Karlsruhe Institute of Technology

Geophysical Institute KIT-Dept. of Physics

2D ELASTIC FULL-WAVEFORM INVERSION OF LAND SEISMIC DATA WITH TOPOGRAPHIC VARIATIONS

2D ELASTISCHE INVERSION DER WELLENFORM LANDSEISMISCHER DATEN BEI TOPOGRAPHIE DER ERDOBERFLÄCHE

Master's Thesis of

Daniel Krieger

at the Department of Physics

Geophysical Institute (GPI)

Date of submission:

15.04.2019

Reviewer: Prof. Dr. Thomas Bohlen
Second reviewer: Prof. Dr. Andreas Rietbrock

Abstract

Areas with strong topography and complex subsurface geology like the Canadian foothills are of interest for exploration geophysics. They pose particular processing and imaging challenges, though. In order to create under these circumstances subsurface images with high resolution, sophisticated methods such as full-waveform inversion (FWI) have to be applied. FWI is, in principle, capable of producing highly detailed parameter models of the subsurface by iteratively updating a starting model to match both traveltimes and amplitudes of simulated data with data recorded in the field. When using land data, in particular when strong surface waves are observed that lead in combination with topographic variations to significant surface wave scattering effects, elastic wave propagation has to be considered. Previous works studying FWI in areas with topographic variations used, e.g., the spectral element method. While this method has the advantage that it has the intrinsic ability to meet the free surface condition in case of a non-planar top surface, it is also hard to implement and hard to use efficiently. Hence, the aim of this work is to investigate, if the use of the finite-difference method paired with the improved vacuum formulation that fully satisfies the free surface boundary condition is also justifyable. The finite difference method has the advantage that it is widely used and much simpler than the spectral element method. I show that the staircase effect has a major influence on Rayleigh waves but not on body waves. This influence can be reduced by choosing a discretization which is much finer than that needed for body waves. However, even without such fine discretization my approach is able to simulate wavefields with sufficient accuracy for FWI applications. I perform simultaneous reconstruction of P-wave velocity, S-wave velocity and density using synthetic data simulated with the spectral element method. I demonstrate that errors in my scheme result only in a thin high-velocity layer below the topographic surface of the P-wave velocity model. Furthermore, I investigate the impact of strong 3D effects, like they occur in mountainous regions, on 2D FWI. The result of this investigations is that in case of a 3D model which is constant in one dimension the recorded data can be easily transformed such that it is almost eqivalent to the 2D data regardless of the surface topography. In this case 2D FWI still produces sufficiently accurate results. In case of topographic variations perpendicular to the acquisition line, however, this transformation is no longer applicable. Thus, 2D FWI fails under such conditions.

Zusammenfassung

Regionen mit starker Topographie und komplexer Geologie, wie das Vorgebirge der kanadischen Rocky Mountains, sind von Interesse für die Explorationsgeophysik. Solche Regionen stellen allerdings eine große Herausforderung für geophysikalische Bildgebungsverfahren dar. Um unter solchen Umständen hochauflösende Abbildungen des Untergrunds zu generieren, sind komplexe Methoden wie die Wellenforminversion (WFI) von Nöten. Die WFI ist, im Prinzip, dazu fähig hochgradig detaillierte Parametermodelle des Untergrunds durch iterative Verbesserung eines Startmodells zu erzeugen, sodass sowohl die Laufzeiten als auch die Amplituden von simulierten Daten mit den im Feld gemessenen Daten übereinstimmen. Bei Landdaten, insbesondere wenn starke Oberflächenwellen beobachtbar sind, welche, in Kombination mit topographischen Variationen, zu signifikanten gestreuten Oberflächenwellen führen, muss elastische Wellenausbreitung berücksichtigt werden. Vorangegangene Studien nutzten z.B. die Spektralelementemethode. Während diese den Vorteil hat, dass die Randbedingung der freien Oberfläche auch bei nicht-planarer Oberfläche intrinsisch erfüllt ist, ist sie schwieriger zu implementieren und auch schwieriger effektiv nutzbar. Daher ist das Ziel dieser Arbeit zu untersuchen, ob die Nutzung der Finite-Differenzen-Methode in Kombination mit der verbesserten Vakuumformulierung ebenfalls effizient ist. Mit der verbesserten Vakuumformulierung ist die Randbedingung der freien Oberfläche ebenfalls erfüllt. Die Finite-Differenzen-Methode hat den Vorteil, dass sie bereits weit verbreitet und außerdem einfach zu nutzen ist im Vergleich zur Spektralelementemethode. Ich zeige, dass der Treppeneffekt einen starken Einfluss auf Rayleigh- aber nicht auf Raumwellen hat. Dieser Einfluss kann verringert werden, indem eine Diskretisierung gewählt wird, welche wesentlich feiner ist, als die Kriterien für Raumwellen verlangen. Allerdings ist meine Herangehensweise auch ohne solche feinen Diskretisierungen in der Lage, Wellenfelder mit ausreichender Genauigkeit für die Anwendung in der WFI zu simulieren. In meiner Arbeit nutze ich simultane Inversion der P-Wellengeschwindigkeit, der S-Wellengeschwindigkeit und der Dichte unter Verwendung synthetischer Daten, welche mit Spektralelementemethode simuliert sind. Ich demonstriere, dass Fehler in meinem Schema nur zu einer dünnen Hochgeschwindigkeitszone unter der Oberfläche des Modells der P-Wellengeschwindigkeit führen. Darüber hinaus untersuche ich den Einfluss von starken 3D-Effekten, wie sie in bergigen Regionen auftreten, auf die 2D WFI. Das Ergebnis dieser Untersuchung ist, dass im Falle eines 3D Modells, welches konstant in einer Dimension ist, die aufgezeichneten Seismogramme leicht transformiert werden können, sodass diese äquivalent zu 2D Daten sind, unabhängig von der Topographie. In diesem Fall produziert die 2D WFI ausreichend gute Ergebnisse. Im Falle von topographischen Variationen senkrecht zur Akquisitionsgeometrie allerdings, ist die verwendete Transformation nicht mehr anwendbar, weshalb die 2D Inversion unter diesen Bedingungen versagt.

Contents

Chapter 1

Introduction

Land seismic data becomes more and more interesting for explorations seismics because easy marine targets are already explored and new technical advances, like drones and wireless geophones, make seismic acquisitions on land easier and cheaper. Unlike in marine environments, the surface where the acquisition geometry is placed on can have significant height differences. One such region are the Canadian Foothills where, as of today, oil and gas are produced for decades. One of the earliest investigations of the Canadian Foothills was done by Gray and Marfurt (1995) which studied the performance of Kirchhoff depth migration in mountainous areas and created the well known Canadian Foothill model. For a successful migration one requires a velocity model of the subsurface. In the past velocity analysis is a foothills setting was, e.g., done using manual velocity picking in a geologically constrained time-migration workflow (Vestrum, 2007). Such methods require a lot of manual work and can only produce very smooth velocity models which are not sufficient for imaging of complex geologies.

A method which is capable of producing highly resolved parameter models of the subsurface is full-waveform inversion (FWI). As the name already suggests, goal of the method is to find a subsurface model which explains both amplitude and phase of the recorded seismograms. This means that the method uses all of the information content inside the recorded data other than, e.g., traveltime tomography which uses only arrival times. The theoretical concepts of FWI were already developed in the early 1980s by Tarantola (1984). However, this method is computational so expensive that large data sets can only be processed on high-performance computers since the last decade (Brossier et al., 2009; Sears et al., 2010; Prieux et al., 2013). The high computational cost is still the limiting factor and restricts the use of FWI often to 2D applications. The main contributor to this cost is the forward problem, that is to simulate synthetic data given a subsurface model, which needs to be solved a lot of times during FWI. There exist various different numerical methods to solve this forward problem. The method which is most commonly used in FWI applications is the finite-difference (FD) method because of its simplicity and effectivity.

There are only a few FWI applications on land data so far but it becomes increasingly popular over the last few years. Examples are Plessix et al. (2013); Mei and Tong (2015); Zheng et al. (2016). However, all these authors use the accoustic approximation which requires massive preprocessing to remove elastic effects from the data, most notably the surface waves. In case of strong surface

topography removal of the elastic effects becomes even more challenging because of surface wave scattering and, to my knowledge, there is no example where this has been done successfully. Hence, elastic FWI has to be used in such a setting. To my knowledge the only study which used elastic FWI on land data with strong topographic variations is Trinh (2018). The author used the spectral element method (SEM) to solve the forward problem in 3D. This method is known to produce very accurate results in case of surface topography (Komatitsch and Tromp, 1999) but is also harder to implement and more difficult to use than the FD method. Thus, the goal of my work is to investigate, if the FD method can also be used for FWI applications in regions with strong topography. Since 2D FWI is currently more popular than 3D FWI because of the immensly reduced computational cost, I also investigate if a setting with topography can be handled as such.

Chapter 2

Theory

In this chapter I would like to introduce the basic theory used in this work. It is divided into two parts. In the first part I introduce the most relevant concepts which are needed to solve the elastic forward problem in seismics. Forward problem in this context means, to calculate the seismic wavefield given a certain distribution of model parameters. In the second part I discuss the inverse problem whose goal is the opposite of the forward problem, that is, to find a model that explains the partly measured wavefield.

2.1 Forward problem

Before solving the inverse problem of FWI it is necessary to understand the forward problem. First, I introduce the elastic wave equation and the concepts needed to derive it. After that I explain the transition from the continuous equation to its discretized form which is required for numerical simulations.

2.1.1 Elastic wave equation

Derivation

In this section the equations for elastic wave propagation are derived. Conceptually, the derivation follows closely Aki and Richards (2002) to which I also refer the reader for a more thorough explanation. In an elastic medium a linear relation between stress and strain is assumed. This relation is described by the generalized Hooke's law. Using Einstein notation this can be written as

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

Stress σ and strain ϵ are second-order tensors while the stiffness tensor C is of fourth order. This means C has 81 components; however, not all of them are independent. As σ and ϵ are symmetric, the number reduces to 36 and for energy reasons, it is further reduced to 21 independent components.

For isotropic media the stress-strain relation can be instread described with only two parameters, called Lamé parameters:

$$\sigma_{ij} = \lambda \operatorname{Tr}(\epsilon)\delta_{ij} + 2\mu\epsilon_{ij} \tag{2.2}$$

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

The constant λ is the first Lamé parameter and μ is the second Lamé parameter, also known as shear modulus. Both parameters are considered time-invariant. The symbol δ_{ij} is the Kronecker delta and u_i is the displacement of a particle in direction of the ith dimension of a Cartesian coordinate system. The relation between the Lamé parameters and seismic velocities is

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

The parameter v_P denotes the velocity of a P-wave and v_S that of a S-wave. In these relations the mass density ρ also occurs which is, like the Lamé parameters, considered time-invariant. Besides the stress-strain relation another important concept for elastic wave propagation is Newton's second law. Considering an arbitrary finite volume V with bounding surface S inside a deformable body yields

$$\int_V \rho\frac{\partial^2 \vec{u}}{\partial t^2}\,\mathrm{d}V = \int_V \vec{f}\,\mathrm{d}V + \int_S \vec{T}(\vec{n})\,\mathrm{d}S \tag{2.5}$$

The left-hand side of 2.5 represents the inertial forces of the volume V and the right-hand side represents the sum of the body and surface forces acting on the volume. This means, that $\vec{f}$ are the body forces acting on a unit volume of V and $\vec{T}$ are the forces acting on a unit area of S. The vector $\vec{T}$ is also called traction. The traction can be expressed in terms of the stress tensor σ as

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

The vector $\vec{n}$ is of unit length and pointing outward normal to the surface S. Using 2.5 and 2.6 together with Gauss' divergence theorem yields

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

Since this expression has to be independent of the volume V, the integrand has to be zero. This gives the equation of motion

$$\rho\frac{\partial v_i}{\partial t} = f_i + \frac{\partial \sigma_{ij}}{\partial x_j}\,. \tag{2.8}$$

Here, the equation is written in terms of the particle velocity $\vec{v}$. Equations 2.2, 2.3 and 2.8 completely describe wave propagation in linear elastic media. Taking the time derivative of equations 2.2 and 2.3

is the final step towards the velocity-stress formulation of elastic wave propagation:

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

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

Matrix-vector formulation

A more compact formulation of the forward problem can be achieved by writing equations 2.8 - 2.10 in terms of matrices and vectors. By rearranging the equations they can be written in the form:

$$\vec{F}(\vec{\xi}, m) = \mathbf{M}^{-1}\left(\frac{\partial \vec{\xi}}{\partial t} - \vec{b}\right) - \mathbf{Q}\vec{\xi} = 0\ . \tag{2.11}$$

The new vectors $\vec{\xi}$ and $\vec{b}$ represent all the different wavefields and the corresponding sources, respectively. Considering only two spatial dimensions, they can be written as

$$\vec{\xi} = \begin{pmatrix} v_x \\ v_y \\ \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{pmatrix} \qquad \vec{b} = \begin{pmatrix} f_x/\rho \\ f_y/\rho \\ \frac{\partial \sigma^{(0)}_{xx}}{\partial t} \\ \frac{\partial \sigma^{(0)}_{yy}}{\partial t} \\ \frac{\partial \sigma^{(0)}_{xy}}{\partial t} \end{pmatrix} \tag{2.12}$$

A small change to the formulation in the previous sections is the introduction of source terms for the stress components $\sigma^{(0)}_{ij}$. These are useful for acoustic sources. The first matrix $\mathbf{M}$ in 2.11 consists of the model parameters m which are λ, μ and ρ. The second matrix $\mathbf{Q}$ in the equation are the spatial derivatives:

$$\mathbf{M} = \begin{pmatrix} \rho & 0 & 0 & 0 & 0 \\ 0 & \rho & 0 & 0 & 0 \\ 0 & 0 & A & B & 0 \\ 0 & 0 & B & A & 0 \\ 0 & 0 & 0 & 0 & \mu^{-1} \end{pmatrix} \qquad \mathbf{Q} = \begin{pmatrix} 0 & 0 & \frac{\partial}{\partial x} & 0 & \frac{\partial}{\partial y} \\ 0 & 0 & 0 & \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \\ \frac{\partial}{\partial x} & 0 & 0 & 0 & 0 \\ 0 & \frac{\partial}{\partial y} & 0 & 0 & 0 \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} & 0 & 0 & 0 \end{pmatrix} \tag{2.13}$$

$$A = \frac{\lambda + 2\mu}{4\mu(\lambda + \mu)} \qquad B = -\frac{\lambda}{4\mu(\lambda + \mu)} \tag{2.14}$$

The third spatial dimension can easily be added by extending the vectors and matrices with the z-components. Equation 2.11 does not have to be changed for that. I, however, only write down the 2D representations because they are shorter and for the majority of my work I use 2D forward modeling.

2.1.2 Standard staggered grid (SSG)

In simple cases the system of equations derived in chapter 2.1.1 can be solved analytically. For complex media, however, a numerical approach is necessary. In most of this work finite differences

are used. This requires a discretization of the physical system in space and time. Model parameters λ, ρ and μ, particle velocities and stresses are all discretized on Cartesian grids. The grid spacing is called dh and the interval between two time steps is dt. The parameter dh is equal for all spatial dimensions. By writing the derivatives in equations 2.8, 2.9 and 2.10 as finite differences, Virieux (1986) showed that the spatial grids of v_x, v_y and σ_{xy} have to be staggered to ensure that the partial differential equations refer to the same grid point. The standard staggered grid in two dimensions is shown in figure 2.1. One can see that the stress components σ_{xx} and σ_{yy} and the model parameters are placed on full grid points. The x-component of the particle velocity v_x is shifted by half a grid point in x-direction and v_y is shifted by half a grid point in y-direction. The stress component σ_{xy} is shifted in both directions by half a grid point. Using second-order centered finite differences the time derivatives of σ and $\vec{v}$ can be approximated by

$$\left.\frac{\partial v_i}{\partial t}\right|^{n+\frac{1}{2}} \approx \frac{v_i^{n+1} - v_i^n}{\mathrm{dt}} \tag{2.15}$$

$$\left.\frac{\partial \sigma_{ij}}{\partial t}\right|^{n} \approx \frac{\sigma_{ij}^{n+\frac{1}{2}} - \sigma_{ij}^{n-\frac{1}{2}}}{\mathrm{dt}} \tag{2.16}$$

Regarding the notation, upper indices indicate the time step. Substituting the derivatives in 2.2 and 2.8 with the FD approximation yields the following scheme for updating the two parameters:

$$v_i^{n+1} \approx v_i^n + \frac{\mathrm{dt}}{\rho}\left.\frac{\partial \sigma_{ji}}{\partial x_j}\right|^{n+\frac{1}{2}} \tag{2.17}$$

$$\sigma_{ij}^{n+\frac{1}{2}} \approx \sigma_{ij}^{n-\frac{1}{2}} + \mathrm{dt}\,\lambda\,\frac{\partial}{\partial t}\,\mathrm{Tr}\,\epsilon\Big|^{n}\,\delta_{ij} + 2\,\mathrm{dt}\,\mu\left.\frac{\partial \epsilon_{ij}}{\partial t}\right|^{n} \tag{2.18}$$

The next step is to also discretize the spatial derivatives. In the following part indices i and j refer to full spatial grid points while the components of stress and velocity are listed explicitly for each dimension. In order to keep a consistent notation, indices which are used to address grid points are separated with a comma while indices which indicate a component of a tensor are not separated. By combining equations 2.17 and 2.18 with 2.10 one can write the discretized elastic wave equation as follows:

$$\sigma_{xx,i,j}^{n+\frac{1}{2}} = \sigma_{xx,i,j}^{n-\frac{1}{2}} + \Delta t \cdot \lambda_{i,j}\left(\left.\frac{\partial v_x}{\partial x}\right|_{i,j}^{n} + \left.\frac{\partial v_y}{\partial y}\right|_{i,j}^{n}\right) + 2 \cdot \Delta t \cdot \mu_{i,j}\left.\frac{\partial v_x}{\partial x}\right|_{i,j}^{n} \tag{2.19}$$

$$\sigma_{yy,i,j}^{n+\frac{1}{2}} = \sigma_{yy,i,j}^{n-\frac{1}{2}} + \Delta t \cdot \lambda_{i,j}\left(\left.\frac{\partial v_x}{\partial x}\right|_{i,j}^{n} + \left.\frac{\partial v_y}{\partial y}\right|_{i,j}^{n}\right) + 2 \cdot \Delta t \cdot \mu_{i,j}\left.\frac{\partial v_y}{\partial y}\right|_{i,j}^{n} \tag{2.20}$$

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

$$v_{x,i+\frac{1}{2},j}^{n+1} = v_{x,i+\frac{1}{2},j}^{n} + \frac{\Delta t}{\rho_{i+\frac{1}{2},j}}\left(\left.\frac{\partial \sigma_{xx}}{\partial x}\right|_{i+\frac{1}{2},j}^{n+\frac{1}{2}} + \left.\frac{\partial \sigma_{xy}}{\partial y}\right|_{i+\frac{1}{2},j}^{n+\frac{1}{2}}\right) \tag{2.22}$$

$$v_{y,i,j+\frac{1}{2}}^{n+1} = v_{y,i,j+\frac{1}{2}}^{n} + \frac{\Delta t}{\rho_{i,j+\frac{1}{2}}}\left(\left.\frac{\partial \sigma_{yx}}{\partial x}\right|_{i,j+\frac{1}{2}}^{n+\frac{1}{2}} + \left.\frac{\partial \sigma_{yy}}{\partial y}\right|_{i,j+\frac{1}{2}}^{n+\frac{1}{2}}\right) \tag{2.23}$$

Table 2.1: Taylor coefficients for second- to eighth- order FD operators.

Order N	$\beta_{1,N}$	$\beta_{2,N}$	$\beta_{3,N}$	$\beta_{4,N}$
2	1			
4	9 / 8	-1 / 24		
6	75 / 64	-25 / 384	3 / 640	
8	1225 / 1024	-245 / 3072	49 / 5120	-5 / 7168

One can see that in equations 2.21, 2.22 and 2.23 density and shear modulus need to be known on half grid points. Moczo et al. (2002) showed, that the accuracy of the finite-difference modeling can be increased by arithmetically averaging the density of two adjacent grid points and harmonically averaging the shear modulus of four adjacent grid points:

$$\rho_{i+\frac{1}{2},j} = \frac{\rho_{i,j} + \rho_{i+1,j}}{2} \qquad \rho_{i,j+\frac{1}{2}} = \frac{\rho_{i,j} + \rho_{i,j+1}}{2} \tag{2.24}$$

$$\mu_{i+\frac{1}{2},j+\frac{1}{2}} = \frac{4}{\mu_{i,j}^{-1} + \mu_{i+1,j}^{-1} + \mu_{i,j+1}^{-1} + \mu_{i+1,j+1}^{-1}} \tag{2.25}$$

Time indices are omitted here because model parameters are invariant in time.

2.1.3 FD operators

While for computational reasons time derivatives are only approximated by second-order FD operators, the spatial derivatives can be approximated by higher-order operators. These consider more surrounding grid points for approximating the derivatives which gives more exact results but comes also with more computational cost. The derivative of a variable f at grid point i can be approximated by:

$$\left.\frac{\partial f}{\partial x}\right|_i \approx \frac{1}{\mathrm{dh}} \left[\sum_{n=1}^{N} \beta_{n,N} \left(f_{i+\frac{1}{2}+n} - f_{i-\frac{1}{2}-n} \right) \right] . \tag{2.26}$$

The parameter N is the order of the FD-operator and $\beta_{n,N}$ are the coefficients of the operator. Note that the coefficients depend on N, so they have to be estimated for each order separately. This estimation can be done via Taylor series expansion (Jastram, 1992). The Taylor coefficients are listed in table 2.1 for orders ranging from 2 to 8.

2.1.4 Numerical stability

Finite-difference modeling uses a discretization in space and time. To minimize computational cost it is desirable to choose the distance between two adjacent points on the spatial grid dh and the time difference between two consecutive time steps dt as large as possible. However, these parameters cannot be arbitrary. They have to be chosen dependent of various other parameters, or else the simulation is numerically unstable. There are two kinds of instabilities. The first kind is grid dispersion and it

Table 2.2: Required points per wavelength for Taylor operators as calculated by Köhn (2011).

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

occurs if dh is too large. To avoid grid dispersion the following criterion has to be met (Köhn, 2011):

$$\text{dh} \leq \frac{\lambda_{\text{min}}}{n} = \frac{v_{\text{min}}}{n f_{\text{max}}} \tag{2.27}$$

This means that dh has to be smaller than n points per minimum wavelength λ_{min}. The minimum wavelength in turn depends on the maximum frequency of the source signal and the minimum seismic velocity of the medium. The required value of n depends on the order and type of the FD operator. For Taylor operators n is listed for different orders in table 2.2.
The second type of numerical instability is Courant instability. It occurs if the temporal sampling dt is too coarse. Courant et al. (1928) show, that in 2D dt has to meet the following criterion:

$$\text{dt} \leq \frac{\text{dh}}{h\sqrt{2} v_{\text{max}}} \tag{2.28}$$

This means, that dt has to be smaller than the time required for a wave to propagate between two adjacent grid points. The parameter h is the sum of the FD-coefficients, so it is dependent on the FD order:

$$h = \sum_{n=1}^{N} \beta_{n,N} \tag{2.29}$$

2.1.5 Initial and boundary conditions

The forward problem described by equations 2.8, 2.9 and 2.10 is ill-posed. This means it has many solutions. To obtain a unique solution initial and boundary conditions are needed. In the elastic case the initial conditions are

$$v_x|_{t=0} = v_y|_{t=0} = 0 \,, \tag{2.30}$$

$$\sigma_{xx}|_{t=0} = \sigma_{yy}|_{t=0} = \sigma_{xy}|_{t=0} = 0 \tag{2.31}$$

at every grid point (Virieux, 1986). This means the system is in equilibrium at time $t = 0$. Since we can not and also don not need to model the entire Earth in applied seismics, boundary conditions at the edges of the model are necessary. The top of the model is usually the Earth's surface which is a special kind of boundary. The other three edges of the model have to absorb the wavefield in order to mimic a quasi-infinite medium. Without this absorption the wavefield would be reflected at the model boundary which is not what happens in the real world. In the following, two boundary conditions are presented which absorb the wavefield and the free surface boundary is discussed in more detail.

Absorbing boundaries

A very easy way of damping the wavefield is a simple taper. Cerjan et al. (1985) suggested a taper of the form:

$$G_i = \exp\left(-[0.015(w_b - i)^2]\right) \qquad i \in [1, w_b] \ . \tag{2.32}$$

The parameter w_b is the boundary width in grid points. The procedure works by tapering the wavefield in the w_b outer most grid points of the model with coefficients G_i. At $i = 1$ the damping is strongest while at $i = w_b$ the wavefield remains unchanged. A much more effective damping method are convolutional perfectly matched layers (C-PML) introduced by Komatitsch and Martin (2007). They stretch the coordinates in the frequency domain in a way that incoming plane waves never reach the actual boundary but decay exponentially. Since this only works perfectly for exact solutions of the wave equation, some reflections occur when working with a discretized solution. To mitigate this effect additional damping is introduced. In practice C-PMLs require replacing the spatial derivatives described by equations 2.19 - 2.23 inside the boundary region.

