Tomographic full waveform inversion: Practical and computationally feasible approach

Ali Almomin and Biondo Biondi

ABSTRACT

We provide an alternative formulation to tomographic full-waveform inversion which computes the backscattering and the forward scattering components of the model separately. To maintain high resolution results of tomographic full-waveform inversion, the two components of the gradient are mixed based on a Fourier domain scale separation. This formulation is based on the Born approximation where the medium parameters are broken into a long wavelength and short wavelength components. This approximation has an underlying assumption that the data contain primaries only without multiples. The results of the Marmousi model show that convergence is possible even with large errors in the initial model that would have prevented convergence to conventional full-waveform inversion.

INTRODUCTION

Seismic velocity-analysis methods can be divided into two major groups. First, there are techniques that aim at minimizing the misfit in the data domain, such as full-waveform inversion (Tarantola, 1984; Pratt, 1999; Luo and Schuster, 1991). Second, there are other techniques that aim at improving the quality in the image domain, such as migration velocity analysis (MVA) (Symes and Carazzone, 1991; Biondi and Sava, 1999; Shen, 2004; Zhang et al., 2012). These techniques try to measure the quality of the image in several ways and then invert the estimated image perturbation using a linearized wave-equation operator.

There are several advantages to minimizing the residual in the image-space, such as global convergence, increasing the signal-to-noise ratio, and decreasing the complexity of the data (Tang et al., 2008). However, a common drawback in doing velocity analysis in the image domain is that only the transmission effect of the velocity are used. This results in a loss of vertical resolution in the estimated model updates. On the other hand, Full-Waveform Inversion (FWI) does not have that problem, since it utilizes the information from both the forward-scattered and back-scattered wavefields. This results in higher resolution in the model estimates. Moreover, the data misfit is computed in the data spaces directly without the need to go to another domain or to separate the data into several components. This direct computation of the errors results in a relatively simple relationship between the data residuals and the model updates. However, FWI has the disadvantage that its objective function

is far from being smooth and convex; it requires the starting model to be very close to the true model to avoid converging to local minima.

The conventional solution is to invert first for the velocity model using MVA techniques and then use the output as the initial model for FWI. However, this practice might not work if the results of MVA are not accurate enough for FWI to start. This could be a result of the larger null space that forward-scattered wavefields do not constrain. Moreover, the convergence rate of the MVA techniques is going to be sub-optimal, since they do not use all of the information in the data.

In a companion paper (Biondi and Almomin, 2012), we present a generalize framework called Tomographic Full Waveform Inversion (TFWI) that combines both FWI and WEMVA techniques. This generalized approach utilizes all the components of seismic data to invert for the medium parameters without the cycle-skipping problem. This is achieved in two steps: first by extending the wave equation and adding an offset axis to the velocity model, and second by adding a regularization term that drives the solution towards the zero subsurface offset (Symes, 2008). However, this velocity extension makes the propagation considerably more expensive because each multiplication by velocity becomes a convolution over the subsurface offset axis.

In this paper, we present an approximation that significantly reduces the computational cost of TFWI while maintaining its desirable characteristic of enabling the simultaneous inversion of all wavelengths of the model. First, we use the Born approximation to break the extended velocity model into two components: a background component affecting transmission and a perturbation component affecting reflections. Second, we reduce the background component to zero subsurface offset while keeping the perturbation component at all offsets. This simplification of the background component is the key in reducing the computation cost since the convolution with velocity in propagation becomes a multiplication instead of convolution. However, breaking the model into two components hinders the simultaneous inversion of different wavelengths of the model. Moreover, the scale separation by Born approximation is not perfect since the reflection component of a certain frequency or angle can affect the transmission component of another frequency or angle. Therefore, we add a final step where we mix the gradients of each component and then separate them in Fourier domain.

Another potential advantage of the method presented in this paper compared to the complete method presented in Biondi and Almomin (2012) is the ability to handle variable density without having to adapt a more complicated form of the wave-equation. In the complete method, the inversion tries to fit the data using a velocity model only. Variations in the velocity model will cause both a transmission and a reflection effect. This becomes an issue in the presence of density variations that only change the reflections of the data without affecting the transmission. In the efficient method, this issue can be avoided since the reflectivity is separated from the background. Therefore, the reflectivity part of the model can absorb any reflection effects that do not fit the background part of the model.

