



The influence of the absorbing boundary condition on full-waveform inversion: a GPU implementation

Victor Koehne*, CPGG/UFBA

Copyright 2018, SBGf - Sociedade Brasileira de Geofísica.

This paper was prepared for presentation at the 8th Symposium of the Brazilian Geophysical Society, held in Salinópolis, Brazil, on 18 to 20 September, 2018.

Contents of this paper were reviewed by the Technical Committee of the 8th Symposium of The Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of The Brazilian Geophysical Society is prohibited.

Abstract

In many test cases of time-domain full-waveform inversion (FWI) methodologies, both the observed and calculated seismic data are synthetic. This paper aims to analyse the impact on the final inversion if different absorbing boundary conditions are used in observed and calculated data. The inversion is also analysed for an observed data free of boundary reflections (closer to a real dataset), and its inversion using the "taper" and a perfectly matching layer (PML) boundaries. To carry out the tests in a timely manner, a GPU implementation of multiscale FWI with a finite-differences scheme for modeling and a L-BFGS algorithm with variable step-length for inversion were used. We show that the inversion is valid if different boundary conditions on observed and calculated data are used, but that the best final velocity model is obtained when the same boundary condition is applied; and that for a dataset free of boundary-reflection events, the PML guarantees a better final inverted model when compared to the taper.

Introduction

Full-waveform inversion (FWI) is becoming established as one of the main techniques for high resolution velocity estimation. The time domain formulation proposed by Lailly (1983) and Tarantola (1984) consists of updating the velocity model by minimizing a misfit function between recorded and observed data. They show that the gradient of the misfit function along which the perturbation model is searched can be built by crosscorrelating the incident wavefield emitted from the source and the back-propagated wavefields (Virieux and Operto, 2009). The basis of every iterative time domain FWI minimization is the calculation of this gradient at each iteration, which require wavefield extrapolation of the source and residue. The computation of the wavefields numerically, in finite grids, presents the persistent problem of artificial reflections at the edges of the model (Clayton and Engquist, 1977; Engquist and Majda, 1977). This paper aims to study the effects of these border effects on time domain FWI.

As booming field of study, many methodologies have been analysed for time domain FWI. Many of the papers in the literature use synthetic data to represent the observed data \mathbf{d}_{obs} , often calculated with the same modeling operator

and absorbing boundary condition used to calculate the modeled wavefield yielding \mathbf{d}_{cal} required for the gradient. Among many examples, we could cite Yang et al. (2015) and Chen and Wang (2017) which uses in \mathbf{d}_{obs} and \mathbf{d}_{cal} the same absorbing boundary condition (ABC) proposed by (Clayton and Engquist, 1977; Engquist and Majda, 1977). The results in this paper show that the quality of the final velocity model estimated by FWI using synthetic \mathbf{d}_{obs} is greatly affected by the edge events, and the absorbing boundary condition used.

To explore this influence we ran the FWI for two different boundary conditions: one that tapers the amplitudes of the wavefields at the border (taper, proposed by Cerjan et al., 1985), and one using a version of the perfectly matched layer (PML, Collino and Tsogka, 2001). We also modeled a \mathbf{d}_{obs} free of edge reflections and tested FWI with \mathbf{d}_{cal} using each of these absorbing boundary conditions.

As time domain FWI is a local optimization problem, it is prone to convergence to local minima. An efficient way to avoid this problem is to use the multiscale approach (Bunks et al., 1995), where the first iterations are done for the lower frequency content, and higher frequencies are included gradually in the process. Estimating a good approximation of the Hessian, applied to the gradient in the Newton's formulation of the problem, also plays an important role on avoiding local minima. To ensure a good Hessian approximation, we used the L-BFGS (Nocedal and Wright, 2006) algorithm.

Wavefield extrapolation is the core of every iteration of time domain FWI, and is a highly intensive computational process. Fortunately, it is a highly parallelizable problem. It can be parallelized for each shot or for each point of the wavefield. The first type of parallelization (on shots) is usually done using several CPUs, for example using a MPI implementation (dos Santos, 2013). The second type, wavefield points on parallel, and one shot at a time, can be better implemented by using the graphics processing unit (GPU), the method of choice in this paper.