Free surface

Above the free surface the particle displacement is not defined and the displacement of the free surface itself is not constrained like internal interfaces (Aki and Richards, 2002). This means, that the boundary condition is

$$\sigma_{xy} = \sigma_{yy} = 0 \ . \tag{2.33}$$

In case of a planar free surface the numerical implementation is often done via the stress image method first proposed by Levander (1988). This method is obtained by considering equations 2.2 and 2.3 and solving them under condition 2.33:

$$\sigma_{yy} = 0 = \lambda \frac{\partial u_x}{\partial x} + (\lambda + \mu) \frac{\partial u_y}{\partial y} \tag{2.34}$$

$$\sigma_{xy} = 0 = \mu \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right) \tag{2.35}$$

These equations can be fulfilled by extending the grid above the free surface by half the length of the FD operator and setting the stresses there accordingly. However, for an inclined free surface this is more complicated. Robertson (1996) generalized this approach for a topographic surface. However this approach requires the categorization of each surface grid node. Depending on that there are, in the 2D case, seven strategies how stresses and velocities have to be updated which reduces the efficiency of the forward calculation. Besides this approach there are many more. A much more elegant method was proposed by Zahradník et al. (1993) and Graves (1996). They suggested the discretization of space above the free surface as vacuum. Hence, this method is called the vacuum formulation. In vacuum all model parameters are zero and the free surface is treated like an internal interface. This has the advantage that no special treatment of the free surface is necessary which saves computational

time. The only requirement is that the grid is extended at the top by half the size of the FD operator. The downside of this formulation is that it is only stable for second-order FD operators (Graves, 1996) and that condition 2.33 is not completely fulfilled. To tackle this problem Zeng et al. (2012) proposed an improved vacuum formulation. It combines the suggestion by Mittet (2002) that the averaged shear modulus at free-surface grid nodes should be zero if any of the nodes participating in the averaging is zero (equation 2.25) with the averaging scheme by Moczo et al. (2002). Since density occurs in the denominator in equations 2.22 and 2.23 the averaging has to be adapted accordingly:

$$\rho^{-1}_{i+\frac{1}{2},j} = \begin{cases} 0 & \text{if } \rho_{i,j} = 0 \text{ and } \rho_{i+1,j} = 0 \\ \frac{2}{\rho_{i,j}+\rho_{i+1,j}} & \text{otherwise} \end{cases} \tag{2.36}$$

$$\rho^{-1}_{i,j+\frac{1}{2}} = \begin{cases} 0 & \text{if } \rho_{i,j} = 0 \text{ and } \rho_{i,j+1} = 0 \\ \frac{2}{\rho_{i,j}+\rho_{i,j+1}} & \text{otherwise} \end{cases} \tag{2.37}$$

$$\mu_{i+\frac{1}{2},j+\frac{1}{2}} = \begin{cases} \frac{4}{\mu^{-1}_{i,j}+\mu^{-1}_{i+1,j}+\mu^{-1}_{i,j+1}+\mu^{-1}_{i+1,j+1}} & \text{if } \mu_{i,j}, \mu_{i+1,j}, \mu_{i,j+1}, \mu_{i+1,j+1} \neq 0 \\ 0 & \text{otherwise} \end{cases} \tag{2.38}$$

This modification ensures that the traction-free boundary condition is fulfilled for an arbitrary free surface topography. It also stabilizes the modeling with higher-order FD operators. The proposed averaging moves the free surface upward by half a grid point as visualized in figure 2.2. This means that the only parameters which lay on the free surface are the stress component σ_{xy} and the particle velocities. All other parameters are either above the free surface where they are zero or in the elastic medium where they do not need to be considered for the free surface. Hence, the only condition to check is if $\sigma_{xy} = 0$ on all surface grid points. However, this is always the case since the averaging scheme sets $\mu_{i+\frac{1}{2},j+\frac{1}{2}}$ to zero on the free surface.

2.1.6 Spectral-element method

The finite-difference method used in the majority of tests in this work because of its simplicity has a major drawback. Interfaces between different media which are tilted relative to the regular grid are approximated by staircases. At strong interfaces like, e.g., the free surface this leads to significant scattering at the edges of the staircases. A method which is able to represent these interfaces more accurately is the spectral-element method since it fulfills the free-surface condition implicitly (Igel, 2017). Like in the finite-element method these elements can be skewed and have curved boundaries. Figure 2.3 shows a qualitative comparison of the free-surface representation in the FD model and in the SEM model. Inside each spectral element the wavefield is approximated by Lagrange polynomials. In combination with a special interpolation technique this simplifies the computation in a way that makes it very efficient on parallel hardware. While irregular elements that match the interfaces inside the model improve the accuracy of the method they also make model generation harder since most popular models are defined on regular grids. Generating good meshes for these models often requires professional software. For tests that involve the spectral-element method I use the software specfem2d which was initially developed by Komatitsch and Vilotte (1998) but nowadays has a large number of contributors. For instance the C-PMLs were implemented by Xie et al. (2014) and the

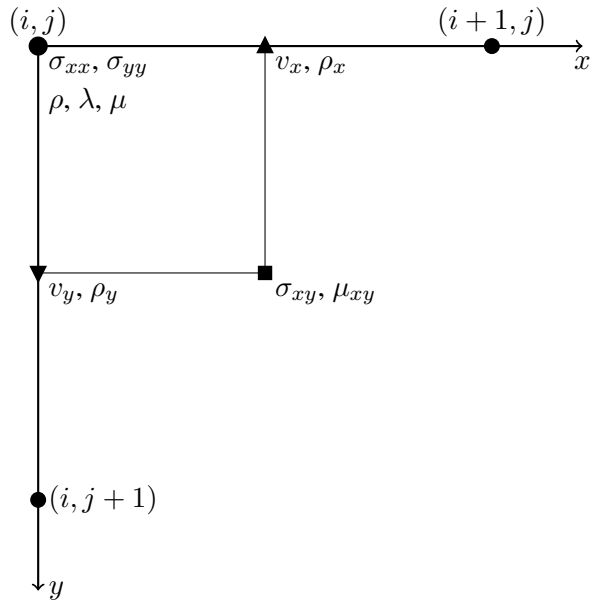

Figure 2.1: Standard staggered grid (SSG) as proposed by Virieux (1986). Some parameters are not defined on full grid points (i, j) but on half grid points which are shifted by dh/2.

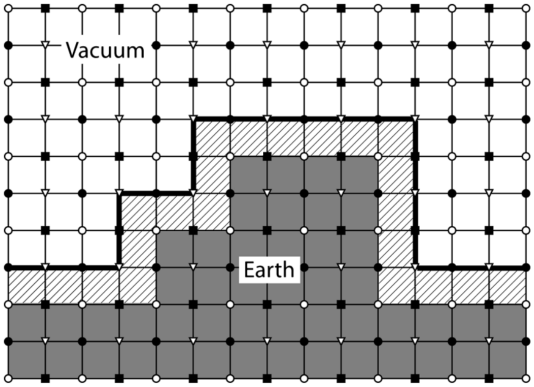

Figure 2.2: Shift of the free surface by the improved vacuum formulation as shown by Zeng et al. (2012). White circles represent full grid points where σ_{xx} and σ_{yy} are specified. Velocities v_x and v_y are located on black squares and black circles, respectively. The stress component σ_{xy} is located on white triangles.

SCOTCH partitioning method was implemented by Martin et al. (2008). Both these features are used during my tests.

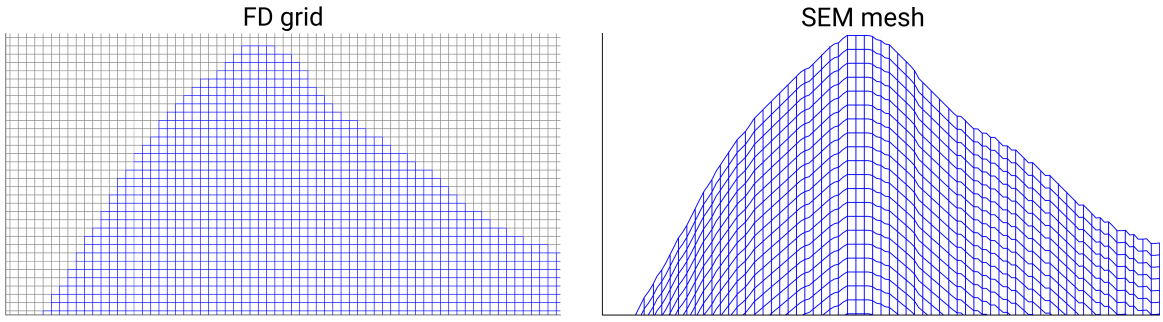

Figure 2.3: Qualitative comparison of the free surface in the FD model which is defined on a regular grid (left) and the SEM model which is defined on an irregular mesh (right). In the FD model grey represents the vacuum grid-points.

2.2 Full-waveform inversion

While the goal of the forward problem is to obtain data from a given model, the goal of the inverse problem is to find a model which explains the data. However, unlike the forward problem the inverse problem in seismics is highly ambiguous, i.e. the result of inversion is only one possible realization of a model which explains the data. Classical inversion methods like, e.g., traveltime tomography only use a fraction of the information contained in the data. The advantage of full-waveform inversion is

that it uses both amplitude and phase information of the recorded wavefield. In this section I first motivate the theoretical concepts of FWI and then show how it can be implemented as algorithm.

2.2.1 Misfit function

In section 2.1 I described how to solve the forward problem of seismic wave propagation. This problem can generally be expressed as applying a nonlinear forward operator $\mathfrak{F}$ on a model m which yields the synthetic data $\vec{d}_{syn}$:

$$\vec{d}_{syn} = \mathfrak{F}(m) \tag{2.39}$$

The model m is an adequate parametrization of the physical system under investigation. In my case it is $m = \{\rho; \lambda; \mu\}$. The goal of FWI is to find a model m_{opt} such that $\vec{d}_{syn}$ produced by this model matches the observed data $\vec{d}_{obs}$. Mathematically, the difference between observed and synthetic data can be expressed as a scalar misfit function $J(\vec{d}_{obs}, \vec{d}_{syn})$. In principle, the misfit function can be defined arbitrarily, in this case, however, I use the L_2 norm of the data residuals $\Delta\vec{d}$:

$$J(m) = \frac{1}{2} \int_{\mathbf{x}} \int_0^T (\Delta\vec{d})^2 \, \mathrm{d}t \, \mathrm{d}\mathbf{x} \tag{2.40}$$

$$\Delta\vec{d} = \vec{d}_{syn} - \vec{d}_{obs} \tag{2.41}$$

When using multiple shots equation 2.40 also includes a sum over all shots, so the total misfit is the sum off all misfits per shot. This means, finding m_{opt} is equivalent to finding the global minimum of J. For this type of problem there exist many algorithms which minimize J on a local or global scale. However since the computational effort to solve the forward problem is still relatively large, gradient-based algorithms are the only ones feasible. They start with an initial model m_0 and iteratively improve it by calculating the gradient of J with respect to m. This gives the following general rule for updating the model in each iteration (Fichtner, 2011):

$$m_{i+1} = m_i - \alpha \mathbf{A}_i \nabla_m J(m_i) \tag{2.42}$$

The parameter α is some step length which with the gradient is scaled and $\mathbf{A}$ is a positive definite matrix which depends on the optimization method used. Since $\mathfrak{F}$ is nonlinear, J is also nonlinear, which means by simply following the gradient it can easily happen that only a local minimum of J is found by this process. Approaches how this can be avoided are presented later. Another challenge is that equation 2.42 requires the calculation of the gradient. Calculating this gradient explicitly by changing each model parameter on each grid point by a tiny distortion δm and then evaluating the misfit is an impossible task because of the enormous number of model parameters and again the time required to solve the forward problem. The most efficient way of calculating the gradient is the adjoint method, which are presented in the next section.

2.2.2 Adjoint-state method

The adjoint-state method for calculating gradients of an objective function in a seismic context was first introduced by Tarantola (1986). I do not show the complete derivation of the gradients since they do not play a key role in the topic of my work. However, I would like to give a brief motivation how they can be calculated for a single shot using the method of augmented Lagrangian function following Plessix (2006). This method minimizes an objective function with a set of constraints. In this case the objective function is equation 2.40 and the constraints are the forward problem. Here the matrix-vector formulation of the forward problem 2.11 is used since it combines the first-order partial differential equations in velocity-stress formulation and one side of the equation is set to zero. This allows to construct the Lagrangian $\mathcal{L}$ by subtracting the constraints multiplied with the Lagrange multiplier $\vec{\psi}$ from the misfit function without changing the expression:

$$\mathcal{L}(\vec{\xi}, \vec{\psi}, m) = J(\vec{\xi}, m) - \vec{\psi} \cdot \vec{F}(\vec{\xi}, m) \tag{2.43}$$

$$= \frac{1}{2} \int_{\mathbf{x}} \int_0^T (\Delta \vec{d})^2 - \int_0^T \int_{\mathbf{x}} \vec{\psi} \left(M^{-1} \left(\frac{\partial \vec{\xi}}{\partial t} - \vec{b} \right) - \mathbf{Q} \vec{\xi} \right) \mathrm{d}t \, \mathrm{d}\mathbf{x} \tag{2.44}$$

Since $\vec{F}(\vec{\xi}, m)$ is zero by construction one can write

$$\mathcal{L}(\vec{\xi}, \vec{\psi}, m) = J(m) \; . \tag{2.45}$$

This means that finding the desired gradient $\frac{\mathrm{d}J}{\mathrm{d}m}$ is equivalent to finding $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}m}$. Applying the chain rule yields

$$\frac{\mathrm{d}\mathcal{L}(\vec{\xi}, \vec{\psi}, m)}{\mathrm{d}m} = \frac{\partial \mathcal{L}(\vec{\xi}, \vec{\psi}, m)}{\partial \vec{\xi}} \frac{\partial \vec{\xi}}{\partial m} + \frac{\partial \mathcal{L}(\vec{\xi}, \vec{\psi}, m)}{\partial m} \; . \tag{2.46}$$

The goal is now to eliminate the first summand which can be done by finding a $\vec{\psi}$ such that

$$\frac{\partial \mathcal{L}(\vec{\xi}, \vec{\psi}, m)}{\partial \vec{\xi}} = \frac{\partial J}{\partial \vec{\xi}} - \vec{\psi} \frac{\partial \vec{F}}{\partial \vec{\xi}} = 0 \; . \tag{2.47}$$

I skip solving this equation step by step and jump right to its solution:

$$\mathbf{M}^{-1} \frac{\partial \vec{\psi}(\tau)}{\partial \tau} - \Delta \vec{d} = -\mathbf{Q} \vec{\psi}(\tau) \tag{2.48}$$

$$\tau = T - t \tag{2.49}$$

One can see that equation 2.48 looks very similar to the forward equation 2.11 with the differences being the signs and that the sources are now the data residuals. This equation is called the anti self-adjoint equation. Hence $\vec{\psi}$ is the solution to the forward problem with sources $\Delta \vec{d}$, they are called adjoint wavefield and adjoint sources, respectively. The introduction of τ means that this forward problem has to be solved backward in time with terminal conditions equal to the initial conditions of the traditional forward problem. In order to use the same implementation for calculating $\vec{\psi}$ as for

calculating $\vec{\xi}$ the signs have to be replaced by substitution:

$$\vec{\psi} = \mathbf{B}\vec{\Phi}\,, \qquad \mathbf{B} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 \end{pmatrix} \tag{2.50}$$

After finding $\vec{\psi}$ the last step is to use equation 2.43 to write the gradients as

$$\frac{\partial J}{\partial m} = -\mathbf{B}\vec{\Phi} \int_0^T \left(\frac{\partial \mathbf{M}^{-1}}{\partial m} \frac{\partial \vec{\xi}}{\partial t} - \Delta\vec{d} \right) \mathrm{d}t\,. \tag{2.51}$$

When using multiple shots, the gradients of all the individual shots are summed to form the final gradient.

2.2.3 Additional methods used in FWI

Step-length estimation

As already shown in equation 2.42, the gradient of the misfit function needs to be scaled with a scalar α. This scalar is called the step-length since it determines how large the allowed model updates are. The step-length has to be chosen carefully because choosing a step-length too large results in huge jumps through the parameter space of the misfit function so that the actual minimum is never reached. On the other hand, if the step-length is too small there is a very slow convergence and since FWI is computationally very expensive this is not desirable either. Another complication is that different iterations of FWI need different step-lengths. At the beginning when the misfit is large there should also be a large step length to invert the general structure of the model. At the end of FWI, when only small features need to be inverted a smaller step length is required. Hence in every iteration a search for the optimal step length is conducted. This step length estimation is based on a parabolic line search suggested by Nocedal and Wright (1999). This algorithm theoretically requires two additional forward calculations to calculate a parabolic fit. However, I use an improved version where we use more forward calculations for a better fit but do not use all shots.

Conjugate-gradient method

Instead of always updating the model with equation 2.42 I use the preconditioned conjugate-gradient method (PCG) by Nocedal and Wright (1999). This method uses equation 2.42 only in the first iteration. Subsequently, it updates the model via:

$$m_{i+1} = m_i - \alpha_i \mathbf{A}_i \nabla_m \bar{J}_i \tag{2.52}$$

$$\nabla_m \bar{J}_i = \nabla_m J_i - \beta_i \nabla_m \bar{J}_{i-1} \tag{2.53}$$

Various authors have proposed different choices of β_i. In the software I use the definition by Polak and Ribière (1969) is implemented:

$$\beta_i = \frac{(\nabla_m J_i)^T (\nabla_m J_i - \nabla_m J_{i-1})}{||\nabla_m J_{i-1}||^2} \tag{2.54}$$

The preconditioning matrix $\mathbf{A}$ is calculated with the K1 method by Plessix and Mulder (2004).

Source time function inversion

In many cases, e.g., if explosive sources are used, the source signature is unknown. To overcome this problem I use the source time function inversion (STF) following Pratt (1999). The conceptual starting point of this technique is the modified discrete forward-modeling equation in the frequency domain:

$$\mathbf{S}\vec{\tilde{u}}(\omega_k) = s(\omega_k)\vec{\tilde{f}}(\omega_k) \tag{2.55}$$

The tilde represents the Fourier transform of the corresponding quantities and $\mathbf{S}$ is a complex-valued impedance matrix which represents the model parameters. In the elastic case all parameters except $\mathbf{S}$ depend on the discrete frequencies ω_k. The goal is to find the complex scalar s such that its product with $\vec{\tilde{f}}$ corresponds to the real source signature at ω_k. This is achieved by minimizing the misfit function 2.40 which can be done analytically. The result is:

$$s(\omega_k) = \frac{\vec{d}^T \vec{u}^*}{\vec{u}^T \vec{u}^* + \varepsilon} \tag{2.56}$$

The asterisks in equation 2.56 represents the complex conjugate. The parameter ε is not part of the analytical solution but has been added to stabilize the equation if the denominator is small. This so-called water level can help significantly, especially in the presence of noise.

Multi-scale approach

A common problem of gradient-based optimization as in FWI is that the algorithm converges towards a local minimum. While almost never the true global minimum is found, the goal should be to find at least a local minimum in the proximity of the global minimum. One way to avoid undesirable local minima is choosing a good starting model m_0 so there is a high chance that by going down the gradient of the misfit function the global minimum is reached. In practice this is rarely a viable option because the knowledge of the site under investigation is usually very limited. One can, however, do not only one but multiple FWIs consecutively and use that smooth models generate smooth misfit functions (Fichtner, 2011). Each of these FWIs is called a stage of the FWI workflow and uses a certain subset of the data. By starting to invert the low-frequency content which is only sensitive to the large-scale variations in the model one can obtain a model $\tilde{m}$ which is closer to the true model than m_0 and hopefully also the global minimum of that frequency range. After that higher frequencies are introduced stage by stage. The idea is that these new frequencies compared to the previous stage

modify the misfit function only so much that $\tilde{m}$ still lies near the new global minimum which can then be found by another FWI.

2.2.4 FWI algorithm

After introducing the theoretical concepts of FWI the purpose of this section is to combine them to form the FWI algorithm. A visual representation of the simplified FWI algorithm is shown in figure 2.4. In this figure there are two starting points, the true model m_{true} and the starting model m_0. In the real world the true model is unknown but we can measure it by acquiring seismic data $\vec{d}_{obs}$. The synthetic data $\vec{d}_{syn}$ belonging to m_0 can be obtained by forward modeling (see section 2.1). During this forward modeling the preconditioning matrix $\mathbf{A}$ is also calculated. By subtracting $\vec{d}_{obs}$ and $\vec{d}_{syn}$ one obtains the data residuals $\Delta\vec{d}$ which are used in the misfit function J described in section 2.2.1 to get a scalar value quantifying the difference between the data sets. With this misfit one has to decide if the data fit is good enough. If not m_0 has to be updated. For this the gradient of J has to be calculated according to section 2.2.2 which requires the forward modeling of the adjoint wavefield. These gradients are calculated for each shot, then multiplied by $\mathbf{A}$ and summed. From this summed gradient the conjugate gradient $\nabla_m \bar{J}$ is calculated. With $\nabla_m \bar{J}$ an optimal step-length α is being searched. In the last step the product of α and $\nabla_m \bar{J}$ is subtracted from m_0 and the next iteration starts with the new model m_1.

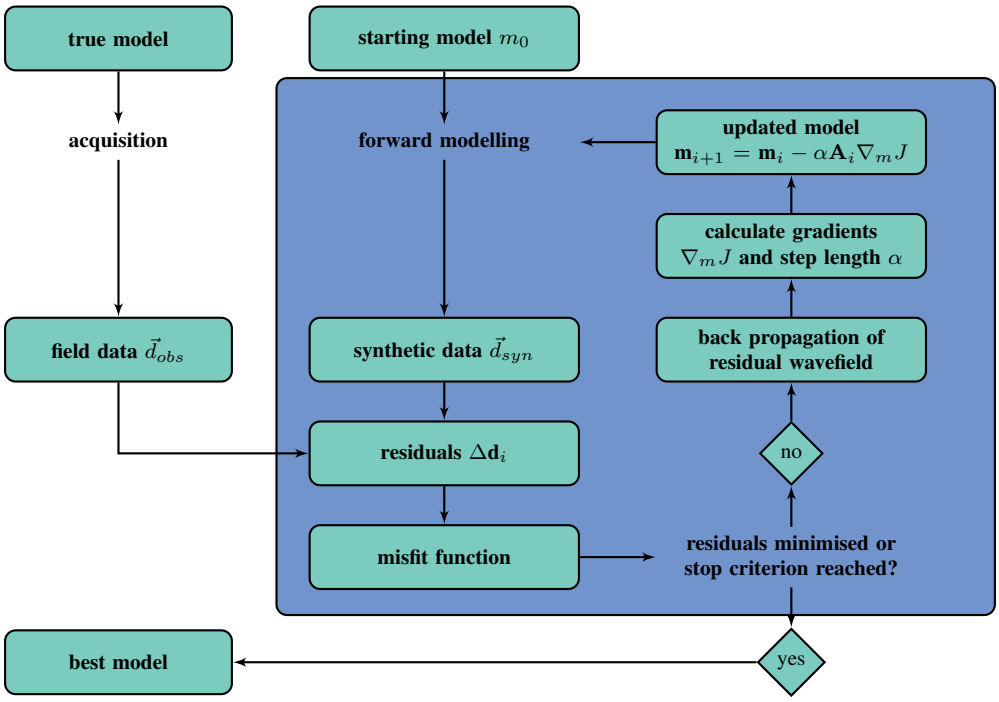

Figure 2.4: Workflow of the FWI process. The blue box indicates the loop over the iteration steps. In the case of a purely synthetic FWI the acquisition is replaced by a forward modeling.

Chapter 3

Model

The models I use in my work are all part of the 2D synthetic model by Gray and Marfurt (1995) which is known as the Canadian Foothills model. In its original form the model describes only the distribution of v_P because it was intended for acoustic modeling with constant density. This distribution is shown in figure 3.1. The highest velocity is $5900\,\mathrm{m\,s^{-1}}$ and the lowest velocity is $3500\,\mathrm{m\,s^{-1}}$. It is discretized with a grid spacing of $10\,\mathrm{m}$ in vertical direction and $15\,\mathrm{m}$ in horizontal direction. The structure of the model can be described as a linear velocity gradient in vertical direction overlayed with complex features of various sizes and shapes. These features have different, but mostly constant, velocities. The structure resembles the geology beneath the foothills of the Canadian Rocky Mountains in the area of northeastern British Columbia. The dimensions of the model are $25\,\mathrm{km}$ in length and $10\,\mathrm{km}$ in depth with the sea level located $2\,\mathrm{km}$ below the top of the model. The free-surface, which is the uppermost interface, shows strong topographic variations with height differences of up to $1.5\,\mathrm{km}$. In the original definition of this model all velocities above the top interface are set to $4000\,\mathrm{m\,s^{-1}}$. This is convenient for acoustic modeling because it minimizes the interaction of the wavefield with this interface. In my work this interaction is very important so, according to the vacuum formulation, the velocities are changed to $0\,\mathrm{m\,s^{-1}}$. Since I use elastic modeling in this work, the remaining parameters have to be derived from the P-wave velocity. A common way to get the S-wave velocity is to assume a constant Poisson's ratio ν. I choose ν to be 0.25 which is a common value for crystalline rocks like, e.g., granite and limestone (Gercek, 2007). The relation between ν and the Lamé parameters is

$$\nu = \frac{\lambda}{2(\lambda + \mu)} \,. \tag{3.1}$$

Together with equations 2.4 a Poisson's ratio of 0.25 results in the relation

$$\frac{v_P}{v_S} = \sqrt{3} \,. \tag{3.2}$$

To derive the density I use the empiric Gardner relation (Gardner et al., 1974) which is

$$\rho = 310 \frac{\mathrm{kg\,s^{\frac{1}{4}}}}{\mathrm{m^{\frac{13}{4}}}} v_P^{\frac{1}{4}} \,. \tag{3.3}$$

The v_S and ρ models obtained with these relations are also shown in figure 3.1. Since they are both based on the v_P model all parameter models are spatially correlated.

The model is accompanied by the pressure seismograms modeled with the acoustic approximation by Gray and Marfurt (1995). I will not use these seismograms because they do not account for most phases found in seismograms one would record in the real world. It can be expected that surface waves, which do not occur when the acoustic wave equation is used, dominate the recorded seismograms in such a setting. Removing them with sufficiently small error from the seismograms in a way which is also applicable during FWI is an impossible task. Beyond that, using only phases which arrive before the surface waves is also no valid strategy because in complex media conversions between P- and S-waves have an effect which is not negligible. However, I will use the same receiver spacing like the authors throughout all my tests, i.e. $15\,\mathrm{m}$. The original source spacing is $90\,\mathrm{m}$. In my work I use an increased source spacing of $150\,\mathrm{m}$ to save computing time.

