University of Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Petroleum Geosciences Engineering	Spring semester, 2018 Open
Author: Mohamed Radwan	(signature of author)

Supervisor: Wiktor Waldemar Weibull

Title of master's thesis:
Application of Elastic Full-Waveform Inversion on Blackfoot Data

Credits: 30 ECTS

Keywords: Elastic Full-Waveform Inversion Elastic Reverse Time Migration Inverse problem Blackfoot Velocity modeling Finite Difference Multiscale approach L-BFGS Isotropic media Perfectly Matching Layer Cycle-skipping Imaging conditions	Number of pages: 82 Stavanger, June, 20th, 2018

UNIVERSITY OF STAVANGER

Faculty of Science and Technology

Department of Energy Resources

MASTER THESIS

Application of Elastic Full-Waveform Inversion on Blackfoot data

by

Mohamed Radwan

supervised by

Prof. Wiktor Waldemar Weibull

June 2018

Abstract

Estimating a high resolution image of the subsurface has been always a challenge in the oilfield exploration. With advances in computational power, Full Waveform Inversion (FWI) has been proven as an efficient imaging tool of the subsurface since the first paper of Tarantola A. (1984). FWI consist basically of two major steps: Forward modeling and backward propagation of data residuals. Modelling of the seismic wave-field requires a numerical solution of the partial differential equations like finite difference method, finite element method and spectral elment method. In this research, We implement finite difference method for modeling the seismic data. The used seismic modeling methodology is described in details in chapter 2. Numerical optimization methods like quasi-Newton, Conjugate gradient and Steepest Descent have been used for fitting the synthetics with data to improve the accuracy of the solution (Chapter 3).

We implemented the method using a real recorded seismic data from the Blackfoot oilfield in Alberta, Canada. The major challenge we faced in this research is to provide an accurate initial model to optimize better results. We applied FWI on the borehole seismic data and a new model was estimated from the borehole seismic data to be used as a starting model for inverting surface data. Chapter 4 shows, in details, the inversion of borehole seismic data while inversion of surface seismic data is described, in details, in chapter 5. This methodology helps to provide an accurate solution and better results. The resultant velocity models obtained from the application FWI on surface seismic data were used for Elastic Reverse Time Migration (ERTM) on the 2D seismic section and it was capable of developing a clearer image.

Acknowledgement

I would like to express my sincere gratitude to my supervisor Professor Wiktor Waldemar Weibull for encouraging me to express my ideas through this project. I will be always grateful for the perfect opportunity that Prof.Weibull gave me to work in this project that brought countless number of discussions that shaped my knowledge in Geophysics and developed my skills. Also, I am thankful to Prof. Weibull for the time and great help to find solutions for the challenges we faced in this project and the guidance in presenting the work at conferences.

I am grateful to University of Stavanger for giving me this amazing study opportunity and experience and for enabling the facilities to finish this thesis. And many thanks to my study colleagues for the the scientific discussions and the friendly atmosphere.

Also i would like to thank Khan Academy, edX, Quora and StackOverflow as they are making the knowledge available for everyone. Using these platforms, I was able to mitigate the challenges i meet and it helped me to expand my knowledge in programming, inverse theory and numerical optimization.

I'd like to thank UNINETT Sigma2 for making Abel, Fram and Stallo clusters available for us as it saved us the time of very expensive processes and many thanks to the support team for the great help.

I'd like to thank my friends Amr, Farid, Hisham, Ali, Ashraf And Amro for the great time we had during the last two years.

Finally I want to thank my family for the patience, encouragement and support.

Table of Contents

iv

List of Figures

vii

1 Introduction

1.1 Background

In the inverse problem, we aim to build a model from measurements. On the other hand, the forward problem is used to forecast measurements from a set of physical parameters. Simply, the forward problem is finding the appropriate physical formulation that allows predicting the measurements to an acceptable accuracy given a set of physical parameters.
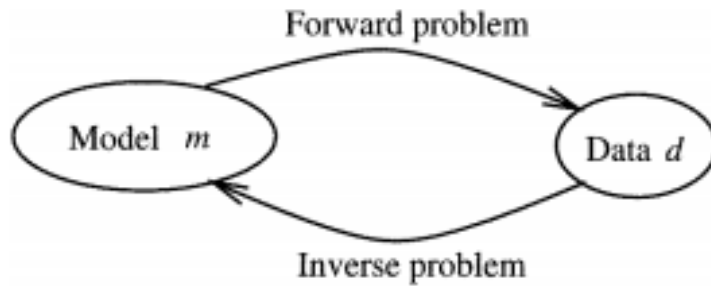
Figure 1.1: Conventional division of inverse and forward problem (Sneider, 1998)

1.2 Objectives

The purpose of this thesis is to implement Elastic Full Waveform Inversion (EFWI) on surface and borehole seismic data (VSP). We are looking to show how the accuracy of the initial model will maximize the potential of FWI and help obtain accurate results. Two initial models are used here: the first is a model linearly increasing in depth and the second is inverted VSP model. The updated models are used as an input for Elastic Reverse Time Migration (ERTM) for a better imaging of seismic data.

1.3 Basic Theory

FWI is a way to estimate an accurate model of the subsurface. It is an numerical optimization problem which aims to estimate a model that is close to reality by fitting the computed synthetics to the seismic data. FWI utilizes all content of the waveform, the amplitude and phase from the data. Having rich content of information is useful and challenging at the same time because that would make

the problem more complicated. Finding a solution for the complicated non-linear relationship between the model and the data is not an easy task as the gradient optimizer might get stuck in local minima and a slight change in the initial velocity value will result in a cycle skipping problem.

Full-waveform inversion (FWI) has been used as a major tool for high resolution imaging since the first paper of (Tarantola, 1984). It uses all measured events such as diving waves, critical reflections and multiples to develop an earth model out of these data. In this paper, We use elastic FWI to simultaneously estimate P-wave and S-wave velocities from real field data. FWI consist basically of two major steps: Forward modeling and backward propagation of data residuals. Modelling of the seismic wave-field requires a solution of the partial differential equations. Velocity models obtained through FWI were promising as they show the complexities of channels in the shallow part.

The elastodynamic equation can be expressed in the compact form as

$$M(x)\frac{d^2 u(x,t)}{dt^2} = K(x)u(x,t) + s(x,t) \tag{1.1}$$

Where M and K are mass and stiffness matrices (Marfurt, 1984), $s(x,t)$ is the seismic source and $u(x,t)$ is the elastic wavefield which represents horizontal and vertical particle velocities. The seismic wavefield $u(x,t)$ is related to the seismic source $s(x,t)$ by the impedance matrix $I(x,t)$

$$I(x,t)u(x,t) = s(x,t) \tag{1.2}$$

The misfit is the difference between the recorded and calculated data at each receiver positions. So, the application of the reduction operator R on the modeled wave-field u results in $d_{cal}(m)$.

$$d_{cal} = Ru \tag{1.3}$$

The main objective here is to reduce the misfit Δd.

$$\Delta d = d_{obs} - d_{cal}(m) \tag{1.4}$$

The least-squares formulation of this problem is to find a model that minimizes the objective function (Tarantola, 1987b). By taking the L2-Norm of misfit Δd.

$$f(m) = \frac{1}{2}\Delta d^{\dagger}\Delta d = \frac{1}{2}\sum_{r=1}^{N_r}\sum_{s=1}^{N_s}\int_0^{t_{max}} d\,t\,|d_{cal}(X_r, t; X_s) - d_{obs}(X_r, t; X_s)|^2 \quad (1.5)$$

Where $\dagger$ is adjoint operator of the matrix, N_r and N_s are number of receivers and sources. That leads to the perturbation model Δm which is added later to the previous model and it results in a new updated model as shown in Figure 1.2.

$$m = m_o + \Delta m \tag{1.6}$$

$$\Delta m = -\left(\frac{\partial^2 f(m_o)}{\partial m^2}\right)^{-1}\frac{\partial f(m_o)}{\partial m} = -H^{-1}\nabla f(m) \tag{1.7}$$

$$\nabla f(m) = \frac{\partial f(m_o)}{\partial m} \tag{1.8}$$

$$H = \frac{\partial^2 f(m_o)}{\partial m^2} = \frac{\partial^2 f(m_o)}{\partial m_j \partial m_i} \tag{1.9}$$

Where $\nabla f(m)$ and H are the gradient and hessian matrix. Misfit decreases along the conjugate gradient direction using Conjugate-gradient (Mora, 1987; Tarantola, 1987. Approximate evaluations of the hessian H or it's inverse (step length) is computed using quasi-Newton method like BFGS which take into account information from gradient $\nabla f(m)$ to find an approximation of the inverse hessian through iterations (See Chapter.3).

1.4 Workflow

This research can be subdivided into two major steps. The first step in this project is to invert the borehole seismic to get an accurate model of the subsurface since the borehole seismic data provides better constrains for vertical variations in velocities. The second step is to use the updated models resulting from the borehole seismic data inversion as a starting point for application of FWI on the surface seismic data, as explained in Figure 1.2, and to compare the quality of these results relative to

the results obtained using an initial model linearly increasing with depth..
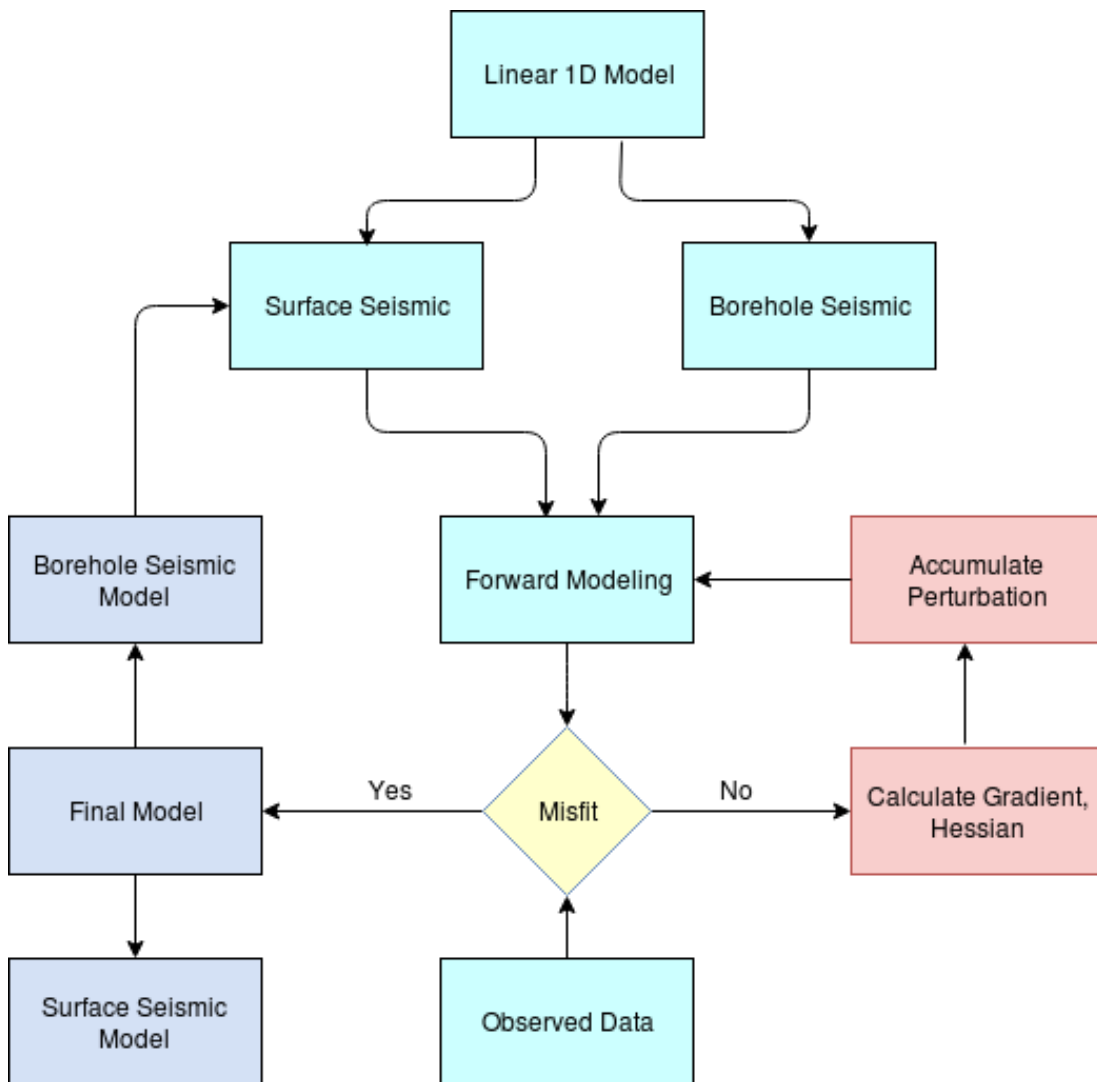
Figure 1.2: Used workflow

1.5 Dataset

The used data in this research is the Blackfoot 3C-3D data. The Blackfoot field is a series of stratigraphic traps in the western canadian basin in south central Alberta (Figure 1.3). The primary target in this field is the Glauconitic Sands (Depth=1550m) which is a fill valleys were incised channel filled with porous cemented sand and shale into the regional Lower Manneville Group. Stratigrahpic column of the cretaceous sequence in the Blackfoot oilfield is shown in Figure 1.4. The porosity in this sandstone is around 18% and it has cumulative production of more than 200

MMbbls[1] oil and 400 BCF gas (Miller etal., 1995).
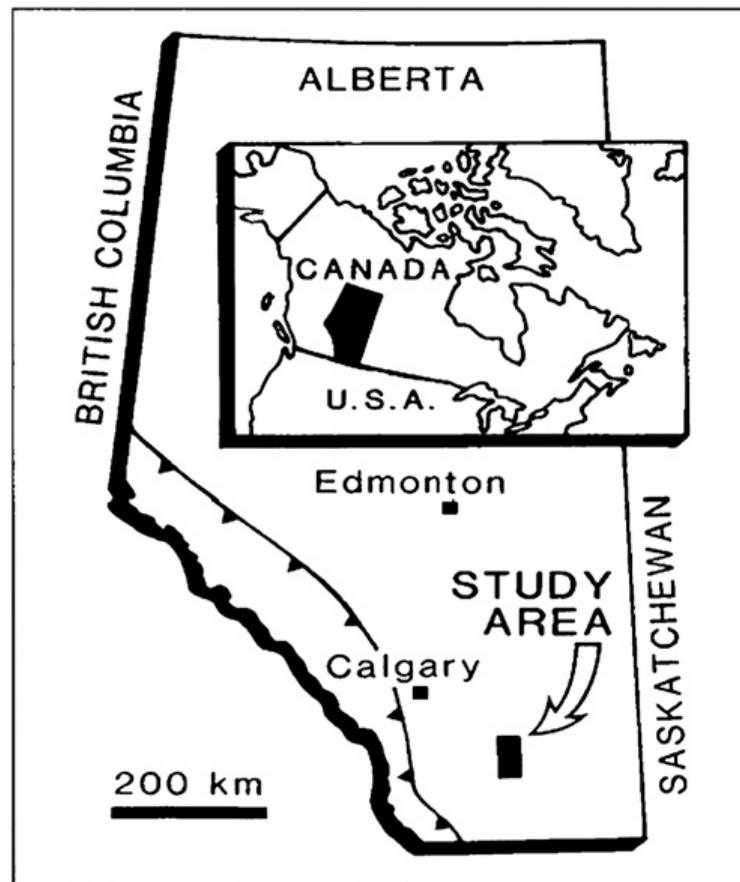
Figure 1.3: Location map of the study area in southern Alberta, Canada. From (Wood & Hopkins, 1992)

1.5.1 Surface Seismic

Acquiring Blackfoot 3D multicomponent seismic survey was sponsored by a group of Calgary based exploration companies and was planned and conducted by CREWES consortium and VectorSeis in 1990s hoping to identify channels and better characterize the reservoir in the field. Survey design is shown in figure 1.5. Sources are oriented North-South and receivers are East-West oriented. The Blackfoot data acquisition parameters are summarized in table 1.1. Figure 1.6 shows the fold coverage of the seismic survey.

---

[1]Millions of Barrels

5
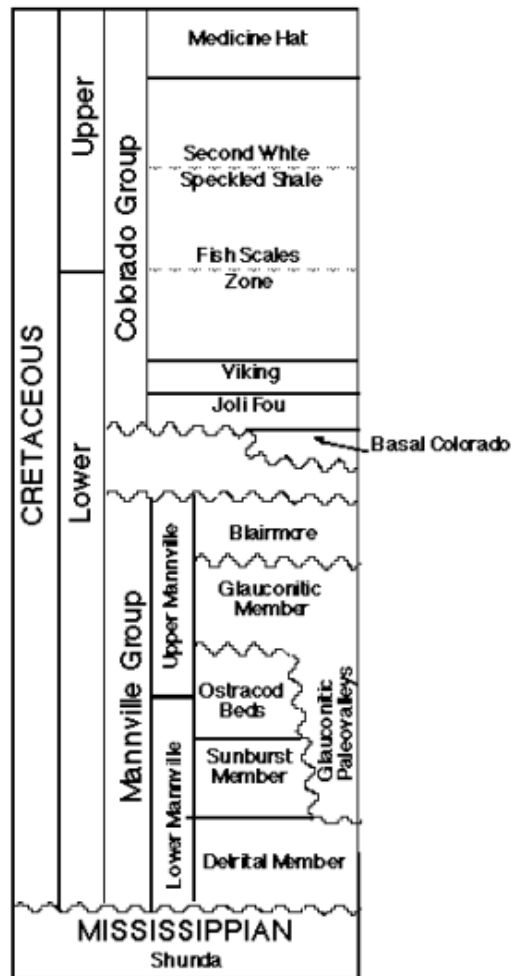
Figure 1.4: Stratigraphic column for the Cretaceous sequence (Wood and Hopkins, 1992)

1.5.2 Borehole Seismic

The 3D VSP was recorded at the same time as the surface 3D seismic in the PCP Blackfoot 12-16-23-23 well (Gulati, et al., 1998). The Blackfoot VSP survey shooting covers 2.5 km in the east-west direction and 2.6 km in the north-south direction. The experiment was acquired using 431 dynamite shot points on 12 north-south shooting lines with line spacing 210 m and shot interval of 60 m (Figure 1.7).

The data was recorded for 4 seconds with a sampling rate of 1 ms. Figure 1.8 shows the recorded data for vertical and radial components (East and North). An important phenomena of the borehole data is that the vertical component is well coupled compared to the radial components.

6
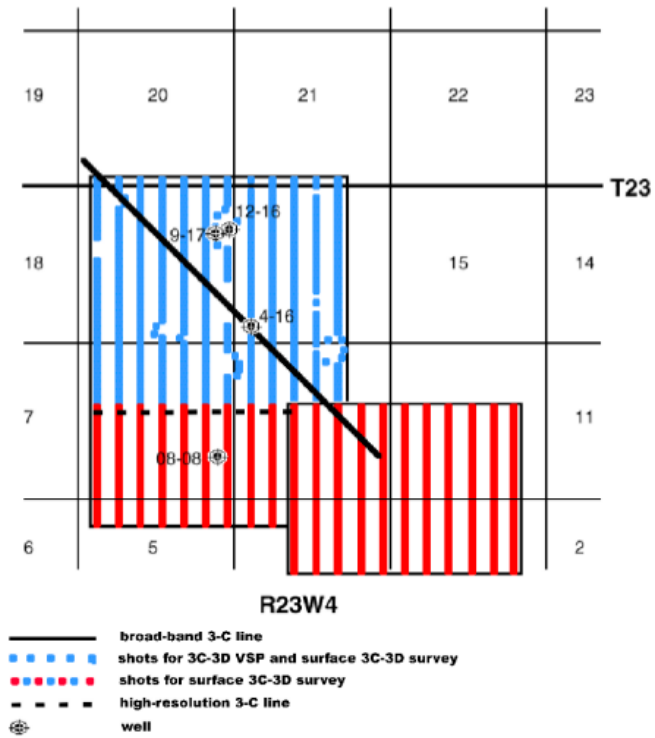
Figure 1.5: Surface and borehole survey design (Stewart, et. al.,1998))

P-S fold: offsets 300-1700 m
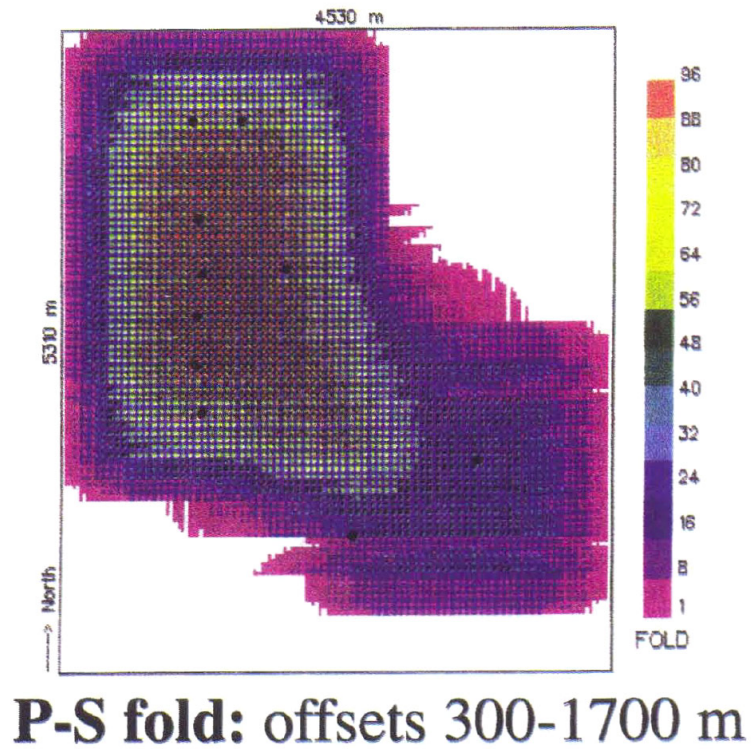
Figure 1.6: Blackfoot survey fold coverage (Stewart, et. al.,1996)

Energy source	dynamite
Source pattern	single hole, 4 kg @ 18 m
Sampling	2 msec
Number of channels	2070 channels, 690 channels per geophone component
Record length	2.0 seconds for vertical, 3.0 sec for horizontal
Receiver interval	60 m
Source interval	60 m
Source-line spacing	210m
Receiver-line spacing	255m
Spread	fixed

Table 1.1: Field data acquisition Parameters for the Blackfoot survey
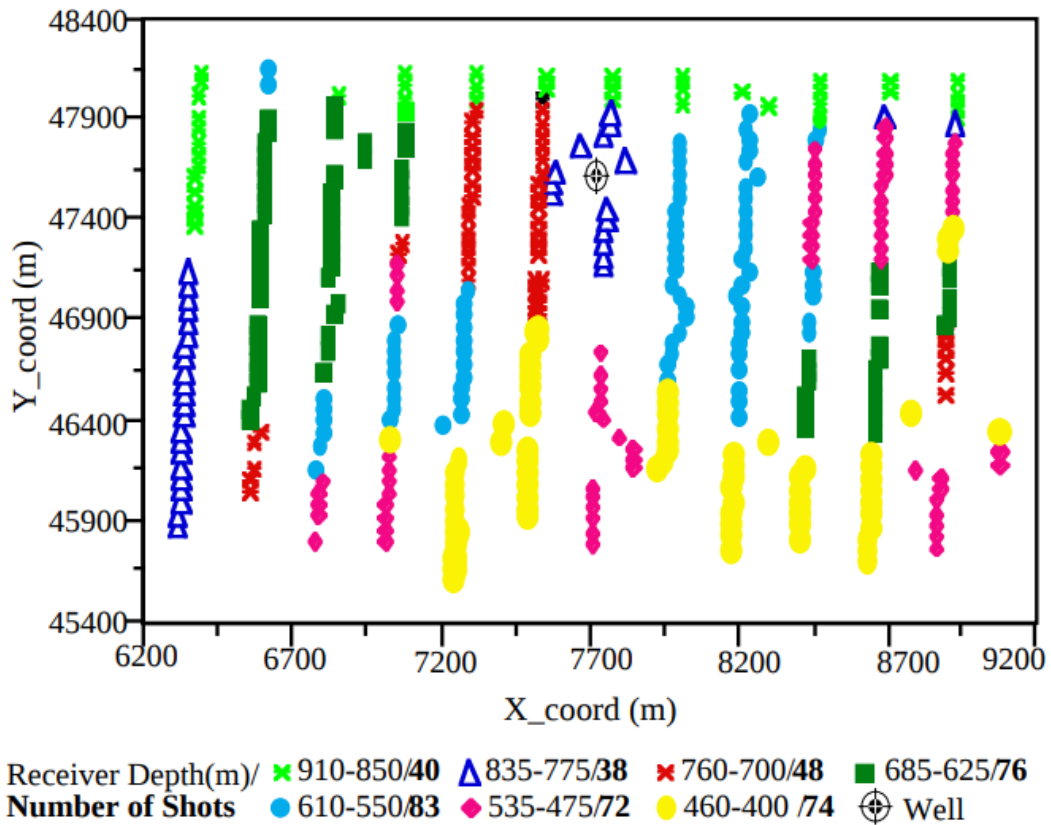
Figure 1.7: Survey shots and receivers depth (Zhang, et. al.,1996)
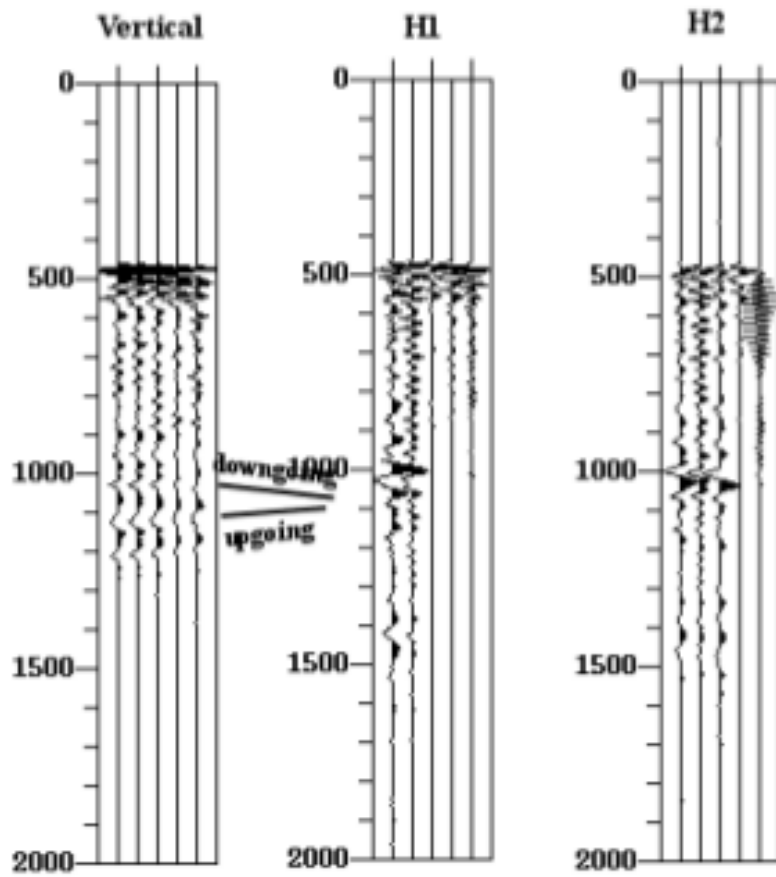
Figure 1.8: VSP Raw vertical and horizontal components of a shot at 115 m offset from the well with receiver tool at 850 m depth (Stewart, et. al.,1998)

2 Numerical Modeling

2.1 Background

The forward problem is dedicated to the estimation of seismic wavefields taking into account medium properties. Acoustic formulation is based on that wave propagation is defined by P-wave velocity V_p and density ρ while in the elastic formulation, S-wave velocity V_s is included. Other parameters can be included in case of anistropic modeling (Thomsen, 1986). Models described by these parameters are called vertically transverse isotropic (VTI) models. But with such kind of complexities, it will be very hard to model all properties as it's computationally very intensive and strongly non-unique to estimate. So, the strategy is to simplify the problem as possible using an isotropic model.

Many methods can be used for modeling seismic data. Some of these methods are very simple and do not need huge computational power like ray tracing which model the ray paths of waves based on the Snell's law. Other methods are more complicated, more accurate and more expensive like finite difference methods (Virieux, 1986; Levander, 1988), finite-element methods(Marfurt, 1984; Min, et al.,2003), finite-volume methods (Brossier, et al., 2008) and pseudo-spectral methods (Danecek & Seriani, 2008). We will focus out attention on the finite difference methods as it is the method that is implemented here since it's easy to implement and does not require the inversion of a very large impedance matrix.

2.2 Finite Difference method (FD)

Seismic waves verify partial differential equations (PDE) locally. So, one way of reconstructing the seismic wavefield u is FD by PDE discretization of the wavefield u_i with respect to time t and cartesian coordinates x,y and z. FD replaces the continuous partial derivatives by discrete approximations. So, the partial derivative of the discrecritized wavefield u_i with respect to x is approximated in the simple form from the centered FD with three nodes (second-order stencil) as follows:

$$\frac{\partial u_i^n}{\partial x} = \frac{u_{i+1}^n - u_{i-1}^n}{2\Delta x} + O(\Delta x^2) \tag{2.1}$$

Truncation error $O(\Delta x^n)$ which depends on the spacing Δx and the order of the approximation n. Accurate approximations can be obtained through the use of polynomials bases. Taylor series polynomial form the basis for more accurate solutions of PDE. Second-order central FD of taylor series (Bednar, 2013) is given by:

$$\frac{u(x+\Delta x)-2u(x)+u(x-\Delta x)}{\Delta x^2} = \frac{\partial^2 u}{\partial^2 m} + [\frac{\partial^4 u}{\partial x^4}\frac{\Delta x^2}{4!} + \frac{\partial^6 u}{\partial x^6}\frac{\Delta x^4}{6!} + O(\Delta x^2) \quad (2.2)$$

Similarly, higher order FD can be estimated for given integer n. It can be shown in the general form as:

$$\frac{u(x+k\Delta x)+u(x-k\Delta x)}{\Delta x^2} = u(x) + k^2\frac{\partial^2 u}{\partial x^2}\frac{\Delta x^2}{2!} + k^4[\frac{\partial^4 u}{\partial x^4}\frac{\Delta x^4}{4!}$$
$$+k^6\frac{\partial^6 u}{\partial x^6}\frac{\Delta x^4}{6!} + k^8\frac{\partial^8 u}{\partial x^8}\frac{\Delta x^8}{8!} + O(\Delta x^n) \quad (2.3)$$

If we chose to use eighth order scheme, the propagator will be nine grid nodes wide in each direction and eight volumes in memory for each step for $n = -4, +4$ which makes the solution very expensive in many cases. All waveform styles (for example, refractions, free-surface, and peg-leg multiples) are possible in this setting since these propagators synthesize full waveform data. Efficiency might be improved using smaller time and space sampling (Staggered Grid). So it means that the variables are not defined at the same node as in case of the collocated grid (Virieux, 1984 & 1986; Levander, 1988). However, size of the grid is still the same as the traditional equally spaced grid. So, staggering does not change the overall size of the problem. Using the higher order difference formulas (equation 2.3) solution for the system (equation A.19)

$$\frac{u(x+\frac{\Delta x}{2})+u(x-\frac{\Delta x}{2})}{2} = u(x) + \frac{1}{4}\frac{\partial^2 u}{\partial x^2}\frac{\Delta x^2}{2!} + \frac{1}{16}\frac{\partial^4 u}{\partial x^4}\frac{\Delta x^4}{4!} \quad (2.4)$$

2.3 Stability

"Small errors in the measurement data may lead to indefinitely large errors in the solutions" (Kabanikhin, 2008). When the solution is not stable, it will produce very large numbers that exceed numerical accuracy, i.e. dipping interfaces will appear

as stair-steps, where the edge of each step acts as a strong diffractor. A higher-order FD solutions requires 5 points per wavelength in a homogeneous medium (Levander, 1989) and 10-15 points per wavelength in case of heterogeneous medium. In other words, the ratio $\Delta t / \Delta x < \frac{1}{c}$ where c is the actual velocity as shown in Figure 2.1 which known as Courant-Friedrichs-Lewy stability condition (CFL).
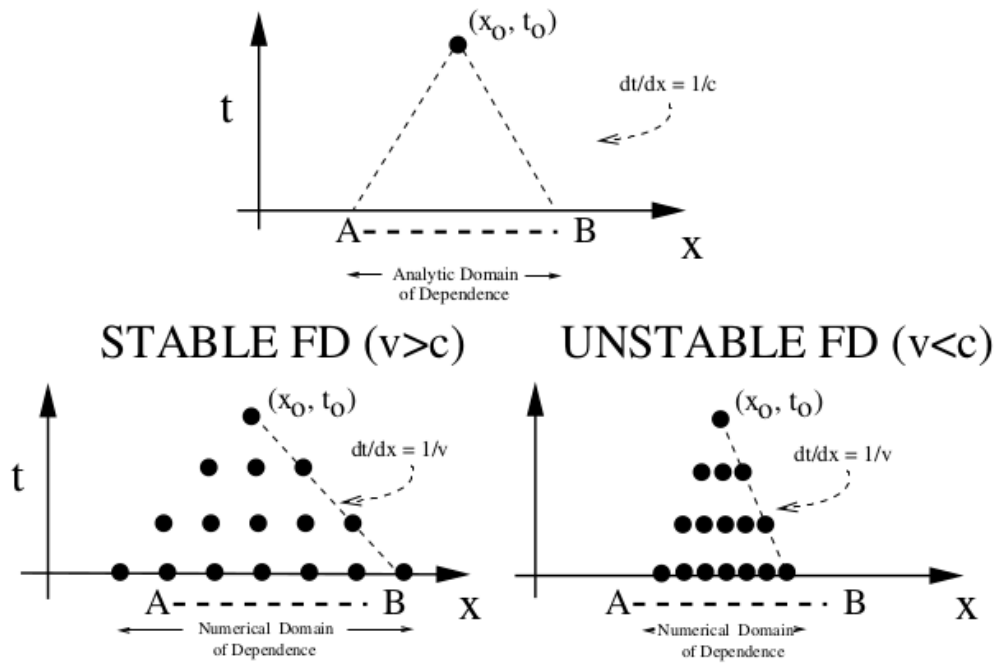
Figure 2.1: Effect of spacing on the stability of modeling. $\Delta t / \Delta x$ is selected to satisfy the stability condition, CFL satisfy initial conditions at $t = 0$ must have a numerical domain of dependence (DOD) larger than the analytical DOD (Mitchell & Griffiths, 1980) to avoid unstable FD solutions. Bottom left shows that numerical domain of dependence is wider than the analytical domain which is not the case in the bottom right, From (Schuster, 2015)

