Karlsruhe Institute of Technology

Geophysical Institute KIT-Dept. of Physics

KARLSRUHE INSTITUTE OF TECHNOLOGY

MASTER THESIS

Random-shot full-waveform inversion of shallow seismic surface waves

---

ZUFALLSSCHUSS VOLLE WELLENFORMINVERSION VON FLACHSEISMISCHEN OBERFLÄCHENWELLEN

Author:
Duy Hoang HOANG

Reviewer:
Prof. Dr. Thomas BOHLEN
Co-Reviewer:
Prof. Dr. Andreas RIETBROCK
Supervisor:
Dr. Yudi PAN

Als Prüfungsexemplar genehmigt

A thesis submitted in fulfillment of the requirements for the degree of Master of Science

in the

Applied Geophyics
Department of Physics

March 5, 2021

Declaration on oath

I declare that I have developed and written the enclosed thesis completely by myself and not used sources or means without declaration in the text.

Signed: _____

Date: _____

Abstract

Seismic acquisitions in urban areas represent a challenge due to the noisy environment during the recording of the data. This often prevents obtaining high-resolution images of the subsurface in the area of interest. One way to achieve better results, despite the challenging circumstances, is full-waveform inversion (FWI). This method iteratively updates a starting model until the simulated data and the observed field data matches to a satisfactory level. This eventually yields the desired subsurface model. However, FWI comes with some major limitations. Because each iteration requires a calculation of the wave propagation several times for each shot, it is highly computationally expensive. Furthermore, due to storage limitations, especially if high performance computing (HPC) clusters are not available, the model size is restricted as well.
This work aims to find a solution for these limitations in an elastic 2D case by implementing the random-shot workflow (RSW). In a conventional FWI workflow which is called full-shot workflow (FSW) in this work, each iteration uses all shots of the recorded data before updating the whole model at once which leads to the high computational and storage costs. RSW randomly picks only one shot, applies FWI within the area of the shot and its corresponding receivers and then updates the model only within this subset. This should, in my expectation, split the workload into several smaller subsets without sacrificing the effectiveness of FWI.
After the implementation, I compare RSW to FSW in a synthetic benchmark where I verify the performance. Afterwards, the workflows are compared to each other by using a fraction of a field data set that was acquired in Salt Lake City. These data represents a large urban seismic land data set with a low signal-to-noise ratio. They also have suboptimal conditions such as the end-on spread geometry due to the use of a land streamer and occasionally bad coupling of the geophones. This turns out to be a big challenge for the FWI algorithm. The starting model used for the field data inversion is based on the multichannel analysis of surface waves which deals with the inversion of surface-wave dispersion curves. It is a cost-efficient way to obtain an initial model for FWI. With a subset of this model and the data, I investigate the influence of traces near the sources which suffer heavily from the suboptimal acquisition conditions, and if more iterations can improve the results. Removing such traces improve the major part of the results, while more iterations are not beneficial with this particular data set. At the end I show a final comparison of both workflows by applying them to the complete Salt Lake City field data set. An application of FWI on such a large-scale land streamer acquisition with such a large number of shots has, to the best of my knowledge, not been done yet, most likely because of the large computational costs, even in 2D.
The comparisons and investigations show that while RSW has a major computational advantage in storage usage and per-CPU performance, it is too sensitive to bad traces and shots which significantly degrade the final model. Therefore, it is unsuitable for field data with a low signal-to-noise ratio or that have not been preprocessed well. Furthermore, RSW cannot utilise the whole power of HPC clusters while FSW is able to do so by optimal shot parallelisation. Thus, FSW could potentially outperform RSW by raw computing power if such excessive power is available.

Acknowledgements

I would like to sincerely thank Prof. Thomas Bohlen for the suggestion of this interesting topic and the guidance of my work. I would also like to thank Prof. Andreas Rietbrock for agreeing to be the second reviewer.
Moreover, I would like to thank Yudi Pan for guiding me through the major part of the thesis and for motivating me to keep up my work. I would also like to thank Tilman Steinweg for introducing me to the code and for teaching me everything I needed to know to handle the implementation of RSW.
Next, I would like to thank Thomas Hertweck for proofreading my thesis and for his general support to us students. Also, special thanks to Mark Wienöbst for helping me running the code on HPC clusters and to Priscilla Ntengue with whom I shared my office.
Furthermore, I would like to thank the rest of the working group "Applied Geophysics" at KIT for the pleasant working atmosphere during my work.
Last but not least, I would like to thank my parents for supporting me throughout my thesis, studies and life.

Contents

List of Figures

List of Tables

List of Abbreviations

FWI	**F**ull-**W**aveform **I**nversion
HPC	**H**igh **P**erformance **C**omputing
FSW	**F**ull-**S**hot **W**orkflow
RSW	**R**andom-**S**hot **W**orkflow
FD	**F**inite **D**ifference
C-PML	**C**onvolutional **P**erfectly **M**atched **L**ayers
STF	**S**ource **T**ime **F**unction
NEHRP	**N**ational **E**arthquake **H**azards **R**eduction **P**rogram
SU	**S**eismic **U**nix
GPIHPC	**G**eo**p**hysical **I**nstitute **H**igh **P**erformance **C**omputer

Chapter 1

Introduction

Urban seismic acquisitions have always been challenging in comparison to rural ones and even more so to marine acquisitions where noise is not as prominent, albeit still challenging and sometimes difficult to remove (Elboth, Geoteam, and Hermansen, 2009). However, if the interest is less about exploring resources far away from city areas or on sea and more about investigating the urban subsurface, e. g. to identify potential hazard areas, there is no way around an urban seismic acquisition. There are many methods available to get an image of the subsurface, such as seismic migration methods. One requirement to perform them, though, is to know the velocity model of the subsurface.

One method that has found its place in many geophysical studies and applications is full-waveform inversion (FWI) which is able to yield a high-resolution parameter model of the subsurface. This advantage is achieved by using the information of both the amplitude and the phase of a seismogram. Thus, all the available information of the recorded data is utilised instead of just one parameter such as in traveltime tomography. First theoretical concepts of FWI were researched in (Tarantola, 1984) and successful applications can be found in various fields such as in laboratory investigations in the sub-metre range (Bretaudeau et al., 2013), exploration problems (Jaiswal et al., 2009), crustal-scale passive seismic investigations (Fichtner et al., 2013) and also medical imaging (Guasch et al., 2020).

The general idea behind FWI is to minimise the misfit between observed field data and calculated synthetic data. The synthetic data are modelled in a starting model which represents the subsurface as a first educated guess. Because the observed data come from the true subsurface model which is unknown to us, there are differences between field and modelled data, quantised as a misfit function. FWI updates the starting model based on this misfit with the aim to recalculate the synthetic data with an updated model, resulting in a smaller misfit. The starting model is consequently iteratively updated until the misfit reaches a satisfactorily small value.

However, this process also comes with limitations, especially where large computational power is not available. FWI requires powerful computational resources because it simulates the wave propagation in the subsurface model several times per iteration. Additionally, in a full workflow, many iterations are required to improve the starting model that ideally converges towards the true model which represents the observed data. Therefore, only in recent years it was feasible to apply FWI to large data sets as in Brossier, Operto, and Virieux (2009), Sears, Barton, and Singh (2010), Prieux et al. (2013). Furthermore, the model size is also often a restricting factor. Not only are the applications often limited to 2D because of the much higher computational cost in 3D, but large models often require large storage capacities that are only available in high performance computing (HPC). Hence, FWI cannot be universally used to determine a large velocity model if computational power is limited.

In this thesis I try to broaden the use case by finding a more efficient way to perform FWI for large near-surface data sets in an elastic 2D case. The idea is to modify the conventional FWI workflow (Kurzmann et al., 2013), that I call the full-shot workflow (FSW) in this work, and introduce the random-shot workflow (RSW) where I try to bypass the computing power and storage limitations of FWI. In FSW each iteration uses all shots available from the recorded data before the model is updated. This results in simulating the wave propagation of all shots several times per iteration and leading to the high costs. The RSW, however, randomly picks only one shot and applies the FWI workflow within the area of the chosen shot and its corresponding receivers. This means that the wave propagation is simulated only for a single shot per iteration. At the end of the iterations, the model is updated in the area of the chosen shot, thus the whole model is iteratively improved step by step in subsets instead of all at once. This should, in my expectation, split the workload into several smaller subsets without sacrificing the effectiveness of FWI.
After the implementation of RSW I compare it to FSW with synthetic data and field data that were acquired in Salt Lake City (Liberty, St Clair, and Gribler, 2018). These data represent a difficult data set due to the use of a land streamer and its acquisition in an urban area. It is also a very large data set with a measured profile of approximately 1300 m and a total of 655 shots which is quite challenging to deal with in a conventional FWI workflow like in the FSW. Therefore, most near-surface applications of FWI deal with a shorter profile or less shots within a profile as in Romdhane et al. (2011), Adamczyk, Malinowski, and Malehmir (2014) and Dokter et al. (2017). A near-surface application of FWI on a scale of the Salt Lake City data set has, to the best of my knowledge, not been done yet.

Chapter 2

Theory

The numerical simulation in my work can be split into two major parts. The first part is the forward problem which deals with elastic wave equations and their solution. This allows for a simulation of wave propagation in an elastic medium that is defined by certain model parameters. Thus, the solution of the forward problem produces synthetic seismograms. The second part is full-waveform inversion that consists of the inverse problem which tries to find the correct velocity model based on the observed seismograms from the field. If the synthetically produced seismograms from the forward problem match the real ones from the field data, then the inverted velocity model that was used to generate the synthetic data represents be the subsurface correctly.

2.1 Forward problem

2.1.1 Wave equation

Derivation

In the following the elastic wave equation is derived that explain the behaviour of the wavefields. The derivation closely follows to approach of Aki and Richards (2002).
In an elastic medium the relationship between stress σ and strain ϵ can be described using the Einstein notation by the linear equation

$$\sigma_{ij} = C_{ijkl}\epsilon_{kl} \ , \quad i,j,k,l \in [1,3]. \tag{2.1}$$

Both stress and strain are tensors of second order, whereas the stiffness tensor C is of fourth order with 81 components. These components, however, consist of only 21 independent elements due to the symmetry of σ and ϵ and energy reasons. If, furthermore, the medium is isotropic, the relation simplifies to just two independent parameters

$$\sigma_{ij} = \lambda\Theta\delta_{ij} + 2\mu\epsilon_{ij} \tag{2.2}$$

with

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right). \tag{2.3}$$

The constants λ and μ are the first and second Lamé parameters, respectively. The parameter Θ is the trace of the strain tensor ϵ, δ is the Kronecker delta and u_i is the particle displacement. From the Lamé parameters, the seismic velocities v_p and v_s of

the P-wave and S-wave, respectively, are defined as

$$v_p = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad v_s = \sqrt{\frac{\mu}{\rho}} \tag{2.4}$$

with the mass density ρ.
In addition to the stress-strain relation, the equation of motion plays an important role in elastic wave propagation. This equation can be derived from Newton's second law (Aki and Richards, 2002). With an arbitrary finite volume V and bounding surface S it holds

$$\int_V \rho \frac{\partial^2 u_i}{\partial t^2} \, \mathrm{d}V = \int_V f_i \, \mathrm{d}V + \int_S T_i(n_i) \, \mathrm{d}S \tag{2.5}$$

where the inertia forces of the volume V equals the sum of body and surface forces which act on that volume or its surface. The vector $\vec{f}$ represents the body forces acting on a unit of V and $\vec{T}$ are the forces acting on a unit of S. By using the stress tensor σ, $\vec{T}$, also known as the traction, can be expressed as

$$T_i = \sigma_{ij} n_j \tag{2.6}$$

with the vector $\vec{n}$ being normal to the surface S pointing outwards. If the Gauss' divergence theorem is used, we get

$$\int_V \left(\rho \frac{\partial^2 u_i}{\partial t^2} - f_i - \frac{\partial \sigma_{ij}}{\partial x_j} \right) \mathrm{d}V = 0. \tag{2.7}$$

Right now, the expression depends on the chosen volume unless the integrand itself is zero. With the particle velocity $\vec{v} = \frac{\partial \vec{u}}{\partial t}$, this gives

$$\rho \frac{\partial v_i}{\partial t} - f_i - \frac{\partial \sigma_{ij}}{\partial x_j} = 0 \tag{2.8}$$

which is the equation of motion. Together with the time derivatives of equations 2.2 and 2.3, the wave propagation in an elastic medium is explained by velocity-stress formulation

$$\frac{\partial \sigma_{ij}}{\partial t} = \lambda \frac{\partial \Theta}{\partial t} \delta_{ij} + 2\mu \frac{\partial \epsilon_{ij}}{\partial t} \tag{2.9}$$

$$\frac{\partial \epsilon_{ij}}{\partial t} = \frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right). \tag{2.10}$$

Matrix-vector formulation

Equations 2.8, 2.9 and 2.10 can also be expressed in matrix-vector form

$$\mathbf{M}^{-1} \left(\frac{\partial \vec{\zeta}}{\partial t} - \vec{b} \right) = \mathbf{Q} \vec{\zeta}. \tag{2.11}$$

In the 2D case with $x_1 = x$ being the horizontal component and $x_2 = y$ the vertical component, the matrix $\mathbf{M}$ in equation 2.11 contains the model parameters λ, μ and ρ

and the the matrix $\mathbf{Q}$ the spatial derivatives. The matrices are given by

$$\mathbf{M} = \begin{pmatrix} \rho & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & A & B & 0 \\ 0 & 0 & B & A & 0 \\ 0 & 0 & 0 & 0 & \mu^{-1} \end{pmatrix} \qquad \mathbf{Q} = \begin{pmatrix} 0 & 0 & \frac{\partial}{\partial x} & 0 & \frac{\partial}{\partial y} \\ 0 & 0 & 0 & \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \\ \frac{\partial}{\partial x} & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 & 0 & 0 \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 & 0 & 0 \end{pmatrix} \tag{2.12}$$

with

$$A = \frac{\lambda + 2\mu}{4\mu(\lambda + \mu)} \qquad \text{and} \qquad B = -\frac{\lambda}{4\mu(\lambda + \mu)}. \tag{2.13}$$

The vector $\vec{\zeta}$ in equation 2.11 denotes the different wavefields and the vector $\vec{b}$ denotes their corresponding sources. In a 2D case, they read

$$\vec{\zeta} = \begin{pmatrix} v_x \\ v_y \\ \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{pmatrix} \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} f_x/\rho \\ f_y/\rho \\ \frac{\partial \sigma_{xx}^{(0)}}{\partial t} \\ \frac{\partial \sigma_{yy}^{(0)}}{\partial t} \\ \frac{\partial \sigma_{xy}^{(0)}}{\partial t} \end{pmatrix}. \tag{2.14}$$

In a 3D case, the z-component must be added in the vectors and matrices of equation 2.11. However, since my forward modelling stays in two spatial dimensions in the scope of this thesis, I do not show their formulations.
The equations from the matrix-vector in equation 2.11 are implemented in the WAVE-Toolbox. This toolbox is the framework for the numerical simulation in this work. The reason for the choice of the WAVE-Toolbox is the availability of the high-performance computing library LAMA from the Fraunhofer Institute for algorithms and scientific computing (SCAI) that allows me to run matrix-vector operations parallel. Thus, the derived formulations can be implemented in the form $y = \mathbf{A} \cdot x$, allowing for very efficient simulations (Brandes, Schricker, and Soddemann, 2017).

2.1.2 Finite-difference method

The derived system of partial differential equations from section 2.1.1 can be solved analytically or numerically. Due to the complex media in realistic cases, only the numerical approach is feasible. Thus, the finite-difference (FD) method is used.
The aim is to approximate the derivatives in equations 2.8, 2.9 and 2.10 by finite differences. To do so, space and time must be discretised. In terms of space, the discretisation allows each specific grid point to hold the model parameters which are in elastic modelling the P- and S-wave velocity and the density. The spatial discretisation is

$$x = i \cdot \Delta x \quad \text{and} \quad y = j \cdot \Delta y \tag{2.15}$$

with (i, j) referring to grid points on a Cartesian grid. Since an equidistant grid is chosen, the grid spacings in both directions are equal, i. e. $\Delta x = \Delta y$, and can be written as $\Delta x = \Delta y = \Delta h$. The discretisation in time is defined by the interval Δt between two time steps, hence

$$t = n \cdot \Delta t \tag{2.16}$$

where n denotes a specific time step.

Staggered grid

An important aspect for efficiency is the use of a staggered grid, which means that some model parameters are shifted by half a grid point. Virieux (1986) also showed this method ensures that the partial differential equations in 2.8, 2.9 and 2.10 refer to the same grid point. An illustration of the staggered grid is show in figure 2.1.
To approximate the temporal derivatives of σ and $\vec{v}$ in equations 2.8 and 2.9, the second-order centred finite difference is used

$$\left.\frac{\partial v_i}{\partial t}\right|^{n+\frac{1}{2}} \approx \frac{v_i^{n+1} - v_i^n}{\mathrm{d}t} \tag{2.17}$$

$$\left.\frac{\partial \sigma_{ij}}{\partial t}\right|^{n} \approx \frac{\sigma_{ij}^{n+\frac{1}{2}} - \sigma_{ij}^{n-\frac{1}{2}}}{\mathrm{d}t}, \tag{2.18}$$

where the time step is denoted as superscript. Accordingly, by approximating the temporal derivatives in 2.2 and 2.8, the two parameters are updated as

$$v_i^{n+1} \approx v_i^n + \frac{\mathrm{d}t}{\rho}\left.\frac{\partial \sigma_{ji}}{\partial x_j}\right|^{n+\frac{1}{2}} \tag{2.19}$$

$$\sigma_{ij}^{n+\frac{1}{2}} \approx \sigma_{ij}^{n-\frac{1}{2}} + \mathrm{d}t\lambda\left.\frac{\partial}{\partial t}\Theta\right|^{n}\delta_{ij} + 2\,\mathrm{d}t\mu\left.\frac{\partial \epsilon_{ij}}{\partial t}\right|^{n}. \tag{2.20}$$

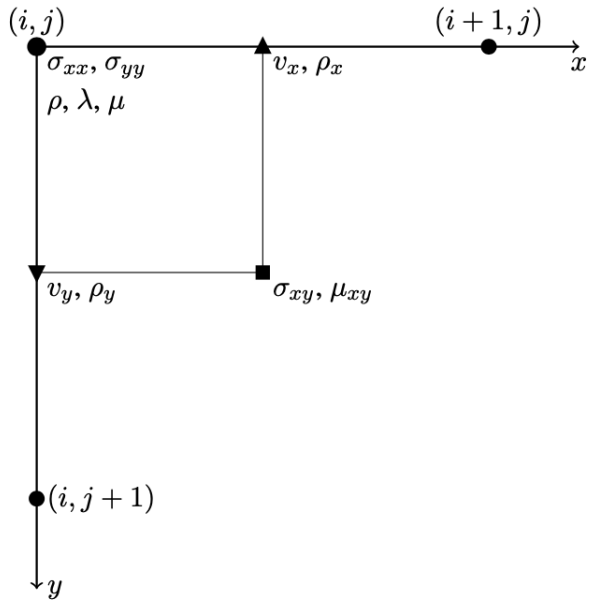


FIGURE 2.1: Standard staggered grid by Virieux (1986). Some parameters are shifted by half a grid point $dh/2$, others are defined on full grid points (i, j).

These can be combined with equation 2.10 to yield the discretised elastic wave equation

$$\sigma_{xx,i,j}^{n+\frac{1}{2}} = \sigma_{xx,i,j}^{n-\frac{1}{2}} + \Delta t \cdot \lambda_{i,j} \left(\frac{\partial v_x}{\partial x}\bigg|_{i,j}^{n} + \frac{\partial v_y}{\partial y}\bigg|_{i,j}^{n} \right) + 2 \cdot \Delta t \cdot \mu_{i,j} \frac{\partial v_x}{\partial x}\bigg|_{i,j}^{n} \tag{2.21}$$

$$\sigma_{yy,i,j}^{n+\frac{1}{2}} = \sigma_{yy,i,j}^{n-\frac{1}{2}} + \Delta t \cdot \lambda_{i,j} \left(\frac{\partial v_x}{\partial x}\bigg|_{i,j}^{n} + \frac{\partial v_y}{\partial y}\bigg|_{i,j}^{n} \right) + 2 \cdot \Delta t \cdot \mu_{i,j} \frac{\partial v_y}{\partial y}\bigg|_{i,j}^{n} \tag{2.22}$$

$$\sigma_{xy,i+\frac{1}{2},j+\frac{1}{2}}^{n+\frac{1}{2}} = \sigma_{xy,i+\frac{1}{2},j+\frac{1}{2}}^{n-\frac{1}{2}} + \Delta t \cdot \mu_{i+\frac{1}{2},j+\frac{1}{2}} \left(\frac{\partial v_x}{\partial y}\bigg|_{i+\frac{1}{2},j+\frac{1}{2}}^{n} + \frac{\partial v_y}{\partial x}\bigg|_{i+\frac{1}{2},j+\frac{1}{2}}^{n} \right) \tag{2.23}$$

$$v_{x,i+\frac{1}{2},j}^{n+1} = v_{x,i+\frac{1}{2},j}^{n} + \frac{\Delta t}{\rho_{i+\frac{1}{2},j}} \left(\frac{\partial \sigma_{xx}}{\partial x}\bigg|_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + \frac{\partial \sigma_{xy}}{\partial y}\bigg|_{i+\frac{1}{2},j}^{n+\frac{1}{2}} \right) \tag{2.24}$$

$$v_{y,i,j+\frac{1}{2}}^{n+1} = v_{y,i,j+\frac{1}{2}}^{n} + \frac{\Delta t}{\rho_{i,j+\frac{1}{2}}} \left(\frac{\partial \sigma_{yx}}{\partial x}\bigg|_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} + \frac{\partial \sigma_{yy}}{\partial y}\bigg|_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} \right) \tag{2.25}$$

Numerical stability

The choice of the spatial and temporal discretisation Δh and Δt is not arbitrary but must be chosen to ensure a precise and stable numerical simulation. If their choice exceeds a certain upper threshold, numerical dispersion and instabilities will spoil the results, whereas choosing too small values leads to longer computation times. Numerical dispersion appears if Δh is too large. The threshold is defined by the criterion (Köhn, 2011)

$$\Delta h \leq \frac{\lambda_{min}}{n} = \frac{v_{min}}{n f_{max}}. \tag{2.26}$$

It depends on the minimum wavelength λ_{min} which is determined by the quotient of the minimum seismic velocity v_{min} divided by the maximum frequency f_{max} of the source signal. The number n depends on the FD order N shown in table 2.1.
Numerical instability is related to the choice of time discretisation Δt, in particular the traveltime between two adjacent grid points. The threshold to ensure a stable simulation in 2-D is definied by the Courant-Friedrichs-Lewy criterion (Courant, Friedrichs, and Lewy, 1928)

$$\Delta t \leq \frac{\Delta h}{\sqrt{2} k v_{max}} \tag{2.27}$$

with k dependent on the FD order N and being the sum of the FD-coefficients $\beta_{n,N}$

$$k = \sum_{n=1}^{N} \beta_{n,N}. \tag{2.28}$$

TABLE 2.1: Dependence of number of grid points per minimum wavelength on FD order for Taylor operators (Köhn, 2011).

FD order N	2	4	6	8	10	12
n	12	8	6	5	5	4

TABLE 2.2: Values of the sum of the FD-coefficients for different FD orders (Köhn, 2011).

FD order N	2	4	6	8	10	12
k	1	7/6	146/20	2161/1680	53089/40320	1187803/887040

The FD coefficients themselves are also dependent on the FD order. Their estimation can be done with a Taylor series expansion (Jastram, 1992). The values of h for different orders are shown in table 2.2.

2.1.3 Boundary conditions

Free surface

In general, the equation of motion 2.8 and the velocity-stress formulation 2.9 and 2.10 explain the wave propagation in an elastic medium. However, the Earth's surface requires special conditions to allow for precise modelling. The initial conditions in an elastic case at the free surface interface $y = 0$ are

$$\sigma_{xy} = \sigma_{yy} = 0. \tag{2.29}$$

This means that the stress in normal direction above the surface disappear since there is no particle displacement in the same sense as within the medium. To deal with this, the mirroring technique by Levander (1988) can be used by solving equations 2.2 and 2.3 under the condition 2.29, in particular

$$\sigma_{yy} = \lambda \frac{\partial u_x}{\partial x} + (\lambda + \mu) \frac{\partial u_y}{\partial y} = 0 \tag{2.30}$$

$$\sigma_{xy} = \mu \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) = 0. \tag{2.31}$$

Mirroring the stresses right below the surface to the opposite side above the surface with an inverse sign will fulfil these equations. This method, however, only works if the free surface is planar.

Model boundaries

After solving the boundary condition at the free surface, there are still the lateral and bottom boundaries of the model. These do not play a role if the entire Earth is used as the model; however, in simulations only finite models are used. Thus, waves reach the edge of the model and create artificial reflections. To mitigate this phenomenon, a boundary layer is used to absorb the incoming waves.
In this work, the convolutional perfectly matched layers method (C-PML) by Komatitsch and Martin (2007) is used. It deals with the reflections by stretching the coordinates in the frequency domain, causing the waves to actually not reach the boundary because they decay exponentially. One caveat, however, is that this only works flawlessly for the exact solution of the wave equation and not for the approximated wave equation by the finite-difference method. Although some reflections still occur, they can be dealt with by the usage of additional damping to suppress them.

2.2 Inverse problem

The inverse problem consists of the challenge to find a model that explains the observed data. This task is not straight forward due to the ambiguous nature of inverse problems in seismics, meaning there are many possible models that could explain one set of data. Full-waveform inversion is an algorithm that tries to tackle this problem. The theory was introduced by Tarantola (1984). In comparison to other inversion methods such as traveltime tomography, FWI makes use of the whole information within the recorded data, both the amplitude and phase, instead of just parts of the information.

2.2.1 Misfit function

Full-waveform inversion comes from the deterministic inverse theory where the aim is to define a misfit function J and to minimize it (Fichtner, 2011). The misfit represents the difference between observed field data and the modelled synthetic data and can be quantified by the L_2-norm

$$J = \frac{1}{2} \int_x \int_0^T (\vec{d}_{syn} - \vec{d}_{obs})^2 \, \mathrm{d}t \, \mathrm{d}x \tag{2.32}$$

with the total recording time T. This represents the misfit of all shots summed together. To find an optimal model m_{opt}, it shall be minimised. Therefore, the global minimum of J needs to be found. This can be done iteratively by gradient-based algorithms where a starting model m_0 is iteratively improved until the minimum of its corresponding misfit J is reached and thus the optimal model is found. Fichtner (2011) shows a general rule to update a model

$$m_{i+1} = m_i - \alpha \mathbf{A}_i \nabla_m J(m_i) \tag{2.33}$$

with a step length α that scales the gradient and a preconditioning matrix $\mathbf{A}$ that depends on the chosen optimisation method. Equation 2.33 is, however, just used in the first iteration. Afterwards, the model updates are

$$m_{i+1} = m_i - \alpha_i \mathbf{A}_i \nabla_m \bar{J}_i \tag{2.34}$$

$$\nabla_m \bar{J}_i = \nabla_m J_i - \beta_i \nabla_m \bar{J}_{i-1} \tag{2.35}$$

which is the preconditioned conjugate-gradient method by Nocedal and Wright (1999). The preconditioning matrix $\mathbf{A}$ is calculated with the K1 method by Plessix and Mulder (2004), whereas the parameter β_n is chosen in the WAVE-Toolbox according to Polak and Ribière (1969):

$$\beta_i = \frac{(\nabla_m J_i)^T (\nabla_m J_i - \nabla_m J_{i-1})}{||\nabla_m J_{i-1}||^2} \tag{2.36}$$

Due to the non-linear nature of the forward problem that can be expressed as a non-linear operator being applied to a model to output the synthetic field data, the misfit function J is non-linear as well. This leads to the potential problem of finding only a local minimum instead of the global minimum. Differentiating a local minimum from a global one, though, is practically not trivial. Moreover, often any minimum is not even reached in a feasible number of iterations. Hence, the final model is already obtained and the inversion aborted if the misfit or its change compared to previous

iteration is sufficiently small enough. A way to avoid stopping at a local minimum is presented in chapter 2.2.4.

