Acoustic Full Waveform Inversion of Marine Reflection Seismic Data

Zur Erlangung des akademischen Grades eines
DOKTORS DER NATURWISSENSCHAFTEN
von der Fakultät Physik des
Karlsruher Instituts für Technologie

genehmigte

DISSERTATION

von

Dipl.-Ing. Anna Przebindowska
aus Krakau

Tag der mündlichen Prüfung	21.06.2013
Referent:	Prof. Dr. Thomas Bohlen
Korreferent:	Prof. Dr. Florian Bleibinhaus

Abstract

The objective of full waveform inversion (FWI) is to reconstruct a model of the subsurface by minimizing the residuals between observed and modelled seismic data. Because of the high-nonlinearity of the inverse problem, the success of the waveform inversion depends mainly on the accuracy of the starting model and on the presence of low frequencies and sufficient offsets in the recorded data. In recent years, many synthetic studies have shown that FWI has a great potential to become an important method for determining high-resolution, multi-parameter models of complex subsurface structures, however, application to field data is not yet a common practice. When inverting marine field seismic data, we have to deal with some additional challenges including seismic noise, limited offsets, lack of low frequencies, or the unknown source signal. These factors lead to poor convergence of the inversion algorithm and deteriorate the recovery of subsurface models. Therefore, some extra processing steps and a careful choice of inversion strategies are required for the successful inversion of field data.

The main objective of the thesis is the application of the 2D acoustic time-domain full waveform inversion to conventional streamer data recorded in the North Sea. In addition to the field data example, the synthetic studies and experiments presented in this work aim to investigate the possibilities and limitations of the method and to provide the guidelines on how to design an efficient inversion scheme. Furthermore, this work discusses several practical aspects related to the application of 2D acoustic FWI to field data, such as the 3D to 2D transformation, source wavelet estimation, and acquisition effects.

The numerical tests are based on a modified acoustic Marmousi2 model and they aim to reflect the real measurement conditions of marine reflection seismics. The first series of inversion experiments investigates the performance of different gradient preconditioning operators, the implementation of the multi-scale inversion approach, and the choice of the objective function. Another aspects investigated in this work is the influence of different model parameterizations on the multi-parameter acoustic waveform inversion. A set of inversion tests is performed to assess the different parameterization sets in terms of the quality of the reconstructed models and of the convergence rate of the inversion. Whereas the resolution of velocity and impedance models is comparable, the reconstruction of density structures strongly depends on the model parameterization.

To investigate the role of density on the recovery of velocity models, I perform a series of numerical experiments testing various initial density models and different strategies for the density update. The inversion results show that the realistic density information should be directly included in the inversion process to improve the accuracy of the velocity reconstruction.

Finally, I present the successful application of 2D acoustic FWI to marine field data. I introduce a workflow for the field data inversion and discuss the key aspects related to the field data processing, starting model generation, initial modelling, resolution study, and choice of the inversion parameters. To reduce the high complexity of the inverse problem, the combination of the multi-scale inversion with the time and offset windowing of the data is applied. The validation of inversion results is achieved by both a comparison of synthetic and observed waveforms and by a comparison of the inverted model with the migrated seismic section.

Contents

List of Figures

Chapter 1

Introduction

The overall goal of marine exploration geophysics is to image geological structures. Among other methods, the full waveform inversion (FWI) proved to be an efficient tool for determining high-resolution multi-parameter models (P-wave velocity, S-wave velocity, density) of complex subsurface structures. In contrast to the traveltime tomography, waveform inversion utilizes the full information content of the observed seismograms, i.e. amplitudes and phases, thus it has a potential to image structures that are smaller than the seismic wavelength. Advances in parallel computing technology, numerical methods, and the improving quality of seismic data make the application of FWI feasible today.

The objective of FWI is to find a model of the subsurface that explains the observed data. For that purpose, the algorithm has to minimize the residuals between modeled and observed data in an iterative process. To solve this problem we need to generate synthetic data from a starting model (the forward problem) and apply an efficient method for minimizing the data misfit function. The inversion strategy is based on the adjoint method. The forward problem and backpropagation of the residual wavefield are solved using a parallel time domain finite difference code, [Bohlen, 2002]. The forward modeling code applies perfectly matched layers to suppress artificial reflections from the model boundaries. The waveform inversion scheme is based on the general approach of Tarantola [1984] and Mora [1987] formulated in the space-time domain. The inversion problem can be addressed in an elastic or acoustic manner by utilizing the elastic or acoustic wave equation, respectively. The elastic FWI is more complicated than the acoustic approach, because it has to simultaneously optimize three coupled elastic parameters and, as a result, it requires more computational power.

It is a common practice to apply the acoustic approach for seismic data from marine exploration, i. e. to streamer or to OBC data, as it leads to a significant reduction in computational cost. However, the elastic simulation would provide a better match to the acquired marine data, since they contain many elastic signals such as surface waves and mode converted waves. Nevertheless, it has been shown that the result of the acoustic inversion is reasonable when using marine seismic data.

Synthetic studies show the great resolution potential of the full waveform inversion, however, application to field data is not a common standard yet. This thesis discusses some of the problems related with the inversion of marine data in the acoustic approximation. The application of waveform inversion to field data is an extremely nonlinear problem. The lack of low frequency information makes the waveform inversion strongly depending on the initial model. Moreover, elastic effects, attenuation,

noise present in the data, as well as an unknown source signature additionally reduce the performance of the inversion algorithm. To mitigate some of these problems, it is necessary to preprocess the data, estimate an accurate starting model and source wavelet. To enable the convergence and to eliminate some of the artifacts, it is essential to apply various preconditioning methods, and to include any available a priori information on the model parameters, including the density information.

1.1 FWI - state of the art

The acoustic full waveform inversion method was originally developed in the time domain in the early 80's by Tarantola [1984] as a local optimization method that aims to minimize the least-squares misfit between observed and modelled seismograms. The theory was extended to the elastic case by Mora [1987]; Tarantola [1986]. The frequency-domain full waveform inversion was introduced in early 90's [Pratt, 1990; Pratt and Worthington, 1990]

Although FWI is computationally intensive, it is of particular interest, because it enables the high-resolution multi-parameter inversion of subsurface structures. The method can be applied to a wide range of geophysical problems:

- ultrasonic data [Dessa and Pascal, 2003; Pratt, 1999];
- crosshole seismic data [Pratt and Shipp, 1999; Wang and Rao, 2006];
- crosshole radar data [Ernst et al., 2007];
- onshore wide-aperture seismic data [Bleibinhaus et al., 2009; Malinowski and Operto, 2008; Ravaut et al., 2004];
- wide-angle refraction data recorded with ocean-bottom stations or ocean bottom cables [Dessa et al., 2004; Operto et al., 2006; Sirgue et al., 2009];
- marine streamer data [Boonyasiriwat et al., 2010; Delescluse et al., 2011; Hicks and Pratt, 2001; Shin and Min, 2006; Shipp and Singh, 2002; Takam Takougang and Calvert, 2011; Vigh and Starr, 2008; Wang and Rao, 2009];
- seismological imaging [Fichtner et al., 2009].

1.2 Objective of the thesis

The aim of this thesis is the application of the 2D acoustic full waveform inversion to marine streamer data. The main goal is to develop a workflow for the field data inversion that includes field data preprocessing, generation of the starting models, source wavelet estimation, and the choice of appropriate inversion strategies. The possibilities and limitations of the inversion method are investigated in details in order to provide some guidance on the appropriate choice of the data processing routines and on the design of an efficient inversion scheme. Furthermore, the thesis discusses several practical issues related to the field data inversion. The FWI is eventually applied to conventional marine streamer data acquired in the North Sea.

1.3 Overview of the thesis

The thesis is divided into six main chapters:

- **Chapter 2** FWI - theory and implementation

  In the first section, I provide a general introduction into the inversion of the non-linear problems. This is followed by the introduction of the full waveform inversion approach in the time-domain. Finally, I describe the key elements of the 2D acoustic time-domain full waveform inversion code, which is used in this thesis.

- **Chapter 3** Choice of inversion strategies

  In this chapter, I illustrate the application of acoustic full waveform inversion to synthetic streamer data. First, I investigate the performance of different preconditioning operators applied to the gradient of the misfit function. Afterwards, I discuss the implementation of the multi-scale inversion approach with a special focus on the efficient selection of frequency bands. Finally, I assess the performance of FWI for different minimization criteria by inverting synthetic data affected by noise.

- **Chapter 4** Parameterization study

  The aim of this chapter is to find the most suitable parameterization for the multi-parameter time-domain acoustic inversion of marine reflection seismics. I investigate three different combinations of parameters: P-wave velocity and density, acoustic impedance and density, P-wave velocity and acoustic impedance. First, I analyse the coupling between the different parameters, when the simultaneous multi-parameter inversion is performed. Finally, I perform the multi-parameter inversion of synthetic data to assess the different model parameterizations in terms of the quality of the reconstructed models and of the convergence rate of the inversion.

- **Chapter 5** The role of density in acoustic full waveform inversion

  The objective of this chapter is to analyse the role of density on the recovery of P-wave velocity models in the marine environment. To investigate the footprint of density on FWI, I perform a series of numerical experiments, testing various initial density models and different strategies for the density update.

- **Chapter 6** Some practical aspects

  In this chapter, I investigates several aspects related to the application of the 2D FWI to real data, such as the 3D to 2D transformation, source wavelet estimation, and acquisition effects.

- **Chapter 7** Field data example from the North Sea

  In the final chapter, I apply the 2D acoustic FWI to seismic streamer data acquired in the North Sea. I present the workflow for the field data inversion that includes field data preprocessing, starting model generation, initial modelling, choice of the inversion parameters, resolution study and the application of the waveform inversion. I present results obtained with different multi-stage inversion strategies, different objective functions, starting models, and parameterization sets. Finally, I apply a conventional reflection data processing to obtain the time-migrated section and to validate the FWI results.

Chapter 2

FWI - theory and implementation

2.1 Inversion of non-linear problems

The aim of the seismic inverse problem is to estimate the physical parameters of the Earth's subsurface from a seismic recording. The seismic inverse problem is complicated because seismic waves are nonlinearly related to the unknown parameters of the medium.

The solution of an inverse problem generally comprises of three steps:

- **Model parameterization**. Choice of an optimal set of model parameters that characterize the model and that can be reconstructed from available data. The selection of model parameters, i.e. the unknowns that we try to estimate, is usually problem dependent. A particular set of model parameters is called a parameterization of the system.

- **Forward modelling**. It is the process of predicting the data based on some physical or mathematical model with a given set of model parameters. This step requires the knowledge of the physical laws that allow to predict the data. For example, the acoustic wave equation allows us to predict seismic data from a given model of acoustic parameters consisting of the bulk modulus and density.

  In general, the relationship between seismic data $\mathbf{d}_{\text{mod}}$ recorded at receivers and the Earth represented by model parameters $\mathbf{m}$ is expressed by

  $$\mathbf{d}_{\text{mod}} = f(\mathbf{m}) \,, \tag{2.1}$$

  where f is the nonlinear forward operator.

- **Inverse problem**. It is the process of estimating the model parameters from the observed data $\mathbf{d}_{\text{obs}}$. Most criteria for selecting the best model of the subsurface is based on fitting the observed data $\mathbf{d}_{\text{obs}}$ and the data predicted by the forward modelling $\mathbf{d}_{\text{mod}}$ for a given model of the subsurface $\mathbf{m}$.

  In general, the predicted values cannot be identical to the observed values due to data uncertainties and theoretical uncertainties. The observed data are recorded by instruments and contain noise (data uncertainties). Furthermore, the physical theory, that allow us to predict the data

for a given parameter of the model, may be inaccurate (theoretical uncertainties). The forward problem is an approximation and the model is a simplified representation of the true system.

Inverse problems are usually ill-posed. The concept of an ill-posed problem was introduced to distinguish between the forward or direct problem and the inverse problem. A problem is called well-posed if [Hadamard, 1902]: 1) there exists a solution to the problem (existence), 2) there is at most one solution to the problem (uniqueness), 3) the solution depends continuosly on the data, i.e. a small perturbation of the data does not produce large perturbations of the inverse problem solution (stability). A problem that lacks any of these properties is called ill-posed. While the forward problem has a unique solution, the inverse problem is non-unique, e.g. there is more than one model that can fit the observed data.

The methods used for solving the non-linear inverse problems can be divided in two main categories:

- **Global methods**. The aim of global optimization is to search the model space in order to find the best solution. These methods allow strongly non-linear problems but small number of parameters. (e.g. *grid search*: systematic search, *Monte Carlo method*: random exploration, *Simulated annealing*, *Genetic algorithm*).
- **Local methods**. These methods as based on using a local information about the gradient of the objective function to iteratively improve the initial model. The objective function is defined as a quantitative measure of the agreement between the observed data and the predicted data. Since an extensive exploration of the model space is avoided, the local methods can be applied to large number of parameters. But on the other hand, the local methods are suitable only for weakly non-linear problems, because the local optimization is based on successive linearization of the problem. As a consequence, the solution strongly depends on the choice of the starting model. (e.g. *Full waveform inversion*).

2.1.1 Local optimization methods

The goal of the seismic inversion is to find a set of model parameters $\mathbf{m}$, which describes the Earth's subsurface and provides a good fit between the observed data, $\mathbf{d}_{\text{obs}}$, and the data predicted by the forward problem for a given model of the subsurface, $\mathbf{d}_{\text{mod}}(\mathbf{m})$.

The data misfit is defined as the difference between the modelled data and the observed data

$$\delta\mathbf{d} = \mathbf{d}_{\text{mod}} - \mathbf{d}_{\text{obs}} = f(\mathbf{m}) - \mathbf{d}_{\text{obs}} \,, \tag{2.2}$$

where $\delta\mathbf{d}$ is the data misfit vector of dimension N, and $\mathbf{m}$ is the model vector of dimension M. A very common criterion for fitting observed and predicted data is to minimize the objective function described by the L_2-norm of the error between the predicted and observed data

$$E(\mathbf{m}) = \frac{1}{2}\delta\mathbf{d}^T\delta\mathbf{d} = \frac{1}{2}\sum_{i=1}^{N}\delta d_i^2 \,, \tag{2.3}$$

where $E(\mathbf{m})$ is the objective function (also called misfit function), and the superscript ^{T} denotes the matrix transpose. The summation is performed over the number of source-receiver pairs and the number of time samples.

If the non-linearity between the data and the model is weak, it is possible to linearize the inverse problem. The most popular approach is the Born approximation, e.g. we search for the best fitting model $\mathbf{m}$ in the vicinity of the starting model $\mathbf{m}_0$ such that

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m} \tag{2.4}$$

where $\delta\mathbf{m}$ is a perturbed model. It means that we consider our medium $\mathbf{m}$ as a perturbation of a reference medium $\mathbf{m}_0$.

To linearize a system, the first-order Taylor series expansion of the misfit function around the starting model is used

$$E(\mathbf{m}_0 + \delta\mathbf{m}) = E(\mathbf{m}) = E(\mathbf{m}_0) + \frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} + \mathcal{O}(\mathbf{m}^2) \,, \tag{2.5}$$

the residuals of higher orders, $\mathcal{O}(\mathbf{m}^2)$, are neglected.

To find the minimum of the objective function $E(\mathbf{m})$ the derivative of the misfit function (Eq. 2.5) with respect to the model parameters must vanish, such that

$$\frac{\partial E(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}} + \frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2}\delta\mathbf{m} = 0 \,, \tag{2.6}$$

this leads to the perturbation model

$$\delta\mathbf{m} = -\left[\frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2}\right]^{-1} \frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}} \,, \tag{2.7}$$

The first derivative of the objective function (Eq. 2.3) with respect to the model parameters is given by

$$\frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}} = \left(\frac{\partial \mathbf{d}_{\mathrm{mod}}(\mathbf{m}_0)}{\partial \mathbf{m}}\right)^T (\mathbf{d}_{\mathrm{mod}}(\mathbf{m}_0) - \mathbf{d}_{\mathrm{obs}}) = \mathbf{J}_0^T \delta\mathbf{d} \tag{2.8}$$

where $\mathbf{J}^T$ is the transpose of the Jacobian matrix $\mathbf{J}$ also called the Fréchet derivate matrix. The individual elements of the matrix are

$$J_{ij} = \frac{\partial d_{\mathrm{mod}_i}}{\partial m_j} \qquad (i = 1, 2, \ldots, N),\ (j = 1, 2, \ldots, M) \,. \tag{2.9}$$

The second derivative of the objective function with respect to the model parameters gives

$$\frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2} = \mathbf{H}_0 = \mathbf{J}_0^T \mathbf{J}_0 + \left(\frac{\partial \mathbf{J}_0}{\partial \mathbf{m}}\right)^T \delta\mathbf{d} \,, \tag{2.10}$$

where $\mathbf{H}$ is the Hessian matrix of dimension MxM, which defines the curvature of the misfit function.

Substituting Eq. 2.8 and Eq. 2.10 into Eq. 2.7 finally gives

$$\delta\mathbf{m} = -\mathbf{H}_0^{-1}\mathbf{J}_0^T \delta\mathbf{d} \,. \tag{2.11}$$

If the misfit function is not exactly quadratic, the Newton method will not converge in one iteration. Therefore, an iterative process is necessary and the model update at iteration n using the **Newton**

method is given by

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mathbf{H}_n^{-1} \left(\frac{\partial E}{\partial \mathbf{m}} \right)_n . \tag{2.12}$$

Because the complete, inverse Hessian is very large and difficult to calculate, the Newton methods are rarely used to solve large inverse problems (Pratt 98). The more common method is the **Gauss-Newton method**, which applies the approximate Hessian matrix $\mathbf{H}_a$ for updating the model parameters

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \left(\mathbf{H}_a^{-1} \right)_n \left(\frac{\partial E}{\partial \mathbf{m}} \right)_n , \tag{2.13}$$

Because the second term in Eq. 2.10 is small (Tarantola 1987) it is neglected and the approximate Hessian matrix is defined as

$$\mathbf{H}_a = \mathbf{J}^T \mathbf{J} . \tag{2.14}$$

Another method for minimizing the objective function and updating the model is the **gradient method**. The gradient is given by

$$\frac{\partial E}{\partial \mathbf{m}} = \mathbf{J}^T \delta \mathbf{d} \tag{2.15}$$

The negative gradient of the misfit function represents the steepest descent direction and the model is updated using the following equation

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mu_n \left(\frac{\partial E}{\partial \mathbf{m}} \right)_n , \tag{2.16}$$

where μ is the step length, which replace the inverse of the Hessian in Eq. 2.12. The main disadvantage of the classical steepest descent gradient is the slow convergence rate, which is related to the estimation of a reliable step length [Virieux and Operto, 2009].

2.2 Time domain full waveform inversion

The acoustic full waveform inversion code, which is used in this thesis, is based on the general approach of [Tarantola, 1984] and [Mora, 1987] formulated in the time domain. In order to find a model of the subsurface, the residuals between the predicted and observed data are minimized in an iterative process using the gradient method. The most critical part of the gradient method is the computation of the Fréchet derivative matrix $\mathbf{J}$. The explicit calculation of the partial derivative matrix would require a perturbation of each model parameter separately and the modelling of the wavefield. For a single shot this would require M forward modellings, where M is the number of unknowns in the model.

In the full waveform inversion, the gradient of the misfit function $\nabla E_{\mathbf{m}}$ is calculated using the adjoint approach [Mora, 1987; Plessix, 2006; Tarantola, 1984]. The method does not require the explicit calculation of the Fréchet matrix, i.e. the gradient direction is obtained by the cross-correlation of the forward propagated wavefield and the backpropagated residual wavefield. Such an efficient calculation of the gradient of the misfit function, which requires only $2 \times N_{shots}$ forward simulations, is an essential feature of the FWI method.

2.2.1 The adjoint approach

To solve the inverse problem, the initial non-linear forward problem is linearised in the vicinity of the background model. Because the new model $\mathbf{m}$ is described as a linear combination of the background model $\mathbf{m}_0$ and the perturbed model $\delta\mathbf{m}$ (Eq. 2.5), the relationship between the modelled data and the model (Eq. 2.1) can be written as

$$\mathbf{d}_{\text{mod}} = f(\mathbf{m}) = f(\mathbf{m}_0 + \delta\mathbf{m}) \tag{2.17}$$

The nonlinear forward operator f can approximated using the first-order Taylor series expansion in a similar way to Eq. 2.5. This gives

$$f(\mathbf{m}_0 + \delta\mathbf{m}) = f(\mathbf{m}) = f(\mathbf{m}_0) + \frac{\partial f(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} + \mathcal{O}(\mathbf{m}^2) \, , \tag{2.18}$$

A small perturbation in the data space $\delta\mathbf{d}$ which result from a small perturbation in the model parameters $\delta\mathbf{m}$ can be defined as

$$\delta\mathbf{d} = f(\mathbf{m}_0 + \delta\mathbf{m}) - f(\mathbf{m}_0) \, , \tag{2.19}$$

Substituting Eq. 2.18 into Eq. 2.19 gives

$$\delta\mathbf{d} = f(\mathbf{m}_0) + \frac{\partial f(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} - f(\mathbf{m}_0) = \frac{\partial f(\mathbf{m}_0)}{\partial \mathbf{m}}\delta\mathbf{m} = \mathbf{J}_0\delta\mathbf{m} \, , \tag{2.20}$$

Since we assumed that the best fitting model $\mathbf{m}$ provides a good fit between the observed data $\mathbf{d}_{\text{obs}}$ and the data predicted by the forward problem $\mathbf{d}_{\text{mod}}(\mathbf{m})$, $\delta\mathbf{d}$ represents the data residuals. Eq. 2.20 means that the data residuals $\delta\mathbf{d}$ have a linear relation with the model perturbation $\delta\mathbf{m}$, and the linear operator $\mathbf{J}$ is the Fréchet derivative matrix.

Eq. 2.20 can be written in a continuous form [Mora, 1987]

$$\delta\mathbf{d}(D) = \int_M dM \frac{\partial \mathbf{d}(D)}{\partial \mathbf{m}}\delta\mathbf{m}(M) \, , \tag{2.21}$$

where M and D indicates the model and the data space. If the Fréchet derivative matrix is known, then all small perturbations in the model parameters can be integrated to calculate the resulting perturbations in the data space.

The gradient of the objective function (Eq. 2.15), which is used in the gradient method to update the model parameters, can be written in a continuous form

$$\nabla E_{\mathbf{m}}(M) = \int_D dD \left[\frac{\partial \mathbf{d}(D)}{\partial \mathbf{m}}\right]^* \delta\mathbf{d}(D) \, , \tag{2.22}$$

where, the superscript (*) denotes the adjoint. It is important to notice that $\delta\mathbf{m}$ and $\nabla E_{\mathbf{m}}$ are not the same.

2.2.1.1 The adjoint problem in the acoustic approximation

The 2D inhomogeneous second-order variable density acoustic wave equation is given by

$$\frac{1}{K(x,z)}\frac{\partial^2 p(x,z,t)}{\partial t^2} - \nabla\cdot\left(\frac{1}{\rho(x,z)}\nabla p(x,z,t)\right) = s(x,z,t)\ , \tag{2.23}$$

where $K = \rho V_P^2$ is the bulk modulus, ρ is the density, V_P is the P-wave velocity, p is the pressure field, and s is the source term. The measurements of the pressure field p for a given model $\mathbf{m}$ and a source located at $\mathbf{x} = \mathbf{x}_s$ are performed at a receiver location $\mathbf{x} = \mathbf{x}_r$, such that the ith component of the pressure seismogram at time t is defined as $d_i = f(\mathbf{m}) = p(\mathbf{x}_r, t; \mathbf{x}_s)$. The source term is

$$s(\mathbf{x}, t; \mathbf{x}_s) = \delta(\mathbf{x} - \mathbf{x}_s)S(t)\ , \tag{2.24}$$

where $S(t)$ is the source time function, and $\delta(\mathbf{x})$ is the Dirac delta function.

In the acoustic formulation, Eq. 2.21 takes the form

$$\delta d_i(\mathbf{x}_r, \mathbf{x}_s, t) = \int_V dV(\mathbf{x})\frac{\partial d_i(\mathbf{x}_r, \mathbf{x}_s, t)}{\partial \mathbf{m}(\mathbf{x})}\delta\mathbf{m}(\mathbf{x})\ , \tag{2.25}$$

whereas the adjoint problem (Eq. 2.22) is

$$\nabla E_{\mathbf{m}}(\mathbf{x}) = \sum_s \int dt \sum_r \frac{\partial d_i(\mathbf{x}_r, \mathbf{x}_s, t)}{\partial \mathbf{m}(\mathbf{x})}\delta d_i(\mathbf{x}_r, \mathbf{x}_s, t) \tag{2.26}$$

The gradient of the objective function is the integral over the data space of the Fréchet derivative multiplied with the data residuals. In order to obtain the adjoint formulation in Eq. 2.26, it is enough to obtain the integral expression in Eq. 2.25 for the forward problem, which defines the perturbation in the pressure field δd_i that corresponds to some perturbations in the model parameters $\delta\mathbf{m}$. The desired integral can be supplied by the linearised solution of the acoustic wave equation in terms of Green's functions.

The full derivation of the adjoint state method in the acoustic approximation is given by Tarantola [1984]. The final gradients for bulk modulus K and density ρ are calculated using the following equations

$$\begin{aligned}
\delta K(\mathbf{x}) &= \frac{1}{K^2(\mathbf{x})}\sum_s \int_t dt\ \frac{\partial p(\mathbf{x}, t; \mathbf{x}_s)}{\partial t}\ \frac{\partial p^{'}(\mathbf{x}, t; \mathbf{x}_s)}{\partial t}\ , \\
\delta\rho(\mathbf{x}) &= \frac{1}{\rho^2(\mathbf{x})}\sum_s \int_t dt\ \nabla p(\mathbf{x}, t; \mathbf{x}_s)\cdot\nabla p^{'}(\mathbf{x}, t; \mathbf{x}_s)
\end{aligned} \tag{2.27}$$

where $p(\mathbf{x}, t; \mathbf{x}_s)$ represents the pressure wavefield excited by a source located at $\mathbf{x}_s$, which propagates in the current model, and $p^{'}(\mathbf{x}, t; \mathbf{x}_s)$ is the backpropagated residual wavefield.

The wavefield $p(\mathbf{x}, t; \mathbf{x}_s)$ is defined by

$$p(\mathbf{x}, t; \mathbf{x}_s) = \int_V dV\ G(\mathbf{x}, t; \mathbf{x}_s, 0) * s(\mathbf{x}, t; \mathbf{x}_s) \tag{2.28}$$

where $G(\mathbf{x}, t; \mathbf{x}_s, 0)$ represents the acoustic Green's function, which is associated with Eq. 2.23, and

$(*)$ denotes time convolution.

The backpropagated residual wavefield $p^{'}(\mathbf{x}, t; \mathbf{x}_s)$ is calculated by propagating the data residuals backward in time

$$p^{'}(\mathbf{x}, t; \mathbf{x}_s) = \sum_r G(\mathbf{x}, -t; \mathbf{x}_r, 0) * \delta d_i(\mathbf{x}_r, \mathbf{x}_s, t) \tag{2.29}$$

The residuals are backpropagated from all receiver locations that correspond to the source $\mathbf{x}_s$. The residual wavefield is referred to as the missing diffracted field for the current model. The computation of the residual wavefield requires a single forward modelling, where the source term is replaced by the data residuals. For a single source, the gradient of bulk modulus (Eq. 2.27) is calculated by the zero-lag cross-correlation of the time derivative of the forward modelled wavefield p and the time derivative of the backpropagated residual wavefield p'. The computation of the density gradient is similar, except that the temporal derivatives of the forward and residuals wavefields are replaced with spatial derivatives. The final gradient is obtained by the summation of gradients calculated for individual shots. The derivation of gradients with respect to other material parameters is shown in Chapter 4.

2.3 Full waveform inversion algorithm

In this section, I describe the key elements of the 2D acoustic time-domain full waveform inversion code, which is used in this thesis. The code originates from the PhD thesis of André Kurzmann [Kurzmann, 2012].

2.3.1 Forward modelling

The first-order acoustic wave equation with a variable density is expressed as

$$\begin{aligned} \frac{\partial p}{\partial t} &= K \, \nabla \mathbf{w} \\ \frac{\partial \mathbf{w}}{\partial t} &= \frac{1}{\rho} \nabla p \end{aligned} \tag{2.30}$$

where $\mathbf{w}$ is the particle velocity, p is the pressure field, K is the bulk modulus, and ρ is the density. For the 2D acoustic wave equation, we get

$$\begin{aligned} \frac{\partial p(x,z,t)}{\partial t} &= K(x,z) \left(\frac{\partial w_x(x,z,t)}{\partial x} + \frac{\partial w_z(x,z,t)}{\partial z} \right) \\ \frac{\partial w_x(x,z,t)}{\partial t} &= \frac{1}{\rho(x,z)} \frac{\partial p(x,z,t)}{\partial x} \\ \frac{\partial w_z(x,z,t)}{\partial t} &= \frac{1}{\rho(x,z)} \frac{\partial p(x,z,t)}{\partial z} \, . \end{aligned} \tag{2.31}$$

The forward problem is solved using a second-order finite-difference (FD) approximation of the 2D acoustic wave equation (Eq. 2.31), which is a widely used technique for numerical modelling of seismic wave propagation [Alford et al., 1974]. The first step in the finite difference modelling is the

discretization in the space and time domains, such that

$$\begin{aligned} x &= i\Delta h \quad i = (1, 2, \ldots, N_x) \,, \\ z &= j\Delta h \quad j = (1, 2, \ldots, N_z) \,, \\ t &= n\Delta t \quad k = (1, 2, \ldots, N_t) \,, \end{aligned} \tag{2.32}$$

where N_x, N_z denote the number of grid points in x and in z direction, and N_t is the total number of time steps. This discretization is referred to as the reference grid.

To calculate the spatial derivatives of the wavefield variables at the correct position, the staggered-grid technique is used [Levander, 1988; Virieux, 1986]. It means that not all quantities are defined at the points of the reference grid, but some quantities are defined as being half a grid point off the reference grid. The pressure p is defined at the points of the reference grid (j, i), whereas the particle velocity $\mathbf{w}$ is defined at the points half a grid off the reference grid. Furthermore, the temporal derivatives of the pressure are calculated at time step n, and the particle velocity is computed at a half time step.

The discretization of the wave equation means that the partial derivatives are replaced by the finite-difference operators. The discretization of the 2D acoustic wave equation leads to the following system of equations

$$\begin{aligned} \frac{\partial w_x(x,z,t)}{\partial t} &\approx \frac{w^{n+1/2}_{x|j,i+1/2} - w^{n-1/2}_{x|j,i+1/2}}{\Delta t} = \frac{1}{\rho_{j,i+1/2}} \frac{p^n_{j,i+1} - p^n_{j,i}}{\Delta h} \\ \frac{\partial w_z(x,z,t)}{\partial t} &\approx \frac{w^{n+1/2}_{z|j+1/2,i} - w^{n-1/2}_{z|j+1/2,i}}{\Delta t} = \frac{1}{\rho_{j+1/2,i}} \frac{p^n_{j+1,i} - p^n_{j,i}}{\Delta h} \\ \frac{\partial p(x,z,t)}{\partial t} \approx \frac{p^{n+1}_{j,i} - p^n_{j,i}}{\Delta t} &= K_{j,i} \frac{w^{n+1/2}_{x|j,i+1/2} - w^{n+1/2}_{x|j,i-1/2} + w^{n+1/2}_{z|j+1/2,i} - w^{n+1/2}_{z|j-1/2,i}}{\Delta h} \end{aligned} \tag{2.33}$$

The first order spatial and time derivatives are evaluated by the second-order finite-difference operators. This means that the Taylor series expansion is truncated after the first term.

2.3.1.1 Initial and boundary conditions

To solve the wave equation, the appropriate initial and boundary conditions must be specified. The initial conditions that describe the medium at rest before the excitation are

$$p(x, z, t = 0) = \partial_t p(x, z, t = 0) = 0 \,. \tag{2.34}$$

Furthermore, there are two types of boundary conditions, i.e. the free surface boundary conditions and the absorbing boundary.

Free surface condition. In all experiments presented in this thesis, the free surface is modelled using the so-called vacuum formalism [Bohlen and Saenger, 2006; Zahradnik et al., 1993]. It means that the medium above the free surface is treated as vacuum, i.e. the zero values are assigned to acoustic parameters and the density is close to zero to avoid division by zero. The vacuum formalism is attractive for its simplicity, especially when applied to the planar free surface. In the marine case,

which is considered in this work, the air-water interface is almost flat such that the planar surface assumption is enough. On the other hand, if the vacuum formalism is applied at the nonplanar free surface, the free surface gets an unfavorable staircase form. The theoretical proof that vacuum formalism fulfills the free-surface conditions is shown in Oprsal and Zahradnik [1999].

Another popular method for implementing the explicit free surface is to apply the mirroring technique [Levander, 1988]. The main reason for not using the explicit free surface formulation is that there are some differences in the implementation of the mirroring technique in various forward modelling codes, which I use to generate synthetic data (2D acoustic, 3D acoustic, 2D elastic). These differences are related to the location of the free surface on the standard staggered grid, which lead to discrepancies in the modelling of the source and receiver ghost effects.

Absorbing boundary. To attenuate the unwanted reflections from the computation edges of the model, the absorbing boundary conditions must be applied. The acoustic FD modelling code includes perfectly matched layers (PML) as boundary condition. The approach is based on the application of the complex coordinate stretching [Berenger, 1994; Chew and Weedon, 1994]. The complete update of the pressure wavefield with PML implementation requires additional equations to compute auxiliary PML variables [Kurzmann, 2012].

2.3.1.2 Accuracy and stability

To avoid numerical artefacts and instabilities, both the spatial and temporal sampling conditions must be satisfied. The grid dispersion is the result of the truncation of the Taylor series, which is made when we approximate spatial derivatives. The condition to avoid grid dispersion is related to the number of grid points n per minimum wavelength $\lambda_{\min}$ such that

$$\Delta h \leq \frac{\lambda_{\min}}{n} = \frac{V_{P_{\min}}}{n f_{\max}} \tag{2.35}$$

where $f_{\max}$ is the maximum frequency of the wavefield, and $V_{P_{\min}}$ is the minimum P-wave velocity. The minimum number of grid points per wavelength depends on the choice of the length and type of the FD operator [Köhn, 2011]. Because all numerical simulations are performed using the second-order FD operators in space and in time using the Taylor coefficients, the number of grid points per minimum wavelength is set to $n = 12$.

To ensure the stability of the finite-difference modelling, it is also necessary to satisfy the sampling criterion. For the 2D modelling, the ratio between the temporal and spatial sampling intervals is defined as

$$\Delta t \leq \frac{\Delta h}{\sqrt{2} V_{P_{\max}}} \tag{2.36}$$

where $V_{P_{\max}}$ is the maximum P-wave velocity in the model. The sampling criterion, the so-called Courant condition [Courant et al., 1928], means that the time step Δt must be less than the propagation time between two neighboring grid points.

2.3.2 Inversion

2.3.2.1 Conjugate gradient method

In order to improve the convergence rate of the gradient method, the conjugate gradient approach is applied [Mora, 1987]. The conjugate direction is a linear combination of the previous and the current steepest descent direction, such that

$$\delta \mathbf{c}_n = \delta \mathbf{m}_n + \beta_n \delta \mathbf{c}_{n-1} \tag{2.37}$$

where $\mathbf{c}_n$ is the conjugate gradient, $\delta \mathbf{m}_n$ is the steepest descent gradient, and β is the scalar that ensures that $\delta \mathbf{c}_n$ and $\delta \mathbf{c}_{n-1}$ are conjugate.

The weighting factor β is computed using the Polak-Ribiere method [Nocedal and Wright, 1999]

$$\beta_n^{PR} = \frac{\delta \mathbf{m}_n^T (\delta \mathbf{m}_n - \delta \mathbf{m}_{n-1})}{\delta \mathbf{m}_{n-1}^T \delta \mathbf{m}_{n-1}} \tag{2.38}$$

To prevent the situation that the second term in Eq. 2.37 may dominate the first term, the β parameter is defined as

$$\beta_n = \max \left\{ \beta_n^{PR}, 0 \right\}. \tag{2.39}$$

This feature provides the restart of the conjugate gradient method, if the objective function is very non-linear. If $\beta = 0$, a steepest descent direction is taken. The advantage of the conjugate gradient approach is that it can be applied without additional computational efforts.

2.3.2.2 Step length estimation

An estimation of an optimal step length is very important for the convergence of the gradient method. One method to estimate the step length is to use the line-search algorithm [Mora, 1987; Nocedal and Wright, 1999]. An alternative is to use the adaptive step length algorithm based on the parabolic curve fitting method [Kurzmann et al., 2009], which is implemented in the FWI code.

The estimation of an optimal step length μ_n comprises the following steps:

1. Define test step lengths. The three test step lengths at iteration n are defined as

$$\begin{aligned} \mu_{1,n} &= s \mu_{init,n} \\ \mu_{2,n} &= \mu_{init,n} \\ \mu_{3,n} &= \frac{\mu_{init,n}}{s} \end{aligned} \tag{2.40}$$

   where s is the scaling factor. The initial test step length must be defined at the first iteration, whereas for iterations $n > 1$, the optimal step length from the previous iteration becomes the initial test step length at the next iteration, such that $\mu_{init,n} = \mu_{n-1}$.

   In this work, $s = 0.5$, and $\mu_{init,1} = 0.01$, which gives the following set of test step lengths at iteration $n = 1$: $\mu_{1,n} = 0.005$, $\mu_{2,n} = 0.01$, $\mu_{3,n} = 0.02$. There are other possible choices for s and $\mu_{init,1}$, which may also provide a stable and efficient reduction of the data misfit function.

However, the choice of high test step lengths should be avoided, because it may lead to wrong model update [Kurzmann, 2012].

2. Calculate the model update. The current model $\mathbf{m}_n$ is updated using the conjugate gradient $\mathbf{c}_n$ and test step lengths, such that we obtain three different model updates

$$\mathbf{m}_{\mu_1,n} = \mathbf{m}_n - \mu_{1,n}\delta\mathbf{c}_n \tag{2.41}$$
$$\mathbf{m}_{\mu_2,n} = \mathbf{m}_n - \mu_{2,n}\delta\mathbf{c}_n$$
$$\mathbf{m}_{\mu_3,n} = \mathbf{m}_n - \mu_{3,n}\delta\mathbf{c}_n$$
$$\tag{2.42}$$

3. Compute the data misfit. The test forward simulations are performed to obtain the predicted data for different model updates: $\mathbf{d}_{\mathrm{mod}(\mu_1,n)}$, $\mathbf{d}_{\mathrm{mod}(\mu_2,n)}$, $\mathbf{d}_{\mathrm{mod}(\mu_3,n)}$, and to calculate the corresponding data misfit: $E(\mathbf{m}_{\mu_1,n})$, $E(\mathbf{m}_{\mu_2,n})$, $E(\mathbf{m}_{\mu_3,n})$.

4. Find the optimal step length. The true misfit function is approximated by fitting a parabola through the three points $(\mu_{i,n}, E(\mathbf{m}_{\mu_i,n}))$, where $i \in \{1,2,3\}$, such that the following system of equation is solved

$$E(\mathbf{m}_{\mu_1,n}) = a\mu_{1,n}^2 + b\mu_{1,n} + c \tag{2.43}$$
$$E(\mathbf{m}_{\mu_2,n}) = a\mu_{2,n}^2 + b\mu_{2,n} + c \tag{2.44}$$
$$E(\mathbf{m}_{\mu_3,n}) = a\mu_{3,n}^2 + b\mu_{3,n} + c \tag{2.45}$$
$$\tag{2.46}$$

where a, b, c are the unknowns. The minimum of the parabola defines the optimal step length μ_n at iteration n (Figure 2.1a) with

$$\mu_n = \frac{-b}{2a} \tag{2.47}$$

The step length estimation method assumes that the data misfit function has a parabolic shape as shown in Figure 2.1a. Because of the high nonlinearity of the inverse problem, the shape of the objective function is usually more complex and additional cases, illustrated in Figure 2.1, must be considered.

- The parabola is concave down (Figure 2.1b). The optimal step length is the test step length, which corresponds to the lowest data misfit: $\mu_n = \min\{E(\mathbf{m}_{\mu_1,n}), E(\mathbf{m}_{\mu_2,n}), E(\mathbf{m}_{\mu_3,n})\}$
- The minimum of the parabola is negative (Figure 2.1c). The optimal step length is the lowest test step length: $\mu_n = \min\{\mu_{1,n}, \mu_{2,n}, \mu_{3,n}\}$.
- The minimum of the parabola exceeds the upper limit $\mu_{\max}$ (Figure 2.1d). The optimal step length is set to the upper limit: $\mu_n = \mu_{\max} = 0.02$. The upper limit is defined to prevent the choice of very high step lengths and to stabilise the inversion process.

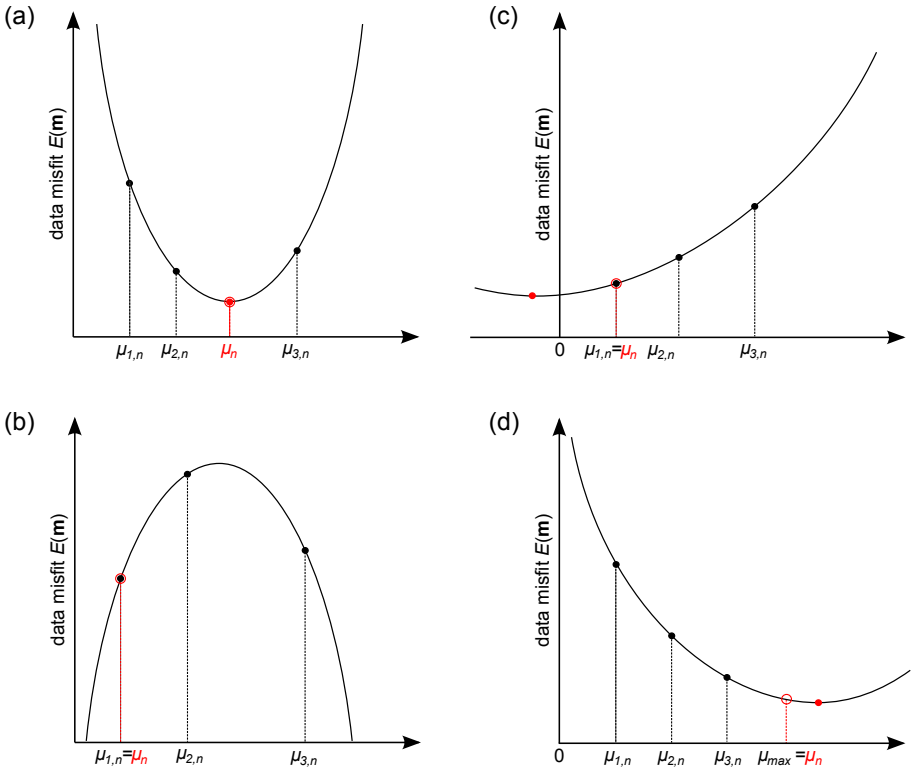


Figure 2.1: Step length estimation.

On the one hand, the application of the adaptive step length algorithm requires additional forward simulations, what results in an increase of the computational cost. But on the other hand, it improves the convergence rate of the gradient method, such that the number of iterations and the total computation time is reduced in comparison to the usage of a constant step length. Furthermore, it is not necessary to use all shots for test modellings to find an optimal step length, but it is enough to select a subset of shots [Kurzmann, 2012]. In the experiments presented in this work, approximately 20 % of available shots were used for test modellings to estimate an optimal step length.

2.3.2.3 Model update

The actual update of the model parameters is based on Pica et al. [1990], with an additional factor used to scale the gradient to the maximum of the model parameter. The model update is given by

$$\mu_n \delta \mathbf{c}_n = \mu_n \frac{\max(\mathbf{m}_n)}{\max(\mathbf{c}_n)} \mathbf{c}_n \ , \tag{2.48}$$

which means that the gradient of each model parameter is normalised to its maximum value and the optimal step length is related to the maximum of the actual model. This should assure the proper units of the model update and a proper distance of the gradient. In addition, there are hard constraints imposed on the model update to stabilise the inversion and to avoid unphysical negative parameter values.

2.3.2.4 Parallelization

To reduce the computational cost and the storage requirements of FWI algorithm, parallel computing is widely used [Virieux and Operto, 2009]. The computational costs of FWI are primarily related to the efficiency of the forward modelling, to the size of the computational domain, and to the size of the data. The size of the computational domain is controlled by the maximum offset of the data, the maximum imaging depth, and the frequency range. The size of the data, i.e. the number of shots, affects the total number of forward simluations required at each iteration step.

There are two levels of parallelization implemented in the acoustic FWI code, which are based on the principle of the domain decomposition (e.g. Bohlen [2002]) and shot parallelization [Kurzmann et al., 2009]. In the domain decomposition strategy, the model is divided into several subdomains, which are distributed to multiple processors. It requires a point-to-point communication to exchange wavefield values between neighboring subdomains, which is achieved with the Message Passing Interface (MPI). With the second level of parallelization, i.e. the shot parallelization, the shots are distributed to groups of processors, such that there is no communication required. The combination of both parallelization types leads to a significant speed-up of the inversion algorithm.

2.3.3 FWI algorithm

FWI requires a definition of the external input:

- Observed pressure data $\mathbf{d}_{\text{obs}}$
- Starting model $\mathbf{m}_0$
- Initial source time function

The acoustic FWI algorithm at iteration n comprises the following steps:

For each source located at $\mathbf{x}_s$

- Solve the forward problem (Eq. 2.23) for the current model $\mathbf{m}_n$ to generate and the pressure wavefield $p(\mathbf{x}, t; \mathbf{x}_s)$ and the predicted data $\mathbf{d}_{\text{mod},n}$
- Calculate the data residuals $\delta\mathbf{d}_n = \mathbf{d}_{\text{mod},n} - \mathbf{d}_{\text{obs}}$ and the misfit function $E(\mathbf{m}_n) = \frac{1}{2}\delta\mathbf{d}_n^T\delta\mathbf{d}_n$
- Calculate the residual wavefield $p^{'}(\mathbf{x}, t; \mathbf{x}_s)$, by backpropagating the data residuals $\delta\mathbf{d}_n$ from the receiver positions
- Calculate the gradient for each material parameter $\delta\mathbf{m}_{n,s}$ according to Eq. 2.27

For all sources

- Sum the gradients over all shots

  $\delta\mathbf{m}_n = \sum_{s=1}^{N_s} \delta\mathbf{m}_{n,s}$
- Apply the preconditioning operator P to the gradient

  $\mathbf{p}_n = P\delta\mathbf{m}_n$
- For all iterations $n > 1$ calculate the conjugate gradient direction $\mathbf{c}_n$

- Estimate the step length μ_n by a parabolic fit
- Update the model parameters

  $\mathbf{m}_{n+1} = \mathbf{m}_n - \mu_n \delta \mathbf{c}_n$

2.4 FWI requirements

Full waveform inversion is a highly non-linear inverse problem and the objective function contains several local minima. Because FWI is a local optimization method, we look for a solution that is only locally optimal, but gives no guarantee of finding a global minimum. Therefore, the choice of a staring model is of crucial importance for the success of the method. A good initial model for FWI should ensure convergence of the algorithm into the global minimum of the objective function. This can be achieved, when the predicted data generated for the starting model match the observed data within half a cycle of the minimum considered frequency [Sun and McMechan, 1992]. If the traveltime mismatch is greater than half a period, the cycle-skipping occurs. In such a case, the inversion will try to fit the calculated data to the wrong cycle of the observed data (Figure 2.2) and it will cause the objective function to converge to a local rather than a global minimum [Sirgue, 2003]. The existence of local minima are mainly due to a mismatch of traveltime between events, and the non-linearity of the inverse problem depends on the minimum frequency in the data and on the starting velocity model.

The convergence criterion can be also expressed in terms of traveltime errors and propagation distances measured in wavelengths [Pratt, 2008]. The traveltime error δt for a given seismic event has to be less than half a wavelength, therefore

$$\frac{\delta t}{T} < \frac{\lambda}{2cT} \quad \text{or} \quad \frac{\delta t}{T} < \frac{1}{2N_\lambda} \tag{2.49}$$

where T is the total arrival time of the event, N_λ is the propagation distance in wavelengths, and c is the background velocity. This strict condition can be relaxed if the number of wavelengths N_λ between the source and receiver is reduced. This, in turn, can be achieved by either using lower frequencies or by reducing the offset.

To summarize, there are three ways to reduce the traveltime error δt and to improve the convergence of FWI:

- improve the accuracy of the starting model,
- reduce the data offsets,
- reduce the starting frequencies.

Because the lack of very low frequencies is a general problem in real data applications, the reduction of the traveltime error by using lower frequencies is limited. The elimination of far offset information and the usage of near offset data only is also not a solution. As shown by Sirgue [2003], the long offset data are useful, because they provide information on the low wavenumber. Long-offset refracted arrivals are more sensitive to the macro variations in seismic velocity than the near offset data and they allow a better quantitative estimation of the velocity. But at the same time, the far offsets are

the most non-linear components of the data with the risk of cycle-skipping. This is related to the fact that the far offsets correspond to longer propagation distance and the traveltime error δt resulting from an incorrect velocity model is larger than for the shorter propagation distance. Therefore, large offsets should be carefully included into the inversion process.

The requirements on the starting model for FWI depends on the minimum frequency available in the data. This means that if the low frequencies are not present in the data, the starting model needs to be more accurate and must contain the long-wavelength structures. There are different methods that can be used to obtain an intial model for the full waveform tomography:

- First-arrival traveltime tomography (FATT) - FATT is often used to obtain a starting model for FWI [e.g. Bleibinhaus et al., 2009; Brenders and Pratt, 2007b; Dessa and Pascal, 2003; Malinowski and Operto, 2008; Pratt, 1999; Ravaut et al., 2004]. The traveltime tomography methods provide robust but low-resolution image [Pratt, 1999]. The spatial resolution of FATT is limited to the width of the first Fresnel zone, which is approximately given by $\sqrt{\lambda L}$, where λ is the dominant wavelength, and L is the propagation distance [Williamson, 1991]. Furthermore, the first-arrival traveltime tomography has limited penetration in presence of low velocity zones and it requires a wide range of offsets to generate sufficient ray coverage in the deeper parts of the model.

- Reflection traveltime tomography (RTT) - In RTT the reflection traveltimes are used to build velocity models. These models have higher resolution than from FATT, however the fit between the predicted and observed direct arrivals may be insufficient for FWI [Prieux et al., 2010]. Application of RTT to real data is shown in Sirgue et al. [2009]; Wang and Rao [2009]. A combination of FATT and RTT to build a starting model for FWI is shown in Prieux et al. [2010]; Shipp and Singh [2002].

- Stereotomography - It is a slope tomographic method that is based on the inversion of traveltimes and slopes of locally-coherent events in the pre-stack data cube. The advantage of stereotomography in comparison to the reflection traveltime tomography is the semi-automatic picking procedure, which is easier than the picking of continuous events. The joint use of refraction and reflection traveltimes in a multiscale stereotomography to build a starting model for FWI is shown in Prieux et al. [2010].

- Laplace domain inversion - The waveform in the Laplace domain can be regarded as a zero frequency component of a damped wavefield in the time domain. The Laplace domain inversion introduced by Shin and Cha [2008] can recover a smooth, long-wavelength velocity model from data without low frequency components. The main drawback of the method is that the penetration depth of the Laplace domain inversion depends on the maximum offset of the data and on the choice of Laplace domain constants. The smooth model from the Laplace domain inversion can be used as a starting model for Laplace-Fourier inversion or a conventional frequency-domain FWI [Shin and Cha, 2009].

The strong non-linearity of the inverse problem can be also reduced by applying a hierarchical multi-stage inversion based on time dependent, offset dependent, and frequency dependent data regularization [Virieux and Operto, 2009]. However, there is no general rule on how to design such a multi-stage inversion approach and it must be adapted to the particular inverse problem.

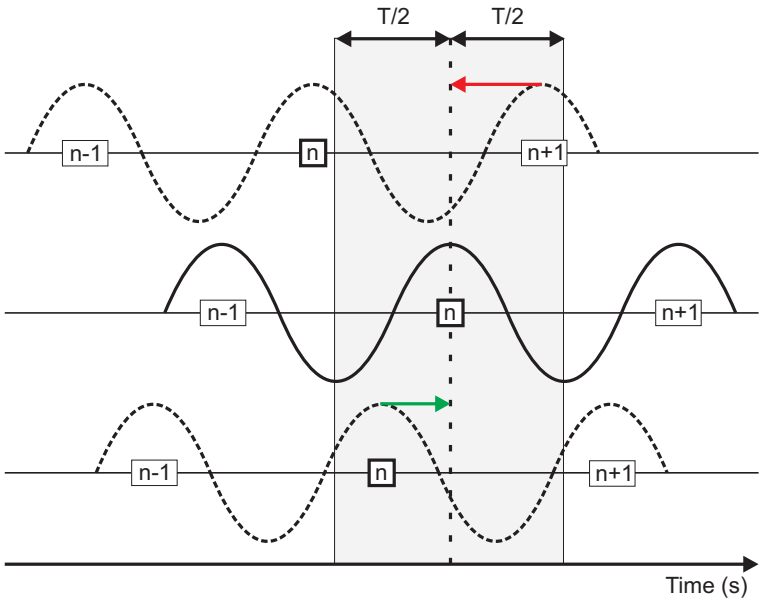


Figure 2.2: The cycle-skipping problem. The solid line represents a monochromatic seismograms. The dashed lines represents the predicted monochromatic seismograms. The predicted seismogram at the top has a time delay greater than $T/2$ and the FWI will try to fit the calculated data to the wrong cycle of the observed seismogram. After Virieux and Operto [2009]

Chapter 3

Choice of inversion strategies

In this Chapter, I illustrate the application of acoustic full waveform inversion to synthetic streamer data. The synthetic data set is generated for a modified acoustic Marmousi2 model. The acquisition geometry of the numerical experiment is based on the acquisition setup of the field data acquired in the North Sea and presented in Chapter 7.

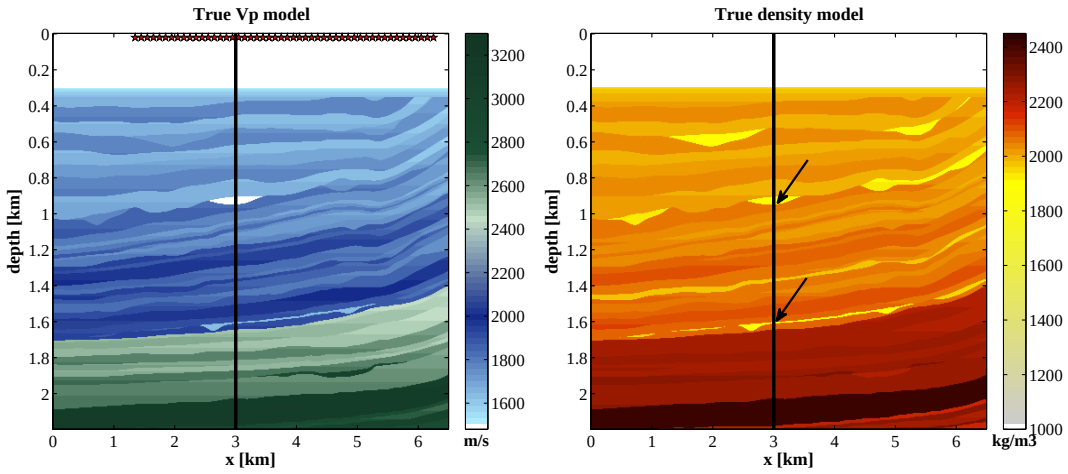
First, I investigate the performance of different preconditioning operators applied to the gradient of the misfit function. The aim of the gradient preconditioning is to suppress the strong artefacts in the vicinity of sources and receivers and to correct for the amplitude loss with depth due to geometrical spreading.

Afterwards, I discuss the implementation of the multi-scale inversion approach with a special focus on the efficient selection of frequency bands. To verify the method, I invert the synthetic data using a poor starting velocity model.

Finally, I assess the performance of FWI for different minimization criteria by inverting two synthetic data sets affected by noise. In the first experiment, I perform acoustic inversion of elastic data. The second data set is generated by adding the swell noise to the acoustic noise-free data.

3.1 Numerical experiment settings - modified Marmousi2 model

The numerical tests are based on a part of the Marmousi2 model [Martin et al., 2006]. The P-wave velocity and density models are shown in Figure 3.1. The model consists of a water layer above layered sediments with two hydrocarbon reservoirs: a gas lens at a depth of 900 m and a thin oil sand layer at 1600 m, with an average vertical thickness of 38 and 40 m, respectively. Sedimentary layers are primarily composed of shale. The average P-wave velocity is 2050 m/s, with the minimum velocity of 1480 m/s in the water layer and the maximum velocity of 3150 m/s. The model is 6.5 km long and 2.35 km deep. The water layer has a thickness of 300 m. A free surface boundary condition is applied at the top of the model, thus the simulated pressure waveform contains both free surface multiples, as well as the source and receiver ghost. The motivation to select this particular subregion of the original Marmousi2 model is that the real data set analysed in this thesis was recorded in an area with similar geological properties.



(a) (b)

Figure 3.1: Modified acoustic Marmousi2 model. (a) True V_P model, and (b) density model. The stars denote the shot point locations. The arrows indicate the gas lens and the thin oil sand layer.

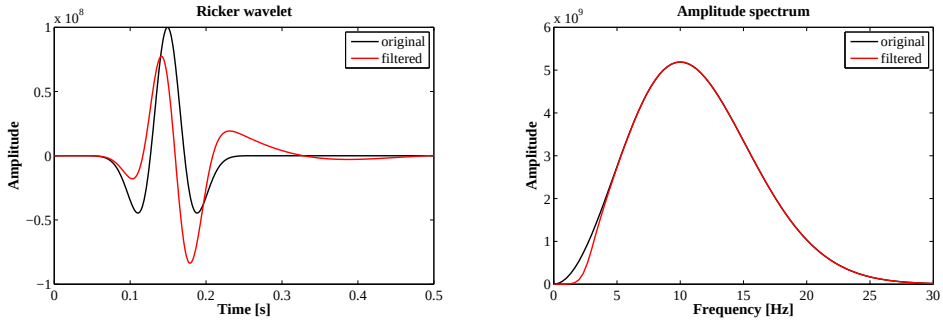


Figure 3.2: Ricker source wavelet. Original Ricker wavelet with the central frequency of 10 Hz (black line) and the wavelet after applying the high-pass Butterworth filter with a cut-off frequency of 3 Hz (red line). (a) Time domain representation, (b) amplitude spectrum.

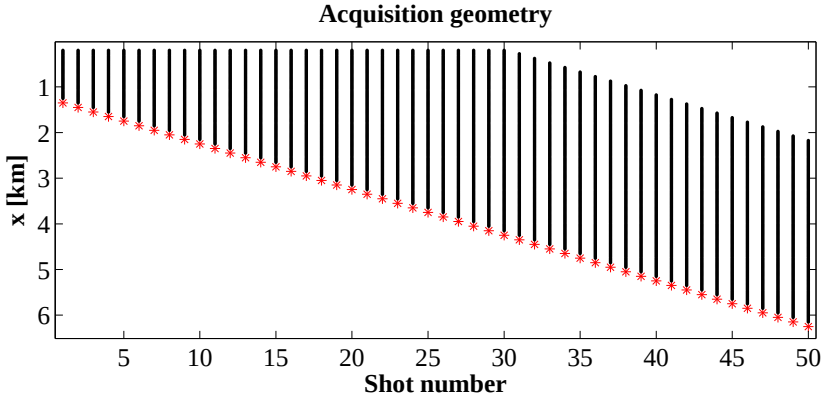


Figure 3.3: Acquisition geometry of the Marmousi2 data set. Stars denote shot location and points represent hydrophones.

Modified Marmousi2 model	
Model size	6.5 km x 2.5 km
Average V_P	2050 m/s
Minimum V_P	1480 m/s
Maximum V_P	3150 m/s
Water layer thickness	300 m
Acquisition parameters	
Number of shots	50
Shot spacing	100 m
Shot depth	7.5 m
Max number of hydrophones in a streamer	160
Hydrophone spacing	25 m
Hydrophone depth	7.5 m
Minimum offset	100 m
Maximum offset	4075 m
Modelling parameters	
Grid size	1300 x 500
Grid spacing	5 m
Time sampling	8e-4 s
Number of time samples	3750
Recording length	3 s
Source wavelet	Ricker 3-20 Hz

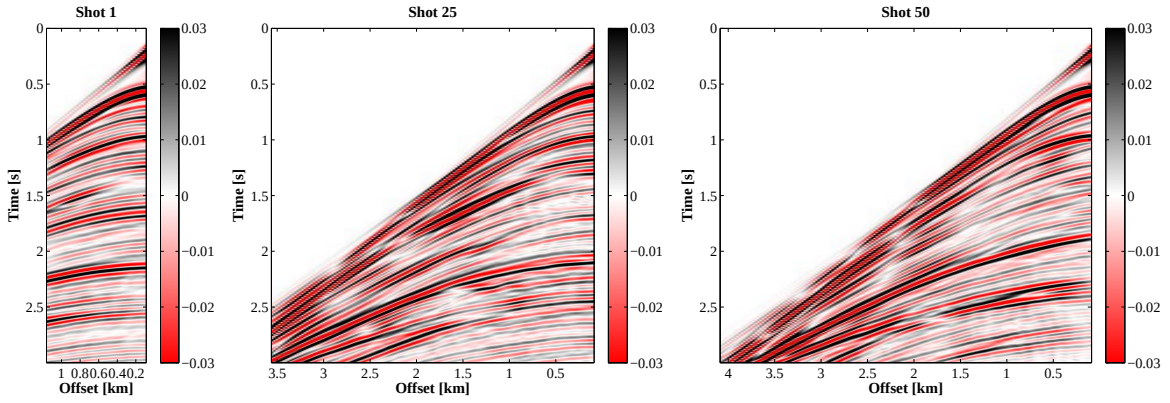
Table 3.1: Modelling and acquisition parameters of the Marmousi2 data set.

3.1.1 Modelling parameters and seismic data

The design of the numerical experiment aims to reflect the real measurement conditions of marine reflection seismics. The pressure source is located 7.5 m below the air-water interface. The source time function is a Ricker wavelet with a central frequency of 10 Hz. Since the lack of low frequencies is a general problem in seismic recording, the original wavelet was filtered with a high-pass Butterworth filter with a cut-off frequency of 3 Hz (Figure 3.2).

The acquisition setup mimics the conventional single-component streamer survey. The streamer consists of up to 160 hydrophones with a spacing of 25 m located at 7.5 m depth. I have simulated a moving streamer acquisition with the source points moving from the left to the right of the model and towing a streamer behind. The near offset is 100 m, and the far offset is increasing from 1150 m, for the first shot located in the left part of the seismic line at 1350 m, to the maximum of 4075 m. As a consequence, the receiver fold is higher in the left and central part and lower in the right part of the model (Figure 3.3).

The synthetic noise-free data is calculated using the 2D acoustic finite-difference forward modelling. A total of 50 shot gathers are generated at a 100 m interval, with 3 seconds of data. The modelling and acquisition parameters are listed in Table 3.1. The exemplary common shot gathers are shown in Figure 3.4.



(a) (b) (c)

Figure 3.4: Modified acoustic Marmousi2 model. Exemplary shot gathers.(a) Shot 1 located at $x = 1.35$ km, (b) Shot 25 located at $x = 3.75$ km, (c) Shot 50 located at $x = 6.25$ km.

3.1.2 Error analysis

In order to quantitatively assess the inversion results, I measure initial and final errors, both in the data and in the model space. Since the waveform inversion is an ill-posed and non-unique problem, this will help to determine whether the reduction of residuals is followed by a better estimation of model parameters.

The error between the modelled d_{mod} and observed data d_{obs} is measured by the least-squares error

$$L_{2\text{error}} = \frac{1}{2}\sum_{i=1}^{N}(d_{\text{mod}|i} - d_{\text{obs}|i})^2 \ . \tag{3.1}$$

where N is the size of the data vector.

To assess the quality of the final velocity models, the relative error between the real and the reconstructed P-velocities within the area of the model update is calculated as

$$V_{P\text{error}} = \frac{1}{M}\sum_{i=1}^{M}\frac{\mid V_{P\text{inv}|i} - V_{P\text{true}|i} \mid}{\mid V_{P\text{true}|i} \mid} \ , \tag{3.2}$$

where $V_{P\text{inv}|i}$ is the inverted P-wave velocity for the ith grid point, $V_{P\text{true}|i}$ is the true velocity value, and M is the size of the model vector within the area of the model update.

3.2 Gradient preconditioning

In the conventional gradient method, the model update is estimated by a scalar step length, μ_n, multiplied by the steepest-descent or conjugate gradient $\delta\mathbf{m}_n$

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mu_n\delta\mathbf{m}_n \tag{3.3}$$

However, the gradient approach can be significantly improved when the gradient is preconditioned with the inverse Hessian, or an approximation of the inverse Hessian [Pratt et al., 1998]. Such a gradient preconditioning improves the resolving power and convergence properties of the algorithm. Furthermore, it removes from the gradient the effect of geometrical spreading of wave propagation from the source. In this way, the preconditioned gradient provides a good balance between shallow and deep perturbations [Virieux and Operto, 2009] .

Due to the extensive computational and memory requirements, the inverse Hessian is not often used in geophysical inverse problems [Sheen et al., 2006]. Because the computation of the approximate Hessian matrix is expensive as well, Shin et al. [2001] proposed a less computationally demanding approximation to the Hessian, which is called the pseudo-Hessian matrix. Another approach is to use a quasi-Newton method such as the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method [Brossier et al., 2009; Nocedal and Wright, 1999]. In the L-BFGS methods an improved estimate of the inverse Hessian matrix is computed at each iteration without forming or storing the Hessian matrix explicitly. The method requires an initial inverse Hessian matrix, which can be provided by the diagonal terms of the pseudo-Hessian matrix, and it maintains a history of the previous updates, which means that a few gradients from the previous iterations need to be stored. As shown by Brossier et al. [2009] the L-BFGS algorithm provides an improved convergence when compared to the conjugate gradient method.

On the other hand, Köhn [2011]; Wang and Rao [2009] showed that the effect of geometrical spreading of the waveform amplitudes can be also effectively compensated by multiplying the gradient with a depth dependent scaling factor.

In this section, I compare the effectiveness of different linear scaling functions used as a preconditioning operator. Furthermore, I test the wavefield-based preconditioning proposed by Fichtner et al. [2009]. The performance of these different preconditioning methods is illustrated with an acoustic inversion applied to the synthetic Marmousi2 data.

3.2.1 Linear scaling functions

Due to the geometrical spreading effect of the propagating wavefront, the amplitude of the signal is very high in the vicinity of the source and it is decreasing with the distance. This results in high values of the cross-correlation of the forward and backpropagated wavefields at the grid points around the source, whereas the deep part of the model gives a weaker contribution to the gradient [Causse et al., 1999].

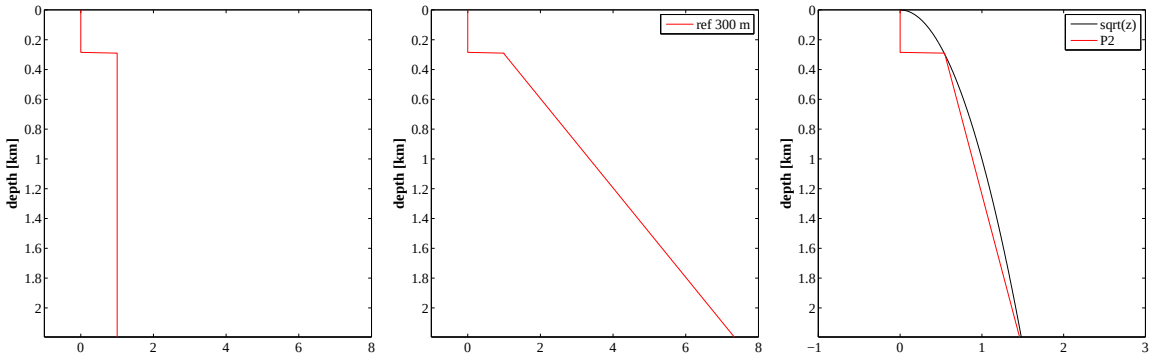
3.2.1.1 P_1 operator

To correct for the amplitude loss with depth and to enhance deeper parts of the model, a simple depth-dependent linear preconditioning operator can be applied [Köhn, 2011] such that the updated model is defined as

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \mu_n P \delta \mathbf{c}_n \tag{3.4}$$

where P is the preconditioning operator, and $\delta \mathbf{c}_n$ is the conjugate gradient at iteration n.

A spatial preconditioning operator P_1 can be defined as



(a) Test 2 (b) Test 3 (c) Test 4

Figure 3.5: Gradient tapers. (a) A taper in the water layer, (b) the preconditioning operator P_1, (c) the preconditioning operator P_2. Horizontal axis represents the scaling amplitude.

$$P = \begin{cases} 0, & \text{if } z \leq z_{\text{ref}} \\ \dfrac{z}{z_{\text{ref}}}, & \text{if } z > z_{\text{ref}} \end{cases} \tag{3.5}$$

where z_{ref} is the reference depth, which can be related to the depth of the seafloor. This preconditioning combines a taper in the water layer with a linear gradient scaling (Figure 3.5b). Applying a taper in the water layer prevents velocity perturbation of this part of the model and eliminates high gradient values around the acquisition geometry. The main disadvantage of this preconditiong operator is that it depends on the choice of the reference depth z_{ref}.

3.2.1.2 P_2 operator

I designed a modified preconditioning operator P_2, which is related to decay of amplitudes with increasing depth. In a homogeneous 2D medium, the amplitude decay of the wavefront is proportional to $1/\sqrt{r}$, where r is the distance from the source. Within the area of the model update, a linear scaling function is calculated, such that it approximates the $1/\sqrt{z}$ amplitude decay (Figure 3.5c). It can be seen, that the operator P_2 has a much smaller scaling ratio than the operator P_1.

3.2.2 Wavefield-based preconditioning

The wavefield-based preconditioning proposed by Fichtner et al. [2009] aims to reduce the high sensitivity of the gradient to the amplitudes near sources and receivers. The preconditioning operator is applied to the gradients of the individual shots such that

$$P\delta m_n = \sum_{s=1}^{N_s} P_s \delta m_{n,s} \ , \tag{3.6}$$

where s is the shot index, N_s is the total number of shots, P_s is the preconditioning operator for shot s, and $\delta m_{n,s}$ is the gradient of shot s at iteration n.

The preconditioning operator P_s is given by

$$P_s(\mathbf{x}) = \frac{1}{Q_s(\mathbf{x}) + Q'_s(\mathbf{x})} \tag{3.7}$$

where

$$Q_s(x) = q\frac{\max_t|p(\mathbf{x},t)| - \min_{\mathbf{x}}(\max_t|p(\mathbf{x},t)|)}{\max_{\mathbf{x}}(\max_t|p(\mathbf{x},t)|) - \min_{\mathbf{x}}(\max_t|p(\mathbf{x},t)|)} + 1\ , \tag{3.8}$$

and

$$Q'_s(x) = q'\frac{\max_t|p'(\mathbf{x},t)| - \min_{\mathbf{x}}(\max_t|p'(\mathbf{x},t)|)}{\max_{\mathbf{x}}(\max_t|p'(\mathbf{x},t)|) - \min_{\mathbf{x}}(\max_t|p'(\mathbf{x},t)|)} + 1\ , \tag{3.9}$$

where $\max_t|p(\mathbf{x},t)|$ and $\max_t|p'(\mathbf{x},t)|$ are the temporal maxima of the forward propagated wavefield $p(\mathbf{x},t)$ and the backpropagated wavefield $p'(\mathbf{x},t)$, respectively, which are defined as functions of the position $\mathbf{x}$. Function $Q_s(x)$ is proportional to the amplitude of the forward propagated wavefield and it ranges between 1 and $q+1$, and $Q'_s(x)$ is proportional to the amplitude of the backpropagated wavefield, and it ranges between 1 and $q'+1$. In Fichtner et al. [2009], the scaling factors q and q' are empirically defined to match their specific application, such that $q = 10$ and $q' = 1$.

3.2.3 Inversion results

To demonstrate the effect of different preconditioning methods, I performed various synthetic inversion tests using the acoustic data set generated for the Marmousi2 model. To focus only on the V_P model reconstruction, the density model and the source time function are assumed to be known. The starting V_P model is a 1D smooth representation of the true velocity model.