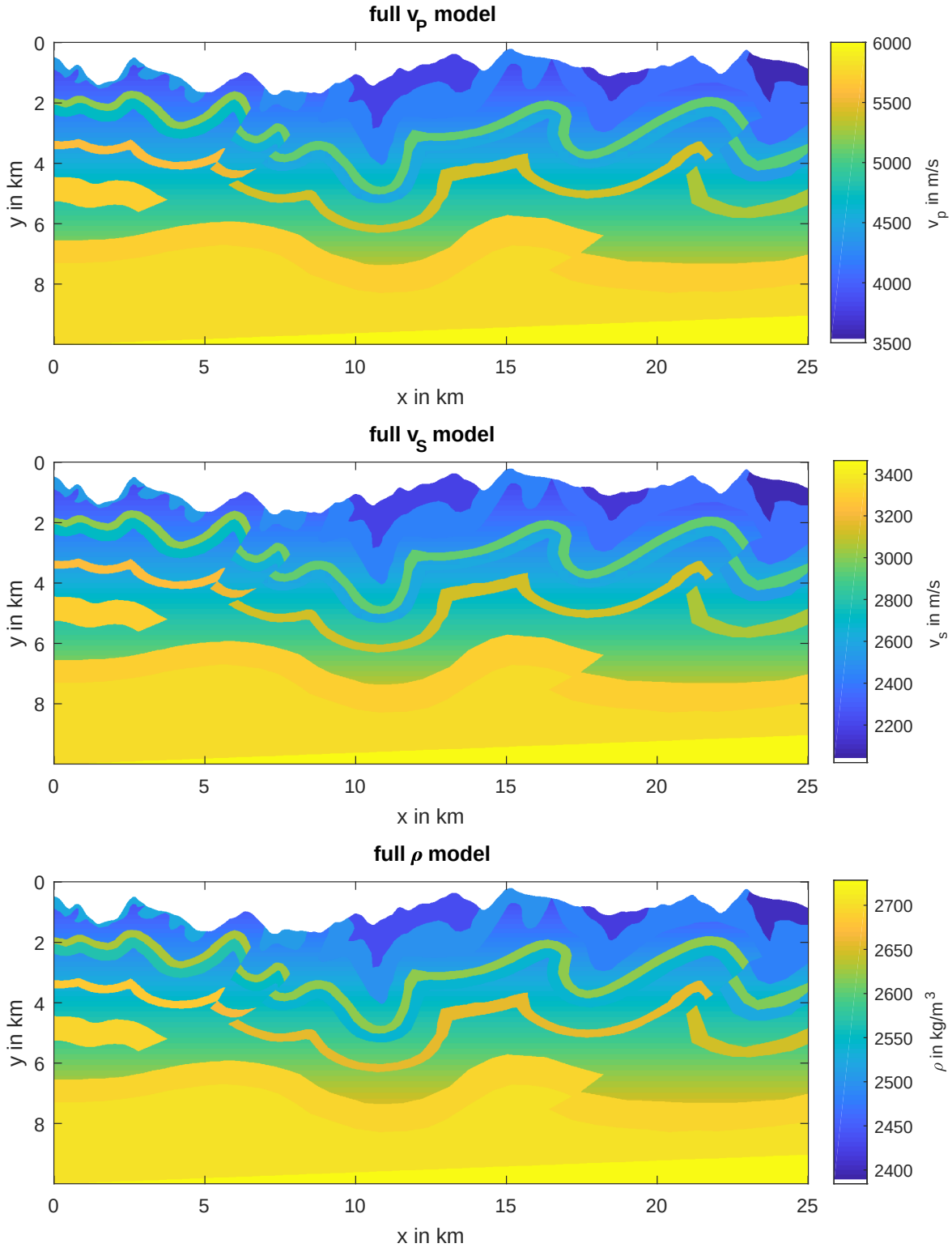

Figure 3.1: The Canadian Foothills model which in its original form only describes the v_P distribution but has been extended for elastic modeling using equations 3.2 and 3.3.

Chapter 4

Forward modeling

The goal of this chapter is to analyze how well a 2D FD solver can model the wavefield in a medium with strong surface topography on the scale of exploration seismics. This analysis is split into three parts. In the first part I test the validity of the 2D FD method using the improved vacuum formulation (IVF) on a scale of exploration seismics. The IVF as presented in section 2.1.5 was introduced by Zeng et al. (2012) in a near-surface context. Such near-surface studies are usually conducted with a large number of grid points per wavelength and only a small number of wavelengths are propagated. This means, the applicability in the setting of exploration seismics is not guaranteed. This test is done with a planar free surface. The goal of the second test is to analyze the influence of a surface with topographic variations. The question that should be answered is how discretization and large offsets affect the accuracy of the FD simulation in such a setting. Discretization plays an important role because FWI is computational already very costly. Therefore, the discretization should be as coarse as possible. In the last part of this chapter I answer the question how well 3D wave propagation found in the real world can be considered a 2D problem in the case of no variation of the model in the third dimension and in case of variations of the free surface.

4.1 Comparison to analytical solution

The goal of this test is to compare the FD simulation using the IVF with an analytical solution to investigate if Rayleigh waves are modeled correctly at large offsets. Analytical solutions for seismic wave propagation are only possible with very simple models. The procedure to obtain an analytical solution is to calculate the Green's function describing the model and then convolving it numerically with a source wavelet.

4.1.1 Methodology

The model I consider in this test is that of a 2D elastic half-space with a free surface. For my tests I use the implementation by Berg (1994). This program computes the response of a vertically-oriented directional point source. The parameters of the half-space are chosen similar to the parameters found

near the surface of the Canadian Foothills model. They are $v_P = 3800\,\mathrm{m\,s^{-1}}$, $v_S = 2200\,\mathrm{m\,s^{-1}}$ and $\rho = 3000\,\mathrm{kg\,m^{-3}}$. Before setting up the acquisition geometry, one has to define the discretization of the FD model. In order to calculate the spatial discretization needed, all parameters required in equation 2.27 have to be defined. The minimum velocity is the S-wave velocity of $2200\,\mathrm{m\,s^{-1}}$ which is approximately the minimum velocity of the Canadian Foothills model described in chapter 3. Since this test is done with a future FWI application in mind and FWI works best with low frequencies, I choose a time-shifted Ricker wavelet with a $5\,\mathrm{Hz}$ center frequency f_c as my source signal. The time-shifted Ricker wavelet s_R with amplitude A is defined as:

$$s_R = A(1 - 2\tau^2)e^{(-\tau^2)} \tag{4.1}$$

$$\tau = \pi f_c \left(t - \frac{1.5}{f_c}\right) \tag{4.2}$$

This wavelet and its amplitude spectrum are shown in figure 4.1. By looking at the amplitude spectrum one can observe that the maximum frequency is about $15\,\mathrm{Hz}$. Besides, I choose a fourth-order FD operator for all of my tests. With equation 2.27 the resulting spatial discretization needed is about $18\,\mathrm{m}$. As another limiting factor Zeng et al. (2012) show that the required number of grid points per wavelength is larger for correct surface wave simulations using the IVF compared to traditional body-wave simulations. Beyond that, the discretization should allow for lower velocities than the ones in the model because during inversion it uncertain if the minimum velocity of the true model is not undercut. Because of all these circumstances, I choose a spatial discretization of dh $= 7.5\,\mathrm{m}$. This grid spacing has the additional advantage that no receiver interpolation is needed for a receiver spacing of $15\,\mathrm{m}$. By using this values, equation 2.28 yields a temporal sampling of dt $= 1\,\mathrm{ms}$. Because the models in later tests have velocities up to about $6000\,\mathrm{m\,s^{-1}}$, I already set my temporal sampling in this test to dt $= 0.7\,\mathrm{ms}$. This reduces the grid dispersion even further and also makes the tests more comparable. At this point I wold like to point out that only in recent years frequencies lower than $5\,\mathrm{Hz}$ can be acquired on land (Mahrooqi et al., 2012). However, results do not depend directly on frequency but on grid points per wavelength. So in case only high frequencies are available, the same results can be achieved by using a finer discretization to balance the grid points per wavelength. Since the ultimate goal is to invert the Canadian Foothills model, the acquisition geometry of this test should match that needed for the model. The vertical size of the model is $10\,\mathrm{km}$, so I choose a maximum offset which is similar to that value. For a good visualization I choose 28 receivers starting at $375\,\mathrm{m}$ offset with a spacing of also $375\,\mathrm{m}$. The resulting maximum offset is $10.5\,\mathrm{km}$. The acquisition geometry is located one grid point below the free surface because the free surface is shifted upward by half a grid point by the IVF and the vertical component of my receivers is shifted downward by half a grid point because of the staggered grid. Placing the receivers on the free surface is not an option in the FD case because then the horizontal component would be inside the vacuum. To mimic an infinite half-space in the FD simulation I extend the model at both sides in the horizontal direction by a buffer zone with a width of 20 grid points and then C-PMLs with a width of another 20 grid points. The depth of the model is set to $3750\,\mathrm{m}$. This is beyond the penetration depth of surface waves. So the total grid size is 1481 $\times$ 500 grid points. The 20 grid points at the bottom of the model are also part of a C-PML in order to dampen reflections from the bottom edge. As time interval for the simulation I choose $6\,\mathrm{s}$, so the number of time steps is 8572. This way the Rayleigh wave is recorded at all receivers.

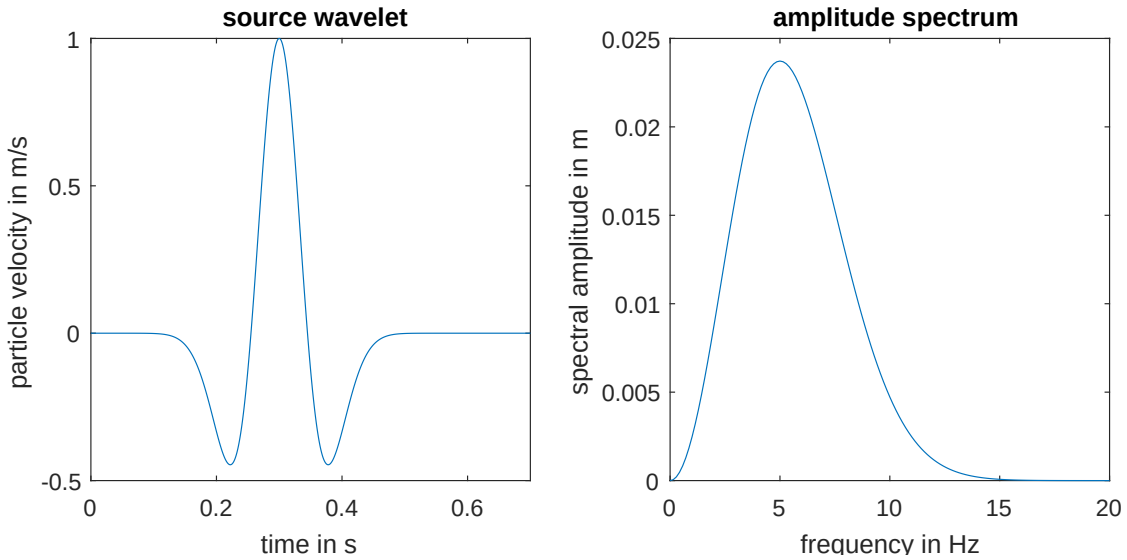

Figure 4.1: Left: A Ricker wavelet defined by equation 4.1 with a center frequency of 5 Hz. Right: Its amplitude spectrum in the range from 0 Hz to 20 Hz.

4.1.2 Wavefield discussion

In this section I discuss the simulated wavefields for a better understanding of the phases found in the seismograms. Instead of showing the snapshots of the velocity fields, I show their spatial divergence and curl. This has the advantage that the divergence is related to the compressional energy (E_P) and the curl is related to the shear energy (E_S). Following Morse and Feshbach (1953) and Dougherty and Stephen (1988) they can be calculated as follows:

$$E_P = (\lambda + 2\mu)(\text{div } \vec{v})^2 \tag{4.3}$$

$$E_S = \mu(-\text{curl } \vec{v})^2 \tag{4.4}$$

Figure 4.2 shows the two wavefield components after 1 s and 3 s. Since the Rayleigh wave is a P-SV wave it contains compressional and shear energy, and hence it is visible in both wavefield components shown here. Unique to this setup is that in a homogeneous half-space the Rayleigh wave is not dispersive. The dispersive property arises only in media with depth-dependent velocity. Furthermore, one can observe that the P-wave separates quickly from the Rayleigh wave because of the velocity contrast between v_P and v_S while, even after 3 s, the Rayleigh wave and S-wave have barely separated. In case the relation between v_P and v_S can be described by equation 3.2, which is the case in all of my tests, the Rayleigh wave propagates with a velocity of 92% of v_S (Müller, 2007). It can also be observed that the amplitude of the Rayleigh wave decays much slower than that of the body waves. Theory predicts that the amplitude of the Rayleigh wave decays $\propto \frac{1}{\sqrt{r}}$ with distance r while that of body waves decays $\propto \frac{1}{r}$ (Aki and Richards, 2002). A positive consequence in the real world is that Rayleigh waves have a better signal-to-noise ratio at far offsets compared to body waves. The last wave type observable is the P-S wave. It is a refracted S-wave excited by the P-wave at the free surface.

4.1.3 Results

The vertical velocity component of the receivers is displayed in figure 4.3. To identify the visible phases, it is useful to look at the snapshots of the wavefield as shown in figure 4.2. The most prominent phase is the Rayleigh wave while the P-wave and S-wave are only faintly visible. Regarding the P-wave, the reason for this is that it propagates and oscillates parallel to the free surface, and hence cannot be recorded with the vertical component of a receiver. Furthermore, the source radiates P-waves mostly in the vertical direction which also affects their recorded amplitudes.
From figure 4.3 one can observe that the difference between the FD and the analytical solution is not very significant compared to the amplitude of the Rayleigh wave. A possible explanation for the difference is the fictitious layer caused by the IVF since the difference does not seem to be a product of grid dispersion or violation of the CFL condition. These two reasons can be excluded because the phase would be shifted consistently and increasing with offset towards either earlier or later arrival times than the analytical solution. The conclusion drawn from this test is that a grid spacing of $7.5\,\mathrm{m}$ and a time step interval of $0.7\,\mathrm{ms}$ is sufficient to model Rayleigh waves at large offsets if there is no topography present.

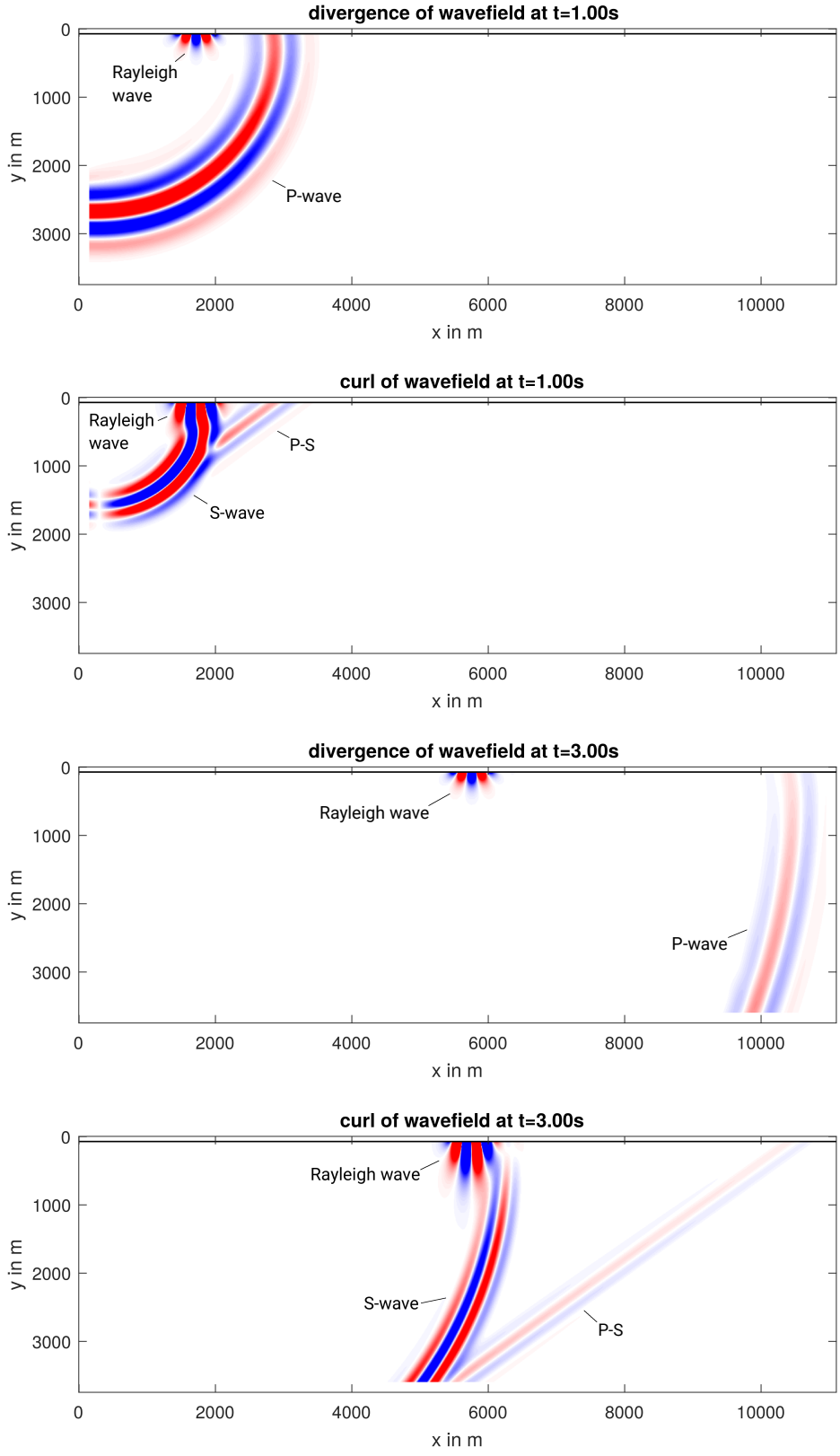

Figure 4.2: Spatial divergence and curl of the FD velocity wavefield for two points in time. The different phases visible in each plot are labeled accordingly. The horizontal black line near $y = 0\,\mathrm{m}$ is the free surface. I show no legend for the colors because they are clipped for better visualization and my simulations are not conducted with true physical units.

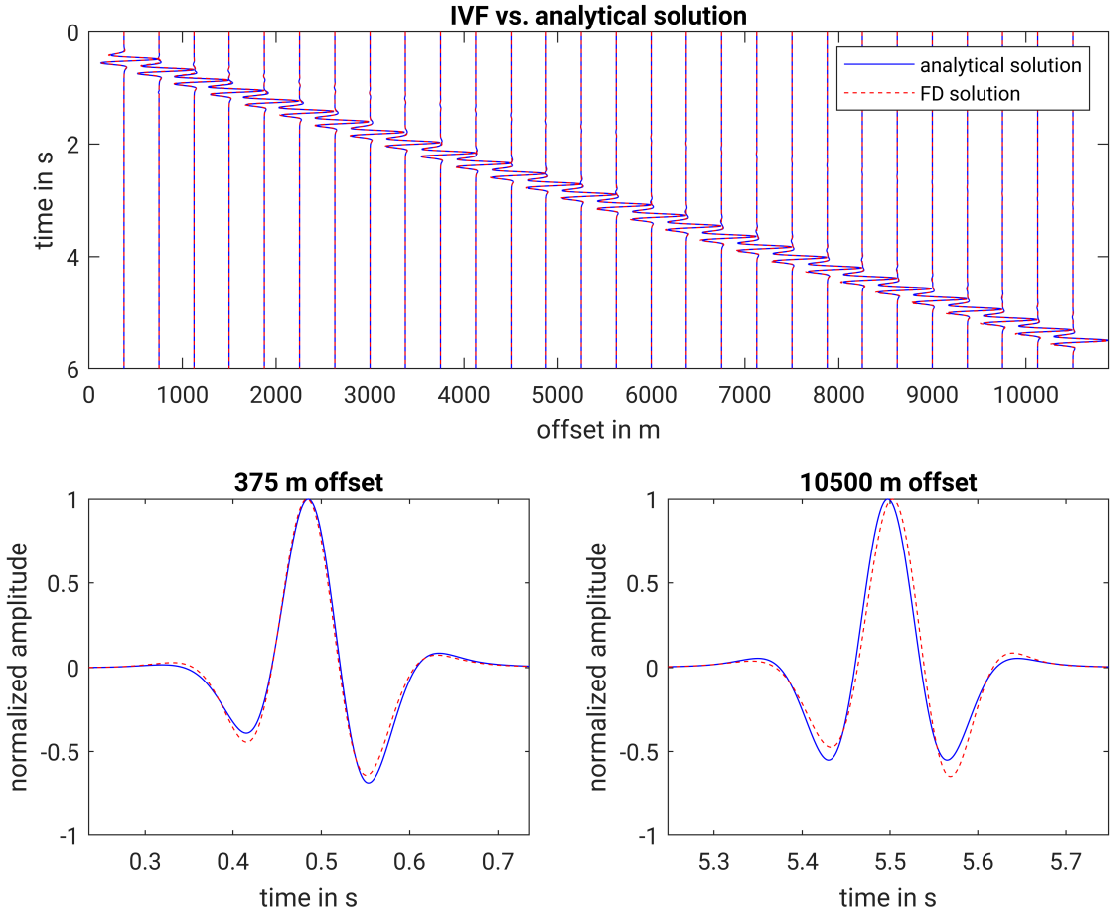

Figure 4.3: Top: The vertical velocity components of all 28 receivers calculated analytically and by FD simulation with IVF. All traces are normalized to their maximum. The prominent event visible is the Rayleigh wave. Bottom: A zoom in on this wave for the nearest offset in the left and for the farthest offset in the right.

4.2 Comparison to spectral element solution

After investigating the requirements to model Rayleigh waves at large offsets with a flat free surface, the next step is to check the requirements needed in the presence of topography. In complex media there is no analytical solution available and I have to use another numerical method as a reference. The numerical method I use as a reference is the spectral element method. The spectral element method is known to produce accurate results in case of surface topography because the shape of the elements can be adapted to the shape of the free surface and the free surface condition is fulfilled implicitly. The accurate results come, however, at the cost of a more challenging implementation and a more complex model generation procedure. Generating good spectral element meshes requires, especially in complex models, a lot of work and expertise. In my experience the spectral element method has a higher computational cost than the FD method if the mesh is not perfectly adapted to the model. Hence, the goal of this test is to find the minimum requirements to model Rayleigh waves sufficiently accurately in this scenario and then to decide, weather these requirements are acceptable for an FWI application.

4.2.1 Methodology

In order to test the IVF in the presence of topography I use the spectral element method as a reference. Since it is challenging to ensure that the models used for the two types of forward modeling are equivalent, I generate a simplified version of the Canadian Foothills model. The simplified version reduces the leftmost $7.8\,\mathrm{km}$ of the model to its most characteristic features. These features are in some cases also extended to form a layered model. This allows for an easy parametrization of the interfaces which are needed to generate the irregular mesh for the spectral element method. The original part of the model and the simplified version are shown in figure 4.5. The layers are generated by picking a few points where an interface is located and then using spline interpolation to get the y-values for the remaining grid points. These interfaces are then modified to be flat near the edges of the model. This improves the damping by the C-PMLs. Damping is crucial in this test because reflections at the boundaries which occur in only one of the two methods can distort the result. The exact model parameters of the simplified model are listed in table 4.2. Since most of the features in the Canadian Foothills model have a constant velocity, the velocities of the corresponding features in the simplified model are equivalent. In the other regions where there is a velocity gradient, I chose a single velocity which represents that section on average reasonably well. The parameters v_s and ρ are derived with equations 3.2 and 3.3, respectively.

For FD modeling I generated four versions of the simplified model with different discretizations via different spline interpolations of the interfaces, starting with the interfaces defining the SEM model. The spline interpolation is used because it generates smoother interfaces the finer the chosen discretization. The chosen discretizations are dh $= 3\,\mathrm{m}$, dh $= 3.75\,\mathrm{m}$, dh $= 5\,\mathrm{m}$ and dh $= 7.5\,\mathrm{m}$. These values all have in common that they do not require receiver interpolation at a receiver spacing of $15\,\mathrm{m}$. The FD simulation is done with a 6th-order FD operator. The spectral element model is discretized with a constant number of elements in the x-direction and a constant number of elements per layer in the y-direction. The number of element in x-direction is 520 and the number of elements

in y-direction in each layer is listed in table 4.2. The total number of elements along y is 380. These numbers are chosen in a way such that there are smaller elements near the surface and larger elements in the lower parts. With this number of elements the solution is definitely converged. There is a good chance that the number of elements can be reduced significantly but the priority of this test is on making sure that the FD model and the SEM model are, in fact, equivalent. For this reason I would like to keep the size of the spectral elements comparable to the size of the FD grid spacing.
The source is a vertical force placed on the surface in the center of the model. As source signal I choose again the $5\,\mathrm{Hz}$ time-shifted Ricker wavelet. This wavelet has a maximum frequency of approximately $15\,\mathrm{Hz}$ (see figure 4.1). For this test I choose a simulation time of $5\,\mathrm{s}$ with a time step interval of dt $= 0.1\,\mathrm{ms}$. The simulation time is shorter than in the previous test because of the shorter offsets. Furthermore, dt is smaller because, given the spatial discretization, it is required by the SEM. The width of the C-PMLs in the spectral element model is 10 elements and the length of one element is approximately $15\,\mathrm{m}$. The width of the boundaries in the FD simulation is chosen such that it is the same. While the C-PMLs do not behave the same in the two methods, they are thick enough so that edge reflections are almost completely dampened in both cases. Reflections within the boundary zone play no role since they are outside the scope of the acquisition geometry and also dampened by the C-PMLs. Regarding the source and receiver positions in the SEM model, the shift of the free surface introduced by the IVF and the shift of the v_y receivers by the staggered grid have both to be taken into account. To be precise, receivers cannot be placed at the exact same location in both methods because the SEM automatically places the receivers on the most appropriate control node inside an element, and hence the exact position is unknown. However, this location error is much smaller than dh because one element has an edge length of approximately $15\,\mathrm{m}$ and I use 9-node elements. Using the minimum velocity in the model and the maximum frequency of the source signal, table 4.1 shows the chosen discretizations in terms of minimum points per wavelength (ppw) for both P- and S-waves. Measuring discretization in ppw has the advantage that it unifies the source signal frequency and the velocity of the medium which makes it easier to compare discretizations of FD simulations on different scales.