SCALE SEPARATION

Although the next derivation is done in frequency domain, the actual implementation is done in time domain. First, we start with the tomographic full waveform objective function J_{TFWI} which can be written as:

$$J_{\text{TFWI}}(\mathbf{v}(\mathbf{h})) = \|\mathbf{d}\left(\mathbf{v}(\mathbf{h})\right) - \mathbf{d}_{\text{obs}}\|_2^2 + \epsilon\|\mathbf{A}\mathbf{v}(\mathbf{h})\|_2^2, \tag{1}$$

where $\mathbf{h}$ is the offset, $\mathbf{v}(\mathbf{h})$ is the extended velocity model, $\mathbf{d}(\mathbf{v}(\mathbf{h}))$ is the modeled data, $\mathbf{d}_{\text{obs}}$ is the observed surface data, ϵ is a scalar weight of the regularization term and $\mathbf{A}$ is a regularization operator. The modeled data $\mathbf{d}(\mathbf{v}(\mathbf{h}))$ is computed as:

$$d(\mathbf{x}_s, \mathbf{x}_r, \omega; \mathbf{v}(\mathbf{h})) = f(\mathbf{x}_s, \omega)G(\mathbf{x}_s, \mathbf{x}, \omega; \mathbf{v}(\mathbf{h}))\delta(\mathbf{x}_r - \mathbf{x}), \tag{2}$$

where $f(\omega)$ is the source function, ω is frequency, $\mathbf{x}_s$ and $\mathbf{x}_r$ are the source and receiver coordinates, and $\mathbf{x}$ is the model coordinate. In the acoustic, constant-density case the Green's function $G(\mathbf{x}_s, \mathbf{x}, \omega; \mathbf{v}(\mathbf{h}))$ satisfies:

$$\left(v^{-2}(\mathbf{x}, \mathbf{h}) * \omega^2 + \nabla^2\right) G(\mathbf{x}_s, \mathbf{x}, \omega) = \delta(\mathbf{x}_s - \mathbf{x}), \tag{3}$$

where $*$ denotes a convolution operator over the subsurface offset axis (Symes, 2008; Biondi and Almomin, 2012). The first simplification of the extended velocity model is to separate it into a background and a perturbation as follows:

$$v^{-2}(\mathbf{x}, \mathbf{h}) = b(\mathbf{x}, \mathbf{h}) + r(\mathbf{x}, \mathbf{h}), \tag{4}$$

where $b(\mathbf{x})$ is the background component, which is a smooth version of the slowness squared and $r(\mathbf{x})$ is the perturbation component. This separation assumes that $b(\mathbf{x})$ will contain the transmission effects and $r(\mathbf{x})$ will contain the reflection effects. Depending on the error of the initial background velocity, the perturbation component can extend across several subsurface offsets so it is important to keep its offset axis. On the other hand, the background component is not expected to generate reflections that would be grossly time shifted with respect to the recorded data, and it thus safe to restrict it to zero offset. A physical interpretation is that the wave speed is not expected to vary much across different angles, at least in the isotropic case that we are analyzing. Therefore, the extent of background component across subsurface offsets can be reduced. If the velocity is expected to vary, the same reduction can be applied while keeping more than zero subsurface offset. In our derivation, we reduce the background to only the zero subsurface offset as follows

$$v^{-2}(\mathbf{x}, \mathbf{h}) \approx b(\mathbf{x}) + r(\mathbf{x}, \mathbf{h}). \tag{5}$$

