of the inversion (see Sec. 6.2.2). The covariance matrix estimates for reflectors one, two, three and seven are shown in Fig. 7.7.

The Bayesian inversion (Chapter 4) was then applied to the travel time data weighted according to the estimated data covariance matrices. The resulting marginal distributions for layer thicknesses and sound velocities are given in Fig. 7.8. It can be seen that all layer thicknesses are resolved well within the prior bounds (which correspond to the plot widths). Sound velocities had a tighter prior relative to their numerical values and thus appear wider (prior bounds for sound velocities are given in Table 7.1). This means that the sound velocities are not as well resolved relative to their prior bounds as the layer thicknesses. However, they are reasonably well

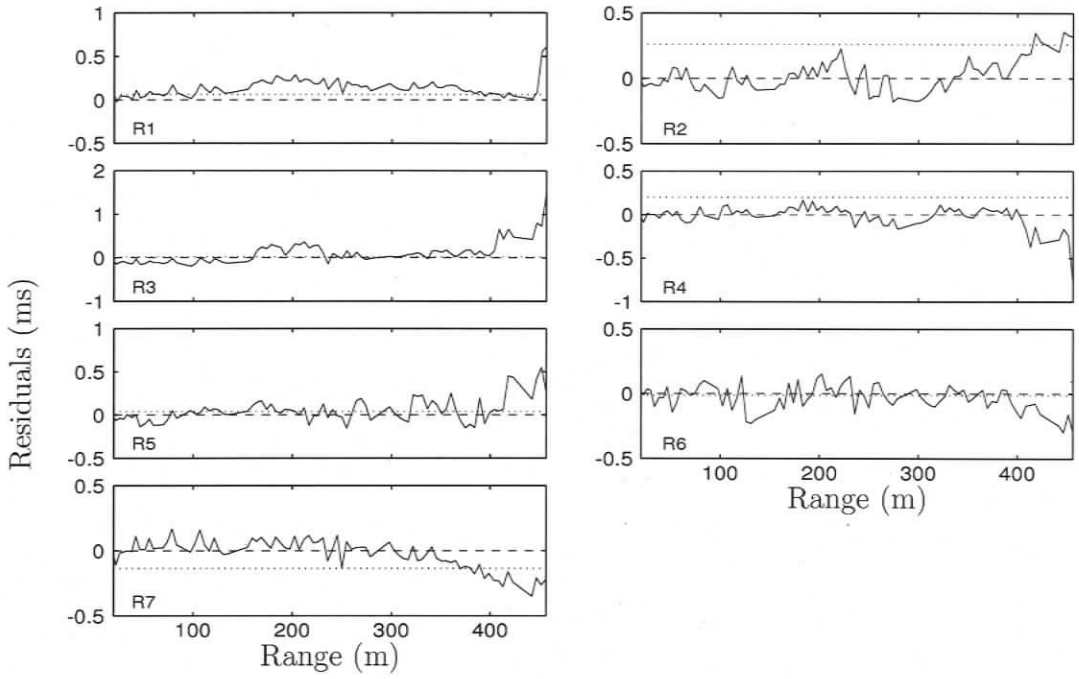

Fig. 7.5: Data residuals for picked and MAP replica data for 7 reflectors. Dotted lines indicate the recovered time offset parameters.

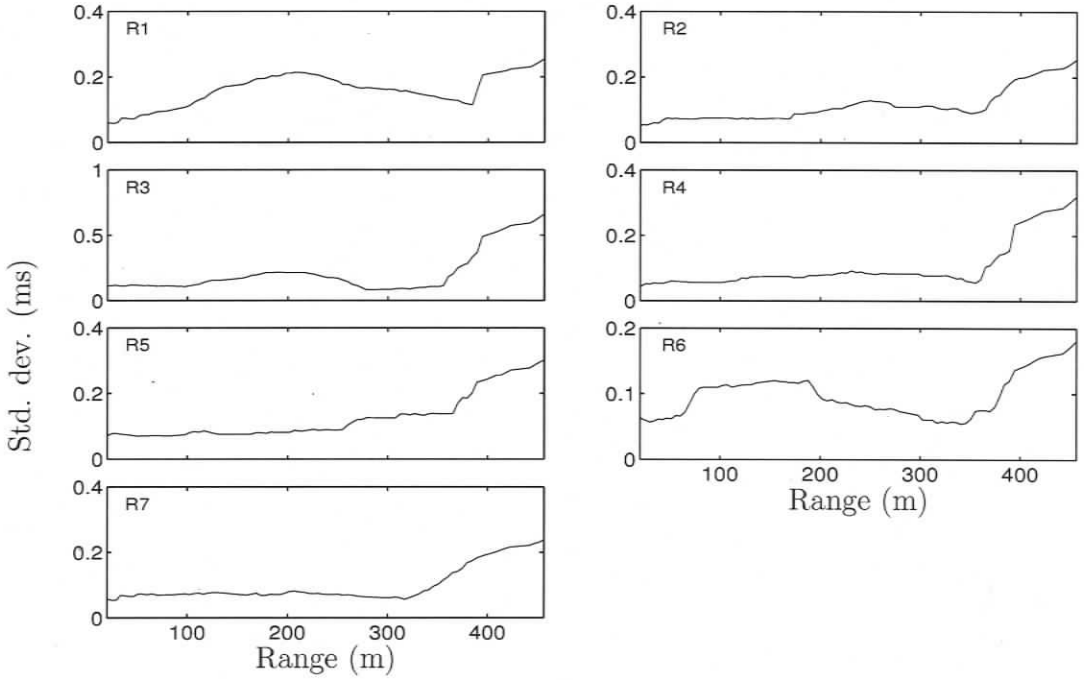

Fig. 7.6: Estimate of the standard deviation with 20 point running RMS filter. Significant changes in standard deviation across range indicate non-stationarity of the residuals.

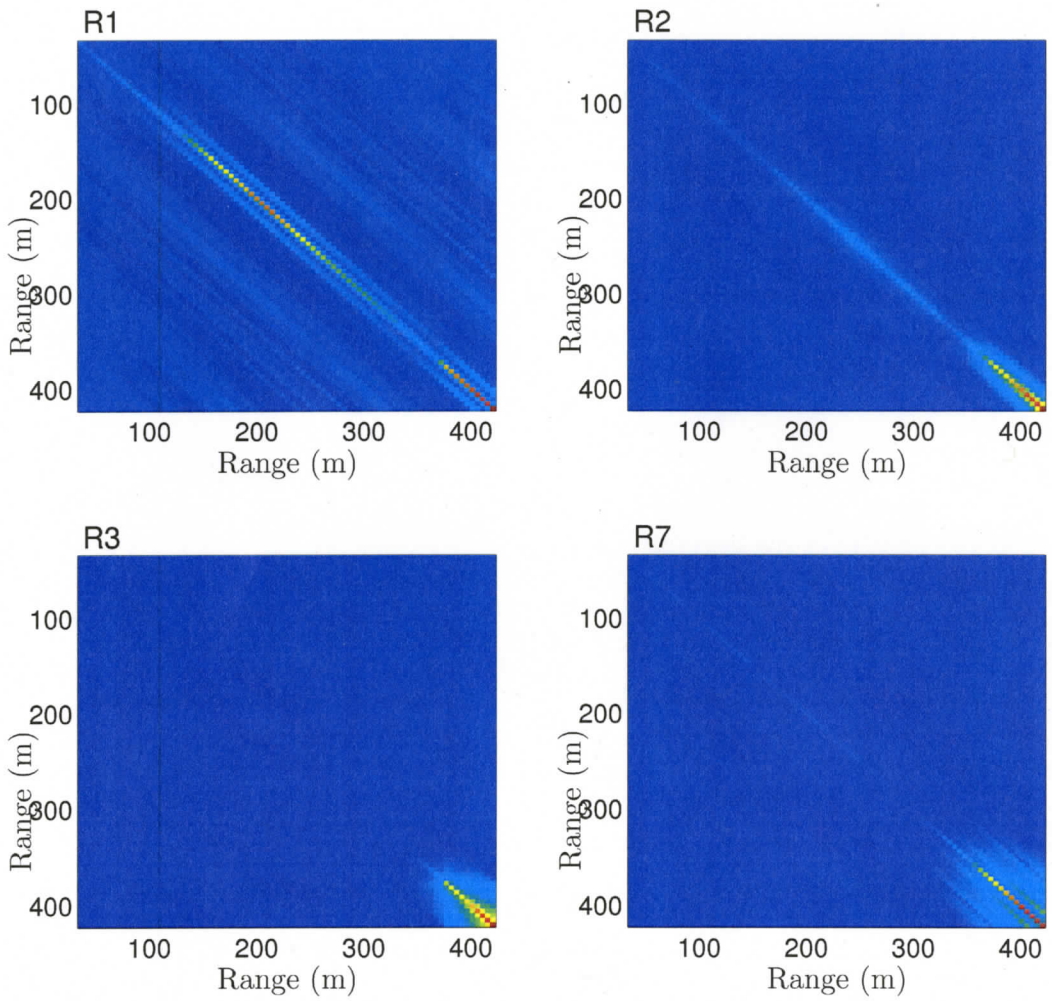

Fig. 7.7: Selected travel time data covariance matrix estimates for reflectors one, two, three, and seven.

resolved within their bounds. The numerical values for 95% HPD credibility intervals and the MAP model for the physical parameters are given in Table 7.1. Finally, Table 7.2 gives the numerical values for the total width (100% HPD credibility intervals) of the marginal distributions. The values for layer thicknesses from Table 7.2 will be used as prior information in the reflection coefficient inversion in the following section. The sound speed information in Table 7.2 is not significantly different from the prior bounds in Table 7.1 and is not applied in the frequency domain inversion.

The recovered marginal probability distributions for the reflector time offset parameters are shown in Fig. 7.9. The prior bounds for the pick offsets were set arbitrarily, depending on the confidence level in the picks for a certain reflector. The first six reflectors were assigned interval of $[-0.75, 0.75]$ ms. For the seventh reflector, a tighter interval of $[-0.3, 0.3]$ was chosen, since the arrival was well separated from other events and the picking was comparably straightforward. The marginal probability distributions indicate good resolution for all offsets, although the offset last reflector is constrained by the prior bounds.

The model parameter correlation matrix for both physical and nuisance parameters is given in Fig. 7.10 and shows strong correlations between numerous parameters. The seabed parameters show several strong correlations, which explain the high computational times required for this inversion. Figure 7.11 shows the joint marginal probability distributions for the six strongest correlations of seabed parameters. It can be seen that the layer thicknesses of several adjacent layers are strongly correlated (negative correlation). This is due to the fact that the negative correlation conserves the total layer thickness of the two layers. The same is true for sound velocities of adjacent layers, as can be seen in Fig. 7.11. The strongest cross correlations, however, are between physical parameters and the picking offset nuisance parameters, as shown in Fig. 7.12. The two largest cross correlations are between h_1 and a_2 (0.96) and between c_2 and a_3 (-0.75). These pronounced correlations are due to the strong coupling between a time offset parameter and a layer thickness or sound velocity. The strong correlations are mainly responsible for widening the marginal distributions for layer thicknesses and sound velocities after including the time offsets as nuisance parameters. Hence, including the time offsets is considered crucial to obtain meaningful uncertainty estimates and avoid overly optimistic (misleading) results.

The strong correlations resulting from the introduction of time offset parameters made the burn-in phase of the Gibbs sampler challenging for this problem. To ef-

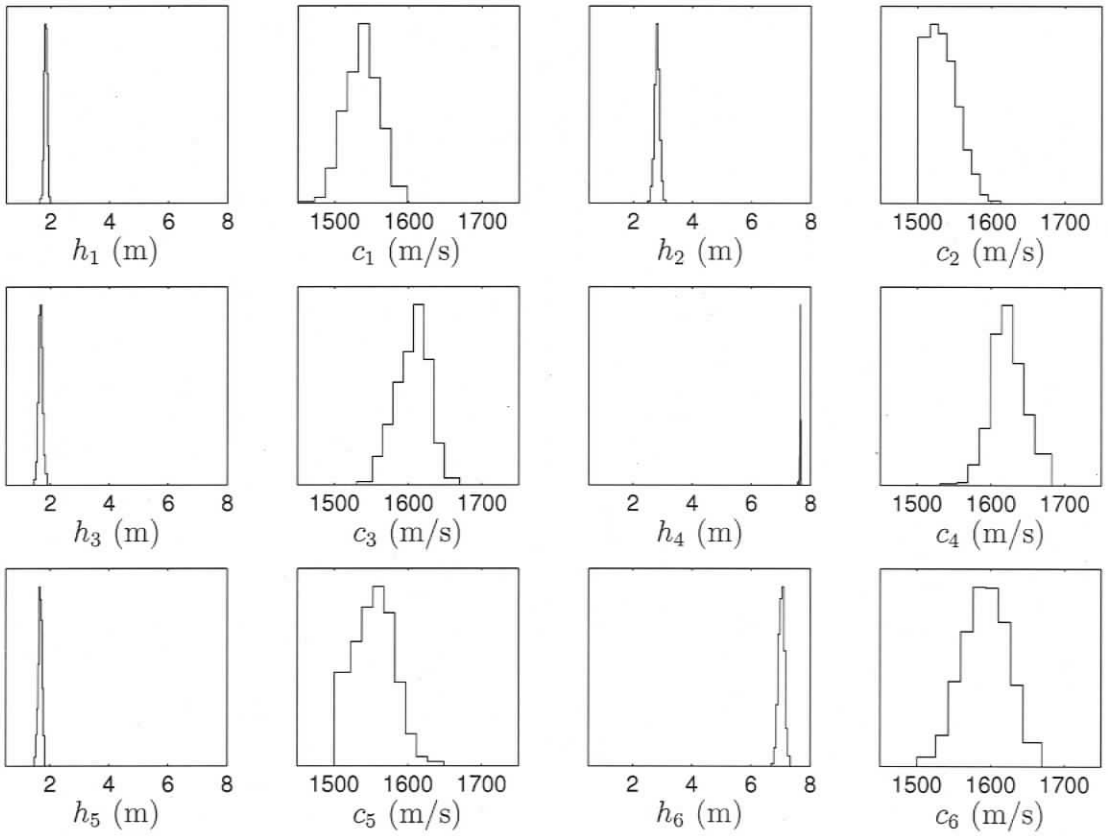

Fig. 7.8: Recovered marginal probability distributions of physical model parameters for the inversion of picked travel time data. The prior bounds for layer thicknesses h_i correspond to the width of the plots.

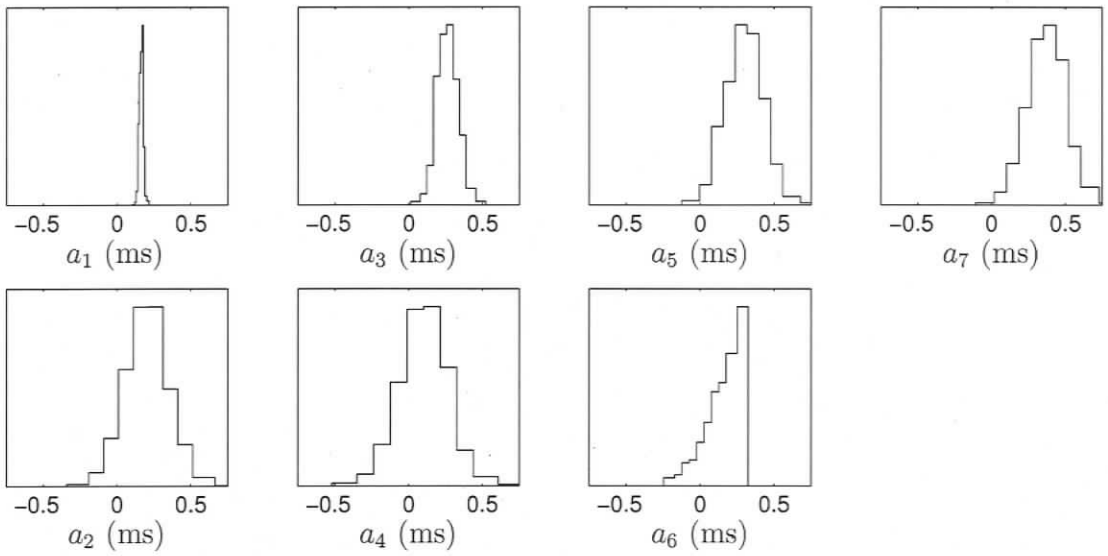

Fig. 7.9: Recovered marginal probability distributions of picking offset parameters of 7 reflectors for the inversion of picked travel time data.

Layer	h (m)			c (m/s)		
	Prior	MAP	95% HPD	Prior	MAP	95% HPD
1	0.10–8.00	1.79	1.63–1.97	1450–1600	1530	1457–1595
2	0.10–8.00	2.80	2.51–3.05	1500–1600	1508	1500–1596
3	0.10–8.00	1.66	1.45–1.87	1500–1700	1581	1537–1662
4	0.10–8.00	7.65	7.56–7.67	1500–1700	1637	1537–1679
5	0.10–8.00	1.61	1.47–1.82	1500–1650	1556	1500–1628
6	0.10–8.00	7.06	6.68–7.28	1500–1700	1569	1500–1664

Table 7.1: Numerical values of prior information and inversion results (MAP model and 95% HPD credibility intervals) for picked travel time data.

fectively sample the complicated parameter space, a linearised parameter covariance matrix estimate was used to compute an initial rotation matrix for the sampling (see Sec. 4.3). The covariance matrix was then regularly updated through the nonlinear sample until convergence of the rotation matrix was obtained according to Sec. 4.3.

It is of fundamental importance to check the quantitative inversion results by examining the assumptions made for the inversion (see Sec. 5.5). For the travel time inversion, the misfit function was derived from a likelihood for Gaussian errors with data covariance $\mathbf{C}^{(\mathbf{d})}$. Therefore, the standardised residuals $(\mathbf{C}^{(\mathbf{d})})^{-1/2}(\mathbf{d}(\mathbf{m}) - \mathbf{d})$ should resemble an uncorrelated Gaussian process and statistical tests should show

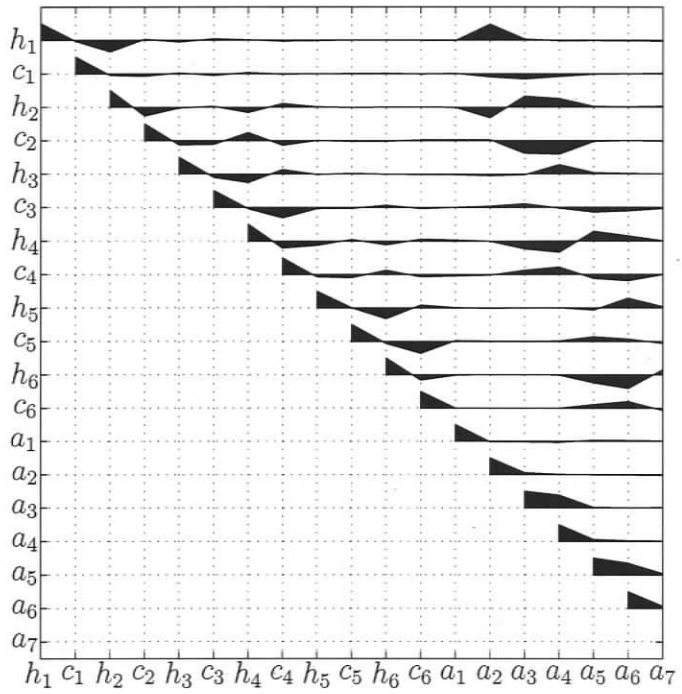

Fig. 7.10: Model parameter correlation matrix. For simplicity, only the upper triangular matrix of the symmetric matrix is plotted.

an improvement over the results for the raw residuals (scaled only by their standard deviation). To examine the issue of randomness, Fig. 7.13 shows the autocovariance function of the data residuals (raw and standardised) for the seven reflectors. Figure 7.13 (a) shows the autocovariance functions for the raw residuals and (b) shows for the standardised residuals. It can be seen that the centre peak at lag zero for the standardised residuals is considerably narrower than for the raw residuals. This indicates that the data covariance matrix estimates reduce data covariances. To quantify

Layer	h (m)	c (m/s)
1	1.63–1.98	1450–1598
2	2.47–3.09	1500–1612
3	1.44–1.90	1530–1669
4	7.56–7.67	1531–1682
5	1.46–1.82	1500–1649
6	6.68–7.33	1500–1669

Table 7.2: Marginal probability distribution width (100 % HPD) recovered from travel time inversion.

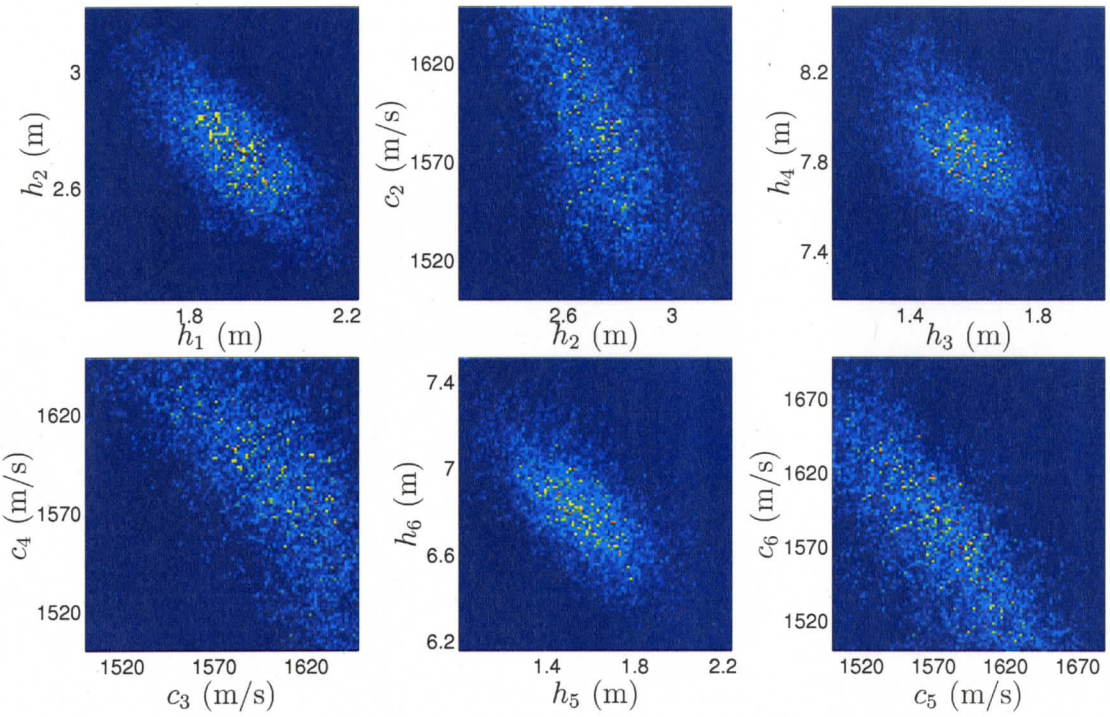

Fig. 7.11: Selected joint marginal distributions for physical model parameters.

this result, a runs test was performed and the resulting p-values are shown in Fig. 7.14. The raw residuals fail the runs test for every reflector with p-values $\sim 10^{-4}$. After applying the data covariance matrix estimate, five out of seven reflectors pass the runs test at a 0.05 level.

To examine the Gaussianity of the residuals, Fig. 7.15 shows histograms of raw and standardised residuals. In addition, the theoretical Gaussian distribution is included on each plot. It can be seen that the data residuals are reasonably Gaussian for most reflectors (with the notable exception of R1). To quantify the Gaussianity of the residuals, a KS test was performed. The cumulative probability distributions that form the base of this test are shown in Fig. 7.16. The p-values were evaluated from a table of critical values that was computed for the appropriate number of data (see Sec. 5.5.2). The p-values for the test are shown in Fig 7.17. It can be seen that applying the covariance matrix does not significantly change the Gaussianity of the residuals. For the standardised residuals, four out of the seven reflectors pass at an 0.05 level and reflector five almost passes with a p-value of 0.035. Reflector R1 strongly failed the validation process. This is surprising, since R1 (the water-

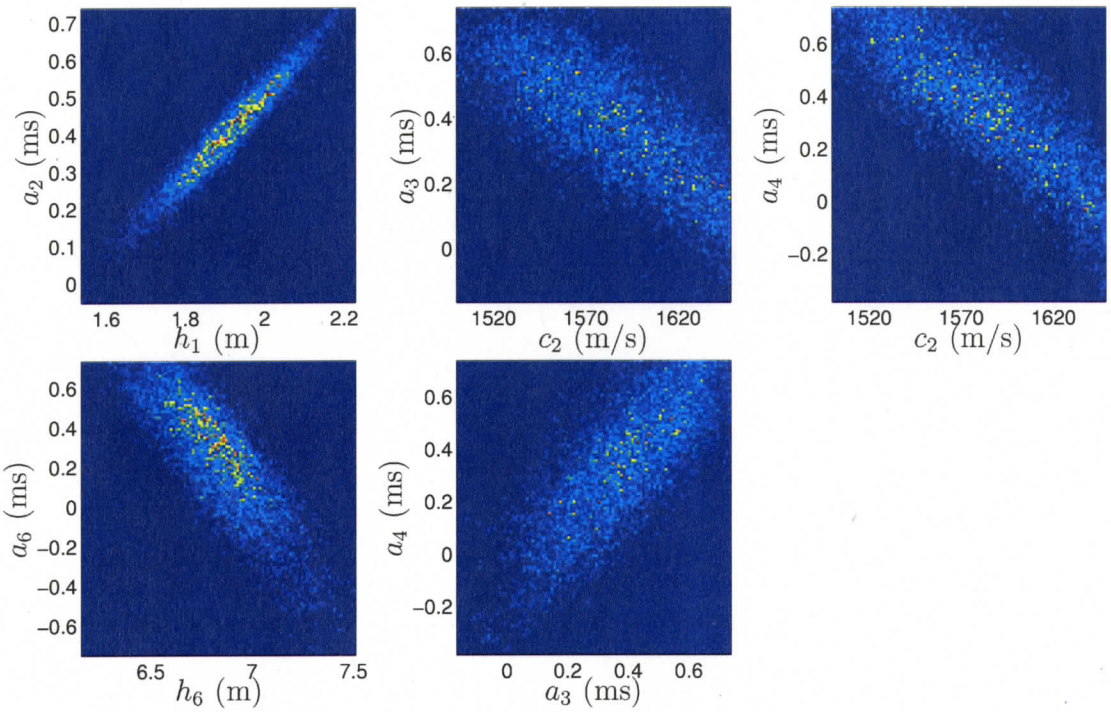

Fig. 7.12: Selected joint marginal distributions between physical parameters (h_i and c_i) and nuisance parameters (constant time picking offsets, a_i).

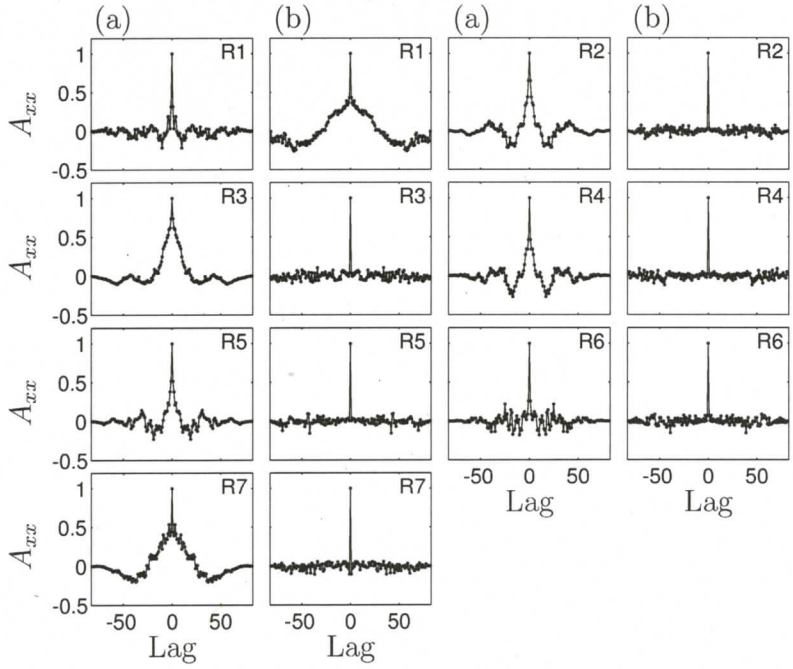

Fig. 7.13: Autocovariance function for raw (a) and standardised (b) residuals as a qualitative measure of data error correlation lengths.

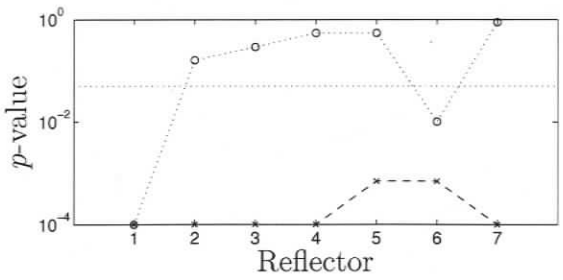

Fig. 7.14: Travel time inversion runs test p-values. The crosses show the p-values for the raw residuals, while the open circles show the p-values for the standardised residuals. The dotted line indicates the 0.05 confidence level.

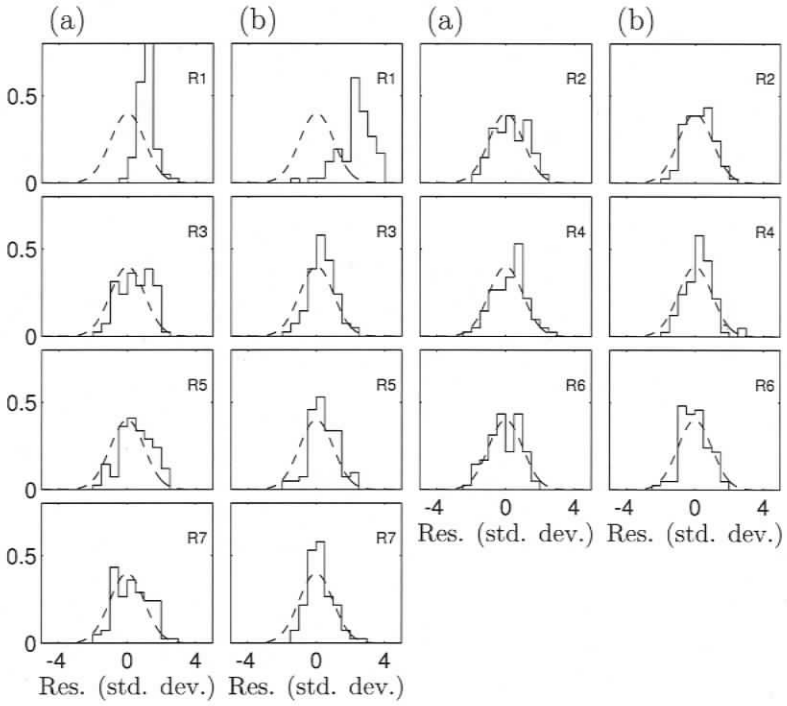

Fig. 7.15: Histograms of raw (a) and standardised (b) residuals as a qualitative measure of Gaussianity of data error.

sediment interface) was particularly clear to pick. However, it should be noted that the inversion evaluates all reflectors at once and that good fit of several reflectors can be at the cost of one particular reflector and is not uncommon.

While the above validation of assumptions is not completely satisfactory in a statistical sense, the tests show that the assumptions appear reasonable and that the inversion likely gives useful estimates and uncertainties for the geoacoustic parameters given the model parameterisation. Hence, the recovered layer thickness information are applied as prior knowledge for the following spherical reflection coefficient inversion in the frequency domain. Further, to be conservative about the prior distributions used in the subsequent frequency domain inversion, 100% HPD credibility intervals

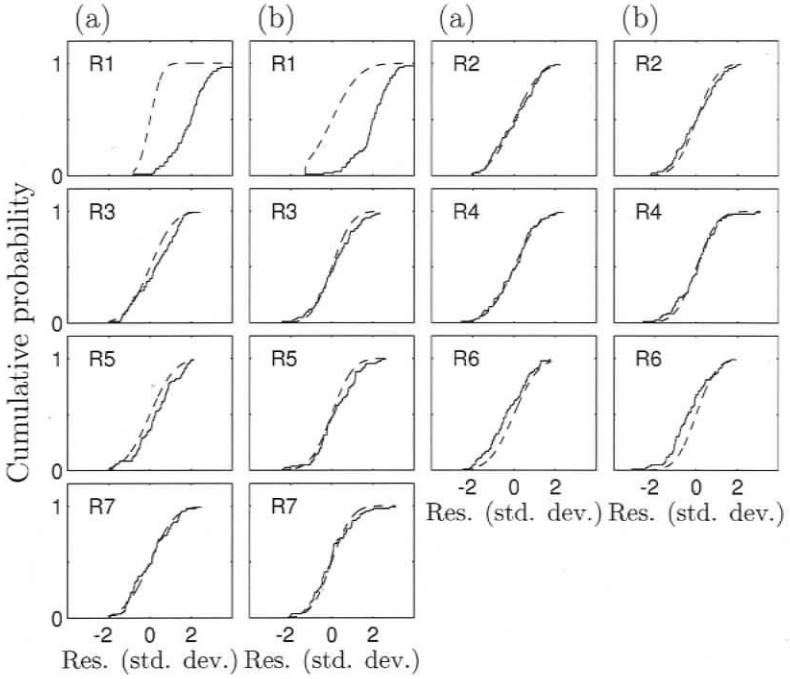

Fig. 7.16: Cumulative probability distributions for (a) raw and (b) standardised residuals. The dashed line represents the theoretical Gaussian distribution.

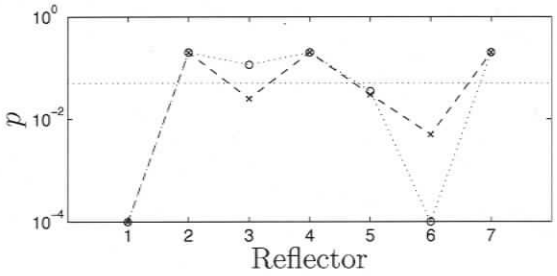

Fig. 7.17: Travel time inversion KS test p-values. The crosses show the p-values for the raw residuals, while the open circles show the p-values for the standardised residuals. The dotted line is the 0.05 level of confidence.

(given in Table 7.2) are used for layer thicknesses. Sound velocities are not further constrained and the same priors as used here will be applied in the reflection coefficient inversion. This conservative approach acknowledges the not perfectly satisfied statistical tests.