2.4   Model boundary

If the boundary condition is free, strong reflections will occur on the sides that interfere with the desired modeled waves. Absorbing silent boundary (Sponge boundaries) is needed to be applied to the edges of the model to transmit waves and not to reflect back inside. There are two popular methods are used to handle the free surface effect: sponge boundary conditions (Cerjan, et al., 1985) and perfectly

matched layers (PML) approach (Berenger, 1994; Chew & Liu, 1996; Festa & Vilotte, 2005).
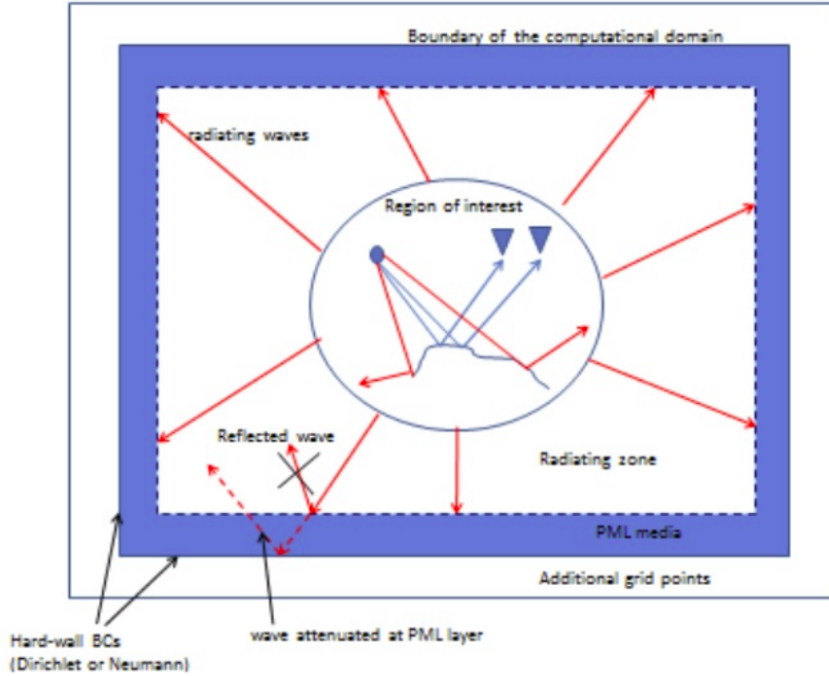
Figure 2.2: Finite Difference of PML scheme on 2D (Herrata & Weglein 2013)

PML is absorbing layers that are padded at the edges of the numerical model. The PML model can be easily formulated through equation splitting procedure. Each wavefield equation (Equation A.19) is split into perpendicular and parallel components based on spatial derivative separation and adding damping term to the perpendicular term (Hastings, et al., 1996; Collino & Tsogka, 2001). For instance, splitting and application of PML on the first equation on the x direction in the system (equation A.19) will result in the following equation.

$$\frac{\partial v_1}{\partial t} + d(x)v_1 = \frac{1}{\rho}\frac{\partial \tau_{11}}{\partial x_1}$$
$$\frac{\partial v_1}{\partial t} = \frac{1}{\rho}(\frac{\partial \tau_{12}}{\partial x_2} + \frac{\partial \tau_{13}}{\partial x_3}) \tag{2.5}$$

An example of damping term is from the work of (Marcinkovich & Olsen, 2013) using V_s is $d_{max} = \log(\frac{1}{R})\frac{tV_s}{nh}$ while R is theoretical reflection coefficient, h is PML thickness, n is model spacing and t is tuning parameter(3:4). The damping is from 0 inside the model to maximum value at the edge (Collino & Tsogka, 2001).
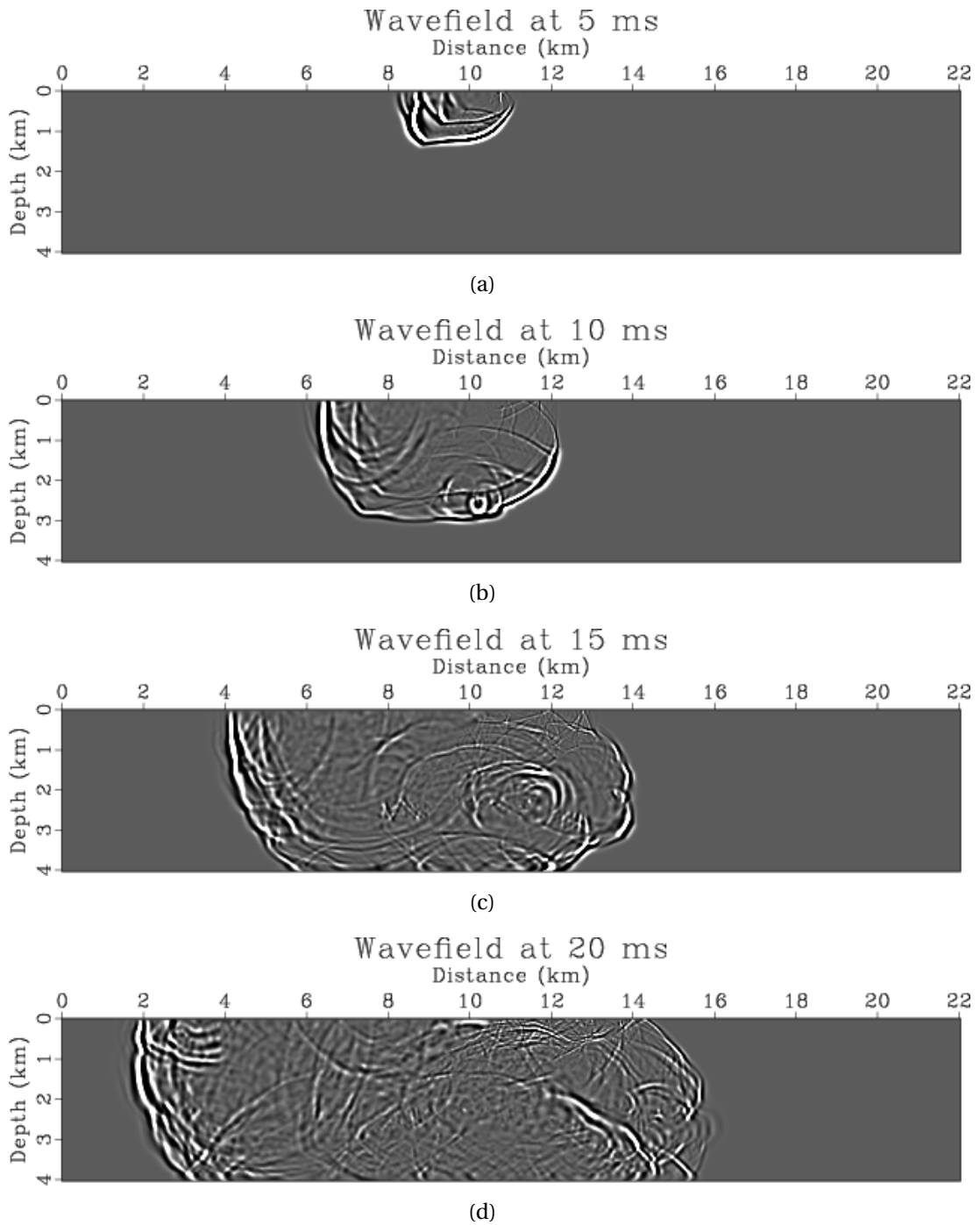
Figure 2.3: Amoco model propagating Wavefield using Finite difference modeling
(Source code from Madagascar repository (Irons, 2014))

3   Inverse Problem

The Inverse problem can be generally defined as inverting for the model m from the recorded data d.

$$d = Lm \tag{3.1}$$

where L is the forward modeling operator which is non-linear operator in the case in question. Having a non accurate starting model will give non-accurate wavefield calculation which will give unacceptable solution.

3.1   Non Linearity

If we assume that problem is simplified linear as shown in Figure 3.1.a, the solution will be very simple also. For instance, inverting the velocity taking into account that velocity increases linearly with depth, but the real case is not really well-posed. Lots of information and noises are included in the seismic data, which represents inconsistent system of equations, that make finding a solution for the inverse problem that justify these details very challenging and expensive. The wavefield representation is obtained that is the sum of of a term that depends linearly on the perturbation in the medium parameters plus a term that depends non-linearly on those parameters (Bleistein, 2000). Ill-posed problem doesn't have a unique solution so, it has many possible solutions. Many models can explain the data equally well or the solution is unstable "i.e. arbitrarily small errors in the measurements may lead to indefinitely large errors in the solutions (Kabanikhin, 200)". The remedy to this is to start with a model m_o close to the actual model and linearize as possible the relationship between the data and model, so the data can be inverted with acceptable convergence (Fletcher, 1980; Meyer, et al., 2004). One example of how to do this, is by using borehole data to create a good initial model at the vicinity of the well. Borehole seismic data is usually easier to invert than surface seismic data, since it is dominated by one way propagation, instead of two way as in short offset surface seismic data. Figure 3.2 shows that, in a non-linear problem, a sufficient initial model is needed. Model 1 will help to find the solution easier than models 2 and 3.
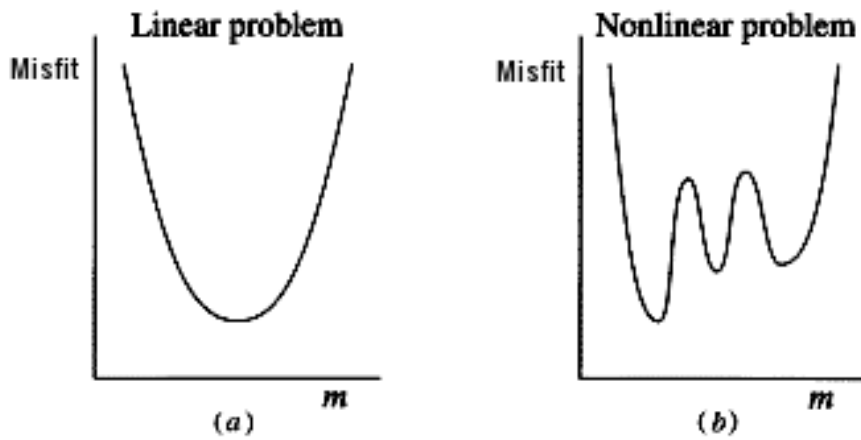
Figure 3.1: Effect of non-linearity results in having many local minima which is challenging for gradient optimizer to find the solution (Sneider, 1998)
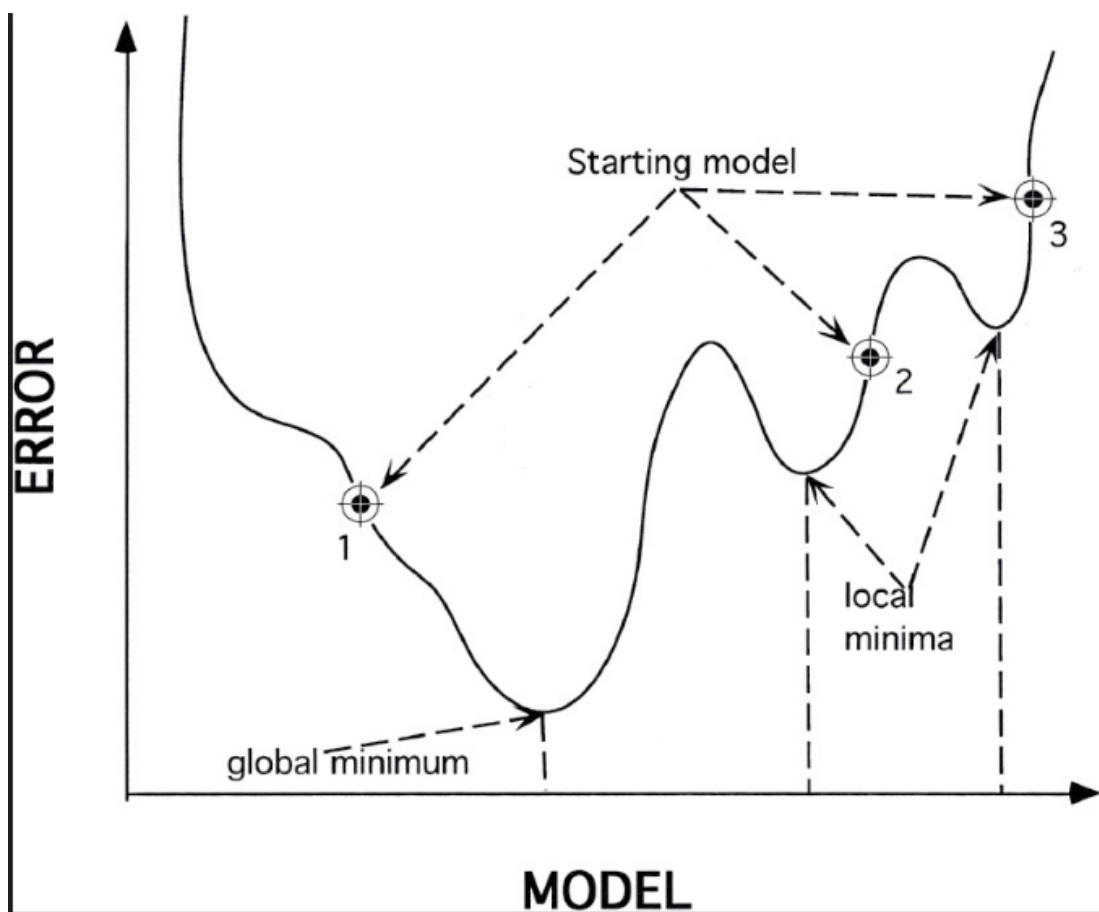
Figure 3.2: Effect of choosing initial model on the convergence, Note the model 1 will make it easier for the optimizer to find the solution (Sen & Stoffa, 2013)

Another solution for a complicated inverse problem is by using Low Frequencies as shown in figure 3.3. In the thesis, Low pass filter of 7.5 hrz was used on borehole and surface data.

Considering a discretized velocity model in the j_n direction denoted as m_j and
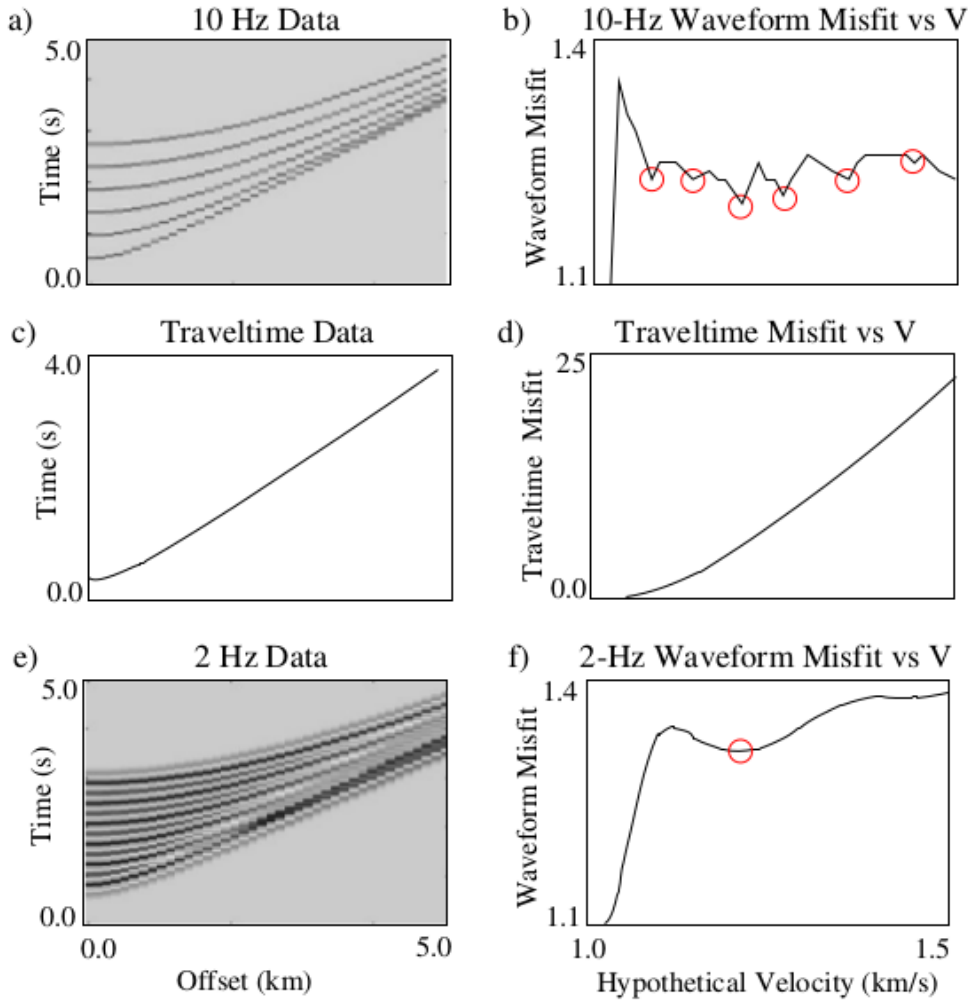
Figure 3.3: The effect of using low frequency on the optimization convergence. Inversion of the high frequency components will results in many local minima (red circles). Gradient optimization will get stuck in a local minima before it reach the global minimum, from (Schuster, 2017).

discretized data in the i_n direction denoted as d_i. The linearization starts with expanding the data d_i to first order Taylor series about initial model m_o close to the true model.

$$d_i(m) \approx d_i(m_o) + \sum_j [\frac{\partial d_j(m)}{\partial m_j}]|_{m_o} \delta m_j \tag{3.2}$$

model perturbation $\delta m = m - m_o$. Equation.3.2 can be rearranged to the following linear form.

$$\delta d_i(m) = \sum_j [\frac{\partial d_j(m)}{\partial m_j}]|_{m_o} \delta m_j \qquad (3.3)$$

And in the general matrix notation as

$$\delta d(m) = L \delta m \qquad (3.4)$$

and data residuals $\delta d_i = [d_i(m) - d_i(m_o)]$ is related to the model purturbation δm by the Jacobian matrix L (Fréchet derivative). i.e. raypath length l_{ij} in the j_n cell for the i_n ray for travel time tomography in the simplest case.

$$l_{ij} = \frac{\partial d_i(m_o)}{\partial m_j} \qquad (3.5)$$

So, We seek the model that best minimizes the objective function ϵ ,using adjoint-state method (Plessix, 2006), which is the sum of norm of data residual taken to the p power and the penalty term.

$$\epsilon = \frac{1}{p} \| L \delta m - \delta d \|^p + \eta g(m) \qquad (3.6)$$

Where $g(m)$ is the penalty function and η is small scalar that decides the tradeoff between reducing the penalty function at the expense of an increased value of the objective function. Many methods of regularization (i.e. Levenberg-Marquardt or Tikhonov) can be used to replace ill-posed problem with a well-posed problem to make the solution becomes closer to the priori. In our case, Total Variation regularization (TV) has been applied to resolve sharp interfaces while edges or discontinuities are preserved (Rudin et al., 1992). TV regularization is based on the principle that noisy data have higher level of total variations that is needed to be reduced (smoothed). TV regularization uses L1 norm of the gradient of objective function (second term in Equation 3.6) while η is the tradeoff parameter to be set that controls how much smoothing is performed to determines the relative balance of two terms of the equation.

3.2 Numerical Optimization

One way of dealing with the non-linearity challenge is to linearize the solution about background model m_o (Tarantola, 1987) using approximations like Taylor expansion and fitting the model to the data by minimizing the residual as expressed in Equation 1.5, or in other words, obtain an estimate of the least-squares solution $f(m^*)$ while m^* is the optimal solution. This is usually included regularization to penalize the models. Gradient optimization searches locally downhill and stops at the first local minimum which is cost wise way of convergence compared to the global optimization [2] methods. However, Gradient methods might get stuck in local minimum (Figure 3.4).
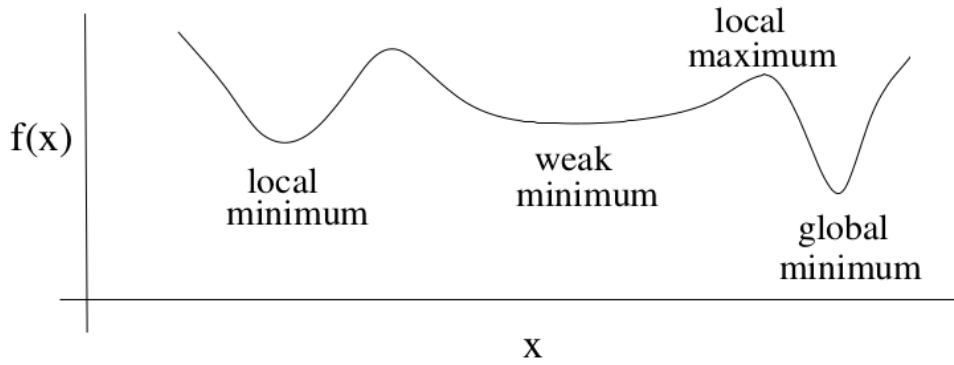
Figure 3.4: Gradient optimization gets stuck in the local minimum on the left along the objective function

3.2.1 Newton Method

Considering that the function $f(m)$ is differentiable with respect to model parameters and has unique global minimum. The gradient $\nabla f(m)$ is a vector containing the partial derivatives of the misfit function with respect to each model parameter. The negative of the gradient points to the steepest descent direction of $f(m)$ at some particular point m_o.

$$\nabla f(m) = \left(\frac{\partial f}{\partial m_1} \quad \frac{\partial f}{\partial m_2} \quad \cdots \quad \frac{\partial f}{\partial m_n} \right) \tag{3.7}$$

---

[2] Global optimization is to find m^* all over the model space and it avoids getting stuck in local minima. i.e. simulated annealing algorithms (Sen & Stoffa, 1992) and (Sen & Stoffa, 1995). However, this way is very expensive which make it an impractical solution for this case. In General, FWI is essentially a local optimization problem

And the curvature is quantified by the second derivative or Hessian[3] of the function $f(m)$.

$$H_{ij} = [\nabla\nabla^T f(m)]_{ij} = \frac{\partial^2 f(m)}{\partial m_i \partial m_j} = \begin{pmatrix} \frac{\partial^2 f}{\partial m_1^2} & \frac{\partial^2 f}{\partial m_1 \partial m_2} & \cdots & \frac{\partial^2 f}{\partial m_1 \partial m_n} \\ \frac{\partial^2 f}{\partial m_2 \partial m_1} & \frac{\partial^2 f}{\partial m_2^2} & \cdots & \frac{\partial^2 f}{\partial m_2 \partial m_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial m_n \partial m_1} & \frac{\partial^2 f}{\partial m_n \partial m_2} & \cdots & \frac{\partial^2 f}{\partial m_n^2} \end{pmatrix} \quad (3.8)$$

The misfit is the difference between the recorded and calculated data at each receiver positions. So, the application of the reduction operator R on the modeled wave-field u results in $d_{cal}(m)$.

$$d_{cal} = R u \qquad (3.9)$$

The goal here is to reduce the misfit Δd.

$$\Delta d = d_{obs} - d_{cal}(m) \qquad (3.10)$$

The least-squares formulation of this problem is to find a model that minimizes the objective function (Tarantola, 1987b). By taking the L2-Norm of misfit Δd.

$$f(m) = \frac{1}{2}\Delta d^\dagger \Delta d = \frac{1}{2}\sum_{r=1}^{N_r}\sum_{s=1}^{N_s}\int_0^{t_{max}} dt |d_{cal}(X_r, t; X_s) - d_{obs}(X_r, t; X_s)|^2 \quad (3.11)$$

Where † is adjoint operator of the matrix, N_r and N_s are number of receivers and sources. That leads to the perturbation model Δm which is later added to the previous model results in a new updated model.

$$m = m_o + \Delta m \qquad (3.12)$$

In the framework of the Born approximation, we assume that the updated model can be written as the sum of the starting model m_o plus a perturbation model(Δm). Taking the second order Taylor Lagrange Theorem of $f(m)$ will give the expression:

---

[3]Hessian matrix is symmetric matrix with all positive eigenvalues (PSD) which make it invertable. It's symmetric since the order of differentiation does not matter for second order derivative

$$f(m) = E(m_o + \Delta m) = f(m_o) + G^T \Delta m + \frac{1}{2} \Delta m^T \nabla \nabla^T f(m_o) \Delta m + O\|\Delta m\|^3 \quad (3.13)$$

While G is the gradient.

$$f(m_o + \Delta m) = f(m_o) + \sum_{i=1}^{N} \frac{\partial f(m_o)}{\partial m_i} \Delta m_i + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 f(m_o)}{\partial m_i \partial m_j} \Delta m_i \Delta m_j + O\|\Delta m\|^3$$
$$(3.14)$$

While $O\|\Delta m\|^3$ is error term and N is number of element in the model m. Truncating the last equation after the third term gives the quadratic model approximation (Linear). The error term is zero when the misfit function is a quadratic function of the model m. In FWI, since the relationship between the data and the model is non-linear so the inversion will take several iterations to converge reach the minimum. At the $k-th$ iteration, the misfit function is differentiated as follows:

$$\frac{\partial f(m)}{\partial m_k} = \frac{\partial f(m_o)}{\partial m_k} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 f(m_o)}{\partial m_i \partial m_j} \Delta m_i \delta_{ki} + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\partial^2 f(m_o)}{\partial m_i \partial m_j} \Delta m_i \delta_{kj} \quad (3.15)$$

$$\frac{\partial f(m)}{\partial m_k} = \frac{\partial f(m_o)}{\partial m_k} + \sum_{j=1}^{N} \frac{\partial^2 f(m_o)}{\partial m_j \partial m_k} \Delta m_j, k = 1, 2, ..., N \quad (3.16)$$

Equation 3.16 can be easily expressed in

$$\frac{\partial f(m)}{\partial m} = \frac{\partial f(m_o)}{\partial m} + \frac{\partial^2 f(m_o)}{\partial m^2} \Delta m \quad (3.17)$$

If m is reduced to the minimum point of $f(m)$, then $G = 0$ which leads to the these linear system of equations. In other words, if m is evaluated at the minimum point of $f(m)$ so equation 3.17 reduced to the following.

$$\Delta m = -\left(\frac{\partial^2 f(m_o)}{\partial m^2} \right)^{-1} \frac{\partial f(m_o)}{\partial m} = -H^{-1} G \quad (3.18)$$

$$G = \nabla f_m = \frac{\partial f(m_o)}{\partial m} \qquad (3.19)$$

Therefore, with a starting model m_o, gradient optimization method can solve equation 3.17 for Δm to minimizes the objective function $f(m)$.

$$H = \frac{\partial^2 f(m_o)}{\partial m^2} = \frac{\partial^2 f(m_o)}{\partial m_j \partial m_i} \qquad (3.20)$$

It's clear that Newton method [4] uses linear second order quadratic approximation which means that it takes into account the first and second order derivative. Gradient descent optimization is hessian free[5]. Gradient descent method searches for function minima based on the information of first order derivative. Newton method uses LU decomposition to find a solution trying to find where the derivative is zero. But, this solution is too expensive, iterative solution of m_i^k by scaled gradients sum is advised to save computation needs.

$$m_i^{k+1} = m_i^k - \sum \beta_{ij}^k G_j^k \qquad (3.21)$$

while β_{ij}^k is scaling of gradients (i.e. step length in steepest descent $\beta_{ij}^k = \delta_{ij} \alpha^k$ where α^k is the step length at the $k-th$ iterate). Then, quadratic approximation can be extended to higher order when the problem is more complicated.

$$m_i^{k+1} = m_i^k - \alpha (H^k)^{-1} G^k \qquad (3.22)$$

Where α is the scalar step length that is equal to one for a quadratic objective function and is be determined at each iteration for non-quadratic functions.

3.2.2 Steepest Descent Method

Steepest descent (SD) optimization, as shown in Figure 3.5, search iteratively along the negative gradient $-\nabla f(m^k)$. SD is computationally inexpensive because the hessian inverse is not needed, but it can suffer from slow convergence with ill-

---

[4]Newton method for simple 1D mathematical problem is $x_{t+1} = x_t - \frac{f'(x_t)}{f''(x_t)}$

[5]For a velocity model associated with 3D seismic needs $1000 * 1000 * 1000 = 10^9$ gridpoints of unknown velocities, so H calculation will require the storage and computation of 10^{18} elements.

conditioned hessians because it does not take into account information about the curvature. From Equation 3.22, Assuming the identity [6] hessian matrix is used, SD equation can be expressed as:

$$m^{k+1} = m^k - \alpha G^k \qquad (3.23)$$

While α is the the step length. The misfit function needs to be circular in order to approximate the Hessian to identity matrix. It means that the condition number $\theta = \frac{largesteigenvalue}{smallesteigenvalue} = \frac{\lambda_{max}}{\lambda_{min}} \approx 1$.
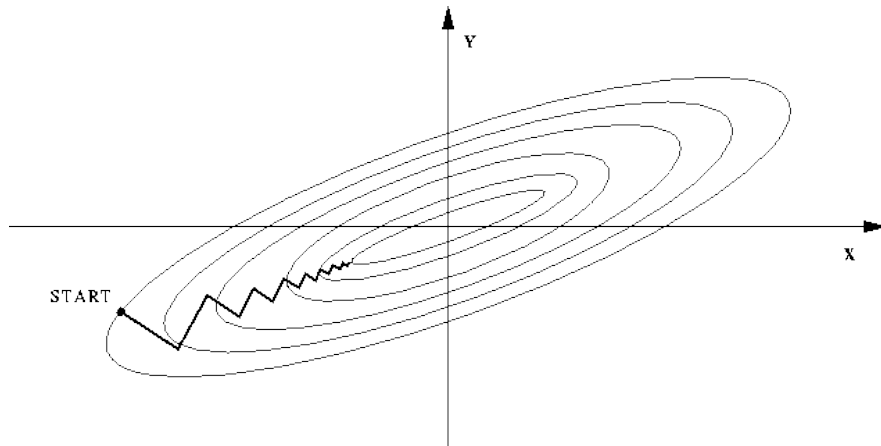

Figure 3.5: Misfit contours (after from Hjorteland, 1999). Search directions using SD method are orthogonal to one another and the step size gets smaller and smaller. The eigenvalues here represents elliptical function and this leads less accurate approximate and will slow down the convergence

3.2.3 Step Length

Having a fixed too large or too small [7] step size α will not give acceptable convergence (Figure 3.6). Convergence analysis will give a better idea which one is right. To compute α in the SD equation 3.23, there are several line search algorithms:

---

[6] Identity matrix is where diagonal elements are ones and other elements are zeros, Approximating identity hessian is when Hessian is diagonally dominant.

[7] Nocedal & Wright (1999) proposed a trust region for optimal model $m^k + \alpha \Delta m$ and restrict step size with radius r centered at m^k such that $\|\alpha \Delta m\| < r$. This restriction might be helpful for non-linear objective function that the gradient and curvature estimate might only be appropriate near m
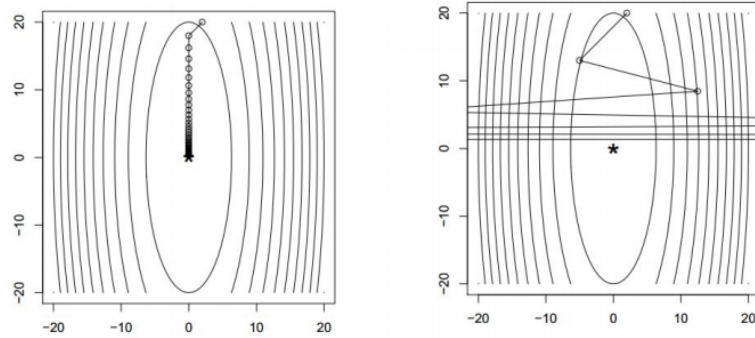
Figure 3.6: Too small step size that leads to slow convergence, however it found the optimal solution but with very slow convergence(left); and too large step size leads to divergence(right). (Boyd & Vandenberghe, 2004)

exact and numerical line search. We will focus here on numerical line search.

$$\arg\min f(m^k - \alpha \Delta m) \tag{3.24}$$

Backtrack line search: it adaptively choose the step length and repeatedly shrinks α by a factor t in each iteration, starting from 1 at first, until the Armijo condition is fulfilled to make sure step size is not exaggerated as shown in Figure 3.7. Armijo condition checks whether step from the current position to the next position reduces the objective function or not.
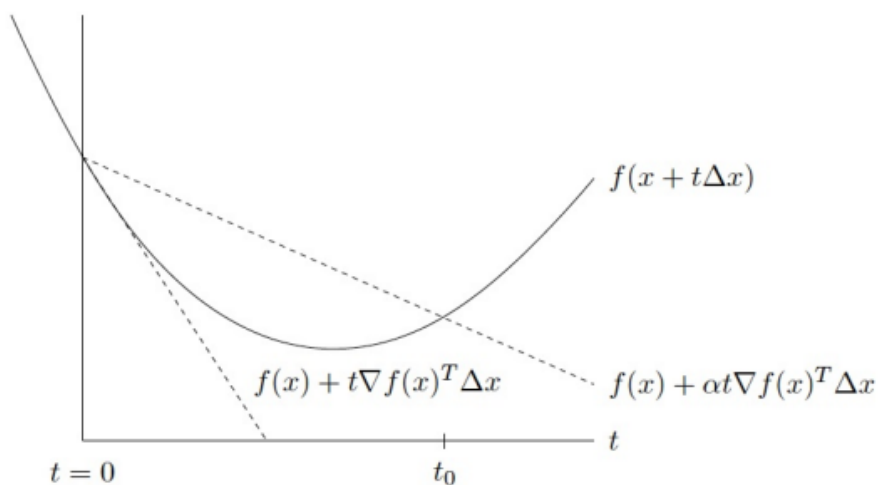

Figure 3.7: Backtracking line search (Note: x in figure refers to the model), Armijo condition is fulfilled when $t < t_0$. (Boyd & Vandenberghe, 2004)

$$f(m + \alpha \Delta m) \leq f(m) + \alpha t_1 \nabla f(m)^T \Delta m \tag{3.25}$$

Pairing this equation with the curvature condition (expression 3.26), to ensure that the slope has to be reduced in an acceptable way, forms the Wolfe conditions.

$$-\Delta m^T \nabla f(m + \alpha \Delta m) \leq -t_2 \Delta m^T \nabla f(m) \tag{3.26}$$

While $0 < t_1 < t_2 < 1$, Modifying the curvature condition gives:

$$|\Delta m^T \nabla f(m + \alpha \Delta m)| \leq t_2 |\Delta m^T \nabla f(m)| \tag{3.27}$$

The combination of 3.25 and 3.27 is the Strong Wolfe conditions.