After the separation and reduction, the Born approximation can be used to linearize wave equation where the data is assumed to contain primaries only. The linearized wave equation defines the data as follows:

$$d(\mathbf{x}_s, \mathbf{x}_r, \omega; \mathbf{b}, \mathbf{r}(\mathbf{h})) = -\omega^2 f(\omega) \sum_{\mathbf{x},\mathbf{h}} G(\mathbf{x}_s, \mathbf{x} - \mathbf{h}, \omega; \mathbf{b})r(\mathbf{x}, \mathbf{h})G(\mathbf{x} + \mathbf{h}, \mathbf{x}_r, \omega; \mathbf{b}), \tag{6}$$

where the Green's functions now satisfy the conventional acoustic wave equation as follows:

$$\left(\omega^2 b(\mathbf{x}) + \nabla^2\right) G(\mathbf{x}_s, \mathbf{x}, \omega) = \delta(\mathbf{x}_s - \mathbf{x}), \tag{7}$$

$$\left(\omega^2 b(\mathbf{x}) + \nabla^2\right) G(\mathbf{x}, \mathbf{x}_r, \omega) = \delta(\mathbf{x} - \mathbf{x}_r).. \tag{8}$$

The forward modeling can be written in a compact notation as follows:

$$\mathbf{d} = \mathbf{L}(\mathbf{b})\mathbf{r}(\mathbf{h}), \tag{9}$$

where $\mathbf{L}$ is the Born modeling operator. We now define a new objective function for the efficient TFWI inversion as:

$$J_{\text{ETFWI}}(\mathbf{b}, \mathbf{r}(\mathbf{h})) = \|\mathbf{L}(\mathbf{b})\mathbf{r}(\mathbf{h}) - \mathbf{d}_{\text{obs}}\|_2^2 + \epsilon\|\mathbf{A}\mathbf{r}(\mathbf{h})\|_2^2. \tag{10}$$

Notice that there are similarity with the regularized linearized inversion proposed by Clapp (2005), with the important difference that here we are including both the background and the perturbation components as variables in this objective function. The Born modeling operator is linear with respect to perturbation but nonlinear with respect to the background component. Therefore, another linearization around the "background" background is required to compute the gradient. First, we rewrite the background as the sum of two components as follows:

$$b(\mathbf{x}) = b_0(\mathbf{x}) + \Delta b(\mathbf{x}), \tag{11}$$

where $b_0(\mathbf{x})$ is the current background model and $\Delta b(\mathbf{x})$ is the perturbation of the background. The Born approximation is used again to linearize the $\mathbf{L}$ operator with respect to the background resulting in a data-space tomographic operator. The data perturbation with respect to the background perturbation is now defined as:

$$\begin{aligned}\Delta d(\mathbf{x}_s, \mathbf{x}_r, \omega; \mathbf{b}_0, \mathbf{r}(\mathbf{h})) = &\sum_{\mathbf{x},\mathbf{y},\mathbf{h}} \\ &\omega^4 f(\omega) G(\mathbf{x}_s, \mathbf{y} - \mathbf{h}, \omega; \mathbf{b}_0) r(\mathbf{y}, \mathbf{h}) G(\mathbf{y} + \mathbf{h}, \mathbf{x}, \omega; \mathbf{b}_0) \Delta b(\mathbf{x}) G(\mathbf{x}, \mathbf{x}_r, \omega; \mathbf{b}_0) + \\ &\omega^4 f(\omega) G(\mathbf{x}_s, \mathbf{x}, \omega; \mathbf{b}_0) \Delta b(\mathbf{x}) G(\mathbf{x}, \mathbf{y} - \mathbf{h}, \omega; \mathbf{b}_0) r(\mathbf{y}, \mathbf{h}) G(\mathbf{y} + \mathbf{h}, \mathbf{x}_r, \omega; \mathbf{b}_0),\end{aligned} \tag{12}$$

where $\mathbf{y}$ is the perturbation coordinates. As we can see in the previous equation, the tomographic operator correlates a background and a scattered wavefield from both source and receiver sides. The scattered wavefields are computed by correlating a background wavefield with the reflectivity model $\mathbf{r}$ and then propagating again to all model locations. The forward tomographic operator can be written in a compact notation as follows:

$$\Delta\mathbf{d} = \frac{\partial \mathbf{L}}{\partial \mathbf{b}}\mathbf{r}(\mathbf{h})\Delta\mathbf{b} = \mathbf{T}\Delta\mathbf{b}, \tag{13}$$