7.3. Spherical Wave Reflection Coefficient Inversion

The travel time inversion for the Malta Plateau data set showed that layer thicknesses for six layers can be resolved. Hence, in the frequency domain, six sediment layers over a sediment halfspace are inverted for. The model consists of layers defined by thickness, sound velocity, density and attenuation. To avoid ambiguity in the inversion, a Bayesian layer stripping approach (see Sec. 6.6) is applied to recover sound velocity and density profiles. The attenuation is regarded as an effective measure of the overall loss processes (including scattering and environmental inhomogeneities) not the intrinsic attenuation of the sediment (Dosso and Holland, 2006). Therefore, recovering attenuation profiles is not the focus of this inversion. Figure 7.18 shows the seismo-acoustic data and the packet boundaries that were used in the inversion, including the time window for the direct arrival. The layer stripping approach also helps keep the numerical effort in a manageable frame. Due to the large number of parameters (27 parameters for 6 layers and a halfspace) and the computationally intensive spherical wave forward model, the ASSA optimisation is only feasible when used in a layer stripping approach. However, due to the very coarse granularity of FGS (see Sec. 4.4), it does not have the same computational limitations.

The first layer packet is time windowed to about 3 m depth below the seafloor. The second packet contains the response from the seafloor to about 8 m depth and contains the effects of the uppermost 3 layers which were identified in the travel time inversion. The final packet includes the effects of all 6 layers that were identified in the travel time inversion. These three time windows were then processed according to Sec. 6.3, to yield the reflection coefficient data in five frequency bands from 400–1600 Hz shown in Fig. 7.19. In the following, each frequency represents the centre of a Gaussian frequency average (see Appendix 6.5) with a fractional bandwidth of 1/20. This bandwidth was found to be appropriate for this data set as it retained most structure in the signal and reduced noise significantly. For the computation of replica data, the frequency average is replaced by an equivalent range average as

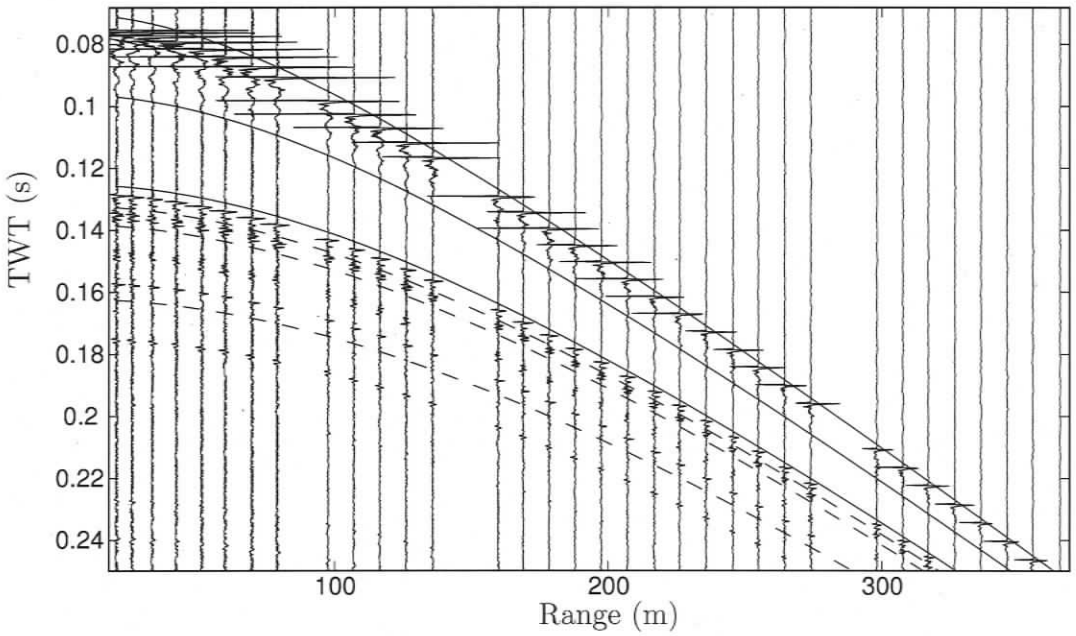

Fig. 7.18: Time windows for the layer packet stripping processing. The uppermost two black lines window the direct arrival. The lower limits of the three basement windows are shown as dashed lines.

discussed in Sec. 6.5 to save computational time. The data are also interpolated onto a uniform spacing in angle with 130 data points at each frequency. Data with a signal to noise ratio of less than 6 dB were excluded. Further, interpolated data that fall into the recording gaps are excluded from computing the data misfit. Figure 7.19 shows that the reflection coefficient becomes more complicated with greater depth and that the level generally increases with depth. This behaviour is expected since more energy is reflected upwards for packets of increasing depth, and the additional layers complicate the shape of the reflection coefficient curve due to interference effects. Figure 7.19 also illustrates the general idea of the layer packet stripping approach. The uppermost layer packet has less information content and also less noise. Detail and information content as well as noise increase with increasing depth. Hence, extracting the information should be more straightforward for shallow packets than for the more complicated deeper packets (i.e., the inversion of shallow packets is not as non-unique).

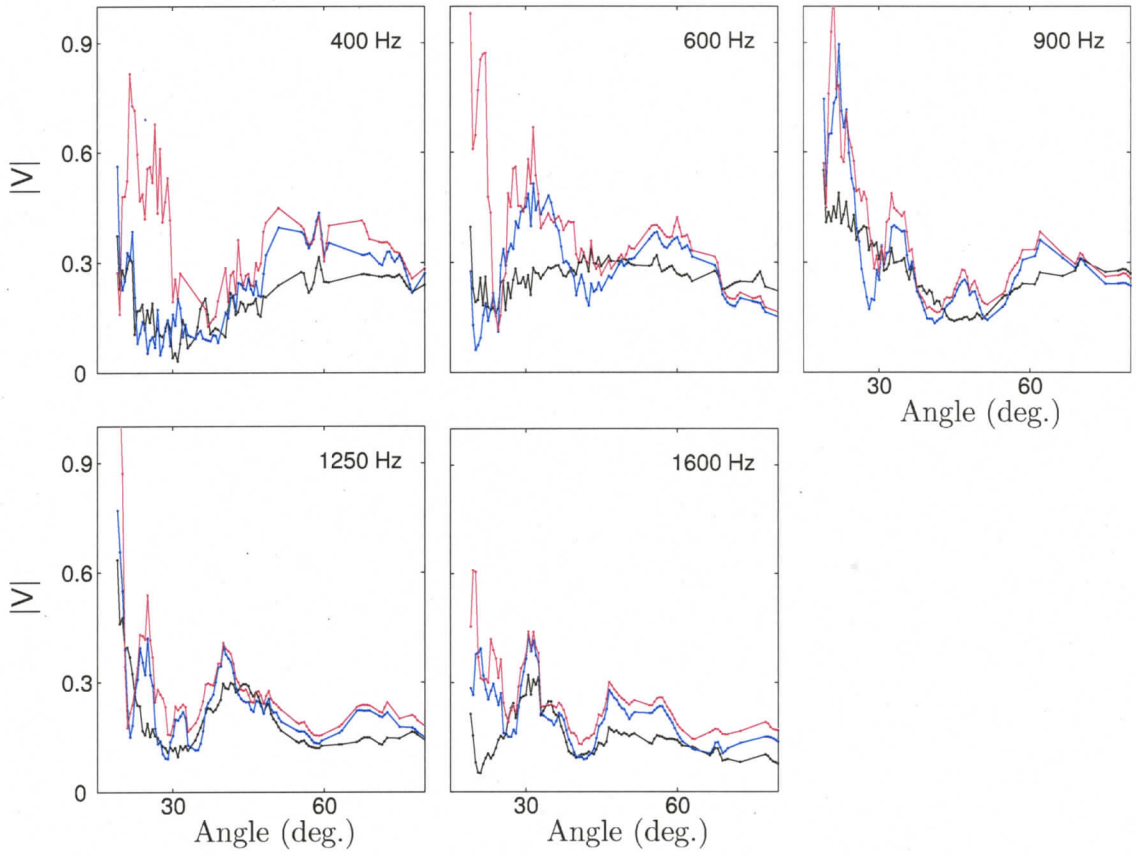

Fig. 7.19: Reflection coefficient data processed into three different packets that reach to increasing depth (packet 1 black, packet 2 blue, packet 3 red).

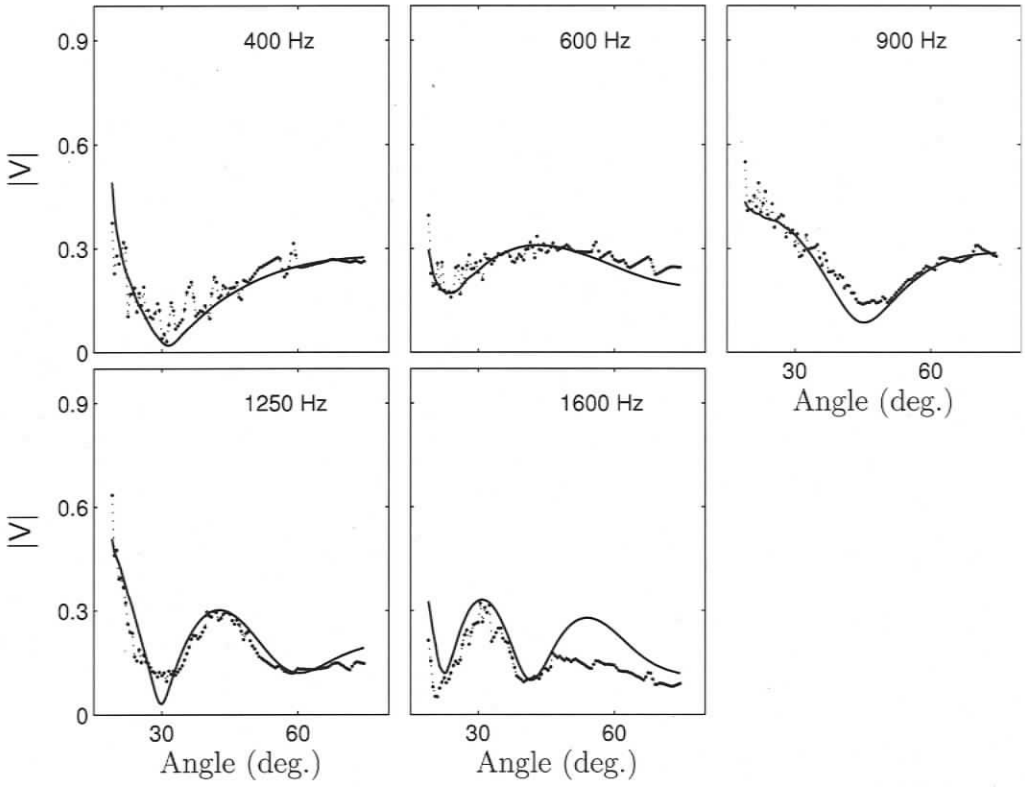

Fig. 7.20: Measured reflection coefficients and MAP model replica (solid line) for the first layer packet consisting of one layer over a halfspace.

7.3.1. First Packet

The inversion was performed in two stages. First, a best fit model was estimated by assuming data errors to be independent from frequency to frequency, and at each frequency to have unknown standard deviation that is stationary across angle. The best fit model was then used to compute replica data as shown in Fig. 7.20 which were then used to estimate data error statistics from the data residuals (Fig. 7.21). To address non-stationarity and correlated errors, a non-Toeplitz data covariance matrix was estimated for each frequency according to Eq. 5.19 (see Fig. 7.22). The data covariance matrices were then applied in sampling the PPD. Finally, the MAP model was extracted from the PPD and used to perform a second iteration of estimating a non-Toeplitz data covariance matrix and applying the second matrix in sampling the final PPD.

Once the final PPD for the first packet was obtained, it was processed to be used as prior information in the inversion for the second packet. In this manner, subse-

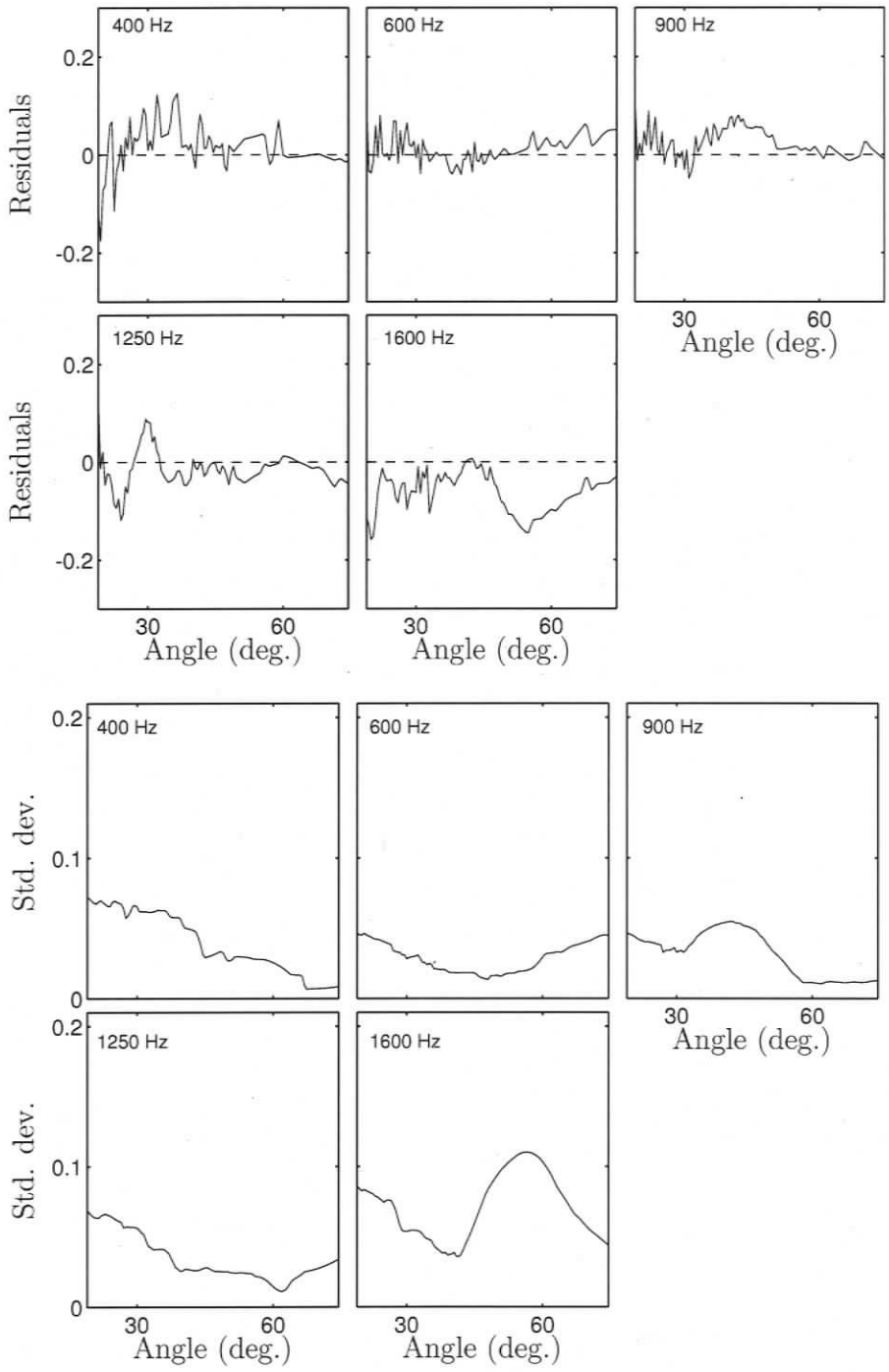

Fig. 7.21: Data residuals (top) and estimated running RMS average standard deviation (bottom) for frequencies from 400 to 1600 Hz, first packet. The data are strongly correlated and non-stationary.

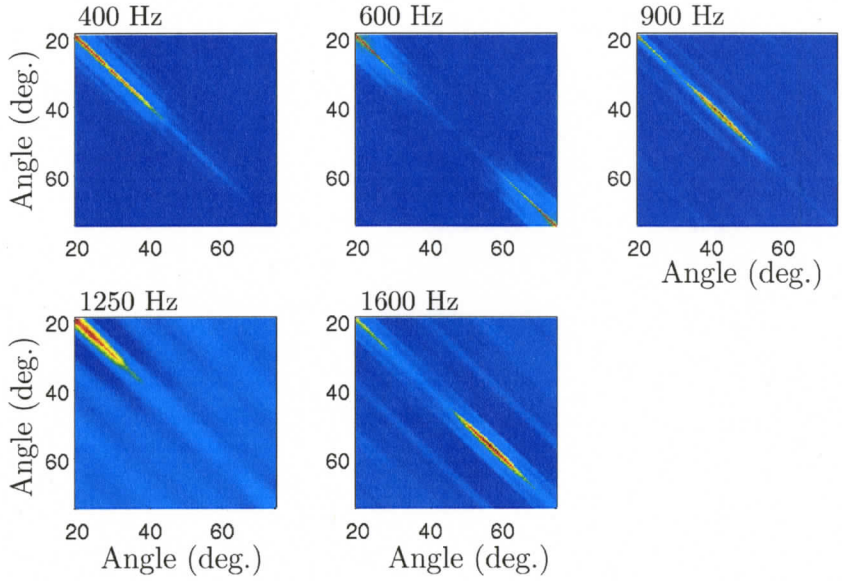

Fig. 7.22: Estimated non-Toeplitz covariance matrices for all frequencies, packet 1.

quent packets always include the layers of the previous packet as well as additional deeper layers (Fig. 7.18). The layers that were included in the previous packet are constrained by the prior knowledge extracted from the PPD. The information about the halfspace is generally discarded since the halfspace is not as well determined as the overlying layers (as there is no travel time information for the halfspace). For this data set, the first packet contained one layer over a halfspace. To determine the prior knowledge for the second packet, the one dimensional marginal probability distributions for the parameters of the first layer were calculated and then rotated into a principal component space according to Eq. 6.24. This ensures that correlations between parameters are taken into account (to first order).

The fit to the data obtained in the inversion for the first packet is shown in Fig. 7.20. In general, the fit to the data are good, particularly at low frequencies. At 1600 Hz, the replica data are offset towards higher reflection coefficient values. Several possible reasons for this were considered. One reason for this could be scattering which is neglected in the forward model. To investigate the effect of scattering, a rough interface with a Gaussian roughness spectrum (Kuperman and Schmidt, 1989; Jensen et al., 1993) was included in the forward model between the first layer and the halfspace of the first packet with an RMS roughness length of 0.05 m. Figure 7.23 shows the

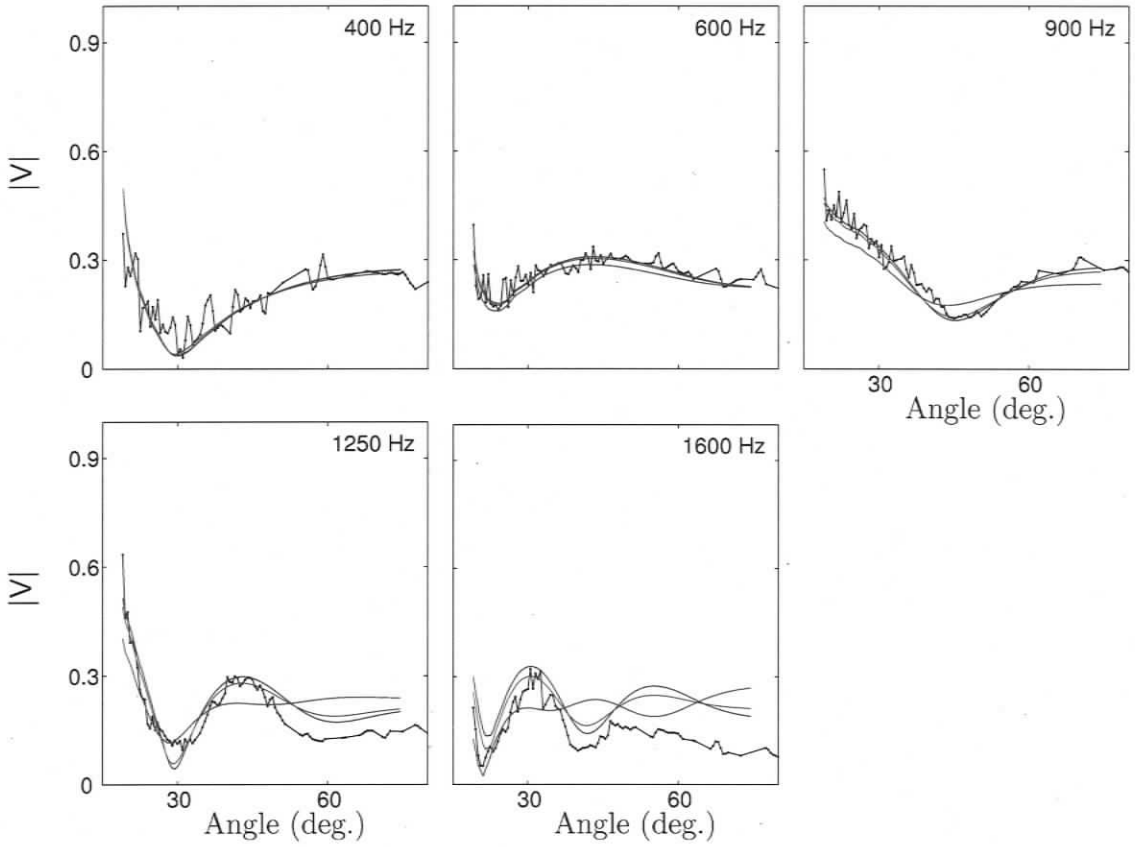

Fig. 7.23: Measured reflection coefficients (black) and MAP model replica for three different values of RMS interface roughness (0.05 blue, 0.1 red, 0.2 green).

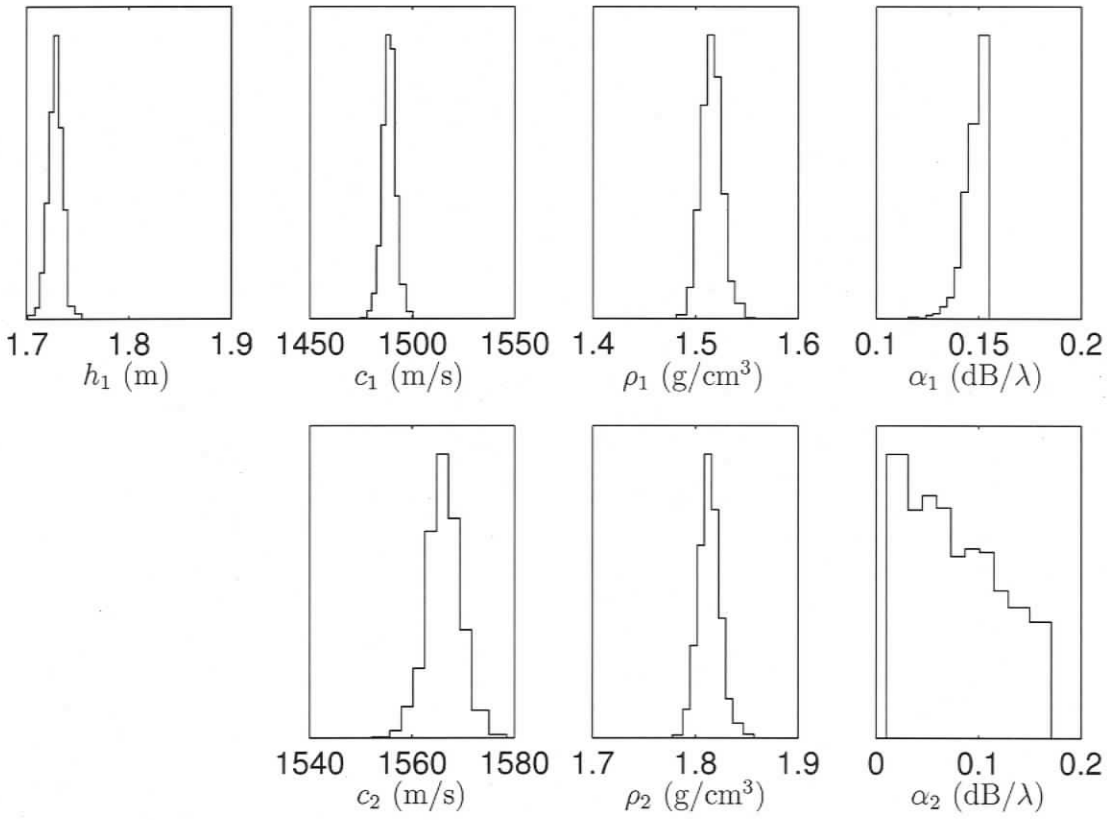

Fig. 7.24: Marginal probability distributions recovered from reflection coefficient inversion of first layer packet.

replica data for the MAP model for this case. With this MAP model, a sensitivity study was performed. Replica fields were calculated by using the MAP model but changing the RMS roughness length as shown in Fig. 7.20. It appears that roughness cannot account for the misfits since the data misfit did not improve significantly. In particular, roughness does not seem to be able to account for the discrepancy in level at high angles at 1600 Hz. Another possible explanation is that gradients in density and sound velocity might be present in the uppermost part of the sediment and might affect the reflection coefficients in a way that is not properly modelled by a single layer. Recovering gradients in multi-layer inversions of reflectivity data are an interesting and challenging problem, but is not pursued for this data set. Chapter 8 considers the problem of velocity and density gradient inversion for a single layer (the transition layer) for data collected over water-saturated fine grained sediments. The reflection data at that site include an angle of intromission, which is a particularly informative feature.

The final result for packet one is presented in terms of marginal probability distributions in Fig. 7.24. The first layer thickness, the sound velocities and densities are resolved well within the prior bounds that are given in Table 7.3. The total widths of the distributions appear to be a reasonable basis for the rotated priors of subsequent inversions.

To validate the assumptions made while inverting the data of the first packet, a runs test and a KS test were performed and qualitative plots of residual autocovariances and histograms were examined at each frequency. Figure 7.25 shows the normalised autocovariance function of the raw and standardised data residuals. It is evident that the centre peak is considerably narrower for the standardised residuals. Figure 7.26 quantifies this result in terms of p-values of the runs test. The raw residuals fail the test at an 0.05 level for each frequency, while the standardised residuals pass at three out of five frequencies. The p-values for the remaining two frequencies improve considerably by applying the data covariance matrices.

Figure 7.27 shows a qualitative test for Gaussianity by plotting histograms of the residuals superimposed with a theoretical Gaussian distribution. In general the raw and standardised residuals agree to about the same degree with the theoretical curves. Figure 7.28 compares cumulative probability distributions for the residuals to theoretical Gaussian cumulative distributions (the basis for the KS test). The resulting p-values for the KS test are given in Fig. 7.29. The two lowest frequencies pass the

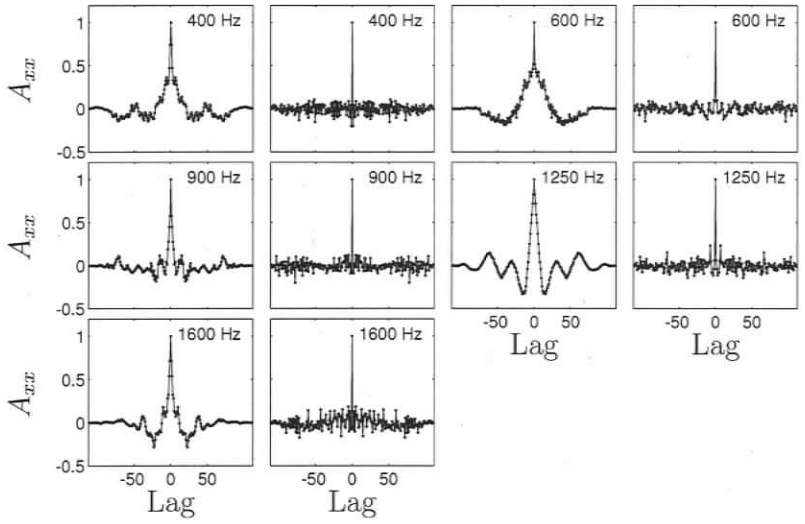

Fig. 7.25: Autocovariance function for raw (columns one and three) and standardised (columns two and four) residuals as a qualitative measure of data error covariances, packet 1.

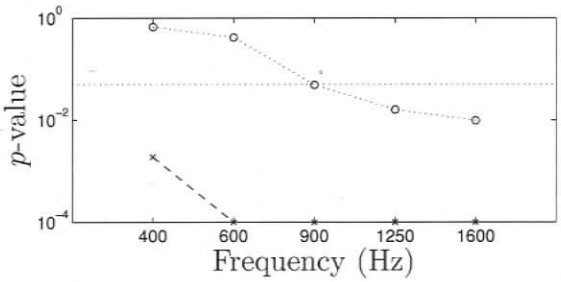

Fig. 7.26: Runs test p-values: The crosses show the p-values for the raw residuals, while open circles show the p-values for the standardised residuals.

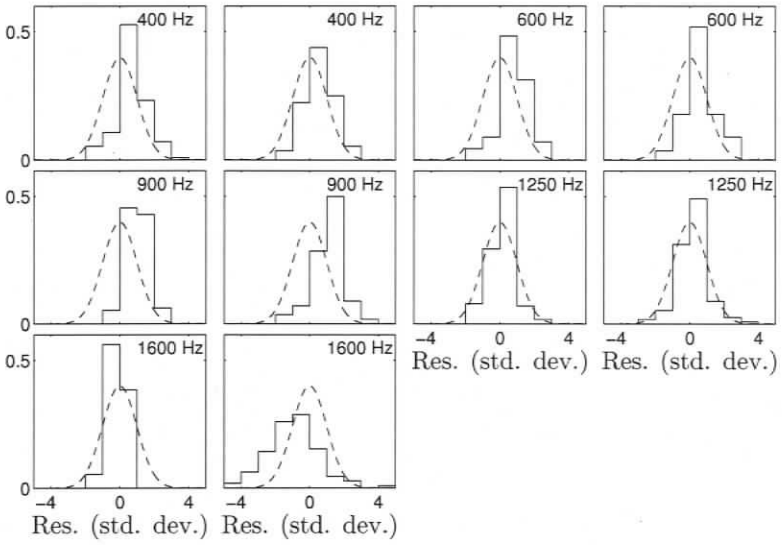

Fig. 7.27: Histograms of raw (columns one and three) and standardised (columns two and four) residuals as a qualitative measure of Gaussianity of data error, packet 1.

test after applying the data covariance matrices; three higher frequencies fail the test with small p-values. This indicates that the data errors are likely not truly Gaussian distributed as assumed. However, the residual histograms in Fig. 7.27 indicate that there is not a substantial difference between the residual distributions and the theoretical Gaussian. In particular, there is no evidence of large outliers in the residuals which can cause major problems in L_2-norm inversions. Hence, the uncertainty distributions derived from this inversion are expected to be reasonably reliable. Further, the residual histograms do not suggest that any other simple distribution (e.g., exponential) would be more appropriate.

7.3.2. Second Packet

In the inversion of the second packet, the parameters of the first layer were constrained using one-dimensional rotated prior distributions from the first packet inversion. In addition to the first layer, two more layers and a halfspace are included in the model, resulting in a total of 15 parameters in the inversion. The data misfit of the MAP model is shown in Fig. 7.30. The general structure of the data is represented well by the replica data and the main features of both data sets agree. At low frequencies (400 and 600 Hz) the fit is not as good as at higher frequencies. This is an indication that the model is better at smaller depth than at greater depths (since lower frequencies

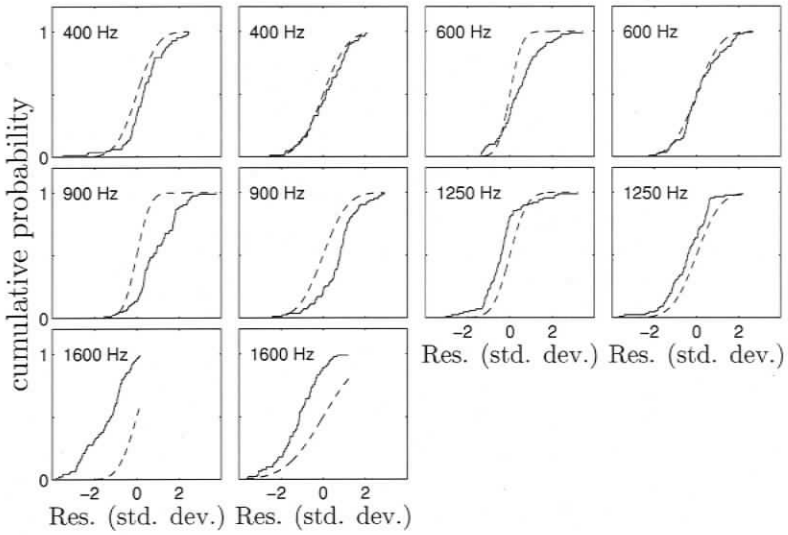

Fig. 7.28: Cumulative probability distributions for raw (columns one and three) and standardised (columns two and four) residuals, packet 1.

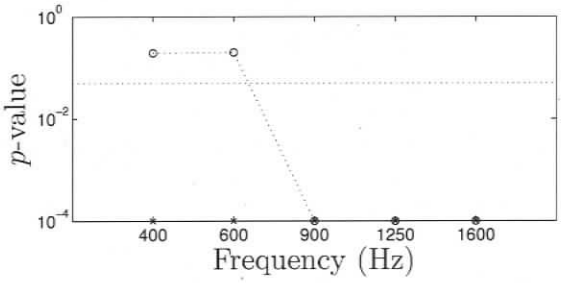

Fig. 7.29: KS test p-values: Crosses show p-values for the raw residuals, while the open circles show the p-values for the standardised residuals.