2.2.2 Adjoint-state method

The adjoint-state method is used to calculate the gradient of an objective function which in this case is the misfit function 2.32. In the following I will briefly show how this method is used to calculate the gradient of a single shot using Lagrange multipliers (Plessix, 2006) by minimising the objective function with a set of constraints.
As mentioned before, the objective function is the misfit function $J(\vec{\zeta}, m)$ that depends on the wavefields $\vec{\zeta}$ from equation 2.14 and the model m. The constraints are the forward problem $F(\vec{\zeta}, m)$ from equation 2.11:

$$\vec{F}(\vec{\zeta}, m) = \mathbf{M}^{-1}\left(\frac{\partial}{\partial t}\vec{\zeta} - \vec{b}\right) - \mathbf{Q}\vec{\zeta} = 0 \tag{2.37}$$

The construction of the Lagrangian L with the Lagrange multiplier $\vec{\psi}$ is

$$\begin{aligned} L(\vec{\zeta}, \vec{\psi}, m) &= J(\vec{\zeta}, m) - \vec{\psi}\vec{F}(\vec{\zeta}, m) && (2.38)\\ &= \frac{1}{2}\int_x \int_0^T (\vec{d}_{syn} - \vec{d}_{obs})^2 \,\mathrm{d}t\,\mathrm{d}x - \vec{\psi}\int_0^T \int_x \left(\mathbf{M}^{-1}\left(\frac{\partial\vec{\zeta}}{\partial t} - \vec{b}\right) - \mathbf{Q}\vec{\zeta}\right)\mathrm{d}t\,\mathrm{d}x && (2.39)\end{aligned}$$

which can be written as

$$L(\vec{\zeta}, \vec{\psi}, m) = J(\vec{\zeta}, m) \tag{2.40}$$

because $\vec{F}(\vec{\zeta}, m)$ is zero. Since the aim is to get the gradient of the objective function $\frac{\mathrm{d}J}{\mathrm{d}m}$, the gradient of the Lagrangian $\frac{\mathrm{d}L}{\mathrm{d}m}$ can be calculated instead. Thus, the partial derivatives are calculated with the chain rule

$$\frac{\mathrm{d}J(\vec{\zeta}, m)}{\mathrm{d}m} = \frac{\mathrm{d}L(\vec{\zeta}, \vec{\psi}, m)}{\mathrm{d}m} = \frac{\partial L(\vec{\zeta}, \vec{\psi}, m)}{\partial\vec{\zeta}}\frac{\partial\vec{\zeta}}{\partial m} + \frac{\partial L(\vec{\zeta}, \vec{\psi}, m)}{\partial m}. \tag{2.41}$$

This is solved if the first term is zero. Hence, I choose $\vec{\psi}$ such that

$$\frac{\partial L(\vec{\zeta}, \vec{\psi}, m)}{\partial\vec{\zeta}} = \frac{\partial J}{\partial\vec{\zeta}} - \vec{\psi}\frac{\partial\vec{F}}{\partial\vec{\zeta}} = 0. \tag{2.42}$$

This is the adjoint-state equation. Its solution is

$$\mathbf{M}^{-1}\frac{\partial\vec{\psi}(\tau)}{\partial\tau} - (\vec{d}_{syn} - \vec{d}_{obs}) = -\mathbf{Q}\vec{\psi}(\tau) \tag{2.43}$$

with $\tau = T - t$. This is the anti-self-adjoint equation. It has a similar form to the forward equation 2.11 where the data residuals are the sources, called adjoint sources. The corresponding wavefields are called adjoint wavefields, which is $\vec{\psi}$ if equation 2.43 is compared to 2.11. Furthermore, with the substitution of τ, it shows that the adjoint forward problem has to be solved backwards in time. Therefore, the initial conditions from the traditional forward problem are the terminal conditions of the adjoint forward problem. The gradient of the objective function is finally obtained by the cross-correlation of the forward wavefields and adjoint wavefields.

2.2.3 Step-length search

In equation 2.33 the step length α is introduced that scales the gradient. Depending on how much it increases or decreases the gradient, the model updates are stronger or weaker, respectively. The key is to find a good compromise. On the one hand, a model update too strong after one iteration leads to large jumps in the misfit function. This can complicate the search for a minimum because the minimum is skipped over and thus never reached. On the other hand, a step length that is too small stretches the convergence towards a minimum and more iterations are needed, hence the computational cost is increased. This is further complicated because the optimal step length changes throughout the inversion. At the beginning, the step length should be larger since the misfit is still large and its minimum probably still far away, whereas near the end, the misfit is near its minimum, therefore smaller step lengths are required to not skip it over.
This problem is tackled by Nocedal and Wright (1999) with a step length estimation based on a parabolic line. In each iteration, at least another two forward calculations are simulated to apply a parabolic fit. A more detailed explanation is given by Köhn (2011).

2.2.4 Multi-stage approach

I first mentioned the non-linearity in chapter 2.2.1. This problem causes the FWI algorithm to converge towards a local minimum, especially if the starting model differs a lot from the true model. In practice a good starting model is not always given, therefore a solution must be found. The multi-stage approach tackles this problem. As the name suggest, there are different stages. Each stage is a full FWI workflow and performed consecutively, where the next stage takes over the model result of the previous stage. The difference between each stage is the frequency range. At the beginning, only low frequencies are used which smooths the misfit function, thus the convergence moves in the direction of the global minimum. Afterwards, higher frequencies are introduced stage by stage. Figure 2.2 illustrates the objective function for different scale lengths. It shows that higher the frequency, the higher the possibility to be stuck in a local minimum due to the more complex functions at the top of the figure (Bunks et al., 1995).

2.2.5 Source time function inversion

With the inversion of field data, certain parameters are unknown such as the source signature. Often it can be approximated or an idealised signature such as the Ricker wavelet can be used, e. g. when synthetic data are created. But to obtain the source signature that created the observed data, I used the source time function inversion (STF) following Pratt (1999). In the following, a tilde denotes the Fourier transform of the corresponding quantity. The modified discrete forward modelling equation in the frequency domain is

$$\mathbf{S}\vec{\tilde{u}}(\omega_k) = s(\omega_k)\vec{\tilde{f}}(\omega_k). \tag{2.44}$$

$\mathbf{S}$ represents the model parameters as a complex-valued impedance matrix and ω_k the discrete frequencies on which all parameters except $\mathbf{S}$ depend on. By minimising the misfit function 2.32, a complex scalar s can be found that fulfils equation 2.44. Thus,

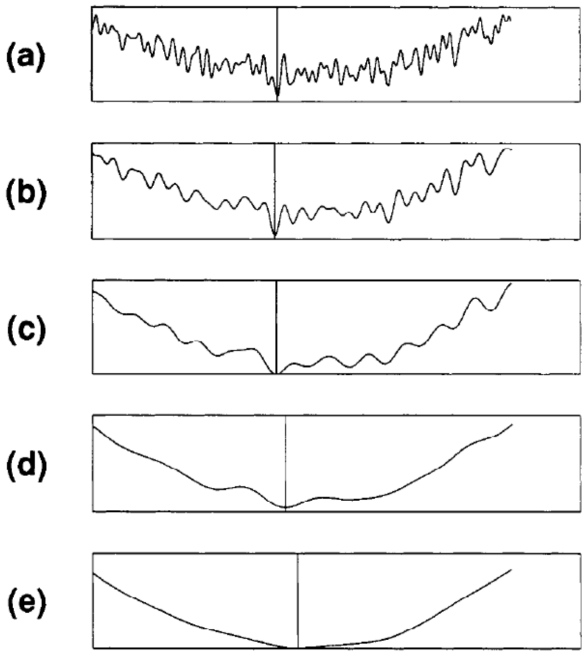


FIGURE 2.2: Objective functions at different scale lengths by Bunks et al. (1995). The top is more complex due to a higher frequencies, thus filled with more local minima.

its product with $\vec{f}$ corresponds to the source signature at ω_k. The analytical result is

$$s(\omega_k) = \frac{\vec{d}^T \vec{u}^*}{\vec{u}^T \vec{u}^* + \epsilon} \tag{2.45}$$

with $\vec{u}^*$ being the complex conjugate of $\vec{u}$. This is also known as the result of the Wiener filter. The Wiener filter is calculated and applied to the synthetic source signature to minimise the misfit between synthetic and observed source signal. The addition of ϵ is the so-called water level and helps to stabilise the equation in case the denominator is small which happens in the presence of a lot noise.

Chapter 3

Random-shot workflow

In chapter 2 I introduced the theoretical concepts and methodology of full-waveform inversion. These concepts are combined to a workflow, the FWI algorithm, which I explain more thoroughly in this chapter. In the first part, I discuss two different workflows. At first, I explain the conventional workflow, which I call in this work full-shot workflow (FSW), and discuss its limitations. To bypass them, I introduce the random-shot workflow (RSW) and discuss the idea behind it and the expectations to utilise the computational power more efficiently. In the second part I briefly explain how RSW is implemented in the WAVE-Toolbox.

3.1 FWI algorithm

3.1.1 Full-shot workflow

The full-shot workflow is the conventional algorithm used, such as in Groos (2013) and Krieger (2019). A visual representation is shown in figure 3.1. There are two starting points, on the top left is the true model which represents the earth model where the observed data were recorded. In the top center is the starting model m_0 which is improved iteratively until the misfit reaches a satisfactory value. To do so, the source time function inversion (STF) described in chapter 2.2.5 is applied in the first iteration for the first shot. The calculated source signature is then used to perform the finite-difference forward modelling from chapter 2.1 to obtain the synthetic data d_{syn} of one shot. These data are compared with the observed data d_{obs} from the acquisition and subtracting from each other. The obtained residuals are used in the misfit function J from chapter 2.2.1 to get a value that quantifies the difference of synthetic and observed data of this particular shot. Then, the adjoint wavefields are forward modelled to obtain the gradient of the misfit function. These steps need to be done for every shot, thus a shot loop is applied. Since the inversion is still in the first iteration, the STF inversion is applied for each shot within the shot loop. At the end of every shot loop iteration, the misfit and gradient of the corresponding shot are stored. In addition, the gradients are multiplied by the preconditioning matrix $\mathbf{A}$. Once all shots are run through, the stored misfits and gradients are summed.
At this step, a stop criterion is implemented. If the summed misfit or its change compared to previous iterations reaches a satisfactory value, the inversion aborts. If it is not sufficiently small yet, the starting model needs to be updated.
Therefore, the gradient $\nabla_m J$ is calculated and an optimal step length α is being searched (see chapter 2.2.3). Thereupon, the model is updated in the negative direction of the gradient, thus the product of α and $\nabla_m J$ is subtracted from the starting model.
This concludes one iteration. The next iteration starts with the updated model as the new starting model and the FWI steps are applied again until the stop criterion is fulfilled. One difference, however, is that the source time function inversion does not

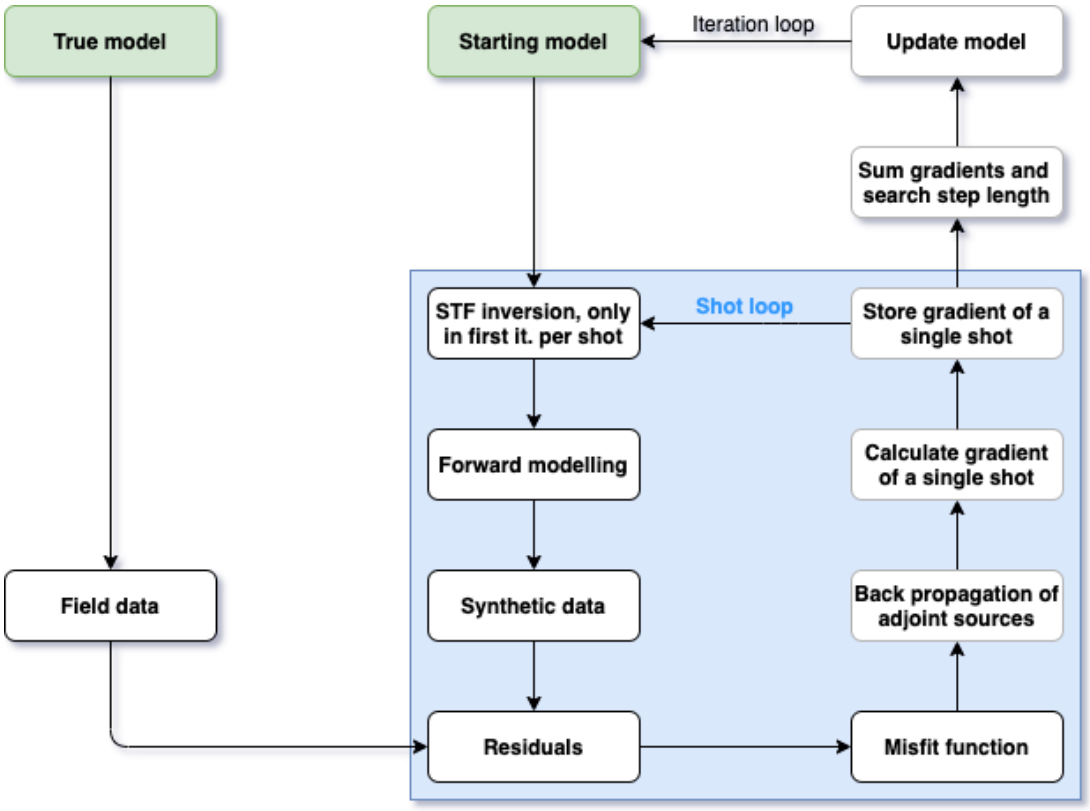


FIGURE 3.1: The full-shot workflow. The blue box highlights the loop over every single shot which happens within one iteration and the green boxes represent the starting points.

calculate the source signature again, but reuses the previously calculated signature of the corresponding shot from the first iteration. If the multi-stage approach is used from chapter 2.2.4, the STF inversion is applied at the first iteration of each stage so that the source signature's frequency is updated.

3.1.2 Random-shot workflow

Idea

In the explanation of the full-shot workflow I focus on the implementation of the shot loop and how the gradients are calculated for every shot and summed once all shots are run through. Thus, FSW has to forward model the wavefields of all shots two times per shot iteration, the first time in the forward modelling and another time in the back propagation of the adjoint sources. This means if a large number of shots are used, the computational costs become relatively large. Moreover, the step length search requires two to three more forward calculations to ensure a parabolic fit is found. Thus, the number of forward calculations per inversion iteration equals to at least four times the number of shots. This workload can be addressed with the use of shot parallelisation that is implemented in the WAVE-Toolbox. Here, each shot is calculated in a separate process. This is especially useful if a lot of nodes or computer cores are available. But because each shot uses a certain amount of storage, the storage requirements are still a limiting factor if the model size of one shot gather is medium to large.
Therefore, I try to split the single workload into smaller parts with the introduction of the random-shot workflow. The idea is to only use one shot per inversion iteration and update the model in the corresponding surrounds of this shot gather. Thus, the model

is updated in subsets instead of the whole model at once. The subsets are chosen in random order and each subset occupies the model size of one shot gather. Hence, one subset contains only one shot and the corresponding receivers. The premise is to be able to apply full-waveform inversion on a streamer acquisition with a large number of shots and a long acquisition profile. This large scale would bring huge computational costs if FSW was used. In figure 3.2 I visualise the basic concept behind RSW. The red box is the chosen subset with its corresponding source and receivers in black. The grey sources and receiver represent all remaining ones from the acquisition which are spread over the whole model. After the starting model is updated within the subset, another shot gather is chosen and the model is updated in this new area. If the new subset interlaces with the old subset, it takes over the updated area from it. This process causes sharp edges at the subset's boundary which I try to deal with by applying some smoothing on the edges, see chapter 3.2.2.

Workflow

In order to implement the idea of a subset loop and the inversion of only a single shot, the existing workflow in the WAVE-Toolbox shown in figure 3.1 must be modified. In figure 3.3 I show the implementations that are added in the toolbox. The starting points in green are the complete true Earth model and a starting model that represents the complete size which I call big model. In the first step, a shot is chosen which defines the subset with its corresponding receivers. The following steps all operate within this subset. As in FSW, a source time function inversion is applied where the Wiener filter is calculated only once at the first iteration of a stage. It is stored to be reused again if the chosen subset is iterated a second time within the same stage. Only in the next stage, the filter is calculated again with the new frequencies. After the STF inversion, the wavefields are forward calculated which yields the synthetic data. The residuals between both data are the results of the subtraction of the observed data from the synthetic data. I would like to emphasise again that this is still done within a single subset. Hence, the data of just one shot are compared to each other and the shot loop from FSW (see grey path in figure 3.3) is bypassed. The misfit and the resulting gradient are therefore not stored or summed, but directly used in the

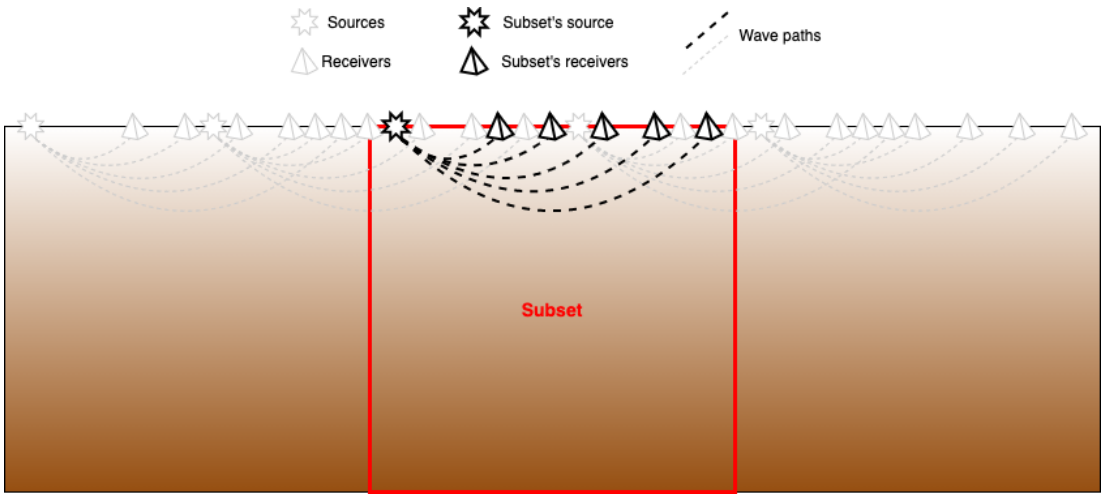


FIGURE 3.2: The concept behind RSW. Out of all available sources and receivers, one source and its corresponding receivers are randomly chosen which yields the subset (red box) where the inversion is applied on. The remaining sources and receivers from the whole acquisition are shown in grey.

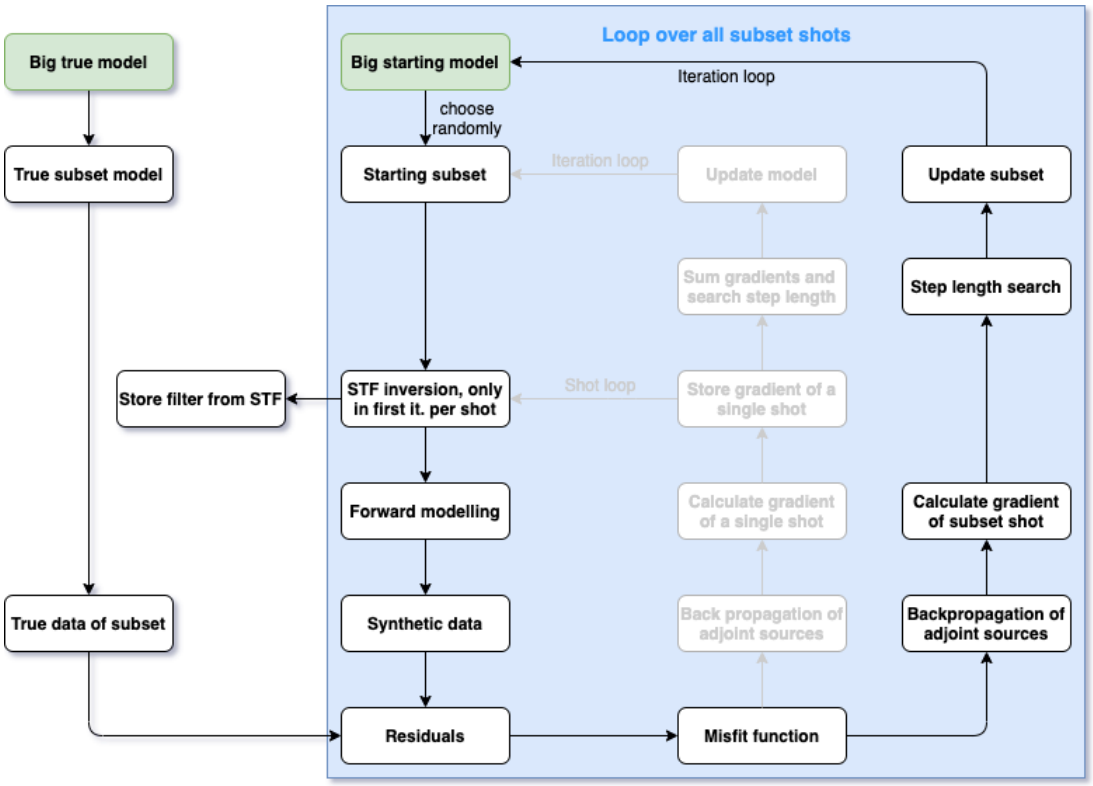


FIGURE 3.3: The random-shot workflow. The green boxes represent the starting points and the blue box highlights the loop over every single shot that corresponds to the chosen subset. This is also the loop over the iteration steps because one shot loop is one FWI iteration. The grey path, which is the full-shot workflow, is bypassed in the random-shot workflow.

step length search to update the subset model. This subset model is inserted in the big starting model which concludes one subset, shot and inversion iteration. With this method the big starting model is iteratively updated, subset by subset. This leads to a problem of a stop criterion. Because the misfit is not summed over all shots but just calculated for a single subset's shot, it does not represent the misfit of the whole model. Therefore, the inversion stops and proceeds to the next stage either after a certain number of iterations are done per stage or after each shot has been iterated sufficiently enough to produce a resolving model.

3.2 Implementation

3.2.1 Subset extraction and insertion

The WAVE-Toolbox saves the big model parameters P-wave velocity v_p, S-wave velocity v_s and density ρ in one dimensional vectors $\vec{d}_{big}$. Each vector element contains the parameter value at a certain grid point. The placement of the grid point inside the model is defined by the model coordinates. To extract a subset from the big model, a mask matrix $\mathbf{M}_{mask}$ is used. Its number of columns is the number of elements in $\vec{d}_{big}$ corresponding to the big model's size. Its number of rows equals the subset's size, such that the subset model parameter vector $\vec{d}_{subset}$ is

$$\vec{d}_{subset} = \mathbf{M}_{mask} \cdot \vec{d}_{big}. \tag{3.1}$$

The mask matrix values consists of ones at the indices that corresponds to the subset model coordinates and of zeros everywhere else. Thus, the subset vector $\vec{d}_{subset}$ contains the elements of $\vec{d}_{big}$ within the chosen subset coordinates.
The insertion of a subset in the big model is done with the transformed mask matrix $\mathbf{M}^T_{mask}$ such that

$$\mathbf{M}^T_{mask} \cdot \vec{d}_{subset} = \vec{d}_{big,update}. \tag{3.2}$$

As the values of the mask matrix do not change, the resulting big model parameter vector $\vec{d}_{big,update}$ contains zeros at the indices that correspond to the model coordinates outside of the subset. These need to be filled with the elements of the existing big model vector $\vec{d}_{big,old}$ by its summation with $\vec{d}_{big,update}$

$$\vec{d}_{big,final} = \vec{d}_{big,old} + \vec{d}_{big,update}. \tag{3.3}$$

However, this would add both values of $\vec{d}_{big,old}$ and $\vec{d}_{big,update}$ within the subset coordinates in the updated big model vector $\vec{d}_{big,final}$. Therefore, the values in the subset in $\vec{d}_{big,old}$ are erased with an erase vector $\vec{m}_{erase}$ which is filled with zeros at the indices corresponding to the subset coordinates and with ones everywhere else such that its multiplication with the big model vector erases the values within the subset

$$\vec{d}_{big,erased} = \vec{d}_{big,old} \cdot \vec{m}_{erase}. \tag{3.4}$$

The correct final big model after inserting the subset is

$$\vec{d}_{big,final} = \vec{d}_{big,erased} + \vec{d}_{big}. \tag{3.5}$$

3.2.2 Smoothing

After the subset is inserted in the big model, an artefact visible as a sharp line appears at the boundary between subset and big model. This is prevented by three smoothing implementations. The first one is a damping at the outer grid points of the subset by partially muting the subset model parameters before they are inserted. The mute is linearly increased from the inside to the outside grid point at the boundary within 10 grid points.
The second and third smoothing implementation is a convolution of the gradients and the parameters near the boundaries with a discrete approximation of the 2D-Gaussian function as the kernel following Fisher et al. (1996). The Gaussian function $G(x, y)$ in 2D is shown in figure 3.4 and has the form of

$$G(x, y) = \frac{1}{2\pi\sigma} \exp^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3.6}$$

with a variance of $\sigma = 1$. A discrete approximation by integrating the Gaussian over the whole discrete area is shown in figure 3.5 as a 7x7-matrix without the outermost rows and columns. They are truncated because their values are nearly zero, thus the matrix becomes a 5x5-matrix. The values are divided by 273, the sum of all values, to normalise them. Since the Gaussian function in equation 3.6 can be separated into x and y components, the application of the 2D convolution happens by convolving first in x direction with a 1D Gaussian approximation, shown in figure 3.6, followed by a second convolution with the 1D Gaussian in y direction.

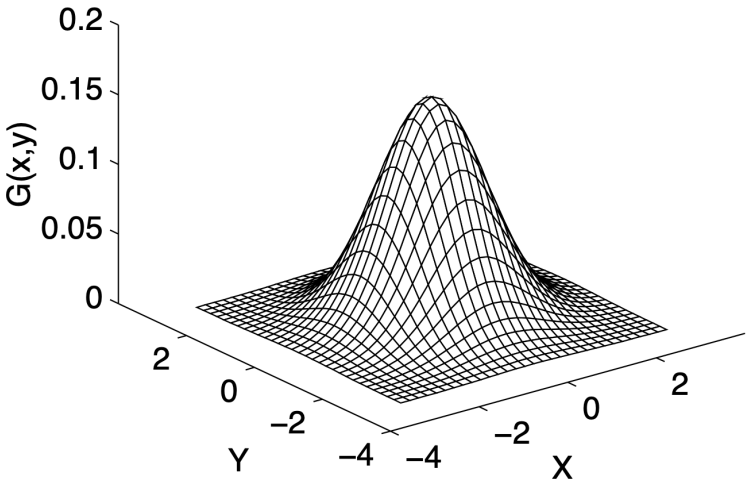


FIGURE 3.4: Gaussian function in 2-D with a variance of $\sigma = 1$by Fisher et al. (1996).

$$\frac{1}{273}\begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix}$$

FIGURE 3.5: Approximation of the Gaussian function in 2D as a 7x7-matrix without the outermost rows and columns by Fisher et al. (1996). The values are calculated by summing the Gaussian at 0.001 increments over the whole pixel. To normalize the values, they are divided by 273, the sum of all values.

.006	.061	.242	.383	.242	.061	.006

FIGURE 3.6: The 1D convolution kernel to obtain the approximation of the Gaussian function in 2-D as shown in figure 3.5 by Fisher et al. (1996).

Chapter 4

Synthetic data inversion

The purpose of full-waveform inversion with synthetic data is the controlled environment with a free choice of model parameters, where a benchmark can be configured. In this chapter I would like to show such a benchmark, where I define a true model that is used to create synthetic seismograms which are used as the "field" data in the inversion. This is done with the full-shot workflow as well as with the random-shot workflow which allows me to compare their results. The goal is to verify the functionality of RSW and ensure that it produces comparable results to FSW.

