Full waveform inversion - a synthetic test using the PSPI migration

Marcelo Guarido[*], Laurence Lines[*], Robert Ferguson

ABSTRACT

Full Waveform Inversion (FWI) is a tool for the estimation of P-wave velocity. We demonstrate, for a 2D acoustic marine synthetic case, that the methodology can be applied with reasonable results, using the PSPI migration with a deconvolution imaging condition and a line search to calculate the step length to invert P-wave velocity. We test different mutes applied in the migrated residuals, and we achieve a high resolution result with what we demonstrate is a moderate mute. We analyze the importance of the initial velocity model for convergence and we obtain good responses for velocity models with smoothed flat layers based on simulated wells in the real velocity model. We find that, using only one well to generate the initial model, the best answer in when that well does not sample a high velocity body in the model. Our results suggest that we can have a high resolution inversion if an initial model is created by an interpolation of two or more wells. The test in the Marmousi 2D model shows encouraging results but more tests are necessary.

INTRODUCTION

Full waveform inversion, or FWI, is an inverse method that uses the least-square technique to minimize the objective function (the difference between a real seismic data and a synthetic seismic data) by updating the initial parameters model until convergence (Margrave et al., 2011). The update is obtained from a reverse-time migration of the data residual.

The full waveform inversion was proposed in the early 80's (Pratt et al., 1998)) but the technique was considered too expensive in computational terms. Lailly (1983) and Tarantola (1984) simplified the methodology by using the steepest-descent method (or gradient method) in the time domain to minimize the objective function without calculate, explicitly, the partial derivatives. They compute the gradient of the misfit between real and synthetic data by pre-stack migration of the data residuals with reverse-time migration (RTM). Pratt et al. (1998) develop a matrix formulation for the full waveform inversion in the frequency domain and show how to compute the inverse of the Hessian matrix (the step length for convergence in the FWI) by calculating the full-Hessian (Gauss-Newton method) or an approximation of the Hessian (quasi-Newton). The Hessian computation is one of the study areas for the FWI improvement, as it is a large matrix and its computation requires lots of computer power. An overview of the FWI theory and studies are compiled by Virieux and Operto (2009). Margrave et al. (2010) use well control to calibrate the model update (known as *step length*) and pre-stack wave-equation migration (Ferguson and Margrave, 2005) instead of the RTM, so the iterations are done in time domain but only selected frequency bands are migrated, using a deconvolution imaging condition (Margrave et al., 2011; Wenyong et al., 2013) as a better reflectivity estimation. Warner and Guasch (2014)

[*]CREWES - University of Calgary

use the deviation of the Weiner filters of the real and estimated data as the object function with great results.

In this work we investigate the FWI in a similar methodology proposed by Margrave et al. (2010), but not using well control as a model calibrator. The velocity residual is calculated using the wave-equation migration with a deconvolution image condition, so the inversion is done in time domain and only selected frequency bands are migrated. The step length is determined by a line search. The selection of the frequency bands to migrate is also studied in this paper, with the objective to find a better practice.

THEORY

This section presents the basic mathematical calculation of the full waveform inversion, showing how to compute the gradient from the objective function, the matrix representation for the Newton, Gauss-Newton and gradient (steepest descent) methods and the deconvolution imaging conditions for the gradient.

The objective function

The goal of the FWI is to minimize the difference between the acquired data and the synthetic data obtained from a "guessed" model and iteratively update the model. The misfit function, or residual wavefield, is:

$$\delta P(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)}) \equiv P(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega) - G(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)}) \tag{1}$$

where $P(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega)$ is the acquired data, $G(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)})$ is the synthetic data, $s_0^{(n)}(\boldsymbol{r}) \equiv 1/c_{0,n}^2(\boldsymbol{r})$ is the squared-slowness in the *nth* iteration, $c_0^2(\boldsymbol{r})$ is the velocity, $\boldsymbol{r}_g$ and $\boldsymbol{r}_s$ are the receiver and source position vectors respectively, and ω is the angular frequency. On each iteration, we update s_0 until the objective function Φ is minimized:

$$\Phi\left(s_0^{(n)}\right) \equiv \frac{1}{2} \int \left(\sum_{s,g} |\delta P|^2\right) d\omega \tag{2}$$

Margrave et al. (2011) derive the parameter update $\delta s_0^{(n)}(\boldsymbol{r''})$ by calculating the first derivative of $\Phi(s_0^{(n)} + \delta s_0^{(n)})$ and setting it to zero, and the expression is:

$$\delta s_0^{(n)}(\boldsymbol{r''}) = -\int H^{(n)-1}(\boldsymbol{r''}, \boldsymbol{r'}) g^{(n)}(\boldsymbol{r'}) d\boldsymbol{r'} \tag{3}$$

where

$$g^{(n)}(\boldsymbol{r'}) = \frac{\partial \Phi \left(s_0^{(n)} \right)}{\partial s_0^{(n)}(\boldsymbol{r'})} \tag{4}$$

and

$$
\begin{aligned}
g^{(n)}(\boldsymbol{r'}) &= \frac{\partial \Phi \left(s_0^{(n)} \right)}{\partial s_0^{(n)}(\boldsymbol{r'})} \\
&= \frac{1}{2} \sum_{s,g} \int \Re \left\{ \frac{\partial G(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)})}{\partial s_0^{(n)}(\boldsymbol{r})} \delta P^*(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)}) \right\} d\omega
\end{aligned} \tag{5}
$$

is the gradient. The Hessian is

$$H^{(n)}(\boldsymbol{r''}, \boldsymbol{r'}) = \frac{\partial^2 \Phi \left(s_0^{(n)} \right)}{\partial s_0^{(n)}(\boldsymbol{r''}) \partial s_0^{(n)}(\boldsymbol{r'})} \tag{6}$$

and $\boldsymbol{r}$, $\boldsymbol{r'}$ and $\boldsymbol{r''}$ are arbitrary positions into the slowness model.

Another way to compute the gradient (Margrave et al., 2011) is in terms of the Green's function applying a small perturbation $\delta s_0^{(n)}(\boldsymbol{r})$ leading to small changes in the field $\delta G(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)})$. Now, we have the relationship:

$$\lim_{\delta s \to 0} \frac{\delta G}{\delta s} = \frac{\partial G(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)})}{\partial s_0^{(n)}(\boldsymbol{r})} = -\omega^2 G(\boldsymbol{r}_g, \boldsymbol{r}, \omega | s_0^{(n)}) G(\boldsymbol{r}, \boldsymbol{r}_s, \omega | s_0^{(n)}) \tag{7}$$

Replacing equation 7 in equation 5, the gradient can be written:

$$g^{(n)}(\boldsymbol{r}) = \sum_{s,g} \int \omega^2 [G(\boldsymbol{r}, \boldsymbol{r}_s, \omega | s_0^{(n)})] \times [G(\boldsymbol{r}_g, \boldsymbol{r}, \omega | s_0^{(n)}) P^*(\boldsymbol{r}_g, \boldsymbol{r}_s, \omega | s_0^{(n)})] d\omega \tag{8}$$

Equation 8 can be interpreted as a depth migration of the data residuals. This result shows that we can use our processing tools to do a FWI iteration. Margrave et al. (2010) use a similar representation for the gradient but for the time domain and recognize that the gradient is a pre-stack reverse-time migration (RTM) with a cross-correlation imaging condition.

In the next section, we present a matrix approximation for the FWI and the Newton, Gauss-Newton and gradient approximations of the Hessian.

The Newton, Gauss-Newton and gradient methods

The objective of the FWI is to find a model that generates a synthetic data equals to the observed data (where the misfit the misfit function vanishes).

Rewriting equation 1 as a misfit vector, we have $\Delta \mathbf{d} = \mathbf{d}_{obs} - \mathbf{d}_{syn}(\mathbf{m})$ of dimension N, where $\mathbf{d}_{obs}$ is the acquired data, $\mathbf{d}_{syn}$ is the data prediction (forward modeling) and $\mathbf{m}$ is the model. The minimization of the misfit function is done using least-squares method. The norm-2 of the misfit is:

$$C(\mathbf{m}) = \frac{1}{2}\Delta \mathbf{d}^{\dagger}\Delta \mathbf{d} \qquad (9)$$

where $\dagger$ denotes the transpose conjugate.

Virieux and Operto, 2009 minimize equation 9 with a second-order Taylor-Lagrange expansion around a initial model $\mathbf{m}_0$ plus a perturbation model $\Delta \mathbf{m}$ (the updated model is $\mathbf{m} = \mathbf{m}_0 + \Delta \mathbf{m}$), with norm-2 as $C(\mathbf{m}_0 + \Delta \mathbf{m})$. They compute the derivative with respect to the model parameter m and equals to zero to find the minimum. The solution is written as:

$$\Delta \mathbf{m} = -\left[\frac{\partial^2 C(\mathbf{m}_0)}{\partial \mathbf{m}^2}\right]^{-1} \frac{\partial C(\mathbf{m}_0)}{\partial \mathbf{m}} \qquad (10)$$

The second part of the equation 10, the first derivative of the objective function is the gradient and is:

$$\nabla C_{\mathbf{m}} = \frac{\partial C(\mathbf{m}_0)}{\partial \mathbf{m}} = -\Re\left[\mathbf{J}^{\dagger}\Delta \mathbf{d}\right] \qquad (11)$$

where $\mathbf{J}$ is the sensitivity or the Frèchet derivative matrix (Jacobian) and $\Re$ denotes the real part of the gradient.