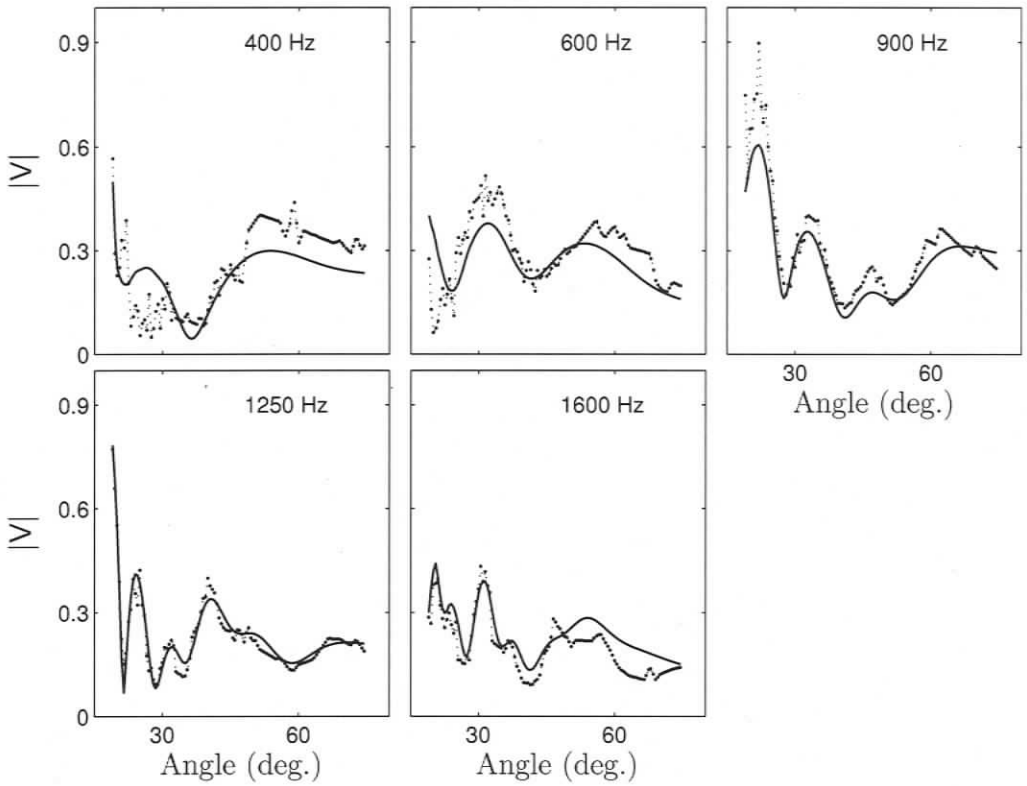

Fig. 7.30: Measured reflection coefficient and MAP model replica (solid line) for the second layer packet consisting of three layers over a halfspace.

are attenuated more slowly with depth). At these low frequencies and large grazing angles, the level of the reflection coefficient is consistently too low in the modelled data.

While the structure (i.e., interference pattern) of the reflection coefficient curves is reasonably well fit, the level, especially at high angles, is not. Since the level is determined by the densities to a large degree, this suggests possible problems with the recovered densities at depth. This can be an indication that the model parameterisation is not entirely correct, as discussed later.

The data errors estimated for the inversion were described by non-Toeplitz covariance matrices that were derived from the data residuals and RMS standard deviations (Fig. 7.31). The data residuals indicate correlated errors as well as errors of different magnitude across angle (non-stationarity). The non-stationarity is clearly illustrated by the RMS average standard deviations and is accounted for by scaling the covariance matrices accordingly (Fig. 7.32).

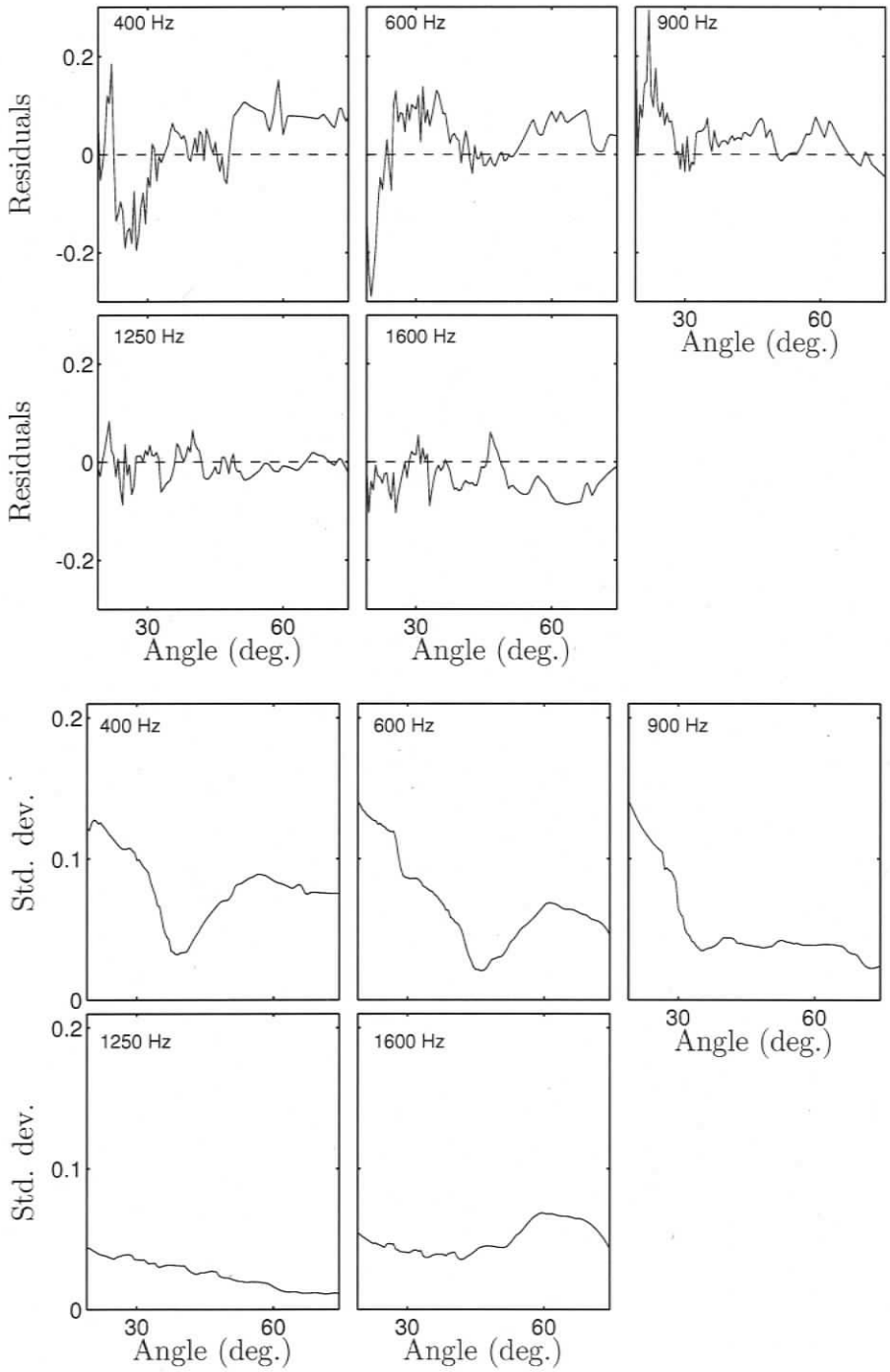

Fig. 7.31: Data residuals (top) and estimated running RMS average standard deviation (bottom) for frequencies from 400 to 1600 Hz, second packet. The data are strongly correlated and non-stationary.

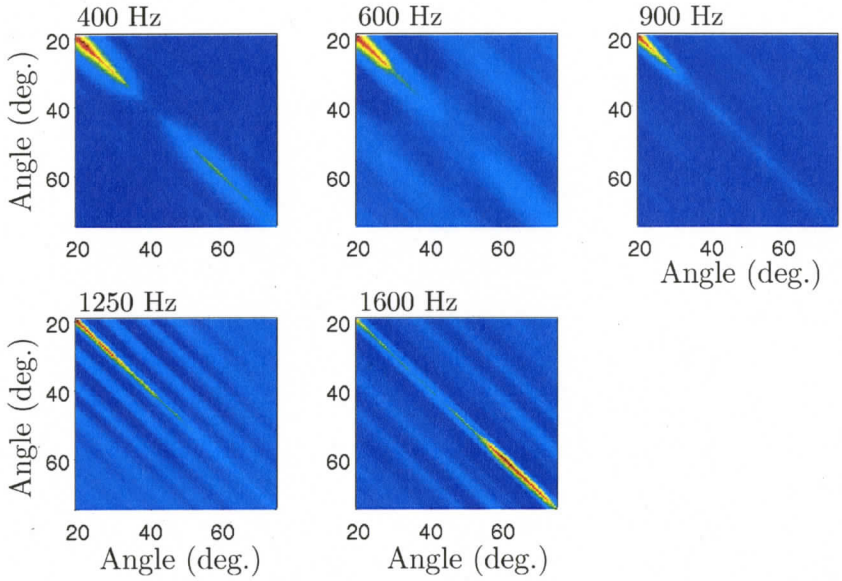

Fig. 7.32: Estimated non-Toeplitz covariance matrices for all frequencies, packet 2.

The marginal probability distributions for packet 2 are shown in Fig. 7.33. The layer thicknesses and sound velocities for all 3 layers and the halfspace are resolved well within the prior bounds (given in Table 7.3). The densities of the first two layers are also well resolved. In the uppermost layer, the density peaks at the upper prior bound. This is due to the constraint from the first layer packet inversion and is not regarded as a problem. However, the third layer and the halfspace show distributions for their densities that peak at very high values that seem to be physically unreasonable. Several other inversions were run with higher upper bounds and showed that these densities continued to increase as the bounds were opened up. It is not considered physically reasonable to allow higher densities than given in Table 7.3 and the inversion results were computed using those values. This problem with the fitting means that the inversion cannot find a good solution for the densities within the expected range of values for typical seabed sediments. Hence, a problem with either the forward model (theory error), the representation of statistical assumptions or with the model parameterisation exists. In the following, the statistical assumptions are validated to ensure proper representation of the data error statistics.

The statistical assumptions were validated (Fig. 7.34–7.38) in the same way as for the first packet. The standardised residuals passed the runs test at all frequen-

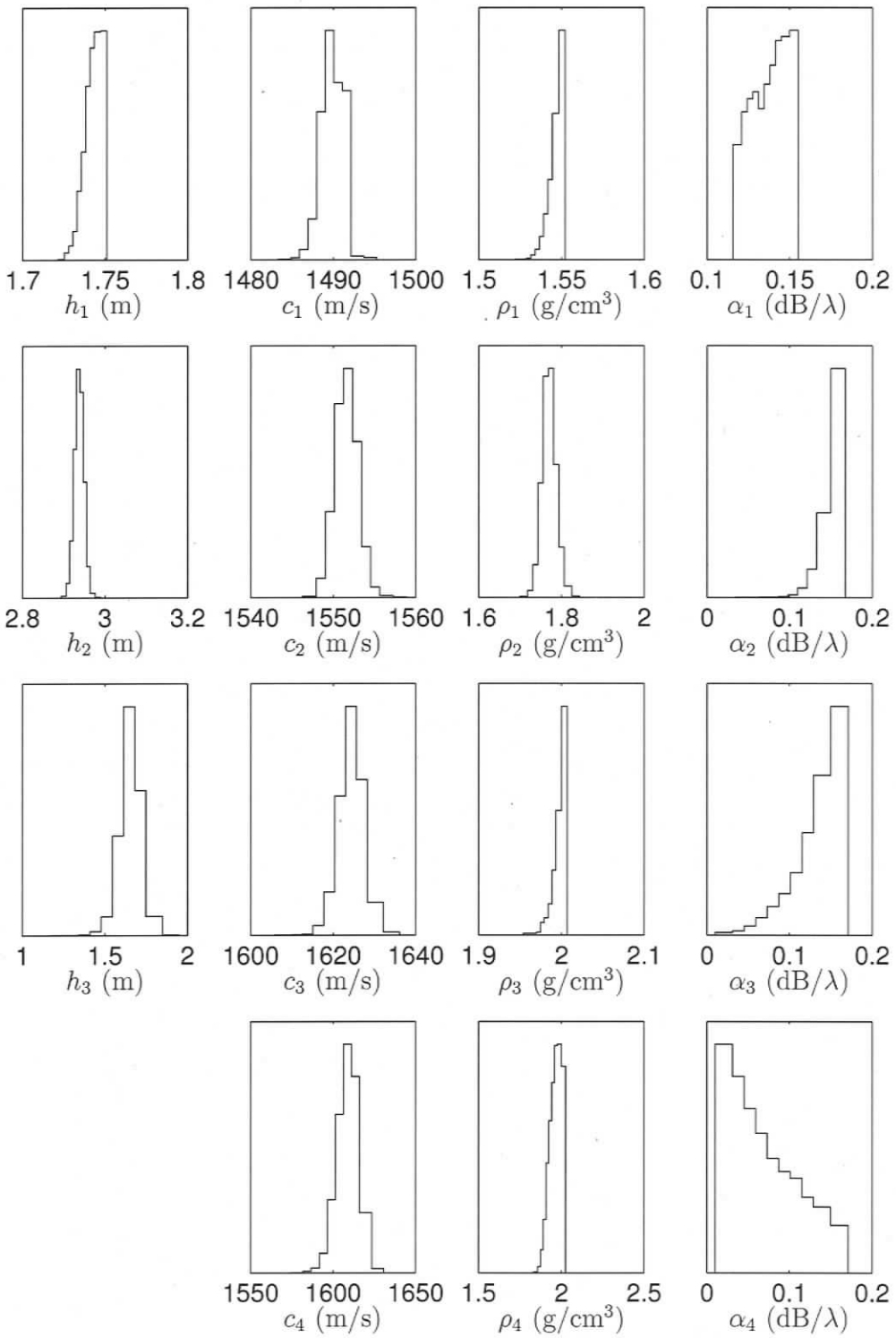

Fig. 7.33: Marginal probability distributions recovered from reflection coefficient inversion of second layer packet.

Layer	h (m)	Sound velocity (m/s)	Density (g/cm^3)	Attenuation (dB/λ)
1	1.53–2.23	1450–1550	1.4–1.8	0.01–0.15
2	2.21–3.22	1500–1600	1.5–2.0	0.01–0.15
3	1.15–2.01	1500–1700	1.5–2.0	0.01–0.15
4	7.18–8.50	1550–1650	1.6–2.0	0.01–0.15
5	1.00–2.25	1650–1750	1.7–2.1	0.01–0.15
6	6.15–7.52	1600–1750	1.7–2.1	0.01–0.15
7		1600–1750	1.7–2.1	0.01–0.15

Table 7.3: Prior bounds for all layers.

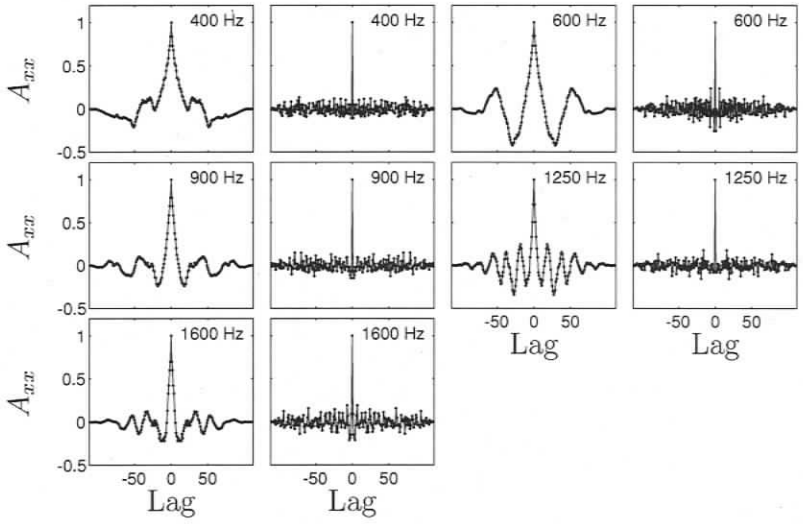

Fig. 7.34: Autocovariance function for raw (columns one and three) and standardised (columns two and four) residuals as a qualitative measure of data error covariances, packet 2.

cies, whereas the raw residuals failed the test at all frequencies at a $p = 0.05$ level (Fig. 7.35). The KS test (Fig. 7.37 and Fig. 7.38) is passed at 2 frequencies for the standardised residuals and a third frequency almost passes, while the raw residuals fail with very small p-values at all frequencies. The residual histograms shown in Fig. 7.36 do not appear strongly non-Gaussian and do not include outliers. It can be concluded that while the covariance matrix estimates do not fully describe the data error statistics, they do improve the inversion. Hence, the underlying statistical assumptions are likely not the source of the problem with the densities noted above. After examining the PPD of the third layer packet, this problem will be further discussed and the model parameterisation examined.

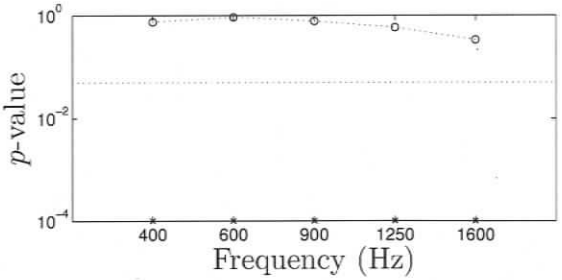

Fig. 7.35: Runs test *p*-values: The crosses show the *p*-values for the raw residuals, while open circles show the *p*-values for the standardised residuals.

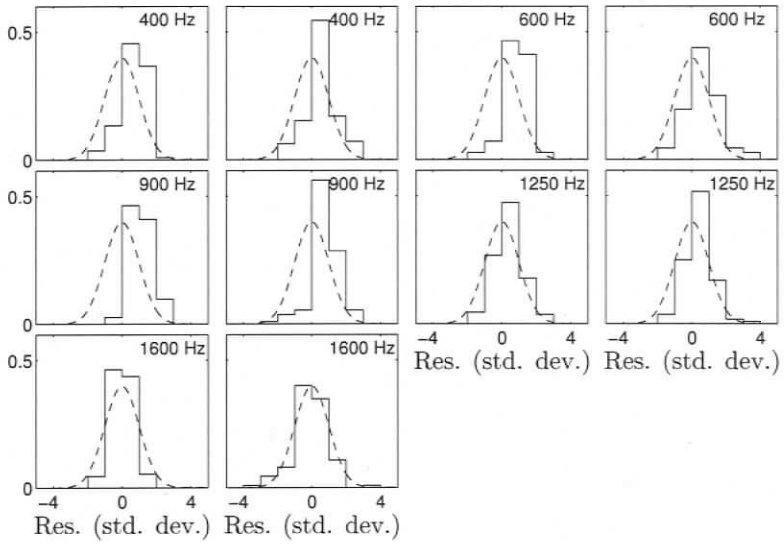

Fig. 7.36: Histograms of raw (columns one and three) and standardised (columns two and four) residuals as a qualitative measure of Gaussianity of data error, packet 2.

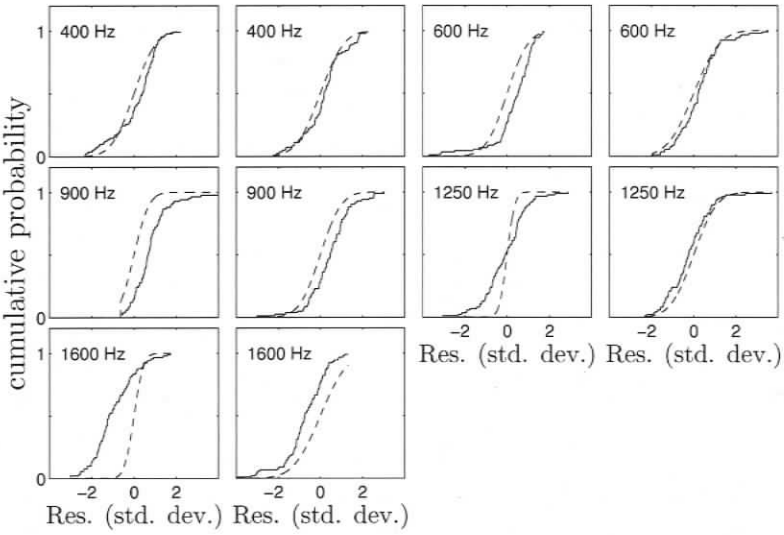

Fig. 7.37: Cumulative probability distributions for raw (columns one and three) and standardised (columns two and four) residuals, packet 2.

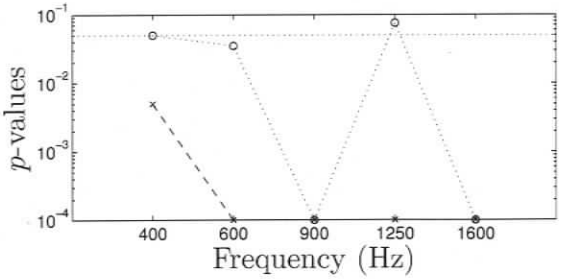

Fig. 7.38: KS test p-values: Crosses show p-values for the raw residuals, while the open circles show the p-values for the standardised residuals.

7.3.3. Third Packet

The third and final packet contained three layers that were constrained by the first and second packet inversion results, as well as three additional layers and a halfspace. Hence, the Gibbs sampling inversion included 27 parameters. Figure 7.39 shows the measured and replica data. The measured reflectivity data show complicated structure as a result of the multiple seabed layers. However, the quality of the data are remarkable and the noise level is low. The general structure of the data are captured well by the MAP model replica with good fit at low angles. Again, the reflectivity level of seems to be biased at high angles. The best fit is at the two highest frequencies (1250 and 1600 Hz).

To investigate the quality of the recovered geoacoustic parameters, the MAP model was used to generate a broad band time series response using the OASES time series model OASP which computes the broad band transfer functions (see Appendix B). These functions were then convoluted with a source wavelet taken from the measured data. The wavelet was obtained by time windowing the direct arrival and then averaging over three traces recorded at close range. The resulting wavelet was then corrected for geometrical spreading and low pass filtered to the frequency band that was used in OASP (100 to 3000 Hz in ~1.5 Hz steps). The results are shown in two way time (TWT) in Fig. 7.40, and in reduced time for $c = 1511$ m/s (Fig. 7.41) according to

$$t_{rd} = \sqrt{t^2 - r^2/c^2}\,, \tag{7.1}$$

where r is the range. Reduced time causes the time axis to spread significantly with range which has to be taken into account when interpreting plots in reduced time. The range step in the modelled data were chosen to offset the traces slightly for better comparison. The modelled data closely resemble the measured data. As expected from the travel time inversion, the first break for each reflector is well captured by

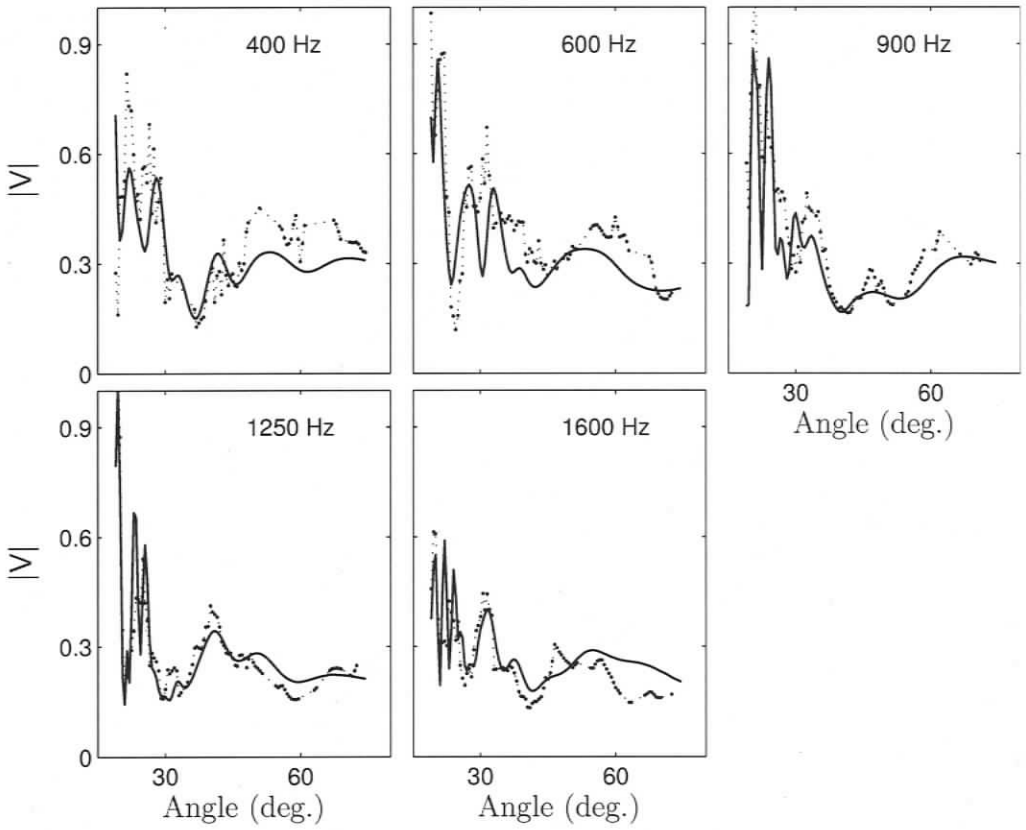

Fig. 7.39: Measured reflection coefficient and MAP model replica (solid line) for the third layer packet consisting of six layers over a halfspace.

the replica. Further, the modelled amplitudes agree well (qualitatively) with the measured data.

Figure 7.42 shows the data residuals and the RMS average standard deviation for the last packet. As before, the errors show correlations across angle as well as non-stationarity. Therefore, a non-Toeplitz covariance matrix estimate (Fig. 7.43) was based on these residuals and standard deviations and then used in the sampling of the PPD.

The marginal distributions of the final PPD for all 27 parameters are shown in Fig. 7.44 and Fig. 7.45. All six layer thicknesses and five of the seven sound velocities are well resolved. In layer five (Fig. 7.45), the sound velocity peaks at 1750 m/s, the upper prior bound, but it is not expected that the actual sound velocity is this high.

Similar problems exist with the densities. The density of layer three peaks at its upper bound as before for the inversion of the second packet. However, the density of layer four (which corresponds to the halfspace in packet two) peaks at a more reasonable value of 1.9 g/cm^3. Below this, the recovered densities jump back and forth between upper and lower bounds which is not considered to be physically meaningful. Rather this seems to be the same problem as was encountered in packet two, but becoming more severe with greater depth, and likely indicates a problem with the model parameterisation. To further consider the problems that existed with the data, the complete frequency domain inversion was carried out for data that were collected for both incoming and outgoing source traces for the two perpendicular tracks (i.e., four data sets). The results were very similar in all cases with the same problems and will not be discussed further here. In the following, the statistical assumptions are examined for this last packet.

Figures 7.46–7.50 show the results of the validation process. The randomness of the residuals is satisfied well, as can be seen from the autocovariance functions as well as from the runs test results (passed by the standardises residuals at four out of five frequencies while the raw residuals failed at all frequencies). The Gaussianity of the standardised residuals is not satisfied as well. The KS test was only passed at 1250 Hz, although the data at 400 and 1600 Hz just fail. The residuals (raw and standardised) at most frequencies seem to be biased to positive values (Fig. 7.48). This is a result of a systematic misfit, especially at high angles. It is possible that problems in the model parameterisation cause a fit to the data that cannot capture the actual environment. This results in residuals that cannot satisfy the statistical assumptions. Therefore,

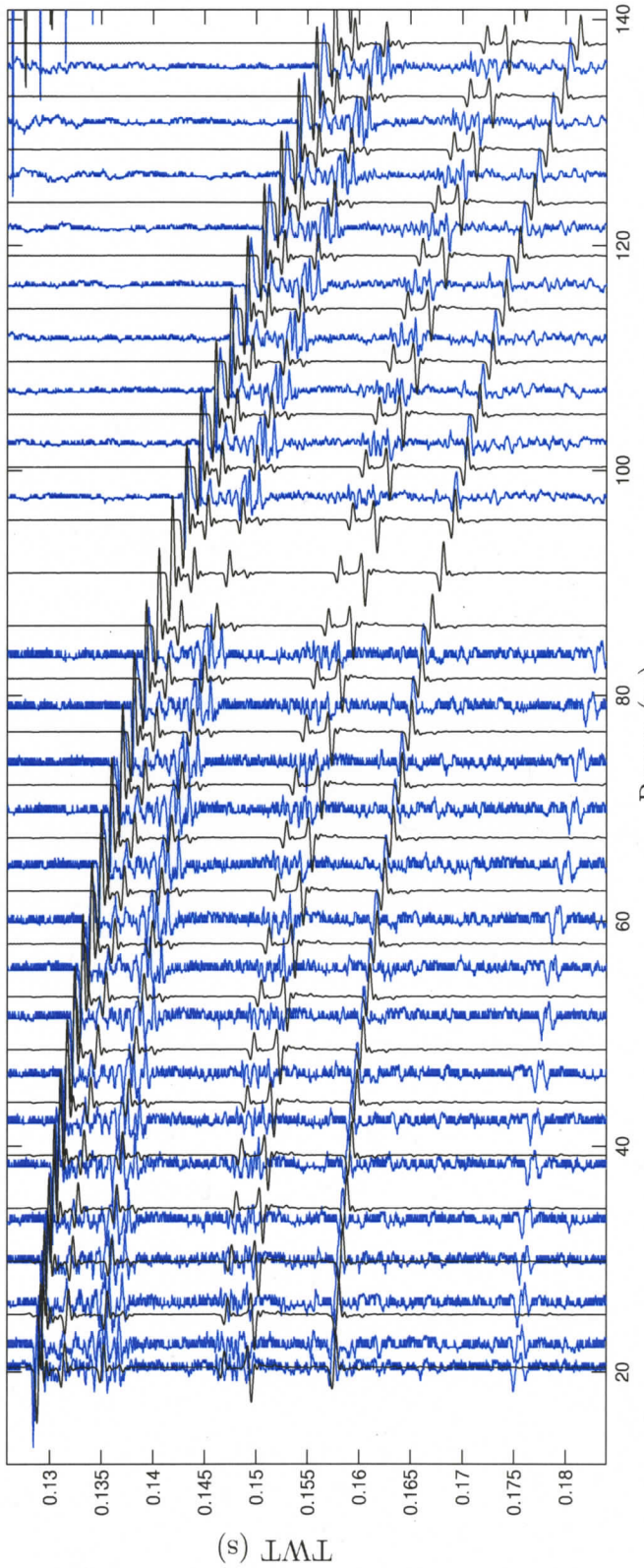

Fig. 7.40: Comparison of measured (blue) and modelled (black) broad band time series (crop shown for better visual impact).

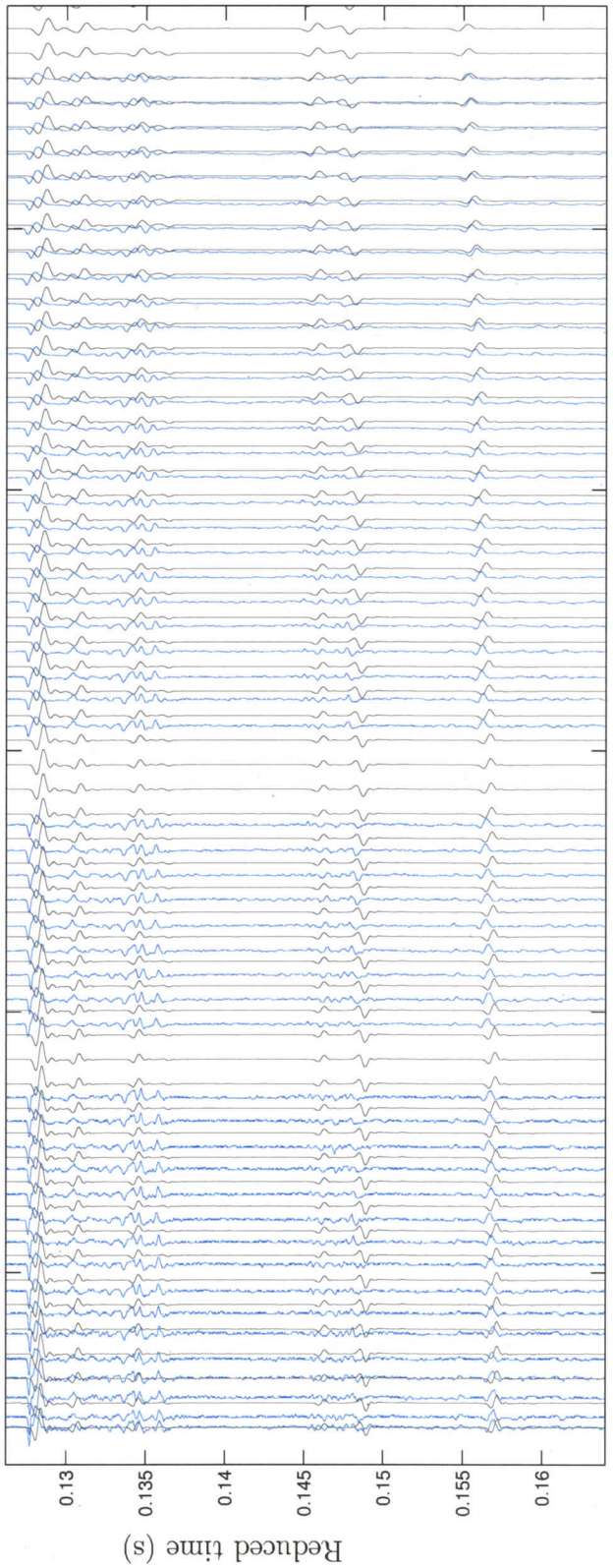

Fig. 7.41: Comparison of measured (blue) and modelled (black) broad band time series data in reduced time.

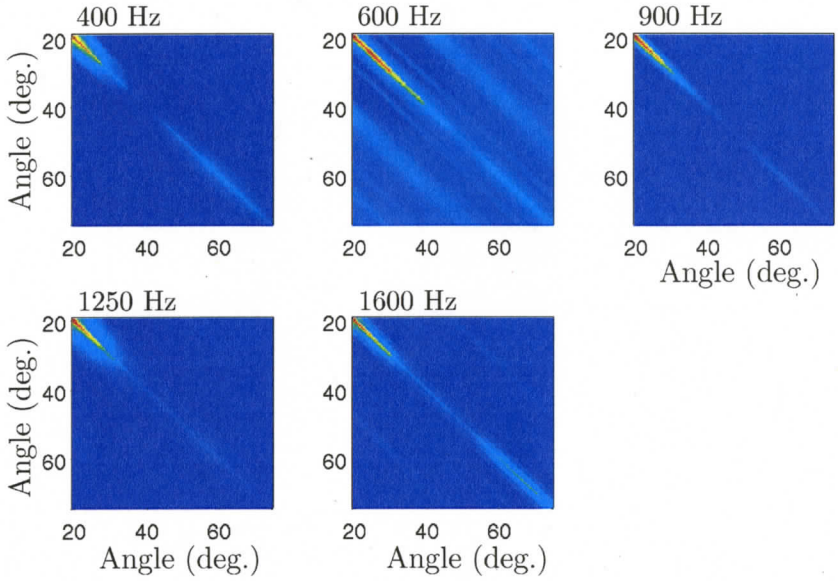

Fig. 7.43: Estimated non-Toeplitz covariance matrices for all frequencies, packet 3.

the model parameterisation is questionable and will be examined in the following section.

7.3.4. Comparison to High Resolution Seismics

After most of the above work was done, a high resolution seismic section that was recorded at the experiment site (perpendicular to the track considered before) became available (Fig. 7.51). The location of the receiver and the approximate depth of the packets are indicated in the figure. It can be seen that below the area where the array was moored, the seabed shows a complicated structure. The depth range included in the second packet shows a layer that pinches out almost directly below the array. To the right of where the layer pinches out, the layering is more complicated than in other areas. This shows that the model of flat lying layers does not appropriately represent the environment at depth. It is also expected that the problems become more significant with greater depth as everything below a certain depth will be affected by the complicated structure.