We study the influence of different absorbing boundary conditions on FWI in two set of tests. In the first set we compare the inversion for \mathbf{d}_{obs} and \mathbf{d}_{cal} with the same ABCs, and each with a different ABC. In the second set we test FWI for a \mathbf{d}_{obs} modeled with no border reflection events, and inversions done for \mathbf{d}_{cal} with taper, and PML.

Overview of time-domain acoustic FWI

In the classical formulation given by Lailly (1983) and Tarantola (1984), FWI is an optimization problem that seeks to find the best velocity model \mathbf{m} that minimizes in a least-squares manner a misfit function E , which measures the error between an observed data \mathbf{d}_{obs} and calculated data \mathbf{d}_{cal} obtained from solving the forward problem over \mathbf{m} :

$$E(\mathbf{m}) = \frac{1}{2} \|\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}\|^2 = \|\mathbf{d}_{\text{obs}} - F(\mathbf{m})\|^2, \quad (1)$$

where $\|\cdot\|^2$ is the l^2 -norm and $F(\cdot)$ is the forward modeling operator, which in an isotropic, acoustic and constant-density 2-D medium is given by:

$$\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = v^2(\mathbf{x}) \nabla^2 p(\mathbf{x}, t) + s(\mathbf{x}, t), \quad (2)$$

with $p(\mathbf{x}, t)$ being the pressure wavefield, $v(\mathbf{x})$ is the acoustic velocity (representing the model \mathbf{m}), ∇^2 the Laplacian operator, and $s(\mathbf{x}, t)$ is a source term. $F(\cdot)$ can then be interpreted as the operator that finds p for each source over time and restrains it to receivers geometry (typically over the surface), matching \mathbf{d}_{cal} to \mathbf{d}_{obs} .

The minimization of the misfit function 1 is then achieved iteratively, due to its highly nonlinear behavior. Following Newton's method formulation, the model update is given by (Ma and Hale, 2012):

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{H}_k^{-1} \mathbf{g}_k. \quad (3)$$

The gradient \mathbf{g} can be efficiently computed using the adjoint-state method (Lailly, 1983; Tarantola, 1984; Plessix, 2006):

$$\mathbf{g} = \frac{2}{v^3} \int_0^{t_{\text{max}}} \lambda \frac{\partial^2 p}{\partial t^2} dt, \quad (4)$$

with t_{max} being the maximum time recorded on the data set, p the forward propagated source wavefield, and λ is the back-propagation of the residues, using $s(\mathbf{x}, t) = (\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}})$ as source for equation 2. The calculation of \mathbf{g} is accumulated shot by shot, correlating fields λ and $\frac{\partial^2 p}{\partial t^2}$ in the same fashion as in a reverse time migration. The derivative field $\frac{\partial^2 p}{\partial t^2}$ is usually calculated as $v^2 \nabla^2 p$ (equation 2), as the Laplacian is already available from the numerical extrapolation scheme.

The Hessian \mathbf{H} has a large size and its explicit numeric calculation is very costly, as it requires too many extrapolation steps to be computed (Virieux and Operto, 2009). To circumvent this problem, the iterative update equation 3 can be re-written in a generalized form as (Ma and Hale, 2012):

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{h}_k, \quad (5)$$

In this paper we follow the development of Yang et al. (2015) for α_k , using an extra modeling process with a tentative velocity model to in order to accurately calculate the step length (Pica et al., 1990). The term \mathbf{h}_k is a search direction that can be determined by gradient methods, such as steepest-descent or nonlinear conjugate gradient (Yang et al., 2015) or quasi-Newton methods, such as the L-BFGS (Liu and Nocedal, 1989). In this paper the L-BFGS was the method of choice.

The L-BFGS method was designed to directly compute the search direction, without explicitly storing the Hessian matrix approximation (dos Santos and Pestana, 2015). The method initializes direction \mathbf{h}_k using the current gradient \mathbf{g}_k

as given by equation 4. It also makes use of the previous models and gradients in the last m iterations, in order to approximate the Hessian (Nocedal and Wright, 2006). For the tests in this paper, 10 past iterations were used ($m = 10$).

