Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Geoscience and Petroleum

Enrico Prandi

Full Waveform Inversion of Wide-Angle Ocean Bottom Seismometer data of a synthetic subduction zone model

Master's thesis in Petroleum Geophysics
Supervisor: Børge Arntsen
Co-supervisor: Umedzhon Kakhkhorov
July 2021

NTNU
Norwegian University of
Science and Technology

Enrico Prandi

Full Waveform Inversion of Wide-Angle Ocean Bottom Seismometer data of a synthetic subduction zone model

Master's thesis in Petroleum Geophysics
Supervisor: Børge Arntsen
Co-supervisor: Umedzhon Kakhkhorov
July 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Geoscience and Petroleum

NTNU
Norwegian University of
Science and Technology

Acknowledgments

This thesis was written at the Department of Geoscience and Petroleum (IGP) at the Norwegian University of Science and Technology (NTNU) in the spring of 2021.

I would like to thank my supervisor, Professor Børge Arnsten and my colleague Umed Kakhkhorov for outstanding guidance and support throughout the project.

Last but not least i would like to thank myself, for not giving up despite the many hardship faced.

Abstract

The water, which penetrated the upper part of the crust gets dragged down in a subduction zone, consecutively it gets released due to increasing heat and migrates in the above tectonic plate. This is believed to be the main cause for tremors, earthquakes and volcanism. Until now these regions were investigated mostly using tomography or full waveform inversion (fwi) on dense data gathers. A possible way to increase our knowledge of the water cycle in those subduction zones is to apply full waveform inversion even on sparse data. The goal of this thesis is to show that low velocity zones are better imaged using fwi and that sparse data gathers can still produce usable high quality results.
Synthetic data is generated to simulate a subduction zone with water infiltration and an integer subduction zone without any modification. The models are inspired by a study carried on in the Japan Kuril trench in 2013. By comparing simple first arrival tomography with multiple full waveform inversion (on dense, medium and sparse data) outputs, we show that a sensible increase in quality of the results is obtained by running fwi. We also highlight how a sparse data gather is enough to obtain a sound increase in precision and sharpness compared to only running tomography. Considering the computational cost and the complexity of fwi we discussed how the medium spacing data gather should be optimal for processing a seismic survey and obtaining high definition results.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Dehydration of subducting tectonic plates and consequent water migration into the upper laying mantle wedge is believed to be one of the controlling factor for tremors, volcanism and generation of earthquakes (Fujie, S. Kodaira, Yamashita et al. 2013). Although oceanic plates acquire water through various mechanisms and the water cycle inside it is still poorly understood, a study by Ranero et al. 2003 on the subducting pacific plate in central America, states that bending related faulting of the incoming plate is responsible for the hydration of the crust down to uppermantle depths. Assuming faulting to be the greatest contributor of water injection into the plate, the questions coming exactly after are, how much water is absorbed and how deep the faults roots reach, a 2017 study in the Lesser Antilles (Paulatto and Laigle 2017) concluded that water can reach a depth of 140 km.

The deep depths of the interest zone make it impossible yet to obtain samples through drilling, meaning that only seismic waves from active or passive sources can be exploited to gain information on the study zone. Tomography and Rayleigh wave analysis were used by Cai et al. 2018 who studied the water input into the Mariana subduction zone using both ocean bottom seismometers (OBS) and earthquakes data.
Fujie, S. Kodaira, Yamashita et al. 2013 also applied tomography to OBS data to map velocity anomalies, in the Kuril trench subduction zone, correlating it with water content.
S. E. Kodaira et al. 2002 mapped the Nankai Trough using 2-D seismic reflection data and obtained a model through forward modeling and traveltime inversion.

Tomography is widely used in these type of studies, however an approach based on Full Waveform Inversion (FWI) could generate higher resolution and higher quality models, that could help us in understanding the water cycle in the subduction zones. FWI has already been tested on dense (1km spacing) OBS data by Operto 2006 and later Górszczyk et al. 2017 reprocessed the same seismic line proving the higher yields of FWI compared to tomography.
Amoux et al. 2017 obtained high quality data applying FWI to sparse (6km spacing) 3D OBS gathers in the study of heat flux on the Juan de fuca plate.

The study of Fujie, S. Kodaira, Kaiho et al. 2018 based on Vp/Vs ratios after tomo-

graphy was able to conclude that water infiltration in the subducting plate could reach a 3 %wt corresponding of a 15-20 % degree of rock serpentinization.

The scope of this paper is to provide a solid foundation towards the application of FWI on sparse (6km spacing) 2D seismic line gathers, in order to do so a set of sensibility tests will be performed on a synthetic data set modelled after the Fujie, S. Kodaira, Kaiho et al. 2018 study of the Kuril subduction zone. We seek to process a synthetic version of the Kuril trench seismic line using at first FAT (first arrival tomography) and later FWI (Full Waveform Inversion) on different OBS spacing (1-3-6km). We aim to check if denser data gather could be justified by a higher quality result and in addition we will also check how much improvement FWI will yield relative to tomography. We expect that fwi, even on sparse data, will be able to detect low velocity zones, (5% decrease) caused by water infiltration, better than tomography. Hopefully enough this study will provide the theoretical tools that will help in imaging and then understanding the water cycle in the subduction zones.

Chapter 2

Crustal Hydration

2.1 Tectonic Settings

In geology, subduction is meant to be the process in which the oceanic lithosphere is recycled into the Earth's mantle, this happens when two or more converging plates meet. The difference in density, between the oceaninc lithosphere of the converging plates determines which one will subduct, as the heavier plate will dive beneath the second plate and sink into the mantle dragged by his own weight. Figure 2.1 gives a general description on how subduction zones are composed. Often the term subduction zone is used interchangeably with convergent (destructive) plate margin or island arc, and while all three terms are intimately related, they are not synonymous. Subduction zones are the three-dimensional manifestation of convective downwelling, convergent plate margins are the superficial manifestations of downwelling, and arcs (better referred to as arc-trench complexes) are superficial and crustal manifestations of a subduction zone that is operating beneath it (Stern 2002). In addition to playing the central role in Earth tectonics, melt generation, and crustal evolution, subduction zones profoundly impact society to our benefit and detriment. Although most of this planet's ore deposits formed as "specialty" distillates of subduction zones, earthquakes and violent eruptions associated with subduction zones can cause widespread and unpredictable death and destruction.

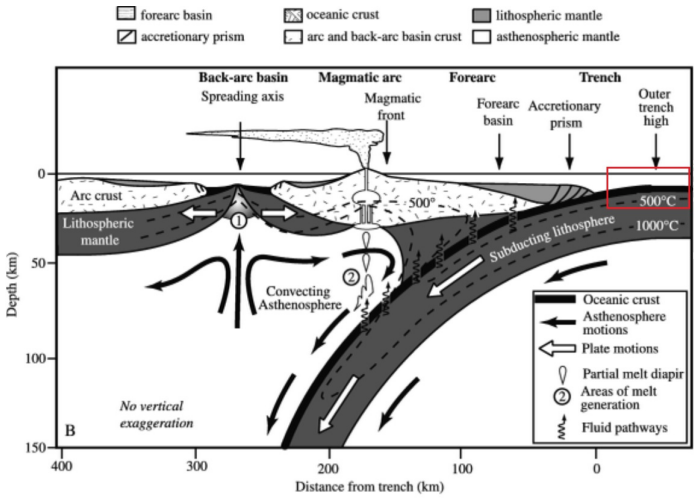


Figure 2.1: Subduction zone components from Stern 2002, in red the zone considered in this study

In Figure 2.1 we can see how deep the subducting lithosphere goes, this means that mostly geophysical data is used to study these phenomenons. Seismic tomography for example is one of the most used methods, it consists in exploiting seismic velocity information contained in many ray paths, that crisscross the Earth between various points near Earth's surface and down to different depths in its interior, in order to produce a model of relative velocity. Regions of anomalously fast mantle corresponding to the earthquake plane define the downgoing lithospheric slab and the regions of anomalously slow mantle lie above. Subducted lithosphere sometimes can be traced this way past the 660 km discontinuity into the lower mantle (Stern 2002).

2.2 Water effects on Rock Physics

The incoming plate begins to bend when it is forced to subduct, this physical process causes a series of fractures called "bend related faulting". Water can penetrate into the cracks and faults present at subduction zones and cause hydration of the lower crust and upper mantle, those hydrated rocks are later dragged down with the downgoing plate. The water will subsequently be released by increasing pressure and temperature conditions. The consequent dehydration reactions at depth affects the composition of the mantle wedge, triggers partial melting and affects subduction zone seismicity. Partially serpentinized peridotite may be a significant reservoir for water in the subducted plate, the mantle wedge and the overriding plate.
Being able to determine the amount of water present could prove to be helpful in understanding the water cycle in the subduction zones, and the direct consequences linked to these processes.

Different early investigations already found that serpentinites have significantly lower compressional wave velocities than unaltered peridotites and dunites. A study done in 1966 (N. Christensen 1966) measured densities, compressional (Vp) and shear (Vs)

waves velocity for a suite of ultramafic rocks ranging from relatively pure lizardite-chrysotile serpentinites to unaltered peridotite. Velocities were measured at confining pressures of atmospheric to 1 GPa and in multiple directions to account for possible anisotropy. The conclusions were that a linear relation exists between the velocity of seismic waves and degree of serpentinization. A More recent study (N. L. Christensen 2004) found that a major factor affecting these velocities is the variety of serpentine present. Antigorite, the serpentine species stable at high temperatures, has higher velocities than the serpentine polymorphs lizardite and chrysotile. Accessory minerals could also play an important role in defining the physics proprieties of the rock. For the sake of optimization in our study we consider serpentinized zones, to be made by lizardite, as the only form of water bearing formations plus we exclude accessory minerals.

The linear relation between water present inside the rocks and wave velocity was also the idea behind a study (Carlson and Miller 2003) which developed a model that relates the degree of serpentinization and water content of partially serpentinized peridotites to their seismic P-wave velocities (figure 2.2).

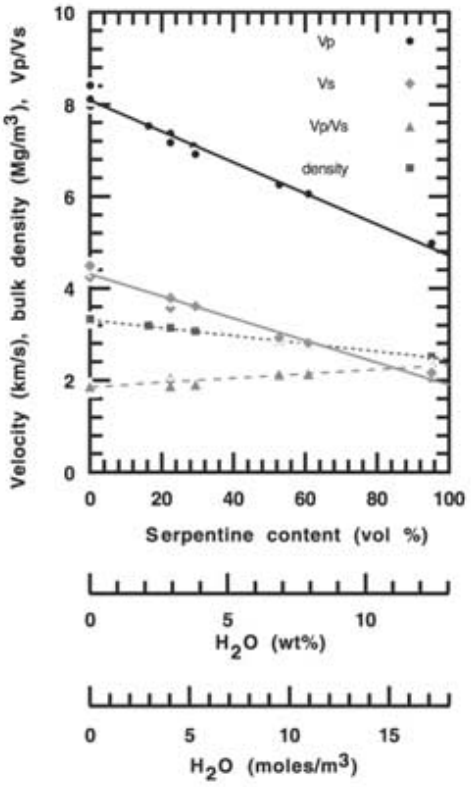


Figure 2.2: Estimated variation of seismic velocities, Vp/Vs ratio and bulk density at 1000 MPa and 400 C. (from Carlson and Miller 2003)

Chapter 3

Theory

3.1 Tomography Theory

Seismic Tomography is a method widely used in probing Earth internal structures, it consists in generating images of zones of complex, heterogeneous velocity and/or attenuation structure of the Earth by using seismic waves. While different types of data (body and shear waves, waveform amplitude, etc) can be exploited, the most used method consists in building images by using spatial distribution of seismic wave velocities determined using travel time data. Tomography aims to minimize the difference in velocity between the real observed traveltimes and the synthetic seismograms (equation 3.1).

$$\begin{aligned} &\mathbf{s} = s_0 + \delta k \\ &\mathbf{s} = \text{slowness from the true model (1/v)} \\ &s_0 = \text{slowness computed from the tomography model} \\ &\delta k = \text{difference between true model and tomography model} \end{aligned} \tag{3.1}$$

It's important to say that δk is not obtained by subtraction, but it is the least square difference (norm of the difference) between the real and the synthetic model summed over all (receivers, sources, ...). The least square difference aims to minimize, with help of conjugate gradients or other descent methods, the sum of the squares of the residuals made in the results of every single equation. By rewriting equation 3.1 as a sum, we get equation 3.2, the matrix A is defined as an overdetermined "n (rows) by m (columns)" with $m < n$. We can solve equation 3.2 by minimizing the norm of the

error (δk) as in equation (3.3).

$$s = \sum_{i=1}^{m} a_i s_{0i}$$

$$\begin{aligned} &\text{s} = \text{measured values} \\ &a_i = \text{column vectors of matrix A} \\ &s_{0i} = \text{fitting value} \end{aligned} \tag{3.2}$$

$$|| \delta k|| = ||s - As_0||^2 \tag{3.3}$$

To solve equation 3.3 and indirectly solving equation 3.1 the main steps, summarized by Q. Liu and Gu 2012, are:

1. Writing the problem based on a set of discrete model coefficients.
2. Computing the predicted data based on the choice of model parameters for an a priori structure, the majority being known 1D model structures.
3. Defining an objective function and adjusting the model parameters to meet the pre-defined goodness-of-fit criteria.
4. Estimating the accuracy and resolution of the inversion outcome, repeating the above steps when necessary.

The first two points are responsible for the forward problem, which makes informed predictions of the observation based on laws of physic. The last two points aim to recover and refine the physical parameters, by minimizing the differences between observations and predictions. This process could be repeated for continued refinement of the model and objective function.

3.1.1 First Arrival Tomography

As the name suggest first arrival tomography specifically utilize only the first wave that reaches the receiver. Since we are working with short and long offsets we use both direct, refracted and diving waves. In figure 4.7 we can clearly see the different wavepaths. The muting function removes everything outside the first arrival signals and is necessary to avoid multiples and unwanted noise that could mess with the algorithm during the processing of the data.

3.1.2 The travel time inversion

Ray-based traveltime tomography has been largely used due to its simplicity. The forward problem involving the computation of the traveltime t of a ray can be expressed

as:

$$t = \int_{x_s}^{x_r} s \cdot dx \tag{3.4}$$

Where dx is the segment of the ray path s is the slowness vector and x_r and x_s are the spatial coordinates of receivers and sources. By tracing a given ray between the two end points we can use Snell's law in order to solve equation 3.4. The main objective of the inverse problem is to determine the fractional change of velocity for all solvable segments along the ray path (Q. Liu and Gu 2012).

3.1.3 Linearization

Basis functions are later adapted to equation 3.2 and we expand the slowness perturbation as a set of linear equations with discrete unknown weights. This parameterization is very simple and results in the natural extension of the discrete form of traveltime (eq. 3.5).

$$\delta t_i = \sum_{j=1}^{N_p} \delta s_j l_{ij} \tag{3.5}$$

where δti is the traveltime difference between observation and reference model prediction for ray i , δs_j is the slowness perturbation of the j-th block in the medium, and l_{ij} is the length of the i-th ray in the j-th block. The summation can be performed over the entire study region, though only a small number of blocks (Np) for a given source–station pair are sampled and contribute to the traveltime residual (Aki and Lee 1976, Ho-Liu et al. 1989). The objective of the inverse problem is to determine δs_j for the majority, if not all, of the blocks through inversions (same way as equation 3.3). Uniform grid is the most obvious choice due to it's simplistic assumption of straight line ray paths. More complicated approaches exists and they implement properties of orthogonality and smoothness made spherical harmonics, in order to include geometrical spreading, latitude, longitude and depth dependant velocity.