The first term of equation 10 is found by a differentiation of equation 11 and evaluating it at $\mathbf{m} = \mathbf{m}_0$:

$$\frac{\partial^2 C(\mathbf{m}_0)}{\partial \mathbf{m}^2} = \Re\left[\mathbf{J}_0^{\dagger}\mathbf{J}_0\right] + \Re\left[\frac{\partial \mathbf{J}_0^T}{\partial \mathbf{m}^T}(\Delta \mathbf{d}_0^* \ldots \Delta \mathbf{d}_0^*)\right] \qquad (12)$$

where the symbol $*$ denotes complex number and $(\Delta \mathbf{d}_0^* \ldots \Delta \mathbf{d}_0^*)$ is a vector with the same size of parameters $\mathbf{m}$. More details can be found in Pratt et al. (1998). Substituting equations 11 and 12 into equation 10:

$$\Delta\mathbf{m} = - \left\{ \underbrace{\Re\left[\mathbf{J}_0^\dagger\mathbf{J}_0 + [\frac{\partial\mathbf{J}_0^T}{\partial\mathbf{m}^T}(\Delta\mathbf{d}_0^* \ldots \Delta\mathbf{d}_0^*)]\right]}_{\text{Hessian}} \right\}^{-1} \underbrace{\Re\left[\mathbf{J}^\dagger\Delta\mathbf{d}\right]}_{\text{Gradient}} \tag{13}$$

This result is known as *Newton method* and has a quadratic local convergence. This solution is great for nonlinear problems (as in seismic), but demands a big effort to be calculated (Pratt et al., 1998).

As an approximation of equation 13, the problem is assumed to be linear and the second term of the Hessian vanishes, leading to the solution:

$$\Delta\mathbf{m} = - \left\{ \Re\left[\mathbf{J}_0^\dagger\mathbf{J}_0\right] \right\}^{-1} \Re\left[\mathbf{J}^\dagger\Delta\mathbf{d}\right] \tag{14}$$

This approximation is known as the *Gauss-Newton* method.

The gradient method is another and simple approximation of the solution of the equation 13 where the Hessian is replaced by a scalar α, which is determined, in general, by a line search (Pratt et al., 1998). The gradient method can be written as:

$$\Delta\mathbf{m} = -\alpha\Re\left[\mathbf{J}^\dagger\Delta\mathbf{d}\right] \tag{15}$$

In the next section, the Gauss-Newton solution is used to demonstrate the deconvolution imaging condition in the migration (gradient) as another type of Hessian approximation.

Deconvolution imaging conditions

Usually a correlation imaging condition is used during migration due to its stability, but the result has a lack in gain terms. Margrave et al. (2010) use a deconvolution imaging condition in the migration of the residuals and Margrave et al. (2011) develop a numerical assumption. Let's rewrite the equation 14 as:

$$\begin{aligned} \Delta\mathbf{m} &= -\Re\left[\left(\mathbf{J}^\dagger\mathbf{W}_d\mathbf{J}\right) + \epsilon\mathbf{W}_m\right]^{-1}\Re\left[\mathbf{J}^\dagger\mathbf{W}_d\Delta\mathbf{d}\right] \\ &= -\underbrace{\Re\left[\left(\mathbf{J}^\dagger\mathbf{W}_d\mathbf{J}\right) + \epsilon\mathbf{W}_m\right]^{-1}}_{\text{Inverse of Hessian term}}\underbrace{\Re\left[\mathbf{J}^T\mathbf{W}_d\Delta\mathbf{d}^*\right]}_{\text{Gradient term}} \end{aligned} \tag{16}$$

where $\mathbf{W}_d = \mathbf{S}_d^t\mathbf{S}_d$ is a data-weighting matrix, $\mathbf{W}_m = \mathbf{S}_m^t\mathbf{S}_m$ is a regularization matrix and ϵ is a stabilization factor (to avoid zeros during computation). The operator $\mathbf{S}_d$ can be used as a diagonal weighting operator to control the weight of the elements of the misfit vector.

The operator $\mathbf{S}_m$ is, in general, used as a roughness operator, to penalize the roughness of the model $\mathbf{m}$. The symbol * is the complex conjugate.

The Jacobian can be explicitly written as:

$$\mathbf{J} = \frac{\partial \mathbf{u}}{\partial \mathbf{p}} = \mathbf{B}^{-1} \frac{\partial \mathbf{B}}{\partial \mathbf{p}} \mathbf{u} \tag{17}$$

where $\mathbf{B}$ is the forward modelling operator, $\mathbf{B}^{-1}$ is the Green's operator and $\mathbf{u}$ is the wavefield.

Now, substituting the Jacobian of equation 17 in gradient term of equation 16, expanding and letting $\mathbf{W}_d = \mathbf{I}$, the gradient is:

$$\nabla C_{\mathbf{m}} = \Re \left[\mathbf{u}^T \frac{\partial \mathbf{B}^T}{\partial \mathbf{p}} \underbrace{\mathbf{B}^{-1^T} \Delta \mathbf{d}^*}_{\substack{\text{back-propagated time} \\ \text{reversed residual}}} \right] \tag{18}$$

The gradient of the equation 18 is a reverse time migration with a cross-correlation imaging condition with no gain correction.

Doing the same in the Hessian term, we have:

$$\mathbf{H} = \Re \left[\mathbf{u}^\dagger \overbrace{\left(\frac{\partial \mathbf{B}}{\partial \mathbf{p}} \right)^\dagger}^{\text{scatterer weighting}} \underbrace{\mathbf{B}^{-1\dagger} \mathbf{B}^{-1}}_{\text{geometrical spreading}} \frac{\partial \mathbf{B}}{\partial \mathbf{p}} \mathbf{u} \right] \tag{19}$$

The term $(\partial \mathbf{B}/\partial \mathbf{p})^T$ is equals to ω^2 and $(\mathbf{B}^{-1})^\dagger (\mathbf{B}^{-1})$ is, for a homogeneous medium, equals to r^{-2} (Margrave et al., 2011). Substituting equations 18 and 19 into equation 16, we get:

$$\Delta \mathbf{m} = \left(\frac{r}{\omega} \right)^2 \underbrace{\frac{\Re \left(\mathbf{u}^T \mathbf{BPTR} \right)}{\Re \left(\mathbf{u}^\dagger \mathbf{u} \right)}}_{\substack{\text{deconvolution} \\ \text{imaging condition}}} \tag{20}$$

where the **BPTR** is the backpropagated time-reversed data residual $\mathbf{B}^{-1^T} \delta \mathbf{P}^*$ which is divided by the autocorrelation of the downward propagated field. This is equivalent to a deconvolution imaging condition.

Equation 20 shows that the model update can basically be computed by a reverse-time migration of the residuals with a deconvolution imaging condition to get a better estimation

of the reflections in the update. Margrave et al. (2010) used this estimation, but with a WEM, combined with well information to do a velocity inversion. In this work we are trying to estimate the velocity model using the same approximation, but using the WEM instead of the RTM.

TEST 1: MUTE

The FWI with a deconvolution imaging condition is tested on synthetic data only. A velocity model (figure 1) is used for the forward modeling, simulating a seismic marine acquisition, and has a resolution of 10x10 meters. The red line is the water bottom picking and the vertical lines (solid and dashed) are simulations of well positions. The water bottom picking is used as a mute edge (everything above the picking is muted, even for initial model as for each update iteration) and the well positions are used in the next session to create initial velocity models.
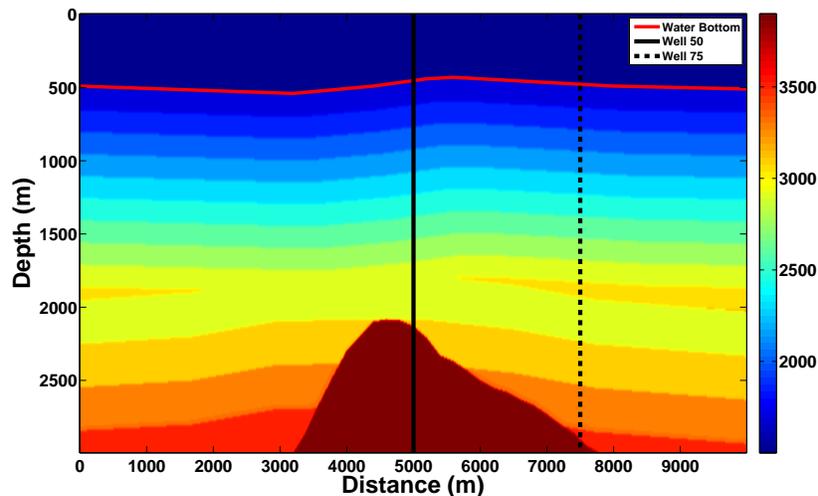
FIG. 1: Velocity model used to create the shots for the FWI test. The color bar represents the wave propagation velocity in m/s. The red line is the water bottom picking and the vertical lines (solid and dashed) are simulations of well positions.

Synthetic shots are generated to be used as the "recorded" data. To simulate a seismic acquisition, 100 shots were created using the velocity model of the figure 1 with an acoustic forward modeling code. The shots 3, 50 and 98 are shown on figure 2:

The shots have a split-spread receiver pattern and they may have different numbers of receivers that go from 201 in the borders to 401 in the center, with 3s of recorded time. The shot spacing is 100m and the receiver spacing is 10m (maximum offsets of -2000m to 2000m). However, for the FWI code, only 96 shots were used (the same number of "workers" in the computer to run the shots in parallel). The dominant frequency of the source wavelet used is 10Hz, but the dominant frequency of the recorded data is about 8Hz, as is possible to see in figure 3.