Because of the problem of a layer pinch out below the receiver, the final inversion results discussed here are limited to the upper parts of the seabed. Figure 7.52 shows the final inversion results for the upper four layers with 95% HPD credibility intervals

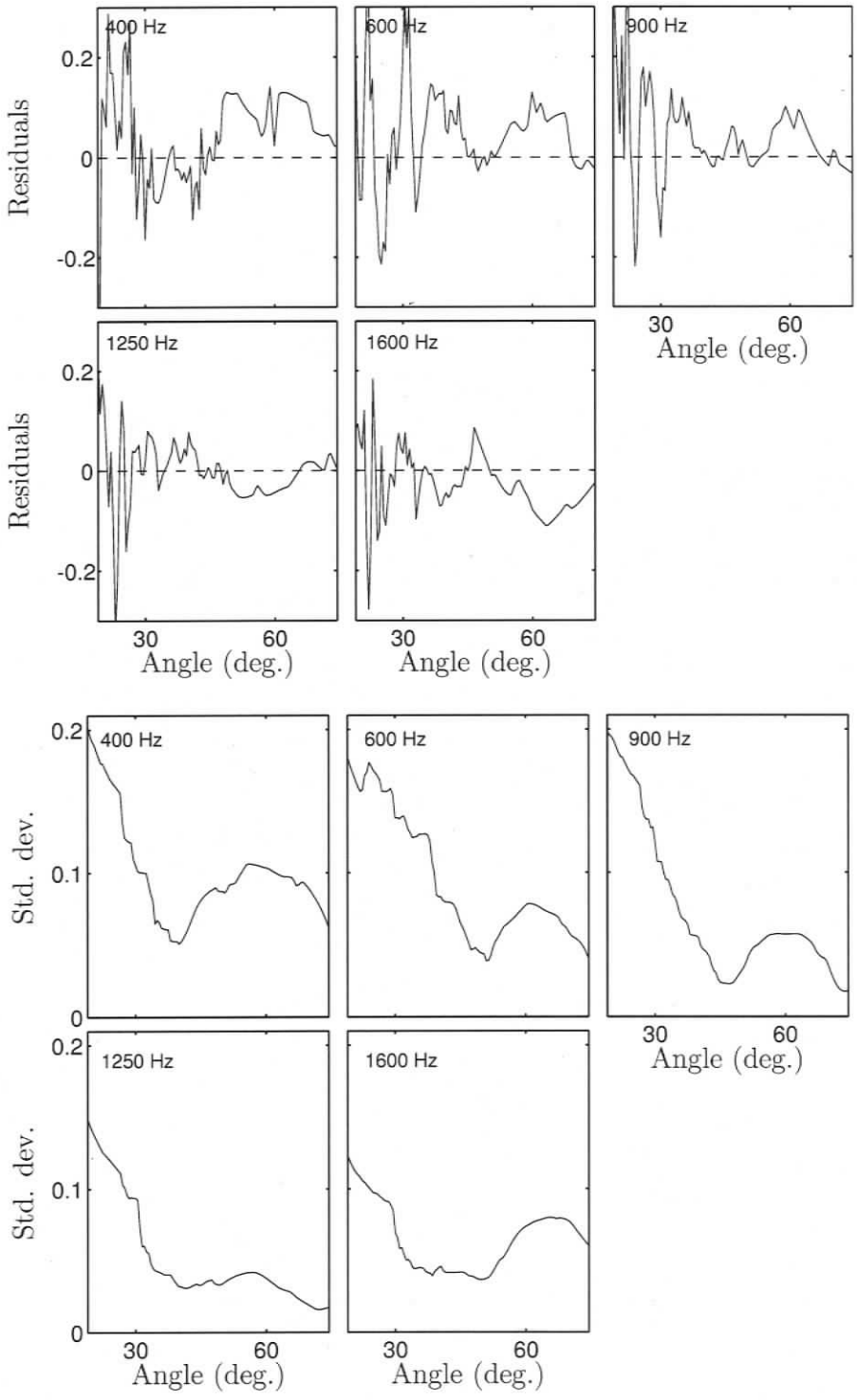

Fig. 7.42: Data residuals (top) and estimated running RMS average standard deviation (bottom) for frequencies from 400 to 1600 Hz, third packet. The data are strongly correlated and non-stationary.

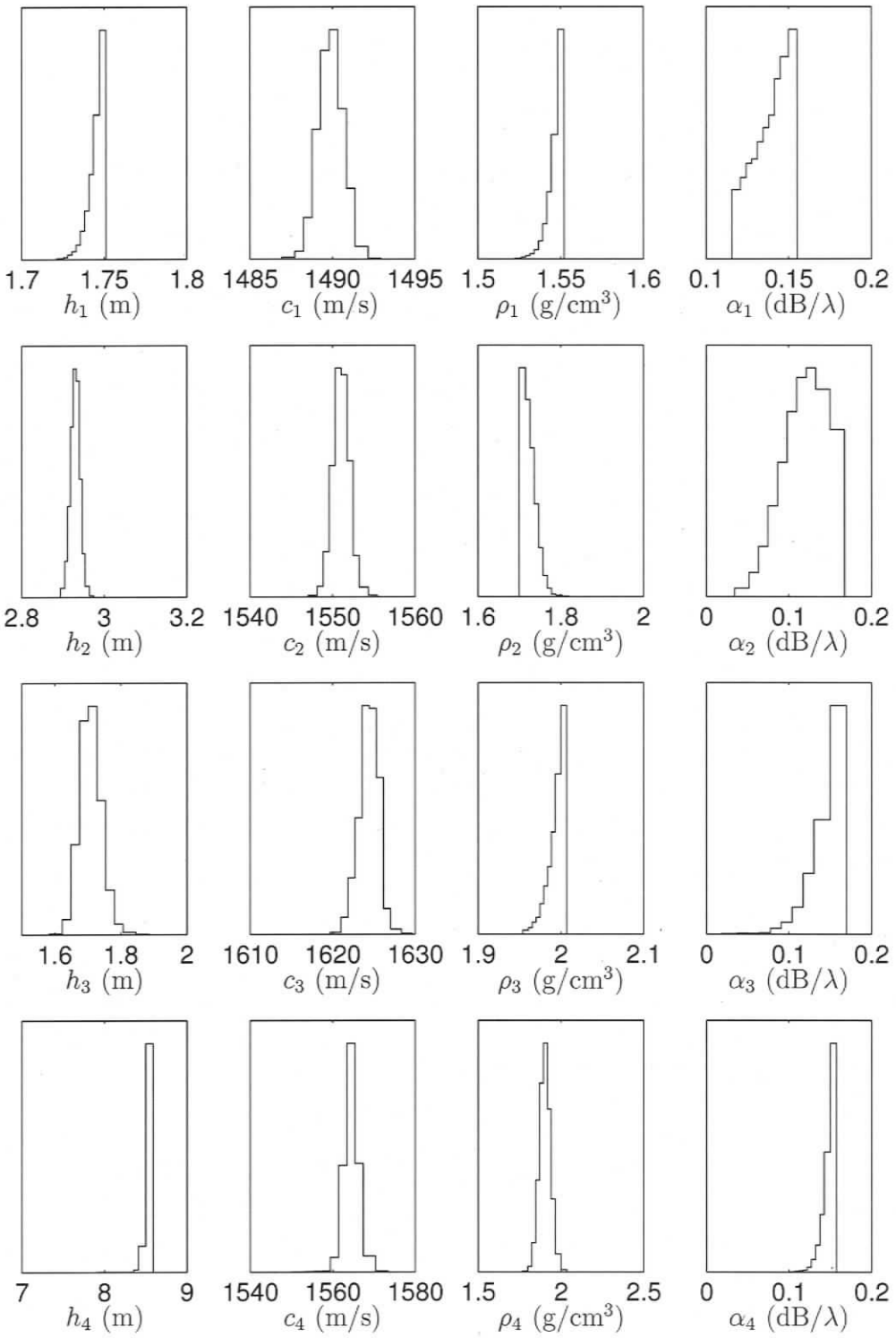

Fig. 7.44: Marginal probability distributions recovered from reflection coefficient inversion of third layer packet (a).

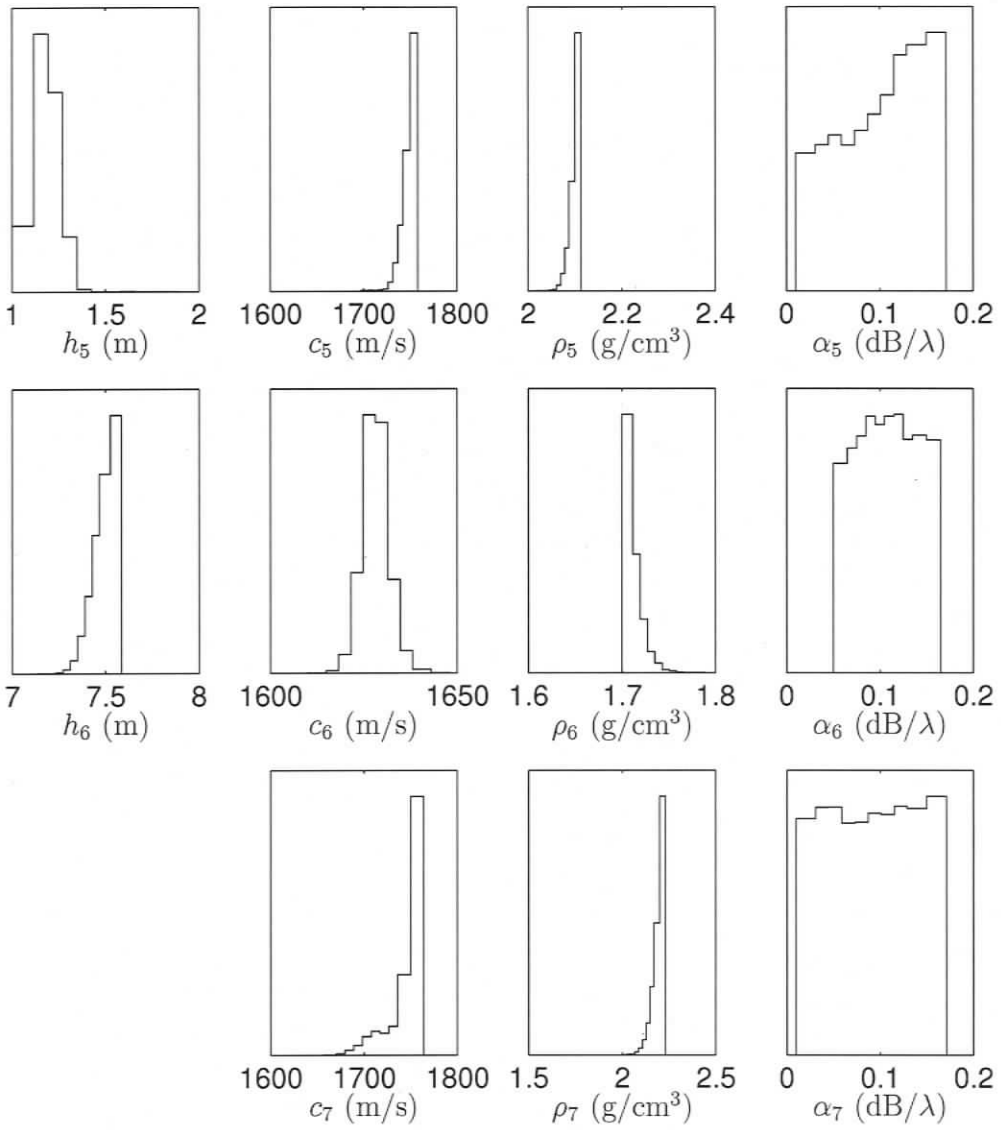

Fig. 7.45: Marginal probability distributions recovered from reflection coefficient inversion of third layer packet (b).

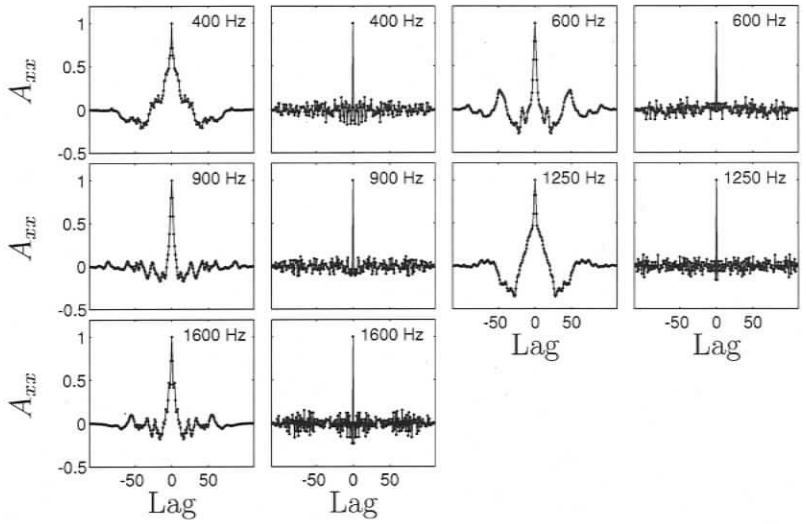

Fig. 7.46: Autocovariance function for raw (columns one and three) and standardised (columns two and four) residuals as a qualitative measure of data error covariances, packet 3.

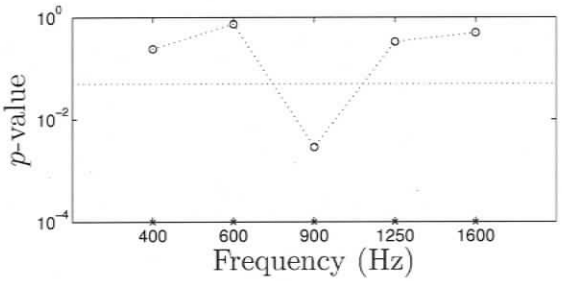

Fig. 7.47: Runs test p-values: The crosses show the p-values for the raw residuals, while open circles show the p-values for the standardised residuals.

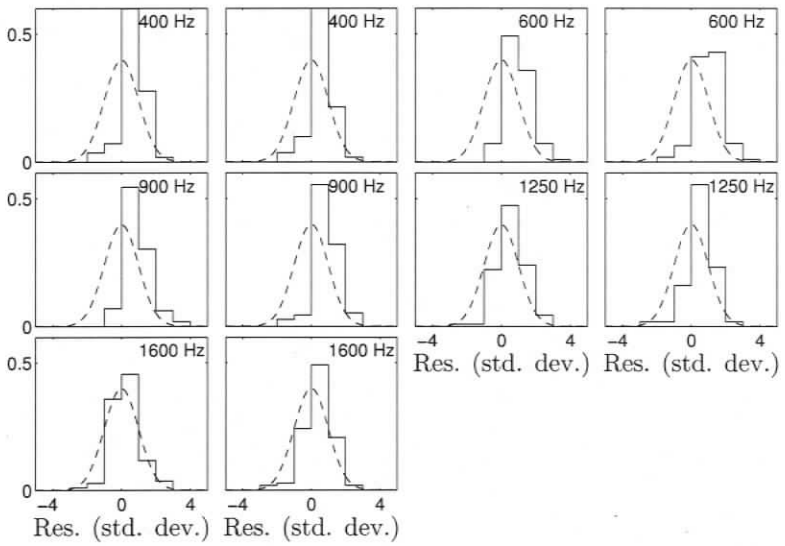

Fig. 7.48: Histograms of raw (columns one and three) and standardised (columns two and four) residuals as a qualitative measure of Gaussianity of data error, packet 3.

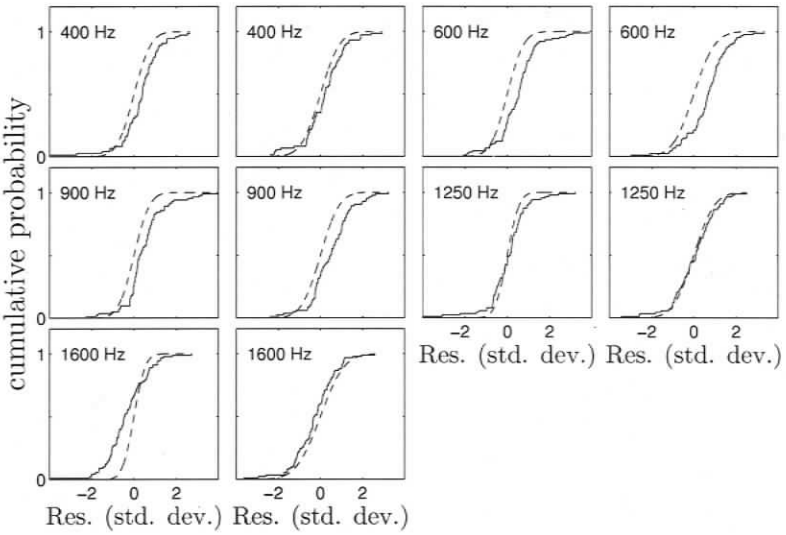

Fig. 7.49: Cumulative probability distributions for raw (columns one and three) and standardised (columns two and four) residuals, packet 3.

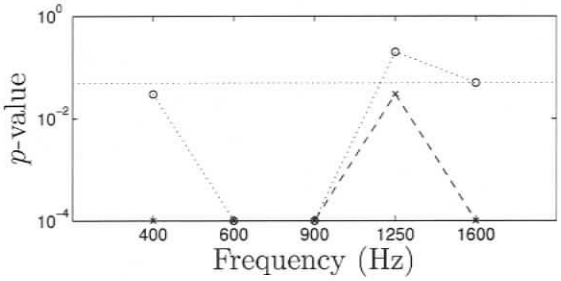

Fig. 7.50: KS test p-values: Crosses show p-values for the raw residuals, while the open circles show the p-values for the standardised residuals.

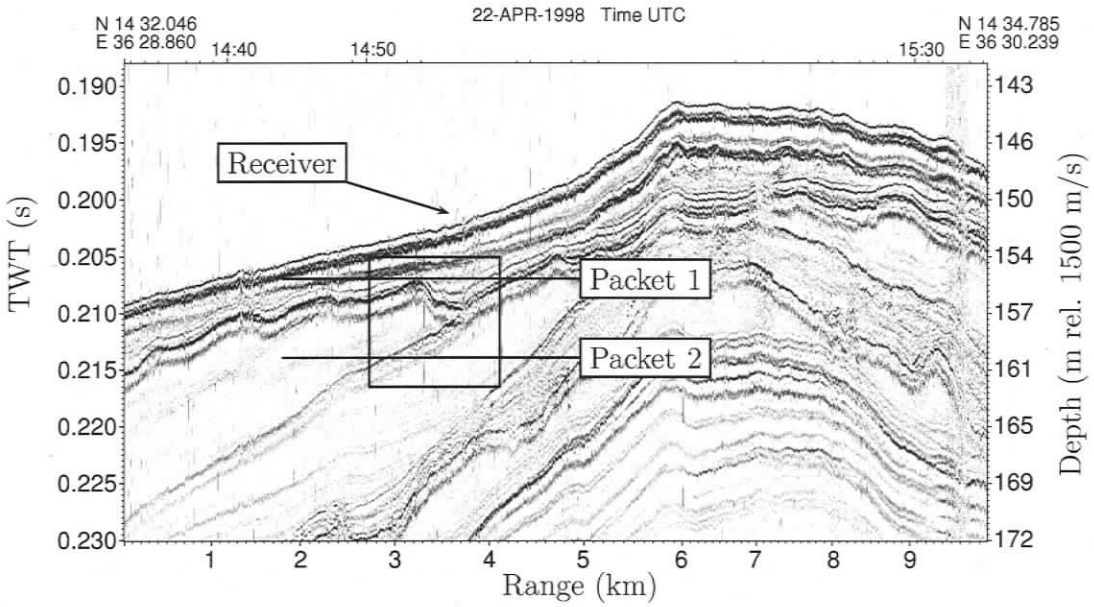

Fig. 7.51: Malta Plateau seismic section. The approximate lower limits of packets 1 and 2 are indicated by solid lines. The box highlights the area below the array where a layer pinches out. (Note: Vertical exaggeration is ~ 220)

are given as shaded areas. In the third layer, the MAP estimate for the density is hitting the upper prior bound (see Fig. 7.33). This density value is not regarded to be well constrained as it peaked at a high value at the upper bound. However, the other parameters seem to give reasonable results. Figure 7.52 compares the inversion results to a piston core that was taken at the experiment site (numerical values are given in Table 7.4). The core reaches to 4 m below the sea-bottom. A second, shallower gravity core is also plotted that largely agrees with the piston core. The inversion results agree well with the core for both density and sound velocity. It should also be noted that the density and sound velocity values at the lower end of the core are likely not realistic. Often, cores are disturbed towards the lower end which yields unrealistic values.

Overall it appears that the inversion method works well when the model parameterisation is applicable. Areas of complicated geology cannot be described with the current model of flat lying layers and would need a much more complicated two-dimensional or even three-dimensional model which is not feasible with current computational capacity. Further, range dependent or three-dimensional modelling of acoustic fields is still limited to fairly simple models and no extensive benchmarks

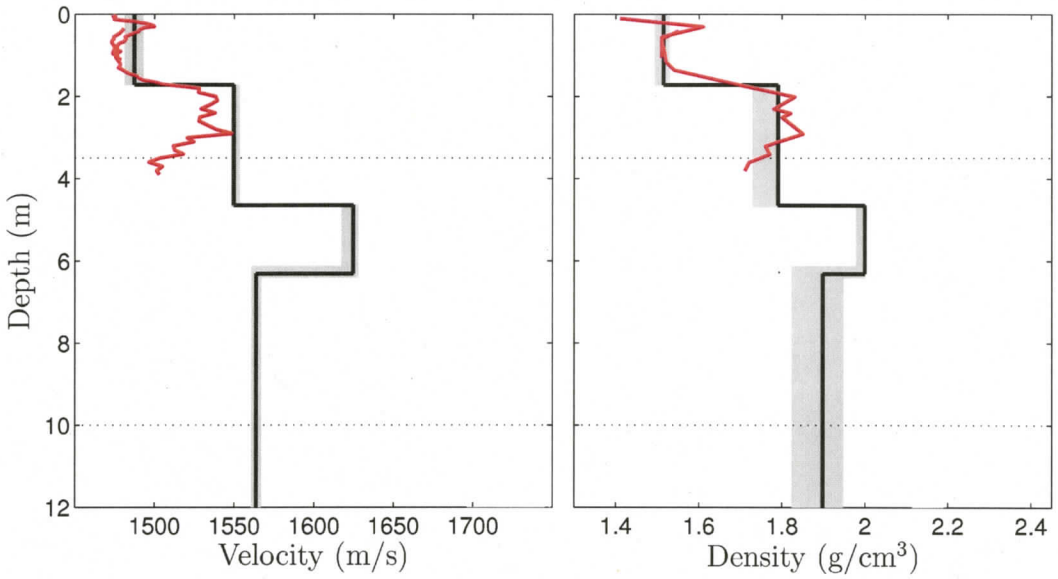

Fig. 7.52: Final inversion results compared to a piston core (red) taken at the experiment site. The piston core was limited to the upper 4 m. The solid black line is the final MAP model and the shaded area represents the 95% HPD credibility intervals.

Layer	h (m)		Sound velocity (m/s)		Density in g/cm^3	
	MAP	95% HPD	MAP	95% HPD	MAP	95% HPD
1	1.74	1.71–1.73	1488	1481–1493	1.514	1.493–1.530
2	2.94	2.91–2.95	1551	1548–1553	1.790	1.728–1.793
3	1.68	1.49–1.72	1624	1617–1628	1.997	1.978–2.000
4			1563	1560–1566	1.898	1.823–1.948

Table 7.4: Numerical values of final inversion results (MAP model and 95% HPD credibility intervals).

exist (Jensen et al., 2006; Chapman and Jiang, 2006).

7.4. Summary and Discussion

This chapter applied the joint time/frequency domain inversion developed in Chapter 6 to data collected on the Malta Plateau. First, travel times for seven individual reflectors were picked in the seismo-acoustic data. The picking was performed in two iterations according to Sec. 6.2.2. The travel times were then inverted using the Bayesian scheme from Chapters 4 and 5. The model parameters consisted of layer thicknesses and sound velocities for six sediment layers (fluid approximation). Since data residuals indicated strongly correlated, non-stationary errors, the errors were addressed in several steps. First, time offset parameters for each reflector were included in the misfit function. Prior bounds for the offset parameters were assigned depending on the confidence in picking particular reflectors. Next, an inversion was carried out assuming uniform uncorrelated errors to provide initial data residuals. These residuals were used to estimate non-stationary standard deviations and a non-Toeplitz covariance matrix according to Sec. 6.2.2.

Gibbs sampling inversion of the travel-time data (weighted by the covariance matrices) yielded a PPD that was interpreted in terms of marginal probability distributions, joint marginal probability distributions and parameter cross correlations. Both physical model parameters and nuisance parameters (representing picking offsets) were examined. The recovered offsets were consistent with the picking confidence. The results for the physical parameters showed good resolution for layer thicknesses. The marginal probability distribution widths were used as prior knowledge for the reflection coefficient inversion. For sound velocities, the total width of the marginal probability distributions was not considerably narrower than the prior bounds and were therefore not used in the reflection coefficient inversion. To validate the travel time inversion results, the statistical assumptions were validated *a posteriori* with quantitative statistical tests (runs test and KS test) and with qualitative plots (autocovariance function and histograms of the data residuals).

For the frequency domain inversion, the seismo-acoustic traces were processed into five frequency bands according to Sec. 6.3. Since the seismo-acoustic data and the travel time inversion indicated a complicated layered system, the reflection coefficient data were processed in three distinct packets each containing the layers of the previ-

ous packet as well as additional deeper layers. The three packets were inverted in a layer packet stripping approach as explained in Sec. 6.6. Due to the geometry of the experiment and the large maximum depth below the seafloor, the full wave forward model was used to model spherical reflection coefficients. To address the computational constraints of the spherical wave model, frequency averaging was replaced by a range average according to Sec. 6.5.

The Bayesian inversion for all packets included full data covariance matrix estimates and accounting for non-stationary data residuals. The results of the Bayesian inversion for each packet were interpreted in terms of marginal probability distributions and the statistical assumptions were examined *a posteriori*. The final MAP model was used to compute synthetic seismograms that were qualitatively compared to the data and showed good agreement. This indicates that the joint time/frequency domain inversion reasonably combines travel time and amplitude information from the seismo-acoustic data.

While inverting the second and third packet, problems with recovered densities became evident. After testing several possibilities (including interface roughness and different tracks for the same site) to account for this problem, high resolution seismic data became available for the site and showed complicated three-dimensional geological structure immediately below the array that violates the assumption of a horizontally stratified environment. Therefore, only the upper 10 meters of the inversion results are considered reliable. In this range, the recovered parameters are consistent with the guidelines in Hamilton (1980). The available piston core data for the upper 4 m agrees well with the inversion results.

8. Remote Sensing of Density and Velocity Profiles in the Transition Layer

8.1. Introduction

This chapter applies Bayesian inversion to seabed reflection coefficient data with the goal of determining density gradients in fine grained sediments close to the water-sediment interface (Holland, Dettmer, and Dosso, 2005). The uppermost sediment layer that often exhibits strong density gradients is only in the order of a few meters thick and is referred to as the transition layer. Knowledge about the transition layer is important to model and understand underwater sound propagation in shallow water. Physical properties of the layer can be measured in situ or from cores, but these methods are costly. Further, coring usually disturbs the sediment and numerous other effects such as settling and compaction during coring and storage can introduce errors which are difficult to quantify. Bayesian inversion of seabed reflection data allows for estimating geoacoustic parameters and quantifying parameter uncertainties while not disturbing the sediment, and reflection measurements can be made at reasonable cost. For the small penetration depth and frequencies of interest in this study, spherical wave effects are negligible. In this chapter, the plane wave forward model is used as developed in Sec. 6.4.1. The reflectivity measurements average over an area of approximately 100 m diameter and are not affected by small scale anomalies (as is coring), but still provide local sampling compared to matched-field inversion methods.

The acoustic data considered here are from the SCARAB98 experiment, conducted in April 1998 in the Strait of Sicily (Holland and Osler, 2000). The experiment location is shown in Fig. 8.1. To separate the effects of the transition layer, the recorded time series were windowed (see Sec. 6.3) to only include effects from the

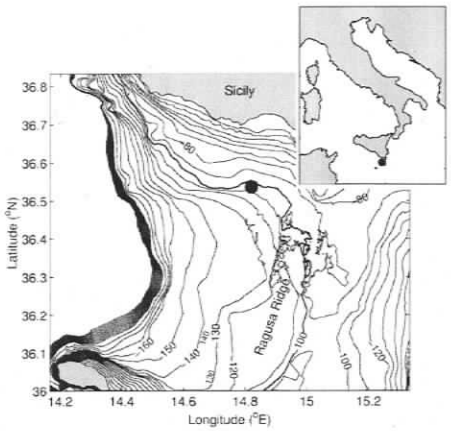

Fig. 8.1: The site of the 1998 SCARAB98 experiment in the Mediterranean Sea.

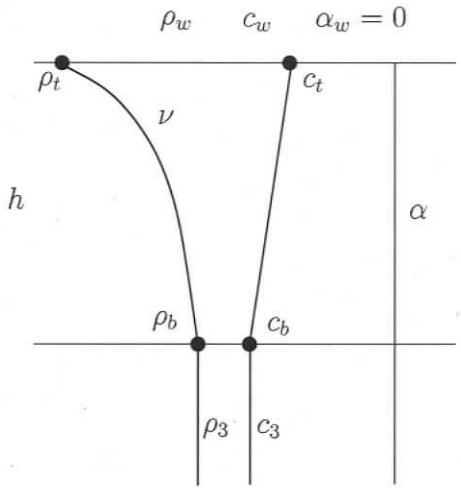

Fig. 8.2: Model parameterisation for density and sound velocity gradients inversion; the nonlinear density profile is represented by parameters ρ_t, ρ_b, ν and h and the linear sound velocity profile by c_t, c_b and h. Attenuation is assumed to be constant in the sub-bottom.

uppermost ~ 1.5 m of the seabed. This uppermost layer did not contain any strong reflectors and is therefore analysed in the frequency domain only (i.e., only reflection coefficient inversion is carried out).

The inversion in this chapter aims at recovery of high resolution structure and requires a special parameterisation that represents detailed information about the density and velocity profiles with a small number of parameters. Figure 8.2 shows the model parameterisation of density and sound velocity gradients between two half-spaces, the upper one being water and the lower one being the basement. The sound velocity gradient is assumed to be linear with velocities c_t and c_b at the top and the base of the layer of thickness h. The nonlinear density gradient is described by four parameters: top and bottom densities ρ_t and ρ_b, layer thickness h and a shape parameter ν according to

$$\rho(\tilde{z}) = \rho_t + \sin^{\nu}(\tilde{z}\frac{\pi}{2})(\rho_b - \rho_t), \tag{8.1}$$

where $\tilde{z}$ is a normalised depth $\tilde{z} = z/h$, and z is the depth into the seabed. Since

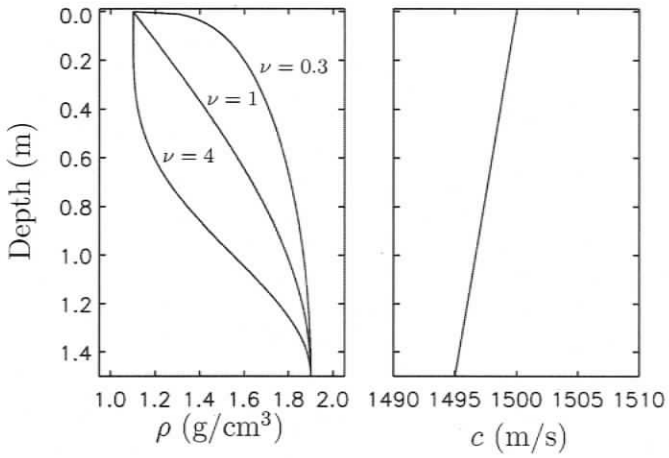

Fig. 8.3: Density and sound velocity parameterisation for measuring density gradients in the transition layer. The different values of ν indicate the possible range of density profiles.

the basement values for density and sound velocity are constrained to the values at the bottom of the gradient layer, and the attenuation is constant in the sediment, the model consists of 7 parameters $\mathbf{m} = [h, c_t, c_b, \rho_t, \rho_b, \nu, \alpha]^\mathsf{T}$. Figure 8.3 illustrates the range of profiles that is included in this parameterisation. The density gradients of transition layers observed in cores often show a low density at the top and then a rapid increase over the first few centimetres. Below this, the density increases more slowly with depth (Holland, Dettmer, and Dosso, 2005). This typical profile shape is well described by the chosen parameterisation when values of ν are less than 1. Since there is no analytic solution for the reflection coefficient for this type of parameterisation, the forward model uses a discretised profile (usually $\sim$ 10–20 layers) calculated from the above functions and recursively calculates the reflection coefficient.

It is important to note that h is a shape parameter (i.e., influences the shape of the density profile) and does not necessarily indicate the base of a distinct layer, since the parameterisation does not allow for a reflector. This is justified here because the data show no evidence of a prominent reflector and appropriate time windowing can ensure that no deeper layers are important. If a reflector is present, the method can also treat the properties of the basement as unknowns by adding more parameters to the inversion.

