University of Stavanger

**FACULTY OF SCIENCE AND TECHNOLOGY**

# MASTER'S THESIS

| | |
|---|---|
| Study programme/specialisation:<br>Petroleum Geosciences Engineering | Spring semester, 2018<br>Open |
| Author:<br>Geng Li | ……<br>(signature of author) |

Programme coordinator:

Supervisor(s): Wiktor Waldemar Weibull (UiS)

Xiaodong Luo (IRIS-Bergen)

Title of master's thesis:

**Data Assimilation Methods in Seismic Inversion**

Credits: 30

| | |
|---|---|
| Keywords:<br><br>Full-waveform inversion (FWI)<br>Data assimilation<br>Acoustic<br>Multiscale<br>Ensemble<br>Uncertainty | Number of pages: 45<br><br>+ supplemental material/other: N/A<br><br>Stavanger, June 15th, 2018 |

# Data Assimilation Methods in Seismic Inversion

by

Geng Li

**MSc. Thesis**

Presented to the Faculty of Science and Technology
The University of Stavanger

**The University of Stavanger**

2018

# Acknowledgments

I would first like to thank my supervisor Wiktor Waldemar Weibull at University of Stavanger and the co-supervisor Xiaodong Luo at IRIS in Bergen. They are always patient and support for the trouble spots or questions about my research and writing. Their consistent guidance allows this thesis to become true, and makes me grow so much both personally and professionally.

Furthermore, I am thankful to Dmitry Shogin who provides wonderful LaTeX introduction lectures. I acknowledge the support teams of Abel, Stallo, and Fram clusters for making it possible to operate the HPCs. I also would like to thank Andreas Habel for providing technical support.

I would like to express my gratitude to my classmates Mohamed Radwan, Signe Kristoffersen, and PhD candidate Karen Synnøve Ohm who give me a lot of supports and help.

Last but not the least, I am very grateful towards my family and friends both in Norway and China for their supports for making my thesis successful.

# Abstract

Full-waverform inversion (FWI) is a powerful tool for getting high resolution images from seismic data to understand formations and the reservoir characteristics. But the inversion problems usually do not have unique solutions which means several inversion results could fit the same true model. In this case, it is quite important to estimate the uncertainty of the inversion solutions and their accuracy. Moreover, the cycle skipping artifact is the main challenge of getting a robust FWI result.

In this thesis, the goal is to study on the FWI and introduce the multiscale method and the ensemble method which are based on the data assimilation method into FWI algorithm. The multiscale method will use different scale of offsets and travel times in iteration steps to prevent the cycle skipping artifacts, and the ensemble method will run multiple models to estimate the uncertainty of the FWI results. The application of this method is first implemented in two different scales of Marmousi model with 2D synthetic seismic data, and then applied on the 2D line NSR-31150R2, Sequence 124 seismic line from the North Sea.

It is shown that applying multiscale and ensemble methods in FWI have a satisfied result of estimating the uncertainty for the inversion solutions both in synthetic data and real field data. The cycle skipping artifacts are limited which makes the algorithm more stable.

# Table of Contents

# List of Figures

# 1 Introduction

The basic concept of geophysics is using the different physical fields to reveal the inner structures and properties of the Earth and its surrounding space environment. Seismic inversion is one of the most powerful geophysical methods used to understand the Earth's subsurface morphological properties and calculate the geophysical parameters from the observations of seismic surveys (Pereira, Nunes, Azevedo, Guerreiro, and Soares, 2017). These results can provide information about the structures and properties beneath the Earth's surface, such as lithology, porosity, saturation, rock mechanics, etc. These can be used to improve the accuracy of the reservoir model.

Since the modern inversion theory was first introduced by Backus and Gilbert in the 1960s (Backus and Gilbert, 1967), lots of methods of seismic inversion have been developed. Especially, in recent decades, methods of seismic inversion have become a powerful tool for exploration and are widely used in the petroleum industry. The targets of explorations, however, are getting more complicated than before. This means that the resolution and accuracy of the inversion play an important role in getting a better understanding of reservoirs. In this thesis, we will use the full-waveform inversion (FWI) which was introduced in 1980s (Tarantola, 1984) and has been getting more popular in recent years. In FWI method, the full-wave equation modeling is performed at each iteration of the optimization using the model of the previous iteration (Virieux and Operto, 2009). All types of waves are involved in the optimization, including diving waves, supercritical reflections and multi-scattered waves, such as free surface related multiples (Virieux and Operto, 2009).

In seismic exploration, the fundamental task is using the measured data and parameters such as velocity, density, etc., to study or extract the physical characters of the formations, and aid in the prospect or reservoir evaluation. Usually, this can be approached by forward modeling or inversion. The forward problem is using the wave equation and a physical model to predict the outcome of measurements, so it has a unique solution. Unlike the forward problem, inversion problems are using the observed seismic waves to build a model and match the actual data (Tarantola, 2005). So, as an inversion problem, seismic inversion does not have unique solutions, and it's necessary to take all the parameters which can influence the solution (such as noise and their uncertainties) into consideration. In this project, the data assimilation method is tested to estimate the uncertainties in the models derived from FWI with the theory of data assimilation.

# 2 Full Waveform Inversion

In this section, a review about the theoretical backgrounds of FWI is given. FWI uses the forward-modeling and the iterative nonlinear optimization to solve the inversion problem. A representative iteration step in FWI procedure includes solving the wave equation, calculating the misfit between the observed and predicted data and applying an optimization method to update the model.

## 2.1 Forward Modeling

The FWI can be implemented both in the time domain (Tarantola, 1984) and frequency domain (Pratt, 1999). In FWI, the objective is to estimate a velocity model which can best fit the observed seismic data. The fundamental step of this process is the forward modeling. By using forward modeling, the simulated seismic data from a starting velocity model can be calculated, and subsequently the misfit between this seismic data and the observed data at the receiver positions. The nonlinear iteration optimization method is implemented to minimize the misfit and reach the final inversion model that includes the geophysical parameters.

The typical forward modeling problem means to forecast the observation data $\mathbf{d}_{cal}$ from the giving starting model $\mathbf{m}$ by using the forward operator $\mathbf{G}$. Following Tarantola (2005), the expression can be written:

$$\mathbf{d}_{cal} = \mathbf{G}(\mathbf{m}) \tag{2.1.1}$$

### 2.1.1 Wave Equation

In this section, the two dimension acoustic wave equation will be reviewed. Solving the acoustic wave equation, one can calculate the wavefield at any time and any position. In an acoustic medium with variable velocity and density, the two dimension acoustic wave equation with source presentation in the time domain can be written by Equation 2.1.2 (Kosloff(1983) and Jensen(2011) ).

$$\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial p(\mathbf{x},t)}{\partial x}\right)+\frac{\partial}{\partial z}\left(\frac{1}{\rho}\frac{\partial p(\mathbf{x},t)}{\partial z}\right)-\frac{1}{c^2(\mathbf{x})\rho}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2}=s(\mathbf{x}_s,t) \tag{2.1.2}$$

Where $p(\mathbf{x},t)$ is the acoustic pressure at position $\mathbf{x}$ and time $t$, in 2D dimension, $\mathbf{x}$ represents the Cartesian coordinate vector of $(x,z)$ in 2D space. $\rho$ is the density at $\mathbf{x}$. $c(\mathbf{x})$ is the

acoustic velocity at position $\mathbf{x}$ in 2D space. $s(\mathbf{x}_s, t)$ is the source position at point $\mathbf{x}_s$ and time $t$.

If ignore the influence of density change or assume the density has a negligible impact on wavefield, the density in the wave equation will not change in the 2D space and become constant. The acoustic wave equation can be written in the constant density form as Equation 2.1.3.

$$\nabla^2 p - \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2} = s(\mathbf{x}_s, t) \tag{2.1.3}$$

Where $\nabla$ is the Laplace operator. $p$ and $c$ is the same as $p(\mathbf{x}, t)$ and $c(\mathbf{x})$ in Equation 2.1.2 respectively.

Applying the Fourier transformation on Equation 2.1.3, the wave equation in frequency domain can be derived, which is called Helmholtz equation (Eq. **??**):

$$\nabla^2 p(\mathbf{x}, \omega) - \frac{\omega^2}{c^2(\mathbf{x})} p(\mathbf{x}, \omega) = s(\mathbf{x}_s, \omega) \tag{2.1.4}$$

Where $\omega$ denotes the angular frequency. $p(\mathbf{x}, \omega)$ and $s(\mathbf{x}_s, \omega)$ represent the pressure field and the source, respectively in the frequency domain.

### 2.1.2 Absorbing Boundary Conditions

To solve the wave equation and calculate the wavefield, the initial conditions and boundary conditions of the partial differential equation should be given. The initial condition should be the state of the entire system at time $t = 0$, which means the two distributions, initial pressure field $p(\mathbf{x}, 0)$ and initial term of partial derivative in time $\frac{\partial^2 p(\mathbf{x},0)}{\partial t^2}$, need to be known.

In addition, to solve the acoustic wave equation, the reflection at the boundaries should be taken into consideration which is called *boundary conditions*. Because this kind of reflection can add plenty of noise and cause some problems for computing the wave propagation in the limit domain. Here, we introduce the absorbing boundary conditions. The main idea of this boundary condition is to truncate infinite domains to make it proper for performing the computation of wave propagation in the medium. The boundary is no longer a barrier for the wave and it can cross the truncated domain as if the wave propagated in an infinite domain. The absorbing boundary condition can create a result that is close to the solution took from a large open space. In this thesis, the Perfect Matched Layers (PML) absorbing boundary conditions (Berenger, 1994) is used as the boundary condition.

### 2.1.3 Solution of Wave Equation

There are several numerical methods applying discretization on the wave equations attempt to solve them with associated initial and boundary conditions. These methods can be implemented on different forms of the wave equation. The Finite Difference Method (FDM) (Jean Virieux, 1986) is the most direct approach among these methods (Jensen et al., 2011). It directly discretizes the continuous values into grid points to calculate partial derivatives through the differential operators. Other two methods that are commonly used are the Finite Element Method (FEM) (Fichtner, 2011) and Finite Volume Method (FVM) (LeVeque, 2002). In this thesis, the FDM will be used to solve the wave equation for the forward modeling.

For one-dimension problems, the derivation of the second order numerical derivative follows Jensen(2011). The solution function $f$ close to the point $x$ can be implemented Taylor series as:

$$f(x+h) = f(x) + hf'(x) + \frac{1}{2}h^2 f''(x) + \frac{1}{6}h^3 f'''(x) + \cdots \tag{2.1.5}$$

$$f(x-h) = f(x) - hf'(x) + \frac{1}{2}h^2 f''(x) - \frac{1}{6}h^3 f'''(x) + \cdots \tag{2.1.6}$$

Addition of (2.1.5) and (2.1.6), and take the second order central finite difference. The approximation of the second derivative is obtained,

$$f''(x) = \frac{d^2 f}{dx^2} = \frac{f(x-h) - 2f(x) + f(x+h)}{h^2} \tag{2.1.7}$$

Adding another term, the equation can be expand into 2D,

$$f''(x,y) = \frac{f(x-h,y) - 2f(x,y) + f(x+h,y)}{h^2} + \frac{f(x,y-h) - 2f(x,y) + f(x,y+h)}{h^2} \tag{2.1.8}$$

Then this second derivative can replace the second derivative of pressure $p$ in the wave equation. (Tengesdal, 2013)

$$\frac{p_{i,j}^{n+1} - 2p_{i,j}^{n} + p_{i,j}^{n-1}}{\Delta t^2 v^2} - \nabla^2 p_{i,j}^{n} = 0 \tag{2.1.9}$$

then the pressure field can be derived from this equation 2.1.9

$$p_{i,j}^{n+1} = 2p_{i,j}^{n} - p_{i,j}^{n-1} + \Delta t^2 v^2 \nabla^2 p_{i,j}^{n} \tag{2.1.10}$$

where

$$\nabla^2 p = \frac{p_{i+1,j}^{n} - 2p_{i,j}^{n} + p_{i-1,j}^{n}}{\Delta x} + \frac{p_{i,j+1}^{n} - 2p_{i,j}^{n} + p_{i,j-1}^{n}}{\Delta z} \tag{2.1.11}$$

and the $p_{i,j}^{n+1}$, $p_{i,j}^{n}$, $p_{i,j}^{n-1}$ are the pressure field in time steps $n+1$, $n$, $n-1$, $\Delta t$ is the sampling frequency in time. $\Delta x$, $\Delta z$ are the mesh sizes in $x$ and $z$ direction.

### 2.1.4 Stability and Accuracy

The accuracy of finite difference methods contains two components which are linked to each other. When the discretization has been implemented, in a certain order and mesh size, this algorithm introduces an error which is between the approximate solution of the discrete equation and the original differential equation. This error is known as the discretization error ($e_d$) (Jensen et al., 2011). When the mesh size tends to be zero, the solution of the finite difference method approaches an analytical result, in other word, the $e_d$ approaches zero, can be called convergent.