3.2.4 Conjugate Gradient (CG)

Conjugate gradient CG and quasi-Newton methods don't need to calculate inverse of the hessian. CG searches along mutually conjugate directions pointing towards the misfit bullseye. Taking the FD approximation for the second derivative of the function gives.

$$\frac{d^2 f(m)}{d m^2} \Delta m = \frac{d f(m + \Delta m)}{d m} - \frac{d f(m)}{d m} \tag{3.28}$$

that can be rephrased as:

$$H(m^{k+1} - m^k) = g^{k+1} - g^k \tag{3.29}$$

Equation 3.29 is the quasi-Newton (QN) formula which relates the hessian to the difference between the misfit gradients over the two iterations. In CG, an initial search direction guess is determined based on SD based on the concept that previous iteration direction d^{k-1} and it's associated gradient g^k is perpendicular to each other. The direction to misfit minimum is on the plane spanned by d^{k-1} and g^k. So, the new iteration gradient g^{k+1} is perpendicular to any vector on the local plane, contains local minimum, between d^{k-1} and g^k while d^k is the conjugate direction

of d^{k-1} as illustrated in Figure 3.8.

$$g^k \perp d^{k-1} \tag{3.30}$$

$$g^{k+1} \perp d^{k-1} \tag{3.31}$$

Polak-Ribiere CG method can be explained in the following (Gill, et al., 1981; Nocedal & Wright, 1999)

$$m^{k+1} = m^k + \alpha^k d^k \tag{3.32}$$

$$d^{k+1} = g^{k+1} + \beta d^k \tag{3.33}$$

$$d^{k+1} = g^{k+1} + \beta d^k \tag{3.34}$$

where β is a scaler value to ensure conjugacy between d^k and d^{k-1}.

3.2.5 Quasi Newton (Q-N)

Q-N methods are used to save the cost since only first derivatives are computed and the hessian is estimated based on the computation of gradient. Equation 1.7 can be rephrased as

$$m^{k+1} = m^k - H_k^{-1} \nabla f_m \tag{3.35}$$

While k is the iteration and H_k is the previous iteration hessian and it can be the initial matrix estimate if it's equal to H_0 and α is the step length computed by line search. The matrix is updated by requiring that the updated hessian satisfy the Secant conditions which means quasi-Newton relation is valid for the updated hessian due to the fact that hessian matrix is invertable or, in other words, is PSD.

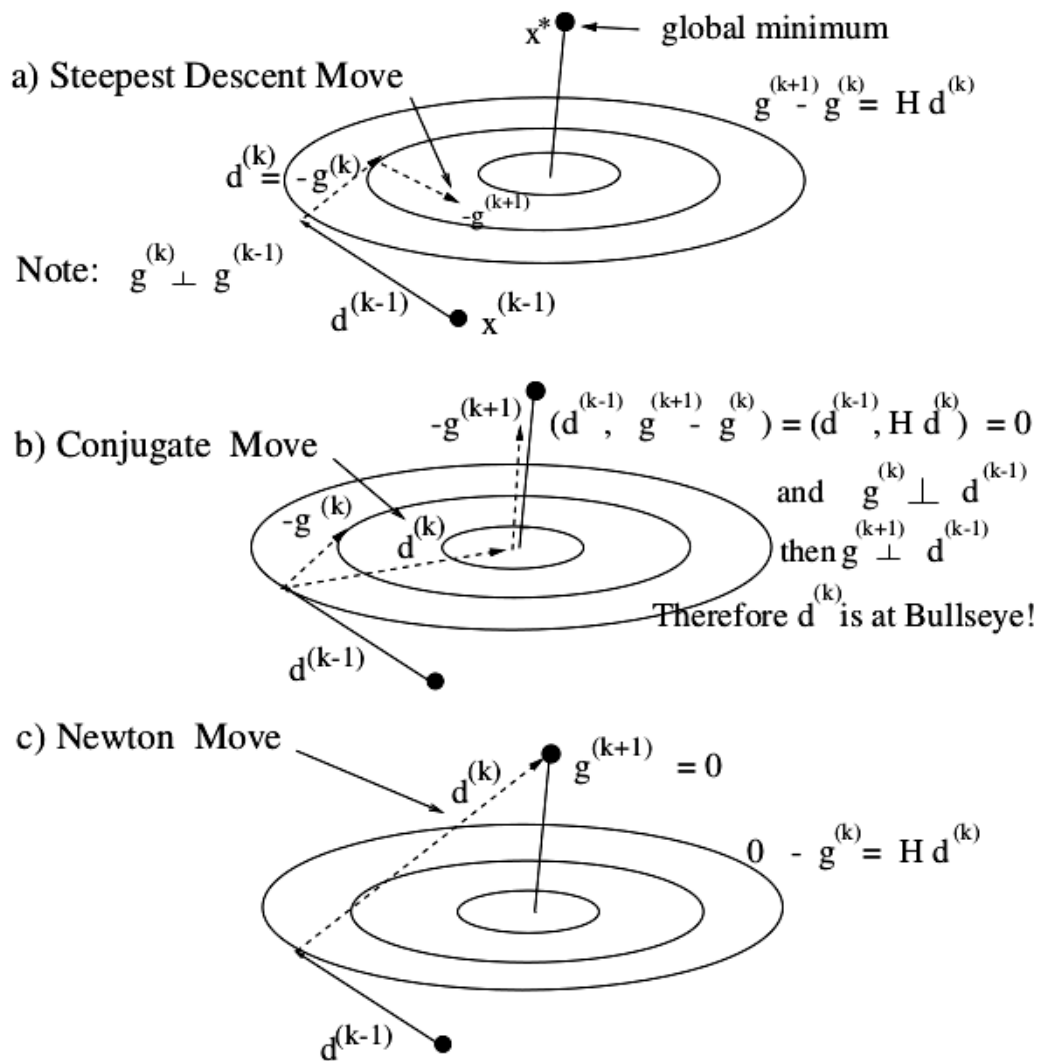$$m^1 - m^0 = H_1^{-1} (\nabla E_m^1 - \nabla E_m^0) \tag{3.36}$$

Figure 3.8: Comparison between Steepest Descent, Conjugate Gradient and Newton method, from (Schuster, 2015).

Taking into consideration that H_{k+1} satisfies the Secant conditions and it is a symmetric matrix, a solution can be found that gives the next expression (Schuster, 2015) which the BFGS solution. To find the optimal solution m^*, it must follows these two conditions $\nabla f(m^*) = 0$ and symmetric positive definite(SPD) hessian matrix $\Delta x^T \nabla \nabla^T f(m^*) \Delta x > 0$ (Kelly, 1999).

$$H_{k+1} = H_k + \frac{\Delta \nabla E_m^k (\Delta \nabla E_m^k)^T}{(\Delta m^k)^T (\Delta \nabla E_m)^k} - \frac{H_k \Delta m (\Delta m)^T H_k}{(\Delta m)^T H_k \Delta m} \tag{3.37}$$

While $\Delta \nabla E_m^k$ is the difference between the two gradients, and Δm is the difference the two models. Limited-memory BFGS (L-BFGS) was used here as it enables faster convergence and to reduce the cost since it doesn't save the matrices (Nocedal & Wright, 1999) but only save the model and gradient differences. The non-linear conjugate gradient and limited memory quasi-Newton methods can offer considerable speedup compared to the steepest descent method for FWI (Mora, 1987, 1989; Zhou, et al., 1995, 1997; Pratt, et al., 1998; Plessix, 2006; Plessix, et al., 2010; Métivier, et al., 2012).

4 Case Study: Borehole Seismic Data

4.1 Multiscale Approach

We will start with application of Elastic FWI on VSP seismic data. Multiscale approach was applied by gradually open the offset and time windows of the data until the whole data is inverted without cycle-skipping (Al–Yaqoobi & Warner, 2013). Multiscale method aims to invert the near source-receiver offset traces to find the velocity model that predicts data with no cycle-skipping due to poor initial model. For this data, the used multiscale strategy is based on the macro window as shown in figure 4.1.
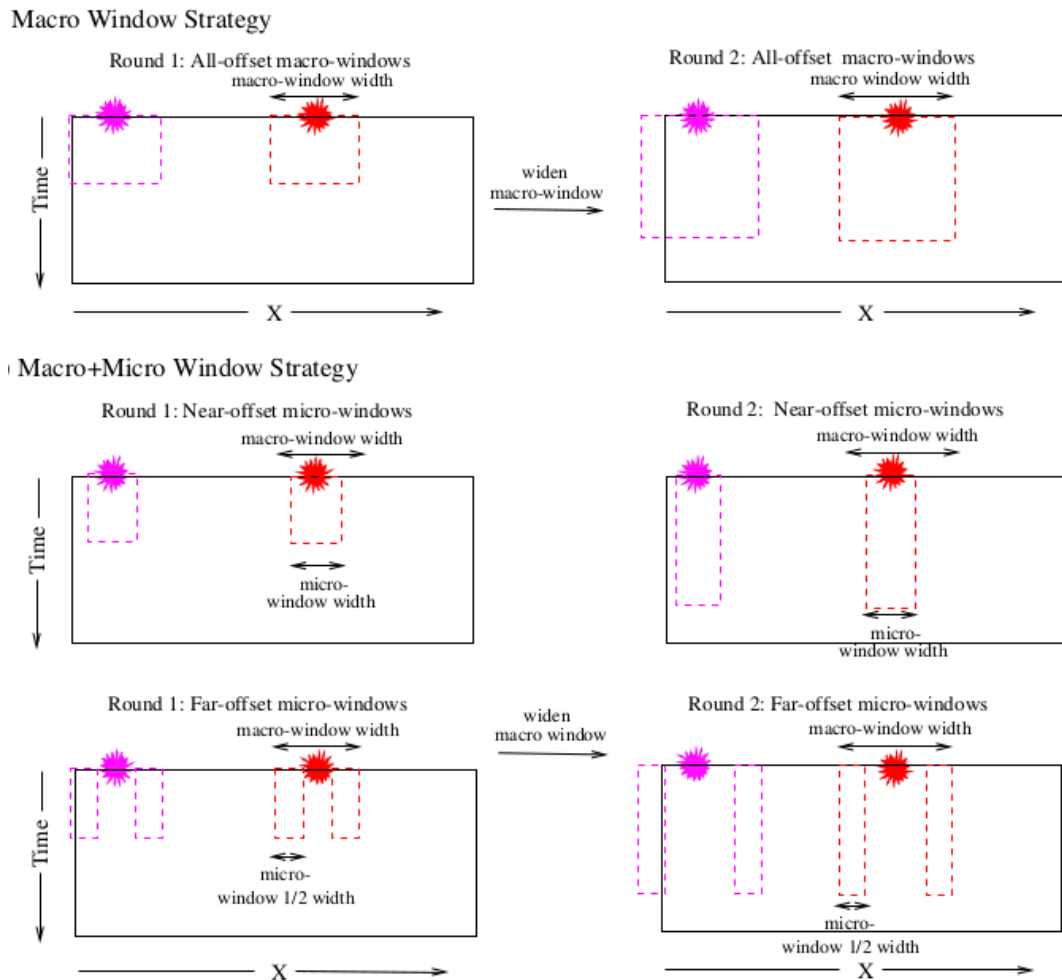
Figure 4.1: The windowing strategy: Macro window from (Warner, et al. 2013) and Micro-Macro windowing from (AlTheyab & Dutta, 2014) of rolling offset strategy. After (Schuster, 2015)

4.2   Results

Wavelet extraction (Figure 4.2) is based on the concept that it's the wavelet that should have been used if L2 difference between observed and synthetic is to be minimized. Low Pass filter of 7.5 hrz was applied to the wavelet and the data. FWI updates were muted for the very shallow part of the model to reduce the effect of non-accurate free surface. The used initial models (Figure 4.3) for FWI is a 1D linear layered model which was built on the basis of the matching travel time ray tracing with the real travel times (Gulati, Stewart & Parkin, 1998). Modeling of seismic data is by using Finite Difference with 8 stencils (see FWI configuration file in Appendix D). Quasi-Newton (L-BFGS) optimization algorithm was used here with a line search based on the Wolfe condition.
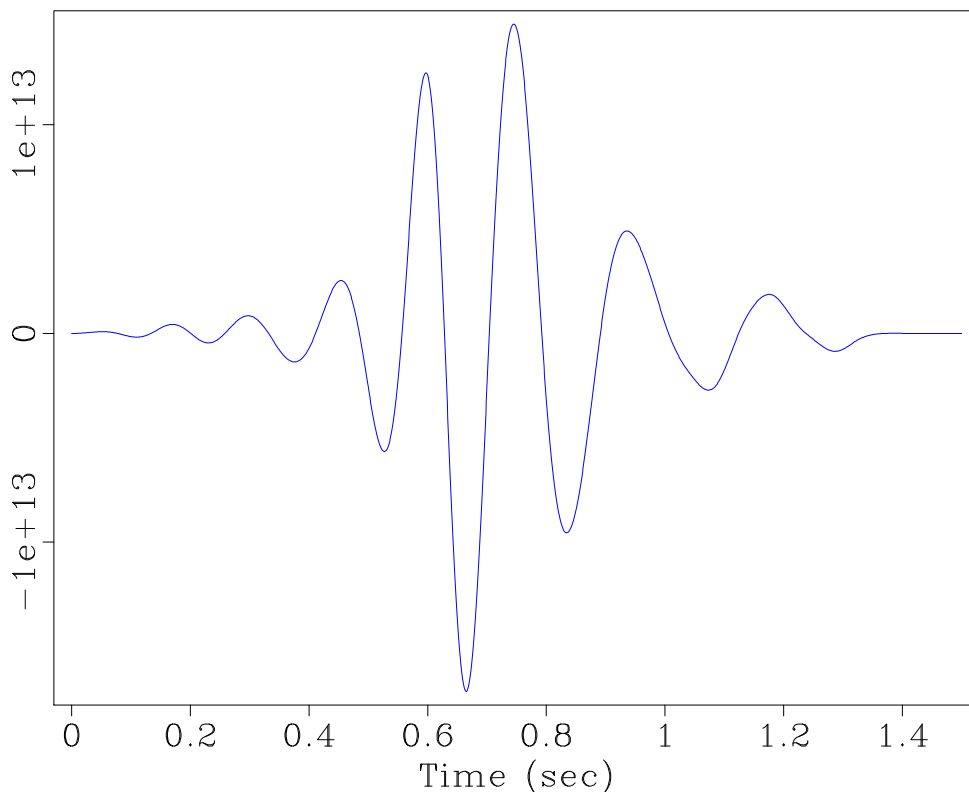
Figure 4.2: Estimated Wavelet

Figures 4.4, 4.5 and 4.6 show a comparison between the whole dataset and synthetics for vertical and radial components. Figures 4.7, 4.8, 4.9 show subset of the data and synthetics and it's clear that synthetics show a good matching with the data for P-wave (recorded around 1sec) and S-wave events (recorded around 1.5sec).
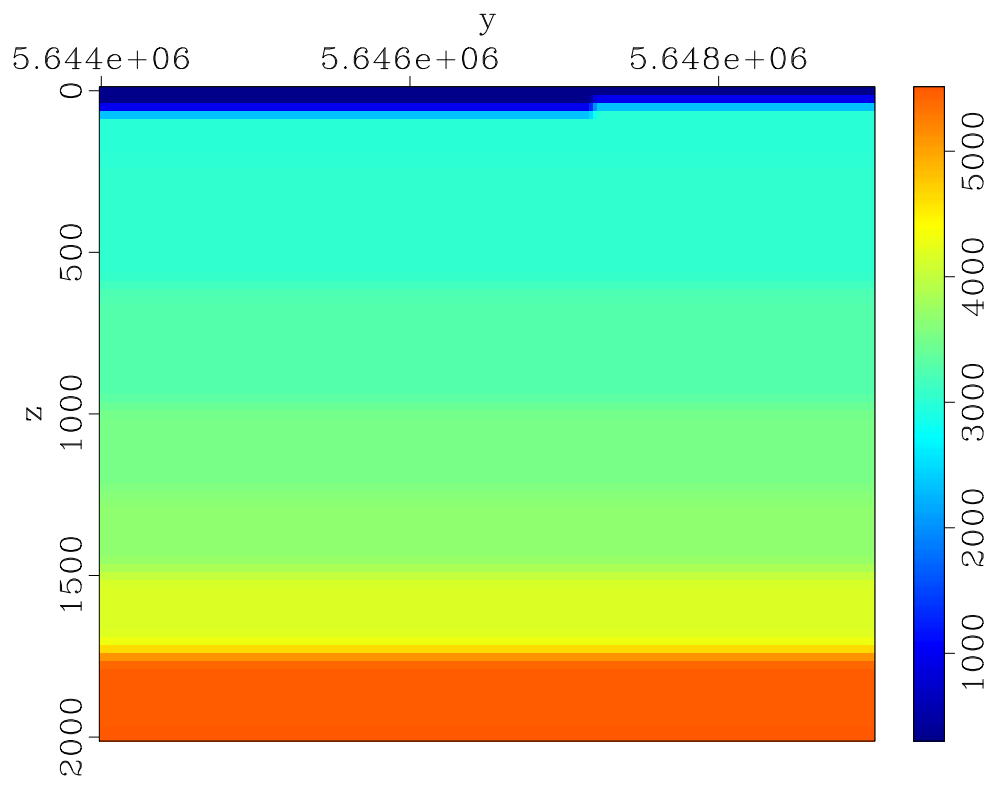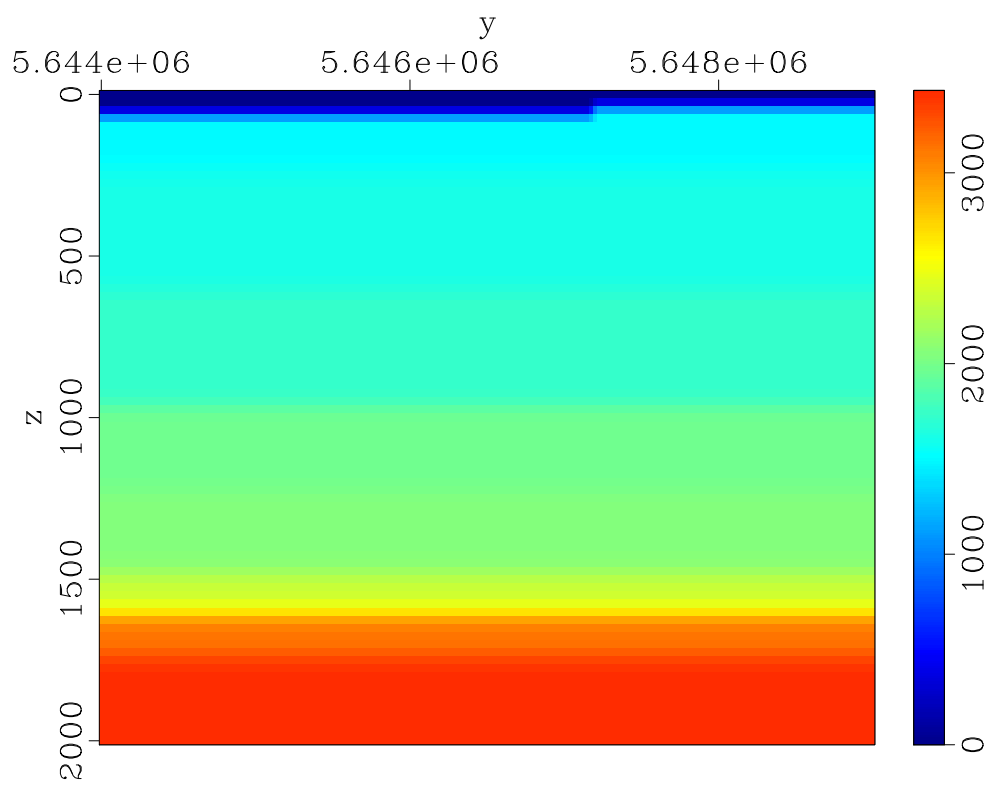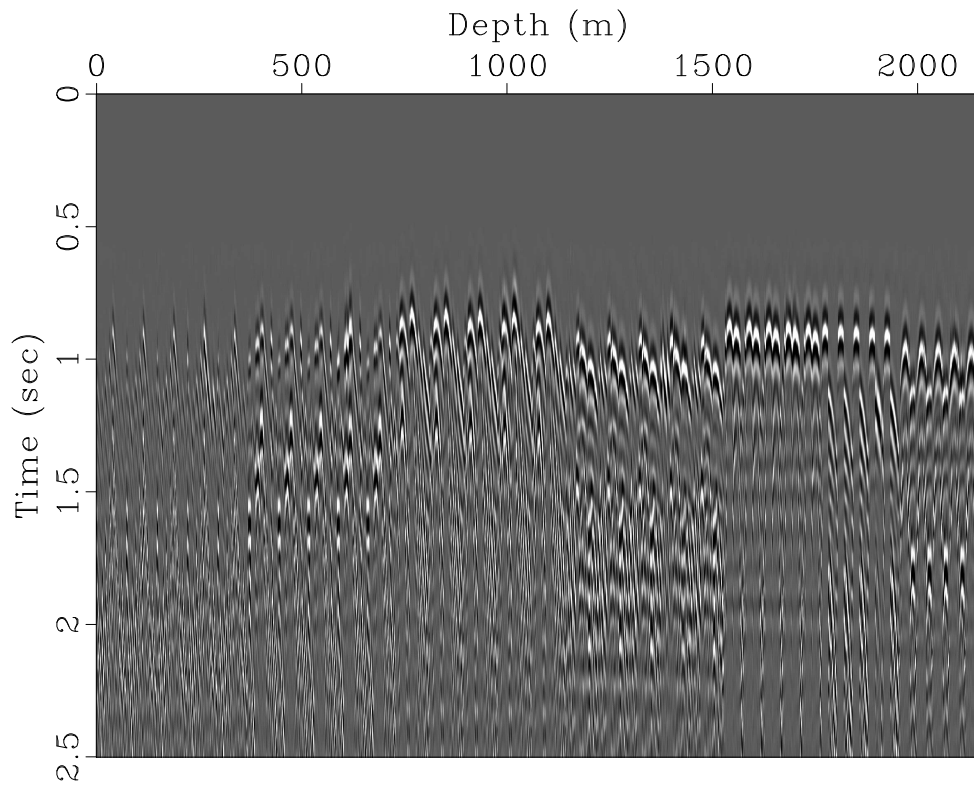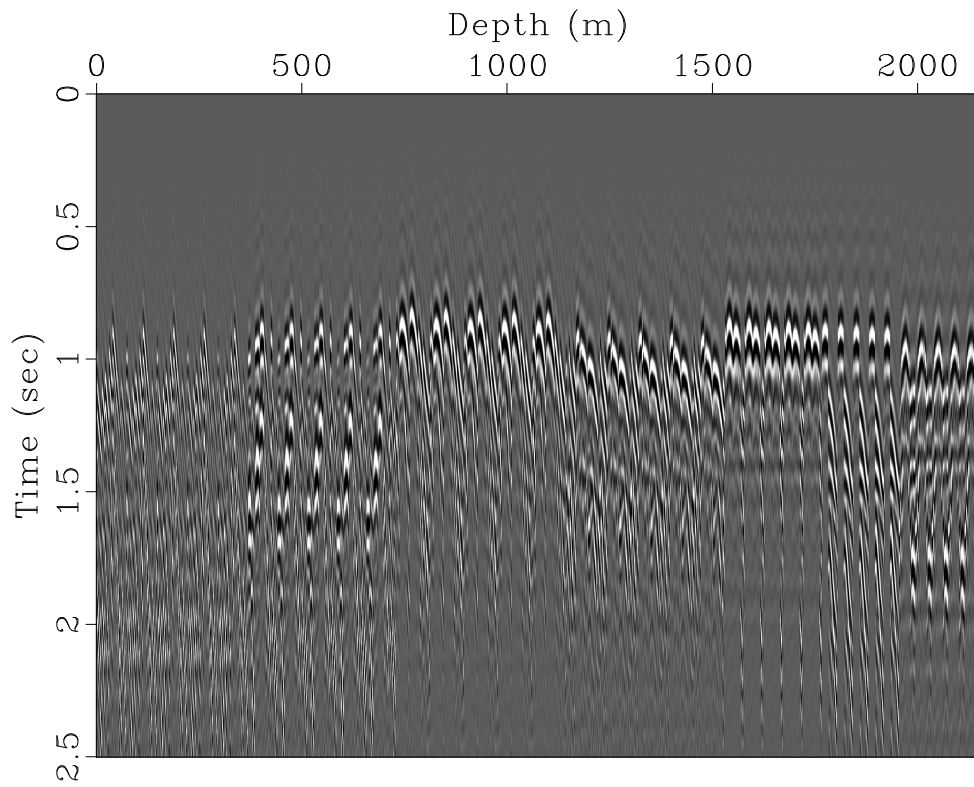
Figure 4.3: 1D linear models a)P-wave velocity b)S-wave velocity

Figure 4.4: Comparison between vertical component: a)Seismic data and b)Computed synthetics
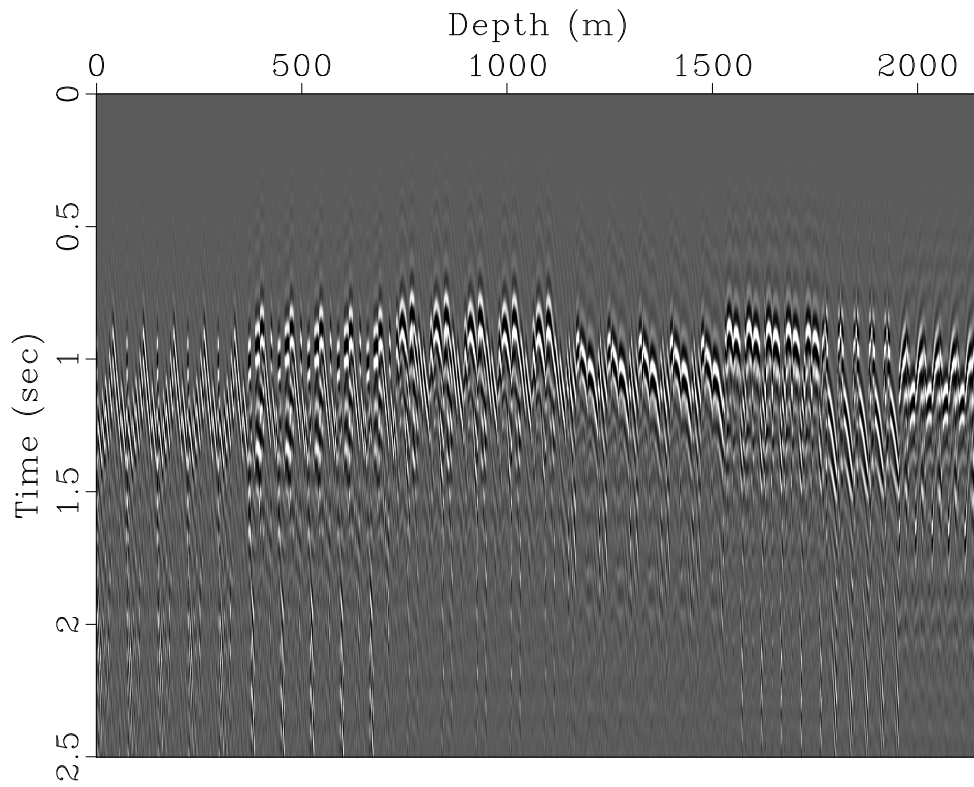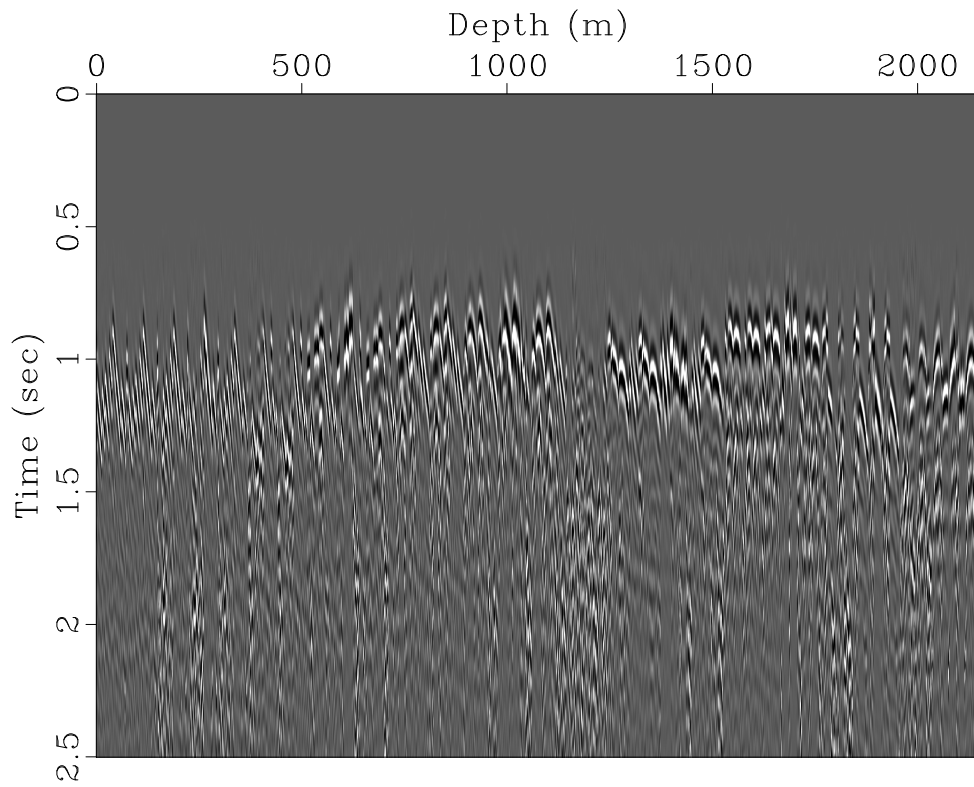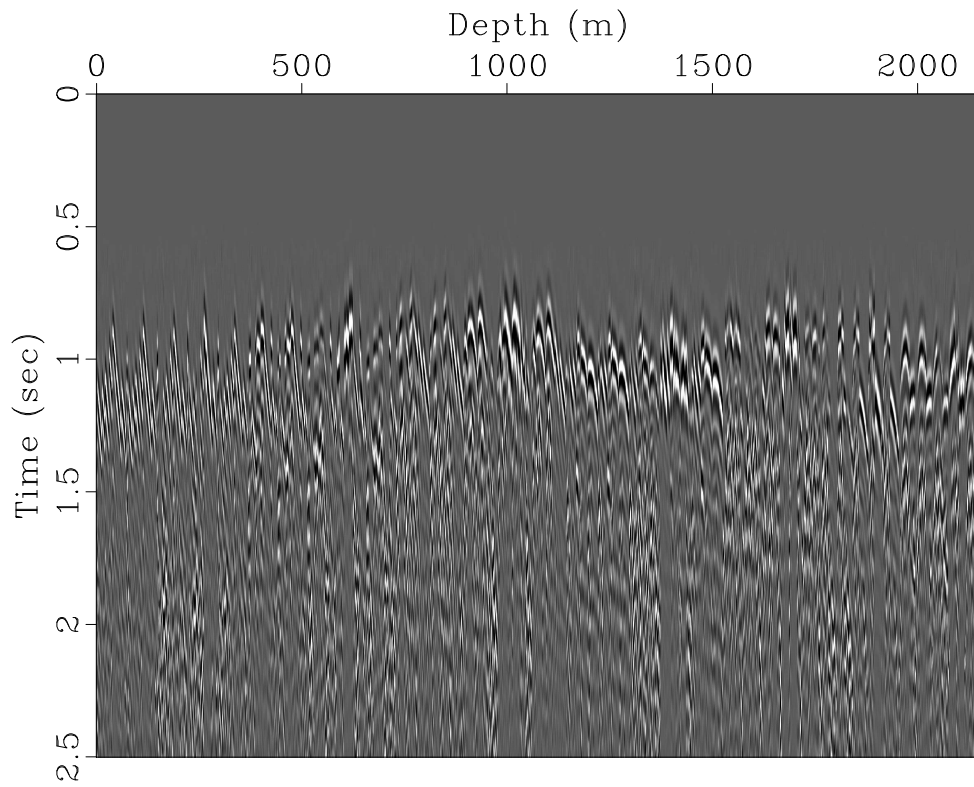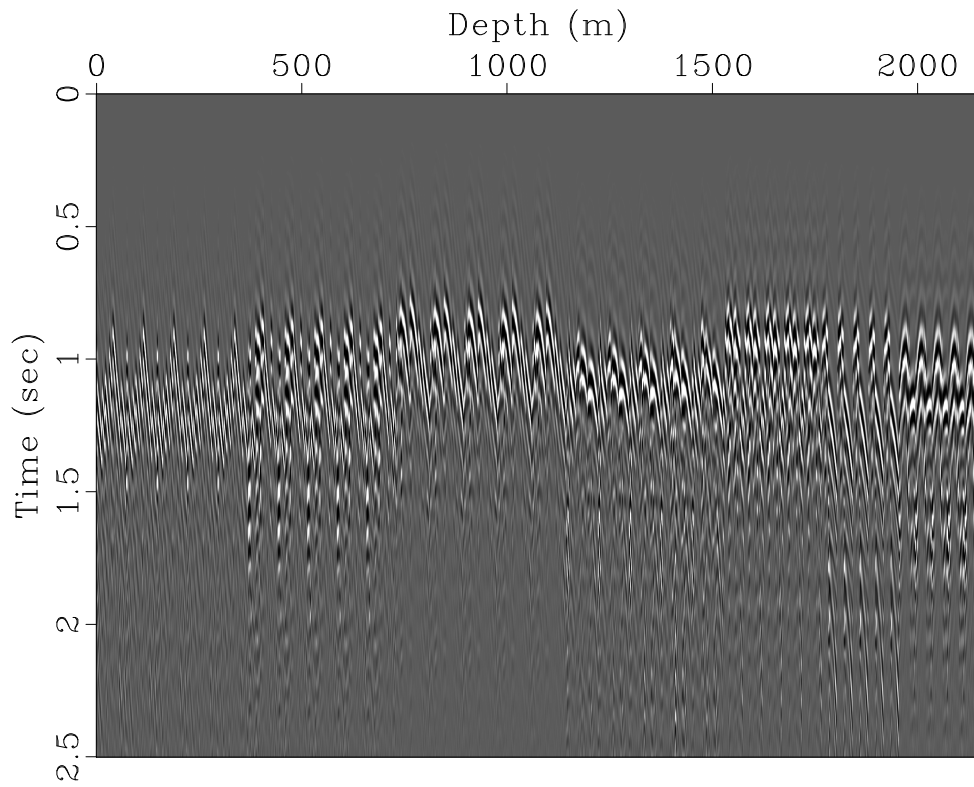
32