Where $\mathbf{T}$ is the tomographic operator that relates changes in the background model to changes in the data. We can now compute the reflectivity gradient as follows:

$$g_{\mathbf{r}}(\mathbf{x}, \mathbf{h}) = \frac{\partial J}{\partial \mathbf{r}(\mathbf{h})} = \mathbf{L}^*\Delta\mathbf{d}, \tag{14}$$

which is simply migration of the residuals. Then, we compute the background gradient as follows:

$$g_{\mathbf{b}}(\mathbf{x}) = \frac{\partial J}{\partial \mathbf{b}} = \mathbf{T}^* \Delta \mathbf{d}. \tag{15}$$

It is important to notice that reflectivity has two roles in computing the background model gradient. First, it is used to computed the data residuals. Second, it is used to scatter the background wavefields and compute the scattered wavefields of the tomographic operator.

SCALE MIXING

A direct use of the gradients is to update their corresponding models directly. However, this would hinders the simultaneous inversion of different wavelengths of the model. Hence, we first mix the two gradients and then separate them in Fourier domain to get the update of each model as follows:

$$s_{\mathbf{b}}(\mathbf{x}) = \mathbf{C_b}(g_{\mathbf{b}}(\mathbf{x}) + g_{\mathbf{r}}(\mathbf{x}, \mathbf{h} = 0)), \tag{16}$$

where $s_{\mathbf{b}}(\mathbf{x})$ is the search direction of the background model and $\mathbf{C_b}$ is a low-pass filter. Similarly, we can compute the update of the perturbation model as:

$$s_{\mathbf{r}}(\mathbf{x}, \mathbf{h}) = \mathbf{C_r}(g_{\mathbf{b}}(\mathbf{x}) + g_{\mathbf{r}}(\mathbf{x}, \mathbf{h})), \tag{17}$$

where $s_{\mathbf{r}}(\mathbf{x}, \mathbf{h})$ is the search direction of the perturbation model and $\mathbf{C_r}$ is a high-pass filter. In order to sum the two gradients properly, both of them need to have the same units as well as the same scale. This requires careful implementation of each operator at each linearization.

For the examples shown in this paper, we used a radial cut-off in the Fourier domain with a cosine squared taper. The wavelength cut-off is based on the dominant frequency in the data as well as the average velocity of the initial model. Also, the two filters were designed such that they always sum to one at all wavelength to maintain the energy of the gradients.

SYNTHETIC EXAMPLES

Gaussian Model

For the first of synthetic example, we use a two-layer model with a Gaussian anomaly. The velocities for the top layer, bottom layer and peak of the Guassian anomaly are 3000 m/s, 3500 m/s and 3500 m/s, respectively. Figure 1 shows the true velocity model. A Ricker wavelet with a fundamental frequency of 15 Hz and temporal sampling of 1.5 ms is used as a source function to model the data. There are 151 fixed

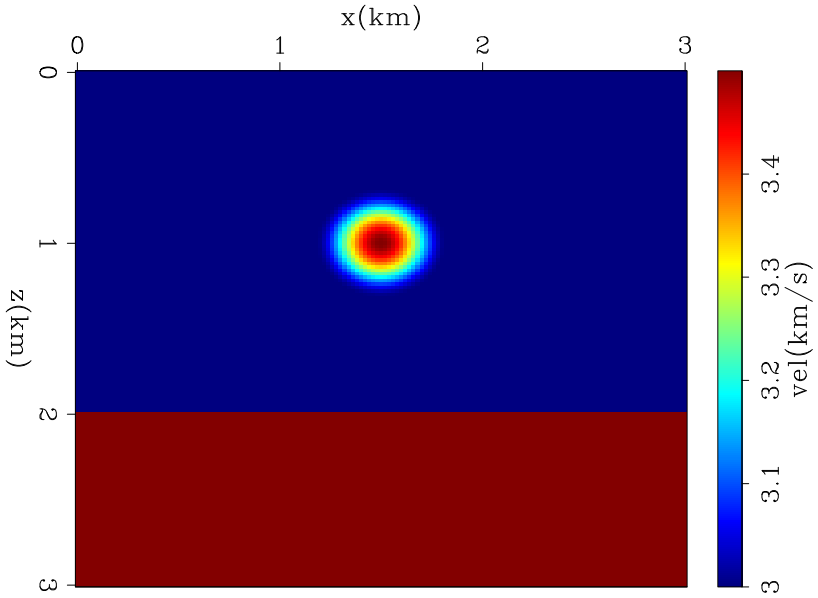


Figure 1: The true velocity of Gaussian model. [**ER**]

receivers with a spacing of 20 m and 31 sources with a spacing of 100 m. The maximum offset used is 1500 m in both sides and the initial model is a constant model of 3000 m/s velocity.