The initial velocity model (figure 4) for this test is a smoothed version of the real velocity model (figure 1). The Gaussian window for smoothing is calculated as $\lambda_{nom} =$
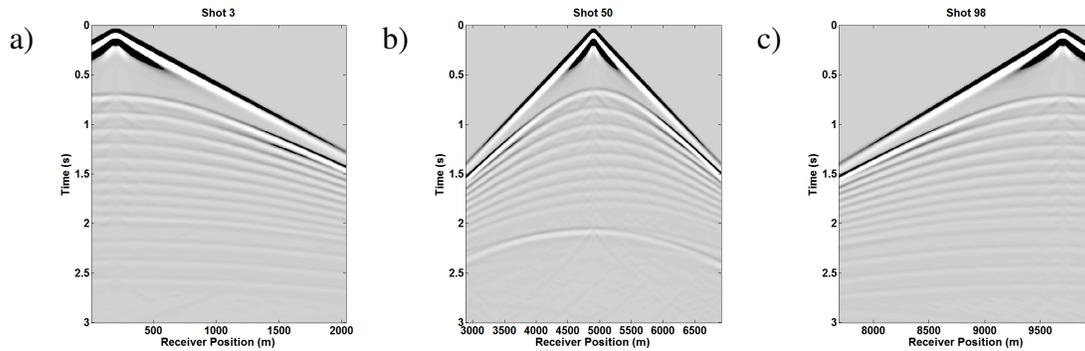
FIG. 2: Shots created using the velocity model of the figure 1. a) Shot 3, b) shot 50 and c) shot 98.
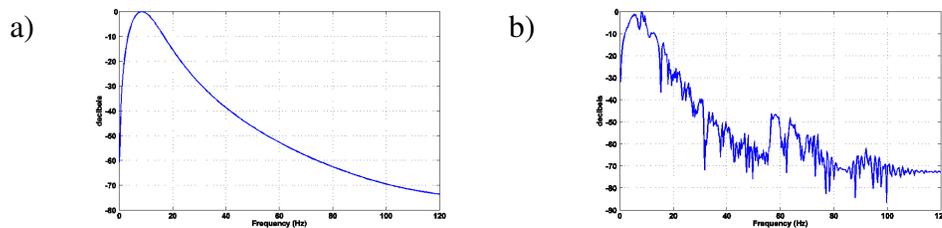
FIG. 3: Amplitude spectrum for the a) wavelet and b) shot 50.

f_{dom}/V_{mean}, where V_{mean} is the mean velocity of the real velocity model, equal to 2539 m/s. This equation is the same as used by Margrave et al. (2010). A mute is applied above the water bottom (red line) to eliminate any event above it. This mute is also used at each iteration in the stacked migrated residuals.
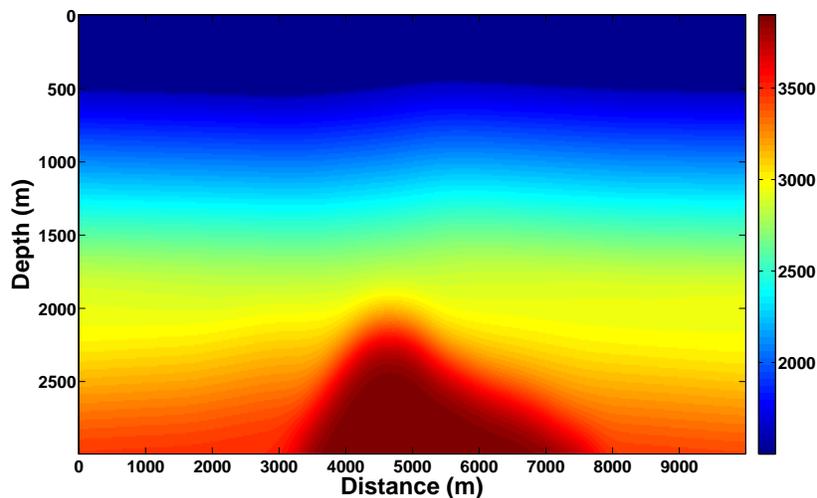
FIG. 4: The initial model for the test is a smoothed version of the model of the figure 1.

New shots are generated with the velocity model of the figure 4, and the same wavelet used to create the real data, as shown in figure 5: a) is the real shot, b) is the synthetic data and c) are the residuals. It is imported to remember that the same wavelet of the input shots

is being used at each iteration. That is the reason the direct wave is perfectly removed in the residuals. We also note a small WB in b), due to the WB mute in the initial model.
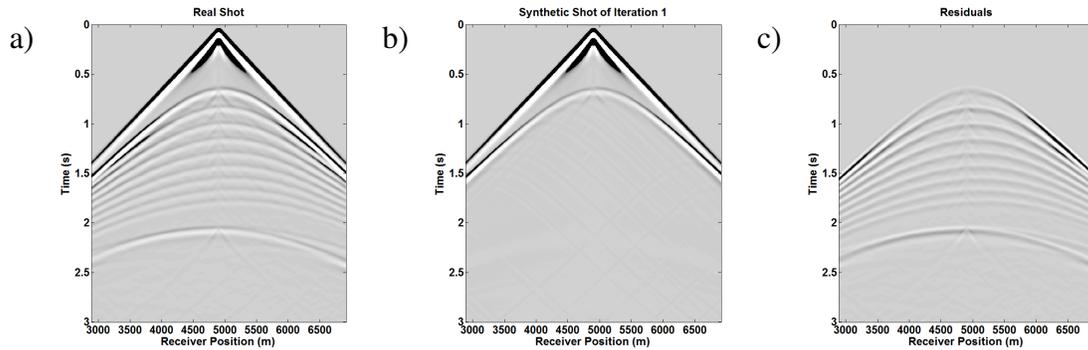
FIG. 5: Shot number 50 in the first iteration (synthetic shots created using initial velocity model of figure 4a), where a) is the real shot, b) is the synthetic data and c) are the residuals.

The residuals are migrated using the PSPI migration (Ferguson and Margrave, 2005) in frequency bands, as used by Margrave et al. (2010). However, we use different frequency bands. It is usual to start the inversion with low-frequency content (Pratt et al., 1998; Plessix et al., 2010), to avoid a local minimum, and then invert higher frequencies. In an absence of low frequencies in the data, a deconvolution can be applied to recover and invert the pseudo-low frequencies (Fei et al., 2012). Warner et al. (2013) invert only the frequencies with a high signal to noise ratio, starting with the lowest frequency possible. In this work, we are using frequencies above -10dB. Looking back at figure 3 b, we decided to use frequencies 6 to 15Hz. The PSPI migration code allows the use of frequency bands to migrate. Here we are using the frequencies above -10dB as the maximum frequency of the band, and the minimum will always be 1Hz. This way, the first iteration has the residuals migrated in the frequency band of 1 to 6Hz. 96 residuals are migrated and then stacked to create the model update. Pratt et al. (1998) inverts the same frequency until stability. A line search with 21 tests is used to calculate the step length, and it is controlled by the shot in the middle position of the model. Plessix et al. (2010) repeat same frequency band 20 times before change to next frequency band. Here, we are repeating the same frequency band 10 times before increase the maximum frequency by 1Hz. In other words, we repeat the inversion in the frequency band of 1-6Hz 10 times, then 1-7Hz 10 times and keep this process until reach the maximum frequency of 15Hz, or until we decide to stop the process.

Margrave et al. (2010) suggest the use of a mute in the migrated residuals to avoid artifacts. Figure 6 shows 4 different mute tests applied in the migrated residuals during the process: a) no mute, b) narrow mute, c) normal mute and d) wide mute, for the data inverted using the initial velocity model of the figure 4.

The absence of mute drive the inversion to a final model with a good resolution (figure 7), but allow the appearance of artifacts in the model, as the almost vertical object in the figures 7b, c and d, leading the line search to stop updating the model by selecting a step length close to zero. These artifacts may be the sum of some far offset effects in the migrated residuals and a mute looks to be a good strategy of improvement. It is possible to observe some shadows around and inside the high velocity body. They show to be larger for the first iterations, but start to decrease as the migrated frequencies increase.
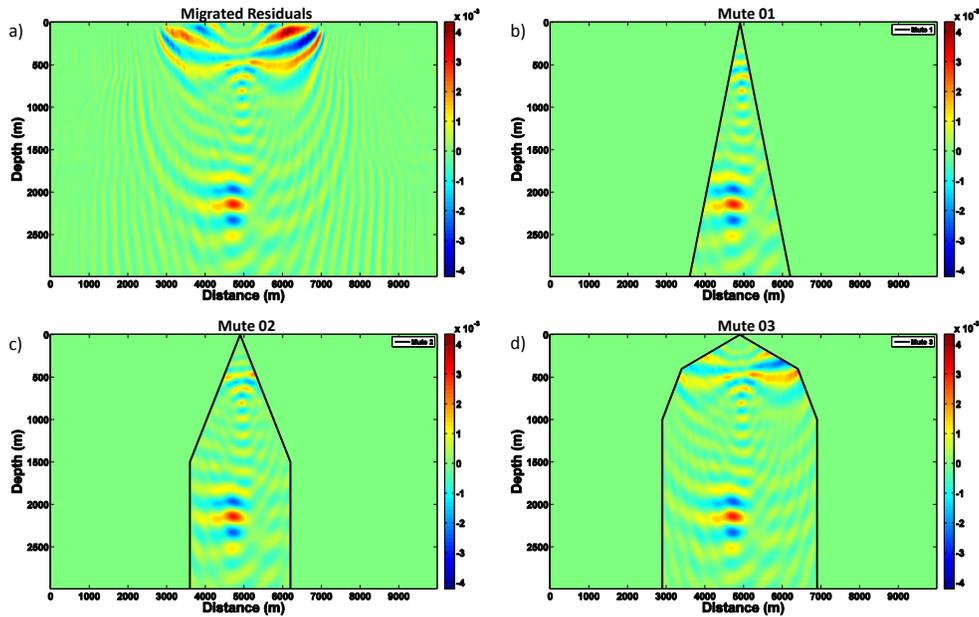
FIG. 6: Differents mutes applied in the migrated residuals: a) no mute, b) narrow mute, c) normal mute and d) wide mute.
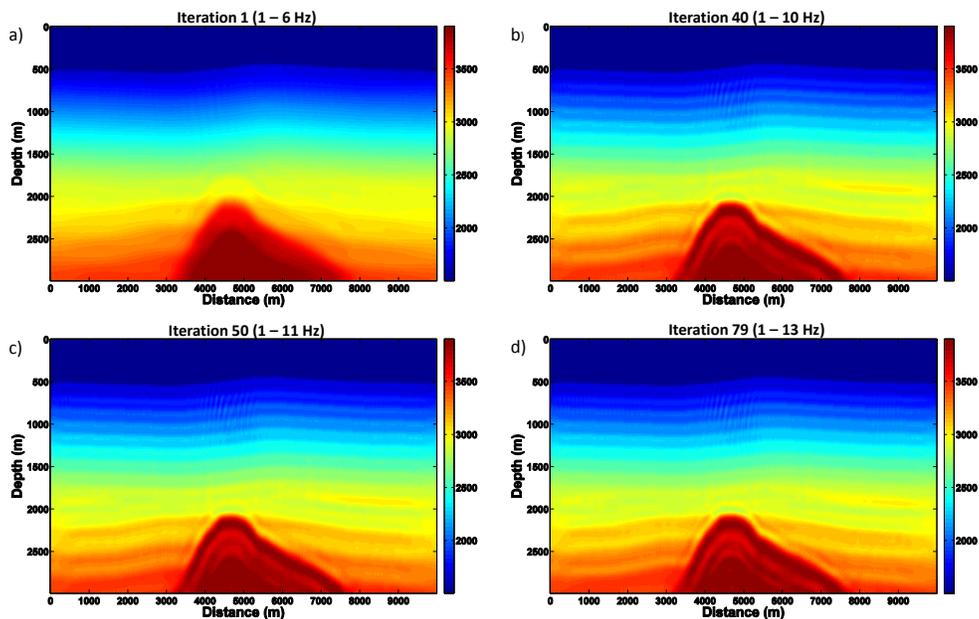
FIG. 7: Inverted velocities with no mute applied in the migrated residuals at a) iteration 1, b) iteration 40, c) iteration 50 and d) iteration 79. Artifacts are observed in the data.

