Imperial College London
Department of Earth Science and Engineering

Full waveform inversion of narrow-azimuth towed-streamer seismic data

Tatiana I. Kalinicheva

October 2019

Supervised by Prof. Michael R. Warner

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy of Imperial College London and
the Diploma of Imperial College London

Declaration of Originality

I herewith certify that all material within this thesis that is not my own research has been fully acknowledged.

Tatiana I. Kalinicheva

Copyright Declaration

Abstract

Full waveform inversion (FWI) is a computational scheme that produces high-fidelity, high-resolution models of the Earth's subsurface from surface seismic data. FWI has become a standard tool in velocity-model building and performs well on full-azimuth long-offset ocean-bottom seismic datasets. However, the majority of marine seismic datasets use narrow-azimuth towed streamers (NATS) which often lack long-offset refracted energy. Here I explore the capability of conventional FWI when it is applied to marine deep-water reflection-dominated NATS field data. I applied FWI to three datasets: the first used a deep-towed 10-km cable and was specifically acquired for 2D FWI; the other two datasets were both 3D reflection-dominated surveys to which FWI had been previously applied with limited success - these datasets were from Gabon and Brazil, and were chosen specifically because FWI had been previously tried and had failed. Applying FWI to these datasets, I reached the following conclusions: 1) When the input data have adequate turning energy and adequate low-frequency energy, acoustic anisotropic FWI can generate accurate high-resolution velocity models of increasing complexity and resolution up to about 40 Hz. 2) Extending FWI to the full bandwidth of the field data produces minimal further change in the macro-velocity model, but nonetheless continues to improve resolution up to and perhaps beyond that which can be recovered by conventional Kirchhoff-based pre-stack depth migration. 3) Applying 2D and full-3D FWI to a single 2D sail line produces similar outcomes. 4) The Gabon dataset proved almost entirely resistant to FWI; the available evidence suggests that the nominally-raw field data were corrupt in some unknown way. 5) The Brazilian dataset was inverted using a third-party FWI code that assumed constant density; this assumption is detrimental to FWI.

Acknowledgements

First and foremost, I would like to thank my supervisor, Mike Warner, for constant support, guidance, long discussions and for making this thesis possible at all. Thank you so very much.

I would like to express my gratitude to every member of the FULLWAVE research group, who all contributed in various ways to this research: Adrian Umpleby, Lluís Guasch, Gang Yao, Tim Burgess, Nuno Vieira da Silva, Henry Debens, Jiashun Yao, Tim Lin, Tenice Nangoo, Michele Paulatto, Kajetan Chrapkiewicz, Jack Ashley and Melissa Gray. In particular, I would like to say special thank you to Melissa Gray for her moral support and laughs throughout this time.

I also would like to thank the members of the FULLWAVE research consortium who provided the sponsorship that allowed me to conduct this research.

Thank you ION Technology for giving me the internship opportunity and kindly providing the permissions to use the materials for this thesis. I would like to say thank you to John Brittan, Ian Jones, Claudia Hagen and Stuart Greenwood for their support and valuable advices.

For providing me with many laughs and great time I would like to thank Francesco Palci, Pete Gordon and Mosayeb Shams.

I would like to say special thank you to Matthieu Zandecki, for giving moral support and help during this research.

To my parents: thank you so much for your love and support that have made it possible for me to have this opportunity. You have contributed more than you will ever know.

To my brother: thank you for always being there and our many many laughs.

To my grandma: thank you for your care and belief in me.

To my uncle: thank you so very much for always encouraging me.

Dedication

To my mum, dad and uncle who always encouraged curiosity.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Chapter overview

This thesis is organised in 8 chapters. Chapter 1 presents an overview of full waveform inversion (FWI) and its historical evolution, demonstrates the value of the method in the context of seismic imaging and, finally, presents the motivation for this research. Chapter 2 describes the mathematical basis of FWI and outlines its implementation. It briefly describes the challenges of conventional FWI and a range of alternative FWI formulations. Finally, it gives practical recommendations for successful FWI application to marine datasets. Chapter 3 introduces three narrow-azimuth towed-streamer datasets, that will be used to conduct all the studies for this thesis. Chapter 4, 5, 6 and 7 present the main body of the thesis. Chapter 4 details the application of acoustic anisotropic time-domain FWI to a towed-streamer dataset in both 2D and 3D and examines the utility and limitations of the former. Chapter 5 explores the utility and cost of conventional FWI at very-high frequencies. Chapter 6 explores a dataset that seems to be suitable for FWI, but that proves to be “difficult” for no obvious reason. Chapter 7 is devoted to another “difficult” dataset which was tackled using a selection of alternative FWI types during my 3-month internship at ION Geophysical. Chapter 8 summarises the conclusions of the research and gives a possible scope for further work.

1.2 Objectives of study

Over the last decade full waveform inversion (FWI) has become part of the standard workflow for velocity model building. It usually works well on ocean-bottom cable datasets (OBC) as they naturally contain high quality low-frequency information and transmitted arrivals (Plessix et al. 2010; Sirgue et al. 2009; Warner et al. 2013), beneficial for successful FWI implementation (Sirgue 2006). However, the majority of acquired datasets use narrow-azimuth towed streamers (NATS). Because they usually contain lower quality low-frequency information, limited transmitted arrivals due to relatively short offsets, and reduced azimuthal coverage, they are not ideal for conventional least-squares FWI. The general scope of this thesis is to explore the capacity of acoustic anisotropic time-domain least-squares FWI applied to reflection dominated NATS datasets, and especially to those datasets that have already proven to be difficult for FWI.

A set of three NATS datasets was explored. All the datasets have reasonably deep targets, principally dominated by reflection data, contain a relatively low amount of noise and seem to be suitable for FWI. However, two out of three were proven to be “difficult” for the inversion.

The first dataset (the Pivot dataset) was a 2D NATS regional seismic line with unusually long offsets, a deep-towed streamer and a large deep-towed source array. This dataset was subjected to two independent studies. Firstly, by inverting the dataset in both 2D and 3D to explore the differences in performance and effective cost between 2D and 3D FWI. It was interesting to find out whether 2D FWI has a role to play in the world of large 3D seismic surveys, especially taking into account extensive cost of 3D FWI applied to large data volumes. Secondly, it was set to explore the capacity of conventional FWI applied to full-bandwidth seismic data. To date, only FWI models at 40 Hz can be found in literature (Lu et al. 2016; Ruth et al. 2017). Consequently, the question arises whether using higher frequencies brings any benefit and if so, what is the cost.

The second dataset (the Gabon dataset) was inverted previously by two other groups and both times the inversion has failed to produce a sensible result. The goal of the research was ultimately to successfully invert this dataset using conventional FWI only and if FWI fails, understand the reasons.

The third dataset (the Picanha dataset) was explored during my 3-month internship at ION Geophysical. This dataset was part of a large seismic survey which was divided into two phases, “west” and “east”. Both phases had a similar geological environment and principally similar acquisition. Although conventional FWI worked well on the “west” dataset, it did not significantly improve the starting model of the “east” dataset. The goal was to explore the capacity of a series of different FWI types in order to successfully invert the data. All inversions of this dataset were carried out using a different to $fullwave3D$ code.

1.3 Full waveform inversion

Conventional seismic imaging relies on having the correct velocity to appropriately migrate the reflectivity in the corresponding model. Without the correct velocity model, reflectors will be positioned at the wrong depth and have incorrect geometry. In addition, horizontal resolution on the scale of the local seismic wavelength may not be possible at all. Thus, the geological interpretation will be compromised.

Reflectivity R, which is in essence a derivative of impedance $I = \rho V$, where ρ is density and V is velocity

$$R = \frac{\rho_1 V_1 - \rho_2 V_2}{\rho_1 V_1 + \rho_2 V_2} \tag{1.1}$$

depends on density and is observed in the acquired seismic data across the whole available spectrum.

Typically, frequencies range from 10 to 100 Hz for old seismic surveys with the tendency to widen the low end of the spectrum down to 5-6 Hz in modern low-frequency acquisition of marine streamer data.

Velocity can be obtained from the study of travel times. These are not sensitive to density variations in the model and provide the spatial resolution restricted by the size of a local Fresnel zone (Williamson 1991). Such spatial resolution is roughly equivalent to that of seismic wavelengths below about 2 Hz (Claerbout 1985).

For these reasons, velocity is known to be a long-wavelength component and reflectivity is a short-wavelength component of “real” velocity, which in principle can be measured in a borehole (Biondi et al. 2013). Figure 1.1 demonstrates a simplified 1D graphical representation of this concept (Claerbout 1985). Note, that the spatial resolution discussed previously addresses only vertical direction because reflections are normally horizontal or sub-horizontal and the largest part of energy travels primarily in a near-vertical direction. Horizontal resolution is generally far less than vertical for both reflectivity and velocity estimated from the travel times.

Therefore, there is a distinct gap in information between the velocity and reflectivity. As the industry strives to narrow the information gap for the benefits of migration and, as a result, geological interpretation, growing computational power has allowed for the development of the full-waveform inversion technique, that simultaneously estimates the velocity and reflectivity by taking advantage of all the information in the data. FWI attempts to overcome conventional imaging limitations and improve the quality of migrated data.

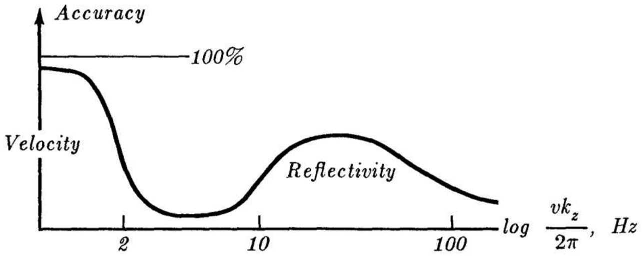


Figure 1.1: 1D graphical representation showing information obtained from surface seismic measurements. FWI attempts to fill the gap between the velocity and reflectivity (after Claerbout (1985)).

1.3.1 FWI and Traveltime Tomography

The word “tomography” derives from the Greek from “tomo” for “slice” and “graph” for “draw”. In other words, it describes an object based on a collection of slices through it. In seismics, tomography sets out to obtain a velocity distribution of the medium based on a series of measurements made of travel times and/or amplitudes associated with reflections, refractions and transmissions (Jones 2010). It also can be used to estimate attenuation, or absorption (Hung et al. 2008; Xin et al. 2008). Strictly speaking, FWI is a type of tomographical algorithm, that utilises the wave-equation and full waveforms (both travel times and amplitudes of the arrivals).

Here, I will briefly discuss ray-based reflection tomography (or traveltime tomography, often shortened to tomography) as this technique is widely used for generating starting velocity models for FWI.

Traveltime tomography, as the name suggests, uses only travel times and does not deal with amplitudes. It is based on the assumption that all the frequencies in the signal travel with the same speed so that there is no frequency dispersion (Luo et al. 1991). It seeks to determine the velocity distribution along each individual ray path. In order to do this, it solves a ray approximation to the scalar wave equation - Eikolal equation given by Yilmaz (2001)

$$\nabla^2 T(\mathbf{r}) = \frac{1}{V_p^2(\mathbf{r})} \tag{1.2}$$

where T is spatially varying traveltime, V_p is the velocity of the medium and $\mathbf{r}$ denotes coordinates. The arrival time T along the path s can be computed as (Yilmaz 2001)

$$T = \int_s \frac{1}{V_p(s)} ds \tag{1.3}$$

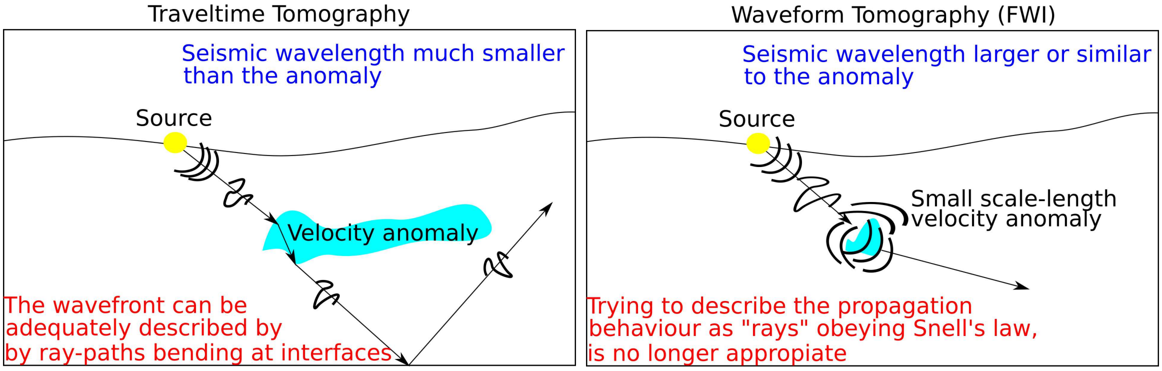
Tomography typically computes the arrival times for every offset of every common image gather (CIG) location, for all the CIG and for all reflections that can be identified. Having enough equations, the algorithm will try to generate a model which best fits the observed data.

Using the ray-paths to describe the wavefront is only an acceptable approximation if the seismic wavelength is much smaller than the anomaly which needs to be resolved (Fig. 1.2(a)). Once the wavelength size is comparable to or larger than the anomaly size, then this anomaly will act as a scatter and diffraction will occur instead of refraction (Fig. 1.2(b)). Snell's law fails to describe diffractions and, therefore, is no longer applicable. In this case, the wave equation must be used. Consequently, traveltime tomography is known to be a "high frequency approximation".

Traveltime tomography and FWI provide principally different resolutions. As a ray-based technique, the spatial resolution of tomography depends on the diameter of the first local Fresnel zone Fresnel$k \sim \sqrt{\lambda L}$, where L is the distance along the ray path from source to receiver (Williamson 1991; Williamson et al. 1993).

In contrast, theoretical and synthetic studies show that the spatial resolution of FWI with good data coverage is only limited by the size of a local seismic wavelength λ (Schuster 1996; Williamson 1991). Pratt (2003) conducted a synthetic study in order to illustrate the difference between the two (Fig. 1.3). He shows, that the traveltime tomography provides resolution much worse than the size of the dominant wavelength, whereas waveform tomography (FWI) captures the small scale features in the velocity model that are much smaller than a dominant wavelength.

In practice, it means that FWI models are always better resolved than models obtained using conventional methods, such as traveltime tomography.



(a) (b)

Figure 1.2: Comparison of resolution scale length of (a) Travel-time tomography and (b) FWI. Velocity anomaly greater than the seismic wavelength - ray theory works.Velocity anomaly comparable to the seismic wavelength - ray theory fails and scattering theory is better for describing the wave propagation (Adapted from Jones (2010)).

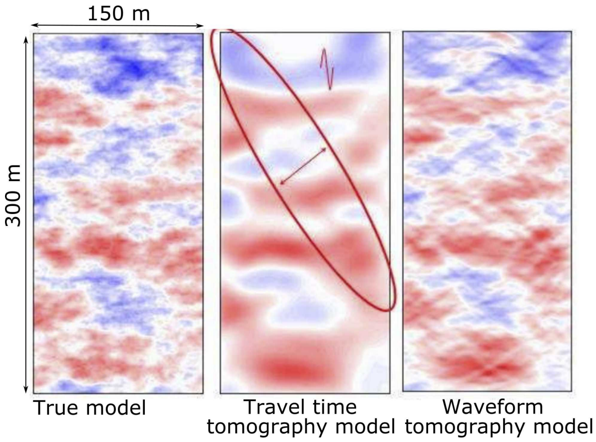


Figure 1.3: Comparison of resolution of traveltime tomography and wavefield tomography (FWI) on a synthetic crosshole example. Left: the true velocity model, middle: the traveltime tomography model and right: the waveform tomography (FWI) model (after Pratt (2003)).

1.3.2 FWI evolution

Full waveform inversion (FWI), also known as wavefield inversion or waveform tomography, is a computational scheme that seeks to minimise the difference between the field and the predicted data in order to produce a high-resolution, high-fidelity model of physical properties of the subsurface (Warner et al. 2013). It attempts to match the field and calculated data wiggle-by-wiggle, trace-by-trace. The core assumption of the method is that if the match between the field and the modelled data is sufficient, then the earth model that was used to generate the modelled data is considered to be the true model of the subsurface. Because FWI is a local optimisation problem, it iterates from a starting guess model that needs to lie within the basin of attraction of the global minimum in order to avoid convergence to a local minimum. In principle, any physical property that has an impact on the wavefield can be recovered by FWI.

First attempts to quantify the physical properties of the subsurface were carried out by using tomography in the early 1980's (Lo et al. 1994). The principle differences between tomography and FWI are introduced in section 1.3.1.

In the mid 1980's, Lailly (1983), Santosa (1984) and Tarantola (1984b)established the theory of FWI by formulating the seismic inverse problem for reflection data in the time domain. A few years later, Gauthier et al. (1986) presented the first computer implementation in 2D on a series of simple synthetic models. Despite the computational limitations of that time, this study has successfully demonstrated the great potential and capacity of the method. First FWI implementation in 2D on real crosshole data were presented by Crase et al. (1990) and Mora (1987, 1988) while Bunks et al. (1995) successfully inverted the Marmousi model using a multi-scale approach.

In 1999, Pratt (1999) formulated the theory of FWI in the frequency domain and Pratt et al. (1999) presented its application on a crosshole field data in 2D. The main advantage of using the frequency domain is reduced computational effort (Vigh et al. 2008), that was crucial considering computational capabilities at that time. It is achieved by limiting the number of frequencies needed for the inversion given wide-angle coverage is provided by the data (Pratt et al. 1990; Sirgue et al. 2004; Yao et al. 2004). In addition, anelastic attenuation and dispersion can be effectively and correctly incorporated in the frequency-domain inversion by introducing complex velocity (Pratt 1999) that allows for the recovery of absorption (Wang 2008). However, the frequency domain demands large memory requirements (Operto et al. 2007), which restricts it to 2D or small-scale 3D problems; and since conventional acquisition configuration have small finite offsets, using a low number of frequencies to recover desired wave numbers is not always possible (Vigh et al. 2008).

In 2008, the theory of FWI was re-formulated in Laplace domain by Shin et al. (2008) who showed its benefits in avoiding local minima.

There have been a number of studies comparing benefits and drawbacks of FWI implementation in the time, frequency and Laplace domains, including those of Umpleby et al. (2010) and Vigh et al. (2008). While the time domain does not take the advantage of wave number redundancy and its cost rises proportionally with the number of shots in the data (Vigh et al. 2008), it is straightforward in its implementation and requires far less memory than the frequency domain if implemented in 3D (Umpleby et al. 2010; Vigh et al. 2008). This makes the time domain a preferable choice for large real-data applications, especially in 3D. Time domain allows pre-processing of the field data to be performed easily prior to the inversion, such as muting and windowing. It also makes quality control fast and easy to carry out at every step of the inversion.

With growing computational power, it became possible to attempt to account for elasticity. In 2002, Shipp et al. (2002) presented a solution for an elastic wave equation in the time domain. A few years later, Gelis et al. (2007) and Sears et al. (2008) demonstrated the first application of elastic time-domain FWI in 2D. However, nowadays elastic FWI proves to be computationally expensive for large 2D problems and normally unaffordable for commercial 3D field datasets.

Ben-Hadj-Ali et al. (2008), Sirgue et al. (2008), Stekl et al. (2008), and Warner et al. (2007, 2008) presented the first practical applications of acoustic FWI in 3D followed by other successful applications of FWI to 3D field datasets (Plessix et al. 2010; Sirgue et al. 2010).

In 2009, Sirgue et al. (2009) demonstrated a successful application of FWI on Valhall oil field located in the North sea. This study showed the great advantages that FWI can provide for geological interpretation and escalated the interest in both the scientific community and industry.

In 2016 and 2017, Lu et al. (2016) and Ruth et al. (2017) presented a study of acoustic time-domain FWI on NATS field data and showed high-resolution FWI impedance models that were obtained using frequencies much higher than typical - 40 Hz. As part of this thesis, the question of how high in terms of frequencies it is possible and feasible to go with conventional FWI is addressed in order to fill this gap.

Over the last decade, acoustic time-domain FWI, that typically utilises frequencies below 10 HZ, became a relatively routine component of velocity model building workflow for marine field datasets in both 2D and 3D.

It usually works well on marine seismic data as problems like cycle skipping and operator error in most cases can be identified and dealt with. But FWI can produce an incorrect model or fail completely for less obvious reasons as shown by Warner (2018). As part of this research, the difficulties of applying FWI to NATS datasets will be addressed.

Chapter 2

Theoretical background, implementation and data requirements

Geophysical methods study different physical fields of the Earth, such as magnetic, gravity, electromagnetic and seismic wavefields. The numerical values of the recorded fields depend, first, on the physical properties of the subsurface as these fields are being propagated through the Earth's interior. The ultimate goal of all geophysical methods is to determine the geological properties from the collected data.

The common approach to tackle this problem consists of constructing different models, predicting the data for each model and finding a model that provides the closest match to the data collected in the experiment. Calculating geophysical data d for a given model m is known as a forward problem

$$d = A(m) \tag{2.1}$$

where A is a forward problem operator that represents a linear or non-linear relationship between the data and the model. Assuming that this relationship A describes the physical laws in a sufficient manner, it can be considered that when the calculated data matches the observed data, the theoretical model may match the real distribution of rock properties fairly well.

Considering the forward problem (Eq. 2.1), the model parameters can be determined from the recorded data by solving the following expression:

$$m = A^{-1}(d) \tag{2.2}$$

where A^{-1} is the inverse problem operator. Equation 2.2 defines the inverse problem of geophysics. A schematic representation of the forward and the inverse problems is shown in Figure 2.1.

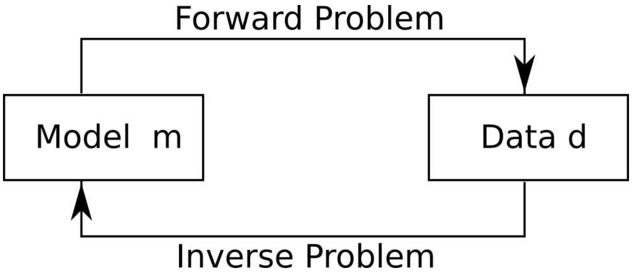


Figure 2.1: The traditional definition of the forward and inverse problems.

For every inverse problem, there are three important questions to answer:

1. Does the solution exist?
2. Is the solution unique?
3. Is the solution stable?

The existence of the solution is related to the mathematical formulation of the inverse problem. From the physical point of view, the solution exists because the geological structure exists within the Earth's subsurface. However, mathematically, there might not be a model that fits the observed data since this always contains noise.

The uniqueness of the solution means there should not be two or more different geological models that generate the same observed data. For example, if there are two different sets of model parameters m_1 and m_2, two different sources s_1 and s_2 and the solution is not unique, then both models produce the same data d

$$\begin{aligned} A(m_1, s_1) &= d \\ A(m_2, s_2) &= d \end{aligned} \tag{2.3}$$

In this case, given the observed data d it is impossible to identify whether it was generated by the model m_1 or m_2.

Finally, another important issue within the inverse problem theory is the stability of the solution m. The stability means that a small perturbation within the observed data corresponds to a small variation in the model parameters. In real world, all geophysical data contain some level of noise δd. It is possible that the difference between the two observed datasets d_1 and d_2 generated using two significantly different geological models m_1 and m_2 and two significantly different sources s_1 and s_2

$$\begin{aligned} A(m_1, s_1) &= d_1 \\ A(m_2, s_2) &= d_2 \end{aligned} \tag{2.4}$$

lies within the noise level δd

$$||d_1 - d_2|| \leq ||\delta d|| \tag{2.5}$$

where $||...||$ measures the magnitude of the difference between the two datasets. In this case, a small change in the noise level results in a large change in the model, therefore, the solution is unstable.

If the solution of the inverse problem does not satisfy all three criteria, i.e. it does not exist, or it is not unique, or it is not stable, the inverse problem is considered to be “ill-posed”. The majority of geophysical problems, and natural science problems in general, are “ill-posed”. A French mathematician Jacques Hadamard (1902) suggested that every “ill-posed” problem does not carry physical or/and mathematical meaning and cannot be solved. Fortunately, his opinion was found to be wrong: “ill-posed” problems are physically and mathematically meaningful and can be solved. In the second half of 20th Century a Russian mathematician Andrey N. Tikhonov established the foundation of the theory of the “ill-posed” problem solutions (Tikhonov et al. 1977). This is beyond the scope of my thesis, but the mathematically rigorous methods of the solutions of the inverse problems in geophysics are fully described by Zhdanov (2002).

Inverse problems still remain challenging due to the complexity of the Earth’s interior and current computational constrains. However, the success of the geological interpretation relies on the ability to provide reasonable geological models by adequately solving inverse problems.

FWI is one of the geophysical inverse problems. In this chapter, I will, first, present the underlying mathematical principles of FWI for its simplest implementation, second, give an overview of the numerical scheme *fullwave3D* developed over the years by my research group and used in this thesis (apart from Chapter 7) and, finally, discuss the data requirements and give some practical recommendations for successful application of FWI to real data.

2.1 Theory

2.1.1 The wave equation

The one-dimensional wave equation was derived in 1746 by a French mathematician Jean-Baptiste le Rond d'Alembert as a first-order nonlinear ordinary differential equation, which is known in mathematics as d'Alembert's equation. Eleven years later, in 1757, the three-dimensional wave equation was derived by a Swiss mathematician Leonhard Euler who described the fluid movement in a flow where the viscosity of the fluid is equal to zero. Mathematically, it is a second-order partial differential equation that contains derivatives in time and space. The wave equation is a very important equation in physics as the behaviour of both mechanical and electromagnetic waves honour it. This makes it a fundamental equation in such disciplines as acoustics, electromagnetics and fluid dynamics. In seismic imaging only mechanical waves are involved.

The wave equation can vary depending on the assumptions on physics, for instance, an acoustic or elastic, anisotropic or isotropic medium being considered, with or without density impact on the propagation. Most commonly, the Earth is treated as an acoustic, isotropic, inhomogeneous, non-attenuating, non-dispersive, stationary medium, and in this case the acoustic pressure waves honour the simplest form of the wave equation, presented here in the time domain

$$\left[\frac{1}{V_p^2(\mathbf{r})}\frac{\partial^2}{\partial t^2} - \rho(\mathbf{r})\nabla\cdot\left(\frac{1}{\rho(\mathbf{r})}\nabla\right)\right]p(\mathbf{r},t) = s(\mathbf{r},t) \tag{2.6}$$

where V_p is the acoustic velocity, ρ is the density of the medium, p is the pressure wavefield and s is the source. The time is denoted by t and the spatial coordinates by $\mathbf{r}$. The properties of the medium vary in space while the source varies both in space and time.

2.1.2 The forward problem

Predicting the seismic wavefield for a given model of the subsurface properties is a forward problem. It can be seen that the wave equation (Eq. 2.6) establishes a linear relationship between the propagated acoustic wavefield and the source through the linear operator A and can be written in a matrix form as

$$\mathbf{Ap} = \mathbf{s} \tag{2.7}$$

where $\mathbf{p}$ is a vector of acoustic wavefield calculated at every position $\mathbf{r}$ for every time value t and $\mathbf{s}$ is a vectorised source function that varies both in time and space. For the simplicity, I set up a vector of model parameters

$$\mathbf{m} = \begin{bmatrix} V_p \\ \rho \end{bmatrix} \tag{2.8}$$

and formulate the forward problem for the wave equation in a following way:

$$\mathbf{p} = G(\mathbf{m}) \tag{2.9}$$

where $\mathbf{m}$ is a vector of model parameters (Eq. 2.8) and G is an operator that, as can be easily seen from the wave equation (Eq. 2.6), establishes a non-linear relationship between the model and the calculated wavefield $\mathbf{p}$ and, therefore, cannot be presented as a matrix. Note, that operator G depends on the source $\mathbf{s}$.

The solution of the forward problem is obtained by solving the wave equation. For the simplest acoustic wave equation with neglected density the solution is presented by Tarantola (1984).

2.1.3 The inverse problem: FWI

FWI is an inverse problem of seismic exploration and seismology. A formal description of the inverse problem in context of geophysical observations was given by Backus et al. (1968) and Backus (1970).

Recalling the forward problem (Eq. 2.9), the inverse problem can be formulated as

$$\mathbf{m} = G^{-1}(\mathbf{p}) \tag{2.10}$$

where G^{-1} is the inverse of forward operator. Hence the forward operator is a non-linear function, the inverse G^{-1} cannot be expressed in a matrix form and the solution cannot be found explicitly. Therefore, it must be approximated using different optimisation methods, for example gradient-type methods that belong to a family of the most powerful optimisation techniques (Zhdanov 2002).

Here I present the most common approach based on gradient-type methods, focusing, in particular, on the less computationally expensive steepest descent, known as well as the gradient method. I largely adopt the compact matrix formalism used by Pratt et al. (1998). Although this paper specifically refers to the frequency domain, I perform the derivations in the time domain. I also follow the general approach of Virieux et al. (2009).

Least-squares formulation

Gradient-type methods attempt to find a model that minimises the difference between the predicted and the observed data. The better this model approximates the reality, the smaller is the difference between the calculated and observed data. This difference $\Delta\mathbf{d}$ can be defined as follows

$$\Delta\mathbf{d} = \mathbf{d}_{calc}(\mathbf{m}) - \mathbf{d}_{obs} \tag{2.11}$$

where $\mathbf{d}_{obs}$ is the collected at receivers seismic data and $\mathbf{d}_{calc}$ is the predicted pressure wavefield at receiver locations only, so that $\mathbf{d}_{calc}$ is extracted from the full predicted wavefield $\mathbf{p}$ using a projection matrix $\mathbf{R}$:

$$\mathbf{d}_{calc} = \mathbf{R}(\mathbf{p}) \tag{2.12}$$

Mathematically, the measure of the difference between the two datasets $\Delta\mathbf{d}$ can be expressed using many different norms, for example least-absolute-values norm L_1, least-squares norm L_2 and various mixed $L_1 - L_2$ norms (Virieux et al. 2009). Least-absolute-values norm L_1 has some advantages for working with seismic data (Claerbout et al. 1973), however, the least-squares norm L_2 is easier to manipulate mathematically (Tarantola 1984) and is the most often used norm for FWI. It assumes a Gaussian distribution of the data errors and is robust when there is no large-amplitude outliers in the data (Tarantola 2005). Thus, the difference $\Delta\mathbf{d}$ in least-squares formulation is given by

$$\phi(\mathbf{m}) = \frac{1}{2}||\mathbf{d}_{calc} - \mathbf{d}_{obs}||_2^2 = \frac{1}{2}(\mathbf{d}_{calc} - \mathbf{d}_{obs})^T(\mathbf{d}_{calc} - \mathbf{d}_{obs}) = \frac{1}{2}\Delta\mathbf{d}^T\Delta\mathbf{d} \tag{2.13}$$

It minimises the sum of the difference between the two datasets over all sources, all receivers and all time. It is a function of model parameters $\mathbf{m}$ and it is a positive scalar. It can be referred to in literature as the "misfit function", the "cost function", the "objective function", or just the "functional". In order to minimise the misfit function $\phi(\mathbf{m})$, its partial derivatives with respect to the model parameters need to be set to zero and solved for $\mathbf{m}$. Differentiating the misfit function with respect to a model parameter m_i results in

$$\begin{aligned} \frac{\partial\phi(\mathbf{m})}{\partial m_i} &= \frac{1}{2}\frac{\partial}{\partial m_i}\sum_j (d_{calc}(\mathbf{m})_j - d_{obs_j})^2 \\ &= \frac{1}{2}\sum_j 2\left(d_{calc}(\mathbf{m})_j - d_{obs_j}\right)\left(\frac{\partial}{\partial m_i}(d_{calc}(\mathbf{m})_j - d_{obs_j})\right) \\ &= \sum_j \frac{\partial d_{calc}(\mathbf{m})_j}{\partial m_i}(d_{calc}(\mathbf{m})_j - d_{obs_j}) \end{aligned} \tag{2.14}$$

where j denotes receivers. In this derivation, I am using the time domain formulation of the wave equation, that is why the values of the misfit function derivative are always real numbers. However, in the frequency domain formulation, the derivative of the misfit function contains complex numbers and mathematics demands only the real part of a complex number to be taken (Pratt et al. 1998). Re-writing equation 2.14 in a matrix form gives the expression:

$$\frac{\partial\phi(\mathbf{m})}{\partial m_i} = \left(\frac{\partial\mathbf{d}_{calc}(\mathbf{m})}{\partial m_i}\right)^T(\mathbf{d}_{calc}(\mathbf{m}) - \mathbf{d}_{obs}) \tag{2.15}$$

Therefore, the derivative of the misfit function with respect to all the model parameters can be written as follows:

$$\begin{aligned}\frac{\partial \phi(\mathbf{m})}{\partial \mathbf{m}} &= \begin{bmatrix} (\frac{\partial \mathbf{d}_{calc}(\mathbf{m})}{\partial m_1})^T \\ ... \\ (\frac{\partial \mathbf{d}_{calc}(\mathbf{m})}{\partial m_N})^T \end{bmatrix} (\mathbf{d}_{calc}(\mathbf{m}) - \mathbf{d}_{obs}) \\ &= \begin{bmatrix} (\frac{\partial \mathbf{d}_{calc}(\mathbf{m})}{\partial m_1}) & ... & (\frac{\partial \mathbf{d}_{calc}(\mathbf{m})}{\partial m_N}) \end{bmatrix}^T (\mathbf{d}_{calc}(\mathbf{m}) - \mathbf{d}_{obs})\end{aligned} \tag{2.16}$$

where N is a number of model parameters. By the definition of vector calculus, a matrix of all first-order partial derivatives of a vector-valued function is called the Jacobian matrix $\mathbf{J}$, also referred to as the Fréchet derivative matrix or the sensitivity matrix:

$$\mathbf{J} := \begin{bmatrix} (\frac{\partial \mathbf{d}_{calc}(\mathbf{m})}{\partial m_1}) & ... & (\frac{\partial \mathbf{d}_{calc}(\mathbf{m})}{\partial m_N}) \end{bmatrix} \tag{2.17}$$

Using the Jacobian matrix, equation 2.16 can be written compactly as

$$\frac{\partial \phi(\mathbf{m})}{\partial \mathbf{m}} = \mathbf{J}^T (\mathbf{d}_{calc}(\mathbf{m}) - \mathbf{d}_{obs}) = \mathbf{J}^T \Delta \mathbf{d} \tag{2.18}$$

Finally, the derivative of the misfit function with respect to the model parameters is obtained (Eq. 2.18).

Gradient-type methods

The core idea of gradient-type methods is to start with some known model $\mathbf{m}_0$ that is believed to be sufficiently close to the true model and make a series of updates to the model $\mathbf{m}_0$ that reduces the misfit function towards zero. Thus, FWI is a local iterative scheme.

The Born approximation states that a change in the model is linearly related to a change in the wavefield produced by this perturbation in the model. In this framework, the updated model $\mathbf{m}$ can be expressed as follows (Virieux et al. 2009)

$$\mathbf{m} = \mathbf{m}_0 + \Delta \mathbf{m} \tag{2.19}$$

where $\mathbf{m}_0$ is the starting model. Approximating the misfit function at the starting point $\mathbf{m}_0$ using the Taylor expansion results in

$$\phi(\mathbf{m}_0 + \Delta\mathbf{m}) = \phi(\mathbf{m}_0) + \Delta\mathbf{m}^T \frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}} + \frac{1}{2}\Delta\mathbf{m}^T \frac{\partial^2\phi(\mathbf{m}_0)}{\partial^2\mathbf{m}}\Delta\mathbf{m} + \mathcal{O}(\Delta^3\mathbf{m}) \tag{2.20}$$

Now, differentiating the approximation of the misfit function (Eq. 2.20) with respect to the model parameters $\mathbf{m}$ (instead of differentiating the misfit function itself), truncating the result to the second order term and setting it to zero to find the minimum value for ϕ gives

$$\begin{split}\frac{\partial\phi(\mathbf{m}_0 + \Delta\mathbf{m})}{\partial\mathbf{m}} = \frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}} + \frac{\partial\Delta\mathbf{m}^T}{\partial\mathbf{m}}\frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}} + \Delta\mathbf{m}^T\frac{\partial}{\partial\mathbf{m}}\left(\frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}}\right) + \\ \frac{\partial^2\phi(\mathbf{m}_0)}{\partial^2\mathbf{m}}\Delta\mathbf{m} + \frac{1}{2}\Delta\mathbf{m}^T\frac{\partial}{\partial\mathbf{m}}\left(\frac{\partial^2\phi(\mathbf{m}_0)}{\partial^2\mathbf{m}}\right)\Delta\mathbf{m} + \mathcal{O}(\Delta^2\mathbf{m}) = 0\end{split} \tag{2.21}$$

Taking into account that $\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}$, where $\mathbf{m}_0$ is a fixed starting velocity model, it can be written that

$$\frac{\partial}{\partial\mathbf{m}} = \frac{\partial}{\partial(\mathbf{m}_0 + \Delta\mathbf{m})} = \frac{\partial}{\partial\Delta\mathbf{m}} \tag{2.22}$$

Recognising that

$$\frac{\partial\Delta\mathbf{m}^T}{\partial\mathbf{m}} = \frac{\partial\Delta\mathbf{m}^T}{\partial\Delta\mathbf{m}} = 1 \tag{2.23}$$

and considering Equation 2.22, it can be seen that the first, the third and the fifth terms in equation 2.21 equal to zero because the value of the misfit function $\phi(\Delta\mathbf{m}_0)$ does not depend upon $\Delta\mathbf{m}$. Equation 2.21 can be re-written as

$$0 + \frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}} + 0 + \frac{\partial^2\phi(\mathbf{m}_0)}{\partial\mathbf{m}^2}\Delta\mathbf{m} + 0 + \mathcal{O}(\Delta^2\mathbf{m}) = 0 \tag{2.24}$$

Re-arranging the terms of equation 2.24 results in

$$\Delta\mathbf{m} = -\left(\frac{\partial^2\phi(\mathbf{m}_0)}{\partial\mathbf{m}^2}\right)^{-1}\left(\frac{\partial\phi(\mathbf{m}_0)}{\partial\mathbf{m}}\right) \tag{2.25}$$

$$\Delta\mathbf{m} = -\mathbf{H}^{-1}\nabla_{\mathbf{m}}\phi \tag{2.26}$$

where $\Delta\mathbf{m}$ is a model update and $\mathbf{H}$ is the Hessian matrix. Both the Hessian and the gradient are evaluated at $\mathbf{m}_0$. Therefore, to calculate the model update $\Delta\mathbf{m}$ given by the equation 2.26, the Hessian and the gradient must be calculated.

Gradient

Re-calling equation 2.18, in order to compute the gradient of the objective function with respect to the model parameters, the Jacobian matrix (or Fréchet derivative matrix, or sensitivity matrix) must be calculated at every iteration. The Jacobian matrix contains N elements, where N is the number of model parameters, which is typically a few million elements (Warner et al. 2013). Thus, at ever iteration computation of the gradient requires $N+1$ calculation of the very large Jacobian matrix; this is not computationally feasible. An alternative way to obtain the gradient without directly calculating the Fréchet derivatives should be used, for example the adjoint method introduced by Tarantola (1984).

Re-calling again the wave equation in its matrix formulation (Eq. 2.7) and differentiating it with respect to the model parameters $\mathbf{m}$ results in

$$\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} + \mathbf{A}\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = 0 \tag{2.27}$$

The right-hand side equals zero hence the source $\mathbf{s}$ does not depend upon the model parameters. Re-arranging the terms in equation 2.27 and multiplying both sides by the inverse of the matrix $\mathbf{A}$ gives

$$\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} \tag{2.28}$$

Applying to equation 2.28 so-called projection matrix $\mathbf{R}$ that contains ones at receiver locations and zeroes everywhere else to compute

$$\mathbf{R}\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} \tag{2.29}$$

and considering that $\mathbf{d}_{calc}$ is the calculated wavefield $\mathbf{p}$ at receiver locations, re-write equation 2.29 as follows

$$\frac{\partial \mathbf{d}_{calc}}{\partial \mathbf{m}} = -\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} \tag{2.30}$$

Comparing equation 2.30 to the definition of the Jacobian matrix (Eq. 2.17), it can be noticed that a new way of calculating the Jacobian is obtained

$$\mathbf{J} = -\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} \tag{2.31}$$

Substitution of equation 2.31 into equation for the gradient calculation (Eq. 2.18) gives the new equation to compute the gradient

$$\nabla_{\mathbf{m}}\phi = -\left(\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p}\right)^T \Delta\mathbf{d} \tag{2.32}$$

Re-arranging the terms in the equation above results in

$$\nabla_{\mathbf{m}}\phi = -\mathbf{p}^T\left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\right)^T (\mathbf{A}^{-1})^T\mathbf{R}^T\Delta\mathbf{d} = -\mathbf{p}^T\left(\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\right)^T \mathbf{r}_b \tag{2.33}$$

where $\mathbf{r}_b$ is back propagated residual wavefield. As can be seen, the gradient computation involves calculation of the forward wavefield $\mathbf{p}$, differentiation of each coefficient of the matrix $\mathbf{A}$ with respect to the model parameters and, finally, back propagation of the residuals $\Delta\mathbf{d}$ using the inverse of the matrix $\mathbf{A}^T$ (the matrix $\mathbf{A}^T$ is called the adjoint matrix to $\mathbf{A}$), zero-lag cross-correlation these two real-valued time series, i.e. computed wavefields, in the time domain, and, finally, all the individual gradients for every source must be stacked together to obtain the global gradient for equation 2.26 (Warner et al. 2013).

In practice, the numerical operator $\mathbf{A}$ is designed to be symmetric in space and anti-symmetric in time, so that $\mathbf{A}^{-T}$ is a back propagator in time. That allows use of the same numerical code to propagate both wavefields. For one source, calculating the gradient using equation 2.33 requires two forward modelling runs compared to $N + 1$ modelling runs required using the equation 2.18 that involves direct calculation of the Fréchet derivatives.

A clear physical interpretation of the gradient is well described by Pratt et al. (1998). To summarise it here, the gradient can be seen to represent an image of the original diffracting point, with the location of diffracting heterogeneity receiving the maximum gradient contribution (Pratt et al. 1998; Tarantola 1984; Virieux et al. 2009).

Hessian

Now, the gradient is obtained and in order to compute the model update (Eq. 2.26), the second term $\mathbf{H}^{-1}$, or the Hessian matrix, must be calculated. By definition, the Hessian matrix is a square matrix of second-order partial derivatives of a scalar-valued function

$$\mathbf{H} := \frac{\partial^2 \phi(\mathbf{m})}{\partial \mathbf{m}^2} \tag{2.34}$$

Differentiating the expression for the gradient (Eq. 2.14), a single component of the Hessian is given by

$$\begin{aligned} H_{ij} &= \frac{\partial}{\partial m_i} \sum_k \frac{\partial d_{calc}(\mathbf{m})_k}{\partial m_j} (d_{calc}(\mathbf{m})_k - d_{obs_k}) \\ &= \sum_k (\frac{\partial^2 d_{calc}(\mathbf{m})_k}{\partial m_i \partial m_j} (d_{calc}(\mathbf{m})_k - d_{obs_k}) + \frac{\partial d_{calc}(\mathbf{m})_k}{\partial m_i} \frac{\partial d_{calc}(\mathbf{m})_k}{\partial m_j}) \\ &= \sum_k (\frac{\partial d_{calc}(\mathbf{m})_k}{\partial m_i} \frac{\partial d_{calc}(\mathbf{m})_k}{\partial m_j} + \frac{\partial^2 d_{calc}(\mathbf{m})_k}{\partial m_i \partial m_j} (d_{calc}(\mathbf{m})_k) - d_{obs_k}) \end{aligned} \tag{2.35}$$

Re-calling the definition of the Jacobian (Eq. 2.17) and re-writing equation 2.35 in a matrix form gives the expression for the Hessian, evaluated at $\mathbf{m}_0$

$$\mathbf{H}_{m_0} = \mathbf{J}_{m_0}^T \mathbf{J}_{m_0} + \frac{\partial \mathbf{J}_{m_0}^T}{\partial \mathbf{m}^T} (\Delta \mathbf{d}_0 ... \Delta \mathbf{d}_0) \tag{2.36}$$

where both the Jacobian $\mathbf{J}_{m_0}$ and the data residuals $\Delta \mathbf{d}_0$ are evaluated at $\mathbf{m}_0$.

If the model has N parameters then the Hessian will have N^2 elements. Typically, the Hessian is a very large matrix and can have a few trillion values (Warner et al. 2013). While the computation of the matrix $\mathbf{H}$ itself might still be within current computational power, the direct calculation of the inverse $\mathbf{H}^{-1}$ is computationally not feasible. Fortunately, the gradient-type methods allow the Hessian matrix to be computed in different ways.

The direct computation of the Hessian is called Newton method and can be found in a paper by Pratt et al. (1998). This method attempts to solve the minimisation problem in one step (in case of a linear problem) and is rarely used. The Hessian can be approximated in different ways, thus Gauss-Newton and quasi-Newton method appear. Finally, replacing the Hessian with a scalar is called the steepest descent method. For non-linear problems all gradient-type methods, including the quasi-Newton and the steepest descent, are local iterative methods, therefore, FWI is a local iterative scheme.

Steepest descent

For simplicity, I will start with the steepest descent method, also referred to as the gradient method. It is one of the most powerful gradient-type methods and is widely used in a variety of inverse problems (Zhdanov 2002). This method approximates the Hessian with the constant number α called the step length. Therefore, the model update $\Delta\mathbf{m}$ in equation 2.26 can be re-written as

$$\Delta\mathbf{m} = -\alpha\nabla_{\mathbf{m}}\phi \tag{2.37}$$

where the gradient identifies the direction for the update and the step length indicates the amplitude of the movement in the opposite direction to the gradient.

The idea of the steepest descent is illustrated in Figure 2.2. For example, the model contains parameters m_1 and m_2 and function $\Phi(m_1, m_2)$ is the misfit which represents the difference between the calculated and observed data. As there are only two model parameters, the misfit Φ is geometrically defined by a surface. The aim of the gradient descent method is to minimise the misfit between the observed and real data. Ideally, it is seeking the global minimum of the surface. In order to find a minimum of a surface, one has to have an initial location on the surface to start with (starting model), a direction to move (the gradient) and a step to take in this direction (the step length) until the minimum is reached. The starting point and the step length are vital components in order for the scheme to succeed. The importance of the starting point is quite obvious - consider two starting points that are not close to each other to have the same gradient direction. Each point will lead to a different minimum (Fig. 2.2), one of which is a local miminum and is not giving the best solution to the problem. It does minimise the cost function compared to its initial value, however, in real seismic problems, it still may generate a velocity model, that will be worse than the starting model.

Do we actually want the global minimum? Under assumption that our measurements of observed data are absolutely correct and forward modelling takes into account the full physics of the seismic wave propagation, then the answer is yes. However, in reality there are bias in both acquisition and modelling and restrictions on physics of wave propagation. In practice, the best fit to the observed data given by the global minimum does not necessarily mean the best geological solution, so the misfit should not be exactly zero but less than some reasonably small positive value ϵ that takes the errors into account.

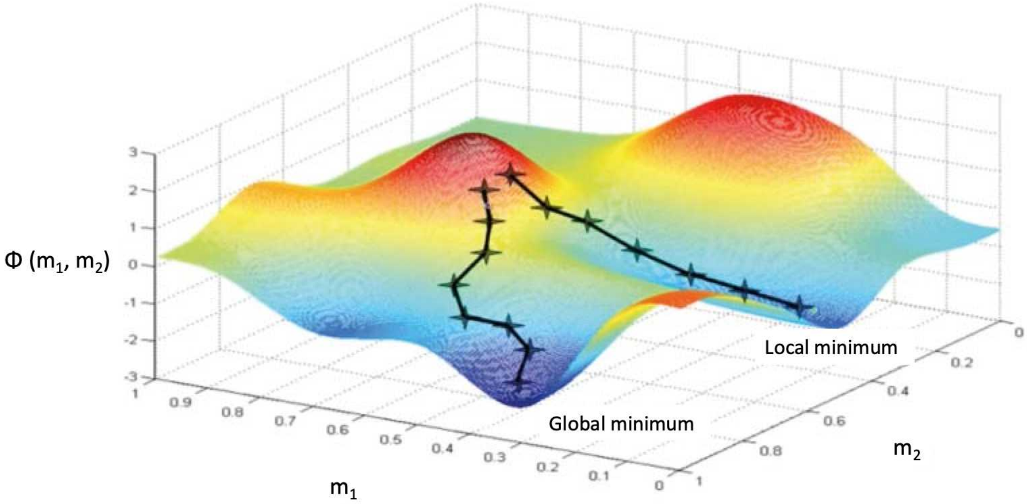


Figure 2.2: Schematic illustration of the gradient descent method. Φ is the misfit function of the two model parameters m_1 and m_2. The background figure is from `https://www.coursera.org/learn/machine-learning/lecture/8SpIM/gradient-descent`

The importance of the step is demonstrated in Figure 2.3. If the step length is too small, the gradient descent method can be slow, which is unacceptable for commercial purposes. If the step is too large, the gradient descent can overshoot minimum, it may fail to converge, or even diverge.

In order to find the model update (Eq. 2.37), the step length α needs to be calculated. This leads to another optimisation problem that seeks for such a constant α

$$\mathbf{m}_\alpha = \mathbf{m}_0 + \alpha\delta\mathbf{m} \tag{2.38}$$

where $\mathbf{m}_0$ is the starting velocity model and $\delta\mathbf{m}$ is a small step taken in the opposite direction

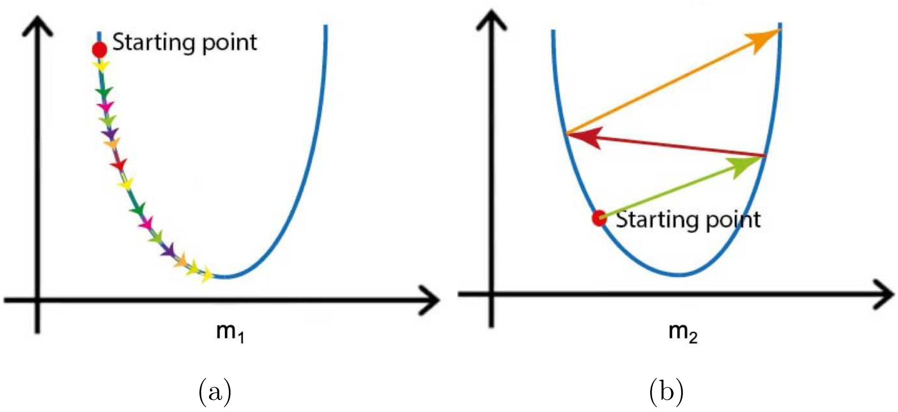


Figure 2.3: Optimal step length in the gradient descent method: (a) if the step is too small, gradient descent can be slow, which is unacceptable for commercial purposes; (b) if the step is too large, gradient descent can overshoot minimum and the scheme may fail to converge, or even diverge. The background figure is from `https://www.coursera.org/learn/machine-learning/lecture/8SpIM/gradient-descent`

to the gradient, that minimises the residuals

$$\Delta \mathbf{d}_\alpha = \mathbf{d}_\alpha^{calc} - \mathbf{d}^{obs} \tag{2.39}$$

where $\mathbf{d}_\alpha^{calc}$ is wavefield associated with the model $\mathbf{m}_\alpha$ and $\mathbf{d}^{obs}$ is the observed data. In least-squares sense, this can be formulated as minimisation of the following misfit function

$$\phi(\mathbf{m}_\alpha) = \frac{1}{2}||\Delta \mathbf{d}_\alpha||^2 \tag{2.40}$$

To solve the problem, the linear search algorithm is applied:

1. Compute the wavefield $\mathbf{d}_0^{calc}$ associated with starting velocity model $\mathbf{m}_0$
2. Compute new model $\mathbf{m}_1 = \mathbf{m}_0 + \delta\mathbf{m}$ where $\delta\mathbf{m}$ is a small step taken in the opposite direction to the gradient
3. Compute the wavefield $\mathbf{d}_1^{calc}$ corresponding to the new model $\mathbf{m}_1$
4. Compute the corresponding residuals $\Delta\mathbf{d}_0 = \mathbf{d}_0^{calc} - \mathbf{d}^{obs}$ and $\Delta\mathbf{d}_1 = \mathbf{d}_1^{calc} - \mathbf{d}^{obs}$
5. Under assumption of Born approximation that implies that a change in the model leads to a linear change in the residuals, it can be considered that the best fit model $\mathbf{m}_\alpha$ lies in

between $\mathbf{m}_0$ and $\mathbf{m}_1$:

$$\Delta\mathbf{d}_{\alpha}^{calc} = \Delta\mathbf{d}_0^{calc} + \alpha(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc}) \tag{2.41}$$

6. Substitution of equation 2.41 into the misfit function 2.40 results in

$$\begin{aligned}\phi(\mathbf{m}_\alpha) &= \frac{1}{2}(\Delta\mathbf{d}_0^{calc} + \alpha(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc}))^T(\Delta\mathbf{d}_0^{cal} + \alpha(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc}))\\ &= \frac{1}{2}(\Delta\mathbf{d}_0^{calc})^T\Delta\mathbf{d}_0^{calc} - \alpha(\Delta\mathbf{d}_0^{calc})^T(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc})\\ &+ \frac{\alpha^2}{2}(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc})^T(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc})\end{aligned} \tag{2.42}$$

Differentiating equation 2.42 and setting the result equals zero gives

$$\frac{\partial\phi}{\partial\alpha}(m_\alpha) = -\Delta d_0^T(\Delta d_1 - \Delta d_0) + \alpha(\Delta d_1 - \Delta d_0)^T(\Delta d_1 - \Delta d_0) = 0 \tag{2.43}$$

Re-arranging the terms gives the expression for the step length for linear forward problem, also given by Pratt et al. (1998)(eq.12)

$$\alpha = \frac{(\Delta\mathbf{d}_0^{calc})^T(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc})}{(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc})^T(\Delta\mathbf{d}_1^{calc} - \Delta\mathbf{d}_0^{calc})} \tag{2.44}$$

7. Once the step length α is found, the original model $\mathbf{m}_0$ is replaced with the new model $\mathbf{m}_\alpha$ and the process repeats until the steepest descent algorithm converges.

The initial non-linear inverse FWI problem was linearised using the Taylor expansion and Born approximation and is solved correctly at every iteration. Therefore, the full non-linear inverse problem can be solved correctly using multiple iterations.

Pseudo-Hessian matrix, or steepest descent with spatial preconditioning

In the steepest descent method, the model update Δm in equation 2.19 is given by the gradient multiplied by a constant instead of the Hessian matrix, which is a brutal approximation. Initial equation for the Hessian matrix (Eq. 2.36) consists of two terms, where the second term can be neglected as it tends towards zero when $||\Delta\mathbf{d}_0||$ approaches the minimum. Hence, the Hessian can be approximated as follows

$$\mathbf{H}_{m_0} \approx \mathbf{J}_{m_0}^T \mathbf{J}_{m_0} \tag{2.45}$$

and used for the Gauss-Newton method. Pratt et al. (1998) demonstrated that Gauss-Newton method considerably sharpens the gradient image compared to the less expensive steepest descent for a fixed number of iterations. However, the Gauss-Newton method requires to approximate the Hessian matrix (as shown in Eq. 2.45) and find its inverse that can be very large to compute. Shin et al. (2001) demonstrated a method of Hessian approximation that does not require extra computational effort on an example of a 2D case. This method is called pseudo-Hessian and is adopted by *fullwave3D*. I give an overview here.

Substitution of expression for the Jacobian (Eq. 2.31) into equation 2.45

$$\begin{aligned} \mathbf{H}_{m_0} = \mathbf{H} &\approx (\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p})^T(\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p}) \\ &= \mathbf{p}^T\frac{\partial \mathbf{A}^T}{\partial \mathbf{m}}\mathbf{A}^{-T}\mathbf{R}^T\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial \mathbf{m}}\mathbf{p} \end{aligned} \tag{2.46}$$

and neglecting the off-diagonal elements results in

$$H_{ii} \approx \mathbf{p}^T\frac{\partial \mathbf{A}^T}{\partial m_i}\mathbf{A}^{-T}\mathbf{R}^T\mathbf{R}\mathbf{A}^{-1}\frac{\partial \mathbf{A}}{\partial m_i}\mathbf{p} \tag{2.47}$$

where the term $\mathbf{A}^{-T}\mathbf{R}^T\mathbf{R}\mathbf{A}^{-1}$ is too demanding to compute as typically there is a high number of receivers. Instead, it is approximated by a scalar value β that is taken care of during the step length calculation in the way described earlier

$$H_{ii} \approx \beta\mathbf{p}^T\frac{\partial \mathbf{A}^T}{\partial m_i}\frac{\partial \mathbf{A}}{\partial m_i}\mathbf{p} \tag{2.48}$$

Replacing the middle term with the scalar assumes similar distribution of sources and receivers over the area. Therefore, this method works fairly well in cases of most surface seismic datasets and, probably, is not the best solution in cases where sources and receivers occupy different areas (for instance, VSP datasets).

Defining a new symbol

$$\mathbf{e}_i := \frac{\partial \mathbf{A}}{\partial m_k}\mathbf{p} \tag{2.49}$$

equation 2.48 for the i^{th} component of the pseudo-Hessian can be written compactly as

$$H_{ii} \approx \beta \mathbf{e}_i^T \mathbf{e}_i \tag{2.50}$$

as well as the equation 2.33 for the i^{th} component of gradient can be re-written as

$$\nabla_{\mathbf{m}_i} \phi = -\mathbf{p}^T (\frac{\partial \mathbf{A}}{\partial m_i})^T \mathbf{r}_b = -\mathbf{e}_i^T \mathbf{r}_b \tag{2.51}$$

Using the pseudo-Hessian matrix and the gradient, the model update (Eq. 2.25) gives the expression

$$\Delta \mathbf{m}_i \approx -\frac{\mathbf{e}_i^T \mathbf{r}_b}{\beta \mathbf{e}_i^T \mathbf{e}_i} \tag{2.52}$$

where the term $\mathbf{e}_i^T \mathbf{e}_i$ can be seen as a preconditioning to the gradient that is defined by a simple scalar, and depends on the location in the model and the constant β is the reciprocal of the step length.

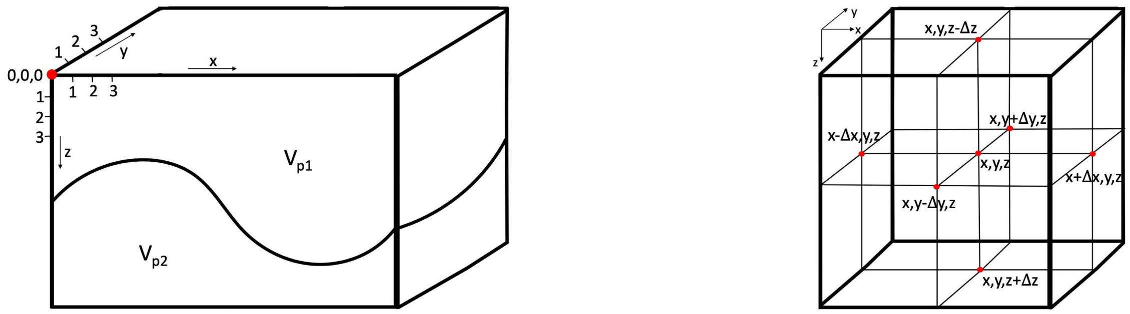
In this section, the core mathematical theory of FWI in its least-squares formulation is presented with the solution choices adopted by the numerical scheme *fullwave3D* used in the major part of this thesis.

2.2 Implementation

Efficiently solving the numerical wave equation of the form 2.9 is the key component of any FWI code. As any numerical partial differential equation (PDE), the wave equation can be solved using one of the various numerical methods, for example finite difference, finite elements, spectral, pseudo-spectral and others. In science, the most popular and commonly used are finite difference (FD) methods as they are the simplest and the fastest (Guash 2011). They replace differential equation by difference equations and, as any other numerical methods, require that PDE becomes discretised on a grid. I will give a short overview of finite difference method as *fullwave3D*, used throughout this thesis, utilises finite-difference scheme.

2.2.1 Finite difference scheme

In finite difference modelling, the medium is divided into a large number of elementary blocks, where the properties of the model, for example velocities, are constant within a block and may vary from one block to another. These blocks can be rectangular prisms regular in size, *fullwave3D* uses regular cubes as it is the most efficient choice (Mufti 1990) and will be discussed in more detail later in this section. Figure 2.4(a) shows an example of a 3D velocity model consisting of two different zones with velocities V_{p1} and V_{p2}. The top left corner of the model has coordinates $(0, 0, 0)$. The model is divided by a set of vertical planes parallel to axes x and y and horizontal planes parallel to axis z. These planes are uniformly spaced along the axes. Intersects of the planes are called grid points. Figure 2.4(b) demonstrates a set of grid points within a 3D model, discretised into a number of rectangular prisms at regular intervals Δx, Δy and Δz along axis x, y and z respectively.



(a) (b)

Figure 2.4: (a) An example of a 3D finite difference model and (b) a set of grid points within a 3D model.

By definition of vector calculus, the first-order derivative of a one-dimensional function f at point x_0 is given by

$$f'(x_0) = lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{2.53}$$

Considering that the function f is smooth in the vicinity of point x_0, the derivative can be approximated using values of the function at a given mesh with a distance between the grid points, i.e. grid size, of Δx using the Taylor expansion

$$f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + \mathcal{O}(\Delta x^2) \tag{2.54}$$

Re-arranging the terms in equation 2.54 gives the approximation of the derivative at x_0

$$\begin{aligned} f'(x_0) =& \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} + \mathcal{O}(\Delta x) \\ f'(x_0) \approx& \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \end{aligned} \tag{2.55}$$

Obtained approximation is called the forward difference. It has an error of order Δx and provides the approximation of the derivative not exactly at point $x = x_0$ but at $x = x_0 + \frac{1}{2}\Delta x$. There are two other types of approximation - the central and backward differences. Extending the Taylor approximation as

$$\begin{aligned} f(x_0 + \Delta x) = f(x_0) + f'(x_0)\Delta x + \frac{1}{2}f''(x_0)\Delta x^2 + \mathcal{O}(\Delta x^3) \\ f(x_0 - \Delta x) = f(x_0) - f'(x_0)\Delta x + \frac{1}{2}f''(x_0)\Delta x^2 - \mathcal{O}(\Delta x^3) \end{aligned} \tag{2.56}$$

and differing the two equations results in

$$\begin{aligned} f'(x_0) =& \frac{f(x_0 + \Delta x) - f(x_0 - \Delta x)}{2\Delta x} + \mathcal{O}(\Delta x^2) \\ f'(x_0) \approx& \frac{f(x_0 + \Delta x) - f(x_0 - \Delta x)}{2\Delta x} \end{aligned} \tag{2.57}$$

Equation 2.57 is approximation of the first-order derivative using the central difference. It is more accurate compared to the forward difference as it computes the derivative precisely at point x_0 and has an error of order Δx^2. Summing equations in 2.56 and re-arranging the terms gives the approximation of the second-order derivative, as well with an error of order Δx^2

$$\begin{aligned} f''(x_0) =& \frac{f(x_0 + \Delta x) - 2f(x_0) + f(x_0 - \Delta x)}{2\Delta x^2} + \mathcal{O}(\Delta x^2) \\ f''(x_0) \approx& \frac{f(x_0 + \Delta x) - 2f(x_0) + f(x_0 - \Delta x)}{2\Delta x^2} \end{aligned} \tag{2.58}$$

The central difference derived to the accuracy of 4^{th} order in space for a one-dimensional function is given by (Mufti 1990)

$$f''(x_0) \approx \frac{-f(x_0 - 2\Delta x) + 16f(x_0 - \Delta x) - 30f(x_0) + 16f(x_0 + \Delta x) - f(x_0 + 2\Delta x)}{12\Delta x^2} \tag{2.59}$$

Now, the equations 2.57, 2.58 and 2.59 can be used to solve the wave equation. The first FD schemes in 2D were presented by Levander (1988) and Virieux (1986). I will present the solution for a 3D case.

For the simplicity, neglecting the density in the wave equation 2.6 and re-writing it in a matrix form to obtain

$$\frac{1}{V_p^2(\mathbf{r})}\frac{\partial^2 \mathbf{p}}{\partial t^2} - \nabla^2 \mathbf{p} = \mathbf{s}(t, \mathbf{r}) \tag{2.60}$$

where $\mathbf{r} = (x, y, z)$ denotes coordinates, $\mathbf{p}$ is the acoustic wavefield, V_p is the velocities defined as a constant value at every elementary block of the size Δx x Δy x Δz, t is time and $\mathbf{s}$ is the source function, defined only at the grid point where the source is located and will be dropped in further calculations. Considering the regular cubic mesh, where $\Delta x = \Delta y = \Delta z = h$, the central difference solution of fourth-order accurate in space and second-order accurate in time for such wave equation is given by Mufti (1990)

$$\begin{aligned} p^{n+1}(x,y,z) = & g(x,y,z)[p^n(x-2h,y,z) + p^n(x+2h,y,z) + p^n(x,y-2h,z) + \\ & p^n(x,y+2h,z) + p^n(x,y,z-2h) + p^n(x,y,z+2h) - \\ & 16(p^n(x-h,y,z) + p^n(x+h,y,z) + p^n(x,y-h,z) + \\ & p^n(x,y+h,z) + p^n(x,y,z-h) + p^n(x,y,z+h)) + \\ & 90p^n(x,y,z)] + 2p^n(x,y,z) - p^{n-1}(x,y,z) \end{aligned} \tag{2.61}$$

where n is the current time t, with Δt being a time sampling, $n-1$ and $n+1$ are the previous and next time steps and $g(x,y,z) = (V_p(x,y,z)\Delta t/h)^2/12$. As can be noticed, in order to obtain the solution at the future time step, only the solutions for current and previous time steps must be known.

Throughout this thesis two versions of the code were used to solve the anisotropic acoustic wave equation in both 2D and 3D in time domain on a regular cubic mesh: (1) the low-order kernel, that is 4^{th} order accurate in space and 2^{nd} order accurate in time and uses an optimised 53-point finite-difference stencil within 5 x 5 x 5 cube giving the performance approximately equivalent to 6^{th} order in space and 4^{th} order in time (Umpleby et al. 2010; Warner et al. 2013) and (2) the high-order kernel, that is 10^{th} order accurate in space and either 4^{th} or 2^{nd} order accurate in time (Silverton 2015).

2.2.2 Numerical stability and grid dispersion

Finite-difference solutions to wave propagation problems have been intensively studied in order to avoid generating erroneous seismic wavefield (Alterman et al. 1968; Boore 1970). Due to approximation of the derivatives by finite-differences, the simulation produces correct result only if certain criteria are met, known as the dispersion criterion and the stability criterion. Both criteria depend on the order of accuracy of the scheme and dimension of the problem.

In the case where the grid is too coarse, the wavefield computed by finite-difference become heavily dispersed (Alford et al. 1974). For the fourth-order accurate in space and second-order accurate in time finite-difference scheme, to avoid significant numerical dispersion, no less than 5 grid point per wavelet must be used (Alford et al. 1974; Dablain 1986; Levander 1988). Considering a regular cubic mesh, where $\Delta x = \Delta y = \Delta z = h$, the dispersion condition can be written as

$$h \leq \frac{\lambda_{min}}{5} \tag{2.62}$$

where the shortest wavelength λ_{min} corresponds to the lowest P-wave area of the model in acoustic case or the lowest S-wave area in elastic case. Using the relationship between frequency f, velocity V and wavelenght λ

$$f = \frac{V}{\lambda} \tag{2.63}$$

the grid spacing h for 4^{th} order in space and 2^{nd} order in time finite-difference scheme is limited by minimum velocity V_{min} and maximum frequency f_{max} as

$$h \leq \frac{V_{min}}{5 f_{max}} \tag{2.64}$$

Note, that for acoustic marine case studies, minimum seismic velocity V_{min} usually corresponds to the minimum velocity in the water column.

The stability condition, also known as Courant-Friedrichs-Lewy (CFL) condition, was first described in 1928 by Richard Courant, Kurt Friedrichs, and Hans Lewy and states that the time step for many explicit time-domain finite-difference schemes must be less than a certain value, otherwise the simulation provides the incorrect results (Courant et al. 1967). The idea behind this condition is that if a wave is propagated within spatially discretised medium and the amplitudes must be calculated at every time step, then the wave must cross less than one cell at every time step. For the 4^{th} order accurate in space and 2^{nd} order accurate in time finite-difference scheme, Graves (1996) demonstrates that a wave must cross less than half a cell at every time step, which means that the time sampling depends on the spatial sampling h and maximum velocity V_{max}:

$$\Delta t < 0.495 \frac{h}{V_{max}} \tag{2.65}$$

The described above conditions applicable to the low kernel code as it is 4^{th} order accurate in space and 2^{nd} order accurate in time.

The first version of the high kernel code, that is 10^{th} order accurate in space and 2^{nd} order accurate in time requires 3.6 grid points per wavelet and allows for a wave to cross no more than 38% of every cell while the second, more accurate in time version, that is of the 4^{th} order in time and 10^{th} order accurate in space, requires 2.6 grid points per wavelet and allows to cross no more than 66% of the cell (Silverton 2015).

2.2.3 Absorbing boundary conditions and source injection

Numerical simulation of the wavefield introduces a problem associated with the model boundaries. It is vital to impose appropriate boundary conditions in an effort to avoid artificial reflections from the model boundaries as those interfere with the desired wavefield.

The obvious solution to this problem seems to be simply enlarging the model enough for the wavelet to get attenuated by the time it reaches the boundaries. This method requires a large amount of additional grid cells to be added to the model and, therefore, is computationally not feasible.

A variety of ways to set up absorbing boundary conditions have been proposed (Berenger

1994; Cerjan et al. 1985; Clayton et al. 1977; Smith 1974). Here I describe only two types, implemented in *fullwave3D*.

The first boundary condition is one of the simplest schemes available and was used throughout this thesis. Proposed by Cerjan et al. (1985), this scheme pads each model boundary with extra cells, where the amplitudes of the wavefield are multiplied by a negative exponential function

$$G = e^{-(0.015(n-i))} \tag{2.66}$$

where n is the amount of cells and i is the position of a cell withing this amount. Typically, n is set to 20 in order for the boundary to be efficient and to not cause excessive computational effort (Cerjan et al. 1985). In this thesis, the amount of cells was chosen for every problem individually to form boundaries twice as large as the longest wavelength.

The second boundary condition, available in *fullwave3D*, attempts to propagate the wavefield beyond the boundaries, thus avoiding reflections from them. This is done by quadratic approximation of the wavefield outside the modelling domain, allowing wave propagation within the modelling domain up to the boundaries (Debens 2016). This boundary condition was not used in this thesis.

Note, that there is a special type of boundary conditions required for solid-air and liquid-air interfaces that are used in land and marine datasets respectively. This boundary condition needs to be implemented with great care. In case of marine data, it is common to use so-called “free-surface” boundary condition for the water-air interface in an attempt to create almost a perfect reflection from this interface to simulate source and receiver ghosts and sea-surface multiples. This is achieved using zero-stress boundary conditions that set stresses σ_{zz}, σ_{zx} and σ_{yz} at the interface to zero and implements a symmetry for the stress components around the interface (Graves 1996) .

The term which was ignored until now in the derivation of different equations but is essential to the simulation is how to include the source in the modelling. In case of a staggered-grid implementation, this can be accomplished in 2 different ways: either by simulating the acoustic pressure at one point of the model (which uses the underlying assumption that we are modelling the source as a single source point), or by using velocities (Graves 1996). Either way, sources are represented as time series which should have the same time sampling that is used for the data and that respects the stability and dispersion conditions. Implementing the source in finite difference is described in detail by Guash (2011).

2.3 Application to real data

A generic velocity model building workflows consists of acquisition, processing, model constructing and depth imaging, where FWI takes the role to improve a velocity model, typically in top 3 km of heterogeneous overburden, prior to re-migration. Re-migration with a high-resolution FWI velocity model allows for significant improvment within migrated image at greater depths and correctly positions reflectors under complex overburden. Over the last 10 years FWI has become a routine component of the conventional velocity model building workflow for 3D field data.

In practice, FWI can provide spectacular results only if the data meet certain requirements and the starting model is accurate enough. It makes use of low frequencies, transmitted arrivals, iterates from a best-guess starting velocity model and details can be critical - if it fails, it does it completely. Most of the conventional imaging techniques will not apply to FWI workflows as FWI often uses the full wavefield including, for instance, multiples (Warner et al. 2013).

This section is devoted to the practical requirements for successful FWI implementation, provides reasoning to this requirements, discusses the main challenges of FWI and gives a brief overview of possible solutions.

2.3.1 What wavenumbers can be resolved

When a seismic wave enters one medium from another, part of the wave is reflected at the interface, and part of it enters the medium, or transmits. If the wave is incident at an oblique angle, then the transmitted wave can change its direction of propagation and this phenomena is called refraction. Refraction can occur only in heterogeneous media. In case of heterogeneous media where the velocity increases with depth, refracted waves are called “diving” or “turning” waves. In seismology, refracted waves are often referred to as "transmitted". Seismic wavefield usually contains both refracted and reflected energy. These two types of seismic waves are shown in Figure 2.5. Use of transmission and reflection energy allows FWI to recover both low and high wave-number components of velocity model. Mora (1989) shows that, in principle, FWI is able to recover all wave-number components that are present in the data, with the upper limit set by the wavelet bandwidth. The relationship between the energy type and wave numbers they recover is essential to understanding the potential and limitations of FWI.

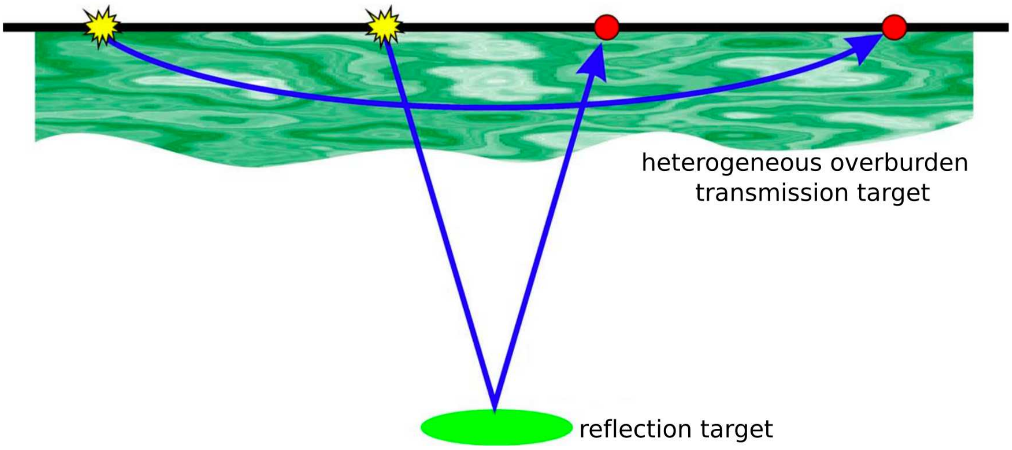


Figure 2.5: Schematic showing transmitted energy through heterogeneous shallow overburden is recorded by long offsets and reflected energy penetrating at depth is recorded by short offsets. Yellow and red circles represent sources and receivers respectively (Morgan 2013).

There are a few different ways to approach this same relationship. I will start with considering a basic 1D plane wave scattering experiment in a homogeneous medium (Fig. 2.6). In this case, the incident and scattering angles are equal, so that the wavenumber vectors of the

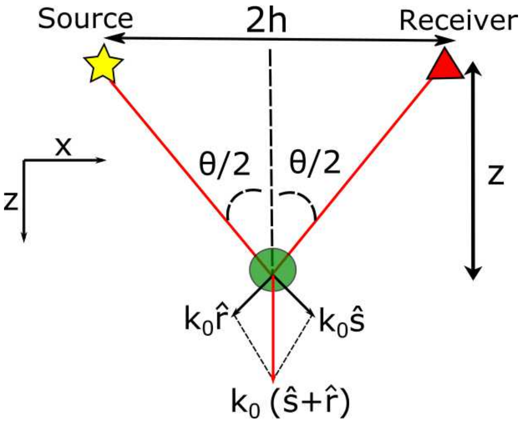


Figure 2.6: A 1-D scattering experiment in a homogeneous medium.

incident and scattered wave, $\mathbf{k}_s = k_0\hat{\mathbf{s}}$ and $\mathbf{k}_r = k_0\hat{\mathbf{r}}$ respectively, can be expressed as (Sirgue et al. 2004)

$$\begin{aligned} \mathbf{k}_s &= k_0\hat{\mathbf{s}} = (k_0 \sin\frac{\theta}{2}, k_0 \cos\frac{\theta}{2}) \\ \mathbf{k}_r &= k_0\hat{\mathbf{r}} = (-k_0 \sin\frac{\theta}{2}, k_0 \cos\frac{\theta}{2}) \end{aligned} \tag{2.67}$$

where $k_0 = \frac{2\pi f}{V}$ is a wavenumber, defined by frequency of the wave f and velocity of the medium V, $\hat{\mathbf{s}}$ and $\hat{\mathbf{r}}$ are source-to-scatter and receiver-to-scatter unit vectors, θ is angle between the incident and scattered propagation directions. From Figure 2.6 it can be observed that

$$\begin{aligned} \cos\frac{\theta}{2} &= \frac{z}{\sqrt{h^2 + z^2}} \\ \sin\frac{\theta}{2} &= \frac{h}{\sqrt{h^2 + z^2}} \end{aligned} \tag{2.68}$$

where $2h$ is the offset and z is the depth of the scatter. Summing equation 2.67 gives the horizontal and vertical components of the vector $\mathbf{k} = k_0(\hat{\mathbf{s}} + \hat{\mathbf{r}})$ as

$$\begin{aligned} k_x &= 0 \\ k_z &= 2k_0 \cos\frac{\theta}{2} \end{aligned} \tag{2.69}$$

Equation 2.69 defines the wavenumber that can be resolved in the experiment. Note, that the horizontal component k_x of the wave number vector $\mathbf{k}$ equals zero due to the equal magnitudes of the incident and scattered angles. Introducing offset-to-depth ratio as a new variable

$$R = \frac{h}{z} \tag{2.70}$$

and re-writing the equation 2.68 as

$$\cos\frac{\theta}{2} = \frac{z}{\sqrt{h^2 + z^2}} = \frac{1}{\sqrt{1 + R}} \tag{2.71}$$

shows that the relationship between the scattering angle θ and the offset-to-depth ratio R demonstrates that as the scattering angle theta increases the offset-to-depth ratio R also increases: the larger is R, the bigger is the value of the angle θ and, consequently, the smaller is the value of $\cos\frac{\theta}{2}$. Large value of R, in its turn, means long offsets $2h$ and shallow depth

z. Keeping that in mind and looking back at equation 2.69, it can be seen that for a given frequency f and velocity V the very low wavenumber values can be resolved when the scattering angle θ is very close to 180°. This is why diving waves, that have large scattering angles and are typically recorded by long offsets, are crucial for resolving low wavenumbers, and are a vital component for successful FWI implementation.

The main observations from presented analysis are: (1) one frequency and one aperture map one wavenumber in the model, (2) the larger the offset-to-depth ratio, and, hence, the associated angle between the incident and scattered waves, the smaller wavenumber is recovered, therefore (3) diving waves are vital for resolving low wavenumbers, (4) maximum wavenumber that can be resolved is $k_{max} = 2k_0$ and it is limited by the frequency f and velocity V.

The second observation might lead to the conclusion, that reflected energy, typically recorded by short offsets, recovers high and intermediate wavenumbers and refracted energy, typically recorded by long offsets, recovers low wavenumebrs. This conclusion is, however, incomplete. Reflection energy also contains the low wavenumber information (Yao et al. 2014a) as illustrated in Figure 2.7. This is a model with a point scatterer and a sub-critical reflector beneath it. The blue reflected wave has angle α much less than 180°, therefore it meets the criteria for the recovery of the high wavenumber components of the velocity model. The second wave, shown in red dashed line, has the scattering angle β close to 180°. This means that it meets the criteria for the recovery of the low wavenumbers at the depth of the scatter. Note, that this transmitted-at-the-scatterer wave is also reflected at the deeper reflector and at this deeper depth it meets the criteria only for the recovery of the high wavenumber components.

Why then does FWI struggle to use reflection energy to extract the low wavenumbers and uses it for high wavenumbers recovery only? The qualitative explanation to that, which describes the behaviour of standard FWI for pure reflection data using a smooth starting model, is given by Yao et al. (2017). I will present its short overview here. Figure 2.8 demonstrates how standard FWI algorithm will proceed for the first two iterations using pure reflection data and a starting model with no reflectors in it.

At the first iteration, forward modelling consists only of transmitted arrivals because of the fact, that the starting velocity model is smooth. The residual wavefield, which is the

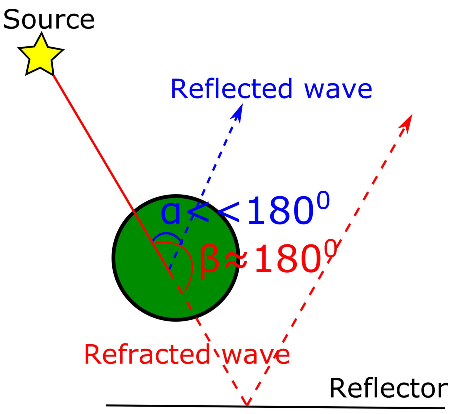


Figure 2.7: Scattering at a point scatterer and a deeper reflector. If the angle is small, only the high wavenumber components are recoverable; if the angle is large, the low wavenumber components are recoverable.

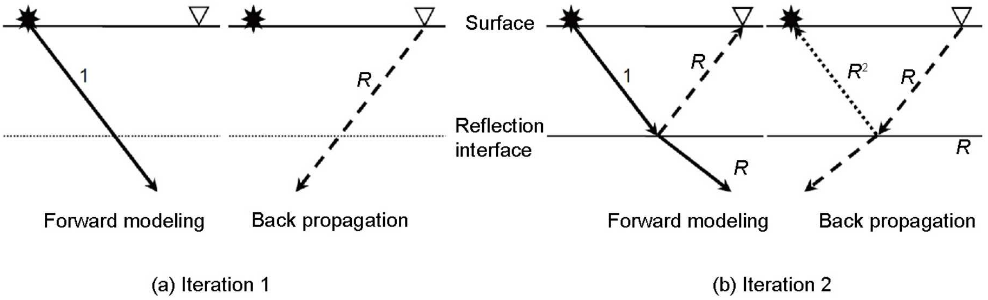


Figure 2.8: Behaviour of FWI algorithm during (a) first and (b) second iteration when using purely reflection data and smooth starting model (Yao et al. 2014b).

difference between the given reflection data and modelled transmission data, contains only reflection energy present in the data. Setting the amplitude of propagated by the source waves to 1, this residual reflection energy will be of magnitude R, where R is some reflection coefficient. Cross-correlating these two wavefields builds a reflector with the same reflection coefficient R into the smooth starting model, therefore, updating the high wavenumbers. This performance is similar to that of RTM (reverse time migration) and that is why this effect is referred to as migration effect of FWI (Mora 1989).

At the second iteration, the smooth velocity model contains the reflector with reflection coefficient R, so that forward propagated wavefield contains both transmitted energy with a magnitude approximately remaining 1 and reflected energy of magnitude R. The back propagated wavefield now contains transmitted energy of magnitude R and reflection energy of magnitude R^2. Cross-correlation of these two wavefields continues to build the reflector of magnitude R and updates the smooth starting model with the updates of magnitude proportional to R^2. Typically, $|R| << 1$, which results in the much stronger reflectivity update (migration term) compared to the background model update (tomography term).

To conclude, FWI is implemented by minimisation of the difference between calculated and observed data, where the difference is formulated using L_2 norm. This conventional approach is an iterative process that updates the velocity model in the direction indicated by the gradient. It successfully uses: (1) refracted energy from target depth for velocity model updates with both low and high wavenumbers and (2) sub-critical reflected energy for velocity model updates with only high wavenumbers (Yao et al. 2014a).

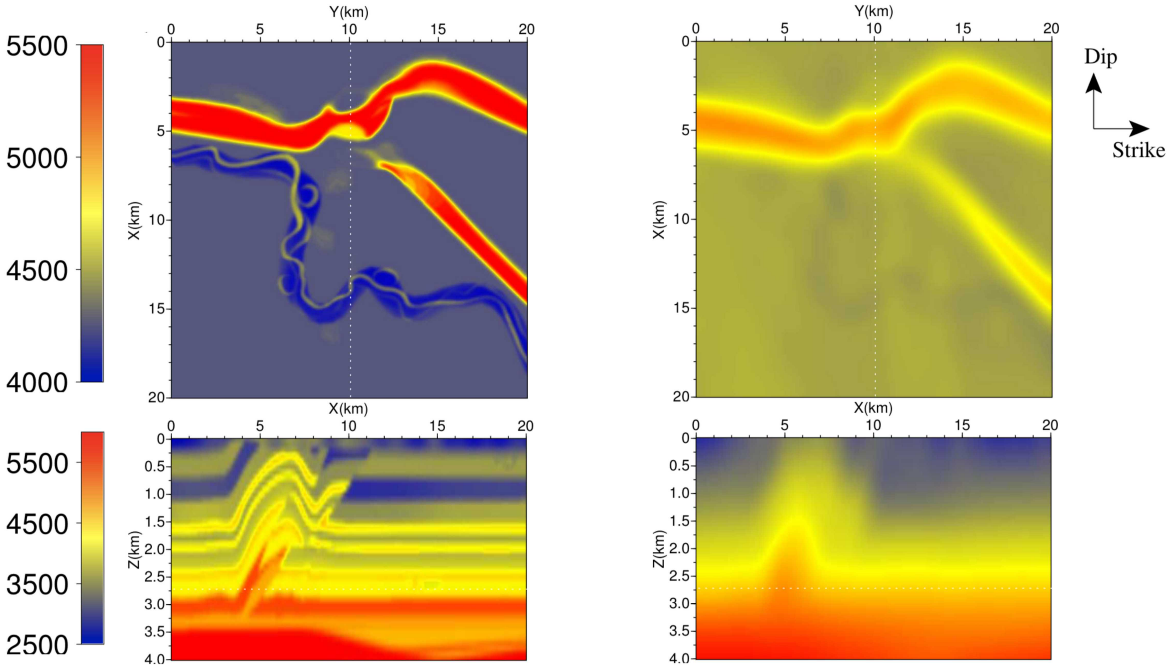
Because conventional FWI updates long wavelength using transmitted arrivals, it needs long offsets with turning energy penetrating the target, close enough starting velocity model and frequencies as low as 3-4 Hz (Yao et al. 2014b). I will discuss the importance of the long offsets, low frequencies and starting velocity model and their influence on the inversion result later in the chapter. I also will briefly review some of the alternative formulations of FWI.

2.3.2 Acquisition: Narrow versus Wide Azimuth

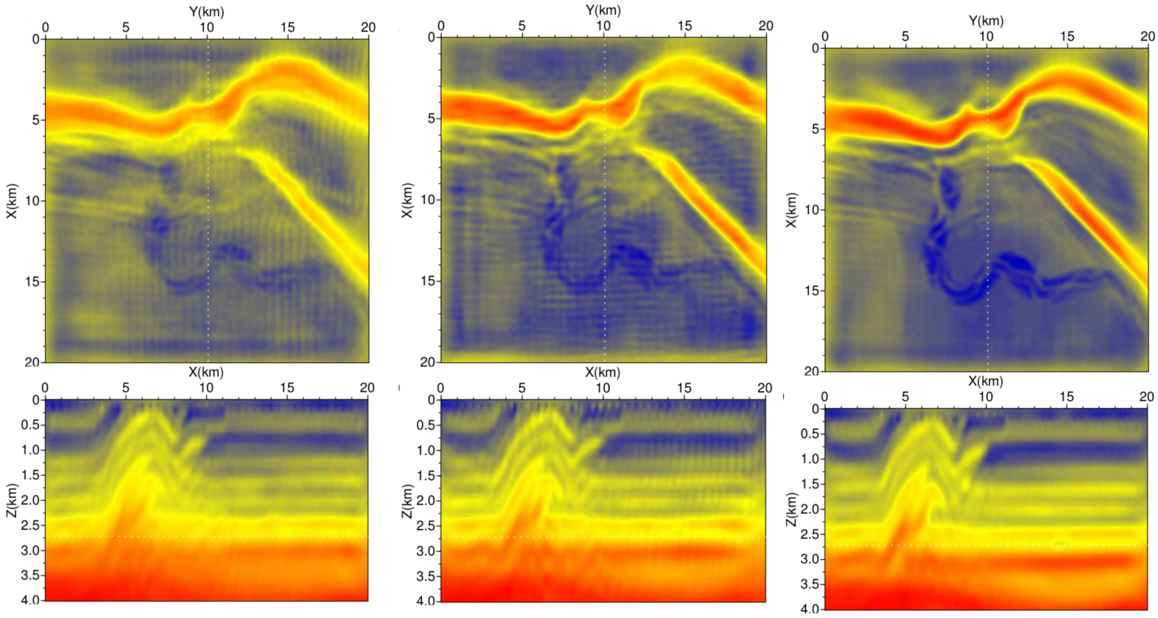
The quality of a depth migrated image may be strongly impacted by the acquisition type. For example, Etgen et al. (1998) explores the impact of the orientation of narrow azimuth surveys on the quality of Pre-Stack Kirchoff Depth migration on a 2.5D synthetic model. They show the substantial quality difference in images obtained using the same migration algorithm but different acquisition type. Regone (2006) has compared Pre-Stack Kirchhoff Depth migration results obtained using narrow-azimuth towed streamer (NATS) and with wide-azimuth towed streamer (WATS) data. It is clear from this 3D synthetic study that wide azimuth data provide a superior image compared to the narrow azimuth data.

The influence of acquisition type on FWI results had remained unassessed until Sirgue et al. (2007) set up a number of synthetic experiments to evaluate the impact of acquisition azimuth on 3D FWI. In the study they use SEG/EAGE over-thrust velocity model (Fig. 2.9(a)) and conduct three numerical tests to model a set of narrow and wide-azimuth experiments. Two narrow-azimuth acquisitions were modelled in strike and in dip direction with the maximum offset of 8 km and a cable spread of 1km. Wide-azimuth acquisition has both inline and crossline offsets limited to 8 km. The starting velocity model for all three tests remains the same and is shown in Figure 2.9(b). The 3D FWI velocity models obtained with a maximum frequency of 5 Hz clearly show that the acquisition type has a large impact on FWI results (Fig. 2.9(c),(d),(e)). It can be seen that the narrow-azimuth acquisition oriented in dip direction (Fig. 2.9(d)) provides better results than the one oriented in strike direction (Fig. 2.9(c)). This result is also consistent with the conclusions of Etgen et al. (1998). Comparing the best narrow-azimuth experiment (Fig. 2.9(d)) with the wide-azimuth experiment (Fig. 2.9(e)), it is obvious that wide-azimuth acquisition type yields superior depth image which does not contain the stripy artefacts. These artefacts can be removed using smoothing during inversion. However, this smoothing may damage the resolution of objects of a comparable size to that of the artefacts (Sirgue et al. 2007).

That makes more expensive wide-azimuth surveys, for instance ocean bottom cable (OBC) acquisition and wide-azimuth towed-streamer (WATS) more desirable for FWI.



(a) True velocity model (b) Starting velocity model



(c) NAZ in strike direction (d) NAZ in dip direction (e) WAZ

Figure 2.9: Horizontal slice at the depth of 2.7km and cross-section: (a) True velocity model, (b) Starting velocity model. 3D FWI results at 5 Hz with different acquisition type: (c) narrow-azimuth acquisition in strike direction, (d) narrow-azimuth acquisition in dip direction, (e) wide-azimuth acquisition (after Sirgue et al. (2007)).

2.3.3 Low frequencies and starting earth model

Vital ingredients of successful FWI are: (1) availability of a reasonable starting earth model and (2) low frequencies present in the data. Good starting earth model lies within the proximity of the global minimum and produces the data that are matching the acquired seismic data within half a cycle. In this circumstances, FWI is likely to converge to the true earth model.

In many cases, such a starting model may not be available (Alkhalifah 2016) and the inversion algorithm may converge to a local minimum solution as demonstrated in Figure 2.10. Here the initial guess model, indicated by the blue circle, used in combination with the data of given frequency content shown by green dashed line will drive the algorithm towards a local minimum solution indicated by the orange circle (Fig. 2.10). The problem of local minima is commonly referred to as a cycle skipping problem.