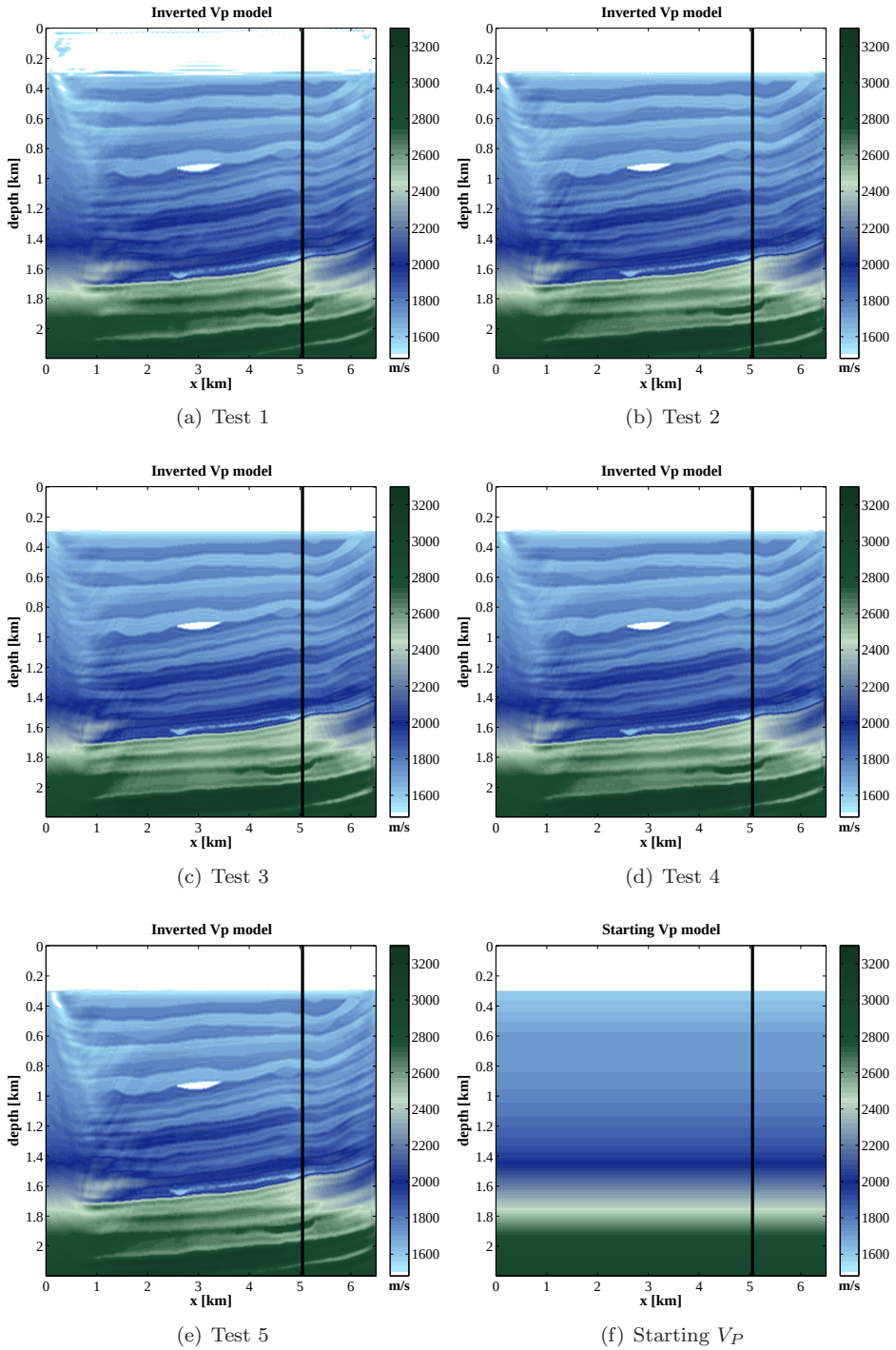
- **TEST 1** - there is no preconditioning applied to the gradient;
- **TEST 2** - a spatial taper is applied to the gradient, such that it sets the gradient in the water layer to zero, i.e. it turns off the velocity update in this part of the model (Figure 3.5a);
- **TEST 3** - the preconditioning operator P_1 is applied (Figure 3.5b). The reference depth is defined at 300 m, which correspons to the depth of seafloor in the true model;
- **TEST 4** - the preconditioning operator P_2 is applied (Figure 3.5c). The reference depth is defined at 300 m;
- **TEST 5** - the wavefield-based preconditiong is combined with a spatial taper in the water layer;

The inverted velocity models are shown in Figure 3.6. The velocity structures are very well reconstructed and apart from the artefacts in the water column, when no preconditioning is applied, the inverted models are quite similar. However, there are some discrepancies between results especially in the deep part of the model and at the sides of the model, where the illumination is poor. A comparison of the depth profiles at $x = 5$ km of the inverted velocity models and the true V_P shown in Figure 3.7 reveals more differences between various preconditioning methods. Without gradient preconditioning (Figure 3.7a), there are artefacts in the water column, especially close to the source position. The accuracy of the reconstructed V_P is good in the shallow part of the model, up to a depth of 1 km, but the resolution is decreasing with depth. Applying a taper in the water layer turns off the velocity update in this part of the model and eliminates the artefacts related to the acquisition

geometry (Figure 3.7b). Furthermore, the quality and the resolution of the final velocity image is improved. With an additional gradient scaling with depth, P_1 and P_2 operators, we can observe further improvements in the reconstruction of deeper structures (Figure 3.7c,d). The accuracy of the inverted velocity model for the wavefield-based preconditioning (Figure 3.7e) is comparable with the result obtained with the gradient tapering in the water layer. This relatively poor performance might be related to the choice of the scaling factors Q_s and Q'_s (Eq 3.8 and Eq 3.9).

The effect of various preconditioning methods is very well illustrated by the velocity gradient. Figure 3.8 shows the V_P gradients without preconditioning and after applying various preconditioning operators at the first iteration step. In addition, the depth profiles of V_P gradients at $x = 5$ km are shown in Figure 3.9. The velocity gradient without preconditioning is dominated by high values near sources and receivers (Figure 3.8a, Figure 3.9a). When the depth-dependent gradient scaling is applied (operator P_1), the contribution of deeper structures to the gradient is significantly enhanced (compare Figure 3.8b with Figure 3.8c, and Figure 3.9b with Figure 3.9c). This effect is smaller for the preconditioning operator P_2, which provided a more uniform distribution of the gradient amplitude with depth (Figure 3.9d). In the case of the wavefield-based preconditioning, the gradient is dominated by shallow structures.

Finally the evolution of the V_P model error for all inversion experiments is illustrated in Figure 3.10. The relative error between the real and the reconstructed P-velocities are calculated with Eq 3.2. For the inversion of the noise-free Marmousi2 data set with a good starting V_P model, the highest final model error is obtained when no preconditioning is applied. On the other hand, the best reduction of the model error and the best performance of the inversion is related to the usage of the preconditioning operator P_1 (TEST 3).

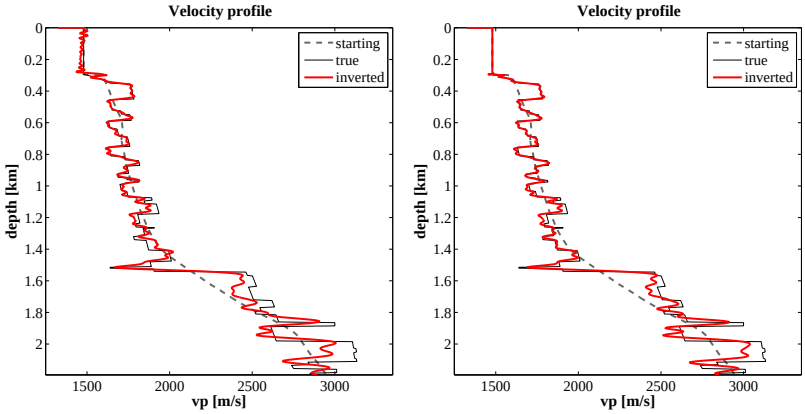


(a) Test 1 (b) Test 2

(c) Test 3 (d) Test 4

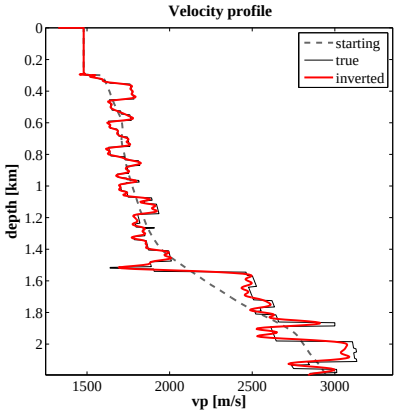
(e) Test 5 (f) Starting V_P

Figure 3.6: Marmousi2 example. FWI results. (a)-(e) Comparison of inverted V_P models for different gradient tapers. (f) Starting V_P model. The black line indicates the location of the V_P profiles shown in Figure 3.7.

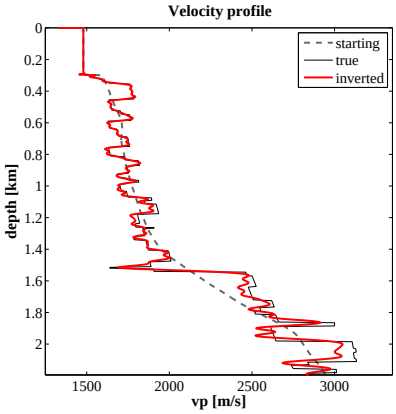


(a) Test 1

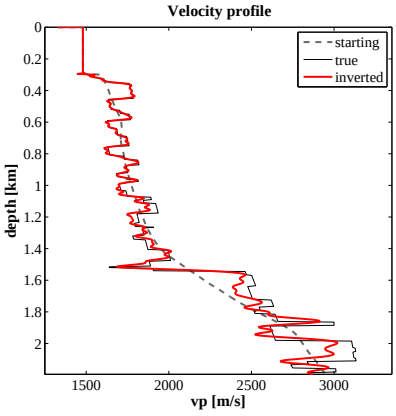
(b) Test 2



(c) Test 3

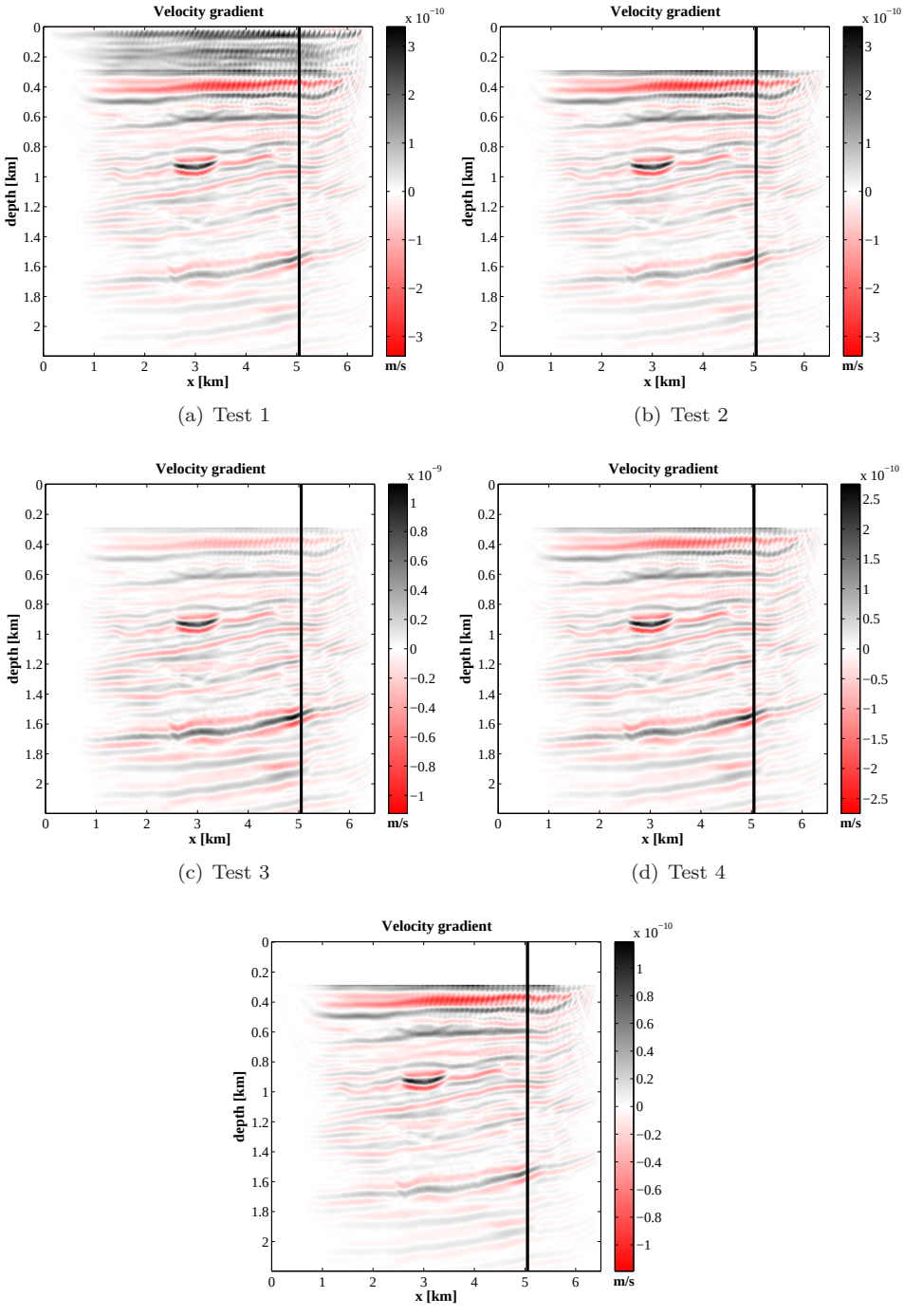


(d) Test 4



(e) Test 5

Figure 3.7: FWI results. Comparison of V_P profiles at $x = 5$ km for different gradient tapers; the black solid line indicates the true model, the gray dashed line indicates the initial model, and the red solid line indicates the inverted V_P model.



(a) Test 1

(b) Test 2

(c) Test 3

(d) Test 4

(e) Test 5

Figure 3.8: FWI results. Comparison of V_P gradients at the first iteration step for different gradient tapers.

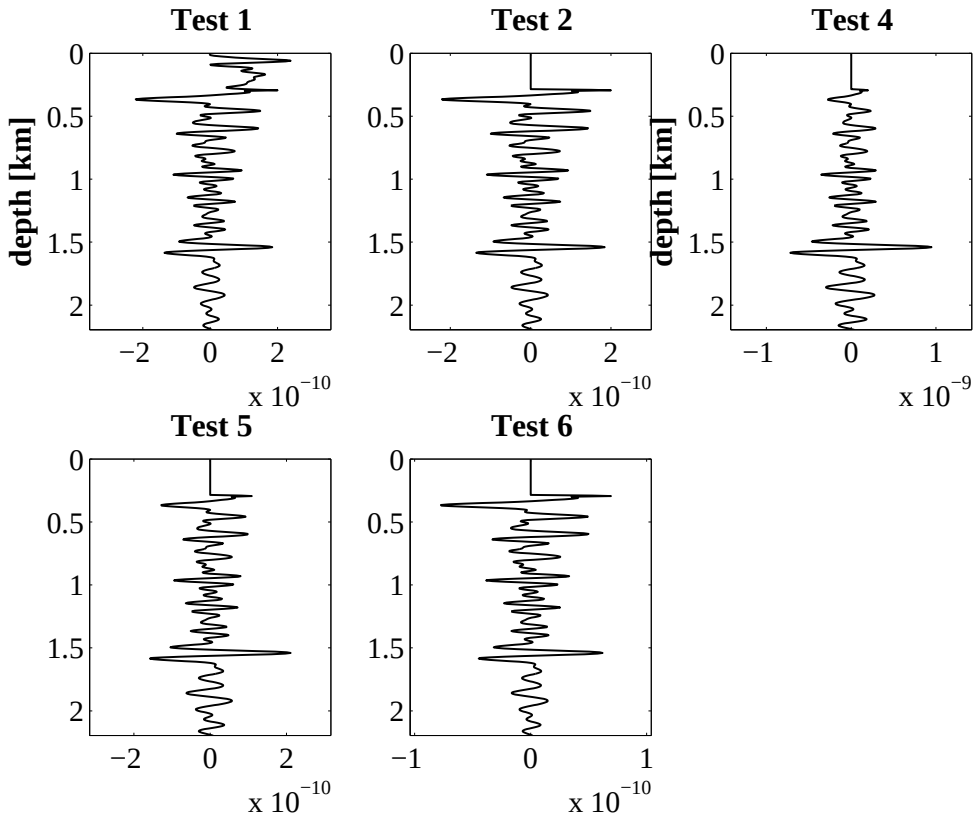


Figure 3.9: Comparison of preconditioned velocity gradients at the first iteration step for different gradient tapers. The profiles are intersecting the velocity gradients shown in Figure 3.8 at $x = 5$ km. The x-axis corresponds to the gradient amplitude.

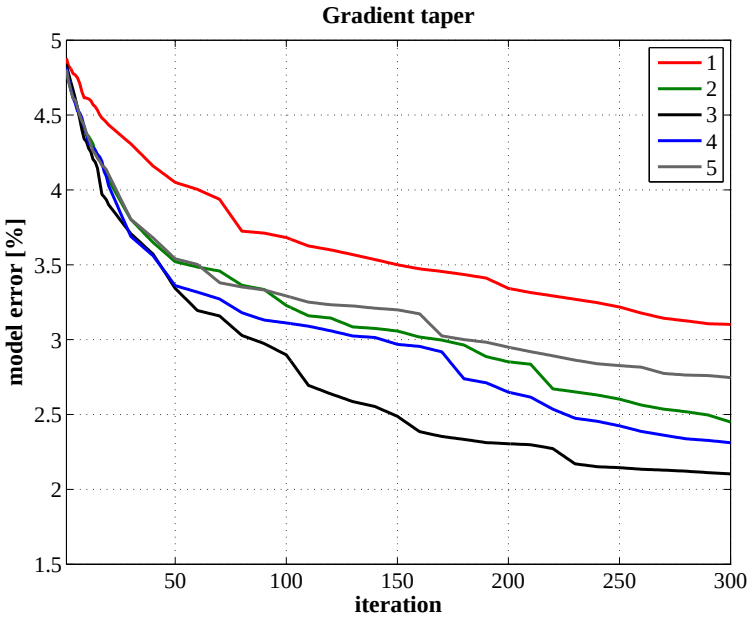


Figure 3.10: Summary of FWI results. V_P error of the inverted models for different gradient tapers. The numbers refer to inversion tests.

3.3 Multi-scale inversion

Due to the high non-linearity of the full waveform inversion, there are numerous local minima present in the objective function. The Born approximation, which is used to estimate partial-derivative wavefields, requires that the predicted data for the starting model should match the observed traveltimes within half a cycle of the dominant frequency [Beydoun and Tarantola, 1988]. If the half-wavelength criterion is not fulfilled, cycle skipping occurs, and the waveform inversion will converge toward a local minimum.

To reduce the non-linearity of the inverse problem and to mitigate the local minima problem, the multi-scale inversion approach can be applied [Bunks et al., 1995]. This means that the inversion starts at low frequencies and higher frequency content is gradually added. Because the objective function at low frequencies is more linear with respect to the model perturbations than at high frequencies, such a multi-scale strategy improves the chance of the inversion to reach the global minimum [Sirgue and Pratt, 2004]. In other words, it is easier to satisfy the half-wavelength condition imposed on the starting model, when the dominant frequency present in the data is low.

The multi-scale strategy in the time-domain waveform inversion is realized by low-pass filtering [Bunks et al., 1995] or band-pass filtering [Boonyasiriwat et al., 2009] of the data. In the frequency-domain, it is common to apply the sequential or the multiple frequency approach [Virieux and Operto, 2009]. In the sequential approach, only one frequency is inverted at a time and higher frequencies are successively included [Brenders and Pratt, 2007b; Operto et al., 2006; Ravaut et al., 2004]. Whereas in the multiple frequency approach, a group of frequencies is inverted simultaneously at each inversion stage [Brossier et al., 2009; Sourbier et al., 2009]. In comparison to the sequential, single-frequency method, inversion of multiple frequencies improves the signal-to-noise ratio and the robustness of FWI, but it is also computationally more expensive [Brossier et al., 2009]. The advantage of the time domain FWI is that the inversion of multiple, overlapping frequencies at each iteration step is carried out without extra computational costs.

In this section, I verify the effectiveness of the multi-scale inversion by inverting the synthetic Marmousi2 data with a poor starting velocity model. In addition, I investigate the problem of selecting the first frequency band and show the impact of the number of iterations per inversion stage on the FWI results.

3.3.1 Selection of frequency bands

The selection of frequency bands for the multi-scale inversion is an important aspect. If there are too many frequency bands, the total computation time of the method will increase. On the other hand, the selection of too few frequency bands may not be enough to avoid the cycle-skipping artefacts. Therefore, the choice of optimal frequencies is of high importance to reduce the computational cost of the waveform inversion and to take full advantage of the multi-scale approach.

To design the multi-scale inversion in the time-domain, I adopted the strategy proposed by Sirgue and Pratt [2004], which was originally developed for the frequency-domain waveform inversion. The approach for selecting frequencies is based on a continuous coverage of vertical wavenumbers.

For a given frequency f_n, and a range of offsets, the verical wavenumber coverage k_z for a 1D case is limited to the range $[k_{z\,\min}(f_n), k_{z\,\max}(f_n)]$, with the minimum and maximum wavenumbers, $k_{z\,\min}$

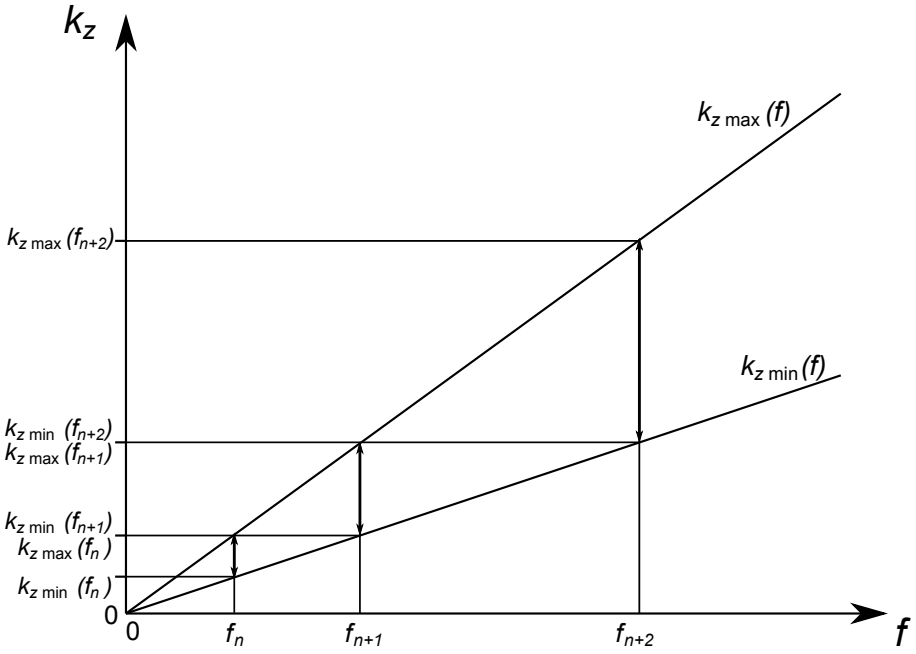


Figure 3.11: Strategy for choosing optimal frequencies in the frequency-domain waveform inversion (after Sirgue and Pratt [2004]).

and $k_{z\,\max}$ given by

$$k_{z\,\min}(f_n) = 4\pi f_n \alpha_{\min}/c_0 \ ,$$
$$k_{z\,\max}(f_n) = 4\pi f_n/c_0 \ , \tag{3.10}$$

where c_0 is the homogeneous background velocity, and

$$\alpha_{\min} = \frac{1}{\sqrt{1+R_{\max}^2}} \tag{3.11}$$

where $R_{\max} = h_{\max}/z$ is the half-offset to depth ratio, $h_{\max}$ is the maximum half-offset, and z is the depth to the target. Since each source-receiver pair contributes a different vertical wavenumber to the gradient image, the minimum wavenumber is produced by the furtherst offset, and the maximum wavenumber is the contribution from the nearest offsets.

Because each frequency has a limited contribution to the image, the frequency selection should allow continuous sampling of the wavenumber spectrum such that

$$k_{z\,\min}(f_{n+1}) = k_{z\,\max}(f_n) \tag{3.12}$$

where f_{n+1} is the next frequency to be selected. This strategy is illustrated in Figure 3.11. Using Eq. 3.10 and Eq. 3.12, finally we get

$$f_{n+1} = \frac{f_n}{\alpha_{\min}} \tag{3.13}$$

Eq. 3.13 shows that the optimum frequency increment is not constant and it increases with frequency.

3.3.1.1 Implementation

The low-pass filtering is performed using the recursive IIR (infinite impulse response) Butterworth filter, which I implemented in the waveform inversion code. The Butterworth filter is applied to the source wavelet and to the data in the time domain, so there is no need to transform the data to the frequency domain.

The Butterworth filter is characterized by an amplitude response that is maximally flat in the passband and monotonic overall. Another advantage is the causality of the filter, which is helpful for the filtering of first arrivals. Furthermore, because of the gentle slopes of the Butterworth filter, the undesirable ringing does not occur in the time domain (the so-called "Gibbs effect"), so there is no additional taper function required.

The maximum frequency of every frequency band is chosen with the above described strategy, where f_n defines the cut-off frequency for the Butterworth low-pass filter for the nth frequency band. This filtering method corresponds to the inversion of multiple overlapping frequencies in the frequency domain FWI.

3.3.2 Poor starting velocity model - inversion of all frequencies

If the very low frequencies are not present in the data, like in the Marmousi2 test case, the success of the waveform inversion strongly depends on the choice of the starting model. In such a case a good starting model containing the long-wavelength structures of the subsurface is required. Otherwise, the inverse problem tends to converge to a local minimum and produces unrealistic results. In the inversion tests performed in the previous section, the starting V_P model was a 1D smooth representation of the true velocity model (Figure 3.6f). However, it is very unlikely to have such a good initial information in the case of real data applications.

In this inversion test, I modified the good initial V_P model by adding +100 m/s to the background velocity. The full frequency content of the data and the preconditioning operator P_2 are used at each iteration step. The final FWI result is shown in Figure 3.12. Because the long wavelength structures of the true model are not present in the starting model, the single-scale waveform inversion converged to a local minimum and failed to reconstruct the true velocity structures.

The failure of this inversion test is related to the cycle-skipping problem. Figure 3.13 shows a comparison between the observed and initial data for some exemplary traces. We can observe a significant mismatch between the observed and modelled data, which mainly concerns the diving/refracted wave at middle and far offsets. Because the initial velocity is too high, the predicted traveltimes of the starting model do not match the observed traveltimes within half a cycle of the observed data. As a consequence, the waveform inversion ended up in a local minimum of the misfit function.

3.3.3 Poor starting velocity model - multi-scale inversion

In order to reduce the non-linearity of the inverse problem, I modified the previous inversion experiment by applying the multi-scale approach. The maximum frequency for each frequency band was calculated using Eq. 3.13. The depth to the target, $z = 2.35$ km, is the maximum depth of the model and the maximum half-offset $h_{\max} = 2.0$ km, which results in $\alpha_{\min} = 0.75$. The frequency bands are applied sequentially with the following maximum frequencies $f_{\max} = (3, 4, 5.3, 7.1, 9.4, 12.5, 16.6, 22)$,

which define the cutoff frequency for the Butterworth low-pass filter, with an order $n = 6$. There are altogether eight frequency bands, and except for the last inversion stage that covers the full frequency content of the data, 20 iterations are performed for each frequency band. The maximum iteration number is 300.

The minimum and maximum vertical wavenumbers can be computed using Eq. 3.10, with $c_0 = 2050$ m/s, which corresponds to the average velocity of the true model. For the first maximum frequency $f_1 = 3$ Hz, we obtain

$$k_{z\ \min}(f_1) = 0.0138 \text{ rad} \cdot \text{m}^{-1} ,$$
$$k_{z\ \max}(f_1) = 0.0184 \text{ rad} \cdot \text{m}^{-1} .$$

With $k = 2\pi/\lambda$, where λ is the wavelength, the minimum and maximum wavelengths related to the vertical resolution are

$$\lambda_{\min}(f_1) = 341 \text{ m} ,$$
$$\lambda_{\max}(f_1) = 455 \text{ m} .$$

For the maximum frequency of the data $f_8 = 22$ Hz, we get

$$k_{z\ \min}(f_8) = 0.1011 \text{ rad} \cdot \text{m}^{-1} ,$$
$$k_{z\ \max}(f_8) = 0.1348 \text{ rad} \cdot \text{m}^{-1} .$$

and

$$\lambda_{\min}(f_8) = 46 \text{ m} ,$$
$$\lambda_{\max}(f_8) = 62 \text{ m} .$$

Figure 3.14 shows both the indermediate inversion results at the end of selected inversion stages and the final FWI result. The result from each lower frequency band is used as the starting model for the next higher frequency inversion. We can observe a dramatic improvement in the reconstruction of V_P structures with comparison to the previous experiment. Although the absolute amplitude of the starting frequency band is very low (Figure 3.14a), because the original wavelet was filtered with the high-pass Butterworth filter with a cut-off frequency of 3 Hz, the very low frequencies contributed to the reconstruction of the large-scale structures. Each higher frequency band yields a recovery of a more detailed V_P image. Most of the structures in the final V_P model are reconstructed successfully, however the accuracy is worse in the deeper part of the model. In general, the quality of the inverted model obtained from the multi-scale inversion of a poor initial V_P model is comparable to the result obtained with a good starting velocity information (Figure 3.6).

Figure 3.15a shows a comparison between the observed and initial data for the first frequency band. We can observe that the initial data fulfill the half-wavelength criterion and the cycle-skipping problem, which caused a failure of the previous single-scale inversion experiment, is not present in this case. The comparison between the observed data and predicted data generated for the final FWI result is illustrated in Figure 3.15b.

3.3.3.1 Impact of the number of iterations per frequency band

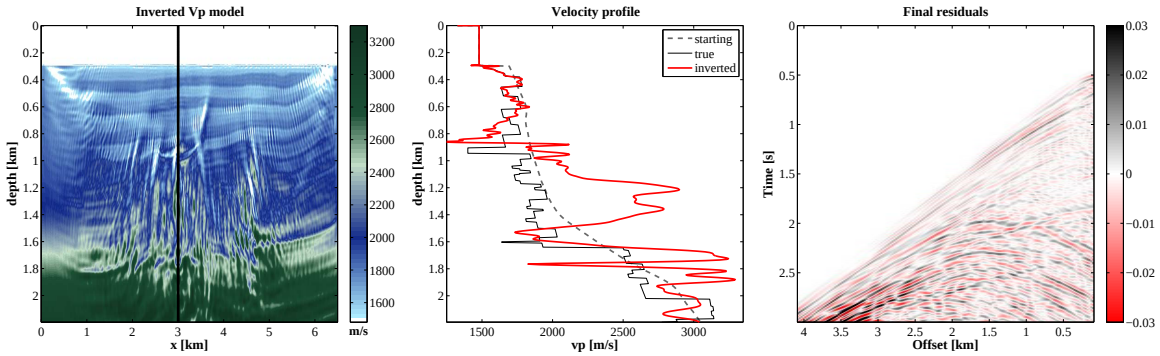
I repeated the previous multi-scale experiment, except that the number of iterations carried out for each frequency band was reduced from 20 to 10 iterations. The final FWI V_P model (Figure 3.16a) is similar to the result with 20 iterations (Figure 3.14d). However, a comparison of the V_P profiles extracted from the inverted FWI models (Figure 3.16b and Figure 3.14d) shows that with 10 iterations the reconstruction in the deep parts of the model is slightly worse.

Figure 3.16c illustrates the evolution of the data misfit functions as a function of iteration number for 10 and 20 iterations per frequency band. FWI with 20 iterations per inversion stage yields a very effective and stable reduction of the data misfit and it requires less iteration steps to reach the same data error as the FWI with 10 iterations. The most significant reduction of the data misfit is achieved after the first two, low frequency, bands, which correspond to the update of large-scale structures in the model. The convergence rate of FWI with 10 iterations is lower, especially at intermediate frequency bands, where the misfit function is much higher than that of FWI with 20 iterations. This indicates that for this particular inverse problem 10 iterations per frequency band are not enough to take a full advantage of the multi-scale inversion approach.

3.3.3.2 Choice of the first frequency band

To illustrate the effect of the starting frequency band, I performed two additional inversion test. In the first experiment, the first maximum frequency $f_{\max}(1) = 4$ Hz and there are seven frequency bands: $f_{\max} = (4, 5.3, 7.1, 9.4, 12.5, 16.6, 22)$. In the second experiment, the first maximum frequency is increased to $f_{\max}(1) = 5$ Hz with six frequency bands: $f_{\max} = (5, 6.7, 8.9, 11.9, 15.9, 22)$. There are 20 iterations carried out for each inversion step, except for the last frequency band, which covers the full frequency content of the data. The maximum iteration number is 300.

The final FWI V_P models and V_P profiles are shown in Figure 3.17a,b. The inversion starting with $f_{\max}(1) = 4$ Hz, provided a good reconstruction of the V_P model. However, there are some inaccuracies especially below the gas lens and the overall model quality is worse than for the inversion starting with a very low frequency band $f_{\max}(1) = 3$ Hz (Figure 3.14d). The further increase of the frequency content of the first band to $f_{\max}(1) = 5$ Hz results in an even worse reconstruction of velocity structures. The final V_P model is affected by many artefacts, which is evidence of the cycle-skipping problem. These results suggest that the success of the multi-scale inversion depends on the choice of the first frequency band, which should be very narrow and contain the lowest available frequencies with a high signal-to-noise ratio.



(a) (b) (c)

Figure 3.12: Marmousi2 example. FWI results for a poor starting velocity model without frequency filtering. (a) Inverted V_P model, (b) V_P profiles at $x = 3$ km: the black solid line represents the true model, the gray dashed line represents the initial model, and the red solid line represents the inverted V_P model. (c) Final data residuals for shot 50 located at $x = 6.25$ km

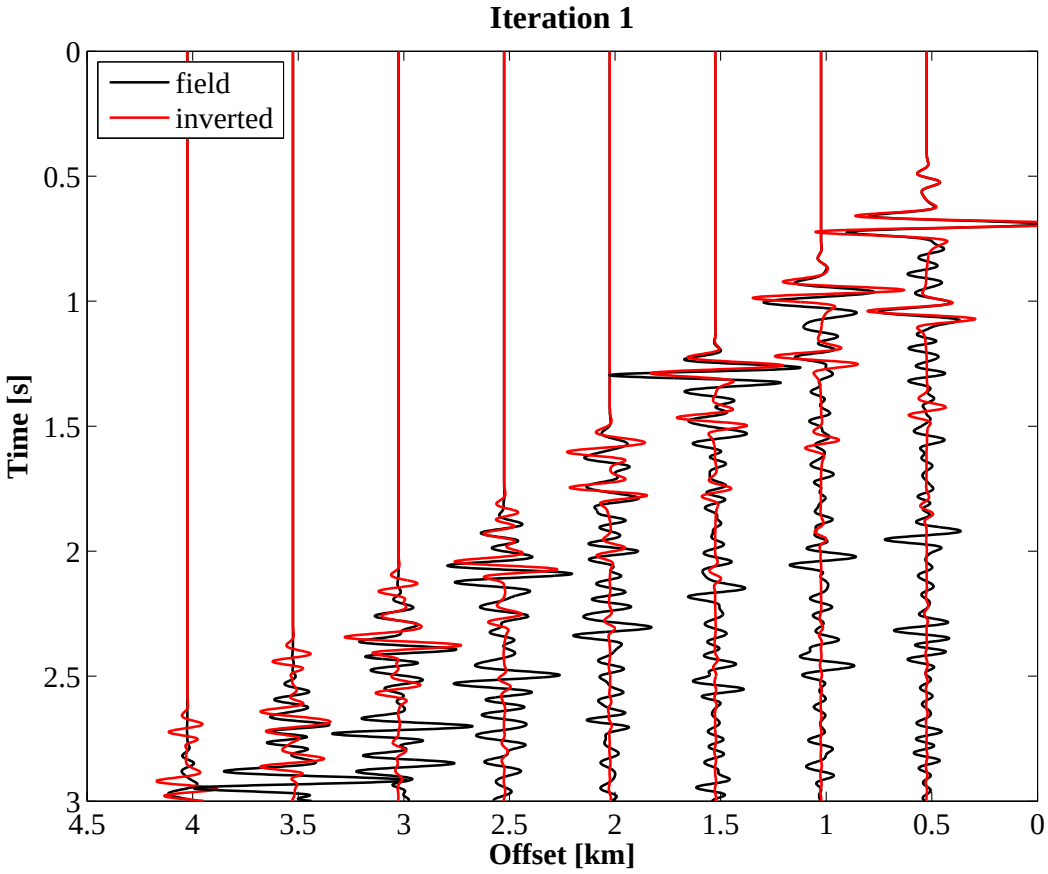


Figure 3.13: Marmousi2 example. Trace comparison between the observed data (black) and the initial data computed for the poor starting V_P model (red). Full frequency content. Shot 50 located at $x = 6.25$ km

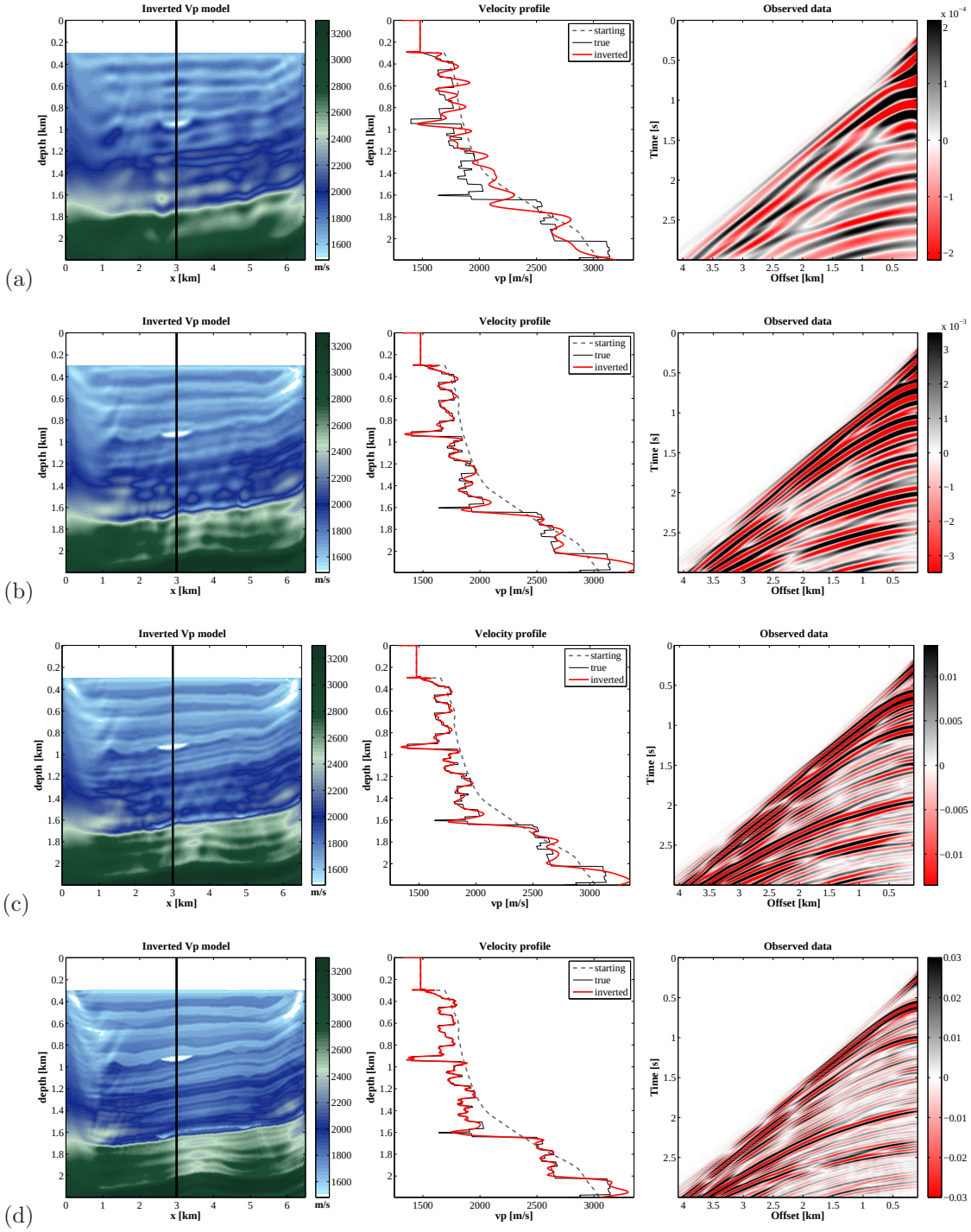
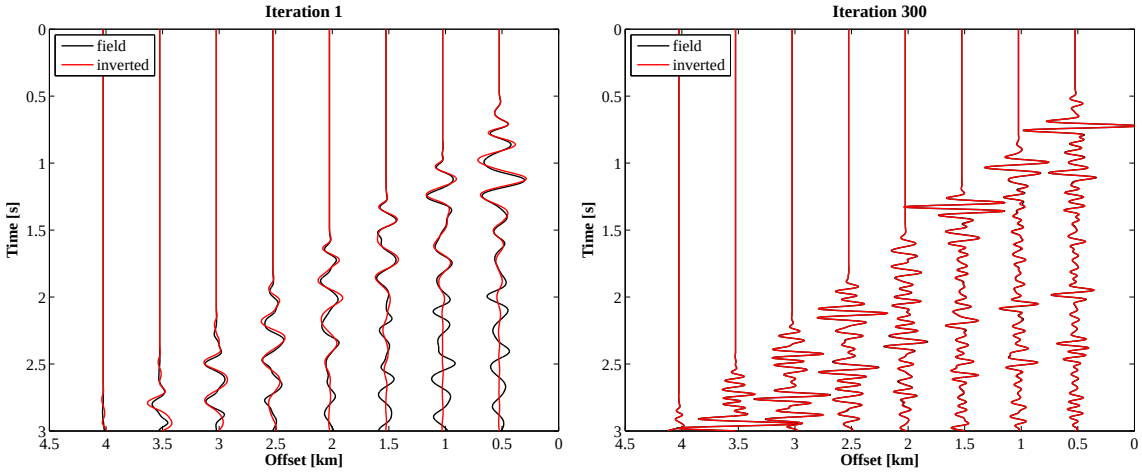
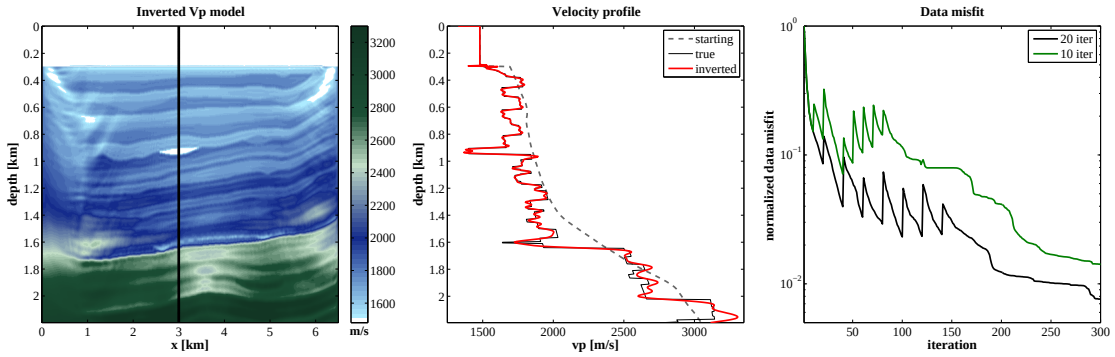


Figure 3.14: Marmousi2 example. FWI results for a poor starting velocity model with frequency filtering. 20 iteration steps per frequency band. Intermediate inversion results: (a) $f_{\max} = 3$ Hz, (b) $f_{\max} = 7.1$ Hz, (c) $f_{\max} = 12.5$ Hz, and (d) the final inversion result $f_{\max} = 22$ Hz. (left) Inverted V_P models, (center) V_P profiles at $x = 3$ km, (right) low-passed filtered original data.



(a) (b)

Figure 3.15: Marmousi2 example. Trace comparison between the observed Marmousi2 data (black) and the synthetic data (red) at the first and at the final iteration step. (a) Synthetic data are generated for the poor starting V_P model, $f_{\max} = 3$ Hz, (b) synthetic data are generated for the final FWI result (Figure 3.14d), full frequency content. Shot 50 located at $x = 6.25$ km



(a) (b) (c)

Figure 3.16: Marmousi2 example. FWI results for a poor starting velocity model with frequency filtering. 10 iteration steps per frequency band. (a) Inverted V_P model, (b) V_P profiles at $x = 3$ km: the black solid line represents the true model, the gray dashed line represents the initial model, and the red solid line represents the inverted V_P model. (c) The evolution of the data misfit function: with 20 iteration steps per frequency band (black) and with 10 iteration steps per frequency band (green). The data misfit functions are normalised with respect to their maximum value, which is identical at the first iteration.

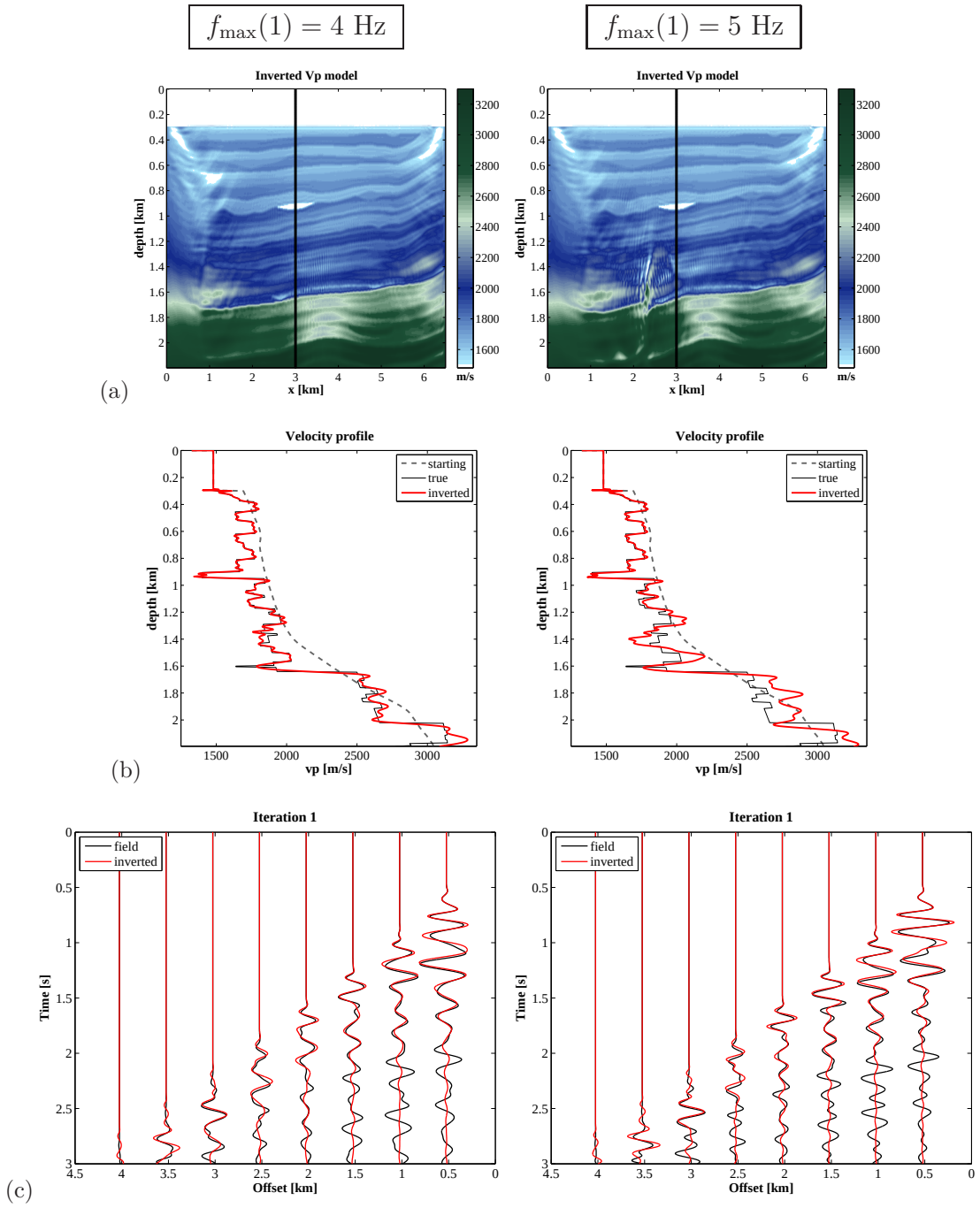


Figure 3.17: Marmousi2 example. FWI results for a poor starting velocity model with different starting frequency bands. The maximum frequency of the first band: (left) $f_{\max}(1) = 4$, (right) $f_{\max}(1) = 5$. (a) Inverted V_P models, (b) V_P profiles at $x = 3$ km, (c) trace comparison between the observed Marmousi2 data (black) and the initial data (red) computed for the poor starting V_P model.

3.4 Choice of the objective function

The aim of FWI is to find a model of subsurface that minimizes the errors between the observed data and data predicted by the forward problem for a given model of the subsurface. One possible choice is to select a model that minimizes the least-squares error between the observed data and predicted data (L_2 norm). Because the L_2 norm assumes a Gaussian distribution of the misfit, this norm is not robust in the presence of large, isolated amplitudes in the data (outliers) [Tarantola, 2005]. To overcome this problem, we can consider other minimization criteria, which are more robust in the presence of noise, including the L_1 norm [Brossier et al., 2010; Crase et al., 1990; Djikpéssé and Tarantola, 1999], the Cauchy criterion [Amundsen, 1991; Crase et al., 1990], the sech criterion [Crase et al., 1990; Monteiller et al., 2005], the Huber criterion [Guitton and Symes, 2003; Ha et al., 2009], or the objective function constructed by taking a logarithm of wavefields [Shin and Min, 2006].

To assess the performance of FWI for different minimization criteria, I invert two synthetic data sets affected by noise. In the first test, I consider acoustic inversion of elastic data. The second data set is generated by adding the realistic swell noise to the acoustic noise-free data.

3.4.1 Minimization criteria

I consider five different minimization criteria:

The least-squares criterion L_2

$$E_{L_2} = \sum_{i=1}^{N} \frac{1}{2} \left[\frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right]^2 , \tag{3.14}$$

the summation is performed over the number of source-reciver pairs and the number of time samples. σ^i is the estimated error deviation, which is specific for every shot and is calculated as the average absolute value of all samples of each shot [Pica et al., 1990]. The model minimizing E_{L_2} (Eq. 3.14) is called the best model with respect to the least-squares criterion.

The least-absolute-values criterion L_1

$$E_{L_1} = \sum_{i=1}^{N} \left| \frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right| . \tag{3.15}$$

The Cauchy criterion

$$E_{\text{Cauchy}} = \sum_{i=1}^{N} \frac{1}{2} \ln \left\{ 1 + \left[\frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right]^2 \right\} . \tag{3.16}$$

The hyperbolic secant (sech) criterion

$$E_{\text{sech}} = \sum_{i=1}^{N} \ln \left\{ \cosh \left[\frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right] \right\} . \tag{3.17}$$

The L_2, L_1, Cauchy and sech criteria are shown in Crase et al. [1990]. Finally, I consider the approximated version of the L_2 norm using the normalized wavefields proposed by Choi and Alkhalifah

[2012]

$$E_{L_2\text{norm}} = \sum_{i=1}^{N} \frac{1}{2} \left[\frac{f^i(\mathbf{m})}{\|f^j(\mathbf{m})\|} - \frac{d^i_{\text{obs}}}{\|d^j_{\text{obs}}\|} \right]^2 \tag{3.18}$$

where $\|d^j_{\text{obs}}\|$ is the root-mean-square (RMS) value calculated for every jth trace of the observed data, and $\|f^j(\mathbf{m})\|$ is the RMS value of every jth trace of modelled data, which is defined by

$$\|f^j(\mathbf{m})\| = \sqrt{\frac{1}{NT} \sum_{k=1}^{NT} f^k(\mathbf{m})^2}, \qquad j = 1, \ldots, \text{ntr} \tag{3.19}$$

The gradient of the misfit function, i.e. the derivative of the objective function with respect to the model parameters, can be computed using the chain rule

$$\frac{\partial E(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial E_{\text{criterion}}\{f(\mathbf{m})\}}{\partial \mathbf{m}} = \frac{\partial E_{\text{criterion}}\{f(\mathbf{m})\}}{\partial f(\mathbf{m})} \frac{\partial f(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{J}^T \mathbf{r} \,, \tag{3.20}$$

where $E_{\text{criterion}}$ represents the objective function for an arbitrary minimization criterion. The gradient is composed of two terms: the Jacobian matrix $\mathbf{J}$, which is independent of the choice of a minimization criterion, and the second term is the residual vector $\mathbf{r}$, which defines the source term in the calculation of backpropagated residual wavefields. The residual vector $\mathbf{r}$ is defined by

$$\mathbf{r} = \frac{\partial E_{\text{criterion}}\{f(\mathbf{m})\}}{\partial f(\mathbf{m})} \,. \tag{3.21}$$

Because the choice of the minimization criterion affects the residual vector, it also affects the computation of the residual wavefield backpropagated from the receivers positions. The ith component of the residual vectors $\mathbf{r}$ for different minimization criteria are

$$r^i_{L_2} = \frac{1}{\sigma^i} \left[\frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right] \tag{3.22}$$

$$r^i_{L_1} = \frac{1}{\sigma^i} \text{sgn} \left[\frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right] \tag{3.23}$$

where sgn is the signum function defined as follows

$$\text{sgn}(x) = \begin{cases} -1, & \text{if } x < 0 \,, \\ 0, & \text{if } x = 0 \,, \\ 1, & \text{if } x > 0 \,. \end{cases} \tag{3.24}$$

$$r^i_{\text{Cauchy}} = \frac{1}{\sigma^i} \left\{ \frac{\dfrac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i}}{1 + \left[\dfrac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right]^2} \right\} \tag{3.25}$$

$$r^i_{\text{sech}} = \frac{1}{\sigma^i} \tanh \left[\frac{f^i(\mathbf{m}) - d^i_{\text{obs}}}{\sigma^i} \right] \tag{3.26}$$

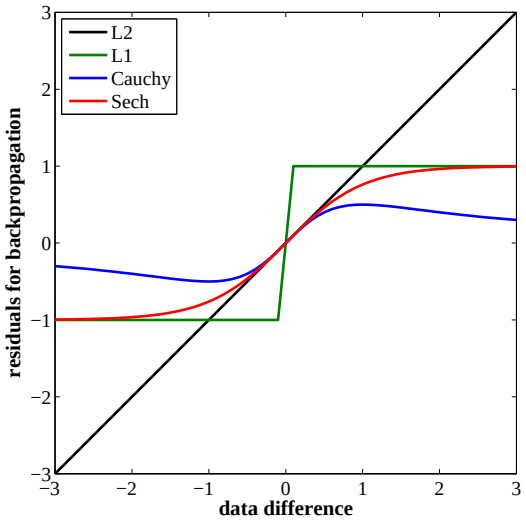


Figure 3.18: The residuals of L_2 norm (black), L_1 norm (green), Cauchy norm (blue), and sech norm (red) as a function of the data difference $f(\mathbf{m}) - \mathbf{d}_{\text{obs}}$.

$$r^i_{L_2\text{norm}} = \frac{f^i(\mathbf{m})}{\|f^j(\mathbf{m})\|} - \frac{d^i_{\text{obs}}}{\|d^j_{\text{obs}}\|} \tag{3.27}$$

Figure 3.18 illustrates the residuals of different minimization criteria. The horizontal axis represents the theoretical data difference $f(\mathbf{m}) - \mathbf{d}_{\text{obs}}$, and the vertical axis shows the corresponding residuals computed for different minimization criteria with Eqs. 3.22- 3.26, where $\sigma_i = 1$. Whereas, the L_2 norm preserves the difference between the predicted and observed data, the L_1 norm ignores the true amplitudes of the data difference. For that reason, the L_1 norm is more robust in the presence of large amplitude noise in the data than the L_2 norm. The Cauchy and sech criteria can be considered as hybrid criteria. The sech criterion behaves as the L_2 norm for small data difference, but for large data errors it acts as the L_1 norm. The transition between the L_2-norm-behaviour and the L_1-norm-behaviour is controlled by the choice of a threshold value σ. In case of the Cauchy criterion, the errors which are small compared to the estimated error deviation σ are treated like the L_2 norm, but the large data errors are weighted towards zero. These properties of the Cauchy and sech norm indicate that both minimization criteria are less sensitive to data outliers than the L_2 norm. The approximated version of L_2 norm using the normalized wavefields $L_{2\text{norm}}$ is not shown on this theoretical plot, because it acts as the conventional L_2 norm, when the RMS value of the modelled and observed traces is identical. The advantage of the $L_{2\text{norm}}$ minimization criterion is that it can account for the difference in the energy level between the predicted and observed wavefields [Choi and Alkhalifah, 2012].

The impact of the minimization criteria on the residual trace is shown in Figure 3.19. The residual trace represents the source term, which is used to calculate the backpropagated residual wavefields. Note the different amplitudes of the residual traces. The maximum value occurs when the L_2 norm is used. The residual traces of the L_1 norm and the sech norm have the same maximum amplitudes, because the sech norm acts as the L_1 norm for data errors higher than the estimated error deviation σ. In case of the $L_{2\text{norm}}$ criterion, the modelled and observed data are normalized to their RMS values before computing the residuals for every trace.

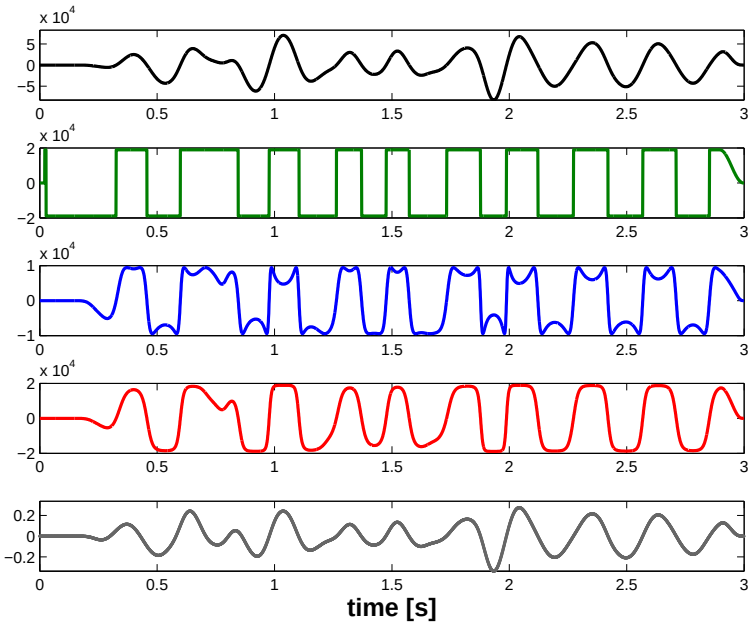


Figure 3.19: Impact of the minimization criteria on the residual trace. From top to bottom: L_2 norm, L_1 norm, Cauchy, sech, L_2 norm using the normalized wavefields. The figure represents the same residual trace at near offset, computed for different minimization criteria with Eqs. 3.22- 3.27

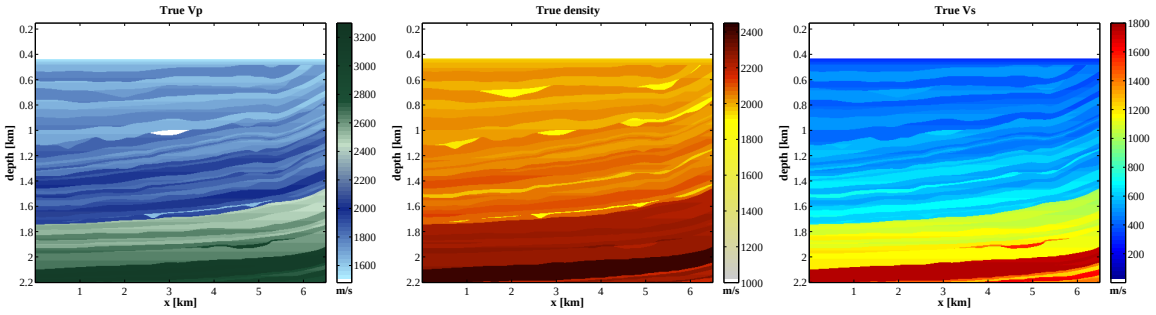
3.4.2 Acoustic inversion of elastic data

In the first numerical experiment, I perform acoustic FWI of elastic data using different minimization criteria. The aim of this test is not only to compare the performance of various objective functions, but also to assess the limits of the acoustic approximation when inverting elastic data.

Because marine streamer data are usually dominated by unconverted P-waves [Djikpéssé and Tarantola, 1999], it is a common practice to apply the acousic FWI to seismic data from marine exploration. Even though the elastic FWI would provide a better match to the recorded seismic data, the main advantage of the acoustic approach is related to a significant reduction of the computational costs. However, there are some issues associated with the acoustic approximation of the wave equation. Acoustic modelling will not predict elastic effects such as non-acoustic reflection amplitudes or mode-converted waves. As shown by Barner and Charara [2009]; Mulder and Plessix [2008] the successful application of acoustic FWI is very difficult when the long offset data are included or when the strong discontinuities are present in the S-wave velocity. Strong impedance contrasts for S-waves generate mode-converted waves, which serve as a strong coherent noise for the acoustic FWI. But even if the converted (e.g. PSP, PSSP) waves are weak or not present in the recorded data, the main problem of the acoustic approximation is an incorrect modelling of the amplitude-versus-offset (AVO) effects.

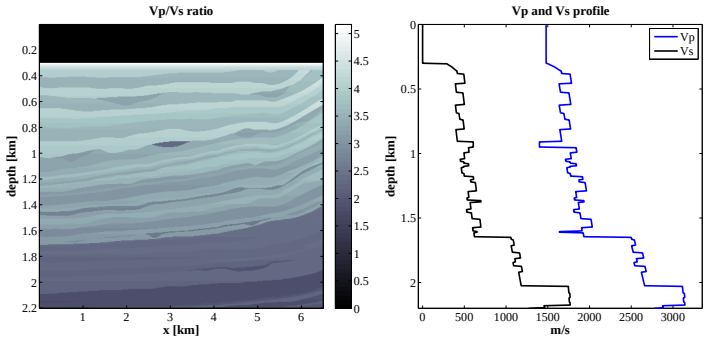
3.4.2.1 Elastic data vs acoustic data

The elastic data set is generated for the modified elastic Marmousi2 model shown in Figure 3.20 using a 2D elastic finite-difference modelling code developed by [Bohlen, 2002]. To generate elastic data, I used the same acquisition geometry, the source wavelet and the modelling parameters as for the



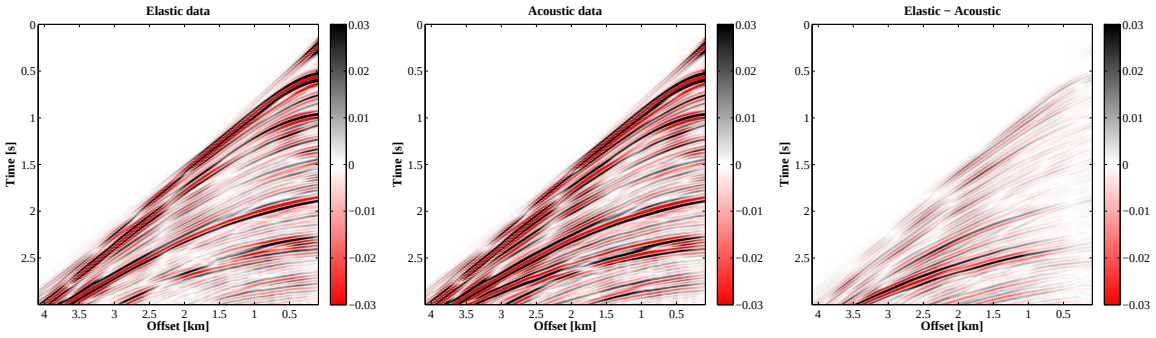
(a) (b) (c)

Figure 3.20: Modified elastic Marmousi2 model. True (a) V_P, (b) density, and (c) V_S models used to generate the elastic data set.



(a) (b)

Figure 3.21: Modified elastic Marmousi2 model. (a) V_P/V_S ratio. (b) V_P and V_S profiles extracted from the true V_P and V_S models at $x = 3$ km.



(a) (b) (c)

Figure 3.22: (a) Elastic data, (b) acoustic data, (c) the difference between elastic and acoustic data. Shot 50 located at $x = 6.25$ km.

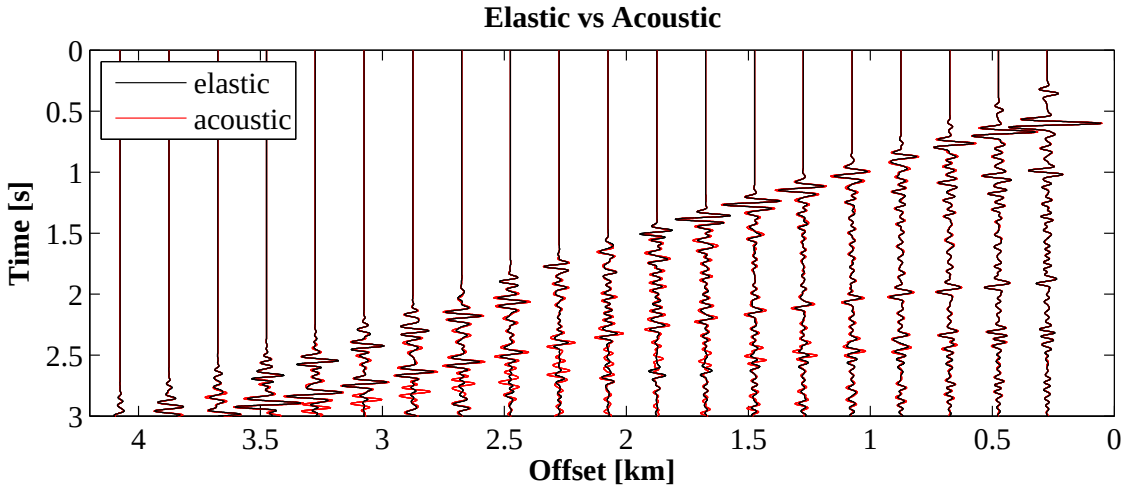


Figure 3.23: Elastic data vs acoustic data. Comparison of exemplary traces for shot 50 located at $x = 6.25$ km.

acoustic data. The distribution of the V_P to V_S ratio in the model, as well as the V_P and V_S profiles are shown in Figure 3.21. The S-wave velocity contrasts are weak in sedimentary layers up to a deph of approximately 1.6 km, and the V_P to V_S ratio varies between 5 and 3. In contrast to the P-wave velocity, there is no strong discontinuity in the V_S at the gas lens (at a depth of 900 m). However, there is a strong velocity contrast both in P-wave and in S-wave at a depth of 1.6 km. The V_P/V_S ratio is decreasing to approximately $2.3 - 1.77$, which is related to the transition from soft rock to hard rock.

The modelled acoustic and elastic data are illustrated in Figure 3.22, both seismograms are normalized to the maximum of the direct wave at the near offset trace. The difference between elastic and acoustic data is very small at near offsets, but it is increasing for middle and far offsets. The main discrepancy is related to the wave reflected from the layer with a strong impedance contrast in P- and S-wave at a depth of 1.6 km. Amplitude discrepancies resulting from the incorrect modelling of the AVO effects are well visible in Figure 3.23. There is a significant loss of amplitude in elastic data for late arrivals at middle and far offsets.

To reduce the high complexity of the inverse problem, the multi-scale inversion approach is applied. The frequency bands are applied sequentially with the following maximum frequencies $f_{\max} = (3, 4, 5.3, 7.1, 9.4, 12.5, 16.6, 20)$ Hz; that define the cutoff frequency for the Butterworth low-pass filter. Furthermore, to correct for the amplitude loss with depth due to geometrical spreading and to enhance deeper parts of the model, the preconditioning operator P_2 is applied. The starting V_P model is a 1D very smooth representation of the true velocity distribution. To focus only on the V_P model reconstruction, the density model and the source time function are assumed to be known.

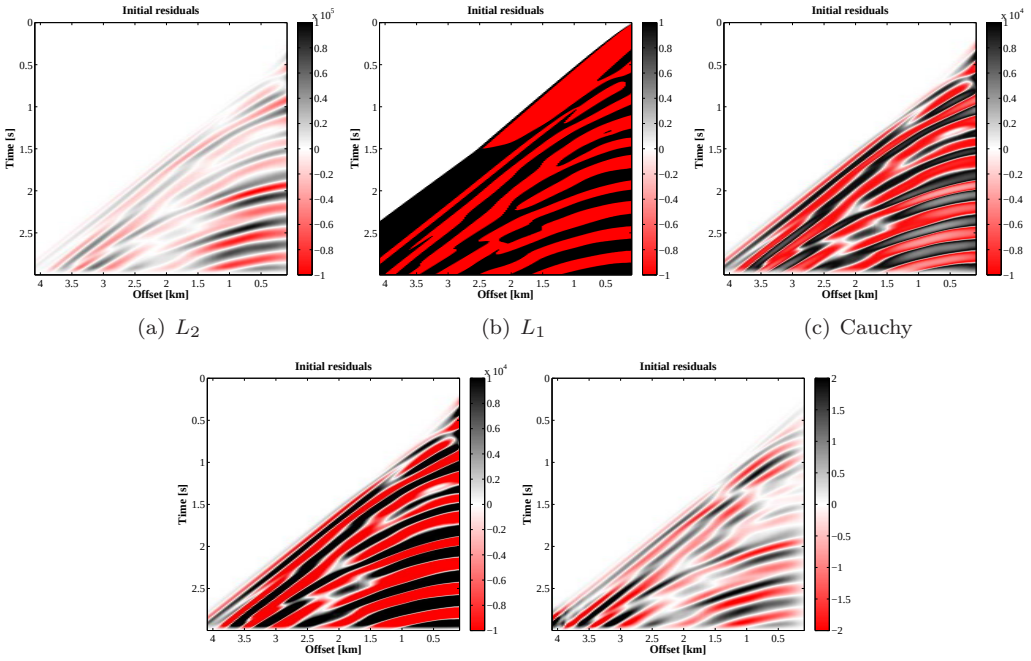
3.4.2.2 Inversion results

I inverted the elastic data with the acoustic FWI using five different minimization criteria. The aim of this experiment is to evaluate the effect of the AVO errors resulting from the acoustic approximation on the inversion results and to compare the performance of various objective functions.

Figure 3.24 compares the residuals for backpropagation at the first iteration step for different minimization criteria. We can observe that the residuals are dominated by strong, near offset reflections when the L_2 norm is used. The residuals for the Cauchy and sech criteria are very similar and they do not possess the blocky nature of the L_1 norm. The L_2 norm using the normalized wavefields has a horizontal trace balancing effect, i.e. the residuals are not dominated by near offset traces as for the conventional L_2 norm. But unlike the Cauchy, sech, or L_1 criteria, the $L_{2\text{norm}}$ norm has no vertical (i.e. time) balancing effect on the residuals.

The final inversion results are shown in Figure 3.25. The sedimentary structures in the upper part of the model are well reconstructed for all minimization criteria. However, there is a lot of artefacts in the deeper part of the model, especially above the interface with high contrast in the S-wave velocity. The artefacts are weaker for the $L_{2\text{norm}}$ criterion, but in general the acoustic inversion failed to reconstruct the deep structures of the true model due to significant elastic effects present in the data. The comparison of V_P profiles Figure 3.26 shows the poor quality of inverted V_P in the deep part of the model. Furthermore, there are strong artefacts at the seafloor, because the inversion tries to compensate the wrong AVO effects in the acoustic modelling. These artefacts are relatively weak for the $L_{2\text{norm}}$ criterion. In addition, the accuracy of the recovered V_P is better for the $L_{2\text{norm}}$ norm than for other minimization criteria, which is well visible at a depth from 1.4 km to 1.6 km. The final data difference for different minimization criteria is shown in Figure 3.27. This is the actual difference between the predicted data and observed data, not the data residuals for backpropagation. The final residuals are relatively large especially for the L_1 norm and the Cauchy norm.

