Applications of 2D and 3D full waveform tomography in acoustic and viscoacoustic complex media

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

der Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

**Dipl.-Geophys. André Kurzmann
aus Marienberg**

Tag der mündlichen Prüfung: 30. November 2012

Referent: Prof. Dr. Thomas Bohlen

Korreferent: Prof. Dr. Jan van der Kruk

Contents

Chapter 1

Introduction

Full waveform tomography (FWT) is a high-resolution imaging method to exploit the full richness of recorded waveforms. In contrast to conventional imaging methods, such as first-arrival tomography, the FWT is able to resolve subsurface structures smaller than a wavelength. The fundamental method was developed by the pioneer A. Tarantola who proposed the theory of acoustic FWT in the time domain (Tarantola, 1984). It represents a local optimization method with the aim to find a subsurface model which explains the full waveforms of the recorded data. It involves the least-squares objective function and employs the method of steepest-descent gradients. Further developments comprise elastic time-domain FWT (e. g., Tarantola, 1986a,b; Mora, 1987; Crase et al., 1990) and the establishment of the frequency-domain FWT by R. G. Pratt (e. g., Pratt and Worthington, 1990; Pratt, 1990; Pratt and Goulty, 1991; Pratt, 1999) representing a commonly used approach in various applications.

Whether applied in time domain or frequency domain, there are numerous fields of application for waveform tomography. Most of the applications concern geophysical problems, i. e., investigating subsurface structures to improve our knowledge of the Earth's interior, to enhance subsurface imaging in geotechnical applications or to obtain detailed images of geological structures for the exploration of oil and gas deposits. The fields of application are manifold covering a wide range from large-scale seismological problems to applications at small scale lengths in ultrasound measurements. For example, FWT is applicable in cross-well experiments involving seismic measurements (e. g., Pratt, 1999), in ground-penetrating-radar measurements with regard to engineering (e. g., Kalogeropoulos et al., 2011) or near-surface investigations (e. g., Ernst et al., 2007), in analyzing near-surface structures by applications of acoustic and elastic shallow seismics (e. g., Smithyman et al., 2009; Romdhane et al., 2011), in on-shore seismics (e. g., lithospheric imaging done by Brenders and Pratt, 2007a), in marine seismics (e. g., Shipp and Singh, 2002) and in seismology (e. g., Fichtner et al., 2009), among others. Apart from geophysics, in recent years the FWT becomes more popular in several sciences, such as human medicine (e. g., Pratt et al., 2007) or helioseismology (e. g., Cobden et al., 2011).

This thesis employs the time-domain FWT proposed by Tarantola (1984). It is targeted to problems allowing a meaningful application of acoustic waveform tomography, such as cross-well measurements or marine-seismic surveys. In general, seismic waveforms are affected by several physical parameters: P-wave velocity, S-wave velocity, density, attenuation or anisotropy. While marine reflection data are dominated by P-waves with

no direct measurement of surface waves or S-waves (apart from converted waves), cross-well data generally reveal well-separated P-waves allowing a simple cut-off of S-wave events. I exploit the acoustic FWT, because it shows a good trade-off between computational efforts and accuracy with respect to the fields of application mentioned above. Furthermore, within the scope of this thesis, I carry out numerous studies, which exclusively consider the impact of several methodological strategies on the reconstruction of the isotropic P-wave velocity. Although the underlying methodology considers the full acoustic parameterization, I assume a known density distribution and a known source-time function. Hence, the single-parameter inversion allows to focus on the main purposes of this work: investigating and improving the performance of the acoustic FWT. The implementation character of this work is related to both computational and methodological optimization. That comprises the efficient application of FWT on parallel supercomputers as well as the reduction of computational efforts by increasing the convergence of the gradient algorithm. Nevertheless, all strategies aim to obtain a satisfactory P-wave velocity model with reasonable efforts.

This work consists of two main objectives: the development of 2D and 3D acoustic FWT implementations in the time domain as well as their synthetic application to complex geological models. With this focus, I perform cross-well and marine reflection experiments including fictional and realistic acquisition geometries.

1.1   Outline of this thesis

This thesis is divided into two parts. This involves the basic theory of acoustic FWT as well as issues concerning methodology and implementation. I perform several numerical studies with applications of 2D and 3D waveform tomography based on those fundamentals. The first part consists of chapters 2 and 3, while the second part includes chapters 4 to 7. The detailed structure is as follows.

Chapter 2

The second chapter describes the basic inversion scheme and the fundamentals of acoustic FWT. That includes an overview of the underlying theory and comprehensive derivations of the required wave equations and their finite-difference solutions. I focus on strategies affecting the convergence and the model reconstruction. In particular, I explain the adaptive step length estimation using a parabolic curve-fitting method and illustrate its notable advantage as well as possible drawbacks.

Chapter 3

In the first part of chapter 3, I describe technical methods which improve the computational performance of the FWT and allow an efficient application to complex problems. The explanations incorporate the massive parallelization comprising domain decomposition (Bohlen, 1998) and shot parallelization (Kurzmann et al., 2009). Additionally, I perform a benchmark to emphasize advantages and problems of these techniques.

In the second part of this chapter, I discuss strategies to reduce the consumption of resources allowing the feasibility of a pure time-domain FWT in 3D. The last subject is the workflow implementation representing the key feature of all applications in this work. It combines existing multi-stage approaches (e. g., Bunks et al., 1995; Shipp and Singh, 2002; Sirgue and Pratt, 2004; Shin and Cha, 2009) with additional features, such as the availability of pure time-domain inversion and the so-called single-frequency method (Sirgue et al., 2008) within the same FWT scheme. The intention of the multi-stage approaches is to mitigate the ill-posedness of the inverse problem. They contain time windowing and offset windowing (e. g., Shipp and Singh, 2002; Wang and Rao, 2009) as well as the sequential inversion for different frequency contents (e. g., Bunks et al., 1995; Sirgue and Pratt, 2004; Brossier et al., 2009; Shin and Cha, 2009). However, the workflow mainly focuses on the frequency selection due to its highest importance in a successful model recovery (Sirgue, 2006).

Chapter 4

Within the scope of a parameter study, this chapter is composed of numerous experiments. I investigate methodological impacts of important steps of the inversion scheme on the outcome of the FWT. Based on the application of multiple FWTs in each experiment, this study aims to give an impression, how the objective function of the inverse problem is affected. This investigation is concluded by a brute-force experiment showing the cross sections of the objective function in dependence of considering different frequency contents.

Chapter 5

In this chapter, I intend to classify the pure time-domain approach used in this work, i. e., the recovered P-wave velocity model is compared with the results of the single-frequency method (Sirgue et al., 2008) and a pure frequency-domain FWT (application of the implementation developed by R. G. Pratt; Pratt, 1999).

Chapter 6

While most of the applications in this work only consider purely acoustic problems, it is interesting to investigate phenomena which cannot be explained by a 2D acoustic FWT. For example, this relates to seismic data recorded in a 3D environment, the occurrence of elastic effects or the appearance of attenuation. In this work I investigate the feasibility of acoustic FWT in presence of attenuation. In general, the consideration of attenuation involves multi-parameter viscoacoustic inversions commonly applied in the frequency domain. The intention is either to obtain a reliable attenuation model or to improve the recovery of the velocity model at the expense of an artificial attenuation model. Applications are published by several authors, such as Hak and Mulder (2008, 2011) and Kamei and Pratt (2008). However, attenuation is also used as a passive parameter in an acoustic FWT with the aim to improve the velocity model (e. g., Brenders and Pratt, 2007a).

Based on two marine reflection experiments with realistic acquisition geometries, I apply the acoustic time-domain FWT to seismic data affected by the intrinsic attenuation of the subsurface. This synthetic study aims to quantify the error in the recovered velocity model

by assuming both "good" and "poor" passive attenuation models, i. e., attenuation is not subject to the inversion. Furthermore, I identify, on which condition the acoustic inversion of viscoacoustic data yields an acceptable velocity model.

Chapter 7

In recent years, increasing computational power and resources allow the application of 3D FWT. In general, it is preferably performed in the frequency domain (e. g., Warner et al., 2007; Ben Hadj Ali et al., 2008; Plessix, 2009) but seems to be limited to frequency contents lower than 7 Hz as summarized by Virieux and Operto (2009). Furthermore, the 3D FWT in the frequency domain is still highly demanding due to expensive frequency-domain modelings. There are also 3D applications in the time domain, such as Vigh and Starr (2008) who developed a plane-wave implementation (instead of commonly used shot gathers) to reduce computational efforts. A third possibility is a hybrid method combining efficient time-domain modeling and frequency-domain inversion – the so-called single-frequency method (Sirgue et al., 2008). Furthermore, several authors apply different strategies to reduce the efforts of 3D modelings, such as phase encoding by using super shots (e. g., Vigh et al., 2009; Ben Hadj Ali et al., 2009). Although this is realized at the expense of the accuracy of the resulting model, they additionally apply techniques to mitigate the appearance of artifacts.

In this chapter, I apply pure time-domain 3D inversions to two experiments revealing a cross-well configuration and a marine reflection geometry in case of very complex 3D media. The aim is to demonstrate its feasibility benefiting from efficient time-domain modeling, the reasonable resource consumption and a reliable model reconstruction. I choose both a random medium and a 3D expansion of the Marmousi-II model (Martin, 2002). Apart from strategies discussed in chapter 3, no further optimization techniques are used. The application of 3D FWT concludes this work and represents an outlook for future applications.

Chapter 2

Methodology of acoustic full waveform tomography

This chapter explains the fundamental theory of acoustic full waveform tomography (FWT) mainly based on the work of Tarantola (1984) and Mora (1987). Section 2.1 illustrates the basic FWT scheme, while sections 2.2 to 2.6 give a brief overview of the FWT steps including methods proposed by other authors or developed within the scope of this work. The corresponding appendices A, B and C elaborately provide detailed descriptions.

2.1 The general scheme of full waveform tomography

Basically, full waveform tomography is an iterative inversion method. Its aim is to find an optimal parameter model, e.g. a subsurface model of seismic velocities, which explains the observed seismic data, i.e., the difference of observed and synthetic data has to be minimized. This iterative optimization problem comprises general steps illustrated by Figure 2.1. While the following sections describe the FWT steps, details of the inversion over multiple stages can be found in chapter 3. Apart from the pure time-domain FWT, I additionally discuss the single-frequency method (a slight modification of the method described by Sirgue et al., 2008, see comparison of inversion methods in chapter 5) representing a combination of time-domain and frequency-domain FWT. It differs from a pure time-domain FWT with respect to a few steps of the inversion scheme:

- on-the-fly transformation of the source wavefields and residual wavefields to the frequency domain by application of a discrete Fourier transform for a predefined set of N_f frequencies,

- only N_f frequency-domain wavefield snapshots have to be stored at forward-propagation,

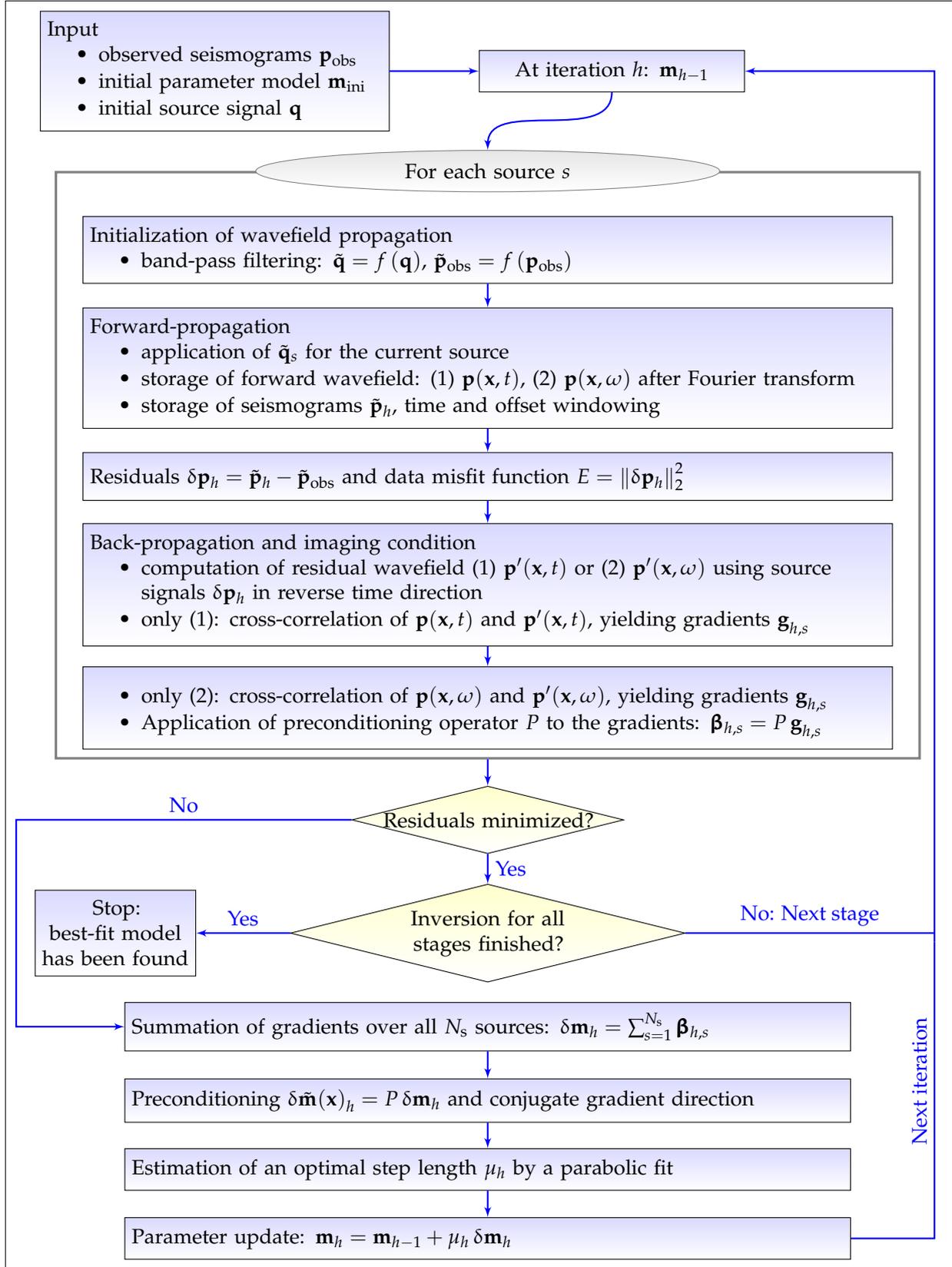- the imaging condition is performed in the frequency domain.

Figure 2.1: FWT scheme of (1) a pure time-domain FWT (Tarantola, 1984) and (2) the single-frequency method (similar to Sirgue et al., 2008). Steps being identical for both methods are not highlighted. The inversion stages refer to different frequency selections and data windows.

2.2   Input of full waveform tomography

The input of full waveform tomography mainly consists of three parts: the observed seismic data, the initial source signal and P-wave velocity model.

Observed data

Within the scope of this work all investigations are restricted to the usage of synthetic data, which are computed by the modeling implementation of the full waveform tomography code. For all applications synthetic "recorded" pressure data is computed by the forward modeling implementation of the FWT codes using the second derivative of the Gaussian distribution (Ricker wavelet; Ricker, 1953) as a source-time function.

Initial source signal

The FWT applications of this work concentrate on the inversion for the model parameter P-wave velocity (v_P), where the true source signal – used to compute the observed data – is assumed as initial wavelet. However, in practice, there are different possibilities to assume an initial source wavelet (which have to be considered in field data applications of the FWT implementation developed in this work):

- estimation from direct wave in the recorded data (e. g., Kravis, 1985),

- estimation from stacked seafloor reflections of the field data at near offsets (e. g., Vigh and Starr, 2008),

- assumption of a synthetic source-time function and initial application of a source-signal inversion using a least-squares method (Pratt, 1999, but with modifications for time-domain applications),

- frequency-domain averaging and phase deconvolution (Hicks and Pratt, 2001),

- employing a separate source signal measurement quite close to the source location.

Initial model

In general, the choice of an appropriate starting model is quite difficult. The data misfit function can be used to define the quality of this model. An intuitive assumption is that a "good" starting model, located quite close to the true model, results in a low misfit, and vice versa. However, especially the choice of the acquisition geometry can intensify unwanted physical effects like cycle skipping (e. g., Ravaut et al., 2004). Consequently, the shape of the misfit function might be more complicated. There are methods exploiting the misfit information to improve the initial model for FWT, such as evolution strategies (Köhn, 2011), or to evaluate the quality of initial models (see the application chapter 4). However, in real data applications, at least a rudimentary initial model has to be provided

by exploiting a priori information or applying conventional imaging methods, such as first-arrival travel-time tomography.

Depending on the purpose and the acquisition geometry used in the synthetic applications, I choose different kinds of initial models. In general, cross-well experiments use homogeneous initial models, while reflection experiments are performed on the basis of smooth or 1D-gradient models. On the one hand, smooth initial models are obtained by applying low-pass filters to the true model, i. e., applying the Gaussian function

$$G(x,y) \sim e^{-\frac{x^2+y^2}{2\sigma^2}} \quad \text{(for a 2D filter)} \tag{2.1}$$

with the spatial coordinates x and y as well as the standard deviation σ. They represent a quite realistic initial model, which might be computed by ray-based methods, such as travel-time tomography. On the other hand, linear-gradient models are considered to be poor initial models due to the usage of very simple assumptions.

2.3 Seismic modeling and imaging condition

2.3.1 Forward modeling

At forward-modeling the source wavefield propagates across the subsurface medium, i. e., the initial model at the first iteration. It is computed by a time-domain finite difference time-stepping method. The synthetic data is recorded by the receivers. For each shot of the acquisition geometry the forward-propagation has to be performed separately. At all or selected time steps (discussed in chapter 3) the spatial distribution of the synthetic forward wavefield is saved in memory for subsequent usage.

In FWT most of the computational efforts are required for seismic modeling due to its separate application for every source at each iteration. In dependence of initial model and acquisition geometry, it is useful or necessary to reduce these efforts and to improve the progress of the FWT by mitigating the ambiguity of the inverse problem, i. e., the nonlinearity of the misfit function. Before forward modeling different techniques can be applied, such as:

- **Time windowing**: The reduction of the information content of the data may also decrease the ambiguity of FWT. On the one hand, it is directly applied to the observed pressure data and on the other hand, the propagation time of modeling is reduced from $0 \leq t \leq T_{\text{total}}$ to $0 \leq t \leq T_{\text{window}}$.

- **Offset windowing**: Only desired offset ranges of observed and forward data are taken into account by this alternative method of data reduction.

- **Band-pass filtering**: Especially low-pass filters are applied to both observed data and initial source signal. There is no positive effect on the efforts of forward-propagation. On the contrary, due to broadening the wavelet, additional propagation time is necessary to avoid acausal signals. However, in comparison with time windowing. this method improves the nature of the misfit function more efficiently. According to Bunks et al. (1995) and Sirgue and Pratt (2004), it is suitable to start the FWT with low frequencies and include higher frequency contents at later inversion stages.

After these preparations seismic modeling is performed, i. e., at each source location the source signal $\mathbf{q}$ is applied and synthetic pressure data is acquired at receiver locations. Within the scope of this work, the acoustic or viscoacoustic wave equation is solved by a finite difference (FD) time-stepping method, i. e., the pressure wavefield $\mathbf{p}(\mathbf{x}, t)$ is computed (see sections 2.3.1.1, 2.3.1.2 and 2.3.1.3).

2.3.1.1 Acoustic wave equation with initial and boundary conditions

The acoustic wave equation represents the simplest case of seismic wave equations. The basis is a medium with only one elastic property, the bulk modulus

$$\kappa = \rho \, v_{\mathrm{P}}^2, \tag{2.2}$$

where ρ is density and v_{P} is the P-wave velocity. In general the homogeneous acoustic wave equation is a second-order partial differential equation:

$$\frac{1}{\kappa(\mathbf{x})} \ddot{p}(\mathbf{x}, t) = \nabla \cdot \left(\frac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t) \right), \tag{2.3}$$

where $p(\mathbf{x}, t)$ is the pressure field. Additionally, initial and boundary conditions have to be defined to find an accurate solution of the given problem. For all spatial locations $\mathbf{x}$ the initial conditions are

$$p(\mathbf{x}, t = 0) = \dot{p}(\mathbf{x}, t = 0) = \ddot{p}(\mathbf{x}, t = 0) = 0. \tag{2.4}$$

The boundary conditions can be divided into two types. In dependence of the problem a free surface is applied, i.e., in the acoustic case an air layer is placed on top of the model. The second type is the absorbing boundary. Due to the finite model size, forward modeling produces unwanted artificial reflections at the model edges. The computations of FWT include synthetic data from forward modeling and observed data which is free from artificial events. To avoid undesirable results it is essential to model as accurate as possible. Thus, this can be realized by perfectly matched layers (PML, see Grote and Sim, 2009), which are implemented in the wave equation. Therefore, an additional layer with attenuating properties is added to the model and the extended 3D wave equation can be written as

$$
\begin{aligned}
\ddot{p} + \left(\sigma_x + \sigma_y + \sigma_z \right) \dot{p} + \left(\sigma_y \sigma_z + \sigma_x \sigma_y + \sigma_x \sigma_z \right) p &= v_{\mathrm{P}}^2 \left[\Delta p + \nabla \cdot \mathbf{w} - \frac{1}{\rho} \nabla \rho \cdot \left(\nabla p + \mathbf{w} \right) \right] \\
&\quad - \sigma_x \, \sigma_y \, \sigma_z \, u, \\
\dot{w}_x &= \left(\sigma_y + \sigma_z - \sigma_x \right) \partial_x p + \sigma_y \sigma_z \partial_x u - \sigma_x w_x, \\
\dot{w}_y &= \left(\sigma_x + \sigma_z - \sigma_y \right) \partial_y p + \sigma_x \sigma_z \partial_y u - \sigma_y w_y, \\
\dot{w}_z &= \left(\sigma_x + \sigma_y - \sigma_z \right) \partial_z p + \sigma_x \sigma_y \partial_z u - \sigma_z w_z, \\
\dot{u} &= p, \tag{2.5}
\end{aligned}
$$

where $\mathbf{w}$ as well as u are auxiliary variables and σ_x, σ_y and σ_z are the attenuation coefficients of the PML. The subscripts x, y and z denote the spatial components in lateral and vertical directions. The definition of the PML coefficients is as follows:

$$
\begin{aligned}
\sigma_x = \sigma_y = \sigma_z = 0 &\qquad \text{within the interior of the model,} \\
\sigma_x > 0, \ \sigma_y > 0, \ \sigma_z > 0 &\qquad \text{within the PML boundary.} \tag{2.6}
\end{aligned}
$$

The complete derivation of the full acoustic wave equation with boundary condition and simplifications, such as the homogeneous-density case or the 2D wave equation, can be found in appendix A.1.

2.3.1.2 Viscoacoustic wave equation with initial and boundary conditions

Within the scope of this work the investigations of attenuation in acoustic FWT are limited to 2D cases. The viscoacoustic wave equation bases on the first-order pressure-velocity formulation of the acoustic wave equation:

$$\dot{p}(\mathbf{x}, t) = \kappa(\mathbf{x}) \ \nabla \cdot \mathbf{w}(\mathbf{x}, t),$$
$$\dot{\mathbf{w}}(\mathbf{x}, t) = \frac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t) \qquad\qquad (2.7)$$

with the particle velocities $\mathbf{w}$. The incorporation of attenuation (based on the generalized standard linear solid with L relaxation mechanisms) results in the following system of partial differential equations:

$$\dot{p}(\mathbf{x}, t) = \kappa_\mathrm{r}(\mathbf{x}) \ \nabla \cdot \mathbf{w}(\mathbf{x}, t) \ [1 + L \, \tau_\mathrm{P}(\mathbf{x})] + \sum_{l=1}^{L} r_l(\mathbf{x}, t),$$
$$\dot{r}_l(\mathbf{x}, t) = -\frac{1}{\tau_{p,l}} \left[\kappa_\mathrm{r} \, \tau_\mathrm{P}(\mathbf{x}) \ \nabla \cdot \mathbf{w}(\mathbf{x}, t) + r_l(\mathbf{x}, t) \right] \quad \text{with} \quad l = \{1, \dots, L\},$$
$$\dot{\mathbf{w}}(\mathbf{x}, t) = \frac{1}{\rho(\mathbf{x})} \nabla p(\mathbf{x}, t). \qquad\qquad (2.8)$$

with the so-called memory variables r_l (which represent the relaxation mechanisms), the relaxed bulk modulus κ_r, the relaxation times $\tau_{p,l}$ and the parameter τ_P which is related to the quality factor Q. The initial conditions of (2.8) are

$$p(\mathbf{x}, t = 0) = \dot{p}(\mathbf{x}, t = 0) = 0,$$
$$\mathbf{w}(\mathbf{x}, t = 0) = \dot{\mathbf{w}}(\mathbf{x}, t = 0) = 0. \qquad\qquad (2.9)$$

To suppress artificial reflections at model boundaries the system of 2D viscoacoustic wave equations includes a PML boundary condition:

$$\dot{p} = \kappa_\mathrm{r} (1 + L \, \tau_\mathrm{P}) (\nabla \cdot \mathbf{w} + u_x + u_y) + (1 + \phi_x + \phi_y + \varphi) \sum_{l=1}^{L} r_l - (\sigma_x + \sigma_y + \theta) \, p, \quad (2.10\mathrm{a})$$

$$\dot{r}_l = -\frac{1}{\tau_{p,l}} \left[\kappa_\mathrm{r} \tau_\mathrm{P} (\nabla \cdot \mathbf{w} + u_x + u_y) + (1 + \phi_x + \phi_y + \varphi) \, r_l \right] - (\sigma_x + \sigma_y + \theta) \, r_l, \quad (2.10\mathrm{b})$$

$$\dot{\mathbf{w}} = \frac{1}{\rho} \nabla p - \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix} \mathbf{w}, \qquad\qquad (2.10\mathrm{c})$$

$$\dot{u}_x = \sigma_y \, \partial_\mathrm{x} w_x, \qquad\qquad (2.10\mathrm{d})$$

$$\dot{u}_y = \sigma_x \, \partial_\mathrm{y} w_y, \qquad\qquad (2.10\mathrm{e})$$

$$\dot{\theta} = \sigma_x \sigma_y, \qquad\qquad (2.10\mathrm{f})$$

$$\dot{\phi}_x = \sigma_x, \qquad\qquad (2.10\mathrm{g})$$

$$\dot{\phi}_y = \sigma_y, \qquad\qquad (2.10\mathrm{h})$$

$$\dot{\varphi} = \psi, \tag{2.10i}$$

$$\dot{\psi} = \sigma_x \sigma_y \tag{2.10j}$$

with $l = \{1, \ldots, L\}$, additional auxiliary variables u_x, u_y, θ, ϕ_x, ϕ_y, φ and ψ as well as PML coefficients σ_x and σ_y (defined by (2.6)). The derivation of (2.10) can be found in appendices B.1 and B.2.

2.3.1.3 Finite difference solution of the wave equation

Homogeneous wave equation

The systems of partial differential equations (2.5) and (2.10) are solved by utilization of a finite-difference (FD) time-stepping scheme (e. g., Alford et al., 1974). Therefore, a time and space discretization has to be applied to the wave equations. In case of an equidistant spatial grid with grid spacing Δh, the coordinates x, y and z are replaced by the subscripts i, j and k. The resulting model dimensions are

$$x = i\,\Delta h, \quad y = j\,\Delta h \quad \text{and} \quad z = k\,\Delta h. \tag{2.11a}$$

The parameters N_x, N_y and N_z denote the the total number of grid points. The time discretization is realized by

$$t = n\,\Delta t, \tag{2.11b}$$

where n represents the time step and the constant Δt is the difference of two successive time steps. The total number of time steps is denoted by N_t.

In case of the second-order acoustic wave equation (2.5) the pressure, its second-order spatial and temporal derivatives are computed at grid location (k, j, i) and time step n. Due to the existence of first-order auxiliary equations, a staggered grid in space and time has to be used within the PML boundary. The location of the update of the auxiliary variables w_x, w_y and w_z is shifted by a half grid point and a half time step, i.e., they are computed at $(k + \frac{1}{2}, j + \frac{1}{2}, i + \frac{1}{2}, n + \frac{1}{2})$.

Furthermore, the viscoacoustic wave equation generally requires a staggered grid formulation. Depending on given wave equations and boundary conditions, the partial derivatives have to be approximated by discrete FD operators of first or second order in space and time. The FD schemes of acoustic and viscoacoustic partial differential equations (2.5) and (2.10) can be found in appendices A.2 and B.3, respectively.

Source implementation

The source signal q is applied at each time step n and pre-defined locations (x_s, y_s, z_s) with $s = 1...N$ ($N =$ number of sources). The continuous signal input is implemented as follows:

$$p_{z_s, y_s, x_s}^n := p_{z_s, y_s, x_s}^n + q_{z_s, y_s, x_s}^n. \tag{2.12}$$

Grid dispersion and instability

The approximation of the wave equation by finite-difference discretization in space and time may result in numerical artifacts and instabilities. On the one hand, FD modeling with a too coarse grid (i.e., a too high Δh) causes a dispersive artificial wavefield. This phenomenon is called grid dispersion. This effect can be avoided by considering a minimum number m of grid points per minimum wavelength $\lambda_{\min}$ (e. g. between 4 and 16 grid points per wavelength for a second-order spatial FD operator). Thus, the following criterion has to be fulfilled (a detailed derivation can be found in Köhn, 2011):

$$\Delta h \leq \frac{\lambda_{\min}}{m} = \frac{v_{\mathrm{P,min}}}{m\, f_{\max}} \approx \frac{v_{\mathrm{P,min}}}{m\, 2 f_{\mathrm{peak}}}, \tag{2.13}$$

where $v_{\mathrm{P,min}}$ is the minimum P-wave velocity as well as $f_{\max}$ and f_{peak} are the maximum and dominant frequencies of the wavefield, respectively. On the other hand, the usage of too big time steps Δt causes numerical instabilities. On the analogy of spatial discretization the time discretization with a second-order FD operator has to fulfill the so-called Courant stability condition (Courant et al., 1928, 1967):

$$\Delta t \leq \frac{\Delta h}{\sqrt{2}\, v_{\mathrm{P,max}}} \quad \text{(in 2D)} \quad \text{and} \quad \Delta t \leq \frac{\Delta h}{\sqrt{3}\, v_{\mathrm{P,max}}} \quad \text{(in 3D)}, \tag{2.14}$$

i.e., within one time step a maximum wavefield propagation of one grid point is allowed.

2.3.2   Residuals

As described by Tarantola (1984) or Mora (1987) (among others; see appendix C.1), the imaging condition of FWT requires the determination of the residual wavefield. Hence, the residuals of synthetic forward and observed data have to be computed for all source wavefields at all receiver locations. However, this depends on the choice of data misfit function which has to be minimized by the inversion scheme. Within this work the misfit function is related to the least squares norm (L_2). On the one hand, its property of incorporating amplitude and phase misfit may impair the robustness of the FWT. But, on the other hand, the main advantage is its simple implementation (Tarantola, 1984).

The derivation of the FWT algorithm is based on the L_2 misfit function (see Tarantola, 1984; Crase et al., 1990, cp. appendix C.2):

$$E\left[\mathbf{p}_h\left(\mathbf{m}_h\right), \mathbf{m}_h\right] = \|(\Delta \mathbf{p}_h, \Delta \mathbf{m}_h)\|_2^2 = \|\Delta \mathbf{p}_h\|_2^2 + \|\Delta \mathbf{m}_h\|_2^2, \tag{2.15}$$

with the pressure data $\mathbf{p}_h$ and the model parameter $\mathbf{m}_h$ at iteration h as well as the general definition of the residuals:

$$\Delta \mathbf{p}_h = \mathbf{p}_h\left(\mathbf{m}_h\right) - \mathbf{p}_{\mathrm{obs}},$$

$$\Delta \mathbf{m}_h = \mathbf{m}_h - \mathbf{m}_{\mathrm{apr}}.$$

Here, $\mathbf{p}_{\mathrm{obs}}$ and $\mathbf{m}_{\mathrm{apr}}$ denote the observed data and the a priori model. In this work the neglection of $\mathbf{m}_{\mathrm{apr}}$ simplifies the least-squares misfit definition. For an exemplary source the

objective function E is given by:

$$E\left[\mathbf{p}_h\left(\mathbf{m}_h\right)\right] = \|\Delta\mathbf{p}_h\|_2^2 = \frac{1}{2}\sum_{n=1}^{N_t}\left(\Delta p_h^n\right)^2 . \qquad (2.16)$$

The residuals are simply defined by Δp_h^n, which represent the so-called "missing diffracted wavefield", i. e. the data residuals computed from the difference of observed and synthetic data represent the mismatch which is not explained by the initial or current model.

2.3.3   Back-propagation and imaging condition

For back-propagation the original receivers act as new sources and vice versa. In analogy to forward modeling in section 2.3.1, the residual data Δp_h^n is simultaneously applied at all original receiver locations and back-propagated (in reverse time direction) to the sources. However, due to the symmetry of the wave equation and its FD solution with respect to time, this method is a common forward modeling.

Using the time step selection from forward-propagation (cp. relation (3.1)) the FWT imaging condition is applied to forward and residual wavefields at corresponding time steps. This yields following model corrections for bulk modulus $\delta\hat{\kappa}(\mathbf{x})$, density $\delta\hat{\rho}(\mathbf{x})$ and source wavelet $\delta\hat{q}(\mathbf{x})$ (Tarantola, 1984) which represent the steepest ascent gradients of the FWT optimization problem:

$$\delta\hat{\kappa}(\mathbf{x}) = \frac{1}{\kappa^2(\mathbf{x})}\sum_{N_s}\int_t dt\, \dot{p}(\mathbf{x}_s, \mathbf{x}_r, t)\ \dot{p}'(\mathbf{x}_s, \mathbf{x}_r, t)\,, \qquad (2.17a)$$

$$\delta\hat{\rho}(\mathbf{x}) = \frac{1}{\rho^2(\mathbf{x})}\sum_{N_s}\int_t dt\, \nabla p(\mathbf{x}_s, \mathbf{x}_r, t)\cdot\nabla p'(\mathbf{x}_s, \mathbf{x}_r, t)\,, \qquad (2.17b)$$

$$\delta\hat{q}(\mathbf{x}) = \sum_{N_s} p'(\mathbf{x}_s, \mathbf{x}_r, t) \qquad (2.17c)$$

with the coordinates of sources $\mathbf{x}_s$ and receivers $\mathbf{x}_r$. The symbol "p'" denotes the back-propagated residual wavefield. Hence, zero-lag cross-correlations of time derivatives and spatial derivatives of the pressure wavefields give the desired model corrections for bulk modulus and density, respectively. The back-propagated wavefield is recorded at original source positions and is treated as correction of the source time function. The derivation of equations (2.17) is described in appendix C.1.

2.4   Gradient computations

This section describes methods to enhance convergence and reduce ambiguity of the inverse problem by preconditioning of the model corrections (2.17a) and (2.17b). In general, the preconditioned gradient $\boldsymbol{\beta}_h$ is obtained by applying a preconditioning operator P_h to the steepest ascent gradient $\mathbf{g}_h$ at iteration h:

$$\boldsymbol{\beta}_h = P_h\,\mathbf{g}_h = P_h\,\delta\hat{\mathbf{m}}_h \quad\text{with respect to}\quad \mathbf{m}(\mathbf{x}) = \begin{pmatrix}\boldsymbol{\kappa}(\mathbf{x})\\ \boldsymbol{\rho}(\mathbf{x})\end{pmatrix}. \qquad (2.18)$$

Within the scope of this work, $P = P(\mathbf{x})$ is a spatial distribution of coefficients and an element-wise application to the gradient is performed. Possible choices are

- **median filter** to remove unrealistic components within the gradient, e.g., artifacts shaped like delta functions,

- **Gaussian averaging operator** to apply a simple smoothness,

- **band-pass filter in the wavenumber-domain** based on the frequencies chosen for filtering of wavelet and data (see 2.3.1),

- **radial taper at source and receiver locations** to suppress corresponding high-amplitude artifacts due to spreading properties of the acoustic wavefield,

- **wavefield-based taper** using the maximum-amplitude distribution of forward- and back-propagated wavefields (Igel et al., 1996; Fichtner et al., 2009),

- **user-defined taper** to suppress model corrections in areas with known parameters or to amplify model updates.

Most of the FWT experiments in chapters 3 to 7 make use of radial and wavefield-based taper geometries (discussed subsequently) as well as user-defined tapers. Apart from median filters, the remaining methods are omitted in all applications. They have been found to be less helpful due to their strong manipulation of the gradient resulting in disadvantageous effects on the computation of the conjugate-gradient direction.

2.4.1 Taper at sources and receivers

Due to the application of source signal in forward- and residuals in back-propagation, the gradient is dominated by high-amplitude artifacts around the acquisition geometry related to geometrical spreading of the acoustic wavefield. In particular the source locations are affected by this problem which prevents a suitable update of the model parameters by exploiting useful gradient information. The artifacts can be mitigated by empirically estimated taper functions $f(r_\mathrm{s}) = f(\mathbf{x} - \mathbf{x}_\mathrm{s})$ and $f(r_\mathrm{r}) = f(\mathbf{x} - \mathbf{x}_\mathrm{r})$ with a circular (2D) or spherical (3D) shape. However, to avoid additional artifacts the taper geometry has to ensure a smooth transition from tapered to the untreated regions.

The implementation of a flexible taper with radius R allows differently steep transitions. The usage of a cosine-based power function with user-defined taper strengths at distances $r = 0$ (taper center at source or receiver location) and $r = R/2$ gives a satisfactory result:

$$ f(r) = \left[\frac{1}{2} - \frac{1}{2} \cos\left(\frac{\pi r}{R} \right) \right]^a [1 - f(r{=}0)] + f(r{=}0) \quad \text{with} \quad r = [-R, R], \qquad (2.19) $$

where

$$ a = -\frac{\ln\left[f\left(r{=}\tfrac{R}{2}\right) - f(r{=}0) \right] - \ln\left[1 - f(r{=}0) \right]}{\ln 2}. $$
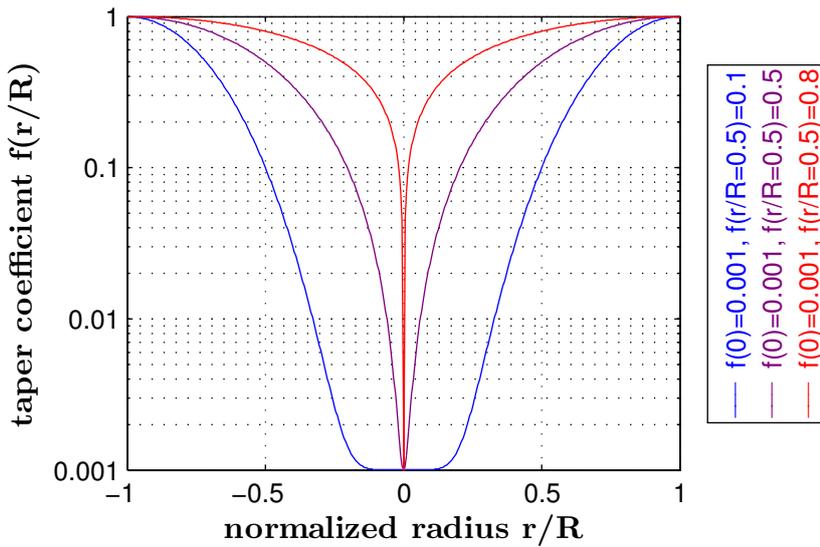
Figure 2.2: Coefficients of a circular or spherical taper with radius R using a cosine-based power function. While the coefficient at the center of the taper is fix, the plots are shown for different coefficients at $r = \frac{R}{2}$.

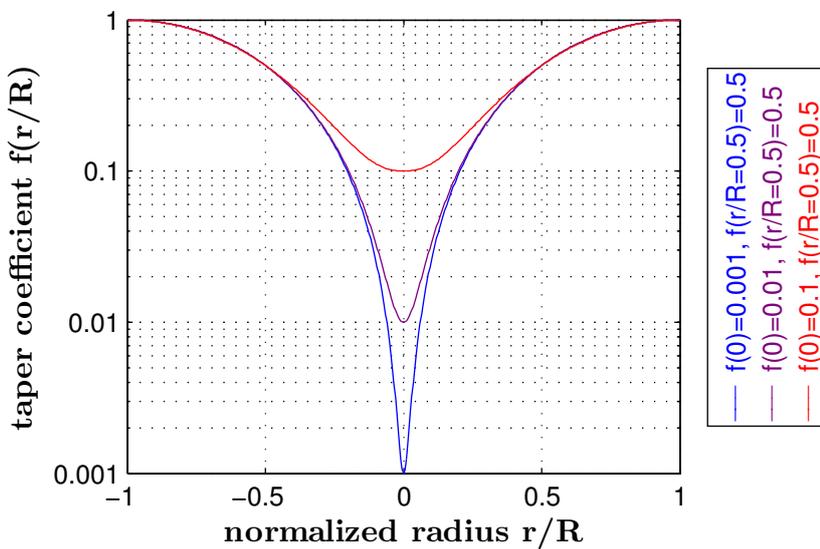**Figure 2.3:** Coefficients of a circular or spherical taper with radius R using a cosine-based power function. While the coefficient at $r = \frac{R}{2}$ is fix, the plots are shown for different coefficients at the center of the taper.
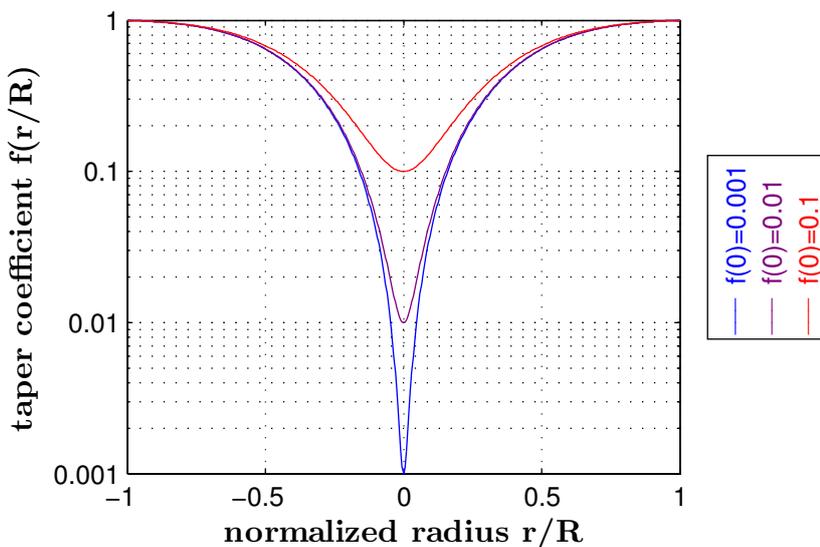

Figure 2.4: Coefficients of a circular or spherical taper with radius R based on a Gaussian function. The plots are shown for different coefficients at the center of the taper.

Equation (2.19) includes the condition $f(r \geq R) = 1$, i.e., the gradient outside the taper area is not altered. While Figure 2.2 shows exemplary taper functions with variable taper strengths $f\left(r = \frac{R}{2}\right)$ and a constant strength $f(r=0)$, Figure 2.3 illustrates variable strengths $f(r=0)$ in conjunction with a constant value $f\left(r = \frac{R}{2}\right)$. Hence, user-defined parameters $f(r=0)$ and $f\left(r = \frac{R}{2}\right)$ can be used to adjust the width and the dynamic range of the taper. An alternative choice is an exponential function which is defined by R and $f(r=0)$ only. It contains a part of the Gaussian function where the decay ratio with respect to the maximum amplitude is larger than $1/e^4$:

$$f(r) = \left[1 - \frac{e^{-4\frac{r^2}{R^2}} - e^{-4}}{1 - e^{-4}} \right] [1 - f(r=0)] + f(r=0) \quad \text{with} \quad r = [-R, R]. \tag{2.20}$$

Due to the condition $f(r=R) = 1$, there is the same behavior outside the taper area. The exponential taper only requires one user-defined parameter $f(r=0)$. The taper geometry is comparable to those of the cosine-based taper with $f\left(r = \frac{R}{2}\right) = 0.5$ (Figure 2.4).

The preconditioning operator can be applied to the gradient $\mathbf{g}_{h+1}$ after summation over all shots or to the shot-specific gradients $\mathbf{g}_{h+1,s}$. In the latter case, the preconditioning is limited to the area around the current source s. The final preconditioning operator is given by

$$P = \begin{cases} f(r_s) & \text{for } |\mathbf{x} - \mathbf{x}_s| \leq R_s \quad \text{(for a single source or all sources)} \\ f(r_r) & \text{for } |\mathbf{x} - \mathbf{x}_r| \leq R_r \quad \text{(for all receivers)} \\ 1 & \text{for } |\mathbf{x} - \mathbf{x}_s| > R_s \text{ and } |\mathbf{x} - \mathbf{x}_r| > R_r \quad \text{(remaining area)} \end{cases} \tag{2.21}$$

with the radii of source tapers (R_s) and receiver tapers (R_r).

2.4.2 Wavefield based preconditioning

This method is intended for the mitigation of strong artificial gradient amplitudes. In particular, they are related to source and receiver locations. In contrast to the previous method, this method does not apply a fix radial taper geometry. Rather, it takes the wavefield propagation into account. Thus, the resulting preconditioning operator is adjusted to the radiation pattern of the wavefield, whose shape depends on inhomogeneity of the model and the location of the acquisition geometry. Especially in reflection seismics, there is no radial pattern. Hence, this preconditioning provides a better performance at the expense of low additional computational costs.

The basic principle is to compute the spatial distribution of maximum absolute wavefield amplitudes, i.e., at every model location the maximum amplitude is computed from the corresponding time series of the pressure wavefield. There are several possibilities to construct the preconditioning operator P. A possible algorithm is given by Igel et al. (1996) and Fichtner et al. (2009). Their preconditioning operator is given by

$$P = \frac{1}{q_f + q_b}. \tag{2.22}$$

The variables q_f and q_b are related to forward-propagated wavefield $p(\mathbf{x}, t)$ and back-propagated wavefield $p'(\mathbf{x}, t)$, respectively:

$$q_f = C_f \frac{\max_t |p| - \min_{\mathbf{x}} (\max_t |p|)}{\max_{\mathbf{x}} (\max_t |p|) - \min_{\mathbf{x}} (\max_t |p|)}, \tag{2.23a}$$

$$q_b = C_b \frac{\max_t |p'| - \min_{\mathbf{x}} (\max_t |p'|)}{\max_{\mathbf{x}} (\max_t |p'|) - \min_{\mathbf{x}} (\max_t |p'|)}. \tag{2.23b}$$

The coefficients C_f and C_b are used to weight amplitudes of the wavefields $p(\mathbf{x}, t)$ and $p'(\mathbf{x}, t)$. On the one hand, they can be chosen manually, e. g., $C_f = 10$ and $C_b = 1$ (Igel et al., 1996; Fichtner et al., 2009). On the other hand, an automatic estimation of C_f and C_b considers changing radiation patterns due to the altered model during the iterative inversion progress:

$$C_f = \frac{\max_{\mathbf{x}} (\max_t |p|) - \min_{\mathbf{x}} (\max_t |p|)}{\max_{\mathbf{x}} (\max_t |p'|) - \min_{\mathbf{x}} (\max_t |p'|)}, \tag{2.24a}$$

$$C_b = 1. \tag{2.24b}$$

Additionally, this work proposes another algorithm to estimate an appropriate preconditioning operator. It is given by

$$P = \frac{b}{\max_{\mathbf{x}} b} \quad \text{with} \quad b = \frac{1}{a + C_{\text{stab}}\, \bar{a}} \quad \text{and} \quad a = \max_t |p| \, \max_t |p'|. \tag{2.25}$$

The auxiliary parameter $\bar{a}$ denotes an average of a. The user-defined coefficient C_{stab} stabilizes the computation of P and thus scales its strength and dynamic range. In case of both algorithms the application of P (2.18) is an element-wise operation at iteration h and grid location (i, j, k):

$$\boldsymbol{\beta}_{h|k,j,i} = P_{k,j,i}\, \mathbf{g}_{h|k,j,i} = P_{k,j,i}\, \delta\hat{\mathbf{m}}_{h|k,j,i}. \tag{2.26}$$

2.5   Model update

The following section summarizes the update of the acoustic parameter model. It considers pure acoustic inversion and acoustic inversion with viscoacoustic modeling. Apart from gradient preconditioning, additional constraints can be directly applied to the model.

2.5.1   Computation of the model update

The preceding step of the model update is the summation of all shot-specific gradients resulting in the global gradient of the entire acquisition geometry. Depending on the preconditioning, two possibilities can be distinguished. On the one hand, preconditioning is applied after the summation of the steepest ascent gradients which is applicable in case of using circular tapers. On the other hand, preconditioning is applied before the summation, which is possible in case of circular tapers but mandatory for wavefield-based preconditioning. An optimized gradient is then computed by the subsequent application of the method of conjugate gradient.

Model update in the acoustic inversion

The general update of the model parameters $\mathbf{m}_h$ at iteration h is derived in appendix C.2 and is given by

$$\mathbf{m}_h = \mathbf{m}_{h-1} - \mu_h \, \mathbf{g}_h = \mathbf{m}_{h-1} - \mu_h \, \delta\hat{\mathbf{m}}_h \tag{2.27}$$

with model corrections $\delta\hat{\mathbf{m}}$ and step length μ of the gradient method. However, instead of using the steepest ascent gradient $\mathbf{g}$, the acoustic FWT scheme employs the conjugate-gradient method. The computation of the conjugate gradient $\mathbf{c}$ requires $\mathbf{g}$ and the preconditioned gradient $\boldsymbol{\beta}$ (see appendix C.2 and section 2.4). The model update can be rewritten in terms of the acoustic parameterization $\mathbf{m} = (\boldsymbol{\kappa}, \boldsymbol{\rho}, \mathbf{q})^T$:

$$\boldsymbol{\kappa}_h = \boldsymbol{\kappa}_{h-1} - \mu_{\boldsymbol{\kappa}|h} \, \mathbf{c}_{\boldsymbol{\kappa}|h}, \tag{2.28a}$$

$$\boldsymbol{\rho}_h = \boldsymbol{\rho}_{h-1} - \mu_{\boldsymbol{\rho}|h} \, \mathbf{c}_{\boldsymbol{\rho}|h}, \tag{2.28b}$$

As an exception the update of the source term is directly computed from the parameter corrections:

$$\mathbf{q}_h = \mathbf{q}_{h-1} - \mu_{\mathbf{q}|h} \, \delta\hat{\mathbf{q}}_h, \tag{2.28c}$$

where $\delta\hat{\mathbf{q}}$ represent the back-propagated residual data recorded at source locations. Apart from the parameterization mentioned above, it is desirable to have P-wave velocity $\mathbf{v}_P$ instead of bulk modulus $\boldsymbol{\kappa}$. The model update in terms of the new parameter $\mathbf{v}_P$ is described by Mora (1987) (see appendix C.2):

$$\mathbf{v}_{P|h} = \mathbf{v}_{P|h-1} - \mu_h \, \text{diag}\left(\mathbf{J}_{\mathbf{v}_P|h}\right) \, \mathbf{c}_{\boldsymbol{\kappa}|h} \tag{2.29}$$

with the definition of the Jacobian

$$J_{v_P} := \frac{\partial \kappa}{\partial v_P} = 2\,\rho\,v_P.$$

Model update in an acoustic inversion with viscoacoustic modeling

The following paragraph summarizes the consideration of attenuation in the inversion scheme. The acoustic inversion with viscoacoustic modeling includes attenuation as a passive modeling parameter. While the parameterization for modeling is $\mathbf{m} = (\boldsymbol{\kappa}_r, \boldsymbol{\rho}, \mathbf{Q}_P, \mathbf{q})^T$, the inversion parameters still are $\mathbf{m} = (\boldsymbol{\kappa}_r, \boldsymbol{\rho}, \mathbf{q})^T$.

At the first iteration, the method requires an initial model κ_0. It is obtained from the user-defined initial acoustic reference model $\mathbf{v}_{P,ref|0}$ by applying the model relaxation (B.6). On the one hand, the relaxed bulk modulus is required for viscoacoustic modeling and on the other hand, it is treated as an acoustic parameter by the inversion algorithm (equation C.39b). However, the desired final result is the acoustic reference velocity model $\mathbf{v}_{P,ref|h}$. It is physically meaningful and is comparable to the initial acoustic velocity model. First, one has to obtain the relaxed velocity model from (C.44). Then, $\mathbf{v}_{P,ref|h}$ is computed by revoking the model relaxation (B.6), which can be rewritten in terms of P-wave velocity as well as parameters ω_0, $\omega_{r,l}$ and $\boldsymbol{\tau}_P$ discussed in section 2.3.1.2:

$$\mathbf{v}_{P,ref|h} = \mathbf{v}_{P|h} \sqrt{1 + \sum_{l=1}^{L} \frac{\omega_0^2/\omega_{r,l}^2}{1 + \omega_0^2/\omega_{r,l}^2} \boldsymbol{\tau}_P}. \tag{2.30}$$

2.5.2 Application of constraints to the model update

Apart from preconditioning, the application of additional constraints is useful to compute physically meaningful models or to stabilize the inversion. The inversion scheme comprises following constraints at iteration h (exemplary for the 2D case):

- The inversion is not allowed to compute models $m_{j,i|h}$ which fall below a lower limit or exceed an upper limit:

$$m_{j,i|\min} \leq m_{j,i|h} \leq m_{j,i|\max}. \tag{2.31}$$

- The initial model $\mathbf{m}_{ini}$ is used as a priori information. The updated model does not exceed the maximum allowable relative distance $d_{h|ini}$ to the initial model:

$$\frac{m_{j,i|h} - m_{j,i|ini}}{m_{j,i|ini}} \leq d_{h|ini}. \tag{2.32}$$

- The updated model does not exceed the maximum allowable relative deviation $d_{h|h-1}$ from the model of previous iteration $h - 1$:

$$\frac{m_{j,i|h} - m_{j,i|h-1}}{m_{j,i|h-1}} \leq d_{h|h-1}. \tag{2.33}$$

- The inversion is forced to produce a 1D model. On the one hand, at each depth j, the average value $\overline{m}_{j,i|h}$ can be assigned to all lateral locations indexed by i. On the other hand, a relative threshold value d_{1D} (with $0 \leq d_{1D} \leq 1$) can be applied to tolerate lateral deviations from the average value, i. e., it represents a transition to a 2D model:

$$
m_{j,i|h} = \begin{cases} \overline{m}_{j|h}\,(1 + d_{1D}) & \frac{m_{j,i|h}}{\overline{m}_{j|h}} > 1 + d_{1D} \\[2ex] m_{j,i|h} & 1 - d_{1D} \leq \frac{m_{j,i|h}}{\overline{m}_{j|h}} \leq 1 + d_{1D} \\[2ex] \overline{m}_{j|h}\,(1 - d_{1D}) & \frac{m_{j,i|h}}{\overline{m}_{j|h}} < 1 - d_{1D} \end{cases} \tag{2.34}
$$

with

$$
\overline{m}_{j|h} = \frac{1}{N_x} \sum_{i=1}^{N_x} m_{j,i|h}.
$$

2.5.3   Choice of parameterization

Within the scope of this thesis, I focus on the reconstruction of P-wave velocity models. This assumption eases to answer the question of an appropriate parameterization. In general, the desired output of imaging techniques are subsurface models containing seismic velocities and density, i. e., the description of the perfect earth requires three (elastic approximation) or two parameters (acoustic approximation). However, depending on the field of application, a certain parameterization is recommended. While a theoretical consideration on this was done by Tarantola (1986a), several authors investigated inversions including different acoustic or elastic parameter couplings. The main conclusion is that the most reliable reconstruction of subsurface models requires a strong parameter decoupling. This applies to the combination of impedance and seismic velocity (proposed by Kolb and Canadas (1986) for a 1D inversion in the acoustic approximation and Assous and Collino (1990) for the inversion of near-offset data in the elastic approximation; verified by Przebindowska et al. for application of the acoustic FWT to multi-offset marine seismic data). Apart from "impedance–velocity", the parameterization with respect to impedances and density is another appropriate choice, which has been found by Assous and Collino (1990) for a layered 2D medium and Köhn et al. (2012) in case of very complex 2D elastic problems explicitly designed for parameterization studies. In contrast, the choice of Lamé parameters is suboptimal. Although the parameterization of the acoustic approximation in this work reveals the first Lamé parameter (see update of the bulk modulus in equation (2.28); Tarantola, 1984), the desired P-wave velocity is directly obtained after a change of parameterization (see equation (2.29); Mora, 1987).

2.6 Step length optimization

In contrast to using the inverse Hessian matrix $\mathbf{H}$, the simplified gradient algorithm (2.28) requires the determination of a step length to scale the gradient.

2.6.1 Methods of step length estimation

At each iteration the estimation of the optimal step length μ_h is performed. It bases on Pica (1990) and is composed of two parts: a user-defined relative factor $\mu_{\mathrm{rel}|h}$ and a factor used to scale the gradient to the maximum range of the model parameter. This allows a meaningful physical unit and a proper distance of the gradient. For example, equation for the model update (C.39b) is rewritten as

$$\mathbf{m}_h = \mathbf{m}_{h-1} - \mu_{\mathrm{rel}|h} \frac{\max |\mathbf{m}_h|}{\max |\mathbf{g}_h|} \mathbf{g}_h. \tag{2.35}$$

The relative factor $\mu_{\mathrm{rel}|h}$ is approximated by a parabolic curve fitting method (compare the line-search method with quadratic or cubic interpolation in Nocedal and Wright, 1999). An initial step length $\mu_{\mathrm{rel,ini}}$ has to be provided at iteration $h = 1$. Two additional values are computed by applying a constant factor $a > 1$: $\mu_{\mathrm{rel,low}} = \dfrac{\mu_{\mathrm{rel,ini}}}{a}$ and $\mu_{\mathrm{rel,low}} = a\,\mu_{\mathrm{rel,ini}}$. Using $(\mu_{\mathrm{rel,ini}}, \mu_{\mathrm{rel,low}}, \mu_{\mathrm{rel,high}})$, three test modelings are computed for a representative subset of sources and a parabola is fitted to the corresponding misfits $[E(\mu_{\mathrm{rel,ini}}), E(\mu_{\mathrm{rel,low}}), E(\mu_{\mathrm{rel,high}})]$. The minimum of the parabola indicates the optimal step length at iteration h, which is also used as initial value at the next iteration. In contrast to simple step length estimations, such as using a constant value, the adaptive estimation causes a significant and stable reduction of the data misfit function (Kurzmann et al., 2008).

2.6.2 Implementation of adaptive step length

In this work, the method of adaptive step length assumes a locally parabolic shape of the data misfit function. However, in dependence of several circumstances, such as choice of initial model, application data and gradient preconditioning as well as the choice of test step lengths, the shape might satisfy or violate this assumption. In general, the data misfit function is highly non-linear. Consequently, the implementation has to consider different exceptions.

Figure 2.5 shows the data misfit function discussed in the context of experiment X in section 4.2. At the first iteration, it is obtained as a function of different step lengths. This misfit function is used to demonstrate the effects of the adaptive step length estimation. The choice of test step lengths $\mu_{\mathrm{rel,ini}}$, $\mu_{\mathrm{rel,low}}$ and $\mu_{\mathrm{rel,high}}$ yields a parabolic fit with optimal step length μ_{opt}. It is based on a scaling factor of $a = 2$. Especially at the first iterations, there might occur extensive local minima at large step lengths. However, in general, they belong to artificially altered models. Regarding Figure 2.5, the aim is to estimate a step length μ within the left local minimum. The right local minimum is considered to be the wrong choice.

Apart from the optimal estimation, some special cases may occur (Figures 2.6 to 2.8). Figure 2.6 shows parabolic fits for a fix initial step length $\mu_{\mathrm{rel,ini}} = 0.05$ and a variable factor a

resulting in different values $\mu_{rel,high}$ and $\mu_{rel,low}$. In all three cases the parabolic curve fitting is able to compute a plausible step length. However, despite using a small $\mu_{rel,ini}$, the application of large scaling factors results in inappropriate μ estimations (Figures 2.6b,c). Instead of using a too high step length μ_{opt}, it is useful to apply an upper limit μ_{limit}. This may avoid the production of artificial model updates.
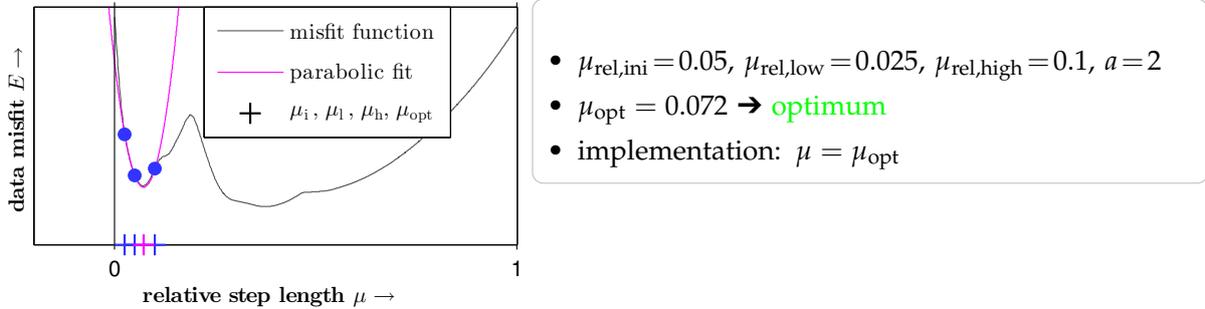

Figure 2.5: Step length estimation: Data misfit at the first FWT iteration as a function of the step length and application of a parabolic curve fitting to a set of test step lengths $\mu_{rel,ini}$, $\mu_{rel,low}$ and $\mu_{rel,high}$ ("+") resulting in an estimation of the optimal step length μ_{opt} ("+"). The circular markers "●" represent the data misfits corresponding to the test step lengths.
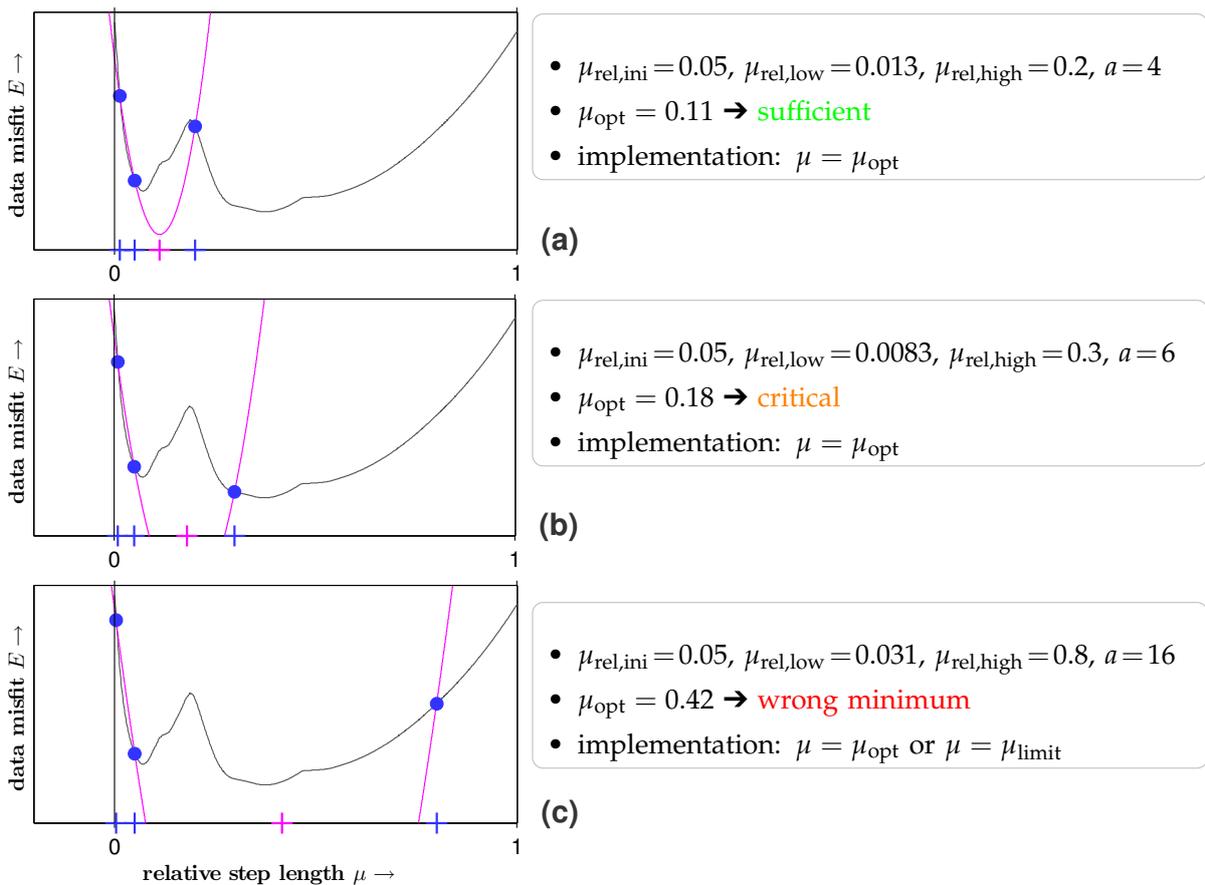

Figure 2.6: Step length estimation: Different cases of a parabolic curve fitting using the same initial step length $\mu_{rel,ini}$ but different test step lengths $\mu_{rel,low}$ and $\mu_{rel,high}$.
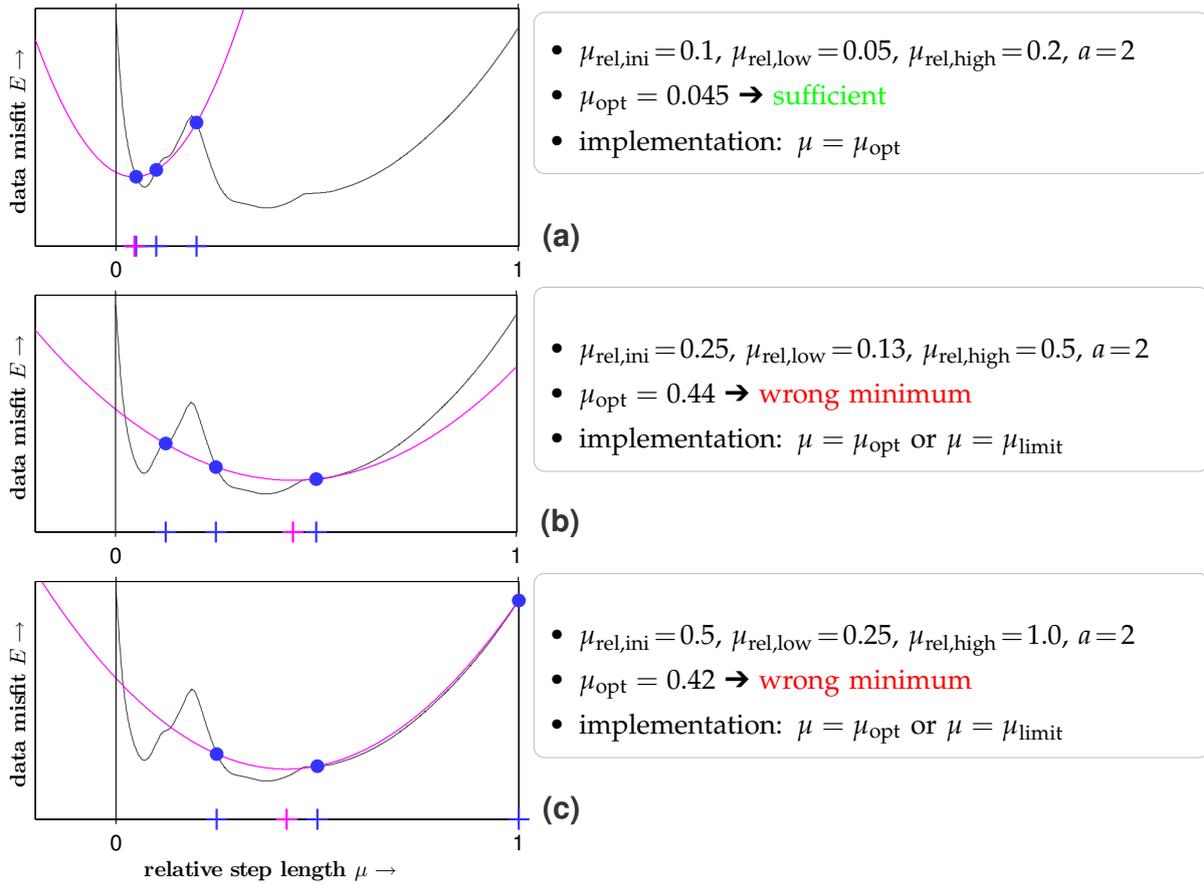
Figure 2.7: Step length estimation: Different cases of a parabolic curve fitting using different initial step lengths $\mu_{\text{rel,ini}}$ in conjunction with the same scaling factor a.

Additionally, while applying a variable $\mu_{\text{rel,ini}}$, a is fix (see Figure 2.7), i. e., the same ratio $a = \frac{\mu_{\text{rel,high}}}{\mu_{\text{rel,ini}}} = \frac{\mu_{\text{rel,ini}}}{\mu_{\text{rel,low}}}$ is used for all examples. Again, only a combination of sufficiently low test step lengths yield a proper parabolic estimation (Figure 2.7a). Interestingly, Figures 2.7b and c demonstrate the computation of comparable improper step lengths by using completely different initial values.

Finally, there are a few exceptions which may have a negative impact on the progress of the FWT. Figure 2.8 summarizes common cases. For example, the quadratic approximation may be reduced to a linear function or just a constant. Consequently the "minimum" is located at infinity or it is undefined, respectively (see Figures 2.8a and b). On account of this, the algorithm is forced to use a very small step length. The calculation of a concave parabola is one of the most problematic cases (Figure 2.8c). Although the test step length with lowest corresponding data misfit represents the most intuitive choice, a small step length (or the minimum step length) would be more useful, as indicated by Figure 2.8c. Furthermore, the estimation of too high or negative μ_{opt} can be handled more robustly (Figures 2.8d,e). On the one hand, the upper limit $\mu = \mu_{\text{limit}}$ is applied. On the other hand, a very small step length can be chosen. Finally, it is possible to compute a reliable step length using improperly high test step lengths (Figure 2.8f), which, however, may happen by accident.
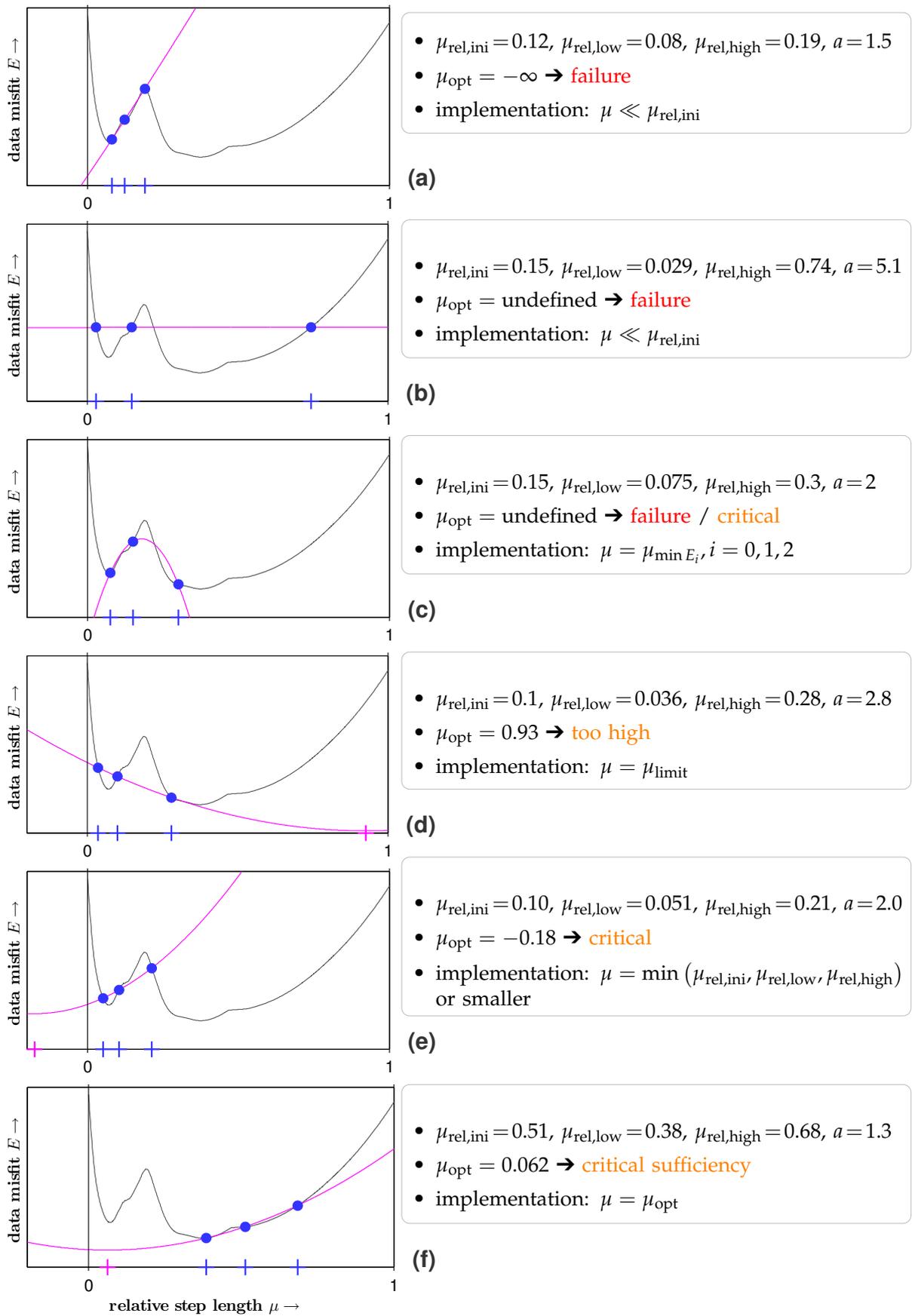
Figure 2.8: Step length estimation: Common exceptional cases where the parabolic curve fitting computes unsatisfactory step lengths.

Chapter 3

Optimizations of the FWT implementation

Most of the computational efforts of time-domain FWT account for seismic modeling. Thus, a meaningful FWT application relies on the availability of numerous computational resources with a sufficiently fast interconnection. The limited access to high-performance supercomputers and the occasional usage of local workstation computers with a nearly unacceptable performance brought up the need of the optimization of the FWT implementation with respect to efficiency. This chapter discusses the optimal combination of parallelization methods – improving the performance of the modeling part – and recombines existing FWT strategies mentioned in section 2.3.1 – directly affecting performance and progress of the inversion part.

3.1 Parallelization

3.1.1 Implementation

The computational efforts of time-domain FWT are very expensive which is caused by the finite-difference modeling. To obtain a tomographic image of the underground, the seismic wavefields have to be computed for a certain number of sources. Each source requires a separate simulation. To improve the performance of the modeling part, I choose two kinds of parallel implementations: the model decomposition (hereinafter referred to as "domain decomposition") and the shot parallelization. Apart from that, Akcelik (2002) comprehensively recapitulates efficient strategies of performing modeling and Virieux and Operto (2009) give a brief overview.

Domain decomposition

In general, 2D modelings do not consume a lot of memory or computation time. However, at every time step the entire wavefield has to be saved in memory. Additionally, low P-wave velocities (e. g., water velocity) and high-frequency data require a small grid spacing Δh. The resulting larger grid size of the model causes a rapid increase of the memory consumption. Hence, it is meaningful to divide the model into several equally sized subdomains, which are distributed among different CPUs (domain decomposition, see Bohlen, 1998). At each
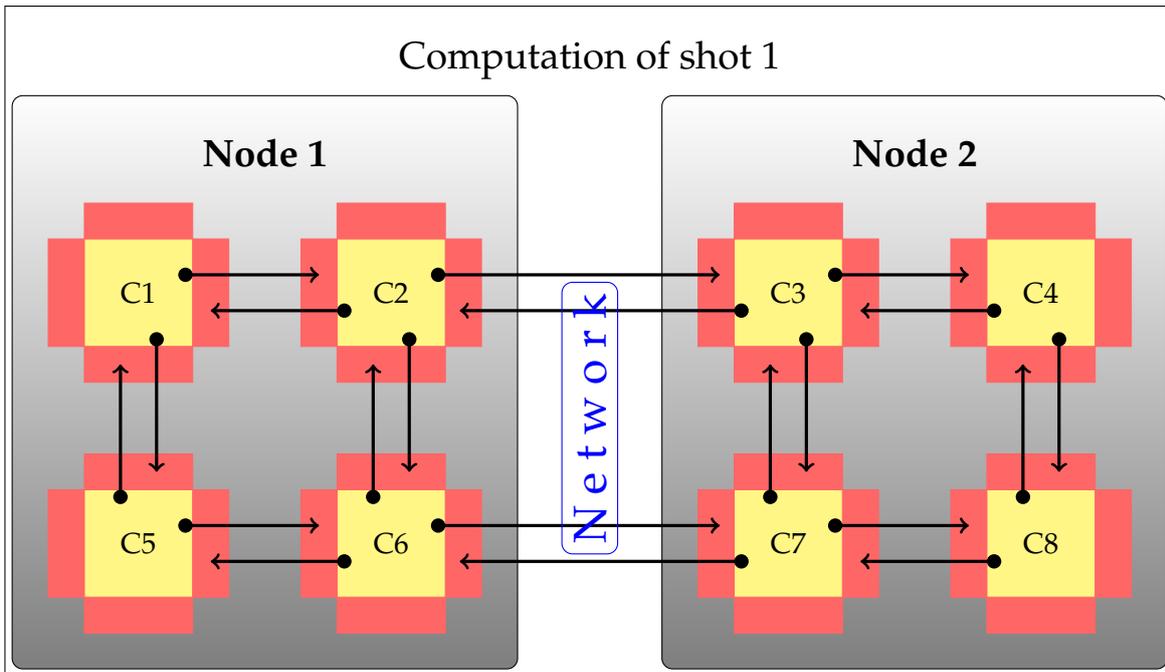
Figure 3.1: Domain decomposition on a cluster computer which consists of 2 nodes with 4 cores per node. Here the model is divided into 8 subdomains (yellow). The corresponding padding layers are colored in red. The exchange of the wavefield (here pressure) requires intra- and inter-node communication (arrows) using *MPI*.
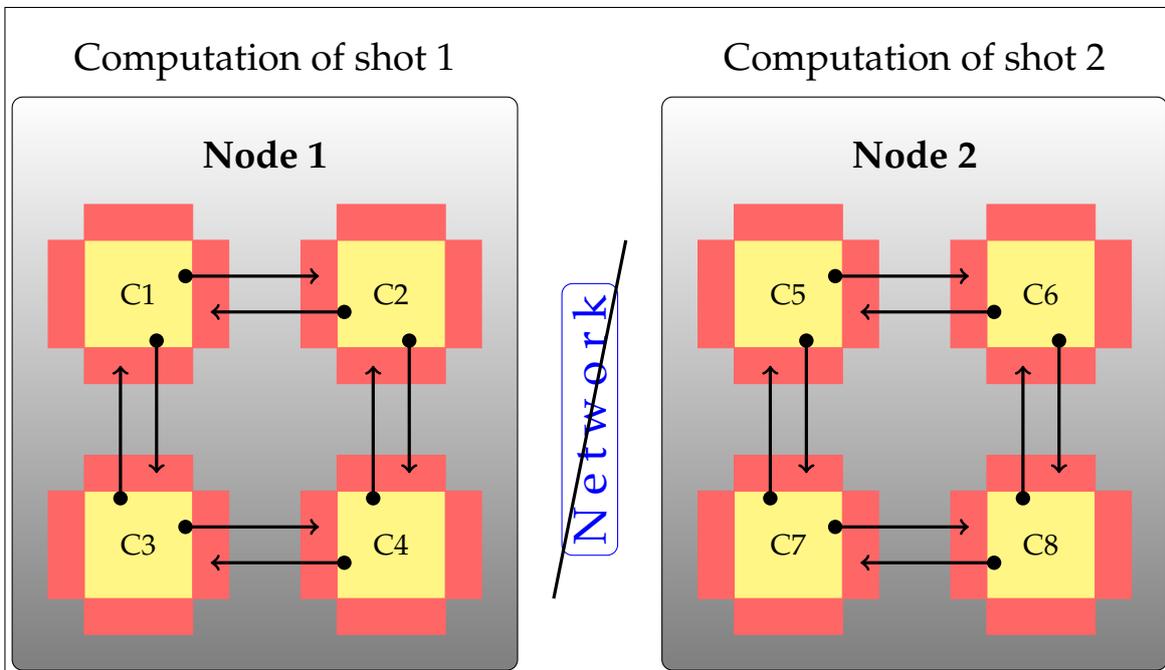

Figure 3.2: Shot parallelization on a cluster computer. The computer architecture is identical to Figure 3.1. However, in contrast to Figure 3.1, the model is divided into 4 subdomains on each node. The exchange of the wavefield at the model boundaries is reduced to intra-node communication using *MPI*.

time step the exchange of wavefield information at the inner boundaries of the subdomains ensures an accurate wavefield propagation. The exchange requires communication among all neighboring processing elements. This is realized by using the *Message Passing Interface* (*MPI*, designed for distributed-memory machines) in FWT implementations developed within the scope of this work. Due to the usage of discrete spatial operators, a padding layer with the half size of the finite-difference operator is added at the model boundaries. In the acoustic approximation the pressure field p has to be exchanged in all directions at all inner boundaries. In contrast, due to first-order derivatives of the auxiliary wavefield $\mathbf{w}$ (used in PML implementations or first-order viscoacoustic wave equation), it is sufficient to apply a one-way exchange of the components of $\mathbf{w}$ from one subdomain to its adjacent subdomains at the top and the left boundary.

The advantage of domain decomposition is the reduction of memory usage and computation time per processing element. However, the unavoidable communication requires a fast network connection. Figure 3.1 illustrates domain decomposition on a cluster computer which consists of two quad-core computers (nodes). Especially in 2D finite-difference simulations with small models the ratio of model size and padding layer size is very inefficient. This results in a very fast computation of the wavefields within every domain but requires a high rate of data exchanges which easily exceeds the capacity of the network and slows down the entire computation.

Shot parallelization

For each source, forward modelings and back-propagations have to be computed separately. Hence, on condition of the availability of computers (nodes) with enough random access memory, it is meaningful to distribute shots among all available nodes (Kurzmann et al., 2009). In case of multi-core computers, domain decomposition can be done internally. As a consequence, a tremendous reduction of network traffic and a speedup of the modeling algorithm can be achieved. By using the cluster computer mentioned above, Figure 3.2 shows the combination of domain decomposition and shot parallelization. The simultaneous computation of two shots reduces the number of cores per forward modeling and omits network communication. In contrast to domain decomposition, the achievement of computational improvements with increasing intensity of shot parallelization is realized at the expense of a huge memory consumption. This represents the only limitation to this type of parallelization. Nevertheless, this method is designed for applications on supercomputers providing a sufficiently high amount of random access memory.

3.1.2 Benchmark of domain decomposition and shot parallelization

In the following, the parallelization methods described by previous paragraphs are compared with each other. In detail, the benchmark compares a pure domain decomposition with a combination of domain decomposition and shot parallelization. The corresponding FWTs were performed on the supercomputers *JUROPA* at the Jülich Supercomputing Centre and *HERMIT* at the High Performance Computing Center Stuttgart. Table 3.1 lists all relevant information about the test configuration, such as modeling parameters as well as the distribution of compute cores with respect to each parallelization type. The benchmark

represents a speedup test, which analyzes the code performance in dependence on the amount of computational resources, i. e., the number of cores varies from 1 to 1024.

The benchmark configuration is oriented towards the architecture of *HERMIT*. The suitability of all configurations is visualized by the traffic-light colors in Table 3.1, where "red" symbolizes an inefficient combination of both parallelization techniques, "green" represents an optimal choice and "yellow" a moderate suitability. In particular, the domain decomposition of the combined parallelization includes 32 subdomains. Since all computations use the same configuration, this involves network communication on *JUROPA*. However, the benchmarks apply the most optimal domain decomposition, which minimizes inter-node communication. Finally, pure domain decomposition is applied to all available cores, while the combined parallelization is limited to core numbers varying from 32 to 1024, while the total grid size remains constant.

The evaluation of the performance is based on computing times expressed in terms of "time per modeling" or "time per iteration". For each computation, ten iterations were performed to obtain reliable values. The benchmark results shown in Figure 3.3 base on the median of the computing times. They clearly point out the communication problem with respect to the pure domain decomposition (see Figures 3.3a,c). In particular, the computations on *HERMIT* can be divided into three stages:

1. **Usage of 1 to 16 cores**: Although the entire computation is limited to one compute node, the doubling of core numbers yields a fluctuating speedup within the range from 1.2 to 3.9 (where "speedup" ≥ 2 is considered to be perfect and "speedup" $= 1$ represents no speedup). Apparently, this is caused by an inefficient domain decomposition with respect to the arrangement of cores on the CPU sockets and corresponding inter-core communication within a node.

2. **Usage of 32 to 256 cores**: Both domain decomposition and shot parallelization reveal a satisfactory performance. While the speedup of domain decomposition decreases from 2.4 to 1.6, the shot parallelization shows a very robust speedup (decreasing from 2.4 to 1.9). Obviously, the pure domain decomposition is characterized by a good trade-off between model size of the subdomains and the amount of data, which has to be exchanged via inter-node communication.

3. **Usage of 512 and 1024 cores**: While the comparison of 256 and 512 cores exhibits a poor speedup of domain decomposition (1.2), there is nearly no further improvement in case of using 1024 cores (speedup of 1.1). The amount of exchanged data bears no relation to the model size of the subdomains. In contrast, the shot parallelization is able to preserve the high speedups (1.8 and 1.7). In case of using 1024 cores, the exploitation of shot parallelization reduces the computing time by 71 % in comparison to only considering domain decomposition.

Apart from the computing times per modeling and per iteration, the performance can be expressed in terms of modelings per time or FWT iterations per time. The corresponding Figures 3.3b,d provide a more demonstrative illustration of performance differences. While the pure domain decomposition computes up to 1.4 iterations per minute, the shot parallelization is able to handle 4.9 iterations per minute.

Table 3.1: Configuration of parallelization benchmarks. The 2D FWTs were performed on the supercomputers *JUROPA* and *HERMIT*. The suitability of a certain configuration with respect to the architecture is colored in green ("good"), yellow ("moderate") and red ("bad"). The number of cores for domain decomposition is given by $N_{core|x}$ and $N_{core|y}$. $N_{sim\ shots}$ denotes the number of simultaneous shots in shot parallelization.