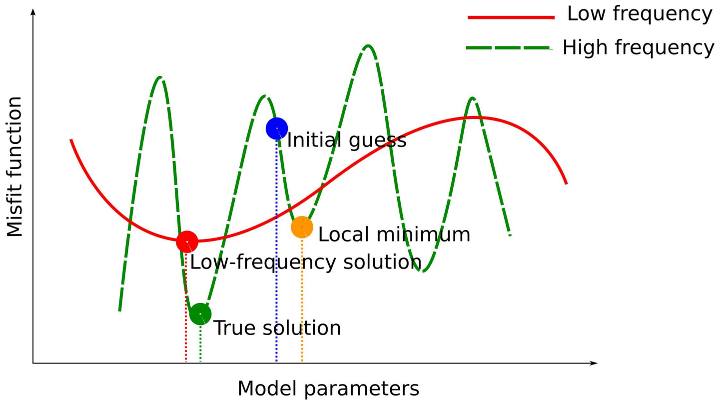


Figure 2.10: Illustration of the local minima for a one-variable optimisation problem.

The standard approach of avoiding local minima is to begin the inversion with the lowest possible frequency. The advantage of this strategy is illustrated in Figure 2.10. The initial guess earth model shown in the blue circle, used with the high frequency data in green, leads the inversion towards a local minimum. Whereas the very same initial guess model, used with the low frequency data in red, enables the algorithm to converge towards the global minimum for the low frequency. This low-frequency solution lies in the vicinity of the true solution. Once the low frequency solution shown in the red circle is found, it can be used as a starting model for the high frequency as, unlike the initial guess, it lies in the basin of attraction of the global minimum. That is why starting the inversion with the low frequencies and gradually including the high frequencies seems to be a reasonable approach to successfully perform FWI.

The data rich in low frequencies allows the relaxation of the constraint on the initial guess velocity model that is commonly produced using tomography and pre-processed data, and enables the recovery of low wave number model updates. For this reason, in exploration geophysics there is a tendency to acquire seismic data as rich as possible in low frequency content with as high as possible signal-to-noise ratio at these frequencies (Mancini et al. 2015).

2.3.4 Long offsets

Unfortunately, sometimes low-enough frequency data may not to be available. However, the wavenumber resolved by FWI depends not only on the frequency content, but on the scattering angle as well (Hu et al. 2018)

$$k = \frac{4\pi f}{V} \cos\frac{\theta}{2} \tag{2.72}$$

where k is the wavenumber dictating the model reconstruction spatial resolution, f is the frequency and θ is the subsurface scattering angle (the angle between the incident and scattered wave) that, in its turn, depends on sources and receivers located around a target and is shown in Figure 2.11. From the equation 2.72 it can be seen that the scattering angle θ plays an important role in defining the wavenumber k: the larger the angle, the lower wavenumber can be resolved given a certain frequency. With this observation in mind and looking at Figure 2.11, one can conclude that the longer the distance between the source and receiver (given a target depth), the greater the angle θ. Consequently, longer offsets are another component

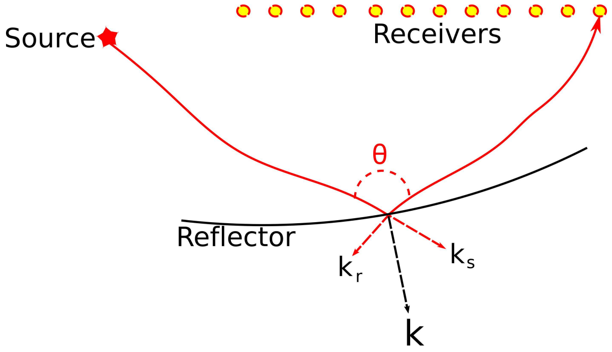
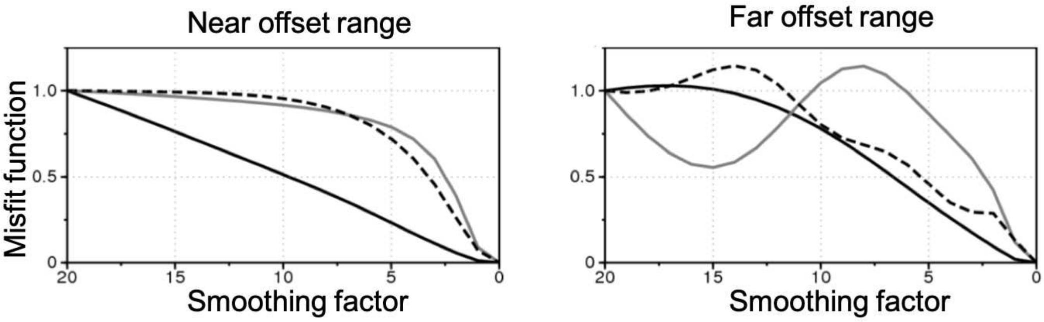


Figure 2.11: Schematic illustration of a relationship between scattering angle θ and long offsets: defining the wavenumber $\mathbf{k}$ that can be resolved during the inversion. The wavenumber $\mathbf{k}$ is a sum of the wavenumbers of the incident wave and the scattered wave, the vectors $\mathbf{k}_s$ and $\mathbf{k}_r$ respectively.

of successful FWI because they record energy paths with larger scattering angles compared to short offsets. Investigations of Sirgue et al. (2004) are consistent with this conclusion as they demonstrate that the short offsets provide information about the high wavenumbers and the long offsets provide information about the low wavenumbers through wide-angle illumination. The synthetic experiment of Brenders et al. (2007) shows that the accuracy of the acquired model is limited by the maximum offset in seismic survey. However, Sirgue (2006) performs an analysis of the complex relationship between offsets, low frequency and non-linearity and demonstrates that, unlike the near offset data, the long offset data are highly non linear with respect to the low wave numbers. This is shown in Figure 2.12. The misfit function is highly non-linear even at the frequencies as low as 3 Hz and 5 Hz for the long offset data (Fig. 2.12(b)), whereas it appears to be more linear for the short offset data (Fig. 2.12(a)). That simply means that the long offset data are the most likely to be cycle skipped as they are the most non-linear component. Therefore, the beneficial long offsets data should be included in the inversion with extra care.



(a) (b)

Figure 2.12: Normalised misfit functions at 1 Hz (solid black), 3 Hz (grey) and 5 Hz (dashed black) as a function of smoothness of the velocity model for (a) a near offset range 2-2.5 km and (b) a far offset range 10-10.5 km (after Sirgue (2006)).

2.4 Alternative formulations of FWI

Conventional FWI has been an effective tool for generating high-resolution velocity models of the subsurface. The scheme seeks to obtain an earth model that minimises the misfit function given by the difference between the observed and modelled data mathematically formulated in a least square sense. Sometimes, the misfit function is set to be the zero lag of the cross-correlation of the observed and modelled data and in such cases FWI algorithm is configured to maximise that misfit (Warner et al. 2016). In either case, the success heavily relies on presence of transmitted arrivals and low frequencies as well as a good starting earth model. Transmitted arrivals, recorded typically by long offsets, define the maximum depth of low wave number updates, and low frequencies in combination with an accurate starting model ensure cycle skipping does not occur. In other words, conventional FWI faces two challenges: (1) limited range of offsets or lack of transmitted arrivals limit the low wave-number model update as well as constrain the depth penetration of the update and (2) an inaccurate starting earth model along with limited low frequencies causes cycle skipping.

Over the last decade, in an effort to tackle these two major problems several other waveform inversion methods have been suggested. As there has been a need for getting the updates beyond the reach of diving waves, and because it may be possible for the data not to contain refracted energy, FWI needs to make use of reflection energy. Reflection FWI schemes that enable recovery of low wave-number deep model update in pure reflection data were proposed by Irabor et al. (2016), Xu et al. (2012), and Yao et al. (2017). Irabor et al. (2016) demonstrated successful application of isotropic Reflection FWI on the Marmousi model.

There has also been a number of methods to overcome cycle skipping suggested (Biondi et al. 2012; Leeuwen et al. 2013; Luo et al. 2011; Ma et al. 2013; Metivier et al. 2016; Warner et al. 2014a; b; 2015). All these works re-formulate the underlying problem to expand the dimensionality of the unknown model space by using different mathematical approaches so that the non-physical model converges to the physical one.

The detailed overview of different groups of waveform inversion methods with extensive list of references is presented by Hu et al. (2018).

In this section I give a short theoretical overview of a few different types of FWI that were used during my internship. They are different to conventional FWI implemented in the major part of the thesis. I briefly summarise the main features of these types and compare the advantages and disadvantages of each approach. The results of the internship project are described in Chapter 7.

2.4.1 Travel Time FWI

Conventional least-squares FWI solves the wave equation by optimising model parameters (typically velocity) to minimise the difference between the observed and calculated data, known as the misfit function. A good starting model and presence of low frequencies are essential to avoid cycle skipping, and consequently, local mimima. However, a sensible starting model, that provides the mismatch between the calculated and observed data within half a cycle, may not be available and/or required low frequencies may be absent in the data. Under such conditions, lest-squares FWI will lead to a local minimum solution.

First introduced by Luo et al. (1991) and lately modified by Wang et al. (2018a), a travel time based FWI suitable for complicated scenarios was suggested as a solution to the cycle skipping problem. Instead of minimising the misfit function based on the data difference, a new misfit function has been proposed. This function is based on the weighted travel time shifts with the correction of the kinematic error between the observed and calculated data and is formulated as (Wang et al. 2018a)

$$\phi_{tt}(\mathbf{m}) = \frac{1}{2}||A\tau_0||_2^2 \tag{2.73}$$

such that

$$\tau_0(\mathbf{x}_r, t; \mathbf{x}_s) = argmax_\tau \, c(\tau, \mathbf{x}_r, t; \mathbf{s}_s) \tag{2.74}$$

where A is the weighting operator and τ_0 is maximising a window cross-correlated function c given by (Wang et al. 2018a)

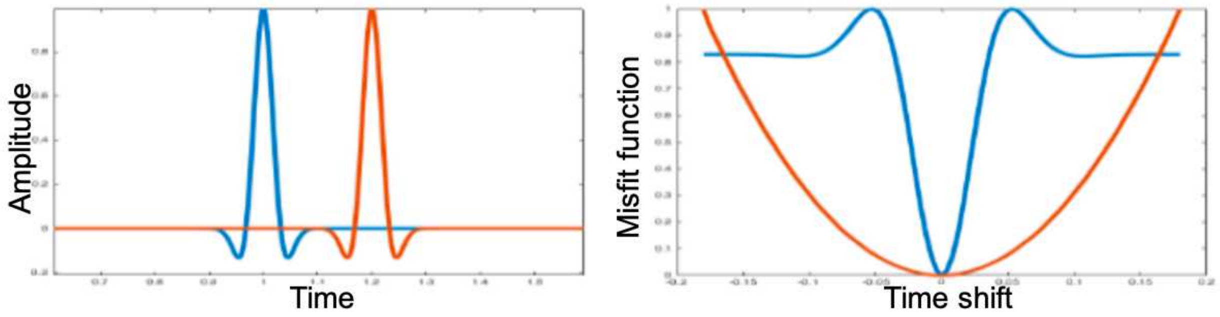
$$c(\tau, \mathbf{x}_r, t; \mathbf{x}_s) = \int_{-\omega}^{\omega} \mathbf{R}\mathbf{p}(\mathbf{x}_r, t+s; \mathbf{x}_s)\mathbf{d}_{obs}(\mathbf{x}_r, t+\tau+s; \mathbf{x}_s)\, ds \tag{2.75}$$

where $\mathbf{R}$ is the projection matrix, applied to the calculated wavefield $\mathbf{p}$ so that $\mathbf{Rp}$ gives the calculated wavefield at source and receiver locations only, $\mathbf{x}_s$ and $\mathbf{x}_r$ respectively, at every time sample t. s is time, τ is a time lag and τ_0 is the travel time shift that is a function of time and space for small window size ω for each source $\mathbf{x}_s$.

Figure 2.13 illustrates the advantage of the travel time based function. Considering a

Ricker wavelet in blue being an observed signal and in red being calculated data as shown in Figure 2.13(a). Figure 2.13(b) demonstrates two differently calculated misfit functions, both depending on the time shift between the two wavelets. One is the conventional misfit function and another is the travel time based misfit function. It can be seen that the least squares misfit function shown in blue has local minima as well as the global minimum and will converge to a local minimum solution if the time shift between the wavelets is more that half a cycle. On the other hand, the travel time based misfit function shown in red is convex with respect to time shifts and will converge to the global minimum in case the time shift between the wavelets is more than half a cycle.

This travel time objective function, combined with the conventional way of calculating the gradient, which is naturally suitable for diving waves, provides depth penetration still limited by offset range. However, it generates an adjoint source that is less likely to be cycle-skipped from correcting the travel times errors and somewhat relaxes the requirement for good starting model and low frequency data. This method can be efficiently used as a first step in FWI workflow in order to provide a better starting model to mitigate cycle skipping.



(a) (b)

Figure 2.13: Comparison of the misfit functions for Travel Time FWI and conventional FWI. (a) Observed wavelet in blue and synthetic wavelet in red; (b) Travel Time misfit function in red and Least Squares misfit function in blue (After Wang et al. (2018a)).

2.4.2 Extended Parameter FWI

Consider the following acoustic wave equation in its simplest form

$$\mathbf{Ap} = \mathbf{s} \tag{2.76}$$

where $\mathbf{A}$ is a linear operator (see Eq. 2.6), $\mathbf{p}$ is the the forward propagated wavefield and $\mathbf{s}$ is the source signature. Conventional FWI scheme seeks to obtain an earth model that minimises the misfit given by the sum of the squares of the sample-by-sample difference between the observed and modelled data, predicted by applying the wave equation to that earth model with a given source. In least square sense that misfit function can be written as

$$\phi(\mathbf{m}) = \frac{1}{2}||\mathbf{R}\mathbf{p} - \mathbf{d}_{obs}||_2^2 \tag{2.77}$$

where $\mathbf{m}$ is the earth model parameters, $\mathbf{R}$ is the projection matrix that, multiplied with the forward propagated wavefield $\mathbf{p}$, gives the wavefield at the receiver locations only. When the earth model $\mathbf{m}$ does not provide a match in time between the observed and modelled data within half a cycle, cycle skipping occurs. Such a cycle skipped model will result in obtaining a local minimum solution.

In order to mitigate the problem of local minima, a different approach was adopted by Wang et al. (2016), the main idea of which is to re-formulate the misfit function 2.77 by adding extra parameters to extend the solution space.

Extended parameter FWI solves the acoustic wave equation in a L_2 sense, as well as conventional FWI does, but expands the solution space by introducing a penalty term into the misfit function 2.77 in order to relax the constraint on the earth model. The penalty term represents a wave equation error that is weighted by a carefully chosen scalar λ. The new misfit function is introduced by Wang et al. (2016) as following

$$\phi_\lambda(\mathbf{m},\mathbf{g}) = \frac{1}{2}||\mathbf{R}\mathbf{p}_g - \mathbf{d}_{obs}||_2^2 + \frac{\lambda^2}{2}||\mathbf{A}(\mathbf{p}_g) - \mathbf{s}||_2^2 \tag{2.78}$$

and depends now on two sets of parameters - the earth model $\mathbf{m}$ and the new source $\mathbf{g}$ that is used for calculating the forward wavefield $\mathbf{p}_g$. Note, that conventional FWI minimises the misfit function 2.77, such that forward propagated wavefield $\mathbf{p}$ is calculated using strictly given source $\mathbf{s}$. The new source $\mathbf{g}$ is a perturbation of the given source $\mathbf{s}$, such that $\mathbf{g} = \mathbf{s} + \delta\mathbf{s}$. Considering that $\mathbf{A}(\mathbf{p}_g) = \mathbf{g}$ and the $G(\mathbf{m})$ is the solution operator of the forward propagated wave equation 2.76, the misfit function 2.78 can be re-written as

$$\phi_\lambda(\mathbf{m},\mathbf{g}) = \frac{1}{2}||\mathbf{R}G(\mathbf{m})_g - \mathbf{d}_{obs}||_2^2 + \frac{\lambda^2}{2}||\mathbf{g} - \mathbf{s}||_2^2 \tag{2.79}$$

This newly formulated problem is an optimisation problem with two sets of variables. It is computationally not feasible to search for both **m** and **g** simultaneously, therefore, there are two optimisation steps implemented. The first step is to approximate the new source **g** as (Wang et al. 2016)

$$\begin{aligned} \mathbf{g} =& \mathbf{s} + \delta\mathbf{s} \\ \delta\mathbf{s} \approx& \frac{1}{\lambda^2}G^T\mathbf{R}^T(\mathbf{d}_{obs} - \mathbf{R}G\mathbf{s}) \end{aligned} \tag{2.80}$$

fix the starting earth model **m** and seek for the source **g** that minimises the misfit function 2.79. Once this source **g** is found, as the second step, the algorithm fixes it and seeks for the optimal earth parameters **m** that minimise the misfit function 2.79. The parameter λ controls how much the first term in the misfit function impacts the inversion result. In case of λ set to a large value, the algorithm should produce the outcome similar to conventional FWI. More detailed explanations along with both synthetic and real data examples are presented by Wang et al. (2018a,b).

2.4.3 Born FWI

Because there has been a desire to obtain the earth model updates at depths greater than standard 2 to 3 km that are covered by refracted energy, FWI has been challenged to make use of reflection energy recorded by typical exploration offsets. However, using reflection dominated or purely reflection data with conventional FWI, only the high wave numbers can be recovered due to the dominant effect of reflectivity migration over velocity tomography (Yao et al. 2014b). In order to avoid cycle skipping, getting updates of the smooth background model is necessary prior to updating high wave numbers. The solution to this problem lies in separating migration and tomographic component, that can be achieved in several ways:

- using Born approximation
- using a local Poynting vector

- muting out the incident wavefield on migrated gathers
- using filtering in the wave-number domain

The first way, i.e. the idea of splitting the earth model (commonly velocity) into two, a long wavelength component and a short wavelength component models, and apply the first order Born approximation to the acoustic wave equation was presented by Hudson et al. (1980)

$$A(\mathbf{m})\mathbf{p}_S = s \tag{2.81}$$

$$A(\mathbf{m})\mathbf{p}_R = \mathbf{r}\,\partial_t^2\mathbf{p}_S \tag{2.82}$$

where A is the wave equation operator that depends on the smooth long wavelength component model $\mathbf{m}$, $\mathbf{r}$ is the given reflectivity model, $\mathbf{p}_S$ is the forward propagated transmitted wavefield generated by the source signature s and $\mathbf{p}_R$ is the forward propagated reflected wavefield. After the wave equation has been linearised into the two equations 2.81 and 2.82, that separate the low wavenumber tomographic component and high wavenumber migration component during forward modelling, the algorithm minimises conventional misfit function of L_2 norm of the data residuals, where the misfit function ϕ can be defined as

$$\phi(\mathbf{m},\mathbf{r}) = \frac{1}{2}||\mathbf{R}\mathbf{p}_R - \mathbf{d}_{obs}||_2^2 \tag{2.83}$$

and depends on both the smooth background model $\mathbf{m}$ and the reflectivity model $\mathbf{r}$.

Note, that the background wavefield $\mathbf{p}_S$ is free of multiples. Hence, the data must be de-multipled for this method. However, the system becomes more linear compared to conventional FWI and convergence of the inversion is boosted. Another advantages of using Born approximation are increased depth penetration as limits go beyond the reach of turning waves and generating low wavenumber updates using reflection energy, that is vital for the datasets with no or little presence of refracted energy.

2.4.4 Travel Time Born FWI

Born FWI, as described in the section above, makes use of reflection energy in order to obtain updates at depth greater than conventional FWI which is limited by shallow reach of refraction energy. It uses Born approximation to linearise the wave equation and, hence, improving chances of converging towards the global minimum given an accurate enough starting earth model. However, avoiding the solution getting trapped in a local minimum is still a challenge and requires a good low wavenumber starting earth model. In the absence of such a model, the problem of cycle skipping needs to be mitigated.

Over the last few years, a variety of solutions were proposed to tackle the problem of cycle skipping and associated with it local minima. Combining the Born FWI and travel time FWI described above, Wang et al. (2018a) suggested travel time based Born waveform inversion as a solution to a cycle skipping problem for reflection waveform inversion. This algorithm uses Born approximation to linearise the wave equation into a system of the two following equations

$$A(\mathbf{m})\mathbf{p}_S = s \tag{2.84}$$

$$A(\mathbf{m})\mathbf{p}_R = \mathbf{r}\,\partial_t^2\mathbf{p}_S \tag{2.85}$$

where A is the wave equation operator that depends on the smooth long wavelength component model $\mathbf{m}$, $\mathbf{r}$ is the given reflectivity model, $\mathbf{p}_S$ is the forward propagated transmitted wavefield generated by the source signature s and $\mathbf{p}_R$ is the forward propagated reflected wavefield. Let $G(\mathbf{m})$ be the solution operator for the linearised wave equation 2.84. Then the forward transmitted wavefield $\mathbf{p}_S$ generated by source s can be computed as (Wang et al. 2019a)

$$\mathbf{p}_S = S(\mathbf{m})s \tag{2.86}$$

and the forward reflected wavefield $\mathbf{p}_R$ can be calculated as (Wang et al. 2019a)

$$\mathbf{p}_R = G(\mathbf{m})\mathbf{r}\,\partial_t^2 G(\mathbf{m})s \tag{2.87}$$

Given a starting low wavelength model $\mathbf{m}$, the algorithm minimised the local travel time error between the calculated and observed data, mathematically expressed as least squares norm (Wang et al. 2019a)

$$\phi_{tt}(\mathbf{m}) = \frac{1}{2}||A\tau_0||_2^2 \tag{2.88}$$

such that (Wang et al. 2019a)

$$\tau_0(\mathbf{x}_r, t; \mathbf{x}_s) = \text{argmax}_\tau\, c(\tau, \mathbf{x}_r, t; \mathbf{s}_s) \tag{2.89}$$

where A is the weighting operator and τ_0 is maximising a window cross-correlated function c given by (Wang et al. 2019a)

$$c(\tau, \mathbf{x}_r, t; \mathbf{x}_s) = \int_{-\omega}^{\omega} \mathbf{Rp}(\mathbf{x}_r, t+s; \mathbf{x}_s)\, \mathbf{d}_{obs}(\mathbf{x}_r, t+\tau+s; \mathbf{x}_s)\, ds \tag{2.90}$$

where $\mathbf{R}$ is the projection matrix, applied to the calculated wavefield $\mathbf{p}$ so that $\mathbf{Rp}$ gives the calculated wavefield at source and receiver locations only, $\mathbf{x}_s$ and $\mathbf{x}_r$ respectively, at every time sample t. s is time, τ is a time lag and τ_0 is the travel time shift that is a function of time and space for small window size ω for each source $\mathbf{x}_s$.

Travel time Born FWI uses reflection energy to generate deep low wavenumber update and increase depth penetration compare to those of conventional FWI. It uses travel time based objective function for the relaxing of the constrains on an accurate starting earth model, therefore improving chances of convergence to the global minimum. It is important to note, that TT Born FWI requires the data to be pre-processed prior to the inversion. It is crucial for the data to be multiple-free as propagator for Born modelling does not predict the multiples. It is important to have a good starting reflectivity model as the reflectors must be placed at the correct depth in the earth model (Irabor et al. 2016).

Chapter 3

Datasets

3.1 Pivot dataset

3.1.1 Location and geology

The Pivot dataset was acquired in the northern part of the Carnarvon basin (The Northern Carnarvon Basin) on the north-west Australian shelf (Fig. 3.1). It consists of 7 regional seismic lines. In this thesis, all the results are shown for one line only.

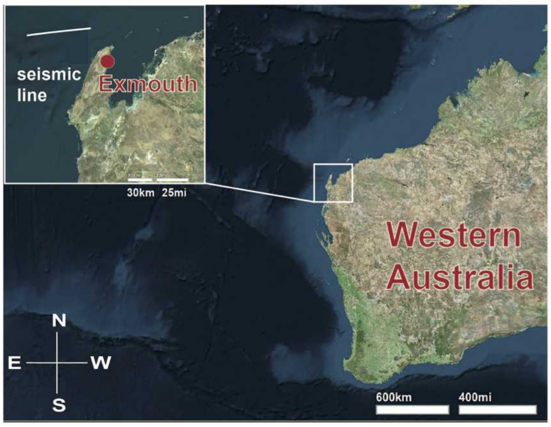


Figure 3.1: Location of the Pivot dataset.

The geological evolution and petroleum geology of the north-west Australian shelf is extensively described in the proceedings of three Petroleum Exploration Society of Australia symposia edited by Purcell et al. (1988, 1994, 1998). I will give a short overview of the geology based mainly on these works.

This basin is regarded as the premier hydrocarbon basin of Australia as its numerous oilfields and gasfields demonstrate the petroleum potential of the region, particularly offshore (Purcell et al. 1994). It is a sedimentary basin that covers a total area of about $5,000,000\,\mathrm{km}^2$ (*Western Australia's Petroleum and Geothermal Explorer's Guide* 2014) with water depths up to 3500 m (*Carnarvon Basin* 2019). It has been explored since 1953 (Longley et al. 2002) and the exploration history is presented in great detail by Purcell et al. (1994).

The basin initiated during break-up of Pangea (Black et al. 2017). It has a post-rift sedimentary architecture and was developed during four successive periods of extension and thermal subsidence (Purcell et al. 1998), which resulted in accumulation of up to 15 km of sedimentary infill (Black et al. 2017). Laterally extensive faults oriented towards the north-northeast divide the area and define a series of structural sub-basins (*Carnarvon Basin* 2019).

Figure 3.2 displays the seismo-chronostratigraphic chart of the area comprising of tectonic events following sequence stratigraphic nomenclature and constrained seismic interpretation (Black et al. 2017). The well $Falcone1A$ displayed on the seismic section is located approximately 10 km north from the seismic line used in this study. The well is not shown in the thesis because geological sequences change its depth quite significantly over this 10 km which makes the comparison to an FWI model of the studied line difficult. The stratigraphy of the area is composed of Triassic to Early Cretaceous silicate delta and marine deposits and Mid-Cretaceous to Cenozoic depositions dominated by carbonates. The carbonate sediments were deposited as a set of north-oriented pro-grading wedges as this region was cooling and subsiding (Purcell et al. 1988). This resulted in the deep burial of the underlying Mesozoic source and reservoir rocks in the inboard part of the basin. The majority of hydrocarbons are stored within the upper Triassic, Jurassic and Lower Cretaceous sandstones with the regional Early Cretaceous seal (*Carnarvon Basin* 2019).

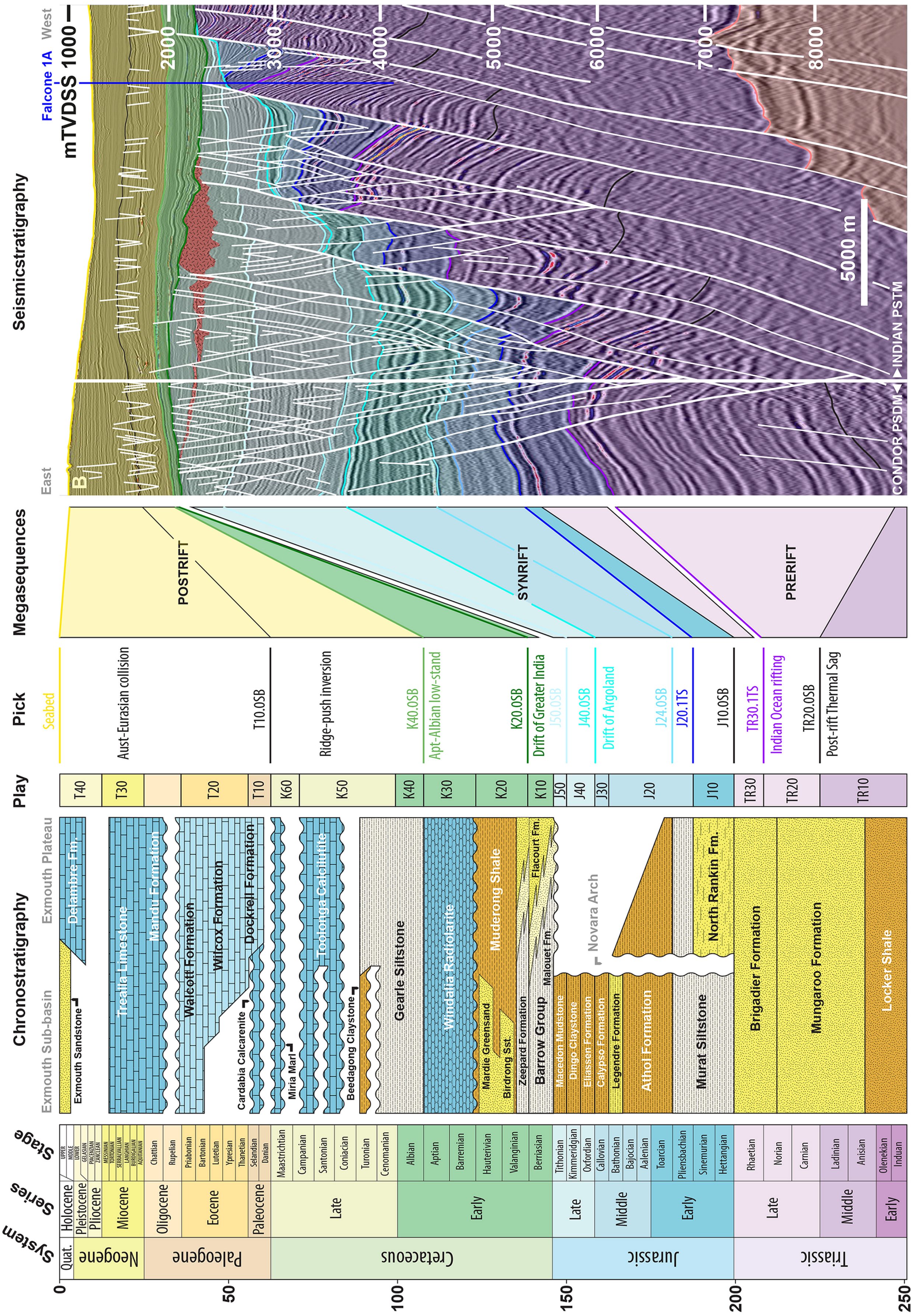


Figure 3.2: Seismo-chronostratigraphic chart for the southern part of the Northern Carnarvon Basin. The well Falcone-1A is located 10 km north from the regional seismic line in the Pivot dataset (after Black et al. (2017)).

3.1.2 PSDM reflection section

The PSDM section from the Pivot dataset was provided by the original contractor and is shown in Figure 3.3.

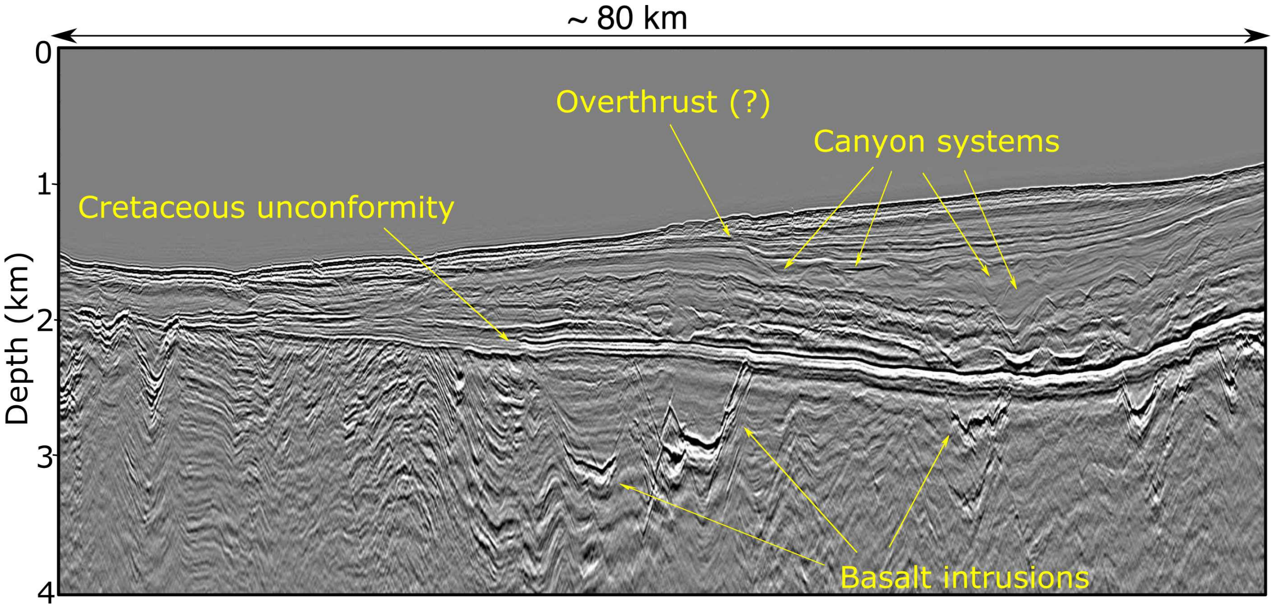


Figure 3.3: PSDM section from the Pivot dataset provided by the original processing contractor. The water depth varies from 600 m to 1600 m.

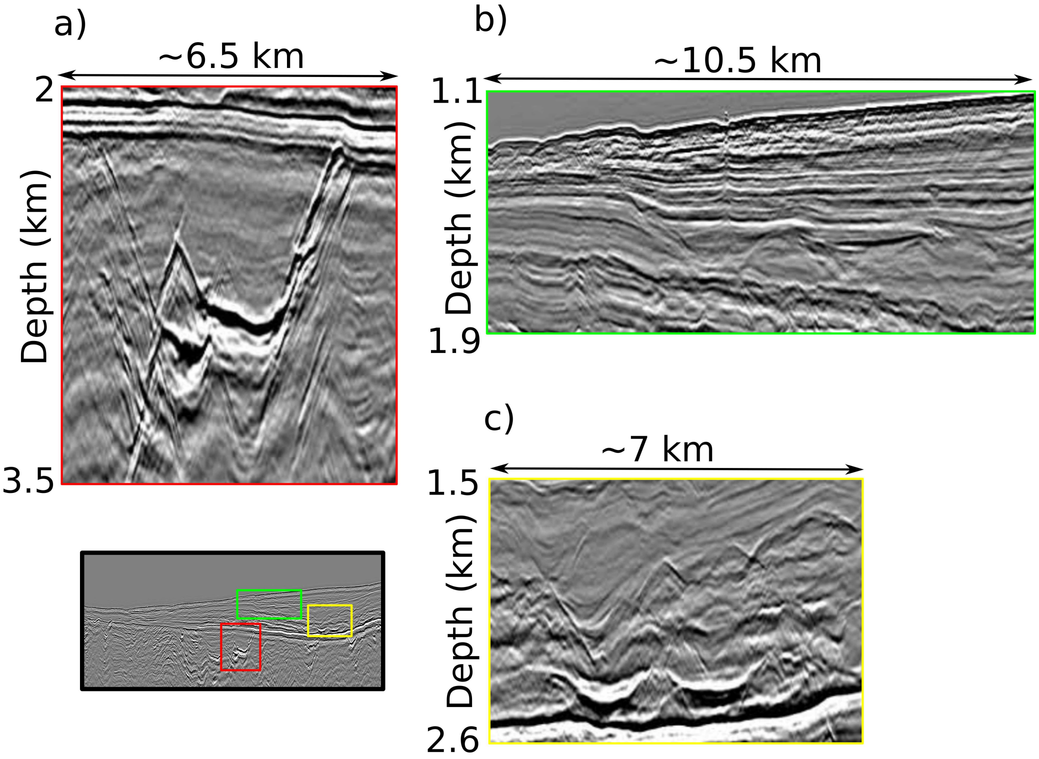


Figure 3.4: Close up of Figure 3.3 showing (a) basalt intrusion, (b) overthrust and sand channels and (c) large canyon system.

It demonstrates the relatively complex geology of the area with two major geological areas, separated by a major Cretaceous unconformity. It includes pre-Cretaceous deposits with intrusion bodies, which are likely to be basalt intrusions, large tilted fault blocks and Cretaceous and post-Cretaceous deposits forming multiple canyon systems and valleys (Longley et al. 2002). In the shallow central part of the section a relatively large displacement can be observed, this is likely to be an overthrust. The water depth in the study area varies from 600 m to 1600 m.

Figure 3.4 shows three close ups from the PSDM section. Figure 3.4(a) displays two bright reflectors at the depth of approximately 3 km, these are probably associated with igneous intrusions. Figure 3.4(b) demonstrates the shallow depth area in the central part of the section with lateral extension of approximately 10 km. In this area two sand channels and the overthrust can be identified. Figure 3.4(c) shows the area just above the major Cretaceous unconformity in the right-hand side of the PSDM section. Here, the mini-basin can be observed; it is probably not a sand channel due to its relatively large width of approximately 3 km. Two bright reflectors immediately below this mini-basin are also quite noticeable features.

3.1.3 Acquisition parameters and shot records

A fairly deep-water 2D narrow-azimuth towed streamer dataset was acquired in 2013. The survey consisted of 7 regional 2D seismic lines. In this thesis only one line is presented.

The line of nearly 80 km length was acquired using a single towed streamer and a single source. Data acquisition was optimised for FWI by employing a long 10 km deep-tow cable, towed at 25 m depth, and a large low frequency source array towed at 10 m depth. The acquisition configuration is shown in Figure 3.5.

Note, that having such a long cable may lead to cable feathering due to the tidal effects and the vessel turning or changing direction. In this particular case, the cable feathering along the line is on average around 400 m and is quite significant at the very beginning of the line, where both long and short offsets are 700 m off the planned line track. This can be easily seen from the diagram presented in Figure 3.6. It demonstrates the receiver distribution in space, where short offsets are highlighted in red and the long offsets are highlighted in blue; the shooting direction is indicated by an arrow. Consequently, the geometry of this 2D line is not exactly 2-dimensional, especially at the very beginning of the line.

The source array consisted of 4 gun arrays with a separation of 10 m between them; each gun array had 11 airguns, arranged in a group 14 m long group (Fig. 3.5). The nominal source spacing was 50 m and was regular throughout the survey; the receiver separation was 12.5 m. No low-cut filter was applied in the recording system.

In total, the line contains 1480 shot points with 804 channels per shot and has 1,189,920 receiver points. All of the sources and receivers were retained for the inversion. The acquisition parameters are summarised in Table 3.1.

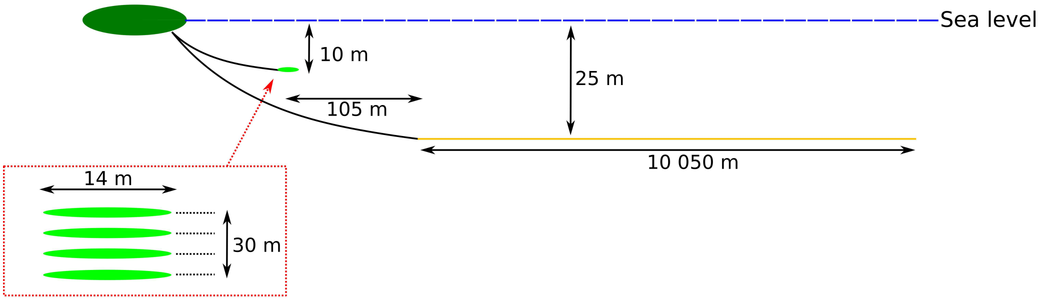


Figure 3.5: Towing offset diagram for the Pivot dataset.

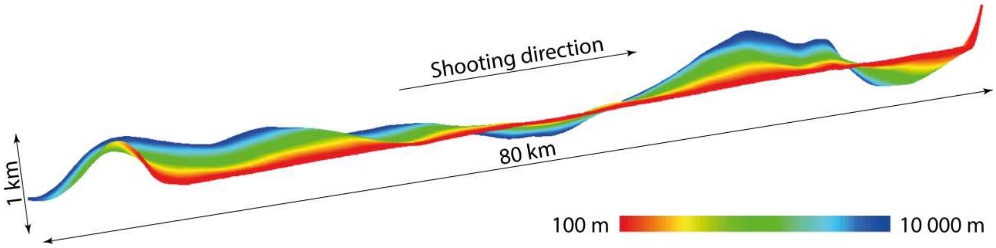


Figure 3.6: Feathering diagram for the Pivot dataset. Note the significant cable feathering in the very beginning of the line.

Typical raw shot gathers are shown in Figure 3.7. The data have some strong very low-frequency noise that can be seen across all offsets, on every shot. This noise is potentially associated with the background ocean swell and the cable tugging and must be removed prior to the inversion as it will not be simulated during forward modelling. The data also have strong first-order water-bottom multiple, significant amount of beneficial refraction energy at long offsets, and relatively low noise levels.

	Actual parameters	Standart parameters
Streamers active length	10050 m	
Nearest offset	105 m	
Furthest offset	10142.5 m	6000 m
Number of shots	1480	
Number of channels per shot	804	
Total number of receiver points	1,189,920	
Group interval / Length	12.5/12.95 m	
Fold	100	
Distance between CDP's	6.25 m	
Record Length	18.432 s	
Sample rate	2 ms	
Streamer Depth	25 m	7 m
SP Interval	50 meters	
Array Volume	5040 cu.in	2000 cu.in
Array Depth	10 m	5 m
Nominal pressure	2000 psi	

Table 3.1: Acquisition parameters for the Pivot dataset and typical parameters for a 2D towed streamer dataset. Parameters optimised for FWI are highlighted in purple.

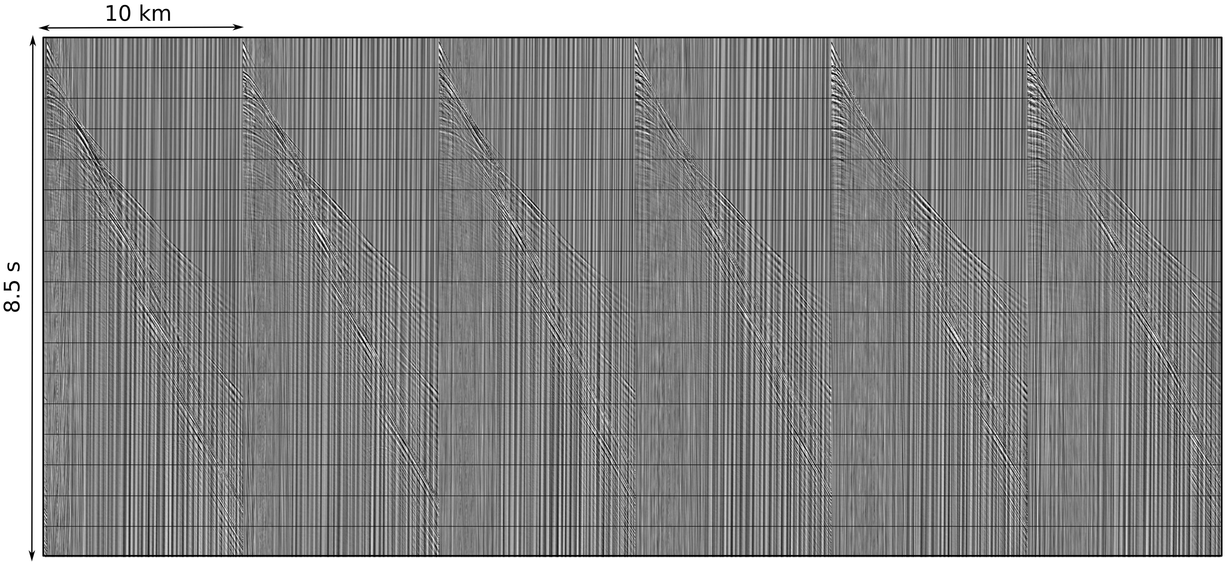


Figure 3.7: Raw shot gathers from the Pivot dataset.

The spectrum of the raw data is displayed in Figure 3.8. There is a significant amount of low-frequency energy below 2 Hz which is due to the cable-tugging and ocean swell. The amplitudes drop from about 8 to 3 Hz, but then start to rise again from 3 to 2 Hz. The drop is very likely to be the effect of source and receiver ghosts which remove low frequencies. The amplitudes rise below 3 Hz is very likely ambient noise which tends to increase at low frequencies. The amplitude drops at 30 and 75 Hz are likely to be the source and receiver notches. To summarise, the Pivot dataset contains low frequencies, long offsets and diving waves, all that makes it favourable for conventional FWI.

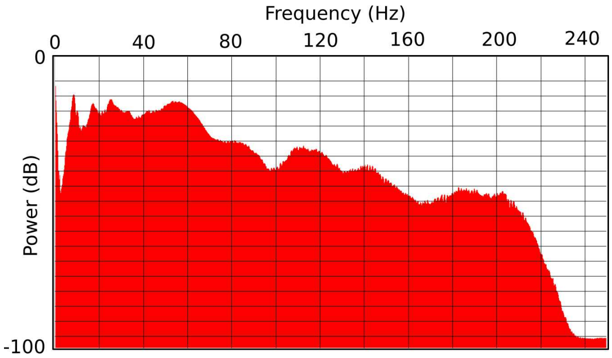


Figure 3.8: Amplitude spectrum of the raw data from the Pivot dataset.

3.2 Gabon dataset

3.2.1 Location and geology

The Gabon dataset was acquired in the Gabon basin in the middle portion of the West African passive continental margin (Fig. 3.9).

Figure 3.9: Location of the Gabon dataset. The dataset is indicated by the solid white line; the whole seismic survey is outlined in white.

The Gabon basin is divided into the three sub-basins: the North Gabon basin and the South Gabon basin, separated by the N'Komi Fracture zone, and the Interior basin, each of which has a different geological history (Brownfield et al. 2006). Figure 3.10 shows the main tectonic elements of offshore Gabon. The Gabon dataset extends across the southern part of the North Gabon Basin and the north-western part of the South Gabon Basin.

The Gabon basin originated during the rifting phase as the Atlantic opened and developed during the post-rifting phase (Tamannai et al. 2013). Figure 3.11 shows the stratigraphy of the study area. No pre-rift deposits are known to be present in both the South and the North Gabon basins (Tamannai et al. 2013). Syn-rift deposits are characterised by shales and sandstones and the post-rift sediments form the basin's sedimentary fill (Brownfield et al. 2006). These deposits range from Aptian to Holocene in age and are characterised by shelf clastics and carbonates with pro-gradational units along the continental margin as well as transgressive units (Brownfield et al. 2006).

The area is bisected by a major NNE-SSW trending strike-slip fault system and the N'Komi fracture zone (Tamannai et al. 2013). Whilst the eastern part of the Gabon basin is expected to contain a relatively thin layer of salt, underlain by lower Cretaceous rift and sag sequences, the western half of the area is thought to lie beyond the limit of the Aptian salt and to comprise of the Aptian marine shales directly overlying basaltic oceanic crust (Brownfield et al. 2006).

The exploration of the Gabon basin has started in the 1920's (Tamannai et al. 2013). Although production offshore has not yet been extended to the deeper water, both pre- and post-salt deposits are found to be perspective (Tamannai et al. 2013).

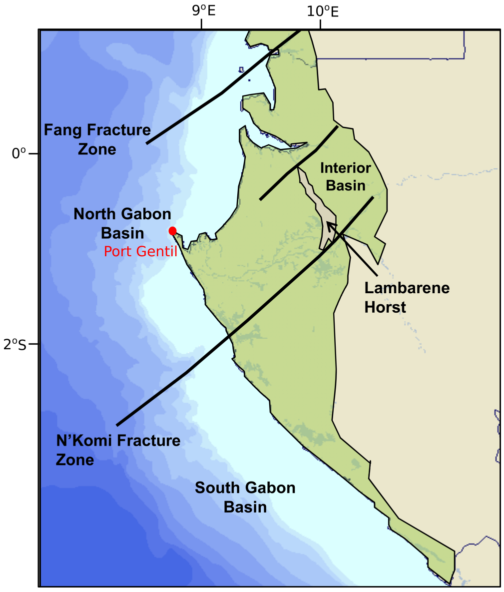


Figure 3.10: Tectonic elements of offshore Gabon. Modified after Tamannai et al. (2013).

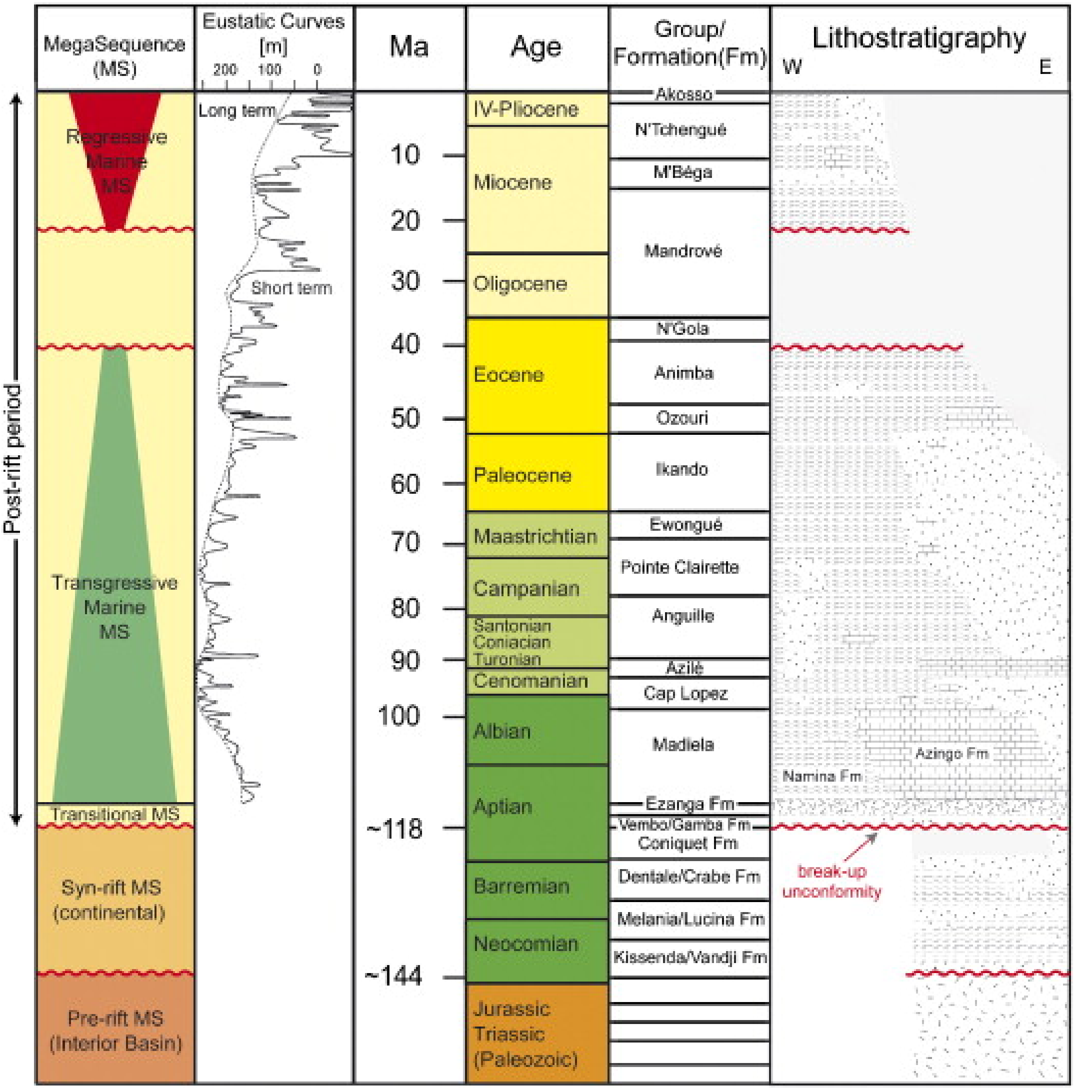


Figure 3.11: Stratigraphy of offshore Gabon (from Dupre et al. (2007)).

3.2.2 PSDM reflection section

The original contractor has provided geological interpretation of a time section from the study area. This PSTM section without the interpretation is displayed in Figure 3.12. It shows 8 s of two-way travel time with black being an increase in acoustic impedance. The lateral extension is approximately 50 km.

The same section with the geological interpretation is shown in Figure 3.13. It contains distinct faulted competent blocks and four main zones of lithology and structural uncertainties:

1. The main uncertainty in lithology indicated as (1) in Figure 3.13. The structure at the depth of approximately 4.5 s has a distinctly different seismic response from the

surrounding deposits. This structure may be: (1) carbonate deposits, (2) mass flow deposits or (3) volcanic sediments. These options will lead to principally different decisions during exploration. Carbonate deposits have high porosity and good permeability which makes them good reservoir; mass flow deposits have high porosity and high permeability which is a perfect reservoir; volcanic deposits, on the contrast, can not be a reservoir at all. Both carbonate and mass flow deposits have low velocity, whereas volcanic deposits have high velocity. Consequently, FWI should be able to differentiate potential reservoir characterised by low velocity from non-reservoir rocks characterised by high velocity and prevent unnecessary drilling in case this is volcanic deposits.

2. The next uncertainty at the depth of approximately 3 s is indicated as (2) in Figure 3.13. These are transparent facies just below the major unconformity and have soft seismic response. To be able to identify these facies the velocity information is essential.

3. At the depth of 5.5 s indicated as (3) and located below the pink horizon in Figure 3.13, there are transparent facies with inversion features. It is believed that there might be salt present in this area, however, it is not certain without additional analysis.

4. In the deep part of the section, at the depth of about 7 s, there are two strong bright reflections indicated as (4) in Figure 3.13. These reflections can be observed throughout the section at approximately the same time of 7 s. These might be intrusive bodies with very high velocities, typically more than 4000 m/s. However, conventional FWI would struggle to image this target as there is no penetration by transmitted arrivals at such depth.

At the bottom, on the right-hand side in Figure 3.13, there is a map showing the location of the interpreted time section in relation to a seismic line studied in this thesis. The interpreted PSTM section is indicated in red and it intersects the line from the Gabon dataset highlighted in blue. Note, that the interpreted PSTM section is oriented practically perpendicular to the studied seismic line. Consequently, multiple faults in the deeper part of the section that separate it into blocks will not be visible on the studied line.

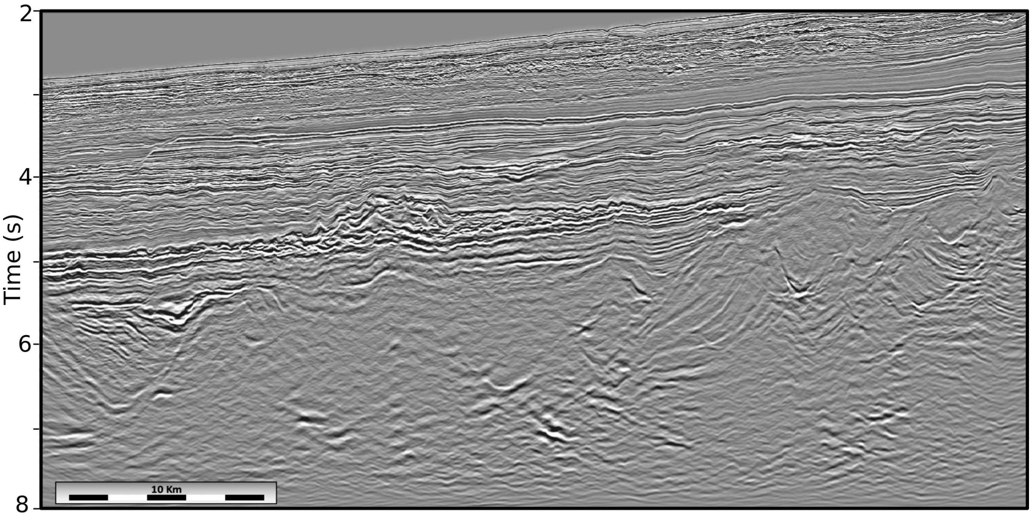


Figure 3.12: PSTM section from the Gabon dataset. Black denotes an increase in acoustic impedance.

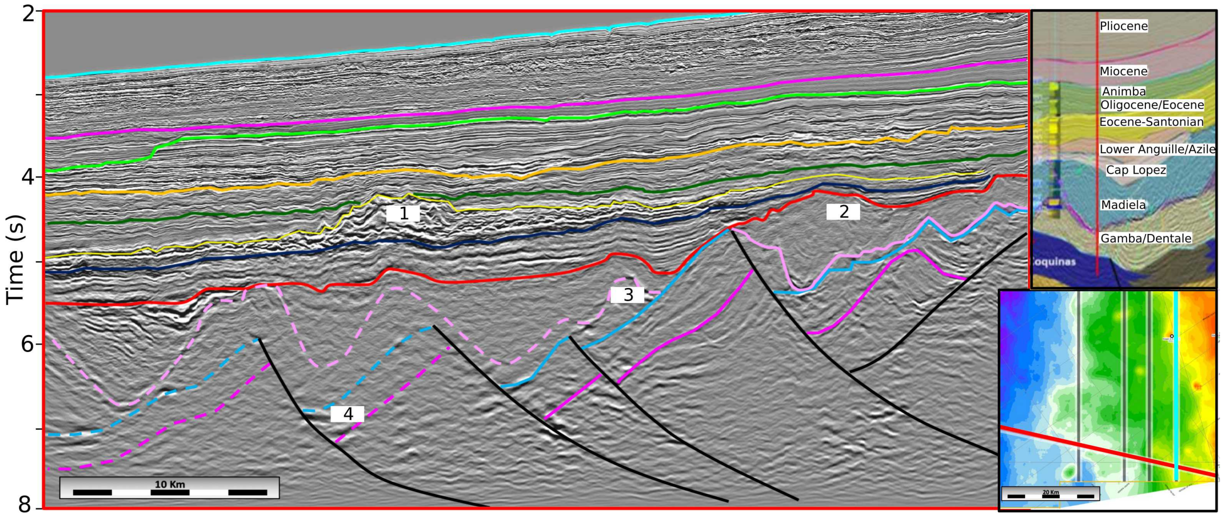


Figure 3.13: PSTM section from the Gabon dataset with the contractor's geological interpretation. Lithology and structural uncertainties: (1) carbonate, mass flow or volcanics? (2) what are transparent facies rugose at the unconformity and with a soft seismic responce? (3) could this be salt? (4) are these deep intrusives?

The seismic line for the Gabon dataset displayed in depth is presented in Figure 3.14. This depth section was obtained from the PSTM contractor's seismic volume cube converted to depth using stacking velocities provided by the processing contractor. Figure 3.14 displays the section with maximum depth of 7 km showing the primary target of FWI circled in yellow, which is the main uncertainty on lithology and is crucial to be identified for exploration purposes.

Figure 3.15 shows close-ups from this section. In the green box, a shallow part of the data immediately below the seabed with the depth ranging from 1 to 2 km and lateral extension of approximately of 12 km is displayed. There are multiple shallow gas pockets throughout the area. They are potential drilling hazards and should be identified during FWI. In the red box, the feature circled in yellow in Figure 3.14 is displayed. It is approximately 600 m in height and 7 km in length; it has a chaotic seismic response with bright reflections. It is a potential reservoir and, therefore, is a primary target of FWI.

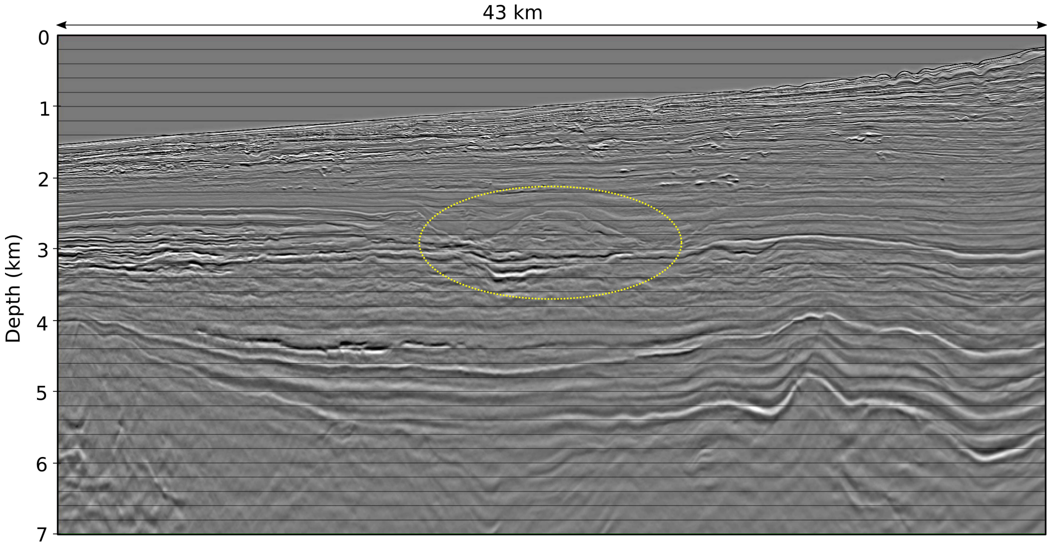


Figure 3.14: PSDM section from the Gabon dataset. The priority area is circled in yellow: carbonate, mass flow deposit or volcanics?

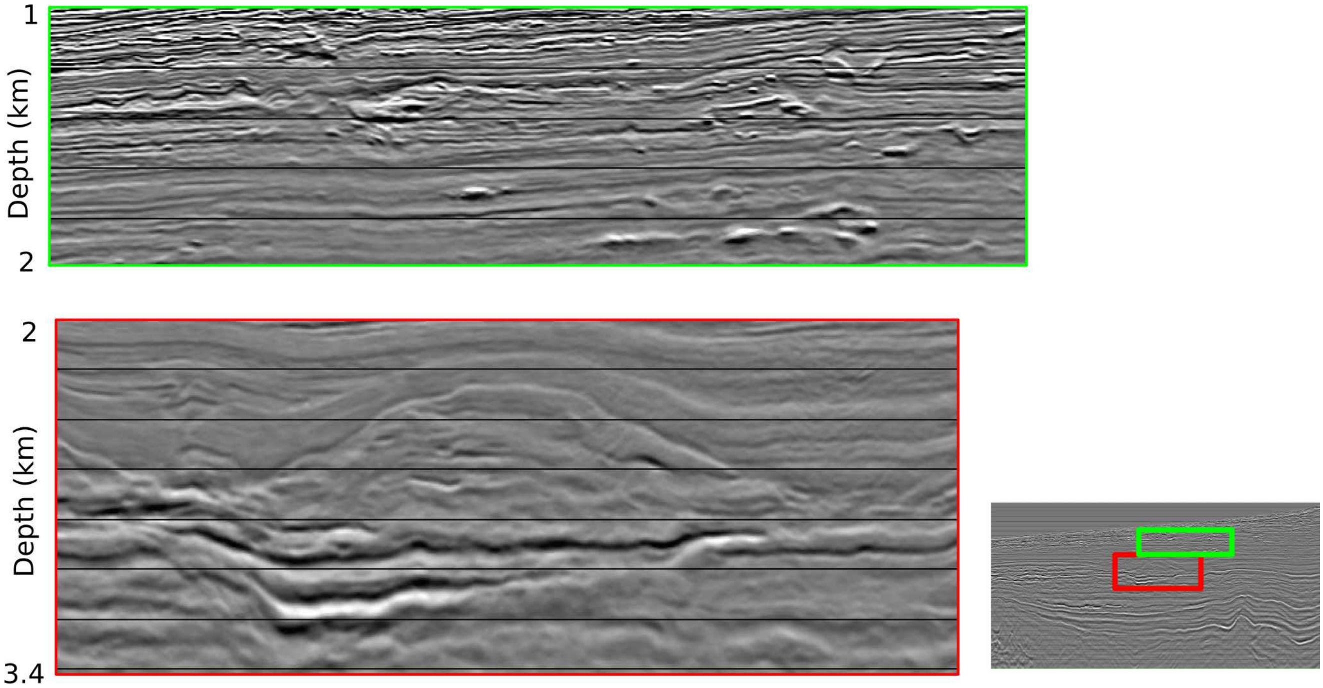


Figure 3.15: Close up of Figure 3.14. In the green box - multiple shallow gas pockets, which are potential drilling hazards, should be identified by FWI. In the red box - the primary target of the survey, uncertainty on lithology - carbonate, mass flow deposit or volcanics?

3.2.3 Acquisition parameters and shot records

A 3D narrow-azimuth towed-streamer seismic dataset was acquired offshore Gabon in November 2014 - February 2015. The scope of the survey was approximately 2285 km^2 and was situated in waters from about 180 m to 2300 m depth, corresponding to two-way travel times of approximately 150-3000 ms, with the shallowest region of the survey being the North-Eastern corner. In this thesis, a single swath from the survey was used. This swath covered an area of 86 km^2 with the dimensions of 2 x 43 km. Figure 3.16 indicates the location of the swath in the dataset and shows the water depth within the area.

Data were acquired using 10 relatively long streamers towed at 14 m depth, each of 8100 m length, and a dual source airgun array towed at 8 m depth. The two airgun arrays, separated at 50 m, operated flip-flop with the interval of 25 m. Each airgun array had 3 sub-arrays and 33 airguns in total. The acquisition configuration and the airgun array configuration is shown in Figure 3.17.

The total amount of shots for the swath was 1922 and there were 12,454,560 receiver points. The acquisition parameters are summarised in Table 3.2.

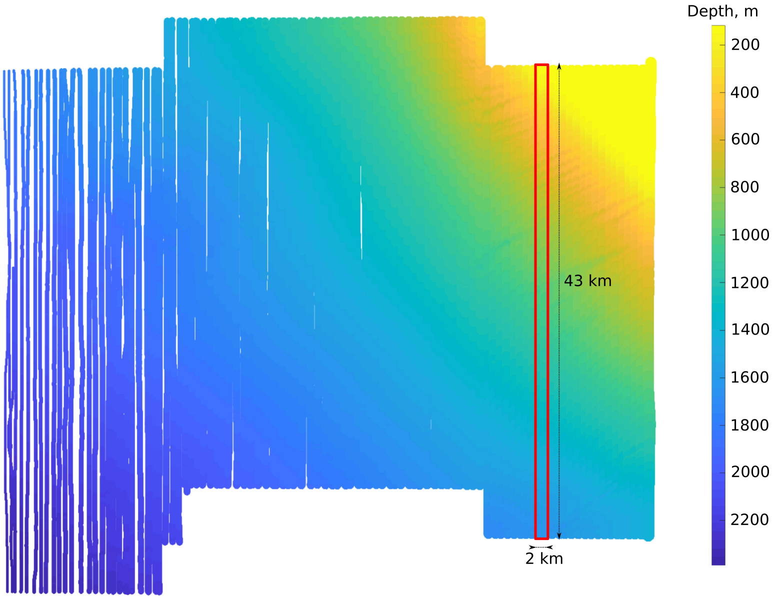


Figure 3.16: Bathymetry map for the Gabon dataset. The single swath is indicated by the red rectangle.

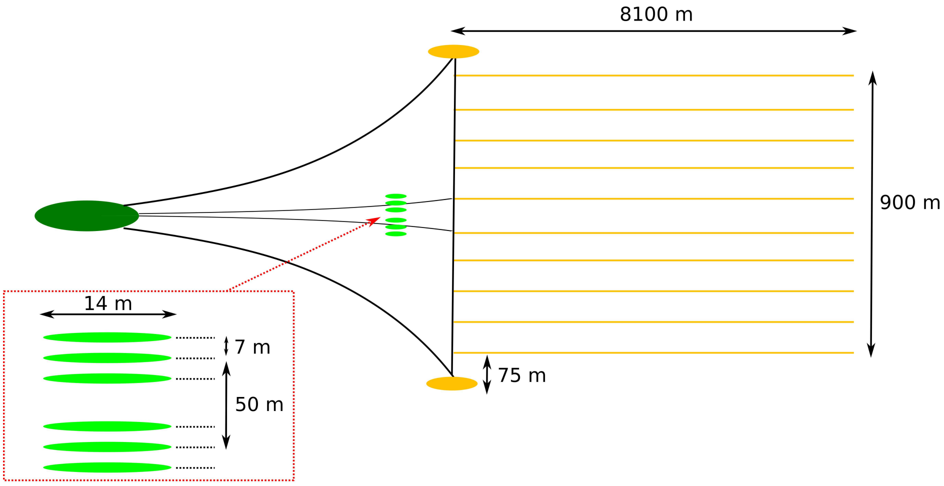
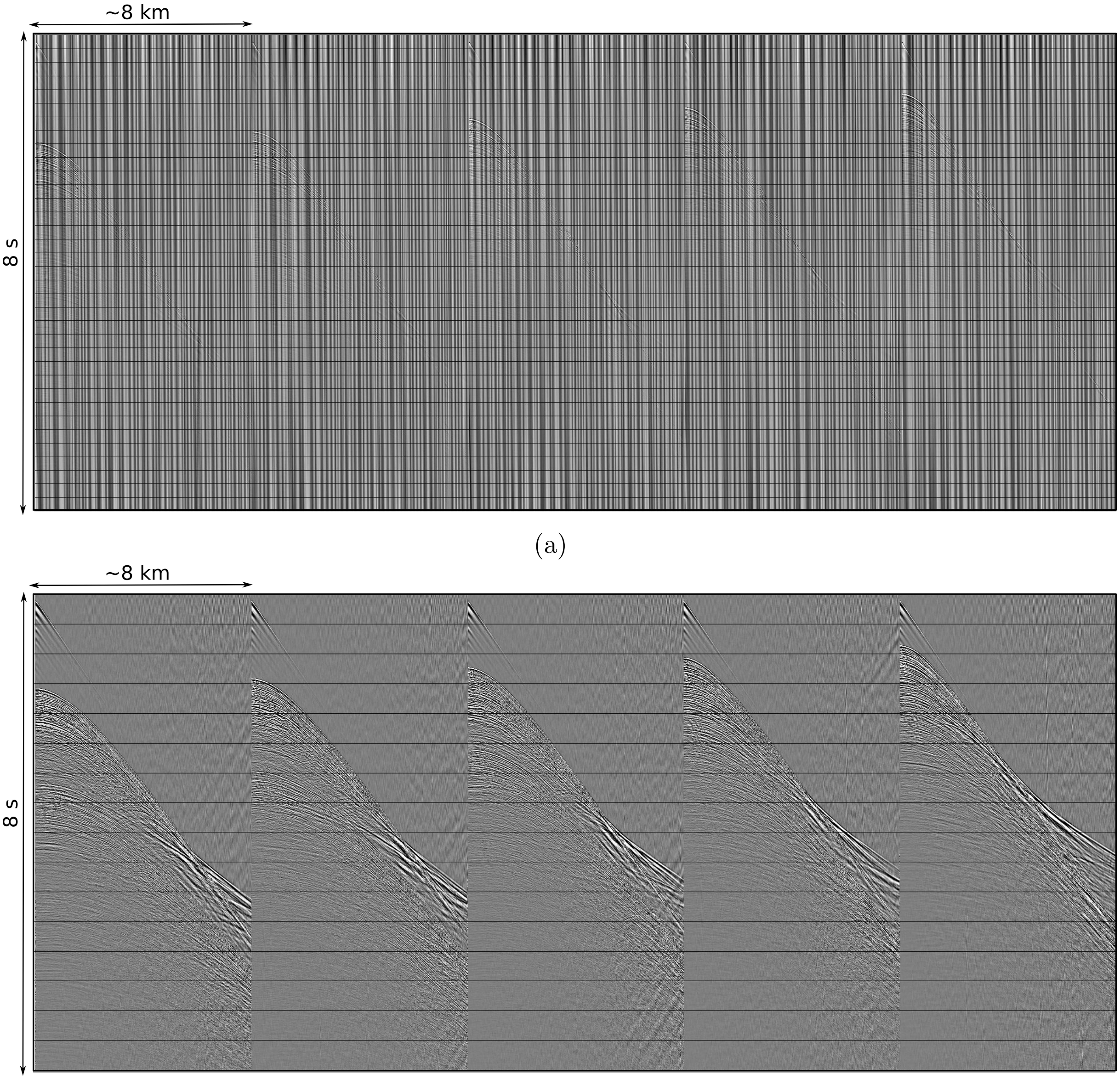


Figure 3.17: Acquisition configuration for the Gabon dataset. Drawing not to scale.

Cable length	8100 m
Number of cables	10
Distance between cables	100 m
Number of channels per cable	648
Group interval	12.5 m
Cable depth	14 m
Source depth	8 m
Number of gun arrays	2
Total source capacity	2 x 3480 cu.in
Gun array separation	50 m
Shot point interval flip flop	25 m
Number of shots	1922
Trace length	8 s
Trace sampling	2 ms

Table 3.2: Acquisition parameters for the Gabon dataset.

A few raw shot gathers from several shots are shown in Figure 3.18(a). These are five different shots but recorded on the same cable positioned in the middle of the 10 streamer arrray. The data are dominated by very low-frequency "stripy" noise associated with the background ocean swell and cable-tugging. The same shot records filtered using a minimum-phase high-pass Ormsby filter rolling off from 2.4 to 3.0 Hz are shown in Figure 3.18(b). The data are reflection dominated, contain strong refraction energy at long offsets, have distinct first order multiples and low-frequency noise present at the tail of the streamer. This noise can be seen starting at behind the streamer and dipping back in the opposite direction to the seismic data; it is not present throughout the survey, however, can be observed at some shots.



(a)

(b)

Figure 3.18: Shot records from the Gabon dataset (a) raw and (b) filtered to retain data above 3 Hz.

The amplitude spectrum of the raw data is displayed in Figure 3.19. It has high-amplitude ultra-low frequencies below 3 Hz which are associated with the cable-tugging and ocean swell. The notches from source and receiver ghosts should be seen at 93 Hz and 54 Hz respectively, but are observed at slightly higher frequencies due to offset variations.

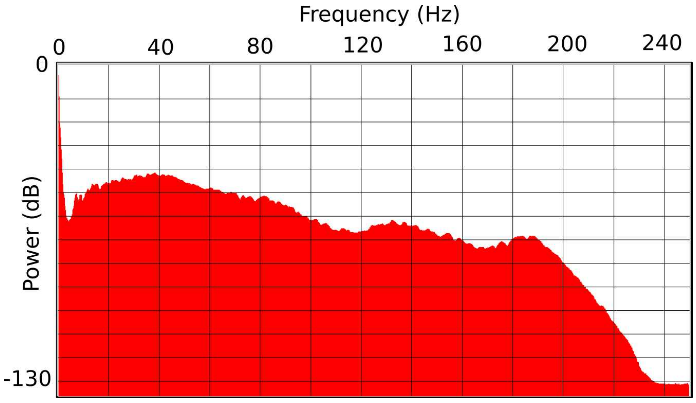


Figure 3.19: Amplitude spectrum of the raw data from the Gabon dataset.

At the first sight, this dataset seems to be suitable for FWI, where FWI could serve three purposes:

- Identify the velocity of the primary target at the depth of approximately 3 km, and therefore establish whether this target is a reservoir or not
- Improve shallow velocity model for further migration
- Identify the residual gas pockets, potentially eliminating drilling hazards

However, this dataset has proven to be difficult as two groups have tried to invert it before and failed.

3.3 Picanha dataset

3.3.1 Location and geology

The Picanha dataset extends across the southern part of the Campos Basin offshore Brazil where the water depth is over 2,000 m. The location of the dataset and bathymetry is displayed in Figure 3.20. The geology of the area is described by a vast number of publications, including those of Castro et al. (2016) and Mohriak et al. (1990). Here, I will present a brief geological introduction based on the work of Mohriak et al. (1990).

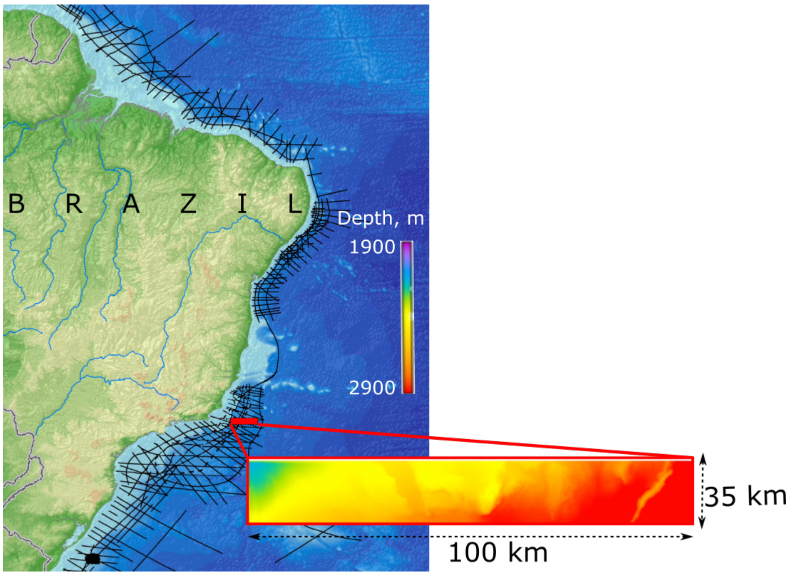


Figure 3.20: Location of the Picanha dataset and bathymetry map.

The basin originates from the break up of Gondwana in the Mesozoic and covers an area of $100,000\,\mathrm{km}^2$. It is one of the most oil prolific basins in Brazil.

The stratigraphy of the study area is composed by Neocomian (Lower Cretaceous) clastic sediments deposited on basalts, followed by a sequence of evaporitic rocks that have undergone intense diapiric activity. The post-salt sequence consists of Albian/Cenomanian (middle Cretaceous) limestones, marls and shales deposited in deep water environment. The post-salt deposits are generally cross-cut by listric normal faults formed in response to the salt movements. From the Upper Cretaceous to present-day the sedimentary sequence is composed by marine clastic sediments. The stratigraphic column for the Campos basin is displayed in Figure 3.22.

The main reservoirs are distributed throughout all the stratigraphy of the basin, from pre-salt fractured basalts and limestone formations to post-salt limestone and sandstone deposits (Mohriak et al. 1990).

3.3.2 PSDM reflection section

The PSDM section across the area is shown in Figure 3.23 with black denoting an increase in acoustic impedance. It is a deep water dataset with the depth in range from 2 to 3 km.

The section demonstrates the presence of salt with three distinct salt diapirs penetrating the post-salt sediments to the depth of 3 km. Immediately above the salt, a very strong reflector represents the top of carbonate deposits. Above the salt and carbonates, post-salt sedimentary layers are penetrated by multiple faults throughout the section, forming mini-grabens and horsts. On the left-hand side of the section, there is a bright reflector at the depth of approximately 3 km. This is an unconformity that pinches out to the left hand-side of the section. Overall, the geology of the region is characterised by complex overburden and presence of salt and carbonates.

The same PSDM section overlaid with the salt flood velocity model is shown in Figure 3.21. Both were provided to me during the internship. It can be seen, that the carbonate deposits have the highest velocity with the maximum of approximately 5890 m/s, whilst based on well logs the salt velocity was assumed to be 4500 m/s. Above these high velocity sediments, the model is approximately one-dimensional and contains no additional structures in it.

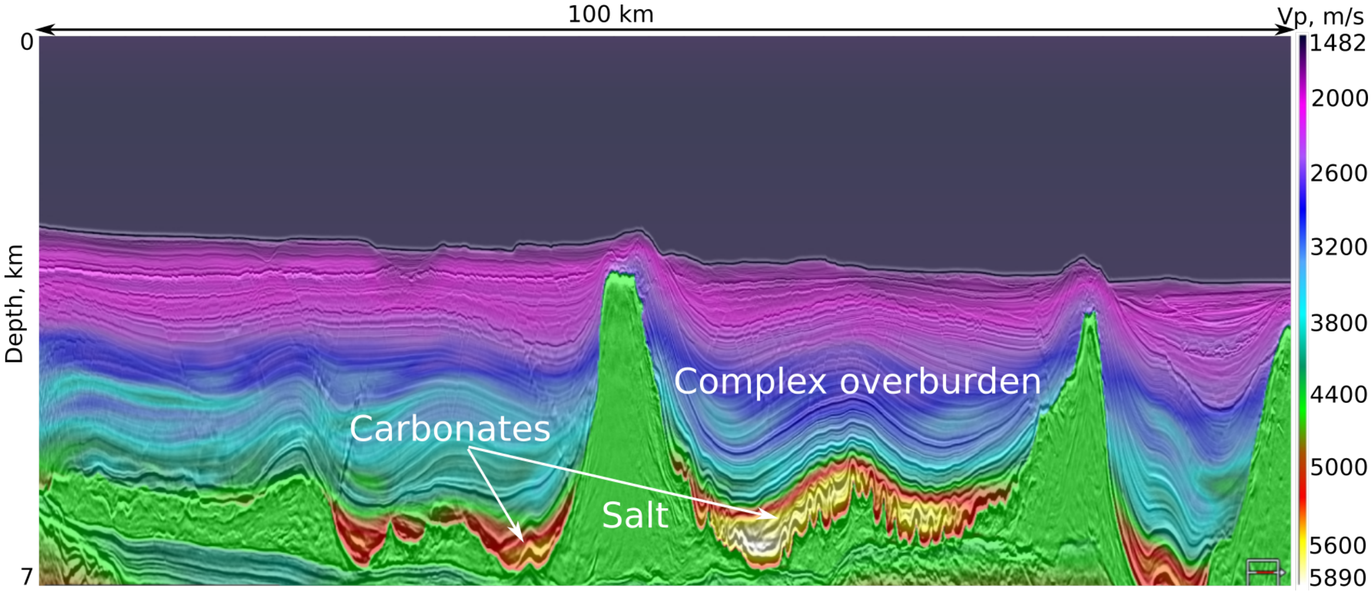


Figure 3.21: PSDM section for the Picanha dataset overlaid with velocities based upon tomography. A constant of 4500 m/s is used for the salt deposits.

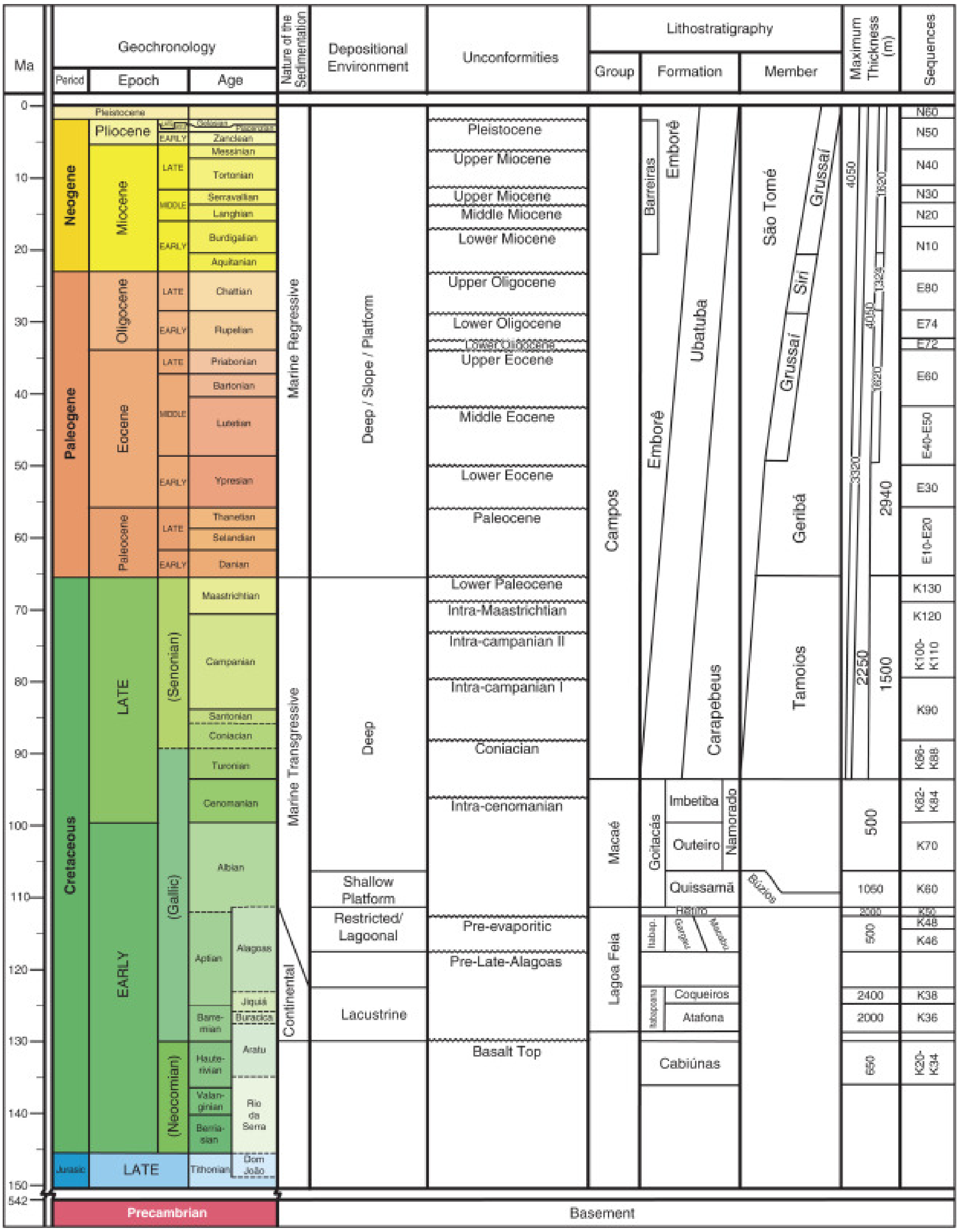


Figure 3.22: Stratigraphy of the Campos basin (from Castro et al. (2016)).

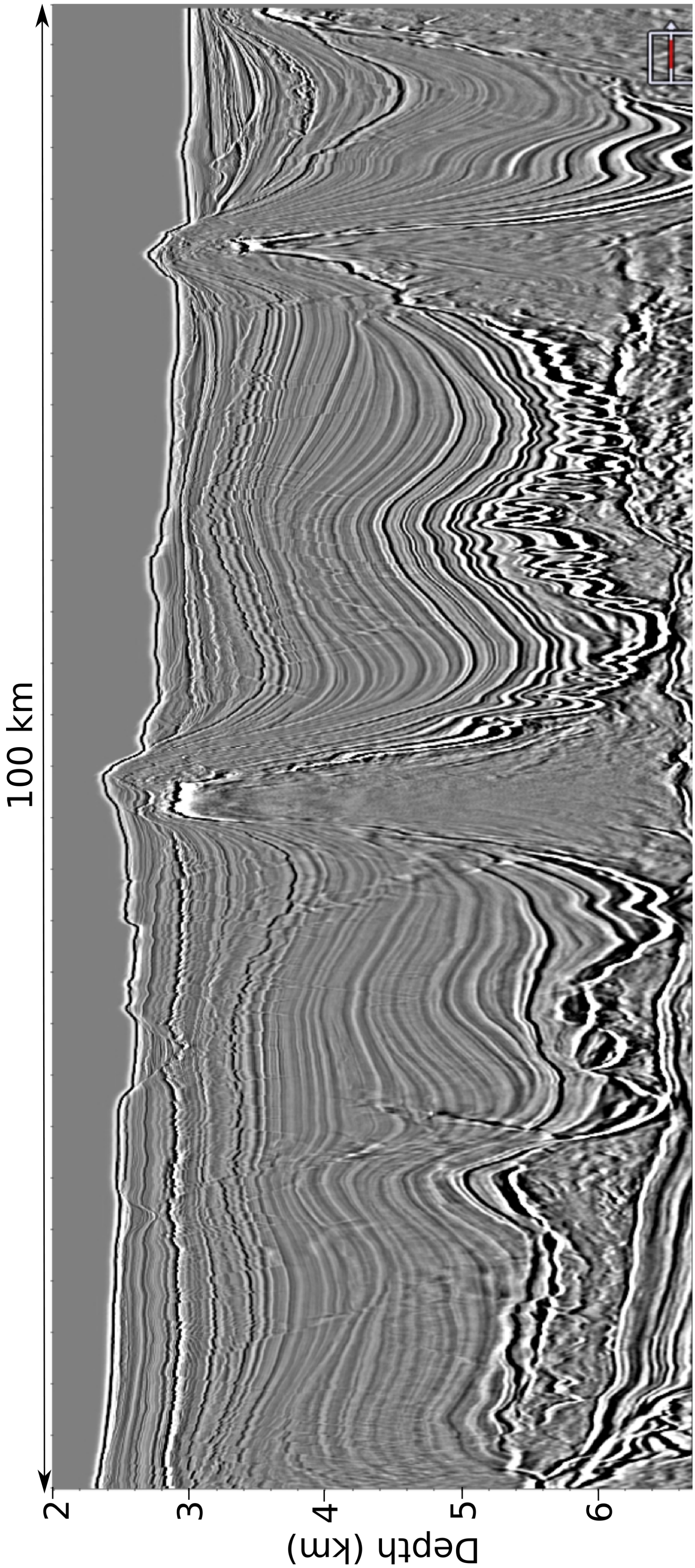


Figure 3.23: PSDM section for the Picanha dataset.

3.3.3 Acquisition parameters and shot records

A deep water 3D narrow-azimuth seismic dataset was acquired offshore Brazil in January-July 2000 by PGS. The survey covers an area of $12,000\,\mathrm{km}^2$ and was shot using CLO (continuous long offset) configuration shown in Figure 3.24.

CLO was a solution to create long offset acquisition for deep reservoir illumination without the need of long towed streamers. It was done using two vessels. The master vessel was towing a source at 6.5 m depth and 10 short streamers of 4 km length each at 8 m depth. The streamer length is half the desired spread length. The second, remote vessel was towing a source at 6. 5m depth and was positioned one streamer length ahead. Both vessels had two gun arrays. The gun arrays for each vessel were operating flip flop with a shot point interval of 37.5 m. The firing order between the four gun arrays is shown in Figure 3.24. Shot records for both vessels were recorded, one from the near vessel, one from the far, to create a long offset shot. No de-blending techniques were applied. CLO timing summary is presented in Table 3.3. Time sampling when recording was 2ms. Acquisition parameters are summarised in Table 3.4. CLO configuration improves streamer control and positioning and increases operational efficiency as less time is required to deploy the equipment and turn the vessels. However, given the short delay between the near and far gun arrays firing, there is a possibility of energy from the near shot on the far gun's record with significant amplitude, especially water layer multiple bounces. In this study we used a single swath with a total number of 5070 shots, covering an area of $100\,\mathrm{km}^2$.

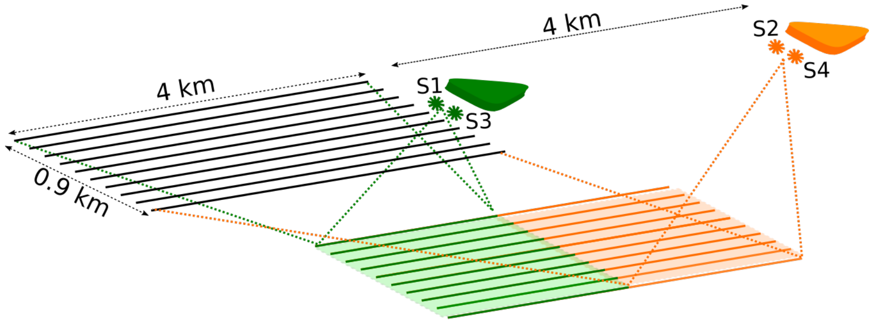


Figure 3.24: Acquisition configuration for the Picanha dataset. The seismic data with offsets below 4 km were acquired with the sources from the streamer vessel (green subsurface tracks) and the long offsets from 4 to 8 km were recorded with the remote vessel (orange subsurface tracks).

Time,s	
0	S1 fires
6	S2 fires
9	End of shot record for S1
15	End of shot record for S2

Table 3.3: CLO timing summary for the Picanha dataset. Four gun arrays are operating flip flop with a shot interval of 37.5 m in the order indicated in Figure 3.24. The lengths of shot record for each gun is 9 s. Due to the short delay between near guns (S1 and S3) and far guns (S2 and S4 respectively) some far shots contain energy from water bottom multiples from the near shots.

Cable length	4050 m
Number of cables	10
Distance between cables	100 m
Number of channels per cable	324
Group interval	12.5 m
Cable depth	8 m
Source depth	6.5 m
Number of gun arrays	4
Gun array separation	50 m
Shot point interval flip flop	37.5 m
Number of shots	5070
Trace length	9 s
Trace sampling	2 ms

Table 3.4: Acquisition parameters for the Picanha dataset.

No raw shot gathers from this datasets were available. Instead, the raw data were pre-processed to two different stages and were provided as two separate sets. The pre-processing is described in chapter 7.

The reason behind using two separate datsets was that different FWI types require different pre-processing, for instance conventional FWI utilises the full wavefield whereas reflection FWI requires the data to be de-ghosted and de-multipled as Born approximation does not predict these events during forward propagation. Therefore, in order to apply different FWI types to the same field dataset, it must be appropriately pre-processed for each FWI type.