FWI is a highly nonlinear problem, and the local minimization procedure is prone to converge to local minima. To avoid that we used a multiscale frequency approach, where the first iterations are done for the lower frequency content, and higher frequencies are included gradually in the process. A detailed explanation of the multiscale process can be found in Bunks et al. (1995).

Regular grid solutions to the edge reflection problem

Solving the Laplacian in 2 numerically (using the neighboring points) causes the wavefield to reflect when it reaches the limits of the model; since there are no neighboring points to calculate the Laplacian, the process must be stopped, causing the wavefield to reflect back to the model (this also happens in the pseudo-spectral method, since the spatial wavenumbers are discretized). This long known problem in seismic modeling have since been solved by extending the velocity and pressure fields, and at these extended boundaries, applying some amplitude attenuation condition to the amplitudes. Greatly extending the grid so that the boundary reflections could not interfere with target events also solves the problem, but is unfeasible from a computational point of view (since the bigger the grid, the longer the processing). One of the most used techniques to attenuate border events on a regular finite-difference scheme is to gradually taper the amplitudes at the extended border with a factor (Cerjan et al., 1985):

$$T(ib) = e^{-[0.015(nb-ib)]^2}, \quad (6)$$

with ib being the distance from model edge to current point inside extended border, and nb the size of the extended border.

Another powerful technique well-established by Collino and Tsogka (2001) for attenuation of the edge events on the numerical solution of the elastodynamic problem is the perfectly matched layer (PML). Interested readers are referred to Collino and Tsogka (2001). Here we develop an adaptation of this method for the regular grid acoustic case. Our approach to the problem is to use, at the extended boundaries, a damped acoustic wave equation (Moreira et al., 2011):

$$\frac{\partial^2 P}{\partial t^2} + 2\alpha B \frac{\partial P}{\partial t} + \alpha^2 B^2 P = c_0^2 \nabla^2 P, \quad (7)$$

which when solved with a 2^{nd} order centered finite-difference for $\frac{\partial^2 P}{\partial t^2}$, and forward finite-difference for $\frac{\partial P}{\partial t} \approx \frac{P(t+\Delta t) - P(t)}{\Delta t}$, gives:

$$P(t + \Delta t) = \left[\frac{1}{1 + 2\alpha B \Delta t} \right] \left\{ 2P(t) - P(t - \Delta t) + c_0^2 \Delta t^2 \nabla^2 P(t) + [2\alpha B \Delta t - \alpha^2 B^2 \Delta t^2] P(t) \right\}, \quad (8)$$

where (Moreira et al., 2011):

$$B = c_0^2$$

$$\alpha(ib) = \frac{1}{c_0^2 \Delta t} \ln \left(\frac{1}{r_{pml}} \right) \left(\frac{ib}{nb} \right)^k, \quad (9)$$

and the best results were obtained for r_{pml} , the maximum absorption coefficient, being equal to 0.8, and the exponent k being equal to 4. In this paper we tested the proposed by Moreira et al. (2011) and a modified form for α that provided better results, adapted from Collino and Tsogka (2001) where:

$$\alpha(ib) = \frac{3}{2 nb c_0} \ln \left(\frac{1}{r_{pml}} \right) \left(\frac{ib}{nb} \right)^k, \quad (10)$$

with $r_{pml} = 0.5$ and $k = 2$.

In our implementations the PML is slightly more costly than the taper, since the latter can be applied by a single step - multiplying the amplitudes by the factor T^- , while PML requires two steps - calculating the laplacian of $p(\mathbf{x}, t)$ and its damped version (last term in equation 8), and then compute $p(\mathbf{x}, t + \Delta t)$.

Figure 1 allows a visual comparison between propagations: without absorbing boundaries (a), with taper (b), with PML (Moreira) (c), and with PML (modified Collino-Tsogka) (d). Analysing the reflection in the bottom edge, for this 2-layer model, the taper was the one that most attenuated the lower edge event. Since the modified-Collino-Tsogka approach showed slightly better results compared to the Moreira-approach, in the FWI tests we only compared taper and modified-Collino-Tsogka boundary conditions.