We first modeled the observed data with the conventional acoustic wave-equation nonlinear modeling operator. Then, we computed the initial data using the same nonlinear operator on the initial model. The two datasets are subtracted from each other to remove the direct arrivals. We then use the subtracted data as the "observed data" in equation 10.

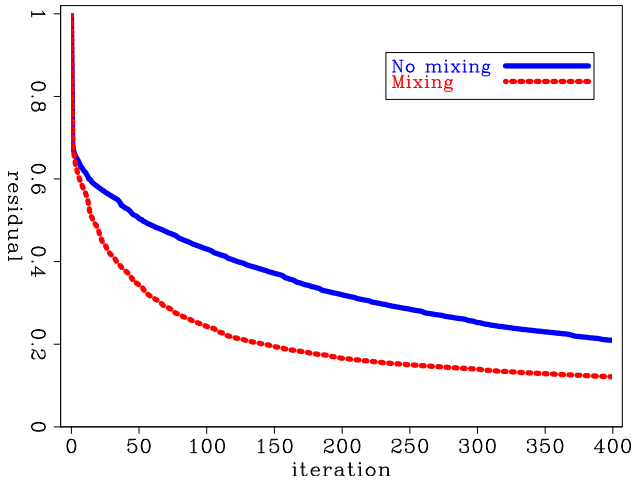


Figure 2: The data residual norm as a function of iterations of Gaussian model inversion. [**CR**]

We show the results of running 400 iterations of minimizing equation 10 with and without the scale mixing described in equations 16 and 17. Figure 2 shows the normalized residual of the data fitting part, described by the first term of equation 10,

as a function of iterations. The residuals of both inversions decrease monotonically without getting stuck in a local minima. However, scale mixing shows a much faster convergence rate.

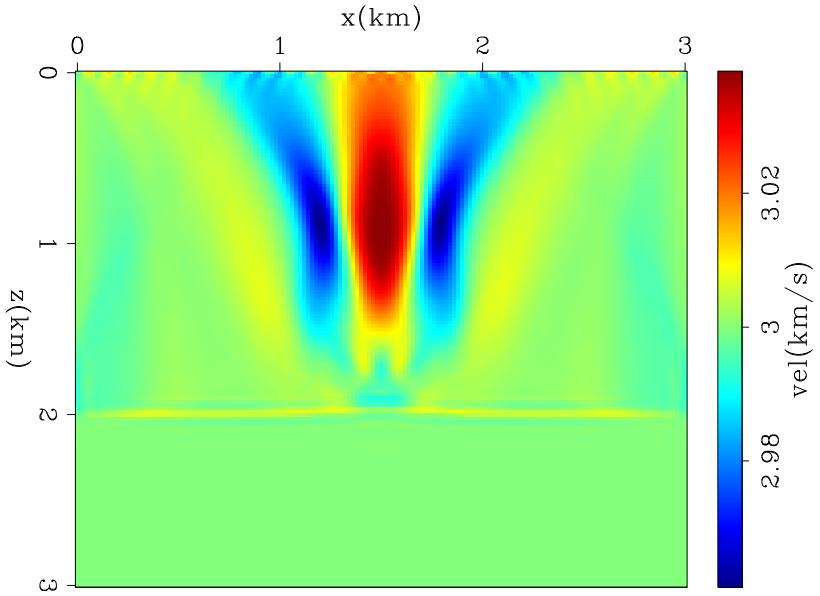


Figure 3: The inverted background of the Gaussian model example. [**CR**]

Figures 3 and 4 shows the final background and reflectivity models without scale mixing. The background has a low vertical resolution with relatively strong side-lobes around the anomaly location. The reflectivity model has a low horizontal resolution where the sides of the anomaly are not well illuminated. This lack of resolution in both models is expected due to the limited acquisition.