The reflectivity data recovered from the recorded seismo-acoustic time series are of high quality and are shown in Fig. 8.4. This figure shows the data represented as

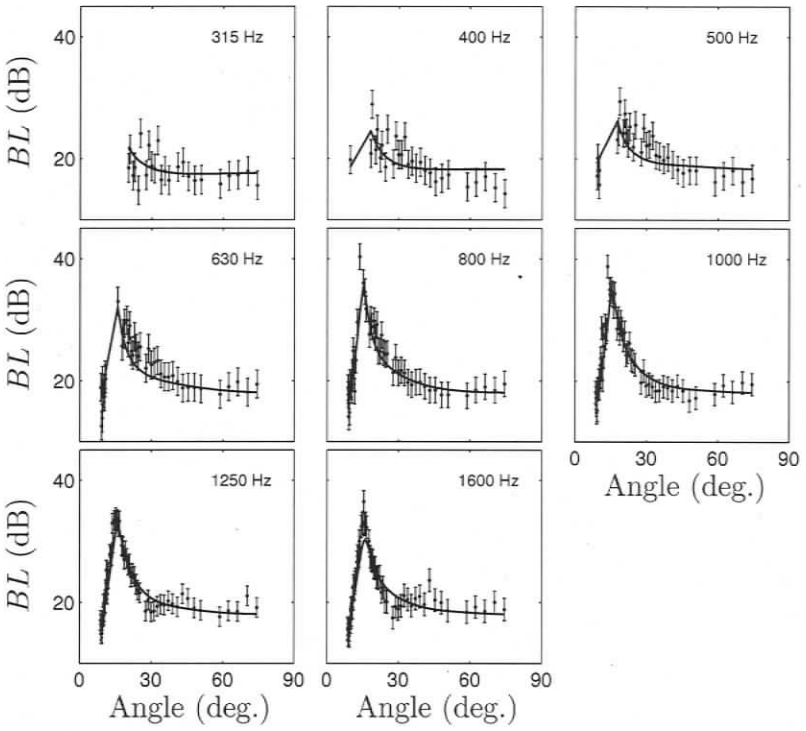

Fig. 8.4: Bottom loss data at the transition layer site (for clarity, only every second datum is shown) with one standard deviation ML error estimates. Note that the angle of intromission is clearly visible as a sharp peak in BL. Replica data are shown as solid lines.

bottom loss (BL), derived from reflection coefficients V by

$$BL = -20\log_{10}(|\mathsf{V}|). \tag{8.2}$$

Nine frequencies from 315 to 1600 Hz were chosen, with each frequency representing the centre of a 1/3-octave band average. The angles range from approximately 12–80° (grazing angle), and, because the source moves at a constant speed, the angular spacing of the data are not uniform. The data at most frequencies show a very prominent angle of intromission that is characterised by a sharp peak of high bottom loss. The angle of intromission lies at around 17° and is most evident at frequencies above 630 Hz. Figure 8.4 also shows replica bottom loss data computed for the MAP model. One standard deviation error estimates, derived from the data residuals, are included on the measured data. It can be seen that the fit to the measured data are good and that the angle of intromission is well represented by the replica data. However, the measured data show more structure than the replica data which results in correlated data errors. The correlated data errors become obvious when the data covariance matrix is computed. Estimating the covariance matrices is complicated by the non-uniform data spacing and follows the technique outlined in Sec. 5.3 to avoid resampling the data onto an evenly spaced grid. In this case, resampling the data is considered disadvantageous since the angular spacing is higher at small grazing angles, where the angle of intromission is located. Having high angular resolution around the angle of intromission ensures its position is well matched , providing considerable information about the physical seabed properties. For the case of uncorrelated errors, the covariance matrix would be diagonal. However, Fig. 8.5 shows that the estimated covariance matrix at 630 Hz displays significant off-diagonal terms (similar results at other frequencies). The non-Toeplitz form that results from the nonuniform sampling is shown in Fig. 8.5 (a). Fig. 8.5 (b) shows the same matrix but plotted against angle. The bands of large pixels are due to gaps in the recording when data were written to disk.

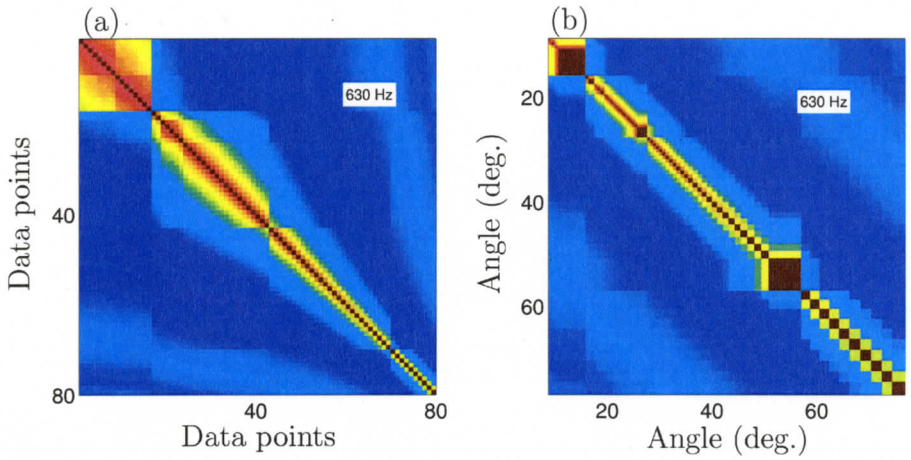

Fig. 8.5: Example of an estimated covariance matrix with nonuniform angle spacing at 630 Hz. (a) Plot of the matrix and (b) stretched to actual angles.

8.2. Broad Band Reflectivity Inversion with Correlated Errors

Before presenting the inversion results for the measured transition layer data, this section considers inversion of synthetic bottom loss data from a simulated experiment designed to replicate the data described in Sec. 8.1. Simulations allow the performance of various components of the inversion algorithm to be examined individually. In simulations, the data errors are exactly known and controlled, so the validity of assumptions and approximations in the analysis is evident. In addition, the final solution can be compared to the true model parameters to unambiguously quantify performance.

It is common in underwater acoustics to assume the data errors to be uncorrelated. However, while working on the broad band reflection loss data (see Fig. 8.4), it became obvious that strong error correlations existed and could not be ignored. This section considers pre-processing of simulated, uniformly sampled single bounce reflection-loss data with strongly correlated data errors to improve the nonlinear Bayesian inversion. Correlated Gaussian errors are generated using realistic data covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ at each frequency derived from experimental measurements (Sec. 8.1).

As a test case, a typical soft bottom environment is chosen. Only the uppermost few metres are used in the simulation with a sediment sound velocity lower than that of water at the sediment-water interface and a strong density gradient in the

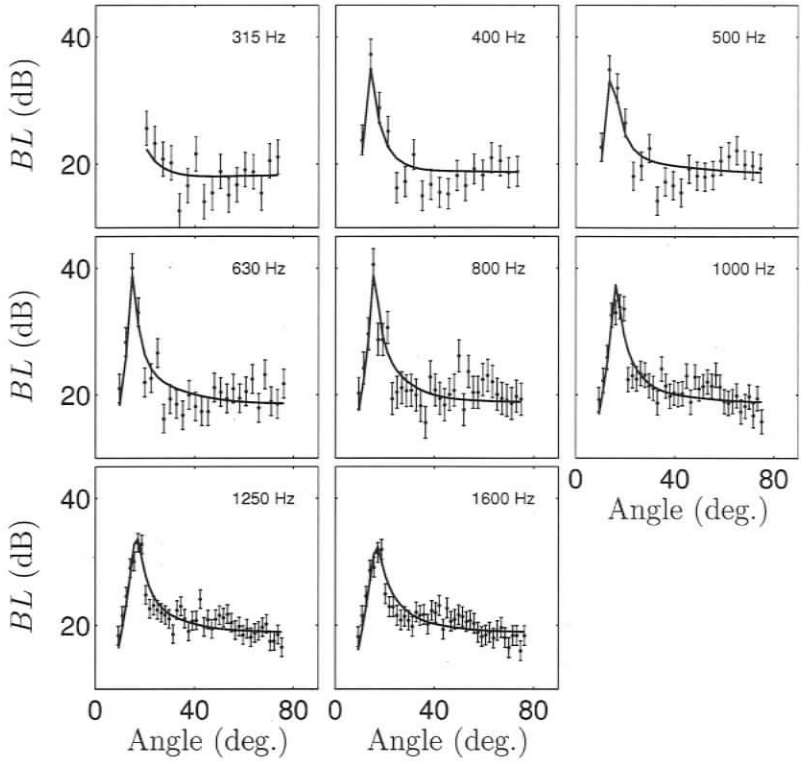

Fig. 8.6: Simulated reflectivity data for a transition layer and strongly correlated data errors (for clarity, only every second datum is shown) with one standard deviation ML error estimates. MAP replica data are shown as solid lines.

uppermost layer. The true parameter values are given in Table 8.1. The inverse problem of estimating the model parameters from the BL data are solved with Gibbs sampling (see Sec. 4.1), which provides parameter estimates and credibility intervals by sampling the posterior probability density.

The simulated data set was generated for eight frequencies f_i in the range of 315–1600 Hz, where each frequency represents the centre of a 1/3 octave band average and the data are given in units of bottom loss (see Fig. 8.6). There are a different number of data N_i for each frequency band, analogous to the observed data set. To generate the noisy data, a noiseless data set $\bar{\mathbf{d}}_i$ was first computed with the plane-wave forward model (see Sec. 6.4.1 for details). A error vector $\bar{\mathbf{n}}_i$ for every frequency was generated from a Gaussian distribution with zero mean and unit standard deviation. Noisy data with correlated errors were then formed using the Cholesky decomposition (Eq. 5.4)

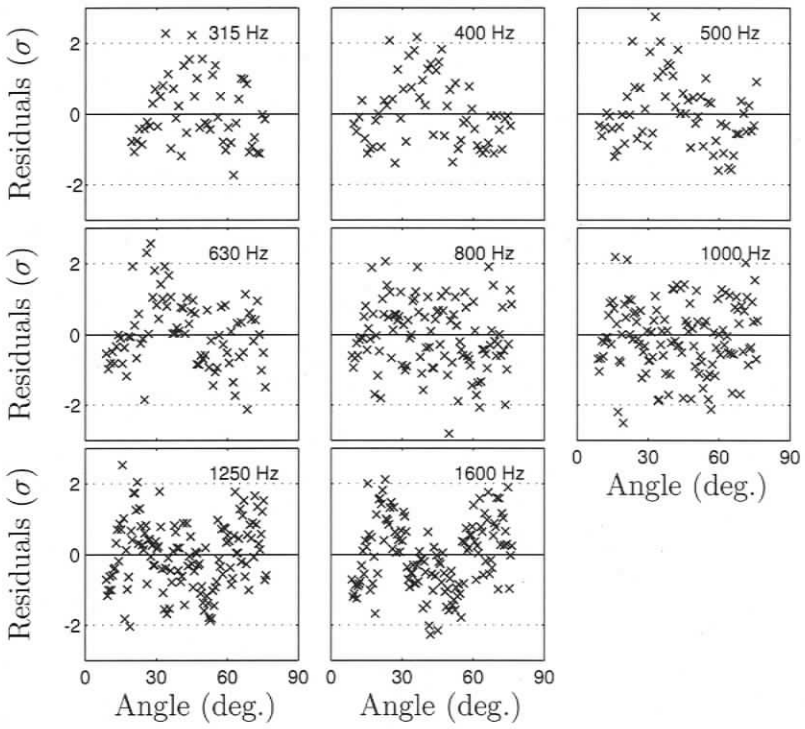

Fig. 8.7: Data residuals in units of standard deviations. The residuals show strong correlations at certain frequencies.

of the covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ according

$$\mathbf{d}_i = \bar{\mathbf{d}}_i + \mathbf{L}_i \bar{\mathbf{n}}_i \,, \tag{8.3}$$

where $\mathbf{L}_i$ is the lower triangular matrix of the Cholesky decomposition of the data covariance matrix. The covariance matrices are derived from the observed data under the assumption of uniformly spaced data.

In order to apply Bayesian inversion to the noisy data, error covariance matrices were first estimated from the data as follows. Using the noisy synthetic data set, a MAP estimate $\hat{\mathbf{m}}$ was computed via ASSA inversion and residuals $\mathbf{n}_i$ were calculated as the difference of simulated data $\mathbf{d}_i$ and replica data $\mathbf{d}_i(\hat{\mathbf{m}})$. The residuals, scaled by the estimated standard deviations, are plotted in Fig. 8.7, and show strong serial correlations (i.e., the residuals are not randomly distributed about zero). To take the error correlations into account in the inversion, a set of data covariance matrices (one at each frequency) with Toeplitz form was estimated recursively from the residuals using the approach developed in Sec. 5.3. Figure 8.8 shows the true covariance matrices $\mathbf{C}_i^{(\mathbf{d})}$ and the recovered (after two iterations) covariance matrices $\hat{\mathbf{C}}_{(i,2)}^{(\mathbf{d})}$. For simplicity, only the first row of each matrix is plotted (since the matrix has Toeplitz form, all information is included in the first row). The recovered covariance matrices match the true covariances well at all frequencies. In some cases, especially when the correlations are not very strong, the recovered covariance includes some oscillations that are not present in the true covariance. Since these oscillations do not appear to have physical meaning, smoothing and damping could be applied.

To qualitatively examine the effect of taking error covariance into account, Eq. 5.20 was applied with $\hat{\mathbf{C}}_{(i,2)}^{(\mathbf{d})}$ and the normalised autocovariance functions were calculated for the raw residuals $\mathbf{n}_i$ and for the standardised residuals $\tilde{\mathbf{n}}_i = \mathbf{L}_i \mathbf{n}_i$, as shown in Fig. 8.9. In the autocovariance function, a wide centre peak indicates strong serial correlations. It can be seen that applying the recovered covariance matrix $\hat{\mathbf{C}}_{(i,2)}^{(\mathbf{d})}$ substantially reduces the width of the centre peak. Figure 8.10 shows histograms of the raw and standardised residuals at all frequencies. The histograms are compared to a theoretical Gaussian distribution with zero mean and unit standard deviation. All frequencies appear reasonably Gaussian for raw as well as standardised residuals, which is consistent with the Gaussian random errors that were used to generate the correlated errors.

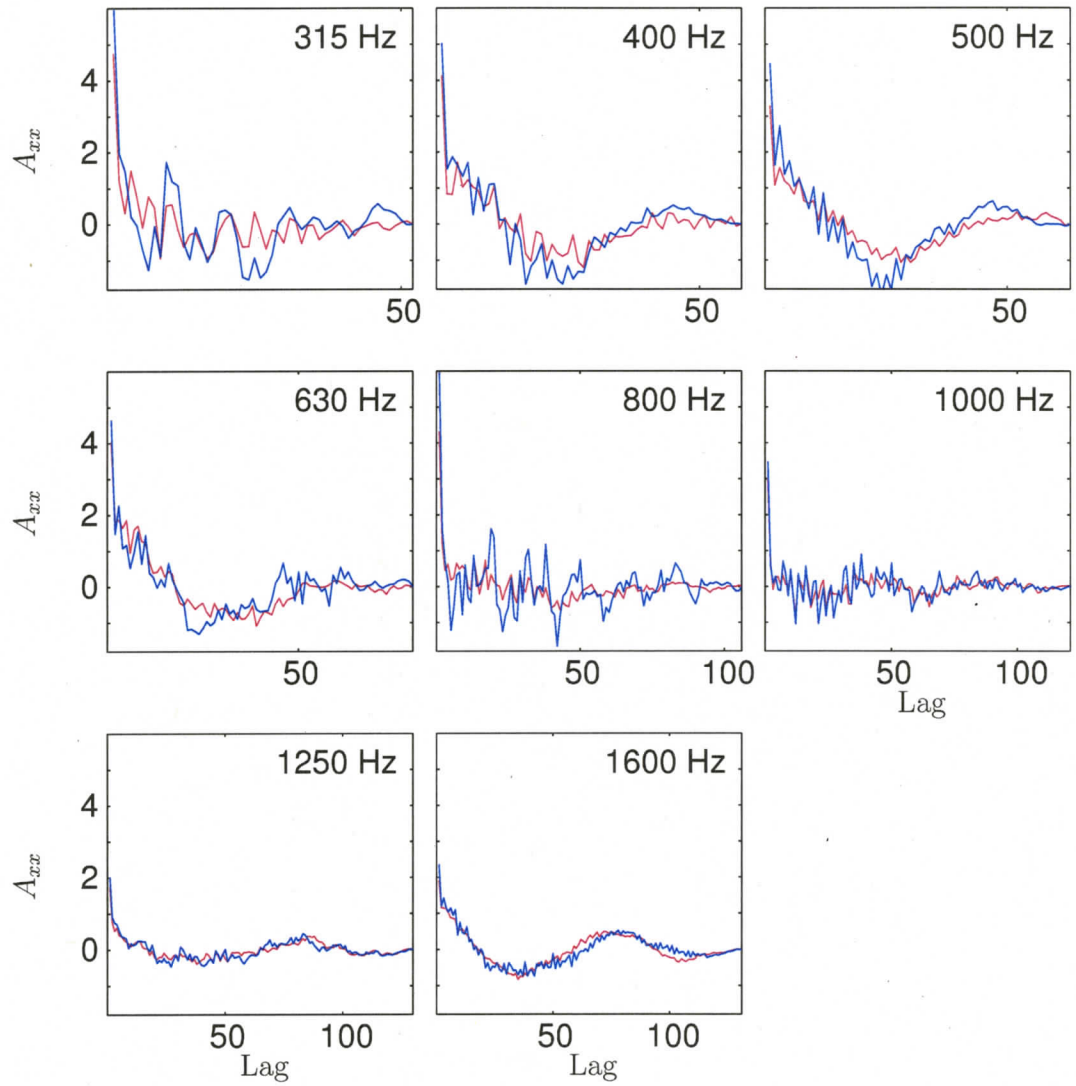

Fig. 8.8: Comparison of true (red) and recovered (blue) covariance matrices. Only the first row of each matrix is shown for simplicity, since this contains all the information for a Toeplitz matrix.

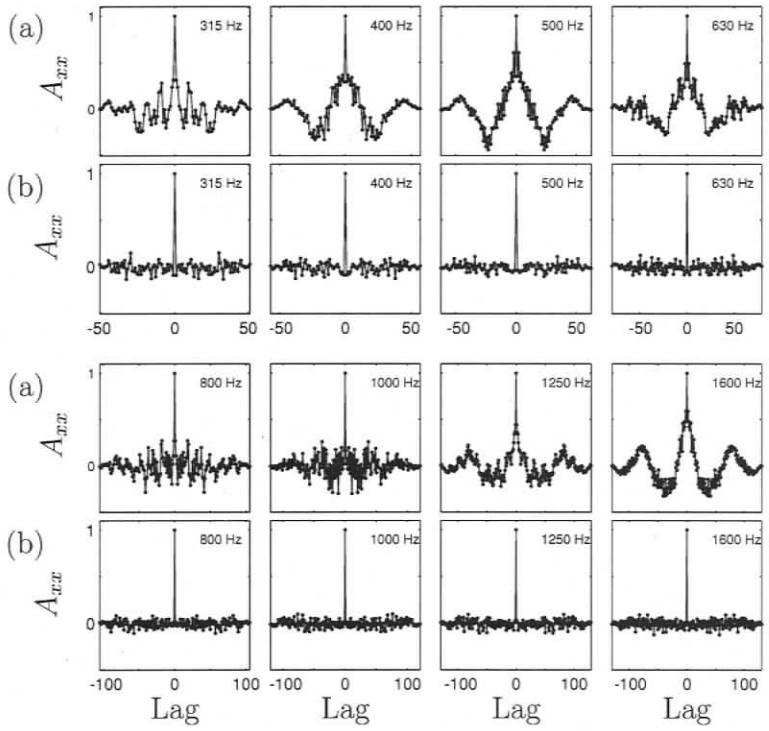

Fig. 8.9: Qualitative examination of the effect of applying a data covariance matrix. (a) Show the autocovariance function of raw residuals $\mathbf{n}_i$ and (b) show the autocovariance function of standardised residuals $\tilde{\mathbf{n}}_i$.

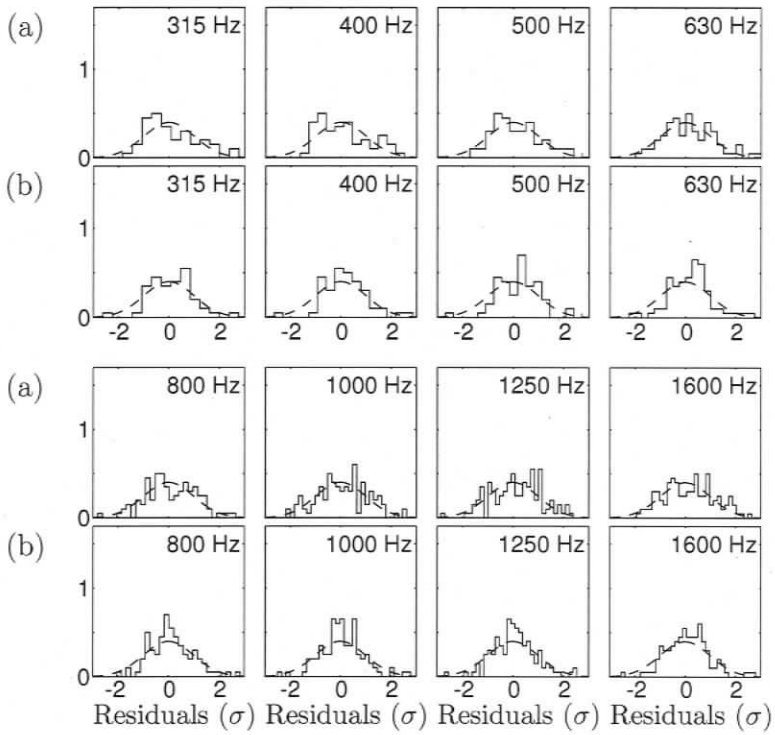

Fig. 8.10: Qualitative examination of the effect of applying a data covariance matrix. (a) Shows the histograms of raw residuals ($\mathbf{n}_i$) and (b) shows histograms of standardised residuals ($\tilde{\mathbf{n}}_i$).

To quantify randomness, a runs test (see Sec. 5.5) was applied to both sets of residuals. For $\mathbf{n}_i$, the p-values were $10^{-4} < p < 0.16$ with residuals at 4 of 8 frequencies passing at the 0.05 level (a value of $p > 0.05$ is generally accepted as little or no evidence against randomness). For $\tilde{\mathbf{n}}_i$, $0.05 < p < 0.78$ with the standardised residuals passing at all frequencies. To quantify Gaussianity of the residuals, a KS test was performed for raw and standardised residuals. The p-values were computed from critical values for the appropriate number of data according to Sec. 5.5.2 and the test was passed at all frequencies for raw and standardised residuals with $p >> 0.05$. The result of the statistical test provides confidence in the inversion results, since the estimated covariance matrix properly describes the data error statistics.

The recovered covariance matrices $\hat{\mathbf{C}}^{(\mathbf{d})}_{(i,2)}$ were applied in a Gibbs sampler to estimate density and sound velocity profiles from the data. Figure 8.2 shows a comparison of the marginal distributions computed when (a) data error correlations are not taken into account, and (b) applying $\hat{\mathbf{C}}^{(\mathbf{d})}_{(i,2)}$ in the inversion. While marginal distributions are generally useful, in this case, due to the model parameterisation, it is more useful to plot full profiles as final results. These final results are plotted in Fig. 8.12 which shows a comparison of the profiles and uncertainties computed both with and without taking error correlations into account. In Fig. 8.12 (a), the true model lies outside of the 95% HPD credibility interval for both the density and velocity profiles, indicating an inversion result that is inconsistent with the true model. In Fig. 8.12 (b) the true model is within the 95% HPD credibility interval at all depth for both density and velocity. Further, the recorded profile shapes closely represent the true shapes. Kinks in the 95% HPD credibility intervals are due to the model parameterisation and have no physical meaning. Numerical values for the true model parameters, prior bounds, MAP models, and 95% credibility intervals can be found in Tables 8.1 and 8.2.

8.3. Inversion with Gaussian Error Statistics

This section presents the inversion results for the measured data (Fig. 8.4) in terms of density and and sound velocity gradients under the plane wave assumption. This section inverts the non-uniformly spaced data including full data covariance matrix estimates. The statistics of the data residuals and the results of statistical tests are discussed first. Then the parameter estimates and uncertainties are presented and compared to two gravity cores which were collected in the area.

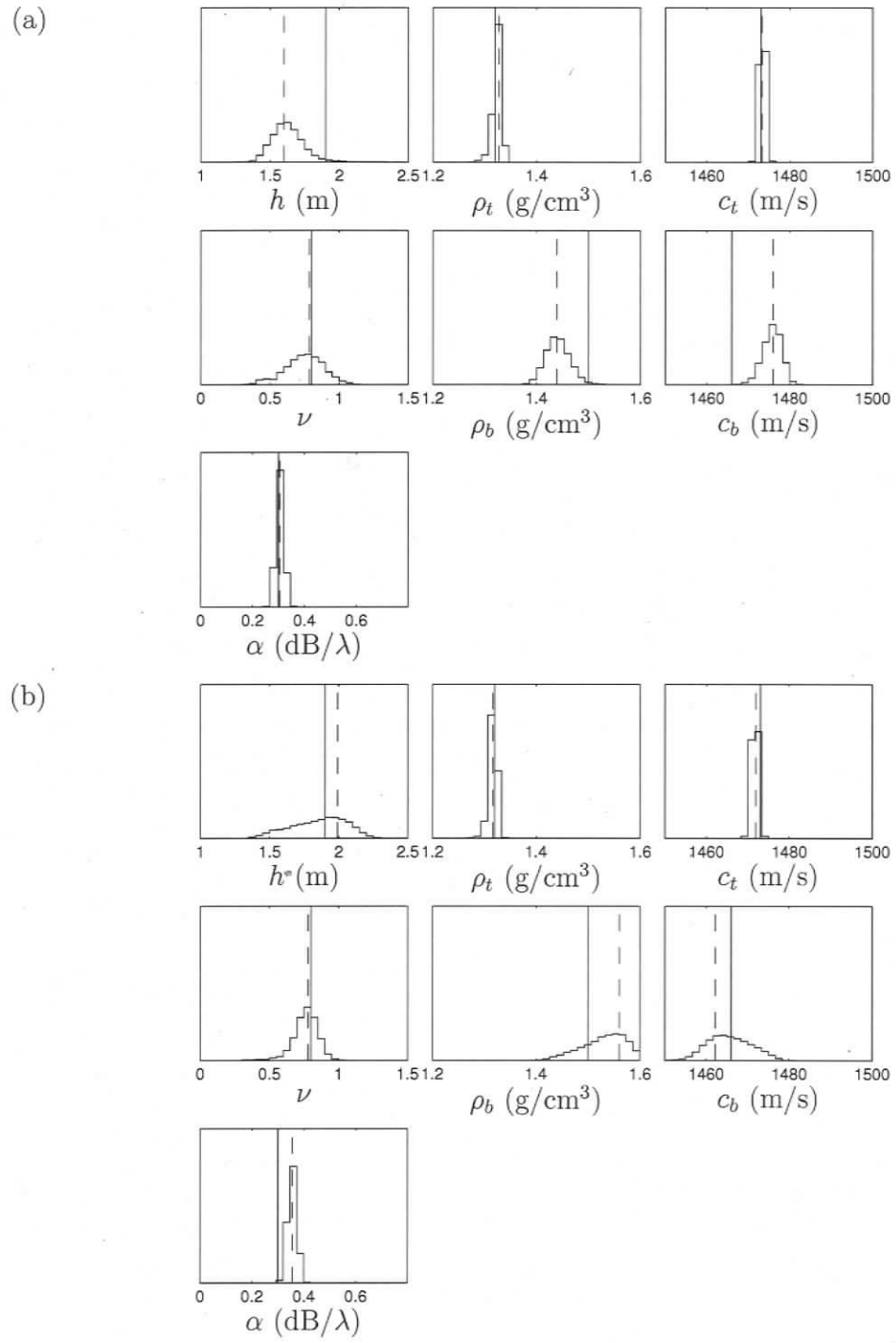

Fig. 8.11: Comparing marginal probability distributions when (a) no covariance matrix applied, and (b) estimated covariance $\hat{\mathbf{C}}^{(\mathrm{d})}_{(i,2)}$ applied. The true model is shown as solid lines and MAP estimates are shown as dashed black lines.

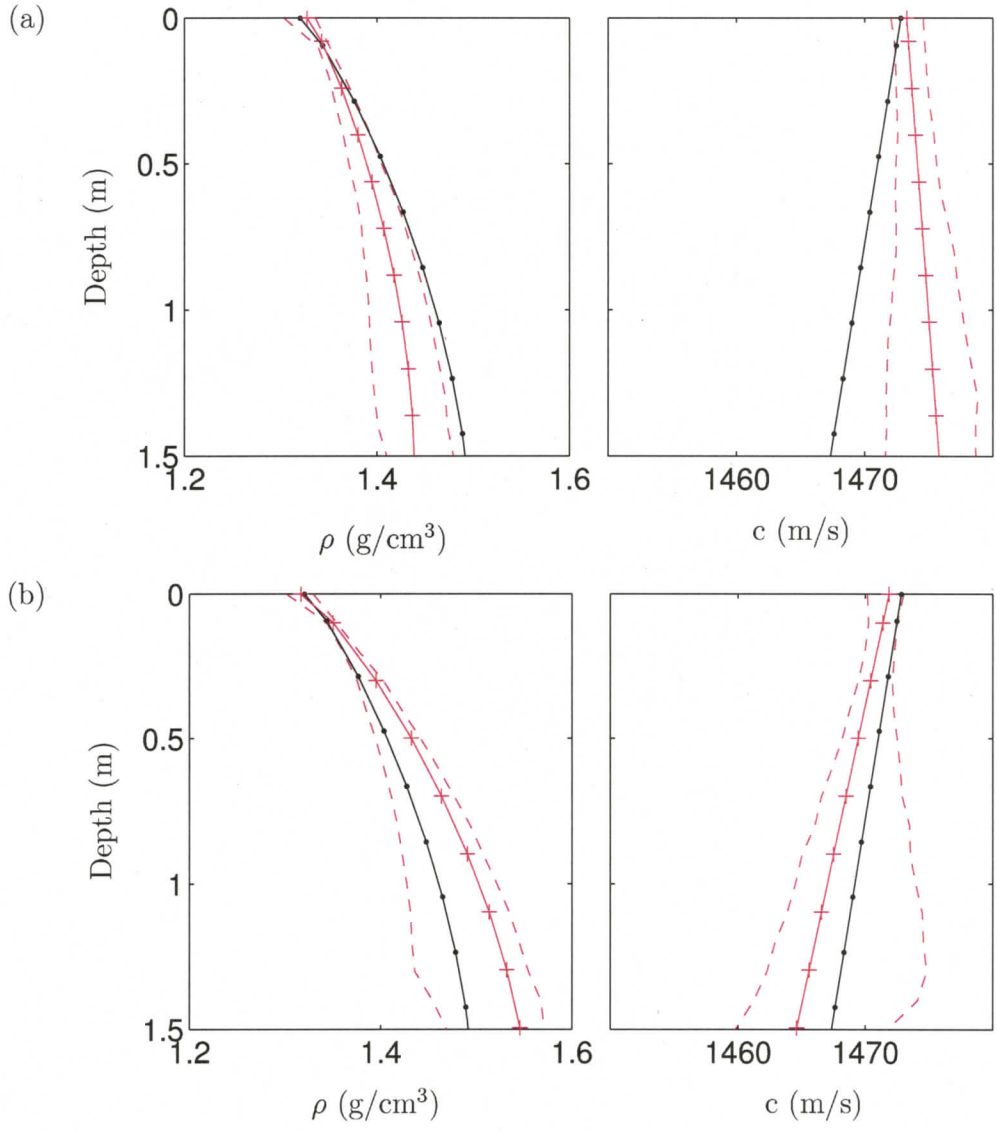

Fig. 8.12: Comparing inversion results for the simulated experiment when (a) only standard deviation applied, and (b) estimated covariance $\hat{\mathbf{C}}^{(\mathbf{d})}_{(i,2)}$ applied. The true model is shown as a solid black line, MAP estimates and 95% HPD intervals are shown as solid red lines and dashed red lines, respectively.

Parameter	true	prior bounds	MAP	95% HPD
h(m)	1.9	1.0–2.5	1.60	1.41–1.84
ρ_t(g/cm^3)	1.32	1.1–1.8	1.327	1.305–1.337
c_t(m/s)	1473.0	1450–1550	1473.27	1472.0–1474.6
ν	0.8	0.0–1.5	0.78	0.42–0.98
ρ_b(g/cm^3)	1.5	1.3–1.8	1.439	1.396–1.485
c_b(m/s)	1466.0	1450–1550	1475.98	1471.2–1479.8
α(dB/λ)	0.3	0.0–0.8	0.35	0.27– 0.33

Table 8.1: Summary of inversion results for simulation, not taking covariances into account.

Parameter	true	prior bounds	MAP	95% HPD
h(m)	1.9	1.0–2.5	1.99	1.46–2.17
ρ_t(g/cm^3)	1.32	1.1–1.8	1.316	1.30–1.32
c_t(m/s)	1473.0	1450–1550	1471.85	1470.1–1472.9
ν	0.8	0.0–1.5	1.00	0.58–0.93
ρ_b(g/cm^3)	1.5	1.3–1.8	1.560	1.44–1.59
c_b(m/s)	1466.0	1450–1550	1462.20	1456.4–1475.5
α(dB/λ)	0.3	0.0–0.8	0.35	0.32–0.38

Table 8.2: Summary of inversion results for simulation, taking covariances into account.

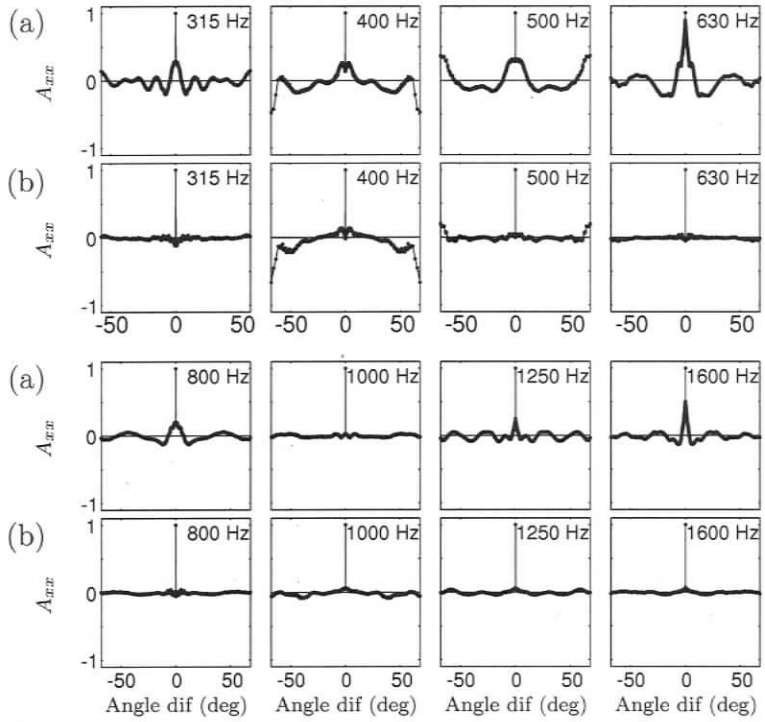

Fig. 8.13: Estimate of the autocovariance function for (a) raw residuals $\mathbf{n}_i$ and (b) standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$.

8.3.1. Data Residuals and Covariance

In the inversion, uniform priors were applied over a wide range of values for all parameters. Figure 8.4 shows the fit of the replica data, generated for the MAP model $\hat{\mathbf{m}}$, to the measured data. The measured data include one standard deviation ML error estimates quantifying both theory and measurement errors. The good agreement between measured and replica data indicates a reasonable parameterisation. However, the plot also indicates correlated errors. Rows 1 and 3 in Fig. 8.13 give the autocovariance function of the raw data residuals $\mathbf{n}_i$ at each frequency and show that the function at every frequency features a wide peak around lag zero indicating strong error correlations. Hence, a full covariance matrix for every frequency (see Fig. 8.5) was estimated as described in Sec. 5 and applied in the misfit function. This produced a second MAP model $\hat{\mathbf{m}}^{(2)}$ and residuals $\mathbf{n}_i^{(2)}$. The result of applying the data covariance matrix estimate to the data is shown in rows 2 and 4 of Fig. 8.13 for the residuals $\tilde{\mathbf{n}}_i^{(2)}$, pre-whitened according to Eq. 5.20. The peaks of the autocovariance function of $\tilde{\mathbf{n}}_i^{(2)}$ are narrow and there are virtually no oscillations at greater lags.