3.2 Full Waveform Inversion Theory

Full waveform inversion (FWI) is one of the newest contribution in our investigation into Earth's structure. Even if the theory was developed in 1984, based on the groundbreaking work of A.Tarantola (Tarantola 1984), the fact that it requires a high computational cost and is very sensible to interferences (noise, source calibration, ground-receiver coupling) made it hard to use it (Operto 2006). Recent advances in processing machines and optimization of algorithms speed made possible to utilize this technology to investigate Earth's interior. FWI is mainly composed of two parts, the first part consist of solving the elastic wave equation numerically and the second of solving the full waveform inverse problem though non-linear optimization.

3.2.1 Numerical solution of the Wave equation

Equation of motion

The first part of FWI is to model the wave propagation in heterogeneous media, to do so we solve a forward problem by simulating a seismic wave propagating though the Earth and computing the resulting synthetic seismograms. The solution of the elastic wave (equation 3.6) allows us to know what happens to the wave at every time (t) and and every position (x).

$$\begin{aligned} &\rho(x)\ddot{u}(x,t) - \nabla\sigma(x,t) = f(x,t) \\ &\\ &\rho = \text{density} \\ &\ddot{u} = \text{acceleration} \\ &\sigma = \text{stress tensor} \\ &\text{f} = \text{external force density} \end{aligned} \tag{3.6}$$

In order to solve this equation the stress tensor σ must be related to the displacement filed u, assuming that σ depends linearly on the history of the strain tensor, we get the definition (equation 3.7) of the linear visco-elastic rheology.

$$\begin{aligned} &\sigma(x,t) = \int_{-\infty}^{\infty} \dot{C}(x,t-t') : \xi(x,t')dx \\ &\\ &\dot{C} = 4^{th} \text{ order elastic tensor} \\ &\xi = \frac{1}{2}(\nabla u + \nabla u^T) \end{aligned} \tag{3.7}$$

The symmetric of ξ makes possible to write equation 3.7 directly as:

$$\sigma(x,t) = \int_{-\infty}^{\infty} \dot{C}(x,t-t') : \nabla u(x,t')dx \tag{3.8}$$

The general simplicity of the Earth at macroscopic scale allows us to describe the elastic moduli (equation 3.9) using only 2 of the original 21 components of $\dot{C}$, those 2 parameters are called Lamé parameters (λ and μ).

$$C_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu\delta_{ik}\delta_{jl} + \mu\delta_{il}\delta_{jk} \tag{3.9}$$

The time dependence of the elastic tensor is responsible for the visco-elastic dissipation, part of the pressure transmitted by the wave is converted into heat. For the sake of optimizing computing velocity we consider our model to be non dissipative, this turns equation 3.8 into:

$$\sigma(x,t) = C(x) : \nabla u(x,t) \tag{3.10}$$

Reinserting equation 3.10 into equation 3.6 we obtain the so called *displacement formulation* of the elastic wave:

$$\rho(x)\ddot{u}(x,t) - \nabla[C(x) : \nabla u(x,t)] = f(x,t) \tag{3.11}$$

Another relevant form, especially in numerical modelling is the *velocity-stress formulation*, where the equation taken as a first order system of equations in both space and time.

$$\begin{cases} \rho(x)\dot{v}(x,t) - \nabla\sigma(x,t) = f(x,t) \\ \dot{\sigma} - C(x) : \nabla v(x,t) = 0 \end{cases} . \tag{3.12}$$

Discretization in Space

Equations 3.11 and 3.12 are applicable in continuous spaces, analytical solutions only exists for comparatively simple models that will not be representative in our study. The continuous wave field u(x,t) is transformed into his discrete spacial approximation $\bar{u}(t)$. The discretization of space turns the elastic wave equation 3.11 into:

$$M\ddot{\bar{u}}(t) + K\bar{u}(t) = \bar{f}(t)$$

$$\begin{aligned} &M = \text{mass matrix} \\ &K = \text{stiffness matrix} \\ &\bar{f}(t) = \text{vector form of discrete force density f} \end{aligned} \tag{3.13}$$

Both M and K tend to be sparse, the mass and stiffness matrices contains information about the structure, for example K contains the information about the stiffness of the material, it contains the derivatives of Lame parameters with respect to the gradient of the displacement field u(x,t) ($\nabla[C(x) : \nabla u(x,t)]$ from equation 3.11) while M contains the density parameters ($\rho(x)$). By using equation 3.13 the forward problem is reduced to solving a large algebraic system of ordinary differential equation in time. The time dependant ODE that arise form the discretisation of space can be solved in both frequency or time domain, depending on user preferences, in our project we used time-domain modelling.

The discrete displacement formulation (equation 3.13) contains the second order derivative of the displacement field $\ddot{\bar{u}}$ which is explicitly given by:

$$\ddot{\bar{u}} \approx M^{-1} * [\bar{f}(t) - K * \bar{u}(t)] \tag{3.14}$$

Expanding $\ddot{\bar{u}}$ into a truncated Taylor series we obtain the approximated second-order central finite difference:

$$\ddot{\bar{u}} \approx \frac{1}{\Delta t^2}[\bar{u}(t + \Delta t) - 2\bar{u}(t) + \bar{u}(t - \Delta t)] \tag{3.15}$$

Depending on when you decide to truncate the Taylor series you obtain higher precision approximation, in this example the second order central finite difference scheme holds a precision of O (h^2).

Rearranging equation 3.15 leads to the explicit time stepping scheme (equation 3.16) that enable us to compute the displacement at future time ($t+\Delta t$) from displacements at past ($t-\Delta t$) and present times (t):

$$\bar{u}(t+\Delta t) \approx 2\bar{u}(t) - \bar{u}(t-\Delta t) + \Delta t^2 M^{-1} * [\bar{f}(t) - K\bar{u}(t)] \tag{3.16}$$

The Finite Difference above described is just one method to numerically solve the wave equation, other techniques could include spectral element, finite volume or numerical element methods. In general FD remains the most used as it's a good compromise between accuracy, computational cost, simplicity of implementation and ability to describe heterogeneous models (Operto 2006).

Stability Condition

Multiple variants of equation 3.16 exists, for example the *Newark scheme* (Newmark 1959) or the *leap-frog method* which is a staggered grid method. The common link between all the explicit different schemes is the stability criteria, since the stress at $t+\Delta t$ depends only on the stress field prior to $t+\Delta t$, it implies that algorithms are only conditionally stable. The stability criterion, the CFL condition, named after (Courant et al. 1967), takes the form of:

$$\Delta t \leq CFL * \frac{\Delta x}{c}$$

$$\begin{aligned} &\Delta t = \text{time sampling interval} \\ &\text{CFL} = \text{depends on method used, normally magnitude of 1} \\ &\Delta x = \text{minimum grid size} \\ &\text{c} = \text{propagation speed of fastest P wave} \end{aligned} \tag{3.17}$$

3.2.2 Full Waveform inverse problem

The synthetic seismograms computed with the forward simulation are a product of an Earth model **D**, comparing them with the real observed seismograms allows us to physically quantify the difference between them (equation 3.18). This approach is the same as tomography but instead of considering only velocity, **D** and D_0 can contain multiple parameters of our choice (velocity, amplitude, density, ...). Mathematically this difference is defined by a misfit functionl χ, usually it depends non-linearly on the Earth model (**D**).

$$\mathbf{D} = D_0 + \chi_0$$

$$\begin{aligned} &\mathbf{D} = \text{real observed seismograms} \qquad \mathbf{D(x)} = [m_1(x), m_2(x), m_3(x), ...] \\ &D_0 = \text{synthetic seismograms computed} \\ &\chi_0 = \text{misfit functional} \end{aligned} \tag{3.18}$$

Same as for tomography χ is not obtained by subtraction in equation 3.18 but is the sum over everything of the least square difference between the real observed data **D** and the synthetic data D_0.
Solving the full waveform inverse problem means to find and optimal Earth model D_{opt} so that $\chi(D_{opt})$ is the global minimum of χ. In this sense FWI is just a special non linear optimization problem, due to the large amount of model parameters (thousands to millions).

$$D_{i+1} = D_i + \gamma_i h_i \qquad with \qquad \chi(D_{i+1}) < \chi(D_i) \tag{3.19}$$

The optimization (equation 3.19) proceeds by iterations with the help of gradient-based descends, such as Newton methods, Conjugate gradients, Adjoint methods Steepest descend or else. All those gradient minimisation algorithms critically rely on the derivative of the misfit functional with respect to the model parameters, they compute the *descent direction*, h_i and the *step length* $\gamma_i > 0$, which are dependant on the approximation scheme used to approximate D_{opt}.

3.2.3 The L-BFGS algorithm

In this study to optimize the model misfit, for both Tomography and FWI, we utilized a limited memory Broyden – Fletcher – Goldfarb – Shanno (L-BFGS) algorithm which belongs to the quasi-Newton methods family (D. C. Liu and Nocedal 1989). L-BFGS uses an estimate of the inverse Hessian matrix to guide its search through variable space, it stores only a few vectors that represent the approximation implicitly. Due to its resulting linear memory requirement, the L-BFGS method is particularly well suited for optimization problems with many variables. The algorithm follows the same principle as his "brother" BFGS, but instead of storing every value separately only saves his last updated overwriting the previous one (Li and Fukushima 2001).
To understand the algorithm itself we first introduce some notation: the iterates are in the form of m_i and we define $s_i = m_{i+1} - m_i$ and $y_i = g_{i+1} - g_i$.
The method uses the inverse BFGS formula:

$$H_{i+1} = V_i^T H_i V_i + \rho_i s_i s_i^T \tag{3.20}$$

where $\rho_i = I/y_i^T s_i$ and so

$$V_i = I - \rho_i y_i s_i^T \tag{3.21}$$

More specifically the L-BFGS follows these step (D. C. Liu and Nocedal 1989):

1. Choose a $m, 0 < \beta' < \frac{1}{2}, \beta' < \beta < 1$ and a symmetric and positive definite starting matrix H_0. Set i=0.
2. Compute the descent direction

$$\begin{aligned} h_i &= -H_i g_i \\ m_{i+1} &= m_i + h_i \gamma_i \end{aligned} \tag{3.22}$$

3. Let $\tilde{m} = min(i, m-1)$, this way we update the matrix H_0, $\tilde{m}+1$ times using (s_i and y_i).

$$\begin{aligned} H_{i+1} =&(V_i^T ... V_{i-\tilde{m}}^T) H_0 (V_{i-\tilde{m}} ... V_i) \\ &+ \rho_{i-\tilde{m}} (V_i^T ... V_{i-\tilde{m}+1}^T) s_{i-\tilde{m}} s_{i-\tilde{m}}^T (V_{i-\tilde{m}+1} ... V_i) \\ &+ \rho_{i-\tilde{m}} (V_i^T ... V_{i-\tilde{m}+2}^T) s_{i-\tilde{m}} s_{i-\tilde{m}}^T (V_{i-\tilde{m}+2} ... V_i) \\ &+ ... \\ &+ \rho_i s_i s_i^T \end{aligned} \tag{3.23}$$

4. Set i=i+1 and repeat step 2.

Wolfe-type inexact line search

The line search of the algorithm was coupled with the Wolfe minimizing conditions. These conditions states that to find a descent direction h_i (needed for equation 3.22), the algorithm needs to reduce the objective function "sufficiently", rather than minimizing the objective function over $h \in R^+$ exactly (D. C. Liu and Nocedal 1989). The idea is to find the minimum for some smooth function $f : R^n \to R$ that satisfies the conditions:

$$\begin{aligned} &f(m_i + \gamma_i h_i) \leq f(m_i) + \beta' \alpha_i g_i h_i \\ &g(m_i + \gamma_i h_i)^T h_i \geq \beta g_i^T h_i \end{aligned} \tag{3.24}$$

equation 3.24 is solved every new iteration and h_i is inserted into equation 3.22.

3.2.4 Local and Global Minima

The high computational cost of the forward problem solution makes impractical to implement inverse probabilistic schemes. We are therefore limited to the deterministic solution of the inverse problem, meaning that we wish to find the Earth model $\widetilde{D_0} \in$ D (where D is the space that contains all admissible Earth models) such that the misfit functional $\chi(D)$ is minimal. If in the neighbourhood $N_r(\widetilde{D_0})$ around our minimum, there exists at least another point where $\chi(\widetilde{D_0})$ is smaller or equal to any $\chi(D_0)$ that is:

$$\chi(\widetilde{D_0}) \leq \chi(D_0), \qquad \text{for all} \quad D_0 \in N_r(\widetilde{D_0}) \tag{3.25}$$

We call that point a *local minimum*. If the case is that $\chi(\widetilde{D_0})$ is strictly smaller than any $\chi(D_0)$ in the neighbouring points N_r:

$$\chi(\widetilde{D_0}) < \chi(D_0), \qquad \text{for all} \quad D_0 \in N_r(\widetilde{D_0}) \tag{3.26}$$

We call $\widetilde{D_0}$ a *strict local minimum*.
A misfit functional might have multiple local minima and some may produce larger misfits than others. The best case is $\chi(\widetilde{D_0})$ to be a *global minimum*, in that case $\chi(\widetilde{D_0})$ will be smaller or equal to $\chi(D_0)$ for any $D_0 \in$ D and not only for neighbouring points.

$$\chi(\widetilde{D_0}) \leq \chi(D_0), \qquad \text{for all} \quad D_0 \in D \tag{3.27}$$

and similarly a *strictly global minimum* will be defined as:

$$\chi(\widetilde{D_0}) < \chi(D_0), \qquad \text{for all} \quad D_0 \in D \tag{3.28}$$

Figure 3.1 exemplify the different types of minima.

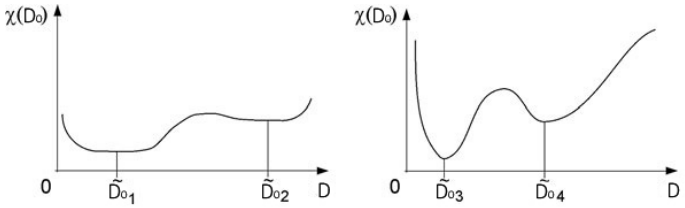


Figure 3.1: Left: Global minimum $\widetilde{D_{01}}$ and local minimum $\widetilde{D_{02}}$. Right: Strict global minimum $\widetilde{D_{03}}$ and strict local minimum $\widetilde{D_{04}}$. From Fichtner 2011

The distinction between a minimum and a strict minimum is subtle but important because it directly relates to the uniqueness or non-uniqueness of the optimal model. In the case of a strict global minimum, the model $\widetilde{D_0}$ describes the data better than any other model in M, and the solution to the inverse problem is said to be unique. When the minimum is not strict, then we can find many models that describe the data equally well, meaning that the solution to the inverse problem is non-unique. Using this terminology, we can say that we are interested in finding the strict global minimum of χ, if it exists. Despite being extremely important, the existence of a strict global minimum has received very little attention in the geophysical literature; and mathematics provides few practical tools that allow us to address this issue in realistic applications. Part of the problem is that χ is not given explicitly, meaning that each evaluation of $\chi(\widetilde{D_0})$ for a specific model m requires the solution of a computationally intense forward problem. The existence of a global minimum – strict or not – is often conjectural, and the conjecture is based on the assumption that the physics of the problem are modelled sufficiently well (Fichtner 2011).