Table 4.1: Chosen spatial discretizations in terms of minimum points per wavelength (ppw) for P- and S-waves.

dh in m	minimum ppw (P)	minimum ppw (S)
7.50	36	21
5.00	53	31
3.75	71	41
3.00	89	51

4.2.2 Wavefield discussion

In the presence of topography the dominant phases of the wavefield are mostly caused by interactions with the free surface. To visualize the wavefields I again use equations 4.3 and 4.4 which separate the compressional and the shear component. Figure 4.4 shows that, except for the direct and P-S converted waves, all visible phases are surface waves or converted from surface waves. It can be

observed that the mountain peak located at x≈2800 m acts as a diffraction point where P- and S-waves are being scattered. Phase number 6 is a back-scattered surface wave coming from the right and reflected from the left side of the mountain. Phase number 7 is a doubly scattered Rayleigh wave, which is reflected first at the left and then at the right flank of the mountain, and hence propagating to the left again but with a small delay compared to the original Rayleigh wave. Phase number 10 is an S-wave converted from the Rayleigh wave when it propagated down the left mountain flank and reached the base which acts as another diffraction point. Since the compressional content of the Rayleigh wave is smaller than the shear content the P-wave equivalent of this wave is not directly visible. As the sixth layer has a high velocity compared to the adjacent layers there is a guided P-wave observable (phase number 8). All in all, there are a lot of high-amplitude surface-wave phases observable which partly overlap each other. This makes the identification of body-wave phases in the seismograms impossible after the first arrival of the Rayleigh wave.

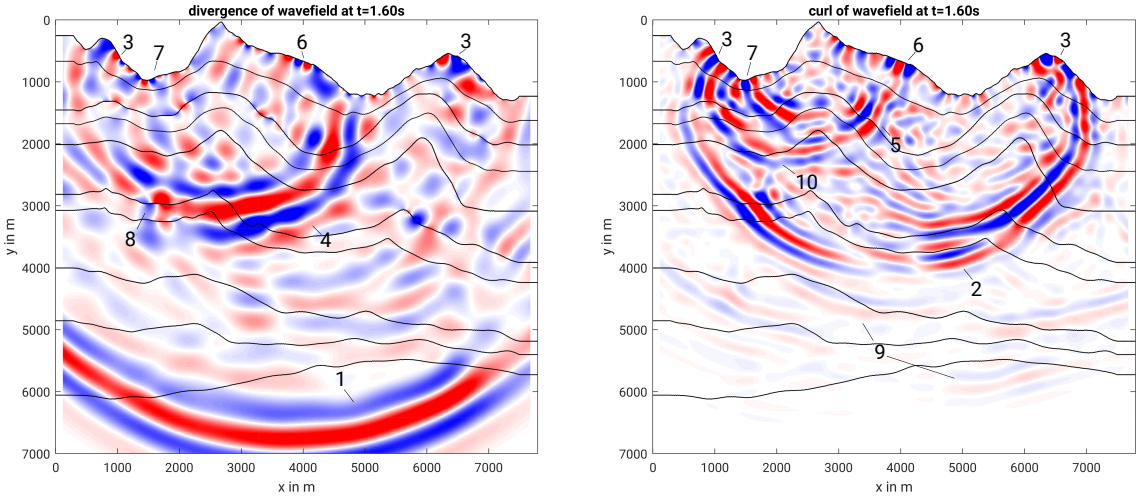

Figure 4.4: Spatial divergence and curl of the FD velocity wavefield after 1.6 s with the contour of the underlying velocity model. The source is located in the center of the model at the surface and the model is discretized with dh = 3 m and dt = 0.1 ms. Amplitudes are clipped such that the phases in each plot are visible in the best way. 1: Direct P-wave, 2: Direct S-wave, 3: Rayleigh wave, 4: P-wave diffracted at mountain peak, 5: S-wave diffracted at mountain peak, 6: Backscattered Rayleigh wave, 7: Doubly scattered Rayleigh wave, 8: Guided P-wave, 9: P-S converted waves, 10: S-wave diffracted at mountain base.

4.2.3 Results

Figure 4.7 shows a detailed comparison of some selected traces. For a broad overview with less detail see the seismograms in figures A.1 to A.4 in the appendix. In all seismograms the phase with the highest amplitude is the Rayleigh wave. Before their first arrival some body waves are observable and after their first arrival is a superposition of body waves and reflected Rayleigh waves. The reflected Rayleigh waves dominate in amplitude over the body waves. In this figure one can clearly observe that the waveforms of the FD simulations become increasingly similar to the waveform of the spectral element simulation for smaller values of dh. However, even at dh = 3 m the waveforms show some differences. It seems that especially the surface waves arrive slightly earlier in the spectral element seismograms and smaller dh also leads to earlier arrivals in the FD seismograms. When looking at the

body waves which arrive earlier than the surface waves, some phases can be found which match well regardless of discretization and some show the same trend as the surface waves. Since body waves can only be directly observed in the seismograms before the onset of the surface wave it is hard to make any assertions about them. However, the observations that can be made lead to the hypothesis that the body wave phases which are influenced by discretization propagate parallel to the free surface like the surface waves. Because discretization of the free surface and direction of propagation relative to the free surface seem to be key factors the reason for this observations has to be the staircase effect. Since the smooth topography of the spectral elements has to be approximated by staircases in the FD method each stair-step acts as a diffractor. This effect varies with the inclination of the interface since up to a dipping angle of $45°$ the number of diffractors per unit length of the free surface increases. This is shown e.g. by Kunz and Luebbers (1993). The staircase effect also depends on dh because at infinitely small dh the dipping surface is smooth again. Thus, the effect has to vanish. Since the difference in model parameters is particularly strong at the free surface these diffractions have higher amplitudes than the diffractions occurring at dipping internal interfaces. The difference between the methods also increases with offset. The offset dependence of the difference between FD and SEM seismograms is better visible in figure 4.6. Here the RMS difference between the methods is shown as function of dh and offset. In this figure the general observable trend is that the difference increases with offset and with the value of dh. This leads to the conclusion that the error resulting from the staircase effect accumulates with offset. However, there is no strict increase with offset, e.g., at offset $3000\,\mathrm{m}$ there is a spike in the RMS difference which suggests that it also depends on the shape of the topography at each receiver position. This also makes sense because, e.g., receivers located in the valley at around $x = 2000\,\mathrm{m}$ certainly record more energy from reflected body waves while much of the surface wave energy gets reflected at the mountain flank to the right. On the other hand, receivers on the mountain flank to the left of the valley record mostly surface waves while body wave paths are being blocked by the topography. From the results it is apparent that a discretization with dh $= 7.5\,\mathrm{m}$ is not enough to reach convergence of the near-surface wavefield.

In conclusion, one would need a very fine discretization to model surface waves correctly. However, such a fine grid would need a tremendous amount of computational resources. Since FWI is already quite costly, in most cases such discretizations are not feasible. This raises the question, how FWI deals with the error in the forward simulation and if these resulting errors are small enough to justify the use of FD-based FWI. This is investigated in the next chapter. Finally, it can be said that such smooth interfaces as produced by the spectral element method in this test are not a perfect representation of real topography either. So the differences shown in this test are a worst-case scenario. One also has to be aware of the fact that the spectral element method is not free of errors either. But, since the FD solution seems to converge towards the SEM solution for small grid spacings and small offsets, it is proven to be an adequate reference.

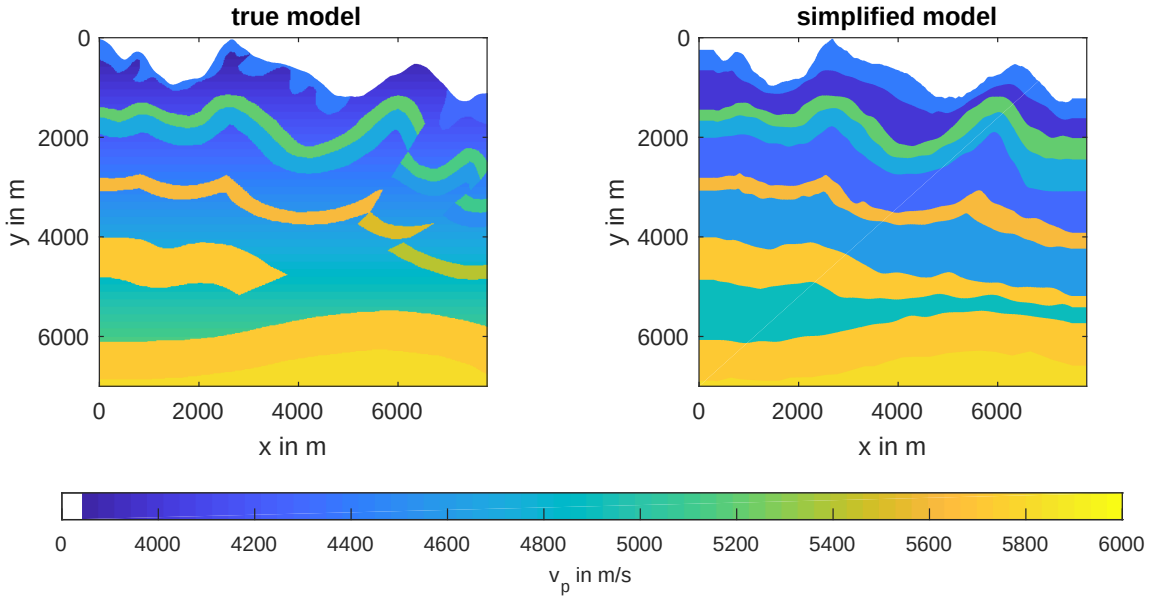

Figure 4.5: Left: The P-wave velocity of the left part of the Canadian Foothills model up to $x = 7.8\,\mathrm{km}$. The y-axis is limited to $7\,\mathrm{km}$. Right: The simplified version of this section which is a layered model containing only the most relevant interfaces.

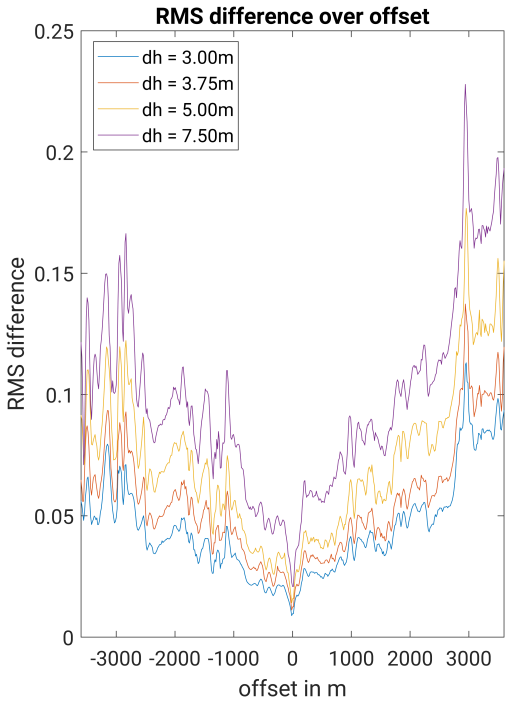

Figure 4.6: The RMS error between the vertical velocity components recorded at all receivers simulated with the SEM and with the FD-method at different discretizations.

Table 4.2: Model parameters of the simplified model shown in figure 4.5. The S-wave velocity is calculated with equation 3.2 and the density with equation 3.3. Both parameters are then rounded to the closest integer value.

Layer	v_P in $\mathrm{m\,s^{-1}}$	# of spectral elements along y
1	4400	50
2	4000	50
3	5200	50
4	4700	30
5	4300	40
6	5600	20
7	4600	50
8	5700	20
9	4900	20
10	5700	20
11	5800	30

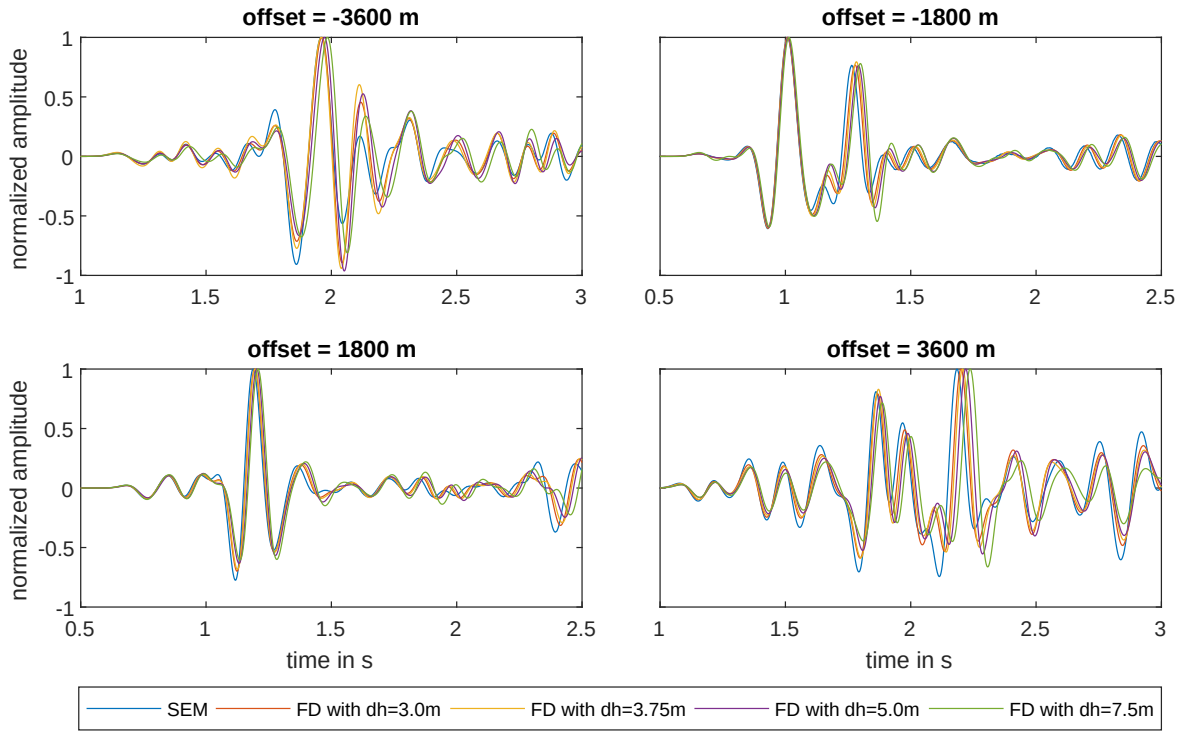

Figure 4.7: Selected time intervals of traces at different offsets modeled with SEM and the FD-method using different discretizations. The traces shown are all velocity seismograms recording the vertical component. All the time windows show the first arriving body waves followed by the high-amplitude surface wave.

4.3 3D effects

So far I have only considered 2D models with 2D wave propagation. However, the real world is 3D. Compared to 2D, waves have different amplitudes because of stronger geometrical spreading and also different phases. As 2D FWI is already costly enough, doing a 3D inversion is not an option, at least for me. The goal of this section is to assess the impact of 3D effects on the recorded seismograms. To do this I conduct two tests. In the first test I investigate only the effect of the different wave propagation in 3D compared to 2D and test a method to mitigate these differences. In the second test I investigate the effect of 3D out-of-plane reflections of surface waves and body waves.

4.3.1 3D wave propagation

To investigate the effect of 3D wave propagation alone, i.e. without out of plane reflections, I first select a subdomain of the model introduced in chapter 3 because 3D simulations are computationally expensive and it is not necessary to consider the whole model to see 3D effects. This test is conducted only with the FD-method and not with the SEM to focus on the influences of topography rather than the different forward solvers. I use a 4th-order FD operator, a grid spacing of dh $= 7.5\,\mathrm{m}$ and a time step interval of dt $= 0.7\,\mathrm{ms}$. The selected subdomain is the upper left part of the model and has a width of 1040 grid points and a depth of 740 grid points which is equal to $7800\,\mathrm{m} \times 5500\,\mathrm{m}$.

I choose this region because it has one of the strongest height variation of the free surface found in the model. Then I continue this model in the third dimension which is done in a symmetric way, so that there are no heterogeneities in the added dimension. This is often called a 2.5D model. I set the length of this new dimension to 400 grid points, which equals $3\,\mathrm{km}$. The acquisition geometry, which I also would like to use in later inversion tests, is set at the center of the new axis. This way influences from the boundaries are not an issue. While I only discuss the center shot in this section, overall I simulate 48 shots with a spacing of $150\,\mathrm{m}$ and a receiver spacing of $15\,\mathrm{m}$. The first shot is located $375\,\mathrm{m}$ away from the left model boundary to ensure no significant influence from boundary reflections. With the given model dimensions and shot spacing the same distance is between the last shot and the right boundary. o absorb boundary reflections I use C-PMLs with a width of 20 grid points. The two component receivers start at the first shot position and end at the last shot position which results in 470 receivers in total. This setup is shown in figure 4.8. Here, I do not show the other parameter models to avoid redundancy but they are calculated with equations 3.2 and 3.3 from the v_P model. As a source signal I use a Ricker wavelet with a center frequency of $10\,\mathrm{Hz}$ which is set off as a vertical source. With this new elastic model and acquisition geometry I perform a 3D forward simulation. I also calculate the 2D seismograms with the corresponding 2D model as a reference.
As I do both, 2D simulations and 3D simulations, in this test, I can also compare the computational resources needed for the two settings. The simulation of a single shot in the 2D setup took about $37\,\mathrm{s}$ on 20 cores which corresponds to 0.2 core-hours. In the 3D setup a single shot took about 11 minutes and $36\,\mathrm{s}$ on 2240 cores which corresponds to 433 core-hours. This is an increase by a factor of more than 2000 which gives a good insight why 3D elastic inversions are rarely done. At first this difference seems unrealistic because the model size increased only by a factor of 400. However, in 3D there is massively more communication between CPUs needed and using less CPUs for less communication is also not an option because of memory limitations. I obtained these values with the hardware of the ForHLR II cluster. The CPUs of this cluster have a base clock-speed of $2.6\,\mathrm{GHz}$.
The 3D equivalent of the seismic source in a 2D simulation would be an infinitely long line source along the third dimension. In a real measurement we deal, however, at least approximately with point sources. To solve this problem I use the reflected wave transformation proposed, e.g., by Forbriger et al. (2014) which converts a point source measurement to that of a line source. The procedure suggested by them is first to convolve each trace of the seismogram with $\sqrt{t^{-1}}$ to transform the phases and then taper each trace with $\sqrt{t}$ to correct the amplitudes. The exact amplitude correction for a homogeneous medium would be a multiplication with $v_{\mathrm{ph}}\sqrt{2t}$ where v_{ph} is the phase velocity. However, the constants are omitted because I do not use true amplitudes but only relative amplitude information. This method has to be tested first because it has never before been applied to a model of that size and that complexity. Applying the transformation to the shot gather of shot 23 results in the transformed seismograms shown in figure 4.9. Although the transformation is meant for body waves the figure shows that it works surprisingly well for surface waves and even reflected surface waves. While the phase shift needed in the transformation is the same for body and surface waves, the amplitude correction factor is different. For surface waves the amplitude correction has to be $\propto \sqrt{t^{-1}}$. Since there is only a difference in amplitude and not in phase, the direct surface waves are automatically "corrected" with the right amplitude factor by trace normalization which is also used during FWI. One would expect that this causes problems because the wrongly corrected amplitudes of the surface

waves offset the amplitude correction of all other phases. Around $t = 1\,\mathrm{s}$ the factors $\sqrt{t}$ and $\sqrt{t^{-1}}$ do not differ by much. So, as long as the first arriving surface wave arrives close to that time, the body waves are also correct. A larger discrepancy is observable at larger offsets. Here, the difference between the two factors causes the transformed body waves to have a smaller amplitude than they should have. The reflected wave transformation also does not correct the direct wave properly, this can be seen in zoom 1 of figure 4.9. On the other hand, it seems that the reflected wave transformation in combination with the constant amplitude offset introduced by trace normalization works really well for reflected surface waves as can be seen in zoom 2 of figure 4.9. The conclusion of the visual interpretation of the seismograms is that overall the reflected wave correction in combination with trace normalization works well for direct and scattered surface waves and at near offsets also for body waves. At far offsets the amplitude transformation of the body waves has some shortcomings but, since the phases are transformed correctly, there is a good chance that this has no big influence on FWI.

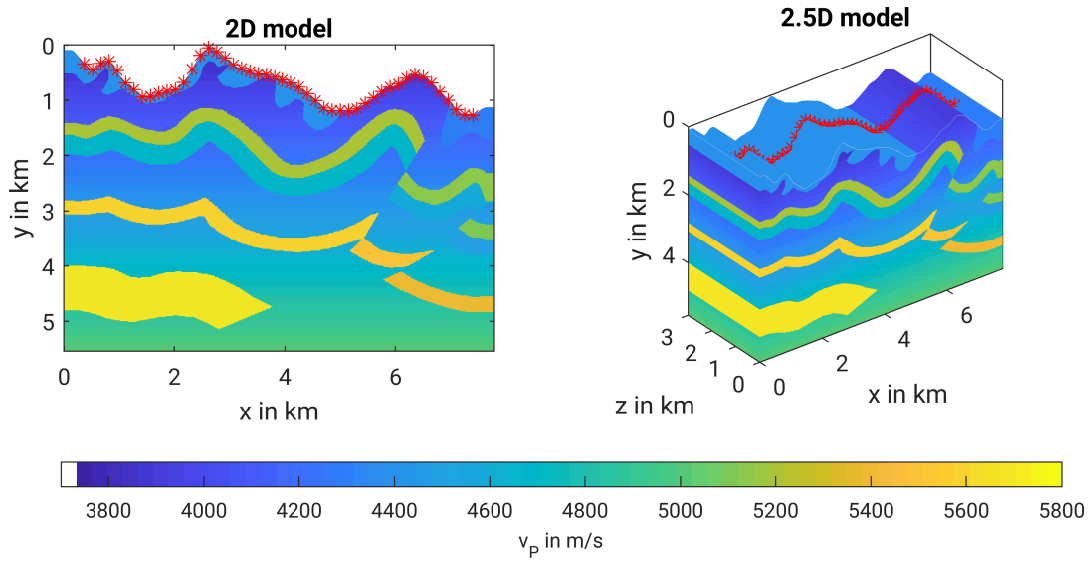

Figure 4.8: On the right is the 2.5D extension of the true v_P model shown on the left with the shot positions marked with red stars.

4.3.2 3D topography

After showing in the previous section that the reflected wave transformation yields sufficiently good results in case of a 2.5D medium, the goal of this section is to see its performance in a 3D medium. For this I construct a 3D model from the 2.5D model which is shown in figure 4.10. A more detailed description of how I constructed this model can be found in appendix B. The basic idea of this 3D model is that the deeper geology can still be approximated by a 2.5D medium and only the surface topography, and hence also the near-surface geology, are truly 3D heterogeneous. I constructed this model also in a way that the slice in the xy-plane along the acquisition geometry is equivalent to the 2D model used previously. This means, the acquisition geometry is at exactly the same position as in the previous test. The only thing different to the previous test is the surface shape to the left and right of the acquisition line, the size and everything else is the same. Applying the reflected

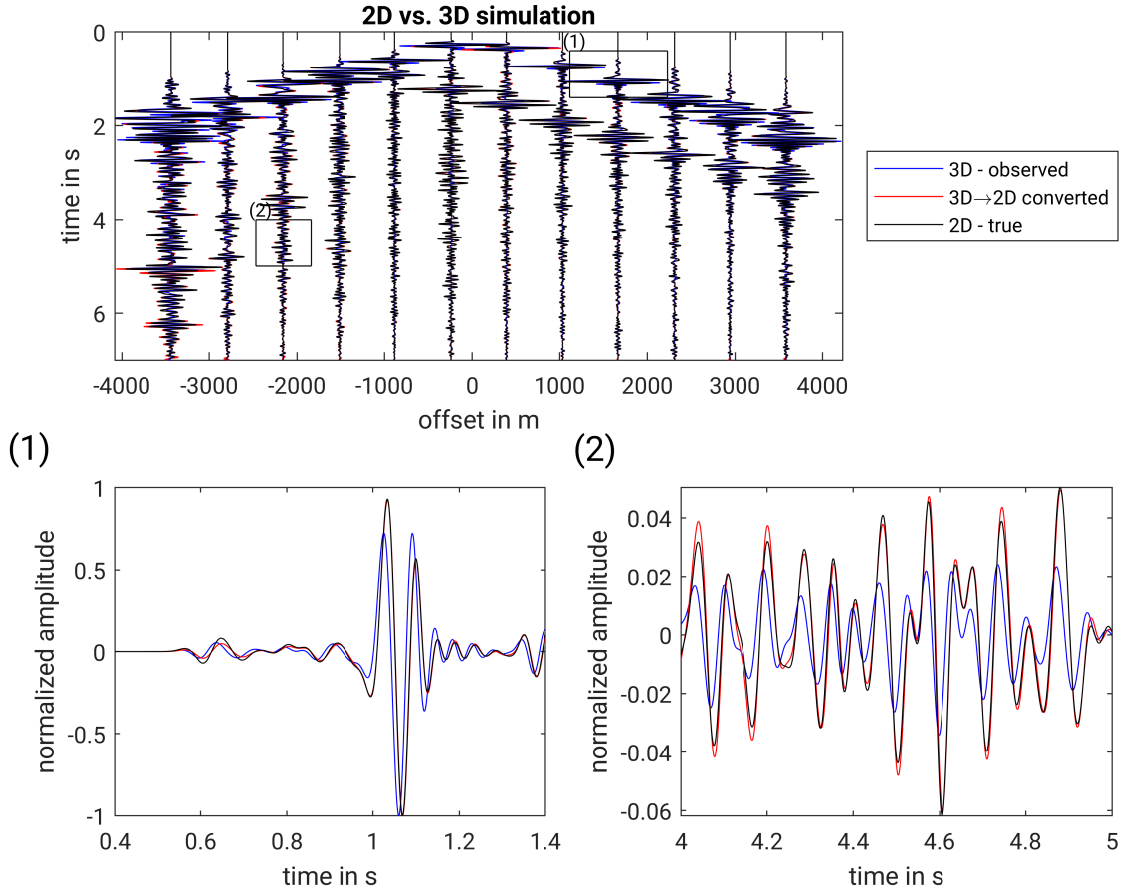

Figure 4.9: Result of the 3D to 2D conversion for the 2.5D model. The upper plot shows 12 equidistant traces in the v_y shot gather of shot 23. The traces are scaled with t and normalized for better visualization of late arrivals. The two plots at the bottom show a zoom in the locations marked by black boxes in the upper plot. The left zoom shows the first arrivals of the body waves and the surface wave. The right zoom shows a superposition of body waves and scattered surface waves.

wave transformation yields the result shown in figure 4.11. One can observe that the transformation performs very poorly in this case. Neither surface waves nor body waves are transformed correctly regardless of the arrival time. There are large differences both in amplitude and in phase. It is also noticeable that the amplitude of later arriving phases is greatly reduced compared to the 2.5D case which means that much of the wave energy is scattering at the topography away from the receivers. This result is not surprising because it is obvious that in case of strong 3D variations of the free surface a line source transformation makes no sense. Out-of-plane reflections of surface waves can also have a huge influence on the recorded seismograms. All in all, this 3D setting is probably beyond the limit of what can be reduced to a 2D problem. Although the subsurface geology in still 2.5D, the influence of surface reflections is too strong. In conclusion, there is only a slim chance that FWI is capable to invert body waves which penetrate inside the 2.5D subsurface of the model. Even though they are transformed correctly, the superimposing surface waves probably have too much of an effect.