General parameters

Implementation	2D acoustic FWT, PML boundary, model update with constant step length
Model size	1024×1024 grid points
Time steps	10000
Sources	32, i.e., 64 modelings per FWT iteration
Snapshots	2000, i.e., wavefields of every fifth time step are involved in cross-correlation
Hardware	*JUROPA*: 8 cores per compute node (2 CPU sockets with 4 cores each), clock speed = 2.93 GHz; 24 GiB memory per compute node
	HERMIT: 32 cores per node (2 CPU sockets with 16 cores each), clock speed = 2.3 GHz; 32 GiB memory per compute node

Parallelization

Pure domain decomposition

N_{core}	$N_{core|x} \times N_{core|y}$	$N_{sim\ shots}$	*JUROPA*	*HERMIT*
1	1×1		🟢	🟢
2	2×1		🟢	🟢
4	2×2		🟢	🟢
8	4×2		🟢	🟢
16	4×4	1	🟡	🟢
32	8×4		🟡	🟢
64	8×8		🟡	🟡
128	16×8		🟡	🟡
256	16×16		🟡	🟡
512	32×16		🔴	🔴
1024	32×32		🔴	🔴

Domain decomposition + shot parallelization

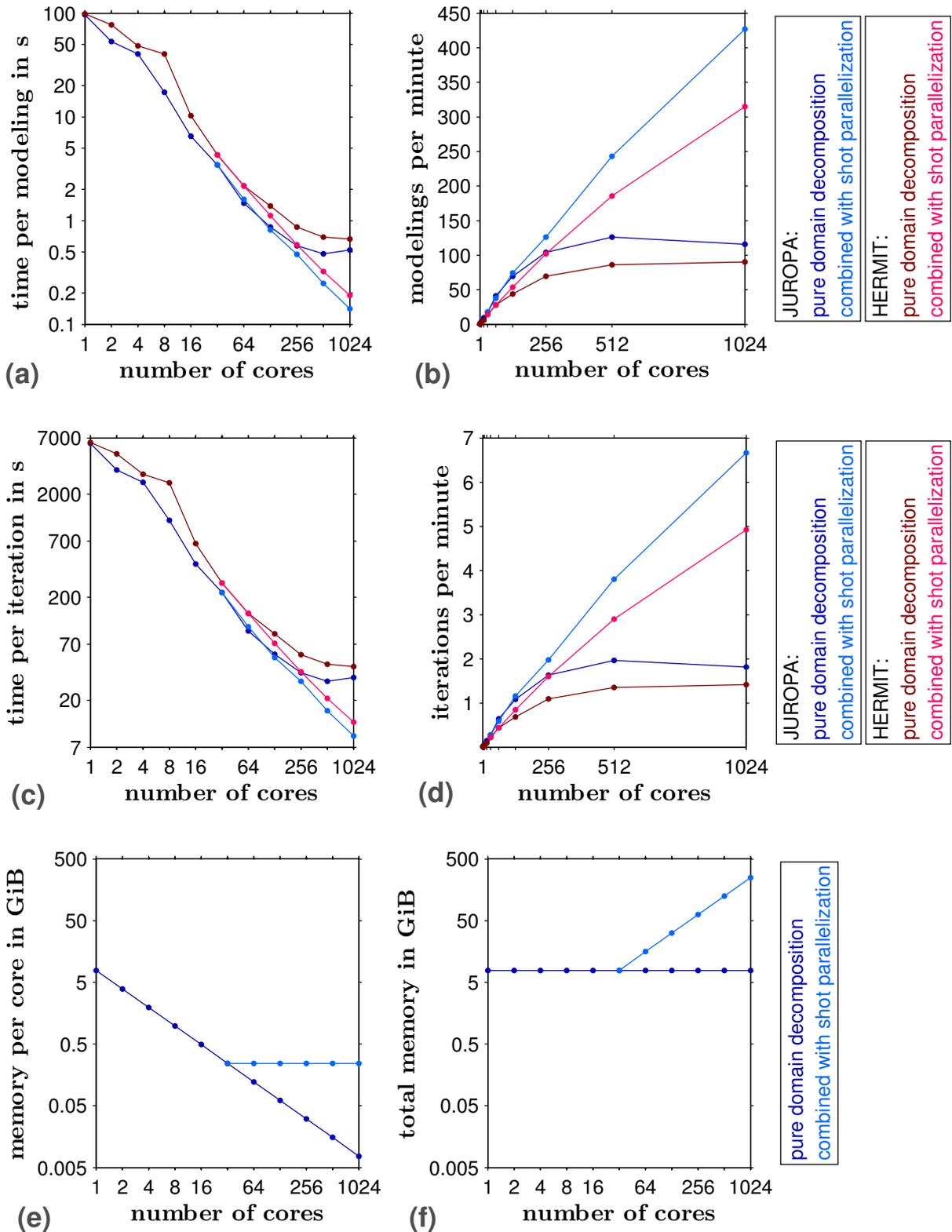| N_{core}	$N_{core|x} \times N_{core|y}$	$N_{sim\ shots}$	*JUROPA*	*HERMIT*
32	8×4	1	🟡	🟢
64		2	🟡	🟢
128		4	🟡	🟢
256		8	🟡	🟢
512		16	🟡	🟢
1024		32	🟡	🟢

Figure 3.3: Comparison of parallel implementations inside a 2D FWT tested on the supercomputers *JUROPA* and *HERMIT*. The performance of pure domain decomposition and a combination of domain decomposition with shot parallelization is shown in (a) and (b) with respect to modelings as well as in (c) and (d) with respect to FWT iterations. The memory consumption is illustrated by (e) and (f).

However, the increase of performance by exploiting the shot parallelization is done at the expense of the memory consumption (see Figures 3.3e,f). Here, the memory consumption is calculated from the product of the model size (of a subdomain or of the entire model), number of wavefield snapshots required in cross-correlation and the number of simultaneously computed sources. On the one hand, in case of domain decomposition the memory consumption per compute core continually decreases with increasing number of cores. The total memory amounts to a constant value. On the other hand, in case of shot parallelization the memory consumption per core is constant but scales with increasing number of simultaneous sources. For this benchmark, the total memory consumption rises up to 250 GiB.

Furthermore, there is a good qualitative match between the benchmark results for *JUROPA* and *HERMIT*. Although *JUROPA* shows a better absolute performance, the application of a corrective factor with respect to the ratio of clock speeds, $\frac{2.93 \text{ GHz}}{2.3 \text{ GHz}} = 1.27$, gives a partial explanation of the performance deviations between both supercomputers.

Apart from the benchmark discussed above, the performance can also be investigated in dependence on the model size. The bigger the entire model, the lower the ratio $\frac{\text{total size of areas to be exchanged}}{\text{"total model size"}}$ and the better the performance. Based on domain decompositions listed in Table 3.1, this ratio is illustrated for five different model sizes ranging from 256×256 to 4096×4096. As expected, the largest model reveals the best trade-off of model size and exchange area (see Figure 3.4a for pure domain decomposition and Figure 3.4b for shot parallelization). In contrast, in the combination of pure domain decomposition with the smallest model, approximately 48 % of the model area account for exchange areas. In this case an extensive domain decomposition does not make any sense. Not only intra-node communication but also inter-node communication increase significantly.
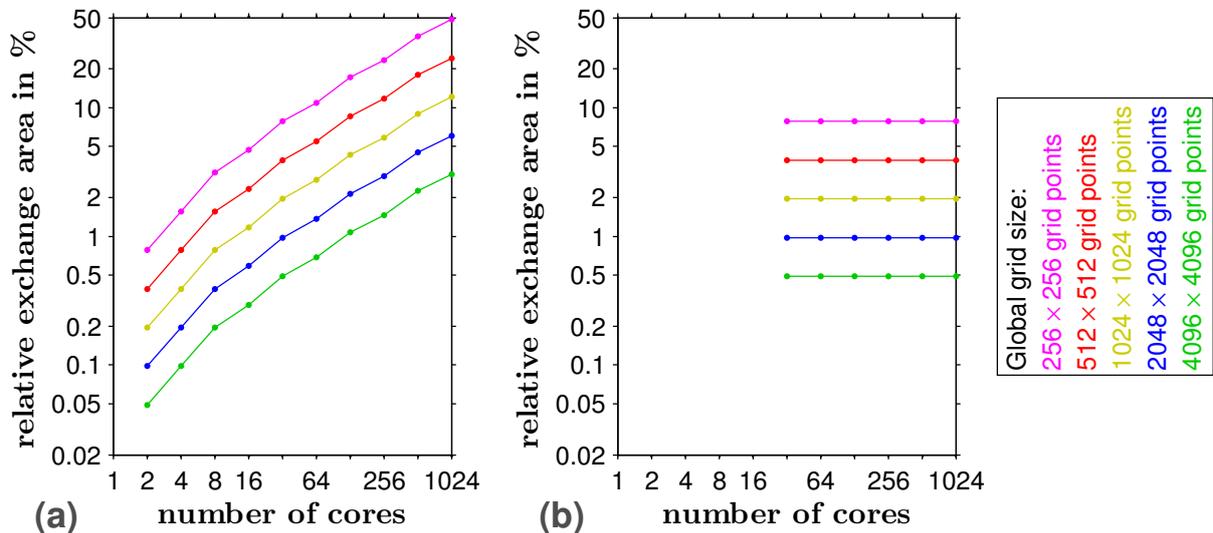

Figure 3.4: Relative amount of pressure-wavefield data which has to be exchanged in case of pure domain decomposition (a) and a combination with shot parallelization (b). The plots are related to 2D acoustic modeling using second-order spatial finite-difference operators. The relative exchange area represents the ratio of absolute exchange area to the total model size. The ratios are computed for different model sizes (color-coded).

3.2   Improvement of code efficiency

3.2.1   Memory consumption

Due to the implementation of shot parallelization, the memory consumption of time-domain FWT is of particular interest (see Figure 3.3f in previous section). In time-domain FWT most of the memory is required for the storage of forward-propagated wavefields. Hence, the following considerations are limited to this problem and omit other issues, such as storage of seismic data.

On the one hand, the finite-difference scheme prescribes the discretization in time and space which is defined by the stability and grid-dispersion criteria (see section 2.3.1.3). On the other hand, the corresponding time step Δt and grid spacing Δh do not have to match the requirements of FWT. In particular, this affects the choice of wavefield snapshots for the imaging condition and the spatial resolution:

- It is not necessary to store wavefield snapshots at all time steps. Based on the maximum frequency f_{max} of the propagating waveforms, a very strict Nyquist criterion can be employed to compute a new sampling interval

$$\Delta t_{samp} \leq \frac{1}{8 f_{max}}, \quad \Delta t_{samp} \overset{!}{\geq} \Delta t. \tag{3.1}$$

  Consequently, it is sufficient to use a subset of time steps allowing a reasonable memory consumption.

- The desired resolution of the model might be coarser than the spatial discretization of the finite-difference model. For example, the forward-propagated wavefield is stored in memory at every i-th grid point in all spatial directions – revealing a new spacing $\Delta h_{samp} = i \, \Delta h$. The imaging condition is then applied to every i^2-th (2D) or i^3-th (3D) grid point yielding a "down-scaled" model as the desired output. For subsequent modelings the model is scaled up using a trilinear interpolation to fulfill the requirements of the finite-difference scheme. This method results in a tremendous reduction of memory consumption.

Table 3.2 summarizes the memory consumption of an exemplary 3D FWT which either omits any optimization of the wavefield storage or involves the reduction techniques to allow a tolerable memory consumption. Although this example represents a small-scale problem, the usage of the finite-difference discretization Δt and Δh causes an excessive memory consumption. This clarifies that the application of such a FWT to a large-scale problem easily exhausts the available resources. The exploitation of the strategies mentioned above is able to reduce the memory consumption by several orders of magnitude.

Due to the availability of present-day supercomputers, a pure time-domain 3D FWT becomes feasible. Chapter 7 shows an application of 3D FWT to two small-scaled problems. This method might become an alternative to strategies, such as the single-frequency method (Sirgue et al., 2008) or pure frequency-domain implementations. In particular, 3D modeling in the frequency domain is highly demanding. However, recent works, such as Sourbier et al. (2009a,b) show remarkable improvements of parallel implementations in the frequency domain.

Table 3.2: Resource requirements for an exemplary 3D FWT using different methods of reducing the memory consumption: (A) storage of wavefield snapshots at all time steps and grid points, (B) snapshot storage for a subset of time steps at all grid points, (C) storage at all time steps and for a subset of grid points (every second grid point in each spatial direction), (D) storage for subsets of time steps and grid points.

General parameters				
Model size	$560 \times 240 \times 240$ grid points ($\hat{=} 123$ MiB)			
Modeling parameters	• 24 sources: peak frequency $f_{\mathrm{peak}} = 9$ Hz ($f_{\mathrm{max}} \approx 20$ Hz) • propagation time: $T = 5.6$ s • $\Delta t = 1.4 \cdot 10^{-3}$ s ($N_t = 4000$ time steps), $\Delta h = 10$ m • domain decomposition: $8 \times 4 \times 4$ • shot parallelization: all sources simultaneously • total number of cores: 3072			

Computational requirements for storage of wavefield snapshots				
Method	A	B	C	D
Discretization: Δt_{samp}	$1.4 \cdot 10^{-3}$ s	$1.25 \cdot 10^{-2}$ s	$1.4 \cdot 10^{-3}$ s	$1.25 \cdot 10^{-2}$ s
Number of snapshots	4000	438	4000	438
Discretization: Δh_{samp}	10 m	10 m	20 m	20 m
Grid size per core	$70 \times 60 \times 60$	$70 \times 60 \times 60$	$35 \times 30 \times 30$	$35 \times 30 \times 30$
Memory consumption per core	3.76 GiB	421 MiB	481 MiB	52.6 MiB
Overall memory consumption	11.3 TiB	1.23 TiB	1.41 TiB	158 GiB

3.2.2 Inversion workflow

Usually, FWT applications include well-known strategies to reduce the non-linearity of the inverse problem, such as frequency filtering over multiple stages (e. g., Bunks et al., 1995; Sirgue and Pratt, 2004) or the so-called "layer-stripping" effect (which is a gradual model recovery with increasing depth caused by application of data windowing; discussed by Shipp and Singh, 2002; Wang and Rao, 2009), within a pure time-domain FWT or the single-frequency method (shown in Figure 2.1). A detailed investigation of these methods can be found in chapter 4. Basically, each stage would require a separate FWT application to obtain intermediate v_{P} models, which are used as input for the successive stage. The workflow avoids this problem by allowing a sophisticated serial execution of these FWT runs within the framework of one FWT. Virieux and Operto (2009) give a brief overview of existing work, mentioning the "multiloop FWI workflow" revived in this chapter. Exemplary applications of multi-stage inversions are shown by Shin and Cha (2009) or Brossier et al. (2009). The workflow proves to be useful in case of very ill-posed problems, where the neglection of the abovementioned strategies results in a failure.

Apart from the features mentioned above, this work proposes a workflow implementation providing a variety of user-defined settings, such as the choice of a stop criterion for an automatic shifting between workflow stages or the possibility to switch between time-domain FWT and single-frequency method. An exemplary workflow input of the current FWT implementation is illustrated in Figure 3.5. The workflow may consist of an arbitrary number of stages represented by one input line. Every column is defined as follows:

- **Column 1**: This value sets the mandatory number of iterations per stage.

- **Column 2**: After completion of the mandatory iterations, additional iterations might be applied until the data misfit converges. If the stop criterion is fulfilled then the FWT continues with the next stage. The stop criterion is the relative misfit deviation between successive iterations. If the value is set to 0.0 then the FWT is restricted to the mandatory iterations.

- **Column 3**: This switch allows the application of a pure time-domain FWT (1) or the single-frequency method (2).

- **Column 4**: In case of a multi-parameter FWT this switch defines the parameter which is subject to the inversion, e. g., P-wave velocity (1), density (2), source-time function (3). This is a rudimentary implementation. Further improvements, such as parameter combinations, are necessary.

- **Column 5**: Data windowing is enabled (1) or disabled (0) by this switch. If it is enabled the following four columns are evaluated.

- **Columns 6 to 9**: Columns 6 and 7 define time windowing using a constant window for all offsets (begin end end of the time window in seconds). The offset window is analogously defined by columns 8 and 9.

- **Column 10**: This switch enables (1) or disables (0) frequency filtering of the recorded data or the source-time function. If it is enabled the following four columns are evaluated.

- **Columns 11 to 14**: Frequency filtering using a band-pass requires the specification of four corner frequencies.

- **Column 15**: This value is evaluated in case of applying the single-frequency method. Due to the possibility of a simultaneous inversion for an arbitrary number of frequencies, this column defines the corresponding number.

- **Columns 16+**: This is the list of frequencies required for the single-frequency method.

In the following, the workflow implementation is demonstrated using the example of a reflection experiment which involves the Marmousi model illustrated in Figure 3.6a. This model and all corresponding specifications are described in chapter 6. The workflow demonstration bases on the same configuration (see Table 6.3). However, the following case study aims to use the Marmousi experiment as a pure test problem rather than an systematic investigation of physical issues.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16+
4	0.03	1	1	1	0.0	2.5	0.0	-1.0	1	0.0	0.0	1.0	3.0	1	0.0
4	0.03	1	1	1	0.0	4.0	0.0	-1.0	1	0.0	0.0	1.0	3.0	1	0.0
8	0.03	1	1	0	0.0	0.0	0.0	0	1	0.0	0.0	1.0	3.0	1	0.0
8	0.03	1	1	1	0.0	2.5	0.0	-1.0	1	0.0	0.0	2.0	5.0	1	0.0
8	0.03	1	1	1	0.0	4.0	0.0	-1.0	1	0.0	0.0	2.0	5.0	1	0.0
5	0.02	1	1	1	0.0	-1.0	4500.0	6000.0	1	0.0	0.0	2.0	5.0	1	0.0
5	0.02	1	1	1	0.0	-1.0	2000.0	6000.0	1	0.0	0.0	2.0	5.0	1	0.0
5	0.02	1	1	0	0.0	0.0	0.0	0	1	0.0	0.0	2.0	5.0	1	0.0
4	0.01	2	1	0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	1	2.0
4	0.005	2	1	0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	3	2.0 3.0 4.0
20	0.00	1	1	0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	1	0.0

Figure 3.5: Exemplary content of a workflow input file used in the 2D and 3D acoustic FWT implementation. Each line represents an inversion stage. For a better recognizability the workflow is divided into five blocks. The first block applies different numbers of mandatory iterations (column 1). The stop criterion is set to a relative deviation of 3 % between successive iterations (column 2). A pure time-domain FWT (column 3) inverts for v_P only (column 4). Data windowing is partly enabled (column 5). Different time windows (in seconds) are taken into account (columns 6 and 7). Offset windowing is disabled (columns 8 and 9). Frequency filtering is enabled (column 10) and uses the band-pass frequencies (in Hz) in columns 11 to 14. Due to the pure time-domain FWT, columns 15+ are not required. While the first block focuses on time windowing and filtering, the remaining blocks have different tasks. The second block involves a broader frequency range. The third block deals with offset windowing (values are specified in meters) and filtering. The fourth block applies the single-frequency method (column 3) by using different number of frequencies (columns 15 and 16+). The last block disables all data computations and completes the FWT after exactly 20 additional iterations.

In contrast to section 6.2.2, a very simple initial model is used. It consists of a known water layer on top and a homogeneous half-space with $v_P = 3000 \frac{m}{s}$ (Figure 3.6b). Consequently, it reveals a huge model error of 25.2 % (computed by relation 6.1a). The corresponding recorded data and synthetic data obtained for the initial model significantly differ from each other (compare Figures 3.6a 3.6b with exemplary data for a shot located at $x = 2.555$ km). Due to the high initial velocity, the initial data is characterized by notable refraction events which do not exist in the recorded data as well as strong seafloor reflections.

Numerous possibilities of workflow schemes are considered. The following pure time-domain applications clearly show its impact on the resulting v_P model. While a simple scheme performs a plain inversion, the remaining tests involve more or less complex combinations of strategies to mitigate the ambiguity of the inverse problem:

(A) The first example is a plain FWT consisting of one stage:

- FWT is forced to perform 80 mandatory iterations,
- stop criterion: 1 %,
- neither data windowing nor filtering are applied.

(B) The second example involves time windowing over four stages:

- FWT is forced to perform 20 mandatory iterations per stage,
- stop criterion: 1 %,
- application of four time windows ranging from time $t = 0$ s to the maximum time $t_{max} = (1.75, 2.55, 3.55, 5.15)$ s.

(C) The third example involves offset windowing over four stages:

- FWT is forced to perform 20 mandatory iterations per stage,
- stop criterion: 1 %,
- application of four offset windows ranging from minimum offsets (5.0, 3.5, 2.0, 0.0) km to full streamer length.

(D) The fourth example involves time windowing and filtering over 16 stages divided into four blocks:

- FWT is forced to perform 5 mandatory iterations per stage,
- stop criterion: 1 %,
- application of four time windows ranging from time $t = 0$ s to the maximum time $t_{max} = (1.75, 2.55, 3.55, 5.15)$ s,
- for each time window the FWT inverts for four different frequency ranges resulting in peak frequencies $f_{peak} = (1.7, 2.9, 4.4, 9.0)$ Hz, where $f_{peak} = 9.0$ Hz represents the peak frequency of the unfiltered recorded data.

(E) The fifth example is similar to example "D" but with permuted order of time window-ing and filtering:

- application of four frequency ranges,

- for each frequency range the FWT applies a sequential time windowing with peak frequencies and time windows mentioned above.

(F) The sixth example resembles example "E" but with offset windowing instead of time windowing:

- application of four frequency ranges,

- for each frequency range a sequential offset windowing with windows ranging from minimum offsets (5.0, 3.5, 2.0, 0.0) km to full streamer length is applied.

While the plain FWT in example "A" is completely unable to recover the desired v_P model (Figure 3.7a), the exclusive applications of time windowing (example "B") or offset window-ing (example "C") do not significantly improve the result (Figure 3.7b,c). The model errors are imperceptibly reduced to 23.3 % ("A"), 19.4 % ("B") and 24.4 % ("C") and there is still a remarkable misfit between recorded data and final synthetic data.

However, the more complex strategies provided by examples "D", "E" and "F" produce satisfactory reconstructions of the v_P model. Figures 3.8, 3.9 and 3.10 illustrate the corres-ponding FWT progresses by showing v_P models and synthetic data at selected intermediate workflow stages. In all cases the last stage represents the final FWT result. Different com-binations of frequency filtering and time windowing ("D" and "E") yield the most robust inversion progresses. The model errors decrease down to 5.51 % (example "D", Figure 3.8) and 6.32 % (example "E", Figure 3.9), respectively. Although example "F" sufficiently recovers the desired v_P model, it reveals a quite unstable inversion progress by involving offset windowing (Figure 3.10). In particular, the early stages produce artificial interme-diate results. Apparently, the application of frequency filtering is able to compensate that disadvantageous effect. The final model error amounts to 5.73 %.

Apparently, offset windowing has to be handled with care to mitigate the ill-posedness of the inverse problem using a very poor initial model. However, depending on the application, time windowing is considered to be useful. For example, Shipp and Singh (2002) and Wang and Rao (2009) applied both time windowing and offset windowing to achieve the "layer-stripping" effect. In contrast, the most remarkable improvements are achieved by frequency filtering. This is also found by other authors, such as Sirgue (2006), who discusses the importance of inverting for low frequencies within the scope of the so-called multi-scale FWT.

The FWT applications in following chapters take advantage of the frequency-filtering method within the scope of the workflow implementation. In particular, the FWT parameter study in chapter 4 carries out detailed investigations on windowing and filtering.
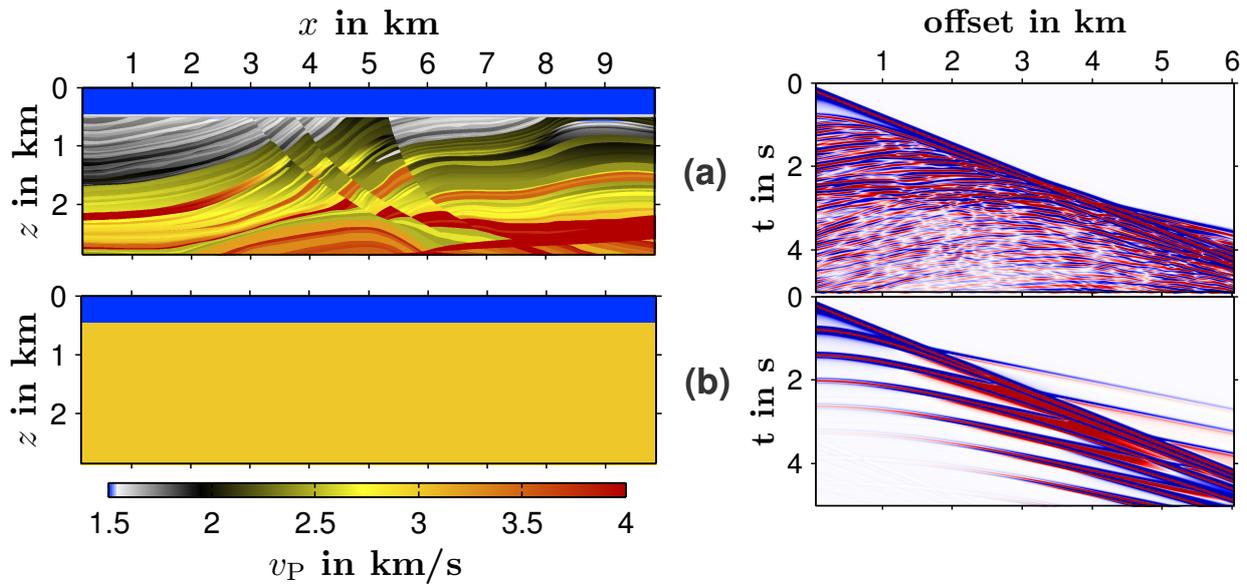
Figure 3.6: (a) shows the true v_P model (left), while (b) illustrates the initial v_P model for FWT. The right plots represent the corresponding recorded data (a) and initial synthetic data (b) for an exemplary shot located at $x = 2.555$ km. The data are individually normalized and clipped. Thus, amplitudes are not comparable.
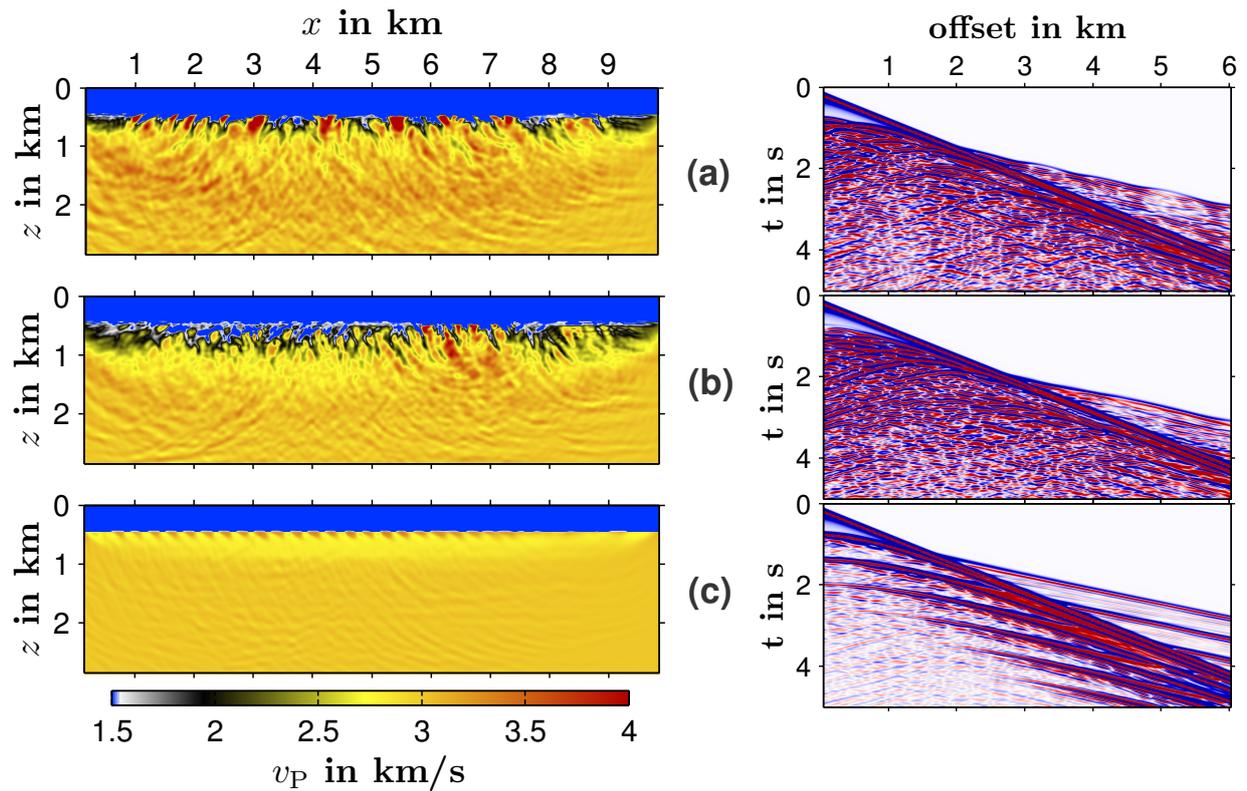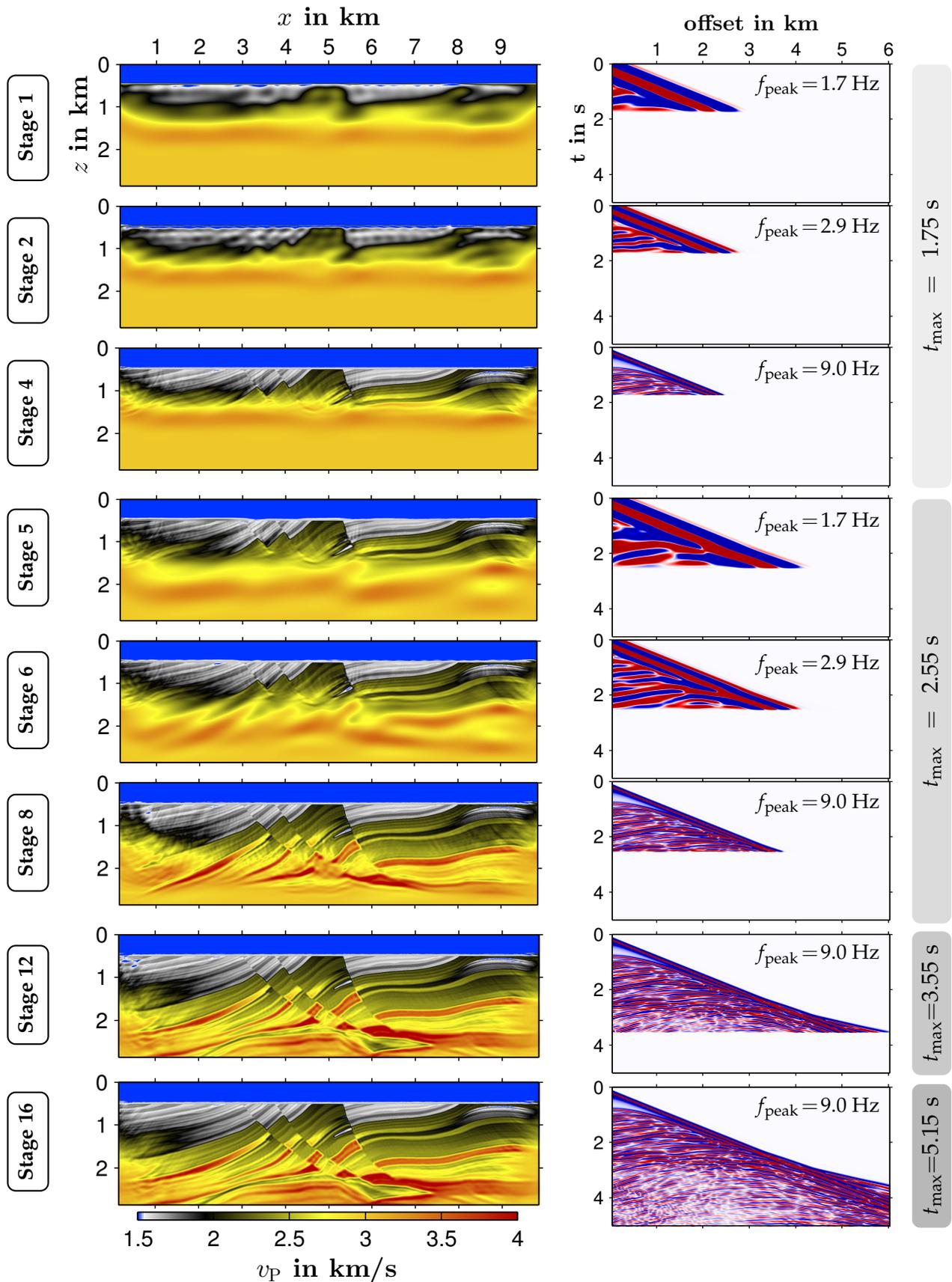

Figure 3.7: Inverted v_P models (left) and corresponding final synthetic data (right) for the workflow examples "A" (a), "B" (b) and "C" (c). The seismograms are obtained for an exemplary shot located at $x = 2.555$ km and are individually normalized and clipped. Thus, amplitudes are not comparable.

Figure 3.8: FWT progress of example "D" with v_P results (left) at the end of exemplary stages using time windowing and frequency filtering. Due to normalization and clipping, corresponding synthetic data (right, shot 9 at $x \approx 2.6$ km) are not comparable. Stage 16 is the final result.
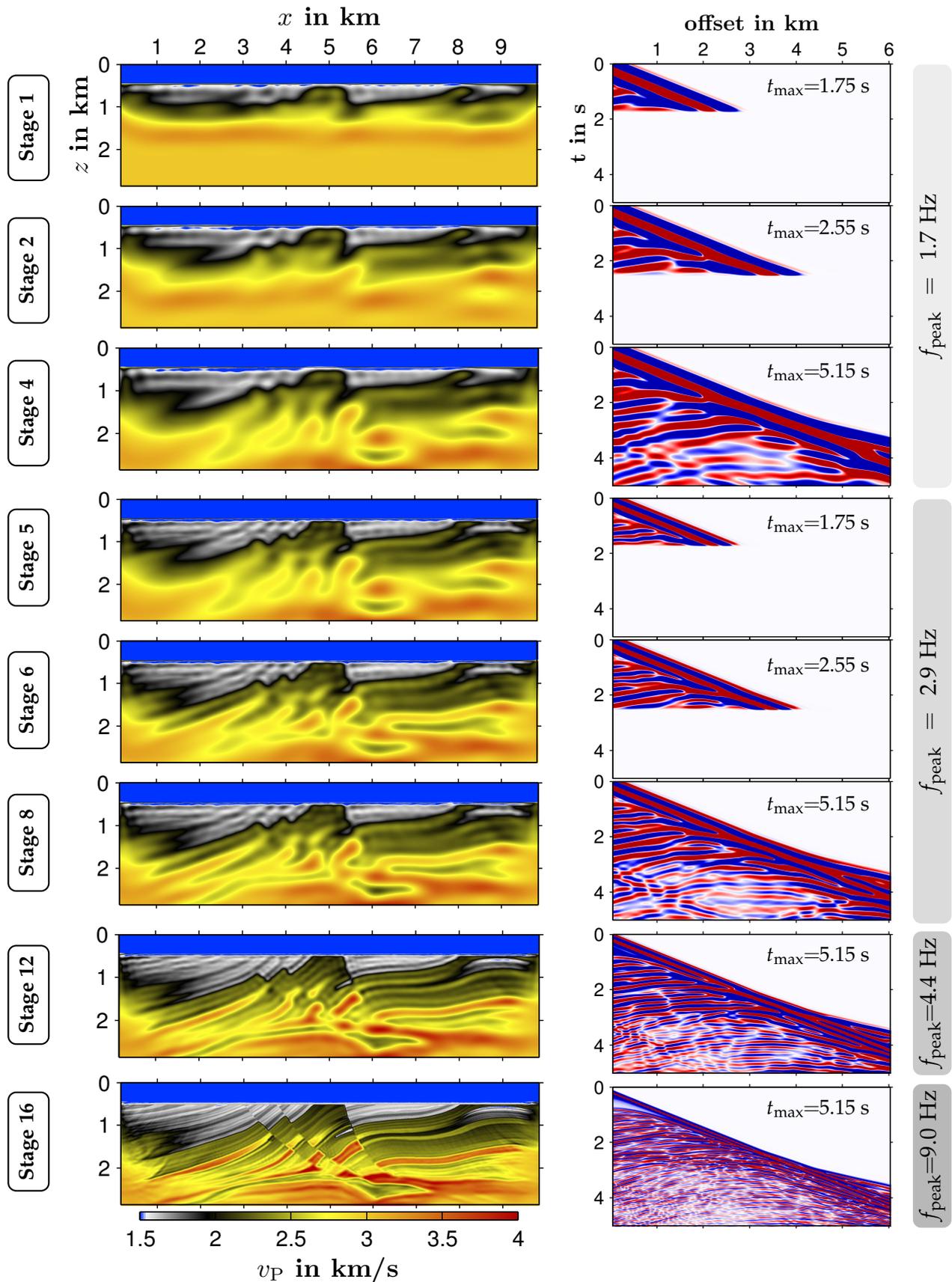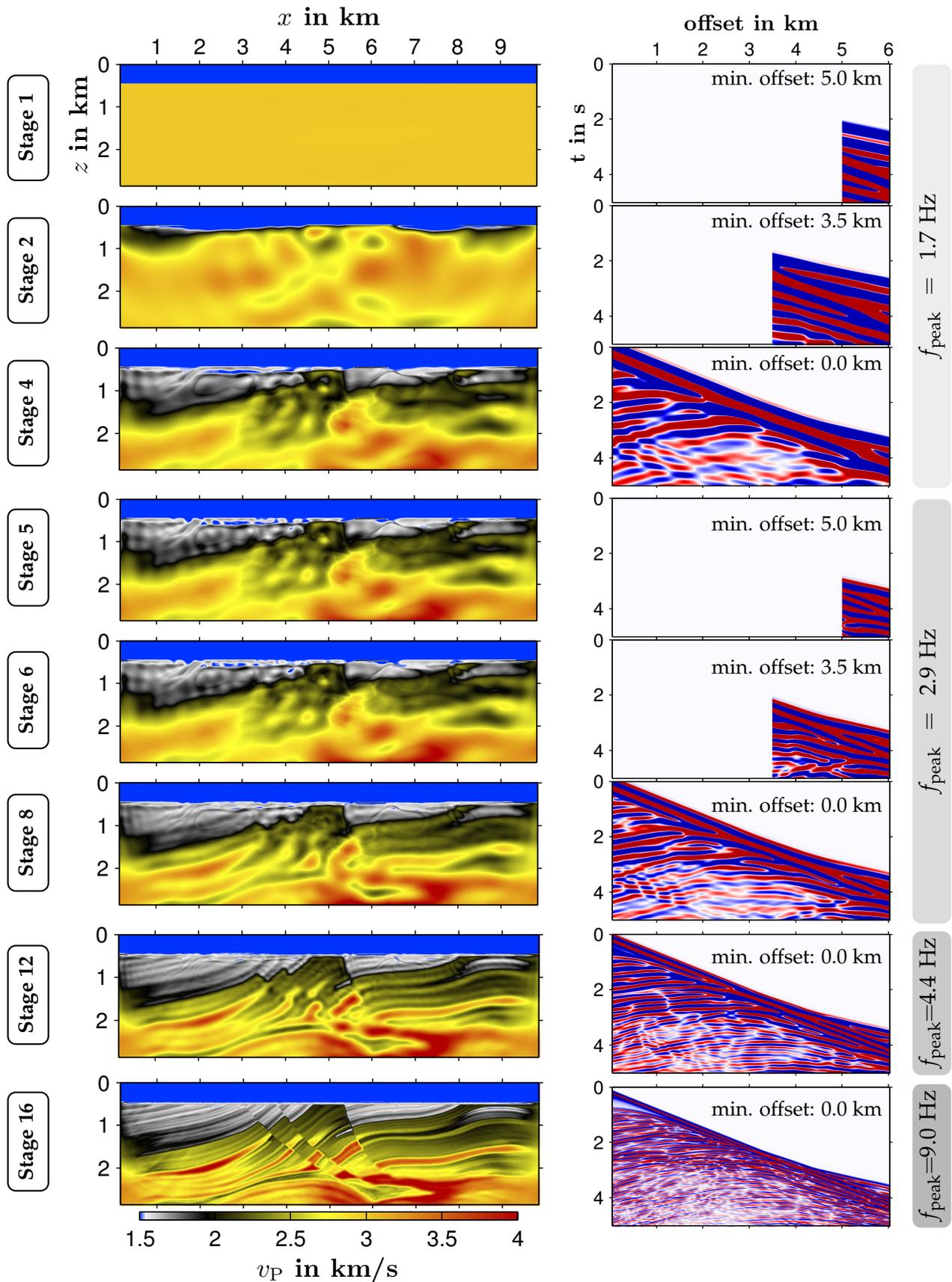
Figure 3.9: FWT progress of example "E" with v_P results (left) at the end of exemplary stages using frequency filtering and time windowing. Due to normalization and clipping, corresponding synthetic data (right, shot 9 at $x \approx 2.6$ km) are not comparable. Stage 16 is the final result.

Figure 3.10: FWT progress of example "F" with v_P results (left) at the end of exemplary stages using frequency filtering and offset windowing. Due to normalization and clipping, corresponding synthetic data (right, shot 9 at $x \approx 2.6$ km) are not comparable. Stage 16 is the final result.

3.3 Summary

Parallelization

Although computing power of cluster computers increases continuously, the overall performance is still limited to the speed of network-based communication. In time-domain FWT, most of the computing time is required for seismic forward modeling. Especially the 2D FWT applying a pure domain decomposition can not benefit from extensive computing resources. Due to a bad trade-off between model size and model area involved in communication, this parallelization method might be inefficient. However, parallelization can exploit the fact that FWT requires the modeling of more than one source per iteration. Hence, the domain decomposition is combined with shot parallelization to reduce or to omit inter-node communication. I tested both parallelization methods on the supercomputers *HERMIT* and *JUROPA*. Especially in case of using a high number of computational cores, the shot parallelization gains advantage over a pure domain decomposition. The shot parallelization allows a nearly linear scaling, while the pure domain decomposition yields a less significant performance improvement. Thus, I recommend an optimal combination of hardware architecture and both parallelization techniques to obtain the best performance. In the course of this thesis, all 2D applications involve a combined parallelization, where domain decomposition is limited to the CPU cores of one compute node.

Memory consumption

The usage of shot parallelization comes along with a high memory consumption, which is related to the storage of forward-propagated wavefield snapshots required for the imaging condition of FWT. In particular, 3D applications might become very expensive. However, it is not necessary to store wavefields with respect to the finite-difference discretization. On the one hand, the exploitation of the Nyquist criterion results in a coarser sampling of spatial snapshots with respect to time. On the other hand, the desired spatial discretization might be larger than the finite-difference grid spacing. Thus, the wavefield storage can be reduced to a subset of all finite-difference time steps and grid points. Consequently, the combination of paring down communication to the minimum by employing shot parallelization and the reduction of memory consumption is feasible in 2D and 3D applications – representing a very efficient time-domain FWT implementation.

Workflow

Finally, the usage of a workflow simplifies the handling of the FWT and avoids to carry out numerous single inversion runs. In general, the ill-posedness of inverse problems has to be mitigated by applying multi-stage approaches. The workflow allows a sequential execution of an arbitrary number of those stages involving data windowing or frequency filtering, among others. Concluding, the workflow simply reduces the amount of "manual work" and is thus extensively used in the following application chapters 4, 5, 6 and 7.

Chapter 4

Application 1: Parameter study

The aim of this chapter is the recapitulation of FWT theory described in chapter 2. Using a synthetic subsurface model with a cross-well geometry, it investigates the effect of following aspects on FWT:

- choice of initial model,

- data regularizations (hereinafter referred to as "data computations"): frequency filtering, time windowing, offset windowing,

- gradient computations: taper at source and receiver locations as well as wavefield-based and user-defined preconditioning,

- step length estimation,

- different acquisition geometries, i. e., varying number of sources within the transmission configuration or applying a reflection geometry.

Furthermore, the following constraints are applied to avoid unwanted side effects:

- inversion for v_P only,

- neglection of density,

- known source-time function.

This investigation employs a combination of conventional FWT and a brute-force search, i. e., inversions are performed for numerous initial models. Hereinafter, it is referred to as "multiple FWT". Therefore, this case study analyses the impact of the methods mentioned above on data and model misfit. On the one hand, the aim is to examine the effect of every single feature with omission of the remaining features. On the other hand, the final FWT tests summarize optimal parameter combinations.

4.1   Basic setup

4.1.1   General parameters

A comprehensive overview of the general setup, including acquisition geometry as well as general parameters for finite-difference modeling and inversion, can be found in Table 4.1. The following paragraphs give detailed information.

The true v_P model consists of a random medium with a circular inclusion (see Figure 4.1a). The choice of this model accounts for the requirements of corresponding experimental setups. The random background medium contains small-scale and large-scale structures with respect to the occurring wavelengths. The model represents a mixture of soft and hard rocks and less soft rocks. Additionally, there is a circular anomaly filled with water ($v_P = 1500 \frac{m}{s}$).

Within the scope of the inversion experiments, three different types of initial models are used (see Figures 4.1b to d). In dependence of the experimental aims, they contain a correct anomaly, no anomaly or a wrong assumption of its location. The models in Figures 4.1b to 4.1d are based on the central background velocity $v_P = 2200 \frac{m}{s}$ representing an exemplary case. The experiments discussed in section 4.2 carry out investigations on a wider range of initial background velocities.
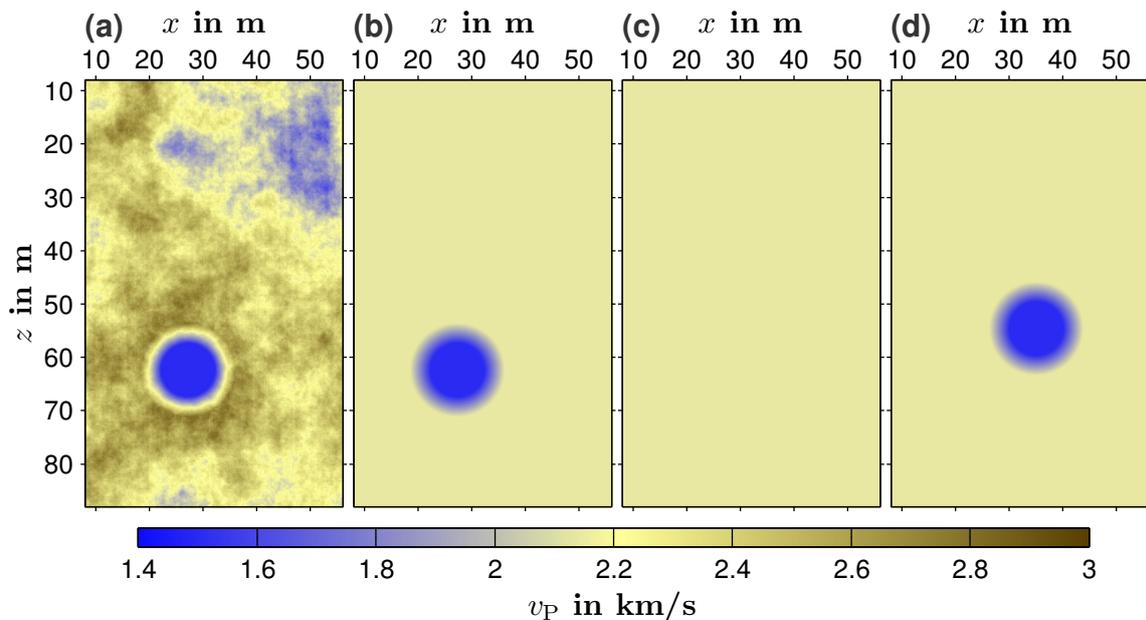

Figure 4.1: Application 1: (a) shows the true v_P model. (b), (c) and (d) illustrate the initial models **A**, **B** and **C** using an exemplary background velocity $v_P = 2200 \frac{m}{s}$.
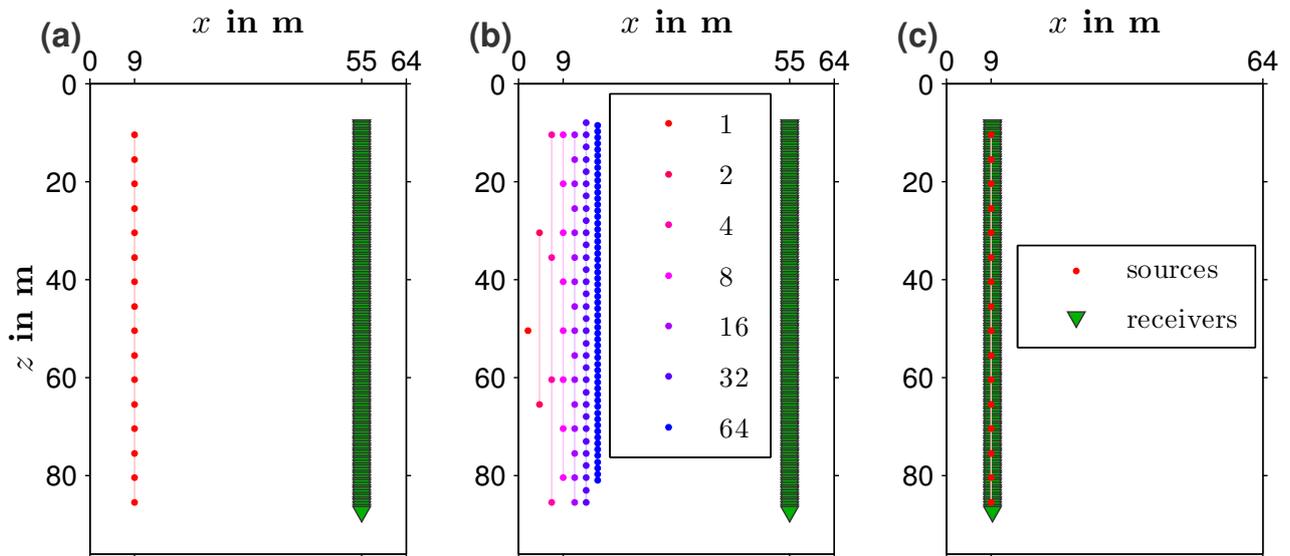
Figure 4.2: Application 1: Acquisition geometries. (a) shows the transmission geometry used in experiments I-IV, VI, VII/1, VII/2, VIII, IX and X. In contains 16 sources numbered from top to bottom. (b) shows the transmission geometry with variable number of sources used in experiment V/1. All sources are located at lateral location $x = 9$ m. (c) shows the reflection geometry required by experiments V/2, VII/3 and VII/4. All geometries contain 159 receivers.
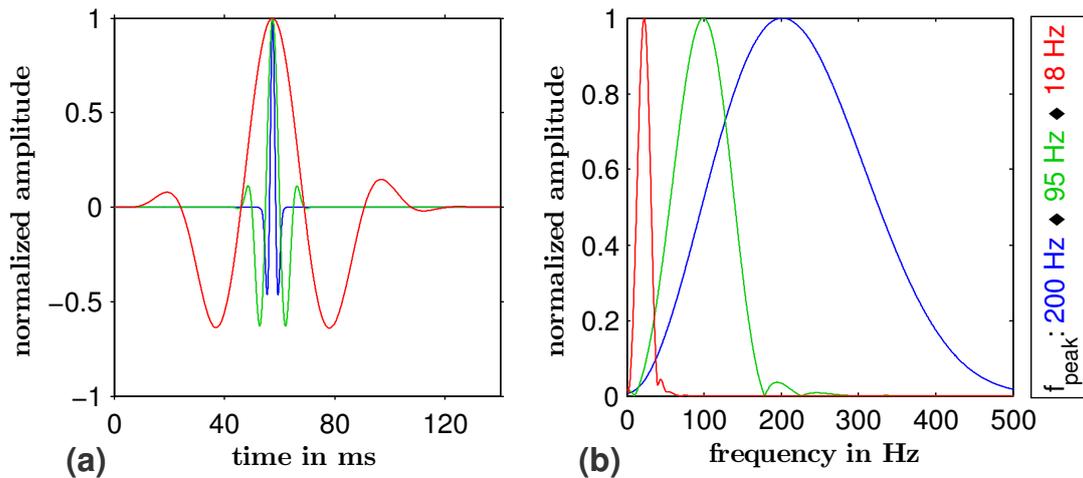

Figure 4.3: Application 1: (a) shows the source time functions used in the FWT experiments. Based on the Ricker wavelet (blue), two additional wavelets are computed by low-pass filtering. Due to visualization reasons, amplitudes are normalized individually. (b) illustrates corresponding amplitude spectra which are normalized individually.

The experiments apply several acquisition geometries. On the one hand, sources and receivers are located in two vertical boreholes forming a transmission acquisition geometry. On the other hand, they are arranged in one borehole forming a reflection geometry. The transmission configuration with 16 sources represents the usual choice (Figure 4.2a). Apart from that, some experiments focus on the influence of the acquisition geometry. Therefore, several transmission configurations with source numbers varying from 1 to 64 are used (Figure 4.2b). Furthermore, the source locations of the reflection geometry in Figure 4.2c and

the "standard" geometry in Figure 4.2a are identical. In all cases the receiver array consists of 159 hydrophones. Further details are listed in Table 4.1.

For all experiments, FD modelings use the Ricker wavelet with the peak frequency $f_{peak} = 200$ Hz as source time function (Figure 4.3a) to generate the observed data. Some experiments require frequency filtering. Hence, based on the wavelet containing the full frequency content, two additional low-pass filtered signals with $f_{peak} = 18$ Hz and $f_{peak} = 95$ Hz are computed (compare Figures 4.3a and b). Apart from a few weak side lobes in amplitude spectra, they resemble the characteristics of the original Ricker wavelet. Furthermore, all source signals are shifted in time by 0.05 seconds to ensure causal wavelets in case of low-pass filtering.

Using the "standard" acquisition geometry (Figure 4.2a), observed pressure seismograms are depicted in Figure 4.4. One can see the data for the uppermost, central and bottommost shot locations. Due to the transmission geometry, the seismograms mainly contain the direct wave and diffraction events caused by the circular anomaly in Figure 4.1.
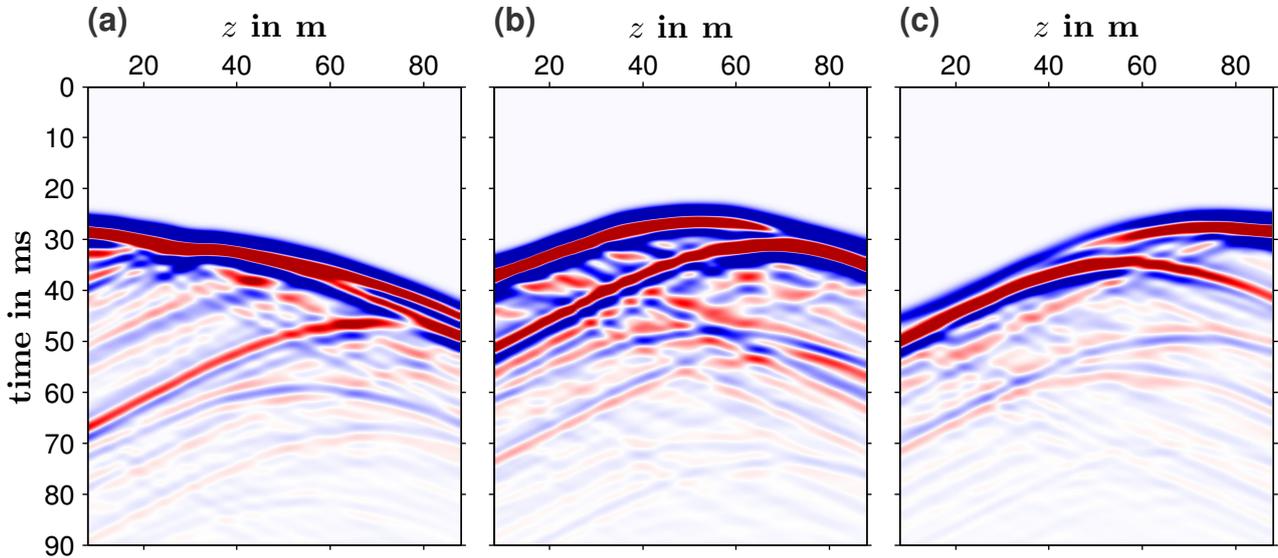

Figure 4.4: Application 1: Exemplary observed seismograms for the true model as well as shots 1 (located at $z = 10.5$ m), 9 ($z = 50.5$ m) and 16 ($z = 85.5$ m) of the acquisition geometry shown in Figure 4.2a.

4.1.2   Configuration of all tests

In total, ten experiments have been performed to investigate the aspects mentioned above. The compilation and setup of all tests can be found in Tables 4.2 to 4.4 and in Table D.1 in appendix D.1. Each test deals with a different aspect highlighted in green. The case study is divided into three parts grouping experiments I to V, VI to IX and X.

Table 4.1: Application 1: General setup for all experiments.

Application 1: Parameter study		
Attributes		**Specifications**
Model	**Size**	width: 64 m, height: 96 m
	Structure	random background medium	· average v_p = 2200 m/s · maximum range: v_p = [1600, 2800] m/s · standard deviation σ = 0.1 · correlation length: 16 m · exponential autocorrelation function
		circular anomaly	· v_p = 1500 m/s (water) · diameter: 17.5 m
	Figure	Figure 4.1
Acquisition geometry	**Location**	full-space, borehole
	Configurations	· **T** = transmission geometry (cross-hole) · **R** = reflection geometry (1 borehole)
	Sources	· explosive sources · source-time function: Ricker wavelet · peak frequency: 200 Hz · time delay of source signal: 0.05 s
	Receivers	· 159 hydrophones · recording length: 0.14 s
	Offsets	· **T**: minimum: 46 m, maximum: 90 m · **R**: minimum: 0.25 m, maximum: 77 m
	Figure	Figure 4.2
Parameters for FD-modeling	**Domain**	time domain
	Model size	· width: 256 grid points, height: 384 grid points · grid spacing: 0.25 m
	Propagation time	· 3500 time steps · time discretization: $4 \cdot 10^{-5}$ s
	PML boundary	width: 8 m (at all model boundaries)
General inversion parameters	**Initial models**	background medium	homogeneous: v_p = (1430, 1461, ..., 2939, 2970] m/s
		circular anomaly	· **A**: correct anomaly · **B**: without anomaly · **C**: anomaly at wrong location
		Figure 4.2
	Domain	imaging condition in the time domain

Table 4.2: Application 1: Basic setup of FWT parameters. "A", "B", "C", "T" and "R" refer to terms given by Table 4.1.

Experiment		Attribute of interest		Initial model	Acquisition geometry		Figure(s)
ID	FWT type	Type	Parameter		Type	Sources	
I	multiple FWT	none	none	A	T	16	4.5, 4.6, 4.7, D.1
II/1	multiple FWT	data	time or offset windowing	A	T	16	4.8, 4.12, D.2, D.3
II/2			low-pass filtering				4.9, 4.10, 4.12, D.4
II/3			low-pass filtering				4.11, 4.12, D.5
III/1	multiple FWT	gradient	taper at geometry	A	T	16	4.13, 4.14, D.6
III/2			wavefield-based taper				4.13, 4.14, D.7
III/3			user-defined taper				4.13, 4.14, D.8
IV/1	multiple FWT	model update	step length estimation	A	T	16	4.15, 4.16, D.9
IV/2	single FWT						4.17
IV/3	single FWT						4.18
IV/4	single FWT						4.19
V/1	multiple FWT	acquisition geometry	variable number of sources	A	T	1, 2, 4, 8, 16, 64	4.20, 4.21, D.10 … D.15
V/2			reflection geometry	A	R	16	4.22, D.16
VI/1	multiple FWT	optimal parameters	optimal parameters	A	T	16	4.23
VI/2			apparently most optimal parameters			64	D.17
VII/1	multiple FWT	none	none	B	T	16	D.18
VII/2		optimal parameters	optimal parameters				D.19
VII/3		geometry	reflection geometry		R		D.20
VII/4		optimal parameters	optimal parameters				D.21
VIII/1	multiple FWT	none	none	C	T	16	D.22
VIII/2		optimal parameters	optimal parameters				D.23
IX	multiple FWT	wrong assumption of user-defined gradient taper		C	T	16	D.24
X	brute force	model update	test of 1001 different step lengths	B	T	16	4.24, 4.25, 4.26, D.25

Table 4.3: Application 1: Detailed setup of all FWT experiments (part 1).

| Experiment | Data computations | | | Gradient computations | | | | | |
	Filtering Frequencies f_{peak} in Hz	Maximum window size: Time in s	Offset in m	Cosine taper at locations of each… source R in m	source $f(r=0,\frac{R}{2})$	receiver R in m	receiver $f(r=0,\frac{R}{2})$	Wavefield-based	User-defined
I	✗	✗	✗			✗		✗	✗
II/1	✗	0.08, 0.1, 0.14	48, 60, max. offset						
II/2	95	✗	✗			✗		✗	✗
II/3	18, 95, 200	✗	✗						
III/1				8	(0.5, 0.9)	2	(0.5, 0.9)	✗	✗
III/2	✗	✗	✗			✗		✓	✗
III/3						✗		✗	✓
IV/1									
IV/2	✗	✗	✗			✗		✗	✗
IV/3									
IV/4									
V/1									
V/2	✗	✗	✗			✗		✗	✗
VI/1	18, 95, 200	✗	✗	8	(0.5, 0.9)	2	(0.5, 0.9)	✗	✗
VI/2									
VII/1	✗					✗			
VII/2	18, 95, 200	✗	✗	8	(0.5, 0.9)	2	(0.5, 0.9)	✗	✗
VII/3	✗					✗			
VII/4	18, 95, 200			8	(0.5, 0.9)	2	(0.5, 0.9)		
VIII/1	✗					✗			
VIII/2	18, 95, 200	✗	✗	8	(0.5, 0.9)	2	(0.5, 0.9)	✗	✗
IX	18, 95, 200	✗	✗	8	(0.5, 0.9)	2	(0.5, 0.9)	✗	✓
X	18, 95, 200	✗	✗			✗		✓	✗

Table 4.4: Application 1: Detailed setup of all FWT experiments (part 2).

| Experiment | Model update: step length μ | | | | Model update: constraints | | | |
	Type	Initial $\mu_{\mathrm{rel,ini}}$	Factor a	Number of test shots	v_{P} limits in $\frac{\mathrm{m}}{\mathrm{s}}$ Minimum	Maximum	Maximum relative deviation from… $v_{\mathrm{P|ini}}$	$v_{\mathrm{P}|h-1}$
I	constant	1	✗	✗	1400	3000	100 %	5 %
II/1								
II/2	constant	1	✗	✗	1400	3000	100 %	5 %
II/3								
III/1								
III/2	constant	1	✗	✗	1400	3000	100 %	5 %
III/3								
IV/1	adaptive	1	8	8				
IV/2	constant	4, 1, 0.25	✗	✗	1400	3000	100 %	5 %
IV/3	adaptive	1	2, 4, 8	8				
IV/4	adaptive	1	8	1, 2, 4, 8, 16				
V/1	constant	1	✗	✗	1400	3000	100 %	5 %
V/2								
VI/1	adaptive	1	4	8	1400	3000	100 %	5 %
VI/2								
VII/1	constant		✗	✗				
VII/2	adaptive	1	4	8	1400	3000	100 %	5 %
VII/3	constant		✗	✗				
VII/4	adaptive		4	8				
VIII/1	constant	1	✗	✗	1400	3000	100 %	5 %
VIII/2	adaptive		4	8				
IX	adaptive	1	4	8	1400	3000	100 %	5 %
X	step length scan	[0, 1000] %	✗	8	1000	3600	✗	✗

Experiment I represents the reference test. It comprises a basic FWT without the application of any features. Experiments II to V investigate the impact of attributes, such as data computations (II), gradient computations (III), methods directly affecting the model update (IV) as well as the acquisition geometry (V). Basically, each study focuses on a single attribute (see Table 4.3), while all other parameters remain unchanged.

Furthermore, experiments VI to IX involve selected parameter combinations. On the one hand, the aim is to obtain the most optimal inversion results (VI, VII/2, VIII/2). On the other hand, the influence of initial velocity models B and C (VII/1, VIII/1) is analyzed. In case of experiment VII, selected experiments (I, VI/1 and V/2) are repeated in conjunction with initial model B. Experiment IX demonstrates the failure of FWT caused by wrong model assumptions. The last experiment (X) represents a pure brute-force search, which scans the data misfit function for several initial models and step lengths.

In detail, Table 4.2 summarizes general information about the experiment, such as its aim and the initial background model (see Table 4.1). In particular, Tables 4.3 and 4.4 provide a list of FWT parameters, such as data and gradient computations as well as step length estimation and model-update constraints. A computational summary of all experiments can be found in Table D.1.

In general, the multiple FWTs are applied to 51 equidistantly spaced initial v_P background models. They range from 1430 to 2970 $\frac{m}{s}$ with the "central" model at 2200 $\frac{m}{s}$ which is the average of the random background medium. For each model, complete FWTs are performed resulting in 51 different inversion results as well as evolutions of data and model error.

Experiment X represents a special case. Instead of performing the iterative inversion scheme, the misfit function is scanned at the first iteration. For this purpose, the brute-force search varies both the initial background model (as mentioned in the previous paragraph) and the step lengths (1001 values ranging from 0 to 1000 %).

Apart from the multiple FWT experiment IV/1, another exceptional case is represented by experiments IV/2 to IV/4. They consists of conventional FWTs using one initial background model and perform 100 iterations to investigate different step-length strategies.

4.2   Results of FWT experiments

This section presents and discusses the results of all FWT experiments. The discussion is associated with Figures containing the key information. Additional illustrations can be found in appendix D.2.

Quantification of errors

The performance of all tests is quantified by the calculation of data misfits ϵ_{data} and model errors ϵ_{model}. They are computed with respect to the true model $\mathbf{v}_{\text{P|true}}$ and to the observed data $\mathbf{p}_{\text{obs}}$. Due to the usage of the least-squares norm in the FWT algorithm, the data misfits are expressed as normalized squared L_2 norms, whereas the model errors are normalized L_1

norms to demonstrate the relative percentage deviation between models:

$$\epsilon_{\text{data}} := \epsilon(\mathbf{p}_h) = \frac{\|\mathbf{p}_h - \mathbf{p}_{\text{obs}}\|_2^2}{\|\mathbf{p}_{\text{obs}}\|_2^2} \, c_{\text{geo}} \qquad \text{(data misfit)}, \qquad (4.1a)$$

$$\epsilon_{\text{model}} := \epsilon\left(\mathbf{v}_{\text{P}|h}\right) = \frac{\left\|\mathbf{v}_{\text{P}|h} - \mathbf{v}_{\text{P}|\text{true}}\right\|_1}{\left\|\mathbf{v}_{\text{P}|\text{true}}\right\|_1} \qquad \text{(model error)}, \qquad (4.1b)$$

where $\mathbf{p}_h$ denotes the synthetic data for the intermediate model $\mathbf{v}_{\text{P}|h}$ at iteration h. The division by data norm $\|\mathbf{p}_{\text{obs}}\|_2^2$ accounts for content-related changes in the data, such as offset windowing, time windowing or frequency filtering. These methods are applied to both synthetic data and observed data. In case of using acquisition geometries deviating from experiment I, i. e., usage of different number of sources, the data misfit is corrected by the factor $c_{\text{geo}} = \frac{N_s}{16}$.

Illustration of multiple-FWT results

To accentuate the results of both multiple FWTs in experiments I to IX, figure compounds with relevant plots are added to subsequent sections 4.2.1 to 4.2.9 as well as to appendix D.2. A figure compound, such as Figure 4.5, is divided into three parts:

- The top row shows exemplary initial models selected from all 51 initial background velocities. All of them are numbered as follows: ① ($v_{\text{P}} = 1430 \, \frac{\text{m}}{\text{s}}$), ② ($v_{\text{P}} = 1461 \, \frac{\text{m}}{\text{s}}$), ..., ㉖ ($v_{\text{P}} = 2200 \, \frac{\text{m}}{\text{s}}$), ..., ㊿ ($v_{\text{P}} = 2939 \, \frac{\text{m}}{\text{s}}$), �① ($v_{\text{P}} = 2970 \, \frac{\text{m}}{\text{s}}$).

- The progress of the inversion is depicted in the middle part. In general, it contains the evolution of the data misfit function and the corresponding model error for all 51 initial models. For each type of initial models (**A**, **B** or **C**), the data misfits are normalized to the maximum among all 51 test computations of the respective reference FWT (experiments I, VII/1 or VIII/1). Hence, in case of "**A**" experiments I, II, III, IV/1, V/1 and VI are comparable. Due to a completely different acquisition geometry, the data misfit of experiment V/2 is normalized to its own maximum. Furthermore, in case of "**B**" all experiments VII and in case of "**C**" experiments VIII as well as IX are comparable.
  The FWT progress that belongs to exemplary initial models (shown in the top row) is highlighted by dashed lines. Occasionally, this part features an additional plot showing the evolution of the step length.

- The bottom row illustrates the inversion result after a certain number of iterations. The inverted models belong to their counterparts in the top row.

For all experiments, the inverted v_{P} model with the lowest model error and its corresponding initial model are highlighted by a green frame. Furthermore, in experiments I to VI most of the tests are illustrated and described using a particular model featuring a initial background velocity of $v_{\text{P}} = 2138 \, \frac{\text{m}}{\text{s}}$ (model ㉔), which is a crucial choice to demonstrate several methodological effects. A red frame is used to highlight this exemplary test case.

4.2.1   Experiment I: Basic FWT

Experiment I represents the reference FWT consisting of a basic inversion configuration without application of any features improving the performance of the FWT. The test setup comprises the initial model type A (Figure 4.1b) and the transmission geometry (Figure 4.2b). Figure 4.5 illustrates the results of the multiple FWT. The initial models [19], [24], [28], [30], [32] and [39] with background velocities $v_P = (1984, 2138, 2262, 2323, 2385, 2600) \frac{m}{s}$ are utilized to demonstrate the problems of such a plain inversion.

Obviously, there is a very small range of initial background velocities (hereinafter referred to as "model window") allowing a successful inversion. Especially between initial models [24] ($v_P = 2138 \frac{m}{s}$) and [25] ($v_P = 2169 \frac{m}{s}$) both the data misfit function and the model errors exhibit a high-contrast transition. Figure D.1 provides a different selection of initial models to focus on that sharp transition area. In case of model [24] a slightly decreasing data misfit is opposed to a slightly increasing model error, i. e., the reduction of the data misfit does not improve the velocity model. The FWT might got stuck in a local minimum of the objective function. Starting with the initial data (Figure 4.6b), the inversion obtained the final data (Figure 4.6c) which does not fit the observed data (Figure 4.6a) very well. Especially the artificial recovery of the lower model areas cause a significantly high final residual data (Figure 4.6d). In contrast, the usage of model [25] yields a tremendous reduction of both data misfit and model error.

The occurrence of the clearly delimited "model window" – ranging from model [25] to [31] – is caused by the cycle-skipping effect (e. g., Ravaut et al., 2004). Figure 4.7 shows exemplary traces for shot number 16 (see Figure 4.4) and demonstrates the offset dependency of this issue. Especially at the largest offset (Figures 4.7a,b) the data are characterized by a significant phase shift. While the data with respect to initial models [24] and [32] is shifted by approximately half a cycle, the data for initial models [19] and [39] exhibit a phase shift of nearly one cycle (Figure 4.7a). In case of the half-cycle shift, the data of the inversion result shows a notable similarity of observed and synthetic direct wave. Some later events already exhibit more or less remarkable phase shifts. In contrast, in case of the nearest offset (Figures 4.7c,d) the initial data is characterized by much smaller phase shifts. Hence, the FWT with small offsets seems to be less ill-posed. Apart from initial models [19] and [24], the choice of the remaining cases results in a satisfactory fit of observed and final synthetic data (Figure 4.7d). In particular, the FWT using model [19] tries to fit the wrong minimum phase, i. e., it gets stuck in a local minimum of the data misfit function due to the cycle-skipping effect. Additionally, in this case the FWT with higher initial velocities, such as model [39], is less sensitive to phase distortions. On the one hand, the good fit of small-offset data causes a partially sufficient recovery of the v_P model (see final results [24] and [32] in Figure 4.5). On the other hand, the bad data misfit is obtained by generating significant artifacts.

The preceding observation exactly coincides with a condition proposed by Beydoun and Tarantola (1988). According to them, the initial model may cause a time shift of up to half a cycle. Otherwise, the FWT ends up in a local minimum. Additional to that time-domain condition, a frequency-domain FWT requires an analogous formulation (Pratt et al., 2008).

Furthermore, experiment I illustrates the problem of choosing an optimal initial v_P model. A reasonable strategy is to find the initial model providing the minimum initial data misfit. A search method, such as a bisection algorithm, can be used to reduce the number of models
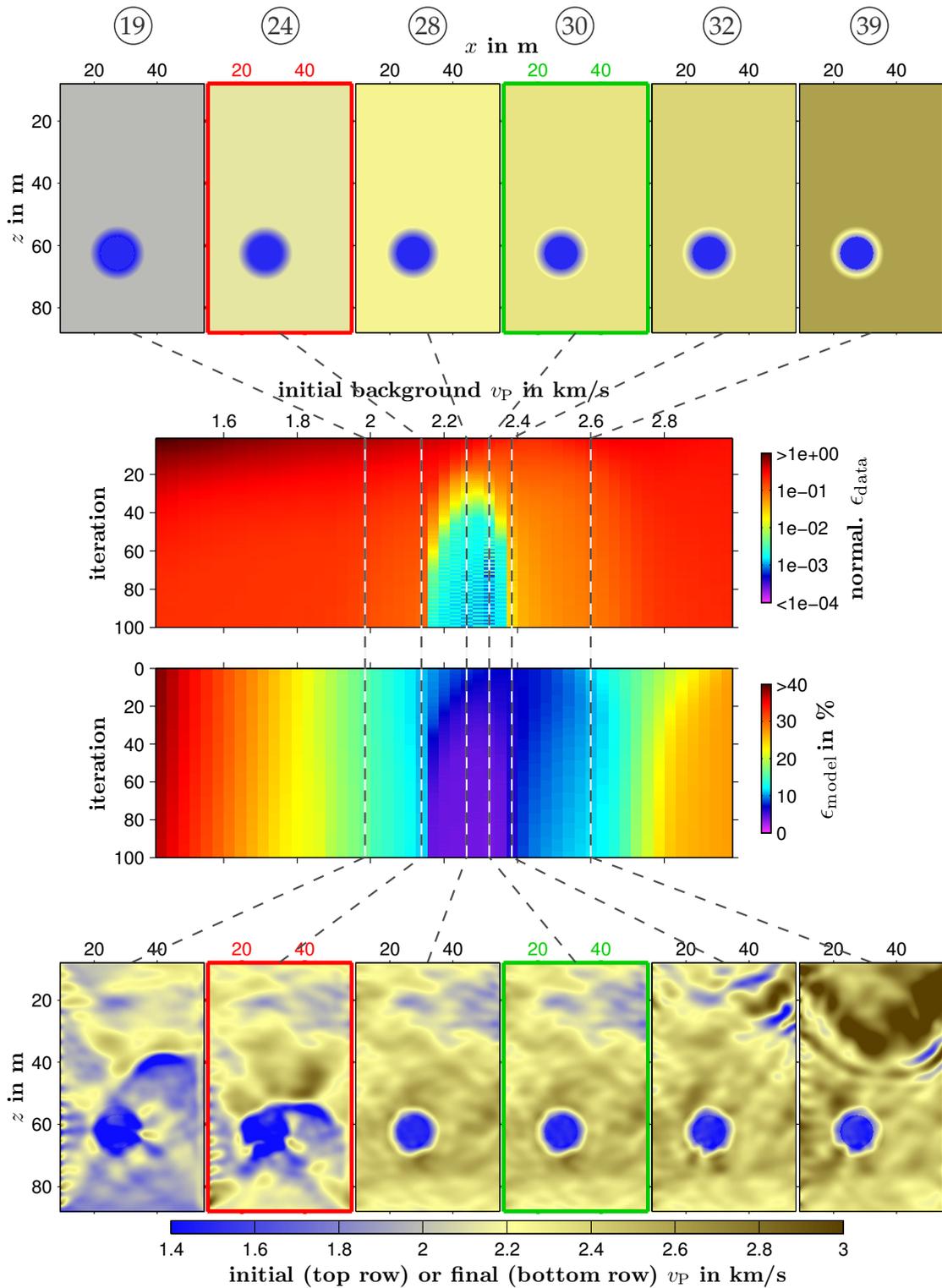
Figure 4.5: Application 1, experiment I: selected initial models (top, with model ID $\overset{\text{X}}{\bigcirc}$) and corresponding v_P results (bottom) as well as the progress of the reference FWT for all 51 initial background models (central part). The upper plot shows the data misfit normalized to the maximum value of all tests. The lower plot illustrates the model error. The extreme values are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 6.7 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 3.8)$ %. The red frame indicates the initial model $\overset{24}{\bigcirc}$ used to demonstrate the outcome of the experiments. The green frame highlights the best model reconstruction.
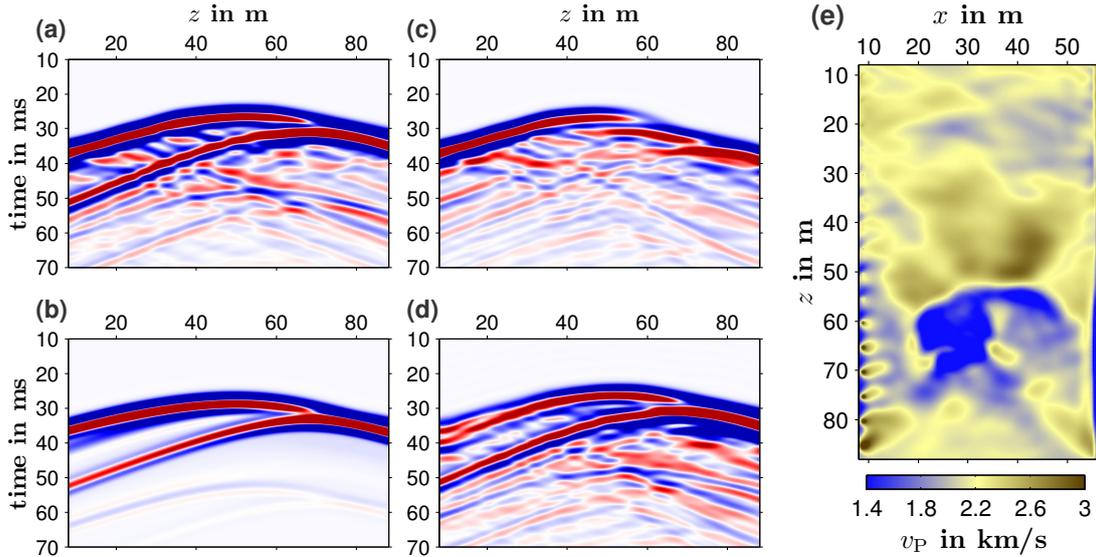
Figure 4.6: Application 1, experiment I: (a) to (d) illustrate seismograms for the central shot 9: observed data (a), the synthetic data for the initial model (b), the final synthetic data (c) and the residuals (d). (e) shows the final inversion result.
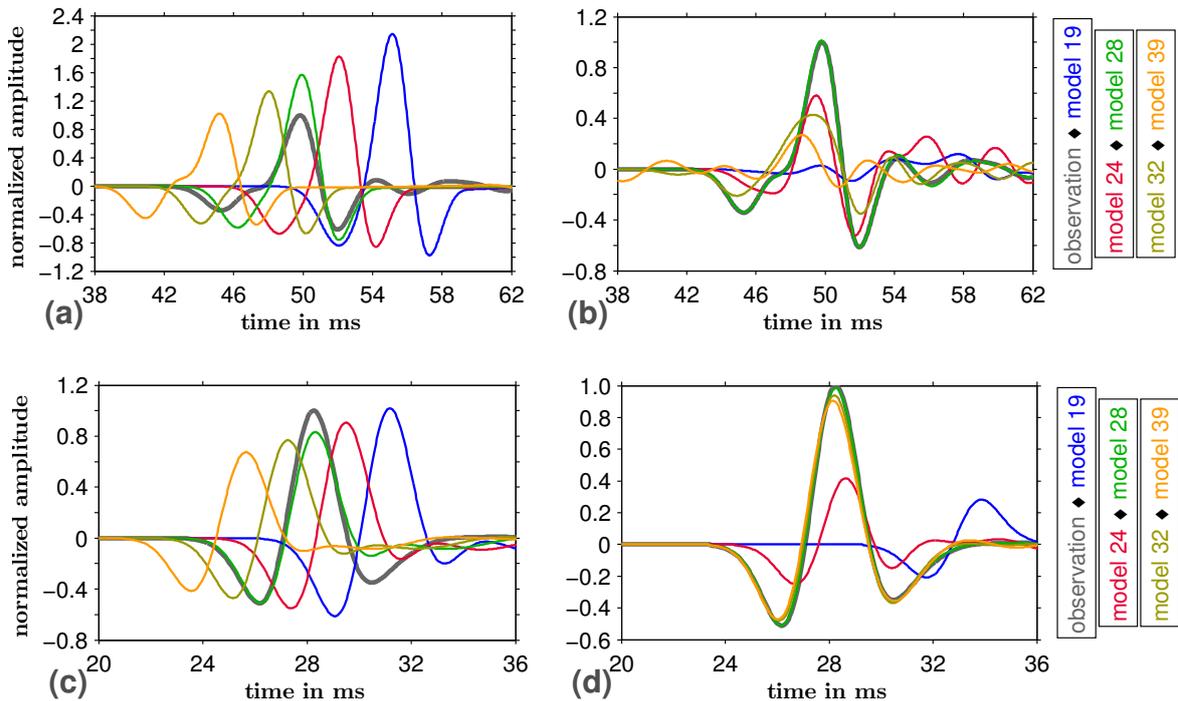
Figure 4.7: Application 1, experiment I: Exemplary data for the deepest source location at $z = 85.5$ m (shot 16) as well as different initial models (19), (24), (28), (32) and (39). In each plot all traces are normalized with respect to the maximum of the observed data. The upper row shows the data for the largest offset. The lower row depicts the data for the smallest offset. While the left figure column shows the initial data, the right column represents the inversion result.

to be tested. The "optimal" choice of an initial model is defined by the minimum initial data misfit which does not necessarily match the minimum initial model error. In case of experiment I model ③④ ($v_P = 2446\,\frac{m}{s}$) represents such an "optimal" choice. However, the progress of the inversion gets stuck in a local minimum resulting in an artificially reconstructed v_P model.

The application of all FWTs reveals that both the minimum data misfit and the minimum model error are obtained for model ③⓪ ($v_P = 2323\,\frac{m}{s}$). The usage of model ③④ causes a data-misfit reduction of less than one order of magnitude. In contrast, the usage of model ③⓪ yields a misfit reduction of more than two orders of magnitude. This observation is verified by the changes of model errors between initial and final model. In case of model ③④ the model errors increases from 7.6 % to 9.5 %! In contrast, the usage of model ③⓪ shows a remarkable reduction from 7.3 % to 3.8 % – representing the best inversion result of all 51 tests.

4.2.2   Experiment II: Data computations

Experiment II investigates the influence of different data-selection methods applied during the FWT. In the following, offset windowing, time windowing and frequency filtering are discussed.

Experiment II/1: Offset and time windowing

As mentioned in previous section 4.2.1, the consideration of different offsets may affect the outcome of the FWT – especially for unfavorable choices of initial models. Within a multi-stage inversion, offset windowing can be applied to improve the performance of the FWT. As a direct consequence of the observation in experiment I, the application of FWT to cross-well data requires a sequential windowing from near to far offsets. In the current experiment the FWT is divided into three stages with different offset ranges given by Table 4.3. The total amount of 100 iterations per FWT test is equally split into (34, 33, 33) iterations.

Figures 4.8a,b,c,d illustrate the observed data with offset windows, the final data, residuals and the inverted velocity model in case of the initial test model ②④. In comparison to the corresponding result of experiment I (see Figure 4.5), the v_P model shows significant improvements. The model error decreases from 10.8 % to 6.3 %, while it increases from 10.8 % to 11.2 % in case of experiment I. Especially the area around the anomaly is reconstructed less artificially. Furthermore, the fit between final synthetic data and observed data is slightly better (compare final synthetic and residual seismograms in Figures 4.8b,c and 4.6c,d).

However, the abovementioned "model window" of successful initial models is broadened insignificantly, i. e., the FWT including offset windowing is slightly less sensitive to the choice of initial background models (see Figure D.2). Consequently, offset windowing is useful to avoid cycle skipping – with respect to large offsets – at the beginning of the inversion.
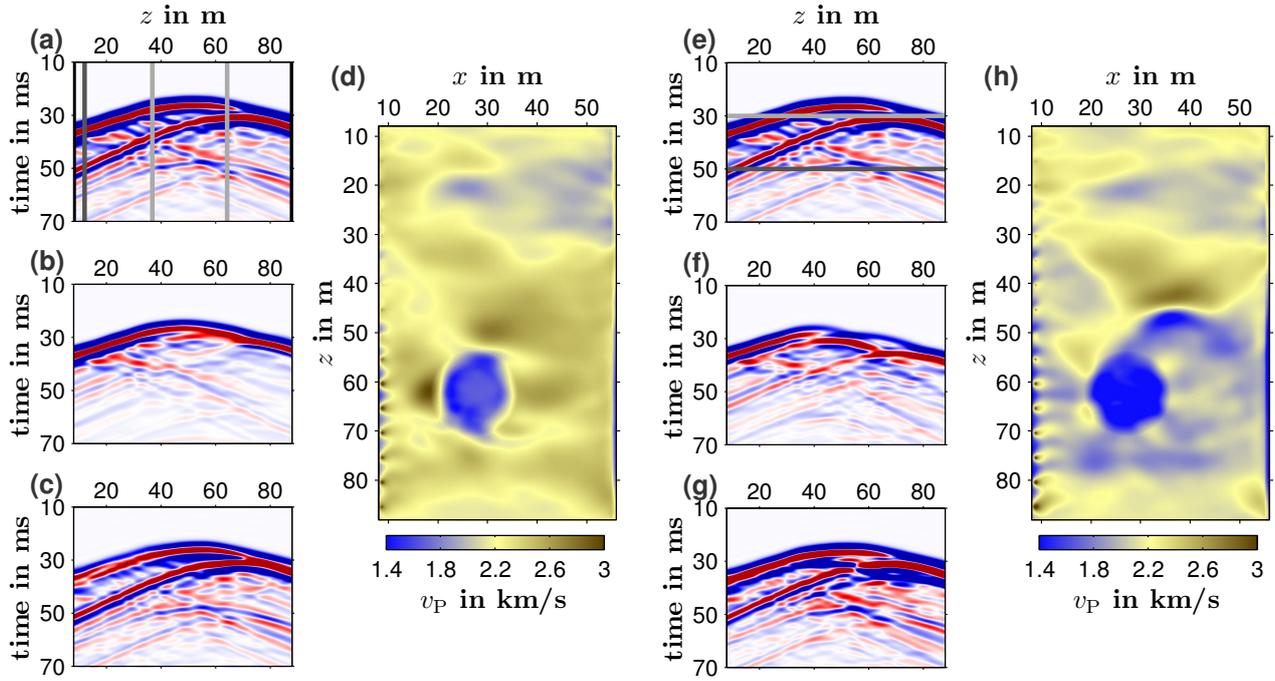
Figure 4.8: Application 1, experiment II/1: Seismograms and v_P results for application of offset windowing (a ... d) and time windowing (e ... h). In detail, the seismograms show observed data (a, e), final synthetic data (b, f) and final residuals (c, g) for the central shot 9. All data amplitudes are comparable. Gray-colored lines indicate offset and time windows, respectively.

Additionally, this experiment investigates the impact of a simple time windowing on the model reconstruction. On the analogy of offset windowing, the same multi stage inversion with time windows according to Table 4.3 is applied.

Figures 4.8e,f,g,h depict the observed data with time windows, the final data, residuals and the inverted velocity model in case of the initial test model ㉔. Obviously, both final seismograms and inverted v_P model are very similar to the result of reference experiment I. Thus, in case of the given cross-well geometry, the application of time windowing does not improve the performance of the FWT, which is emphasized by Figure D.3. Compared with Figure D.1, the initial "model window" is even narrower. Rather, it is shifted to higher velocities, so that the critical choice of model ㉔ results in an artificially recovered v_P model (evolution of model error: 10.8 % → 13 %), while a satisfactory inversion result is obtained from model ㉜ (model error: 7.2 % → 4.1 %).