3.2.5 Why FWI over Tomography

Pratt (Pratt et al. 1996) proved on synthetic data that FWI is superior to First Arrival Tomography (FAT), and recent studies (Operto 2006 and Górszczyk et al. 2017) confirmed it on real data too. A research carried out on black smokers (Amoux et al. 2017) concluded that FWI has an improvement on spatial resolution up to four fold the FAT approach. Even if FWI has higher resolution and produces higher quality models is far from perfect. The first problem, as discussed before, is a common problem for optimization problems, we don't know if the minimum reached, is the global minimum or a local minimum, as the gradient responsible for the descent direction of the misfit functional χ consider only neighbouring points of fixed radius (Fichtner 2011). In order to avoid running into these problems, a multi-scale approach can be implemented (as by Operto 2006 and Dessa et al. 2004) and an optimal starting model is necessary (Amoux et al. 2017, Operto 2006, DeGrande and Mochizuki 2020) even better if it's smoothed (Morgan et al. 2016, DeGrande and Mochizuki 2020).

The starting velocity model needs to predict the traveltime of any arrival involved in the inversion to within half a period of the lowest inverted frequency (equation 3.29 needs to be respected) when a classical misfit functional based on the data difference is used, otherwise we risk converging to a local minimum (Virieux and Operto 2009).

$$\frac{\Delta t}{T} < \frac{1}{2N_\lambda}$$

$$\begin{aligned} &\Delta t = \text{Traveltime error} \\ &T = \text{Period} \\ &N_\lambda = \text{Number of propagated wavelenghts} \end{aligned} \tag{3.29}$$

3.2.6 Mulstiscale approach

Data-driven strategies are conventionally designed to take the edge of the non-linearity of FWI. In time domain this takes the form of windowing. We implemented the multiscale approach for both tomography and FWI by matching first short offset traveltimes (20km) and by increasing the data inclusion (window size was 10km) every iteration. We stopped after 4 iterations, as our goal was to mimic real life data and offsets larger than 50km are usually discarded because they become littered by noise. This method proved to be effective, the feeding of the whole datasets caused the failure in minimizing the misfit and could not produce a single iteration.

Chapter 4

Modelling

4.1 Programs

In order to complete this study we relied on a multitude of programs. Matlab was used to build the initial models, successively Madagascar, an open source script based program for seismic processing, was used to generate wavelets, smooth the models and generate muting and weighting filters. The bulk of the work, synthetic datasets, control surveys, tomography and full waveform inversion was done using another script based program called "rockseis" developed by Wiktor Waldemar Weibull and used at NTNU. Other minor adjustments to the scripts were done using "csim", the plotting was done with a combination of Python and Matlab.
To face the massive computing power necessary to run multiple FWI's we depended on the supercomputer "Betzy", part of the Uniett Sigma2 project. Betzy is a Bull-Sequana XH2000, provided by Atos, with a theoretical peak performance of 6.2 Peta-Flops is ranked as the 56th fastest cluster in the world. The supercomputer is placed at NTNU in Trondheim and has been in production since 24 November 2020. For more specifics about Betzy we refer to the official website (https://www.sigma2.no/systems).

4.2 Modelling Parameters

In order to check if large receiver spacing (6km) is enough to produce meaningful results, we performed a series of sensibility tests using FWI on a set of synthetic data modelled after Fujie, S. Kodaira, Yamashita et al. 2013 and Fujie, S. Kodaira, Kaiho et al. 2018.
The model's area was of 160km length with a 15 km depth, the grid spacing was set at 20m. To solve the forward problem we used a Finite Difference scheme in time domain with a precision of O (h^8). In order to optimize the error produced by the FD scheme itself and the numerical scattering of the so called *staircase effect* we needed at least 3 or more grid points per wavelength for maximum frequency and minimum velocity (Morgan et al. 2016). With a CFL of 0.5 we have a maximum time sample of $\leq$ 0.98 ms. Table 4.1 contains a resume of all parameters used in the modelling.

Grid size Δx	20 meters
Grid size Δy	20 meters
Time step Δt	0.5 ms
Wavelet type	Ricker
Peak frequency	5 hz
Peak time	0.1 s
OBS spacing	1-3-6 km
Source spacing	200 meters
Source depth	10 meters

Table 4.1: Gridding parameters

4.3 Standard Model

The model geometry (as seen in figure 1 in the appendix) was based on bathimetry data and displays an overall flat seafloor dipping towards north (left). The internal layering was also based on figure 1 and consisted of 4 different layers: seafloor sediments, upper basement, lower basement and upper mantle. The synthetic model (figure 4.1) consists of regular grid nodes with 20m spacing. The data population of the models was done so that fixed values were given at layers boundaries, the inside layer data was conditioned by a logarithmic function in order to account for overburden corrections. The model was in the end smoothed to limit the presence of artifacts.

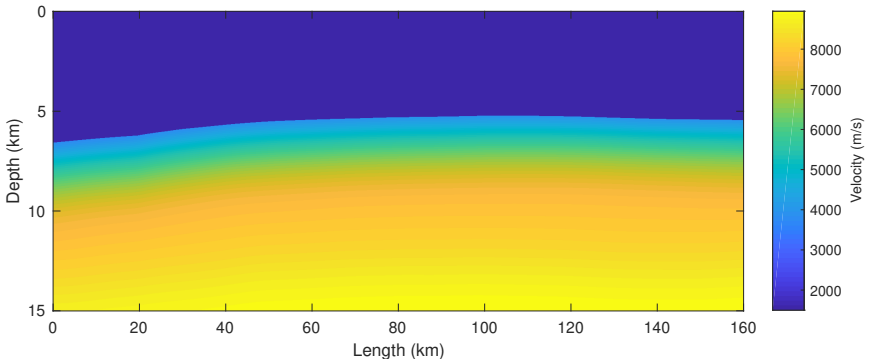


Figure 4.1: Initial Model

The data values for Vp and density were taken by correlating globe-scale surveys (N. Christensen and Salisbury 1975) with local studies (Oikawa et al. 2010 and Shinohara et al. 2008). Figures 2, 3 and 4 highlight how the depth-velocities trends are compared to each other.

4.4 Model with low velocity zones

In order to check if FWI with 6km obs spacing is sensible enough to detect possible water penetration down to a deep of 15km, we built a model with matching geometries with the standard (figure 4.1), and manually inserted low velocity zones. These zones were considered to have a 0.1 - 0.2 degree of serpentinization, with a water content in range of 1-3 of wt%. As explained in the first chapter, a linear relation exists between degree of serpentinization, velocity, density and ratio of pressure over shear wave. Geometry and unaltered zones were kept the same for the two models.
The presence of water affects seismic velocities, in figure 4.3 we see the difference between the standard (figure 4.1) and the model with low velocities (figure 4.2), the presence of serpentinized rocks causes Vp to deacrease by 5% (a maximum of 350 m/s).

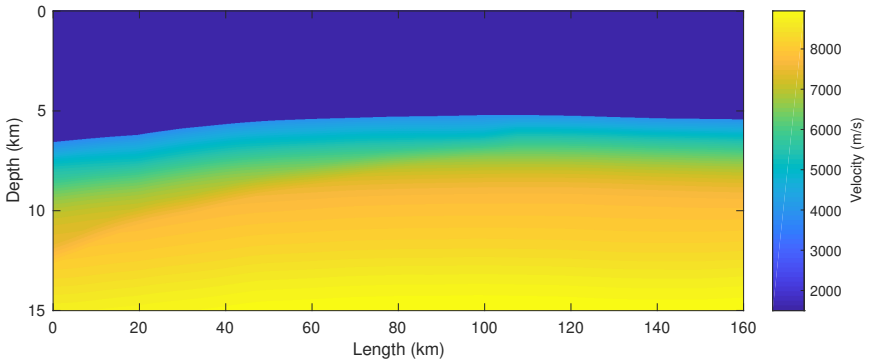


Figure 4.2: Initial Model with water inflitration

4.5 Data quality control

The difference between the starting models is shown in figure 4.3, and the difference in velocity profiles in figure 4.4.

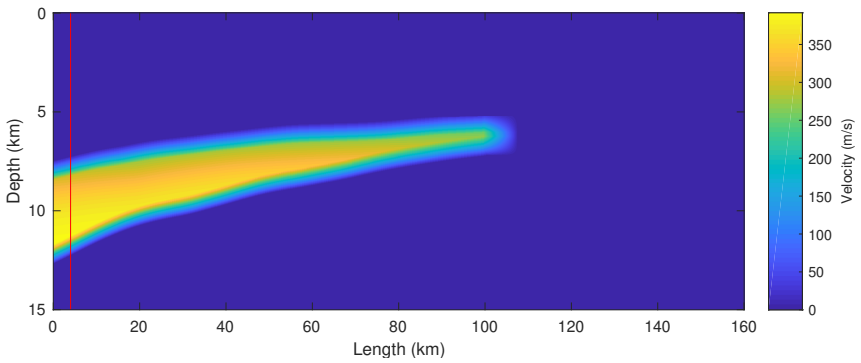


Figure 4.3: Difference between starting models, the red line in where the velocity profiles for figure 4.4 are taken

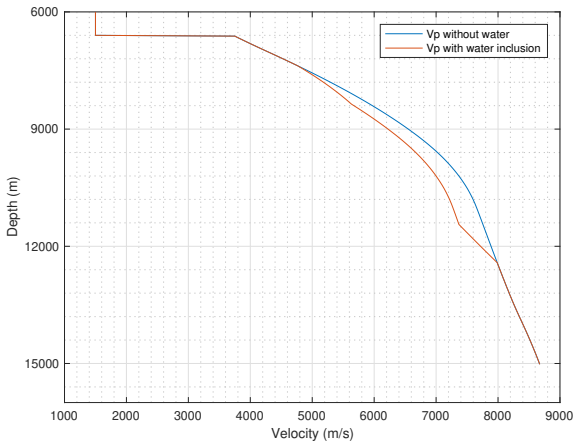


Figure 4.4: Velocity profiles

After generating different models, we built a synthetic dataset by forward propagating a wave and recording the pressure at fixed positions. To optimize computing we had to exploit the reverse principle and invert source and receiver positions, our "virtual" sources were placed on the seafloor and the "virtual" receivers were located at seatop. Figure 4.7 on the right is a single shot gather (common source) but in reality it would be a common receiver gather.

To check the stability of the numerical scheme we plotted a trace belonging the the receiver present at km 24 (figure 4.5). From that it is evident that outside of wave arrivals we do not have disturbances, neither before or after the wavefront. A zoom in onto the direct wave arrival (highest amplitude wave) highlights how the shape of the wavelet (Ricker) is preserved. The difference of the low velocity zones is easily spotted on figure 4.5, as expected the water infiltration present as zones with a 5% lower velocity causes a small delay in the peak of the first arrival wave. The shape of the wavefront is mostly untouched.

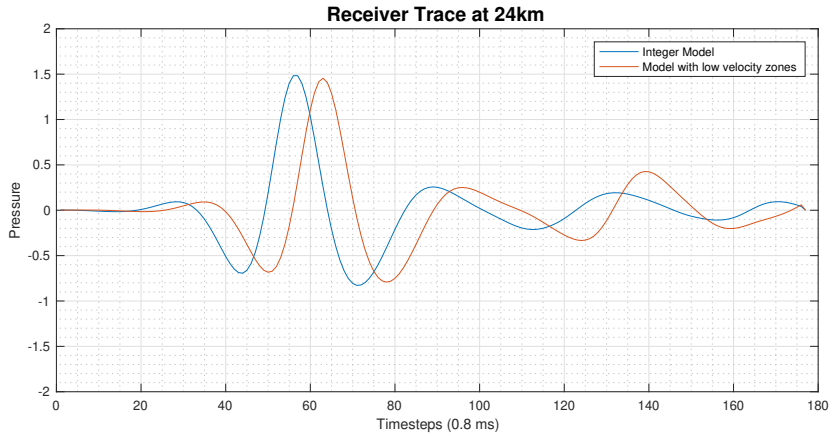


Figure 4.5: Single traces present at a single receiver (the one at 24km) for both starting models.

The input model for FWI was generated using first arrival tomography (FAT), after picking the first arrival waves from the dataset generated from our standard model (figure 4.2) we generated a 1D tomography model by horizontally stacking all the 1D traces (figure 4.6). The 1D Tomographic model output was clearly just a general velocity gradient of the standard model, the upper and lower limit velocities cutoffs and the smoothing applied caused the low velocity zones to be optimized, resulting in comparable models. The fact that both starting models had negligible differences prompted us to only run tomography for the model with low velocity zones. This way we could see if tomography could already detect small velocity anomalies, and in which measure it could be differentiated from FWI. The dataset used to run tomography was from the survey of 6km spacing ocean bottom seismometers.

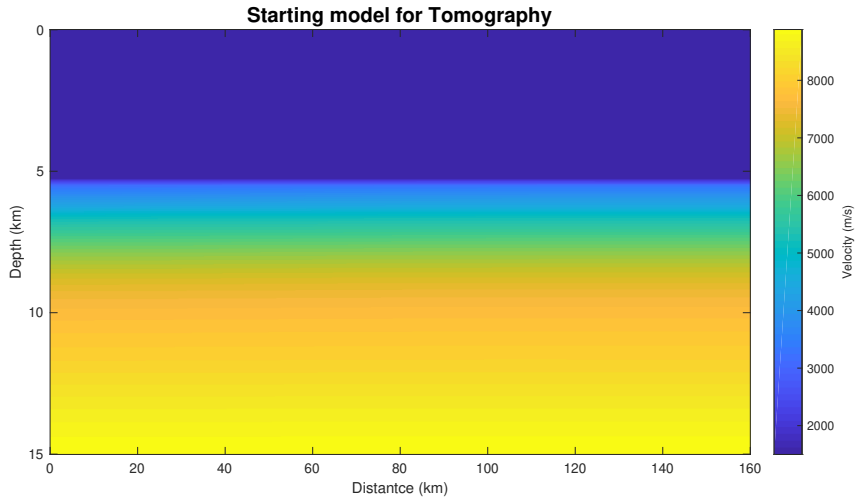


Figure 4.6: Input model before tomography

An important factor when working with tomography and fwi is how well the muting filter fits the data. Figure 4.7 highlights how a single shot containing multiple reflections (right) is muted and only holds first arrival waves (left). As anticipated in chapter 3 the first arrival waves are composed by direct and refracted for short offsets, and by the largest part of diving waves. This phenomenon is clearly visible in figure 4.7 where the different apparent dip changes with trace number.