Figure 3.28a illustrates the evolution of the model error as a function of iteration number for different minimization criteria. The choice of the $L_{2\text{norm}}$ criterion results in the lowest model error. The performance of the L_2, Cauchy and sech criteria is comparable, but worse than the performance of the L_1 norm, which ignores the amplitude information of the data misfit. The evolution of the data misfit is shown in Figure 3.28b. Because the data misfit is calculated for different minimization criteria, the plots are not directly comparable.



(a) L_2 (b) L_1 (c) Cauchy

(d) sech (e) L_2norm

Figure 3.24: Acoustic inversion of elastic data. Residuals for backpropagation at the first iteration step for different minimization criteria.

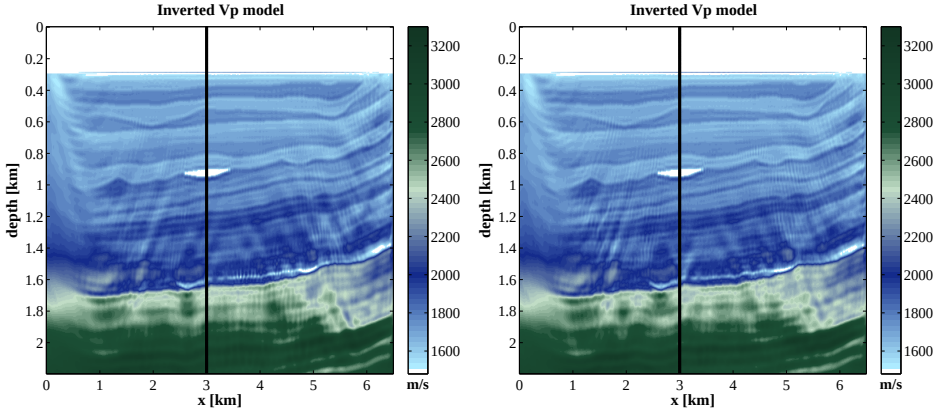
3.4.3 Swell noise data

In the second experiment, I introduced swell noise to synthetic acoustic noise-free data. The swell noise was extracted from the real data presented in Chapter 7. Swell noise originates from the sea-surface waves and vibrations in the streamer caused by turbulent water related to these surface waves. On streamer operating during bad weather conditions, the swell noise is so strong that the useful signal cannot be identified.

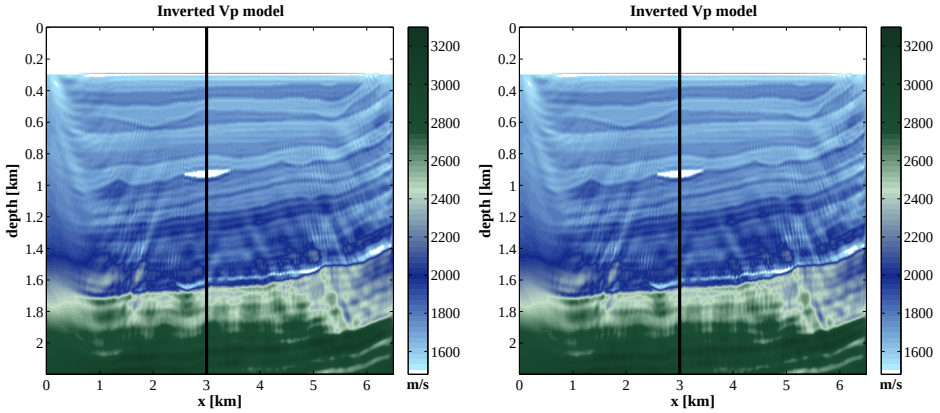
Figure 3.29 shows the acoustic noise-free data, the swell noise, and the noisy-data for three exemplary shot gathers. Swell noise has large amplitudes at relatively low frequencies (up to 5 Hz) and it affects a number of neighboring traces. It can be observed as vertical stripes in seismic data. The maximum amplitude of the swell noise is at 2 Hz and it is 2.3 times higher than the amplitude of the useful signal at 2 Hz.

The intermediate FWI results after inverting the third frequency band with $f_{\max} = 5.3$ Hz are shown in Figure 3.30 for different minimization criteria. The high-amplitude swell noise produced strong artefacts in the V_P model when the L_2 norm was used. This is the known behaviour of the L_2 criterion, which is very sensitive to non-Gaussian errors [Brossier et al., 2010; Crase et al., 1990]. The L_1, Cauchy and sech criteria are more robust in the presence of swell noise in the data and provided a good reconstruction of the V_P models. The poor performance of the $L_{2\text{norm}}$ criterion was caused by difficulties in estimating a reliable step length using the parabolic curve fitting method. The negative minima estimated in the initial iteration steps resulted in the choice of the very small step lenghts and yielded a weak update of the V_P model.

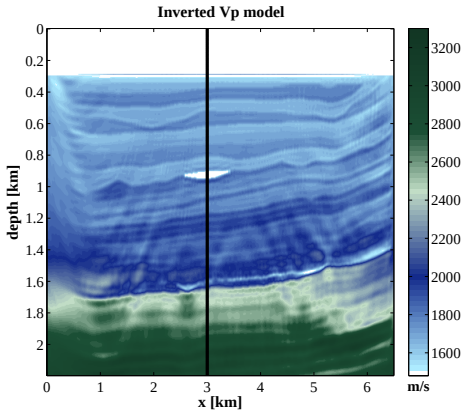
The final FWI results for $f_{\max} = 20$ Hz are illustrated in Figure 3.31. The effect of the swell noise is



(a) L_2 (b) L_1

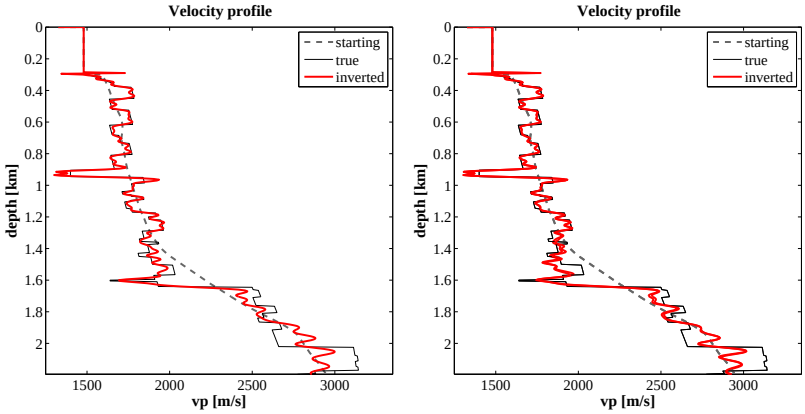


(c) Cauchy (d) sech



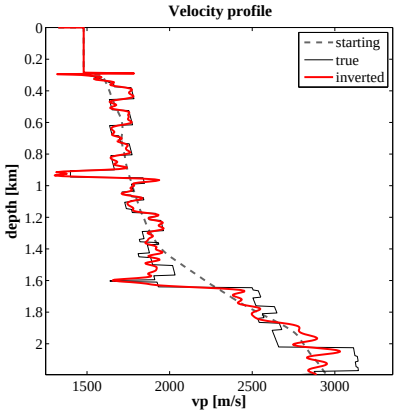
(e) L_2norm

Figure 3.25: Acoustic inversion of elastic data. FWI results. (a)-(e) Comparison of inverted V_P models for different minimization criteria.

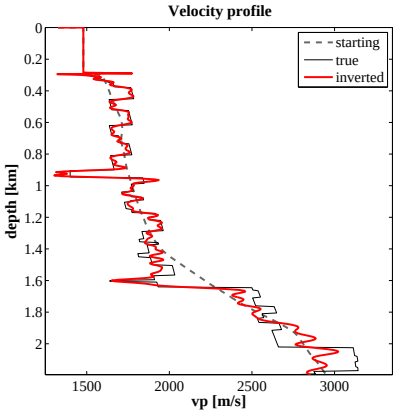


(a) L_2

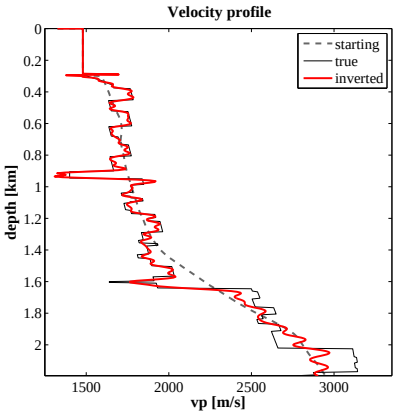
(b) L_1



(c) Cauchy

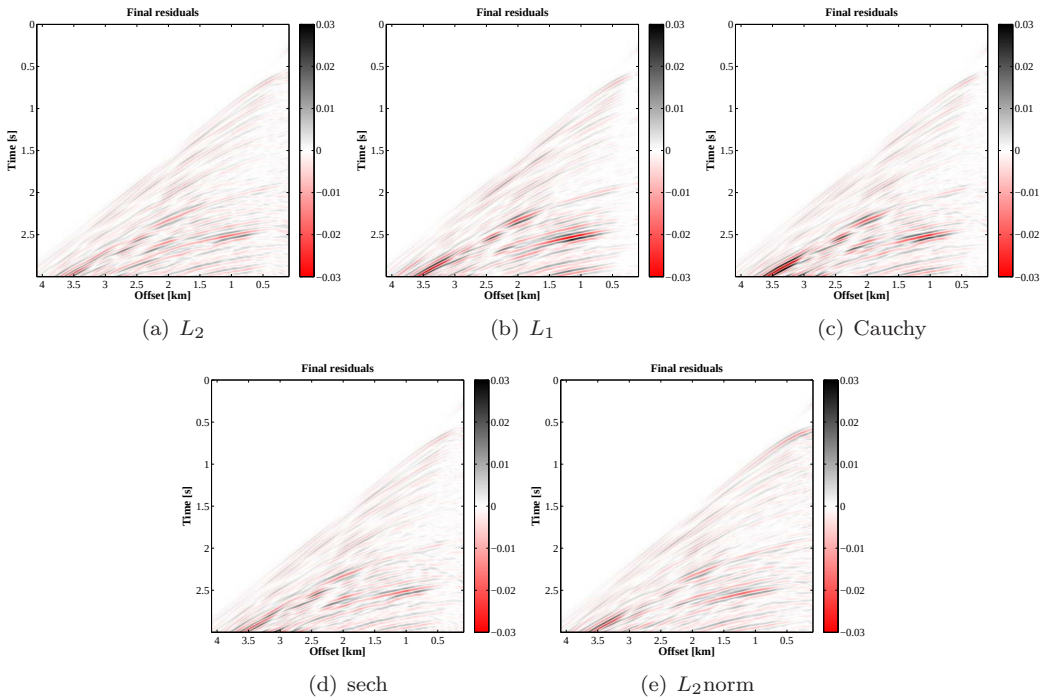


(d) sech



(e) L_2norm

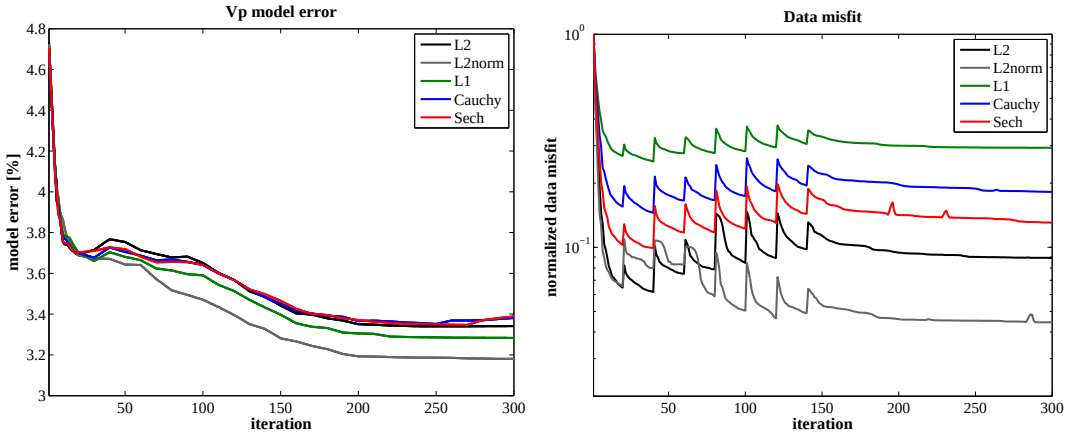
Figure 3.26: Acoustic inversion of elastic data. Comparison of V_P profiles at $x = 3$ km for different minimization criteria; the black solid line indicates the true model, the gray dashed line indicates the initial model, and the red solid line indicates the inverted V_P model.



(a) L_2 (b) L_1 (c) Cauchy

(d) sech (e) L_2norm

Figure 3.27: Acoustic inversion of elastic data. Final data residuals computed as the difference between the predicted data for the final FWI result and the observed data.



(a) (b)

Figure 3.28: Acoustic inversion of elastic data. Summary of FWI results. (a) V_P model error of the inverted models (Eq. 3.2) for different minimization criteria. (b) The evolution of the data misfit functions, which are calculated with Eqs. 3.14- 3.18. The data misfit functions are normalized with respect to their maximum value, which are not identical.

no longer visible in the reconstructed P-wave velocity for the L_2 norm, except for some weak artefacts at the sides of the model. Because the swell noise was dominant only at very low frequencies and the higher frequencies were not affected by noise, the application of the multi-scale inversion strategy reduced the noise effect on inversion results. The accuracy of reconstructed V_P models is very good and comparable for all minimization criteria (Figure 3.32). The final data residuals are dominated by the swell noise (Figure 3.33).

The evolution of the V_P model error shows the best performance of the L_1, Cauchy and sech criteria when inverting the swell noise data (Figure 3.34). The L_2 norm results in a much higher error, especially for the inversion of low frequency data dominated by swell noise. The $L_{2\text{norm}}$ criterion, which provided the best results for the acoustic inversion of elastic data, had significant problems to provide a stable reduction of the model error.

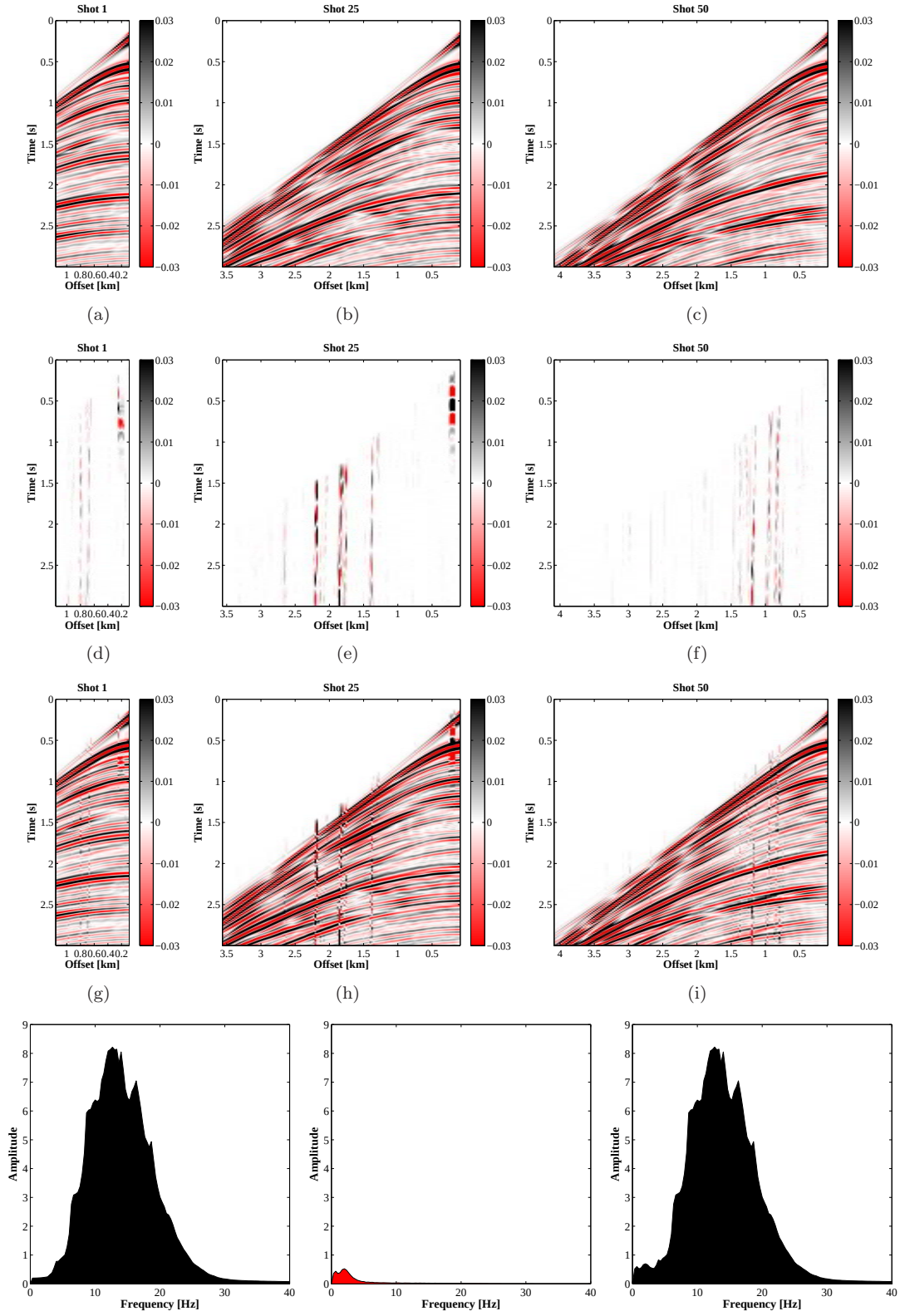
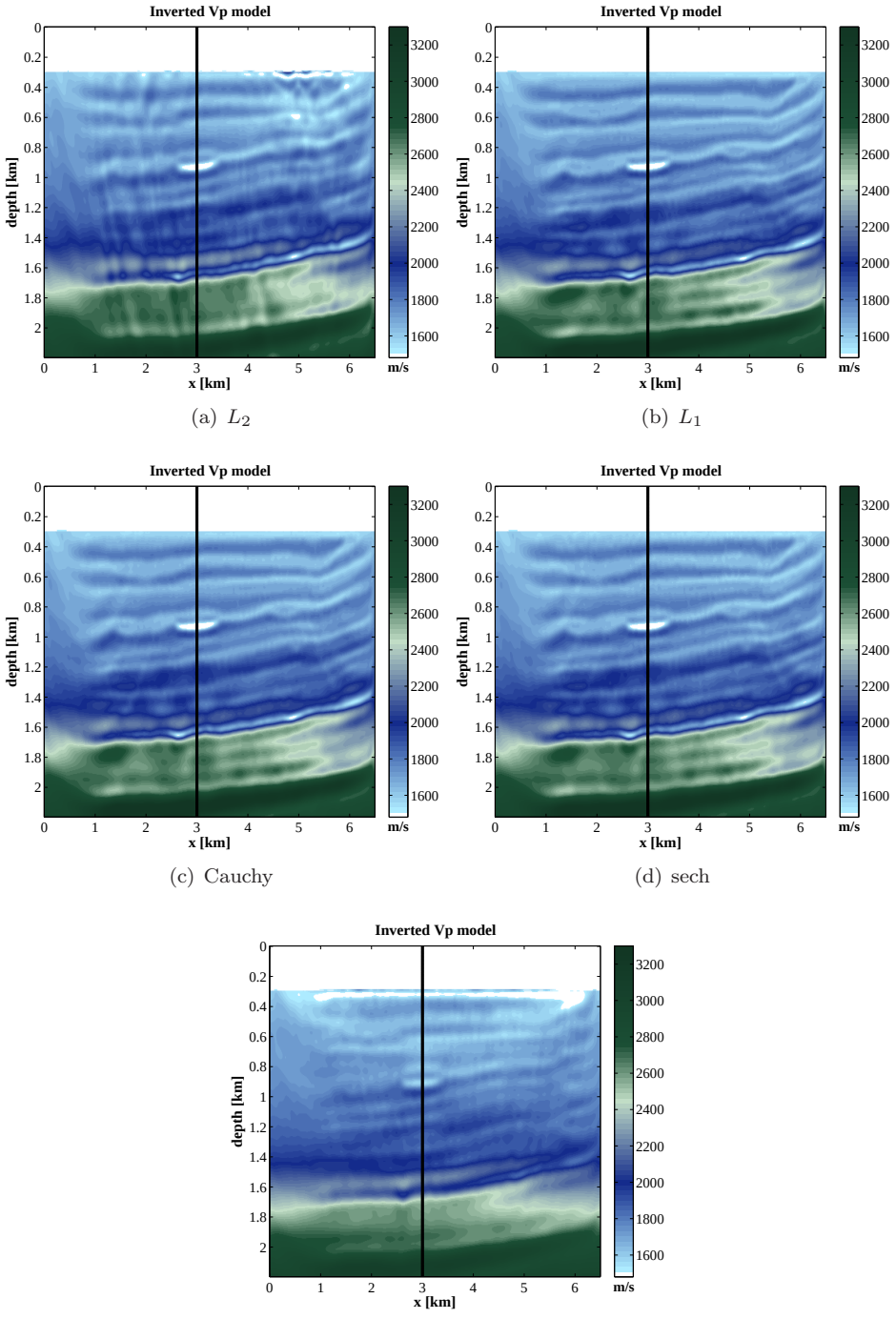


Figure 3.29: Swell noise data. Exemplary noise-free shot gathers: (a) Shot 1 located at $x = 1.35$ km, (b) Shot 25 located at $x = 3.75$ km, (c) Shot 50 located at $x = 6.25$ km. (d)-(f) swell noise, (g)-(i) noisy data. Amplitude spectra of the (j) noise-free data, (k) swell noise, (l) noisy data for shot 50.

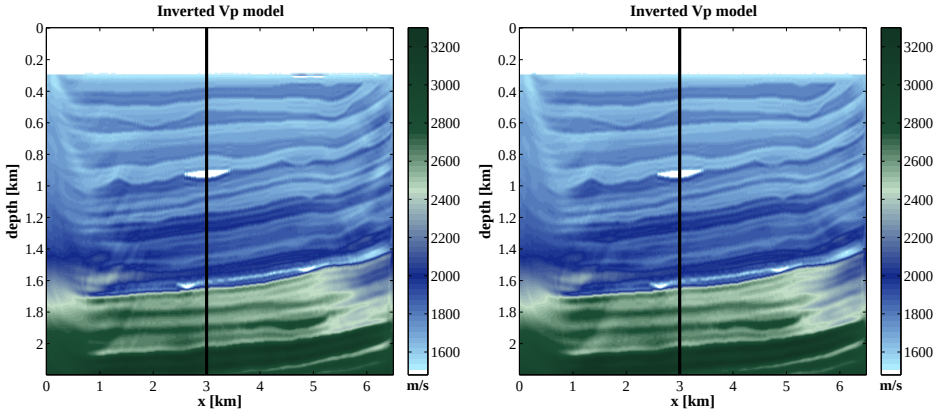


(a) L_2 (b) L_1

(c) Cauchy (d) sech

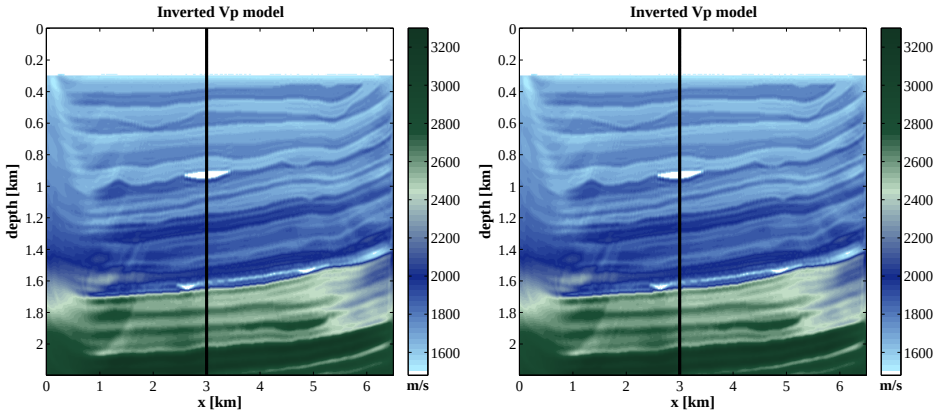
(e) L_2norm

Figure 3.30: Inversion of swell noise data. Intermediate FWI results after the third frequency band with $f_{\max} = 5.3$ Hz for different minimization criteria.



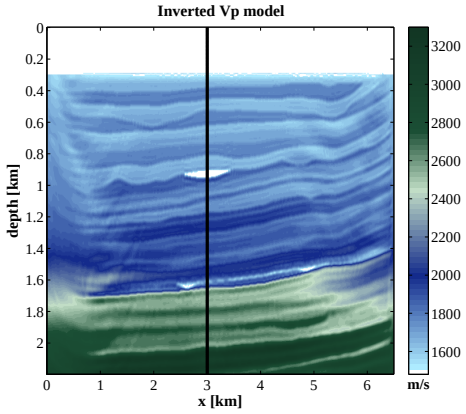
(a) L_2

(b) L_1



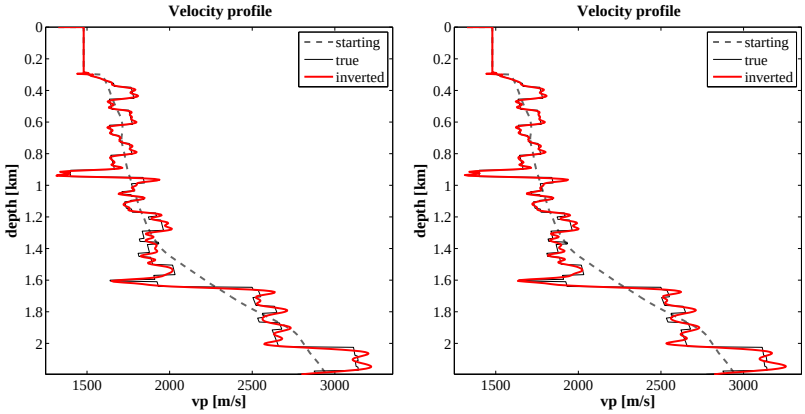
(c) Cauchy

(d) sech



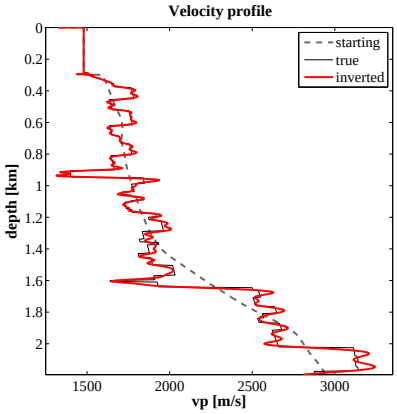
(e) L_2norm

Figure 3.31: Inversion of swell noise data. Comparison of final inverted V_P models for different minimization criteria.

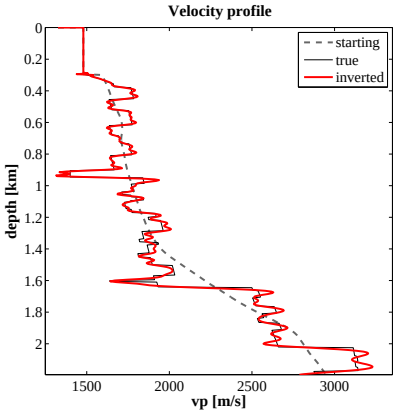


(a) L_2

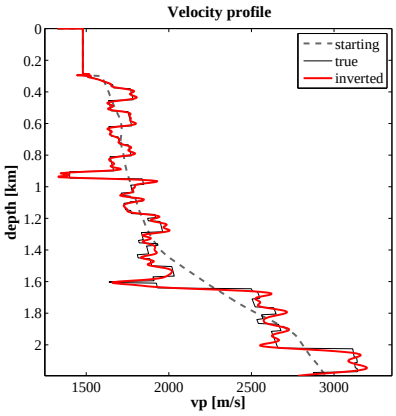
(b) L_1



(c) Cauchy

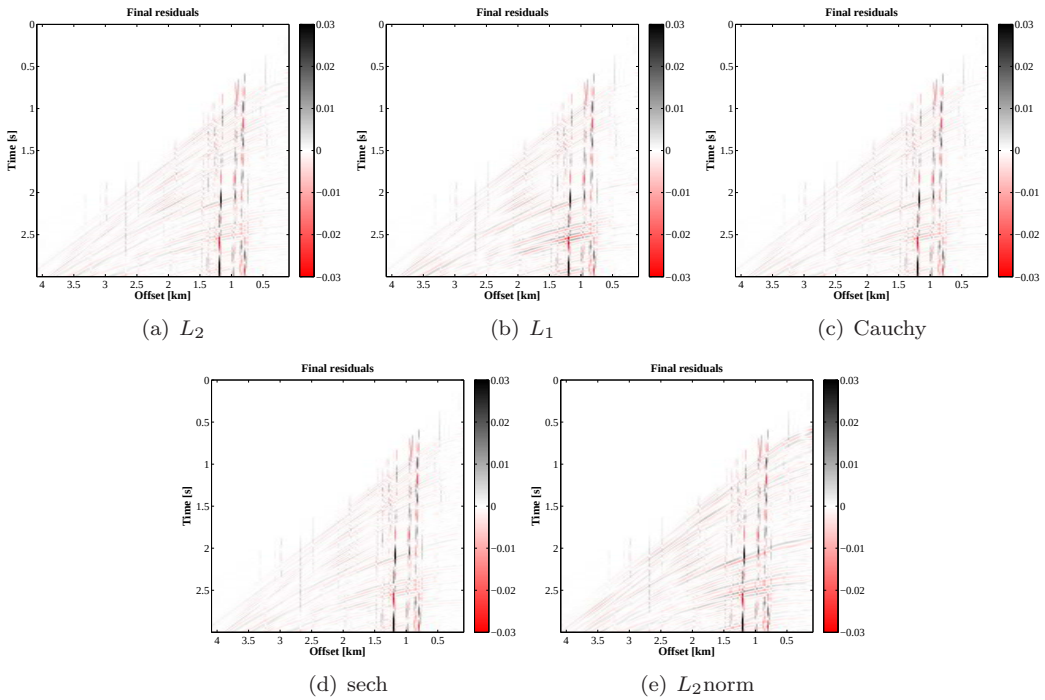


(d) sech



(e) L_2norm

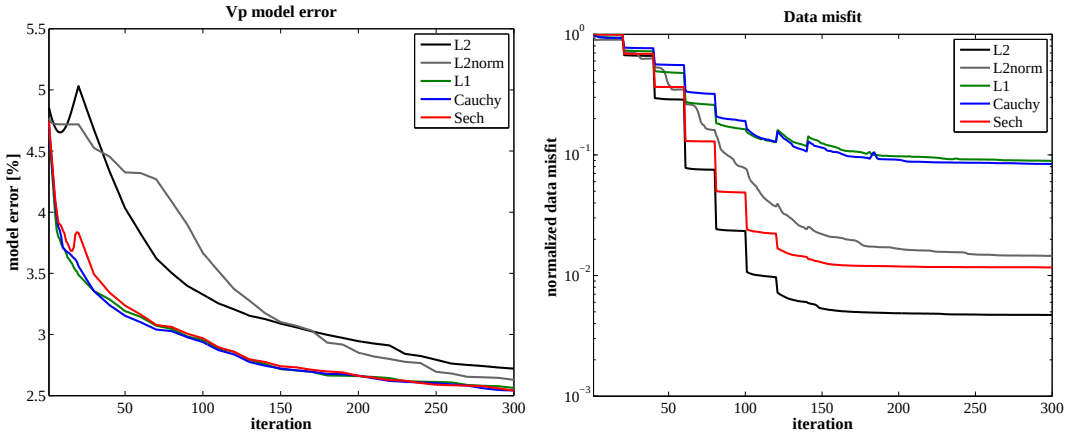
Figure 3.32: Inversion of swell noise data. Comparison of V_P profiles at $x = 3$ km for different minimization criteria; the black solid line indicates the true model, the gray dashed line indicates the initial model, and the red solid line indicates the final inverted V_P model.



(a) L_2 (b) L_1 (c) Cauchy

(d) sech (e) L_2norm

Figure 3.33: Inversion of swell noise data. Final data residuals computed as the difference between the predicted data for the final FWI result and the observed data.



(a) (b)

Figure 3.34: Inversion of swell noise data. Summary of FWI results. (a) V_P model error of the inverted models (Eq. 3.2) for different minimization criteria. (b) The evolution of the data misfit functions, which are calculated with Eqs. 3.14- 3.18. The data misfit functions are normalized with respect to their maximum value, which are not identical.

3.5 Summary

- **Gradient preconditioning**

  The comparison of various preconditioning methods shows that the satisfactory performance of the inversion algorithm can be achieved by applying a preconditioning operator that combines a spatial taper with a linear scaling function. A spatial taper, which turns off the model update in the water layer, allows to suppress the artefacts related to the acquisition geometry and to improve the quality of inverted models. In addition, the effect of geometrical spreading of the waveform amplitudes can be effectively compensated by applying a simple depth-dependent linear preconditioning operator to the gradient of the objective function.

- **Multi-scale inversion**

  The success of the FWI depends not only on the choice of the starting model, but also on methods which may reduce the complexity of the inverse problem. The multi-scale inversion is a very effective method to reduce the non-linearity of the inverse problem and to mitigate the problem of local minima in the objective function. The selection of optimal frequency bands is an important aspect of the multi-scale inversion, because it affects the computational cost of FWI and it is related to the problem of cycle-skipping. The implemented multi-scale approach is based on the continuous coverage of vertical wavenumbers. The success of the multi-scale inversion depends on the choice of the first frequency band, which should be relatively narrow and contain the lowest available frequencies with a high signal-to-noise ratio.

- **Choice of the objective function**

  The elastic effects limits the applicability of the acoustic FWI especially in the presence of strong contrasts in the S-wave velocity. The acoustic inversion failed to reconstruct the deep structures of the true V_P model due to significant elastic effects present in the data. In general, the quality of reconstructed V_P models from acoustic inversion of elastic data is comparable for different minimization criteria, with the lowest model error obtained for the approximated version of the L_2 norm using the normalized wavefields.

  The numerical example with the high-amplitude swell noise added to the data showed the robust behaviour of the L_1, Cauchy and sech criteria. On the other hand, the L_2 norm and the $L_{2\text{norm}}$ criterion, which are more sensitive to noise present in the data, produced worse inversion results at low frequencies dominated by the swell noise. Because the higher frequencies were not affected by noise, the final FWI results for the maximum frequency of 20 Hz are comparable for all minimization criteria. Apparently, the multi-scale inversion strategy reduced the effect of the low-frequency noise on the final inversion results.

Chapter 4

Parameterization study

The aim of this study is to find the most suitable parameterization for the multi-parameter time-domain acoustic inversion of marine reflection seismic data. I investigate three different combinations of parameters: P-wave velocity and density, acoustic impedance and density, P-wave velocity and acoustic impedance. Since coupling between parameter pairs is azimuth-dependent, I consider separately the inversion of the near-offset (0.1 km - 1.15 km) and the full-offset data (0.1 km - 4 km). The acoustic Marmousi2 model with a conventional streamer geometry, and a frequency range from 2.5 to 20 Hz is used. A set of inversion tests is performed to assess the different parameterizations in terms of the quality of the reconstructed images and the convergence rate of the inversion.

4.1 Introduction

The aim of the full waveform inversion (FWI) is to estimate the physical properties of the Earth by minimizing the misfit between observed and predicted seismic data. To reconstruct reliable models of the subsurface structures from field measurements, the waveform inversion should correctly account for the most significant wave propagation phenomena present in the data. However, numerous approximations are usually made to limit the number of physical parameters and to reduce the computational cost of the method. For instance, it is common practice to use the acoustic approximation when inverting marine seismic data. In most marine field data applications the authors only invert for the P-wave velocity [Boonyasiriwat et al., 2010; Delescluse et al., 2011; Kelly et al., 2010; Operto et al., 2004; Shipp and Singh, 2002]. This mono-parameter acoustic waveform inversion is not correctly describing the amplitudes of field data. In order to improve the accuracy of seismic amplitude modeling, an extension from the mono-parameter inversion to the multi-parameter inversion is a straightforward solution. However, the joint reconstruction of more parameters is more expensive and increases the ill-posedness of the inverse problem [Virieux and Operto, 2009].

In the context of the multi-parameter inversion, an important factor is the choice of the parameters describing the medium [Tarantola, 1986]. The acoustic medium can be described by P-wave velocity V_P and density ρ, or the acoustic impedance I_P. Thus, the possible parameterization sets are: P-wave velocity and density, acoustic impedance and density, P-wave velocity and acoustic impedance. The choice of the acoustic parameters may influence the convergence rate and the ambiguity of the inverse

problem, and affects the final results [Kolb and Canadas, 1986].

The choice of an adequate parameterization has been mainly investigated for an elastic medium [Assous and Collino, 1990; Debski and Tarantola, 1995; Köhn et al., 2012; Tarantola, 1986]. Tarantola [1986] has shown that the model parameters should be related to the parameters of the Earth that can be resolved by a certain acquisition. The typical seismic reflection data contain two different kinds of information: on the long-wavelength seismic wave velocities, and on the short-wavelength impedances [Jannane et al., 1989]. For the long wavelengths ($\lambda \geq 300$ m), the misfit function mainly depends on traveltimes of the main reflections, and thus the velocity of the medium. Whereas, for the short wavelengths ($\lambda \leq 60$ m), the data misfit is mainly sensitive to impedance constrasts (the reflection amplitudes). So, the arrival time of the waves provides the information on the velocity distribution and the amplitude information allows to resolve the high-resolution impedance contrasts. Tarantola [1986] suggests that for long wavelenths the P-wave and S-wave velocities are the most adequate parameters and for short wavelengths the seismic impedances and density are more suitable.

Another important aspect is the coupling between different parameters. It would be favorable to select parameters that are as uncorrelated as possible. The trade-off is often investigated by considering the energy radiation patterns of a point diffractor [Assous and Collino, 1990; Tarantola, 1986; Virieux and Operto, 2009]. In this case a homogeneous Earth is assumed and each parameter is individually perturbed. Acoustic radiation patterns for different parameter sets are studied in Virieux and Operto [2009]. An important observation is that the coupling is a function of the offset. For example, using P-waves only and small offsets it is difficult to distinguish between a density diffractor and a P-wave velocity diffractor. On the other hand, the V_P and I_P point diffractors scatter energy for different apertures, V_P for wide apertures and I_P for short apertures, which might suggest that this is the most adequate parameter set. For instance, this parameterization has been chosen to define the medium in the acoustic multi-parameter inversion of sesimic reflection data performed by Kolb and Canadas [1986].

In this study, I test three different acoustic parameterizations, which I implemented in the waveform inversion code. The first parameter set is the P-wave velocity and density $\mathbf{m_1} = [V_P, \rho]$, the second is the P-wave impedance and density $\mathbf{m_2} = [I_P, \rho]$, and the third parameter set is the P-wave velocity and P-wave impedance $\mathbf{m_3} = [V_P, I_P]$. First, I investigate the coupling between the different parameters, when the simultaneous multi-parameter inversion is performed. If there is a strong trade-off between two parameters, this may result in an incorrect solution and may lead to a wrong interpretation of the inversion results. Furthermore, the coupling effects for the same parameter set may vary depending on the maximum offset of the seismic data used in the inversion. For that reason, I investigate separately the multi-parameter inversion of near-offset and full-offset data.

In the second set of experiments, I perform the acoustic waveform inversion of marine reflection seismic data simulated for realistic P-wave velocity and density models. Here, I assess the different model parameterizations in terms of the quality of the reconstructed models and of the convergence rate of the inversion.

4.2 Theoretical background

The gradient expression in terms of new model parameters $\mathbf{m}_{\text{new}}$ can be derived as follows [Mora, 1987]

$$\delta \mathbf{m}_{\text{new}} = \frac{\partial \mathbf{m}}{\partial \mathbf{m}_{\text{new}}} \delta \mathbf{m} \tag{4.1}$$

It requires the computation of the Jacobian $\partial \mathbf{m}/\partial \mathbf{m}_{\text{new}}$ and the gradient of the original model parameters. The gradient for bulk modulus K and density ρ at iteration n can be written as [Tarantola, 1984]

$$\begin{aligned} \delta K_n &= \frac{1}{K_n^2} \sum_{\text{shots}} \int_t dt \; \frac{\partial p_n}{\partial t} \; \frac{\partial p_n^{'}}{\partial t} \; , \\ \delta \rho_n &= \frac{1}{\rho_n^2} \sum_{\text{shots}} \int_t dt \; \nabla p_n \cdot \nabla p_n^{'} \; , \end{aligned} \tag{4.2}$$

where $p_n(x, z, t)$ is the forward propagated field in the current model, and $p_n^{'}(x, z, t)$ is generated by propagating the residual data from all receiver positions backward in time.

To evaluate the gradient in terms of P-wave velocity and density, we need the relationship between the P-wave velocity V_P, the bulk modulus K and density ρ, which is $K = \rho V_P^2$. The gradient for the P-wave velocity can be expressed as

$$\delta V_P = \frac{\partial K}{\partial V_P} \; \delta K + \frac{\partial \rho}{\partial V_P} \; \delta \rho = 2 \rho V_P \; \delta K \; . \tag{4.3}$$

The gradient for the density $\delta \rho_{v_p}$ for the parameterization $\mathbf{m_1} = [V_P, \rho]$ can be written as

$$\delta \rho_{v_p} = \frac{\partial K}{\partial \rho} \; \delta K + \frac{\partial \rho}{\partial \rho} \; \delta \rho = V_P^2 \; \delta K + \delta \rho \; . \tag{4.4}$$

To derive the gradients with respect to the P-wave impedance I_P and density, we use the relationship $K = {I_P}^2/\rho$. The gradients for the parameterization $\mathbf{m_2} = [I_P, \rho]$ are thus

$$\begin{aligned} \delta I_P &= \frac{2 I_P}{\rho} \; \delta K \; , \\ \delta \rho_{imp} &= -\frac{1}{{I_P}^2} \; \delta K + \delta \rho \; . \end{aligned} \tag{4.5}$$

Using the equation relating the P-wave impedance I_P and P-wave velocity V_P to the bulk modulus, which is $K = V_P I_P$, the gradients in terms of I_P and V_P can be written as

$$\begin{aligned} \delta I_P &= \frac{1}{V_P {I_P}^2} \; \delta K \; , \\ \delta V_{P imp} &= \frac{1}{{V_P}^2 I_P} \; \delta K \; . \end{aligned} \tag{4.6}$$

4.3 Numerical experiment settings

The numerical tests presented in this study are based on a subset of the acoustic Marmousi2 model. The P-wave velocity and density models, as well as the resulting P-wave impedance model, are shown in Figure 4.1. The source is a pressure source, located 7.5 m below the air-water interface, with the Ricker wavelet time function. The frequency content of a signal is limited to a bandwidth from 3 to 20 Hz. The near offset is 100 m and the maximum offset is 4 km. The total of 50 shot gathers are generated at a 50 m interval with 3 seconds of data.

To reduce the high complexity of the inverse problem, the multi-scale inversion approach is applied. The frequency bands are applied sequentially with the following maximum frequencies $f_{\max} = (3, 4, 5.3, 7.1, 9.4; 12.5, 16.6, 20)$ Hz; that define the cutoff frequency for the Butterworth low-pass filter. Furthermore, to correct for the amplitude loss with depth due to geometrical spreading and to enhance deeper parts of the model, the linear gradient scaling with depth is implemented [Mora, 1987]. The parameters of the water layer and the source time function are assumed to be known. The starting model of any perturbed parameter is a 1D smooth representation of the true model.

4.4 The coupling effects

As mentioned above, the coupling is a function of the offset. To investigate this relationship, I perform separately the multi-parameter inversion of the near-offset and of the full-offset data. I define the long-offset data as those data acquired with source-receiver offset greater than the depth to the imaging targets. The assumed imaging target is a gas lens located at the depth of 1 km. Therefore, the maximum offset of the short-offset data is set to 1.15 km, whereas the full-offset data contains all offsets. Because of the moving streamer acquisition, the far offset is ranging from 1.125 km to the maximum of 4 km. The waves of the long-offset data propagate more horizontally than that of the near-offset data, thus they illuminate the subsurface in a different way. The maximum angle of incidence of a ray reflected at the gas lens is approximately 26° for the short-offset data and 63° in case of the full-offset data.

In this section, I investigate the trade-off between different parameter sets. This is achieved by performing the multi-parameter inversion for different combinations of parameters $\mathbf{m_1} = [V_P, \rho]$, $\mathbf{m_2} = [I_P, \rho]$, and $\mathbf{m_3} = [V_P, I_P]$. For each configuration, the model parameters are perturbed individually, which results in six combinations of acoustic models. Synthetic data are generated for each set of acoustic models. The modelling and acquisition parameters are listed in Table 3.1.

4.4.1 Inversion results

The inversion results for the first parameter set $\mathbf{m_1} = [V_P, \rho]$ are shown in Figure 4.3 for a variable V_P model and constant density. The true models and starting models are shown in Figure 4.2). Figure 4.5 shows the FWI results for the same parameters set, but here V_P is constant and variable density model is used. We can observe a quite strong coupling between V_P and ρ, especially when the near-offset data is inverted. There are velocity structures present in the density models, and strong cross-talk artefacts are visible in the inverted velocity model. These artefacts reflect the interfaces of the true density structures (Figure 4.5). However, the coupling between V_P and ρ, and thereby the

ambiguity of the inversion, is significantly decreasing once the information from far offsets is included. It is clearly visible on the velocity and density profiles (Figure 4.3d and Figure 4.5d).

The parameterization with $\mathbf{m_2} = [I_P, \rho]$ is illustrated in Figure 4.7 and in Figure 4.9. When considering a perturbation in P-wave impedance with a homogeneous density model, it can be seen that the acoustic impedance model is very well resolved from both near-offset and far-offset data. The final density models contain only weak artefacts resulting from the impedance structures. This indicates that both parameters are not strongly coupled. On the other hand, if we want to consider a perturbation in density only and preserve a zero impedance variation, this will result in unrealistic, negative velocity contrasts. To avoid this unplausible assumption, the true I_P is defined as $I_P = V_{P(\text{constant})} \cdot \rho_{(\text{perturbed})}$. The true I_P model is used as a starting model for the inversion. We can observe that the original impedance structures are hardly influenced by the density, both for the near-offset and for the full-offset data (Figure 4.9d). The density model is fairly well reconstructed, but still the accuracy is not as good as for $\mathbf{m_1} = [V_P, \rho]$. This agrees with the amplitude diffraction pattern analysis shown by Tarantola [1986]. A diffractor with a perturbation of density but constant impedances would be hardly visible using surface seismic reflection data with moderate offsets, because it only scatters energy downwards into the medium.

In the third parameterization, the medium is described by the P-wave velocity and the P-wave impedance $\mathbf{m_3} = [V_P, I_P]$. Here, we also face the problem of negative density contrast, which would be introduced if a constant impedance model and a perturbed velocity was assumed. For that reason, I define the true I_P as $I_P = V_{P(\text{perturbed})} \cdot \rho_{(\text{constant})}$, which is also an initial model for the inversion (Figure 4.11a). Relatively strong high-frequency artefacts are present in the inverted velocity model, especially when the near-offset data is inverted. Inversion of the full-offset data improves the quality of the reconstructed V_P model, however at the same time it negatively affects the I_P model (Figure 4.11d). This means that the trade-off between V_P and I_P is generally increasing while the information from far-offsets is included. Furthermore, the estimated optimum step-lengths are very variable and have much higher values than the step-lengths estimated for other parameterizations. This may indicate that the inverse problem is running into a local minimum of the misfit function. Additionally, the velocity and impedance errors are only decreasing up to a certain iteration number, afterwards the model errors are increasing, which is apparently associated with the inversion of the high-frequency data. However, at the same time, the data misfit function is gradually decreasing. Figure 4.13 shows the FWI results for a perturbed acoustic impedance with a homogeneous velocity model. The impedance structures are fairly well resolved, particularly from the short-offset data. Nevertheless, the velocities contain structural I_P information, which may lead to the incorrect interpretation of the inversion results.

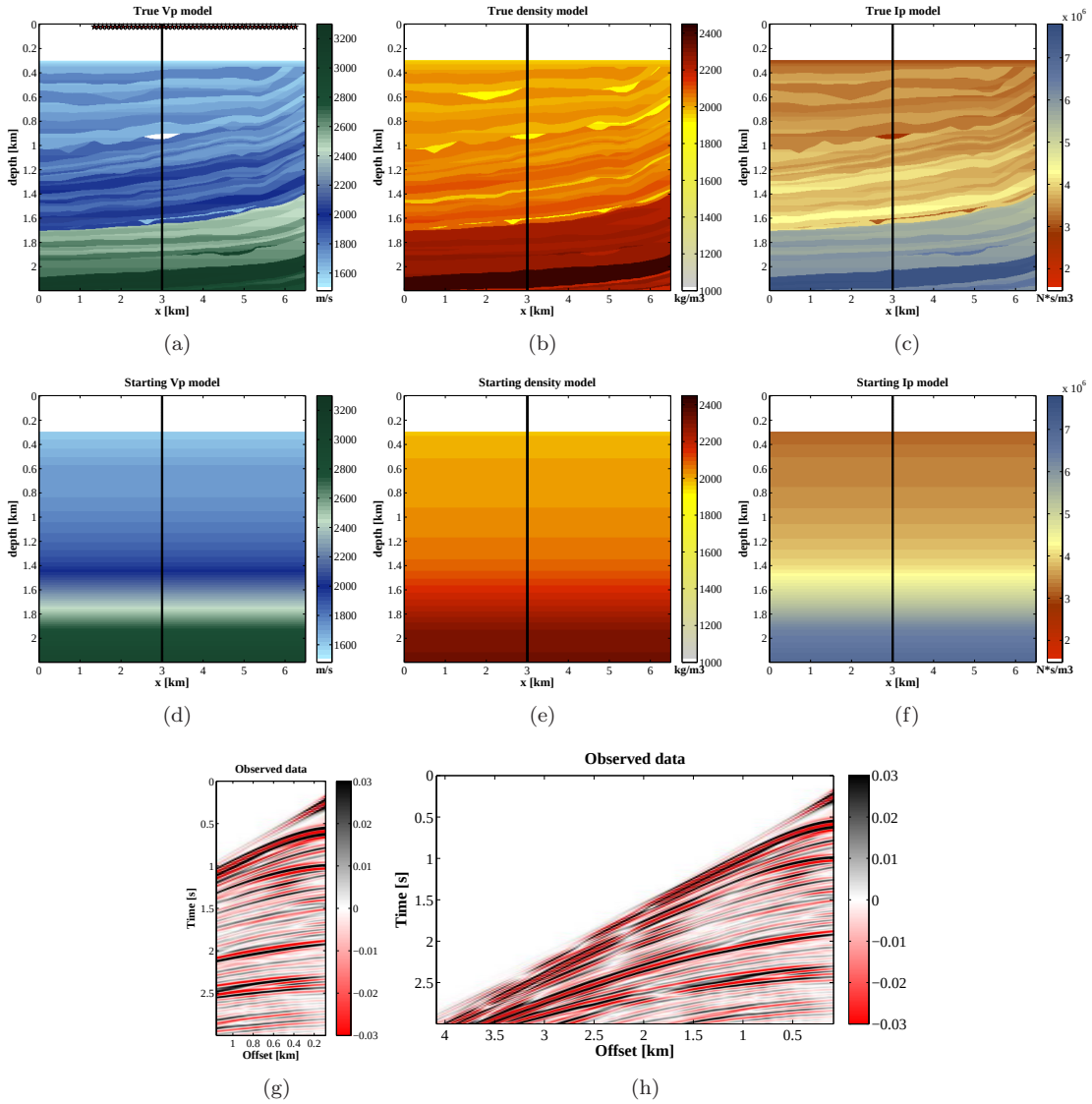


Figure 4.1: (a) The true P-wave velocity and (b) density models. The stars denote shot point locations. (c) The resulting P-wave impedance model $I_P = V_P\rho$. Initial (d) P-wave velocity, (e) density, and (f) impedance models. Acoustic data for true V_P and ρ model, shot 50 located at $x = 6.25$ km: (g) the near-offset data (0.1 km - 1.15 km), (h) the full-offset data (0.1 km - 4 km).

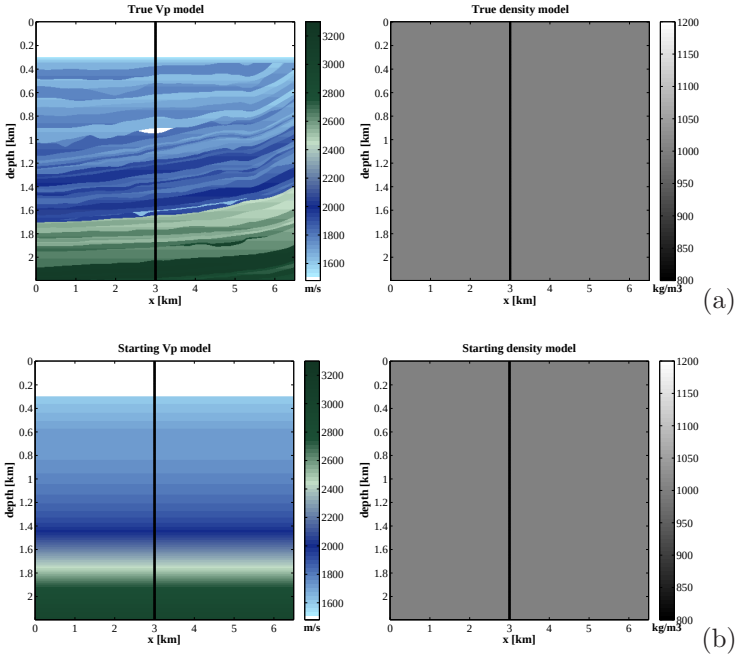


Figure 4.2: Parameter set $\mathbf{m_1} = [V_P, \rho]$. Perturbation in V_P, homogeneous ρ. (a) True V_P and ρ model; (b) Starting V_P and ρ model.

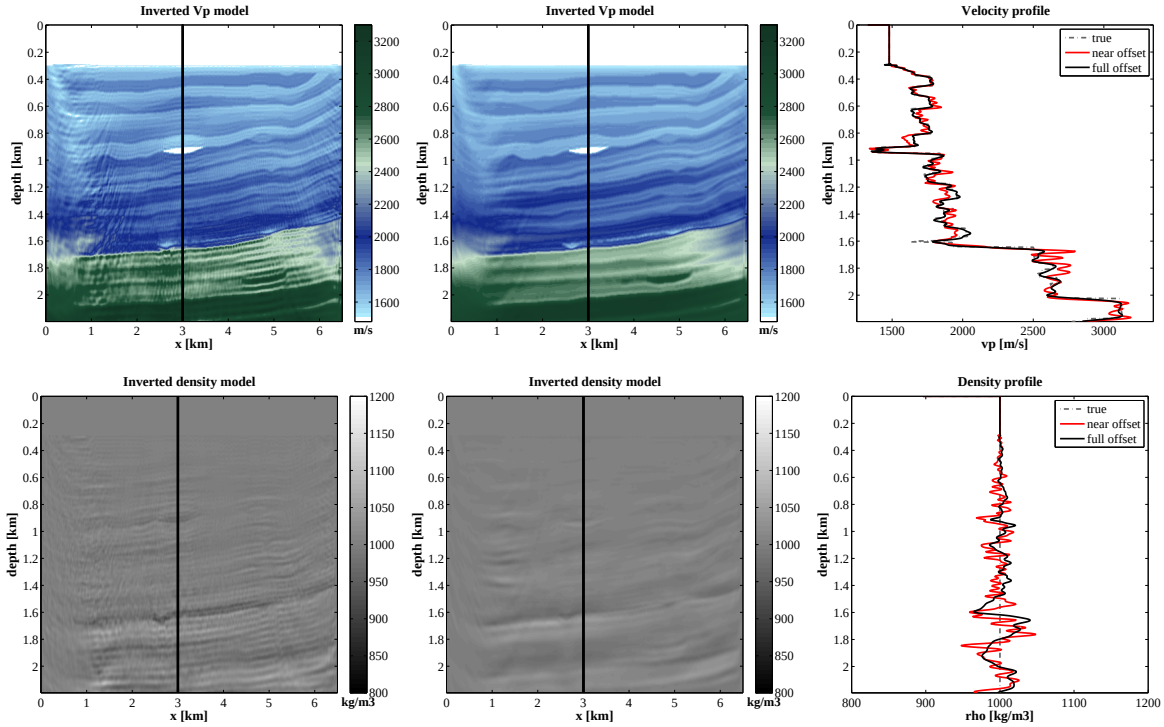


Figure 4.3: Parameter set $\mathbf{m_1} = [V_P, \rho]$. Perturbation in V_P, homogeneous ρ. (left) FWI results of the near-offset data (0.1 km - 1.15 km); (middle) FWI results of the full-offset data (0.1 km - 4 km); (right) V_P and ρ profiles at $x = 3$ km of the true models (dash-dot line) and of the inversion results for the near-offset data (red line) and for the full-offset data (black line).

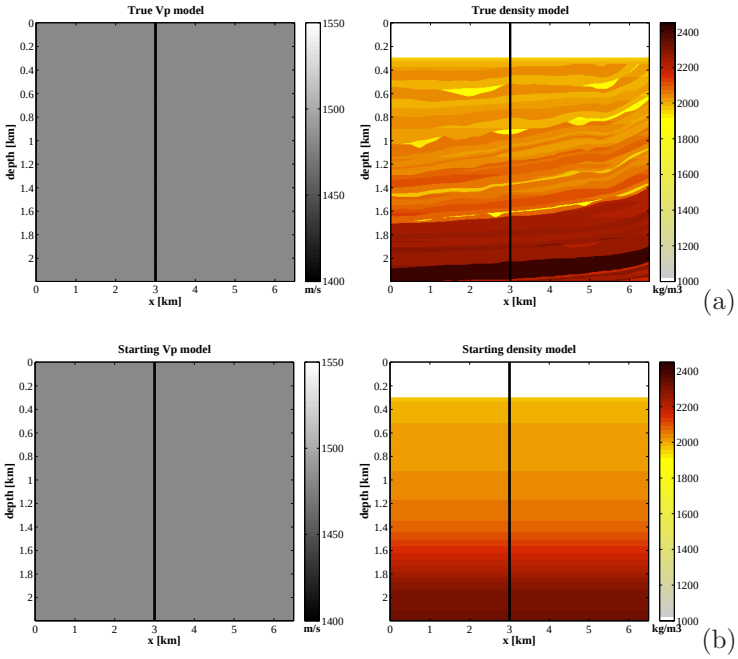


Figure 4.4: Parameter set $\mathbf{m_1} = [V_P, \rho]$. Perturbation in ρ, homogeneous V_P. (a) True V_P and ρ model; (b) Starting V_P and ρ model.

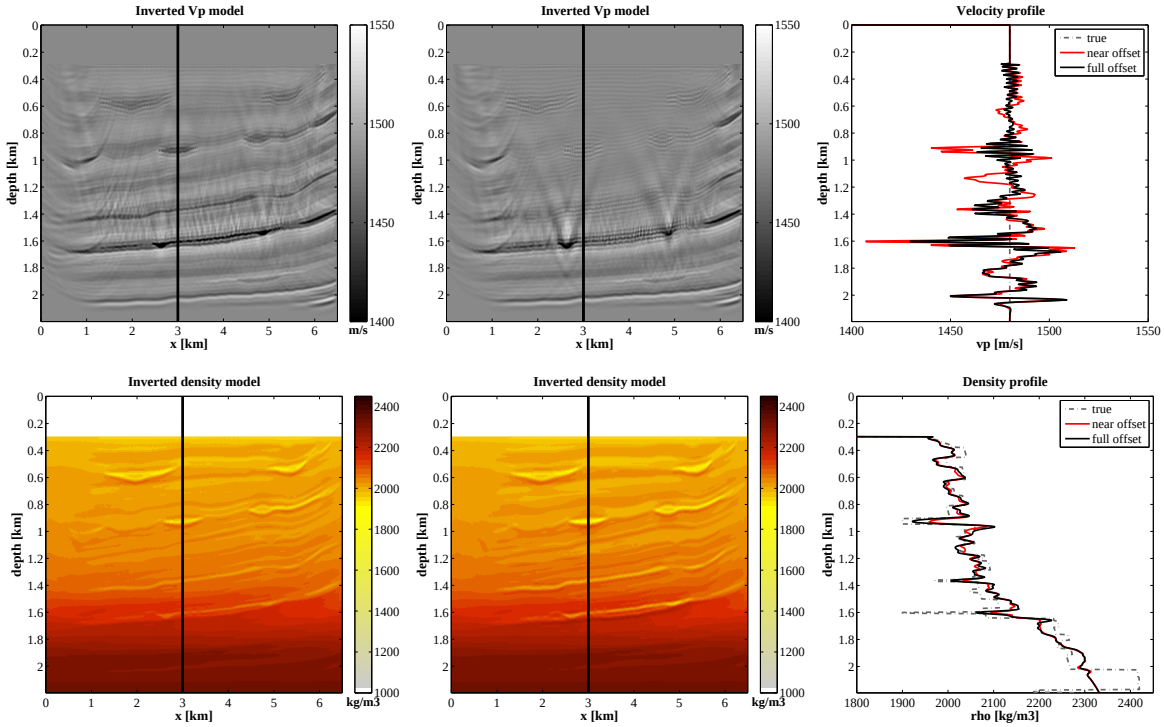


Figure 4.5: Parameter set $\mathbf{m_1} = [V_P, \rho]$. Perturbation in ρ, homogeneous V_P. (left) FWI results of the near-offset data (0.1 km - 1.15 km); (middle) FWI results of the full-offset data (0.1 km - 4 km); (right) V_P and ρ profiles of the true models and of the inversion results.

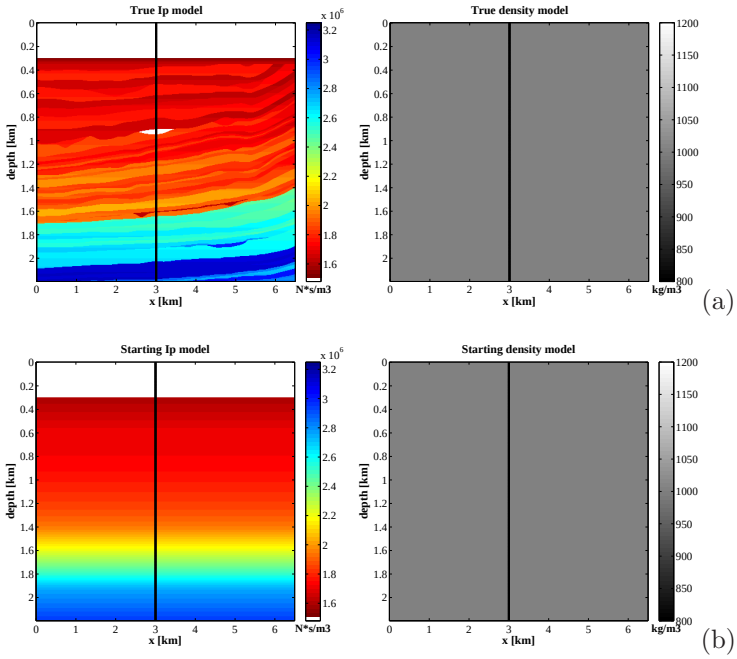


Figure 4.6: Parameter set $\mathbf{m_2} = [I_P, \rho]$. Perturbation in I_P, homogeneous ρ. (a) True I_P and ρ model; (b) Starting I_P and ρ model.

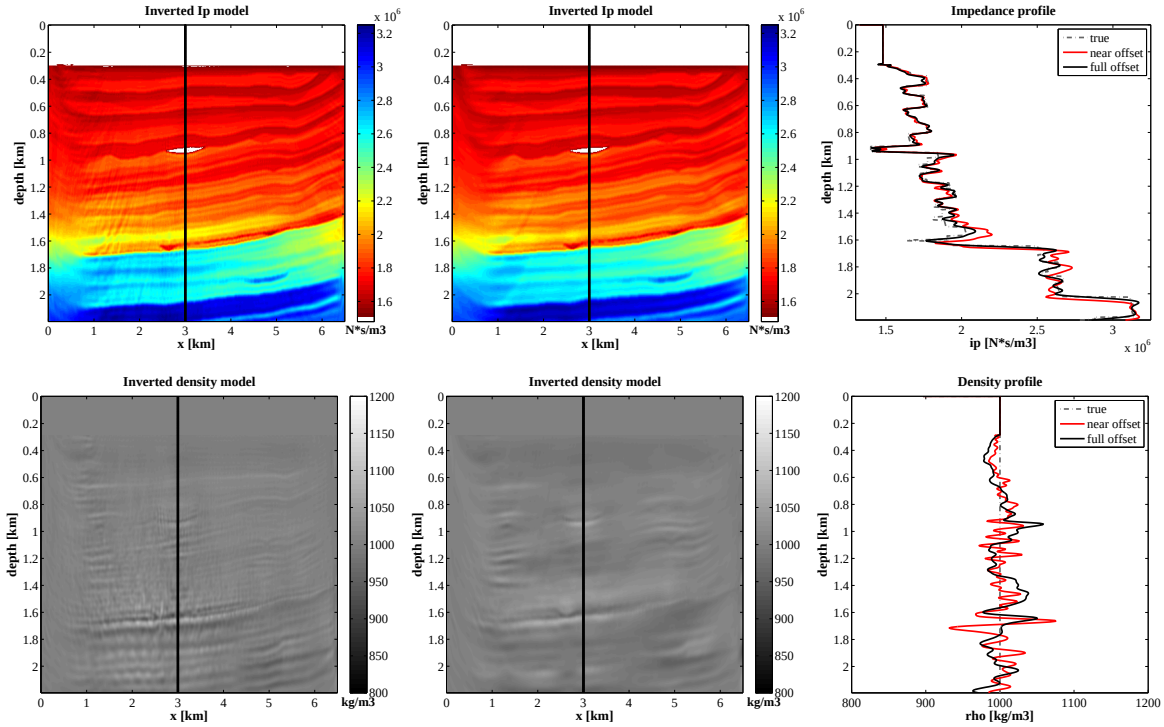


Figure 4.7: Parameter set $\mathbf{m_2} = [I_P, \rho]$. Perturbation in I_P, homogeneous ρ. (left) FWI results of the near-offset data (0.1 km - 1.15 km); (middle) FWI results of the full-offset data (0.1 km - 4 km); (right) I_P and ρ profiles of the true models and of the inversion results.

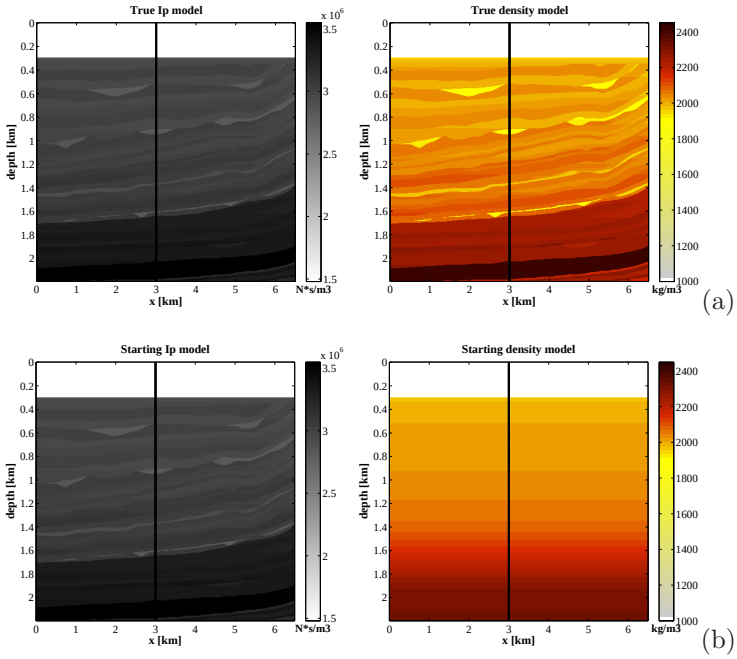


Figure 4.8: Parameter set $\mathbf{m_2} = [I_P, \rho]$. Perturbation in ρ, $I_P = V_{P(\text{constant})} \cdot \rho_{(\text{perturbed})}$. (a) True I_P and ρ model; (b) Starting I_P and ρ model.

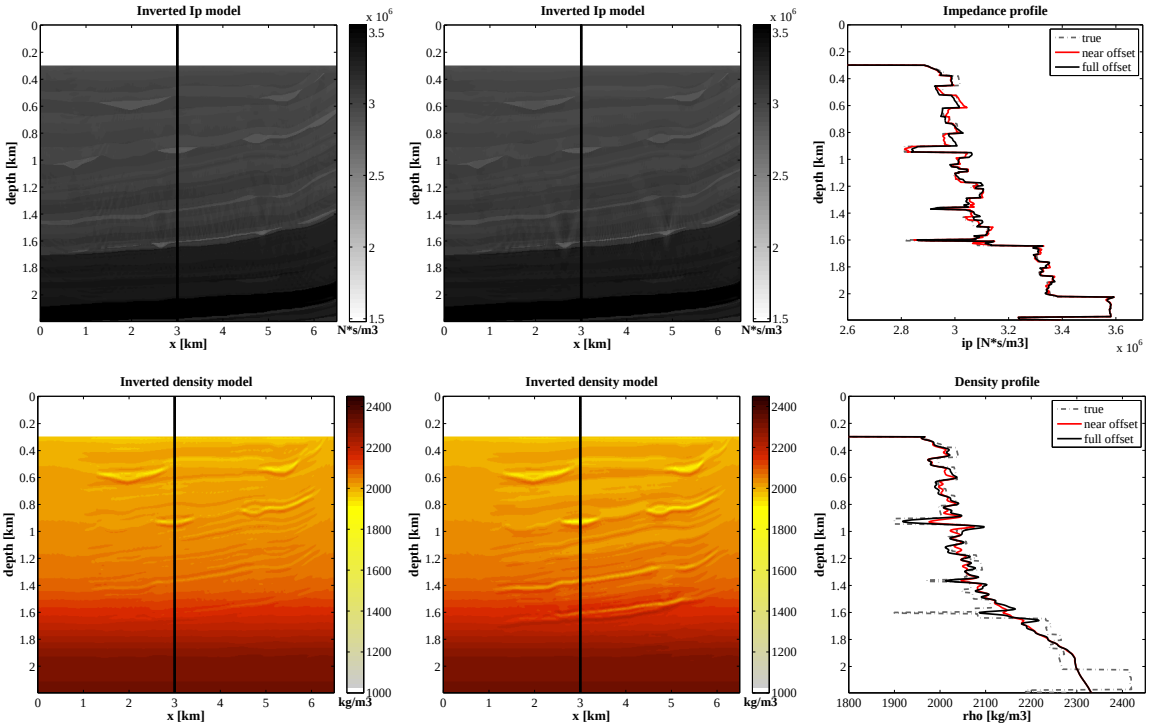


Figure 4.9: Parameter set $\mathbf{m_2} = [I_P, \rho]$. Perturbation in ρ, $I_P = V_{P(\text{constant})} \cdot \rho_{(\text{perturbed})}$. (left) FWI results of the near-offset data (0.1 km - 1.15 km); (middle) FWI results of the full-offset data (0.1 km - 4 km); (right) I_P and ρ profiles of the true models and of the inversion results.

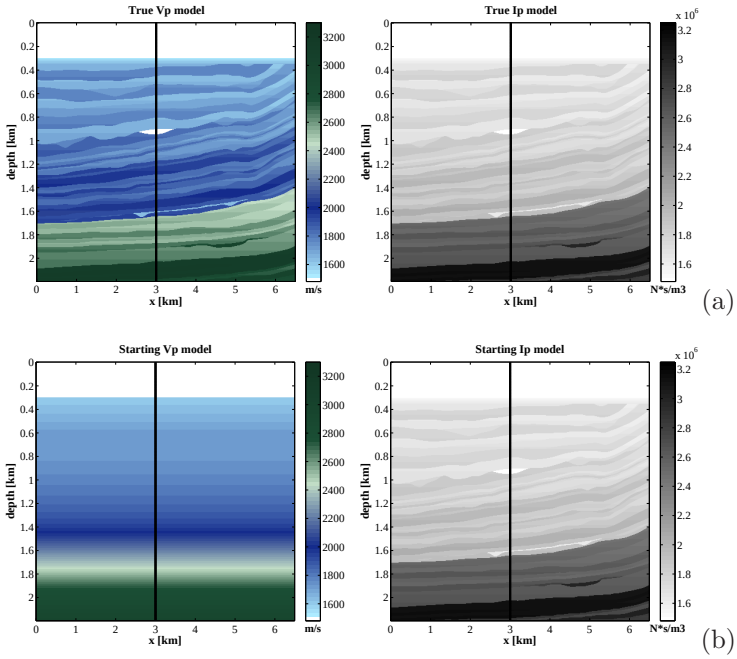


Figure 4.10: Parameter set $\mathbf{m_3} = [V_P, I_P]$. Perturbation in V_P, $I_P = V_{P(\text{perturbed})} \cdot \rho_{(\text{constant})}$. (a) True V_P and I_P model; (b) Starting V_P and I_P model.

