3D ELASTIC FULL WAVEFORM INVERSION OF RANDOM MEDIUM - FIRST RESULTS

S. Dunkl, A. Kurzmann, and T. Bohlen

email: *simone.dunkl@kit.edu*

keywords: *full waveform inversion, elastic, 3D, adjoint, gradient, random medium*

ABSTRACT

Full waveform inversion (FWI) with its high resolution and its ability to receive multiparameter images belongs to the most promising technologies in seismic tomography. We developed a 3D FWI code based on the conjugate adjoint gradient method. The wavefields are simulated using finite differences. To minimize runtime and storage, the simulation of the wavefields is performed in the time domain, whereas the gradients are calculated in the frequency domain. This can be done with the single frequency method, where only few discrete frequencies are employed. We performed a FWI of complex random medium, inverting for v_p and v_s in transmission geometry, using homogeneous starting models. 33 iteration steps with increasing frequencies were computed. The results show the ability of the code to resolve different sized features within the random medium. Hereby, the v_p model is less resolved and smoother than the v_s model due to the larger P-wavelengths. Furthermore, the innner part of the model, where wave path coverage is higher, is better resolved.

INTRODUCTION

In the 80's Tarantola (1984) and Mora (1987) developed a new inversion strategy known as full waveform inversion (FWI). By iteratively minimizing the misfit between observed and modeled seismograms this technique uses the full information content given by the waveforms. However, the method is computationally highly demanding and only in recent years, with heavy developments in computer science, FWI came back to view. Nowadays research concentrates on FWI and by now, the potential of FWI showed in synthetic studies (e.g. Brossier and Virieux (2011) and a few applications to field data (e.g. Brenders and Pratt (2007) and Sirgue et al. (2010)). It was shown, that detailed images can be achieved with resolution down to half a wavelength. Additionally, multiparameter inversion is possible, including seismic velocities and density as well as first approaches to invert for Q-factors and anelasticity (Virieux and Operto, 2009).
Today, most applications are still limited to the acoustic approximation. This can be sufficient in marine seismics for streamer data where P- to S-wave conversion can be neglected, however, in land seismics or ocean bottom seismics elastic effects have to be taken into account to fully recover the wavefield. A further advantage of the elastic approach is of course its potential to invert for the S-wave velocity.
Another confinement generally used is the restriction to the 2D inversion. 3D FWIs are computationally much more expensive and are thus limited to small problem sizes (Brossier and Virieux, 2011). In exchange, detailed images of 3D structures can be achieved. Furthermore, in highly heterogenous 3D medium, the 2D approximation is not valid anymore. With increasing comutational power, 3D FWI will become more and more important.
In this report, we will present a 3D elastic inversion code and present a first application to synthetic data. First, we will shortly review the underlying theory of FWI. Second, the implementation of the 3D FWI code will be explained. Last, the inversion of random medium in transmission geometry will be shown as an example.

THEORY OF FWI

The method we use to solve the inverse problem is the adjoint conjugate gradient method. Here we will only provide an overview about its theory and refer to Tarantola (1988), Mora (1987) and Köhn (2011) for detailed discussions.

Wave propagation

We use the elastic wave equation in the velocity-stress formulation:

$$\rho\frac{\partial v_i}{\partial t} - \frac{\partial \sigma_{ij}}{\partial x_j} = f_i \qquad \sigma_{ij} - c_{ijkl}\epsilon_{kl} = T_{ij} \qquad \dot{\epsilon}_{ij} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right) \qquad v_i = \dot{u}_i \tag{1}$$

Hereby ρ is the density and c_{ijkl} is the stiffness tensor. The wavefield is described by the particle velocity **v**, the displacement **u**, the stress tensor σ_{ij} and the strain tensor ϵ_{ij}. The source terms are the body force **f** and surface force T_{ij}. Additionally, initial and boundary conditions are applied. This forward problem can be solved very accurately with numerical methods like finite difference or finite element modeling.