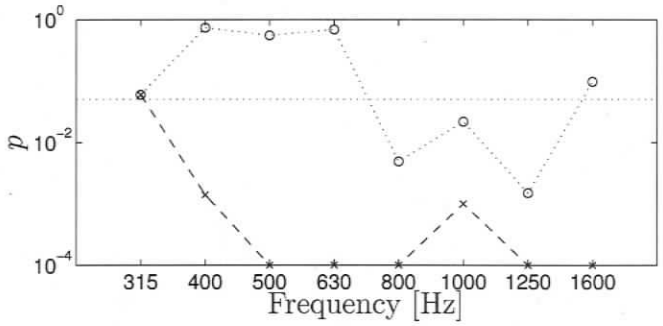

Fig. 8.14: Runs test p-values for the inversion with Gaussian statistics. The dotted line indicates the $p = 0.05$ level of significance. Crosses show the p-values for raw residuals $\mathbf{n}_i$ and circles shoe the p-values for standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$.

To quantify the randomness of the residuals, a runs test was performed and p-values calculated for every frequency, both before and after applying the data covariance matrix estimate according to Eq. 5.20. Before taking covariances into account, the residuals had p-values of $p < 10^{-2}$ to $p = 0.06$ with all but one frequencies fail at the $p > 0.05$ confidence level. After taking correlations into account, the test produced values between $p = 0.002$ and $p = 0.737$; five out of eight frequencies pass the test at the 0.05 confidence level. The p-values are given in Fig. 8.14.

To quantify the effect of the data covariance matrix estimate on the statistical distribution of the residuals, the KS test was applied to $\mathbf{n}_i$ and $\tilde{\mathbf{n}}_i^{(2)}$. Figure 8.15 shows cumulative distributions for the data residuals (solid line) and a theoretical Gaussian distribution (dashed lines) that are the basis of the KS test. Figure 8.15 (a) shows cumulative probability distributions for the raw residuals $\mathbf{n}_i$ and Fig. 8.15 (b) shows cumulative probability distributions for the standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$, pre-whitened according to Eq. 5.20. The cumulative probabilities are compared to theoretical Gaussian distributions. The match between the theoretical distributions and the data residuals appears to be reasonably good for most frequencies before and after applying the covariances. The p-values from the KS test, given in Fig. 8.17, show that four of eight frequencies pass at the $p > 0.05$ confidence. In one case, $f = 1000$ Hz, the p-value changes from $p = 0.13$ for $\mathbf{n}_i$ to $p = 0.03$ for $\tilde{\mathbf{n}}_i^{(2)}$. In all other cases, applying the covariance matrix does not change the statistics of the residuals significantly. As a qualitative test, Fig. 8.16 shows histograms of raw and standardised residuals which are compares to theoretical Gaussian distributions ($G(0, 1)$). It can be seen that the agreement is reasonably good at all frequencies with no large outliers;

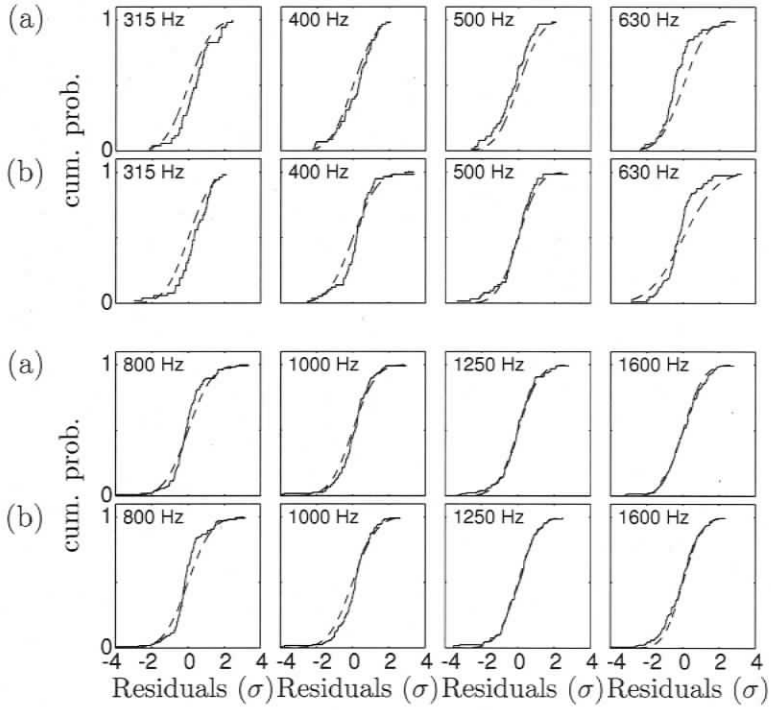

Fig. 8.15: Cumulative probability distributions for (a) raw residuals $\mathbf{n}_i$ and (b) standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$.

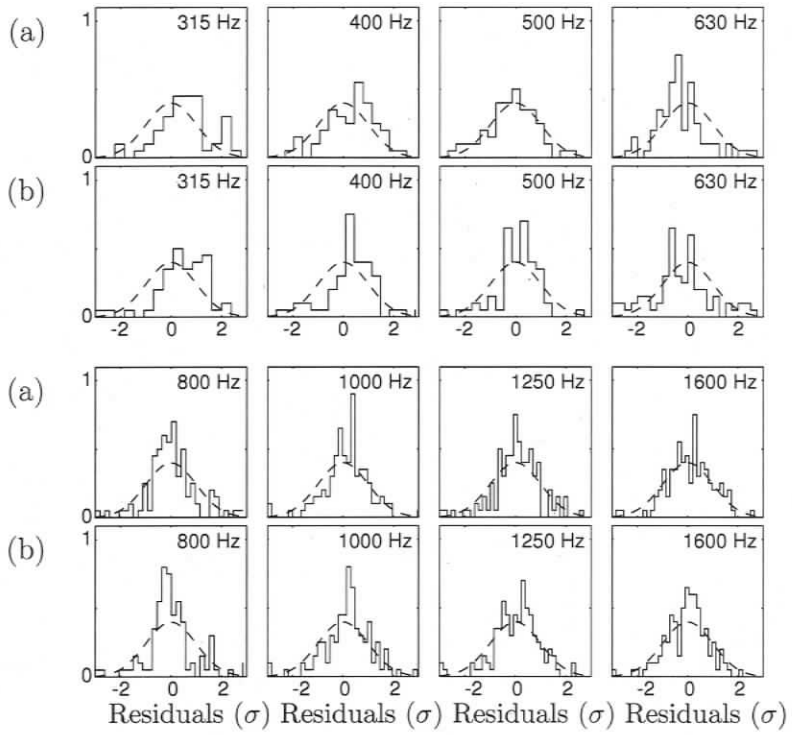

Fig. 8.16: Histograms for (a) raw residuals $\mathbf{n}_i$ and (b) standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$ compared to theoretical Gaussian distributions (dashed lines).

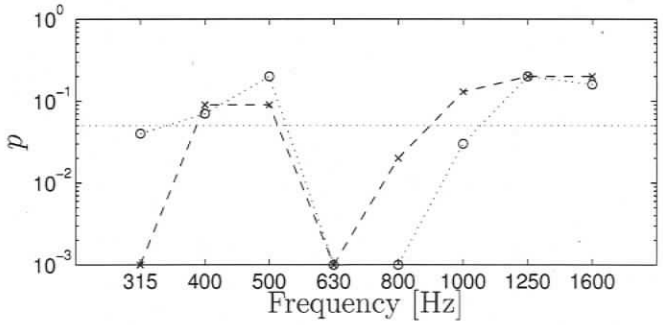

Fig. 8.17: Results of the KS test for the inversion with Gaussian statistics. The dotted line indicates the $p = 0.05$ level of significance. Crosses show the p-values for raw residuals $\mathbf{n}_i$ and circles show the p-values for standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$.

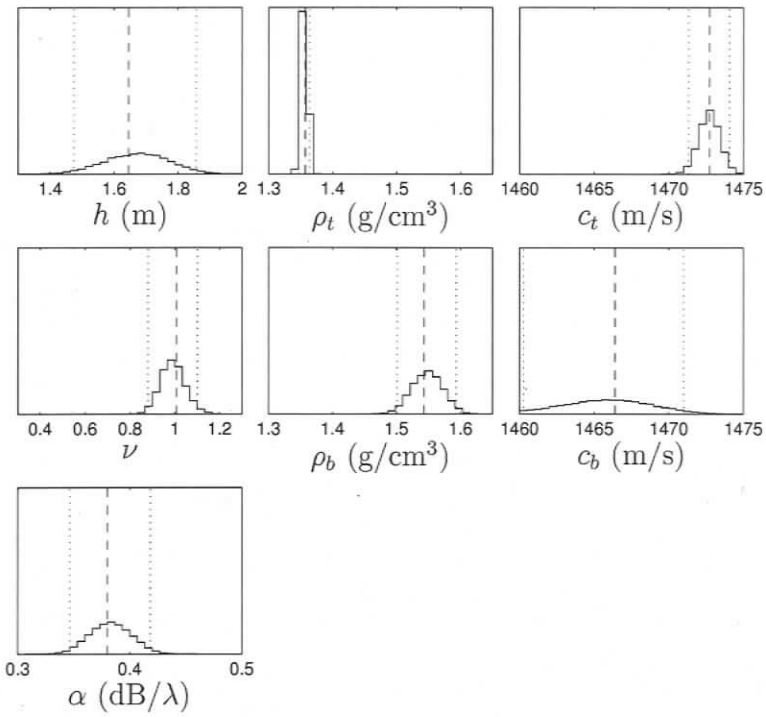

Fig. 8.18: Marginal probability distributions for an estimate of the PPD that incorporated $\hat{\mathbf{C}}^{(\mathbf{d})}_{(i,2)}$. The MAP estimate is shown as a dashed line, 95% HPD credibility intervals as dotted lines.

however, at some frequencies the distribution of data residuals are more peaked and show somewhat heavier tails than the theoretical Gaussian.

8.3.2. Parameter Estimates and Uncertainties

This section presents the inversion results that were obtained from applying the Bayesian inversion with the assumption of Gaussian errors to the measured data. Figure 8.18 shows the marginal probability distributions for the seven geoacoustic parameters. Numerical values for model priors, 95% HPD credibility bounds and MAP estimates are given in Table 8.3. It is striking that the density distributions ρ_t and ρ_b do not overlap, which means that a density gradient is clearly resolvable from the bottom loss data.

Figure 8.19 shows four selected joint marginal distributions. Figure 8.19 (a) indicates a strong positive correlation between layer thickness h and base density ρ_b. Another strong, but negative, correlation can be seen in Fig. 8.19 (b) between the

Parameter	prior bounds	MAP	95% HPD
h(m)	0.5–3.5	1.67	1.48–1.86
ρ_t(g/cm^3)	1.1–1.6	1.356	1.35–1.36
c_t(m/s)	1450.0–1550.0	1472.7	1471.3–1474.1
ν	0.0–1.5	1.00	0.88–1.10
ρ_b(g/cm^3)	1.2–1.8	1.55	1.50–1.60
c_b(m/s)	1450.0–1550.0	1465.9	1460.3–1471.0
α(dB/λ)	0.0–0.8	0.38	0.35–0.42

Table 8.3: Numerical values of inversion results for inversion with likelihood for Gaussian errors.

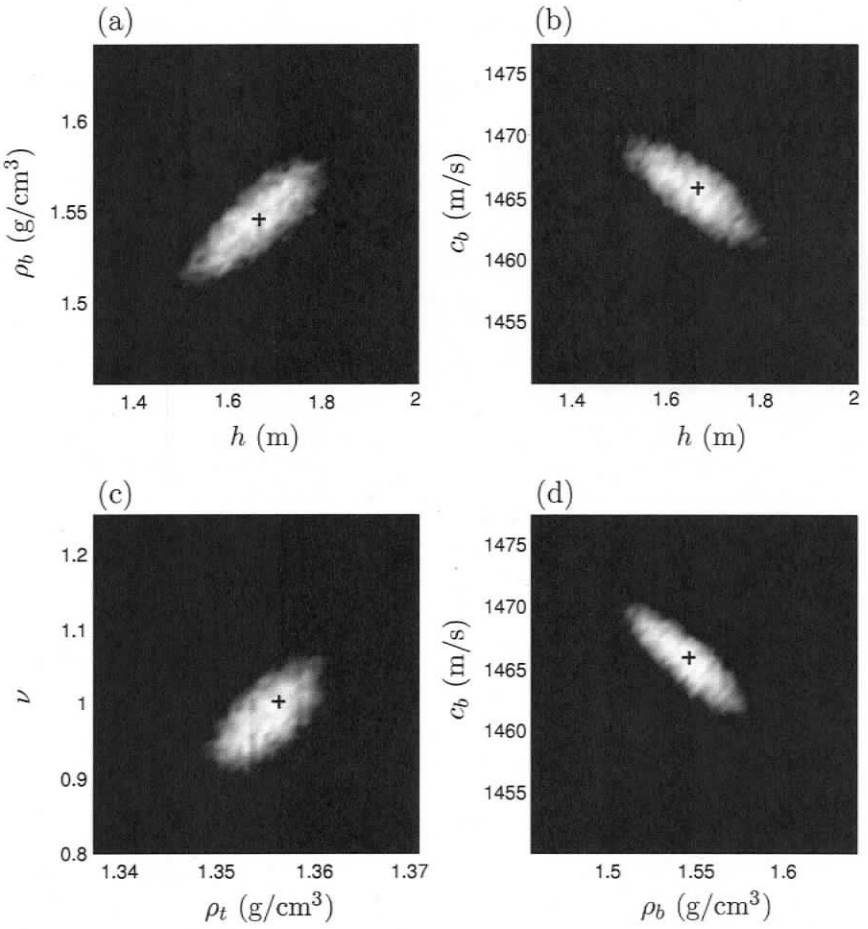

Fig. 8.19: Joint marginal distributions for selected parameter pairs. The black "+" indicates the MAP model.

base sound velocity and layer thickness. The correlations involving h come from the fact that the layer thickness is not constrained by a reflector but acts as a shape parameter. Hence, a strong trade off occurs between density, sound velocity and layer thickness. The correlation between ν and ρ_t in Fig. 8.19 (c) is similar to the above case, since ν is a shape parameter as well that determines the shape of the density profile which goes into the forward model with 10 discrete layers. Therefore, ν constrains the discrete layers below the uppermost value that is given by ρ_t. A higher value of ν produces a smaller gradient in the upper part of the profile. To conserve the average density, a higher ν can be offset by adjusting the top density to a higher value. The coupling between lower sound speed and lower density in Fig. 8.19 (d) is inherent in reflection coefficient calculations as the product of the two forms the impedance of a medium.

At the experiment site, two gravity cores were collected that sampled the uppermost metre of the sediment. The cores were 11 cm in diameter and the core barrel was 1.3 m long. The cores were adjusted until approximately 10 cm of sea water was preserved on top of the core to minimise disturbance of the fragile water-sediment interface. After settling for several hours, compressional sound velocity (at 200 kHz) and density (gamma ray attenuation) were measured. The manufacturer of the core logger estimated the two standard deviation errors in sound velocity estimation to be 6 m/s and 2% in the density.

Figure 8.20 shows a comparison between the inversion results and two piston cores taken at the experiment site. The uncertainties from the manufacturer of the data logger are included on every tenth core value. However, instrumental uncertainties do not represent the total error of the core measurement, which include experimental effects. Sampling the seabed with cores is a complicated procedure and it is difficult to avoid biases due to core compaction, decompression, and temperature changes. Further, piston cores often disturb the sediment close to the water-sediment interface. In particular, the sediments at this location are soft and highly water saturated which makes it difficult to sample the sediment without introducing disturbance.

In Fig. 8.20 the profiles computed from the MAP parameter estimates are shown as solid black lines. The dashed black lines represent 95% HPD intervals (obtained from the sample of models collected by the Gibbs sampler). Finally, the 95% HPD credibility intervals were calculated for a set of discrete layers. The final inversion result shows excellent agreement between the profiles and the core measurements.

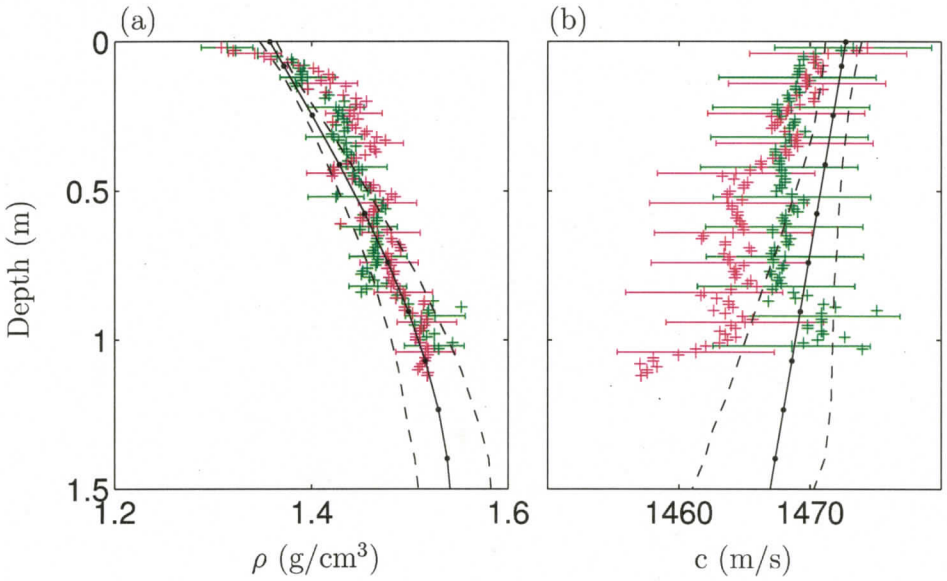

Fig. 8.20: Comparison between final inversion results and core data for two different core sites (red and green). Every tenth datum of the cores includes error bars of two standard deviations. The MAP estimate profile is shown as a solid line, 95% HPD intervals as dashed lines.

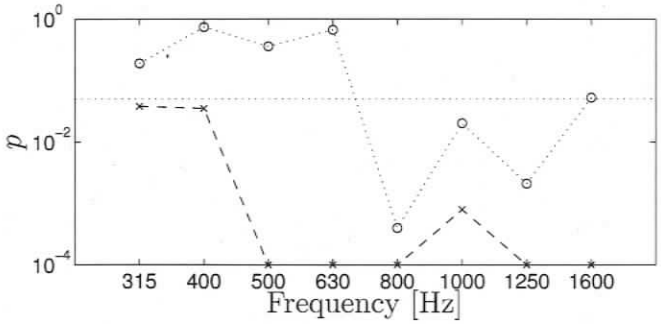

Fig. 8.21: Results of the runs test for the inversion with the assumption of double-exponential error statistics. The dotted line indicates the $p = 0.05$ level of significance. Crosses show the p-values for raw residuals $\mathbf{n}_i$ and circles show the p-values for standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$.

8.4. Inversion with Exponential Distributed Errors

In the previous section, examination of the assumed data error statistics yielded good results for the randomness of the standardised residuals (Figs. 8.14 and 8.17). However, in the case of the KS test for Gaussianity, four out of eight frequencies did not pass the test at the 0.05 level. Further, as shown in Fig. 8.16, the histograms of the data residuals at some frequencies appear more peaked than the theoretical Gaussian distribution with somewhat heavier tails. This suggests that the data errors may be better described by another statistical distribution. This section considers the same measured data, but the inversion is carried out with the assumption of double sided exponential data errors, including covariance estimation for this assumption (see Chapter 5). In examining the residuals computed for the MAP model, the runs test yielded improved p-values at all frequencies once the covariance matrix was applied (Fig. 8.21). The KS test for double exponential statistice produced higher p-values (Fig. 8.22) for residuals at some frequencies and lower values at other frequencies compared to the inversion for likelihood Gaussian errors described in the previous section. In both cases, five of eight frequencies pass the test at a 0.05 level. However, the KS test is not very sensitive to the tails of a distribution. The double-sided exponential distribution features heavier tails than the Gaussian distribution. The histograms of the data residuals (Fig. 8.23) show that the peaks are better represented by the double-sided exponential distribution but the tails are not heavy enough to fully justify this distribution.

Figure 8.24 shows the marginal probability distributions for the inversion with

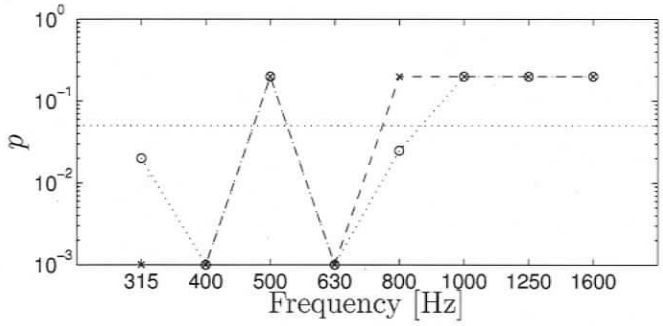

Fig. 8.22: Results of the KS test for the inversion with the assumption of double-exponential error statistics. The dotted line indicates the $p = 0.05$ level of significance. Crosses show the p-values for raw residuals $\mathbf{n}_i$ and circles show the p-values for standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$.

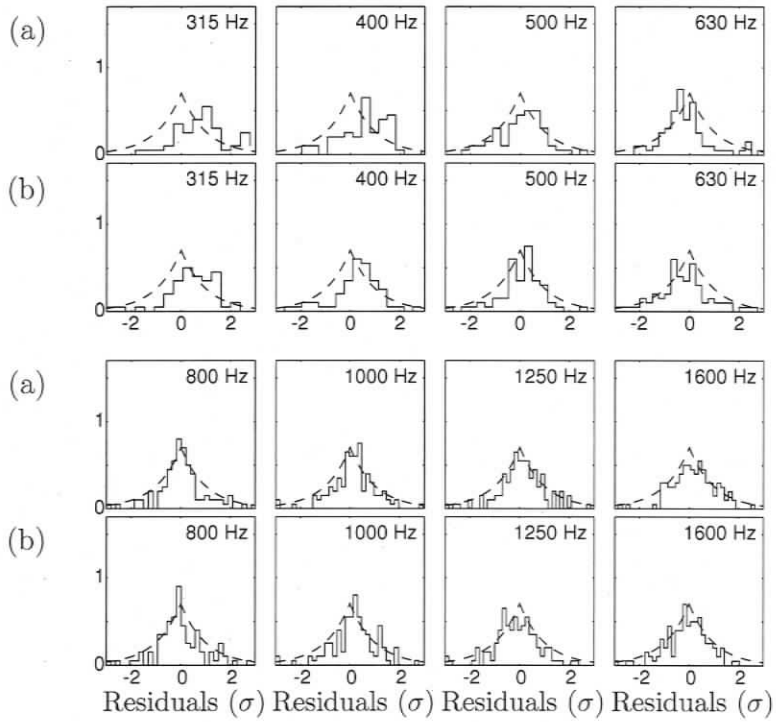

Fig. 8.23: Histograms for (a) raw residuals $\mathbf{n}_i$ and (b) standardised residuals $\tilde{\mathbf{n}}_i^{(2)}$ compared to theoretical double-exponential distributions (dashed lines).

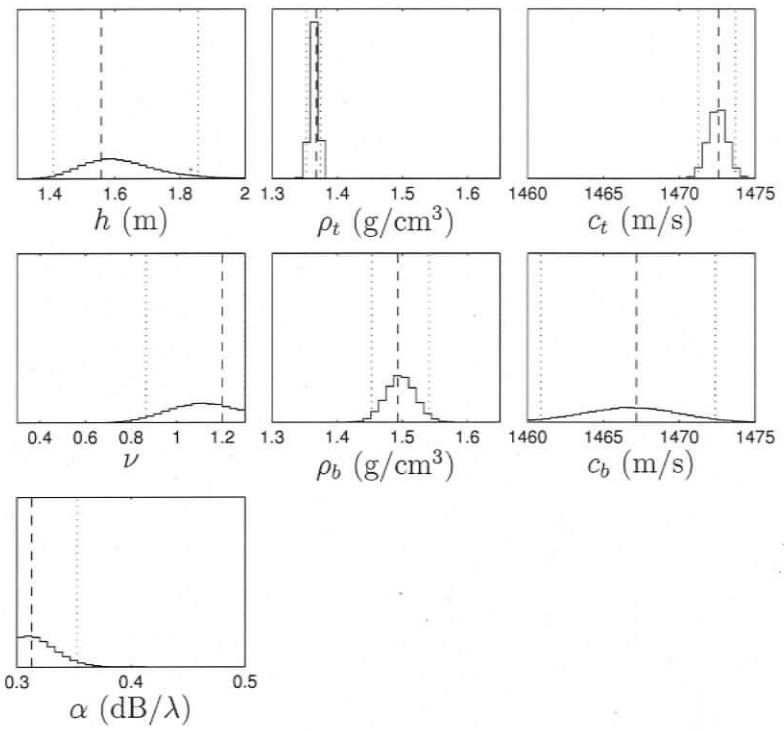

Fig. 8.24: Marginal probability distributions for an inversion that used $\hat{\mathbf{C}}_{(i,2)}^{(\mathbf{d})}$ and the assumption of exponential errors. The MAP estimate is shown as a dashed line, 95% HPD credibility intervals as dotted lines.

double exponential errors assumptions. The density gradient is clearly resolved by the inversion. The inversion results are compared to the core data in Fig. 8.25 with excellent agreement. Numerical values for prior bounds, MAP model estimate and HPD credibility intervals are given in Table 8.4. Overall, the results for the two assumptions of Gaussian and exponential errors agree well with each other. The density values at greater depth in Fig. 8.23 are slightly lower than for the inversion with Gaussian statistics. However, the core values still fall into the range of the estimated uncertainties.

Comparing the results for the assumptions of Gaussian and double-exponential data error statistics provides confidence in the inversion result, as it appears to be more important to incorporate error correlations than finding the exact form for the error distribution (likelihood function). Both inversions yield results that are in good agreement with the core data and are also consistent with each other.

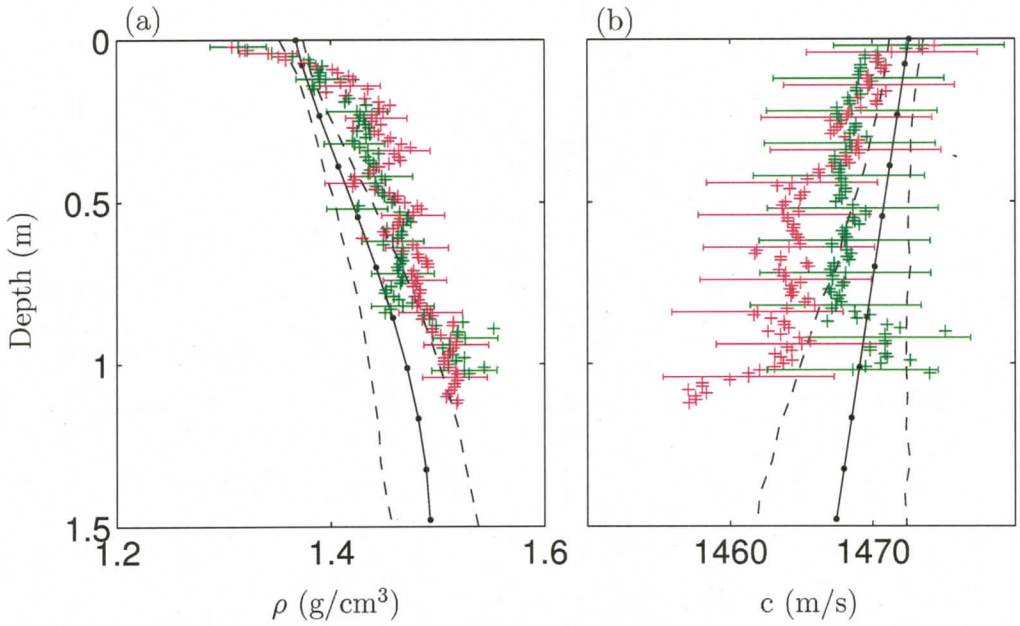

Fig. 8.25: Comparison between final inversion results and core data for two different core sites (red and green) for double-exponential data error statistics. Every tenth datum of the cores includes error bars of two standard deviations. The MAP estimate profile is shown as a solid line, 95% HPD credibility intervals as dashed lines.

Parameter	prior bounds	MAP	95% HPD
h(m)	0.5–3.5	1.56	1.36–1.99
ρ_t(g/cm^3)	1.1–1.6	1.368	1.35–1.38
c_t(m/s)	1450.0–1550.0	1472.6	1470.7–1474.0
ν	0.0–1.5	1.20	0.77–1.47
ρ_b(g/cm^3)	1.2–1.8	1.49	1.43–1.55
c_b(m/s)	1450.0–1550.0	1472.5	1457.9–1474.3
α(dB/λ)	0.0–0.8	0.31	0.26–0.37

Table 8.4: Numerical values of inversion results for inversion with double-exponential error statistics.

8.5. Discussion of Results

This chapter applied the Bayesian inversion approach (Chapters 4 and 5) to bottom loss data collected on the Malta Plateau at a site consisting of water saturated, fine grained sediments. The high quality data featured a prominent angle of intromission and the inversion aimed at recovering high-resolution density and sound velocity structure of the uppermost ($\sim$ 1.5 m) sediment layer, the transition layer. To limit the number of parameters, a linear sound velocity gradient was used. The density profile was defined through an upper and lower density and a shape parameter that determined the gradient between the two values. The semi-infinite basement was fixed to the values at the base of the gradient layer.

Early inversion results for this data indicated that the errors are correlated. A simulated experiment was used to examine the effect of including data covariance matrices in the inversion when correlated data errors are present. It was shown that data covariances can be estimated and that it is essential to account for covariances to obtain correct inversion results.

The measured data were then inverted assuming a likelihood function based on Gaussian distributed errors and full error covariance matrices estimated for non-uniform data sampling. The statistical assumptions were quantitatively and qualitatively examined *a posteriori*. The Bayesian inversion results were then interpreted in terms of MAP estimates, parameter cross correlations and credibility intervals for gradient profiles. The inversion results showed excellent resolution and high sensitivity to detailed density information. The recovered density and sound velocity structures were compared to piston cores taken at the experiment site and showed good agreement.

The complete inversion was also performed for a likelihood function based on the assumption of exponentially distributed errors which yielded results that were consistent with the results based on the Gaussian errors assumption. In both cases, the *a posteriori* validation of the statistical assumptions yielded similar results, which is regarded an indication for the stability of the inversion.

Seabed density gradients have not been recovered from reflection coefficient data with such resolution previously. The reason for the high information content of this data set is largely dependent on the good resolution of the angle of intromission. The angle of intromission is a prominent feature in the data that contains information

about densities which are well constrained by the location (angle) of this feature. However, it should be kept in mind that the estimates for the uncertainty bounds are for the model of the environment, not the environment itself. This means that it is possible that the true density can be outside these bounds.

In principle, the method is also applicable to deeper layers and the parameterisation can be embedded into a more complex layered model. However, the computational expense is considerable and may or may not justify application in a complex model with many sediment layers. For sediments that do not feature an angle of intromission, the sensitivity to sediment density structure is expected to be lower than in the example shown here. Dosso and Holland (2006) compared the inversion of angle of intromission data to the inversion of critical angle data for a simple semi-infinite half space seabed and found the critical angle data to be less informative about density.

9. Reflectivity from Ambient Noise

9.1. Introduction

An interesting and promising recent alternative to the single bounce reflectivity measurements considered in Chapters 6–8 of this thesis is the extraction of reflectivity measurements from oceanic ambient noise (Harrison and Simons, 2002; Harrison, 2004; Siderius and Harrison, 2004). This approach is based on measuring the ambient noise (AN) field on a vertical line array (VLA) of hydrophones (Fig. 9.1). This strikingly simple experiment can recover seabed reflection coefficient data without an active source, simply by separating the up-going and down-going noise fields with beam-forming. This chapter briefly considers this method in a simulated study designed to quantify the geoacoustic resolving power of ambient noise data. The assumption in this section is that the ambient noise field is mainly generated by wind and breaking waves, forming a sheet source at the ocean surface. In this case, the reflection coefficient (or bottom loss) as a function of angle and frequency can be deduced from ambient noise as described below (Harrison and Simons, 2002).

Considering a vertically stratified volume of sea water bounded by the sea surface and the sea bottom (Fig. 9.2), with a sheet source located just below the surface, the flux through this plane can be analysed, ignoring horizontal fluxes because of

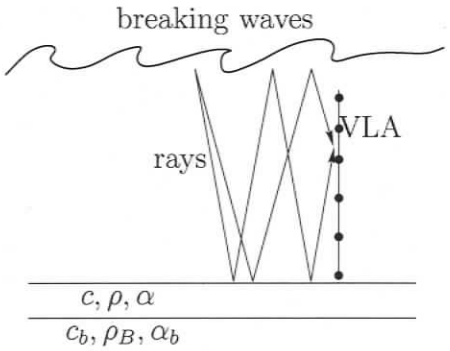

Fig. 9.1: The ambient noise experiment uses surface wave generated noise recorded at a vertical line array and separates the upward and downward fields by beam-forming.

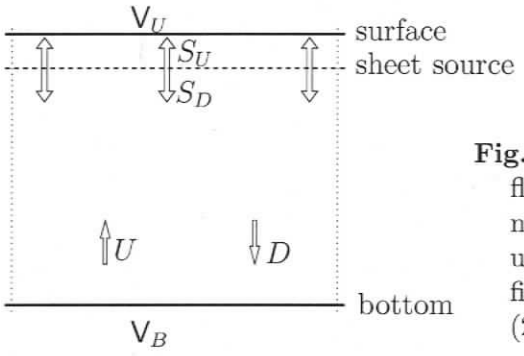

Fig. 9.2: Deducing plane wave reflection coefficient from ambient noise is done by comparing the up-going (U) and down-going (D) fields, after Harrison and Simons (2002).

horizontal uniformity. The sheet source emits the upwards flux S_U and the downwards flux S_D. Beneath the source depth, the downward field D can be written as

$$D = S_D + \mathsf{V}_U(S_U + U), \tag{9.1}$$

with V_U being the surface power reflection coefficient, and U the upward field. If V_B is the bottom power reflection coefficient, the upward field is

$$U = D\mathsf{V}_B. \tag{9.2}$$

With an effective source strength $S = S_D + V_U S_U$, we can write

$$D = S/(1 - \mathsf{V}_U\mathsf{V}_B) \tag{9.3}$$

$$U = V_B S/(1 - \mathsf{V}_U V_B) \tag{9.4}$$

$$\Longrightarrow V_B = \frac{U}{D}. \tag{9.5}$$

The upward field always has one more bottom reflection than the down-going field. Since a sheet source is considered, the field consist of plane waves and plane wave reflection coefficients are measured.