Experiment II/2: Frequency filtering

The inversion with respect to single frequencies or several frequency ranges is common practice in FWT applications (e. g., Bunks et al., 1995; Sirgue and Pratt, 2004). The aim is to reduce the ambiguity of the inverse problem and to prevent the FWT getting stuck in local minima of the data misfit function. The general strategy is to start at low frequencies related to large-scale structures of the model and to move to higher frequencies related to small-scale structures.

In this experiment, a low-pass filter is used to reduce the frequency content from the peak frequency $f_{peak} = 200$ Hz (full content) to $f_{peak} = 95$ Hz. It is applied to the source time function (Figure 4.3) and the observed data (Figure 4.9a). Again, initial model ㉔ is utilized to demonstrate the effect of frequency filtering. In contrast to full-content data in experiment I, the observed data is dominated by the direct wave. Consequently, the conformity of filtered observed data and initial synthetic data (Figure 4.9b) is already quite high. Thus, at early iterations the FWT is able to correct the wrong background velocity of the initial model. Due to the limited frequency content, only medium-scale structures are added later on, which results in the final model (Figure 4.9e). There is a high similarity between observed data and final synthetic data (Figure 4.9c) yielding in very low residuals (Figure 4.9d).

In case of choosing initial model ㉔, the data misfit is reduced by more than two orders of magnitude – contrasting to experiment I. The model error decreases from 10.8 % to 3.9 %. In spite of removing high-frequency contents and neglecting the recovery of small-scale structures, the allover minimum model error of all 51 tests is 3.7 %. This is even better
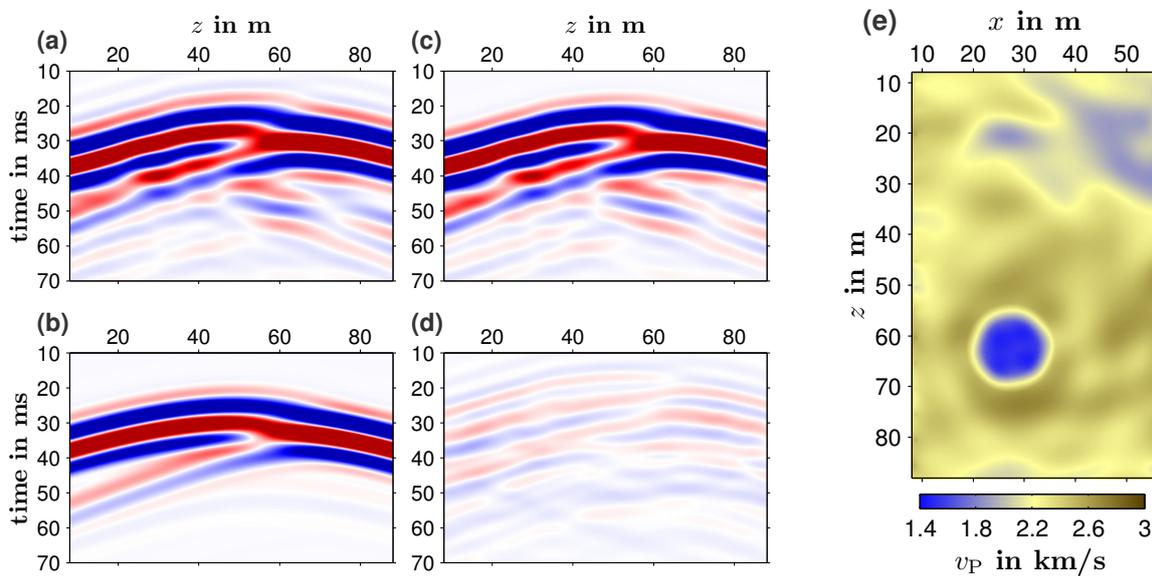
Figure 4.9: Application 1, experiment II/2: Seismograms for the central shot 9 (a . . . d) and v_P result (e) in case of applying frequency-filtering. (a) to (d) show the filtered observed data, initial synthetic data, the final synthetic data and the final residuals.
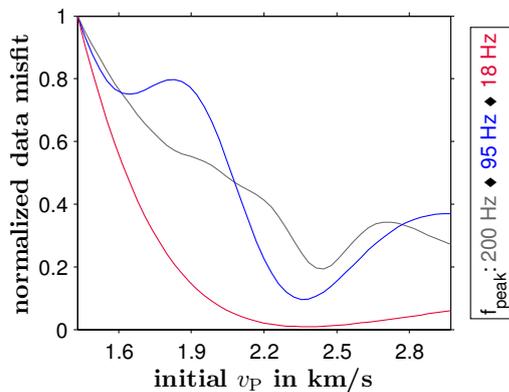
Figure 4.10: Application 1, experiment II/2: Comparison of data misfit functions for application of low-pass filters with different frequency ranges. For all 51 background velocities, the misfit values are obtained from the initial model. All graphs are normalized individually.

than the minimum value of experiment I (3.8 %). Although, a dominant wavelength of approximately 25 m is derived from the average v_P and the peak frequency, structures at sub-wavelength scales are recovered. For example, the FWT reconstructed structures at length scales of roughly half a wavelength, such as details of the low velocity zone.

Due to frequency filtering, the FWT is much less sensitive to the choice of initial models. For example, in case of model ㉔, the FWT not only performs successfully, but it also shows a much stronger speed of convergence. The "model window" allowing the computation of a satisfactory v_P result is broadened tremendously (see Figure D.4). It ranges from initial model ⑱ ($v_P = 1954\,\frac{m}{s}$) to initial model ㊸ ($v_P = 2724\,\frac{m}{s}$).

Apart from choosing a certain frequency range with $f_{peak} = 95$ Hz, Figure 4.10 demonstrates the effect of applying low-pass filters with respect to different peak frequencies. The comparison comprises the impact of $f_{peak} = (18, 95, 200)$ Hz on the data misfit of the initial models. Especially the efficiency of search algorithms, such as a bisection method as mentioned in section 4.2.1, are affected by the shape of the initial misfit function. The aim is to find the optimal initial model. However, in case of using the full frequency content ($f_{peak} = 200$ Hz), the ambiguity of the inverse problem seems to be too high. The FWT results in an artificially recovered v_P model (see section 4.2.1). On the one hand, a quite complex misfit function is still obtained by choosing an intermediate frequency range with ($f_{peak} = 95$ Hz). On the other hand, its minimum is located within the "model window" of promising initial models (see Figure D.4). An application of a strong low-pass filter ($f_{peak} = 18$ Hz) significantly mitigates the ambiguity of the inverse problem.

Experiment II/3: Frequency filtering over multiple stages

As mentioned in the previous section, the application of frequency-filtering is extended to a sequential procedure over three stages. The associated peak frequencies are $f_{peak} = (18, 95, 200)$ Hz. On the analogy of experiment II/1, the available amount of 100
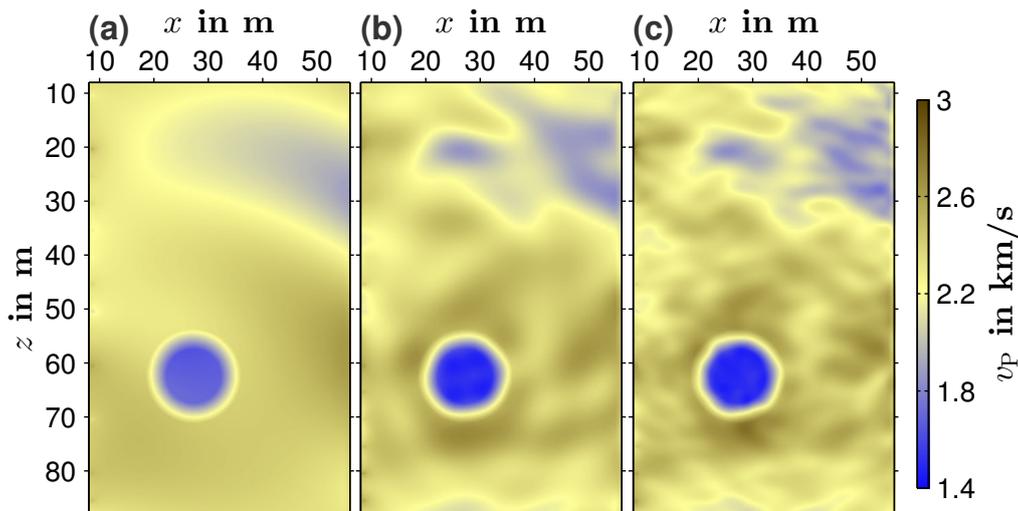
Figure 4.11: Application 1, experiment II/3: Inverted v_P models computed by a FWT with frequency filtering over multiple stages. The initial model is given by model ㉔. The intermediate models are obtained after the first and second stage (a, b), while the final result is shown in (c).

iterations per FWT test is equally split into (34, 33, 33) iterations. Figure 4.11 illustrates inverted v_P models with respect to the choice of initial model [24]. Figures 4.11a,b represent intermediate results after the first and second stage, while Figure 4.11c depicts the final v_P model. The sequential model reconstruction, beginning with large-scale structures and ending with small-scaled details, can be clearly seen. The comparison of the final result with the true model (Figure 4.1a) reveals a good concordance of structures down to half a wavelength which is approximately 11 m.

Due to the inversion for different frequency ranges, the comparability of data misfits among the stages is difficult. Hence, in the following paragraphs only model errors are discussed. In comparison with the single-stage FWT in experiment II/2, the multi-stage method applied to initial model [24] shows a slightly better performance (evolution of model error: 10.8 % → 3.7 %).

The application of the multi-stage inversion further reduces the sensitivity of the FWT to the choice of the initial model. The initial "model window" allowing a successful FWT extends over a wide range of background velocities (see Figure D.5). Even the choice of model [10] ($v_P = 1707 \frac{m}{s}$) results in a satisfactory v_P model, where the model error is significantly reduced from 26.8 % to 4.7 %. Additionally, the "model window" exceeds the upper limit of the initial-velocity range ($v_P = 2970 \frac{m}{s}$). Again, this indicates that – in case of given model and geometry – the successful FWT prefers initial models with higher velocities. In contrast to all previous experiments, the model error for all 51 initial models decreases during the inversion. Probably, the FWT might also succeed using low-velocity models, but only in conjunction with a poor convergence.

Comparative summary of multi-stage FWTs in experiments II/1 and II/3

Figure 4.12 summarizes the performance of multi-stage methods, such as offset windowing, time windowing or frequency filtering. Obviously, its efficiency depends on the initial model. For all methods, model [28] represents an appropriate choice with respect to the outcome of the FWT (see Figures D.2, D.3 and D.5). However, the application of windowing in conjunction with models [24] and [39] is characterized by unsuccessful FWTs. In any case, the application of frequency filtering reveals satisfactory results.

Apparently, FWTs including frequency filtering show the most stable performance (Figures 4.12c,f,i). During all three stages, the misfit is reduced significantly. However, no stage has the precedence with respect to the strength of reduction. For all initial models, the strongest misfit reduction is obtained in stages one and two involving lower-frequency contents.

In case of model [28], the application of excessive windowing seems to be unnecessary. The strongest misfit reduction is observed in stages one and two (Figures 4.12d,e). Thus, the combination of a very good initial model and a sufficient fit of near offsets or the direct wave, respectively, is enough to recover the v_P model properly. Especially in case of time windowing the contribution of the last stage is negligible. In contrast, it is even possible to increase the data misfit. Additionally, the alternating misfit plots indicate difficulties in minimizing the data misfit (due to the choice of a constant step length).

The combination of suboptimal initial models and offset windowing reveals a very poor misfit reduction (Figures 4.12a,d,g). On the one hand, this might indicate that the inversion gets stuck in a local minimum of the misfit function. On the other hand, it is possible to
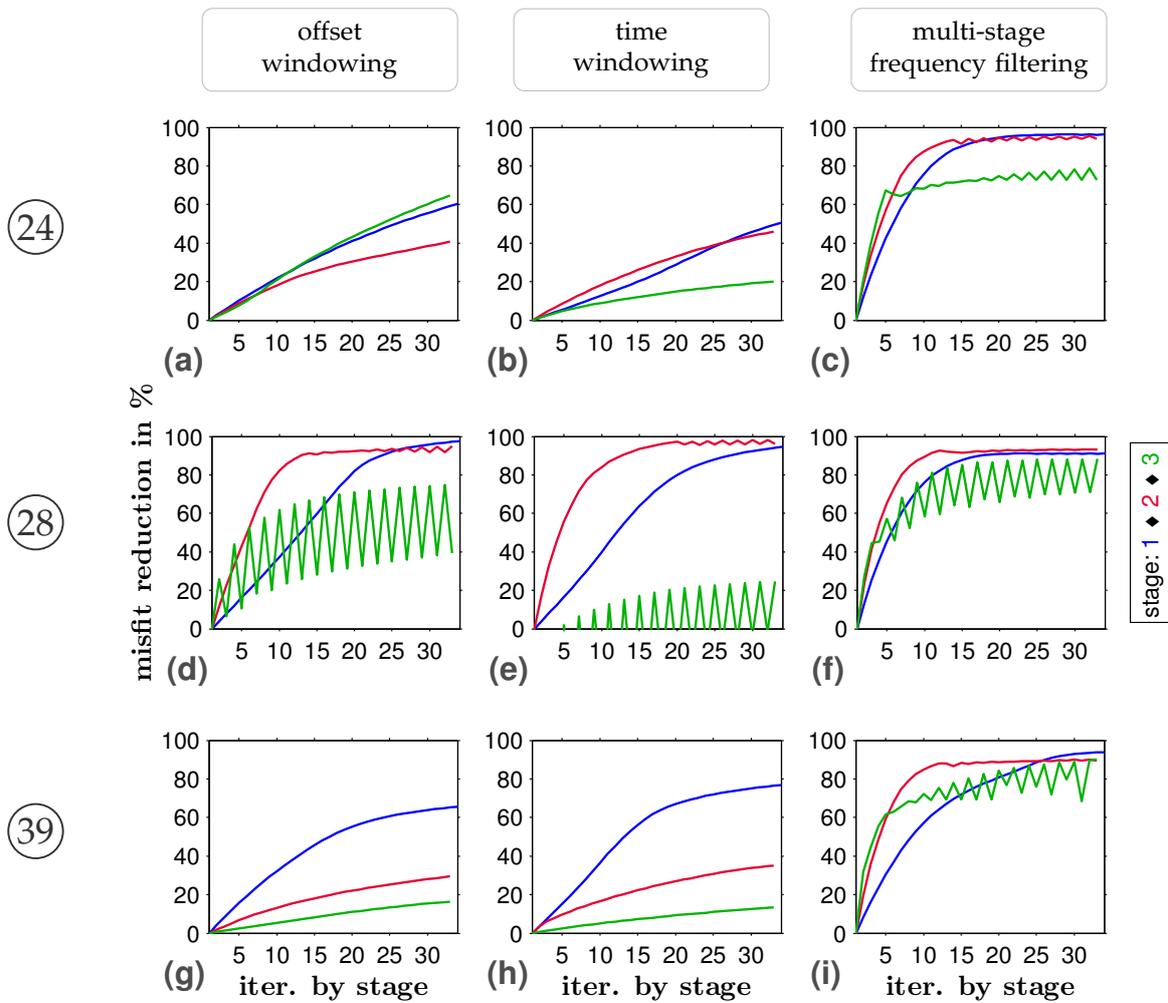
Figure 4.12: Application 1, experiment II: Data-misfit reduction for different combinations of initial models (rows) and multi-stage methods (columns). FWTs including offset windowing, time windowing and frequency filtering are applied to initial models ㉔, ㉘ and ㊴. (a) to (i) show the misfit reduction obtained within each stage and with respect to the first iteration of each stage.

obtain a satisfactory result after a huge number of iterations due to a very weak convergence. The misfit reduction can be used as a measure to evaluate the performance of the FWT. Only a few combinations of initial model and multi-stage method prove to be useful, whereas other choices are unable to improve the inversion result.

4.2.3 Experiment III: Gradient preconditioning

Experiment III involves gradient computations. Exploiting the exemplary initial model ㉔, it investigates the effect of three preconditioning methods on the reconstruction of the v_P model: application of circular gradient tapers around source and receiver locations (experiment III/1, Figure 4.13a), wavefield-based preconditioning (experiment III/2, Figure 4.13b) and application of user-defined tapers (experiment III/3, Figure 4.13c). All preconditioning methods are applied to the raw gradient **g**, i.e., the direct output of cross-correlation (2.17a).
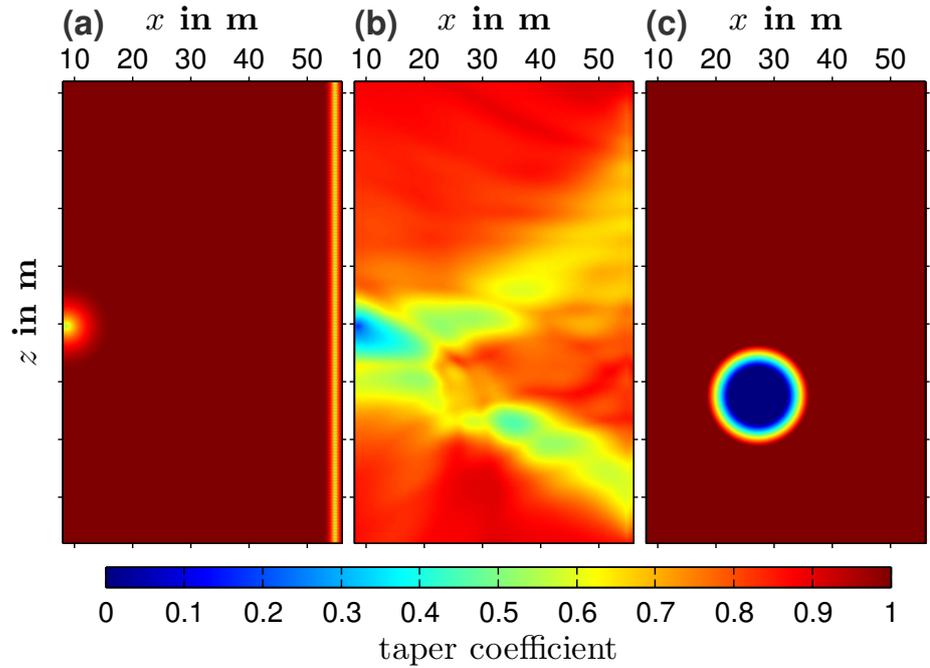
Figure 4.13: Application 1, experiment III: Taper matrices for gradient preconditioning in case of choosing initial model ㉔. (a) shows the circular taper geometry around sources and receivers with respect to the central source 9. (b) illustrates the preconditioning matrix obtained from wavefield propagation at the first iteration. (c) shows a user-defined taper used to nullify the gradient within the circular anomaly.
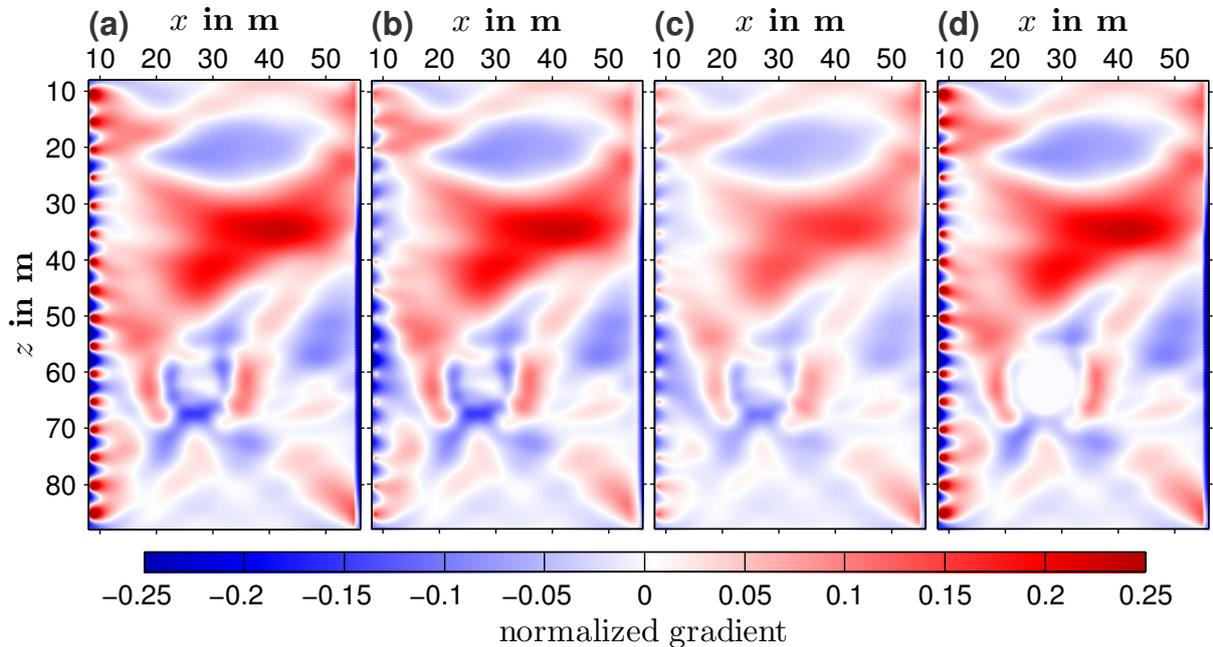
Figure 4.14: Application 1, experiment III: Gradients at the first iteration in case of choosing initial model ㉔. (a) depicts the raw gradient obtained from the imaging condition. (b) to (d) correspond to the tapers in Figure 4.13 and represent preconditioned gradients with a separate application of circular tapers around sources and receivers (b), a wavefield-based taper (c) or a user-defined taper (d).

Figure 4.14a illustrates the raw gradient at the first iteration. Apart from features to correct the interior of the initial velocity model, it is characterized by strong artifacts around the sources. In particular, this complicates the model update in equation (2.35). Due to the gradient normalization to max $|\mathbf{g}|$, a meaningful relative step length is not calculable. A huge step length would be necessary to enhance model-related structures, resulting in a remarkable amplification of the source artifacts. Furthermore, artifacts with smaller amplitudes appear around receiver locations, too. However, for the given example they are hardly observable. Due to the very dense receiver array, artifacts with alternating positive and negative amplitudes annihilate caused by interference.

The circular tapers around the acquisition geometry are obtained by (2.19). Due to stronger artifacts around the sources, different specifications for cosine-shaped functions are defined to distinguish between sources and receivers. At source locations, the parameters for equation (2.19) are: $R = 8$ m, $f(r{=}0) = 0.5$ and $f(r{=}\frac{R}{2}) = 0.9$. For receiver locations, the taper size is reduced to $R = 2$ m, while the remaining coefficients are identical. Within the loop over shots (see inversion scheme in Figure 2.1), the source-receiver taper is separately computed for each source and all receivers. Figure 4.13a shows the taper for a representative shot. The preconditioned gradient in Figure 4.14b exhibits significantly damped source artifacts, while around receiver locations, the gradient is modified inadequately.

The wavefield-based taper is computed by equation (2.25). The user-defined coefficient amounts to $C_{\mathrm{stab}} = 0.15$. The resulting preconditioned gradient for the central shot is depicted in Figure 4.14c. In comparison with Figure 4.14b, the suppression of source artifacts is stronger, while there are only minor changes around receiver locations. However, the taper also affects the model interior. A partial mitigation of this problem can be achieved by adjusting the coefficient $C_{\mathrm{stab}} = 0.15$. Furthermore, the amplitudes of acquisition-geometry artifacts vary during the inversion. Hence, the wavefield-based taper – recomputed for every shot at each iteration – seems to be advantageous over tapers with a fixed circular shape.

The user-defined taper represents a hard constraint (see Figure 4.13c). On the one hand, it is used to weight the gradient, i.e., to damp or amplify amplitudes. On the other hand, selected model areas can be excluded from the model update by nullifying corresponding parts of the gradient. In this experiment, the circular water-filled anomaly is assumed to be known. The taper is used to prevent model updates within this area (see Figure 4.14d).

Although, the tapers reduce the strength of geometry-related artifacts, their exclusive application does not improve the recovered v_P models. The inverted models in Figures D.6, D.7 and D.8 for the initial model [24] show a high similarity with the reference result of experiment I (Figure 4.5). The corresponding data misfits are reduced by less than one order of magnitude, while the model errors decreased slightly or increased from 10.8 % up to 12.6 %. Furthermore, there are nearly no differences between the best possible results of experiments I and III. Consequently, gradient preconditioning can only be useful in conjunction with other methods discussed in this chapter.

4.2.4 Experiment IV: Step length estimation

Experiment IV investigates the impact of the choice of step length on the model update. In particular, it comprises a comparison of experiment I featuring a constant step length

and the application of the adaptive step length (experiment IV/1). Additionally, different choices of both constant (IV/2) or adaptive step lengths (IV/3) are analyzed. Experiment IV/4 investigates the robustness of the adaptive step length method.

Experiment IV/1: Adaptive step length

In contrast to experiment I, experiment IV/1 involves the adaptive step length method described in section 2.6 by following the step-length computations described in section 2.6.1. The initial relative step length is $\mu_{\text{rel,ini}} = 1\,\%$ (see Table 4.4) and the remaining test step lengths are obtained by applying the coefficient $a = 8$ resulting in $\mu_{\text{rel,low}} = \frac{1}{8}$ and $\mu_{\text{rel,high}} = 8$. The optimal step length $\mu_{\text{rel,opt}}$ is defined as $\mu_{\text{rel,ini}}$ at next iteration. In this experiment, it is not allowed to calculate step lengths larger than $\mu_{\text{rel,high}}$. To avoid computations of excessively high values, the overall upper limit is set to 8 %. Furthermore, additional forward-propagations are required to obtain the data misfits for all test step lengths. The FWT uses eight representative sources to reduce computational efforts.

Figure D.9 illustrates the performance of multiple FWTs including the evolution of adaptive step lengths. This experiment reveals some interesting observations:

- The adaptive step length does not improve the sensitivity of the FWT with respect to the initial model. A successful inversion is still limited to the same narrow "model window" of models shown in experiment I.

- Within the "model window" the data misfit is characterized by both a drastic reduction and a stable evolution without an alternating shape. The maximum reduction of the data misfit amounts to 3.2 orders of magnitudes (compared to 2.5 orders of magnitude in experiment I).

- Accordingly, the model-error evolution shows a similar behavior. The allover minimum model error is 3.4 %.

- The adaptive step lengths vary over a broad range of approximately five orders of magnitude to account for the complex shape of the data misfit function.

- In dependence of the initial-model choice, the algorithm is more or less able to compute reliable relative step lengths. Within the "model window" there appears a typical inversion progress starting with quite high step lengths and ending with intermediate or low values. In general, the average step length is less than 1 % (compare Pica (1990) who proposed a meaningful step length of about 1 %). Permanently huge or tiny step lengths indicate a failure of the FWT. In this experiment, this is related to too high or too low initial background velocities. On the one hand, the algorithm tries to provide negative step lengths which are handled as described in section 2.6.2. On the other hand, too high values, such as exceptionally huge step lengths up to 47000 %, are truncated as mentioned above.

- In conjunction with the very strong convergence, the model reconstruction is improved significantly. In contrast to experiment I, within the limit of 100 iterations much more small-scaled details can be recovered.
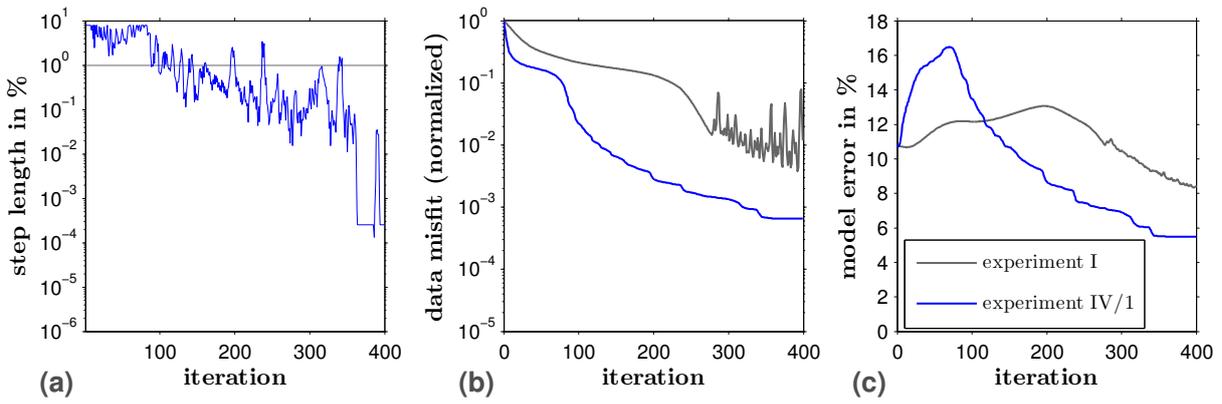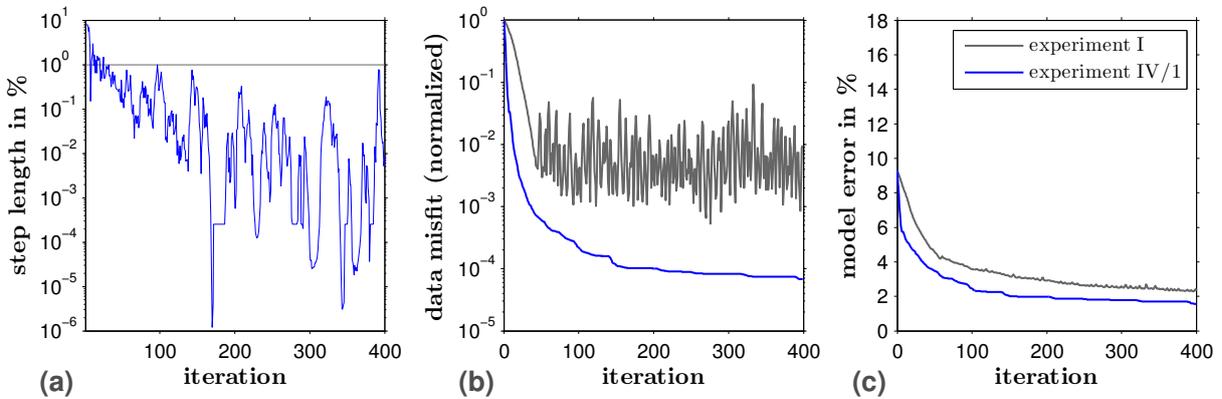
Figure 4.15: Application 1, experiment IV/1: Inversion progress for experiments I (constant step length) and experiment IV/1 (adaptive step length) for initial model $\overset{24}{\bigcirc}$ ($v_\mathrm{P} = 2138 \; \frac{\mathrm{m}}{\mathrm{s}}$). The individual figures contain the evolutions of the relative step length (a), normalized data misfit (b) and model error (c).

Figure 4.16: Application 1, experiment IV/1: Inversion progress for experiments I (constant step length) and experiment IV/1 (adaptive step length) for initial model $\overset{26}{\bigcirc}$ ($v_\mathrm{P} = 2200 \; \frac{\mathrm{m}}{\mathrm{s}}$). The individual figures contain the evolutions of the relative step length (a), normalized data misfit (b) and model error (c).

Furthermore, the FWTs of experiment IV/1 and I are repeated for the exemplary initial model $\overset{24}{\bigcirc}$ ($v_\mathrm{P} = 2138 \; \frac{\mathrm{m}}{\mathrm{s}}$) and a more optimal initial model $\overset{26}{\bigcirc}$ ($v_\mathrm{P} = 2200 \; \frac{\mathrm{m}}{\mathrm{s}}$). The number of iterations is extended to 400. The inversion progresses are illustrated in Figures 4.15 and 4.16 (graphs in blue for experiment IV/1 and gray for experiment I).

As mentioned above, the model located outside the optimal "model window" causes a specific inversion progress. The dynamic range of step lengths is small, i. e., the FWT permanently uses quite high values (see Figure 4.15a). However, this does not result in a strong misfit or model-error reduction (Figures 4.15b,c). In contrast, the data misfit decreases by approximately three orders of magnitude, while the model error shows an unsteady evolution resulting in a final error of 5.5 %. The inversion using model $\overset{26}{\bigcirc}$ is characterized by a continuous decrease of the adaptive step lengths covering a broad dynamic range of almost seven orders of magnitude (see Figure 4.16a). This accounts for the increasing complexity of the data misfit function. It is robustly reduced by approximately four orders of magnitude (Figure 4.16b). In this case, a nearly optimal-fit model is found and the FWT is only recovering very small scaled details. The final model error is to 1.6 % (Figure 4.16c).

Experiments I and IV/1 reveal the significant advantages of the adaptive step length method compared to constant step lengths. In case of a constant step length, the data misfit evolution shows an erratic behavior. It is not able to account for the shape of the misfit function by computing appropriate gradient directions. Consequently, the model error decreases very slowly – compared to experiment IV. The minimum model error of experiment I (2.3 %) is obtained after 400 iterations, while the same value of experiment IV/1 corresponds to iteration 105. The method of adaptive step length significantly increases the convergence of the FWT. In spite of accepting additional forward modelings, the computational efforts are reduced tremendously.

Experiments IV/2 and IV/3: Comparison of constant and adaptive step lengths

In addition to experiment IV/1, experiments IV/2 and IV/3 perform further investigations related to different constant step lengths and different parameters for the parabolic fit of the adaptive step length method (see discussion of different choices of $\mu_{\mathrm{rel,ini}}$, $\mu_{\mathrm{rel,low}}$ and $\mu_{\mathrm{rel,high}}$ in section 2.6.2). All tests are applied to the optimal initial model (26).

Figure 4.17 illustrates the inversion progress for three different constant step lengths. The choice $\mu_{\mathrm{rel}|h} = 1$ % resembles the computation discussed in the previous paragraph and shown in Figure 4.16 (gray graph). While the usage of a smaller step length $\mu_{\mathrm{rel}|h} = \frac{1}{4}$ % stalls the erratic behavior of the misfit evolution, a high value of $\mu_{\mathrm{rel}|h} = 4$ % does not allow a reliable inversion (Figure 4.17b). Until iteration 70 the step length seems to be sufficiently low. However, later on it causes an irreversible jump within the model space producing an artificially altered v_{P} model (see increasing model error in Figure 4.17c). For the current example, the choice of $\mu_{\mathrm{rel}|h} = \frac{1}{4}$ % yields the most stable progress. On the one hand, the convergence of the model error is quite weak. But on the other hand, similar v_{P} results are obtained from FWTs with $\mu_{\mathrm{rel}|h} = 1$ % and $\mu_{\mathrm{rel}|h} = \frac{1}{4}$ %.

Figure 4.18 shows the inversion progress for the adaptive step length method with three different sets of test step lengths. The step lengths $\mu_{\mathrm{rel,low}}$ and $\mu_{\mathrm{rel,high}}$ are obtained by the use of $\mu_{\mathrm{rel,ini}} = 1$ % and $a = (2, 4, 8)$. Even the choice of $\mu_{\mathrm{rel}|h} = \left(\frac{1}{8}, 1, 8\right)$ % allows the calculation of a meaningful optimal step length (as illustrated by 2.5). Obviously, in case of initial model (26), the inverse problem is quite well-posed.

Experiment IV/4: Computational efforts for adaptive step lengths

This experiment is related to the adaptive step length method. As described above, it is necessary to apply additional forward-propagations to compute data misfits (E_1, E_2, E_3) corresponding to the test step lengths $(\mu_{\mathrm{rel,ini}}, \mu_{\mathrm{rel,low}}, \mu_{\mathrm{rel,high}})$. In order to limit computational efforts, it is not advisable to use all sources $(N_{\mathrm{s}} = 16)$ for test modelings. This would result in $3 \times N_{\mathrm{s}}$ additional modelings, compared to $2 \times N_{\mathrm{s}}$ modelings for forward-propagation and back-propagation. The question is, how to define a reasonable number of sources to enable a stable inversion.

Previous experiments IV/1 and IV/3 applied a default value of 8 sources. This experiment investigates the impact of using 1, 4 and all 16 sources on the inversion progress (see Figure 4.19). Apparently, with decreasing number of sources the adaptive step length estimation tends to fail finding a proper optimal step length $\mu_{\mathrm{rel,opt}}$ (see Figure 4.19a).

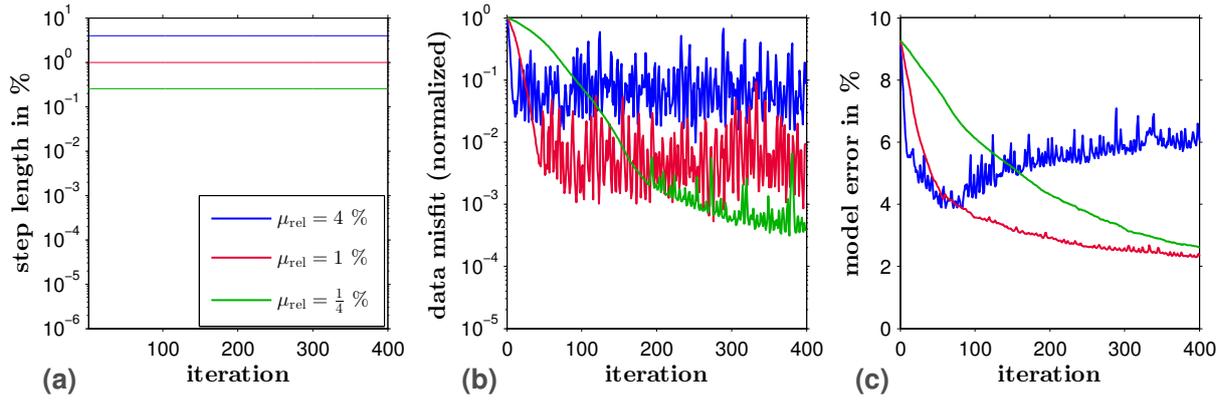**Figure 4.17:** Application 1, experiment IV/2: Inversion progress for different constant step lengths in case of using the optimal initial model $\textcircled{26}$. The individual figures contain the evolutions of the relative step length (a), normalized data misfit (b) and model error (c).
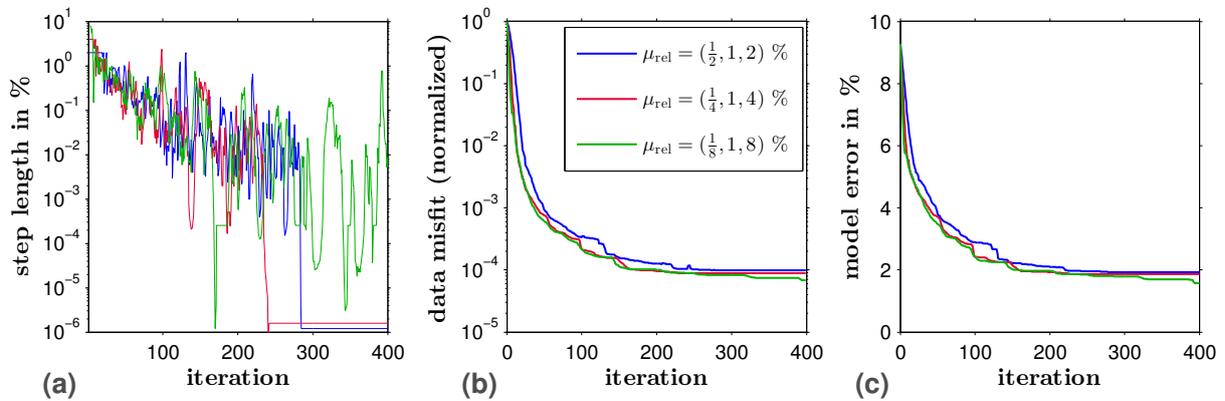

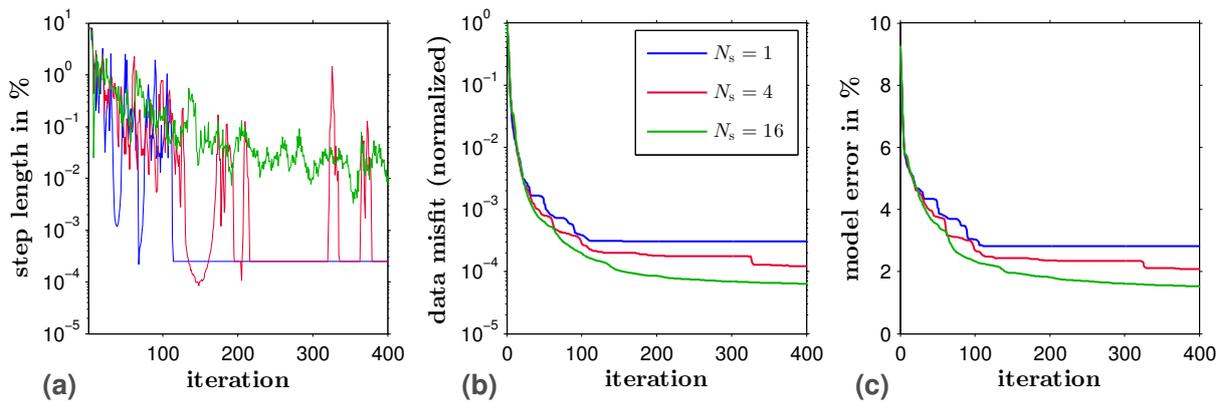**Figure 4.18:** Application 1, experiment IV/3: Inversion progress for the adaptive step length method with different sets of test step lengths (in case of using the optimal initial model $\textcircled{26}$). The individual figures contain the evolutions of the relative step length (a), normalized data misfit (b) and model error (c).

Figure 4.19: Application 1, experiment IV/4: Inversion progress for the adaptive step length method with different numbers of sources for test modelings (in case of using the optimal initial model $\textcircled{26}$). The individual figures contain the evolutions of the relative step length (a), normalized data misfit (b) and model error (c).

The algorithm provides $\mu_{\mathrm{rel,opt}} < 0$ after 230 iterations (in case of $N_s = 1$) or 260 iterations ($N_s = 4$), respectively. The algorithm tries to handle the problem by applying small step lengths (see section 2.6.2). Finally, it gets stuck.

Obviously, the usage of $N_s = 1$ or $N_s = 4$ results in data misfits (E_1, E_2, E_3), which do not represent the data misfit of all sources. Consequently, the deviating shape of the misfit function affects the parabolic fit. Inappropriate "optimal" step lengths might cause a jump into a local minimum.

With increasing number of sources for test modelings, the model error is decreasing (see evolutions of data misfit and model error in Figures 4.19b,c). The final model errors are: $\epsilon_{\mathrm{model}} = 2.8\ \%$ ($N_s = 1$), $\epsilon_{\mathrm{model}} = 2.1\ \%$ ($N_s = 4$), $\epsilon_{\mathrm{model}} = 1.8\ \%$ ($N_s = 8$) and $\epsilon_{\mathrm{model}} = 1.5\ \%$ ($N_s = 16$). The trade-off between accuracy and computational efforts justifies the usage of $N_s = 8$. It is applied in all subsequent experiments including the adaptive step length method.

4.2.5 Experiment V: Influence of the acquisition geometry

Apart from the previous experiments, experiment V investigates different acquisition geometries. In particular, experiment V/1 applies a transmission geometry with variable number of sources (see Figure 4.2b). Experiment V/2 investigates the impact of a reflection geometry on the model reconstruction (Figure 4.2c). Several multiple FWTs are performed (results shown in Figures D.10 to D.16) and exemplary results are discussed using the initial model ㉔. Apart from varying the acquisition geometry, experiments V/1 and V/II are once again plain FWTs that do not apply any special inversion strategies.

Experiment V/1: Transmission geometry with a variable number of sources

It is well-known that a successful model reconstruction requires a sufficiently high illumination of the subsurface. While the illumination is improved by increasing numbers of sources (N_s) or receivers, only N_s affects the computational performance of a FWT. This experiment tests, if the FWT is able to get along with a poor illumination, or, if a very high number of sources improves the inversion result. Figure 4.20 compares the results of the extreme cases using $N_s = 1$ (Figures 4.20a to d) or $N_s = 64$ (Figures 4.20e to h).

Obviously, the inversion using $N_s = 1$ is not able to explain the data of a single shot by reconstructing a satisfactory v_P model (Figure 4.20d). After 100 iteration there are still significantly high data residuals (Figure 4.20c).

In comparison with the result of experiment I ($N_s = 16$), a higher illumination using $N_s = 64$ does not mitigate the ambiguity of the inverse problem or the dependency on the initial model. The final model shows some improvements with respect to details within the upper structures (compare Figures 4.6e and 4.20h), while there is a bad fit of final synthetic and observed data (Figures 4.20e,f). Consequently, a successful inversion is still restricted to the "model window" discussed in previous experiments.

In addition to the results of experiment I in Figure 4.5 ($N_s = 16$), Figures D.10 to D.15 (in appendix D.2) illustrate the multiple-FWT results for $N_s = 1$, $N_s = 2$, $N_s = 4$, $N_s = 8$, $N_s = 32$ and $N_s = 64$. On the one hand, in case of exemplary initial model ㉔, the increase

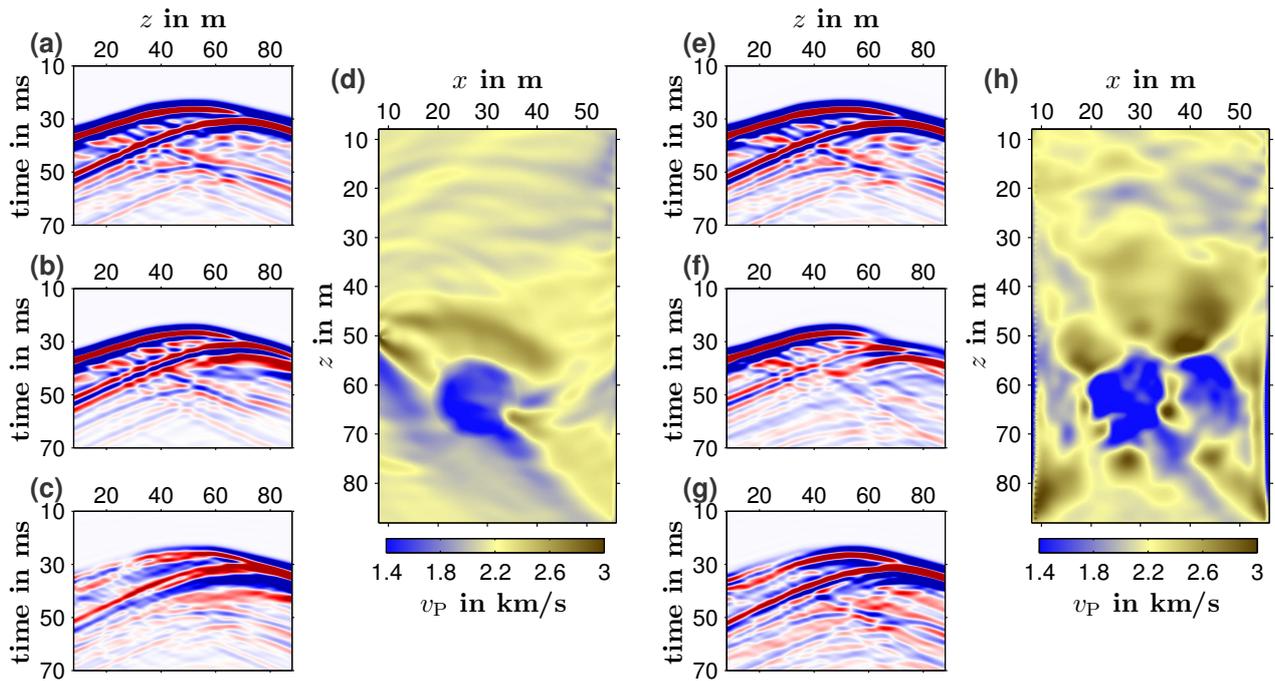**Figure 4.20:** Application 1, experiment V/1: Seismograms and models for FWTs using $N_s = 1$ (a to d) and $N_s = 64$ (e to h), respectively. All inversions are performed for the initial model ㉔. The figure comprises observed data (a and e), final synthetic data (b and f), final residuals (c and g) as well as the final v_P model (d and h).
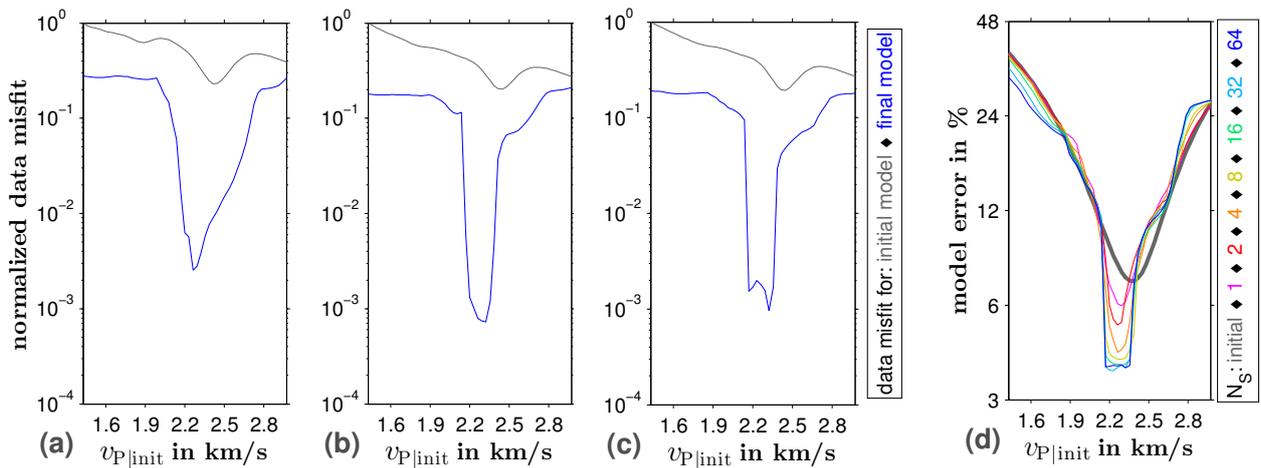

Figure 4.21: Application 1, experiment V/1: Initial and final data misfits (a to c) as well as model errors (d) with respect to inversions using different numbers of sources, N_s. While (a) to (c) are representative choices $N_s = 1$, $N_s = 8$ and $N_s = 64$, (d) summarizes model errors. Due to different numbers of sources, the data misfits of all computations are normalized individually.

of N_s improves the v_P recovery within the upper model areas only. On the other hand, there is a significant impact on the inverted v_P model in case of choosing the initial model ㉘. Provided that $N_s = 2$, it results in the corresponding optimal result (Figure D.11). In spite of a quite poor illumination, the choice of the initial model is crucial in computing a reasonable v_P model.

Additionally, Figures 4.21a to c visualize the initial and final data misfits as a function of all initial background velocities as well as for $N_s = 1$ (a), $N_s = 8$ (b) and $N_s = 64$ (c). The associated multiple-FWT Figures are D.10, D.13 and D.15. The shape of the initial data misfit is not affected by different choices of N_s. Only in case of $N_s = 1$, a local minimum appears at quite low initial velocities. Apparently, for $N_s = 1$ the "model window" of successful inversions seems to be broader. However, the corresponding misfit reduction is approximately 2.5 orders of magnitude. Furthermore, the FWTs for $N_s = 8$ and $N_s = 64$ exhibit comparable final data misfits. However, the usage of $N_s = 64$ reveals a stronger convergence (compare Figures D.13 and D.15).

Apart from the data misfit, the model error is a more instructive measure to describe the quality of the results. Especially within the "model window" one can observe: the higher N_s, the lower the final model error (Figure 4.21d). The optimum model errors amount to:

- $N_s = 1$: $\epsilon_{model} = 6.0\,\%$, see also Figure D.10;

- $N_s = 2$: $\epsilon_{model} = 5.2\,\%$, see also Figure D.11;

- $N_s = 4$: $\epsilon_{model} = 4.2\,\%$, see also Figure D.12;

- $N_s = 8$: $\epsilon_{model} = 4.0\,\%$, see also Figure D.13;

- $N_s = 16$: $\epsilon_{model} = 3.8\,\%$, see also Figure 4.5;

- $N_s = 32$: $\epsilon_{model} = 3.7\,\%$, see also Figure D.14;

- $N_s = 64$: $\epsilon_{model} = 3.8\,\%$, see also Figure D.15.

In comparison with the inversion result for $N_s = 8$, all FWTs with $N_s > 8$ are characterized by marginal model improvements. This justifies the choice of $N_s = 16$ in most of the experiments of this chapter, i.e., this avoids side effects caused by a poor ray coverage.

Experiment V/2: Reflection geometry

In contrast to all previous tests, this experiment uses a reflection geometry, where sources and receivers are located in one borehole (see Figure 4.2c). Here, it is regarded as a disadvantageous acquisition geometry. Due to the random medium geology, this is not a common reflection seismic application and it might be difficult to recover the subsurface structures. Furthermore, in comparison to the transmission geometry, a worse illumination complicates the reconstruction of the subsurface model. Again, the circular anomaly is assumed to be known. It is the only object being relevant for a reflection seismic application. On average, its diameter corresponds to 1.3 wavelengths (with respect to employing the full frequency content). Thus, it can be classified as a reflector or diffractor.
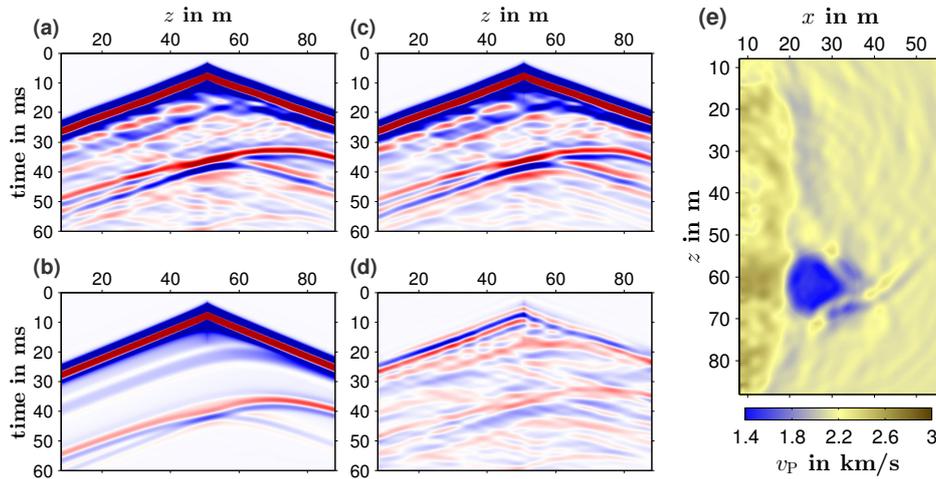
Figure 4.22: Application 1, experiment V/2: FWT result with respect to the reflection geometry and initial model ㉔. The seismograms in figures (a) to (d) are related to the central shot 9. Their amplitudes are comparable. The figures show the observed data (a), initial synthetic data (b), final synthetic data (c), final residuals (d) and the final v_P model (e).

The application of the FWT to the given reflection seismic problem reveals an unsatisfactory result (see Figure 4.22). Based on the exemplary initial model ㉔. The FWT is unable to reconstruct reasonable structures. The best model recovery can be found around the acquisition geometry (Figure 4.22e). However, even the direct waves of observed data and final synthetic data show a significant difference (compare Figures 4.22a,c,d).

The optimal inversion result is obtained in case of initial model ㉛ ($v_P = 2354\,\frac{m}{s}$). The model areas along the left boundary are recovered quite well, whereas the remaining model areas resemble the corresponding initial model (see Figure D.16). While the final model error amounts to 6.2 %, the data misfit is reduced by more than two orders of magnitude. The sufficient model reconstruction around the acquisition geometry yields a good fit of the dominant direct wave and thus explains the quite strong reduction of the data misfit.

Regarding the choice of model ㉛, a recomputation of the FWT with more iterations (400) does not reveal an improved result. The model error slightly decreased down to 6.1 %.

In comparison to experiments with acquisition geometry, the multiple FWT of experiment V/2 reveals some differences. On the one hand, the initial "model window", which is characterized by a significant misfit reduction, extends over a broader velocity range. But, on the other hand, it does not fully match the progress of the model error at higher initial velocities. Despite that, the inversion tends to prefer higher-velocity initial models to compute less artificial results. Furthermore, there is no clear delimitation of the "model window". For the given model, effects which impact negatively, such as cycle-skipping, are of minor importance.

4.2.6   Experiment VI: Optimal parameter configuration

This experiment completes the investigations of experiments I to V. Consequently, experiments VI/1 and VI/2 comprise an optimal parameter configuration. Due to the importance

of the acquisition geometry, they additionally compare different number of sources (VI/1: $N_s = 16$, VI/2: $N_s = 64$). The following configuration has been found to be useful:

- transmission geometry (see Figures 4.2a,b),

- frequency filtering over multiple stages (see section 4.2.2),

- gradient preconditioning by applying circular tapers at source and receiver locations (see experiment III/1),

- adaptive step length estimation (see experiment IV).

The most optimal result is obtained by experiment VI/1. Combining advantageous features, it is able to apply a successful FWT to almost all initial models (see Figure 4.23). Especially the adaptive step length estimation can be regarded as a reliable indicator for a stable inversion. Starting with quite high values, the inversion continues with intermediate step lengths confirming the results of experiment IV.

However, this experiment reveals a huge discrepancy between optimal initial model and the best possible result. Both the minimum initial data misfit and initial model error can be found for initial model ㉜ ($v_P = 2385 \frac{m}{s}$). In contrast, the allover minimum data misfit and minimum model error do not correspond to the same initial model. The minimum data misfit belongs to initial model ㉔ ($v_P = 2138 \frac{m}{s}$), while the usage of initial model ⑩ ($v_P = 1707 \frac{m}{s}$) yields the minimum model error. In detail, the initial and final model errors are:

- for initial model ㉜: 7.16 % $\rightarrow$ 3.49 %,

- for initial model ㉔: 10.8 % $\rightarrow$ 3.27 %,

- for initial model ⑩: 26.8 % $\rightarrow$ 3.15 %.

The optimum initial model does not necessarily correspond to the allover optimum inversion result. Apparently, this characteristic depends on the multiple-stage method. With increasing frequency content, the optimum choice of initial models tends towards lower initial velocities. With respect to the minimum data misfit within each stage, the optimal initial model moves from ㉜ (initial) to ㉚ (stage 1), ㉙ (stage 2) and ㉔ (stage 3 = final). Accordingly, the optimum initial model, corresponding to the minimum model error within each stage, moves from ㉜ to ㉓, ㉑ and ⑩.

Furthermore, for the optimal choice of model ⑩, the FWT has been recomputed. The number of iterations is increased up to 232 iterations. An automatic stop criterion is used to switch to the next stage or to stop the inversion. It is based on a threshold value of 0.5 % applied to the relative change of the data misfit at three successive iterations. Thus, the particular stages of the FWT require 36 (stage 1), 86 (stage 2) and 110 (stage 3) iterations. Apparently, the inversion for low-frequency contents reveals a much stronger convergence.

Although, additional features, such as the multiple-stage approach, improve the performance of the FWT, a modified acquisition geometry including 64 sources does not increase convergence or accuracy of the optimum result (Figure D.17). In contrast, during the first
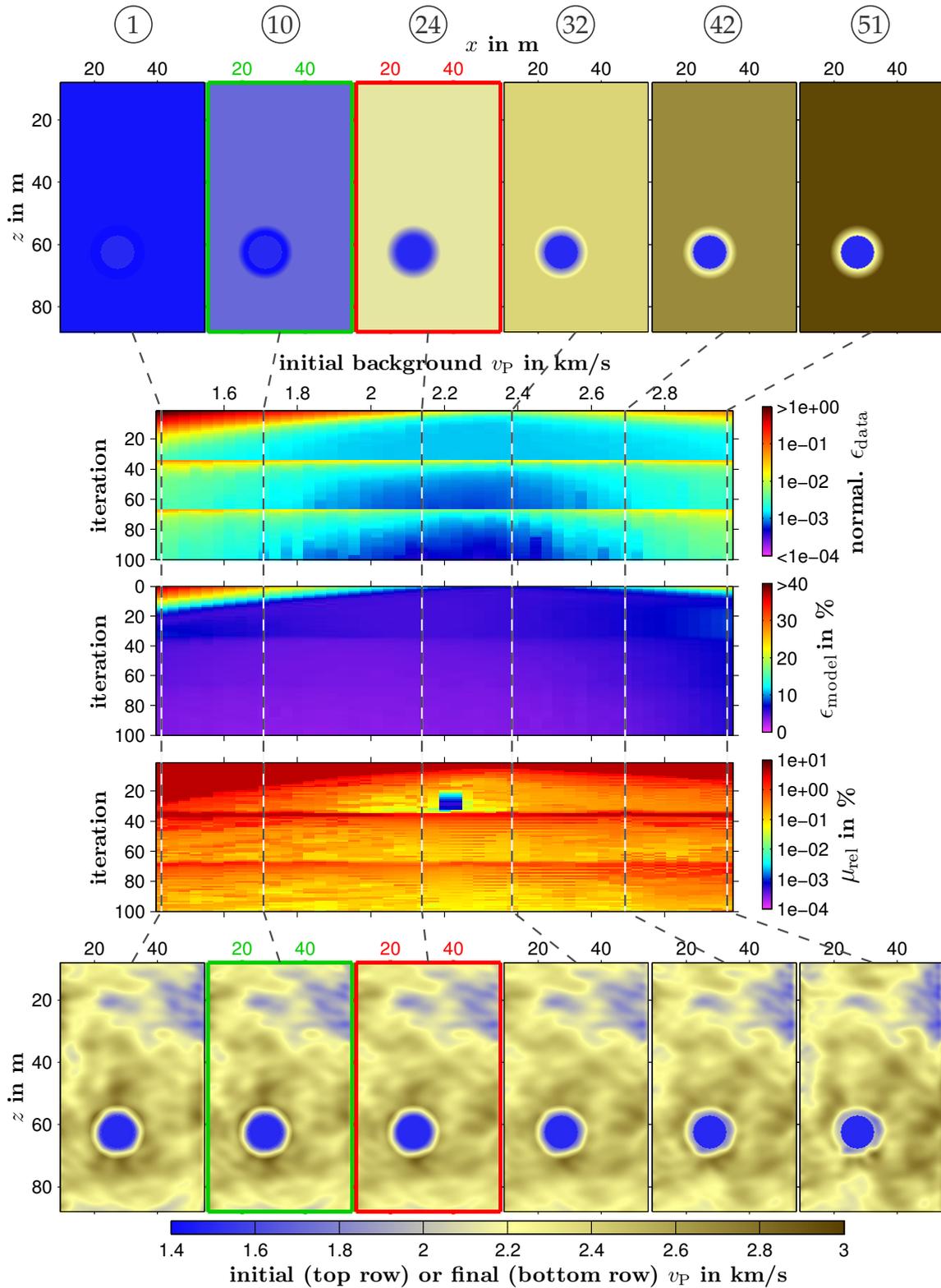
Figure 4.23: Application 1, experiment VI/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of an optimal parameter combination and "standard" transmission acquisition geometry including 16 sources. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 3.2)$ %.

stage, the inversion struggles to compute reasonable step lengths. In particular, at higher initial velocities, the model error is partially increasing. Consequently, the resulting v_P models are less accurate – compared to corresponding results of experiment VI/1. However, both experiment VI/1 and VI/2 are able to reconstruct nearly identical optimum v_P models.

4.2.7 Experiment VII: Rerun for the homogeneous initial model (type B)

While the previous experiments were applied to the initial-model type A, the experiment VII investigates the influence of a purely homogeneous initial model (type B, see Figure 4.1c) on the inversion progress. For this purpose, I choose inversion settings of selected experiments. In detail, I repeat following experiments:

VII/1:   • rerun of experiment I,
           • transmission geometry,
           • basic FWT without application of methodical improvements,

VII/2:   • rerun of experiment VI/1,
           • transmission geometry,
           • optimal choice of parameters,

VII/3:   • rerun of experiment I,
           • reflection geometry,
           • basic FWT,

VII/4:   • rerun of experiment VI/1,
           • reflection geometry,
           • optimal choice of parameters.

Experiments VII/1 and VII/2: transmission geometry

The inversion progress of experiment VII/1 Figure D.18 resembles experiment I (Figure 4.5). However, it reveals a slower convergence and the results in more artificial velocity models. The following list summarizes the most important observations.

- Surprisingly, in comparison with initial-model type A, the absence of the anomaly increases the absolute initial data misfit insignificantly (by 0.2 % only). Accordingly, the initial model error shows a similar behavior.

- Due to inversion of the full frequency content, the FWT tries to recover the high contrast of the anomaly. Especially outside the "model window" (between initial models ㉔ and ㉛) cycle-skipping might cause model corrections into a completely wrong direction.

- In spite of the comparable initial misfits and errors of experiments I and VII/1, the attempt of reconstructing both the anomaly and the remaining random-medium structures results in a less satisfactory performance.

- The optimum inversion result is obtained in case of initial model ㉙ ($v_P = 2292\ \frac{m}{s}$). The corresponding data misfit is reduced by two orders of magnitude, while the model error decreases from 8.6 % to 5.5 %.

The experiment VII/2 involves an optimal parameter combination including frequency filtering over multiple stages, gradient preconditioning and adaptive step length estimation. Figure D.19 shows the results of the multiple FWTs, while relevant observations are summarized below.

- Due to application of frequency filtering, the FWT first recovers long-wavelength informations of the anomaly. Later on, the high-contrast boundary of the anomaly is partially reconstructed by adding higher-frequency contents.

- Similarly to experiment VI, the broader range of initial models allows a successful FWT.

- The optimum inversion result is obtained in case of initial model ㉑ ($v_P = 2046\ \frac{m}{s}$). The corresponding model error decreases from 14 % to 4.7 %.

The final result of experiment VII/2 shows significant improvements in the reconstruction of the v_P model. However, it is unable to obtain a perfect recovery of the high-contrast anomaly. Thus, the initial model containing a priori information, such as the anomaly, significantly improves the outcome of the FWT.

Experiments VII/3 and VII/4: reflection geometry

Compared to experiment V/2 which investigates the influence of a reflection geometry using the initial model of type A, the choice of the simplest initial model amplifies the ill-posedness of the inverse problem. Apart from the direct wave, the initial data does not provide any reflection or diffraction events. Hence, all the information to reconstruct the v_P model has to be provided by the observed or residual data, respectively. Figure D.20 shows the the multiple-FWT results of experiment VII/3. The following list summarizes mentionable observations.

- Due to a plain inversion without any improvements (VII/3) and the choice of homogeneous initial models, the FWT is not able to reconstruct a satisfactory v_P model. Either the recovery is restricted to the area around the acquisition geometry trying to explain the direct wave, or it ends up with producing very strong artifacts (Figure D.20) by "misinterpreting" the observed data.

- The most optimal result is obtained for initial model ㉚ ($v_P = 2323\ \frac{m}{s}$). But, in contrast to experiment V/2, it represents a poor model reconstruction. The data misfit is reduced by 1.2 orders of magnitude (1.9 in experiment V/2), while the model error decreases from 8.3 % to 7.3 % (from 10.6 % to 5.25 % in experiment V/2).

- Using the initial model ㉚, I extended the FWT computation to 400 iterations. Compared to the optimal result of the multiple FWT with 100 iterations only, a slightly clearer indication of the anomaly can be observed. However, the model error of 7.2 % is insignificantly better.

Again, the ill-posedness can be mitigated by methods used in experiment VI/1 (optimal choice of parameters). Figure D.21 shows the multiple-FWT results of experiment VII/4 and relevant facts are listed below:

- On the analogy of experiment VII/2, in particular, the multi-stage approach drastically improves the performance of the FWT (see Figure D.21). The "model window" of low model errors verifies that the computation of quite reliable inversion results is less sensitive to the initial model (compared to experiment VII/3).

- The optimal result is obtained for initial model $^{\text{②⑦}}$ ($v_P = 2231 \, \frac{m}{s}$). In contrast to experiments V/2 and VII/3, it represents a sufficient model reconstruction. The model error decreases from 9.4 % to 4.6 %.

- The robust estimation of intermediate step lengths also indicates a stable inversion with respect to a wide range of initial models.

- Due to the slow convergence, especially initial models with $v_P > 1800 \, \frac{m}{s}$ show the potential of inverting a satisfactory v_P model after a high number of iterations. The corresponding FWT uses an automatic stop criterion and finishes after 133 iterations. However, there are no further improvements due to the limitations of the reflection geometry. Again, the final model error is 4.6 %.

4.2.8 Experiment VIII: Rerun for the initial model with wrong assumptions

Here, I continue to investigate the influence of the initial model. I chose the initial-model type C which is characterized by wrong assumptions (see Figure 4.1d). This model contains a circular anomaly, which is identical to the true anomaly in terms of size and velocity. But, it is located incorrectly. For the purpose of experiment VIII, I choose inversion settings of a selection of previously performed experiments I and VI. In detail, I repeat following experiments:

VIII/1:
- rerun of experiment I,
- transmission geometry,
- basic FWT without application of methodical improvements,

VIII/2:
- rerun of experiment VI/1,
- transmission geometry,
- optimal choice of parameters.

The usage of an initial model containing a wrong a priori information with regard to a high-contrast structure negatively affects the inversion progress. Figure D.22 shows the multiple-FWT result for a plain inversion of experiment VIII/1. The following list provides additional information.

- Obviously, for all initial models, the FWT is unable to move the incorrectly placed anomaly. Within a narrow range of initial models (from ㉓ to ㉛), the FWT produces the most reliable v_P models. While the upper parts are recovered well, the lower areas are severely affected by the anomaly.

- The optimal result is obtained for initial model ㉚ ($v_P = 2323 \frac{m}{s}$). Even in this case the inversion tries to compensate the wrong low-velocity zone by creating high-velocity artifacts.

- With respect to the optimal result, the data misfit is reduced by 1.3 orders of magnitude, while the model error decreases from 9.4 % (initial) to 7.4 % (final value at iteration 100). Due to the ill-posedness of the inverse problem, the minimum model error of 7.2 % is obtained at iteration 43. While the model error increases after iteration 43, the data misfit is reduced continuously.

In accordance with previous experiments, a more appropriate parameter configuration tremendously affects the outcome of the inversion. Figure D.23 shows the multiple-FWT result of experiment VIII/2, while additional facts are listed subsequently.

- Even the multi-stage inversion is not fully able to correct the location of the anomaly. Although, the model error is significantly reduced over a broad range of initial models, the smallest errors tend to occur at low-velocity initial models. In other words, the FWT prefers the usage of initial models, which do not exhibit incorrectly placed high-contrast structures.

- The optimum result is obtained for initial model ③ ($v_P = 1492 \frac{m}{s}$, i. e., it is nearly homogeneous).

- With respect to the optimal result, the data misfit is reduced by more than 2 orders of magnitude, while the model error drastically decreases from 36 % to 5.6 %.

On the one hand, methods, such as multi-stage frequency filtering, significantly mitigate the ambiguity of the inverse problem. On the other hand, experiment VIII demonstrates, that the presence of wrong a priori information might cause an insufficient performance of the FWT. Even the usage of the full frequency content is not enough to account for high-contrast interfaces.

4.2.9   Experiment IX: Initial model and FWT with wrong assumptions

Experiment IX is the logical continuation of experiment VIII/2. It uses the parameter configuration of the last experiment but adds the user-defined gradient preconditioning discussed in experiment III/3. In contrast to experiment VIII/2, the FWT is forced to omit the model update within the incorrectly located anomaly of the initial model (type C). The results are shown in Figure D.24 and important observations are listed below.

- Again, the application of features, such as the multi-stage approach, seems to annihilate the dependency of the inverted models on the initial model. But, in this case the inversion results in very similar artificial v_P models.

- Due to the application of a wrong constraint, the inversion is forced to compensate the forbidden model update by producing huge artifacts.

- The optimal result is obtained in case of the initial model ㊲ ($v_P = 2539 \, \frac{m}{s}$). The model error shows a marginal decrease from 11.9 % to 11.6 %.

- Within each stage the data misfit is reduced continuously, whereas the progress of the model error reveals an unusual behavior: 11.9 % (initial) $\rightarrow$ 7.6 % (minimum stage 1) $\rightarrow$ 8.1 % (minimum stage 2) $\rightarrow$ 10.5 % (minimum stage 3) $\rightarrow$ 11.6 % (final). At the early stage, the inversion reconstructs large-scale structures due to the application of frequency-filtering. At the lowest peak frequency $f_{peak} = 18 \, \text{Hz}$ of stage 1, the dominant wavelength is larger than 100 m. The inner diameter of the taper, where taper coefficients are equal to zero, amounts to 10 m. Consequently, due to the small-sized taper geometry, the FWT does not take the incorrectly applied constraint into account. After adding higher frequency contents and thus decreasing wavelengths, the negative impact of the taper on the model reconstruction is amplified continuously.

- Again, the adaptive step length proves to be a useful parameter. In general, during the inversion, the step length tends to decrease. However, in this experiment the evolution of the step length shows the opposite behavior. Furthermore, quite often the parabolic fit provides invalid values ranging from -460 % to 66000 %. This is an additional indication of severe problems during the inversion.

4.2.10   Experiment X: Brute-force search

Experiment X contrasts with all previous experiments. It does not perform complete inversions. Rather, it represents a brute-force search. In accordance with previous multiple FWTs, data misfits and model errors are computed for 51 initial models (homogeneous, type B). In contrast to the conventional estimation of adaptive step lengths using three test values, 1001 relative step lengths ranging from 0 % to 1000 % are evaluated at the first iteration.

This experiment investigates the frequency dependency of both the initial data misfit function and initial model errors. Figures 4.24 and 4.25 visualize cross sections of both data misfit functions and model error functions. With respect to the usage of the full frequency content or applying frequency filtering with peak frequencies $f_{peak} = (95, 18) \, \text{Hz}$, the data misfit functions reveal significant differences. With decreasing frequency content the shape of the data misfit function becomes less complex.

Around the optimal initial model (see previous experiments) and for low step lengths, the minimum misfit is clearly recognizable. However, with lower frequency contents it becomes more dominant and extends over a broad range of initial models verifying previous observations (compare top parts of Figures 4.24a,b,c). In principle, it is possible to achieve a tremendous reduction of the model error at the first iteration – if both the corresponding optimal initial model and an appropriate step length are known. Depending on the frequency content, the following optimal choices of initial model and step length are found at the first FWT iteration:

$$f_{\text{peak}} = 200\,\text{Hz}: \quad v_{\text{P|ini}} = \text{㉙}, \; \mu_{\text{rel|}h} = 17\,\% \; \rightarrow \; \epsilon_{\text{model}} = 7.5\,\% \;(\text{initial } \epsilon_{\text{model}} = 8.5\,\%),$$

$$f_{\text{peak}} = 95\,\text{Hz}: \quad v_{\text{P|ini}} = \text{㉘}, \; \mu_{\text{rel|}h} = 19\,\% \; \rightarrow \; \epsilon_{\text{model}} = 7.0\,\% \;(\text{initial } \epsilon_{\text{model}} = 8.9\,\%),$$

$$f_{\text{peak}} = 18\,\text{Hz}: \quad v_{\text{P|ini}} = \text{㉗}, \; \mu_{\text{rel|}h} = 13\,\% \; \rightarrow \; \epsilon_{\text{model}} = 7.3\,\% \;(\text{initial } \epsilon_{\text{model}} = 9.4\,\%),$$

where the initial models are ㉙ $= 2292\,\frac{\text{m}}{\text{s}}$, ㉘ $= 2262\,\frac{\text{m}}{\text{s}}$ and ㉗ $= 2231\,\frac{\text{m}}{\text{s}}$. Hence, the exclusive consideration of lower-frequency contents yields the highest efficiency in finding the best model. The strongest model-error reduction at the first iteration is obtained for $f_{\text{peak}} = 18\,\text{Hz}$. In all three cases, the data misfit and the model error show a quite good match with respect to the location of the local minimum (compare minima in Figures 4.26a,b,c). On account of this and due to the well-posed shape of the misfit functions, the automatic estimation of an adaptive step length is able to compute reliable values.

In general, there is a high risk to produce artificial models by applying a very high step length to the gradient direction. Although, this example shows a perfect case, limitations should be applied to the step lengths – accepting a slower convergence but preventively stabilizing the inversion.

Furthermore, this experiment demonstrates the problematic nature of the data misfit function with respect to the estimation of adaptive step lengths (see discussion in section 2.6.2). Figure D.25 compares data misfit function, corresponding model error and the step length estimation (using the common choice of $\mu_0 = 1$ and $a = 1$ of the other tests) at the first iteration. It combines essential observations of previous experiments with respect to frequency filtering (experiment II) and step length estimation (experiment IV):

- Due to the usage of intermediate initial models with a reasonable background velocity (such as Figures D.25b,e,h), the shape of data misfit function and model error show a good agreement. In other words, if a step length estimation successfully locates the desired local minimum of the data misfit, then this should automatically corresponds to the optimum model error. However, in case of inverting for the full frequency content, the parabolic function is too inaccurate to account for the more complex shape of the misfit function. Consequently, an inappropriate step length is computed. The decrease of the frequency content significantly mitigates the problem (Figures D.25e,h).

- Especially the usage of too low initial velocities (Figures D.25a,d,g) increases the qualitative mismatch of data misfit and model error. Either the algorithm is unable to compute an optimal step length (Figure D.25d), or the optimal step length corresponds to the local minimum of the data misfit but misses the optimal model error (e. g., Figure D.25a).

- Figures D.25c,f,i illustrate quite well-posed shapes of data misfit and model error in case of initial models with a higher velocity. However, the step length estimation exhibits an unpredictable behavior. Figure D.25c demonstrates a perfect local fit of the misfit function. Again, Figure D.25f represents a failure of estimating the step length. Apparently, Figure D.25i shows a nearly perfect fit. However, the huge aperture of the parabola and the usage of test step lengths $\left(\frac{1}{8}, 1, 8\right)$ % cause a high sensitivity of the parabolic shape. Marginal variations of the misfit function within the range of the test step lengths could cause improper step lengths or a destabilization of the algorithm.
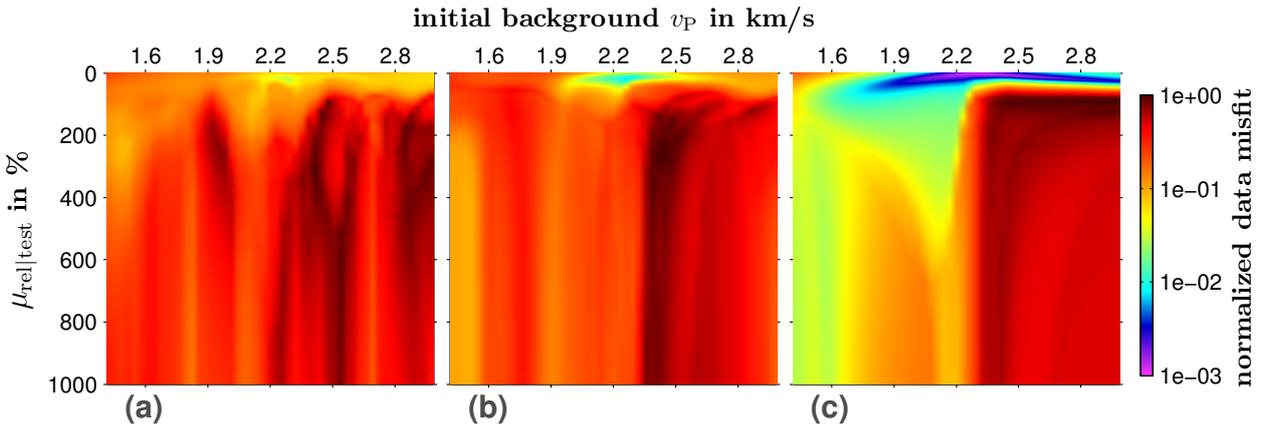
Figure 4.24: Application 1, experiment X: The normalized data misfit as a function of initial model and step length at the first FWT iteration. Frequency filtering for different frequency contents is applied: (a) $f_{peak} = 200$ Hz, (b) $f_{peak} = 95$ Hz, (c) $f_{peak} = 18$ Hz.
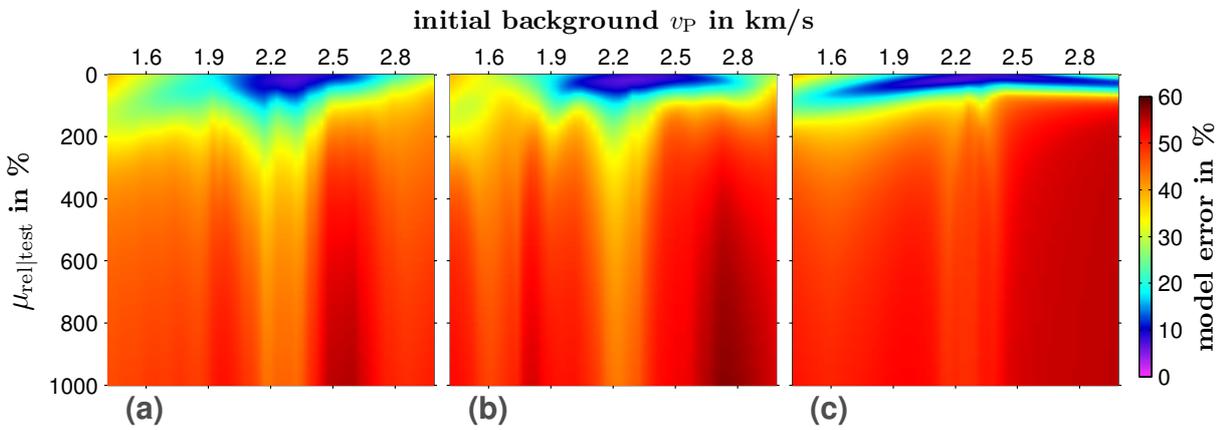

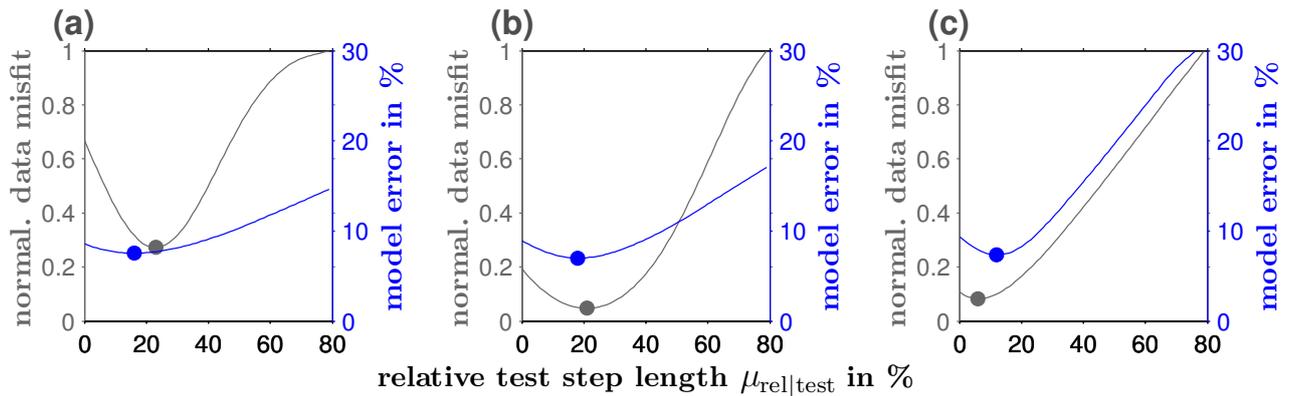**Figure 4.25:** Application 1, experiment X: The model error as a function of initial model and step length at the first FWT iteration. Frequency filtering for different frequency contents is applied: (a) $f_{peak} = 200$ Hz, (b) $f_{peak} = 95$ Hz, (c) $f_{peak} = 18$ Hz.

Figure 4.26: Application 1, experiment X: Cross sections from data misfit functions (Figure 4.24) and model errors (Figure 4.25) for different initial models. (a) corresponds to $v_{P|ini} = 2292 \frac{m}{s}$ in Figures 4.24a and 4.25a. (b) corresponds to $v_{P|ini} = 2262 \frac{m}{s}$ in Figures 4.24b and 4.25b. (c) corresponds to $v_{P|ini} = 2231 \frac{m}{s}$ in Figures 4.24c and 4.25c.

4.3 Summary of experiments

In general, the FWT represents an ill-posed inverse problem, which can be mitigated by several strategies. This chapter discussed methods which can be applied at suitable stages of the FWT scheme. With the focus on the dependence on the initial model, specific experiments investigated the impact of these methods on the progress of the FWT and the resulting v_P reconstruction. The summarizing comparison of all techniques with respect to the plain reference computation (experiment I) and an optimized FWT test (experiment VI/1) alleviates the classification of their influence. The following considerations limit to the model error to quantify the FWT progress.

Based on a full-space velocity model with borehole acquisition geometries, the experiments of this chapter reveal a specific dependence on the initial model (see Figure 4.27). The initial model error is characterized by a broad range from 7.16 % ($v_{P|ini} = 2380 \frac{m}{s}$) up to 38.2 % ($v_{P|ini} = 1430 \frac{m}{s}$).

The final model error of the reference experiment widely reproduces the initial error. Only a narrow window of initial models allows a successful inversion. At this, the optimum initial model does not necessarily yield the best inversion result (final error $\epsilon_{model} = 3.79$ % for $v_{P|ini} = 2323 \frac{m}{s}$). Rather, the corresponding final v_P model is reconstructed artificially. In contrast, the optimized experiment VI/1 significantly reduces both the dependence on the initial model and the final model error (final error $\epsilon_{model} = 3.15$ % for $v_{P|ini} = 1707 \frac{m}{s}$).

Due to the usage of full data and full frequency content in the reference experiment, especially cycle-skipping increases the ambiguity of the inverse problem. This issue might be eliminated by the application of several techniques, in particular frequency filtering.

Figure 4.28 confronts the individual impact of all methods in case of two slightly different initial models with significantly different model errors: $v_{P|ini} = 2138 \frac{m}{s}$ (outside the model window shown in Figure 4.27 with $\epsilon_{model} = 10.8$ %, referred to as "bad" model) and $v_{P|ini} = 2200 \frac{m}{s}$ (within the model window with $\epsilon_{model} = 9.27$ %, referred to as "good" model). Depending on their impact, the techniques can be divided into several groups:
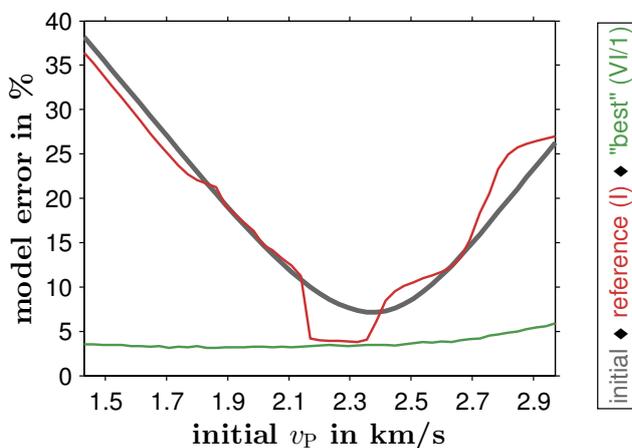
Figure 4.27: Application 1: Summary of FWT experiments. The plot shows the model error as a function of the initial velocity model. It is a comparison of initial model error with the final errors after 100 iterations for the reference experiment I and the optimal experiment VI/1.

Group 1: Negative or no significant impact

- Especially **time windowing** does not effectuate notable improvements (Figures 4.28a,e). In contrast, in case of the "bad" initial model, an increase of the model error can be observed. However, in case of using the "good" model, time windowing does not positively affect the outcome of the FWT but decreases the convergence.

Group 2: Partially positive effect

- Obviously, **offset windowing** can be applied to account for cycle-skipping. Consequently, in case of the "bad" model, it helps to obtain a remarkable reduction of the model error (Figure 4.28a). With respect to the usage of the "good" model, there is an analogy of offset windowing with time windowing. Due to the absence of the cycle-skipping effect, there are nearly no improvements by the application of offset windowing (Figure 4.28e).

