

## MULTISCALE FULL WAVEFORM INVERSION USING A FILTERING APPROACH AND GRADIENT PRECONDITIONING

Rafael Abreu de Cristo<sup>1</sup> and Milton J. Porsani<sup>2</sup>

**ABSTRACT.** The full waveform inversion (FWI) multiscale approach in data domain produces better results because it gets closer to the overall minimum avoiding the local minima. The method works solving the problem in different scales, avoiding the interference of the initial velocity model choice as well as the cycle skipping. The multiscale approach was done after choosing frequency bands using Wiener filter and SVD filter trace by trace, both in data domain. The trace by trace SVD filter assembles each trace of the gradient onto the shifted matrix traces and do the decomposition from low to high frequencies. In addition this multiscale approach in data domain was compared to another multiscale approach using damping filters on the objective function (MDFOF). Due to the geometrical spreading, during the propagation of the wave field, the deeper regions of the model are not well illuminated, hence the preconditioning of the gradient was done in order to eliminate this issue and allow the deeper regions to be compared.

**Keywords:** SVD filter, pseudo-Hessian diagonal, geometrical spreading, Wiener filter.

**RESUMO.** A abordagem multiescala na inversão da forma de onda completa (FWI) produz melhores resultados pois consegue convergir para o mínimo global, evitando o problema do mínimo local. O método funciona em diferentes escalas, evitando a interferência da escolha do modelo inicial de velocidades, bem como o efeito de salto de ciclo. A abordagem multiescala foi realizada escolhendo-se bandas de frequências usando-se os filtros de Wiener e de SVD traço a traço, ambos no domínio do dado. O filtro SVD traço a traço insere cada traço do gradiente na matriz de traços deslocados e faz a decomposição das baixas às altas frequências. Outra abordagem multiescala usando filtros atenuantes foi comparada com a anterior, no domínio do dado. Devido à divergência esférica, durante a propagação da onda, as regiões mais profundas do modelo não são corretamente imageadas, portanto faz-se necessário o condicionamento do gradiente com intuito de eliminar esse problema e permitir a comparação das duas abordagens nas regiões mais profundas do modelo.

**Palavras-chave:** filtro SVD, diagonal pseudo-Hessian, divergência esférica, filtro de Wiener.

---

<sup>1</sup>Universidade Federal da Bahia (UFBA), Center for Research in Geophysics and Geology (CPGG), Rua Barão de Geremoabo, s/n, Salvador, BA, Brazil  
– E-mail: rafael.abreudecristo@hotmail.com

<sup>2</sup>Universidade Federal da Bahia (UFBA), Center for Research in Geophysics and Geology (CPGG), National Institute of Science and Technology of Petroleum Geophysics (INCCT-GP), Rua Barão de Geremoabo, s/n, Salvador, BA, Brazil – E-mail: porsani@cpgg.ufba.br

## INTRODUCTION

From the physical properties of the rocks in subsurface it is possible to estimate the presence of hydrocarbons in an indirect way. Seismic velocity measurements of density and impedance, for example, are physical properties of the rocks which may assist in the determination of oil and gas plays. One strategy used to estimate these physical properties of rocks was the CDP (Yilmaz, 2001) technique which is a product from the velocity field, obtained from an NMO correction for hyperbolic events. In this technique the subsurface geology was considered flat and stratified. Because the layers are not stratified and have a very large lateral variation precluding a coherent velocity analysis can generate some challenge problems.

The development of imaging techniques allowed achieve seismic velocity models more accurately according the subsurface geology, allowing a greater accuracy in the determination of reservoirs, source rocks that previously those structures were undetectable. Regarding these new technologies, the seismic tomography (Stork, 1992) is a technique used to estimate the field of optimized seismic velocities, from a travel time table. The inversion is performed only by the travel time without taking into account the amplitude and the phase. On the other hand, other techniques using travel time are used as pre-migration seismic velocity analysis which can also be obtained by a seismic velocity field (Lambaré et al., 2007).

Since the development of high performance computing, Tarantola (1984) in its series of papers proposed a linearised inversion using the full wave equation as the forward problem solution. The inversion of the waveform arises from this initial idea of Tarantola, and the main objective is not only to make an inversion considering the travel time, but the phase and amplitude of the propagated wave field. Nowadays, the FWI technique has been widely used in 2-D and 3-D problems producing fields of high-resolution seismic velocities in order to estimate more accurately subsurface geological structures (Virieux & Operto, 2009).

FWI is a non-linear and ill posed problem, i.e, there are more parameters to be determined than observed data taking into account, of course, because the limited surface geometry. The waveform inversion can be solved by minimizing the objective function with respect to the parameters of the model to be determined.

Due to the high non-linearity of the objective function, the FWI might converges to local minima producing unreal models of seismic velocities. The absence of low frequency content in the observed data is also one of the reasons for the objective function may converges to local minima. Some assumptions are made try to overcome this problem, for example, to use an initial model

relatively close to the true model (Shin & Cha, 2008), or adopting the multiscale approach (Bunks et al., 1995), using different frequencies bands. In the multiscale approach using frequencies bands, a fixed number of iterations is adopted for each frequency band and the inversion is performed until the entire frequency spectrum of the observed data is scanned (Bunks et al., 1995). The multiscale approach in this article has been described by Boonyasirawat et al. (2009) using the Wiener filter. Besides that, the multiscale approach was performed using the SVD filter trace by trace to overcome the local minimum problem. The big picture using the SVD filter trace by trace method is, basically, create a matrix with shifted traces using the entire trace of the observed data. Hence, the SVD decomposition is performed in this matrix creating from low to high frequency traces containing the entire frequency band of the data. Another multiscale approach also was performed. This multiscale approach uses damping filters (Chen et al., 2015) on the objective function (MDFOF) and recover the shallower parts of the model until it reaches the deeper parts. Performing the inversion through this way the error on the shallow part is not carried over to the deeper parts, providing a more stable inversion.

Although avoiding convergence to the local minima the multiscale approach did not perform very well, due to geometrical spreading, at the deeper regions of the model; to avoid this problem, the gradient precoditioning was done using the main diagonal of the pseudo-Hessian (Dai & Chen, 2016).

The gradient preconditioning is one of most important factor determining the convergence rate of the non-linear inversion. The precondition factor can remove the effect of geometric spreading from source position to the deep part of model, rebalance the deep and shallow scatterers contribution to gradient, and make the inversion converge faster. Shin et al. (2001) proposed a virtual-source preconditioning, which is constituted by the reverse of the diagonal of the pseudo-Hessian matrix. Shin et al. (2001) applied virtual-source preconditioning in Laplace domain and Laplace-Fourier domain FWI, and achieved a better result. This preconditioning was tested in FWI gradient (Dai & Chen, 2016).

## THEORY

### Full Waveform Inversion

The waveform inversion is a minimization problem between observed data (seismograms) and the data synthesized from a model. The synthetic data is generated from a seismic velocity model and this model is updated iteratively as long as the misfit function of the current iteration is necessarily smaller than the last iteration (Wolfe, 1969). The misfit function,  $L^2$  norm, is written

as follow:

$$E(\mathbf{m}) = \frac{1}{2} \|\mathbf{d} - \mathbf{G}\mathbf{m}\|^2, \quad (1)$$

$\mathbf{d}$  is the observed data,  $\mathbf{G}\mathbf{m}$  is the synthetic data and  $E$  is the misfit function. The Taylor series expansion of the objective function around the neighbourhood of  $\mathbf{m}_0$ , we get:

$$\begin{aligned} E(\mathbf{m}) &= E(\mathbf{m}_0 + \Delta\mathbf{m}) \\ &= E(\mathbf{m}_0) + \sum_{i=1}^M \frac{\partial E(\mathbf{m}_0)}{\partial m_i} \Delta m_i \\ &+ \frac{1}{2} \sum_{i=1}^M \sum_{j=1}^M \frac{\partial^2 E(\mathbf{m}_0)}{\partial m_i \partial m_j} \Delta m_i \Delta m_j + \mathcal{O}(\|\mathbf{m}^3\|). \end{aligned} \quad (2)$$