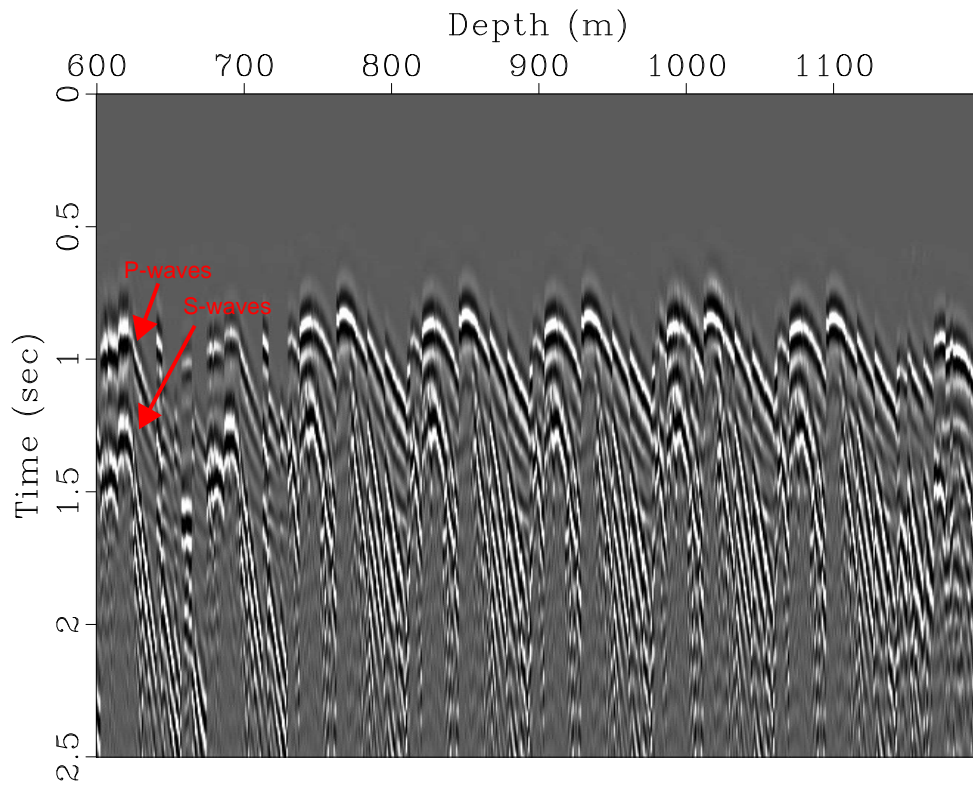**Figure 4.5:** Comparison between x-horizontal component (Hx): a)Seismic data and b)Computed synthetics

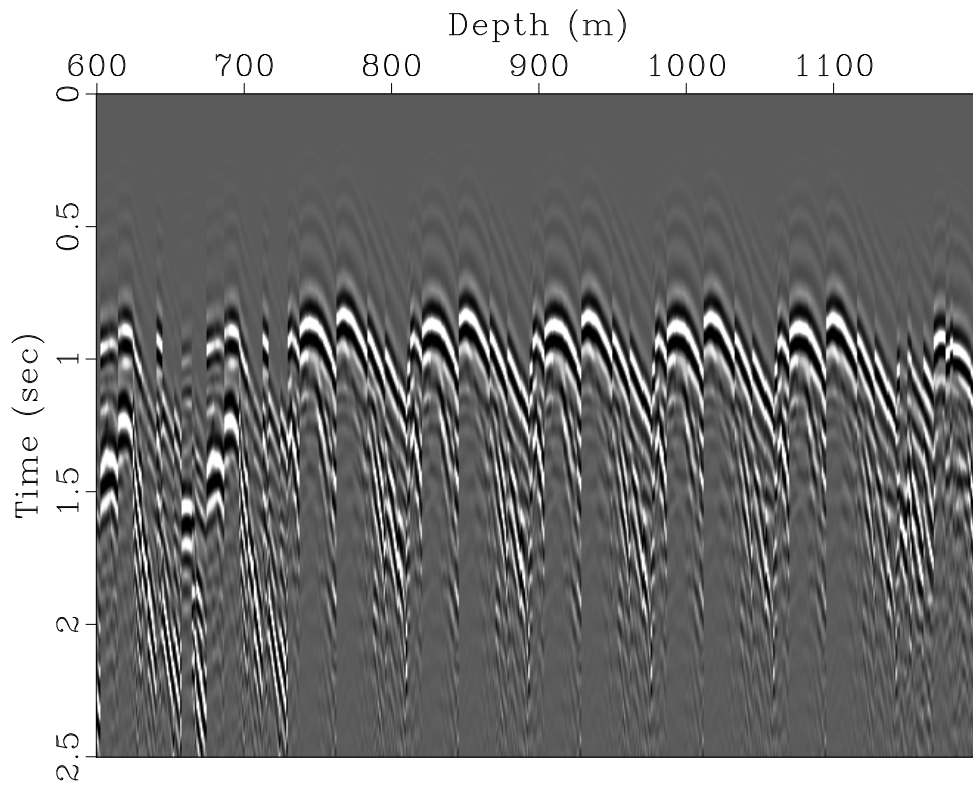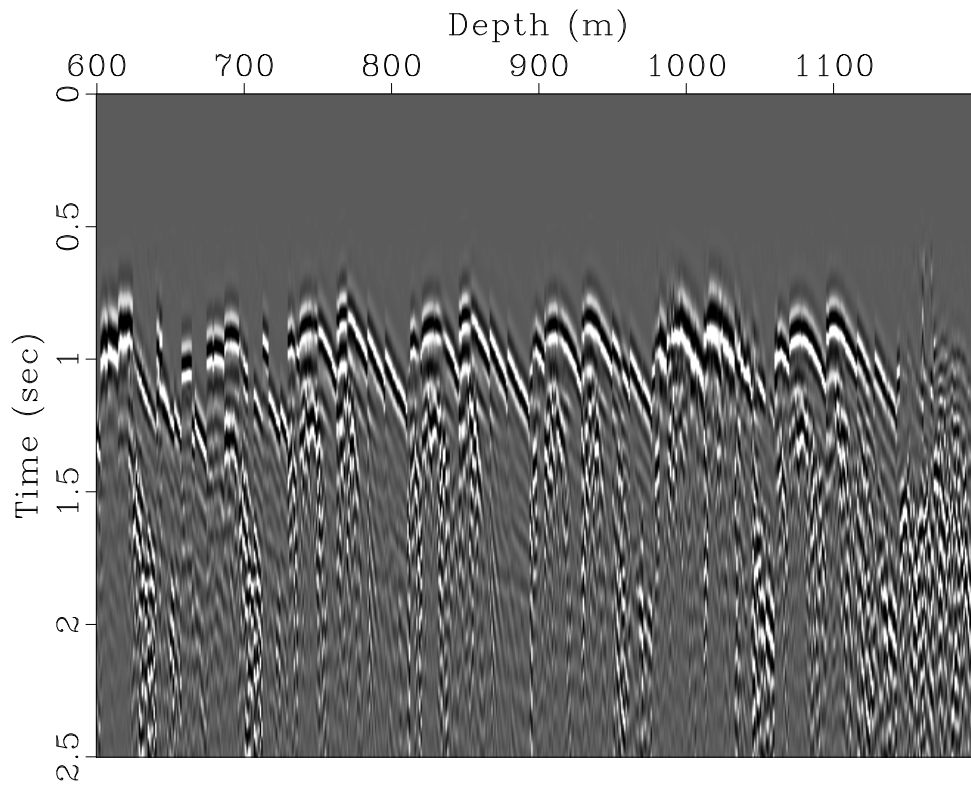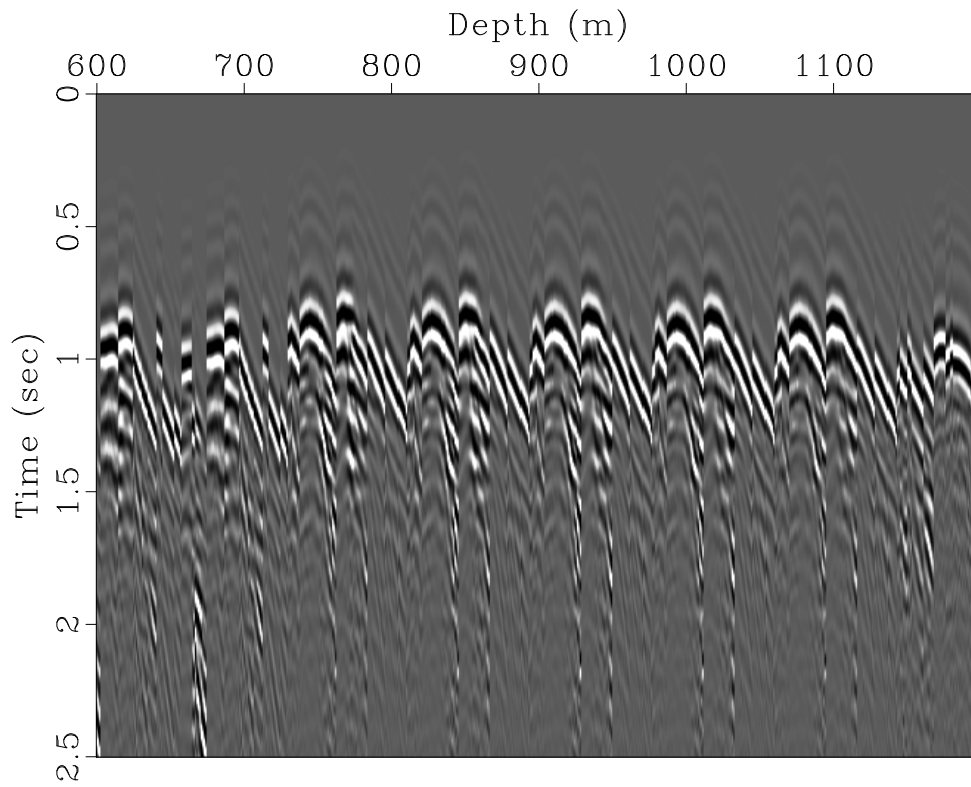**Figure 4.6:** Comparison between y-horizontal component (Hy): a)Seismic data and b)Computed synthetics

34

Figure 4.7: Subset of vertical component:a)Seismic data and b)Computed synthetics for depth between 600m and 1200m, Note the recording of P-waves and S-waves

35

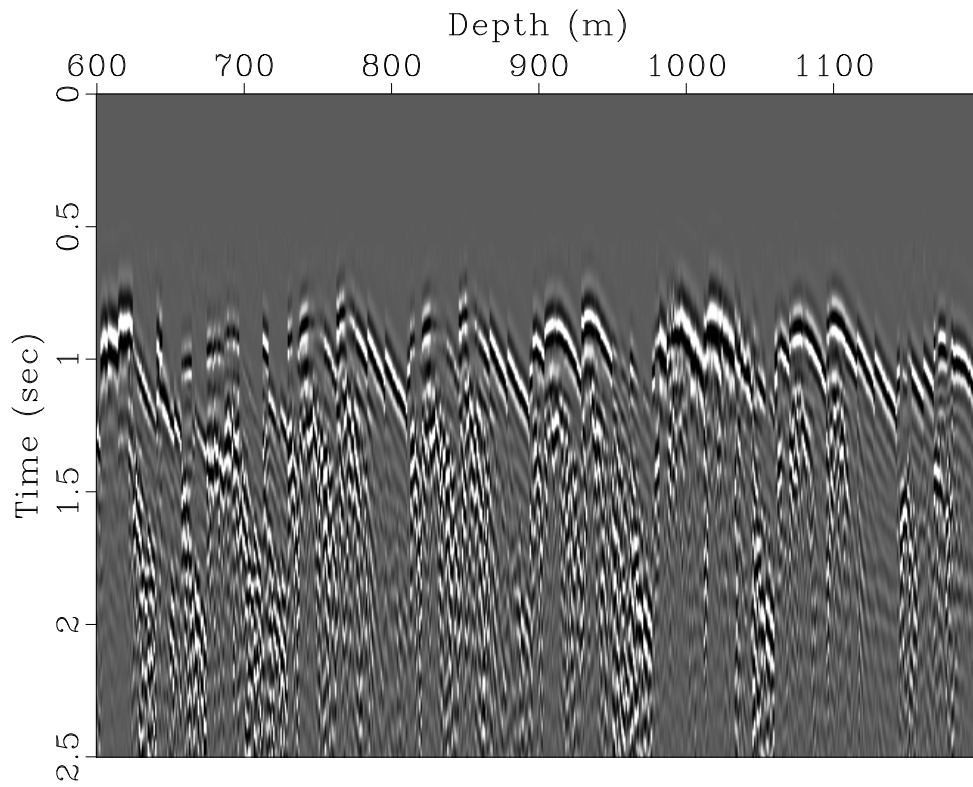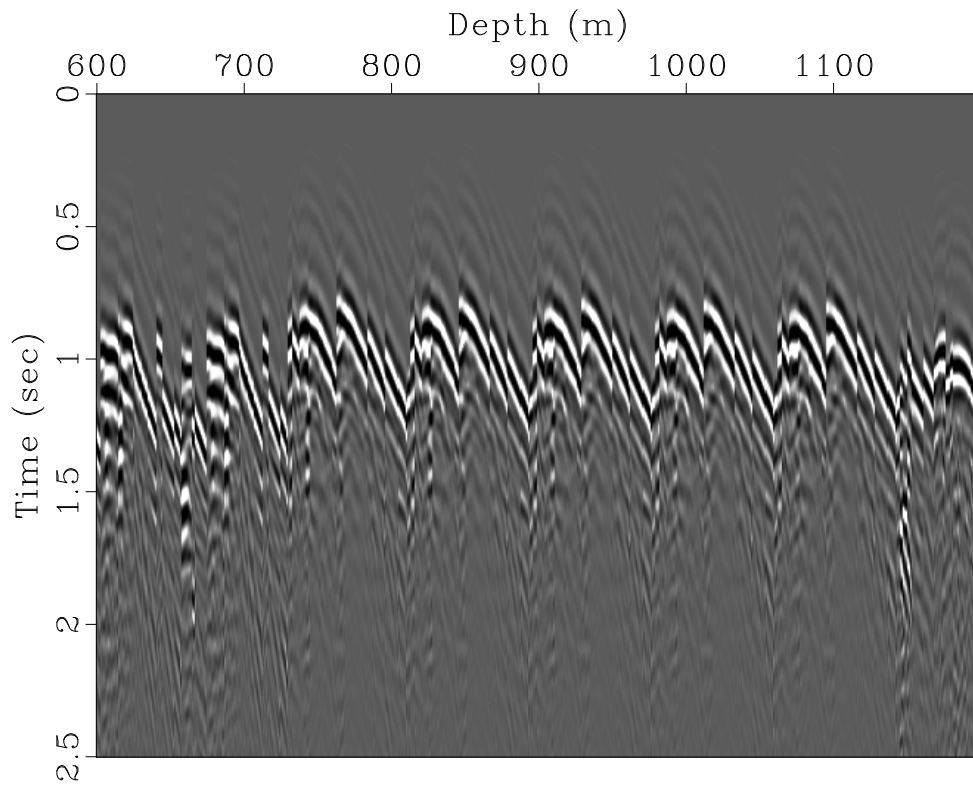**Figure 4.8:** Subset of horizontal (Hx) component: a)Seismic data and b)Computed synthetics for depth between 600m and 1200m

Figure 4.9: Subset of horizontal (Hy) component: a)Seismic data and b)Computed synthetics for depth between 600m and 1200m

37

The data waveform residuals are shown in figures 4.10, 4.11, 4.12 after 5 iterations for each multiscale step. It shows that vertical and x-horizontal component(Hx) residuals were reduced compared to the y-horizontal component(Hy).
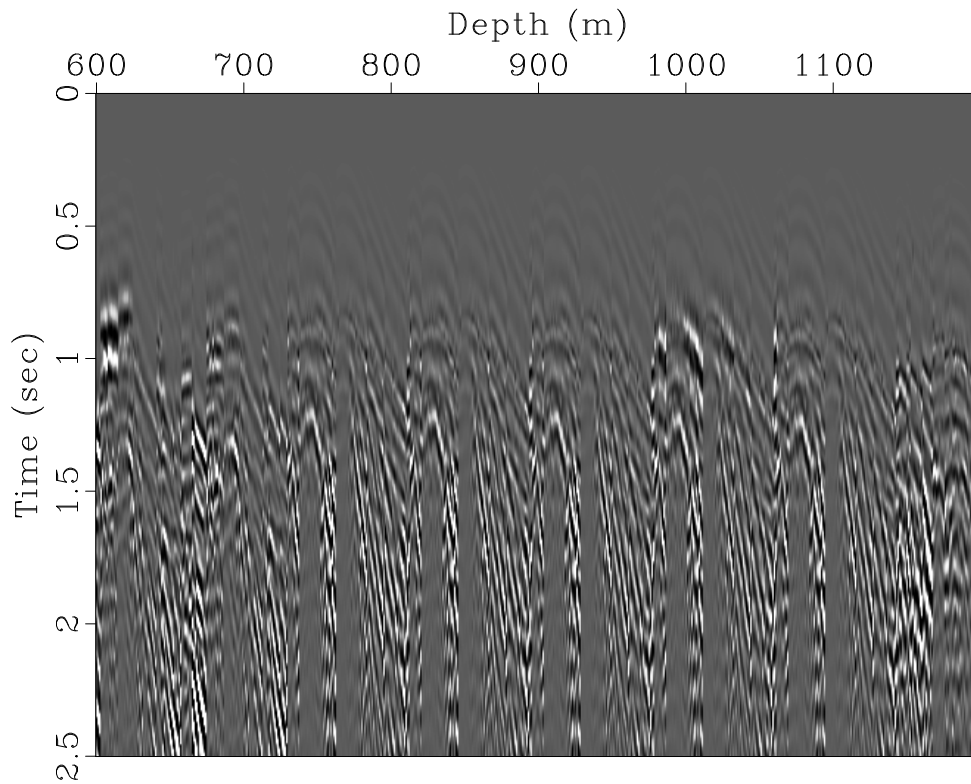

Figure 4.10: Seismic data residual for the Vertical component

Figures 4.13 shows the updated P-wave and S-wave velocity models. The velocity updates were limited to the location of the PCP Blackfoot 12-16-23-23 well. To further understand more details about the velocity updates, the initial velocity models were subtracted from updated models. Figure 4.14 shows the accumulations of updates over 5 iterations for each step in the multiscale approach. Also, It shows strong positive update for Vp and Vs which might be attributed to the cement casing of the well that causes exaggeration is P-wave and S-wave velocities. Figure 4.15 shows a depth slice at 900m of the perturbations accumulation. 1D velocity model was extracted from the updated model at the well location to be compared with sonic log, shown in Figure 4.16, and it shows very minor change in S-wave velocity from the initial model at the well depth interval.

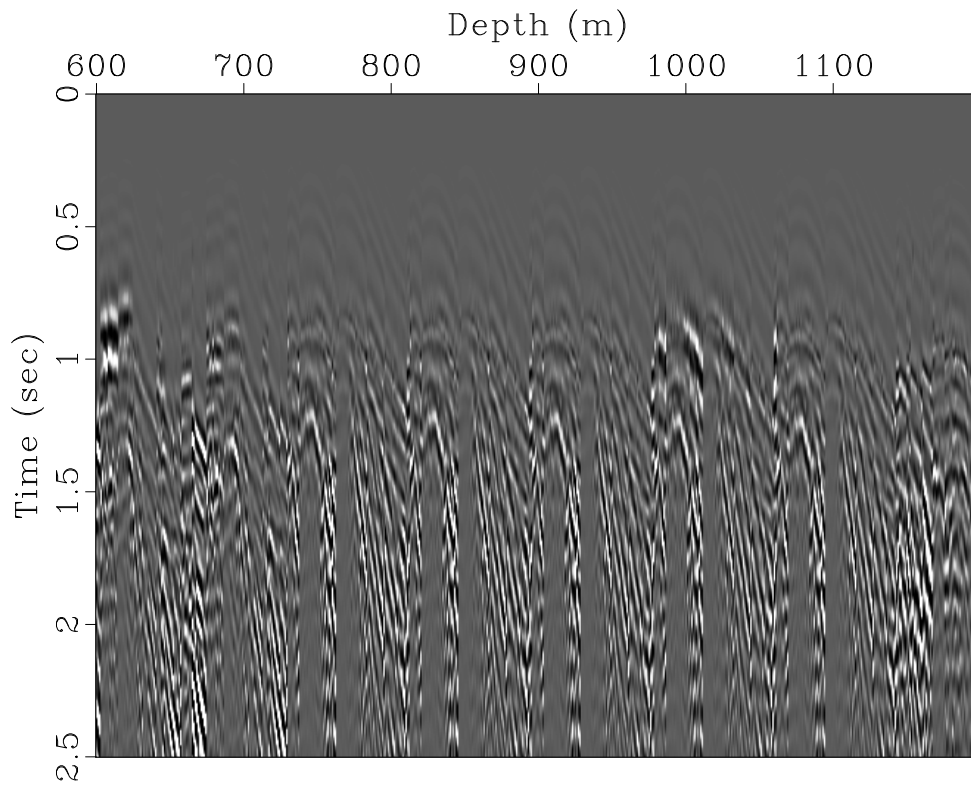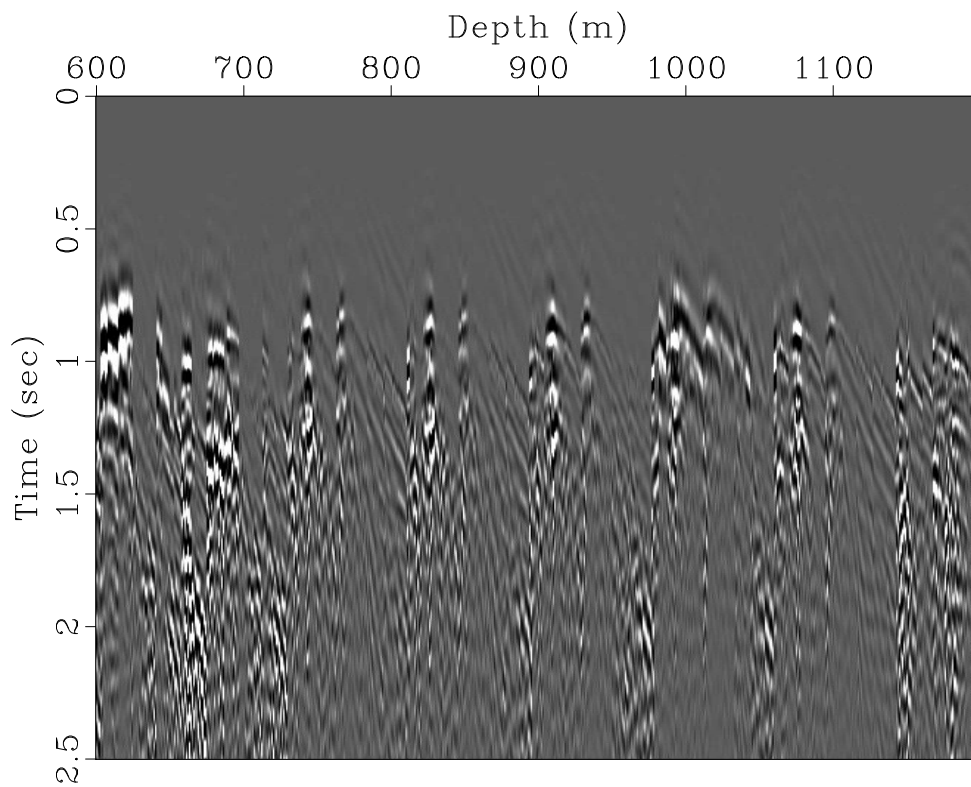**Figure 4.11:** Seismic data residual for the x-horizontal(Hx) component

Figure 4.12: Seismic data residual for the y-horizontal(Hy) component
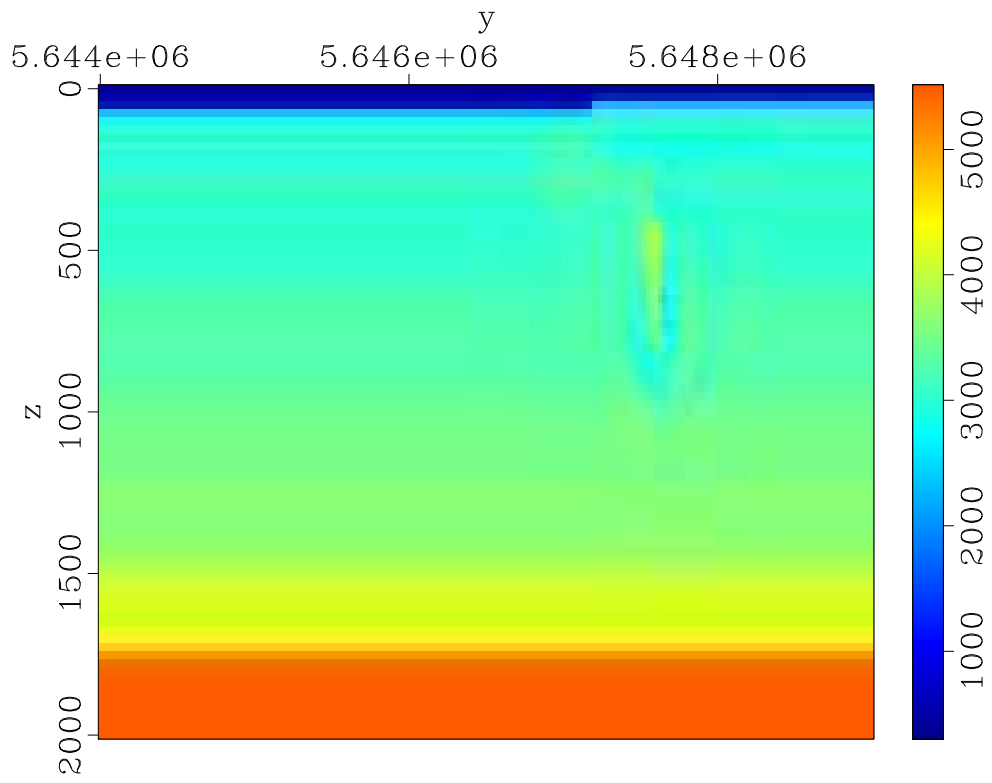
(a)
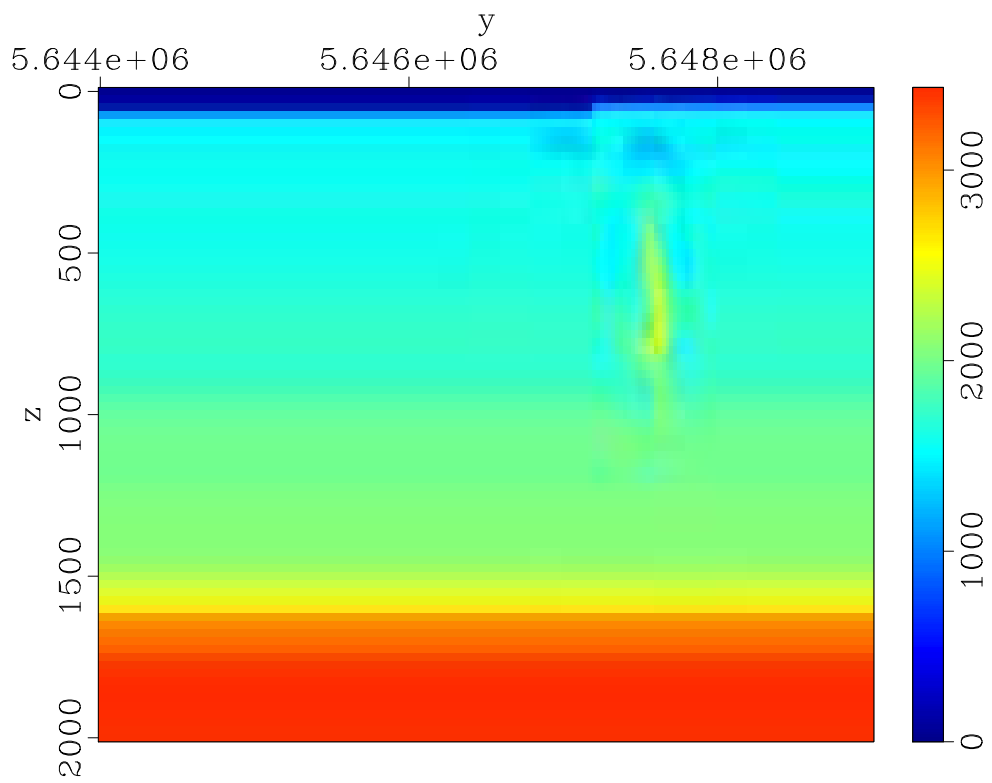
(b)