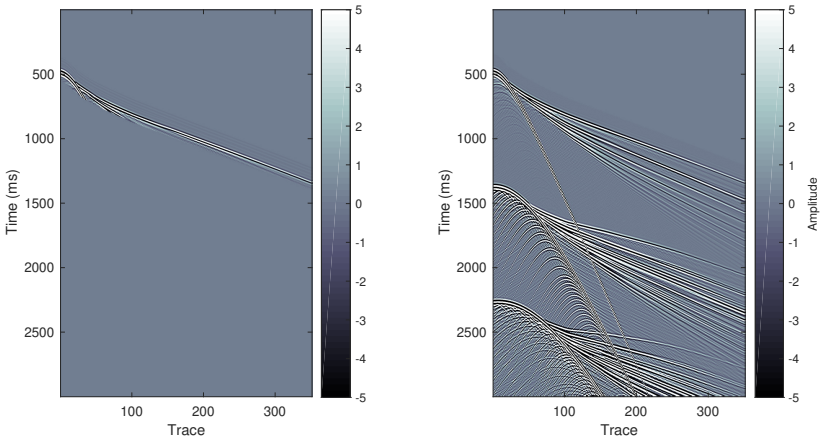


Figure 4.7: shot gather after applying of a muting function (left) and the full shot gather (right)

Chapter 5

Results

The general workflow is described by figure 6. The first step in this process was the building, based on real data analogues, of the two initial models for both density and velocity (figures 4.1 and 4.2). Immediately after we generated multiple synthetic data sets based on those first models, with different ocean bottom seismometers spacing. Those data sets have the shape of pressure recordings, figure 4.7 is an example of a shot generated on the initial model. The successive step, was the generation of the input model for tomography (figure 4.6). The input model for tomography was created based on the stacking of 1D tomography with lower and upper velocity constraints, the model was later smoothed and the result was a generally smoothed velocity gradient. We only ran tomography utilizing the first arrival waves extrapolated from the synthetic data set generated from the initial model containing low velocity anomalies. The resulting model is shown in figure 5. It's important to clarify that we ran tomography using a general velocity model paired with first arrival picks from the 6km spacing obs survey.

The output model after tomography was then checked by simulating a seismic survey, utilizing the same geometries and wavelet from table 4.1. This allowed us to crosscheck it with the same data from the initial model. In order to utilize the output model of tomography for FWI, the first arrival wave's peaks have to be shifted by less than half of the maximum frequency wave period, else the FWI algorithm will not converge (Operto 2006) or it will converge to a local minimum (Virieux and Operto 2009). The first arrival waves for the tomography model were compared with the ones from the true model (figure 5.1). The direct comparison allows us to see how small offsets matches almost perfectly and how on large offsets the first arrivals are separated by less than 50 time steps (40ms, as each time step is 0.8 ms). Since the first arrival's peaks are close, the running of full waveform inversion will not cause convergence problems.

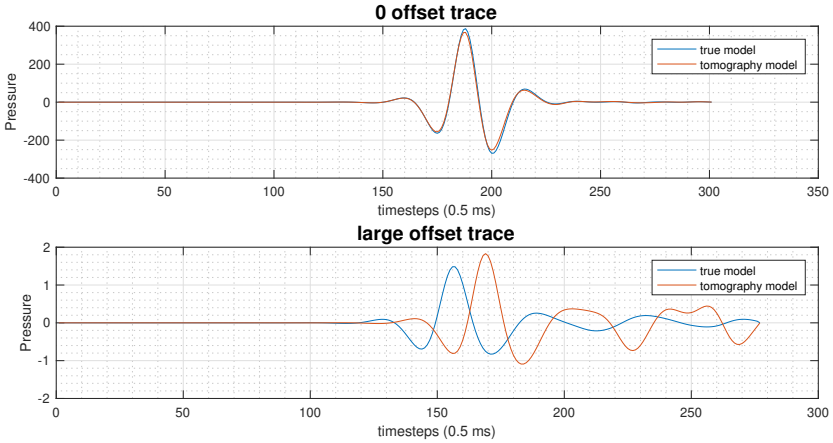


Figure 5.1: Comparison between starting model and model after tomography of 0 and far offset receivers

The tomography model was used as starting velocity model for both the cases, the inversion based on the model with low velocity zones and the inversion on the standard integer model.
The results of FWI were compared with the initial models using four different type of analysis: calculating RMS values, generate synthetic data sets to check first arrival waves, calculating the resulting misfits and directly comparing the results with the true models.

5.1 FWI on the Standard Model

The output model generated with tomography (figure 5) was fed to the FWI algorithm multiple times using different OBS geometries. The first result analysis was to generate a data set to check first arrival waves and compare them with the initial model. Figures 5.2 and 5.3 are traces generated from a single shot with source at seafloor at 24 km and receivers placed every 200 meters at a depth of 10 meters, the maximum traveltime was fixed at 20 seconds. Geometries of source-receivers, wavelet type and intensity were kept the same as the one utilized for the data generated on the initial model. The shortening of the maximum traveltime was to minimize computational cost.

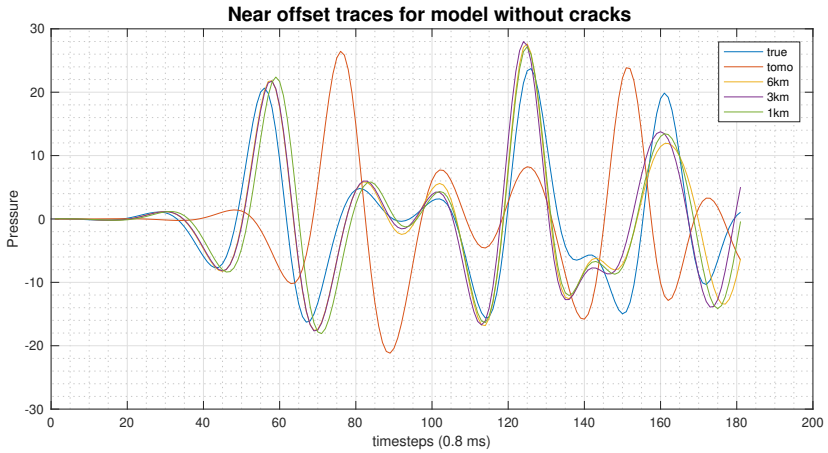


Figure 5.2: Near offset traces comparison between different FWI outputs for the standard model

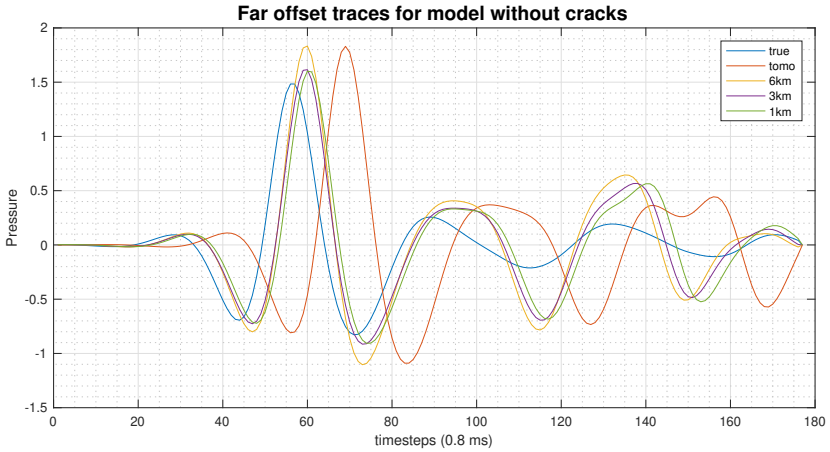


Figure 5.3: Far offset traces comparison between different FWI outputs for the standard model

Another way of analyzing the output data was to directly compare the initial and the final models. Figure 5.12 shows the difference between the standard initial model without low velocity zones and the model obtained through tomography, the same model used for input for the FWI process. The different FWI velocity outputs are figures 5.4, 5.5 and 5.6. To better highlight the differences we also plotted vertical velocities taken along 20 60 and 100km (figures 8, 9 and 10 in the appendix).

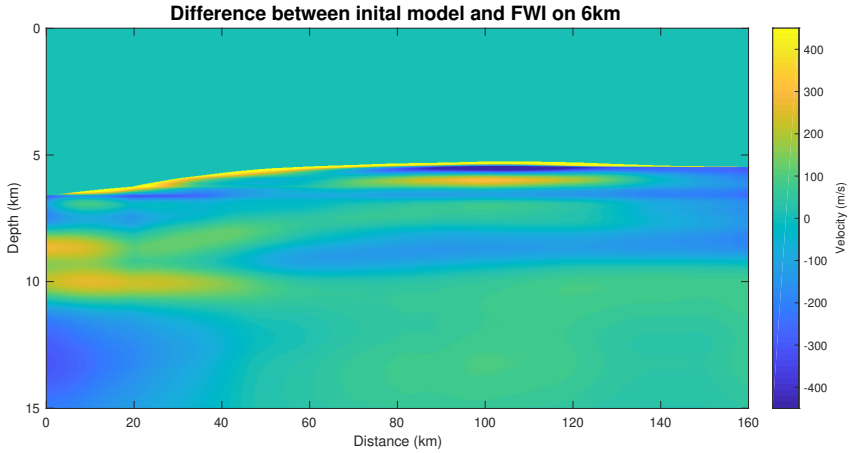


Figure 5.4: Difference between the initial model and the output of FWI using 6km obs spacing

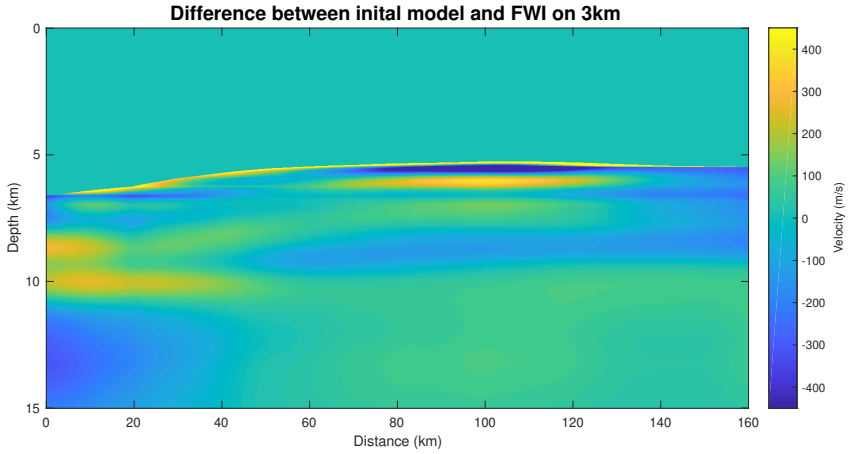


Figure 5.5: Difference between the initial model and the output of FWI using 3km obs spacing

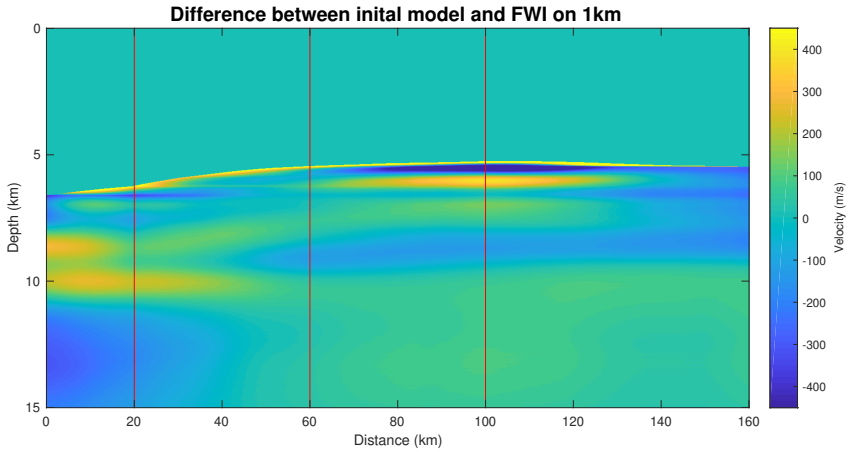


Figure 5.6: Difference between the initial model and the output of FWI using 1km obs spacing, the red lines are where the velocities profiles are taken in figures 8,9 and 10

The third form of result analysis was the computation of the root mean square (RMS) velocity for the output models and then calculating the difference with the initial models. The RMS (root mean square) velocity consider the model as a sequence of parallel horizontal layers of interval velocity V_{int}, in the i-th layer. We label layers from 1 to N, the two-way traveltime for seismic energy propagating perpendicularly through the i-th layer is Δt_i. The RMS velocity at the N-th reflector, for travel perpendicular to the layers is:

$$V_{rms}N = \sqrt{\frac{\sum_{i=1}^{N} V_i^2 \Delta t_i}{\sum_{i=1}^{N} \Delta t_i}} \tag{5.1}$$

Equation 5.1 was used to compute RMS values assuming horizontal planes of thickness of 20m (same as the spacing of the grid points). The difference between RMS values of the standard model and the final outputs is shown in figure 5.7.

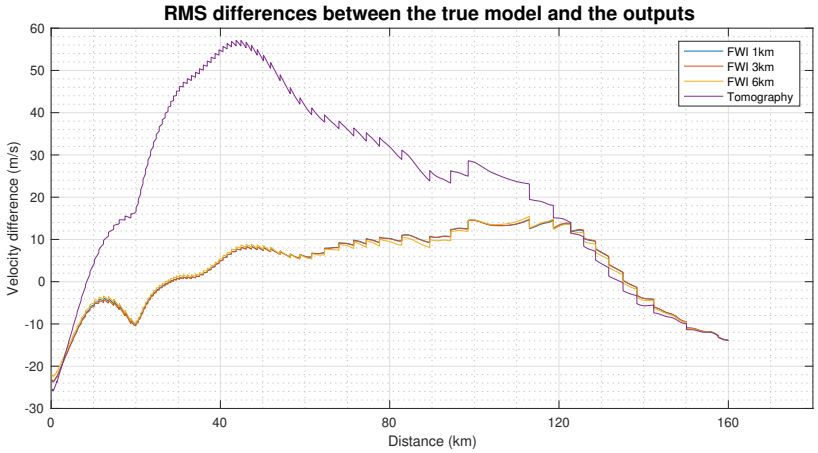


Figure 5.7: Difference between the true model and the outputs

The last type of analysis was done by comparing the misfits. The non-linearity and different data gathers made the direct comparison non representative, instead of directly comparing misfits between the different tests we compared the rate of decrease within the same run. The first value of the misfit was computed when the algorithm performed the first linesearch and was set as 100%. In figure 5.8 we can see that 6km FWI only decreases the misfit by 10-12 % over four iterations while medium and dense grid obs gathers reduce the misfit by more than 60%.

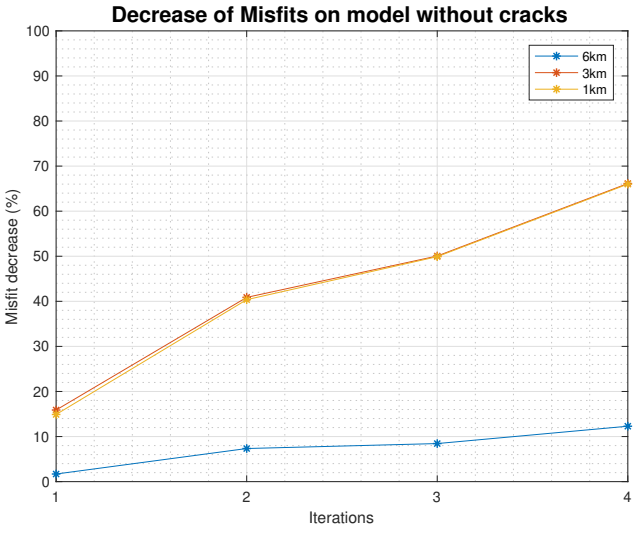


Figure 5.8: Misfit functional decrease in percentage from the first computed value

5.2 Model with low velocity zones

While we were running FWI on the standard model to check the sensibility of receiver spacing on the results, we were running another FWI test in parallel. The test being on, how good a low velocity zone, caused by water infiltration, could be mapped using FWI.

The same comparing methods used for the standard model were utilized for the low velocity analysis. Near (figure 5.9) and far (figure 5.10) offset traces were simulated and plotted with the model containing low velocity zones.