Taking the derivative with respect to the parameter  $m_i$  we obtain:

$$\begin{aligned} \frac{\partial E(\mathbf{m})}{\partial m_i} &= \frac{\partial E(\mathbf{m}_0)}{\partial m_i} \\ &+ \sum_{j=1}^M \frac{\partial^2 E(\mathbf{m}_0)}{\partial m_j \partial m_i} \Delta m_j, \quad i = 1, 2, \dots, M \end{aligned} \quad (3)$$

rewriting:

$$\frac{\partial E(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}_0} + \frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2} \Delta \mathbf{m}, \quad (4)$$

imposing that the objective function is minimal, we can determine  $\Delta \mathbf{m}$

$$\begin{aligned} \Delta \mathbf{m} &= - \left[ \frac{\partial^2 E(\mathbf{m}_0)}{\partial \mathbf{m}^2} \right]^{-1} \frac{\partial E(\mathbf{m}_0)}{\partial \mathbf{m}} \\ &= -\mathbf{H}^{-1} \nabla E_m. \end{aligned} \quad (5)$$

Where  $\nabla E_m$  is the gradient vector and can be represented by:

$$\nabla E_m = \left[ \frac{\partial E(\mathbf{m}_0)}{\partial m_1}, \frac{\partial E(\mathbf{m}_0)}{\partial m_2}, \dots, \frac{\partial E(\mathbf{m}_0)}{\partial m_M} \right]^T. \quad (6)$$

We have,  $\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta \mathbf{m}$ , therefore the update equation for the model is:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \mathbf{H}^{-1} \nabla E_m, \quad (7)$$

or

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k. \quad (8)$$

Equation (7) would update the model accurately if it was possible to calculate exactly the inverse value of the Hessian. Equation (8) refers to a generic update, considering the  $\mathbf{p}_k$  the search

direction. In the steepest descent method the search direction is the gradient vector, the inverse of the Hessian is approximately the identity and Eq. (7) looks like:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \nabla E_m. \quad (9)$$

### Forward Problem

In the case of constant density, the acoustic wave equation is specified by:

$$\frac{1}{\mathbf{v}^2} \frac{\partial^2 \mathbf{P}}{\partial t^2} = \nabla^2 \mathbf{P} + \mathbf{S}(x, t), \quad (10)$$

where  $\mathbf{P}$  is the pressure field and  $\mathbf{S}$  is the source,  $\mathbf{v}$  is the velocity field. Solving the acoustic wave equation using the finite difference method, iteratively:

$$\mathbf{P}^{n+1} = 2\mathbf{P}^n - \mathbf{P}^{n-1} + c^2 \Delta t^2 \nabla^2 \mathbf{P} + \mathbf{S}(x, t), \quad (11)$$

The time discretization<sup>3</sup> was 2nd order and in space was 8th order.

### Numerical Stability Criteria and Numerical Dispersion

Whenever the medium is discretized by the Eq. (11) two problems arising from this discretization of the continuous problem: instability and numerical dispersion. The solution should be unstable numerically when its values grows indefinitely, generating spurious results (dos Santos, 2013). To avoid instability  $\Delta t$  needs to satisfy the equation below:

$$\Delta t \leq \frac{1}{\mathbf{v} \sqrt{\frac{1}{\Delta x^2} + \frac{1}{\Delta z^2}}}, \quad (12)$$

numerical dispersion occurs when the phase velocity of the numerical wave is different from the propagation velocity of the medium, damaging the results obtained by numerical modelling.

This problem arises due to the spatial discretization of the wave equation, therefore the phase and group velocities become a function of the spacing between the points of the mesh, the frequency and the angle of propagation (Araújo, 2009). To overcome this difficulty, a relation must be respected which depends on the frequency source, the mesh and the velocity field, which is described below,

$$f_{\max} \leq \frac{\min(\mathbf{v})}{F \max(\Delta z, \Delta x)}, \quad (13)$$

where  $F$  depends on the finite difference operator in space. For example if the operator is of 2nd order  $F = 10$ , if the operator is 8th order  $F = 2$ . In the Fourier domain  $F = 2$  always, since in the Fourier method the discretization is more precise.

<sup>3</sup>The  $n$  index is related to the time discretization.

### Gradient Computation

The adjoint method (Plessix, 2006) computes the gradient from a reverse time migration (RTM) from the residue between the calculated data and the observed data. The gradient computation is done by the following equation:

$$\nabla E_{\mathbf{m}} = \frac{2}{\mathbf{v}^3(x)} \sum_{r=1}^{nr} \sum_{s=1}^{ns} \int_0^T \frac{\partial \mathbf{P}_{calc}^2(\mathbf{x}, t; \mathbf{x}_s)}{\partial t^2} \mathbf{P}_{res}(\mathbf{x}_r, t; \mathbf{x}_s) dt, \quad (14)$$

where  $ns$  represents the number of shots,  $nr$  represents the number of receivers/geophones and  $\mathbf{v}$  is the seismic velocity field,  $\mathbf{d}_{res}$  is residue backward propagation. For each FWI iteration, the gradient is computed cross correlating the second derivative of the calculated and data with the residue for all times therefore sums up for all shots and for all receivers. The misfit function is also done in the same way,

$$E(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{nr} \sum_{s=1}^{ns} \int_0^T |\mathbf{P}_{calc} - \mathbf{P}_{obs}|^2 dt. \quad (15)$$

The residue backward propagation is obtained from the wave equation, using as source, the residue between the synthetic data and the observed data and can be obtained by doing the reverse modelling of the residue, represented by the following equation:

$$\frac{1}{\mathbf{v}^2} \frac{\partial^2 \mathbf{P}_{res}}{\partial t^2} = \nabla^2 \mathbf{P}_{res} + \overbrace{(\mathbf{P}_{calc} - \mathbf{P}_{obs})}^{residue}. \quad (16)$$

### Backtracking Line Search Method

Somehow whether the objective function of the current iteration is higher than the previous one the method enters into a line search method in order to allow FWI to converge to the global minimum. This method works decreasing the step length and keeping the current model and the search direction constant. The line search method finishes if the objective function is reduced compared to the previous one. The pseudo code shows how the line search method works. In this paper the stability of the FWI was measured by the number of iterations of the backtracking line search method.

Algorithm 1: Backtracking line search method.

```

1 begin
2    $\alpha \leftarrow \bar{\alpha}, \tau \epsilon(0, 1)$ 
3   if
4      $E(\mathbf{m}_k + \alpha_k \mathbf{p}_k) \leq E(\mathbf{m}_k) + c_1 \alpha \nabla E(\mathbf{m}_k^T) \mathbf{p}_k$ 
5     then
6        $\alpha_k = \alpha$ 
7     else
8        $\alpha \leftarrow \tau \alpha$ 
9   end
10 return  $\alpha_k = \alpha$ 

```

### MULTISCALE APPROACH

The presence of local minima at all scales in the waveform inversion prevents a reasonable convergence to the global minimum (Bunks et al., 1995). The multiscale approach decomposes the waveform inversion problem into scales. At large scales there are fewer local minima, so for each iteration the problem is taken from the large scales to the small scales then the final result is closer to the global minimum.

Convergence to local minima is basically due to many problems but the mainly is: the choice of the initial velocity model. A good initial velocity model close to the true model, makes the FWI converge to the global minimum. The problem in the choice of the initial model lies in the cycle skipping problem (Virieux & Operto, 2009); the signals propagated in subsurface are at times different from the modelled data. When the time lag between the model data is greater than half of the period, the model data cycle  $n$  will fit with the  $(n - 1)$  observed data cycle generating problems in the velocity model, also another problem is the presence of high frequencies in the wavelet components.