A first test of mute is the narrow and very aggressive mute of the figure 6b, allowing only a small amount of data to pass the selection. The results are presented in figure 8. Artifacts due to far offset effects are not observed anymore. However, the mute is too aggressive at lower depth, and the borders of the mute windows are present in the velocity model. We also see the presence of the shadows close to the high velocity body, same as seen in figure 7. Even so, the result is encouraging.
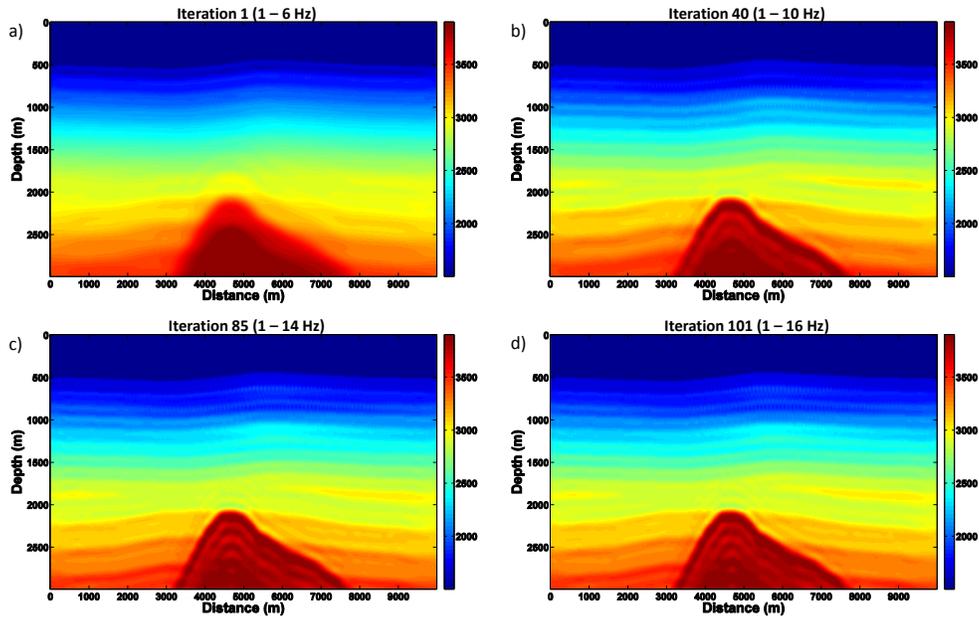
FIG. 8: Inverted velocities with the mute 01 (figure 6b) applied in the migrated residuals at a) iteration 1, b) iteration 40, c) iteration 85 and d) iteration 101.

Figure 9 shows the results of the less aggressive mute of the figure 6c. No artifacts are observed and, differently of the results of the narrow mute, the lower depth parts of the inverted velocities show a better resolution. The shadows around the high velocity bodies are still present. It is possible to say that we had a good improvement by using a not so aggressive mute.
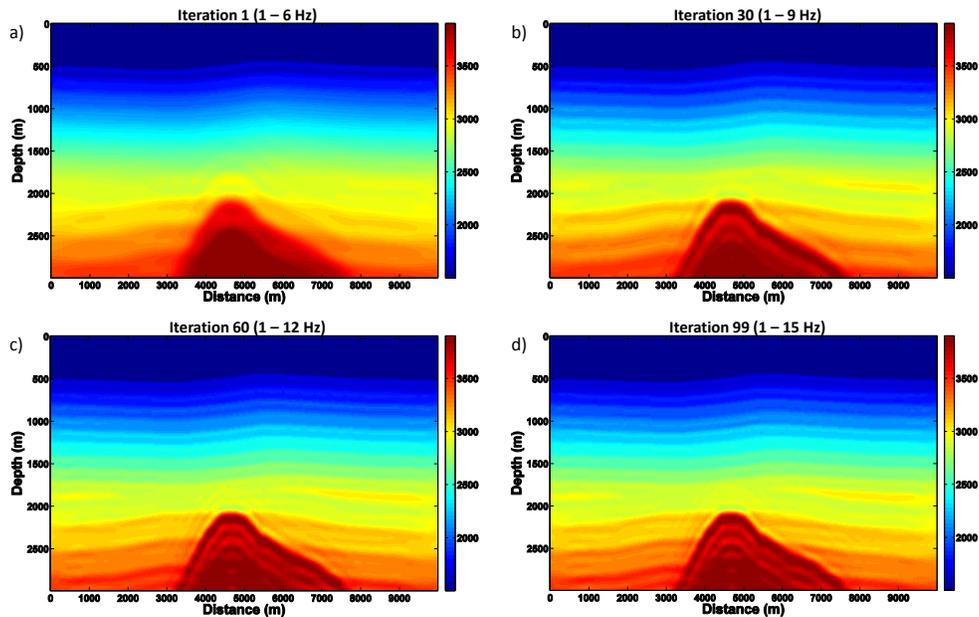
FIG. 9: Inverted velocities with the mute 02 (figure 6c) applied in the migrated residuals at a) iteration 1, b) iteration 30, c) iteration 60 and d) iteration 99.

Now, more data is allowed to pass in a conservative mute (figure 6d). It is a try to

have a better stack but without artifacts. However, the results of figure 10 show that this objective is not reached. Strong artifacts appeared in the edges of the inverted velocity and artifacts in the middle of the model, similar of the artifacts in the figure 7, is also observed, suggesting that the mute cannot be very conservative, or artifacts can appear in the inverted model.
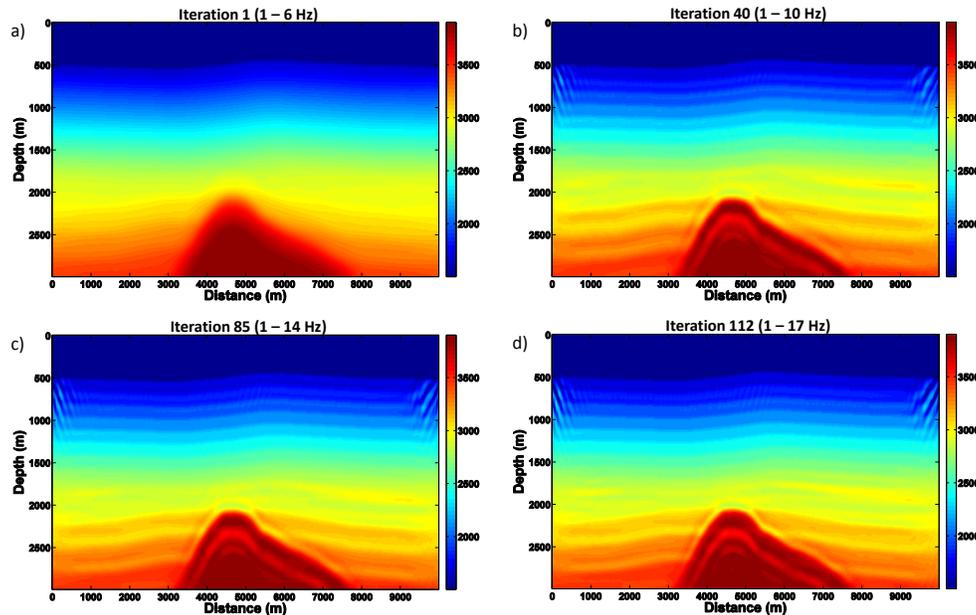
FIG. 10: Inverted velocities with the mute 03 (figure 6d) applied in the migrated residuals at a) iteration 1, b) iteration 40, c) iteration 85 and d) iteration 112.

Figure 11a compares the shots errors (the average residuals of all the synthetic shots relative to the real shots, per iteration) for all the mute tests done, showing a best behavior for the inversion with the normal mute (blue line). As this is a synthetic test, it is possible to compare each inverted velocity model per iteration to the real velocity model (it is not possible in the real world), what is plotted in figure 11b. Again, the normal mute has a better behavior, with the lowest errors values. Curiously, the worst behavior happens with the wide mute, due to the strong artifacts in the borders. The test with no mute applied has a better behavior than the narrow mute test, even with the artifacts that appeared in the middle of the model. The low depth stripes in the narrow mute test must have a high weight in the errors calculations.

The narrow and normal mute tests do not look to have the errors stabilized (figures 11 a and b), suggesting that we must keep with the iterations at higher frequency. In the next session, it is done for the normal mute, until reach 40Hz as the maximum inverted frequency.