Figure 4.13: Updated velocity models a)P-wave velocity and b)S-wave velocity
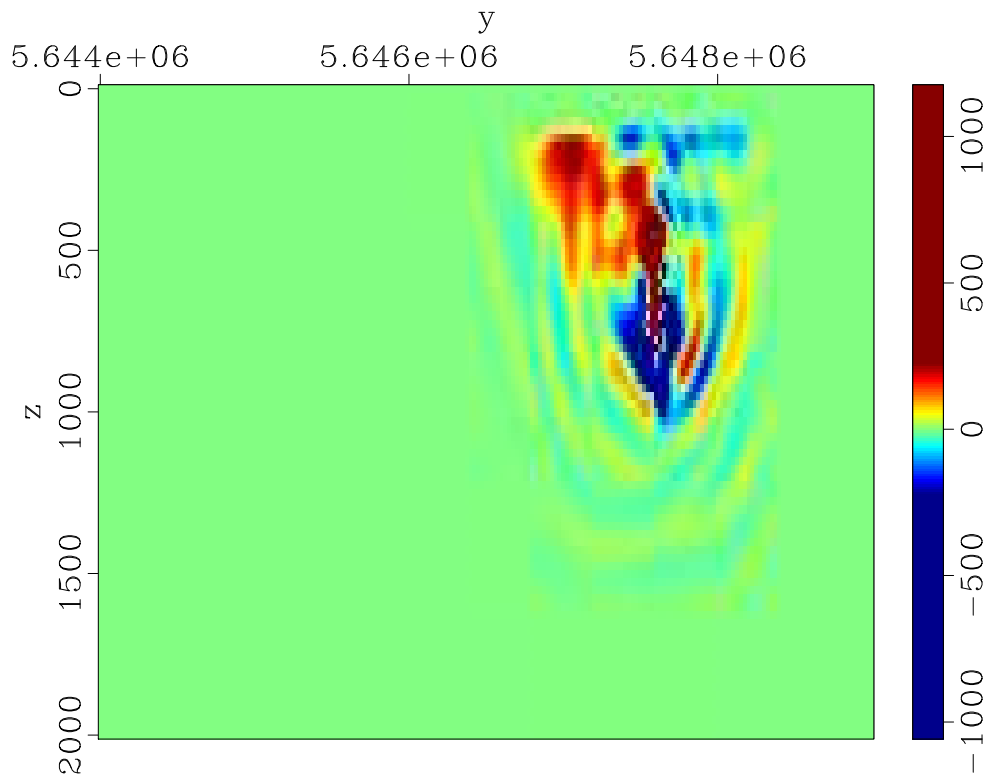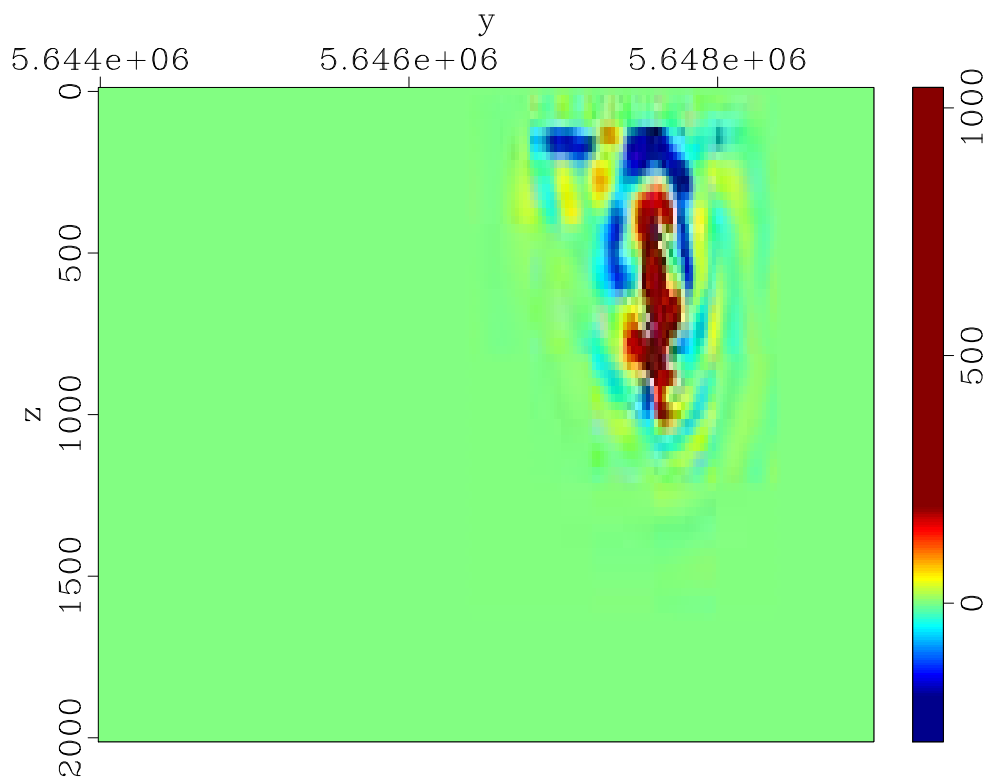
Figure 4.14: Accumulated Velocity perturbations a)Vp and b)Vs
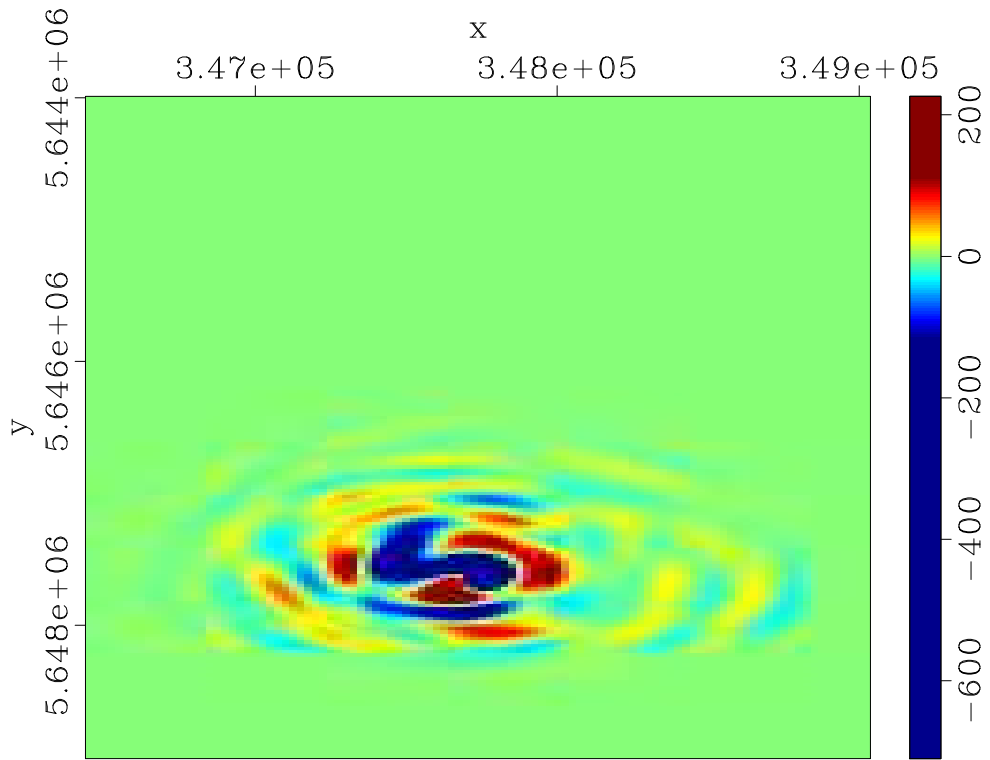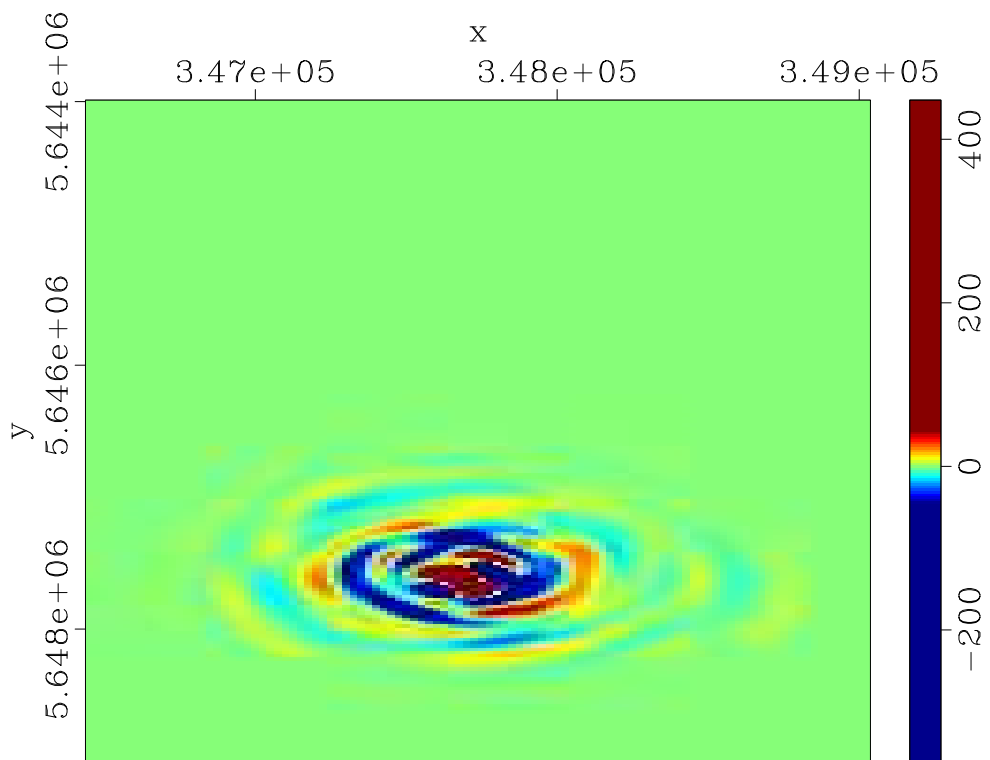
Figure 4.15: Depth slice at 900m of Accumulation of velocity updates a)Vp and b)Vs
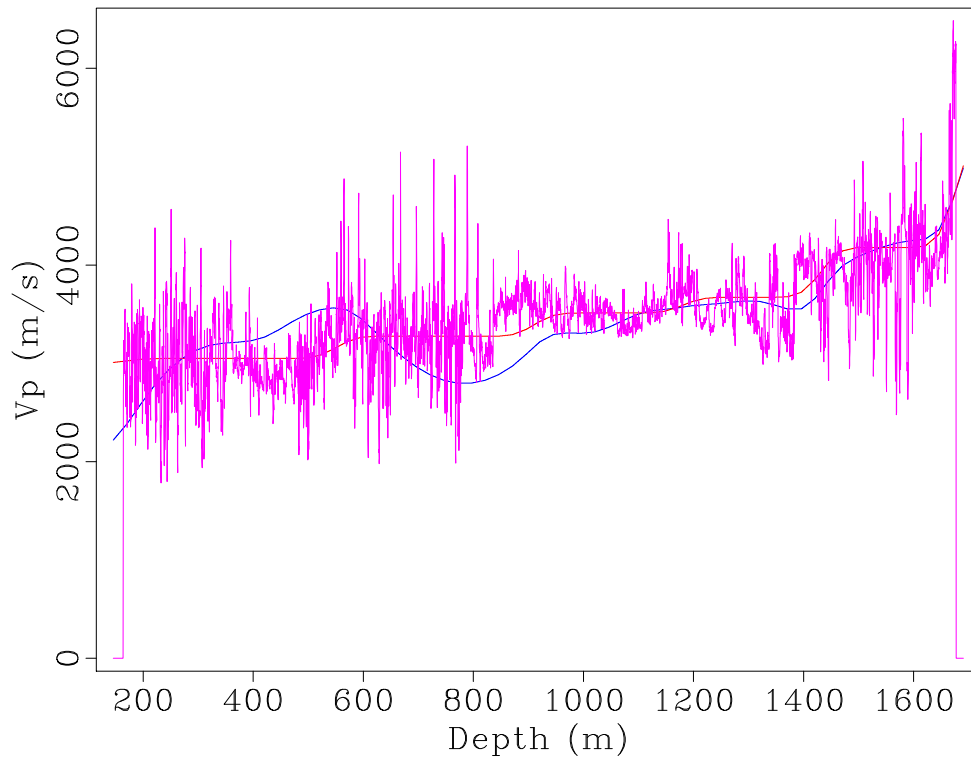
(a)
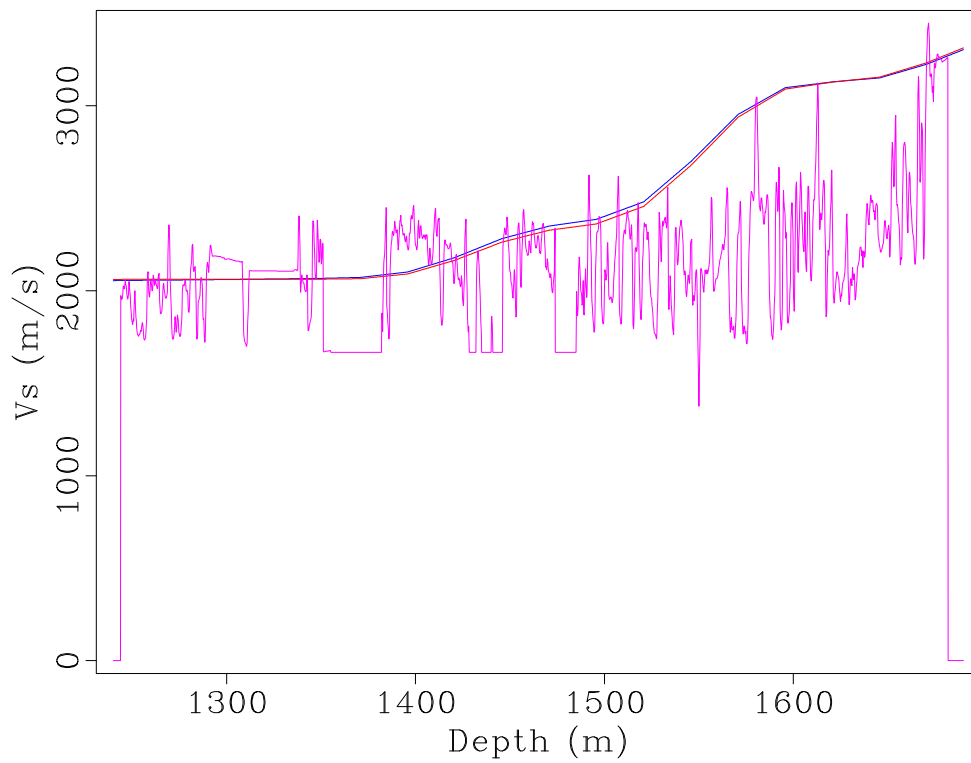
(b)

Figure 4.16: Comparison between updated velocity (blue), initial velocity (red) and sonic log velocity (purple) a)Vp, and b)Vs. Notice that the sonic logs were recorded at limited depths in the well (PCP Blackfoot 12-16-23-23)

43

5 Case Study: Surface Seismic Data

5.1 Results

5.1.1 Elastic FWI

The second major step in this research is applying the FWI on seismic surface data using the 1D linear model as initial model (Figures 4.3) and using the updated VSP model as initial model (Figure 5.1). Since the VSP updated models show matching with the sonic logs at the well location and better constraints in the vertical variations in velocities, the model updates were windowed to the well location and extended back into the 2D domain prior to FWI. The methodology was applied on extracted North-South section from the 3D data (60 shot gathers). A subset of the 2D section (7 shot gathers) for the 2D North-South section are shown in figure 5.2. The used wavelet is the same wavelet used for inverting VSP events, shown in Figure 4.1, since the VSP and surface seismic data acquisition shared the same source wavelet.

Different strategies were applied on the data to estimate the best results. Using Low Pass filter of 10 hrz and 7.5 depends on the how fast the convergence is at each multiscale step. Also, changing the initial model sampling from fine (25m) to coarse (50m) depends on the speed of convergence at each multiscale step. However, the results shown here is by using A low pass filter with a high cut of 7.5 Hz on the data and the wavelet to enable a faster convergence.

Also, muting of very shallow FWI updates were applied helped to mitigate the inaccuracy in modeling of the free surface and topography. The velocities were gradually increases according to topography as shown in Figures 4.3 and 5.1.

Different optimization algorithms, i.e. L-BFGS, steepest descent and conjugate gradient, were used here using line search based on Wolfe and Armijo conditions. Reducing the objective function was very challenging task in this data. L-BFGS using line search based on Wolfe condition tends to make the convergence faster that results in reducing the misfit in a way faster than the steepest descent and conjugate gradient. Five iterations were used for each multiscale-step by increasing the depth of 100m at each step.
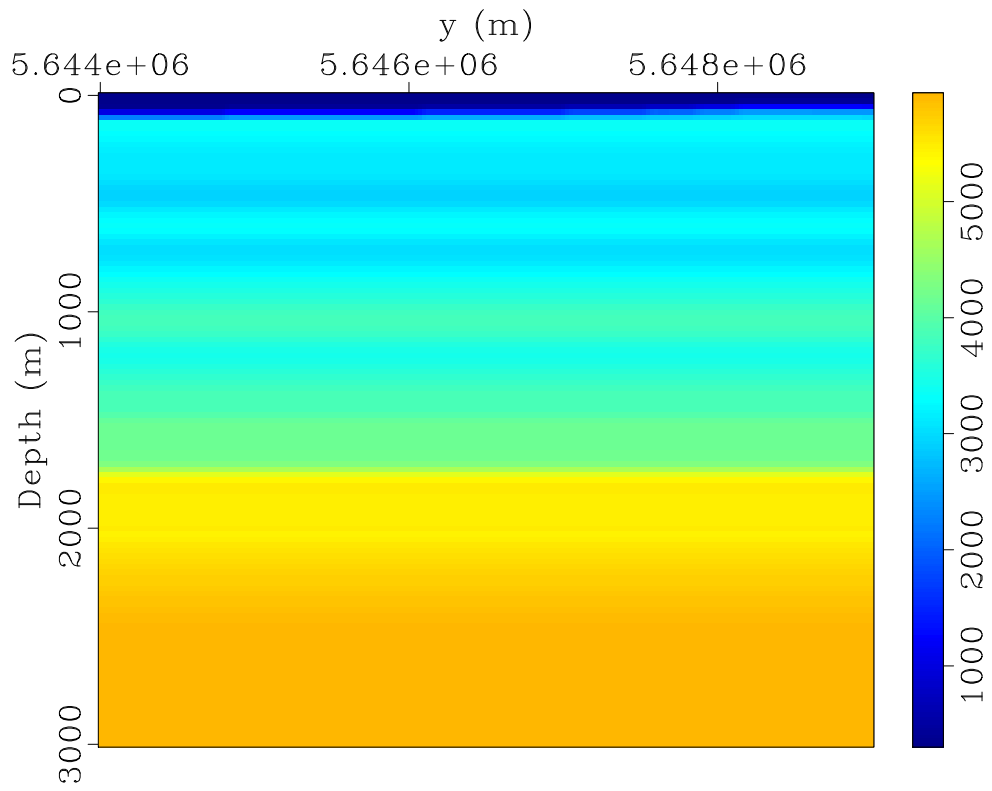
(a)
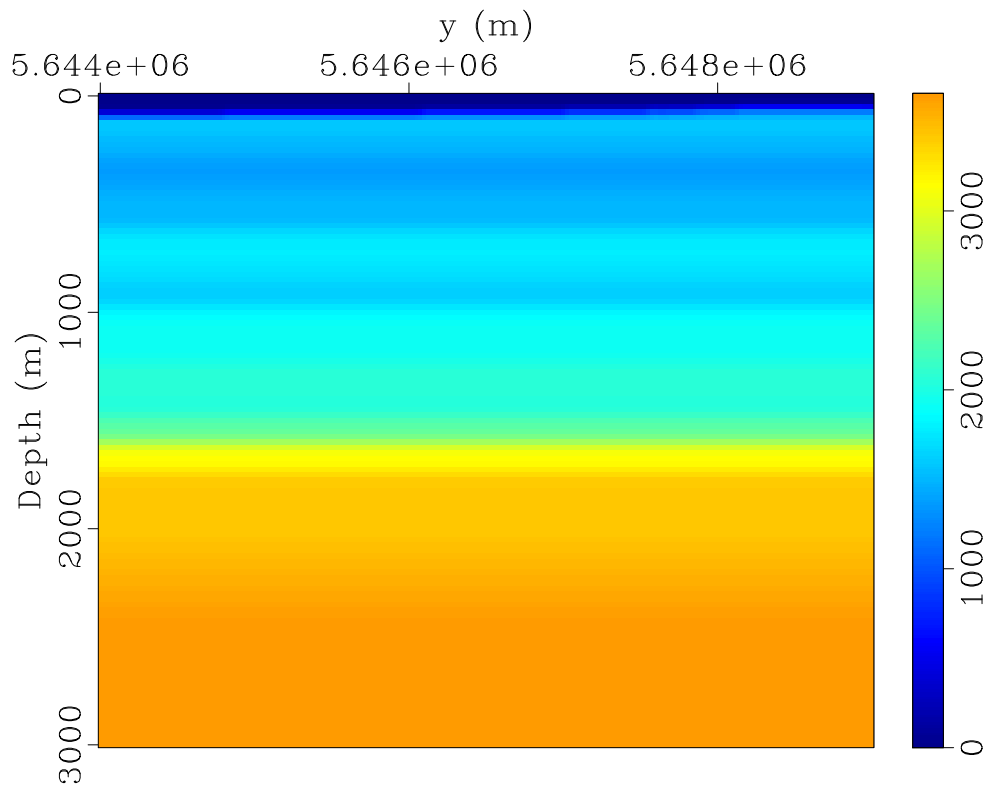
(b)

Figure 5.1: Initial models from VSP updated models used for inverting surface seismic data:a)Vp and c)Vs

(a)

(b)

Figure 5.2: Subset (7 shot gathers) of the surface recorded seismic data a)Vertical component and b)Radial component

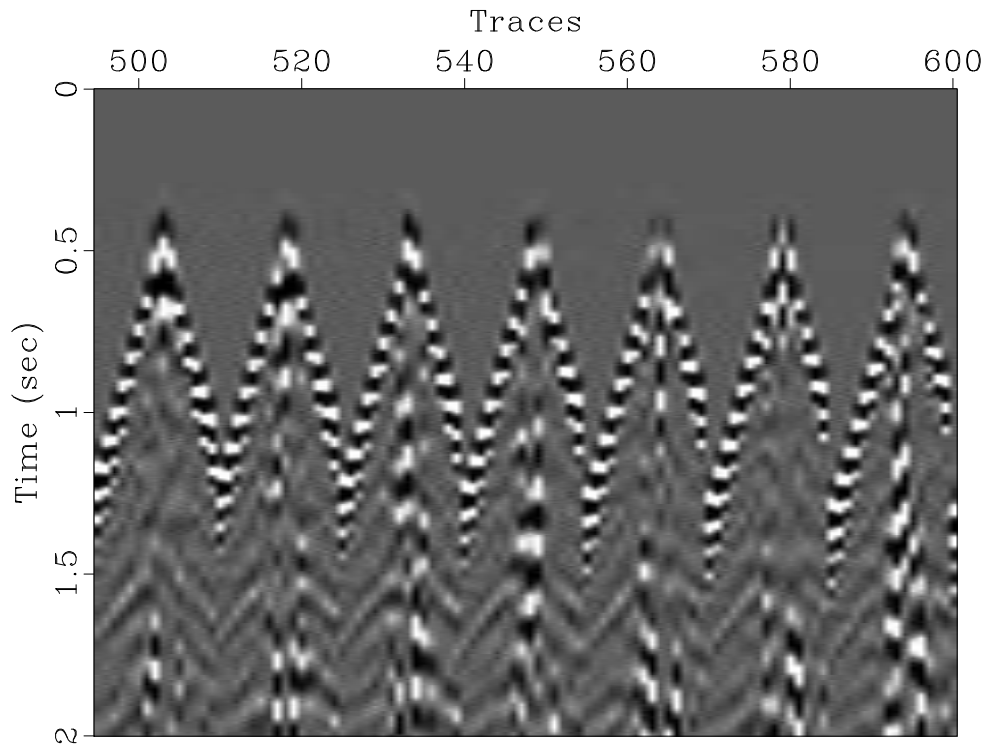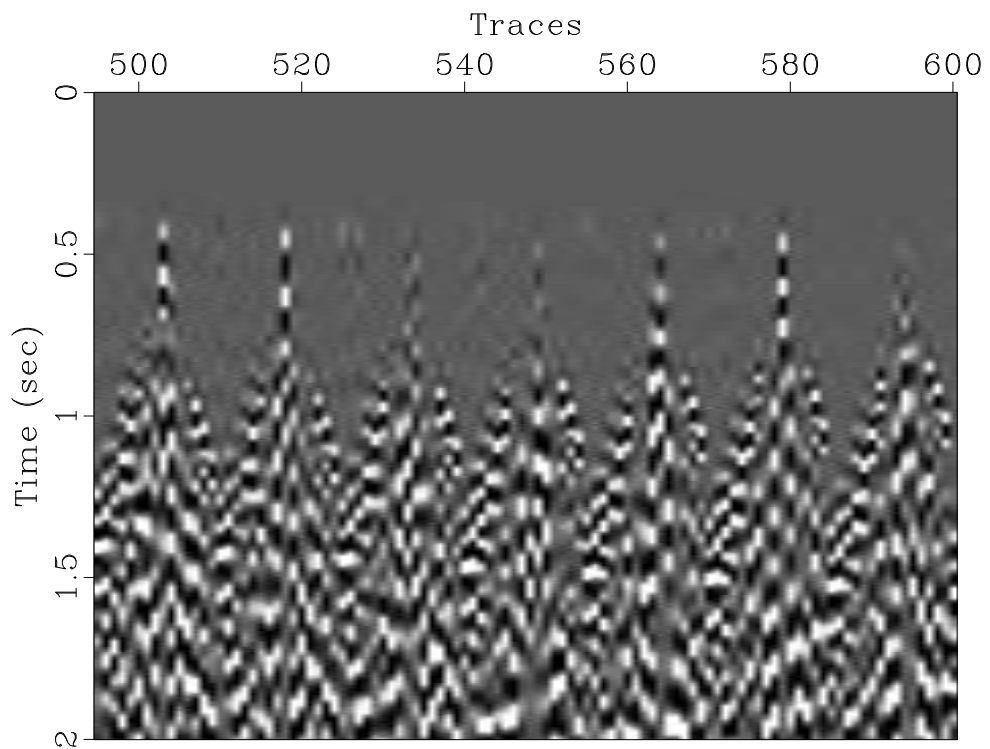Figure 5.3 shows the computed synthetics using the 1D linear velocity model as initial model while Figure 5.4 show the synthetics using the VSP updated model. For more details, we took the far left shot gather in figure 5.2, 5.3, 5.4 to make the comparison. The shot gather shown in figure 5.5 is the recorded data while the shot gathers are shown in figures 5.6 and 5.7 show the synthetics. Data residuals are shown in figures 5.8 and 5.9. It's clear from the figures that data residuals were reduced in the radial component using the VSP updated model as initial model compared to the residuals that used the linearly increasing velocity model as initial model. For the vertical component, the residual doesn't change much using the two different models. On the other hand, the radial component residuals show reductions which means that the resultant model is more accurate.

Figure 5.10 shows an updated velocity models using the 1D linear model as initial model for FWI using L-BFGS method for optimization. It is clear that the model was not updated enough to build the subsurface geology compared to the models in Figure 5.11. The first was just able to add perturbations just in the shallow part of the model. On the other hand, Figure 5.11 shows stronger velocity updates in deeper section. Figures 5.12, 5.13, 5.14 and 5.15 show the velocity Updates for different optimization methods to show perturbations over iterations. Accumulations of updates are calculated by the difference between the initial models and updated models. L-BFGS is shown in figure 5.12 using linear initial model it was able to update the very shallow part for P-wave velocity model and it shows relatively stronger updates for S-wave velocity model in the shallow part. Figures 5.13 and 5.14 show weaker updates for conjugate gradient and steepest descent. On the other hand, Figure 5.15 shows much stronger in the shallow and deeper sections and it shows geological layerings specially in the areas with higher coverage in the middle of the section for both P-wave and S-wave velocity. It means that the gradient optimizer was able to converge faster and estimate a more accurate model as the initial model is more close to reality.
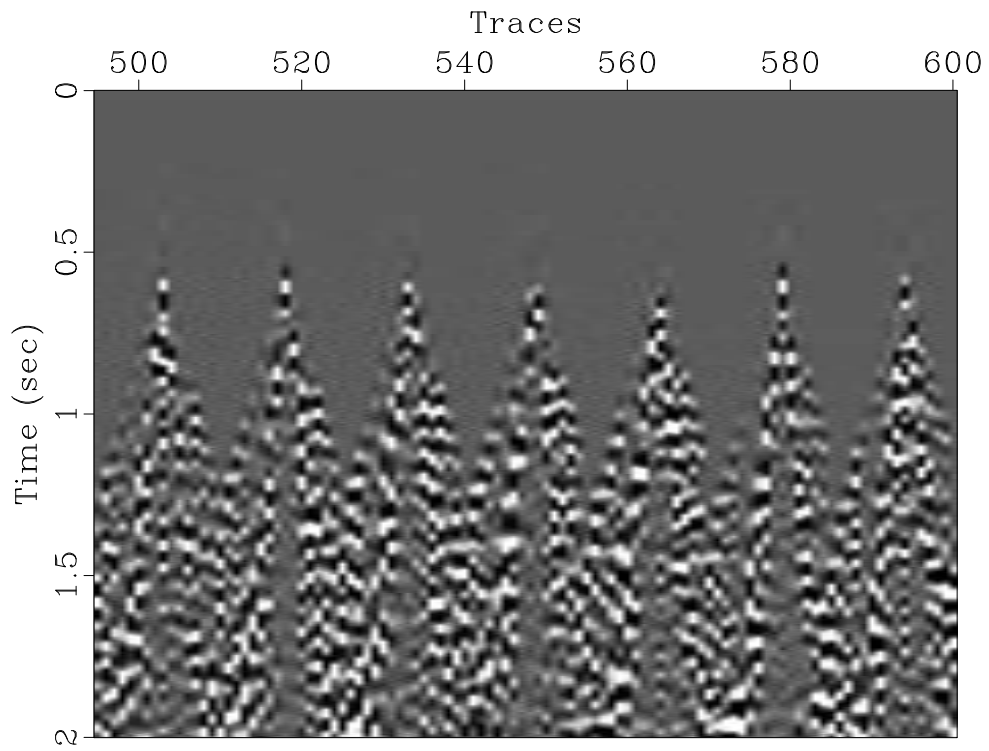
(a)
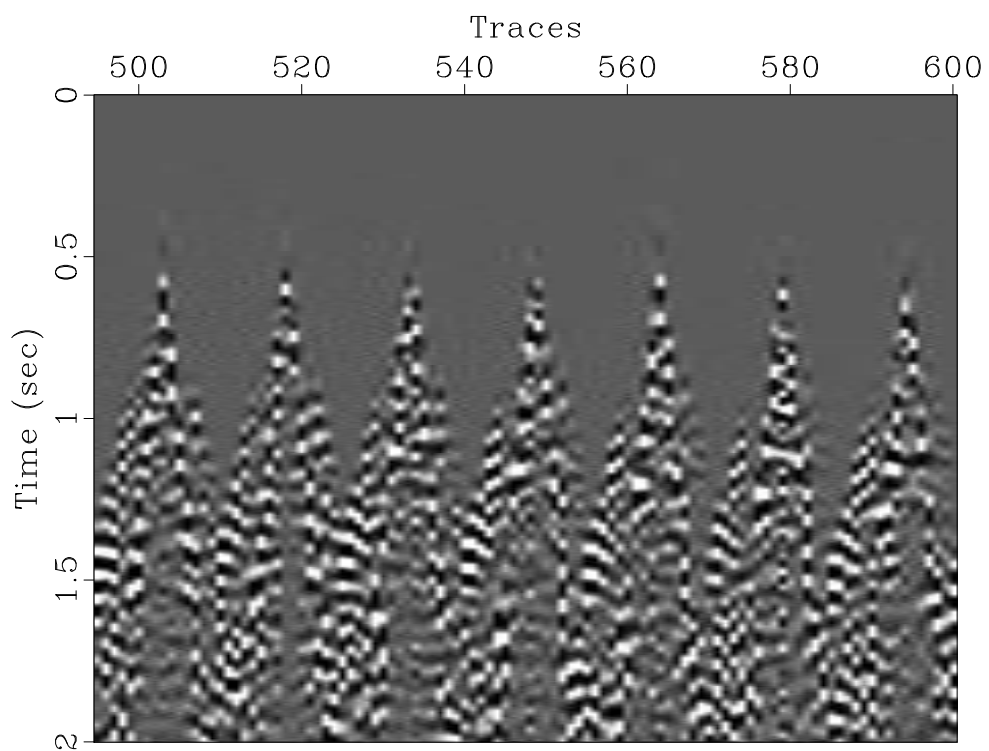
(b)

Figure 5.3: Subset (7 shot gathers) of the surface seismic synthetics using 1D lin-
early increasing velocity model as initial model a)Vertical component
and b)Radial component
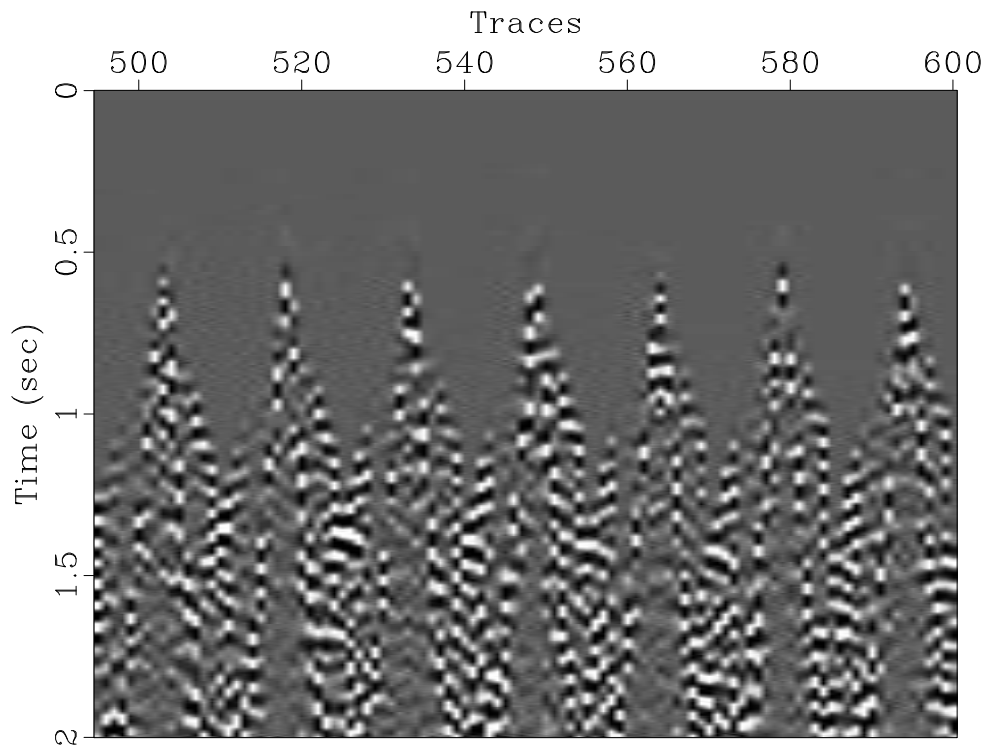
(a)
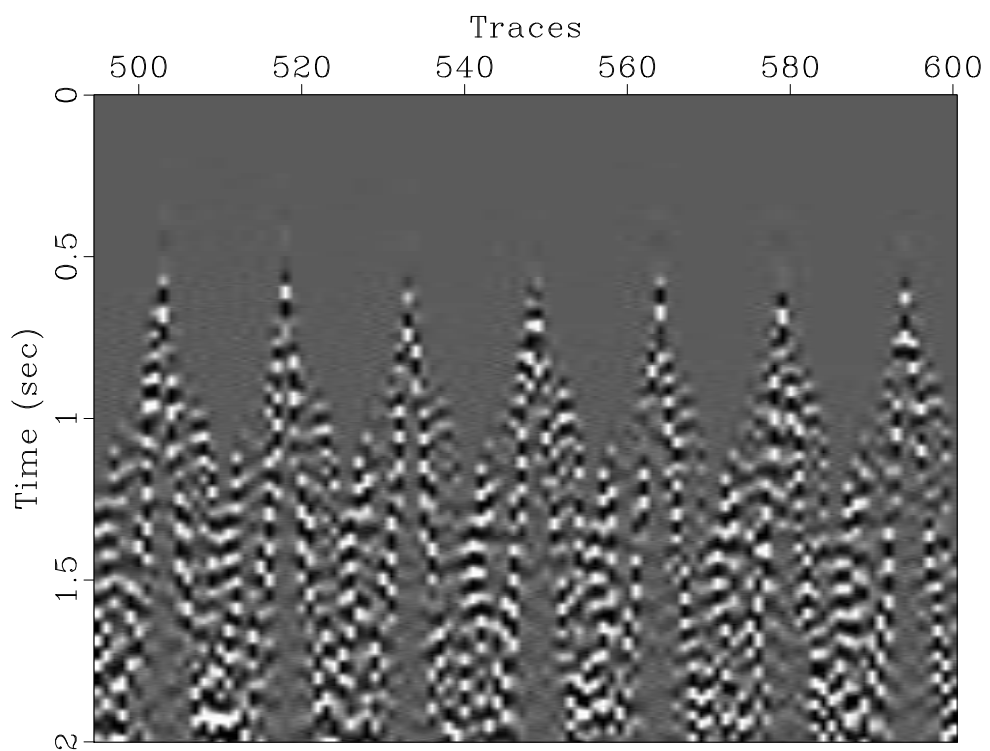
(b)