In time domain, one of each multiscale approach is performed by the selection of a frequency band, which must be done on the data and on the source. The choice of a frequency band is related to how much detail should be recovered. It is necessary, strictly, the frequency band of the source be equal to the band of the data. Several filters can be used to select the frequency band for the data and for the sources, the one used in this paper was the Boonyasiriwat et al. (2009) filter. The equation that defines the filter is:

$$f_{Wiener}(\omega) = \frac{W_{target}(\omega) W_{original}^\dagger(\omega)}{|W_{original}|^2 + \epsilon^2}, \quad (17)$$

where  $f_{Wiener}$  is the filter that will be applied on the data,  $\omega$  is the temporal frequency,  $W_{original}$  is the wavelet of the data,  $W_{target}$  is the wavelet with limited frequency content,  $\dagger$  is the multiscale approach operator of complex conjugate and  $\epsilon$  is a small number used to stabilize the result. Boonyasiriwat et al. (2009) showed how the Wiener filter can obtain better results than conventional filters, such as Hanning, when the original wavelet is known.

### SVD Filter Trace by Trace in Multiscale Approach

SVD filtering was also used to filter the source and data from the low frequencies to the high frequencies. The SVD decomposition methodology is simple and uses the standard approach of SVD decomposition. Considering  $\mathbf{d} = (d_1 \dots d_M)^T$  the seismic traces and the matrix of the shifted traces  $\mathbf{D}_N$ , it is possible to construct a matrix with the seismic shifted traces (Silva et al., 2016) in time of  $\tau = 0, \dots, N - 1$ , defined below.

$$\mathbf{D}_N = \begin{bmatrix} \mathbf{d} & 0 & \dots & 0 \\ 0 & \mathbf{d} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & \mathbf{d}_s \end{bmatrix} \quad (18)$$

using SVD decomposition (Golub & Van Loan, 2012; Silva et al., 2016), we obtain:

$$\mathbf{D}_N = \sum_{\tau=0}^{N-1} \sigma_{\tau} \mathbf{u}_{\tau} \mathbf{v}_{\tau}^T = \sum_{\tau=0}^{N-1} \mathbf{D}_{N\tau}, \quad (19)$$

### Multiscale Approach using Damping Filters on the Objective Function (MDFOF)

The wave field recorded at the receivers is basically influenced by the diving waves or by the first reflections on the surface of the velocity model (Klimm, 2013). The wave field recorded at the receivers for long travel times correspond to the reflections in the deeper parts of the model or even multiple reflections. The problems occurred in the superficial reflections are carried over to the deeper parts, causing a barely update of the velocity model in the deeper parts. So it is reasonable to update first the shallower parts of the model and later the deeper parts, in this case the overall update is better than the update performed globally (Klimm, 2013). One way to do this is to set up an attenuation factor on the gradient (Klimm, 2013):

$$\nabla E_{\mathbf{m}} = \frac{2}{v^3(x)} \sum_{r=1}^{nr} \sum_{s=1}^{ns} \int_0^T \frac{\partial d_{calc}^2(\mathbf{x}, t; \mathbf{x}_s)}{\partial t^2} d_{res}(\mathbf{x}_r, t; \mathbf{x}_s) \exp^{-kt} dt, \quad (21)$$

where  $k$  is the attenuation parameter. The problem is solved starting from high values for  $k$  until it reaches zero. With large values the model is updated only on the shallower parts and throughout the iterations  $k$  reduces until becomes zero allowing the complete update. This method using an attenuation parameter is called inversion in the Laplace domain, since the Laplace transform of  $p$  is (Klimm, 2013), but does not require the low frequency information in the seismic data:

$$\mathbf{p}(k) = \int_0^T \mathbf{p}(t) \exp^{-kt} dt. \quad (22)$$

the matrix  $\hat{\mathbf{D}}$  represent the eigenimage of the trace. To obtain the original trace is necessary to use a linear operator  $\mathbf{J}$  will recover the original trace decomposed in your eigenimage,

$$\mathbf{J} \{ \mathbf{D}_N \} = \sum_{\tau=0}^{N-1} \mathbf{J} [ \hat{\mathbf{D}}_{N\tau} ] = \sum_{\tau=0}^{N-1} \hat{\mathbf{d}}_{N\tau} = \mathbf{d}. \quad (20)$$

---

#### Algorithm 2: SVD decomposition trace by trace.

---

```

1 begin
2    $\mathbf{d}_0 \leftarrow \mathbf{d}$ 
3   for  $\tau = 1, N - 1$ 
4     do
5        $\mathbf{d}_{\tau-1} \leftarrow \mathbf{D}_{\tau}$ 
6       computes the first eigenimage  $\hat{\mathbf{D}}_0 \leftarrow \sigma_0 \mathbf{u}_0 \mathbf{v}_0^T$ 
7       computes the first eigentrace  $\hat{\mathbf{D}}_{\tau 0} \leftarrow \mathbf{J} [ \hat{\mathbf{D}}_{\tau 0} ]$ 
8       update the low frequency components
9        $\mathbf{d}_{\tau} \leftarrow \hat{\mathbf{D}}_{\tau 0}$ 
9     end
10  end
11  return Low frequency trace  $\mathbf{d}_{N-1}$ 
12  return
    High frequency trace  $\hat{\mathbf{d}}_{N-1} \leftarrow \mathbf{d} - \mathbf{d}_{N-1}$ 

```

---

In order to explain better how the method works, one-shot filtering using the SVD filter is shown in Figure 1. The decomposition a trace of the seismogram is represented in Figure 2 and the amplitude spectra of the low frequency component, the original data and the high frequency component Figure 3.

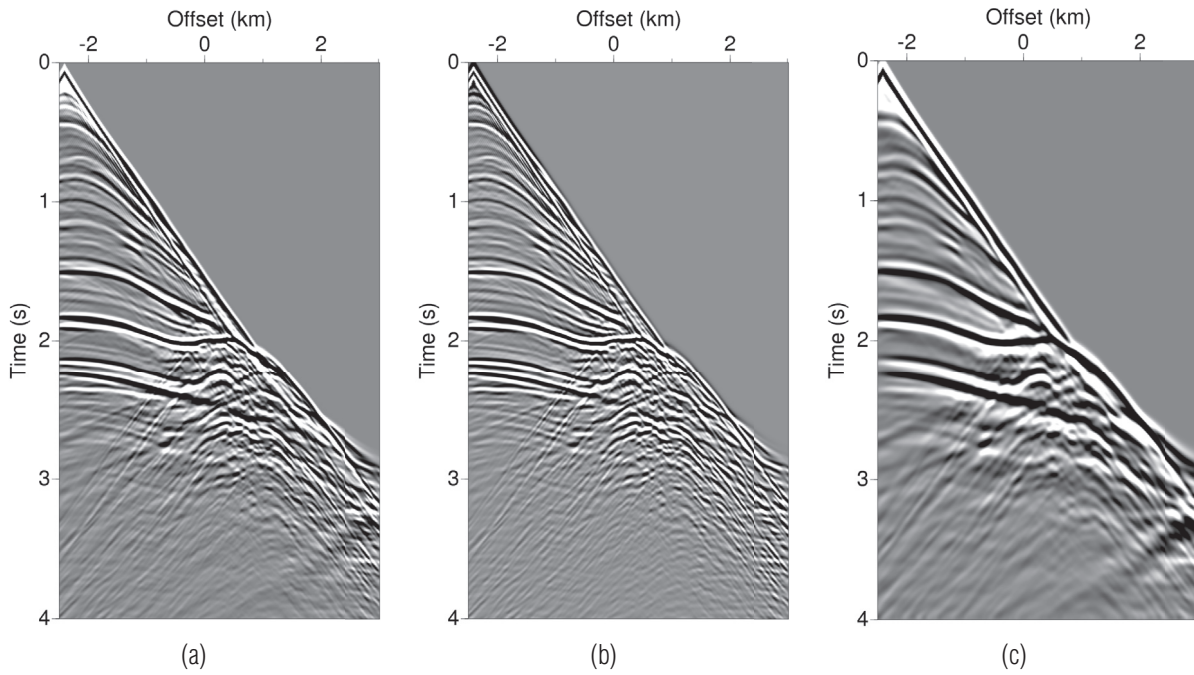


Figure 1 – (a) original shot, (b) low frequency from the original shot, (c) high frequency from the original shot.