Comparing all the resulted velocity models with the initial (dashed line) and real (black line) velocity models in figure 12, the normal mute test (red line) placed the reflectors (layers interfaces) in a more precise position than the others tests. Not muting the artifacts (magenta line) shows a good result, but, as noticed in figure 7, artifacts are present in the model. The narrow mute (blue line) do not match the reflectors properly, as the wide mute shows higher effect of the shadows around the high velocity body (around 2000m).
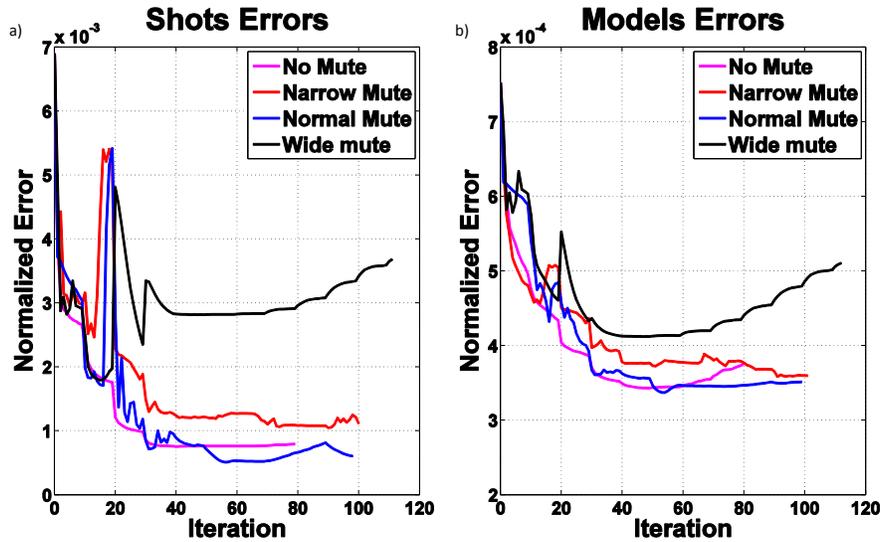
FIG. 11: Comparison of the errors of the a) synthetic shots to the real shots and b) inverted velocity models to the real velocity model. In both plots, the normal mute (blue line) shows the best behavior.
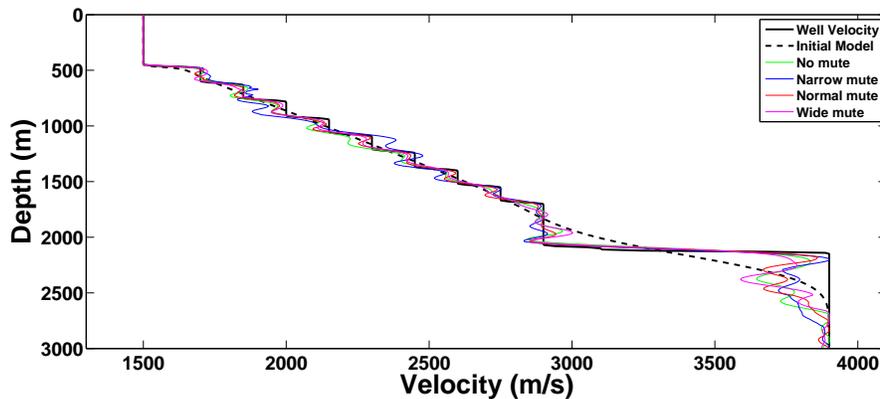
FIG. 12: Comparing the velocities models in the center position of the real model (black line), initial model (dashed line) and the last iteration of all the tests. The normal mute test (red line) placed the reflectors (layers interfaces) in a more precise position than the others tests.

The results so far show that a not too aggressive and not too conservative mute is a good choice, and showed that we must keep inverting higher frequencies. However, the initial velocity model is a "good" start, as it is a smoothed version of the real velocity model. In the next section, the smoothed and others initial velocity models are tested.

TEST 2: INITIAL VELOCITY MODEL

Warner et al. (2013) use a PSDM velocity as initial velocity model for the FWI. The reason is to start the process as close as possible to the global minimum and to avoid be trapped in a local minimum. We are testing this effect using different initial velocity models. All the tests here are done using the normal mute (figure 6c). Also, as noticed

in the previous section, we are keeping the iterations for higher frequencies, stopping in 40Hz. However, for frequencies higher than 15Hz the frequency range is repeat 5 times and, for frequencies higher than 25Hz the maximum frequency of the range is increased by 5Hz. This was chosen due to the low convergence rate.

Different initial velocity models were tested. One is simply the same as used for the mute test, but using only the normal mute and continuing the iterations for higher frequencies. Others two are initial velocity model based in well simulations, drawn in figure 1. One is based on the "Well 50" (in a position at 50% of the total length of the model) and the other is based on the "Well 75" (in a position at 75% of the total length of the model). In both case, the initial velocity model consist on flat layers, which were smoothed and plotted at figure 13.
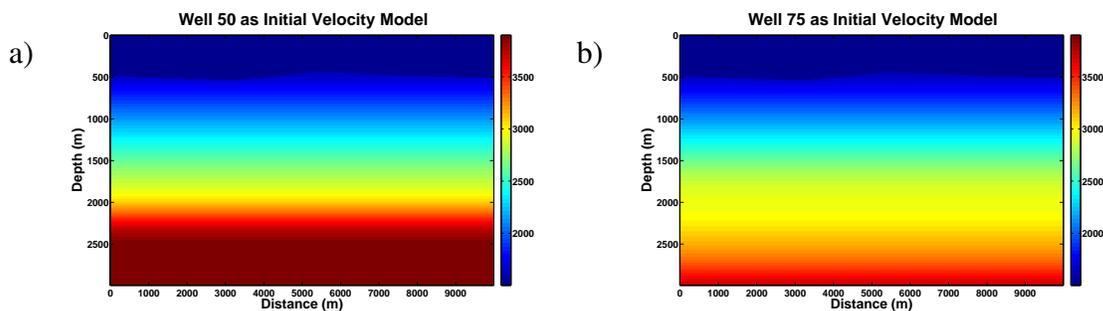
FIG. 13: Initial velocity models using as base a) well 50 and b) well 75 in the right (as draw in figure 1).

The results are plotted in figure 14, the iteration 55 for all the cases and in figure 15. Is is visible that the inversion is more successful when using a smoothed initial velocity model (figures 14b and 15b). We also have a reasonable result when using the initial velocity model based on Well 75 (figures 14d and 15d), where the shallow and deep layers have a good resolution, as the high velocity body edges, but the correct velocity could not be "filled" inside it. Maybe the results could be improved, as we are starting with a not so good solution, if we start the inversion with lower frequencies and do more repetitions per frequency range. Even the sharp layers around 2000m have a better resolution. The results of the inversion using the initial velocity based on Well 50 (figures 14c) and 15c look less precise at high depth, where the interface of the layers and the high velocity body edges were found, but the correct velocity could not be calculated. For the shallow part, the resolution looks better than when the smoothed initial velocity model is used.

We also can note that the inversions of figure 15 starts to show some high sharp layers that do not exist (for example, around 2500m at figures 15b and 15d). We also observed that, in both tests using the simulated wells, artifacts started to appear in the borders of the models. This might be due to line search be driven by the shot in the middle of the model. The step length calculated is the best option for the shot in the middle, not worrying about effects outside its length. This suggests that the line search must be improved, using more shots in different positions.

When we compare the shots and models errors per iteration (figure 16) we check a very interesting result. If we look to the models errors (figure 16b), we have the same conclusion
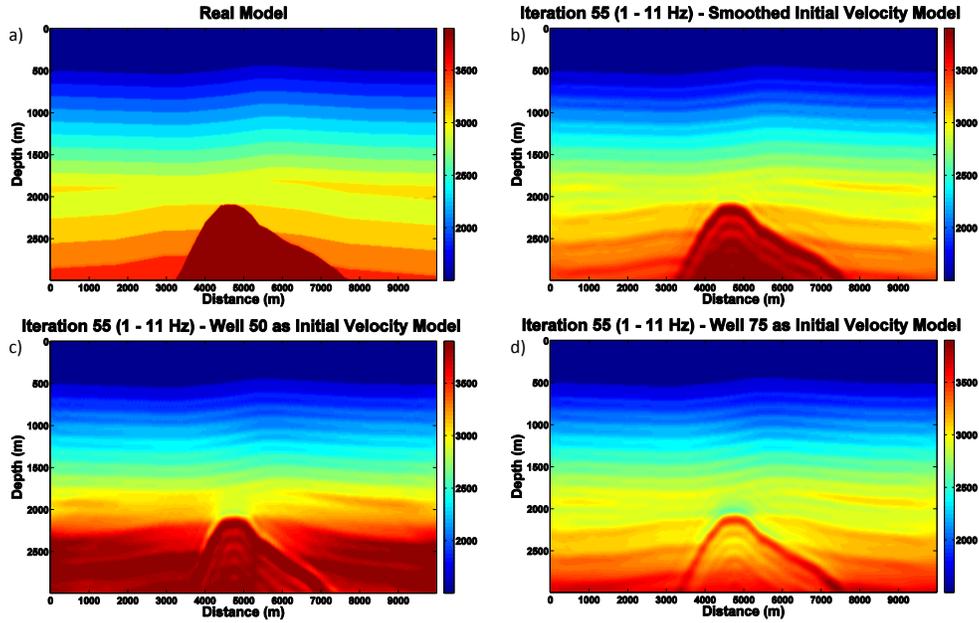
FIG. 14: Inverted velocities with different initial velocity models, where a) is the real model, b) is the iteration 55 of the inversion using the smoothed initial velocity model (figure 4), c) is the iteration 55 of the inversion using the well 50 as the initial velocity model (figure 13a) and d) is the iteration 55 of the inversion using the well 75 as the initial velocity model (figure 13b).
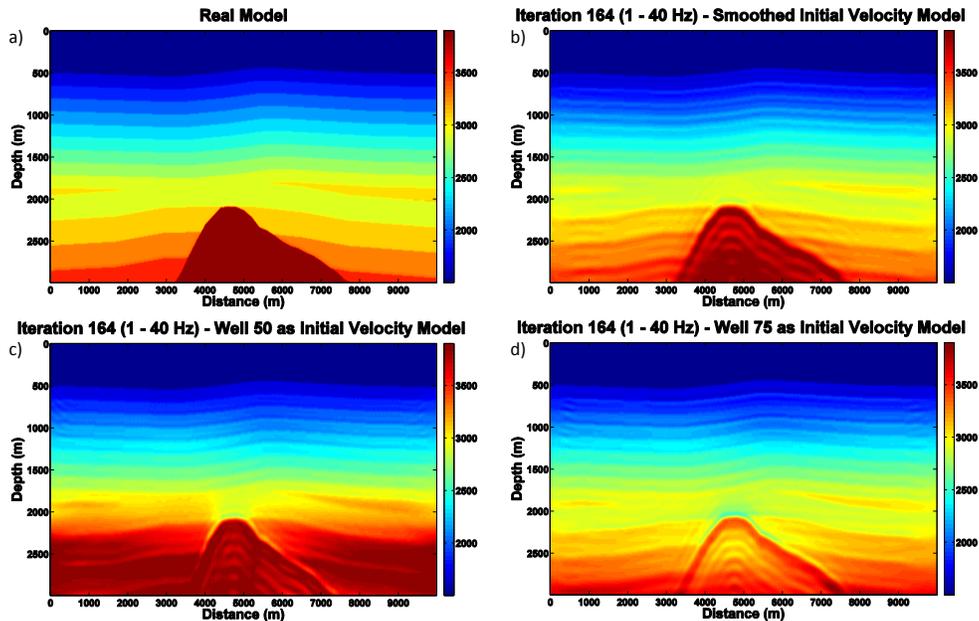
FIG. 15: Same as figure 14, but for the iteration 164. The inversion started to diverge after $\sim$60 iterations for almost all the cases. For the inversion using Well 75 (c), the inversion is still converging.