Figures 5 and 6 shows the final background and reflectivity models with scale mixing. The background has a much better vertical resolution that locates the anomaly at the correct depth. Moreover, the peak amplitude of the anomaly is more than three times stronger compared to Figure 3 while the side-lobes remained at the same strength. The reflectivity model shows substantial improvements in resolution as well. The sides of the anomaly seem to be better illuminated after scale mixing. Although reflectivity is linear with respect to the data, the inverted model seems to benefit from the scale mixing. Therefore, the contribution of the background gradient seem to add to the null space components of the reflectivity model, and vice-versa.

Marmousi Model

A modified Marmousi model is used for the next synthetic example where 500m of water layer is added to the top. Figure 7 shows the true velocity model. A Ricker wavelet with a fundamental frequency of 15 Hz and temporal sampling of 1.5 ms is used as a source function to model the data. There are 461 fixed receivers with a

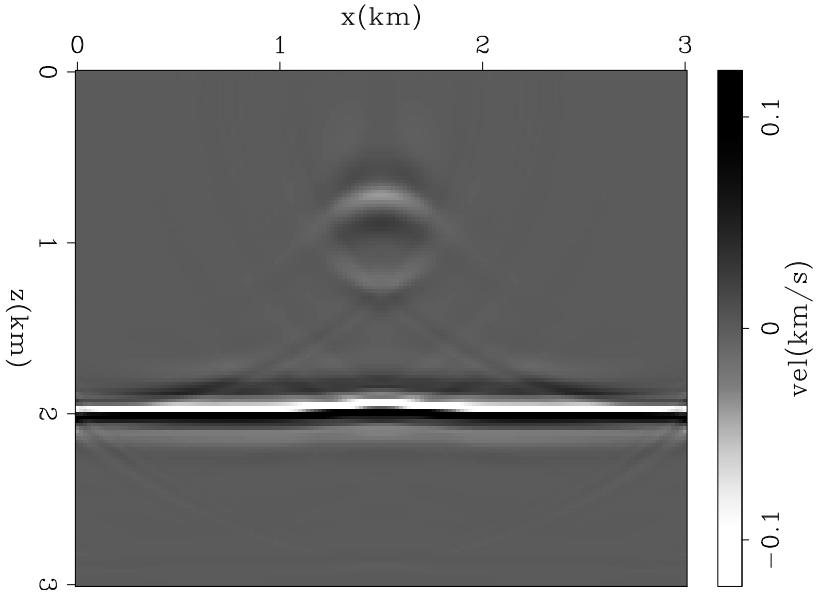


Figure 4: The inverted reflectivity of the Gaussian model example. [**CR**]

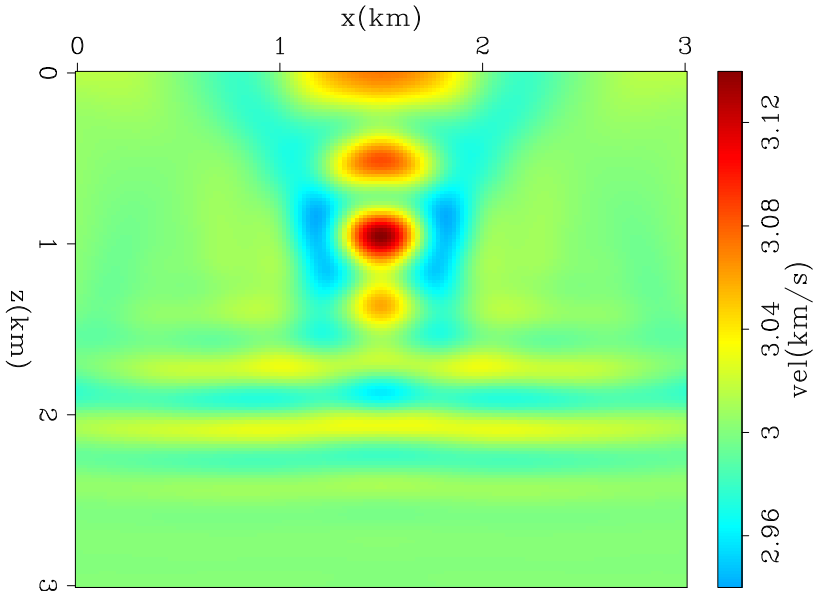


Figure 5: The inverted background of the Gaussian model example with mixing of scales. [**CR**]