Figure 5.4: Subset (7 shot gathers) of the surface seismic synthetics using the VSP updated model as initial model a)Vertical component and b)Radial component
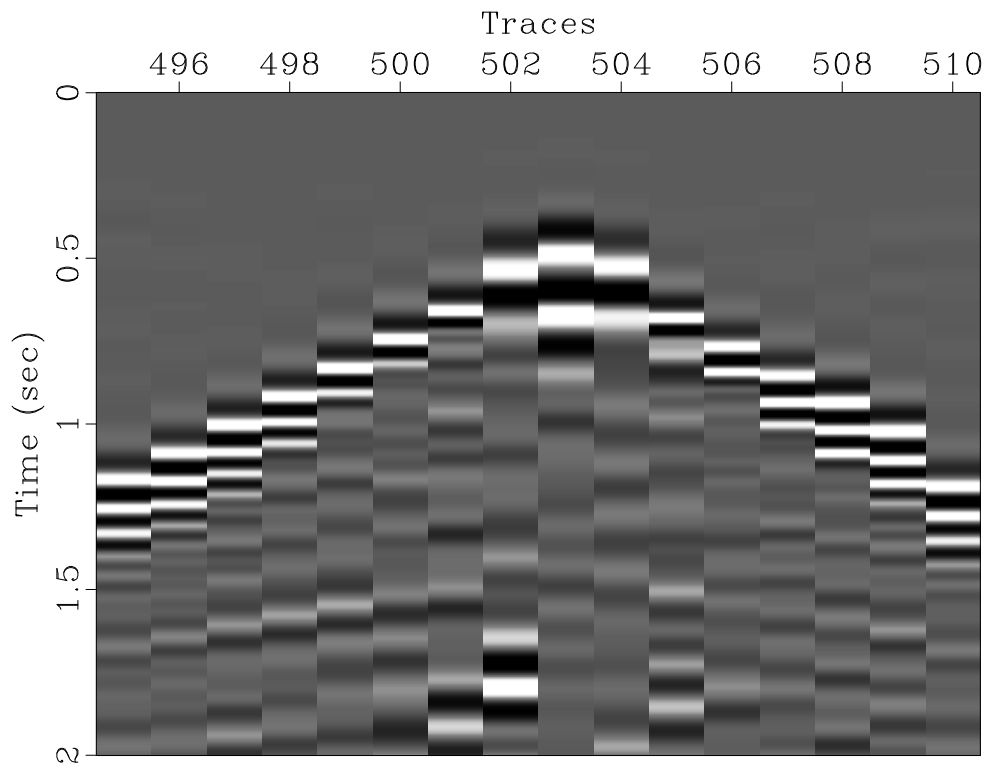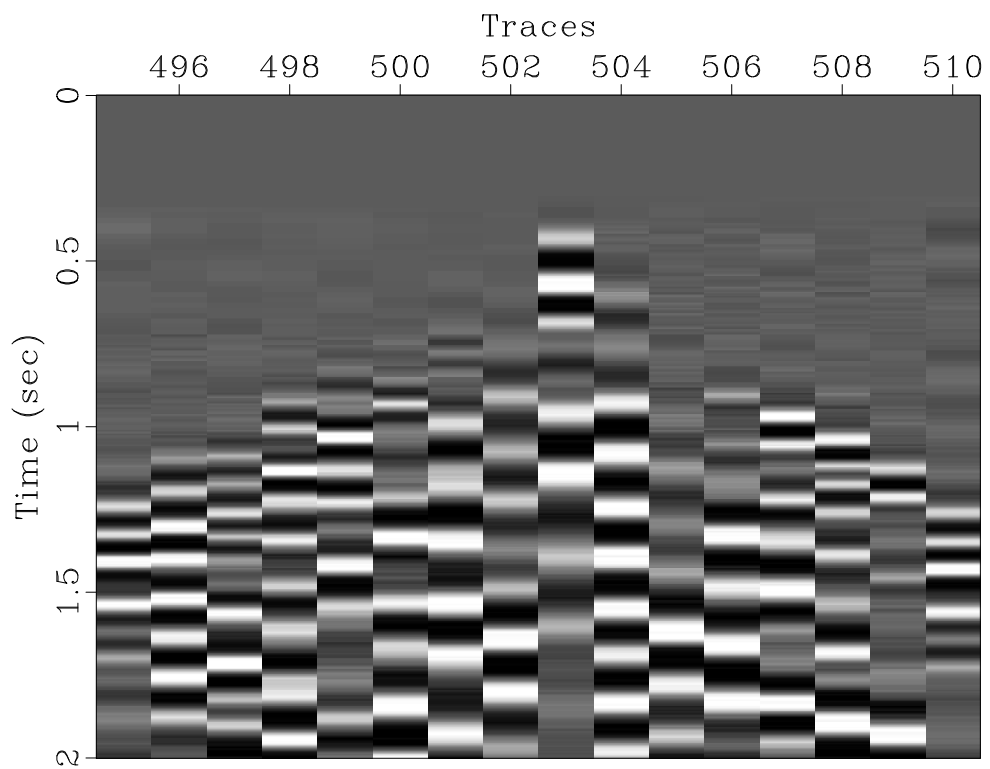
Figure 5.5: Single shot gather of surface recorded seismic data a)Vertical compo-
nent and b)Radial component
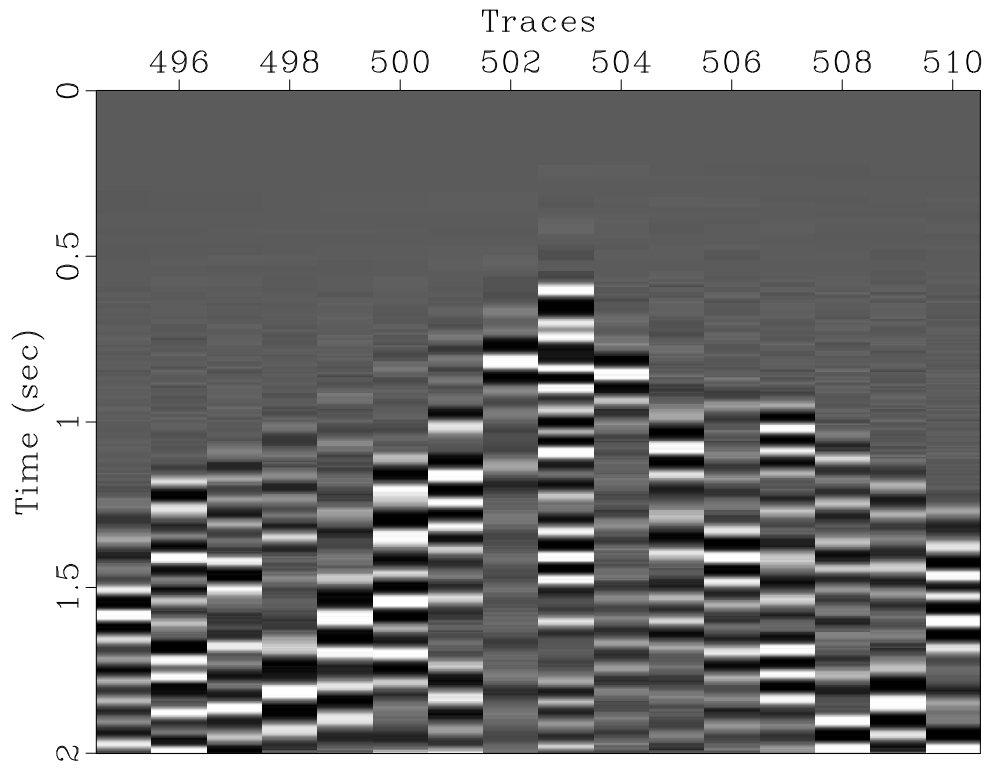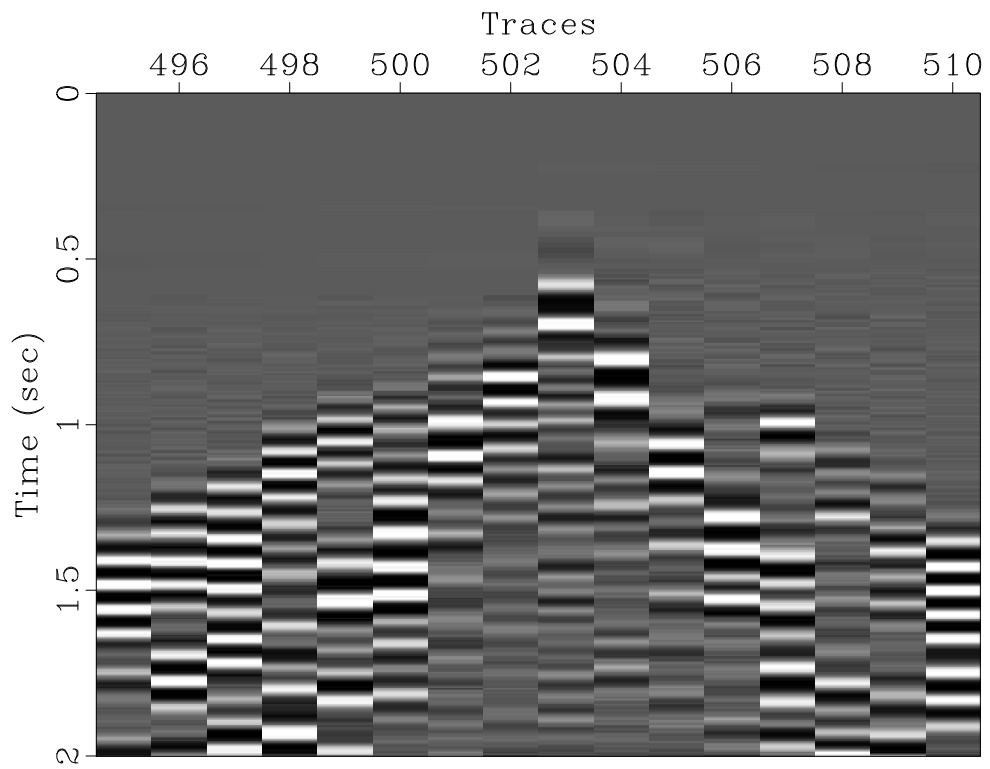
Figure 5.6: Single shot gather of surface seismic synthetics using 1D linear model as initial model a)Vertical component and b)Radial component
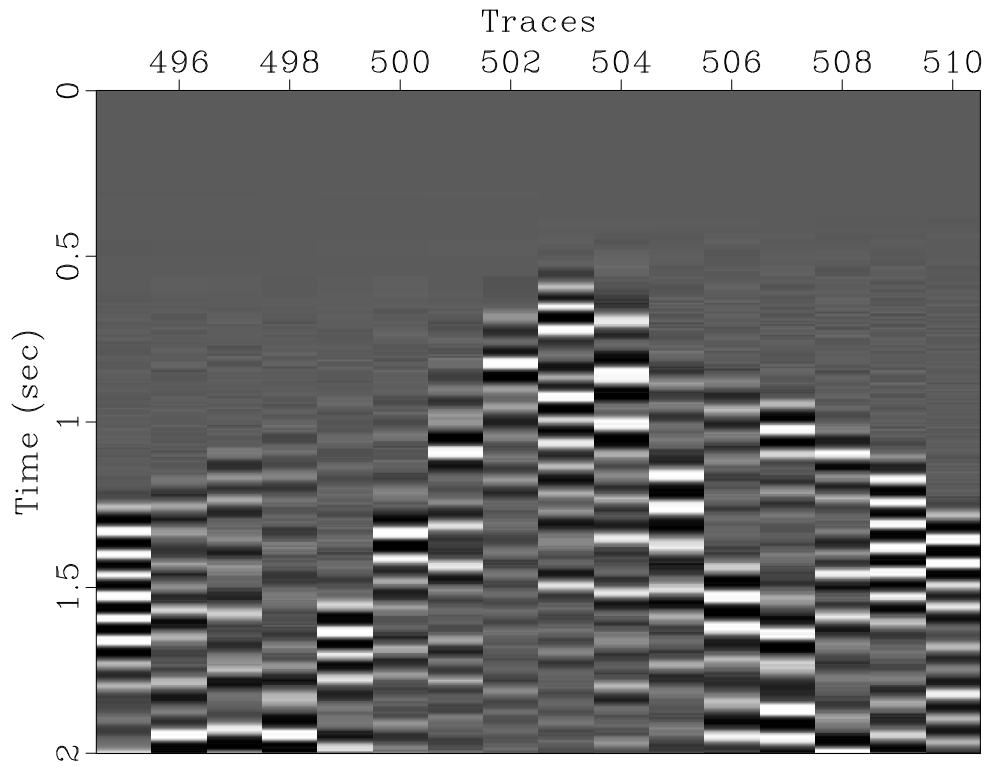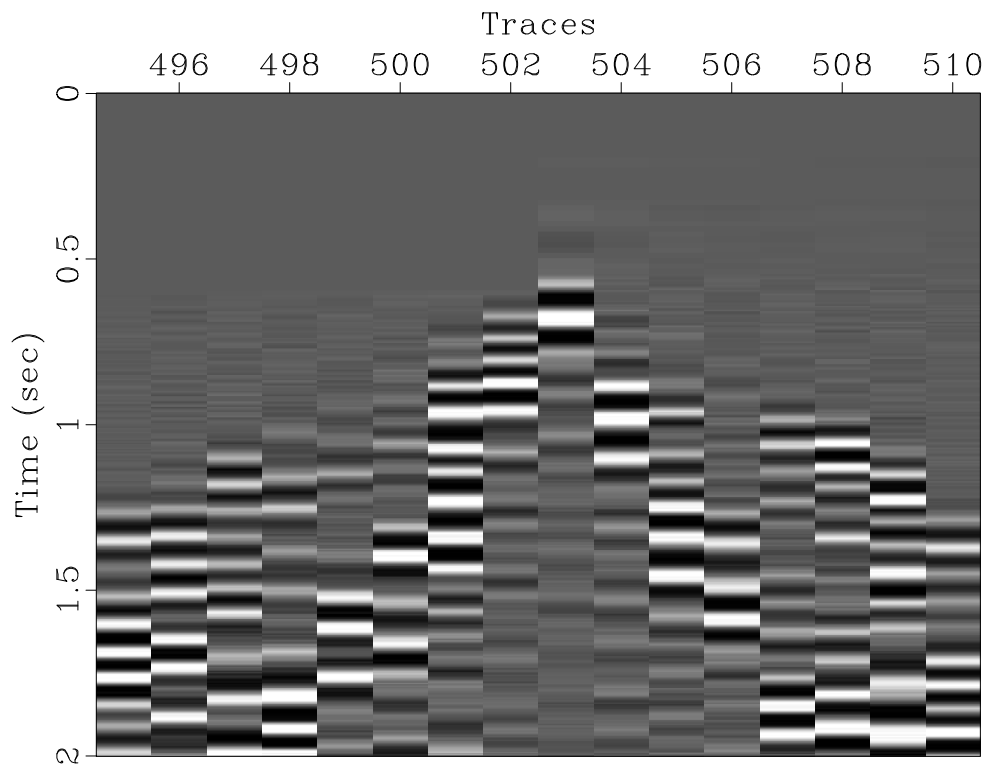
Figure 5.7: Single shot gather of surface seismic synthetics using the VSP updated model as initial model a)Vertical component and b)Radial component
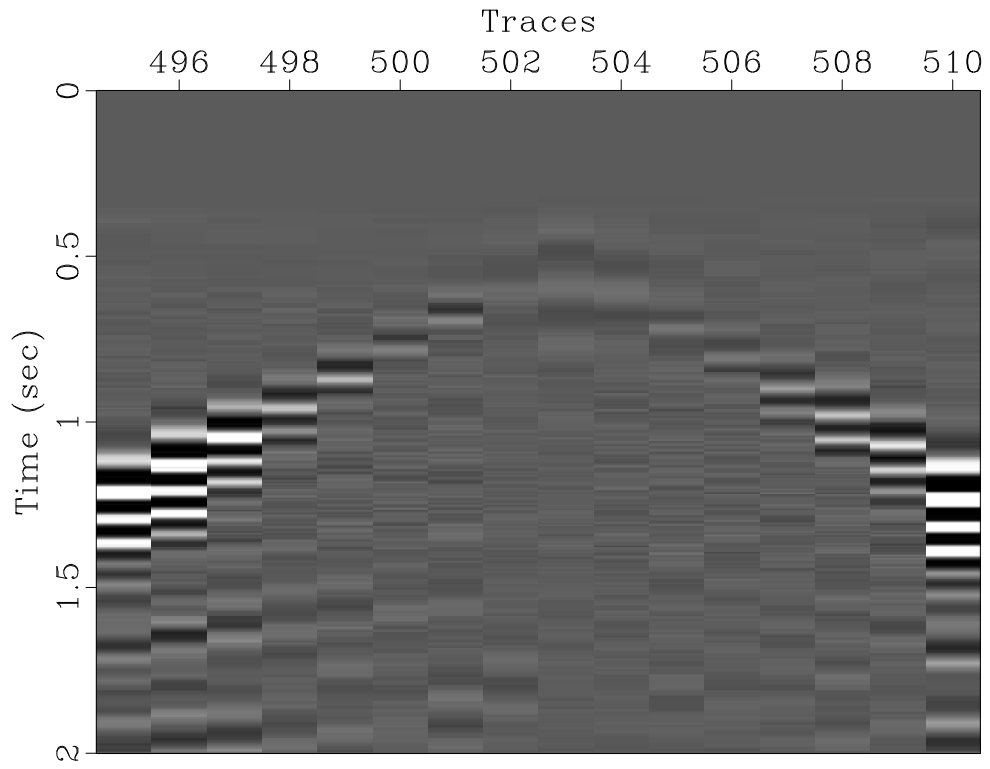
(a)
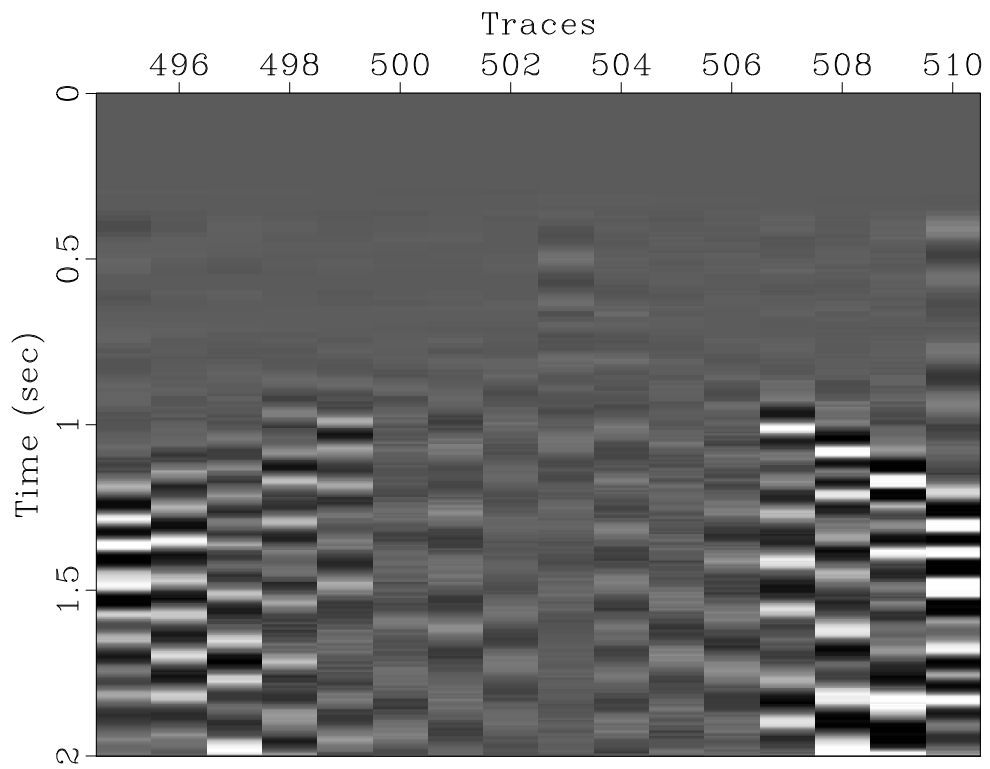
(b)

Figure 5.8: Data residuals using initial linear model a)Vertical component and b)Radial component
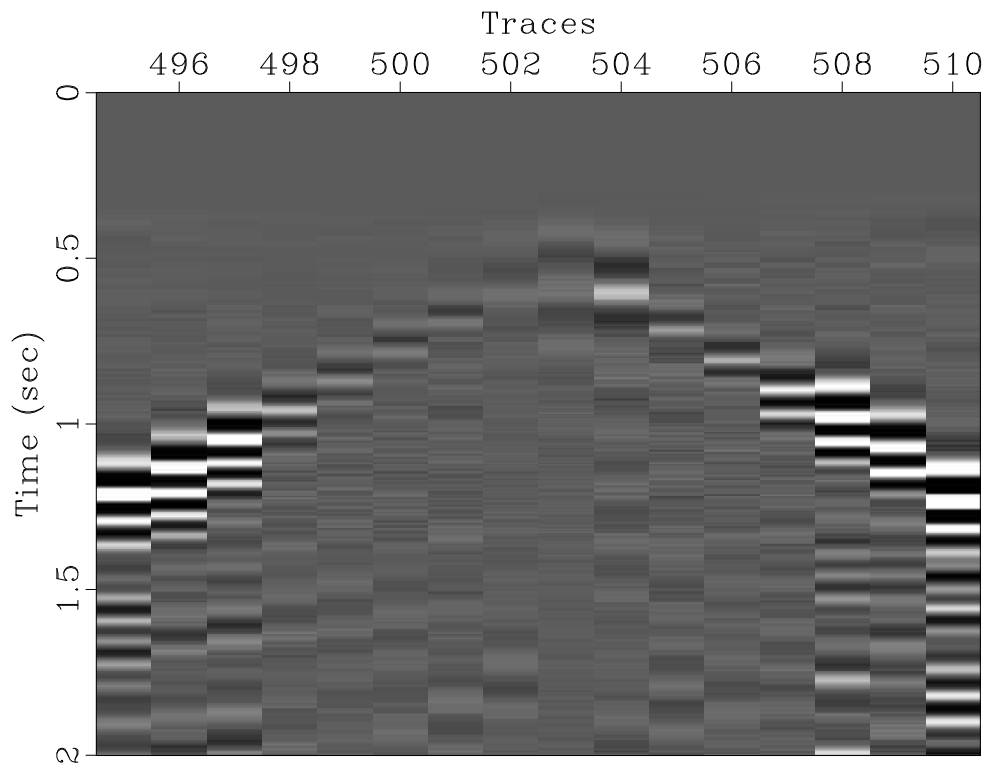
(a)
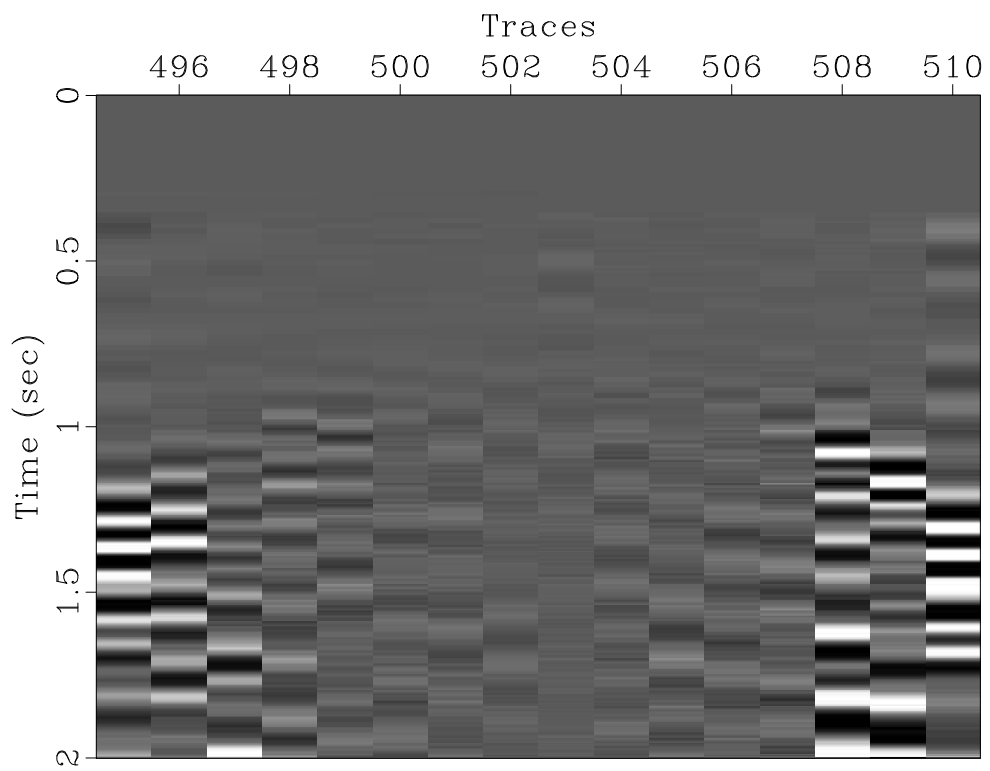
(b)

Figure 5.9: Data residuals using VSP updated model a)Vertical component and b)Radial component
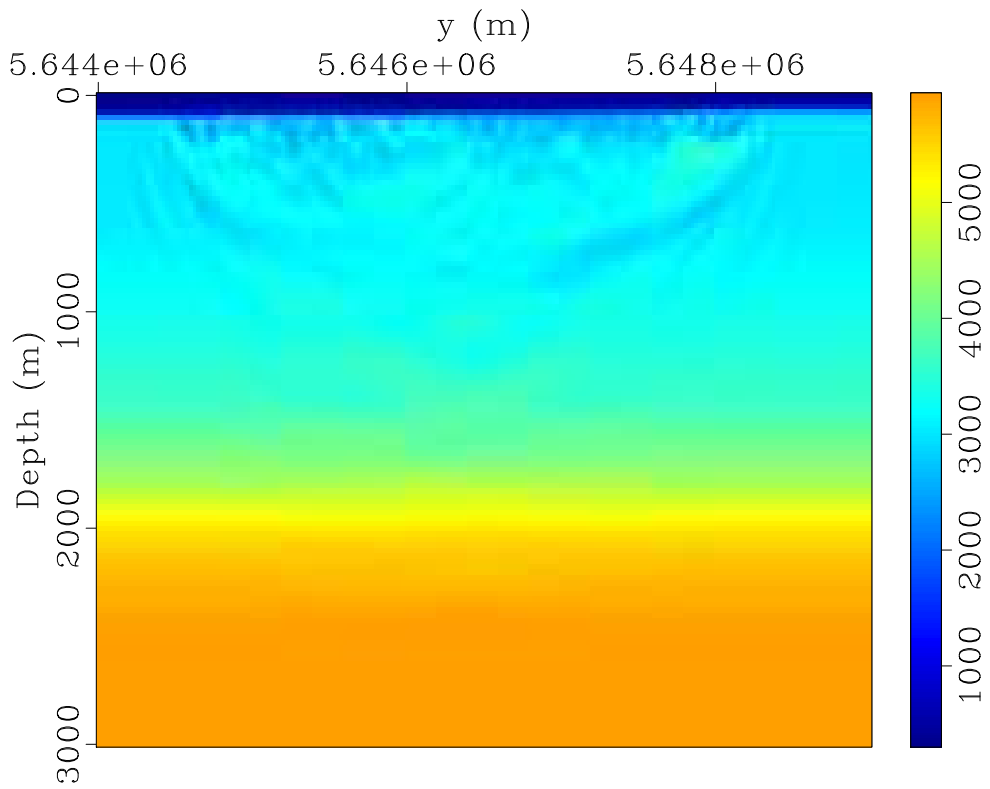
(a)
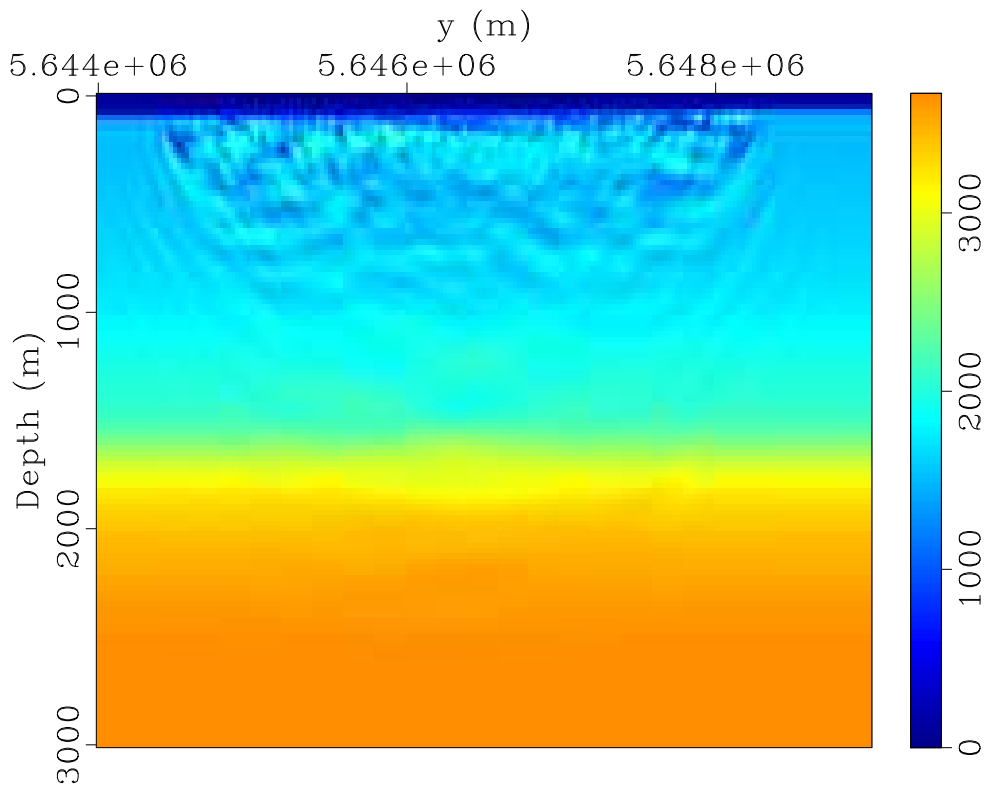
(b)

Figure 5.10: FWI updated models after using linearly increasing velocity model as initial model after 20 iterations: a)Vp and b)Vs using LBFGS Optimization

55
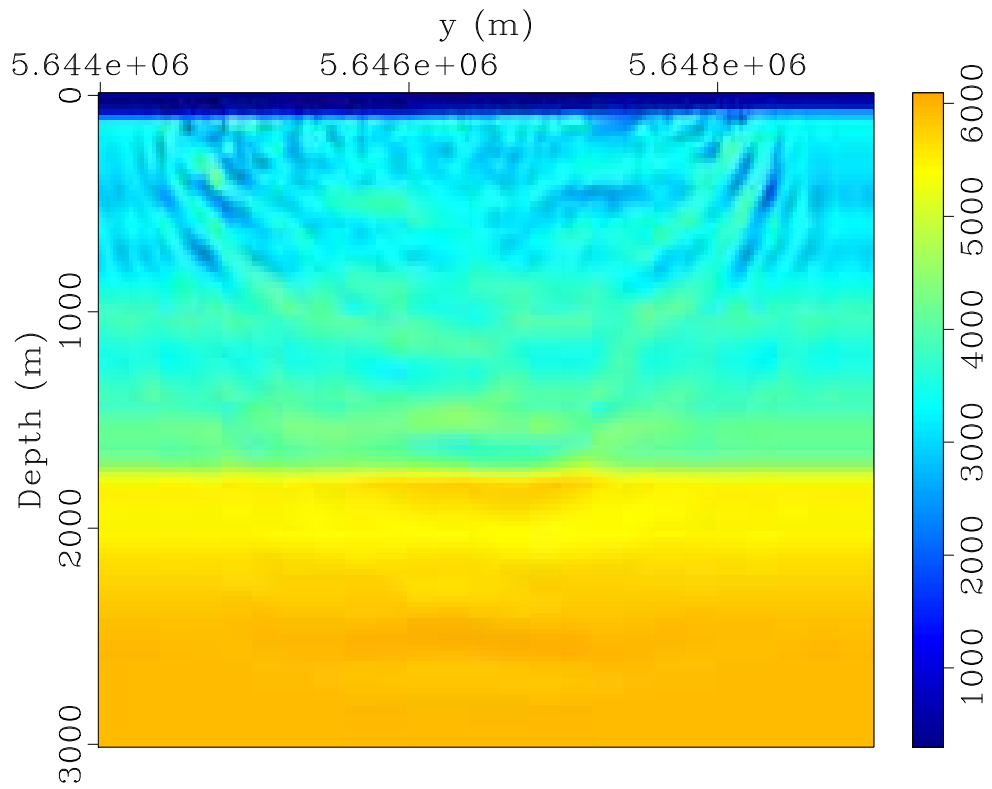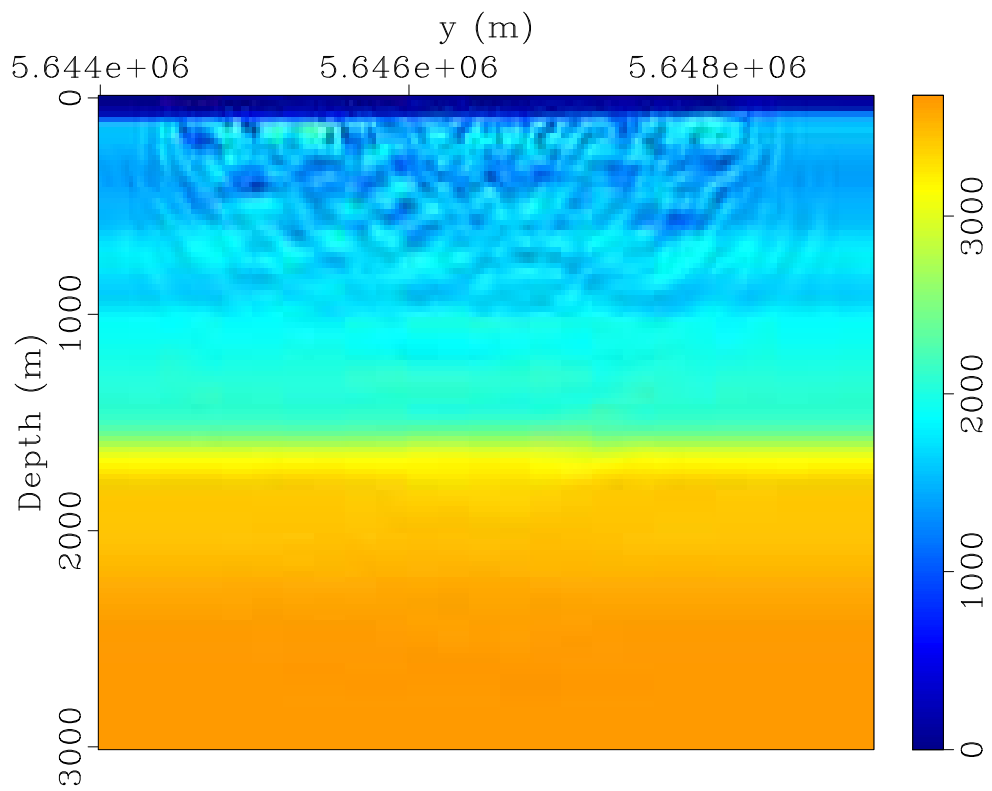
Figure 5.11: FWI updated models after using the updated VSP model as initial model after 20 iterations: a)Vp and b)Vs using L-BFGS Optimization
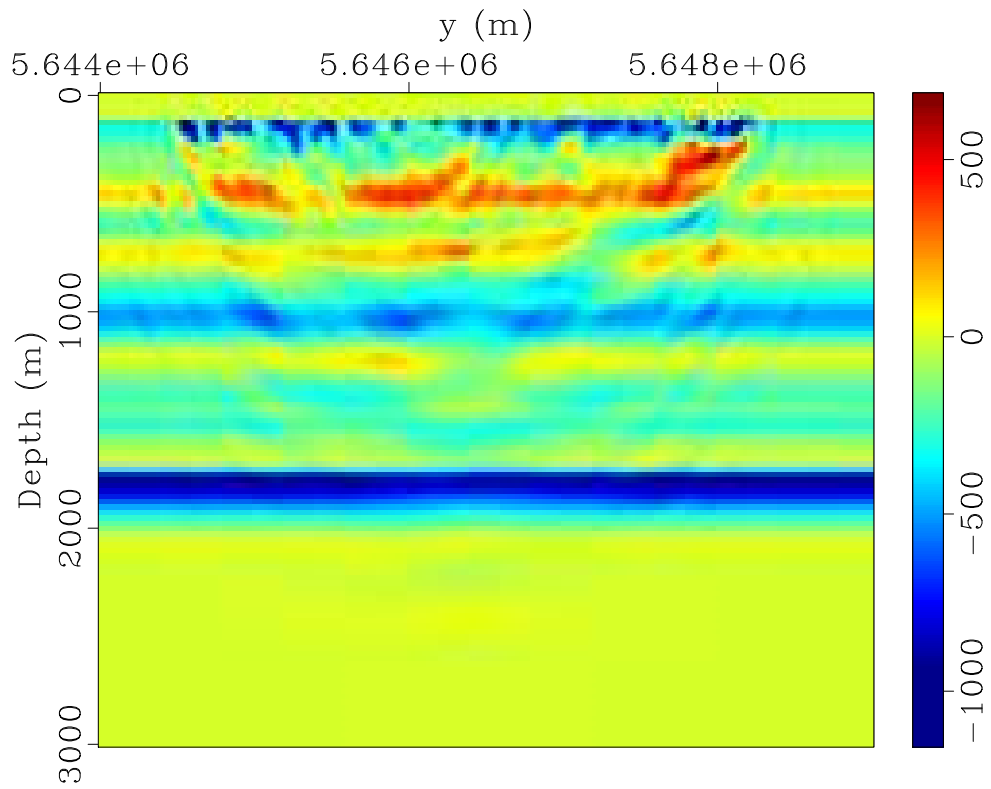
(a)
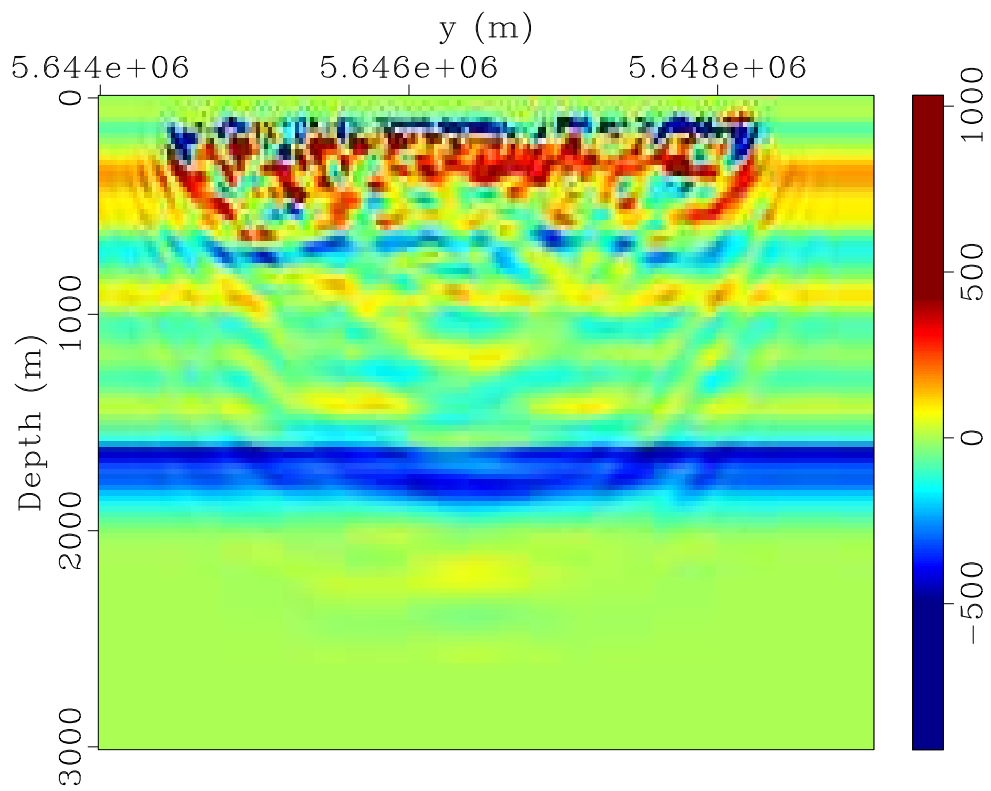
(b)