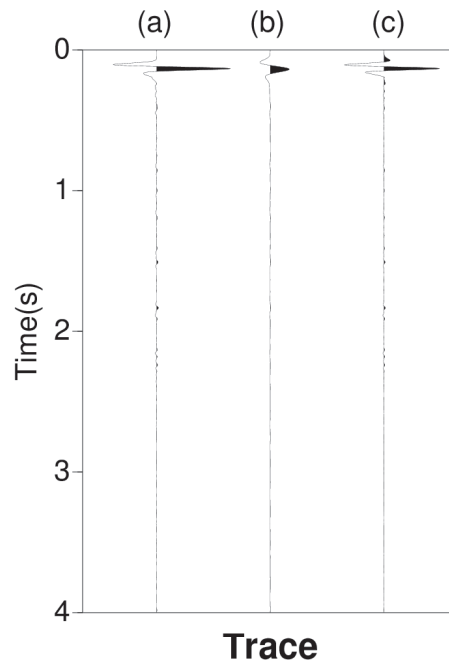
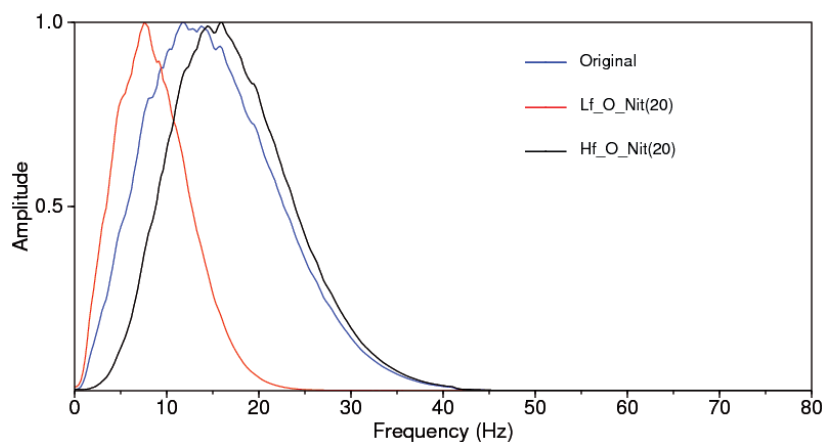


Figure 2 – (a) original trace, (b) low frequency component, (c) high frequency component.

Another approach using time-attenuation filters can be applied over the objective function (Chen et al., 2015):

$$E(\mathbf{m}) = \frac{1}{2} \sum_{r=1}^{nr} \sum_{s=1}^{ns} \int_0^T |\mathbf{P}_{calc} - \mathbf{P}_{obs}|^2 \exp^{-kt} dt. \tag{23}$$





**Figure 3** – The frequency spectrum of the original data and the decompositions in low and high frequency.

These two approaches try to reduce the possibility to the convergence a local minimum. First of all, as mentioned, the error in updating shallow parts will not be carried over to the deeper parts and in addition the set of parameters to be inverted will be smaller, due to the attenuation parameter increasing the immunity with respect to the local minimum (Chen et al., 2015). In this paper the second approach, using the decay factor in the objective function, was adopted.

### Virtual-Source in the Preconditioning of the Gradient

The gradient preconditioning is one of the most important factors determining the convergence rate of the non-linear inversion. The precondition factor can remove the effect of geometric spreading from source to the deep part of model, rebalance the deep and shallow scatterers contribution to gradient, and make the inversion converge faster (Dai & Chen, 2016). Shin et al. (2001) proposed virtual-source precondition, which is constituted by the reverse of the diagonal of the pseudo-Hessian matrix. For this, they considered the use of the virtual-source in the calculation of the pseudo-Hessian. Shin et al. (2001) estimated the virtual-source using the time domain wave equation. In the Laplace domain, the virtual-source can be calculated as (Dai & Chen, 2016):

$$\mathbf{f} = -\frac{\partial \mathbf{Z}}{\partial \mathbf{v}(z, x)} \mathbf{p}, \quad (24)$$

where  $\mathbf{Z}$  is the impedance matrix,  $\mathbf{v}$  is the velocity field and  $\mathbf{p}$  is the wave field. The pseudo-Hessian matrix for the Gauss-Newton method is:

$$\mathbf{H} \approx \mathbf{f}^T \mathbf{f}, \quad (25)$$

in the time domain the virtual-source preconditioning is found by the derivative of the wave equation with respect to the velocity

(Dai & Chen, 2016):

$$\mathbf{f} = \frac{\partial \left( \frac{1}{\mathbf{v}^2} \frac{\partial^2 \mathbf{P}}{\partial t^2} - \nabla^2 \mathbf{P} - \mathbf{S}(x, t) \right)}{\partial \mathbf{v}}, \quad (26)$$

considering that the wave field does not vary with velocity, we find the virtual-source equation:

$$\mathbf{f} = \frac{2}{\mathbf{v}^3} \frac{\partial^2 \mathbf{P}}{\partial t^2}, \quad (27)$$

then the Hessian's diagonal in the time domain is written as follows:

$$\mathbf{H} = \left( \frac{2}{\mathbf{v}^3} \frac{\partial^2 \mathbf{P}}{\partial t^2} \right)^2, \quad (28)$$

summing for all over shots:

$$\mathbf{H}^* \approx \sum_{i=1}^{ns} \left( \frac{2}{\mathbf{v}^3} \frac{\partial^2 \mathbf{P}}{\partial t^2} \right)^2 + \beta, \quad (29)$$

where  $\beta$  is a stabilizing factor. The preconditioned gradient using the pseudo-Hessian can be written by:

$$\nabla E(\mathbf{m})^* = \frac{\frac{2}{\mathbf{v}^3(x, z)} \sum_{r=1}^{nr} \sum_{s=1}^{ns} \sum_{t=0}^T \frac{\partial \mathbf{d}_{calc}^2(\mathbf{x}, t; \mathbf{x}_s)}{\partial t^2} \mathbf{d}_{res}}{\sum_{i=1}^{ns} \sum_{t=0}^T \mathbf{H}^* + \beta}, \quad (30)$$

### NUMERICAL EXAMPLE

The comparison between the multiscale approaches was done initially by comparing the Wiener filter and the SVD filter trace by trace. The velocity model used to the FWI was the Marmousi which try to simulate a complex geological structure. The inversion input parameters are in the Table 1.

**Table 1** – Input inversion parameters of the Marmousi model in multiscale approach.

Input	Value
dt (sampling interval)	0.001 s
nz (depth sample)	375
nx	369
Number of shots	62
nt	3201
dx	25 m
dz	8 m
Distance between shots	6 m
Number of border points	40
Iterations	100
Initial frequency	3 Hz
Interval between frequencies	3 Hz
Number of iterations per frequency	20

In the multiscale approach using the SVD filter trace by trace there is no interval between frequencies, being valid only for the multiscale approach using the Wiener filter. The frequency band used in each method was compared, therefore is a fair comparison. The initial model and the inverted model using the Wiener filter and SVD filter is represented in the Figure 4.

The Wiener filter showed to be little better in the deeper regions of the model compared to the new SVD filter trace by trace. This is going to be more evident in the seismic velocity profile. The comparison between the objective functions, Figure 5, using the two filters are represented in 5 figures since every 20 iterations the FWI changes the frequency band, therefore a new inverse problem. The normalized error of the model was defined to estimate which of the two methods is the best. The equation that defines it is:

$$NRMS = \frac{\sqrt{\sum_{i=1}^n (\mathbf{m}^i - \mathbf{m}_{est}^i)^2}}{\max(\mathbf{m}) - \min(\mathbf{m})}, \quad (31)$$

the comparison of two filters using a velocity profile is in the Figure 6.

The number of iterations of the backtracking line search method, showed that the SVD filter was more stable in the first iterations than Wiener filter and the last iterations the Wiener filter proved to be better as shown in the Figure 7. The comparison

between the deeper parts of the velocity model was damaged due to geometrical spreading, so the preconditioning of the main diagonal of the pseudo-Hessian improved the deeper regions of the model represented by the Figure 9, and the velocity profile on the deeper regions became better and showed the supremacy of the Wiener filter. The model error with the preconditioning showed the superiority of the Wiener filter in relation to the SVD filter even in the deeper regions of the model as presented in Figure 11.