as looking the figures 14 and 15: the best inversion comes when using a smoothed initial velocity model (black line at figure 16b), and the worst result comes when using the Well

50 as initial velocity model (red line). However, if we look the shots errors in figure 16a, the errors that we can calculate in a real data inversion, we have a different result: the best shot errors comes when using the Well 50 as initial velocity model (red line). In general, we note a minimum value around iteration 55 for all the cases and divergence follows. However, for Well 75 initial model, the shots errors continue to decrease as we do more iterations.
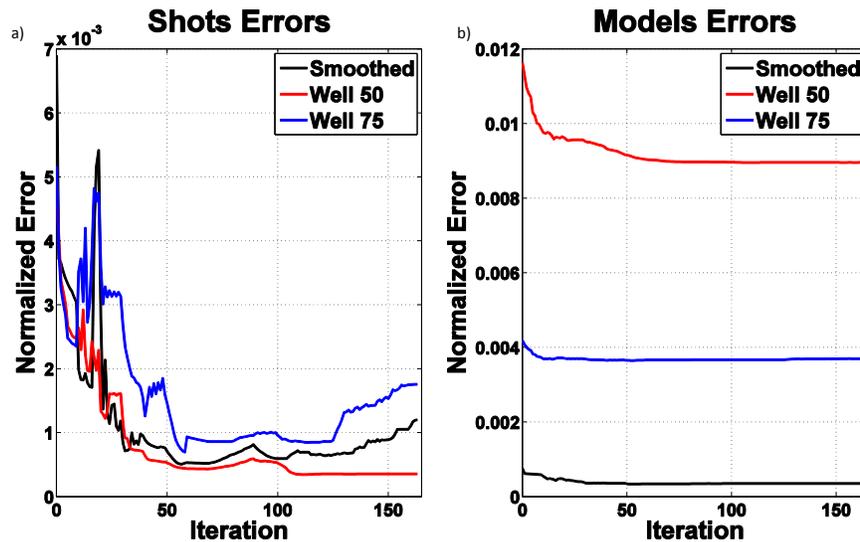
FIG. 16: Comparison of the errors of the a) synthetic shots to the real shots and b) inverted velocity models to the real velocity model, with different initial velocity models. For the shots errors, the inversion using the well 50 as initial velocity model looks more stable. However, comparing the models errors, the best result is the one using the smoothed velocity model as the initial model.

Looking at the 1D velocity check at figure 17, we observe that the inversion using the Well 75 as initial velocity model (red and blue dashed lines in figure 17b) is more accurate in the shallow part than the inversion when using the smoothed initial velocity model (red line in figure 17a). This information suggests that the average errors of the shots have a higher influence for the shallow area than for the deep area.

Comparing the differences between real and synthetic shots of the iteration 164 of the inversion using smoothed initial velocity model (figure 18, difference in c) and inversion using the Well 50 as the initial velocity model (figure 19, difference in c), it is notable that the final model of the Well 50 inversion leads to a synthetic data more similar to the real one than with the final model of the smoothed inversion. In others words, we found a bad solution that can generate synthetic shots very similar to the real ones. Treitel (1989) had a similar experience and suggested that a good fit of the shots is necessary but it is not a sufficient condition for a good inversion. It can be avoided if the initial guess is not too far off the real one.

Another test is trying the inversion using a constant velocity model as the initial velocity model. Figure 20 shows the result when we apply the FWI in the dataset using an initial velocity model with a constant velocity of 1500m/s (water velocity). The inversion stopped
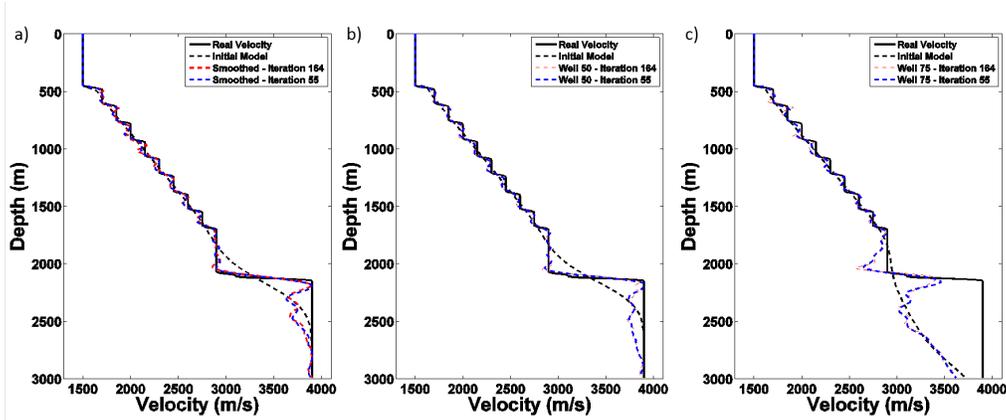
FIG. 17: Comparing the velocities models of iteration 164 (red dashed line) and iteration 55 (blue dashed line) in the center position of the real model (black line) and the initial model (dashed line) with the a) last iteration of the smoothed initial model test, b) the well 50 as initial velocity model test and c) the well 75 as initial velocity model test.
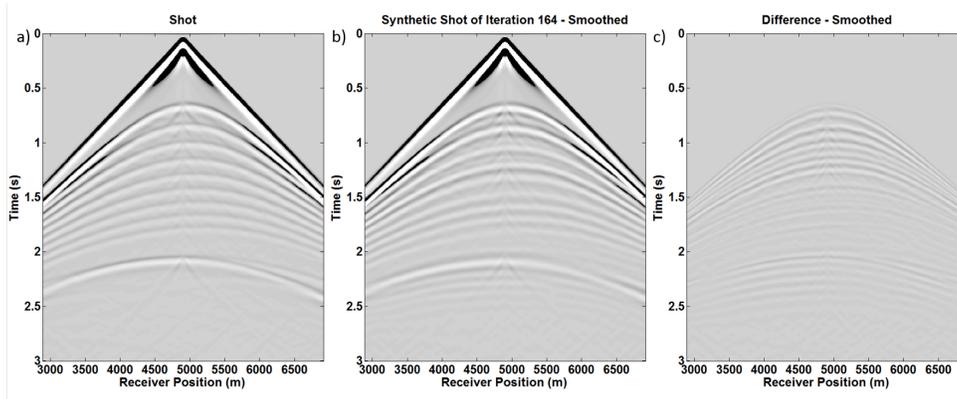
FIG. 18: Comparing a) the real shot, b) synthetic shot of iteration 164 and c) the difference, with the smoothed initial velocity test.
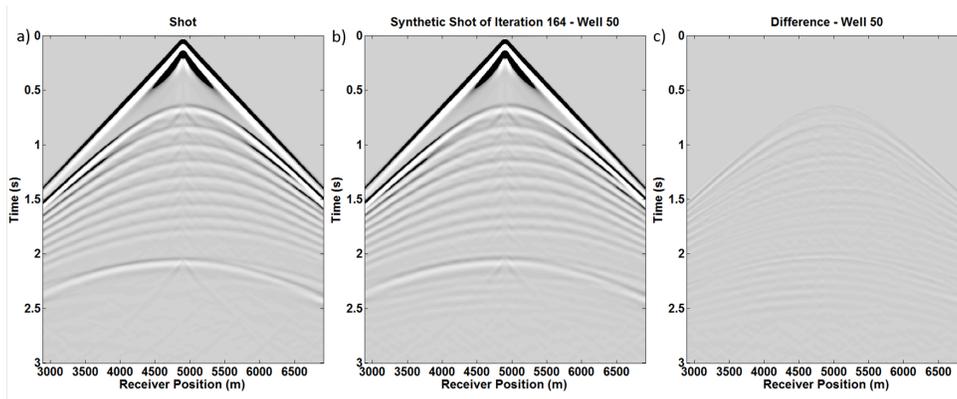
FIG. 19: Comparing a) the real shot, b) synthetic shot of iteration 164 and c) the difference, with the well 50 as the initial velocity test.

in a local minimum in the very first iterations and didn't change at all, even doing the same process as for the wells tests and reaching 164 iteration.
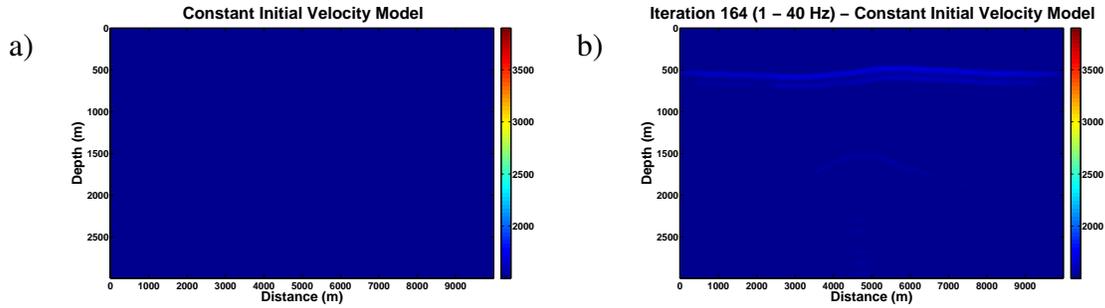
FIG. 20: Results for the a) constant velocity initial model with b) 164 iterations. The inversion stopped in a local minimum.