- In comparison with the reference experiment, **gradient preconditioning** is useful to increase the convergence (Figure 4.28f). The best performance is achieved by applying tapers at sources and receivers. However, gradient preconditioning does not result in relevant enhancements of the final v_P model.

- Due to the simplicity of the initial models and the high complexity of the true model, the application of a **reflection geometry** does not yield satisfactory inversion results.

Group 3: Significant positive effect

- In case of using the "good" initial model, the advantages of the **adaptive step length estimation** can be exploited (Figure 4.28g). Although, both experiment VI/1 and experiment IV/3 involve this technique, the individual application of the adaptive step length yields the strongest convergence and the lowest model error of all experiments. Thus, the involvement of additional methods seems to be unnecessary. Apart from that issue, the adaptive step length estimation does not reduce the dependence on the initial model. While the data misfit decreases, the model error shows a contrary behavior (Figure 4.28c).
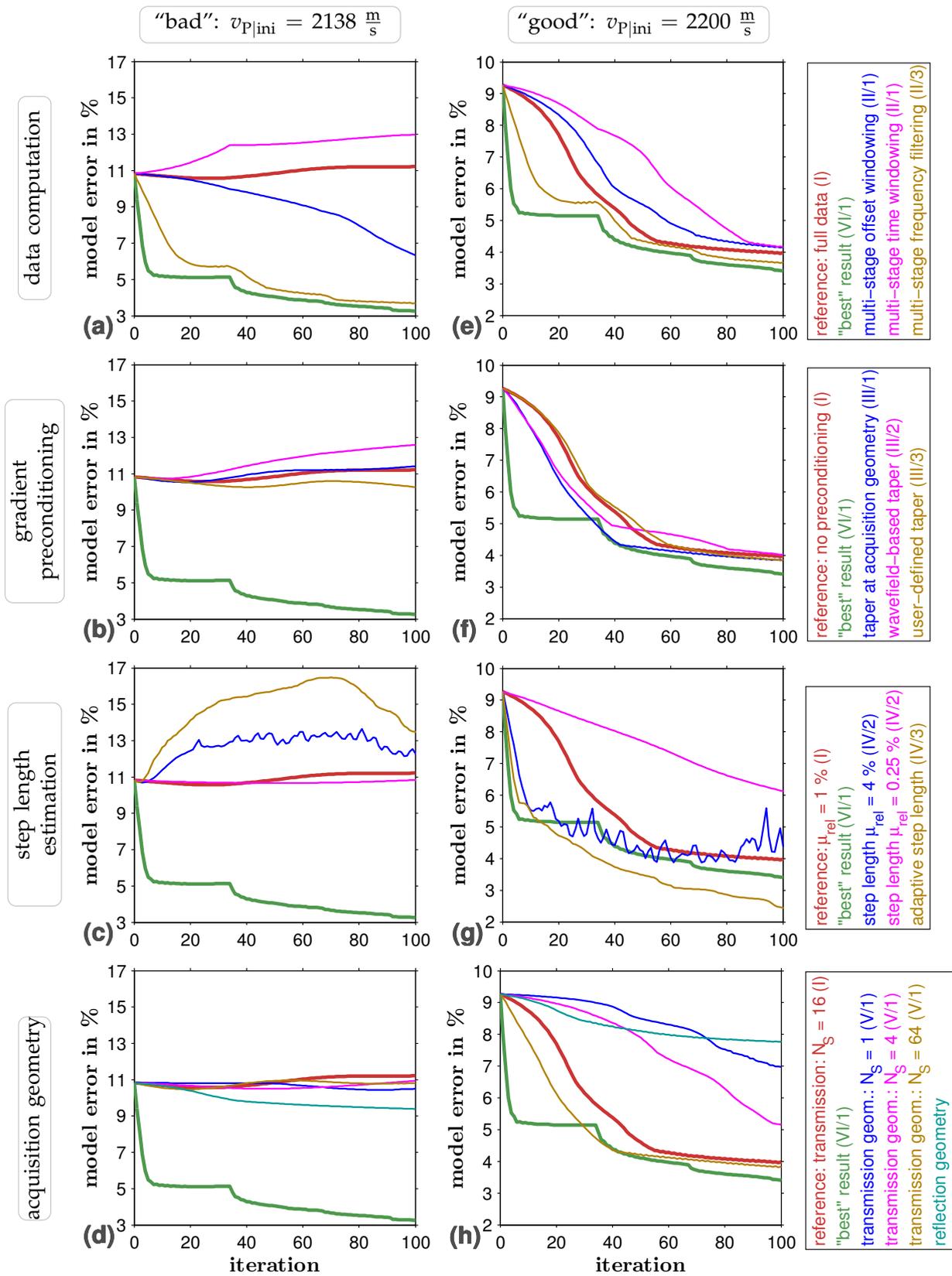
Figure 4.28: Application 1: Compilation of FWT performances of reference and optimal experiment as well as several experiments with individual application of several techniques (see rows). The plots show the model errors with respect to a "bad" and a "good" initial model. The progresses of the model errors represent cross sections of the multiple FWT results (see Figures in appendix D.2).

- The **acquisition geometry** plays a major role with respect to model accuracy, convergence and computational efforts of the FWT. A general rule of thumb is: The higher the illumination of the subsurface, the better the performance of the FWT. But, this principle is limited by some circumstances. The inversion requires a "good" initial model to benefit from a high illumination (compare Figures 4.28d,h). On this understanding and considering a transmission geometry, the quality of the inverted model definitely depends on the number of sources, N_s: The higher N_s, the lower the model error (Figure 4.28h). However, this rule does not cause an unlimited improvement of the model. Due to physical conditions, such as maximum available frequency contents or minimum available wavelengths, the ability of reconstructing the velocity model is limited. Within the scope of the given example, a sufficient performance is obtained by using 16 sources. An increase up to 64 sources does not effectuate further improvements – but tremendously increases the computational efforts.

Group 4: Significant positive effect with high independence on the initial model

- **Frequency filtering** can be regarded as the most reliable method to mitigate the ambiguity of the inverse problem. In case of using the full frequency content (related to short cycles), small variations of the initial background velocity are enough to induce the cycle-skipping effect. The restriction to low frequencies increases the cycle length. Thus, the appearance of the cycle-skipping effect requires a larger time-shift of the seismic waveforms. The exploitation of low frequencies proves to be very important (Sirgue, 2006). Consequently, a broader range of initial velocity models, such as examples in Figures 4.28a,e, allows a successful inversion.

Chapter 5

Application 2: Comparison of FWT in the time domain and the frequency domain

This chapter follows the final conclusions of section 4.3. As mentioned in chapter 4, the progress of the inversion depends on the frequency content of the data or source signal, respectively. The higher the frequency, the more information is included in the data and the more complicated the objective function is. In case of using the full frequency content the application of FWT to a bad starting model may fail, whereas a good starting model, which already contains the long-wavelength structures, may be enough to succeed. Thus, a high-frequency dataset for a velocity model, which contains both small- and large-scale structures, is suitable for a comparison of the time-domain implementation with the frequency-domain implementation *FULLWV* developed by R. G. Pratt (Pratt, 1999) and the single-frequency method proposed by Sirgue et al. (2008).

5.1 Emulation of frequency-domain inversion

The most important advantage of frequency-domain FWT is the inversion of selected frequencies in the frequency domain. This method starts at low frequencies to resolve the basic large-scale structures and adds higher frequency contents later on. Thus, the suppression of high-frequency contents at the beginning results in a simplification of the objective function and increases the chance of computing reliable results.

In the time-domain FWT, there are different possibilities to emulate the advantages of frequency-domain FWT. On the one hand, frequency filtering can be applied to mitigate the ambiguity of the inverse problem (see chapter 4). On the other hand, the single-frequency method combines the advantages of time-domain and frequency-domain FWT. While forward-propagations and back-propagations are performed in the time domain, the imaging condition is evaluated in the frequency domain. Although, 2D frequency-domain FWT reveals a high computational efficiency, 3D modelings in the frequency domain are much more expensive. Thus, the single-frequency method represents a good trade-off of time-domain and frequency-domain FWT.

The fundamental idea of the single-frequency method is the selection of frequencies by applying the discrete Fourier transform to the time-domain wavefields. The approach

used in this work is a minor modification of the method published by Sirgue et al. (2008). Contrary to that, the source time function is applied at time-domain forward modeling. The imaging condition only consists of a zero-lag cross-correlation of the forward-propagated and the back-propagated wavefields. In detail, the time-domain source wavefield $\mathbf{p}(\mathbf{x}, t|\mathrm{s})$ and the residual wavefield $\mathbf{p}^*(\mathbf{x}, t|\mathrm{r})$ are transformed to the frequency domain for selected frequencies f:

$$\underline{\mathbf{p}}(\mathbf{x}, f|\mathrm{s}) = \sum_{t=0}^{T} e^{-i2\pi ft}\mathbf{p}(\mathbf{x}, t|\mathrm{s}) \quad \text{and} \quad \underline{\mathbf{p}}^*(\mathbf{x}, f|\mathrm{r}) = \sum_{t=0}^{T} e^{-i2\pi ft}\mathbf{p}^*(\mathbf{x}, t|\mathrm{r}), \qquad (5.1)$$

where $\underline{\mathbf{p}}(\mathbf{x}, f|\mathrm{s})$ and $\underline{\mathbf{p}}^*(\mathbf{x}, f|\mathrm{r})$ denote the frequency-domain wavefields and T is the maximum propagation time. The gradient $\mathbf{g}(\mathbf{x})$ is obtained by frequency-domain cross-correlation, summation over selected frequencies, N_f and sources N_s:

$$\mathbf{g}(\mathbf{x}) = \sum_{N_\mathrm{s}} \sum_{N_f} \left(-4\pi f^2 \underline{\mathbf{p}}(\mathbf{x}, f|\mathrm{s}) \, \underline{\mathbf{p}}^*(\mathbf{x}, f|\mathrm{r}) \right). \qquad (5.2)$$

According to pure frequency-domain methods, such as Pratt (1999), the inversion starts with the extraction of low-frequency contents and moves to higher-frequency information at later iterations. A common approach is the manual definition of individual frequencies and the corresponding number of required iterations for each frequency selection. However, in general the convergence of the misfit function is unknown and might vary among different frequencies. Hence, the implementation comprises an automatic selection realized by the workflow (see section 3.2.2).

5.2   Basic setup

5.2.1   General parameters

The comparison of the FWT in time domain and frequency domain employs the random medium model with a size of 320 × 210 meters and includes a cross-well acquisition geometry. It is generated by applying an exponential autocorrelation function with a correlation length of 52 m (Figure 5.1a), where v_P varies between 1600 $\frac{\mathrm{m}}{\mathrm{s}}$ and 2400 $\frac{\mathrm{m}}{\mathrm{s}}$. The resulting model is a self-similar medium and includes structures on different length scales. For waveform tomography, two different initial models are used. The first one is homogeneous with $v_\mathrm{P} = 2000 \, \frac{\mathrm{m}}{\mathrm{s}}$ (Figure 5.1b). The second initial model is computed from the true model by applying a 2D Gaussian filter (see equation (2.1)) with the constant $\sigma = 35$ and a size of 80 × 80 m (Figure 5.1c).

The acquisition geometry is arranged in a cross-hole configuration, which includes 24 explosive sources and 277 receivers. The source spacing is 12 m and the receiver spacing is 1 m. Time-domain forward modeling involves a Ricker-wavelet with a peak frequency of $f_\mathrm{peak} = 150$ Hz and a time discretization of $\Delta t = 0.1$ ms. The record length is set to $T = 0.26$ s, where the wavelet is shifted by 0.08 s to account for wavelet "stretching" due to low-pass filtering. The grid spacing is set to $DH = 0.5$ m to fulfill the grid dispersion criterion (2.13). This results in a grid size of 640 × 420 grid points. The model is located

Table 5.1: Application 2: General setup.

Application 2: Comparison of FWT in the time domain and the frequency domain			
Attributes		**Specifications**	
Model	**Size**	width: 210 m, height: 320 m	
	Structure	random medium	· average v_P = 2000 m/s · maximum range: v_P = [1600, 2400] m/s · standard deviation σ = 0.075 · correlation length: 52 m · exponential autocorrelation function
	Figure	Figure 5.1	
Acquisition geometry	**Location**	full-space, borehole	
	Configuration	transmission geometry (cross-hole)	
	Sources	· 24 explosive sources · source-time function: Ricker wavelet · peak frequency: 150 Hz · time delay of source signal: 0.08 s (time domain only)	
	Receivers	· 277 hydrophones · effective recording length: 0.18 s	
Modeling parameters	**Model size**	time domain: · width: 420 grid points, height: 640 grid points · grid spacing: 0.5 m frequency domain: · width: 210 grid points, height: 320 grid points · grid spacing: 1 m	
	Propagation time	time domain only: · 2600 time steps (effective recording length + time delay) · time discretization: $1 \cdot 10^{-4}$ s	
	PML boundary	width: 19 m (time domain only)	
General inversion parameters	**Initial models**	homogeneous	v_P = 2000 m/s
		smooth	application of a Gaussian filter to the true model: · width: 80 m, height: 80 m · σ = 35
		Figure 5.1	
	Model update	hard constraint: v_P range: [1400, 2600] m/s	
	Frequency selection	time domain	frequency filtering
			single frequencies
		freq. domain	single frequencies
		Figure 5.5	

within the full space, surrounded by perfectly matched layers with a width of 19 m. All general modeling and inversion parameters are summarized by Table 5.1.

Exemplary observed data is illustrated in Figure 5.2. Due to different wavelet definitions and modeling parameters, the inversion results separately refer to the observed data obtained from time-domain modeling (Figure 5.2a) and frequency-domain modeling (Figure 5.2b).

Furthermore, the initial seismograms (Figures 5.3a,b, obtained from time-domain modeling) demonstrate potential issues with respect to the choice of the initial model (see chapter 4). Despite a plausible average velocity model, the homogeneous initial model does not explain the observed data. The corresponding initial data is characterized by a huge phase shift (compare Figures 5.2a and 5.3a). Probably, it is too far away from the true model which may cause a deadlock in a local minimum of the data misfit function and a failure of the inversion. In contrast, the smooth model represents a much better initial model. It is already a good representation of the long-wavelength structures (compare smooth initial model in
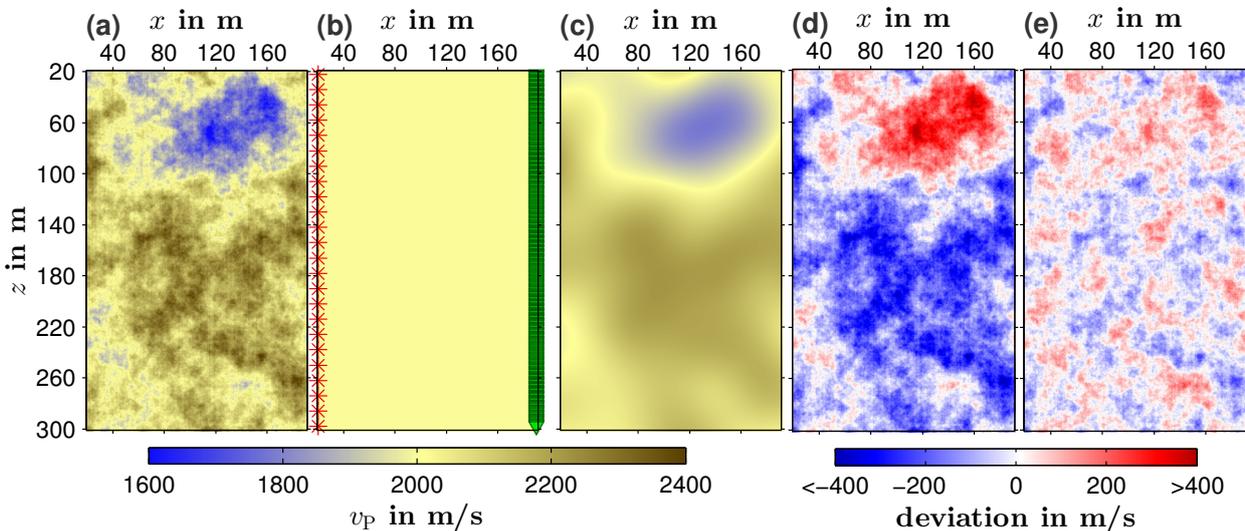
Figure 5.1: Application 2: Random medium velocity model: (a) true model, (b) homogeneous initial model with acquisition geometry including 24 sources (red markers) and 277 receivers (green markers), (c) smooth starting model computed from the true model, (d) and (e) show deviations of the initial models (c and d) from the true model.
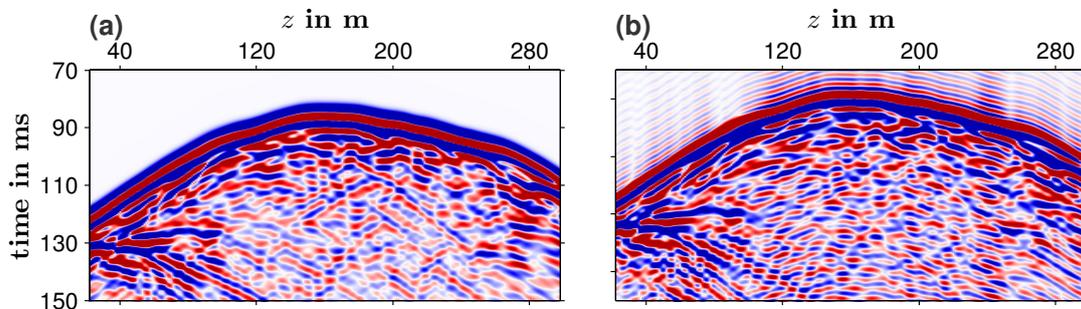
Figure 5.2: Application 2: Observed seismograms belonging to the true velocity model in Figure 5.1a. (a) is computed in the time domain, while (b) is synthesized from frequency-domain Fourier coefficients using the modeling software *OMEGA* developed by R. G. Pratt.

Figure 5.1c with the residual model in Figure 5.1e). Hence, this model is able to give a better explanation of the direct wave (Figure 5.3b) resulting in a lower data misfit (compare initial data residuals in Figures 5.3c,d).
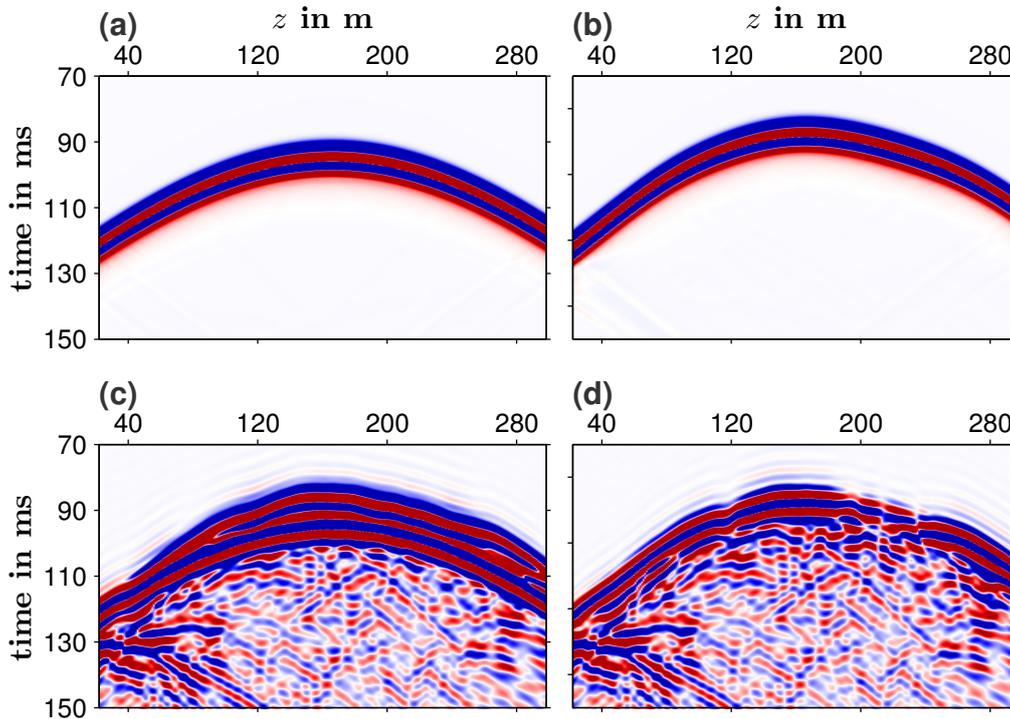
Figure 5.3: Application 2: Initial seismograms (a and b) belonging to the initial velocity models in Figure 5.1b,c and corresponding residuals (c and d) computed with respect to the observed data in Figure 5.2a. The data are obtained from time-domain modeling.

5.2.2 Configuration of comparative experiments

The comparison of FWT in the time-domain (TD) and frequency-domain (FD) comprises two consecutive experiments. The aim of the preliminary experiment is to find an appropriate initial model for the second experiment. In other words, the initial model has to be "bad" enough to cause a failure of the FWT which simultaneously inverts for the full frequency content of the data. The second experiment compares the performance of frequency-domain FWT with time-domain FWT involving emulation techniques (see section 5.2.1).

Table 5.2 gives an overview of the configuration of both experiments. The preliminary experiment performs inversions in the time domain and frequency domain with respect to the homogeneous and smooth initial model. While the TD FWT involves the full time-domain data, the FD FWT simultaneously inverts for a selection of 50 equidistantly spaced frequencies ranging from 8 Hz to 302 Hz. The frequency selection appropriately represents the maximum frequency range of the data (see Figure 5.4). Based on the usage of the homogeneous initial model, the consecutive experiment investigates the performance of two emulation techniques in the TD FWT. On the one hand, 50 stages of low-pass filters with ascending upper corner frequencies are applied (see blue graphs in Figure 5.5). However, low-pass filters do not perfectly reproduce the effect of the single-frequency selection. Already at stage 30 the filtered data exhibits the maximum peak frequency of approximately

145 Hz. On the other hand, both the TD FWT and the FD FWT exploit the same serial selection of single frequencies (see red graph in Figure 5.5).

Table 5.2: Application 2: Detailed setup of comparative experiments.

Preliminary experiment				
Initial model	homogeneous		smooth	
FWT	TD	FD	TD	FD
Frequency selection	full data	simultaneous inversion for 50 frequencies $f = (8, 14, \ldots, 302)$ Hz	full data	simultaneous inversion for 50 frequencies $f = (8, 14, \ldots, 302)$ Hz
Figures	5.6a,e,i,m	5.6b,f,j,n	5.6c,g,k,o	5.6d,h,l,p

Consecutive experiment			
Initial model	homogeneous		
FWT	TD	TD	FD
Frequency selection	serial application of 50 low-pass filters with ascending corner frequencies	serial application of 50 ascending single frequencies	
	$\dfrac{4 \text{ iterations}}{\text{frequency range}}$	$\dfrac{4 \text{ iterations}}{\text{frequency}}$	
Figures	5.7a,d,g,j	5.7b,e,h,k	5.7c,f,i,l
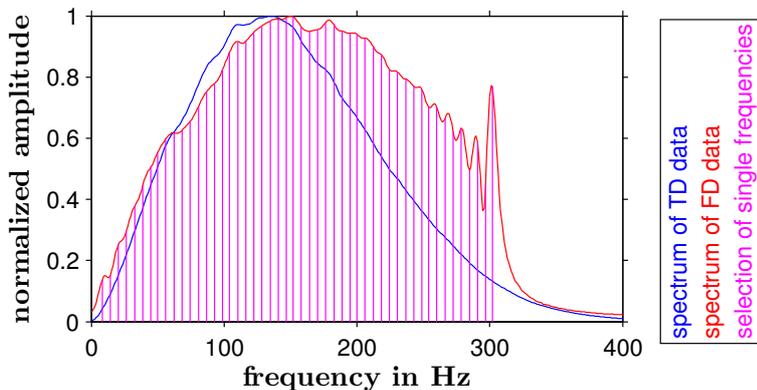
Figure 5.4: Application 2: Amplitude spectra of observed TD and FD data as well as frequency selection of FD FWT.
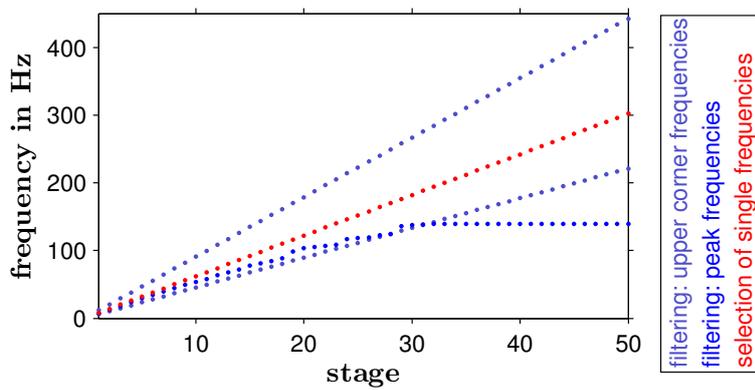
Figure 5.5: Application 2: Frequency selection for emulation using low-pass frequency filtering in TD FWT (blue graphs) or applying single frequencies in TD FWT and FD FWT (red graph). Both corner frequencies and single frequencies are equidistantly spaced.

5.3   Results of time-domain and frequency-domain FWT

The results of the first experiment are assembled in Figure 5.6. Obviously, the homogeneous initial model fulfills the requirements of a "bad" model. Both TD FWT and FD FWT fail to reconstruct a satisfactory v_P model. While the TD FWT produces an artificial result (see inverted model and deviation from the true model in Figures 5.6a,e), the FD FWT computes a v_P model showing a high similarity to the true model (compare Figures 5.6b,f). In both cases, the FWT is unable to explain the observed data. The final synthetic data does not match the observed data at all (compare seismograms in Figures 5.6i,m and 5.6j,n with corresponding observed data in Figures 5.2).

In contrast, the smooth initial model is not suited for the following experiment. Due to its advantages mentioned in section 5.2.1, the FWTs produce significantly better inversion results (see v_P models in Figures 5.6c,d and associated deviations from the true model in Figures 5.6g,h). Accordingly, this observation is verified by the final data fit (see final synthetic and residual seismograms in Figures 5.6k,l as well as 5.6o,p). However, once a good model is found, the TD FWT not only computes a reliable v_P model, but also reveals a better performance than the FD FWT. Probably this is caused by the different consideration of frequency contents. While the FD FWT uses single frequencies, the TD FWT involves the full frequency range (see Figure 5.4).

On the basis of the preliminary experiment, the second experiment compares the inversion results of the FD implementation and TD implementation including emulation techniques, such as frequency filtering and the single-frequency method. Figure 5.7 summarizes the inverted v_P models and corresponding deviations from the true model as well as final synthetic data and residuals. Apart from minor overestimated velocities within the v_P model computed by the FD FWT (Figures 5.7c,f), all results resemble each other (compare Figures 5.7a,b,c). The TD FWT involving the emulation techniques is able to overcome the disadvantages of the inversion for the full frequency content of the data (see first and second column in Figures 5.7).
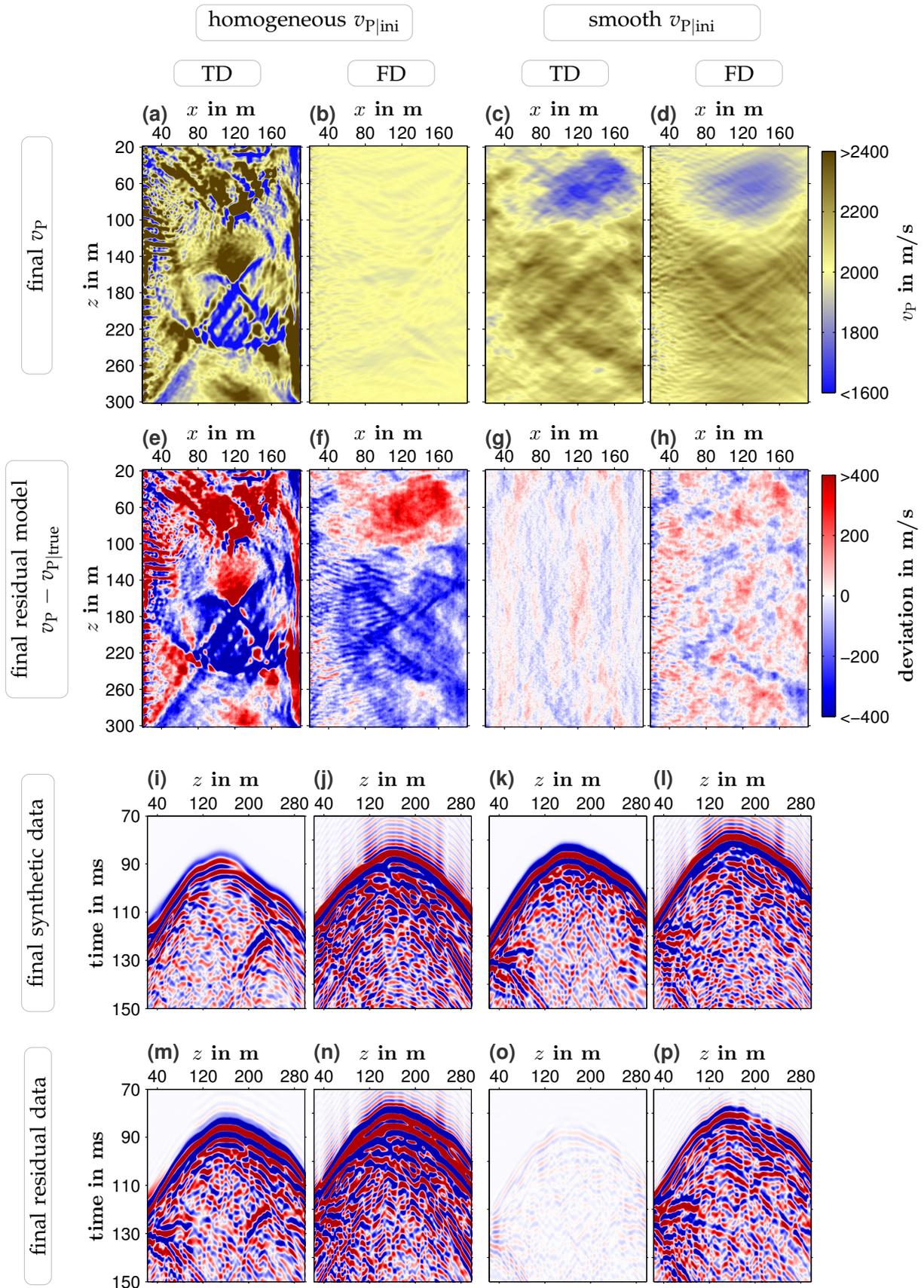
Figure 5.6: Application 2: Results of time-domain FWT (TD) and frequency-domain FWT (FD) with respect to the simultaneous inversion for the full frequency content. The data plots represent the central shot located at $z = 166$ m.
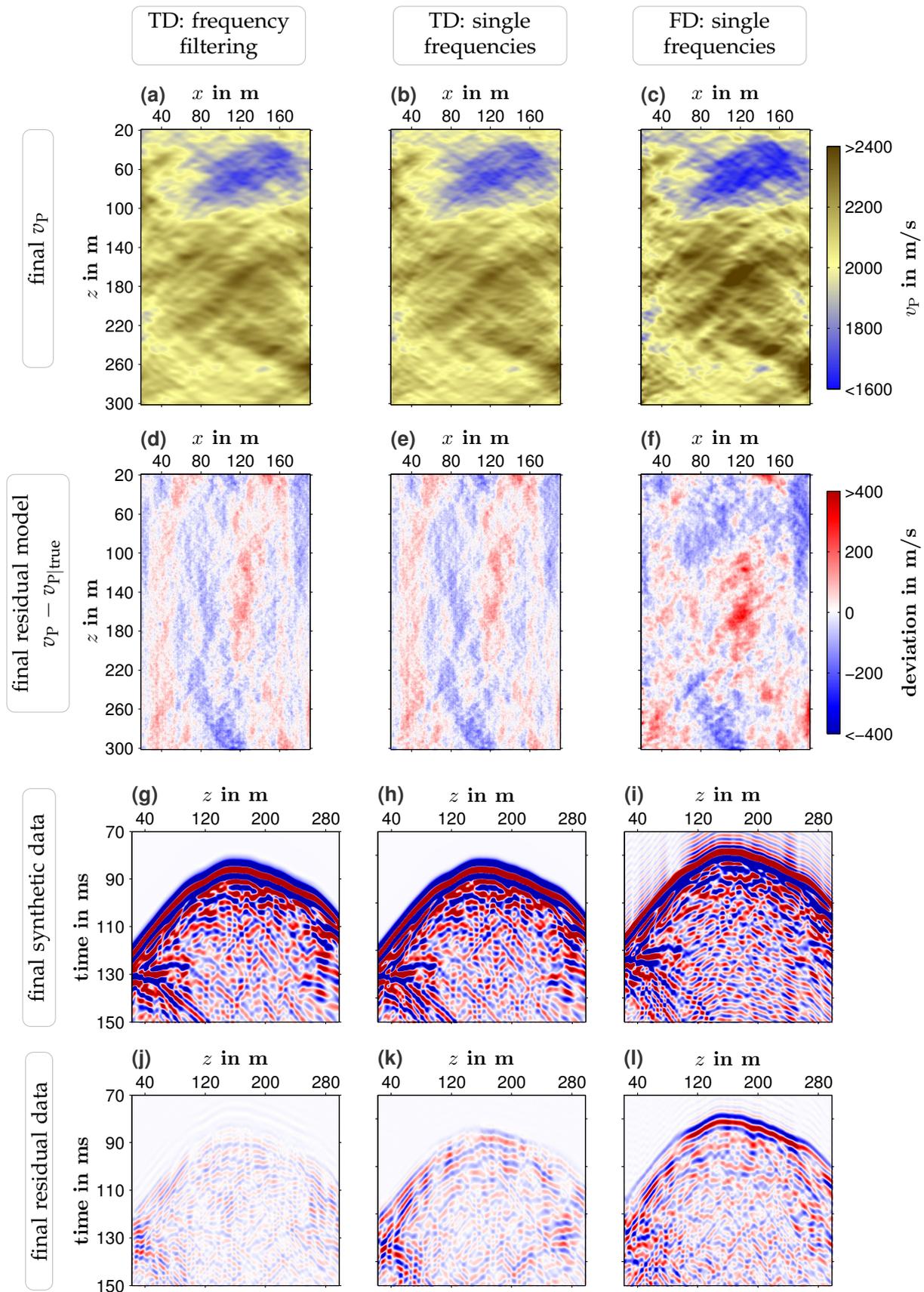
Figure 5.7: Application 2: Results of time-domain (TD) FWT strategies (first and second column) and frequency-domain (FD) FWT with incremental frequency selection (right column). The homogeneous initial v_P model is used. The data are plotted for the shot located at $z = 166$ m.

5.4 Summary

The comparison of frequency-domain FWT with time-domain FWT involving different techniques to emulate the frequency selection reveals reliable final v_P models. Apart from minor inaccuracies of the FD result, the application of frequency filtering and single frequency method resemble the result of the FD implementation *FULLWV* developed by Pratt (1999). Although, the emulation techniques additionally require fast Fourier transforms (filtering) or discrete Fourier transforms (single frequency method), the computational efforts increase insignificantly. However, frequency filtering is only applied to the source time function and to the observed data. Due to the focus on certain frequency ranges or single frequencies, it is not necessary to include wavefield "snapshots" at all finite-difference time steps into the cross-correlation. A strict Nyquist criterion is applied to choose a subset of "snapshots" – tremendously reducing the essential number of discrete Fourier transforms. Consequently, especially in 3D applications the emulation techniques prove to be very useful (as performed by Sirgue et al., 2008).

Chapter 6

Application 3: Acoustic FWT in the presence of attenuation: a quantitative study

The aim of FWT is to find a subsurface model which explains the recorded seismic data, i. e., it iteratively minimizes the difference between observed and synthetic seismograms. The majority of FWT applications faces a multi-parameter problem. In particular, attenuation affects both amplitude and phase of seismic signals causing amplitude changes and frequency-dependent phase velocity dispersion (Causse et al., 1999).

Nowadays, a purely acoustic FWT is usually applied to recover the P-wave velocity model in transmission and reflection seismic configurations. However, most applications neglect the impact of intrinsic attenuation on seismic waveforms and thus the model reconstruction. The purpose of this work is to investigate the influence of attenuation on the recovered velocity model and the validity of an acoustic FWT in presence of attenuation.

There are two different ways to take attenuation into account. On the one hand, attenuation is used as a passive parameter, i. e., attenuation is a modeling parameter only. The aim is to improve the performance of the inversion for desired parameters, such as P-wave velocity v_P (e. g., Brenders and Pratt, 2007b). On the other hand, a multi-parameter FWT can involve attenuation as an additional inversion parameter (e. g., Hak and Mulder, 2008, 2011; Kamei and Pratt, 2008).

Field data applications are mainly conducted in the frequency domain. Hicks and Pratt (2001) obtained reliable quality factor subsurface models from shallow seismic data recorded in the North Sea. Takam Takougang and Calvert (2011) reconstructed realistic velocity models from marine reflection data. In order to achieve satisfactory results, they combined different inversion strategies, such as a multi-stage approach with incremental frequency selection (compare Bunks et al., 1995; Sirgue and Pratt, 2004) and separate inversion of near and far offsets. The v_P-only inversion at low frequencies connected with the joint inversion for v_P and Q_P at higher frequencies improved the recovery of both v_P and Q_P models. Malinowski et al. (2011) applied a frequency-domain viscoacoustic FWT to wide-aperture seismic field data recorded in the Polish basin. The focus of their investigation is the applicability of a viscoacoustic joint inversion for P-wave velocity v_P and Q_P. They reconstructed satisfactory subsurface models for both v_P and Q_P coinciding with the expected geology. In this context,

Mulder and Hak (2009) discussed problems of a joint inversion with respect to short-aperture data in reflection seismics. They found that the ill-posedness of the inverse problem causes a very poor recovery of both phase velocity and attenuation. A brief overview of recent publications in this area is given by Virieux and Operto (2009).

In this chapter, I use the time-domain implementation to study the role of attenuation for weakly and strongly attenuative media in marine environments. I investigate the validity of an acoustic inversion scheme in presence of attenuation by systematically quantifying the errors in the inverted velocity model. Attenuation is incorporated as a passive parameter in forward modeling only. I apply acoustic FWT including purely acoustic or viscoacoustic modeling to analyze and compare the impact of attenuation on the reconstruction of the velocity models from viscoacoustic reflection datasets.

Although this study concentrates on 2D acoustic FWT in the time domain, it is also targeted to 3D applications. In contrast to the high efficiency of 2D frequency-domain FWT due to straightforward implementation of attenuation, 3D frequency-domain modeling is highly demanding. Hence, in 3D FWT applications time-domain modeling is commonly used. Furthermore, the time-domain approach has some advantageous features, such as straightforward time-windowing of data and the consideration of broad frequency bands (instead of single frequencies).

This chapter deals with the application of acoustic FWT to viscoacoustic data. The first section verifies the viscoacoustic time-domain modeling by comparing it with a semi-analytical solution. The second section discusses the impact of attenuation by choosing two examples: a simple 1D medium with a reflection acquisition geometry providing a high ray coverage and the 2D Marmousi model with a towed streamer geometry. For both experiments the same inversion tests are performed. The investigation comprises the evaluation of the resulting data misfits and model errors to demonstrate the footprint of attenuation on the recovered velocity models.

6.1   Accuracy of viscoacoustic modeling in the time domain

This section employs 2D time-domain acoustic as well as viscoacoustic modeling to a homogeneous full-space example. The results are compared with a semi-analytical solution. The model and its parameters are shown in Figure 6.1. For finite-difference modeling I use a grid size of 1000×1600 grid points with a spacing of $\Delta h = 0.5$ m, a time step of $\Delta t = 10^{-4}$ s and a Ricker wavelet with a peak source frequency of $f_{\text{peak}} = 80$ Hz. To suppress artificial reflections from model boundaries, the acquisition geometry is surrounded by a broad perfectly matched layer (PML).

In the following time-domain viscoacoustic approach the attenuation of a medium is approximated by $L = 3$ relaxation mechanisms within the desired frequency range of $f = [0, 200]$ Hz, which corresponds to the relevant frequency content of the source wavelet. L relaxation frequencies $f_{\text{r},l}$ ($l = \{1, \ldots, L\}$) are estimated from $Q_{\text{P},0}$ and the reference frequency $f_0 := f_{\text{peak}}$. This is realized by the Q_{P}-approximation method mentioned above. The result of this optimization is a logarithmically spaced set $f_{\text{r},l} = (1.470, 21.40, 199.6)$ Hz. The approximation of the quality factor $Q_{\text{P}}(f)$ and the resulting dispersion $v_{\text{P}}(f)$ is illustrated by Figures 6.2 (b) and (c). At frequency f_0 I assume the acoustic reference
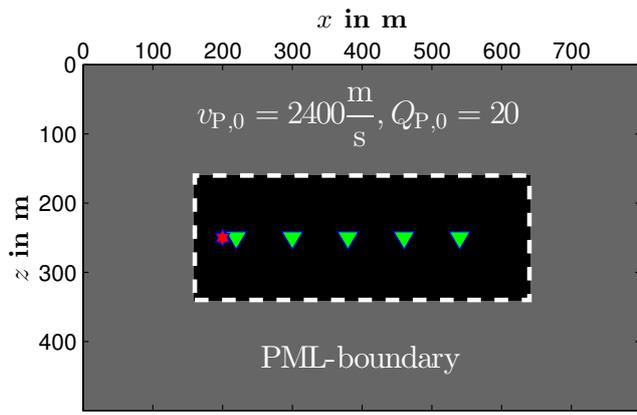
Figure 6.1: Application 3: Viscoacoustic homogeneous full-space model used for time-domain finite-difference modeling. The acquisition geometry consists of one directed force source in vertical direction (red star) and five receivers (green triangles).
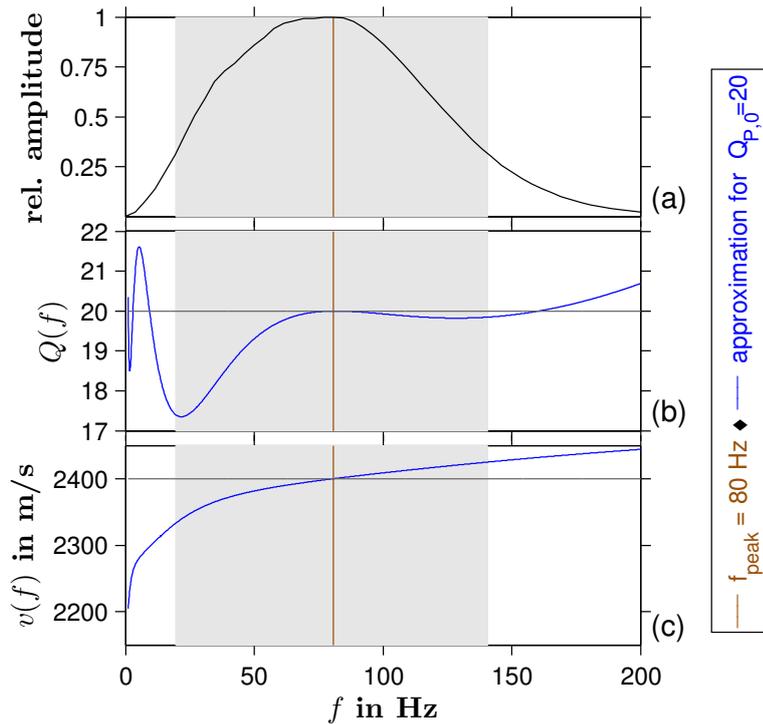

Figure 6.2: Application 3: Approximation of the quality factor $Q_P(f, L = 3)$ for the model shown in Figure 6.1 and the desired frequency range of the underlying source signal (b) as well as corresponding dispersion curve based on an acoustic reference velocity model $v(f_0) = v_{P,0} = 2400 \frac{m}{s}$ (c). The normalized amplitude spectrum of the observed data and corresponding bandwidth (shaded area) are shown in (a)
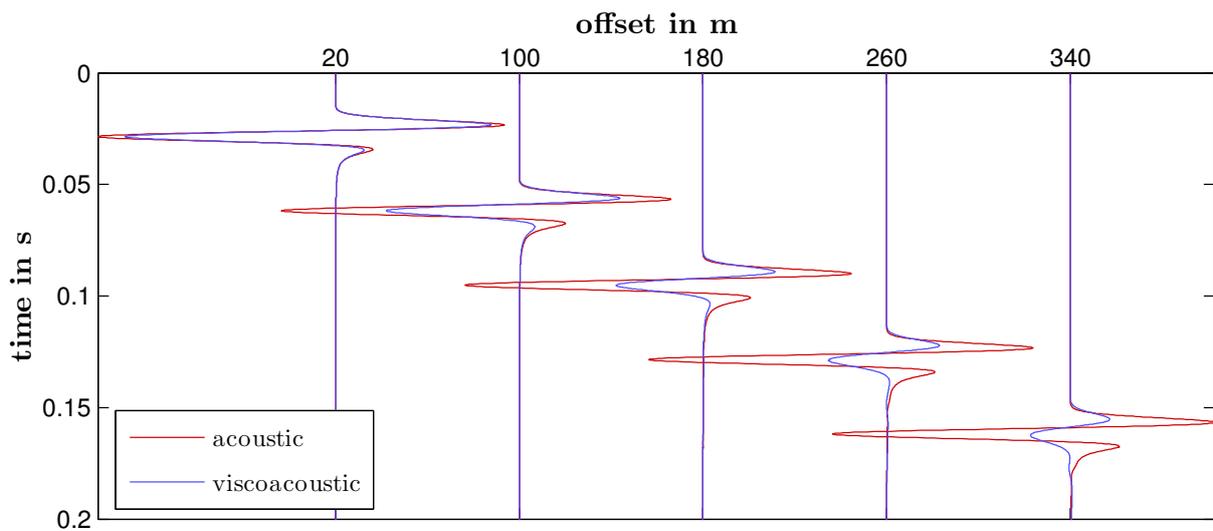
Figure 6.3: Application 3: Vertical particle velocity obtained for geometry in Figure 6.1. Due to plotting issues, all traces are normalized to maximum amplitude of the acoustic signals.

phase velocity $v_{P,0} = 2400 \frac{m}{s}$. Thus, $v_P(f)$ ranges from $v_{P,min} = v_P(f = 0) \approx 2163 \frac{m}{s}$ to $v_{P,max} = v_P(f \to \infty) \approx 2499 \frac{m}{s}$, whereas in finite-difference modeling $v_P(f \to \infty)$ is computed at the Nyquist frequency $f_{Ny} = 1/(2\Delta t) = 5000$ Hz (with respect to time step Δt). Both $v_{P,min}$ and $v_{P,max}$ have to be considered to avoid grid dispersion or instability. Hence, in forward modeling minimum and maximum velocities are computed by equations B.8. In case of the homogeneous model experiment, the velocity limits amount to $v_{min} \approx 2184 \frac{m}{s}$ and $v_{max} \approx 2491 \frac{m}{s}$, which slightly differ from the theoretical values.

The effect of attenuation on seismic data is demonstrated by a comparison of acoustic and viscoacoustic modeling based on the given acquisition geometry. Figure 6.3 shows the resulting seismograms. As expected with increasing offset a significant amplitude decay and a modification of phases can be observed. Especially the amplitudes of high frequency contents are subject to strong attenuation (cp. amplitude spectrum in Figure 6.4a). Furthermore, Figure 6.4b visualizes a comparison of acoustic and viscoacoustic phase spectra. The frequency-dependent phase distortions mainly occur at frequencies $f > f_0$ and disappear around the peak frequency.

Hereafter, the results of viscoacoustic finite-difference modeling are verified using the program *AnalyticAnelastic* (implemented by Josep de la Puente, LMU Geophysics, Munich; based on Carcione et al., 1988b; Carcione, 2001; Emmerich and Korn, 1987). It provides the same rheology as mentioned above and is utilized for computation of a semi-analytical solution of the wave equation in a viscoelastic and consequently viscoacoustic or acoustic medium. Figure 6.5a shows the acoustic reference comparison for an exemplary trace at farthest offset ($x = 340$ m). Both semi-analytical and finite-difference solution fit very well. The viscoacoustic data at near offset ($x = 20$ m) and far offset ($x = 340$ m) are depicted in Figures 6.5b and 6.5c. Apart from minor deviations, the finite-difference solution resembles the semi-analytical solution.
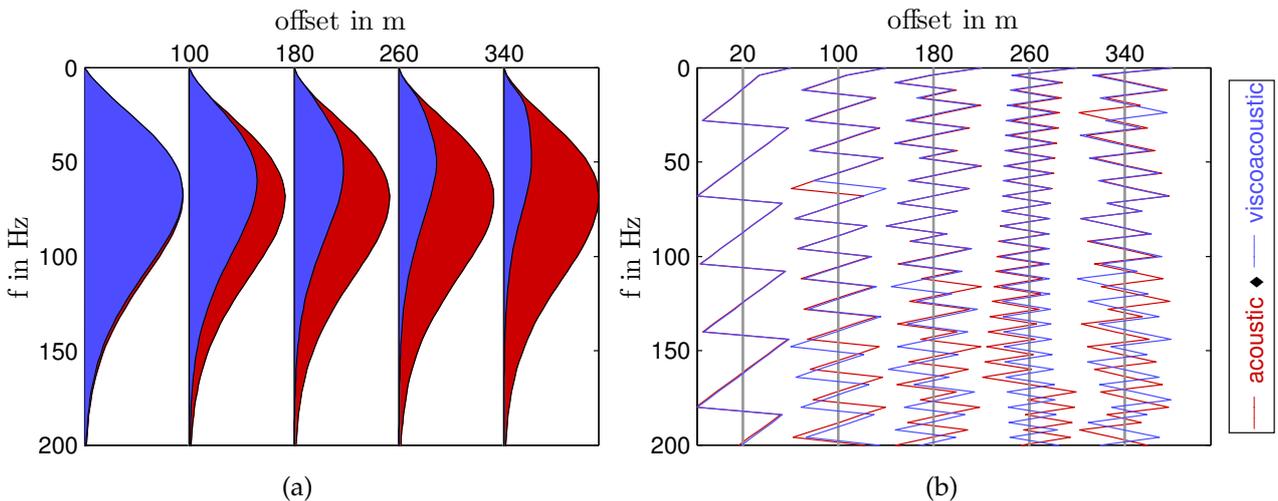
Figure 6.4: Application 3: Amplitude (a) and phase spectra (b) of the seismic data shown in Figure 6.3. All traces in (a) are normalized to maximum values of the acoustic spectra.
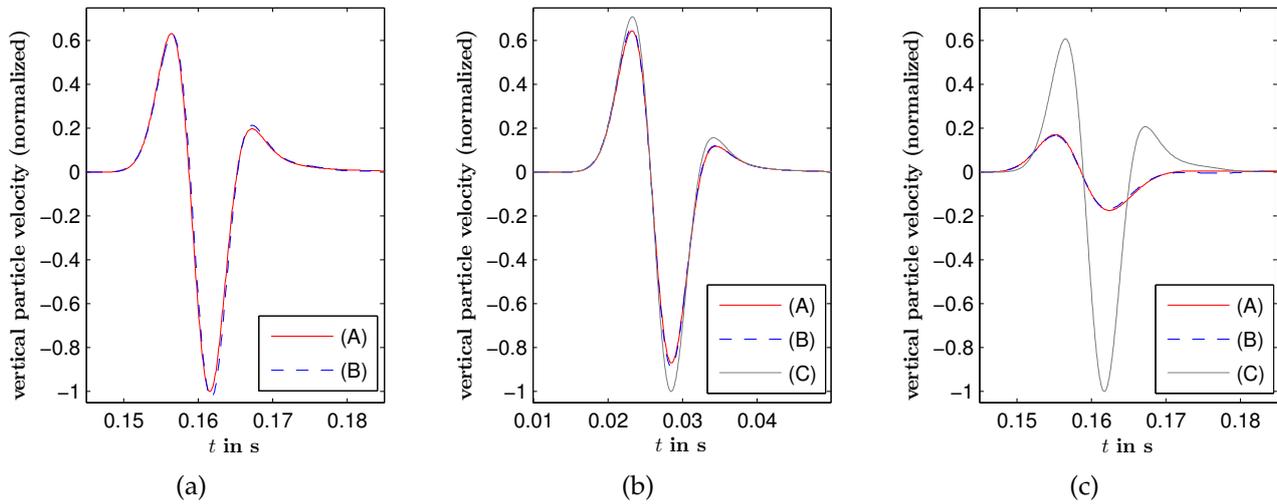
Figure 6.5: Application 3: Comparison of semi-analytical and finite-difference solutions. Exemplary data is shown for the acoustic case at far offset (a), for viscoacoustic data at near offset (b) and far offset (c). Traces (A) are represented by the semi-analytical solution, traces (B) by finite-difference solution and traces (C) by the corresponding acoustic solution in (a).

6.2   Synthetic experiments

I investigate the impact of attenuation in application of acoustic FWT to viscoacoustic data with and without a passive Q_P model. Several acoustic full waveform inversion tests are employed to analyze the footprint of attenuation on v_P inversion results. I analyze the effects of both viscoacoustic and acoustic modeling in acoustic FWT. The tests are applied to both a simple 1D medium and the more complex Marmousi model using different initial v_P and passive Q_P models. A list of all tests can be found in Table 6.1. To avoid unwanted side effects, some general restrictions and only necessary preconditioning features are applied. The FWT tests comprise the following setup and constraints:

(1) the P-wave velocity v_P is the only inversion parameter,

(2) to focus on attenuation-related effects, the density model is neglected,

(3) the true source signal is used, i. e., no inversion for source time function takes place,

(4) in case of viscoacoustic modeling, a Q_P model is used as passive model parameter, i. e., it is not an inversion parameter,

(5) Marmousi experiment: low-pass filtering over multiple stages (Bunks et al., 1995; Sirgue and Pratt, 2004),

(6) wavefield-based gradient preconditioning (see 2.4.2) to suppress source artifacts,

(7) no model update within the water layer due to known parameters v_P and Q_P,

(8) stop criterion for 1D experiment to obtain most optimal results: FWT is unable to reduce data misfit (relative threshold value between two successive iterations is 0.0001 %), or it is not possible to compute a meaningful step length,

(9) stop criterion for shifting within multiple stages in the Marmousi experiment: due to high computational efforts, the relative threshold value of the data misfit between two successive iterations is 1 %.

The reference computation applies acoustic inversion to acoustic data. Its aims to show the performance of FWT for a given geology and geometry of both experiments. Tests 1 and 2 analyze the effect of attenuation on an acoustic inversion with acoustic modeling. Tests 3 to 5 involve viscoacoustic modeling in acoustic FWT and investigate three different passive Q_P models in conjunction with the more realistic initial v_P model.

Data misfits and model errors (ϵ) are calculated to quantify the performance of all tests. They are computed with respect to the true model $\mathbf{v}_{P|true}$ and to the observed data $\mathbf{p}_{obs}$. Due to the usage of the least-squares norm in the FWT algorithm, the data misfits are expressed as normalized squared L_2 norms. The model errors are normalized L_1 norms to ensure comparability with relative model deviations. In the following relations $\mathbf{p}_{init}$ and $\mathbf{p}_{result}$ denote the synthetic data for initial model $\mathbf{v}_{P|init}$ and final model $\mathbf{v}_{P|result}$, respectively:

$$\epsilon\left(\mathbf{v}_{P|init}\right) = \frac{\left\|\mathbf{v}_{P|init} - \mathbf{v}_{P|true}\right\|_1}{\left\|\mathbf{v}_{P|true}\right\|_1} \qquad \text{(initial model error)}, \qquad (6.1a)$$

$$\epsilon\left(\mathbf{v}_{P|result}\right) = \frac{\left\|\mathbf{v}_{P|result} - \mathbf{v}_{P|true}\right\|_1}{\left\|\mathbf{v}_{P|true}\right\|_1} \qquad \text{(final model error)}, \qquad (6.1b)$$

$$\epsilon(\mathbf{p}_{init}) = \frac{\|\mathbf{p}_{init} - \mathbf{p}_{obs}\|_2^2}{\|\mathbf{p}_{obs}\|_2^2} \qquad \text{(initial data misfit)}, \qquad (6.1c)$$

$$\epsilon(\mathbf{p}_{result}) = \frac{\|\mathbf{p}_{result} - \mathbf{p}_{obs}\|_2^2}{\|\mathbf{p}_{obs}\|_2^2} \qquad \text{(final data misfit)}. \qquad (6.1d)$$

Table 6.1: Application 3: List of all FWT tests for both the 1D and the Marmousi experiment. The terms "acoustic" and "viscoacoustic" are denoted by abbreviations "A" and "V".

FWT test	Figures of v_P results		Data	Modeling in FWT	Initial v_P-model	Passive Q_P-model
	1D model	Marmousi				
Reference	6.9a	6.15a	A	A	smooth	–
Test 1	6.9b	6.15b	V	A	true	–
Test 2	6.9c	6.15c	V	A	smooth	–
Test 3	6.9d	6.15d	V	V	smooth	true
Test 4	6.9e	6.15e	V	V	smooth	smooth
Test 5	6.9f	6.15f	V	V	smooth	homogeneous

6.2.1 Synthetic experiment: layered 1D model

The first experiment uses a 2D model with a 1D geology (hereinafter referred to as "1D model") consisting of four layers over a half-space: a water layer on top, followed by highly and weakly attenuative sedimentary rocks. The corresponding v_P and Q_P models are shown in Figure 6.6. Due to the occurrence of a thick layer with high attenuation on top of the sediments, this model represents a shallow marine environment. The acquisition geometry is located at the water surface and consists of 24 explosive sources with a spacing of 12 m as well as 278 hydrophones with a spacing of 1 m. For each shot gather all receivers are used. The resulting offsets range from 0.5 m to 292.5 m. The source signal is a Ricker wavelet with a peak frequency $f_{\mathrm{peak}} = 80$ Hz and the recording time of synthetic seismic data is 0.21 s. The total model size is 130×308 m with a spatial discretization of $\Delta h = 0.5$ m. However, due to the application of a perfectly matched layer (width = 15 m) in finite-difference modeling, all model related figures are limited to the relevant area (excluding the PML layer). General parameters are summarized by Table 6.2.

For viscoacoustic modeling I determine the relaxation parameters such that I get an optimal representation of constant Q_P within the bandwidth of the seismic data (see Figure 6.7b). Here, I define the bandwidth as the contiguous frequency range, wherein the decay of the amplitude spectrum with respect to the maximum amplitude is less than 10 dB (shaded areas in Figure 6.7). The range of phase velocity dispersion of all layers is estimated within this bandwidth (see exemplary dispersion for the second layer with $Q_{P,0} = 10$ in Figure 6.7c) and visualized by shaded areas in vertical sections across the 1D medium (Figure 6.6a). The aim is to analyze if the recovered v_P model can be explained by the minimum and maximum velocity dispersion.

The approximation of relaxation parameters is based on the reference frequency $f_0 := f_{\mathrm{peak}}$ and an average $Q_{P,0} = 74$ computed from the true quality factor model within the sub-seafloor area (Figure 6.6b). The acoustic velocity model $v_{P,\mathrm{ref}}$ (Figure 6.6a) is defined at f_0, i.e., no dispersion occurs at f_0. The bandwidth of the seismic data is limited to $\tilde{\Delta} f = [19.6, 141]$ Hz. This corresponds to a dynamic range of 2.8 octaves. Based on the rule of one relaxation mechanism per octave (Blanch et al., 1995), I use three relaxation mechanisms. The resulting optimal set of relaxation frequencies is $f_{\mathrm{r},l} = (1.202, 17.62, 179.4)$ Hz. I found that it is not necessary to obtain relaxation frequencies for all quality factors of the true model, i.e., it is not essential to provide models containing relaxation frequencies. Using the relaxation parameters computed from the average $Q_{P,0} = 74$, I obtain quite accurate approximations for all quality factors given in the 1D model. An exemplary quality factor approximation is shown for the second layer with $Q_P = 10$ (Figure 6.7b). Obviously, the deviation in corresponding phase velocity dispersion is negligible (compare Figure 6.7c).

The L_1 based Q_P approximation error is given with respect to constant $Q_{P,0}$ and is quantified within the seismic bandwidth. For all layers, I obtain acceptable approximation errors of about 3 %. While the approximation at reference frequency is perfect, the largest errors can be observed at the upper end of the bandwidth. These errors are nearly identical to those of the accurate Q_P approximation using exact Q_P values of each layer.

A comparison of acoustic and viscoacoustic data computed from the true model at a central shot is shown in Figure 6.8. Due to very weak attenuation in water, the direct waves are nearly identical. At near offsets one can observe a quite good match in phases and amplitudes of the seafloor reflection. However, at larger offsets and for all later reflection

Table 6.2: Application 3, 1D experiment: General setup.

Application 3: 1D experiment		
Attributes		**Specifications**
Model	**Size**	width: 308 m, height: 130 m
	Structure	· 1D geology (v_p = [1700, 2600] m/s) · with water layer (v_p = 1500 m/s)
	Figure	Figure 6.6
Acquisition geometry	**Location**	on top of half-space, sea surface
	Configuration	reflection geometry
	Sources	· 24 explosive sources · source-time function: Ricker wavelet · peak frequency: 80 Hz
	Receivers	· 278 hydrophones · recording length: 0.21 s
	Offsets	minimum: 0.5 m, maximum: 292.5 m
Modeling parameters	**Model size**	· width: 616 grid points, height: 260 grid points · grid spacing: 0.5 m
	Propagation time	· 2625 time steps · time discretization: $8 \cdot 10^{-5}$ s
	PML boundary	· width: 15 m · free surface
	Relaxation parameters	· relaxation frequencies: (1.20, 17.6, 179) Hz · reference frequency: 80 Hz
	Dispersion	Figure 6.6a
General inversion parameters	**Initial models**	Figure 6.6a
	Passive models	Figure 6.6b
	Preconditioning	· wave-field based · user-defined: water layer is known
	Model update	· hard constraint: v_p limit: [1300, 3150] m/s · maximum deviation from the initial model: 25 %
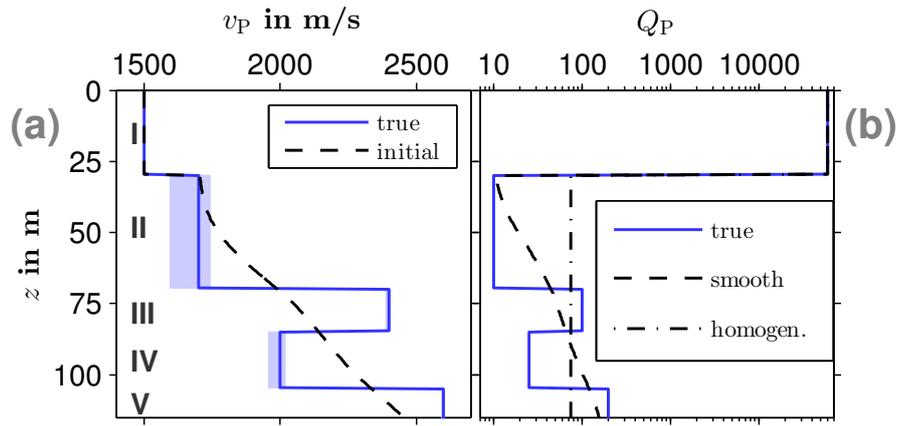
Figure 6.6: Application 3, 1D experiment: Vertical cross sections of the 1D models: (a) true and initial velocity model. The range of phase velocity dispersion due to attenuation is illustrated by shaded areas. The layers are labeled with Roman numbers (water layer and half-space are denoted by "I" and "V", respectively). (b) shows true as well as smooth and homogeneous passive Q_P models ("homogeneous" with respect to the sub-seafloor area).
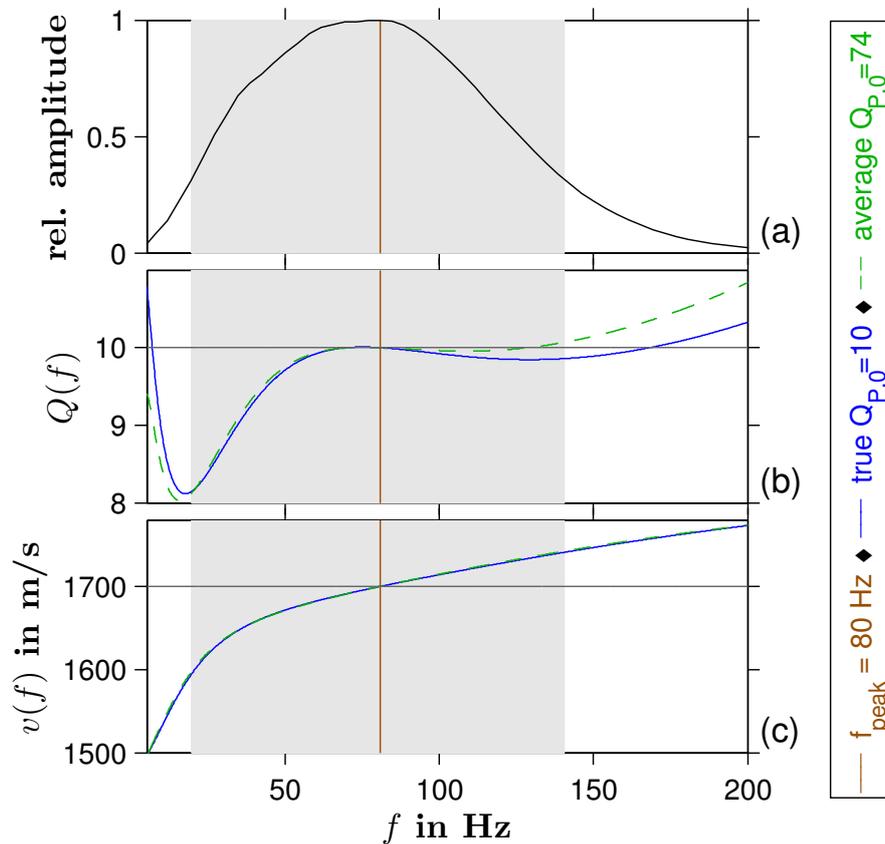
Figure 6.7: Application 3, 1D experiment: Approximation of the quality factor. (a) shows the normalized amplitude spectrum of all observed viscoacoustic data. (b) and (c) illustrate the quality factor approximation and phase velocity dispersion for the second layer using the relaxation frequencies for $Q_{P,0} = 10$ and the average $Q_{P,0} = 74$. The shaded areas represent the bandwidth with respect to the observed data. The solid and dashed lines show $Q(f)$ and corresponding phase velocity dispersion for the same relaxation frequencies $f_{r,l} = (1.202, 17.62, 179.4)$ Hz but true and average Q_P, respectively.
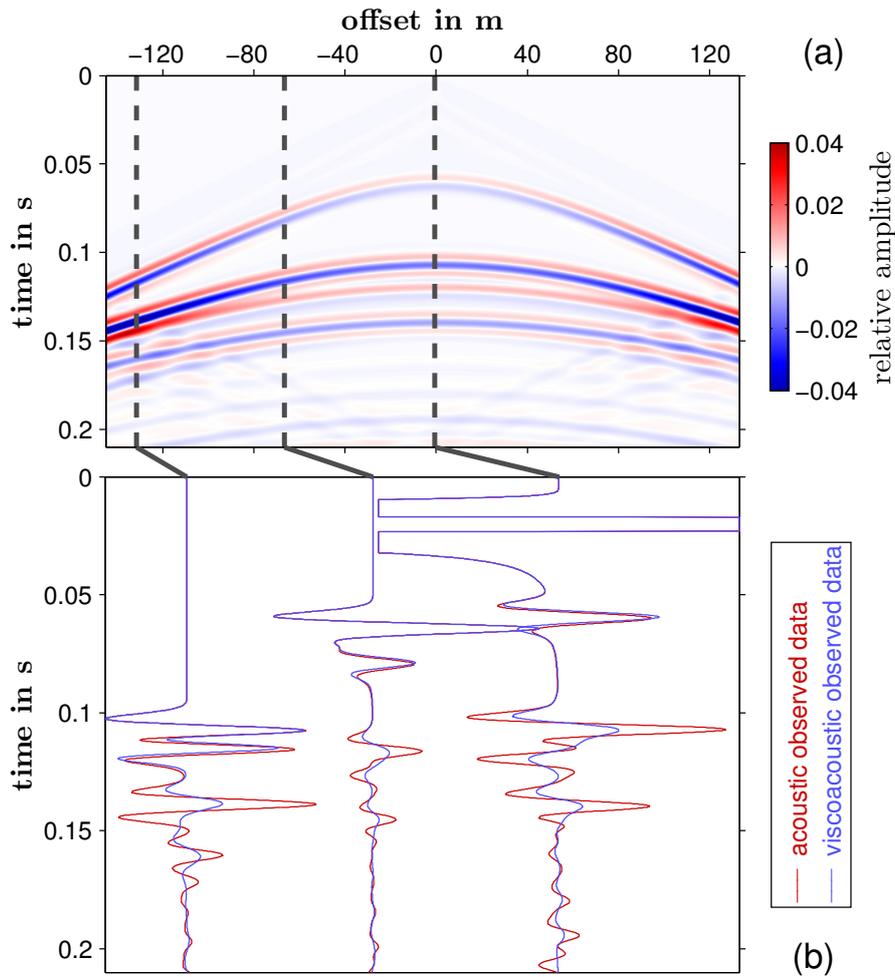
Figure 6.8: Application 3, 1D experiment: (a) shows the difference between acoustic and viscoacoustic data for the central shot located at $x = 160$ m. It exhibits true amplitudes which are clipped to ± 4 % of the maximum acoustic amplitude. (b) illustrates traces at exemplary offsets. For a better visualization, they are normalized independently for each offset and the direct wave of the zero-offset trace is clipped. Acoustic and viscoacoustic amplitudes are still comparable.

events there are significant differences between acoustic and viscoacoustic data. The phase misfit is explained by the highly dispersive properties of the second layer.

The smooth initial v_P and passive Q_P models for waveform tomography are generated by the application of a 2D Gaussian filter (size 100×100 m, $\sigma = 33$) to the sub-seafloor area of the true model (see Figures 6.6a and 6.6b). Figures 6.9a to 6.9f show the results obtained by the reference FWT as well as from acoustic FWT of viscoacoustic data (compare Table 6.1), which will be discussed in the following. All sub-figures contain auxiliary plots of minimum and maximum velocity dispersion with respect to the true v_P and Q_P models. The vertical section of all inversion results is computed by lateral averaging of a representative model area within the interval $x = [80, 228]$ m. This avoids the involvement of unreliably recovered velocities, mainly related to the areas close to lateral model boundaries. Figure 6.10 depicts the evolution of the corresponding data misfits and model errors.

A reliable interpretation of the effects caused by attenuation can be done by computing a reference result which comprises acoustic inversion of pure acoustic observed data. The

nearly optimal conformity of the true and the final model (Figure 6.9a) ensures the resolving power of FWT with respect to the given model and geometry. The reference FWT is characterized by the strongest reduction of both data and model error (see Figure 6.10). It stopped after 1762 iterations. Due to the computation of too small step lengths and the limited accuracy of single precision, the model update stagnated.

Test 1 investigates the effect of neglecting Q_P information in an acoustic FWT applied to viscoacoustic data (Figure 6.9b). In spite of using the true v_P model as initial model, the FWT starts at the highest initial data misfit (Figure 6.10a) which is reduced at the expense of the accuracy of the velocity model. On the one hand, the interface locations are still recognizable. But, on the other hand, both a significant model error (increasing from 0 % to 4.5 %, see Figure 6.10b) and an artificial alteration of the velocity model can be observed. In particular, layer III is smeared heavily. Omitting Q_P information results in a failure of the FWT. Only layer II is recovered within the maximum range of velocity dispersion. This indicates that the effects of attenuation might be negligible in near-surface areas of the given model. The FWT of test 1 stops after 61 iterations due to the inability of computing a meaningful step length.

Test 2 represents a common FWT application (Figure 6.9c). In this case I neglect Q_P and use the smooth initial v_P model (Figure 6.6a). The final velocity model is recovered insufficiently, which is caused by a bad fit of observed and synthetic data. While the final v_P model shows some improvements, it is still very similar to the initial model. Both the data and the model error are reduced slightly producing a smooth velocity model. Furthermore, the large-scale structures are comparable to the result of test 1. This implies that both inversions got stuck in a neighboring local minimum of the data misfit function. The FWT of test 2 stops after 77 iterations due to the inability of computing a meaningful step length.

Test 3 includes the true Q_P model (Figure 6.6b) as a passive model parameter. It can be clearly seen that the result resembles the acoustic reference result (Figure 6.9d). The velocity of layer II is explained within the range of relevant phase velocity dispersion (see Figure 6.7c). Especially in case of low attenuation (layer III and half-space), one does not observe this effect. Apart from the reference FWT, test 3 is characterized by the strongest reduction of both the data misfit and model error (see green plots in Figure 6.10a and b). This verifies the methodology of combining viscoacoustic modeling with the acoustic inversion scheme, i.e., the gradient computation (C.25a) and model update (C.44) use the relaxed model parameter but are based on the acoustic equations without any modification. The revocation of relaxation (2.30) yields an inversion result comparable to the reference result. The FWT of test 3 stops after 4191 iterations due to the threshold stop criterion mentioned above.

Test 4 shows a more realistic FWT application (Figure 6.9e). I use the smooth passive Q_P model (Figure 6.6b). The upper model areas are recovered quite well. In contrast to the previous result, one can observe a decreasing resolution of the velocity model with increasing depth. However, there is a remarkable reduction of both data and model error (see Figure 6.10) resulting in a qualitatively good identification of layers and interfaces. The FWT of test 4 stops after 195 iterations due to the inability of computing a meaningful step length.

Test 5 deals with the simplest case of implementing attenuation (Figure 6.9f). Here, I use a homogeneous passive Q_P model, i.e., this model consists of the water layer over a homogeneous half-space with the average quality factor $Q_P = 74$. This test yields the smallest misfit reduction among all tests with viscoacoustic modeling (Figure 6.10a) and a poor recovery of the velocity model which still is characterized by a high similarity to the initial model (compare Figure 6.10b). The FWT of test 5 stops after 111 iterations due to the inability of computing a meaningful step length.

The observations for all tests coincide with the evolution of both the data misfit and the model error with respect to FWT iterations (Figure 6.10a). The neglection of attenuation in test 1 and test 2 results in inappropriate velocity models. Surprisingly, the usage of true v_P as initial model in test 2 causes a higher initial data misfit. However, both FWT tests end at the same high misfit level failing in the attempt to explain viscoacoustic data with acoustic
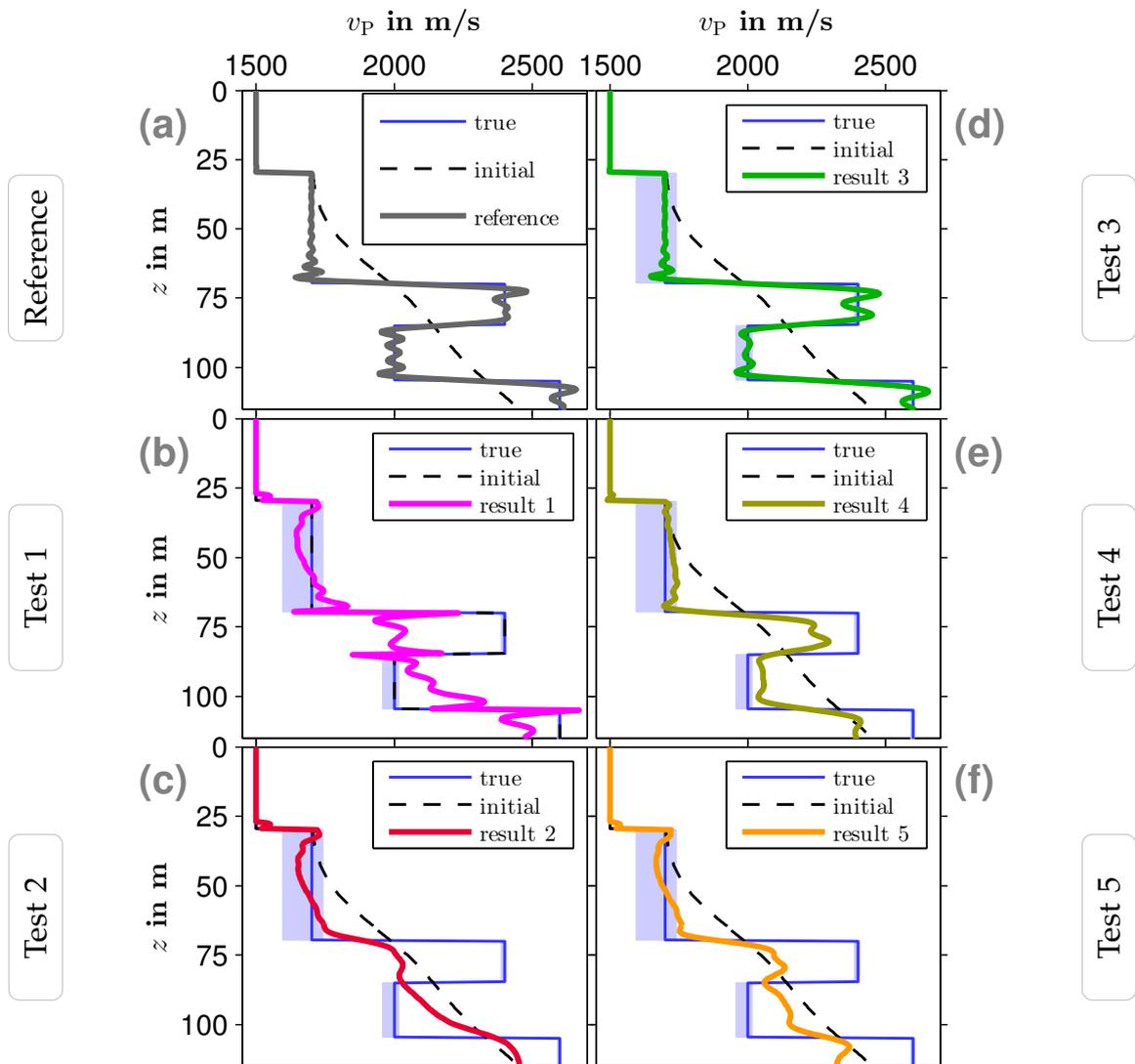
Figure 6.9: Application 3, 1D experiment: (a) shows the FWT result for the acoustic reference computation. (b) to (f) illustrate the results of acoustic FWT applied to viscoacoustic data (tests 1 to 5). The shaded areas denote phase velocity dispersion.

modeling. Furthermore, the homogeneous Q_P model in test 5 yields a misfit evolution being nearly identical to test 2. In spite of incorporating attenuation, the result is very similar to the velocity model obtained in test 2. Obviously, in case of the 1D experiment, the application of a homogenous passive $Q_P = 74$ – which is an incorrect representation of the subsurface – is insufficient for a successful v_P reconstruction. In contrast, the usage of the true Q_P model in test 3 results in a continuous reduction of the data misfit. This indicates the most optimal convergence to the global minimum of the misfit function. An acceptable trade-off is achieved by using the smooth Q_P model in test 4. In practice, a good Q_P model is usually unknown. Empirical relations can be used to derive a Q_P model from the initial velocity model. However, in general, they do not account for all rock types and physical conditions occurring in the given subsurface.

As mentioned above, a certain choice of a homogeneous passive Q_P model might cause unsatisfactory results. Additionally, I performed an inversion using a homogeneous passive model with $Q_P = 10$ (representing layer II in Figure 6.6b). In contrast to test 5, the inverted v_P model shows a much better recovery of the upper layers. The reconstruction of layer II is nearly identical to the reference result. Layers III and IV show a small v_P overestimation and underestimation, respectively. The average deviation from the true model is less than 3 %. However, the velocity of layer V is characterized by a large overestimation of about 13 %. Consequently, the passive Q_P model should rather be a good representation of the upper structures of the model.

Furthermore, apart from layer II, velocity fluctuations can not be explained by velocity dispersion. In the most optimal result of test 3, they rather show a high similarity to those of the acoustic reference result. Hence, they are caused by other effects, such as a worse illumination with increasing depth. Additionally, this might indicate the resolution limit of FWT. For all layers the extension of one fluctuation "cycle" is within the sub-wavelength area (approximately one-fourth of the dominant wavelength).
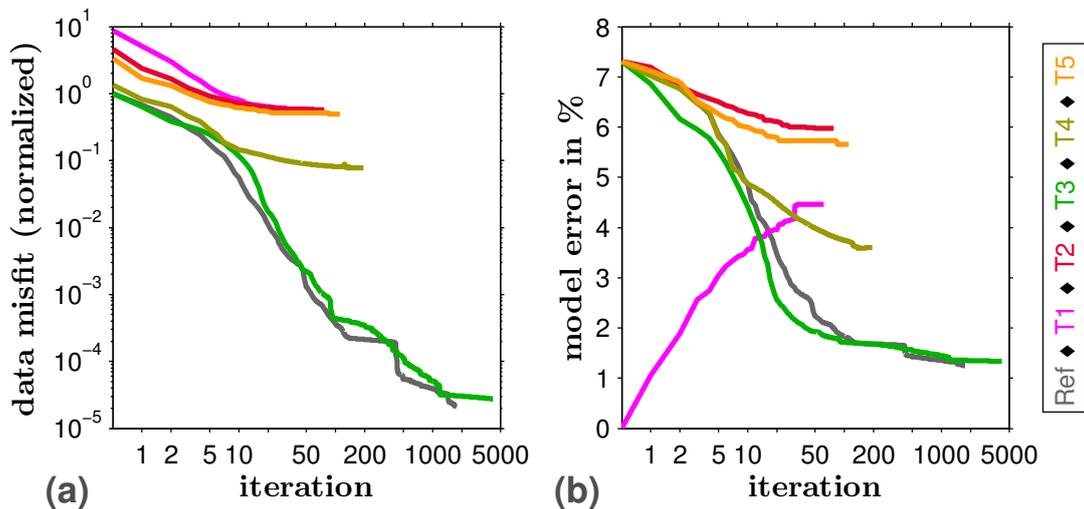
Figure 6.10: Application 3, 1D experiment: (a) shows the evolution of the L_2 based data misfit with respect to the number of iteration. The data misfit of the reference FWT ("R") is normalized to its own initial value. The data misfits of tests 1 to 5 ("T1" to "T5") are normalized to the initial value of test 3 due to its best comparability with the reference FWT. (b) illustrates the evolution of corresponding L_1 based model errors with respect to the true v_P model. Color coding is equivalent to Figure 6.9.

6.2.2   Synthetic experiment: Marmousi model

Using a modified section of the Marmousi-II model (Martin, 2002; Martin et al., 2006, shown in Figure 6.11) (based on Versteeg, 1994), I repeat all previous investigations. To reduce computational efforts velocities are clipped to $v_P = [1.5, 4]$ km/s (Figure 6.11) which affects the deep salt layer extended from $x \approx 6$ km to $x = 10$ km. The acquisition geometry consists of 32 explosive sources and a maximum number of 300 hydrophones per source. It forms a marine streamer geometry at the water surface moving from the right to the left model boundary (Figure 6.12). I choose a streamer length of 5980 m, a receiver spacing of 20 m and a nearest offset of 45 m. However, due to the existence of the right model boundary, only the receiver arrays for shots 20 to 32 provide the full streamer length, while shot one is equipped with the shortest streamer containing 18 receivers only. The source time function is a Ricker wavelet with a peak source frequency $f_{\text{peak}} = 9$ Hz. The recording time of seismic data is
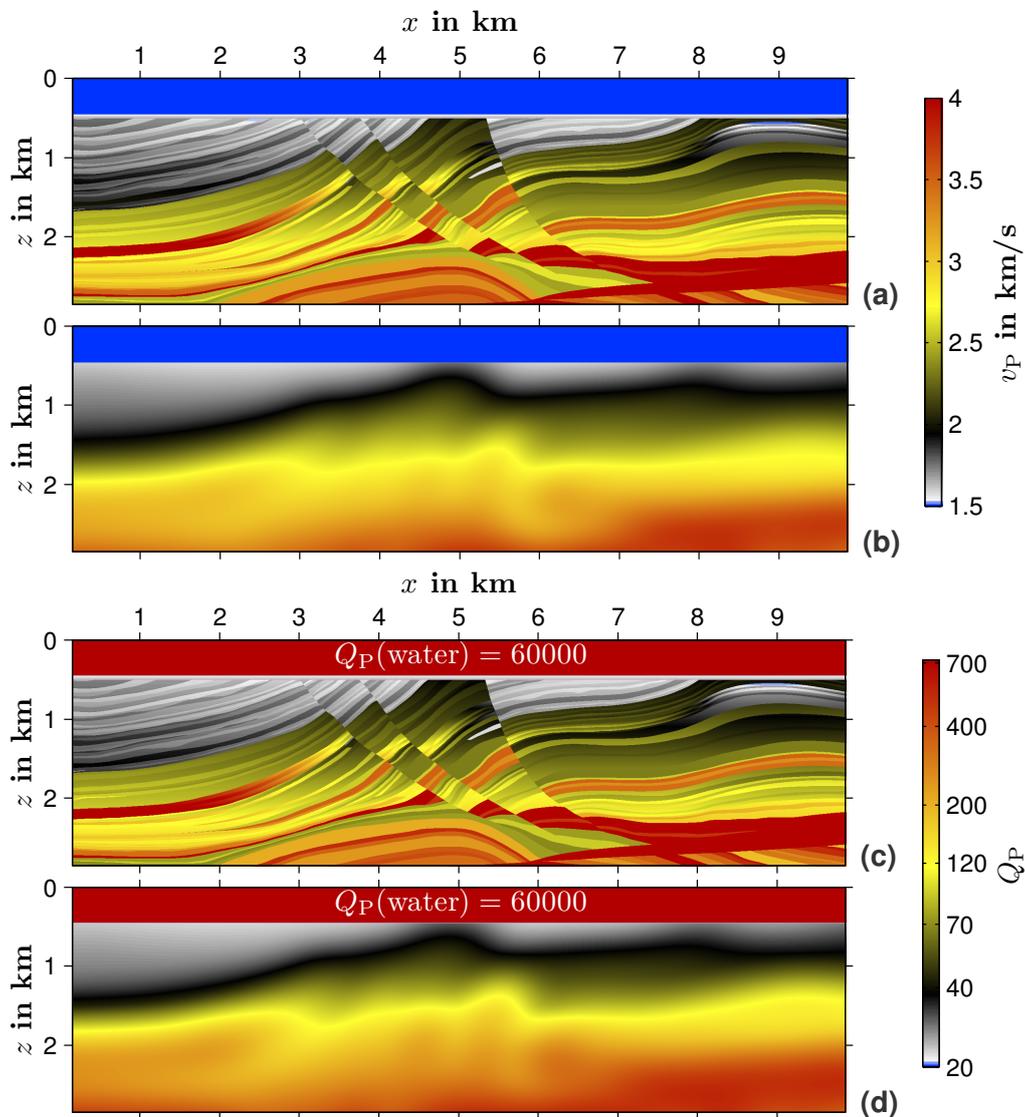
Figure 6.11: Application 3, Marmousi experiment: (a) shows the true v_P model and (b) depicts the initial v_P model. (c) and (d) illustrate the true Q_P model and the smooth passive Q_P model.

lateral distance in km
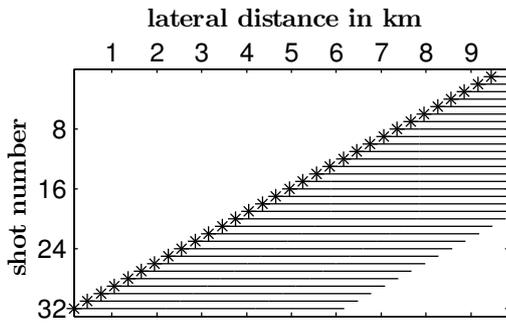
Figure 6.12: Application 3, Marmousi experiment: Marine streamer acquisition geometry equipped with a maximum number of 300 hydrophones per receiver line. While the receiver line of shot 1 exhibits the shortest length (18 receivers), the full streamer length is available in case of shots 20 to 32.
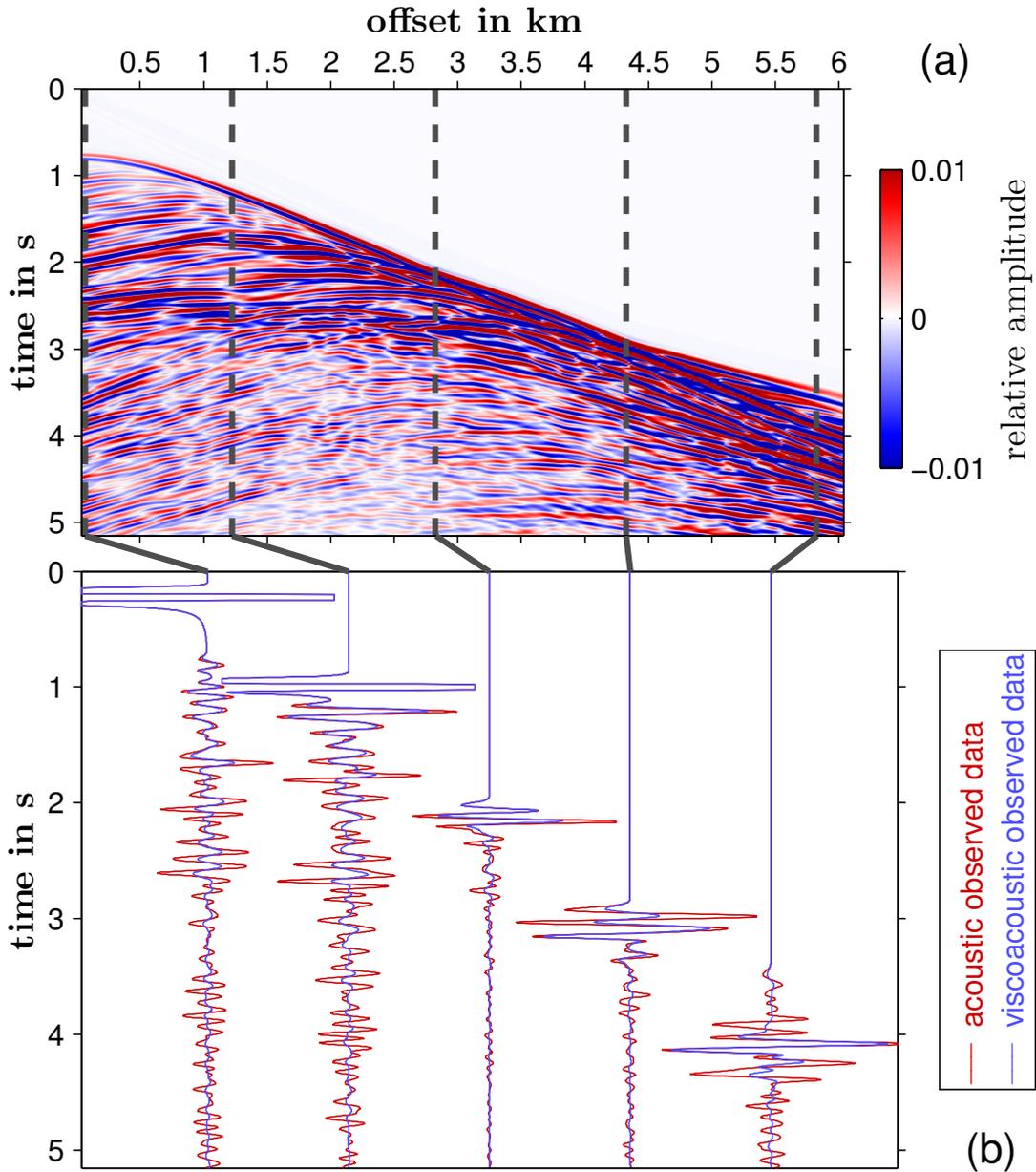
Figure 6.13: Application 3, Marmousi experiment: (a) shows the residual seismogram with true amplitudes which are clipped to $\pm 1\%$ of the maximum acoustic amplitude. The residuals are computed from acoustic and viscoacoustic data recorded at shot location $x = 2.6$ km. (b) illustrates observed traces at exemplary offsets. For a better visualization they are normalized independently for each offset. Acoustic and viscoacoustic amplitudes are still comparable.