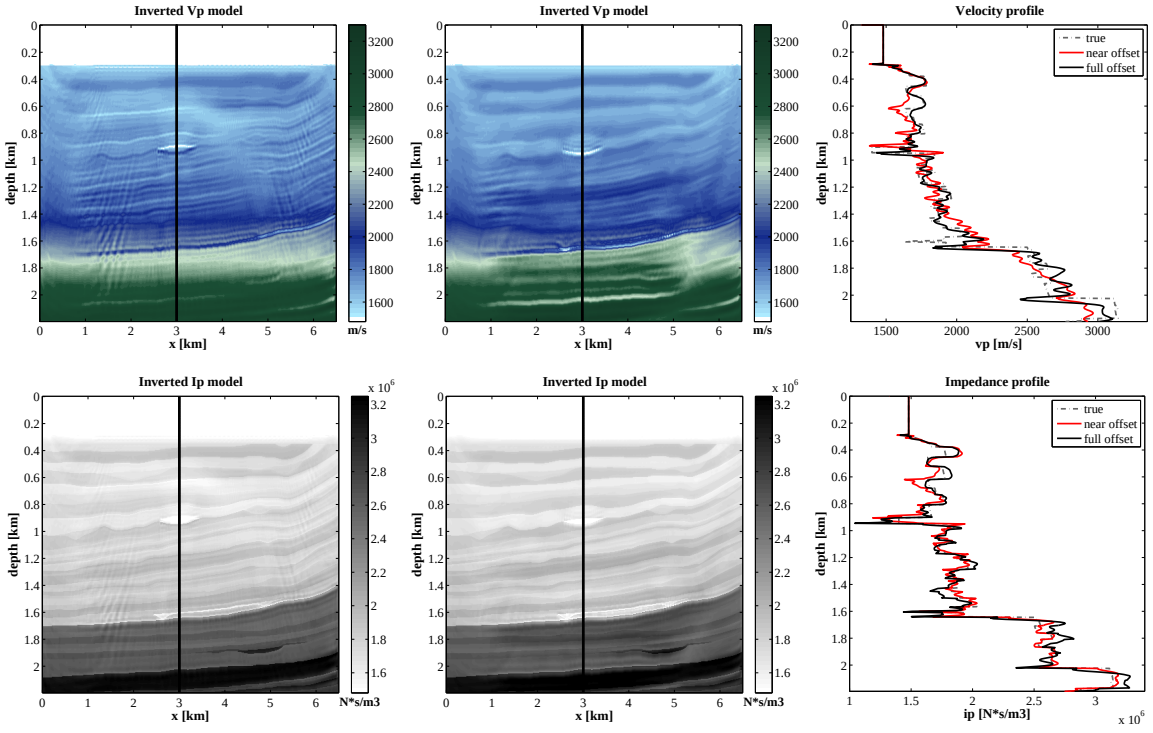


Figure 4.11: Parameter set $\mathbf{m_3} = [V_P, I_P]$. Perturbation in V_P, $I_P = V_{P(\text{perturbed})} \cdot \rho_{(\text{constant})}$. (left) FWI results of the near-offset data (0.1 km - 1.15 km); (middle) FWI results of the full-offset data (0.1 km - 4 km); (right) V_P and I_P profiles of the true models and of the inversion results.

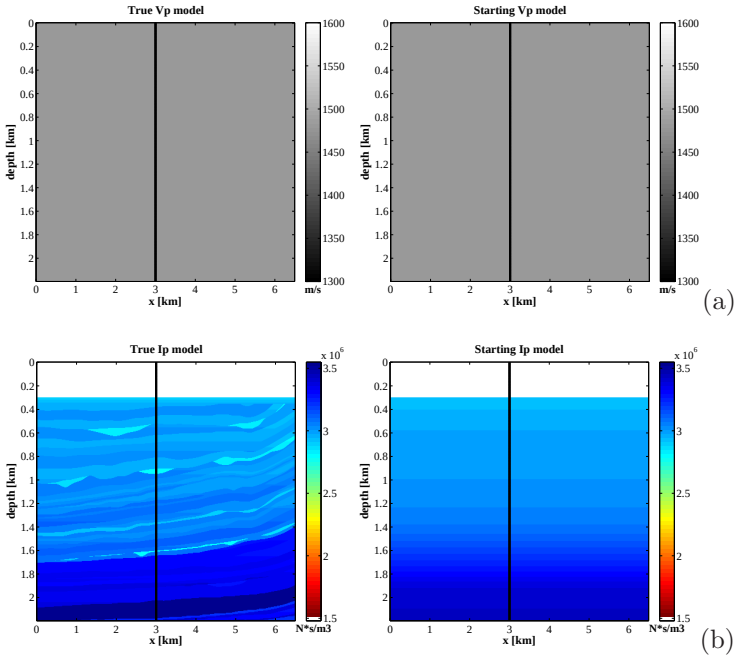


Figure 4.12: Parameter set $\mathbf{m_3} = [V_P, I_P]$. Perturbation in I_P, homogeneous V_P. True and starting models. (a) True V_P and I_P model; (b) Starting V_P and I_P model.

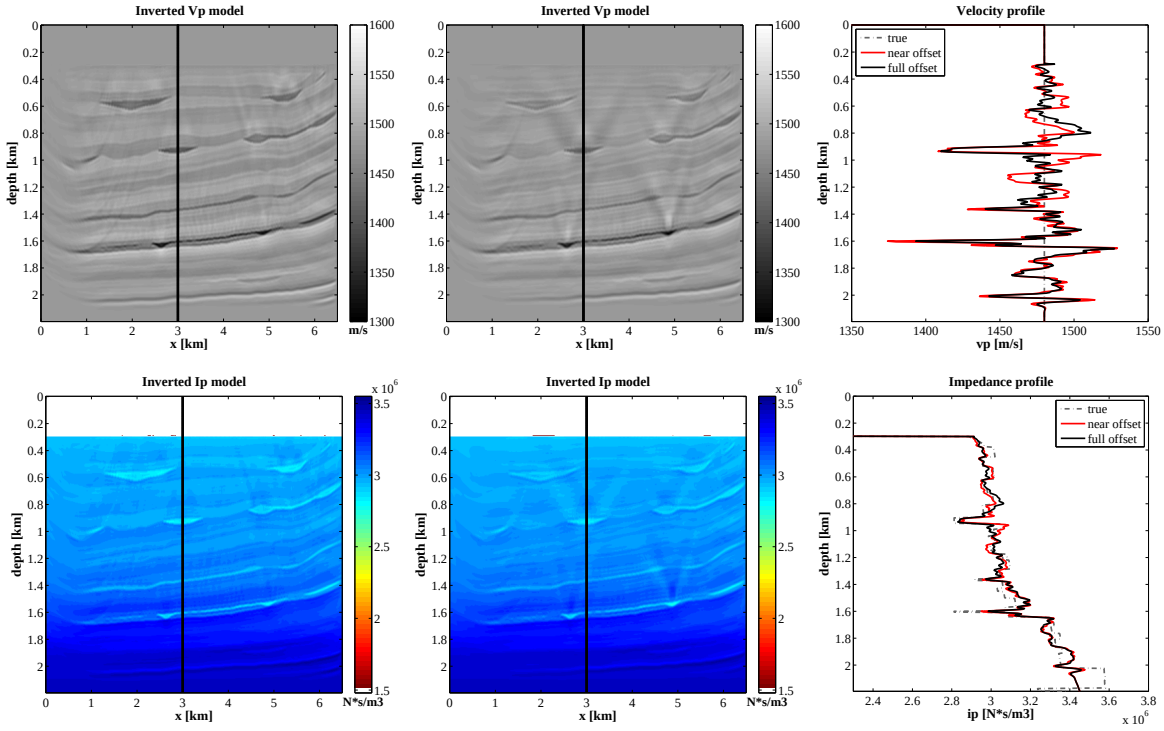


Figure 4.13: Parameter set $\mathbf{m_3} = [V_P, I_P]$. Perturbation in I_P, homogeneous V_P. (left) FWI results of the near-offset data (0.1 km - 1.15 km); (middle) FWI results of the full-offset data (0.1 km - 4 km); (right) V_P and I_P profiles of the true models and of the inversion results.

4.5 Gradients comparison

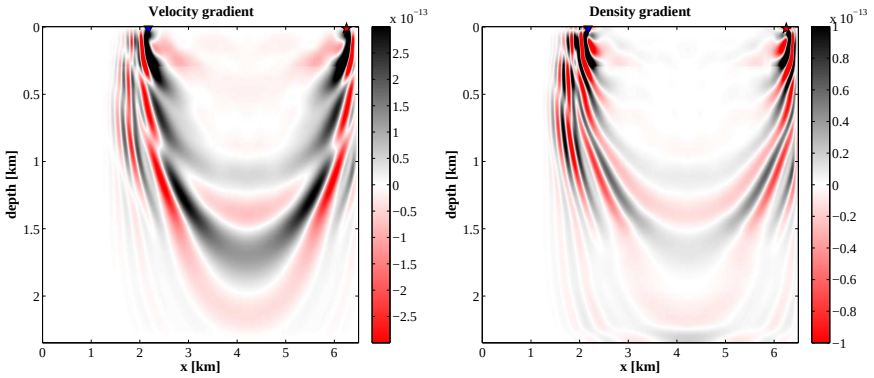
In order to better understand the difference between different parameterizations, I compare the gradients of V_P, I_P and density computed with $\mathbf{m_1} = [V_P, \rho]$, $\mathbf{m_2} = [I_P, \rho]$, and $\mathbf{m_3} = [V_P, I_P]$. Gradients are calculated with Eq.4.3 - Eq.4.6.

Gradients are calculated at the first iteration for the starting model (Figure 4.1d-f) using a source-receiver pair. The source is located at $x = 6.250$ km and the receiver is located at $x = 2.175$ km. The offset of 4.075 km corresponds to the maximum offset of the data set. Gradients are computed for two frequency bands: from 3 to 5 Hz, and from 3 to 20 Hz. Such a band-limited, multi-frequency gradient calculated in the time domain can be formed by integrating monochromatic gradients calculated in the frequency-domain for the same frequency range. The total gradient, that defines the model update, is the superposition of gradients for all source-receiver pairs.

Figure 4.14 illustrates gradients computed for different parameter sets for the frequency range from 3 to 5 Hz. There is no preconditioning applied to the gradient, except for a circular taper around the source and the receiver. Because the amplitude of the wavefield is very high in the vicinity of the source and the receiver, the gradient has large values in these parts of the model and it would overshadow the rest of the model. The colorbar scales of the same model parameter are identical, such that V_P gradient from parameterization $\mathbf{m_1} = [V_P, \rho]$ can be directly compared with the V_P gradient from parameterization $\mathbf{m_3} = [V_P, I_P]$. The general observation is that the gradients computed for different parameter sets are relatively similar to each other.

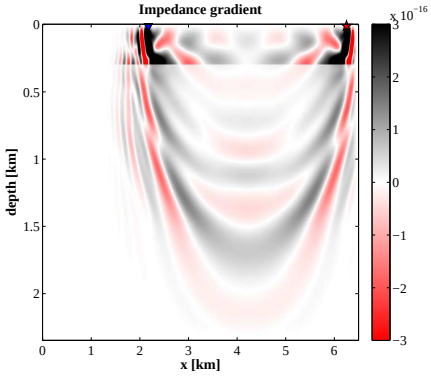
Due to the presence of primary and multiple reflections in the data set and multiple frequencies, the gradients for one source-receiver pair are more complex that the gradients computed in a homogeneous or a smooth background model [Bleibinhaus and Rondenay, 2009; Operto et al., 2006; Pratt et al., 1996; Sirgue, 2003], where the first Fresnel zone can be easily identified. A monochromatic gradient computed for a given source-receiver pair, which is referred to as a wavepath [Woodward, 1992], is often used to explain the imaging limitations of the waveform tomography for a given acquisition geometry [Pratt et al., 1996]. The wavepath computed in a background medium contains distorted elliptical lobes, where the central zone corresponds to the Fresnel zone. The arrivals within the first Fresnel zone correspond mainly to transmitted events, whereas the reflected waves map the outer Fresnel zones [Sirgue, 2003].

The complexity of gradient images increases significantly when the frequency content of the data is higher. The gradients computed for the frequency range from 3 to 20 Hz are shown in Figure 4.15. Here, we can observe more discrepancies between different parameterizations than in the case of the low frequency band data. First of all, there is a difference between the V_P gradient calculated in the $\mathbf{m_1} = [V_P, \rho]$ parameterization and in the $\mathbf{m_3} = [V_P, I_P]$ parameterization (Figure 4.15a,e). Gradients computed within the same parameterization are very similar to each other in the case of $\mathbf{m_1} = [V_P, \rho]$ and $\mathbf{m_3} = [V_P, I_P]$, but the density gradient differs significantly from the impedance gradient in the parameterization $\mathbf{m_2} = [I_P, \rho]$ (Figure 4.15c,d). On the other hand both density gradients are very similar (Figure 4.15b,d) and they resemble the V_P gradient in the $\mathbf{m_1} = [V_P, \rho]$ parameterization.

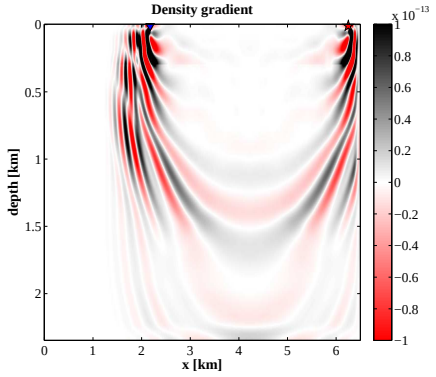


(a) $\mathbf{m_1}$, V_P gradient

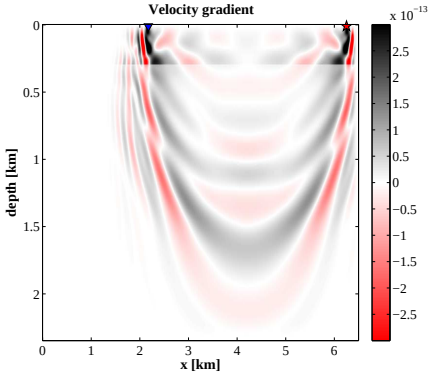
(b) $\mathbf{m_1}$, ρ gradient



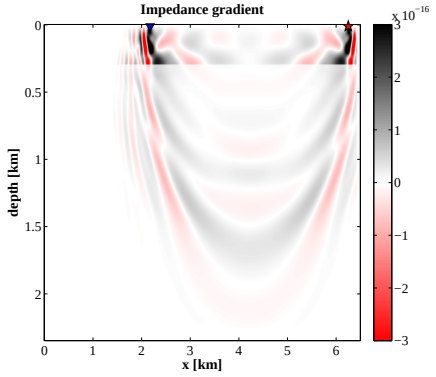
(c) $\mathbf{m_2}$, I_P gradient



(d) $\mathbf{m_2}$, ρ gradient

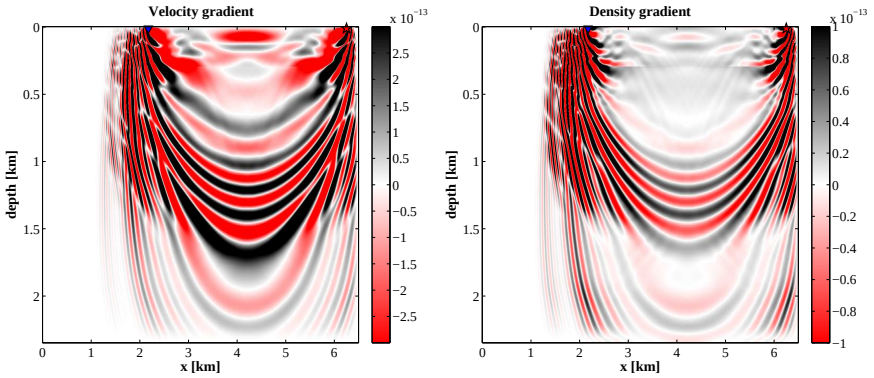


(e) $\mathbf{m_3}$, V_P gradient



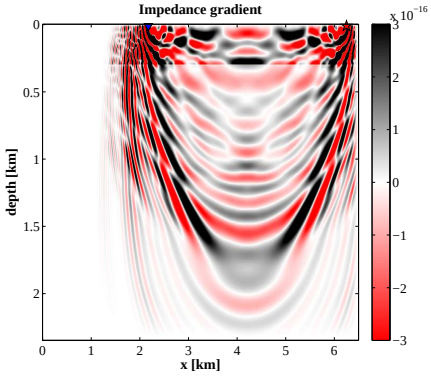
(f) $\mathbf{m_3}$, I_P gradient

Figure 4.14: Marmousi2 model. Gradient at the first iteration computed for a source-receiver pair for a frequency band from 3 to 5 Hz. Star denotes shot location, triangle denotes receiver location.

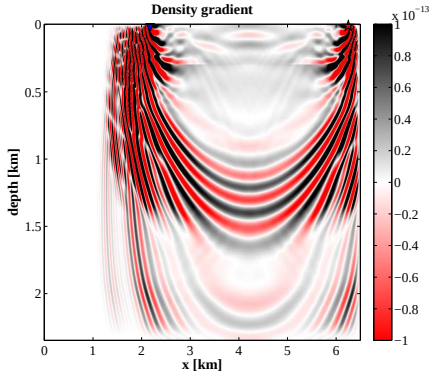


(a) $\mathbf{m_1}$, V_P gradient

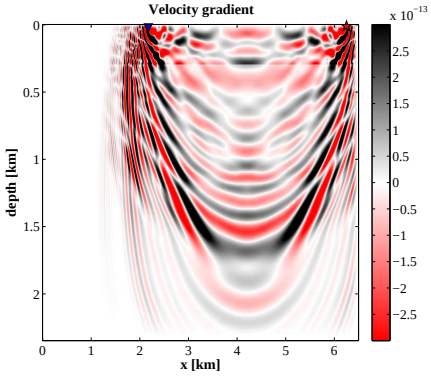
(b) $\mathbf{m_1}$, ρ gradient



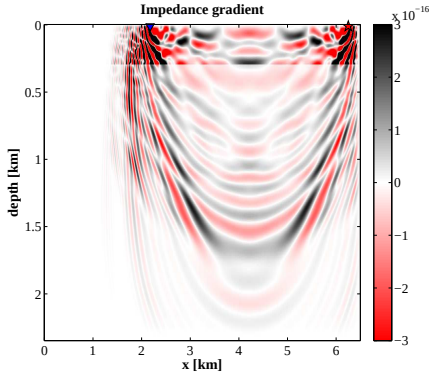
(c) $\mathbf{m_2}$, I_P gradient



(d) $\mathbf{m_2}$, ρ gradient



(e) $\mathbf{m_3}$, V_P gradient



(f) $\mathbf{m_3}$, I_P gradient

Figure 4.15: Marmousi2 model. Gradient at the first iteration computed for a source-receiver pair for a frequency band from 3 to 20 Hz. Star denotes shot location, triangle denotes receiver location.

4.6 Acoustic multi-parameter inversion - Marmousi2 model

The acoustic data set is generated for the Marmousi2 model (Figure 4.1). The near-offset and full-offset data are shown in Figure 4.1g and 4.1h respectively. Figure 4.1d-f represents initial models used in the full waveform inversion. The aim of this experiment is to assess the different model parameterizations in terms of the quality of the reconstructed models and of the convergence rate of the inversion.

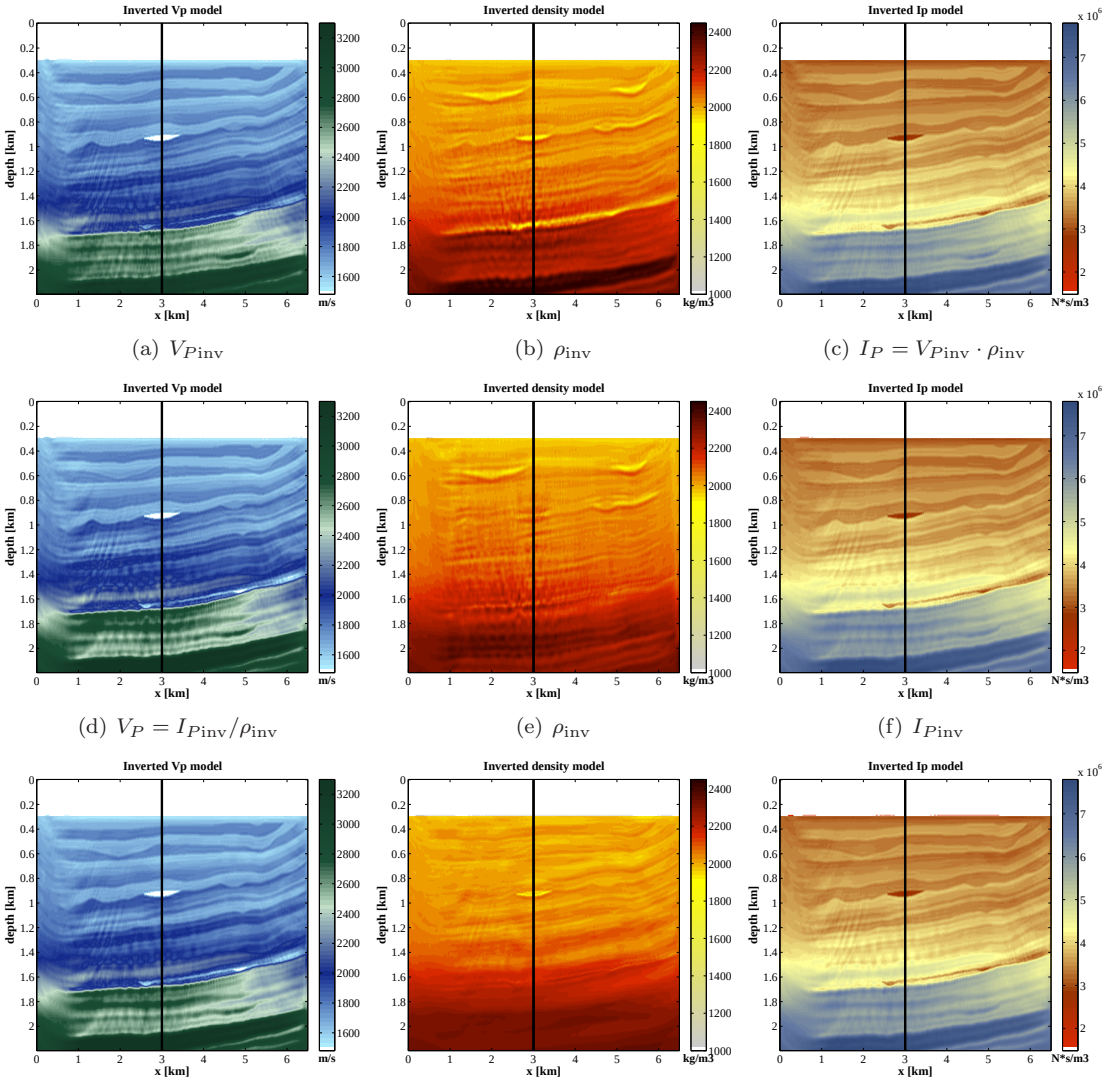
The multi-parameter FWI results using the near-offset information are shown in Figure 4.16. The same data set is inverted using different parameterizations. Additionally, the third parameter, that is not represented by a particular parameter set, is recalculated from the final inverted models. This allows for a comparison of all considered parameter classes. The upper structures of the velocity and impedance models, down to the depth of about 1 km, are well reconstructed. However, the deeper parts of the inverted models show strong high-frequency artefacts. The density is well resolved only in the $\mathbf{m_1} = [V_P, \rho]$ parameterization.

The inversion results of the full-offset data are presented in Figure 4.17. Here, we can observe that the recovery of all parameters has improved significantly. This suggests, that the far-offset data reduces the ambiguity of the inversion and provides an important information for the reconstruction of the subsurface structures.

The depth profiles intersecting the gas lens are shown in Figure 4.18. They compare the true model with the inversion results for the near-offset and full-offset data. The resolution of the P-wave velocity and acoustic impedance is comparable for every parameterization. But the main difference concerns the density reconstruction. The density structures are fairly well resolved only in the $\mathbf{m_1} = [V_P, \rho]$ parameterization. The worst density image is obtained, when the parameterization $\mathbf{m_2} = [I_P, \rho]$ is used. This poor result has been confirmed by the coupling effects analysis (see Figure 4.9). On the other hand, the density values recalculated from the inverted velocity and impedance models ($\mathbf{m_3} = [V_P, I_P]$) are significantly underestimated. Some density structures are not reconstructed at all and only the location of density interfaces can be recognized.

Figure 4.19 illustrates the comparison of the inversion progress for different parameterizations between the near-offset and full-offset data inversion. The evolution of the data misfit function (Figure 4.19a,b) shows the most stable progress and the most effective reduction of the data residuals for parameter set $\mathbf{m_1} = [V_P, \rho]$. The jumps in the misfit function for iterations 1-140 are related to the multi-scale inversion approach. At these iterations the next frequency band is included. The data misfit reduction is much worse for $\mathbf{m_2} = [I_P, \rho]$ and $\mathbf{m_3} = [V_P, I_P]$.

In order to quantitatively assess the inversion results, the L_1-based model error between the true and reconstructed velocity models is measured. We can observe, that the significant reduction of data residuals is followed by the best reconstruction of the model parameters for $\mathbf{m_1} = [V_P, \rho]$. The convergence of the model errors is much slower, when we choose the parameter set $\mathbf{m_2} = [I_P, \rho]$. The worst quality of the final models is provided by the parameterization $\mathbf{m_3} = [V_P, I_P]$. The model errors of all parameters are significantly reduced, when the full-offset data is included.

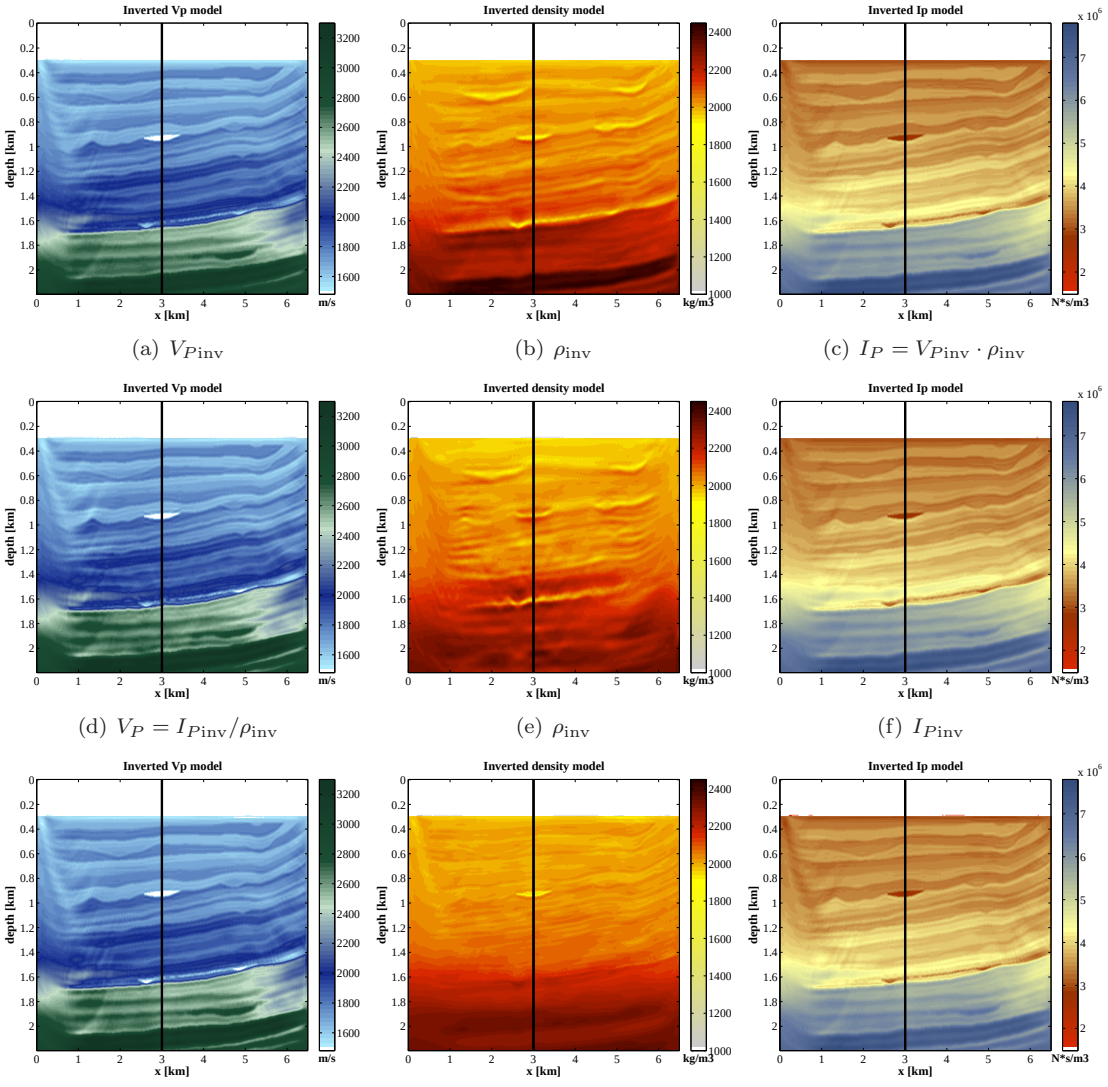


(a) $V_{P\text{inv}}$ (b) ρ_{inv} (c) $I_P = V_{P\text{inv}} \cdot \rho_{\text{inv}}$

(d) $V_P = I_{P\text{inv}}/\rho_{\text{inv}}$ (e) ρ_{inv} (f) $I_{P\text{inv}}$

(g) $V_{P\text{inv}}$ (h) $\rho = I_{P\text{inv}}/V_{P\text{inv}}$ (i) $I_{P\text{inv}}$

Figure 4.16: Marmousi2 model. FWI results for the near-offset data. Top row $\mathbf{m_1} = [V_P, \rho]$, middle row $\mathbf{m_2} = [I_P, \rho]$, bottom row $\mathbf{m_3} = [V_P, I_P]$. Left: P-wave velocity, middle: density, right: acoustic impedance.

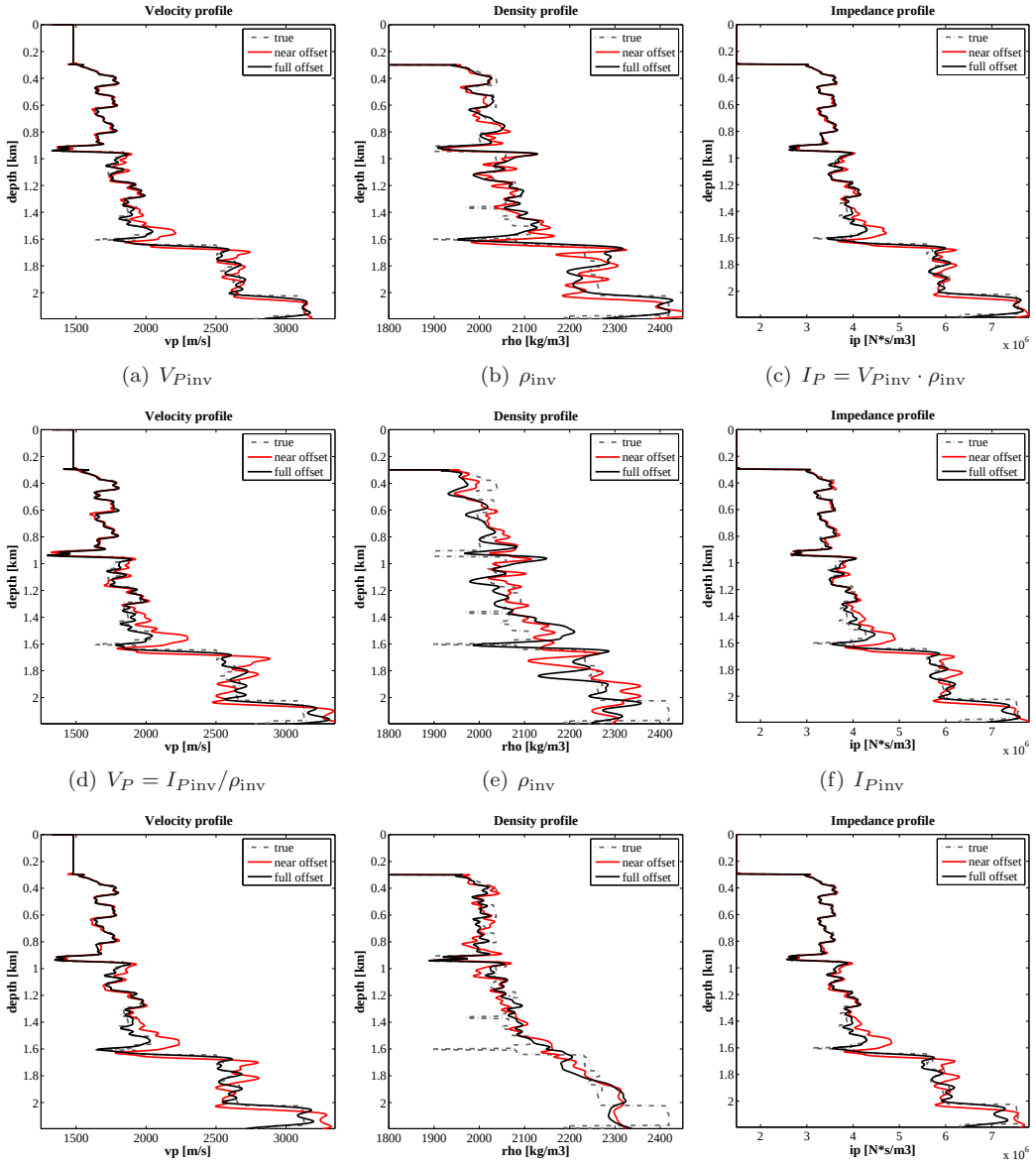


(a) $V_{P\text{inv}}$ (b) ρ_{inv} (c) $I_P = V_{P\text{inv}} \cdot \rho_{\text{inv}}$

(d) $V_P = I_{P\text{inv}}/\rho_{\text{inv}}$ (e) ρ_{inv} (f) $I_{P\text{inv}}$

(g) $V_{P\text{inv}}$ (h) $\rho = I_{P\text{inv}}/V_{P\text{inv}}$ (i) $I_{P\text{inv}}$

Figure 4.17: Marmousi2 model. FWI results for the full-offset data. Top row $\mathbf{m_1} = [V_P, \rho]$, middle row $\mathbf{m_2} = [I_P, \rho]$, bottom row $\mathbf{m_3} = [V_P, I_P]$. Left: P-wave velocity, middle: density, right: acoustic impedance.

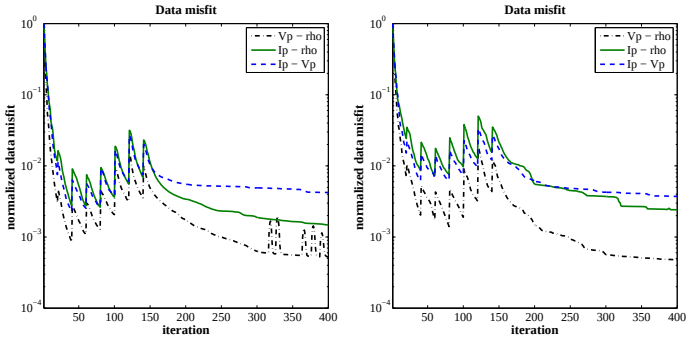


(a) $V_{P\text{inv}}$ (b) ρ_{inv} (c) $I_P = V_{P\text{inv}} \cdot \rho_{\text{inv}}$

(d) $V_P = I_{P\text{inv}}/\rho_{\text{inv}}$ (e) ρ_{inv} (f) $I_{P\text{inv}}$

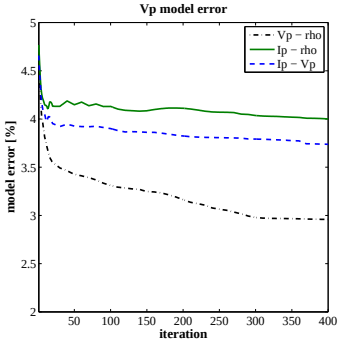
(g) $V_{P\text{inv}}$ (h) $\rho = I_{P\text{inv}}/V_{P\text{inv}}$ (i) $I_{P\text{inv}}$

Figure 4.18: Marmousi2 model. Depth profiles at $x = 3$ km of the true models (dash-dot line) and of the FWI results for the near-offset (red solid line) and full-offset data inversion (black solid line). Top row $\mathbf{m_1} = [V_P, \rho]$, middle row $\mathbf{m_2} = [I_P, \rho]$, bottom row $\mathbf{m_3} = [V_P, I_P]$. Left: P-wave velocity, middle: density, right: acoustic impedance.

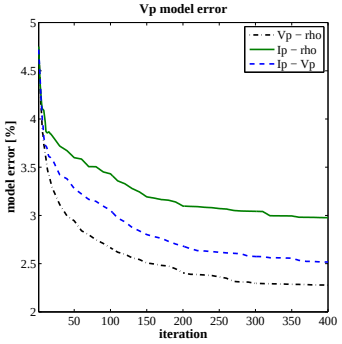


(a) data misfit - near offset

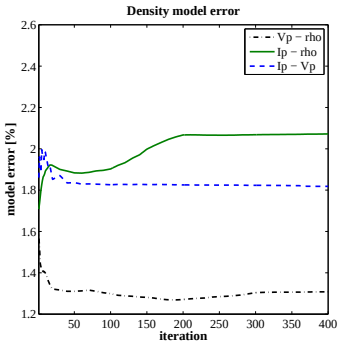
(b) data misfit - full offset



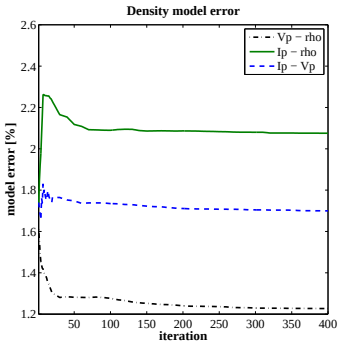
(c) V_P error - near offset



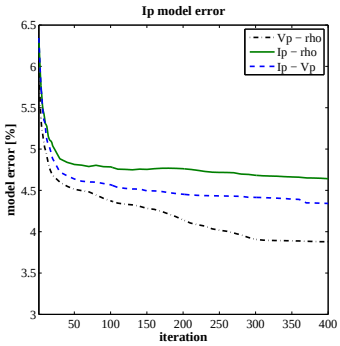
(d) V_P error - full offset



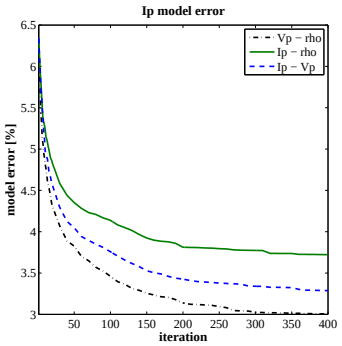
(e) ρ error - near offset



(f) ρ error - full offset



(g) I_P error - near offset



(h) I_P error - full offset

Figure 4.19: Marmousi2 model. Summary of FWI results for the near-offset (left) and the full-offset data (right). $\mathbf{m_1} = [V_P, \rho]$ (black dash-dotted line), $\mathbf{m_2} = [I_P, \rho]$ (green solid line), $\mathbf{m_3} = [V_P, I_P]$ (blue dashed line). The data misfit functions are normalized with respect to their maximum value.

4.7 Conclusions

In this study, I have investigated the influence of different model parameterizations on the multi-parameter acoustic waveform inversion. The numerical experiments have been designed to reflect the field measurement conditions of marine reflection seismics. Since the lack of low frequencies is a general problem in seismic recording, the frequency content of the data is limited to a range from 3 to 20 Hz.

First I analyzed the coupling between different parameters and considered separately the information from the near-offset and far-offset data. The far-offset information used in the inversion reduces the coupling between different parameters, i.e. it reduces the ambiguity of the inversion. When ambiguity is present, it results in the incorrect solution and may lead to a wrong interpretation of the inversion results.

The choice of the model parameters affects the convergence and the accuracy of final models, when inverting the marine reflection seismic data. Whereas the resolution of velocity and impedance models is comparable, the reconstruction of density structures strongly depends on the model parameterization. Out of the investigated parameter sets, the velocity and density provided the best convergence rate and the best accuracy of the inverted results for both the near-offset and full-offset data inversion. However, when the far-offset information is included, it reduces the ambiguity of the inverse problem and improves the quantitative accuracy of the inverted structures, especially for deeper targets.

Chapter 5

The role of density in acoustic full waveform inversion

To reconstruct reliable models of the subsurface structures from the field measurements, the forward modelling should correctly account for wave propagation phenomena present in the recorded data. This mainly concerns the correct modelling of seismic amplitudes that are sensitive not only to the velocity variations, but also to the density, attenuation, source directivity effects, and to the seismic noise.

The objective of this study is to analyse the role of density on the recovery of P-wave velocity models in the marine environment. To investigate the footprint of density on FWI, I performed a series of numerical experiments, testing various initial density models and different strategies for the density update. The results show that it is important to include realistic density information into the inversion scheme. Furthermore, the more accurate a density model is, the better the velocity estimate will be. Additionally, I have investigated the importance of density in the inversion of noisy data, by considering acoustic inversion of elastic data.

5.1 Introduction

Full waveform inversion is a nonlinear inversion method to extract the information about the subsurface structure and physical parameters from seismic recordings. It is a least-squares data-fitting procedure; in order to find a model that fits the real seismograms the misfit between the real data and modelled data is minimized. Therefore, the accurate modelling of seismic waveforms is critical for the success of the method.

The least-squares norm, which is the most commonly used misfit function (Virieux and Operto, 2009), exploits both the kinematic (travel times) and the dynamic (amplitudes) information contained in the seismic data. The travel times are sensitive to seismic velocity distribution, whereas the amplitudes contain information about the reflection coefficients. The amplitude of reflected seismic waves is dominantly affected by contrasts in acoustic impedance, which is a product of velocity and density. Therefore, if the P-wave velocity is the only variable parameter in the FWI, the seismic amplitudes are not correctly modelled. For this reason, it would be useful to introduce density as an

additional parameter in the inversion of the field data, even at the expense of higher computational costs. Moreover, the density is a useful parameter for hydrocarbon characterization and the multi-parameter V_P and ρ inversion could provide additional information on the reservoir properties.

The full waveform inversion of marine seismic data is mainly performed using the acoustic approximation [Virieux and Operto, 2009]. Barner and Charara [2009] have shown that this approximation can lead to reliable results, when the S-wave velocity contrast and density contrast are not strong, or when only near offsets are used. Moreover, they have shown that the constant-density acoustic inversion leads to wrong impedance contrasts and produces worse inversion results. Since the density is a difficult parameter to reconstruct [Forgues and Lambaré, 1997], in most of the case studies of real marine reflection data the authors only invert for the P-wave velocity. Density is usually estimated using an empirical formula [Boonyasiriwat et al., 2010; Delescluse et al., 2011; Hicks and Pratt, 2001; Kelly et al., 2010; Shipp and Singh, 2002] or is fixed at a constant value [Bae et al., 2010; Operto et al., 2004].

In this study I investigate the effect of density on the reconstruction of P-wave velocity models from marine seismic data. I consider 2D acoustic time-domain inversion of a synthetic streamer data generated from heterogeneous P-wave velocity (V_P) and density (ρ) models. I run series of inversion tests assuming different initial density information and diverse strategies for density updates within the inversion scheme to analyse the density footprint on the waveform inversion. These strategies comprise: disregarding the density information, keeping an initial, realistic density model fixed, density model update at each iteration step using an empirical velocity-density relation and density inversion. To provide a quantitative estimate of the inversion results I compare the data misfit and measure the model error between the true model and the inverted velocity models.

In the second set of inversion tests, the waveform inversion is applied to elastic data. If the P-wave velocity is the only unknown in acoustic FWI, then all amplitude discrepancies resulting from seismic noise and nonacoustic factors (elastic effects, attenuation, anisotropy) will be projected into the V_P model. To mitigate this problem FWI could account for additional inversion parameters such as density. To investigate the potential benefits of the multi-parameter inversion in the presence of elastic effects in the data, I apply the V_P only and combined V_P and ρ inversion to synthetic elastic data. The goal of this numerical experiment is to show, whether the density inversion can partly compensate for the inversion artefacts due to the coherent noise present in the data.

5.2 Density inversion

The gradient for the P-wave velocity at iteration n can be expressed as

$$\delta V_{Pn} = 2\rho_n V_{P_n} \frac{1}{K_n^2} \sum_{\text{shots}} \int_t dt \, \frac{\partial p_n}{\partial t} \frac{\partial p_n'}{\partial t} \, , \tag{5.1}$$

where $K = \rho V_P^2$ is the bulk modulus, $p_n(x, z, t)$ is the forward propagated field in the current model, and $p_n^{'}(x, z, t)$ is generated by propagating the residual data from all receiver positions backward in time.

The gradient for the density $\delta\rho_{v_p}$ in the parameterization $\mathbf{m} = [V_P, \rho]$ can be written as

$$\delta\rho_{v_p} = V_{P_n}^2 \frac{1}{K_n^2} \sum_{\text{shots}} \int_t dt \, \frac{\partial p_n}{\partial t} \, \frac{\partial p_n^{'}}{\partial t} \; + \; \frac{1}{\rho_n^2} \sum_{\text{shots}} \int_t dt \, \nabla p_n \cdot \nabla p_n^{'} \, . \tag{5.2}$$

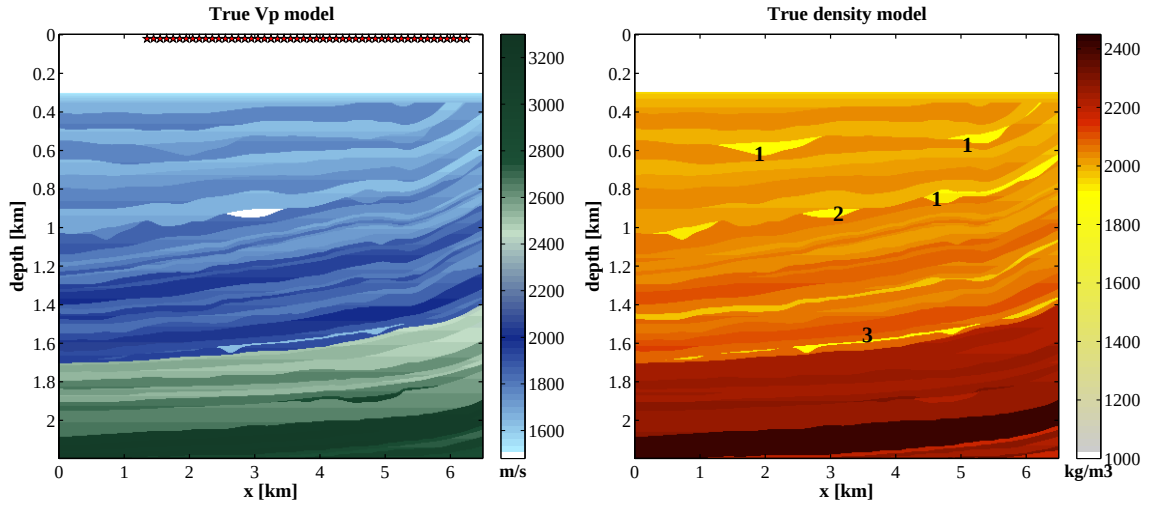
For the details of the gradient derivation, see Chapter 4.

Estimating density values from seismic data is an ill-posed inverse problem [Debski and Tarantola, 1995]. The density is poorly resolved especially from the short-offset P-wave velocity, because the amplitude diffraction pattern related to a perturbation of ρ is very similar to the diffraction pattern caused by a perturbation of V_P [Tarantola, 1986]. The influence of the density contrast on the P-wave reflection amplitude becomes more visible at larger angles. Therefore, wide-angle recordings might be required in order to reduce the uncertainty of the density inversion from seismic data. As shown by Amundsen and Ursin [1991], velocity model converges faster than density, because most of the velocity information comes from the arrival times of the seismic reflections, whereas the density is derived from the amplitude information only. The general velocity trend should be therefore correct before the density model can be recovered by fitting the amplitudes.

The full waveform inversion is a highly nonlinear problem. To avoid the convergence to local minima an accurate starting model is required. Initial density models can be provided from well log data, AVO analysis (amplitude variation with offset), or derived from the velocity model using empirical relationships between seismic velocity and density. Although the density logs provide the most accurate estimation of the bulk density, the information is available only along the borehole.

The density information can also be derived from the AVO analysis, which is based on the inversion of the Zoeppritz equations [Zoeppritz, 1919]. These equations describe the amplitude variation with offset as a function of P-wave velocity contrast, S-wave velocity contrast, and density contrast across the interface. However, the reliable density extraction from AVO is difficult [Behura et al., 2010]. It requires either a wide range of incidence angles [Li, 2005], or strong density contrast along the interface if only the near and middle offsets are available [Kabir et al., 2006].

The starting density model can be build using one of the empirical relationships between velocity and density. The drawback of this method is that all empirical relations are only valid for certain rock types.



(a) (b)

Figure 5.1: (a) The true P-wave velocity and (b) density models. The stars denote the shot point locations. **1** water wet sand, **2** a gas charged sand channel, **3** a thin oil sand layer.

5.3 Numerical experiment settings

5.3.1 Marmousi2 model

The numerical tests are based on a part of the Marmousi2 model [Martin et al., 2006]. The P-wave velocity and density models are shown in Figure 5.1. Sedimentary layers are primarily composed of shale. There are two hydrocarbon reservoirs embedded in the model: a gas charged sand channel (2 in Figure 5.1b) and a thin oil sand layer (3). In addition, there are some water wet sand structures (1).

Density values are assigned to each layer based on the lithology [Martin et al., 2006]. The empirical relationship used to relate density to P-wave velocity was proposed by Castagna et al. [1993] for various lithologies. Generally, sandstone is less dense than shales, therefore there are two different relations linking the V_P to the density for shale and for sand layers.

The $V_p - \rho$ relationship for sandstone is

$$\rho\ (\mathrm{g/cm^3}) = 0.2736 \cdot V_P^{0.261}\ , \tag{5.3}$$

whereas the $V_p - \rho$ relationship for shales is given by

$$\rho\ (\mathrm{g/cm^3}) = 0.2806 \cdot V_P^{0.265}\ . \tag{5.4}$$

5.3.2 Modelling and inversion parameters

The synthetic, noise-free data set is used in the first part of inversion tests and it was calculated using the finite-difference solution of the 2D acoustic wave equation. The total of 50 shot gathers

are generated at a 50 m interval with 3 seconds of data. The source is a pressure source, located 7.5 m below the air-water interface, with the Ricker wavelet time function. The frequency content of a signal is limited to a bandwidth from 3 to 20 Hz. The near offset is 100 m and the maximum offset is 4 km. The detailed modelling and acquisition parameters are listed in Table 3.1.

To reduce the high complexity of the inverse problem, the multi-scale inversion approach is applied. The frequency bands are applied sequentially with the following maximum frequencies $f_{\max} = (3, 4, 5.3, 7.1, 9.4; 12.5, 16.6, 20)$ Hz; that define the cutoff frequency for the Butterworth low-pass filter. Furthermore, to correct for the amplitude loss with depth due to geometrical spreading and to enhance deeper parts of the model, the linear gradient scaling with depth is implemented.

The starting V_P model is a 1D very smooth representation of the true velocity distribution. The depth of the seafloor in the starting model is correct, but the initial velocity value at the sea bottom differs from the true velocity. The seafloor parameters (V_P, density, depth) are updated during inversion. To allow for a direct comparison of the results, the same inversion scheme and the initial V_P model are used in all experiments. For the inversion of acoustic data, I use the least-squares minimization criterion.

5.3.3 Error analysis

In order to quantitatively assess the inversion results, I measure initial and final errors, both in the data and in the model space. Since the waveform inversion is an ill-posed and non-unique problem, this will help to determine whether the reduction of residuals is followed by a better estimation of model parameters.

The error between the modelled d_{mod} and observed data d_{obs} is measured by the least-squares error

$$L_{2\text{error}} = \frac{1}{2}\sum_{i=1}^{N}(d_{\text{mod}|i} - d_{\text{obs}|i})^2 \ . \tag{5.5}$$

where N is the size of the data vector.

To assess the quality of the final velocity models, the relative error between the real and the reconstructed P-velocities within the area of the model update is calculated as

$$V_{P\text{error}} = \frac{1}{M}\sum_{i=1}^{M}\frac{\mid V_{P\text{inv}|i} - V_{P\text{true}|i} \mid}{\mid V_{P\text{true}|i} \mid} \ , \tag{5.6}$$

where $V_{P\text{inv}|i}$ is the inverted P-wave velocity for the ith grid point, $V_{P\text{true}|i}$ is the true velocity value, and M is the size of the model vector within the area of the model update.

Test	starting density model	density model in FWI
A	true rho	fixed
B	constant rho	fixed
C	from V_P-ρ Brocher's relation (Eq.5.8)	fixed
D	from V_P-ρ Gardner's relation (Eq.5.7)	fixed
E	from V_P-ρ Brocher's relation	updated with Brocher's relation
F	from V_P-ρ Gardner's relation	updated with Gardner's relation
G	from V_P-ρ Brocher's relation	inverted
H	from V_P-ρ Gardner's relation	inverted

Table 5.1: List of inversion tests.

5.4 Inversion strategies

To investigate the effect of density on P-wave model reconstructions, I first applied acoustic waveform inversion to noise-free data. I performed eight inversion tests using different starting density models and strategies for incorporating density information into the inversion process. The inversion tests are listed in Table 5.1.

For the acoustic inversion of noise-free data four different starting density models were considered:

- true density model (Test A);
- homogeneous density model (Test B);
- in sediments: density model linked with the starting V_P model using Brocher's relationship, in water: 1000 kg/m^3 (Test C, E, G);
- in sediments: density model linked with the starting V_P model using Gardner's relationship, in water: 1000 kg/m^3 (Test D, F, H).

To find the most efficient strategy for including density information into the inversion scheme, I tested the following strategies:

- **FIXED** – starting density model is fixed during the inversion (Test A, B, C, D);
- **UPDATED** – density is updated after each iteration using one of the $V_p - \rho$ relationships (Test E, F);
- **INVERTED** – multi-parameter inversion for V_P and ρ (Test G, H).

The choice of the relationship between V_P and density controls the amplitude of reflections. It should be noted that these relationships are not valid in the presence of hydrocarbons. Gardner's $V_P - \rho$ relationship [Gardner et al., 1974] is an approximate average of the relations for a number of sedimentary rock types, weighted toward shales. This relationship gives relatively good density estimates for sediment layers in our true model (Figure 5.2). The relation takes the following form:

$$\rho \text{ (kg/m}^3) = 0.31 \cdot 1000 \cdot V_P^{0.25} \tag{5.7}$$

Additionally, I selected Brocher's density-velocity relationship [Brocher, 2005]. It is an empirical formula for the Nafe-Drake curve [Nafe and Drake, 1963], which is relating seismic velocity to density

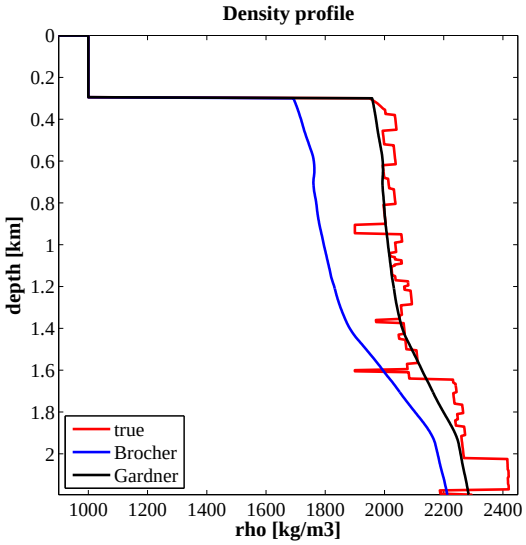


Figure 5.2: Comparison of the true density model (red line) with different initial density models. Blue line indicated a starting density model from V_P-ρ Brocher's relation, black line represnts a starting density model from V_P-ρ Gardner's relation. The density profile is located at $x = 3$ km.

in the sedimentary rock samples. The Brocher equation can be used in the V_P range from 1500 to 8500 m/s and it takes the following form:

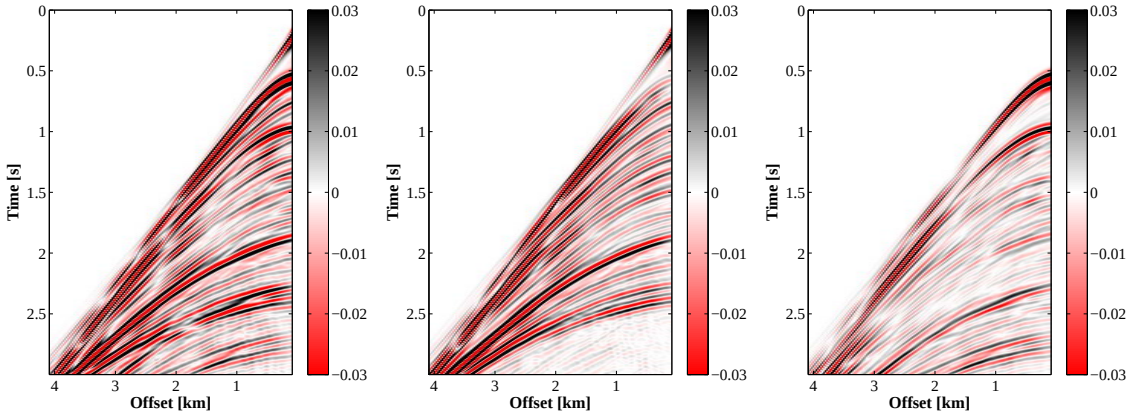
$$\rho\ (\mathrm{g/cm^3}) = 1.6612V_P - 0.4721V_P^2 + 0.0671V_P^3 - 0.0043V_P^4 + 0.000106V_P^5 \tag{5.8}$$

This relation systematically underestimates density with respect to the true values and provides a poor initial density model for the inversion tests (Figure 5.2).

5.5 Effect of density on the waveform

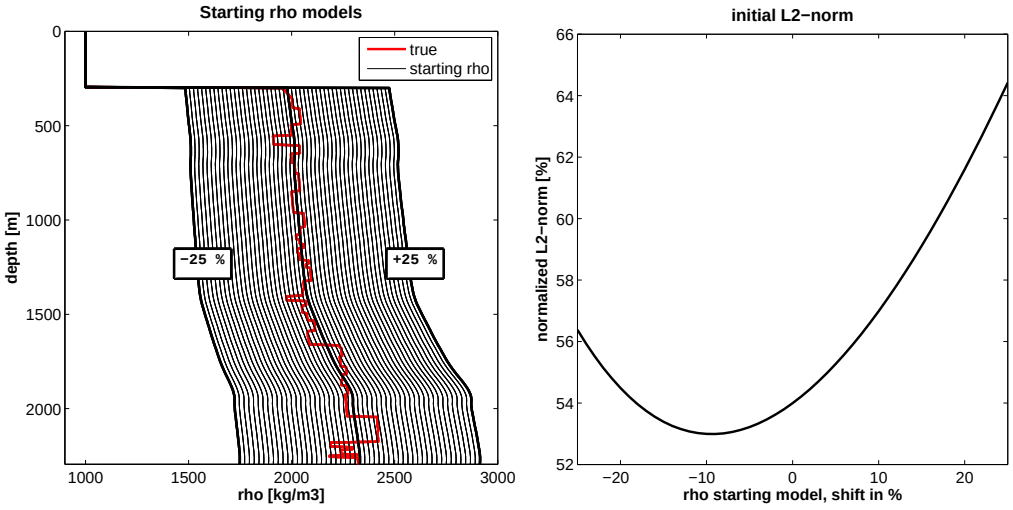
To analyse the effect of density on acoustic data, I compare the seismogram generated for the true V_P and true density models with the shot gather generated for the same true V_P but with constant density model. Figure 5.3c shows that a significant part of the waveform is due to the reflection of seismic energy at density contrasts. The reflection coefficient at the seafloor is dominated by the density contrast and thus most of the amplitudes of the primary and multiple water bottom reflections are caused by the strong density contrast. On the other hand, the reflection coefficients of the discontinuities inside the medium are influenced by both the P-wave and density contrasts.

In order to quantify the influence of the starting density model on the data misfit, I selected various initial density models (Figure 5.4a). The reference initial density is a 1D smoothed true density model. It was varied by adding a maximum $\pm$ 25 % density perturbations along the entire profile, disregarding the water layer, to create a set of 51 initial density models. Then the data misfit between the observed seismograms is calculated (for the true model shown in Figure 5.1) and the initial data computed for different starting density models. The resulting L_2-norms vary from 53 % to 64 % of the observed data energy (Figure 5.4b). The lowest misfit value corresponds to the density model, which together with the initial V_P, gives the closest match to the real amplitude of the primary and secondary seafloor reflections, which are the most energetic events in the initial synthetics.



(a) (b) (c)

Figure 5.3: Shot gathers for (a) the true V_P and true density models shown in Figure 5.1, (b) the true V_P model but constant density model, (c) the difference between (a) and (b). Seismograms correspond to shot location at $x = 6.25$ km.



(a) (b)

Figure 5.4: Influence of the starting density model on the initial data misfit. (a) Set of starting density models generated by adding a maximum ± 25 % density perturbations with a 1 percent shift to the smoothed 1D true density model. The true density model at $x = 3000$ m is indicated by the red line. (b) Initial L_2-norm calculated for all seismograms that corresponds to selected starting density models.

5.6 Inversion results - noise-free acoustic data

Inversion results of the noise-free data are summarized in Table 5.2 (data misfit) and Table 5.3 (V_P model error) and can be found in Figures 5.5 - 5.7.

5.6.1 Fixed starting density model

Figure 5.5 shows the inversion results when using fixed density models, i.e. the density models remain unchanged during the inversion. To examine the resolution and accuracy of the FWI, I performed an inversion test with the true density model (Figure 5.5a, Test A). The final velocity model is very well reconstructed, the V_P error was reduced from 4.91 % to 1.92 % (Table 5.3). After 400 iterations, the final data residuals are very small (Figure 5.9a).

The highest misfit values both in the data space and in the model space ($V_{P\,\text{error}} = 3.55$ %) result from the constant density assumption (Test B). The final velocity model (Figure 5.5b) shows strong artefacts around the seafloor and deviations from the true velocity model in deeper regions. The lack of density information produced significant amplitude errors in the modelling of the seafloor reflections. In order to compensate for these errors, the inverted P-wave velocities at the sea bottom are significantly overestimated. Due to the wrong velocity in the uppermost part of the model, the velocity reconstruction in the subseafloor area is negatively affected as well.

If we include more accurate fixed density models using an empirical velocity-density formula, we can observe an improvement in the recovery of the P-wave model (Figure 5.5c, d). The V_P error was reduced to 2.77 % for the poor density model, and to 2.59 % for the good density model. Even if the density model is poor it contains information about an impedance contrast at the seafloor (Figure 5.5c). Therefore the reconstructed velocity model is more accurate than with the constant density assumption. Thus the most important information is the density at the seafloor, as the data misfit is dominated by the seabed reflection and its multiples.

5.6.2 Density update using P-wave velocity and density relations

The inversion results with the density update at each iteration step are shown in Figure 5.6a-f for Brocher's $V_P - \rho$ relationship (Test E), and in Figure 5.6g-l for Gardner's relationship (Test F).

The density update strategy generally provides a slightly better velocity reconstruction than the fixed density approach (Figure 5.8a,b). In the fixed density strategy, the smooth density models produce lower impedance contrast for the P-wave. To compensate for weaker amplitudes, the recovered velocities are overestimated. On the other hand, the updated density models have stronger contrasts and result in higher reflection coefficients. Therefore, the corresponding amplitude errors are smaller and yield a slightly better reconstruction of the velocity model (compare velocity profiles in Figure 5.5d and Figure 5.6f).

5.6.3 Multi-parameter inversion for P-wave velocity and density

The multi-parameter inversion results are shown in Figure 5.7a-f – for the poor initial density model (Test G), and in Figure 5.7g-l – for the good initial density model (Test H). The V_P models are well reconstructed, but density values of individual layers differ from the true density model, even if

the good initial density is used. If the very poor starting model is assumed, the waveform inversion cannot recover correct background densities. However, the location of interfaces and the density contrast across the interfaces are fairly well estimated. The final model errors ($V_{P\text{error}} = 2.32$ % for poor initial ρ, and $V_{P\text{error}} = 2.28$ % for good initial ρ). A combination of the good initial density model with the simultaneous inversion for velocity and density produced the best final result (apart from using the true density model). The minimum model error is also reflected in very small data residuals. All reflection events are well fitted, and the data misfit is significantly reduced (Table 5.2).

The hydrocarbon layers, which are characterized by a high impedance contrast, are well reconstructed both in the V_P and in the density model. There is one essential difference between the FWI results from the multi-parameter inversion strategy and the density update strategy. The water wet sand structures marked in the true density model with number **1**, are very well visible in the inverted density model (Figure 5.7j). But on the other hand, these structures are missing, when the density update strategy is used (Figure 5.6j).

This result is due to the fact that there are two different $V_p - \rho$ relations used for shale layers and sand layers in the true model. But as there is only one empirical relation used in the update strategy, it is obviously not valid for all lithologies assigned to the true density model.

5.6.4 Model and data error

Figure 5.8 illustrates the evolution of the V_P model error for all inversion experiments. The plots are subdivided into categories according to the strategy for including density information into the inversion scheme. The highest error resulting from using a constant density model, is significantly reduced when a realistic density information is included. The initial V_P model error is reduced by 43 % for a poor density model and by 47 % for a good density model. There is also a small improvement in the accuracy of inverted velocity models, when a density model is updated after each iteration step instead of keeping a fixed initial ρ model (Figure 5.8a,b). The initial V_P model error is reduced by 46 % (poor density model) and by 49 % (good density model), which gives an additional model error reduction of approximately 3 % with respect to the fixed strategy. The multi-parameter inversion (Figure 5.8c) results in the lowest model error. The initial V_P error is reduced by 52 % for a poor initial density model and by 53.5 % for a good initial density model. Furthermore, for the multi-parameter inversion strategy the choice of the starting density model has a smaller effect on the final velocity model error than in the case of using the fixed or updated strategy.

The final data residuals are shown in Figure 5.9. According to expectations, the highest data residuals are generated when a constant density model is used (Figure 5.9b). Because the reflection coefficient at the seafloor is primarily governed by a high density contrast and this information is not available, the main data misfit is observed for the seafloor reflection. The best reduction of the data residuals is achieved for the multi-parameter inversion strategy (Figure 5.9g,h). Especially, when a good initial density model is combined with the multi-parameter inversion this results in the best V_P error reduction and in the lowest data misfit.

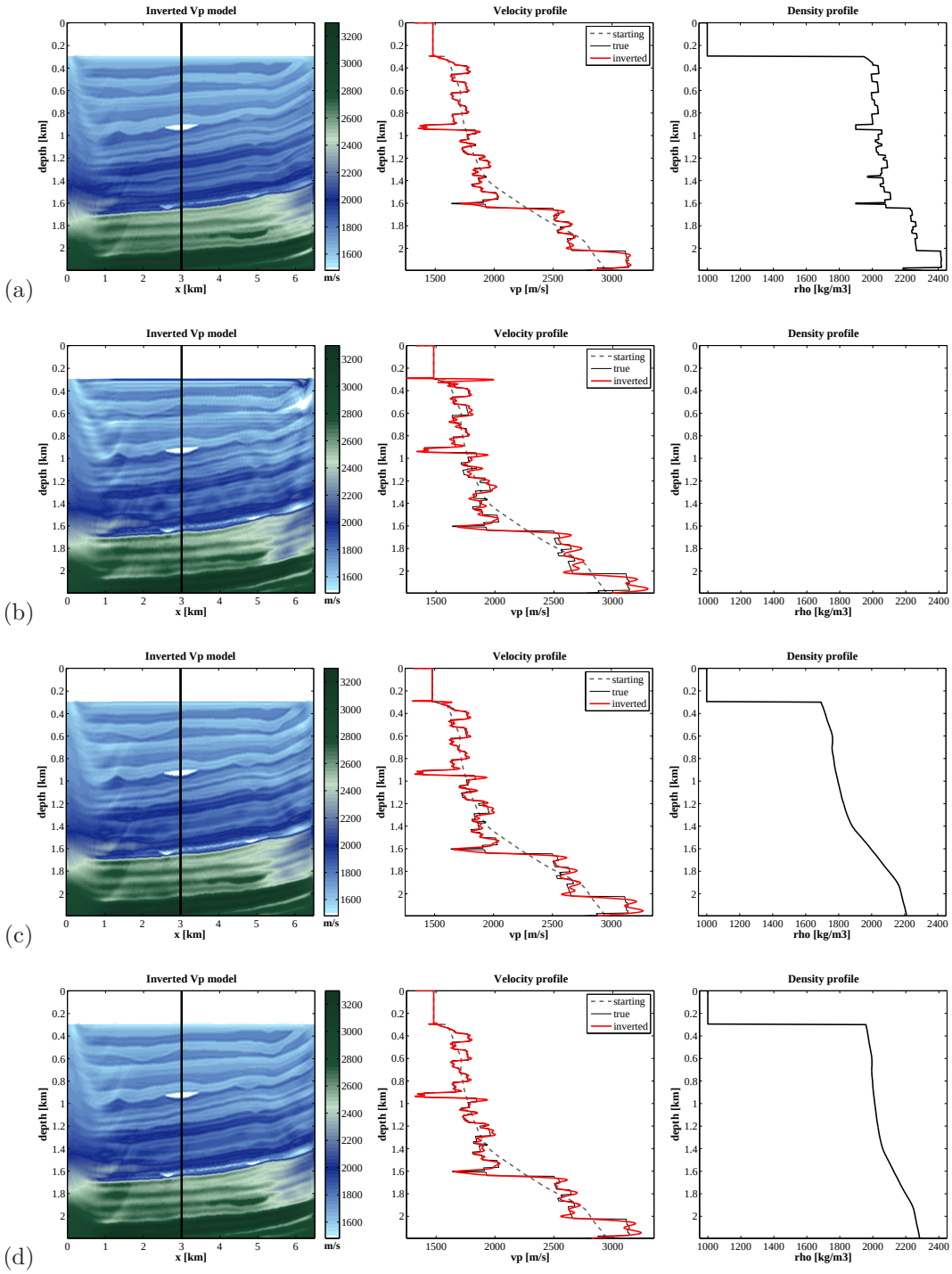
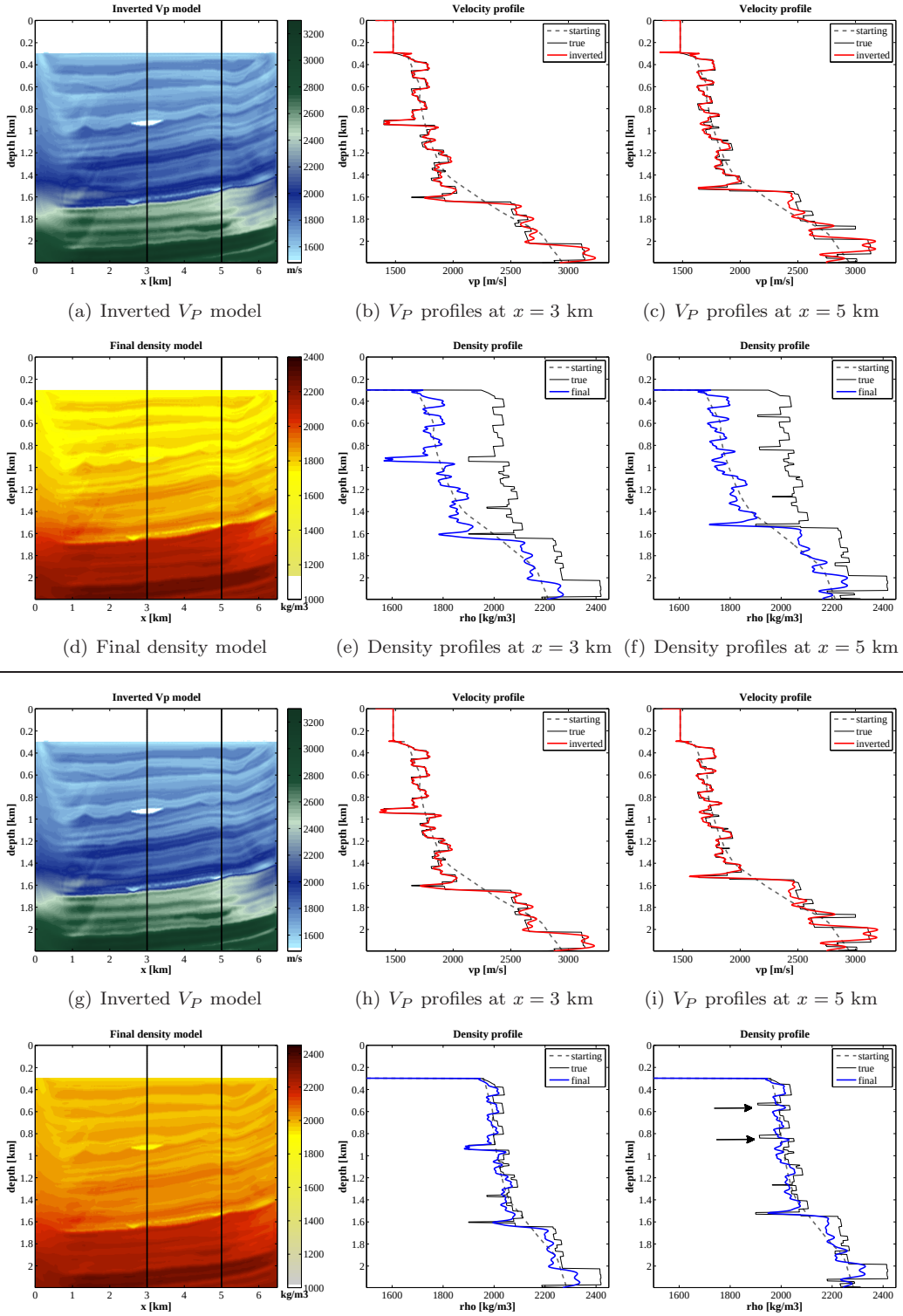


Figure 5.5: FWI results of acoustic data; fixed density models. Inverted V_P models (left); V_P profiles (center): the black solid line indicates the true model, the gray dashed line indicates the initial model, and the red solid line indicates the inverted V_P model. Density profiles of the fixed models (right). (a) Test A – true density model, (b) Test B – constant density model, (c) Test C - poor 1D smooth density model, (d) Test D - good 1D smooth density model.