Another error is the numerical error. Because of the discretization of the quation, there is limited numerical precision in the calculation compared to the analytical solution. This will introduce the second error which is called stability error or rounding error ($e_s$) (Jensen et al., 2011). This error could grow magnificently and not remain bound as computation goes on. In this case, the algorithm of the finite difference method is not stable. To get stability, the application of the finite difference method in the wave propagation should be both convergent and stable.

Clearly, the accuracy of the finite difference solution depends on the total error of the solution, and it should be the sum of the two errors (Jensen et al., 2011).

$$e = e_d + e_s \tag{2.1.12}$$

## 2.2 The Theory of FWI

### 2.2.1 Inversion

As mentioned in chapter 2.1, the forward modeling is predicting the observation form the giving starting model by using the forward operator. In the reality, the seismic survey has a similar concept as the forward modeling which is

$$\mathbf{d}_{obs} = \mathbf{g}(\mathbf{m_m}) \tag{2.2.1}$$

where the $\mathbf{d}_{obs}$ presents the observation data, the $\mathbf{g}$ indicates the physical laws and the $\mathbf{m_m}$ is the physical model of the medium.

The inversion means to calculate the unknown physical parameters of the medium by using the observation data and physical laws. But unlike the forward modeling, the inversion problem has several difficulties. The most common issue of the problem is that there may be abundant of models which satisfy the same observation data. The uniqueness makes it difficult to decide how accurate one solution is. In addition, the presence of noise is also an obstacle of using the forward modeling solution to solve the inversion problem. Furthermore, the slight change of the observation data may dramatically change the solution of inversion. This kind of ill-conditioned problem makes it more difficult to get the solution and evaluate its accuracy.

### 2.2.2 Objective Function of Optimization

The objective function of FWI is a measure of how well the calculated data match the observed data for a physical medium model $\mathbf{m_m}$ the best. To achieve this, it defines the misfit computing between the forward modeling data and the observation data of seismic acquisition and tries to minimize this misfit. The misfit vector is defined as

$$\mathbf{\Delta d} = \mathbf{d}_{obs} - \mathbf{d}_{cal} \tag{2.2.2}$$

where the $\mathbf{d}_{obs}$ is the observation date getting from the seismic survey, the $\mathbf{d}_{cal}$ is the synthetic data simulated from forward modeling.

The most used objective function is the least squares norm (2 norm)which is given by

$$E(\mathbf{m}) = \frac{1}{2}\left\|\mathbf{d}_{obs} - \mathbf{d}_{cal}\right\|^2 = \frac{1}{2}\mathbf{\Delta d}^{\mathrm{T}}\mathbf{\Delta d}^{*} \tag{2.2.3}$$

where $\mathbf{\Delta d}$ is the residual of wavefield, $T$ represents the transposition operator, $*$ denotes the conjugate operator.

Taking the derivative with respect to one of the parameter $m_l$ in the objective function $E(\mathbf{m})$ gives

$$\begin{aligned}\frac{\partial E(\mathbf{m})}{\partial m_l} &= -\frac{1}{2}\sum_{i=1}^{N}\left[\frac{\partial d_{\mathrm{cal}i}}{\partial m_l}\left(d_{\mathrm{obs}i} - d_{\mathrm{cal}i}\right)^{*} + \left(d_{\mathrm{obs}i} - d_{\mathrm{cal}i}\right)\frac{\partial d_{\mathrm{cal}i}^{*}}{\partial m_l}\right] \\ &= -\sum_{i=1}^{N}\Re\left[\left(\frac{\partial d_{\mathrm{cal}i}}{\partial m_l}\right)^{*}\left(d_{\mathrm{obs}i} - d_{\mathrm{cal}i}\right)\right]\end{aligned} \tag{2.2.4}$$

where $\Re$ and $*$ denote the real part and the conjugate of a complex number, respectively.

### 2.2.3 The Gradient and Hessian Matrix in FWI

The gradient of the objective function (or misfit function, Eq. 2.2.3) indicates the fastest increasing direction of the misfit. The solution can be reached after several iterations by taking the opposite direction to reduce the misfit.

From the definition of gradient in the objective function, we can take the derivative of the equation 2.2.4 to get the gradient direction

$$\begin{aligned}\nabla E_{\mathbf{m}} = \frac{\partial E(\mathbf{m})}{\partial \mathbf{m}} &= -\Re\left[\left(\frac{\partial \mathbf{d}_{\mathrm{cal}}}{\partial \mathbf{m}}\right)^{\mathrm{T}}\left(\mathbf{d}_{\mathrm{obs}i} - \mathbf{d}_{\mathrm{cal}}\right)^{*}\right] \\ &= -\Re\left[\mathbf{J}^{\mathrm{T}}\mathbf{\Delta d}^{*}\right]\end{aligned} \tag{2.2.5}$$

where $\mathbf{J}$ is the Jacobian or the $N \times M$ Frechet derivative matrix, the elements of which are given by

$$J_{ij} = \frac{\partial p_i}{\partial m_j}, \qquad i = (1, 2, ..., N); j = (1, 2, ..., M) \tag{2.2.6}$$

And the expression in matrix form for the Hessian is written as

$$\frac{\partial^2 E(\mathbf{m})}{\partial \mathbf{m}^2} = -\Re\left[\mathbf{J}^{\mathrm{T}}\mathbf{J}^*\right] + \Re\left[\frac{\partial \mathbf{J}^{\mathrm{T}}}{\partial \mathbf{m}}(\mathbf{\Delta d}...\mathbf{\Delta d})\right] \tag{2.2.7}$$

To present all the node points, the additional terms are added into the matrix $\mathbf{J}$ and a new $M \times L$ matrix $\hat{\mathbf{J}}$ is obtained, the element of which is given by

$$\hat{J_{ij}} = \frac{\partial p_i}{\partial m_j}, \qquad i = (1, 2, ..., M); j = (1, 2, ..., L) \tag{2.2.8}$$

Take the new elements into equation 2.2.5, a new equation is yielded:

$$\nabla E_{\mathbf{m}} = -\Re\left[\hat{\mathbf{J}}^{\mathrm{T}}\mathbf{\Delta}\hat{\mathbf{d}}^*\right] \tag{2.2.9}$$

where $\Delta\hat{d}$ is the data residual vector which is caused by adding new terms in the original residual vector. We add $(q-k)$ zero values to generate the new vector of length $q$. The equation 2.2.9 is written

$$\begin{aligned}
\begin{bmatrix} \frac{\partial E}{\partial m_1} \\ \frac{\partial E}{\partial m_2} \\ \vdots \\ \frac{\partial E}{\partial m_M} \end{bmatrix} &= \left\{ \begin{bmatrix} \frac{\partial p_1}{\partial m_1} & \cdots & \frac{\partial p_N}{\partial m_1} & \frac{\partial p_{N+1}}{\partial m_1} & \cdots & \frac{\partial p_L}{\partial m_1} \\ \frac{\partial p_1}{\partial m_2} & \cdots & \frac{\partial p_N}{\partial m_2} & \frac{\partial p_{N+1}}{\partial m_2} & \cdots & \frac{\partial p_L}{\partial m_2} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial p_1}{\partial m_M} & \cdots & \frac{\partial p_N}{\partial m_M} & \frac{\partial p_{N+1}}{\partial m_M} & \cdots & \frac{\partial p_L}{\partial m_M} \end{bmatrix} \begin{bmatrix} \Delta d_1^* \\ \vdots \\ \Delta d_N^* \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\} \\
&= \Re\left\{ \begin{bmatrix} \frac{\partial \mathbf{p}^{\mathrm{T}}}{\partial m_1} \\ \frac{\partial \mathbf{p}^{\mathrm{T}}}{\partial m_2} \\ \vdots \\ \frac{\partial \mathbf{p}^{\mathrm{T}}}{\partial m_M} \end{bmatrix} \begin{bmatrix} \Delta d_1^* \\ \vdots \\ \Delta d_N^* \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}
\end{aligned} \tag{2.2.10}$$

From equation 2.2.9, the gradient can be calculated when the derivative of Frechet matrix and residual of the wavefield are known. The solution of the derivative of Frechet will be discussed in the following paragraph.

Take the derivative of the matrix form wave equation $\mathbf{Bu} = \mathbf{s}$ with respect to the $i$th parameter (Pratt, Shin, and Hick, 1998):

$$\frac{\partial \mathbf{B}}{\partial m_i}\mathbf{u} + \mathbf{B}\frac{\partial \mathbf{u}}{\partial m_i} = 0 \tag{2.2.11}$$

where $\mathbf{B}$ denotes the impedance matrix (Marfurt, 1984), $\mathbf{u}$ presents the seismic wavefield and $\mathbf{s}$ is the source term.

The $i$th virtual source term is introduced as:

$$\mathbf{s}^{(i)} = -\frac{\partial \mathbf{B}}{\partial m_i}\mathbf{u} \tag{2.2.12}$$

Then the equation 2.2.11 can be transformed into

$$\mathbf{B}\frac{\partial \mathbf{u}}{\partial m_i} = \mathbf{s}^{(i)} \tag{2.2.13}$$

For all node points of every parameter, they should have a virtual source for themselves. The virtual source should be an $L \times L$ vector. By solving the equation 2.2.13, the partial derivative wavefield of the $i$th node can be obtained. Thus, the partial derivative wavefield of the entire model or the extended Frechet matrix is in the form of expression

$$\hat{\mathbf{J}} = \begin{bmatrix} \frac{\partial \mathbf{u}}{\partial m_1} & \frac{\partial \mathbf{u}}{\partial m_2} & \cdots & \frac{\partial \mathbf{u}}{\partial m_M} \end{bmatrix} = \mathbf{B}^{-1}\begin{bmatrix} \mathbf{s}^{(1)} & \mathbf{s}^{(2)} & \cdots & \mathbf{s}^{(M)} \end{bmatrix} \tag{2.2.14}$$

or

$$\hat{\mathbf{J}} = \mathbf{B}^{-1}\mathbf{S} \tag{2.2.15}$$

where $\mathbf{S} = \begin{bmatrix} \mathbf{s}^{(1)} & \mathbf{s}^{(2)} & \cdots & \mathbf{s}^{(M)} \end{bmatrix}$, is an $L \times M$ matrix.

From the equation 2.2.5 we can get the solution of gradient depends on the derivative of Frechet and the residuals of the wavefield. The residual of the wavefield is calculated from the result of forward modeling, and we just reviewed the solution of Frechet matrix. In this

case, the gradient is obtained.

In Newton methods of inversion, the main concept is expanding the objective function in equation 2.2.3 as a Taylor series and retaining terms up to quadratic order (Tarantola, 1987):

$$E(\mathbf{m}+\delta\mathbf{m}) = E(\mathbf{m}) + \delta\mathbf{m}^{\mathrm{T}}\nabla E(\mathbf{m}) + \frac{1}{2}\delta\mathbf{m}^{\mathrm{T}}\mathbf{H}\delta\mathbf{m} + O(|\delta\mathbf{m}|^3) \tag{2.2.16}$$

where $\mathbf{m}$ denotes the parameters in the objective function. $\delta\mathbf{m}$ is the parameters with a tiny perturbation. $\mathbf{H}$ represents the $m \times m$ second-derivative objective function Hessian matrix, the elements of which are given by

$$H_{ij} = \frac{\partial E(\mathbf{m})}{\partial m_i \partial m_j} \quad i,j = (1,2,...,m) \tag{2.2.17}$$

Then the Newton method for iterative solution can be derived by using the vector $\delta\mathbf{m} = -\mathbf{H}^{-1}\nabla E$ which is minimized within the quadratic approximation.

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} - \mathbf{H}^{-1}\nabla E \tag{2.2.18}$$

In quasi-Newton method, the Hessian matrix is replaced by the finite approximation of the Hessian. The most used method of computing the approximation of Hessian and its inverse is the BFGS algorithm (named by the inventors Broyden, Flethcer, Goldfarb, and Shanno). The main idea is to use the approximation of Hessian $\mathbf{H}^{(k)}$ and the gradient $\nabla E^{(k)}$ at iteration $k$ to compute the Hessian at the next iteration. But for the FWI problem, solving the approximation Hessian and its inverse requires a huge storage space and computing time. To reduce the storage, Nocedal (1980) introduced a limited storage variants of quasi-Newton BFGS method known as the L-BFGS algorithm. Instead of computing the Hessian matrix directly, it uses updated gradient and velocity from several most recent iterations (typically, 3-20 iterations)to build up the Hessian matrix. In this case, the Hessian doesn't need to be stored. The only information needs to be stored is the iteration length and its computing results. If we set the first guess as $H^{(0)}$, the L-BFGS approximation $H_k$ follows Nocedal (1999) is given by:

$$
\begin{aligned}
H_k =&(V_{k-1}^T \cdots V_{k-m}^T) H^{(0)} (V_{k-m} \cdots V_{k-1}) \\
&+ \rho_{k-m} (V_{k-1}^T \cdots V_{k-m+1}^T) s_{k-m} s_{k-m}^T (V_{k-m+1} \cdots V_{k-1}) \\
&+ \rho_{k-m+1} (V_{k-1}^T \cdots V_{k-m+2}^T) s_{k-m+1} s_{k-m+1}^T (V_{k-m+2} \cdots V_{k-1}) \\
&+ \cdots \\
&+ \rho_{k-1} s_{k-1} s_{k-1}^T
\end{aligned} \tag{2.2.19}
$$

where

$$
\rho_k = \frac{1}{y_k^T s_k}, \quad V_k = I - \rho_k y_k s_k^T, \tag{2.2.20}
$$

and

$$
s_k = m_{k+1} - m_k, \quad y_k = \nabla E_{k+1} - \nabla E_k. \tag{2.2.21}
$$

With the $m+1$ vector $\{s_i, y_j\}$ in storage, the $H_k \nabla E_k$ is computed by using this expression in a recursive procedure.