### Multiscale Approach in Data Domain × Multiscale Approach using Damping Filters on the Objective Function (MDFOF)

The same parameters were adopted in the two multiscale approach comparison, as well as the initial velocity field and the true model field. The multiscale in data domain was compared using the two filters: Wiener and SVD trace by trace and also was compared to multiscale approach with damping filters in objective function. The estimated model was in Figure 12. The MDFOF results were worse than the multiscale approach results in data domain, because the immunity to the local minima is greater when the multiscale is performed on data domain. The model error prove multiscale in data domain is better than MDFOF showed in the Figure 13. A velocity profile comparing the results show the supremacy of the multiscale approach in data domain using represented by the Figure 14. The result shown in the profile only assure the benefits using multiscale approach in the data domain.

## CONCLUSION

The full waveform inversion using the SVD filter and Wiener filter produced good results, however the Wiener filter showed up more efficient in terms of the achieved model. Although in terms of stability the SVD filter was more stable on the first iterations than the Wiener filter. The gradient preconditioning is essential to show the filters work even in deeper parts of the model. Comparing multiscale approaches in the data domain and MDFOF, the multiscale approach in data domain produced a better estimated model in comparison with MDFOF, but MDFOF was more stable than multiscale approach in data domain.

## ACKNOWLEDGEMENTS

This research was supported by CNPq and INCT-GP/CNPq. The facility support from CPGG/UFBA is also acknowledged.



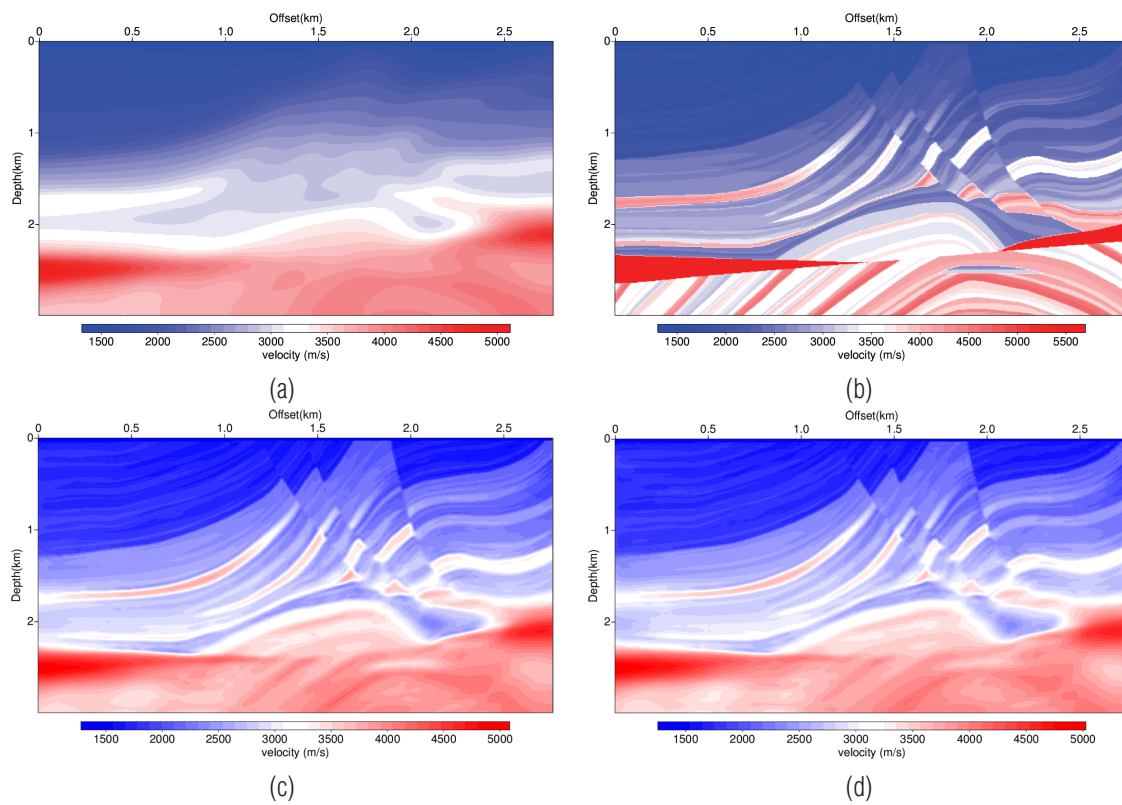


Figure 4 – (a) initial model, (b) true velocity model, (c) estimated model by Wiener filter, (d) estimated model using SVD filter.

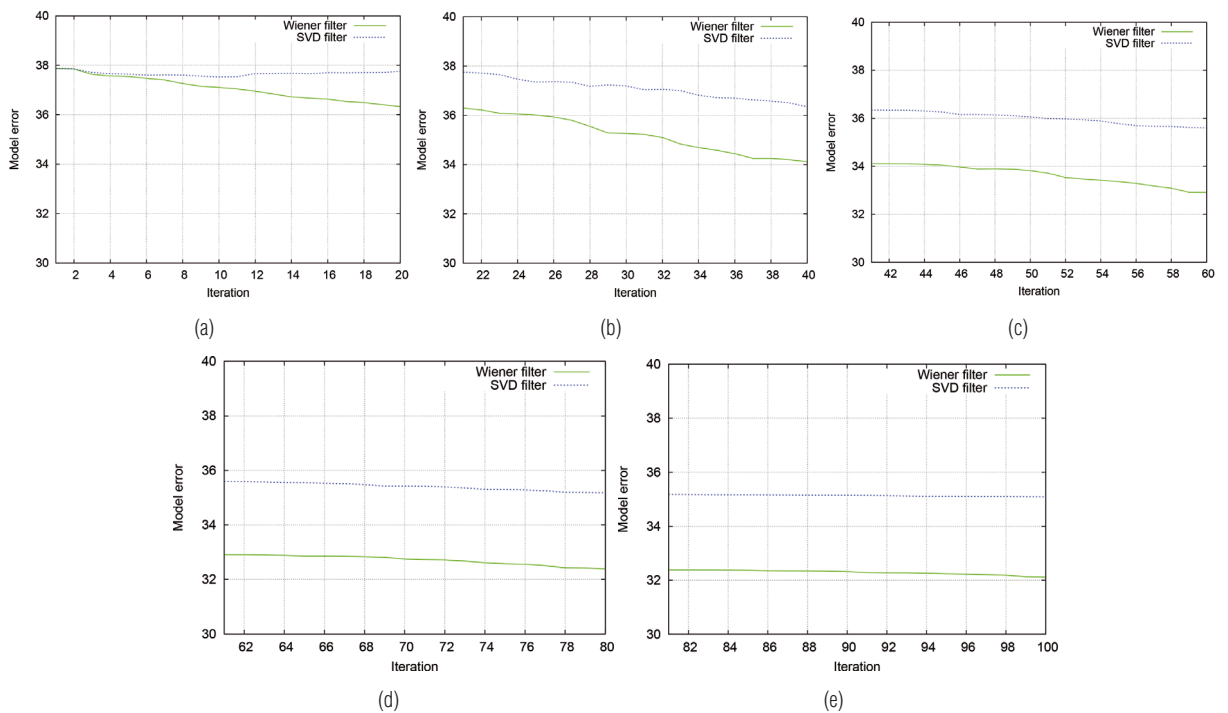
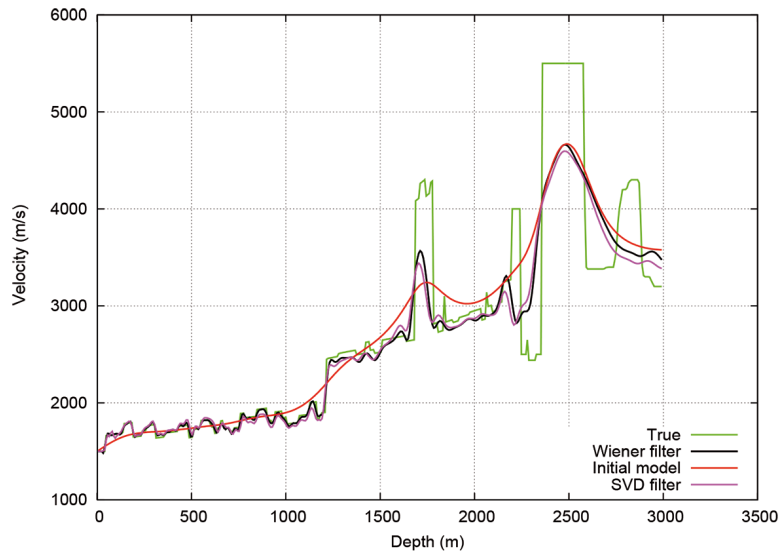
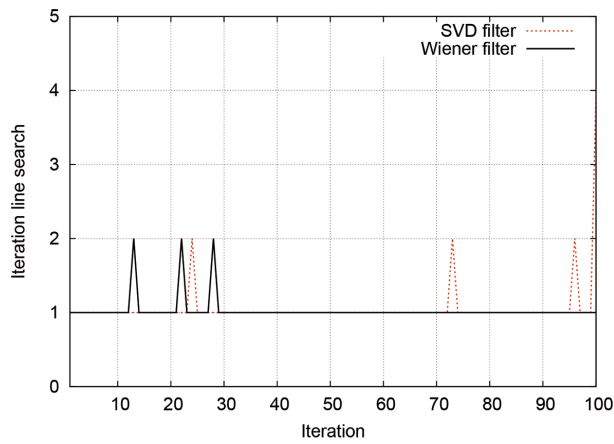