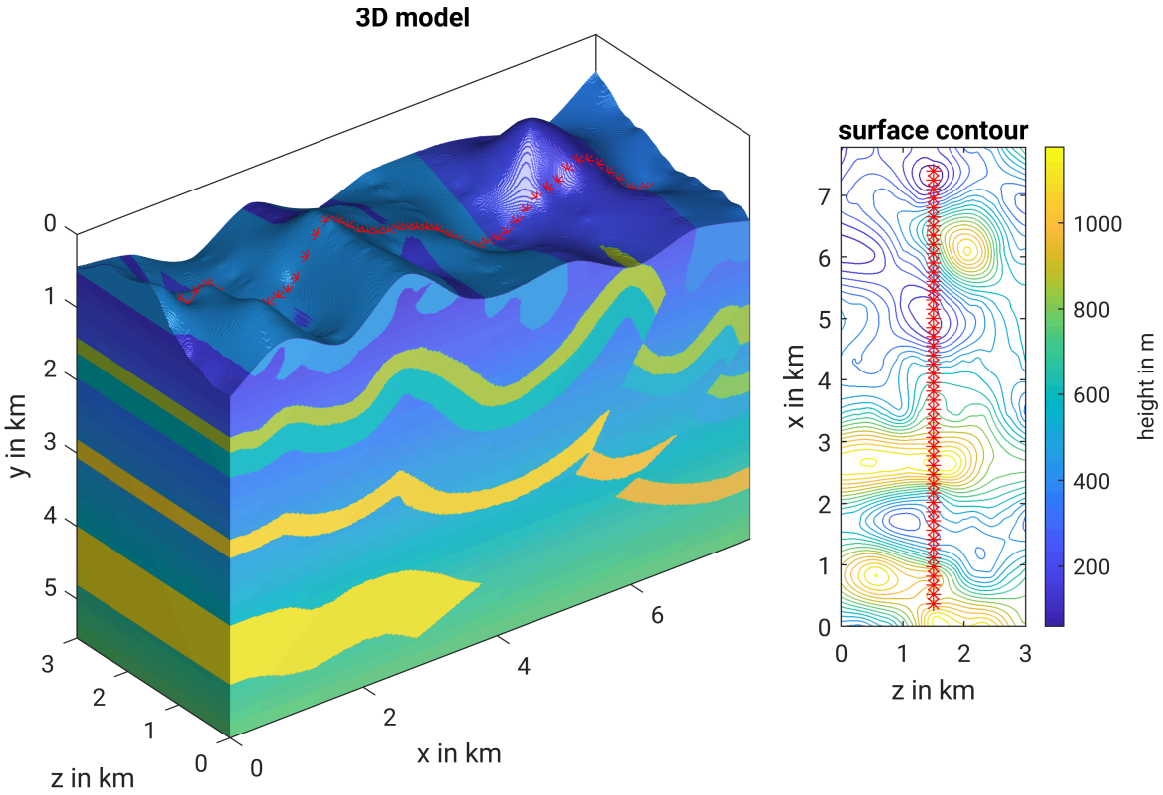

Figure 4.10: Left: The visualization of the 3D model which is generated by combining a heterogeneous surface with the 2.5D model shown in figure 4.8. As I use lighting in this figure to improve the visualization of the surface topography I do not show a legend for the colors. Right: The contour of the surface topography. The height shown in this plot is relative to the lowest point of the surface. In both plots the shot positions are indicated by red stars.

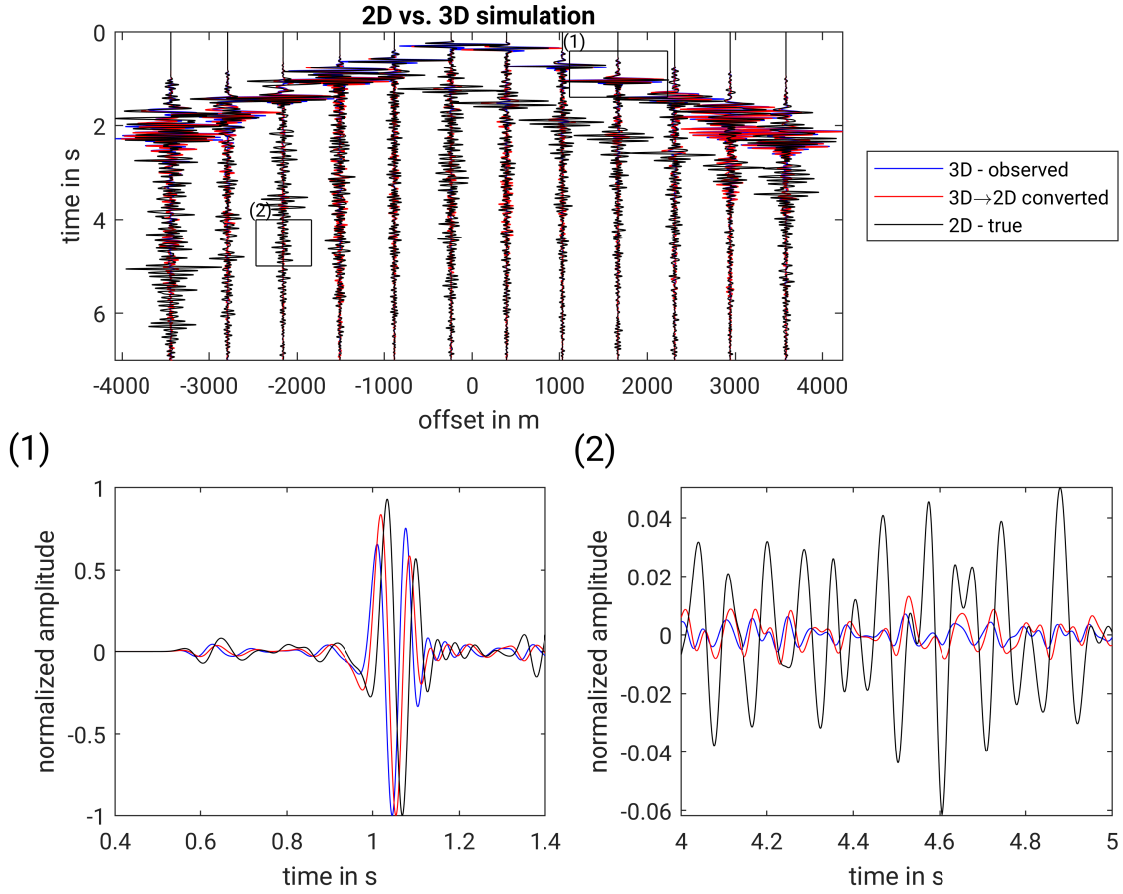

Figure 4.11: Result of the 3D to 2D conversion for the 3D model. The upper plot shows 12 equidistant traces in the v_y shot gather of shot 23. The traces are scaled with t and normalized for better visualization of late arrivals. The two plots at the bottom show a zoom in the locations marked by black boxes in the upper plot. The left zoom shows the first arrivals of the body waves and the surface wave. The right zoom shows a superposition of body waves and scattered surface waves.

Chapter 5

Inversion with topography

In this chapter I use FWI on synthetic land data in the presence of topographic variations. In general this chapter is divided into two parts. In the first part I investigate the consequences of the error done by using FD solvers on FWI. This section is based on the results obtained by the forward modeling test in section 4.2. In the second part I investigate how 3D effects like shown in section 4.3 affect 2D FWI.

5.1 Applicability of FD solvers

Goal of this section is to test the applicability of FD solvers for FWI in regions with strong topographic variations. As seen in section 4.2, especially for surface waves, there are significant discrepancies between the SEM solution and the FD solution. These discrepancies increase with increasing grid spacing dh. To test the applicability of FD solvers for FWI in regions with strong topographic variations I invert SEM seismograms with a FD-based FWI in this section. To evaluate the result of this inversion I use the same settings to invert pseudo-observed data simulated with the same FD solver as used in FWI.

5.1.1 Low-frequency inversion

In this section I first introduce the general workflow, techniques and parameters used in the inversion tests. With these settings I then conduct two inversions. In the first inversion I invert pseudo-observed data modeled with the same FD solver as used in FWI. This means, the impairments coming from incorrectly modeled surface waves do not have any effect in this test. The result is used as a reference solution for the second inversion in which I use pseudo-observed data simulated with the SEM. As shown in section 4.2, discretization is a critical issue. To reduce the computational cost I assume a modern acquisition where low frequencies are recorded in this test. Lower frequencies mean longer wavelengths and thus less grid points needed in total to obtain a certain number of grid points per wavelength. Low frequencies also means there is a reduced risk for cycle skipping. This describes the unwanted process that an event in the synthetic data is fit to a wrong event in the observed data and

in a consequence of the local optimization used in FWI. Cycle skipping occurs if the corresponding events in the synthetic data and field data are separated by more than half a period. To avoid cycle skipping multi-scale FWI as described in section 2.2.3 is used. However, this only helps if the starting model fulfills the half a period criterion of the lowest frequencies recorded.

Methodology

In this section I introduce the most important parameters of the FWI algorithm. I start with the parameters I use for the FD forward solver. I use a FD solver which is of second order in time and fourth order in space because higher orders do not bring any significant improvements since the spatial discretization is still relatively fine. Since the model shown in section 4.2 is based on the model I want to invert in this test and which I show later, the impairment of the surface waves depends on dh in the same way as in the forward modeling test. The discretization I choose for my inversion test is dh $= 7.5\,\mathrm{m}$ for two reasons. The first is that the forward modeling test shows large mismatches between SEM and FD seismograms when using this dh. This means, the effect of the poor discretization will definitely be observable in the FWI result. Second, because this dh is the largest of the tested discretizations in section 4.2 it requires the least computational effort. FWI itself is already costly so the goal is to reduce the required computational resources whenever possible. The size of the models is similar to that of the model used in section 4.2 to keep the offset-dependent error of the FD solution in the same range. The exact dimensions are $7800\,\mathrm{m}$ in length and $7050\,\mathrm{m}$ in depth which is equal to 1040 grid points and 940 grid points, respectively. Following criterion 2.28, I set dt to $0.7\,\mathrm{ms}$. With the average velocities in my model I estimate the simulation time needed to be $7\,\mathrm{s}$. This should be more than enough for an S-wave to travel from a source in the upper left corner of the model to the lowest reflector and then to a receiver at the upper right corner and also should be enough to record most multiples.

For the SEM simulations, which are needed to obtain the pseudo-observed seismograms for the second inversion, different discretizations are required. In this test I do not adapt the SEM mesh to the internal interfaces of the model but only to the surface. Since it is known that staircase effects at internal interfaces are negligible in body wave simulations with the FD method, they are also negligible in this case. To save some computational resources I choose a fine discretization of the SEM mesh near the surface and a coarser discretization in the deeper part. Up to $y = 2.5\,\mathrm{km}$ the SEM mesh is 100 elements in height. Below this depth the mesh has another 150 elements in height and the elements have a rectangular shape. So the total number of elements in the vertical direction is 250. In horizontal direction there are 520 elements in both regions of the mesh. Over all, this results in a relatively high number of spectral elements. The reason is that I want to ensure the similarity of the SEM model and the FD model. The time discretization needed with this spatial discretization is $0.1\,\mathrm{ms}$.

The true and starting models used in this test are shown in figure 5.1. The true models are the upper-left sub-region of the Canadian Foothills model introduced in chapter 3. I do not use the full model because the chosen sub-region has one of the strongest topographic variations of the free surface and also has a reasonable size. Hence, if the inversion is successful in this region, it can be applied to all other regions of the model. The models shown in the figure are the models used for the FD method.

To obtain the SEM version of the true model, I use nearest-neighbor interpolation to deduce the material parameters of the spectral elements from the FD model. This interpolation preserves the sharp interfaces of the model. To minimize the effect of boundary reflections I use C-PMLs which have a thickness of 20 grid points in the FD case and 10 elements in the SEM case. I use C-PMLs in this test rather than absorbing boundaries because the absorption works different in the two methods I want to compare, so I need to make sure it works almost perfectly. Reflections occurring in one method but not the other could result in artifacts in the inverted models. C-PMLs work best when there are no reflectors parallel to the edge they cover inside them because in their theoretical derivation a constant model along the normal direction is assumed (Komatitsch and Martin, 2007). Hence, I remove all variations perpendicular to the model boundaries inside the C-PMLs. The starting model is a linear velocity gradient. This gradient is different to the true gradient which underlies the Canadian Foothills model because true velocities are unknown in the real world. The assumption that seismic velocities and density increase with depth holds in most cases so this starting model requires only little a-priori information.

Regarding the acquisition geometry I place all sources and receivers just below the free surface such that they are not located within the half grid point layer introduced by the IVF (see section 2.1.5). I use 470 two-component receivers which record the particle velocity. For the inversion I use 48 vertical velocity sources. To further minimize the effect of boundary reflections I keep a distance of 50 grid points between the start and end of the acquisition geometry and the model edges. This results in a horizontal receiver spacing of $15\,\mathrm{m}$ and a horizontal shot spacing of $150\,\mathrm{m}$. Shot positions are also indicated in figure 5.1.The first reciever is placed at the position of the first shot and the last receiver at the position of the last shot. As I want to use the STF inversion introduced in section 2.2.3, I need to define two source signals. A true source signal which is used to produce the pseudo-observed data and a synthetic source signal which is used as a starting signal for the STF inversion. As true source signal I use the same $5\,\mathrm{Hz}$ Ricker wavelet as in the forward modeling tests. As synthetic wavelet I choose as $\sin^3$ wavelet because of its low-frequency content which guarantees the stability of the simulation needed for STF inversion. The $\sin^3$ wavelet is defined as:

$$s_{sin} = \begin{cases} A\sin(\pi f_c(t - t_{shift}))^3 & \text{if } t_{shift} \leq t \leq t_{shift} + f_c^{-1} \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

Like in the definition of the Ricker wavelet A describes the amplitude, f_c the center frequency and t_{shift} the time shift. All required parameters for the two sources are listed in table 5.1. The frequencies of the two wavelets are very similar since the frequency needed for the synthetic source can be estimated easily from the amplitude spectrum of the observed seismograms.

For the STF inversion I do not use the entire shot gathers because the strong reflected surface waves are otherwise visible in the inverted source signals. I rather use tapered gathers where only the first arriving body waves and the direct surface wave is used. For shot 24 the full and the tapered version of the shot gather are shown in the top row of figure 5.2. To generate this taper I automatically pick the first onset of each seismogram trace. Up to $0.5\,\mathrm{s}$ after each picked onset the taper is 1 and after that a cosine taper with the length of another $0.5\,\mathrm{s}$ is used which acts as the transition to the muted part. Since I use synthetic seismograms the automated picking can be done easily. In the presence of noise the definition of the muting taper probably has to be done manually for each shot.

During inversion I use the preconditioned conjugate gradient method, step length estimation and the multi-scale approach which all have been introduced in section 2.2.3. For the multi-scale approach I use a $2\,\mathrm{Hz}$ and $4\,\mathrm{Hz}$ low-pass filter consecutively in the first stages before I use the entire frequency content. The filter type I use is a fourth-order causal Butterworth filter. The filter is causal rather than zero-phase to avoid wrap-around effects of the first arrivals at low corner-frequencies. Figure 5.3 shows how this filter affects the source signal in time and frequency domain.
In section 4.2 I show that surface waves suffer most from the staircase effect. As of their high amplitudes, however, they dominate the misfit function. To mitigate their influence, and thus boost the influence of body waves, I also use a fk filter in every second stage. Body waves should not be affected much by the staircase effect near the surface and also allow for better model updates in the deeper regions of the model. I use this filter in every second stage because the inversion of surface waves is more stable and I want to use the information about the subsurface contained in the surface waves. The effect of the filter is exemplary shown in figure 5.4 for shot 24. In the spectrum the separation of body and surface waves is not very clear. However, there are clear energy maxima observable which have to correspond to the surface waves but the transition to the body waves which have to be located at higher frequencies and lower wave numbers is continuous. The reason for the poor separation is that the receivers are not on a straight line because of the topography which distorts the k part of the spectrum. A velocity filter with a corner velocity of $2000\,\mathrm{m\,s^{-1}}$ has its transition zone right below the energy maxima in the spectrum as is indicated by the red lines in the figure. Comparing the seismograms where this filter is applied to the original one in time domain which is shown on the right of the figure one can see that in most traces the separation of body waves and surface waves works surprisingly well. To minimize the effect of the fk transform on the body waves I mute them in the fk domain instead of the surface waves and then subtract the resulting filtered seismograms from the original ones in time domain. With this result I can be confident to use the filter during FWI. As misfit function I use the L_2 norm with normalized seismograms which increases the weight of far offset traces. This also gives larger model updates in the deeper parts and thus weakens the impact of surface waves on the gradient. Finally the complete workflow is shown in table 5.2. I conduct a STF inversion at the beginning of each stage and then again every tenth iteration. This way the increase of computational cost coming from the additional forward modelling needed for the STF inversion is not too significant but large changes in the inverted model which happen inside a stage can still be used for improving the STF.

Table 5.1: Parameters of the synthetic and the true source signal. The true signal is calculated with equation 4.1 and the synthetic signal with equation 5.1. Amplitude A has no unit because our code does not scale the sources to their true physical units. This is not problematic because recording true amplitudes is very difficult in practice, and hence rarely done.

	type	A	f_c in Hz	t_{shift} in s
true source	Ricker	1.0	5.0	0.00
synthetic source	$\sin^3$	1.5	6.0	0.15

Table 5.2: Inversion workflow. In every all three model parameters are beeing inverted. Higher stages use higher frequency content until the last stages use the entire frequency spectrum of the source. The fk filter is used in every second stage.

Stage	v_P	v_S	ρ	filter edge in Hz	fk filter
1	✓	✓	✓	2	–
2	✓	✓	✓	2	✓
3	✓	✓	✓	4	–
4	✓	✓	✓	4	✓
5	✓	✓	✓	-	–
6	✓	✓	✓	-	✓

Results and Discussion

The six inverted models resulting from the two inversions are shown in figure 5.1. One can observe that the inverted v_P and v_S models of both inversions match the true models to a very high degree. The edges of the model features are slightly smoother than in the true model but this is a result of the low frequencies used. One can also observe that the inverted v_S models appear to be a little better resolved than the v_P models. This is reasonable because S-waves have a shorter wavelength than P-waves. Both inverted velocity models fail to reconstruct the true velocity of the high velocity feature on the left just below a depth of $4\,\mathrm{km}$ but only reconstruct the increase in velocity at the edges of the feature. The reason for that is that only some reflected waves illuminate this feature because it is not in the center of the acquisition geometry. In general, the limit with this acquisition geometry seems to be around a depth of $4\,\mathrm{km}$ because there appear to be more random fluctuations in the smooth background velocity gradient below this depth. On the first look, the resulting v_P and v_S models are very similar for both the inversions. In the inverted v_P model where the SEM data is used, however, high velocity artifacts occur near the surface, especially in regions with strong topography. For better visualization zooms of these artifacts are shown in figure 5.5. For reference the true model and the inverted model for which FD data is used are also shown in the figure. This figure proves that the high velocity zones cannot be found in the true model or the inverted model which is obtained using pseudo-observed data from the FD method. This means, the artifacts are connected to the SEM data. This is plausible because, as seen in section 4.2, the surface waves are slightly faster in the SEM seismograms compared to the FD seismograms even though the underlying models are equivalent. To balance this velocity difference the inversion introduces high velocity zones. Interestingly, the high velocity artifacts cannot be observed in the v_S model. Out of the three parameters I invert for the inverted density model has by far the largest error. In general, areas where the true model has higher densities than the starting model are overestimated and regions where it is the other way around are underestimated. The inversion seems not to be capable to invert the true values but only density contrasts. This behavior is expected and also common for FWI. The problem with inverting density is that it mostly affects the amplitudes of the reflected waves rather than their phases. This makes the inversion more difficult. Some authors like, e.g., Brossier et al. (2009) do not invert for density because of that reason. Nevertheless, I will do so because it yields the advantage that artifacts from non-physical modeling can be collected in the density model which improves inversion of the other parameters as shown, e.g., by Przebindowska et al. (2012). As a result, I show inverted density models

from now on for the sake of completeness but I do not discuss them.
With this high velocity zones in the v_P model the inverted data matches the pseudo-observed data to a very high degree. The data fit is equivalent to that of the inversion which uses FD data as is shown in figure 5.6. The figure shows that in both inversions the inverted seismograms almost perfectly align with the pseudo-observed seismograms in all parts of the shot gather. There are some errors observable at large offsets and at very late arrival times but these errors are very localized, and hence considered minor. The good data fit is also a consequence of the STF inversion result which is shown in figure 5.2 for shot 24 of the inversion which uses FD data. The inverted source signal matches the true source signal to a point where they cannot be distinct visually when using tapered seismograms. The inverted source signal which uses no taper and is shown for reference has some oscillations at around $1.5\,\mathrm{s}$ which are caused by the high amplitude reflected surface wave. Without tapering, the amplitude of the first minimum of the inverted signal has also a slight error. The reason for this are probably notches in the amplitude spectrum of late arriving phases which are caused by interference. Finally, the misfit development for both inversions is shown in figure 5.7. It can be observed that the final misfit is below 20% of the initial misfit which is that obtained with the starting model. The inversion which uses SEM data takes about 50 iterations longer than the one which uses FD data. While the misfits after the fifth stage are almost equivalent in both inversions, the misfit after the sixth stage of the SEM data inversion is ca. 5% points lower than that of the FD data inversion. While this sounds surprising at first, the difference is not really significant. When comparing previous stages of the inversions, the differences are in the same order of magnitude and regardless the relative misfits are the same after stage 5. In general, this comparison of misfits is only possible to a limited degree because the STF inversion also has a considerable influence on it. Although body waves have smaller amplitudes the misfit in stages where I use a fk filter is higher than in stages where I do not because I normalize each trace before calculating the adjoint sources. Furthermore, the misfit without fk filter is dominated by surface waves. The inversion of surface waves is more stable because their sensitivity is significantly higher and focused on a smaller region compared to body waves. The figure of the misfit development also justifies the use of the STF inversion in every tenth iteration of each stage. E.g. in the first stages there are iterations where the misfit decrease stagnates but then, after a new STF inversion is executed, it decreases again rapidly.