### 2.2.4 The Work-flow of FWI

The workflow of full waveform inversion is shown as follows:

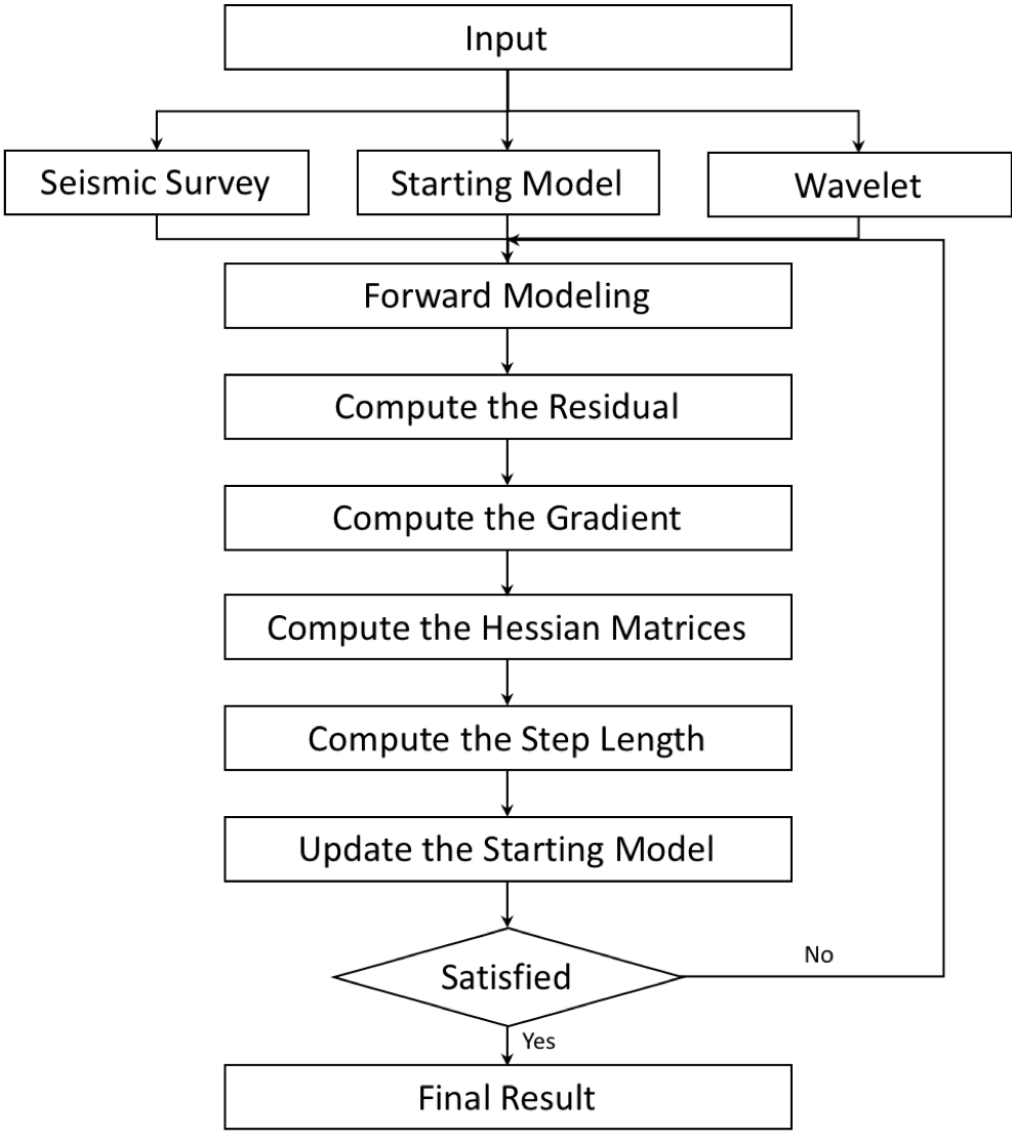



Figure 1: The work-flow of full waveform inversion

# 3 Applications of FWI in Marmousi model

## 3.1 Outline

The main objective of this thesis is to analyze the accuracy and uncertainty for the results full waveform inversion, and to see the feasibility of the method by applying it to a real field data set. The model used to investigate the method in this thesis is the Marmousi model (Figure 2a). Instead of using the typical FWI workflow, the multiscale FWI is used in this thesis. To test the workflow, the application starts with a part of Marmousi model, and then investigate the entire model. And to study the uncertainties and accuracy of inversion results, a series of starting models has to be set up. In the following sections, the key parameters and workflow will be discussed, such as the Marmousi model, setting up the starting models, the multiscale FWI procedure, and the final results of partial and whole Marmousi model.

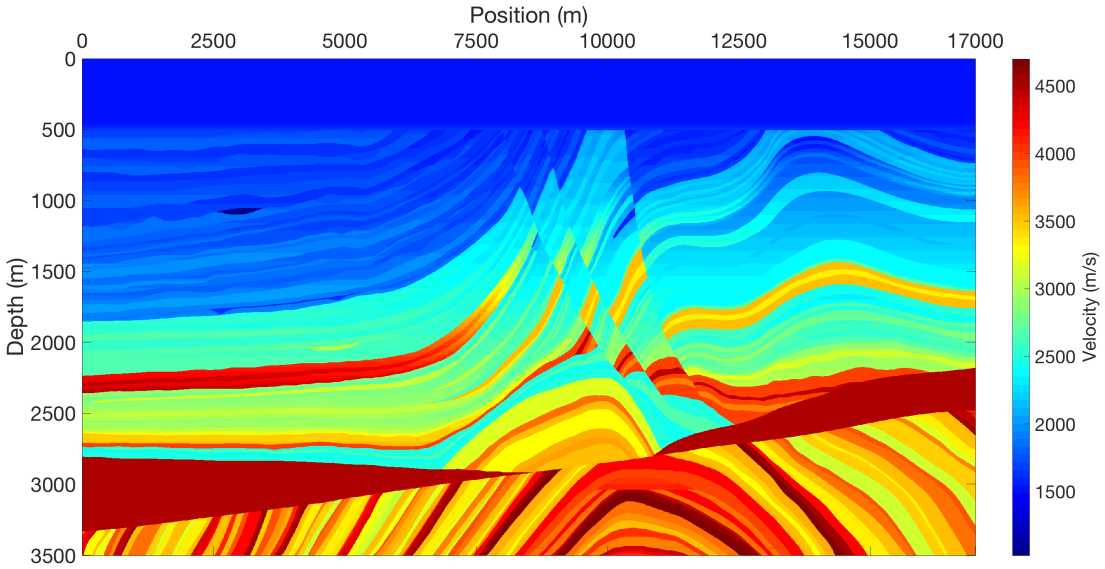
## 3.2 Model

### 3.2.1 Marmousi Model

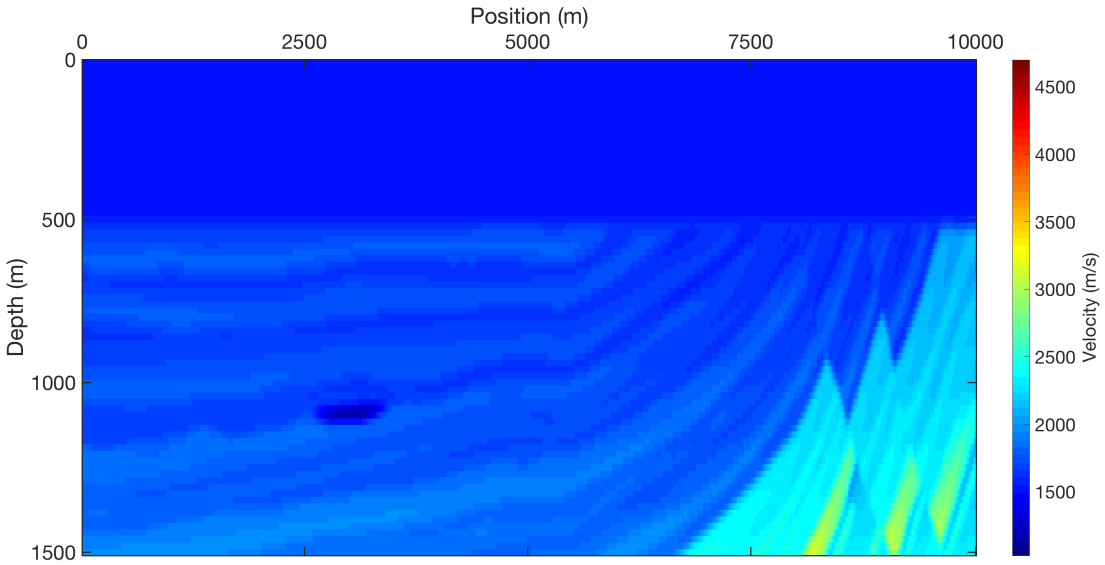
The Marmousi model was created in 1988 by the Geophysics Group in Institut Français du Pétrole (IFP). In 1990, the Marmousi dataset was created and used for the workshop on practical aspects of seismic data inversion at the 52nd EAEG meeting in Copenhagen (Versteeg, 1994). The structure of this model is based on a profile through the North Quenguela trough in the Cuanza basin, Angola. It has significant changes of velocities in both vertical and lateral. The complexity of geometry and velocity model requires advanced velocity estimation method and inversion techniques to reveal a good quality geophysical model of the earth.

The Marmousi model which is 17000 m in width and 3500 m in depth and contains 158 horizontally layered horizons is used in this thesis (Figure 2a). A series of normal faults and resulting tilted blocks complicates the model towards its center. The water depth of the model is approximately 500 m.

The part of Marmousi model used to test the workflow of multiscale FWI is illustrated in Figure 2b. It is taken from the upper left corner of the Marmousi model. This part contains part of the series normal faults and low velocity gas indicator. If the workflow worked well in this small part, the method can be applied to the entire model.




**(a)**




**(b)**

Figure 2: The Marmousi model. (a) the real entire Marmousi velocity model, (b) the upper left part of the Marmousi model to test the workflow

### 3.2.2 Synthetic Seismic Data

As mentioned in Chapter 2.1, the seismology is simulated by the forward modeling method with the physical model of the formation. The partial Marmousi model shown in Figure 2b is implemented for this investigation. The model consists of $501 \times 76$ grid-cells in the horizontal and vertical directions, respectively. Both the horizontal and vertical grid spacings are 20 m. There are 246 receivers evenly located on the ocean bed cable with fixed-spread acquisition geometry. The 32 sources are located at starting point (0,0) with the interval of 250 m in x direction and at the water depth of 20 m. The wavelet is modeled by the Ricker wavelet with the dominant frequency of 20 Hz (Figure 3). The recording length is 5 s and sampling interval is 1 ms. The synthetic seismic data is shown as Figure 4. The bandpass filter of 7.5 Hz is applied after the synthetic data has been created.