The inverse problem and gradient calculation

Based on the L2 norm, the misfit is given by

$$M = \sum_{\substack{receiver\\source}} \sum_{timesteps} |(\mathbf{u}_{mod} - \mathbf{u}_{obs})|^2 \tag{2}$$

where $\mathbf{u}_{obs}$ and $\mathbf{u}_{mod}$ correspond to the real data and modeled data, respectively. Starting from some starting model $\mathbf{m}_0 = (\delta\rho_0, \delta c^0_{ijkl})$ the real model $\mathbf{m} = (\delta\rho, \delta c_{ijkl})$ is approached by iteratively minimizing the misfit function, which means adapting the modeled to the real data.
The gradient of M can be calculated with the adjoint method. We consider the linearized forward problem, which calculates the effects of a perturbation in the model space $\delta\mathbf{m}$ on the data space:

$$\delta\mathbf{u}_i = \mathbf{L}\delta\mathbf{m} = \int dV \frac{\partial u_i}{\partial \mathbf{m}}\delta\mathbf{m} \tag{3}$$

Using the fact, that the linear operator **L** is self-adjoint the inverse problem can be written as

$$\delta\mathbf{m'} = \sum_{source}\int dt \sum_{receiver} \frac{\partial u_i}{\partial \mathbf{m}}\delta u_i \tag{4}$$

For further calculations the Born approximation is used, where δu_i is approximated as first order perturbation. In the Born approximation the perturbed wavefield can be written as

$$\delta u_i(\mathbf{x}_r, t) = \int_V dV(\mathbf{x}) \int d\tau[-G^0_{ij}(\mathbf{x}_r, t-\tau; \mathbf{x})\frac{\partial^2 u_i}{\partial t^2}(\mathbf{x}, \tau)\delta\rho - \frac{\partial G^0_{ij}}{\partial x_k}(\mathbf{x}_r, t-\tau; \mathbf{x})\epsilon_{lm}(\mathbf{x}, \tau)\delta c_{jklm}(\mathbf{x})] \tag{5}$$

G^0_{ij} indicates the Green's function of the unperturbed medium, $\mathbf{x}_r$ the receiver position. Comparing equation (3) and equation (5) we find the expressions for $\frac{\partial u_i}{\partial \mathbf{m}}$ which can be inserted into equation (4). This gives us the following expression for the gradients:

$$\delta\rho' = -\sum_{sources}\int d\tau \frac{\partial^2 u_i}{\partial \tau^2}(\mathbf{x}, \tau)\Psi_j(\mathbf{x}, \tau) \tag{6}$$

$$\delta c'_{jklm} = -\sum_{sources}\int d\tau \epsilon_{lm}(\mathbf{x}, \tau)\frac{\partial \Psi_j}{\partial x_k}(\mathbf{x}, \tau) \tag{7}$$

with

$$\Psi_j(\mathbf{x}, \tau) = \sum_{receiver}\int dt G^0_{ji}(\mathbf{x}, t-\tau; \mathbf{x}_r)\delta u_i(\mathbf{x}_r, t) \tag{8}$$

Thus, the gradients can be calculated as the correlation between the forward propagating wavefield and the so called backpropagated wavefield Ψ_i. The latter propagates from the receiver into the medium, with the residual $\delta\mathbf{u}$ in reverse time acting as source time function.
In the frequency domain, the correlation is replaced by a simple product between forward propagated wavefield and complex conjugated backpropagated wavefield.

The conjugate gradient method

We use the conjugate gradient method, as described in Mora (1987). The model obtained in iteration k is given by:

$$\mathbf{m}_k = \mathbf{m}_{k-1} + \alpha \mathbf{p}_k \tag{9}$$

$\mathbf{p}_k$ denotes the search direction. In gradient methods it shows in the direction of steepest descent of the misfit function $\delta\mathbf{m}_k$, as calculated in equation 6-8. Additionally, to obtain the conjugate gradient direction some part of $\mathbf{p}_{k-1}$ is added

$$\mathbf{p}_k = \delta\mathbf{m}_k + \beta_k \mathbf{p}_{k-1} \tag{10}$$

resulting in a better convergence. The factor β_k is defined as given in Mora (1987). As the gradient only gives the direction of update, some steplength α has to be chosen to find the absolute model update.