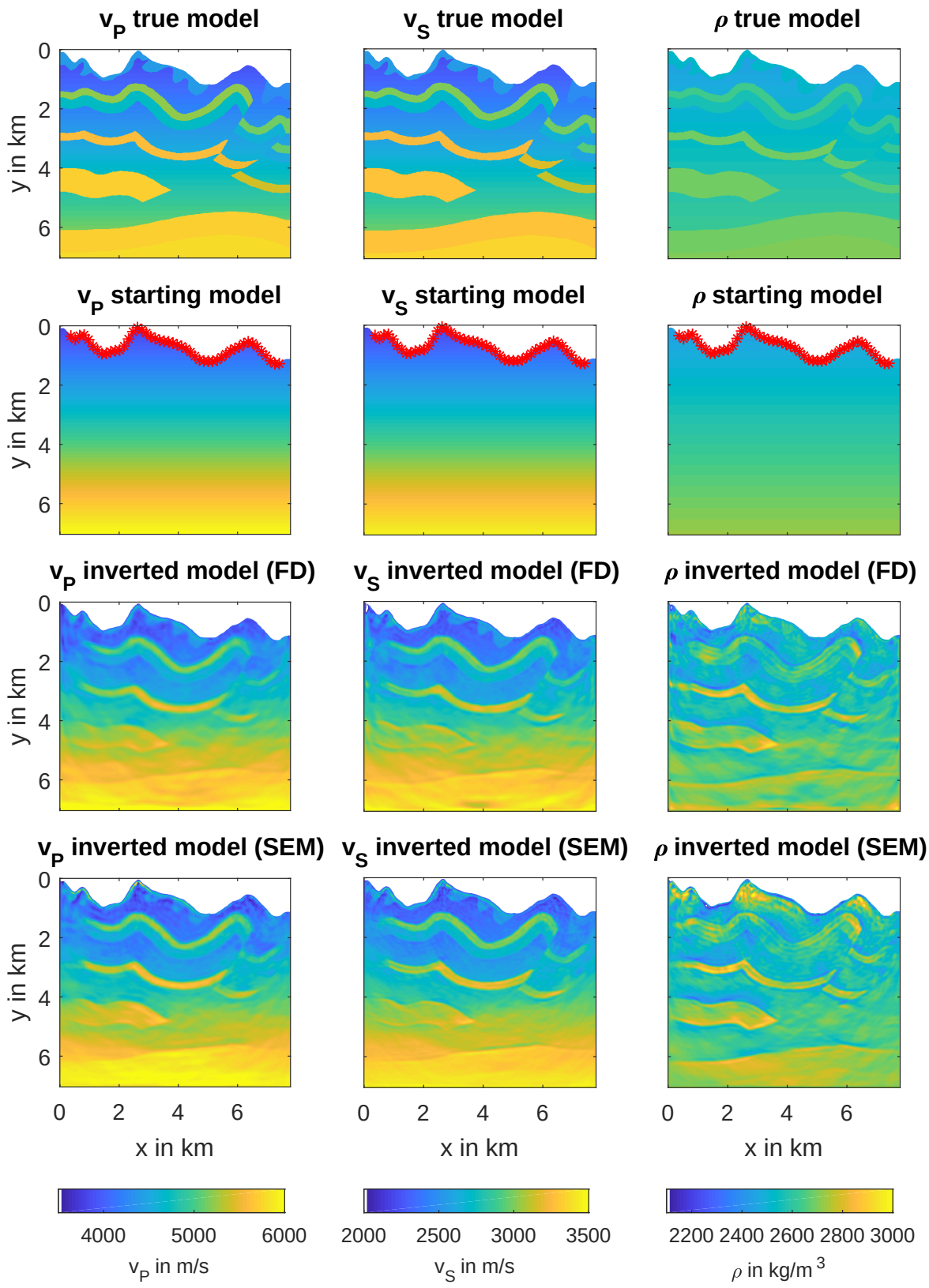

Figure 5.1: The first row shows the true parameter models used for the inversion tests. In the second row the corresponding starting models are shown. In this models I also show the source positions indicated with red stars. The last two rows show the inversion results. In the penultimate row are shown the inverted models when using pseudo-observed data coming from a FD solver. The last row shows the result when using pseudo-observed data coming from a SEM solver.

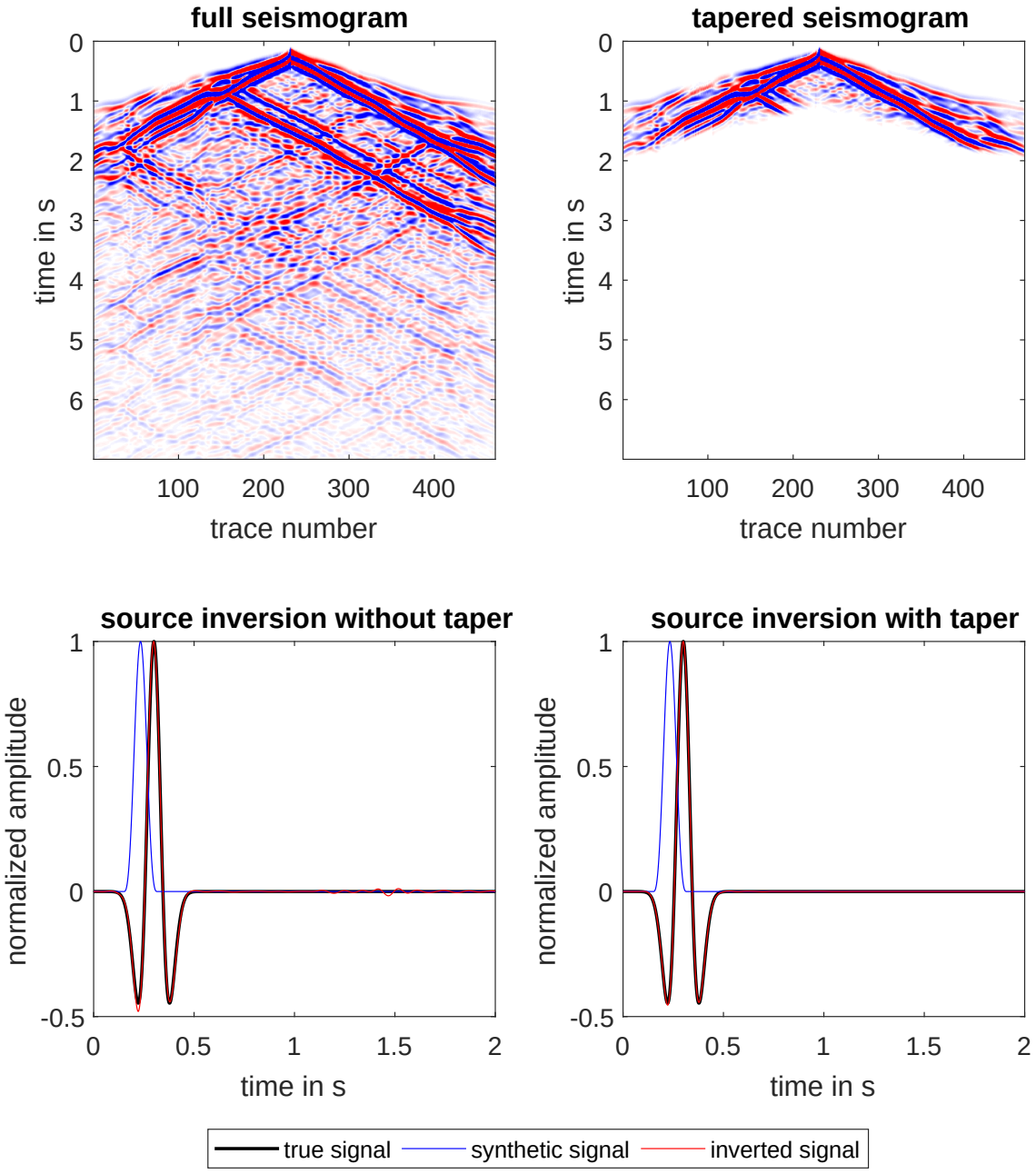

Figure 5.2: The upper row shows image plots of the raw and tapered shot gather for shot 24. These image plots are clipped and scaled linear with time to improve the visualization. The tapered shot gathers are used for STF inversion. The lower row shows the inverted source signatures used in the fifth stage of the inversion which uses FD data. Shown is a comparison of the inverted signal when a the taper is used (right) and when it not (left).

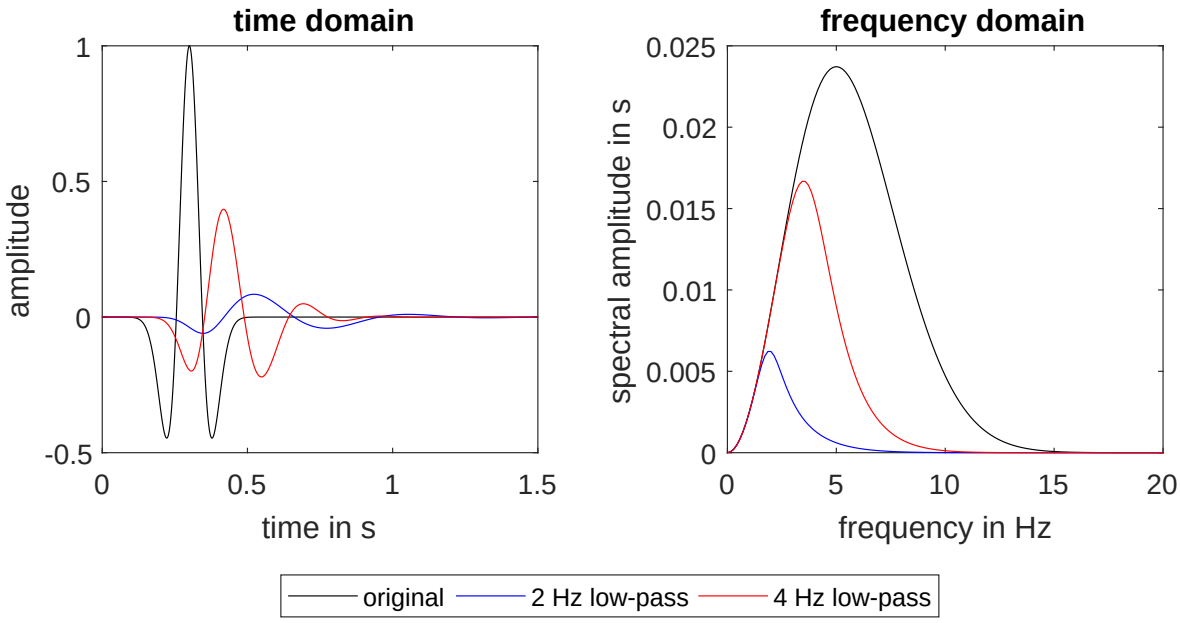

Figure 5.3: Source wavelet in time and frequency domain with and without frequency filters used during the inversion. With lower corner frequency of the filter the amplitude of the wavelet decreases and the length increases.

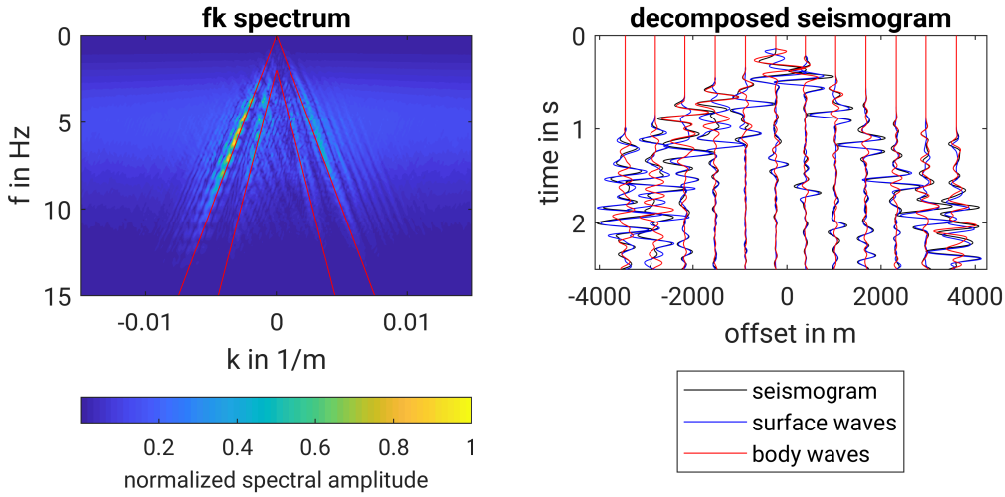

Figure 5.4: The plot on the left hand side shows the fk spectrum of the v_x receivers at shot 24 which is located in the center of the acquisition geometry before filtering. The red lines are the borders of the fk-filter flanks. The plot on the right hand side shows the result of the filtering in time domain. The plotted surface wave seismograms correspond to the inverse 2D Fourier transform of the filtered fk spectrum and the plotted body wave seismograms correspond to the difference between the original seismograms and the surface wave seismograms. Shown are only 12 equidistant traces of the original 469 traces for better visualization. For the same reason I also show only the first 2.5 s of the total time which is 7 s.

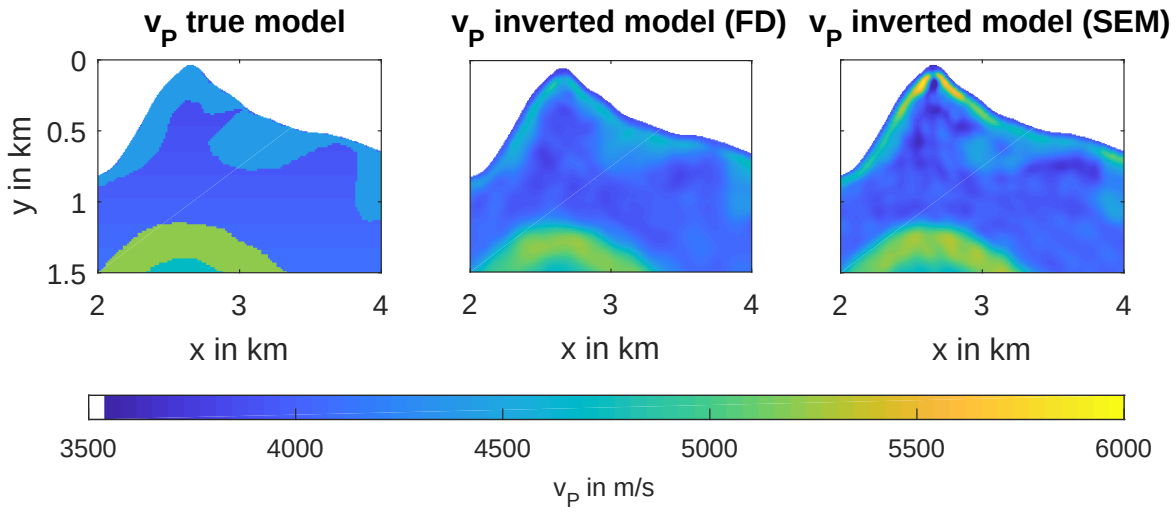

Figure 5.5: Comparison of the central near surface region between the true v_P model, the inverted model using FD data and the inverted model using SEM data. In the SEM data inversion strong high velocity artifacts are observable.

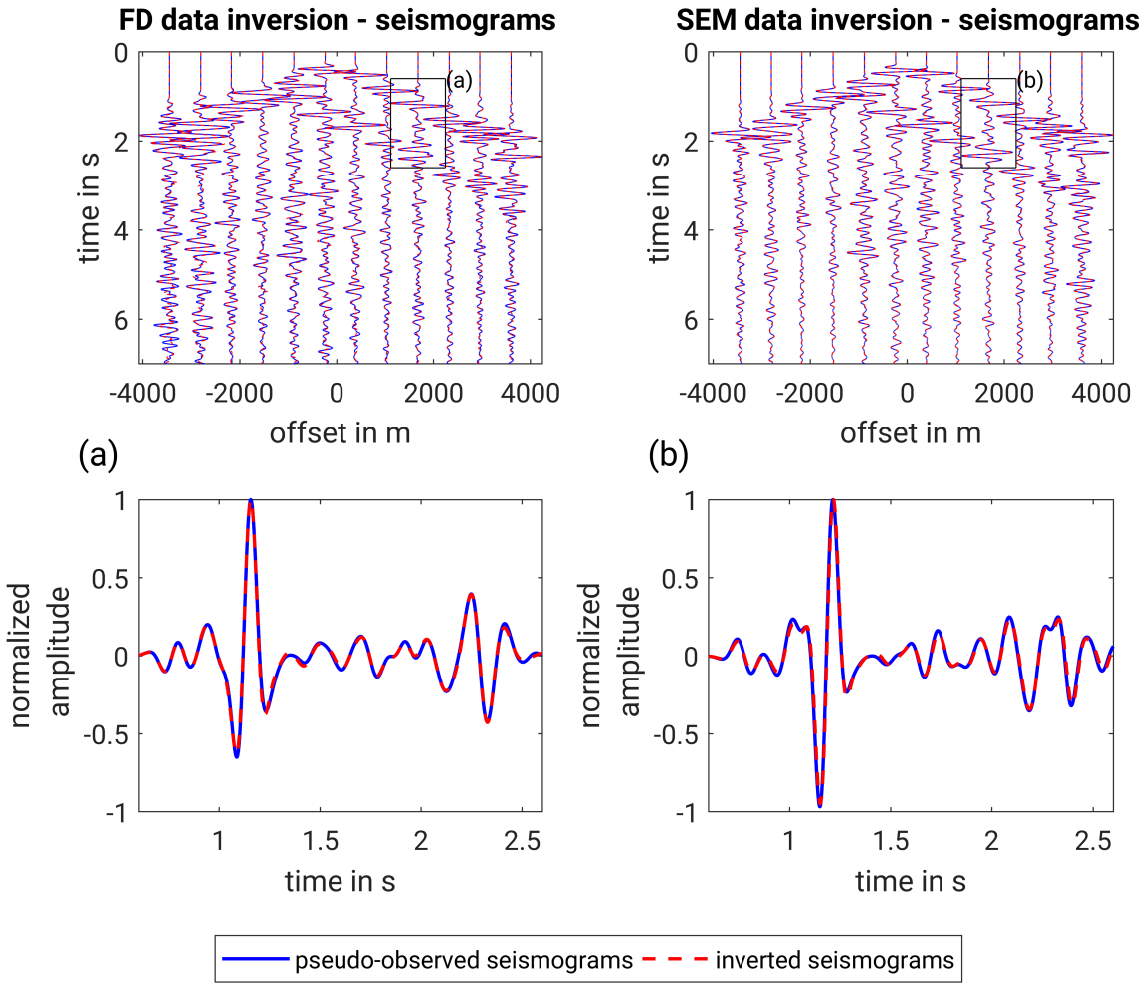

Figure 5.6: Comparison of the pseudo-observed and inverted v_y seismograms for shot 24. On the left is shown the result of the FD data inversion and on the right the result of the SEM data inversion. The upper plots shows an overview of the shot gather of shot 24 and the lower ones shows a zoom on the trace which is indicated by the black box in the overview.

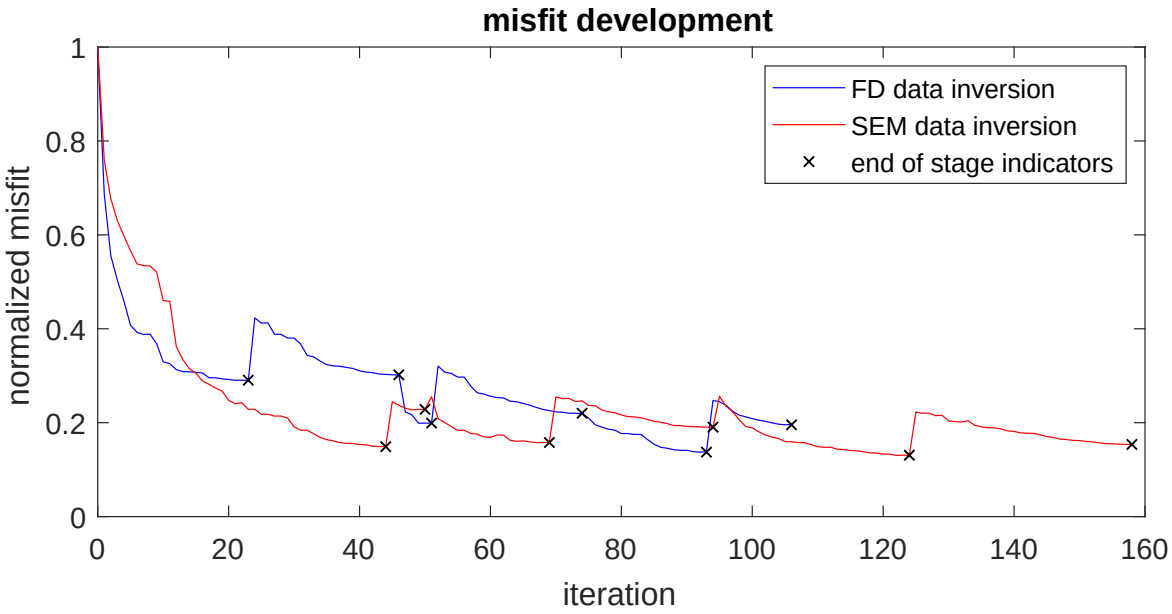

Figure 5.7: Shown is the misfit development for the inversion which uses FD data and the inversion which uses SEM data. Plotted on top of the curves are black crosses which indicate the end of a stage. The misfits are normalized to that of the starting model.

5.1.2 Moving to higher frequencies

The previous tests are conducted with very low frequencies because of the inaccurate modeling of surface waves at fewer grid points per wavelength, as shown in section 4.2. Since the result of section 5.1.1 are far better than expected the next step is a test with higher frequencies. Frequencies below $5\,\mathrm{Hz}$ can only be acquired on land since a few years (Mahrooqi et al., 2012). If this test is successful older data sets can also be used for FWI.

Methodology

The methodology of this test is almost the same as the one of the test in section 5.1.1. Only the source signal and starting model are adjusted. The new source parameters are listed in table 5.3. The center frequency of the source in this test is double that of the last test. Using criterion 2.27 yields that with a spatial discretization of dh $= 7.5\,\mathrm{m}$ the grid dispersion is still acceptable. Starting the inversion with higher frequencies requires a better starting model to prevent cycle-skipping. Instead of a linear gradient I use a strongly smoothed version of the true model. The smoothing is done in a three step process. The first step is setting all vacuum grid points in the true v_P model to $4000\,\mathrm{m\,s^{-1}}$. This is the approximate velocity near the surface and prevents a strong low-velocity zone near the top surface after the smoothing. The second step is the actual smoothing. For this I use a spatial filter which is applied in the kk domain after a 2D Fourier transform of the substituted model. The spatial filter removes all features which are smaller than $1600\,\mathrm{m}$ which is approximately two wavelengths close to the surface and less than that in the deeper regions. The last step is replacing the previously substituted grid points with vacuum to restore the free surface. The starting models for v_S and ρ are derived from the starting model of v_P with equations 3.2 and 3.3, respectively. The new starting models are shown in figure 5.9. As an alternative, if no good starting model is available, Warner et al. (2013) presented a way to down-shift the frequencies of the field data to prevent cycle-skipping. Since this method requires considerable additional computational resources I will conduct this test under the assumption that a sufficient starting model is available. Such a starting model can be obtained, e.g., by travel-time tomography.
The new frequency band used in this test also requires a change in the frequency filters used. Field data often demands the use of small frequency steps (Thiel, 2018). Hence, I use the frequency band from $5\,\mathrm{Hz}$ to $11\,\mathrm{Hz}$ with a step interval of $1\,\mathrm{Hz}$. I also increase the filter order from 4 to 8. With the higher corner frequencies, compared to the previous test, this filter order becomes stable and the increase in filter order improved the convergence of the FWI significantly. The new workflow is shown in table 5.4.

Results and Discussion

The misfit development with this new frequency band is shown in figure 5.8. The misfit decreases significantly in each stage, so the chosen frequency steps are justifiable. The inverted models are shown in figure 5.9. In general they look very similar to the inverted models of the previous test (see figure 5.1). Up to a depth of ca. $3\,\mathrm{km}$ there are no obvious differences observable. The high velocity

Table 5.3: Parameters of the synthetic and the true source signal with realistic frequencies. The true signal is calculated with equation 4.1 and the synthetic signal with equation 5.1. Amplitude A has no unit because our code does not scale the sources to their true physical units. This is not problematic because recording true amplitudes is very difficult in practice, and hence rarely done.

	type	A	f_c in Hz	t_{shift} in s
true source	Ricker	1.0	10.0	0.00
synthetic source	$\sin^3$	1.5	12.0	0.15

Table 5.4: Inversion workflow. In every all three model parameters are beeing inverted. Higher stages use higher frequency content. The fk filter is used in every second stage.

Stage	v_P	v_S	ρ	filter edge in Hz	fk filter
1	✓	✓	✓	5	–
2	✓	✓	✓	5	✓
3	✓	✓	✓	6	–
4	✓	✓	✓	6	✓
5	✓	✓	✓	7	–
6	✓	✓	✓	7	✓
7	✓	✓	✓	8	–
8	✓	✓	✓	8	✓
9	✓	✓	✓	9	–
10	✓	✓	✓	9	✓
11	✓	✓	✓	10	–
12	✓	✓	✓	10	✓
13	✓	✓	✓	11	–
14	✓	✓	✓	11	✓

feature in the lower left below $4\,\mathrm{km}$ seems to be reconstructed better now but the reason for this is that it is already present in the starting model. The high velocity feature at a depth of ca. $3.5\,\mathrm{km}$ seems to have lower velocities now than the true model. In the previous test the inverted velocities of this feature match the true velocities better. To quantify this difference and to compare the inverted models of this test with the inverted models of the previous test regarding their absolute amplitudes figure 5.10 shows a vertical slice through the parameter models. One can observe that up to $y \approx 3\,\mathrm{km}$ the inverted velocity models of this test match the inverted models of the previous test. On the other hand, below $y \approx 3\,\mathrm{km}$ the inverted velocity models of this test show significant shortcomings compared to the models of the previous test. While the P-wave velocity of the feature at $y \approx 3.5\,\mathrm{km}$ is inverted almost perfectly in the previous test, the inverted velocity in this test is ca. $300\,\mathrm{m\,s^{-1}}$ less than the true velocity. This difference has to be related directly to the used frequencies. Since the high velocity zone above, which is broader in y-direction, is reconstructed well, wavelengths which are too short cannot be the reason. I assume the reason to be the different sensitivity of the higher frequencies. One important wave type for FWI are diving waves because they contain long wavelength information (Zhou et al., 2015). Their rays are bent upwards again because of the velocity gradient. Diving waves have a Fresnel zone in the shape of a "banana doughnut" (Dahlen et al., 2000). At low frequencies these are wider than at high frequencies where they ultimately converge to a ray. It is possible that the high sensitivity zone of the banana doughnut with a $5\,\mathrm{Hz}$ source barely intersects the discussed model feature but the one of a $10\,\mathrm{Hz}$ source is too thin which results in low sensitivity for the feature.