Figure 5.12: Accumulated FWI perturbations using the linearly increasing velocity model as initial model : a)Vp and b)Vs using L-BFGS Optimization
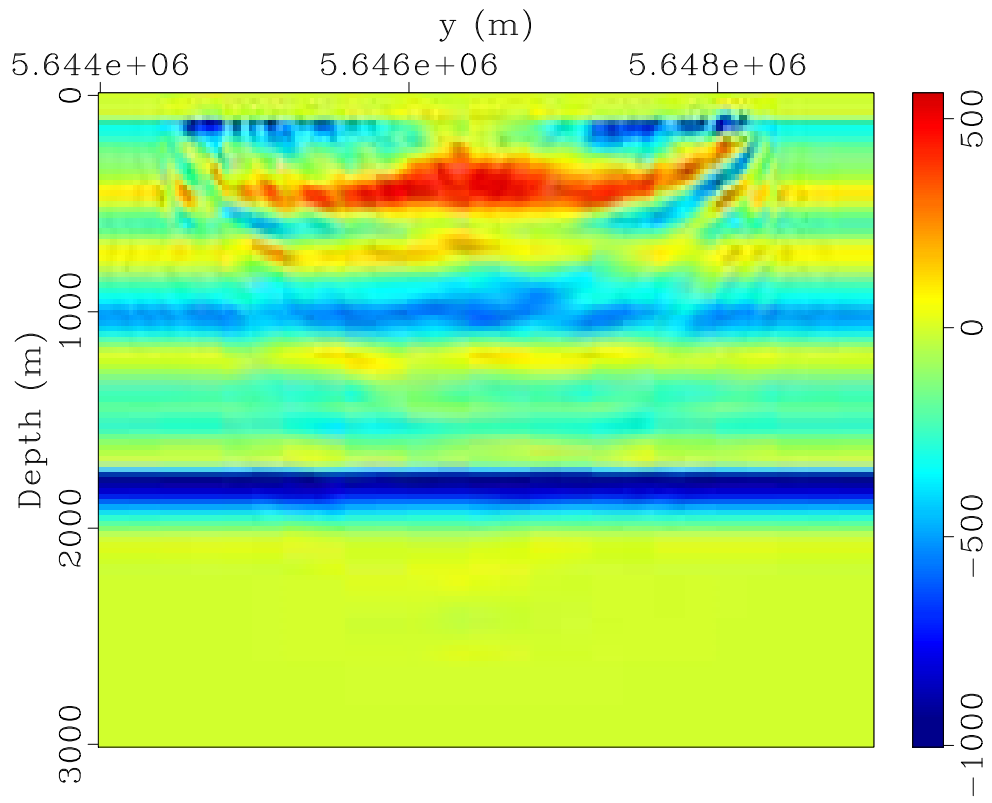
(a)
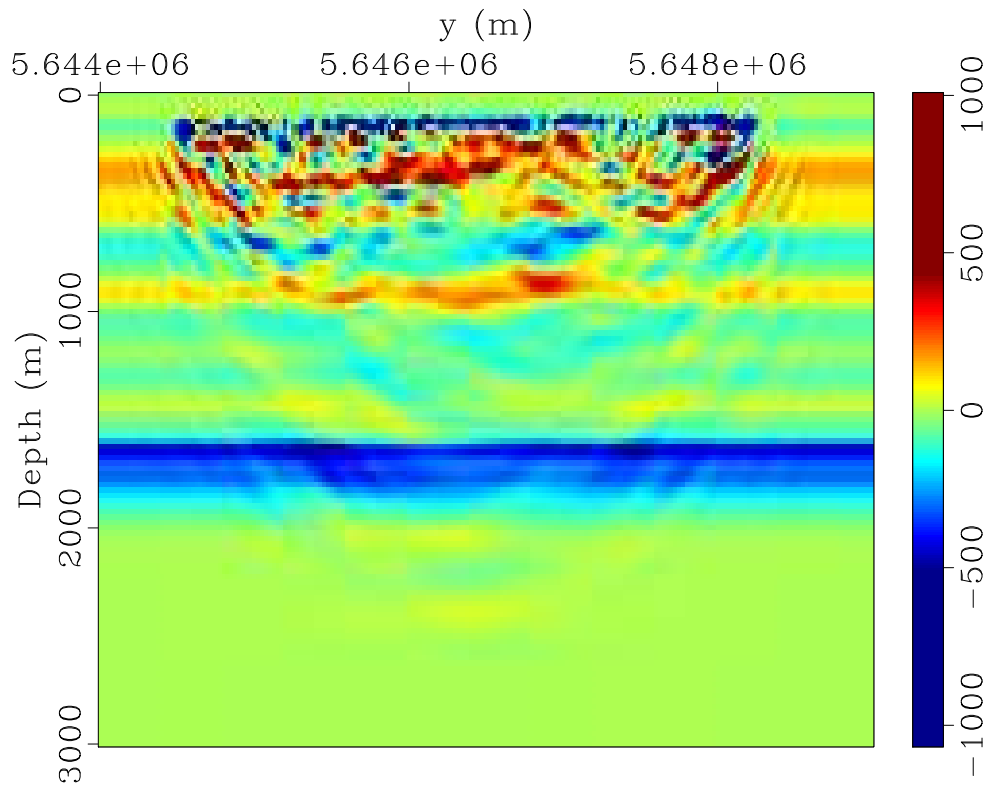
(b)

Figure 5.13: Accumulated FWI perturbations using the linearly increasing velocity model as initial model: a)Vp and b)Vs using Conjugate Gradient Optimization

58
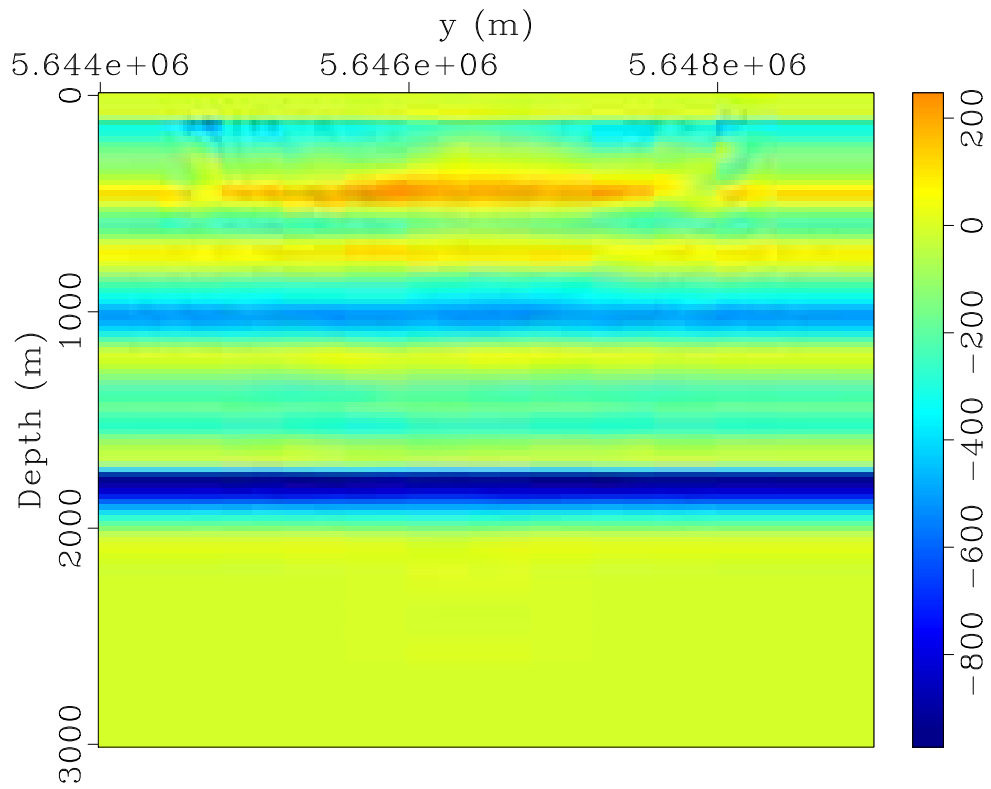
(a)
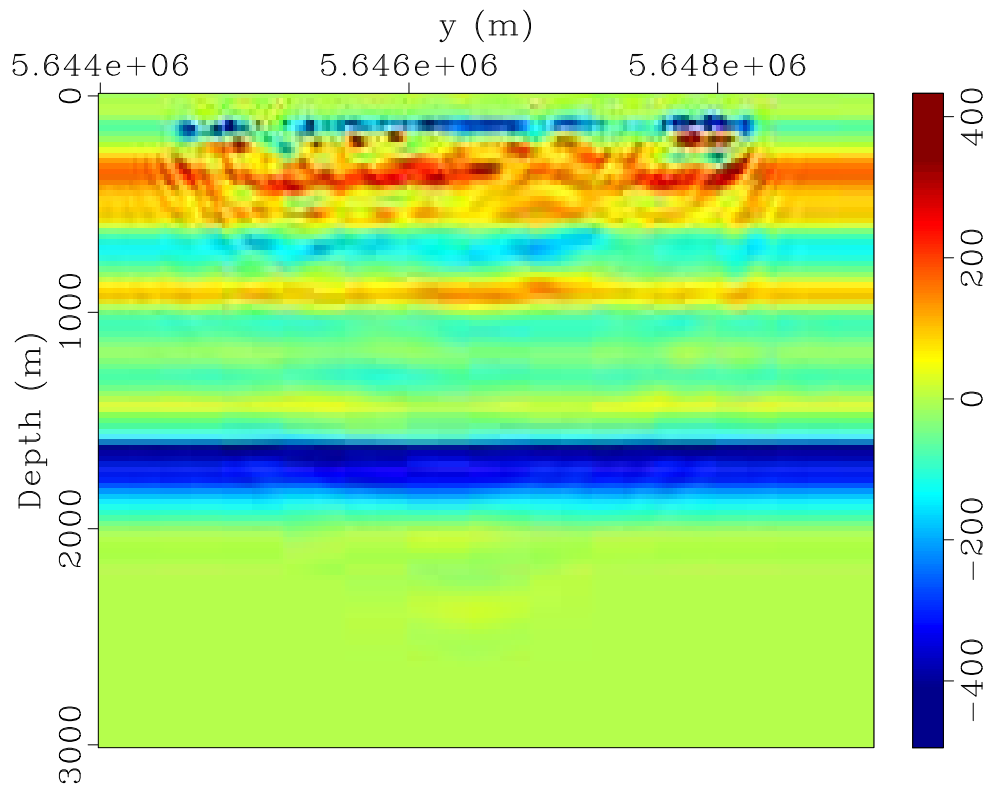
(b)

Figure 5.14: Accumulated FWI perturbations using the linearly increasing velocity model as initial model: a)Vp and b)Vs using Steepest Descent
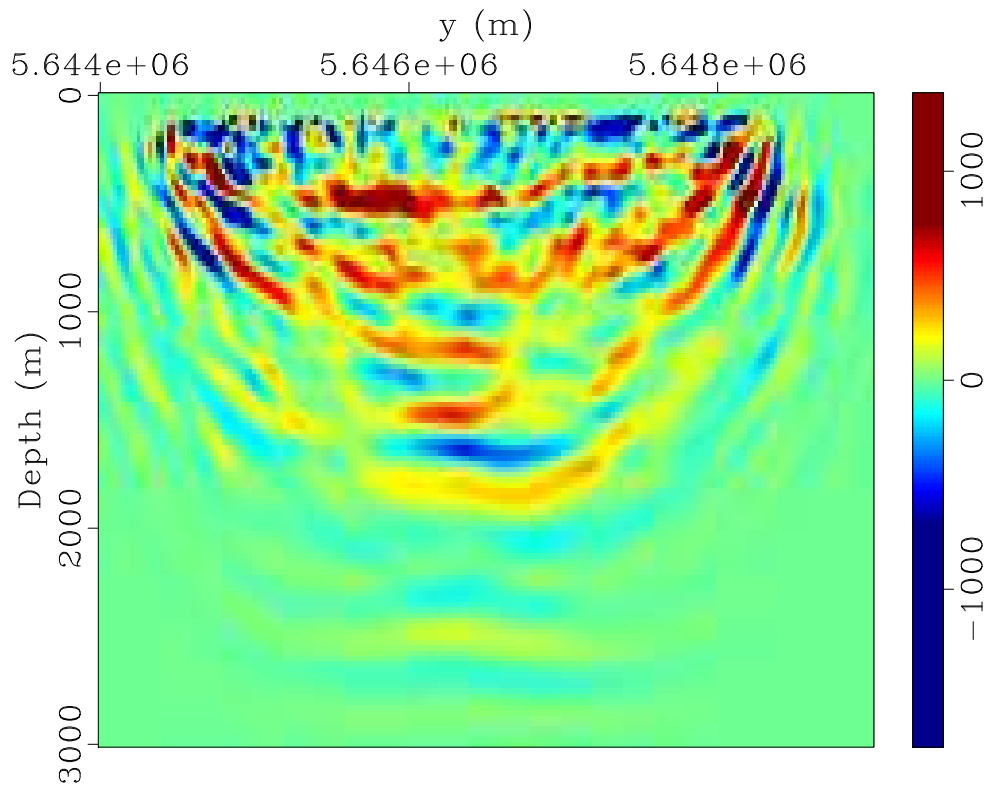
(a)
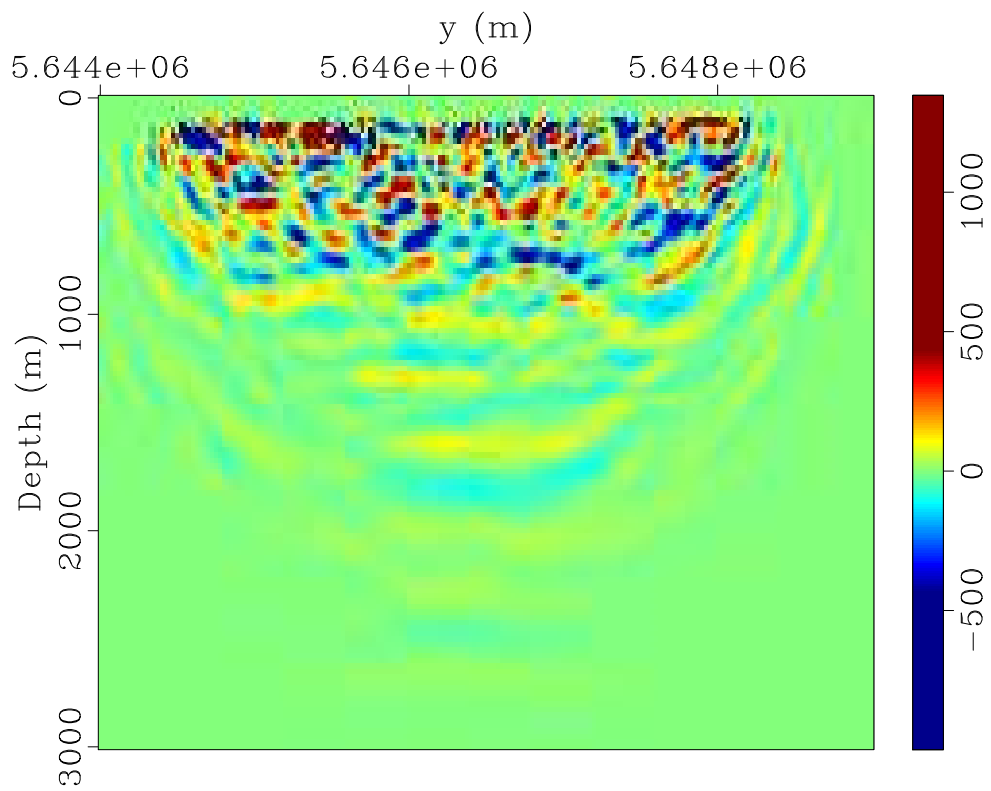
(b)

Figure 5.15: Accumulated FWI perturbations after 20 iterations using VSP updated model as initial model: a)Vp and b)Vs using L-BFGS optimization method   60

5.1.2 Elastic Reverse Time Migration

Elastic Reverse Time Migration (ERTM) was used to to check the accuracy the resultant models and compare the images based on the quality of the images. The updated models, either from using linear initial models (Figure 5.10) or VSP updated models using L-BFGS optimization (Figure 5.11), were used for application of ERTM on the 2D seismic section. The used Imaging condition (IC) is based on (Sava & Fomel, 2006; Yan & Sava, 2008) Extended Elastic IC (See Appendix B). Figures 5.16 and 5.17 show the P-wave and S-wave RTM image using the updated velocity model that resulted from application of FWI using linear 1D model as initial model. On the other hand, Figures 5.18 and 5.19 show the P-wave and S-wave using the updated models from application of FWI using VSP updated models as starting point. The resultant images (Figures 5.18 and 5.19) were more clear, as indicates by arrows in the figures, and that indicates that the VSP updated models help the FWI to optimize better results.
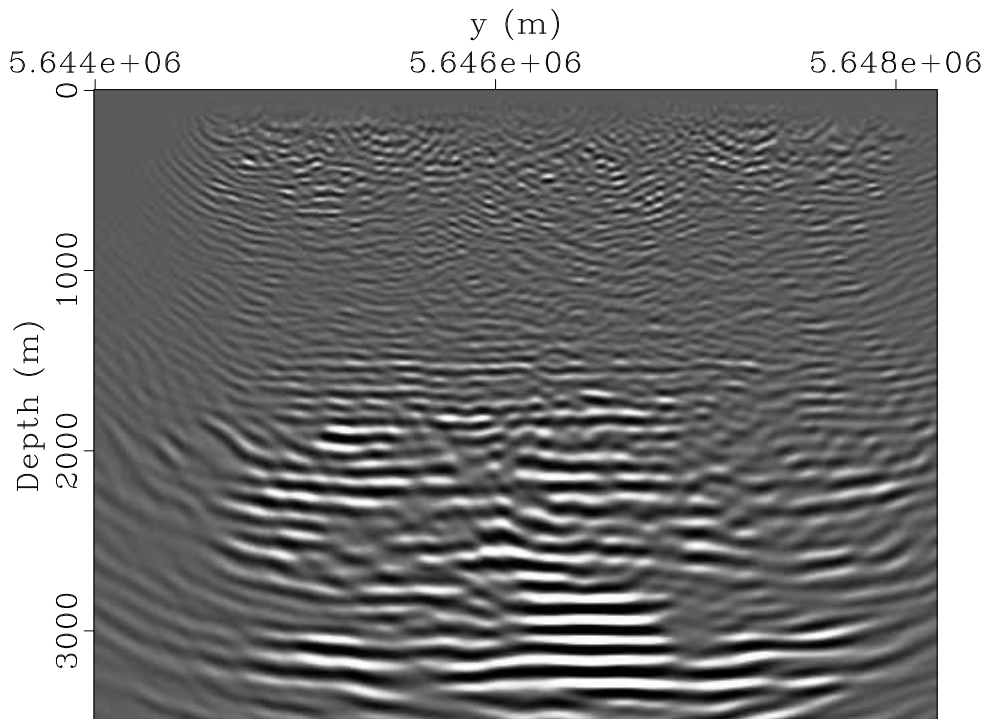
Figure 5.16: P-wave RTM image using linear model as initial model.
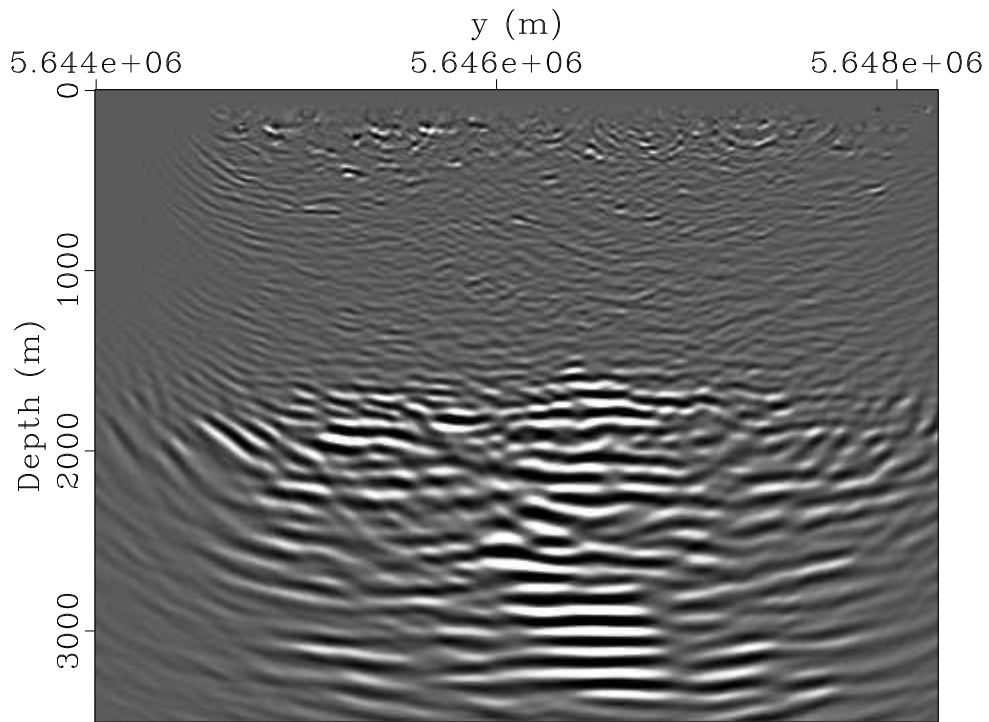
Figure 5.17: S-wave RTM image using linear model as initial model.
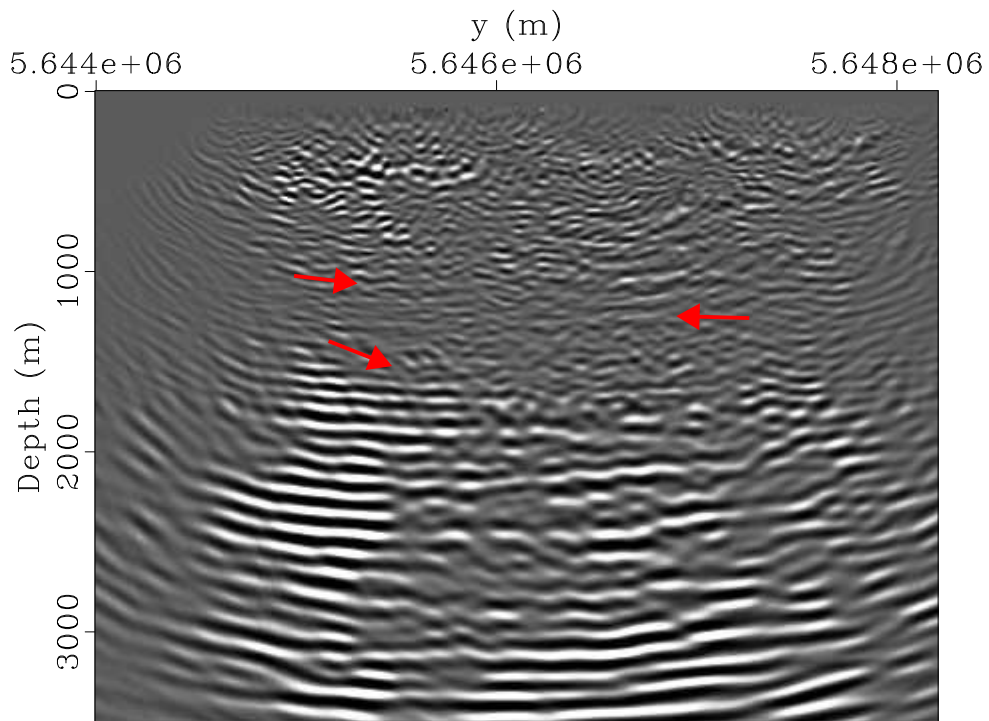
Figure 5.18: P-wave RTM image using VSP updated model as initial model.
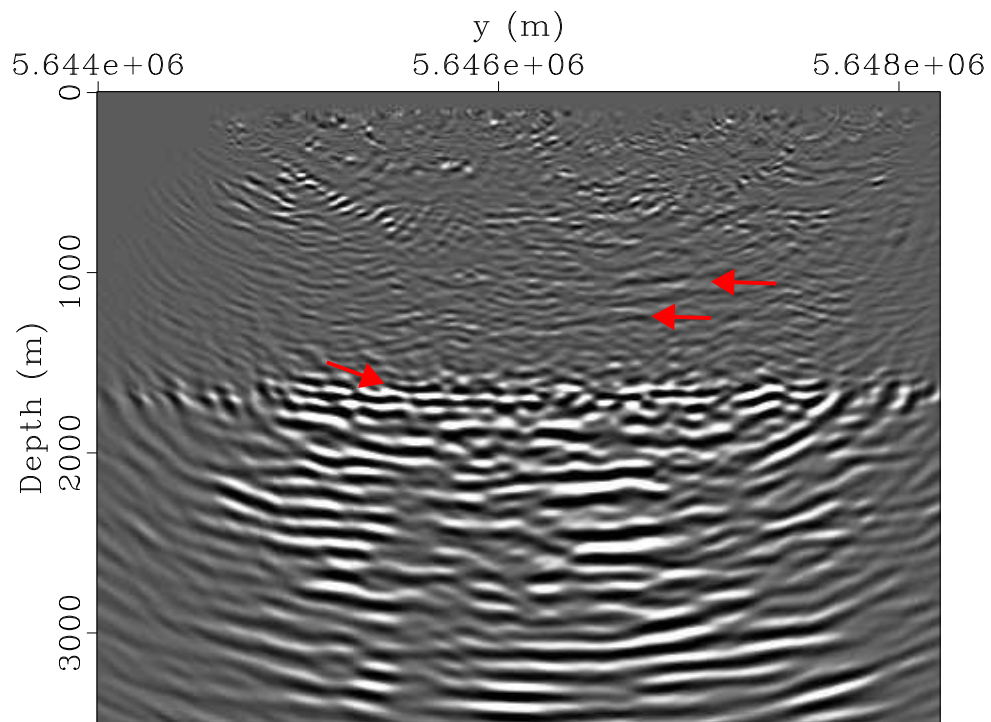
Figure 5.19: S-wave RTM image using VSP updated model as initial model. Notice that the reflectors are better focused

6 Conclusion

This study shows an application of elastic FWI to Blackfoot multicomponent VSP and surface data from Western Canadian Basin. FWI simultaneously updates the P-wave and S-wave velocity models in order to match the observed data. This FWI study faced many challenges that mainly result from the computation limitations, non-linearity of inverse problem, inadequate initial model and poor wavelet estimate.

Application of FWI on the VSP data gives an accurate model which can be used as starting point for many methods from estimating rock physical properties and inversion of surface data. Having more accurate initial velocity model results in synthetics that is closer to the real data. Also, it helps to improve the resolution of migrated seismic section.

Cutting high frequency was a robust idea to reduce the effect of cycle-skipping. Also, using a multiscale approach was very helpful to reduce the inaccuracy in shallow model and to enable further updating the deeper section. Different optimization methods like steepest descant, conjugate gradient and qausi-Newton were used. In this work, the quasi-Newton method was the best choice for optimization as it tends to save computational power and speed up convergence.

Application of Elastic FWI on the whole 3D volume is a very expensive process to apply, however, 3D FWI provides a better constrains on the velocity models as the 3D volume has higher fold and angle coverage than the extracted 2D line.

Through this project, we wanted to keep the method as simple as possible and avoid using highly complicated method to save the computation cost. However, it might be helpful to include further details to the model. The elastic media was assumed to be isotropic. But considering the anisotropy in this case might be helpful for accurate modeling the synthetics to estimate better results. Also implementing the absorption of amplitudes could be helpful however modeling of amplitudes and anisotropy needs a huge computation cost as it will make the solution more complicated.

The finite difference, with perfectly matching layer (PML) condition, was able to compute accurate synthetics. However, using spectral element and finite element might be helpful to more accurately model the topography and free surface.

References

[1] Aki K. and Richards P., 1980, Quantitative Seismology, 2nd Edition, W. H. Freeman and Co., San Francisco.

[2] Anagaw A. & Sacchi M., 2011, Full Waveform Inversion with Total Variation Regularization, 2011 CSPG CSEG CWLS Convention.

[3] Baysal, E., Kosloff, D., and Sherwood, J., 1983, Reverse time migration, GEOPHYSICS, VOL.48, NO.II, 1514-1524

[4] Boyd S. & Vandenberghe L., 2004, Convex Optimization, Cambridge University Press, ISBN 9780521833783.

[5] Charara, M., Barnes, C. and Tarantola, A., 1996, The state of affairs in inversion of seismic data: an OVSP example, 66th Annual International Meeting, SEG, Expanded Abstracts, 1999–2002.

[6] Claerbout, J., 1971, Toward a a unified theory of reflector mapping, GEOPHYSICS, 36(3), 467-481.

[7] Dai M. & Chen J., 2016, Studies on multi-scale full waveform inversion for time-domain acoustic wave equation based on virtual-source precondition, 7th International Conference on Environmental and Engineering Geophysics, Atlantis Press.

[8] Claerbout, J., 2010, Basic Earth Imaging, Stanford Exploration Project.

[9] Du, Q., Zhu, Y., and Ba, J., 2012, Polarity reversal correction for elastic reverse time migration: Geophysics, 77, No. 2, S31–S41.

[10] Fernandez-Granda C., 2015, Statistical and Mathematical Methods, DS-GA 1002, New York University.

[11] Esser† E., Guasch L., Leeuwen T., Aravkin A. and Herrmann F., 2016, Total-variation regularization strategies in full-waveform inversion.

[12] Fomel S., 2000, Three-dimensional seismic data regularization, PhD dissertation, Stanford Exploration Project, CA.

[13] Guevara S. and Stewart R., 1998, Multicomponent seismic polarization analysis, CREWES Research Report — Volume 10.