A few shots from the first pre-processed dataset are displayed in Figure 3.25(a); following procedures were applied to this dataset:

- linear de-noise
- de-bubble operator
- low-pass filtration using a zero-phase Ormsby filter of bandwidth 0-2-8-12 Hz
- muting ahead of direct arrivals

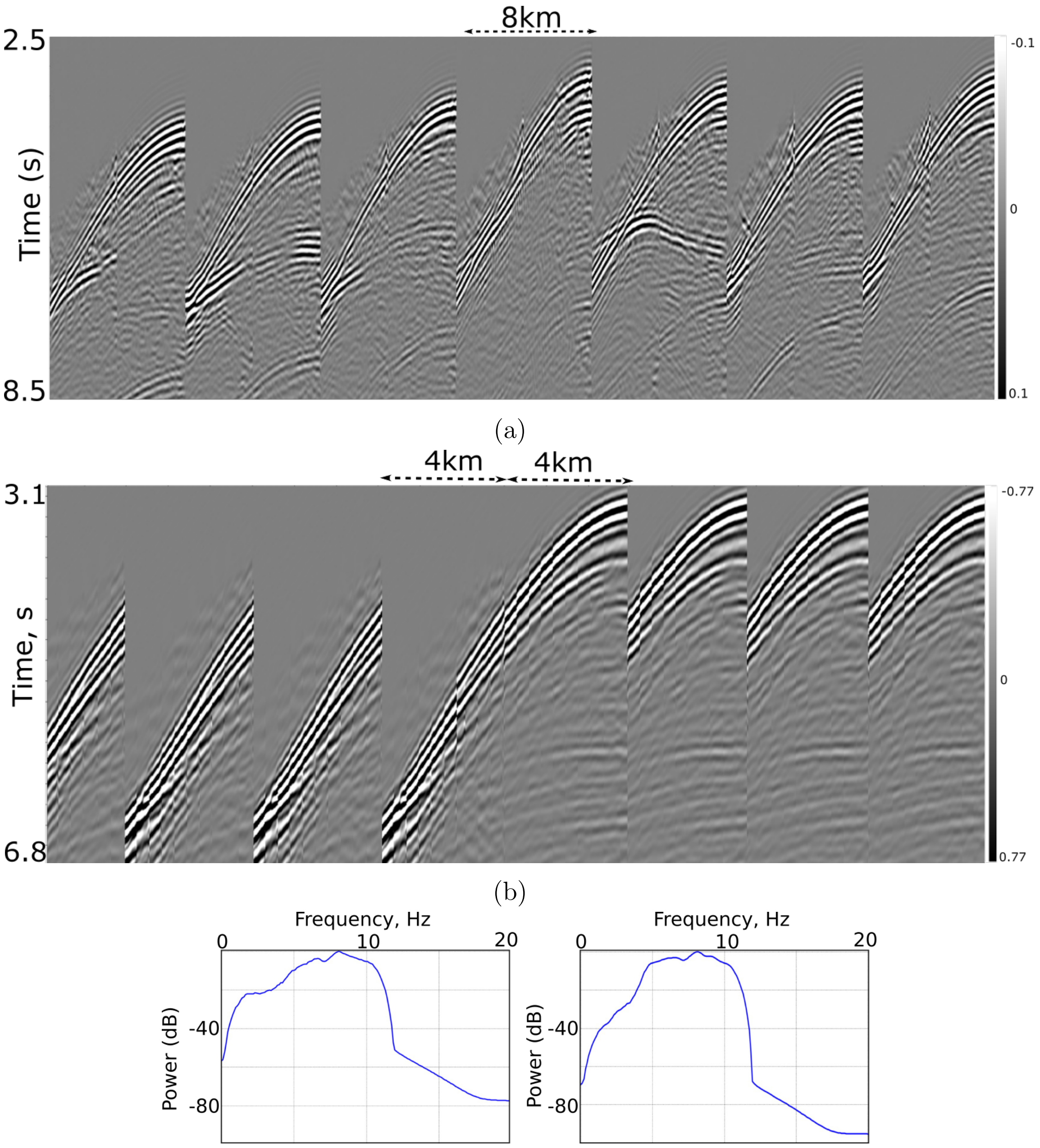
A few different shots from the second pre-processed dataset are displayed in Figure 3.25(d) and following procedures were applied to it:

- linear de-noise
- de-bubble operator
- mute ahead of direct arrivals
- de-multiple
- de-ghosting
- zero-phasing
- 4D regularisation (described in chapter 7)
- low-pass filtration using a zero-phase Ormsby filter of bandwidth 0-2-8-12 Hz

It can be seen that the data are reflection-dominated with no refraction energy present, contain weak first-order multiples and have low noise level.

Note, that due to CLO acquisition, Figure 3.25(a) shows not 7, but 14 shots gathers. The near and long offsets are acquired separately, which doubles the amount of shots and, consequently, increases the cost of the inversion. Figure 3.25(b) shows “super shots” created by 4D regularisation, which effectively reduces the number of shots by replacing several sources with one “effective” source. 4D regularisation is described in chapter 7.

Figure 3.25(c) and Figure 3.25(d) demonstrate the spectra of the volume 1 and volume 2 respectively. They have similar shape but differ at the low frequencies. This is caused by the different pre-processing applied to the data as de-multiple, de-ghosting, zero-phasing and 4D-regularisation affect both amplitude and phase spectra of the data.



(a)

(b)

(c)

(d)

Figure 3.25: Shot gathers from two differently pre-processed volumes for the Picanha dataset: (a) Set 1 and (b) Set 2; the amplitude spectra for (c) Set 1 and (d) Set 2.

3.4 Summary

To summarise, in this thesis the capability of conventional FWI when it is applied to marine deep-water reflection-dominated NATS field data was explored using different datasets. I applied FWI to three datasets: the first used a deep-towed 10-km cable and was specifically acquired for 2D FWI (the Pivot dataset); the other two datasets were both 3D reflection-dominated surveys to which FWI had been previously applied with limited success - these datasets were from Gabon (the Gabon dataset) and Brazil (the Brazil dataset), and were chosen specifically because FWI had been previously tried and had failed. A short summary is presented in Table 3.5.

NATS dataset	Pivot	Gabon	Picanha
Type	2D	3D	3D
Cable depth	25 m	14 m	8 m
Array depth	10 m	8 m	6.5 m
Maximum offset	10 km	8 km	8 km
Acquisition filters	Analogue low-cut 3 Hz	Analogue low-cut 2 Hz	Information not available

Table 3.5: Summary of the datasets.

Chapter 4

2D vs. 3D FWI

Three-dimensional anisotropic acoustic FWI has become a relatively routine component of depth-model building for PSDM and shallow hazard identification, and is increasingly being used for pore-pressure-prediction and reservoir characterisation. However, 3D FWI is relatively resource intensive, especially at higher frequencies and when extensive initial testing is required for parameter selection, data preconditioning and quality control of starting models. During early exploration, full 3D coverage may not be available everywhere, and long 2D lines are commonly used to tie new 3D surveys to more-distant wells and provide regional context. Consequently, 2D FWI may have a role to play as part of a larger 3D FWI workflow.

4.1 Rational

3D FWI allows the use of velocity models that vary in three dimensions, but it also allows sources and receivers to be properly distributed in space and it allows those sources and receivers to act as points rather than the lines that 2D wave propagation assumes. The latter changes both the amplitude and phase spectra of the predicted data and influences its temporal decay. Consequently, 3D simulation provides different results and is superior to simple 2D modelling of the same data. Some of these differences can be mitigated during FWI and some cannot.

Most algorithms that are used to simulate the seismic wavefields required for FWI scale as n^3 in 2D and as n^4 in 3D for a single source where n is a measure of the linear dimensions of the model in mesh points. Furthermore, 3D modelling requires additional source coverage in the third dimension so that in 3D the number of sources required scales as n^2 whereas in 2D it scales as n. Consequently the computational cost of 3D FWI is around n^2 times greater than in 2D FWI; typically n is a few hundred.

In practice, depending in part upon the acquisition geometry, it is often possible to reduce the source density in 3D FWI below that required in pure 2D so that this scaling is not quite as severe as suggested, but 3D FWI is always significantly more expensive than 2D. In addition, as the maximum frequency of the data increases, the number of mesh points required to capture the wavefield accurately also increases so that the difference in cost between 2D and 3D FWI becomes more marked.

Here acoustic anisotropic FWI in both 2D and 3D was applied to the same 2D seismic dataset in order to explore the utility, accuracy and limitations of the former. I used the workflow described by Warner et al. (2013).

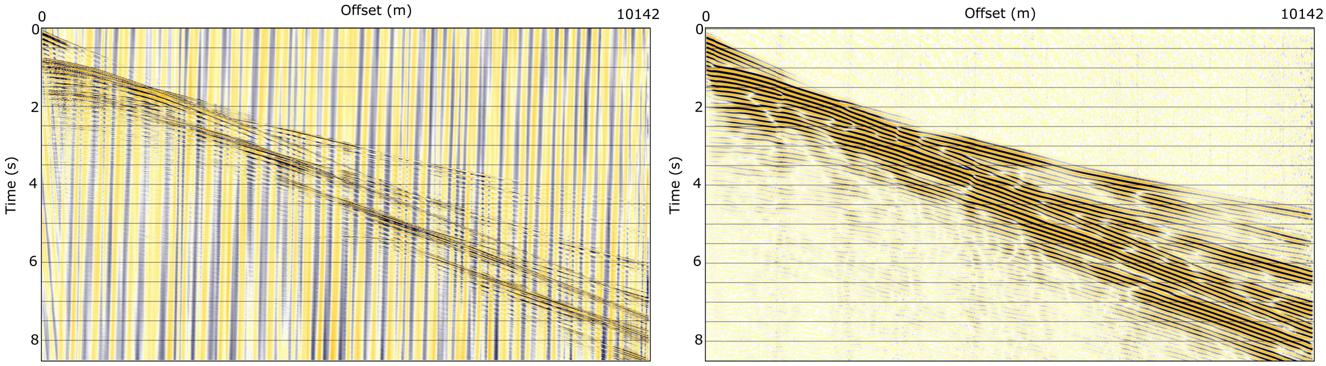
4.2 Pre-processing

The field data for this study was taken from the Pivot dataset from Carnarvon basin on the north-west Australian shelf. It is an 80-km 2D seismic line acquired with a single streamer and a single source; the description of the dataset is given in chapter 3.

Figure 4.1 shows a typical record. The data contain significant refracted energy at long offsets, significant low-frequency energy and strong very-low frequency noise associated with the background ocean swell and the cable tugging.

The pre-processing applied to the dataset was minimal and consisted of two passes of a minimum-phased band-pass filter: (1) low-cut filter to remove the very low-frequency noise below 2 Hz and (2) low-pass filter to retain frequencies below 12 Hz that is the maximum frequency used during the inversion. No other pre-processing procedures were applied.

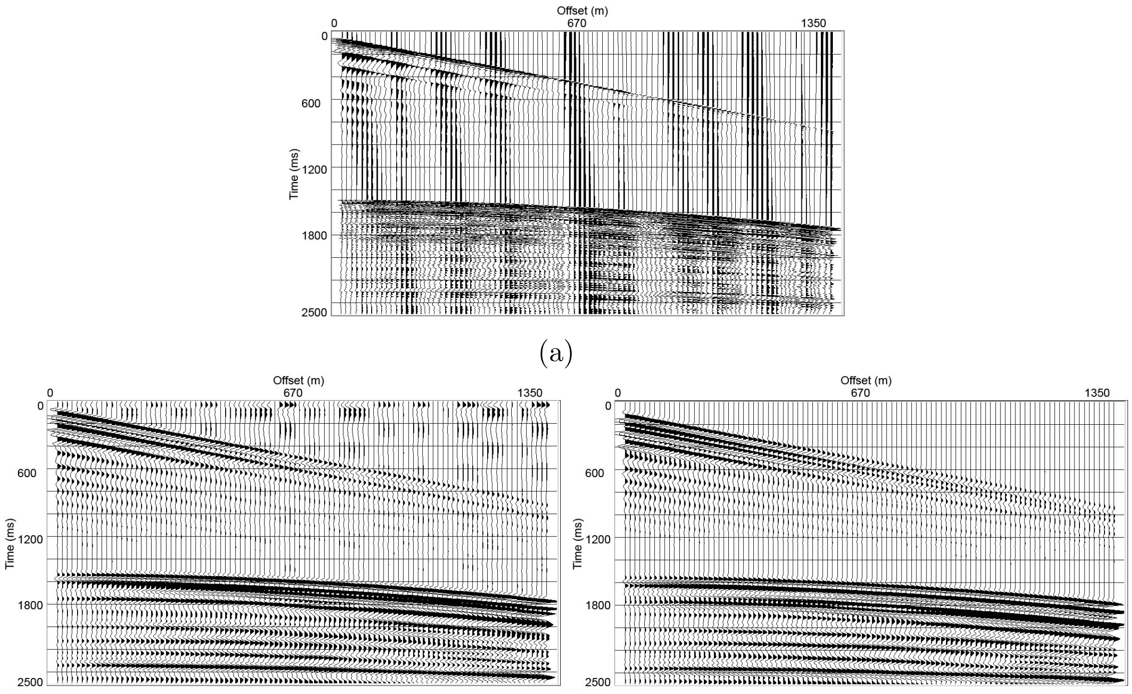
Most processing softwares implement band-pass filters by applying Fourier transform trace-by-trace, cutting off the desired frequencies and applying inverse Fourier transform. Because the noise is a signal that starts at zero time with a relatively large non-zero value, application of such filter causes a chaotic signal in the beginning of the trace. Therefore, eliminating such noise using a low-cut bandpass filter must be implemented with care should the chaotic noise be undesired. In principle, the chaotic noise produced by a filter can be simply muted out, especially in fairly deep water datasets as the noise will be ahead of the water bottom arrivals. However, sometimes it is desired to preserve clean direct arrivals, for instance, for source estimation. In such cases an extra step must be implemented to achieve the desirable outcome. The simplest solution is to pad the beginning of each trace prior to bandpass filtration with extra samples given, for example, by a symmetric projection of the trace. This allows the chaotic noise to be generated in the padded area only. After the filter has been applied, the padded area with the noise must be removed. I have implemented this technique and used padding of 2 s length. Figure 4.2 demonstrates first 2 s recorded by short offsets from a typical raw shot gather and the same shot with bandpass filter applied, with and without padding the trace length prior to filtration. It can be easily seen that padding prior the filtration provides superior noise removal compared to that of conventional filtration.



(a) (b)

Figure 4.1: Shot records (a) raw and (b) filtered below 12 Hz.

Figure 4.1(b) shows a pre-processed shot record of the field data went into the inversion. Both refracted and reflected arrivals were used throughout the inversion. The multiples and surface ghosts were retained within the data and were used at all stages of the inversion. Attenuation and elastic affects were not explicitly taken into account. However, $fullwave3D$ is designed to be robust against systematic amplitude variations in the field data that do not match acoustic assumptions.



(a)

(b)

(c)

Figure 4.2: Shot records (a) raw and band-passed: (b) without trace padding prior to the filter application, (c) with trace padding prior to the filter application. Shot record (b) has significant amount of chaotic noise ahead of the direct arrivals whilst shot record (c) has no noise ahead of the direct arrivals.

4.3 Choosing the starting frequency

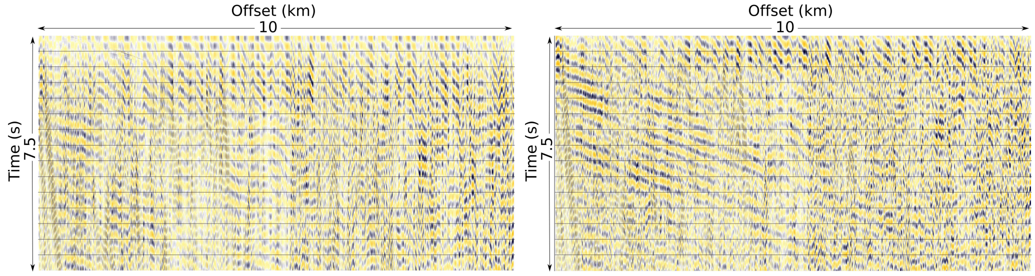
Presence of low frequencies in the field data is a key requirement for FWI implementation. Because FWI is a local iterative scheme, the starting velocity model must predict the data, that differs from the field data by no more than half a cycle. The lower the frequencies available in the data, the easier this condition is to meet.

Consequently, successful FWI ideally requires very low frequencies. Bunks et al. (1995) suggested the strategy to begin the inversion with the lowest frequency present in the field data. However, it is important that this frequency has signal-to-noise ratio at least 1:1 rather than large amplitudes. Ideally, signal-to-noise ratio should be even higher. If the selected starting frequency is too low, then the field data will contain too much noise that will be introduces in the results compromising the inversion quality. If the selected starting frequency is too high, then cycle-skipping may occur, potentially leading to the inversion failing or producing an erroneous model.

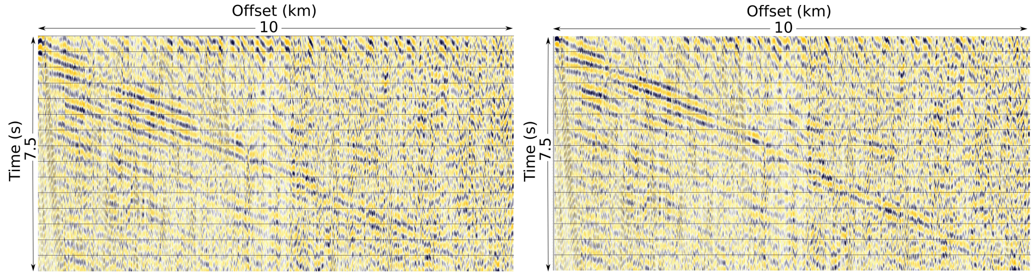
In practice, the starting frequency for FWI varies between 2 and 5 Hz, where the 2-3

Hz is a typical value for ocean bottom cable (OBC) datasets and the higher values are common for towed-streamer datasets.

A series of low-pass filters was applied to time-domain shot gathers to asses the low-frequency content of the field data. Figure 4.3 shows an example full-offset shot gather low-passed with an Ormsby filter below 2.5, 3.0, 3.5 and 4.0 Hz. The gathers are scaled trace-by-trace to allow visually evaluate signal-to-noise ratio at different offsets. At 2.5 Hz, the field data show some evidence of coherent signal, however, the data are dominated by both coherent and incoherent noise with very low signal-to-noise ratio, especially at mid and far offsets. At 3.0 Hz, the data clearly demonstrate presence of coherent signal at near and mid offsets with little evidence of signal at the very far offsets. At 3.5 Hz, coherent energy is present everywhere within the shot gather with signal-to-noise ratio similar to that at 3.0 Hz. By 4.0 Hz, it can be observed that the data has slightly higher signal-to-noise ratio with the coherent energy present along the full offset range.



(a) Filtered below 2.5 Hz (b) Filtered below 3.0 Hz



(c) Filtered below 3.5 Hz (d) Filtered below 4.0 Hz

Figure 4.3: An example full-offset shot gather filtered at (a) 2.5 Hz, (b) 3.0 Hz, (c) 3.5 Hz and (d) 4.0 Hz. The gathers are scaled trace-by-trace.

It is also good practice to evaluate the amplitude spectrum of the data. Figure 4.4 shows the amplitude spectrum of the same example shot gather low-passed with an Ormsby filter below 12 Hz. Frequencies below 2 Hz are practically absent in the field data. This is explained by using a low-cut Ormsby filter rolling off from 1.5 to 1.9 Hz during the pre-

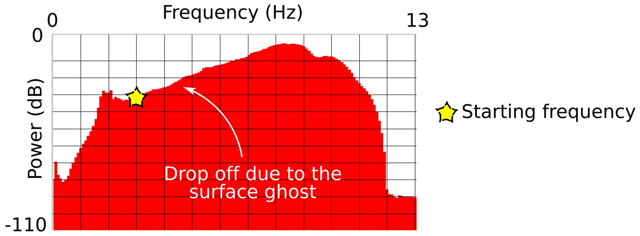


Figure 4.4: The amplitude spectrum of the low-passed below 12 Hz field data for the Pivot dataset. Drop off is due to the surface ghost. The starting frequency of 3 Hz is indicated by the yellow star.

processing with the purpose of eliminating the coherent high-amplitude noise. Amplitudes drop from 8 to 3 Hz, but then start to rise again from 3 to 2 Hz. The drop is very likely the effect of source and receiver ghosts which remove low frequencies, and the rise below 3 Hz is very likely ambient noise which tends to increase at low frequencies. The amplitude spectrum therefore also suggests that around 3 Hz is likely to be the optimal starting frequency. Because there was no low-cut filter applied during the acquisition, it is possible that even the frequencies lower than 2 Hz would be accessible for the inversion if there was no need to apply a low-cut filter during the pre-processing.

4.4 Starting models

4.4.1 Velocity and anisotropy

Successful FWI requires a sufficiently accurate starting velocity model. For the inversion scheme used in this thesis, all major arrivals in the modelled data, generated using the starting velocity model, must match the field data within half a cycle at the lowest inversion frequency to avoid a local minimum solution.

For 2D FWI

A 2D anisotropic velocity model generated for PSDM was used as the starting velocity model for 2D FWI. This model was provided by the processing contractor and was obtained using stacking velocities picked with a lateral interval of 1 km (160 CDP's). The contractor's model is relatively simple and broadly one-dimensional. It is shown in Figure 4.5(a). It is smooth and has no structures sharper than half of a wavelet at the lowest frequency used in the inversion, which is an essential requirement for the starting model, unless the structure is true in both depth and all other dimensions (Warner et al. 2013). The sea bed is the only sharp boundary present in the model. It is crucial for the inversion quality to have this reflector at the exact position in depth. Any inaccuracies can result in model being cycle-skipped and/or the inversion failing completely.

The velocity gradually increases with depth from 1495 m/s to 4300 m/s. The water velocity is constant everywhere with the value of 1495 m/s. The 2D model measures 80 x 4 km with a grid spacing of 25 m in both vertical and horizontal directions.

The Pivot dataset has a vertical transverse isotropic (VTI) model since there is no tilted symmetry axis or strong evidence to suggest azimuthal p-wave anisotropy. Horizontal and vertical velocities differ for VTI media and are related by Thompson's parameters, delta δ and epsilon ϵ as (Thomsen 1986)

$$V_p^{horizontal} = V_p^{vertical}\sqrt{1+2\epsilon} \tag{4.1}$$

This relationship is known as "pseudo-acoustic anisotropy". In geophysics, however, it is often shortened to "acoustic anisotropy".

The delta and epsilon models were provided by the processing contractor and are shown in Figure 4.5(b) and Figure 4.5(c). The dataset is highly anisotropic with the values of delta reaching 11% and the values of epsilon reaching 16%. Both models are relatively one-dimensional. However, there are noticeable lateral variation associated with the stratigraphy that are consistent between the two models.

For 3D FWI

Strictly speaking, the 3D FWI conducted in this study is essentially a 2.5-dimensional experiment due to the nature of the velocity and anisotropy starting models. However, I will refer to it as a 3D FWI because (1) the 3D forward propagator was used and (2) 3-dimensional acquisition geometry was properly taken into account.

The 2D starting models provided by the processing contractor and described in the

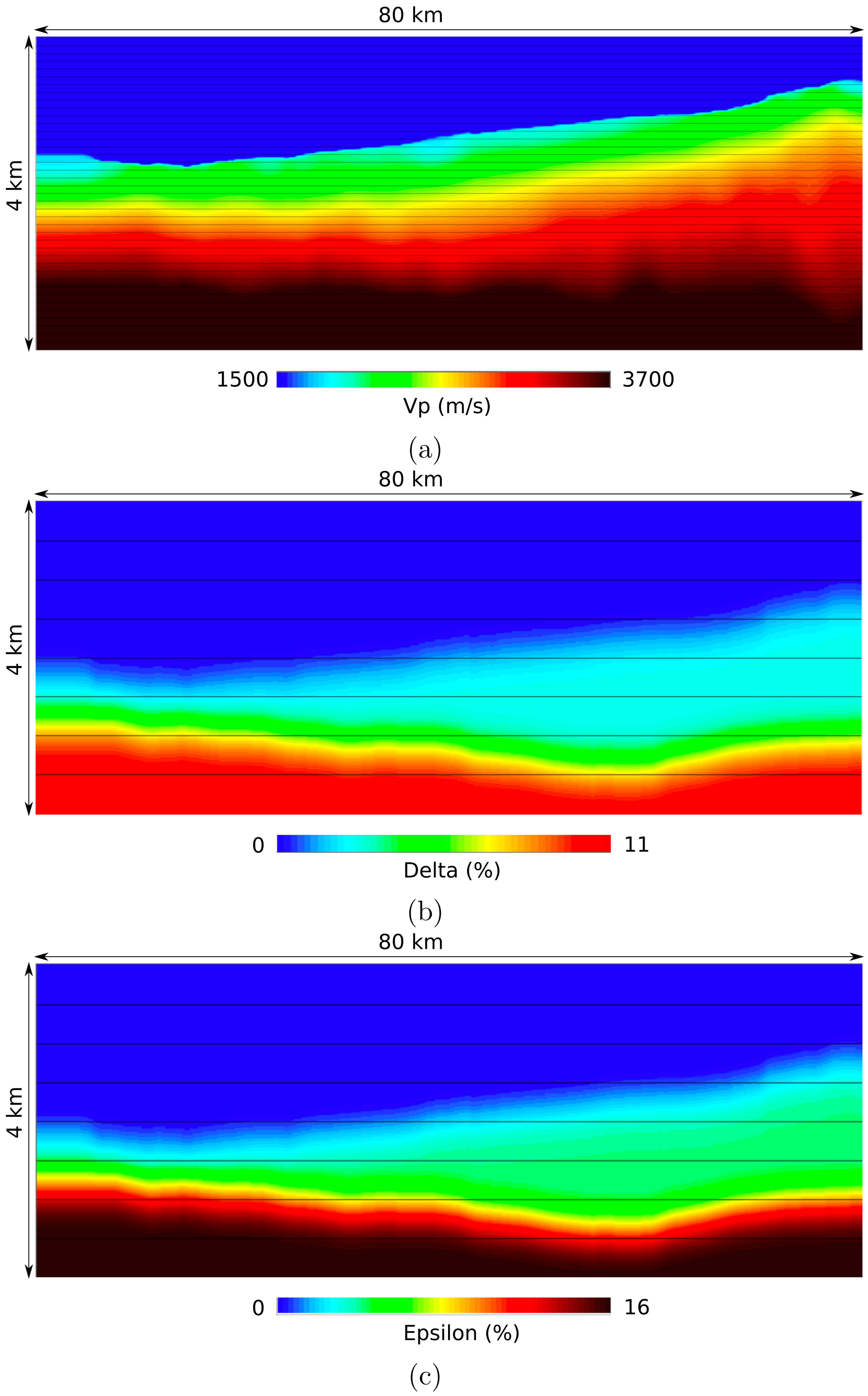


Figure 4.5: Starting (a) velocity, (b) delta and (c) epsilon models for the Pivot dataset.

previous section were used to generate the starting models appropriate for 3D FWI.

Figure 4.6 shows the true distribution of sources (red) and receivers (grey) in space mapped onto grid coordinates with a grid spacing of 25 m. Although the Pivot survey is a 2D seismic line, it is obvious that sources and, especially, receivers are not located in-line with each other and have non-constant y-coordinates. The black line denotes the 2D starting model

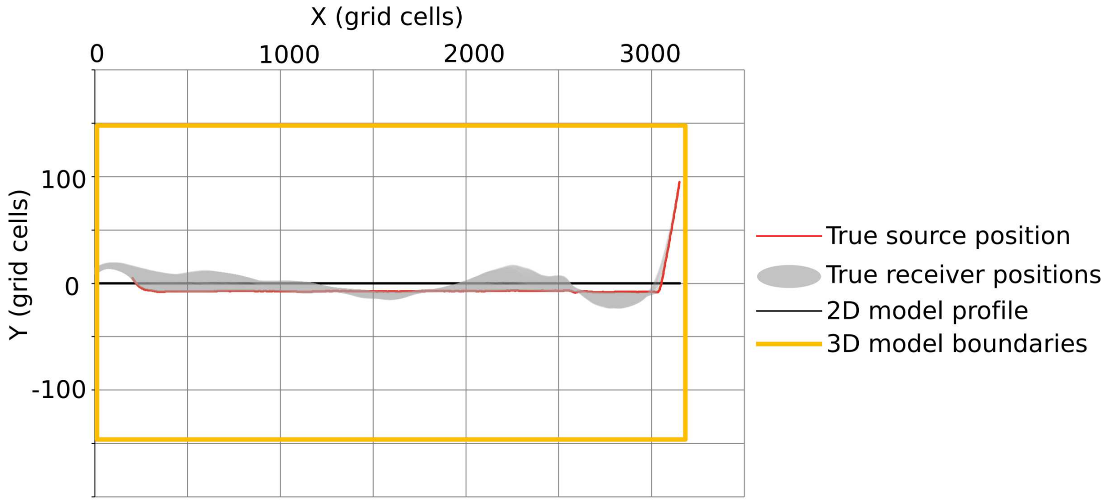


Figure 4.6: The true source and receiver positions mapped onto grid coordinates with the grid spacing of 25m. Sources are shown in red, receivers in grey. The distribution of the receivers is far from 2-dimensional. The black line denotes the 2D model profile and the orange rectangle shows the 3D model boundaries.

profile with 3153 cells along the x-axis and only 1 cell along the y-axis.

For 2D FWI, all sources and receivers were projected onto this 2D profile. There are two obvious ways to project the coordinates. First, a simple orthogonal projection and second, a projection with rotation along the circle with the radius equal to the offset; the latter retains the true offset value. These two ways are schematically illustrated in Figure 4.7. Both were tested and in this particular case no substantial difference between the inversion results were observed.

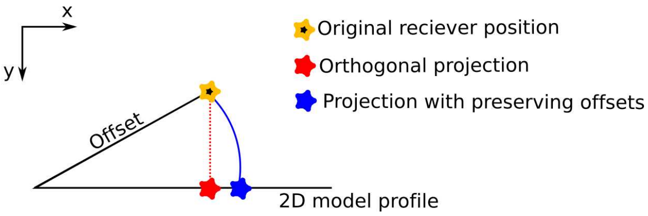


Figure 4.7: Schematic illustration of transforming geographical coordinates to model coordinates. An example receiver is shown in yellow. The red receiver is projected using orthogonal projection, that keeps the x-coordinate but does not keep offset. The blue receiver is projected by rotation along the circle with the radius equal to the offset and the centre located at the corresponding source.

For 3D FWI, the true source and receiver positions were retained within the starting model. The 3D starting model was generated by extending the y-dimension of the 2D model simply by keeping the velocity and anisotropy values the same along the same depth. Figure 4.6

shows the boundaries of the obtained 3D model.

There is a trade off between the size of the model and computational cost of the inversion. The size of the boundaries was set to the minimum that covers all sources and receivers and, in addition, avoids artificial boundary reflections. The 3D starting model has dimensions of 3151 x 105 x 300 grid cells whilst the 2D model has dimensions of 3151 x 300 grid cells. The former was heavily smoothed along the y-direction.

4.4.2 Density

The seismic wavefield depends upon the density of the medium in which it is travelling as well as the velocity. When a seismic wave encounters an interface between the two media, the portion of energy that reflects from the interface depends upon the seismic, or acoustic, impedance contrast between the two. In seismology, the acoustic impedance Z is defined as

$$Z = V\rho \tag{4.2}$$

where V is the velocity of the medium and ρ is the density of the medium. In principle, the density model should be required for the inversion as it governs the acoustic impedance, although it is possible to ignore the density in a wave equation and use only the velocity contrast rather than the impedance contrast to form the reflections.

In the case of marine datasets, the only sharp boundary always present in the starting model is the water bottom reflector. The impedance contrast at the water bottom caused by large variation between the water and sediments densities whilst the compressional velocity contrast is negligible. Therefore, neglecting density by using a constant density model is proven to be insufficient for such a dataset because it may result in inaccurate modelling of water bottom reflection and water-bottom multiples (Nangoo 2013). In order to mitigate this effect, (1) the water bottom and its multiple can be removed from the field data or, alternatively (2) the gradient at the water bottom can be fixed so that the velocity updates will start just below the water bottom.

The acoustic wave equation 2.6 used throughout this thesis contains the density as one of the model parameters. The pre-processing was minimal and the water bottom reflection and its multiple were retained in the field data. Therefore, the density model must be obtained prior to the inversion. There are different methods to generate an appropriate density model suitable for FWI, including those non-physical (Silverton et al. 2014; 2015). Conventionally, the density model is derived from the velocity using an empirically established relationship.

In 1974, Gardner et al. (1974) has formulated an empirical relationship between the density ρ and the velocity V for different rock types as

$$\rho = \alpha V^{\beta} \tag{4.3}$$

where α and β are some constants that depend upon both the lithology and the units in which density and velocity are measured. This relationship, also known as Gardner's law, with $\alpha = 310$ and $\beta = 0.25$, was used to obtain the density model in the sediments from the velocity model given in m/s at every iteration during the inversion.

Figure 4.8 illustrates how the density model was generated in $fullwave3D$ on an example of a single trace. The left hand-side shows a trace from the given velocity model and the right hand-side shows the generated from it density profile. The blue curves on both indicate the water column whilst the green curves indicate the sediments below the seabed. Both models are discretised in space with the crosses denoting the grid cells. Here it is assumed that the seabed in the velocity model has the correct position in depth. Everywhere in the water column the density value was set to 1000 kg/m^3. That was achieved by telling $fullwave3D$ the maximum water velocity. For the Pivot dataset, where the water column with the constant value was used, it was 1495 m/s. Gardner's law was applied to the velocity values equal and above the "lowest sediment velocity" $V_{sediment\_min}$, that was carefully examined and set to 1560 m/s. Everywhere in between, the density was calculated by linear interpolation.

Figure 4.8 demonstrates that correct identification of the lowest sediment velocity value $V_{sediment\_min}$ is of vital importance. In purple, some sediment velocity value is incorrectly set as the minimum sediment velocity and Gardner's law applied to this value gives the density value highlighted in bold purple. Then the linear interpolation of the density values is performed between the bold purple and black crosses. The effective seabed is now shifted from its true

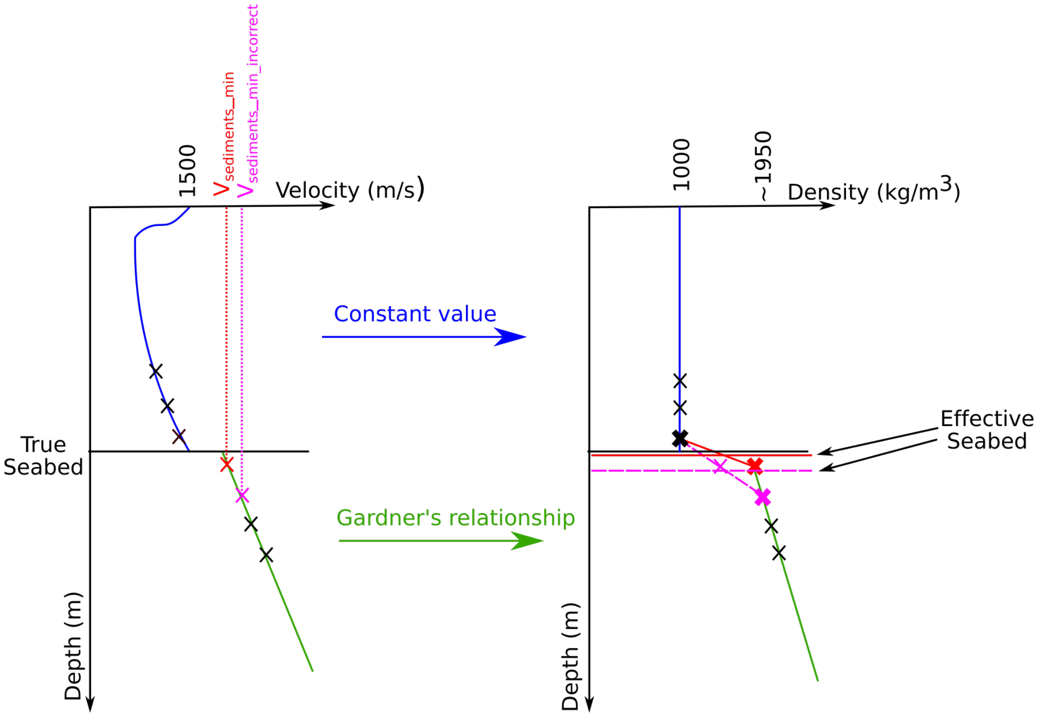


Figure 4.8: Schematic illustration showing the density model generation. Crosses denote the grid cells. In the water column, the density is kept constant with the value of 1000 kg/m^3. Below the lowest velocity value of the sediments, the density is computed using Gardner's law. The bold crosses on the density profile indicate the values between which the density is linearly interpolated. The effective seabed is defined by the chosen lowest velocity value of the sediments.

position by one cell. This will result in incorrect water bottom modelling and, eventually, the inversion failing. Once the minimum sediment velocity $V_{sediment\_min}$ is identified correctly as shown in red, then the depth of the effective seabed in the density model will correspond to the true seabed position in the velocity model.

4.5 Source wavefield

Estimating an accurate source signature is a vital component of successful FWI. Unlike conventional processing, where estimating the source wavelet is also important, for FWI it is crucial for the wavelet to be accurate at the lowest frequencies. For marine field data, FWI is often performed using explicit free-surface boundary condition in the modelling that allows for the reflection from the sea surface. Therefore, it is also important for the wavelet to contain no source ghost in it. A significantly incorrect wavelet will lead to a local minimum during the inversion or will force the inversion to fail completely.

There are many possible ways to estimate the source wavelet:

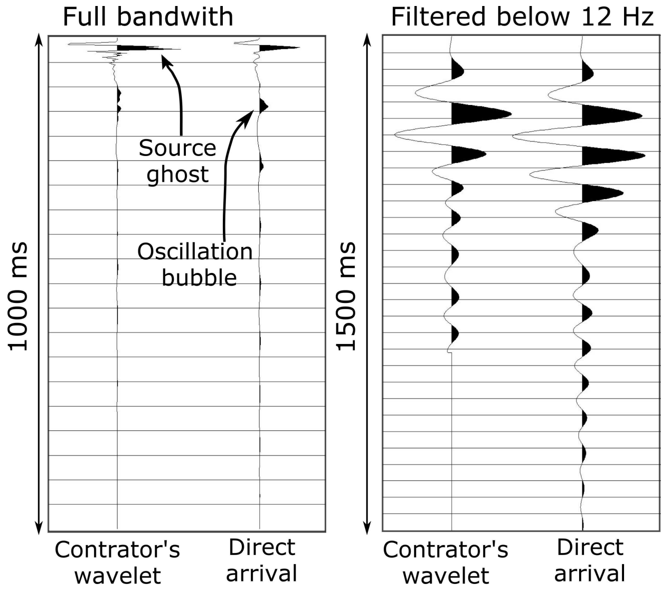
- **Extract the wavelet from the data**. In conventional processing it is not uncommon to use water bottom arrivals and direct arrivals for the source estimation. This usually works well when extracting from the direct arrivals in a deep-water field dataset, where there is no influence of geology on the source. Warner et al. (2013) has demonstrated successful estimation of the source wavelet from the shallow-water field dataset. However, this method requires the wavelet to be de-ghosted and, in some cases, de-multiplied prior the inversion.

- **Record the wavelet on a near-gun hydrophone** during acquisition. Such a wavelet needs to be de-ghosted.

- **Estimate the source signature using a numerical simulation**, where the wavelet depends on the gun configuration and appropriate physical properties of medium, such as water velocity, density, temperature and salinity. The left trace in Figure 4.9(a) shows the contractor's estimate of the far-field source wavelet that contains the source ghost. This wavelet was generated by a marine source modelling software. The right trace in Figure 4.9(a) shows a direct arrival recorded by a near-offset hydrophone at an offset of about 100 m in 1500 m of water, shifted to zero time. Figure 4.9(b) shows the same two wavelets low-passed using an Ormsby filter rolling off from 9 to 12 Hz; note that the time axis in Figure 4.9(a) and Figure 4.9(b) differ. The physics simulation approach usually works well for source estimation at medium and high frequencies. In this particular case, perhaps, it works well enough for the source estimation at the crucial low frequencies as well. However, Warner et al. (2013) has demonstrated that it is not always the case and wavelets estimated using such techniques should be examined carefully. In addition, wavelets produces using this method need to be de-ghosted.

- **Inverting for the source**. Pratt (1999) shows that the source wavelet can be estimated during the inversion. If the source varies from shot to shot, this method provides an advantage over other methods. However, any systematic errors in the velocity model may be mapped into the estimated source and will result in time shifts between the predicted and observed data, thus, compromising the quality of the inversion.

- **Using a matching filter to derive the source**. Described by Robinson et al. (1967), this method involves making an appropriate initial guess (for instance, low-pass filtered delta function), predicting the direct arrival through the water and using a Wiener filter to match the predicted and observed direct arrival. This method deals correctly with the surface ghost and, therefore, requires no de-ghosting. It also automatically corrects the amplitude and phase spectra for the 2D source if 2D modelling is used.



(a) (b)

Figure 4.9: Comparison of the data-derived and the contractors source signatures for the Pivot dataset. (a) Full-bandwidth waveforms. (b) The same waveforms filtered below 12 Hz. The leftmost waveform is the contractor's supplied source signature and the rightmost waveform is the direct arrival.

I consequently did not use the contractor's wavelet. Rather, the latter approach was followed as it would produce an already de-ghosted source signature that is accurate at low frequencies. This method was particularly attractive for the Pivot dataset because the deep-water field data contained direct arrivals not contaminated by interference with the water bottom arrivals.

First, I will describe step-by-step approach that was used to estimate the source signature for 2D FWI and, second, for 3D FWI.

4.5.1 Source estimation for 2D FWI

The first step in estimating the source signature involved building an accurate water velocity model. The direct arrival travels from a source to a receiver through the shallow part of the water column and the gradient of this arrival determines the water velocity. The gradient was estimated carefully, on a number of randomly selected shots across the model. Figure 4.10 shows a close up of a shot gather filtered below 12 Hz focusing on the direct arrival. The water velocity was found to be 1524 m/s.

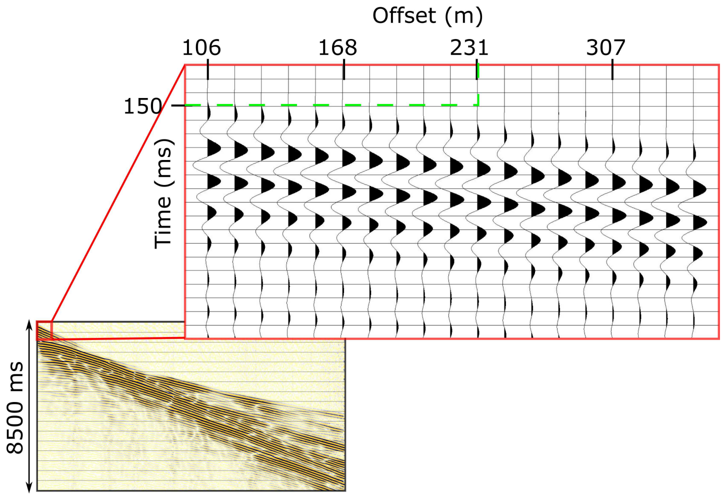


Figure 4.10: Close up of a shot gather filtered below 12 Hz. The shallow water velocity is determined using the gradient of the direct arrival at near-offsets and is found to be 1524 m/s.

Once the water velocity has been calculated, a 2D water model with a constant velocity of 1524 m/s was created. To avoid numerical dispersion and allow stable propagation of the wavefield up to a frequency of 12 Hz, a grid size of 25 m was used. The model had dimensions of 3 x 3 km and an explicit free surface top boundary condition. The model was large enough so that the single source positioned in the centre of the model would not cause any reflections from the remaining boundaries to be recorded at the receivers.

The next step involved extracting the direct arrival from the field data. In principle, any chosen near-offset trace can be used if the direct arrival is relatively noise-free and its shape is consistent from trace to trace. Sometimes, an average of a few traces and/or across a few shots might be used in order to improve signal-to-noise ratio (Ashley 2019). I used a single trace with an offset of 168 m and I will demonstrate the rational for this later in this section.

The next step required a reasonable guess source signature. This was achieved by applying a minimum-phase Ormsby low-pass filter rolling off from 9 to 12 Hz to a delta function (Fig. 4.11(a)) and re-sampling it in time to the same sampling rate of 2 ms as field data. Note, that this exact filter was applied to the field data.This guess wavelet would contain the same frequencies as the field data so it was a suitable initial guess source. This source was propagated in the shallow water model using the 2D acoustic modelling code and recorded at the receivers. Figure 4.12(a) shows the observed direct arrival and calculated direct arrival using the initial guess source. It can be seen that the predicted direct arrival does not have quite the same phase as the observed direct arrival and the energy distribution within the two wavefields differ significantly, although the first packs of energy arrive at approximately the same time. The mismatch between the modelled and field data appeared to increase with increasing offsets. Both calculated and observed direct arrival were consistent from trace to trace so it was decided to use a single trace with an offset of 168 m. This trace shifted in time by 1000 ms is shown in Figure 4.11(b).

Finally, having obtained the initial guess source, the predicted direct arrival and the observed direct arrival, the final source was generated using a matching Wiener filter (Yilmaz 2001). It is well established that: (1) convolution of a guess source with the impulse response of the velocity model produces the modelled data and (2) convolution of the true source with the impulse response of the velocity model results in the observed field data:

$$\begin{gathered} guess\_source * velocity\_model = modelled\_data \\ true\_source * velocity\_model = field\_data \end{gathered} \tag{4.4}$$

where the impulse response of the velocity model is shortened to *velocity_model*. If the field data, the guess source and the modelled data are known, and the velocity model is assumed to be true, than the inverse of the modelled data convolved with the guess source and the impulse response of the model is approximately equal to the delta function:

$$(synthetic\_data)^{-1} * guess\_source * velocity\_model \approx \delta \tag{4.5}$$

Thus, the following can be obtained:

$$[observed\_data * (synthetic\_data)^{-1} * guess\_source] * velocity\_model \approx \delta * observed\_data$$
$$= observed\_data \quad (4.6)$$

where

$$observed\_data * (synthetic\_data)^{-1} * guess\_source = true\_source \quad (4.7)$$

That is the Wiener filter that I applied to the guess source. Although an important requirement must be met. In signal processing, a signal is “causal” if it equals 0 at every time before zero-time. The convolution product must be causal, otherwise the information will be lost and it will compromise the final source signature. That is why a time shift of 1000 ms was applied to the observed trace prior to convolution. The same time shift was applied back to the final source signature. The traces involved into the matching filter are shown in Figure 4.11. Once

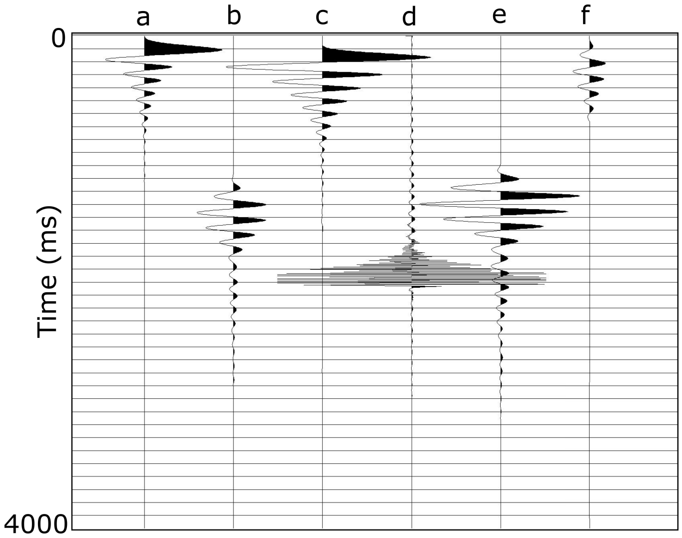
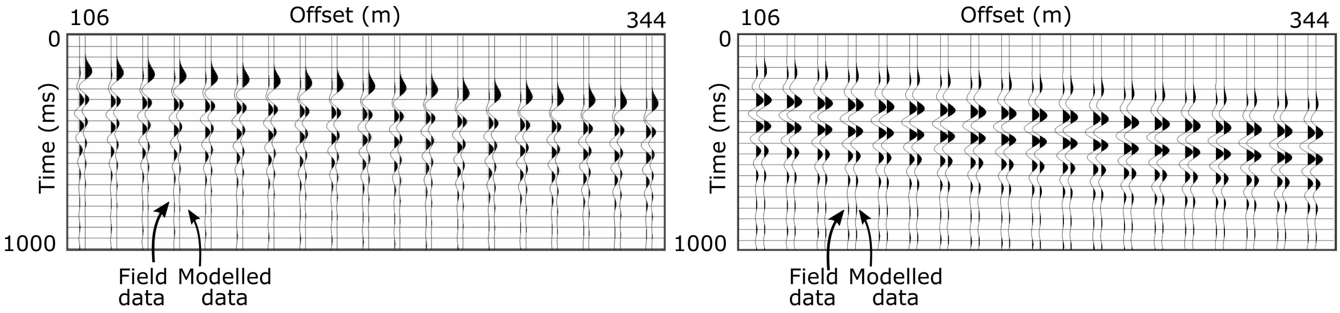


Figure 4.11: Step-by-step generation of a source wavelet from the direct arrival using a Wiener filter. (a) The initial guess source; (b) the direct arrival with an offset of 168 m and with applied time shift of 1000 ms; (c) the modelled direct arrival with the same offset of 168 m using the initial guess source; (d) the inverse of the modelled direct arrival; (e) the convolution product of (b) the time shifted direct arrival and (d) the inverse of the modelled trace. (f) The convolution product of (e) and (a) is the final source signature.

obtained, the source signature (Fig. 4.11(f)) must be validated. In some surveys the source wavelet may vary from shot to shot due to unstable gun array performance or systematic errors during the acquisition. Therefore, it is reasonable to test the source signature on a few randomly selected shots across the model. Here I will show only one shot, but the validation was performed using multiple shot gathers from different parts of the dataset and the results appeared to be consistent within those selected shot records. An example shot gather of the direct arrivals comparing the modelled data using the derived source wavelet and the field data are shown in Figure 4.12(b). Note, that the data have been trace-equalised, so that only the relative amplitudes within a trace can be compared. Overall, the match between the modelled and observed direct arrivals appeared to be adequate within all selection of the shot gathers. Therefore, the derived 2D source wavelet (Fig. 4.11(f)) is accurate for 2D FWI.



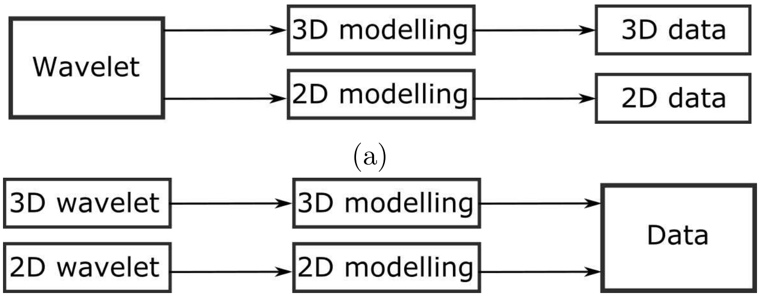
(a) (b)

Figure 4.12: Trace-by-trace comparison of the field and modelled direct arrivals using: (a) the guess source signature and (b) the derived signature using a matching filter. The match between the field data and modelled data significantly improves from (a) to (b).

4.5.2 2D vs. 3D

In order to perform the inversion on the same dataset using 2D or 3D modelling, it is important to recognise that the same source signature can not be used for both cases. This is because the same wavelet, propagated through a model will generate data with different amplitude and phase spectra depending upon the type of propagation. In FWI, the modelled data must match the field data no matter whether the inversion is in 2D or 3D. In order to generate the same data from both 2D and 3D modelling, different source signatures must be used. I perform two synthetic tests to demonstrate this.

Firstly, I propagated the same source signature in a homogeneous model in both 2D and 3D in order to compare the modelled data. Secondly, recognising the difference between the two, I made adjustments to the source signature which are appropriate for 2D and propagated two different sources in the same homogeneous model again in order to obtain the same data. A summary of the experiment is outlined in Figure 4.13.

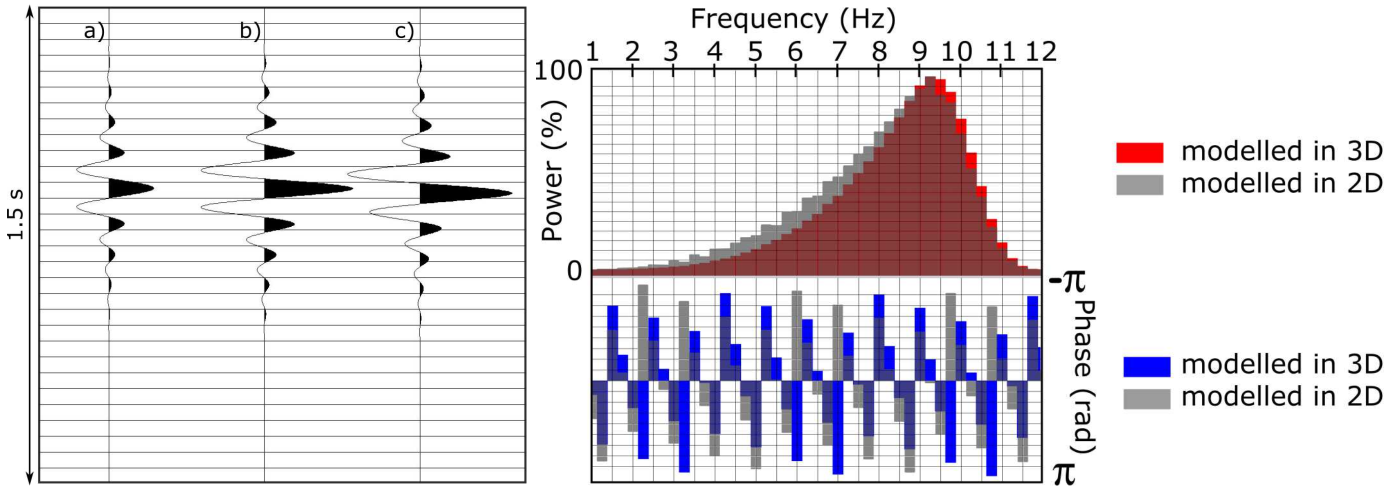


(a)

(b)

Figure 4.13: Schematic diagram illustrating a synthetic experiment: (a) the same wavelet propagated using 2D and 3D modelling codes in the same homogeneous model generates different data and (b) to obtain the same data whilst modelling in 2D and 3D, different wavelets must be used.

For this experiment I used a homogeneous velocity model with the velocity of 1520 m/s and the dimensions of 3 x 3 x 3 km for 3D and 3 x 1 x 3 km for 2D, a single source and a single receiver with an offset of 107 m. The source signature given by a spike low-passed below 12 Hz is displayed in Figure 4.14(a.a). I will refer to this source signature as the “3D source”. The grid spacing was 12.5 m, the time sampling was 1 s and the high kernel version of the code was used.



(a)

(b)

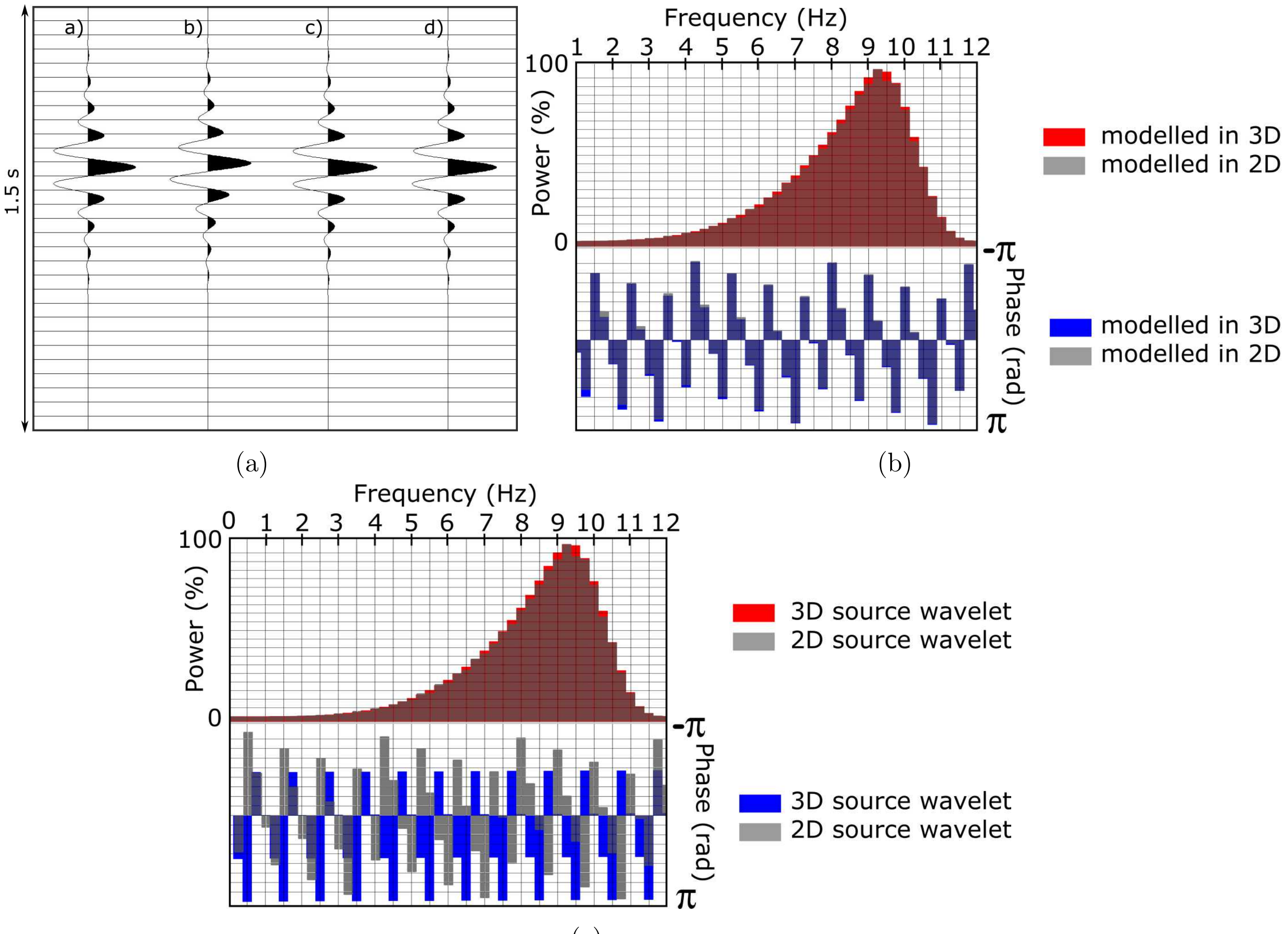
Figure 4.14: Modelling in 2D and 3D using the same source. (a.a) The source signature, (a.b) modelled in 3D data, (a.c) modelled in 2D data. (b) The amplitude and phase spectra of the modelled in 2D and 3D data.

Once the source is propagated in 3D, the signal of the same shape, i.e. of the same amplitude and phase spectra should be recorded at the receiver. However, in 2D it is not the case. Figure 4.14(a) shows three traces: the source signature, that is propagated in 2D and 3D; a trace recorded at the receiver in 3D and a trace recorded at the receiver in 2D. The traces are normalised and the absolute amplitudes can not be compared. It can be seen, that the trace modelled in 3D is identical to the source signature, while the trace modelled in 2D has an obvious phase shift. The amplitude and phase spectra of the modelled traces are shown in Figure 4.14(b). The differences between the modelled traces are: (1) phase shift of $\pi/4$ and (2) amplitude spectra for 2D data is stronger at lower frequencies than that of 3D.

Due to the nature of wave propagation, in different domains the plane wave is propagated differently. In two-dimensional simulations, the plane wave source is injected along a line, whilst in three-dimensional simulations the plane wave source is injected along a plane. In other words, the source is treated as a point in 3D and as a line in 2D. Hence, 2D propagation acts as a high-cut filter of $1/\sqrt{f}$ (Mike Warner, personal communication, October 1^{st}, 2019). It also shifts the frequency spectrum of the data which can be explained by interference of the frequencies when they travel along the line in the 2D case whilst in the 3D case there is no interference.

Consequently, in order to obtain the same data whilst modelling in 2D or 3D, there must be appropriate adjustments made for 2D to the source signature: (1) a positive phase shift of $\pi/4$ and (2) a low-cut filter of $\sqrt{f}$. Figure 4.15(a) shows the 3D source, the 2D source and the corresponding modelled traces. It can be seen, that now the data modelled in 2D are identical to the data modelled in 3D, which is also demonstrated by their identical amplitude and phase spectra (Fig. 4.15(b)).

To summarise, the source signatures for the same dataset differ if the inversion is performed in 2D and 3D. It is vital to use an appropriate source signature for successful FWI .



(a) (b)

(c)

Figure 4.15: Modelling in 2D and 3D using appropriate for 2D and 3D sources. (a.a) The 3D source, (a.b) the 2D source, (a.c) modelled in 3D data, (a.d) modelled in 2D data. Note, that the sources are different whilst the modelled data are identical. (b) The amplitude and phase spectra of the modelled in 2D and 3D data. (c) The amplitude and phase spectra of the 2D and 3D source signatures.

4.5.3 Source estimation for 3D FWI

The appropriate signature for 2D FWI was obtained using an initial estimate to model the direct arrival through water, subsequently using a matching Wiener filter to match this to the observed direct arrival. This approach automatically corrects for the phase and amplitude spectra of the source wavelet for 2D; it also deals correctly with the surface ghost so that no de-ghosting is required. The method is fully described in the section above.

However, the effective source wavelet required for FWI differs between 2D and 3D (see the analysis in the section above). In principle, it is possible to obtain a 3D source signature with a simple phase rotation of $\pi/4$ and a low-cut filter of $1/\sqrt{\omega}$ applied to a 2D source wavelet with $\omega = 2\pi f$ being an angular frequency.

However, the appropriate source for 3D FWI was obtained using exactly the same strategy as for the source derivation for 2D FWI but using the 3D wavefield propagator instead of the 2D wavefield propagator. The 2D water model with a constant velocity value of 1524 m/s was extended by an extra dimension; the final size was 3 x 3 x 3 km.

The obtained source signatures for both 2D and 3D FWI are shown in Figure 4.16; It can be easily noticed, that they differ. Figure 4.17 displays the phase and amplitude spectrum for both 2D and 3D source wavefields. It is easy to observe that the 2D source wavelet is lacking lower frequencies compare to the 3D source, in this case especially noticeable at the range from 5 to 6.5 Hz; as if it was through a low-cut filter. The phase spectrum shows the phase difference of about $\pi/4$ for a single frequency in the spectrum. These differences between the two sources are consistent with the synthetic experiment presented earlier. Hence, the appropriate source signatures for the same field dataset have different phase and amplitude spectra. It only depends upon whether the source wavelet is designed for 2D or 3D FWI.

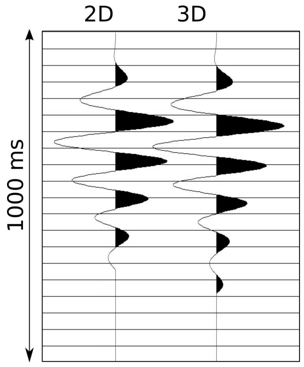


Figure 4.16: Source signatures for 2D and 3D FWI. The leftmost waveform shows the 2D-derived source wavelet and the rightmost waveform shows the 3D-derived wavelet. The phase shift between the two waveforms is $\pi/4$.

Also, in 3D the energy of the wave-front is spread out over the spherical surface area $4\pi R^2$, where R is the distance from the source. This means that the energy per unit area of an expanding spherical wave decreases as $1/R^2$. Since energy is proportional to amplitude squared, an inverse square law for energy translates to a $1/R$ decay law for amplitude for three-dimensional simulations. In contrast, in 2D the energy of the wave-front per unit area of an expanding wave decreases as $1/R$, which translates into a $1/\sqrt{R}$ decay law for amplitude. Fig-

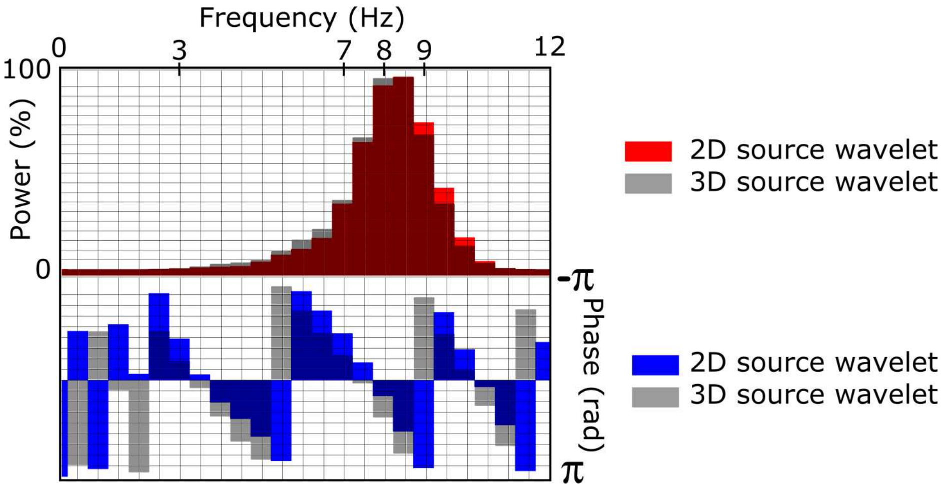


Figure 4.17: Comparison of the amplitude and phase spectra of the 2D and 3D source signatures.

ure 4.18 demonstrates this on an example of two traces, simulated in 3D and 2D respectively. The traces are oriented horizontally rather than vertically with the zero-time on the left hand side of each trace. It can be easily seen, that the amplitudes in 3D decay much faster rather

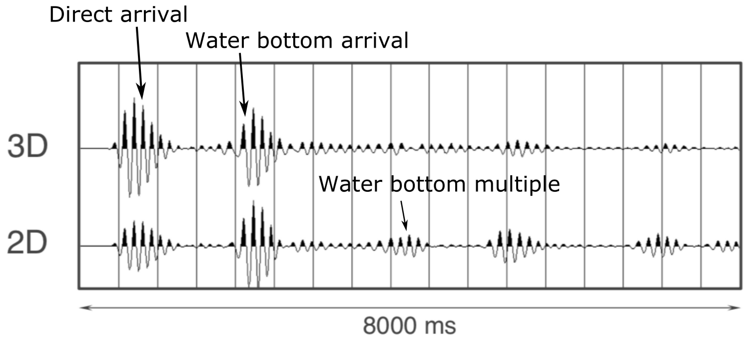


Figure 4.18: Comparison of the traces simulated using the 3D (top) and the 2D (bottom) propagators. The amplitudes in 3D decay as $1/R$; in 2D as $1/\sqrt{R}$, where R is the distance from the source. The multiples become more pronounced in 2D.

than in 2D. For this reason, in 2D multiples become more pronounced than in 3D and, consequently, play larger role in the inversion. Although 3D simulation provides a more accurate real world prediction, relative amplitudes crucially do not impact the inversion result in 2D. I did not explicitly correct the temporal decay of amplitude for 2D FWI because $fullwave3D$ does not require the amplitudes being scaled for 2D. Instead, this was corrected heuristically by matching RMS amplitudes between the predicted and field data, suitably stabilised, at each time slice within each shot record. This forms part of the default parameterisation of $fullwave3D$ for marine field data where “it is designed principally to deal with the amplitude effects of anelasticity and sub-grid-cell scattering, but it has the useful side effect of also dealing reasonably effectively with the amplitude differences introduced by 2D simulation" (Kalinicheva et al. 2017).

4.6 Modelling strategy

FWI can be computationally expensive; 3D FWI is always significantly more expensive than 2D. For a 2D model with a dimension of n mesh points in each direction and n time steps, the runtime of the inversion is proportional to n^3. For a 3D model with an extra third dimension of n mesh points, the runtime is proportional to n^4. For a time-domain algorithm used in this thesis the cost of FWI also depends upon the number of sources per iteration and the number of iterations (Warner et al. 2013). Consequently, to minimise the cost of the inversion the number of sources and/or the number of mesh points might be reduced.

Here, the full number of sources available in the data was used for both 2D and 3D FWI. The amount of mesh points is associated with the size of the grid cells in the model. In its turn, the grid size is defined by the frequency content of the field data, velocity values in the starting model and accuracy of the modelling code. For this study, for both 2D and 3D cases, the field data contained frequencies with the maximum of 12 Hz, the minimum velocity of the starting model was equal to water velocity of 1495 m/s, and the modelling code accurate at 5 grid points per wavelength was used. Recalling formula 2.64, the following grid size h was calculated

$$h \leq \frac{1}{5}\frac{1495m/s}{12Hz} \approx 25m \tag{4.8}$$

At maximum frequency of 12 Hz, given the velocity model and the modelling code, the grid spacing must be no more than 25 m. This would provide 5 grid points per wavelength in the water column and even more everywhere else in the model because the velocities in the sediments are higher than 1495 m/s. The grid spacing of 25 m is optimal for this study as it gives minimal numerical dispersion for frequencies below 12 Hz and the lowest computational cost possible. At the grid spacing of 25 m, the starting models in 2D and 3D have dimensions of 3153 x 300 and 3153 x 105 x 300 grid cells respectively. Allowing for boundaries, these models arrive to dimensions of 3193 x 320 and 3193 x 145 x 320 grid cells for 2D and 3D respectively.

A finer grid size would allow modelling higher frequencies, although increase the cost of the inversion. For a 2D problem it would still be feasible and will be discussed in another chapter of this thesis. However, modelling frequencies over 12 Hz in 3D for the large Pivot dataset would result in non feasible computational cost.

The chosen gird size and the maximum velocity value in the starting model define the time step required for the stable wavefield propagation. The modelling code used for this study is 4^{th} order accurate in space and 2^{nd} order accurate in time, therefore allows the wavefield to cross no more than half a grid cell at every time step. Recalling the equation 2.65, the time step can be calculated as

$$\Delta t < 0.495 \frac{25m}{5500m/s} \approx 2.25ms \tag{4.9}$$

The time step of 2 ms was chosen.

4.7 Validation of the starting models

Assuring the quality of the starting model, the source wavelet and the starting frequency prior to the inversion is potentially the most important step to take to ensure the desired outcome. For the Pivot dataset, the quality of the source signature and frequency analysis were addressed in sections 4.5 and 4.3 respectively. Now, the requirement that must be tested is whether the starting model predicts the data that match the field data within half a cycle at the starting frequency.

Because FWI is a local iterative scheme, it suffers from cycle-skipping if this condition is not satisfied. Due to the oscillatory nature of seismic data, it is also possible that the mismatch between the modelled and field data greater than 180° might reduce the data misfit (Beydoun et al. 1998) . In this case, cycle-skipping may not even be noticed, nonetheless the obtained model will be wrong.

To ensure that the starting model is not cycle-skipped, Shah et al. (2012) suggested a rigorous analysis by diagnosing the spatial continuity of the phase difference between the modelled and field data, extracted at the starting frequency after windowing the data in time around arrivals of interest. Alternatively, after the forward modelling, a number of time-domain shot gathers randomly selected across the model and filtered down to the starting frequency may be assessed. This method requires no additional computational effort, is easy to perform and is presented here.

4.7.1 For 2D FWI

To validate the 2D starting velocity and anisotropy models, a range of randomly selected shot gathers across the model was generated using the 2D modelling code and the 2D source signature. Note, that the obtained 2D source signature contains the frequencies with the maximum of 12 Hz. Consequently, the shot gathers will also contain the frequencies with the maximum of 12 Hz.

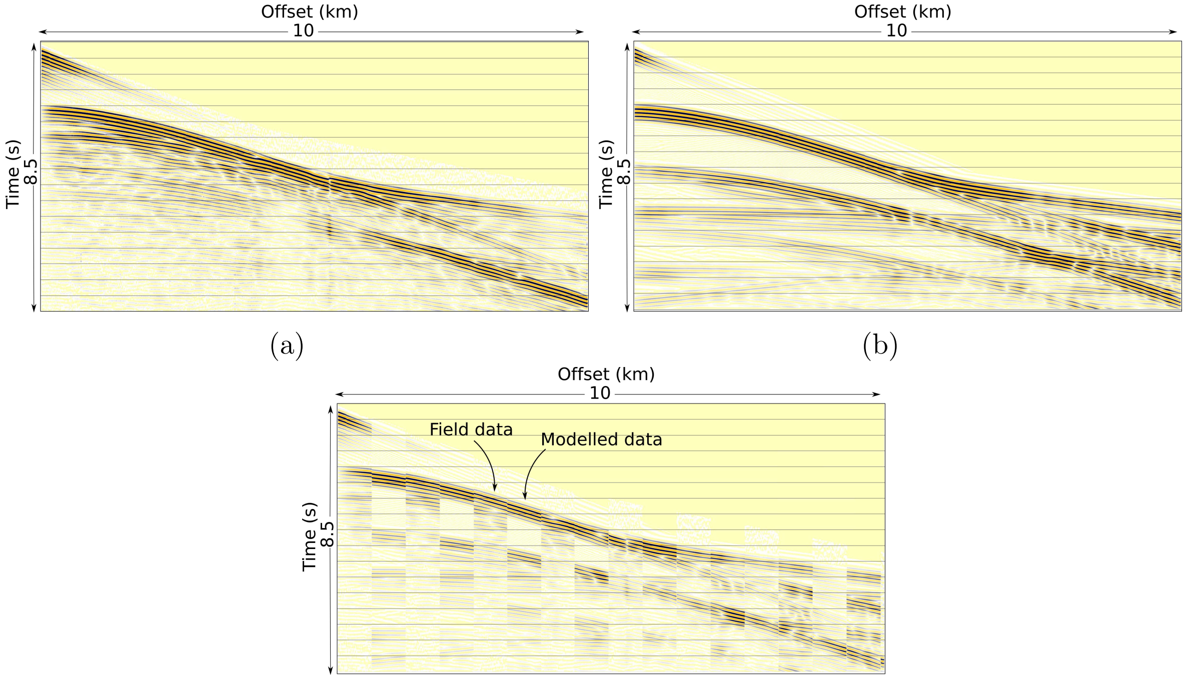
Figure 4.19 shows an example raw shot gather, the corresponding modelled shot and the interleaved shot gather of the two. At the presented range of frequencies the main reflections observed in the field data and the modelled data are the water bottom arrival, its multiple and some refraction energy at long offsets.

The 2D starting model provides the mismatch between the modelled and the field data approximately of half a cycle for the water bottom arrival and the refracted energy ahead of the water bottom arrival. The water bottom multiple appears to be completely cycle-skipped. This does not mean that the starting velocity model is inaccurate for the inversion. It is simply not accurate enough to start the inversion at 12 Hz.

Figure 4.20 shows the same shot gathers low-passed filtered to the starting frequency of 3 Hz. It can be seen that at the starting frequency, the mismatch between the modelled and the field data is within half a cycle everywhere along the first arrivals. Although the raw shot gather contains relatively large amount of incoherent noise, that will not affect the quality of the inversion as $fullwave3D$ is rather robust against most sources of noise.

4.7.2 For 3D FWI

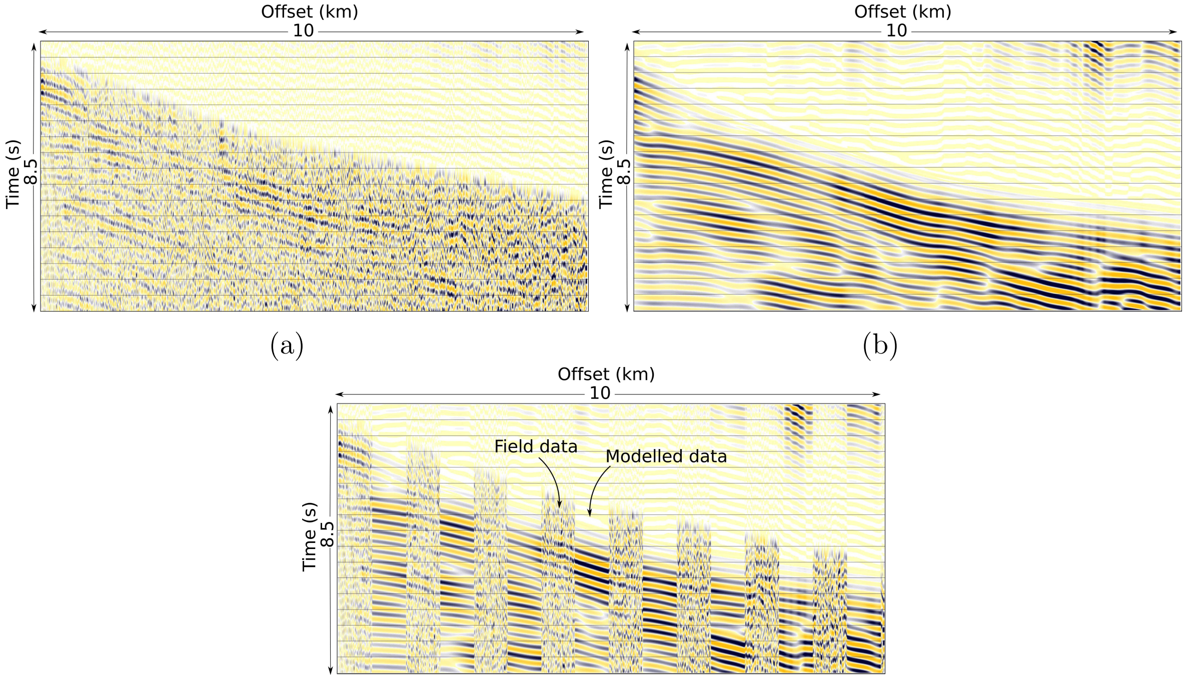
To validate the 3D starting velocity and anisotropy models an analysis similar to that of the 2D model was performed. Same randomly selected shot gathers across the model were generated using the 3D modelling code and the 3D source signature. Note, that the obtained 3D source



(a) (b)

(c)

Figure 4.19: An example (a) field and (b) modelled in 2D shot gather low-passed below 12 Hz. (c) Interleaved (a) and (b) shot gathers. The blocks of 50 traces are sequentially displayed. The shots are scaled trace-by-trace.



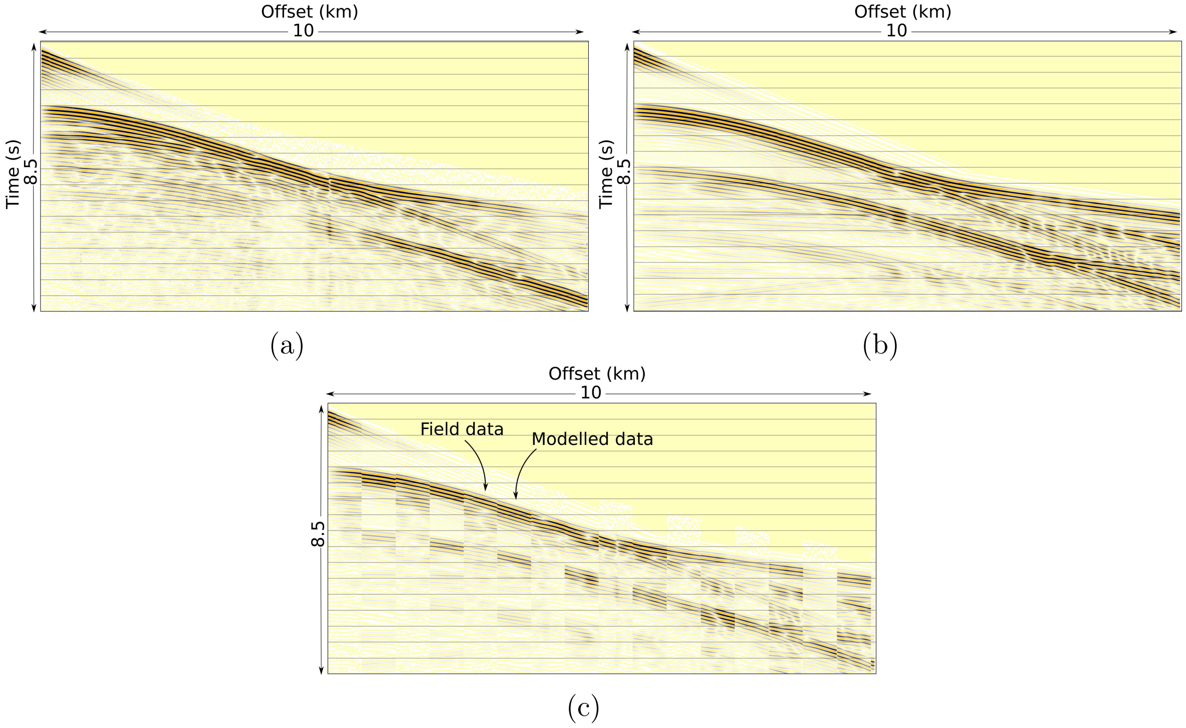
(a) (b)

(c)

Figure 4.20: Validation of the 2D starting velocity model. An example (a) field and (b) modelled in 2D shot gather low-passed filtered using Ormsby filter rolling off from 3.0 to 3.5 Hz. (c) Interleaved (a) and (b) shot gathers. The blocks of 50 traces are sequentially displayed. The shots are scaled trace-by-trace.

signature contains frequencies with the maximum of 12 Hz. Consequently, the shot gathers will also contain frequencies with the maximum of 12 Hz.

Figure 4.21 shows one of the raw shot gathers, the corresponding modelled shot and the interleaved shot gather of the two. Similar to the 2D case, the mismatch between the field and modelled data appears to be equal or slightly less than half a cycle for the fist arrivals. The water bottom multiple is also cycle-skipped at the nearest offsets. However, this only means that the inversion is likely to suffer from a local minimum if started from the frequency of 12 Hz. The same shot gathers filtered to the starting frequency of 3 Hz are displayed in Figure 4.22.



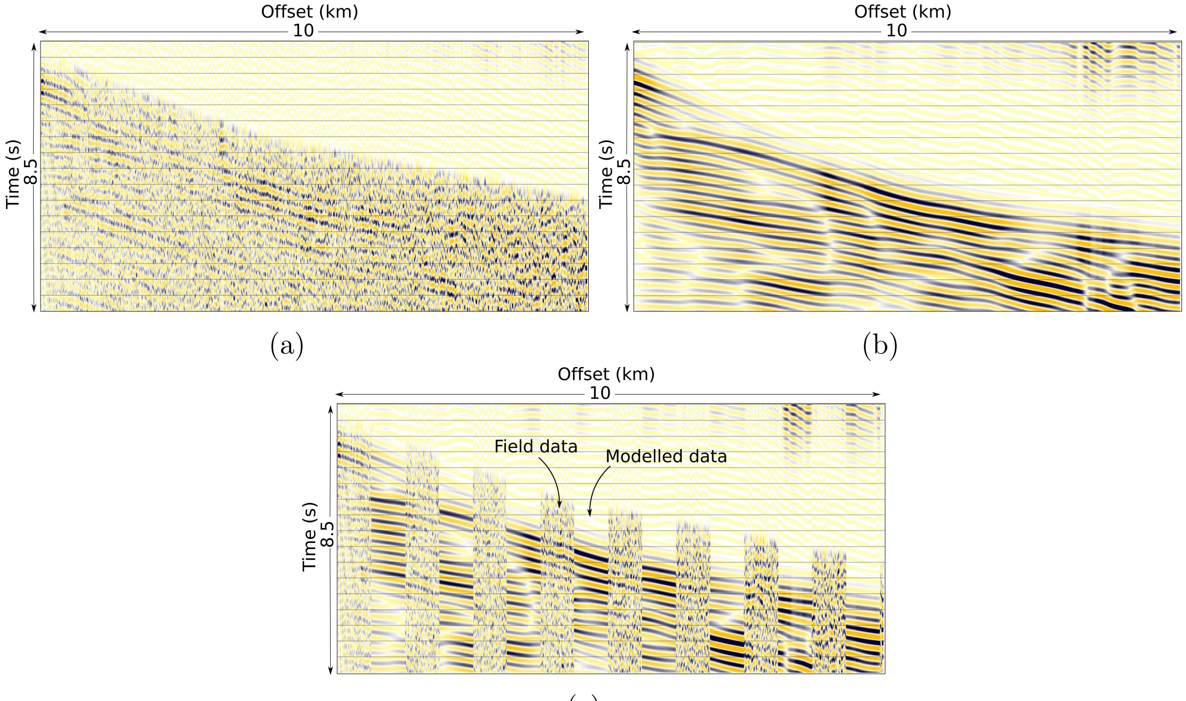
(a) (b)

(c)

Figure 4.21: An example (a) field and (b) modelled in 3D shot gather low-passed below 12 Hz. (c) Interleaved (a) and (b) shot gathers. The blocks of 50 traces are sequentially displayed. The shots are scaled trace-by-trace.

It can be easily seen, that the 3D starting velocity model predicts the data within half a cycle at the starting frequency.

Consequently, the starting velocity model and anisotropy models are adequate for FWI with the starting frequency of 3 Hz.



(a) (b)

(c)

Figure 4.22: Validation of the 3D starting velocity model. An example (a) field and (b) modelled in 3D shot gather low-passed filtered using Ormsby filter rolling off from 3.0 to 3.5 Hz. (c) Interleaved (a) and (b) shot gathers. The blocks of 50 traces are sequentially displayed. The shots are scaled trace-by-trace.

4.8 Inversion strategy

The inversion strategy designed for this study and described below was exactly the same for the 2D and 3D FWI.

In principle, it was possible to invert all 1480 sources simultaneously during every iteration. However, it would demand sufficient computational cost per iteration, especially in 3D. Instead, less shots per iteration were used and more iterations were performed. Such a strategy provides better results at a given computational cost (Leeuwen et al. 2013). The idea behind it is simple: a subset of sources would generate an update in the starting model during the first iteration so that another subset of sources would update the already improved model during the second iteration. This would produce better result rather than using all the sources during the first iteration simultaneously.

When designing a subset of sources, especially for sparse datasets, normally it is beneficial to choose sources randomly distributed in the model to avoid the interference pattern (Diaz et al. 2011). However, for this study, the subsets were formed simply by setting a number of shots to skip prior to assigning a shot to the subset. Thus, if there are 5 sources and the number of shots to skip is set to 1, then sources 1, 3 and 5 will form the first subset and sources 2 and 4 will form the second subset. In this study, the number of shots to skip was set to 1. The source separation was 50 m. Gaussian smoothing with the length of the dominant wavelength was performed after every iteration along the direction of the source line. No interference patterns were observed.

In total, 10 frequency bands were used with 4 iterations per each frequency band. Each source was used twice per each frequency band. The lowest frequency of 3 Hz was determined by the field data and the highest frequency of 12 Hz was set by computational effort required for 3D FWI and available for this study. The benefits of performing FWI at higher frequencies are discussed in details in another chapter of this thesis.

Modelling was acoustic rather than elastic with attenuation and elastic effects not explicitly taken into account. The inversion for slowness rather than for velocity was performed because some evidence suggest that it improves resolution and convergence rate (Warner et al. 2013). The anisotropy was fixed at every iteration during the inversion.

The inversion strategy is summarised in table 4.1.

4.9 Results and discussion

Anisotropic full-waveform inversion with the starting model based upon smoothed PSTM stacking velocities shown in Figure 4.23(a) was applied to the Pivot dataset in both 2D and 3D. The inversion began at the starting frequency of 3 Hz and was performed to the maximum frequency of 12 Hz. The strategy used for both inversions was exactly the same.

The final 2D-derived velocity model is displayed in Figure 4.23(b). Whilst the starting velocity model is smooth and does not contain any structures, the 2D FWI model has recovered many features. The strong irregular reflections in the second half of the section are basaltic intrusions, that appear as high velocity features in the FWI velocity model. Both the FWI velocity model and PSDM pick out the major unconformity that traverses the section, and both show shallow channels in the upper part of the section.

The final 3D-derived FWI model is displayed in Figure 4.23(c). Since there was little

	2D FWI	3D FWI
Starting frequency, Hz	3	
Maximum frequency, Hz	12	
Number of frequency bands	10	
Frequency bands	3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12 Hz	
Iterations per frequency	4	
Total number of iterations	40	
Total number of shots	1480	
Number of shots per iteration	740	
Invert for slowness	yes	
Normalise amplitudes	yes	
Spatial preconditioning	yes	
	Model parameters	
Size, km	78.8 x 7.475	78.8 x 2.6 x 7.475
Grid spacing, m	25	
Size, cells	3153 x 300	3153 x 105 x 300
	Data parameters	
Time sampling, ms	2	
Trace length, ms	8500	

Table 4.1: Summary of the inversion strategies for 2D and 3D FWI.

control over the velocity variations in the cross-line direction, the velocity model was strongly regularised in the cross-line direction during FWI. Apart from this regularisation, the inversion was run fully in 3D with sources and receivers properly distributed in space.

The final 3D FWI model is qualitatively very similar to the 2D FWI model. They both recover the same features, that also correspond to the ones observed in the PSDM section(Fig. 4.23(d)). Although these models are remarkably similar, they are not identical. For example, the local absolute velocity values do not always agree, however, there is no systematic differences between them. Both models match the PSDM section closely, with very subtle differences between them.

The only area with significant differences between the 2D and 3D FWI models is circled in Figure 4.23. For this area, the streamer was not straight due to the cable feathering of over 700 m (Fig. 3.6). The strong feathering possibly made a great complication for 2D FWI because it makes it difficult to account for the geometry properly, which in its turn is not a problem for a 3D domain. Thus, here, the 2D assumption was significantly in error. In these particular circumstances, 3D FWI that takes into account proper distribution of sources and receivers provides a superior outcome compared to 2D.

To further access the quality of both 2D and 3D FWI models, forward modelling was performed using the starting, 2D FWI and 3D FWI velocity models. The shot gathers are compared in Figure 4.24. It can be seen, that the shots modelled using the 2D and the 3D FWI models are similar and contain the same energy patches at the same locations, for instance the water bottom, its multiple and refraction energy arriving prior to the water bottom reflection. For both 2D and 3D FWI models these reflections are not cycle skipped with respect to the field data. However, there are evidence suggesting that in places the 3D FWI velocity model predicts the data that match the field data somewhat closer than the 2D FWI model. For instance, the mismatch between the modelled first water bottom multiple at mid-offsets and the modelled water-bottom arrival at the very long offsets become less pronounced if predicted with the 3D FWI; these differences are indicated by red arrows in Figure 4.24(d) and 4.24(c).

To further assure the quality of the FWI models and analyse differences between them anisotropic 2D Kirchhoff pre-stack depth migration was performed. The common image gathers generated with the starting velocity model, the 2D FWI and the 3D FWI are shown in Figure 4.25. The starting velocity model appears to be reasonably good, but it is too slow in most places (which is demonstrated by migrated common image gathers). Both 2D and 3D FWI models demonstrate significant improvement compared to the starting model. There are practically no visual difference between the 2D and 3D FWI.

Figure 4.23 displays the depth migrated stack sections with the starting velocity model, the 2D and 3D FWI velocity models. These PSDM sections show the central 25 km of the data. It can be seen, that both 2D and 3D FWI models provide flatter reflectors at approximately 2 km depth compared to that of the starting model and a better focused basaltic intrusion at the depth of approximately 3 km.

Although the migration was not perfect as it was performed in 2D for both 2D and 3D derived FWI models and cable feathering was not dealt with correctly, it is clear that: (1) 2D FWI provides significant improvement compared to the starting velocity model and (2) there is no or little difference between the image generated with 2D and 3D FWI.

4.10 Conclusions

Anisotropic full-waveform inversion was successfully applied to a narrow-azimuth marine-streamer dataset in both 2D and 3D domains with the maximum frequency of 12 Hz. It demonstrated an excellent resolution which can be obtained with FWI in 2D. Although the 3D FWI indicates superior results where the 2D assumptions were violated by the geometry of the dataset, it was clear that provided 2D distribution of sources and receivers 2D FWI proves to be useful for parameter testing, earlier QC while working with 3D datasets and is a good enough result for further depth migration. For the dataset in this study, computational effort of 2D FWI was at about 400 times less expensive than 3D FWI.