To validate that theory I calculate the finite-frequency sensitivity kernels for the two different source signals using the implementation in specfem2d which follows Liu and Tromp (2008). The sums of these kernels for all shots are shown in figure 5.12. To display these kernels I use a logarithmic color scale because of the large range of sensitivities. Regarding the previous assumption two observations are important. The first is that indeed there is an arc of high sensitivity observable in the v_P and v_S kernels with a $10\,\mathrm{Hz}$ source which has its apex just above the discussed model feature. The second observation is that the sensitivity in both the v_P and the v_S kernel is higher in $3\,\mathrm{km}$ depth when using a $5\,\mathrm{Hz}$ source than when using a $10\,\mathrm{Hz}$ source. In general the sensitivity in the v_P kernel with a $5\,\mathrm{Hz}$ source is more evenly distributed compared to $10\,\mathrm{Hz}$, so that the arc is not observable at that frequency. The even distribution has two reasons. The first is that the sensitivity of the diving wave is spread over a larger area with the $5\,\mathrm{Hz}$ source, and hence the sensitivity per unit area is smaller. The second reason is that the background sensitivity is higher in the $5\,\mathrm{Hz}$ case because surface waves penetrate deeper into the earth. Diving waves are only recorded at the maximum offset possible with this acquisition geometry. So, in order to utilize diving waves for the inversion of deep regions in the model, one needs a larger acquisition geometry.

There are some additional observations that can be made from the sensitivity kernels. In all kernels shown the by far highest sensitivity is near the surface. In the first kilometer below the surface the v_S sensitivity decreases by two orders of magnitude. The v_P sensitivity decreases even faster. Only about $100\,\mathrm{m}$ below the surface the sensitivity is already two magnitudes lower than at the surface. This directly shows the zone dominated by surface waves. The thickness of this high-sensitivity zone also depends on the frequency of the source signal. Below the high-sensitivity zone the values are relatively stable up to a depth of ca. $3\,\mathrm{km}$. Especially with the $10\,\mathrm{Hz}$ source one can see from the density kernel that there is a high sensitivity at the model interfaces but not in the large regions in between which explains why the inverted density models shown so far correctly display interfaces but completely fail to match the absolute values of the true models.

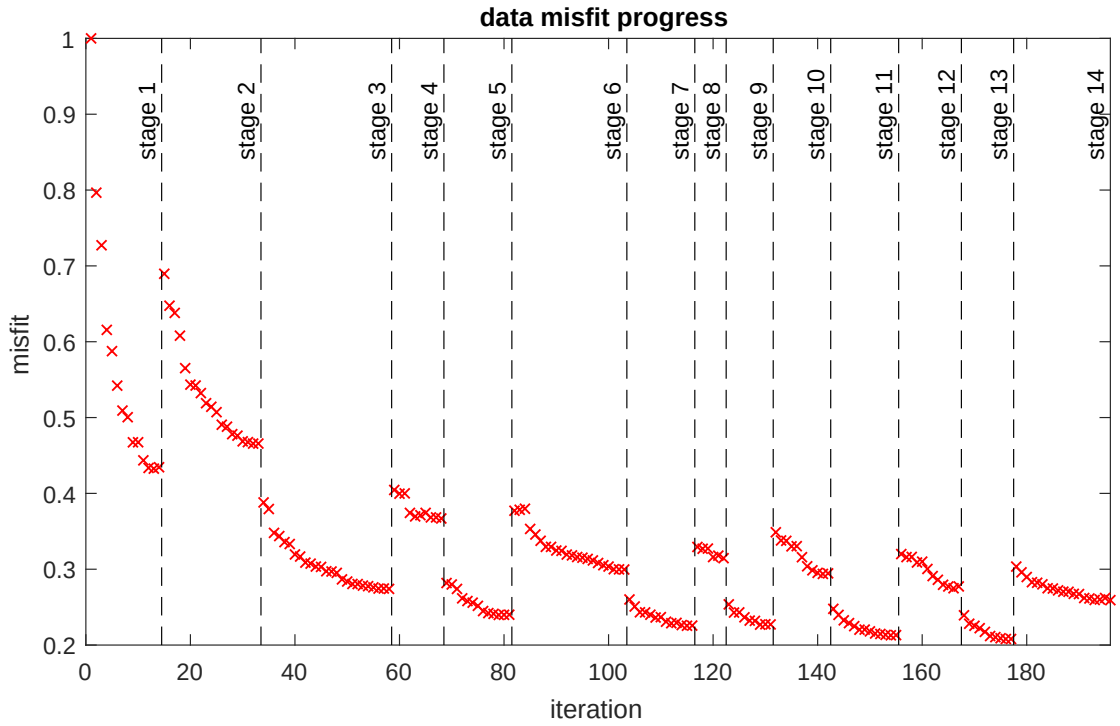

Figure 5.8: Progression of the data misfit for all iterations during the inversion of SEM data with a center frequency of 10 Hz. Stage changes are indicated by vertical dashed lines. The misfit is normalized to that of iteration 0 witch is the starting model. In all stages the inversion is done for all three model parameters. In each stage with an even number an fk filter is being used.

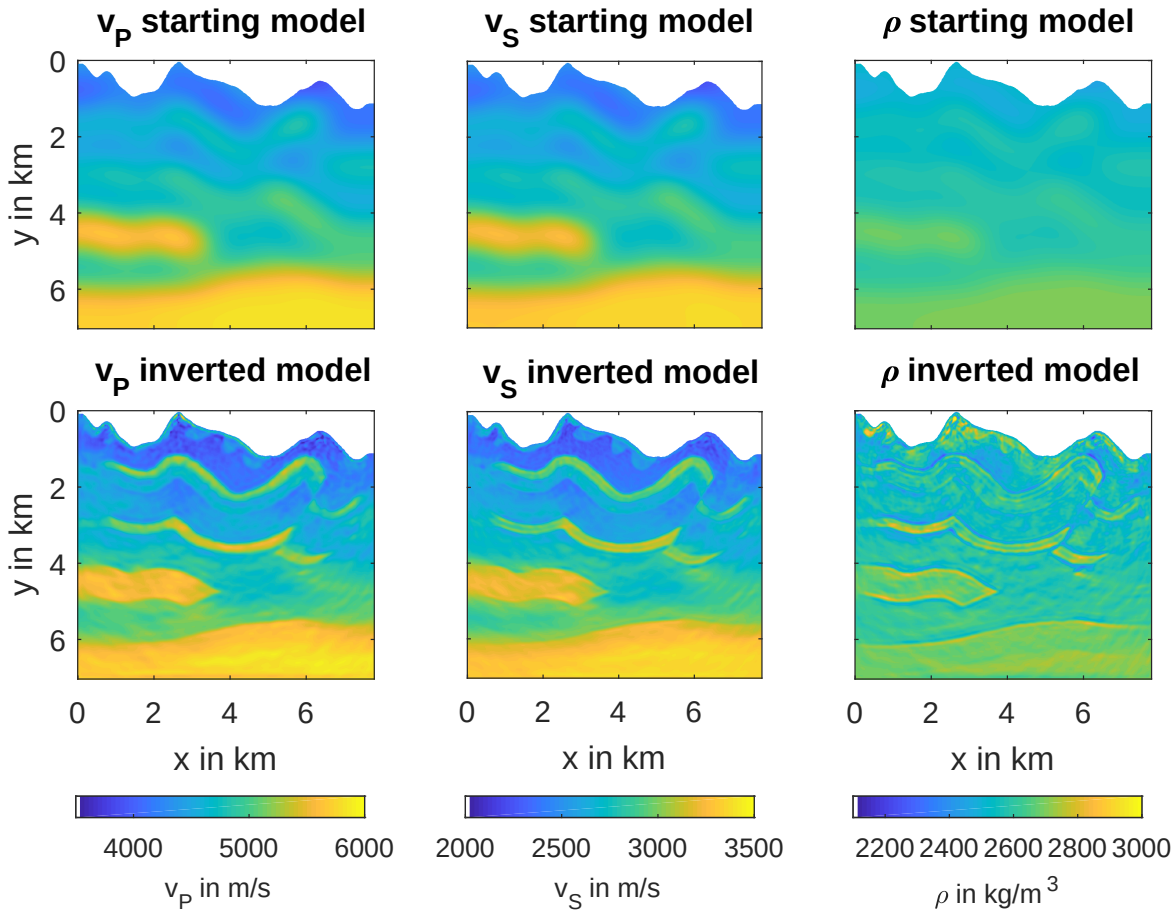

Figure 5.9: The top row shows the starting models used for the test with a 10 Hz source wavelet. The bottom row shows the resulting inverted models. White represents the area above the free surface where all parameters are set to vacuum.

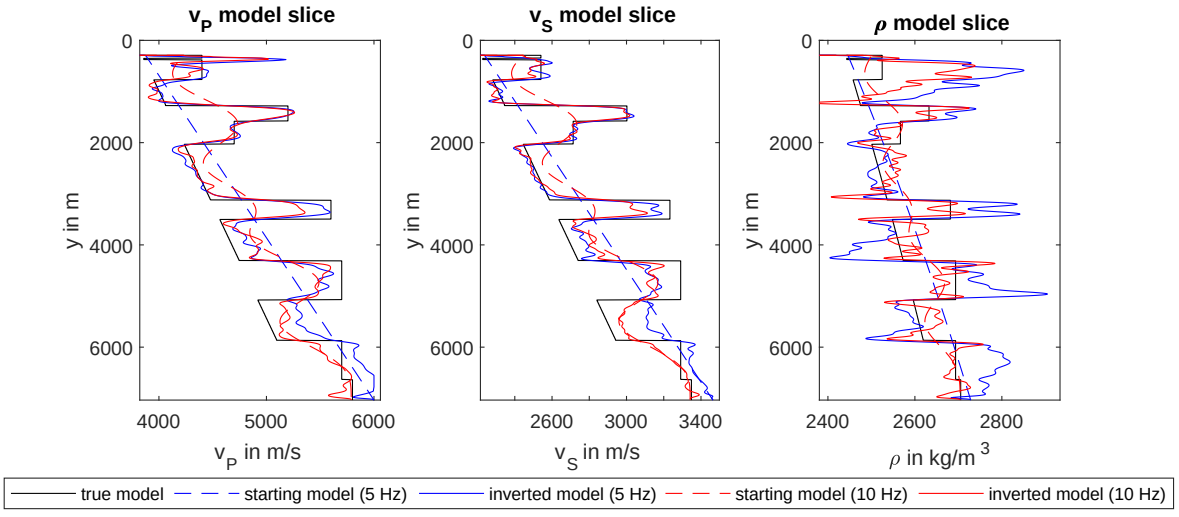

Figure 5.10: Vertical slices through parameter models at $x = 3000\,\mathrm{m}$. This figure shows a comparison of the inverted models using SEM data of the test in section 5.1.1 (5 Hz) and the test in section 5.1.2 (10 Hz). The corresponding starting models are also shown for reference.

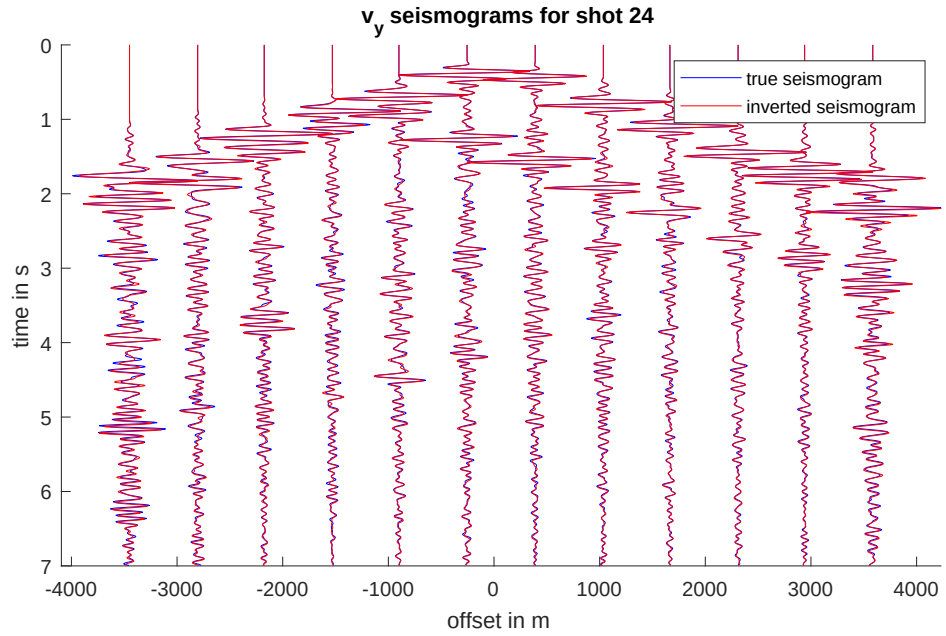

Figure 5.11: Inverted and observed seismograms after the last iteration. The displayed shot is at $x = 3817.5\,\mathrm{m}$ which is almost in the center of the model. For better visualization there is shown the vertical component of 12 equidistant receivers which is only a small subset of the 469 2-component receivers used. For the same reason traces are normalized and scaled linear with time to boost late arrivals.

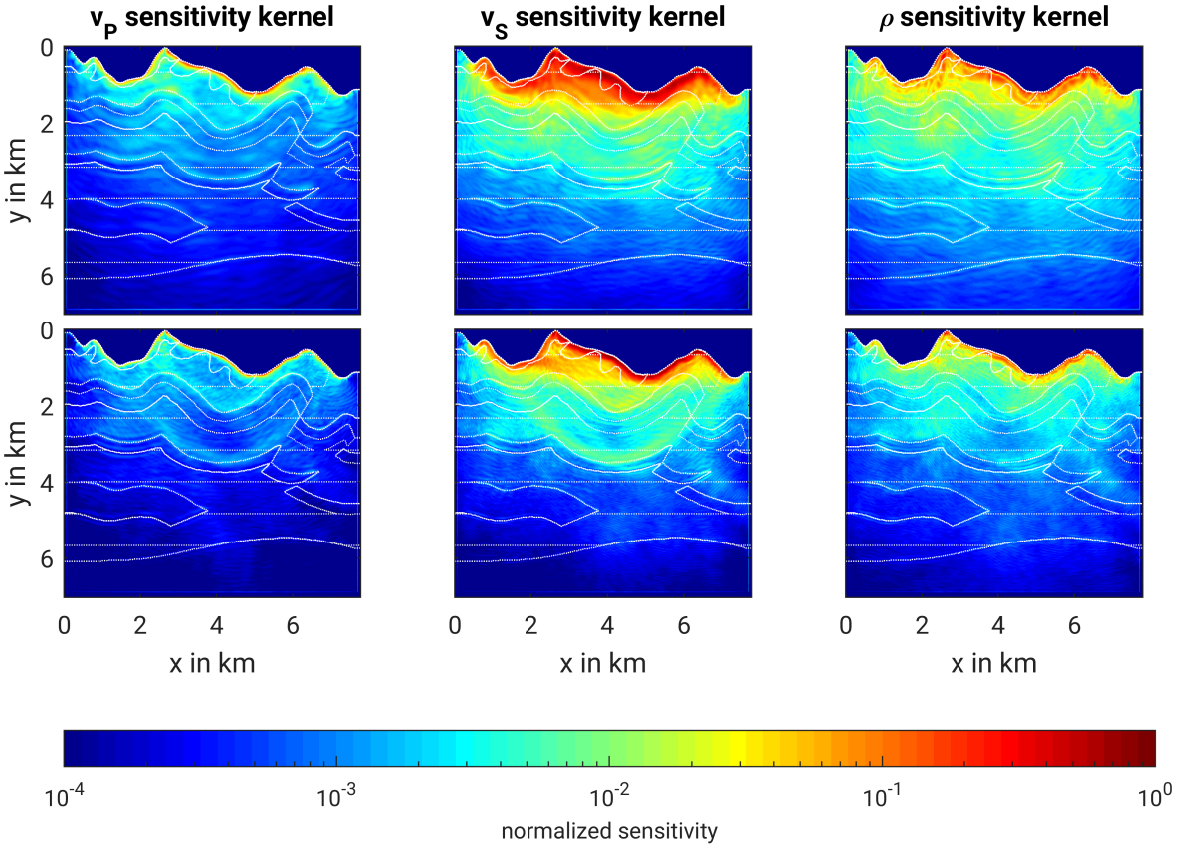

Figure 5.12: Finite frequency sensitivity kernels for a source with a center frequency of 5 Hz (top row) and a center frequency of 10 Hz (bottom row). Shown are the absolute values of the sensitivity normalized to their maximum. For better orientation the model is shown in white dotted contour lines.

5.2 Inversion with 3D effects

Goal of this section is to invert the 2.5D model and 3D model presented in section 4.3 using 2D FWI. The results from the comparison of the 2D and 3D seismograms were very similar in case of the 2.5D model, so it is expected that this test will be successful. For the 3D model this is not the case, so it is unlikely for the inversion to be successful.

Methodology

As observed data I use the converted seismograms of the test in section 4.3. For a detailed description of the acquisition geometry I also refer the reader to this section. Like in the previous inversion tests I use STF inversion to better mimic realistic conditions. The source signal used for the STF inversion is described in section 5.1.2. Since the pseudo-observed data is generated with the FD method again, the near surface artifacts observed in the tests with SEM should now be vanished again. As a starting model I choose the same starting model used in section 5.1.2 because the center frequency of my source signal is still $10\,\mathrm{Hz}$. Since the true 3D models are slightly smaller in the vertical direction the starting model is adjusted accordingly. Also the inversion strategy is the same as in the previous section.

Results

The inversion result when using the 2.5D true model is shown in figure 5.13. Compared to previous inversion results the inverted v_P model has a deteriorated quality. While the high velocity artifact near the surface vanished as predicted, the rest of the model seems to have more fluctuations than before. Also the velocities of the model feature at ca. $2\,\mathrm{km}$ depth are not accurately reconstructed. This cannot be said about the inverted v_S model. Here up to a depth of ca.$3\,\mathrm{km}$ the absolute values of the inverted velocities almost perfectly match the true velocities. A likely reason for this behavior is the incorrect amplitude normalization discussed in section 4.3.1. This mostly affects the v_P model since the amplitudes of body waves in the converted 3D seismograms are off by a constant factor. The relative amplitudes of the surface waves are correct because of trace normalization and as shown in section 5.1.2 their sensitivity for v_S goes much deeper than their sensitivity for v_P. However, a decreased quality of the inverted v_P model and an unaffected inverted v_S model matches the results of Auer et al. (2013) who investigated asymptotic 3D to 2D transformations for crosshole FWI. They use the same phase correction but different amplitude factors and probably no normalized seismograms during the inversion. Since in a crosshole experiment no surface waves occur there might be other effects contributing to the decreased quality of the inverted v_P model.
The result when using the 3D model is shown in figure 5.14. One can see that this inversion fails completely. There is only a slight improvement to the starting model and strong artifacts are now observable in the v_P and v_S models. This result is expected because of the poor similarity between the 2D and 3D seismograms which is shown in section 4.3.
The conclusion drawn from this test is that 2D FWI is only possible if the topography of the free

surface is homogeneous in at least one dimension. In case of a surface which varies in all directions 2D FWI fails and 3D wave propagation has to be considered. One could test, if slight 3D variations of the free surface topography lead to errors which are acceptable. However, even finding an area which has topography but only weak 3D variations of it in the scale of exploration seismics in the real world is nearly impossible. Hence, I do not investigate this any further.

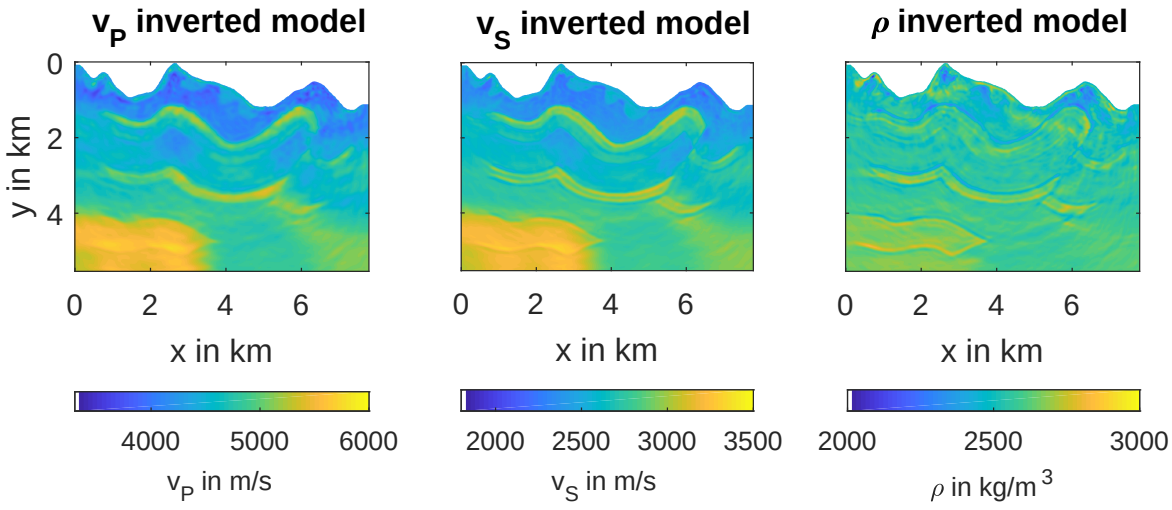

Figure 5.13: Inverted models with a 2.5D true model. White represents the area above the free surface where all parameters are set to vacuum.

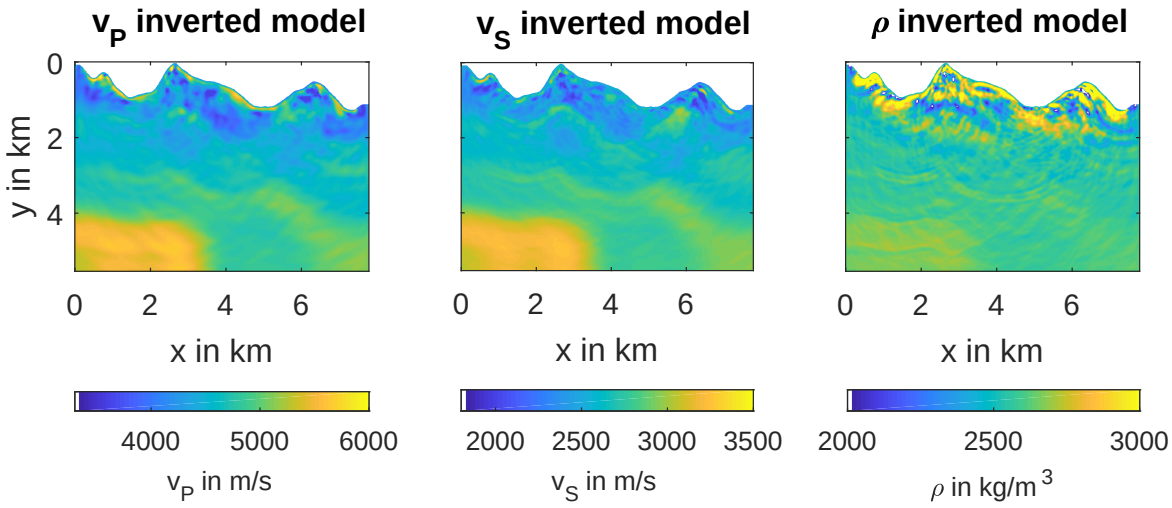

Figure 5.14: Inverted models with a 3D true model. White represents the area above the free surface where all parameters are set to vacuum.

Chapter 6

Summary & Conclusions

The goal of this work is to investigate the feasibility of the 2D FD method for FWI applications in regions with strong topographic variations of the Earth's surface on the scale of exploration seismics. Because one of the most characteristic features of land data are strong surface waves, which are also scattered at the surface topography, I use the elastic wave equation in my work. The free surface is implemented using the improved vacuum formulation which allows the creation of models with, in principle, arbitrarily shaped free surfaces. Since this formulation fulfills the traction-free boundary condition via the model parameters and no special treatment of the free surface grid points is needed during the simulation it is very efficient and thus well suited for FWI applications. The only considerable drawback of this formulation is that continuous shaped surfaces are approximated by staircases. I show that these staircases can have a large influence on waves which travel parallel to the free surface such as surface waves. It is shown that coarser discretizations shift the arrival of the main energy of surface waves to later times. The effect of the staircases is reduced the finer the model is discretized. However, for an efficient use of the FD method such fine discretizations are not an option. In case one wants to model surface waves with a very high accuracy the spectral element method, which is used as a reference in my work, is, although more difficult to implement and to use, the more suitable method. The FD solution seems to converge towards the SEM solution for very fine discretizations, although, complete convergence is not reached in my tests. In my inversion tests I use FD-based FWI to invert synthetic data simulated with the SEM and show that the differences in the seismograms, even when coarse discretizations are used, don't have a major influence on the inversion result. The inverted models of the P-wave velocity and S-wave velocity are very similar to the true models used. The only notable artifact the use of SEM data produces is a thin high-velocity zone just below the free surface of the v_P-model which balances the observed time shift. In my inversion tests I also show that because of surface waves the near-surface region of the v_S-model is especially well reconstructed and highly resolved. However, because of their large amplitudes they dominate the inversion so they should, if possible, also be filtered in some stages of the inversion to obtain a better reconstruction of deeper model regions. In all of my inversion tests I was not able to invert the true densities but only density contrasts. To mimic more realistic conditions I also invert the source signal. These inverted signals match the true signals almost perfectly. In addition to the validation of the FD method I also investigate the influence of 3D effects, such as they occur in regions with strong topography, on 2D

FWI. Here, I am able to show that phases and relative amplitudes of 3D point sources can be easily transformed to their 2D equivalent in case the model is 2.5D i.e., has no variations perpendicular to the acquisition line. If such variations are present, which is probably always the case in the real world, the transformation fails and hence also 2D FWI. Only in case of 2.5D models the 2D FWI yields sufficient inverted models. However, even in this case the errors done by the transformation lead to a deterioration of the inverted models.

Bibliography

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

Auer, L., Nuber, A. M., Greenhalgh, S. A., Maurer, H., and Marelli, S. (2013). A critical appraisal of asymptotic 3D-to-2D data transformation in full-waveform seismic crosshole tomography. *Geophysics*, 78(6):R235–R247.

Berg, P. (1994). EX2DDIR. `http://www.quest-itn.org/library/software/ex2ddir-ex2delel-ex2dvael`.