[14] Gulati, J., Stewart, R., and Parkin, J., 1998, 3C-3D VSP: The Blackfoot experiment, CREWES Research Report Volume 10.

[15] Gulati J., Stewart R. and Parkin J., 2004, Analyzing three-component 3D

vertical seismic profiling data, GEOPHYSICS, VOL.69, NO.2, P.386–392.

[16] Hinterműller M., 2013, Non-Linear Optimization, Berlin School of Mathematics.

[17] Hua, H., Zoub Z., Woa, Y., Youna, O., 2018, Reverse time migration: A prospect of seismic imaging methodology, Earth-science Reviews, 197, 207-227.

[18] Huang Y., 2013, Least-squares Migration and Full Waveform Inversion with Multisource Frequency Selection, King Abdullah University of Science and Technology

[19] Huang Y., Schuster G., 2004, Resolution limits for wave equation imaging, Journal of Applied Geophysics Volume 107, Pages 137-148.

[20] Ibrahim, F., Weibull, W., 2017, Converted wave imaging and velocity analysis using elastic reverse time migration, University of Stavanger.

[21] Jiang, Z., Bancroft, J., and Lines L., 2012, Elastic imaging conditions based on Helmholtz decomposition, CREWES Research Report — Volume 24.

[22] Jiang, Z., Bancroft, J., and Lines, L., 2012, Elastic reverse-time migration, CREWES Research Report-Volume 24.

[23] Kabanikhin S., 2008, Definitions and examples of inverse and ill-posed problems, J. Inv. Ill-Posed Problems 16 (2008), 317–357.

[24] Lawton, D., Robert, R., Stewart, Cordsen, A., and Hrycak, S., 2009, Advances in 3C-3D design for converted waves.

[25] Lu H. and Margrave G., 1999, A study of multi-mode conversions in the Blackfoot 3C seismic data, CREWES Research Report — Volume 11.

[26] Liu Y. and Sen M., 2009, Advanced finite-difference methods for seismic modeling, GEOHORIZONS.

[27] Marcinkovich, C. and Olsen, K., 2013, on the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme, Journal Of Geophysical Research, VOL.108,NO.B5,2276.

[28] Matignon R., 2005, Neural Network Modeling Using SAS Enterprise Miner, ISBN:1418423416.

[29] Miller S., Aydemir E. and Margrave G., 1995, Preliminary interpretation of P-P and P-S seismic data from the Blackfoot broad-band survey, CREWES Research Report - Volume 7.

[30] Miller S., 1996, Multicomponent Seismic Data Interpretation, The University of Calgary.

[31] Mora P., 1989, "Inversion = migration + tomography" GEOPHYSICS, 54(12), 1575-1586.

[32] Nocedal, J.and Wright, S., 1999, Numerical optimization, Springer.

[33] Pan, W., Innanen, K., Geng, Y., 2018, Elastic full-waveform inversion and parameterization analysis applied to walk-away vertical seismic profile data for unconventional (heavy oil) reservoir characterization, Geophysical Journal International.

[34] Plessix R., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, Geophys. J. Int. 167, 495–503.

[35] Podgornova, O., Leaney, S., Charara, M. and Lunen E., 2014, Elastic full waveform Inversion for land walkaway VSP data, GeoConvention 2014: FOCUS.

[36] Pumplin J., Stump D. and Tunga W., 2000, Multivariate Fitting and the Error Matrix in Global Analysis of Data, MSU-HEP-07100, CERN-TH/2000-249.

[37] Richardson A., 2018, Seismic Full-Waveform Inversion Using Deep Learning Tools and Techniques, Cornell University, arXiv:1801.07232.

Ramos-Martinez J., Crawley S., Zou Z., Valenciano A., Qiu L. & Chemingui N., A Robust Gradient for Long Wavelength FWI Updates, 2016.

[38] Sambridge, M., Inverse problems in a nutshell. Center for Advanced Data Inference, Research School of Earth Sciences, Australian National University, ACT 0200, Australia.

[39] Schuster G., 2015, Seismic Inversion,Investigations in Geophysics series n.20, Society of Exploration Geophysicists.

[40] Selesnick I. and Bayram I., 2010, Total Variation Filtering, New York University.

[41] Sen M. and Stoffa P., 2013 , Global Optimization Methods in Geophysical Inversion, Cambridge University Press.

[42] Simin V., Harrison M. and Lorentz G., 1996, Processing the Blackfoot 3C-3D seismic survey, CREWES Research Report — Volume 8.

[43] Snieder R., 1998, The role of nonlinearity in inverse problems, Inverse

Problems 14, 387–404.

[44]Spetzler J. and Snieder R., 2004 The Fresnel volume and transmitted waves, GEOPHYSICS, VOL. 69, NO. 3 (MAY-JUNE 2004); P. 653–663, 12 FIGS.10.1190/1.1759451.

[45] Stewart R., Gaiser J., Brown R. and Lawton D., 2002 Tutorial Converted-wave seismic exploration: Methods, GEOPHYSICS, VOL. 67, NO. 5, P. 1348–1363, 18 FIGS. 10.1190/1.1512781.

[46] Stewart, R., Gulati, J. , Zhang, Q., Parkin, J.,and Peron, J., Analysing 3C-3D VSP data: The Blackfoot, Alberta survey, SEG Technical Program Expanded Abstracts 1998: pp. 381-384.

[47] Stewart R., Hoffe B., Bland H., Margrave G., Gallant E., Bertram M., 1997, The Blackfoot high-resolution 3-C seismic survey: design and initial results, CREWES Research Report — Volume 9.

[48] Sun, W. & Yuan, Y., 2006, Optimization Theory and Methods: Nonlinear Programming, Springer US.

[49] Tarantola A., 1984, Inversion of seismic reflection data in the acoustic approximation, GEOPHYSICS, 49(8), 1259-1266.

[50] Trampert J., 1998 ,Global seismic tomography: the inverse problem and beyond, Inverse Problems 14 (1998) 371–385.

[51] VanDecar J. and Snieder R., 1994, Obtaining smooth solutions to large, linear, inverse problems, GEOPHYSICS, VOL. 59, NO. 5 (MAY 1994); P. 818-829, 10 FIGS.

[52] Vigh D., Jiao K., and Watts D., 2012, Elastic full-waveform inversion using 4C data acquisition. SEG Technical Program Expanded Abstracts 2012.

[53] Virieux, J., Etienne, V., Cruz-Atienza, V., Brossier, R., Chaljub, E., et al.. Modelling Seismic Wave Propagation for Geophysical Imaging. Seismic Waves - Research and Analysis, Masaki Kanao, 253-304, Chap.13, 2012, 978-953-307-944-8. <hal-00682707>.

[54] Virieux J. and Operto S., 2009, An overview of full-waveform inversion in exploration geophysics., GEOPHYSICS, 74(6), WCC1-WCC26.

[55] Virieux J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method, GEOPHYSICS, 51(4), 889-901.

[56] Weglein, A., Herreta, W., Initial study and implementation of the con-

volitional Perfectly Matched Layer for modeling of the 2D acoustic wave equation, University of Houston, 2013

[57] Whitmore, N., 1983, Iterative depth migration by backward time propagation: 53rd Annual International Meeting, SEG, Expanded Abstracts, 382–385

[58] Wolfe P., 1969, Convergence Conditions for Ascent Methods. SIAM Review 11(2):226-235.

[59] Weibull W., 2016, 4D elastic full-waveform inversion Making full use of the life of field seismic datasets, Workshop on 4D seismic and history matching, IOR NORWAY 2016, University of Stavanger.

[60] Yan J. and Sava P., 2008, Isotropic angle-domain elastic reverse-time migration, GEOPHYSICS, VOL.73, NO.6, 229–239

[61] Yang, P., 2014, A numerical tour of wave propagation, Technical report, Xi'an Jiaotong University.

[62] Zhang Q., Stewart R., Parkin J., Sun Z., 1996, Analysis of the Blackfoot 3C-3D VSP survey, CREWES Research Report Volume 8.

Appendix A   Waves in Elastic media

A.1   Waves propagation

Describing particle movement is accomplished using two fundamental laws, Newton's second law of motion and and Hooke's Law. The mathematical derivation in this setting is based on 1D acoustic model (horizontal movements of particles).
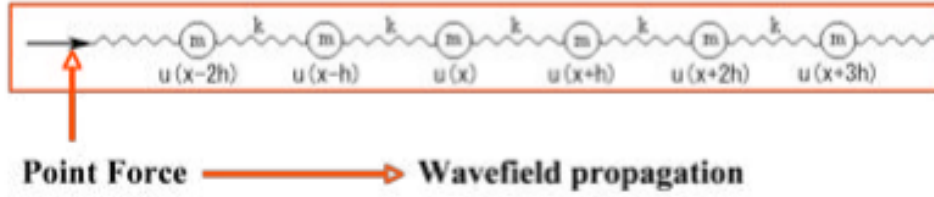
Figure A.1: One dimensional model.

Newton's law, force equals mass m times acceleration, is expressed in the following expression, while the acceleration is the rate of velocity change.

$$\tau(x,t) = m \frac{v(x, t+\Delta t) - v(x,t)}{\Delta t} \tag{A.1}$$

While τ is the stress or force per unit area.

$$\tau(x,t) = m \frac{u(x, t+\Delta t) - 2u(x,t) + u(x, t-\Delta t) - v(x,t)}{\Delta t^2} \tag{A.2}$$

While v is the change in particle position u over time $v_i = \frac{\partial u_i}{\partial t}$.

$$\tau(x,t) = \tau(x+h,t) - \tau(x-h,t) = k \frac{u(x+h,t) - u(x,t)}{h} + k \frac{u(x,t) - u(x-h,t)}{h} \tag{A.3}$$

$$\frac{\tau(x+h,t) - \tau(x-h,t)}{h} = \rho \frac{u(x, t+\Delta t) - 2u(x,t) + u(x, t-\Delta t)}{\Delta t^2} \tag{A.4}$$

$$\frac{\partial \tau}{\partial x} = \rho \frac{\partial^2 u}{\partial t^2} \tag{A.5}$$

Equation A.5 is the first order partial-differential equation relating second order change in time to a first order change in force per unit area (stress). Stepping up to

70

the simple isotropic elastic models, equation A.5 can be written in the the following form in 3D with 9 stresses τ_{ij}.

$$\frac{\partial^2 u_i}{\partial t^2} = \frac{1}{\rho} \sum_{j=1}^{3} \frac{\partial \tau_{ij}}{\partial x_j} \tag{A.6}$$

Since Hook's law relates τ_{ij} to u_i.

$$\tau_{ij} = \sum_{m,n} c_{ijmn} E_{mn} \tag{A.7}$$

Then, equation A.6 can be expressed as the follwoing complex system (i=1,2,3) of anistropic equations of motion.

$$\frac{\partial^2 u_i}{\partial t^2} = \sum_{m,n,j} \frac{c_{ijmn}}{\rho} \frac{\partial^2 u_m}{\partial x_n \partial x_j} \tag{A.8}$$

A.2 Anisotropy

Relationship between the stress τ and strain u for polar anisotropic media can be expressed as the following while z-axes as the symmetry axis.

$$\begin{pmatrix} \tau_{xx} \\ \tau_{yy} \\ \tau_{zz} \\ \tau_{xy} \\ \tau_{yz} \\ \tau_{zx} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{11}-2c_{66} & c_{13} & 0 & 0 & 0 \\ c_{11}-2c_{66} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{66} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{pmatrix} * \begin{pmatrix} u_{xx} \\ u_{yy} \\ u_{zz} \\ u_{xy} \\ u_{yz} \\ u_{zx} \end{pmatrix} \tag{A.9}$$

while c is the stiffness tensor

$$V_p = \sqrt{\frac{c_{33}}{\rho}} \tag{A.10}$$

$$V_s = \sqrt{\frac{c_{44}}{\rho}} \tag{A.11}$$

And the measure of anisotropy in the simplest form can be expressed in the following form (Thomsen parameters).

$$\varepsilon = \frac{c_{11} - c_{33}}{2c_{33}} \tag{A.12}$$

$$\gamma = \frac{c_{66} - c_{44}}{2c_{44}} \tag{A.13}$$

$$\delta = \frac{1}{2} \frac{(c_{13} + c_{44})^2 - (c_{33} - c_{44})^2}{c_{33}(c_{33} - c_{44})} \tag{A.14}$$

A.3 Isotropic Elastic symmetry

If the subsurface is assumed isotropic, the elastic constants reduced to two parameters, the Lamé parameters, which depend on V_p and the V_s. the c can be reduced to

$$c = \begin{pmatrix} \lambda+2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda+2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda+2\mu & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix} \tag{A.15}$$

$$V_p = \sqrt{\frac{\lambda+2\mu}{\rho}} \tag{A.16}$$

$$V_s = \sqrt{\frac{\mu}{\rho}} \tag{A.17}$$

Given the matrix (equation A.15) and hooke's law (equation A.8):

$$\frac{\partial^2 u_1}{\partial t^2} = \frac{1}{\rho}\left(\frac{\partial \tau_{11}}{\partial x_1} + \frac{\partial \tau_{12}}{\partial x_2} + \frac{\partial \tau_{13}}{\partial x_3}\right)$$

$$\frac{\partial^2 u_2}{\partial t^2} = \frac{1}{\rho}\left(\frac{\partial \tau_{21}}{\partial x_1} + \frac{\partial \tau_{22}}{\partial x_2} + \frac{\partial \tau_{23}}{\partial x_3}\right)$$

$$\frac{\partial^2 u_3}{\partial t^2} = \frac{1}{\rho}\left(\frac{\partial \tau_{31}}{\partial x_1} + \frac{\partial \tau_{22}}{\partial x_2} + \frac{\partial \tau_{33}}{\partial x_3}\right)$$

$$\tau_{11} = \frac{\lambda+2\mu}{\rho}\frac{\partial u_1}{\partial x_1} + \frac{\lambda}{\rho}\frac{\partial u_2}{\partial x_2} + \frac{\lambda}{\rho}\frac{\partial u_3}{\partial x_3}$$

$$\tau_{12} = \frac{\mu}{\rho}\left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1}\right)$$

$$\tau_{13} = \frac{\mu}{\rho}\left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1}\right)$$

$$\tau_{22} = \frac{\lambda}{\rho}\frac{\partial u_1}{\partial x_1} + \frac{\lambda+2\mu}{\rho}\frac{\partial u_2}{\partial x_2} + \frac{\lambda}{\rho}\frac{\partial u_3}{\partial x_3}$$

$$\tau_{23} = \frac{\mu}{\rho}\left(\frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3}\right)$$

$$\tau_{33} = \frac{\lambda}{\rho}\frac{\partial u_1}{\partial x_1} + \frac{\lambda}{\rho}\frac{\partial u_2}{\partial x_2} + \frac{\lambda+2\mu}{\rho}\frac{\partial u_3}{\partial x_3} \tag{A.18}$$

Taking the partial derivatives of stress-strain terms gives the Velocity-stress system of equations in 3D.

$$\frac{\partial v_1}{\partial t} = \frac{1}{\rho}\left(\frac{\partial \tau_{11}}{\partial x_1} + \frac{\partial \tau_{12}}{\partial x_2} + \frac{\partial \tau_{13}}{\partial x_3}\right)$$

$$\frac{\partial v_2}{\partial t} = \frac{1}{\rho}\left(\frac{\partial \tau_{21}}{\partial x_1} + \frac{\partial \tau_{22}}{\partial x_2} + \frac{\partial \tau_{23}}{\partial x_3}\right)$$

$$\frac{\partial v_3}{\partial t} = \frac{1}{\rho}\left(\frac{\partial \tau_{31}}{\partial x_1} + \frac{\partial \tau_{22}}{\partial x_2} + \frac{\partial \tau_{33}}{\partial x_3}\right)$$

$$\frac{\partial \tau_{11}}{\partial t} = \frac{\lambda+2\mu}{\rho}\frac{\partial v_1}{\partial x_1} + \frac{\lambda}{\rho}\frac{\partial v_2}{\partial x_2} + \frac{\lambda}{\rho}\frac{\partial v_3}{\partial x_3}$$

$$\frac{\partial \tau_{12}}{\partial t} = \frac{\mu}{\rho}\left(\frac{\partial v_1}{\partial x_2} + \frac{\partial v_2}{\partial x_1}\right)$$

$$\frac{\partial \tau_{13}}{\partial t} = \frac{\mu}{\rho}\left(\frac{\partial v_1}{\partial x_3} + \frac{\partial v_3}{\partial x_1}\right)$$

$$\frac{\partial \tau_{22}}{\partial t} = \frac{\lambda}{\rho}\frac{\partial v_1}{\partial x_1} + \frac{\lambda+2\mu}{\rho}\frac{\partial v_2}{\partial x_2} + \frac{\lambda}{\rho}\frac{\partial v_3}{\partial x_3}$$

$$\frac{\partial \tau_{23}}{\partial t} = \frac{\mu}{\rho}\left(\frac{\partial v_3}{\partial x_2} + \frac{\partial v_2}{\partial x_3}\right)$$

$$\frac{\partial \tau_{33}}{\partial t} = \frac{\lambda}{\rho}\frac{\partial v_1}{\partial x_1} + \frac{\lambda}{\rho}\frac{\partial v_2}{\partial x_2} + \frac{\lambda+2\mu}{\rho}\frac{\partial v_3}{\partial x_3} \tag{A.19}$$

Other kind of data can be modeled based on it's complexity.

- Vertical Transverse Isotropy (VTI): Where waves exhibit a symmetry around the vertical or depth axis like shale or thin bed behaviour. VTI is abit more complicated than the elastic isotropic case.

- Polar Isotropy Symmetry (TTI): Where the symmetry axis is tilted. due to the tectonic stresses, relative to the vertical axis orthogonal to the dip or rock in the simple case, Symmetry can be estimated by rotating the tensor c_{ijmn} of VTI model.

- Orthorhombic Isotropy Symmetry: which is more realistic case. It has three mutually orthogonal planes of symmetry result from combination of sedimenation transversely isotropic media (TI) and further effect of Tectonic forces that cause fractures result in another transversely isotropic media (TI) with a symmetry axis parallel to the stress direction and usually normal to the sedimentation TI (Song & Alkhalifah, 2013).

Appendix B   Elastic Reverse Time Migration (ERTM)

B.1   Introduction

Reflectors exist at points in the ground where the first arrival of downgoing wave is time coincident with the upgoing wave (Claerbout, 1971). RTM can be separated into two parts: wavefield reconstruction by elastic wave equation using the recorded vector data as boundary conditions followed by application of an imaging condition (IC) from (Yan & Sava, 2008). Elastic wave equation for isotropic elastic media (Aki & Richards, 2002) given by:

$$\rho \frac{\partial^2 u}{\partial^2} = f + (\lambda + 2\mu)\nabla(\nabla.u) - \mu\nabla \times \nabla \times u \tag{B.1}$$

Where u is vector displacement wavefield and f is body source force. The procedure is to reverse the recorded data in time, use the reversed data as sources along the recording surface to propagate the wavefields and apply the IC at zero time. Many literature have been published to study IC (Claerbout, 1971; McMechan, 1983; Baysal et al., 1983; Whitmore, 1983; Biondi & Chan, 2002) using zero-lag cross correlation or deconvolution of source and receiver (Back extrapolated) acoustic wavefields to map strong imaging energies. For Elastic RTM,(Yan & Sava, 2008) used wavefield decomposition using Helmholtz decomposition to separate the components prior to applying the ICs.

$$u = \nabla\phi + \nabla \times \psi \tag{B.2}$$

$$P = \nabla.u = \nabla^2\phi \tag{B.3}$$

$$S = \nabla \times u = -\nabla^2\psi \tag{B.4}$$

Where ϕ and ψ is the scaler and vector potential of extrapolated wavefield u. ϕ and ψ are obtained by applying div and curl of the wavefield u. For the isotropic elastic case(Jiang, Bancroft & Lines, 2012), simply imagining a simple point reflector, using Lamé parameter and density as modeling parameters and taking the cross correlation of source wavefield $S(x_1, x_3, t)$ and receiver wavefield $R(x_1, x_3, t)$,

one can have four ICs. Various ICs can be a simple approximation for reflection coefficients (RC) for various components.

$$I_{VV}(x_1, x_3) = \sum_t S_V(x_1, x_3, t) R_V(x_1, x_2, t)$$

$$I_{VH}(x_1, x_3) = \sum_t S_V(x_1, x_3, t) R_H(x_1, x_2, t)$$

$$I_{HV}(x_1, x_3) = \sum_t S_H(x_1, x_3, t) R_V(x_1, x_2, t)$$

$$I_{HH}(x_1, x_3) = \sum_t S_H(x_1, x_3, t) R_H(x_1, x_2, t) \tag{B.5}$$

But that doesn't correct for the effects of source nearby strong and weak energy. It's supposed that the source energy is very strong in the vertical component (strong source effect) of downgoing waves and horizontal component is very weak near the source(weak source effect). So, source normalized cross-correlation IC should be applied to eliminate source undesirable effects (Du, et al., 2012; Jiang, Bancroft & Lines, 2012; Whitmore and Lines, 1986; Kaelin & Guitton, 2006). Taking the above IC equations and divide it by sum of the zero lag autocorrelation of source components wavefield. And the result image, i.e. shot gather, can be estimated by stacking all the product of cross correlations snapshots.

$$I_{VV}(x_1, x_3) = \frac{\sum_t S_V(x_1, x_3, t) R_V(x_1, x_2, t)}{\sum_t (S_V^2(x_1, x_3, t + S_H^2(x_1, x_3, t))}$$

$$(I_{VH}(x_1, x_3) = \frac{\sum_t S_V(x_1, x_3, t) R_H(x_1, x_2, t)}{\sum_t S_V^2(x_1, x_3, t) + S_H^2(x_1, x_3, t))}$$

$$I_{HV}(x_1, x_3) = \frac{\sum_t S_H(x_1, x_3, t) R_V(x_1, x_2, t)}{\sum_t (S_V^2(x_1, x_3, t) + S_H^2(x_1, x_3, t))}$$

$$I_{HH}(x_1, x_3) = \frac{\sum_t S_H(x_1, x_3, t) R_H(x_1, x_2, t)}{\sum_t (S_V^2(x_1, x_3, t) + S_H^2(x_1, x_3, t))} \tag{B.6}$$
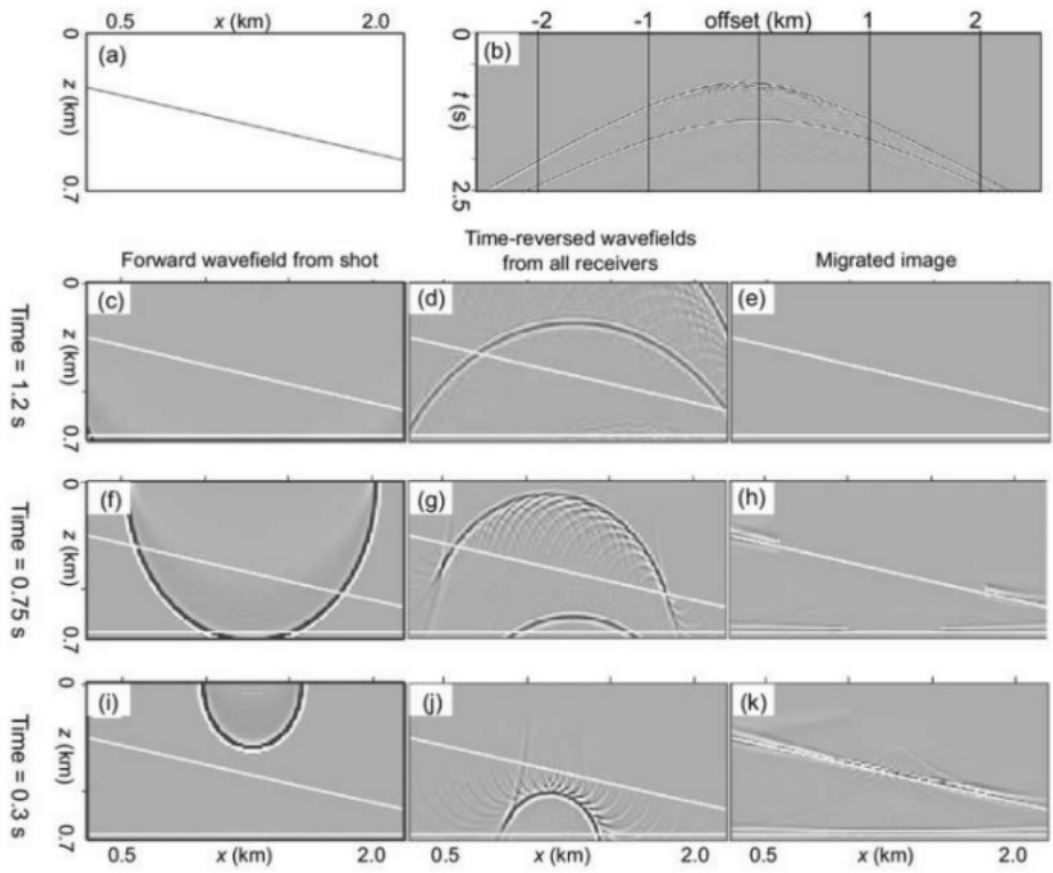
Figure B.1: Simple reflector reconstruction using RTM (Zhou, 2014).

Following the work of (Ibrahim & Weibull, 2017) using (Sava & Fomel, 2006; Yan, & Sava, 2008) extended elastic IC. The ICs, in the compact form, are:

$$IC_{PP} = \int C_P S_{ii}(x,t)R_{jj}(x,t)dt$$

$$IC_{PS} = \int (C_S S_{ij}(x,t)R_{ij}(x,t)dt - C_P S_{ii}(x,t)R_{jj}(x,t)dt) \tag{B.7}$$

Where $C_P = V_p^4$ and $C_S = V_s^4$ are scaling factors

Appendix C   Resolution

Understanding the spatial resolution is crucial part to optimize the use of FWI. In travel time tomography, fresnel zone maximum width is proportional to $\sqrt{\lambda L}$ where λ is the dominant wavelength and L is the source-receiver distance (Spetzler & Snieder, 2004). Considering a trace $d(g|s)$, where g is the geophone and s is the source, $\delta d(g|s)$ is given by the Born equation (Huang & Schuster, 2014):

$$\delta d(g|s) = \omega^2 \int_{\Omega} G(g|y) \Delta m G(y|s) \; dy^2 \qquad (C.1)$$

$$\delta d(g|s) = d(g|s)^{mod} - d(g|s)^{obs} \qquad (C.2)$$

Where $G(g)$ and $G(s)$ is the Green's function for the background model which is decomposed to direct and reflection (Figure C.1), $\Delta m(x)$ is perturbation model, ω is the angular frequency and Ω defines the integration points in the region that is being updated (yellow ray-path regions in figure C.1). The misfit gradient given by the adjoint modeling equation:

$$\Delta m = \omega^2 \int_{D} G(g|x)^* \delta d(g|s)^* G(x|s) dx_g dx_s \qquad (C.3)$$

While D is points of integration along sources and receivers. Combining the two equations gives:

$$\Delta m^{mig} = \omega^4 \int_{\Omega} \overbrace{G(g|x)^* G(x|s)^*}^{\text{migration kernel}} dx_g dx_s \int_{D} \overbrace{G(g|y) G(y|s)}^{\text{data kernel}} \Delta m \; dy^2 \qquad (C.4)$$

The gradient of misfit is also the migration of residuals. The migration kernel, Figure C.1, is to smear residuals along the yellow regions. For instance, migration kernel updates the low-wavenumber (high-wavenumber) portion of the velocity model (Mora P., 1989) if the model is to updated by smearing the residual along the rabbit ears (Figure C.1.b and c).
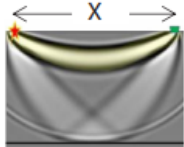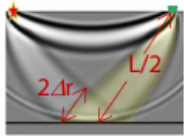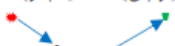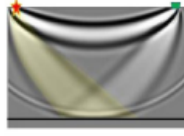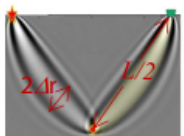
Figure C.1: Migration, data kernel, the associated wavepaths and resolution limit(after Huang & Schuster, 2014)

Appendix D   Configuration Files

D.1   3D Elastic Full Waveform Inversion

```
# MPI 3d elastic full-waveform inversion configuration file

# Modelling parameters
freesurface = "false"; # True if free surface should be on
order = "8"; # Order of finite difference stencil
lpml = "10"; # Size of pml absorbing boundary (should be larger than order + 5 )
snapinc = "1"; # Snap interval in multiples of modelling interval
apertx = "-250"; # Aperture for local model (source is in the middle)
aperty = "-250"; # Aperture for local model (source is in the middle)
source_type = "0"; # Dynamite

# Checkpointing parameters
snapmethod = "1"; # 0- Full checkpointing; 1- Optimal checkpointing
nsnaps = "8";
incore = "true";

#Fwi parameters
misfit_type = "1"; # 0- Difference; 1- Correlation 2-Adaptive
dataweightx = "false";
Dataweightxfile = "xweight.rss";
dataweighty = "false";
Dataweightyfile = "yweight.rss";
dataweightz = "false";
Dataweightzfile = "zweight.rss";
mute = "true"; # Mute gradient and updates
Mutefile = "mute.rss"; # File with mute weights
max_linesearch = "3"; # maximum number of linesearches
max_iterations = "5"; # maximum number of iterations
update_vp = "true"; # Update vp
update_vs = "true"; # Update vs
update_rho = "false"; # Update rho
update_source = "false"; # Update source
optmethod = "1"; # 1-L-BFGS; 2-CG_FR; 3-STEEPEST DESCENT; 4-CG_PR
linesearch = "3"; # 1-Decrease; 2-Armijo; 3-Wolfe; 4-Strong Wolfe
reciprocity = "true"; # Use receiver gathers instead of source gathers

# Diagonal scaling parameters
kvp = "1000.0";
kvs = "500.0";
krho = "50.0";
ksource = "1.0";

#Parameterisation
paramtype = "1"; # 0- grid; 1- B-spline;
dtx = "100.0"; # knot sampling in B-spline
dty = "100.0"; # knot sampling in B-spline
dtz = "100.0"; # knot sampling in B-spline

#Regularisation
vpregalpha = "1.0e-5";
vsregalpha = "1.0e-5";
rhoregalpha = "0.0e-5";

# Files
Vp = "vp.rss";
Vs = "vs.rss";
Rho = "rho.rss";
Wavelet = "wav.rss";
Uxrecordfile = "Hx.rss";
Uyrecordfile = "Hy.rss";
Uzrecordfile = "V.rss";
Snapfile = "Local/Snap.rss";
```

Figure D.1: Sample of 3D EFWI Configuration File

D.2 2D Elastic Full Waveform Inversion

```
# MPI 2d elastic full-waveform inversion configuration file

# Modelling parameters
freesurface = "false"; # True if free surface should be on
order = "4"; # Order of finite difference stencil
lpml = "8"; # Size of pml absorbing boundary (should be larger than order + 5 )
snapinc = "1"; # Snap interval in multiples of modelling interval
apertx = "7150"; # Aperture for local model (source is in the middle)
source_type = "0";

# Checkpointing parameters
snapmethod = "1"; # 0- Full checkpointing; 1- Optimal checkpointing
nsnaps = "15";
incore = "true";

#Fwi parameters

misfit_type = "1"; # 0- Difference; 1- Correlation; 2- Adaptive with Gaussian;  3- Adaptive with linear
dataweightx = "false";
dataweightz = "false";
Dataweightxfile = "weight.rss";
Dataweightzfile = "weight.rss";
mute = "true"; # Mute gradient and updates
Mutefile = "mute.rss"; # File with mute weights
max_linesearch = "5"; # maximum number of linesearches
max_iterations = "50"; # maximum number of iterations
optmethod = "1"; # 1-L-BFGS; 2-CG_FR; 3-STEEPEST DESCENT; 4-CG_PR
linesearch = "3"; # 1-Decrease; 2-Armijo; 3-Wolfe; 4-Strong Wolfe
update_vp = "true"; # Update vp
update_vs = "true"; # Update vs
update_rho = "false"; # Update rho
update_source = "false"; # Update source
reciprocity = "true"; # Use receiver gathers instead of source gathers


# Diagonal scaling parameters
kvp = "100000";
kvs = "50000";
krho = "1000.0";
ksource = "1.0e08";

#Parameterisation
paramtype = "1"; # 0- grid; 1- B-spline;
dtx = "25.0"; # knot sampling in B-spline
dtz = "25.0"; # knot sampling in B-spline

#Regularisation
vpregalpha = "0.0";
vsregalpha = "0.0";
rhoregalpha = "0.0";

# Files
Vp = "vp_vsp.rss";
Vs = "vs_vsp.rss";
Rho = "rho_vsp.rss";
Wavelet = "wavelet_estim.rss";
Uxrecordfile = "data-vx-reversed.rss";
Uzrecordfile = "data-vz.rss";
Snapfile = "Local/Snap.rss";
```

Figure D.2: Sample of 2D EFWI Configuration File

D.3 Reverse Time Migration

```
# MPI 2d elastic reverse-time migration configuration file

# Modelling parameters
freesurface = "false"; # True if free surface should be on
order = "8"; # Order of finite difference stencil
lpml = "18"; # Size of pml absorbing boundary (should be larger than order + 5 )
source_type = "0"; # Source type 0 - pressure. 1 for Vx. 3 for Vz.
snapinc = "4"; # Snap interval in multiples of modelling interval
apertx = "5000"; # Aperture for local model (source is in the middle)

# Checkpointing parameters
snapmethod = "1";
nsnaps = "40";
incore = "true";

# Migration parameters
nhx = "1";
nhz = "1";

# Booleans
Pimaging = "true"; # Set these to true if imaging of these events is to be made.
Pgather = "true"; # If surface gathers are to be output
Simaging = "true"; # Set these to true if imaging of these events is to be made.
Sgather = "true"; # If surface gathers are to be output

# Files
Vp = "vp_hybrid.rss";
Vs = "vs_hybrid.rss";
Rho = "rho_hybrid.rss";
Wavelet = "wavelet.rss";
Pimagefile = "Pimage2d.rss";
Pgatherfile = "Pgather2d.rss";
Sgatherfile = "Sgather2d.rss";
Simagefile = "Simage2d.rss";
Vxrecordfile = "H2-2d.rss";
Vzrecordfile = "V-2d.rss";
Snapfile = "snaps.rss";
```

Figure D.3: Sample of 2D ERTM Configuration File