IMPLEMEMTATION

Forward modeling with SOFI3D

The modeling of forward and backpropagated wavefield is based on the elastic version of the SOFI3D code, which was developed by Bohlen (2002). It is a finite difference code of 2nd to 12th order in space and 2nd order in time. Calculations take place on a staggerded grid system. Boundary reflections are efficiently reduced by using convolutional perfectly matched layers as introduced by Kamatitsch and Martin (2007). The code is parallelized using MPI.

Time-frequency FWI

For 3D problems time domain modeling is faster than frequency modeling. However, time domain inversion requires the storage of the wavefields for the whole time series. By contrast, for inversions in the frequency domain it was shown, that it is suffient to use only one or few frequencies for the gradient calculations, known as single frequency inversion (e.g. Sirgue and Pratt (2004)). Thus, storage costs decrease dramatically, as only the forward wavefields for these frequencies have to be stored, This means, that the most efficient method regarding runtime and storage needs is a combination of forward modeling in the time domain and inversion in the frequency domain. This was first suggested by Sirgue et al. (2008). Hereby the forward and backpropagated wavefields are calculated in the time domain and transformed into the frequency domain via discrete Fourier transformations, that is

$$v_i(\mathbf{x}, w) = \sum_{t=0}^{t_{max}} \exp(i\omega t) v_i(\mathbf{x}, t) \tag{11}$$

This transformation can be implemented very efficiently. The gradients are then calculated in the frequency domain.
There are two additional advantages of this strategy. First, it is possible to increase the frequency during the inversion without further effort. To start from low freqeuncies to higher frequencies is extremly important due to effects of nonlinearity. FWI is a local inversion method which bears the danger of ending in some local minima of the misfit function instead of reaching the global minima. For higher frequencies the number of local minima increases because data becomes increasingly nonlinear with respect to model perurbations (Sirgue and Pratt, 2004). It is consequently important to start from low freqencies, especially if the starting model is poor. Second, doing the forward modeling in the time domain enables an easy implementation of time windowing.

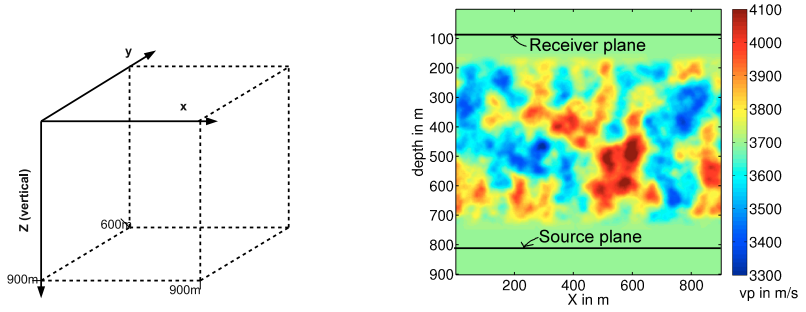
Step length calculation

After calculating the gradient, the optimal step length has to be found. This is done using a parabola method as proposed by Köhn (2011) and Kurzmann et al. (2009). Additionally to the misfit calculated in the current iteration step, which means for a steplength of zero, the misfit for two test step lengths is calculated. A parabola can be fitted to these three misfit values. The location of its minimum is then used as optimal steplength and the model is updated accordingly.