5.15 s. The total model size is 3×10 km with a spatial discretization of $\Delta h = 5$ m. General parameters are summarized by Table 6.3.

The true Q_P-model (Figure 6.11c) is derived from the velocity model (Figure 6.11a) by applying an empirical v_P-Q_P relation (Hamilton, 1972):

$$\frac{1}{Q_P} = \alpha_P \frac{v_P}{\pi f - \frac{\alpha_P^2 v_P^2}{4\pi f}}. \tag{6.2}$$

The intrinsic attenuation α_P is assigned to the structures of the velocity model. I use laboratory α_P values for the frequency range of the given example in marine sedimentary layers (Attewell and Ramana, 1966) – ranging from $10^{-5} \frac{1}{m}$ to $10^{-3} \frac{1}{m}$. The quality factor ranges from 21 in the upper sedimentary layers to 707 in deeper high velocity zones.

The approximation of relaxation parameters is based on the reference frequency $f_0 := f_{peak}$, $L = 3$ and an average $Q_{P,0} = 62$ (harmonic mean) within the area beneath the seafloor. The bandwidth of the seismic data is limited to $\tilde{\Delta} f = [3.3, 16.5]$ Hz. This corresponds to a dynamic range of 2.3 octaves. Consequently, I use a sufficiently high number of three relaxation mechanisms. The resulting optimal set of relaxation frequencies is $f_{r,l} = (0.1513, 1.925, 18.94)$ Hz. Based on the true model, both acoustic and viscoacoustic data as well as their residuals are shown for shot 24 located at $x = 2.6$ km (Figure 6.13). For a better illustration, the residual seismogram is clipped to ± 1 % of maximum acoustic amplitude (Figure 6.13a) and data traces are normalized to the maximum amplitude of viscoacoustic observed data (Figure 6.13b). Especially with increasing offset or travel time, the misfit of acoustic and viscoacoustic waveforms is increasing, too. In particular, the amplitudes are significantly affected by attenuation. Furthermore, especially within the upper rock layers phase velocity dispersion can be observed. However, the minimum and maximum dispersion computed from (B.8) and shown in Figure 6.14 is related to zero frequency and infinite frequency. The dispersion with respect to the bandwidth is negligible.

For the Marmousi model I perform the same inversion tests as for the 1D experiment (Table 6.1). The smooth initial v_P model is generated by application of a 2D Gaussian filter (size 1250×1250 m, $\sigma = 51$) to the sub-seafloor area of the true model $v_{P|true}$ (see Figure 6.11b). While I use relation (6.2) to compute the smooth Q_P model (see Figure 6.11d) from the initial smoothed v_P model, the homogeneous passive Q_P model consists of the water layer over a half-space with the average quality factor $Q_P = 62$. All inversion tests are decomposed into multiple stages to reduce nonlinearity of the inverse problem: By application of low-pass filters the inversion is performed for five frequency ranges, moving from low to high frequencies (Sirgue and Pratt, 2004). The peak frequencies of low-pass-filtered data are $f_{peak} = (0.96, 1.7, 3.3, 4.4, 9.0)$ Hz.

The inverted velocity models are shown in Figure 6.15 and the corresponding relative deviations from the true model can be found in Figure 6.16. For a better visualization, the deviation images are clipped to ± 20 %. The actual maximum range is up to ± 50 %. This is caused by the lack of very high frequencies which are necessary to achieve a resolution at the scale of the finite-difference grid spacing. Consequently, very small scale structures and high-contrast interfaces cannot be recovered. Furthermore, Table 6.4 summarizes the data and model errors computed by equations (6.1). For all FWT tests, it compares the change between initial and final errors.

Table 6.3: Application 3, Marmousi experiment: General setup.

Application 3: Marmousi experiment		
Attributes		**Specifications**
Model	**Size**	width: 10 km, height: 3 km
	Structure	· modified Marmousi-II geology (v_p = [1550, 4000] m/s) · with water layer (v_p = 1500 m/s)
	Figure	Figure 6.11
Acquisition geometry	**Location**	on top of half-space, sea surface
	Configuration	marine reflection geometry
	Sources	· 32 explosive sources · source-time function: Ricker wavelet · peak frequency: 9 Hz · time delay of source signal: 0.45 s
	Receivers	· towed streamer · minimum: 18 hydrophones · maximum: 300 hydrophones · recording length: 5.6 s
	Offsets	minimum: 45 m, maximum: 5980 m
	Figure	Figure 6.12
Modeling parameters	**Model size**	· width: 2000 grid points, height: 600 grid points · grid spacing: 5 m
	Propagation time	· 8000 time steps · time discretization: $7 \cdot 10^{-4}$ s
	PML boundary	· width: 150 m · free surface
	Relaxation parameters	· relaxation frequencies: (0.151, 1.93, 18.9) Hz · reference frequency: 9 Hz
	Dispersion	Figure 6.14
General inversion parameters	**Initial models**	Figure 6.11
	Passive models	Figure 6.11
	Preconditioning	· wave-field based · user-defined: water layer is known
	Model update	· hard constraint: v_p limit: [1500, 4000] m/s · maximum deviation from the initial model: 50 %

Figure 6.14: Application 3, Marmousi experiment: Minimum dispersion (a) and maximum dispersion (b) with respect to the true v_P model in Figure 6.11a as well as zero frequency and infinite frequency. Due to high attenuation, the upper sedimentary rocks cause significant dispersion. The water layer is free from dispersion.

Again, as described for the 1D experiment, an acoustic inversion of pure acoustic data is used as the reference result (Figure 6.15a). Considering the quite low peak frequency of the data, I can observe a good match of the reconstructed model and the true model (compare relative model deviation in Figure 6.16a). Both data and model errors are reduced significantly (see Table 6.4). Apart from some notable data residuals at near offsets, the final synthetic data and the observed data match very well (see Figure E.1).

In case of neglecting Q_P information (test 1 and test 2) the FWT is unable to recover subsurface structures properly from viscoacoustic data (Figures 6.15b and 6.15c). The resolution decreases dramatically with increasing depth causing a poor final v_P model with high model and data errors (compare initial and final errors for test 1 and test 2 in Table 6.4 as well as in Figures 6.16b and 6.16c).

In the following, I discuss the inversion progress of test 2 in more detail. Figure 6.17 depicts the intermediate inversion results at the end of every frequency-filtering stage. Test 2 shows a satisfactory reconstruction of large-scale structures during the inversion of the low-frequency content (see Figures 6.17a to 6.17c). By including higher frequencies artifacts appear in the upper sedimentary structures, while there are no improvements in deeper regions. In contrast, one can observe a destruction of large-scale structures which have already been recovered (Figures 6.17e and 6.17f). The minimum model error is obtained within the fourth stage with a peak frequency $f_{peak} = 5.0$ Hz (Figure 6.17d). Throughout the remaining inversion progress the model error is increasing continuously, while the data misfit is decreasing. For a further investigation of this phenomenon I repeat test 2 without multi-stage approach. I invert for the full frequency content at once and obtain a similar final result. However, the FWT skips the computation of an accurate intermediate result and directly produces an artificial model. Only until the sixth iteration there is a marginal improvement of the model. Apparently, the inversion of low-frequency contents is less
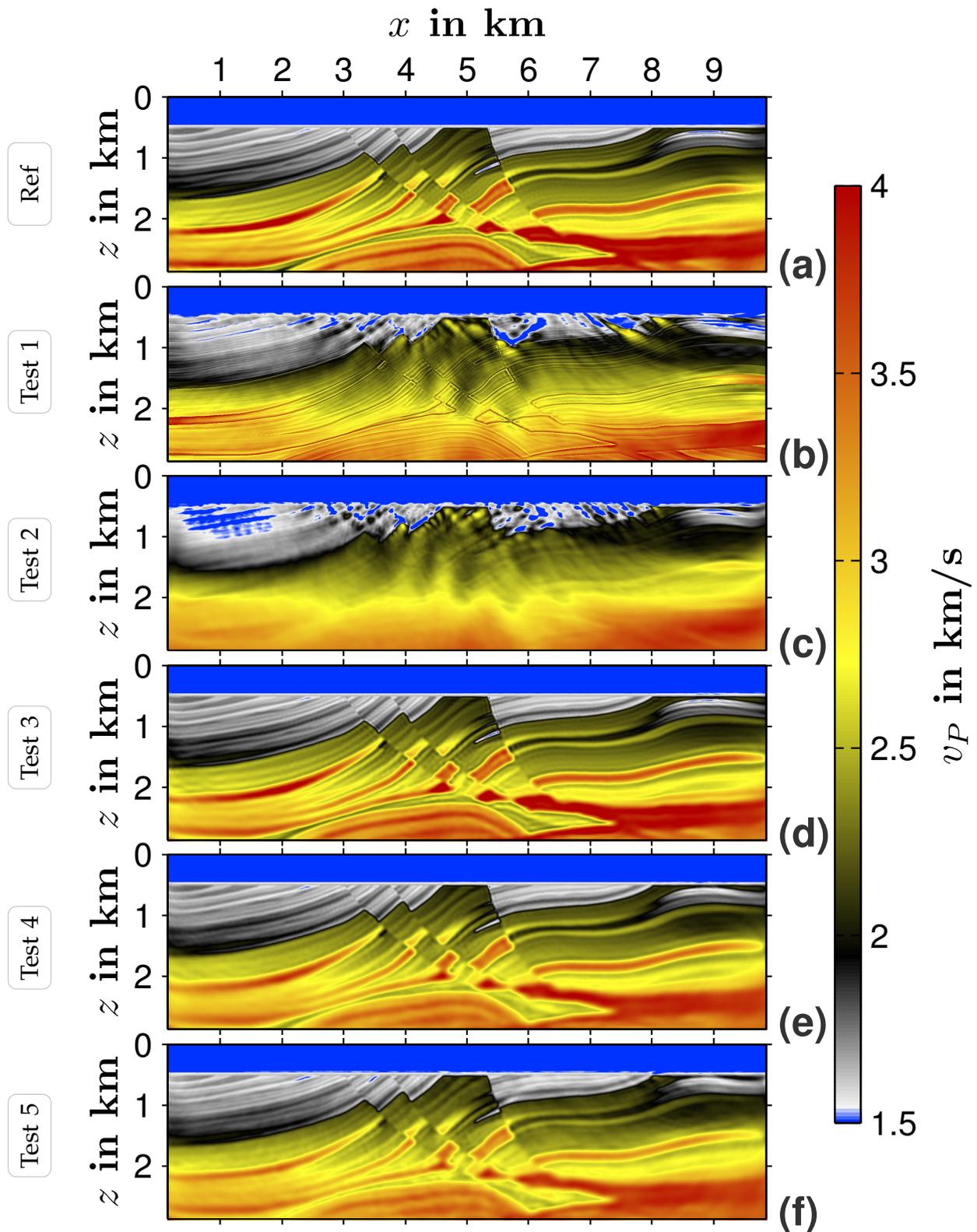
Figure 6.15: Application 3, Marmousi experiment: (a) to (f) show the recovered v_P models for acoustic reference FWT of pure acoustic data (a) and acoustic FWT of viscoacoustic data for tests 1 to 5 (b) to (f).
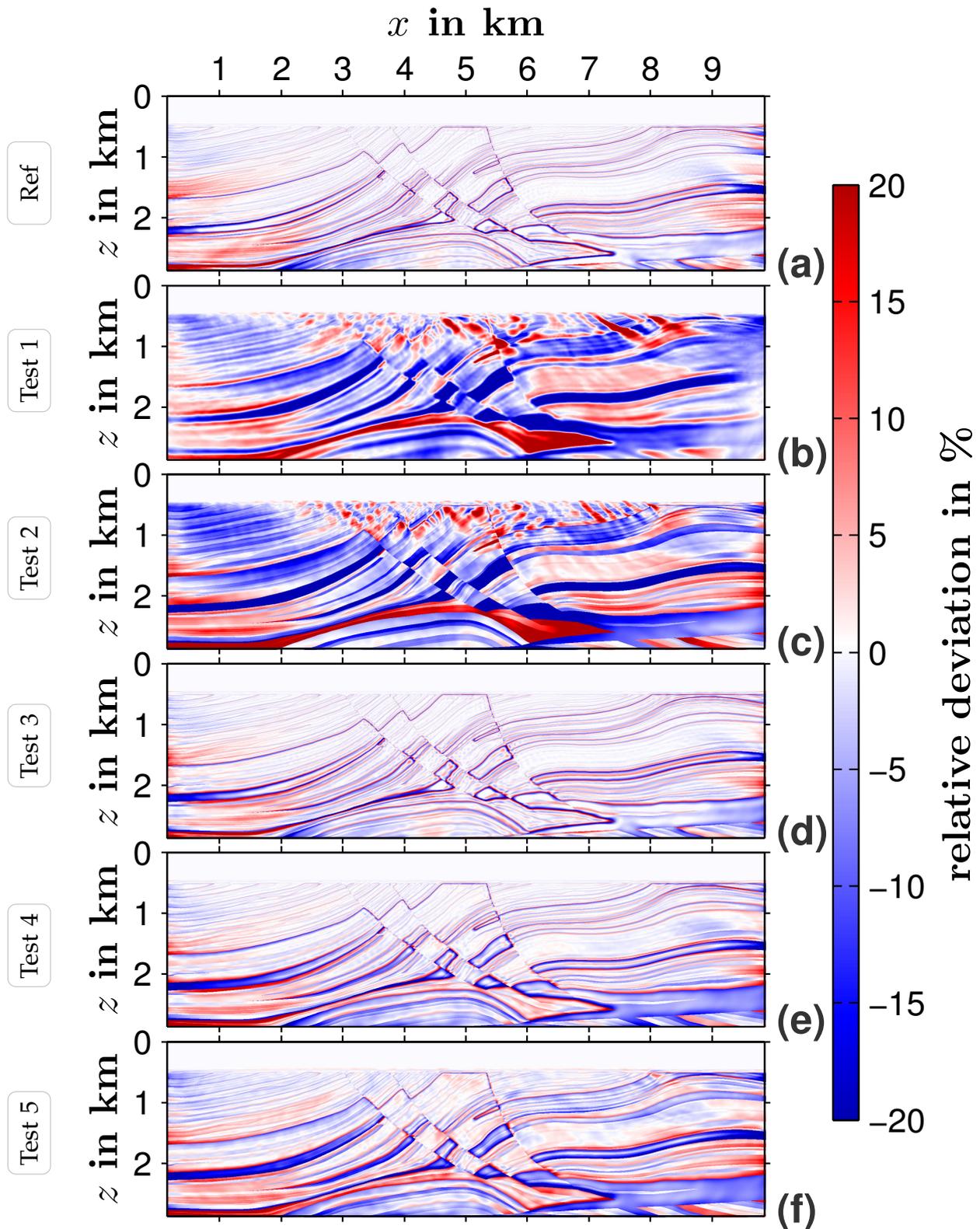
Figure 6.16: Marmousi experiment: (a) to (f) show the relative deviation of the results in Figure 6.15 with respect to the true model in Figure 6.11a. For a better visualization, the images are clipped at ± 20 percent of the relative deviation.
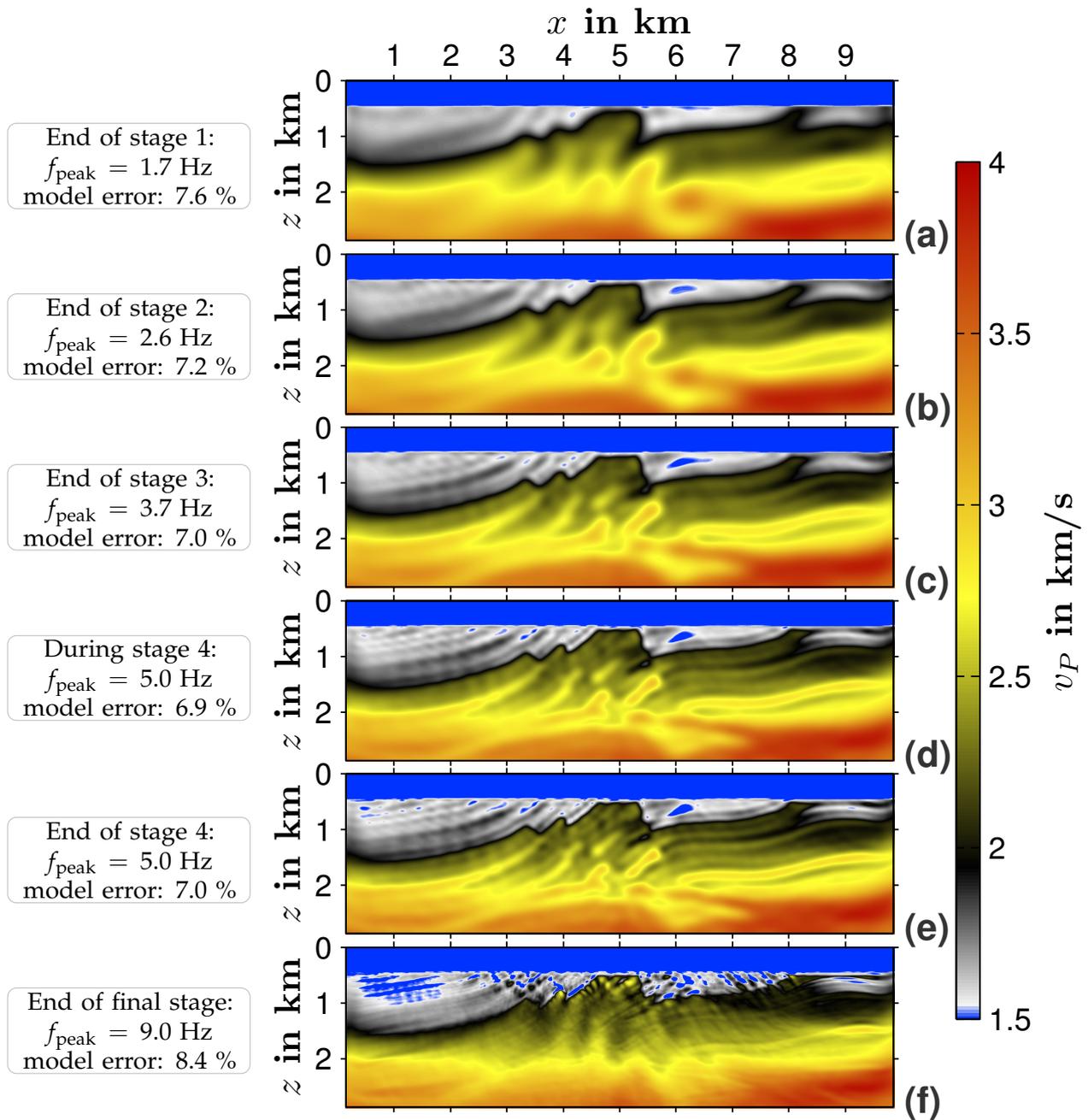
Figure 6.17: Application 3, Marmousi experiment: (a) to (f) show the evolution of the v_P model with respect to the inversion progress for the test 2. (a), (b), (c), (e) and (f) illustrate intermediate v_P models at the end of each stage of low-pass filtering. (d) shows the model with the lowest model error (for comparison: the model error of the initial model is 8.6 %). (f) corresponds to Figure 6.15c.

sensitive to attenuation. This observation coincides with inversion strategies in viscoacoustic frequency-domain FWT applied by several authors, such as Takam Takougang and Calvert (2011). They use a similar multi-stage approach. At low frequencies they only invert for v_P. Later on, a combined inversion for v_P and Q_P is applied.

Obviously, test 1 and test 2 demonstrate that the acoustic inversion of viscoacoustic data is independent of the initial v_P model within the framework of the Marmousi experiment. In both cases it ends at comparable artificial final v_P models with low reduction of data errors and an increase of model errors (see Table 6.4). The misfit between viscoacoustic observed data and acoustic synthetic data is mapped to the velocity model. While the final v_P model particularly explains phases of the seafloor reflection and some later reflection events, it causes significant amplitude mismatches for these events in the seismic data (see seismograms in Figures E.2 and E.3). However, later events and especially the refracted wave show remarkable phase deviations between synthetic and observed data. Obviously, the influence of phase velocity dispersion seems to increase with increasing depth or offset.

In contrast to test 1 and test 2, the involvement of a Q_P model improves the v_P recovery significantly. Using $Q_{P,true}$ as passive parameter, as performed in test 3, the reconstructed v_P model is comparable to the optimal acoustic reference result (Figures 6.15d and 6.16d) – proved by an excellent data fit (Figure E.4). Test 4 employs the smooth Q_P model and, thus, is the most realistic case. Considering this imperfect Q_P information, the FWT produces a satisfactory v_P model (see final result in Figure 6.15e and the model deviation in Figure 6.16e). Although the smooth Q_P model does not allow the reconstruction of a v_P model with high resolution, the comparison of test 4 with test 3 only shows a minor increase of the model error (Table 6.4). Even the implementation of a homogeneous Q_P information within the sub-seafloor region yields a surprisingly good result (Figures 6.15f and 6.16f). Apart

Table 6.4: Application 3, Marmousi experiment: List of errors ϵ with respect to initial model $\mathbf{v}_{P|init}$ and corresponding initial data $\mathbf{p}_{init}$ as well as for the resulting model $\mathbf{v}_{P|result}$ and corresponding data $\mathbf{p}_{result}$. Using equations (6.1) the errors are computed with respect to the true model $\mathbf{v}_{P|true}$ and observed data $\mathbf{p}_{obs}$. The arrows indicate the strength of error ratios $\epsilon(\mathbf{p}_{result})/\epsilon(\mathbf{p}_{init})$ and $\epsilon(\mathbf{v}_{P|result})/\epsilon(\mathbf{v}_{P|init})$.

FWT	Data error with respect to observed data (in %)		Model error with respect to true model (in %)		Change of	
	$\epsilon(\mathbf{p}_{init})$	$\epsilon(\mathbf{p}_{result})$	$\epsilon(\mathbf{v}_{P|init})$	$\epsilon(\mathbf{v}_{P|result})$	data error	model error
Reference	53.4	0.0659	8.6	2.9	▼	▼
Test 1	28.8	14.2	0.0	7.4	◤	▲
Test 2	45.2	16.1	8.6	8.4	◤	➤
Test 3	12.4	0.0204	8.6	3.1	▼	▼
Test 4	13.2	0.186	8.6	4.1	▼	▼
Test 5	19.2	2.69	8.6	5.0	◤	◤

from decreasing v_P resolution with increasing depth, the qualitative Marmousi geology is clearly noticeable. Both the smooth and homogeneous Q_P models allow a quite good explanation of phases of the seismic waveforms (see seismograms in Figures E.5 and E.6 for a streamer laterally crossing the fault structures of the Marmousi model). However, especially the homogeneous Q_P model causes significant amplitude misfits at large offsets. This model does not properly take account of the 2D structured subsurface. Thus, amplitude errors accumulate with increasing offset.

In general, the Marmousi experiment resembles the results of the 1D experiment. In case of test 5, I observe different performances of the FWT. For both examples, there is a nearly identical reduction of the data misfit. However, in case of the Marmousi experiment, there is a significantly stronger reduction of the model error (42 % versus 25 % for the 1D example). This observation is explained by the choice of the homogeneous Q_P model. For example, there is a huge Q_P discrepancy between the second layer of the true 1D model and homogeneous model (true $Q_P = 10$ versus passive $Q_P = 74$). This causes incorrect data and an artificial model reconstruction. In contrast, there is a much better match of the homogeneous model and the upper structures of the Marmousi model (the arithmetic mean value of the true model and homogeneous $Q_P = 62$ are equivalent for depths 470 m $\leq z \leq$ 1830 m). Concluding, a good Q_P model should be a good representation of the "near-seafloor" regions.

6.2.3 Computational efforts

The implementation of viscoacoustic time-domain modeling in acoustic FWT comes along with increased computational efforts. The FWT computations are performed on the high-performance computer *JUROPA* at the Jülich Supercomputing Centre. Its compute nodes consist of eight CPU cores. Thus, for most optimal performance I decomposed the Marmousi model into eight sub-domains and computed 16 shots at once. This resulted in an allocation of 128 CPU cores. For each finite-difference modeling, I used 8000 time steps and a model dimension of 600×2000 grid points. In time-domain FWT, the forward-propagated wavefield has to be stored in advance. During the back-propagation, it is cross-correlated with the back-propagated wavefield. However, it is not necessary to store the spatial wavefield snapshots of forward-propagation at all time steps. A subset can be selected by using a very strict Nyquist criterion. The corresponding sampling rate is estimated as follows:

$$\Delta t_{\text{samp}} = \frac{1}{8 f_{\text{max}}} \quad \text{with} \quad f_{\text{max}} \approx 2 f_{\text{peak}}. \tag{6.3}$$

Consequently, the FWT required 728 snapshots resulting in a feasible memory consumption of 416 MB per CPU core. In total, the FWT applied 112 modelings per iteration. They are composed of 64 forward- and back-propagations as well as 48 test modelings for step length optimization (three test modelings and application of every second shot). In case of the acoustic FWT with only acoustic modeling, the computational time of one iteration amounted to 142 s. In contrast, the computational costs of viscoacoustic modeling were significantly higher (340 s per iteration).

6.3   Summary

In this work, I investigate the impact of intrinsic attenuation on 2D acoustic full waveform tomography in the time domain. The acoustic inversion scheme is applied to two viscoacoustic datasets generated for a 1D structured model and the Marmousi model. Using these examples, I assigned realistic quality factors to the subsurface models.

In the presence of soft rocks with pronounced absorption I observe a poor recovery of the velocity model. The neglection of attenuation causes an unsuccessful recovery of the v_P model. The attenuation-related data misfit is mapped to the velocity model by generating remarkable artifacts. If I use the true velocity model as an initial model, then the footprint of attenuation can be clearly observed in the artificially altered v_P model.

By considering an appropriate attenuation model – i.e., applying a passive Q_P model – in the forward modeling of the FWT, the accuracy of the reconstructed velocity model improves significantly in both cases. The usage of a smooth Q_P model results in a sufficient near-surface recovery of v_P but the resolution is decreasing with increasing depth. This is due to an attenuation-related loss of high-frequency information with increasing depth. Depending on the deviation from the true Q_P model, the choice of a homogeneous Q_P model increases the risk of an unsatisfactory v_P recovery (see 1D experiment).

In case of the appearance of soft sediments, the FWT has to take attenuation into consideration. The availability of a sufficiently good passive quality factor model allows the reconstruction of a reliable velocity model by applying the acoustic inversion scheme. However, such an appropriate good model does not necessarily have to be characterized by a high complexity being close to the true model. For example, the passive involvement of Q_P, which is derived from the initial v_P model, might significantly improve the resolution of the v_P model, provided that the Q_P model is at least an appropriate representation of the uppermost subsurface structures. In conclusion, it is not advisable to neglect attenuation or to use potentially poor attenuation information in FWT applications to real data recorded in marine environments with soft sediments. The results of this study suggest to consider attenuation as a smooth background modeling parameter to improve velocity model building by a purely acoustic inversion scheme in reflection seismic configurations.

Chapter 7

Application 4: 3D acoustic FWT in the time domain

The 2D applications discussed in previous chapters assume subsurface models with a 2D geology. However, in practice, the general structure of the subsurface is unknown. The occurrence of a 1D or 2D geology allows a meaningful application of a 2D FWT, which only requires a 3D-to-2D transformation of the recorded data to account for the different geometrical spreading. However, a 2D FWT yields an inaccurate reconstruction of the velocity model by projecting events in the seismic data, which arise from significant 3D heterogeneities, to the 2D model. Under these circumstances, the application of a 3D FWT is mandatory. In this chapter I present the preliminary results of a 3D acoustic FWT in the time domain applied to two experiments containing a 3D geology: a random medium with cross-well configuration and a 3D expansion of the Marmousi model with reflection configuration.

The implementation of 3D FWT is still in development and shows the following setup with respect to both experiments:

The application of certain FWT features and existing limitations of the current implementation due to the preliminary state of development result in the following inversion setup:

1. density is neglected, the FWT only inverts for P-wave velocity v_P,

2. the true source signal is used, i. e., no inversion for source time function takes place,

3. application of gradient preconditioning to suppress artifacts related to the acquisition geometry (spherical tapers in case of the cross-well experiment [using the cosine-taper 2.19] and wavefield-based preconditioning in case of the Marmousi experiment [using equation 2.25]),

4. application of a user-defined gradient preconditioning: no model update within the water layer of the Marmousi model due to known P-wave velocity v_P,

5. application of multiple-stage approaches within the workflow implementation, consideration of sequential frequency filtering,

6. usage of automatic criteria with respect to the misfit reduction for shifting within the workflow stages or to stop the FWT.

123

7.1   3D FWT for a cross-well geometry

7.1.1   Random medium model and geometry setup

The transmission experiment employs the random medium model with a size of $160 \times 160 \times 256$ meters. The model is generated by applying an exponential autocorrelation function with a correlation length of 40 m (Figure 7.1), where v_P varies between $1600 \frac{m}{s}$ and $2400 \frac{m}{s}$. The resulting model is a self-similar medium and includes structures on different scale lengths.
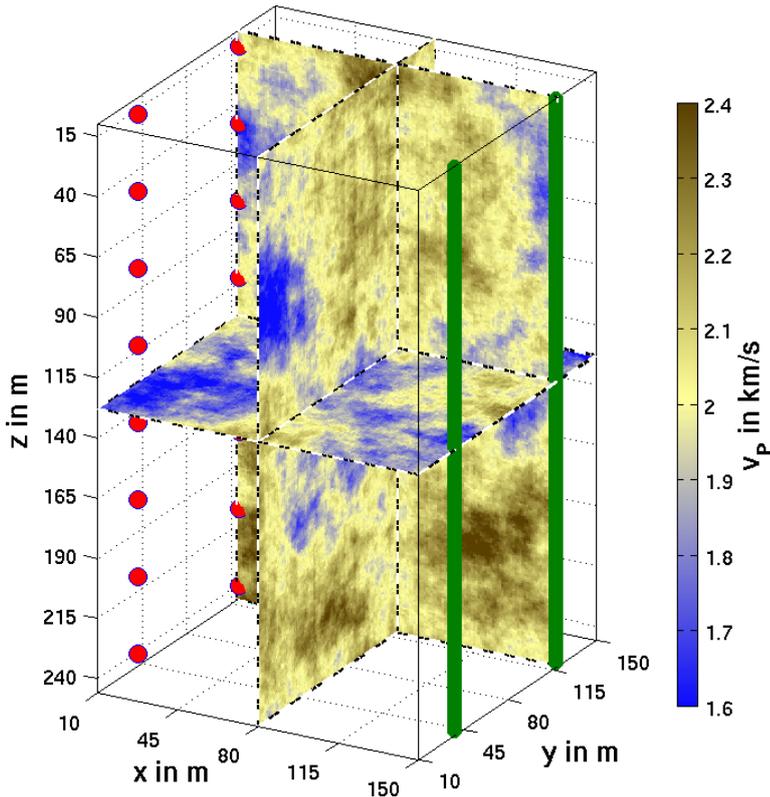
Figure 7.1: Application 4, cross-well experiment: True v_P model consisting of a random medium. Exemplary cross sections are shown at $x = 80$ m, $y = 120$ m and $z = 128$ m 16 explosive sources (red markers) and 470 hydrophones (green lines) are arranged in a transmission configuration.
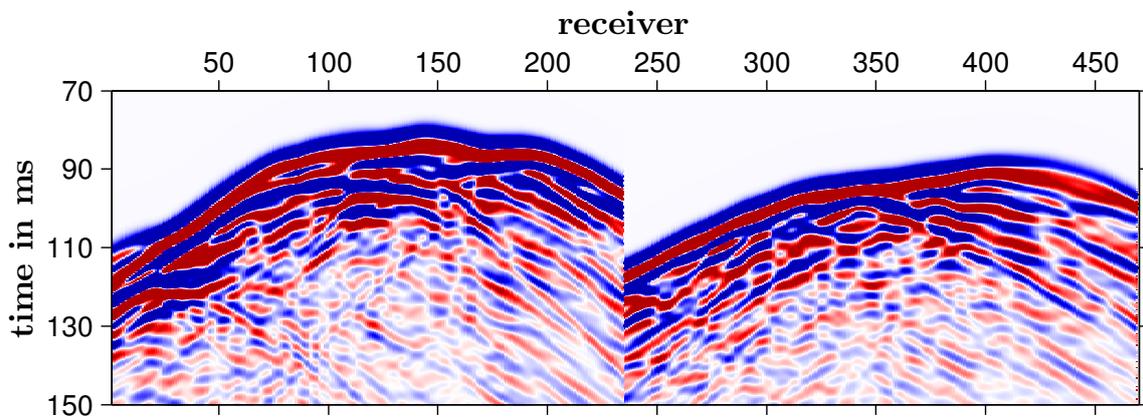
Figure 7.2: Application 4, cross-well experiment: Pressure data for an exemplary source located at $(x, y, z) = (11, 40, 144)$ m. The data are recorded in both receiver boreholes at $y = 40$ m (left part) and $y = 120$ m (right part). For a better visualization, the data are normalized and clipped.

Table 7.1: Application 4: General setup of the 3D cross-well experiment.

Application 4: 3D random medium		
Attributes		**Specifications**
Model	**Size**	lateral dimensions: 160 x 160 m, height: 256 m
	Structure	random medium
		average v_P = 2000 m/s maximum range: v_P = [1600, 2400] m/s standard deviation σ = 0.08 correlation length: 40 m exponential autocorrelation function
	Figure	Figure 7.1
Acquisition geometry	**Location**	full-space, borehole
	Configuration	transmission geometry (crosshole)
	Sources	· 16 explosive sources distributed to 2 boreholes · source-time function: Ricker wavelet · peak frequency: 150 Hz · time delay of source signal: 0.02 s
	Receivers	· 470 hydrophones distributed to 2 boreholes · effective recording length: 0.18 s
	Offsets	minimum: 138 m, maximum: 279 m
	Figure	Figure 7.1
Modeling parameters	**Model size**	· lateral: 320 x 320 grid points, height: 512 grid points · grid spacing: 0.5 m
	Propagation time	· 2000 time steps (effective recording length + time delay) · time discretization: $1 \cdot 10^{-4}$ s
	PML boundary	width: 10 m
General inversion parameters	**Initial model**	homogeneous: v_P = 2000 m/s
	Gradient computations	· imaging condition in time domain · taper at sources and receivers · additional taper within PML boundary
	Model update	· hard constraint: v_P limit: [1600, 2400] m/s · maximum deviation from the initial model: 25 %
	Workflow	multi-stage frequency filtering

The acquisition geometry is arranged in a cross-hole configuration, which includes two boreholes with 8 explosive sources each and two boreholes with 235 receivers each. The source spacing is 32 m and the receiver spacing is 1 m. Time-domain forward modeling involves a Ricker-wavelet with a peak frequency of $f_{peak} = 150$ Hz and a time discretization of $\Delta t = 0.1$ ms. The record length is set to $T = 0.2$ s, where the wavelet is shifted by 0.02 s to account for wavelet "stretching" due to low-pass filtering. Figure 7.2 illustrates observed data for the exemplary shot 5 located at $(x, y, z) = (11, 40, 144)$ m. The data are characterized by a dominant direct P-wave and numerous scattering events caused by the differently scaled structures of the random medium.

The grid spacing is set to $DH = 1.0$ m to fulfill the grid dispersion criterion (2.13). This results in a grid size of $320 \times 320 \times 512$ grid points. The model is located within the full space, surrounded by perfectly matched layers with a width of 10 m. All general modeling and inversion parameters can be found in Table 7.1.

In consequence of cross-well experiments in chapters 4 and 5, I choose a homogeneous initial model with $v_P = 2000 \frac{m}{s}$ for waveform tomography. It matches the average velocity of the true model. Thus, it already gives a quite good explanation of the direct wave. While its maximum deviation from the true model is ± 20 %, the average model error (calculated with relation 6.1a) amounts to 6.04 % corresponding to an average absolute deviation of approximately $\pm 121 \frac{m}{s}$.

The technical feasibility of FWT is supported by the application of a multi-stage approach within the workflow implementation, which comprises a sequential frequency filtering over seven stages with associated peak frequencies $f_{peak} = (20, 25, 30, 40, 70, 80, 150)$ Hz. Each stage consists of a certain number of mandatory iterations: 5 (stage one), 10 (stages two to six) and 15 (last stage). Additional iterations are applied unless the misfit reduction between two successive iterations is less than 0.5 %.

7.1.2   Results

The results of the cross-well experiment are illustrated by cross sections across the 3D model. Representative central slices are shown in Figure 7.3. Additional slices can be found in Figures F.1 (x-z planes), F.2 (y-z planes) and F.3 (x-y planes) in appendix F.1. The final synthetic data for an exemplary shot (compare Figure 7.2) is shown in Figure 7.4.

Obviously, the 3D FWT in the time domain is able to compute a meaningful velocity model from a homogeneous initial model. While large-scale structures are recovered well, the resolution of small-scale structures reveals some uncertainties. Based on the average velocity and the peak frequency, the dominant wavelength amounts to approximately 13 m. The FWT reconstructs structures at the scale of a seismic wavelength.

In contrast to a 2D FWT employing a comparable model (chapter 5), the reconstruction of the 3D velocity model is less accurate. The acquisition geometry is not able to illuminate the entire model. Consequently, illumination gaps occur between source boreholes and between receiver boreholes. In particular, all model boundaries exhibit a poor v_P reconstruction (for example top row in Figure F.1, top and bottom rows in Figure F.2 as well as Figure F.3). The x-z plane at $y = 80$ m reveals a bad illumination at small and large x-coordinates. Here, the resulting v_P model is characterized by significant artifacts (see top row in Figure
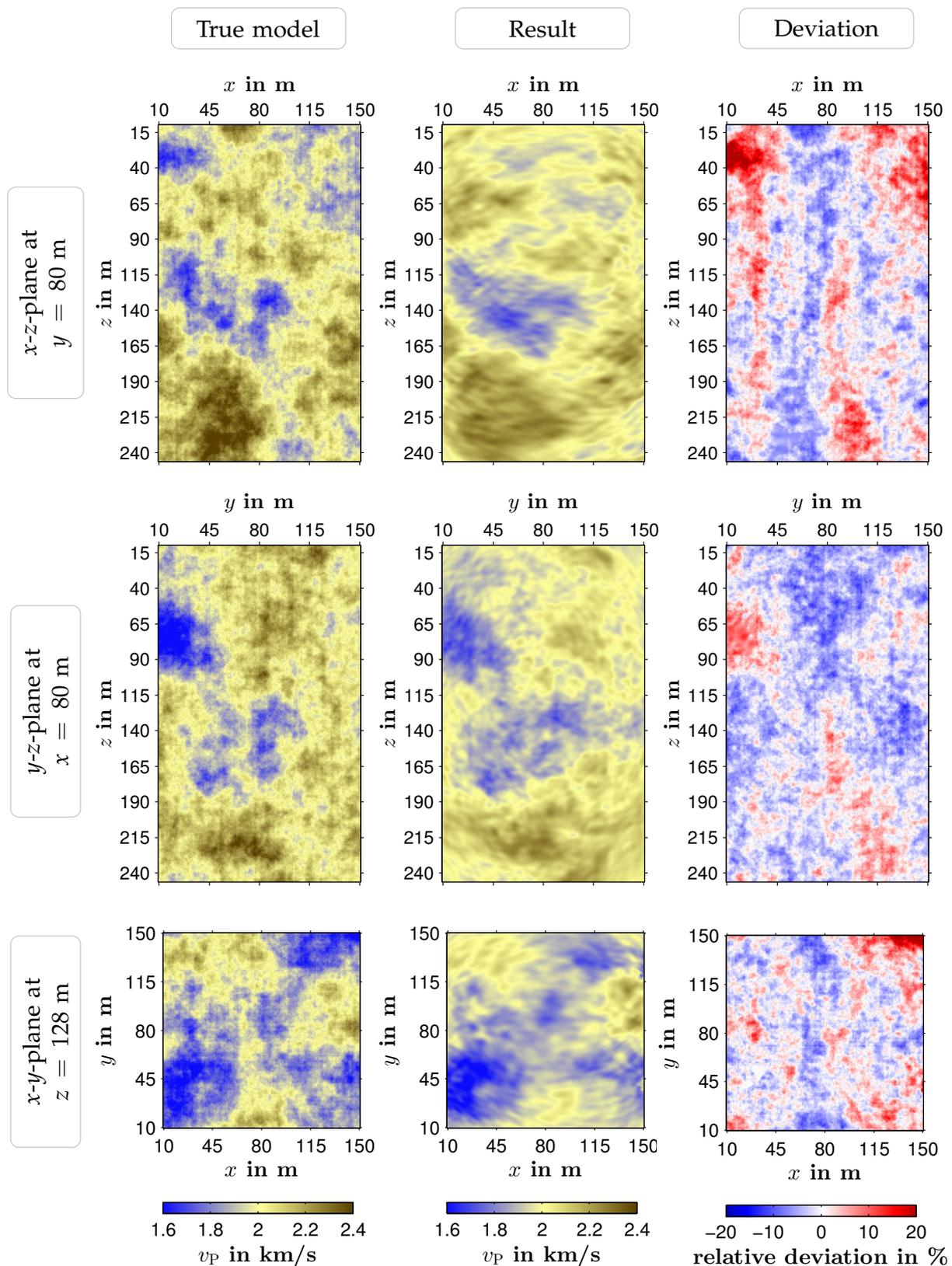
Figure 7.3: Application 4, cross-well experiment: Exemplary vertical (top and middle row) and lateral (bottom row) cross sections for the true random medium model (left column), the inverted v_P model (middle column) and corresponding relative deviations (right column). All planes are located centrally.

7.3). Apparently, the seismic wavefields are not sensitive to the structures in these areas. Nevertheless, the FWT is able to explain the recorded data (see Figure 7.4). This comes along with a model-error reduction down to 3.44 % which corresponds to an average absolute deviation of approximately $\pm 69 \, \frac{m}{s}$.

The v_P model exhibits a clearly visible footprint of an ellipsoidal shape, which

The cross sections with respect to the x-z plane and x-y plane show a smearing of the recovered random medium structures along ellipsoidal shapes (e. g., top and bottom rows in Figure 7.3, top row in Figure F.1 as well as top and middle rows in Figure F.3). This might be related to the superposition of the first Fresnel zones caused by the source-receiver coverage. In particular, the multi-stage approach affects size and shape of these sensitivity areas. At the inversion of low-frequency contents, the Fresnel zones show the largest spatial extent. This allows the reconstruction of very large scale structures close to the model boundaries, even between the source boreholes or the between receiver boreholes. The inversion of higher-frequency contents causes an increased narrowing of the Fresnel zones. Consequently, the sensitivity of the wavefields on model perturbations decreases towards the model boundaries (see y-z plane in the bottom row of Figure F.2 or Figure F.3). Usually, smaller-scale structures are artificially added within these regions. The optimal illumination is achieved within the acquisition geometry, particularly with regard to the x-z planes at $y = 40$ m and $y = 120$ m which are virtually spanned by the boreholes. The average model error is computed as a function of the y direction for all x-z planes. It verifies the observation by showing significant minima around the locations of the "source-receiver planes" (see Figure 7.5). This indicates a suboptimal arrangement of the acquisition geometry with irregularly distributed sources and receivers. The most reasonable recovery of structures at the scale of the seismic wavelength is obtained within the central parts of the optimal planes (see middle and bottom rows in Figure F.1). The resolving power seems to be limited to the minimum available wavelength, leaving the smallest features out of consideration (see the relative deviation between v_P result and true model in the middle and bottom rows of Figure F.2).
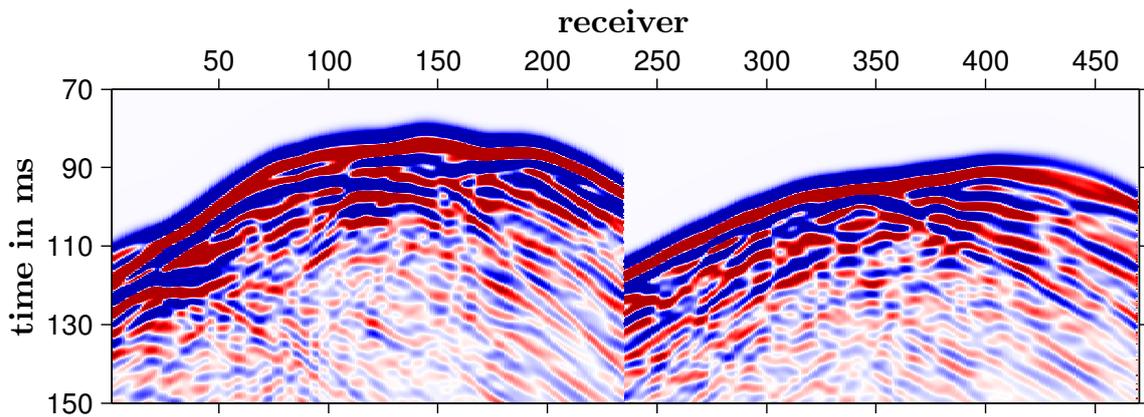
Figure 7.4: Application 4, cross-well experiment: Final synthetic data for an exemplary source located at $(x, y, z) = (11, 40, 144)$ m. The data are recorded in both receiver boreholes at $y = 40$ m (left part) and $y = 120$ m (right part). The seismogram can be compared with the observed data in Figure 7.2 due to an identical application of normalization and clipping.
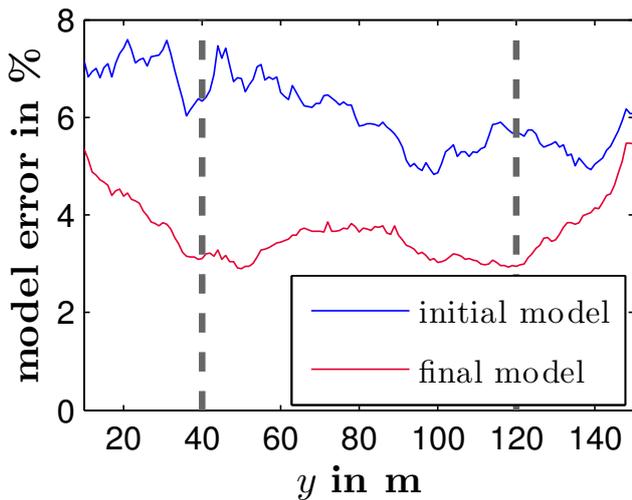
Figure 7.5: Application 4, cross-well experiment: Model error for the initial and final v_P model with respect to the true v_P model; it represents the average error computed from vertical x-z planes at all y positions. The lowest final model errors can be found inside the areas between source boreholes and receiver boreholes with y locations highlighted by the dashed lines. On average, the model errors amount to 6.04 % (initial) and 3.44 % (final).

7.2 3D FWT for a reflection geometry

7.2.1 Marmousi expansion and geometry setup

In the analogy of the cross-well experiment, the reflection experiment demonstrates the characteristics and difficulties arising from a 3D tomography.

The experiment employs a section of the 2D Marmousi-II model (Martin et al., 2006). It is replicated with respect to the second lateral direction y. An additional lateral shift in x direction using a combination of a linear trend and a cosine function ensures the generation of 3D structures (Figure 7.6). The true v_P model is thus characterized by complex fault structures which are embedded into a layered sedimentary geology. The model size amounts to $5600 \times 2400 \times 2400$ meters. The size of the finite-difference grid amounts to $560 \times 240 \times 240$ grid points. Table 7.2 summarizes general parameters.

The reflection acquisition geometry is located at the sea surface and consists of 24 sources along three lines as well as a total number of 1778 hydrophones along seven lines. Although, this experiment has to get along with 24 sources only, the large receiver array is used to ensure an adequate illumination of the subsurface (compare with 32 sources and marine streamers used in 2D applications in sections 6.2.2 and 3.2.2). Time-domain forward modeling involves a Ricker-wavelet with a peak frequency of $f_{\text{peak}} = 9$ Hz and an effective record length of $T = 4.85$ s. Figure 7.7 illustrates observed data for the exemplary shot 5 located at $(x, y, z) = (3170, 520, 10)$ m. Apart from direct waves and dominant seafloor reflections, the data are characterized by numerous diffractions caused by small structures within the fault zones.

For waveform tomography, I choose a linear 1D v_P gradient as an initial model (Figure 7.6). It contains the water layer of the true model, which is excluded from the model update. On the analogy of previous reflection experiments in chapter 6 and section 3.2.2, the multi-stage approach is applied. It consists of low-pass frequency filtering over five stages of inverting seismic data with peak frequencies which are not equidistantly spaced Sirgue and Pratt (2004): $f_{\text{peak}} = (1.4, 2.0, 2.9, 4.5, 9.0)$ Hz. All five stages have to apply a certain number of mandatory iterations: 10, 10, 15, 20, 50. Additional iterations are applied unless the misfit reduction between two successive iterations is less than 0.5 %.

Figure 7.6: Application 4: True (top) and initial (bottom) v_P models for the Marmousi experiment. The true model consists of a complex 3D geology and a water layer on top. The acquisition geometry is located at the sea surface and consists of 24 explosive sources (red markers) and 1778 hydrophones (green lines).
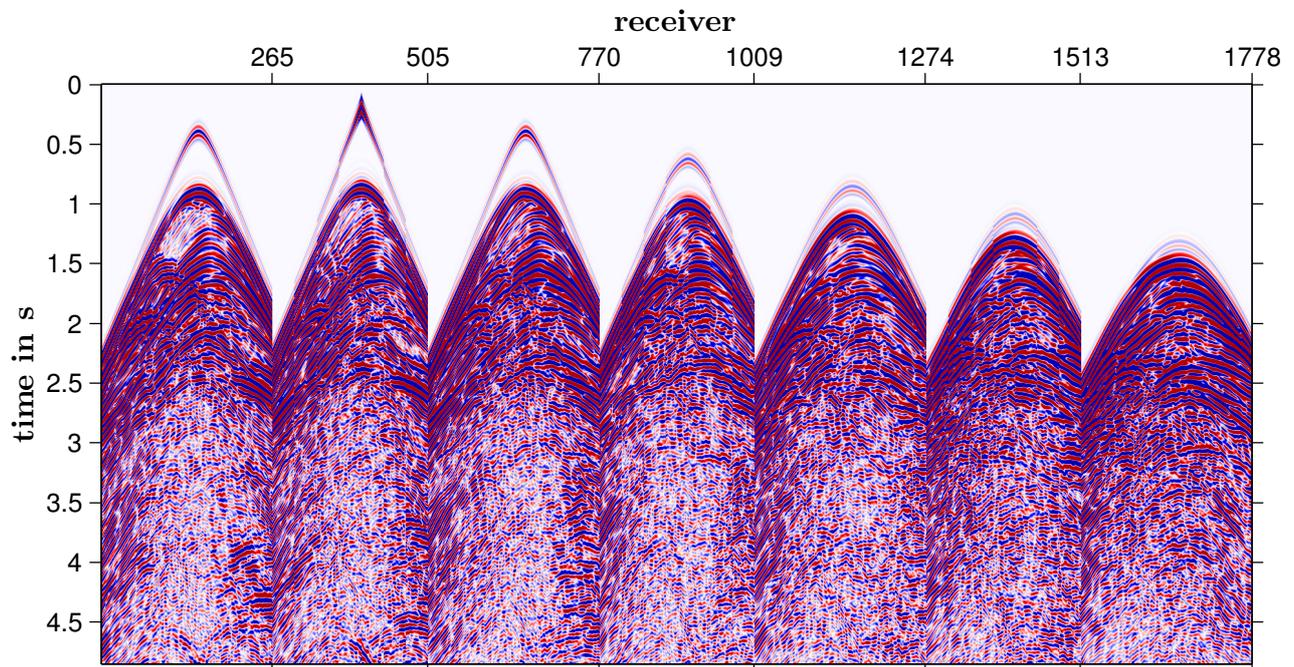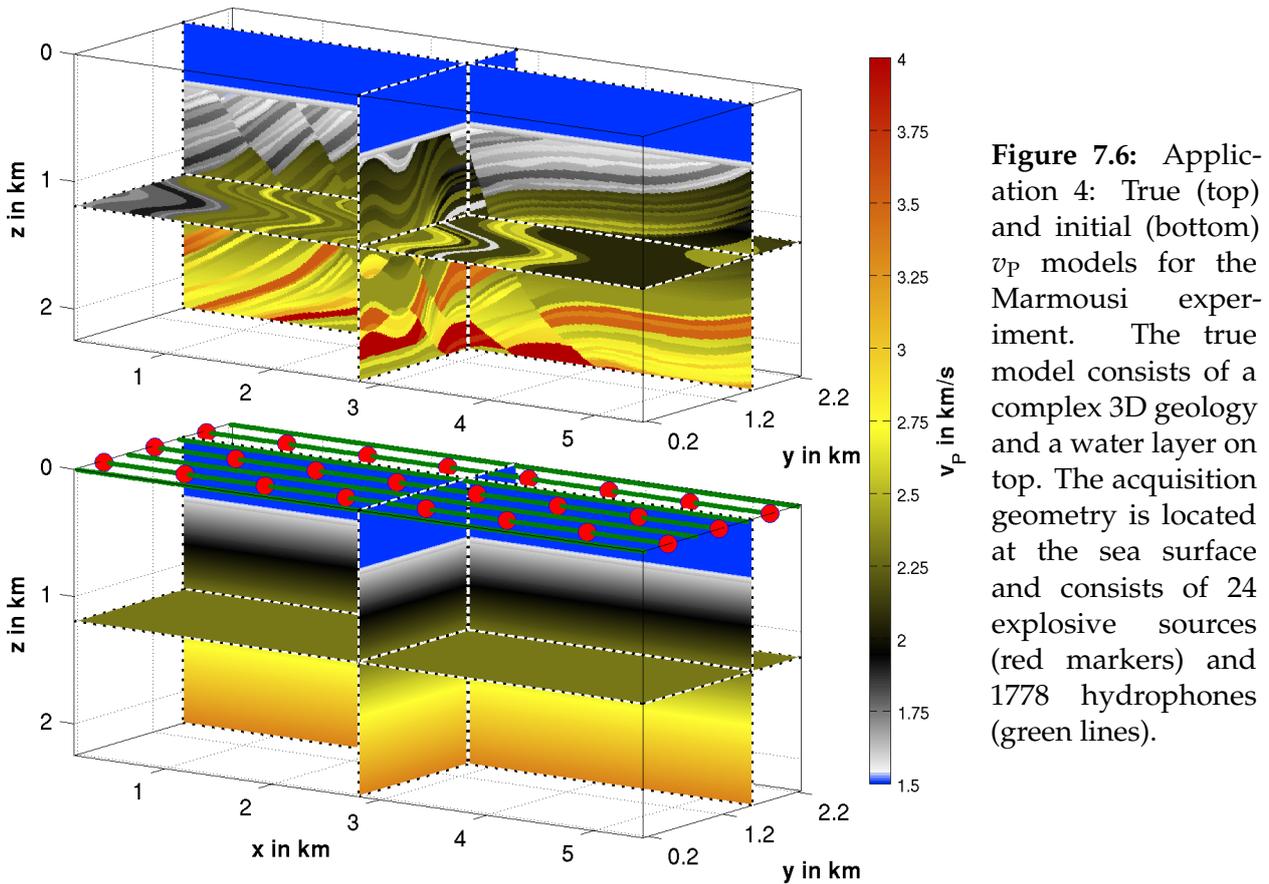
Figure 7.7: Application 4: Recorded pressure data for the Marmousi model in Figure 7.6 (top) and an exemplary shot located at $(x, y, z) = (3170, 520, 10)$ m. The seismogram comprises all seven receiver lines, where the lines are located at $y = (180, 520, 860, 1200, 1540, 1880, 2220)$ m. For a better visualization, a time-dependent amplitude gain of $t^{1.4}$ is applied. Furthermore, the data are normalized to the maximum of all traces and clipped to ± 1.5 %.

Table 7.2: Application 4: General setup of the 3D Marmousi experiment.

Application 4: 3D Marmousi experiment		
Attributes		**Specifications**
Model	**Size**	lateral dimensions: 5.6 x 2.4 km, height: 2.4 km
	Structure	· modified Marmousi-II geology (v_p = [1550, 4000] m/s) · with water layer (v_p = 1500 m/s)
	Figure	Figure 7.6 (top)
Acquisition geometry	**Location**	on top of half-space, sea surface
	Configuration	reflection geometry
	Sources	· 24 explosive sources along 3 lines · source-time function: Ricker wavelet · peak frequency: 9 Hz · time delay of source signal: 0.75 s
	Receivers	· 1778 hydrophones along 7 lines · recording length: 5.6 s
	Offsets	minimum: 30 m, maximum: 5537 m
	Figure	Figure 7.6 (bottom)
Modeling parameters	**Model size**	· lateral: 560 x 240 grid points, height: 240 grid points · grid spacing: 10 m
	Propagation time	· 4000 time steps · time discretization: $1.4 \cdot 10^{-3}$ s
	PML boundary	· width: 150 m · free surface
General inversion parameters	**Initial model**	Figure 7.6 (bottom)
	Gradient computations	· imaging condition in time domain · wave-field based preconditioning · user-defined preconditioning: water layer is known · additional taper within PML boundary
	Model update	· hard constraint: v_p limit: [1500, 4000] m/s · maximum deviation from the initial model: 50 %
	Workflow	frequency filtering over multiple stages

7.2.2 Results

The results of the Marmousi experiment are illustrated by cross sections across the 3D model. Representative central slices are shown in Figure 7.8. Additional slices can be found in Figures F.4 (x-z planes), F.5 (x-y planes) and F.6 (y-z planes) in appendix F.2. The final synthetic data for an exemplary shot (compare Figure 7.7) is shown in Figure 7.11.
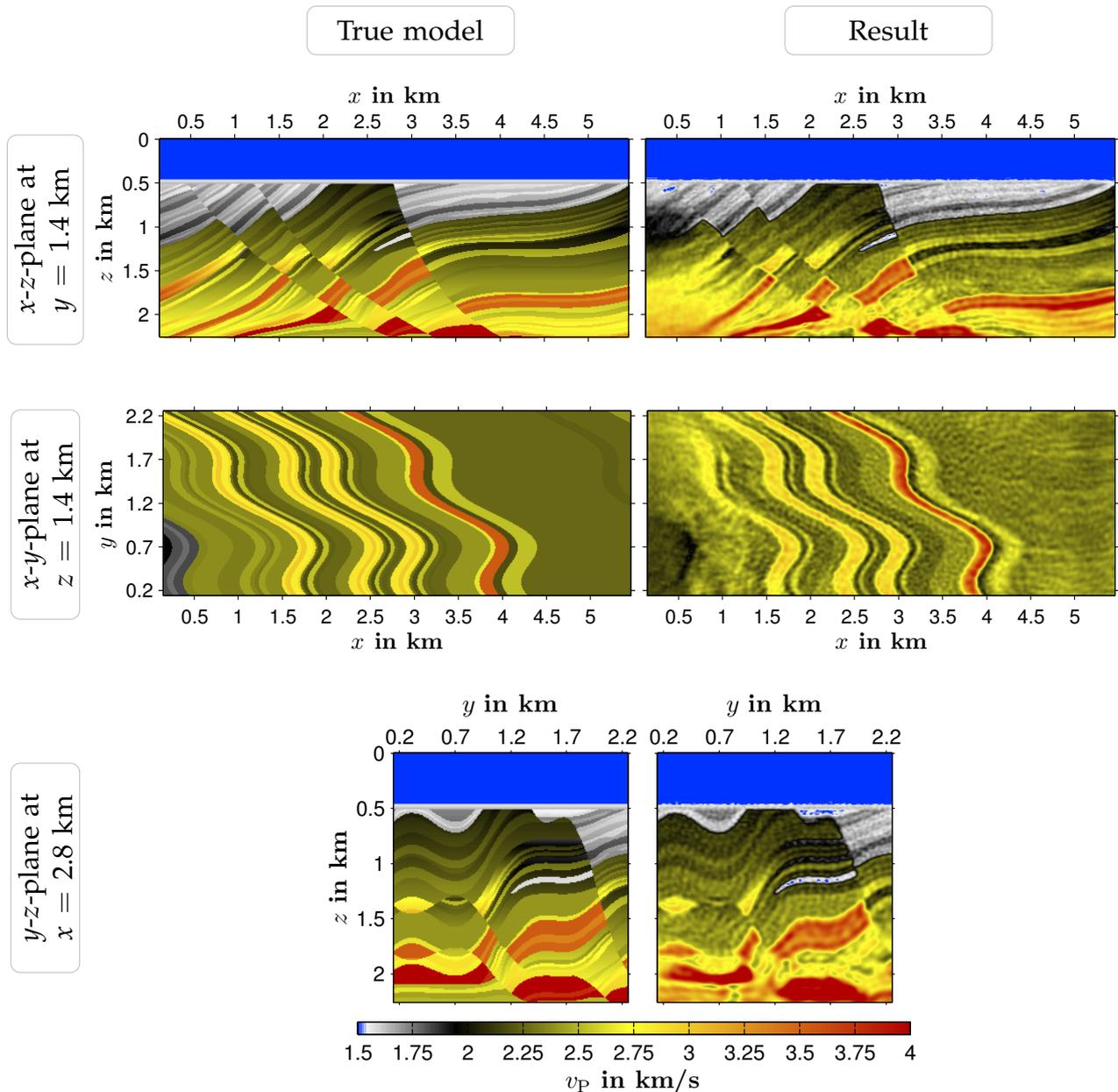
Figure 7.8: Application 4: Exemplary vertical (top and bottom row) and lateral (middle row) cross sections for the true Marmousi model (left column) and the inverted v_P model (right column). All planes are located centrally (compare Figure 7.6).

First of all, the 3D FWT sufficiently recovers complex 3D structures of the subsurface model. However, in contrast to a 2D FWT, the resulting v_P model is less accurate. The FWT produces very smooth intermediate v_P models at early stages with the inversion of low-frequency contents. Then it refines the v_P model, but also adds significant noise to the model during the inversion of the full frequency content (see inversion progress in Figure 7.9). Despite remarkable model improvements during the early stages, the existing coverage seems to become insufficient at higher frequencies due to an increasing ill-posedness of the inverse problem. In particular, the sources are spatially undersampled, which has already been discussed by Sun and McMechan (1992) (for FWT in the time domain) and Brenders and Pratt (2007a) (for an application of 2D frequency-domain FWT), among others. However, further investigations – including more sophisticated multi-stage strategies or using more appropriate acquisition geometries to provide a better illumination – are necessary.

In contrast to the 2D Marmousi experiments, the limitations of the resolving power are clearly visible. It is well-known that both the lateral and vertical resolution of a seismic tomography using a reflection geometry decreases with increasing depth.

As mentioned in section 7.1.2, the sensitivities on model perturbations are strongly affected by the frequency content. In case of low frequencies, the Fresnel zones extend down to the bottom of the model and thus the FWT is able to recover 3D structures at the scale of long wavelengths. In contrast, the involvement of higher frequencies causes a focusing of the sensitivities on near-surface areas resulting in the resolution of small-scale structures (for example, compare lateral cross sections for different depths in Figure F.5 and Figure 7.8).

Additionally, this comes along with a defocusing effect towards the model boundaries and a best resolution at the lateral center of the acquisition geometry (in particular, see cross sections for $z \geq 1.1$ km in Figure F.5 and top row in Figure 7.8). This observation is verified by a depth-dependent model error computed for x-y planes at all depths (see Figure 7.10). The comparison of initial and final model shows a strong reduction of the model error for depths of $z \lessgtr 1.7$ km emerging from a quite good recovery of large-scaled and small-scaled structures. However, deeper regions reveal an erratic behavior of an increasing model error indicating the lack of higher-contrast features, such as interfaces.

Despite usage of a homogeneously distributed acquisition geometry, the FWT fails to sufficiently recover structures at the "left" end of the model with $x \lessgtr 1.5$ km (see vertical cross sections in Figures F.4 and 7.8). In contrast, the v_P reconstruction along the "right" model boundary is much more satisfactory. Apparently, this might be caused by the closeness of the complex fault structures to the model boundary combined with a decreased illumination.

Apart from problems mentioned above, the FWT reveals a proper reconstruction of small-scaled structures, such as the low-velocity inclusion at $z \approx 1.1$ km which probably contains hydrocarbons (see Figures 7.8 [top and bottom rows], F.4, F.5 [middle row] and F.6 [y-z plane at $x = 3.65$ km]) or anticlinal deformations (see bottom row in Figure 7.8). Although the initial model does not seem to be the most optimal linear gradient model (see increasing depth-dependent model error in Figure 7.10), the FWT reduces the global model error – excluding the known water layer – from 12.1 % (initial) to 5.76 % (final). In spite of a differently good quality of v_P reconstruction, the recorded seismic data is explained quite well (Figure 7.11b). While the most significant residuals appear in case of some reflection events (Figure 7.11a), the residual data mostly represents a fraction of the recorded data.
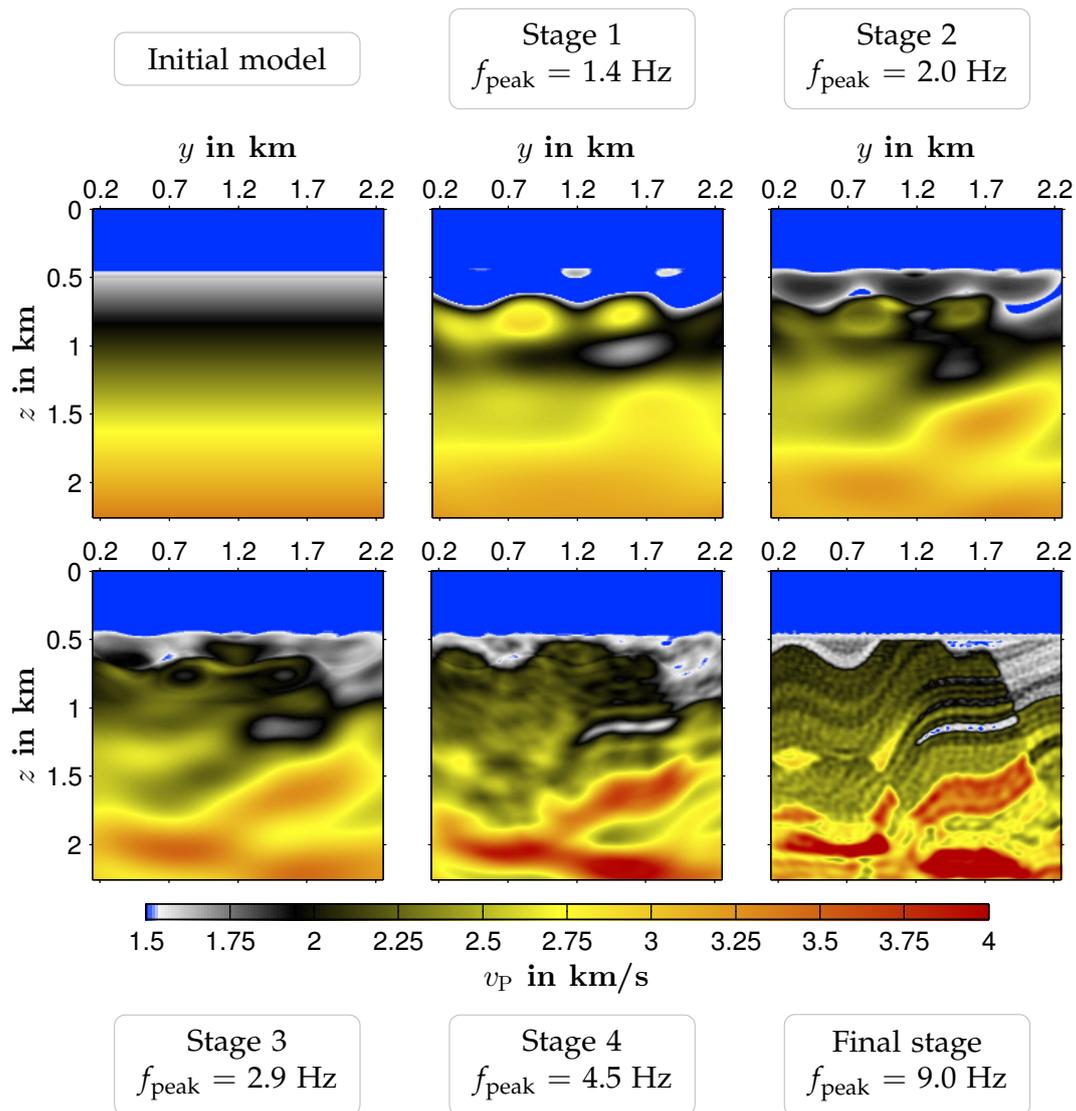
Figure 7.9: Application 4, Marmousi experiment: Cross sections at the central position $x = 2.8$ km for the initial v_P model and intermediate v_P results obtained at the end of each frequency-filtering stage. While the inversion of low-frequency contents produces reliably smooth models, the consideration of the full frequency content results in a quite noisy v_P model.
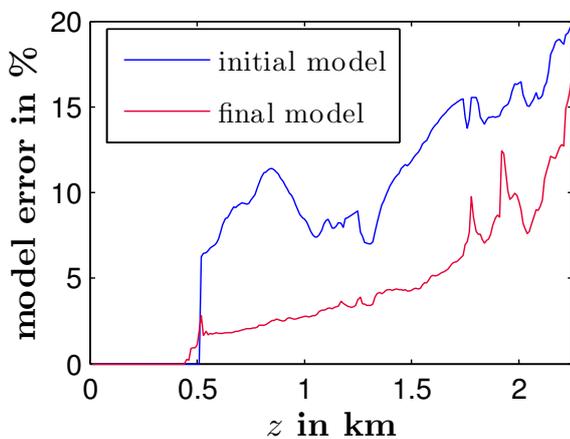

Figure 7.10: Application 4, Marmousi experiment: Model error for the initial and final v_P model with respect to the true v_P model; it represents the average error computed from lateral x-y planes at all depths z. On average, the model errors amount to 12.1 % (initial) and 5.76 % (final).
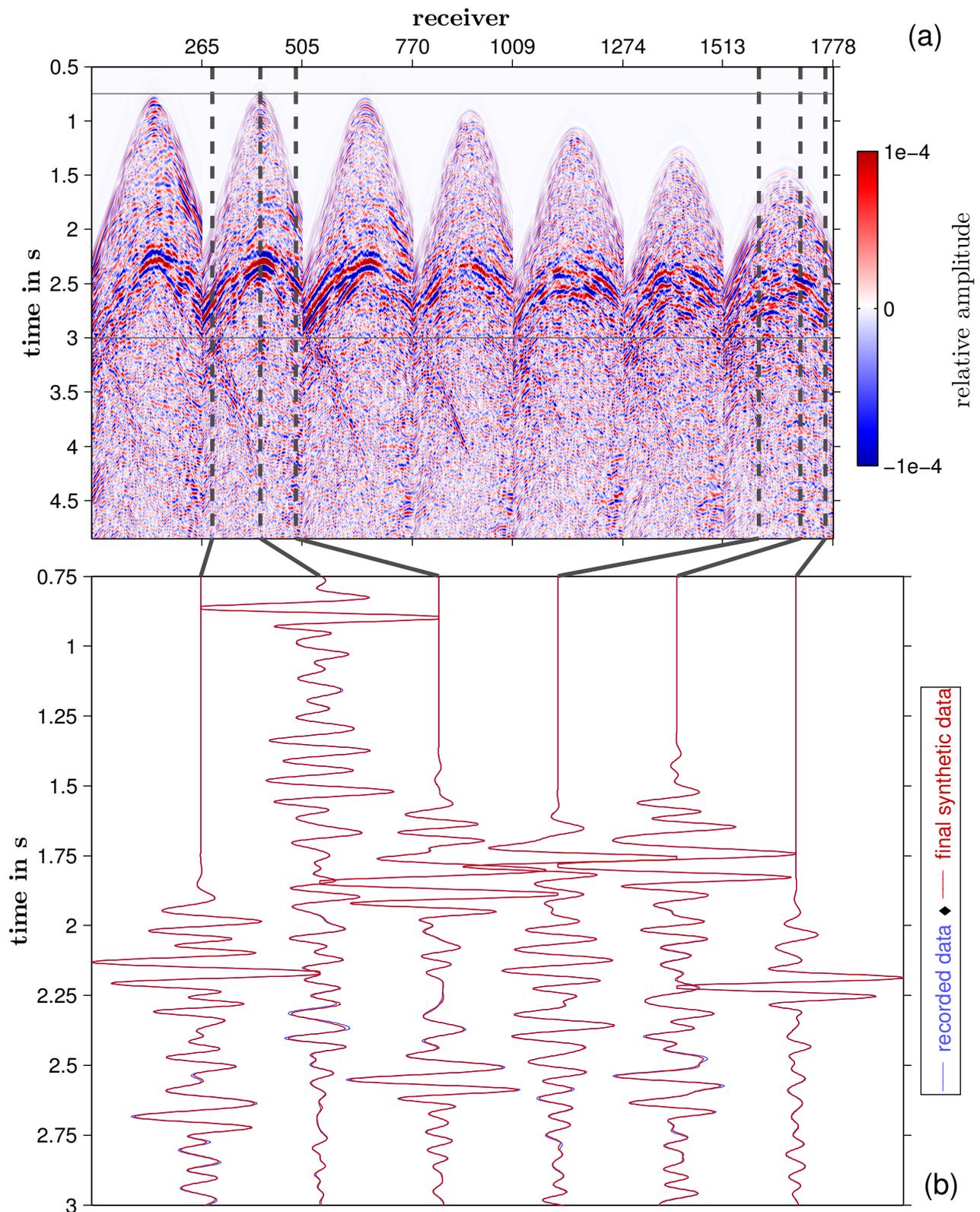
Figure 7.11: Application 4, Marmousi experiment: Both plots show pressure data for an exemplary shot located at $(x, y, z) = (3170, 520, 10)$ m. (a) illustrates the residual data obtained for the final v_P model. The seismogram is normalized to the maximum amplitude of the recorded data and clipped to ± 0.01 %. (b) shows the corresponding recorded data and final synthetic data for selected traces along the receiver lines 2 (at $y = 520$ m) and 7 (at $y = 2220$ m). For a better visualization, a time window of $[0.75\ 3]$ s is chosen (highlighted by horizontal lines in (a)).

7.3   Computational efforts of 3D acoustic FWT

In the course of the 3D-FWT development, I decided to apply a pure time-domain FWT. I performed preliminary tests employing the single-frequency method (see Figure 2.1 and chapter 5). While I obtained a reasonable result for the cross-well experiment which is comparable to the result shown in section 7.1.2, the FWT failed in case of the reflection experiment. A possible reason is an unfavorable combination of using an unusually low coverage (only 24 sources), choosing a single frequency and the appearance of the cycle-skipping effect in conjunction with the existence of seafloor and corresponding (multiple) reflections. The single-frequency method resulted in an artificial model construction. A higher number of sources would probably improve the performance of the FWT. Obviously, the application of a pure time-domain FWT with exploitation of broad frequency bands and usage of a sparse acquisition geometry seems to be more robust. To keep consistence, both experiments employ a pure time-domain FWT. Although the computational times of both methods are comparable, the time-domain method reveals an extensive memory consumption (see section 3.2.1).

Table 7.3: Application 4: Computational efforts of 3D acoustic FWT applications with respect to a cross-well experiment performed on the supercomputer *JUROPA* and a reflection experiment performed on the supercomputer *HERMIT* (see Table 3.1 for hardware details).

	Cross-well experiment	**Reflection experiment**
Grid size in modeling	$320 \times 320 \times 512$	$560 \times 240 \times 240$
Grid spacing in modeling: Δh	0.5 m	10 m
Number of time steps in modeling	2 000	4 000
Time step in modeling: Δt	$1.0 \cdot 10^{-4}$ s	$1.4 \cdot 10^{-3}$ s
Grid size for wavefield storage	$160 \times 160 \times 256$	$560 \times 240 \times 240$
Grid spacing for storage: Δh_{samp}	1 m	10 m
Number of wavefield snapshots	499	666
Time step for storage: Δt_{samp}	$4.0 \cdot 10^{-4}$ s	$8.4 \cdot 10^{-3}$ s
Domain decomposition	$4 \times 4 \times 4$	$8 \times 4 \times 4$
Shot parallelization	all 16 sources	all 24 sources
Number of cores	1 024	3 072
Memory consumption per core	195 MiB	640 MiB
Overall memory usage (snapshots)	195 GiB	1.88 TiB
Number of FWT iterations	245	281
Average time per FWT iteration	187 seconds	366 seconds
Total amount of forward modelings	19 600	33 720
Total computational time	12.7 hours	28.6 hours
Total amount of core time	$\approx 13\,000$ core hours	$\approx 88\,000$ core hours

Both experiments apply methods described in section 3.2.1 to ensure the computational feasibility of time-domain FWT. The experiments are performed on the supercomputers *JUROPA* (cross-well experiment) and *HERMIT* (reflection experiment). The optimization settings are adapted to the extent of the computations and the resources available on these supercomputers (see Table 3.2.1 for details). Table 7.3 makes clear that a pure time-domain FWT is highly demanding. While the cross-well experiment employs the reduction of both time steps and grid size for storage of pressure wavefields, the reflection experiment uses the full model grid.

It is possible to reduce the costs of wavefield storage by using a coarser inversion grid (see section 3.2.1). I repeated the reflection experiment using a grid size of $280 \times 120 \times 120$ grid points for storage, i. e., the model is reduced down to $\frac{1}{8}$ of its original size exhibiting a grid spacing of $\Delta h_{\mathrm{samp}} = 20$ m (finite-difference modeling still requires the full grid with $\Delta h = 10$ m). Consequently, the FWT has to perform less computations with respect to the imaging condition. Furthermore, this strategy also accelerates the overall memory management and the I/O. The average computational time for one iteration decreases from 366 seconds (see Table 7.3) to 255 seconds. Apart from the coarser grid, both FWTs produce comparable results. Exemplary cross-sections of the final v_{P} model for FWT test using the full or reduced grid size are illustrated by Figure 7.12. However, the coarse model is characterized by some more inaccuracies, such as higher "noise" and slightly misplaced interfaces within the top-right part of the model. This is probably caused by the application of the trilinear interpolation method to rescale the velocity model to the size of the finite-difference grid. Apparently, a meaningful strategy might comprise several FWT stages – starting with coarse grids and using an increased grid resolution at later stages.
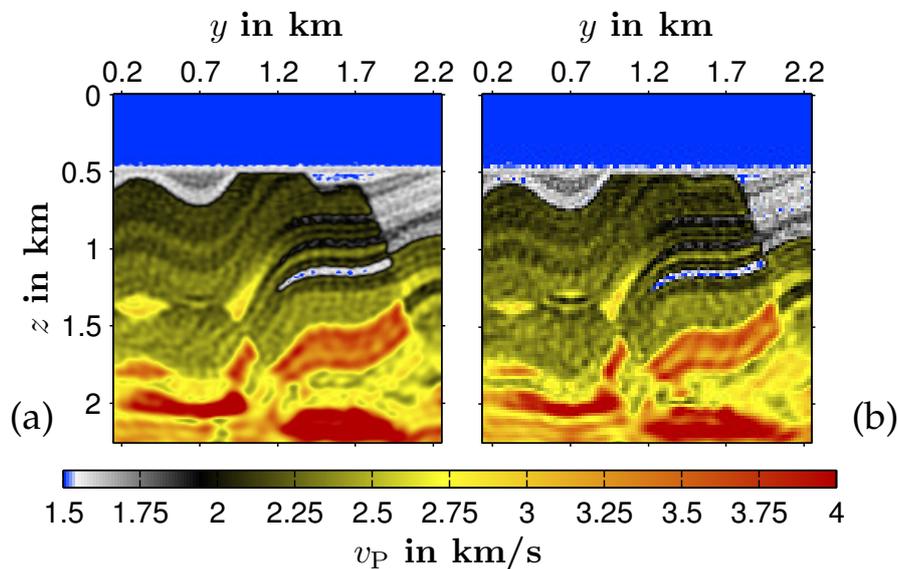
Figure 7.12: Application 4: Comparison of an exemplary cross section of final v_{P} results, which are produced by inversions using different grid sizes for the imaging condition. (a) is obtained from the FWT discussed in section 7.2 involving the full grid size and is identical with Figure 7.8. (b) is computed by a FWT using a reduced grid with the size of $\frac{1}{8}$ of the full grid, i. e., every other grid point along x, y and z directions is skipped.

7.4 Summary

In general, the applicability of a 2D FWT is restricted to fields of application exhibiting quite simple subsurface structures, such as 1D or 2D geological models. As an outlook for future applications, the scope of this work comprises the development of a 3D acoustic FWT implementation in the time domain and its application to two experiments including subsurface models with a 3D geology. In the style of 2D applications presented in chapters 5 and 6, the 3D FWT is applied to both a cross-well experiment and a reflection experiment. Although the 3D experiments have to get along with quite sparse acquisition geometries resulting in a limited illumination of the subsurface, they are able to produce reliable v_P models. Similarly to the 2D applications, the inverse problem of the cross-well experiment is less ambiguous, which allows a satisfactory recovery of complex velocity models by using the simply homogeneous initial model. The worse illumination causes a model resolution which is not as good as in a 2D FWT.

The FWT implementation used in this chapter takes advantage of the pure time-domain inversion scheme by considering broad frequency ranges of the seismic data instead of using single frequencies. In particular, this method facilitates performance and outcome of the reflection experiment, which has to handle disadvantageous circumstances, such as the sparse acquisition geometry and a poor initial velocity model. The most reliable resolution is achieved by the inversion of low-frequency contents – producing realistic large-scale structures throughout the entire model. Due to increasing frequencies, the sensitivity on model perturbations and the corresponding consideration of low-wavelength structures are limited to near-surface areas. Due to the sparse acquisition geometry, the near-surface sensitivity becomes more fragmentary at high frequencies causing artificial model noise at small scale lengths.

The 3D acoustic FWT proves to be feasible in academic applications to small-scale problems. However, the combination of a pure time-domain approach and optimization strategies allows reasonable computing times but is realized at the expense of a huge memory consumption.

Beyond the scope of this thesis, the high computational costs raise the question of the necessity of 3D FWT in dependence of model complexity and acquisition geometry. A synthetic study involving the evaluation of 3D cross-well data for a complex medium with focus on recovering the S-wave velocity model is carried out by Dunkl et al.. This study compares the performances of 2D elastic FWT after a 3D-to-2D transformation of the data and 3D elastic FWT. Although the velocity model is sufficiently recovered, both cases show a lower resolution compared to a 2D inversion of pure 2D data. On the one hand, the 3D FWT suffers from a relatively poor coverage (as discussed in this chapter). On the other hand, the 2D FWT maps the scattered-wavefield information caused by 3D structures to the 2D model by producing artifacts. Nevertheless, a transmission experiment seems to be the most promising field of application for a 2D FWT. However, in real-data applications of reflection experiments, a successful performance of a 2D FWT requires a very simple geology, such as a 1D layered medium (also suggested by Köhn, 2011). Additionally, the application of acoustic FWT in case of marine reflection experiments involves the consideration of 3D geology and elastic effects. This requires a trade-off between computational efforts and accuracy with respect to the applicability of 2D/3D acoustic/elastic FWT.

Chapter 8

Summary

In this work I developed 2D and 3D time-domain implementations of acoustic full waveform tomography (FWT) including its optimization with respect to methodological and technical aspects. I accomplished several synthetic feasibility studies with applications to seismic data computed in cross-well or reflection experiments.