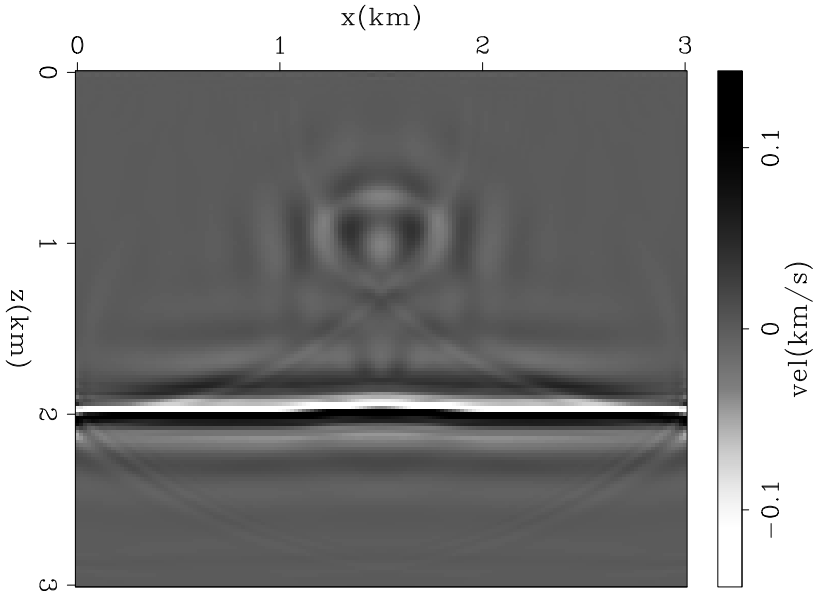


Figure 6: The inverted reflectivity of the Gaussian model example with mixing of scales. [**CR**]

spacing of 20 m and 93 sources with a spacing of 100 m. The initial model is shown in Figure 8 which is obtained by strongly smoothing the true model laterally. The conventional FWI fails when starting from this initial model (Biondi and Almomin, 2012).

Similar to the Gaussian model, we modeled the observed data with the conventional acoustic wave-equation nonlinear modeling operator. Then, we computed the initial data using the same nonlinear operator on the initial model shown in Figure 8. The two datasets are subtracted from each other to remove the direct arrivals. We then use the subtracted data as the "observed data" in equation 10.

We show the results of running 600 iterations of minimizing equation 10 where the gradients are mixed as described in equations 16 and 17. Figure 9 shows the normalized residual of the data fitting part, described by the first term of equation 10, as a function of iterations. The residuals decrease monotonically without getting stuck in a local minima. Figure 10 shows three constant-location sections of the final reflectivity model that are a function of depth and subsurface offset. The three images are focused around the middle, i.e. the zero subsurface offset, which indicates convergence towards the correct background model.

Figure 11 shows the final background model. It shows great resolution in both the vertical and horizontal directions that captures the main features of the true model. Figure 12 shows zero subsurface offset of the final reflectivity model. Finally, Figure 13 shows the total inverted model which is the sum of the background and reflectivity. It shows remarkable convergence towards the true model.