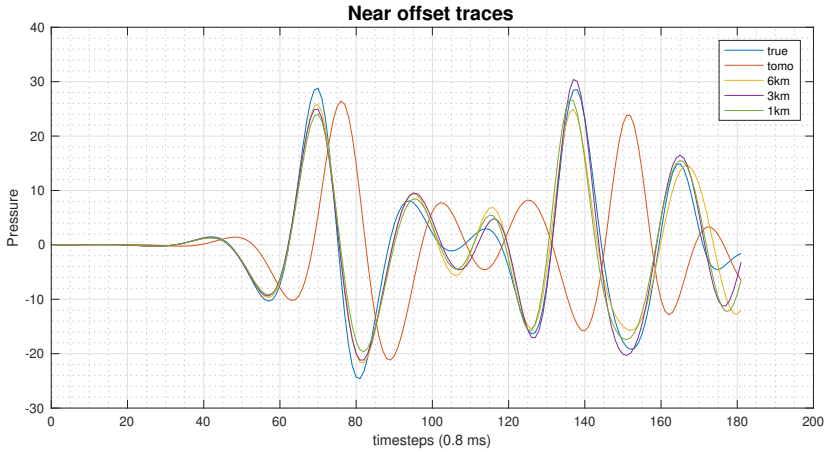


Figure 5.9: Near offset traces comparison between different FWI outputs

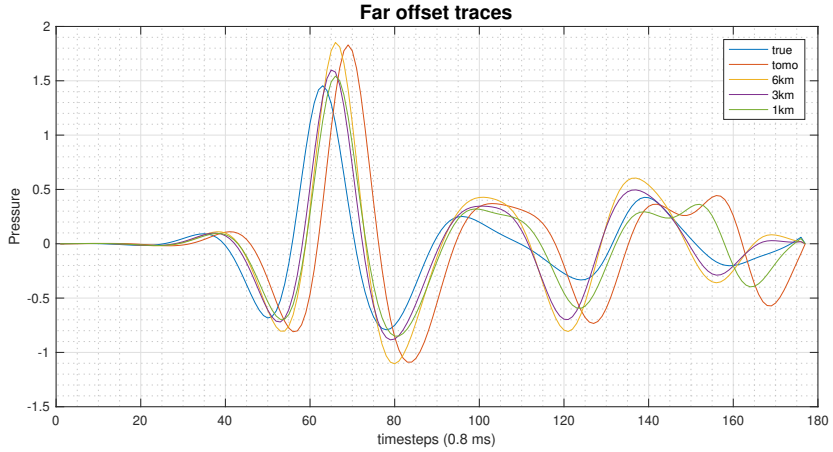


Figure 5.10: Far offset traces comparison between different FWI outputs

The difference between the output models is a very simple yet effective way to check the difference. Figure 5.11 shows how the low velocity should appear, it is the differ-

ence between the initial models. Figure 5.12 shows the difference between the true initial model without low velocity zones and the model obtained through tomography. In comparison figures 5.13, 5.14 and 5.15 highlight how well the low velocity zone is reconstructed by subtracting the initial standard model to the fwi outputs. Vertical velocity profiles were taken in zones of particular interest, results are shown in figures 11, 12 and 13.

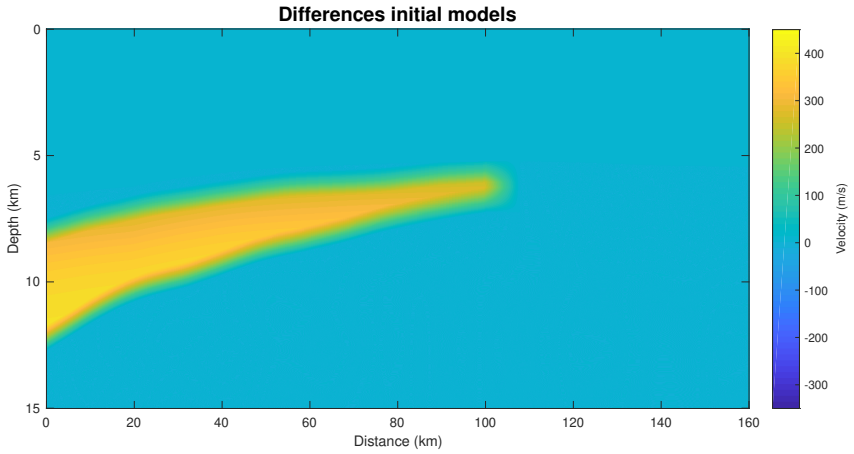


Figure 5.11: Difference between the standard initial model and the modified versions with low velocity zone

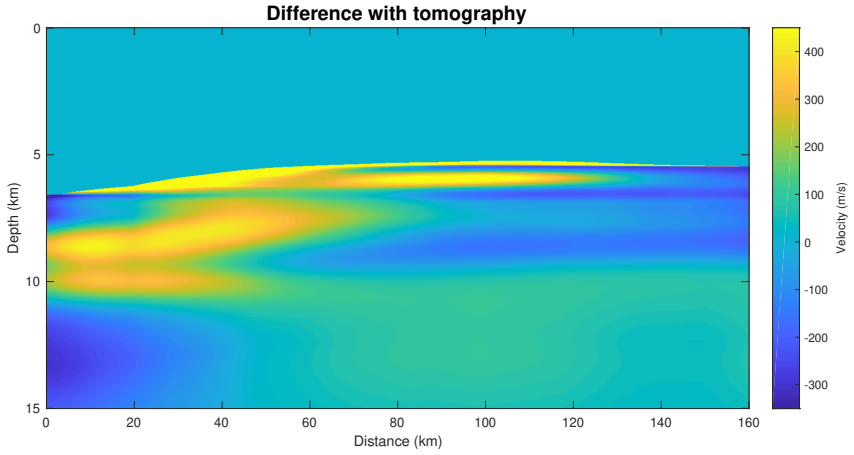


Figure 5.12: Difference between standard model and tomography model

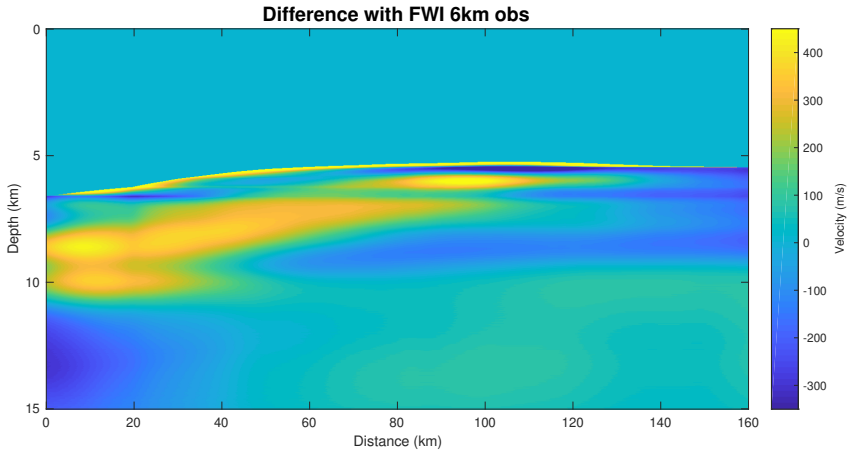


Figure 5.13: Difference between standard model and output model from FWI on 6km obs data

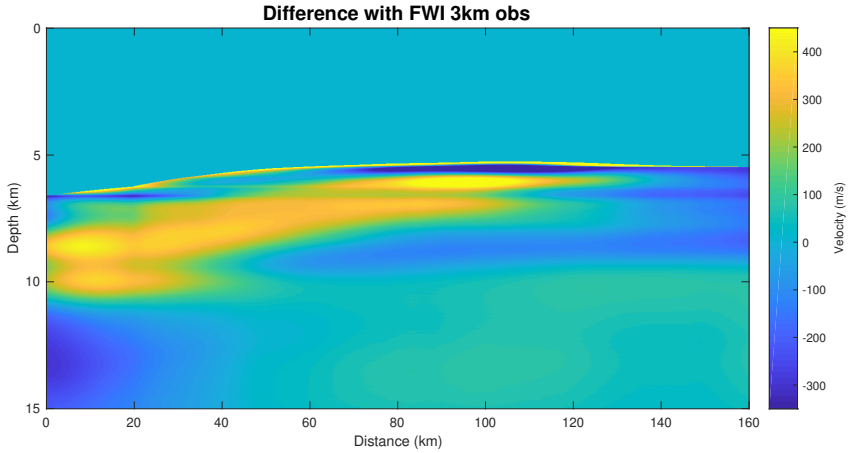


Figure 5.14: Difference between standard model and output model from FWI on 3km obs data

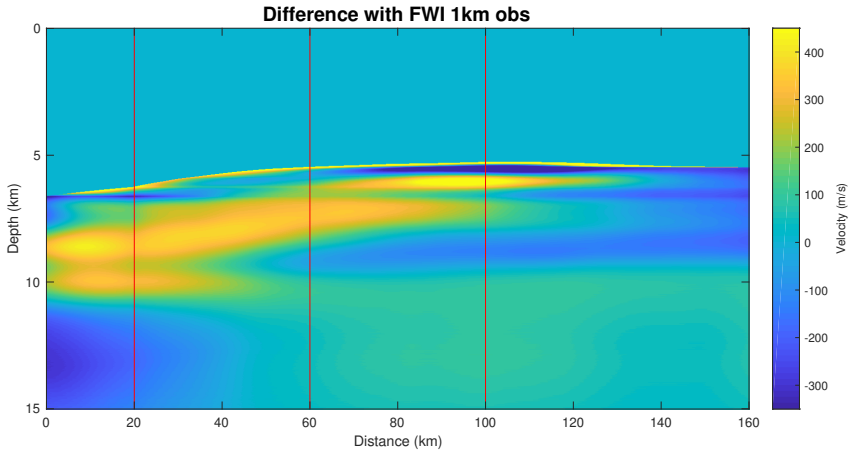


Figure 5.15: Difference between standard model and output model from FWI on 1km obs data, the red lines are where the velocities profiles are taken in figures 11,12,13

As for the standard model RMS values were computed using equation 5.1 and then subtracted from the initial model. Results are shown in figure 5.16.

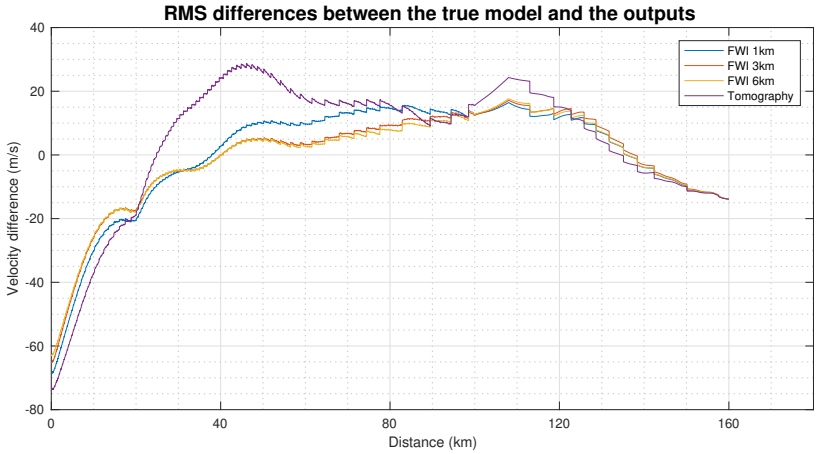


Figure 5.16: Difference between the true model with low velocity zones and the outputs

Misfits percentage decreases are shown in figure 5.17 and it's clear that, same as the integer model, the 6km FWI decreases the misfit value by very little (not even 10%). These misfit functionals decreases in smaller magnitudes compared with the integer model.

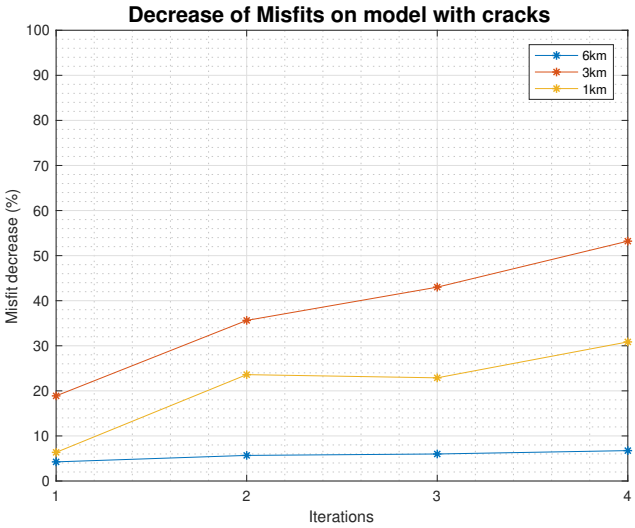


Figure 5.17: Misfit functional decrease in percentage from the first computed value

Chapter 6

Discussion

The objective of Full Waveform inversion is to generate high resolution models. This is done by first solving numerically the elastic wave equation in the discretized space and then by solving the full waveform inversion problem though non-linear optimization. The results are then utilized as input for the next iteration, this guarantees that the misfit functional becomes smaller at every iteration. The multiscale approach, to first match short offset data (from 20km) and increasing the data inclusion (10km) at every iteration, proved to be crucial, as without it the limited memory BFGS algorithm was finding difficulties in minimizing the error, due to the non-linear structure of the problem and the large amount of data. The zero and close to zero offset receivers signal were discarded as in reality those shots only contain water column information and mess with the data linearization problem.
The dual scope expressed at the beginning of this paper was to first check if the spacing of the seismometers placed on the seafloor could play a major role in resulting image quality. At the same time we checked if small velocity changes (5% decrease) such as low velocity zones caused by water infiltration in the subducting plates, could better be detected by FWI algorithms with respect to the mostly used method for this type of studies (tomography).
The low velocity zones are definitely detectable, different result analysis had similar results. The first control was to generate a test shot on the output velocity models and to compare it to the true model with low velocity zones, to only control the velocity the density was kept constant (the same density model was used for all the synthetics). In figure 5.9 we see how the short offset (20 km) receiver registers different traces. FWI model outputs are able to generate traces that almost perfectly match the true model (with slight differences in amplitude), while tomography has a small time delay. On the far offset trace (figure 5.10) FWI reduces the traveltime error compared to tomography but it's still 3-6 ms behind the true model. Overall FWI methods have a good trace fit, amplitude match is best obtained by 1km FWI, closely followed by 3km FWI, while 6km FWI tends to have higher peaks and lower troughs.
To directly compare the results with the initial model we subtracted them, figure 5.11 shows how the shape of the zone should look like, figure 5.12 shows the tomography output and figures 5.13, 5.14 and 5.15 display different FWI outputs. To better investigate the differences we took vertical velocity profiles at 20 , 60 and 100 kilometers,

these profiles are figure 11,12 and 13. This type of analysis clearly showed how FWI methods have better resolution and detection on the low velocity zones compared to tomography, a particular increase in the sharpness of the seafloor is visible in figure 12. Although FWI overall better matches the true model it also generates artifacts of high data variance, those artifacts are visible as abnormally high velocity followed by a sharp decrease into a lower velocity zone (for example the blue and yellow zone at 100km in 5.15). Those zones are clearly highlighted in figure 13 at 5.5km depth, immediately after the seafloor, and are visible in lower degree in figure 11. Manually smoothing those zones in between iteration could increase even more the result quality.
The RMS computation was an ulterior control method, assuming flat layers with the same thickness as the grid point spacing (20 meters) we can compare nominal 1D velocity traveltime and minimize the anomalies by stacking all 750 resulting layers. The best way to visualize (as in figure 5.16 the difference is to subtract it to the initial model and see how close it is to the zero line. Again FWI is proved to have better results compared to tomography, especially for the first 3000 points (60km) where the geometry of the area is irregular, slightly dipping. In the rest of the model all methods are close to zero and are overall equiparable, surprisingly 1km FWI had the worst fit (higher difference) in most of the model compared to 3 and 6 kilometer FWI.