At the present time, the FWT evolves as a powerful method in seismic imaging. One can find a high diversity of different approaches. Usually, the relevant literature distinguishes between FWT in time domain and frequency domain. Although frequency-domain FWT is commonly used, several reasons induced the decision to focus on the development of a time-domain implementation. In the long term, the 3D FWT will become more important, in particular, if a (real-data) application of a 2D FWT is insufficient due to its inability to take 3D effects into account. Thus, one motivation is to avoid general drawbacks of (3D) frequency-domain FWT and to benefit from the advantages of time-domain FWT, such as the consideration of broad frequency bands instead of single frequencies or the straightforward 3D modeling including a variety of efficient parallelization techniques.

8.1   Significant contributions of this work

Implementation

The acoustic FWT implementation bases on existing fundamentals of the FWT methodology and finite-difference modeling (described in chapter 2). Due to its high computational efforts and the limited availability of computational resources at early stages, the optimization emerged as an essential task of my work to overcome these drawbacks. This comprises methodological improvements (M) to increase the convergence of the iterative gradient algorithm (described in section 2.6) and technical features (T) (discussed in chapter 3):

(M) The adaptive step length estimation involving a parabolic fit represents the most significant methodological feature. The sophisticated implementation is able to handle numerous exceptional cases. Depending on the application or the iteration-dependent complexity of the data misfit function, it may provide large or very small step lengths – ranging over several orders of magnitude – to ensure a continuous misfit reduction. In comparison to the usage of simply constant step lengths, the exemplary application in chapter 4 demonstrates a reduction of essential iterations by 75 %.

(T) Most of the computational time is consumed by modeling, i.e., by forward-propagations of the source wavefields, the back-propagation of the residual wavefields and additional modelings for the step length estimation. Hence, I combined the simultaneous modeling of shots with the existing domain decomposition to get a massive parallelization exhibiting a high efficiency. In particular, 2D applications benefit from this method by overcoming the disadvantages of a pure domain decomposition. Even on supercomputers a pure domain decomposition easily exhausts communication capacities. The more intense the decomposition, the smaller the subdomains of the model area, the faster the modeling computations and the more data has to be communicated within a certain period of time. The combined parallelization is extremely adaptable to the hardware architecture and allows the usage of extensive resources by minimizing the domain-decomposition efforts and maximizing the number of modelings computed at once. The first tests were performed on a local workstation cluster. Due to the limitation of modeling to each workstation and the resulting elimination of network communication, the computing time was reduced by almost 90 %. Even on supercomputers there is a tremendous discrepancy between a combined parallelization – allowing a nearly linear speedup as shown by two benchmarks in chapter 3 – and the pure domain decomposition.

(T) As a consequence of the massive parallelization, the memory consumption of the time-domain implementation increased significantly. Although, the forward-propagated wavefields have to be stored in memory, it is not necessary to sample wavefield snapshots with respect to the finite-difference time step. The application of the Nyquist sampling criterion results in storing a small subset of snapshots. In case of 3D FWT, this is also applied to the spatial sampling which allows a significant reduction of memory consumption along with the exploitation of shot parallelization.

(T) The implementation of a multi-purpose workflow scheme increases the efficiency by simply improving the handling of the FWT implementation. It allows

– arbitrary combinations of common strategies to mitigate the ill-posedness of the inverse problem, i.e., it is a flexible inversion scheme over several stages,

– to benefit from advantages of inverting in the time-domain and frequency-domain due to the free choice of time-domain FWT or single-frequency method,

– the minimization of efforts per stage by employing stop criteria (in order to achieve convergence in the course of a stage, the number of essential iterations might vary significantly),

– the potential of arbitrary expansions, such as applying different methods of gradient preconditioning, among others.

Concluding, the combination of these strategies increases the convergence and reduces computational efforts. In the course of this work, these improvements allowed numerous synthetic studies and were the essential condition to prove the feasibility of 3D FWT in the time domain.

Parameter study

Within the scope of an extensive quantitative study including numerous cross-well experiments (see chapter 4), I examined the impact of methodological strategies and the choice of the initial model on the progress of the FWT and the corresponding ability to reconstruct the velocity model. In general, a usual experiment consisted of FWT runs for 51 different initial velocity models with 100 iterations each to visualize the evolution of the adaptive step length as well as cross sections of the data misfit function and the model error function. The impact of methodological problems, such as the choice of the initial model, the so-called cycle-skipping effect, frequency filtering, usage of an appropriate acquisition geometry (among others), was already taken into account by the relevant literature. However, this work presents the first comprehensive study (to the best of my knowledge) which elaborately visualizes their footprint on the inversion progress. An assortment of interesting observations is listed below.

- A plain FWT, that does not apply any strategy to improve its outcome, significantly suffers from cycle-skipping. Only a small range of initial models allows the reconstruction of a satisfactory velocity model. Interestingly, the choice of two very similar initial models might cause completely different FWT progresses. In other words, the deviation of their background velocity with respect to the average P-wave velocity (2200 $\frac{m}{s}$) of the true model amounts to ($-62 \frac{m}{s}$) and ($-31 \frac{m}{s}$), respectively. In the first case, the FWT gets stuck in a local minimum. In the second case, a satisfactory velocity model is recovered.

- The most impressive impact is achieved by frequency filtering. The lower the frequency content considered in FWT, the weaker the dependency on the initial model. In other words, the mitigation of the cycle-skipping effect is realized by broadening the range of initial models allowing a successful inversion.

- In contrast, other factors, such as step length estimation or the acquisition geometry, have an influence on the convergence but not on the cycle-skipping effect. The strongest increase of the convergence is achieved by exploiting the adaptive step length estimation. Concerning the acquisition geometry, the essential number of sources represents a very relevant problem. The cross-well experiment shows that a trade-off between "undersampling & low computational efforts" and "oversampling & high computational efforts" has to be found, e. g., the optimal choice is the usage of 16 sources. While using a few sources (e. g., 4) results in a spatial undersampling and an unsatisfactory model, an arbitrary increase of the number of sources (e. g., 64) does not necessarily yield further improvements.

Comparison of approaches in the time domain and the frequency domain

To identify assets and drawbacks of the time-domain FWT, I classified its performance with regard to other approaches, such as the frequency-domain implementation *FULLWV* developed by R. G. Pratt and the single-frequency method. In chapter 5, I used a cross-well experiment to perform a comparison of all 2D approaches with the aim to obtain the most optimal results. I involved time-domain frequency filtering as the counterpart to the selection of single frequencies. The outcome of the comparison validated the focus on the development of a time-domain implementation. The reasons are:

- Within the scope of the given experiment, all techniques produced comparable results. However, the most optimal result was obtained from the pure time-domain approach. Apparently, it took advantage of the consideration of broad frequency bands – due to low-pass filtering – instead of using single frequencies.

- The 2D frequency-domain FWT proved to be most efficient with respect to the computational performance. While it runs on a single CPU core, the parallelized time-domain approach needs a multiple of these computational resources to get along with a comparable computing time. But, an efficient parallelization and the availability of present-day (super)computers compensate this disadvantage. Concerning the development of 3D FWT, the 3D time-domain implementation fully benefits from the achievements of the 2D implementation, while a 3D frequency-domain FWT is still highly demanding.

Acoustic FWT in the presence of attenuation

In a further synthetic feasibility study in chapter 6, I investigated the applicability of acoustic FWT in the presence of attenuation. I employed two marine reflection experiments using a simple 1D geology and the Marmousi model. The latter case included a realistic towed-streamer acquisition geometry. While the focus was on the recovery of the velocity model, the experiments either ignored the occurrence of attenuation or took it into account by assuming passive quality-factor models. The most relevant findings are:

- The application of a purely acoustic inversion scheme to seismic data affected by significant attenuation results in unsatisfactory velocity models. The attenuation-related data misfit which cannot be explained by the acoustic modeling, is partly mapped to the velocity model by producing notable artifacts.

- The reconstruction of the velocity model benefits from the consideration of attenuation in the modeling of the FWT. If the passive quality-factor model is an appropriate representation of the near-surface areas, then even a simple assumption, such as a homogeneous model, yields a reliable velocity model.

In particular, the inversion of seismic data acquired in marine environments with soft sediments requires the consideration of attenuation. The most realistic choice of including attenuation is the assumption of a smooth quality-factor model, which might be obtained from the initial velocity model.

Furthermore, a paper manuscript with these results has been prepared for submission.

3D acoustic FWT in the time domain

As a consequence of the experiences with the 2D FWT, I developed a pure time-domain implementation of 3D acoustic FWT and carried out a feasibility study with regard to the reconstruction of the velocity model in two synthetic experiments. I applied the 3D FWT to small-scale but very complex subsurface models shown in chapter 7: a random medium with a transmission geometry and a 3D expansion of the Marmousi model with a marine reflection geometry. In contrast to a 2D FWT, a different consideration of some aspects is required:

- First of all, the 3D FWT is able to recover reliable velocity models for both the transmission and the reflection experiment.

- The 3D FWT shows a stronger dependency on the acquisition geometry which particularly affects the reflection experiment. While the size of both subsurface models is comparable, the Marmousi experiment is equipped with a more complex acquisition geometry composed of 3 lines with 24 sources and 7 hydrophone lines at the sea surface (compared to 2 boreholes with 16 sources and 2 receiver boreholes in the transmission geometry). However, the reflection experiment seems to be spatially undersampled with respect to the inversion of higher-frequency contents. A limited model resolution and the appearance of artificial noise at very small scale lengths are the immediate consequences.

- Additionally, I implemented the single-frequency method proposed by Sirgue et al. (2008) to carry out a comparison. While this method sufficiently reconstructed the random medium, it was not able to recover the Marmousi model. Even the inversion of lower frequencies got stuck in a local minimum. Obviously, the consideration of broad frequency bands by time-domain FWT mitigates problems caused by the involvement of relatively sparse acquisition geometries.

Apart from the ability to reconstruct the velocity model, the consideration of computational efforts is very important. The feasibility of a 3D FWT does not only depend on the appropriate choices of models and acquisition geometries, but also on the exploitation of helpful features, such as the efficient parallelization and the workflow scheme. In particular, the shot parallelization results in very high computational demands. However, due to the high priority of minimizing computing times, the 3D FWT is targeted to high-performance supercomputers providing a sufficiently high amount of random access memory. Nevertheless, concerning small-scale problems, this work has proven the feasibility of the 3D acoustic FWT in the time domain.

8.2   Conclusion and outlook

This thesis presents a multi-purpose feasibility study of acoustic full waveform tomography in the time domain with focus on the reconstruction of P-wave velocity models. Based on an early motivation to improve the performance of the FWT, the implementation of technical features, such as the massive parallelization and the comprehensive workflow scheme, or methodological improvements, such as the adaptive step length estimation, emerged as important tasks. This paved the way to carry out numerous 2D FWT experiments and to allow the application of the 3D FWT – considering the computational limitations of academic resources.

The FWT implementation developed in the course of this work is the starting point for different investigations. Przebindowska et al. (2012) particularly deal with the inversion of real streamer data recorded in a marine environment. This requires the consideration of essential problems, such as data-preprocessing, the impact of density, an appropriate parameterization or the inversion of the source wavelet, among others. Furthermore, the acoustic reverse time migration (RTM) is an additional component of the 2D implementation (necessary modifications done by Baumann-Wilke, 2009) which takes advantage of the very high methodical similarity of FWT and RTM. For example, Müller et al. (2012) focus on the application of RTM to ultrasonic data by exploiting the RTM capabilities of that implementation.

In particular, the findings of the parameter study, of the investigation of attenuation and of the 3D application will be considered in a subsequent real-data application. The intention is to apply the acoustic FWT to a dataset obtained from a cross-well measurement. However, that requires further development, such as the consideration of attenuation by the 3D FWT, multi-parameter inversions or the usage of different data-misfit definitions.

Bibliography

Akcelik, V. (2002). Multiscale Newton-Krylov methods for inverse acoustic wave propagation. PhD thesis, Carnegie Mellon University.

Aki, K. and Richards, P. (1980). Quantitative seismology. W. H. Freeman and Co.

Alford, R. M., Kelly, K. R., and Boore, D. M. (1974). Accuracy of finite-difference modeling of the acoustic wave equation. Geophysics, 39(6):834–842.

Assous, F. and Collino, F. (1990). A numerical method for the exploration of sensitivity: the case of the identification of the 2D stratified elastic medium. Inverse Problems, 6:487–513.

Attewell, P. B. and Ramana, Y. V. (1966). Wave attenuation and internal friction as functions of frequency in rocks. Geophysics, 31:1049–1056.

Baumann-Wilke, M. (2009). Amplitudenbewahrende akustische Reverse-Time Migration: Theorie und numerische Anwendung. diploma thesis, TU Bergakademie Freiberg.

Ben Hadj Ali, H., Operto, S., and Virieux, J. (2008). Velocity model building by 3D frequency-domain full-waveform inversion of wide-aperture seismic data. Geophysics, 73(5):VE101–VE117.

Ben Hadj Ali, H., Operto, S., and Virieux, J. (2009). Three-dimensional frequency-domain full waveform inversion with phase encoding. In Extended Abstracts, 71st Conference and Technical Exhibition. EAGE. 5812.

Berenger, J. P. (1994). A Perfectly Matched Layer for the Absorption of Electromagnetic Waves. Journal of Computational Physics, 114:185–200.

Beydoun, W. B. and Tarantola, A. (1988). First Born and Rytov approximation: Modeling and inversion conditions in a canonical example. Journal of the Acoustical Society of America, 83:1045–1055.

Blanch, J. O., Robertsson, J. O. A., and Symes, W. W. (1995). Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. Geophysics, 60(1):176–184.

Bohlen, T. (1998). Viskoelastische FD-Modellierung seismischer Wellen zur Interpretation gemessener Seismogramme. PhD thesis, Christian-Albrechts-Universität zu Kiel.

Bohlen, T. (2002). Parallel 3D viscoelastic finite difference seismic modelling. Computers & Geosciences, 28:887–899.

Brenders, A. J. and Pratt, R. G. (2007a). Efficient waveform tomography for lithosperic imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data. Geophysical Journal International, 168:152–170.

Brenders, A. J. and Pratt, R. G. (2007b). Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. Geophysical Journal International, 168:133–151.

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

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

Carcione, J. M. (2001). Wave fields in real media; wave propagation in anisotropic, anelastic and porous media. Amsterdam; New York: Pergamon.

Carcione, J. M., Kosloff, D., and Kosloff, R. (1988a). Wave propagation simulation in a linear viscoacoustic medium. Geophysical Journal of the Royal Astronomical Society, 93(2):393–407.

Carcione, J. M., Kosloff, D., and Kosloff, R. (1988b). Wave propagation simulation in a linear viscoelastic medium. Geophysical Journal of the Royal Astronomical Society, 95(3):597–611.

Causse, E., Mittet, R., and Ursin, B. (1999). Preconditioning of full-waveform inversion in viscoacoustic media. Geophysics, 64(1):130–145.

Chew, W. C. and Weedon, W. H. (1994). A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates. Microwave and Optical Technical Letters, 7(13):599–604.

Cobden, L. J., Tong, C. H., and Warner, M. R. (2011). Inversion of full acoustic wavefield in local helioseismology: A study with synthetic data. The Astrophysical Journal, 727(2):5 pp.

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

Courant, R., Friedrichs, K., and Lewy, H. (1967). On the partial differential equations of mathematical physics. IBM Journal, pages 215–234.

Crase, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A. (1990). Robust elastic nonlinear waveform inversion: Application to real data. Geophysics, 55(5):527–538.

Devaney, A. J. (1984). Geophysical diffraction tomography. Inst. Electr. and Electron. Eng., Trans., Geosci and Remote Sensing, GE-22:3–13.

Dunkl, S., Kurzmann, A., and Bohlen, T. (2013). 3D elastic full waveform inversion of small-scale heterogeneities in transmission geometry. submitted to Geophysical Prospecting, manuscript ID: GP-2012-1429.

Emmerich, H. and Korn, M. (1987). Incorporation of attenuation into time-domain computations of seismic wave fields. Geophysics, 52(9):1252–1264.

Ernst, J. R., Maurer, J. H., Green, A. G., and Holliger, K. (2007). Full-Waveform Inversion of Crosshole Radar Data Based on 2-D Finite-Difference Time-Domain Solutions of Maxwell's Equations. Geoscience and Remote Sensing, 45(9):2807–2828.

Fichtner, A., Kennett, B. L. N., Igel, H., and Bunge, H. (2009). Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods. Geophysical Journal International, 179:1703–1725.

Grote, M. J. and Sim, I. (2009). Perfectly matched layer for the second-order wave equation. In Proc. of 9th Internat. Conf. on Math. and Numerical Aspects of Wave Propagation, pages 370–371. WAVES 2009.

Hak, B. and Mulder, W. (2008). Preconditioning for linearized inversion of attenuation and velocity perturbations. In Expanded abstracts, 78th Annual International Meeting, pages 2036–2040. SEG.

Hak, B. and Mulder, W. A. (2011). Seismic attenuation imaging with causality. Geophysical Journal International, 184:439–451.

Hamilton, E. L. (1972). Compressional-wave attenuation in marine sediments. Geophysics, 37(4):620–646.

Hicks, G. J. and Pratt, R. G. (2001). Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit. Geophysics, 66(2):598–612.

Igel, H., Djikpéssé, H., and Tarantola, A. (1996). Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio. Geophysical Journal International, 124:363–371.

Johnston, D. H. (1981). Attenuation: A state-of-art summary. In Toksöz, M. N. and Johnston, D. H., editors, Seismic Wave attenuation. Geophysics reprint series, no. 2: Society of Exploration Geophysicists.

Kalogeropoulos, A., van der Kruk, J., Hugenschmidt, J., Merz, K., and Busch, S. (2011). Chlorides and moisture assessment in concrete by GPR full waveform inversion. Near Surface Geophysics, 9:277–285.

Kamei, R. and Pratt, R. G. (2008). Waveform tomography strategies for imaging attenuation structure for cross-hole data. In Extended Abstracts, 70th Conference and Technical Exhibition. EAGE. F019.

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

Köhn, D., De Nil, D., Kurzmann, A., Przebindowska, A., and Bohlen, T. (2012). On the influence of model parametrization in elastic full waveform tomography. Geophysical Journal International, 191(1):325–345.

Kolb, P. and Canadas, G. (1986). Least-Squares Inversion of Prestack Data: Simultaneous Identification of Density and Velocity of Stratified Media. In Expanded abstracts, number 5, pages 604–607. SEG.

Kravis, S. P. (1985). Estimation of marine source signatures from direct arrivals to hydrophone groups. Geophysical Prospecting, 33:987–998.

Kurzmann, A., Köhn, D., and Bohlen, T. (2008). Comparison of Acoustic Full Waveform Tomography in the Time- and Frequency-domain. In Extended Abstracts, 70th Conference and Technical Exhibition. EAGE.

Kurzmann, A., Köhn, D., Przebindowska, A., Nguyen, N., and Bohlen, T. (2009). 2D Acoustic Full Waveform Tomography: Performance and Optimization. In Extended Abstracts, 71st Conference and Technical Exhibition. EAGE.

Liu, H. P., Anderson, D. L., and Kanamori, H. (1976). Velocity dispersion due to anelasticity: implications for seismology and mantle composition. Geophysical Journal of the Royal Astronomical Society, 47(1):41–58.

Luenberger, D. (1984). Linear and nonlinear programming, 2nd edition. Addison-Wesley.

Malinowski, M., Operto, S., and Ribodetti, A. (2011). High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full-waveform inversion. Geophysical Journal International, 186:1179–1204.

Martin, G. (2002). Marmousi-2: An updated model for the investigation of AVO in structurally complex areas. In Expanded abstracts, 72nd Annual International Meeting. SEG.

Martin, G. S., Wiley, R., and Marfurt, K. J. (2006). Marmousi2 – An elastic upgrade for Marmousi. The Leading Edge, 25:156–166.

Mora, P. (1987). Nonlinear two-dimensional elastic inversion of multioffset data. Geophysics, 52(9):1211–1228.

Mulder, W. A. and Hak, W. (2009). Simultaneous imaging of velocity and attenuation perturbations from seismic data is nearly impossible. In Extended Abstracts, 71st Conference and Technical Exhibition. EAGE. S043.

Müller, S., Niederleithinger, E., and Bohlen, T. (2012). Reverse Time Migration: A Seismic Imaging Technique Applied to Synthetic Ultrasonic Data. International Journal of Geophysics, 2012:7 pp. Article ID 128465, doi:10.1155/2012/128465.

Nocedal, J. and Wright, S. J. (1999). Numerical Optimization. Springer, New York.

Pica, A. (1990). Nonlinear inversion of seismic reflection data in a laterally invariant medium. Geophysics, 55:284–292.

Plessix, R.-E. (2009). 3D frequency-domain full-waveform inversion with an iterative solver. Geophysics, 74(6):WCC149–WCC157.

Polak, E. and Ribiére, G. (1969). Notes sur la convergence de méthodes de directions conjugées. Revue Fr. Inf. Rech Oper, 16-R1:35–43.

Pratt, R. G. (1990). Inverse theory applied to multi-source cross-hole tomography. Part II: Elastic wave-equation method. Geophysical Prospecting, 38:311–330.

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

Pratt, R. G. and Goulty, N. R. (1991). Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data. Geophysics, 56:204–224.

Pratt, R. G., Huang, L., Duric, N., and Littrup, P. (2007). Sound-speed and attenuation imaging of breast tissue using waveform tomography of transmission ultrasound data. In Proceedings of the Society of Photo-Optical Instrumentation Engineers Conference Series, volume 6510. SPIE. Medical Imaging 2007: Physics of Medical Imaging.

Pratt, R. G., Shin, C., and Hicks, G. J. (1998). Gauss-Newton and Full-Newton methods in frequency-domain seismic waveform inversion. Geophysical Journal International, 133:341–362.

Pratt, R. G., Sirgue, L., Hornby, B., and Wolfe, J. (2008). Cross-well waveform tomography in fine-layered sediments – Meeting the challenges of anisotropy. In Extended Abstracts, 70th Conference and Technical Exhibition. EAGE. F020.

Pratt, R. G. and Worthington, M. H. (1990). Inverse theory applied to multi-source cross-hole tomography, Part I: Acoustic wave-equation method. Geophysical Prospecting, 38:287–310.

Przebindowska, A., Köhn, D., Kurzmann, A., and Bohlen, T. (2012). The Role of Density in Acoustic Full Waveform Inversion of Marine Reflection Seismics. In Extended Abstracts, 74th Conference and Technical Exhibition. EAGE.

Przebindowska, A., Köhn, D., Kurzmann, A., and Bohlen, T. (2013). The role of parametrization in acoustic full waveform inversion of marine reflection seismic data. in preparation, to be submitted to Geophysical Prospecting.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., and Dell'Aversana, P. (2004). Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency domain full-waveform tomography: application to a thrust belt. Geophysical Journal International, 159:1032–1056.

Ricker, N. (1953). The form of laws and propagation of seismic wavelets. Geophysics, 18(1):10–40.

Robertsson, J. O. A., Blanch, J. O., and Symes, W. W. (1994). Viscoelastic finite-difference modeling. Geophysics, 59(9):1444–1456.

Romdhane, G., Grandjean, G., Brossier, R., Rejiba, F., Operto, S., and Virieux, J. (2011). Shallow-structure characterization by 2D elastic full-waveform inversion. Geophysics, 76(3):R81–R93.

Shin, C. and Cha, Y. H. (2009). Waveform inversion in the Laplace-Fourier domain. Geophysical Journal International, 177:1067–1079.

Shipp, R. M. and Singh, S. C. (2002). Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. Geophysical Journal International, 151:325–344.

Sirgue, L. (2006). The importance of low frequency and large offset in waveform inversion. In Extended Abstracts, 68th Conference and Technical Exhibition. EAGE. A037.

Sirgue, L., Etgen, J., and Albertin, U. (2008). 3D frequency domain waveform inversion using time domain finite difference methods. In Extended Abstracts, 70th Conference and Technical Exhibition. EAGE.

Sirgue, L. and Pratt, R. G. (2004). Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. Geophysics, 69(1):231–248.

Smithyman, B., Pratt, R. G., Hayles, J., and Wittebolle, R. (2009). Detecting near-surface objects with seismic waveform tomography. Geophysics, 74(6):WCC119–WCC127.

Sourbier, F., Operto, S., Virieux, J., Amestoy, P., and L'Excellent, J. (2009a). FWT2D: a massively parallel program for frequency domain full-waveform tomography of wide-aperture seismic data – part 1: algorithm. Computer & Geosciences, 35:487–496.

Sourbier, F., Operto, S., Virieux, J., Amestoy, P., and L'Excellent, J. (2009b). FWT2D: a massively parallel program for frequency domain full-waveform tomography of wide-aperture seismic data – part 2: numerical examples and scalability analysis. Computer & Geosciences, 35:496–514.

Sun, R. and McMechan, G. A. (1992). 2-D full-wavefield inversion for wide-aperture, elastic, seismic data. Geophysical Journal International, 111:1–10.

Takam Takougang, E. M. and Calvert, A. J. (2011). Application of waveform tomography to marine seismic reflection data from the Queen Charlotte Basin of western Canada. Geophysics, 76(2):B55–B70.

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

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

Tarantola, A. (1986b). Theoretical Background for the Inversion of Seismic Waveforms, Including Elasticity and Attenuation. Pure and Applied Geophysics, 128(1/2):365–399.

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

Vigh, D. and Starr, E. W. (2008). 3D prestack plane-wave, full-waveform inversion. Geophysics, 73(5):VEI35–VEI44.

Vigh, D., Starr, E. W., and Kapoor, J. (2009). Developing earth models with full waveform inversion. The Leading Edge, 28(4):432–435.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics, 54(4):889–901.

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

Wang, Y. and Rao, Y. (2009). Reflection seismic waveform tomography. Journal of Geophysical Research, 114. B03304, 12 pp., doi: 10.1029/2008JB005916.

Warner, M., Stekl, I., and Umpleby, A. (2007). Full wavefield seismic tomography – Iterative forward modeling in 3D. In Extended Abstracts, 69th Conference and Technical Exhibition. EAGE. C025.

Yuan, X., Borup, D., Wiskin, J., Berggren, M., and Johnson, S. A. (1999). Simulation of Acoustic Wave Propagation in Dispersive Media with Relaxation Losses by Using FDTD Method with PML Absorbing Boundary Condition. IEEE Transactions on ultrasonics, ferroelectrics and frequency control, 46(1):14–23.

List of Tables

List of Figures

Appendix A

The acoustic wave equation

This appendix describes the derivation and finite-difference solution of the 3D and 2D acoustic wave equation. It includes an implementation of perfectly matched layers (PML) proposed by Grote and Sim (2009).

A.1   Continuous formulation

The derivation of the acoustic wave equation with PML boundary condition bases on the homogeneous wave equation

$$\frac{1}{\kappa(\mathbf{x})}\ddot{p}(\mathbf{x},t) = \nabla \cdot \left(\frac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x},t) \right), \tag{2.3}$$

with the bulk modulus

$$\kappa(\mathbf{x}) = \rho(\mathbf{x})\, v_{\mathrm{P}}^2(\mathbf{x}), \tag{2.2}$$

pressure p, P-wave velocity v_{P}, density ρ, the position vector $\mathbf{x}$ and time t. The pressure field and its second time derivative can be transformed to the Laplace-domain yielding the complex wavefield $\underline{p}$:

$$\frac{1}{\rho(\mathbf{x})\, v_{\mathrm{P}}^2(\mathbf{x})}s^2 \underline{p}(\mathbf{x},s) = \nabla \cdot \left(\frac{1}{\rho(\mathbf{x})}\nabla \underline{p}(\mathbf{x},s) \right),$$

$$\frac{1}{v_{\mathrm{P}}^2}s^2 \underline{p} = \left[\partial_x^2 + \partial_y^2 + \partial_z^2 \right] \underline{p} - \frac{1}{\rho}\left[\partial_x\rho\,\partial_x + \partial_y\rho\,\partial_y + \partial_z\rho\,\partial_z \right] \underline{p} \tag{A.1}$$

with $s \in \mathbb{C}$. The next step is a coordinate stretching which takes effect within the absorbing boundary by substituting the differential operators

$$\partial_{x,y,z} \mapsto \frac{1}{1 + \frac{\sigma_{x,y,z}}{s}}\partial_{x,y,z} = \frac{1}{\gamma_{x,y,z}}\partial_{x,y,z}, \tag{A.2}$$

where $\gamma_{x,y,z}$ are auxiliary variables and $\sigma_{x,y,z}$ are the attenuation coefficients in x-, y- and z-direction:

$$\sigma_x = \sigma_y = \sigma_z = 0 \qquad\qquad \text{(domain interior)},$$

$$\sigma_x > 0, \ \sigma_y > 0, \ \sigma_z > 0 \qquad \qquad \text{(PML boundary)}.$$

After application of the new differential operator to equation (A.1)

$$\frac{1}{v_P^2} s^2 \underline{p} = \frac{1}{\gamma_x} \partial_x \left(\frac{1}{\gamma_x} \partial_x \underline{p} \right) + \frac{1}{\gamma_y} \partial_y \left(\frac{1}{\gamma_y} \partial_y \underline{p} \right) + \frac{1}{\gamma_z} \partial_z \left(\frac{1}{\gamma_z} \partial_z \underline{p} \right)$$

$$- \frac{1}{\rho} \left[\frac{1}{\gamma_x^2} \partial_x \rho \cdot \partial_x + \frac{1}{\gamma_y^2} \partial_y \rho \cdot \partial_y + \frac{1}{\gamma_z^2} \partial_z \rho \cdot \partial_z \right] \underline{p}, \qquad (A.3)$$

multiplication of both sides with $\gamma_x \gamma_y \gamma_z$,

$$\frac{1}{v_P^2} \gamma_x \gamma_y \gamma_z s^2 \underline{p} = \partial_x \left(\frac{\gamma_y \gamma_z}{\gamma_x} \partial_x \underline{p} \right) + \partial_y \left(\frac{\gamma_x \gamma_z}{\gamma_y} \partial_y \underline{p} \right) + \partial_z \left(\frac{\gamma_x \gamma_y}{\gamma_z} \partial_z \underline{p} \right)$$

$$- \frac{1}{\rho} \left[\frac{\gamma_y \gamma_z}{\gamma_x} \partial_x \rho \cdot \partial_x \underline{p} + \frac{\gamma_x \gamma_z}{\gamma_y} \partial_y \rho \cdot \partial_y \underline{p} + \frac{\gamma_x \gamma_y}{\gamma_z} \partial_z \rho \cdot \partial_z \underline{p} \right], \qquad (A.4)$$

and evaluating the terms

$$\frac{\gamma_y \gamma_z}{\gamma_x} = 1 + \frac{s \left(\sigma_y + \sigma_z - \sigma_x \right) + \sigma_y \sigma_z}{s \left(s + \sigma_x \right)},$$

$$\frac{\gamma_x \gamma_z}{\gamma_y} = 1 + \frac{s \left(\sigma_x + \sigma_z - \sigma_y \right) + \sigma_x \sigma_z}{s \left(s + \sigma_y \right)},$$

$$\frac{\gamma_x \gamma_y}{\gamma_z} = 1 + \frac{s \left(\sigma_x + \sigma_y - \sigma_z \right) + \sigma_x \sigma_y}{s \left(s + \sigma_z \right)},$$

$$\gamma_x \gamma_y \gamma_z s^2 = s^2 + s \left(\sigma_x + \sigma_y + \sigma_z \right) + \sigma_y \sigma_z + \sigma_x \sigma_y + \sigma_x \sigma_z + \frac{\sigma_x \sigma_y \sigma_z}{s} \qquad (A.5)$$

one can get the following equation:

$$\frac{1}{v_P^2} \left[s^2 + s \left(\sigma_x + \sigma_y + \sigma_z \right) + \sigma_y \sigma_z + \sigma_x \sigma_y + \sigma_x \sigma_z + \frac{\sigma_x \sigma_y \sigma_z}{s} \right] \underline{p} =$$

$$\left[\partial_x^2 + \partial_y^2 + \partial_z^2 \right] \underline{p} - \frac{1}{\rho} \partial_x \rho \cdot \partial_x \underline{p} - \frac{1}{\rho} \partial_y \rho \cdot \partial_y \underline{p} - \frac{1}{\rho} \partial_z \rho \cdot \partial_z \underline{p}$$

$$+ \partial_x \left[\frac{s \left(\sigma_y + \sigma_z - \sigma_x \right) + \sigma_y \sigma_z}{s \left(s + \sigma_x \right)} \partial_x \underline{p} \right] + \partial_y \left[\frac{s \left(\sigma_x + \sigma_z - \sigma_y \right) + \sigma_x \sigma_z}{s \left(s + \sigma_y \right)} \partial_y \underline{p} \right]$$

$$+ \partial_z \left[\frac{s \left(\sigma_x + \sigma_y - \sigma_z \right) + \sigma_x \sigma_y}{s \left(s + \sigma_z \right)} \partial_z \underline{p} \right]$$

$$- \frac{s \left(\sigma_y + \sigma_z - \sigma_x \right) + \sigma_y \sigma_z}{s \left(s + \sigma_x \right)} \frac{1}{\rho} \partial_x \rho \cdot \partial_x \underline{p} - \frac{s \left(\sigma_x + \sigma_z - \sigma_y \right) + \sigma_x \sigma_z}{s \left(s + \sigma_y \right)} \frac{1}{\rho} \partial_y \rho \cdot \partial_y \underline{p}$$

$$- \frac{s \left(\sigma_x + \sigma_y - \sigma_z \right) + \sigma_x \sigma_y}{s \left(s + \sigma_z \right)} \frac{1}{\rho} \partial_z \rho \cdot \partial_z \underline{p}. \qquad (A.6)$$

The replacement of the terms

$$\underline{w}_x := \frac{s\left(\sigma_y + \sigma_z - \sigma_x\right) + \sigma_y\sigma_z}{s\left(s + \sigma_x\right)}\partial_x\underline{p},$$

$$\underline{w}_y := \frac{s\left(\sigma_x + \sigma_z - \sigma_y\right) + \sigma_x\sigma_z}{s\left(s + \sigma_y\right)}\partial_y\underline{p},$$

$$\underline{w}_z := \frac{s\left(\sigma_x + \sigma_y - \sigma_z\right) + \sigma_x\sigma_y}{s\left(s + \sigma_z\right)}\partial_z\underline{p},$$

$$\underline{u} := \frac{\underline{p}}{s} \tag{A.7}$$

by the auxiliary wavefields $\underline{\mathbf{w}}$ and $\underline{u}$ simplifies equation (A.6) and introduces four additional (partial) differential equations. Equations (A.6) and (A.7) are transformed back to the time domain and the system can be written as:

$$\ddot{p} + \left(\sigma_x + \sigma_y + \sigma_z\right)\dot{p} + \left(\sigma_y\sigma_z + \sigma_x\sigma_y + \sigma_x\sigma_z\right)p = v_P^2\left[\Delta p + \nabla\cdot\mathbf{w} - \frac{1}{\rho}\nabla\rho\cdot\left(\nabla p + \mathbf{w}\right)\right]$$
$$- \sigma_x\,\sigma_y\,\sigma_z\,u,$$
$$\dot{w}_x = \left(\sigma_y + \sigma_z - \sigma_x\right)\partial_x p + \sigma_y\sigma_z\partial_x u - \sigma_x w_x,$$
$$\dot{w}_y = \left(\sigma_x + \sigma_z - \sigma_y\right)\partial_y p + \sigma_x\sigma_z\partial_y u - \sigma_y w_y,$$
$$\dot{w}_z = \left(\sigma_x + \sigma_y - \sigma_z\right)\partial_z p + \sigma_x\sigma_y\partial_z u - \sigma_z w_z,$$
$$\dot{u} = p. \tag{A.8}$$

In dependence of the type of application several simplifications are possible. In case of the 2D acoustic wave equation (A.8) is reduced to

$$\ddot{p} + \left(\sigma_x + \sigma_y\right)\dot{p} + \sigma_x\sigma_y\,p = v_P^2\left[\Delta p + \nabla\cdot\mathbf{w} - \frac{1}{\rho}\nabla\rho\cdot\left(\nabla p + \mathbf{w}\right)\right],$$
$$\dot{w}_x = \left(\sigma_y - \sigma_x\right)\partial_x p - \sigma_x w_x,$$
$$\dot{w}_y = \left(\sigma_x - \sigma_y\right)\partial_y p - \sigma_y w_y. \tag{A.9}$$

The assumption of homogeneous density yields modified wave equations in 3D,

$$\ddot{p} + \left(\sigma_x + \sigma_y + \sigma_z\right)\dot{p} + \left(\sigma_y\sigma_z + \sigma_x\sigma_y + \sigma_x\sigma_z\right)p = v_P^2\left[\Delta p + \nabla\cdot\mathbf{w}\right] - \sigma_x\,\sigma_y\,\sigma_z\,u,$$
$$\dot{w}_x = \left(\sigma_y + \sigma_z - \sigma_x\right)\partial_x p + \sigma_y\sigma_z\partial_x u - \sigma_x w_x,$$
$$\dot{w}_y = \left(\sigma_x + \sigma_z - \sigma_y\right)\partial_y p + \sigma_x\sigma_z\partial_y u - \sigma_y w_y,$$
$$\dot{w}_z = \left(\sigma_x + \sigma_y - \sigma_z\right)\partial_z p + \sigma_x\sigma_y\partial_z u - \sigma_z w_z,$$
$$\dot{u} = p, \tag{A.10}$$

and 2D,

$$\ddot{p} + \left(\sigma_x + \sigma_y\right)\dot{p} + \sigma_x\sigma_y\,p = v_P^2\left[\Delta p + \nabla\cdot\mathbf{w}\right],$$

$$\dot{w}_x = (\sigma_y - \sigma_x)\,\partial_x p - \sigma_x w_x,$$
$$\dot{w}_y = (\sigma_x - \sigma_y)\,\partial_y p - \sigma_y w_y. \tag{A.11}$$

The attenuation parameters $\sigma_{x,y,z}$ can be computed from several mathematical functions $f_{x,y,z}$ such as quadratic, exponential or cosine functions. This ensures a smooth transition from the interior of the model domain to the boundary and within the PML boundary. The coefficients are defined as follows:

$$\sigma_{x,y,z} = \sigma_0\, f_{x,y,z} \text{ with } 0 \le f_{x,y,z} \le 1 \text{ and } \sigma_0 = -\frac{\tilde{v}_P\,\ln(R)}{L}, \tag{A.12}$$

where $\tilde{v}_P$ is the average P-wave velocity, L is the width and R is the relative reflection of the absorbing frame, whereas the relative reflection is limited to the range $0 < R \le 1$. Useful relative reflection values have been estimated empirically: $10^{-5} \le R \le 10^{-3}$.

A.2   Finite-difference solution

The finite-difference (FD) solution of the full acoustic wave equation requires the approximation of partial derivatives by using discrete linear FD operators and FD discretization (2.11). The following considerations are limited to second-order FD operators in time and space. For the pressure component p this comprises central derivatives within the interior of the domain:

$$\partial_x^2 p \approx \frac{p_{i+1} + p_{i-1} - 2p_i}{(\Delta h)^2}, \quad \ddot{p} \approx \frac{p^{n+1} + p^{n-1} - 2p^n}{(\Delta t)^2}. \tag{A.13}$$

The usage of PML boundary conditions requires additional first-order equations and corresponding auxiliary variables, such as $\mathbf{w} = (w_x, w_y, w_z)^T$. Exemplary FD approximations are

$$\partial_x w_x \approx \frac{w_{x|i} - w_{x|i-1}}{\Delta h} \quad \text{and} \quad \dot{w}_x \approx \frac{w_x^n - w_x^{n-1}}{\Delta t}. \tag{A.14}$$

The mixture of first-order and second-order equations causes the computations of p at full grid points and time steps, while the components of $\mathbf{w}$ are shifted by half a grid point and half a time step. Thus, additional wavefield averaging is needed within the PML boundary.

The derivation yields the following general relations for pressure field p, auxiliary scalar field u as well as the components of the auxiliary vector field w_x, w_y and w_z. The usage of the auxiliary variables is limited to the PML. The update of the pressure field at time step $n+1$ and grid position (k, j, i) is computed by

$$p_{k,j,i}^{n+1} = s_{k,j,i}\left[v_{P|k,j,i}^2\,\frac{(\Delta t)^2}{(\Delta h)^2}\left(p_{k,j,i-1}^n + p_{k,j,i+1}^n + p_{k,j-1,i}^n + p_{k,j+1,i}^n + p_{k-1,j,i}^n + p_{k+1,j,i}^n - 6p_{k,j,i}^n\right)\right.$$

$$+ v_{P|k,j,i}^2\,\frac{(\Delta t)^2}{\Delta h}\left(\tilde{w}_{x\,k,j,i+\frac{1}{2}}^n - \tilde{w}_{x\,k,j,i-\frac{1}{2}}^n + \tilde{w}_{y\,k,j+\frac{1}{2},i}^n - \tilde{w}_{y\,k,j-\frac{1}{2},i}^n + \tilde{w}_{z\,k+\frac{1}{2},j,i}^n - \tilde{w}_{z\,k-\frac{1}{2},j,i}^n\right)$$

$$+ \left(2 - (\Delta t)^2 \sigma_{y|j}\,\sigma_{z|k} + \sigma_{x|i}\,\sigma_{y|j} + \sigma_{x|i}\,\sigma_{z|k}\right) p_{k,j,i}^n$$

$$+ \left(\frac{\Delta t}{2}\left(\sigma_{x|i} + \sigma_{y|j} + \sigma_{z|k}\right) - 1\right) p_{k,j,i}^{n-1}$$

$$- \sigma_{x|i}\, \sigma_{y|j}\, \sigma_{z|k} \frac{(\Delta t)^2}{2} \left(u_{k,j,i}^{n+\frac{1}{2}} - u_{k,j,i}^{n-\frac{1}{2}} \right)$$

$$+ r_{x|k,j,i} \left(\frac{1}{2\,\Delta h} \left(p_{k,j,i+1}^n - p_{k,j,i-1}^n \right) + \frac{1}{2} \left(\tilde{w}_{x|k,j,i-\frac{1}{2}}^n + \tilde{w}_{x|k,j,i+\frac{1}{2}}^n \right) \right)$$

$$+ r_{y|k,j,i} \left(\frac{1}{2\,\Delta h} \left(p_{k,j+1,i}^n - p_{k,j-1,i}^n \right) + \frac{1}{2} \left(\tilde{w}_{y|k,j-\frac{1}{2},i}^n + \tilde{w}_{y|k,j+\frac{1}{2},i}^n \right) \right)$$

$$\left. + r_{z|k,j,i} \left(\frac{1}{2\,\Delta h} \left(p_{k+1,j,i}^n - p_{k-1,j,i}^n \right) + \frac{1}{2} \left(\tilde{w}_{z|k-\frac{1}{2},j,i}^n + \tilde{w}_{z|k+\frac{1}{2},j,i}^n \right) \right) \right] \tag{A.15}$$

with the PML term

$$s_{k,j,i} = \frac{1}{1 + \frac{\Delta t}{2} \left(\sigma_{x|i} + \sigma_{y|j} + \sigma_{z|k} \right)},$$

the density-related terms

$$r_{x|k,j,i} = -v_{\mathrm{P}|k,j,i}^2 \frac{(\Delta t)^2}{2\,\Delta h} \frac{\rho_{k,j,i+1} - \rho_{k,j,i-1}}{\rho_{k,j,i}},$$

$$r_{y|k,j,i} = -v_{\mathrm{P}|k,j,i}^2 \frac{(\Delta t)^2}{2\,\Delta h} \frac{\rho_{k,j+1,i} - \rho_{k,j-1,i}}{\rho_{k,j,i}},$$

$$r_{z|k,j,i} = -v_{\mathrm{P}|k,j,i}^2 \frac{(\Delta t)^2}{2\,\Delta h} \frac{\rho_{k+1,j,i} - \rho_{k-1,j,i}}{\rho_{k,j,i}}$$

and the averages of the auxiliary wavefields

$$\tilde{w}_{x|k,j,i+\frac{1}{2}}^n = \frac{1}{4} \left(w_{x|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^n + w_{x|k-\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^n + w_{x\,k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}^n + w_{x|k+\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}^n \right),$$

$$\tilde{w}_{x|k,j,i-\frac{1}{2}}^n = \frac{1}{4} \left(w_{x|k+\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}^n + w_{x|k-\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}^n + w_{x\,k-\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}^n + w_{x|k+\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}^n \right),$$

$$\tilde{w}_{y|k,j+\frac{1}{2},i}^n = \frac{1}{4} \left(w_{y|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^n + w_{y|k-\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^n + w_{y\,k-\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}^n + w_{y|k+\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}^n \right),$$

$$\tilde{w}_{y|k,j-\frac{1}{2},i}^n = \frac{1}{4} \left(w_{y|k+\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}^n + w_{y|k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}^n + w_{y\,k-\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}^n + w_{y|k+\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}^n \right),$$

$$\tilde{w}_{z|k+\frac{1}{2},j,i}^n = \frac{1}{4} \left(w_{z|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^n + w_{z|k+\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}^n + w_{z\,k+\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}^n + w_{z|k+\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}^n \right),$$

$$\tilde{w}_{z|k-\frac{1}{2},j,i}^n = \frac{1}{4} \left(w_{z|k-\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^n + w_{z|k-\frac{1}{2},j-\frac{1}{2},i+\frac{1}{2}}^n + w_{z\,k-\frac{1}{2},j-\frac{1}{2},i-\frac{1}{2}}^n + w_{z|k-\frac{1}{2},j+\frac{1}{2},i-\frac{1}{2}}^n \right),$$

$$u_{k,j,i}^n = \frac{1}{2} \left(u_{k,j,i}^{n+\frac{1}{2}} - u_{k,j,i}^{n-\frac{1}{2}} \right).$$

Furthermore, the auxiliary vector wavefield is computed by:

$$w_{x|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^{n+1} = \frac{1}{1 + \frac{\Delta t}{2}\tilde{\sigma}_{x|i+\frac{1}{2}}} \left[\Delta t \left(\tilde{\sigma}_{y|j+\frac{1}{2}} + \tilde{\sigma}_{z|k+\frac{1}{2}} - \tilde{\sigma}_{x|i+\frac{1}{2}} \right) D_x p_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}^{n+\frac{1}{2}} \right.$$

$$+ \Delta t \, \tilde{\sigma}_{y|j+\frac{1}{2}} \tilde{\sigma}_{z|k+\frac{1}{2}} D_x u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \left(1 - \frac{\Delta t}{2} \tilde{\sigma}_{x|i+\frac{1}{2}}\right) w^n_{x|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} \Bigg],$$

$$w^{n+1}_{y|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{1 + \frac{\Delta t}{2} \tilde{\sigma}_{y|j+\frac{1}{2}}} \Bigg[\Delta t \left(\tilde{\sigma}_{x|i+\frac{1}{2}} + \tilde{\sigma}_{z|k+\frac{1}{2}} - \tilde{\sigma}_{y|j+\frac{1}{2}}\right) D_y p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$$

$$+ \Delta t |\tilde{\sigma}_{x|i+\frac{1}{2}} \tilde{\sigma}_{z|k+\frac{1}{2}} D_y u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \left(1 - \frac{\Delta t}{2} \tilde{\sigma}_{y|j+\frac{1}{2}}\right) w^n_{y|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} \Bigg],$$

$$w^{n+1}_{z|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{1 + \frac{\Delta t}{2} \tilde{\sigma}_{z|k+\frac{1}{2}}} \Bigg[\Delta t \left(\tilde{\sigma}_{x|i+\frac{1}{2}} + \tilde{\sigma}_{y|j+\frac{1}{2}} - \tilde{\sigma}_{z|k+\frac{1}{2}}\right) D_z p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$$

$$+ \Delta t \, \tilde{\sigma}_{x|i+\frac{1}{2}} \tilde{\sigma}_{y|j+\frac{1}{2}} D_z u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} + \left(1 - \frac{\Delta t}{2} \tilde{\sigma}_{z|k+\frac{1}{2}}\right) w^n_{z|k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} \Bigg] \quad (A.16)$$

with the averages of the attenuation coefficients

$$\tilde{\sigma}_{x|i+\frac{1}{2}} = \frac{1}{2}\left(\sigma_{x|i} + \sigma_{x|i+1}\right), \quad \tilde{\sigma}_{y|j+\frac{1}{2}} = \frac{1}{2}\left(\sigma_{y|j} + \sigma_{y|j+1}\right), \quad \tilde{\sigma}_{z|k+\frac{1}{2}} = \frac{1}{2}\left(\sigma_{z|k} + \sigma_{z|k+1}\right)$$

and the gradient operators applied to the spatial and temporal averages of the pressure field

$$D_x p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{8 \, \Delta h} \Bigg[\left(p^n_{k,j+1,i} + p^n_{k+1,j+1,i} + p^n_{k+1,j+1,i+1} + p^n_{k,j+1,i+1}\right)$$

$$- \left(p^n_{k,j,i} + p^n_{k+1,j,i} + p^n_{k+1,j,i+1} + p^n_{k,j,i+1}\right)$$

$$+ \left(p^{n+1}_{k,j+1,i} + p^{n+1}_{k+1,j+1,i} + p^{n+1}_{k+1,j+1,i+1} + p^{n+1}_{k,j+1,i+1}\right)$$

$$- \left(p^{n+1}_{k,j,i} + p^{n+1}_{k+1,j,i} + p^{n+1}_{k+1,j,i+1} + p^{n+1}_{k,j,i+1}\right) \Bigg],$$

$$D_y p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{8 \, \Delta h} \Bigg[\left(p^n_{k,j,i+1} + p^n_{k+1,j,i+1} + p^n_{k,j+1,i+1} + p^n_{k+1,j+1,i+1}\right)$$

$$- \left(p^n_{k,j,i} + p^n_{k+1,j,i} + p^n_{k,j+1,i} + p^n_{k+1,j+1,i}\right)$$

$$+ \left(p^{n+1}_{k,j,i+1} + p^{n+1}_{k+1,j,i+1} + p^{n+1}_{k,j+1,i+1} + p^{n+1}_{k+1,j+1,i+1}\right)$$

$$- \left(p^{n+1}_{k,j,i} + p^{n+1}_{k+1,j,i} + p^{n+1}_{k,j+1,i} + p^{n+1}_{k+1,j+1,i}\right) \Bigg],$$

$$D_z p^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}} = \frac{1}{8 \, \Delta h} \Bigg[\left(p^n_{k+1,j,i} + p^n_{k+1,j+1,i} + p^n_{k+1,j+1,i+1} + p^n_{k+1,j,i+1}\right)$$

$$- \left(p^n_{k,j,i} + p^n_{k,j+1,i} + p^n_{k,j+1,i+1} + p^n_{k,j,i+1}\right)$$

$$+ \left(p^{n+1}_{k+1,j,i} + p^{n+1}_{k+1,j+1,i} + p^{n+1}_{k+1,j+1,i+1} + p^{n+1}_{k+1,j,i+1}\right)$$

$$- \left(p^{n+1}_{k,j,i} + p^{n+1}_{k,j+1,i} + p^{n+1}_{k,j+1,i+1} + p^{n+1}_{k,j,i+1}\right) \Bigg].$$

The computation of spatial gradient operators $D_x u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$, $D_y u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$ and $D_z u^{n+\frac{1}{2}}_{k+\frac{1}{2},j+\frac{1}{2},i+\frac{1}{2}}$ is analogous. The auxiliary scalar wavefield is updated by

$$u^{n+\frac{1}{2}}_{k,j,i} = \Delta t\, p^n_{k,j,i} + u^{n-\frac{1}{2}}_{k,j,i}. \tag{A.17}$$

Within the main domain the entire wavefield update simplifies to one second-order equation:

$$
\begin{aligned}
p^{n+1}_{k,j,i} =\, & v^2_{\mathrm{P}|k,j,i}\, \frac{(\Delta t)^2}{(\Delta h)^2} \left(p^n_{k,j,i-1} + p^n_{k,j,i+1} + p^n_{k,j-1,i} + p^n_{k,j+1,i} + p^n_{k-1,j,i} + p^n_{k+1,j,i} - 6 p^n_{k,j,i} \right) \\[2mm]
& + 2 p^n_{k,j,i} - p^{n-1}_{k,j,i} \\[2mm]
& + \frac{r_{x|k,j,i} \left(p^n_{k,j,i+1} - p^n_{k,j,i-1} \right) + r_{y|k,j,i} \left(p^n_{k,j+1,i} - p^n_{k,j-1,i} \right) + r_{z|k,j,i} \left(p^n_{k+1,j,i} - p^n_{k-1,j,i} \right)}{2\,\Delta h}.
\end{aligned}
\tag{A.18}
$$

Appendix B

The viscoacoustic wave equation

This appendix describes the involvement of attenuation into the acoustic wave equation as well as the derivation of its continuous and 2D finite-difference solution. It includes an implementation of perfectly matched layers.

B.1   Attenuation in acoustic media

The implementation of attenuation into acoustic or elastic modeling has been described by numerous authors (compare Carcione et al., 1988a,b; Emmerich and Korn, 1987; Blanch et al., 1995; Liu et al., 1976; Robertsson et al., 1994; Bohlen, 2002). In contrast to modeling in the frequency domain, a direct implementation of attenuation using complex velocities (Johnston, 1981) is not possible. Hence, the attenuative properties of the medium must be approximated by a suitable rheology, which is represented by the generalized standard linear solid (GSLS) (Liu et al., 1976). It consists of a parallel connection of a Hooke and L Maxwell bodies. The Hooke body represents pure acoustic properties. A Maxwell body consists of a Hooke and a Newton body where the latter one characterizes the viscosity of the medium. Thus, attenuation is described by L relaxation mechanisms.

While an acoustic model is characterized by density ρ and bulk modulus κ, the viscoacoustic medium is defined by ρ, the relaxed bulk modulus κ_r and additional $2L$ relaxation parameters $\tau_{p,l}$ and $\tau_{\epsilon,l}$ for each mechanism of the GSLS with $l = \{1, \dots, L\}$. The relaxation times $\tau_{p,l}$ and retardation times $\tau_{\epsilon,l}$ are required for an appropriate approximation of the quality factor Q_P, which is proportional to the reciprocal of attenuation α_P. I only consider constant Q_P models, i.e., $Q_P(f) = const. = Q_{P,0}$, within the desired frequency range $f_{min} \leq f \leq f_{max}$ of the seismic waveforms. The general frequency-dependent relation of $Q_P(\omega)$ (with $\omega = 2\pi f$) and relaxation parameters is given by

$$Q_P(\omega, \tau_{p,l}, \tau_{\epsilon,l}) = \frac{1 - L + \sum_{l=1}^{L} \frac{1 + \omega^2 \tau_{\epsilon,l} \tau_{p,l}}{1 + \omega^2 \tau_{p,l}^2}}{\sum_{l=1}^{L} \frac{\omega(\tau_{\epsilon,l} - \tau_{p,l})}{1 + \omega^2 \tau_{p,l}}}. \tag{B.1}$$

167

This equation is simplified by defining a new L independent parameter ([Blanch et al., 1995](#))

$$\tau_{\mathrm{P}} := \frac{\tau_{\epsilon,l}}{\tau_{p,l}} - 1, \tag{B.2}$$

which replaces the L dependent retardation times, yielding

$$Q_{\mathrm{P}}(\omega, \tau_{p,l}, \tau_{\mathrm{P}}) = \frac{1 + \sum_{l=1}^{L} \frac{\omega^2 \tau_{p,l}^2}{1 + \omega^2 \tau_{p,l}^2} \tau_{\mathrm{P}}}{\sum_{l=1}^{L} \frac{\omega \tau_{p,l}}{1 + \omega^2 \tau_{p,l}^2} \tau_{\mathrm{P}}}. \tag{B.3}$$

To achieve an approximation of $Q_{\mathrm{P}}(\omega) = const. = Q_{\mathrm{P},0}$ the nonlinear equation (B.3) has to be minimized by application of a least-squares optimization algorithm (e. g., [Blanch et al., 1995](#); [Bohlen, 1998](#)). Hence, the constant-Q_{P} model is defined by L relaxation times and $Q_{\mathrm{P},0}$. Instead of defining relaxation times, a more common choice of relaxation frequencies is $\omega_{\mathrm{r},l} = 2\pi f_{\mathrm{r},l} = \frac{1}{\tau_{p,l}}$.

The implementation of attenuation requires τ_{P} rather than Q_{P}. In case of low attenuation, relation (B.3) can be simplified ([Emmerich and Korn, 1987](#); [Blanch et al., 1995](#)) and rewritten in terms of $\tau_{\mathrm{P}}(Q_{\mathrm{P}})$ and $\omega = \omega_0$:

$$\frac{1}{\tau_{\mathrm{P}}} = \sum_{l=1}^{L} \frac{\omega_0/\omega_{\mathrm{r},l}}{1 + \omega_0^2/\omega_{\mathrm{r},l}^2} Q_{\mathrm{P},0} \quad \text{for} \quad Q_{\mathrm{P},0} \gg 1. \tag{B.4}$$

For $\omega_{\mathrm{r},l} := \omega_0$ and $L = 1$ this equation reduces to the approximation

$$\tau_{\mathrm{P}} \approx 2/Q_{\mathrm{P},0}. \tag{B.5}$$

Furthermore, the relaxation parameters are used to compute relaxed frequency-dependent bulk modulus κ_{r} and relaxed P-wave phase velocity v_{P}, respectively:

$$\kappa_{\mathrm{r}} = \rho\, v_{\mathrm{P,ref}}^2 \left(1 + \sum_{l=1}^{L} \frac{\omega_0^2/\omega_{\mathrm{r},l}^2}{1 + \omega_0^2/\omega_{\mathrm{r},l}^2} \tau_{\mathrm{P}} \right)^{-1} \quad \text{for} \quad Q_{\mathrm{P},0} \gg 1 \tag{B.6}$$

with the angular reference frequency $\omega_0 = 2\pi f_0$. At reference frequency ω_0 acoustic reference velocity $v_{\mathrm{P,ref}}$ is defined, i. e., $v_{\mathrm{P,ref}}$ corresponds to the acoustic phase velocity. A useful choice of f_0 is the peak frequency f_{peak} of the source wavelet or seismic data.

Apart from $Q_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right)$, the frequency-dependent dispersion $v_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right)$ is computed from the relaxation parameters. Provided that no dispersion occurs at the reference frequency, it is defined by

$$v_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right) = v_{\mathrm{P,ref}} \left(\sqrt{1 + \sum_{l=1}^{L} \frac{\omega^2 \tau_{p,l}^2}{1 + \omega^2 \tau_{p,l}^2} \tau_{\mathrm{P}}} - \sqrt{1 + \sum_{l=1}^{L} \frac{\omega_0^2 \tau_{p,l}^2}{1 + \omega_0^2 \tau_{p,l}^2} \tau_{\mathrm{P}}} \right). \tag{B.7}$$

The expressions for minimum and maximum dispersion are:

$$\min v_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right) = \lim_{\omega \to 0} v_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right) = v_{\mathrm{P,ref}} \sqrt{\frac{1}{1 + \sum_{l=1}^{L} \frac{\omega_0^2 \tau_{p,l}^2}{1 + \omega_0^2 \tau_{p,l}^2} \tau_{\mathrm{P}}}}, \tag{B.8a}$$

$$\max v_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right) = \lim_{\omega \to \infty} v_{\mathrm{P}}\left(\omega, \tau_{p,l}, \tau_{\mathrm{P}}\right) = v_{\mathrm{P,ref}} \sqrt{\frac{1 + L\,\tau_{\mathrm{P}}}{1 + \sum_{l=1}^{L} \frac{\omega_0^2 \tau_{p,l}^2}{1 + \omega_0^2 \tau_{p,l}^2} \tau_{\mathrm{P}}}}. \tag{B.8b}$$

B.2 Continuous time-domain wave equation

In addition to relaxation of the model parameter, attenuation has to be implemented into the wave equation. This appendix describes the derivation of the viscoacoustic wave equation in the time domain. It includes perfectly matched layers (PML) as boundary condition. This method is based on the application of the so-called complex coordinate stretching (Berenger, 1994; Chew and Weedon, 1994). Although, this implementation uses a similar rheology, it differs from existing work of other authors, such as Yuan et al. (1999).

Due to the requirement of first-order partial differential equations for implementation of attenuation, second-order wave equation (2.3) has to be rewritten to the pressure-velocity formulation which is a system of coupled first-order partial differential equations:

$$\dot{p}\left(\mathbf{x}, t\right) = \kappa\left(\mathbf{x}\right) \, \nabla \cdot \mathbf{w}\left(\mathbf{x}, t\right),$$
$$\dot{\mathbf{w}}\left(\mathbf{x}, t\right) = \frac{1}{\rho\left(\mathbf{x}\right)} \nabla p\left(\mathbf{x}, t\right). \tag{2.7}$$

The auxiliary vector $\mathbf{w}\left(\mathbf{x}, t\right)$ denotes the particle velocity field. The introduction of attenuation yields following wave equation (Emmerich and Korn, 1987; Carcione et al., 1988a; Robertsson et al., 1994):

$$\dot{p}\left(\mathbf{x}, t\right) = \kappa_{\mathrm{r}}\left(\mathbf{x}\right) \, \nabla \cdot \mathbf{w}\left(\mathbf{x}, t\right) \left[1 + \sum_{l=1}^{L}\left(\frac{\tau_{\epsilon,l}}{\tau_{p,l}} - 1\right)\right] + \sum_{l=1}^{L} r_l\left(\mathbf{x}, t\right),$$
$$\dot{r}_l\left(\mathbf{x}, t\right) = -\frac{1}{\tau_{p,l}}\left[\kappa_{\mathrm{r}}\left(\frac{\tau_{\epsilon,l}}{\tau_{p,l}} - 1\right) \, \nabla \cdot \mathbf{w}\left(\mathbf{x}, t\right) + r_l\left(\mathbf{x}, t\right)\right] \quad \text{with} \quad l = \{1, \ldots, L\},$$
$$\dot{\mathbf{w}}\left(\mathbf{x}, t\right) = \frac{1}{\rho\left(\mathbf{x}\right)} \nabla p\left(\mathbf{x}, t\right). \tag{B.9}$$

The substitutions based on relation (B.2) result in

$$\dot{p}\left(\mathbf{x}, t\right) = \kappa_{\mathrm{r}}\left(\mathbf{x}\right) \, \nabla \cdot \mathbf{w}\left(\mathbf{x}, t\right) \left[1 + L\,\tau_{\mathrm{P}}\left(\mathbf{x}\right)\right] + \sum_{l=1}^{L} r_l\left(\mathbf{x}, t\right),$$
$$\dot{r}_l\left(\mathbf{x}, t\right) = -\frac{1}{\tau_{p,l}}\left[\kappa_{\mathrm{r}} \tau_{\mathrm{P}}\left(\mathbf{x}\right) \, \nabla \cdot \mathbf{w}\left(\mathbf{x}, t\right) + r_l\left(\mathbf{x}, t\right)\right] \quad \text{with} \quad l = \{1, \ldots, L\},$$
$$\dot{\mathbf{w}}\left(\mathbf{x}, t\right) = \frac{1}{\rho\left(\mathbf{x}\right)} \nabla p\left(\mathbf{x}, t\right). \tag{2.8}$$

Viscoacoustic modeling in the time domain needs additional L wavefield variables $r_l\left(\mathbf{x}, t\right)$ and L equations. The so-called "memory variables" $r_l\left(\mathbf{x}, t\right)$ characterize the memory of the viscoacoustic medium. According to section 2.3.1.2, initial and PML boundary conditions have to be applied to the wave equation. The initial conditions are

$$p\left(\mathbf{x}, t = 0\right) = \dot{p}\left(\mathbf{x}, t = 0\right) = 0,$$

$$\mathbf{w}\left(\mathbf{x}, t=0\right) = \dot{\mathbf{w}}\left(\mathbf{x}, t=0\right) = 0. \tag{2.9}$$

In the following, the viscoacoustic wave equation with PML boundary condition is derived in 2D based on system (2.8). For subsequent simplifications constant terms are substituted by

$$b := \kappa_{\mathrm{r}}\left(1 + L\,\tau_{\mathrm{P}}\right), \tag{B.10a}$$

$$c := -\frac{1}{\tau_{p,l}}\kappa_{\mathrm{r}}\,\tau_{\mathrm{P}}. \tag{B.10b}$$

In analogy to appendix A.1, the Laplace transform of the wave equation yields the equations with the complex wavefields $\underline{p}$, $\underline{\mathbf{w}}$ and $\underline{r}_l$:

$$s\,\underline{p} = b\left(\partial_{\mathrm{x}}\,\underline{w}_x + \partial_{\mathrm{y}}\,\underline{w}_y\right) + \sum_{l=1}^{L}\underline{r}_l, \tag{B.11a}$$

$$s\,\underline{r}_l = c\left(\partial_{\mathrm{x}}\,\underline{w}_x + \partial_{\mathrm{y}}\,\underline{w}_y\right) - \frac{1}{\tau_{p,l}}\underline{r}_l, \quad l = \{1, \dots, L\}, \tag{B.11b}$$

$$s\,\underline{\mathbf{w}} = \frac{1}{\rho}\nabla\underline{p} \tag{B.11c}$$

A complex coordinate stretching

$$\partial_{x,y} \mapsto \frac{1}{1 + \frac{\sigma_{x,y}}{s}}\partial_{x,y}, \quad s \in \mathbb{C} \tag{B.12}$$

with PML coefficients σ_x and σ_y (defined by (2.6)) is applied. Now equation (B.11a) can be rewritten as

$$s\,\underline{p} = b\left(\frac{1}{1 + \frac{\sigma_x}{s}}\,\partial_{\mathrm{x}}\,\underline{w}_x + \frac{1}{1 + \frac{\sigma_y}{s}}\,\partial_{\mathrm{y}}\,\underline{w}_y\right) + \sum_{l=1}^{L}\underline{r}_l,$$

where the multiplication of both sides with $\left(1 + \frac{\sigma_x}{s}\right)\left(1 + \frac{\sigma_y}{s}\right)$ results in

$$s\,\underline{p} + \left(\sigma_x + \sigma_y\right)\underline{p} + \frac{\sigma_x\sigma_y}{s}\underline{p} = b\left(\partial_{\mathrm{x}}\,\underline{w}_x + \partial_{\mathrm{y}}\,\underline{w}_y + \frac{\sigma_y}{s}\partial_{\mathrm{x}}\,\underline{w}_x + \frac{\sigma_x}{s}\partial_{\mathrm{y}}\,\underline{w}_y\right)$$
$$+ \left(1 + \frac{\sigma_x}{s} + \frac{\sigma_y}{s} + \frac{\sigma_x\sigma_y}{s^2}\right)\sum_{l=1}^{L}\underline{r}_l$$

and the transformation back to time domain yields

$$\dot{p} + \left(\sigma_x + \sigma_y\right)p + \theta p = b\left(\partial_{\mathrm{x}}\,w_x + \partial_{\mathrm{y}}\,w_y + u_x + u_y\right) + \left(1 + \phi_x + \phi_y + \varphi\right)\sum_{l=1}^{L}r_l. \tag{B.13}$$

Relation (B.13) introduces several auxiliary PML variables which are defined by

$$\dot{u}_x = \sigma_y\,\partial_{\mathrm{x}}\,w_x, \tag{B.14a}$$

$$\dot{u}_y = \sigma_x\,\partial_{\mathrm{y}}\,w_y, \tag{B.14b}$$

$$\dot{\theta} = \sigma_x \, \sigma_y, \tag{B.14c}$$

$$\dot{\phi}_x = \sigma_x, \tag{B.14d}$$

$$\dot{\phi}_y = \sigma_y, \tag{B.14e}$$

$$\ddot{\varphi} = \sigma_x \, \sigma_y, \tag{B.14f}$$

where the second-order equation (B.14f) can be written as

$$\dot{\varphi} = \psi \quad \text{with} \quad \dot{\psi} = \sigma_x \, \sigma_y. \tag{B.14f}$$

The relations (B.11b) for memory variables are obtained similarly by using auxiliary variables given by (B.14):

$$\dot{r}_l + \left(\sigma_x + \sigma_y \right) r_l + \theta r_l = c \left(\partial_x \, w_x + \partial_y \, w_y + u_x + u_y \right) - \frac{r_l}{\tau_{p,l}} \left(1 + \phi_x + \phi_y + \varphi \right) \tag{B.15}$$

with $l = \{1, \dots, L\}$. Finally, both components of the vector equation (B.11c) are handled separately. To get the desired partial differential equation for w_x the x-component of relation (B.11c),

$$s \, w_x = \frac{1}{\rho} \frac{1}{1 + \frac{\sigma_x}{s}} \partial_x \, p,$$

has to be multiplied with $\left(1 + \frac{\sigma_x}{s} \right)$ and transformed back to time domain which results in

$$\dot{w}_x = \frac{1}{\rho} \partial_x \, p - \sigma_x w_x. \tag{B.16a}$$

The equation for w_y is computed analogously:

$$\dot{w}_y = \frac{1}{\rho} \partial_y \, p - \sigma_y w_y. \tag{B.16b}$$

The matrix formulation of equations (B.16) can be found in equations (2.10). Furthermore, the choice of appropriate PML coefficients σ_x and σ_y is described in appendix A.1.

B.3   Finite-difference solution

The finite-difference (FD) solution of the full viscoacoustic wave equation requires the approximation of partial derivatives by using discrete linear FD operators and FD discretization (2.11). The following considerations are limited to second-order FD operators in time and space. Furthermore, the pressure-velocity formulation (2.10) requires the staggered-grid implementation (e. g., Virieux, 1986). For pressure p and particle velocities $\mathbf{w}$ this comprises exemplary staggered-grid formulations at grid point i and time step n:

$$\partial_x \, p \approx \frac{p_{i+1} - p_i}{\Delta h}, \qquad\qquad \dot{p} \approx \frac{p^{n+1} - p^n}{\Delta t},$$

$$\partial_x \, w_x \approx \frac{w_{x|i+\frac{1}{2}} - w_{x|i-\frac{1}{2}}}{\Delta h}, \qquad\qquad \dot{w}_x \approx \frac{w_x^{n+\frac{1}{2}} - w_x^{n-\frac{1}{2}}}{\Delta t}, \tag{B.17}$$

Exemplary general relations between f and g within the staggered grid are:

$$f_i \sim \left(g_{i+\frac{1}{2}} - g_{i-\frac{1}{2}} \right) \quad \text{and} \quad g_{i+\frac{1}{2}} \sim \left(f_{i+1} - f_i \right).$$

The complete update of the viscoacoustic wave equation with PML implementation is obtained by the FD solution of relations (2.10) which comprises $(10 + L)$ equations:

- 1 equation for update of pressure p,

- L equations for update of memory variables r_l ($l = \{1, \ldots, L\}$),

- 2 equations for update of particle velocities w_x and w_y,

- 7 equations for additional auxiliary PML variables u_x, u_y, θ, ϕ_x, ϕ_y, φ and ψ.

Within the staggered grid the variables $p, r_l, u_x, u_y, \theta, \phi_x, \phi_y, \varphi$ are computed at grid location and time step $(j, i; n)$, whereas the computation of w_x and w_y is done at $\left(j + \frac{1}{2}, i + \frac{1}{2}; n + \frac{1}{2} \right)$. As an exception the variable ψ is calculated at $\left(j, i; n + \frac{1}{2} \right)$. The FD solution of the pressure wavefield (equation 2.10a) can be written as

$$p_{j,i}^{n+1} = \frac{1}{1 + \frac{\Delta t}{2} e_{j,i}^{n+\frac{1}{2}}} \left\{ b_{j,i} \Delta t \left[\frac{1}{\Delta h} \left(w_{x|j,i+\frac{1}{2}}^{n+\frac{1}{2}} - w_{x|j,i-\frac{1}{2}}^{n+\frac{1}{2}} + w_{y|j+\frac{1}{2},i}^{n+\frac{1}{2}} - w_{y|j-\frac{1}{2},i}^{n+\frac{1}{2}} \right) + \tilde{u}_{x|j,i}^{n+\frac{1}{2}} + \tilde{u}_{y|j,i}^{n+\frac{1}{2}} \right] \right.$$
$$\left. + \Delta t \left(1 + \tilde{\phi}_{x|j,i}^{n+\frac{1}{2}} + \tilde{\phi}_{y|j,i}^{n+\frac{1}{2}} + \tilde{\varphi}_{j,i}^{n+\frac{1}{2}} \right) \sum_{l=1}^{L} \tilde{r}_{l|j,i}^{n+\frac{1}{2}} + p_{j,i}^{n} \left(1 - \frac{\Delta t}{2} e_{j,i}^{n+\frac{1}{2}} \right) \right\}$$

$$\text{(B.18)}$$

with substitution of constant term (see equation B.10a)

$$b_{j,i} = \kappa_{r|j,i} \left(1 + L \, \tau_{P|j,i} \right),$$

the auxiliary term

$$e_{j,i}^{n+\frac{1}{2}} = \left(\sigma_{x|i} + \sigma_{y|j} + \tilde{\theta}_{j,i}^{n+\frac{1}{2}} \right)$$

and the temporal averages

$$\tilde{u}_{x|j,i}^{n+\frac{1}{2}} = \left(u_{x|j,i}^{n} + u_{x|j,i}^{n+1} \right) / 2,$$
$$\tilde{u}_{y|j,i}^{n+\frac{1}{2}} = \left(u_{y|j,i}^{n} + u_{y|j,i}^{n+1} \right) / 2,$$
$$\tilde{\theta}_{j,i}^{n+\frac{1}{2}} = \left(\theta_{j,i}^{n} + \theta_{j,i}^{n+1} \right) / 2,$$
$$\tilde{\phi}_{x|j,i}^{n+\frac{1}{2}} = \left(\phi_{x|j,i}^{n} + \phi_{x|j,i}^{n+1} \right) / 2,$$
$$\tilde{\phi}_{y|j,i}^{n+\frac{1}{2}} = \left(\phi_{y|j,i}^{n} + \phi_{y|j,i}^{n+1} \right) / 2,$$

$$\tilde{\varphi}_{j,i}^{n+\frac{1}{2}} = \left(\varphi_{j,i}^{n} + \varphi_{j,i}^{n+1}\right) / 2,$$

$$\tilde{r}_{l|j,i}^{n+\frac{1}{2}} = \left(r_{l|j,i}^{n} + r_{l|j,i}^{n+1}\right) / 2.$$

The memory variables (equation (2.10b)) are computed by

$$r_{l|j,i}^{n+1} = \frac{1}{1 + \frac{\Delta t}{2\tau_{p,l}}d_{j,i} + \frac{\Delta t}{2}e_{j,i}} \left\{ c_{j,i}\Delta t \left(\frac{w_{x|j,i+\frac{1}{2}}^{n+\frac{1}{2}} - w_{x|j,i-\frac{1}{2}}^{n+\frac{1}{2}} + w_{y|j+\frac{1}{2},i}^{n+\frac{1}{2}} - w_{y|j-\frac{1}{2},i}^{n+\frac{1}{2}}}{\Delta h} + \tilde{u}_{x|j,i}^{n+\frac{1}{2}} + \tilde{u}_{y|j,i}^{n+\frac{1}{2}} \right) \right.$$
$$\left. + \left(1 - \frac{\Delta t}{2\tau_{p,l}}d_{j,i} - \frac{\Delta t}{2}e_{j,i}\right) r_{l|j,i}^{n} \right\} \tag{B.19}$$

with $l = \{1, \ldots, L\}$ and substitution of constant terms (see (B.10b)) as well as the auxiliary term $d_{j,i}$:

$$c_{j,i} = -\frac{1}{\tau_{p,l}}\kappa_{r|j,i}\,\tau_{P|j,i},$$

$$d_{j,i} = \left(1 + \tilde{\phi}_{x|j,i}^{n+\frac{1}{2}} + \tilde{\phi}_{y|j,i}^{n+\frac{1}{2}} + \tilde{\varphi}_{j,i}^{n+\frac{1}{2}}\right).$$

The particle velocities (equation (2.10c)) are obtained by

$$w_{x|j,i+\frac{1}{2}}^{n+\frac{1}{2}} = \frac{1}{1 + \tilde{\sigma}_{x|i+\frac{1}{2}}\Delta t/2} \left[\frac{1}{\tilde{\rho}_{j,i+\frac{1}{2}}} \frac{\Delta t}{\Delta h} \left(p_{j,i+1}^{n} - p_{j,i}^{n}\right) + \left(1 - \tilde{\sigma}_{x|i+\frac{1}{2}}\frac{\Delta t}{2}\right) w_{x|j,i+\frac{1}{2}}^{n-\frac{1}{2}} \right], \tag{B.20a}$$

$$w_{y|j+\frac{1}{2},i}^{n+\frac{1}{2}} = \frac{1}{1 + \tilde{\sigma}_{y|j+\frac{1}{2}}\Delta t/2} \left[\frac{1}{\tilde{\rho}_{j+\frac{1}{2},i}} \frac{\Delta t}{\Delta h} \left(p_{j+1,i}^{n} - p_{j,i}^{n}\right) + \left(1 - \tilde{\sigma}_{y|j+\frac{1}{2}}\frac{\Delta t}{2}\right) w_{y|j+\frac{1}{2},i}^{n-\frac{1}{2}} \right] \tag{B.20b}$$

with the averages of constant terms

$$\tilde{\rho}_{j,i+\frac{1}{2}} = \left(\rho_{j,i} + \rho_{j,i+1}\right) / 2,$$
$$\tilde{\rho}_{j+\frac{1}{2},i} = \left(\rho_{j,i} + \rho_{j+1,i}\right) / 2,$$
$$\tilde{\sigma}_{x|i+\frac{1}{2}} = \left(\sigma_{x|i} + \sigma_{x|i+1}\right) / 2,$$
$$\tilde{\sigma}_{y|j+\frac{1}{2}} = \left(\sigma_{y|j} + \sigma_{y|j+1}\right) / 2.$$

The FD solutions for the remaining wavefield variables (equations (2.10d) ... (2.10j)) are

$$u_{x|j,i}^{n+1} = \sigma_{y|j} \frac{\Delta t}{\Delta h} \left(w_{x|j,i+\frac{1}{2}}^{n+\frac{1}{2}} - w_{x|j,i-\frac{1}{2}}^{n+\frac{1}{2}} \right) + u_{x|j,i}^n, \tag{B.21a}$$

$$u_{y|j,i}^{n+1} = \sigma_{x|i} \frac{\Delta t}{\Delta h} \left(w_{y|j+\frac{1}{2},i}^{n+\frac{1}{2}} - w_{y|j-\frac{1}{2},i}^{n+\frac{1}{2}} \right) + u_{y|j,i}^n, \tag{B.21b}$$

$$\theta_{j,i}^{n+1} = \Delta t\, \sigma_{x|i} \sigma_{y|j} + \theta_{j,i}^n, \tag{B.21c}$$

$$\phi_{x|j,i}^{n+1} = \Delta t\, \sigma_{x|i} + \phi_{x|j,i}^n, \tag{B.21d}$$

$$\phi_{y|j,i}^{n+1} = \Delta t\, \sigma_{y|j} + \phi_{y|j,i}^n, \tag{B.21e}$$

$$\varphi_{j,i}^{n+1} = \Delta t\, \psi_{j,i}^{n+\frac{1}{2}} + \varphi_{j,i}^n, \tag{B.21f}$$

$$\psi_{j,i}^{n+\frac{1}{2}} = \Delta t\, \sigma_{x|i} \sigma_{y|j} + \psi_{j,i}^{n-\frac{1}{2}}. \tag{B.21g}$$

Appendix C

The FWT method

This appendix describes the derivations of model corrections employing the Born approximation (section C.1) and model updates using the conjugate gradient method (section C.2). The subsequent considerations are strongly based on the work of Tarantola (1984) and Mora (1987). They also include aspects of Crase et al. (1990).

C.1 Derivation of model corrections

Definition of forward and adjoint problem

This section discusses the computation of model corrections in the acoustic approximation with respect to the model parameters bulk modulus, density and source time function. To simplify following equations the derivation is carried out for model parameters $\mathbf{m} := \mathbf{m}_h$ at iteration h and a single source. The final equations take all sources into account.

The starting point is the inhomogeneous second-order acoustic wave equation

$$\left[\frac{1}{\kappa} \partial_t^2 - \nabla \cdot \left(\frac{1}{\rho} \nabla \right) \right] p = q, \tag{C.1}$$

where the pressure data $p := p(\mathbf{x}_s, \mathbf{x}_r, t, \mathbf{m})$ is observed at receiver locations $\mathbf{x}_r$ resulting from a source at $\mathbf{x}_s$. The medium is described by the model parameters bulk modulus κ and density ρ. Using the Dirac distribution $\delta(\mathbf{x})$ and the source time function $q(t)$ the source term is given by a point source

$$q := q(\mathbf{x}, t) = \delta(\mathbf{x} - \mathbf{x}_s) \, q(t). \tag{C.2}$$

In the following both continuous and discrete matrix formulations are used, e.g., the seismic forward problem can be written in continuous $p := p(\mathbf{m})$ or discrete form $\mathbf{p} := \mathbf{p}(\mathbf{m})$ with vectorized data. The linearized forward problem can be written as

$$\delta \mathbf{p} = \mathbf{p}(\mathbf{m} + \delta \mathbf{m}) - \mathbf{p}(\mathbf{m}) = \frac{\partial \mathbf{p}}{\partial \mathbf{m}}(\delta \mathbf{m}) = \mathbf{D}\,\delta \mathbf{m} + \mathbf{o}\left(\|\delta \mathbf{m}\|^2 \right) \tag{C.3}$$

with the linear operator $\mathbf{D} = \partial \mathbf{p} / \partial \mathbf{m}$ denoting the Frechét derivatives. Equation (C.3) describes a small change in the pressure wavefield $\delta \mathbf{p}$ due to a small model perturbation

$\delta\mathbf{m}$. Hence, higher-order terms $\mathbf{o}\left(\|\delta\mathbf{m}\|^2\right)$ are neglected. For next steps the continuous formulation of (C.3) will be helpful:

$$\delta\mathbf{p}\left(D\right) = \int_M dM \frac{\partial\mathbf{p}\left(D\right)}{\partial\mathbf{m}\left(M\right)} \delta\mathbf{m}\left(M\right). \tag{C.4}$$

This is an integration over model space M (mapped to data space D) of model perturbations multiplied by Frechét derivatives. It represents the Born approximation of the forward problem (cp. Mora (1987)) by introducing the Green's function. However, the aim of an inversion is the determination of model corrections $\delta\hat{\mathbf{m}}$ obtained by an unknown expression which is equivalent to (C.3) and is called its adjoint operation:

$$\delta\hat{\mathbf{m}} = \mathbf{D}^* \delta\mathbf{p}. \tag{C.5}$$

In analogy to (C.4), the continuous form of (C.5) is

$$\delta\hat{\mathbf{m}}\left(M\right) = \int_D dD \left[\frac{\partial\mathbf{p}\left(D\right)}{\partial\mathbf{m}\left(M\right)}\right]^* \delta\mathbf{p}\left(D\right). \tag{C.6}$$

Here it is more comprehensible to assign a physical meaning to the model parameters, data and model space. The forward and adjoint problems of the acoustic wave equation are parameterized by bulk modulus, density and source term:

$$\mathbf{m}(\mathbf{x}) = \begin{pmatrix} \boldsymbol{\kappa}(\mathbf{x}) \\ \boldsymbol{\rho}(\mathbf{x}) \\ \mathbf{q}\left(\mathbf{x}\right) \end{pmatrix}. \tag{C.7}$$

Equivalent expressions of the forward problem (C.4) and its adjoint operation (C.6) are given by

$$\delta\mathbf{p}(\mathbf{x}_s, \mathbf{x}_r, t) = \int_V dV \frac{\partial\mathbf{p}(\mathbf{x}_s, \mathbf{x}_r, t)}{\partial\mathbf{m}\left(\mathbf{x}\right)} \delta\mathbf{m}\left(\mathbf{x}\right), \tag{C.4}$$

$$\delta\hat{\mathbf{m}}\left(\mathbf{x}\right) = \int_t dt \sum_{N_r} \frac{\partial\mathbf{p}(\mathbf{x}_s, \mathbf{x}_r, t)}{\partial\mathbf{m}\left(\mathbf{x}\right)} \delta\mathbf{p}(\mathbf{x}_s, \mathbf{x}_r, t), \tag{C.6}$$

where N_r denotes the number of receivers. In contrast to the forward problem, the adjoint operation is an integration over data space of data residuals multiplied by Frechét derivatives (cp. Tarantola, 1984; Mora, 1987). To obtain the adjoint operation $\delta\hat{\mathbf{m}}$ the Born approximation of the acoustic wave equation (C.4) is required, because it provides the pressure perturbation $\delta\mathbf{p}$ corresponding to some perturbation in model parameters $\delta\mathbf{m}$.