3D FWI OF RANDOM MEDIUM

Model and geometry

We performed a 3D elastic FWI of random medium. The random medium shows a Gaussian distribution with a standard deviation of 4% in the seismic velocities v_p and v_s, with average velocities of $v_p = 3700\ \frac{\mathrm{m}}{\mathrm{s}}$ and $v_s = 2150\ \frac{\mathrm{m}}{\mathrm{s}}$. A fractal spatial distribution with correlation lengths of $a_s = 35$ m and $a_p = 25$ m and Hurst coefficients of 1.6 and 1.9 was used for the v_s and v_p model, respectively. v_s and v_p velocities are thereby calculated independently, however, they show similar main features. The models show different sized structures as can be seen in the model slices plotted on the left hand of figure 5, figure 6 and figure 7. Thus they offer the possibility to investigate the resolution of different sized structures. The density model is homogeneous with $2000\ \frac{\mathrm{kg}}{\mathrm{m}^3}$ and was kept constant during inversion.
We use a cartesian coordinate system with x and y in horizontal and z showing in vertical direction (see figure 1(a)). The model size is $300 \times 200 \times 300$ gridpoints in x-, y- and z-direction. The grid spacing dh is 3 m with $\frac{a_s}{dh} = 11.7$ and $\frac{a_p}{dh} = 8.3$.



(a) coordinate system and model size

(b) vertical model slice with indication of source and receiver plane

Figure 1: coordinate system and model geometry, including source and receiver positions

A transmission geometry was used, which consists of 16 sources located in a horizontal plane of 810 m depth and 512 receivers located in a horizontal plane of 90 m depth. Source and receiver planes are indicated in figure 1(b). The receivers are situated at every tenth grid point in x- and y-direction to achieve a good coverage. The random medium passes into homogeneous medium towards source and receiver plane which are thus positioned within the homogeneous medium. We applied vertical point sources with a sin^3-function as source time function. The source wavelet and the corresponding spectra are plotted in figure 2.

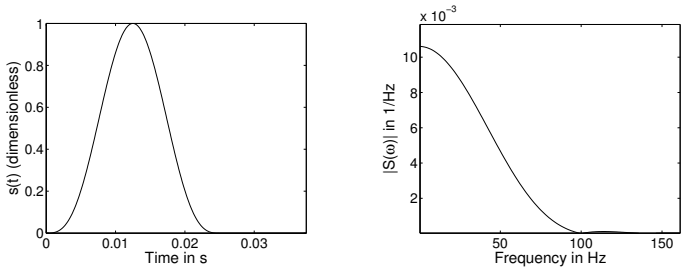


Figure 2: Source time function (left) with corresponding spectra (right) used for random medium model

Inversion procedure

We started with homogeneous starting models, using the average seismic velocities of the random medium. The frequency selection can be seen in table 1 together with the corresponding wavelengths λ_s and λ_p for v_s and v_p. In total 33 iteration steps were computed. Only one frequency per iteration was employed, with frequencies increasing from 20 Hz to 47 Hz. This frequency range corresponds to about 4-10 wavelengths per propagation distance for the P-wave and 7-17 for the S-wave velocity.

number of iterations	frequency in Hz	λ_p in m	λ_s in m
8	20	185	108
6	23	161	93
5	27	137	80
5	32	116	67
5	38	97	57
4	47	79	46

Table 1: Frequency selection for inversion of random medium

Due to high amplitude artefacts around sources and receivers, it was necessary to precondition the gradients. Thus, a Gaussian tapering was applied around source and receiver plane. Additionally, the model boundaries were tapered, as boundary reflections cause artefacts in these areas.
The calculations were performed on the JUROPA cluster in Jülich. With the use of 800 CPUs the total computation time for the 33 iterations amounted to about 17 hours. A total of 924 forward modelings was computed.

Results and discussion

The evolution of the misfit function in figure 3 shows how well the modeled data was adapted to the real data in each iteration step. The misfit is calculated in the time domain as given in equation 2, i.e. summed up over all unfiltered seismograms.