4.1 Inversion of synthetic data

4.1.1 Methodology

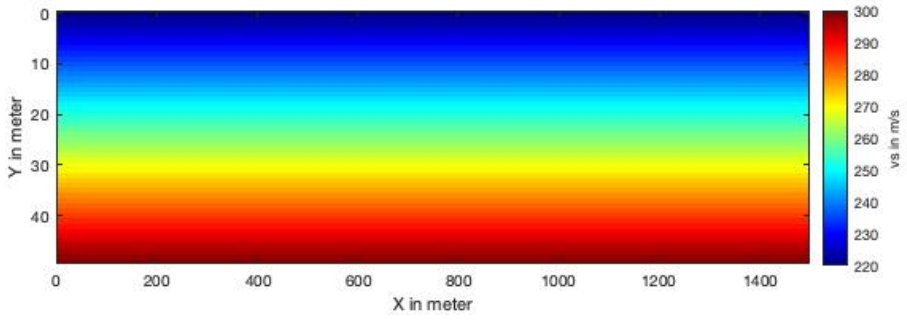
In the following chapter 5 I explain why the penetration depth of the surface waves is limited in the field data and consequently also in the produced synthetic data. Thus, these waves are dominating the data at shallow depth. Therefore, I perform only mono-parameter full-waveform inversions in this work where only one model parameter, the S-wave velocity, is investigated and the other parameters, the P-wave velocity and the density, are constant. This also simplifies the objective function to ensure an easier convergence towards an improved model. To cover different use cases, I perform two inversion runs, each with a different model. The starting models are shown in figure 4.1a and 4.1b which both have a linearly increasing background S-wave velocity v_s with depth from $220\,\mathrm{m\,s^{-1}}$ to $300\,\mathrm{m\,s^{-1}}$ in case of the first model and $200\,\mathrm{m\,s^{-1}}$ to $350\,\mathrm{m\,s^{-1}}$ in case of the second model. The first model consists of several high and low v_s anomalies with circular and rectangular shapes lying in different depths, see in figure 4.2a. The velocities range from $200\,\mathrm{m\,s^{-1}}$ to $320\,\mathrm{m\,s^{-1}}$. Due to the stretched y-axis, the circular shapes appear as ellipses. The background velocity is unchanged from the starting model. The second model is a layer model, shown in figure 4.2b. It consists of four layers which increase with depth by $50\,\mathrm{m\,s^{-1}}$ per layer from $200\,\mathrm{m\,s^{-1}}$ to $350\,\mathrm{m\,s^{-1}}$. Due to the mono-parameter FWI, the P-wave velocity v_p and density ρ parameter do not differ between starting model and true model. They are displayed in figure 4.2c and 4.2d, respectively, where v_p linearly increases with depth from $700\,\mathrm{m\,s^{-1}}$ to $1200\,\mathrm{m\,s^{-1}}$ and the density ρ is homogeneously constant at $2000\,\mathrm{kg/m^3}$. All model parameter values are approximately based on the results of Liberty, St Clair, and Gribler (2018), see chapter 5.2, instead of rough estimations from previous studies, such as the choice of v_p that is higher than the definition

$$\frac{v_p}{v_s} = \sqrt{\frac{1-\nu}{\frac{1}{2}-\nu}} \tag{4.1}$$

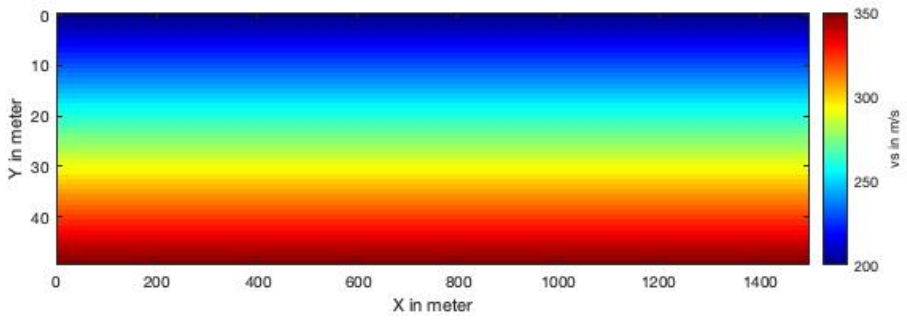
would suggest if the Poisson's ratio ν was chosen to be constant at 0.25 according to Reynolds (1997). This guarantees that the synthetic model is comparable to the Earth model in Salt Lake City. As the parameters for the finite-difference solver, I use a second order in time and an eighth order in space. Based on the order, the model discretisation should follow equation 2.26 to ensure numerical dispersion is minimised. Because of the multi-stage approach, the last stage of the inversion is limited to 50 Hz. Thus, a grid spacing of 0.5 m is chosen for both vertical and horizontal direction in case of RSW. The temporal sampling rate is 0.2 ms and based on equation 2.27. The recording time of the modelling is 0.9 s. The total grid size is 1500×100 grid points of which 10 grid points at the boundary are used as a buffer zone for the damping by C-PML. The subset in the random-shot workflow utilises 168×100 grid points including the 10 grid points as buffer zones. A total of 135 shots with a spacing of 10 m between each shot are used with 48 receivers spaced by 1.25 m resulting in an aperture of 60 m. The geometry is based on the acquisition of the Salt Lake City field data, thus the receivers and sources are placed at the surface. As a source signature, a shifted Ricker wavelet s_{ricker} is used with a center frequency of $f_c = 30$ Hz. In the WAVE-Toolbox, it is defined as

$$s_{ricker} = \left(-2\left(\pi f_c\left(t - \frac{1.5}{f_c}\right)\right)^2\right) \cdot \mathrm{e}^{-\left(\pi f_c\left(t-\frac{1.5}{f_c}\right)\right)^2} \tag{4.2}$$

The choice of 30 Hz is based on the frequency content of the Salt Lake field data that I show in chapter 5.3. Besides, the synthetic inversion uses the multi-stage approach with 5 stages starting from 10 Hz to 50 Hz with 10 Hz intervals. As an abort criterion, FSW stops the stage and proceeds to the next one if the misfit value does not change by more than 0.1 from the second last iteration. In RSW's case, this is done after 700 iterations which means every shot is calculated on average approximately five times.

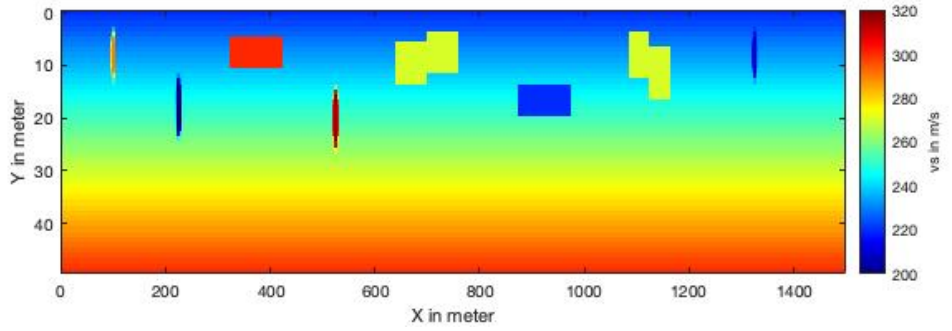


(A) Starting v_s anomaly model

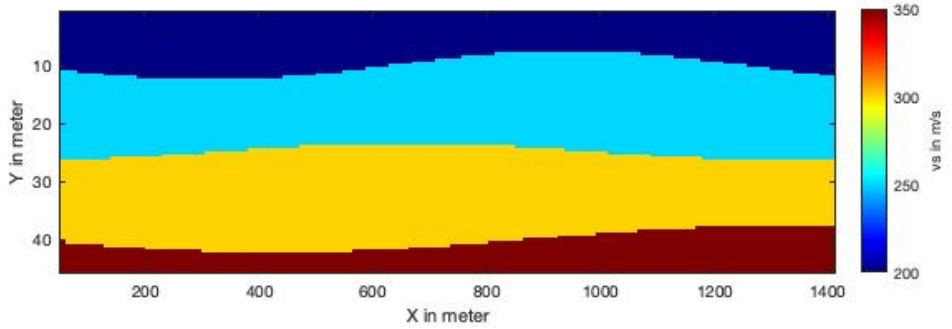


(B) Starting v_s layer model for the synthetic data inversion

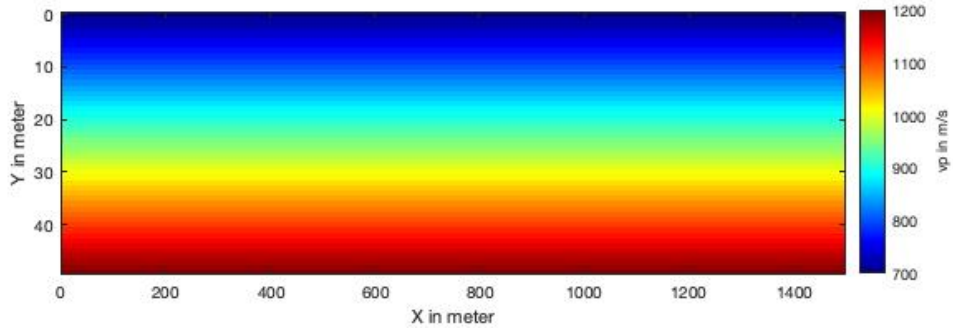
FIGURE 4.1: Starting anomaly (A) and layer (B) v_s models for the synthetic data inversion.



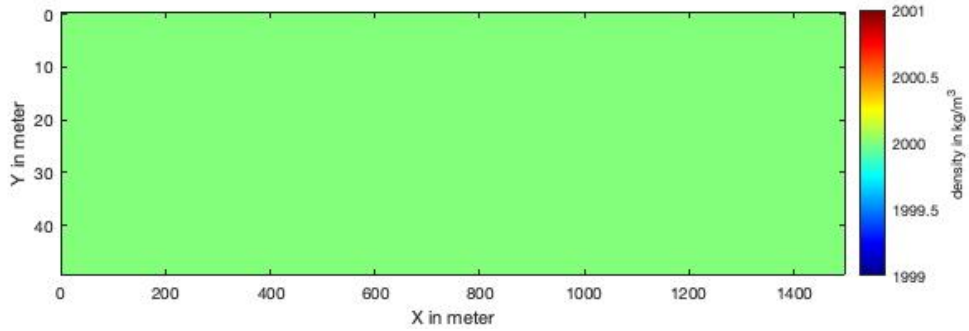
(A) True v_s model with high and low velocity anomalies



(B) True v_s model in form of a layer model



(C) True and starting v_p model



(D) True and starting density model

FIGURE 4.2: True models for the synthetic data inversion. (A) is the v_s anomaly model, (B) the v_s layer model, (C) the v_p model for both and (D) the density model for both.

4.1.2 Full-shot workflow results and discussion

Anomaly model

At first I discuss the results of the anomaly model with the use of FSW. The v_s model after the last iteration and the misfit evolution together with the final seismogram of shot 70 are shown in figure 4.3 and 4.4, respectively. Comparing the model in figure 4.3 with 4.2a, all anomalies are well reconstructed. The small circles are well contoured to be identified as anomalies and the two connecting rectangles at around $x = 700\,\text{m}$ and $x = 1100\,\text{m}$ are also well visible. In greater depths, the background velocity is slightly distorted, but not to a significant amount. The misfit evolution in figure 4.4 supports the good result. It is the misfit summed over all shots and shows a smooth decrease in each stages except the last two ones. In those stages, not many iterations are run and the abort criterion is reached early. With the start of a new stage, the misfit raises significantly. This is expected because a new stage introduces higher frequency content that has not been considered before. Thus, the inversion starts from the beginning again where the misfit is minimised. Figure 4.4 also shows a normalised sample seismogram of shot 70 after the last iteration with its source at $x = 720\,\text{m}$. The modelled data in red fit the synthetic field data in black well with no large difference.

Layer model

The results of the layer model are shown in figure 4.5 and 4.6. The v_s model after the last iteration in figure 4.5 shows the first two layers with their homogeneous velocity quite clearly. The boundaries are well defined and correspond to the true layer model in figure 4.2b. However, the two deeper layers appear a bit undefined due to the remains of the gradient background velocity from the starting model in figure 4.1b. This is most apparent in the deepest layer where the upper boundary is almost not visible. This happens because the amplitudes of the surface waves, that dominate the data due to their large amplitude, decay exponentially with depth, hence greater depths cannot be well imaged. The large amplitudes of the surface waves can be observed in figure 4.6. There, the misfit is plotted together with a normalised sample seismogram of shot 70 after the last iteration with its source at $x = 720\,\text{m}$. As with the anomaly model, the misfit decreases smoothly again in the early stages while not many iterations are performed in the later stages. Moreover, the same phenomenon of higher misfits at the first iteration of a new stage can be seen here as well. The seismograms show a good fit between modelled data in red and the synthetic data traces in black.

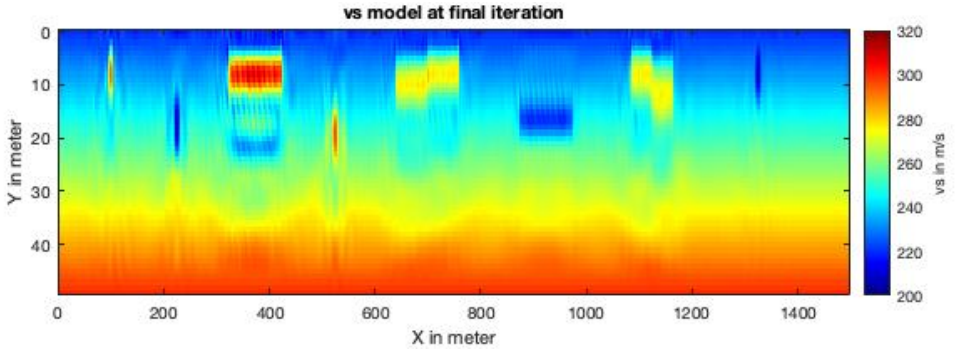
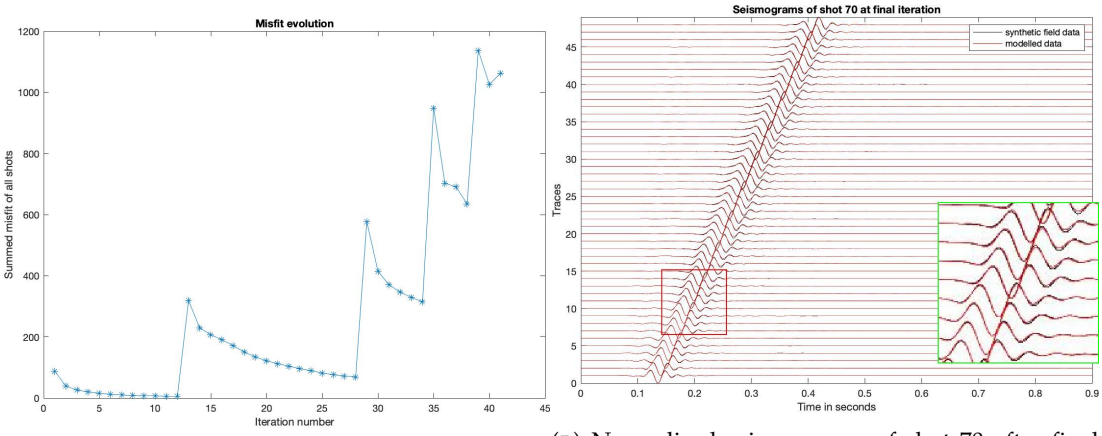


FIGURE 4.3: Final FSW v_s anomaly model of the synthetic data inversion after the last iteration.



(A) Evolution of the misfit summed over all shots

(B) Normalised seismograms of shot 70 after final the iteration

FIGURE 4.4: Evolution of the misfit summed over all shots and normalised seismogram of shot 70 with its source at $x = 720\,\text{m}$ after the final iteration of the FSW inversion of the anomaly model. Both plots correspond to the result in figure 4.3.

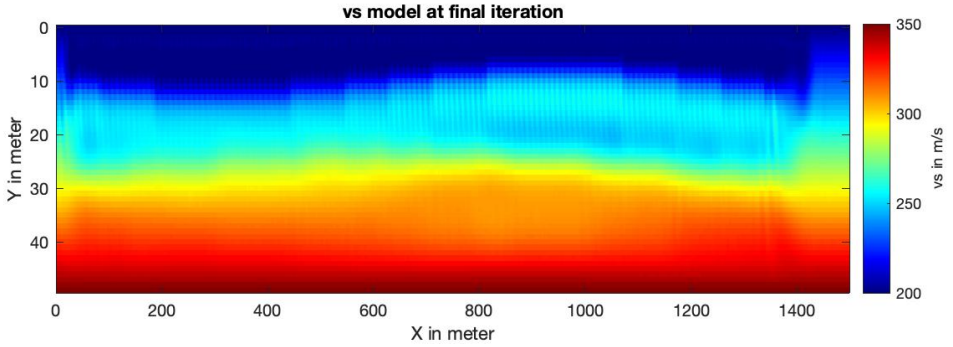
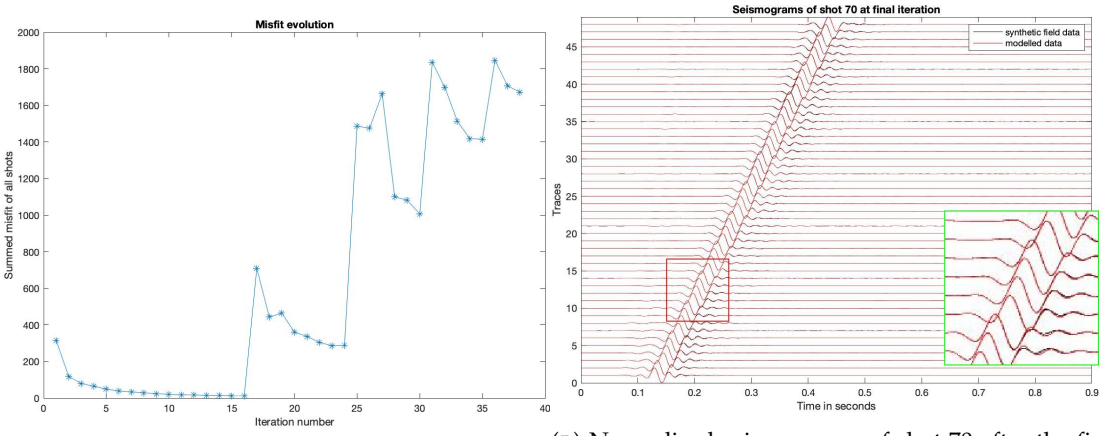


FIGURE 4.5: Final FSW v_s layer model of the synthetic data inversion after the last iteration.



(A) Evolution of the misfit summed over all shots

(B) Normalised seismograms of shot 70 after the final iteration

FIGURE 4.6: Evolution of the misfit summed over all shots and normalised seismogram of shot 70 with its source at $x = 720\,\text{m}$ after the final iteration of the FSW inversion of the layer model. Both plots correspond to the result in figure 4.5.

4.1.3 Random-shot workflow results and discussion

Anomaly model

In this section I would like to discuss the results obtained with the use of the random-shot workflow. The v_s model after the last iteration is shown in figure 4.7. At first glance, most anomalies are reconstructed well, albeit their shapes are not as clear as in the true model in figure 4.2a. This is especially noticeable for the rectangle at $x = 400\,\text{m}$ and for the circle at $x = 500\,\text{m}$. The latter also shows a lower velocity than in the true model. It could be hard to identify if the true model was unknown and noise was apparent. Despite the unclear shapes of the anomalies, the normalised sample seismograms of shot 70 with its source at $x = 720\,\text{m}$ in figure 4.8 shows a good fit between synthetic field data in black and modelled data in red. Almost no difference can be observed which suggests that the model at this particular subset represents the true model well enough.

Layer model

The inversion result of the layer model is shown as the v_s model after the final iteration in figure 4.9. Compared to the true layer model in figure 4.2b, the first boundary between the first two layers is clearly reconstructed. It is noticeable that the curvature of the boundary is well preserved as well as the homogeneity within the first two layers. However, the deeper boundaries between the layers become less defined due to the same reason as in the FSW results in chapter 4.1.2. Starting from 20 m depth, the background velocity from the starting model in figure 4.1b becomes more present and smooths out the boundaries. Moreover, the velocity within the last two layers is not as homogeneous as in the true model. Looking at the normalised sample seismogram in figure 4.10, the fit between synthetic field data in black and modelled data in red is good for the majority of the waveform. Only at the later arrivals the waveform some differences can be observed, albeit arguably negligible.

4.2 Comparison of the synthetic results

As discussed at the beginning of this chapter, the purpose of the comparison is to ensure the results of the random-shot workflow are comparable to the results of the full-shot workflow. Looking at the final v_s models of the anomaly model in figure 4.3 and 4.7, the results indeed look similar. However, FSW yields a sharper image with clearer lines at the boundaries of the anomalies, especially noticeable at the anomaly

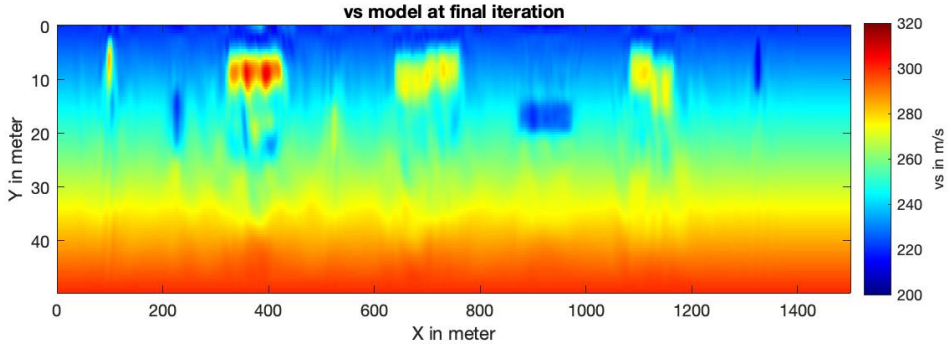


FIGURE 4.7: Final RSW v_s anomaly model of the synthetic data inversion after the last iteration.

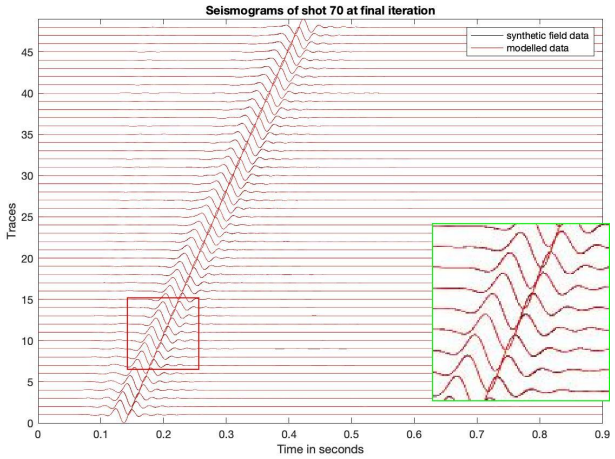


FIGURE 4.8: Normalised seismograms of shot 70 after the final iteration of the RSW inversion of the anomaly model with its source at $x = 720\,\text{m}$. It corresponds to the result in figure 4.7.

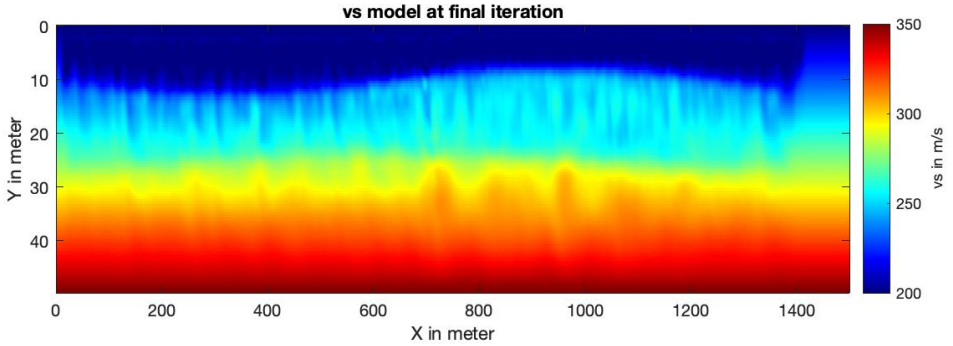


FIGURE 4.9: Final RSW v_s layer model of the synthetic data inversion after the last iteration.

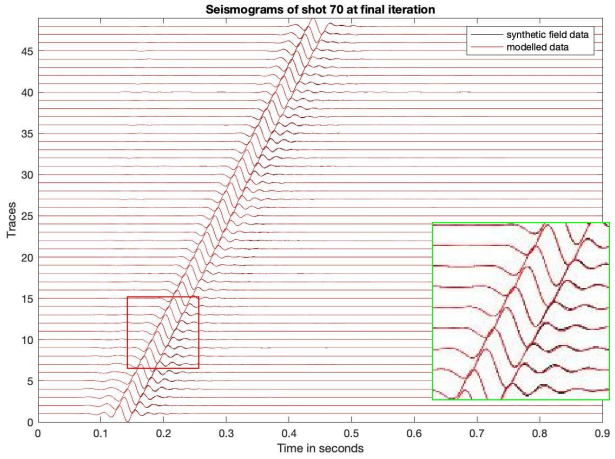


FIGURE 4.10: Normalised seismograms of shot 70 after the final iteration of the RSW inversion of the layer model with its source at $x = 720\,\text{m}$. It corresponds to the result in figure 4.9.

at $x = 500\,\text{m}$. There, FSW can accentuate the circle clearly from the background velocity, whereas RSW struggled to do so, resulting in a blurry image of the anomaly. Nonetheless, RSW yields a comparable model besides this detriment.
The results of the layer model in figure 4.5 and 4.9 show similar conclusions. FSW yields a sharper image again, especially at greater depths where RSW failed to image

the layers' boundaries, albeit FSW also struggled in this regard. Due to the smoothing in the implementation of RSW, the blurriness turns out to be unproblematic for a homogeneous layer model at shallow depths.
In the seismograms of a sample shot in figure 4.4, 4.6, 4.8 and 4.10, both workflows yield satisfactory results with a good fit and little to negligible differences between the synthetic field data and modelled data. On average, though, RSW manages to fit the data in the seismograms slightly better than FSW does, even though the RSW's final model is not as highly resolving. This can probably be explained by the focus of a single shot that happens during the model update of RSW. Therefore, finer adjustments can be done in RSW than in FSW where the focus is not on a single shot, but averaged over all shots and the whole model.
In terms of a computational comparison, RSW manages to finish FWI quicker than FSW. While FSW needs around 25.4 hours, RSW required approximately 18.1 hours. Both workflows utilise the Geophysical Institute high performance computer (GPIHPC). A more in-depth analysis of the computing times of both RSW and FSW with field data is done in chapter 6.2.6.
Before I continue with the inversion of the field data, I would like to show a brief introduction to these data, where and how they were acquired and how I preprocessed them before they can be used in FWI.

Chapter 5

Salt Lake City field data

In this chapter I describe the field data that are used in the full-waveform inversion as the observed data for both the full-shot workflow and the random-shot workflow. I go through the acquisition, explain what makes it a difficult data set for FWI and show briefly the results from Liberty, St Clair, and Gribler (2018) who did a tomography based on the data. At the end of this chapter, I explain how I preprocessed the data to be used in the FWI workflows of this work.

5.1 Acquisition

5.1.1 Geological background

Salt Lake City lies right on top of the Wasatch fault zone, an active normal fault zone with a length of 343 km in the U.S. state of Utah. It is composed of 10 discrete segments of which five are active (Machette et al., 1991). One of the segments is called the Salt Lake City Fault Segment which can be divided into three en-echelon normal faults, the Warm Springs, East Bench, and Cottonwood faults (Liberty, St Clair, and Gribler, 2018). Salt Lake City's downtown, where the acquisition took place, lies right between the East Bench and Warm Springs faults. In this work, I only used part of the data that were acquired at Salt Lake City in 2017. This part is referred to as the 500 South profile and lies in the south-east of the downtown district and upon Holocene lake deposits (McKean, 2014). A map with the profile's position is shown in figure 5.1.

5.1.2 Geometry

The data were acquired by a seismic land streamer and a 200 kg accelerated weight drop source system. The streamer consists of 1.25 m-spaced 48 channels made of two-components 4.5 Hz geophones, resulting in a 60 m aperture of one shot gather with the nearest offset being at 10 m. I only use the vertical component data in this work for FWI. The geophones were mounted on flat baseplates made to be used on paved city streets. The 500 South profile has 655 shots spaced by 2 m, thus a profile length of 1370 m. During the acquisition, the traffic was not stopped and continued along profile.
This induces a lot of noise during the acquisition that I deal with in the preprocessing. Furthermore, the end-on spread geometry of the streamer causes unfavourable conditions for FWI. In each shot gather, the receivers are only on one side of the source, thus the wavepaths always come from the a single side during the acquisition, not from more sides. An illustration of the end-on spread geometry of a single shot is shown in figure 5.2. It also displays the aperture of 60 m which limits the possible depth that can be imaged. The depth is further limited due to the frequency content of the accelerated weight drop source.