To obtain the upward and downward fields, the noise field is measured with a vertical line array and beamforming is applied to separate the upward from the downward fields. The ratio of the upward over downward fields then gives the plane wave reflection coefficient (Fig. 9.3). This represents a local estimate from within one ray cycle of the array. At grazing angles greater than the critical angle, the data show fringe patterns due to interference of reflections from different horizons. The reflection

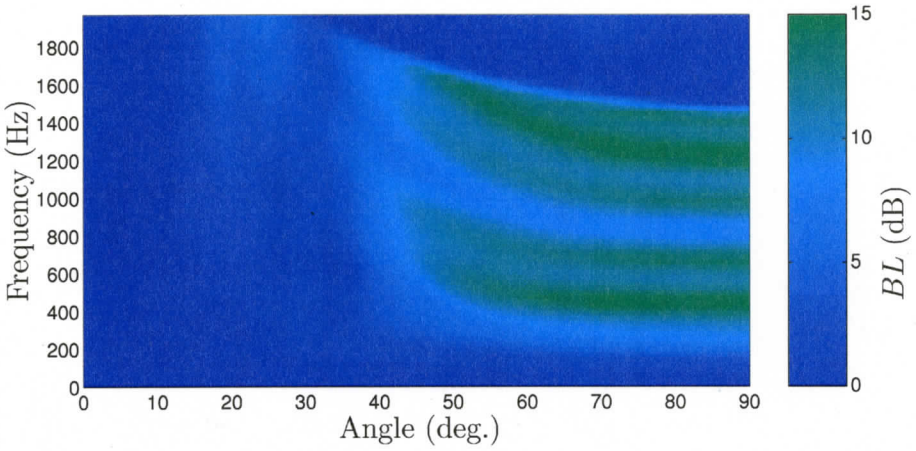

Fig. 9.3: Example of bottom loss vs. angle and frequency for a two layers over a half space model, as generated by the ambient noise forward model.

coefficient data can be used to invert for geoacoustic parameters of the sub-bottom.

9.2. The Forward Model

Measured data are typically collected as a time series by a vertical line array with equal hydrophone spacing (Fig. 9.1). Frequency domain beam forming and power averaging are applied (Harrison and Simons, 2002). The beam angle is mapped onto the sea-floor by Snell's law. For the inversion, this process needs to be emulated by calculating the plane wave response of the seabed for multiple layers. Starting at the lower-most interface, the reflection coefficient for the sea-bed is calculated recursively according to Sec. 6.4.1. This reflection coefficient is then smeared out to simulate the effects of the beamforming process (Harrison and Simons, 2002). The geoacoustic model is parametrised as a set of fluid, homogeneous, layers. Each layer is defined by layer thickness h, sound velocity c, density ρ, and attenuation coefficient α.

9.3. Geoacoustic Resolution Study

This section analyses the potential abilities of ambient noise inversion in terms of depth resolution and the ability to resolve discrete layers using simulated data (Dettmer et al., 2004b). The data is simulated using the forward model described above and unbiased Gaussian distributed random noise with constant standard deviation (1dB)

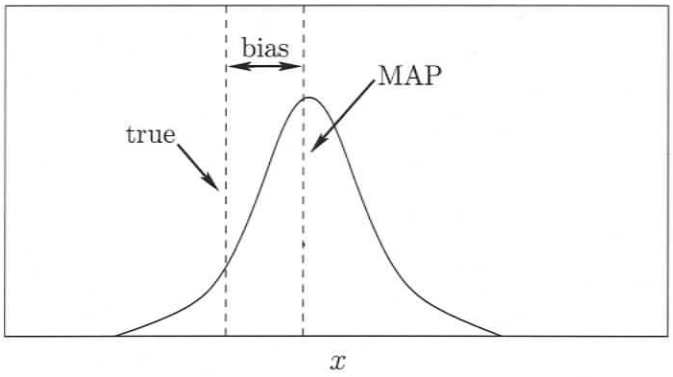

Fig. 9.4: The bias for calculating the standard deviation about the true parameter

across frequency is added to the simulated data. Data are computed at 12 frequencies from 370–1400 Hz. The model parameters that were used to compute the simulated data were chosen based on Harrison and Simons (2002). Fig. 9.3 shows a typical plot of the bottom-loss for a two layer over a half space model. The critical angle is $\sim 30°$ and a fringe pattern is found at higher angles. The low values at high angles and frequencies are due to the upper frequency limit of the array (Eq. 9.7).

The minimum layer thickness resolvable by the ambient noise data is considered by using a single sediment layer of thickness $h_1^{(t)}$ over a half space. The layer thickness $h_1^{(t)}$ is varied and for every thickness considered, simulated data are generated and inverted using Gibbs sampling as discussed in Chapter 4. The resulting PPDs are then interpreted in terms of marginal probability distributions. Standard deviation estimates σ for each parameter are calculated from the marginal distributions. The standard deviations are then corrected for their bias about the true model parameters according to

$$\sigma_x^{(t)} = \sqrt{bias^2 + \sigma_x^2}\,, \tag{9.6}$$

where *bias* is the difference between MAP estimate and true parameter value as illustrated in Fig. 9.4, and σ_x is the standard deviation of some model parameter x as obtained from the marginal probability distributions. Therefore, $\sigma_x^{(t)}$ gives the standard deviation of parameter x about its true value. Figure 9.5 shows the standard deviation of the upper layer model parameters about the true model parameters as a function of the true thickness $h_1^{(t)}$. Hence, each datum in Fig. 9.5 represents the result of one Bayesian inversion. These plots can be interpreted as a measure of when full resolution is obtained. Once the standard deviation does not significantly change with true layer thickness anymore, full resolution is obtained. In this case, all parameters

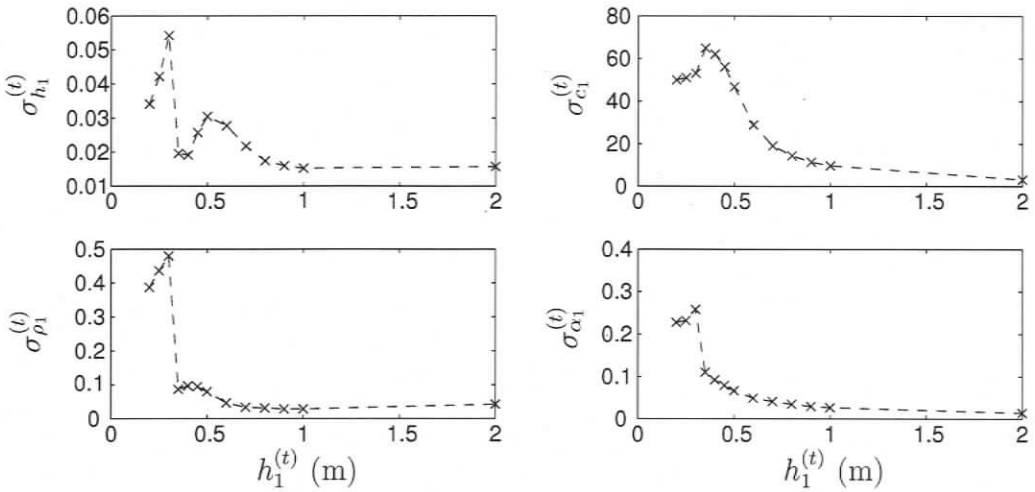

Fig. 9.5: Standard deviations of the estimated model parameters of the first layer about the true model parameter vs. the true layer thickness $h_1^{(t)}$.

reach full resolution at about $h_1^{(t)} = 0.7$ m. For thinner layers, the method cannot resolve the sound velocity. Densities already reach full resolution at $h_1^{(t)} \approx 0.4$ m.

In the following, two sediment layers over a semi-infinite halfspace are considered and inversion results will be discussed in terms of marginal probability distributions and parameter standard deviations about true values, to quantify resolution with depth. The simulated data are computed in the same way as described above. To be consistent with the results from Fig. 9.5, a layer thickness of $h_2 = 1$ m (clearly above the resolution limit of 0.7 m) was chosen for the lower layer. Gibbs sampling was then applied to the simulated data in multiple inversions, where the thickness of the upper layer ($h_1^{(t)}$) was varied and all other parameters were held fixed. The resulting PPD for one of these inversions ($h_1^{(t)} = 2$ m) is shown in terms of marginal probability distributions in Figure 9.6. All parameters are generally well resolved and are in good agreement with the true values (there are no significant biases and the true parameter values lie within the credibility intervals). Numerical values for MAP model and HPD credibility intervals are given in Table 9.1. This verifies that the method can resolve both layers well and the penetration depth versus resolution limit is considered next.

Several Bayesian inversions with increasing upper layer thicknesses h_1 were conducted until the parameters of the second layer started to degrade. For every inversion result, marginal probability distributions were computed. Standard deviations

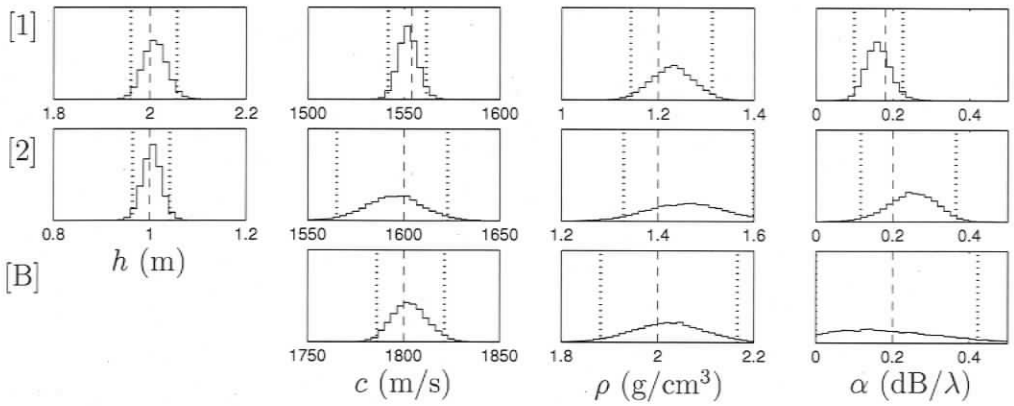

Fig. 9.6: Estimated marginal probability distributions for the data shown in Fig. 9.3. True values are shown as dashed lines, 95% HPD credibility intervals are given as dotted lines.

Parameter	95% HPD	Parameter	95% HPD
h_1(m)	1.97–2.05	h_2(m)	0.97–1.03
c_1(m/s)	1540.00–1560.00	c_2(m/s)	1566.67–1623.33
ρ_1(g/cm^3)	1.15–1.31	ρ_2(g/cm^3)	1.32–1.59
α_1(dB/λ)	0.09–0.22	α_2(dB/λ)	0.13–0.37
c_b(m/s)	1785.00–1820.00	ρ_b(g/cm^3)	1.87–2.15
α_b(dB/λ)	0.00–0.42		

Table 9.1: Highest probability density intervals for the estimated marginal distributions shown in Fig. 9.6.

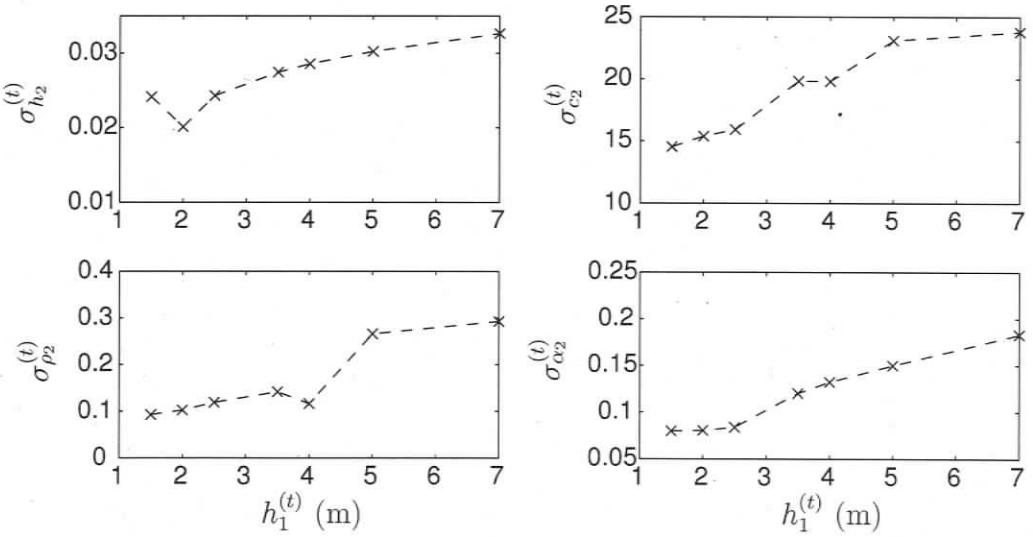

Fig. 9.7: Standard deviations of the estimated model parameters of the second layer about the true model parameters vs the true layer thickness $h_1^{(t)}$ of the first layer.

about the true model parameters were derived from these marginal distributions and are plotted in Fig. 9.7 as a function of the true upper layer thickness $h_1^{(t)}$. At $h_1^{(t)} = 5$ m, the density in the lower layer is no longer resolved. Figure 9.8 shows marginal distributions, true values, and 95% HPD credibility intervals for the case of $h_1^{(t)} = 5$ m. Here the HPD credibility intervals are considerably wider than in Fig. 9.6, particularly for the densities. Resolution of the density in the second layer is completely lost. However, the inversion is still somewhat sensitive to the density in the basement layer. Layer thicknesses are still resolved, which is consistent with them being mainly represented by the fringe spacing (Harrison and Simons, 2002), which is a significant feature in the data. In other situations (e.g., a model with a softer or harder bottom) the resolution could be different.

The above findings of the minimum resolvable layer thickness being 0.7 m and the maximum depth of penetration being $\sim$ 5 m are consistent with what would be expected for the array used in the forward model (Harrison and Baldacci, 2004). Beams were computed for a 32 element array of 16 m total length and 0.5 m hydrophone spacing. The upper frequency limit for such an array is constrained by grating lobes and can be written as

$$f_U = \frac{2f_0}{1 + \sin\theta} \tag{9.7}$$

where $f_0 = c/2a$ is the design frequency of the array with a being the hydrophone spacing, c the sound velocity, and θ the grazing angle. The lower frequency limit depends on the beam angle resolution and can be written as

$$f_L = \frac{2f_0}{N \sin\theta} \tag{9.8}$$

where N is the number of hydrophones. Layer thickness is mainly characterised by the fringe spacing $\Delta f = 1/TWT$, where TWT is the two way time. The thinnest detectable layer should have a fringe spacing $\Delta f \leq f_U$ which results in a layer thickness $h \geq a$. Hence, in this case the minimum detectable layer thickness would be expected to be $\sim$ 0.5 m, which is consistent with the results shown in Fig. 9.5.

Harrison and Baldacci (2004) also found that angle and frequency resolution get poor for a model consisting of one layer over a half space, where the layer thickness exceeds $h_1 = 4$ m. This is approximately consistent with the result in this study, where resolution was lost at $h_1 = \sim 5$ m (see Fig. 9.7).

To consider AN inversion for complex, multiple-layer environments, a Bayesian

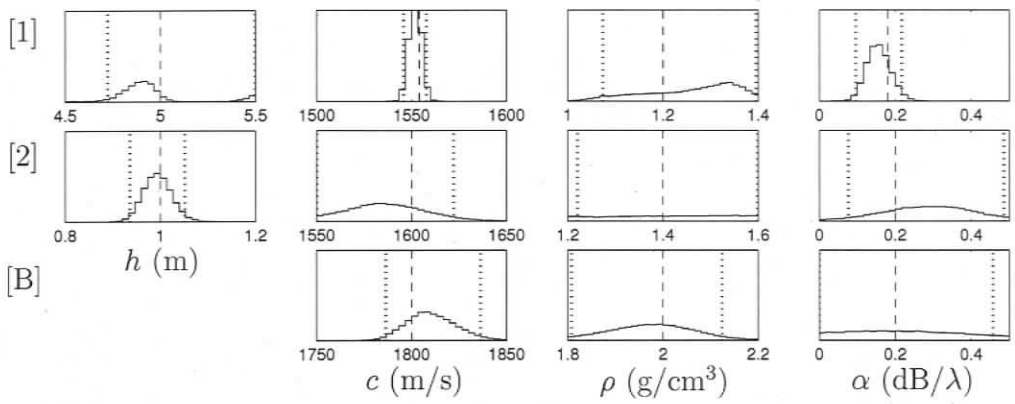

Fig. 9.8: Estimated marginal probability distributions for a 5m and a 1m layer over a half space. True values are shown as dashed lines, 95% HPD credibility intervals are given as dotted lines.

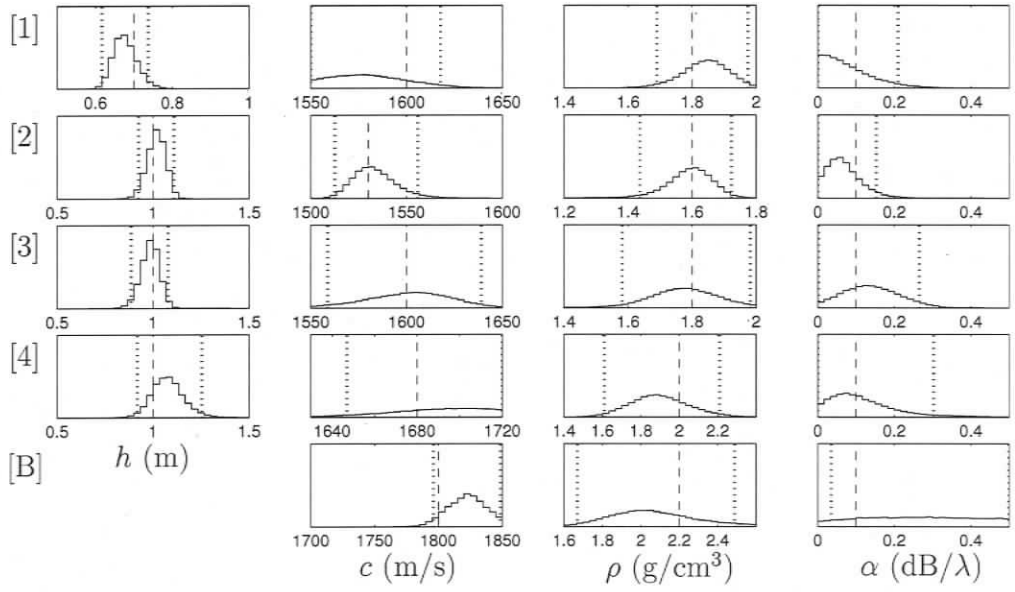

Fig. 9.9: Estimated marginal probability distributions for a model with three homogeneous layers overlying a half space. True values are shown as dashed lines, 95% HPD credibility intervals are shown as dotted lines.

inversion was applied to simulated data for a model with four layers over a semi-infinite halfspace. Figure 9.9 shows estimates of marginal probability distributions for the resulting PPD. All parameters are resolved by the inversion. The first layer has a thickness of $h_1 = 0.7$ m which is at the lower limit of what can be resolved. Thus, the sound velocity has a fairly wide distribution and partly shifts to the sound velocity of the next layer. This example shows that the AN inversion is capable of resolving relatively complex sub-bottom structures.

10. Summary and Discussion

10.1. Bayesian Inversion Methodology

This thesis developed a fully nonlinear Bayesian inversion approach for seabed reflectivity data that included rigorous examination of statistical assumptions. Seismo-acoustic time traces are recorded by a single receiver moored in the lower third of the water-column and an impulsive source is towed along tracks crossing the receiver while transmitting at fixed range increments. The data are inverted with a Bayesian joint time/frequency domain inversion that uses results from the time domain inversion as prior information for the frequency domain inversion.

In the time domain, travel times were picked on the seismo-acoustic traces and inverted for layer thicknesses and sound velocities for several layers. The forward model consists of an efficient two point ray tracing code that provides travel times. In the inversion, particular attention was paid to picking errors. Unknown time offset parameters for each reflector were included in the inversion to account for possible biases (constant over range). Further, a covariance matrix is estimated for each reflector and included in the inversion to treat correlated data errors. Since typical picking errors often vary with range, the covariance matrices are corrected to a non-Toeplitz form that accounts for non-stationary errors. A simulation study showed the importance of correctly accounting for data errors and biases. The time offset parameters introduced to account for biased errors caused strong parameter cross correlations, which slowed convergence of the Gibbs sampling algorithm considerably during burn-in. To obtain an effective rotation matrix with low computational effort, a linearised parameter covariance matrix estimate was used to initialise sampling. Further, the rotation matrix was regularly updated during the burn-in phase.

For the frequency domain inversion, the seismo-acoustic data are processed to yield reflection coefficients as a function of frequency and grazing angle. The data are frequency averaged over finite frequency bands to ensure a high signal to noise ratio.

Two different forward models were developed to model the reflection coefficients. First, a recursive approach was used to calculate the plane wave reflection coefficient. However, in some cases, the plane wave assumption proved to be insufficient. In particular, more complex environmental models could not be resolved by the plane wave model at greater depth, causing misleading results. Therefore, a spherical wave model was developed based on the computation of acoustic fields under the assumption of point sources. The acoustic fields were calculated with a wavenumber integration model (OASES) and include full wave field effects. This approach is very general and can incorporate complex phenomena into the forward model such as arbitrary depth dependence in the water column and seabed shear and interface roughness.

The forward model based on wavenumber integration is computationally intensive. To save computational time, frequency averaging (required by the data processing to reduce noise), was replaced by equivalent range averaging in the forward model. This saves considerable (one order of magnitude) computational time, since the wavenumber integration is already carried out with small range steps. To address complicated multi-layered seabeds with the computationally intensive forward model, both the optimisation and the integration of the PPD were implemented on massively parallel computers. The code was run on up to 80 CPUs with an approximately linear reduction in computation time with number of CPUs.

Given the large number of parameters in some of the reflection coefficient inversions, and to reduce non-uniqueness in certain cases, a layer packet stripping approach was developed to separate different layer packets. The seismo-acoustic data were time windowed into packets of increasing depth, and the packets were inverted sequentially with the results from one packet inversion used as prior knowledge for the subsequent packet inversions. The prior information consisted of marginal probability distributions rotated to principal components to account for parameter correlations (to first order). This layer packet stripping was especially useful for the PPD optimisation, where computational performance was a limiting factor. Due to the small granularity of the optimisation code, it did not parallelise as well as the Gibbs sampling integration. The layer stripping approach helped to keep the computational expense to a reasonable amount.

The nonlinear Bayesian inversion is based on certain assumptions about the form of the data error distribution, and incorrect assumptions can affect the credibility of the results. Since data errors are not directly accessible, data residuals (the difference

between the best fit model and the measured data) were used to evaluate the statistics of the data errors. Analysing data residuals showed that the errors in the travel time and reflection coefficient inversions are often strongly correlated. This was addressed by including correlations in the inversion through data covariance matrix estimates. Depending on the sampling increments and the stationarity of the residuals, Toeplitz or non-Toeplitz covariance matrices were used. The statistical assumptions were then validated *a posteriori* to ensure meaningful and rigorous inversion results. For the validation, the form of the residuals distribution was quantitatively tested with the Kolmogorov-Smirnov test, and randomness was quantified with the runs test. To qualitatively check the assumptions, histograms and autocorrelation functions of the residuals were plotted at all frequencies for all inversions. Simulated experiments were used to show that neglecting error covariances can cause misleading inversion results with unrealistically high confidence.

While validating statistical assumptions about data errors is important to support the inversion results, it is difficult to test for systematic errors and biases. This thesis developed realistic computer simulations to emulate many aspects of the actual experiment. The simulations were largely independent from the forward models used in the inversion and permit the detection of weaknesses in the forward models. This thesis illustrated two cases where systematic errors caused strongly biased inversion results. In the travel time inversion, it was important to include travel time offset parameters in the inversion to account for biased travel time picks. Neglecting the travel time bias caused strongly biased inversion results. In the reflection coefficient inversion, full wave field effects (spherical reflection coefficients) had to be included under certain conditions. While the data could be fit well by plane wave reflection coefficients, the inversion results were strongly biased and the recovered parameter uncertainty estimates were misleading. By taking the full wave field effects into account, complicated environmental models could be correctly recovered.

The inversion results obtained here were interpreted in terms of optimal parameter estimates, marginal probability distributions, credibility intervals and parameter correlations. Final results were presented as depth profiles of sound velocity and density that are extracted from the PPD.

As an alternative experimental technique to measure local reflection coefficients, ambient noise was briefly considered. For simulated data, the resolution power and limits of ambient noise data were examined. The Bayesian inversion was used to

quantify resolution limits as a function of depth and layer thickness. It was shown that even though the data are smeared out due to beam forming, the method has considerable potential to resolve detailed seabed structure.

10.2. Application to Measured Data

The first application used data that were collected on the Malta Plateau in the Mediterranean Sea. The seismo-acoustic data sample the upper 20–30 meters of the sediment and clearly show several reflectors. The complete joint time/frequency domain approach was applied to these data. First, the travel time inversion yielded layer thicknesses and sound velocities. The layer thickness information was then used as prior information in the reflection coefficient inversion. Even though the data contained seven strong reflectors to about 20 m depth, only the layers of the upper few meters were successfully inverted for. Below this, complicated three-dimensional geological structures affected the data in a way that could not be captured by the forward model. However, the inversion for the upper layers were in good agreement with a piston core that was taken at the experiments site.

The second application used measured data collected at another site on the Malta Plateau, but here the data was time windowed to only contain the upper 1.5 m of the sediment. This section contained no prominent reflectors but included the transition layer. The transition layer features strong density and sound velocity gradients in the uppermost part of the seabed. The study aimed at resolving these density and sound velocity gradients with precision. The data were characterised by a strong angle of intromission that was clearly resolved by the data across the frequency band of 300–1600 Hz. Due to the shallow depth of the data and the frequency range used in the inversion, the plane wave reflection coefficient model was sufficient. Sound velocities were parametrised with a linear gradient that was discretised with 10 layers. The density gradient used the same discretisation but was parametrised so as to allow different shapes including linear and curved profiles.

The results showed that the data were able to resolve seabed structure to a very high degree and could recover the shape of the density profile. The results were compared to two cores taken at the experiment site and showed good agreement. The high resolution is related to the high quality of the data as well as to the angle of intromission that constitutes a highly informative feature in the data.

11. Conclusion

A key point in this thesis is that in inverse problems, reliable estimation of posterior probabilities (parameter uncertainties) require a physically meaningful specification of the likelihood function (quantifying data error uncertainties). This involves at least three components. First, an appropriate form for the error distribution (e.g., Gaussian or exponential) must be chosen and validated *a posteriori*. Second, if significant data error correlations are present, they must be quantified and included in the inversion by means of data covariance matrices. Third, possible error biases must be accounted for. All three aspects are straightforward to implement in the numerical Bayesian approach used in this thesis and were applied in the inversion of measured data.

Further, this thesis showed that geoacoustic seabed reflection data has a high information content about geoacoustic parameters. The high angular resolution allows for recovering detailed density and sound velocity profiles with vertical resolution in the order of 10^{-1} m at the frequencies considered (300–2000 Hz). This scale of resolution is important to many underwater acoustic applications (e.g., modelling of the underwater sound field) as well as general geoscientific questions (e.g., understanding sedimentation and quantifying porosity in sediments).

The methodology developed here is applicable to recover layered geoacoustic models. It was also shown that high resolution continuous profiles without hard reflectors can be recovered when an appropriate parametrisation is used. The reflection coefficients for continuous profiles are computed for a discrete layered model where the layers are constrained to represent a general profile shape. With this method, it was possible to recover detailed structure of the transition layer that included the uppermost metre of the seabed. Density gradients were resolved with higher resolution than heretofore possible.

Recovering layer thickness information from reflection coefficient data is often problematic due to interference effects. The multi-modal marginal probabilities that result from interference effects complicate determining the actual layer thickness. It

was shown that by using picked travel times in an initial ray tracing inversion, layer thicknesses can be constrained to avoid ambiguous marginal probabilities in a subsequent reflection coefficient inversion. The Bayesian reflection coefficient inversion was constrained by applying the results from the time domain as prior knowledge.

To address complicated layered models with up to 27 parameters in the reflection coefficient inversion, a layer packet stripping approach was developed. The seismic traces were time windowed into packets of increasing depth and then processed to yield reflection coefficient data. The results of each packet inversion were used as prior knowledge for the subsequent packet. To account for correlations between parameters, rotated priors were used.

Finally, it was shown that full wave effects can be of significant importance in recovering meaningful profiles of geoacoustic parameters, particularly for low frequencies and depths of more than a few metres below the seafloor.

A future goal is to apply the inversion methodology developed in this thesis to a variety of data sets to resolve spatial variability of geoacoustic parameters in seabed sediments. To be able to distinguish between uncertainty in the inversion and actual variation in the sediment, a rigorous data error treatment as outlined in this work is cruicial. The high geoacoustic resolution that was obtained for the data considered in this thesis is promising for quantifying spatial variability.

A. Ray-Series and Ray Tracing

This chapter develops the theory for the ray-series and ray tracing that is used for the travel time inversion in chapter 6.2. The derivation is done for the general elastic case for isotropic media and is then simplified for the acoustic case without shear. The foundation of stress and strain theory is the *elastodynamic equation*. From this equation, the eikonal and transport equations, which are the basis of ray methods, can be derived by a high frequency approximation. The eikonal equations yield the kinematic ray tracing (KRT) system of equations which is used to trace rays. The transport equation can be used to derive amplitudes along rays by calculating the geometrical spreading through dynamic ray tracing. Since amplitudes are not needed for the travel time inversion, the derivation will be omitted. The derivation follows Červený (2001) and Lay and Wallace (1995). Einsteins summation convention is used, hence, if the same index appears twice, summation is done over that index. Further, a comma in the index indicates differentiation with respect to the index following the comma.

A.1. The Elastodynamic Equation

Solid bodies act largely elastic if a wave propagates through them, meaning that the displacement is in the linear part of the stress-strain diagram (exceptions can for example be found close to seismic sources). The displacement of all particles in such a continuum can be described by a vector field $\mathbf{u}(\mathbf{x}, t)$. $\mathbf{u}(\mathbf{x}, t)$ describes the internal elastic deformation of a macroscopic solid caused by stress. The deformation is a superposition of elongations and angular distortions (Fig. A.1):

$$\varepsilon_{normal} = \lim_{\Delta s \to 0} \left(\frac{\Delta s' - \Delta s}{\Delta s} \right) \quad \text{und} \quad \varepsilon_{shear} = \frac{1}{2} \lim_{\substack{\Delta s_1 \to 0 \\ \Delta s_2 \to 0}} \left(\frac{\pi}{2} - \theta' \right). \tag{A.1}$$

All deformed quantities are primed. θ' is the angle that results from the deformation

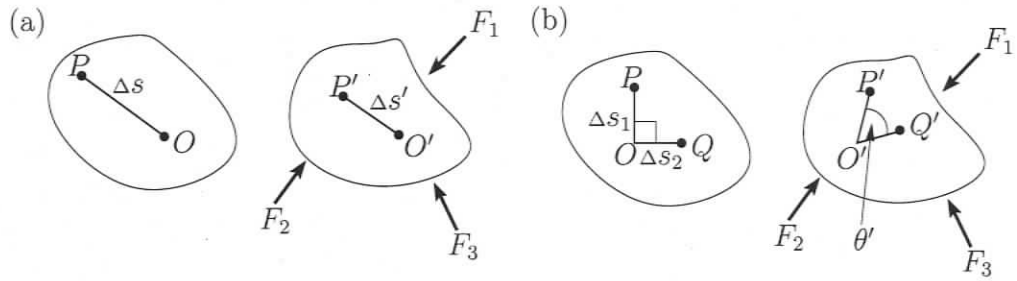

Fig. A.1: Elastic deformation through stress, (a) elongations and (b) angular distortion.

of an angle of 90°. To describe deformation in a three dimensional space three components are needed to describe the longitudinal deformation and another six components to describe angular distortion (two angular distortions per ordinate to describe the relative change with respect to the other two ordinates). These nine quantities define the *strain tensor* and it can be shown that (Lay and Wallace, 1995),

$$\varepsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) . \tag{A.2}$$

If external forces are applied to a continuum, contact forces are generated within the medium. Hence, the three normal planes ΔA_i of the coordinate system experience the components of the force $\Delta\mathbf{F} = (\Delta F_1, \Delta F_2, \Delta F_3)^\top$. This defines the *stress tensor* as

$$\sigma_{ij} = \lim_{\Delta A_i \to 0} \frac{\Delta F_j}{\Delta A_i} .$$

Since both tensors are symmetric, $\varepsilon_{ij} = \varepsilon_{ji}$ and $\sigma_{ij} = \sigma_{ji}$ hold. The *elastodynamic equation* for a perfectly elastic medium is given by

$$\sigma_{ij,j} + f_i = \rho u_{i,tt} , \tag{A.3}$$

where the f_i are volume forces (the *source term*). Since most deformation relevant to seismology is in the linear part of the stress-strain diagram, stress and strain can be related through *Hooke's Law*

$$\sigma_{ij} = c_{ijkl}\varepsilon_{kl} , \tag{A.4}$$

Where c_{ijkl} is the elastic tensor, a tensor of fourth order. This law is valid for arbitrary media, including anisotropic media where the elastic tensor contains 21 independent

components. The elastodynamic equation then becomes

$$\left(c_{ijkl}u_{k,l}\right)_{,j} + f_i = \rho u_{i,tt}\,.$$

For isotropic media, the number of independent parameters in the elastic tensor reduces to nine and can be written as

$$c_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right)\,,$$

where δ_{ij} is the Kronecker-symbol with

$$\delta_{ij} = \begin{cases} 0 & \forall i \neq j \\ 1 & \forall i = j \end{cases} \qquad i,j = 1,2,3$$

and λ and μ are the Lamé parameters. μ is the *shear modulus* and is zero in the acoustic case. Applying the isotropic case to Eq. A.4 and using Eq. A.2 and $\varepsilon_{kk} = u_{k,k} = \nabla\mathbf{u}$ yields

$$\sigma_{ij} = \lambda\delta_{ij}u_{k,k} + \mu\left(u_{i,j} + u_{j,i}\right)\,, \tag{A.5}$$

Hooke's law for the isotropic case. Substituting Eq. A.5 into Eq. A.3 yields the elastodynamic equation for isotropic media

$$\left(\lambda+\mu\right)u_{j,ij} + \mu u_{i,jj} + \lambda_{,i}u_{j,j} + \mu_{,j}\left(u_{i,j} + u_{j,i}\right) + f_i = \rho u_{i,tt}\,. \tag{A.6}$$

To solve the elastodynamic equation (Eq. A.6), the displacement $\mathbf{u}(\mathbf{x},t)$ is expanded into the **ray-series**. The ray series is an infinite series that is asymptotic for the angular frequency ω with $\omega \to \infty$:

$$u_i\left(x_m,t\right) = \sum_{n=0}^{\infty} \frac{U_i^{(n)}(x_m)}{(-\imath\omega)^n} e^{-\imath\omega\left(t-\tau(x_m)\right)}\,. \tag{A.7}$$

The first term in the sum is the amplitude term that is proportional to inverse powers of angular frequency. $U_i^{(n)}$ is the vector amplitude. The eikonal $\tau(x_m)$ represents wavefronts for $\tau(x_m) = const.$. $U_i^{(n)}(x_m)$ and $\tau(x_m)$ only depend on spatial coordinates. Hence, the spatial and time dependence of the displacement were *separated.* Further, since the ray-series is inversely proportional in powers of frequency, the *lead-*

ing term, $n = 0$, of the series is sufficient for high frequencies. The leading term is given by

$$u_i(x_m, t) = U_i^{(0)}(x_m) e^{-\imath\omega(t-\tau(x_m))} . \tag{A.8}$$

This term is also called the zeroth order approximation of the ray-series.