We can say that the better is the initial velocity model, the better will be the inversion. However, we had reasonable results using well simulation as the initial model. We are able to believe that we can use more than one well to generate an initial velocity model. Maybe, if we have an initial model generated by an interpolation of two or more wells, we can have an initial velocity model similar to the smoothed version of the model on figure 4.

MARMOUSI

After good results using a simple velocity model, now we conduct the the tests in the Marmousi model (figure 21). This model is complicated and well known. The model has 3000m depth, 10420m width and 10m of resolution. The minimum wave propagation velocity in the model is 1500m/s and the maximum is 5500m/s
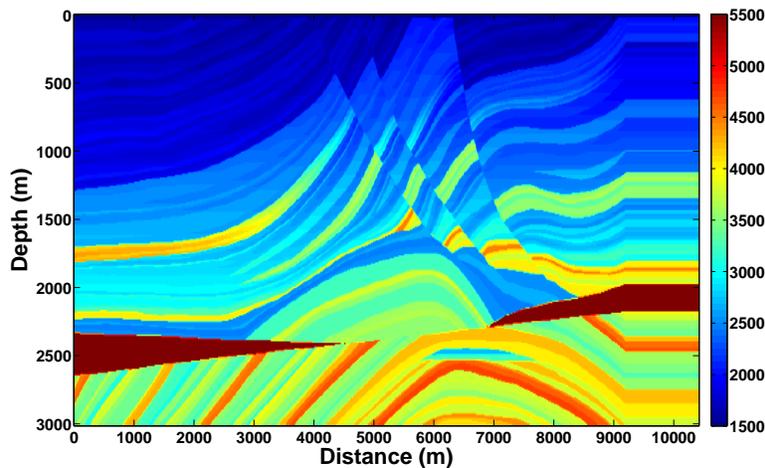
FIG. 21: The Marmousi 2D model. The colors bar indicates the wave propagation speed in m/s.

We use synthetic data as real data and they have a very similar parametrization as used before. 102 shots are created to simulate field data. Shot spacing is 100m. The shots have a maximum of 401 receivers (varying in the borders) totalizing 2000m of maximum offset. The register time is 3s and sample rating of 4ms. Some examples of the shots shown on figure 22. The dominant frequency of the wavelet is 10Hz (figure 23a) which generated data with similar dominant frequency (figure 23b).
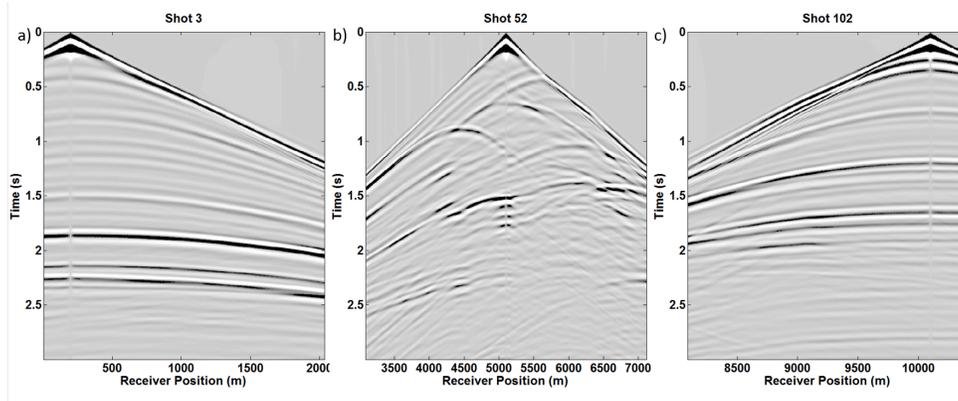
FIG. 22: Three of the 102 shots used in the test. a) shot 3, b) shot 52 and c) shot 102.
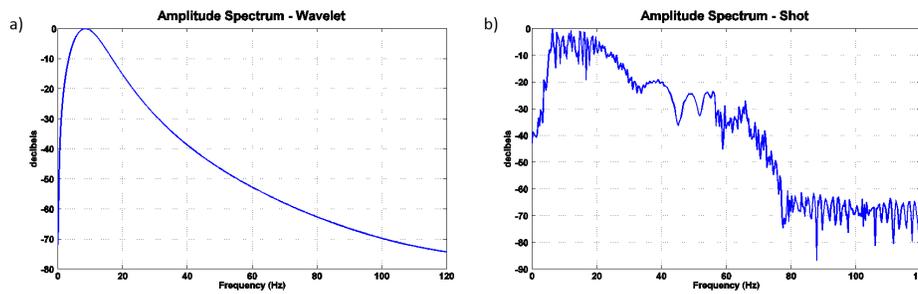
FIG. 23: Amplitude spectrum of the A) wavelet and B) shot.

The starting velocity model (figure 24) is a smoothed version of the model on figure 21 using a Gaussina window calculated by $\lambda_{nom} = f_{dom}/V_{mean}$, where V_{mean} is the mean velocity of the real velocity model, resulting in a Gaussian window.
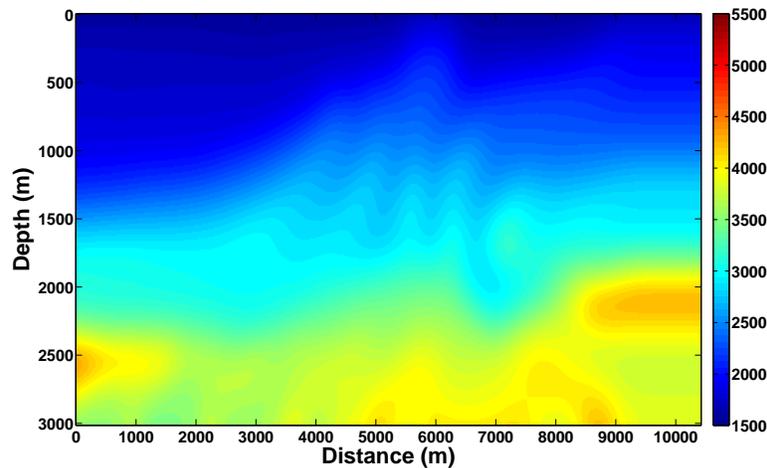
FIG. 24: Initial velocity model is a convolution of the Marmousi model with a Gaussian window of 290m.

For this test, the initial frequency range of the migration starts at 1-4Hz and the maximum frequency of the range is increase by one after 5 iterations (the second range is 1-5Hz) until reach 13Hz. After that the maximum frequency is increased by 5Hz each two itera-

tions. A normal mute similar to the mute of the figure 6c is applied at each migrated shot. The difference of the mute is that it starts at 300m in the surface.

We smooth the migrated stack at each iteration before the line search. The 2D Gaussian smoother is similar to the one use by Margrave et al. (2010), which is calculated to be decrease at each iteration by $h_w = v_{mean}/(6f_{max})$. The difference is that the v_{mean} is the mean velocity of the actual model.
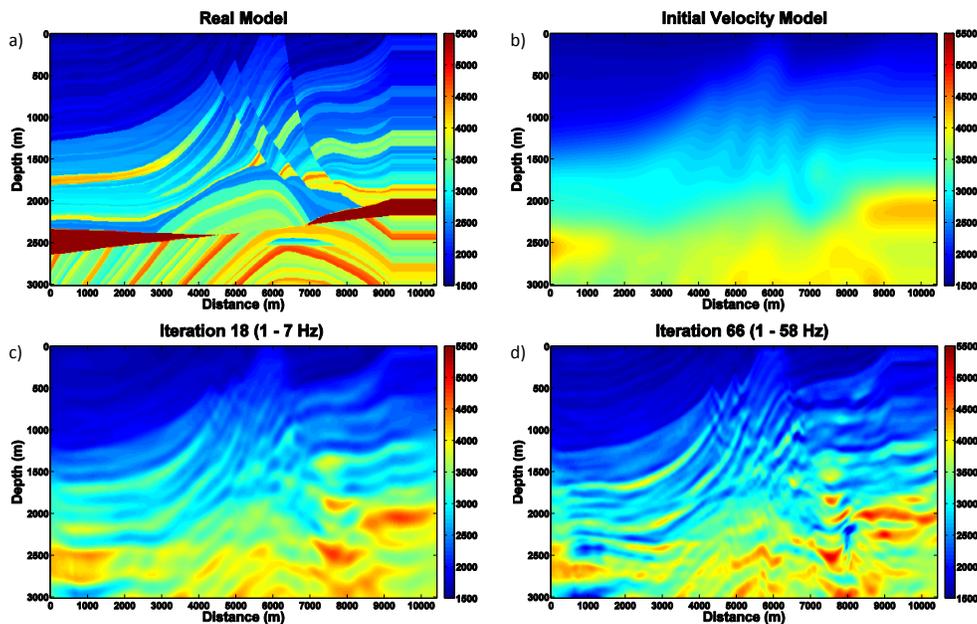
FIG. 25: FWI in the a) Marmousi model. b) is the smoothed initial model, c) is the inversion at iteration 18 (lowest model error) and d) is the iteration 66 (lowest shots errors).