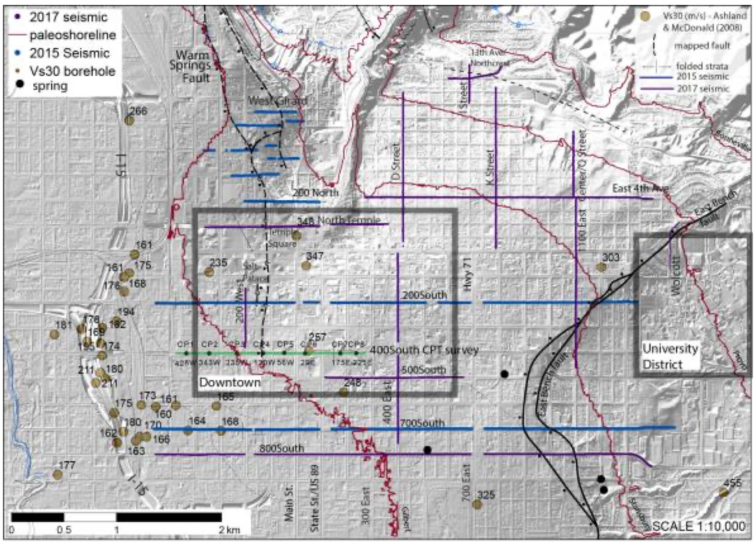


FIGURE 5.1: Lidar map for Salt Lake City with the seismic profile, earlier Vs30 results (circles), and fault locations, taken from Liberty, St Clair, and Gribler (2018). The profile used in this work comes from 500 South within the marked downtown district.

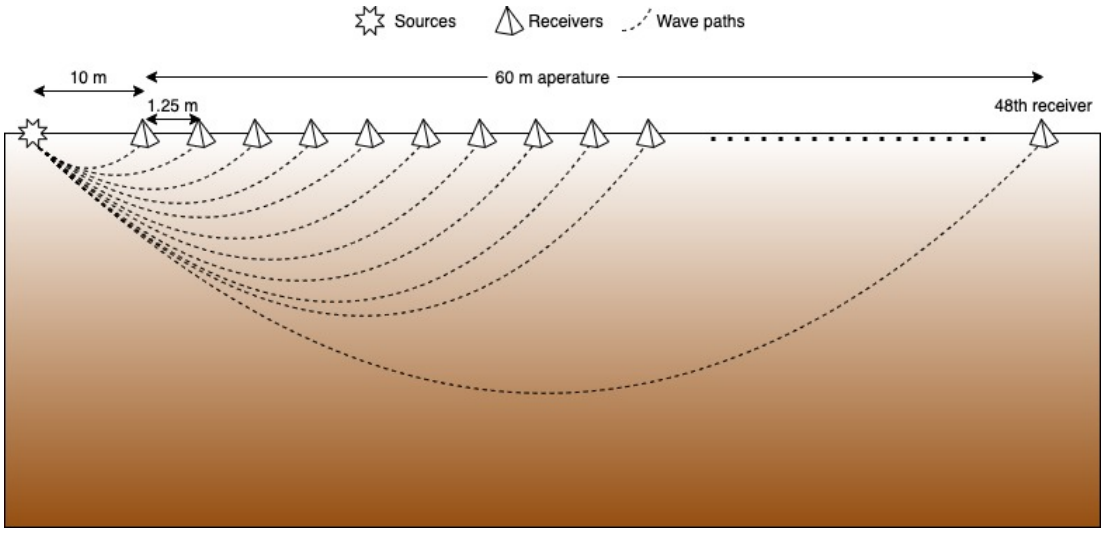


FIGURE 5.2: End-on spread geometry of a single shot used to acquire the Salt Lake City field data.

5.2 Results from Liberty et al.

Liberty, St Clair, and Gribler (2018) use a modified tomography code of St Clair (2015) to image the subsurface of the 500 South profile. Their inversion only uses the travel-time information of the data to update the model where the aim is to reduce the root mean square misfit between field and modelled data. To obtain a starting model for the S-wave velocities v_s, they identified the fundamental mode Rayleigh wave dispersion picks from about 5-50 Hz by using an approach developed by Gribler et al. (2016). The v_s tomography results are shown in figure 5.4. At around 700 m to 1100 m along the profile, the area shows faster S-wave velocities compared to the adjacent areas. This area is marked with the left red arrow and lies where a lateral spread deposit is mapped, see figure 5.3 and McKean (2014). Previous excavations revealed numerous high-angle faults and other deformations in this area that are related to the

lateral spread deposit and not directly to tectonic faults (Scott and Shroba, 1985). Furthermore, tip resistance measurements for the CP-9 and CP-10 surveys also revealed a stiffer ground which correlates with faster v_s values. The velocities in the tomogram are classified by the National Earthquake Hazards Reduction Program (NEHRP) where the D class is specified as stiff soil.
I use the acquired data of the lateral spread deposit area in the following comparisons between FSW and RSW in chapter 6. The result of the whole profile in figure 5.4 is also used to compare the inversion of the complete field data in chapter 6.2.5. The goal is to improve the resolution of the final v_s model.

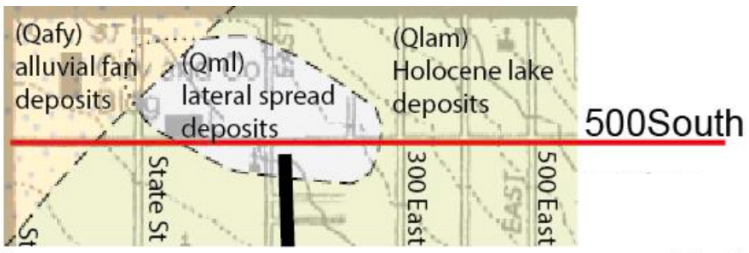


FIGURE 5.3: Mapped geology beneath the 500 South profile by McKean (2014).

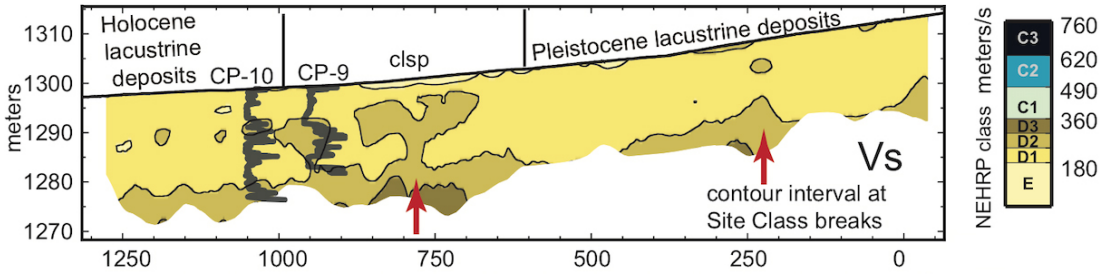


FIGURE 5.4: v_s tomogram with marked tip resistance measurements for CP-9 and CP-10 surveys by Liberty, St Clair, and Gribler (2018). The red arrows show an increase of v_s compared to the adjacent areas.

5.3 Preprocessing

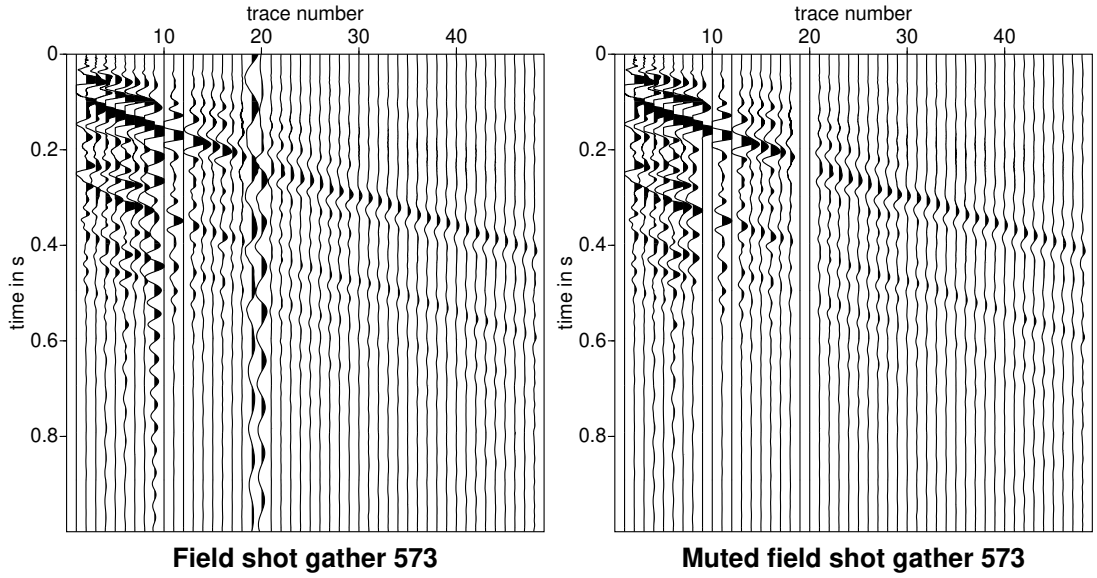
As mentioned earlier, only the vertical component data are used in FWI. Before they can be used, though, several preprocessing steps are applied which is done in Seismic Unix and Matlab. Each SU file corresponds to one shot gather. Each of them is examined individually to quality check the data and to improve the signal-to-noise ratio. This is done by removing noisy signals, for instance by partially muting or killing traces that are badly coupled to the ground or by removing signals in the F-K domain. Two examples of the noise removal are shown in figures 5.5 and 5.6. In figure 5.5, the bad coupling of channel 19 and 20 is easy to identify and kill. Channel 9 shows coherent noise as well starting from approximately 2 s, hence, it is partially muted. Figure 5.6 shows how a signal that comes from the end of the streamer is removed in the F-K domain. The signal's propagation velocity is approximately calculated by its slope in the seismogram. With the velocity, the corresponding area in the F-K domain can therefore be muted to remove the signal in the T-X domain, see figure 5.7. Moreover, the 1st, 10th and 12th traces always show minimal to no recordings in all shots. Thus, they are permanently killed and not used during FWI.

After the quality check of the data, a low-pass filter with a cut-off frequency at 70 Hz and a taper length of 20 Hz is applied to remove noise at higher frequencies. In addition, the data are time shifted by 0.03 s to allow the source time function inversion to apply the filtered source signature at the first few traces without becoming non-causal. This is shift is corrected again in FWI by adding a time shift of 0.03 s to the source signature. The resulting frequency spectrum after the noise removal and low-pass filter is shown in figure 5.8. The peak frequency is at about 24 Hz, while the frequency content below 5 Hz and above 70 Hz, due to the lowpass filter, is strongly attenuated compared to the rest of the spectrum.
Furthermore, another adjustment needs to be applied to the data. I use the 2D adjoint waveform inversion in this work which inherently uses a line source in 3D space. Because the wave simulations are not done in 3-D, the spherical divergence of the waves are not properly simulated in the 2D model. This contradicts the wave propagation in the Earth during the acquisition that happened in 3D and are caused by a point source, the accelerated weight drop. Therefore, the field data are adjusted by converting the waveforms in a such way as if they were generated by line sources. Based on the theoretical background in Forbriger, Groos, and Schäfer (2014), I apply two factors F_{amp} and $\tilde{F}_{\sqrt{t^{-1}}}(\omega)$ to the amplitudes and phases of the data following the suggestions by Schäfer et al. (2014)

$$F_{amp} = r\sqrt{\frac{2}{t}} \tag{5.1}$$

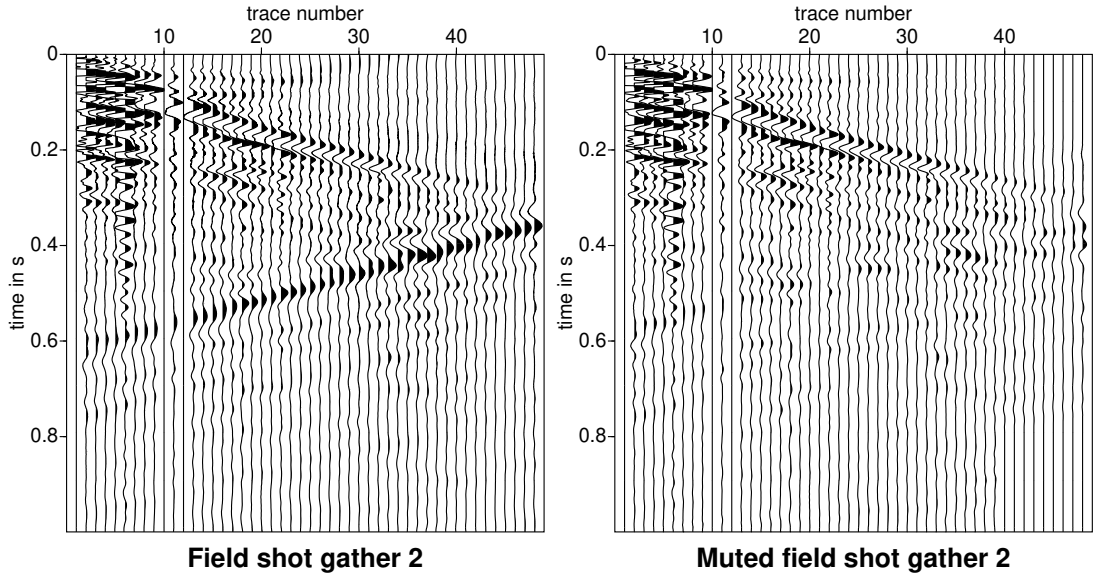
$$\tilde{F}_{\sqrt{t^{-1}}}(\omega) = \sqrt{\frac{\pi}{\omega}} e^{i\pi/4} \tag{5.2}$$

with r being the source to receiver distance and ω the angular frequency. The amplitudes are multiplied by F_{amp} which is the suggested factor for waves travelling at shallow depths, and the factor $\tilde{F}_{\sqrt{t^{-1}}}(\omega)$ is applied in the frequency domain which essentially applies a phase shift of $\pi/4$ and results in a convolution with $\sqrt{t^{-1}}$ in the time domain. An example of this correction on shot 70 after the noise removal is shown in figure 5.9. The amplitudes of the corrected traces in red have a slightly higher amplitude and show a phase shift as expected.



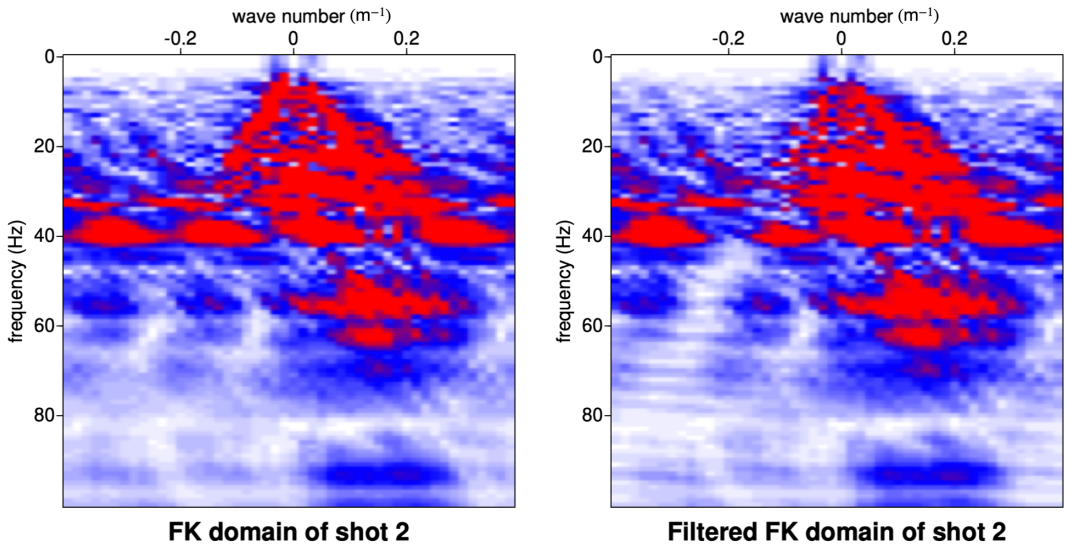
(A) Original seismograms of shot 573 (B) Seismograms of shot 573 with removed noise

FIGURE 5.5: Noise removal of shot 573 from the Salt Lake City field data. (A) is the original seismograms, (B) the preprocessed seismograms where traces 19 and 20 are completely killed and trace 9 partially.



(A) Original seismograms of shot 2 (B) Seismograms of shot 2 with removed noise

FIGURE 5.6: Noise removal of shot 2 from the Salt Lake City field data. (A) is the original seismograms, (B) the preprocessed seismograms where noise is removed in the F-K domain.



(A) Original F-K domain of shot 2

(B) F-K domain of shot 2 with removed noise

FIGURE 5.7: F-K domains shot 2 from the Salt Lake City field data with noise in (A) and with removed noise in (B). The noise is identified by its velocity that can be approximately calculated from its slope in the seismogram in figure 5.6a.

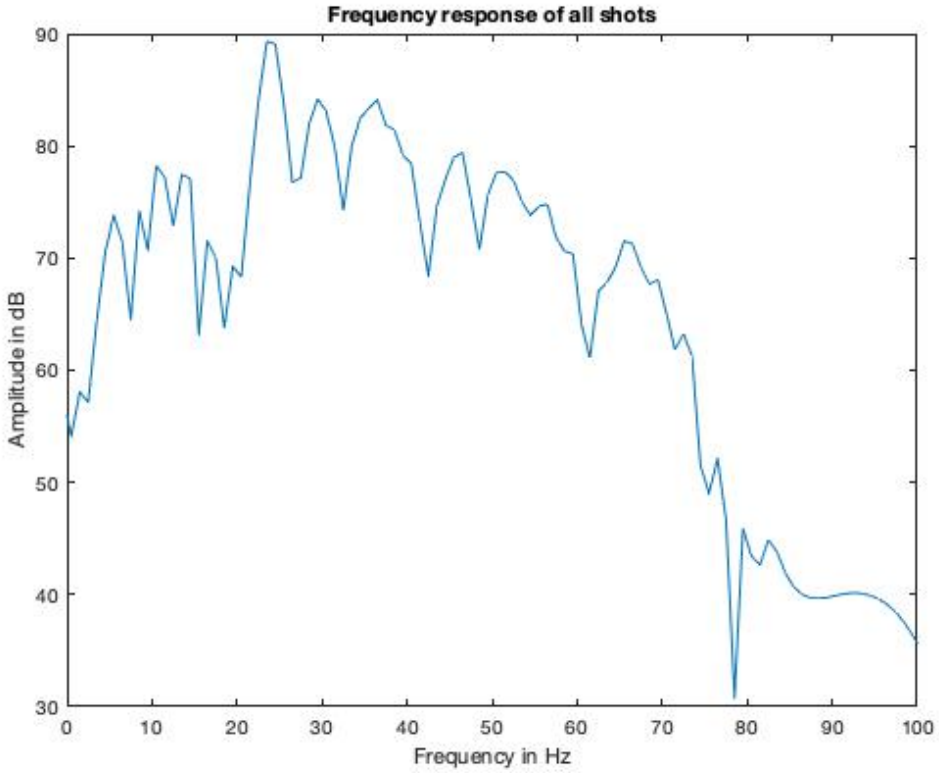


FIGURE 5.8: Amplitude spectrum averaged over all shots from the Salt Lake City field data after noise removal and application of a lowpass filter.

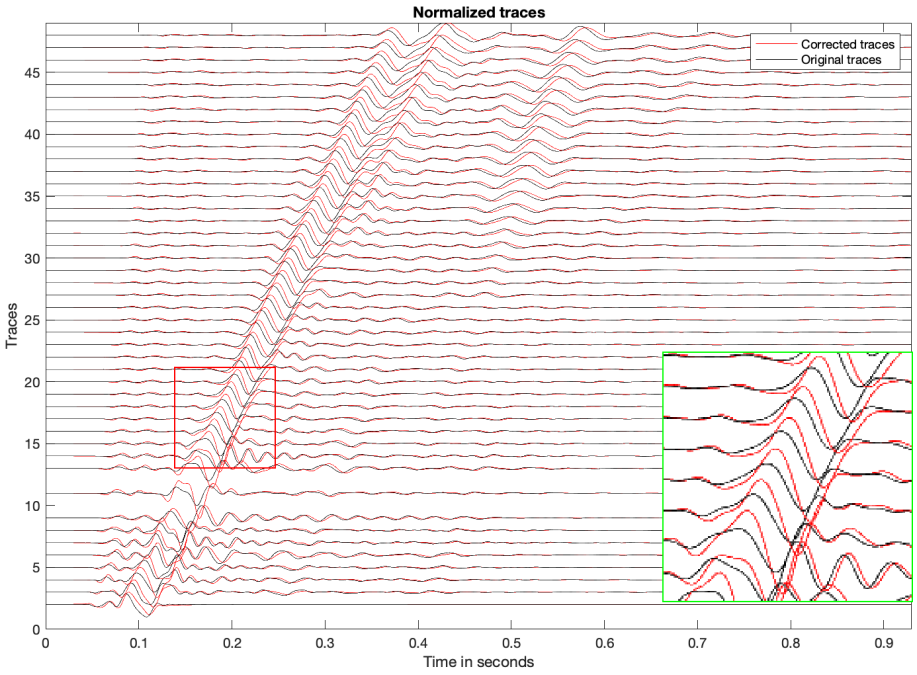


FIGURE 5.9: Point to line source correction of shot 70 with normalised traces. Due to the correction, amplitudes are slightly increased and a phase shift is applied.

Chapter 6

Field data inversion

In this chapter I focus on the comparison between the full-shot workflow and the random-shot workflow using the Salt Lake City field data. At first, I explain the initial starting model that I use for FWI. Afterwards, I use part of the field data to analyse the workflows' general sensitivity towards certain parameters. I also compare the workflows based on the inversion of the complete data set and how their computational costs differ. The goal is to find the best workflow to be used for large land streamer data such as the Salt Lake City data.

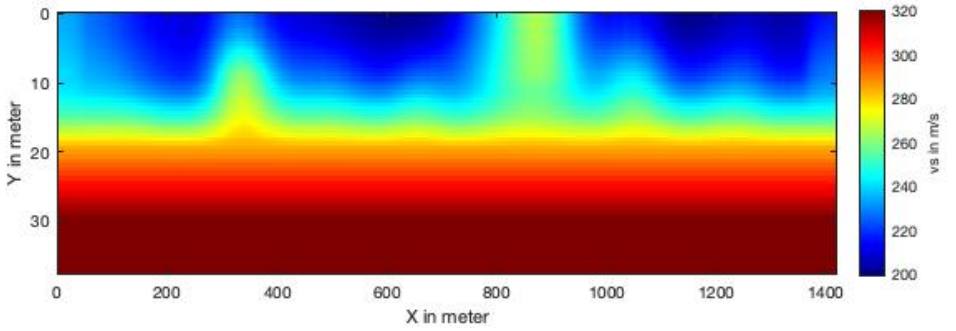
6.1 Initial model

The inversion starts with an initial model. The properties of such a model are ideally as close to the true model as possible to ensure that the FWI algorithm can minimise the objective function. If the starting model vastly differs from the true model, the convergence towards the true model, the global minimum in the objective function, is very difficult. Thus, the multichannel analysis of surface waves is used to build the initial model. This method provides v_s information of the near surface by an inversion of Rayleigh wave dispersion curves whilst being highly cost-effective (Park et al., 2007). Since my work is not focused on the theory behind this method, For a detailed explanation of its sequential application with FWI I refer to Pan, Gao, and Bohlen (2019). The images of the initial v_s, v_p and density models used for both the FSW and RSW are shown in figure 6.1a, 6.1b and 6.1c, respectively. They all share a grid spacing of 0.625 m and a model size of 2273×61 grid points. The v_s model shows similar high-velocity areas at shallow depths at around $x = 250\,\mathrm{m}$ and $x = 850\,\mathrm{m}$ as can be seen in the v_s tomogram from Liberty, St Clair, and Gribler (2018) in figure 5.4. Due to the choice of a mono-parameter FWI, the v_p and density models in 6.1b and 6.1c are laterally homogeneous gradient models. The P-wave velocity ranges from $500\,\mathrm{m\,s^{-1}}$ to $1700\,\mathrm{m\,s^{-1}}$ while the density linearly increases with depth from $1800\,\mathrm{kg/m^3}$ to $2250\,\mathrm{kg/m^3}$ within the first 18 m. From a depth of 18 m and onwards, the density stays homogeneous.

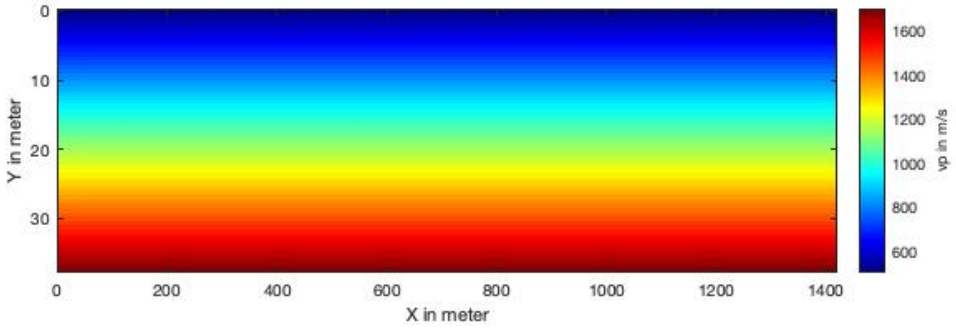
6.2 Comparison of the full-shot workflow and random-shot workflow

6.2.1 Methodology

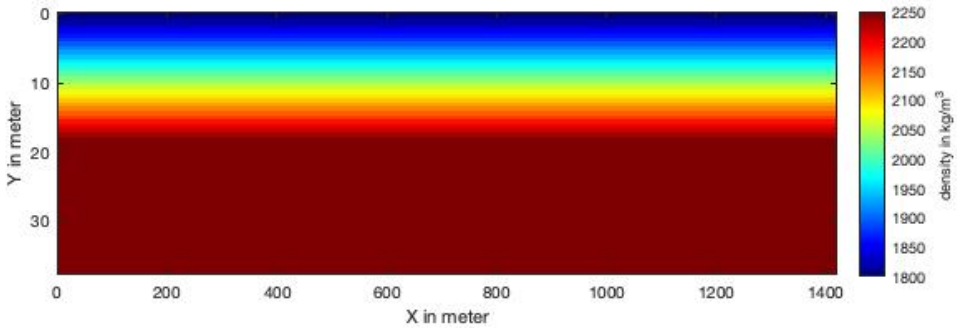
As I mention in chapter 5.1.2, the Salt Lake City field data set contains a total of 655 shots and a profile length of over 1300 m. This large data set would require long computing times due to the high computational costs of FWI. Therefore, I choose to do a major comparison between FSW and RSW only with a subset of the complete data.



(A) Initial v_s model



(B) Initial v_p model



(C) Initial density model

FIGURE 6.1: Initial v_s (A), v_p (B) and density (C) model in the full-waveform inversion of the Salt Lake City field data.