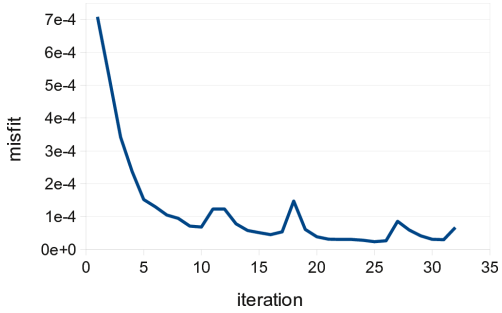
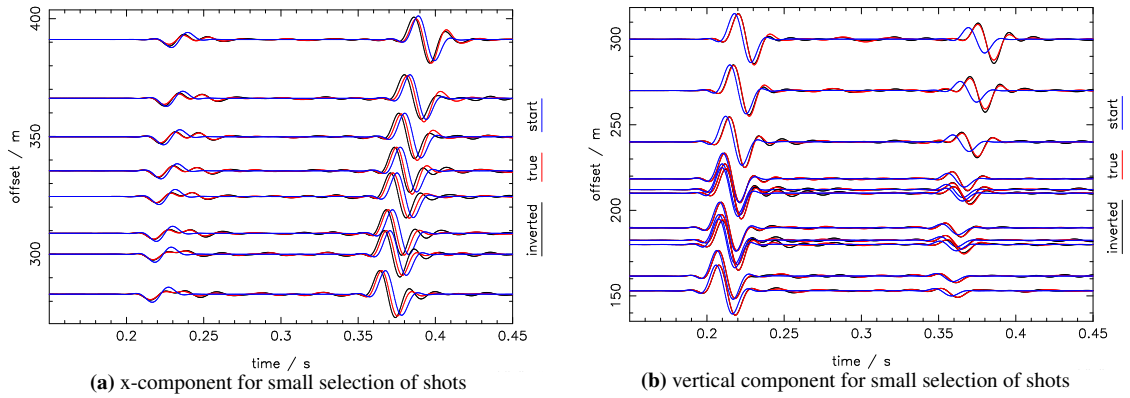


Figure 3: misfit versus iteration; misfit is summed up over all unfiltered seismograms

The misfit decreases very rapidly during the first 10 iterations, in which the gradient for the lowest frequencies was calculated. Here, at small frequencies most of the misfit reduction is done. In the later part of the inversion, there is only a slight decrease of the misfit. Additionally, in this part the misfit curve is not always monotonically decreasing, but shows some jumps to higher values for iterations higher than 10. Here, the steplength calculation apparently failed, which is probably due to the increasing nonlinearity of the misfit function at higher frequencies. Hence, the adaption of the data becomes more difficult for higher frequencies. The inversion result shown in the following is the result after 31 iterations.

In figure 4 two representative selections of seismograms for one shot are shown, with 8 traces of the x-component in figure 4(a) and 11 traces of the vertical component in figure 4(b). The seismograms are trace normalised and filtered with a lowpass filter of 47 Hz, because this is the maximum frequency used during inversion. For comparison we show the waveforms of the real model (red), the starting model (blue) and the inverted model (black).



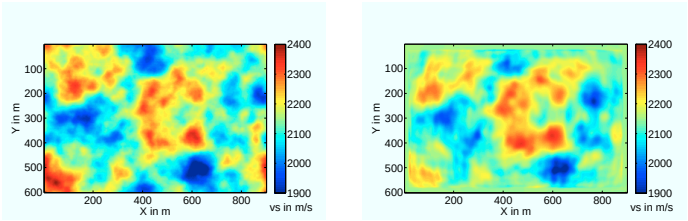
(a) x-component for small selection of shots

(b) vertical component for small selection of shots

Figure 4: Trace normalized seismograms of true, starting and inverted model filtered after application of a 47 Hz low pass filter

The inverted waveforms shown here fit quite nicely to the observed data. There are some differences to be seen, like for example in the S-wave of figure 4(a) or in some low amplitude oszillations. Of course, these traces can only exemplary show the adaption. In the overall 551 traces of each shot, worse and better quality of adaption can be found.