The second objective was to control how much the distance of the receivers could influence the results. The superiority of FWI compared to tomography was already proved by Górszczyk et al. 2017 on data from the Nanaki trough, however he was working with dense data, 1km receiver spacing. Most of ocean bottom seismometers (OBS) surveys use 6km spacing between receivers, if we could prove that 6km is enough to apply FWI and obtain quality results, the whole seismic imaging scene could be revolutionized. The tomography model we utilized as starting model for FWI (figure 5) was also obtained using 6km spacing obs data.
The near and far offset trace comparison between the three FWI models on both models (standard and with low velocity zones) has no showing of a method being far better than the others. In figure 5.2, 1km FWI has a worse overall fit than 3km FWI, with 6km being the worst, but in figure 5.3, 1 and 3km have better amplitude fits than 6km.
From the direct comparison and the velocity profiles it is hard to conclude that a method is superior, true enough 1km FWI has better fits at the seafloor but also generates the stronger artifacts.
RMS values of the outputs of FWI on the standard model have no meaningful difference between them, while in the low velocity zone 3 and 6km FWI have better values compared to 1km FWI, this is related to the artifact generation, as the difference disappears after 100km.
The misfit comparison displays the misfit functional, the difference between the true model and the FWI output model, and how much it decreases with every iteration. In the model without cracks the 6km FWI run is clearly inferior to the other two which are practically overlapped. In the model with low velocity zones 6km FWI is again the worse but 3km is actually the best and shows a 20% higher decrease compared with the 1km run.
Misfit percentage decrease highlights how much the algorithm diminishes the difference between the true model and the computed model. In the model without water 1 and 3km have the same trend and have at least a five time larger decrease than 6km

obs. Surprisingly enough in the model with low velocity zones the 3km run has above 50 % misfit reduction, which is by far greater than the 1km run with a reduction of 30 %. The regular shape of the graph suggest that 3km could steadily continue to decrease the misfit with consequent iterations, while 1km has more an irregular shape and it will be more difficult to predict. The 6km fwi diminishes the misfit constantly but with a small margin, after four iteration the total decrease is less than 10%.
Based on the model with low velocity zone, the 3km FWI has the better optimization of the three runs.

Chapter 7

Conclusions

Studying the effect of low velocity zones shows clear evidence that Full Waveform Inversion (FWI) is able to detect velocity anomalies of just 5%. When we apply tomography in order to image the velocity anomalies, traveltime errors will cause unfocused images especially in the zones of dipping layers and at sharp boundary contrasts. When we further process the tomography output model using FWI we considerably increase the output resolution and greatly improve sharp boundary focusing. Even if we discard density updates like we did, FWI is able to handle sharp velocity changes (like the seafloor) even when paired with irregular geometries. The velocity models obtained with FWI even if they are superior compared to tomography can still be improved. The artifacts, abrupt velocity fluctuations are generated due to the non linear nature of the problem, and a smoothing process after every iteration could improve even more the results.

Furthermore we have shown that FWI does not need dense data gather in order to produce meaningful results. We have tested sparse (6km) medium (3km) and dense (1km) data gathers, and surprisingly the results were for most part comparable. The model generated from dense data gathers, is more similar to the true model in amplitude of the waveform, handles better the velocity jumps and fits the best on vertical velocity profiles. The larger data present however causes problems, during the inversion, the artifacts generated are even stronger and the longer computing time might not make dense data gather processing worthwhile. FWI on sparse data gather is far from unusable, true enough is less precise than the other two but can still provide massive advantages compared to only running tomography. The best compromise between quality of the results, computational cost and complexity of the non-linear inversion problem is the medium spacing (3km) data gather.

In the introduction we highlighted how these low velocity zones caused by water infiltration can influence our lives, our experiment proved that, 3km obs data is the best, but 6km is enough to map these zones. Seismic surveys done with sparse obs gathers can be processed with FWI and produce quality results, and even where tomography was already utilized, a reprocessing of the zone using FWI can provide more details that could help us in better understanding the subduction zone water cycle.

We propose to test this idea on real data and utilize the elastic (or even better visco-elasitc) wave equation, as the limitation of this study was that the synthetic generated data contained only the body wave component and this could prove to be too simple and easy to process compared to real world analogues. A smoothing process could also be included in between different iterations as this could help in canceling out the velocity artifacts.

Bibliography

Aki, K. and W. H. K. Lee (Aug. 1976). 'Determination of three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes: 1. A homogeneous initial model'. In: *Journal of Geophysical Research* 81.23. DOI: https://doi.org/10.1029/JB081i023p04381.

Amoux, G. et al. (2017). 'Seismic evidence that black smoker heat flux is influenced by localized magma replenishment and associated increases in crustal permeability'. In: *Geophysical Research Letters - Agu Pubblications*, pp. 1687–1695.

Cai, C. et al. (Nov. 2018). 'Water input into the Mariana subduction zone estimated from ocean-bottom seismic data'. In: *Nature* 563.7731, pp. 389–392. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0655-4.

Carlson, R. L. and D. J. Miller (2003). 'Mantle wedge water contents estimated from seismic velocities in partially serpentinized peridotites'. In: *Geophysical Research Letters* 30:5. URL: doi:10.1029/2002GL016600.

Christensen, N. (Dec. 1966). 'Elasticity of ultrabasic rocks'. In: *Journal of Geophysical Research* 71.24. DOI: https://doi.org/10.1029/JZ071i024p05921.

Christensen, N. and M. Salisbury (1975). 'Structure and constitution of the lower oceanic crust. IMAGING OF THE EASTERN NANKAI TROUGH'. In: *Reviews of geophysics and space physics* 13.1. DOI: 10.1029/RG013i001p00057.

Christensen, N. L. (2004). 'Serpentinites, Peridotites, and Seismology'. In: *International Geology Review* 46:9, pp. 795–816. URL: doi:10.2747/0020-6814.46.9.795.

Courant, R., K. Friedrichs and H. Lewy (Mar. 1967). 'On the Partial Difference Equations of Mathematical Physics'. In: *IBM Journal of Research and Development* 11.2. DOI: 10.1147/rd.112.0215.

DeGrande, J. and K. Mochizuki (2020). 'Application of full waveform inversion to OBS data acquired in an activesource seismic survey: a synthetic test for the 2020 Bungo-channel profiles'. In: *Earthquake Research Institute, University of Tokyo, Japan, Texas AM University - Corpus Christi.*

Dessa, J. X. et al. (Feb. 2004). 'Multiscale seismic imaging of the eastern Nankai trough by fullwaveform inversion. IMAGING OF THE EASTERN NANKAI TROUGH'. In: *Journal of Geophysical Research: Solid Earth* 109.B2. ISSN: 0148-0227. DOI: 10.1029/2003jb002689.

Fichtner, A. (2011). *Full Seismic Waveform Modelling and Inversion, Advances in Geophysical end Environmental Mechanics and Mathematics*. Springer. URL: doi:10.1007/978-3-642-15807-0_1.

Fujie, G., S. Kodaira, Y. Kaiho et al. (2018). 'Controlling factor of incoming plate hydration at the north-west Pacific margin'. In: *Nature Communications*. URL: doi:10.1038/s41467-018-06320-z.

Fujie, G., S. Kodaira, M. Yamashita et al. (2013). 'Systematic changes in the incoming plate structure at the Kuril trench'. In: *Geophysical Research Letters* 40, pp. 88–93. URL: doi:10.1029/2012GL054340.

Górszczyk, A., S. Operto and M. Malinowski (June 2017). 'Toward a robust workflow for deep crustal imaging by FWI of OBS data:The eastern Nankai Trough revisited. FULL-WAVEFORM INVERSION OF OBS DATA'. In: *Journal of Geophysical Research: Solid Earth* 122.6, pp. 4601–4630. ISSN: 2169-9313. DOI: 10.1002/2016jb013891.

Kodaira, S. E., E. Kurashimo and J. O. Park (June 2002). 'Structural factors controlling the rupture process of a megathrust earthquake at the Nankai Trough seismogenic zone'. In: *Geophysical Journal International* 143.3. DOI: https://doi.org/10.1046/j.1365-246X.2002.01691.x.

Li, D.H. and M. Fukushima (2001). 'On the global convergence of the BFGS method for nonconvex unconstrained optimization problems'. In: *Siam J. Optim* 11.4.

Liu, D. C. and J. Nocedal (1989). 'On the limited memory BFGS Method for large scale optimization'. In: *Mathematical Programming* 45.

Ho-Liu, P., A. L. Bent and D. V. Helmberger (Apr. 1989). 'Waveform modeling of the November 1987 Superstition Hills earthquakes'. In: *Bulletin of the Seismological Society of America* 79.2.

Liu, Q. and Y.J. Gu (2012). 'Seismic imaging: From classical to adjoint tomography'. In: *Tectonophysics* 566-567, pp. 31–66. ISSN: 0040-1951. DOI: https://doi.org/10.1016/j.tecto.2012.07.006. URL: https://www.sciencedirect.com/science/article/pii/S004019511200409X.

Morgan, J. et al. (Jan. 2016). 'Next-generation seismic experiments – II: wide-angle, multi-azimuth, 3-D, full-waveform inversion of sparse field data'. In: *Geophysical Journal International* 204.2. DOI: https://doi.org/10.1093/gji/ggv513.

Newmark, N. (July 1959). 'A method of computation for structural dynamics'. In: *Journal of the Engineering Mechanics Division* 85.3. DOI: https://doi.org/10.1061/JMCEA3.0000098.

Oikawa, M., K. Kaneda and A. Nishizawa (Apr. 2010). 'Seismic structures of the 154–160 Ma oceanic crust and uppermost mantle in the Northwest Pacific Basin'. In: *Earth, Planets and Space* 62.e13–e16. DOI: https://doi.org/10.5047/eps.2010.02.011.

Operto, S. (Feb. 2006). 'Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern Nankai trough. IMAGING OF THE EASTERN NANKAI TROUGH'. In: *Journal of Geophysical Research* 111, B09306. ISSN: 0148-0227. DOI: 10.1029/2003jb002689.

Paulatto, M. and M. Laigle (July 2017). 'Dehydration of subducting slow-spread oceanic lithosphere in the Lesser Antilles'. In: *Nat Commun* 8.15980. DOI: https://doi.org/10.1038/ncomms15980.

Pratt, R. G., Z. M. Song and P. Williamson (Feb. 1996). 'Two-dimensional velocity models from wide-angle seismic data by wavefield inversion'. In: *Geophysical Journal International* 124.2. DOI: https://doi.org/10.1111/j.1365-246X.1996.tb07023.x.

Ranero, C. R. et al. (Sept. 2003). 'Bending-related faulting and mantle serpentinization at the Middle America trench'. In: *Nature* 425. DOI: https://doi.org/10.1029/2001RG000108.

Shinohara, Masanao et al. (2008). 'Upper mantle and crustal seismic structure beneath the Northwestern Pacific Basin using a seafloor borehole broadband seismometer

and ocean bottom seismometers'. In: *Physics of the Earth and Planetary Interiors* 170.1, pp. 95–106. ISSN: 0031-9201. DOI: https://doi.org/10.1016/j.pepi.2008.07.039. URL: https://www.sciencedirect.com/science/article/pii/S003192010800201X.

Stern, R. J. (Dec. 2002). 'Subduction zones'. In: *Reviews of geophysics* 40.4. DOI: https://doi.org/10.1029/2001RG000108.

Tarantola, A. (Dec. 1984). 'Linearized Inversion of seismic reflection data'. In: *Geophysical Prospecting* 32.6. DOI: https://doi.org/10.1111/j.1365-2478.1984.tb00751.x.

Virieux, J. and S. Operto (Nov. 2009). 'An overview of full-waveform inversion in exploration geophysics'. In: *GEOPHYSICS* 74.6. DOI: https://doi.org/10.1190/1.3238367.

Wolfe, P. (Apr. 1969). 'Converge conditions for Ascent methods'. In: *SIAM review* 11.2.

— (Apr. 1971). 'Converge conditions for Ascent methods II, some corrections'. In: *SIAM review* 13.2.

Appendix

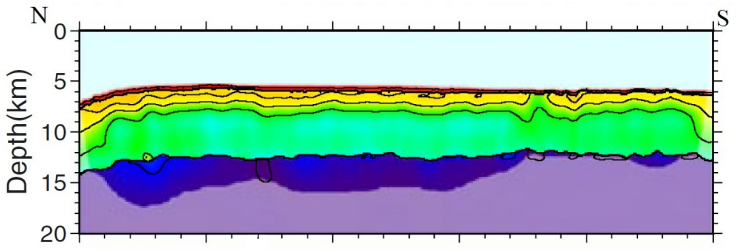


Figure 1: Model from Fujie, S. Kodaira, Kaiho et al. 2018

	Thickness	**Vp**	**Vs**	**Vp/Vs**	**Rho**	**K**
Layer 1	5.5 km	1.5 km/s	0 km/s	-	1	-
Layer 2 top	$\leq$ 0.5 km	1.6 km/s	0.15 km/s	10	1.6	4
Layer 2 bottom		2.6 km/s	1.2 km/s	2.1	1.8	9
Layer 3 top	2 km	5.5 km/s	3 km/s	1.84	2.55	48
Layer 3 bottom		7.1 km/s	3.8 km/s	1.85	2.7	88
Layer 4 top	6 km	7.4 km/s	4.2 km/s	1.76	2.85	94
Layer 4 bottom		7.9 km/s	4.5 km/s	1.75	3.0	112
Layer 5 top	6.5 km	8.4 km/s	4.85 km/s	1.73	3.3	137
Layer 5 bottom		9.1 km/s	5.4 km/s	1.68	3.55	166

Table 1: Fixed Model parameters

	Thickness	**Vp**	**Vs**	**Vp/Vs**	**Rho**	**K**
Layer 1	5.5 km	1.5 km/s	0 km/s	-	1	-
Layer 2 top	$\leq$ 0.5 km	1.6 km/s	0.15 km/s	10	1.6	4
Layer 2 bottom		2.6 km/s	1.2 km/s	2.1	1.8	9
Layer 3 top	2 km	5.225 km/s	2.79 km/s	1.87	2.42	42
Layer 3 bottom		6.745 km/s	3.53 km/s	1.9	2.56	77
Layer 4 top	6 km	7.03 km/s	3.9 km/s	1.799	2.7	83
Layer 4 bottom		7.45 km/s	4.18 km/s	1.78	2.9	99
Layer 5 top	6.5 km	8 km/s	4.51 km/s	1.76	3.135	121
Layer 5 bottom		8.6 km/s	5.02 km/s	1.72	3.37	147