The area of interest is the high-velocity area between $x = 800\,\text{m}$ and $x = 940\,\text{m}$, marked in figure 6.2, and includes the lateral deposit area, see figure 5.3. The subset consists of 28 shots and has a length of approximately 140 m.
To keep the methodology consistent with the synthetic benchmark, I set similar parameter values. The grid spacing Δh is adjusted from 0.5 m to 0.625 m to match the model parameters from the multichannel analysis of surface waves and thus, the FD order is decreased from 8th to 6th order in space and kept as 2nd order in time to avoid numerical dispersion. The temporal sampling rate Δt is kept at 0.2 ms with a simulation time of 0.93 s. The shot and receiver geometry is taken from the field data acquisition that I described in chapter 5.1.2. Hence, shots are spaced by 2 m and each shot has 48 channels spaced by 1.25 m. The model includes 10 grid points at the boundaries as buffer zones for the C-PML damping.
Unlike the synthetic benchmark, the source signature is unknown for these field data. Therefore, the source time function inversion is utilised. The resulting signal is tapered

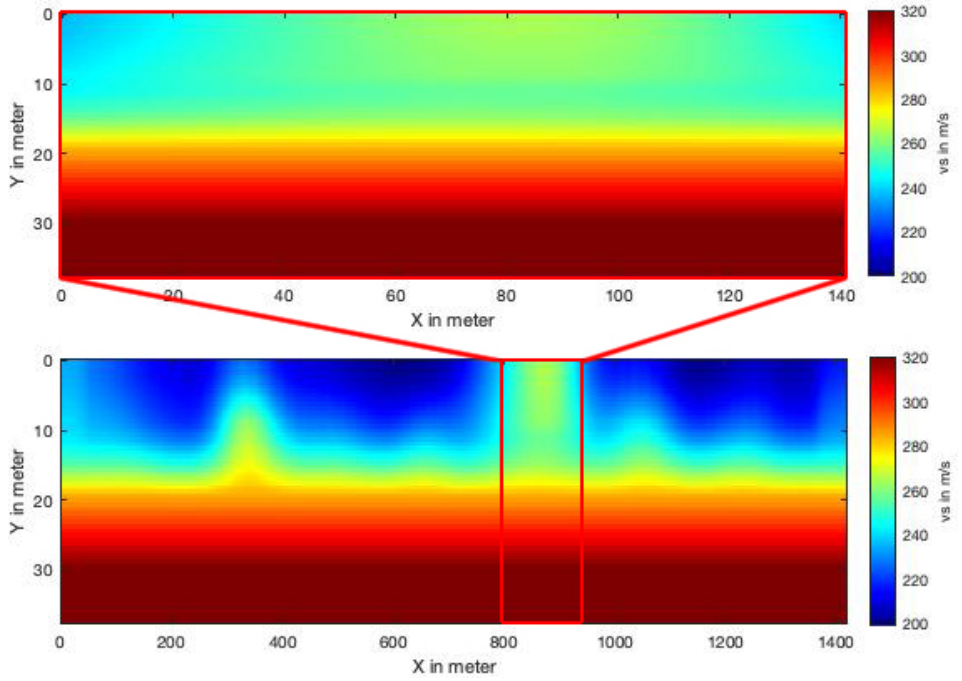


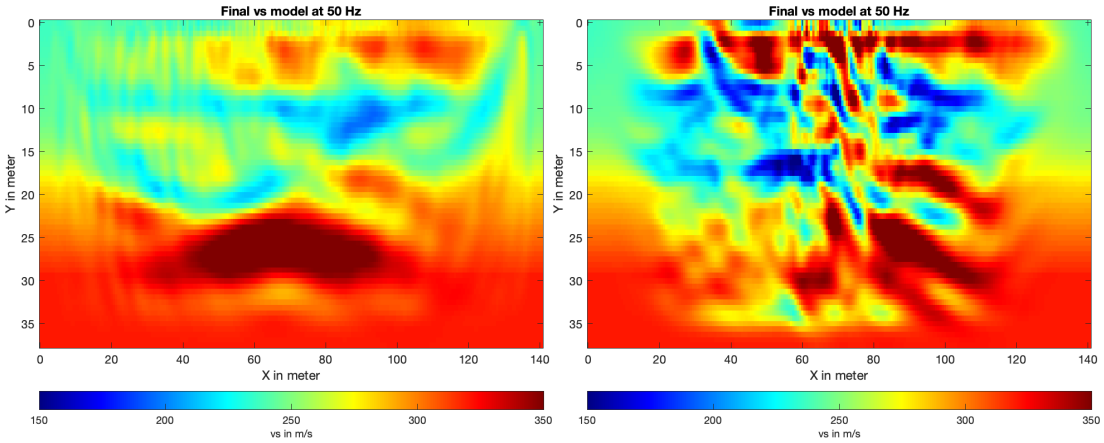
FIGURE 6.2: Initial subset v_s model that is used for the comparison of FSW and RSW.

by a cosine taper from 0 s to 0.02 s and again from 0.24 s to 0.48 s. Furthermore, to minimise strong artefacts near the source, a circular source taper is used with a radius of 40 grid points. The configuration of the multi-stage approach is also changed compared to the synthetic benchmark. Each stage increases the frequency only by 5 Hz instead of 10 Hz, thus a total of 9 stages from 10 Hz to 50 Hz are used. This is done to ensure a less aggressive increase of the misfit at the beginning of the next stage and an easier convergence, albeit the computing time is increased by this choice. The stop criterion for FSW is set to a relative misfit change of less than 0.01 which means that the inversion stops and progresses to the next stage if the misfit changes by less than 0.01 compared to the second last iteration. In case of RSW, I choose to iterate every single shot exactly once before progressing to the next stage because more iterations result degrade the model as I will explain in the following sections.

6.2.2 First results

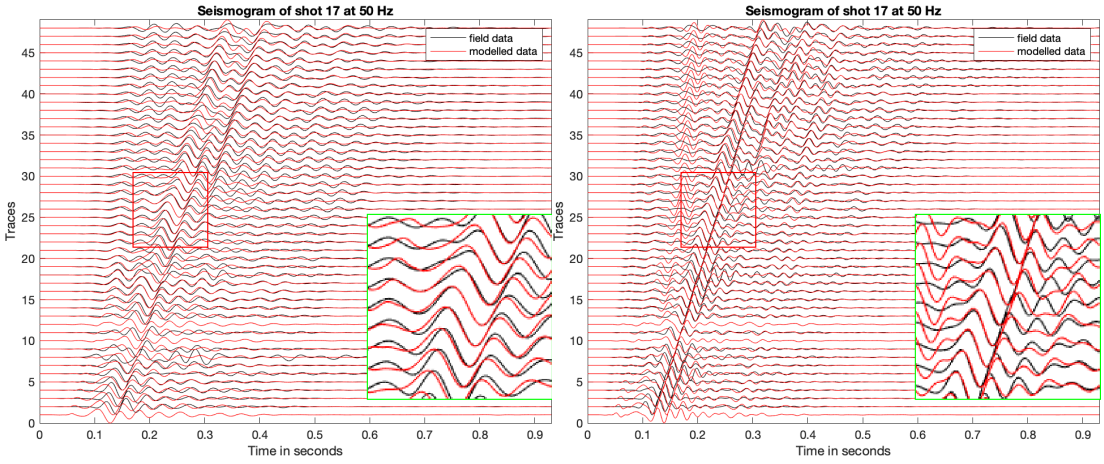
The final v_s subset models of FSW and RSW are shown in figure 6.3 and corresponding sample seismograms of shot 17 are displayed in figure 6.4. Both models show a high-velocity layer at shallow depths up to 8 m. RSW indicates a much higher velocity up to 350 m s^{-1} for this layer while FSW only reaches up to around 310 m s^{-1}. Under it, both workflows show a low-velocity area between 8 m and 15 m depth. FSW shows this as a homogeneous area while RSW shows strong heterogeneity throughout the model. In addition, the velocities in RSW's model reach lower values down to 150 m s^{-1}. A strong velocity contrast with very high and low velocities can be explained by over-iterating the model, however, RSW only iterates every shot once per stage, thus this explanation seems unlikely. I investigate the influence of more iterations in chapther 6.2.4. Nevertheless, there must be other reasons for these extreme values that I investigate in the following sections. Looking at the seismograms of shot 21 in figure 6.4 shows some misfit between the field data in black and the modelled data in red that are larger at some traces such as trace 31 and smaller at others such as trace 42 to 48.

Comparing the seismograms of FSW to RSW shows that FSW constructs the first arrivals at far offset better. RSW models waveforms in the traces 24 to 30 and 35 to 48 that are not apparent in the field data. However, RSW constructs the waveform after the surface wave arrivals with a better fit which is especially noticeable in traces 35 to 38. Thus, based on the seismogram's misfit, neither FSW nor RSW are better, but the model results show that RSW seems to invert a much less coherent model than FSW.



(A) FSW v_S model at 50 Hz (B) RSW v_S model at 50 Hz

FIGURE 6.3: Final v_S subset models of FSW (A) and RSW (B) at the end of stage 9 at 50 Hz.



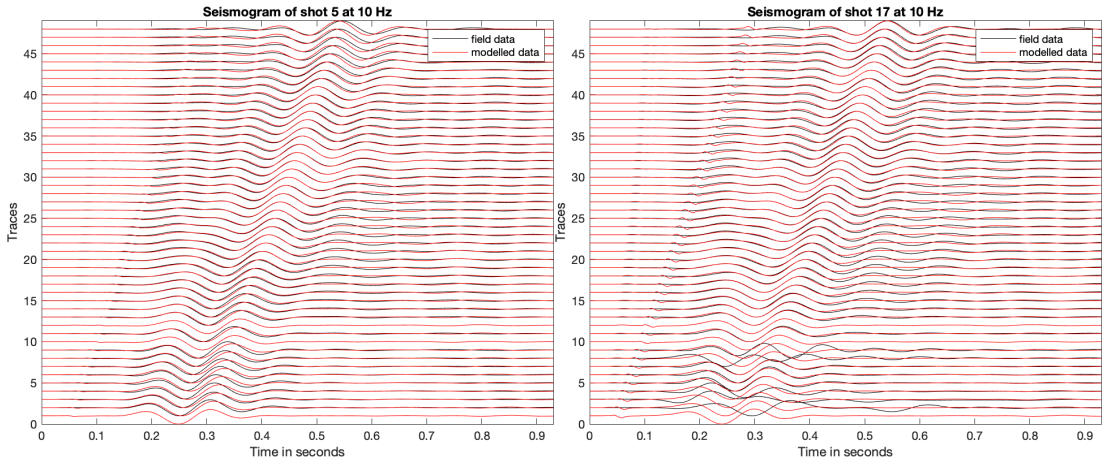
(A) FSW seismograms at 50 Hz (B) RSW seismograms at 50 Hz

FIGURE 6.4: FSW (A) and RSW (B) seismograms of shot 17 from the subset model at the end of stage 9 at 50 Hz.

6.2.3 Influence of trace killing

To investigate the incoherent final model of RSW, I start to look at different seismograms after the first iteration of the first stage at 10 Hz. Two of them, shot 5 and shot 17, are included in figure 6.5. They represent the examples of a good fit as in shot 5 and a bad fit as in shot 17. Shot 17 in figure 6.5b shows traces near the source which do not match with the adjacent traces and are out of phase in comparison. This observation can be seen on many shots, where some of the first 10 traces near the source show incoherent behaviours in comparison to the adjacent traces. It is unlikely that this phenomenon is explained by the underground structure because it appears randomly

throughout some shots within the first 10 traces without a clear pattern, both in FSW and RSW. Thus, the cause might be related to the acquisition such as bad coupling of the geophones or other causes that are related to a near offset.
Therefore, I choose to consistently kill the first ten nearest offset traces in addition to the traces that I killed during the preprocessing steps, albeit not all shots have bad traces near the source such as shot 5 in figure 6.5a shows. Since I remove not just bad but also good data by the trace killing, I inspect the results again of both FSW and RSW where I compare the models and seismograms with and without trace killing after the first stage at 10 Hz, the 5th stage at 30 Hz and the last stage at 50 Hz. The results after the 5th stage are shown in figure 6.6 and 6.7 for FSW and 6.9 and 6.10 for RSW. The results of the first and the last stage are in appendix A in figure A.1 to A.8.



(A) RSW seismograms of shot 5 at 10 Hz (B) RSW seismograms of shot 17 at 10 Hz

FIGURE 6.5: RSW seismograms of shot 5 (A) and 17 (B) of the subset models after the first iteration at 10 Hz.

Full-shot workflow results and discussion

In figure 6.6 the FSW v_s model without trace killing shows a high-velocity area near the sources between 30 m and 60 m at around 2 m to 6 m depth. This is due to the contribution of the near-offset traces that show a significant misfit in figure 6.7a. If they are removed as 6.7b shows, the high-velocity area is also gone and smaller low-velocity areas are inverted in the model in figure 6.6b. The structure in greater depths changes as well, resulting in a more heterogeneous distribution of high- and low-velocity areas. Each high-velocity area also shows a higher value than the corresponding area in the model when trace killing is not utilised, while the low-velocity areas show a lower value, resulting in a more extreme contrast. This happens because the inversion has to update fewer traces when trace killing is used, therefore each trace are more heavily influenced by the model update per iteration. This can be observed well in figure 6.7. The traces 27 to 33 in figure 6.7b show a better fit between field and modelled data. Furthermore, the first arrivals are consistently better fitted. The better fit is also quantitatively illustrated in the misfit evolutions in figure 6.8. There, the misfit values are the result of a summed misfit of all shots divided by the total number of traces. This allows for a quantitative comparison. Without trace killing, the inversion struggles to find a better model in the early stages, thus only a few iterations are performed. With trace killing, the misfit at the first iteration is already more than 10% lower and further

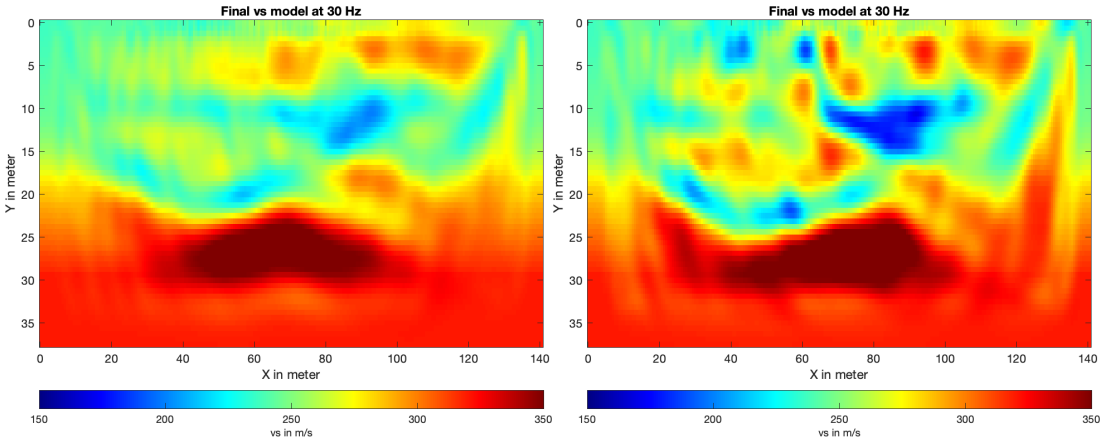
decreases in the following iterations in the same stage. The increase of the misfit at the beginning of the next stage is a common behaviour that I discussed in chapter 4.1.2.

Random-shot workflow results and discussion

The results of RSW behave similarly to trace killing as the FSW results. Both v_s models after stage 5 at 30 Hz in figure 6.9 show a strong velocity contrast with very high and very low velocity values. When trace killing is utilised, the high-velocity anomaly near the sources between 30 m and 60 m at around 2 m to 6 m depth mostly disappears as well. However, the velocities still reach unlikely high values up to 350 m s^{-1} and low values of 150 m s^{-1}. My explanation for this phenomenon is the general sensitivity of RSW towards bad traces and shots. Because each shot is updated individually, their gradients contribute more significantly to the model update than in FSW where the gradients are summed first before the model is updated. As I showed in the preprocessing in chapter 5.3 and also in the sample seismogram of shot 17 in figure 6.5b, the Salt Lake City field data exhibits a rather low quality with strong interference of noise due to the acquisition circumstances. Therefore, those bad data disturb the model and leave strong anomalies after a model update, no matter if near-offset traces are killed or not. The seismograms of shot 17 in figure 6.10 show how trace killing partially improves the fit between field data and modelled data, such as in figure 6.10b at the far-offset traces 40 to 48 where after 0.4 s the waveforms are slightly better fitted. Also, the first arrivals of all traces are better aligned as well. However, in other areas, the fit is better without trace killing, such as between 0.2 s and 0.3 s in trace 35 to 48 in figure 6.10a.

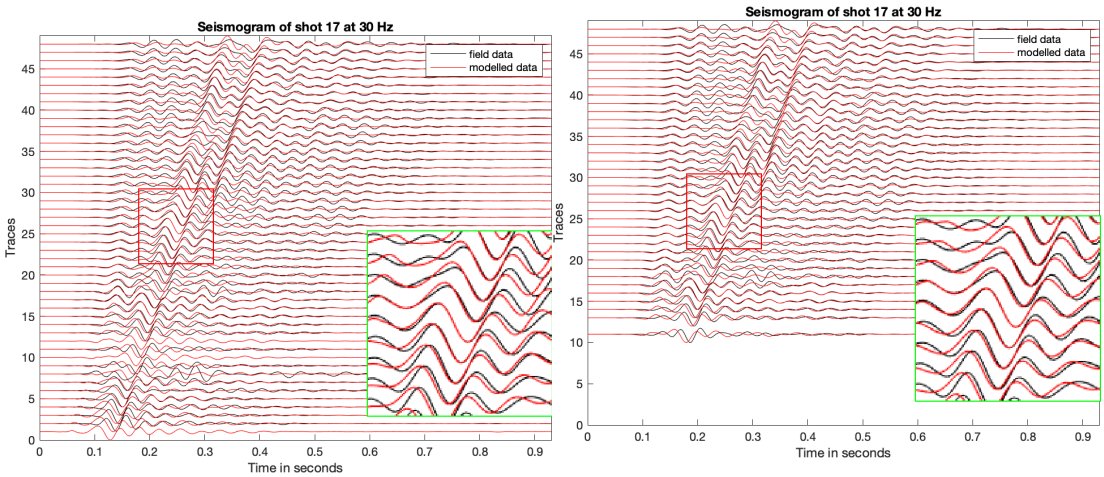
FSW vs RSW

Comparing the seismograms with the seismograms from FSW in figure 6.7, a slightly better fit of RSW's seismograms can be observed. This is surprising because the RSW's model in figure 6.9 shows very strong anomalies. Thus, it is hard to judge whether these anomalies are purely artificial or partially geological. Another way to compare the results is to look at the source time function inversion. Because every shot uses the same source, each source signature in one stage should ideally look the same. Figure 6.11 shows all inverted STFs at 10 Hz, 30 Hz and 50 Hz for both FSW and RSW. The first stage at 10 Hz shows consistent STFs for both FSW and RSW. In stage 5 at 30 Hz, the differences between both workflows become more apparent. In figure 6.11c, the amplitude's global maximum of FSW is higher than the amplitude's global maximum of RSW in figure 6.11d. Moreover, the waveform changes after the global minimum. After it, RSW's waveform raises much higher than FSW's waveform, at both 30 Hz and 50 Hz. Hence, there are apparent differences between FSW and RSW in the STF inversion. However, it is hard to say which workflow produces more constant STFs. Ideally, the source signature of an accelerated weight drop source should be minimum or mixed phased. This attribute is shared among both FSW's and RSW's STFs, albeit in FSW's case more clearly. Switching to the STF inversion results with the use of trace killing in figure 6.12, the differences between FSW and RSW are less apparent, albeit the waveforms at 30 Hz and 50 Hz after their global minimum are still different. All STFs become more and more diverse the higher the frequency becomes which is expected due to finer differences with a higher frequency content that need to be fitted with the STF inversion.



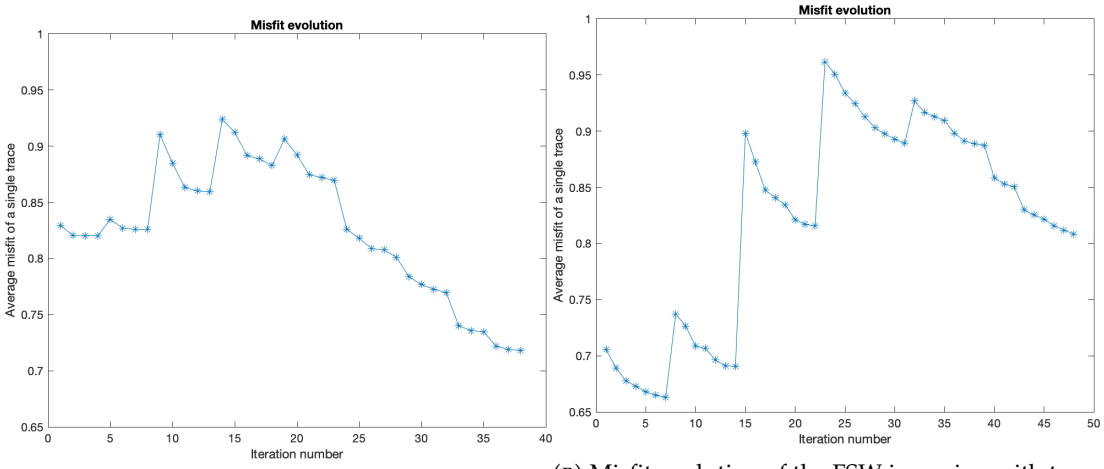
(A) FSW v_s model at 30 Hz (B) FSW v_s model at 30 Hz with trace killing

FIGURE 6.6: FSW v_s subset models after stage 5 at 30 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) FSW seismograms of shot 17 at 30 Hz (B) FSW seismograms of shot 17 at 30 Hz with trace killing

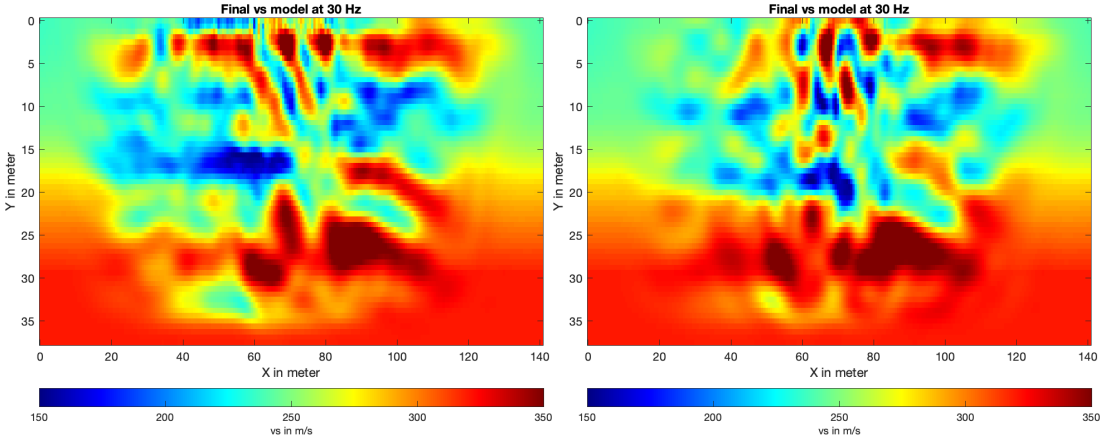
FIGURE 6.7: FSW seismograms of shot 17 of the subset model after stage 5 at 30 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) Misfit evolution of the FSW inversion

(B) Misfit evolution of the FSW inversion with trace killing

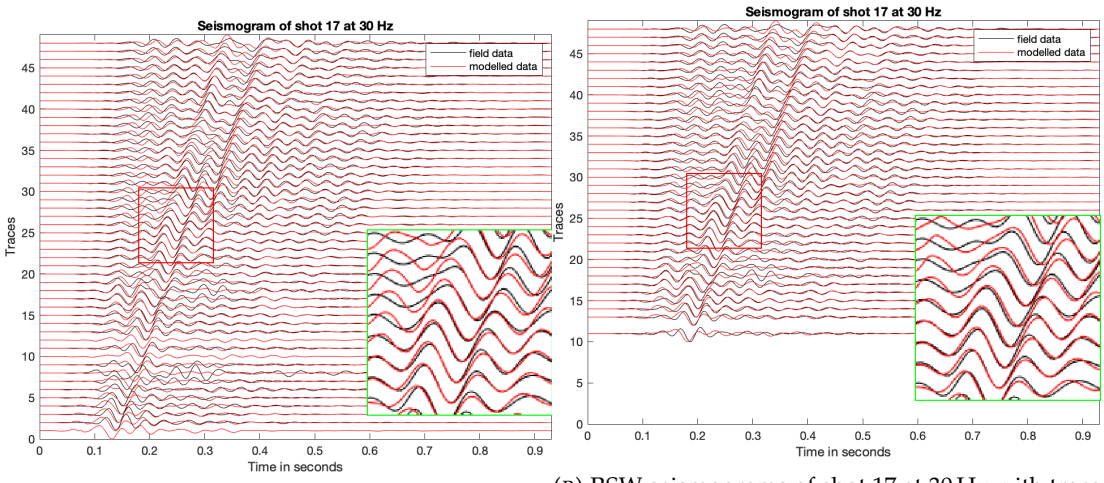
FIGURE 6.8: Misfit evolution of the FSW inversion runs of the subset model without (A) and with trace killing (B). Each misfit value per iteration represents the summed value of all shots divided by the number of all traces to enable a quantitative comparison of their values.



(A) RSW v_S model at 30 Hz

(B) RSW v_S model at 30 Hz with trace killing

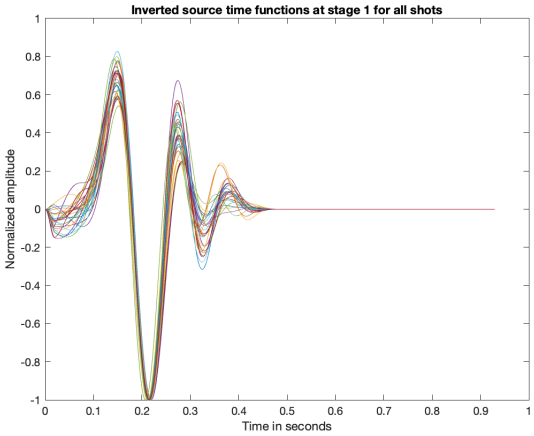
FIGURE 6.9: RSW v_S subset models after stage 5 at 30 Hz with all near-offset traces (A) and with killed near-offset traces (B).



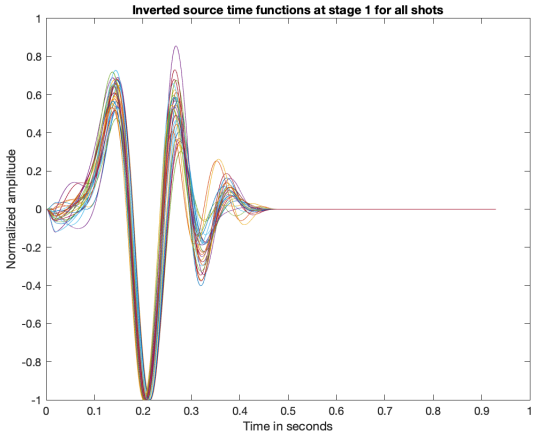
(A) RSW seismograms of shot 17 at 30 Hz

(B) RSW seismograms of shot 17 at 30 Hz with trace killing

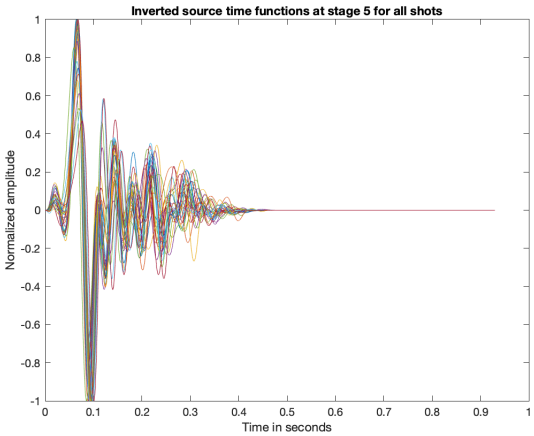
FIGURE 6.10: RSW seismograms of shot 17 of the subset model after stage 5 at 30 Hz with all near-offset traces (A) and with killed near-offset traces (B).



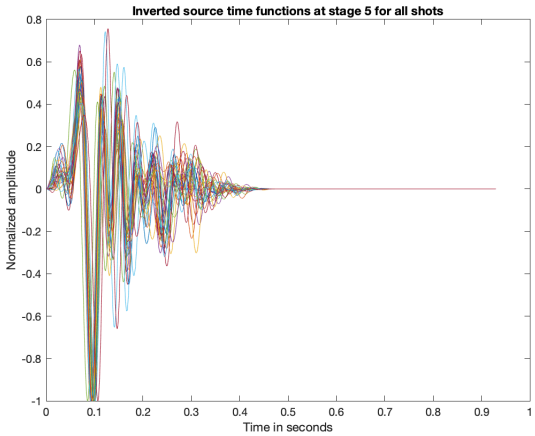
(A) FSW source time function of all shots at 10 Hz



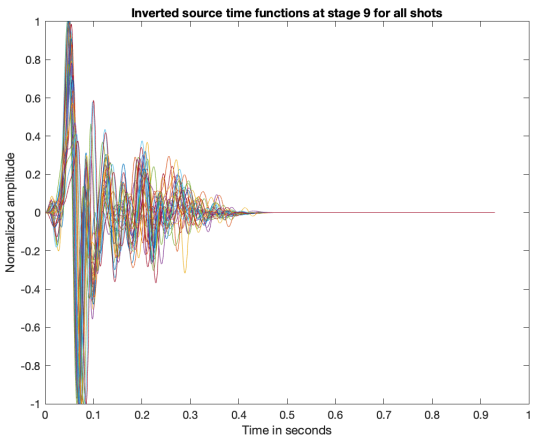
(B) RSW source time function of all shots at 10 Hz



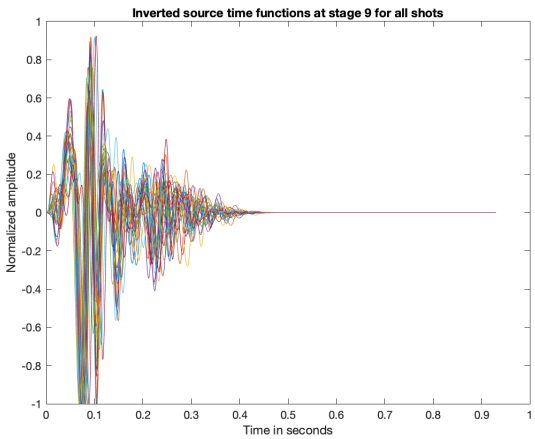
(C) FSW source time function of all shots at 30 Hz



(D) RSW source time function of all shots at 30 Hz

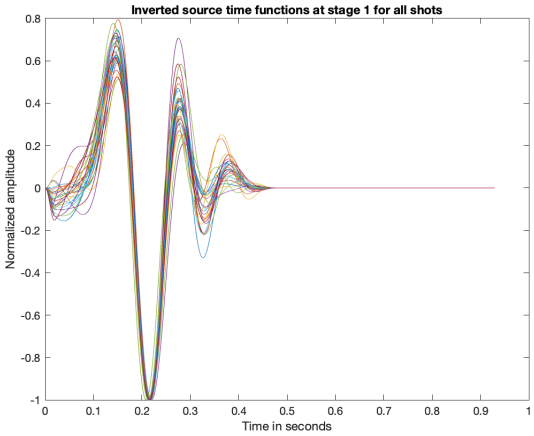


(E) FSW source time function of all shots at 50 Hz

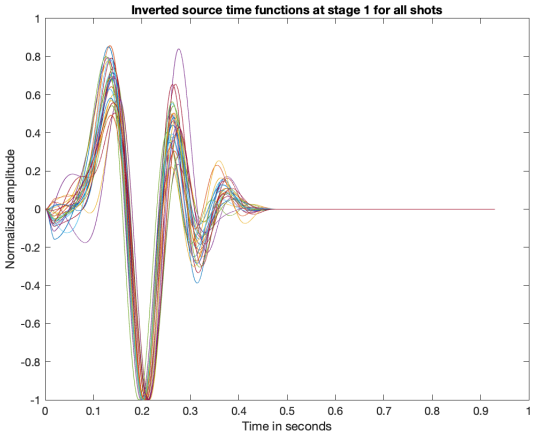


(F) RSW source time function of all shots at 50 Hz

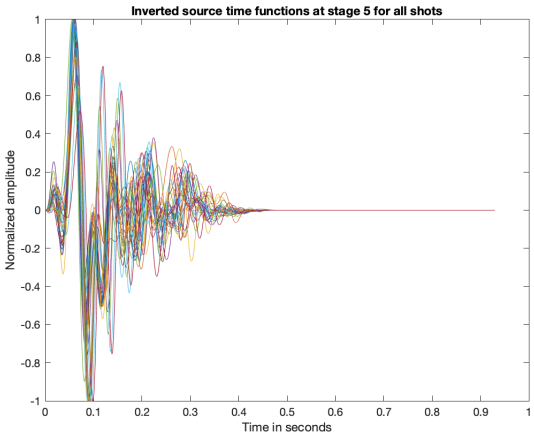
FIGURE 6.11: Source time functions of all shots of the subset model at different stages for both FSW (left side) and RSW (right side).