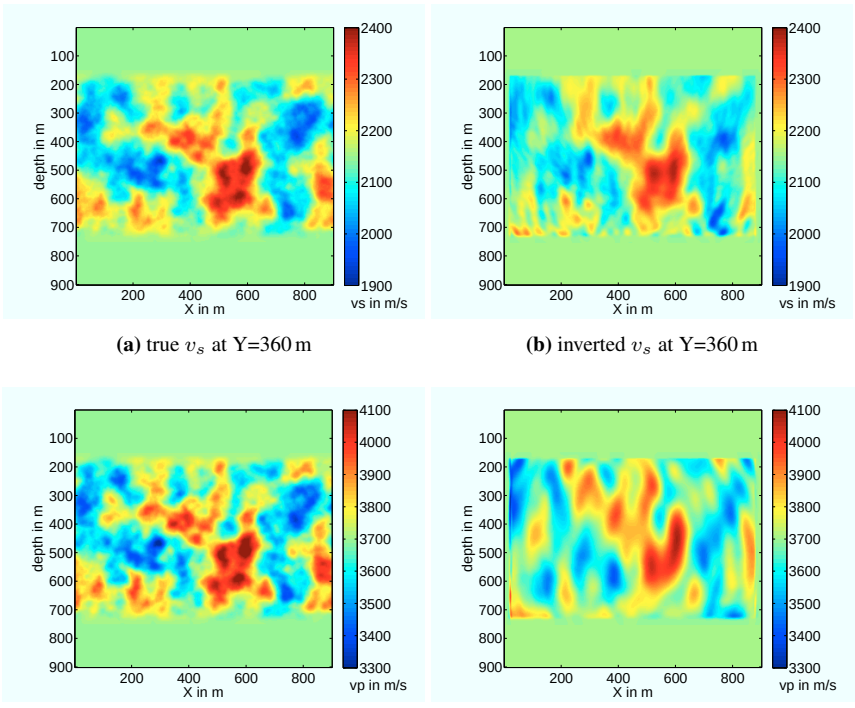
The "final" models as gained after 31 iterations can be seen in figure 5, figure 6 and figure 7. The 3D models are exemplified here by three orthogonal 2D slices. The left pictures display the true model. the right pictures the corresponding inverted model. For each slice the v_p model is plotted below the v_s model.



(a) true v_s **(b)** inverted v_s

(c) true v_p **(d)** inverted v_p

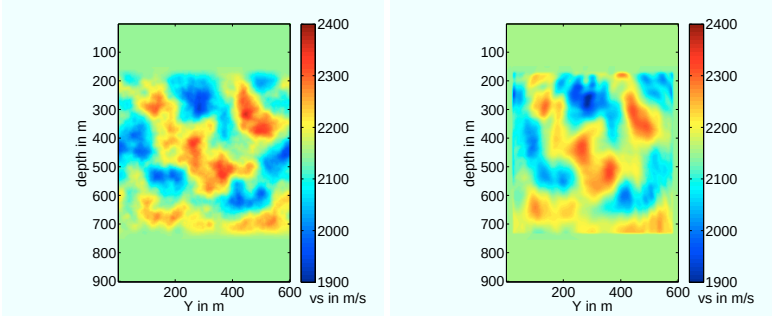
Figure 5: True and inverted (after 31 itearations) model for v_s and v_p at $z = 450\,\mathrm{m}$



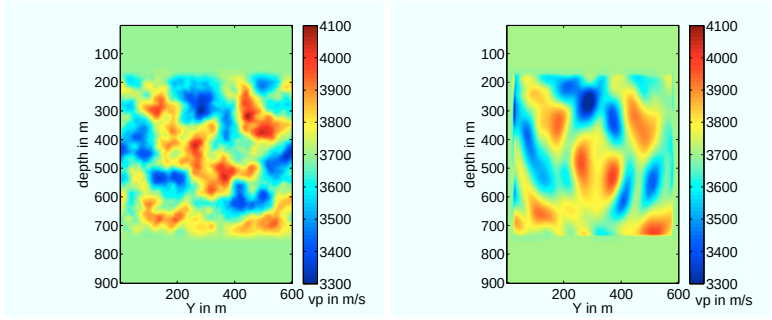
(a) true v_s at Y=360 m **(b)** inverted v_s at Y=360 m