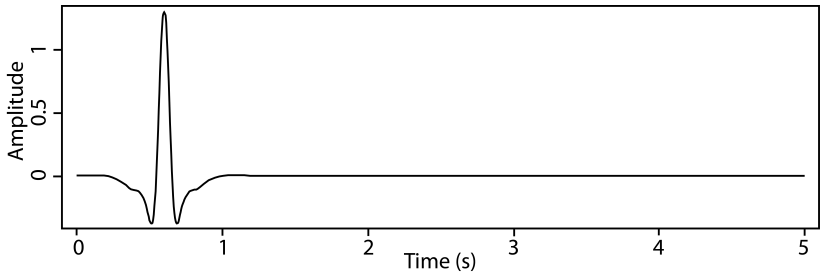



Figure 3: The Ricker wavelet which is used to create the synthetic seismic data

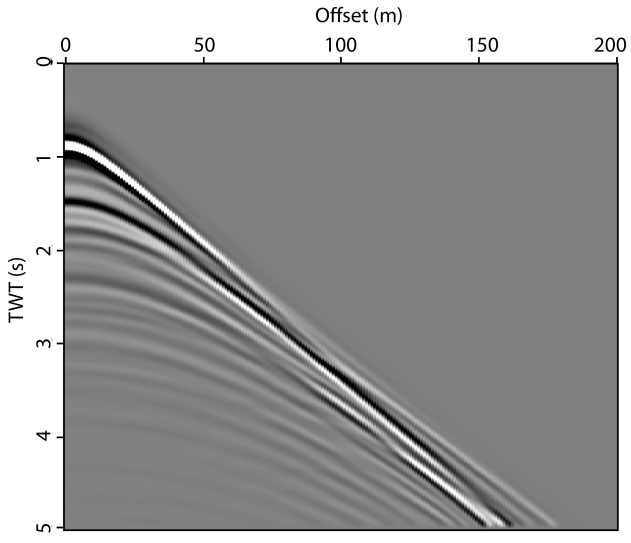



Figure 4: The first shot of synthetic data

### 3.2.3 Starting Models

To create the initial velocity model for the FWI, the triangle filter smoothing is implemented in the real Marmousi velocity model. The starting model created by the triangle filter is shown as Figure 5. The idea of triangle filtering is a rectangle smoothing done twice. The triangle smoother is an approximation of Gaussian smoother. Repeating the triangle smoother will yield a result which rapidly approaches the result of Gaussian function. (Claerbout, 2014) The expression of triangle smoothing in radius is given by

$$T_k(Z) = B_k(\frac{1}{Z})B_k(Z) \tag{3.2.1}$$

where

$$B_k(Z) = \frac{1}{k}\left(1 + Z + Z^2 + \cdots + Z^k\right) = \frac{1}{k}\frac{1 - Z^{k+1}}{1 - Z} \tag{3.2.2}$$

and $k$ is the smoothing radius.

In this case, the smoothing radius $k$ is 30 in both direction for the initial velocity model.

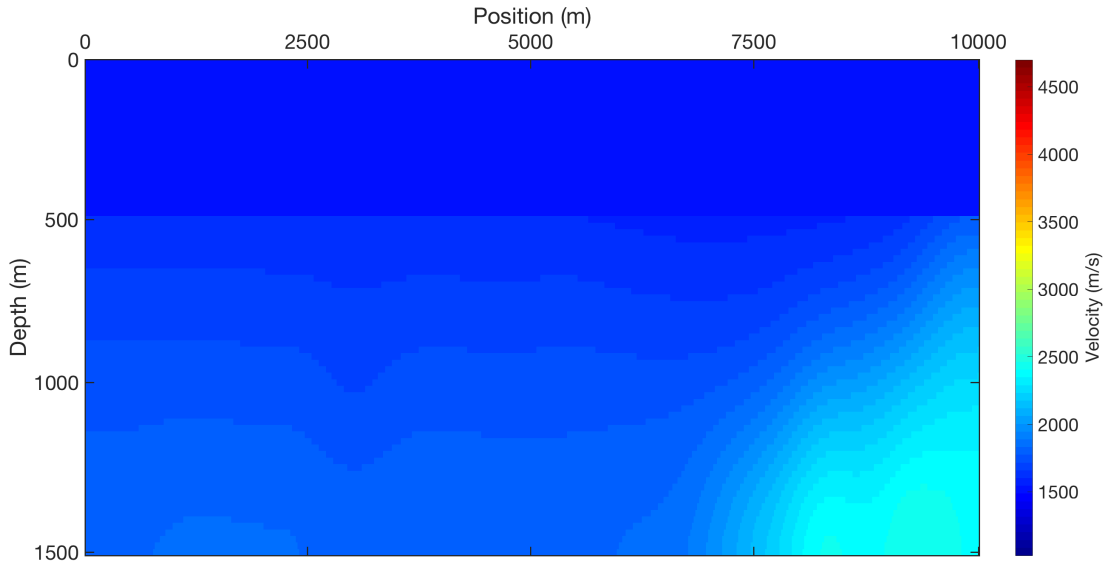



Figure 5: The starting model for FWI.

To estimate the uncertainty and accuracy of FWI, the initial model should be generated into an ensemble, in another word, the initial model should include a series models that

have nearly the same inversion result. To generate the ensemble, there are three methods to implement the procedure: (1) adding perturbation to the initial model, it is implemented either by Mento Carlo method or random perturbation, (2) sensitivity analysis, (3) multi-model ensemble, this changes the model formulation. In this step, the random perturbations are added to the initial models to yield the ensemble. The random perturbation has a uniformly distribution of numbers in the interval between -2000 and 2000 which has also been smoothed by the moving average filter with the radius of 30. The perturbation only implements on the formation not on the water layer. The distribution of perturbation is shown as Figure 6b.

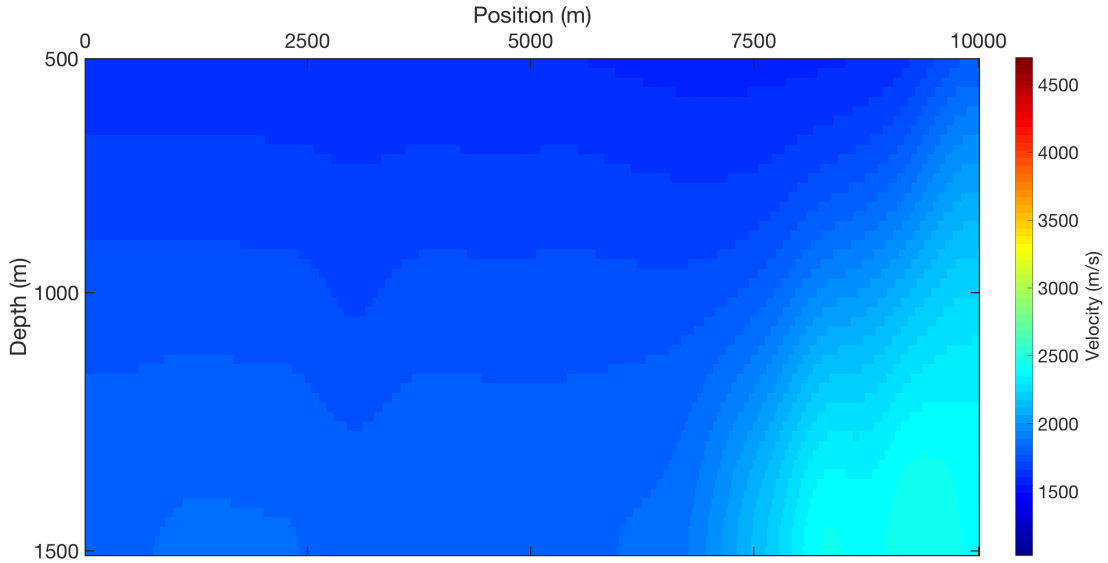
Then the perturbation is added to the initial starting model to generate one member of the ensemble which is illustrated as Figure 7. Repeating this random perturbation method 20 times generates the ensemble with 20 members. That is the starting model for the multiscale FWI.

## 3.3 Multiscale Iteration

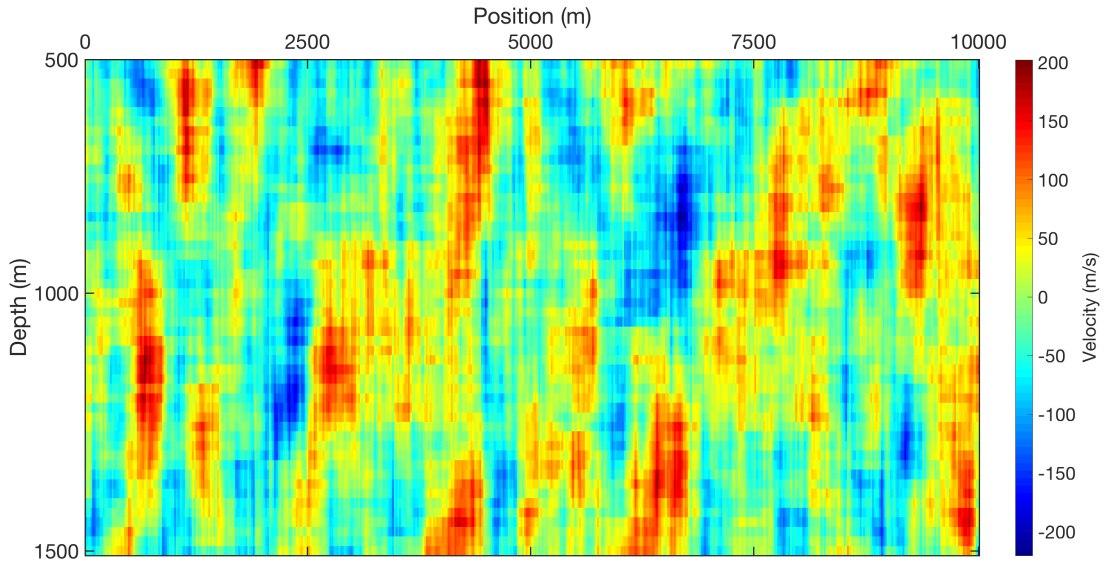
To solve the nonlinear FWI problem, in recent decades, there are many studies to develop some multiscale strategies to accelerate the convergence, alleviate the influence of local minimum. The multiscale strategy successively subdivides the inversion data set into different scales and tries to solve the objective functions based on these scales. Finally, the method can reach the true global solution.

Usually, the starting model is not close enough to the true model. The objective function may converge to a local minimum $v_L$ instead of the true solution $v_T$, as shown in Figure 8 (Hu, Chen, Liu, and Abubakar, 2018). This phenomenon is usually called *cycle skipping* artifacts which are the main challenge of FWI applications. This problem can be prevented by using the lower frequency to perform the FWI because the phase mismatch can be limited in half period. The short offset and time, which has limit wavenumber and time difference, can also prevent the cycle skipping. In the practical seismic exploration, the low frequency data is commonly unavailable. And in some cases, if the starting model is really poor estimated, there may be also some cycle skipping appearance.

To reduce the influence of cycle skipping artifacts, the multiscale iteration is an efficient method in the FWI workflow. In the multiscale method, the seismic data is subdivided into small scales in both offset and time. With the iterations going on, the offset and time are increasing at each step with a fixed step length. The starting model of each step is also updated by using the FWI result from the previous iteration. In this situation, the FWI will be implemented on a smaller scale which contains less wavenumber and has a low possibility




**(a)**




**(b)**

Figure 6: The distribution of random perturbation that is added to the initial model without the water layer. (a) the initial starting model without water layer, (b) one of the random perturbation distribution in the same area.

of mismatching the true minimum solution. With this high accurate solution as the starting model, the following steps also have high possibilities to reach the results close to the true model. For the test partial Marmousi model, the starting offset and time are set as 1000 m

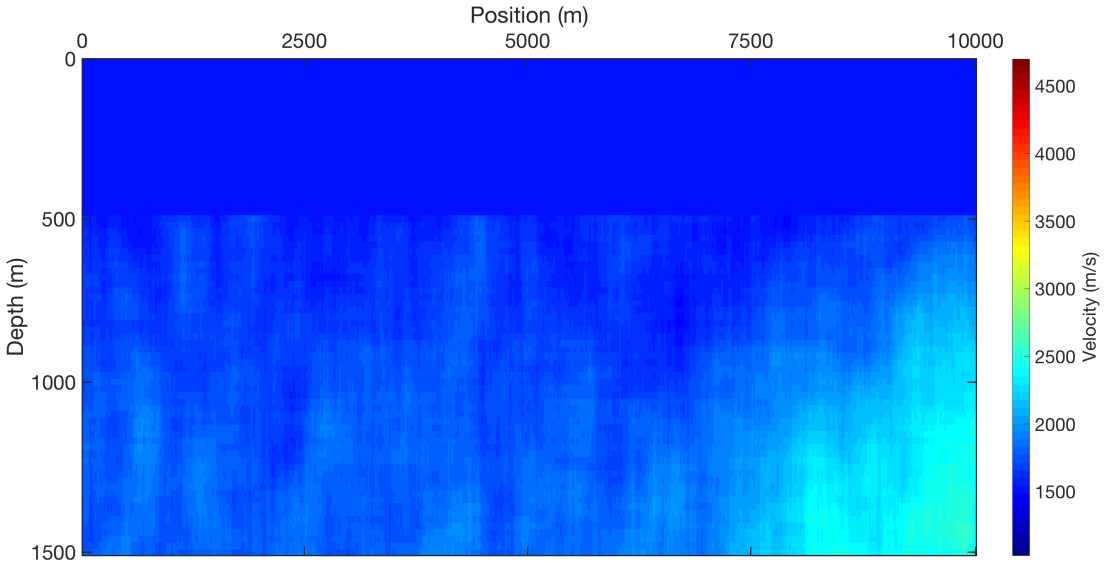



Figure 7: The perturbed starting model

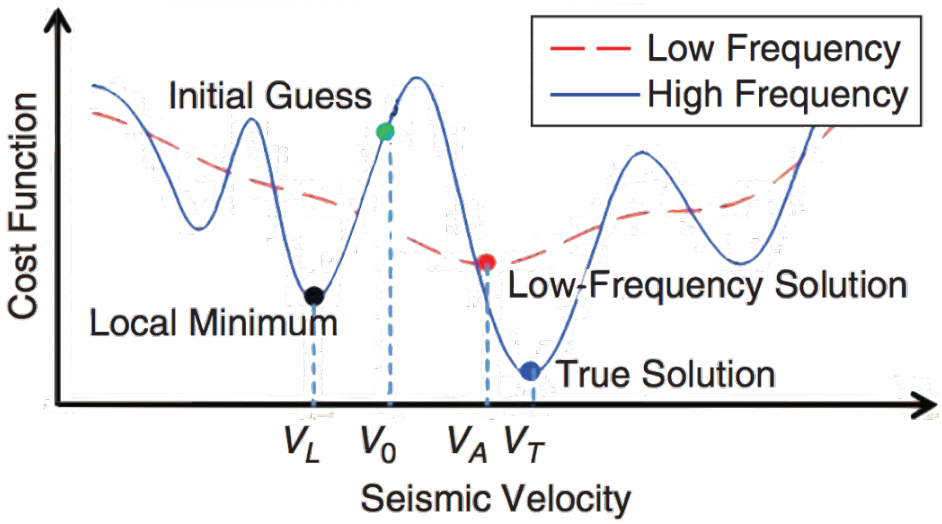



Figure 8: The cycle skipping problem (Hu, Chen, Liu, and Abubakar, 2018)

and 1.20 s, respectively. They are increasing with the step length of 450 m and 0.15 s.