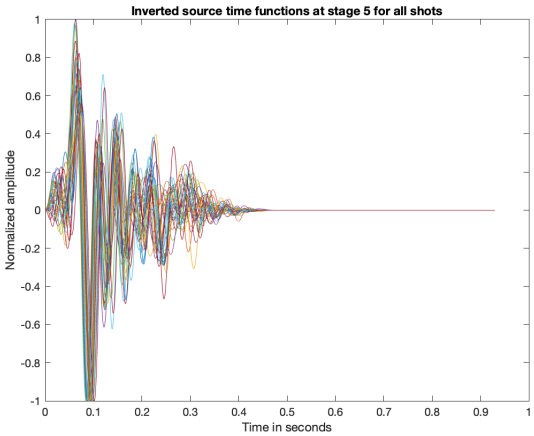
(A) FSW source time function of all shots at 10 Hz



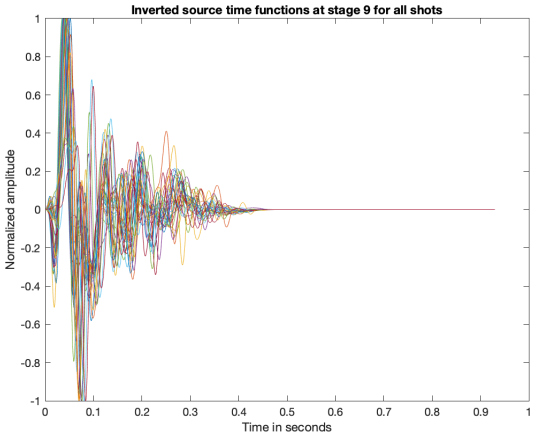
(B) RSW source time function of all shots at 10 Hz



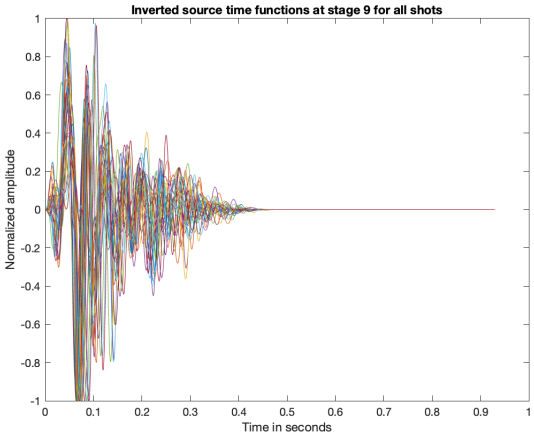
(C) FSW source time function of all shots at 30 Hz



(D) RSW source time function of all shots at 30 Hz



(E) FSW source time function of all shots at 50 Hz



(F) RSW source time function of all shots at 50 Hz

FIGURE 6.12: Source time functions of all shots of the subset model at different stages for both FSW (left side) and RSW (right side) with trace killing.

6.2.4 Influence of higher iteration numbers

For the analysis of the trace killing influences, each shot is iterated once in RSW and on average four to five times in FSW. The misfit evolutions in figure 6.8 also show how later stages are iterated less on average. Thus, I would like to investigate whether higher iteration numbers can improve the model and whether or not the abort criterion is reached too early. The reason to optimise each workflow is to compare the FSW and RSW at their optimal configurations. If one workflow does not utilise its potential because it is more sensitive to certain parameters than the other, such as the number of iterations it must perform, the comparison becomes flawed. Therefore, I increase the iterations per stage from one to three in RSW and set the abort criterion in FSW to a relative misfit change of less than 0.002 instead of the previous less than 0.01. Furthermore, I perform the new inversion runs with and without trace killing to cover all possibilities.

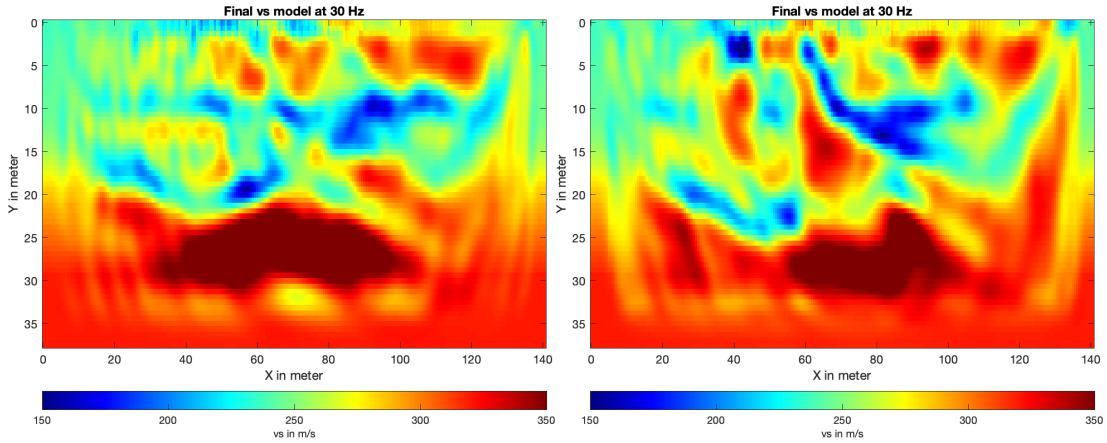
Full-shot workflow results and discussion

The v_s model results from FSW with more iterations are shown in figure 6.13 with and without trace killing after stage 5 at 30 Hz. The results from stage 1 at 10 Hz and stage 9 at 50 Hz are included in appendix B in figures B.1 to B.4. While the total number of iterations before were 38 and 48 without and with trace killing, respectively, with the new abort criterion they now are at 56 and 72 without and with trace killing, respectively. This can be seen in the misfit evolutions shown in figure 6.15. Comparing the model results with the models from lower iteration numbers in figure 6.6, it can be observed that the model has finer details, especially at greater depths between 11 m and 20 m, where finer outlines can be identified. With more iterations, the velocity field exhibits a stronger contrast. This effect is further increased by the usage of trace killing. To keep the comparison consistent, the seismograms of shot 17 are shown in figure 6.14. In comparison with the fit between field and modelled data with less iterations in figure 6.7, it barely changes with more iterations. This is supported by the misfit evolution in figure 6.15. Comparing the misfit values with the values in figure 6.8 where less iterations are done, almost no improvements can be seen. At some stages, the misfit is lower when more iterations are performed, at other stages the misfit is actually higher. Thus, I conclude that higher iteration numbers do not benefit the FSW and no better models yielded. This is also visible in the results in appendix B in figures B.1 and B.2. Although higher frequencies are utilised in later stages, the models at 50 Hz in figure B.4a and B.4b do not significantly differ from the models at 30 Hz in figure 6.13. Therefore, it is possibly better to save computing time and only run the inversion up to 30 Hz for the Salt Lake City field data.

Random-shot workflow results and discussion

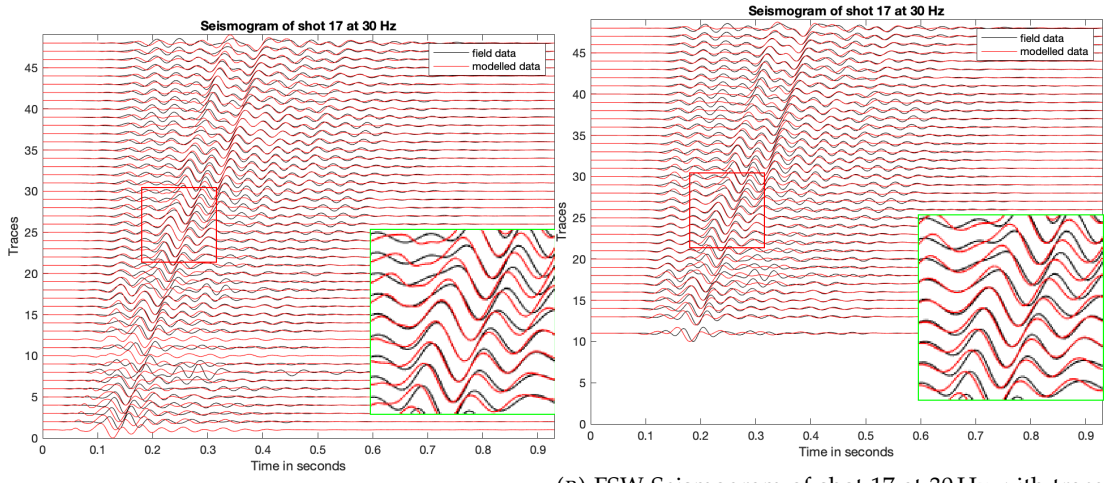
In chapter 6.2.2 I already mentioned how the velocities in RSW's v_s model exhibit a strong velocity constrast compared to FSW's results and also the results from Liberty, St Clair, and Gribler (2018) in figure 5.4. Hence, more iterations unlikely solve this issue. This assumption is confirmed in figure 6.16 where the v_s models of RSW with more iterations are shown at the end of different stage at 10 Hz, 30 Hz and 50 Hz for both with and without trace killing. At 10 Hz, the model without trace killing shows a large high-velocity anomaly near the sources that dominates the rest of the model which was also observed in chapter 6.2.3, see figures A.5 and A.6. Because each shot is iterated three times now instead of just once per stage, the effect of bad near-offset traces are amplified. Looking at the model where those near-offset traces are killed,

the anomaly disappears. When the inversion progresses to later stages with higher frequencies, the model results become very heterogeneous and the velocities further reach very high values up to 350 m s^{-1} and low values down to 150 m s^{-1}. Essentially, the observations that are described in chapter 6.2.2 are amplified due to more iterations. Therefore, the RSW yields worse models if more iterations are performed which supports the conclusion in chapter 6.2.3 where it turns out how sensitive RSW is to bad traces and shots. Because more iterations abuse this sensitivity, the resulting models are further disrupted by strong gradients that are caused by bad quality data. Another explanation for the degraded models is the case of an underdetermined problem. The seismograms the appendix B in figure B.5 show that the misfits are not much different from the misfits in the seismograms of FSW. Thus, many models might explain the data and the final models in figure 6.16 are one of them which do not look geologically reasonable. However, it is still unclear to my knowledge, why this phenomenon only happens with RSW in this subset.



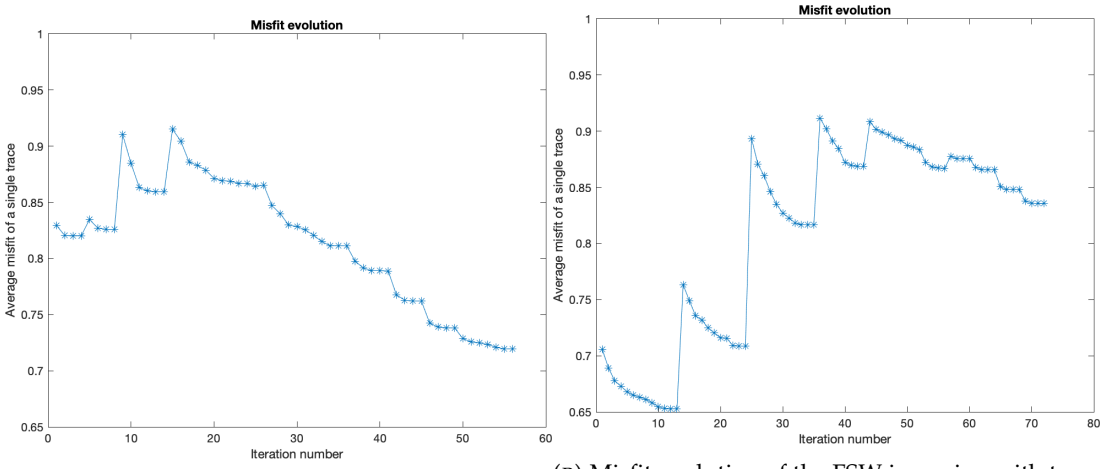
(A) FSW v_S model at 30 Hz (B) FSW v_S model at 30 Hz with trace killing

FIGURE 6.13: FSW v_S subset models after more iterations at stage 5 at 30 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) FSW seismogram of shot 17 at 30 Hz (B) FSW Seismogram of shot 17 at 30 Hz with trace killing

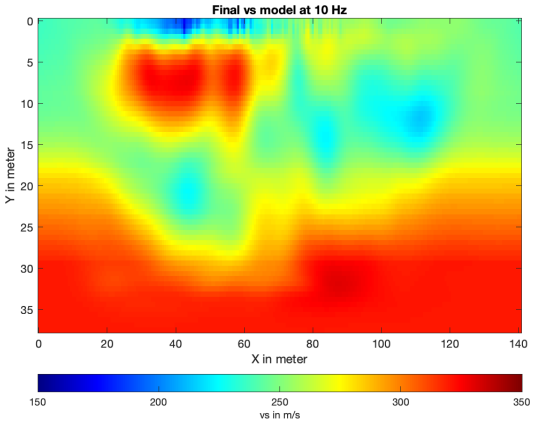
FIGURE 6.14: FSW seismograms of shot 17 of the subset model after more iterations at stage 5 at 30 Hz with all near-offset traces (A) and with killed near-offset traces (B).



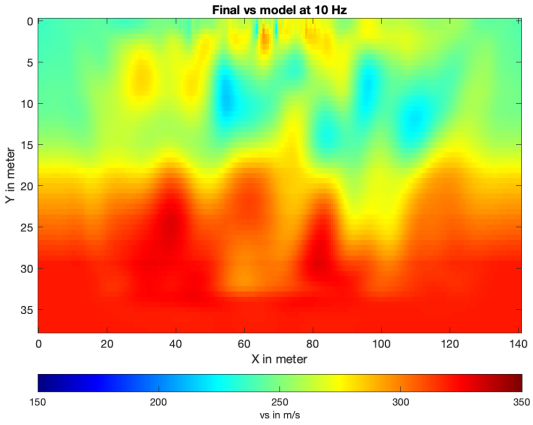
(A) Misfit evolution of the FSW inversion

(B) Misfit evolution of the FSW inversion with trace killing

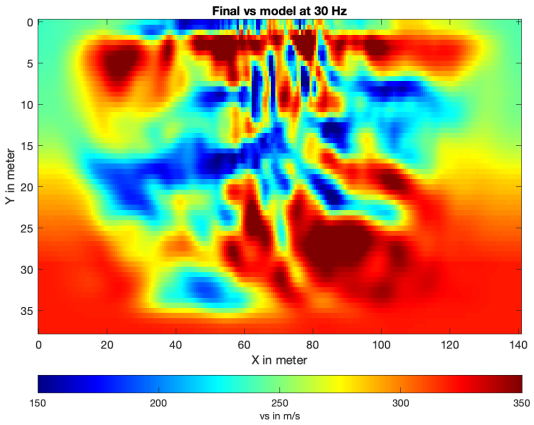
FIGURE 6.15: Misfit evolution of the FSW inversion runs of the subset model without (A) and with trace killing (B) with a more strict abort criterion, thus higher iteration number. Each misfit value per iteration represents the summed value of all shots divided by the number of all traces to enable a quantitative comparison of their values.



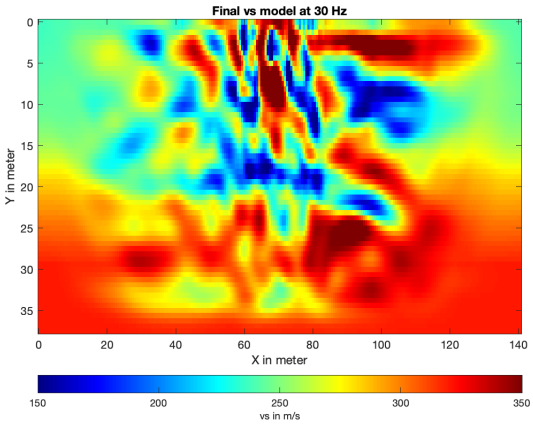
(A) RSW v_s model at 10 Hz



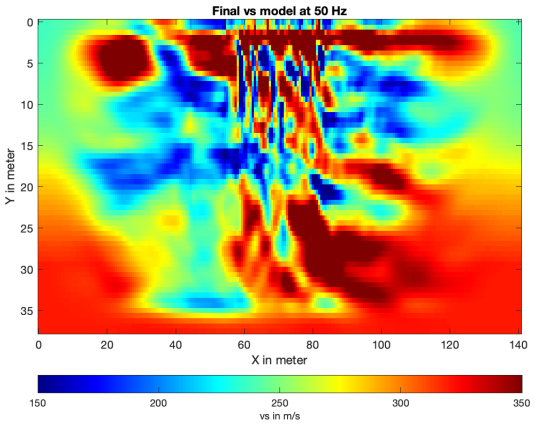
(B) RSW v_s model at 10 Hz with trace killing



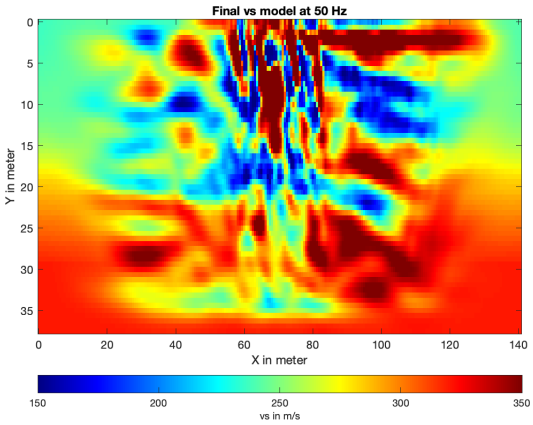
(C) RSW v_s model at 30 Hz



(D) RSW v_s model at 30 Hz with trace killing



(E) RSW v_s model at 50 Hz

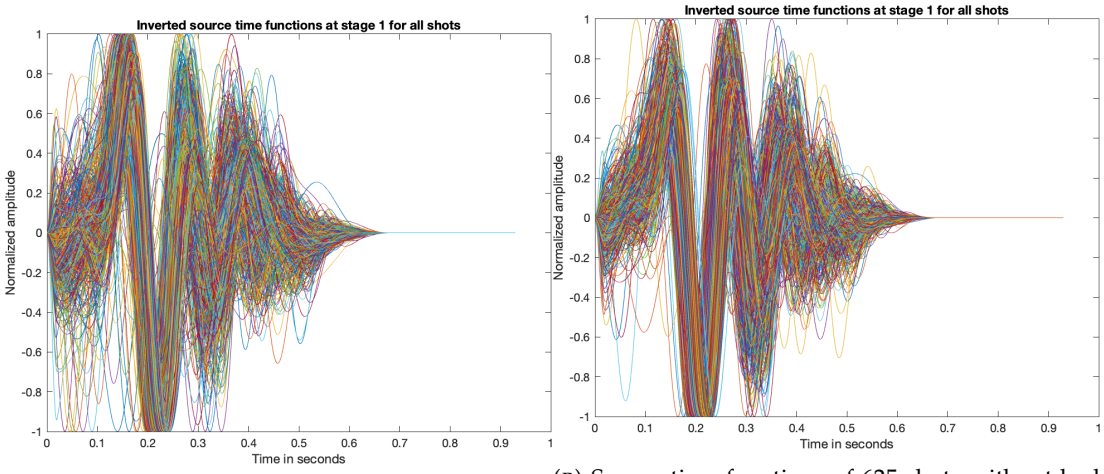


(F) RSW v_s model at 50 Hz with trace killing

FIGURE 6.16: Final v_s subset models of RSW after more iterations without trace killing on the left and with trace killing on the right. Each shot is iterated three times per stage.

6.2.5 Inversion of the complete data set

Until now, the comparison between FSW and RSW is made only with a subset of the Salt Lake City data. The initial goal of this work, however, is to find a method to apply full-waveform inversion on a large data set and not a small one. Therefore, both FSW and RSW are used to invert the complete data set for the final comparison of their model results. Before the inversion, I readjust some parameters based on the conclusion from the previous comparisons in chapter 6.2.3 and 6.2.4. Because the model does not change much beyond 30 Hz, I set the multi-stage configuration to a total of 6 stages from 10 Hz to 35 Hz with a 5 Hz increases after each stage to save computing time. The FSW's stop criterion is adjusted as well and set to a relative misfit change of <0.002. A higher value results in almost no changes of the starting model and only two to three iterations per stage, hence the choice of a stricter abort criterion. RSW proceeds to the next stage after all shots are iterated two times. During the preprocessing, I also identified 30 shots with data that are completely unusable due to their low signal-to-noise ratio. Thus, I remove them for the inversion of the complete data which reduces the total number of shots to 625. Furthermore, I set the cosine taper that I apply to the STFs from 0.34 s to 0.68 s. An illustration of the difference of the inversion of the STFs with and without the bad shots is shown in figure 6.17 where the STFs from all shots are plotted at 10 Hz. With the bad shots, some STFs vastly differ from the others, especially before 0.2 s, whereas without the bad shots, most outliers are removed, albeit not all of them.



(A) Source time functions of all 655 shots

(B) Source time functions of 625 shots without bad shots

FIGURE 6.17: A comparison of the inverted source time functions with all shots (A) and without bad shots (B) at stage 1 at 10 Hz of the Salt Lake City field data.

Full-shot workflow results and discussion

The v_s model at the final iteration of FSW is shown in figure 6.18. The shallow high-velocity anomalies at around $x = 300\,\text{m}$ and $x = 900\,\text{m}$ from the initial model and v_s tomogram from Liberty, St Clair, and Gribler (2018) in figure 5.4 is well reconstructed with seemingly more resolution and detail of their shape, especially between 10 and 20 m depth. Beneath them between around 14 m and 20 m, the velocity decreases again which can also be seen in the v_s tomogram between $x = 700\,\text{m}$ and $x = 750\,\text{m}$ and between $x = 800\,\text{m}$ and $x = 900\,\text{m}$, not at the first one however at around $x = 250\,\text{m}$ where the anomaly is not decoupled from the background velocity

as in the FSW model. Further details can be seen from $x = 1000\,\text{m}$ onwards where a small high-velocity layer can be identified at around 10 m to 14 m depth. This is also sporadically visible in the v_s tomogram, although, less consistent along the profile and not as large. The inverted v_s values also roughly correspond to the velocities in the v_s tomogram. There, they are classified in the NEHRP class D1 and D2 between $180\,\text{m}\,\text{s}^{-1}$ and $300\,\text{m}\,\text{s}^{-1}$ for depths up to 20 m, while the v_s model from the FSW inversion indicates a velocities between approximately $170\,\text{m}\,\text{s}^{-1}$ and $280\,\text{m}\,\text{s}^{-1}$. Below 20 m depth, the velocity abruptly increases to $300\,\text{m}\,\text{s}^{-1}$ and higher which is also seen in the tomogram where v_s reaches up to $360\,\text{m}\,\text{s}^{-1}$ at depths below 25 m.
Based on this comparison with the v_s tomogram, the inverted model from FSW seems to be realistic with maybe greater resolution at shallow depths above 20 m. However, beneath 20 m, the model shows very curvy contours that seem to be shot dependent, especially between $x = 200\,\text{m}$ and $x = 800\,\text{m}$. Therefore, the FSW inversion appears to be unstable at these depths. Moreover, the model seems to contain some sort of numerical noise, visible as fine vertical lines. Although the inversion is capped at 35 Hz with the use of the multi-stage approach and based on equation 2.26 and 2.27 Δh and Δt are chosen sufficiently fine for a 6th order FD scheme, the parameters might have been set too closely to the threshold where numerical noise is induced.
The misfit evolution is shown in figure 6.19. It can be seen that the inversion struggles to decrease the misfit by a more significant amount per iteration, even in early stages below 15 Hz. Despite the little improvements of the misfit, the resulting model is comparable to the v_s tomogram from Liberty, St Clair, and Gribler (2018). The difficulties to minimise the misfit could come from the two dimensional nature of the 2D adjoint FWI. To investigate this possibility, I show the seismograms of shot 121 and 501 in figure 6.20. The first shot is in the first high-velocity area at around $x = 300\,\text{m}$ and the second shot is at around $x = 1050\,\text{m}$ where the CP-9 and CP-10 tip resistance measurements were performed. Shot 501 is also within the lateral spread deposits area that is marked in figure 5.3. The chances of 3D effects are therefore higher in this area. Shot 121 in figure 6.20a shows a satisfactory fit between field and modelled data between trace 22 and 37 with a very good fit between trace 29 and 36. However, starting from trace 38, the surface waves' arrival time diverge more and more with increasing offset. Hence, the velocity model is probably not correct at greater depths because the wavepaths for far-offset receivers can dive deeper into the subsurface. This correlates with the strongly deformed boundary lines in the v_s model at depths between 20 m and 30 m in figure 6.18. Shot 501 in figure 6.20b shows a generally good fit with some differences between trace 10 and 20. However, starting from trace 37, the waveforms after 0.45 s start to differ more significantly. From trace 42 to 48 between 0.6 s and 0.7 s there seems to be a possible reflection of the surface wave in the field data which could come from the third dimension lateral to the acquisition profile. Because the inversion only considers the subsurface in the 2D model, these waveforms in the field data cannot be modelled, thus the modelled waveforms cannot be corrected, which leads to the increased misfit for this particular shot. Because this phenomenon does not only happen once but for many shots, the difficulties to decrease the misfit as seen in its evolution in figure 6.19 could be partially explained by 3D effects.

Random-shot workflow results and discussion

Due to the random choice of subsets and thus a random order of their inversion, different RSW inversion runs will yield different final models even if the parameters are identical. Using several applications of the RSW inversion, model uncertainties can be identified. If a certain area in the model is consistently different among all produced

results, this area could be categorised as a model uncertainty, whether it is because of the method or the data quality. Therefore, two v_s model results from the RSW inversion are shown in figure 6.21. The two high-velocity anomalies at around $x = 300\,\text{m}$ and $x = 900\,\text{m}$ can be identified again as in the FSW results, albeit the first anomaly is significantly smaller in size compared to the FSW results in figure 6.18. Another similarity between all three models are the small high-velocity layer from $x = 1000\,\text{m}$ onwards at around 10 m to 14 m depth. RSW also manages to construct this area as a more consistent layer compared to the v_s tomogram in figure 5.4 from Liberty, St Clair, and Gribler (2018). Thus, the reason for the more consistent image probably comes from the application of full-waveform inversion and is independent from the choice of FSW or RSW. Moreover, RSW seems to construct a slightly more homogeneous layer at depths greater than 20 m, at least in the first inversion run in figure 6.21a. There are still uneven contours with quickly changing directions, but less so than FSW exhibits in its final model. The model in 6.21b has a stronger heterogeneity at depths greater than 20 m.