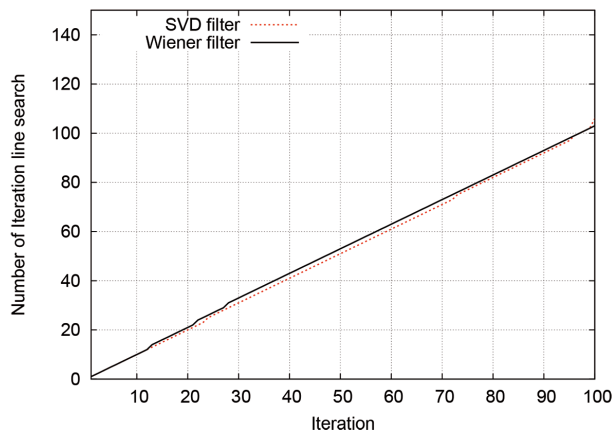
Figure 5 – The behavior of the misfit function comparing both SVD and Wiener filters.



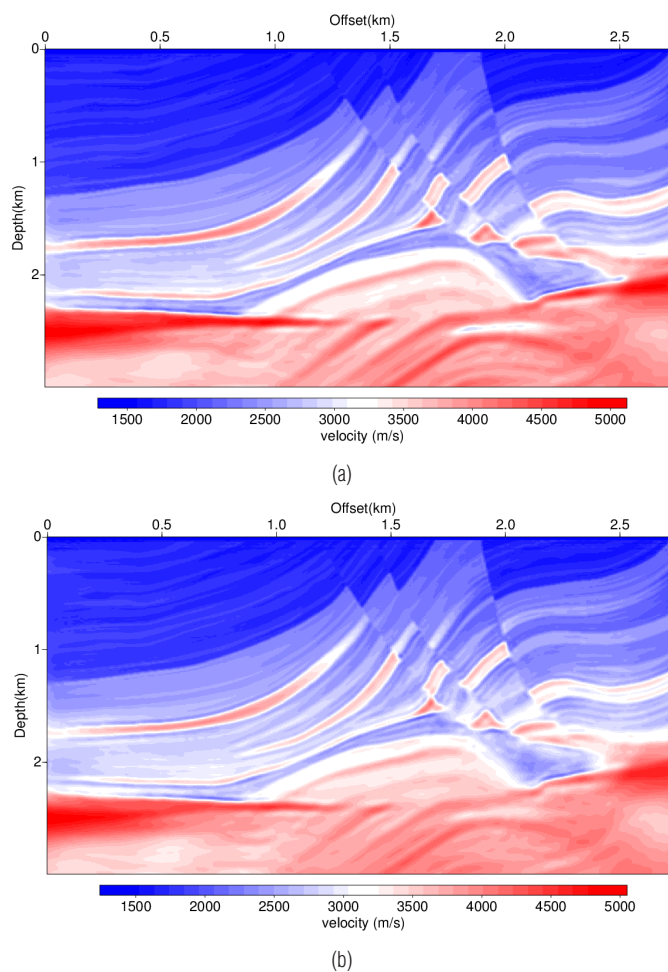
**Figure 6** — Velocity profile found using the two filters: SVD and Wiener. The region of the high depths closely approximates to the initial data because of the geometrical spreading. The profile shows that the Wiener filter is closer to true.



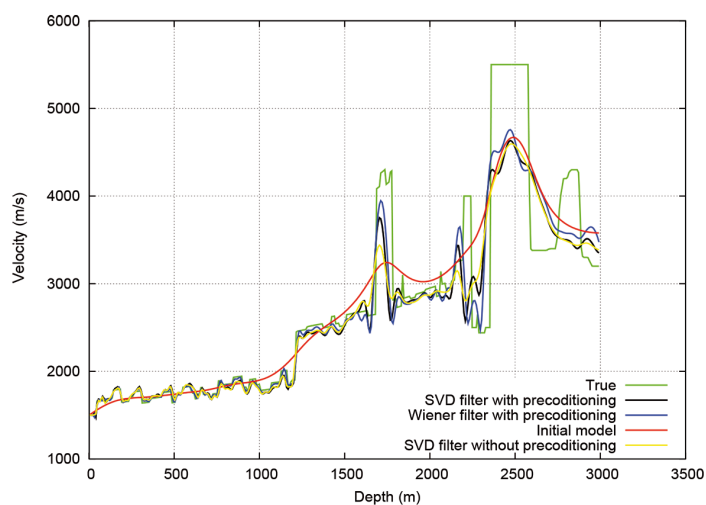
**Figure 7** — Number of iterations of the line search method by regression comparing the two filters in the multiscale approach.



**Figure 8** — Number of iterations of the line search method stacked. This figure shows SVD filter is better in the first iterations and get worse in the last iterations.



**Figure 9** — (a) preconditioned velocity field using Wiener filter, (b) preconditioned velocity field using SVD filter. It is evident that the anomaly of the syncline became more prominent after preconditioning. The Wiener filter shows better.



**Figure 10** — Velocity profile found using the two filters: SVD and Wiener. The region of the high depths closely approximates to the initial data because of the geometrical spreading. The problem was overcome using the preconditioning diagonal Hessian.