At each iteration step, the scaled synthetic seismic data is filtered with a mask which removes all the reflections in the data (Figure 9). The successful FWI applications usually use the refraction with a lower frequency because the reflection FWI has some main difficulties. Firstly, The reflection coefficient is related to the density model which also has real high

uncertainty and needs to be estimated during the FWI. Secondly, the estimation of starting model should be extremely close to the true velocity model to migrate the reflectors to the correct depth. Finally, the gradient generated by FWI typically favors updating reflector amplitudes more than it favors updating the background macro-velocity model (Yao and Wu, 2017).

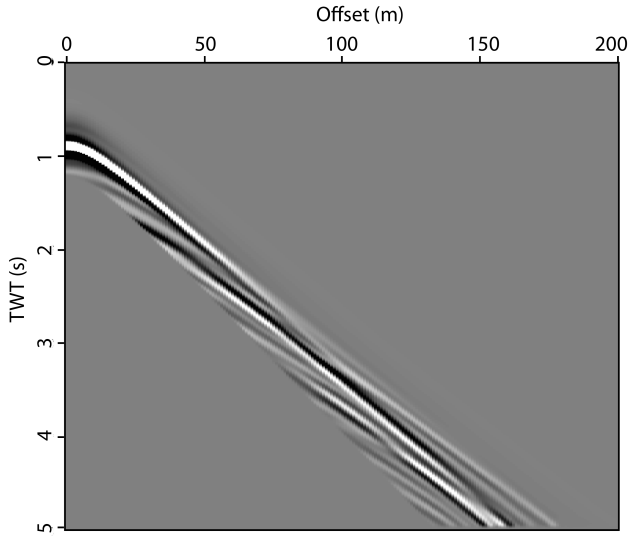



Figure 9: The muted seismic record, which only remains the direct and refracted arrivels.

For the multiscale method, the FWI is applied in every iterative step with the difference of misfit calculation and inner iteration number of maximum 5 times. Because every scale of the velocity model will use as the starting model in the next step. Only the P-wave velocity will be updated during the FWI iteration. The brief workflow of multiscale iteration is shown as Figure 10.

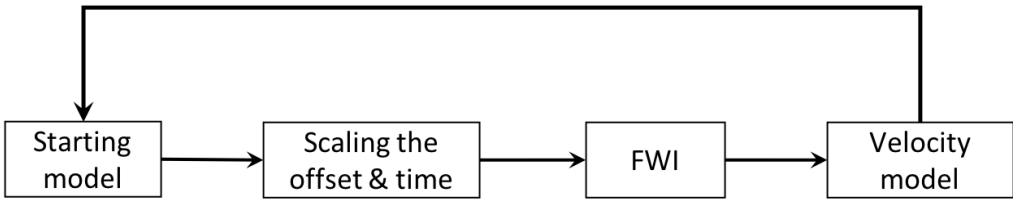
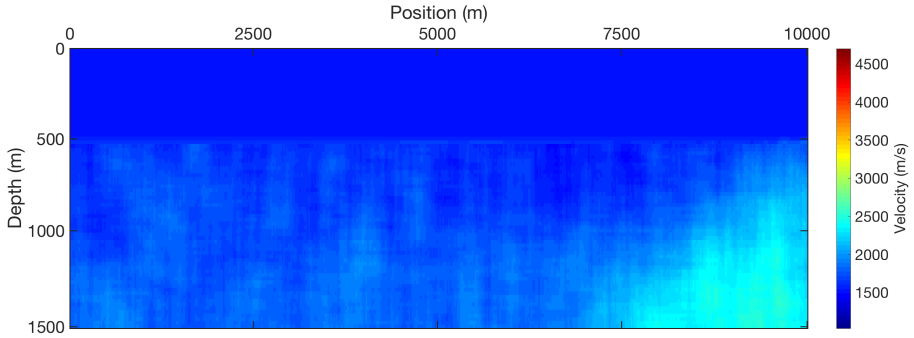



Figure 10: The loop of multiscale interation

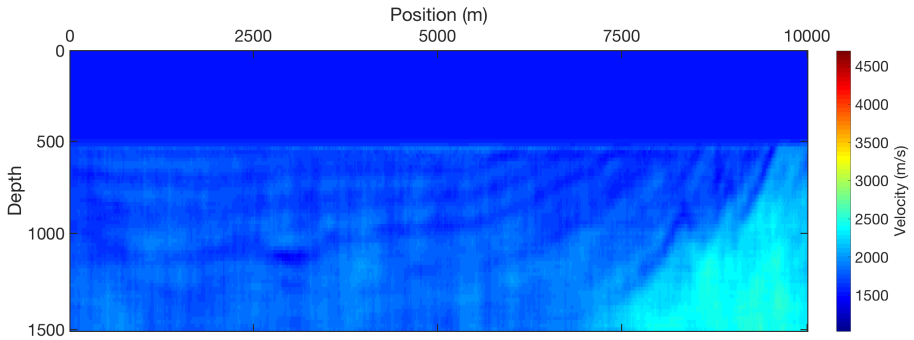
## 3.4 Results and Uncertainty Analysis

The purpose of the partial Marmousi model is to test the method of evaluating the uncertainties of the ensemble multiscale FWI. The multiscale iteration starts with the smallest

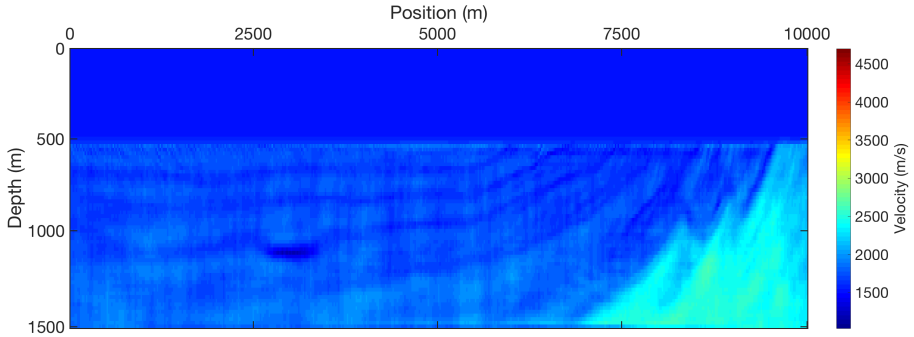
offset and time, 100 m and 1.2 s respectively, and ends with the full size of offset and time. In this case, there are 20 iterations to go through the entire area. The 1st, 10th and 20th iteration results are shown as Figure 11.




**(a)**




**(b)**




**(c)**

Figure 11: The results of partial Marmousi model FWI. (a) the result of 1st iteration (smallest offset and time), (b)the result of 10th iteration (half offset and time), (c)the result of 20th iteration (full offset and time)

After going through the whole offset and time, the outline of main structures have been successively inverted. Compare with the true model (Figure 2b), the formation boundaries,

shallow gas and the series normal faults are shown clearly in the final result (Figure 11c).

This multiscale FWI has been implemented on the ensemble of 20 starting models. Using these 20 results as samples, the uncertainties and accuracy can be evaluated by the statistic method. In the actual geophysical exploration, the real velocity is typically unknown. To evaluate the uncertainty of the FWI, the key value of the accuracy is standard deviation. As it is shown in Figure 12, the standard deviation shows the high values at the deep corner of both sides and the shallow gas position.

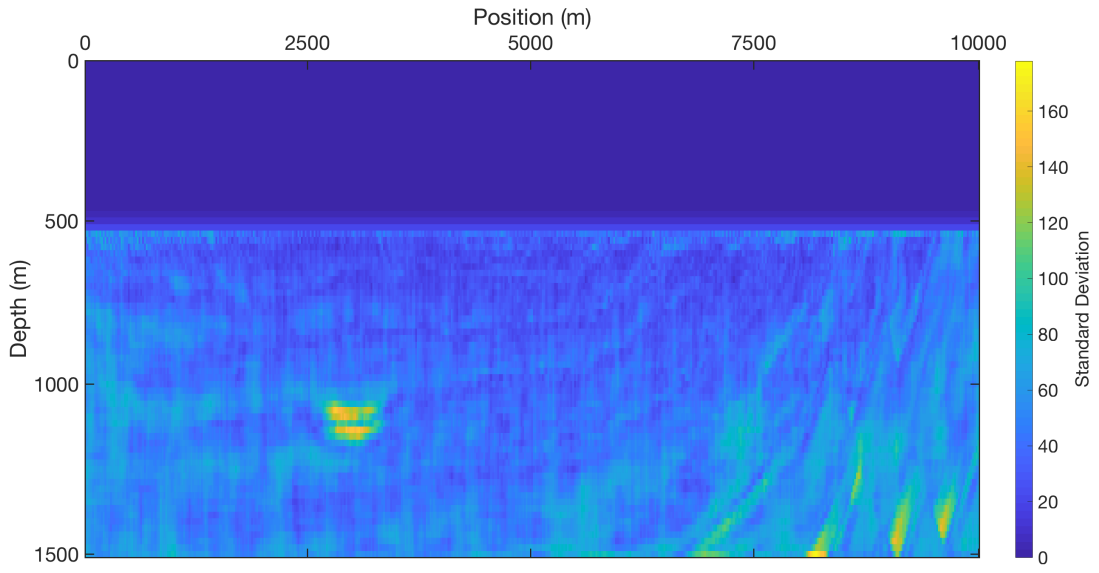



Figure 12: The standard deviation of partial Marmousi multiscale FWI

## 3.5 Full Marmousi Model Application

### 3.5.1 FWI of Full Marmousi Model

The partial Marmousi multiscale FWI has successfully estimated the uncertainty of the inversion result of the ensemble calculation. The multiscale method can be extended to the full scale of the Marmousi velocity model (Figure 2a) to test the flexibility of its application.

The full scale Marmousi model is 17000 m in width and 3500 m in depth. The model consists of $851 \times 176$ grid-cells in the horizontal and vertical directions, respectively. Both the horizontal and vertical grid spacings are 20 m. There 53 sources at the water depth of 20 m and spacing interval of 320 m. The 426 receivers are located on the ocean bed cable with fixed spacing of 40 m. The wavelet is the same one as we described in the partial Marmousi

model which is shown in Figure 3. The recording length is slightly increased which is 6 s and sampling interval is 1 ms. The synthetic seismic data with the bandpass filter of 7.5 Hz is demonstrated as Figure 13a. The reflections are muted during the multiscale iterations (Figure 13b).