Furthermore, RSW inverts a model with higher velocity contrasts. This correlates with the conclusions from the investigation of the influence of trace killing and higher iteration numbers in chapter 6.2.3 and 6.2.4. The velocities reach down to $150\,\text{m}\,\text{s}^{-1}$ and up to $350\,\text{m}\,\text{s}^{-1}$. The high-velocity anomaly at $x = 900\,\text{m}$ also reaches a higher velocity than both FSW and the v_s tomogram indicate which are both approximately $80\,\text{m}\,\text{s}^{-1}$ lower. Besides, on the one hand, the RSW v_s model results look generally more heterogeneous than FSW results, even though the synthetic benchmark shows a smoother RSW model than FSW for both the anomaly and layer model. Thus, I assume that the heterogeneity must come from the use of the field data and not the random-shot workflow. On the other hand, however, the heterogeneity can be due to a potentially higher resolution of RSW. This conclusion is rather unlikely though, else it would have been observed during the synthetic benchmark already.

Based on the conclusions from the random-shot workflow results in 6.2.3 and 6.2.4, that RSW exhibits strong sensitivity to bad traces and shots, it is likely that some anomalies do not come from a geological object but from this sensitivity. This is especially noticeable in figure 6.21a. The v_s model at around $x = 550\,\text{m}$ and around $x = 1150\,\text{m}$ shows a strong anomaly in the shape of a wavepath from a single shot. Inspecting the seismograms in the two areas shows a very strong misfit between field and modelled data as seen in figure 6.22 where the seismograms of shots 244 and 544 are plotted. Shot 243, 244, 245, 543 and 544 all show such a strong misfit, but the shots adjacent to them are significantly better fitted. Because the shots are only 2 m spaced from each other and the acquisition has an aperture of 60 m, a geological anomaly should be identifiable in more seismograms, for instance as an anomaly outside the 2-D model in the third dimension lateral to it. But because the adjacent ones do not exhibit a strong misfit, I assume this might be correlated specifically to only these shots. The first few traces in shot 244 in figure 6.22a are reasonably well fitted. However, from trace 25 onwards waveforms vastly differ. Shot 544 in figure 6.22b shows a constant misfit that increases with offset. I also show the seismograms of shot 121 and 501 in figure 6.23 to compare their fit between field and modelled data with FSW's fit in figure 6.20. RSW has similar difficulties to fit the data at far offsets and differences between RSW and FSW are negligible. Thus, RSW is capable of fitting data as well as FSW. The major difference in RSW happens only in the model update, where the gradient of one shot has a much more significant impact on the model than the gradient of all shots summed together as in FSW. Therefore, the RSW's sensitivity to bad traces and shots still shows in the inversion of the complete data.

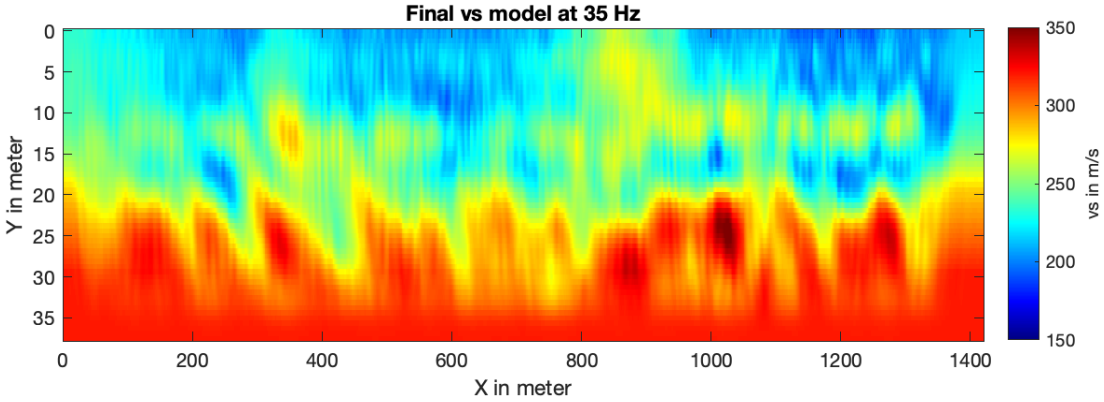


FIGURE 6.18: FSW v_s model of the complete Salt Lake City field data at the final iteration at 35 Hz.

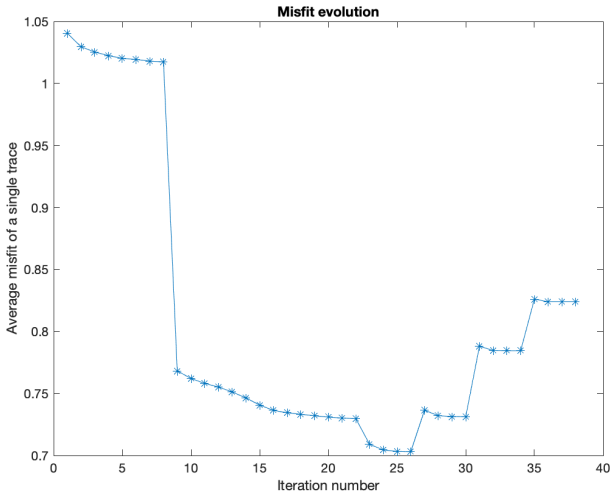
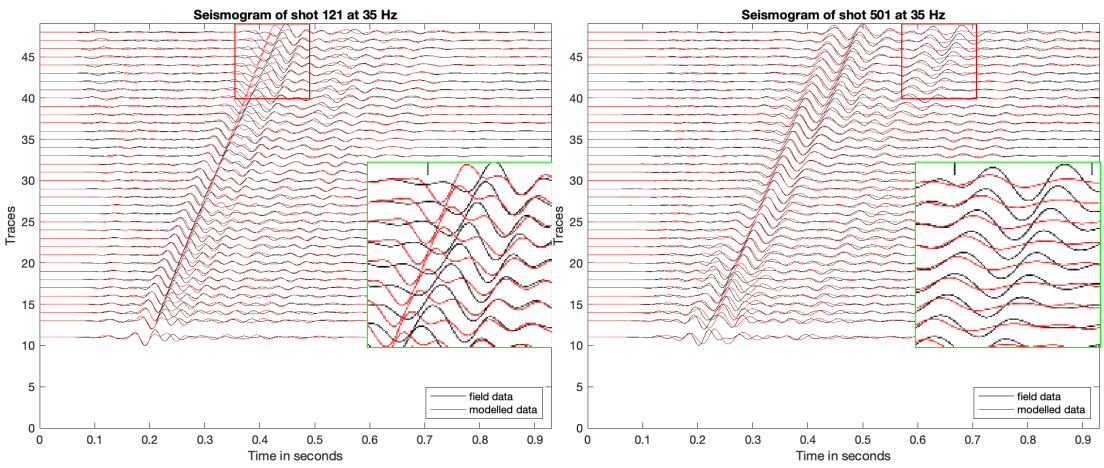
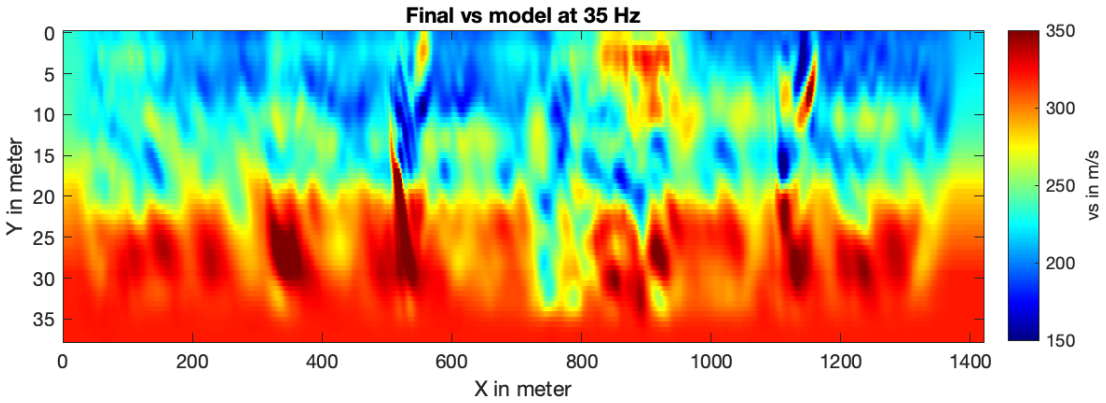


FIGURE 6.19: Misfit evolution of the FSW inversion of the complete Salt Lake City field data.

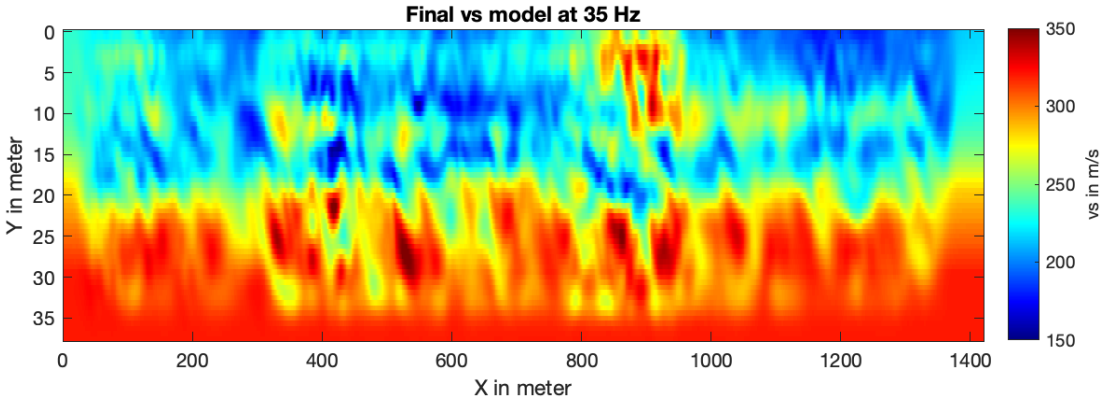


(A) FSW seismogram of shot 121 at 35 Hz (B) FSW seismogram of shot 501 at 35 Hz

FIGURE 6.20: Two sample seismograms of shot 121 and 501 from the FSW results of the complete Salt Lake City field data at 35 Hz. Their source locations are at $x = 254$ m and $x = 1014$ m, respectively.

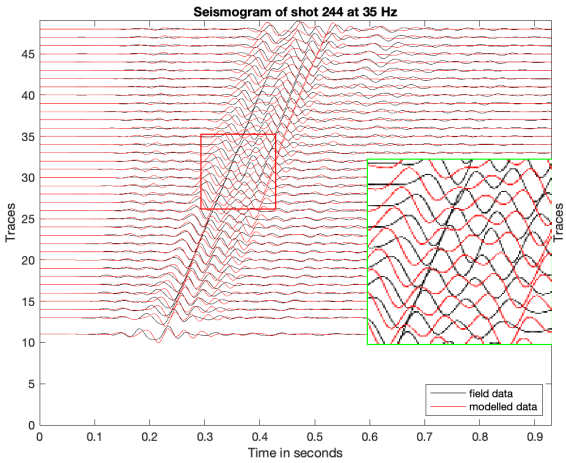


(A) RSW v_S model at the final iteration at 35 Hz from first run

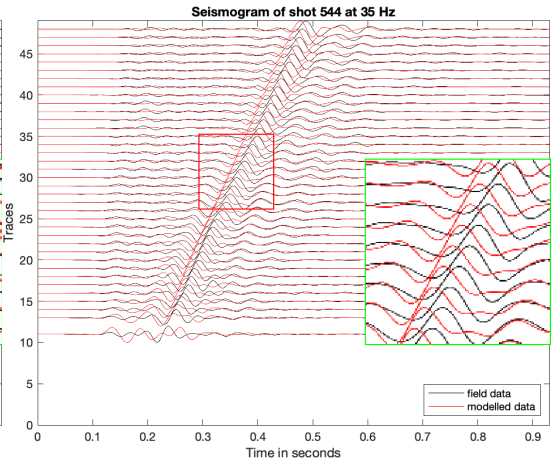


(B) RSW v_S model at the final iteration at 35 Hz from second run

FIGURE 6.21: Two v_S models from two different RSW inversion runs at 35 Hz of the complete Salt Lake City field data.

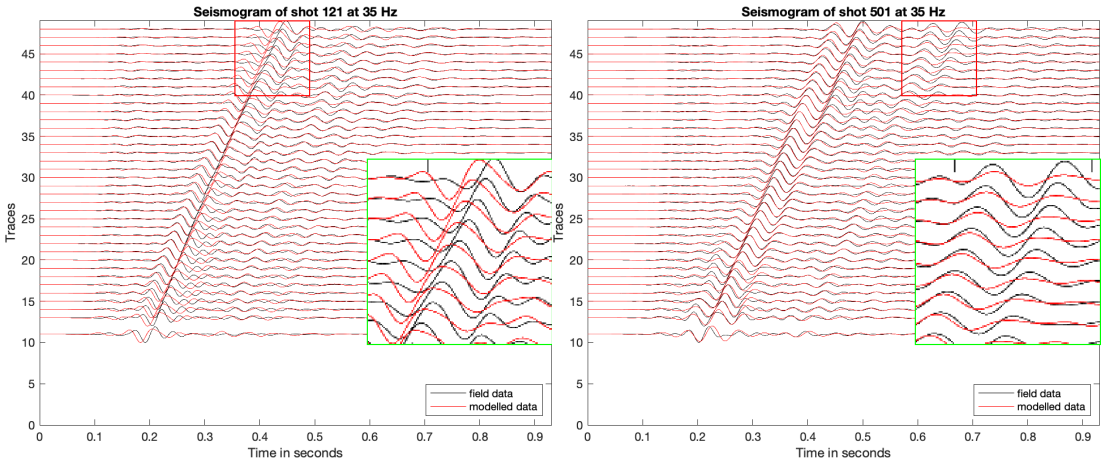


(A) RSW seismogram of shot 121 at 35 Hz



(B) RSW seismogram of shot 501 at 35 Hz

FIGURE 6.22: Two sample seismograms of shot 244 and 544 from the RSW results of the complete Salt Lake City field data at 35 Hz. Their source locations are at $x = 512$ m and $x = 1134$ m, respectively.



(A) RSW seismogram of shot 121 at 35 Hz (B) RSW seismogram of shot 501 at 35 Hz

FIGURE 6.23: Two sample seismograms of shot 121 and 501 from the RSW results of the complete Salt Lake City field data at 35 Hz. Their source locations are at $x = 254$ m and $x = 1014$ m, respectively.

6.2.6 Computational comparison

Part of the goal of this work is to find a method to not only apply full-waveform inversion on a large streamer data set, but also to apply it as efficiently as possible. The initial expectations of RSW are that it can perform faster than FSW with less storage required. This is partially achieved. All inversions were performed on the Geophysical Institute high performance computer (GPIHPC). RSW finishes the inversion of the complete data in around 59 hours which converts to approximately 329 minutes per iteration. FSW finishes the inversion in 110 hours which converts to approximately 203 minutes per iteration. Hence, RSW finishes earlier because it uses less iterations than FSW. Due to the stronger impact of one model update of a single shot in RSW, more iterations do not necessarily yield a better result, as discussed in chapter 6.2.4. Therefore, RSW needs less iterations to yield a comparably resolving image. However, in terms of performance per iteration, it lacks behind FSW. This is due to the use of shot parallelisation in FSW, where several shots are simultaneously forward calculated by different processes. With the use of multi-core processors and HPC clusters that have a large number of processors, shot parallelisation can be optimally utilised. For the FSW inversion on the GPIHPC, I use 25 cores and thus 25 shots can be calculated simultaneously. Currently, shot parallelisation is not implemented in RSW. It would be theoretically possible, though the usage would be restricted. As each inversion is complete within one subset, a simultaneous calculation of subsets requires them to not overlap each other. Depending on the shot spacing and the width of a subset, this restricts the maximum number of shots that can be parallelised in addition to the restriction of the inability to calculate overlapping shots simultaneously. Therefore, RSW only utilises one core and one process. Considering the major computational disadvantage that RSW can only utilise a fraction of the computing power that FSW is able to use, it still manages to beat FSW in terms of total computing time. This is due to the contribution of how efficiently it utilises this one process. Looking at per-CPU computing time per iteration, FSW needs approximately 5000 minutes per iteration while RSW only needs approximately 329 minutes per iteration which is around 15 times faster. Moreover, RSW uses much less storage because the calculations happen

only in a subset and not the complete model. Only 761 Megabytes are used in comparison to 12 Gigabytes that FSW uses per process, resulting in 300 Gigabytes of required disk space during the inversion. Therefore, RSW has the potential to be much faster and more efficient than FSW despite not being able to fully utilise shot parallelisation. However, the use of 25 cores is rather small compared to larger HPC systems such as JUWELS at Jülich Supercomputing Centre where hundreds of cores and terabytes of storage can be utilised. If such a powerful computing facility is available, FSW can make use of it while RSW is not able to take advantage of it. If such power is not available however, FSW cannot feasibly process a large data set in form of the Salt Lake City data while RSW can do it on personal computers with only a small number of cores and little storage.
In summary, the inversion of the Salt Lake field data set shows the advantages and disadvantages of RSW. With the investigation of trace killing near-offset traces, I identified RSW's sensitivity to bad traces and shots. I also examined if both FSW and RSW could yield better results with more iterations which is not the case. Thus, I conclude RSW is capable of yielding a comparably resolving model with less iterations and ability to use little computational resources more efficiently. However, RSW's disadvantages, which are its sensitivity to bad traces and shots that degrade the final model severely and its inability to freely utilise as many multi-core processors as available, speak against RSW's usage for a field data set with a low signal-to-noise ratio or which has not been preprocessed carefully and thoroughly.

Chapter 7

Summary and conclusion

This work aimed to find an efficient way to apply the 2D FWI to a large land streamer data set. As the conventional workflow of FWI, the full-shot workflow, forward calculates the wavefields of all shots simultaneously and updates the starting model as a whole, it is computationally expensive and requires the use of HPC for large models and data. To bypass these restrictions, I introduced the random-shot workflow where only a subset of the data and only one shot with its corresponding receivers are used in a single inversion iteration. The model is therefore iteratively updated in subsets and the workload is divided into smaller portions.
I compared RSW to FSW in several different instances. At first I performed a synthetic benchmark to verify that RSW results are comparable to FSW results. This was done with the use of synthetic data that are generated in two synthetic models, an anomaly model and a layer model. I showed that the final models are indeed similar and that both manage to minimise the misfit to a satisfactory level.
Before I continued with the comparison, I introduced the Salt Lake City field data and showed some results from Liberty, St Clair, and Gribler (2018) who applied a traveltime tomography to the data. This data set represents a large land streamer data set that was acquired under difficult circumstances in downtown Salt Lake City. Hence, it was used to compare both FSW and RSW with real field data.
After its preprocessing, I used a subset of the data to take a look at the first results that FSW and RSW produce. RSW's final model shows a strong velocity contrast with seemingly unnatural anomalies and incoherent structures. To analyse the reason and improve the results, I investigated the influence of near-offset traces which are close to the source location and whether more iterations yield better results. From this investigation, I concluded that RSW is highly sensitive to bad traces and shots which degrade the final model heavily with artificial anomalies and that an inversion without the near offset traces are beneficial for both FSW and RSW. Furthermore, more iterations do not result in an improved final model. Thus, I recommended to use fewer iterations for the Salt Lake City data to save computing times.
The conclusions are used to apply FWI to the complete Salt Lake City field data. Comparisons with results from Liberty, St Clair, and Gribler (2018) showed that FSW yields a slightly more resolving model which is affected by some coherent numerical noise. RSW produces a comparable model, albeit bad shots disrupt it in certain areas which makes it difficult to differentiate between artificial and geological anomalies. Hence, it is unsuitable for data with low signal-to-noise ratio. In addition, both workflows do struggle with the possible 3-D effects that are not considered in the 2-D adjoint FWI and with the construction of the model at greater depths below 20 m.
In terms of the computational comparison, RSW manages to be more efficient with the computing power that it utilises. Its sensitivity to single shots also results in a stronger update of the model. Therefore, it does not require as many iterations as FSW to yield comparably resolving models. However, it cannot utilise computing

power as freely as FSW is able to do. Thus, on the one hand, if large computing power is available, FSW can fully utilise it with shot parallelisation whereas RSW inherently remains restricted. On the other hand, however, RSW has the possibility to be used with very limited computing power where FSW would not be feasible.

Bibliography

Adamczyk, A., M. Malinowski, and A. Malehmir (2014). "High-resolution near-surface velocity model building using full-waveform inversion -a case study from southwest Sweden." *Geophysical Journal International* 197(3), pp. 1693–1704. URL: https://doi.org/10.1093/gji/ggu070.

Aki, K. and P. Richards (2002). "Quantitative Seismology". *University Science Books*.

Brandes, T., E. Schricker, and T. Soddemann (2017). "The LAMA Approach for writing Portable Applications on Heterogeneous Architectures." *Scientific Computing and Algorithms in Industrial Simulations: Projects and Products of Fraunhofer SCAI*. Springer, pp. 181–198.

Bretaudeau, F., R. Brossier, D. Leparoux, O. Abraham, and J. Virieux (2013). "2D elastic full-waveform imaging of the near-surface: application to synthetic and physical modelling data sets." *Near Surf. Geophys.* 11(3), pp. 307–316. URL: https://doi.org/10.3997/1873-0604.2012067.

Brossier, R., S. Operto, and J. Virieux (2009). "Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion." *Geophysics* 74(6), WCC105–WCC118.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent (1995). "Multiscale seismic waveform inversion." *Geophysics* 60(5), pp. 1457–1473.

Courant, R., K. Friedrichs, and H. Lewy (1928). "Über die partiellen Differenzengleichungen der mathematischen Physik." *Mathematische Annalen* 100, pp. 32–74.

Dokter, E., D. Köhn, D. Wilken, D. De Nil, and W. Rabbel (2017). "Full waveform inversion of SH- and Love-wave data in near-surface prospecting." *Geophysical Prospecting* 65(S1), pp. 216–236. URL: https://doi.org/10.1111/1365-2478.12549.

Elboth, T., F. Geoteam, and D. Hermansen (2009). "Attenuation of noise in marine seismic data." *SEG Technical Program Expanded Abstracts*, pp. 3312–3316. URL: https://doi.org/10.1190/1.3255547.

Fichtner, A. (2011). "Full Seismic Waveform Modelling and Inversion." *Springer Science and Business Media*.

Fichtner, A., E. Saygin, T. Taymaz, P. Cupillard, Y. Capdeville, and J. Trampert (2013). "The deep structure of the North Anatolian Fault Zone." *Earth planet. Sci. Lett.* 373, pp. 109–117. URL: https://doi.org/10.1016/j.epsl.2013.04.027.

Fisher, R., S. Perkins, A. Walker, and E. Wolfart (1996). "Hypermedia Image Processing Reference (HIPR)." *J. Whiley and Sons, Ltd*, pp. 156–158.

Forbriger, T., L. Groos, and M. Schäfer (2014). "Line-source simulation for shallow-seismic data. Part 1: theoretical background." *Geophysical Journal International* 198, pp. 1387–1404. URL: https://doi.org/10.1093/gji/ggu199.

Gribler, G., L. Liberty, T. Mikesell, and P. Michaels (2016). "Isolating retrograde and prograde Rayleigh-wave modes using a polarity mute." *Geophysics* 81(5), pp. 379–385. URL: https://doi.org/10.1190/geo2015-0683.1.

Groos, L. (2013). "2D full waveform inversion of shallow seismic Rayleigh waves." Ph.D Thesis, Karlsruhe Institute of Technology.

Guasch, L., O. Calderón Agudo, MX. Tang, P. Nachev, and M. Warner (2020). "Full-waveform inversion imaging of the human brain." *npj Digit. Med. 3* 28. URL: https://doi.org/10.1038/s41746-020-0240.

Jaiswal, P., C.A. Zelt, R. Dasgupta, and K.K. Nath (2009). "Seismic imaging of the Naga thrust using multiscale waveform inversion." *Geophysics* 74, WCC129–WCC140. URL: https://doi.org/10.1190/1.3158602.

Jastram, C. (1992). "Seismische Modellierung mit Finiten Differenzen höherer Ordnung auf einem Gitter mit vertikal variierendem Gitterabstand." *PhD thesis* Universität Hamburg.

Komatitsch, D. and R. Martin (2007). "An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation." *Geophysics* 72(5), SM155–SM167.

Krieger, D. (2019). "2D elastic full-waveform inversion of land seismic data with topographic variations." Master's thesis, Karlsruhe Institute of Technology.

Kurzmann, A., L. Gaßner, R. Shigapov, N. Thiel, N. Athanasopoulos, T. Bohlen, and T. Steinweg (2013). "Real Data Applications of Seismic Full Waveform Inversion." *High Performance Computing in Science and Engineering ' 17* Springer, pp. 467–484. URL: https://doi.org/10.1007/978-3-319-68394-2_28.

Köhn, D. (2011). "Time Domain 2D Elastic Full Waveform Tomography." *PhD thesis* Christian-Albrechts-Universität zu Kiel.

Levander, A. (1988). "Fourth-order finite difference P-SV seismograms." *Geophysics* 53(11).

Liberty, L., J. St Clair, and G. Gribler (2018). "Seismic land-streamer data reveal complex tectonic structures beneath Salt Lake City". *SEG Technical Program Expanded Abstracts 2018*, pp. 2662–2666. URL: https://doi.org/10.1190/segam2018-2998568.1.

Machette, M., S. Personius, A. Nelson, D. Schwartz, and W. Lund (1991). "Wasatch fault zone, Utah - segmentation and history of Holocene earthquakes." *Journal of Structural Geology* 13(2), pp. 151–164. URL: https://doi.org/10.1016/0191-8141(91)90063-O.

McKean, A. (2014). "Interim geologic map of the Salt Lake City North quadrangle, Salt Lake and Davis Counties, Utah, contract deliverable to US Geological Survey NEHRP award G13AC00169.", p. 43.

Nocedal, J. and S. Wright (1999). "Numerical optimization." *Springer Science* 35, pp. 67–68.

Pan, Y., L. Gao, and T. Bohlen (2019). "High-resolution characterization of near-surface structures by surface-wave inversions: From dispersion curve to full waveform." *Karlsruher Institut für Technologie (KIT)* CRC 1173. URL: https://doi.org/10.5445/IR/1000089310.

Park, C., R. Miller, J. Xia, and J. Ivanov (2007). "Multichannel analysis of surface waves (MASW) - Active and passive methods." *Kansas Geological Survey* 26, pp. 60–64. URL: https://doi.org/10.1190/1.2431832.

Plessix, R.-E. (2006). "A review of the adjoint-state method for computing the gradient of a functional with geophysical applications." *Geophysical Journal International* 167(2), pp. 495–503.

Plessix, R.-E. and W. Mulder (2004). "Frequency-domain finite-difference amplitude-preserving migration." *Geophysical Journal International* 157(3), pp. 975–987.

Polak, E. and G. Ribière (1969). "Note sur la convergence de méthodes de directions conjuguées." *Revue Francaise d'Informatique et de Recherche Opérationalle* 16(3), pp. 35–43.

Pratt, R. G. (1999). "Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model." *Geophysics* 64(3), pp. 888–901.

Prieux, V., R. Brossier, S. Operto, and J. Virieux (2013). "Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field." *Geophysical Journal International* 194(3), pp. 1640–1664.

Reynolds, J. (1997). "An Introduction to Applied and Environmental Geophysics." *Wiley*.

Romdhane, A., G. Grandjean, R. Brossier, F. Rejiba, S. Operto, and J. Virieux (2011). "Shallow-structure characterization by 2D elastic full-waveform inversion." *Geophysics* 76, R81–R93. URL: https://doi.org/10.1190/1.3569798.

Schäfer, M., L. Groos, T. Forbriger, and T. Bohlen (2014). "Line-source simulation for shallow-seismic data. Part 2: full-waveform inversion - a synthetic 2-D case study." *Geophysical Journal International* 198, pp. 1405–1418. URL: https://doi.org/10.1093/gji/ggu171.

Scott, W. and R. Shroba (1985). "Surficial geologic map of an area along the Wasatch fault zone in the Salt Lake Valley, Utah, U.S. Geol. Surv." *Open-File Report*, pp. 85–448.

Sears, T. J., P. J. Barton, and S. C. Singh (2010). "Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to Alba Field, UK North Sea." *Geophysics* 75(6), R109–R119.

St Clair, J. (2015). "Geophysical investigations of underplating at the Middle American Trench, weathering in the critical zone, and snow water equivalent in seasonal snow". Ph.D Thesis, Univerisity of Wyoming.