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

Cerjan, C., Kosloff, D., Kosloff, R., and Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):705–708.

Courant, R., Friedrichs, K., and Lewy, H. (1928). Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100(1):32–74.

Dahlen, F., Hung, S.-H., and Nolet, G. (2000). Fréchet kernels for finite-frequency traveltimes – I. theory. *Geophysical Journal International*, 141:157–174.

Dougherty, M. E. and Stephen, R. A. (1988). Seismic energy partitioning and scattering in laterally heterogeneous ocean crust. In *Scattering and Attenuations of Seismic Waves, Part I*, pages 195–229. Springer.

Fichtner, A. (2011). *Full seismic waveform modelling and inversion*. Springer Science & Business Media.

Forbriger, T., Groos, L., and Schäfer, M. (2014). Line-source simulation for shallow-seismic data. part 1: Theoretical background. *Geophysical Journal International*, 198(3):1387–1404.

Gardner, G. H. F., Gardner, L. W., and Gregory, A. R. (1974). Formation Basics for Velocity and Density - The Traps Diagnostic Stratigraphic. *Geophysics*, 39(6):770–780.

Gercek, H. (2007). Poisson's ratio values for rocks. *International Journal of Rock Mechanics and Mining Sciences*, 44(1):1 – 13.

Graves, R. W. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America*, 86(4):1091–1106.

Gray, S. H. and Marfurt, K. J. (1995). Migration from topography: Improving the near-surface image. *Canadian journal of exploration geophysics*, 31(1):18–24.

Igel, H. (2017). *Computational seismology - a practical introduction*. Oxford University Press.

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

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

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

Komatitsch, D. and Tromp, J. (1999). Introduction to the spectral element method for three-dimensional seismic wave propagation. *Geophysical Journal International*, 139(3):806—-822.

Komatitsch, D. and Vilotte, J. P. (1998). The spectral-element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *bssa*, 88(2):368–392.

Kunz, K. S. and Luebbers, R. J. (1993). *The finite difference time domain method for electromagnetics*. CRC Press.

Levander, A. R. (1988). Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53(11):1425–1436.

Liu, Q. and Tromp, J. (2008). Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods. *Geophysical Journal International*, 174:265–286.

Mahrooqi, S., Rawahi, S., Yarubi, S., Abri, S., Yahyai, A., Jahdhami, M., Hunt, K., and Shorter, J. (2012). Land seismic low frequencies: acquisition, processing and full wave inversion of 1.5 - 86 Hz. In *SEG Technical Program Expanded Abstracts 2012*.

Martin, R., Komatitsch, D., Blitz, C., and Le Goff, N. (2008). Simulation of seismic wave propagation in an asteroid based upon an unstructured MPI spectral-element method: blocking and non-blocking communication strategies. 5336:350–363.

Mei, J. and Tong, Q. (2015). A practical acoustic full waveform inversion workflow applied to a 3d land dynamite survey. In *2015 SEG New Orleans Annual Meeting*.

Mittet, R. (2002). Free-surface boundary conditions for elastic staggered-grid modeling schemes. *Geophysics*, 67(5):1616–1623.

Moczo, P., Kristek, J., Vavryčuk, V., Archuleta, R., and Halada, L. (2002). 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities. 92(8):3042–3066.

Morse, P. M. and Feshbach, H. (1953). Methods of theoretical physics. *New York: McGraw-Hill*.

Müller, G. (2007). *Theory of Elastic Waves*. Scientific Technical Report 2007: GeoForschungsZentrum Potsdam.

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

Plessix, R., Milcik, P., Rynja, H., Stopin, A., Matson, K., and Abri, S. (2013). Multiparameter full-waveform inversion: Marine and land examples. *The Leading Edge*, 32(9):1030–1038.

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

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

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

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

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

Przebindowska, A., Kurzmann, A., Köhn, D., and Bohlen, T. (2012). The role of density in acoustic full waveform inversion of marine reflection seismics. In *74th EAGE Conference and Exhibition incorporating EUROPEC 2012*.

Robertson, J. O. (1996). A numerical free-surface condition for elastic/viscoelastic finite difference modeling in the presence of topography. *Geophysics*, 61:1921–1934.

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

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

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, 51(10):1893–1903.

Thiel, N. (2018). *Acoustic and elastic FWI of marine dual-sensor streamer data in the presence of salt*. PhD thesis, Karlsruhe Institute of Technology (KIT).

Trinh, P.-T. (2018). *3D Multi-parameters Full Waveform Inversion for challenging 3D elastic land targets*. PhD thesis, Université Grenoble Alpes.

Vestrum, R. (2007). Geoscience integration for seismic imaging in thrust-belt environments. In *SEG Technical Program Expanded Abstracts 2007*, pages 2792–2796. Society of Exploration Geophysicists.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method. *Geophysics*, 51(4):889–901.

Warner, M., Nangoo, T., Nikhil, S., Umpleby, A., and Morgan, J. (2013). Full-waveform inversion of cycle-skipped seismic data by frequency down-shifting. In *SEG Houston 2013 Annual Meeting*.

Xie, Z., Komatitsch, D., Martin, R., and Matzen, R. (2014). Improved forward wave propagation and adjoint-based sensitivity kernel calculations using a numerically stable finite-element PML. 198(3):1714–1747.

Zahradník, J., Moczo, P., and Hron, F. (1993). Testing four elastic finite-difference schemes for behavior at discontinuities. *Bulletin of the Seismological Society of America*, 83(1):107–129.

Zeng, C., Xia, J., Miller, R. D., and Tsofilas, G. P. (2012). An improved vacuum formulation for 2D finite-difference modeling of rayleigh waves including surface topography and internal discontinuities. *Geophysics*, 77(1):T1–T9.

Zheng, Y., Zhang, W., Xue, Q., and Chang, X. (2016). An application of full-wavform inversion to land seismic data unsing the pseudo-hessian matrix. *Interpretation*, 4(4):2324–8866.

Zhou, W., Brossier, R., Operto, S., and Virieux, J. (2015). Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation. *Geophysical Journal International*, 202(3):1535–1554.

List of Figures

List of Tables

Chapter 7

Acknowledgements

First of all, I want to thank professor Thomas Bohlen for the interesting topic and his guidance during my work. I also would like to thank professor Andreas Rietbrock for accepting to be the second reviewer. Next, I thank Tilman Steinweg for helping me with many technical problems and also for the fruitful discussions about my topic. I would like to thank Yudi Pan for the help and discussions regarding surface waves and topography. In particular I also would like to thank Thomas Hertweck for proofreading my thesis.

Next, I want to thank all other partners of the WAVE Project especially Thomas Brandes from the Fraunhofer Institute for algorithms and scientific computing (SCAI) for his support with their HPC library LAMA. Also special thanks to TEEC GmbH for inviting me to Hannover to get insight into their company and for providing me my model.

Furthermore, I want to thank all other members of the working group Applied Geophysics at KIT for the pleasant working environment and especially thanks to Valérie Krampe and Fabian Kühn with whom I shared the office most of the time. But I also want to thank all other Master's students. Many thanks also to Claudia Payne for the help with all sorts of administrative issues and to Petra Knopf for the excellent IT support. In particular I want to thank Renat Shigapov who shared his computational resources at the ForHLR II cluster with me.

Last but not least, I want to thank my parents and friends for the support not only during my thesis but also during all of my time at KIT.

Chapter 8

Declaration on oath

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

Daniel Krieger

Appendix A

Comparison of FD and SEM seismograms

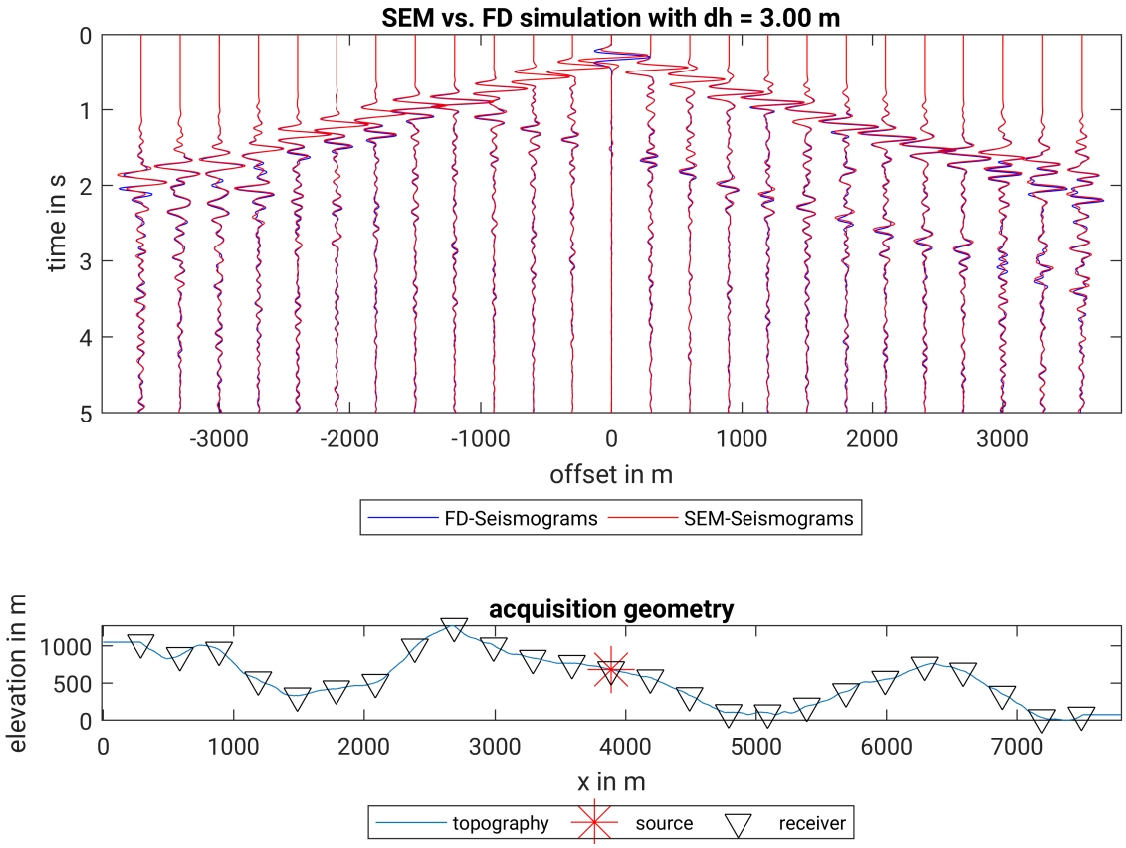

Figure A.1: Comparison of SEM seismograms and FD seismograms generated with a model discretization of dh=3 m. The top plot shows a shot gather with a selection of receivers and the bottom plot shows the elevation for reference. The lowest point of the free surface has an elevation of 0 m.

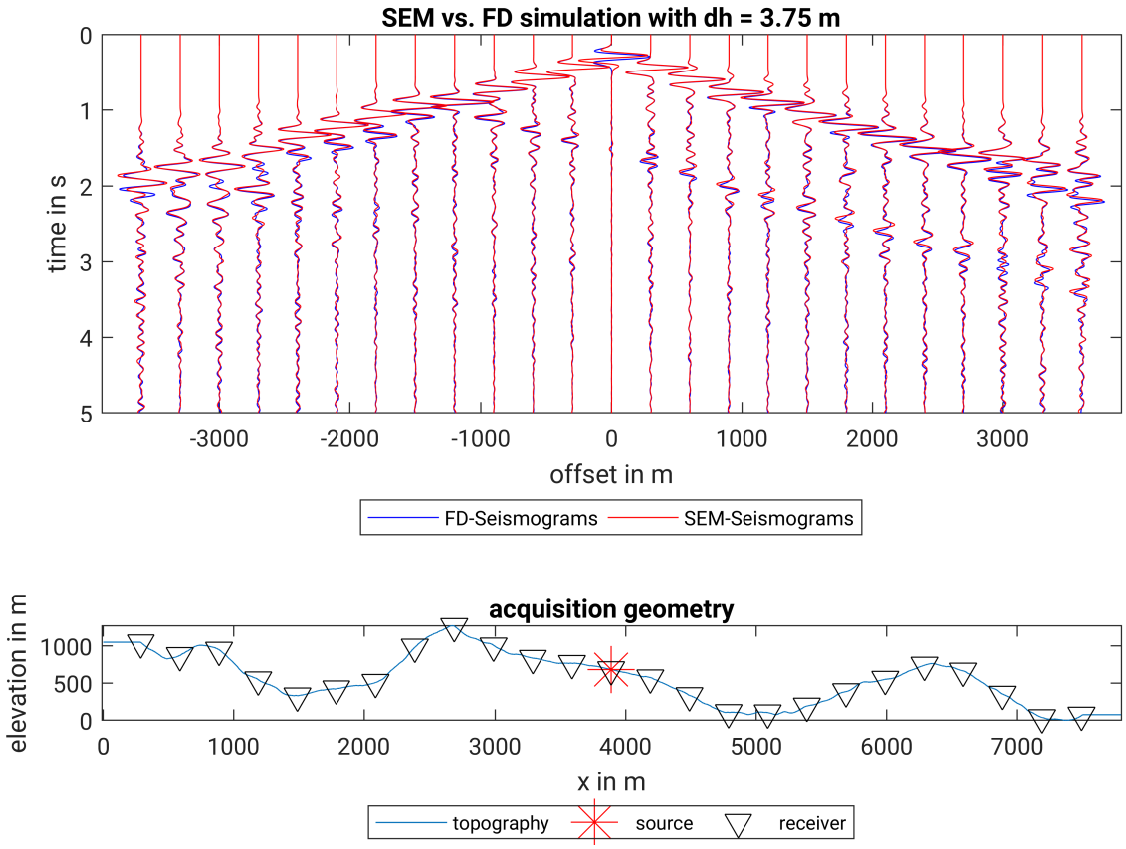

Figure A.2: Comparison of SEM seismograms and FD seismograms generated with a model discretization of dh=3.75 m. The top plot shows a shot gather with a selection of receivers and the bottom plot shows the elevation for reference. The lowest point of the free surface has an elevation of 0 m.

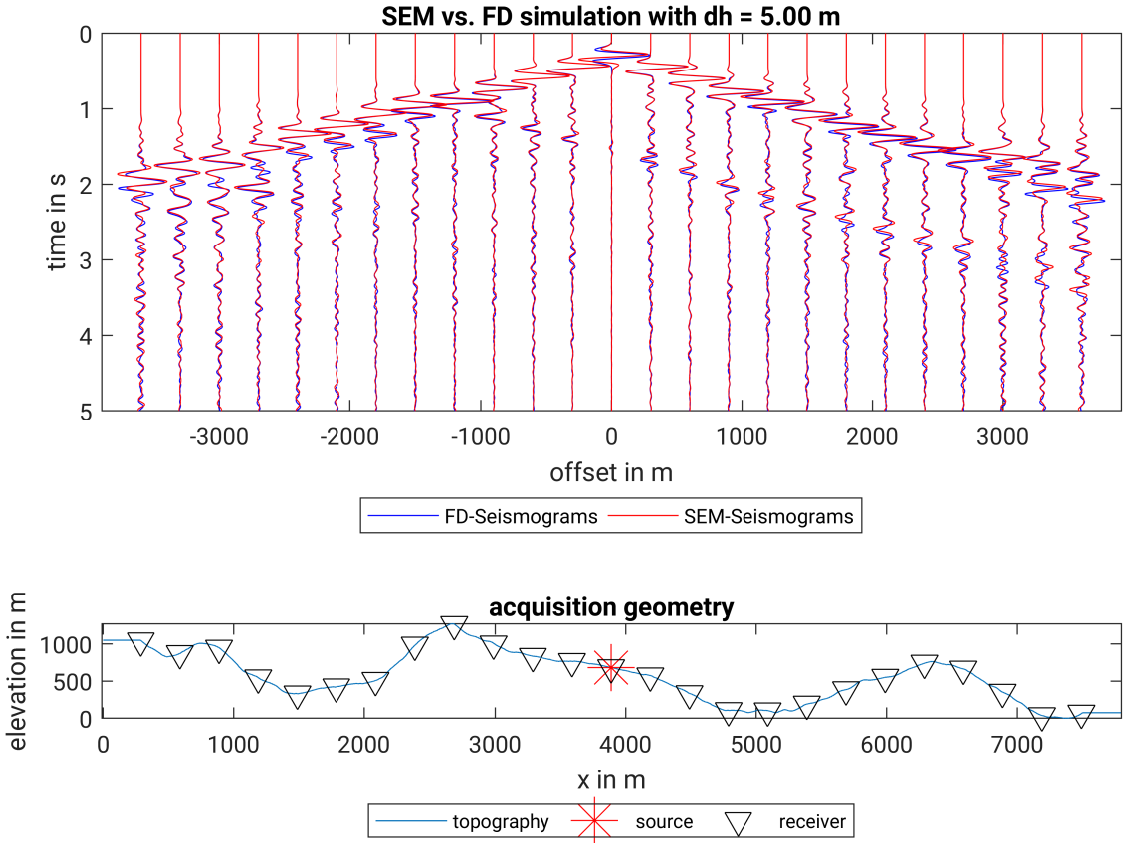

Figure A.3: Comparison of SEM seismograms and FD seismograms generated with a model discretization of dh=5 m. The top plot shows a shot gather with a selection of receivers and the bottom plot shows the elevation for reference. The lowest point of the free surface has an elevation of 0 m.

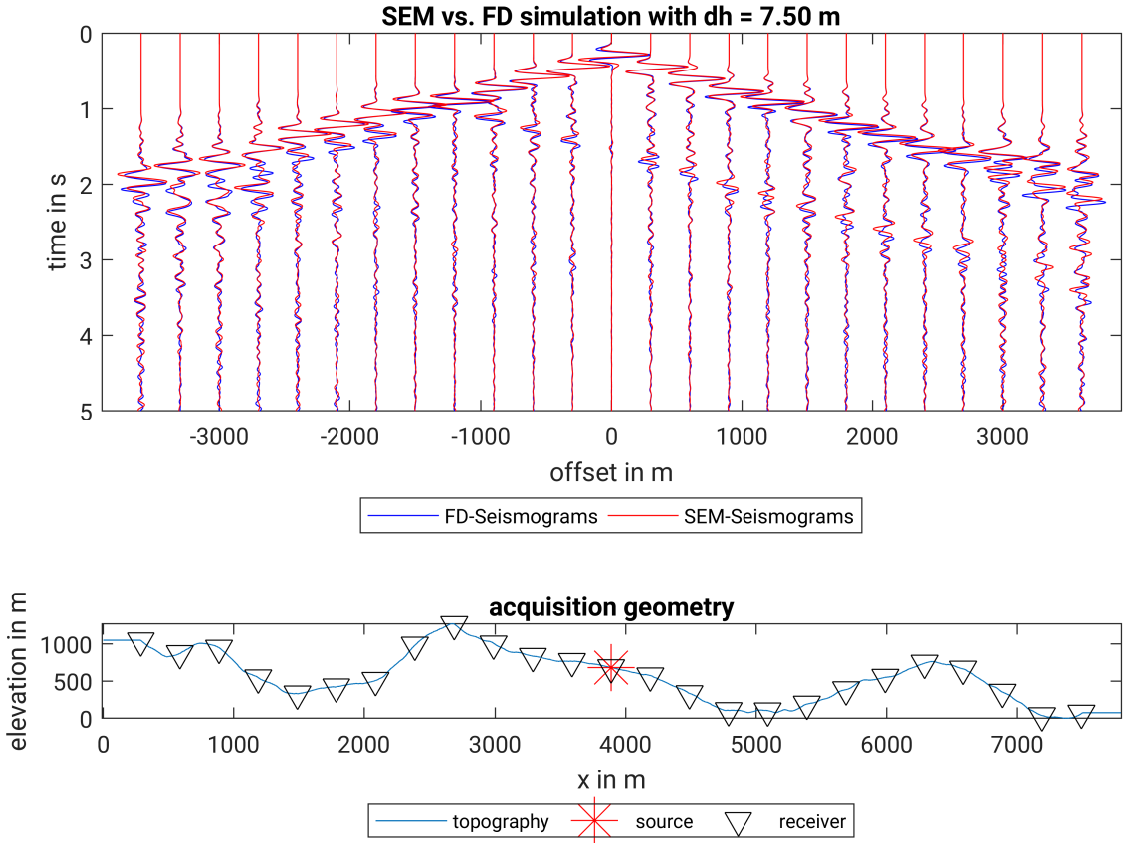

Figure A.4: Comparison of SEM seismograms and FD seismograms generated with a model discretization of dh=7.5 m. The top plot shows a shot gather with a selection of receivers and the bottom plot shows the elevation for reference. The lowest point of the free surface has an elevation of 0 m.

Appendix B

3D model generation

There is no straight forward solution to generate a 3D model from a 2D model. In my case the purpose of the 3D model is to investigate the effect of 3D heterogeneous topography. Hence, the geological features below the surface can and also should be homogeneous in the third dimension. This means, the only heterogeneity I need to add is that of the free surface. As a starting point for the 3D model generation I choose the 2.5D model described in section 4.3.1. Next, I construct a random surface with the constraint that the curve at $z = \mathrm{nz}/2 \cdot \mathrm{dh}$ is the same as the surface of my 2D model. To avoid confusion, from now on I refer to the surface of the 2D model as line surface. As a second constraint I choose that the surface has to be inside the bounds of the line surface. I.e., the surface should at no point be lower/higher than the lowest/highest point of the line surface. This has the advantage that no extremely high mountains or extremely deep valleys are generated. In practice, I construct this surface by generating random curves which are parallel to the z-axis, intersect the line surface and are equally spaced. Then, I interpolated these curves and the line surface on the entire xz-plane to form a surface. As interpolation algorithm I chose the $v4$ method implemented in MATLAB because, unlike cubic interpolation, it causes no strong fluctuations.

With this method I am able to obtain a surface which looks natural and has height differences in the same order of magnitude as the ones present in the 2D model. Now only at a few locations the new surface has the same height as the surface of the 2.5D model. In the next step I set all grid-points above the new surface to vacuum. All grid points which are vacuum in the 2.5D model but lay below the new surface I fill with the value of the grid-point right at the surface of the 2.5D model. The model generation procedure is visualized in figure B.1.

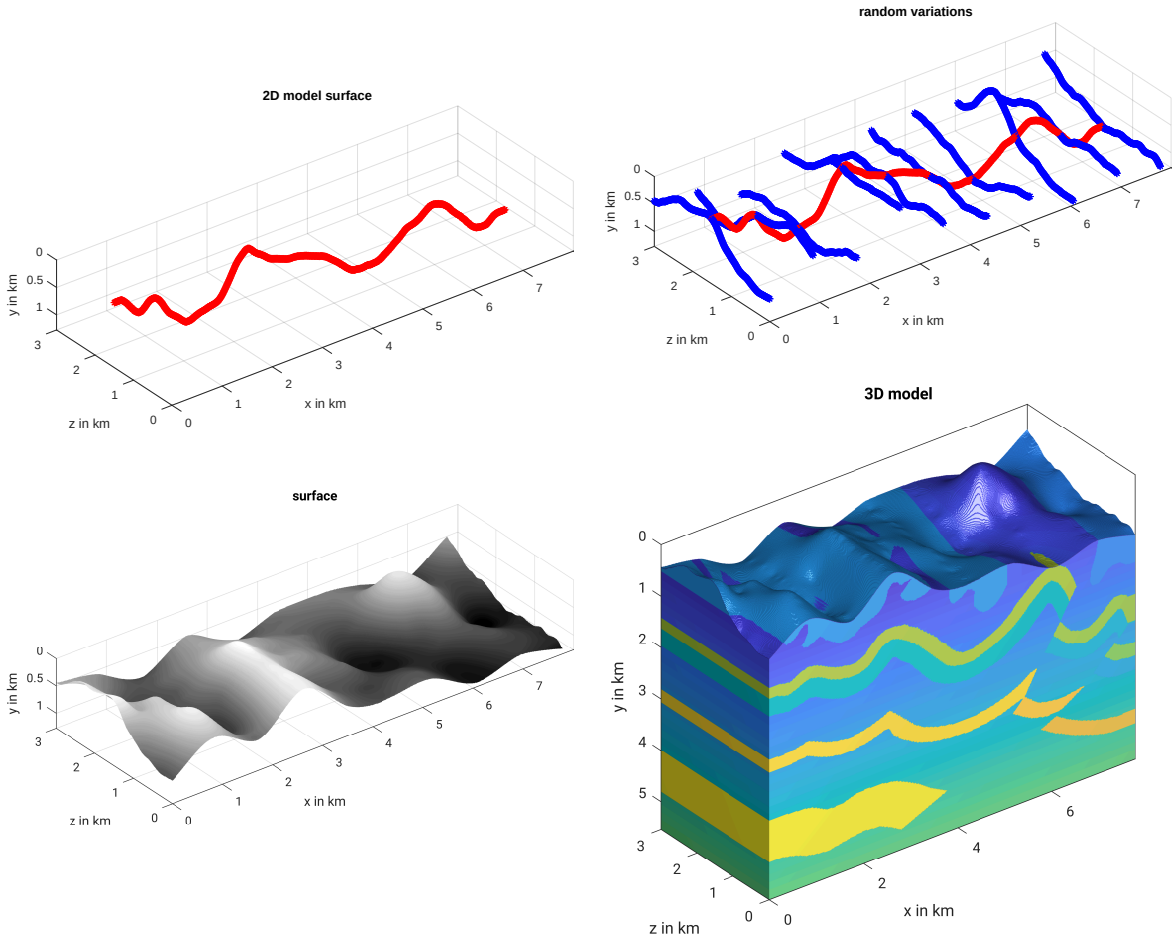

Figure B.1: Illustration of the 3D model creation process. The procedure starts with the surface line of the 2D model (upper left). Then, random curves are added which are parallel to the z-axis, intersect the surface line and have a constant spacing in x-direction (upper right). These curves are then interpolated on the entire xz-plane (lower left). The surface is then combined with the 2.5D model (figure 4.8) by cutting out or filling up the differences (lower right).