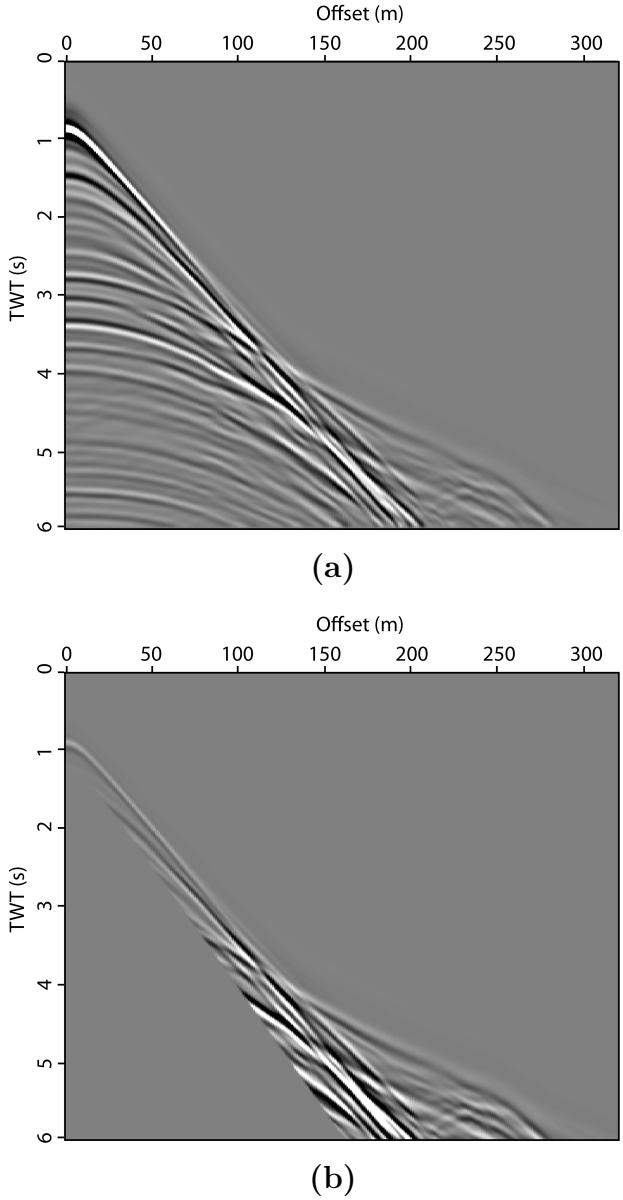
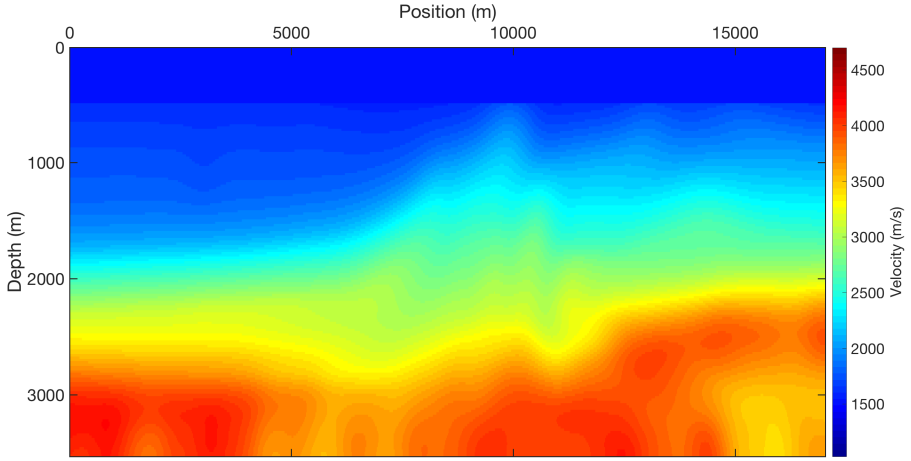


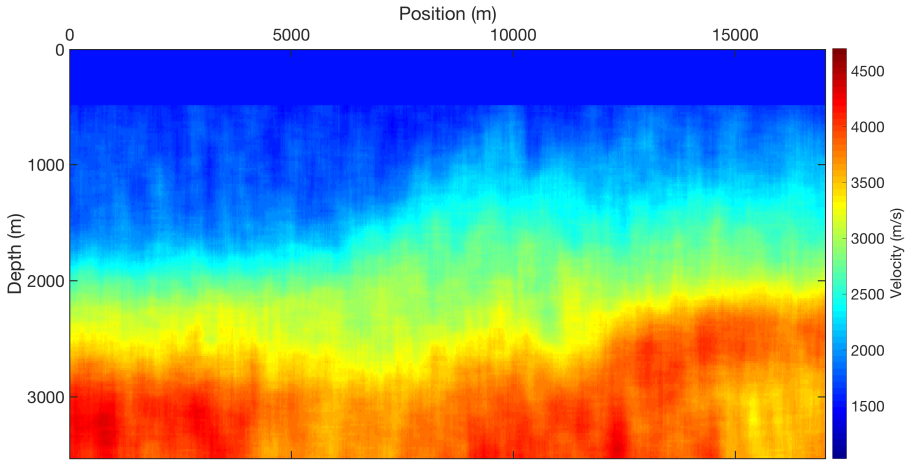

Figure 13: The first short of full Marmousi synthetic data. (a) raw data, (b) muted data

The initial model is smoothed with the radius of 20 from the true velocity(Figure 14a). As mentioned in the partial Marmousi multiscale FWI, the random perturbation method has been used to create the ensemble. In the full Marmousi model case, the same method is implemented in the initial model with a uniformly distribution numbers in the interval between -2000 and 2000. The ensemble includes 20 members and one of them compares to

the initial model is shown as Figure 14b.




**(a)**




**(b)**

Figure 14: The initial model of full Marmousi mode FWI. (a) the initial model smoothed from the true velocity model, (b) one of the perturbed initial model

Next, the multiscale FWI is implemented on this data set. The starting offset and time are 1000 m and 1.20 s with an increment of 450 m and 0.15 s, respectively. The FWI iteration in each multiscale iterative step has a limitation of 5 times. It takes 32 iterative steps to go through the entire area. The results of 1st, 10th, 20th and the final iteration are demonstrated as Figure 15.

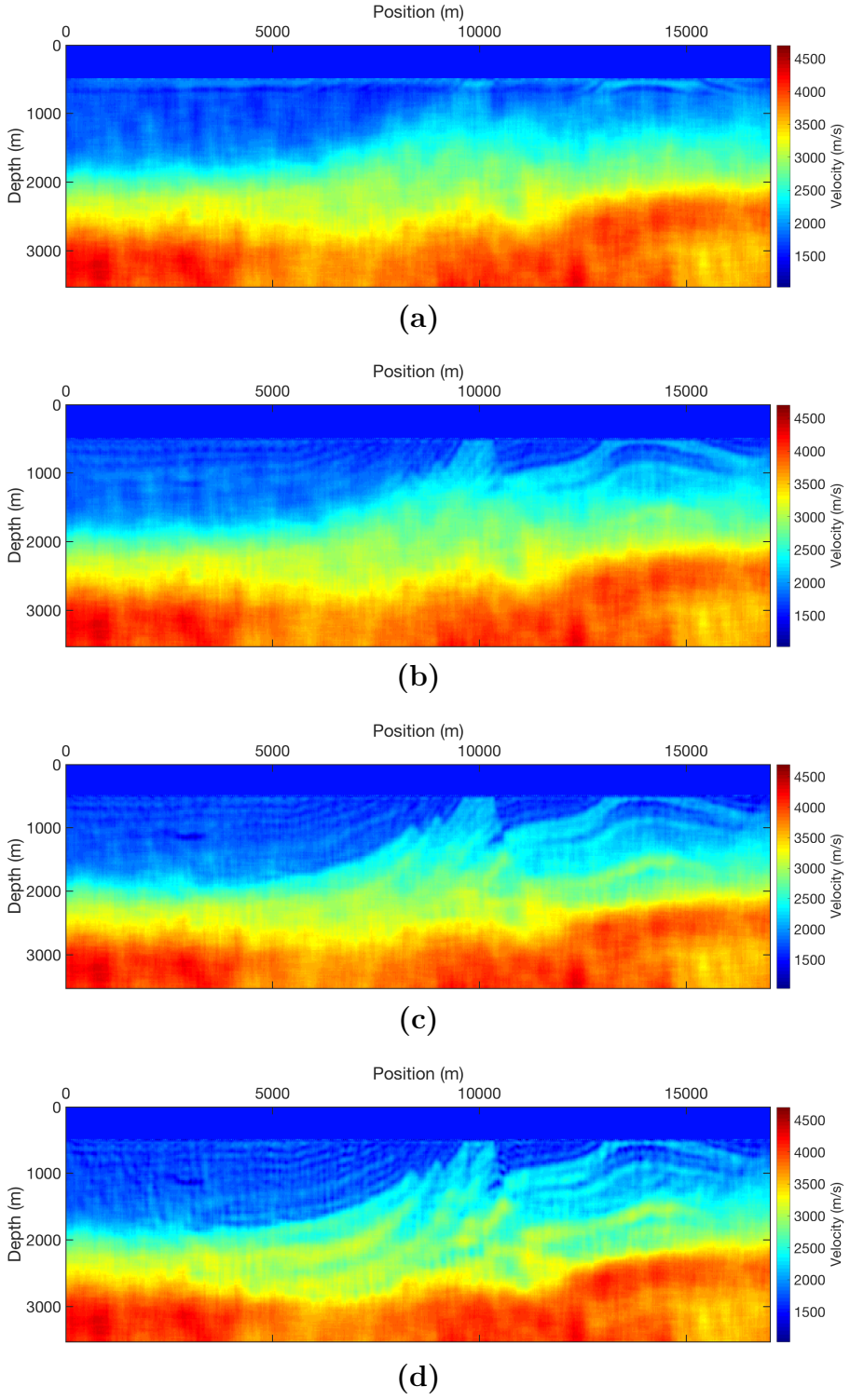
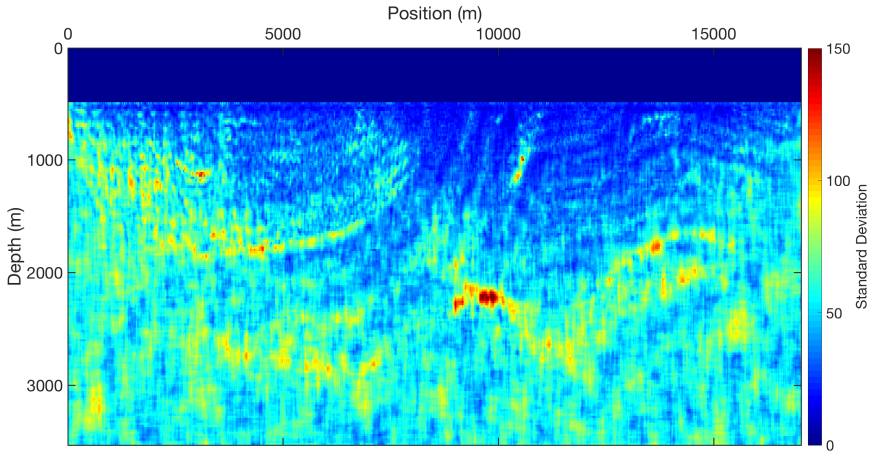
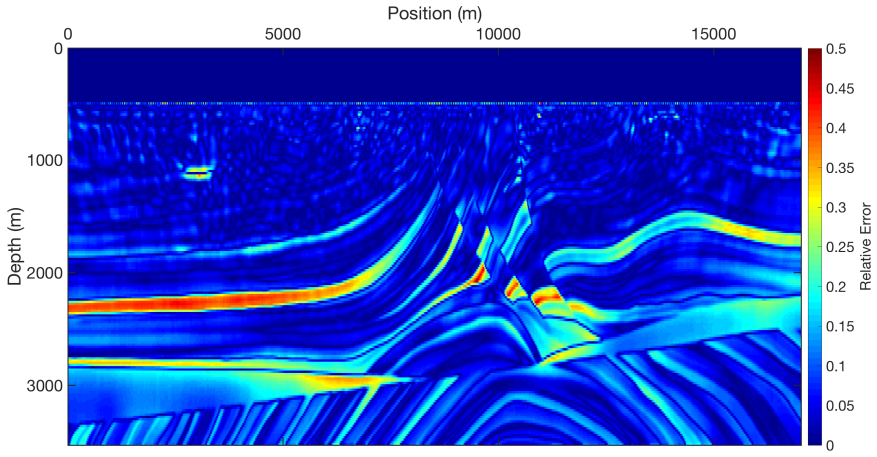



Figure 15: The results of full Marmousi model FWI. (a) the result of 1st iteration (smallest offset and time), (b)the result of 10th iteration (one third offset and time), (c)the result of 20th iteration (two thirds offset and time), (c)the result of 32th iteration (full offset and time)

The result shows the shallow part of the model has been properly inverted. The series of normal faults, the formation boundary and shallow gas are fully illustrated in the result of final iteration. The deep part and the right/left side close to the edge of the model is not completely inverted. One reason for this phenomenon could be lack of FWI iterations at the final step. The test result shows it will add artifacts by increasing the iterations of FWI in the final mutiscale step. The other possible reason is the refraction is limited by the offset and angle. At this certain survey geometry, the refraction cannot reach the deep part and the right/left margin.




**(a)**




**(b)**

Figure 16: The uncertainty analysis of full Marmousi mode FWI. (a) the standard deviation of ensemble full Marmousi mutiscale FWI, (b) the relative error $\frac{|V_{fwi} - V_t|}{V_t}$, $V_{fwi}$ is the average of the ensemble FWI results, $V_t$ is the velocity of true model

The standard deviation and relative error has been estimated by using the ensemble result and true model (Figure 16). An inverse triangle of low values is shown in the distribution of standard deviation. This proves that the failure of deep and right/left margin FWI is caused by the limitation of the offset. There is not enough offset to generate the refraction to reach this aforementioned area. Comparing the standard deviation and relative error, they match extremely well. It appears to be the high standard deviations around the formation between 5000 m and 8000 m with the depth around 2000 m, the area close to points (10000,1000) and (9500, 2100), the shallow gas, and the small anticline at the depth of 1800 m on the right side (Figure 16a). These areas also show high values in the relative error distribution (Figure 16b). This means the accuracy or uncertainty of FWI can be estimated by using the ensemble standard deviation result without using the true velocity model.