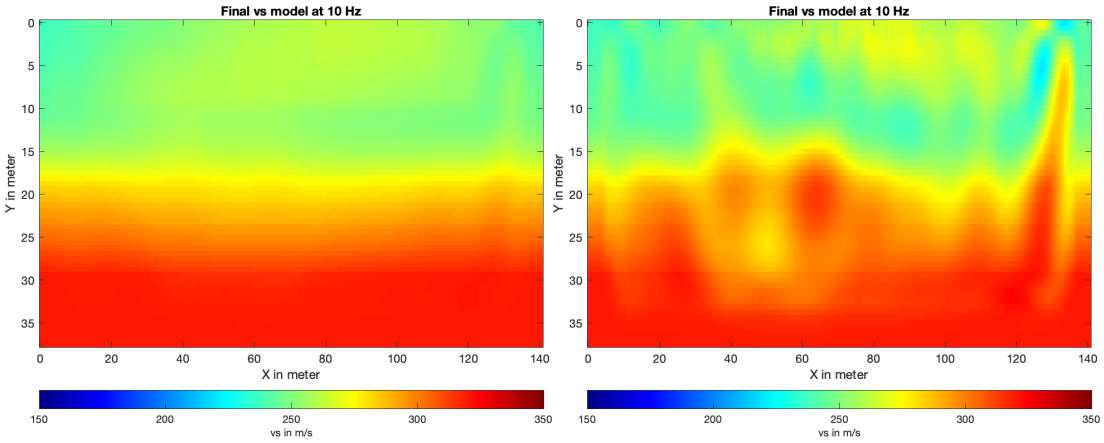
Tarantola, A. (1984). "The deep structure of the North Anatolian Fault Zone." *Inversion of seismic reflection data in the acoustic approximation.* 49(8), pp. 1259–1266.

Virieux, J. (1986). "P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method." *Geophysics* 51(4), pp. 889–901. URL: https://doi.org/10.1190/1.1442147.

Appendix A

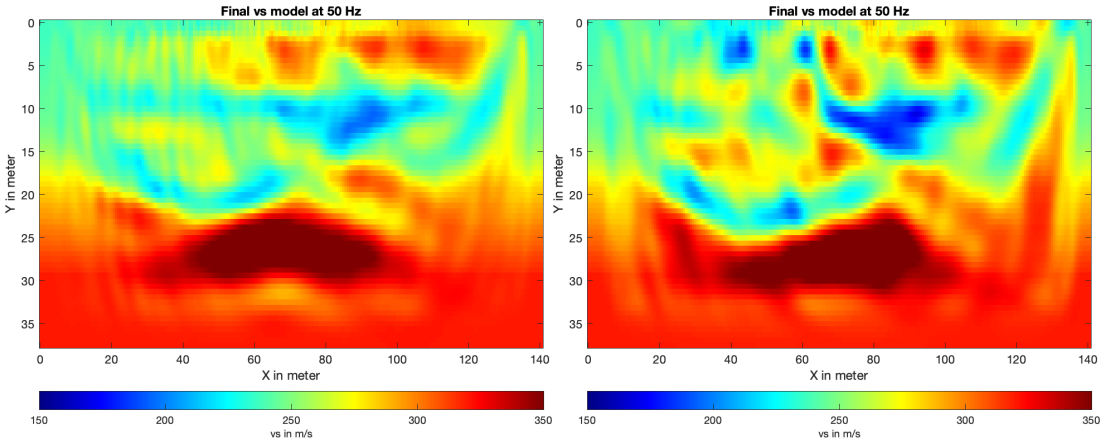
Results from the trace killing comparison

These are the additional results at stage 1 at 10 Hz and stage 9 at 50 Hz from the investigation of trace killing in chapter 6.2.3.



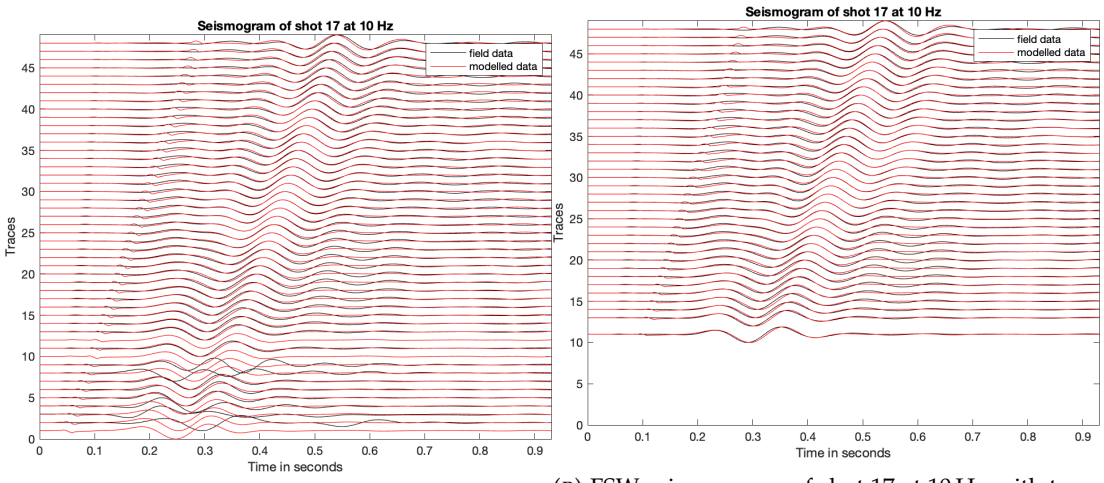
(A) FSW v_s model at 10 Hz (B) FSW v_s model at 10 Hz with trace killing

FIGURE A.1: FSW v_s subset models after stage 1 at 10 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) FSW v_s model at 50 Hz (B) FSW v_s model at 50 Hz with trace killing

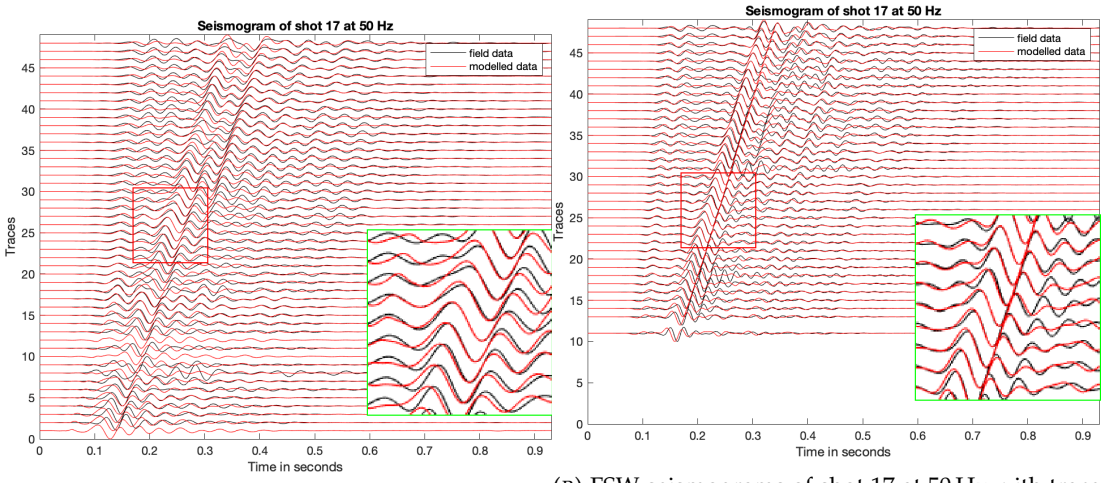
FIGURE A.2: FSW v_s subset models after stage 9 at 50 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) FSW seismograms of shot 17 at 10 Hz

(B) FSW seismograms of shot 17 at 10 Hz with trace killing

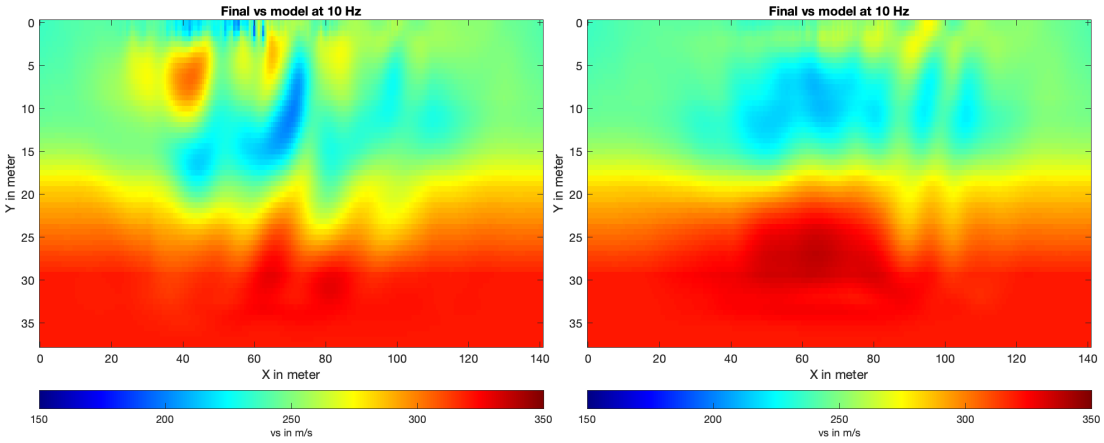
FIGURE A.3: FSW seismograms of shot 17 of the subset model after stage 1 at 10 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) FSW seismograms of shot 17 at 50 Hz

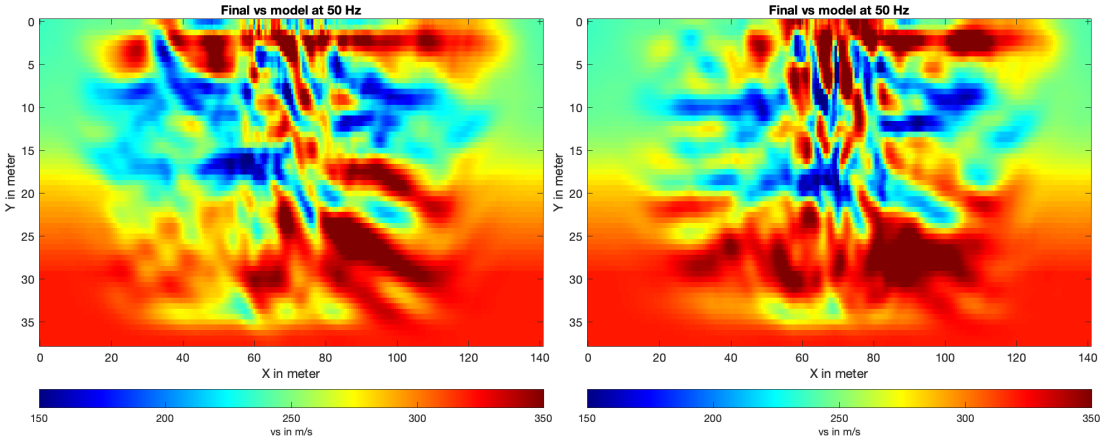
(B) FSW seismograms of shot 17 at 50 Hz with trace killing

FIGURE A.4: FSW seismograms of shot 17 of the subset model after stage 9 at 50 Hz with all near-offset traces (A) and with killed near-offset traces (B).



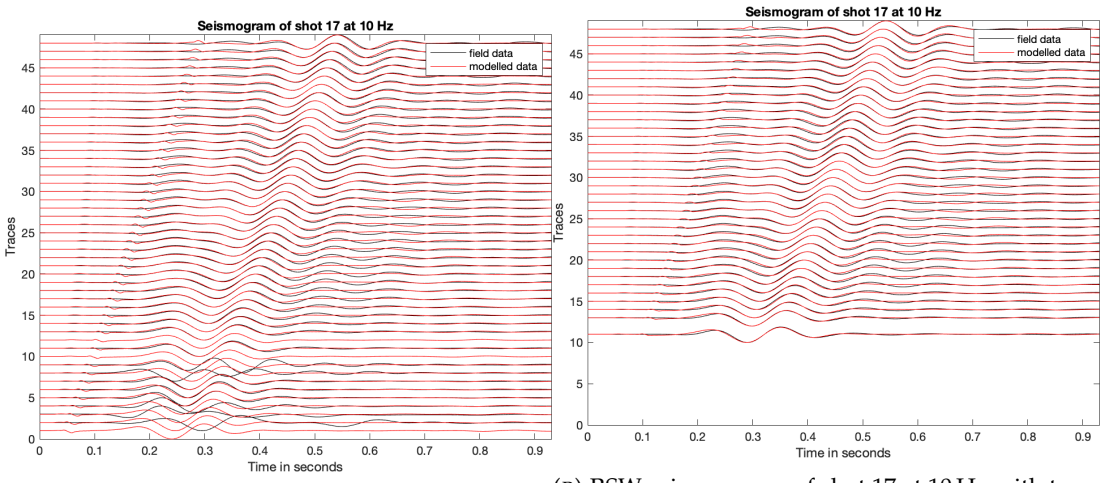
(A) RSW v_s model at 10 Hz (B) RSW v_s model at 10 Hz with trace killing

FIGURE A.5: RSW v_s subset models after stage 1 at 10 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) RSW v_s model at 50 Hz (B) RSW v_s model at 50 Hz with trace killing

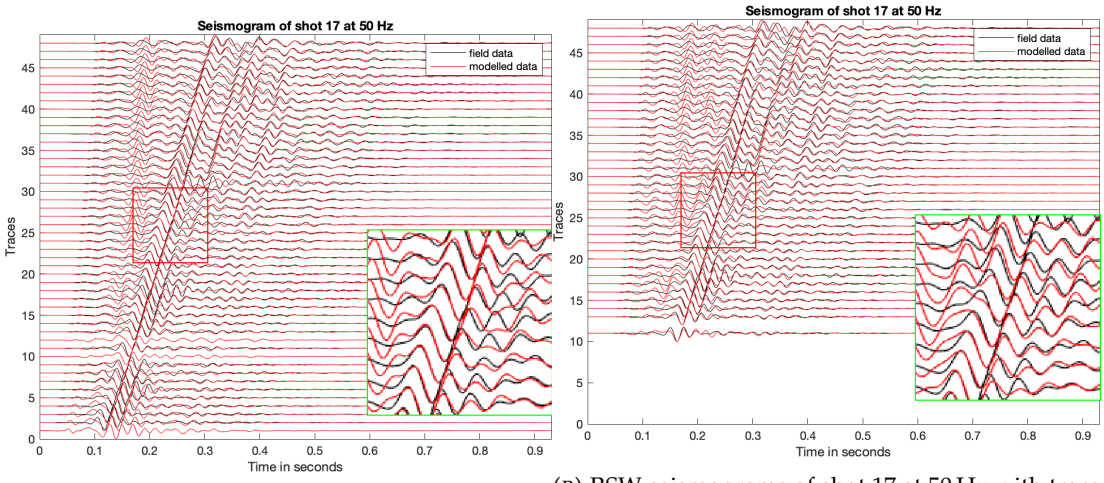
FIGURE A.6: RSW v_s models after stage 9 at 50 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) RSW seismograms of shot 17 at 10 Hz

(B) RSW seismograms of shot 17 at 10 Hz with trace killing

FIGURE A.7: RSW seismograms of shot 17 of the subset model after stage 1 at 10 Hz with all near-offset traces (A) and with killed near-offset traces (B).



(A) RSW seismograms of shot 17 at 50 Hz

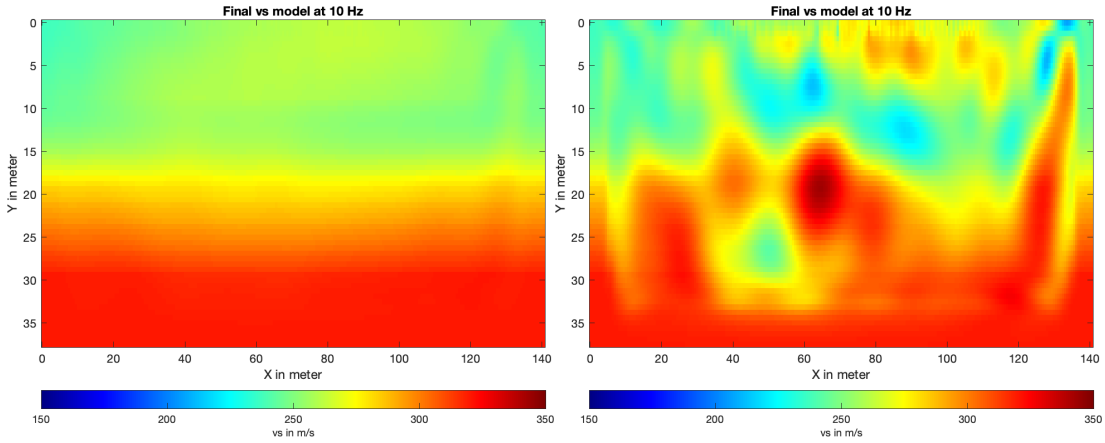
(B) RSW seismograms of shot 17 at 50 Hz with trace killing

FIGURE A.8: RSW seismograms of shot 17 of the subset model after stage 9 at 50 Hz with all near-offset traces (A) and with killed near-offset traces (B).

Appendix B

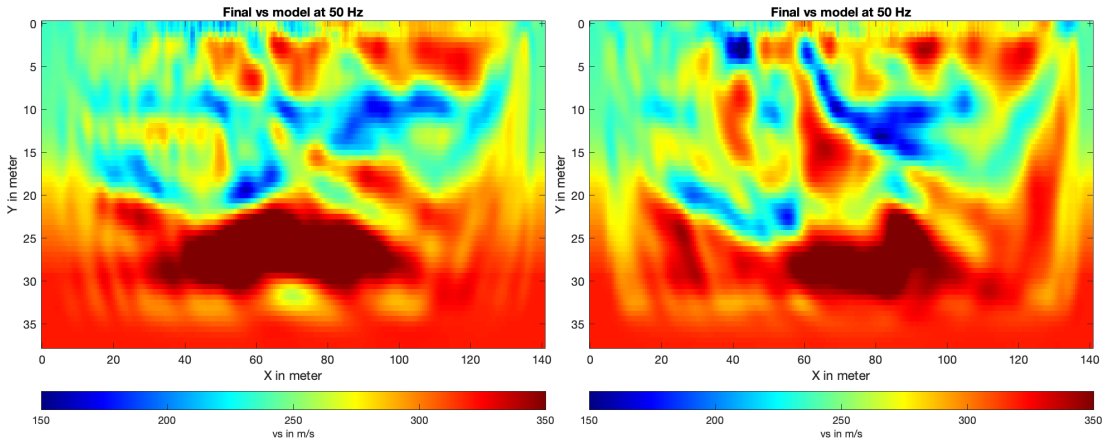
Results from the higher iteration number comparison

These are the additional results at stage 1 at 10 Hz and stage 9 at 50 Hz for FSW and the corresponding seismograms of shot 17 to figure 6.16 for RSW from the investigation if more iterations improve the results or not in chapter 6.2.4.



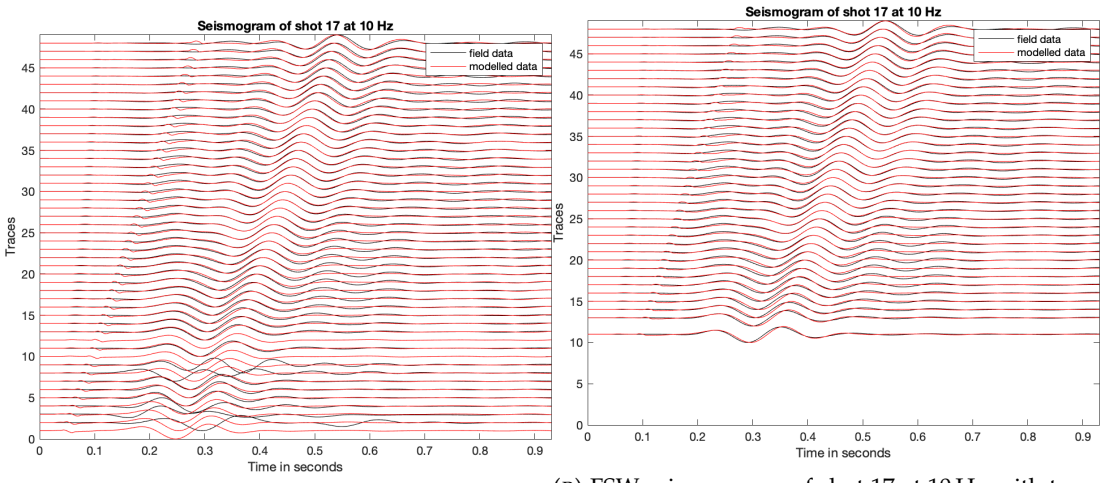
(A) FSW v_S model at 10 Hz (B) FSW v_S model at 10 Hz with trace killing

FIGURE B.1: FSW v_S subset models after stage 1 at 10 Hz with all near-offset traces (A) and with killed near-offset traces (B) after more iterations.



(A) FSW v_S model at 50 Hz (B) FSW v_S model at 50 Hz with trace killing

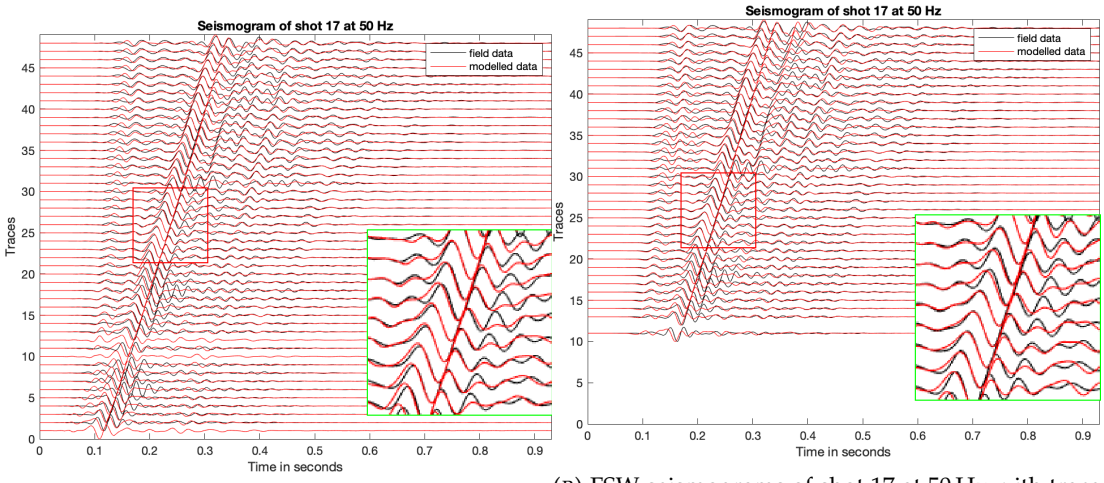
FIGURE B.2: FSW v_S subset models after stage 9 at 50 Hz with all near-offset traces (A) and with killed near-offset traces (B) after more iterations.



(A) FSW seismograms of shot 17 at 10 Hz

(B) FSW seismograms of shot 17 at 10 Hz with trace killing

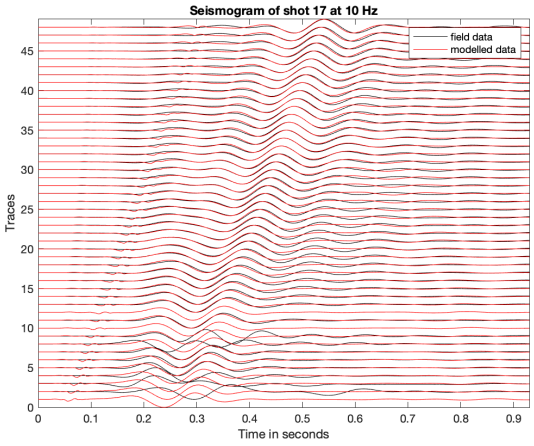
FIGURE B.3: FSW seismograms of shot 17 of the subset model after stage 1 at 10 Hz with all near-offset traces (A) and with killed near-offset traces (B) after more iterations.



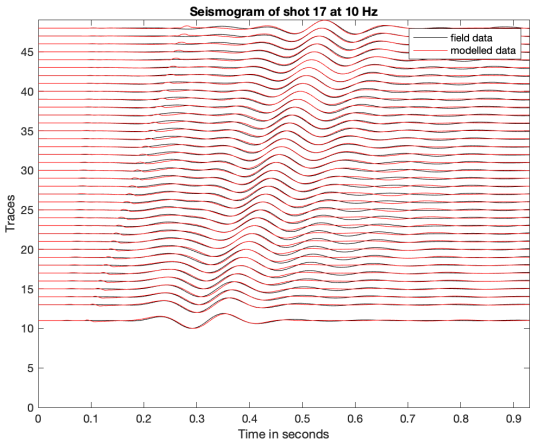
(A) FSW seismograms of shot 17 at 50 Hz

(B) FSW seismograms of shot 17 at 50 Hz with trace killing

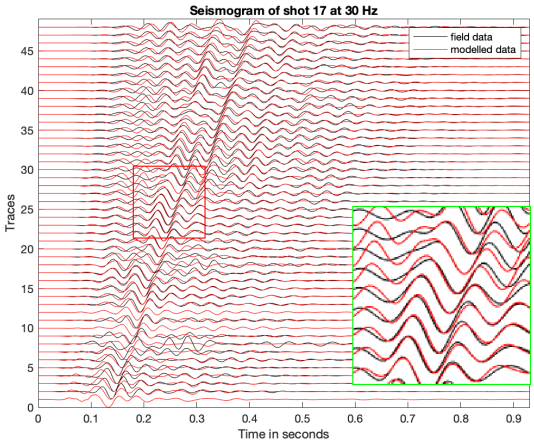
FIGURE B.4: FSW seismograms of shot 17 of the subset model after stage 9 at 50 Hz with all near-offset traces (A) and with killed near-offset traces (B) after more iterations.



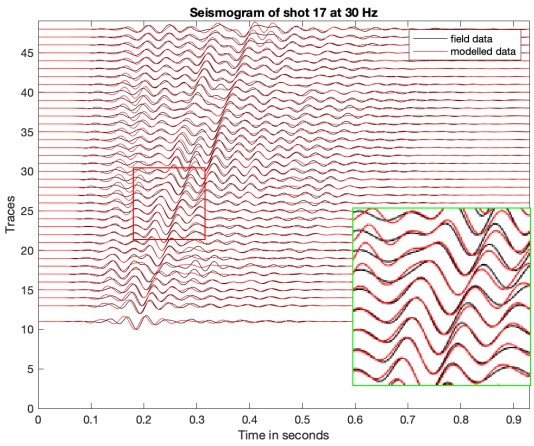
(A) RSW v_s seismograms at 10 Hz



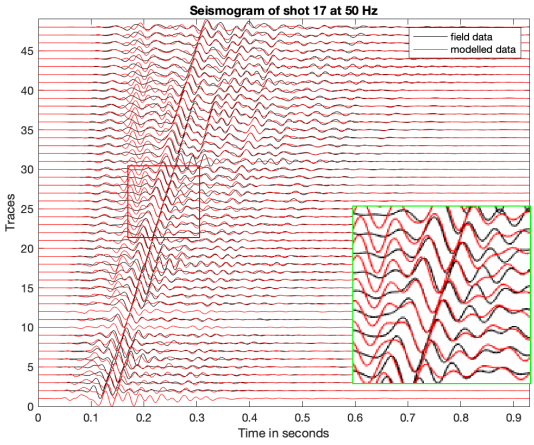
(B) RSW v_s seismograms at 10 Hz with trace killing



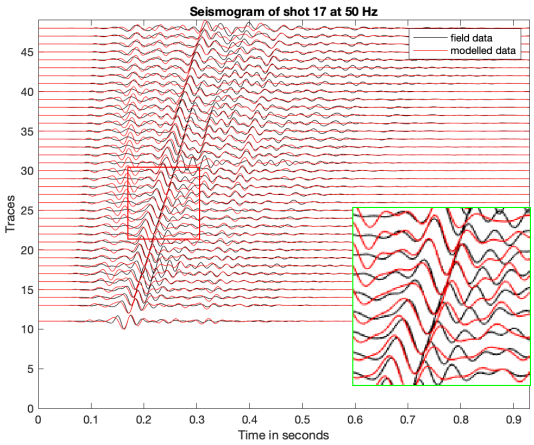
(C) RSW v_s seismograms at 30 Hz



(D) RSW v_s seismograms at 30 Hz with trace killing



(E) RSW v_s seismograms at 50 Hz



(F) RSW v_s seismograms at 50 Hz with trace killing

FIGURE B.5: RSW seismograms of shot 17 of the subset models after more iterations without trace killing on the left and with trace killing on the right. Each shot is iterated three times per stage.