Table 2: Cracked Rocks parameters

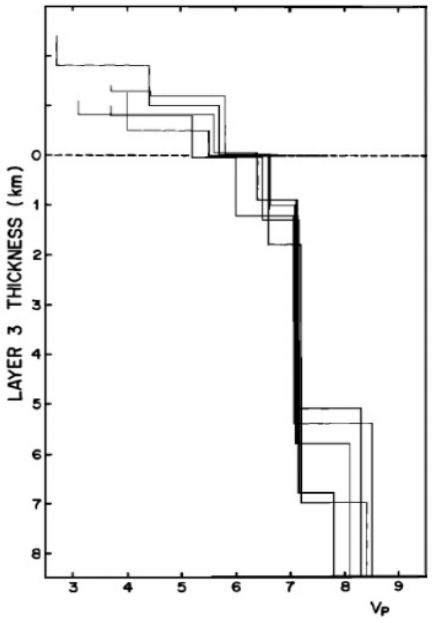


Figure 2: General velocity model from N. Christensen and Salisbury 1975

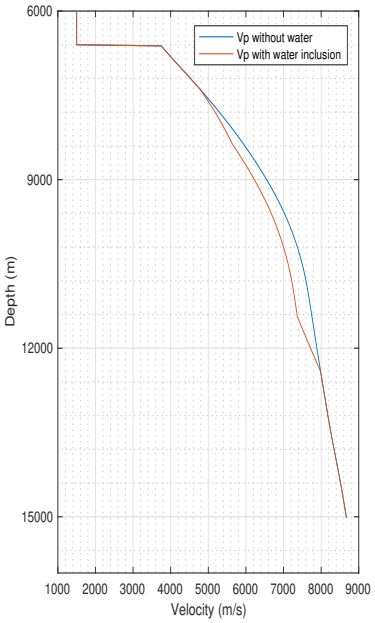


Figure 3: Velocity model for our case

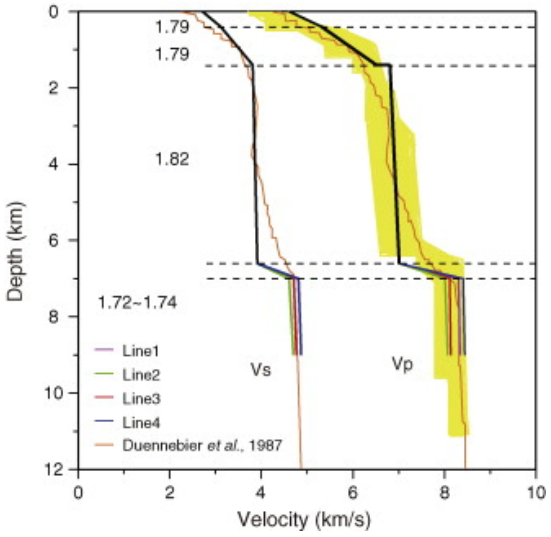


Figure 4: Velocity model for the Japan trench from Shinohara et al. 2008

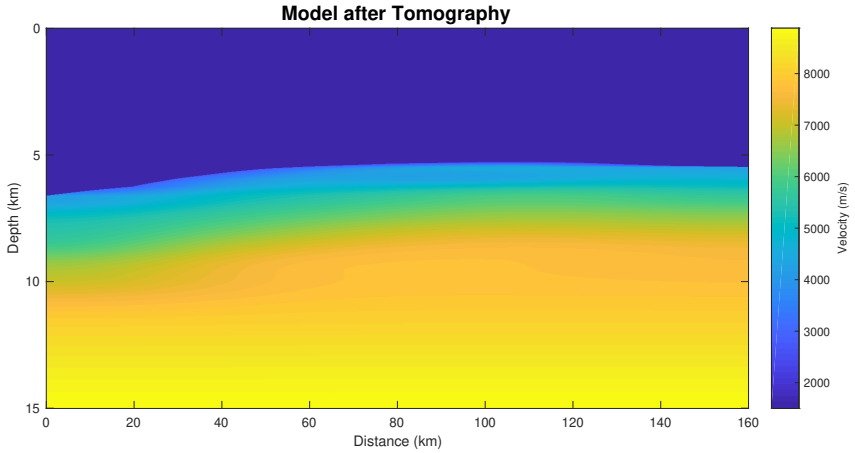


Figure 5: Model obtained using first arrival tomography

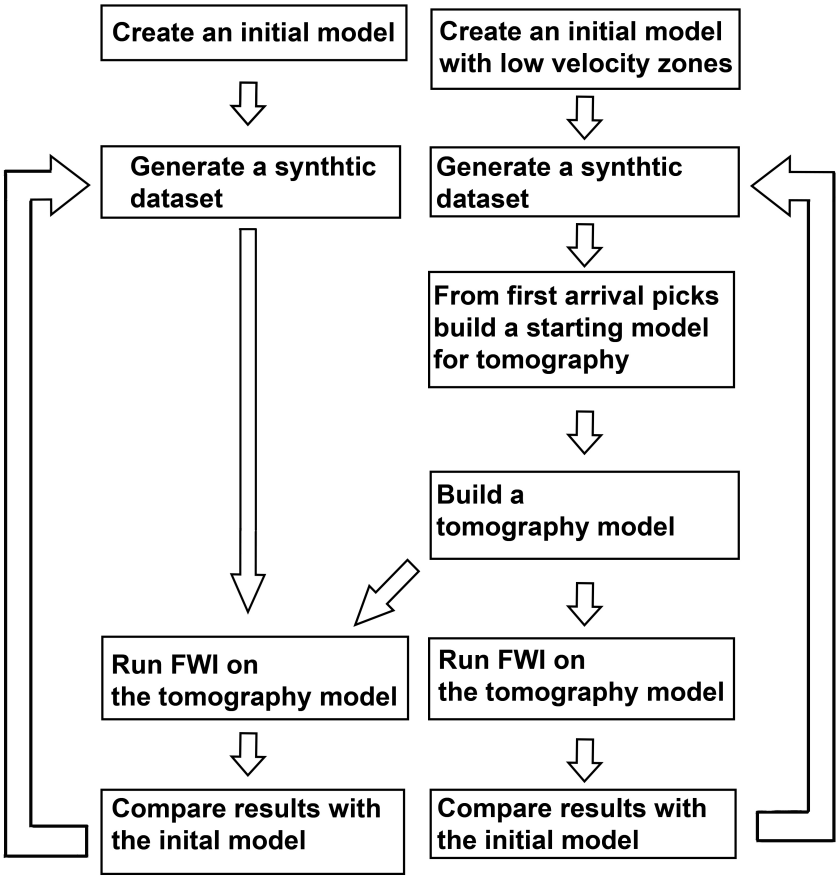


Figure 6: Workflow followed in this study

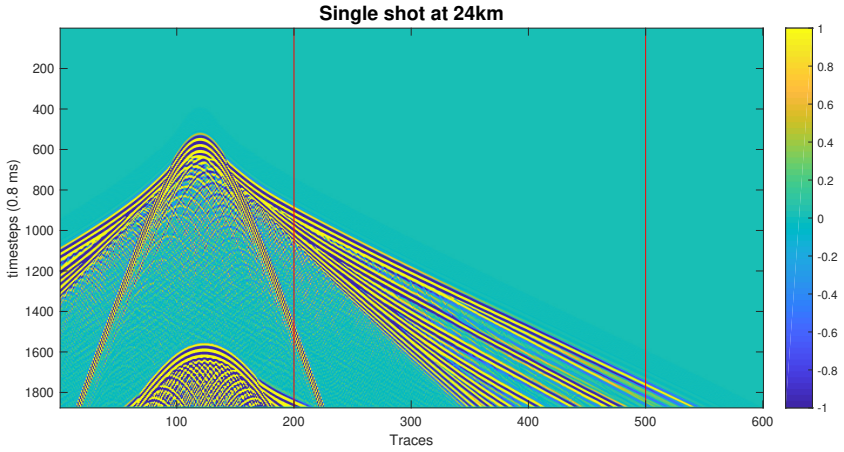


Figure 7: Example of a single shot gather on the final output of FWI using 1km used for comparing FWI results, in red the traces taken for near and far offsets in figures 5.9 and 5.10