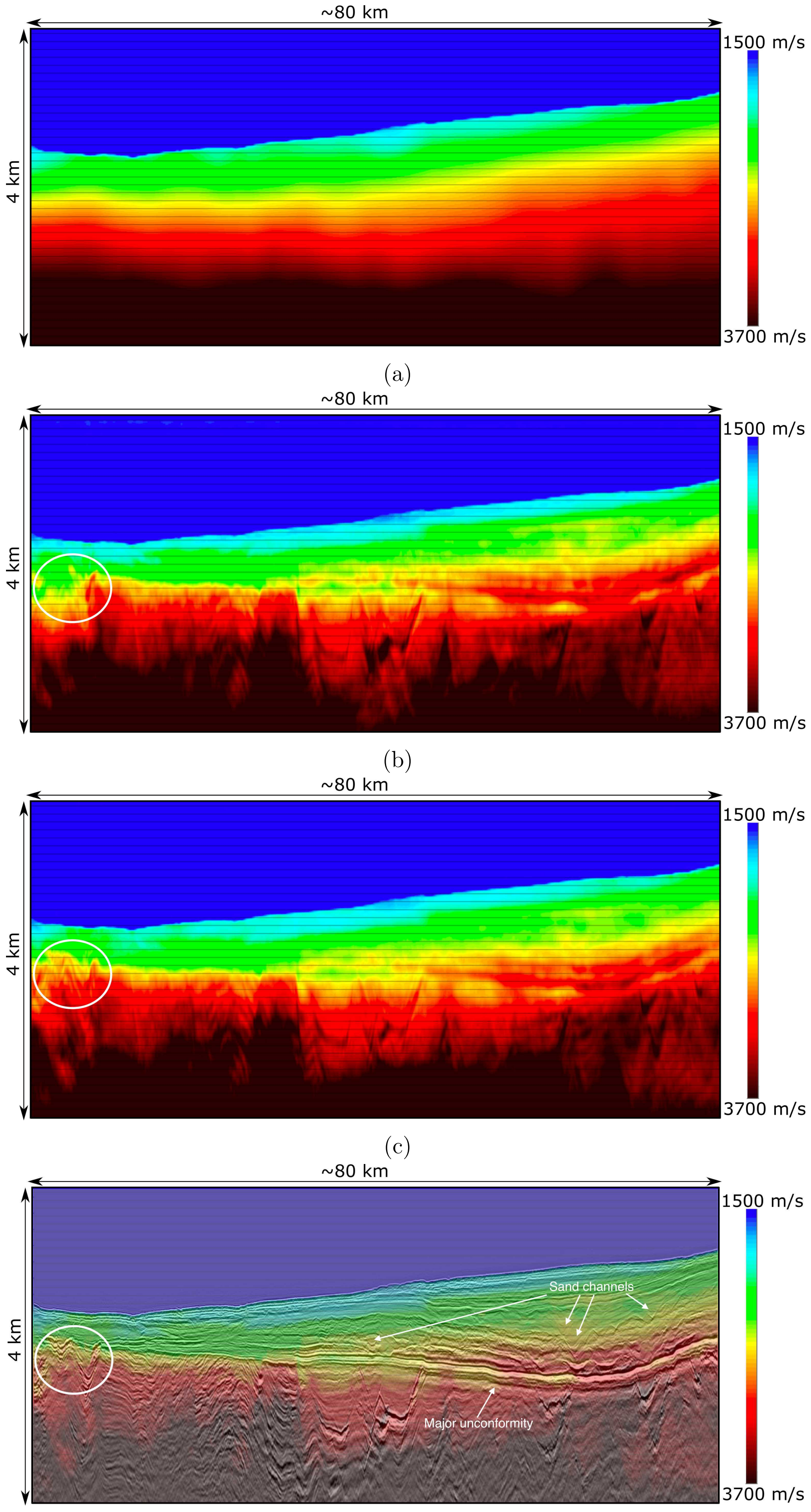
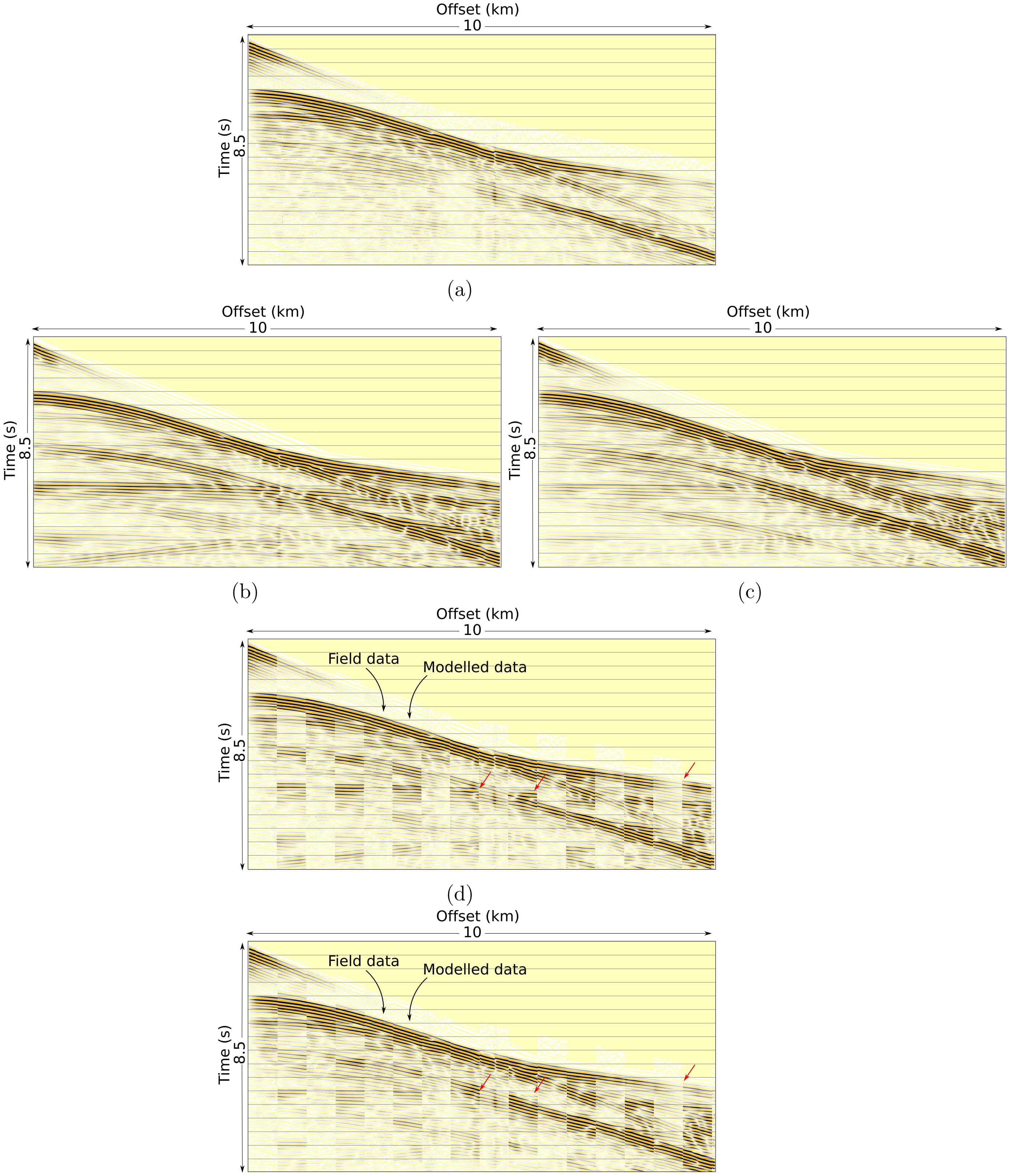


Figure 4.23: 2D vs. 3D: velocity models comparison. (a) The starting model; FWI-derived models at 12 Hz in (b) 2D and (c) 3D; (d) 3D FWI overlaid with PSDM.



(a)

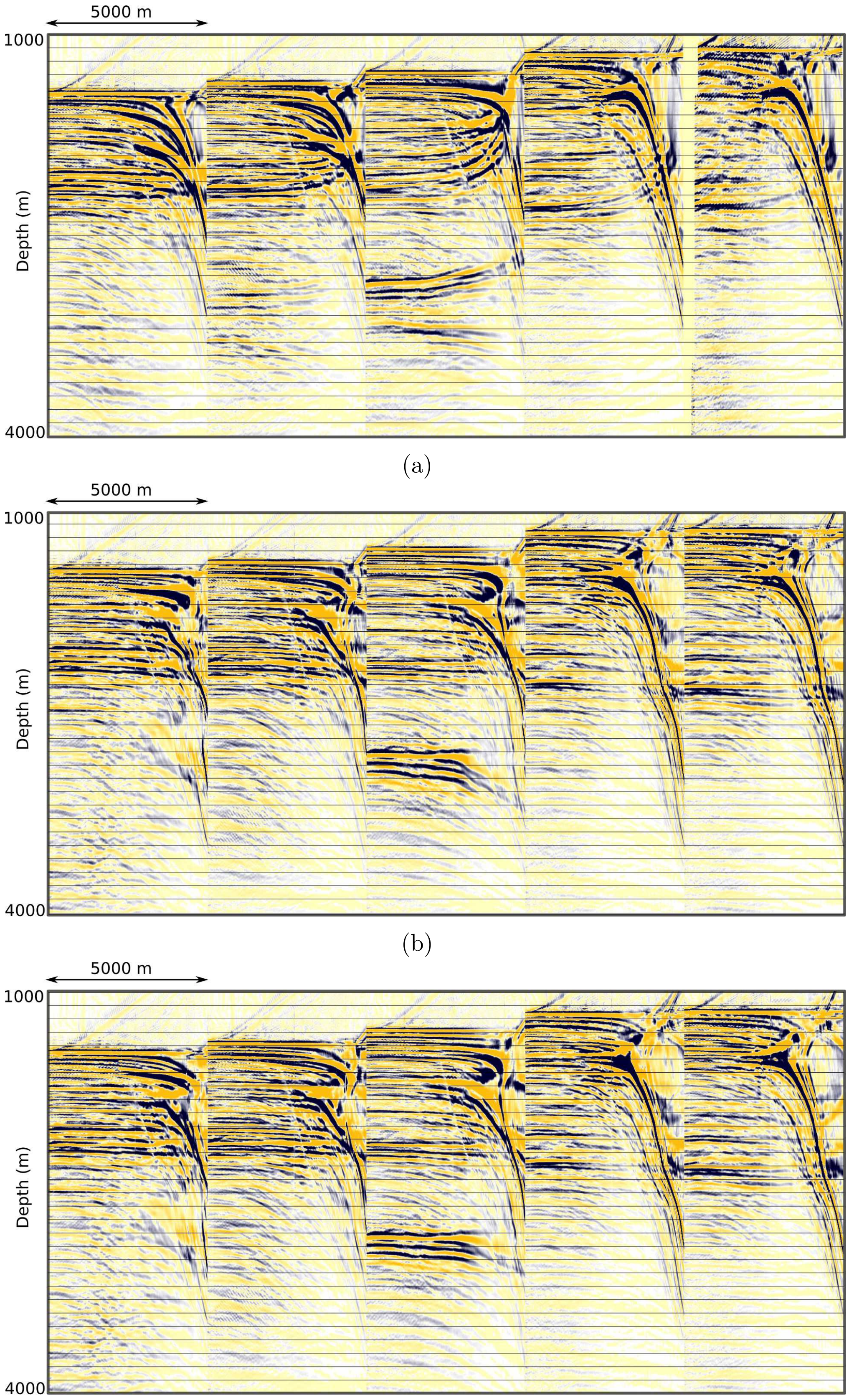
(b)

(c)

(d)

(e) Interleaved (a) and (c)

Figure 4.24: 2D vs. 3D: shot gather comparison. An example (a) field, (b) modelled with 2D-derived FWI velocity model and (c) modelled with 3D-derived FWI velocity model shot gather. (d) Interleaved (a) and (b) shot gathers; (e) Interleaved (a) and (c) shot gathers. The blocks of 50 traces are sequentially displayed. The shots are scaled trace-by-trace. The shot gather with 3D-derived FWI velocity model has a slightly better match to the raw data rather than the shot with 2D-derived FWI velocity model.

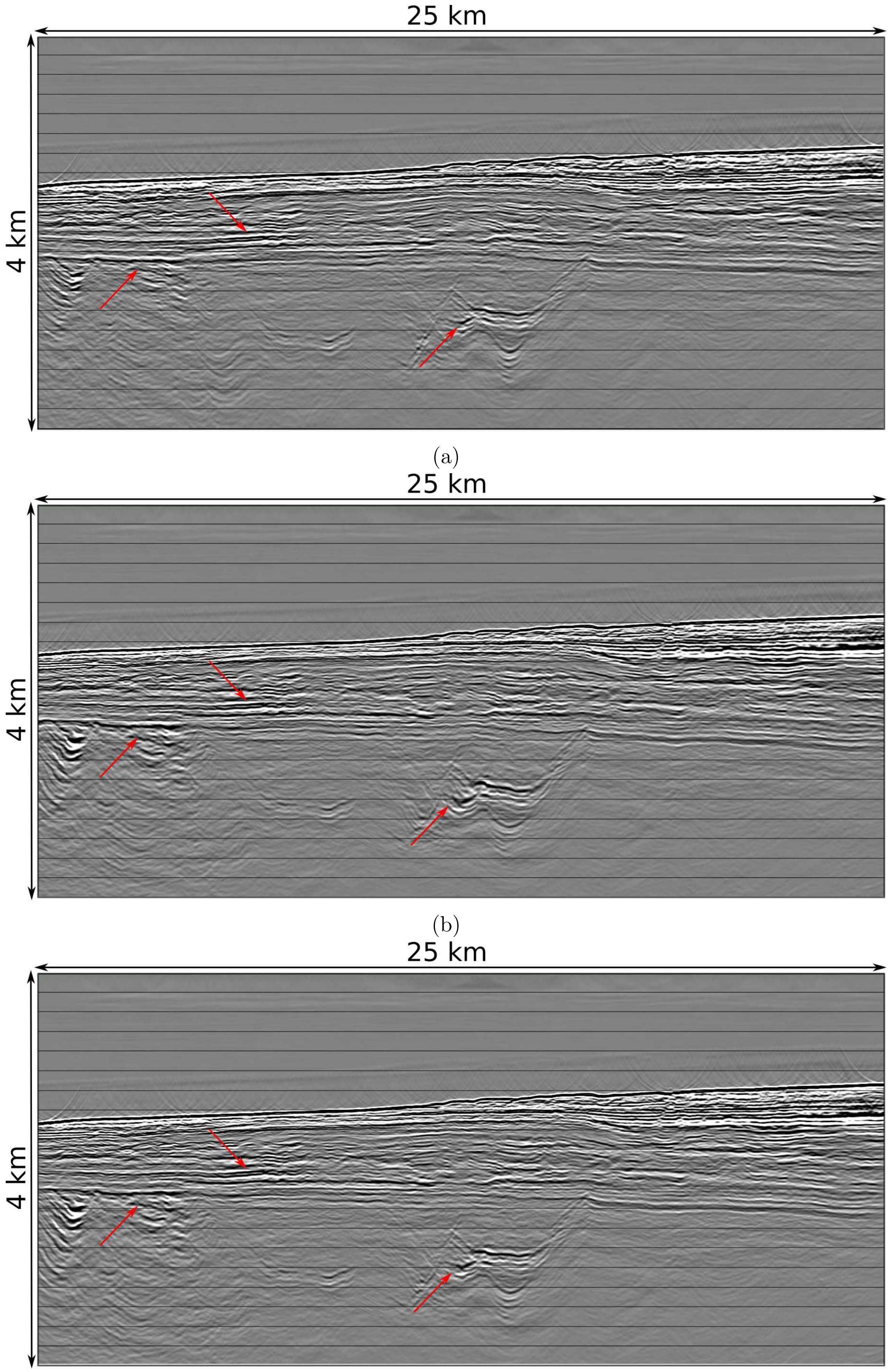


(a)

(b)

(c)

Figure 4.25: Common image gathers generated using: (a) the starting model, (b) the 2D-derived FWI, (c) the 3D-derived FWI. The starting model provides flat gathers at the offsets to about 3 km. FWI greatly improved the migrated gathers at long offsets. There is little or no at all difference between gathers obtained using 2D and 3D FWI models.

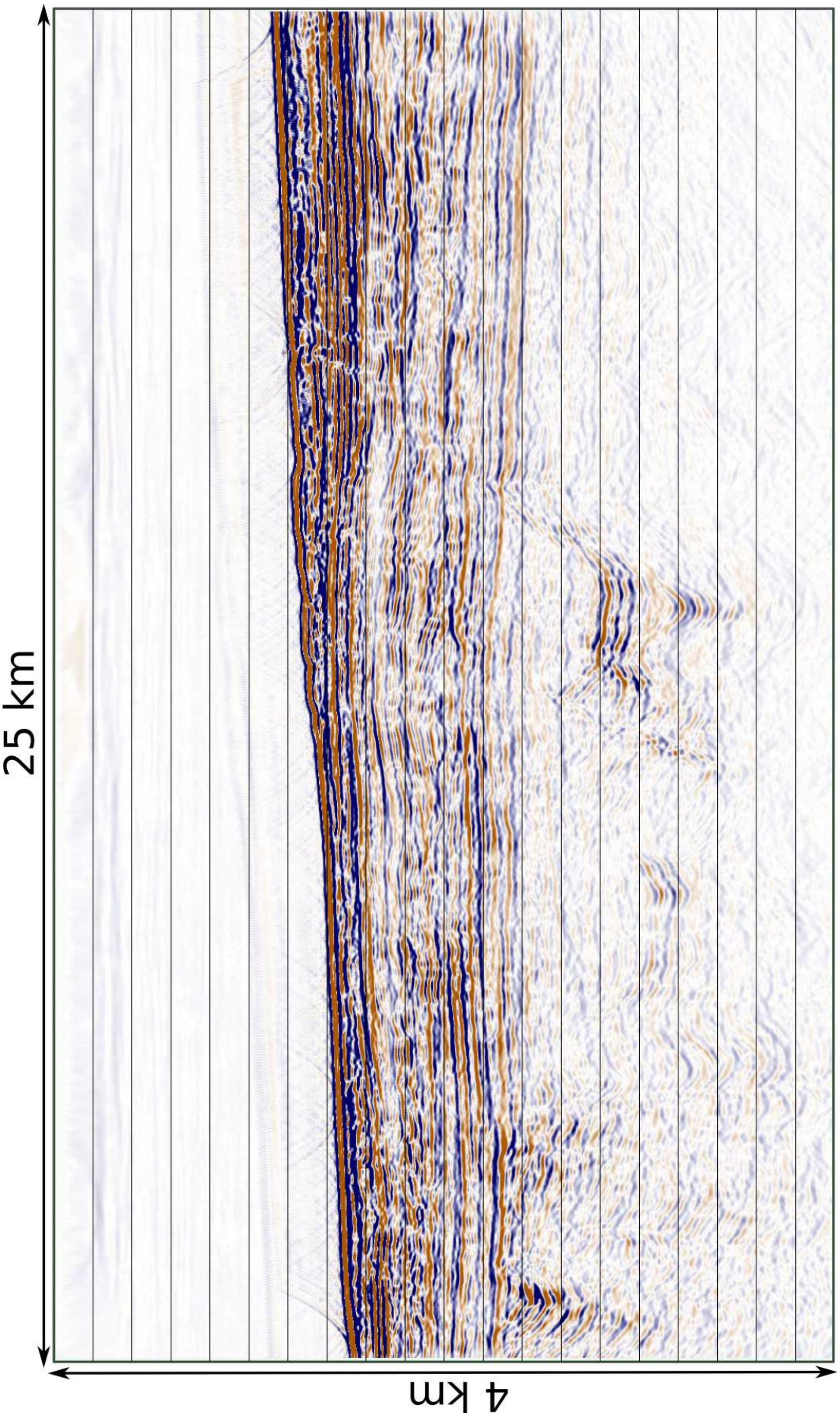


(a)

(b)

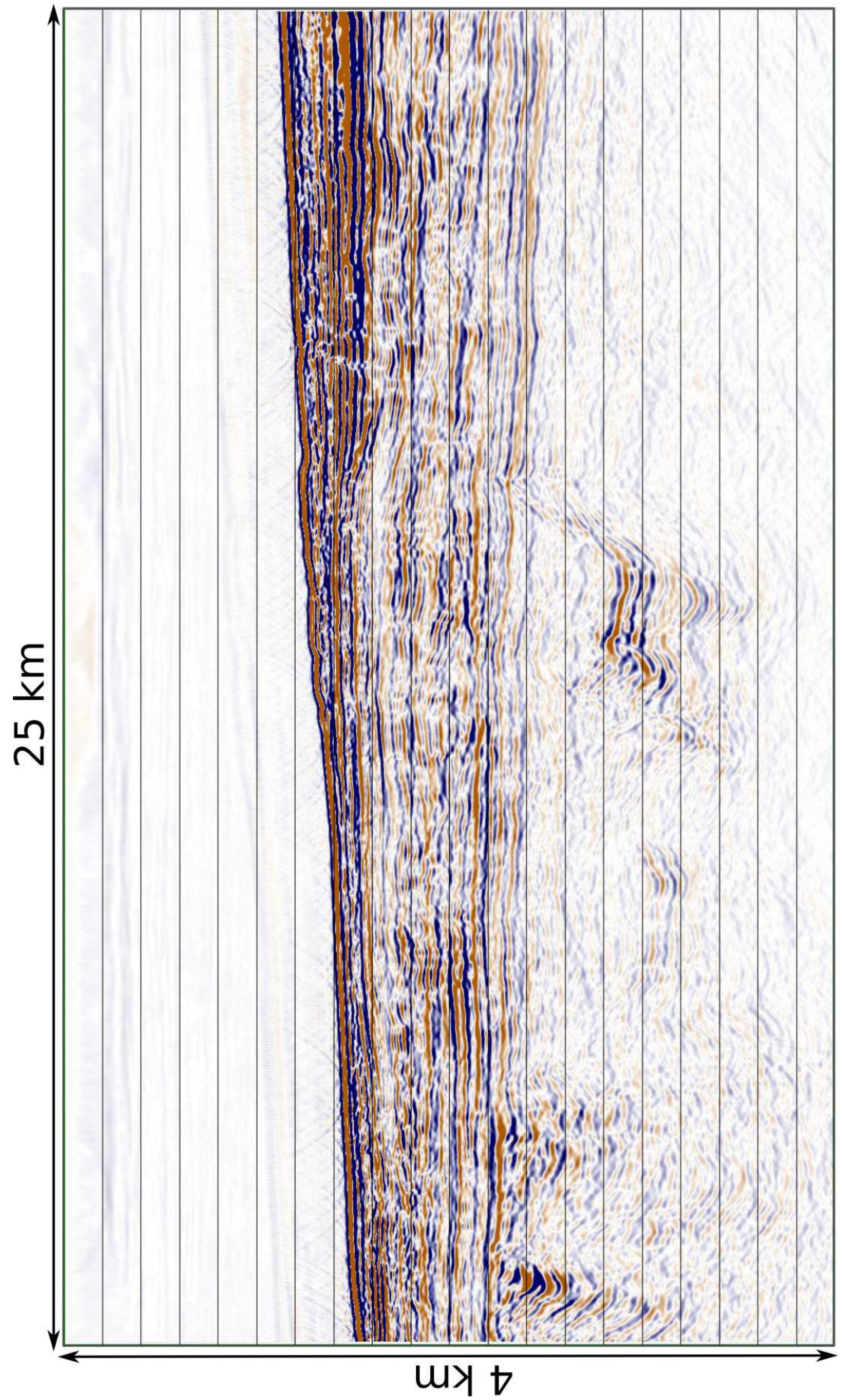
(c)

Figure 4.26: PSDM sections of the central part of the model generated using: (a) the starting model, (b) the 2D-derived FWI, (c) the 3D-derived FWI, (d) the starting model enlarged, (e) the 2D-derived FWI enlarged and (f) the 3D-derived FWI enlarged (cont.).



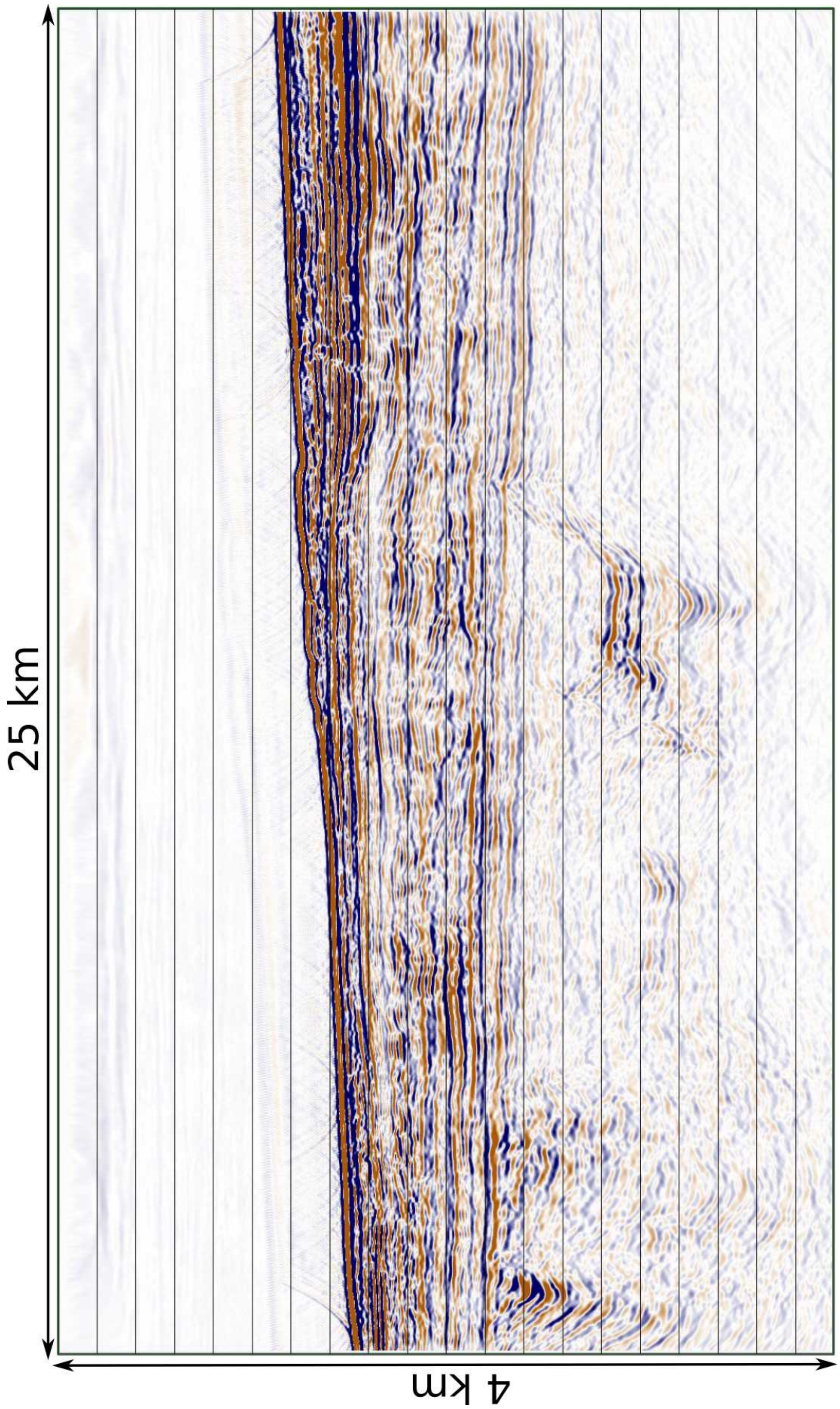
(d) PSDM using the starting model

Figure 4.26: PSDM sections of the central part of the model generated using: (a) the starting model, (b) the 2D-derived FWI, (c) the 3D-derived FWI, (d) the starting model enlarged, (e) the 2D-derived FWI enlarged and (f) the 3D-derived FWI enlarged (cont.).



(e) PSDM using the 2D-derived FWI

Figure 4.26: PSDM sections of the central part of the model generated using: (a) the starting model, (b) the 2D-derived FWI, (c) the 3D-derived FWI, (d) the starting model enlarged, (e) the 2D-derived FWI enlarged and (f) the 3D-derived FWI enlarged (cont.).



(f) PSDM using the 3D-derived FWI

Figure 4.26: PSDM sections of the central part of the model generated using: (a) the starting model, (b) the 2D-derived FWI, (c) the 3D-derived FWI, (d) the starting model enlarged, (e) the 2D-derived FWI enlarged and (f) the 3D-derived FWI enlarged.

Chapter 5

High frequency FWI

In conventional processing, the seismic data are used to produce high-resolutional images of reflections due to the impedance contrast in the medium. These images, however, can not be interpreted directly unlike volumes of the subsurface properties, such as velocity or impedance, produced by FWI. Over the last decade FWI has become a routine component of the velocity model building workflow to predict shallow hazards, derisk prospects, optimise well placements and help to estimate resources. Unlike conventional processing, that utilises the entire bandwidth of the data provided by acquisition, FWI is performed using the low-frequency content of the data, typically less than 10 Hz. To date, most of the published examples present FWI case studies at frequencies below 10 Hz while some present the case studies with the maximum frequency of 40 Hz (for example, Ruth et al. (2017)). Consequently, the question arises about FWI potential at the full bandwidth of the acquired data.

5.1 Rational

FWI is an optimisation technique that produces high-resolutional volumes of physical properties of the subsurface. Because it honours full physics of finite-frequency wave propagation, its spatial resolution is limited only by (Warner et al. 2013):

- the noise level
- source and receiver distribution

- the local seismic wavelength

Theoretically, given the data provide reasonable ray coverage and have relatively low noise level, FWI is able to spatially resolve features as small as half of a local wavelength.

The impact of the noise level is intuitively clear. If there are two objects that generate the data with amplitude of magnitude A and the noise level is the same order as A or higher, than the useful signal will be effectively lost within the noise level; that will make it impossible to resolve the two objects. Although, there are principal differences between coherent and incoherent noise. Whilst the incoherent noise even with relatively high amplitudes is well managed by the algorithm used in this thesis, coherent noise is a problem that must be eliminated prior to the inversion.

There is no sensible theory published about the impact of the geometry, i.e. source spacing and receiver spacing, on the FWI resolution. In practice, one or the other needs to sample half of a wavelength with at least two grid points to successfully resolve an object of half the size of this wavelength. In seismic experiments with reasonable coverage, the smallest of the two spacings matter. In reality, the earth model contains scatterers or reflectors that act as virtual sources. Although they have errors in depth position and change the location at every iteration, they increase the effective source density and the ray coverage.

The final thing that impacts FWI resolution is a local seismic wavelength λ defined as

$$\lambda = \frac{V}{f} \tag{5.1}$$

where V denotes velocity and f denotes frequency. FWI can spatially resolve objects of half a wavelength in size. The resolution varies with depth as at a given maximum frequency, the velocity usually gradually increases with depth, therefore, increasing the wavelength and decreasing the resolution. Consequently, the resolution is defined by the maximum frequency contained in the data.

For the study performed on the Pivot dataset and presented in chapter 4, the maximum frequency used was 12 Hz with the shallow sediment velocity of 1600 m/s and the sediment velocity of 4200 m/s at approximately 4 km depth. Hence, the objects of about 67 m were resolved just below the water bottom and the objects of about 175 m were resolved at the depth of 4 km.

Theoretically, the full frequency bandwidth retained in data prior to the inversion, can provide the same degree of resolution as conventional seismic processing (which provides resolution of half a local wavelength but, unlike FWI, utilises very high frequencies). For example, if inverting the Pivot dataset whilst the frequency of 100 Hz is kept within the data, objects as little as 8 m and 21 m could possibly be resolved immediately below the sea bottom and at 4 km depth respectively.

If possible to achieve, such a high resolution could provide significant benefits for geological interpretation. However, it might imply an increase in computational effort. Thus, an FWI model at 100 Hz will be approximately 4000 times more expensive than FWI model at 12 Hz.

In this chapter, 2D acoustic anisotropic FWI was successfully applied to a marine seismic dataset to explore the tradeoff between the value and the cost of full bandwidth FWI. The dataset was inverted to the maximum frequency of 100 Hz and the vertical derivatives of high-frequency FWI velocity model were compared to a conventional PSDM section. Also, ways to improve FWI results were explored, such as (1) incorporating Q in the inversion and (2) performing spectral shaping of the data.

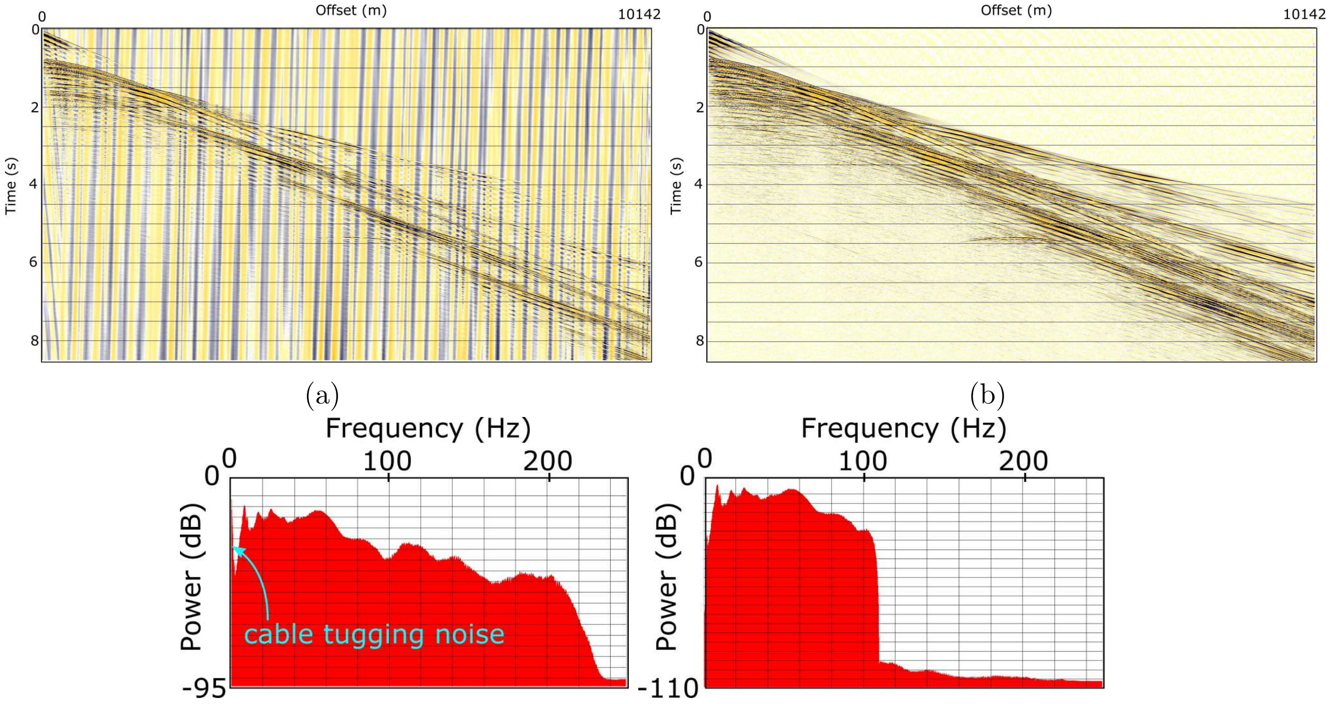
5.2 Inputs: data, source and start models

The study was conducted on the Pivot dataset acquired from the Carnarvon basin on the north-west Australian shelf. It is an 80-km 2D seismic line obtained with a single streamer and a single source; the full description of the dataset is presented in chapter 3.

The Pivot dataset was used for the study in chapter 4. Therefore, the data preprocessing, starting models, choice of the starting frequency, the source signature derivation and validation of the starting model will be discussed in short manner with the only purpose being to highlight the differences to those presented in details in chapter 4.

5.2.1 Data

The data pre-processing was minimal. It contained only two passes of an Ormsby minimum-phase filter in order to mitigate the very low-frequency noise associated with the background ocean swell and cable tugging, as well as to preserve the frequencies below 100 Hz. Figure 5.1(a) and 5.1(b) show a raw shot gather and a filtered shot gather using an Ormsby filter rolling off from 100 to 110 Hz. Figure 5.1(c) and 5.1(d) demonstrate the amplitude spectra of both shot gathers. The spectrum of the low-passed gather shows no presence of the very-low frequency content below 2 Hz and no presence of frequencies above 100 Hz. Otherwise, the two spectra are identical. The data shown in Figure 5.1(b) were used for the inversion. The multiples and surface ghost were retained within the data. Both refracted and reflected energy present in the data were used during the inversion. The inversion was performed in 2D and the amplitudes were not explicitly scaled for 2D.



(a) (b)

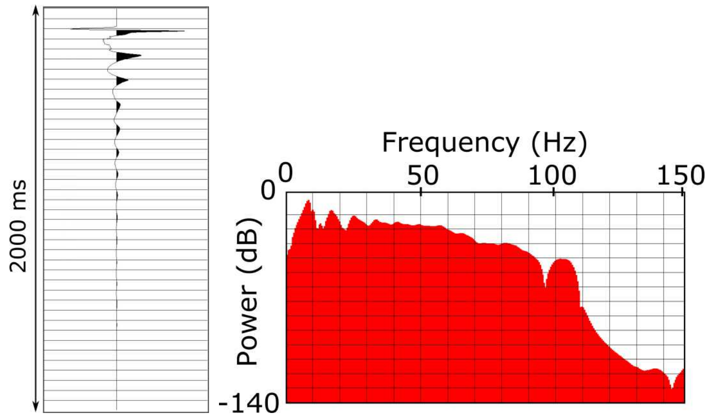
(c) (d)

Figure 5.1: Shot records: (a) raw and (b) filtered below 100 Hz and their amplitude spectra (c) and (d) respectively.

5.2.2 Source wavefield

The appropriate 2D source signature was obtained using an initial estimate to model the direct arrival through the water, subsequently using a matching Wiener filter to match this to the observed direct arrival. The approach is described step-by-step in chapter 4. The only difference for this study is that the filter applied to both direct arrival and the initial guess source was a minimum phase Ormsby filter rolling off from 100 to 110 Hz rather than from 9 to 12 Hz.

Figure 5.2 shows the derived source wavefield and its amplitude spectrum. The source signature contains a distinct oscillation pattern which is due to the bubble oscillation. The data were not de-bubbled prior to the inversion and the derived source signature appropriately accounts for that. The amplitude spectrum of the source is similar to that of the data (Fig. 5.1(d)). Both spectra have the notches at 30 Hz (the receiver notch) and at 75 Hz (the source notch) and a drop off in power above 75 Hz.



(a) (b)

Figure 5.2: (a) High-frequency source wavefield and (b) its amplitude spectrum.

5.2.3 Starting models

The starting velocity and anisotropy models were provided by the processing contractor and were described in chapter 4. The starting models were proven to be appropriate to start the inversion at 3 Hz (see chapter 4).

Because higher frequencies were included in the inversion, the starting models needed to be re-sampled to satisfy the stability and dispersion conditions. The optimal sampling rate was found to be 5 m and will be discussed later in section 5.3.

For successful FWI, the water bottom reflector (the only sharp boundary in the starting model) must be correctly positioned in depth. Because the sampling rate went down from 12.5 m to 5 m and no additional information about water depth was contained in the initial velocity model, the appropriate bathymetry was obtained and inserted in the starting model prior to the inversion.

The water depth at the source given in meters was provided in the seg-y headers of the raw data. No information about how this depth was measured was available. Commonly, bathymetry is obtained using an echo-sounder that measures two-way travel time to the sea bed and then, usually assuming water velocity of 1500 m/s, is converted into meters. The desired grid spacing of 5 m requires the water bottom position in depth to have an error no more than 5m. Consequently, the water depth must be verified one way or another.

Given the offset $Offset$, water velocity V, the source and receiver depth z_1 and z_2 and the sizes of source and receiver group $2x_1$ and $2x_2$, the sea bed position h with desired precision can be theoretically calculated. Figure 5.3 schematically shows a simple geometry problem used for such calculation.

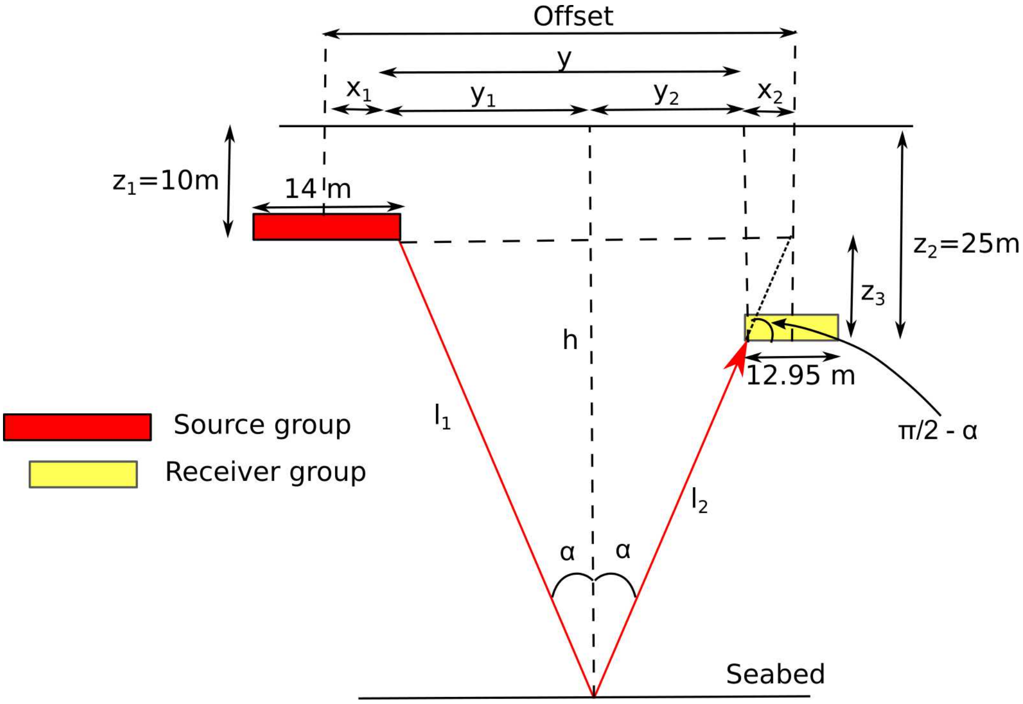


Figure 5.3: Theoretical water depth calculation.

The red and yellow rectangles denote the source and receiver groups of 14 m and 12.95 m in length respectively. The $Offset$ is the distance between the centres of the two. The calculations were performed using the nearest offset trace, therefore it was assumed that the ray travels from the last gun in the group to the first receiver in the group. Consequently, this provides the effective offset $y = Offset - x_1 - x_2$. The length of the ray path L can be written as

$$L = l_1 + l_2 = Vt \tag{5.2}$$

where t is the two-way travel time. Now, to obtain the depth of the seabed h, the ray path is estimated as

$$\begin{aligned} y_1 &= l_1 \sin\alpha \\ y_2 &= l_2 \sin\alpha \end{aligned} \tag{5.3}$$

The summation of the two gives

$$\begin{aligned} (y_1 + y_2) &= (l_1 + l_2) \sin\alpha \\ \sin\alpha &= \frac{y}{L} \end{aligned} \tag{5.4}$$

so that

$$\alpha = \sin^{-1}\left(\frac{y}{L}\right) \tag{5.5}$$

Note, that

$$\tan\left(\frac{\pi}{2} - \alpha\right) = \frac{z_3}{x_2} \tag{5.6}$$

that results in

$$x_2 = \frac{z_3}{\tan\left(\frac{\pi}{2} - \alpha\right)} \tag{5.7}$$

Now, it can be noticed that

$$\tan\alpha = \frac{y_1}{h - z_1} \tag{5.8}$$

Consequently, the sea bottom depth h can be calculated as

$$h = \frac{(Offset - x_1 - x_2)/2}{\tan\alpha} + z_1 \tag{5.9}$$

Using this equation and assuming the water velocity value of 1500 m/s and 1495m/s, the theoretical depth values were compared to the values provided in the headers. Some of the randomly selected control shot points are presented in the table 5.1. The theoretical depth values calculated with the water velocity of 1500 m/s are consistent with the provided depth values in the segy file. Nonetheless, the starting velocity model has a constant water velocity value of 1495 m/s. Thus, the water depth was re-calculated using the velocity of 1495 m/s, inserted in the starting model and smoothed with a Gaussian filter with the size of 2 wavelength.

Control point	Theoretical depth, m with V=1500m/s	Theoretical depth, m with V=1495m/s	Provided depth, m
1	1091.6	1087.9	1092
2	1555.8	1482.7	1555.2
3	992.6	989.3	989.7

Table 5.1: Bathymetry for the Pivot dataset: theoretical vs. provided.

5.3 Modelling strategy

Here, the modelling strategy is similar to the that described in chapter 4. The data were inverted using a starting frequency of 3 Hz, increasing this frequency by stages, to a maximum of 100 Hz. At this maximum frequency, a 5 m grid spacing provides more than 2.6 grid points per wavelet in the water layer and more than 3 grid points per wavelet everywhere that the velocity is above 1560 m/s. The modelling code used here is a high kernel version that is 10^{th} order accurate in space and 4^{th} order accurate in time. It requires at least 2.6 grid points per wavelet and allows the wavefield to cross no more than 66% of a grid cell at every time step. Consequently, a 5 m grid spacing generates minimal numerical dispersion and the code is stable everywhere in the model with a time step of 0.6 ms:

$$\begin{aligned} grid\_spacing &\leq \frac{1}{2.6}\frac{1495m/s}{100Hz} \leq 5.75m \approx 5m \\ time\_sampling &< \frac{2}{3}\frac{5m}{4950m/s} \leq 6.7ms \approx 0.6ms \end{aligned} \tag{5.10}$$

FWI can be computationally expensive, especially at higher frequencies. By allowing the higher frequencies, the sampling of the model becomes finer and, therefore, the model becomes larger in terms of grid cells. As the frequency doubles, the cost becomes 16 times more expensive. Consequently, FWI model at 100 Hz will be approximately 4000 times more expensive than FWI at 12 Hz.

In order to save computational effort, it was decided to reduce the model size. The initial starting model with the size of 78.8 x 7.475 km was cropped to a final model size of 50.25 x 4 km. At the grid spacing of 5 m, the final starting model has dimensions of 10051 x 801 grid cells. Allowing for boundaries, this model arrives to dimensions of 10221 x 886 grid cells.

5.4 Inversion strategy

This dataset has been previously inverted as a part of this thesis. It was inverted multiple times to the maximum frequency of 30 Hz and the results were presented at SEG conference in 2017 by Kalinicheva et al. (2017). I will refer to this run as “Phase 0” and use it as the starting point for further inversions.

I have used the final FWI model from this run, smoothed in both vertical and horizontal directions and cropped horizontally and vertically in order to save computational effort, as the starting model for the following phases. Alongside with the model size, the amount of sources was reduced from 1476 to 1002.

In principle, it was possible to invert all 1002 sources simultaneously during every iteration. However, it would demand sufficient computational cost per iteration due to the model size. Instead, less shots shot per iteration were used and more iterations were performed. Similar strategy was used in chapter 4.

Here, at every iteration 501 shots were inverted simultaneously and two iterations were performed during each frequency band. That means that every shot was used once at every frequency band.

In order to optimise computational cost, the inversion was done in 2 phases. The same version of the code was used throughout the inversion at every stage.

During the first phase, the inversion started from 3 Hz and the data were inverted using 20 frequency bands with the maximum frequency of 41 Hz. This frequency range allowed a coarser grid to be used with the spacing of 12.5 m and a coarser time sampling of 1.4 ms, that significantly reduced computational cost of the 41 Hz FWI model.

The second phase of the inversion used smoothed and re-sampled to the 5 m grid 30 Hz FWI model as the starting model. Here, the data were inverted to the maximum frequency of 100 Hz. For this phase, finer time sampling of 0.6 ms was used. Only the offsets with maximum of 2500 m were used at this stage as the reflections are primary source to the updates at high frequencies rather than refractions.

In total, 56 different frequency bands and 224 iterations were performed. The lowest frequency of 3 Hz was determined by the field data and the highest frequency of 100 Hz was restricted by the chosen high-cut filter.

Modelling was acoustic rather than elastic with attenuation and elastic effects not explicitly taken into account. The inversion for slowness rather than for velocity was performed because some evidence suggests that it improves resolution and convergence rate (Warner et al. 2013). The anisotropy was fixed at every iteration during the inversion.

A summary of the inversion strategy is shown in Table 5.2.

	Phase 0	Phase 1	Phase 2	Overall
Number of shots	1476	1002		
Number of shots per iteration	738	501		
Model size	78.8 x 7.4875 km	50.25 x 4 km		
Trace length	8500 ms			
Starting frequency	3 Hz	3 Hz	30 Hz	3Hz
Maximum frequency	30 Hz	41 Hz	100 Hz	100 Hz
Total number of frequency bands	28	20	36	84
Total number of iterations	112	40	72	224
Offset range	105-10,042 m	105-10,042 m	105-2,500 m	
Grid spacing	12.5 m	12.5 m	5 m	
Model size	6305 x 600 cells	4021 x 320 cells	10051 x 801 cells	
Model size with boundaries	6309 x 604	4521 x 570 cells	10221 x 886 cells	
Time sampling	1 ms	1.4 ms	0.6	
Time samples	8499	6071	14166	
Invert for slowness	yes			
Normalise amplitudes	yes			
Spatial preconditioning	yes			
Kernel	low	high	high	
NCPUs		720	720	
Run time	14 h	3.5 h	46 h	63.5 h

Table 5.2: Summary of the inversion strategy for high frequency FWI.

5.5 Results

Anisotropic acoustic time-domain full-waveform inversion with the starting model based upon smoothed PSTM stacking velocities was applied to a 2D marine dataset. The field data had minimal pre-processing and no mutes applied. The inversion has began with the starting frequency of 3 Hz and was performed to the maximum frequency of 100 Hz. In this section I will present the FWI results and compare the FWI models to conventional PSDM sections provided by different contractors.

5.5.1 FWI models

Figure 5.4 shows the starting velocity model and FWI velocity models at 7, 13, 25, 30, 50 and 100 Hz. Figure 5.5 shows the same velocity models enlarged. The starting velocity model is very smooth, practically one-dimensional, contains no structures in it and has the only sharp reflector being the seabed. In the water column, the velocity is constant everywhere and equals 1495 m/s; in the sediments velocity gradually increases to 4200 m/s.

At 7 Hz, FWI model has recovered the major unconformity at about 2 km depth that traverses the section. It shows shallow channels and a thin layer at about 1.5 km depth in the upper part of the section as high velocity anomalies. The strong irregular reflections in the second part of the section are basaltic intrusions, they appear as high velocity features in the FWI model. All these features are picked up on both the FWI model and the PSDM section (Fig. 5.6(a)).

At 13 Hz, FWI demonstrates the same main features, but in greater detail. The sand channels become more defined with a mini-basin at approximately 2 km depth in the right-hand side of the section becoming more pronounced. The major unconformity and basalt intrusions sharpen.

At 25 Hz, FWI model shows significantly improved resolution compared to the 13 Hz FWI model. The sand channels and the mini-basin are better defined and multiple additional layers can be observed in the upper part of the section. It is important to notice, that at this frequency the shallow features immediately below the seabed appear as low-velocity anomalies in the FWI model. These features can be associated with potential drilling hazards. The major unconformity appears to have more complicated structure alongside the two new basins picked up as low velocity anomalies immediately above it. The basalt intrusions in the lower part of the section are better resolved.

At 50 Hz, FWI provides a great resolution and shows many additional thinner layers that were not picked up on the 25 Hz FWI model. Many low-velocity features and high-velocity anomalies immediately below the seabed are resolved. The sand channels and the mini basin are well-defined with multiple thinner sedimentary layers around them. In the lower part of the section, the basalt intrusions are significantly better defined compared to the 25 Hz FWI model. In the left-hand side of the section, below the major unconformity, multiple layers dipping at great angles are well identified as well as the faults. All these features are present on the PSDM section (Fig. 5.6).

At 100 Hz, the velocity model might seem to be rather similar to that of the 50 Hz FWI model. The question appears to be whether all possible information has already been extracted from the wavefield at about 40-50 Hz. To examine whether there is any use of inverting frequencies higher than that, the FWI models at 40, 50, 60, 80 and 100 Hz were carefully examined.

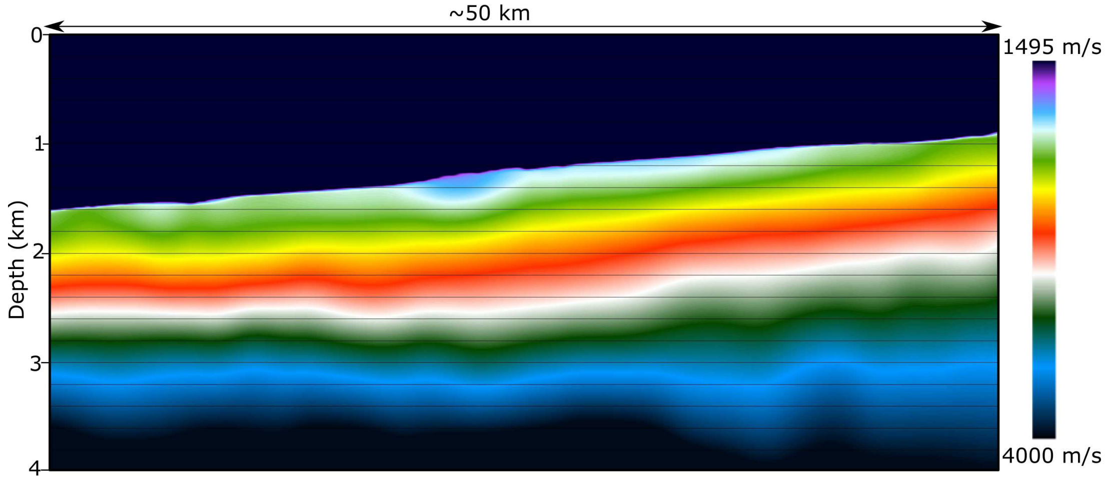
Figure 5.7 shows the comparison between these FWI models displayed in a different manner to those usually used to present such data. Rather than velocities, the derivatives of the velocities (that are in essence the impedance contrasts) are compared. The velocity models were differentiated trace-by-trace in the vertical direction. That does not turn the section into the reflectivity section produced by conventional processing, but somewhat similar and comparable.

Differentiation introduced some amount of sub-horizontal coherent noise, that can be best observed at the very bottom of the model (Fig. 5.7 (a)). This noise is coming from the starting FWI velocity model because this model was re-sampled in a data processing software that does not use linear interpolation and, therefore, introduced this artefact in the model. However, simple subtraction of the final FWI model and the starting model will not illuminate this noise because it is not exactly identical. This is because the inversion attempts to eliminate the noise and, consequently, changes it from iteration to iteration. Nonetheless, the noise might be removed by conventional signal processing techniques, such as filtration or avoid having it at all if using linear interpolation when re-sampling the model prior to the inversion.

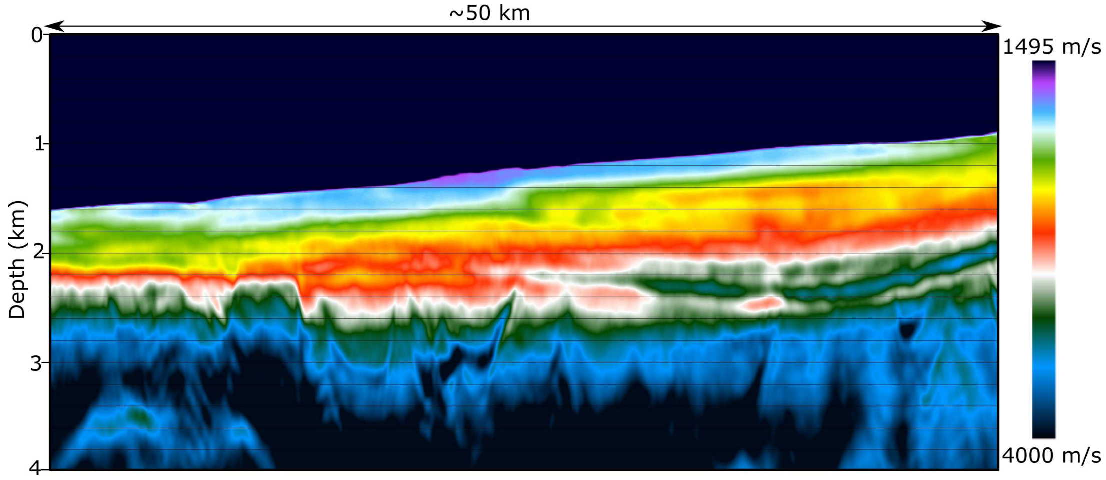
Figure 5.7(a), (b), (c), (d) and (e) compares the 40, 50, 60, 80 and 100 Hz FWI impedance contrast models. It can be seen, that FWI does not stop updating the model beyond conventionally used frequencies of 30-40 Hz. Moreover, the 100 Hz model is significantly more resolved than that of 40 Hz.

So at what frequency is there no more useful information obtained? Both FWI impedance contrast models at 80 Hz and 100 Hz are similar, but they are not identical. Figure 5.8 demonstrates close-ups of Figure 5.7(d) and (e). In red boxes, the area around an igneous intrusion is displayed with the depth ranging from 2.5 to 3 km. It is obvious, that the 100 Hz FWI demonstrates improved resolution compare to the 80 Hz FWI. The red arrow indicates an event shown on the 80 Hz model as a single reflector whilst at 100 Hz, they are two clearly separate reflectors. The yellow arrow shows that three separate reflectors appear at 100 Hz that are not visible at 80 Hz. This is potentially of remarkable value for detailed geological interpretation of reservoirs. The second close-up in Figure 5.8 shown in the green boxes demonstrates the area at the depth of approximately 2.5 km. 100 Hz FWI clearly improves the resolution of the events compared to 80 Hz FWI.

These observations prove that (1) updates in the FWI model do not stop at frequencies around 50 Hz, but continue to happen whilst using frequencies as high as 100 Hz and (2) 100 Hz FWI can bring significant benefits to detailed reservoir interpretation.

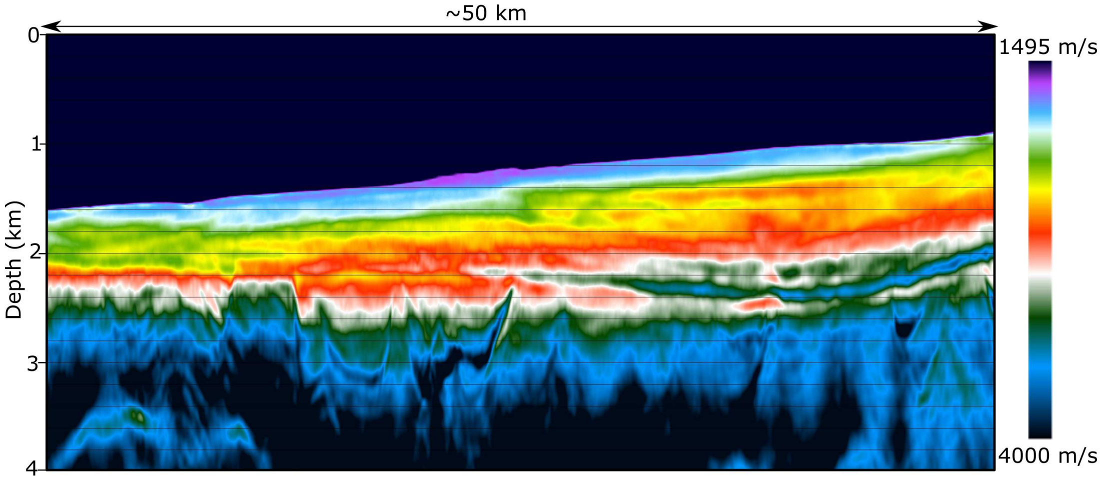


(a) Starting velocity model

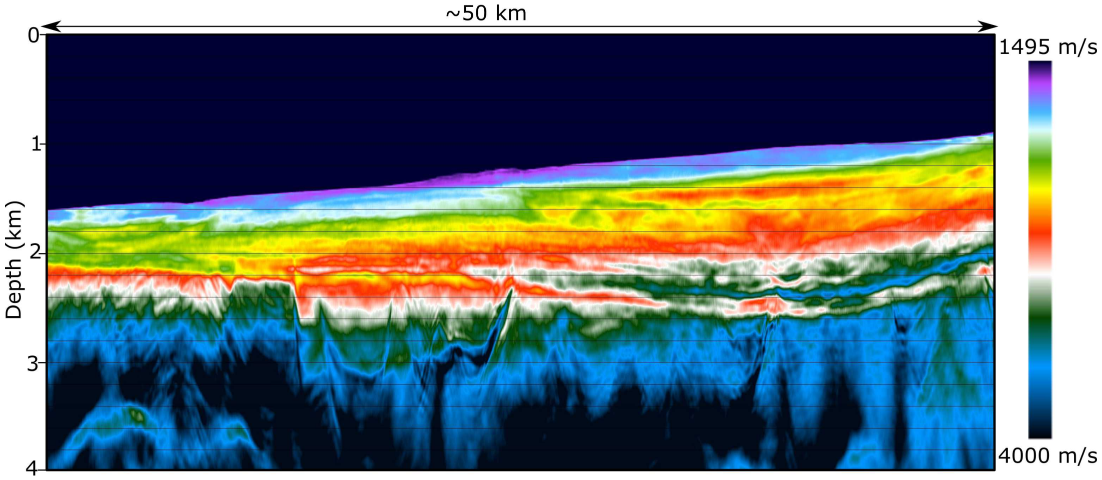


(b) 7 Hz FWI

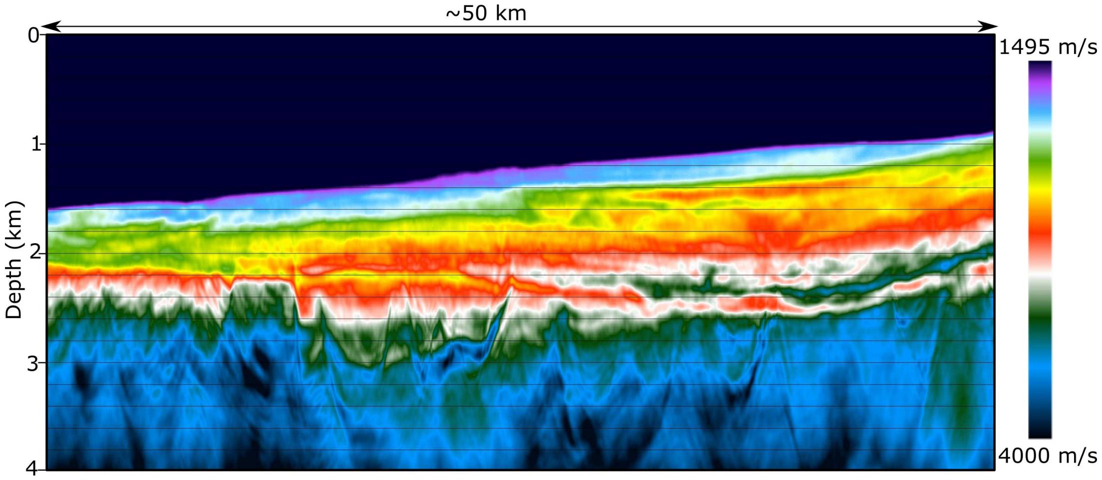
Figure 5.4: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz (cont.).



(c) 13 Hz FWI

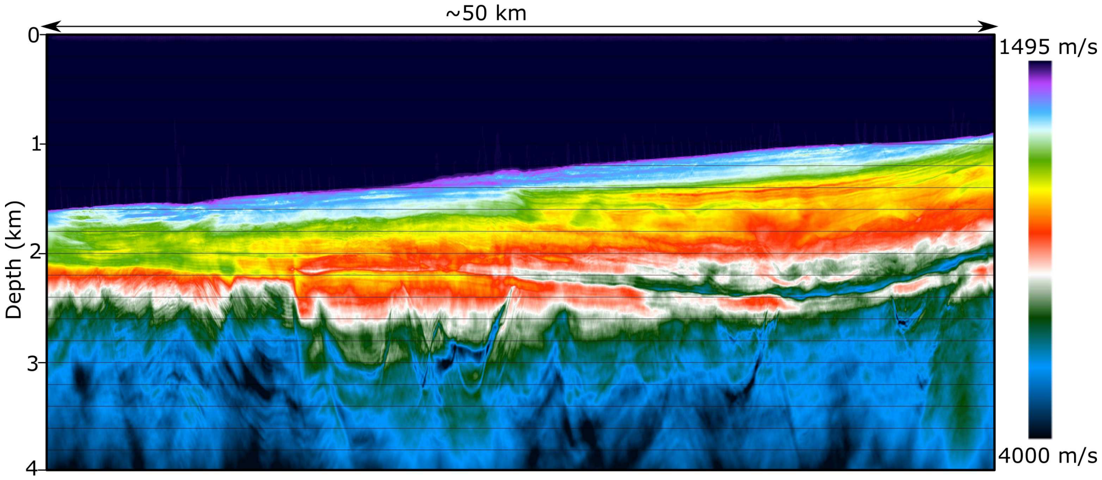


(d) 25 Hz FWI

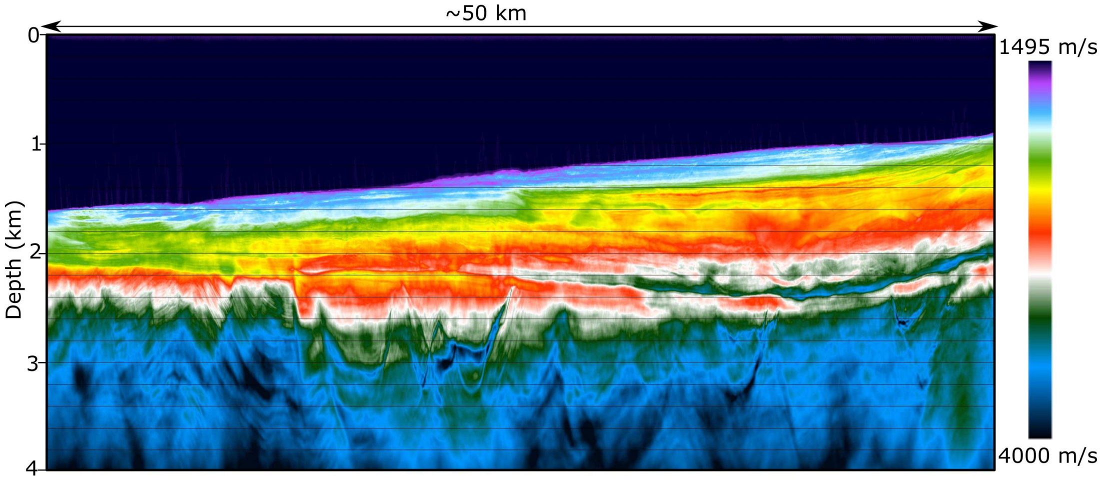


(e) Heavily smoothed 30 Hz FWI model- the starting model for Phase 1 of the inversion

Figure 5.4: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz (cont.).

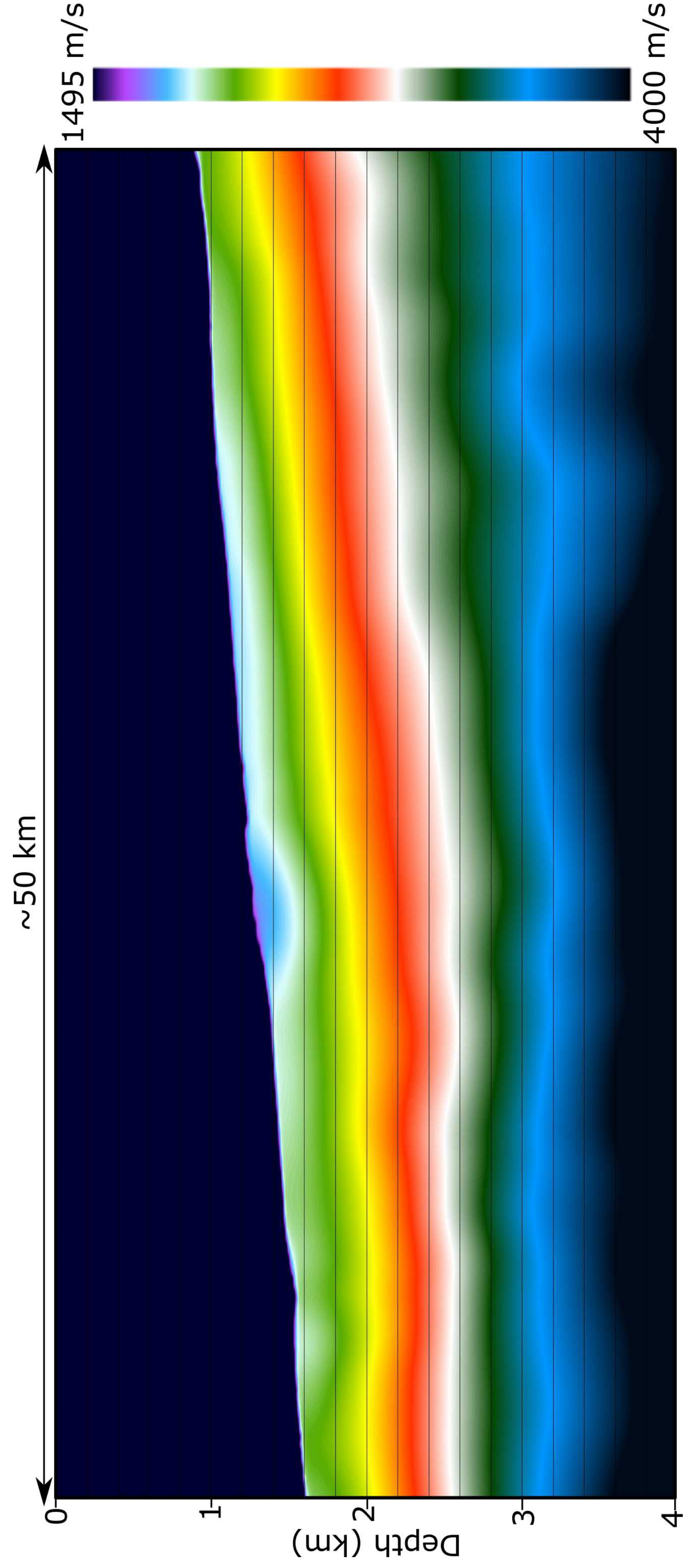


(f) 50 Hz FWI



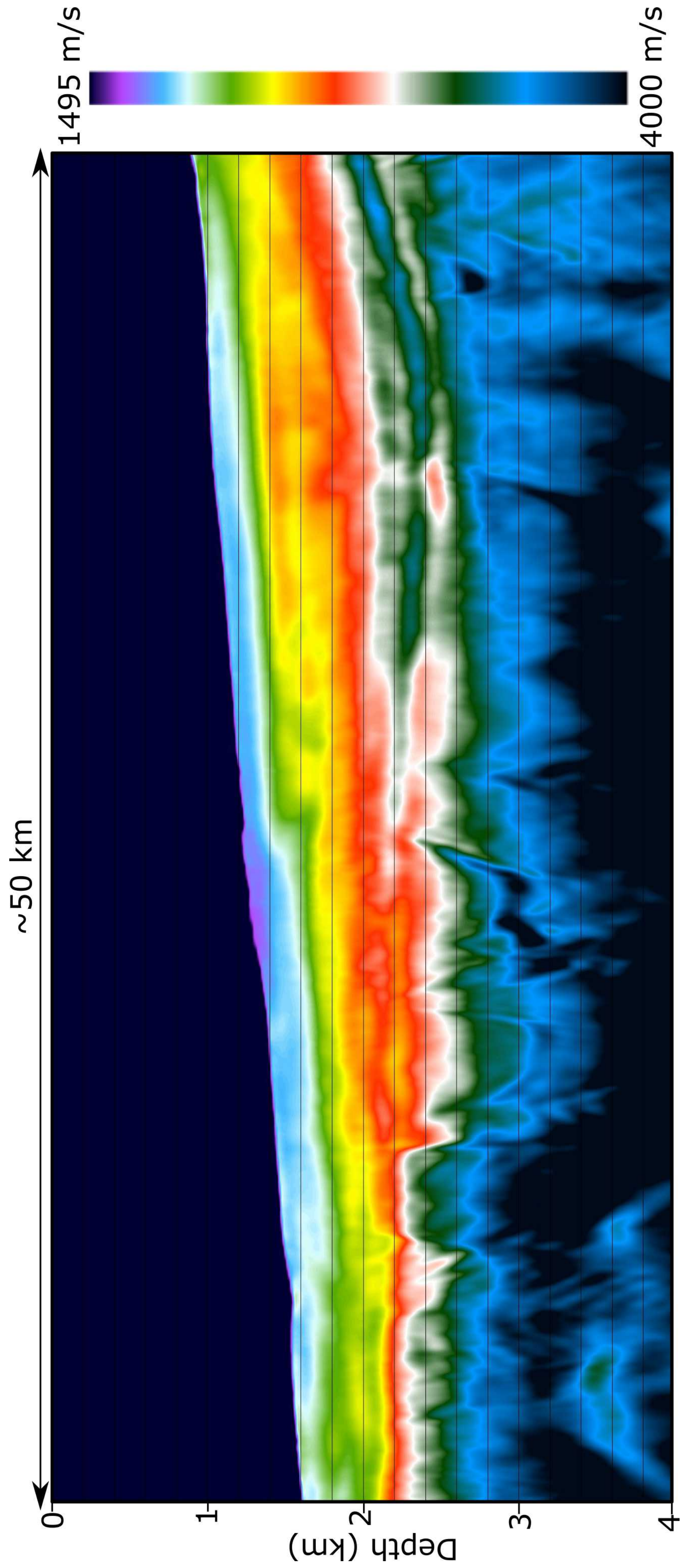
(g) 100 Hz FWI. About 4000 times the cost of the 7 Hz FWI.

Figure 5.4: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz.



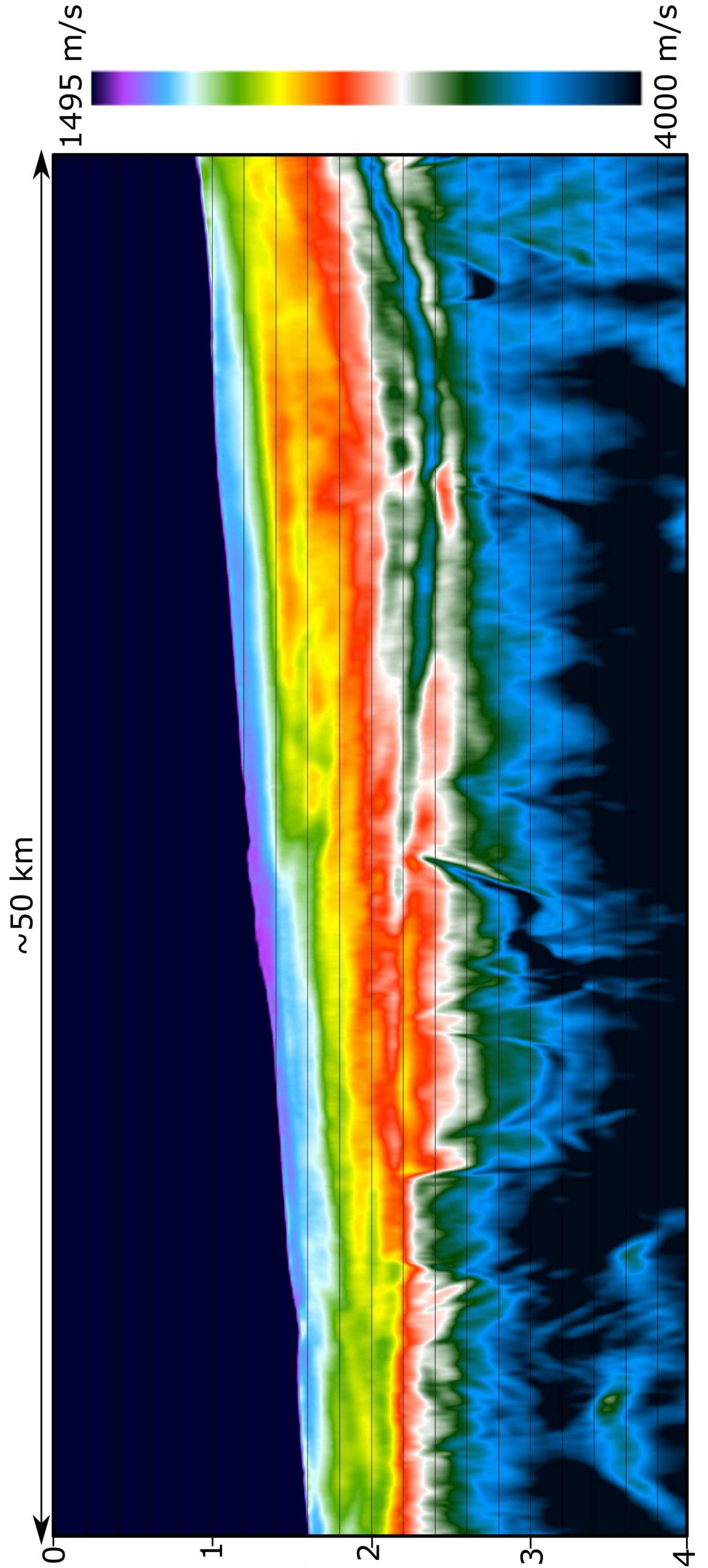
(a) Starting velocity model

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz. (cont.).



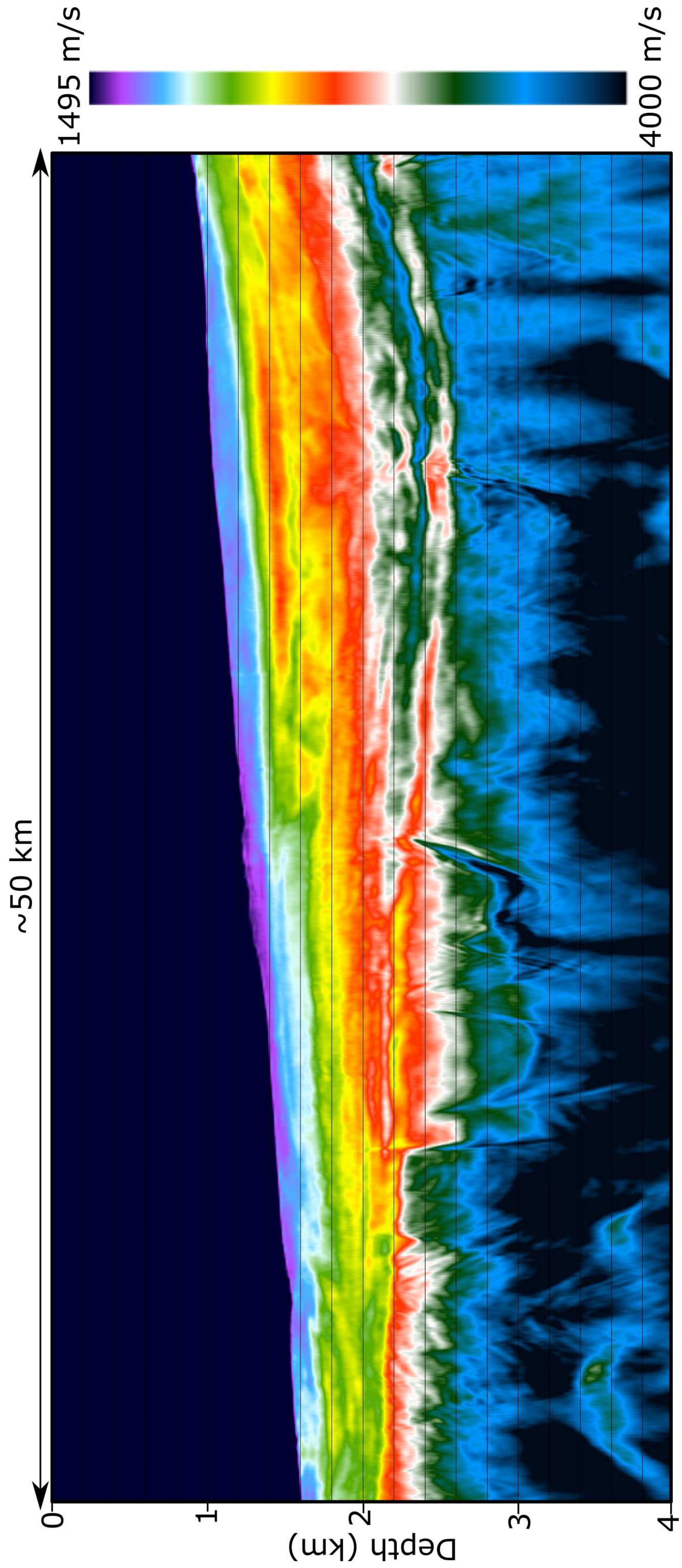
(b) 7 Hz FWI

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz. (cont.).



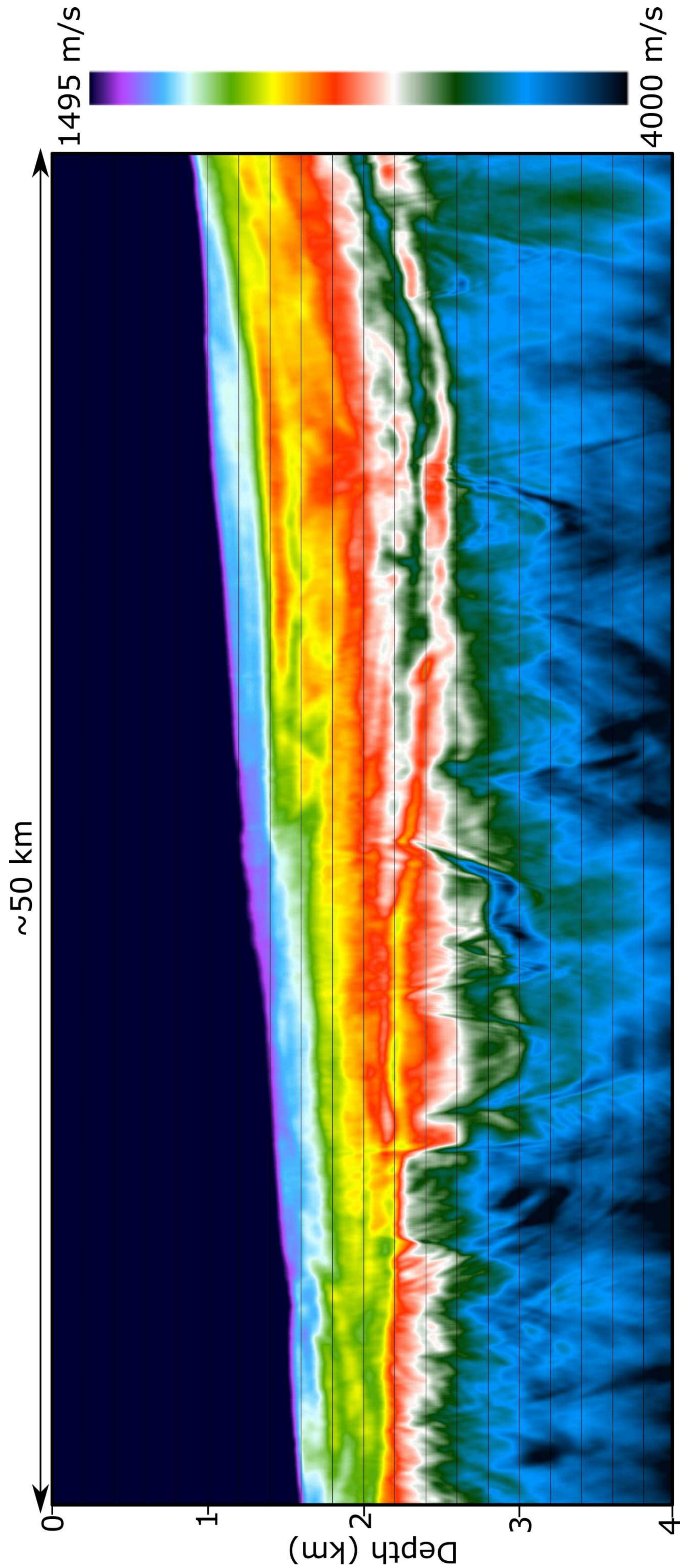
(c) 13 Hz FWI

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz. (cont.).



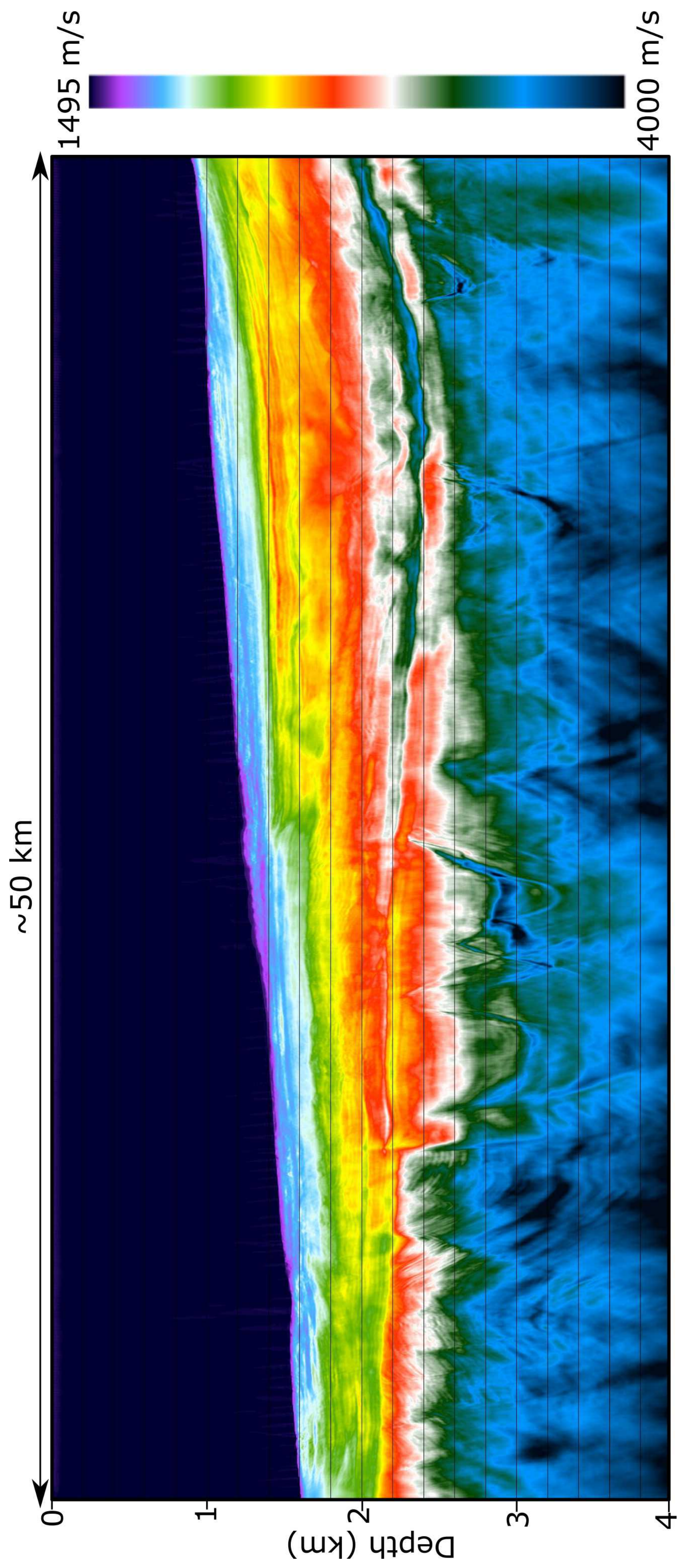
(d) 25 Hz FWI

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz. (cont.).



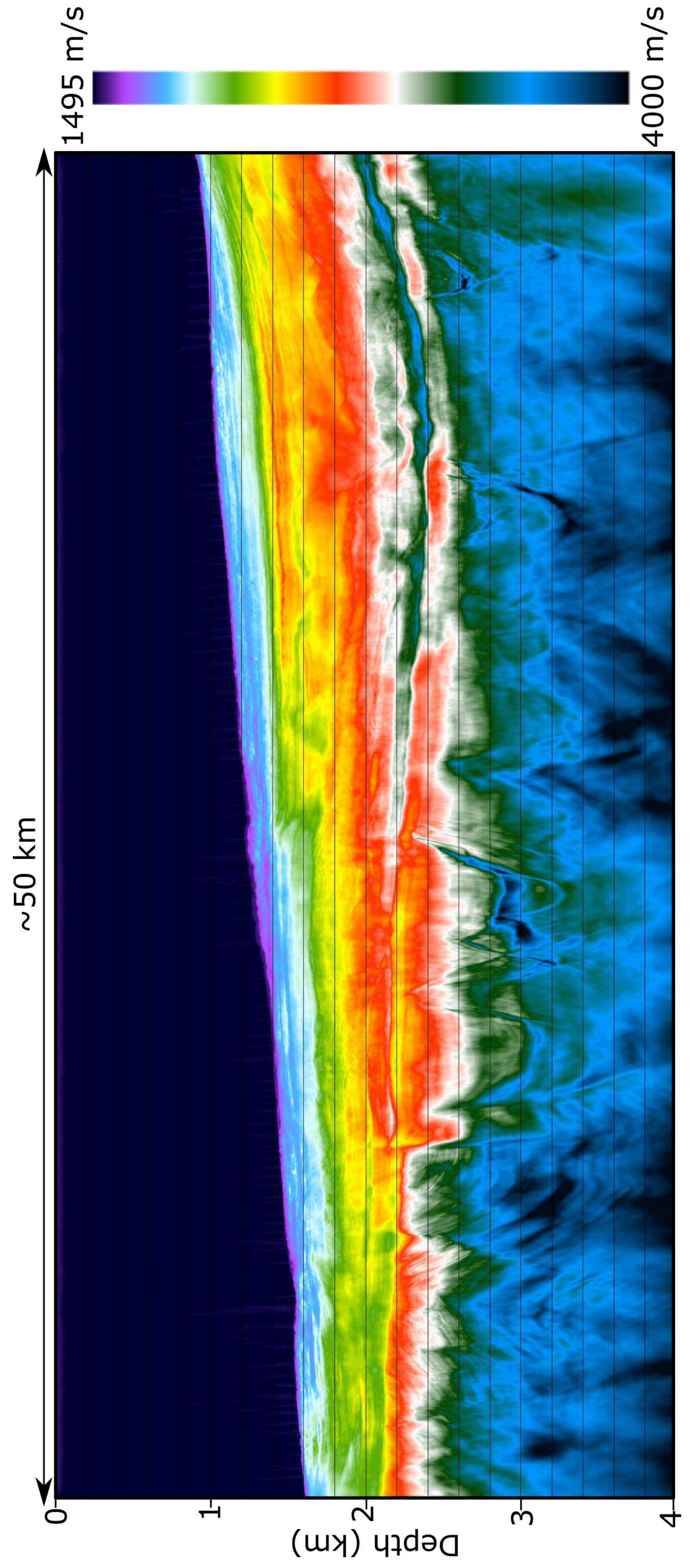
(e) Heavily smoothed 30 Hz FWI model- the starting model for Phase 1 of the inversion

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz. (cont.).