GPU implementation

Seismic modeling is a highly intensive computation process. In this implementation each iteration of FWI roughly requires three seismic modeling processes: two for p, λ (gradient) and one for $F(\mathbf{m}_k + \epsilon \mathbf{h}_k)$ (step length). Fortunately, it is a highly parallelizable problem. It can be parallelized for each shot or for each point of the wavefield. The first type of parallelization (on shots) is usually done using several CPUs, for example using a MPI implementation (dos Santos, 2013). The second type, wavefield points on parallel, and one shot at a time, can be better implemented by using the graphics processing unit (GPU), the method of choice in this paper.

The GPU was initially designed for graphical renderization, where every image pixel is constantly updated in parallel (NVIDIA, 2010). GPU dedicates more transistors to data processing than CPU, and less for data cache and control, thus being faster for floating point calculations, and slower for memory transfer processes. In 2006, NVIDIA started launching hardwares with CUDA (compute unified device architecture), expanding the pixel-oriented calculations to more general types of data, and a C based language, CUDA C (Sanders and Kandrot, 2010).

The GPU is efficient for intensive calculation problems that require few memory transfers. This is the case for seismic modeling, that consists of updating a wavefield over time, and only transferring data back to the CPU when the shot is finished. In this paper the wavefield extrapolations were done on the GPU, one shot at a time. Interested readers

on GPU implementation of seismic modeling are referred to Wang et al. (2011).

Results

To understand the influence of the absorbing boundary in \mathbf{d}_{obs} and \mathbf{d}_{cal} , and later on the FWI, we initiate the analysis using a 28 shots, regular split-spread data, with 101 receptors in each shot, and an interval of 10 samples between shots over a reduced Marmousi model (Irons, 1988) of size 125×383 , modeled with no absorbing boundary condition. The reduced Marmousi model is shown in Figure 3-a, and the shots modeled with no border in Figure 2-a. The most noticeable events arising from reflections on edges are elucidated in Figure 2-b: there is a strong reflection from the top limit, present in all shots; reflections on the left side of the model in the initial shots; and reflections on the right side of the model in the last shots.

An efficient way of generating an "infinite border", which is a real acquisition situation (there is no border, except the top edge corresponding to the air-earth interface), is to model with a boundary thickness two times bigger than the vertical dimension of the model, in a way that the arrival times for the edge reflections are bigger than the maximum recorded time. In this example, $n_z = 125$, and the border thickness is $nb = 256$. The result is a seismogram free of boundary events, but the processing and memory required for this modeling is around 6 times greater than modeling with $nb = 32$ and absorbing boundaries. For this reason, in most cases, modeling without border events is not feasible (in time and/or memory).

Figures 2-c,d highlights the border events for modelings with "infinite boundary" (2-c), without boundaries (2-d). Figures 2-e,f with $nb = 32$ and taper (2-e) and with $nb = 32$ and PML (figure 2-f). The size 32 was chosen empirically, since it provided a good attenuation of the edge events at a low computational/memory cost.

Every result on each set of tests was obtained from 120 iterations of a multiscale FWI, with 4 rounds of 30 iterations, and the frequency bands divided in: 0 – 7.5 Hz, 0 – 14 Hz, 0 – 22.5 Hz and 0 – 30 Hz. For the source we used a Ricker wavelet.

In the first test we ran each of the following cases:

1. \mathbf{d}_{obs} and \mathbf{d}_{cal} with *taper*,
2. \mathbf{d}_{obs} and \mathbf{d}_{cal} with PML,
3. \mathbf{d}_{obs} with PML and \mathbf{d}_{cal} with *taper*.

The results are shown in Figure 5, and are very similar. The total RMS error (comparing result and original model) on figure 5-d for each inversion shows with more detail that "PML-PML" is better in the first 30 iterations, corresponding to the lowest frequency multiscale round; in the following rounds, where greater frequencies are included, "taper-taper" performs better. As expected, the "PML-taper" result is the worst of all, but still produces a reasonable image (figure 5-c), although the oil reservoir in the anticline of the Marmousi model (at $z = 2.5km$ and $x = 6.5km$) is not well imaged.

It shows that FWI for synthetic \mathbf{d}_{obs} and \mathbf{d}_{cal} gives the best

results if the absorbing boundary condition is the same for both datasets; and that taper in both datasets gave the best results.