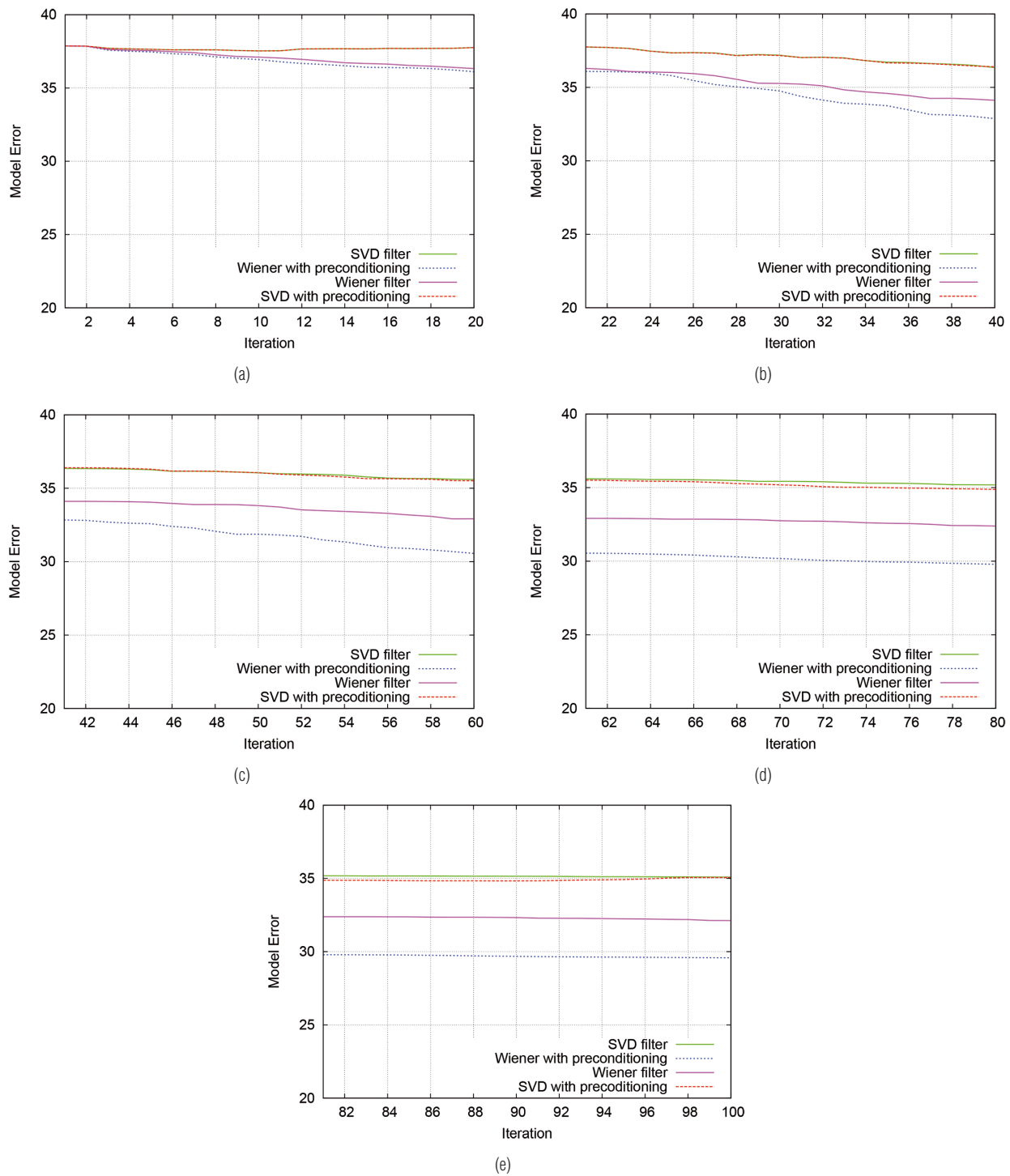
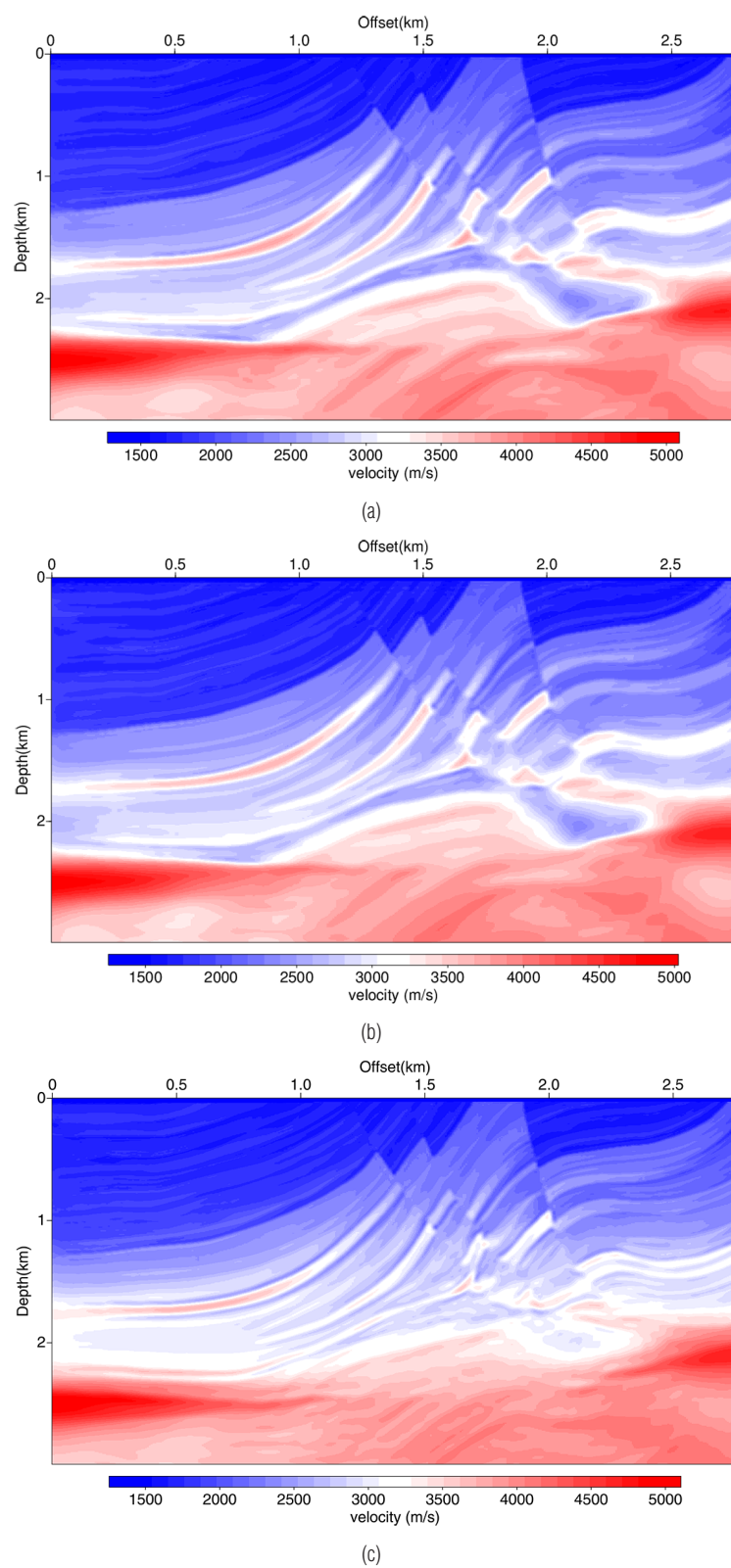


Figure 11 – The behavior of the misfit function comparing both SVD and Wiener preconditioned filters.