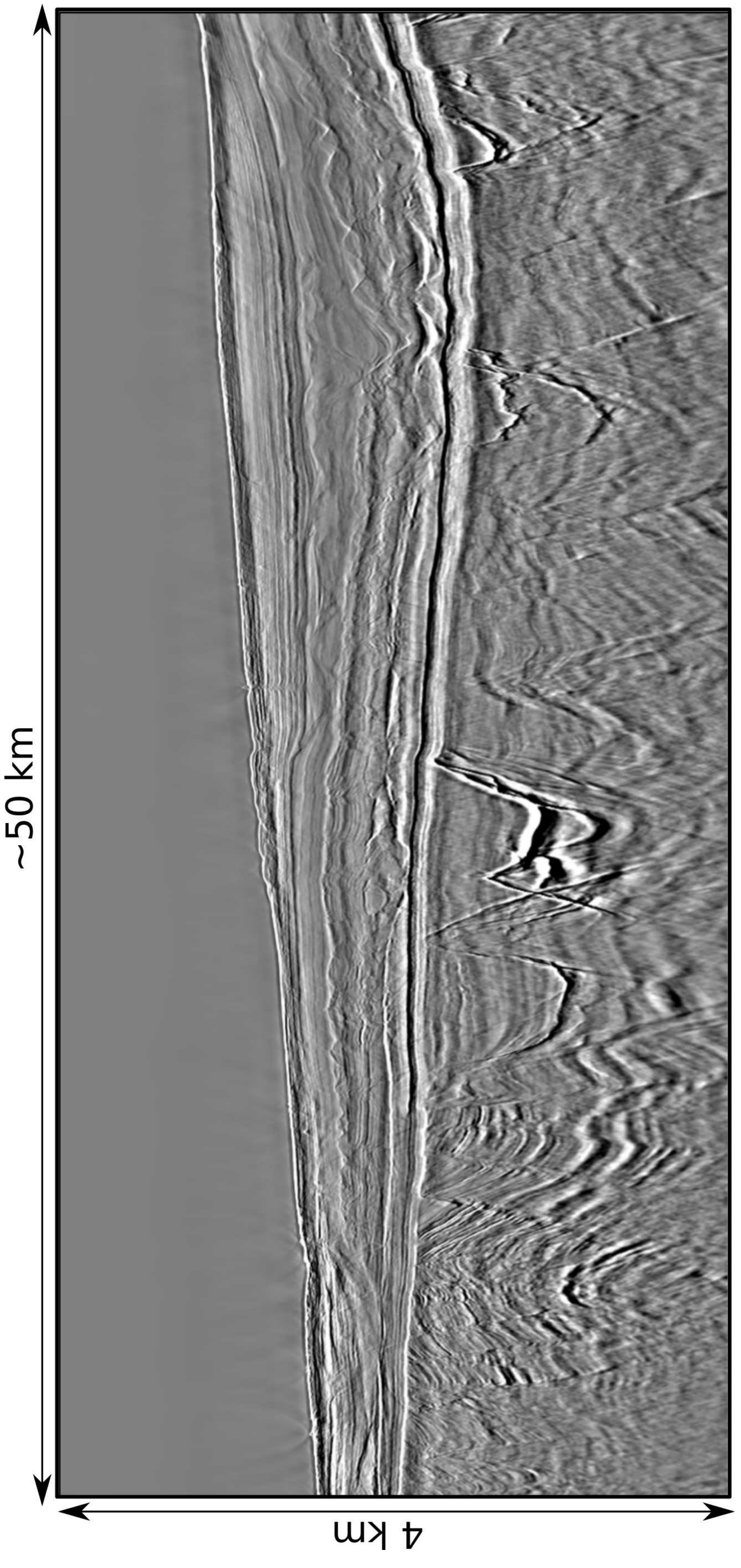
(f) 50 Hz FWI

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz. (cont.).



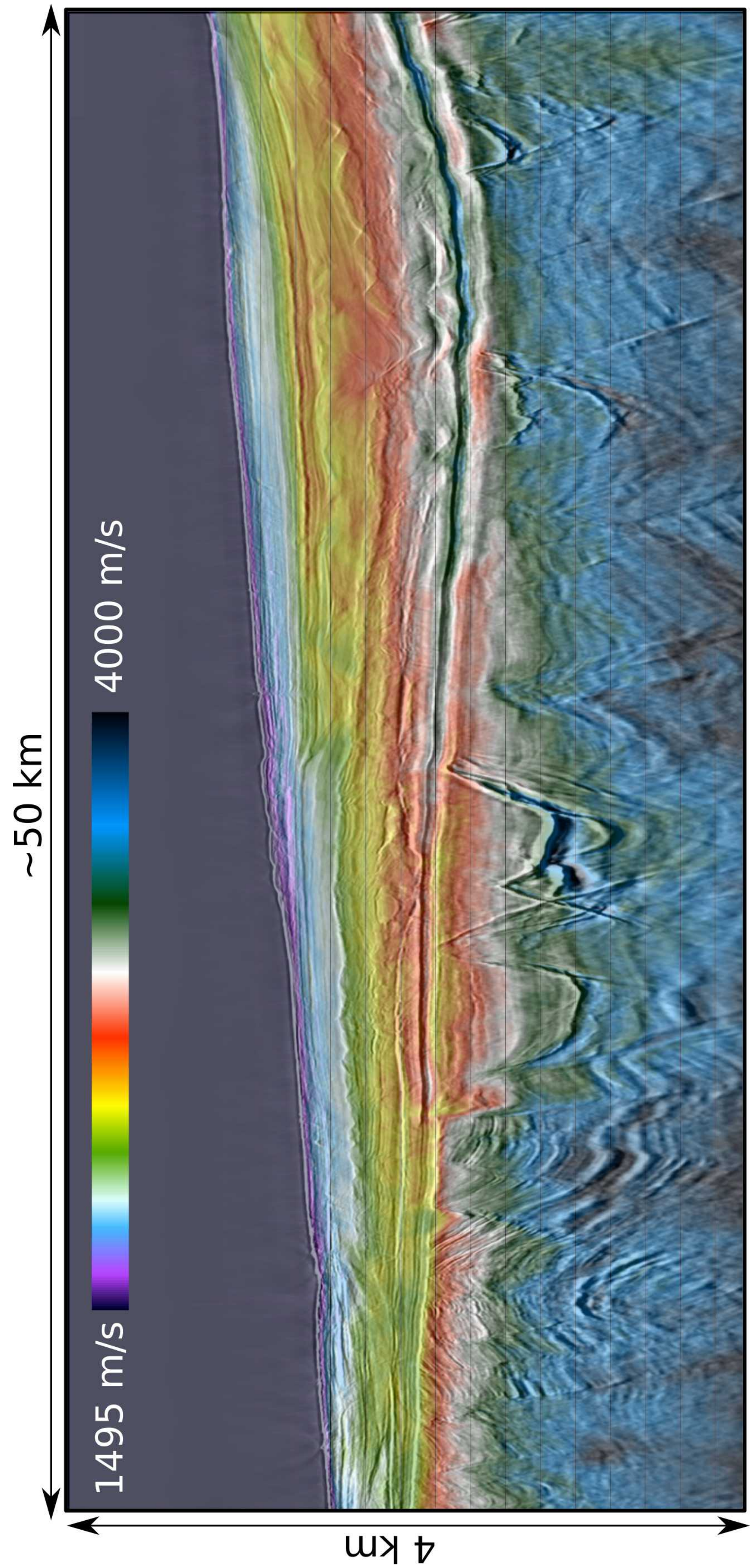
(g) 100 Hz FWI

Figure 5.5: Velocity models: (a) starting and FWI at (b) 7 Hz, (c) 13 Hz, (d) 25 Hz, (e) heavily smoothed 30 Hz model - starting model for the Phase 1, (f) 50 Hz and (g) 100 Hz.



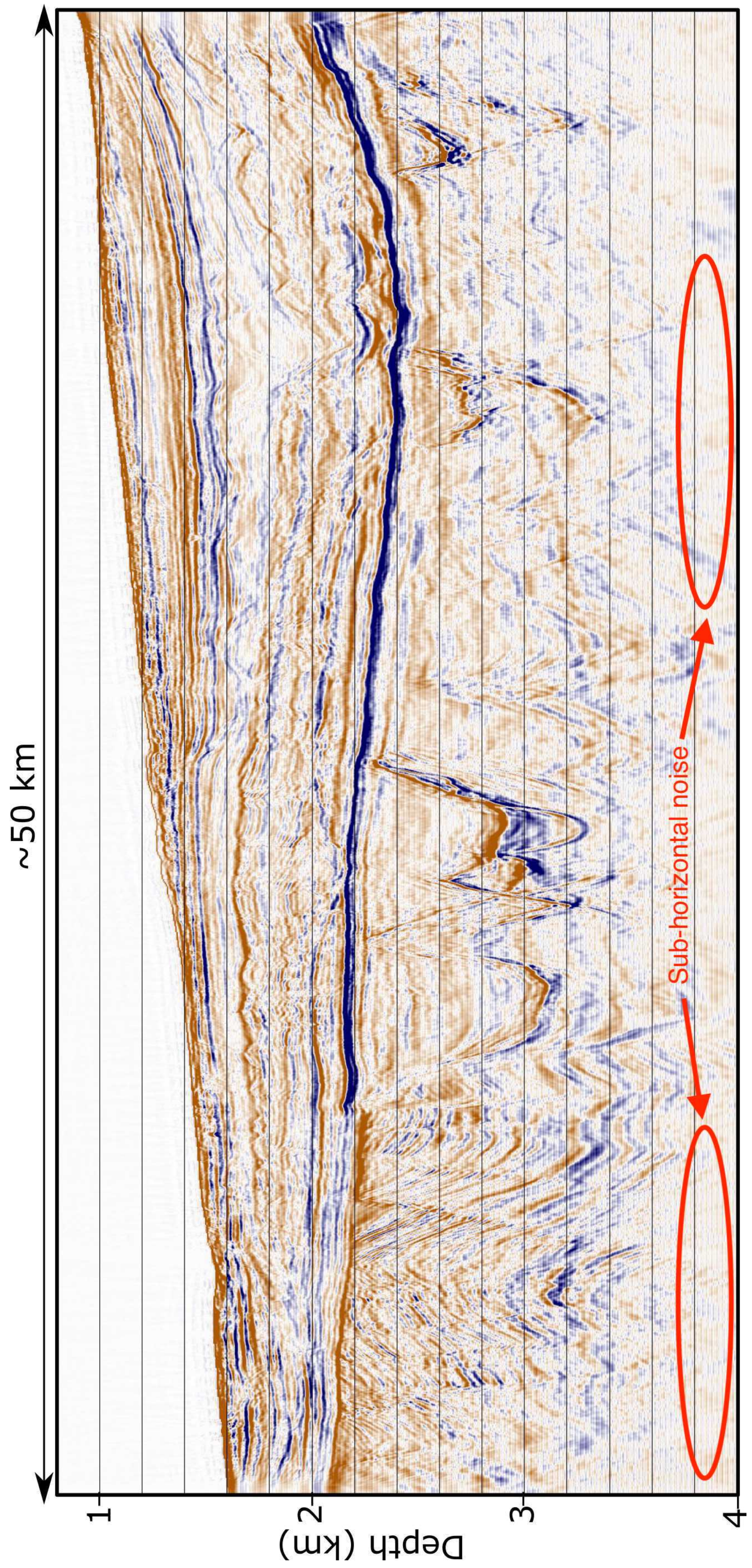
(a) PSDM section

Figure 5.6: (a) PSDM section and (b) 50 Hz FWI model overlaid with (a) (cont.).



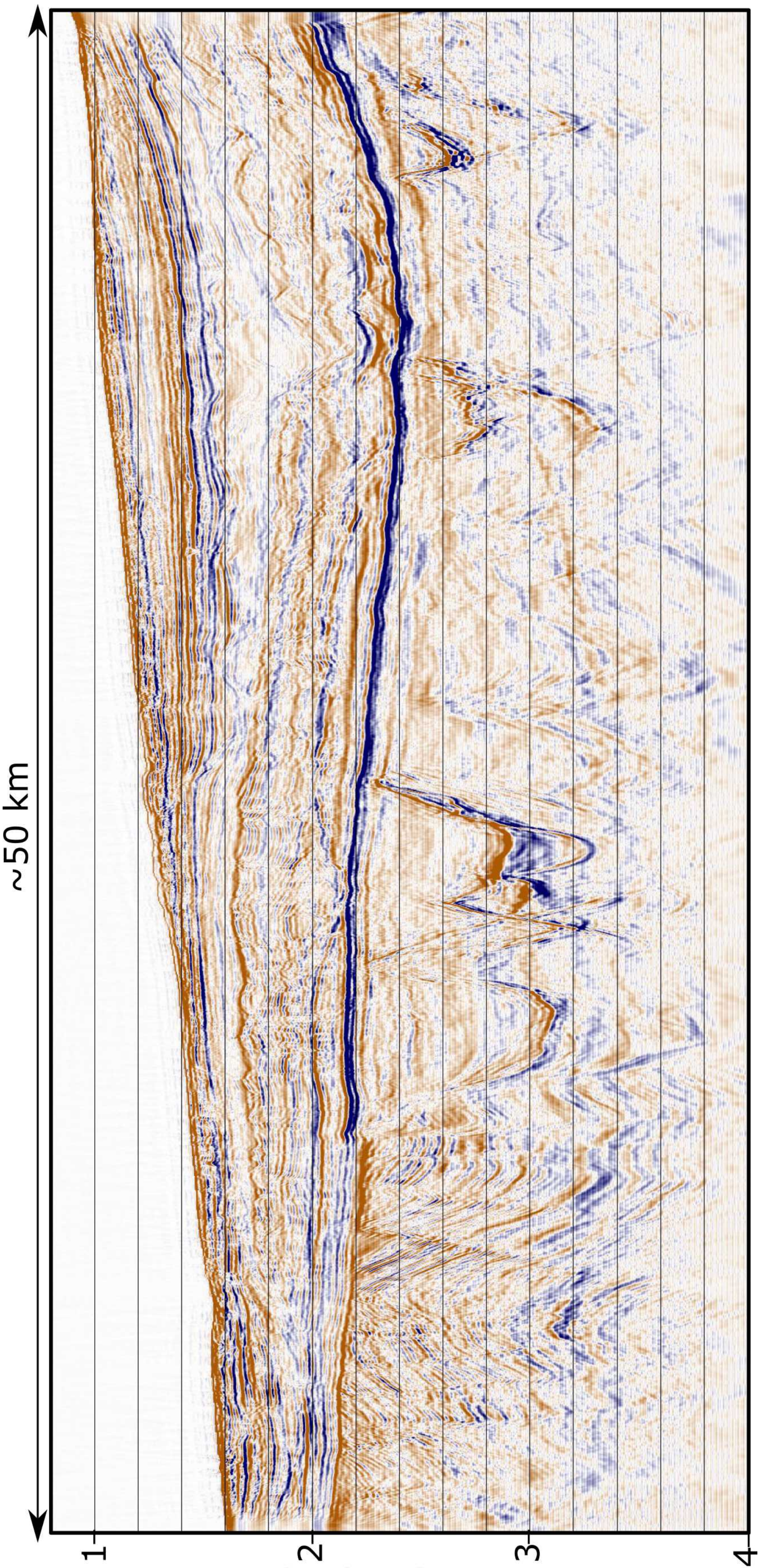
(b) 50 Hz FWI overlaid with PSDM section

Figure 5.6: (a) PSDM section and (b) 50 Hz FWI model overlaid with (a).



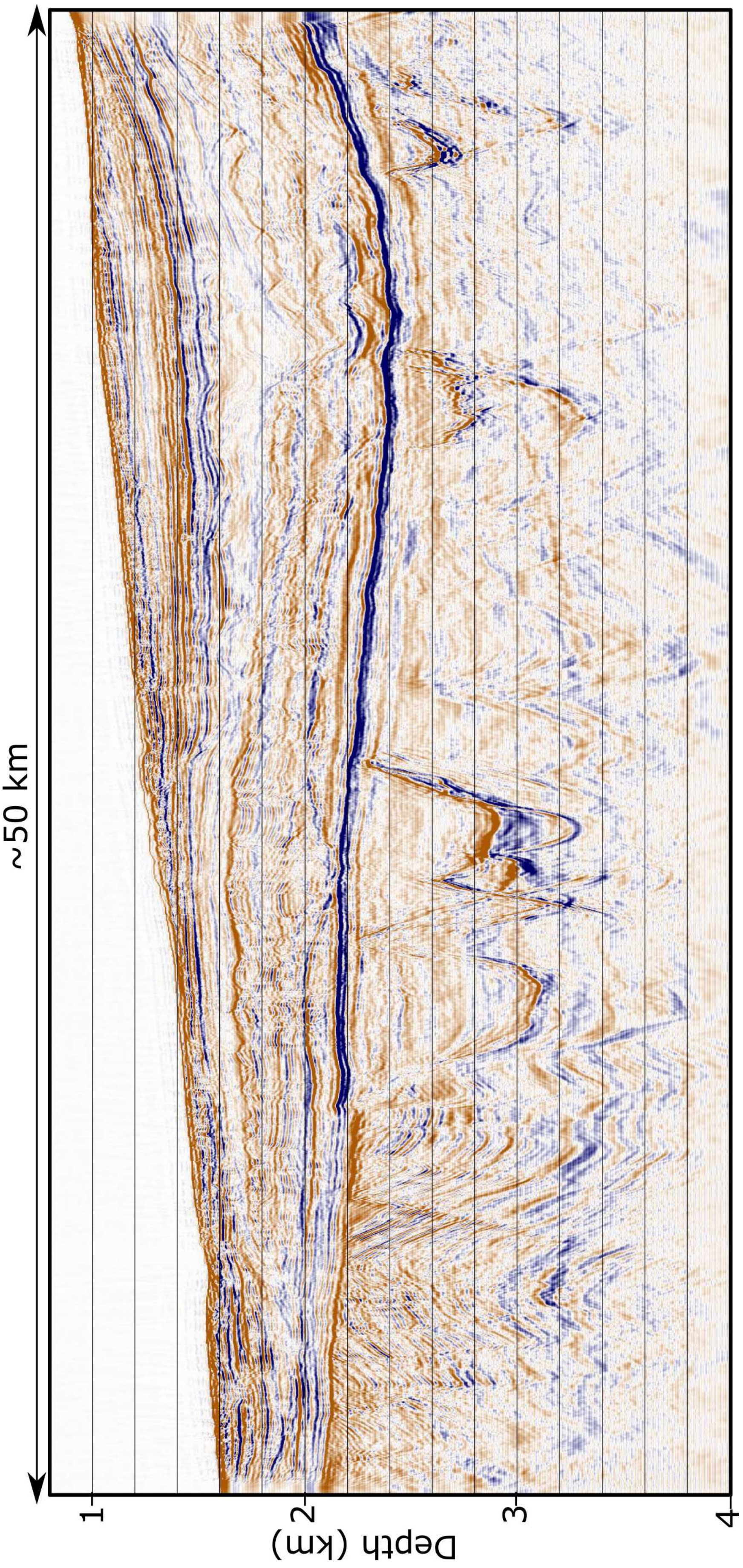
(a) 40 Hz FWI

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease (cont.).



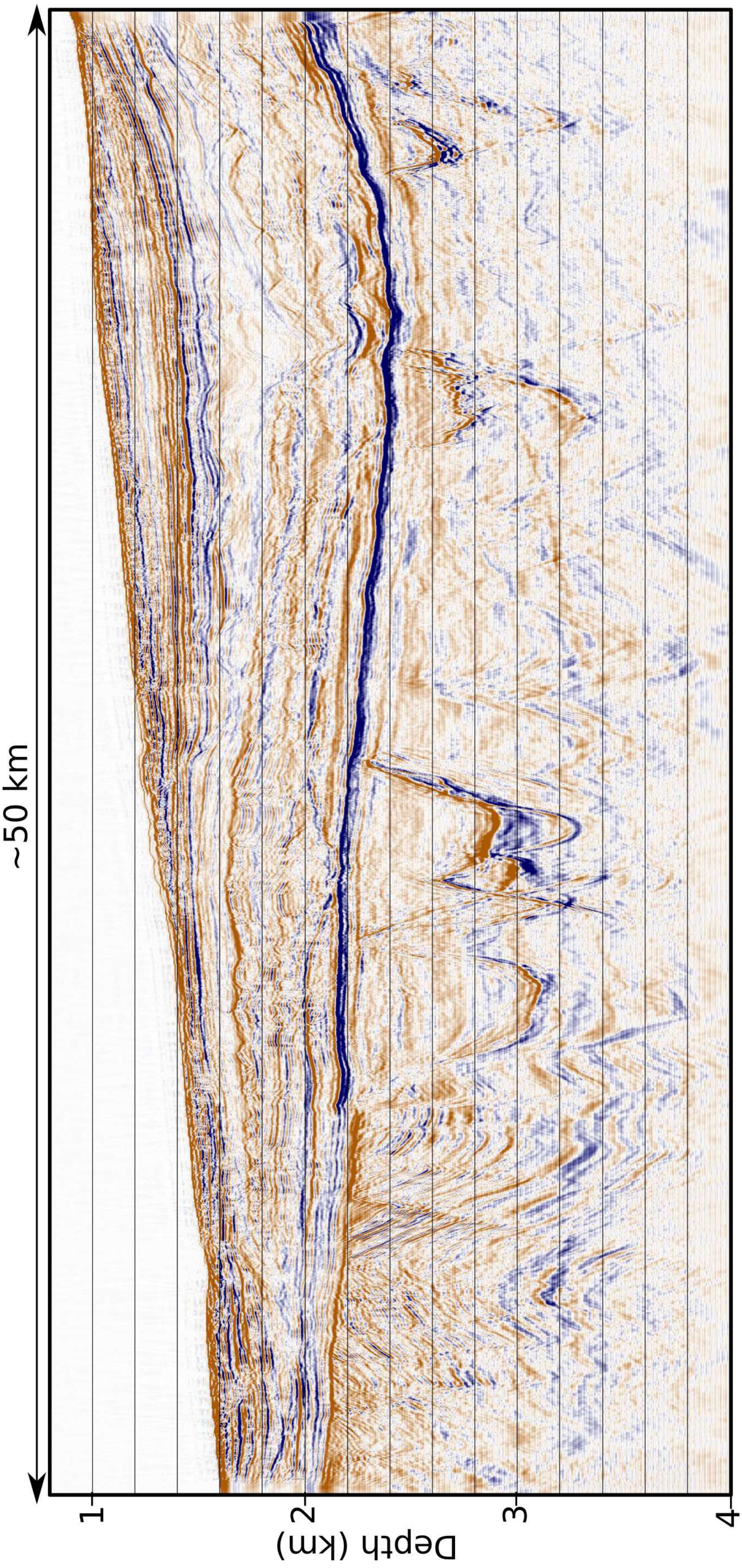
(b) 50 Hz FWI

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease (cont.).



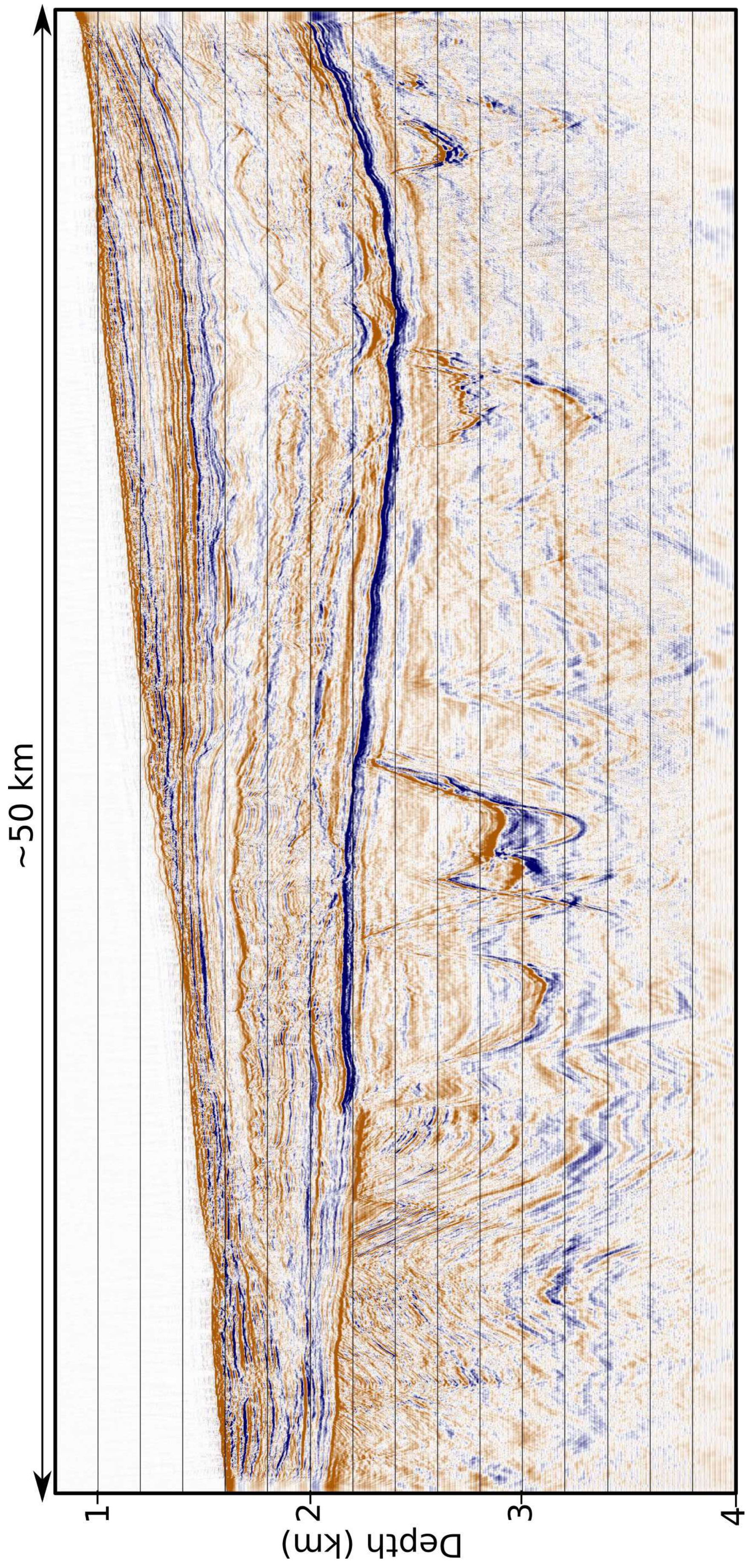
(c) 60 Hz FWI

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease (cont.).



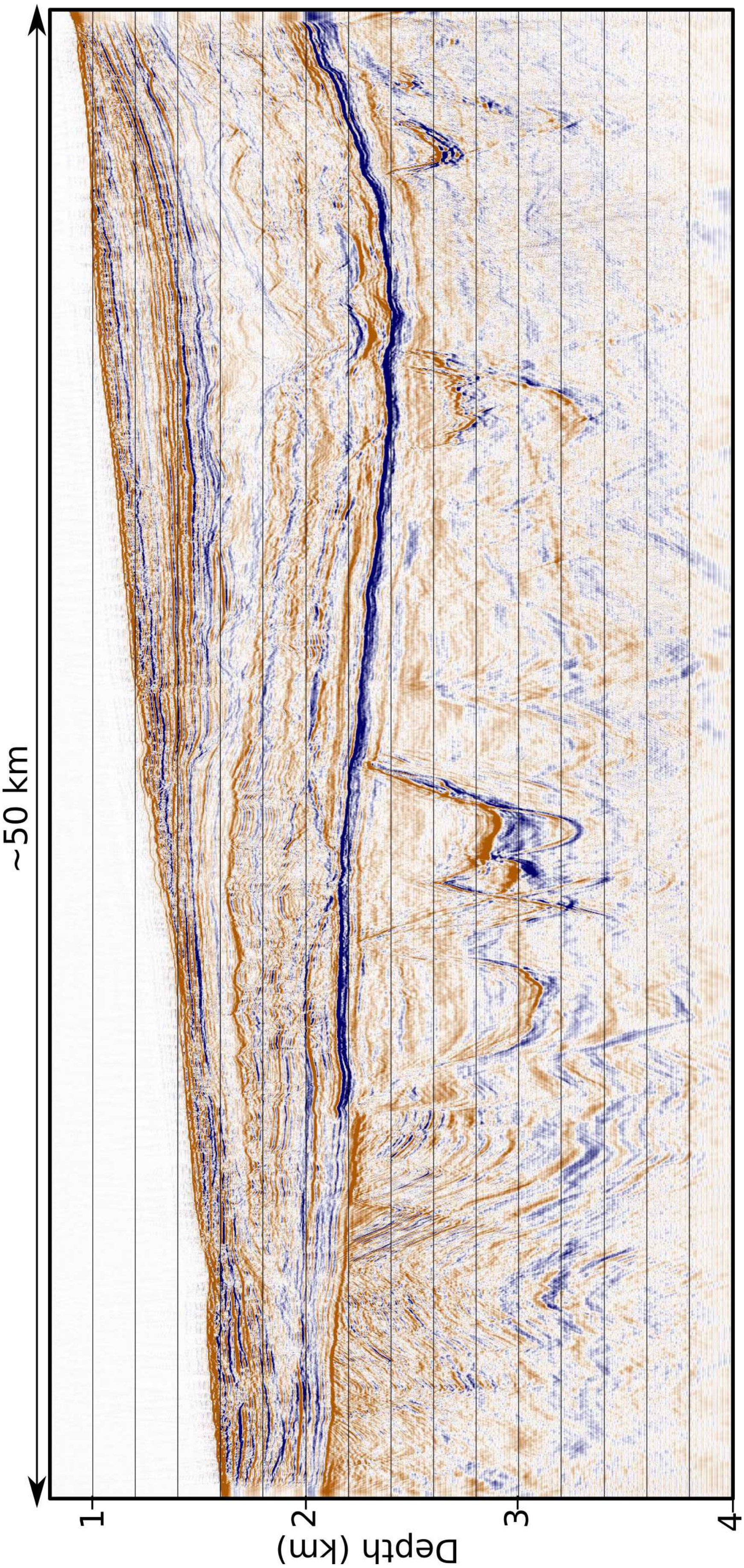
(d) 80 Hz FWI

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease (cont.).



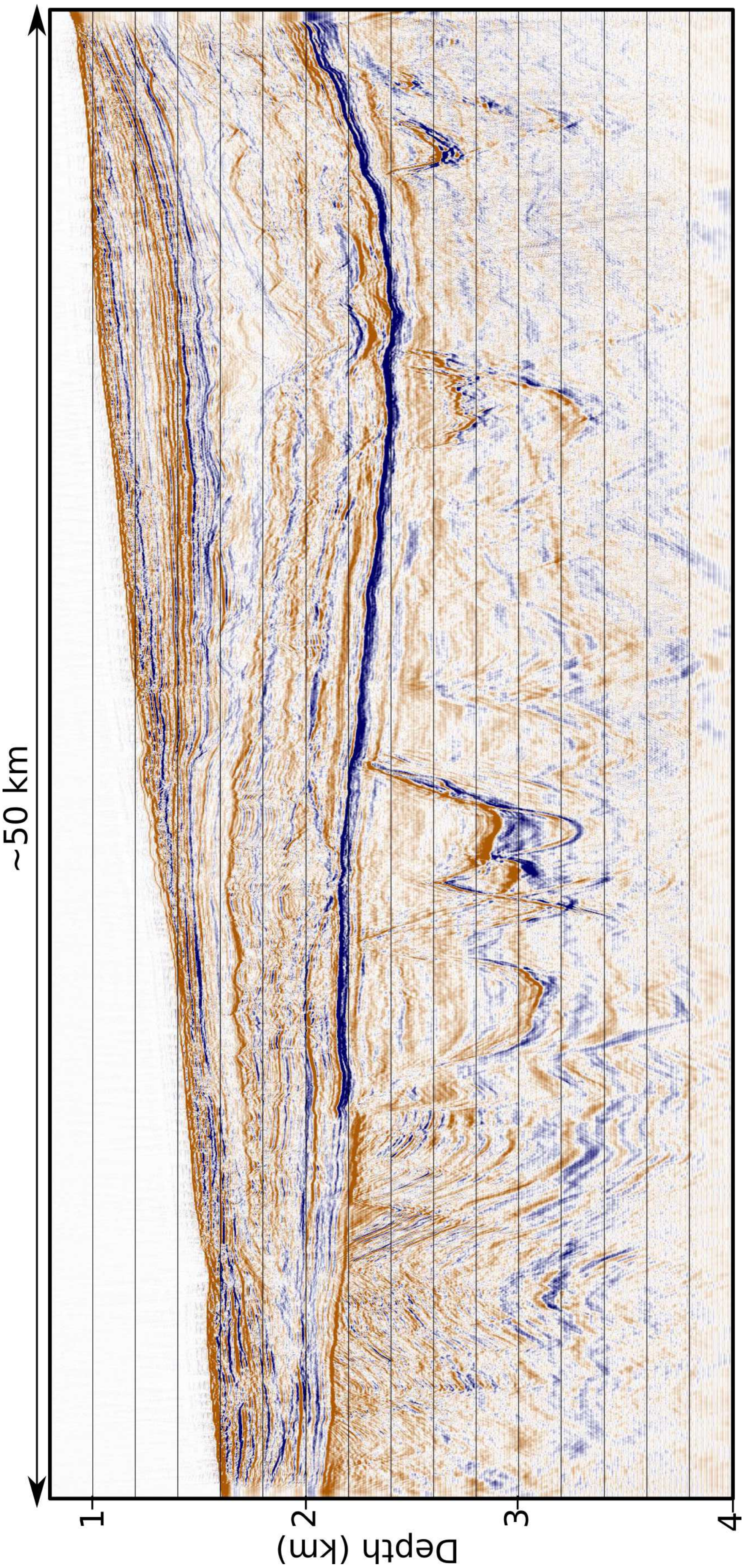
(e) 100 Hz FWI

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease (cont.).



(f) 100 Hz FWI using the spectral shaped field data

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease (cont.).



(g) 100 Hz FWI with included Q in the inversion

Figure 5.7: FWI Impedance contrast models at (a) 40, (b) 50, (c) 60, (d) 80 and (e) 100 Hz; (f) at 100 Hz using the spectral shaped field data and (g) at 100 Hz with Q included in the inversion. Brown denotes an increase in impedance contrast and blue denotes a decrease.

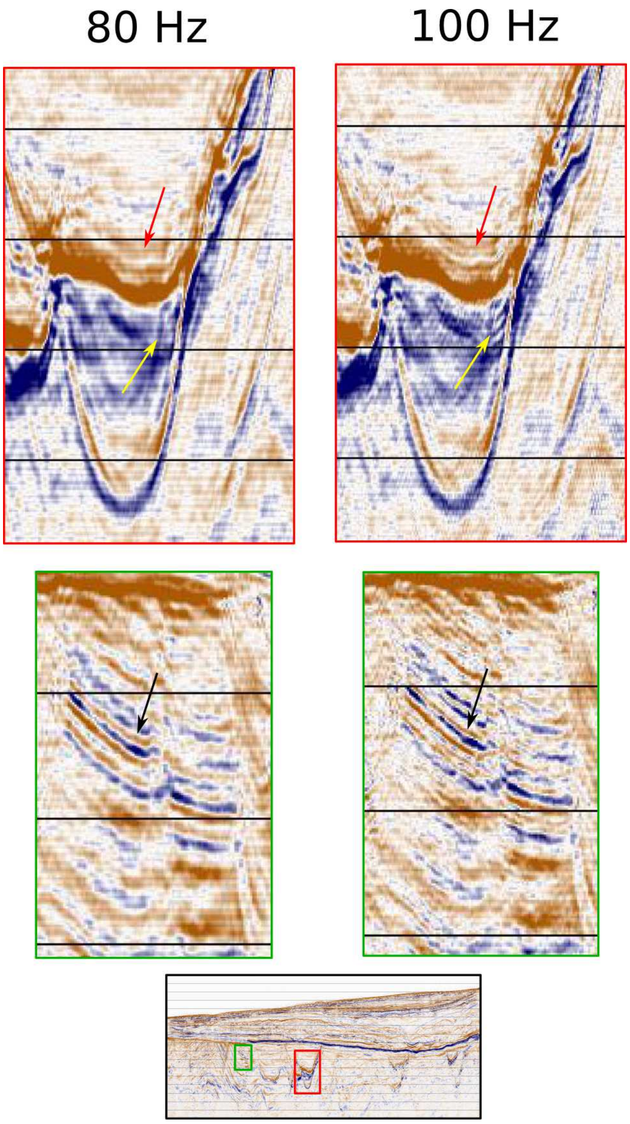


Figure 5.8: Close up of Figure 5.7. Later extent is 3.5 km and 2 km for the red and green boxes respectively.

5.5.2 Computational cost

FWI can be computationally expensive. The cost is proportional to the number of sources, number of iterations and the size of the model in grid cells.

For this study, FWI was performed in 2D and two different hardware were used at different stages of the inversion. The way it was done in this thesis, the 40-Hz FWI model was obtained by using 4 iterations per frequency band, using the low-kernel code and bigger physical size of the model (Phase 0, see in Table 5.2.). The 100-Hz FWI model was obtained using a different machine, performing 2 iteration per each frequency band, using the high-kernel version of the code, smaller physical size of the model, but larger model size in grid cells due to re-sampling of the model. It is not very sensible to compare these run times, but for this experiment 100-Hz model was 4.5 times more expensive than 40-Hz model, required 14 h and 63.5 h respectively.

However, it was possible to optimise the computational effort by starting directly with Phase 1 (see in Table 5.2.) and perform Phase 1 and Phase 2 only. Thus, the same version of the code (high kernel) was run on the same hardware to invert the model of the same physical size. Run in in this settings, 40-Hz FWI model took 3.5 h to compute and 100-Hz FWI model took 49.5 h to compute, making the latter 14 times more expensive.

In principle, it was also possible to reduce the total cost of the inversion even more by using coarser grid spacing and finer time sampling for the earlier iterations. Suggested way of doing that is presented in Table 5.3. This strategy should slightly reduce the cost of 100-Hz 2D FWI and significantly reduce the cost of 100-Hz FWI model if performed in 3D.

Nominal frequency	Max frequency	Grid spacing	Time step	Total iterations	Max offset	2D run time	3D run time
10 Hz	12 Hz	50 m	6 ms				
20 Hz	24 Hz	25 m	3 ms				
40 Hz	48 Hz	12.5 m	1.4 ms	40	10,142 m	1	1
100 Hz	110 Hz	5 m	0.6 ms	72	2,500 m	11	28

Table 5.3: Optimised inversion strategy for high frequency FWI.

5.5.3 100 Hz FWI vs. conventional PSDM

Here, I would like to compare the 100-Hz FWI model to two conventional Kirchhoff PSDM sections; both were provided by different processing contractors and each of them was a result of few month of complicated work.

The first section was obtained using first-pass tomography model and the data with simple de-ghosting applied.

The second section was obtained using a contractor's FWI model and the data with sophisticated de-ghosting applied. It was provided as pictures in two colour schemes - black/white and blue/red.

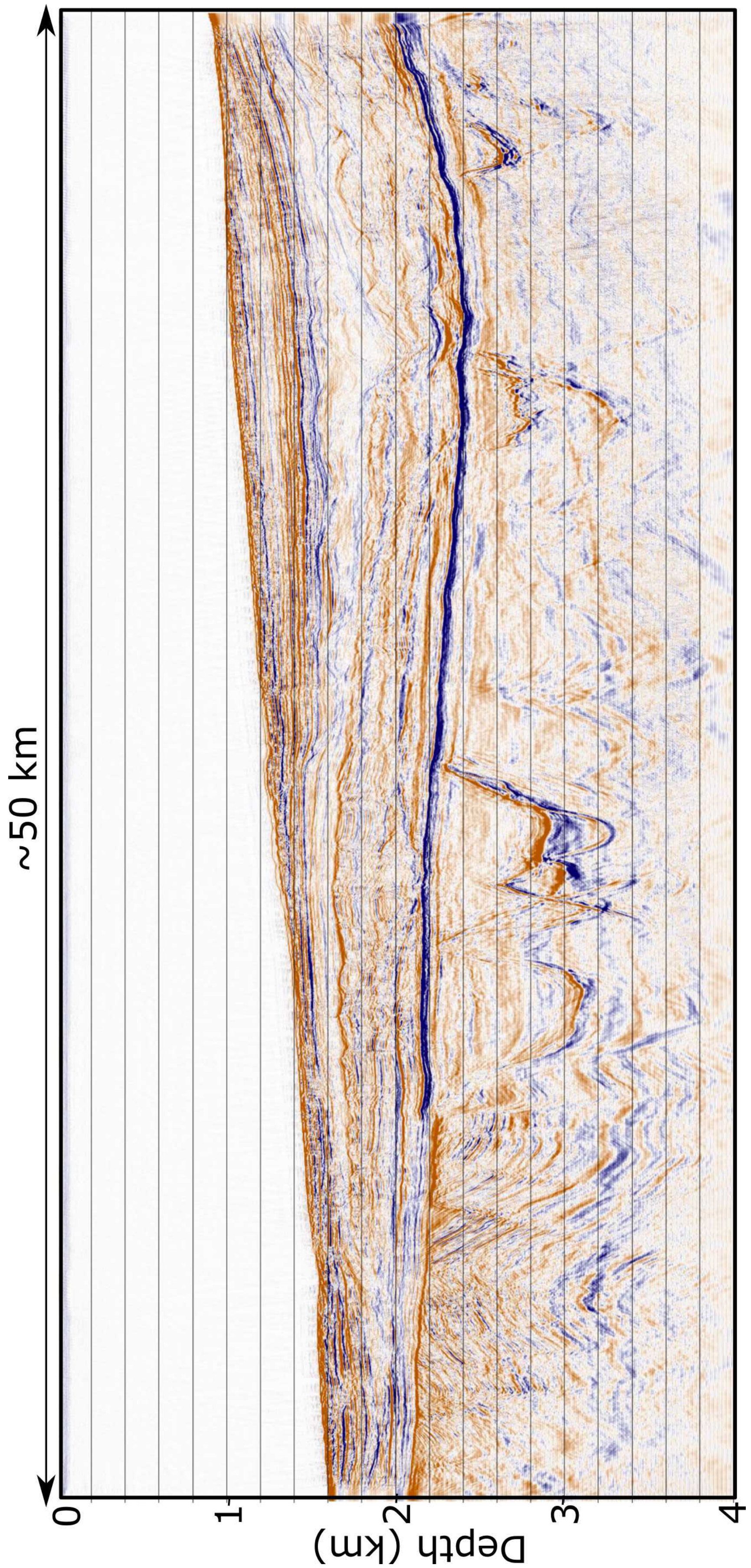
Both PSDM sections have different velocity models, different de-ghosting applied to the data, may have different anisotropy models, different velocities in water and different wavelets. Therefore, the same events are not positioned at exactly the same place on these two models.

Figure 5.9 shows the 100-Hz FWI model on raw field data, displayed as vertical trace-by-trace derivative of 100-Hz FWI velocity model and two Kirchhoff PSDM sections provided by contractors, one of the section is presented in two colour schemes. It can be seen that the PSDM section with sophisticated de-ghosting applied (Fig. 5.9(c) and (d)) is better resolved that the section from another processing contractor (Fig. 5.9(b)). However, the 100-Hz FWI model is at least as resolved as the best contractor's PSDM section and in places, even better.

Figure 5.10 demonstrates close-ups of these four images. The depth position of major events are not exactly the same for the reasons mentioned in the beginning of this section, however, it is easy to identify the corresponding reflectors. It can be seen, that the 100-Hz FWI is much better resolved than the Kirchhoff PSDM obtained using tomography model and better resolved than the Kirchhoff PSDM obtained using an FWI model.

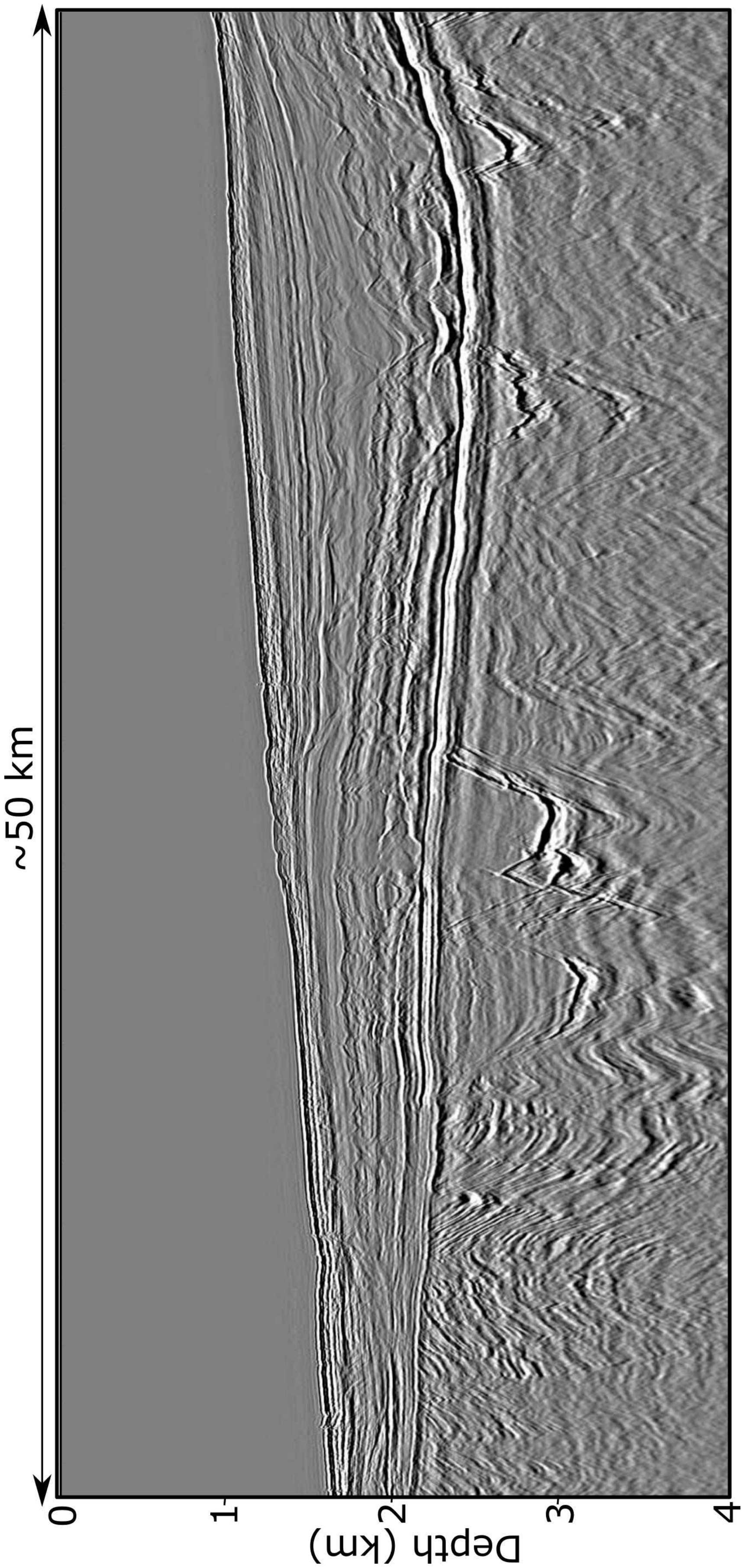
Obtained results show that high-frequency FWI is capable of generating highly resolved depth images directly from raw data. This might suggest a new interpretation suite, consisting not only of conventional processed sections and low-frequency velocity models, but well resolved high-frequency FWI models along with high-frequency velocity models. The close-up of an area at approximately 3 km depth is shown in Figure 5.11. Each of these three images has different physical meaning, such as Kirchhoff PSDM shows acoustic impedance, 100-Hz FWI shows acoustic impedance contrast and 100-Hz FWI velocity model shows compressional velocities of the medium.

Perhaps, for the detailed interpretation it is useful to have all the three volumes rather than just the first one. Especially, taking into account that the highly resolved FWI volumes can be obtained directly from the raw data many months earlier than conventional processing can deliver the results.



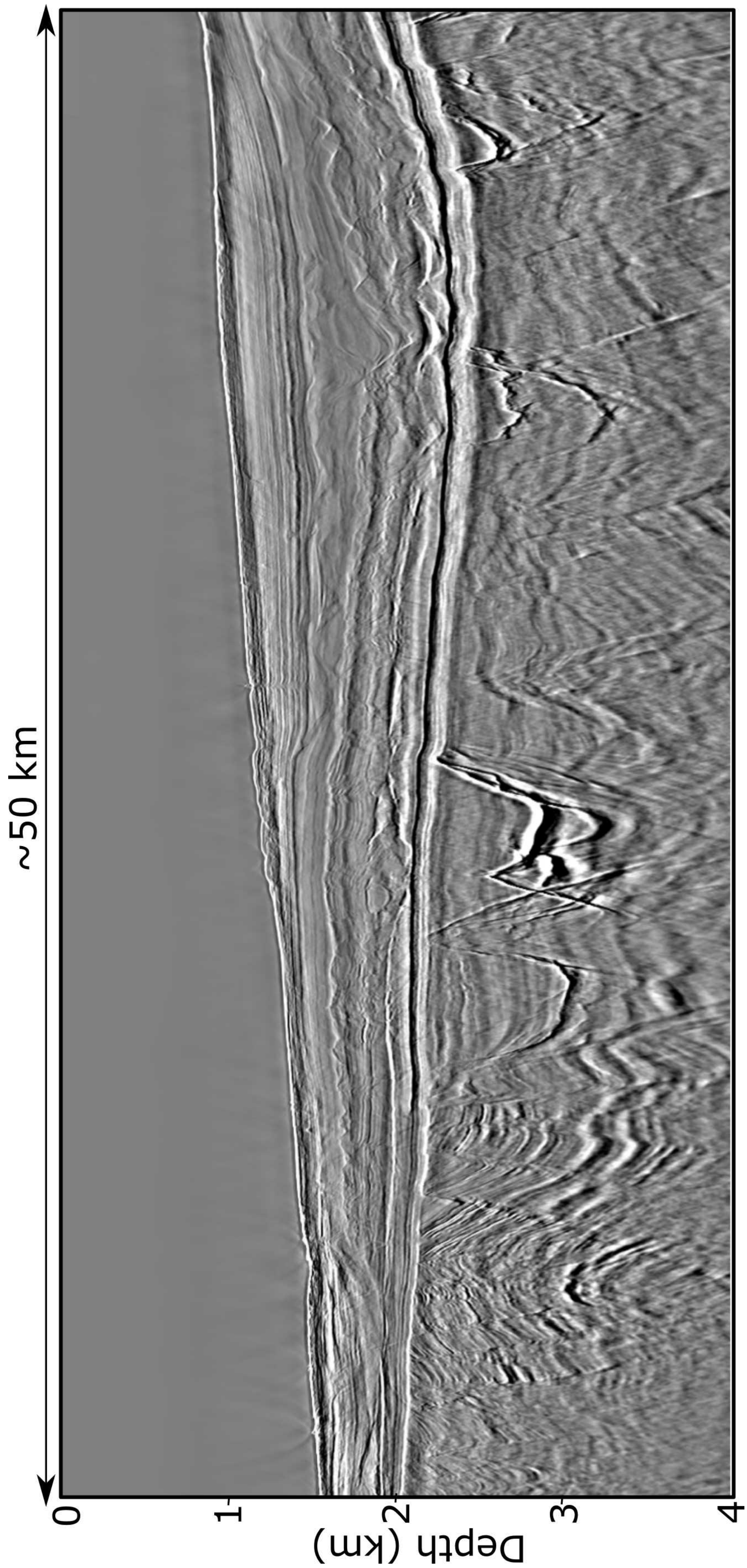
(a) 100-Hz FWI on raw field data

Figure 5.9: Comparison of (a) 100 Hz FWI impedance contrast model compared to two conventional PSDM sections produced by two different contractors. Brown denotes an increase in acoustic impedance contrast and blue denotes a decrease. (b) The Kirchhoff PSDM section obtained using tomography velocities; (c) The Kirchhoff PSDM section obtained using the data with sophisticated de-ghosting and an FWI model; (d) the PSDM section (c) displayed in different colour scale (cont.).



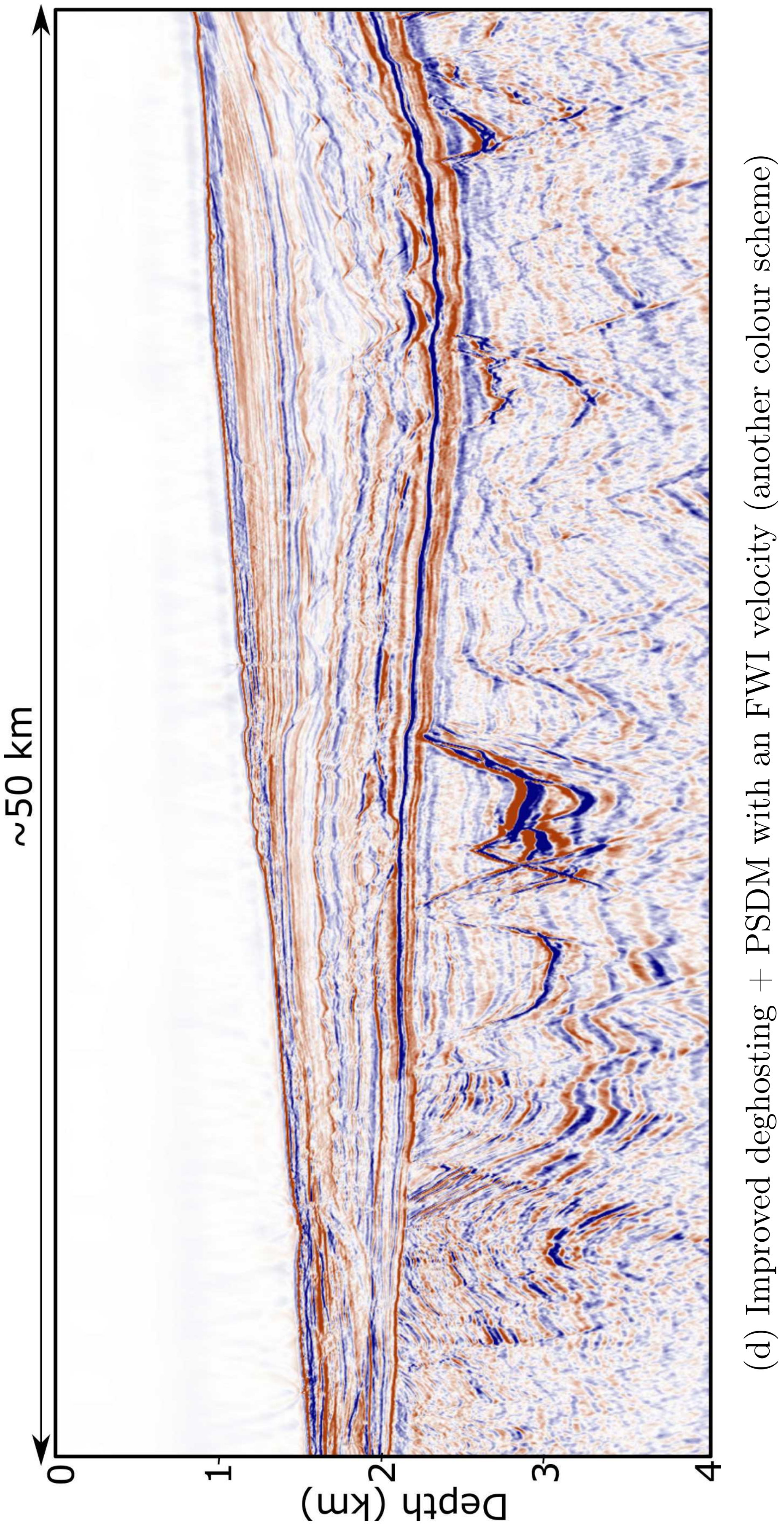
(b) Kirchhoff PSDM after processing + tomo Vp

Figure 5.9: Comparison of (a) 100 Hz FWI impedance contrast model compared to two conventional PSDM sections produced by two different contractors. (b) The Kirchhoff PSDM section obtained using tomography velocities; (c) The Kirchhoff PSDM section obtained using the data with sophisticated de-ghosting and an FWI model; (d) the PSDM section (c) displayed in different colour scale (cont.).



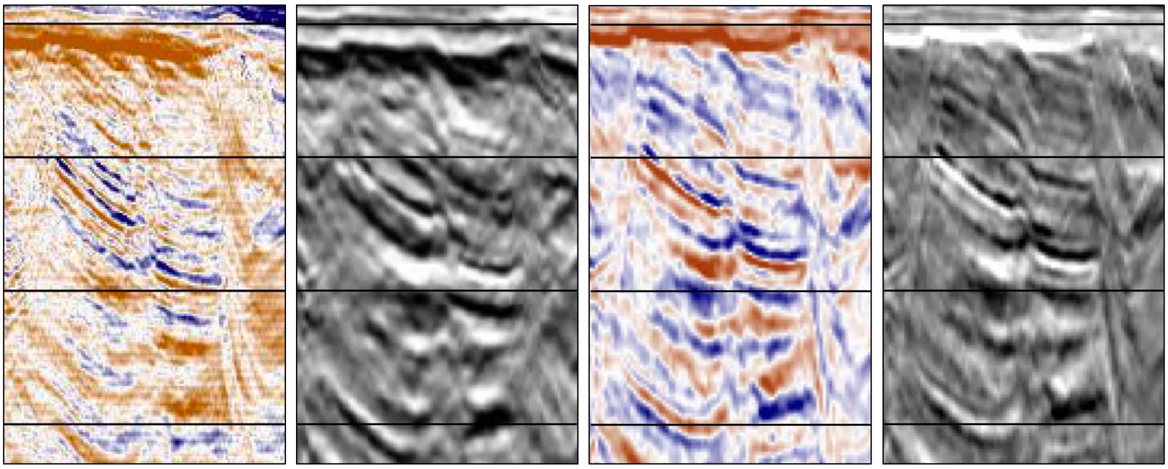
(c) Improved deghosting + PSDM with an FWI velocity

Figure 5.9: Comparison of (a) 100 Hz FWI impedance contrast model compared to two conventional PSDM sections produced by two different contractors. (b) The Kirchhoff PSDM section obtained using tomography velocities; (c) The Kirchhoff PSDM section obtained using the data with sophisticated de-ghosting and an FWI model; (d) the PSDM section (c) displayed in different colour scale (cont.).



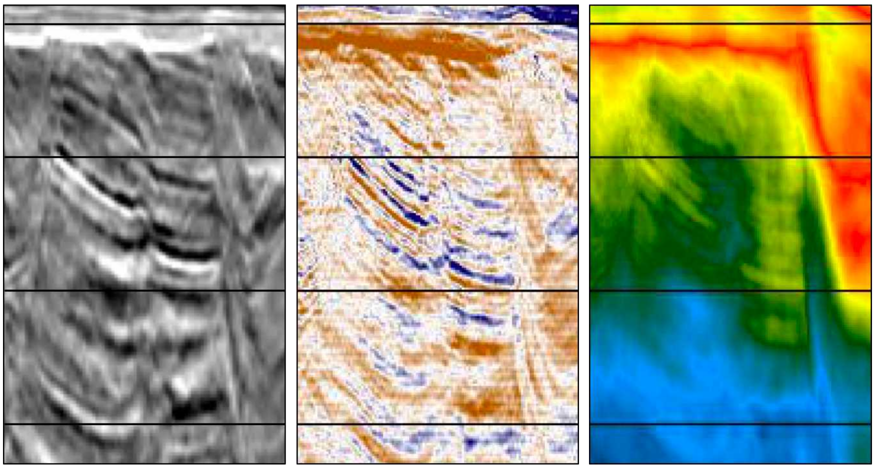
(d) Improved deghosting + PSDM with an FWI velocity (another colour scheme)

Figure 5.9: Comparison of (a) 100 Hz FWI impedance contrast model compared to two conventional PSDM sections produced by two different contractors. (b) The Kirchhoff PSDM section obtained using a tomography model; (c) The Kirchhoff PSDM section obtained using the data with sophisticated de-ghosting and an FWI model; (d) the PSDM section (c) displayed in different colour scale.



(a) 100-Hz FWI (b) Tomo Kirchhoff (c) FWI Kirchhoff (d) FWI Kirchhoff

Figure 5.10: Close-up of Figure 5.9. Comparison of (a) 100-Hz FWI, (b) Kirchhoff PSDM using a tomography model, (c) Kirchhoff PSDM using an FWI model and (d) Kirchhoff PSDM using an FWI model, different colour scheme. Wavelet, de-ghosting, velocity model, anisotropy and water column all potentially different. The position of the close-up is indicated in Figure 5.8.



(a) FWI Kirchhoff, (b) 100-Hz FWI (c) 100-Hz FWI velocity model

Figure 5.11: Close-up showing an interpretation suite: (a) Kirchhoff PSDM using an FWI model, (b) 100-Hz FWI and (c) 100-Hz FWI velocity model. The position of the close-up is indicated in Figure 5.8.

5.6 Potential ways to improve results

In the previous section, it was demonstrated that the updates in FWI model clearly bring value when using frequencies as high as 100 Hz. This result was achieved simply by retaining higher frequencies available in the data and inverting them using exactly the same strategy as for the standard lower frequencies.

However, there might be some factors that could affect FWI at such high frequencies while not having an impact at the lower frequencies. These factors include:

- **Inaccuracy in source estimation at high frequency**. It is well known that estimation of an accurate source is crucial for the successful inversion. An inaccurate source signature at the lowest frequency used in the inversion can lead to cycle skipping and/or fail the inversion completely. At the standard for FWI low frequencies, the source signature estimated is always a far field signature. However, at 100 Hz, the wavelength in the water column is about 15 m while at 25 Hz it is 60 m. For the Pivot dataset, the gun array has dimensions of 14 x 30 m which is very much comparable to the smallest wavelength size. This might mean that using direct arrivals to estimate the source wavelet might give an inaccurate estimation at 100 Hz due to the interference of the higher frequencies.

- **The higher frequencies lacking in the source**. $fullwave3D$ internally applies a bandpass filter both to the field data and the source wavefield with different central frequency at every iteration. This filter is in addition to that applied to the field data and the source signature during the pre-processing. It rolls off rapidly above the central frequency as well as reduces amplitudes below the central frequency (Warner et al. 2013). Consequently, the source wavefield must be accurate at every frequency used in the inversion and, most importantly, contain high frequencies. In case the high frequencies are not present in the source signature, after that internal filter has been applied, there is no appropriate source energy to propagate and generate the modelled data matching the field data.

- **Attenuation**. One of the assumptions commonly made on physics during FWI is neglecting attenuation. There are 4 reason for energy loss during propagation: (1) geometrical spreading, (2) transmission losses due to large-scale scattering, (3) anaelastic losses and (4) sub-resolution elastic scattering. Here, attenuation due to geometrical spreading and transmission loses due to large-scale scattering were taken was taken into account during the inversion. However, attenuation due to anelastic losses and sub-resolution elastic scattering, that depends on the geology of the study area, was not dealt with properly.

Although the earth is an attenuating medium, conventionally used lower frequencies tend to attenuate with a lower rate than the higher frequencies. It is possible, that for the Pivot dataset taking attenuation into account could improve the 100 Hz FWI model.

- **Gun array is not at constant depth**. As has been mentioned before, the gun array dimensions are comparable to the lowest wavelength at the higher frequency used in the study. During the inversion, the gun depth was assumed to be constant everywhere. If in reality this assumption was violated, then the offsets would change by an amount that would not create a kinematic and dynamic difference at the low frequencies but would generate one at the high frequencies.

Some of these factors have been explored in order to investigate whether they have any impact on the high frequency FWI result and whether or not they can be neglected during the inversion.

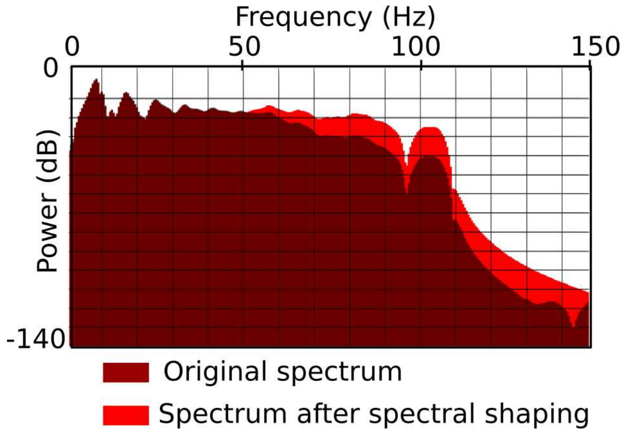
5.6.1 Spectral shaping

Figure 5.12 shows the spectra of the source wavefield and the field data after the initial preprocessing. It can be seen, that the two are remarkably similar and both have a decline in the amplitudes after 60 Hz. The similarity here is to be expected because the source signature was derived from the data.

Theoretically, boosting higher frequencies might improve the FWI model at those frequencies. There are few ways to achieve that, for instance, boosting the high frequencies in the model itself or in the data. For this study, the latter was performed using spectral shaping.

Spectral shaping of the field data prior to the inversion has been successfully used by Lazaratos et al. (2011) and Ruth et al. (2011). It is a way of applying dynamic processing across the frequency spectrum that can help to achieve the desired balance. In order to increase the relative amplitudes of the frequencies above 60 Hz, spectral shaping was applied focusing on just certain areas in the frequency spectrum.

Figure 5.12 compares the spectra before and after the spectral shaping for both the source wavefield and the field data. The same filter was applied to the source signature and the field data in order to maintain consistency between the two. The final spectra is flatter across the whole frequency bandwidth, therefore the relative amplitude of the high-frequency half of both spectra have increased.



(a) (b)

Figure 5.12: Amplitude spectra before and after spectral shaping of the (a) source wavefield and (b) field data.

After the data have been additionally pre-processed, acoustic anisotropic FWI has been performed using the same starting model and exactly the same strategy. The FWI impedance contrast models obtained using the field data with and without spectral shaping are compared in Figure 5.7(e) and Figure 5.7(f). They are very similar, but not exactly the same. Careful examination indicates that boosting the higher frequencies in the data leads to improved resolution in some parts of the FWI impedance contrast model (Fig. 5.13).

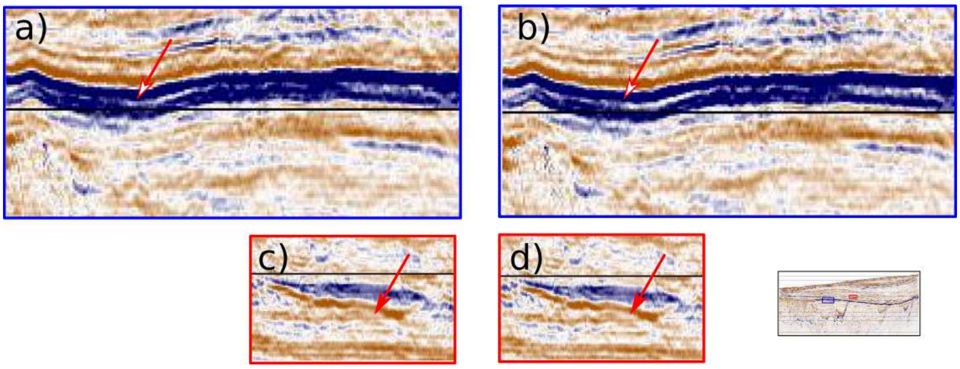


Figure 5.13: Close up of Figure 5.7(e) and (f). FWI impedance contrast models at 100 Hz (a), (c) without and (b), (d) with spectral shaping applied to the field data during the pre-processing. Later extent is 3 km and 1.5 km for the blue and red boxes respectively.

5.6.2 Attenuation

In seismic, attenuation is defined as energy loss of the wave as it propagates through the earth. The energy dissipates due to:

- Geometrical spreading, that is caused by distribution of the energy to greater surface as the wavefront of the wave expands during the propagation. Geometrical spreading is automatically taken into account during the inversion on the forward modelling stage.
- Transmission losses due to scattering from structure at wavelength and longer scale lengths - FWI takes those transmission loses correctly into account.
- Transmission losses due to scattering from sub-resolution-scale structure - that is, from things too small for FWI to recover. FWI does not correct for these, and they look very like viscous attenuation although they are not the same physics.
- Viscous effects, that means that part of energy transforms into heat. The portion of this loss is defined solely by the properties of the medium.

In practice, in seismic processing and inversion, viscous effects and fine-scale scattering are always lumped together and described with a single Q value.

Conventionally, like elasticity, viscous effects and losses due to small-scale scattering are neglected during acoustic time-domain FWI in order to reduce computational effort. However, these effects cause amplitude attenuation and phase dispersion (Causse et al. 2018). The significance of these effects on the wavefield depends upon the geology of the area. Agudo et al. (2017) and Agudo et al. (2018) explore the benefits of including viscous effects into acoustic time domain FWI.

In theory, the higher frequencies attenuate faster than the lower frequencies. Consequently, it was important to understand whether including an attenuation model - described by a quality factor Q - would improve resolution of the high frequency FWI model.

I would like to thank my colleague Oscar Agudo, who has contributed greatly to this study by producing the attenuation model Q as a part of his own research (Agudo 2018).

Here, I will give a brief overview of the technique used for Q model estimation, which is described in detail by Agudo (2018).

Estimate of Q was computed from the near-offset data using Gabor transform (Wang 2004; 2014). Gabor transform is a windowed Fourier transform with the time window given by a Gaussian function. Gabor transform applied to a seismic trace in time domain computes its frequency content as a function of time. This method robustly and automatically, trace-by-trace, estimates Q from reflection data corrected for spherical divergence and is suggested to perform better than alternative methods (Lupinacci et al. 2015). For the Pivot dataset, that contained both reflections and refractions, only the first-offset trace from every shot was used for Q estimation. The data were corrected for spherical divergence using the method described by Newman (1973). Then, the Gabor transform of every trace was calculated. Similar to velocity analysis, for every trace it produces a 2D spectrum with values dependent upon time and frequency. Then the 2D spectrum is collapsed into a 1D spectrum by stacking the equal values $x = \omega t$, where ω is the frequency and t is time. This 1D spectrum is normalised in order to obtain the function everywhere greater than zero. Finally, linear approximation of the natural logarithm of this function gives the reference Q value for the whole trace. In order to produce a Q model that varies in depth and better approximates the geology, computation was carried out in different intervals using the method of Wang (2009).

The final estimated Q model is displayed in Figure 5.14. It shows large values everywhere in the water column, which is due to the absence of the viscous effects in water. The Q values are increasing with depth, indicating the decrease of attenuation with depth. This generally appears to be true for the sedimentary basins as viscous effects are expected to be greater at shallow regions.

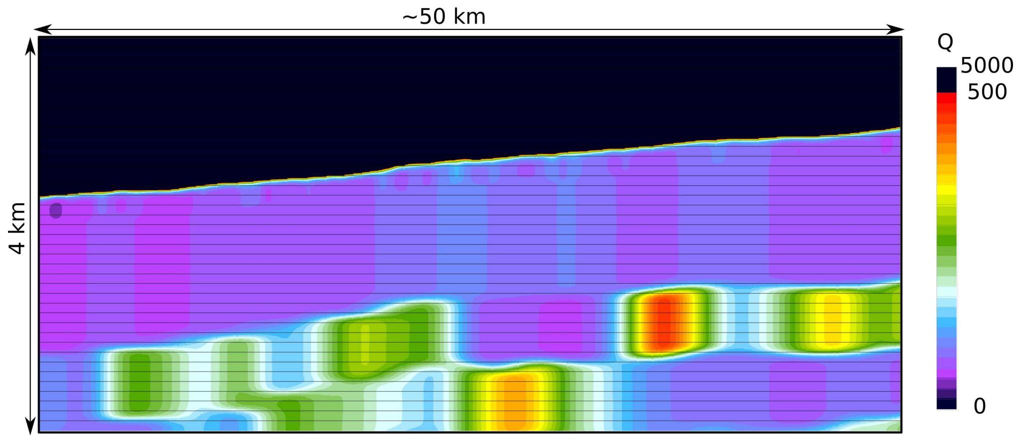


Figure 5.14: Q model estimated from the field data.

After the model has been obtained, the dataset was inverted again using 2D acoustic anisotropic FWI code in time domain using the starting frequency of 3 Hz, the maximum frequency of 100 Hz and the Q model being fixed at every iteration.

The FWI impedance contrasts at 100 Hz obtained without and with Q are compared in Figure 5.7(e) and 5.7(g). After careful examination, it was concluded that in this particular case viscous effects do not seem to have a significant impact on high frequency FWI.

5.7 Conclusions

2D anisotropic full-waveform inversion was successfully applied to a narrow-azimuth marine-streamer dataset with a maximum frequency of 100 Hz. It demonstrated an excellent resolution which can be obtained using higher frequencies than the standard frequencies of 10-12 Hz. It was shown that the useful updates to the model do not stop at frequencies around 40 Hz, which, to date, is the highest frequency that can be found in published materials.

It also was successfully shown that there is a benefit of inverting frequencies as high as 100 Hz as that can bring a significant value for detailed geological interpretation. The FWI model at 100 Hz is 16 times the cost of 50 Hz FWI, that suggests it is a reasonable effort to perform in the areas of high interest, for instance, reservoirs.

In this chapter, the factors that can affect the inversion of the high frequencies were discussed as well as the means of improving the achieved result. It was obvious, that boosting the higher frequencies benefits the inversion whilst including viscous effects in the modelling had little effect in this particular case possibly due to the geological environment.

Most importantly, high-frequency FWI is capable of generating highly resolved depth images directly from raw data very much faster than conventional processing. These images are at least as well resolved as conventional PSDM sections. They are possible to obtain in a matter of a week, potentially directly on a boat, provided raw data, a simple starting velocity model, an appropriate source signature and an access to a cloud to run the inversion. In contrast, conventional processing takes many months of complicated work to produce a high-quality migrated sections. That has long been an aim of FWI, and finally it seems to be materialising.

Chapter 6

Difficulties with NATS

Over the last decade FWI has become a standard tool for velocity model building workflows on a commercial scale. Commercial FWI with cycle-skipping solutions works well at least 80% of the time (Warner 2018). However, for a number of reasons FWI can fail and, sometimes, these reasons might not be obvious. It is important to understand what can go wrong when performing FWI and why.

6.1 Rational

The Gabon dataset was acquired for a small oil company without any direct experience with FWI. They had it processed by a major contractor. The major contractor has significant experience in FWI so they were asked to apply FWI and they have reported that FWI did not work on the dataset. The oil company then sent it to a small company that specialises in FWI. This company could not make FWI work either. Then we obtained the dataset and tried to explore this very unusual situation.

6.2 Reasons FWI fails

Provided that the data have been collected appropriately, there is an adequate starting model and a suitable source signature, FWI commonly works well on marine datasets. However, it also can fail for one of the following reasons:

1. Starting velocity model is not good enough
2. Low frequencies are missing in the collected data
3. The data have been pre-processed in some inappropriate way
4. The data are very noisy
5. The data are very sparse
6. The true model is very complicated, especially if it contains salt or basalts or highly contrasting shallow layers
7. The target is too deep relative to the maximum offset
8. There are strong elastic effects present (when performing acoustic FWI)
9. Starting model is nearly perfect, so there is no room for improvement

When FWI fails, it usually fails in one of the following three ways:

- The FWI model appears to look sensible, but it does not fit the wells and does not improve migration gathers
- FWI goes very obviously wrong and introduces clear artefacts. Usually it happens when data are noisy and/or sparse and/or badly cycle-skipped
- FWI stalls and does not move on from the starting model. This outcome is not very common, however, it is possible

In this study, FWI fails in the most unusual and uncommon way - it stalls.

6.3 Inputs: data and start models

The study was conducted on the Gabon dataset described in chapter 3. This is a 3D narrow-azimuth towed-streamer dataset acquired offshore Gabon with the primary aim of imaging the Upper Cretaceous section of the data, lying between approximately 2.5 and 3.5 s (two-way time). Here, all the inversions were performed on a single 3D swath of the data.

6.3.1 Field data

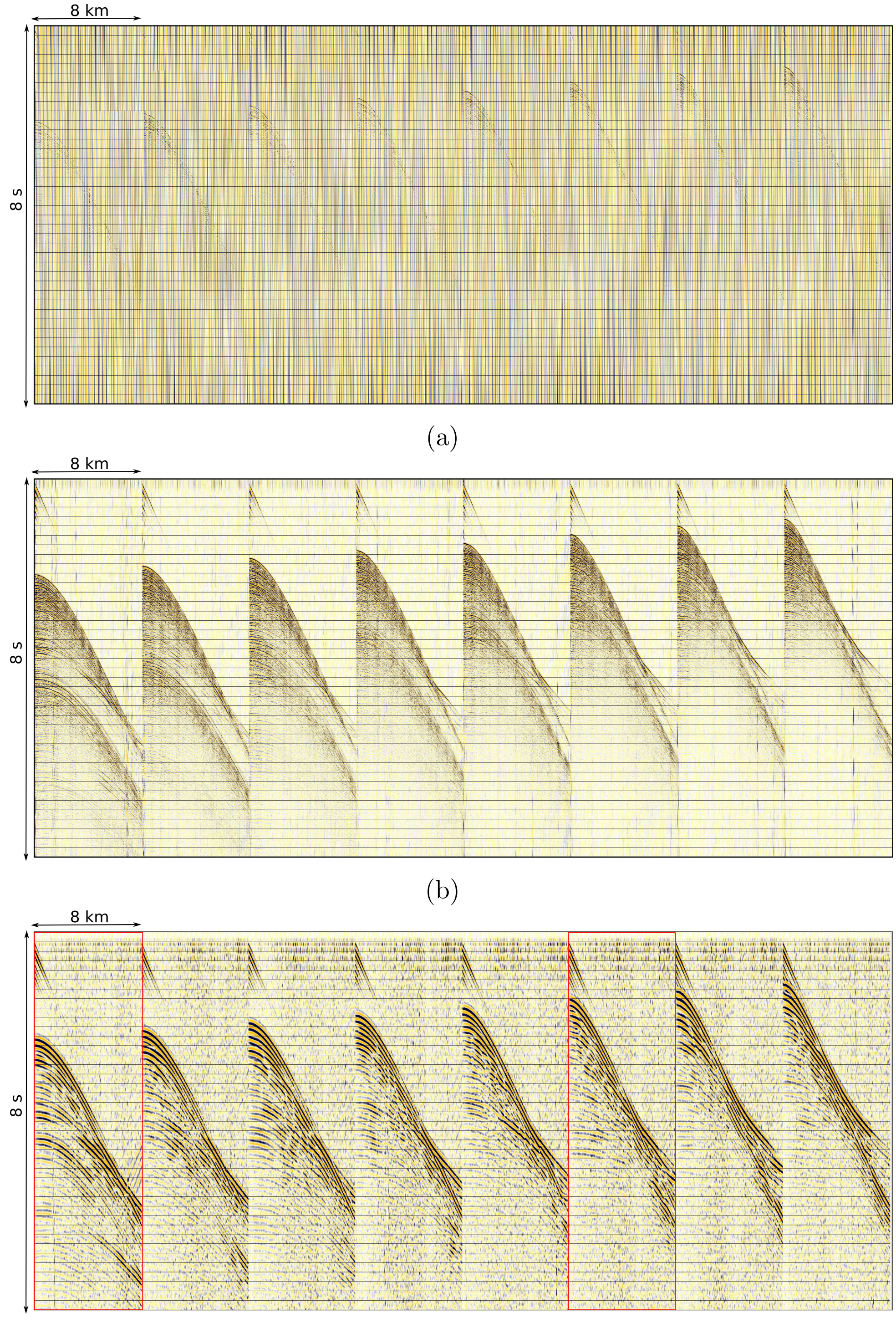
The raw field data are displayed in Figure. 6.1(a). These are eight different shots gathers recorded with the same cable that was positioned in the middle of the 10 streamer array. The gathers are evenly distributed across the model, with the left-hand side gather located in deep water and the right-hand side gather located in relatively shallow water with water depth corresponding to two-way travel times of the seabed arrivals from 2,000 to 850 ms. The raw data are clearly dominated by ultra-low frequency noise associated with the background ocean swell.

The same shot gathers, high-passed using a minimum phase Ormsby filter rolling off from 1.5 to 3 Hz are displayed in Figure 6.1(b) with no scaling applied. It can be seen, that the data are reflection dominated and contain strong first-order multiples in the deep-water area, which are not obviously present in the shallow part of the survey. The data have some refraction energy in the shallow-water area and minimal refraction energy in the deep water area. The data also contain relatively small amount of strong randomly distributed swell-noise, which is more visible at low-frequencies and will be examined later.

Figure 6.1(c) displays the same eight shot records, low-passed with an Ormsby filter below 12 Hz. This data were used for most of the inversion tests presented here. It can be seen, that the first-order multiple is not equally strong throughout the survey and the data are relatively noisy, although the signal-to-noise ratio appears to be adequate for FWI away from the obvious large noise bursts; certainly datasets with this level of apparent noise normally behave sensibly during FWI.

The first and the sixth shot gathers are enlarged in Figure 6.2. Both are displayed with no scaling applied. The shot gather shown in Figure 6.2(a) contains strong swell noise on a few traces at both short and long offsets and some amount of coherent noise, dipping down to the first arrivals, at the very long offsets. The shot gather in Figure 6.2(b) appears to be a relatively noise-free record and shows no evidence of strong swell noise and coherent noise at the long offsets.

Overall, the data contain some amount of incoherent swell noise, distributed randomly across the survey that affects no more than about 20% of the recorded traces. In addition, there is also coherent noise, visible at very long offsets, this noise appears to be visible at relatively low frequencies and it is not present everywhere in the survey and might be associated with the energy recorded from a nearby shooting vessel.

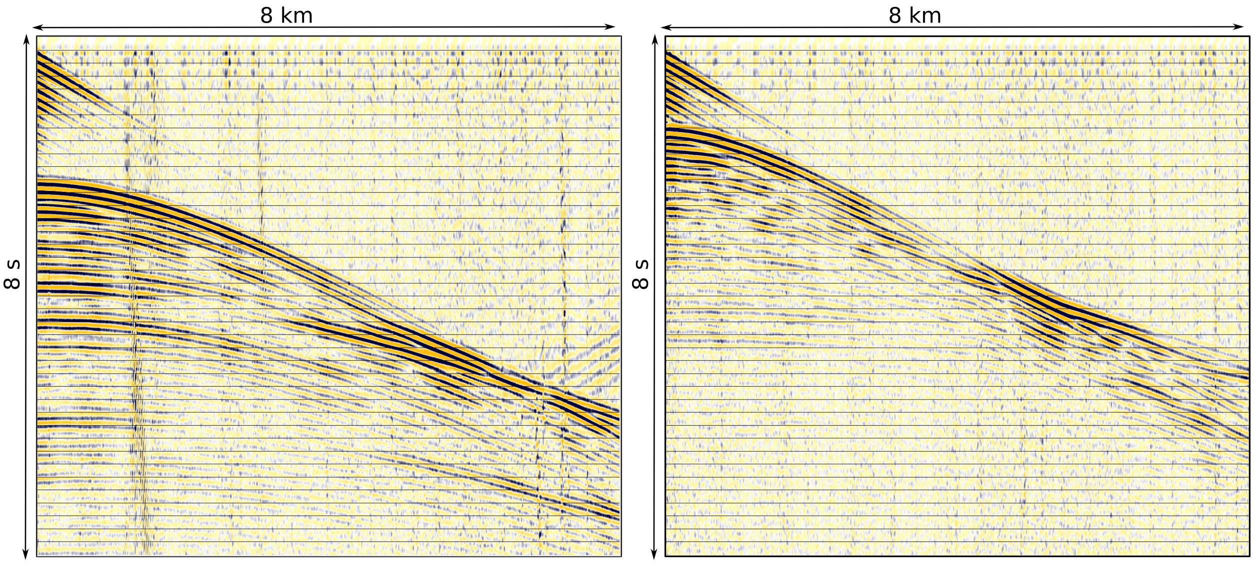


(a)

(b)

(c)

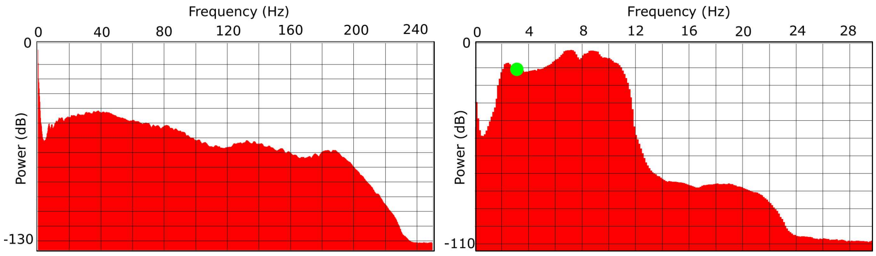
Figure 6.1: Shot records: (a) raw; (b) filtered to retain data above 3 Hz, no scaling applied; (c) Filtered to retain data between 3 and 12 Hz, traces are scaled individually.



(a) (b)

Figure 6.2: Close up of Figure 6.1(c): shot gathers low-pass filtered at 12 Hz. No scaling applied.

The amplitude spectrum of the raw data (Fig. 6.1(a)) is shown in Figure 6.3(a). The high amplitude peak at the very-low frequency, below 2 Hz, is associated with the cable-tugging and background swell noise. The central frequency is about 40 Hz and is followed by a decline of amplitudes to about 200 Hz and then a significant drop in amplitudes for the frequencies above 200 Hz. This drop is due to a high-cut filter applied during the acquisition. Theoretically, the source and receiver notches should be present at about 93 and 54 Hz respectively, but they appear to be slightly shifted to the higher part of the spectrum.



(a) (b)

Figure 6.3: Amplitude spectrum of the data: (a) raw and (b) 3 to 12-Hz band-pass filtered. The green dot marks the starting frequency for FWI.

The amplitude spectrum of the data low-passed below 12 Hz is shown in Figure 6.3(b). The very-low frequencies are absent due to the coherent cable-tugging and ocean swell noise removed using a low-cut Ormsby filter rolling off from 1.5 to 3 Hz. Amplitudes drop from 7 to 3

Hz, but then start to rise again from 3 to 2 Hz. This is very likely to be the effect of the source and receiver ghost that eliminates low frequencies. The rise below 3 Hz is probably ambient noise, which tends to increase at low frequencies. The amplitude spectra of the data suggest, that the frequency of 3 Hz is likely to be the lowest possible frequency to start the inversion.

Overall, the dataset seems to look suitable for conventional FWI because it contains some amount of refracted energy at long offsets, higher than 1:1 signal-to-noise ratio and low frequencies present in the data.

6.3.2 Starting models

An anisotropic migration velocity model was used as the starting model for the inversion. It was provided by the processing contractor. I was not involved in generating this model, but the process is a familiar one.

The velocity model was obtained during migration velocity analysis using a preliminary 1-km velocity field with 2-km smoothing and ray-traced Kirchhoff migration with the migration velocity points generated by migrating the 3D volume, with phase-only Q-compensation applied. Migration velocities were picked on a 1 km grid of inline/crossline locations.

An eta correction for anisotropy was picked simultaneously on the same grid as migration velocities. It was picked in the shallow-mid section of data, using values of 2.5%, 5% and 7.5%.

The original contractor provided both models in the time domain and they were converted into the depth domain using Dix equation (Dix 1955):

$$V_n^{int} = \left(\frac{V_n^2 t_n - V_{n-1}^2 t_{n-1}}{t_n - t_{n-1}} \right)^{0.5} \tag{6.1}$$

where V is the migration velocity, t is time, n is a depth interval, V^{int} is the interval velocity in depth.

The velocity values were then converted into slowness values, linearly interpolated on the 50 m grid in horizontal direction (the choice of the grid spacing and time sampling will be discussed later in this chapter), merged with the bathymetry model defined on a 1 m vertical grid, smoothed with a Gaussian filter using a standard deviation of 150 m (that provides smoothing with a range from 1.2 to 2.7 size of a local wavelength from water to the deepest sediments), then re-sampled in depth to the 50 m grid and, finally, the slowness values were converted back to the velocity values. The model was smoothed in slowness in order to preserve the travel times.

The bathymetry model was provided as depth values, that were initially recorded by echo sounder and converted into depth using a constant velocity of 1500 m/s. The bathymetry profiles are displayed with respect to the geometry of a 3D swath and migration velocity points in Figure 6.4. These values were converted back into two-way travel times, re-calculated to properly adjust for the water column velocities used in the inversion and linearly interpolated on the migration velocity grid.

The provided anisotropy (eta) model, after being converted into depth, was heavily smoothed as it was blocky due to being constructed only from three constant values of 2.5, 5 and 7%. Then, it was also interpolated on the same grid as the velocity model, merged with the bathymetry model and re-sampled in depth. However, $fullwave3D$ requires anisotropy to be set as two separate parameters, delta and epsilon. Considering that the ratio of delta to epsilon δ/ϵ commonly ranges from 0.5 to 0.75 for sedimentary basins (Mike Warner, personal communication, September 15, 2017) and that the area was characterised by low eta values, it was decided to use provided eta values as the epsilon model and populate the delta model with zeros.

The final starting velocity and anisotropy models are shown in Figure 6.5. The velocity model is simple and very smooth, practically one-dimensional with velocities gradually increasing from a constant everywhere in water of 1500 m/s to a maximum of 5000 m/s in the deep part of the section. The anisotropy model is also very simple and one-dimensional with maximum values reaching 2.7%, which suggests low anisotropy in the area. It was kept fixed during every inversion test.

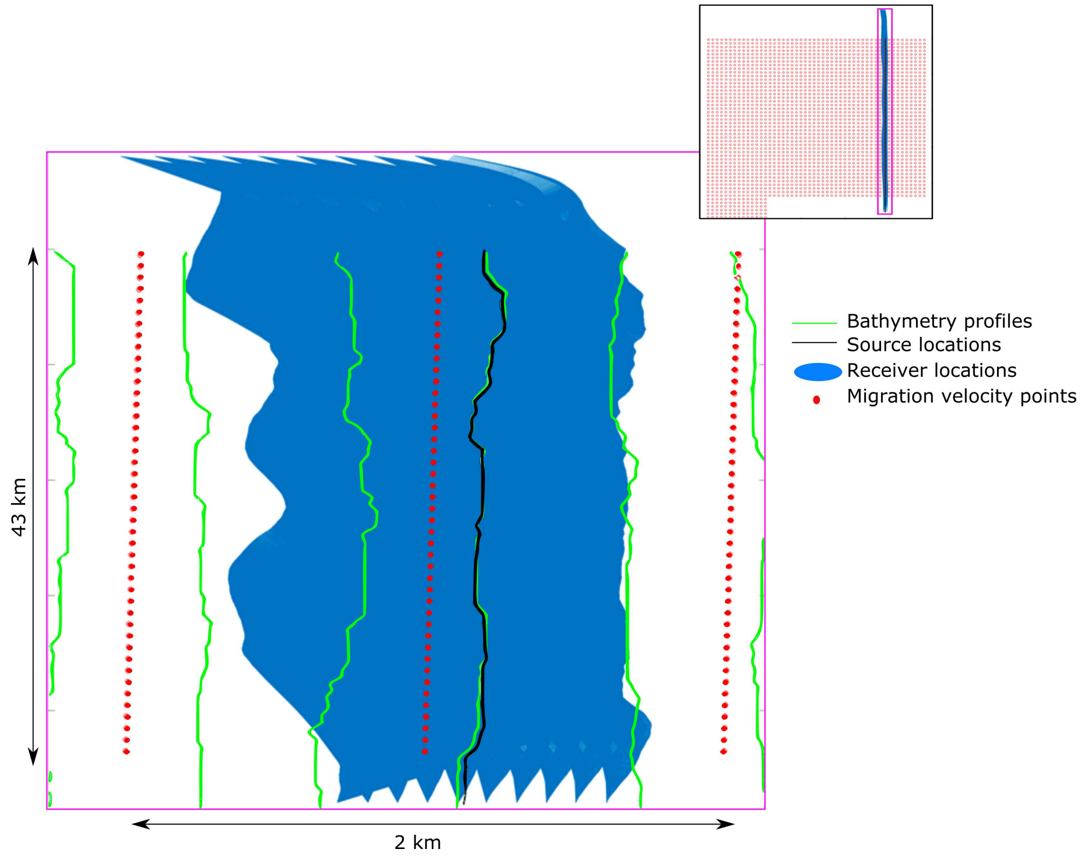
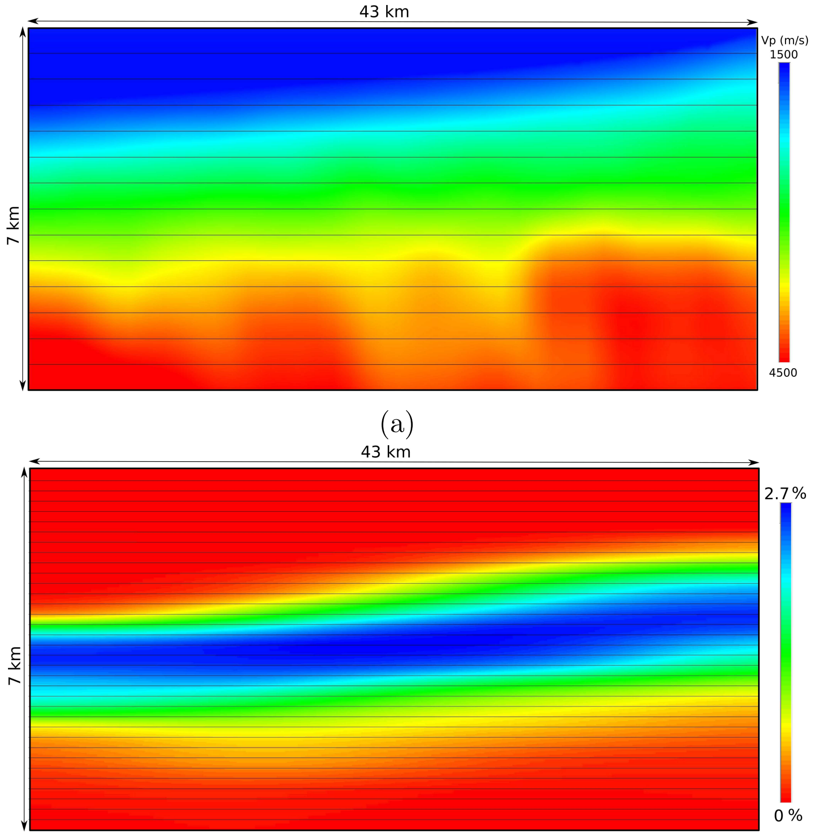


Figure 6.4: Location of a 3D swath with respect to migration velocity points and bathymetry profiles.