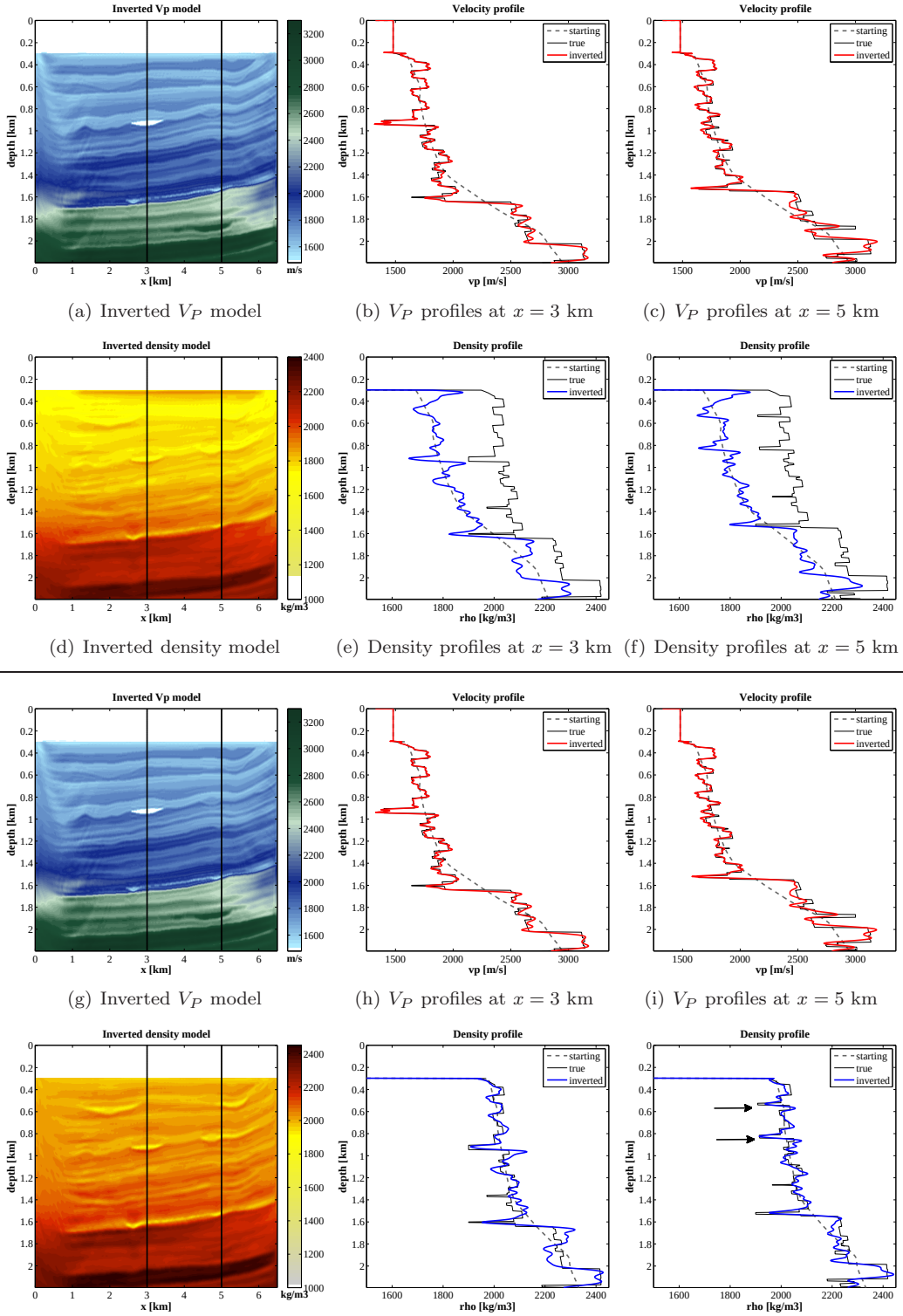
(a) Inverted V_P model (b) V_P profiles at $x = 3$ km (c) V_P profiles at $x = 5$ km

(d) Final density model (e) Density profiles at $x = 3$ km (f) Density profiles at $x = 5$ km

(g) Inverted V_P model (h) V_P profiles at $x = 3$ km (i) V_P profiles at $x = 5$ km

(j) Final density model (k) Density profiles at $x = 3$ km (l) Density profiles at $x = 5$ km

Figure 5.6: FWI results of acoustic data; comparison of different empirical relations for density update at each iteration. (a)-(f) Test E – poor initial density; Brocher's $V_P - \rho$ relationship, (g)-(l) Test F – good initial density, Gardner's $V_P - \rho$ relationship.



(a) Inverted V_P model (b) V_P profiles at $x = 3$ km (c) V_P profiles at $x = 5$ km

(d) Inverted density model (e) Density profiles at $x = 3$ km (f) Density profiles at $x = 5$ km

(g) Inverted V_P model (h) V_P profiles at $x = 3$ km (i) V_P profiles at $x = 5$ km

(j) Inverted density model (k) Density profiles at $x = 3$ km (l) Density profiles at $x = 5$ km

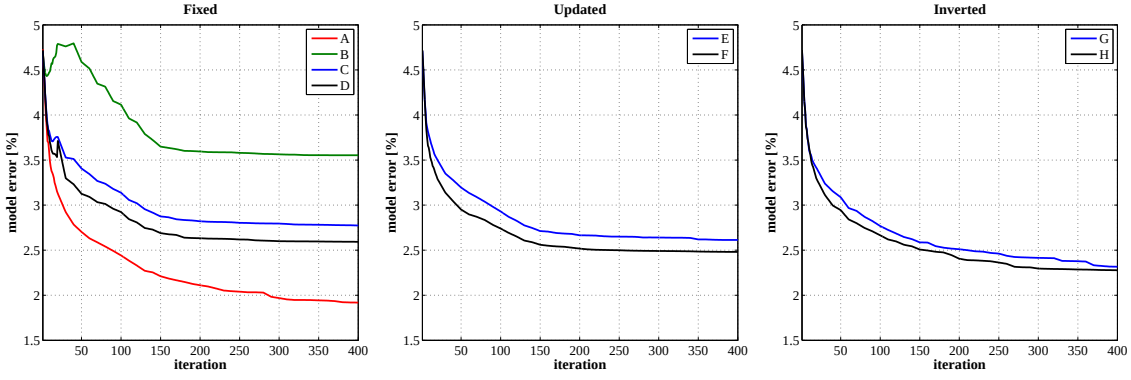
Figure 5.7: FWI results of acoustic data; multi-parameter inversion with different starting density models. (a)-(f) Test G – multi-parameter inversion with a poor initial density model; (g)-(l) Test H – multi-parameter inversion with a good initial density model.

	density update strategy		
starting rho	fixed	updated	inverted
true rho model	3.03e+03 (A)	-	-
constant rho	3.56e+06 (B)	-	-
from Brocher	2.56e+05 (C)	2.41e+05 (E)	1.73e+04 (G)
from Gardner	8.33e+04 (D)	5.27e+04 (F)	6.39e+03 (H)

Table 5.2: FWI results of acoustic data. L_2-norm (equation 5.5) of the final data residuals. Capital letters refer to inversion tests listed in Table 5.1.

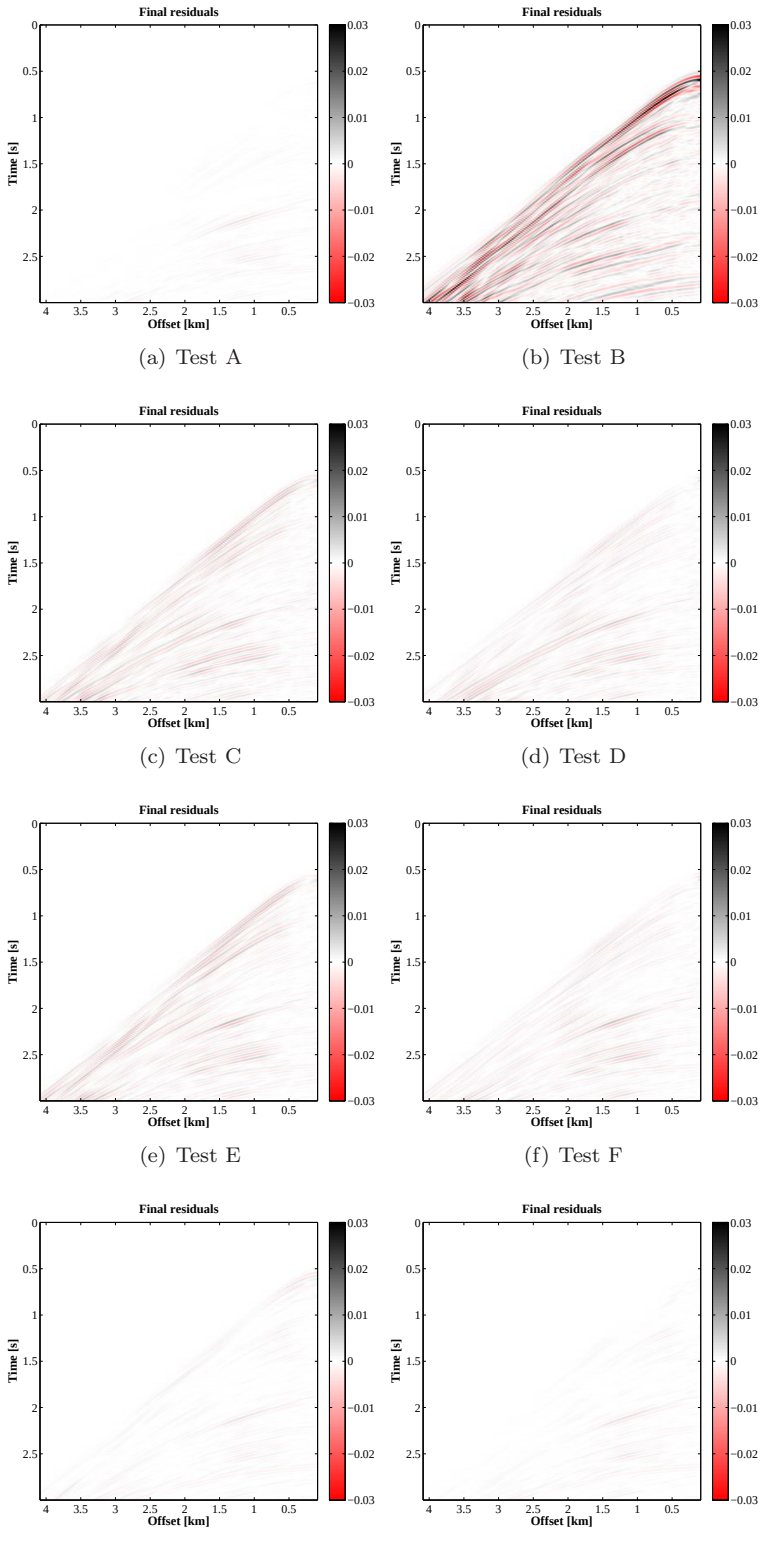
	density update strategy		
starting rho	fixed	updated	inverted
true rho model	1.92 % (A)	-	-
constant rho	3.55 % (B)	-	-
from Brocher	2.77 % (C)	2.61 % (E)	2.32 % (G)
from Gardner	2.59 % (D)	2.48 % (F)	2.28 % (H)

Table 5.3: FWI results of acoustic data. V_P error of the inverted models (equation 5.6). The V_P error of the initial model is 4.91 %. Capital letters refer to inversion tests listed in Table 5.1.



(a) ρ fixed (b) ρ updated (c) ρ inverted

Figure 5.8: FWI results of acoustic data – summary of final V_P model errors for all inversion tests. Red line – true ρ model; green line – constant density model; blue line – poor density model (Brocher); black line – good density model (Gardner). Capital letters refer to inversion tests listed in Table 5.1.



(a) Test A (b) Test B

(c) Test C (d) Test D

(e) Test E (f) Test F

(g) Test G (h) Test H

Figure 5.9: FWI results of acoustic data; final data residuals. (a) Test A – true density model, (b) Test B – constant density model, (c) Test C - poor 1D smooth density model, (d) Test D - good 1D smooth density model, (e) Test E – poor initial density, Brocher's $V_P - \rho$ relationship, (f) Test F – good initial density, Gardner's $V_P - \rho$ relationship, (g) Test G – multi-parameter inversion with a poor initial density model, (h) Test H – multi-parameter inversion with a good initial density model.

5.7 Inversion of elastic data

In the second set of inversion tests, the acoustic inversion is applied to elastic data. The aim is to investigate the performance of different inversion strategies described in this study in the presence of noise associated with elastic wave propagation. I want to verify if the multi-parameter inversion of elastic data offers the same benefits as for the pure acoustic data inversion.

The elastic data set is generated for the part of the elastic Marmousi2 model (for details see Chapter 3). I use the same inversion scheme and the initial V_P model as for the inversion of acoustic data. The only difference is related to the choice of the minimization criterion. Here I apply the L_2 norm using the normalized wavefields, L_2norm, instead of the L_2 norm, because the first one provided the best results for the elastic data inversion (see Chapter 3).

Figure 5.10 shows results for the V_P only inversion using different fixed density information (true, constant, poor density model, good density model). Apart from some artefacts, the velocity models are well reconstructed down to a depth of 1.6 km, but the resolution decreases below that depth. This is related to the interface, located at 1.6 km, with a strong impedance contrast for both P and S-waves, which results in converted S-waves. Even when the true density model is used, there are still some data residuals related to the seafloor reflection (Figure 5.12a). This is due to the fact that the acoustic inversion cannot reproduce the AVO effect (amplitude-verus-offset) of elastic data.

The density update at each iteration step using Gardner's relationship is shown in Figure 5.11a-f. The quality of the inverted model is comparable with the results obtained with the fixed strategy. In contrast to the inversion of acoustic data, the worst result is provided by the multi-parameter inversion strategy (Figure 5.11g-l). The velocity model, as well as the inverted density model, contains a lot of small-scale artefacts. The density values are wrong and significantly overestimated, especially in the deeper parts of the model. Apparently, density tries to compensate for the amplitude and phase errors caused by elastic effects. But since there is a trade-off between velocity and density reconstruction (see Chapter 4), the accuracy of the inverted velocity model is negatively affected by the wrong density reconstruction. The poor recovery of density structures is not related to the selection of the L_2 norm using the normalized wavefields as a misfit function. An identical inversion test, performed with the standard L_2 norm misfit function, resulted in much worse quality of reconstructed models.

Figure 5.13 shows the evolution of the V_P model error for the elastic data inversion. Using a homogeneous density model in the inversion of elastic data produced the highest V_P error. On the other hand, the accuracy of the reconstructed velocity models is almost identical for the rest of other density models, which are kept fixed during inversion (true, poor and good density models). There is also no improvement, when the density is updated at each iteration step using Gardner's relationship (compare Figure 5.13a,b). In contrast to the results of the acoustic data inversion, the multi-parameter inversion for V_P and density results in a very high model error, which is caused by the wrong reconstruction of density structures.

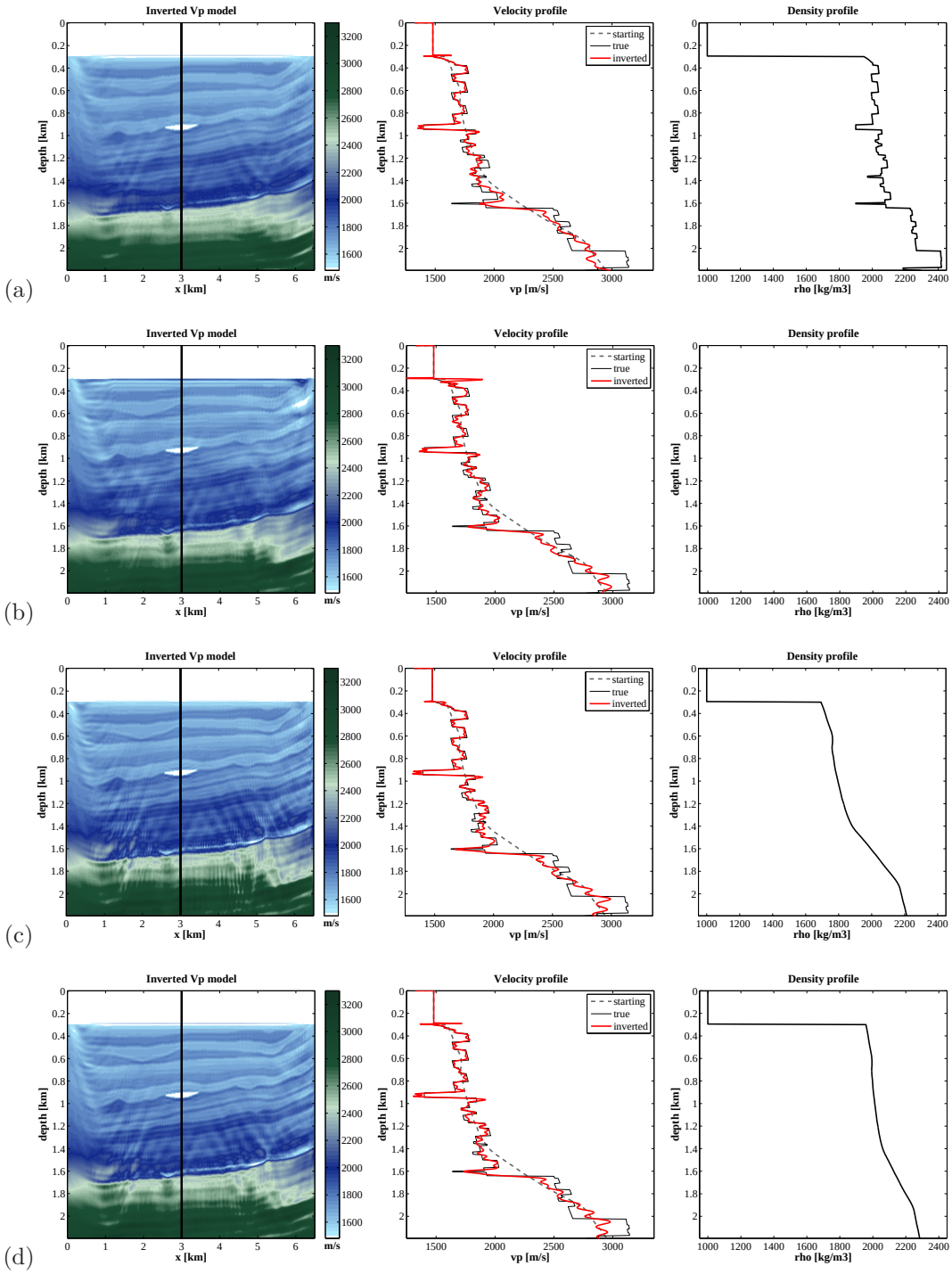
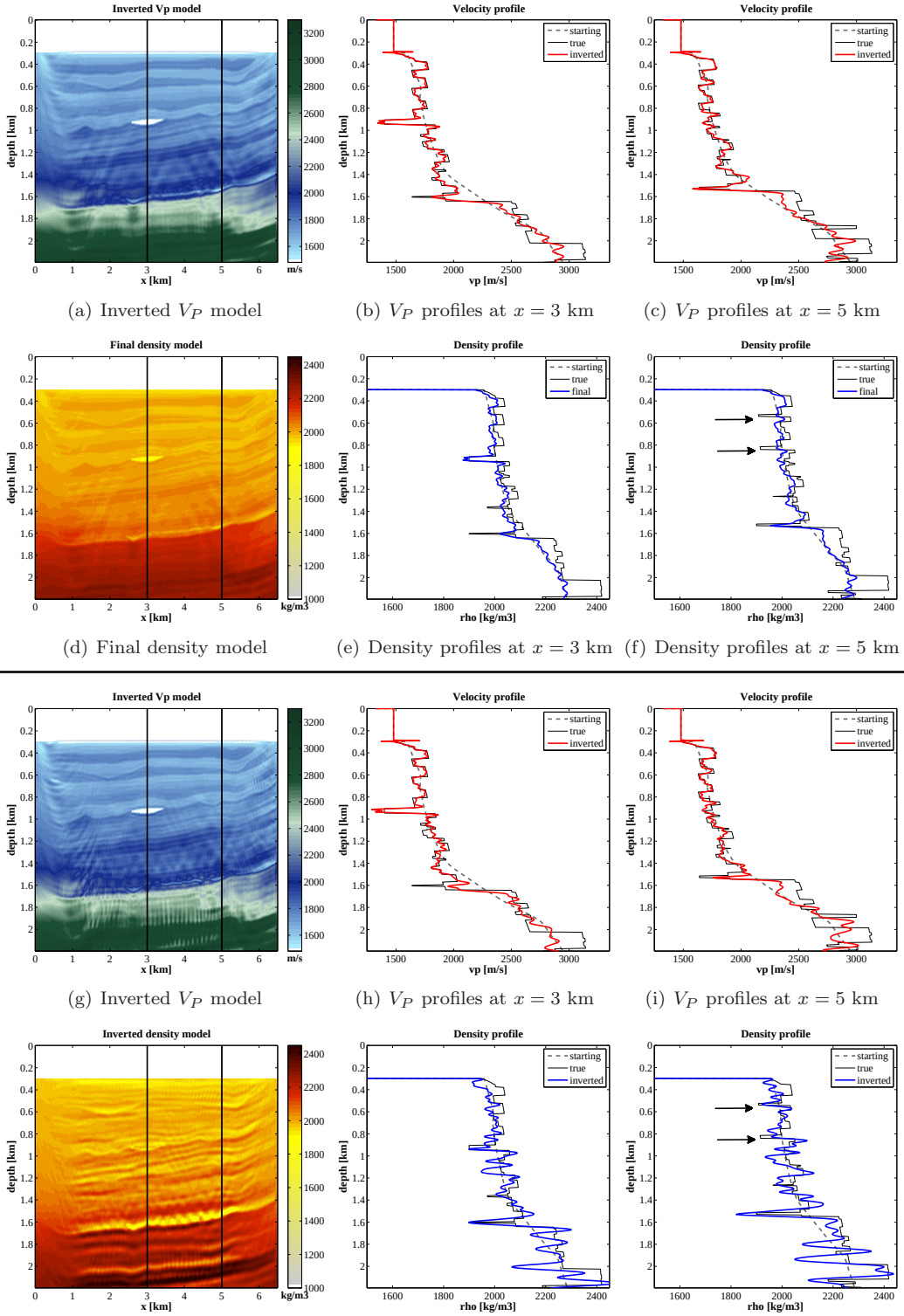


Figure 5.10: FWI results of elastic data – fixed density models. Inverted V_P models (left); V_P profiles (center): the black solid line indicates the true model, the gray dashed line indicates the initial model, and the red solid line indicates the inverted V_P model. Density profiles of the fixed models (right). (a) Test A – true density model, (b) Test B – constant density model, (c) Test C - poor 1D smooth density model, (d) Test D - good 1D smooth density model.



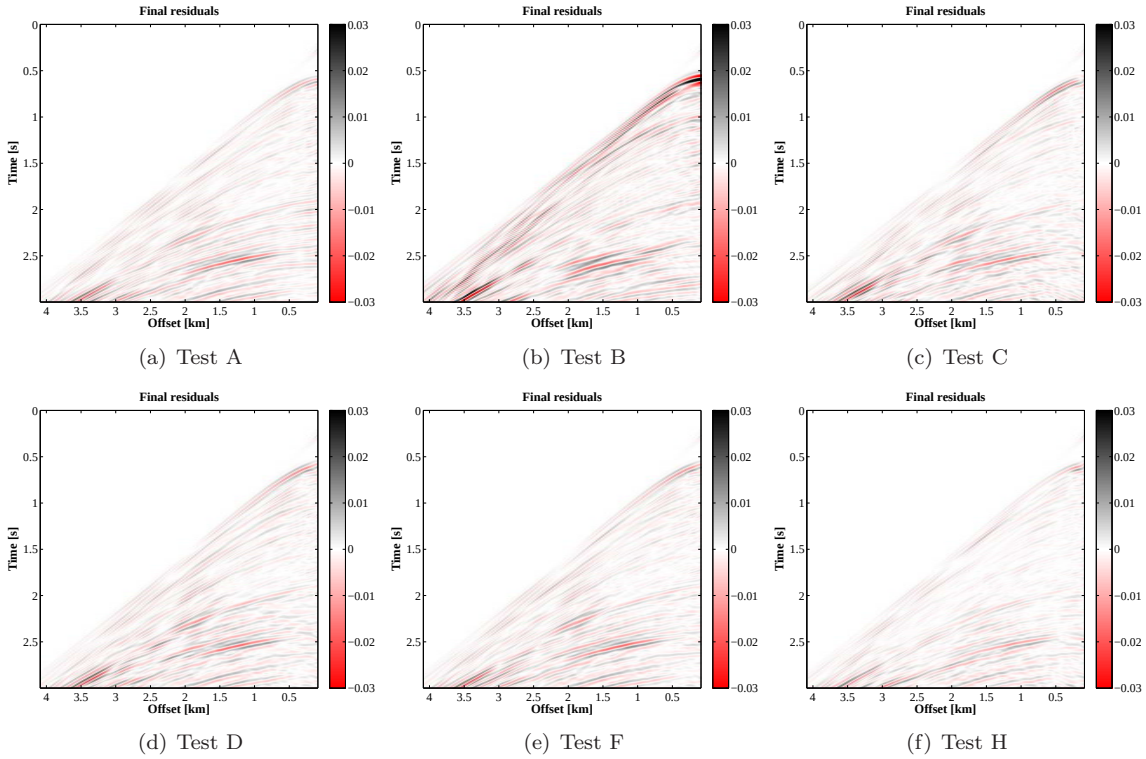
(a) Inverted V_P model (b) V_P profiles at $x = 3$ km (c) V_P profiles at $x = 5$ km

(d) Final density model (e) Density profiles at $x = 3$ km (f) Density profiles at $x = 5$ km

(g) Inverted V_P model (h) V_P profiles at $x = 3$ km (i) V_P profiles at $x = 5$ km

(j) Final density model (k) Density profiles at $x = 3$ km (l) Density profiles at $x = 5$ km

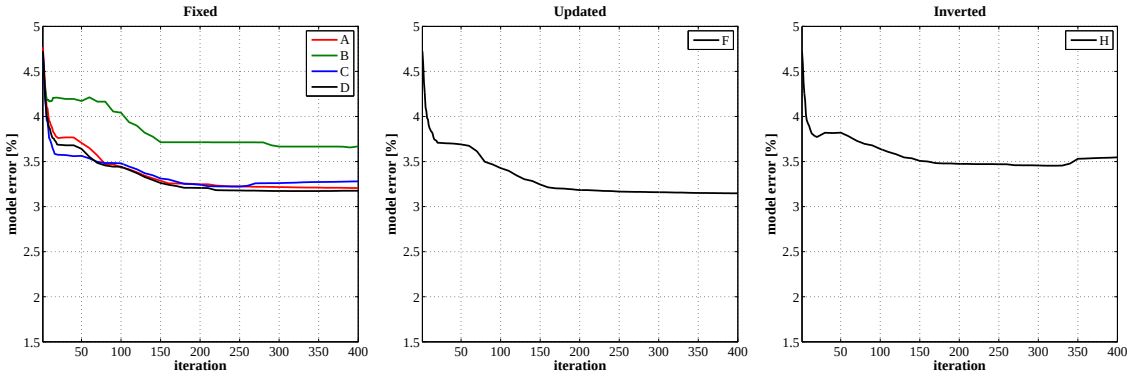
Figure 5.11: FWI results of elastic data. (a)-(f) Test F – density update at each iteration, good initial density, Gardner's $V_P - \rho$ relation. (g)-(l) Test H – multi-parameter inversion with a good initial density model.



(a) Test A (b) Test B (c) Test C

(d) Test D (e) Test F (f) Test H

Figure 5.12: FWI results of elastic data – final data residuals. (a) Test A – true density model, (b) Test B – constant density model, (c) Test C - poor 1D smooth density model, (d) Test D - good 1D smooth density model, (e) Test F – good initial density, Gardner's $V_P - \rho$ relationship, (f) Test H – multi-parameter inversion with a good initial density model.



(a) ρ fixed (b) ρ updated (c) ρ inverted

Figure 5.13: FWI results of elastic data – summary of final V_P model errors. Red line – true ρ model; green line – constant density model; blue line – poor density model (Brocher); black line – good density model (Gardner). Capital letters refer to inversion tests listed in Table 5.1.

5.8 Conclusions

The main goal of this study was to investigate the role of density on the recovery of velocity models. To do so, I generated both the acoustic and the elastic data set for the Marmousi2 model with heterogeneous density distribution and applied a 2D acoustic time-domain FWI.

The inversion results show that the realistic density information should be directly included in the inversion process to improve the accuracy of the velocity reconstruction. This is particularly important, when we take into account both amplitude and phase information of the data. The most significant is the information about density at the seafloor, because the reflection coefficient at this interface is dominated by the strong density contrast. Using a constant density assumption produces artefacts in the recovered velocity models, because then all reflections are interpreted in terms of velocity contrasts only.

Since the more accurate density models improve the V_P estimation, it would be preferable to allow for density inversion, rather than to use a fixed relationship between velocity and density. This is particularly true in light of the fact that all empirical relations are valid only for certain rock types. In the case of the noise-free acoustic data inversion, the multi-parameter V_P and ρ inversion provided an additional information on the lithology of the model. However, a reliable inversion of density turned out to be very difficult with the elastic data. The density inversion is very sensitive to amplitude errors resulting from the acoustic approximation. As a consequence, the incorrect density values negatively affect the velocity reconstruction, because the V_P model compensates the wrong density structures.

Chapter 6

Some practical aspects

Waveform inversion using synthetic data sets offers a number of benefits. In particular, it provides full control over the simulation of seismic data. Usually, the same forward modelling code is used to generate and to invert observed data. It means that the noise due to measurement errors, modelling inaccuracy or approximations in the physics of wave propagation is not considered. On the other hand, there is a number of practical considerations that should be taken into account when inverting field data.

This chapter investigates several aspects related to the application of the 2D FWI to real data. Here, I address the following issues:

- 3D to 2D transformation
- Source wavelet estimation.
- Acquisition effects

6.1 3D to 2D transformation

The seismic waves propagating through the Earth have a three-dimensional characteristic, whereas the inversion algorithm is based on the 2D forward modelling. In three dimensions, seismic waves are excited by a point source and the wavefronts are spherical in a uniform medium. On the other hand, 2D modelling assumes that sources and receivers are constrained to a single straight acquisition line and the seismic wavefield is generated by a line source. A seismic line source is extending parallel to the y-axis, i.e. in the out of plane direction, and it generates cylindrical waves [Chapman, 2004]. The cylindrical solution can be considered in three-dimensions as the superposition of point sources along the y-line. It is obvious that the spherical spreading of 3D data does not correspond with cylindrical spreading implied by the 2D wave equation. The difference in the nature of the source results in a different amplitude decay with offset and in a phase misfit. It means that the field data (point source) cannot be directly compared with the modelled data (line source) and it is necessary to transform field seismograms such that they reflect wave propagation in a 2D medium.

Several authors have investigated the problem of the line source assumption in the 2D modelling. The approach of Song and Williamson [1995] is based on the application of a Fourier transform in the

out-of plane direction, so that the 3D simulation is reduced to multiple 2D problems. The method assumes a 2.5D configuration, i.e. a medium varies only in two dimensions but the source is a 3D point source. It is computationally less intensive than 3D modelling, but more demanding than 2D forward modelling. This approach was used to perform a 2.5D frequency-domain acoustic waveform inversion of crosshole data [Song et al., 1995]. Another group of methods aims to derive filters for 3D to 2D transformation, which are generally formulated as pairs of integral equations [Amundsen, 1993; Amundsen and Reitan, 1994; Roberts, 2005; Wapenaar et al., 1992]. The underlying assumption is that the model is horizontally layered. Williamson and Pratt [1995] gives an overview of filters derived from asymptotic ray theory that can be applied to convert acoustic data from 2.5D configurations to 2D data [e.g. Bleistein, 1986; Deregowski and Brown, 1983]. The drawback of these methods is that they provide a very approximate correction and they are not valid in the presence of large velocity variations.

6.1.1 Filter for the 3D to 2D transformation

In this study, I consider a 3D to 2D transformation derived from Green's functions that allows for the conversion of point source data to line source data [e.g. Wang and Rao, 2009; Williamson and Pratt, 1995].

In a homogeneous unbounded medium, the frequency-domain Green's function in 3D is given by [Ikelle and Amundsen, 2005]

$$G_{3D}(\mathbf{r}, \mathbf{r}_s, \omega) = \frac{1}{4\pi R} \exp\left[\frac{\mathrm{i}\omega R}{c}\right] , \tag{6.1}$$

where $R = |\boldsymbol{R}| = |\mathbf{r} - \mathbf{r}_s|$ is the distance from the source located at $\mathbf{r} = \mathbf{r}_s$, and c is the acoustic velocity.

The 2D Green's function G_{2D} can be obtained from the 3D Green's function by regarding G_{2D} as the signal in three dimensions from an infinitely long line source along the y-axis, i.e. G_{2D} is the sum of the signals from all the 3D point sources constituting the line source

$$G_{2D}(\mathbf{r}, \mathbf{r}_s, \omega) = \frac{1}{4}(H)_0^{(1)}(kR) , \tag{6.2}$$

where $(H)_0^{(1)}(kR)$ is the zero-order Hankel function of the first kind, and $k = \omega/c$ is the wavenumber. In the far-field, this can be approximately evaluated as

$$G_{2D}(\mathbf{r}, \mathbf{r}_s, \omega) \approx \frac{1}{4}\sqrt{\frac{2}{\pi k R}} \exp\left[i(kR - \pi/4)\right] = \sqrt{\frac{-ic}{8\pi\omega R}} \exp\left[\frac{\mathrm{i}\omega R}{c}\right] . \tag{6.3}$$

3 Now, we want to find a filter F from 3D to 2D such that

$$G_{2D}(\mathbf{r}, \mathbf{r}_s, \omega) = F G_{3D}(\mathbf{r}, \mathbf{r}_s, \omega) \tag{6.4}$$

We obtain

$$F \approx \sqrt{\frac{2\pi R c}{i\omega}} = D_{-1/2}(t)\sqrt{2\pi R c} , \tag{6.5}$$

where $D_{-1/2}(t) = \mathcal{F}^{-1}[(i\omega)^{-1/2}]$ is the causal half-integrator [Deregowski and Brown, 1983], which

can be achieved in the time domain by convolution with $1/\sqrt{t}$. When we assume that $R \propto ct$, we get

$$F \propto D_{-1/2}(t)\ c\sqrt{2\pi t}\ . \tag{6.6}$$

This simple correction assumes a constant velocity background. The first term accounts for the phase misfit and the second term is a time-dependent amplitude spreading correction. The amplitude correction term is often reduced to $\sqrt{t}$ and we obtain

$$C_1 \propto D_{-1/2}(t)\sqrt{t}\ . \tag{6.7}$$

This correction, or just $\sqrt{t}$, is widely used as a part of the field data preprocessing for the 2D full waveform inversion [e.g. Crase et al., 1990; Hicks and Pratt, 2001; Operto et al., 2006; Shipp and Singh, 2002; Zhou et al., 1995].

A constant acoustic velocity can be replaced with the time-dependent RMS velocity V_{rms} (the root-mean-square velocity), so that the medium is assumed to be horizontally layered [Wang and Rao, 2009]

$$C_2 \propto D_{-1/2}(t)V_{\mathrm{rms}}\sqrt{t}\ . \tag{6.8}$$

In the next section, I investigate the accuracy of these two spreading corrections, C_1 and C_2, and their effect on the inversion results.

6.1.2 2D inversion of 3D data - numerical example

To test the performance of the described 3D to 2D transformations, I generated both the point source data (3D) and the line source data (2D). The point source data are modelled with the 3D acoustic forward modelling code [Bohlen, 2002]. The acquisition geometry is identical for both models, i.e. the sources and receivers are restricted to the plane. The source function is a Ricker wavelet with a dominant frequency of 10 Hz.

It is assumed that the subsurface is invariant in the off-line direction (y-axis). To create such a 2.5D model, the 2D Marmousi2 model (Figure 6.1c) is duplicated along the y-axis (Figure 6.1a). It means that the 3D wavefields are recorded in 2.5D configuration and they are not influenced by out-of-plane 3D structures.

An exemplary 3D seismogram and 2D seismogram are shown in Figure 6.1b,d, respectively. We can clearly observe a stronger amplitude decay with time and distance of the 3D data with comparison to the 2D data, which is caused by the different geometric spreading of the wavefronts.

C_2 transformation requires a computation of the RMS velocity function. The RMS velocity is the square root of the average squared velocity. For horizontal, isotropic layers and small offsets, the RMS velocity equals the stacking velocity. For a 1D medium with n layers, V_{rms} is given by [Ikelle and Amundsen, 2005]

$$V_{\mathrm{rms}} = \sqrt{\frac{\sum_{i=1}^{n} V_i^2 \Delta t_i}{\sum_{i=1}^{n} \Delta t_i}}\ , \tag{6.9}$$

where

$$\Delta t_i = \frac{h_i}{V_i}\ , \tag{6.10}$$

where V_i is the interval velocity of the ith layer and h_i represents the thickness of the ith layer.

In this example, the RMS velocity is calculated from the initial V_P model (Figure 6.2a). Thickness of every layer equals the grid spacing, such that $h_i = 5$ m, and the total number of layers $n = 470$, which corresponds to the number of grid points in depth. The RMS velocity function is displayed in Figure 6.2b.

Figure 6.3a,b shows the result of the transformation from 3D to 2D for both corrections. The seismograms are normalized to the maximum of the seafloor reflection at the near offset. The difference between the 3D data after correction and the 2D data is presented in Figure 6.3c,d. Both transformation methods yield satisfactory results. However, we can observe that the simple C_1 transformation, that assumes a constant velocity, results in a higher amplitude misfit than the C_2 transformation. In both cases, there is a significant misfit of the direct wave. This is likely due to the far-field approximation of the 2D Green's function. A more detailed waveform comparison reveals that amplitudes after C_1 transformation are underestimated with respect to the 2D data (Figure 6.3e). On the other hand, the C_2 transformation results in the overestimation of amplitudes (Figure 6.3f).

In Figure 6.4 the results of the 2D waveform inversion of the transformed 3D data are compared with the result obtained from the inversion of the 2D data. Inverted V_P models are very well resolved for both corrections and they are comparable to the reference 2D data inversion. However, it can observed that the resolution decreases below a depth of 1.6 km, which is probably due to the strong velocity contrast at this depth. The under- and overestimation of amplitudes, discussed in the previous paragraph, have a direct impact on the inversion results. This effect is well visible in the velocity profiles (Figure 6.4c). The correction C_1 is underestimating amplitudes and as a consequence, the reconstructed velocity values are slightly smaller than the true V_P values. The opposite effect is seen, when the C_2 correction is applied. In this case, the amplitudes and, in effect, the velocities are overestimated. Quite surprisingly, the overall model error is a little bit higher for the C_2 correction, although this transformation provides a better fit to the 2D data.

In the case of a relatively simple geology, the 3D to 2D transformation derived from Green's functions are very efficient. However, as shown by Köhn [2011], the simple C_1 correction produces artefacts in the inversion results when applied to more complex geological settings.

6.1.3 2D inversion of 3D data

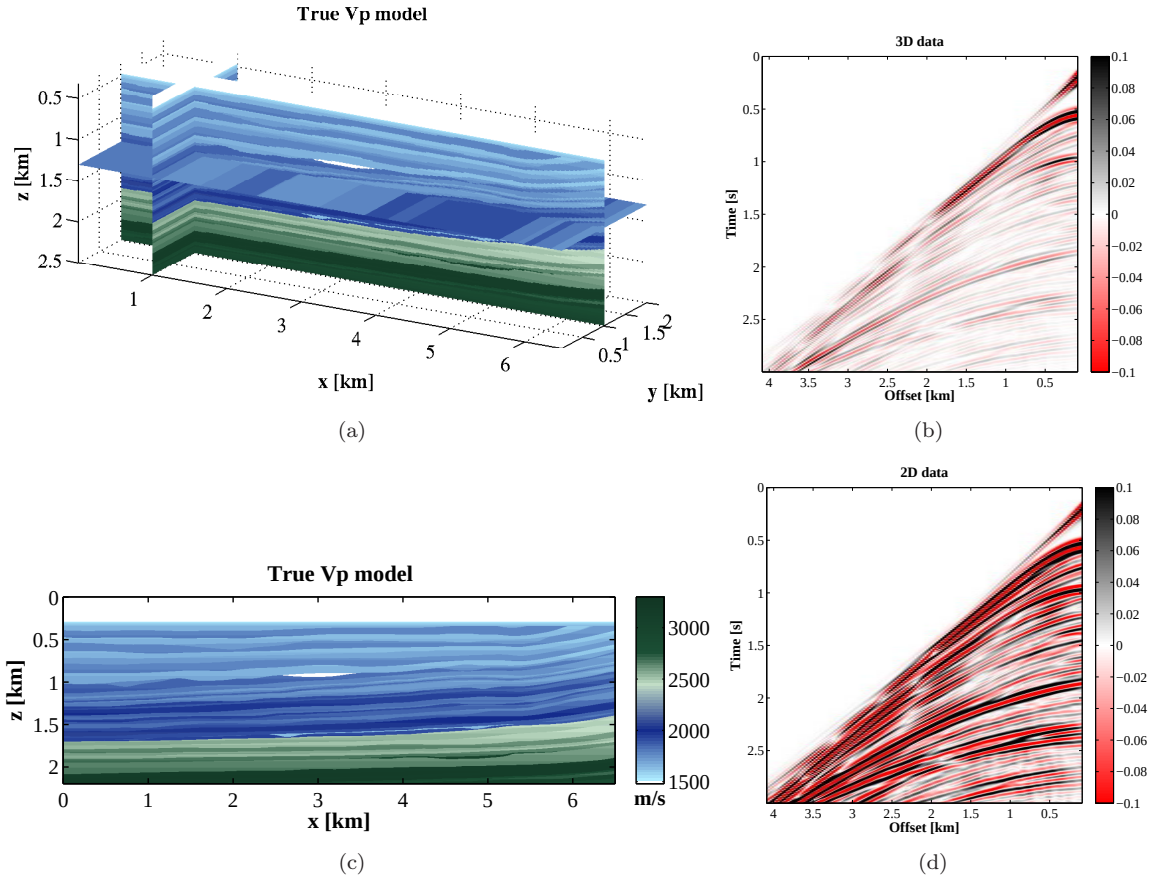


Figure 6.1: (a) True 2.5D velocity model used for the 3D acoustic modelling and (b) the corresponding 3D point source data. (c) True 2D velocity model used for the 2D acoustic modelling and (d) the corresponding 2D line-source data.

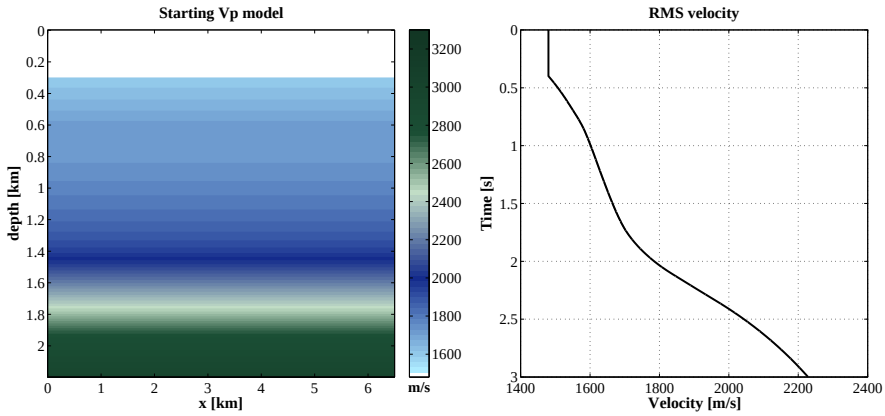


Figure 6.2: (a) Starting V_P model and the corresponding RMS velocity function (b) used in the transformation C_2.

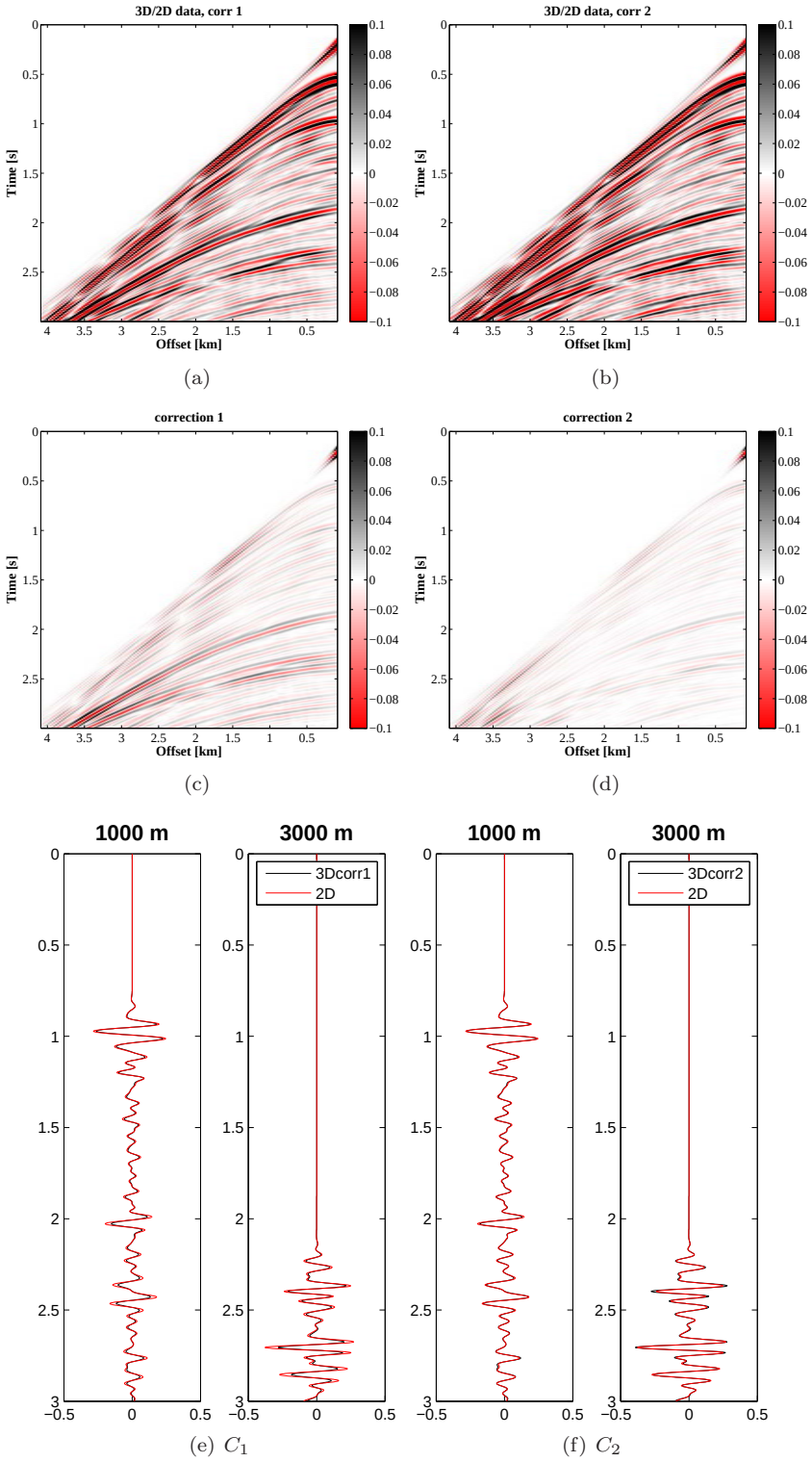


Figure 6.3: Comparison of the 3D data after 3D to 2D transformations. (a) 3D data after correction C_1, (c) residuals between the 3D data after correction C_1 and the 2D data, (e) trace comparison at 1km and at 3 km for the 2D data (red line) and the 3D corrected data (black line). (b) 3D data after correction C_2, (d) residuals between the 3D data after correction C_2 and the 2D data, (f) trace comparison.

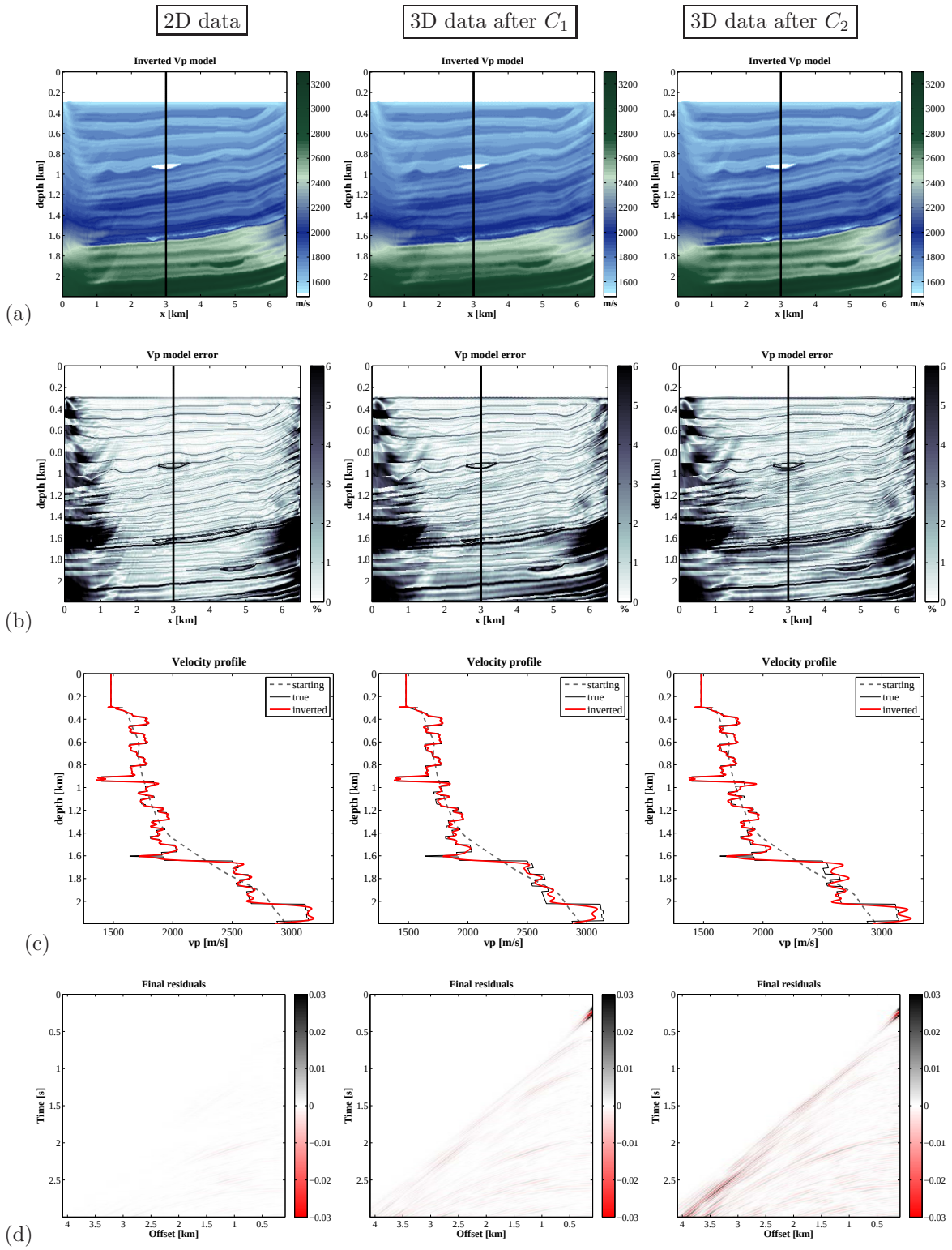


Figure 6.4: 2D inversion of: 2D data (left), 3D data after transformation C_1 (middle), and C_2 (right). (a) Inverted V_P models, (b) relative error between the inverted and the true V_P model, (c) V_P profiles at $x = 3$ km, (d) final data residuals.

6.2 Estimation of the source signature

It is common to assume known source properties in synthetic experiments. It means that the same source wavelet is used to generate synthetic data and in the full waveform inversion. However, an accurate estimation of wavelet is an important problem in field data applications. Wrong wavelet used in forward modelling results in a poor fit of synthetic data to observed data.

There are different methods used to estimate the wavelet signature for the full waveform inversion:

- solving a linear least-squares inverse problem [Pratt, 1999]
- stacking along the water-bottom reflection [Boonyasiriwat et al., 2008]
- near offsets at the water bottom are used to derive the source wavelet and then a shaping filter is designed to transform the estimated wavelet to the Ricker wavelet [Vigh and Starr, 2008]
- source wavelets recorded at the source during the seismic survey [Crase et al., 1990]
- frequency-domain averaging and phase deconvolution algorithm [Hicks and Pratt, 2001]
- airgun source modelling software [Delescluse et al., 2011]
- synthetic estimate of the source signature based upon a cumulative modelled response of the individual elements of the airgun array [Shipp and Singh, 2002]

I have investigated two different methods for source wavelet estimation. The first one is based on the estimation of source signature from direct arrivals to hydrophones [Kravis, 1985]. The second method is a linear-least squares inversion proposed by [Pratt, 1999]. This approach is widely used in frequency-domain full waveform inversion of field data, however it required some modifications to apply it in the time-domain inversion. To compare the performance of these two methods, I present the results from synthetic inversion experiments using the Marmousi2 test case.

6.2.1 Wavelet estimation from the direct wave

The first method, proposed by Kravis [1985], estimates the source wavelet from the direct wave. It is a portion of the data that can be easily used to estimate the reflectivity. The principles of this approach are illustrated on the Marmousi2 test case (for details see Chapter 5.2.3).

The signature estimation is based on the deconvolution of the impulse response of a hydrophone from the observed direct arrivals. In the absence of noise, the direct wave $d(t)$ recorded at the hydrophone is the convolution of the pure seismic source signature $w(t)$ with the impulse response of a hydrophone $p(t)$.

$$d(t) = w(t) * p(t) \ . \tag{6.11}$$

The response of a hydrophone to a pressure impulse transmitted from a source consists of two impulses: 1) ray traveling directly from the source to the hydrophone, 2) source ghost reflection. Source ghost is arising from the water-air interface. The energy from the source is not only propagating horizontally to the hydrophone, but it is also traveling upward to the sea surface, where it is reflected downward (Figure 6.5a). This reflected signal is called the source ghost. It has reversed polarity and it is delayed in time with respect to the primary pulse. Since the source is located near the sea surface, usually at

depths of around 10 m, the source ghost closely follows the primary pulse and changes the effective wave shape [Ikelle and Amundsen, 2005].

The response of a single hydrophone located at position $x = x_h$, at a depth d_h, to an impulse excited at time $t = 0$ from a source located at $x = 0$, at a depth d_s, is given by

$$p(x_h, t) = \frac{1}{r_\mathrm{d}}\,\delta\left(\frac{r_\mathrm{d}}{v_\mathrm{d}} - t\right) + \frac{R}{r_\mathrm{gh}}\,\delta\left(\frac{r_\mathrm{gh}}{v_\mathrm{d}} - t\right)\,, \tag{6.12}$$

where $\delta(t)$ is the Dirac delta function, v_d is the average velocity of the direct ray, R is the reflection coefficient at the sea surface, which is close to -1. r_d and r_gh represent the raypath lengths of the directly traveling ray and the ghost reflection, respectively. The time delay between the two impulses is given by

$$T_\mathrm{d} = \frac{2d_\mathrm{h}d_\mathrm{d}}{v_\mathrm{d}x_n}\,. \tag{6.13}$$

In this test example the source and receiver depth is $d_\mathrm{h} = d_\mathrm{s} = 7.5$ m, source-hydrophone distance is $x_n = 100$ m, and water velocity $v_\mathrm{d} = 1480$ m/s, which gives the time delay of 760 μs.

The deconvolution is performed in the frequency domain by inverse filtering, such that

$$d(t) = w(t) * p(t) \Longrightarrow D(f) = W(f) * P(f) \tag{6.14}$$

$$W(f) = D(f) * \frac{1}{P(f)}\,. \tag{6.15}$$

The aim is to find the stable inverse filter $1/P(f)$. First, the impulse response of a hydrophone is digitized with a very small sampling rate (Figure 6.5b), it this example 0.1 ms, and then the Fourier transform is applied to calculate the resulting spectrum $P(f)$. The reciprocal of the impulse response represents the inverse operator. Finally, the inverse of the impulse response (shown in Figure 6.5d) is convolved with the observed direct arrival (Figure 6.5e) by frequency domain multiplication to provide an estimate of the source signal

$$\widehat{w}(t) = \mathcal{F}^{-1}\left\{D(f)\cdot\frac{1}{P(f)}\right\}\,. \tag{6.16}$$

The estimated source signature $\widehat{w}(t)$ is a very good match to the true wavelet (Figure 6.5f).

Although the method is efficient and simple, there are some drawbacks that may limit its accuracy in field applications:

- It requires that there is no interference between the recorded direct arrivals and the seabed reflections or refractions.
- It assumes a spherically symmetric source. It is possible to extend the method to handle non-spherically symmetric sources, such as source arrays, but then it requires that all source elements are identical. This is rarely the case in marine streamer surveys, which mainly use airgun arrays composed of multiple airgun units with different volumes.
- It requires a determination of the amplitude scaling factor to compensate for the unknown source strength and sensitivity of the recording system. It means that the field shot gathers must be scaled with the corresponding amplitudes of the initial synthetics.

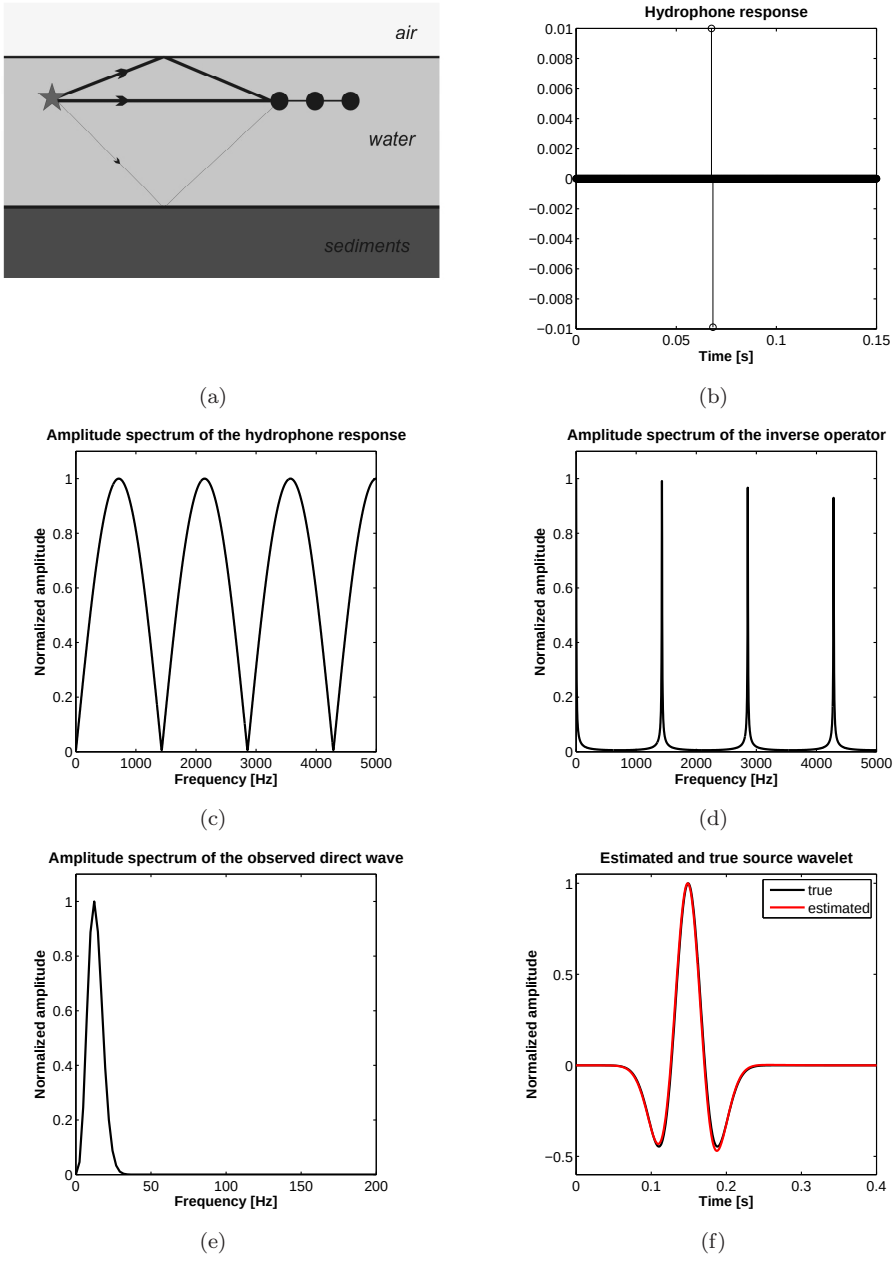


Figure 6.5: Wavelet estimation from the direct wave. a) Raypaths for a direct wave. b) Impulse response of a hydrophone to a horizontally traveling impulse, c) with the corresponding amplitude spectrum. d) Amplitude spectrum of the inverse filter, e) amplitude spectrum of the observed direct arrival waveform. f) red line: estimated source wavelet obtained by convolving the direct arrival waveform with the inverse filter, black line: true source wavelet.

This can be achieved by scaling the amplitudes of the seabed reflection at the nearest offset. However, to use this approach, the starting velocity model must already contain an accurate information on the true seafloor parameters, because the seafloor reflection in each field gather would be equalized to the value of the synthetic data. Amplitude calibration can be also performed by balancing the amplitudes around the direct arrival at the nearest offset channel. This second approach requires a close fit of the modeled direct arrival to the one observed in the field data. Therefore, a proper description of the acquisition parameters, a good estimate of the source signature, and an accurate modelling of the directivity effects is necessary. But on the other hand, this method allows the update in the seafloor region and it normalizes variations in the source strength between the shots.

6.2.2 Source wavelet inversion

Source wavelet can be estimated from the seismic data by solving a linear inversion problem. In this approach the wavelet is the unknown that we want to determine from the known reflectivity series and recorded data. The method involves forward modelling to compute a synthetic seismogram that is then compared to the observed data. The initial wavelet is modified such that synthetics give a satisfactory match to the observed data. This approach can be directly incorporated into the waveform inversion algorithm.

6.2.2.1 Method

According to the convolutional model [Berkhout, 1977], the seismic trace $d_i(t)$ recorded at the i-th receiver is the result of convolution of the earth's reflectivity function $r_i(t)$ with the propagating seismic wavelet $w(t)$ plus a noise component $n(t)$

$$d_i(t) = w(t) * r_i(t) + n(t) \; . \tag{6.17}$$

Once the reflectivity sequence is known, the synthetic trace $u_i(t)$ can be written as

$$u_i(t) = s(t) * r_i(t) \; , \tag{6.18}$$

where $s(t)$ is a wavelet used to generate synthetic data. Given the initial wavelet estimate $s(t)$, we want to find a filter $h(t)$ so that we can estimate $w(t)$ as follows

$$h(t) * s(t) = \widehat{w}(t) \; , \tag{6.19}$$

where $\widehat{w}(t)$ is an estimate of $w(t)$. After applying the filter $h(t)$ to the synthetic seismic trace, we obtain

$$h(t) * u_i(t) = h(t) * [s(t) * r_i(t)] = [h(t) * s(t)] * r_i(t) = \widehat{w}(t) * r_i(t) = \widehat{d_i}(t) \; . \tag{6.20}$$

The inverse problem is formulated in the frequency domain. We compute the filter by minimizing the least-squares error function (Pratt, 1999)

$$\epsilon = ||\mathbf{d}(\omega_j) - h(\omega_j)\mathbf{u}(\omega_j)||_2^2 = \frac{1}{2} \left(\mathbf{d}(\omega_j) - h(\omega_j)\mathbf{u}(\omega_j)\right)^{\mathrm{T}} \left(\mathbf{d}(\omega_j) - h(\omega_j)\mathbf{u}(\omega_j)\right)^{*} \; , \tag{6.21}$$

where $\mathbf{d}(\omega)$, $\mathbf{u}(\omega)$ are the discrete Fourier transforms of recorded and modelled shot gathers, respectively. ω is angular frequency, j the frequency index, the superscript T represents the matrix transpose, and * denotes the complex conjugate. $h(\omega_j)$ is a complex-valued scalar representing an optimal source weight at a given frequency. The least-squares error is minimized when the following condition is satisfied

$$\frac{\partial \epsilon}{\partial h(\omega_j)} = 0 \, . \tag{6.22}$$

Since it is a linear least squares problem, a closed-form solution is given by

$$h(\omega_j) = \frac{\sum_{i=1}^{M} \mathbf{u}_i(\omega_j)^t \mathbf{d}_i(\omega_j)^*}{\sum_{i=1}^{M} \mathbf{u}_i(\omega_j)^t \mathbf{u}_i(\omega_j)^*} \, , \tag{6.23}$$

where M is the number of traces in a shot gather. If the reflection coefficients are assumed known, then the least squares inversion and Wiener filter approach are equivalent (Lines, Treitel).

Straight spectral division causes numerical problems in the calculation. Small or zero values in the denominator results in numerical instability, therefore the deconvolution problem from (6.23) must be regularized. To stabilise a spectral division, I have chosen the water-level deconvolution method [Menke, 1984]. The values in the denominator that are below a specified threshold are replaced with the so-called water-level parameter. Above the water-level, the denominator value is identical to the original value, and only the small values are replaced. This prevents high amplifications of the data spectrum at frequencies that have very small (close to zero) amplitudes. The water-level parameter is chosen by trial and error and it is a fraction of the maximum value of the denominator computed for all frequencies. I have chosen the lowest value that produces acceptable results and the water-level parameter is set to 1e-5 of the maximum amplitude of the observed shot gather.

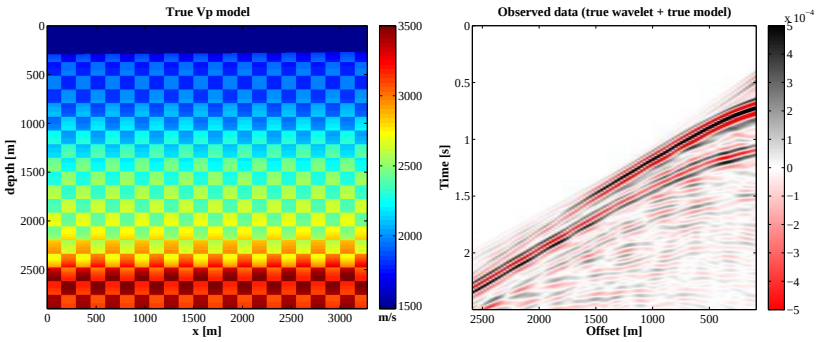
In order to obtain a time-domain representation of the updated source signature, we need to calculate the optimum source weight for every discrete frequency n ($n = 1, \ldots, N$), where N is the number of frequencies represented by the Fourier transform of the seismic data. According to the convolution theorem, convolution in the time domain is mapped as multiplication in the frequency domain. Thus the wavelet estimate in time domain can be obtained by

$$\widehat{w}(t) = \mathcal{F}^{-1} \left\{ \mathbf{h}(\omega) \cdot \mathbf{s}(\omega) \right\}_N \, , \tag{6.24}$$

where $\mathbf{s}(\omega)$ is the Fourier transform of the initial source wavelet $s(t)$.

6.2.2.2 Implementation

I implemented this source wavelet estimation algorithm in the full waveform inversion code. The source signature is estimated separately for every shot gather. There is only one additional forward modelling required prior to the actual inversion of model parameters. The aim is to obtain the first source estimate for the starting velocity model. As an initial source signature, I use a delta function. It provides better results than the Ricker wavelet, mainly because it has constant amplitude at all frequencies (white spectrum). However, the propagation of a delta function on the finite-difference grid is complicated due to its white spectrum. To avoid numerical dispersion effects, an infinitesimal grid spacing would be required. To overcome this problem, a low-pass filtered delta function is used. The high-cut frequency of the Butterworth filter corresponds to the maximum frequency component



(a) (b)

Figure 6.6: (a) True V_P model and (b) the corresponding observed data. Source time function is the Ricker wavelet with the peak frequency of 10 Hz.

of the input data to ensure no loss of accuracy.

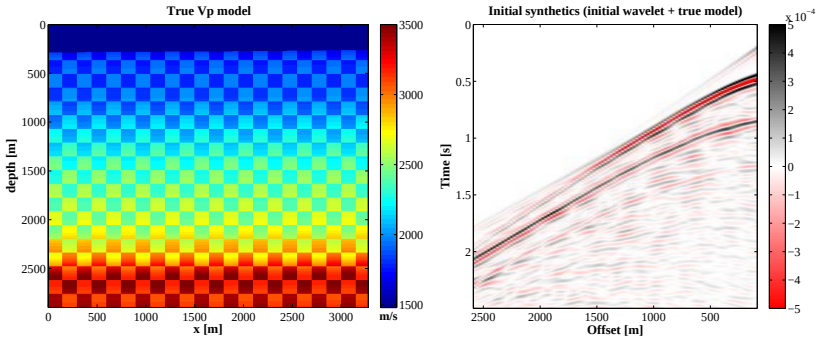
Source wavelet inversion is performed only at iteration steps related with the change of the data content, i.e. the change of frequency band, time window or offset window. Synthetic inversion tests have shown that it is not necessary to update the source signature at every iteration step.

6.2.2.3 Performance

The main factor limiting the accuracy of the wavelet inversion method is the assumption that the velocity model is known. To illustrate the impact of velocity model errors on the accuracy of the wavelet reconstruction, two test cases are considered. In the first case the correct velocity model is used to generate initial synthetic data so that the source wavelet is the only unknown (Figure 6.7a). In the second, more realistic, example a smooth background model is used instead of the true velocity (Figure 6.8a). The observed data are generated for the checkerboard velocity model using the Ricker wavelet with the peak frequency of 10 Hz as the source time function (Figure 6.6). The initial source wavelet is the low-pass filtered delta function.