The second and final test evaluates FWI for a \mathbf{d}_{obs} closer to reality, modeled with $nb = 256$, without border events (Figure 2-c). We then test the inversion for \mathbf{d}_{cal} with taper (Figure 2-e), and for \mathbf{d}_{cal} with PML (Figure 2-f). The result for each test is shown in Figure 4 (a) and (b), respectively. To quantify the results a total RMS error (compared to original model) is presented (Figure 4-c), providing more detail to the analysis.

Conclusions

The inversion using \mathbf{d}_{cal} with PML showed better results than the one with taper, but neither imaged the reservoir at $(z, x) = (2.5\text{km}; 6.5\text{km})$ correctly. We can notice that the edge events influence the inversion process, since they affect the residue $\mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{cal}}$, but not to the point of making it unrealizable. This test, when compared with the previous one, shows that the attenuated edge events present in \mathbf{d}_{obs} and \mathbf{d}_{cal} are assimilated in the inversion process as relevant events in the imaging process, since for example the reservoir on the model is well imaged for the first set of tests, but not for the second. The inversion results for this second test is still considered bad, based on the fact that the final RMS error for the inverted model is greater than the RMS error for the initial model (Figure 4-c; both compared to the original model). In conclusion, all the tests suggest that when dealing with real data, PML usage for \mathbf{d}_{cal} can give better results.

Acknowledgements

This research was supported by CAPES, CNPq and INCT-GP/CNPq in the facilities of CPGG/UFBA. We also thank IFP for providing the Marmousi model.

References

- Bunks, C., F. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473.
- Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition for discrete acoustic and elastic wave equations: *Geophysics*, **50**, 705–708.
- Chen, G. and Z. Wang, 2017, Robust full-waveform inversion based on particle swarm optimization: *SEG Technical Program Expanded Abstracts*, 1302–1306.
- Clayton, R. and B. Engquist, 1977, Absorbing boundary conditions for acoustic and elastic wave equations: *Bulletin of the Seismological Society of America*, **67**, 1529–1540.
- Collino, F. and C. Tsogka, 2001, Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media: *Geophysics*, **66**, 294–307.
- dos Santos, A. W. G., 2013, Inversão de forma de onda aplicada à análise de velocidades sísmicas utilizando uma abordagem multiescala: *Dissert. de Mestrado, Universidade Federal da Bahia, Salvador, Brasil*.
- dos Santos, A. W. G. and R. C. Pestana, 2015, Time-domain multiscale full-waveform inversion using the rapid expansion method and efficient step-length estimation: *Geophysics*, **80(4)**, R203–R216.
- Engquist, B. and A. Majda, 1977, Absorbing boundary conditions for numerical simulation of waves: *Bulletin of the Seismological Society of America*, **67**, 1529–1540.
- Irons, T., 1988, Marmousi model: 52nd eage meeting, Institut Français du Pétrole, Rueil-Malmaison, Paris.
- Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations: *Conference on inverse scattering: Theory and Applications*, 206–220.
- Liu, D. and J. Nocedal, 1989, On the limited memory bfgs method for large scale optimization: *Mathematical Programming*, **45**, 503–528.
- Ma, Y. and D. Hale, 2012, Quasi-Newton full-waveform inversion with a projected Hessian matrix: *Geophysics*, **77**, R207–R216.
- Moreira, R., A. Bulcão, R. Pessolani, T. Kassuga, E. Velandia, and F. Monteiro, 2011, Perfect matched layer method applied to acoustic wave modeling: *Proceedings of COBEM 2011*.
- Nocedal, J. and S. Wright, 2006, *Numerical optimization: Springer series in operations research and financial engineering*.
- NVIDIA, 2010, *NVIDIA CUDA C programming guide: Version 3.1.1*.
- Pica, A., J. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: *Geophysics*, **55**, 284–292.
- Plessix, R., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- Sanders, J. and E. Kandrot, 2010, *CUDA by example: Addison-Wesley*.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, WCC127–WCC152.
- Wang, B., J. Gao, H. Zhang, and W. Zhao, 2011, Cuda based acceleration of full waveform inversion on GPU: *SEG Technical Program Expanded Abstracts*, 2528–2533.
- Yang, P., J. Gao, and B. Wang, 2015, A graphics processing unit implementation of time-domain full-waveform inversion: *Geophysics*, **80(3)**, F31–F39.