**Figure 12** — (a) estimated model by Wiener filter, (b) estimated model by SVD filter, (c) estimated model by MDFOF.

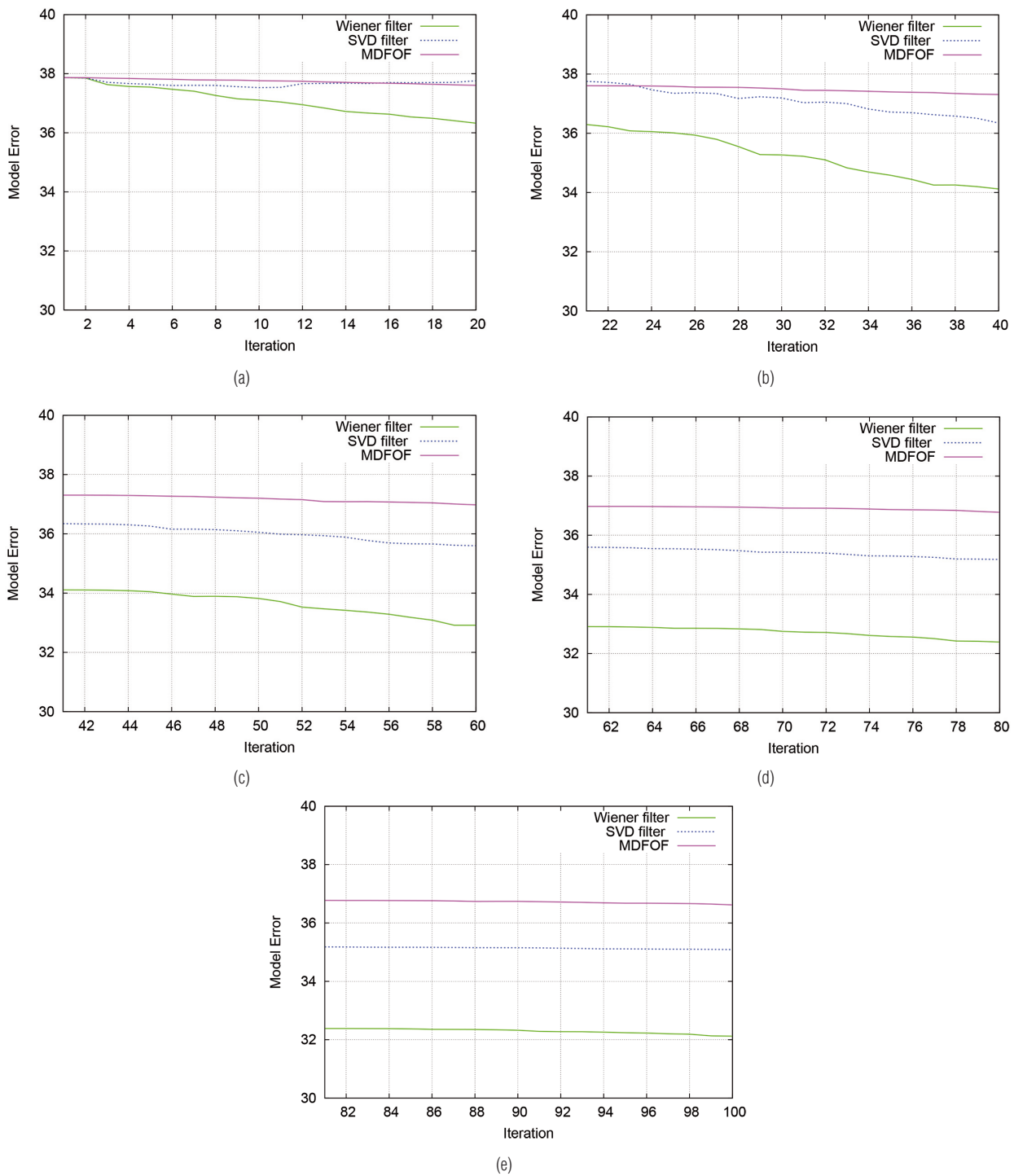


Figure 13 – Model error MDFOF  $\times$  Multiscale approach in data domain.



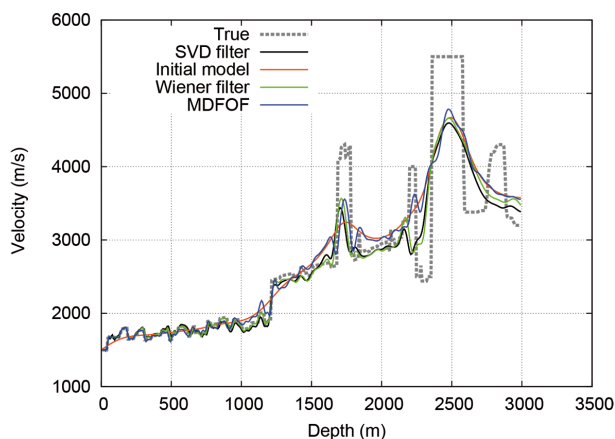


Figure 14 – Velocity profile found using the two multiscale approach.

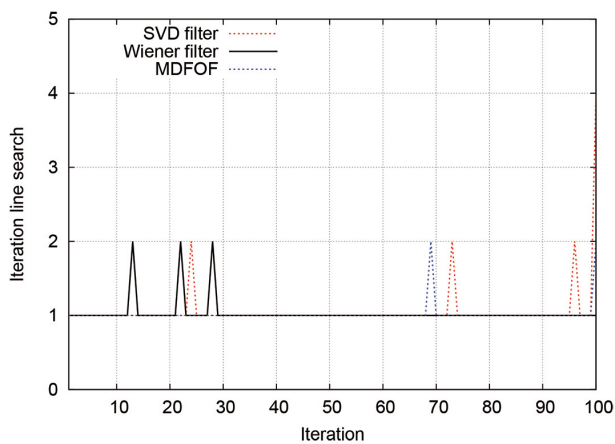


Figure 15 – Number of iterations of the line search method by regression comparing the two multiscale approach.

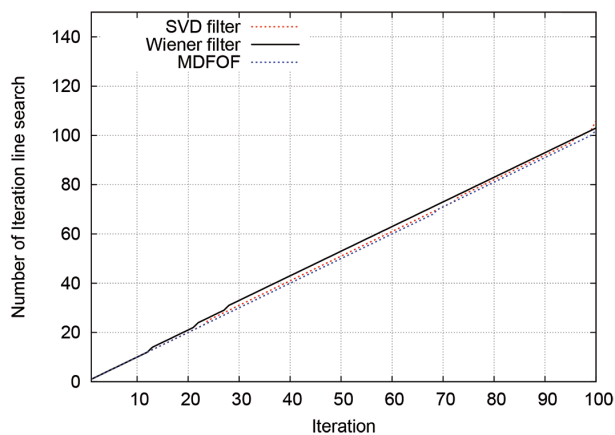


Figure 16 – Number of iterations of the line search method stacked. This figure shows greater MDFOF stability.

## REFERENCES

- ARAÚJO ESD. 2009. Analysis of finite difference methods and rapid expansion method in reverse time migration. Master's thesis, Universidade Federal da Bahia, Brazil. 48 pp.
- BOONYASIRIWAT C, VALASEK P, ROUTH P, CAO W, SCHUSTER GT & MACY B. 2009. An efficient multiscale method for time-domain waveform tomography. *Geophysics*, 74: WCC59–WCC68.
- BUNKS C, SALECK FM, ZALESKI S & CHAVENT G. 1995. Multiscale seismic waveform inversion. *Geophysics*, 60: 1457–1473.
- CHEN G, CHEN S & WU R-S. 2015. Full waveform inversion in time domain using time-damping filters. In: SEG Technical Program Expanded Abstracts 2015. Society of Exploration Geophysicists. p. 1451–1455.
- DAI M & CHEN J. 2016. Studies on multi-scale full waveform inversion for time-domain acoustic wave equation based on virtual-source precondition. In: International Conference on Environmental and Engineering Geophysics, 7., & Summit Forum of Chinese Academy of Engineering on Engineering Science and Technology. p. 269–272.
- DOS SANTOS AWG. 2013. Waveform inversion applied to the analysis of seismic velocities using a multiscale approach. Master's thesis, Universidade Federal da Bahia, Brazil. 144 pp.
- GOLUB GH & VAN LOAN CF. 2012. Matrix computations. Volume 3: JHU Press. 687 pp.
- KLIMM B. 2013. Time domain full waveform inversion using ADI modeling. PhD thesis, Technische Universität Kaiserslautern. 82 pp.
- LAMBARÉ G, HERRMANN P, TOURÉ J-P, CAPAR L, GUILLAUME P & BOUSQUIÉ N. 2007. From time to depth imaging: an accurate workflow. In: SEG Annual Meeting. 2007, 5 pp.
- PLESSIX R-E. 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167: 495–503.
- SHIN C & CHA YH. 2008. Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173: 922–931.
- SHIN C, JANG S & MIN D-J. 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49: 592–606.
- SILVA M, PORSANI M & URSIN B. 2016. A single-trace singular-value decomposition method with application to the ground-roll removal. In: SEG Technical Program Expanded Abstracts 2016. Society of Exploration Geophysicists. p. 4659–4663.
- STORK C. 1992. Reflection tomography in the postmigrated domain. *Geophysics*, 57: 680–692.
- TARANTOLA A. 1984. Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32: 998–1015.
- VIRIEUX J & OPERTO S. 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74: WCC1–WCC26.
- WOLFE P. 1969. Convergence conditions for ascent methods. *SIAM Review*, 11: 226–235.
- YILMAZ Ö. 2001. Seismic data analysis. Volume 1. Society of Exploration Geophysicists, Tulsa. 995 pp.