### 3.5.2 Larger Ensemble Size

The result of full Marmousi model shows a good estimation of the uncertainties, but there are still some disadvantages, such as the size of the ensemble being quite small and the resolution of FWI results being not perfect. To solve these problems, an ensemble with 200 members is generated by using the same starting model and the random perturbation method mentioned in the Chapter 3.5.1. The survey geometry and the synthetic seismic data are the same except the reflections are included to get a better resolution. One of the FWI result is shown as Figure 17. Thereafter the standard deviation is calculated which is shown as Figure 18.

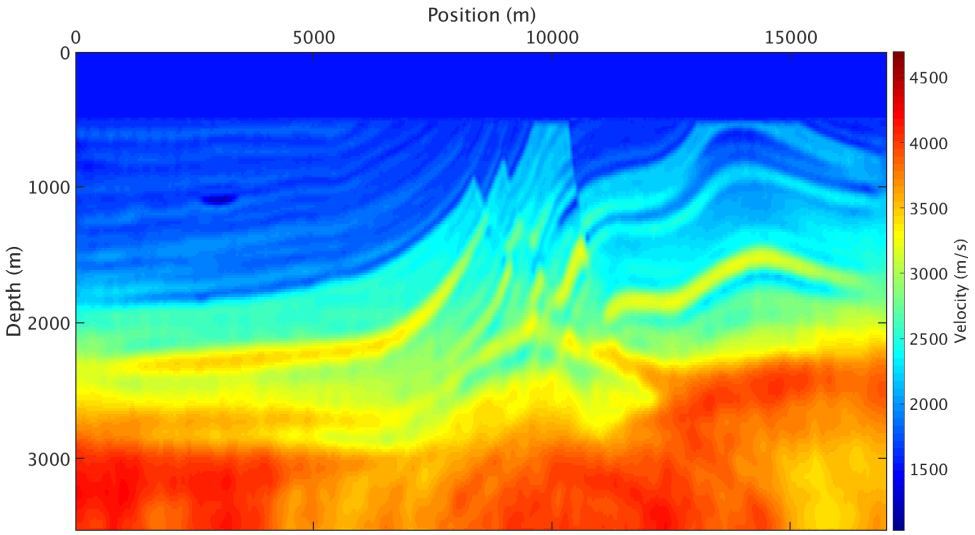



Figure 17: The one of full Marmousi mode FWI results with reflections

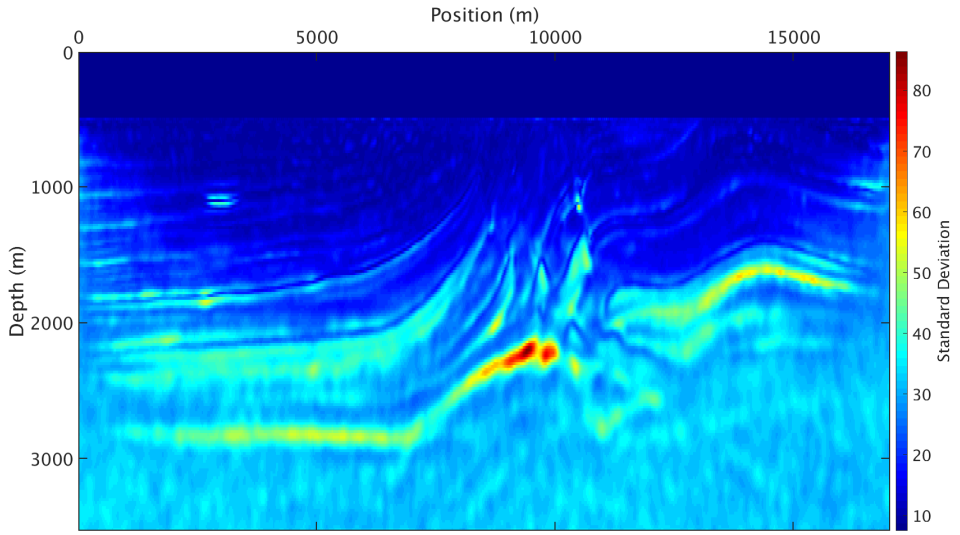



Figure 18: The standard deviation of full Marmousi mode FWI with larger ensemble size

The FWI result using reflections shows a better resolution with sharper formation and structural contacts. It also has more information about the deep parts than the refraction FWI results. The problem at the right/left side and deep part is still there, but it has improved significantly. The standard deviation also has high values in these parts. With larger ensemble size and higher resolution, the standard deviation shows lower values compared to Figure 16b. However, the spatial distribution of the highest values follows the same pattern.

# 4 Applications of FWI in the North Sea 2D Seismic

## 4.1 Dataset

The area of the survey corresponding to this section is the 2D line NSR-31150R2, Sequence 124 from the North Sea. Considering the cost of computation time, we only take a part of the 2D line, which is from 12 km to 26 km of the start of the 2D line and 2 km in depth. There is a channel in the center of this section which is the reference structure for the inversion. There are 118 source positions with a regular interval of 100 m in between them. The source depth is 7 m. The streamer length is 8100 m, and consists of 314 receivers with a regular interval of 25 m and at a water depth of 9 m. The minimum offset is 175 m from the source point. A low pass filter with a high cut of 7.5 Hz is applied on the 2D seismic data before running the FWI.

The wavelet is estimated from the amplitude spectrum of the direct wave arrivals, and then a Wiener filter (Sheriff, 1995) is applied in the estimated wavelet to yield a proper fit of the observation data. The recording length is 5 s and the sampling interval is 2 ms. The final estimated wavelet is shown in Figure 19.

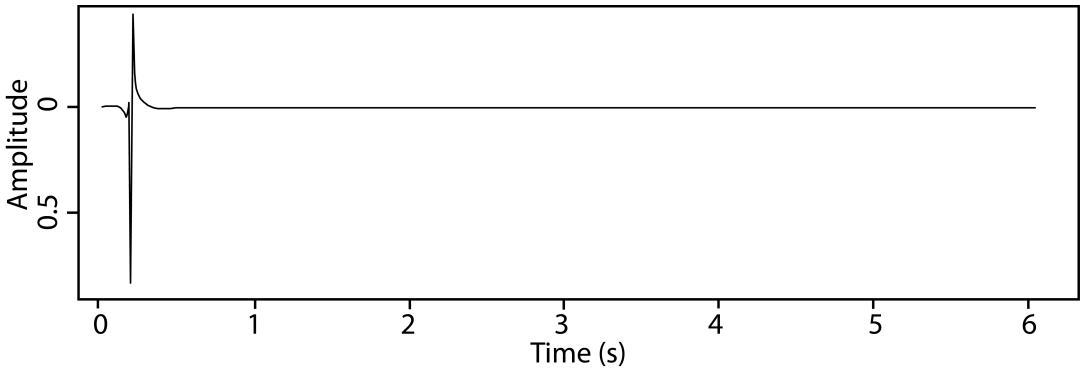
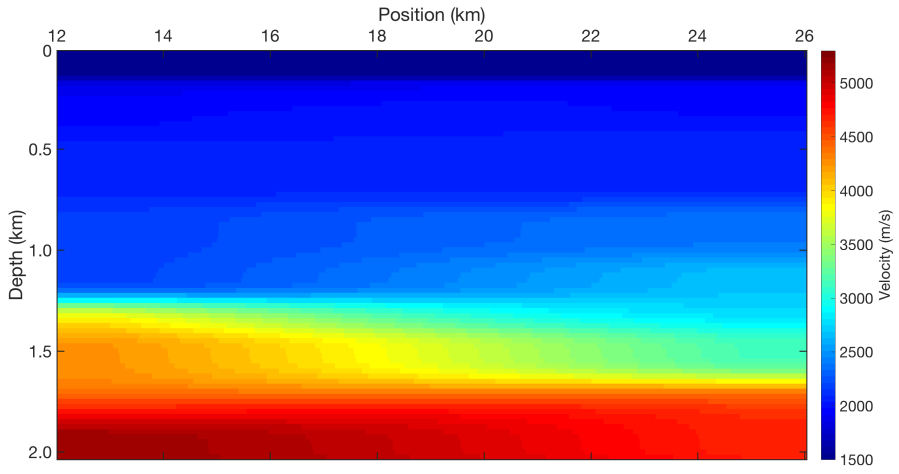


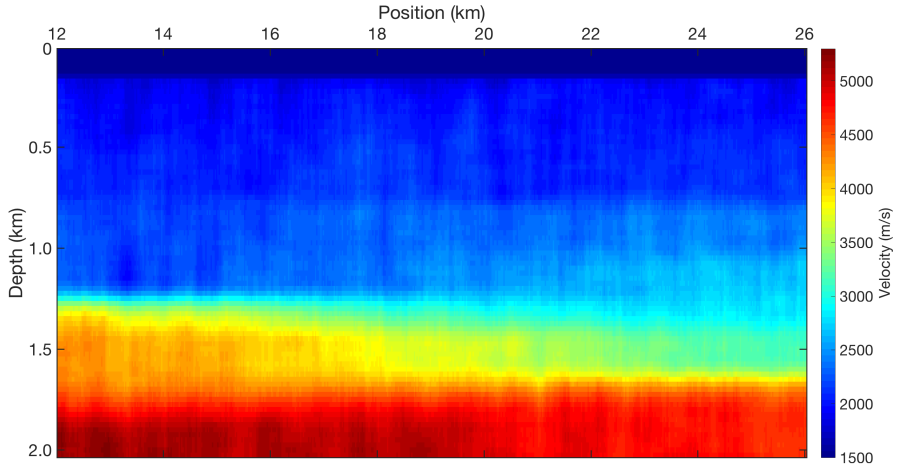

Figure 19: The estimation of wavelet for the field seismic data

The starting model of velocity is given by conventional velocity analysis of common midppoint gathers, and is shown as Figure 20a. The random perturbation is added into the starting model to create the ensemble (Figure 20b). The number of ensembles is 20 in this case.

The multiscale ensemble FWI is then applied to the real field data. The starting offset and time are 1 km and 1.8 s with an increment of 450 m and 0.2 s, respectively. A fixed number of 5 iterations of l-BFGS FWI are done at each multiscale step. The total number of the multiscale steps is 16.




**(a)**




**(b)**

Figure 20: The initial model of NSR-31150R2. (a) the initial model from the velocity picking, (b) one of the perturbed initial model

## 4.2 Results

As the inversion results shown in Figure 21, the channel between 18 km and 20 km is successively getting more details. The deep part is fairly homogeneous, with little reflectivity, and therefore is not well constrained by the seismic data. There are two layers with low velocity at the depth around 0.5 km.