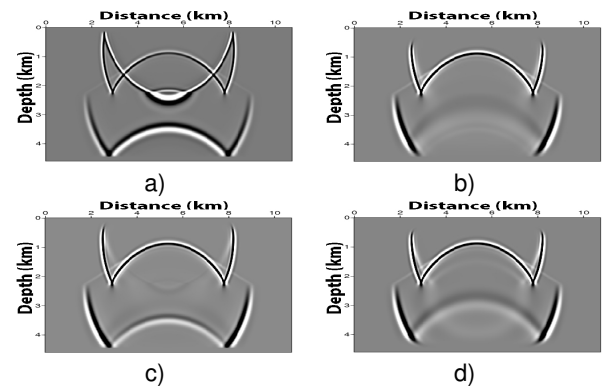


Figure 1: Propagation in a 2-layer model, after 20ms. a) No absorbing boundary; b) with taper; c) with PML (Moreira); d) with PML (modified Collino-Tsogka).

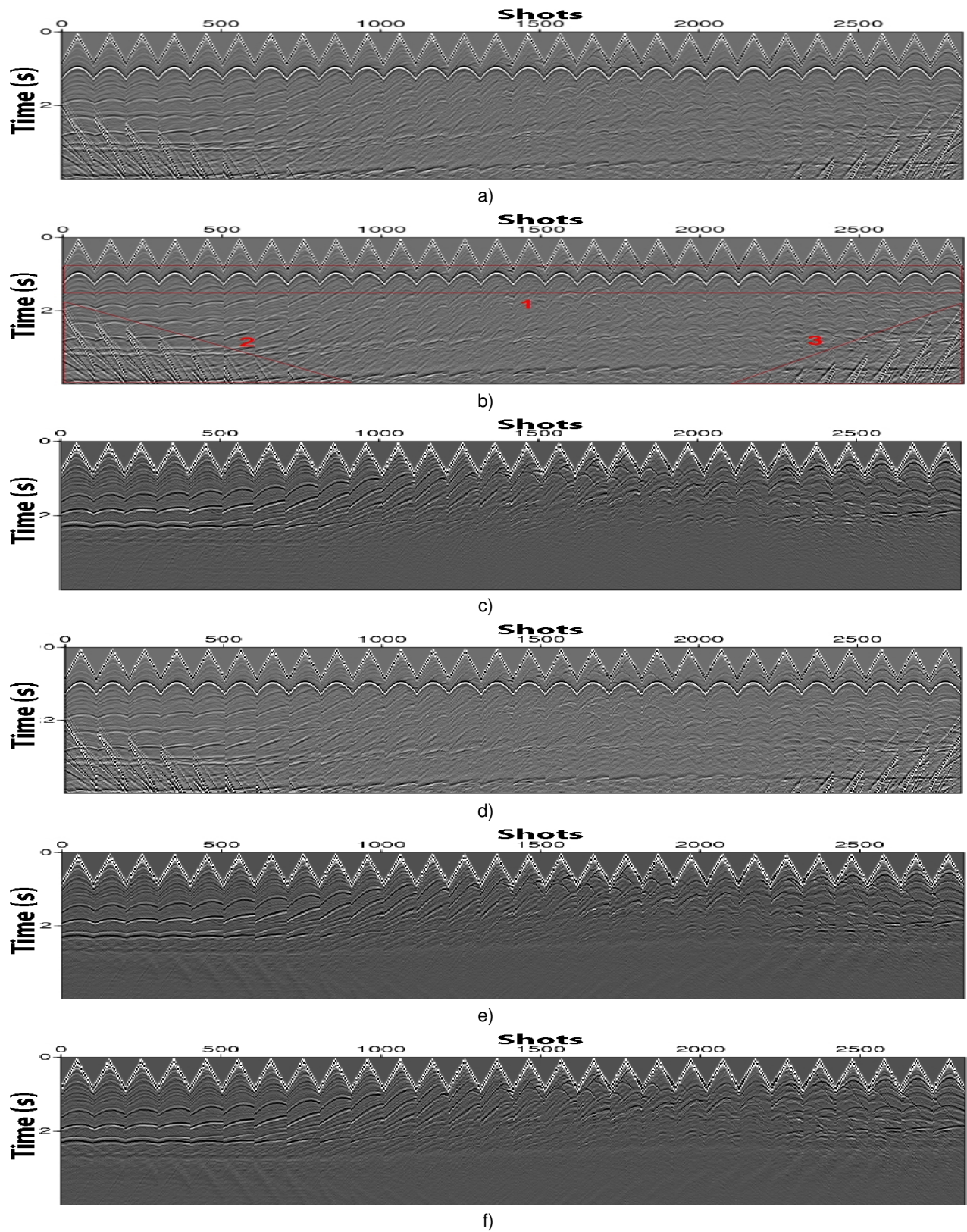


Figure 2: a) 28 shots modeled with no absorbing boundaries. b) Highlighted edge reflection events: reflections on the top border (1); reflections on the left border (2); reflections on the right border (3). c) Shots modeled with border thickness of size 256, without edge reflections. d) Shots modeled without absorbing boundaries. e) Shots modeled with taper. f) Shots modeled with PML.

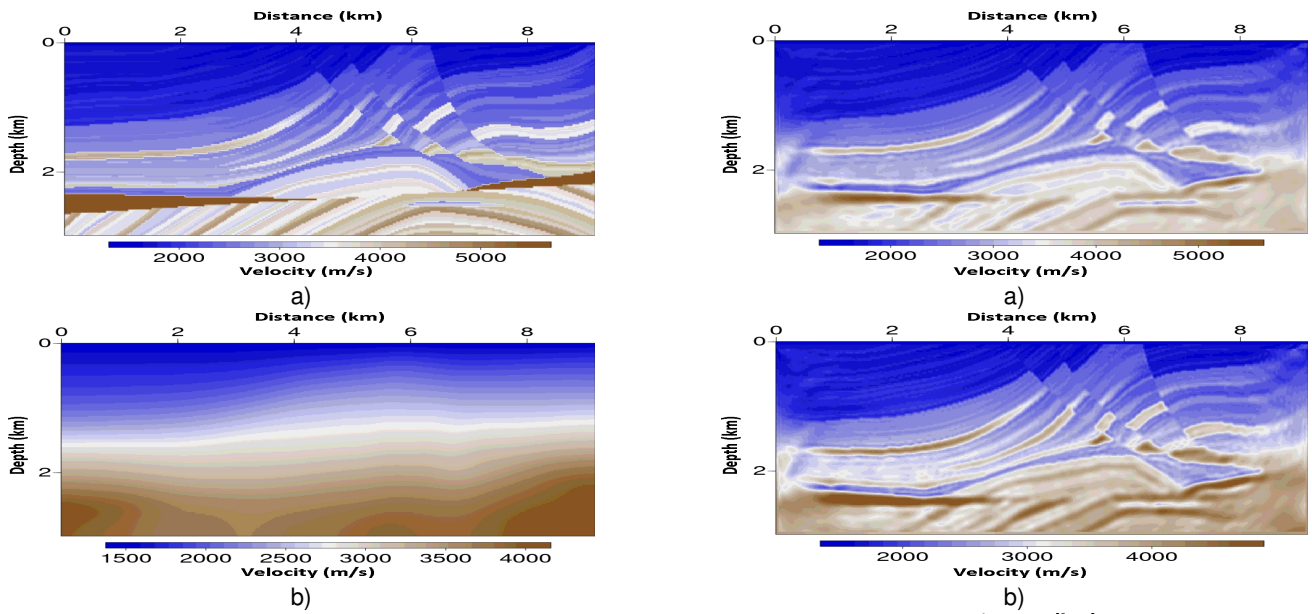


Figure 3: Reduced (125×383) Marmousi model. a) Original. b) Smoothed twice.