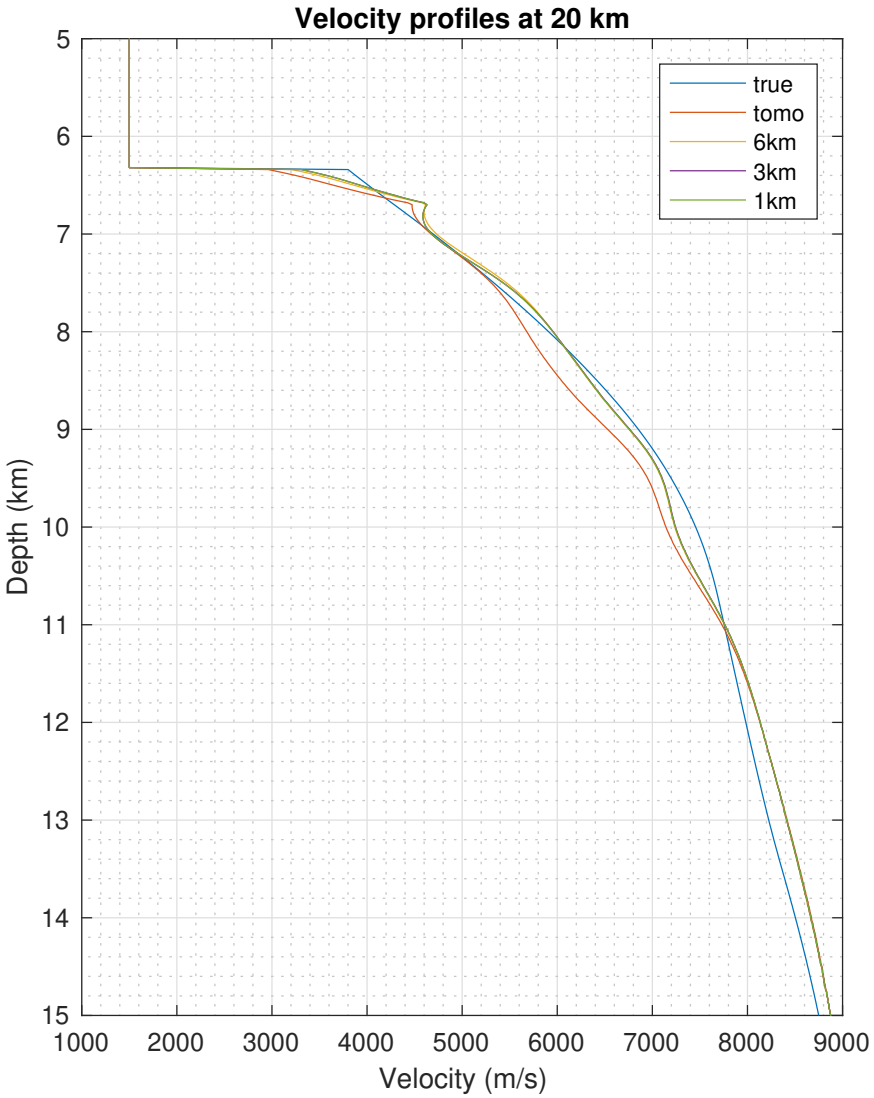


Figure 8: Vertical velocity profiles for all standard models at 20 km

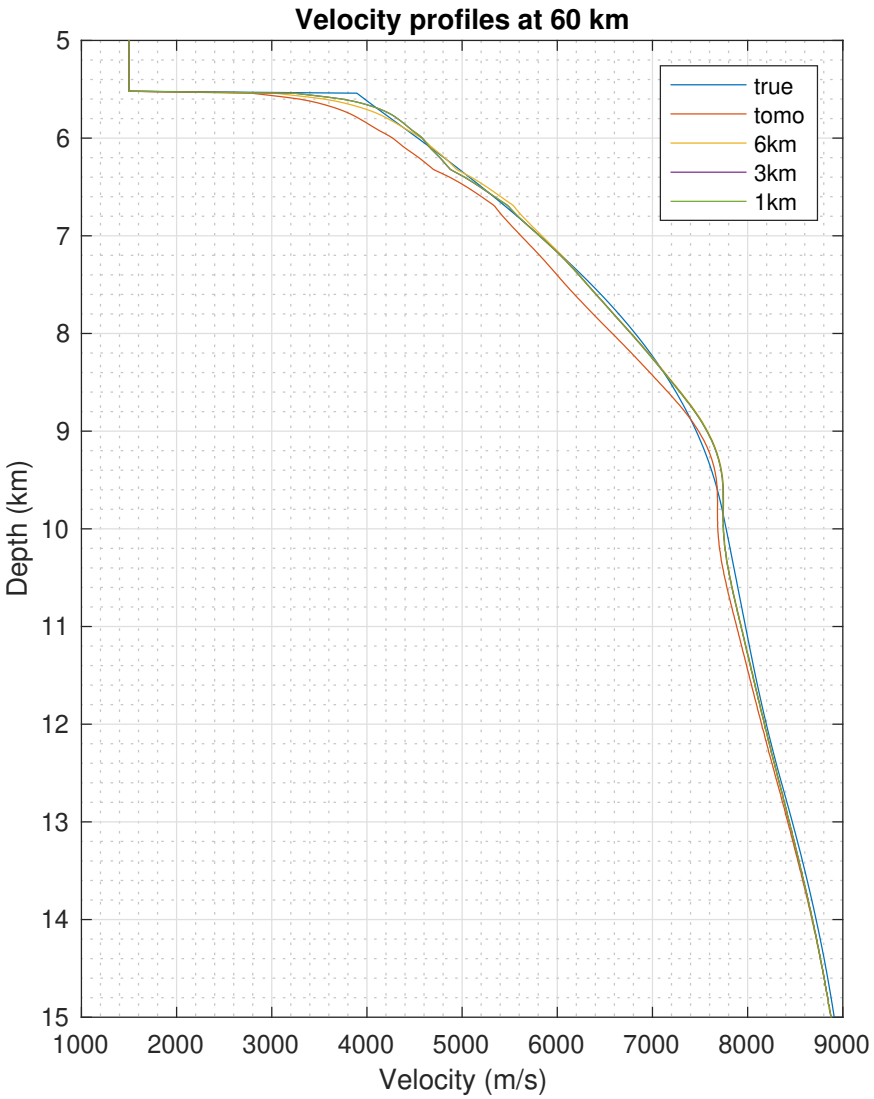


Figure 9: Vertical velocity profiles for all standard models at 60 km

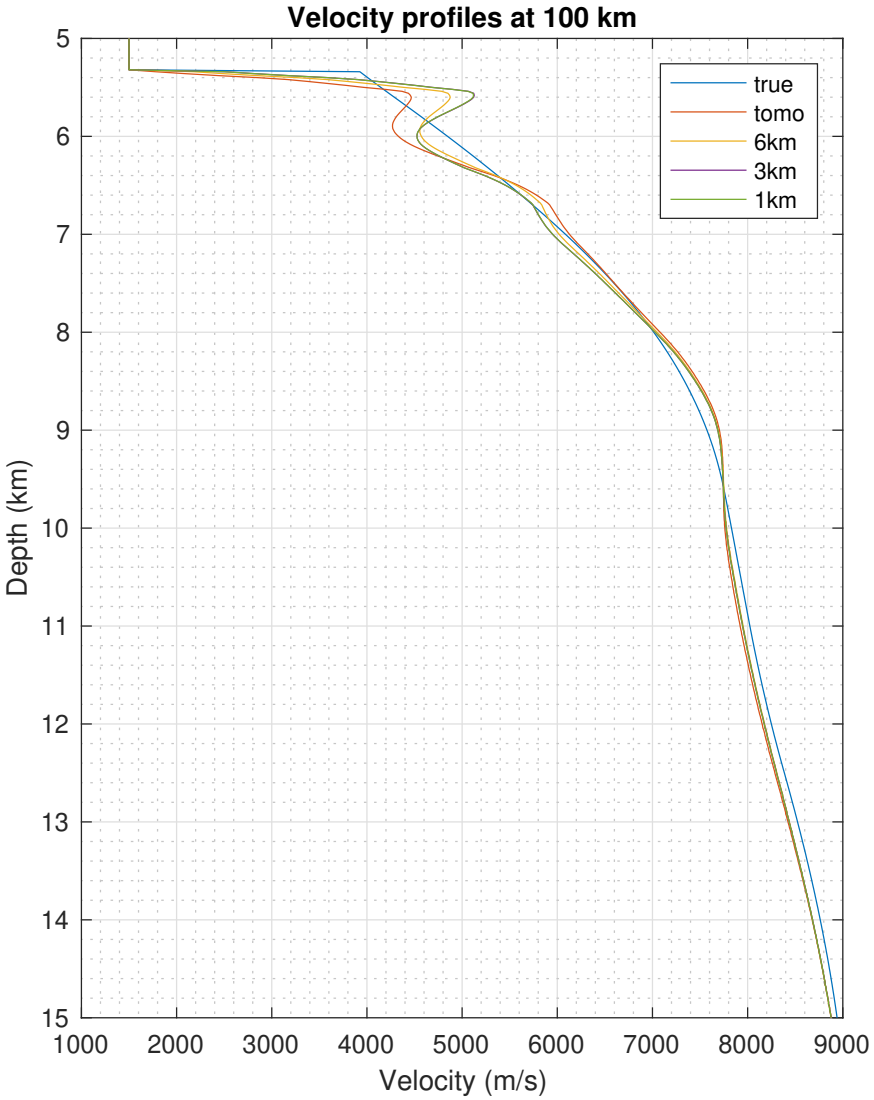


Figure 10: Vertical velocity profiles for all standard models at 100 km

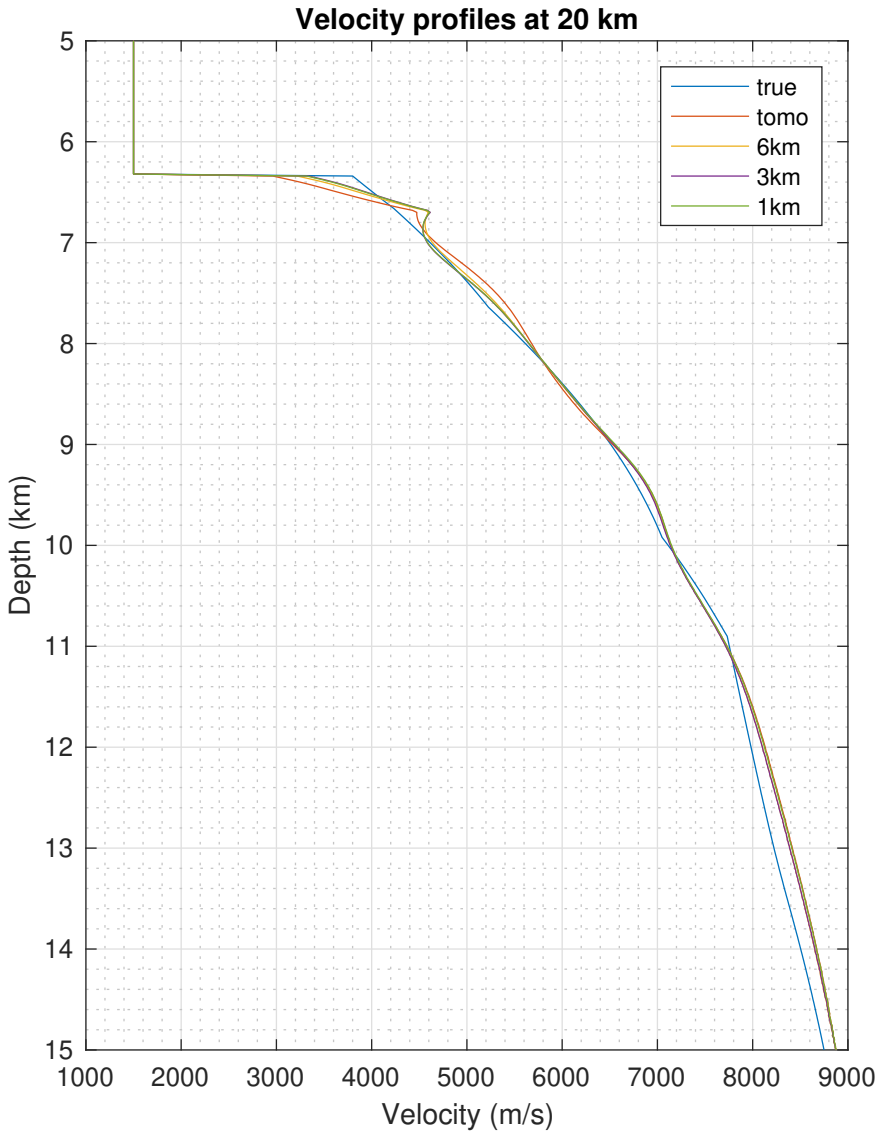


Figure 11: Vertical velocity profiles for all models with low velocity zones at 20 km

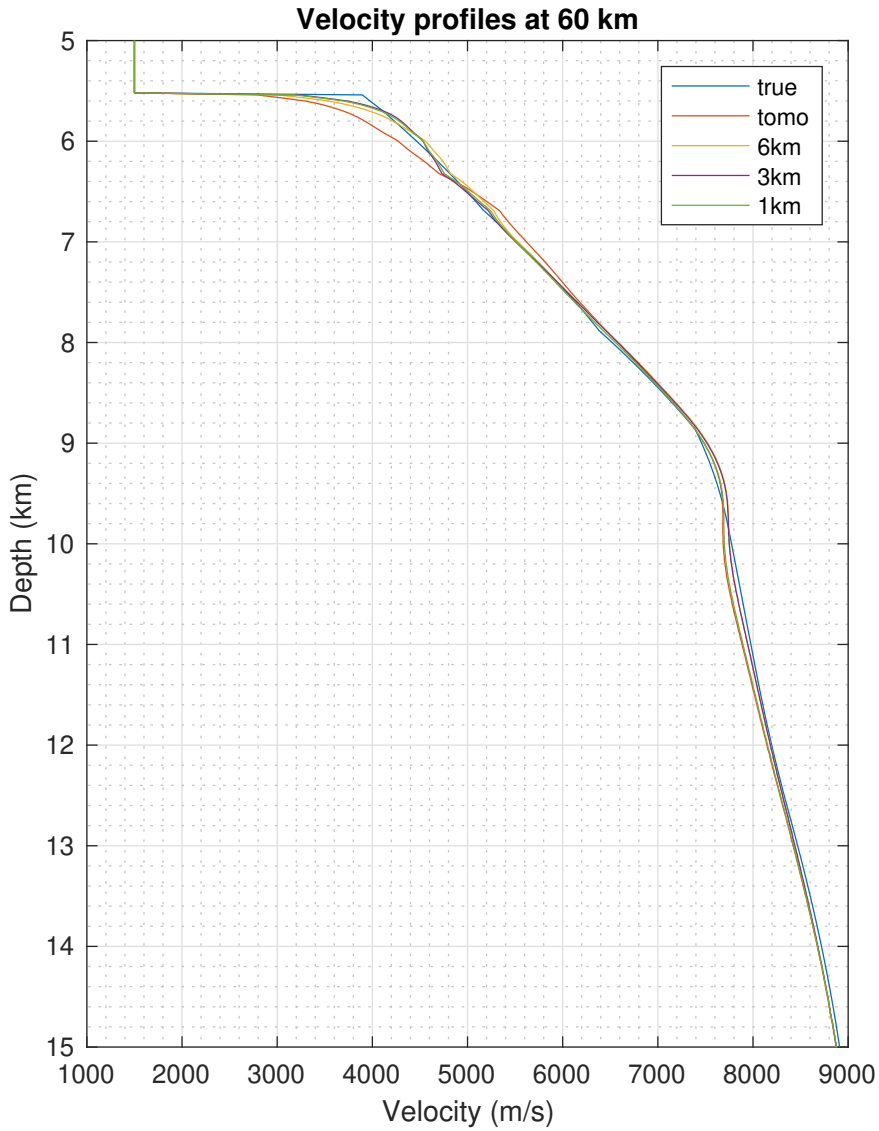


Figure 12: Vertical velocity profiles for all models with low velocity zones at 60 km

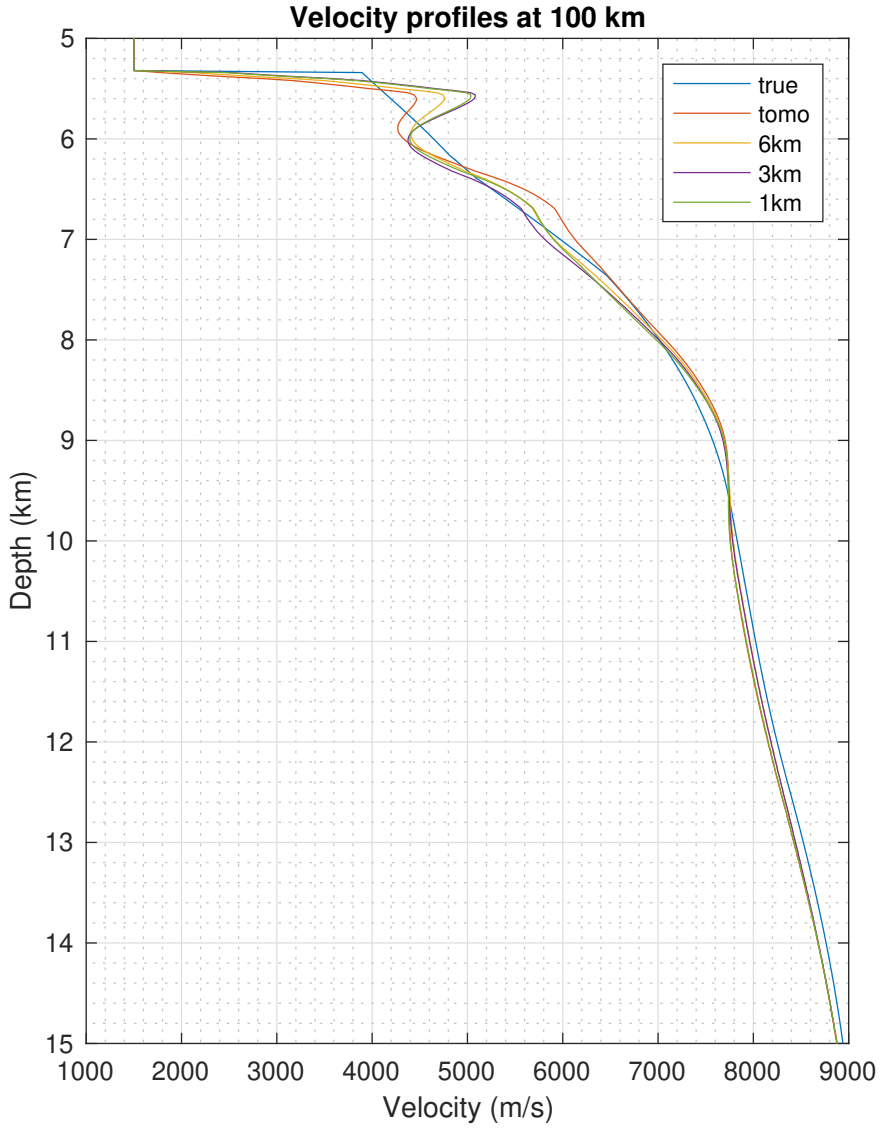
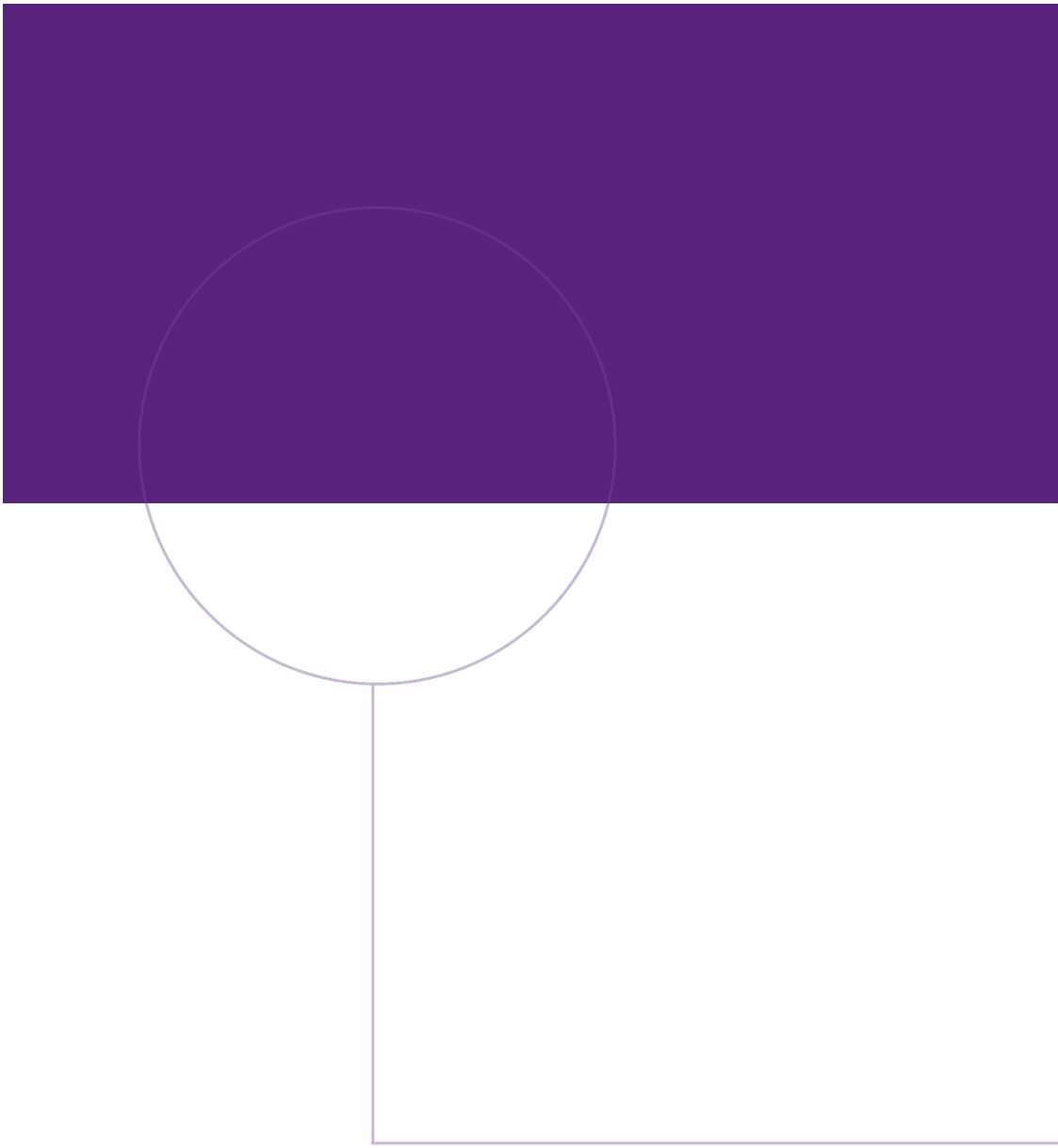


Figure 13: Vertical velocity profiles for all models with low velocity zones at 100 km



NTNU

Norwegian University of
Science and Technology