(a)

(b)

Figure 6.5: Starting models for the Gabon dataset: (a) velocity and (b) anisotropy (Eta).

6.4 Modelling and inversion strategy

The grid size of the model and the time sampling of the data were defined using similar logic as described in the previous chapters. For the Gabon dataset, the starting model velocities ranged from 1490 m/s in the water column to 5000 m/s in the sediments.

The high kernel version of the code, that is 10^{th} order accurate in space and 4^{th} order accurate in time, was used. Initially, it was planned to invert the data with the frequencies at least as high as 12 Hz. Because this code allows the wavefield to cross no more than 66% of a grid cell at every time step, it would provide minimal dispersion of frequencies below 12 Hz and numerical stability using the 50 m grid spacing and 6 ms time sampling.

The low kernel version of the code which was used during one of the tests is 4^{th} order accurate in space, 2^{nd} order accurate in time and requires at least 5 grid points per wavelet. Given the same grid spacing of 50 m and the same time sampling of 6 ms, it would be numerically stable and provide minimal numerical dispersion of frequencies below 6 Hz.

The starting model had dimensions of 2 x 43 x 7 km. These dimensions translate into 41 x 861 x 141 grid cells provided the chosen grid spacing of 50 m. Allowing for the boundaries, the model size arrived to 81 x 901 x 221 grid cells.

The dataset contained more than 1900 sources acquired with dual air-gun array operating flip-flop with the interval of 25 m. In order to save computational effort, which is directly proportional to the number of sources, it was decided to reduce the number of shots and use the shot gathers recorded from one air-gun array. The total number of sources was 770.

During the inversion, 154 shots were inverted simultaneously at every iteration. There were 5 iterations per every frequency band, which allows the inversion to use every shot once per every frequency band. The inversion has started from 3 Hz and in most cases the data were inverted to the maximum frequency of 6 Hz using 6 frequency bands with central frequencies of 3.0, 3.4, 3.9, 4.5, 5.2 and 6.0 Hz. In one of the inversion runs the data were inverted using the maximum frequency of 12 Hz with central frequencies of 3.0, 3.4, 3.9, 4.5, 5.2, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0 and 12.0 Hz. The inversion was performed for slowness rather than velocity and after every iteration the model was smoothed in both inline and crossline directions using a Gaussian filter with the size of 0.5 wavelength. The inversion parameters are summarised in table 6.1.

Model size	2 x 43 x 7 km
Trace length	8000 ms
Number of shots	770
Number of shots per iteration	154
Number of iterations per frequency band	5
Starting frequency	3 Hz
Maximum frequency	12 Hz
Total number of frequency bands	3.0, 3.4, 3.9, 4.5, 5.2, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0 ,12.0 Hz
Frequency bands	12
Total number of iterations	60
Grid spacing	50 m
Model size	41 x 861 x 141 cells
Model size with boundaries	81 x 901 x 221 cells
Time sampling	6 ms
Time samples	1333
Invert for slowness	yes
Normalise amplitudes	yes
Spatial preconditioning	yes

Table 6.1: Summary of the inversion parameters for the Gabon dataset.

6.5 Results

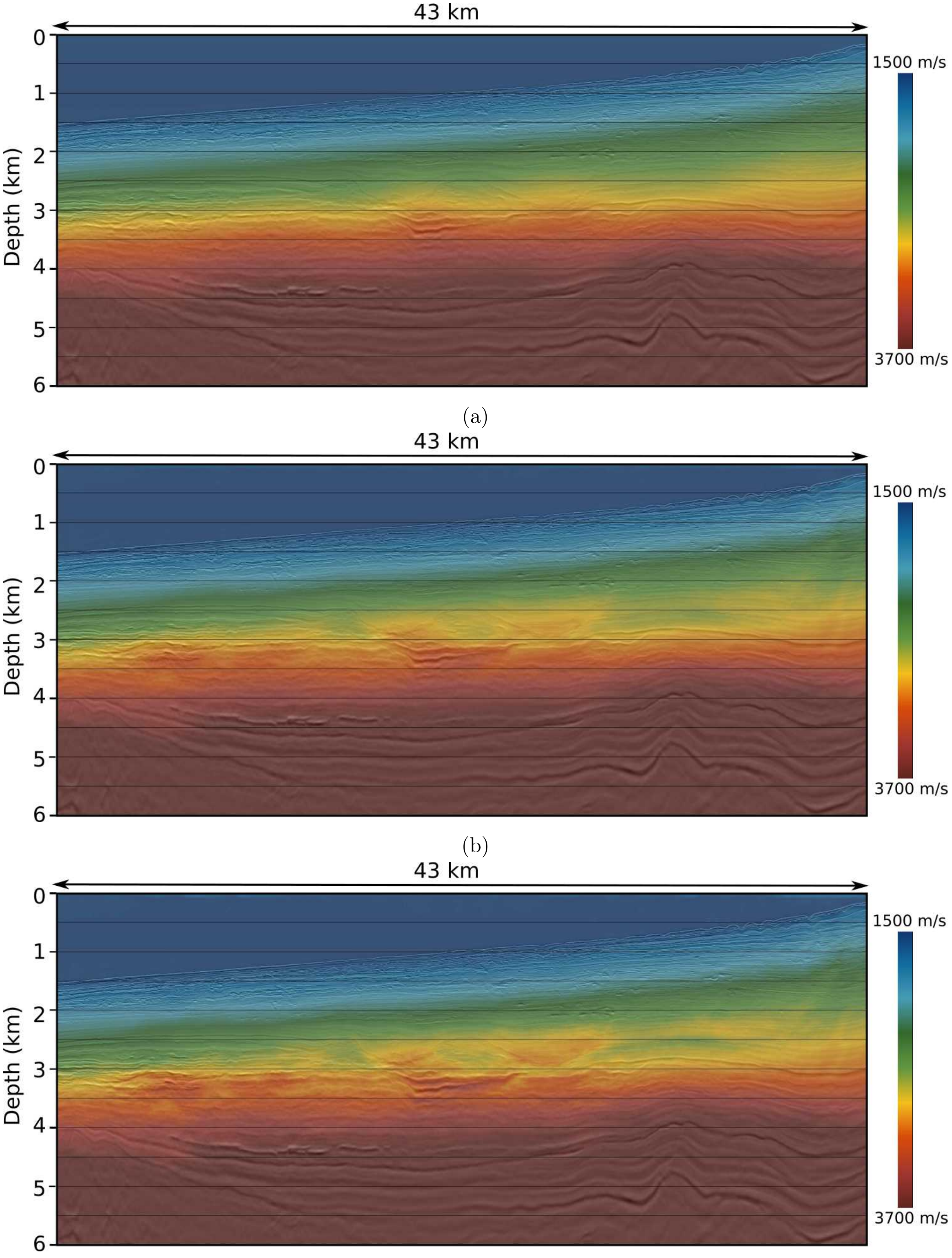
The dataset was inverted many times using anisotropic acoustic FWI with the starting model based upon smoothed migration velocities. The field data were 3-12 Hz band-passed with an Ormsby minimum phase filter and had additional noise reduction applied. The noise reduction consisted of removing bad traces that are contaminated with swell noise, muting the data ahead of first arrivals and below 6s TWT using an offset-variant bottom mute. In total, approximately 30% of traces were removed, reducing the number of receivers from approximately 4.5 to 3.8 million. The noised reduction is described later in this chapter in section 6.6.4. The field data with reduced noise are shown in the same section in Figure 6.17.

The inversion began at 3 Hz and was performed to a maximum frequency of 12 Hz. In total, there were 60 iterations. Figure 6.6 displays the starting velocity model and FWI models at 6 and 12 Hz. In this figure, the velocity models are overlaid with the depth section obtained from the contractor's PSTM volume using the starting velocity model.

The starting model is smooth and relatively one-dimensional with velocities gradually increasing with depth. The lateral velocity variation is due to rapidly changing water depth, that ranges from 1550 m on the left-hand side of the section to 200 m on the right-hand side. The object of interest at the depth of approximately 3 km in the central part of the section is not defined in the starting model. Various shallow geological features, that are demonstrated by the depth section, are also not present in the starting model.

At 6 Hz, FWI seems to pick up velocity anomalies at the depth of about 3 km with the biggest updates produced near the area of interest, located in the central part of the section and representing lithology uncertainty. It can be seen, that the bright reflector immediately below this main central feature is defined as high-velocity anomaly as well as two large basins on the both sides from it. On the left hand-side of the section, the bright reflector at approximately 3 km depth is also picked up by FWI.

At 12 Hz, these same features became more pronounced. Two basins in the central part of the section are more resolved as well as the bright reflector on the left-hand side of the section. The velocity updates in the deeper part of the section are small and, therefore, not visible due to the saturated colour scheme and will be shown later in this section.



(a)

(b)

(c)

Figure 6.6: (a) Starting velocity model and FWI models at (b) 6 Hz and (c) 12 Hz. The velocity models are overlaid with the depth section obtained from the contractor's PSTM volume using the starting velocities.

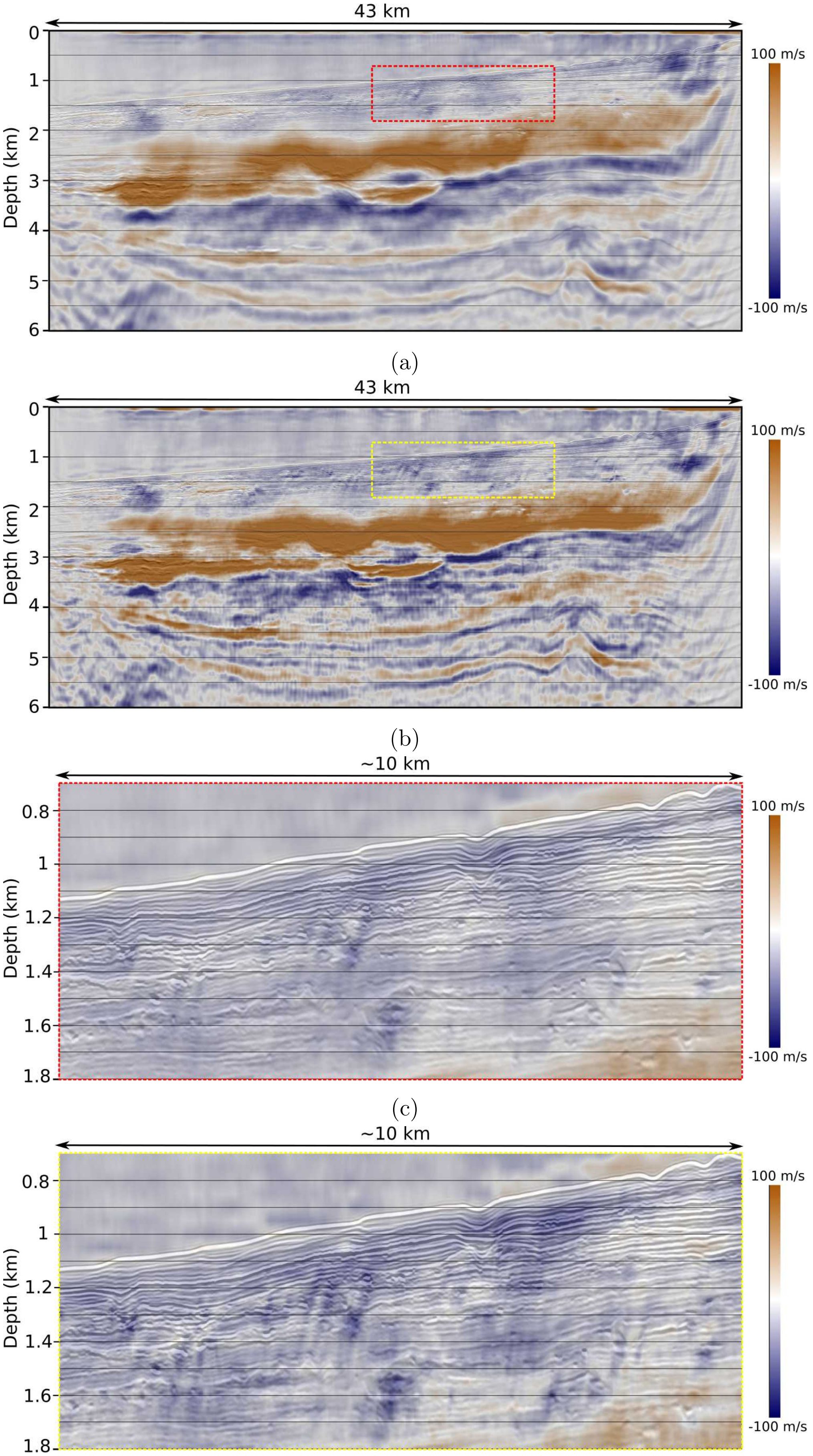
Both FWI models, at 6 and 12 Hz, are remarkably similar and both demonstrate unexpected behaviour in the very shallow part of the model. As can be noticed from Figure 6.6(b) and Figure 6.6(c), FWI models seem to generate very small or no shallow updates at all. This is very unusual for FWI because conventional FWI is perfectly capable of improving a particularly shallow velocity model due to diving-waves penetration of depths usually above 2-3 km.

To eliminate the impact of the colour scale, a series of velocity model differences were produced with close-ups focusing on the shallow depth area of the model. Figure 6.7(a) and Figure 6.7(b) show the FWI velocity updates at 6 and 12 Hz respectively. These updates are the result of subtracting the corresponding FWI model from the starting migration velocities. The models are overlaid with the depth section obtained from the contractor's PSTM volume.

It is obvious, that the majority of the updates lie below 2 km with the main update area being in the range of 2 to 3.5 km. The deeper part of the model seems to be relatively sensibly updated during FWI. The 6 Hz FWI model, shown in Figure 6.7(a), suggests that the updates as deep as 5.5 km are correlated with the depth seismic section. For instance, at the depth range of approximately 4 to 5 km, a thick sedimentary layer is pinching out from the right- to the left-hand side of the model. The boundaries of this layer are defined in FWI models by high-velocity anomalies and follow the reflectors shown by the depth seismic section.

However, FWI seems to fail to resolve the upper part of the section. Figure 6.7(c) and Figure 6.7(d) show the close-ups of Figure 6.7(a) and Figure 6.7(b) with the focused area indicated by the red and yellow rectangles respectively. The lateral extension of the focused area is approximately 10 km, the vertical extension is 1.1 km. The depth section, overlaid with the FWI velocity updates, clearly demonstrates the presence of multiple little features, layers, mini-basins and other complexities in the shallow area. FWI, however, produces small, predominantly negative, updates and seems not to have much of an affect in the shallow part of the model even if pushed as high as 12 Hz in frequencies. The shot gathers for the this run are shown in Figure 6.20 and are discussed within the analysis frame in section 6.6.4.

FWI seems to stall in the upper part of the model and, consequently, does not improve the shallow velocities, which is very unusual. The way FWI stalls is described in section 6.6.6. In the analysis below, I will try to explore possible reasons for such FWI behaviour.



(a)

(b)

(c)

(d)

Figure 6.7: Velocity updates generated by FWI at (a) 6 Hz and (b) 12 Hz with the focus on shallow depth area at (c) 6 Hz and (d) 12 Hz. The close-ups are indicated in red and yellow rectangles for 6 and 12 Hz FWI respectively. The models are overlaid with the depth section obtained from the contractor's PSTM volume using the starting velocity model.

6.6 Analysis and discussions

As has been demonstrated in the previous section, conventional FWI applied to the Gabon dataset seems to fail in the most unusual way - it stalls and does not improve the shallow velocity model. In this section, I will explore possible reasons that could cause such behaviour in the hope to find out why this is happening.

6.6.1 The source accurate?

The importance of having a correct source signature and possible ways to estimate it is discussed in detail in chapter 4. A source that is significantly incorrect can result in FWI heading towards local minima or failing completely.

In this section, I am going to test whether the source signature used for the inversion tests in this study is appropriate for FWI.

The 3D source signature was obtained using an initial estimate to model the direct arrival through the water, subsequently using a matching Wiener filter to match this to the observed direct arrival. Because part of the dataset is situated in deep water, the direct arrival is not interfering with the water-bottom arrival and is easy to use. The approach is described step-by-step in chapter 4. Here, the filter applied to both direct arrivals and the initial guess source, given as a delta function, was a minimum phase Ormsby filter rolling off from 9 to 12 Hz. The water velocity for the modelling was carefully estimated from the gradient of the direct arrival using a number of randomly selected shot gathers.

The derived source wavelet is shown in Figure 6.8(a). It contains frequencies with the maximum of 12 Hz and has a “tail” which is very likely to be the bubble oscillation.

To verify whether the source signature was appropriate for the inversion, a number of shot gathers were modelled using the estimated shallow water velocity. The observed and calculated direct arrivals were compared. Figure 6.8(c) shows a trace-by-trace comparison between the two arrivals with the first trace being the recorded data and the second trace being the calculated data. This figure is a close up of Figure 6.8(b) and focuses on the direct arrival on the short offsets only. The data were trace-equalised, consequently, only the relative amplitudes can be compared. It is obvious that the derived source signature provides a very close match between the observed and the predicted data.

Therefore, it was concluded that the source wavelet is appropriate for FWI and it was used during all the inversions conducted in this study.

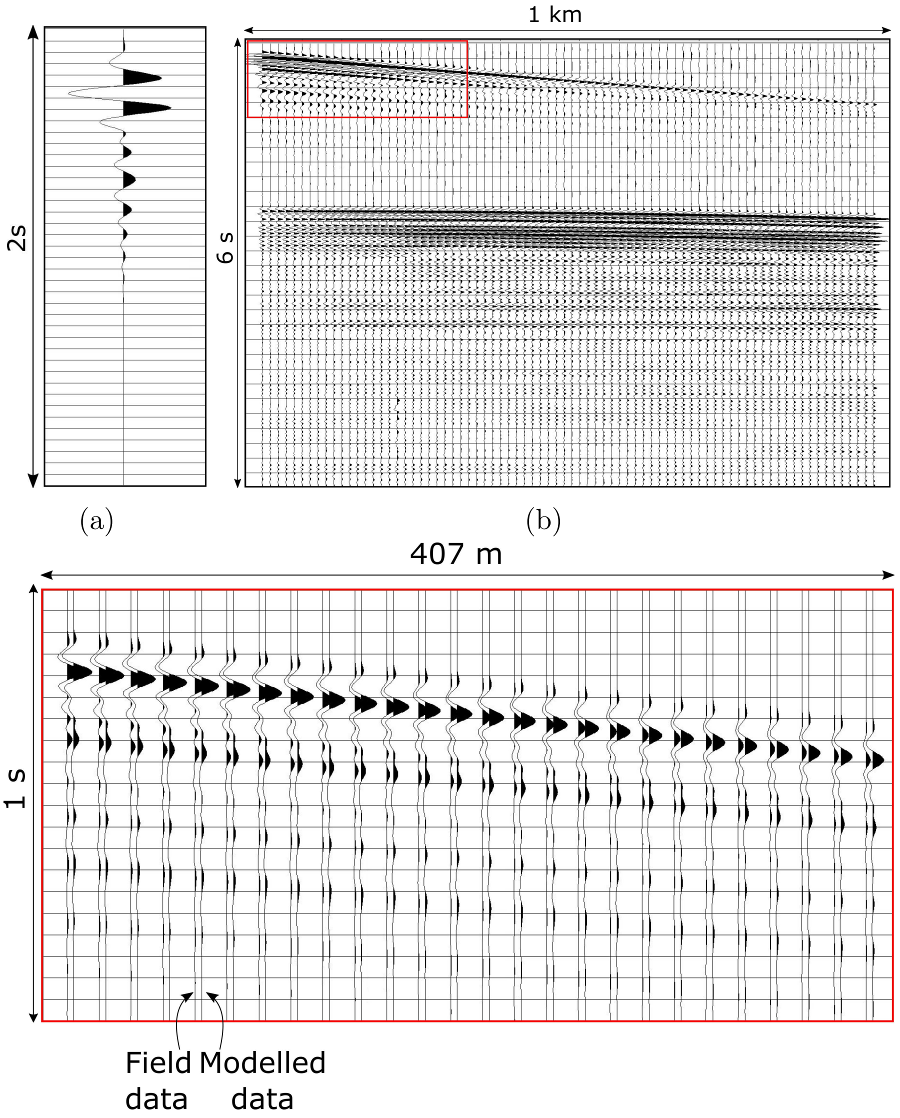


Figure 6.8: Validation of the source signature for the Gabon dataset: (a) the source signature, (b) a raw shot gather filtered at 12 Hz with the short offsets displayed. The shallow water velocity is determined using the gradient of the direct arrival at near-offsets and is found to be 1520 m/s. (c) Trace-by-trace comparison of the observed and modelled direct arrivals using the derived source signature (a). The shot gathers are trace-by-trace normalised.

6.6.2 The starting model cycle skipped?

An accurate starting model is a vital component of successful FWI. For conventional FWI, an accurate starting model is a model that predicts the data that match the field data within half a cycle. In order for the algorithm to converge to the global minimum rather than a local minimum, the model must be accurate at the lowest frequency used in the inversion. The importance of having an accurate starting model is discussed in detail in the theory chapter in section 2.3.

The contractor provided an anisotropic starting model based upon the smoothed migration velocities, with a constant water velocity of 1500 m/s. This model is described earlier in the chapter. Figure 6.10 displays eight raw shot gathers and corresponding modelled gathers using this starting velocity model. The shot gathers are evenly distributed across the model, from deep to shallow water, the location of the shots is indicated in Figure 6.11(a).

It can be seen, that the starting model provides a reasonable match at the very short offsets. At all the displayed shot gathers the water bottom arrival at the nearest offsets is clearly not cycle skipped. However, the long offsets are less obvious for analysis. Because there is a “gap” in the field data at the middle offsets, it is difficult to identify the first phase of the water bottom arrival at the long offsets. At some gathers, where this “gap” is absent or less pronounced, the careful comparison of the field and the modelled data indicates that the water bottom reflection at the long offsets arrives ahead of the field data and, consequently, the long offsets do not provide the required match of half a cycle between the field and the modelled data.

Reasons for the mismatch between the field and modelled energy pack at the long only offsets can be: (1) the starting model is too fast in the deeper part of the section and/or (2) constant water velocity is not good enough for the deep water environment (so that the first seabed arrivals will be cycle skipped). To test the first potential problem, a series of starting models was generated by reducing the sediments velocity of the starting model by 0.95, 0.9, 0.8 and 0.7 from its initial value. The gradient of the obtained starting models is shown in Figure 6.9.

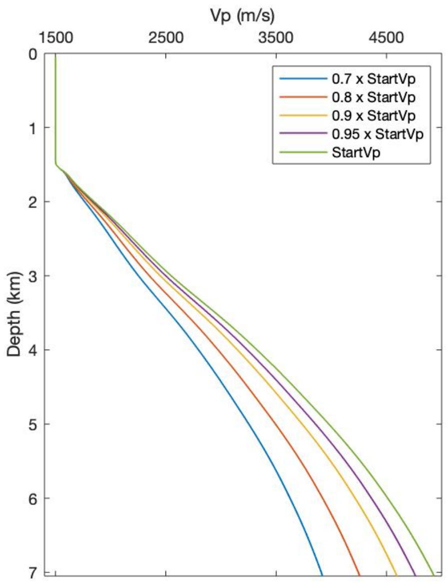


Figure 6.9: A suite of 1-D profiles of the starting velocity models for the Gabon dataset. A trace from the contractor's starting model is shown in green, followed by the suite of its modified versions by changing the velocity gradient below the seabed to 0.95, 0.9, 0.8 and 0.7 from its initial value. Water velocity is constant everywhere and equals to 1500 m/s.

Here, a single trace from the initial velocity model is displayed in green. The same trace, with steepening gradient is shown in purple, yellow, red and blue. The water velocity was kept constant everywhere as it was in the initial starting model.

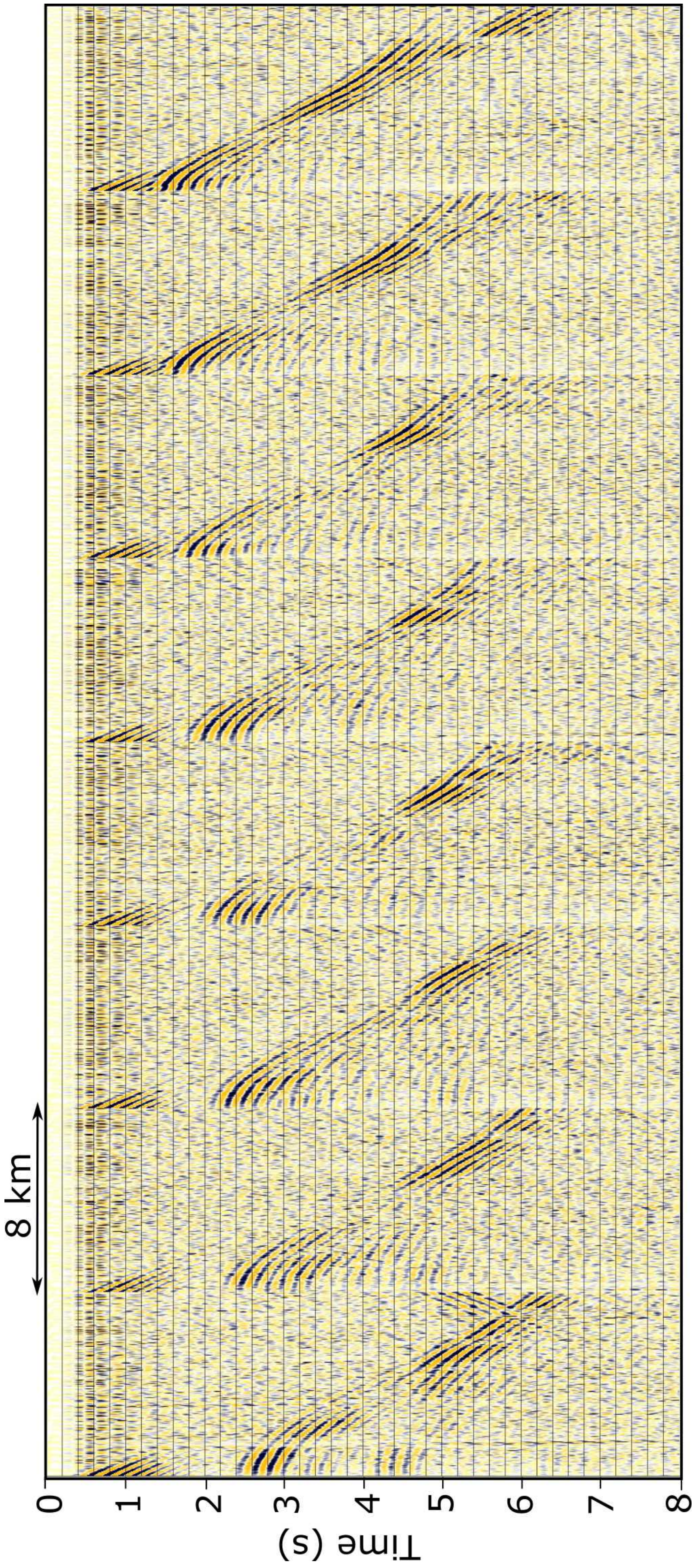
After four new velocity models have been obtained, the synthetic data were generated for each of them and carefully analysed. The models with the gradient of 0.7 and 0.8 of its initial value generated the data that clearly and consistently at every shot gather had larger TWT compared to the field data. This indicates that these models have velocities that are too slow. However, the synthetic gather shown in Figure 6.10(c), which was obtained using the model with the gradient of 0.9 from its initial value, may provide an improved match at the longest offsets (considering the difficulty in identifying the first arrival phase at the long offsets).

The field data were inverted twice to a maximum frequency of 6 Hz using two different starting models: (1) the initial starting model with the constant water velocity of 1500 m/s and (2) its version with the gradient reduced to 90% of its initial value everywhere in the sediments while the water velocity of 1500 m/s was kept constant.

Figure 6.11 shows the depth section and two FWI models at 6 Hz. Both models produce qualitatively somewhat similar velocity anomalies. However, these anomalies have different positions in depth and different velocity values. The depth position of the anomalies does not precisely match the reflectors on the depth section. It is likely, that the strong low-velocity anomaly that traverses the section (displayed in green) is the first water bottom multiple being picked up by the inversion.

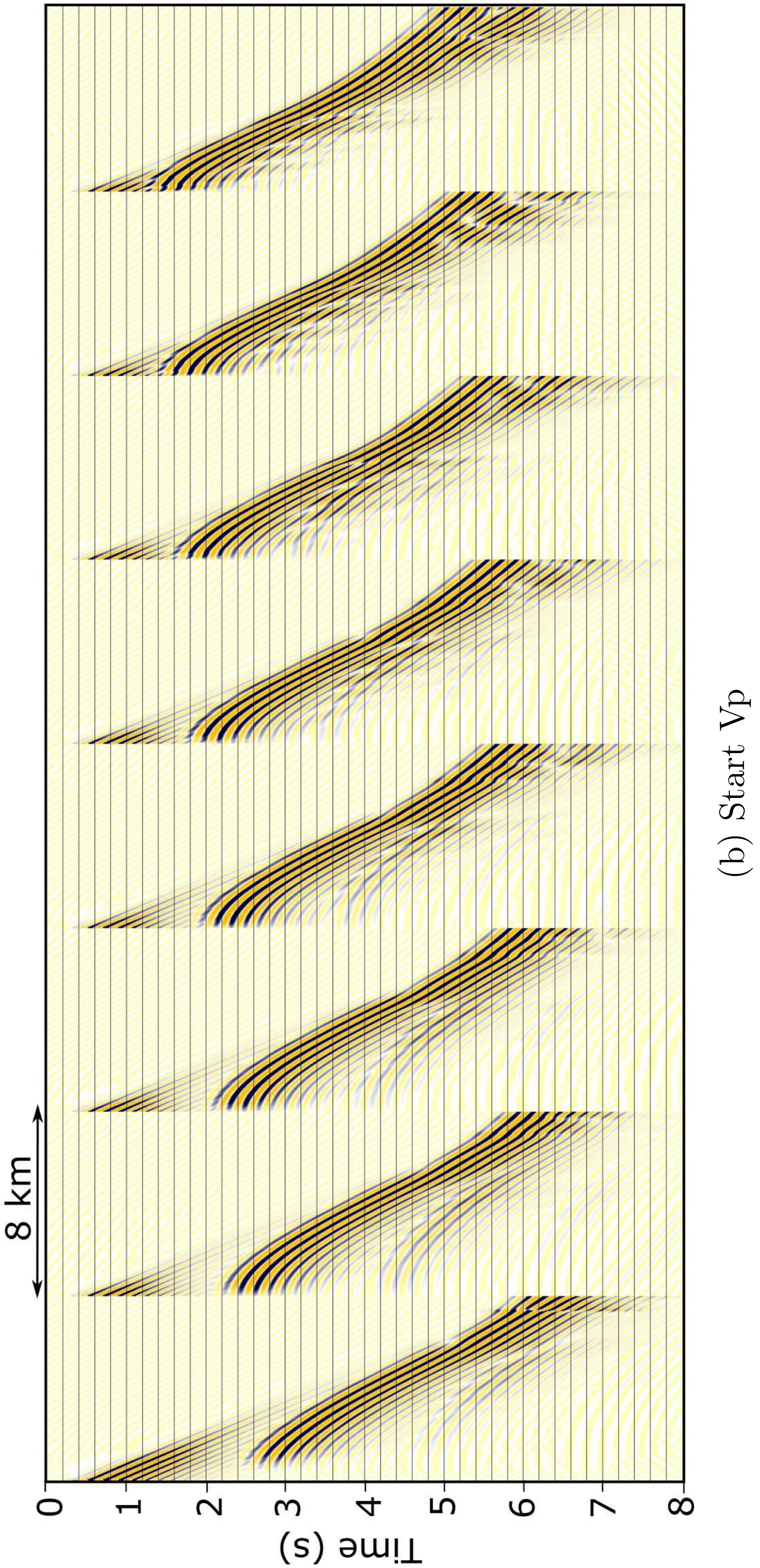
Both models appear to look cycle skipped. The shot gathers obtained using these two models are shown in Figure 6.12. Comparison of the modelled gathers with the raw data indicates that the FWI models are cycle skipped at the long offsets.

This may mean that the original starting velocity model is not cycle skipped due to the “too high” velocities in the deeper part of the section, but the mismatch between the modelled and field data at long offsets might be caused by kinematic errors in water bottom reflection due to ignoring the variable velocities in the water column. Water column impact on the inversion will be explored during the next step of this analysis.



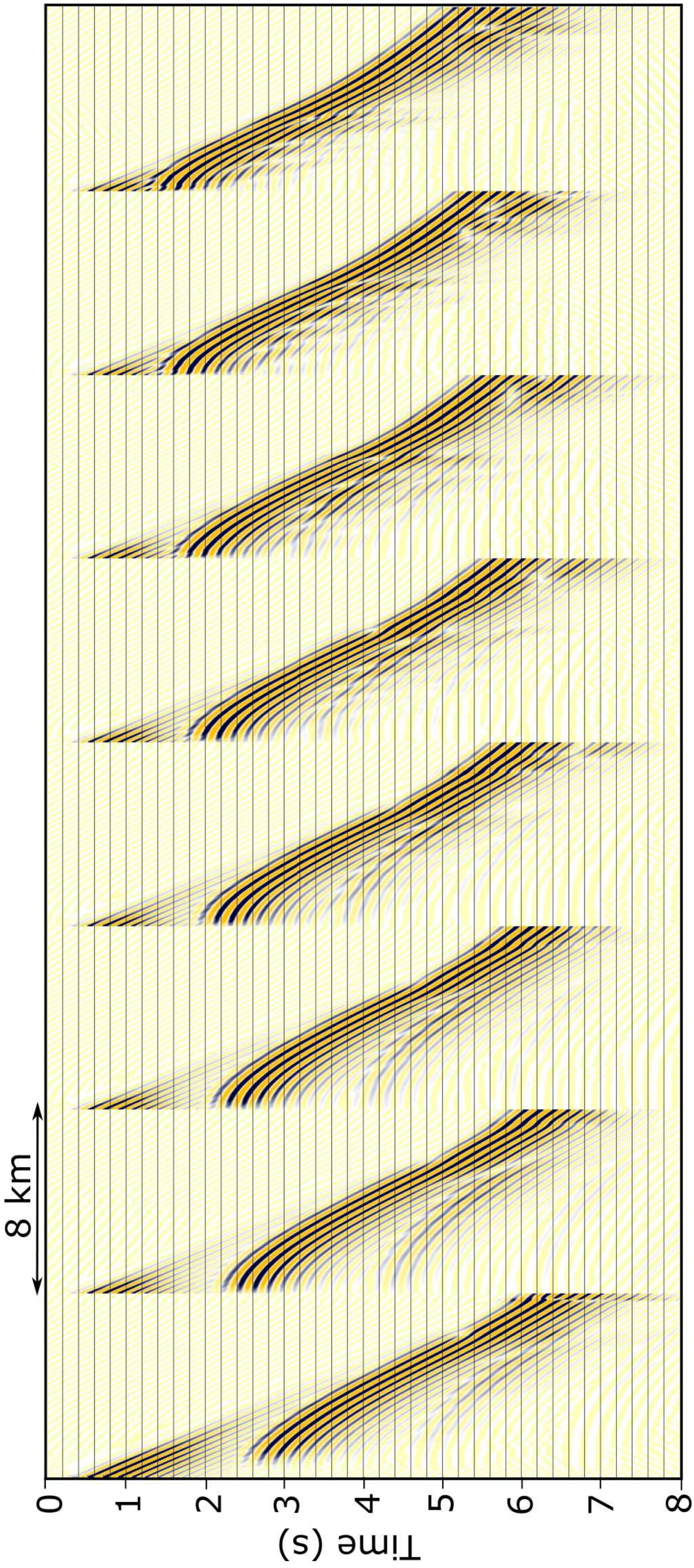
(a) Raw

Figure 6.10: Comparison of the shot gathers: (a) raw and modelled using (b) the initial starting model and (c) the model with the gradient of 0.9 from its initial value (cont.).



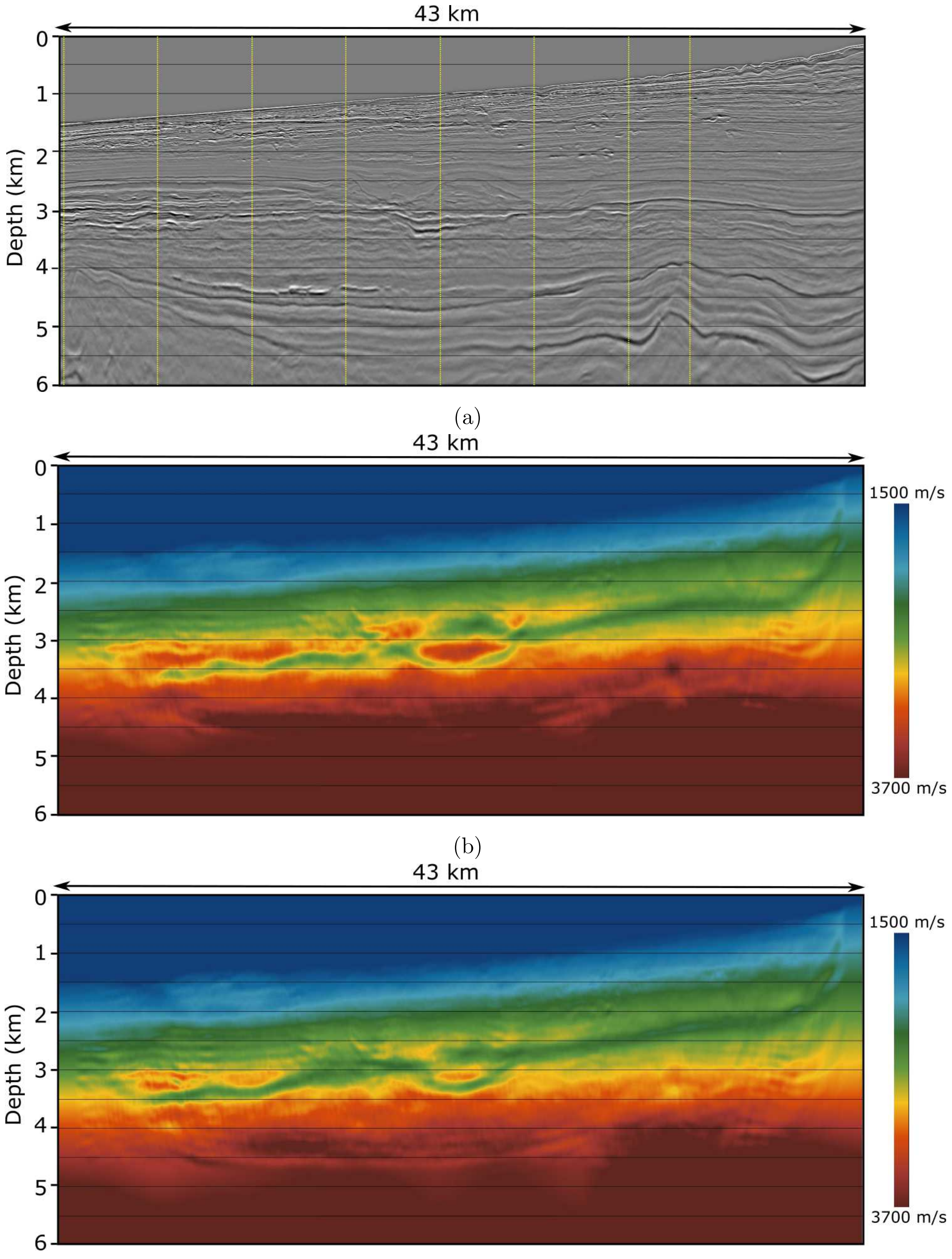
(b) Start Vp

Figure 6.10: Comparison of the shot gathers: (a) raw and modelled using (b) the initial starting model and (c) the model with the gradient of 0.9 from its initial value (cont.).



(c) 0.9 x StartVp

Figure 6.10: Comparison of the shot gathers: (a) raw and modelled using (b) the initial starting model and (c) the model with the gradient of 0.9 from its initial value.

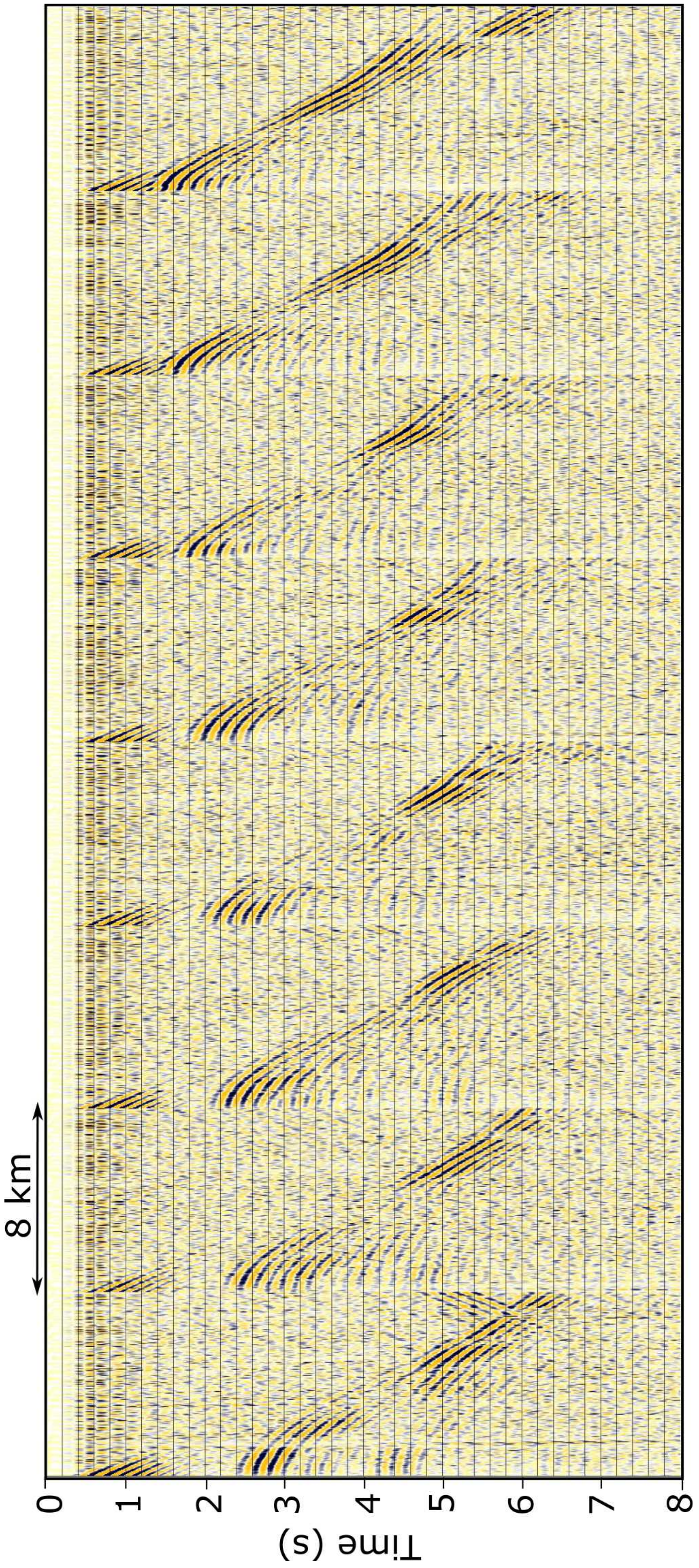


(a)

(b)

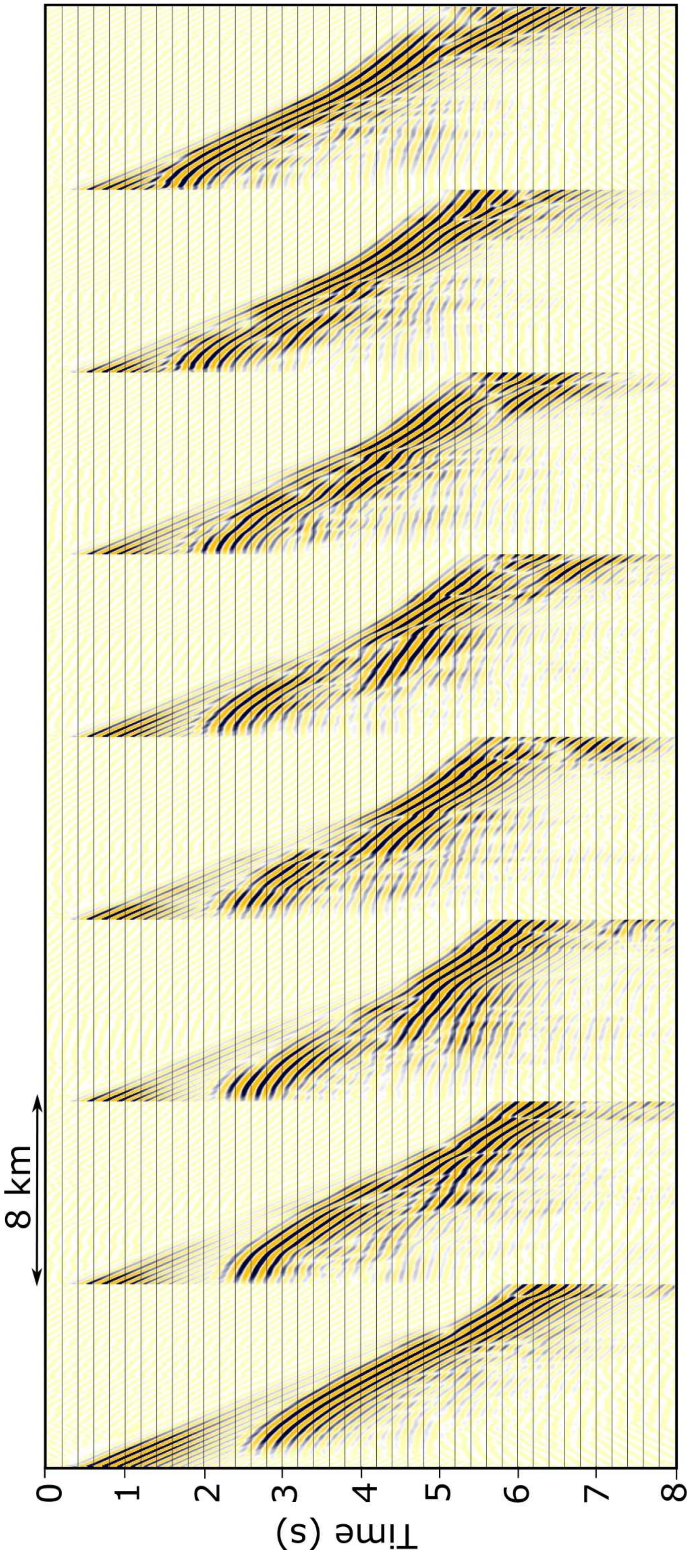
(c)

Figure 6.11: Comparison of (a) the depth section to 6 Hz FWI velocity models obtained using: (b) the initial starting model and (c) 0.9 times slower model. The yellow lines indicate shot locations used for quality control forward modelling.



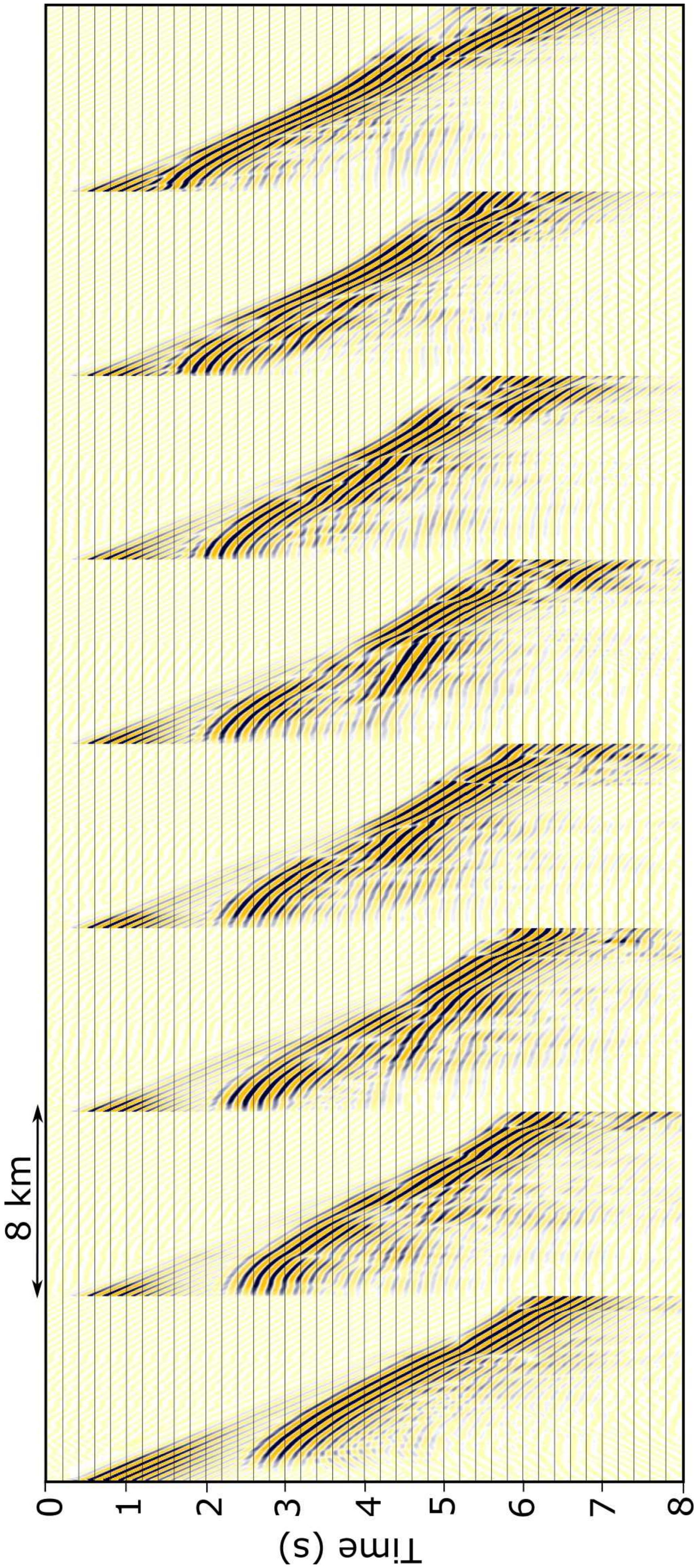
(a) Raw

Figure 6.12: Comparison of the shot gathers (a) raw and modelled with FWI models obtained using: (b) the initial starting model and (c) 0.9 times slower model (cont.).



(b) FWI using StartVp

Figure 6.12: Comparison of the shot gathers (a) raw and modelled with FWI models obtained using: (b) the initial starting model and (c) 0.9 times slower model (cont.).



(c) FWI using 0.9 x StartVp

Figure 6.12: Comparison of the shot gathers (a) raw and modelled with FWI models obtained using: (b) the initial starting model and (c) 0.9 times slower model.

6.6.3 Water column matters?

The original contractor's starting model has a constant water velocity of 1500 m/s. The previous inversions have used these original velocities and shown a kinematic mismatch of the energy recorded at the long offsets. For some of the shot gathers, it is difficult to say whether the first arrival at the long offsets is a purely water bottom arrival or the result of interference between the water bottom and other arrivals. However, the inversion attempted to account for this mismatch by increasing shallow water velocity, which indicates that the mismatch may be due to a constant water velocity assumption. Here, I will try to explore the impact of such an assumption.

In reality, the water velocity in the ocean varies with depth (Urick 1979). While in the shallow water settings it can be roughly approximated to a constant value due to the column being well mixed and, consequently, homogeneous (Kuperman et al. 2004), in deep water settings the column is usually not mixed and therefore velocity is not homogeneous (Urick 1979).

The speed of sound through water depends upon its physical properties, such as temperature and salinity (Kamaluddin et al. 2011). Generally, the very shallow water layer tends to have the highest velocity, with an immediate sudden drop in velocities below this layer, then followed by a slow gradual decrease in velocity and a slow rise again with increasing depth. In other words, the speed of the acoustic wave is not constant during its propagation in the water column. This induces two main problems: (1) the mean speed of sound in water varies with the depth of the seabed and (2) the mean speed of the water bottom arrival recorded by short and long offsets is different. This is because at zero offset, the velocity obtained using the combination of the water depth and time gives the mean velocity; but refraction causes the moveout velocity to be higher than this, equal approximately to the RMS velocity. Refraction makes the rays more vertical in the slow layers, and more oblique in the fast layers, so that the effective move-out velocity increases. This results in a mismatch at the long offsets between the observed and modelled water bottom arrival if the constant water velocity is assumed.

In order to improve the starting model, the depth-variable water velocity profile can be obtained using an empirical formula, that is a function of temperature, salinity and depth.

All these parameters can be measured using different equipments or calculated (Kamaluddin et al. 2011). In this study, the high resolution vertical temperature and salinity profiles with respect to depth were obtained from the National Oceanic and Atmospheric Administration's (NOAA) World Ocean Database (WOD) (Boyer et al. 2013). These profiles were collected in December 2013 using Conductivity Temperature Depth probes (CTD) with the cruise ID FR-15741 and WOD unique cast number 15848413.

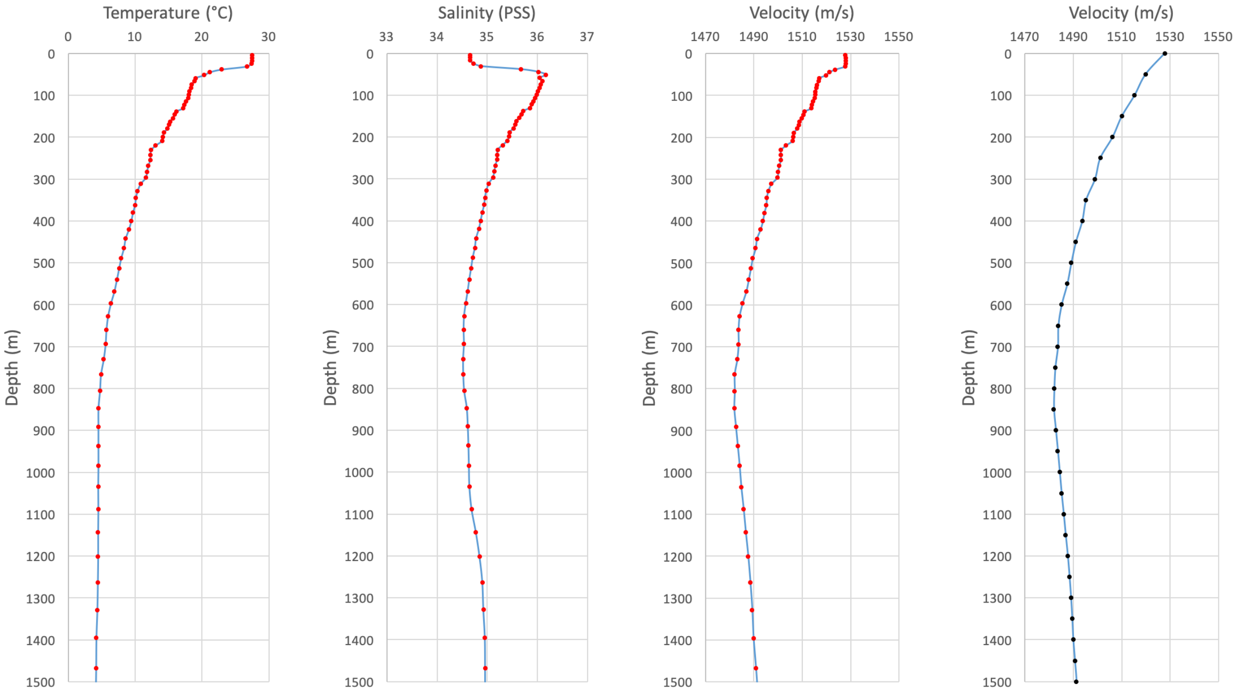
There is a number of empirical relationships available to calculate the speed of sound in water using these profiles. Kamaluddin et al. (2011) suggested that the most popular and accurate are those of Del Grosso (1974), Medwin (1975), Mackenzie (1981)and Chen et al. (1977). For this study, Medwin (1975) relationship was used:

$$V_{water}(t, S, z) = 1449.2 + 4.6t - 0.055t^2 + 0.00029t^3 + (1.34 - 0.01t)(S - 35) + 0.016z \quad (6.2)$$

where t denotes temperature in degrees Celsius, S denotes salinity in parts per thousand and z is the depth in metres. Medwin equation is valid in the ranges for temperature of 0 to 35°C and for salinity of 0 to 40 parts per thousand (Kamaluddin et al. 2011).

The calculated profile of wave propagation velocity in the water column is shown in Figure 6.13(c). In the initial 50 m the velocity is relatively constant at approximately 1529 m/s which then gradually reduces to 1480 m/s at 750 m. The rest of the profile down to the maximum depth is characterised by a gradual velocity increase to 1492 m/s. The average velocity of the calculated water column is 1500.1 m/s, which coincides with the average water column velocity of 1500 m/s estimated during migrational velocity analysis by the processing contractor. This indicates that there will be no kinematic changes in the arrival time of the water bottom reflection at the near offsets when modelling the data. However, the velocity is predominantly lower than 1500 m/s and this will change the arrival time of the water bottom reflection at the long offsets as well as first-order multiple. In addition, this might affect the dynamics of the water bottom arrival by altering the impedance contrast across the seabed.

The calculated velocity profile was then smoothed in slowness, resampled to a grid size of 50 m (Fig. 6.13(d)), extrapolated in the x and y directions and inserted into the starting velocity model to substitute the water velocity.

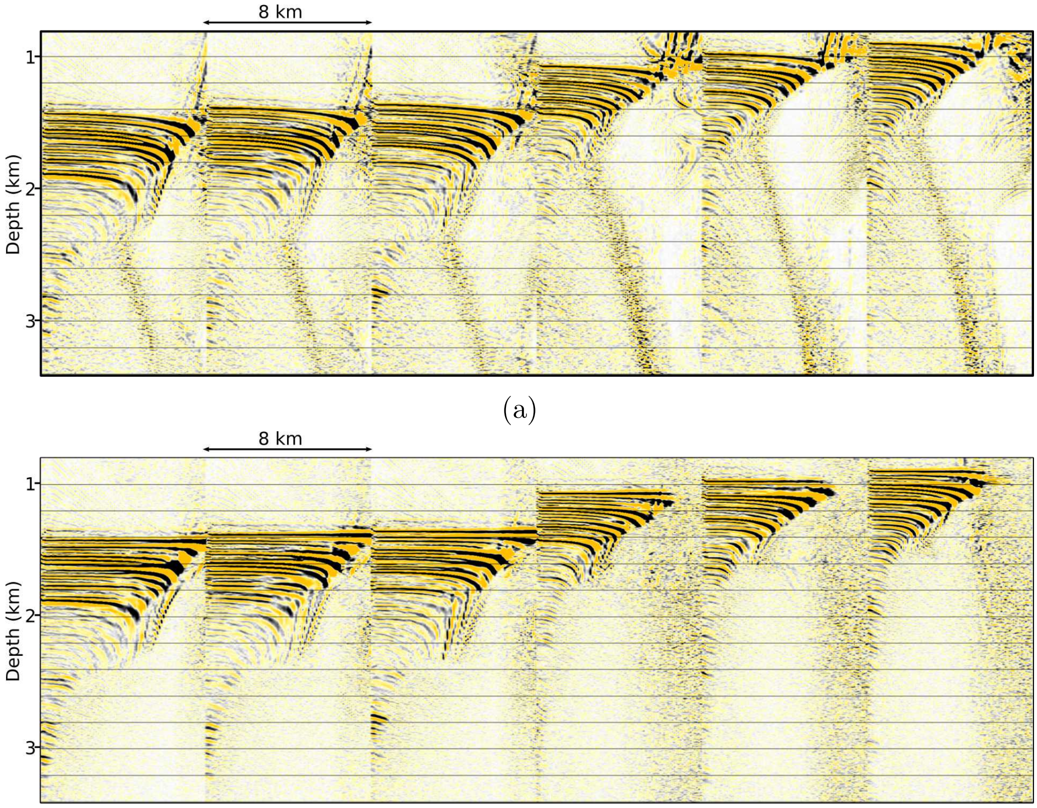


(a) (b) (c) (d)

Figure 6.13: Vertical profiles of the water column: (a) temperature, (b) salinity, (c) velocity calculated from (a) and (b) using equation 6.2. The red markers denote the measurement points. The velocity profile (c) smoothed in slowness and resampled to 50m grid is shown in (d). The black markers denote grid points.

To validate the new starting velocity model and to explore the impact of the depth-variable velocity profile on the travel times Kirchhoff depth-migration can be performed. Figure 6.14 shows the migrated common image gathers for the selection of shots located as indicated in Figure 6.15(a). These gathers are located in the areas with different water depths, decreasing from the left- to the right-hand side. For the model with constant water velocity, the water bottom reflector appears to be flat on the near offsets and has a gradually decreasing move out at long offsets with the depth decrease. For the model with depth-variable velocity profile, in contrast, the water bottom reflector appears to be flat at full offsets range across all the gathers. This indicates, that the depth-variable water velocities, that are commonly neglected during the inversion, provide accurate travel times of the first arrivals for the full offset range in deep water settings.

The inversion was carried out many times using the same data, the same source signature, same inversion strategy but different starting models (with respect to water velocity) and different density parameters.



(a)

(b)

Figure 6.14: Kirchhoff-migrated common image gathers obtained using (a) the starting model with constant water velocity and (b) the starting model with the depth-variable velocity profile. The full offset is shown.

Here I will present 5 inversion tests:

1. *Test 1*

   The inversion uses a starting velocity model with a constant water velocity value being 1500 m/s. This model was initially provided by the processing contractor. The lowest sediment velocity was identified after careful visual examination of the model and set to 1560 m/s, which effectively defines the starting density model. The algorithm will use constant density everywhere in water and Gardner equation everywhere else to calculate the density from the velocity equal and above the minimum sediment velocity. The detailed description of how the density model is computed is given in section 4.4.2. Since no mutes were applied ahead of the water bottom arrival and the direct arrivals were retained in the data, the water column was kept constant during the inversion at every iteration, allowing for the updates only in the sediments.

2. *Test 2*

   The inversion uses a starting model with a constant water velocity of 1495 m/s. The depth position of the water bottom was adjusted appropriately in order to preserve the original travel times. The lowest sediment velocity was kept as 1560 m/s. As in the previous test, the inversion did not update the water velocity. This test was carried out to in order to explore the effect of slightly different average water velocity on the inversion provided that the travel times of the short offsets are identical.

3. *Test 3*

   The inversion uses a starting model with the depth-variable velocity profile described earlier in this section. The seabed depth position was kept the same as in Test 1 because the average of the depth variable profile was found to be 1500.1 m/s. The comparison of this inversion to Test 1 shows the difference between using a constant versus a variable depth-profiles.

4. *Test 4*

   The inversion uses the identical starting density and velocity model to Test 3, where the velocity model has the depth-variable velocity profile. Unlike Test 3, the inversion is allowed to update velocities everywhere including the water column. In theory, if the water velocities are correct and the noise level prior to the seabed arrivals is sufficiently low, then the velocity updates in the water column should be minimal.

5. *Test 5*

   The inversion uses the same starting velocity model used in Test 3. As in Test 3, it also allows for velocity updates below the seabed only. However, this inversion uses a different starting density model, calculated based on the new minimum sediment velocity of 1540 m/s. Compared to Test 3, this inversion demonstrates the impact of the density model.

To summarise, two out of five inversions use a starting model with constant water velocity of 1500 m/s and 1495 m/s. The rest uses the same depth-variable water profile. Effectively, three starting velocity models with different water columns were used for five inversion runs. For the model utilising the depth-variable water velocities, two additional tests were conducted by (1) allowing the inversion to change velocities everywhere in the model, including water and (2) using a different density model. A summary of the inversion parameters for these tests is shown in Table 6.2. These are in essence a replication of the same inversion strategy described previously in section 6.4, however with the differences in starting models and inversion restrictions included.

Five different FWI models obtained during the tests and overlaid with the depth section are presented in Figure 6.15. Note, that the shallow and the deep parts of every model are shown in different colour schemes. Since all of these inversions have stalled in the shallow part of the section, the colour scheme for the shallow depths was chosen to clearly demonstrate the position of the seabed and water velocities.

Figure 6.15(a) and (b) shows the FWI models obtained using the starting models with the different water velocity, both being constant. The shot gathers for one of them (Test 1) are displayed in Figure 6.12(b). Although the combination of water velocity and seabed depth in the starting model provide correct travel times of the first water bottom arrivals recorded by the near offsets (Fig. 6.10(b)), it can be seen that the both resulting FWI models seem to be cycle skipped (for the description refer to section 6.6.2). They both produce a distinct low-velocity anomaly that traverses all the section, does not follow the reflectors shown in the depth section and appears to have depth close to that of first-order water bottom multiple. In both of these models, the position of the seabed is pushed down in depth, in the deep water areas by a significant amount. This will result in an FWI model generating predicted data that are badly cycle skipped.

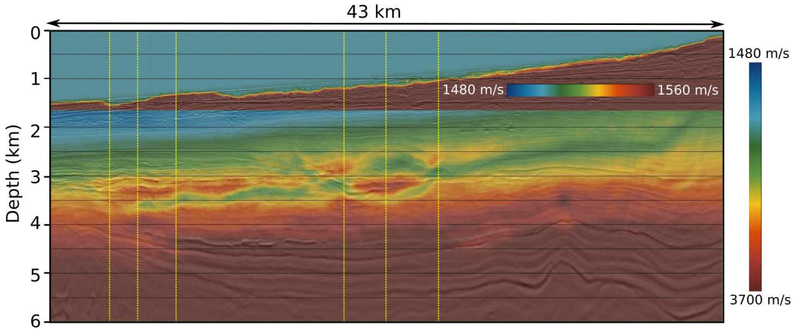
Figure 6.15(c),(d) and (e) shows the FWI models obtained using the starting model with the depth-variable water velocity profile. All these models appear to look sensible and not obviously cycle skipped. They produce qualitatively the same velocity anomalies and in general look very similar to each other. However, they are different in detail and none of them appears to update velocity in the shallow part of the model. All of these inversions have slightly changed the depth of the seabed reflector, which can be seen from the modelled gathers displayed in Figure 6.16. The best fit to the raw data is provided by Test 5 FWI velocity model. Here, water bottom reflector has not been moved down during the inversion due to the inversion restriction on the minimum velocity allowed for the update.

As a conclusion, it was demonstrated that using a depth-variable water velocity profile rather than its average approximation is beneficial for the inversion. It allows kinematics of the first water bottom arrivals at the long offsets to be predicted accurately and therefore, eliminate associated cycle skipping of the first arrival across all the offsets.

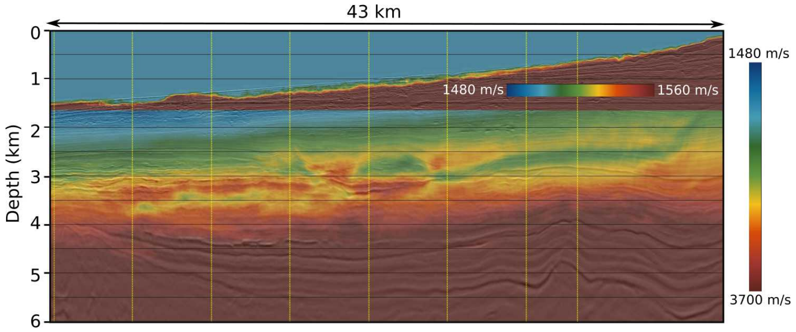
Nonetheless, all the tests showed very little or none updates in the shallow part of the model. This means that FWI stalls for another reason rather than an initially inaccurate water velocity.

	Tests				
	1	2	3	4	5
Starting frequency	3 Hz				
Maximum frequency	6 Hz				
Frequency bands	3, 3.4, 3.9, 4.5, 5.2, 6.0 Hz				
Iterations per frequency	5				
Total iterations	30				
Total number of sources	770				
Sources per iteration	154				
Invert for slowness	yes				
Normalise amplitudes	yes				
Spatial preconditioning	yes				
Water velocity	1500 m/s	1495 m/s	depth-variable profile		
Allow changes in the water	no	no	no	yes	no
Allow to change velocities above	1502 m/s	1497 m/s	1531 m/s	1201 m/s	1542 m/s
Minimum sediment velocity	1560 m/s	1560 m/s	1560 m/s	1560 m/s	1540 m/s

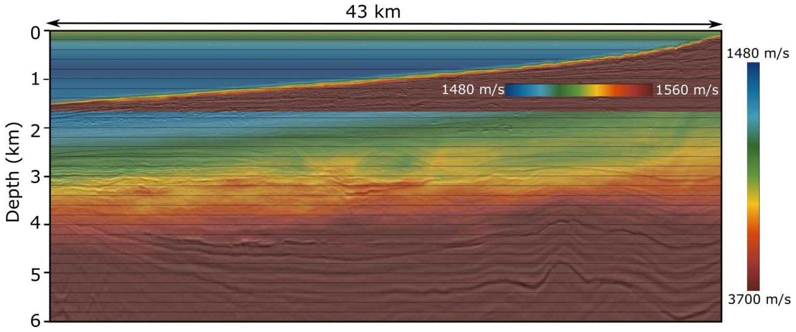
Table 6.2: Summary of the inversion parameters for the Gabon dataset inversion tests.



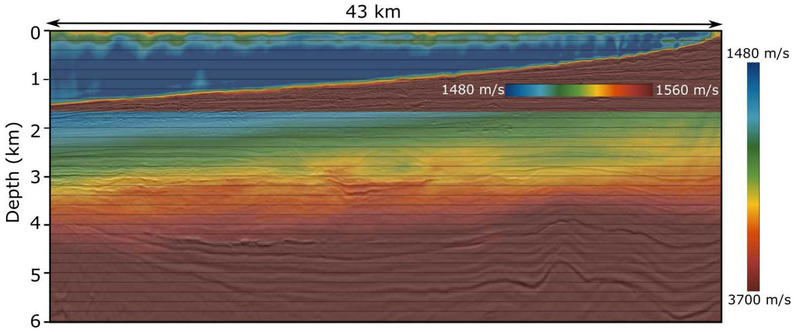
(a) Test 1



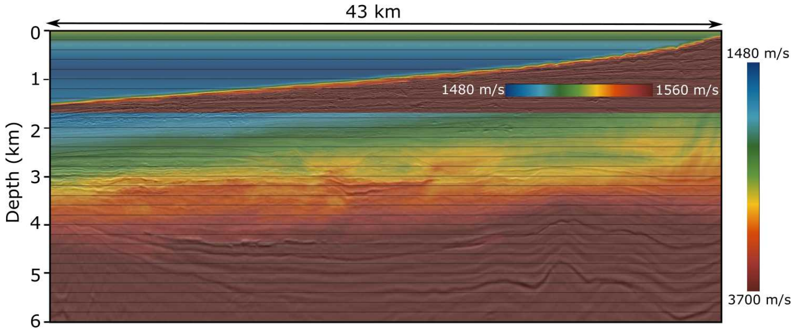
(b) Test 2



(c) Test 3

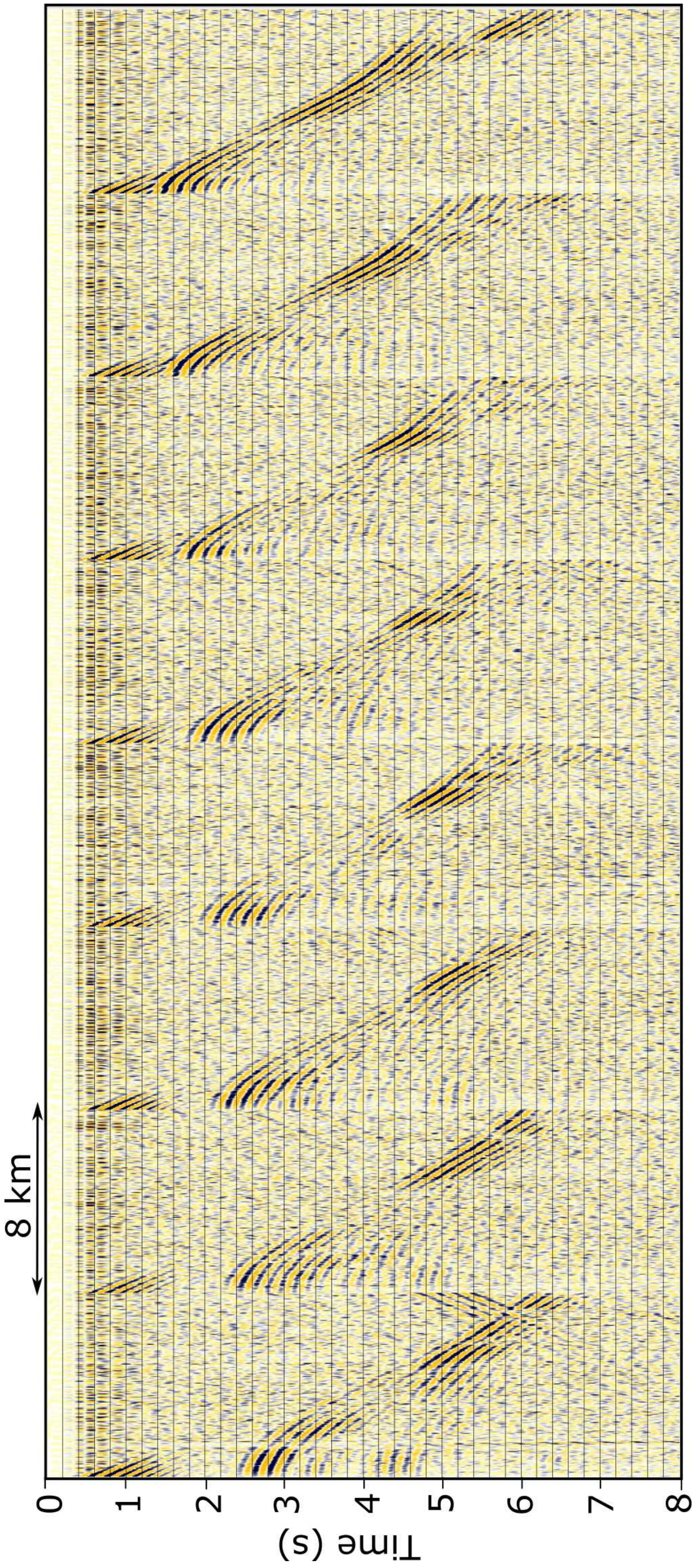


(d) Test 4



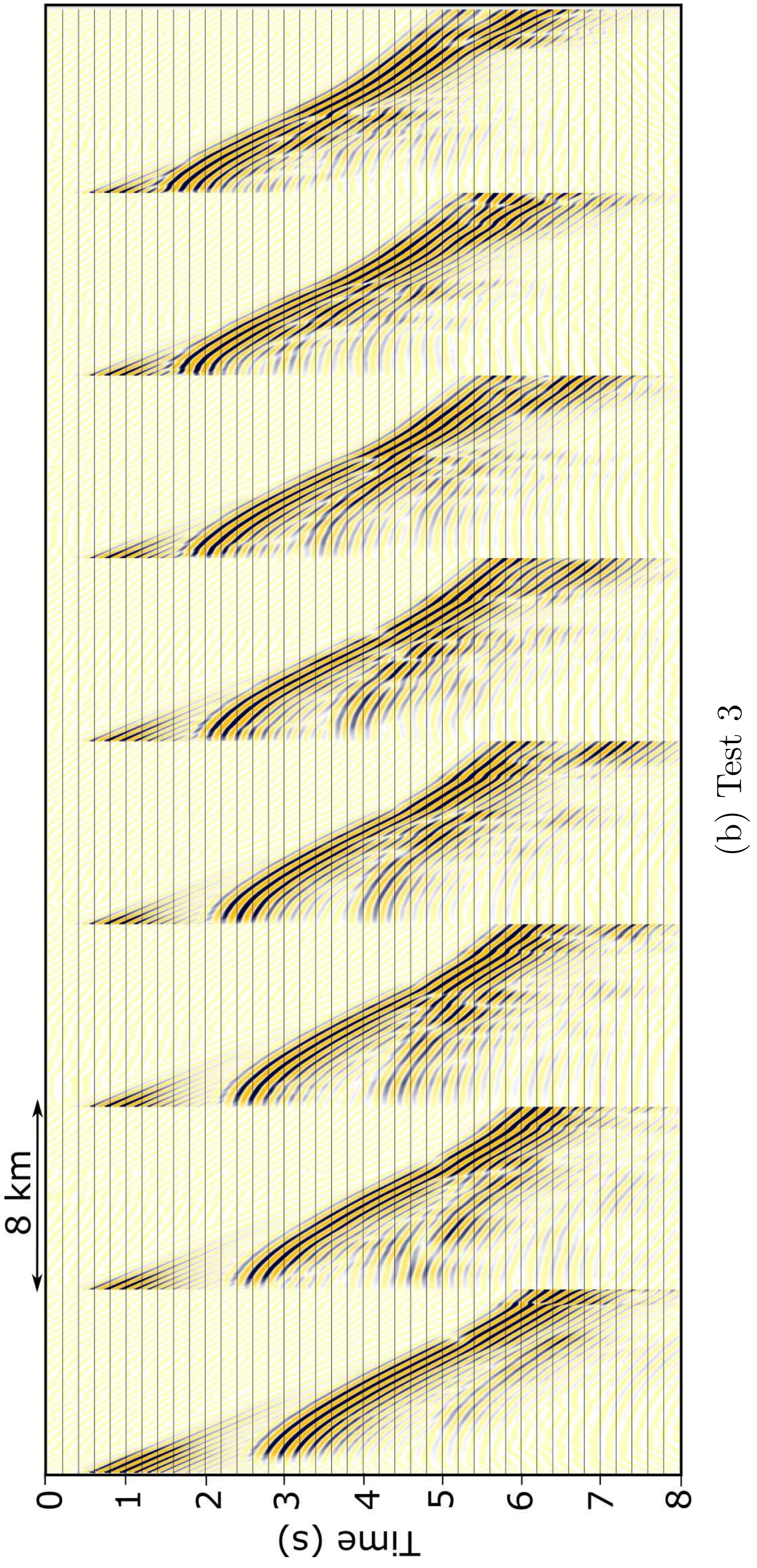
(e) Test 5

Figure 6.15: Comparison of FWI models obtained using starting models with different water velocity profiles and different density parameters. The parameters for each test are shown in Table 6.2.



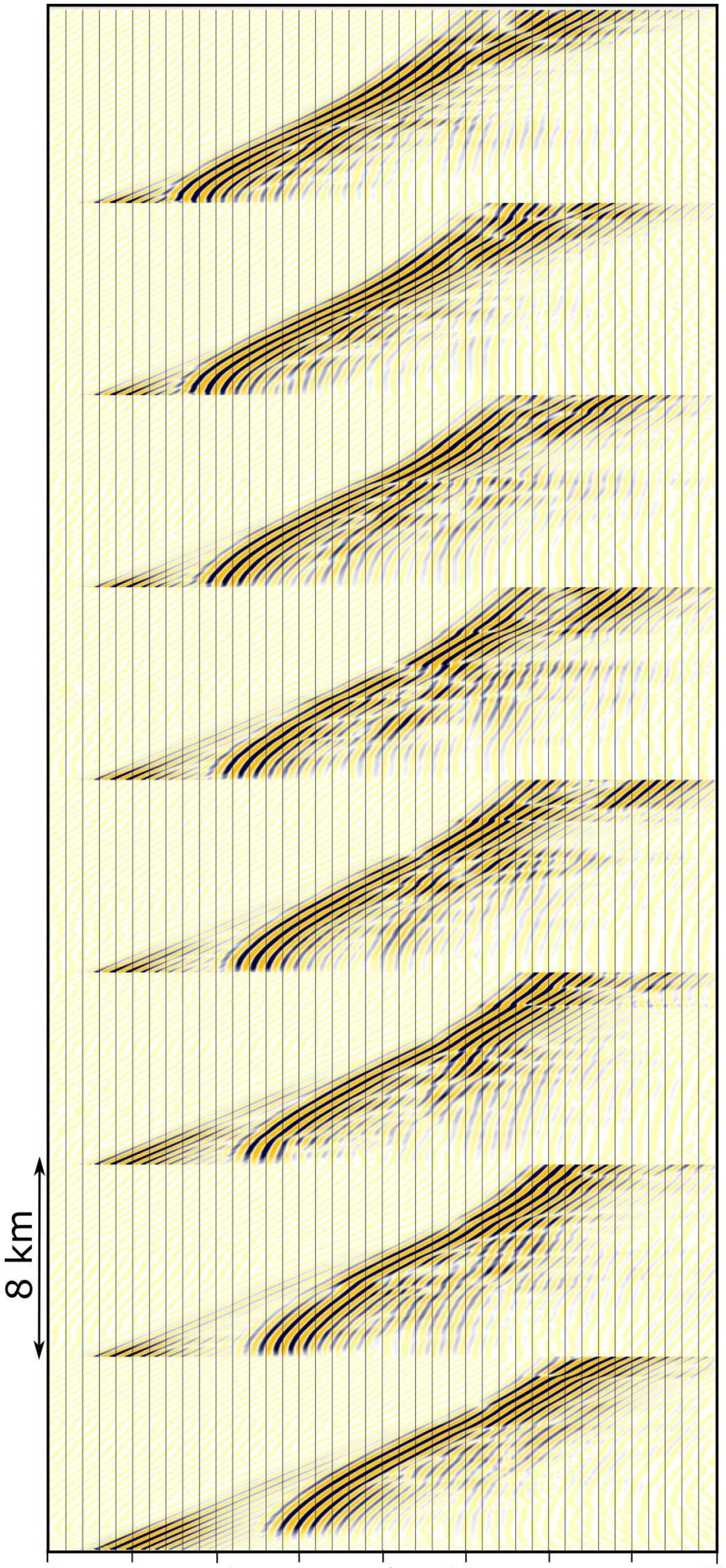
(a) Raw

Figure 6.16: Comparison of the shot gathers (a) raw and modelled with 6 Hz FWI velocity models obtained in (b) Test 3 (c) Test 4 and (d) Test 5 (cont.).



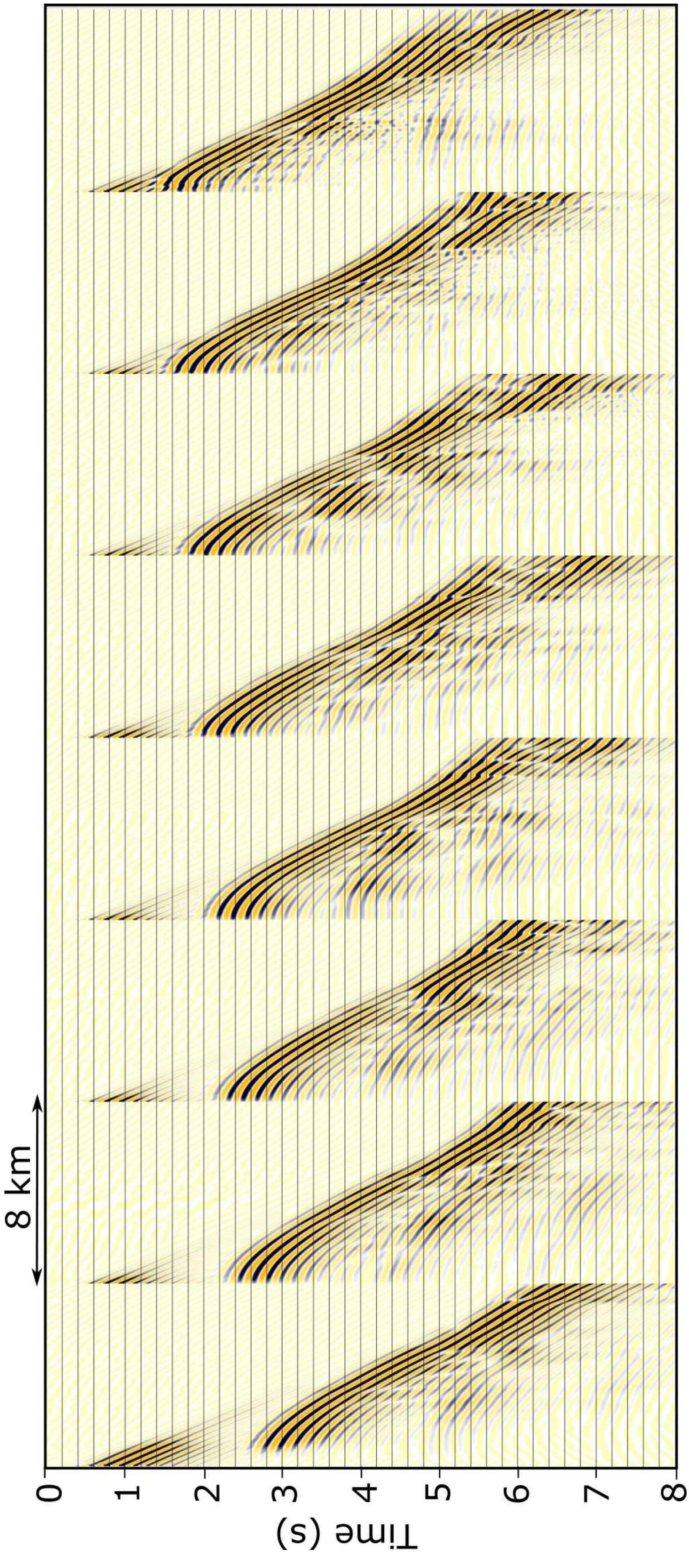
(b) Test 3

Figure 6.16: Comparison of the shot gathers (a) raw and modelled with 6 Hz FWI velocity models obtained in (b) Test 3 (c) Test 4 and (d) Test 5 (cont.).



(c) Test 4

Figure 6.16: Comparison of the shot gathers (a) raw and modelled with 6 Hz FWI velocity models obtained in (b) Test 3 (c) Test 4 and (d) Test 5 (cont.).



(d) Test 5

Figure 6.16: Comparison of the shot gathers (a) raw and modelled with 6 Hz FWI velocity models obtained in (b) Test 3 (c) Test 4 and (d) Test 5.

6.6.4 The data noisy?

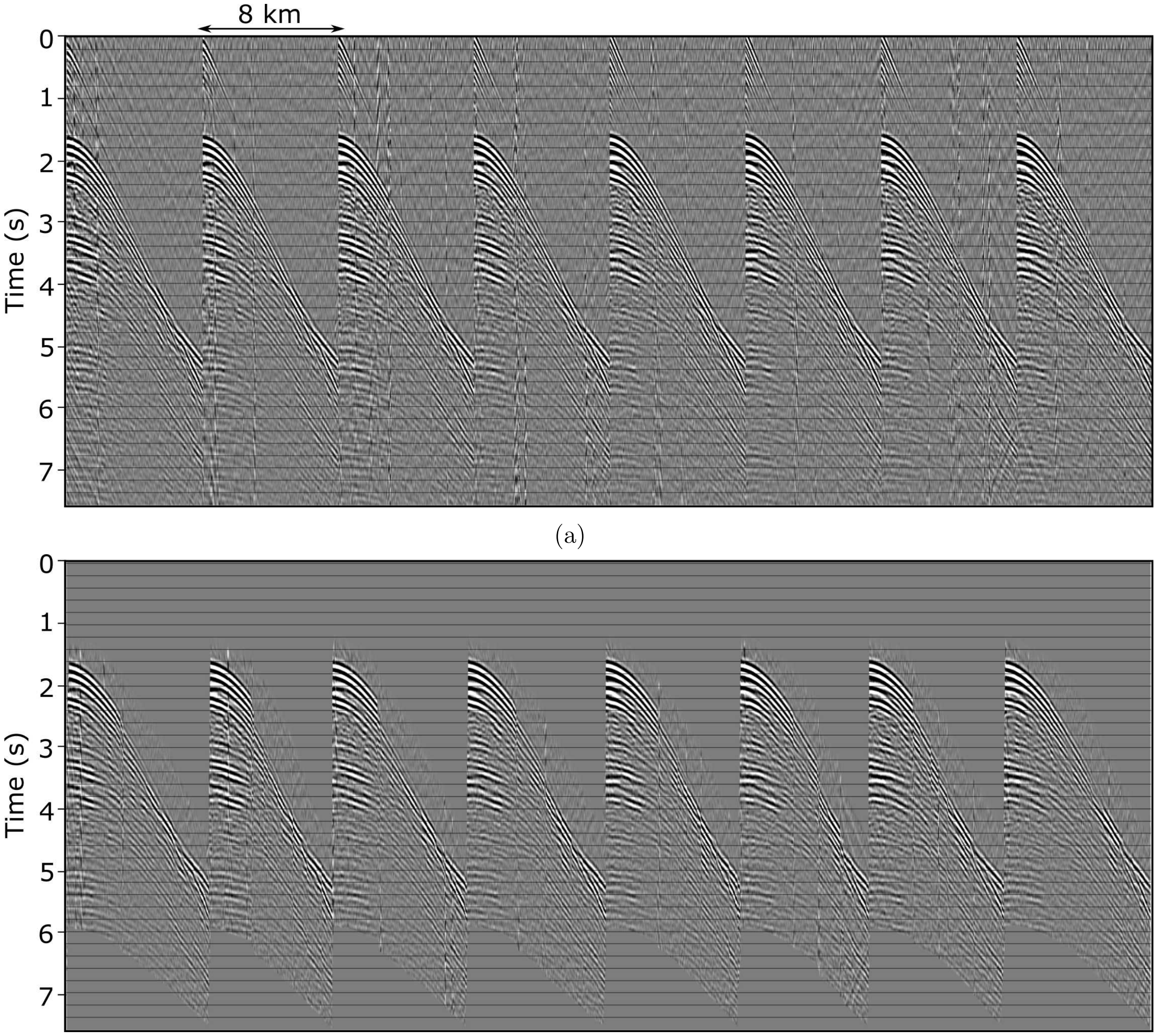
The relatively low amount of noise in the data is vital for successful FWI application. Theoretically, complete absence of noise in the acquired data would make it possible to predict the data that matches the field data perfectly given the true velocity model. Although it is an unrealistic scenario because there is always noise present, this does not mean that the proximity of the global minimum can not be reached. Provided there is a good enough signal-to-noise ratio, which in practice should be 1:1 or higher, and the starting velocity model is lying in the basin of attraction of the global minimum, FWI is capable to converge towards the true solution. This is commonly achieved by using regularisation of the misfit function and/or amplitude normalisation in the algorithm. However, if the signal-to-noise ratio is too low and/or there is a significant amount of "bad" traces, FWI can struggle to converge towards the true solution.

Here, I will explore the impact of noise on the inversion and investigate whether or not the noise is the reason for FWI stalling.

The field shots gathers, 3 to 12 Hz band-passed with a minimum phase Ormsby filter, are displayed in Figure 6.17(a). The data are relatively noisy, although the signal-to-noise ratio appears to be adequate for FWI away from the obvious large noise bursts, which is likely to be the swell noise. This swell noise is more obvious at low frequencies. The very short offsets on a few shots contain cable-tug noise below 4 s TWT. In addition, in all previous inversions no mutes were applied to the data, which means that FWI was trying to fit the noise ahead of the direct arrival. Although usually the datasets with such noise level behave sensibly during FWI and muting ahead the direct arrival is desirable but not necessary (if the data is relatively noise free), it still may be possible that the large noise bursts compromise the inversion.

For this test the data were de-noised. The de-noise consisted of excluding a small number of very noisy shots, removing the traces with large noise bursts, muting the data ahead of water bottom arrival and muting the bottom part of the data using an offset-variable mute. The direct arrival was not retained in the most part of the data away from the very shallow area, where it interferes with the water bottom arrival and is not easy to mute in time domain. In total, approximately 70% of the receiver points were retained for the inversion. No other procedures were applied to the data.

Note, that the previously derived source signature is perfectly adequate for the denoised dataset because neither amplitude nor phase content of the data has been modified in any way. In case of using more sophisticated noise-reduction techniques, for instance F-K filter to suppress the linear cable-tug noise, the new source signature must be obtained.