The forward problem

The involvement of perturbations $\mathbf{p} + \delta\mathbf{p}$, $\boldsymbol{\kappa} + \delta\boldsymbol{\kappa}$, $\boldsymbol{\rho} + \delta\boldsymbol{\rho}$ and $\mathbf{q} + \delta\mathbf{q}$ yields a new acoustic wave equation describing pressure perturbations $\delta\mathbf{p}$:

$$\left[\frac{1}{\kappa + \delta\kappa} \partial_t^2 - \nabla \cdot \left(\frac{1}{\rho + \delta\rho} \nabla\right)\right] (p + \delta p) = q + \delta q. \tag{C.8}$$

The term $\frac{1}{a+\delta a}$ can be approximated by

$$\frac{1}{a+\delta a} = \frac{1}{a} - \frac{\delta a}{a^2} + \mathbf{o}\left(\delta a^2\right)$$

with neglection of higher-order terms due to the assumption of small perturbations in obtainment of Frechét derivatives. The resulting wave equation is

$$\left[\frac{1}{\kappa}\partial_t^2 - \nabla \cdot \frac{1}{\rho}\nabla\right]p - \left[\frac{\delta\kappa}{\kappa^2}\partial_t^2 - \nabla \cdot \frac{\delta\rho}{\rho^2}\nabla\right]p + \left[\frac{1}{\kappa}\partial_t^2 - \nabla \cdot \frac{1}{\rho}\nabla\right]\delta p - \left[\frac{\delta\kappa}{\kappa^2}\partial_t^2 - \nabla \cdot \frac{\delta\rho}{\rho^2}\nabla\right]\delta p = q + \delta q.$$
(C.9)

Simplification by omitting terms of the original wave equation on both sides and neglecting the last summand on the left side containing higher-order terms yields

$$\left[\frac{1}{\kappa}\partial_t^2 - \nabla \cdot \left(\frac{1}{\rho}\nabla\right)\right]\delta p = \left[\frac{\delta\kappa}{\kappa^2}\partial_t^2 - \nabla \cdot \left(\frac{\delta\rho}{\rho^2}\nabla\right)\right]p + \delta q$$
(C.10)

with the new source term

$$\Delta q = \left[\frac{\delta\kappa}{\kappa^2}\partial_t^2 - \nabla \cdot \left(\frac{\delta\rho}{\rho^2}\nabla\right)\right]p + \delta q.$$
(C.11)

Due to the equivalence of equation (C.9) with (C.1), it can be solved in terms of Green's functions $G(\mathbf{x}_r, t; \mathbf{x}, 0)$ (Aki and Richards, 1980) of acoustic wave equation:

$$\delta p(\mathbf{x}_r, t) = \int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \Delta q(\mathbf{x}_r, t)$$
(C.12)

where "$*$" denotes the convolution in time. It resembles the continuous form of equation (C.4). Thus, it can be used to define the Frechét derivatives. Replacing the source term in (C.12) by (C.11) gives

$$\begin{aligned}
\delta p(\mathbf{x}_r, t) &= \int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \left[\delta q + \frac{\delta\kappa}{\kappa^2}\partial_t^2 p - \nabla \cdot \left(\frac{\delta\rho}{\rho^2}\nabla p\right)\right] \\
&= \int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \delta q + \int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \left(\frac{\delta\kappa}{\kappa^2}\partial_t^2 p\right) \\
&\quad - \int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \left[\nabla \cdot \left(\frac{\delta\rho}{\rho^2}\nabla p\right)\right].
\end{aligned}$$
(C.13)

Hence, the solution of the acoustic forward problem is given by the general function

$$\mathbf{p} = \mathbf{p}(\mathbf{m}) = \mathbf{p}(\kappa, \rho, \mathbf{q}).$$
(C.14)

Furthermore, the perturbations of model parameters introduce the operators U, W and T via the nonlinear function $\mathbf{p}$,

$$\mathbf{p}\left[\kappa(\mathbf{x}) + \delta\kappa, \rho(\mathbf{x}) + \delta\rho, \mathbf{q}(t) + \delta\mathbf{q}\right] = \mathbf{p}(\kappa, \rho, \mathbf{q}) + U\delta\kappa + W\delta\rho + T\delta\mathbf{q} + \mathbf{o}\left(\delta\kappa, \delta\rho, \delta\mathbf{q}\right)^2,$$
(C.15)

where U, W and T represent the derivatives of the pressure wavefield with respect to $\boldsymbol{\kappa}$, $\boldsymbol{\rho}$ and $\mathbf{q}$:

$$U(\mathbf{x}_s, \mathbf{x}_r, t) = \frac{\partial p(\mathbf{x}_s, \mathbf{x}_r, t)}{\partial \kappa(\mathbf{x})}, \tag{C.16a}$$

$$W(\mathbf{x}_s, \mathbf{x}_r, t) = \frac{\partial p(\mathbf{x}_s, \mathbf{x}_r, t)}{\partial \rho(\mathbf{x})}, \tag{C.16b}$$

$$T(\mathbf{x}_s, \mathbf{x}_r, t') = \frac{\partial p(\mathbf{x}_s, \mathbf{x}_r, t)}{\partial q(t')}. \tag{C.16c}$$

Using (C.15) the perturbation $\delta\mathbf{p}$ can be written as

$$\begin{aligned}
\delta\mathbf{p} &= \mathbf{p}\left[\boldsymbol{\kappa}(\mathbf{x}) + \delta\boldsymbol{\kappa}, \boldsymbol{\rho}(\mathbf{x}) + \delta\boldsymbol{\rho}, \mathbf{q}(t) + \delta\mathbf{q}\right] - \mathbf{p}(\boldsymbol{\kappa}, \boldsymbol{\rho}, \mathbf{q}) \\
&= U\delta\boldsymbol{\kappa} + W\delta\boldsymbol{\rho} + T\delta\mathbf{q} + \mathbf{o}(\delta\boldsymbol{\kappa}, \delta\boldsymbol{\rho}, \delta\mathbf{q})^2.
\end{aligned} \tag{C.17a}$$

After neglection of higher-order terms this equation can be written as

$$\delta p(\mathbf{x}_s, \mathbf{x}_r, t) = \int_V dV\, U(\mathbf{x}_s, \mathbf{x}_r, t)\, \delta\kappa(\mathbf{x}) + \int_V dV\, W(\mathbf{x}_s, \mathbf{x}_r, t)\, \delta\rho(\mathbf{x}) + \int_t dt'\, T(\mathbf{x}_s, \mathbf{x}_r, t)\, \delta q(t') \tag{C.17b}$$

which resembles equation (C.4) and defines the Frechét derivatives. Relations (C.17b) and (C.13) yield

$$[U\,\delta\kappa](\mathbf{x}_s, \mathbf{x}_r, t) = \int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \left[\frac{\delta\kappa(\mathbf{x})}{\kappa^2(\mathbf{x})} \partial_t^2 p(\mathbf{x}_s, \mathbf{x}_r, t)\right], \tag{C.18a}$$

$$[W\,\delta\rho](\mathbf{x}_s, \mathbf{x}_r, t) = -\int_V dV\, G(\mathbf{x}_r, t; \mathbf{x}, 0) * \left[\nabla \cdot \left(\frac{\delta\rho(\mathbf{x})}{\rho^2(\mathbf{x})} \nabla p(\mathbf{x}_s, \mathbf{x}_r, t)\right)\right], \tag{C.18b}$$

$$[T\,\delta q](\mathbf{x}_s, \mathbf{x}_r, t') = G(\mathbf{x}_r, t; \mathbf{x}, 0) * \delta q(t'). \tag{C.18c}$$

On condition of $\lim_{t\to\infty} g = 0$ and $\lim_{t\to\infty} p = 0$, one can use a property of convolution, $\partial_t f(t) * g(t) = f(t) * \partial_t g(t)$, and rewrite (C.18a):

$$[U\,\delta\kappa](\mathbf{x}_s, \mathbf{x}_r, t) = \int_V dV\, \frac{1}{\kappa^2(\mathbf{x})} \partial_t G(\mathbf{x}_r, t; \mathbf{x}, 0) * \partial_t p(\mathbf{x}_s, \mathbf{x}_r, t)\, \delta\kappa(\mathbf{x}). \tag{C.19a}$$

Applying the product rule

$$\nabla \cdot (a\mathbf{v}) = a\,\nabla \cdot \mathbf{v} + \mathbf{v} \cdot \nabla a$$

to (C.18b) yields

$$g * \left[\nabla \cdot \left(\frac{\delta\rho}{\rho^2}\nabla p\right)\right] = \nabla \cdot \left[g * \left(\frac{\delta\rho}{\rho^2}\nabla p\right)\right] - (\nabla g) * \left(\frac{\delta\rho}{\rho^2}\nabla p\right),$$

which results in

$$[W\,\delta\rho](\mathbf{x}_s, \mathbf{x}_r, t) = -\int_V dV\, \nabla \cdot \left[G(\mathbf{x}_r, t; \mathbf{x}, 0) * \left(\frac{\delta\rho(\mathbf{x})}{\rho^2(\mathbf{x})} \nabla p(\mathbf{x}_s, \mathbf{x}_r, t)\right)\right]$$

$$+ \int\limits_V dV \ (\nabla G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}, 0)) * \left(\frac{\delta\rho(\mathbf{x})}{\rho^2(\mathbf{x})} \nabla p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \right)$$

$$= \int\limits_V dV \ \frac{\delta\rho(\mathbf{x})}{\rho^2} \ \nabla G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}, 0) * \nabla p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \,. \tag{C.19b}$$

This equation is obtained by applying GAUSS' divergence theorem to the first integral over volume V. It can be rewritten as an integral over the surface A of the earth which vanishes due to the assumption of homogeneous boundary conditions:

$$\int\limits_V dV \ \nabla \cdot \left[g * \left(\frac{\delta\rho}{\rho^2} \nabla p \right) \right] = \int\limits_A dA \ \mathbf{n} \cdot \left[g * \left(\frac{\delta\rho}{\rho^2} \right) \right] := 0 .$$

Now the kernels of U, W and T can be deduced from (C.16), (C.17b), (C.19a), (C.19b) and (C.18c):

$$U(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) = \frac{\partial p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)}{\partial\kappa(\mathbf{x})} = \frac{1}{\kappa^2(\mathbf{x})} \partial_t \, G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}, 0) * \partial_t \, p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \,, \tag{C.20a}$$

$$W(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) = \frac{\partial p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)}{\partial\rho(\mathbf{x})} = \frac{1}{\rho^2(\mathbf{x})} \nabla G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}, 0) * \nabla p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \,, \tag{C.20b}$$

$$T(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t') = \frac{\partial p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)}{\partial q(t')} = G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}_\mathrm{s}, t') \,. \tag{C.20c}$$

The adjoint problem

The integration of Frechét derivatives defined by (C.17b) and (C.20) over data space (cp. (C.6)) produces the adjoint operation

$$\hat{\mathbf{m}}(\mathbf{x}) = \begin{pmatrix} \delta\hat{\boldsymbol{\kappa}}(\mathbf{x}) \\ \delta\hat{\boldsymbol{\rho}}(\mathbf{x}) \\ \delta\hat{\mathbf{q}}(\mathbf{x}) \end{pmatrix} \tag{C.21}$$

with the model corrections – incorporating the data residuals $\delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)$ –

$$\delta\hat{\kappa}(\mathbf{x}) = \int\limits_t dt \ \sum_{N_\mathrm{r}} U(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \ \delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)$$

$$= \frac{1}{\kappa^2(\mathbf{x})} \int\limits_t dt \ \sum_{N_\mathrm{r}} \partial_t \, G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}, 0) * \partial_t \, p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \ \delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \,, \tag{C.22a}$$

$$\delta\hat{\rho}(\mathbf{x}) = \int\limits_t dt \ \sum_{N_\mathrm{r}} W(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \ \delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)$$

$$= \frac{1}{\rho^2(\mathbf{x})} \int\limits_t dt \ \sum_{N_\mathrm{r}} \nabla G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}, 0) * \nabla p(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \ \delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \,, \tag{C.22b}$$

$$\delta\hat{q}(\mathbf{x}) = \int\limits_t dt \ \sum_{N_\mathrm{r}} T(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \ \delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t)$$

$$= \int\limits_t dt \ \sum_{N_\mathrm{r}} G(\mathbf{x}_\mathrm{r}, t; \mathbf{x}_\mathrm{s}, t') \ \delta\hat{p}(\mathbf{x}_\mathrm{s}, \mathbf{x}_\mathrm{r}, t) \,. \tag{C.22c}$$

Propagation of residuals backward in time

The new wavefield

$$p'(\mathbf{x}_s, \mathbf{x}_r, t) = \int_V dV \, G(\mathbf{x}, -t; \mathbf{x}_r, 0) * \delta q(\mathbf{x}', \mathbf{x}_s, t) \tag{C.23}$$

for back-propagation is introduced by using the property of convolution to shift its location,

$$\int dt \, [f(t) * g(t)] \, h(t) = \int dt \, f(-t) \, [g(t) * h(-t)] \,,$$

and the reciprocity property of the Green's function, i.e., source and receiver positions can be interchanged with identical responses at $\mathbf{x}$ due to source $\mathbf{x}'$ and vice versa:

$$g(\mathbf{x}, t; \mathbf{x}', 0) = g(\mathbf{x}', t; \mathbf{x}, 0) \,.$$

The source term in (C.23) can be written in terms of data residuals,

$$\delta q(\mathbf{x}', \mathbf{x}_s, t) = \sum_{N_r} \delta(\mathbf{x} - \mathbf{x}_r) \, \delta \hat{p}(\mathbf{x}_s, \mathbf{x}_r, t) \,, \tag{C.24}$$

which represents point sources at receiver positions $\mathbf{x}_r$. Equation (C.23) describes the propagation of the residual wavefield backward in time (indicated by $-t$ in the Green's function) with respect to a certain source location $\mathbf{x}_s$. The residuals contain information of the data which is not explained by the current model parameters of the forward operation, i.e., $p'(\mathbf{x}_s, \mathbf{x}_r, t)$ can be called the "missing diffracted field". Relation (C.24) shows that the source of back-propagation is a superposition of the residuals $p'(\mathbf{x}_s, \mathbf{x}_r, t)$ at all receiver locations. Hence, for each source position $\mathbf{x}_s$ a forward propagation of a single source wavefield and a back-propagation of a superimposed receiver wavefield are required. However, the term "back-propagation" is only a formal description. Due to time symmetry of the wave equation, in forward- and back-propagation the wave equation is solved by forward modeling.

Model corrections

Using (C.23), (C.24) and the commutativity property of the convolution as well as involving all sources N_s, (C.22) can be rewritten yielding final model corrections

$$\delta\hat{\kappa}(\mathbf{x}) = \frac{1}{\kappa^2(\mathbf{x})} \sum_{N_s} \int_t dt \, \dot{p}(\mathbf{x}_s, \mathbf{x}_r, t) \left[\sum_{N_r} \dot{G}(\mathbf{x}, -t; \mathbf{x}_r, 0) * \delta\hat{p}(\mathbf{x}_s, \mathbf{x}_r, t) \right]$$

$$= \frac{1}{\kappa^2(\mathbf{x})} \sum_{N_s} \int_t dt \, \dot{p}(\mathbf{x}_s, \mathbf{x}_r, t) \, \dot{p}'(\mathbf{x}_s, \mathbf{x}_r, t) \,, \tag{C.25a}$$

$$\delta\hat{\rho}(\mathbf{x}) = \frac{1}{\rho^2(\mathbf{x})} \sum_{N_s} \int_t dt \, \nabla p(\mathbf{x}_s, \mathbf{x}_r, t) \cdot \sum_{N_r} \nabla G(\mathbf{x}, -t; \mathbf{x}_r, 0) * \delta\hat{p}(\mathbf{x}_s, \mathbf{x}_r, t)$$

$$= \frac{1}{\rho^2(\mathbf{x})} \sum_{N_s} \int_t dt \, \nabla p(\mathbf{x}_s, \mathbf{x}_r, t) \cdot \nabla p'(\mathbf{x}_s, \mathbf{x}_r, t) \,, \tag{C.25b}$$

$$\delta\hat{q}(\mathbf{x}) = \sum_{N_s} \sum_{N_r} G(\mathbf{x}, -t; \mathbf{x}_r, 0) * \delta\hat{p}(\mathbf{x}_s, \mathbf{x}_r, t)$$

$$= \sum_{N_s} p'(\mathbf{x_s}, \mathbf{x_r}, t). \tag{C.25c}$$

Basically, model corrections (C.25) for bulk modulus and density are obtained by a cross-correlation in time of incident forward wavefield and back-propagated residual ("missing diffracted") wavefield (Devaney, 1984). In general, the imaging condition of inverse problems is represented by a convolution. This coincides with the given FWT problem due to the fact that a cross-correlation corresponds to a convolution in reverse time direction.

C.2   Conjugate gradient method and model update

The misfit function: method of least squares

The gradient method of the inverse problem discussed in this work bases on the method of nonlinear least squares (cp. Mora, 1987). The forward problem is solved by $\mathbf{p} = \mathbf{p}(\mathbf{m})$ (see equation (C.14)) whereas in general, the observed data is not explained by the given initial model $\mathbf{m}_{apr}$. The aim of the nonlinear inversion is to find the pair $(\mathbf{p}, \mathbf{m})$ which minimizes the distance to the observed data as well as the a priori model $(\mathbf{p}_{obs}, \mathbf{m}_{apr})$. Hereby the least squares (L_2) functional is used:

$$E\left[\mathbf{p}\left(\mathbf{m}\right), \mathbf{m}\right] = \|(\Delta\mathbf{p}, \Delta\mathbf{m})\|_2^2 = \|\Delta\mathbf{p}\|_2^2 + \|\Delta\mathbf{m}\|_2^2, \tag{2.15}$$

with

$$\Delta\mathbf{p} = \mathbf{p}\left(\mathbf{m}\right) - \mathbf{p}_{obs},$$
$$\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_{apr}.$$

In (2.15) and subsequent equations the observed pressure data are given by $\mathbf{p}_{obs}$ and the a priori model is denoted by

$$\mathbf{m}_{apr} = \begin{pmatrix} \boldsymbol{\kappa}_{apr} \\ \boldsymbol{\rho}_{apr} \\ \mathbf{q}_{apr} \end{pmatrix}. \tag{C.26}$$

In matrix notation (2.15) can be written as

$$E\left[\mathbf{p}\left(\mathbf{m}\right), \mathbf{m}\right] = \frac{1}{2}\Delta\mathbf{p}^* \, \mathbf{C_p} \, \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{m}^* \, \mathbf{C_m} \, \Delta\mathbf{m}, \tag{C.27}$$

where "*" denotes the conjugate transpose of a vector. $\mathbf{C_p}$ and $\mathbf{C_m}$ represent matrices with respect to data and model space. They weight both data and model residuals and can contain a priori information, such as describing uncorrelated noise in the data ($\mathbf{C_p}$) and including smoothness constraints to the model ($\mathbf{C_m}$) (cp. Mora, 1987). However, in general, only diagonal elements of these matrices are non-zero.

Definition of data residuals

The definition of data residuals depends on the choice of the misfit function. Apart from the least squares norm, there are alternatives, such as L_1- or CAUCHY-criterion (see Crase et al., 1990). On the one hand, they may provide a better robustness of the inversion due to low

amplitude related sensitivity. But, on the other hand, the corresponding residuals may cause problems in time-domain FWT, i.e., especially in modeling. Thus, this work concentrates on the usage of least squares misfit function. A general misfit function E with respect to data and model space can be written as a sum over all data samples N_t and model parameters N_m:

$$E[\mathbf{p(m)}] = \sum_{n=1}^{N_t} Y\left[p^n(\mathbf{m}) - p^n_{\text{apr}}\right] \tag{C.28}$$

with the least-squares kernels of E,

$$\mathbf{Y}_p(\Delta\mathbf{p}) = \frac{1}{2}\Delta\mathbf{p}^* \, \mathbf{C_p} \, \Delta\mathbf{p} \quad \text{and} \quad \mathbf{Y}_m(\Delta\mathbf{m}) = \frac{1}{2}\Delta\mathbf{m}^* \, \mathbf{C_m} \, \Delta\mathbf{m}. \tag{C.29}$$

The residuals $\delta\hat{\mathbf{p}}$ (cp. appendix C.1) are defined by a derivative of the kernels $\mathbf{Y}$ of $E[\mathbf{p(m)}]$ with respect to the data of the forward problem (Crase et al., 1990):

$$\delta\hat{\mathbf{p}} = \frac{\partial\mathbf{Y}_p}{\partial\mathbf{p(m)}} + \frac{\partial\mathbf{Y}_m}{\partial\mathbf{p(m)}} = \Delta\mathbf{p}. \tag{C.30}$$

The resulting least-squares functional is

$$E\left[\mathbf{p}\left(\mathbf{m}\right), \mathbf{m}\right] = \frac{1}{2}\Delta\mathbf{p}^* \, \mathbf{C_p} \, \Delta\mathbf{p} + \frac{1}{2}\Delta\mathbf{m}^* \, \mathbf{C_m} \, \Delta\mathbf{m}. \tag{C.31}$$

The model update: method of steepest descent gradient

The update of the model parameter employs the method of steepest descent gradient. The linearization of the forward problem (Tarantola, 1984; Mora, 1987),

$$\delta\mathbf{p} = \mathbf{p}\left(\mathbf{m} + \delta\mathbf{m}\right) - \mathbf{p}\left(\mathbf{m}\right) = \frac{\partial\mathbf{p}}{\partial\mathbf{m}}\left(\delta\mathbf{m}\right) = \mathbf{D}\,\delta\mathbf{m}, \tag{C.3}$$

is required and involves the Frechét derivatives $\mathbf{D}$ (see appendix C.1). The partial derivation of the functional E with respect to $\mathbf{m}$ yields the steepest ascent gradient vector $\mathbf{g}$:

$$\mathbf{g} = \frac{\partial E}{\partial\mathbf{m}} = \left(\frac{\partial\mathbf{p}}{\partial\mathbf{m}}\right)^* \mathbf{C_p} \, \Delta\mathbf{p} + \mathbf{C_m} \, \Delta\mathbf{m} = \mathbf{D}^* \, \mathbf{C_p} \, \Delta\mathbf{p} + \mathbf{C_m} \, \Delta\mathbf{m}. \tag{C.32}$$

Due to algebraic reasons, the (conjugate) transpose of the Frechét derivatives matrix, $\mathbf{D}^*$, has to be used. In algebra the application of the conjugate transpose of Frechét derivatives is equivalent to the adjoint operation (see equations (C.5) and (C.6)). On condition of $\mathbf{g} = 0$ the minimization problem is solved. Hence, previous expression has the form

$$0 = \mathbf{D}^* \, \mathbf{C_p} \, \Delta\mathbf{p} + \mathbf{C_m} \, \Delta\mathbf{m} \tag{C.33a}$$

$$\mathbf{C_m}\left(\mathbf{m}_{\text{apr}} - \mathbf{m}\right) = \mathbf{D}^* \, \mathbf{C_p} \left[\mathbf{p(m)} - \mathbf{p}_{\text{obs}}\right]. \tag{C.33b}$$

Adding $\mathbf{D}^*\mathbf{D}\left(\mathbf{m}_{\text{apr}} - \mathbf{m}\right)$ to both sides of previous equation yields

$$\left(\mathbf{C_m} + \mathbf{D}^*\mathbf{D}\right)\left(\mathbf{m}_{\text{apr}} - \mathbf{m}\right) = \mathbf{D}^*\mathbf{D}\left(\mathbf{m}_{\text{apr}} - \mathbf{m}\right) + \mathbf{D}^* \, \mathbf{C_p} \left[\mathbf{p(m)} - \mathbf{p}_{\text{obs}}\right]$$

$$= \mathbf{D}^* \left\{ \mathbf{D} \left(\mathbf{m}_{\mathrm{apr}} - \mathbf{m} \right) + \mathbf{C_p} \left[\mathbf{p}(\mathbf{m}) - \mathbf{p}_{\mathrm{obs}} \right] \right\}. \tag{C.33c}$$

Due to positive definiteness of $\mathbf{D}^*\mathbf{D}$, (C.33b) can be rewritten:

$$\mathbf{m}_{\mathrm{apr}} - \mathbf{m} = \left(\mathbf{C_m} + \mathbf{D}^*\mathbf{D} \right)^{-1} \mathbf{D}^* \left\{ \mathbf{D} \left(\mathbf{m}_{\mathrm{apr}} - \mathbf{m} \right) + \mathbf{C_p} \left[\mathbf{p}(\mathbf{m}) - \mathbf{p}_{\mathrm{obs}} \right] \right\}. \tag{C.33d}$$

The solution is the desired model

$$\mathbf{m} = \mathbf{m}_{\mathrm{apr}} - \left(\mathbf{C_m} + \mathbf{D}^*\mathbf{D} \right)^{-1} \mathbf{D}^* \left\{ \mathbf{D} \left(\mathbf{m}_{\mathrm{apr}} - \mathbf{m} \right) + \mathbf{C_p} \left[\mathbf{p}(\mathbf{m}) - \mathbf{p}_{\mathrm{obs}} \right] \right\} \tag{C.33e}$$

which is obtained in one step. However, in a nonlinear inversion this solution is not applicable. Thus, an iterative solution is given by

$$\mathbf{m}_h = \mathbf{m}_{\mathrm{apr}} - \left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right)^{-1} \mathbf{D}_h^* \left\{ \mathbf{D}_h \left(\mathbf{m}_{\mathrm{apr}} - \mathbf{m}_{h-1} \right) + \mathbf{C_p} \left[\mathbf{p}(\mathbf{m}_{h-1}) - \mathbf{p}_{\mathrm{obs}} \right] \right\} \tag{C.34a}$$

where h denotes the iteration. Several algebraic operations simplify expression (C.34a):

$$\mathbf{m}_h = -\left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right)^{-1} \left\{ -\left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right) \mathbf{m}_{\mathrm{apr}} + \mathbf{D}_h^*\mathbf{D}_h \left(\mathbf{m}_{\mathrm{apr}} - \mathbf{m}_{h-1} \right) + \mathbf{D}_h^*\mathbf{C_p} \left[\mathbf{p}(\mathbf{m}_{h-1}) - \mathbf{p}_{\mathrm{obs}} \right] \right\}$$

$$= -\left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right)^{-1} \left\{ -\left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right) \mathbf{m}_{h-1} + \mathbf{C_m} \left[\mathbf{m}_{h-1} - \mathbf{m}_{\mathrm{apr}} \right] + \mathbf{D}_h^*\mathbf{C_p} \left[\mathbf{p}(\mathbf{m}_{h-1}) - \mathbf{p}_{\mathrm{obs}} \right] \right\}$$

$$= \mathbf{m}_{h-1} - \left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right)^{-1} \left\{ \mathbf{D}_h^*\mathbf{C_p} \left[\mathbf{p}(\mathbf{m}_{h-1}) - \mathbf{p}_{\mathrm{obs}} \right] + \mathbf{C_m} \left[\mathbf{m}_{h-1} - \mathbf{m}_{\mathrm{apr}} \right] \right\}$$

$$= \mathbf{m}_{h-1} - \mu_h \mathbf{H}_h \left\{ \mathbf{D}_h^*\mathbf{C_p} \left[\mathbf{p}(\mathbf{m}_{h-1}) - \mathbf{p}_{\mathrm{obs}} \right] + \mathbf{C_m} \left[\mathbf{m}_{h-1} - \mathbf{m}_{\mathrm{apr}} \right] \right\} \tag{C.34b}$$

with step length μ_h of the gradient algorithm and the inverse Hessian matrix

$$\mathbf{H}_h = \left(\mathbf{C_m} + \mathbf{D}_h^*\mathbf{D}_h \right)^{-1}. \tag{C.35}$$

The computation of $\mathbf{H}_h$ is highly demanding (Pratt et al., 1998). Hence, using a gradient method the inverse Hessian matrix can be approximated by

$$\mathbf{H}_h :\approx \mathbf{I}. \tag{C.36}$$

The term $\{\cdot\}$ in equation (C.34b) can be replaced by the steepest ascent gradient $\mathbf{g}_h$ (C.32):

$$\mathbf{m}_h = \mathbf{m}_{h-1} - \mu_h \mathbf{I} \mathbf{g}_h. \tag{C.37}$$

However, the choice of $\mathbf{H}$ may have tremendous effect on the success of the inversion. The approximation (C.36) represents the simplest case. Using (C.5) and (C.32) the steepest ascent gradient can be written in terms of model perturbations $\delta\mathbf{m}$ (derived in appendix C.1). At h^{th} iteration it is given by

$$\mathbf{g}_h = \delta\hat{\mathbf{m}}_h + \left(\mathbf{m}_{h-1} - \mathbf{m}_{\mathrm{apr}} \right). \tag{C.38}$$

Additionally, instead of minimizing the model norm $\|\Delta\mathbf{m}\|_2^2$ of model at h^{th} iteration and an a priori model, the definition $\mathbf{m}_{\mathrm{apr}} := \mathbf{m}_{h-1}$ implies the neglection of an a priori model. Thus, the model parameters are updated by

$$\mathbf{m}_h = \mathbf{m}_{h-1} - \mu_h \, \delta\hat{\mathbf{m}}_h. \tag{C.39a}$$

The explicit expressions for $\boldsymbol{\kappa}, \boldsymbol{\rho}$ and $\mathbf{q}$ are given by

$$\boldsymbol{\kappa}_h = \boldsymbol{\kappa}_{h-1} - \mu_h \, \delta\hat{\boldsymbol{\kappa}}_h,$$
$$\boldsymbol{\rho}_h = \boldsymbol{\rho}_{h-1} - \mu_h \, \delta\hat{\boldsymbol{\rho}}_h,$$
$$\mathbf{q}_h = \mathbf{q}_{h-1} - \mu_h \, \delta\hat{\mathbf{q}}_h. \tag{C.39b}$$

The conjugate gradient

Furthermore, the convergence of steepest descent gradient algorithm is improved by employing the preconditioned conjugate gradient method. At h^{th} iteration a modification of the Polak-Ribiére conjugate direction $\mathbf{c}_h$ is computed by using a linear combination of gradient directions from current and previous iteration (Polak and Ribiére, 1969; Luenberger, 1984; Nocedal and Wright, 1999):

$$\mathbf{c}_h = \boldsymbol{\beta}_h + \frac{\boldsymbol{\beta}_h^* \left(\boldsymbol{\beta}_h - \boldsymbol{\beta}_{h-1} \right)}{\boldsymbol{\beta}_{h-1}^* \boldsymbol{\beta}_{h-1}} \, \mathbf{c}_{h-1}. \tag{C.40}$$

This relation requires the preconditioned gradient $\boldsymbol{\beta}$ which is obtained by application of a preconditioning operator P_h to the steepest ascent gradient:

$$\boldsymbol{\beta}_h = P_h \, \mathbf{g}_h. \tag{C.41}$$

In this work, P_h is used to weight $\mathbf{g}_h$, i. e., it damps source artifacts or excludes the model update at predefined locations.

Change of parameterization

Apart from the parameterization $\mathbf{m} = (\boldsymbol{\kappa}, \boldsymbol{\rho}, \mathbf{q})^T$, it is desirable to have P-wave velocity $\mathbf{v}_{\text{P}}$ instead of bulk modulus $\boldsymbol{\kappa}$. According to equation (C.38), the gradient with respect to $\boldsymbol{\kappa}$ is given by

$$\mathbf{g}_{\boldsymbol{\kappa}} = \delta\hat{\kappa}(\mathbf{x}) + \Delta\boldsymbol{\kappa}. \tag{C.42a}$$

The gradient in terms of the new parameter $\mathbf{v}_{\text{P}}$ is defined by

$$\mathbf{g}_{\mathbf{v}_{\text{P}}} = \frac{\partial \boldsymbol{\kappa}}{\partial \mathbf{v}_{\text{P}}} \, \delta\hat{\kappa}(\mathbf{x}) + \Delta\mathbf{v}_{\text{P}}, \tag{C.42b}$$

involving the Jacobian (Mora, 1987)

$$J_{\mathbf{v}_{\text{P}}} := \frac{\partial \kappa}{\partial v_{\text{P}}} = 2\,\rho\,v_{\text{P}}. \tag{C.43}$$

To avoid the usage of ambiguous mathematical operators, the resulting iterative update of P-wave velocity is written in index notation at spatial coordinates (j, i):

$$v_{\text{P}|h|j,i} = v_{\text{P}|h-1|j,i} \,-\, \mu_h \, J_{\mathbf{v}_{\text{P}}|h|j,i} \, \delta\hat{\kappa}_{h|j,i}. \tag{C.44}$$

Appendix D

Application 1: Parameter study

This appendix contains additional information with respect to the FWT application 1 discussed in chapter 4. Appendix D.1 consists of a table with computational facts of all experiments. Appendix D.2 comprises results of all experiments I to X.

D.1 Basic setup

Table D.1: Application 1: Computational summary of all FWT experiments.

Experiment	Brute-force parameters			FWT iterations	Total number of time-domain …	
	Initial background models		runs / model		finite-difference modelings	finite-difference time steps
	Number	Velocity range				
I					163 200	571 200 000
II/1					163 200	434 112 000, 571 200 000
II/2					163 200	571 200 000
II/3	51	[1430, 2970] $\frac{\text{m}}{\text{s}}$	100 iterations	5 100	163 200	571 200 000
III/1					163 200	571 200 000
III/2					163 200	571 200 000
III/3					163 200	571 200 000
IV/1	51	[1430, 2970] $\frac{\text{m}}{\text{s}}$	100 iterations	5 100	285 600	999 600 000
IV/2	1	2200 $\frac{\text{m}}{\text{s}}$	400 iterations	400	12 800	44 800 000
IV/3	1	2200 $\frac{\text{m}}{\text{s}}$	400 iterations	400	22 400	78 400 000
IV/4	1	2200 $\frac{\text{m}}{\text{s}}$	400 iterations	400	14 000 … 32 000	49 000 000 … 112 000 000
V/1	51	[1430, 2970] $\frac{\text{m}}{\text{s}}$	100 iterations	5 100	10 200 … 652 800	35 700 000 … 2 284 800 000
V/2					163 200	571 200 000
VI/1					285 600	999 600 000
VI/2					1 142 400	3 998 400 000
VII/1					163 200	571 200 000
VII/2	51	[1430, 2970] $\frac{\text{m}}{\text{s}}$	100 iterations	5 100	285 600	999 600 000
VII/3					163 200	571 200 000
VII/4					285 600	999 600 000
VIII/1					163 200	571 200 000
VIII/2					285 600	999 600 000
IX					285 600	999 600 000
X	51	[1430, 2970] $\frac{\text{m}}{\text{s}}$	1001 step lengths	51	410 040	1 435 140 000

D.2   Results of FWT experiments

Experiment I: Basic FWT
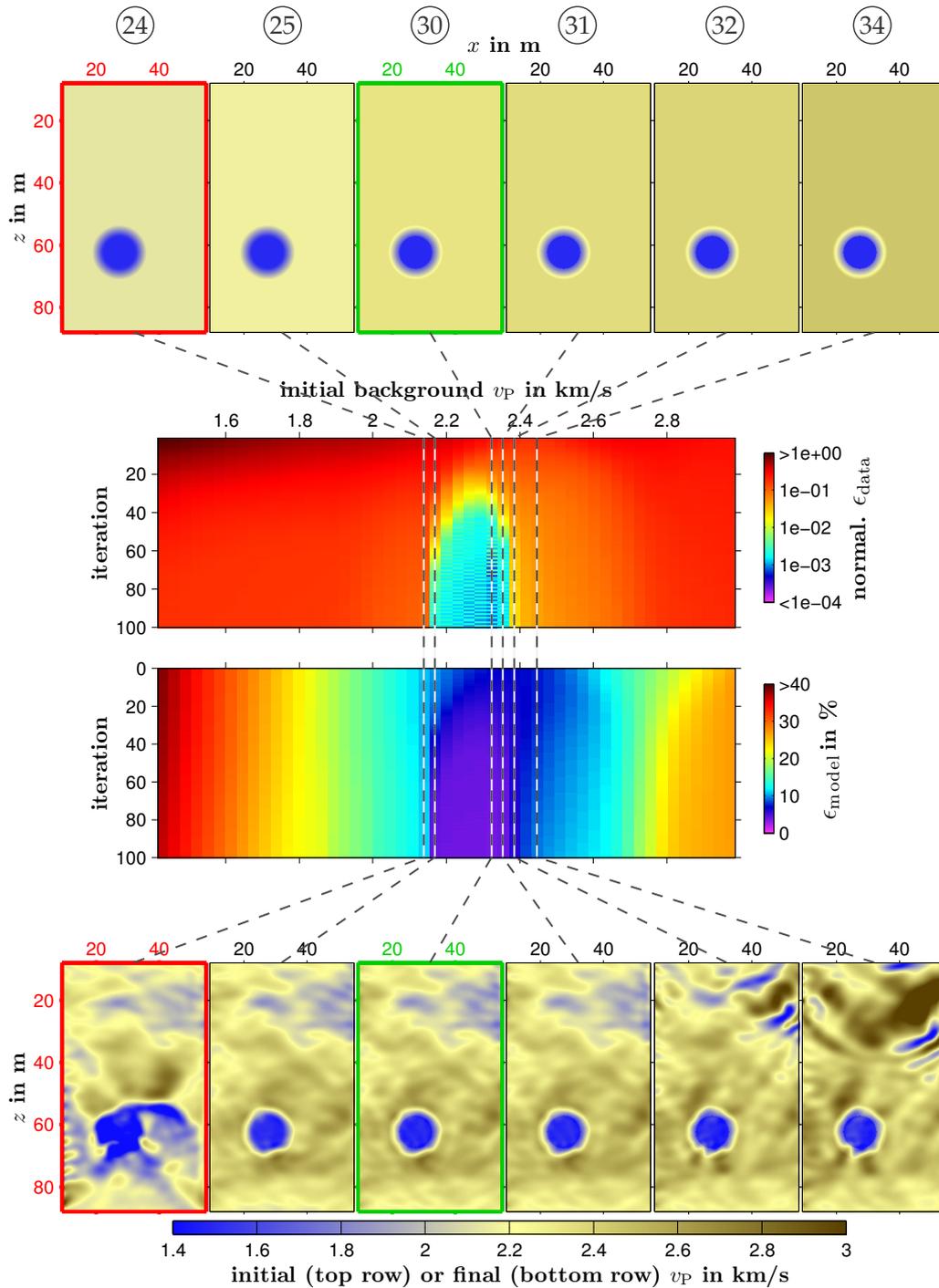
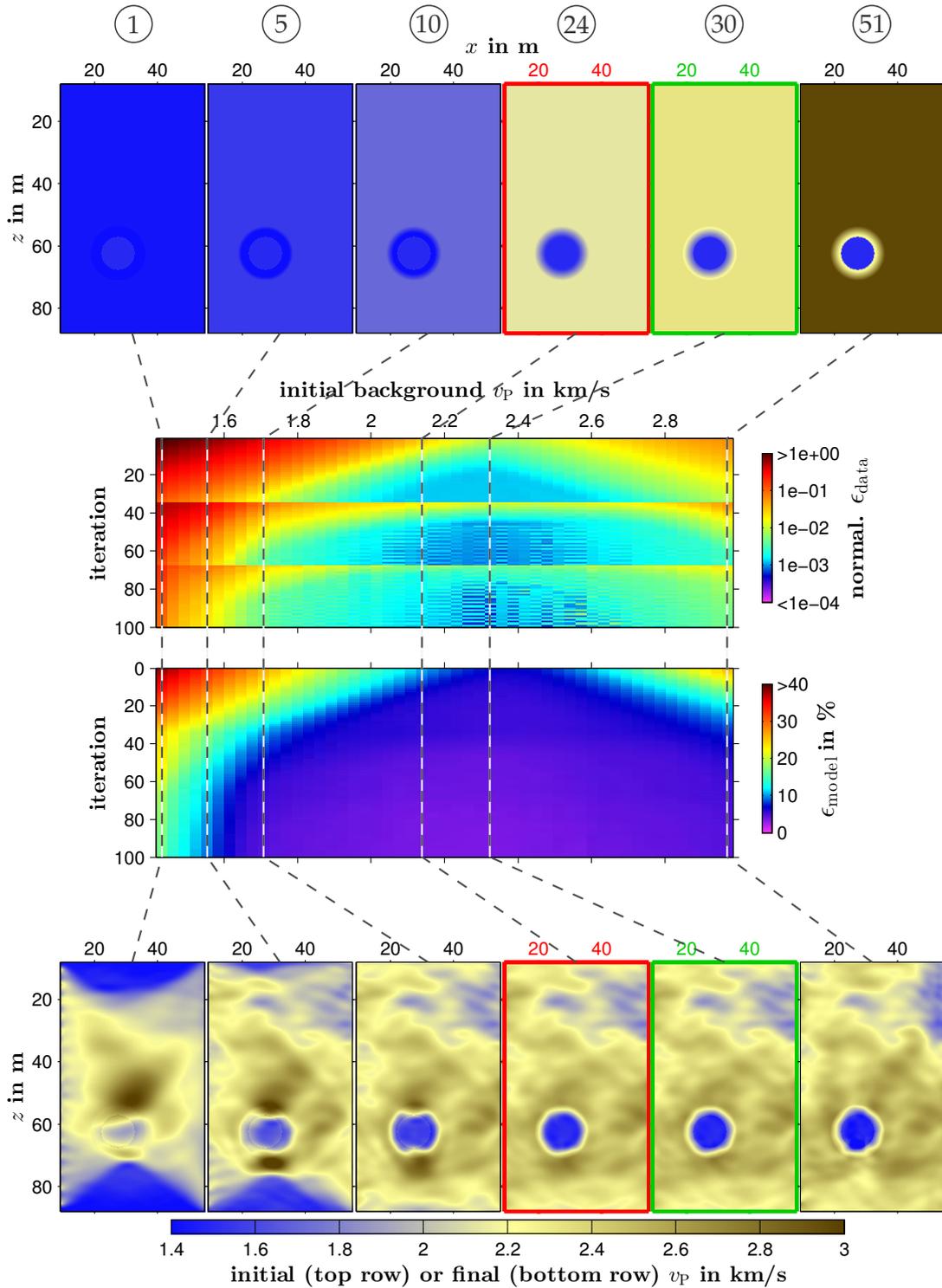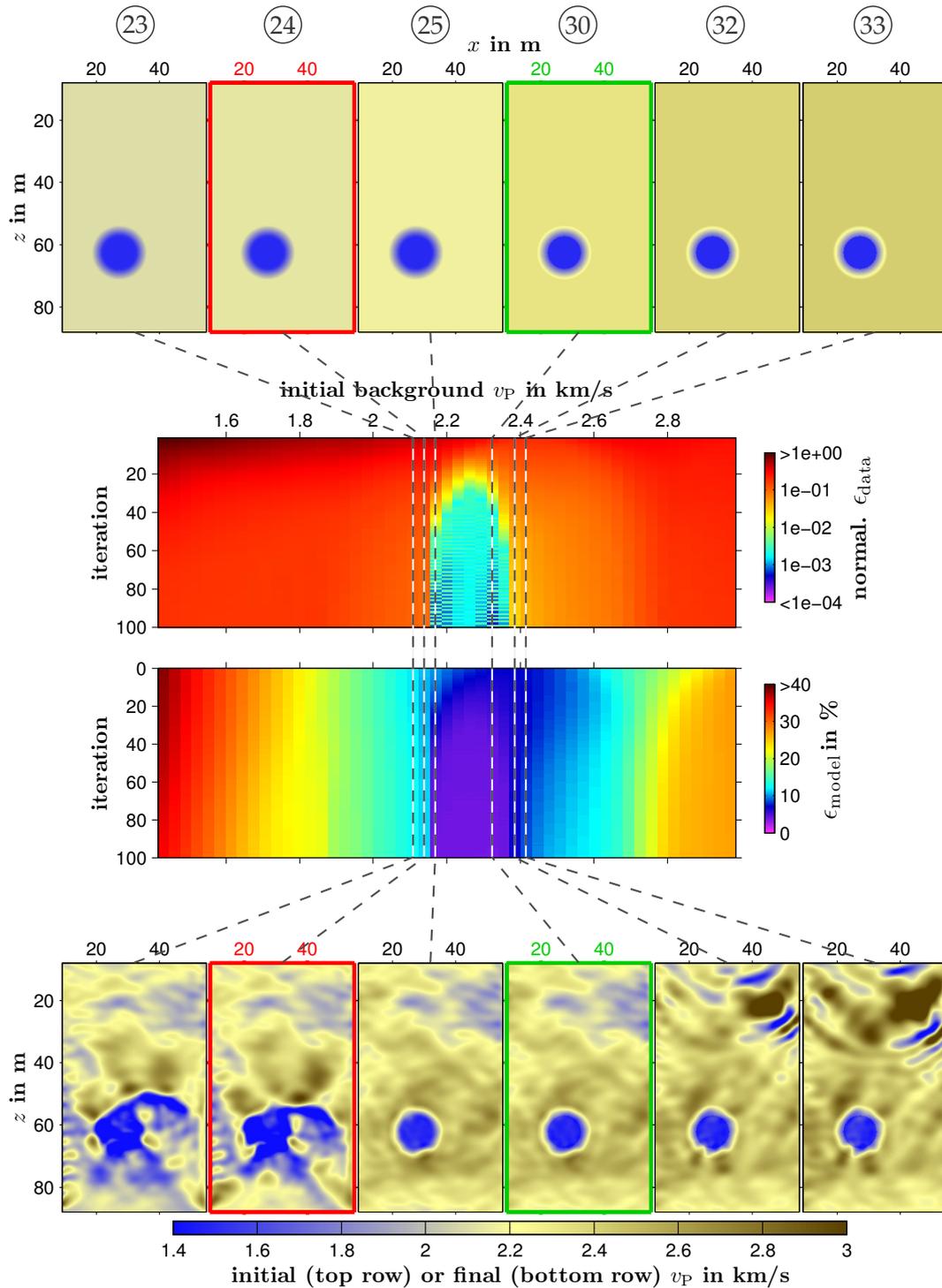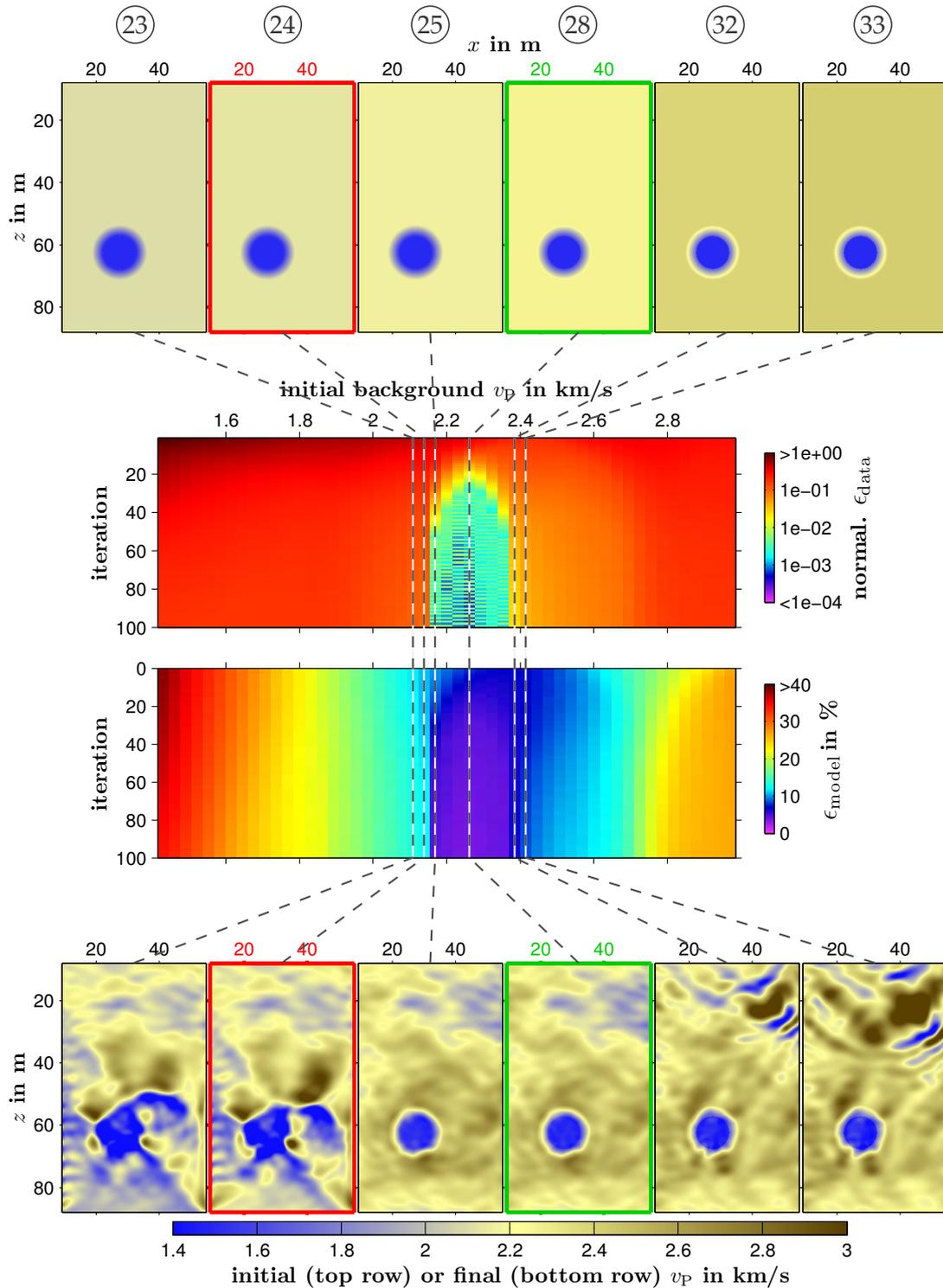**Figure D.1:** Application 1, experiment I: selected initial models (top) and v_P results (bottom) as well as the progress of the reference FWT for all 51 initial background models (central part). The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 6.7 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 3.8)$ %.
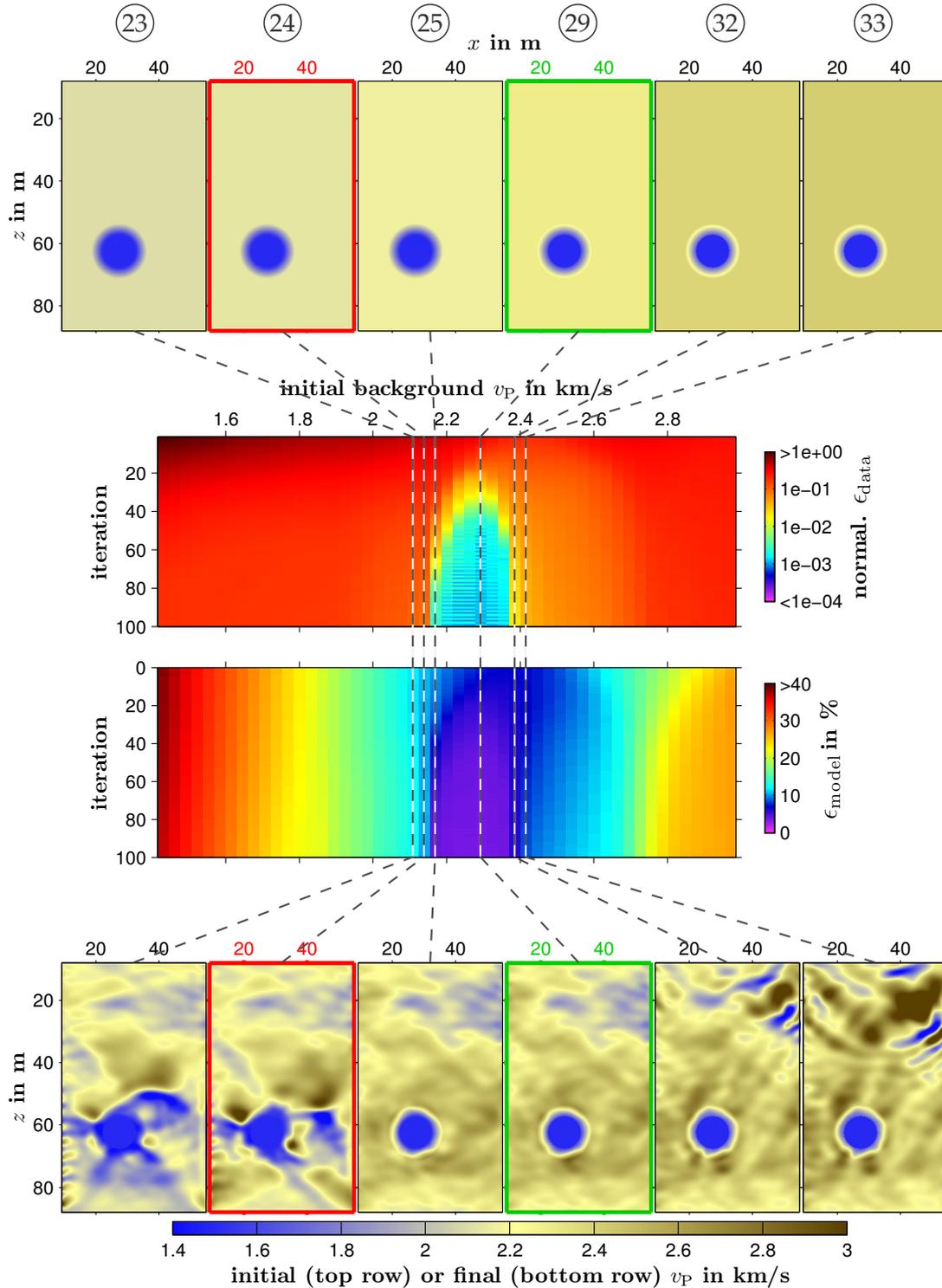
Experiment II: Data computations
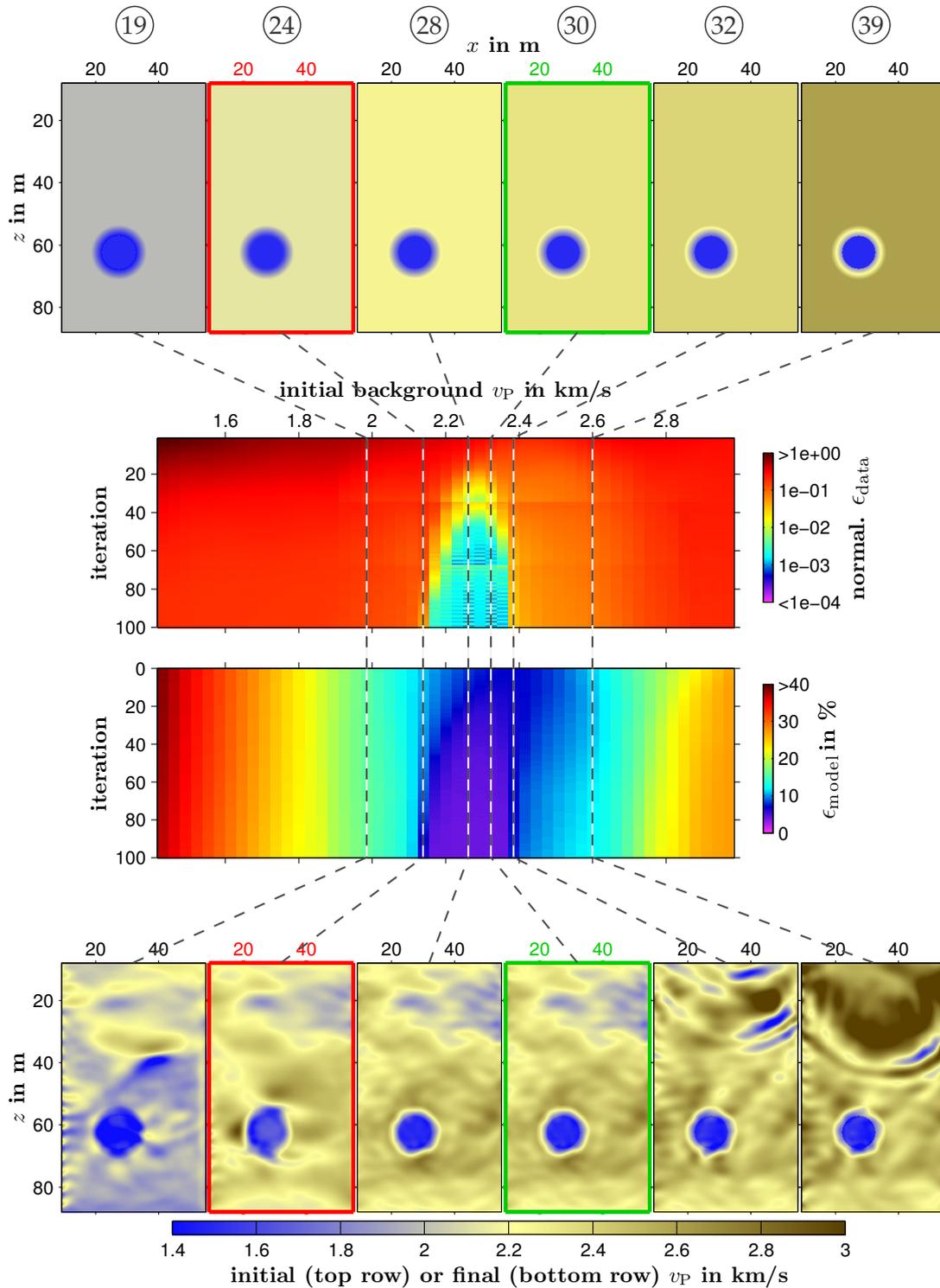
Figure D.2: Application 1, experiment II/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of offset windowing. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (0.98, 9.2 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 4.0)$ %.
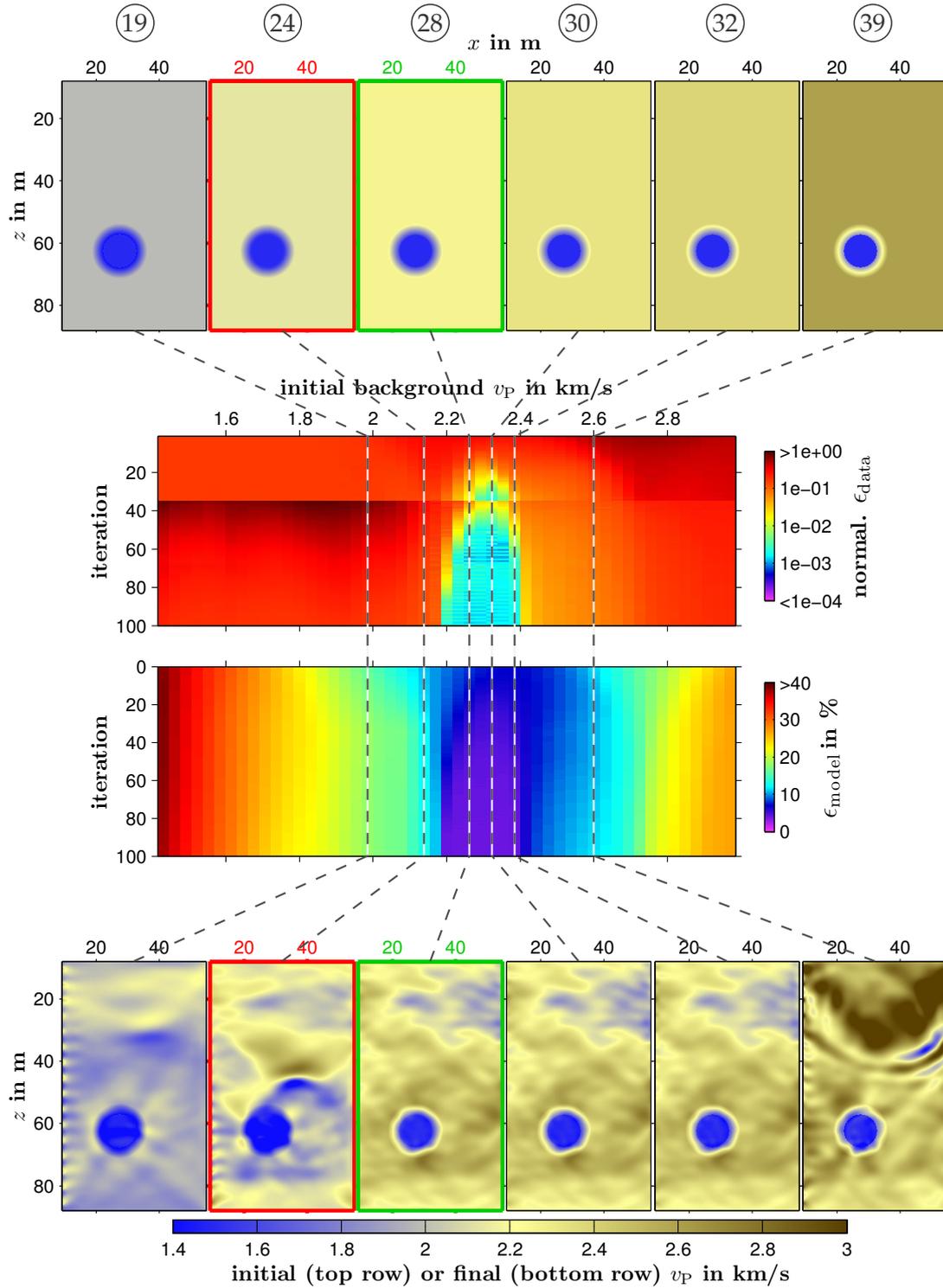
Figure D.3: Application 1, experiment II/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of time windowing. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\mathrm{data}},\ \min \epsilon_{\mathrm{data}}) = (0.89,\ 1.1 \cdot 10^{-3})$ and $(\max \epsilon_{\mathrm{model}},\ \min \epsilon_{\mathrm{model}}) = (38,\ 4.0)$ %.
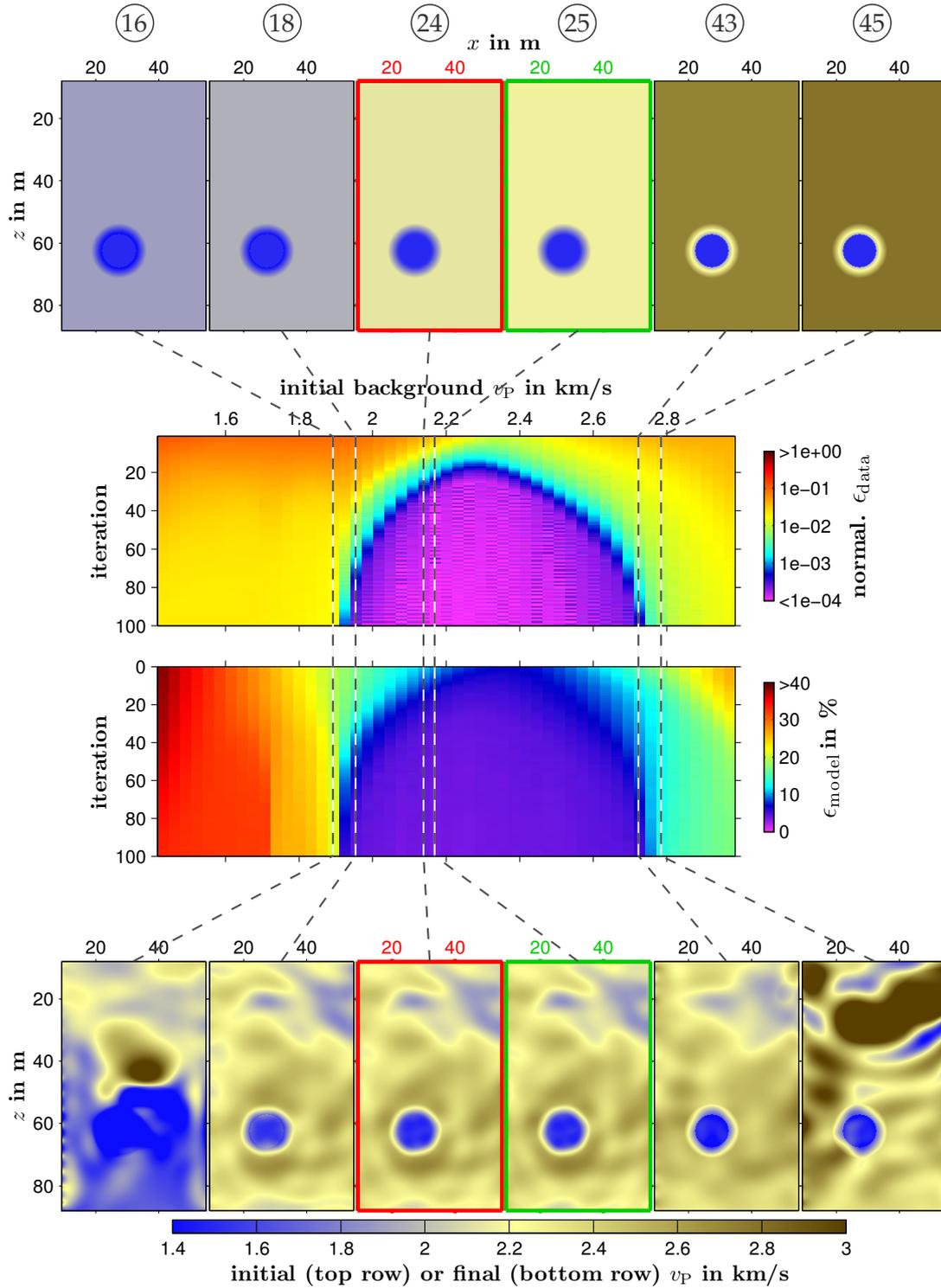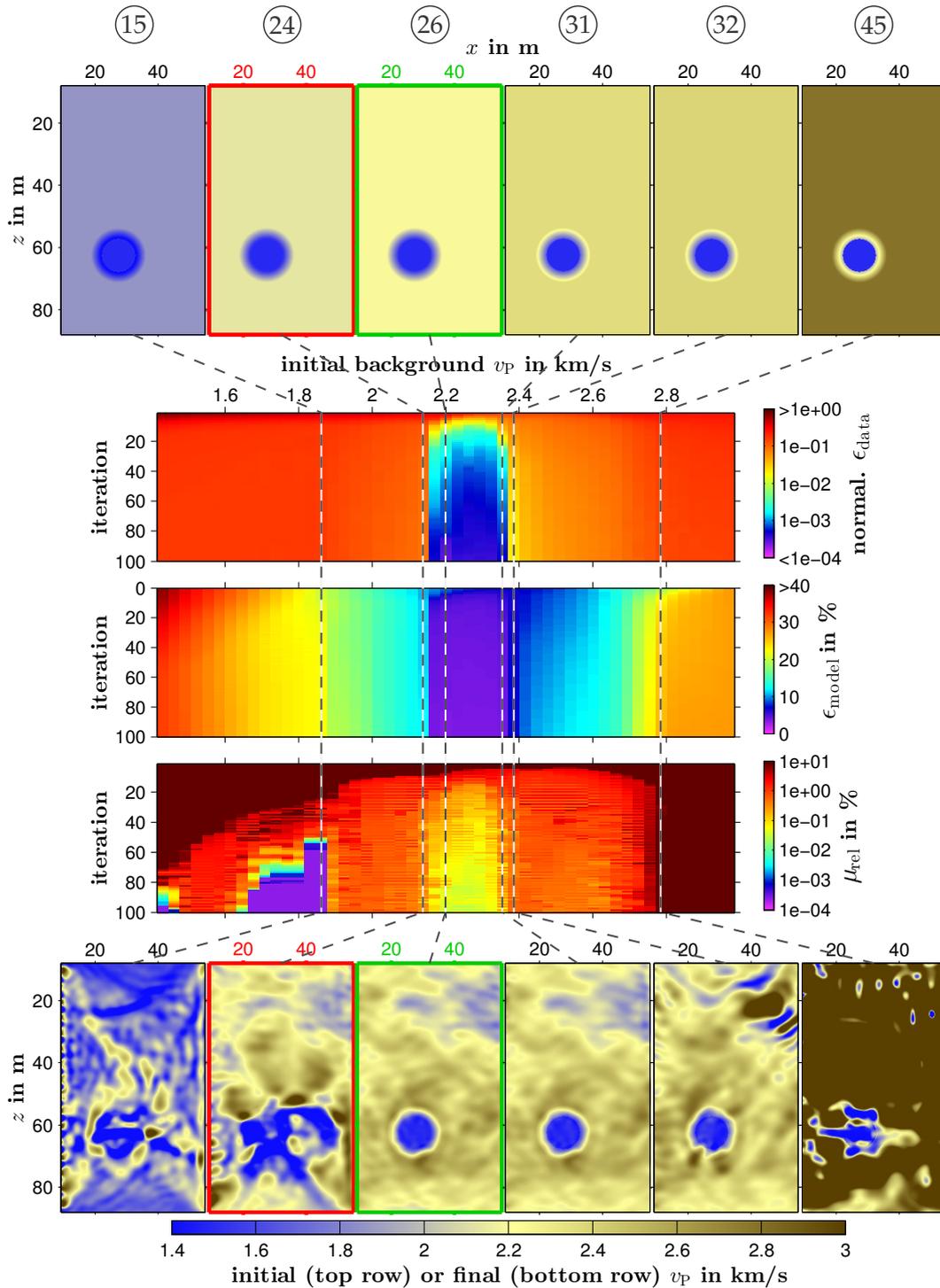
Figure D.4: Application 1, experiment II/2: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of single-stage frequency filtering ($f_{peak} = 95$ Hz). The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 3.7)$ %. The data misfit is clipped at $\cdot 10^{-4}$, whereas the minimum value is $8.1 \cdot 10^{-5}$.
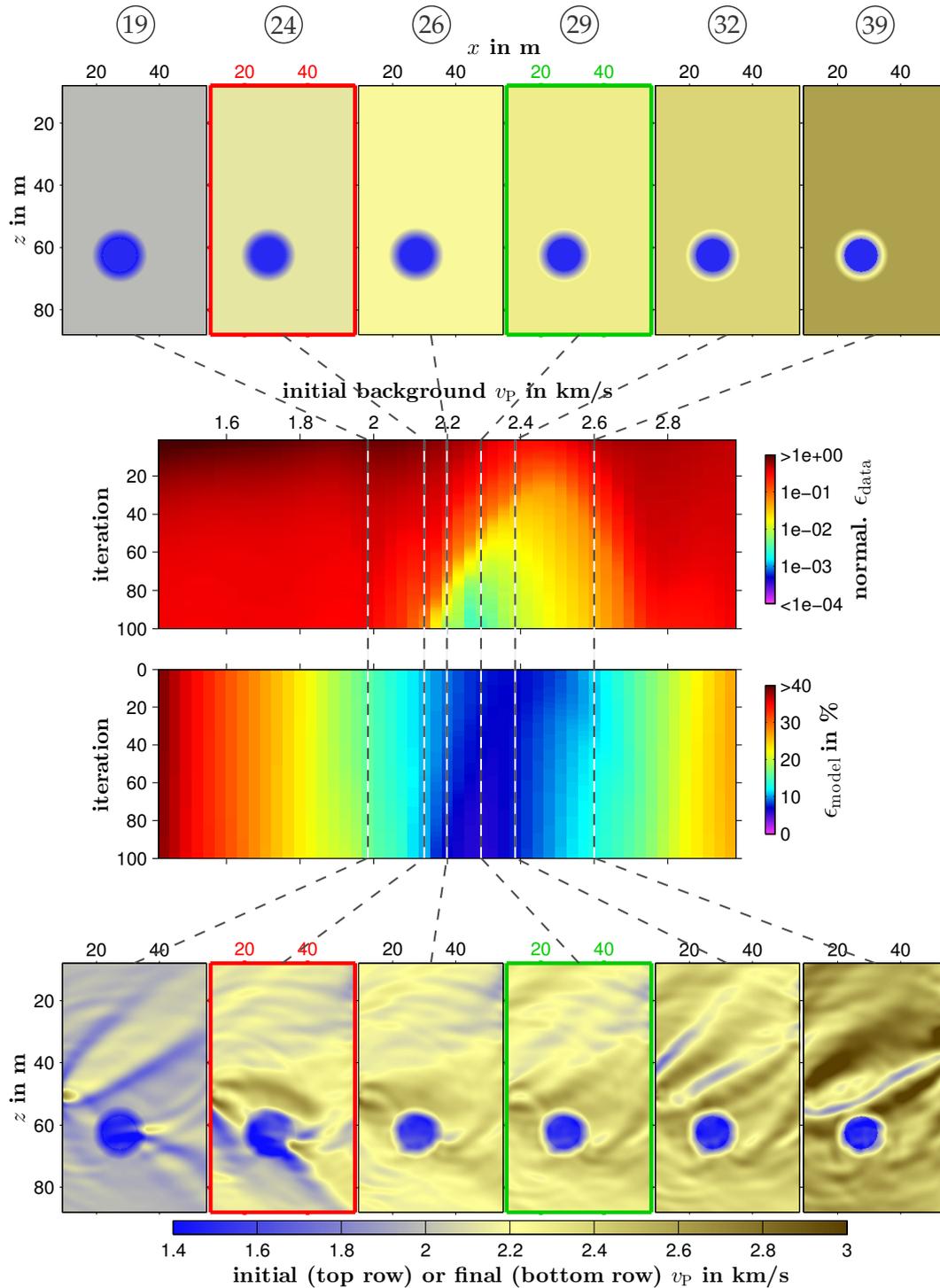
Figure D.5: Application 1, experiment II/3: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of multi-stage frequency filtering ($f_{peak} = (18, 95, 200)$ Hz). The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 3.5)$ %.
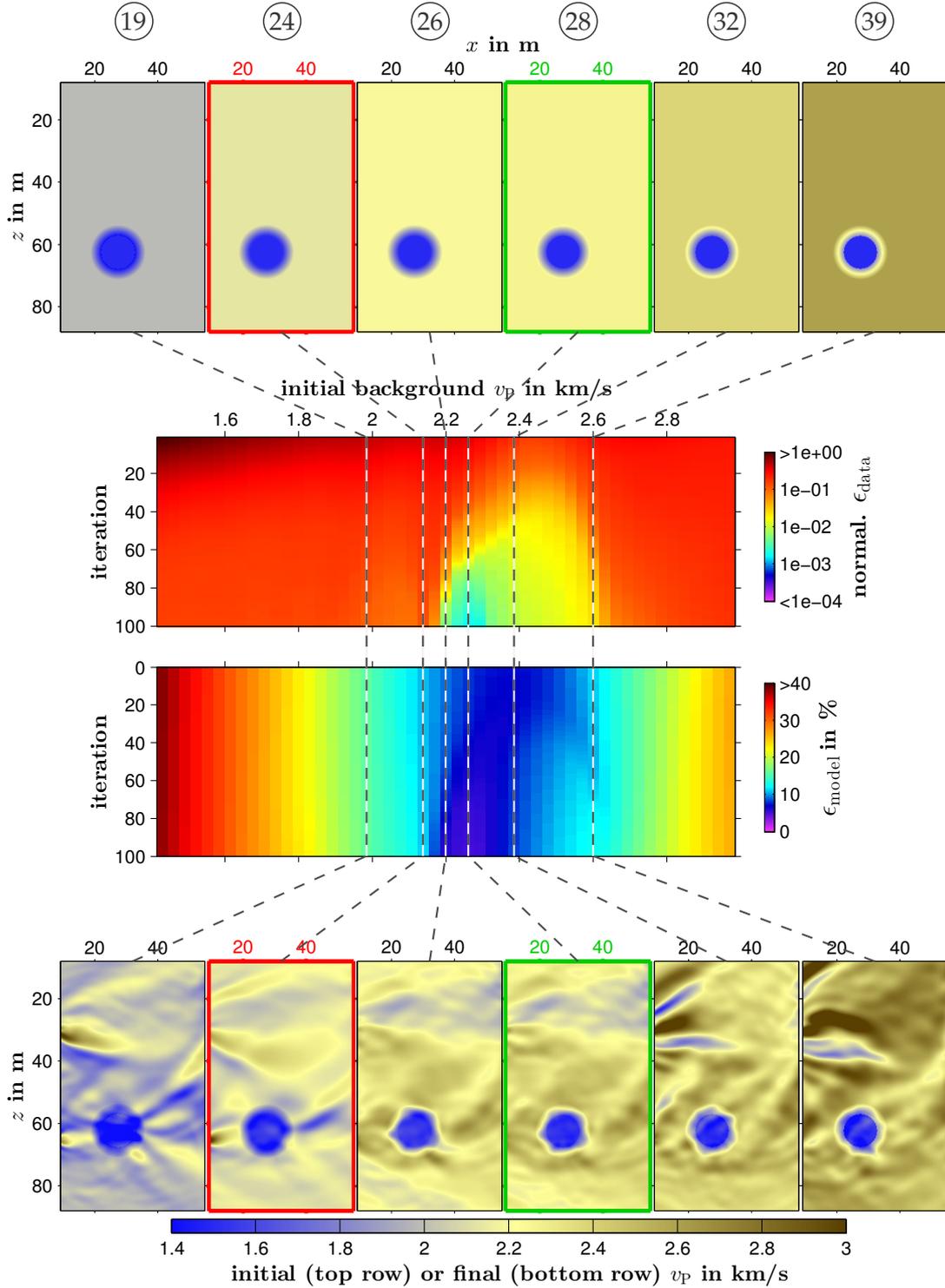
Experiment III: Gradient preconditioning

Figure D.6: Application 1, experiment III/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of gradient preconditioning (circular tapers around sources and receivers). The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \ \min \epsilon_{\text{data}}) = (1.0, \ 5.7 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \ \min \epsilon_{\text{model}}) = (38, \ 3.8)$ %.

Figure D.7: Application 1, experiment III/2: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of gradient preconditioning (wavefield-based taper). The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_\mathrm{data}, \min \epsilon_\mathrm{data}) = (1.0, 6.9 \cdot 10^{-4})$ and $(\max \epsilon_\mathrm{model}, \min \epsilon_\mathrm{model}) = (38, 3.7)$ %.
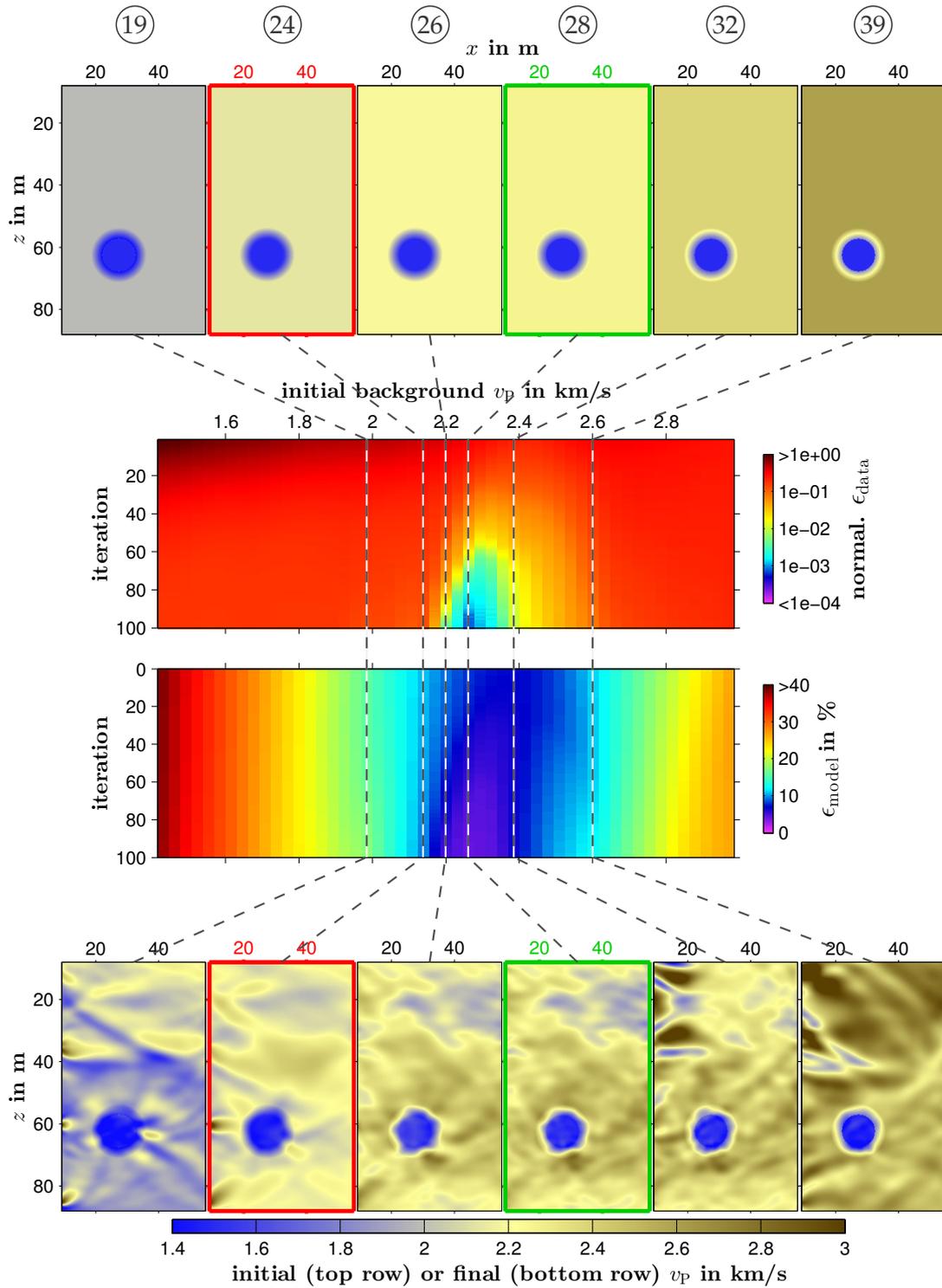
Figure D.8: Application 1, experiment III/3: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of gradient preconditioning (user-defined taper within the water-filled anomalous area). The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{data}, \min \epsilon_{data}) = (1.0, 1.0 \cdot 10^{-3})$ and $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 3.7)$ %.
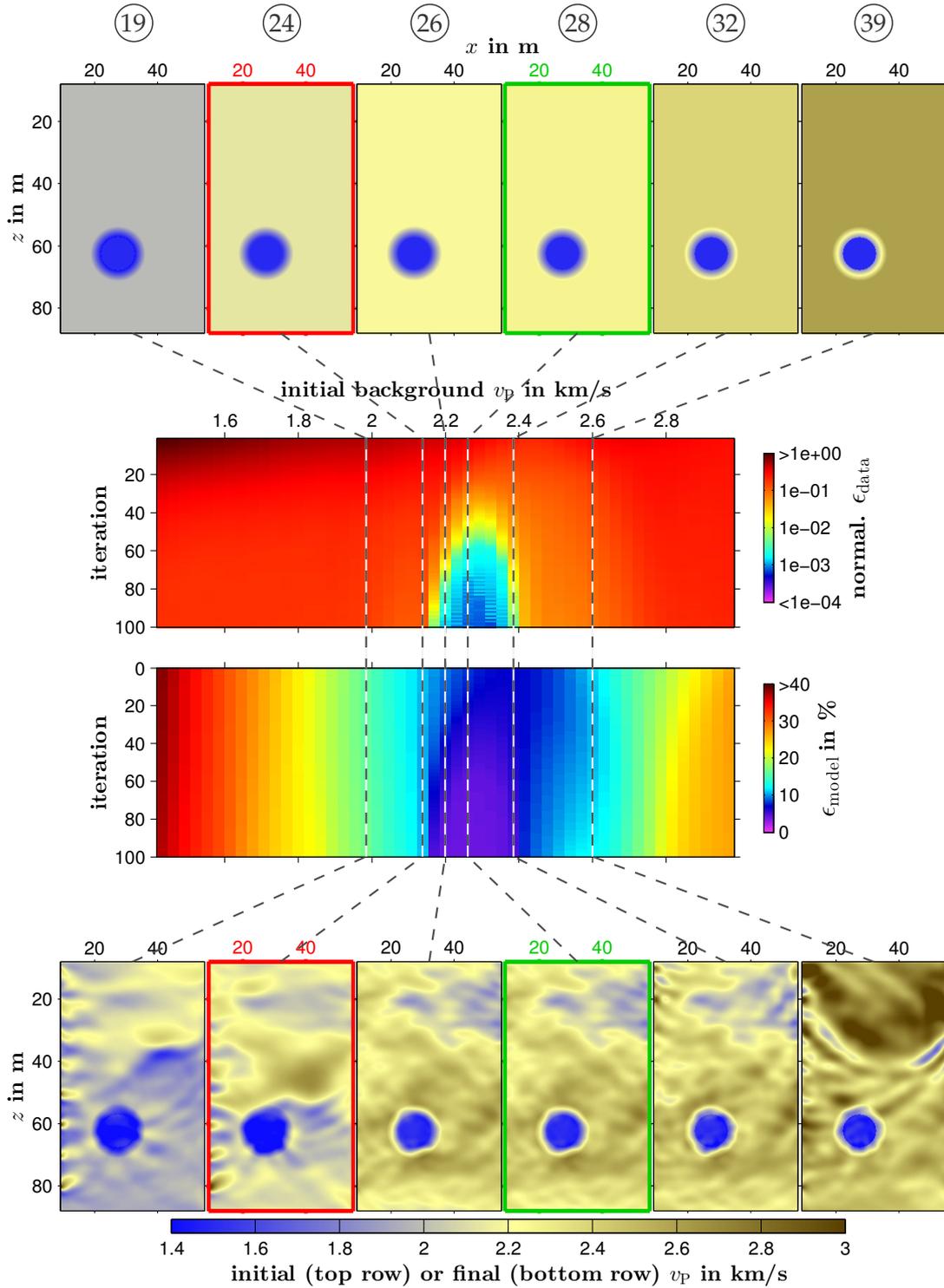
Experiment IV: Step length estimation

Figure D.9: Application 1, experiment IV/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of the adaptive step length method. In addition to data misfit, model error, the evolution of the relative step length $\mu_{rel|h}$ is visualized. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{data}, \min \epsilon_{data}) = (1.0, 2.8 \cdot 10^{-4})$ and $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 3.4)$ %.

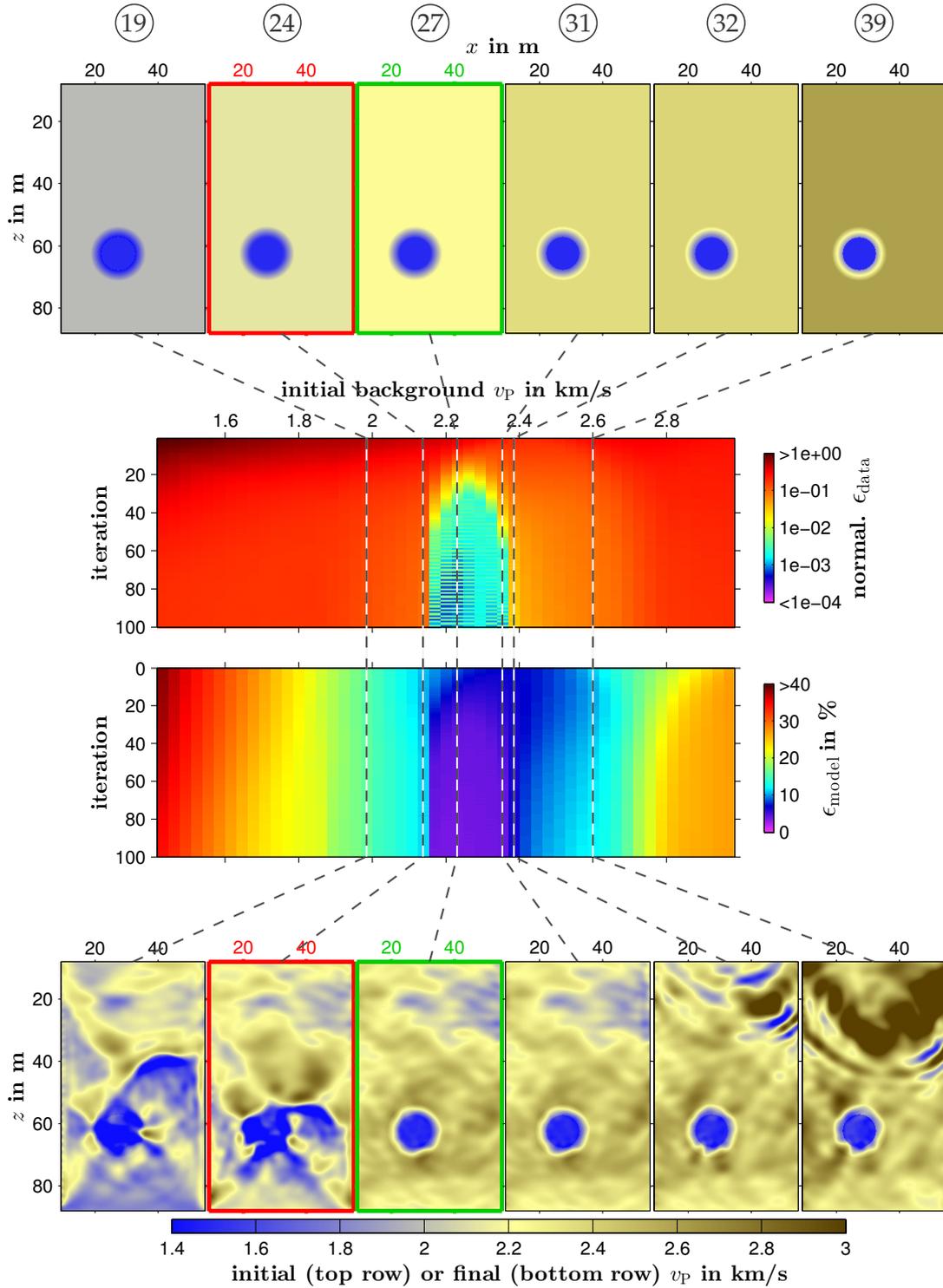## Experiment V: Influence of the acquisition geometry

Figure D.10: Application 1, experiment V/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of 1 source. The data misfit is normalized to the maximum misfit value of experiment I. The auxiliary corrective factor $\frac{N_s=1}{N_s=16}$ is applied to the misfit to account for different numbers of sources, N_s. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.2, 3.0 \cdot 10^{-3})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 6.0)$ %.
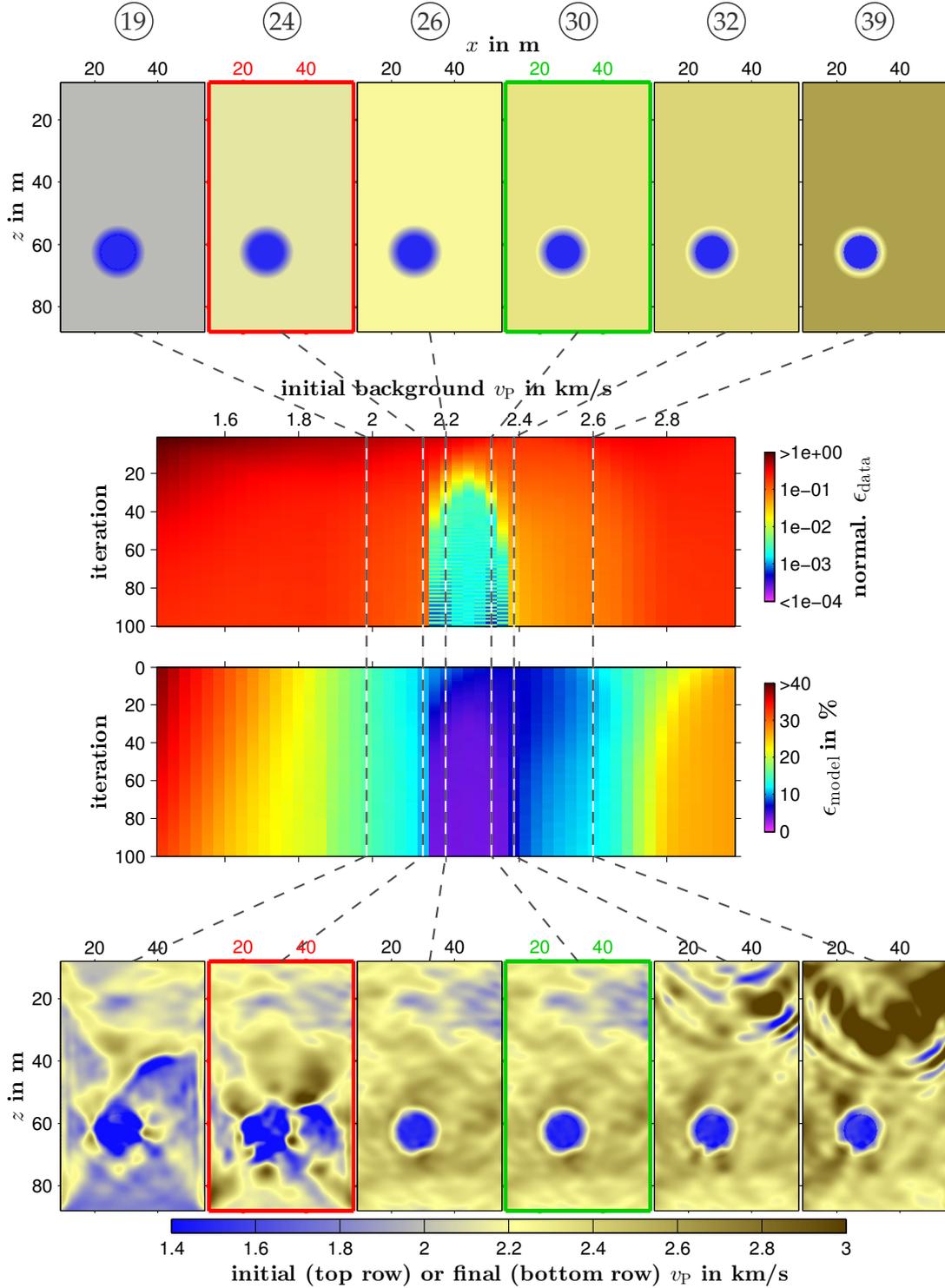
Figure D.11: Application 1, experiment V/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of 2 sources. The data misfit is normalized to the maximum misfit value of experiment I. The auxiliary corrective factor $\frac{N_s=2}{N_s=16}$ is applied to the misfit to account for different numbers of sources, N_s. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 1.7 \cdot 10^{-3})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 5.2)$ %.