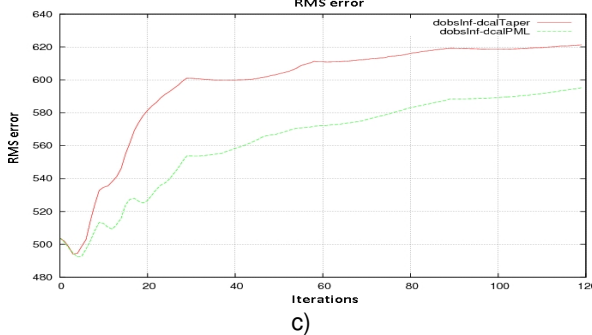
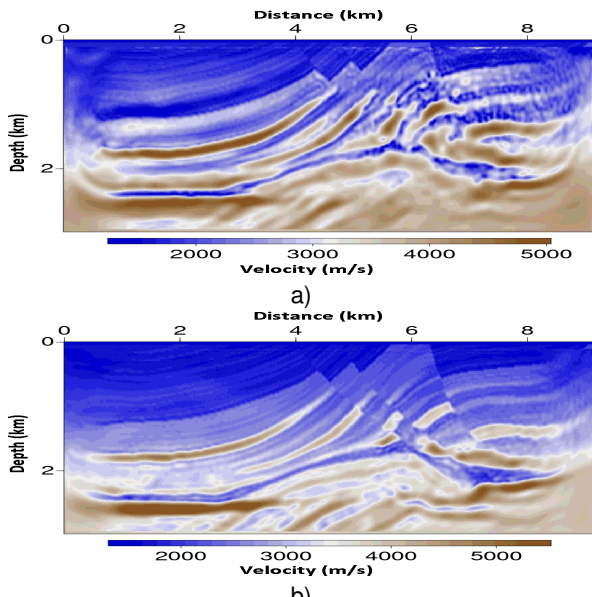


Figure 4: Results of FWI for a d_{obs} with "infinite border" (no edge reflection events). a) d_{cal} with taper. b) d_{cal} with PML. c) Total RMS errors for results (a) and (b).

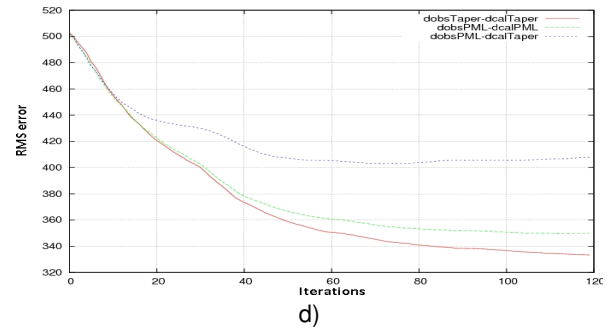
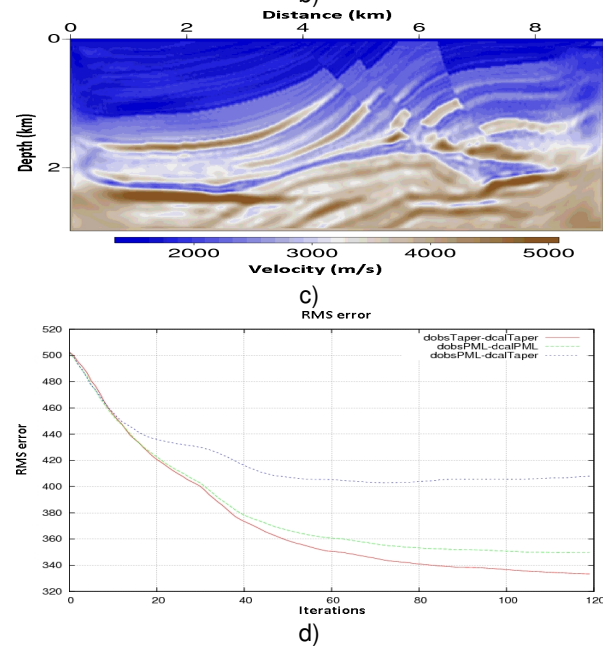


Figure 5: Results of multiscale FWI, with boundary thickness of size 32. a) d_{obs} with taper, d_{cal} with taper. b) d_{obs} with PML, d_{cal} with PML. c) d_{obs} with PML, d_{cal} with taper. d) Total RMS errors for results (a), (b) and (c).