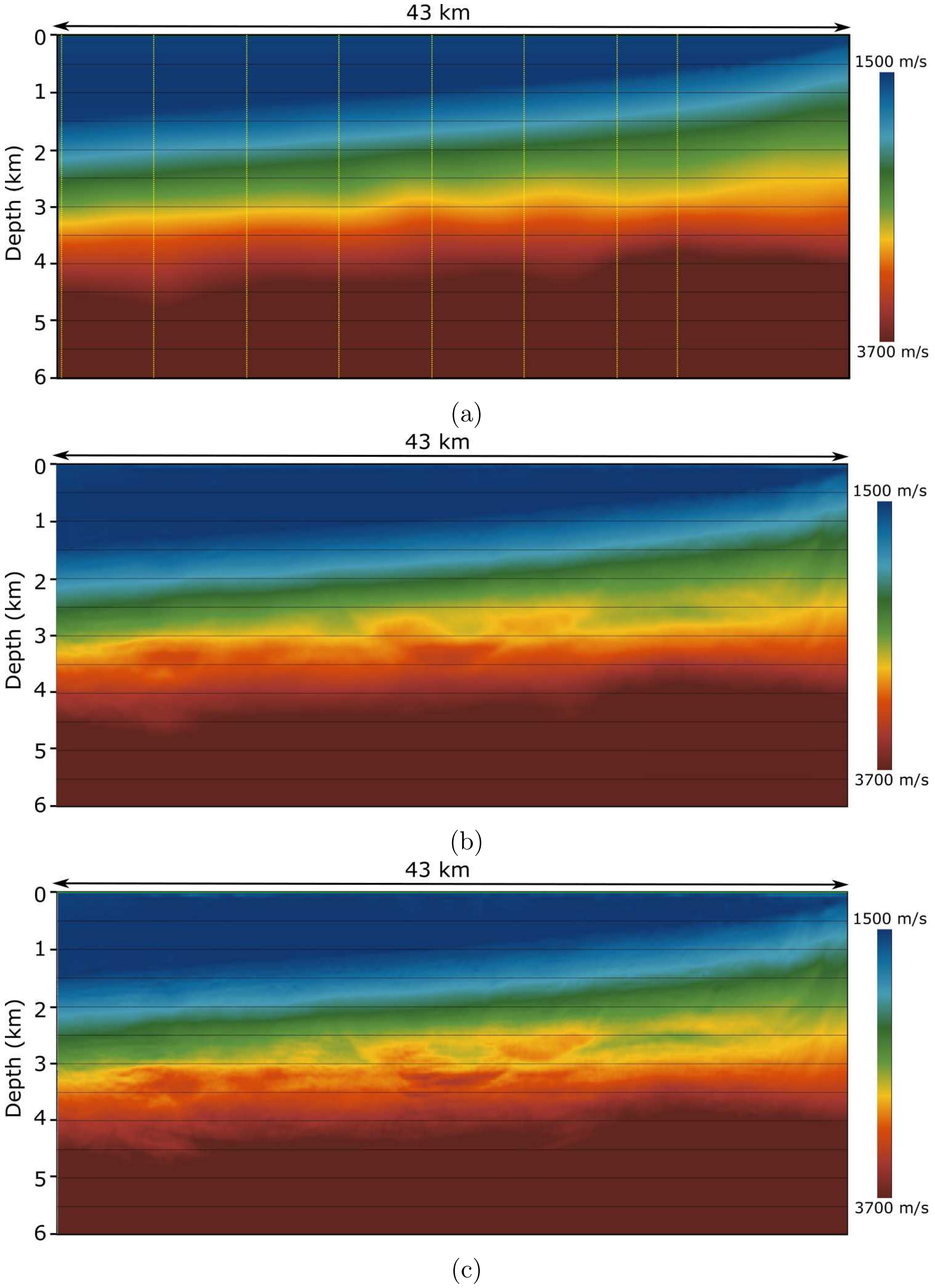


(a)

(b)

Figure 6.17: Shot gathers 3 to 12-Hz band-pass filtered (a) without and (b) with noise reduction. Noise reduction included removing traces with the swell noise, muting ahead of the water-bottom arrival and the bottom part of the data. Traces are individually scaled.

The data were inverted using the starting velocity model with depth-variable water velocity (for the description of the starting model see section 6.3.2). The starting velocity model and FWI models at 6 and 12 Hz are shown in Figure 6.18. The same models overlaid with the depth slice are displayed in Figure 6.6. It can be clearly seen that both models produce the same main features and these features are located predominantly in the middle part of the section. As expected, the 12 Hz model has an improved resolution compared to that of the 6 Hz model. However, even the 12 Hz model does not seem to generate velocity updates in the shallow part of the model.



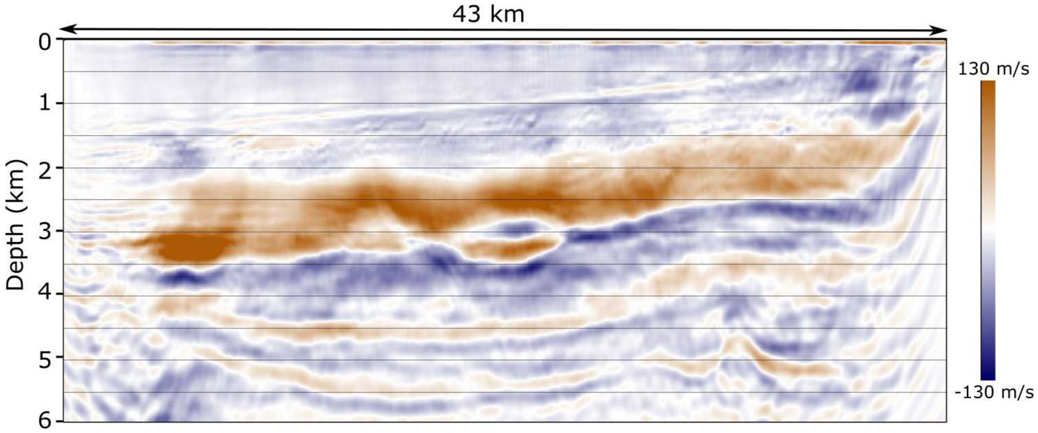
(a)

(b)

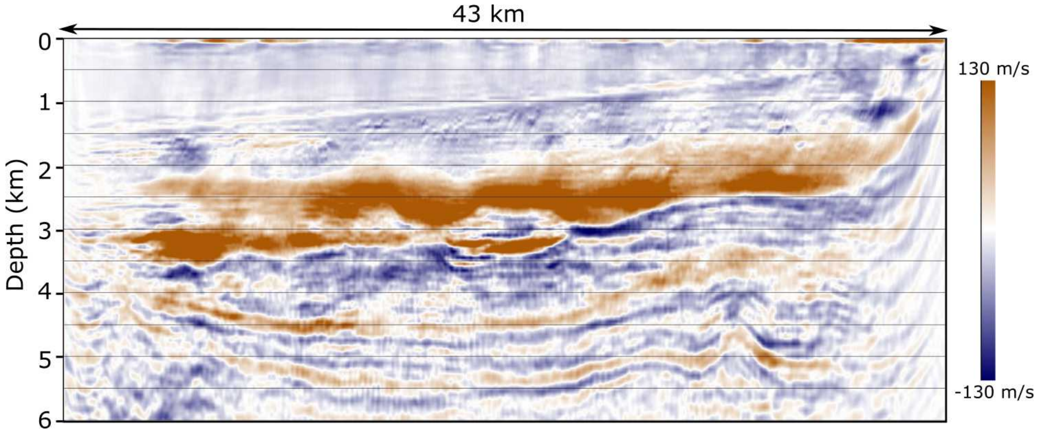
(c)

Figure 6.18: Comparison of (a) starting model and FWI models at (b) 6 Hz and (c) 12 Hz. The FWI was performed on the data with reduced noise.

In order to demonstrate that the upper part of the section seems to have no visibly noticeable velocity updates, Figure 6.19 shows the result of the subtraction of the starting velocities from the FWI models at 6 and 12 Hz. Both models produce updates below approximately 2 km. FWI seems to stall in the upper part of the section even after the noise in the data has been reduced.



(a)

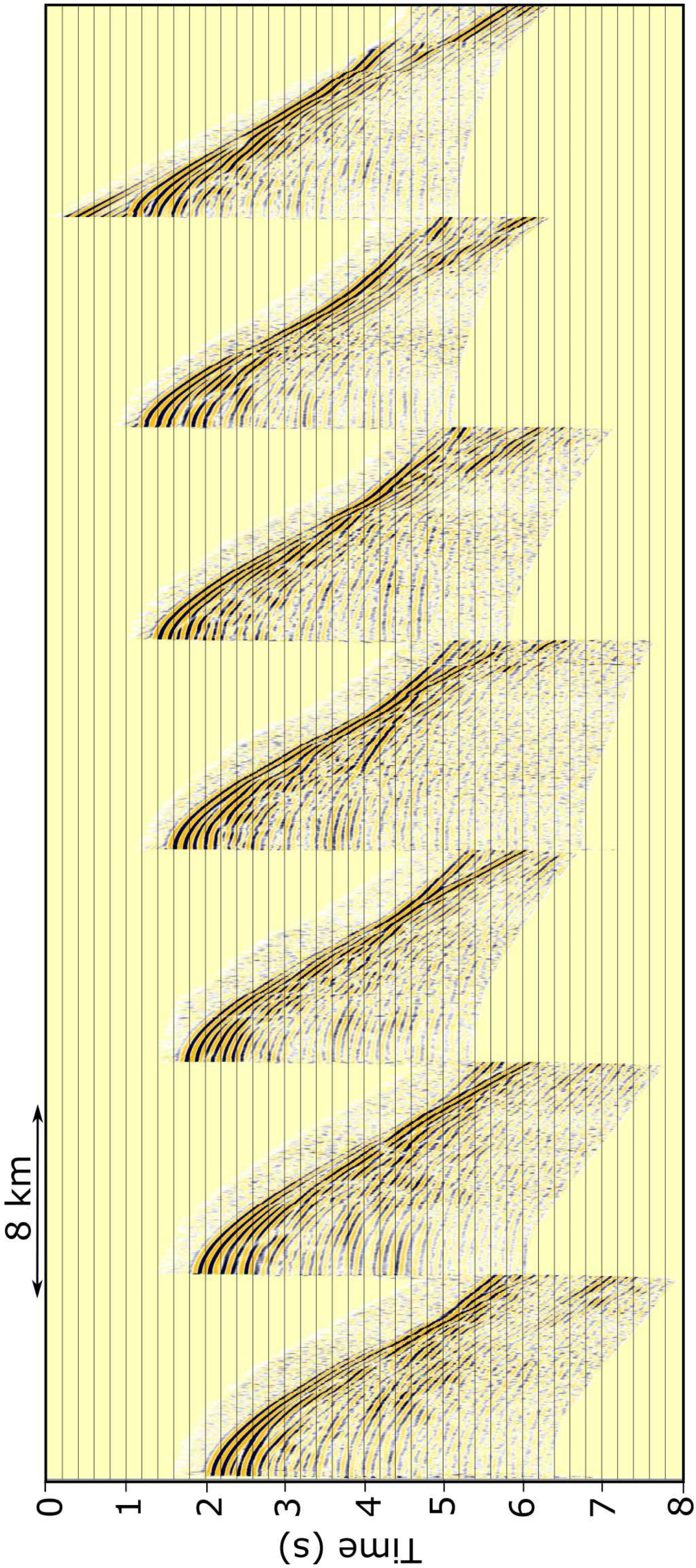


(b)

Figure 6.19: FWI velocity update at (a) 6 Hz and (b) 12 Hz. The FWI was performed on the data with reduced noise.

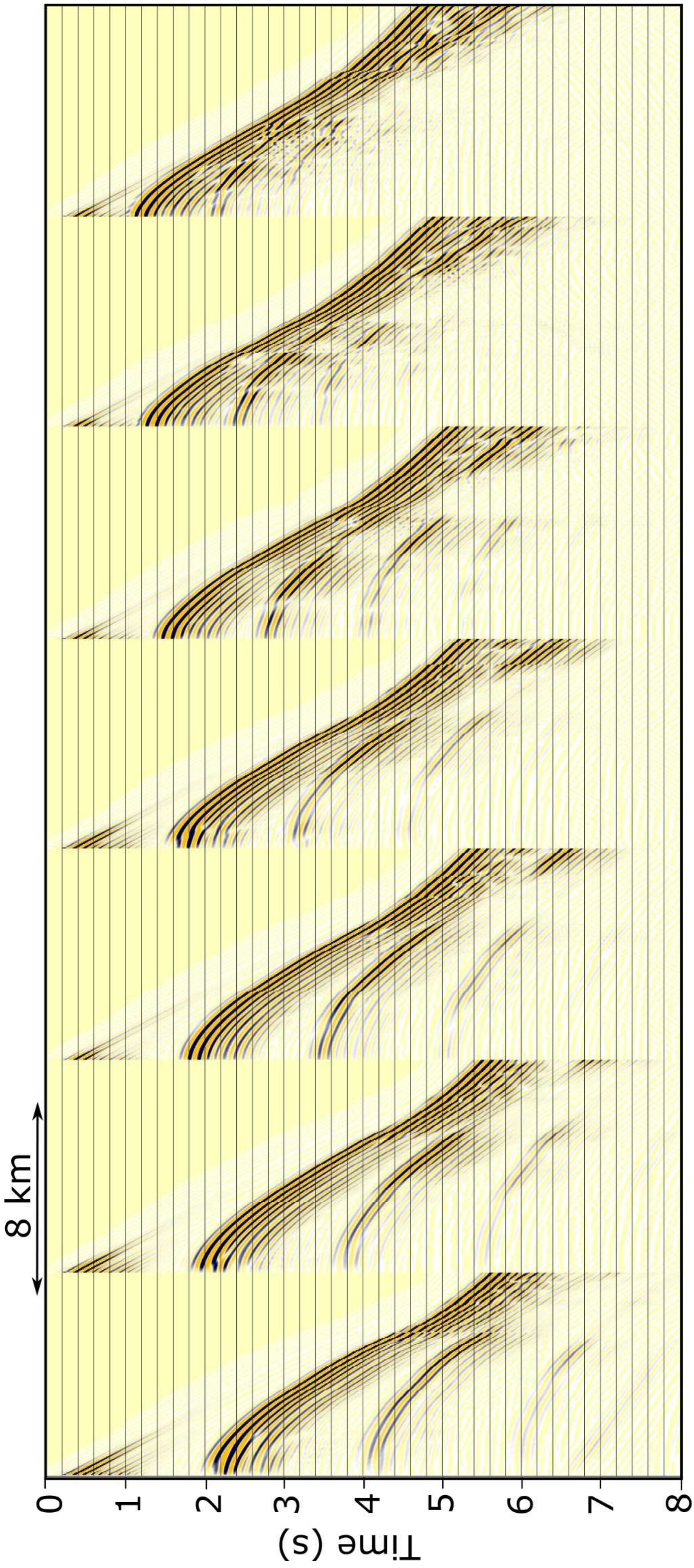
Figure 6.20(a) and (b) compares the raw shot gathers to the gathers modelled using the starting velocity model. The data contain frequencies with the maximum of 12 Hz. While the starting velocity model provides reasonable match at 6 Hz, at 12 Hz some of the shots appear to be cycle skipped. Note, that the raw data have very weak or no distinct multiples whilst the modelled data contain strong first-order multiples. As the inversion progresses, the match between the shot gathers and the field data improves. Figure 6.20(c) shows the raw shot gathers obtained using the 12 Hz FWI model. It can be noticed, that 12 Hz FWI model significantly improved the gathers. The water bottom arrival is no longer cycle skipped at frequencies as high as 12 Hz apart the very last gather in the shallow water region. Most of the displayed gathers demonstrate an adequate match on the long offsets as well. The inversion has suppressed the strong multiple generated by the starting velocity model. The fit in the deeper part of the records seem to improve overall.

As the conclusion, FWI seems to benefit from the noise reduction. However, the noise is not the reason for FWI to stall in the upper part of the section.



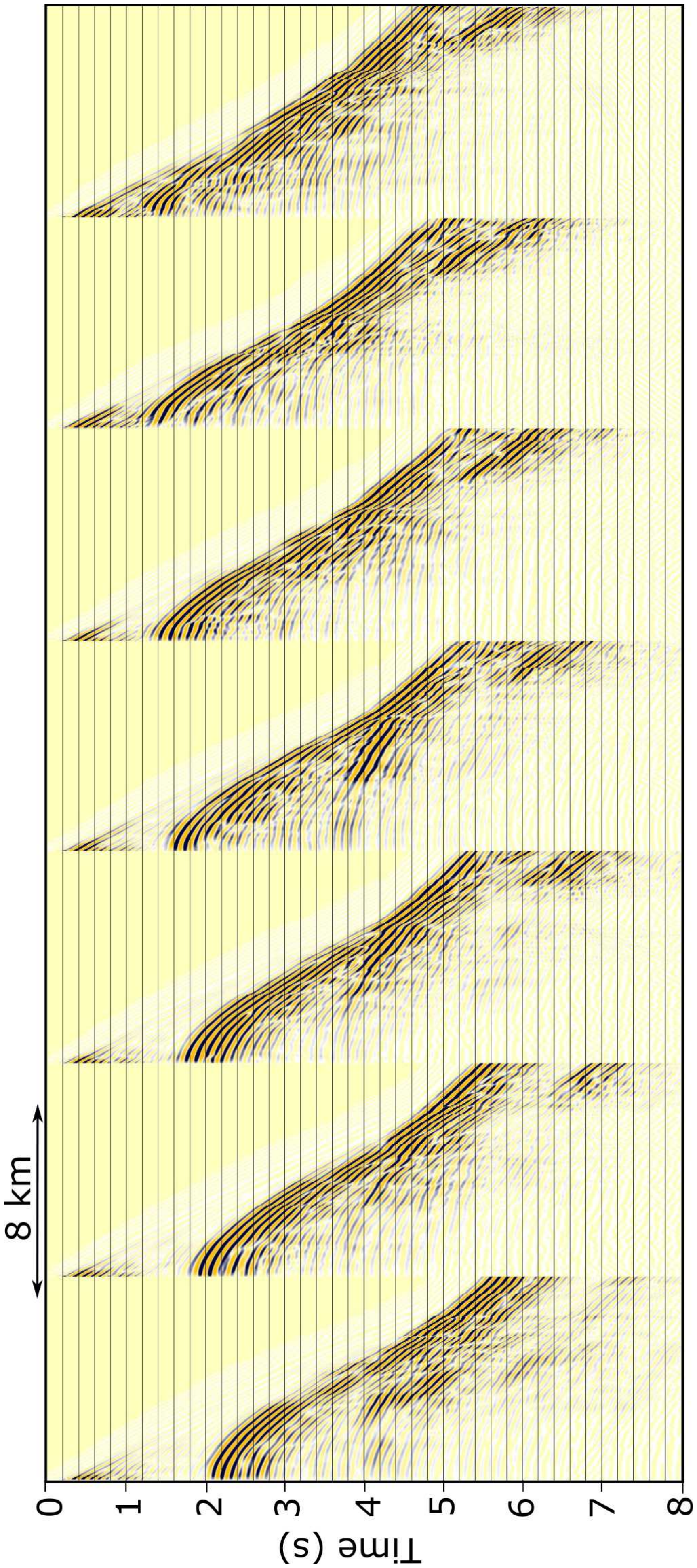
(a) Raw

Figure 6.20: Comparison of the shot gathers (a) raw and modelled obtained using: (b) the starting velocity model and (c) the 12 Hz FWI (cont.).



(b) Starting model

Figure 6.20: Comparison of the shot gathers (a) raw and modelled obtained using: (b) the starting velocity model and (c) the 12 Hz FWI (cont.).



(c) 12 Hz FWI

Figure 6.20: Comparison of the shot gathers (a) raw and modelled obtained using: (b) the starting velocity model and (c) the 12 Hz FWI.

6.6.5 RTM - the seabed correct?

Although the synthetic gathers generated using the starting velocity model seemed not to be cycle skipped at the water bottom reflection, most of the previously presented inversions have attempted to shift the seabed position down in depth. Unlike conventional migration, FWI requires not only correct travel times but also the true depth position of the reflectors. The travel times to the seabed were validated independently by Kirchhoff depth migration and the migrated gathers have demonstrated a flat seabed reflector at all offsets (Fig. 6.14(b)). A possible way to QC the depth position of the seabed, assuming that the water velocity is correct, can be Reverse Time Migration (RTM).

RTM is one of the available migration algorithms, that is in essence a two-way wave equation depth migration that translates shot gathers into an image of reflection coefficients in depth. Unlike other migration algorithms, it can image steep depths and handle complex velocity distributions. However, it is computationally more expensive than other migration algorithms.

RTM is based on a principle established by Claerbout (1971): “reflectors exist at points in the ground where the first arrival of the down-going wave is time coincident with an up-going wave”. Conventional RTM computes the forward wavefield, back propagates the residuals in time, cross-correlates the two wavefields and stacks the result of cross-correlation at every time step for every shot gather.

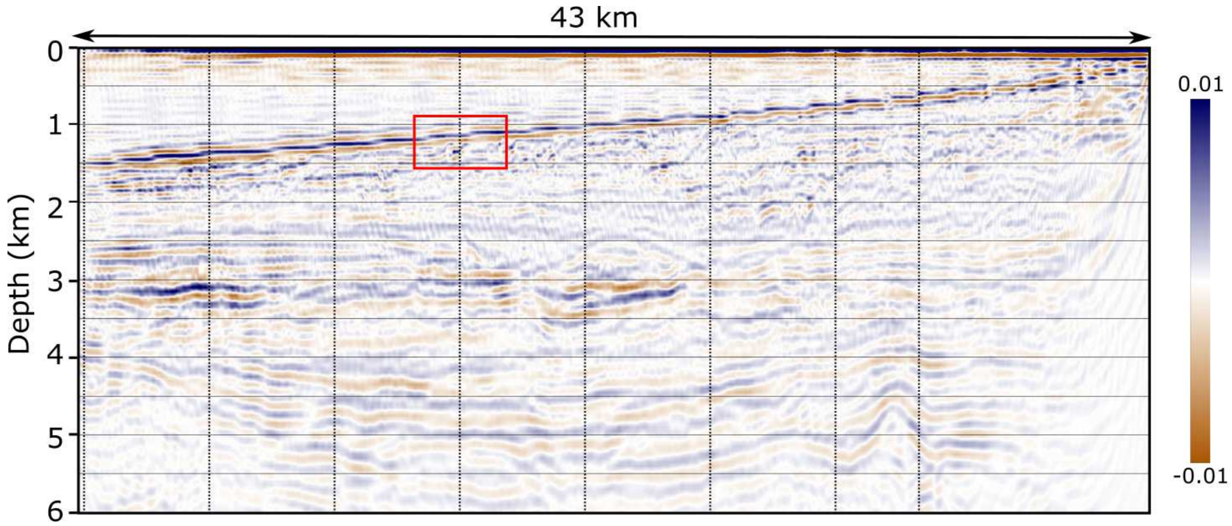
$Fullwave3D$ code performs iterative least-squares non-linear RTM, which is different to conventional RTM. It uses reflection data, therefore utilises the short offsets only. It generates the FWI gradient the same way as described in chapter 2.1.3, filters it using a Laplacian filter to eliminate the long wavelengths and then uses the resulting gradient to update the model. In this case, the RTM image is not strictly speaking a reflection coefficient, but the scaled difference between the starting model V_{start} and the updated model $V_{updated}$:

$$Image = \frac{V_{start} - V_{updated}}{2V_{start}} \tag{6.3}$$

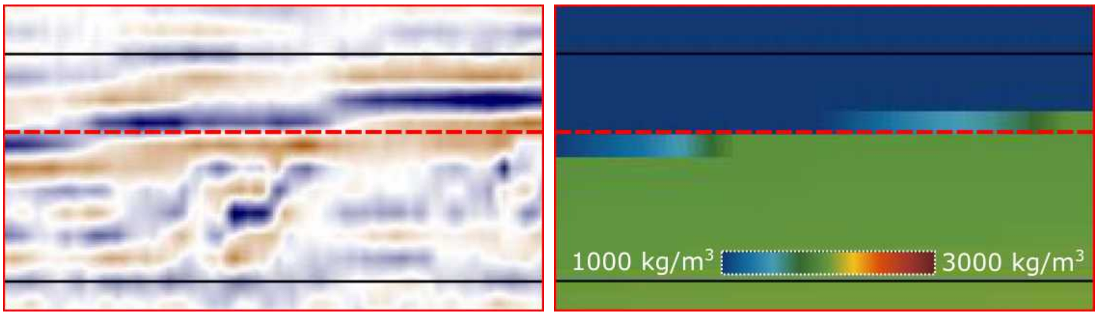
This image appears to look similar to a reflection coefficient image.

Unlike conventional RTM where a reflector is indicated by a zero-phase wavelet, $fullwave3D$ uses a wavelet with the phase of $\pi/2$. This means, that in conventional RTM images reflectors are associated with “peaks” and “troughs” while in $fullwave3D$ RTM reflectors are associated with zero-crossings.

Here, RTM was performed using the starting velocity model with the depth-variable velocity profile. The RTM image produced after 10 iterations is shown in Figure 6.21(a).

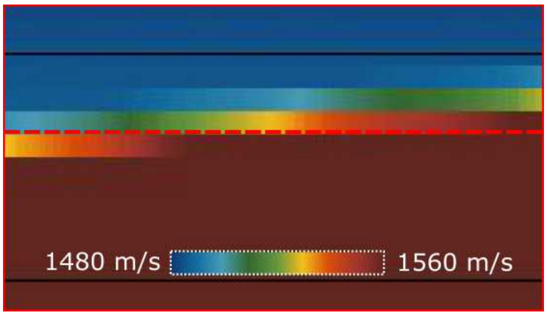


(a)



(b) RTM

(c) Density



(d) Velocity

Figure 6.21: Comparison of the RTM image to the starting models. (a) The RTM image and close-ups of (b) the RTM image, (c) the starting density model and (d) the starting velocity model. The position of the close-up is indicated in the red rectangle in (a). The water bottom is indicated by a dashed red line and its location is consistent through all the images.

It can be seen, that this RTM image picked up the main reflectors as indicated in the depth section (Fig. 6.11(a)) with the water bottom corresponding to a zero-crossing. To verify the position of the water bottom in the starting FWI models, the RTM image, the density model and the velocity model are displayed in Figure 6.21(b),(c) and (d). It is clear that both the velocity and the density model computed from it are consistent with the RTM. Thus, it is very likely that the water bottom reflector is correctly placed in the starting model and it is not the reason causing unusual FWI behaviour.

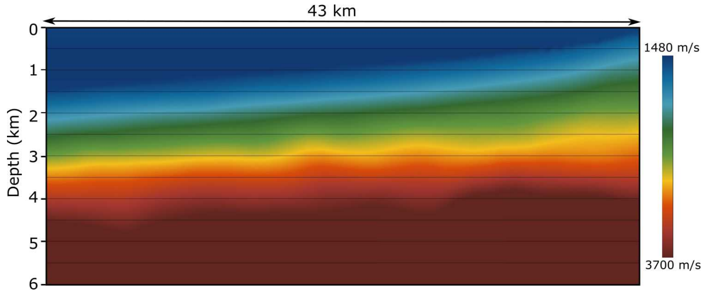
6.6.6 Different code versions

All the previous inversions were performed using the high kernel version of the code. In this section, I have tried to run the inversion on the two versions (both high and low kernels) in order to see how sensitive the results are to the propagator.

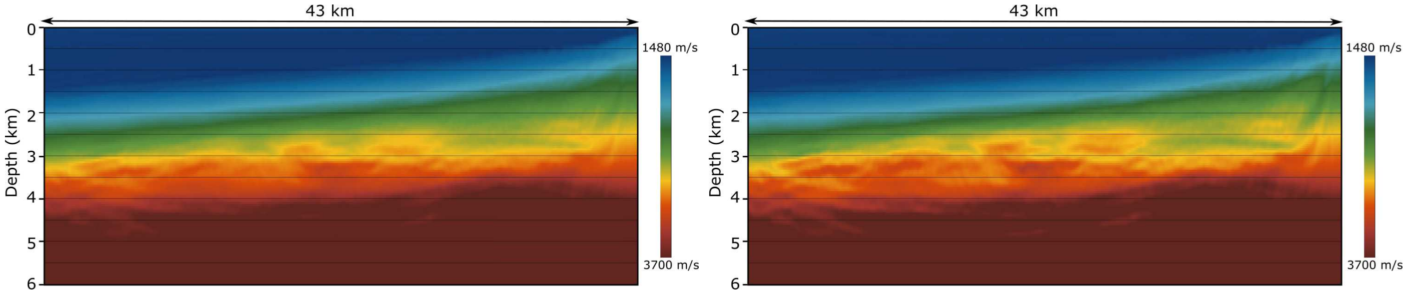
In principle, there should be no difference between the two FWI models produced by the low and the high kernel versions provided they both used the same data, the same source, the same starting models and exactly the same inversion strategy. These two models should be extremely similar, however not identical due to the numerical noise.

Figure 6.22 shows the starting velocity model and compares the FWI models obtained during the first four iterations using the low and the high kernel versions of the code. It can be seen, that the first iteration FWI models are qualitatively very similar to each other. Perhaps, the only difference is the amplitudes of the velocity updates - the low kernel FWI model seems to have approximately twice the amplitude compared to those of the high kernel. The high kernel demonstrates very small updates after iteration two, three and four. It also seems that the first-iteration FWI model, produced by the high kernel, is more resolved than the following models as if the inversion “smoothes” the initial updates.

The second iteration of the low kernel, in contrast, produces a model that is clearly cycle skipped. The inversion on the low kernel goes completely wrong and produces very different results to that of the high kernel. This is a very unexpected result and in order to analyse it, it might be useful to compare the misfit functions and step factors for these two runs.

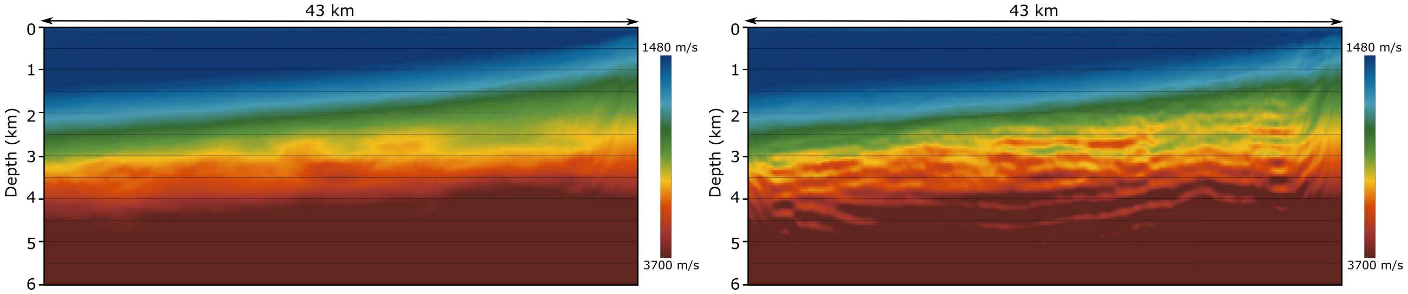


(a) Starting velocity model



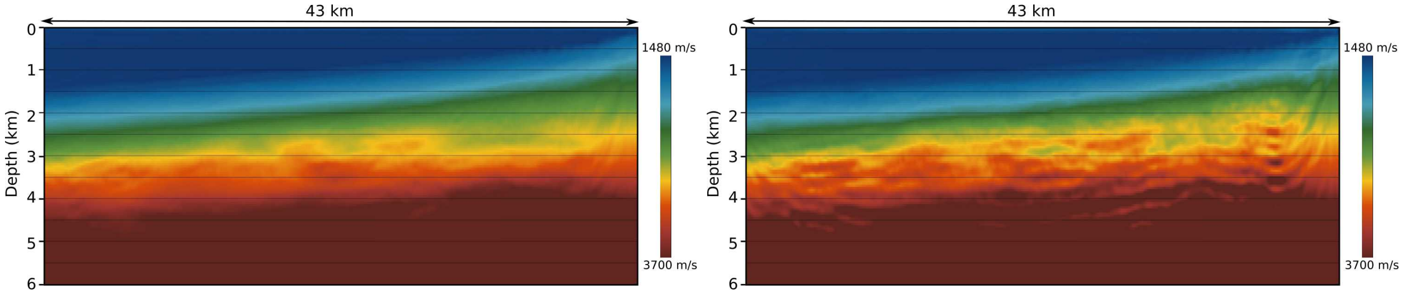
(b) High kernel - iteration 1

(c) Low kernel - iteration 1



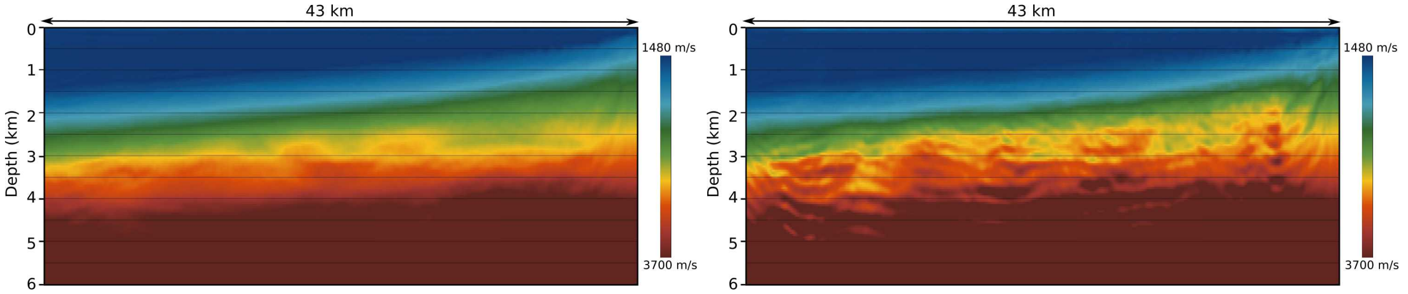
(d) High kernel - iteration 2

(e) Low kernel - iteration 2



(f) High kernel - iteration 3

(g) Low kernel - iteration 3

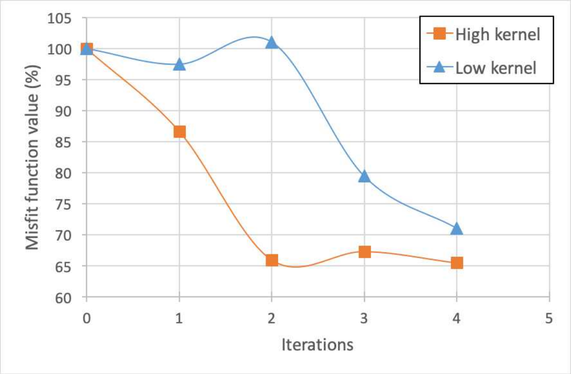


(h) High kernel - iteration 4

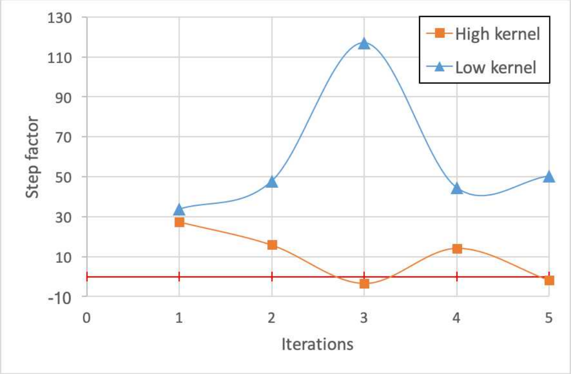
(i) Low kernel - iteration 4

Figure 6.22: Comparison of the FWI models obtained using the low and the high kernel. (a) The starting velocity model. FWI models obtained after 1^{st} iteration using (b) high and (c) low kernels; after 2^{nd} iteration using (d) high and (e) low kernels; after 3^{rd} iteration using (f) high and (g) low kernels; after 4^{th} iteration using (h) high and (i) low kernels.

Figure 6.23 compares the misfit values and the step factors for the first 5 iterations performed on both the low and high kernels. Note, that the misfit function starts with iteration "0" while the step factor starts with iteration "1". This is because the first misfit value is the value representing the difference between the field data and the modelled data using the starting velocity model, in other words, this is the value prior to the first iteration. The step factor, in turn, starts with "1" and shows what step factor was used to update the model that generated the misfit with the first-iteration FWI model.



(a)



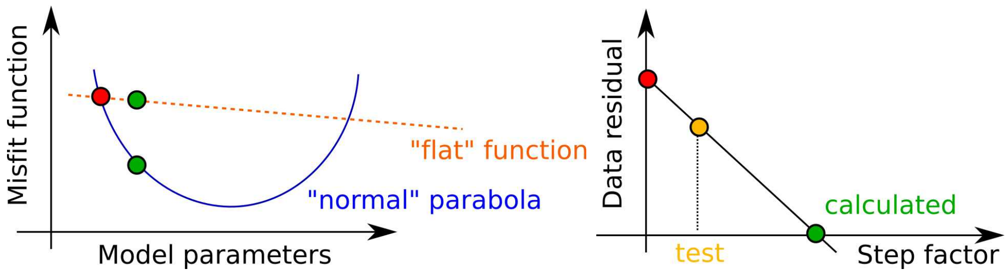
(b)

Figure 6.23: Comparison of (a) the misfit function and (b) the step factor between the high and the low kernel versions of the code. The zero step factor value is indicated by the red line. Corresponding velocity models are shown in Figure 6.22.

It can be seen, that the step factors for both inversion runs are significantly different, especially for iteration "3". Both inversions start by taking a large step. The high kernel gradually decreases the step factor towards a negative value at iteration "3" and then, uses relatively small values, possibly oscillating around zero. The low kernel increases the step factor value to a very large number at iteration "3" and, overall, uses the larger step factors than the high kernel. Because the first step factors are very similar, the first-iteration FWI models look similar to each other. After the first iteration, the step factors are significantly different and therefore so are the second-iteration FWI models.

The misfit functionals for both runs for the first five iterations are displayed in Figure 6.23(a). It can be seen, that both are decreasing, which indicates that both kernels produce FWI models that mathematically provide a better fit to the field data than the starting model. The high kernel decreases the misfit value by 35% after iteration "2" and then does not significantly change it. The low kernel drops the misfit value at the last two iterations. Both kernels produce a very similar misfit value at iteration "4". However, the models look extremely different and the one obtained by the low kernel will clearly generate strongly cycle-skipped data.

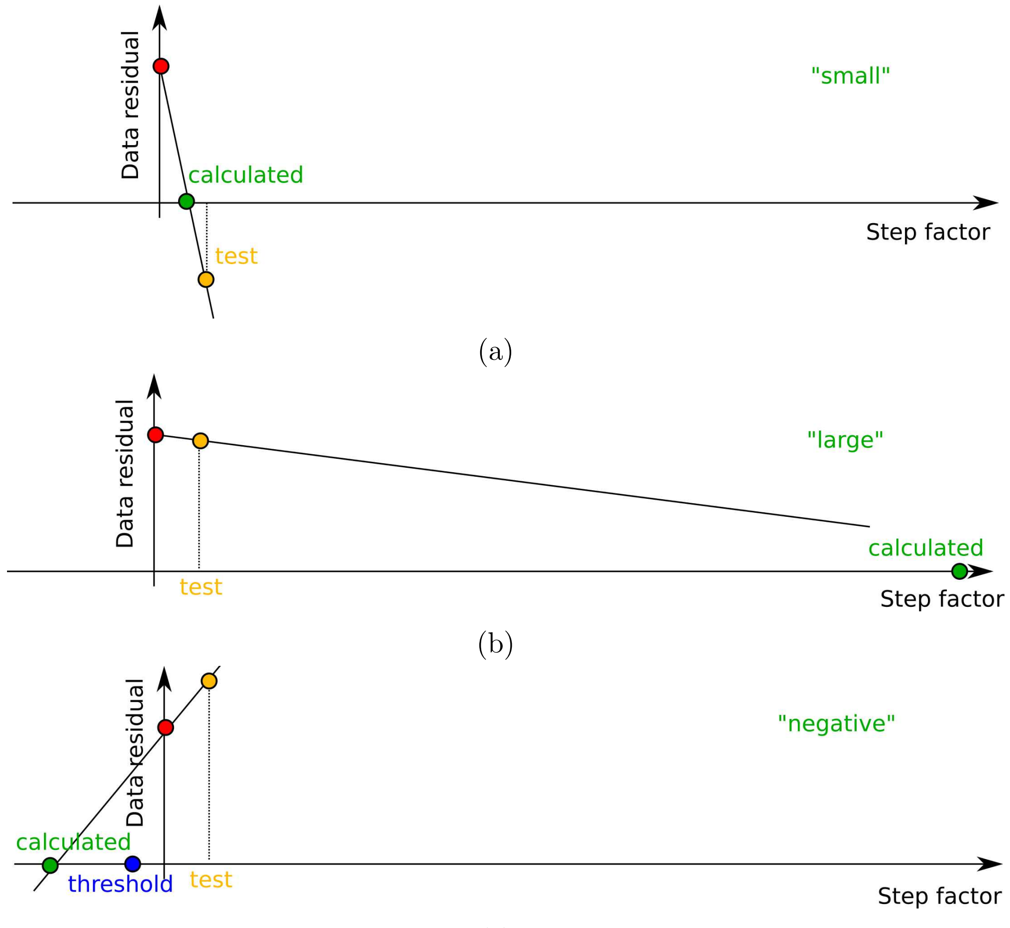
In order to analyse these results, I would like to illustrate how the step factor is calculated. To summarise the detailed description provided in chapter 2.1.3, the step factor is calculated using the linear search algorithm, where the inversion takes a small "test" step in the opposite direction to the gradient, calculates the difference between the field and the modelled data using the new model (I will refer to this difference as the data residuals) and then, assuming that a linear change in the model leads to a linear change in the data residuals, the algorithm searches for the "calculated" step factor that minimises the data residuals (Fig. 6.24(b)). Assuming that the gradient is correct, a relatively large step factor is beneficial as it makes the inversion converge towards the global minima faster and therefore minimises the cost (Fig. 6.24(a)). However, if the misfit function is presented by a flat function instead of a traditional parabolic function, the large step factor will result in minimal changes in the data misfit (Fig. 6.24(a)).



(a) (b)

Figure 6.24: Schematic diagram showing (a) the misfit function and (b) the step length calculation.

Now, I would like to compare the cases when the calculated step factor has a small (less than 1), large or even a negative value. Figure 6.25 illustrates these three cases, where the same "test" step factor is taken. The calculated step factor is small if the "test" step factor produces a model that generates negative data residuals and therefore, the step factor will be smaller than the "test" step factor. The calculated step factor is large if the data residual is very small after updating the model using the "test" step factor. This means that the gradient does not significantly affect the data residuals. Finally, the calculated step factor is negative if the data residual calculated using the model obtained with the "test" step factor is larger than the data residual calculated using the starting velocity model, in other words, the data residual gets worse after taken a small step in the opposite to the gradient direction. In $fullwave3D$ there is a certain threshold on the negative step factor values, which clips a smaller negative step factor to the smallest allowed step factor.



(a)

(b)

(c)

Figure 6.25: Illustration of the step length calculation in case the step lengths is (a) small, (b) large and (c) negative.

Note, that by the data residual I mean a single value, representing the difference between the field and the observed data at a single time sample on a single trace. For the whole dataset, there are millions of such data residuals and consequently the "calculated" step factor is obtained as an average of all the calculated step factors.

Now, looking back at the step factor values for the high and the low kernel runs (Fig. 6.23(b)), it can be concluded that the first large step factor value probably indicates that the initial gradient causes very small changes in the residuals on average and therefore, both kernels take a large step in the opposite direction to the gradient. It is possible, that this small change in the residuals is comparable to the noise amplitude and therefore, the smallest differences in the misfit between the field and the calculated data result in the very different time steps calculated by the two kernels.

In summary, the large step factors and the difference in behaviour of the two kernels is anomalous and significant. It means that the gradient calculated by FWI has only minimal effect on the residuals on average, though changing the model using the gradient very likely produces large changes, both good and bad, in some of the residuals. The large resultant step lengths are spurious, and the resultant update is not useful. Another way to express this is is to recognise that the objective function represents a multi-dimensional error surface in model space; that surface is extremely flat in the region of the starting model, so that the slightest noise (whether in the data or in the numerics) dramatically changes the gradient direction and/or the step length leading to spurious outcomes. This analysis though cannot reveal why the error surface is so flat for this dataset.

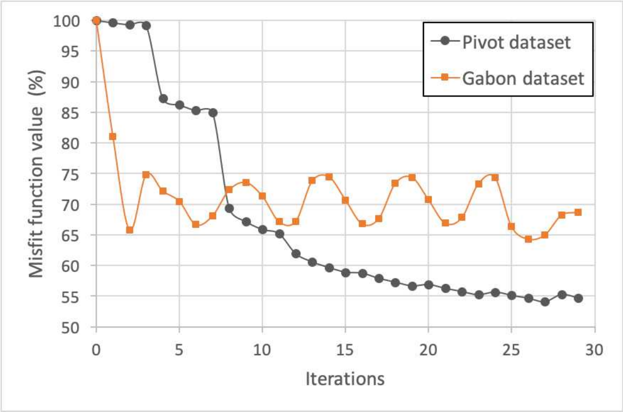
6.7 Conclusions

As it has been shown in this chapter, FWI on the Gabon dataset seems to generate most of the updates during the very first iteration and then it stalls, which results in the final model being similar to the first-iteration model. It also does not update the shallow areas of the model, which is extremely unusual for FWI. All the common reasons that cause FWI problems were carefully examined and it was concluded that it is very unlikely they are the cause of such untypical FWI behaviour.

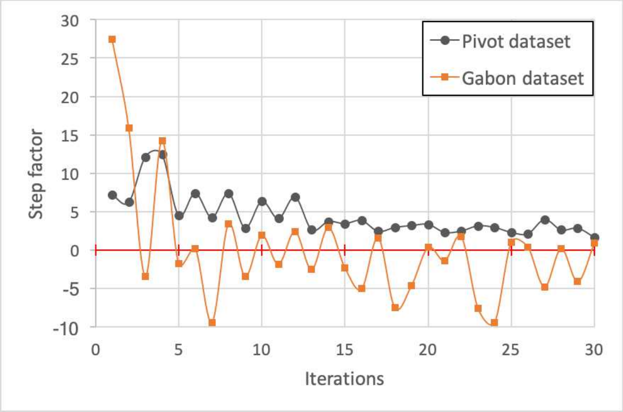
I have compared one of the FWI runs to a successful FWI run on the Pivot dataset, discussed in this thesis. Figure 6.26 demonstrates the misfit functions and the step factors for the Gabon and the Pivot datasets inversions. The inversion of the Pivot dataset was performed using more than 30 iterations, however, only the first 30 iterations are shown here. The Gabon dataset uses 5 iterations per frequency block and the Pivot dataset uses 4 iterations per frequency block.

It can be seen, that for the Pivot dataset the first 30 iterations reduce the misfit function value by 45% while for the Gabon dataset the misfit function reduces by approximately 30%. It is interesting, that the first two iterations reduce the misfit function to 35% and all the iterations after that do not improve the model. The misfit function appears to oscillate around the same value and FWI stalls.

The step factors for both inversions are displayed in Figure 6.26(b).



(a)



(b)

Figure 6.26: Comparison of (a) the misfit functions and (b) the step factors for the Gabon and the Pivot datasets. The zero step factor value is indicated by the red line.

It can be seen, that the Pivot dataset demonstrates classic behaviour - the step length is initially relatively large and then it gradually decreases with the inversion converging towards the minimum. None of the step factor values are below zero, which indicates that every iteration improves the velocity model. The Gabon dataset, however, demonstrates different behaviour. The inversion starts with a large step factor and then the step factor oscillates around zero, with plenty of negative step factor values. This indicates that FWI goes forward and backwards, not changing the velocity model very much as the number of iterations increase.

So, how can the behaviour of this dataset be explained? None of the common explanations for the failure of FWI appear to apply here - the dataset appears suitable for FWI, the start model seems sensible, low frequencies are present, there are clear shallow refractions in the data, the model does not contain shallow salt, basalts or other significant anomalies, rocks at seabed are slow so that elastic effects are unlikely to be significant at the seabed and there are no signs of elastic effects in the data. Sensible changes to water column and the detailed water bottom improve the data fit somewhat, but do not really improve the FWI results. The reflection data show various shallow features, at least some of which must have associated small-scale velocity anomalies. But nothing works - neither for me or for two commercial processing contractors.

Consequently, I have concluded that there must have been some unknown and undeclared prior adjustment made to the data, and/or that the acquisition parameters described in the data headers do not match those that were used in the field. The consequence of either of these situations will be that the raw field data are no longer capable of being described by the wave equation - that is, there is no model that can explain the data accurately because the data are not in fact as they are described.

In geophysics, we are always very reluctant to say “blame the data”, but in this case that would appear to be by far the most likely explanation. So what could be wrong? Nothing is obvious, and conventional reflection processing generates sensible looking migrated sections. It is most likely that something data adaptive has been applied to the data in the field, perhaps directly by the acquisition system, but there is no possible way to find that out without access to completely raw data hard drives and an on-board acquisition report. Even this may well not solve the problem since it is quite likely that the acquisition company were unaware of whatever it is that has been done to this data. Potentially it is an instrumental malfunction or mis-setting, or potentially it is a deliberate action to mitigate some feature of the data that would otherwise cause it to be outside the acquisition specification. The data are noisier than would often be acceptable to a client, and it is possible that something illicit has been done to bring the data back marginally into specification, though this of course can not be proven with the information available to me.

Chapter 7

Variations of FWI

I have completed a 3 month internship with ION Geophysical, during which I have been applying different types of 3D time-domain acoustic FWI to a NATS dataset acquired offshore Brazil.

The code I was working with was not $fullwave3D$ described earlier and used throughout this thesis. This code has been developed at ION. It uses adjoint state method to calculate the gradient, spatially preconditions the gradient using a Gaussian smoother, approximates the Hessian with its diagonal, utilises line search for calculating the step length and normalises amplitudes shot-by-shot. The code automatically stops to perform iterations when the objective function does not change for more than an internally set up limit; this is the programmed specifications of the code and everywhere in this chapter I will refer to this as "convergence". Some others characteristics remain unknown to me due to confidentially.

7.1 Rational

The reflection-dominated NATS dataset had an unusual acquisition configuration and was situated in a complex geological environment that included salt, carbonates and complex post-salt deposits (see Chapter 3). Conventional FWI has previously been applied to this dataset and did not seem to significantly improve a velocity model generated by tomography, although FWI worked well on a neighbouring dataset in the same geological area. The main objective of my internship was to explore whether other FWI types could solve the problem. Consequently, a number of different types of acoustic time-domain anisotropic FWI has been applied to the same 3D NATS dataset, being focused on post-salt deposits, and the results will be presented in this chapter.

7.2 Pre-processing

The raw data consists of the near offset shots with offsets below 4 km and the far offset shots with offsets ranging from 4 to 8 km due to the acquisition type, i.e. there are two separate shots per one full-offset record.

The data were pre-processed by ION in two steps and the two different pre-processed volumes were used for this study:

- The first volume was used for refraction FWI and had minimal conventional pre-processing applied to it. The processing flow contained linear de-noise (several passes of a Radon filter) to eliminate cable-tug noise and swell noise, de-bubble operator to remove the bubble footprint and a minimum-phase operator. Then the data were re-sampled to 16ms, low-pass filtered with a zero-phase Ormsby bandpass filter of bandwidth 0-2-8-12Hz and muted ahead of the water bottom arrivals. No Q-compensation or spherical divergence was applied. Fourteen pre-processed shot gathers evenly distributed across the model are shown in Figure 7.1. The data are reflection dominated with almost no refraction energy observed, it has strong salt reflections and weak water bottom multiples.
- The second volume was used for reflection FWI and had some extra procedures applied in addition to the linear de-noise and de-bubble operators: the data were de-multipled using 3D SRME (Surface-related multiple elimination), de-ghosted, zero-phased (with a positive amplitude (peak) for an increase in acoustic impedance as the achieved polarity) and formed into super-shot gathers using 4-D regularisation.

4D-regularisation is one of the methods used to form super shots in order to reduce the amount of shots and therefore the inversion cost as well as to improve signal-to-noise ratio, which is especially important for lower frequencies. In this case eight sources (four near offsets, four far offsets) that form four full-offset shot gathers were replaced with one effective source and the receivers were output on the original locations (Figure 7.3 and Figure 7.4). Pre-processed super-gathers low-passed using a zero-phase Ormsby filter of bandwidth 0-2-8-12 Hz are shown in Figure 7.2.

The raw data and full processing report were unavailable to me.

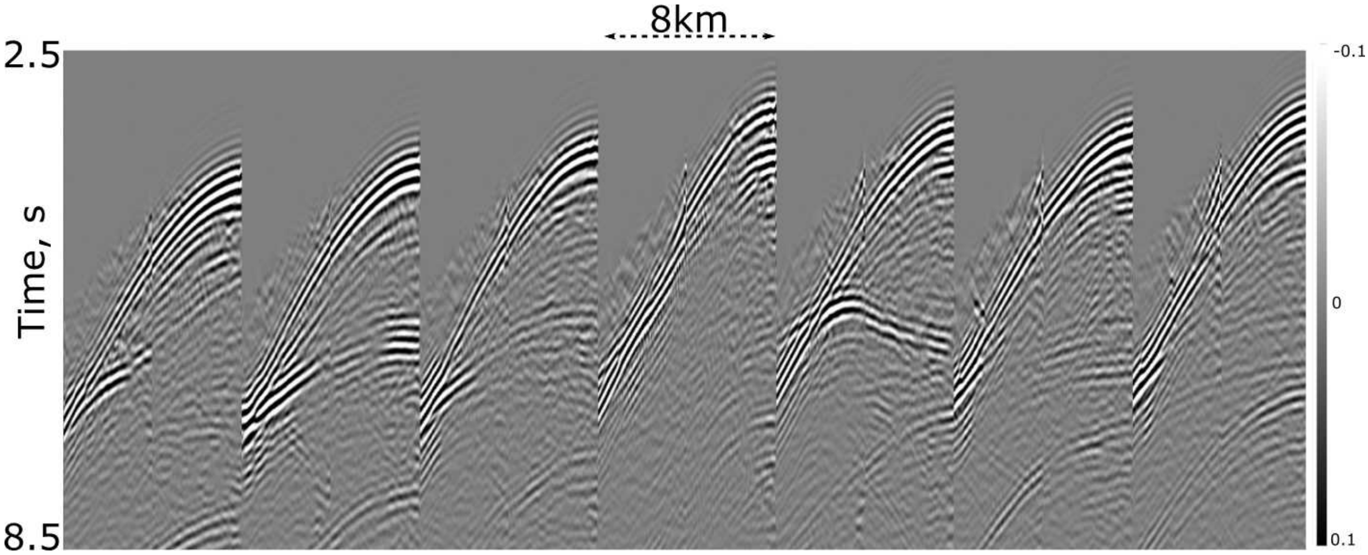


Figure 7.1: Pre-processed shot gathers for refraction FWI. Linearly de-noised, de-bubbled, minimum phased and low passed using a zero-phase Ormsby filter of bandwith 0-2-8-12Hz. No gain applied.

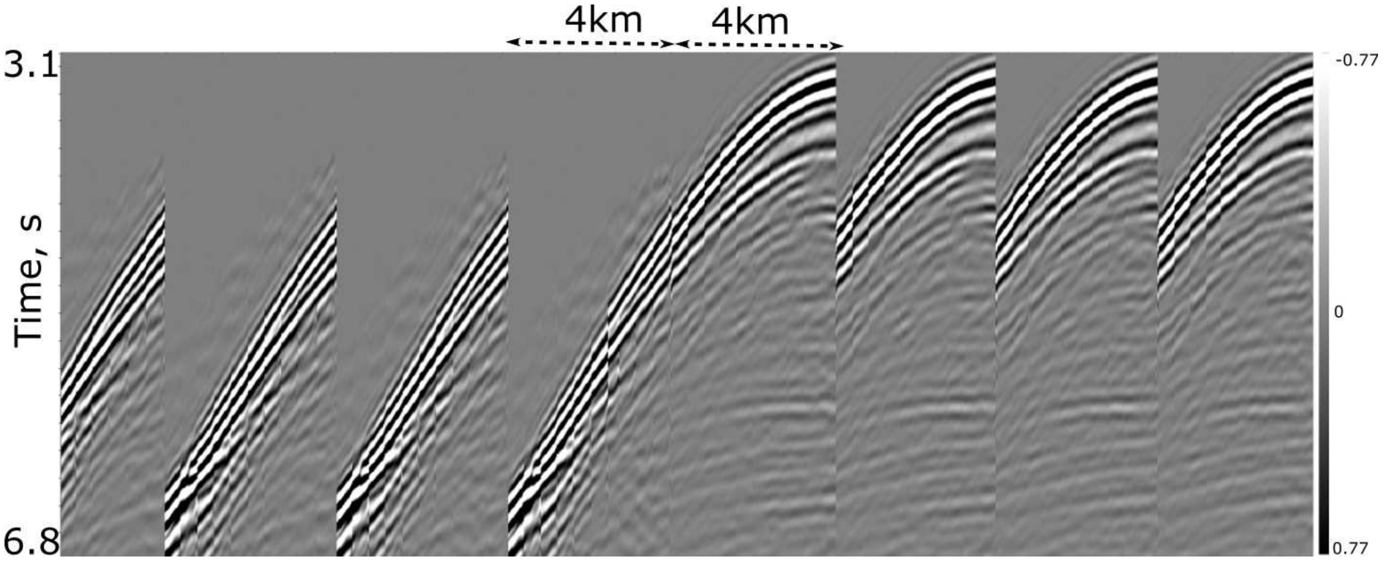


Figure 7.2: Pre-processed shot gathers for reflection FWI. Linearly de-noised, de-bubbled, zero-phased, de-ghosted, de-multipled, 4D-regularised and low-passed using a zero-phase Ormsby filter of bandwidth 0-2-8-12 Hz. No gain applied.

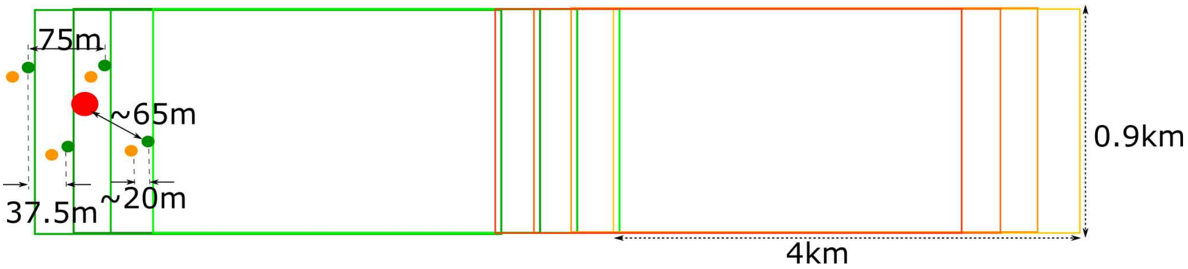


Figure 7.3: Schematic illustration of super shot concept. Not to scale. Near vessel sources are coloured in green, far vessel sources are in orange. Green and orange rectangles are the streamers boundaries for the sources accordingly. The eight sources are replaced using 4D-regularisation with the effective source indicated in red. The receivers are output on the original locations (the green and orange rectangles). The resulting super shot has one source and 8 sets of receivers.

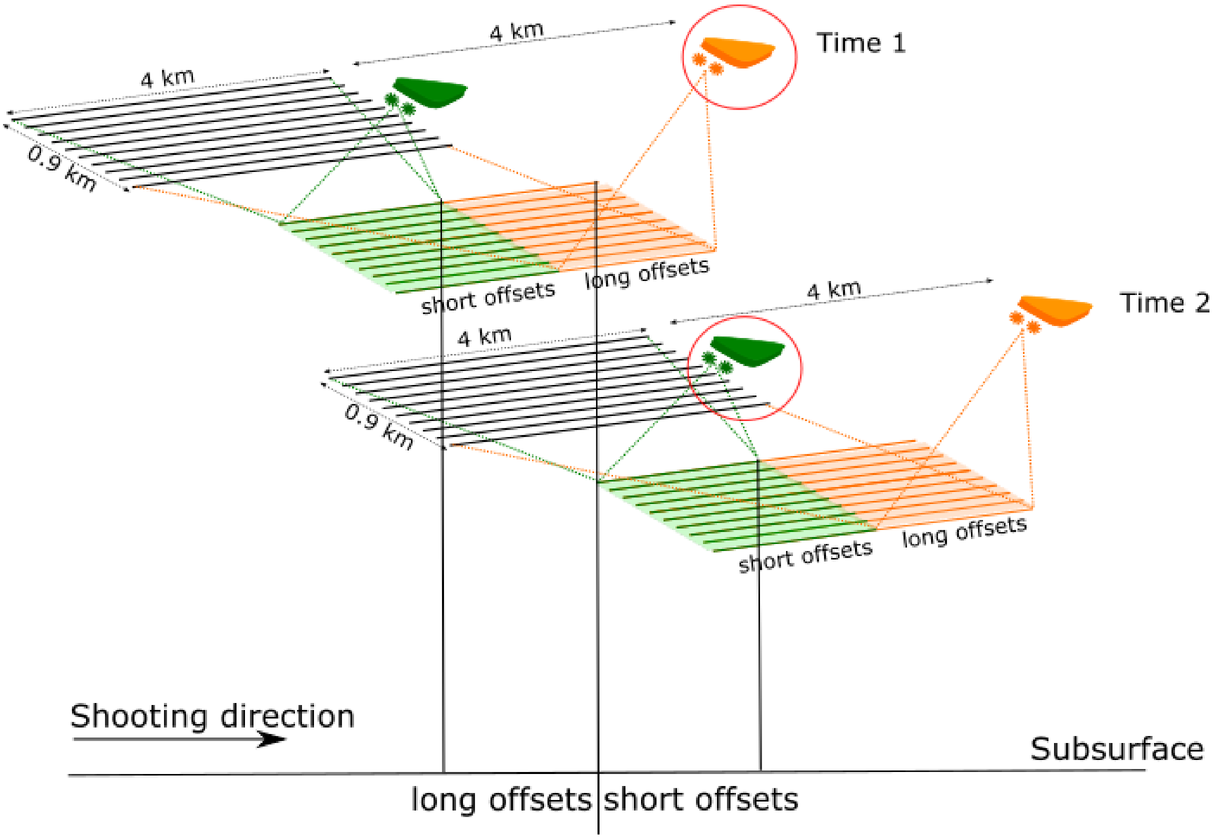


Figure 7.4: Acquisition configuration for the Picanha dataset. Not to scale. Near vessel sources are coloured in green, far vessel sources are in orange. Two different times are shown. The far vessel and the green vessel circled in red are a pair of vessels that from a “full offset” shot gather. Four pairs like that are shown in Figure. 7.3. Ideally, the far and the near vessels should be in the exact same position to form a “full offset” gather; however, in reality, the position is not exactly the same due to tides and GPS errors.

7.3 Source wavefield

It is vital for a source signature to have the same zero-time as the data, have exactly the same filters applied as to the data and have no ghost. The wavelet can be non-casual and have energy before zero time since both the source and the data are shifted down in time to keep that information.

The statistically extracted from the full volume of the data wavelet was provided to me. The wavelet was obtained by stacking the near-offset traces, after the water bottom reflection had been flattered, over the whole volume to eliminate the geological impact. However, this data volume was not the one provided for FWI and had different pre-processing applied to it. The data volume for FWI was de-bubbled and the provided wavelet was extracted from the data that had not been de-bubbled. It is very easy not to notice the difference in pre-processing between the data and the source wavelet and use the incorrect wavelet, which has had different pre-processing applied to it.

Here, I would like to present a test demonstrating the importance of using the correct source signature. Figure 7.5 shows the way two wavelets were obtained. The wavelet A was originally provided to me as a source wavelet. It was statistically extracted from the full data volume. This data volume, de-bubbled and significantly reduced in size, was given as an input data for FWI. The wavelet B was obtained from the source A by applying exactly the same de-bubble operator that was applied to the original data volume. Both sources are displayed in Figure 7.6, that clearly indicates presence of the bubble in the source A and absence of the bubble in the source B.

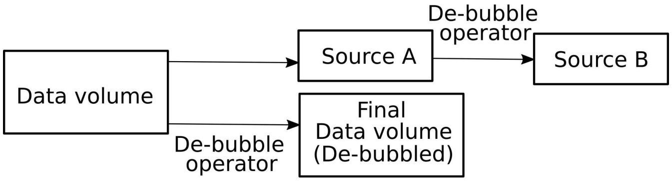


Figure 7.5: Diagram showing the source signatures estimation.

Both sources were propagated through the starting model and the two sets of modelled data were compared to a pre-processed shot gather from the volume available for FWI, that had the de-bubble operator applied to it (Fig. 7.7). It can be seen that the water bottom reflection is different between the two modelled shots (with a trough having maximum amplitude in the modelled shot using the wavelet that was not de-bubbled, and a peak having maximum amplitude in the modelled shot using the de-bubbled wavelet) as well as the shallow reflection under water bottom, indicated by yellow arrows, seems to be absent on the modelled using the incorrect source signature. The impact of an incorrect source is crucial as it makes it impossible to correctly match the synthetic data to the field data.

The source signature statistically derived from the initial pre-processed volume, de-bubbled with the same de-bubble operator applied to the provided data volume, re-sampled to 16 ms and low-passed using a zero-phased Ormsby bandpass of bandwidth of 0-2-8-12 Hz that was applied to the provided data volume was used for refraction FWI (Fig. 7.6(b)).

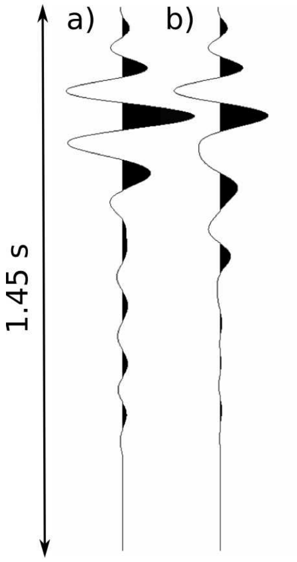


Figure 7.6: Source signatures: (a) not de-bubbled (b) de-bubbled.

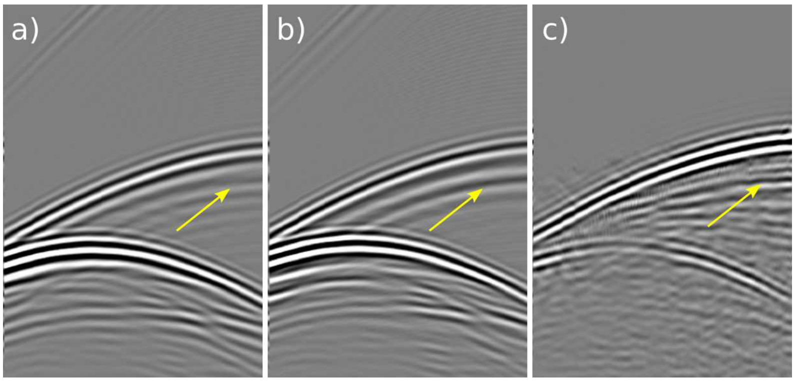


Figure 7.7: Comparison of the modelled shot gathers obtained with (a) de-bubbled source signature and (b) not de-bubbled source signature with (c) the processed data. The processed data were de-bubbled. Shots (a) and (b) are displayed using the same gain.

For reflection FWI, a zero-phased Ricker wavelet with the peak frequency of 7 Hz was used (Fig. 7.8). The Ricker wavelet was chosen as it is zero-phased as the provided pre-processed data for reflection FWI and requires no effort to derive the wavelet from the data. Besides, the available data volume was small and the direct arrivals were not present in the data. Consequently, the source wavefield needed to be obtained from the water bottom reflection, meaning it would be affected by the geology. The peak frequency of 7 Hz was chosen based on a series of different peak frequencies tested. Wavelets with peak frequency from 3 to 12 Hz were used to model the data and these modelled shot gathers were compared to the field data. The wavelet with the peak frequency of 7 Hz provided the best match.

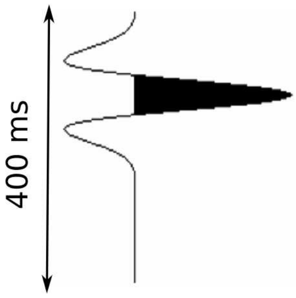


Figure 7.8: Source signature for Born and TT Born FWI.

7.4 Starting models and starting frequency

The starting models were provided by the company. I did not participate in obtaining them.

The velocity model was obtained from tomography and the salt velocity was set to be a constant value of 4500 m/s. This value was chosen after analysis based on well data. The starting model overlapped with PSDM section is shown in Figure 7.9. It is fairly simple and approximately one-dimensional in post-salt sediments.

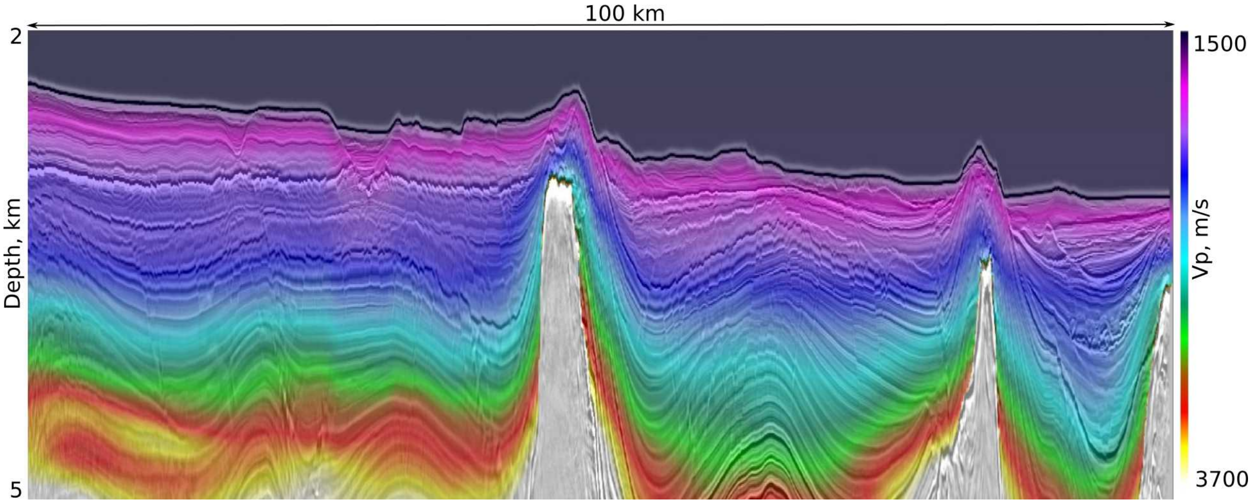
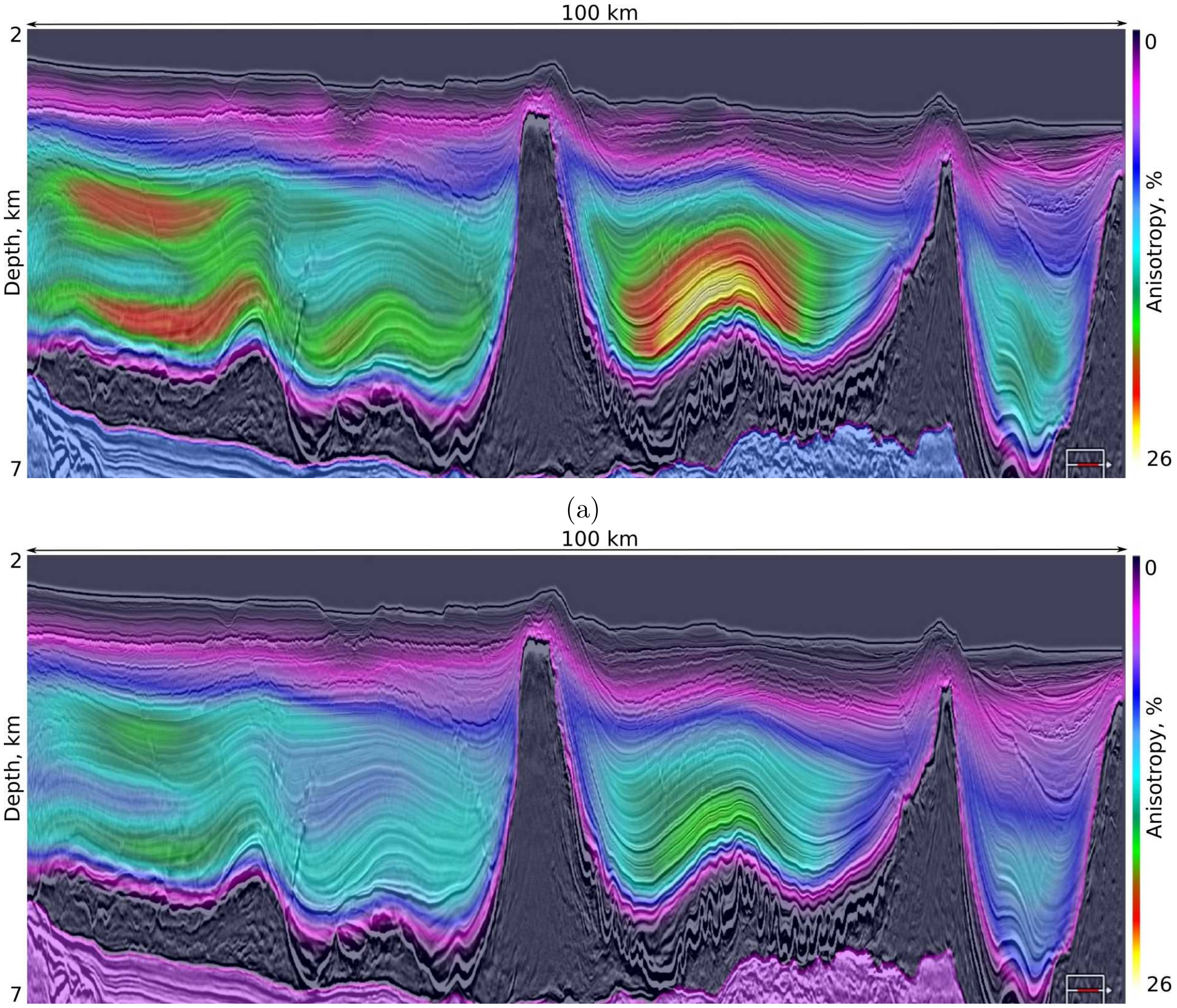


Figure 7.9: Starting velocity model based for the Picanha dataset overlaid with the PSDM section.

The area of study was highly anisotropic with epsilon and delta values up to 26% and 20% respectively (Fig. 7.10(a) and Fig. 7.10(b)). The anisotropy model was derived from 4th order RMO (residual move out) and calibrated to the wells. Both delta and epsilon models appear to have structures at the depth of approximately 4 km. These structures are correlated between the both models.

The reflectivity stack used for Born and TT Born FWI was obtained from Kirchhoff migration, filtered back to 15 Hz to avoid aliasing and then filtered with an F-K filter to eliminate fine details. The Kirchhoff migration was not performed by me, but I have tested different filters in order to obtain an optimal reflectivity stack. The stack shouldn't contain too finer detail as to avoid noise in the modelled wavefield. The final reflectivity stack is shown in Figure 7.11.



(a)

(b)

Figure 7.10: Anisotropy model for the Picanha dataset overlaid with the PSDM section: (a) epsilon and (b) delta. The models were derived from RMO.

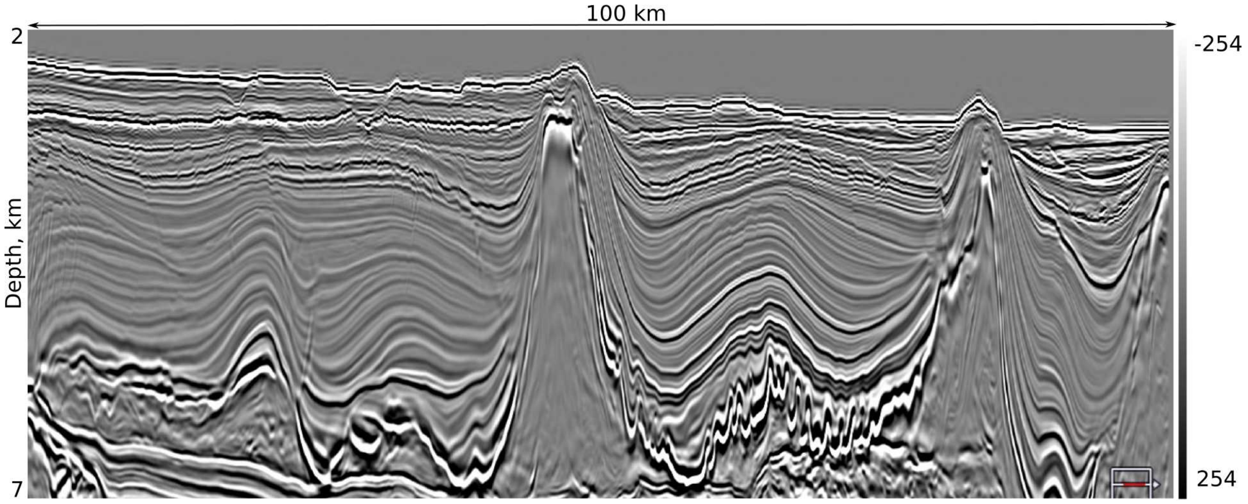
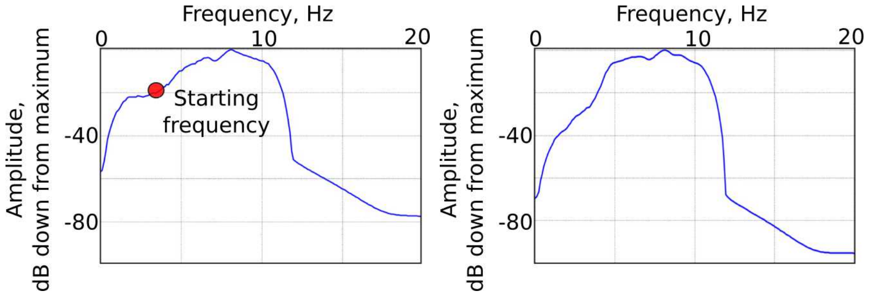


Figure 7.11: Reflectivity stack for Born and TT Born FWI. Obtained from Kirchhoff migration, filtered to 15 Hz and filtered with an F-K filter.

The amplitude spectra of both sets of the pre-processed data are shown in Figure 7.12. Both data volumes were low-passed using a zero-phase Ormsby filter rolling off from 10 to 12 Hz. The low-frequency end of the specta are different between the two volumes due to different processing manipulations applied to the data. For instance, the amplitude spectrum of the reflection-processed volume (Figure 7.12(b)) demonstrates higher signal-to-noise ratio at lower frequencies compared to that of refraction-processed data (Figure 7.12(a)). This effect of improving S/N ratio at low frequencies is caused by 4-D regularisation, that was applied with a primary purpose of forming the super shots.

For refraction FWI, the data were inverted using increasing frequency bandwidth with 4 Hz as the starting frequency and 9 Hz as the maximum frequency.

For reflection FWI the data were inverted in a single bandwidth with a maximum frequency of 12 Hz.



(a) (b)

Figure 7.12: The amplitude spectra of the data pre-processed for (a) refraction and (b) reflection FWI.

7.5 Inversion strategies

A number of different FWI techniques were applied to the dataset, both refraction FWI (Least Squares, Travel Time and Extended Parameter) and reflection FWI (Born and Travel Time Born). Each group required field data pre-processed differently. Whilst the field data for conventional, Travel Time and Extended parameter FWI must have multiples and surface ghosts retained within the data as these FWI types predicts the events during forward modelling, the field data for Born and Travel Time Born FWI must have multiples and surface ghosts removed. Fig. 7.13 demonstrates the FWI types applied to the dataset and highlights the different inputs required for each of them.

Time sampling of the data was 16 ms, however, the wavefield was sampled in time internally on a much finer grid to allow numerical stability of the simulation. The data contained frequencies with the maximum of 12 Hz. Therefore, to minimise the grid dispersion, the grid spacing of 32.5 m was chosen, as the code allows to cross no more than 2/3 of a grid cell at every time step. The strategies for each applied technique are summarised in Table 7.1.

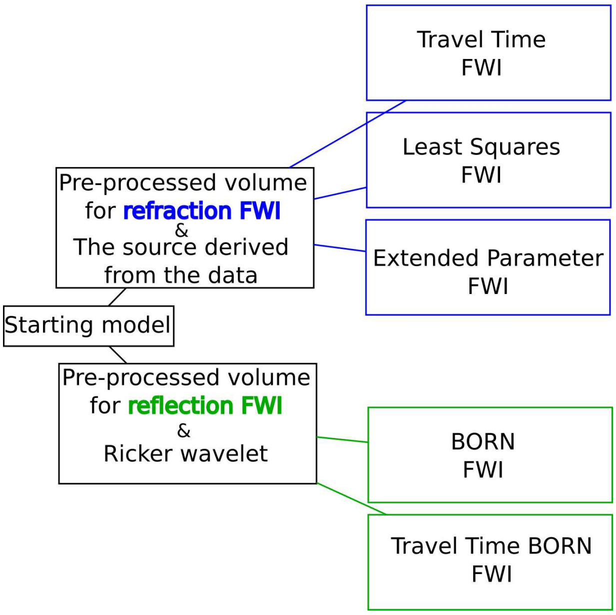


Figure 7.13: A variety of FWI types applied to the Picanha dataset.

	TT	LS	Extended Param.	BORN	TT BORN
Uses	Refractions			Reflections	
Model dimensions	3200 x 1120 x 320 cells				
Space sampling (m)	32.5m				
Time sampling	16ms				
Data volumes	Processed for refraction FWI			Processed for reflection FWI	
Source	Extracted from the data			Ricker wavelet	
No. of shots	5070			662	
No. of shots used	845			331	
No. of frequency bands	1	2	1	1	1
No. of iterations till converged	5	27	11	3	5
Max frequency	6.5Hz	7.5Hz	9Hz	12Hz	12Hz
TT window size	0.5s	n/a	n/a	n/a	0.5s
Max allowed time shift	0.15s	n/a	n/a	n/a	0.05s

Table 7.1: Summary of the inversion strategies for the Picanha dataset.

7.6 Results

Five different FWI types were applied to the dataset. Below I will present the results for every inversion performed.

7.6.1 Least Squares FWI

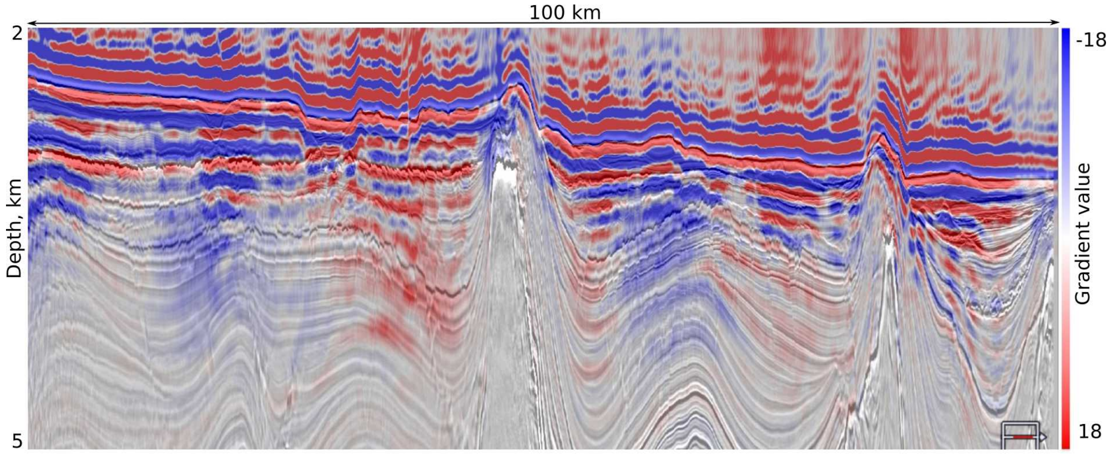
There were two tests conducted with conventional LS FWI. Both used the tomography velocity model (Fig. 7.9) as the starting model with the anisotropy model (Fig. 7.10) fixed during the inversion. Both inversion runs used 845 shots out of 5070 available shot records; the shots were randomised. The inversions had the starting frequency of 3.5 Hz and the maximum frequency of 7.5 Hz. Although both inversions were performed using the field data that were de-noised, de-bubbled, minimum-phased and low-passed, different mutes were applied to the data.

Test 1 uses the data with no mutes applied. The shot gathers are shown in Figure 7.14(c). The raw gradient for Test 1 appears to have a repeatable ringing pattern at the water bottom that can be seen as deep as 700 m below the water bottom (Fig. 7.14(a)). This may be happening because the acoustic propagator assumes a constant density. These artefacts in the gradient may lead to velocity imprints in the updates, hence a number of data mutes were tested and Test 2 was conducted.

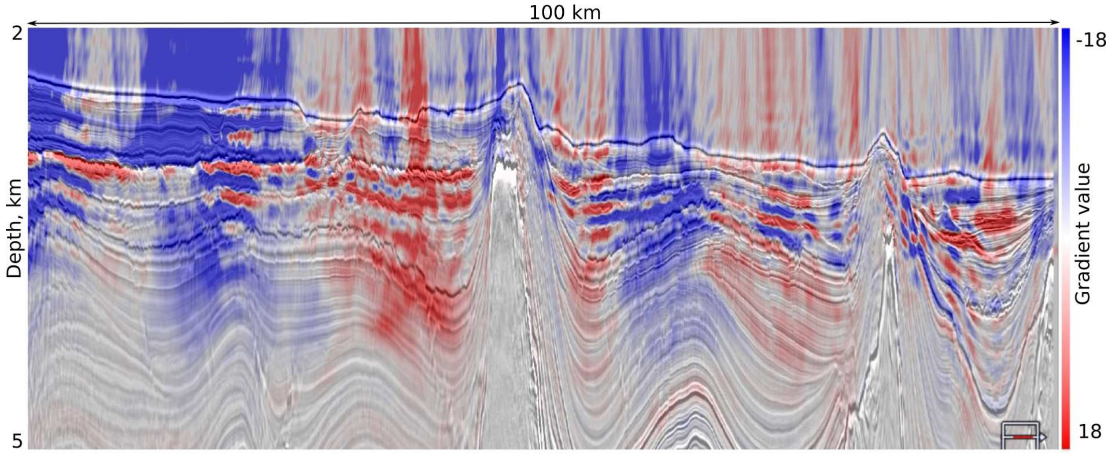
Test 2 also has a mute applied to the data in order to mute out the water bottom reflection and its multiple accordingly (Fig. 7.14(d)). The gradient for Test 2 is displayed in Figure 7.14(b). It does not have a ringing pattern that repeats the shape of the water bottom reflection. Deeper in the section, it follows the reflections and looks similar to that of the Test 1.

The Test 1 inversion has converged (here and everywhere else in this chapter, by convergence I mean that the code automatically stops to perform iterations when the objective function does not change for more than an internally set up limit) after 24 iterations and Test 2 inversion has converged after 27 iterations.

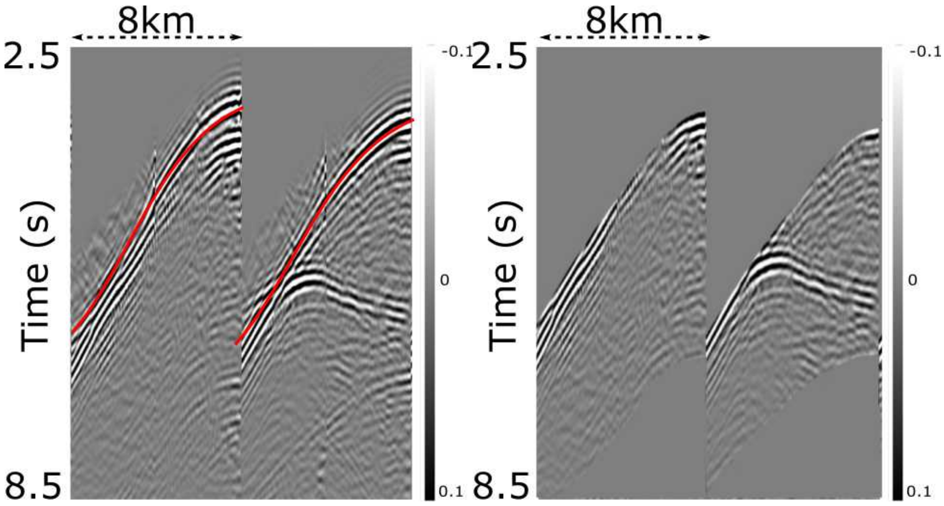
The velocity models obtained with the maximum frequency of 7.5 Hz for the two tests are demonstrated in Figure 7.15. These models are similar qualitatively: the same main velocity anomalies can be observed on both, such as the major reflector on the left-hand side of the model at the depth of approximately 2.8 km (marked as C in Figure 7.15(c)). However, the model from Test 2 (with water bottom mute applied to the data) is resolving the geological structures better and does not have the imprint of the gradient in the shallow part of the model. It can be seen that at the left-hand side of the section, the major bright reflector C, located at the depth of 2.8 km, is better defined; the shallow sedimentary layer just above it is also fully resolved compared to the Test 1 model. The in-between the salt diapirs sedimentary layers are also better resolved, for example the three inclined edge feature marked as A and located to the right from the first salt diapir. On the right-hand side of the model, to the right from the second salt diapir, the triangle-shaped feature marked as B and located at the depth of approximately 3 km is better defined as well as the basin underneath it.



(a)

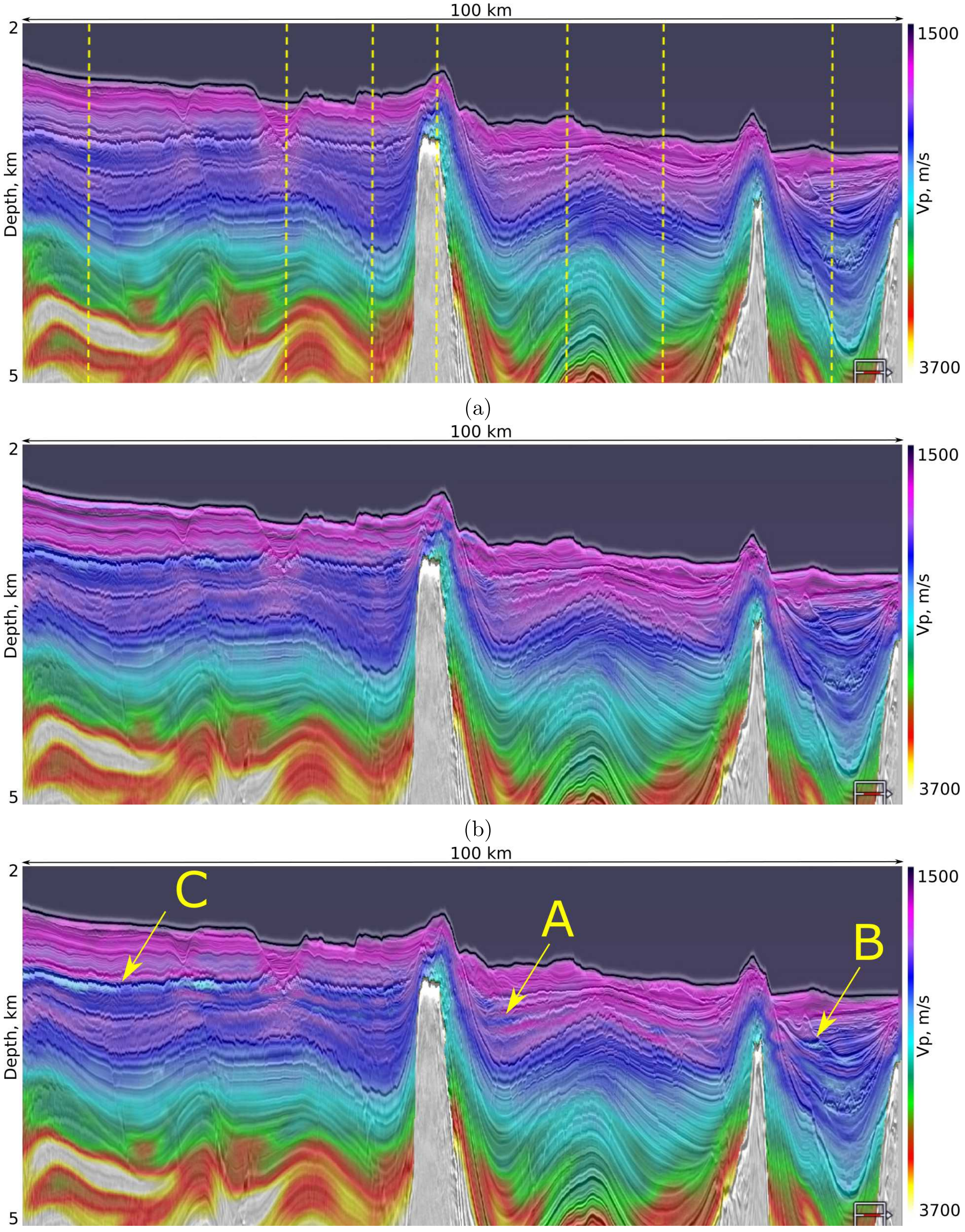


(b)



(c) (d)

Figure 7.14: Gradients calculated using the data (a) without and (b) with water bottom mute applied. The pre-processed data (c) without and (d) with the mute applied. The water bottom mute shown in red.