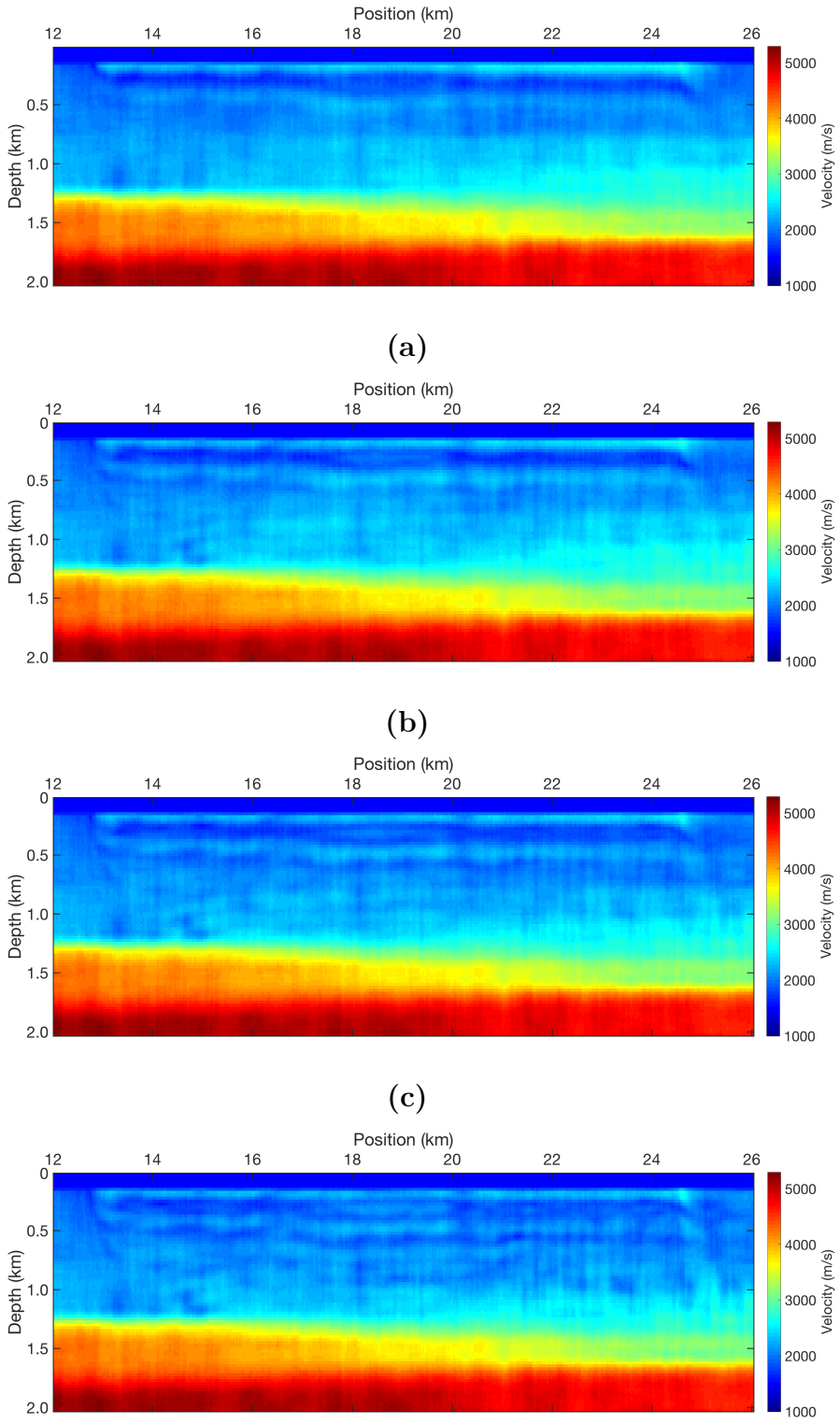



Figure 21: The results of the real field 2D seismic multiscale ensemble FWI. (a) the result of 4th iteration, (b)the result of 8th iteration, (c)the result of 12th iteration, (c)the result of 16th iteration

The standard deviation shows the low values are at the low velocity layers. The channel, the part beneath it and the formation between the low velocity layers have high values of standard deviation. The deep part does not show much differences because the formation is fairly homogeneous.

As we can see from both synthetic data results and real field data results, the uncertainty of FWI at the contact of layered formations is quite low. The complicated structures in the deep part and large changes over smalls spatial scales may lead to high uncertainties in FWI.

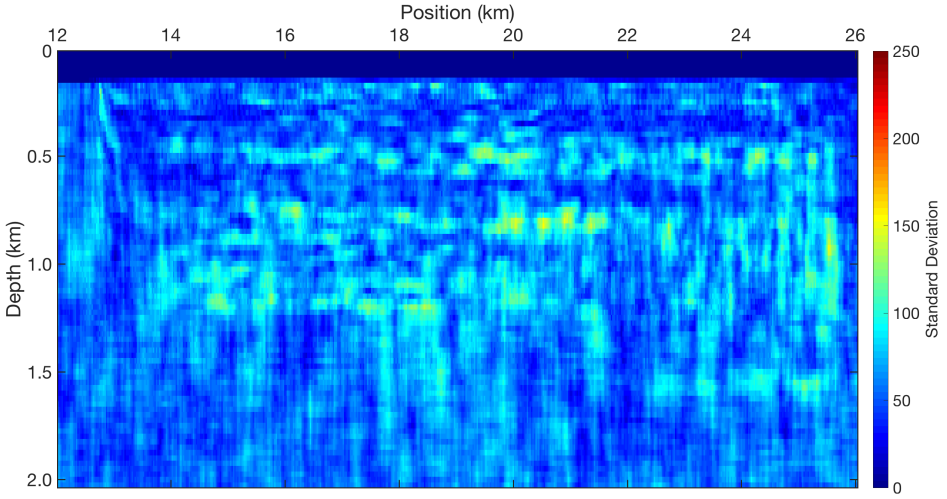



Figure 22: The standard deviation of the real field 2D seismic multiscale ensemble FWI

# 5 Discussion

In this master thesis, we reviewed the FWI theory, which includes the forward modeling and inversion theory and applied two methods combining the usage of multiscale iteration and the ensemble method to estimate the uncertainty and accuracy. The FWI is implemented during each multiscale iteration step, and an ensemble of initial models is inverted independently to get the sample distribution. The multiscale iteration has two main purposes. First, the small offset and short time can prevent the cycle skipping artifacts. Second, the basic concept of data assimilation is using the different observation data to improve the initial starting model and use it as the starting model with new observation. The scaled seismic data with different offset and time are several observations with dissimilar geometries. Using them in each multiscale iteration step matches the basic theory of data assimilation. And the ensemble algorithm is widely used in the recent study of data assimilation algorithm, such as the ensemble Kalman filter (EnKF) (Evensen, 1994, 2009), the Ensemble Kalman Smoother (EnKS) (Evensen and Van Leeuwen, 2000) and the iterative Ensemble Smoother (iES) (Luo, Stordal, Lorentzen, and Naevdal, 2015). The uncertainty can be estimated by using the ensemble simulation.

The Marmousi model and a 2D line of real field seismic data have been used to test the combination of these two methods on FWI. The partial and full Marmousi models are demonstrating proper estimations of the uncertainty in the FWI results. By comparing the relative errors yielded from the inversion result and true velocity model, it is proved that the estimation of standard deviation is approaching the real errors. The results also shows even the FWI fails to update the model in the deep formation and right/left margins due to limitations of the acquisition geometry, the uncertainty estimation still gives high value of standard deviation. This gives more confident to the uncertainty evaluation, because even if the geometry of FWI has some limitations for the FWI, the method still can give the correct uncertainty estimation for the deep part. From the uncertainty results, the high value of standard deviation commonly shows at the position where the velocity changes a lot both laterally and vertically, such as at the shallow gas, series faults, and channel. This uncertainty may be improved by using a more accurate starting model or applying the full data assimilation (i.e. including higher frequencies).

There are still some limitations and disadvantages for this multiscale ensemble FWI method in estimating the uncertainty. First, the entire method implemented in this thesis is one single loop of the data assimilation method. Using the entire data assimilation method may improve the accuracy of the FWI. If we want to apply the entire data assimilation method, there are several challenges. Since implementing the FWI is quite time consuming, even though it was performed on the high performance computer, it still took more than 20 hours to finish a

single 2D section. To estimate the uncertainty, we need to run more than 20 members of the ensemble. This will need days to finish the calculation for just one iteration. As we already proved the method works well with single loop and it can estimate the uncertainty. The iteration of this loop should be working with some adjustment and enough time. Second, the method is applied on 2D acoustic model. The feasibility of 3D and elastic model should be taken into consideration. Third, the survey geometry, e.g. offset, has the influence of the refraction FWI as aforementioned in the full Marmousi model results. The sensitivity of the geometry parameters should be an interesting topic which we didn't take it into account.

# 6 Conclusions

## 6.1 Conclusions

The FWI currently has no clear vision of uncertainty and accuracy of the inversion results. Because there could be many inversion results that fit the same true velocity model for the inversion problem, the uncertainty and accuracy of the results are the most important data need to be known. Another challenge for FWI is about reducing the cycle skipping artifacts. This research tested whether the ensemble multiscale FWI does have a feasibility to estimate the uncertainty of inversion problem and reduce the cycle skipping artifacts. We reviewed the theory of FWI and introduced the ensemble method and multiscale method into the refraction FWI algorithm. The method was implemented on the Marmousi model in different scales and the real field data. The results showed that the FWI with ensemble and multiscale method could estimate the uncertainty within the 2D section. The high uncertainty is particularly notable at the complex structures and the position where the velocity for both lateral and vertical changes dramatically . The research also found when the seismic survey geometry was not properly set, like the offset is not large enough, the high uncertainty will happen in the deep formations and right/left margin of the 2D sections. With the information of uncertainties, the non-unique inversion problem has been solved with a known accuracy. It is also quite important for the geophysicist and geologist to study the geological structures and find reservoirs of hydrocarbon in a model with certain accuracy.

## 6.2 Future Work

Estimating the uncertainty of FWI is an extremely difficult challenge, and this thesis has set quite a good cornerstone of this study and done quite a lot of it. For the further study on this topic, building up a better wavelet and starting model should take more effort in the study. The entire data assimilation method should be tested, which will reduce the requirement of the starting model. The influence of survey geometry could be also a fairly interesting factor. The sensitivity of geometry for estimating the uncertainty should be tested.

The ensemble multiscale FWI method should expand to 3D and elastic wave propagation and inversion. This could be quite time consuming and require the powerful computational ability of the computer.

A study of North Sea 2D seismic has been successfully implemented. More applications on the real dataset with different geological settings could be very interesting to test the feasibility of this method.

# References


Backus, G. E. & Gilbert, J. F. (1967). Numerical Applications of a Formalism for Geophysical Inverse Problems. *Geophysical Journal International*, *13*(1-3), 247–276. doi:10.1111/j.1365-246X.1967.tb02159.x

Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, *114*(2), 185–200. doi:https://doi.org/10.1006/jcph.1994.1159

Claerbout, J. (2014). *Geophysical image estimation by example*. Lulu. com.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, *99*(10), 20. doi:10.1029/94JC00572

Evensen, G. (2009). *Data assimilation : the ensemble Kalman filter* (2nd). Dordrecht: Springer.

Evensen, G. & Van Leeuwen, P. J. (2000). An ensemble Kalman smoother for nonlinear dynamics. *Monthly Weather Review*, *128*(6), 1852–1867. doi:10.1175/1520-0493(2000)128<1852:AEKSFN>2.0.CO2

Fichtner, A. (2011). *Full Seismic Waveform Modelling and Inversion*. Advances in Geophysical and Environmental Mechanics and Mathematics. Springer-Verlag Berlin Heidelberg.

Hu, W., Chen, J., Liu, J., & Abubakar, A. (2018). Retrieving Low Wavenumber Information in FWI: An Overview of the Cycle-Skipping Phenomenon and Solutions. *Signal Processing Magazine, IEEE*, *35*(2), 132–141.

Jensen, F. B., Kuperman, W. A., Porter, M. B., & Schmidt, H. (2011). *Computational Ocean Acoustics*. Modern Acoustics and Signal Processing. New York, NY: Springer New York.

Kosloff, D. D. & Baysal, E. (1983). Migration with the full acoustic wave equation. *GEOPHYSICS*, *48*(6), 677–687.

LeVeque, R. J. (2002). Finite volume methods for hyperbolic problems. Cambridge: Cambridge University Press.

Luo, X., Stordal, A., Lorentzen, R., & Naevdal, G. (2015). Iterative Ensemble Smoother as an Approximate Solution to a Regularized Minimum-Average-Cost Problem: Theory and Applications. *SPE J.* *20*(5), 962–982. doi:10.2118/176023-PA

Marfurt, K. J. (1984). Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations. *GEOPHYSICS*, *49*(5), 533–549.

Nocedal, J. (1980). Updating quasi-Newton matrices with limited storage. *Mathematics of Computation*, *35*(151), 773–782. Retrieved from http://www.ams.org/jourcgi/jourgetitem?pii=S0025-5718-1980-0572855-7

Nocedal, J. (1999). *Numerical optimization*. Springer series in operations research. New York: Springer.

Pereira, Â., Nunes, R., Azevedo, L., Guerreiro, L., & Soares, A. (2017). Geostatistical seismic inversion for frontier exploration. *Interpretation*, *5*(4), T477. doi:10.1190/INT-2016-0171.1

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain; Part 1, Theory and verification in a physical scale model. *Geophysics*, *64*(3), 888–901. doi:10.1190/1.1444597

Pratt, R. G., Shin, C., & Hick, G. J. (1998). Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophysical Journal International*, *133*(2), 341–362. doi:10.1046/j.1365-246X.1998.00498.x. eprint: /oup/backfile/content_public/journal/gji/133/2/10.1046/j.1365-246x.1998.00498.x/2/133-2-341.pdf

Sheriff, R. E. (1995). Exploration seismology. (2nd ed., Chap. 15, pp. 559–560). Cambridge: Cambridge University Press.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, *49*(8), 1259–1266. doi:10.1190/1.1441754

Tarantola, A. (1987). *Inverse problem theory: Methods for data fitting and parameter estimation*. Elsevier, Amsterdam.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Philadelphia: Society for Industrial and Applied Mathematics.

Tengesdal, H. C. (2013). Ray-born modelling and full waveform inversion. Bergen: Department of Earth Science, University of Bergen.

Versteeg, R. (1994). The Marmousi experience: Velocity model determination on a synthetic complex data set. *The Leading Edge*, *13*(9), 927–936.

Virieux, J. & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Virieux, J. [Jean]. (1986). P-SV wave propagation in heterogeneous media; velocity-stress finite-difference method. *Geophysics*, *51*(4), 889–901.

Yao, G. & Wu, D. (2017). Reflection full waveform inversion. *Science China Earth Sciences*, *60*(10), 1783–1794.