So far, it has not been defined what *high frequency* means. The term needs to be understood in a relative sense as follows. A frequency ν with wavelength $\Lambda = v/\nu = 2\pi v/\omega$, where v is the wave velocity in the medium, is sufficiently high for the approximation, if the spacial variation of physical parameters over one wavelength is negligible compared to the absolute values of the parameters. The physical parameters of interest are density ρ, sound velocity v and the elastic tensor (Lamé parameters). The same relation has to also hold for the vector amplitude and for the slowness vector $\tau_{,i}$ (defined later) and the three relationships can be written as

$$\Lambda \ll \frac{U_i}{\mid \nabla U_i \mid} , \qquad \Lambda \ll \frac{v}{\mid \nabla v \mid} , \qquad \Lambda \ll \frac{\tau_{,i}}{\mid \nabla \tau_{,i} \mid} .$$

The last inequality constrain means that no strong changes in direction are allowed, hence does not allow for large ray curvature. Since all inequality constrains contain gradients in the denominator, the dimension of all right sides are lengths. Hence, the *characteristic length* L can be defined as the minimum of the right sides and can be used to define a general rule for the validity of the ray-series:

$$\frac{\Lambda}{L} \ll 1 \qquad \Rightarrow \qquad L \sim \frac{1}{\omega} . \tag{A.9}$$

Equation A.9 requires the wavelength to be small compared to the *characteristic length*. For example, no sharp velocity contrasts are allowed and perturbations must be large compared to the wavelength (unless special boundary conditions are used, i.e., Snell's law at interfaces). In other words, the medium needs to be smooth.

Substituting the *leading term* (Eq. A.8) into the elastodynamic Eq. A.6 and rearranging the terms according to the derivations yields

$$(\imath\omega)^2 N_i(\mathbf{U}) + (\imath\omega)^1 M_i(\mathbf{U}) + L_i(\mathbf{U}) = 0 , \tag{A.10}$$

where

$$\begin{aligned} N_i(\mathbf{U}) &= (\lambda+\mu)\, U_j \tau_{,i} \tau_{,j} + \mu U_i \tau_{,j} \tau_{,j} - \rho U_i \,, \\ M_i(\mathbf{U}) &= (\lambda+\mu)\left[U_{j,i}\tau_{,j} + U_{j,j}\tau_{,i} + U_j \tau_{,ij}\right] + \mu\left[2U_{i,j}\tau_{,j} + U_i \tau_{,jj}\right] \\ &\quad + \lambda_{,i} U_j \tau_{,j} + \mu_{,j}\left[U_i \tau_{,j} + U_j \tau_{,i}\right] \quad \text{und} \\ L_i(\mathbf{U}) &= (\lambda+\mu)\, U_{j,ij} + \mu U_{i,jj} + \lambda_{,i} U_{j,j} + \mu_{,j}\left(U_{i,j} + U_{j,i}\right) . \end{aligned} \tag{A.11}$$

The three terms in Eq. A.10 are of different orders of magnitude at high frequencies. The third term with the smallest power in ω (zeroth power) will be neglected from here on.

The first two terms have to be separately equal to zero to satisfy Eq. A.10, which yields the equations to calculate $\tau(x_m)$ and $\mathbf{U}(x_m)$

$$N_i(\mathbf{U}) = 0 \,, \tag{A.12}$$

$$M_i(\mathbf{U}) = 0 \,. \tag{A.13}$$

Equation A.12 yields the *eikonal equations* for τ and separates the wave field at high frequencies. Therefore, the high frequency approximation allows for separating P and S waves. P (primary) and S (secondary) waves are the most important and common waves in seismology.

A.2. The Eikonal Equations

Equation A.12 can be rearranged to give

$$\left(\frac{\lambda+\mu}{\rho}\tau_{,i}\tau_{,j} + \frac{\mu}{\rho}\tau_{,k}\tau_{,k}\delta_{ij} - \delta_{ij}\right) U_j = 0 \,,$$

which constitutes an eigenvalue problem of the form

$$\left(\Gamma_{ij} - \delta_{ij}\right) U_j = 0 \,, \tag{A.14}$$

where

$$\Gamma_{ij} = \frac{\lambda+\mu}{\rho}\tau_{,i}\tau_{,j} + \frac{\mu}{\rho}\tau_{,k}\tau_{,k}\delta_{ij} \tag{A.15}$$

is the *Christoffel Matrix*. The eigenvalue problem has a non trivial solution iff the determinant of the matrix is zero: $det(\Gamma_{ij} - \delta_{ij}) = 0$. The eigenvalues then follow as the three nulls of the (cubic) characteristic polynomial (Bronstein et al., 1997)

$$\chi = det\left(\Gamma_{ij} - G^{(m)}\delta_{ij}\right) = 0 \qquad m = 1, 2, 3\,.$$

In isotropic media, two of the eigenvalues coincide (degeneration). This degeneration characterises the S-wave. The remaining eigenvalue whose eigenvector is perpendicular to the other one characterises the P-wave. The $G^{(m)}$ are equal to 1. The polarisation of the wave is described by the eigenvectors $g_j^{(m)}$ with

$$\left(\Gamma_{ij} - G^{(m)}\delta_{ij}\right) g_j^{(m)} = 0\,.$$

The eigenvectors are pairwise perpendicular to each other (see Fig. A.2) and satisfy $g_j^{(m)} g_j^{(n)} = \delta_{mn}$. It follows that

$$G^{(m)} = 1 = \Gamma_{ij} g_i^{(m)} g_j^{(m)}\,. \tag{A.16}$$

It follows for the degenerated case (isotropic media) by applying A.15 that

$$G^{(1)} = G^{(2)} = 1 = \frac{\lambda(x_i)+2\mu(x_i)}{\rho(x_i)}\tau_{,i}\tau_{,i} = \beta^2(x_i)p_i p_i \quad \text{und} \tag{A.17}$$

$$G^{(3)} = 1 = \frac{\mu(x_i)}{\rho(x_i)}\tau_{,i}\tau_{,i} = \alpha^2(x_i)p_i p_i. \tag{A.18}$$

This introduces the slowness vector $p_i = \tau_{,i}$ (see Fig. A.2) that is perpendicular to the wavefronts. $\alpha(x_i)$ and $\beta(x_i)$ are the velocities of P and S waves as a function of space. The eigenvector $\mathbf{g}^{(3)}$ points in the direction of the slowness which means that the P wave is linearly polarised in the direction of the wave propagation. The two degenerated eigenvectors fall into a plane perpendicular to the slowness and their direction within the plane cannot be resolved. This means that the S wave is polarised transversal.

The ***eikonal equations*** can now be defined as nonlinear partial differential equations of zeroth order

$$p_i p_i = \tau_{,i}\tau_{,i} = \frac{1}{v^2(x_i)}\,, \tag{A.19}$$

where $v(x_i)$ is the wave velocity of P or S waves. The travel time τ is also called the *eikonal*. Hence, each wave has an eikonal equation. The wave propagation of both

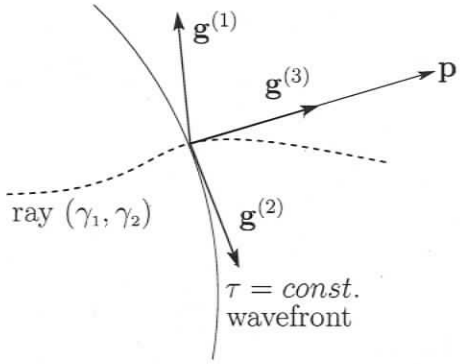

Fig. A.2: A wavefront and the corresponding eigenvectors. The orientation of $\mathbf{g}^{(1)}$ and $\mathbf{g}^{(2)}$ is only constrained to the plane tangential to the wavefront. $\mathbf{g}^{(3)}$ is perpendicular to the wavefront and points in the direction of the slowness $\mathbf{p}$.

waves is decoupled for high frequencies which means that energy transfer between the two types of waves is neglected. The travel time or eikonal τ can be computed through the kinematic ray tracing (KRT) system (see Sec. A.3), which can be derived from Eq. A.19. It is also possible to solve the eikonal equations by finite differences using the *Vidale-scheme* (Vidale, 1988, 1990). However, in that case the computation is restricted to first arrivals. Finally, the computation of amplitudes has not been discussed as it is not important for the basic ray tracing used in this thesis.

Notes on the Leading Term

To derive the elastodynamic equation (Eq. A.10), the term $L_i(\mathbf{U})$ was neglected which results in errors. These errors can be reduces by taking more terms than just the leading term into account. However, since the ray series only converges for $\omega \to \infty$ but diverges otherwise for $N \to \infty$, using many terms in the approximation can deteriorate the approximation.

It is also possible to include the last term in Eq. A.10 with the other terms in the derivation of the eikonal equations. This leads to rays as a function of frequency with frequency dependent eikonal equations and a frequency dependent transport equation.

A.3. The Ray-Tracing System

The Eikonal equation (A.19) is a differential equation of the *Hamilton-Jacobi* type

$$\mathcal{H}(x_i, p_i) = 0.$$

In a six dimensional phase space (the three spatial coordinates and the three slowness components) the Hamiltonian $\mathcal{H}$ constitutes a hyper plane. $\mathcal{H}$ can be written

$$\mathcal{H}(x_i, p_i) = p_i p_i - v^{-2} \quad \text{und}\,. \tag{A.20}$$

The differential equation can be solved with the method of characteristics (e.g., Courant and Hilbert (1966)). Characteristics are trajectories in space that satisfy $\mathcal{H}(x_i, p_i) = 0$. The associated system of differential equations is

$$\frac{dx_i}{du} = \frac{\partial \mathcal{H}}{\partial p_i}\,, \qquad \frac{dp_i}{du} = -\frac{\partial \mathcal{H}}{\partial x_i} \quad \text{und} \quad \frac{d\tau}{du} = p_i \frac{\partial \mathcal{H}}{\partial p_i}\,, \tag{A.21}$$

where u is a monotonous parameter along the trajectory $x_i = x_i(u)$. The equations are coupled for $x_i(u)$ and $p_i(u)$. The solution for $x_i(u)$ constitutes a ray that propagates in the hyper plane $\mathcal{H}(x_i, p_i) = 0$. Hence, the eikonal equation is satisfied along the whole ray. Therefore, Eq. A.21 allows for tracing rays and calculating travel times τ along these rays and is referred to as the **kinematic ray tracing** (KRT) system. $\mathcal{H}(x_i, p_i)$ can be chosen so that $d\tau/du = 1$, hence τ is a monotonous parameter along the ray. The KRT then becomes

$$\frac{dx_i}{d\tau} = v^2 p_i \quad \text{und} \quad \frac{dp_i}{d\tau} = -\frac{1}{v}\frac{\partial v}{\partial x_i}\,. \tag{A.22}$$

Other KRT systems can be derived in the same way. For example, the arc length s (path length) along the ray is often used. The KRT can be solved numerically by Runge-Kutta or similar integration schemes. In isotropic media, the rays are perpendicular to the wavefronts and are independent of the coordinate system (Červený, 2001). Further, the time averaged energy propagates along the ray.

Ray tracing simplifies in a horizontally stratified (range independent) medium. For homogeneous layers, the ray-path in the layer is a straight line and Snell's law is used to calculate reflection and refraction at interfaces. For layers with linear sound

velocity gradients, the ray-path is described by a sector of a circle. These attributes greatly simplify ray tracing in range independent models. A common problem is two point ray tracing where the connecting ray (the eigenray) between two points needs to be found. An efficient way to find eigenrays is to use the method of images (Jensen et al., 1993) and start a Newton's method search with a straight line until the search converges to within an arbitrary, user chosen, bound.

B. Integral Transform Technique and Wavenumber Integration

Wavenumber integration is a popular method to calculate the complex acoustic field in a horizontally stratified ocean. It is analogous to the reflectivity method (Fuchs and Müller, 1971) in the seismological community. In this thesis, the wavenumber integration code OASES was used to compute complex acoustic fields to model reflectivity including spherical wave effects. This section will derive the most important framework for the wavenumber integration method and follows the derivation in Jensen et al. (1993).

The most important (and also most limiting) assumption of wavenumber integration is that it is limited to horizontally stratified environments. This assumption is reasonable in many applications in acoustical oceanography, including all environmental models used in this thesis.

In a range independent, horizontally stratified environment with a omni-directional point source at range $r = 0$ in layer m, the then azimuth independent Helmholtz equation in cylindrical coordinates for this layer is given by

$$\left(\nabla^2 + k_m^2(z)\right) \Psi_m(r,z) = f_s(r,\omega)\frac{\delta(r)}{2\pi r}\,, \tag{B.1}$$

with the Laplace operator in cylindrical coordinates

$$\nabla^2 = \frac{1}{r}\frac{\partial}{\partial r} r \frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \phi^2} + \frac{\partial^2}{\partial z^2} \tag{B.2}$$

and the wavenumber in layer m given by

$$k_m(z) = \frac{\omega}{c(z)}\,. \tag{B.3}$$

For the layers without sources $f_s(z,\omega)$ is set to zero and the field is then described by the homogeneous Helmholtz equation. The Hankel transform can be applied to the Helmholtz equation to obtain a depth separated wave equation for a omni-directional point source in a range independent environment. The transform pair is given by

$$f(r,z) = \int_0^\infty f(k_r,z)J_0(k_r r)k_r dk_r \tag{B.4}$$

$$f(k_r,z) = \int_0^\infty f(r,z)J_0(k_r r)r dr \tag{B.5}$$

Substituting the forward transform (Eq. B.4) into Eq. B.1 yields the depth-separated wave equation

$$\left(\frac{d^2}{dz^2} - \left(k_r^2 - k_m^2(z)\right)\right)\Psi_m(k_r,z) = \frac{f_s(z)}{2\pi}, \tag{B.6}$$

which constitutes an ordinary differential equation in depth. The field $\Psi_m(k_r,z)$ (the solution of Eq. B.6) is a sum of a particular solution $\hat{\Psi}_m(k_r,z)$ and any linear combination of two independent solutions $\Psi_m^+(k_r,z)$ and $\Psi_m^-(k_r,z)$ (the up-going and down-going fields). The complete solution or kernel for the depth dependence is hence given by

$$\Psi_m(k_r,z) = \hat{\Psi}_m(k_r,z) + A_m^+(k_r)\Psi_m^+(k_r,z) + A_m^-(k_r)\Psi_m^-(k_r,z) \tag{B.7}$$

and also referred to as the depth dependent Green's function.

A_m^+ and A_m^- are coefficients that can be found by solving the boundary conditions for the adjacent interfaces of layer m. The particular solution can be chosen to be the field of the source in an unbound medium. The field at nay range can finally be found through the inverse Hankel transform. The depth dependent kernel for azimuth independent cylindrical coordinates is the same as for two dimensional Cartesian coordinates and hence holds for point or line sources.

The computational challenge is to solve the infinite integral for the inverse transform. Since the kernels have asymptotic character, the integration interval can be truncated. Applying the integral transform (Eq. B.4) represents a decomposition of the field into conical waves since

$$J_0(k_r,r) = \frac{1}{2}\left(H_0^{(1)}(k_r r) + H_0^{(2)}(k_r r)\right) \tag{B.8}$$

with (for the asymptotic case)

$$H_0^{(1)}(kr) \approx \sqrt{\tfrac{1}{\pi kr}} e^{i(kr-\pi/4)} \tag{B.9}$$

$$H_0^{(2)}(kr) \approx \sqrt{\tfrac{1}{\pi kr}} e^{-i(kr-\pi/4)} \tag{B.10}$$

for $kr \to \infty$.

For the case of homogeneous fluid layers (used in this thesis), an analytical form can be found for the solution of the depth-separated wave equation. The analytical form of the solution then limits the numerical effort to finding the coefficients A_m^+ and A_m^- from the boundary conditions. A homogeneous layer of an ideal fluid has constant sound velocity $c = \sqrt{k/\rho}$ since the bulk modulus K and the density ρ are both constant, which also results in a constant wavenumber $k_m(z)$. The homogeneous solution is then given by the down-going and up-going conical waves.

$$\phi^+(k,z) = e^{ik_z z} \tag{B.11}$$

$$\phi^-(k,z) = e^{-ik_z z}\,, \tag{B.12}$$

where $k_z = \sqrt{k_m^2 - k_r^2}$ is the vertical wavenumber and ϕ the P wave potential. The complete solution for a layer without sources is then given by solving the wavenumber integral which yields the *transfer function*

$$\phi(r,z) = \int_0^\infty \left(A^- e^{-ik_z z} + A^+ e^{ik_z z}\right) J_0(k_r r) k_r dk_r\,. \tag{B.13}$$

For an omni-directional point source the term of the particular solution has to be added:

$$\hat{\phi}(k_r,z) = \frac{S_\omega}{4\pi} \frac{e^{ik_z|z-z_s|}}{ik_z}\,. \tag{B.14}$$

The Hankel transform is again used to find $\hat{\phi}(r,z)$.

To determine the coefficients for a fluid-fluid interface, two continuous boundary conditions have to be evaluated. Both the vertical displacement w and the normal stress or negative pressure σ_{zz} have to be continuous across the boundary. The vertical

displacement for a layer without source is given by

$$\begin{aligned} w(r,z) =& \frac{\partial \phi}{\partial z} \\ =& \int_0^\infty \left(-ik_z A^- e^{-ik_z z} + ik_z A^+ e^{ik_z z} \right) J_0(k_r r) k_r dk_r \end{aligned} \tag{B.15}$$

and the normal stress σ_{zz} is defined through Hooke's law

$$\begin{aligned} \sigma_{zz}(r,z) =& K \nabla^2 \phi(r,z) \\ =& -\rho\omega^2 \phi(r,z) \\ =& -\rho\omega^2 \int_0^\infty \left(A^- e^{-ik_z z} + A^+ e^{ik_z z} \right) J_0(k_r r) k_r dk_r \,. \end{aligned} \tag{B.16}$$

Similar expressions can be formulated for a layer with a source present.

Since the boundary conditions have to be satisfied at all ranges, the kernels have to satisfy the conditions. Different implementations of wavenumber integration use different approaches to solve the linear system of equations resulting from the boundary conditions and for the numerical evaluation of the Hankel transform, where the latter poses the main problem for this work. For the far field, the inverse Hankel transform can be approximated very efficiently. However, for the very short ranges (as low as 10 m from the source) used in single bounce methods, a full Hankel transformation scheme has to be used for small values of kr. For larger values of kr, the full transform can be tapered into the usual FFP (Jensen et al., 1993) integration. The computation of the full Hankel transform is considerably more expensive. Care has to be taken because of the infinite upper integration limit and it has to be ensured that the rapidly varying Bessel function in the integrand of the wavenumber integral is sufficiently sampled. To constrain the upper limit of the integral, a wavenumber $k_{\max}$ is chosen after which the integrand does not contribute significantly to the integral.

The wavenumber sampling is determined from the wavenumber interval $k_{\max} - k_{\min}$ and the number of samples. The number of samples has to be appropriate to avoid aliasing. The number of wavenumber samples also specifies the number of range steps that the inverse Hankel transform yields.

REFERENCES

T. Akal and J. M. Berkson, editors. *Ocean SeismoAcoustics*, chapter Evaluation of Experimental Techniques for Determining the Plane Wave Reflection Coefficient at the Sea Floor, pages 721–730. Plenum, New York, 1987.

K. Aki and P. G. Richards. *Quantitative Seismology, Theory and Methods.* Freeman, New York, 1980.

R. C. Aster, B. Borchers, and C. H. Thurber. *Parameter Estimation and Inverse Problems.* International Geophysics. Elsevier, Amsterdam, 2005.

L. M. Brekhovskikh. *Waves in Layered Media.* Applied Mathematics and Mechanics. Academic Press, 1st edition, 1980.

L. M. Brekhovskikh and O. A. Godin. *Acoustics of Layered Media I: Plane and Quasi-Plane Waves.* Wave Phenomena. Springer, 1st edition, 1990.

L. M. Brekhovskikh and O. A. Godin. *Acoustics of Layered Media II: Point source and bounded beams.* Wave Phenomena. Springer, 1st edition, 1992.

L. M. Brekhovskikh and Yu. P. Lysanov. *Fundamentals of Ocean Acoustics.* Springer Series on Wave Phenomena. Springer Verlag, 2nd edition, 1991.

I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mhlig. *Taschenbuch der Mathematik.* Verlag Harri Deutsch, 3rd edition, 1997.

B. A. Brooks and L. N. Frazer. Importance reweighting reduces dependence on temperature in Gibbs samplers: an application to the coseismic geodetic inverse problem. *Geophys. J. Int.*, 161:12–20, 2005.

G. M. Bryan. *Physics of Sound in Marine Sediments*, chapter Sonobuoy Measurements in Thin Layers, pages 119–130. Plenum, New York, 1974.

G. M. Bryan. The hydrophone-pinger experiment. *J. Acoust. Soc. Am.*, 68:1403–1408, 1980.

V. Červený. *Seismic Ray Theory.* Cambridge University Press, 2001.

N. R. Chapman. Modelling ocean-bottom reflection loss measurements with the plane-wave reflection coefficient. *J. Acoust. Soc. Am.*, 73:1601–1607, 1983.

N. R. Chapman and Y. Jiang. Geoacoustic inversion in a 3-dimensional shallow water environment. In *Proceedings of the Eighth European Conference on Underwater Acoustics*, pages 521–526, 2006.

R. E. Christensen, J. A. Frank, and W. H. Geddes. Low-frequency propagation via shallow refracted paths through deep ocean unconsolidated sediments. *J. Acoust. Soc. Am.*, 57:1421–1426, 1975.

M. D. Collins, W. A. Kuperman, and H. Schmidt. Nonlinear inversion for ocean-bottom properties. *J. Acoust. Soc. Am.*, 93:2770–2783, 1992.

R. Courant and D. Hilbert. *Methods of mathematical physics.* Wiley, New York, 1966.

J. Dettmer, S. E. Dosso, and C. W. Holland. Geoacoustic inversion with strongly correlated errors. *Can. Acoust.*, 32:194–195, 2004a.

J. Dettmer, S. E. Dosso, M. K. Prior, C. H. Harrison, and N. R. Chapman. Geoacoustic inversion of ambient noise reflection-loss data. In *Proceedings of the Seventh European Conference on Underwater Acoustics*, pages 511–516, 2004b.

J. Dettmer, S. E. Dosso, and C. W. Holland. Bayesian inversion of reflection data for seabed properties of multi-layered systems. In *Proceedings of the Eighth European Conference on Underwater Acoustics*, pages 473–478, 2006.

S. E. Dosso. Quantifying uncertainty in Geoacoustic Inversion. I. A fast Gibbs sampler approach. *J. Acoust. Soc. Am.*, 111:129–142, 2002a.

S. E. Dosso. Benchmarking range-dependent propagation modeling in matched-field inversion. *J. Computat. Acoust.*, 10:231–242, 2002b.

S. E. Dosso. Environmental uncertainty in ocean acoustic source localization. *Inverse Problems*, 19:419–431, 2003.

S. E. Dosso and N. E. Collison. Acoustic tracking of a freely-drifting field of sonobuoys. *J. Acoust. Soc. Am.*, 111:2166–2177, 2002.

S. E. Dosso and G. R. Ebbeson. Array element localisation and survey design. *Can. Acoust.*, 33:16–26, 2005.

S. E. Dosso and C. W. Holland. Geoacoustic uncertainties from visco-elastic inversion of seabed reflection data. *IEEE J. Ocean. Eng.*, 2006. In press.

S. E. Dosso and P.L. Nielsen. Quantifying uncertainty in Geoacoustic Inversion. II. Application to broadband, shallow-water data. *J. Acoust. Soc. Am.*, 111:143–159, 2002.

S. E. Dosso and M. J. Wilmut. Data uncertainty estimation in matched-field geoacoustic inversion. *IEEE J. Ocean. Eng.*, XX:XX, 2005. in press.

S. E. Dosso, M. L. Yeremy, J. M. Ozard, and N. R. Chapman. Estimation of ocean-bottom properties by matched-field inversion of acoustic field data. *IEEE J. Ocean. Eng.*, 18:232–239, 1993.

S. E. Dosso, M. R. Fallat, B. J. Sotirin, and J. L. Newton. Array element localization for horizontal arrays via Occams's inversion. *J. Acoust. Soc. Am.*, 104:846–859, 1998.

S. E. Dosso, M. J. Wilmut, and A.-L. S. Lapinski. An Adaptive-Hybrid Algorithm for geoacoustic inversion. *IEEE J. Ocean. Eng.*, 26:324–336, 2001.

S. E. Dosso, P. L. Nielsen, and M. J. Wilmut. Data error covariance in matched-field geoacoustic inversion. *J. Acoust. Soc. Am.*, 119:208–219, 2006.

M. R. Fallat and S. E. Dosso. Geoacoustic inversion for the workshop 97 benchmark test cases using simulated annealing. *J. Computat. Acoust.*, 6:29–43, 1998.

M. R. Fallat and S. E. Dosso. Geoacoustic inversion via local, global, and hybrid algorithms. *J. Acoust. Soc. Am.*, 105:3219–3230, 1999.

C. M. Ferla and F. B. Jensen. *Are current environmental databases adequate for sonar predictions in shallow water? Impact of littoral environmental variability on acoustic predictions and sonar performance.* Kluwer Academic, Dordrecht, 1987.

J. E. Freund. *Modern Elementary Statistics.* Prentice-Hall, 1967.

K. Fuchs and G. Müller. Computation of synthetic seismograms with the relectivity method and comparision with observations. *GJRAS*, 23:417–433, 1971.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattn. Anal. Mach. Intel.*, 6:721–741, 1984.

P. Gerstoft. Inversion of seismoacoustic data using genetic algorithms and a posteriori probability distributions. *J. Acoust. Soc. Am.*, 95:770–782, 1994.

P. Gerstoft and C. F. Mecklenbräuker. Ocean acoustic inversion with estimation of a posteriori probability distribution. *J. Acoust. Soc. Am.*, 104:808–819, 1998.

W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov Chain Monte Carlo In Practice.* Interdisciplinary Statistics. Chapman & Hall/CRC, Florida, 1996.

D. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading, 1989.

W. Gropp, E. Lusk, and A. Skjellum. *Using MPI, Portable Parallel Programming with the Message-Passing Interface.* MIT Press, Cambridge, 1999.

E. L. Hamilton. Geoacoustic modeling of the sea floor. *J. Acoust. Soc. Am.*, 68: 1313–1340, 1980.

J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods.* Methuen, London, 1964.

C. H. Harrison. Sub-bottom profiling using ocean ambient noise. *J. Acoust. Soc. Am.*, 115:1505–1515, 2004.

C. H. Harrison and A. Baldacci. Bottom reflection properties deduced from ambient noise: Simulation of a processing technique. Technical Report SM-392, SACLANT ASW Research Centre, La Spezia, Italy, November 2002.

C. H. Harrison and A. Baldacci. Bottom reflection properties deduced from ambient noise: Simulation and experiment. *IEEE J. Ocean. Eng.*, 2004. (submitted).

C. H. Harrison and J. A. Harrison. A simple relationship between frequency and range averages for broadband sonar. *J. Acoust. Soc. Am.*, 97:1314–1317, 1995.

C. H. Harrison and P. Nielsen. Plane wave reflection coefficient from near field measurements. *J. Acoust. Soc. Am.*, 116:1355–1361, 2004a.

C. H. Harrison and P. Nielsen. Plane wave reflection coefficient derived from spherical wave measurements. In *Conference Proceedings*, pages 3–8, 2004b.

C. H. Harrison and D. G. Simons. Geoacoustic inversion of ambient noise: A simple method. *J. Acoust. Soc. Am.*, 112:1377–1389, 2002.

O. F. Hastrup. Digital analysis of acoustic reflectivity in the Tyrrhenian Abyssal Plain. *J. Acoust. Soc. Am.*, 42:181–190, 1970.

C. W. Holland. Geoacoustic inversion for fine-grained sediments. *J. Acoust. Soc. Am.*, 111:1560–1564, 2002.

C. W. Holland. Seabed reflection measurement uncertainty. *J. Acoust. Soc. Am.*, 114: 1861–1873, 2003.

C. W. Holland and J. Osler. High-resolution geoacoustic inversion in shallow water: A joint time- and frequency-domain technique. *J. Acoust. Soc. Am.*, 107:1263–1279, 2000.

C. W. Holland, J. Dettmer, and S. E. Dosso. Remote sensing of sediment density and velocity gradients in the transition layer. *J. Acoust. Soc. Am.*, 118:163–177, 2005.

C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss. Uncertainty analysis in matched-field geoacoustic inversions. *J. Acoust. Soc. Am.*, 119:197–207, 2006.

L. Ingber. Very fast simulated reannealing. *Math. Comput. Model.*, 12:967–973, 1989.

L. Jaschke and N. R. Chapman. Matched field inversion of broadband data using the freeze bath method. *J. Acoust. Soc. Am.*, 106:1838–1851, 1999.

F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt. *Computational Ocean Acoustics.* Series in Modern Acoustic and Signal Processing. AIP Press, New York, 1993.

F. B. Jensen, P. L. Nielsen, M. Zampolli, M. D. Collins, and W. L. Siegmann. Benchmarking range-dependent seismo-acoustic propagation problems. In *Proceedings of the Eighth European Conference on Underwater Acoustics*, pages 45–50, 2006.

S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–679, 1983.

H. P. Kuenzi, H. G. Tzschach, and C. A. Zehnder. *Numerical Methods of Mathematical Optimization.* Academic Press, New York, 1971.

W. A. Kuperman and H. Schmidt. Self-sonsistent perturbation approach to rough surface scattering in stratified elastic media. *J. Acoust. Soc. Am.*, 86:1511–1522, 1989.

T. Lay and T.C. Wallace. *Modern Global Seismology.* Academic Press, 1995.

H. W. Lilliefors. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *J. Am. Stat. Assoc.*, pages 399–402, 1967.

C. E. Lindsay and N. R. Chapman. Matched field inversion for geoacoustic model parameters using adaptive simulated annealing. *IEEE J. Ocean. Eng.*, 18:224–231, 1993.

F. J. Massey. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.*, 46:68–78, 1951.

C. F. Mecklenbräuker and P. Gerstoft. Objective functions for ocean acoustic inversion derived by likelihood methods. *J. Computat. Acoust.*, 8:259–270, 2000.

W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory.* Academia Press, New York, 1984.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, and A. Teller A. E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21: 1087–1092, 1953.

K. Mosegaard and M. Sambridge. Monte Carlo analysis of inverse problems. *Inverse Problems*, 18:R29–R54, 2002.

J. A. Nelder and R. Mead. A simplex method for function minimization. *Comp. J.*, 7:308–313, 1965.

J. Nocedal and S. J. Wright. *Numerical Optimisation.* Operations Research. Springer, New York, 1999.

D. B. Owen. *Handbook of Statistical Tables.* Addison-Wesley, Reading, 1962.

R. L. Parker. *Geophysical Inverse Theory.* Princeton University Press, Princeton, 1994.

W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in Fortran 77.* Cambridge University Press, 2nd edition, 1997.

A. C. Rencher. *Multivariate Statistical Inference and Applications.* John Wiley & Sons, Inc., New York, 1998.

M. D. Richardson. *Shallow Water Acoustics*, chapter In-situ shallow water sediment geoacoustic properties. China Ocean Press, 1997.

M. Riedel, S. E. Dosso, and L. Beran. Uncertainty estimation for amplitude variation with offset (AVO) inversion. *Geophysics*, 68:1485–1496, 2003.

D. H. Rothman. Automatic estimation of large residual statistics corrections. *Geophysics*, 51:337–346, 1986.

M. Sambridge and K. Mosegaard. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40:3–1–3–29, 2002.

H. Schmidt. SAFARI: Seismo-acoustic fast field algorithm for range-independent environments. Technical Report SR-113, SACLANT ASW Research Centre, La Spezia, Italy, 1988.

M. K. Sen and P. L. Stoffa. *Global Optimization Methods in Geophysical Inversion.* Elsevier, Amsterdam, 1995.

M. K. Sen and P. L. Stoffa. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. *Geophys. Prosp.*, 44:313, 1996.

P. M. Shearer. *Introduction To Seismology.* Cambridge University Press, 1st edition, 1999.

M. Siderius and C. H. Harrison. High-frequency geoacoustic inversion of ambient noise data using short arrays. *IEEE J. Ocean. Eng.*, 2004. (submitted).

A. Sommerfeld. *Partial Differential Equations in Physics.* Academic, New York, 1949.

D. C. Stickler. Negative bottom loss, critical-angle shift, and the interpretation of the bottom reflection coefficient. *J. Acoust. Soc. Am.*, 61:707–710, 1977.

H. Szu and R. Hartley. Fast simulated annealing. *Phys. Lett.*, A122:157–162, 1987.

A. Tarantola. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation.* Elsevier, Amsterdam, 1987.

W. M. Telford, L. P. Geldart, and R. E. Sheriff. *Applied Geophysics.* Cambridge University Press, 2nd edition, 1995.

D. Tollefsen, S. E. Dosso, and M. J. Wilmut. Matched-field geoacoustic inversion with a horizontal array and low-level source. *J. Acoust. Soc. Am.*, 120:1–10, 2006.

A. Tolstoy. MFP benchmark inversions via the Rigs method. *J. Computat. Acoust.*, 6:185–203, 1998.

J. E. Vidale. Finite-difference calculation of travel times. *Bull., Seis Soc. Am.*, 78: 2062–2076, 1988.

J. E. Vidale. Finite-difference calculation of traveltimes in three dimensions. *Geophysics*, 55:521–526, 1990.