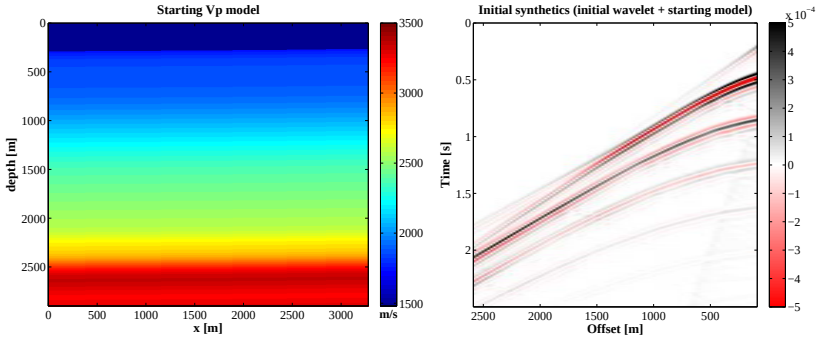
When the true velocity model is used, the inverted source signature is almost identical to the true source wavelet (Figure 6.9). On the other hand, for the smooth velocity model, we still obtain a very good estimate of the source wavelet (Figure 6.9). There are some discrepancies mainly visible in the amplitude spectrum, which are caused by the inaccuracies in the velocity model that are propagating into the source wavelet. This result suggests that an initial velocity model is sufficient to provide a correct representation of the source time function.

As pointed out by Brenders [2011] the inverted source wavelets can also serve as the quality check on the velocity model during the full waveform inversion. At each iteration, the model parameters are updated and the synthetic data is generated. When the residuals between the observed and predicted data are decreasing and the estimated source wavelets improve as well, this indicates an accurate reconstruction of the velocity model. On the contrary, if the reconstructed velocity model is incorrect this would result in a lack of coherency in estimated wavelets.



(a) (b)

Figure 6.7: Test case 1. (a) True V_P model is used to generate initial synthetics (b) with the low-pass filtered delta function as a source wavelet.



(a) (b)

Figure 6.8: Test case 2. (a) Smooth, background V_P model is used to generate initial synthetics (b) with the low-pass filtered delta function as a source wavelet.

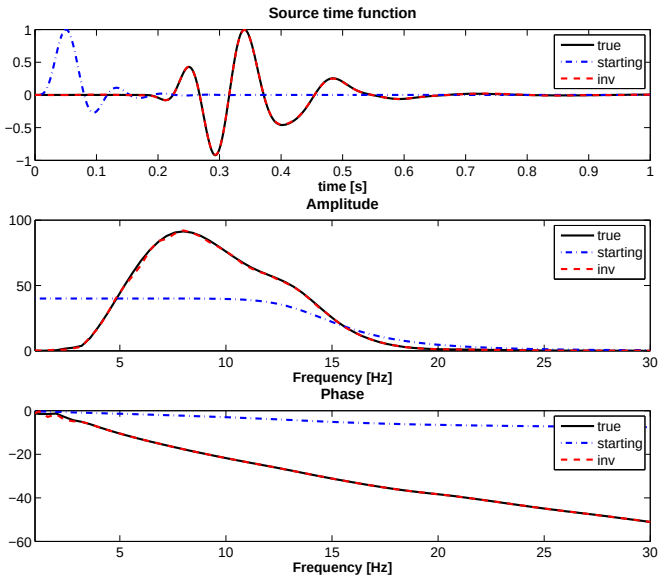


Figure 6.9: Test case 1 - wavelet inversion with the true velocity model. Top: time domain representation of the source wavelet, center: amplitude spectrum, bottom: phase spectrum. Solid line denotes the true Ricker wavelet used to generate observed data; dash-dotted line is an initial, low-pass filtered delta function; dashed line represents the wavelet estimated with least-squares inversion.

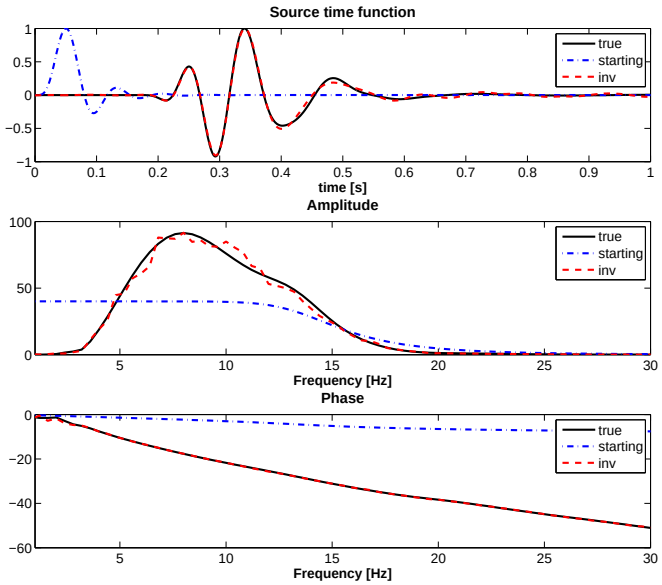


Figure 6.10: Test case 2 - wavelet inversion with the smooth velocity model. For details see caption of Figure 6.9

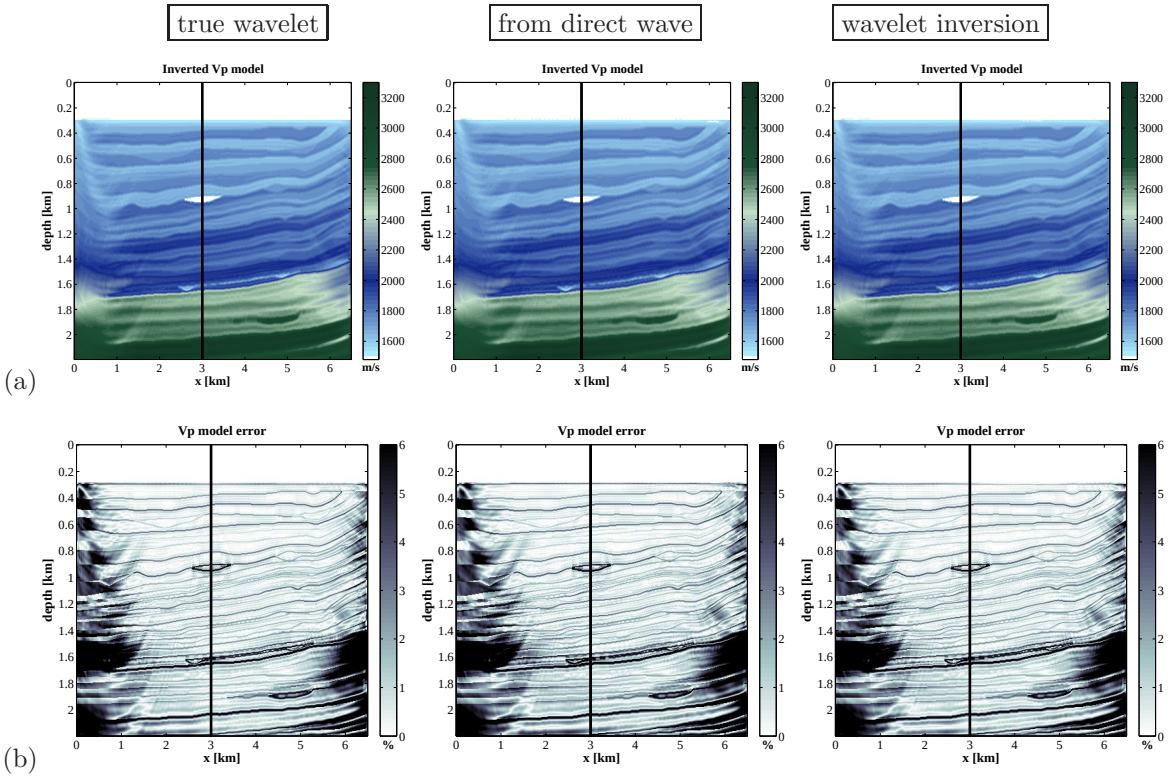


Figure 6.11: FWI case 1 with the correct source location. (a) Inverted V_P models, (b) final V_P model error. Left: reference inversion result with the true wavelet, center: wavelet estimated from the direct wave, right: source wavelet inversion.

6.2.3 Wavelet estimation - inversion tests

The presented wavelet estimation methods were applied to the benchmark Marmousi2 test. The synthetic dataset was generated with the Ricker wavelet as the source time function, with the central frequency of 10 Hz. The starting velocity model is the 1D-smoothed true model. In order to highlight wavelet effects, the density model is assumed to be known. The reference inversion result, computed with the correct source time function, is displayed in Figure 6.11a.

Case 1

The objective of this experiment is to assess the effectiveness of two investigated wavelet estimation methods within the inversion scheme. Since the wavelet is unknown, we must determine both the wavelet and the velocity model from the input data. The inversion result using the first method, i.e. the wavelet estimated from the direct wave, is shown in Figure 6.11b, whereas the FWI result using the least-squares inversion approach for the wavelet estimation is shown in Figure 6.11c. The velocity structures are very well recovered and are nearly identical to those of the reference inversion result. This indicates that both methods provides a very good estimation of the real source signal. In addition, the final relative model error between the real and the reconstructed P-velocities is shown in Figure 6.11b.

Case 2

In the second inversion experiment, I introduced a systematic error in the acquisition geometry.

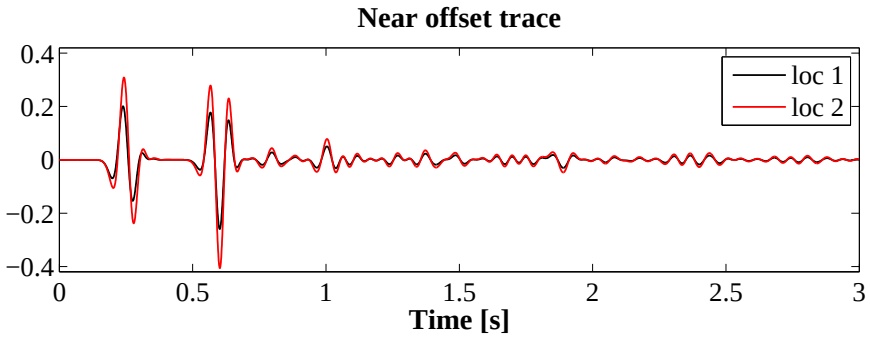
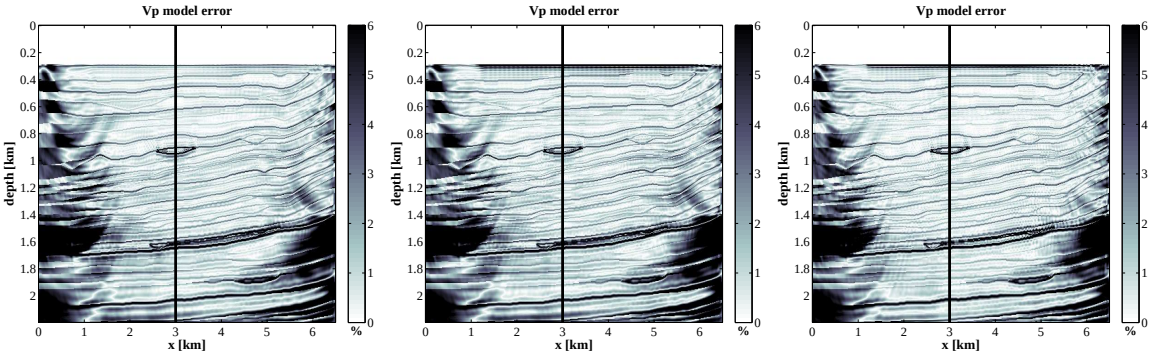


Figure 6.12: The effect of a source location on the waveform recorded at the near offset receiver located at $\mathbf{r} = (6150\text{m}, 7.5\text{m})$. Black trace corresponds to the source located at $\mathbf{s_1} = (6250\text{m}, 7.5\text{m})$, red trace corresponds to the source located at $\mathbf{s_2} = (6255\text{m}, 12.5\text{m})$.

The source location used to generate synthetic data is different from the source location within the inversion scheme. The sources are shifted by 5 m in x- and in y-direction, which means that the "real" streamer is towed at the depth of 12.5 m, whereas the streamer in FWI is towed at 7.5 m below the air-water interface, and the near-offset is 105 m and 100 m, respectively. This relatively small source location error has a significant effect on the observed wavefrom (Figure 6.12), which is mainly caused by the different source ghost effect. The interference between the primary, downgoing wavefield and the ghost wavefield, that is reflected at the air-water interface, varies with the depth of the source. Since the source ghost is almost indistinguishable from the primaries, it is treated as part of an effective-source signature [Ikelle and Amundsen, 2005].

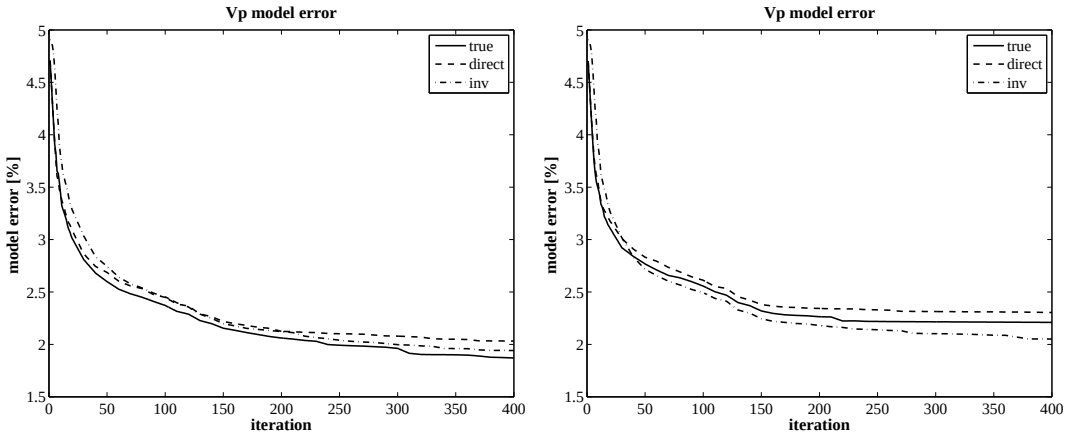
In field applications, this type of acquisition setup errors is easily introduced due to the finite-difference discretization method and due to the source geometry instability. In marine acquisition, the stability of the geometry depends not only on the acquisition system design but mainly on the weather conditions [Parkes and Hatton, 1986]. The source energy and the exact source and receiver location are strongly affected by sea surface waves, which in consequence has a major effect on the effective-source signature.

Within this experiment, I performed three FWI tests using the true source wavelet, the wavelet estimated from the direct wave, and the source wavelet inversion method. The final velocity model deviations from the true model are presented in Figure 6.13. The overall accuracy of the reconstructed images is lower in comparison to the first experiment. This trend can be also observed in Figure 6.14, which represents the evolution of the V_P model error during the inversion process for different source wavelets. However, in contrast to the first test with the correct source acquisition, the wavelet inversion method provides the most accurate inversion result when the incorrect source location is introduced. With the true source signature, the effects of geometry errors are projected directly into the velocity model, whereas the wavelet inversion can partly compensate for source location errors. This results in a better quality of the inverted models. When the wavelet is estimated from the direct wave, the reconstructed model is acceptable, but it contains more noise relative to other two inversion results.



(a) true wavelet (b) from direct wave (c) wavelet inversion

Figure 6.13: FWI case 2 with the incorrect source location. Final V_P model error. Left: reference inversion result with the true wavelet, center: wavelet estimated from the direct wave, right: source wavelet inversion.



(a) correct source location (b) incorrect source location

Figure 6.14: Evolution of the V_P model error. (a) FWI case 1 with the correct source location, (b) FWI case 2 with the incorrect source location. Solid line indicates the reference inversion with the true wavelet, dashed line represents the wavelet estimated from the direct wave, dashed-dotted line represents the source wavelet inversion.

6.3 Acquisition effects - Airgun array versus single airgun approximation

The typical source in marine streamer seismics is an array of airguns that operates as an exploding source. A single airgun is a poor energy source because of the bubble effect. The firing of an airgun generates an oscillating bubble in the water. The initial bubble expansion generates the desired broadband seismic pulse, which is followed by the subsequent damped oscillating bubble pulse that cause source-generated noise. To minimize the bubble effect, airguns of various sizes are combined to form arrays [Gadallah and Fisher, 2005]. Spacing between individual airguns, their volume and the actual fire time is such that the bubble pulses interfere destructively and the resulting array signature is approaching the ideal characteristic of a single impulse. Another reason for using airgun arrays is to increase a source strength. The peak amplitude of an airgun's signature is proportional to the cube root of the volume of air in the airgun

$$A \sim V^{1/3} \,. \tag{6.25}$$

It means that an airgun volume would have to be increased by a factor of 8 in order to generate a two times larger amplitude. The same effect can be achieved when two airguns of the same volume are fired simultaneously.

Since an airgun array has a certain width and length, it shows directivity effects in contrast to a single airgun. Each airgun array has its own specific radiation pattern, i.e. the energy of the airgun array signature changes with the horizontal azimuth angle and with the emission angle (vertical angle) [Li et al., 2011]. It means that the frequency content and the pressure strength of the array signature are different at different locations in the water. An important feature is that the directivity effects are stronger for high frequencies [Sheriff and Geldart, 1982].

To investigate whether the single point source approximation of an airgun array is accurate for waveform inversion, I performed 3D numerical simulations using both a single airgun and an airgun array. The acquisition geometry used in this experiment is based on the field data survey presented in Chapter 7. The airgun array (Figure 6.15) consists of 20 airguns of different volume located at a depth of 7.5 m below the air-water interface. The source array length is 20 m and the width is 30 m. The test model is the 3D-Marmousi2 model presented in Chatper 3 and the maximum frequency of the source signal is 20 Hz.

Waveforms generated with the airgun array and with the point source are shown in Figure 6.16. To allow for a direct comparison, both seismograms are normalized to the maximum of the seabed reflection at the near offset trace. Figure 6.17a shows a comparison of the RMS amplitude of a trace versus offset for the point source data and the airgun array data. Because most of the energy emitted from the airgun array is concentrating close to the vertical emission angle, the amplitudes at wide angles are smaller than the corresponding amplitudes from the point source. The relative error of the RMS amplitudes of point source is less than 11 per cent and it is generally increasing with offset (Figure 6.17b). This relatively small misfit indicates that the point source approximation is sufficient to reproduce the directivity effects of the analysed airgun array for the frequency range from 0 to 20 Hz. This result is consistent with the results of Shipp and Singh [2002], who investigated directivity effects of a horizontal source array for different frequency ranges, but limited the comparison to the seabed reflection only.

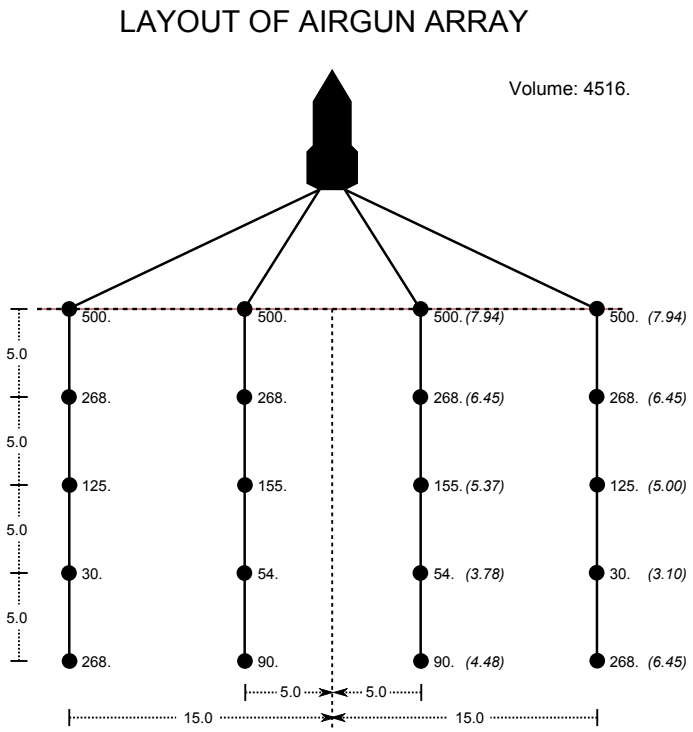
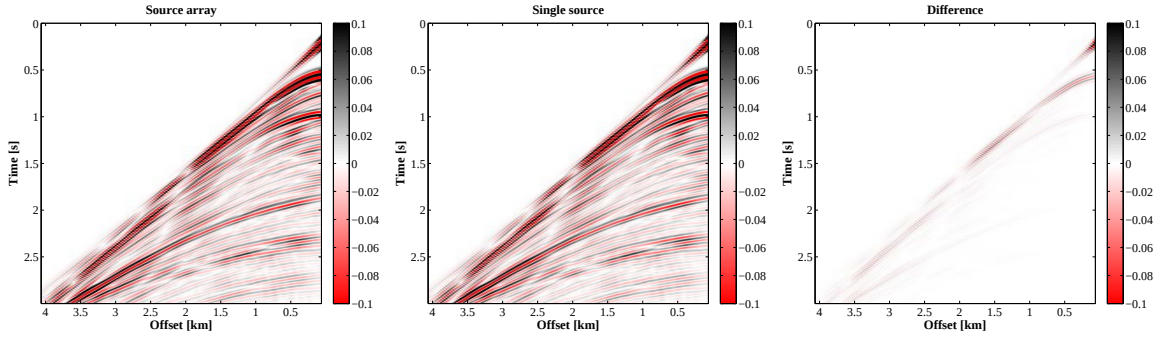


Figure 6.15: Airgun array configuration. There are 20 individual airguns with a total volume of 4516 inch^3. The numbers in brackets represent the peak amplitude of an airgun.



(a) (b) (c)

Figure 6.16: Shot gathers generated with the airgun array (a) and with the single airgun (b). Both seismograms are normalized to the maximum of the seabed reflection at the near offset trace. (c) The difference plot shows that the waveform discrepancy is mainly associated with the strong seafloor reflection and reflections at middle and far offsets.

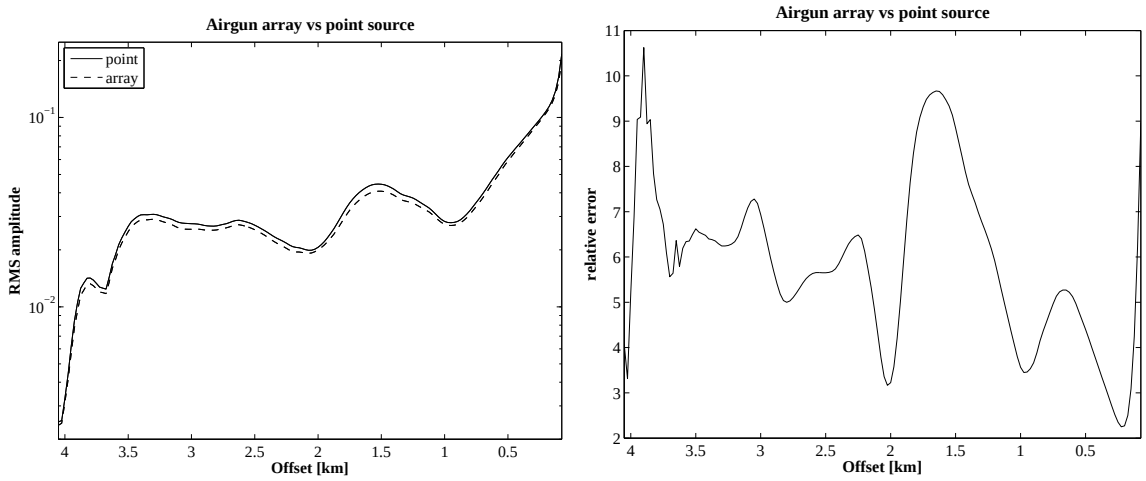


Figure 6.17: (a) RMS amplitude versus offset for the point source and the airgun array, RMS amplitude is computed for every trace.(b) The relative error of the RMS amplitudes of the point source.

6.3.1 Summary

In this chapter, I have discussed several practical aspects related to the FWI of real data. To illustrate numerically the efficiency of different 3D to 2D corrections and the source wavelet estimation methods, I have carried out a series of inversion tests. In the following, I give a brief summary of the main findings.

Due to the difference between 2D and 3D geometrical spreading, it is necessary to transform the 3D point source data to 2D line source data. In this study, I have considered simple 3D to 2D transformations derived from the Green's functions. In spite of strong assumptions on velocity structures, both corrections are efficient for transformation of 3D data generated for a simple geological model.

Accurate source wavelet estimation is a prerequisite for successful application of FWI to field data. In this study, I have investigated two methods: source wavelet estimation from direct arrivals to hydrophones and a linear-least squares inversion. Although the first method can be easily used as part of the data processing, it has some drawbacks that may limit its accuracy in field applications. On the other hand, the source wavelet inversion method provides a very good estimate of the source time function and it offers a number of advantages. It can be easily introduced within the full waveform inversion framework and the source signature is estimated separately for every shot gather. Furthermore, the numerical results show that the wavelet inversion can partly compensate for location errors in the acquisition geometry, which is of importance when inverting real data.

Finally, I have addressed the problem of the directivity of an airgun array. The synthetic tests indicate that the point source approximation is sufficient to reproduce the directivity effects of an airgun array for the frequency range from 0 to 20 Hz.

Chapter 7

Field data example from the North Sea

In this Chapter, I apply the 2D acoustic FWI to seismic streamer data acquired in the North Sea in 1991. First, I discuss the key aspects related to the field data preprocessing, which is required to enable a direct comparison between the observed and modelled data. The main challenge of the field data inversion is the generation of an accurate starting velocity model. In this study, I assess the accuracy of initial models generated with different methods, i.e. from the VSP data, from the NMO velocities, and from the traveltime tomography. Afterwards, I discuss the choice of the inversion parameters that aims to reduce the high complexity of the inverse problem. The resolving power of the FWI is tested with the checkerboard model using the same acquisition setup, source signature and the frequency content as in the field data example. In the next section, I compare the field data inversion results obtained for different multi-stage inversion strategies, I present the multi-parameter inversion results as well as the mono-parameter inversion for different minimization criteria and for different starting models. Finally, I validate the reliability of reconstructed models by a comparison with the migrated seismic section and by the analysis of the data fit.

This Chapter is organized according to the workflow for the field data inversion that comprises the following steps:

1. Data preprocessing
2. Starting model generation
3. Source wavelet estimation
4. Initial modelling
5. Choice of the inversion parameters
6. Resolution study
7. Multi-scale, multi-stage FWI inversion
8. Validation of inversion results

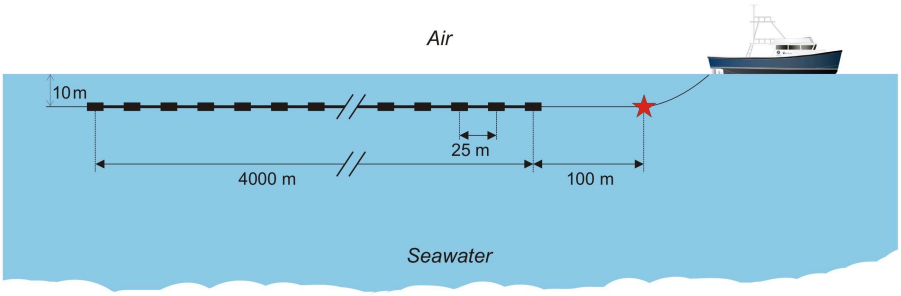


Figure 7.1: The acquisition parameters of the field data.

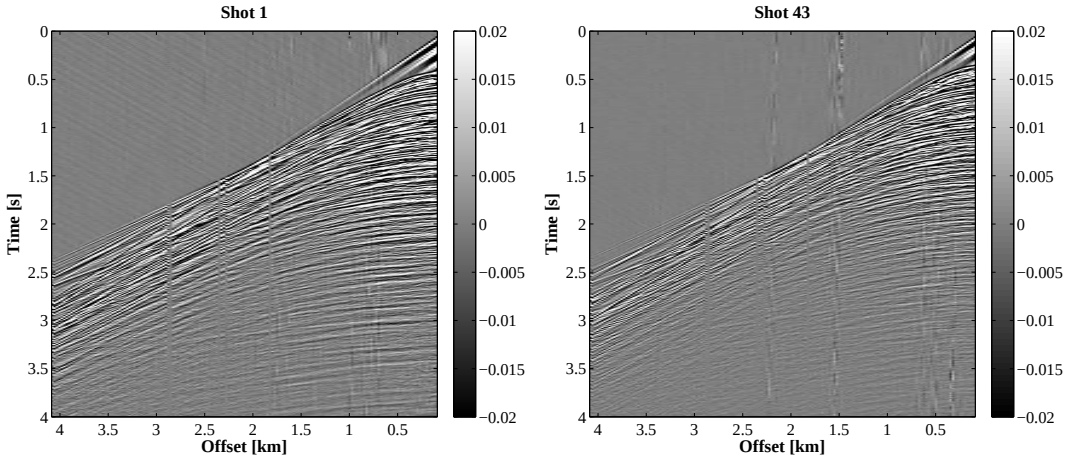
7.1 Field data

The seismic data were acquired in the northern Norwegian North Sea in the Møre Basin, which is a deep Cretaceous basin. The data set was collected by Geco-Prakla in 1991. The survey gathered over 30 km of 2D seismic data. The acquisition geometry is composed of a 160-channel, 4000 m single-sensor towed streamer (Figure 7.1).

The main acquisition parameters are:

Source Type	Airgun array
Volume	4804 C.I.
Source tow depth	7.5 m
Shotpoint interval	25 m
Receiver type	Streamer cable
Recording system	HSSQ
Group interval	25 m
Number of groups	160
Streamer tow depth	10 m
Near offset	100 m
Far offset	4075 m
Sample interval	2 ms
Record length	7000 ms

From the available data, I selected a subregion that extends over 7.5 km, and I used 62 shots spaced every 100 m (every fourth shot). Examplary raw shot gathers are shown in Figure 7.2 and the amplitude spectrum is illustrated in Figure 7.3.



(a) (b)

Figure 7.2: Exemplary raw shot gathers. The vertical stripes in the data represent swell noise that has large amplitudes at low frequencies.

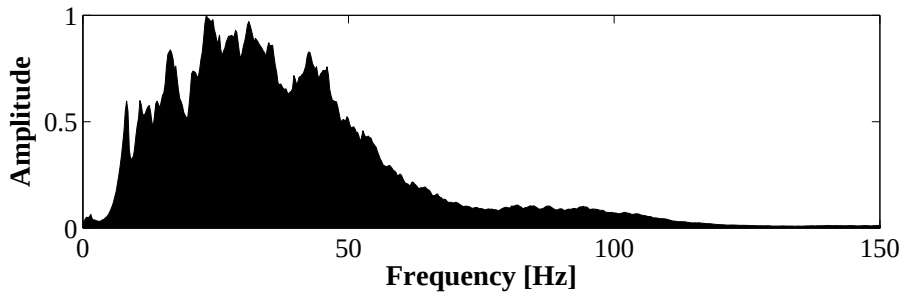


Figure 7.3: Normalized amplitude spectrum of the raw shot gather 1. Spectrum averaged over 160 traces, time window: 0 to 4000 ms.

7.2 Data preprocessing

The data preprocessing is a crucial part of the workflow for full waveform inversion of the field data. The main objectives of the data processing are to improve the signal-to-noise ratio and to transform the field data such that they reflect the approximations made in the forward modelling.

The aim of acoustic FWI is to find a set of model parameters, i.e. P-wave velocity and density, which describes the acoustic properties of the subsurface and provides a good fit to the observed data. Since the criterion for selecting the best model of the subsurface is based on fitting the observed data and the data predicted by the forward modelling, it is important to allow for a direct comparison between the modelled data and observed data. The forward problem is solved using a finite-difference approximation of the 2D acoustic wave equation. Therefore, it is necessary to preprocess the observed data to remove these wave propagation phenomena that cannot be modelled correctly, such as seismic noise, elastic effects, or 3D effects. Otherwise the seismic noise and the nonacoustic factors will be projected into the reconstructed P-wave velocity models.

7.2.1 Quality control

The quality control is be used to monitor amplitudes of the data and to identify bad or noisy traces, or misfired shots. Figure 7.4 shows the root-mean-square (RMS) trace amplitude as a function of shot and channel. Each square cell represents one trace in the input raw data. There are four channels (70, 91, 111, 138) that are quieter than the others. These bad traces can be also identified in the shot gathers (Figure 7.2). The distribution of the RMS amplitude with offset is relatively similar for all shots. However, the amplitudes at near offset traces are higher for shots 50 - 62.

7.2.2 Mute

Muting is applied to the data to remove the noise preceding the first arrivals. Moreover, the late arrivals that correspond to deep reflections coming from outside of the model are muted as well. The original recording length of 7 seconds is limited to 4 seconds two-way time.

7.2.3 Trace interpolation

Interpolation of bad traces is not a prerequisite, because the full waveform inversion can handle any irregular geometry configuration. In this particular case, the four channels (70, 91, 111, 138) recorded the signal but its amplitude was very weak in comparison to the neighboring traces. Therefore, it was enough to balance the amplitudes of these bad traces by estimating the correct energy level from adjacent channels.

7.2.4 Noise suppression

A noise component is always present in seismic data. An efficient removal or attenuation of different noise types can be a challenging process, however it is important to obtain good inversion results. Marine noise comprises mainly noise generated due to hydrostatic pressure fluctuations, swell noise, noise from the vessel, and seismic interference [Elboth et al., 2009]. This high-amplitude noise normally contains frequencies from 0-10(15) Hz, and without an appropriate de-noising of seismic data it would

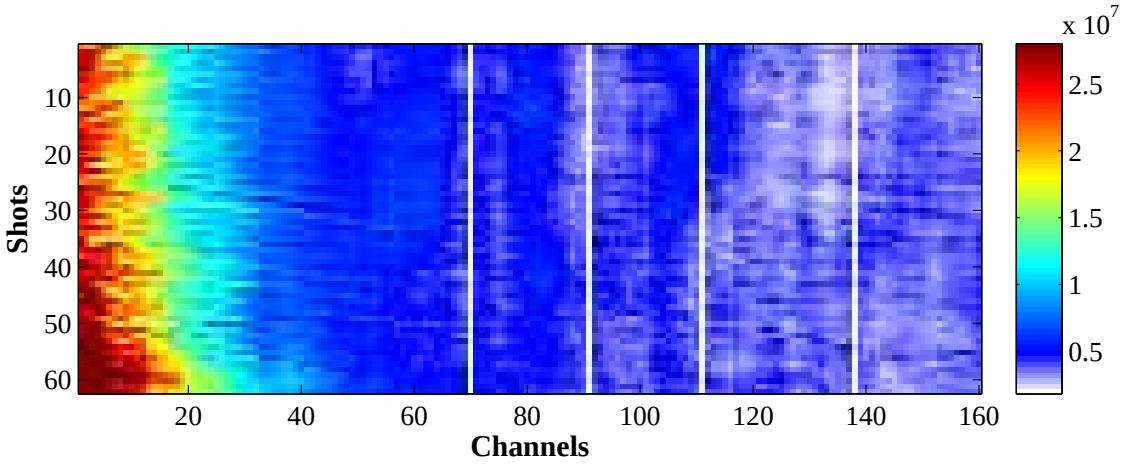
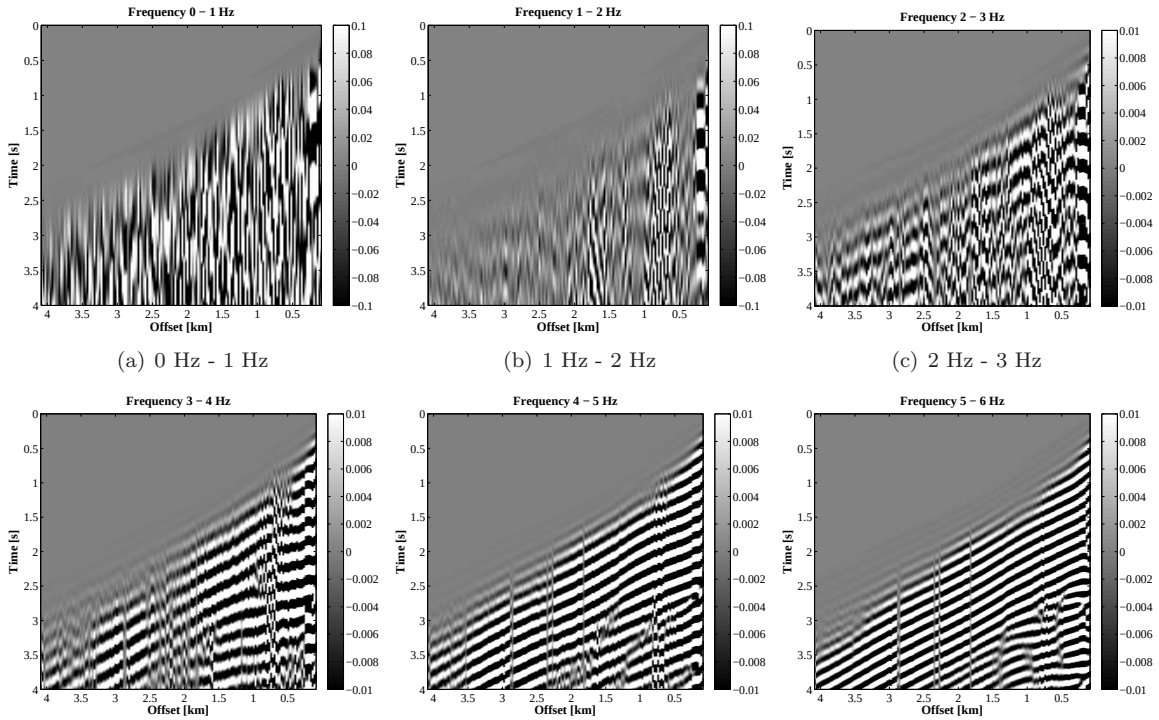


Figure 7.4: The root-mean-square (RMS) trace amplitude of the raw data. The x-axis is the channel number, the y-axis is the shot number. Every fourth shot is plotted, resulting in the shot spacing of 100 m.

produce artefacts in the inverted models. Hydrostatic pressure variations are related to the change of water column height over the seismic streamer and are caused by water swells and the vibration of a streamer. The frequency content of the hydrostatic pressure variations is limited up to 1-2 Hz. The swell noise is the non-coherent high-amplitude noise, that has large amplitude at relatively low frequencies 2-10 (15) Hz. It usually affects a number of neighboring traces and can be observed as vertical stripes in seismic data (Figure 7.2b). The swell noise originates from the sea-surface waves and vibrations in the streamer, which are caused by turbulent water related to the sea-surface waves. Low frequencies dominated by the hydrostatic pressure variations and swell noise could be removed by applying a low-cut filter. However, since the presence of low frequencies in recorded data is crucial for the convergence of FWI, the standard low-cut filters that remove unwanted noise together with the large parts of the useful seismic signal are not suitable.

To analyse the quality of the recorded data and to define the lowest frequency for FWI, it is useful to plot narrow frequency bands of the raw data (Figure 7.5). The data are filtered with the Butterworth band-pass filter, with an order $n = 6$. Very low frequencies, below 3 Hz, are completely dominated by strong noise and therefore not useful for the waveform inversion. The signal-to-noise ratio is better for the frequency range from 3 Hz to 4 Hz, although there is still a significant amount of noise present in the data. When we analyse the amplitude spectrum of the original data (Figure 7.7a), we can observe that the frequency content of the swell noise is limited to frequencies below 8 Hz.

To suppress this low frequency noise, I applied a time-frequency de-noising algorithm (TFDN) proposed by Elboth et al. [2008]. It is a localized method that identifies the parts of the frequency spectrum that are affected by noise, and afterwards estimates the actual amplitude of the affected frequency. Figure 7.6a shows the band-pass filtered (3-8 Hz) original shot gather before noise suppression and after applying the TFDN algorithm (Figure 7.6c). Although the highest amplitudes of the swell noise are attenuated (Figure 7.7b), there is still a significant amount of noise left in the data. Becuase the TFDN algorithm requires a careful choice of parameters, this poor performance might be caused by incorrect parameter settings. The difference plot illustrates the extracted noise (Figure 7.6d). The



(a) 0 Hz - 1 Hz (b) 1 Hz - 2 Hz (c) 2 Hz - 3 Hz

(d) 3 Hz - 4 Hz (e) 4 Hz - 5 Hz (f) 5 Hz - 6 Hz

Figure 7.5: Bandpass filtered raw shot gather 1.

swell noise removed from the field data with the TFDN algorithm was used in Chapter 3 to analyse the performance of different minimization criteria in the presence of noise.

Another possibility to remove the swell noise from seismic data is to apply $f-k$ dip filtering. The noise can be separated from the reflected and refracted waves on the basis of their dips in the frequency-wavenumber $f-k$ domain. To remove the swell noise, the field data were transformed to the $f-k$ domain using the 2D Fourier transformation and a dip filter was applied such that all events with a dip higher than 25 ms/trace were rejected. With the receiver spacing of 25 m, the dip of 25 ms/trace corresonds to rejecting the events with the apparent velocity lower than 1000 m/s. Figure 7.6b shows the shot gather after $f-k$ dip filtering and the corresponding amplitude spectrum is illustrated in Figure 7.7c. We can observe that the $f-k$ dip filter was very efficient in removing the swell noise from the recorded data.

7.2.5 3D to 2D transformation

The field data represent the wave propagation in a 3D medium, whereas the inversion algorithm is based on the 2D forward modelling. In the 2D modelling an explosive source acts not as a point source but as a line source. This difference in the nature of the source results in a different amplitude decay with offset and in a phase misfit. It means that the field data (point source) cannot be directly compared with the modelled data (line source) and it is necessary to transform field seismograms such that they reflect wave propagation in a 2D medium.

To transform the field seismograms, I used the following 3D to 2D correction (for details see

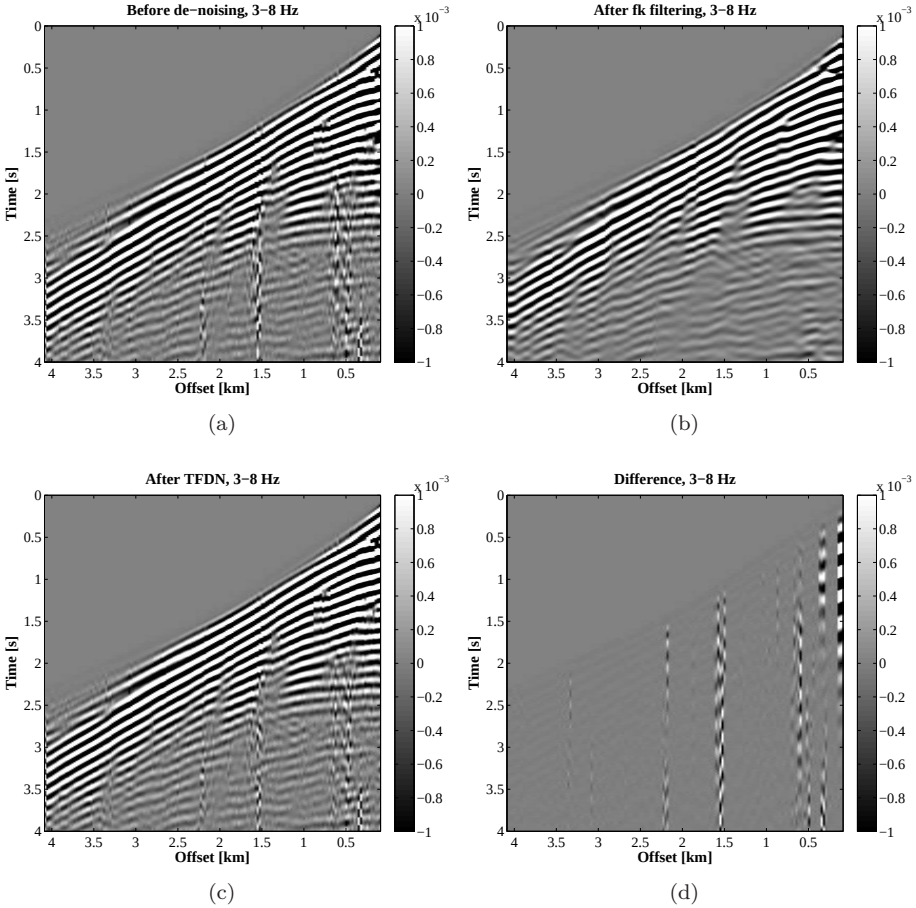
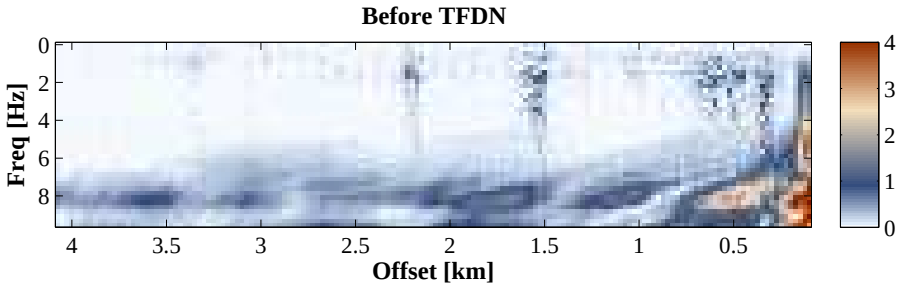
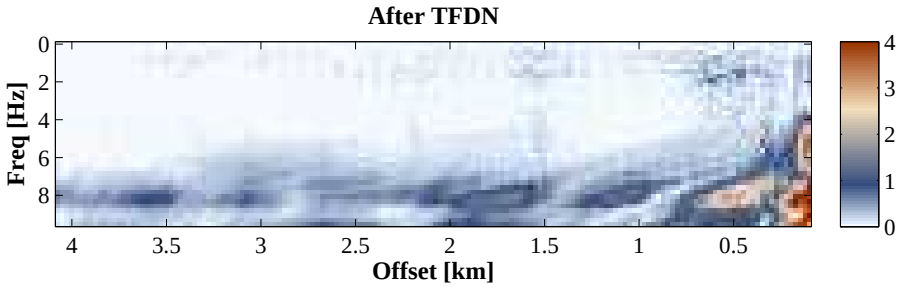


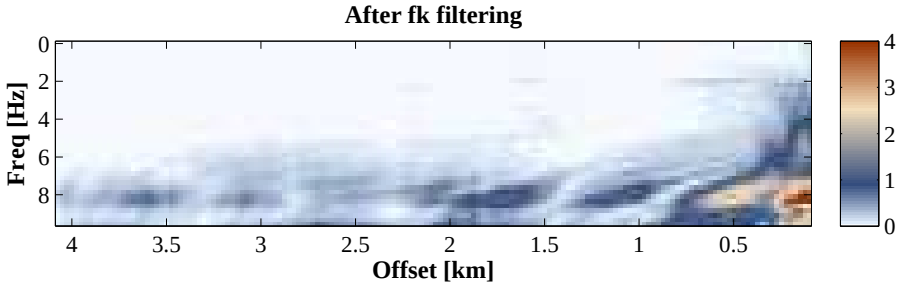
Figure 7.6: Noise suppression. (a) Band-pass filtered shot gather 43 (3 Hz - 8 Hz) before noise suppression, (b) after $f - k$ dip filtering, (c) after applying the TFDN algorithm, (d) the difference between (a) and (c).



(a)

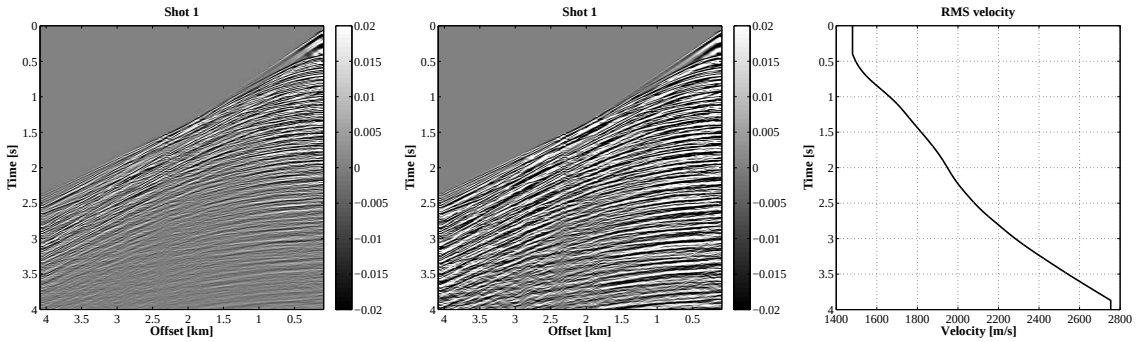


(b)



(c)

Figure 7.7: Amplitude spectrum of the shot gather 43 (a) before noise suppression, (b) after applying the TFDN algorithm, (c) after $f - k$ dip filtering.



(a) (b) (c)

Figure 7.8: 3D to 2D transformation. Shot gather 1 (a) after noise suppression, (b) after 3D to 2D transformation. (c) The RMS velocity function.

Chapter 6)

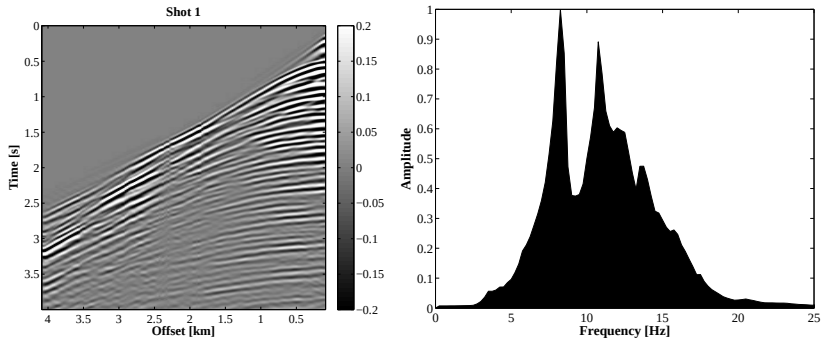
$$C_2 \propto D_{-1/2}(t) V_{\mathrm{rms}} \sqrt{t}\,. \tag{7.1}$$

where $D_{-1/2}(t)$ is the half-integrator, which corresponds to the convolution with $1/\sqrt{t}$ in the time domain, and V_{rms} is the time-dependent root-mean-square velocity. The RMS velocity was calculated from the starting P-wave velocity model (from VSP data) (Figure 7.8c). The first term of the 3D to 2D correction accounts for the phase misfit and the second term is a time-dependent amplitude spreading correction. The field data before and after the 3D to 2D correction are shown in Figure 7.8a,b.

7.2.6 Frequency filtering

Because the attenuation effects are not taken into account in the forward modelling, the use of high frequencies for acoustic FWI is limited. In general, seismic attenuation increases with frequency, i.e. the high-frequency components of seismic waves are attenuated more rapidly during wave propagation than the low-frequency components. The frequency-dependent attenuation produces the change in the amplitude and phase spectra of recorded signal [Gadallah and Fisher, 2005]. As shown by Kurzmann [2012], the pure acoustic inversion of viscoacoustic data produces strong artefacts in the reconstructed V_P models, when the higher frequency components (peak frequency of the signal higher than 5 Hz) are included. Because the low-frequency components are less sensitive to attenuation, Shipp and Singh [2002] limited the frequency content of the marine streamer data up to 15 Hz prior to FWI. To improve the final velocity model, Takam Takougang and Calvert [2011] introduced attenuation for the inversion of frequencies higher than 10.5 Hz.

In this study, the 3D-to-2D transformed data are filtered with a 6th-order bandpass Butterworth filter with a passband from 3 to 14 Hz. The aim of the filter is to remove the very low frequencies with a poor signal-to-noise ratio and to limit the maximum frequency content of the data, such that the attenuation effects are weaker. The exemplary, filtered shot gather and the associated amplitude spectrum are shown in Figure 7.9.



(a) (b)

Figure 7.9: (a) Exemplary shot gather after filtering with a 6th-order bandpass Butterworth filter with a passband from 3 to 14 Hz, (b) the associated normalized amplitude spectrum. Spectrum averaged over 160 traces, time window: 0 to 4000 ms.

7.2.7 Shot balancing

To compensate for the difference in source energy between shots, it is necessary to perform an amplitude balancing of the data. This was achieved by scaling each shot gather by a scalar, so that the amplitude of the direct wave at the near offset trace is constant for all shot gathers. Unlike the offset-dependent amplitude scaling, this approach preserves the true seafloor reflectivity and the variation of amplitudes with offset. The offset-dependent amplitude scaling is often applied to 2D acoustic inversion of real data to compensate for the differences between field and synthetic data [Bleibinhaus and Rondenay, 2009; Brenders and Pratt, 2007b; Jaiswal et al., 2008]. However, the similar effect can be obtained with the L_2 norm using the normalized wavefields, because this minimization criterion accounts for different energy levels between the observed and synthetic data.

7.2.8 Time windowing

To add stability to the inversion, a time window can be applied to the early arrivals [Pratt and Worthington, 1988]. As shown by Sirgue [2003], early arrivals are more linear than the late arrivals. Furthermore, time windowing of the data around the first arrivals is compatible with the reconstruction of the large and intermediate wavelengths of the model and helps to reduce the risk of cycle skipping. In contrast to the frequency domain approach, time windowing of the data is straightforward in the time-domain FWI. To remove the late arrivals, which increase the non-linearity of the inverse problem, a time window of 1.5 s was applied after the first arrival (central panel in Figure 7.24.

7.3 Starting models

As mentioned before, the success of the inversion depends on the accuracy of the initial velocity model. In general, FWI requires the definition of a smooth background velocity model that will fill in the gap of low frequencies missing in the data. Due to the lack of frequencies below 3 Hz in the field data, it is necessary to generate an initial velocity model that already contains the long wavelength features of the subsurface. In this study, I generated three different starting velocity models:

- from the VSP data,
- from the NMO stacking velocities,
- from the refraction traveltime tomography.

As shown in Chapter 5, the density information is also required to provide realistic reflection coefficients in marine environment. The starting density models are computed from initial velocity models using Gardner's velocity-density relationship (Eq. 5.7).

7.3.1 Seafloor topography

A good description of the seafloor topography is required for the correct modelling of the seafloor reflections and the surface-related multiples. These waves have high amplitudes and thus they dominate the data residuals. Because the seabed depth measurements using sound pulses (sonar) are not available for this data, the seabed depth was estimated from the recorded data. The plot of the nearest offset channels (quasi-zero-offset section) indicates a slight variation of the seafloor topography (Figure 7.10a).

Seafloor topography was automatically determined from the arrival time of the seafloor reflection at the near offset trace. Based on the picked travel time T_0 and on the water velocity V_w, the travelpath of the seafloor reflection s is calculated as $s = V_\mathrm{w} T_0$. The corresponding depth of the seafloor H_sea, assigned to the midpoint between the shot and the channel, is calculated as

$$H_\mathrm{sea} = \sqrt{(0.5R)^2 - (0.5s)^2}\,, \tag{7.2}$$

where R is the offset between the shot and the channel. In this study, the water velocity estimated from the direct wave $V_\mathrm{w} = 1480$ m/s. Because of the picking errors the reconstructed seafloor topography is very irregular (Figure 7.10b). Therefore, the original topography was smoothed and then interpolated on the model grid.

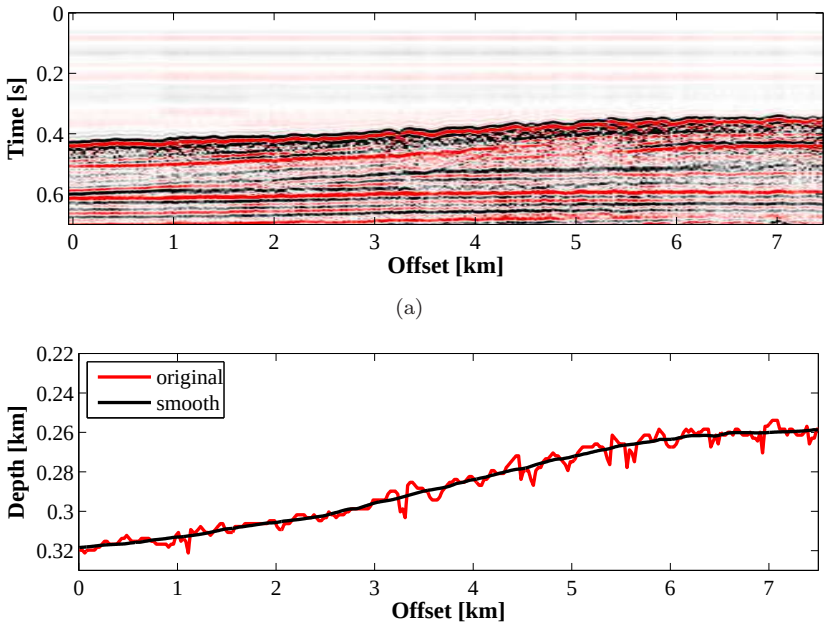
7.3.2 Starting model from the VSP data

The first initial velocity model for the waveform inversion is based on the interval velocities calculated form the vertical seismic profiling (VSP) data available from the wellbore located at $x = 2.3$ km (Figure 7.11a). The original interval velocities were smoothed, such that the individual layers are no longer visible and are replaced by a gradient (Figure 7.11b). The model was filtered with a 2D-Gaussian filter with a size of 100×100 m and $\sigma = 30$. The 2D-Gaussian filter is defined as

$$G(x, z) \sim e^{-\frac{x^2+z^2}{2\sigma^2}}\,, \tag{7.3}$$

where x and z are the spatial coordinates, and σ is the standard deviation.

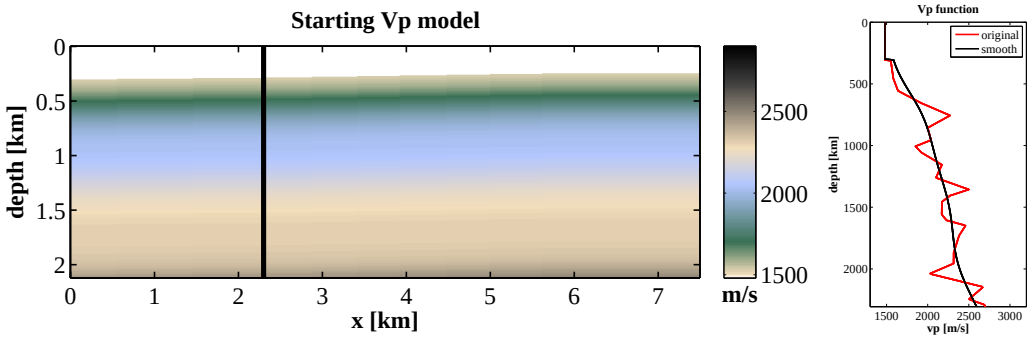
The dimensions of the model are 7.5 km in the x-direction and 2.3 km in the z-direction. P-wave velocity ranges from 1536 m/s at the seafloor to 2600 m/s at a depth of 2.3 km.



(a)

(b)

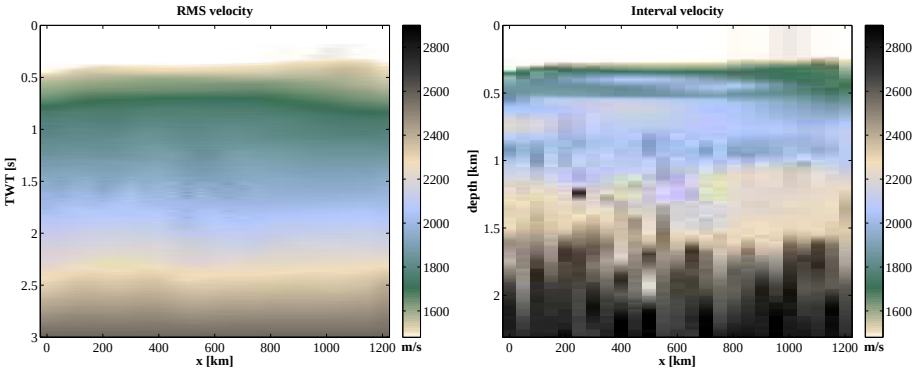
Figure 7.10: Estimation of the seafloor topography.(a) Nearest-offset traces extracted from shots along the profile, (b) estimated seafloor topography; red line - original, black line - smoothed.



(a)

(b)

Figure 7.11: (a) Starting V_P model from the VSP data. The vertical line indicates the borehole location. (b) P-wave velocity function; red line indicates the original interval velocities from the VSP data, black line represents the smoothed velocity function used as a starting model for the inversion.



(a) (b)

Figure 7.12: (a) NMO velocity model, (b) interval velocities computed from the NMO velocity model.

7.3.3 NMO velocity model

The second initial model for FWI is based on the NMO (normal-moveout) velocity. Building a velocity model was part of the conventional processing, which I applied to generate the post-stack time migrated section. The NMO velocity is the velocity required to correct for NMO, i.e. to make the primary reflection on CMP (common-mid-point) gather occur at the same time on all traces. For horizontal layers and assuming that the offset is small compared with the depth, the NMO velocity is equal to the RMS (root-mean-square) velocity.

The NMO/RMS velocities were picked from the semblance plots of 25 CMP (common-mid-point) gathers (Figure 7.12a) and then converted to interval velocities (Figure 7.12b). The interval velocities can be calculated from RMS velocities V_{rms} using Dix formula, such that

$$V_i = \frac{V_{\text{rms};i}^2 T_{0,i} - V_{\text{rms};i-1}^2 T_{0,i-1}}{2h_i} \,, \tag{7.4}$$

and

$$h_i = \frac{T_{0,i} - T_{0,i-1}}{2} V_i \,, \tag{7.5}$$

where $T_{0,i}$ is the two-way traveltime at zero offset for a reflection at ith interface and $T_{0,i-1}$ is the TWT at zero offset for a reflection at $(i-1)$ interface, V_i is the interval velocity of the ith layer, and h_i is the thickness of the ith layer.

To obtain the starting model for FWI (Figure 7.15c), the original interval velocity model was interpolated and smoothed with a 2D Gaussian filter (size 150×150 m, $\sigma = 30$).

7.3.4 Starting model from the refraction traveltime tomography

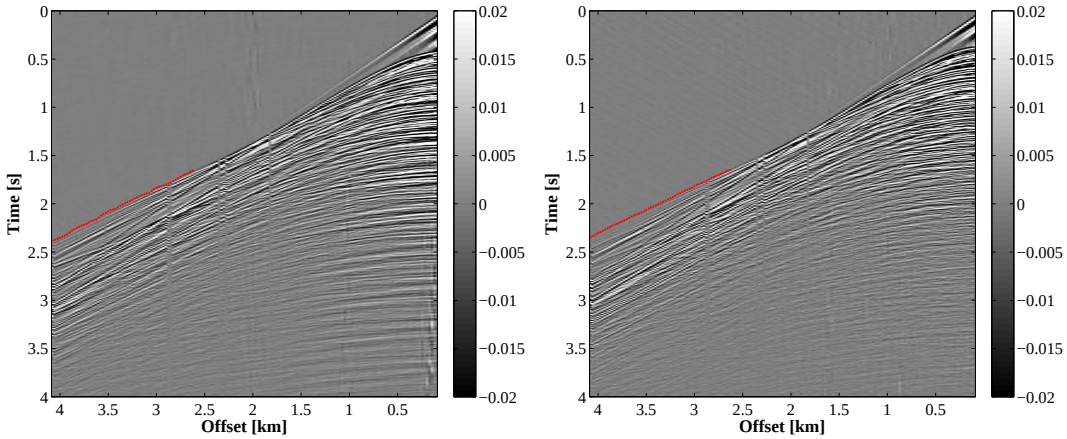
To generate the third starting model for FWI, I performed the refraction traveltime tomography using the ReflexW software. Because all sources and receivers are located within one line at the surface, a curved raytracing must be used for the calculation of the traveltimes. The curved rays are calculated using a finite difference approximation of the Eikonal equation. The calculation of the synthetic traveltimes is restricted to the first arrivals, and no reflections or secondary arrivals are used. The tomographic algorithm is based on an iterative adaptation (SIRT -Simultaneous Iterative

Reconstruction Technique). The model update is derived from the traveltime residuals between the synthetic and picked traveltimes.

As an input for the refraction traveltime tomography, I used the hand-picked first-arrival traveltimes at offsets larger than 2.6 km. These first-arrivals correspond to the refracted/diving waves. The first shot is located at $x = 4.25$ km and the last shot is at $x = 11.65$ km, with the shot spacing of 100 m. The total number of traveltime picks is 3675. Figure 7.13 shows exemplary shot gathers with traveltime picks.

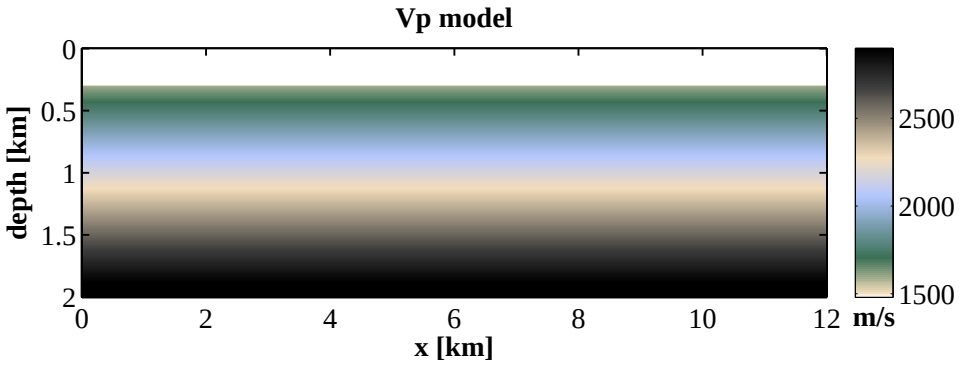
I applied a two stage traveltime tomography. The first starting velocity model is a simple 1D model with a 300 m thick water layer above a linear velocity gradient (Figure 7.14a). The dimensions of the model are 12 km in the x-direction and 2 km in the z-direction. The model for the traveltime tomography was extended to the left by 4.5 km in comparison to the actual model used in FWI. The main reason was to obtain a sufficient ray coverage at the left side of the model and to include far offset traces for all shots. The space increment is set to 50 m both in x- and in y-direction. To allow for a high ray coverage within the medium vertical velocity gradients must be present. The velocity at the seafloor is increasing from 1600 m/s to 3000 m/s at a depth of 2 km. This results in the velocity gradient of $\partial V_P/\partial z = 0.8$. The RMS travetime misfit for the initial model is 163 ms. The ray density for this model is shown in Figure 7.14b. As we can see the ray coverage is limited to approximately 1000 m. Because of the moving streamer geometry (vessel is moving from the left to the right part of the model) the ray coverage is poor from 0 to 3 km. After 15 iterations step, the RMS traveltime misfit was reduced to 55 ms and the resulting velocity model is shown in Figure 7.14c. Due to the lack of ray coverage, the velocity model was not updated at the edges of the model. On the other hand, the update in the central part of the model is very similar and reveals a 1D velocity distribution below the seafloor.

For the second stage of the traveltime tomography, I modified the first starting velocity model. Because the velocities below the seafloor were too low in the original starting model, the new value of 1680 m/s was extracted from the initial traveltime tomography result. Furthermore, the maximum velocity at a depth of 2 km was reduced to 2700 m/s. This value was based on the starting FWI model obtained from the conventional velocity analysis. The final traveltime tomography result shown in Figure 7.14d was obtained after 15 iteration steps and the RMS traveltime misfit was reduced to 52 ms. The starting model for FWI was built by extracting the average 1D velocity profile from the final traveltime tomography result and by applying the seafloor topography at the top.

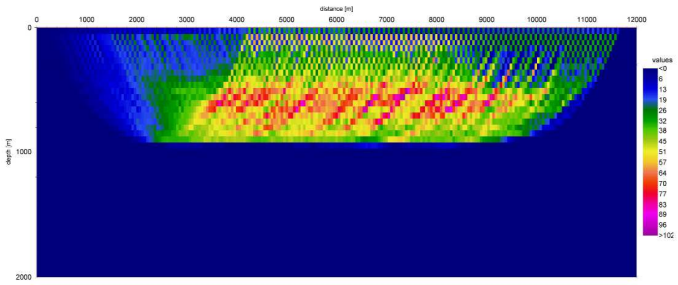


(a) (b)

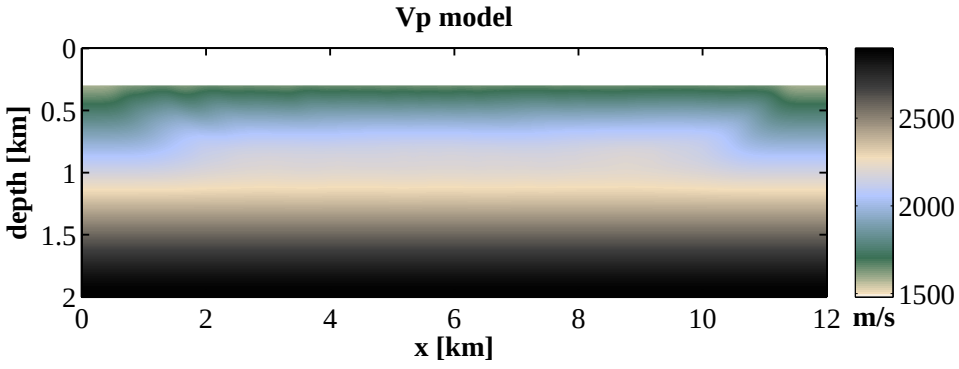
Figure 7.13: Exemplary shot gathers with traveltime picks. (a) Shot gather located at $x = 4.35$ km, (b) shot located at $x = 8.15$ km in the traveltime tomography model.



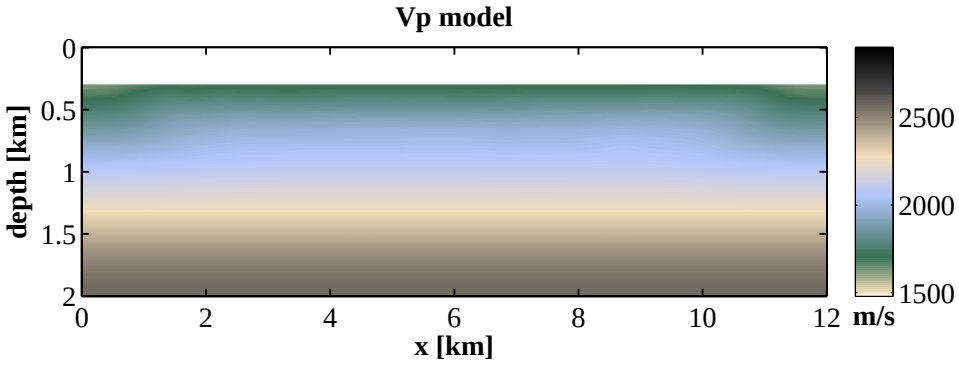
(a)



(b)



(c)



(d)

Figure 7.14: Refraction traveltime tomography. (a) First starting model for the traveltime tomography and (b) the corresponding ray density. (c) The initial traveltime tomography result obtained after 15 iterations with (a) as a starting model. (d) The final traveltime tomography result.