(a)

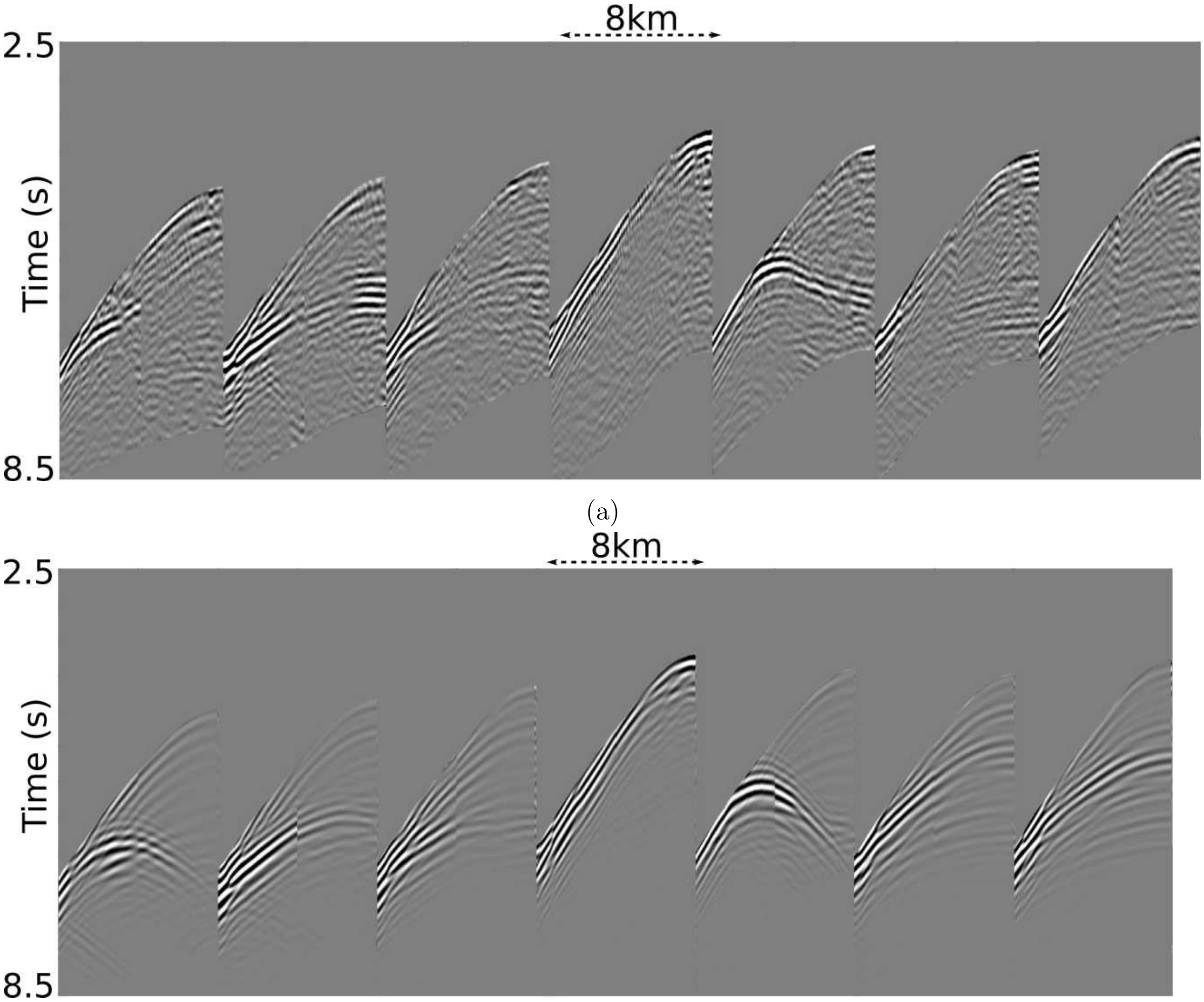
(b)

(c)

Figure 7.15: Comparison of (a) the starting velocity model and 7.5 Hz LS FWI models obtained using the data (b) with no mute applied to the data and (c) with the water bottom mute applied to the data. The PSDM stack has not been re-migrated with new FWI models.

The gathers modelled with the starting tomography model and the 7.5 Hz FWI velocity model (with water bottom mute applied to the data) are displayed in Figure 7.16(b). The location of the shots are shown in Figure 7.15(a) in the following manner: the first yellow dashed line at the left-hand side of the model is the location of the last shot gather displayed in Figure 7.16(b). It can be seen from the modelled shot gathers (Fig. 7.16(b) and Fig. 7.16(c)) that the inversion improved the match to the pre-processed data both on the near and the far offsets.

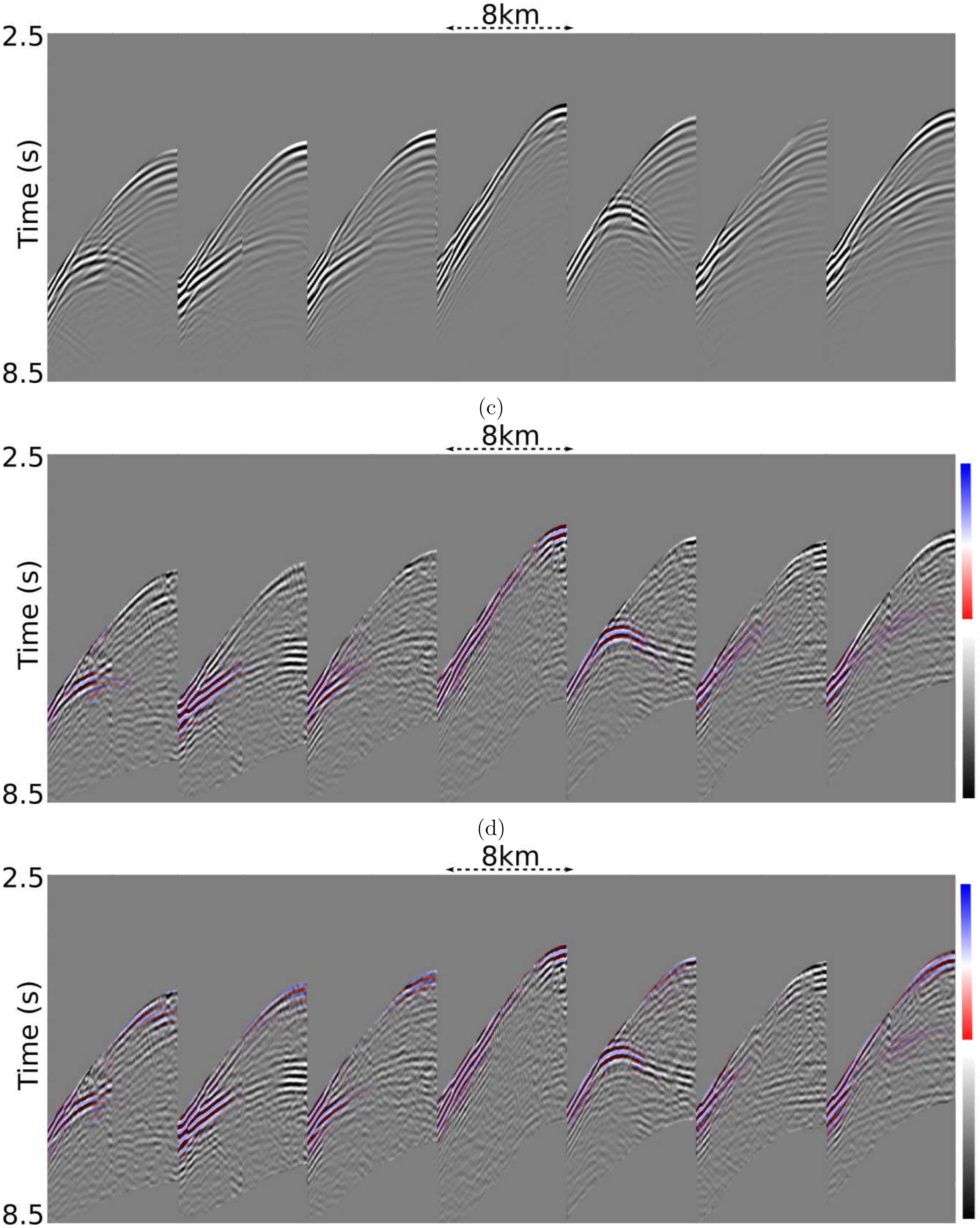
Test 2 produced reasonable updates in the shallow part of the model, improving the migration gathers above the bright horizon at approximately 2.9 km, making the deeper part of the model slightly too fast and improving the bright salt reflections (Fig. 7.17). The areas of improvement in the gathers migrated using the FWI models compared to those migrated using the starting model are circled in green; the areas where migrated gathers degrade are indicated in red. Most of the updates happened within first 3.5 km of the model. Common image gathers show that some parts of the model have been improved compared to the starting velocity model (mostly the very shallow part up to 3.5 km and salt reflections), whereas some areas of the model tend to become too fast.



(a)

(b)

Figure 7.16: QC shot gathers for Least Squares FWI: (a) the field data with the water bottom mute applied; (b) the modelled data using the starting velocity model; (c) the modelled data using the FWI model; (d) overlaid (a) and (b); overlaid (a) and (c). The pre-processed data on (d) and (e) are in grey colour scheme; the modelled data are in blue/red colour scheme. Blue and white denote a decrease in acoustic impedance, red and black denote an increase in acoustic impedance (cont.).

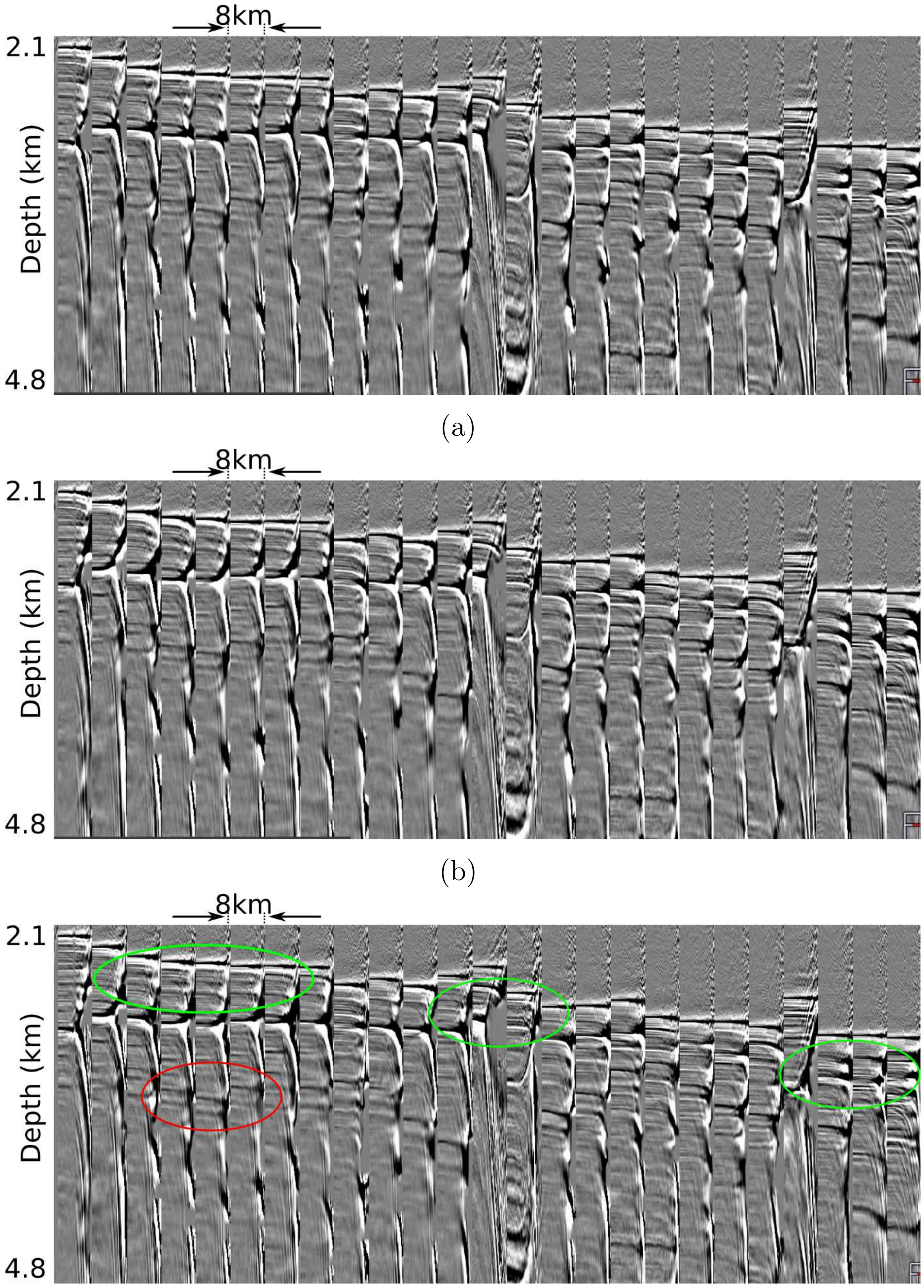


(c)

(d)

(e)

Figure 7.16: QC shot gathers for Least Squares FWI: (a) the field data with the water bottom mute applied; (b) the modelled data using the starting velocity model; (c) the modelled data using the FWI model; (d) overlaid (a) and (b); overlaid (a) and (c). The pre-processed data on (d) and (e) are in grey colour scheme; the modelled data are in blue/red colour scheme. Blue and white denote a decrease in acoustic impedance, red and black denote an increase in acoustic impedance.



(a)

(b)

(c)

Figure 7.17: Full offset common image gathers generated using: (a) the starting velocity model, (b) the LS FWI model obtained using the data with no mute applied and (c) the LS FWI model obtained using the data with water bottom mute applied. The areas of improvements compared to the starting model are highlighted in green.

7.6.2 Travel Time FWI

Travel Time FWI was applied to the data pre-processed for refraction FWI and top and bottom mutes applied to it. Because TT FWI seeks to improve the starting velocity model for further FWI applications, the mutes were applied in order to leave the water bottom and the shallow pack of first arrivals in the data to ensure that TT FWI would adjust the starting model if cycle skipped. The time window between the mutes was 2 s. The input gathers are shown in Figure 7.18.

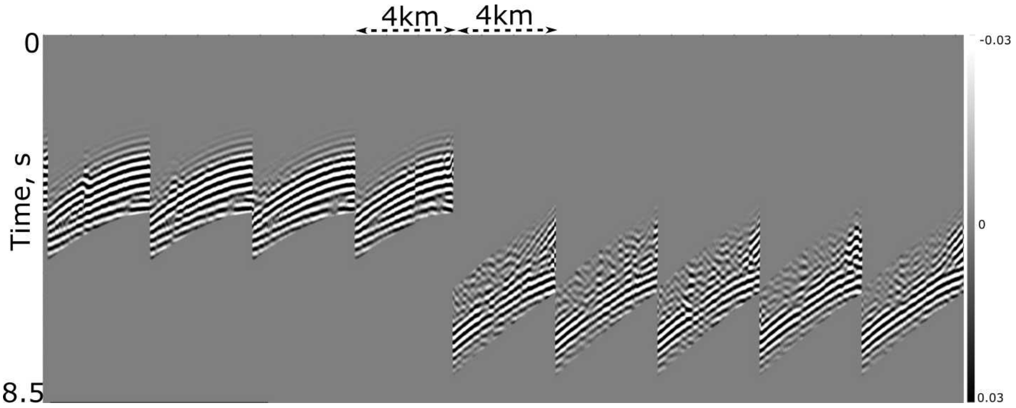
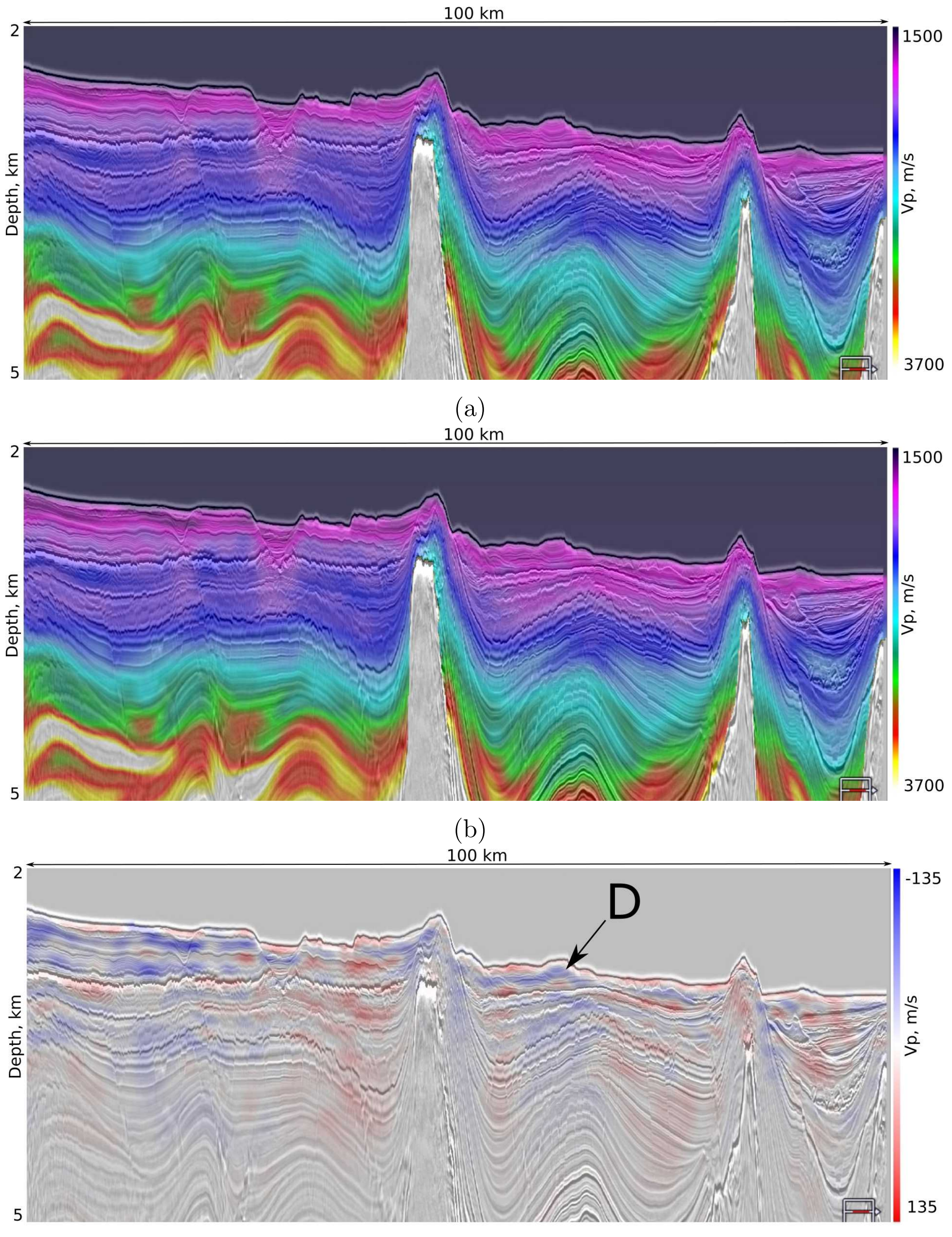


Figure 7.18: Shot gathers pre-processed for Travel Time FWI. The data are filtered at 6.5 Hz and have a top and a bottom mute applied.

The inversion has stopped after 5 iterations, which means that the objective function was not reducing above the internally set value. The maximum velocity update was about 60 m/s. The updates were consistent with geological structures apart from one area near the water bottom located to the right-hand side of the first salt dome, where the negative update (coloured in blue and marked as feature *D*) does not follow the sedimentary layer (Fig. 7.19).

This obtained Travel Time FWI model was used as the starting model for the Least-Squares inversion. The final LS FWI model was obtained after another 11 iterations. LS FWI has updated the model with reasonably large positive and negative values with the maximum velocity update around 140 m/s. The updates were consistent with the geology (Fig. 7.20(a)).

As quality control, Kirchhoff PSDM was performed and common image gathers are presented in Figure 7.21. Figure 7.21(a) and Figure 7.21(b) display the common image gathers obtained using the tomography model and the TT FWI. It can be seen, that the TT FWI improves the very shallow part of the model and flatters the under-migrated events above the first bright horizon at about 3 km depth on the left hand side of the section as indicated in area *E* in Figure 7.21(b). Some of the deeper events at about 3.5 km depth flattens as well compared to the starting velocity model, this is shown in area *F* in the same figure. Perhaps, the near salt events are getting under-migrated which indicates that the model has become too fast in these areas.



(a)

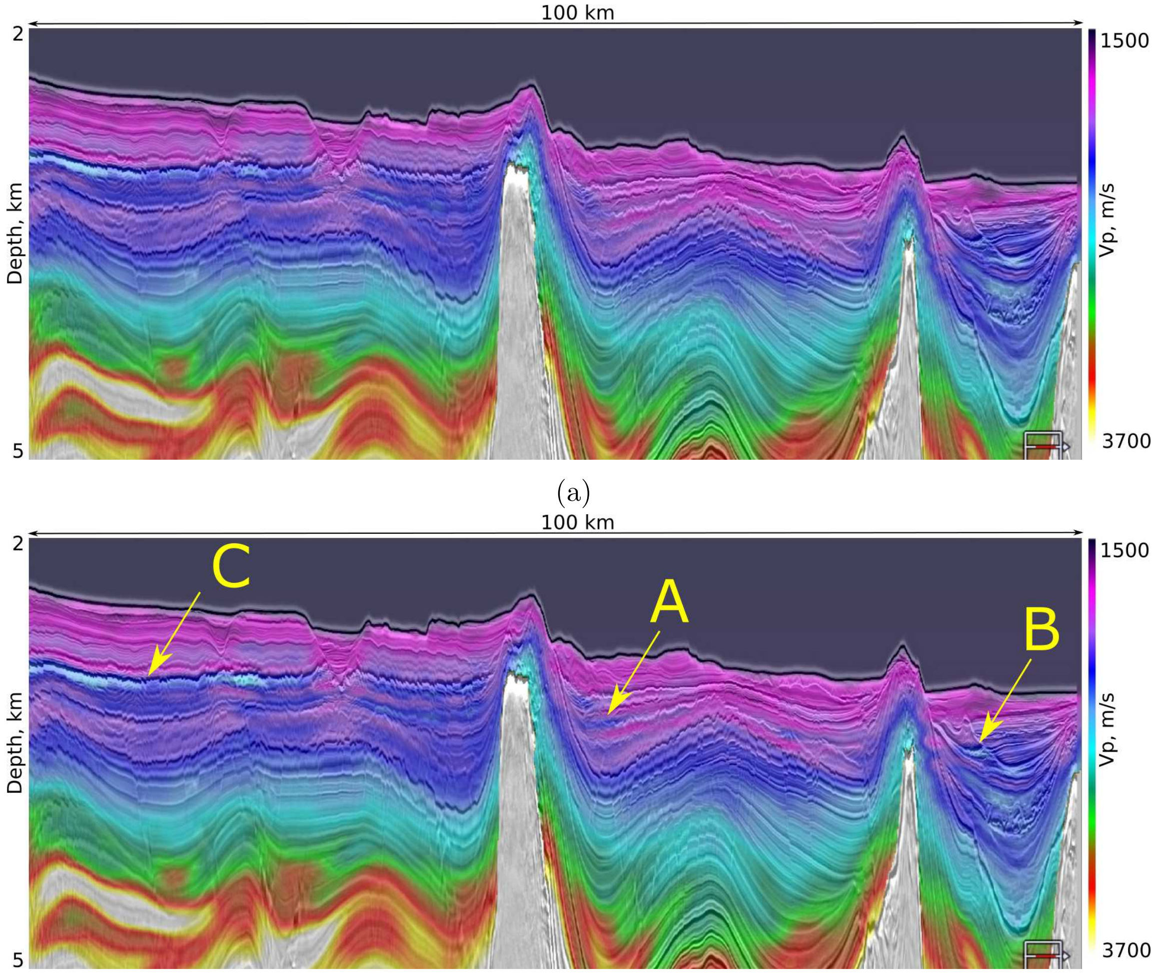
(b)

(c)

Figure 7.19: Comparison of (a) the starting model, (b) the TT FWI model and (c) the difference between them. The maximum velocity update is 60 m/s.

Then the LS FWI was performed twice using two different starting models: (1) the tomography and (2) the TT FWI model. When comparing these two LS FWI velocity models, both with the maximum frequency of 7.5 Hz, it can be seen that the two models are very close in main features but different in small detail only (Fig. 7.20). It can be seen that the reflector at approximately 2.8 km which is indicated as C in Figure 7.20(b), and the triangle shaped feature to the right from the second salt dome at 3 km depth (indicated as B) are more defined if using the tomography model as the starting model.

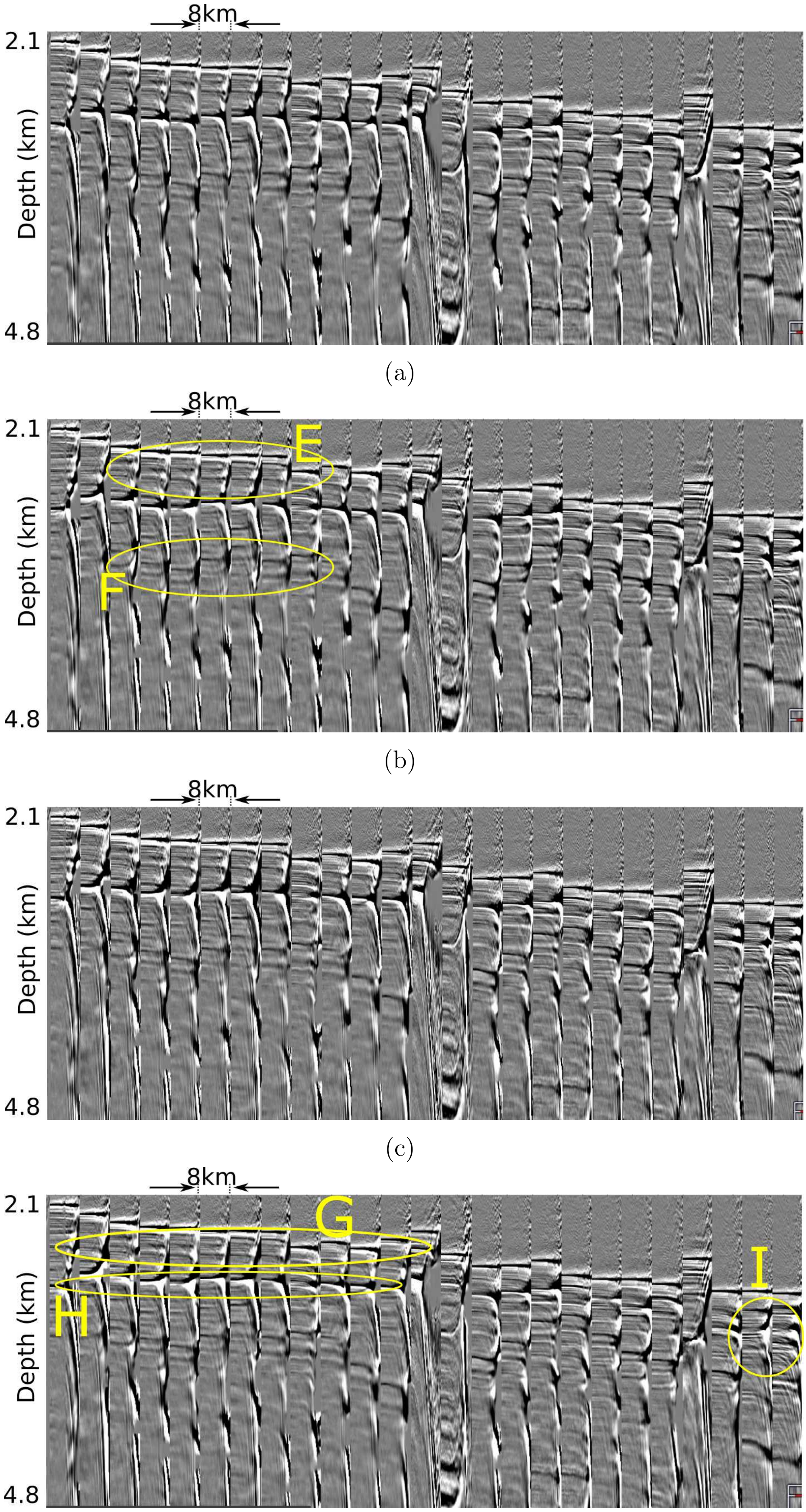
The common image gathers displayed in Figure 7.21(c) and Figure 7.21(d) obtained using the two different LS FWI models show that in some areas, for instance indicated as H and I, using the tomography model as the starting model provides better velocity model improvements. However, in other areas, particularly in the very shallow part of the model, up to 2.8 km depth and indicated as G, the multiple events are flatter if the TT FWI was used as the starting model for the LS FWI; these improvements can be observed at the offsets more than 4 km.



(a)

(b)

Figure 7.20: LS FWI velocity models obtained using: (a) TT FWI model as the starting model and (b) the tomography velocities as the starting velocity model.



(a)

(b)

(c)

(d)

Figure 7.21: Full offset common image gathers generated using (a) the tomography velocities, (b) the TT FWI model, the LS FWI models obtained using (c) the tomography model as the starting model and (d) using the TT FWI model as the starting model.

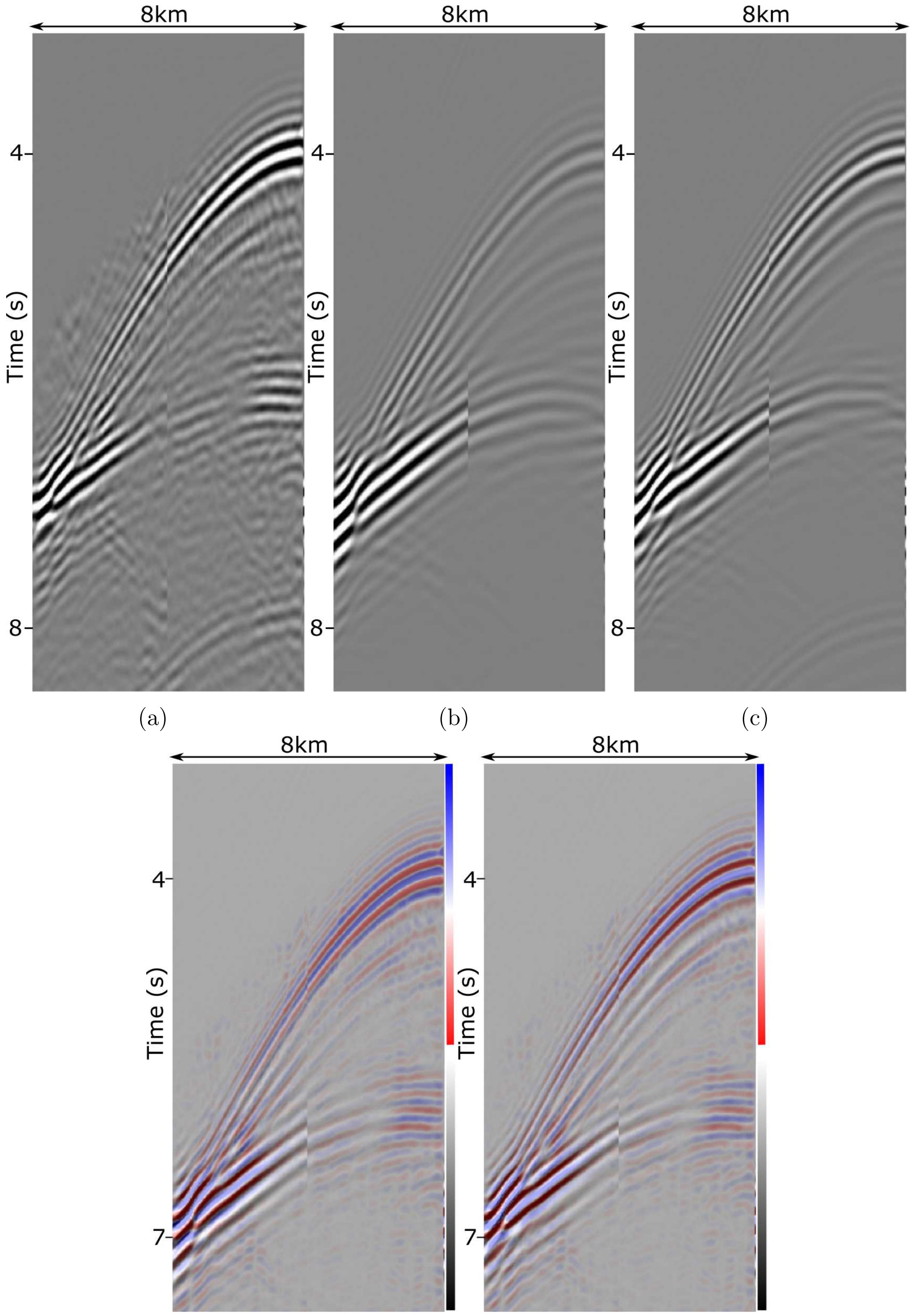
7.6.3 Extended Parameter FWI

After LS and TT FWI, the next type of FWI to test was Extended Parameter FWI. Extended Parameter FWI can be applied for both reflection and refraction case studies. Here, this FWI technique was applied to the data pre-processed for the refraction FWI. Extended Parameter FWI re-formulates the conventional objective function by expanding the solution space of the model and introducing the penalty term into it.

Unlike all the other FWI types, it was performed isotropically only. The starting tomography velocity model was converted to an isotropic model using the epsilon and delta models.

The first gradient had positive values everywhere in the model, which might indicate the presence of a shift between the modelled and the pre-processed data. However, on the synthetic gathers shown in Figure 7.22(d) and Figure 7.22(e) we can see that initially the mismatch at the water bottom on the near offsets as well as on the long offsets between the input and the modelled data is less than half a cycle. On the forward modelled data using the starting velocity model (Fig. 7.22(b)) the water bottom multiple is absent, while the inversion seems to recover and place it at the correct time (Fig. 7.22(c)). The relative amplitude of the modelled water bottom multiple is still not as strong as it is on the pre-processed data.

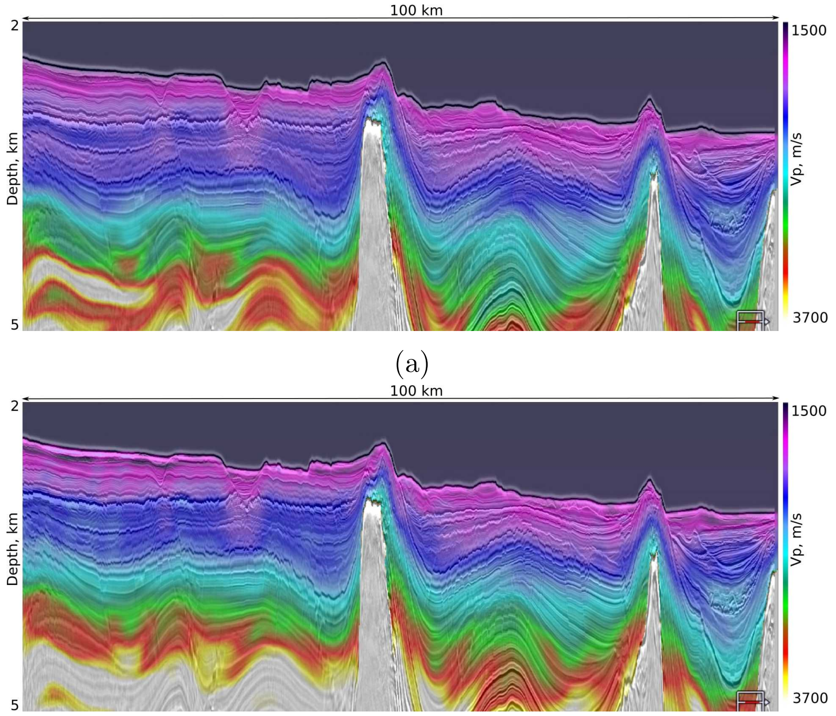
The FWI velocity model and the starting velocity model are displayed in Figure 7.23. The updates are positive everywhere and made the starting velocity model significantly faster, especially in the deep part of the model, below 3.5 km. This update appears to have a negative impact as can be seen from common image gathers that became under-migrated in the deeper part of the model (Fig. 7.24). The main reasons for this is that the inversion has tried to increase the velocities in the model because the anisotropy was not taken into account and the area was highly anisotropic with epsilon and delta values reaching 26% and 20% respectively.



(a) (b) (c)

(d) (e)

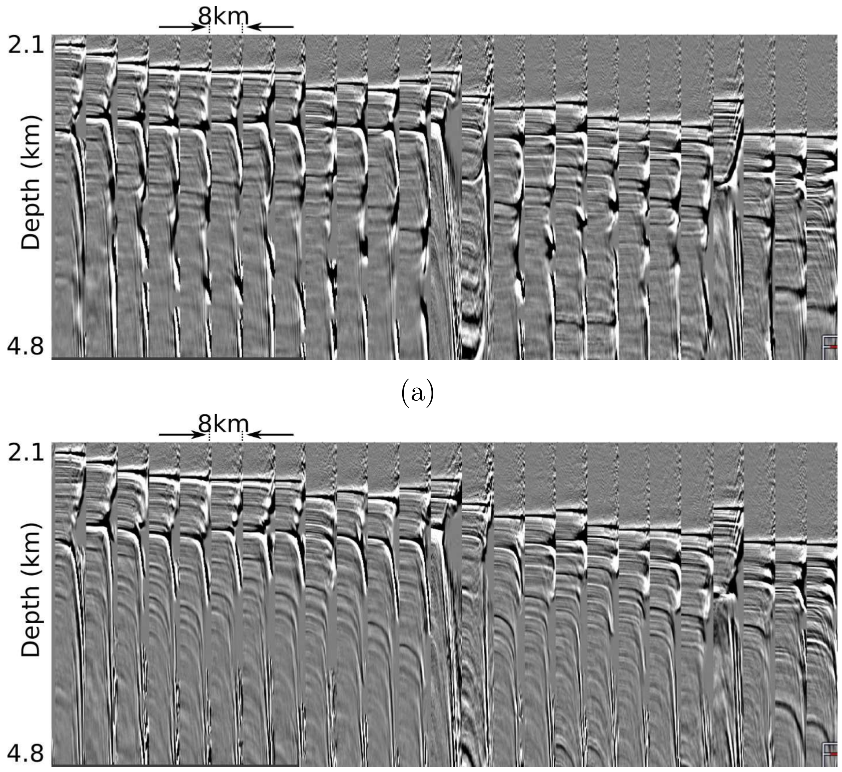
Figure 7.22: QC shot gathers for Extended Parameter FWI. (a) A pre-processed shot gather, (b) a modelled gather using the starting tomography model gather, (c) a modelled gather using the Extended Parameter FWI model; (d) overlaid (a) and (b; (e) overlaid (a) and (c). The pre-processed data on (d) and (e) are in blu-red colour scale. The modelled data on (d) and (e) are in grey colour scale. Blue and white denote a decrease in acoustic impedance, red and black denote an increase in acoustic impedance.



(a)

(b)

Figure 7.23: Comparison of (a) the starting model obtained from tomography and (b) Extended Parameter FWI model.



(a)

(b)

Figure 7.24: Full offset common image gathers generated using (a) the tomography model and (b) the Extended Parameter FWI model.

7.6.4 Born and TT BORN FWI

Born and TT Born FWI are types of reflection FWI that utilise Born approximation. Because reflection dominated or purely reflection data will only provide updates for the high wavenumbers if performing conventional FWI, in order to get updates of the smooth background model and, therefore, avoid cycle skipping, it is necessary to separate the tomographic and migration components.

The Picanha dataset is a reflection dominated dataset with very little, and at some areas no, refraction energy observed. Consequently, it was decided to perform reflection FWI.

Born and TT Born FWI were applied to the pre-processed dataset with multiples and ghosts removed from the field data. Removing these events is vital because Born approximation does not predict them in the wavefield.

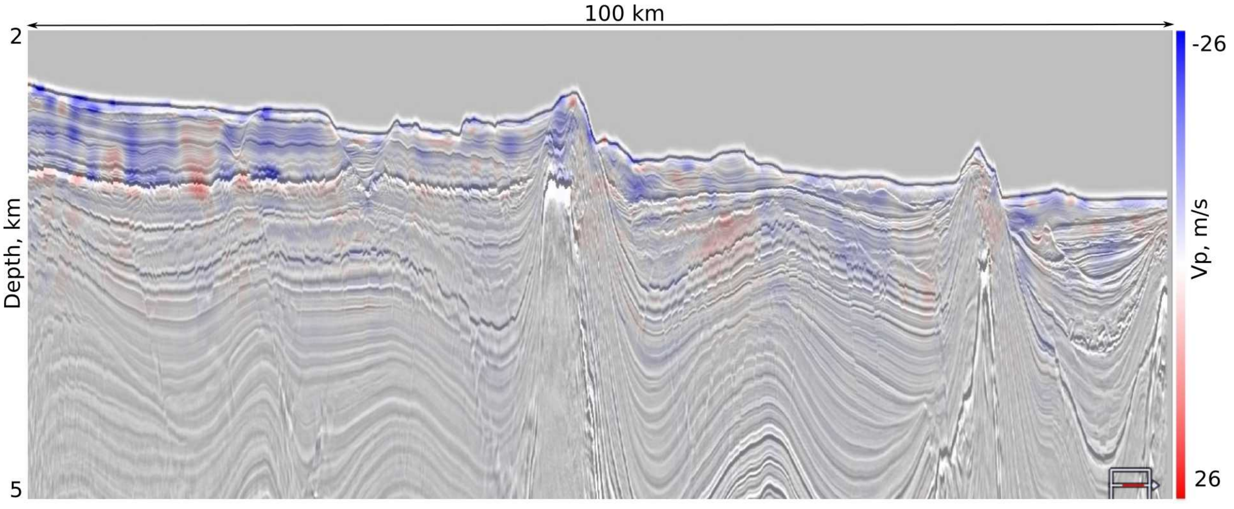
The Born inversion has automatically stopped after 3 iterations, which means that the objective function was not being reduced by more than the internally set up threshold. The inversion seemed to produce very little both positive and negative velocity updates with a maximum of 26 m/s. The majority of the velocity updates are negative and can be seen only up to a bright horizon that is located at the left hand side of the model at approximately 2.8 km depth and is indicated as *G* in Figure 7.25(b).

The shot gathers displayed in Figure 7.26 show that the starting velocity model provides a very close match on the nearest offsets and reasonably good match on the far offsets. There is no kinematic difference between the forward modelled shot using the starting velocities and using the Born FWI velocities on the nearest offsets and very little difference in amplitudes at the far offsets only.

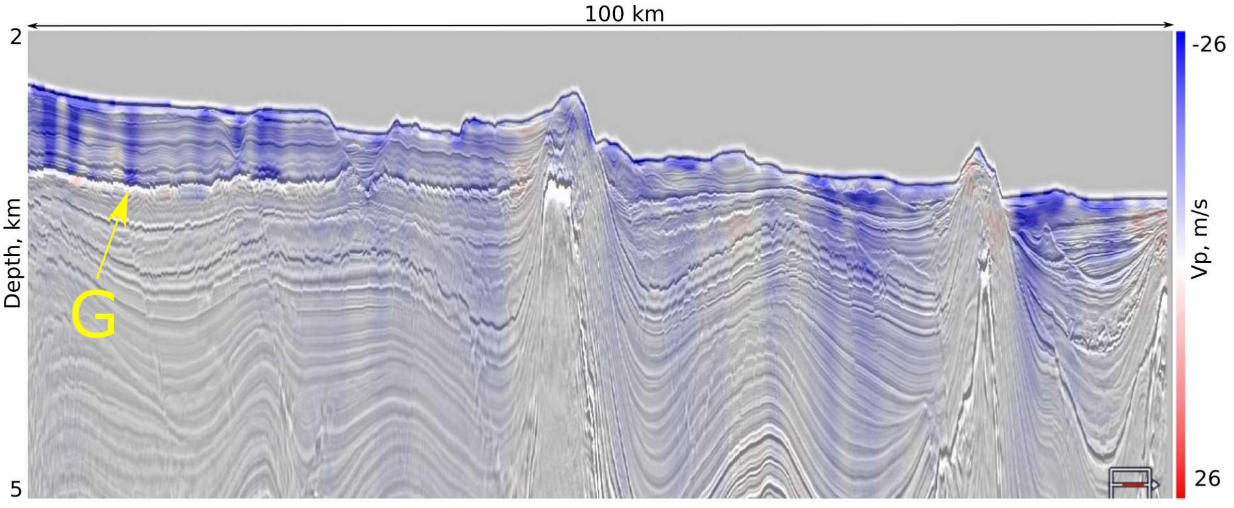
Travel Time Born FWI automatically stopped after 5 iterations and has produced bigger updates compared to the Born FWI with the maximum velocity update of 60 m/s (Fig. 7.25(b)). Most of the updates are seen in the very shallow part of the model, above the same bright horizon *G* located at 2.8 km depth. The updates are predominantly negative.

The updates being negative appears to be in a sensible direction for the inversion according to the migration gathers that indicate that velocities in the starting model are too fast above the major bright horizon *G* in the starting velocity model (Fig. 7.27). It can be seen from the common image gathers that the events above the bright horizon G are flattened as well as some deep events at 4 to 4.5 km depth from both sides of the second salt diapir on the right-hand side of the model, these areas are indicated as J and K.

The modelled shot gathers matched the pre-processed data very well for the reflections above 4.5 s. The deep reflections below 5.5 s were modelled kinematically correct (Fig. 7.26(d)), but the amplitudes in the modelled data are much stronger than they are in the pre-processed data (Fig. 7.26(a) and Fig. 7.26(d)).

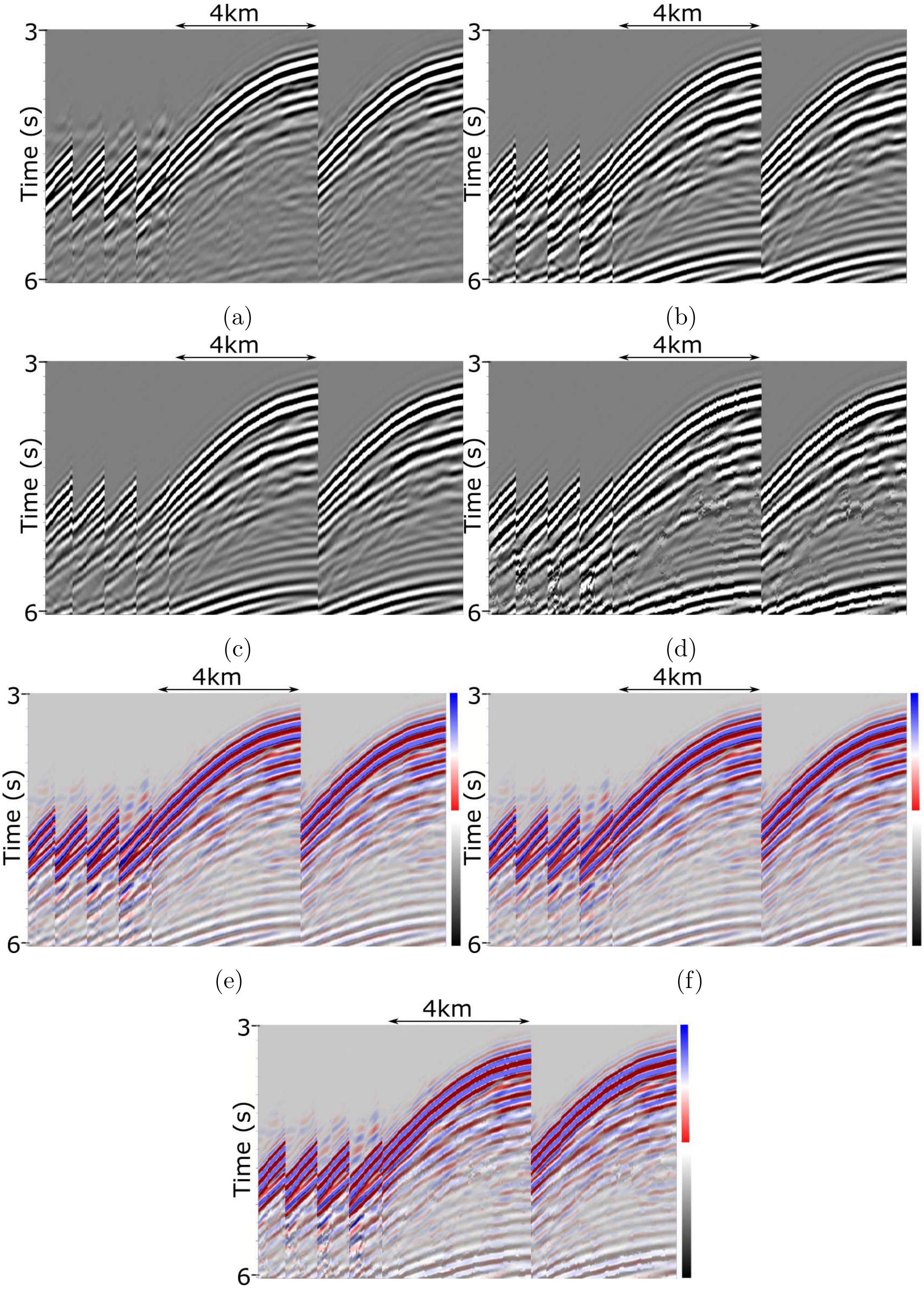


(a)



(b)

Figure 7.25: Comparison of the velocity updates obtained by (a) Born FWI and (b) TT Born FWI.



(a)

(b)

(c)

(d)

(e)

(f)

(g)

Figure 7.26: QC shot gathers for Born FWI and TT Born FWI. (a) Pre-processed raw shot gathers. Modelled shot gathers using (b) the starting model obtained from tomography, (c) the Born FWI model, (d) the TT Born FWI model. (e) Overlaid (a) and (b); (f) overlaid (a) and (c); (g) overlaid (a) and (d). The pre-processed field data on (e), (f) and (g) are in blue-red colour scale. The modelled data on (e), (f), and (g) are in grey colour scale. Blue and white denote a decrease in acoustic impedance, red and black denote an increase in acoustic impedance.

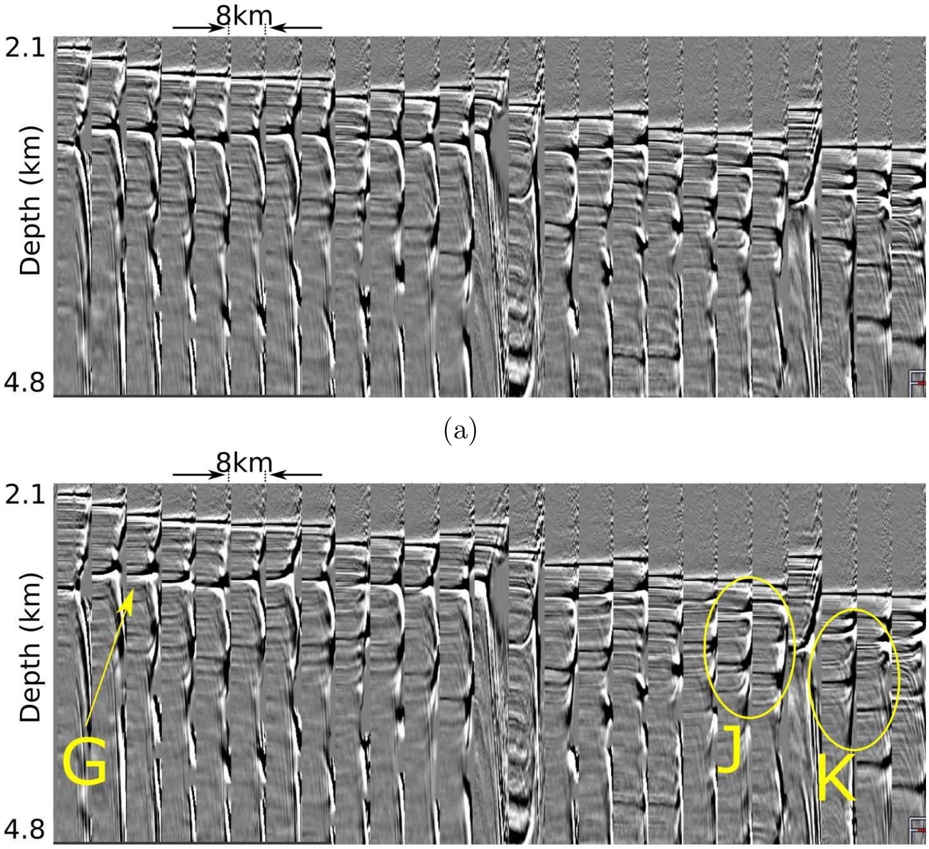


Figure 7.27: Comparison of the full offset common image gathers obtained using (a) the starting velocity model and (b) the TT Born FWI model.

7.7 Conclusions

Most of the applied FWI types improved the starting tomography velocity model, that can be seen from the coherence of the updates with geological structures, the improved fit of the modelled data to the pre-processed data and the overall flatter common image gathers.

Least Squares FWI with the water bottom mute applied to the data seemed to produce the biggest velocity updates, while Born FWI produced the smallest updates of under 30 m/s.

The water bottom mute was a solution to the repeatable pattern in the gradient mimicking the shape of the water bottom reflection and, therefore, contaminating the shallow part of the model. This could be happening because the acoustic propagator did not take density into account. It is well known, that the amplitudes of reflections are defined by the reflection coefficient, which in turn, depends on the acoustic impedance of the two media separated by the reflector. Acoustic impedance is a multiplication product of density and velocity of the medium. Because at the seabed the density contrast between water and sediment is much stronger than the velocity contrast between the two, the acoustic impedance and, therefore, reflection coefficient is predominantly defined by density rather than velocity. Consequently, density introduces a large impact on acoustic impedance at the seabed and by neglecting it, the predicted amplitudes will be wrong with respect to the observed amplitudes.

Travel Time FWI stopped after only 5 iterations with the maximum update about 60 m/s and improved the starting velocity model at the depth above 4 km as the migration gathers flattened, especially on the longest offsets. Travel Time FWI is designed to overcome the cycle-skipping problem, so we have used both the TT FWI velocity model and the tomography model as the starting models for Least-Squares inversion and the results came out very similar, which can indicate that the tomography model was already reasonably good and not cycle skipped.

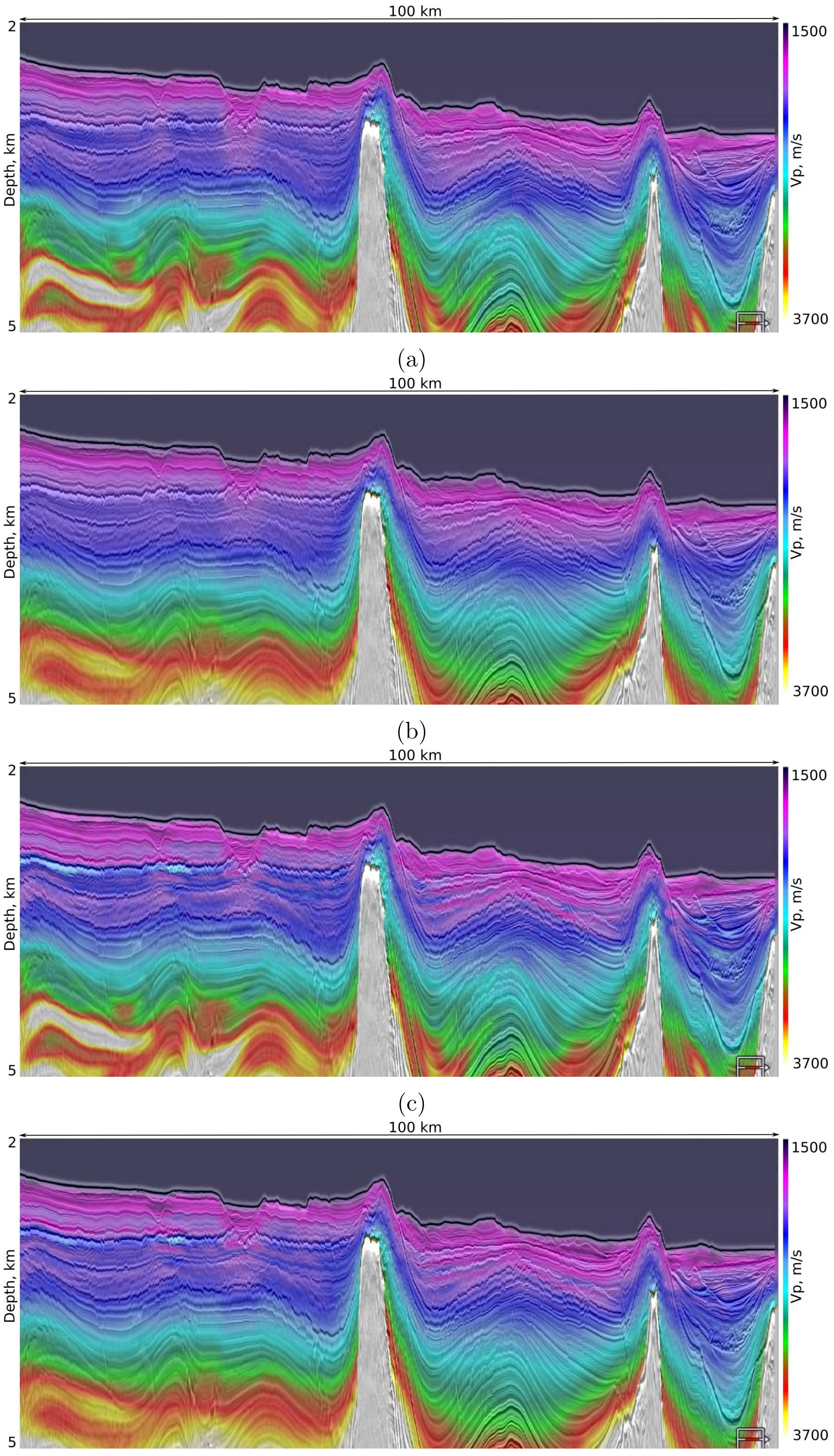
Extended parameter FWI has tried to increase the velocities everywhere in the model, the updates were coherent with the geology but did not seem sensible according to the migration gathers as they became heavily under-migrated below 3.5 km depth. The reason for this could be that the isotropic propagator was used in a highly anisotropic area. By excluding anisotropy, the velocity model has become too slow and the inversion was trying to increase it.

Both Born and Travel Time Born FWI have updated the model in a negative direction only, which was proven to be sensible by the migration gathers indicating that the velocities in the starting model needed to become slower. Born FWI has produced little updates so that there was practically no difference between the modelled shot gathers and common image gathers obtained using the starting tomography and the final Born FWI models. Travel Time Born FWI has produced the biggest updates and has improved the common image gathers in the very shallow part above 3.5 km and the deeper part of the model as well.

All of the FWI techniques applied to this particular dataset have produced qualitatively similar results and have not generated significant amounts of updates, which raises the question: what if our starting model requires a little update for the objective function to converge because it is already quite detailed and provides reasonably flat migration gathers?

To answer this question, we have smoothed the tomography model in slowness, reinserted the sharp top and bottom salt boundaries and filled the salt bodies with a constant velocity value of 4500 m/s (Fig. 7.28(b)). This kept the travel times accurate everywhere in between the salt diapirs. However, the near salt areas became too fast due to the high velocities of the salt becoming merged with the velocities of the surrounding sediments.

Least-Squares inversion from the smoothed tomography model defined the same major velocity anomalies as Least-Squares from the tomography model, but the latest version had more fine detail in it (Fig. 7.28(c) and Fig. 7.28(d)). Neither generated major updates below 4.5km but have updated the model above 4.5 km reasonably well. The two results are qualitatively similar, which may mean that in this particular case of geological environment and collected data, the provided tomography model may not require large updates in order for the algorithms to minimise the objective function.



(a)

(b)

(c)

(d)

Figure 7.28: Comparison of the velocity models: (a) starting velocity model obtained by tomography and (b) smoothed afterwards. (c) LS FWI models obtained using: (a) the tomography model as the starting model and (d) the smoothed tomography model as the starting model.

Chapter 8

Conclusions

The capability of conventional anisotropic acoustic FWI when it is applied to marine deep-water reflection-dominated NATS field datasets, that are not ideal for conventional FWI but represent the majority of acquired marine seismic dataset, were explored.

I applied FWI to three such datasets: the first used a deep-towed 10-km cable and was specifically acquired for 2D FWI on the NW Australian shelf; the second and third datasets were both 3D reflection-dominated surveys to which FWI had been previously applied with limited success by experienced processing companies – the second dataset was from Gabon and the third was from Brazil. These latter two datasets were chosen specifically because FWI had been previously tried and had failed.

A wide range of FWI-related tests and procedures were applied to the three datasets, and a number of conclusions were reached:

- Acoustic anisotropic FWI can generate accurate high-resolution velocity models that are of increasing complexity and resolution up to about 40 Hz where the input data have adequate turning energy and adequate low-frequency energy.

- Extending FWI beyond about 40 Hz to the full bandwidth of the field data produces minimal further change in the macro-velocity model, but nonetheless it continues to improve resolution up to and perhaps even beyond that which can be recovered by conventional Kirchhoff-based pre-stack depth migration. There is clear potential for full-bandwidth FWI applied to unprocessed field data to replace conventional processing and migration in favourable circumstances.

- Applying 2D and full-3D FWI to a single nominally-2D sail line produced similar outcomes though the resultant 2D and 3D models differed in detail and the 3D appeared to be marginally superior. The cost of 3D FWI though is many times higher than that of 2D, and the latter is useful in many situations for initial intensive testing and quality control.

- The Gabon dataset lacked significant refracted energy, but appeared otherwise amenable to FWI; it contained obvious shallow velocity targets, and a larger deep target, that appeared likely to produce a strong local velocity response during FWI. This dataset proved almost entirely resistant to FWI; extensive testing demonstrated that the reason for this was unlikely to be any one of the long list that can typically cause FWI problems. I concluded that there must have been some prior unknown adjustments made to the dataset that rendered it inconsistent with the core underlying assumption of FWI and/or that the specified acquisition geometry is not correct. Either of these would mean that the raw field data were not capable of being produced by a seismic wave equation applied to a model of physical earth properties.

- The Brazilian dataset was inverted using a third-party FWI code that assumed a constant density in the model. This assumption means that predicted amplitudes of the water-bottom reflection and of water-bottom-related multiples are significantly in error – they are too small. On this dataset, these strong observed amplitudes dominate the least-squares data mismatch, and various flavours of FWI are unable to fully recover from that initial mis-assumption.

The most important finding of this research can be summarised as: 3D FWI on NATS data can be made to work well, full-bandwidth FWI can usefully replace conventional pre-stack depth migration in deep water, but FWI on such datasets must be done appropriately if it is to be successful.

8.1 Recommendations and further work

Here, I outline recommendations for improving the results achieved in this thesis and suggest the scope for future work.

For the Pivot dataset, the high-resolution full-bandwidth FWI looks very promising. The demonstrated results can be improved by (1) removing the high-frequency noise observed on the differentiated FWI velocity model and (2) optimising the inversion strategy to reduce the inversion cost.

For the Gabon dataset, where FWI appears to have been compromised by poor industry protocols to assure that the data is really raw and/or that the geometry was correct, the only way to successfully invert the dataset is to find out what has been done to the data and get an access to the raw data.

For the Brazilian dataset, it would be useful to re-invert the data taking density into account; ION is performing this test.

8.1.1 Full-bandwidth FWI

To date, the only published materials of FWI applications that I could find use frequencies up to 40 Hz only. In this thesis, it has been demonstrated that full-bandwidth FWI clearly has great potential. The high-resolution full-bandwidth FWI applied to unprocessed field data has shown impressive results applied to a regional 2D deep-water long-cable seismic line. It would be interesting to:

- test full-bandwidth FWI applied in 3D to a dataset situated in shallower water and acquired with a shorter cable. In the event of success, the long-term goal for FWI to replace conventional processing might be finally becoming a realistic possibility.
- perform full-bandwidth FWI applied in 3D on a cloud rather than a conventional cluster to explore the high-speed potential of FWI and show whether it is possible to obtain high-resolution images of the subsurface from the raw data as fast as 7-10 days after the data has been acquired.

Bibliography

Agudo, O. C. (2018). *Acoustic full-waveform inversion in geophysical and medical imaging.* Ph.D. thesis, Imperial College London.

Agudo, O. C., N. Vieira da Silva, M. Warner, and J. Morgan (2017). “Addressing viscous effects in acoustic full-waveform inversion.” In: *79th EAGE Conference & Exhibition 2017, Extended Abstract.*

Agudo, O. C., N. Vieira da Silva, M. Warner, T. Kalinicheva, and J. Morgan (July 2018). “Addressing viscous effects in acoustic full-waveform inversion”. In: *Geophysics* 83.6, R611–R628.

Alford, R. M., K. R. Kelly, and D. M. Boore (Dec. 1974). “Accuracy of finite-difference modeling of the acoustic wave equation”. In: *Geophysics* 39.6, pp. 834–852.

Alkhalifah, T. A. (2016). *Full Waveform Inversion in an Anisotropic World - Where are the parameters hiding?* EAGE Publications.

Alterman, Z. and F. C. Karal (1968). “Propagation of elastic waves in layered media by finite-difference methods”. In: *Bulletin of the Seismological Society of America* 58, pp. 367–398.

Ashley, J. (2019). *3D full waveform inversion of narrow azimuth towed streamer seismic data.* Ph.D. thesis, Imperial College London.

Backus, G. (1970). “Inference from inadequate and inaccurate data”. In: *Proceedings of the National Academy of Sciences* 65.1, pp. 100–105.

Backus, G. and F. Gilbert (1968). “The resolving power of gross earth data”. In: *Geophysical Journal of the Royal Astronomical Society* 16, pp. 169–205.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux (2008). “Velocity model building by 3D frequency-domain, full waveform inversion of wide-aperture seismic data”. In: *Geophysics* 73.5, VE101–VE117.

Berenger, J. P. (1994). “A perfectly matched layer for the absorption of electromagnetic waves”. In: *Journal of Computational Physics* 114, pp. 185–200.

Beydoun, W. B. and A. Tarantola (1998). “First Born and Rytov approximation: Modelling and inversion conditions in a canonical example”. In: *Journal of the Acoustical Society of America* 83, pp. 1045–1055.

Biondi, B. and A. Almomin (2012). “Tomographic full-waveform inversion (TFWI) by combining full waveform inversion and wave-equation migration velocity analysis”. In: *SEG Technical Program Expanded Abstracts 2012*. DOI: `10.1190/segam2012-0275.1`.

— (Sept. 2013). “Tomographic full-waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis”. In: *The Leading Edge* 32.9, pp. 1074–1080. DOI: `https://doi.org/10.1190/tle32091074.1`.

— (2015). “Simultaneous inversion of full data bandwith by tomographic full-waveform inversion”. In: *Geophysics* 79.3, WA129–WA140. DOI: `10.1190/geo2013-0340.1`.

Black, M., K. D. McCormac, C. Elders, and D. Robertson (2017). “Extensional fault evolution within the Exmouth Sub-basin, North West Shelf, Australia”. In: *Marine and Petroleum Geology* 85, pp. 301–315. DOI: `https://doi.org/10.1016/j.marpetgeo.2017.05.022`.

Boore, D. M. (1970). *Finite-difference solutions to the equations of elastic wave propagation, with application to Love waves over dipping interfaces*. Ph.D. thesis, M.I.T.

Boyer, T. P. et al. (2013). *World Ocean Database 2013*.

Brenders, A. J. and R. G. Pratt (Jan. 2007). “Waveform tomography of marine seismic data: What can limited offset offer?” In: *SEG Technical Program Expanded Abstracts 2007*, pp. 3024–3028.

Brittan, J. and I. Jones (Mar. 2019). “FWI evolution: from a monolith to a toolkit”. In: *The Leading Edge* 38.3, pp. 179–184. DOI: `10.1190/tle38030179.1.`.

Brownfield, M. E. and R. R. Charpentier (2006). “Geology and Total Petroleum Systems of the West Central Coastal Province (7203), West Africa.” In: *U.S. Geological Survey Bulletin* 24.2207–B.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent (Sept. 1995). “Multiscale seismic waveform inversion”. In: *Geophysics* 60.5, pp. 1457–1473. DOI: `https://doi.org/10.1190/1.1443880`.

Carnarvon Basin (2019). URL: `https://www.ga.gov.au/scientific-topics/energy/province-sedimentary-basin-geology/petroleum/offshore-northwest-australia/canarvon` (visited on 09/04/2019).

Castro, R. D. and J. P. Picolini (2016). “1- Main features of the Campos Basin regional geology”. In: *Geology and Geomorphology: Regional Environmental Characterization of the Campos Basin, Southwest Atlantic*, pp. 1–12.

Causse, E., R. Mittet, and B. Ursin (2018). “Preconditioning of full-waveform inversion in viscoacoustic media.” In: *Geophysics* 64.1, pp. 130–145.

Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef (1985). “A nonreflecting boundary condition for discrete acoustic and elastic wave equations”. In: *Geophysics* 50, pp. 705–708.

Chen, C. T. and F. J. Millero (1977). “Speed of sound in seawater at high pressures”. In: *The Journal of the Acoustical Society of America* 62.5, pp. 1129–1135.

Claerbout, J. F. (1971). “Toward a unified theory of reflector mapping”. In: *Geophysics* 36, pp. 467–481.

— (1985). *Imaging the Earth’s interior.* Palo Alto, Calif. : Blackwell Scientific Publications.

Claerbout, J.F. and F. Muir (1973). “Robust modelling with erratic data”. In: *Geophysics* 38, pp. 826–844.

Clayton, R. and B. Engquist (1977). “Absorbing boundary conditions for acoustic and elastic wave equations”. In: *Bulletin of the Seismological Society of America* 67, p. 1529.

Courant, R., K. Friedrichs, and H. Lewy (1967). “On the partial difference equations of mathematical physics”. In: *IBM Journal of Research and Development* 11.2, pp. 215–234. DOI: `10.1147/rd.112.0215`.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola (1990). “Robust elastic nonlinear waveform inversion: application to real data”. In: *Geophysics* 55, p. 527.

Dablain, M. A. (1986). “The application of higher differencing to the scalar wave equation”. In: *Geophysics* 51, pp. 54–66.

Debens, H. A. (2016). *Three-Dimensional Anisotropic Full-Waveform Inversion.* Ph.D. thesis, Imperial College London.

Del Grosso, V. A. (1974). “New Equation for speed of sound in natural waters (with comparison to other Equations)”. In: *The Journal of the Acoustical Society of America* 56.4, pp. 1084–1091.

Diaz, E. and A. Guitton (2011). “Fast full-waveform inversion with random shot decimation.” In: *SEG Technical Program Expanded Abstracts 2011*, pp. 2804–2808.

Dix, C. H. (1955). “Seismic velocities from surface measurements.” In: *Geophysics* 20, pp. 68–86.

Dupre, S., G. Bertotti, and S. Cloetingh (Mar. 2007). “Tectonic history along the South Gabon Basin: Anomalous early post-rift subsidence”. In: *Marine and Petroleum Geology* 24.3, pp. 151–172.

Etgen, J. and C. Regone (1998). “Strike shooting,dips hooting,wide patch shooting - Does prestack depth migration care? A model study.” In: *SEG Technical Program Expanded Abstracts 1998*, pp. 66–69.

Gardner, G., L. Gardner, and A. Gregory (1974). “Formation velocity and density - The diagnostic basics for stratigraphic traps.” In: *Geophysics* 39.6, pp. 770–780.

Gauthier, O., J. Virieux, and A. Tarantola (1986). “Two-dimensional nonlinear inversion of seismic waveforms: numerical results”. In: *Geophysics* 51, p. 1387.

Gelis, C., J. Virieux, and G. Grandjean (2007). “Two-dimensional elastic full waveform inversion using born and rytov formulations in the frequency domain”. In: *Geophysical Journal International* 168, pp. 605–633.

Graves, R. W. (1996). “Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences”. In: *Bulletin of the Seismological Society of America* 86.4, pp. 1091–1106.

Guash, L. (2011). *3D Elastic Full-Waveform Inversion.* Ph.D. thesis, Imperial College London.

Hadamard, J. (1902). “Sur les problème aux derivées partielles et leur signification physique”. In: *Princeton Univ. Bull.* 13, pp. 49–52.

Han, F. X., J. G. Sun, and K. Wang (2012). “The influence of sea water velocity variation on seismic travel times, ray paths and amplitude.” In: *Applied Geophysics* 9.3, pp. 319–325.

Hu, W., J. Chen, J. Liu, and A. Abubakar (2018). “Retrieving low wavenumber information in FWI”. In: *IEEE Signal Processing Magazine* 35.2, pp. 132–141. DOI: `10.1109/MSP.2017.2779165`.

Hudson, J. and J. Heritage (1980). “The use of the Born approximation in seismic scattering problems”. In: *Geophysical Journal of the Royal Astronomical Society* 66.1, pp. 221–240. DOI: `https://doi.org/10.1111/j.1365-246X.1981.tb05954.x`.

Hung, B., K. F. Xin, S. Birdus, and J. Sun (2008). “3-D tomographic amplitude inversion for compensating transmission losses in the overburden”. In: *70th EAGE Conference & Exhibition 2008, Extended Abstract.*

Irabor, K. and M. Warner (Sept. 2016). “Reflection FWI”. In: *SEG Technical Program Expanded Abstracts 2016*, pp. 1136–1140. DOI: `https://doi.org/10.1190/segam2016-13944219.1`.

Jones, I. F. (2010). *An introduction to: Velocity model building.*

Kalinicheva, T., M. Warner, J. Ashley, and F. Mancini (Sept. 2017). “Two- vs three-dimensional full-waveform inversion in a 3D world.” In: *SEG Technical Program Expanded Abstracts 2017*, pp. 1383–1387.

Kamaluddin, T., Y. Othman, S. Saiful, and A. Azani (Mar. 2011). “Determination of speed of sound using empirical equations and SVP”. In: *Proceedings - IEEE 7th International Colloquium on Signal Processing and Its Applications, CSPA 2011*, pp. 252–256. DOI: `10.1109/CSPA.2011.5759882`.

Kuperman, W. A. and J. F. Lynch (2004). “Shallow-water acoustics”. In: *Physics Today* 57.10, pp. 55–61.

Lailly, P. (1983). “The seismic inverse problem as a sequence of before stack migrations”. In: *Conference on Inverse Scattering: Theory and Application*, pp. 206–220.

Lazaratos, S., I. Chickichev, and K. Wang (2011). "Improving the convergence rate of full wavefield inversion using spectral shaping". In: *SEG Technical Program Expanded Abstracts 2011*, pp. 2428–2432.

Leeuwen, T. van and F. J. Hermann (2013). "Mitigating local minima in full-waveform inversion by expanding the search space". In: *Geophysical Journal International* 195, pp. 661–667. DOI: `10.1093/gji/ggt258.`.

Levander, A. (1988). "Fourth-order finite-difference P-SV seismograms". In: *Geophysics* 53, p. 1425.

Lo, T. and P. Inderwiesen (1994). *Fundamentals of seismic tomography.* Society of Explorational Geophysicists.

Longley, I. M. et al. (2002). "The North West Shelf of Australia - A Woodside Perspective". In: *The sedimentary basins of Western Australia 3 : Proceedings of the Petroleum Exploration Society of Australia Symposium, Perth, WA*, pp. 22–88.

Lu, R., S. Lazaratos, S. Hughes, and D. Leslie (2016). "Revealing overburden and reservoir complexity with high-resolution FWI". In: *SEG Technical Program Expanded Abstracts 2016*, pp. 1242–1246.

Luo, S. and P. Sava (2011). "A deconvolution-based objective function for wave-equation inversion". In: *SEG Technical Program Expanded Abstracts 2011*, pp. 2788–2792.

Luo, Y. and G. Schuster (1991). "Wave-equation travel time inversion". In: *Geophysics* 56, pp. 645–653.

Lupinacci, W. M. and S. A. M. Oliveira (2015). "Q factor estimation form the amplitude spectrum of the time-frequency transform of stacked reflection seismic data." In: *Journal of Applied Geophysics* 114, pp. 202–209.

Ma, Y. and D. Hale (2013). "Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion". In: *Geophysics* 78.6, R223–R233. DOI: `10.1190/geo2013-0004.1`.

Mackenzie, K. V. (1981). "Nine-term equation for sound in the oceans". In: *The Journal of the Acoustical Society of America* 70.3, pp. 807–812.

Mancini, F., J. Moss, K. Prindle, and T. Ridssdill-Smith (2015). “Where can Full Waveform Inversion have the biggest impact in the exploration and production cycle?” In: *77th EAGE Conference & Exhibition 2015, Extended Abstract.*

Medwin, H. (1975). “Speed of sound in water: a sample equation for realistic parameters”. In: *The Journal of the Acoustical Society of America* 58.6, pp. 1318–1319.

Metivier, L., R. Brossier, E. Oudet, Q. Merigot, and J. Virieux (2016). “Overcoming Cycle Skipping in FWI - An Optimal Transport Approach”. In: *78th EAGE Conference & Exhibition 2016, Extended Abstract.*

Mohriak, W. U., M. R. Mello, J. F. Dewey, and J. R. Maxwell (Jan. 1990). “Petroleum geology of the Campos Basin, offshore Brazil”. In: *Geological Society* Special Publications, 50, pp. 119–141. DOI: `https://doi.org/10.1144/GSL.SP.1990.050.01.07`.

Mora, P. (1987). “Nonlinear two-dimensional elastic inversion of multioffset seismic data”. In: *Geophysics* 52, pp. 1211–1228.

— (1988). “Elastic wave-field inversion of reflection and transmission data”. In: *Geophysics* 53, pp. 750–759.

— (Dec. 1989). “Inversion = migration + tomography”. In: *Geophysics* 54.12, pp. 1575–1586.

Morgan, J. (2013). “Next generation seismic experiments: wide-angle, multi-azimuth, 3D FW - using OBS”. In: DOI: `http://www.obsip.org/documents/LA-morgan-oct2013_public.pdf`.

Mufti, I. R. (Sept. 1990). “Large scale three-dimensional seismic models and their interpretive significance”. In: *Geophysics* 55.9, pp. 1166–1182.

Nangoo, T. P. (2013). *Seismic full waveform inversion of 3D field data - from the near surface to the reservoir.* Ph.D. thesis, Imperial College London.

Newman, P. (1973). “Divergence effects in a layered earth”. In: *Geophysics* 38.3, pp. 481–488.

Operto, S., J. Virieux, P. Amestoy, J. L’Excellent, L Giraud, and H. Ben Hadj Ali (2007). “3D finite-difference frequency-domain modelling of visco-acoustic wave propagation using a massively parallel direct solver: a feasibility study”. In: *Geophysics* 72.5, SM195–SM211.

Plessix, R. E. and C. Perkins (2010). “Full waveform inversion of a deep water ocean bottom seismometer data set”. In: *First Break* 28, pp. 71–78.

Pratt, R. G. (1999). “Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model.” In: *Geophysics* 64.3, pp. 888–901.

— (2003). “Waveform tomography: theory and practice”. In: 12th *International workshop on controlled source seismology.*

Pratt, R. G., C. Shin, and G. J. Hicks (1998). “Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion”. In: *Geophysical Journal International* 133, pp. 341–362. DOI: `10.1046/j.1365-246X.1998.00498.x.`.

Pratt, R. G. and R. M. Shipp (1999). “Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data”. In: *Geophysics* 64, pp. 902–914.

Pratt, R. G. and M. Worthington (1990). “Inverse theory applied to multi-source cross-hole tomography”. In: *Geophysical Prospecting* 38, pp. 287–310.

Purcell, P. G. and R. R. Purcell (1988). “The North West Shelf, Australia - An Introduction”. In: *The North West Shelf, Australia: Proceedings of the Petroleum Exploration Society of Australia, Perth*, pp. 3–15.

— (1994). “The Sedimentary Basins of Western Australia”. In: *The North West Shelf, Australia: Proceedings of the Petroleum Exploration Society of Australia, Perth*, pp. 1–15.

— (1998). “The Sedimentary Basins of Western Australia 2”. In: *The North West Shelf, Australia: Proceedings of the Petroleum Exploration Society of Australia, Perth*, pp. 1–15.

Regone, C. J. (2006). “A modeling approach to wide-azimuth design for subsalt imaging”. In: *The Leading Edge* 25, pp. 1467–1475. DOI: `https://doi.org/10.1190/1.2405331`.

Robinson, E. A. and S. Treitel (1967). “Principles of digital wiener filtering”. In: *Geophysical prospecting* 15.3, pp. 311–332.

Ruth, P., S. Lazaratos, A. Baumstein, S. Lee, Y. Cha, I. Chikichev, N. Downey, D. Hinkley, and J. Anderson (2011). “Encoded simultaneous source full-wavefield inversion for spectrally shaped marine streamer data.” In: *SEG Technical Program Expanded Abstracts 2011*, pp. 2433–2438.

Ruth, P. et al. (2017). “Impact of high-resolution FWI in the Western Black Sea: Revealing overburden and reservoir complexity”. In: *The Leading Edge* 36.1, pp. 60–66. DOI: `https://doi.org/10.1190/tle36010060.1`.

Santosa, F. (1984). *Inverse problems of acoustic and elastic waves*. Siam.

Schuster, G. T. (1996). “Resolution limits for crosswell migration and traveltime tomography”. In: *Geophysical Journal International* 127, pp. 427–440.

Sears, T., S. Singh, and P. Barton (2008). “Elastic full waveform inversion of multicomponent OBC seismic data”. In: *Geophysical Prospecting* 56, pp. 843–862.

Shah, N., M. Warner, T. Nangoo, A. Umpbley, I. Stekl, J. Morgan, and L. Guasch (2012). “Quality assured full-waveform inversion: Ensuring starting model adequacy”. In: *SEG Technical Program Expanded Abstracts 2012*, S13.5.

Shin, C. and Y. Cha (2008). “Waveform inversion in the laplace domain”. In: *Geophysical Journal International* 173, pp. 922–931.

Shin, C., S. Jang, and D-J. Min (2001). “Improved amplitude preservation for prestack depth migration by inverse scattering theory”. In: *Geophysical Prospecting* 49, pp. 592–606.

Shipp, R. and S. Singh (2002). “Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data”. In: *Geophysical Journal International* 151, pp. 325–344.

Silverton, A. (2015). *Applied 3D Full-Waveform inversion: increasing the resolution and depth penetration*. Ph.D. thesis, Imperial College London.

Silverton, A., M. Warner, A. Umpbley, J. Morgan, and K. Irabor (2014). “Non-physical Water Density as a Proxy to Improve Data Fit during Acoustic FWI”. In: *76th EAGE Conference & Exhibition 2014, Extended Abstract.* DOI: `10.3997/2214-4609.20141126`.

Silverton, A., M. Warner, J. Morgan, and A. Umpbley (2015). “Offset-variable density improves acoustic full-waveform inversion: A shallow marine case study”. In: *Geophysical prospecting* 65.5, pp. 1201–1214. DOI: `10.1111/1365-2478.12336`.

Sirgue, L. (2006). “The importance of low frequency and large offset in waveform inversion”. In: *68th EAGE Conference & Exhibition 2006, Expanded Abstract.*

Sirgue, L., J. T. Etgen, and U. Albertin (2008). “3D frequency domain waveform inversion using time-domain finite difference methods”. In: *70th EAGE Conference & Exhibition 2008, Expanded Abstract.*

Sirgue, L., J. Etgen, and U. Albertin (Jan. 2007). “3D Full Waveform Inversion: Wide versus Narrow Azimuth Acquisitions”. In: *SEG Technical Program Expanded Abstracts 2007*, pp. 1760–1764.

Sirgue, L. and R.G. Pratt (Jan. 2004). “Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies”. In: *Geophysics* 69.1, pp. 231–248.

Sirgue, L., O. Barkved, J. Gestel, O. Askim, and J. Kommedal (2009). “2D waveform inversion on Valhall wide-azimuth OBS”. In: *71th EAGE Conference & Exhibition 2009, Expanded Abstract.*

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommendal (2010). “Full waveform inversion: The next leap forward in imaging at Valhall”. In: *First Break* 28, pp. 65–70.

Smith, W. (1974). “A nonreflecting plane boundary for wave propagation problems”. In: *Journal of Computational Physics* 15, pp. 492–503.

Stekl, I., M Warner, A. Umpleby, J. Morgan, C. Pain, and Y. Wang (2008). “Practical 3D wavefield tomography on field datasets”. In: *70th EAGE Conference & Exhibition 2008, Extended Abstract.*

Tamannai, M S.., T. Hansen, R. Borsato, J. Greenhalgh, M. R. Moussavou, and L. O. Essongue (2013). “Pre- and Post-Salt Prospectivity Analysis of the South Gabon Basin”. In: *75th EAGE Conference & Exhibition 2013, Extended Abstract.*

Tang, Y., S. Lee, A. Baumstein, and D Hinkley (Aug. 2013). “Tomographically enhanced full wavefield inversion”. In: *SEG Technical Program Expanded Abstracts 2013*, pp. 1037–1041.

Tarantola, A. (1984b). “Linearized inversion of reflection seismic data”. In: *Geophysical prospecting* 32, pp. 998–1015.

— (Aug. 1984). “Inversion of seismic reflection data in the acoustic approximation”. In: *Geophysics* 49.8, pp. 1259–1266.

— (2005). *Inverse problem theory: Methods for data fitting and model parameter estimation.*

Thomsen, L. (1986). “Weak elastic anisotropy”. In: *Geophysics* 51.10, pp. 1954–1966.

Tikhonov, A.N. and V.Y. Arsenin (1977). *Solution of ill-posed problems.* V.H. Winston and Sons.

Umpleby, A., M. Warner, and I. Stekl (2010). “Time vs frequency for 3D wavefield tomography”. In: *72th EAGE Conference & Exhibition 2010, Extended Abstract.*

Urick, R. J. (1979). *Sound propagation in the sea.* Technical report, Catholic University of America, Washington DC, Dept. of Mechanical Engineering.

Vigh, D. and E. Starr (2008). “Comparisons for waveform inversion, time or frequency domain?” In: *SEG Technical Program Expanded Abstracts 2008*, pp. 1890–1894.

Virieux, J. (1986). “P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method”. In: *Geophysics* 51, p. 889.

Virieux, J. and S. Operto (Nov. 2009). “An overview of full-waveform inversion in exploration geophysics”. In: *Geophysics* 74.6, WCC1–WCC26.

Wang, C., Y. Yingst, P. Farmer, I. Jones, G. Martin, and J. Leveille (2016). “Full Waveform Inversion with the Reconstructed Wavefield Method”. In: *SEG Technical Program Expanded Abstracts 2016*, pp. 1237–1241. DOI: `10.1190/segam2016-13870317.1`.

Wang, C., P. Farmer, Y. Yingst, I. Jones, J. Brittan, and C. Calderon (Oct. 2018a). “Extended parameter full waveform inversion: increasing depth and resolution”. In: *SEG International Exposition and 88th Annual Meeting Workshop.*

Wang, C., P. Farmer, Y. Yingst, I. Jones, J. Brittan, C. Calderon, and J. Leveille (June 2018b). “Traveltime based reflection full waveform inversion”. In: *80th EAGE Conference and Exhibition 2018 Workshop.*

Wang, C., P. Farmer, D. Yingst, I. Jones, C. Calderon, and J. Brittan (2019a). “Preconditioned Reflection Full Waveform Inversion for Subsalt Imaging”. In: *81th EAGE Conference & Exhibition 2019, Extended Abstract.* DOI: `10.3997/2214-4609.201900868`.

Wang, C., Y. Yingst, P. Farmer, I. Jones, G. Martin, and J. Leveille (2019b). *Reconstructed full waveform inversion with the extended source.* URL: `https://www.iongeo.com/virtuals/ResourceArchives/content/documents/Resource%5C%20Center/Technical%5C%20Papers/TP_SEG_Wang_RFWI_171001.pdf` (visited on 08/22/2019).

Wang, Y. (2004). “Q analysis on reflection seismic data.” In: *Geophysical Research Letters* 31.17, pp. 11–14.

— (2008). “Inverse-Q filtered migration”. In: *Geophysics* 73.1, S1.

— (2009). *Seismic inverse Q filtering.* John Wiley&Sons Ltd.

— (2014). “Stable Q analysis on vertical seismic profiling data.” In: *Geophysical Research Letters* 79.4, pp. D217–D225.

Warner, M. (2018). “When FWI goes wrong”. In: *80th EAGE Conference and Exhibition 2018 Workshop.*

Warner, M. and L. Guash (2014a). “Adaptive waveform inversion”. In: *SEG Technical Program Expanded Abstracts 2014*, pp. 1089–1093.

— (2014b). “Adaptive waveform inversion: Theory”. In: *76th EAGE Conference & Exhibition 2014, Extended Abstract.* DOI: `10.3997/2214-4609.20141092`.

— (2015). “Robust adaptive waveform inversion”. In: *SEG Technical Program Expanded Abstracts 2015*, pp. 1059–1063.

— (Nov. 2016). “Adaptive waveform inversion: Theory”. In: *Geophysics* 81.6, R429–R445.

Warner, M., I. Stekl, and A. Umpleby (2007). “Full waveform seismic tomography-iterative forward modelling in 3D”. In: *69th EAGE Conference & Exhibition 2007, Extended Abstract.*

— (2008). “Efficient and effective 3D wavefield tomography”. In: *70th EAGE Conference & Exhibition 2008, Extended Abstract.*

Warner, M. et al. (Mar. 2013). “Anisotropic 3D full-waveform inversion”. In: *Geophysics* 78.2, R59–R80.

Western Australia’s Petroleum and Geothermal Explorer’s Guide (2014). Petroleum Division and Geological Survey of Western Australia.

Williamson, P. R. (Feb. 1991). “A guide to the limits of resolution imposed by scattering in ray tomography”. In: *Geophysics* 56.2, pp. 202–207.

Williamson, P. R. and M. Worthington (1993). “Resolution limits in ray tomography due to wave behaviour: numerical experiments”. In: *Geophysics* 58, pp. 202–208.

Xin, K. F., B. Hung, S. Birdus, and J. Sun (2008). “3-D tomographic amplitude inversion for compensating amplitude attenuation in the overburden”. In: *70th EAGE Conference & Exhibition 2008, Extended Abstract.*

Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambare (2012). “Full waveform inversion for reflected seismic data”. In: *74th EAGE Conference & Exhibition 2012, Extended Abstract.*

Yao, G. and M. Warner (2004). “Seismic waveform tomography in the frequency-space domain: selection of the optimal temporal frequency for inversion”. In: *Exploration Geophysics* 35, pp. 19–24.

— (2014a). “Bootstrapped Waveform Inversion: Long-wavelength Velocities from Pure Reflection Data”. In: *77th EAGE Conference & Exhibition 2015, Extended Abstract.* DOI: `10.3997/2214-4609.201413208`.

Yao, G., M. Warner, and A. Silverton (2014b). “Reflection FWI for both Reflectivity and Background Velocity”. In: *76th EAGE Conference & Exhibition 2014, Extended Abstract.* DOI: `10.3997/2214-4609.20141089`.

Yao, G. and D. Wu (Oct. 2017). “Reflection full waveform inversion”. In: *Science China Earth Sciences* 60.10, pp. 1783–1794. DOI: `10.1007/s11430-016-9091-9`.

Yilmaz, O. (2001). *Seismic data analysis.* Vol. 1. Society of Exploration Geophysicists Tulsa.

Zhdanov, M.S. (2002). *Geophysical inverse theory and regularization problems*. Elsevier.