Figure D.12: Application 1, experiment V/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of 4 sources. The data misfit is normalized to the maximum misfit value of experiment I. The auxiliary corrective factor $\frac{N_s=4}{N_s=16}$ is applied to the misfit to account for different numbers of sources, N_s. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}},\, \min \epsilon_{\text{data}}) = (0.98,\, 7.3 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}},\, \min \epsilon_{\text{model}}) = (38,\, 4.2)$ %.
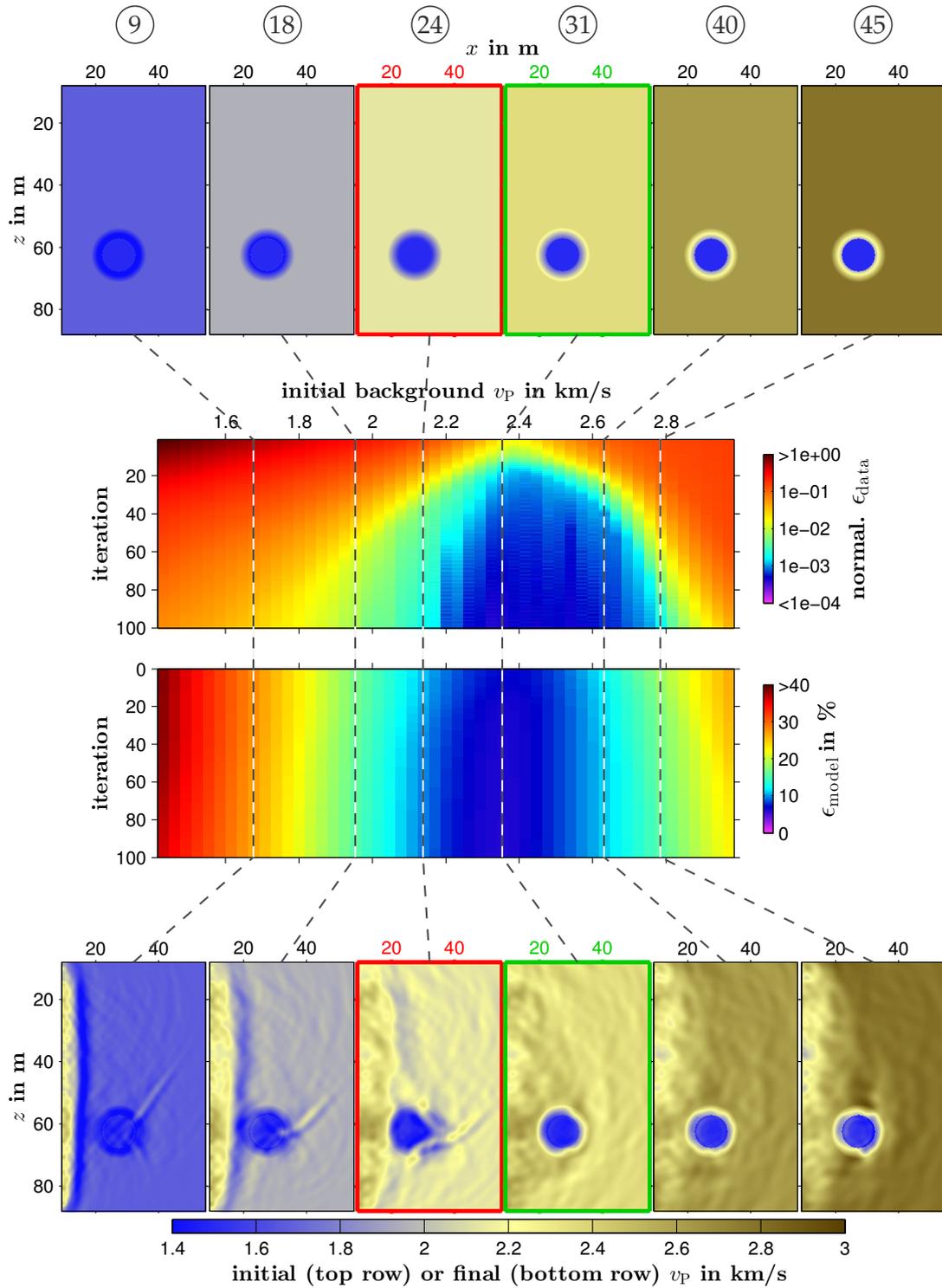
Figure D.13: Application 1, experiment V/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of 8 source. The data misfit is normalized to the maximum misfit value of experiment I. The auxiliary corrective factor $\frac{N_s=8}{N_s=16}$ is applied to the misfit to account for different numbers of sources, N_s. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 6.4 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 4.0)$ %.
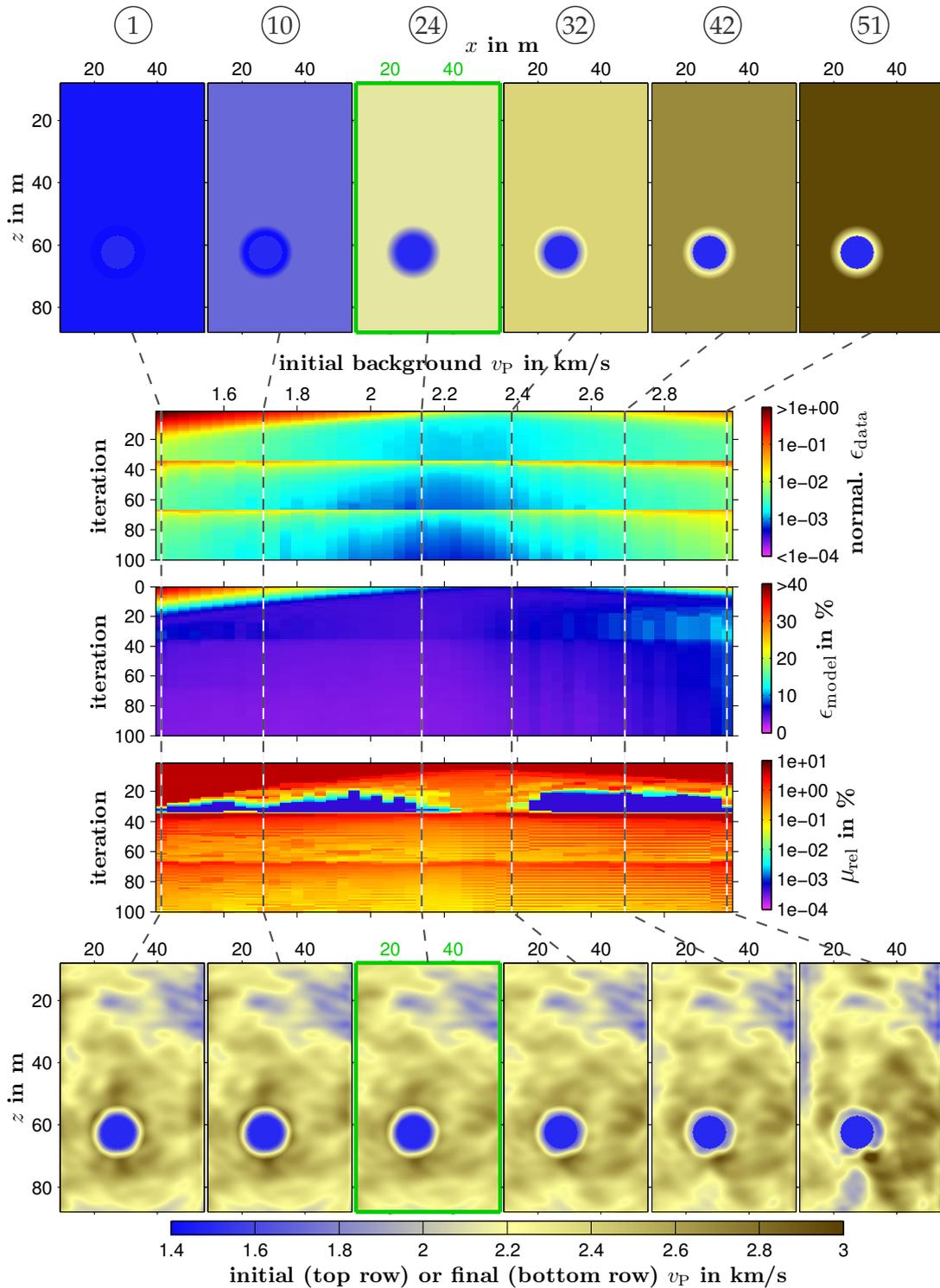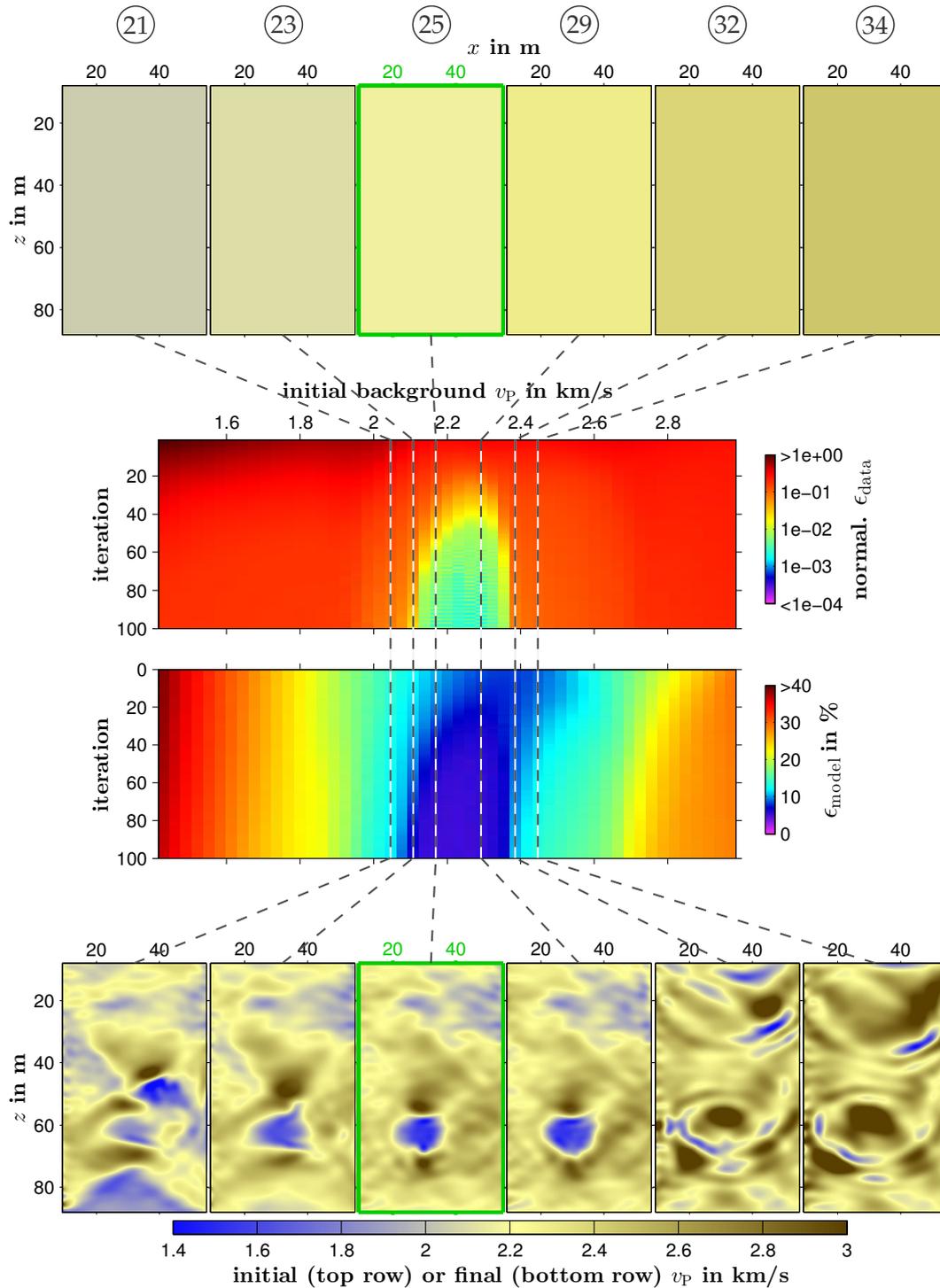
Figure D.14: Application 1, experiment V/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of 32 sources. The data misfit is normalized to the maximum misfit value of experiment I. The auxiliary corrective factor $\frac{N_s=32}{N_s=16}$ is applied to the misfit to account for different numbers of sources, N_s. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 6.9 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 3.7)$ %.
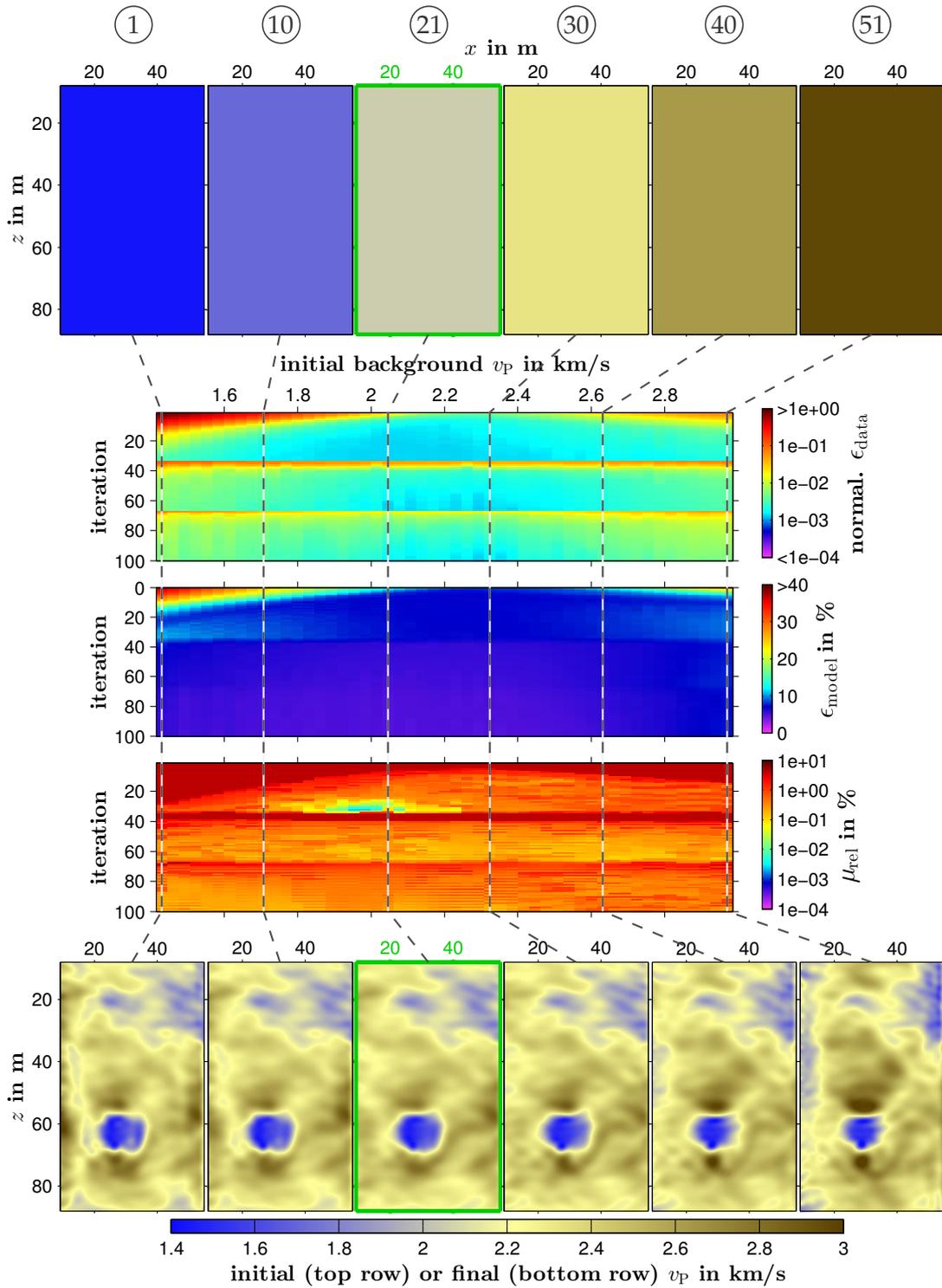
Figure D.15: Application 1, experiment V/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of 64 sources. The data misfit is normalized to the maximum misfit value of experiment I. The auxiliary corrective factor $\frac{N_\mathrm{s}=64}{N_\mathrm{s}=16}$ is applied to the misfit to account for different numbers of sources, N_s. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_\mathrm{data}, \min \epsilon_\mathrm{data}) = (1.0, 6.1 \cdot 10^{-4})$ and $(\max \epsilon_\mathrm{model}, \min \epsilon_\mathrm{model}) = (38, 3.8)$ %.

Figure D.16: Application 1, experiment V/2: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of a reflection geometry. The data misfit is normalized to the maximum misfit value of this multiple FWT. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 3.9 \cdot 10^{-4})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 6.2)$ %.
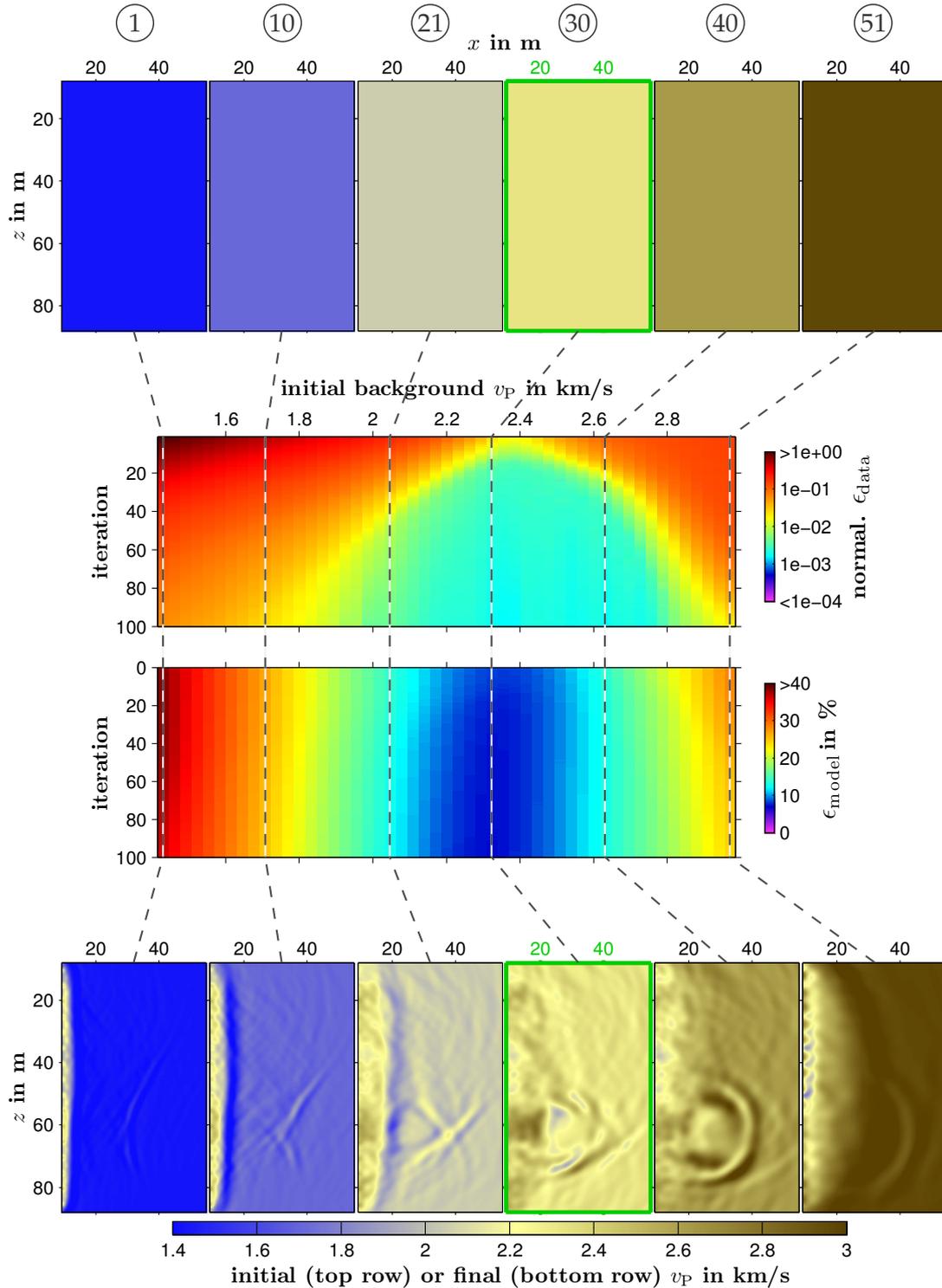
Experiment VI: Optimal parameter configuration

Figure D.17: Application 1, experiment VI/2: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of an optimal parameter combination and an acquisition geometry including 64 sources. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{\mathrm{model}}, \min \epsilon_{\mathrm{model}}) = (38, 3.2)$ %.

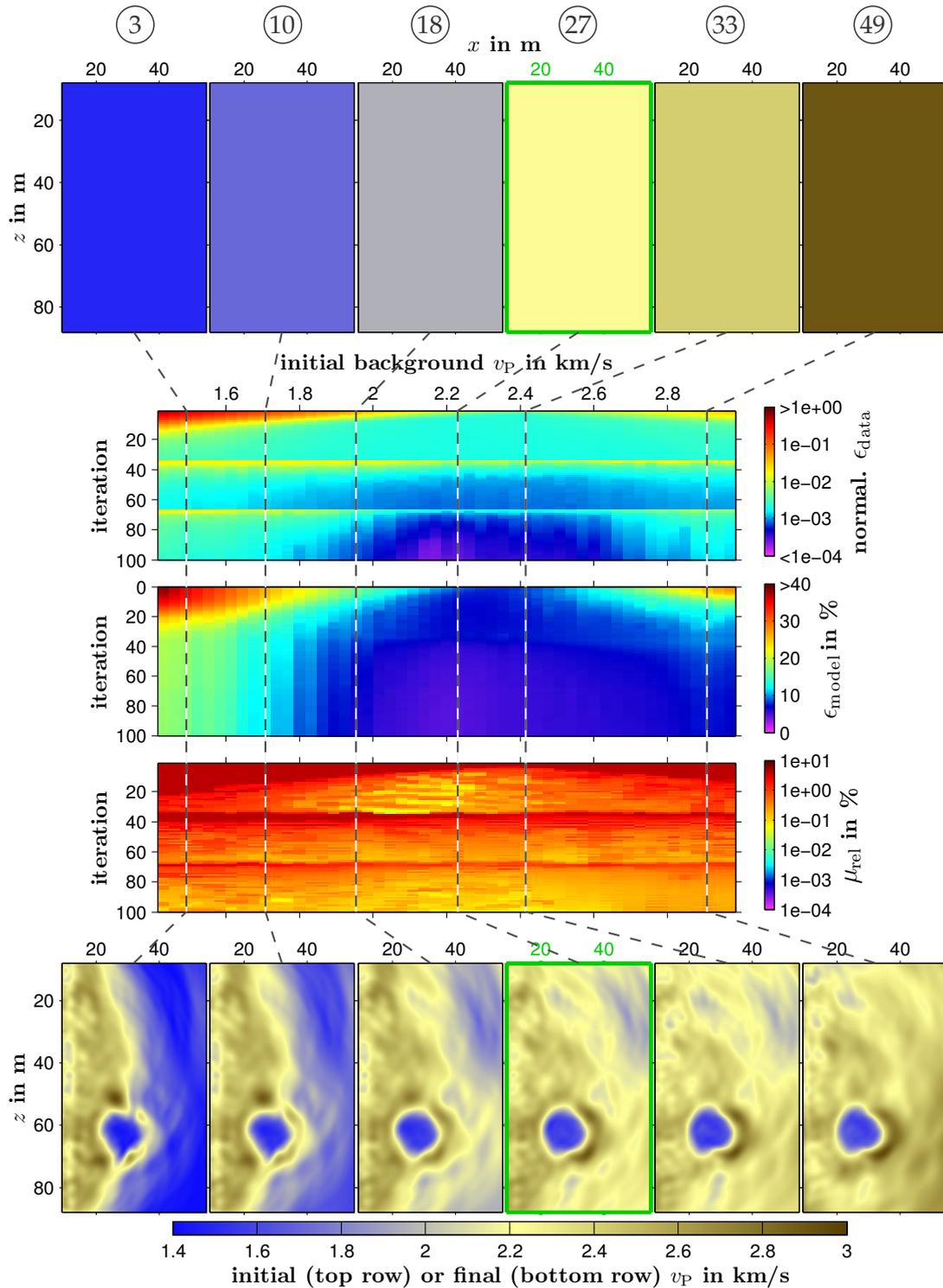**Experiment VII: Recapitulation for a different initial model**

Figure D.18: Application 1, experiment VII/1: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of a basic FWT without any methodical improvements. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\text{data}}, \min \epsilon_{\text{data}}) = (1.0, 2.0 \cdot 10^{-3})$ and $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 5.5)$ %.
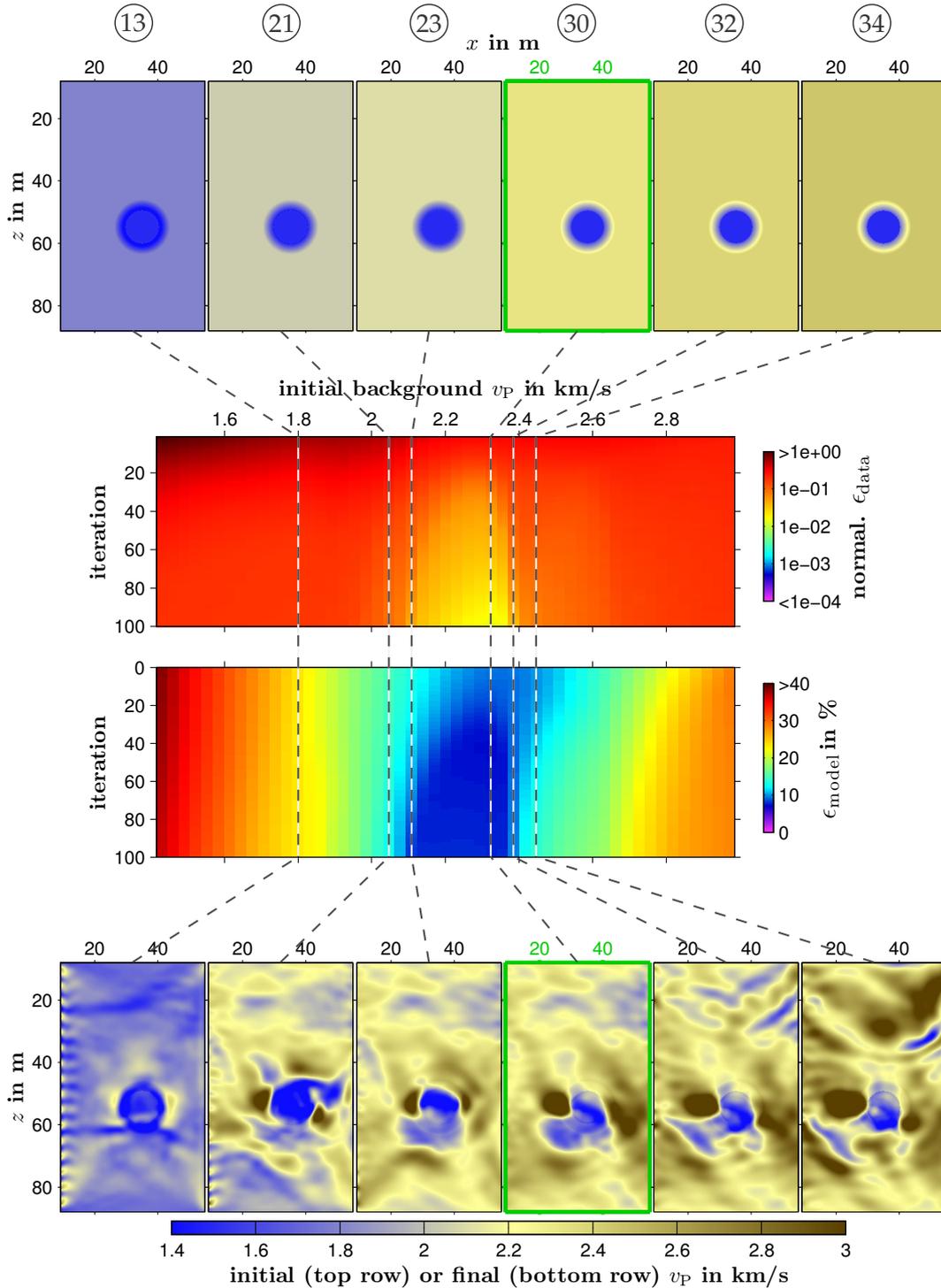
Figure D.19: Application 1, experiment VII/2: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of an optimal parameter combination. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 4.7)$ %.
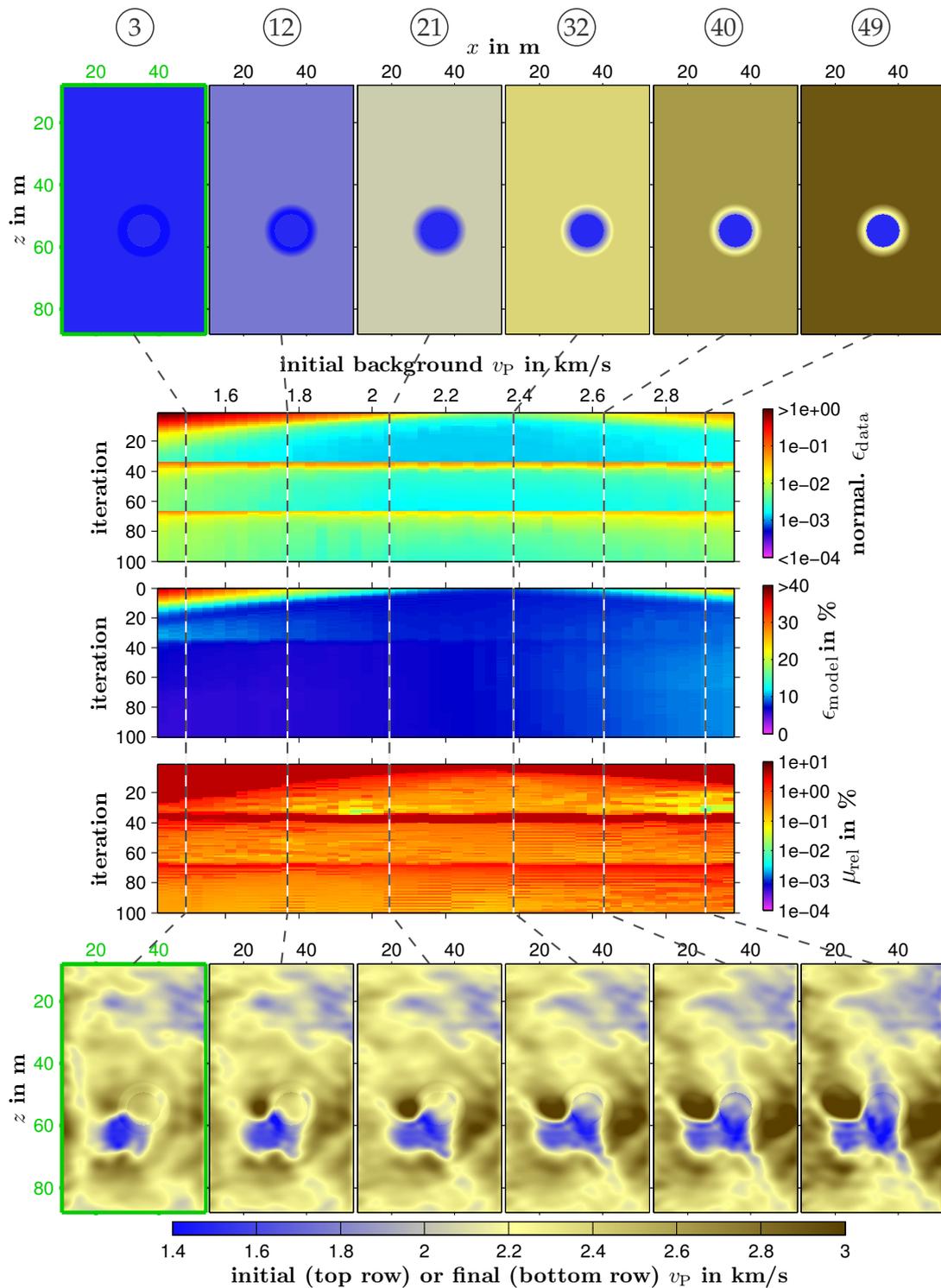
Figure D.20: Application 1, experiment VII/3: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of a basic FWT and a reflection geometry. The data misfit is normalized to the maximum misfit value of experiment V/2. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{data}, \min \epsilon_{data}) = (1.0, 1.6 \cdot 10^{-3})$ and $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 7.3)$ %.

Figure D.21: Application 1, experiment VII/4: selected initial models (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of an optimal parameter combination and a reflection geometry. The data misfit is normalized to the maximum misfit value of experiment V/2. The extreme values of the model error (lower plot) are: $(\max \epsilon_{\text{model}}, \min \epsilon_{\text{model}}) = (38, 4.6)$ %.
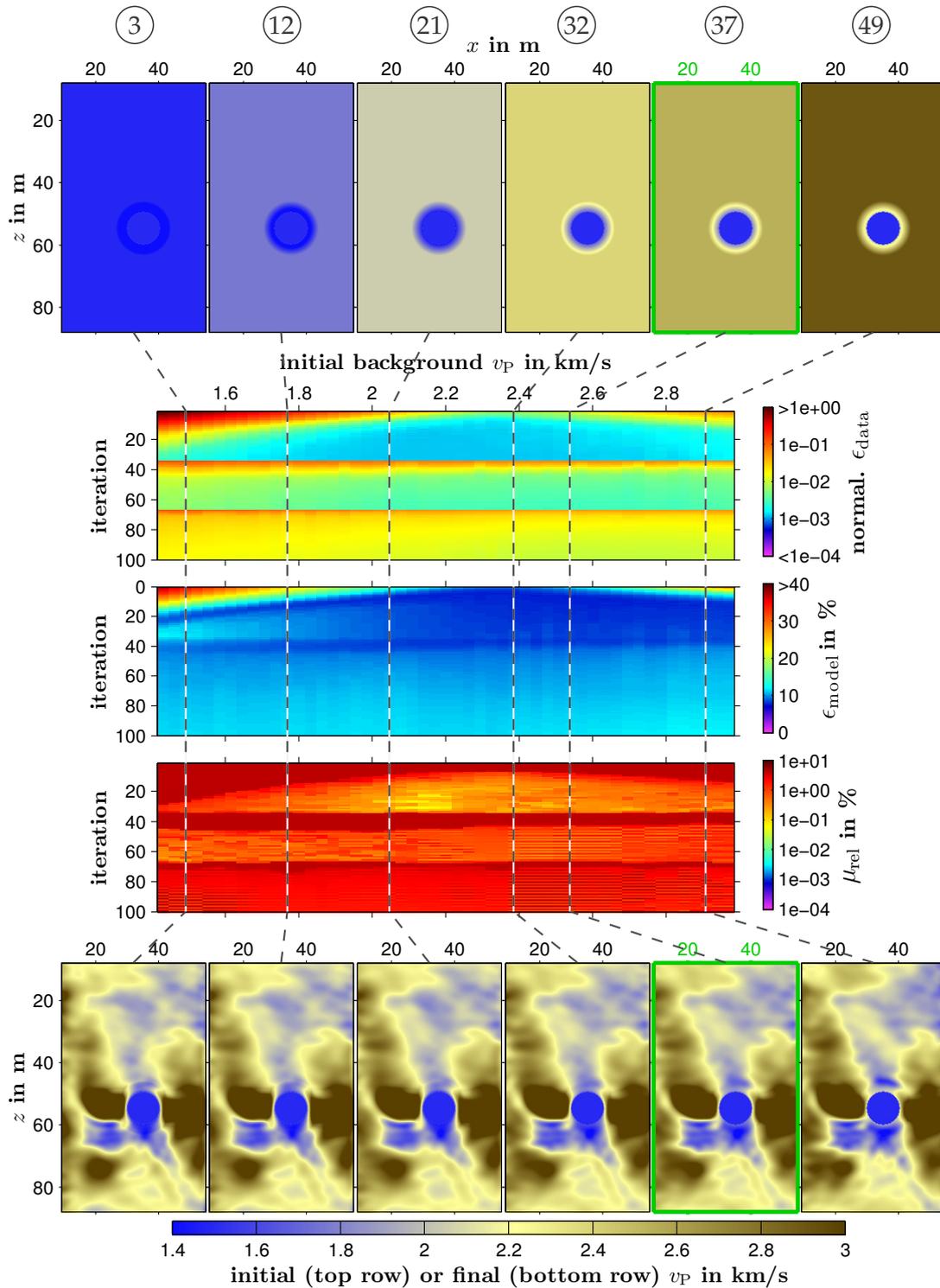
Experiment VIII: Initial model with wrong assumptions

Figure D.22: Application 1, experiment VIII/1: selected initial models of type C (top) and v_{P} results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of a basic FWT without any methodical improvements. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of normalized data misfit (upper plot) and model error (lower plot) are: $(\max \epsilon_{\mathrm{data}}, \min \epsilon_{\mathrm{data}}) = (1.0, 1.4 \cdot 10^{-2})$ and $(\max \epsilon_{\mathrm{model}}, \min \epsilon_{\mathrm{model}}) = (38, 7.2)$ %.

Figure D.23: Application 1, experiment VIII/2: selected initial models of type C (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of an optimal parameter combination. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{model}, \min \epsilon_{model}) = (38, 5.6)$ %.

Experiment IX: Initial model and FWT with wrong assumptions

Figure D.24: Application 1, experiment IX: selected initial models of type C (top) and v_P results (bottom) as well as the FWT progress for all 51 initial background models (central part) with application of an optimal parameter combination and a misplaced user-defined gradient preconditioning. The data misfit is normalized to the maximum misfit value of experiment I. The extreme values of the model error (lower plot) are: $(\max \epsilon_{\mathrm{model}}, \min \epsilon_{\mathrm{model}}) = (38, 7.6)$ %.

Experiment X: Brute-force search
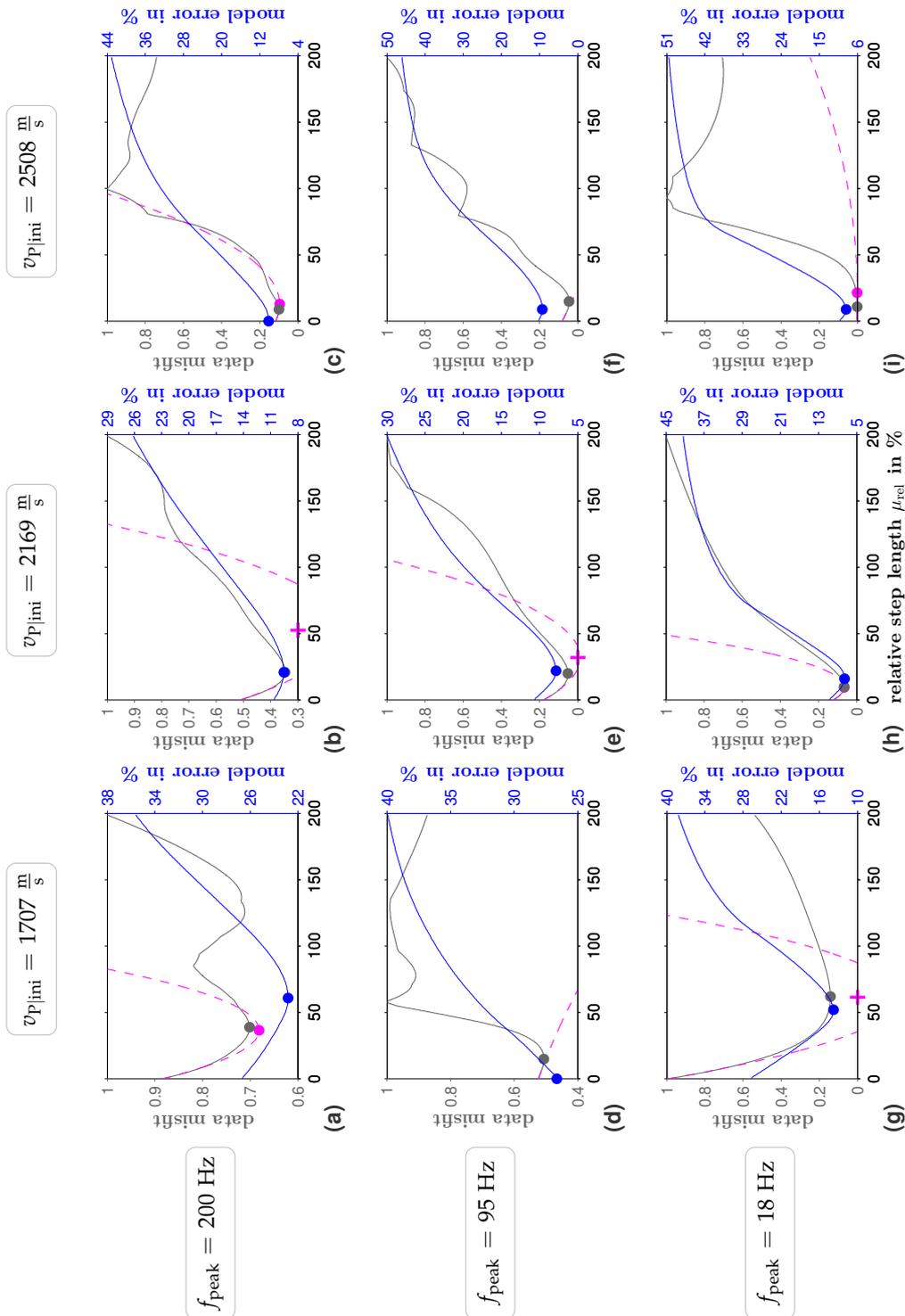
Figure D.25: Application 1, experiment X: Data misfits and model errors as functions of step lengths at the first iteration. The plots show cross sections from Figures 4.24 and 4.25. The rows correspond to the application of frequency filtering. The columns represent the cross sections for different initial models. The magenta graph illustrates the parabolic fit applied to the data misfit function and based on test step lengths $(0.125, 1, 8)$ %. Gray and blue "●" symbolize the minimum data misfit or model error. The magenta "●" or + indicate the optimum step length.

Appendix E

Application 3: Acoustic FWT in the presence of attenuation: a quantitative study

This appendix contains additional seismograms for the Marmousi experiment of FWT application 3 discussed in chapter 6.

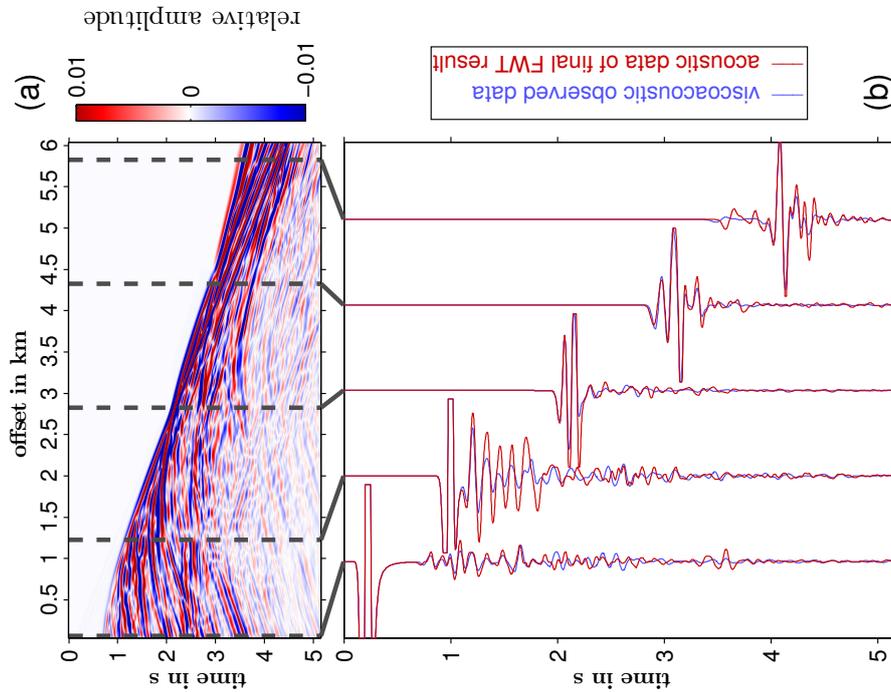# E.1   Synthetic experiment: Marmousi model
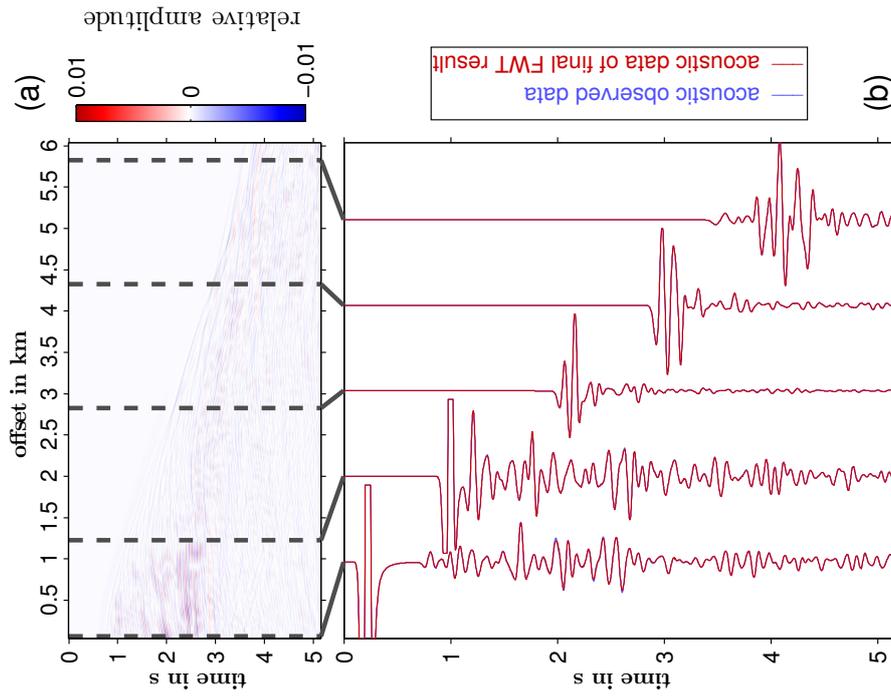
Figure E.1: Application 3, Marmousi model: Final data of reference FWT for a shot at $x = 2.6$ km. (a) shows the final data residuals. Amplitudes are clipped to ± 1 % of the maximum observed amplitude. (b) illustrates normalized traces of observed and final synthetic data belonging to highlighted offsets. Observed and synthetic waveforms can be compared.
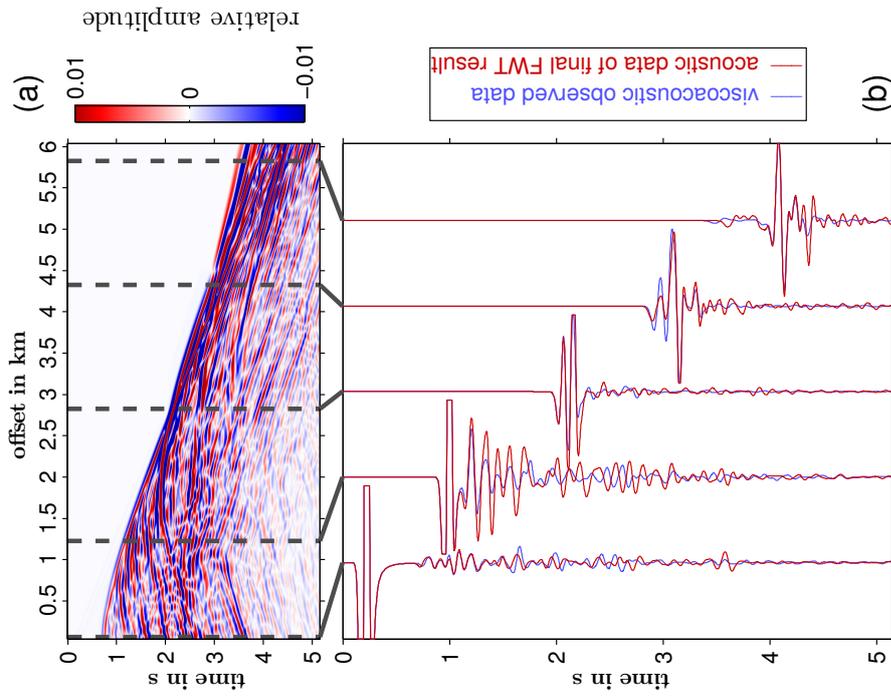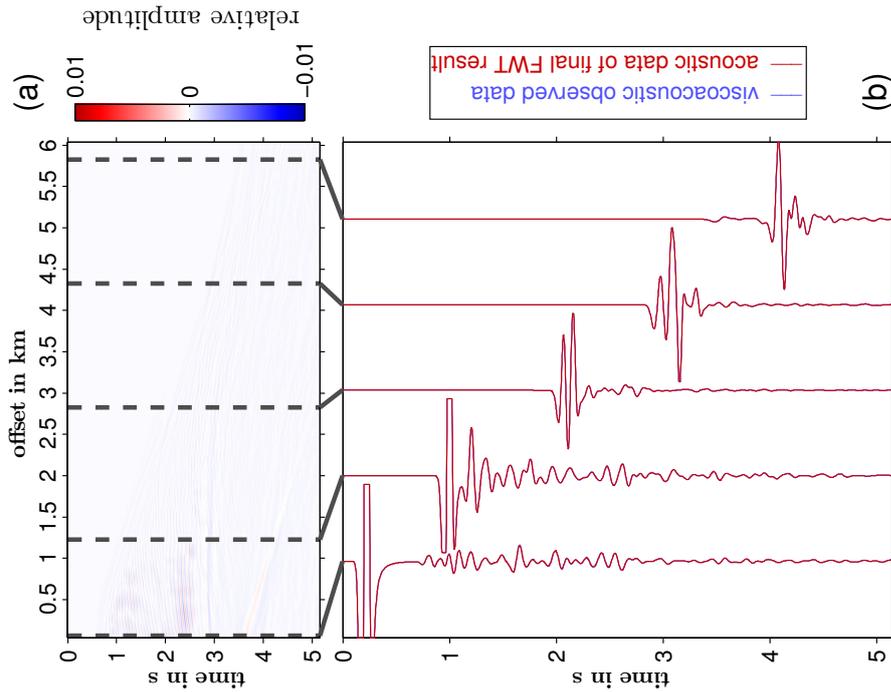

Figure E.2: Application 3, Marmousi model: Final data of test 1 for a shot at $x = 2.6$ km. (a) shows the final data residuals. Amplitudes are clipped to ± 1 % of the maximum observed amplitude. (b) illustrates normalized traces of observed and final synthetic data belonging to highlighted offsets. Observed and synthetic waveforms can be compared.

Figure E.3: Application 3, Marmousi model: Final data of test 2 for a shot at $x = 2.6$ km. (a) shows the final data residuals. Amplitudes are clipped to $\pm 1\%$ of the maximum observed amplitude. (b) illustrates normalized traces of observed and final synthetic data belonging to highlighted offsets. Observed and synthetic waveforms can be compared.
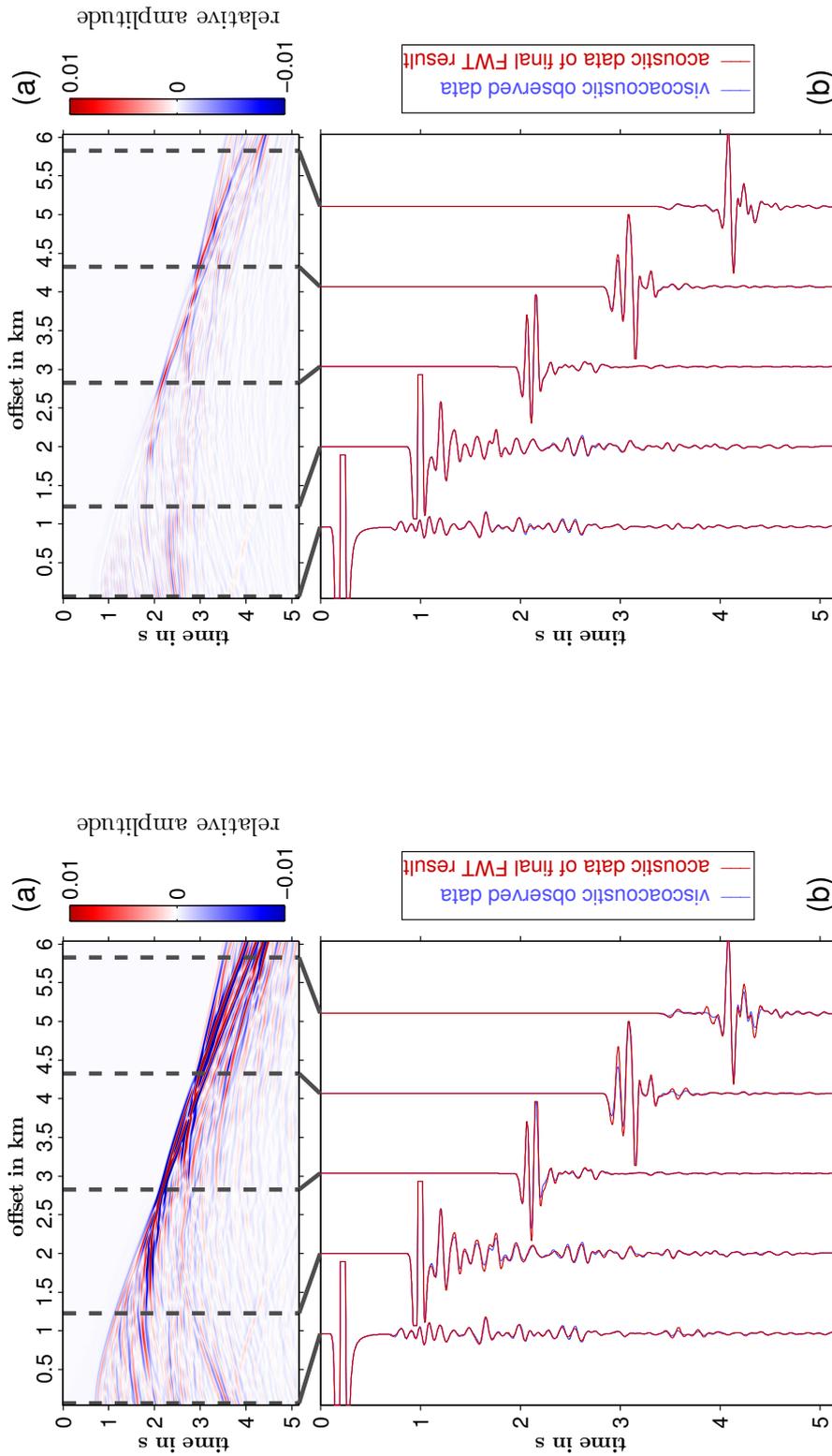

Figure E.4: Application 3, Marmousi model: Final data of test 3 for a shot at $x = 2.6$ km. (a) shows the final data residuals. Amplitudes are clipped to $\pm 1\%$ of the maximum observed amplitude. (b) illustrates normalized traces of observed and final synthetic data belonging to highlighted offsets. Observed and synthetic waveforms can be compared.

Figure E.5: Application 3, Marmousi model: Final data of test 4 for a shot at $x = 2.6$ km. (a) shows the final data residuals. Amplitudes are clipped to $\pm 1\%$ of the maximum observed amplitude. (b) illustrates normalized traces of observed and final synthetic data belonging to highlighted offsets. Observed and synthetic waveforms can be compared.

Figure E.6: Application 3, Marmousi model: Final data of test 5 for a shot at $x = 2.6$ km. (a) shows the final data residuals. Amplitudes are clipped to $\pm 1\%$ of the maximum observed amplitude. (b) illustrates normalized traces of observed and final synthetic data belonging to highlighted offsets. Observed and synthetic waveforms can be compared.

Appendix F

Application 4: 3D acoustic FWT in the time domain

This appendix contains additional figures for the applications of 3D acoustic FWT discussed in chapter 7. It summarizes cross sections of the true and inverted v_P models along all spatial directions with respect to the cross-well experiment in appendix F.1 and the Marmousi experiment in appendix F.2.
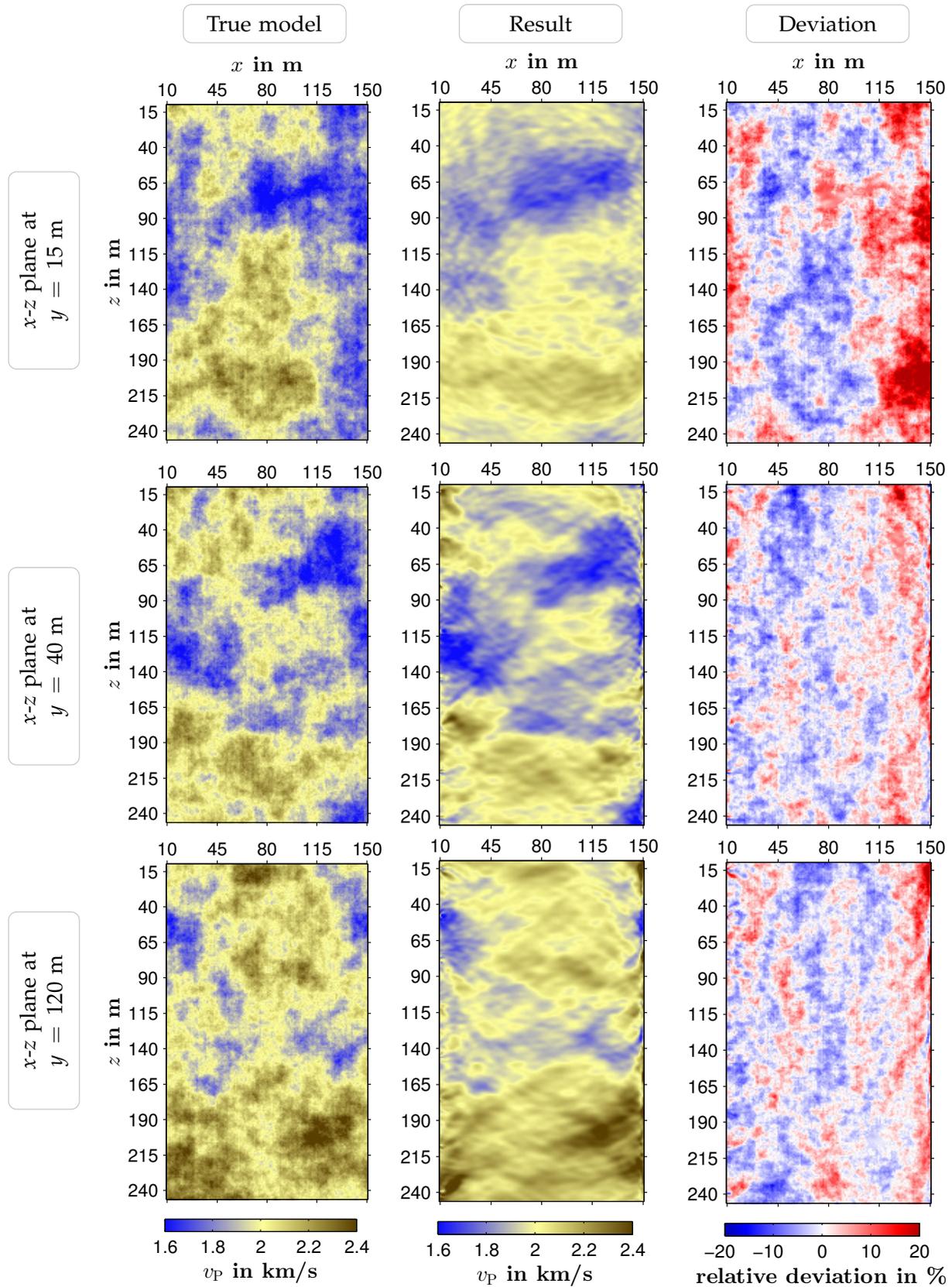
F.1 3D FWT for a cross-well geometry

Figure F.1: Application 4, Cross-well experiment: Vertical (*x-z*) cross sections of true and final v_P models as well as deviations from the true model for exemplary *y* positions.
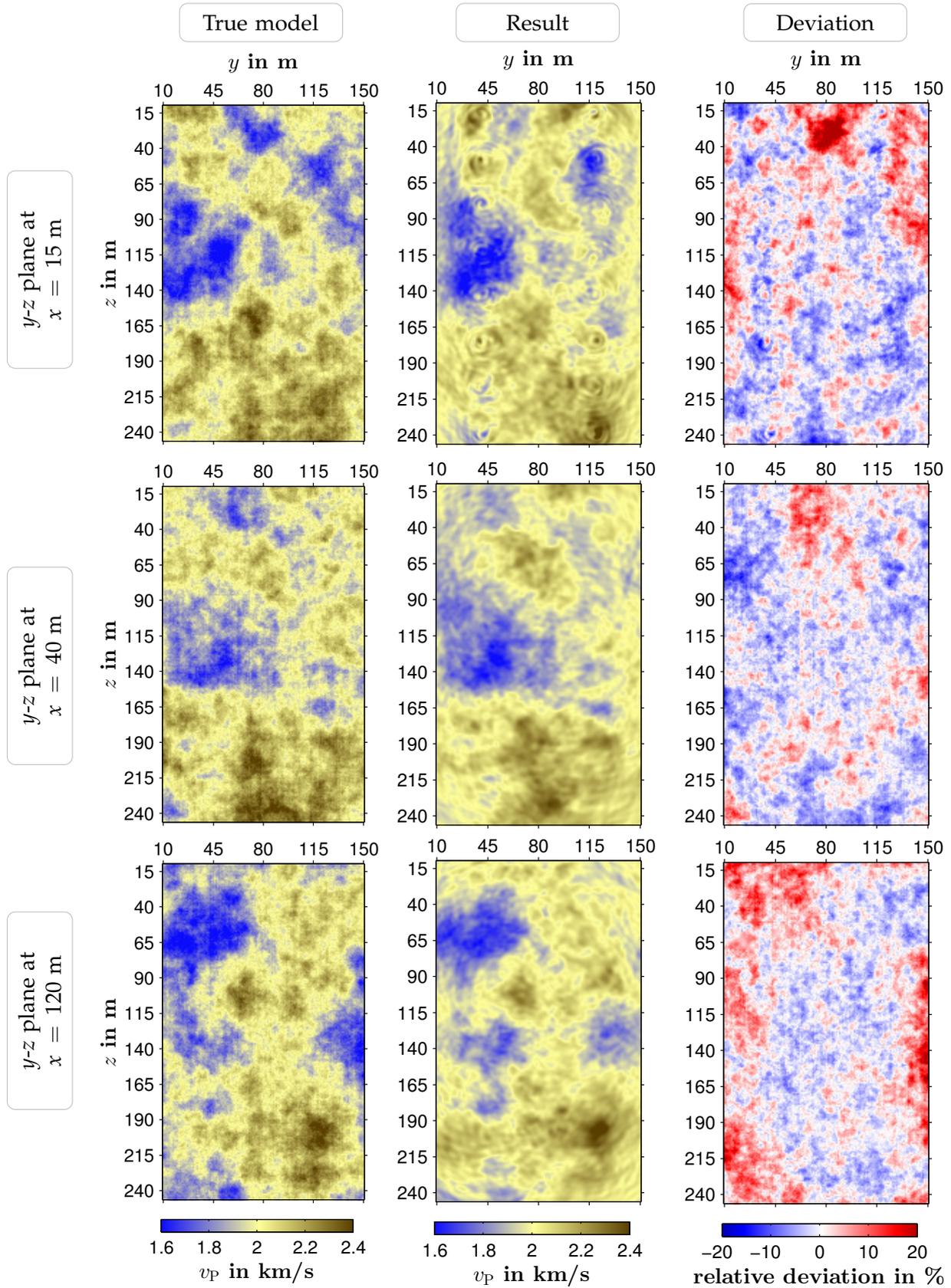
Figure F.2: Application 4, Cross-well experiment: Vertical (*y-z*) cross sections of true and final v_P models as well as deviations from the true model for exemplary *x* positions.
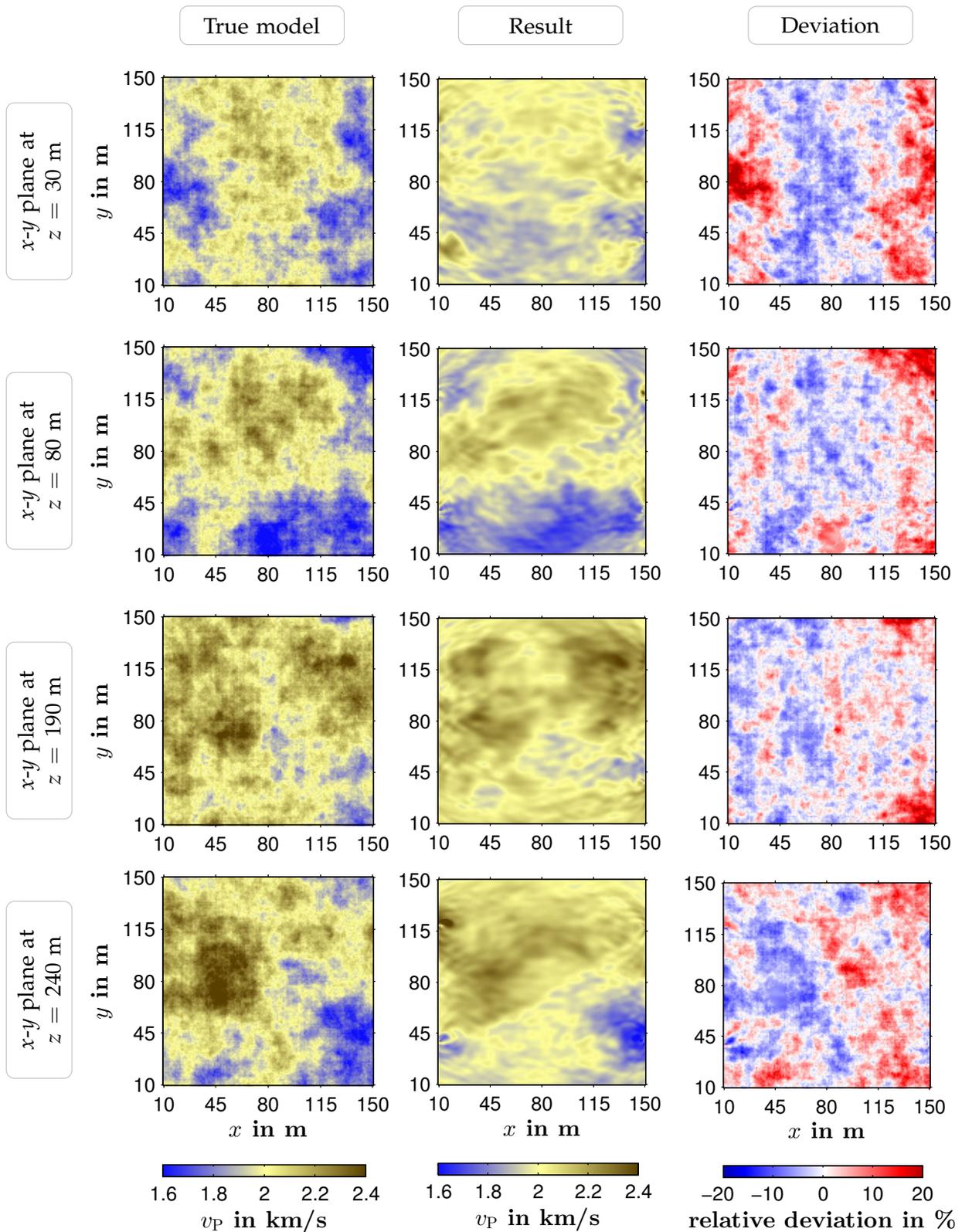
Figure F.3: Application 4, Cross-well experiment: Horizontal (*x-y*) cross sections of true and final v_P models as well as deviations from the true model for exemplary depths z.
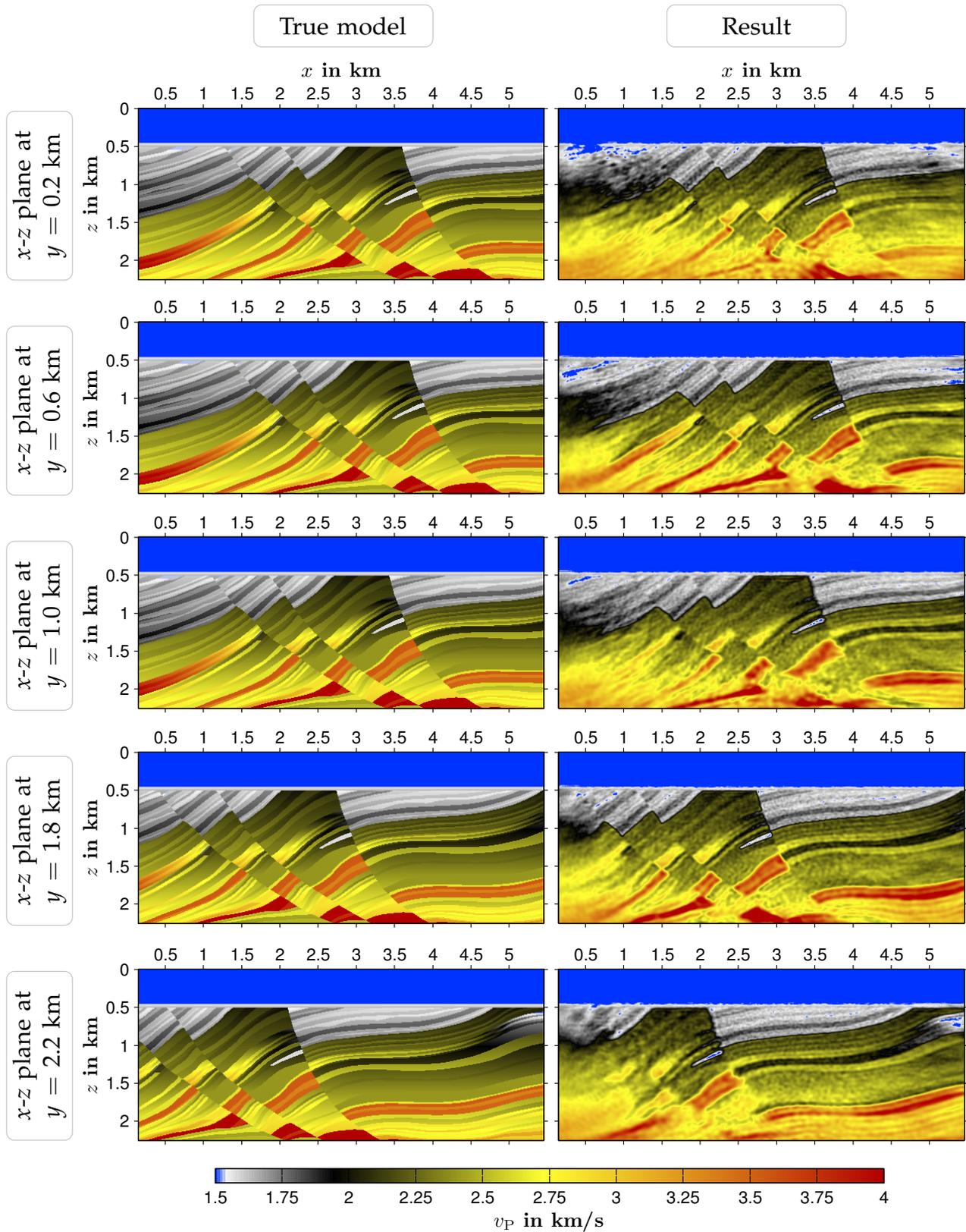
F.2   3D FWT for a reflection geometry

Figure F.4: Application 4, Marmousi experiment: Vertical (*x-z*) cross sections of true and final v_{P} models for exemplary *y* positions.
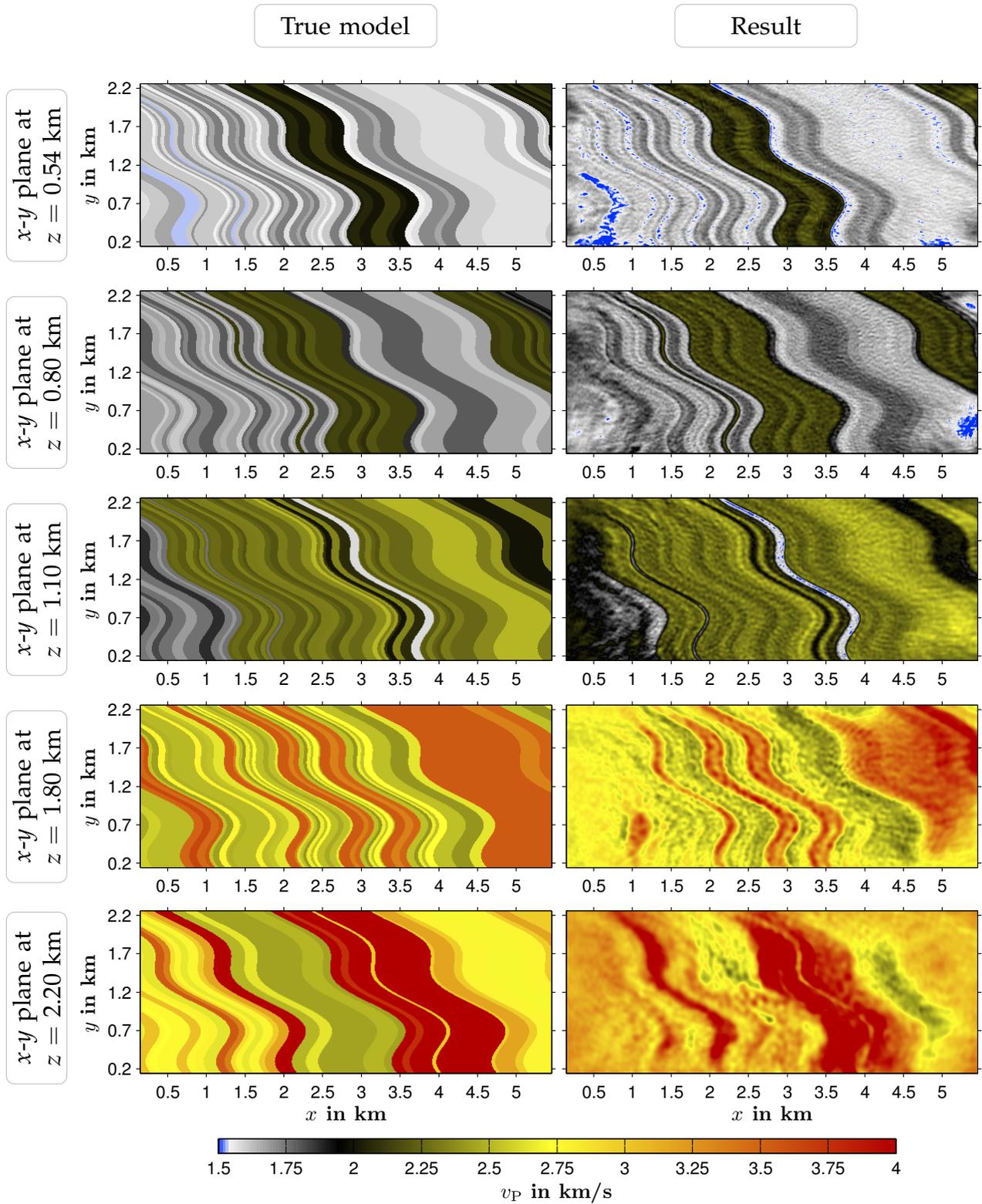
Figure F.5: Application 4, Marmousi experiment: Lateral (*x-y*) cross sections of true and final v_P models for exemplary depths *z*.
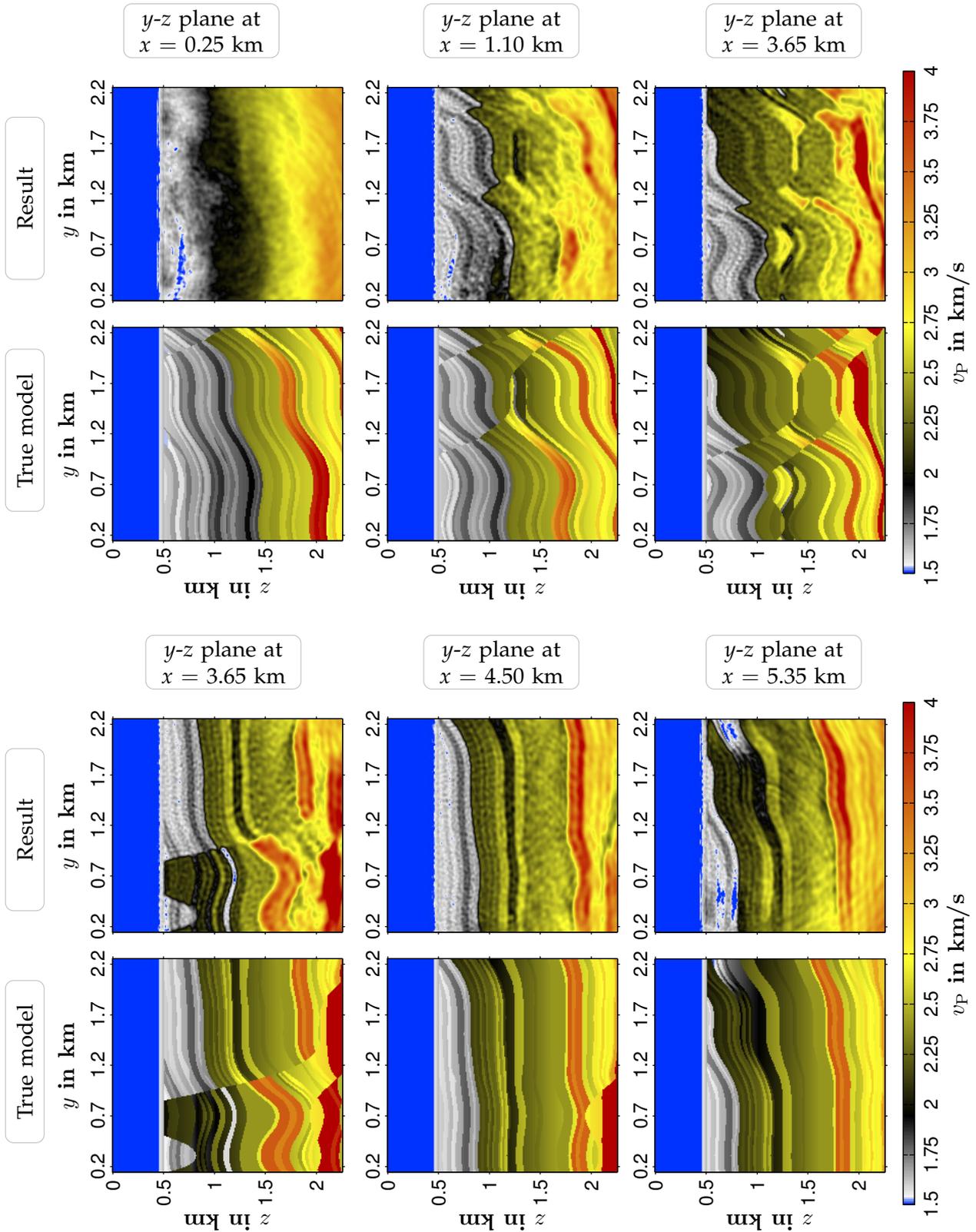
Figure F.6: Application 4, Marmousi experiment: Vertical (*y-z*) cross sections of true and final v_P models for exemplary *x* positions.

Appendix G

Software and hardware

Software

Within the scope of this work, I used several software packages to perform full-waveform inversions, to process the input or output of FWT implementations and to create documents, such as this thesis:

(A) FWT implementations:

- Most of 2D inversion results were computed using the 2D acoustic time-domain FWT implementation *PROTEUS* (written in *C*) which is the essential part of code development in this work. The FWT implementation mainly bases on the theory proposed by Tarantola (1984) and Mora (1987). In particular, the parallel implementation with respect to domain decomposition bases on the 2D finite-difference modeling implementation *FDVEPS* (Bohlen, 1998).

- For comparisons of 2D inversions in time domain and frequency domain, I used the 2D acoustic frequency-domain implementation *FULLWV* developed by R. G. Pratt (Pratt, 1999).

- The 3D inversion results were computed using the 3D acoustic time-domain FWT implementation *POSEIDON* which is developed on the basis of *PROTEUS*.

(B) Pre-processing and Post-processing

- I used *MathWorks MATLAB*® to perform additional scientific computations and to create figures. In particular, I used the *MATLAB*® script *AnalyticAnelastic.m* (implemented by Josep de la Puente, LMU Geophysics, Munich) to compute a semi-analytical reference solution.

- Especially the conversion of seismic file formats was realized by *Seismic Un*x*.

(C) Document generation

- Several tools – developed within the workgroup of applied geophysics at the Geophysical Institute – were deployed to create documents.

- This work is written in $\LaTeX\,2_\varepsilon$ within the operating-system environment *Linux*.

225

Hardware

Within the scope of this work, I employed several workstations or high-performance computers to obtain FWT results. The following list includes all hardware resources required for obtaining preliminary (test) results or the results shown in this thesis:

- workstation cluster at the *Institute of Geophysics, TU Bergakademie Freiberg*,

- cluster computer *CHIC* at *TU Chemnitz*,

- workstation cluster at the *Geophysical Institute, Karlsruhe Institute of Technology (KIT)*,

- high-performance computer *HP XC3000* at the *KIT*,

- *bwGRiD* cluster computers at Baden-Württemberg state universities in Karlsruhe, Stuttgart, Esslingen and Mannheim,

- high-performance supercomputer *JUROPA* at the *Jülich Supercomputing Centre*,

- high-performance supercomputer *HERMIT* at the *High Performance Computing Center Stuttgart*.

Danksagung

Zum pünktlichen Abschluß meiner Arbeit habe ich nun noch diesen Platz für einige dankende Worte reserviert.

Allen voran möchte ich meiner Familie danken, insbesondere meinen Eltern, die mich stets unterstützt haben und sich regelmäßig besorgt nach dem Fortschritt der Arbeit erkundigten. Dies gilt ebenso für meine Großeltern, denen ich den baldigen Doktor versprochen hatte, die aber leider nicht mehr daran teilhaben können.

Auch erinnere ich mich gern an die Zeit und Weggefährten in Freiberg. Das gilt besonders für die schöne WG-Zeit und ihre lustigen Diss-Abende – abseits von inversen Problemen – mit Stefan W., Jörg, Julian und Olaf. Vor allem möchte ich mich bei den ehemaligen Kolleginnen und Kollegen am Institut für Geophysik der TU Freiberg bedanken, die mich als Lehrer geophysikalisch großzogen und mit denen ich sehr gern zusammenarbeitete. Hervorheben möchte ich Holli, Maja, Tino und Holger. Stets stand Holli hilfsbereit zur Seite und erleichterte mir bei der allmorgendlichen Kaffeerunde mit Klatsch und Tratsch den Start in den Arbeitstag. Maja, Tino und Holger bin ich für die moralische Unterstützung in den regelmäßigen und sehr konstruktiven Dienstberatungen dankbar.

Vor allem möchte ich mich aber auch bei meinen Karlsruher Freunden, Kolleginnen und Kollegen bedanken, die mich über den Großteil meiner Promotionszeit begleiteten und mir in der Schlußphase der Doktorarbeit durch fleißiges Lesen geholfen bzw. mich durch den undurchsichtigen Promotions-Dschungel gelotst hatten – danke Anna, Ines, Lisa, Simone und Martin. Das gilt insbesondere auch für Anja's und Sven's "Kantine", wo ich meine allmorgendliche Kaffeetradition wiederbelebt habe. Vielen Dank auch an Claudia, die sich stets mit viel Einsatz um das lästige Drumherum gekümmert hatte.

Um nun den Schwenk zum Fachlichen zu vollziehen, möchte ich Thomas Bohlen danken, der mein großes Interesse an inversen Problemen erkannte. Danke für die Bereitstellung des Promotionsthemas und die fachliche Betreuung. Auch möchte ich ihm für die vielen inspirierenden Eingebungen danken, die mich immer wieder auf neue Ideen brachten. An dieser Stelle möchte ich auch der VNG Leipzig und dem WIT-Konsortium danken, die weitgehend meine Arbeit finanzierten. Insbesondere danke ich Daniel für die vielen und fruchtbaren fachlichen Diskussionen sowie die zahlreichen Hilfestellungen, die mir den Start ins Promotionsthema erleichterten. Abschließend danke ich meinem Korreferenten Jan van der Kruk für die Begutachtung meiner Doktorarbeit und konstruktiver fachlicher Diskussion.

Besonderer Dank gilt aber Stefan J. als verlässlichen Freund und Kollegen, der mich mit Beharrlichkeit und Regelmäßigkeit angestoßen hat, wenn ich mal wieder den Wald vor lauter Bäumen nicht gesehen hatte.