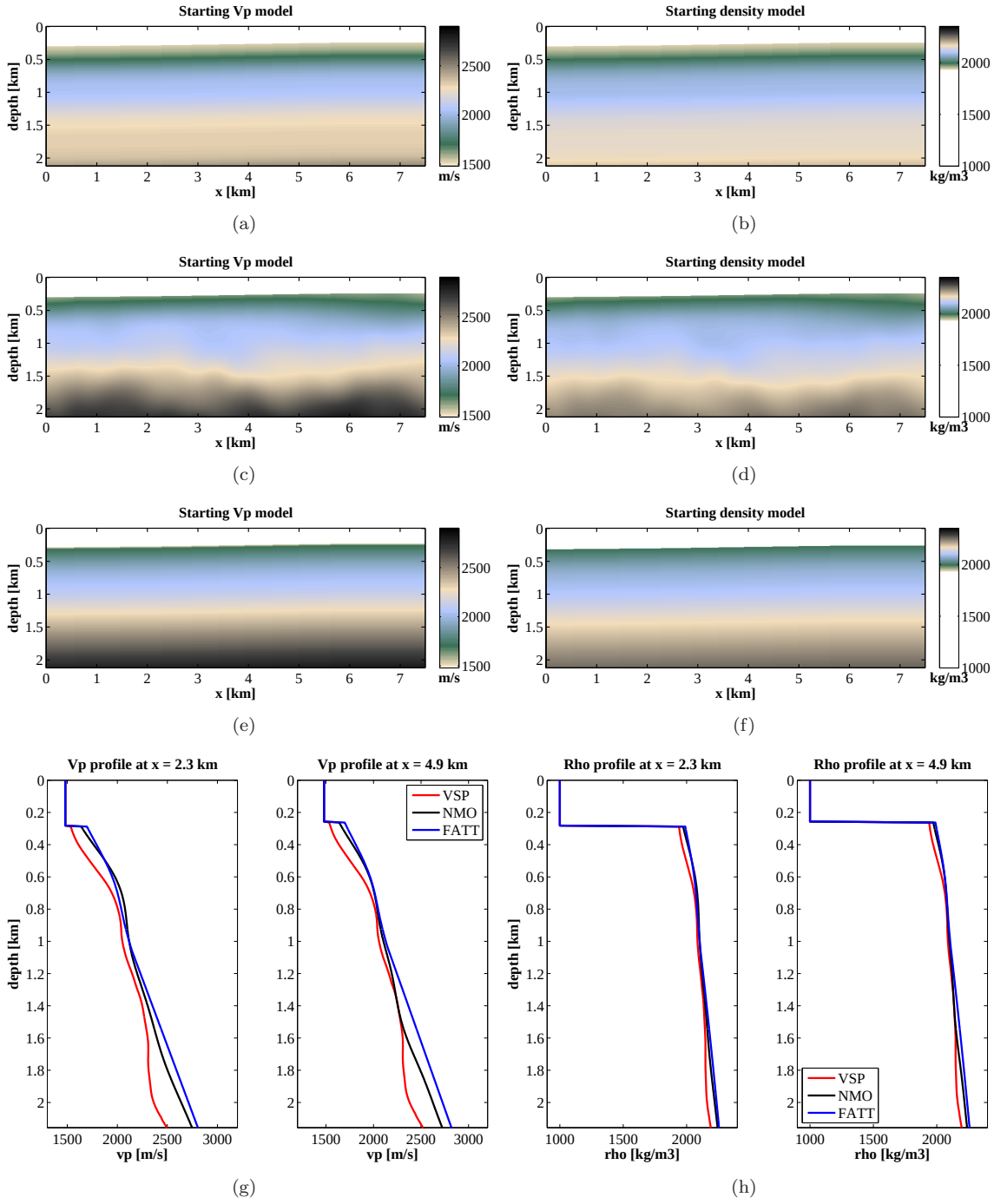


Figure 7.15: Starting velocity models for FWI. (a) Starting V_P model from the VSP data and the corresponding (b) density model. (c) Starting V_P model based on the NMO velocity model and the corresponding (d) density model. (e) Starting V_P model from the refraction traveltime tomography and the corresponding (f) density model. (g) V_P profiles and (h) density profiles at $x = 2.3$ km and at $x = 4.9$ km. The red line corresponds to the initial model from the VSP data, black line corresponds to the model obtained from the NMO velocity, blue line corresponds to the model from the refraction traveltime tomography.

7.4 Initial modelling

Prior to the waveform inversion, the initial forward modelling is performed. It is an important step to evaluate whether the waveform inversion is likely to succeed. If the observed waveform is not matched to within a half-cycle by the synthetic data generated from the starting model, the algorithm will attempt to fit the calculated events to the wrong cycle of the observed data and it will cause the objective function to converge to a local rather than a global minimum.

7.4.1 Modelling parameters

To account for multiple reflections and the ghost effects present in the data, the free surface is modelled using the so-called vacuum formalism. The velocity and density models are discretized on a mesh with 1200×400 grid points, using 6.25 m grid spacing. The detailed modelling and acquisition parameters are listed in Table 7.1.

7.4.2 Starting model requirements

Based on the convergence criterion (Eq. 2.49), it is possible to estimate the maximum traveltime error δt for a given frequency and the propagation distance. The maximum frequency in the first frequency band for the multi-scale FWI is $f_1 = 5$ Hz. For a wave propagating at the maximum offset of 4000 m, an approximate maximum propagation distance expressed in terms of wavelength is given by $N_\lambda = 13.5$. This means that the relative traveltime error $\delta t/T$ in the starting model must be less than 0.037. Taking the arrival time of the refracted/diving wave at the maximum offset, $T = 2.5$ s, we obtain the following condition

$$\delta t \text{ (s)} < \frac{T}{2N_\lambda} = 0.092 \tag{7.6}$$

Because the RMS traveltime misfit of the starting model from the refraction traveltime tomography is 52 ms, it means that this initial model satisfies the convergence requirements for the first frequency band.

7.4.3 Source wavelet

Source wavelets for the initial modelling were estimated using a linear least-squares inversion described in Chapter 6. Initial source wavelets were calculated individually for three different starting models and for two frequency bands: $3 - 5$ Hz, which represents the first frequency band for FWI, and for the full frequency band $3 - 14$ Hz.

We can observe small differences between source signatures estimated for different starting models, especially for the low frequency band (Figure 7.16). The accuracy of the reconstructed source wavelet depends on the inaccuracies in the velocity model, i.e. the velocity errors are propagating into the source wavelet. Because the starting V_P model from the refraction traveltime tomography is very similar to the starting model obtained from the NMO velocity (Figure 7.15g), the corresponding wavelets are similar as well.

Starting model for the field data inversion	
Model size	7.5 km x 2.5 km
Average V_P	2210 m/s
Minimum V_P	1480 m/s
Maximum V_P	2917 m/s
Water layer thickness	260 - 320 m
Acquisition parameters	
Number of shots	62
Shot spacing	100 m
Shot depth	9.375 m
Max number of hydrophones in a streamer	160
Hydrophone spacing	25 m
Hydrophone depth	9.375 m
Minimum offset	75 m
Maximum offset	4050 m
Modelling parameters	
Grid size	1200 x 400
Grid spacing	6.5 m
Time sampling	8e-4 s
Number of time samples	5000
Recording length	4 s
Source wavelet	estimated from the field data

Table 7.1: Modelling and acquisition parameters for the field data inversion. The average, minimum, and maximum V_P values are defiend for the starting model from the refraction traveltime tomography.

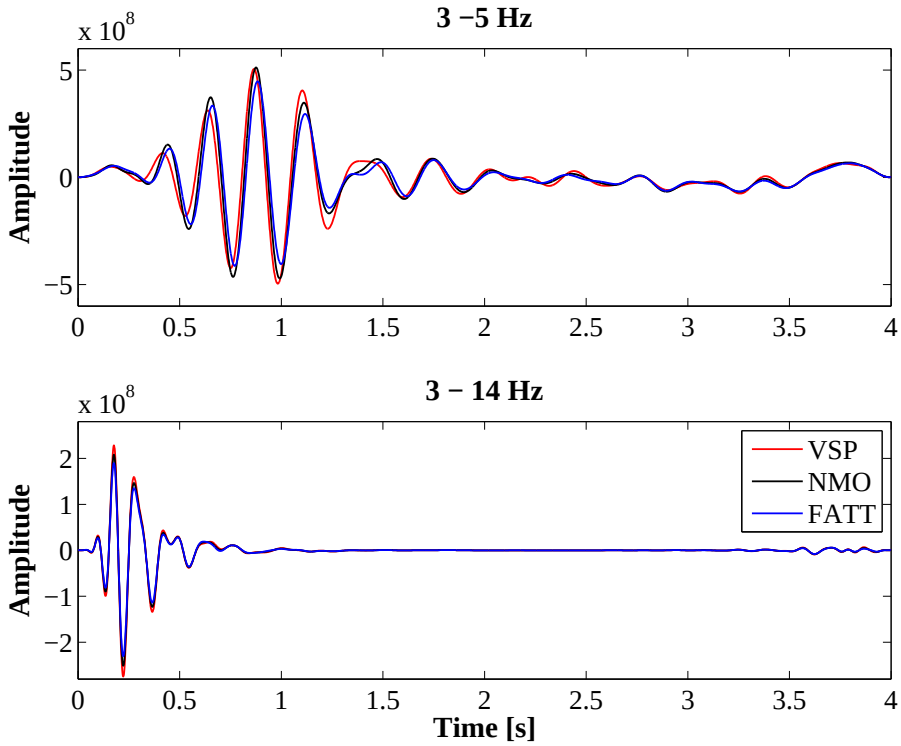


Figure 7.16: Initial wavelets estimated for different starting models. Top: frequency band of $3-5$ Hz; bottom: frequency band of $3-14$ Hz. The red line corresponds to the model the VSP data, black line corresponds to the model obtained from the NMO velocity, blue line corresponds to the model from the refraction traveltime tomography.

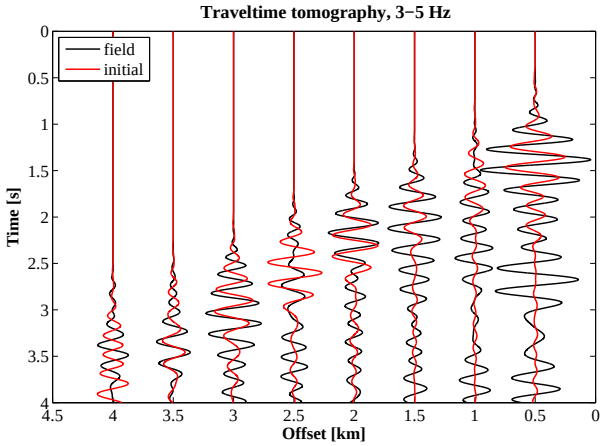
7.4.4 Data comparison

Figure 7.17 shows the comparison between the observed and initial data generated for three different starting models for the frequency band of $3-5$ Hz. We can observe a significant mismatch between the field data and the modelled data at far offsets, when the starting model from VSP data is used (Figure 7.17a). The predicted traveltimes at far offsets are too late, which indicates that the velocities in the starting model are too low. This poor fit indicates that the initial model from the VSP data is not accurate enough for the waveform inversion with the first frequency band of $3-5$ Hz. The waveform fit is much better in the case of the model obtained from the NMO velocity (Figure 7.17b). However, the best match between the field data and initial synthetics is provided by the starting model from the refraction traveltime tomography (Figure 7.17c). The main differences between the initial synthetics concern the modelling of refracted/diving waves at far offsets (Figure 7.19), which are also the most non-linear components of the data and are very likely to be cycle skipped [Sirgue, 2003].

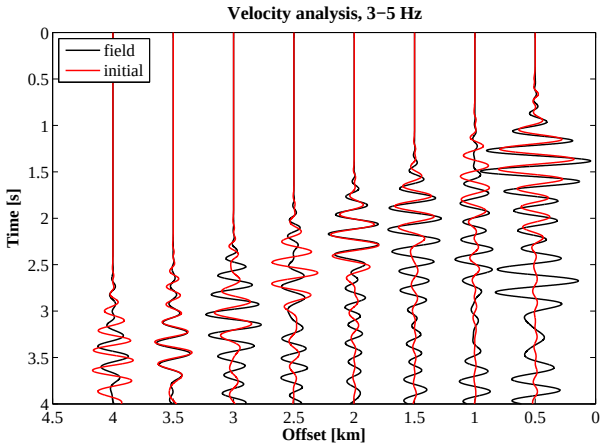
The comparison between the field data and initial synthetics for the full frequency band is shown in Figure 7.18 and in Figure 7.19 (right panel). At higher frequencies the waveform is more complex and the requirements for the starting model are increasing. Even though the early arrivals at near offset are well matched, there is a problem with cycle-skipping at middle and far offsets. It is obvious that the high frequency components can not be included at the early stages of FWI.

Finally, it is useful to compare the variation of the RMS amplitude with offset of the observed data and synthetic data. The RMS amplitude versus offset trend shows a reasonably good fit between the field data after preprocessing and modelled data for the starting frequency band of $3-5$ Hz and offsets from 0.3 km to 4.0 km (Figure 7.20a). This small amplitude discrepancy indicates that no additional offset-dependent amplitude scaling is required for this data set. The significant amplitude misfit at near offset is most likely due to the far-field approximation of the 3D to 2D transformation applied to the field data. To avoid these high residuals, the nearest offset traces (< 300 m) are precluded from inversion.

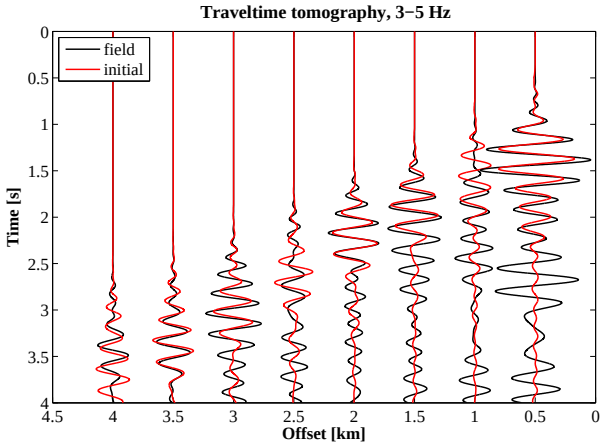
In case of the full frequency band of $3-14$ Hz, the relative RMS amplitude differences between the recorded and modelled data are more significant (Figure 7.20b). We can observed sharp amplitude variations of the field data, which has higher RMS amplitudes than the initial synthetics at near and middle offset below 2 km. The differences between starting models are more evident as well.



(a)

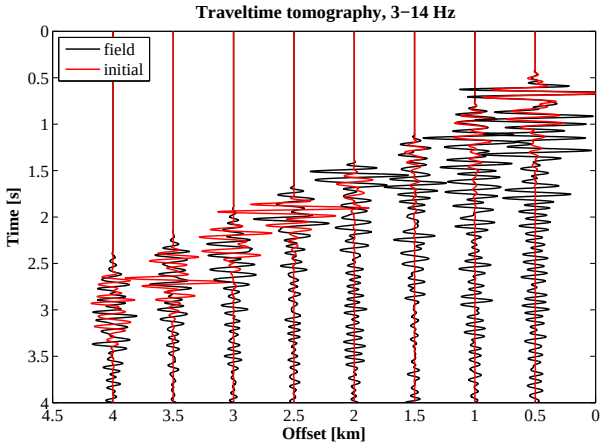


(b)

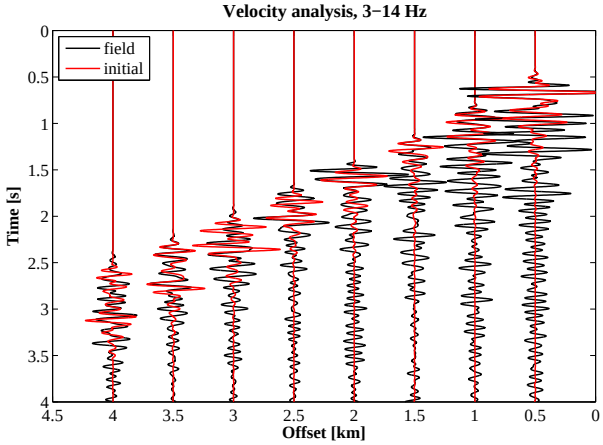


(c)

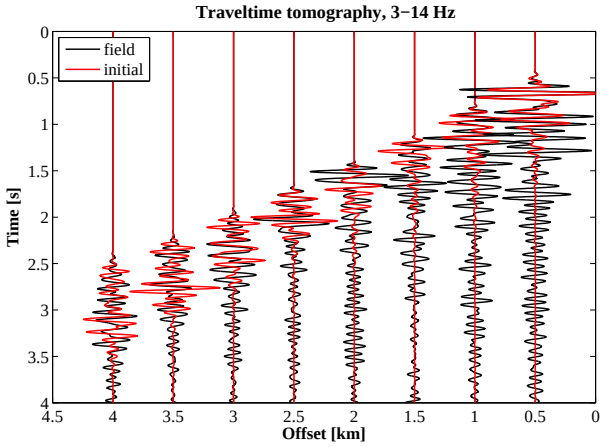
Figure 7.17: Initial modelling. Frequency band of $3-5$ Hz. Trace comparison between the observed data (black) and the initial data (red) computed for: (a) starting V_P model from the VSP data, (b) Starting V_P model based on the NMO velocity model, (c) starting V_P model from the refraction traveltime tomography. Shot 62 located at $x = 7.275$ km.



(a)



(b)



(c)

Figure 7.18: Initial modelling. Frequency band of $3-14$ Hz. Trace comparison between the observed data (black) and the initial data (red) computed for: (a) starting V_P model from the VSP data, (b) starting V_P model based on the NMO velocity model, (c) starting V_P model from the refraction traveltime tomography. Shot 62 located at $x = 7.275$ km.

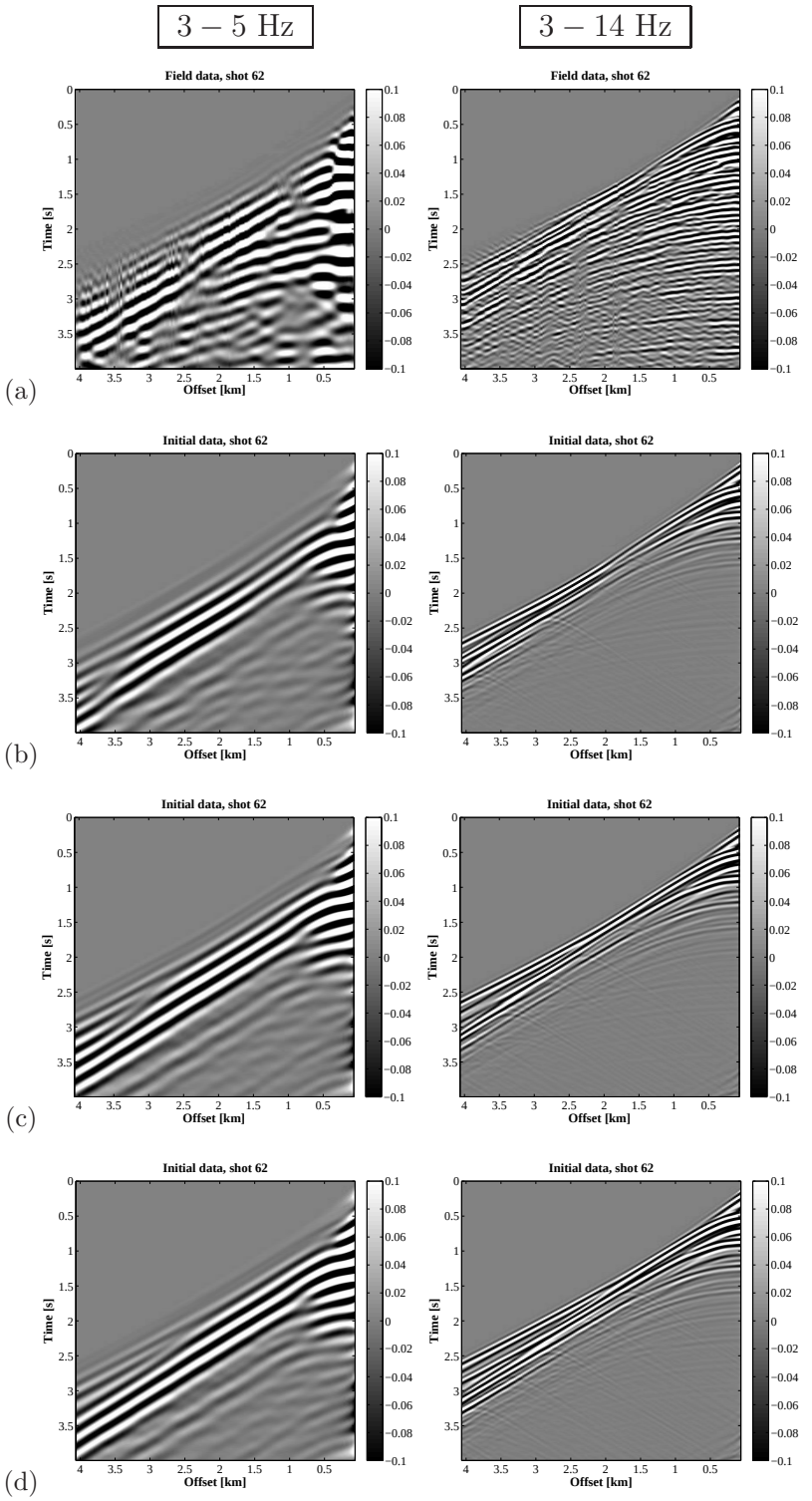
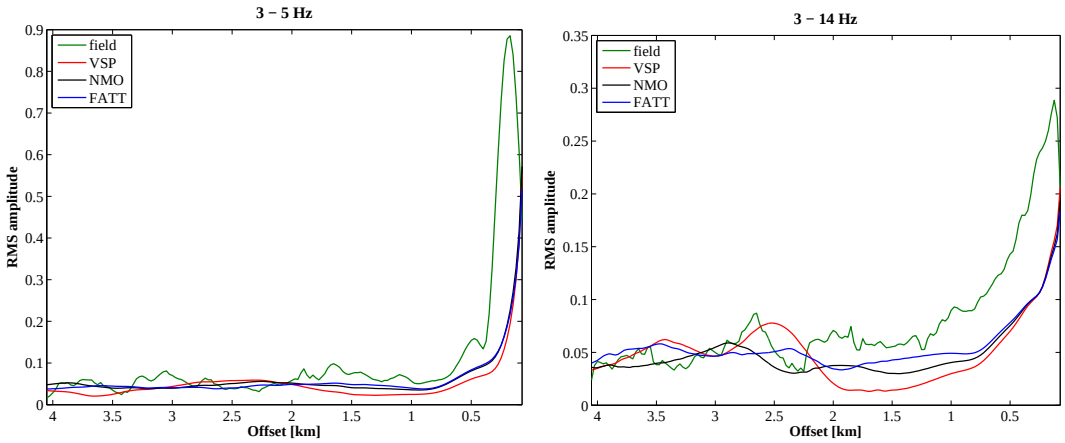


Figure 7.19: Initial modelling - comparison of shot gathers. Left: frequency band of $3-5$ Hz, right: frequency band of $3-14$ Hz. (a) Field data, (b) starting V_P model from the VSP data, (c) starting V_P model based on the NMO velocity model, (d) starting V_P model from the refraction traveltime tomography. Shot 62 located at $x = 7.275$ km.



(a) (b)

Figure 7.20: RMS amplitude variation with offset of the field data (green) and of the initial data computed for three different starting models for shot 62 located at $x = 7.275$ km. (a) Frequency band of $3 - 5$ Hz, (b) frequency band of $3 - 14$ Hz. Shot 62 located at $x = 7.275$ km.

7.5 Inversion parameters

To reduce the high complexity of the inverse problem, the multi-scale inversion approach described in Chapter 3 is applied. The maximum frequency for each frequency band was calculated using Eq. 3.13. With the depth to the target $z = 2.0$ km, and the maximum half-offset $h_{\max} = 2.025$ km, we obtain $\alpha_{\min} = 0.7$. Because the signal-to-noise ratio is relatively poor at low frequencies, the maximum frequency of the first frequency band is set to 5 Hz. The four frequency bands are applied sequentially with the following maximum frequencies $f_{\max} = (5.0, 7.1, 10.1, 14.5)$, which define the cutoff frequency for the $6th$ order low-pass Butterworth filter.

The minimum and maximum vertical wavenumbers can be computed using Eq. 3.10, with $c_0 = 2210$ m/s, which corresponds to the average velocity of the starting model from the refraction traveltime tomography. For the first maximum frequency $f_1 = 5$ Hz, we obtain

$$k_{z\,\min}(f_1) = 0.02 \text{ rad} \cdot \text{m}^{-1} ,$$
$$k_{z\,\max}(f_1) = 0.0284 \text{ rad} \cdot \text{m}^{-1} .$$

With $k = 2\pi/\lambda$, where λ is the wavelength, the minimum and maximum wavelengths related to the vertical resolution are

$$\lambda_{\min}(f_1) = 221 \text{ m} ,$$
$$\lambda_{\max}(f_1) = 314 \text{ m} .$$

For the frequency of $f_3 = 10.1$ Hz, we get

$$\lambda_{\min}(f_3) = 109 \text{ m} ,$$
$$\lambda_{\max}(f_3) = 156 \text{ m} .$$

For the maximum frequency of $f_4 = 14.5$ Hz, the minimum and maximum wavelengths related to the vertical resolution are

$$\lambda_{\min}(f_4) = 76 \text{ m} ,$$
$$\lambda_{\max}(f_4) = 109 \text{ m} .$$

To correct for the amplitude loss with depth due to geometrical spreading and to enhance deeper parts of the model, the linear gradient scaling with depth is implemented (P_2 operator described in Chapter 3). Furthermore, the taper in the water layer is limited to 25 m above the seafloor, such that the seafloor parameters (V_P, depth) can be updated during inversion.

7.5.1 Data decimation and aliasing

A total number of shots, along with a model size and a recording length, are the main parameters affecting the computation time of FWI. If the measured data are densly sampled, they can be decimated for the inversion.

Brenders and Pratt [2007a] investigated the effect of sparse source and receiver sampling on waveform inversion results. They showed that the image quality is reduced by aliasing of the data due to insufficient source or receiver sampling. The condition for unaliased surface sampling is given by

$$\Delta_{\text{samp}} \leq \frac{1}{2k_{\max}} , \tag{7.7}$$

where Δ_{samp} is the spatial sampling interval, and $k_{\max}$ is the maximum horizontal data wavenumber. At a given frequency f and a minimum velocity of a medium $c_{\min}$, the spatial wavenumber is

$$k = \frac{f \sin\theta}{c_{\min}} , \tag{7.8}$$

where θ is the emerging angle of a propagating wave. For an emerging angle $\theta = \pi/2$, we get

$$k = k_{\max} = \frac{f}{c_{\min}} = \frac{1}{\lambda} . \tag{7.9}$$

Finally, we obtain the condition

$$\Delta_{\text{samp}} \leq \frac{\lambda}{2} , \tag{7.10}$$

which means that no spatial aliasing occurs if the wavefield is sampled with a source spacing $\Delta s \leq \Delta_{\text{samp}}$, and a receiver spacing $\Delta r \leq \Delta_{\text{samp}}$.

The sparseness of the data can be also expressed in terms of the aliasing number N_A [Bleibinhaus et al., 2009]

$$N_A = \frac{\max\{\Delta s, \Delta r\}}{\Delta_{\text{samp}}} , \tag{7.11}$$

i.e. the spatial aliasing occurs when $N_A > 1$.

In this study, an unaliased wavefield is obtained when Δs and Δr is less than 148 m for the maximum frequency of the first frequency band $f = 5$ Hz; for the maximum frequency of 10 Hz: $\Delta_{\text{samp}} = 74$ m, and for $f = 14$ Hz: $\Delta_{\text{samp}} = 52$ m. Because the wavefield usually does not propagate

with the horizontal incidence angle, as assumed in Eq. 7.9, the good quality image can be obtained for a shot spacing $\Delta s \leq 3\Delta_{\text{samp}}$ and a receiver spacing $\Delta r \leq \Delta_{\text{samp}}$ [Brenders and Pratt, 2007a]. It means that the receiver spacing of 25 m, and the selection of every fourth shot, which results in the shot spacing of 100 m, are sufficient to obtain good quality inversion results. This acquisition geometry gives the aliasing number $N_A = 0.67$ for the maximum frequency of 5 Hz, $N_A = 1.35$ for $f = 10$ Hz, and $N_A = 1.9$ for $f = 14$ Hz.

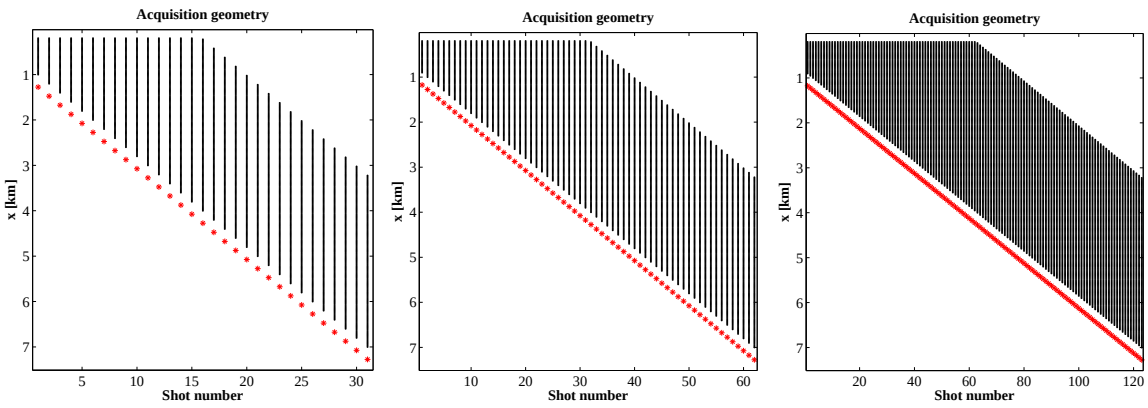
7.6 Resolution study

To test the resolving power and the reliability of the FWI results, I performed a set of synthetic inversion tests using the same acquisition setup, source signature, and frequency content as in the field example. The resolution tests are useful to determine which parts of the model are well-resolved, to identify artefacts due to source-receiver geometry limitations (subsurface illumination), and to examine the footprint of initial models. Moreover, such tests can be used to verify the choice of inversion parameters.

A true V_P model for the resolution study (Figure 7.23a) was generated by adding $+/-5$ % velocity perturbation to the initial model from the traveltime tomography. The checkerboard anomaly size of 100 m corresponds to the average wavelength that can be resolved with the maximum frequency of 14 Hz. The starting model for the checkerboard test is the initial model from the traveltime tomography. The input data are muted, such that a muting window of 1.5 s after the first arrival is applied. The multi-scale inversion approach comprises four frequency bands and 30 iterations for each frequency stage. To investigate the influence of the shot decimation on the waveform inversion results, I performed three inversion tests: using 62 shots with 100 m spacing, using 31 shots with 200 m spacing, and finally using 123 shots with 50 m shot spacing (Figure 7.21). Due to a moving streamer geometry, the maximum offset is increasing from 975 m, for the first shot located in the left part of the seismic line at 1175 m, to the maximum of 4050 m. As a consequence, the receiver fold is higher in the left and central part and lower in the right part of the model (Figure 7.21).

Figure 7.22 shows both the intermediate inversion results at the end of each inversion stage and the final FWI result. Because of a band-limited source signal, the sharp edges of the checkerboard pattern are not resolved. After the first frequency stage, the checkerboard pattern in well recovered in the central part of the model, but with the increasing frequency and the iteration number the V_P reconstruction in the deep part and at the sides of the model is improved. The smearing effects resulting from a poor subsurface illumination are strong at the edges of the model and significantly limit the accuracy of V_P reconstruction with the increasing depth.

The comparison of final V_P models for different shot spacing is shown in Figure 7.23. The velocity structures reproduced with the shot spacing of 100 m and 50 m are almost identical. With the shot spacing of 200 m and 31 shots (Figure 7.23b), the smearing effect is slightly stronger when compared with the results from 100 m and 50 m spacing. This is due to the aliasing effect descibed in the previous section. For the maximum frequency of 14 Hz, the source spacing $\Delta s = 200$ m is four times higher than the spatial sampling interval $\Delta_{\text{samp}} = 52$ m required to obtain an unaliased wavefield.



(a) (b) (c)

Figure 7.21: Acquisition geometries for the resolution study with: (a) 200 m shot spacing (31 shots), (b) 100 m shot spacing (62 shots), (c) 50 m shot spacing (123 shots). Stars denote shot location and points represent hydrophones.

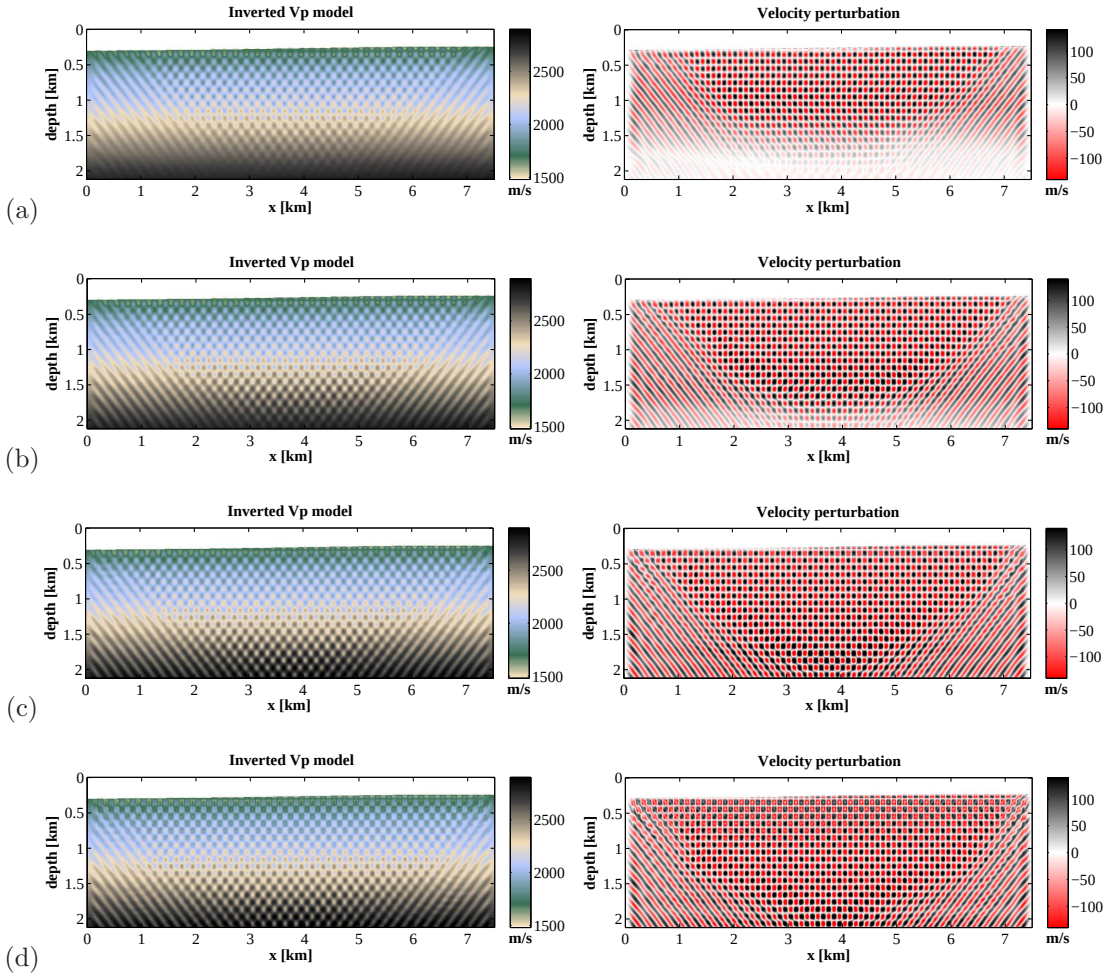


Figure 7.22: Resolution study with a shot spacing of 100 m. Intermediate inversion results: (a) $f_{\max} = 5$ Hz, (b) $f_{\max} = 7.1$ Hz, (c) $f_{\max} = 10.1$ Hz, and (d) the final inversion result $f_{\max} = 14.5$ Hz.

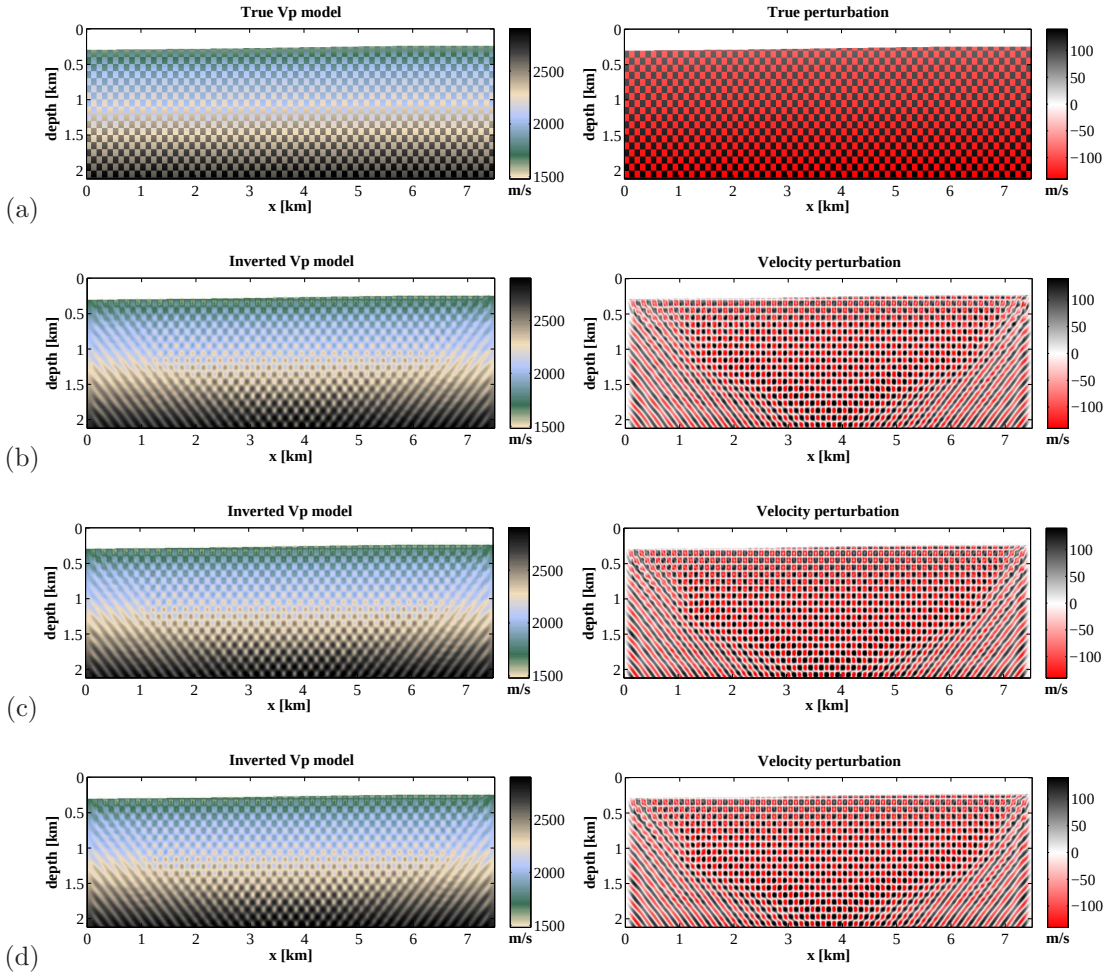


Figure 7.23: Resolution study - the acquisition geometry effect. (a) True model. Final FWI results with a shot spacing of (b) 200 m, (c) 100 m, (d) 50 m.

7.7 Inversion results

7.7.1 Multi-stage inversion

The first set of inversion experiments aims to analyse the effect of different inversion strategies based on time windowing and/or offset windowing of the data on the FWI results. The aim of such hierarchical, multi-stage inversion techniques is to mitigate the non-linearity of the inverse problem. In all inversion tests, the minimization criterion is the approximated version of the L_2 norm using the normalized wavefields (Eq. 3.18). Furthermore, the only inversion parameter is the P-wave velocity and the initial density model is fixed during inversion. The starting model for the inversion is the V_P model obtained from the refraction traveltime tomography.

I performed six inversion tests:

- **Test 1** - there is no time windowing applied, thus the full available data up to 4 s are inverted at every inversion stage. Table 7.2.
- **Test 2** - a fixed time window of 1.5 s is applied to the data to damp the late arrivals (Figure 7.24). This time damping reduces the amount of data at near and at middle offsets. Table 7.3.
- **Test 3** - an extending time windowing. For each frequency band, the time window after the first arrival is sequentially increasing. In this way, the shallow part of the model is reconstructed prior to the recovery of the deeper parts of the model (Figure 7.25). Table 7.4.
- **Test 4** - near offset data inversion combined with the extending time windowing. The maximum offset is limited to 2 km. Table 7.5.
- **Test 5** - offset windowing combined with the fixed time window. For each frequency band, the near offset data are inverted before the far offset information is included (Figure 7.26). Because the near offset data corresponds to the shorter propagation distance they are less likely to be cycle skipped than the far offset data. Table 7.6.
- **Test 6** - combination of the offset windowing with the extending time windowing. First the early arrivals at near offsets are inverted, this part contributes to the update of the long wavelength shallow structures. Afterwards the early arrivals at far offsets are included to update the low wavenumbers in the deeper part of the model. In the next part, the time window is increased and the inversion proceeds from the near offset to the far offset data (Figure 7.27). This strategy is similar to the layer stripping inversion approach applied by Takam Takougang and Calvert [2011]; Wang and Rao [2009]. Table 7.7.

Figure 7.24 shows an increasing resolution of the inverted V_P models when higher frequencies are included. Each higher frequency band yields a recovery of a more detailed V_P image. Because the attenuation is not introduced in the acoustic FWI, the inversion of the last frequency band with the maximum frequency of 14 Hz introduced a lot of small scale artefacts into the reconstructed V_P models (Figure 7.24d). For that reason all inversion results are shown after the third inversion stage with the maximum frequency of 10.1 Hz.

A comparison of final inversion results for Test 1 - Test 6 is illustrated in Figure 7.28. In general, all major structures are similar in the reconstructed velocity models, except for the result from the near

Iteration	Frequency (Hz)	Max time (s)	Offset (km)
1-30	3-5	4.0	0.275 - 4.0
31-60	3-7.1	4.0	0.275 - 4.0
61-90	3-10.1	4.0	0.275 - 4.0
91-120	3-14.5	4.0	0.275 - 4.0

Table 7.2: Field data inversion. Test 1 - no time windowing.

Iteration	Frequency (Hz)	Max time (s)	Offset (km)
1-30	3-5	FA + 1.5	0.275 - 4.0
31-60	3-7.1	FA + 1.5	0.275 - 4.0
61-90	3-10.1	FA + 1.5	0.275 - 4.0
91-120	3-14.5	FA + 1.5	0.275 - 4.0

Table 7.3: Field data inversion. Test 2 - a fixed time window. FA denotes the first-arrival time.

offset inversion - Test 4 (Figure 7.28d). This is related to the fact that the near offset reflections are less sensitive to the long wavelengths of the velocity model than the wide-angle refractions. It means that the reduction of the maximum offset provided the update of the high wavenumber information, i.e. only the short wavelengths were recovered.

When the full data content is inverted at each frequency band, the V_P model is affected by small scale artefacts (Figure 7.28a). The quality of the reconstructed model is better, once a fixed time window (Test 2) is applied to remove the late arrivals (Figure 7.28b). The application of an extending time windowing (Test 3) results in a very similar result to Test 2, however the reconstructed V_P is smoother and contains less artefacts in the shallow part of the model (Figure 7.28c). This result shows a benefit of a layer stripping approach, i.e. the successive inversion from shallow to deep structures.

The combination of the offset windowing with a fixed time window (Test 5) produced a lot of artefacts in the inverted V_P model (Figure 7.28e). Apparently, this multi-stage strategy is not appropriate to provide good quality inversion results. This may be related to the same problem as described for Test 4, i.e. the late arrivals at the near offset provided an update of the small scale structures prior to the inversion of early arrivals at far offsets, which are associated with the low wavenumbers (Figure 7.26).

Test 6 represents the most complex combination of the frequency filtering, time windowing and offset windowing. The inverted V_P model (Figure 7.28f) is similar to the result from Test 2, however the layer contrasts are sharper and there is a lot of fine layers recovered in the shallow part of the model.

Figure 7.29 shows a comparison of the depth profiles extracted from the final inverted FWI models. There is a good agreement between the velocity structures for all inversion tests except for Test 4. We can identify a thick, high velocity layer at a depth of 600 m and a thin low velocity layer at a depth of 800 m, which is clearly visible at the velocity profiles at $x = 2.3$ km. Furthermore, there is a strong velocity contrast in the deeper part of the model, at a depth of approximately 1.6 km. The fine layers in the shallow part of the model, which are present in the V_P models from Test 1 and Tests 4-6, are almost invisible for Test 2 and Test 3. Therefore it is not clear whether these sharp contrasts reflects true velocity structures.

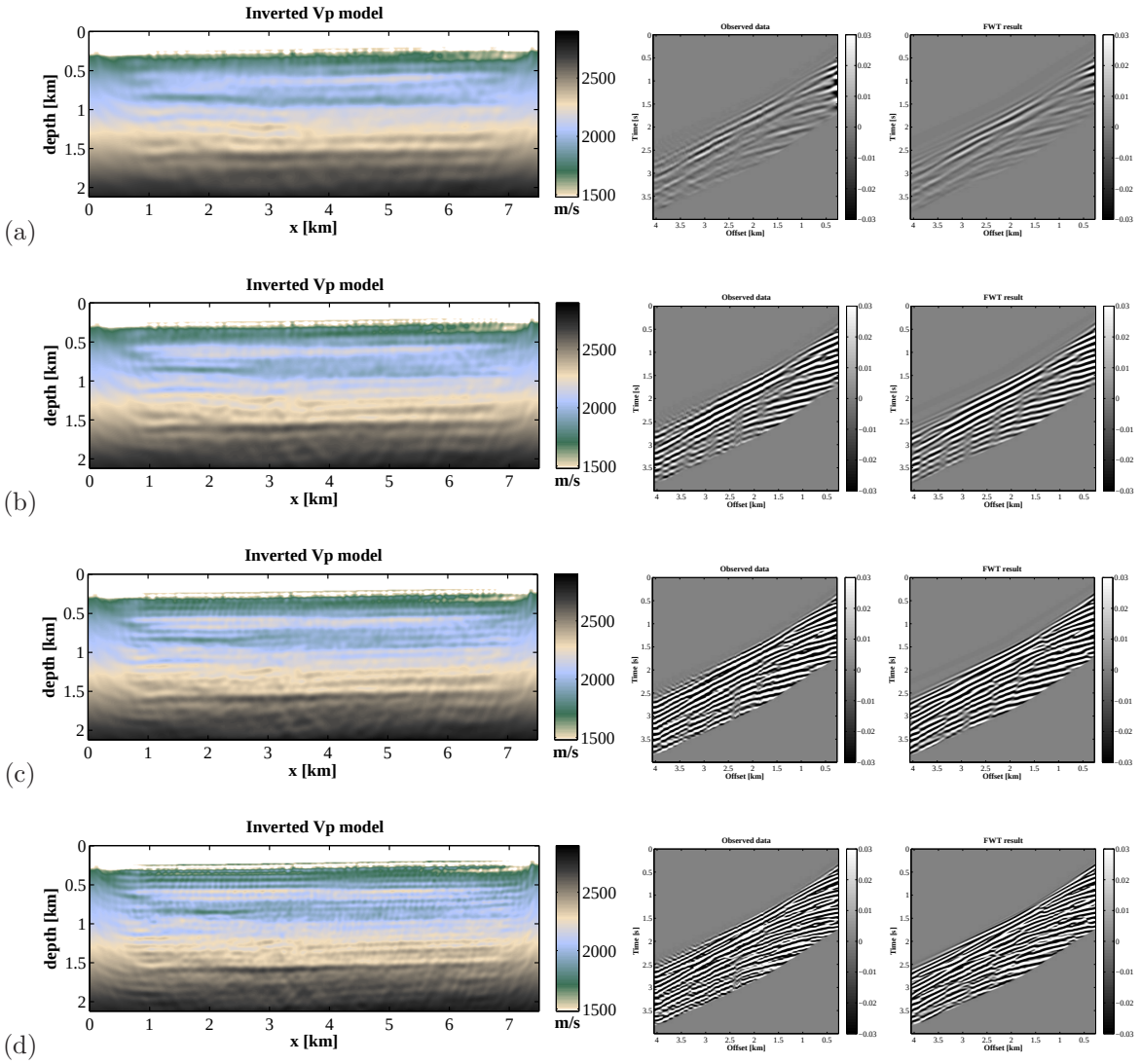


Figure 7.24: Test 2 with a fixed time window (Table 7.3). Intermediate and final FWI results after each frequency stage. (a) $f_{\max} = 5$ Hz, (b) $f_{\max} = 7.1$ Hz, (c) $f_{\max} = 10.1$ Hz, and (d) the final inversion result $f_{\max} = 14.5$ Hz. Left: inverted V_P models, center: the field data, right: the synthetic data generated for the FWI model. Shot 62 located at $x = 7.275$ km. Note the increase of the resolution when higher frequencies are included.

Iteration	Frequency (Hz)	Max time (s)	Offset (km)
1-10	3-5	FA + 0.5	0.275 - 4.0
11-20		FA + 1.0	
21-30		FA + 1.5	
31-40	3-7.1	FA + 0.5	0.275 - 4.0
41-50		FA + 1.0	
51-60		FA + 1.5	
61-70	3-10.1	FA + 0.5	0.275 - 4.0
71-80		FA + 1.0	
81-90		FA + 1.5	
91-100	3-14.1	FA + 0.5	0.275 - 4.0
101-110		FA + 1.0	
111-120		FA + 1.5	

Table 7.4: Field data inversion. Test 3 - extending time windowing. FA denotes the first-arrival time.

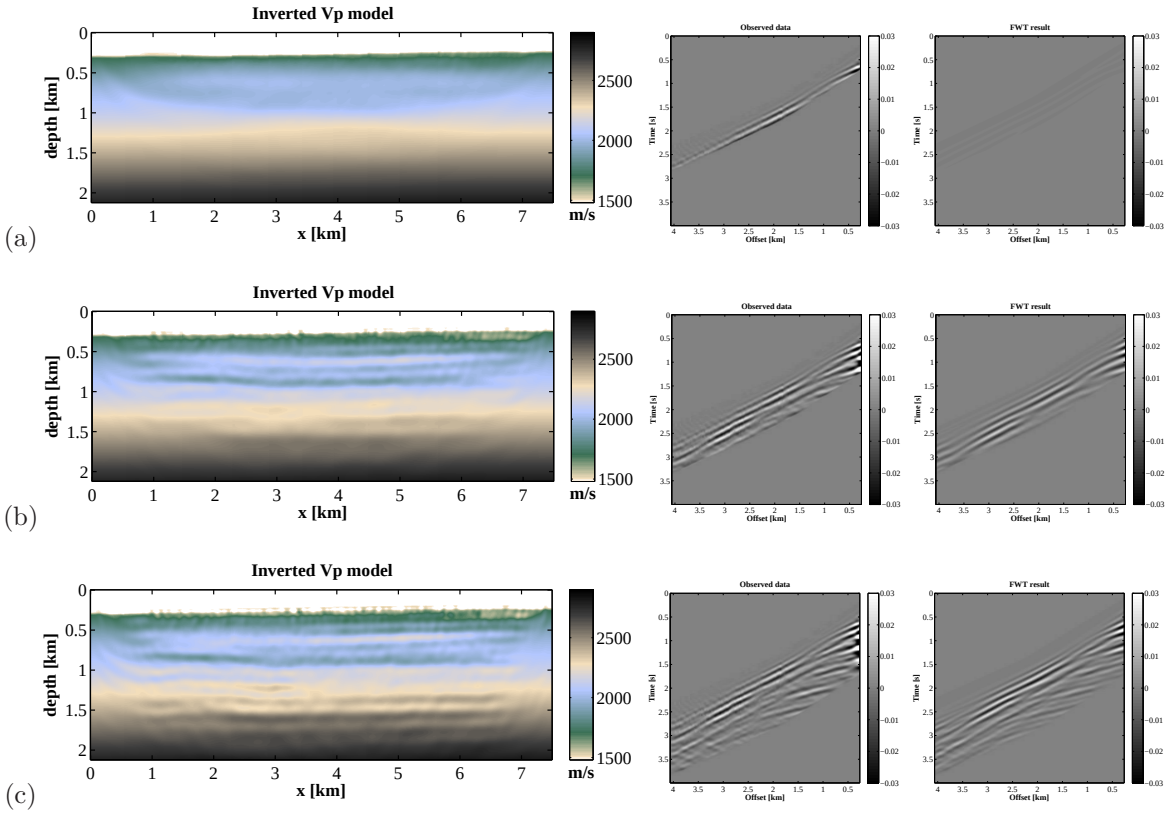


Figure 7.25: Test 3 with an extending time windowing (Table 7.4). Intermediate FWI results for the first frequency band from 3 Hz to 5 Hz, (a) time window of 0.5 s, (b) time window of 1.0 s, (c) time window of 1.5 s. Left: inverted V_P models, center: the field data, right: the synthetic data generated for the FWI model. Shot 62 located at $x = 7.275$ km.

Iteration	Frequency (Hz)	Max time (s)	Offset (km)
1-10	3-5	FA + 0.5	0.275 - 2.0
11-20		FA + 1.0	
21-30		FA + 1.5	
31-40	3-7.1	FA + 0.5	0.275 - 2.0
41-50		FA + 1.0	
51-60		FA + 1.5	
61-70	3-10.1	FA + 0.5	0.275 - 2.0
71-80		FA + 1.0	
81-90		FA + 1.5	
91-100	3-14.1	FA + 0.5	0.275 - 2.0
101-110		FA + 1.0	
111-120		FA + 1.5	

Table 7.5: Field data inversion. Test 4 - near offset data inversion with an extending time windowing. FA denotes the first-arrival time.

Iteration	Frequency (Hz)	Max time (s)	Offset (km)
1-20	3-5	FA + 1.5	0.275 - 2.0
21-40			0.275 - 4.0
41-60	3-7.1	FA + 1.5	0.275 - 2.0
61-80			0.275 - 4.0
81-100	3-10.1	FA + 1.5	0.275 - 2.0
101-120			0.275 - 4.0
121-140	3-14.5	FA + 1.5	0.275 - 2.0
141-160			0.275 - 4.0

Table 7.6: Field data inversion. Test 5 - offset windowing with a fixed time window. FA denotes the first-arrival time.

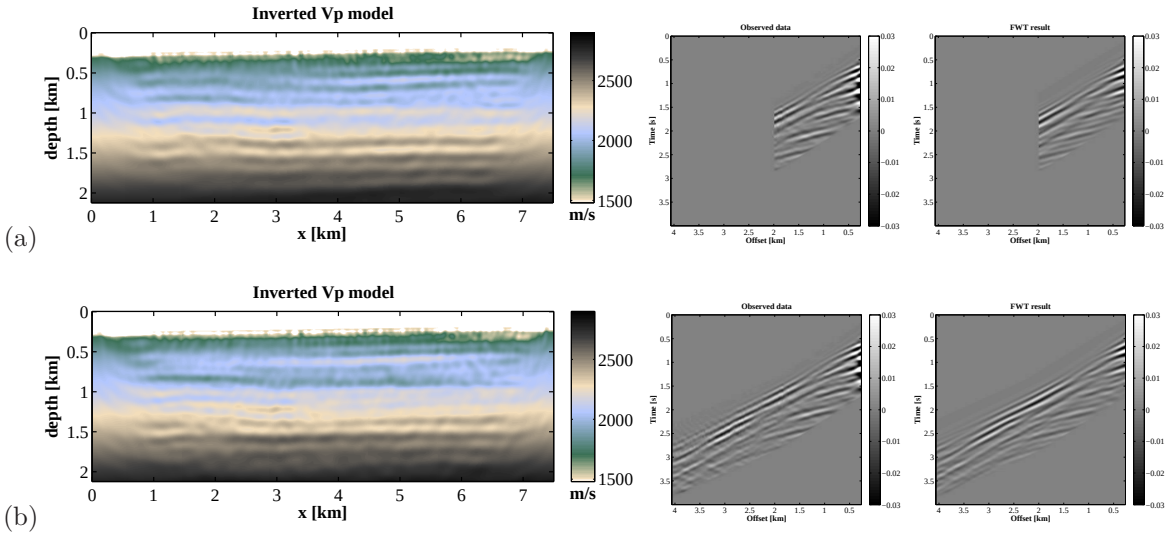


Figure 7.26: Test 5 - offset windowing with a fixed time window (Table 7.6). Intermediate FWI results for the first frequency band from 3 Hz to 5 Hz, (a) near offset window up to 2 km, time window of 1.5 s, (b) full offset data, time window of 1.5 s. Left: inverted V_P models, center: the field data, right: the synthetic data generated for the FWI model. Shot 62 located at $x = 7.275$ km.

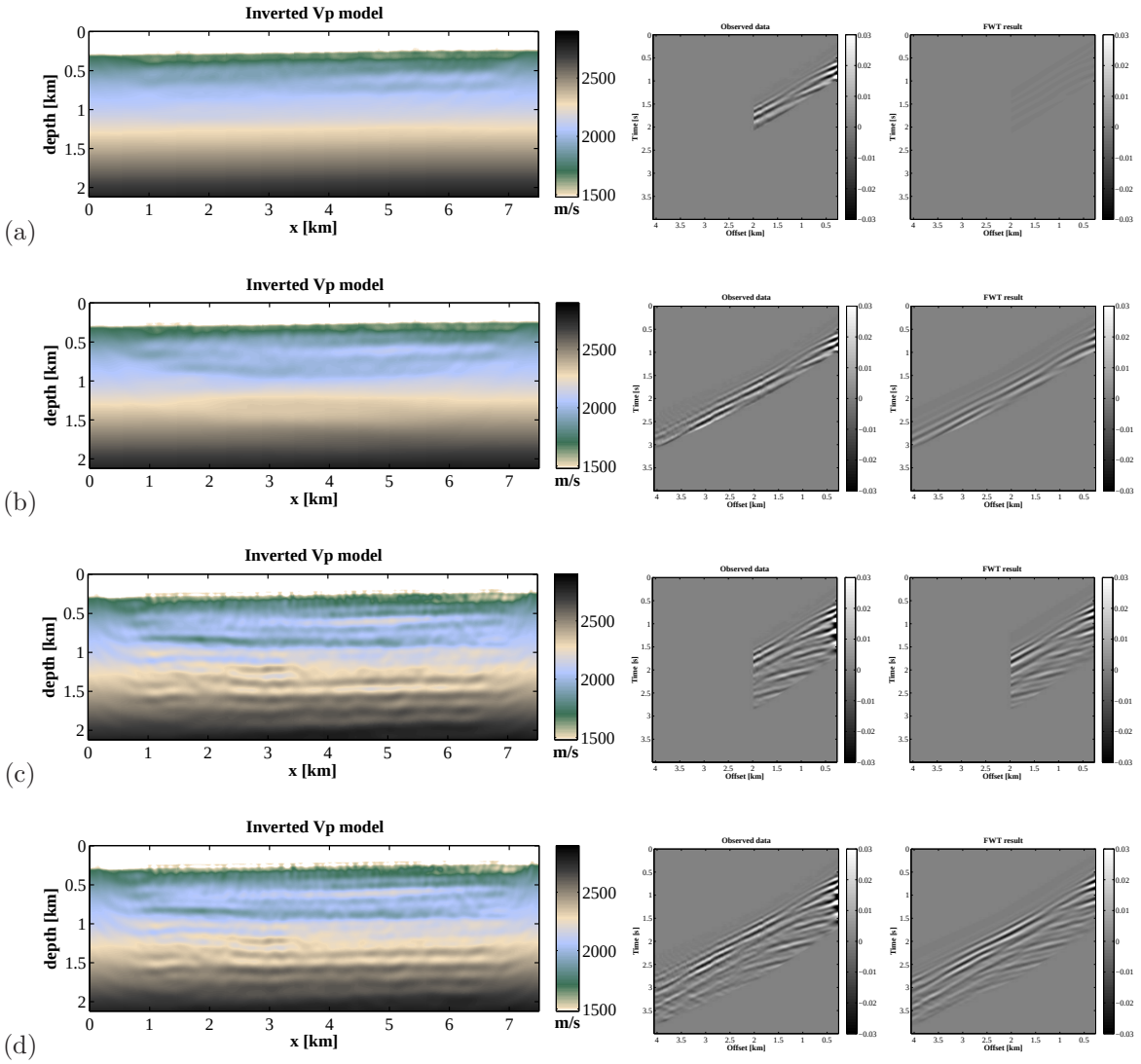
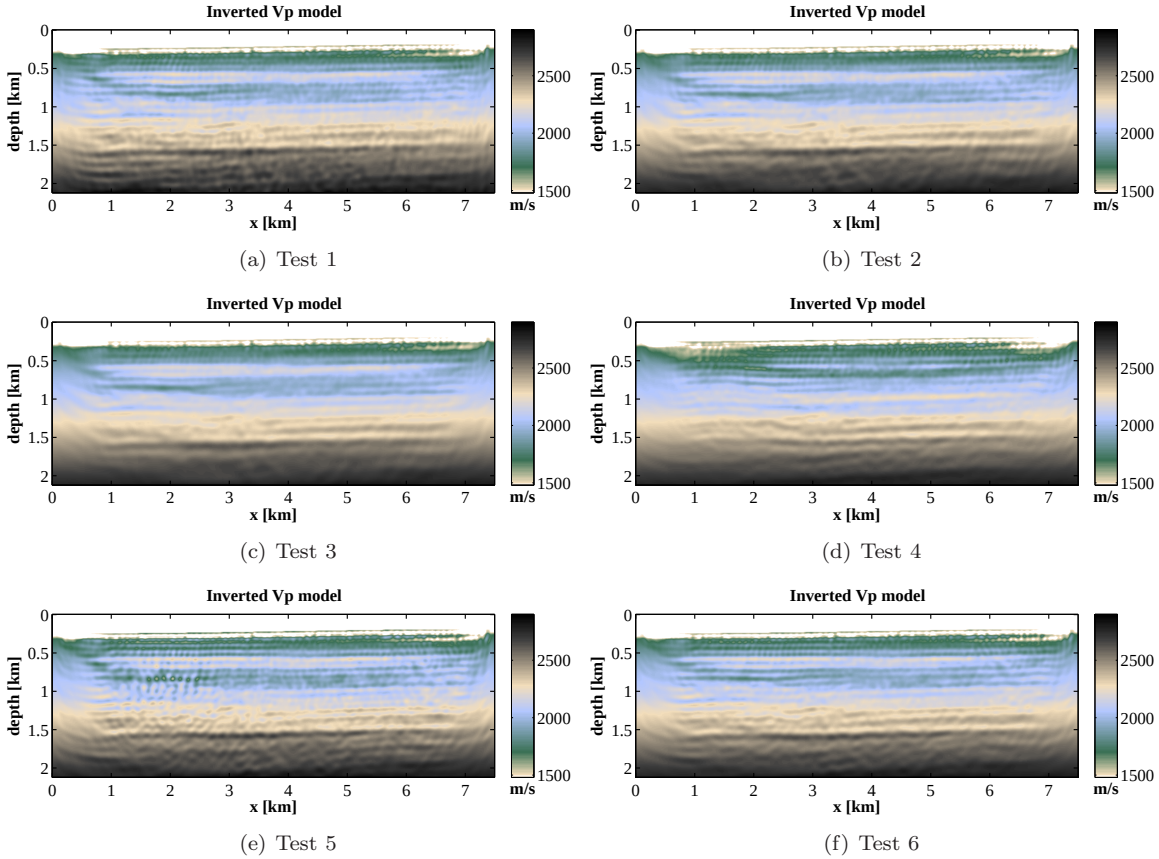


Figure 7.27: Test 6 – offset windowing with extending time window (Table 7.7). Intermediate FWI results for the first frequency band from 3 Hz to 5 Hz, (a) near offset window up to 2 km, time window of 0.75 s, (b) full offset data, time window of 0.75 s, (c) near offset window up to 2 km, time window of 1.5 s, (d) full offset data, time window of 1.5 s. Left: inverted V_P models, center: the field data, right: the synthetic data generated for the FWI model. Shot 62 located at $x = 7.275$ km.

Iteration	Frequency (Hz)	Max time (s)	Offset (km)
1-10	3-5	FA + 0.75	0.275 - 2.0
11-20		FA + 0.75	0.275 - 4.0
21-30		FA + 1.5	0.275 - 2.0
31-40		FA + 1.5	0.275 - 4.0
41-50	3-7.1	FA + 0.75	0.275 - 2.0
51-60		FA + 0.75	0.275 - 4.0
61-70		FA + 1.5	0.275 - 2.0
71-80		FA + 1.5	0.275 - 4.0
81-90	3-10.1	FA + 0.75	0.275 - 2.0
91-100		FA + 0.75	0.275 - 4.0
101-110		FA + 1.5	0.275 - 2.0
111-120		FA + 1.5	0.275 - 4.0
121-130	3-14.5	FA + 0.75	0.275 - 2.0
131-140		FA + 0.75	0.275 - 4.0
141-150		FA + 1.5	0.275 - 2.0
151-160		FA + 1.5	0.275 - 4.0

Table 7.7: Field data inversion. Test 6 - offset windowing combined with an extending time windowing. FA denotes the first-arrival time.



(a) Test 1

(b) Test 2

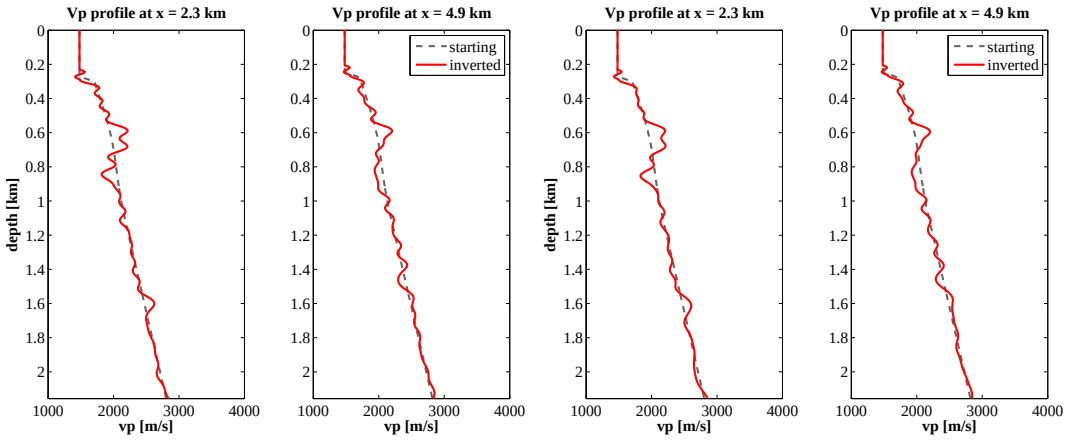
(c) Test 3

(d) Test 4

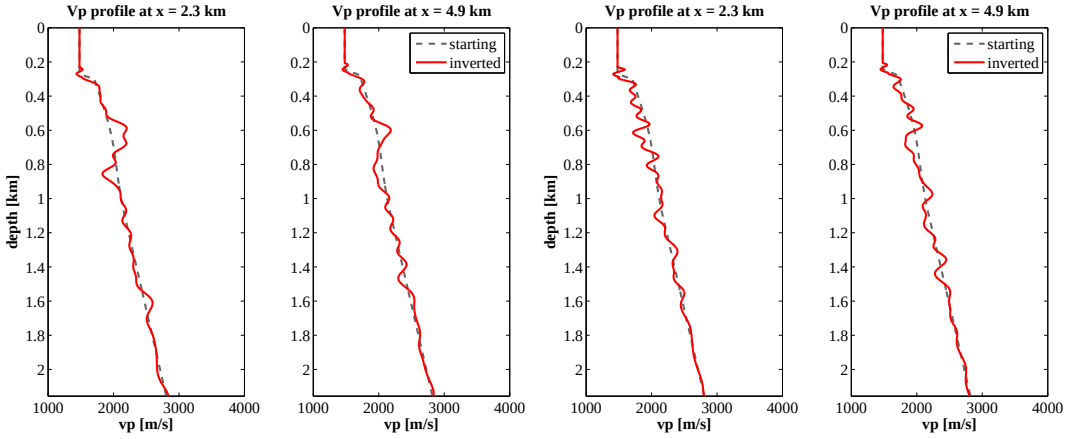
(e) Test 5

(f) Test 6

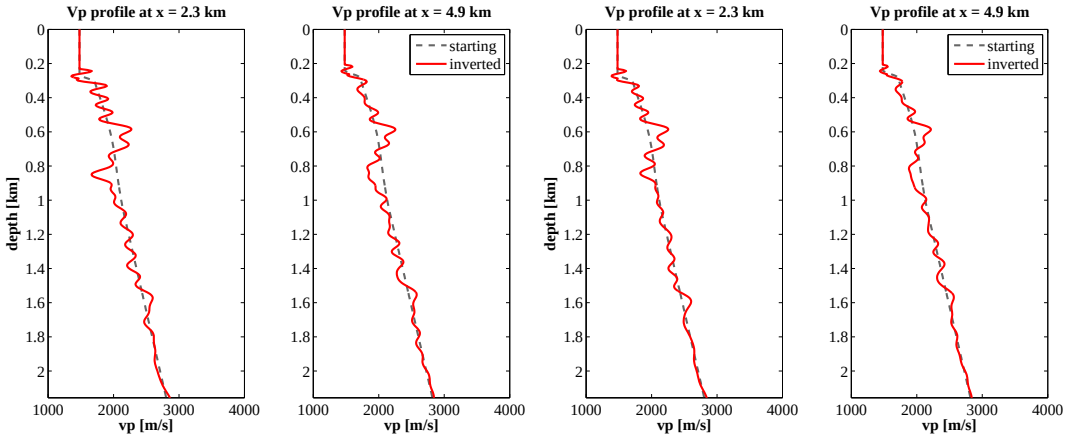
Figure 7.28: Test 1 - Test 6. Comparison of FWI results, $f_{\max} = 10.1$ Hz. Inverted V_P models.



(a) Test 1 (b) Test 2



(c) Test 3 (d) Test 4



(e) Test 5 (f) Test 6

Figure 7.29: Test 1 - Test 6. Comparison of FWI results, $f_{\max} = 10.1$ Hz. V_P profiles at $x = 2.3$ km and at $x = 4.9$ km. The gray dashed line represents the initial model, and the red solid line represents the inverted V_P model.

7.7.2 Choice of the objective function

In the second set of inversion experiments, I inverted field data using different minimization criteria discussed in Chapter 3. The inversion strategy was identical as in Test 3 (Table 7.4). The final inversion results are shown in Figure 7.30. The quality of reconstructed V_P models is comparable for all minimization criteria. There are some minor differences which mainly concerns the low velocity layer at a depth of 800 m. Furthermore, we can observe a difference between the L_2 norm and the other minimization criteria, which is related to the increased amplitude of reflectors (compare the V_P perturbations in Figure 7.30). Because the field data are not affected by high amplitude noise, there is no clear benefit of using the minimization criteria, which are more robust to noise, such as the L_1 norm, Cauchy or sech norm.

7.7.3 Multi-parameter inversion

The preceding tests were based on a mono-parameter inversion with the V_P as the only unknown parameter. The initial density model was fixed and all data residuals were projected to the P-wave velocity model. In this section, I applied a multi-parameter acoustic inversion using three different parameterization sets: P-wave velocity and density $\mathbf{m_1} = [V_P, \rho]$, P-wave impedance and density $\mathbf{m_2} = [I_P, \rho]$, and the third parameter set is the P-wave velocity and P-wave impedance $\mathbf{m_3} = [V_P, I_P]$. The inversion strategy was identical as in Test 3 (Table 7.4), such that full offset data were included at each iteration step and an extending time windowing was applied.

The inverted V_P models for different parameter sets are shown in Figure 7.31. In case of the parameterization with $\mathbf{m_2} = [I_P, \rho]$, the final velocity model is computed as $V_P = I_{P\text{inv}}/\rho_{\text{inv}}$. The resolution of the reconstructed P-wave velocity models is comparable for all parameterizations and the differences are very small.