The FWI process have 66 iterations and the models results are shown at figure 25, where a) is the real model, b) is the initial model, c) is the model at iteration 18 and d) is the model at iteration 66. The inversion shows higher resolution after 66 iterations (figure 25d) as we migrate higher frequencies and with reasonable quality on shallow depth (2000m) but looks to lose resolution for deep events if compared to the model of the iteration 18 (figure 25c).
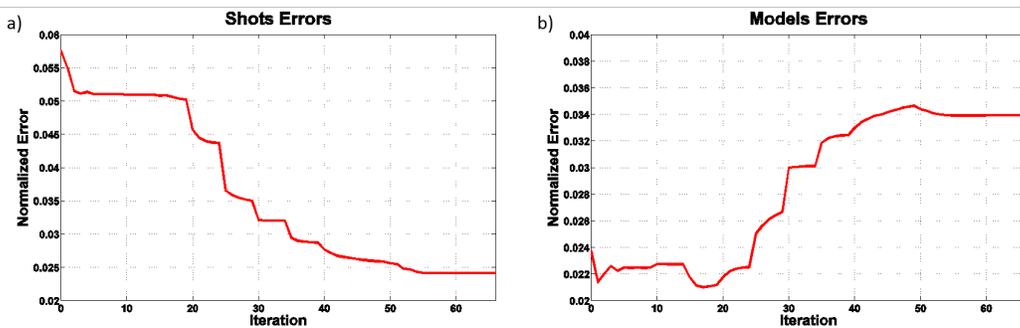
FIG. 26: Comparison of the errors of the a) synthetic shots to the real shots and b) inverted velocity models to the real velocity model. The models look to be fitting better the data at each iteration but the precision of the models decrease starting from iteration 20.

The shots errors and model errors are shown at figure 26. From the shots errors plot (figure 26a) the models are updating in the way to minimize the difference of the synthetic and real shots with the best fit at iteration 66. From the models errors plot (figure 26b), the best model (lowest error) is the model of the iteration 18. We are getting a better fit with a model far from the global minimum. Figure 27 compare the real shot (a) with the synthetic shot of iteration 66 (b) and calculate the difference (c). The fit is still not as good as we expect.
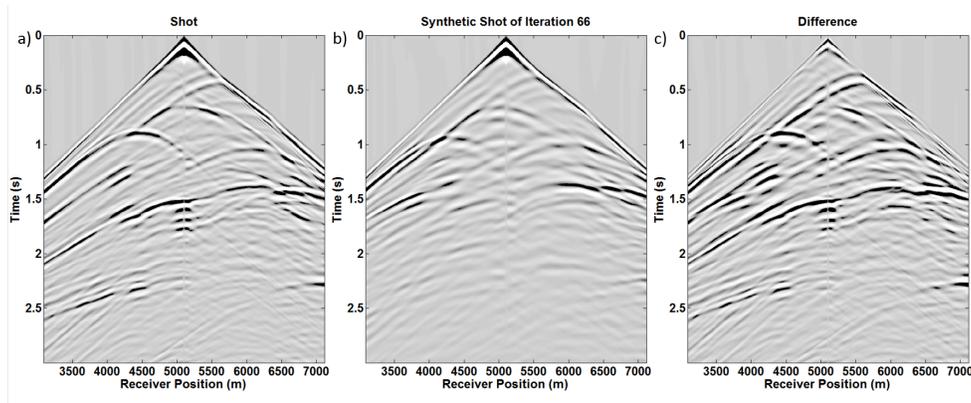
FIG. 27: Comparing shot of the best fit model. a) real shot 3, b) synthetic shot of iteration 66 and c) difference.

The 1D plot of figure 28 compares the inverted velocities (iteration 66 in red and iteration 18 in blue) with the real (black line) and initial (black dashed line) models in the middle of the survey. The FWI routine worked to find several reflectors positions but couldn't invert smaller layers. This must be caused by the absence of higher frequencies in the shots.
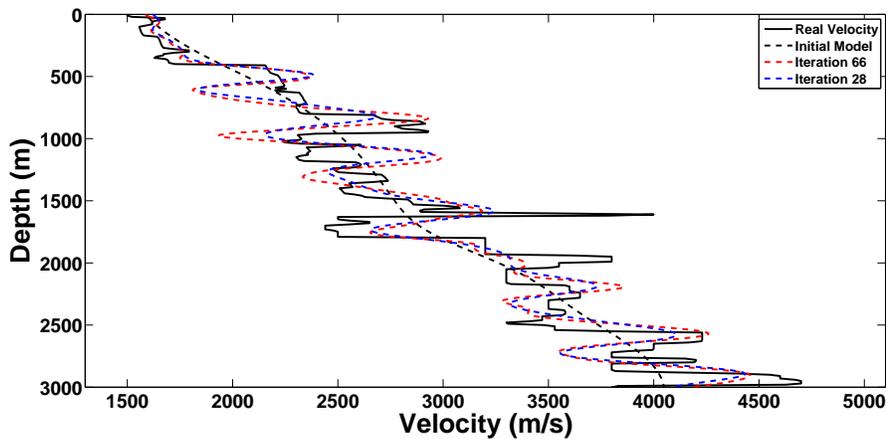
FIG. 28: Comparing the velocities models in the center position of the real model (black line), initial model (dashed line) and the last iteration of all the tests. The normal mute test (red line) placed the reflectors (layers interfaces) in a more precise position than the others tests.

We may get some improvement if in the deconvolution imaging condition step we use the same wavelet of the data. For now, our routine do an estimation of the wavelet and then do the deconvolution.

CONCLUSIONS

We presented a FWI test on acoustic synthetic data with the objective to understand and learn a best practice for application. A simple velocity model with a high velocity body (salt) were used to create shots in the entire length of the velocity model. This helps to avoid strong borders effects. The gradient is composed of the stack of the migrated residuals (difference between real and synthetic shots) and we used the PSPI migration instead of the RTM, showing good results The PSPI migration allowed us to apply the FWI in time domain, but migrate only specific frequency bands. The step length was calculated by a line search driven by a shot in the middle of the model, giving good approximations in most tests, but an improvement (using shots in different locations) looks necessary.

Two main tests were done in this work: finding a more effective mute for the migrated residuals and the effect of the initial velocity model in the convergence process. For the mute test, we used a smoothed version of the real velocity model as the initial model and the iterations were computed to be done until 15Hz. We showed that no mute or wide mute leads to reasonable results, but artifacts are allowed into the inverted model. A very narrow mute results in stripes at shallow depth, due to the low stacking factor. The mute that showed a better behavior was what we called the "normal mute", that is between the narrow and the wide mutes. The normal mute resulted in a high resolution inversion, without the presence of stripes or artifacts.

Four different initial velocity models were used in the second test which are the smoothed velocity model as used in the first test, two velocity models with flat layers and were based in two different wells positions in the real model, and a constant velocity model. For this test, the iterations were computed to run until reach 40Hz. We showed that the best result were obtained with the smoothed velocity test, giving a very good final model. The process using the initial velocity model based in a well that does not pass through the high velocity body (well 75) gave a reasonable result, inverting properly almost all the layers in the original model and the contour of the high velocity body, but could not "fill" it with the correct velocity. When the initial model based in a well that pass through the high velocity body (well 50) were used, all the interfaces were inverted, but, for the deep part of the model, the correct velocities could not be inverted. However, the shallow part had the higher resolution of all the tests and the errors of the shots showed the lowest values, even this not been the best inverted model, suggesting that the error is driven by the low depth reflectors. For the constant velocity model test, the process stopped in a local minimum in the first few iterations and the inversion was not possible. Therefore, not surprising, the best result was obtained with the initial velocity model that is the closest one to the real model. For all the others tests, we can say that the convergence were reached with approximate 50 iterations, around 11Hz, having only slightly changes after and, in mostly cases, starting to diverge.

The test in the Marmousi model shown encouraging results, with higher resolution inversion in some areas of the model but still need more improvements such as increase the frequency band of the data and to use a better estimation of the wavelet in the deconvolution imaging condition.

As future work, we suggest an improvement in the line search, using more shots or be

done around a step length calculated with a well information. We also suggest that, for initial velocity models that are not so close the local minimum (or, that are not the smoothed version of the real model), the iterations should start with lower frequencies and must be repeated only until the convergence for that frequency band is reached. An initial velocity model based on more than one well (interpolated) could be a good improvement. Improvements in the deconvolution imaging condition is necessary, using a better estimation of the wavelet of using the same wavelet of the data.

ACKNOWLEDGEMENTS

REFERENCES

Fei, T. W., Luo, Y., Qin, F., and Kelamis, P. G., 2012, Full waveform inversion without low frequencies: A synthetic study, chap. 484, 1–5, `http://library.seg.org/doi/pdf/10.1190/segam2012-0641.1`.

Ferguson, R., and Margrave, G., 2005, Planned seismic imaging using explicit one-way operators: Geophysics, **70**, No. 5, S101–S109, `http://dx.doi.org/10.1190/1.2073885`.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering, theory and application: Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Margrave, G., Ferguson, R., and Hogan, C., 2010, Full waveform inversion with wave equation migration and well control: CREWES Research Report, **22**, 63.1–63.20.

Margrave, G., Yedlin, M., and Innanen, K., 2011, Full waveform inversion and the inverse hessian: CREWES Research Report, **23**, 77.1–77.13.

Plessix, R., Baeten, G., de Maag, J. W., Klaassen, M., Rujie, Z., and Zhifei, T., 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set, chap. 184, 930–934, `http://library.seg.org/doi/pdf/10.1190/1.3513930`.

Pratt, R. G., Shin, C., and Hick, G. J., 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, No. 2, 341–362, `http://gji.oxfordjournals.org/content/133/2/341.full.pdf+html`.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, No. 8, 1259–1266, `http://dx.doi.org/10.1190/1.1441754`.

Treitel, S., 1989, Quo vadit inversio?: Geophysics: The Leading Edge of Exploration, 38–42.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, WCC1–WCC26, `http://dx.doi.org/10.1190/1.3238367`.

Warner, M., and Guasch, L., 2014, Adaptive waveform inversion: Theory, chap. 207, 1089–1093, `http://library.seg.org/doi/pdf/10.1190/segam2014-0371.1`.

Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., Stekl, I., Guasch, L., Win, C., Conroy, G., and Bertrand, A., 2013, Anisotropic 3d full-waveform inversion: Geophysics, **78**, No. 2, R59–R80, `http://dx.doi.org/10.1190/geo2012-0338.1`.

Wenyong, P., Margrave, G., and Innanen, K., 2013, On the role of the deconvolution imaging condition in full waveform inversion: CREWES Research Report, **25**, 72.1–63.19.