(c) true v_p at Y=360 m **(d)** inverted v_s at Y=360 m

Figure 6: True and inverted (after 31 itearations) model for v_s and v_p at $y = 360\,\mathrm{m}$



(a) true v_s at X=390 m **(b)** inverted v_s at X=390 m



(c) true v_p at X=390 m **(d)** inverted v_p at X=390 m

Figure 7: True and inverted (after 31 itearations) model for v_s and v_p at $x = 390$ m

The main features are resolved quite well for v_p and v_s in all slices. However it can be seen, that the inner parts are resolved best, whereas quality is less towards the boundaries, which is reasonable due to the lower wavepath coverage there. Comparing the v_s and v_p model, it is visible, that v_s models are more detailed and better resolved than the inverted v_p models. This is mainly due to the different wavelengths used in the inversion. The highest frequency of 47 Hz corresponds to main wavelengths of about 79 m and 46 m for v_p and v_s, respectively. With a resolution of about half a wavelength, the model is much smoother for v_p. To receive more detail in v_p it would be necessary to invert for higher frequencies. As these plots show, that even though the decrease of misfit is highest in the lower frequency range, higher frequencies are still necessary to resolve the details and small sized features of the models.

CONCLUSIONS AND OUTLOOK

The inversion of random medium in transmission geometry showed the potential of the elastic 3D FWI code to resolve different sized features both in v_p and v_s in all spatial directions. The quality is especially good in the inner part where wave coverage is high. To invert for different sized structures a higher range of frequencies is necessary. Low frequencies build up a smooth model, which is very important to reduce the nonlinearity of the inverse problem. Higher frequencies offer clearer images with more detailed structures. However, upturns in our misfit curve show that these higher frequencies also complicate the inversion. Frequencies needed for a detailed v_p model are higher and we consequently did not reach the same resolution as in the inverted v_s model.
The inversion results showed, that the inversion succeeds with only one frequency per iteration. However, we plan to repeat the inversion using a group of frequencies, because an improved redundancy can lead to better robustness of FWI and thus more stable results. This might help us to include higher frequencies and to improve the resolution of this example.

ACKNOWLEDGMENTS

The work was performed within the project TOAST which is part of the GEOTECHNOLOGIEN program, funded by the German Ministry of Education and Research (BMBF) and the German Research Foundation (DFG), grant 03G0752A
This work was kindly supported by the sponsors of the *Wave Inversion Technology (WIT) Consortium*.
The calculations were performed on the JUROPA cluster at Jülich supercomputing center.

REFERENCES

Bohlen, T. (2002). Parallel 3-D viscoelastic finite difference seismic modeling. *Computers and Geoscience*, 28:887–889.

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

Brossier, R. and Virieux, J. (2011). Lecture notes on full waveform inversion. *SEISCOPE Consortium, Grenoble, France*.

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

Köhn, D. (2011). Time domain 2d elastic full waveform tomography [dissertation]. *Christian-Albrechts-Universität zu Kiel*.

Kurzmann, A., Köhn, D., Przebindowska, A., Nguyen, N., and Bohlen, T. (2009). 2D acoustic full waveform tomography: performance and optimization. *70th EAGE Conference and Technical Exhibition*.

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

Sirgue, L., Barkved, O., Dellinger, J., Etgen, J.and Albertin, U., and Kommeda, J. (2010). Full waveform inversion: the next leap forward in imaging at valhall. *First Break*, 28:65–70.

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

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

Tarantola, A. (1984). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.

Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. *Pure and Applied Geophysics*, 128(1/2):365–399.

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