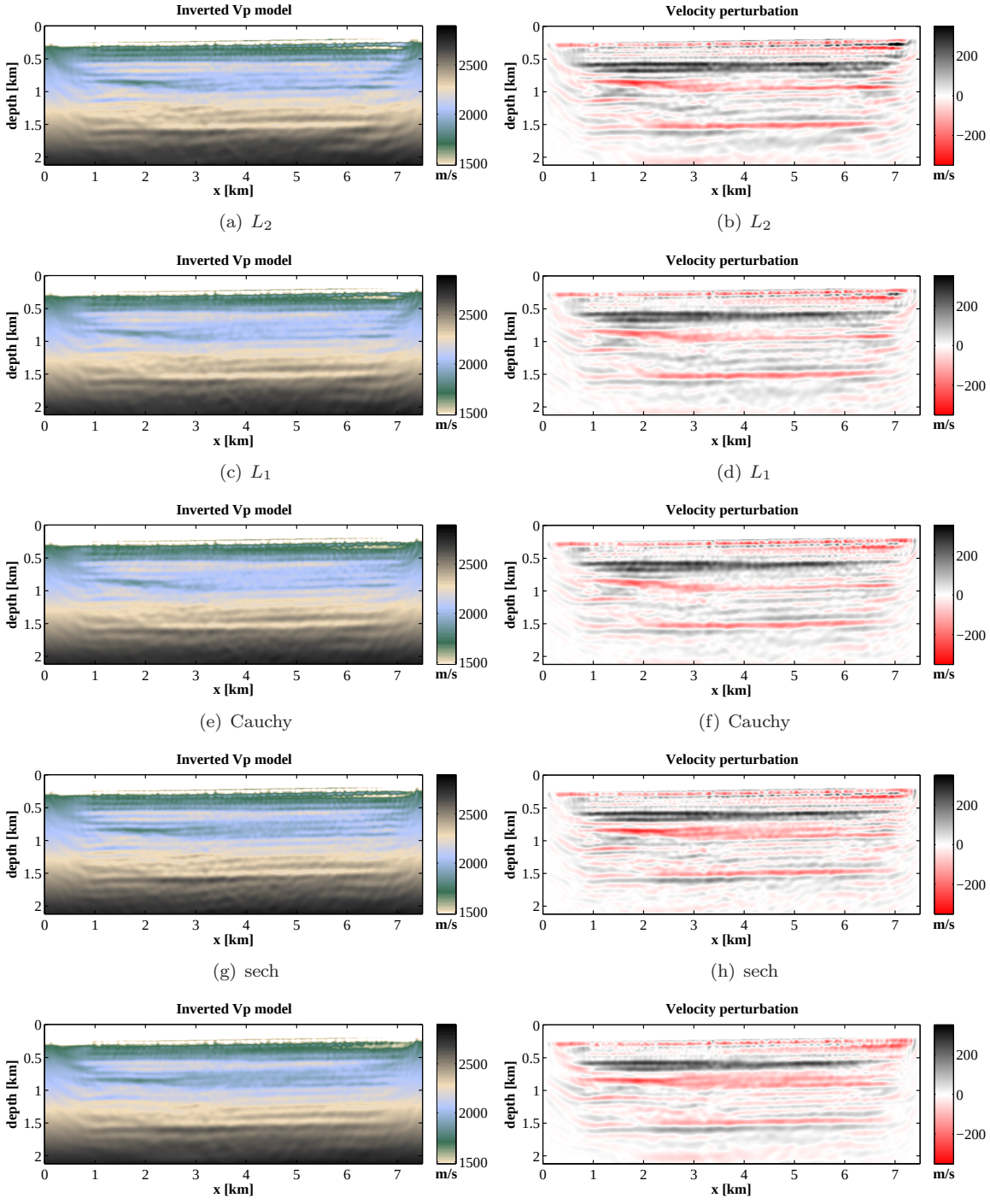
As I have shown in Chapter 4, density is a difficult parameter to reconstruct and its accuracy depends on the model parameterization. A comparison of inverted density models for different parameter sets is shown in Figure 7.32. The poor recovery of density structures is related to the choice of the $\mathbf{m_2} = [I_P, \rho]$ parameterization. On the other hand, when the $\mathbf{m_1} = [V_P, \rho]$ is used, we can recognize some structures that are also present in the velocity models, for example the reflector at a depth of 1.6 km or the low velocity/density layer at a depth of 800 m. The density contrasts are much weaker for the $\mathbf{m_3} = [V_P, I_P]$ parameterization. These results are consistent with those obtained using synthetic data (see Chapter 4).

Finally, the inverted I_P models are illustrated in Figure 7.33. The initial impedance model was calculated as $I_P = V_P\rho$. The reconstructed impedance structures are very similar for all parameterization sets. We can clearly identify the high impedance and the low impedance layers in the shallow part of the model. Furthermore, the fine layers present in the I_P models are consistent with those in the inverted V_P models.

7.7.4 Starting models

In the final set of inversion experiments, I performed FWI using various starting models. The inversion strategy was identical as in Test 3 (Table 7.4) and a mono-parameter inversion for V_P was applied.

The final inversion results are shown in Figure 7.34. A comparison of V_P profiles reveals a good



(a) L_2 (b) L_2

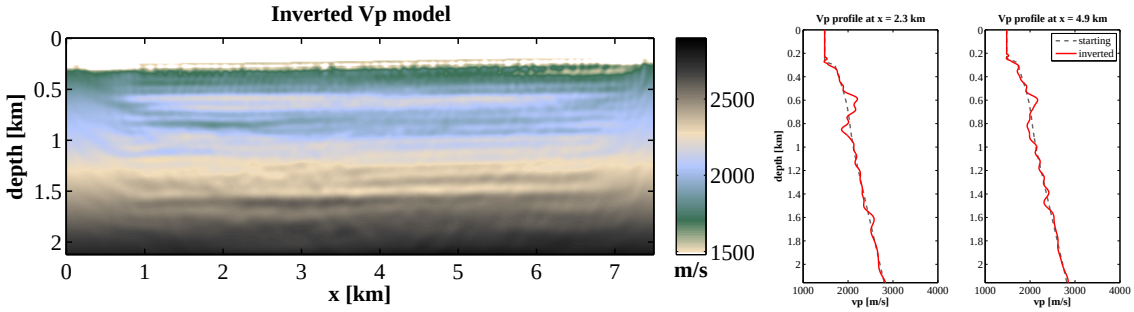
(c) L_1 (d) L_1

(e) Cauchy (f) Cauchy

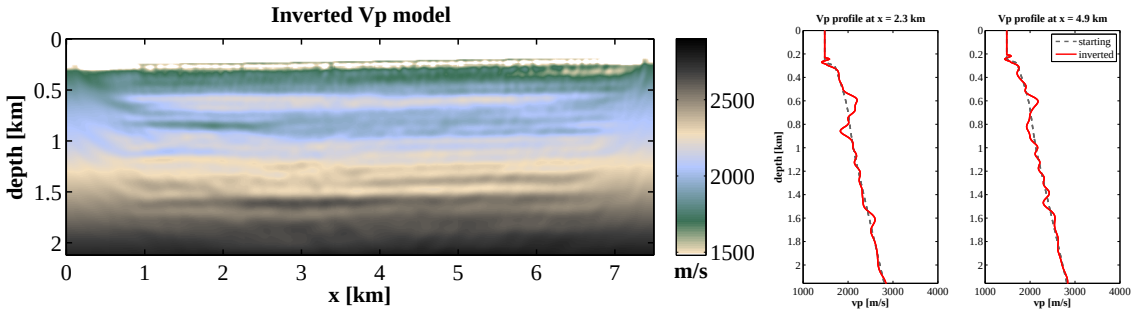
(g) sech (h) sech

(i) L_2norm (j) L_2norm

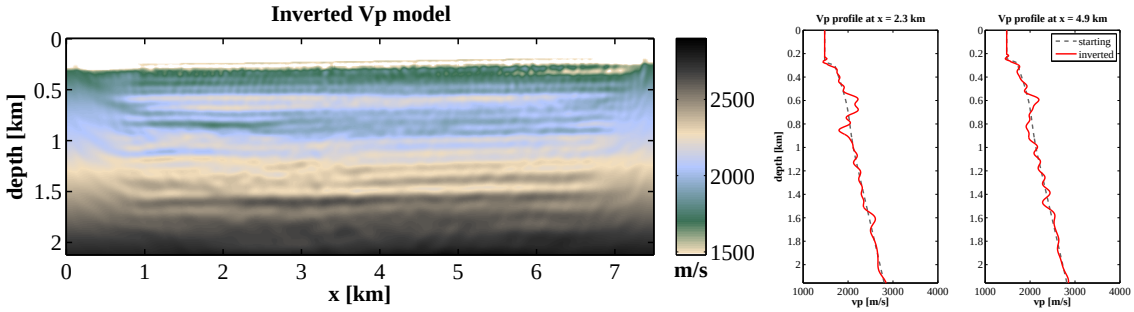
Figure 7.30: Comparison of FWI results for different minimization criteria, $f_{\max} = 10.1$ Hz. Left: inverted V_P models, right: V_P perturbation computed as the difference between the inverted and the starting V_P model.



(a) $\mathbf{m}_1 = (V_P, \rho)$

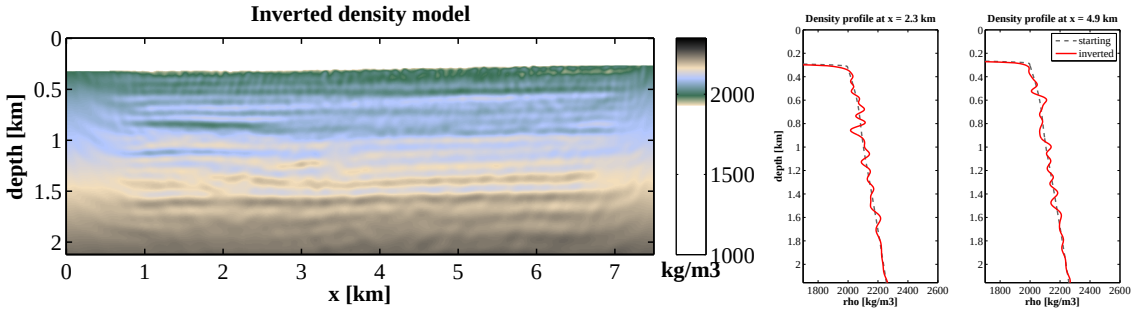


(c) $\mathbf{m}_2 = (I_P, \rho)$

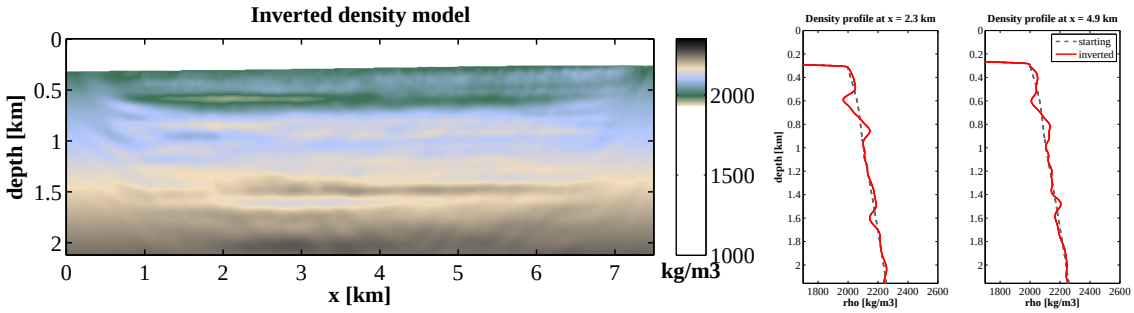


(e) $\mathbf{m}_3 = (V_P, I_P)$

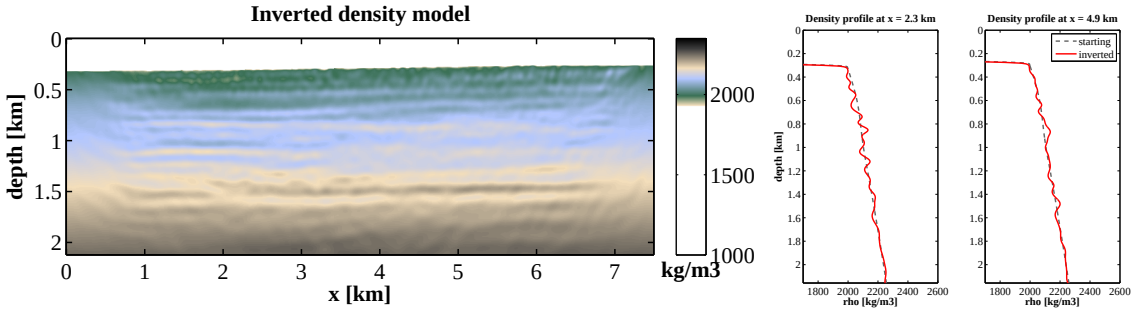
Figure 7.31: Multi-parameter inversion of the field data, $f_{\max} = 10.1$ Hz. Left: inverted V_P models for different parameterization sets. Right: V_P profiles at $x = 2.3$ km and at $x = 4.9$ km. The gray dashed line represents the initial model, and the red solid line represents the inverted V_P model.



(a) $\mathbf{m}_1 = (V_P, \rho)$

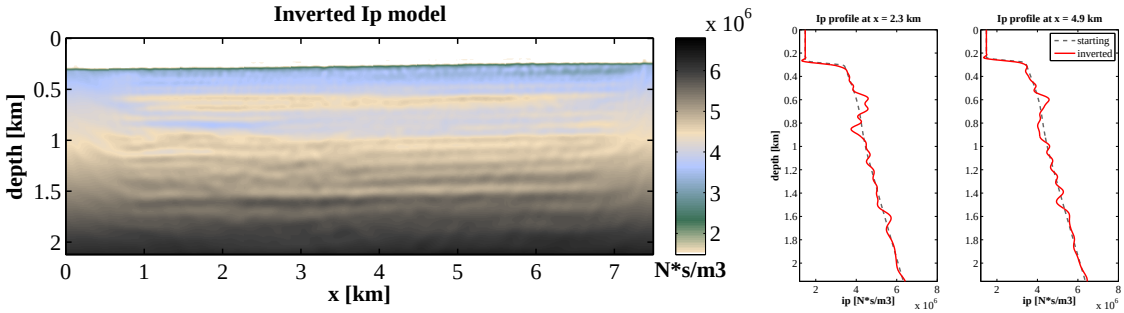


(c) $\mathbf{m}_2 = (I_P, \rho)$

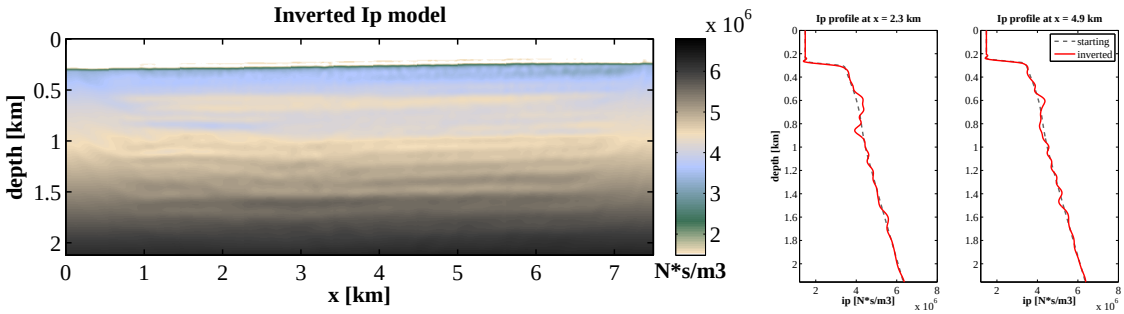


(e) $\mathbf{m}_3 = (V_P, I_P)$

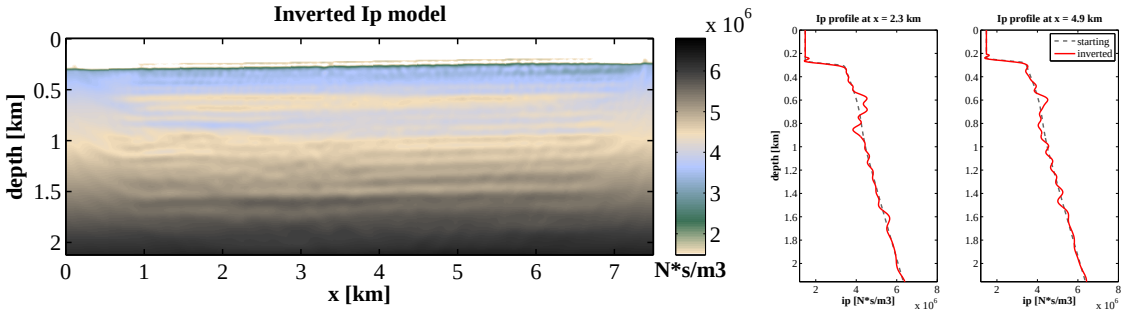
Figure 7.32: Multi-parameter inversion of the field data, $f_{\max} = 10.1$ Hz. Left: inverted density models for different parameterization sets. Right: Density profiles at $x = 2.3$ km and at $x = 4.9$ km. The gray dashed line represents the initial model, and the red solid line represents the inverted density model.



(a) $\mathbf{m}_1 = (V_P, \rho)$



(c) $\mathbf{m}_2 = (I_P, \rho)$



(e) $\mathbf{m}_3 = (V_P, I_P)$

Figure 7.33: Multi-parameter inversion of the field data, $f_{\max} = 10.1$ Hz. Left: inverted I_P models for different parameterization sets. Right: I_P profiles at $x = 2.3$ km and at $x = 4.9$ km. The gray dashed line represents the initial model, and the red solid line represents the inverted I_P model.

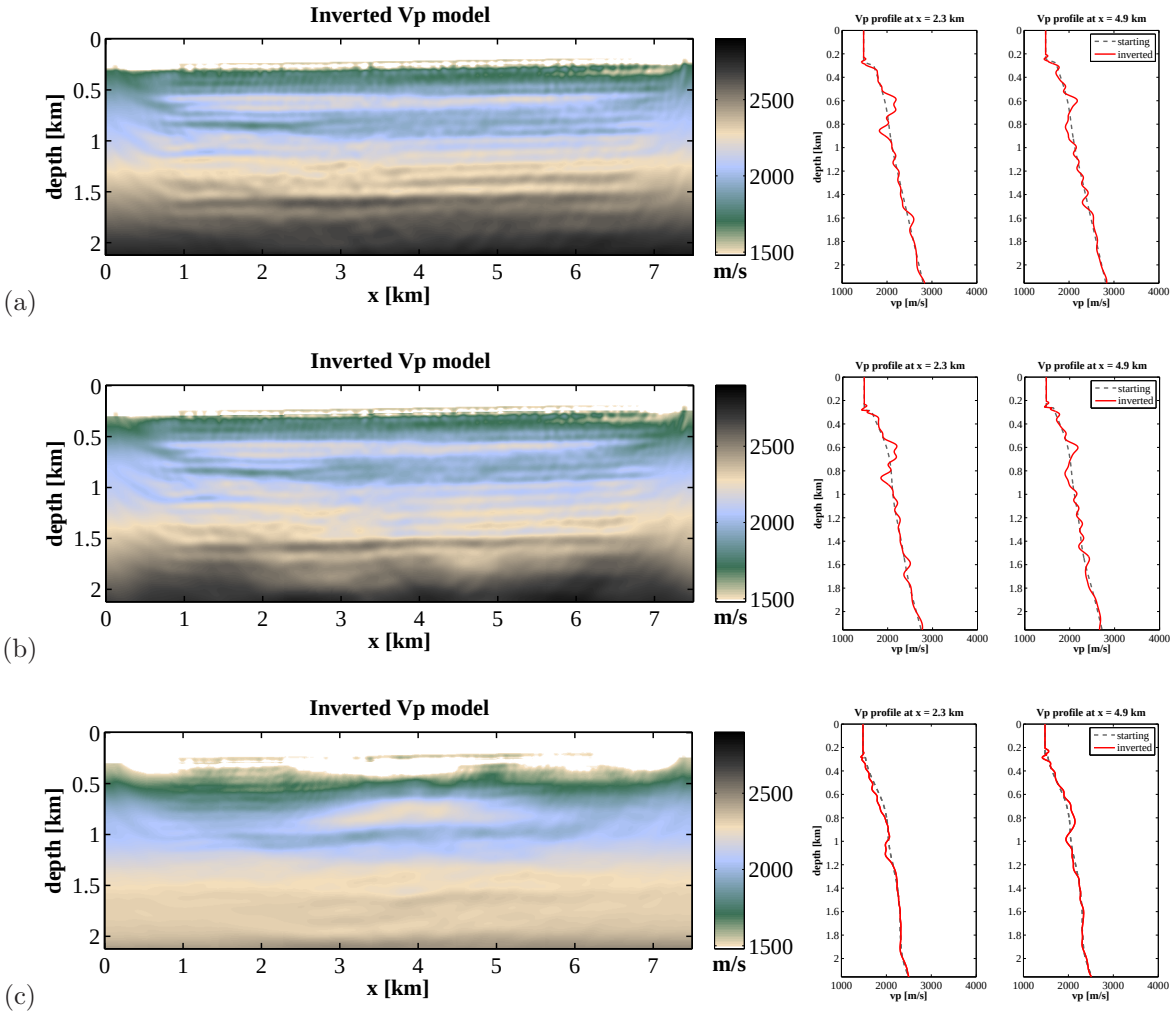


Figure 7.34: Comparison of FWI results for different starting models, $f_{\max} = 10.1$ Hz. Left: inverted V_P models, right: V_P profiles at $x = 2.3$ km and at $x = 4.9$ km. The gray dashed line represents the initial model, and the red solid line represents the inverted I_P model. (a) starting V_P model from the refraction traveltime tomography, (b) starting V_P model based on the NMO velocity model, (c) starting V_P model from the VSP data.

agreement between the FWI results obtained from both the traveltime tomography and NMO starting models. The fit is especially good in the upper part of the V_P models, the reconstructed low velocity and the high velocity layers are very similar. The velocity structures in the deep part of the inverted V_P models do not match exactly. This is caused by the different background velocities in these two initial models below the depth of 1 km (Figure 7.15), i.e. the background velocity in the starting model from the traveltime tomography is higher than in the NMO-based model. This inversion result indicates that the long-wavelength structures of the P-wave velocity are not updated by FWI in a deep part of the model.

Finally, the FWI inversion with the starting model from the VSP data failed to reconstruct the V_P model (Figure 7.34c). This result is not surprising, since the waveform match between the observed and initial data generated for the model from the VSP data is very poor at far offsets (Figure 7.17a). Because of the cycle-skipping problem, the waveform inversion ended up in a local minimum of the misfit function.

7.8 Validation of inversion results

An important step of the workflow for the field data inversion is the validation of inversion results. This can be achieved by a comparison of the observed data with the synthetic data generated for the FWI results. Furthermore, it is useful to analyse the estimated source wavelets, because the may indicate the accuracy of the reconstructed models. The validity of the inverted models can be also confirmed by a comparison with the migrated seismic section.

7.8.1 Data comparison

Figure 7.35 shows a comparison between the observed data and synthetic data generated for the final FWI results for Test 2, Test 3, and Test 6. I selected these inversion tests, because the final V_P models are of good quality and they are least contaminated by artefacts. We can observe a good match between the observed data and the synthetic data generated for the final FWI results. Because the reconstructed velocity structures are very similar in all three V_P models, the data fit is nearly equivalent. There is some data misfit related to the amplitude information at the far offset, which may be due to the elastic effects or attenuation.

7.8.2 Final source wavelets

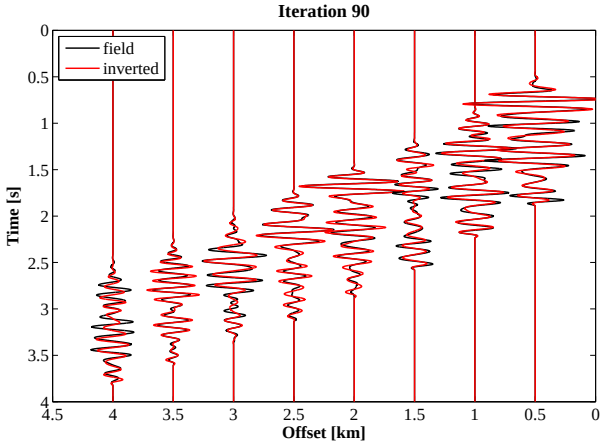
As mentioned in Chapter 6, the inverted source wavelets can serve as the quality check on the velocity model during the full waveform inversion. When the residuals between the observed and predicted data are decreasing and the estimated source wavelets improve as well, this indicates an accurate reconstruction of the velocity model. The initial source wavelet for shot 62 is shown in Figure 7.16 (blue line in the top panel), whereas the final wavelets are illustrated in Figure 7.36. In general, source wavelets produced for different final V_P models are very similar, but the source signature that corresponds to Test 2 exhibits more ringing artefacts than the other wavelets. Figure 7.37 shows the final source wavelets for three inversion tests and for all individual shots. We can observe that the estimated wavelets are coherent and the general shape of source signatures is comparable for all

inversion tests. However, the first ten shots are mory noisy, which is due to the limited maximum offset related to these shots (Figure 7.21b).

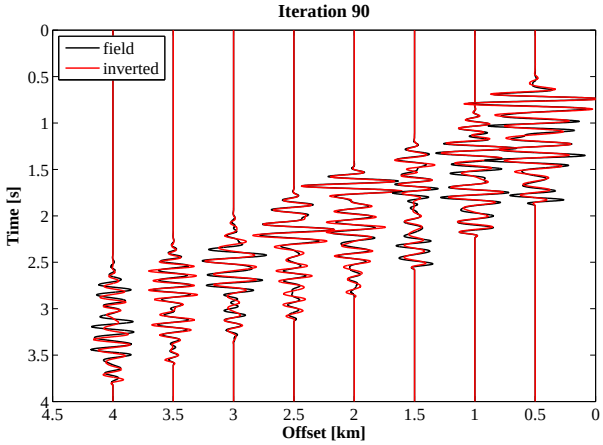
7.8.3 Comparison with the migrated seismic section

Finally, I applied a conventional seismic processing to the field data to obtain a migrated section. The processing sequence is described in Appendix A. Figure 7.38a shows a seismic section after a post-stack Kirchoff time migration. There are a lot of fine, quasi-horizontal sedimentary structures present in the migration image.

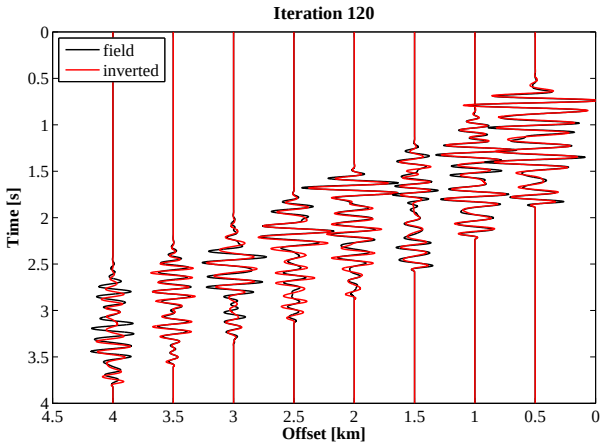
To compare the final V_P model from Test 3 with the migrated seismic section, the inverted model was converted from depth to time (Figure 7.38b). In addition, the vertical velocity gradient of the time-converted V_P model was calculated to allow a better identification of layer interfaces. Because of the limited frequency range, the layer interfaces in the inverted V_P model are not as sharp as in the migrated section. Nevertheless, most of the major structures present in the inverted V_P model are clearly visible in the migrated seismic section. This mainly concerns the strong reflector at 1.6 s, or the low velocity structure at 1 s. Furthermore, the top and the bottom interface of the high velocity layer in the upper part of the model can be identified in the migrated section, this structure is getting thinner in the right part of the model. FWI resolved a lot of fine structures between 1.0 and 1.7 s, but the layering structures are not present below 1.7 s. This result correlates well with the migrated section, where the weak reflectors area is visible below 1.7 s.



(a)



(b)



(c)

Figure 7.35: Trace comparison between the observed data (black) and the synthetic data (red) generated with the final FWI result from: (a) Test 2, (b) Test 3, (c) Test 6. Shot 62 located at $x = 7.275$ km.

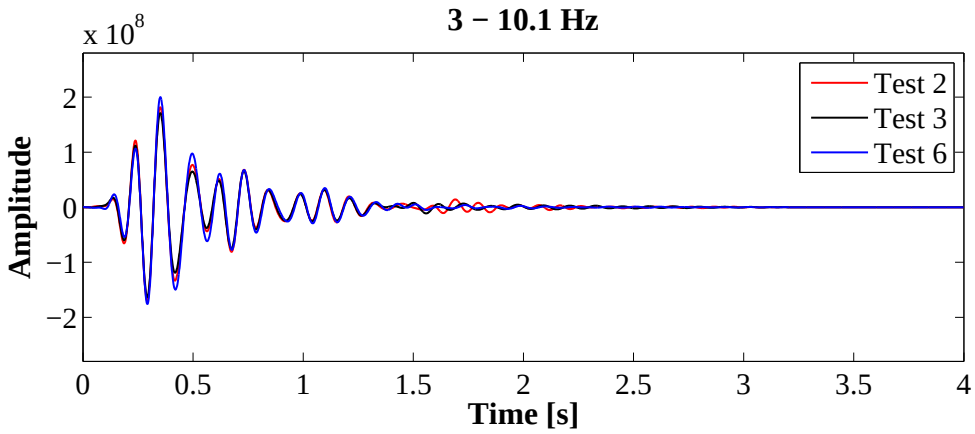
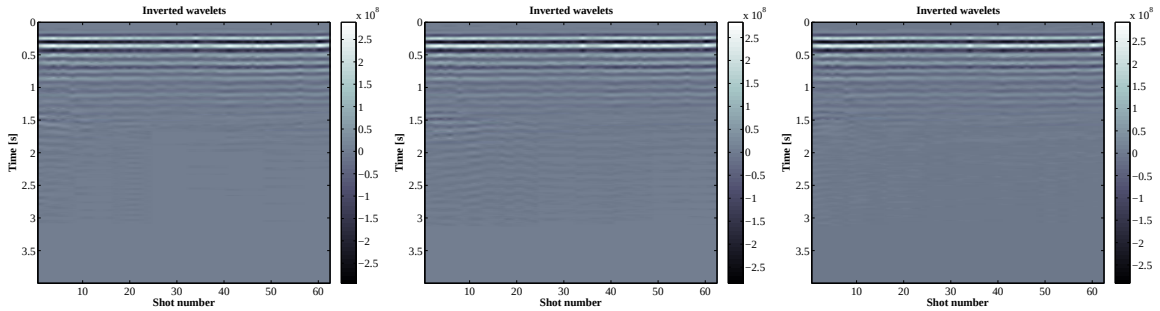
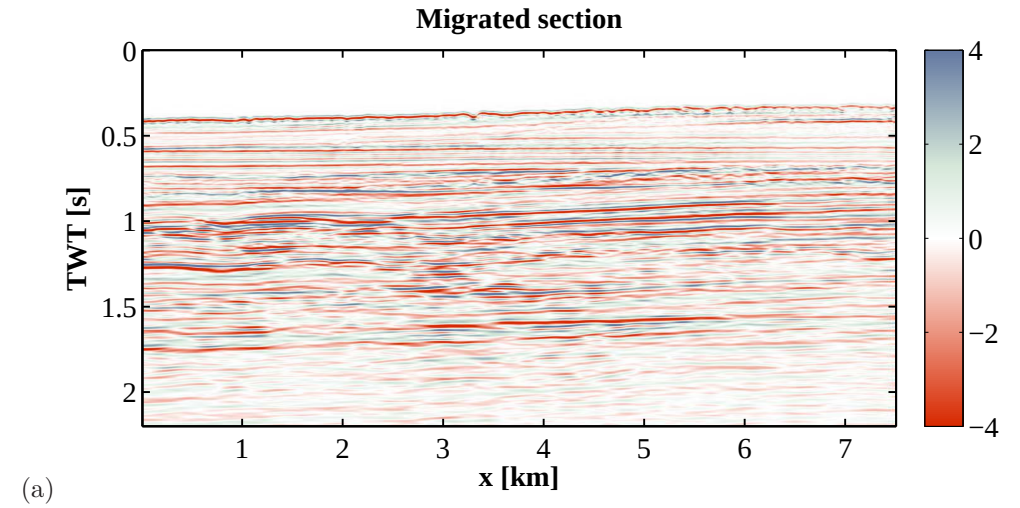


Figure 7.36: Comparison of source wavelets for shot 62 estimated for the final inversion results from: (a) Test 2, (b) Test 3, (c) Test 6.

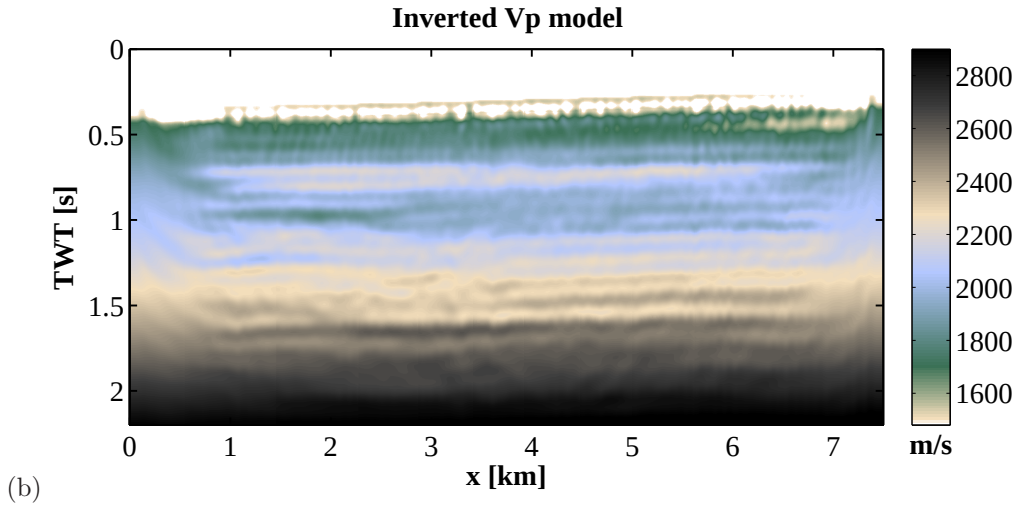


(a) (b) (c)

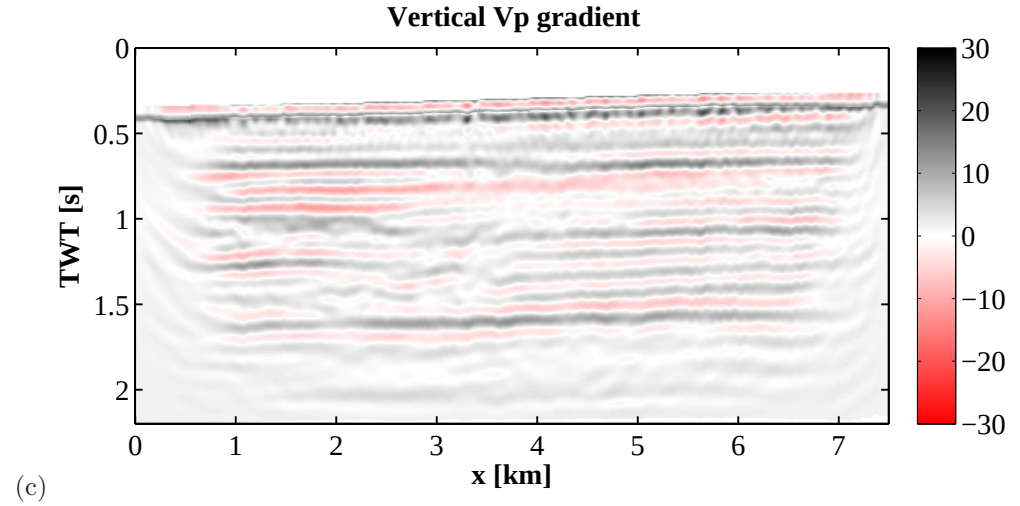
Figure 7.37: Comparison of source wavelets estimated for the final inversion results from: (a) Test 2, (b) Test 3, (c) Test 6.



(a)



(b)



(c)

Figure 7.38: (a) Time-migrated section of the seismic line that corresponds to the subset of the data selected for the FWI inversion. (b) Final V_P model from Test 3 after conversion from depth to time, and (c) after the calculation of the vertical velocity gradient.

7.9 Summary

In this Chapter, I presented the application of the acoustic FWI to marine streamer data acquired in the North Sea. I have shown that the success of the full waveform inversion depends on three main factors: the careful data preprocessing, choice of the starting model, and an appropriate inversion strategy.

The aim of the data preprocessing was to improve the signal-to-noise ratio and to remove these wave propagation phenomena that cannot be modelled correctly, such as seismic noise, elastic effects, or 3D effects. The frequency content of the inverted data was limited to $3 - 10$ Hz. The very low frequencies were removed because of the strong noise component. The maximum frequency was limited to 10 Hz, because the inversion of higher frequencies introduced a lot of artefacts into the reconstructed V_P models. This was mainly caused by the attenuation effects, which are not taken into account in the forward modelling.

Due to the lack of frequencies below 3 Hz in the field data, it was necessary to generate an initial velocity model that already contains the long-wavelength structures of the subsurface. In order to avoid the cycle-skipping artefacts, the predicted data generated for the starting model should match the observed data within half a cycle. Initial modelling performed prior to the waveform inversion helped to evaluate the accuracy of initial P-wave velocity models. The velocity models obtained from both the refraction traveltime tomography and the NMO velocity provided a good match between the field data and the initial synthetics for the first frequency band. Furthermore, the RMS amplitude versus offset trend showed a reasonably good fit between the field data after preprocessing and the modelled data, which indicates that no additional offset-dependent amplitude scaling was required for this data set.

To reduce the high complexity of the inverse problem, the multi-scale inversion approach based on the continuous coverage of vertical wavenumbers was applied. With the source spacing of 100 m and the receiver spacing of 25 m, the aliasing number is $N_A = 0.67$ for the maximum frequency of 5 Hz, and $N_A = 1.35$ for $f = 10$ Hz. It means that the source and the receiver sampling was sufficient to obtain a good quality image not affected by spatial aliasing.

The resolving power of the FWI was tested with the checkerboard model using the same acquisition setup, source signature and the frequency content as in the field data example. This was useful to determine which parts of the model are well-resolved, to identify artefacts due to source-receiver geometry limitations, and to verify the choice of inversion parameters. The checkerboard pattern with the anomaly size of 100 m was well recovered. However due to a poor subsurface illumination at the sides of the model, the strong smearing effects were present in these parts of the inverted V_P models.

The comparison of different inversion strategies based on time windowing and/or offset windowing of the data showed their significant effect on the inversion results. First of all, it is useful to apply a time windowing of the data around the first arrivals and to damp the late arrivals which increase the non-linearity of the inverse problem. Furthermore, the application of the extending time windowing resulted in a good quality of the reconstructed model. This successive inversion from shallow to deep structures is similar to the layer stripping approach. The combination of the frequency filtering, time windowing and offset windowing resulted in no significant improvement of the velocity model.

The FWI using the near-offset data and the full-offset data information produced different V_P

models. Because the near-offset relections are less sensitive to the long wavelengths of the velocity model, the limitation of the maximum offset resulted in the update of the short wavelengths only. This means that the long-offset data are essential to reconstruct the long-wavelength structures of the model, especially when the low frequencies are not available in the data.

Because the preprocessed field data were not affected by high amplitude noise, the inversion results obtained for different minimization criteria were very similar. Even the conventional L_2 norm produced a good quality image.

The multi-parameter inversion showed the difficulty in resolving a reliable density model. On the other hand, the P-wave velocity and impedance models were well reconstructed for all parameterization sets and show the existence of similar structures.

All final V_P models obtained from the mono-parameter and the multi-parameter inversion of the full-offset data contain similar structures. FWI resolved a lot of fine structures that were not present in the starting velocity model. However, the long-wavelength structures in the deep parts of the model were not reconstructed because of the lack of very low frequencies and the limited offset.

The comparison between the observed data and synthetic data generated for the final FWI results showed a good match between waveforms. Furthermore, most of the major structures present in the inverted V_P models can be clearly identified in the migrated seismic section, which indicates the reliability of reconstructed structures. The main advantage of the FWI is that it provided not only the structural image of the subsurface but also the information on the velocity distribution.

Chapter 8

Conclusions

In this thesis, I presented the application of the 2D time-domain acoustic full waveform inversion to marine streamer data. I discussed the problems and pitfalls that concern the field data preprocessing, generation of the starting model, and source wavelet estimation. The aim of the synthetic inversion experiments was to provide the guidelines on how to design an efficient inversion scheme concerning the gradient preconditioning, the multi-scale inversion approach, the choice of the minimization criterion, and the choice of the modelling and inversion parameters. Finally, I introduced a workflow for the field data inversion and presented the successful application of acoustic FWI to marine streamer data acquired in the North Sea. The final P-wave velocity model derived from the acoustic full waveform inversion contains a lot of fine structures that were not present in the starting velocity model. A comparison between the observed data and synthetic data generated for the final FWI result shows a good match between waveforms. Furthermore, the reliability of reconstructed structures was assessed by a comparison with the migrated seismic section.

To reconstruct the distribution of material parameters in the subsurface, the inversion algorithm aims to minimize the residuals between the observed data and synthetic data in an iterative process. Because of the high non-linearity of the inverse problem, the success of waveform inversion depends mainly on the accuracy of the starting model and on the presence of low frequencies and sufficient offset range in the recorded data. In practice, when inverting real marine data, we have to deal with some additional challenges like seismic noise, limited offsets, the lack of very low frequencies, and the unknown source wavelet. Furthermore, because we use the 2D acoustic FWI, the elastic effects, attenuation, or 3D effects, which are present in the data, are not taken into account during the inversion process. These factors can lead to poor convergence of the inversion algorithm and may deteriorate the recovery of velocity model. Thus, it is clear that the reconstruction of realistic velocity models from the field data requires an additional effort to overcome the problem of the local minima and to mitigate the effect of non-acoustic factors. Therefore, some extra processing steps and a careful choice of adequate inversion strategies are required for the successful inversion of real data.

In the following, I will summarize the main results of this thesis.

Data preprocessing

Prior to the full waveform inversion, a specific preprocessing has to be applied to the raw seismic data. The main objectives are to improve the signal-to-noise ratio and to transform the field data

such that they reflect the approximations made in the 2D acoustic modelling, i.e. to remove these wave propagation phenomena that cannot be modelled correctly (seismic noise, elastic effects, or 3D effects).

- **Seismic noise** - An efficient attenuation of seismic noise is important to obtain accurate inversion results. Marine data contain high-amplitude, low-frequency ($0 - 10$ Hz) noise generated due to hydrostatic pressure fluctuations and swell noise. Since the presence of low frequencies in the recorded data is crucial for the convergence of FWI, the standard low-cut filters that remove unwanted noise together with the large parts of the useful sesimic signal are not advisable. As an alternative, a time-frequency de-noising algorithm (TFDN) or the $f - k$ dip filtering can be applied to remove the swell noise from the recorded data. However, the efficiency of these filters is limited at very low frequencies due to a poor signal-to-noise ratio. In the field data example presented in this theses, the minimum useful frequency is 3 Hz.

- **Attenuation** - The use of high frequencies for acoustic FWI is limited, because the attenuation effects are not taken into account in the forward modelling. The pure acoustic inversion of the field data produced artefacts in the reconstructed V_P models, when the frequencies higher than 10 Hz were included.

- **3D effects** - Due to the difference between 3D and 2D geometrical spreading, it is necessary to transform the 3D point source data such that they reflect wave propagation in a 2D medium. Otherwise, the field data (point source) cannot be directly compared with the modelled data (line source). In this study, I considered simple 3D to 2D transformations derived from the Green's functions. In spite of strong assumptions on velocity structures, these corrections are efficient for transformation of 3D data recorded in a simple geological environment.

- **Elastic effects** - The synthetic inversion tests show, that the elastic effects limits the applicability of the acoustic FWI especially in the presence of strong contrasts in the S-wave velocity. But even if the converted waves are weak or not present in the recorded data, the main problem of the acoustic approximation is an incorrect modelling of the amplitude-versus-offset (AVO) effects. This problem can be partly mitigated by choosing an appropriate minimization criterion. The lowest V_P model error resulting from the acoustic inversion of elastic data was obtained for the approximated version of the L_2 norm using the normalized wavefields.

- **Data decimation** - Because the total number of shots is one of the main parameters affecting the computation time of FWI, it is useful to decimate the densly sampled measured data. However, the source and the receiver sampling must be sufficiently small to avoid the spatial aliasing, which can significantly reduce the image quality.

Starting model generation

The choice of a staring model is of crucial importance for the success of the waveform inversion. A good initial model for FWI should ensure convergence of the algorithm into the global minimum of the objective function. This can be achieved when the synthetic data generated for the starting model match the observed traveltimes within half a cycle of the dominant frequency. If the half-wavelength

criterion is not fulfilled, cycle skipping occurs, and the waveform inversion will converge toward a local minimum.

The requirements on the starting model for FWI depends on the minimum frequency available in the data. This means that if the low frequencies are not present in the data, the starting model needs to be more accurate and must already contain the long-wavelength structures of the subsurface. In the field data example, the velocity models obtained from both the refraction traveltime tomography and the NMO velocity provided a good match between the field data and the initial synthetics for the first frequency band. Although, the first-arrival traveltime tomography is a widely used method to generate a starting model for FWI, it provides a low-resolution image and requires a wide range of offsets to generate sufficient ray coverage in the deeper parts of the model. With the maximum offset of 4 km, the ray coverage of the refraction traveltime tomography was limited to approximately 1 km.

In addition, the density information is also required to provide realistic reflection coefficients in marine environment. The most significant is the information about density at the seafloor, because the reflection coefficient at this interface is dominated by the strong density contrast. The starting density model can be derived from the initial velocity information using an empirical velocity-density relationship.

Source wavelet estimation

Accurate source wavelet estimation is a prerequisite for successful application of FWI to field data. Wrong wavelet used in forward modelling results in a poor fit of synthetic data to observed data. There are different methods to estimate the wavelet signature and one of them is based on solving a linear least-squares inverse problem. This method provides a very good estimate of the source time function and it offers a number of advantages. It can be easily introduced within the full waveform inversion framework and the source signature is estimated separately for every shot gather. Furthermore, the numerical experiments show that the wavelet inversion can partly compensate for location errors in the acquisition geometry, which is of importance when inverting real data. The inverted source wavelets can also serve as the quality check on the velocity model during the full waveform inversion. When the residuals between the observed and predicted data are decreasing and the estimated source wavelets improve as well, this indicates an accurate reconstruction of the velocity model. The source wavelet estimation is also necessary to perform an initial modelling, which is an important step to evaluate whether the waveform inversion is likely to succeed. If the observed data are not matched to within half a cycle by the synthetic data generated from the starting model, it is likely that the inversion algorithm will attempt to fit the calculated events to the wrong cycle of the observed data and it will fail to recover an accurate velocity model.

Furthermore, the numerical experiments indicate that the point source approximation is sufficient to reproduce the directivity effects of an airgun array for the frequency range used in the field data inversion.

Choice of the inversion strategies

The high non-linearity of the inverse problem can be mitigated by using different inversion strategies:

- **Gradient preconditioning** - The satisfactory performance of the inversion algorithm can be achieved by applying a preconditioning operator that combines a spatial taper with a linear scaling function. A spatial taper, which turns off the model update in the water layer, allows to suppress the artefacts related to the acquisition geometry and to improve the quality of inverted models. In addition, the effect of geometrical spreading of the waveform amplitudes can be effectively compensated by applying a simple depth-dependent linear preconditioning operator to the gradient of the objective function.

- **Multi-scale inversion** - The multi-scale inversion is a very effective method to reduce the non-linearity of the inverse problem and to mitigate the problem of local minima in the objective function. Because the objective function at low frequencies is more linear with respect to the model perturbations than at high frequencies, the inversion starts at low frequencies and higher frequency content is gradually added. When the dominant frequency present in the data is low, it is also easier to satisfy the half-wavelength condition imposed on the starting model and the chance of the inversion to reach the global minimum is higher.

  In this work, the multi-scale strategy is realized by low-pass filtering of the observed data. The selection of optimal frequency bands is based on the continuous coverage of vertical wavenumbers. The choice of optimal frequencies is of high importance to reduce the computational cost of the waveform inversion and to take full advantage of the multi-scale approach, i.e. to avoid the cycle-skipping problem. The first frequency band should be relatively narrow and contain the lowest available frequencies with a high signal-to-noise ratio.

- **Time windowing + offset windowing** - The combination of time windowing and/or offset windowing of the data has a significant effect on the inversion results. Because the early arrivals are more linear than the late arrivals, a time windowing of the data around the first arrivals helps to reduce the risk of cycle skipping and to add stability to the inversion. Furthermore, the successive inversion from shallow to deep structures, which is similar to the layer stripping approach, results in a good quality of the reconstructed model.

  Although, the far offsets are the most non-linear components of the data with the high risk of cycle-skipping, they are essential to reconstruct the long-wavelength structures of the model, especially when the low frequencies are not available in the data. The far-offset data are more sensitive to the macro variations in seismic velocity than the near offset data and allow a better quantitative estimation of the velocity.

  The combination of an offset windowing with a time windowing for each frequency band is much more complex. If late arrivals are included at early stages of inversion, the reconstructed models contain a lot of artefacts. On the other hand, a combination of the offset windowing with the extending time windowing, such that the inversion proceeds from the near offset to the far offset data and from early arrivals to late arrivals resulted in much better inversion results.

- **Choice of the inversion parameters** - The mono-parameter acoustic FWI with the P-wave velocity as the only unknown can be easily extended to the multi-parameter inversion. An important aspect of the multi-parameter inversion is the choice of the parameters describing the medium, which can influence the convergence rate and the ambiguity of the inverse problem. I

investigated three different combinations of parameters: P-wave velocity and density, acoustic impedance and density, P-wave velocity and acoustic impedance. Whereas the resolution of velocity and impedance models is comparable, the reconstruction of density structures strongly depends on the model parameterization. Out of the investigated parameter sets, the velocity and density provided the best convergence rate and the best accuracy of the inverted results for the inversion of synthetic acoustic data. On the other hand, an accurate reconstruction of the density model turned out to be very difficult with the synthetic elastic data, because the density inversion is very sensitive to amplitude errors resulting from the acoustic approximation. The similar difficulty with resolving a reliable density model was encountered in the case of the multi-parameter inversion of the field data.

- **Minimization criteria** - If the data are dominated by isolated, high amplitude noise, such as the swell noise, it is favorable to apply the minimization criteria which are less sensitive to noise than the conventional L_2 norm. The numerical example with the swell noise added to the data showed the robust behaviour of the L_1, Cauchy and sech criteria. Because the preprocessed field data were not affected by high amplitude noise, the inversion results obtained for different minimization criteria were very similar and even the L_2 norm produced a good quality image.

Outlook

To improve the quality of inverted models and to include higher frequency components, it will be necessary to apply a visco-acoustic FWI to this field data set. Because the time-domain inversion does not allow for a separate inversion of phase and amplitude information, the application of the frequency-domain inversion will be a possible direction for future investigations. Furthermore, to improve the resolving power and convergence properties of the inversion algorithm, a more sophisticated gradient preconditioning should be applied, such as the approximate Hessian matrix or the L-BFGS method.

Appendix A

Marine processing sequence

I applied a conventional seismic data processing to the field data presented in Chapter 7. The 2D marine data processing was performed with GLOBE Claritas. The processing sequence comprised the following steps:

1. Reformat and edit out bad traces
2. Mute to remove refractions and direct arrivals
3. Geometrical spreading correction: spherical divergence and balance (the average amplitude of the output trace is constant for all traces)
4. Bandpass Butterworth filter: low cut 3 Hz, high cut 60 Hz
5. FK-filtering to attenuate swell noise
6. Minimum-phase deconvolution
7. Set up of field geometry
8. CMP sorting
9. Creating semblance spectra and picking velocities
10. NMO correction and brute stack
11. Velocity analysis
12. NMO correction
13. Stacking
14. Post-stack Kirchoff time migration
15. Gain recovery: linear gain to recover the relative amplitudes of the section.

Bibliography

Alford, R. M., Kelly, R. M., and Boore, D. M. (1974). Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, 39(6):834–842. 10

Amundsen, L. (1991). Comparison of the least-squares criterion and the cauchy criterion in frequency-wavenumber inversion. *Geophysics*, 56(12):2027–2035. 41

Amundsen, L. (1993). Wavenumber-based filtering of marine point-source data. *Geophysics*, 58(9):1335–1348. 101

Amundsen, L. and Reitan, A. (1994). Transformation from 2-D to 3-D wave propagation for horizontally layered media. *Geophysics*, 59(12):1920–1926. 101

Amundsen, L. and Ursin, B. (1991). Frequency-wavenumber inversion of acoustic data. *Geophysics*, 56(7):1027–1039. 82

Assous, F. and Collino, F. (1990). A numerical method for the explanation of sensitivity: the case of the identification of the 2d stratified elastic medium. *Inverse Problems*, 6:487–513. 60

Bae, H., Shin, C., Cha, Y., and Choi, Y. (2010). 2D acoustic – elastic coupled waveform inversion in the laplace domain. *Geophysical Prospecting*, 58:997–1010. 81

Barner, C. and Charara, M. (2009). The domain of applicability of acoustic full-waveform inversion for marine seismic data. *Geophysics*, 74(6):WCC91–WCC103. 44, 81

Behura, J., Kabir, N., Crider, R., Jilek, P., and Lake, E. (2010). Density extraction from p-wave avo inversion: Tuscaloosa trend example. *The Leading Edge*, 29(7):772–777. 82

Berenger, J. P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200. 12

Berkhout, A. J. (1977). Least-squares inverse filtering and wavelet deconvolution. *Geophysics*, 42:1369–1383. 110

Beydoun, W. and Tarantola, A. (1988). First Born and Rytov approximation: Modeling and inversion conditions in a canonical example. *Journal of the Acoustical Society of America*, 83:1045–1055. 32

Bleibinhaus, F., Lester, R., and Hole, J. (2009). Applying waveform inversion to wide-angle seismic surveys. *Tectonophysics*, 472:238–248. 2, 18, 146

Bleibinhaus, F. and Rondenay, S. (2009). Effects of surface scattering in full-waveform inversion. *Geophysics*, 74(6):WCC69–WCC77. 71, 130

Bleistein, N. (1986). Two-and-one-half dimensional in-plane wave propagation. *Geophysical Prospecting*, 34:686–703. 101

Bohlen, T. (2002). Parallel 3d viscoelastic finite difference seismic modelling. *Computers and Geosciences*, 28:887–899. 1, 16, 44, 102

Bohlen, T. and Saenger, E. (2006). Accuracy of heterogeneous staggered-grid finite-difference modeling of rayleigh waves. *Geophysics*, 71(4):T109–T115. 11

Boonyasiriwat, C., Schuster, G., Valasek, P., and Cao, W. (2010). Applications of multiscale waveform inversion to marine data using a flooding technique and dynamic early-arrival windows. *Geophysics*, 75(6):R129–R136. 2, 59, 81

Boonyasiriwat, C., Valasek, P., Routh, P., Macy, B., Cao, W., and Schuster, G. (2008). An application of time-domain multiscale waveform tomography to marine data. In *SEG Expanded Abstracts*, volume 27. 107

Boonyasiriwat, C., Valasek, P., Routh, P., Macy, B., Cao, W., and Schuster, G. (2009). An efficient multiscale method for time-domain waveform tomography. *Geophysics*, 74(6):WCC59–WCC68. 32

Brenders, A. and Pratt, R. (2007a). Efficient waveform tomography for lithospheric imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data. *Geophysical Journal International*, 168:152–170. 146, 147

Brenders, A. and Pratt, R. (2007b). Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, 168:133–151. 18, 32, 130

Brenders, A. J. (2011). *Strategies for waveform tomography for long-offset, 2-D exploration seismic data*. PhD thesis, The University of Western Ontario, Canada. 112

Brocher, T. M. (2005). Empirical relations between elastic wavespeeds and density in the Earth's crust. *Bull. Seism. Soc. Am.*, 95:2081–2092. 85

Brossier, R., Operto, S., and Virieux, J. (2009). Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC105–WCC118. 24, 32

Brossier, R., Operto, S., and Virieux, J. (2010). Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75(3):R37–R46. 41, 48

Bunks, C., Saleck, F., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473. 32

Castagna, J., Batzle, M., and Kan, T. (1993). Rock physics - the link between rock properties and avo response. In Castagna, J. and Backus, M., editors, *Offset dependent reflectivity, Theory and Practice of AVO analysis*, pages 135–171. SEG. 83

Causse, E., Mittet, R., and Ursin, B. (1999). Preconditioning of full-waveform inversion in viscoacoustic media. *Geophysics*, 64(1):130–145. 24

Chapman, C. (2004). *Fundamentals of Seismic Wave Propagation*. Cambridge University Press, Cambridge. 100

Chew, W. C. and Weedon, W. H. (1994). A 3-d perfectly matched medium from modified maxwell's equations with stretched coordinates. *Microwave and Optical Tech. Lett.*, 7:599–604. 12

Choi, Y. and Alkhalifah, T. (2012). Multi-source waveform inversion of marine streamer data using the normalized wavefield. In *74th Conference and Technical Exhibition, EAGE, Expanded Abstracts*, page P347. 41, 43

Courant, R., Friedrichs, K., and Lewy, H. (1928). Ueber die partiellen differenzengleichungen der mathematischen physik. *Matchematische Annalen*, 100:32–74. 12

Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic nonlinear waveform inversion: Application to real data. *Geophysics*, 55:527–538. 41, 48, 102, 107

Debski, W. and Tarantola, A. (1995). Information on elastic parameteres obtained from the amplitudes of reflected waves. *Geophysics*, 60(5):1426–1436. 60, 82

Delescluse, M., Nedimović, M., and Louden, K. (2011). 2D waveform tomography applied to long-streamer MCS data from the Scotian Slope. *Geophysics*, 76:B151–B163. 2, 59, 81, 107

Deregowski, S. and Brown, S. (1983). A theory of acoustic diffractors applied to 2-D models. *Geophysical Prospecting*, 31:293–333. 101

Dessa, J.-X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., and Virieux, J. (2004). Multiscale seismic imaging of the eastern nankai trough by full waveform inversion. *Geophysical Research Letters*, 31. 2

Dessa, J.-X. and Pascal, G. (2003). Combined traveltime tomography and frequency-domain seismic waveform inversion: a case study on multi-offset ultrasonic data. *Geophysical Journal International*, 154:117–133. 2, 18

Djikpéssé, H. and Tarantola, A. (1999). Multiparameter l_1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms. *Geophysics*, 64(4):1023–1035. 41, 44

Elboth, T., Pettersson Reif, B., and Andreassen, Ø. (2009). Flow and swell noise in marine seismic data. *Geophysics*, 74(2):Q17–Q25. 124

Elboth, T., Qaisrani, H., and Hertweck, T. (2008). De-noising seismic data in the time-frequency domain. In *SEG Expanded Abstracts*, volume 27. 125

Ernst, J., Green, A., Maurer, H., and Holliger, K. (2007). Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data. *Geophysics*, 72(5):353–364. 2

Fichtner, A., B.L.N., K., Igel, H., and Bunge, H.-P. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint method. *Geophyscial Journal International*, 179:1703–1725. 2, 24, 25, 26

Forgues, E. and Lambaré, G. (1997). Parameterization study for acoustic and elastic ray + born inversion. *Journal of Seismic Exploration*, 6:253–278. 81

Gadallah, M. R. and Fisher, R. L. (2005). *Applied seismology*. PennWell Corporation. 118, 129

Gardner, G., Gardner, L., and Gregory, A. (1974). Formation velocity and density—the diagnostic basics for stratigraphic traps. *Geophysics*, 39:770–780. 85

Guitton, A. and Symes, W. (2003). Robust inversion of seismic data using the huber norm. *Geophysics*, 68(4):1310–1319. 41

Ha, T., Chung, W., and Shin, C. (2009). Waveform inversion using a backpropagation algorithm and a huber function norm. *Geophysics*, 74:R15–R24. 41

Hadamard, J. (1902). Sur les problémes aux dérivées partielles et leur signification physique. *Princeton University Bulletin*, pages 49–52. 5

Hicks, G. and Pratt, G. (2001). Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit. *Geophysics*, 66(2):598–612. 2, 81, 102, 107

Ikelle, L. and Amundsen, L. (2005). *Introduction to petroleum seismology*. Society of Exploration Geophysicists. 101, 102, 108, 116

Jaiswal, P., Zelt, C. A., Bally, A. W., and Dasgupta, R. (2008). 2-D traveltime and waveform inversion for improved seismic imaging: Naga thrust and fold belt, India. *Geophysical Journal International*, 173:642658. 130

Jannane, M., Beydoun, W. B., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., Singh, S., Snieder, R., Tarantola, A., Trezeguet, D., and Xie, M. (1989). Wavelengths of earth structures that can be resolved from seismic reflection data. *Geophysics*, 54:906–910. 60

Kabir, N., Crider, R., Ramkhelawan, R., and Baynes, C. (2006). Can hydrocarbon saturation be estimated using density contrast parameter? *CSEG Recorder*, pages 31–37. 82

Kelly, S., Ramos-Martinez, J., Tsimelzon, B., and Crawley, S. (2010). Application of an impedance-based full-waveform inversion method for dual-sensor, single-streamer field recordings. *72nd Conference and Technical Exhibition, EAGE, Extended Abstracts*, A020. 59, 81

Köhn, D. (2011). *Time Domain 2D Elastic Full Waveform Tomography*. PhD thesis, Christian-Albrechts-Universität zu Kiel. 12, 24, 103

Köhn, D., De Nil, D., Kurzmann, A., Przebindowska, A., and Bohlen, T. (2012). On the influence of model parametrization in elastic full waveform tomography. *Geophyscial Journal International*, 191:325–345. 60

Kolb, P. and Canadas, G. (1986). Least-squares inversion of prestack data: Simultaneous identification of density and velocity of stratified media. *SEG Annual Meeting, November 2 - 6, 1986 , Houston, Texas*. 60

Kravis, S. P. (1985). Estimation of marine source signatures from direct arrivals to hydrophone groups. *Geophysical Prospecting*, 33:987–998. 107

Kurzmann, A. (2012). *Applications of 2D and 3D full waveform tomography in acoustic and viscoacoustic comlex media*. PhD thesis, Karlsruhe Institute of Technology. 10, 12, 14, 15, 129

Kurzmann, A., Köhn, D., Przebindowska, A., Nguyen, N., and Bohlen, T. (2009). 2d acoustic full waveform tomography: Performance and optimization. 71st conference and technical exhibition, eage, extended abstracts, p006. 13, 16

Levander, A. (1988). Fourth-order finite-difference p-sv seismograms. *Geophysics*, 53:1425–1436. 11, 12

Li, G., Cao, M., Chen, H., and Ni, C. (2011). Modelling the signature of clustered airguns and analysis on the directivity of an airgun array. *Journal of Geophysics and Engineering*, 8:92–98. 118

Li, Y. (2005). A study on applicability of density inversion in defining reservoirs. In *SEG Expanded Abstracts*, pages 1646–1649. 82

Malinowski, M. and Operto, S. (2008). Quantitative imaging of the permo-mesozoic complex and its basement by frequency domain waveform tomography of wide-aperture seismic data from the Polish Basin. *Geophysical Prospecting*, 56:805–825. 2, 18

Martin, G., Wiley, R., and Marfurt, K. (2006). Marmousi2 - an elastic upgrade for Marmousi. *The Leading Edge*, 25:156–166. 20, 83

Menke, W. (1984). *Geophysical data analysis: discrete inverse theory*. Academic Press, Inc., New York. 111

Monteiller, V., Got, J.-L., Virieux, J., and Okubo, P. (2005). An efficient algorithm for double-difference tomography and location in heterogeneous media, with an application to the kilauea volcano. *J. Geophys. Res.*, 110. 41

Mora, P. (1987). Nonlinear two-dimensional elastic inversion of multioffset seismic data. *Geophysics*, 52:1211–1228. 1, 2, 7, 8, 13, 61, 62

Mulder, W. and Plessix, R.-E. (2008). Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56:827–841. 44

Nafe, J. E. and Drake, C. L. (1963). Physical properties of marine sediments. In Hill, M. N., editor, *The Sea*, volume 3, pages 794–815. Interscience, New York. 85

Nocedal, J. and Wright, S. (1999). *Numerical optimization*. Springer. 13, 24

Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A., and Dell'Aversana, P. (2004). Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study. *Geophysical Prospecting*, 52:625–651. 59, 81

Operto, S., Virieux, J., Dessa, J.-X., and Pascal, G. (2006). Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern Nankai trough. *Journal of Geophysical Research*, 111. 2, 32, 71, 102

Oprsal, I. and Zahradnik, J. (1999). Elastic finite-difference method for irregular grids. *Geophysics*, 64:240–250. 12

Parkes, G. and Hatton, L. (1986). *The marine seismic source*. D. Reidel Publishing Company, The Netherlandes. 116

Pica, A., Diet, J., and Tarantola, A. (1990). Nonlinear inversion of seismic reflection data in a laterally invariant medium. *Geophysics*, 55(3):284–292. 15, 41

Plessix, R.-E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167:495–503. 7

Pratt, R. (2008). Waveform tomography - successes, cautionary tales, and future directions. In *70th Conference and Technical Exhibition, EAGE, Expanded Abstracts*. 17

Pratt, R. G. (1990). Inverse theory applied to multi-source cross-hole tomography, Part II: Elastic wave-equation method. *Geophysical Prospecting*, 38:311–330. 2

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics*, 64:888–901. 2, 18, 107

Pratt, R. G., Shin, C., and Hicks, G. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133:341–362. 24

Pratt, R. G. and Shipp, R. (1999). Seismic waveform inversion in the frequency domain, Part 2: Fault delineation in sediments using crosshole data. *Geophysics*, 64:902–914. 2

Pratt, R. G., Song, Z.-M., Williamson, P., and Warner, M. (1996). Two-dimensional velocity models from wide-angle seismic data by waveform inversion. *Geophysical Journal International*, 124:323–340. 71

Pratt, R. G. and Worthington, M. (1988). The application of diffraction tomography to cross-hole seismic data. *Geophysics*, 53:1284–1294. 130

Pratt, R. G. and Worthington, M. (1990). Inverse theory applied to multi-source cross-hole tomography, Part I: Acoustic wave-equation method. *Geophysical Prospecting*, 38:287–310. 2

Prieux, V., Lambaré, G., Operto, S., and Virieux, J. (2010). Building starting models for full waveform inversion from wide-aperture data by stereotomography. In *SEG Expanded Abstracts*, volume 29. 18

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and Dell'Aversana, P. (2004). Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International*, 159:1032–1056. 2, 18, 32

Roberts, M. (2005). A line source to point source transformation for seismograms. *Lithos Science Report*, 7:49–52. 101

Sheen, D.-H., Tuncay, K., Baag, C.-E., and Ortoleva, P. (2006). Time domain Gauss-Newton seismic waveform inversion in elastic media. *Geophys. J. Int.*, 167:1373–1384. 24

Sheriff, R. and Geldart, L. (1982). *Exploration Seismology Vol. 1: History, Theory and Data Acquisition*. Cambridge University Press, Cambridge. 118

Shin, C. and Cha, Y. (2008). Waveform inversion in the laplace domain. *Geophysical Journal International*, 173:922–931. 18

Shin, C. and Cha, Y. (2009). Waveform inversion in the laplace-fourier domain. *Geophysical Journal International*, 177:1067–1079. 18

Shin, C. and Min, D.-J. (2006). Waveform inversion using a logarithmic wavefield. *Geophysics*, 71(3):R31–R42. 2, 41

Shin, C., YOOn, K., Marfurt, K. J., Park, K., Yang, D., Lim, H., and Chung, S. (2001). Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion. *Geophysics*, 66:1856–1863. 24

Shipp, R. and Singh, S. (2002). Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. *Geophys. J. Int.*, 151:325–344. 2, 18, 59, 81, 102, 107, 118, 129

Sirgue, L. (2003). *Inversion de la forme d'onde dans le domaine fréquentiel de données sismiques grands offsets*. PhD thesis, Université Paris and Queen's University. 17, 71, 130, 141

Sirgue, L., Barkved, O., Gestel, J., Askim, O., and Kommedal, J. (2009). 3D waveform inversion of on Valhall wide-azimuth OBC. In *71st Conference and Technical Exhibition, EAGE, Expanded Abstracts*. 2, 18

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69:231–248. 32, 33

Song, Z.-M. and Williamson, P. (1995). Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part I–2.5-D modeling method. *Geophysics*, 60(3):784–795. 100

Song, Z.-M., Williamson, P., and Pratt, R. (1995). Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part II-inversion method, synthetic experiments and real-data results. *Geophysics*, 60(3):796–809. 101

Sourbier, F., Operto, S., Virieux, J., Amestoy, P., and L'Excellent, J. (2009). FWT2D:a massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data Part 2: Numerical examples and scalability analysis. *Computers and Geosciences*, 35:496–514. 32

Sun, R. and McMechan, G. (1992). 2-D full-wavefield inversion for wide-aperture, elastic, seismic data. *Geophyscial Journal International*, 111:1–10. 17

Takam Takougang, E. and Calvert, A. (2011). Application of waveform tomography to marine seismic reflection from the Queen Charlotte Basin of western Canada. *Geophysics*, 76(2):B55–B77. 2, 129, 151

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49:1259–1266. 1, 2, 7, 9, 61

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51:1893–1903. 2, 59, 60, 63, 82

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial and Applied Mathematics, Philadelphia. 41

Vigh, D. and Starr, E. (2008). 3D prestack plane-wave, full-waveform inversion. *Geophysics*, 73(5):VE135–VE144. 2, 107

Virieux, J. (1986). P-sv wave propagation in heterogeneous media: velocity-stress finitedifference method. *Geophysics*, 54(4):889–901. 11

Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC127–WCC152. 7, 16, 18, 19, 24, 32, 59, 60, 81

Wang, Y. and Rao, Y. (2006). Crosshole seismic waveform tomography - I. Strategy for real data application. *Geophysical Journal International*, 166:1224–1236. 2

Wang, Y. and Rao, Y. (2009). Reflection seismic waveform tomography. *Journal of Geophysical Research*, 114. 2, 18, 24, 101, 102, 151

Wapenaar, C., Verschuur, D., and Herrmann, P. (1992). Amplitude preprocessing of single and multicomponent seismic data. *Geophysics*, 57(9):1178–1188. 101

Williamson, P. (1991). Resolution limits in ray tomography due to wave behaviour. *Geophysics*, 56:202–207. 18

Williamson, P. and Pratt, R. (1995). A critical review of acoustic wave modeling procedures in 2.5 dimensions. *Geophysics*, 60(2):591–595. 101

Woodward, M. (1992). Wave-equation tomography. *Geophysics*, 57(1):15–26. 71

Zahradnik, J., Moczo, P., and Hron, F. (1993). Testing four elastic finite difference schemes for behaviour at discontinuities. *Bulletin of the Seismological Society of America*, 83:107–129. 11

Zhou, C., Cai, W., Luo, Y., Schuster, G., and Hassanzadeh, S. (1995). Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data. *Geophysics*, 60(3):765–773. 102

Zoeppritz, K. (1919). Erdbebenwellen viii b, on the reflection and penetration of seismic waves. *Goettinger Nachrichten*, 1:66–84. 82

Acknowledgements

Although this is the last page of my thesis, it definitely doesn't mean that it's the least important one. Here, I would like to acknowledge all the people who supported me through my PhD work.

First of all, I would like to thank **Thomas Bohlen** for inviting me to join the working group and for giving me a lot of freedom to pursue this research topic. I am very grateful for many valuable discussions, for the supervision of my thesis, and for giving me an opportunity to present the work at national and international conferences.

I am very thankful to **Daniel Köhn** for guiding me through the meanders of scientific work, especially during my first year in Freiberg. His valuable tips, suggestions, and many geophysical discussions were invaluable and helped me to make this thesis come true.

Special thanks to **André Kurzmann** for many scientific discussions and for his support concerning the inversion code. It was very nice to fight on the same side of the battle: acoustic vs elastic.

I would also like to thank **Florian Bleibinhaus** for agreeing to review my thesis and for discussing the results of my work.

Many thanks to the **Freiberg team**: to Denise De Nil for sharing an office and for many (not always scientific) conversations; to Olaf Hellwig for his infectious positive attitude; to Klaus Spitzer for making my first stay in Freiberg possible.

I would like to thank all colleagues from the **Karlsruhe team**: Lisa Groos for countless discussions on modelling and inversion problems, Simone Dunkl for being a great fellow at many conferences, Claudia Payne for the constant help with all administrative issues, many thanks to Anja Diez, Petra Knopf, Ines Veile, Thomas Forbriger, Jörn Groos, Sven Heider, Stefan Jetschny, Jürgen Mann, and Martin Schäfer.

I want to thank **Verbundnetzgas AG (VNG)** for delivering the field data set and for the financial support of my work. Furthermore, many thanks to the sponsors of **Wave Inversion Technology Consortium (WIT)** for the financial support and for the fruitful discussions during the annual meetings in Hamburg.

Szczegolne podziekowania dla moich rodzicow za wiare we mnie i za to, ze zawsze moglam i moge liczyc na Wasza pomoc i wsparcie. Regularne zapytania jak posuwa sie praca nad doktoratem wyjatkowo dobrze wplywaly na moja mobilizacje do pracy.

Ogromnie podziekowania dla Slawomira: za wyrozumialosc, za dopingowanie mnie do pracy, za berlinskie tygodnie, za dbanie o to, zeby nie wyrosl mi garb od ciaglego siedzenia przed komputerem, za pokazanie mi, ze swiat z perspektywy dwoch kolek wyglada zupelnie inaczej, za ... i tak moglabym dalej wymieniac, ale nie chce przekroczyc 200 stron.