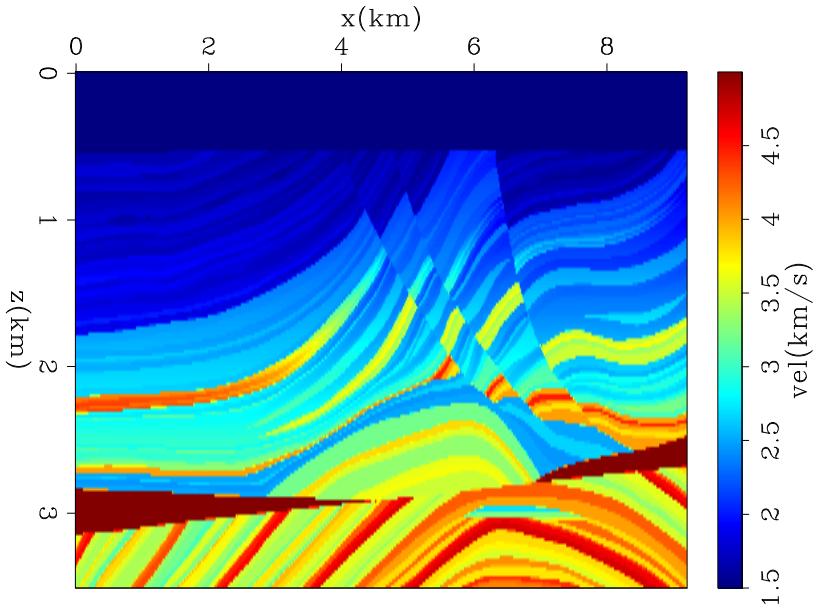


Figure 7: The true velocity of marmousi model. [**ER**]

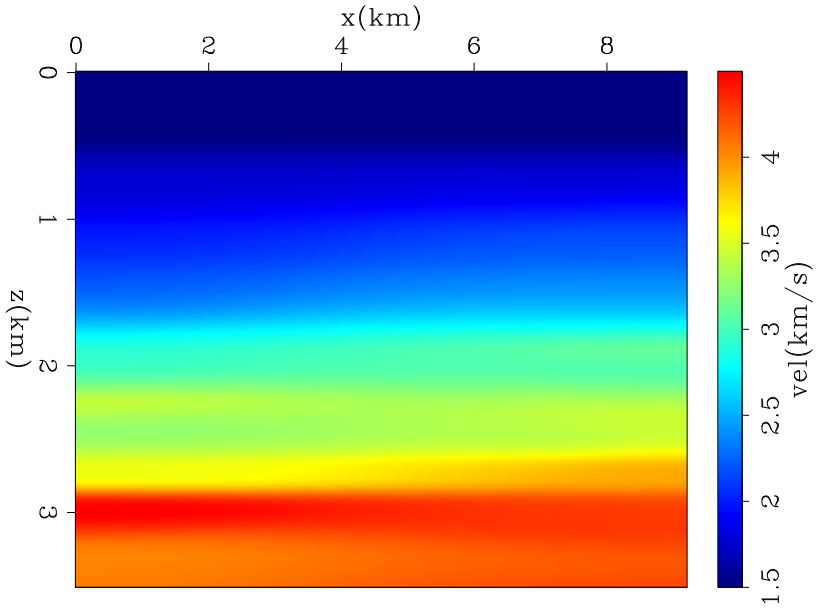


Figure 8: The initial velocity of marmousi model. [**ER**]

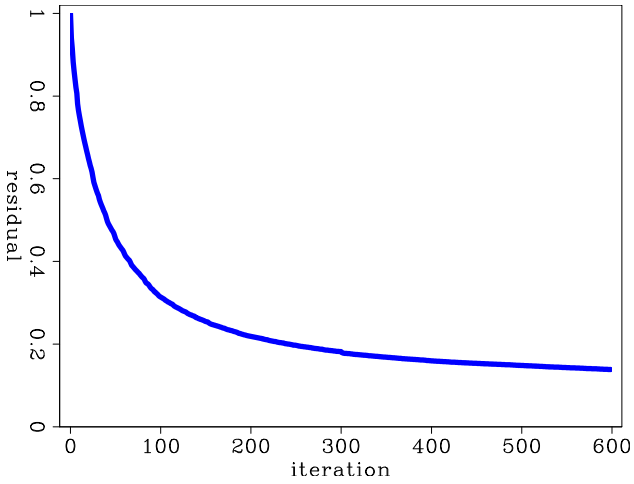


Figure 9: The data residual norm as a function of iterations of marmousi model inversion. [**CR**]

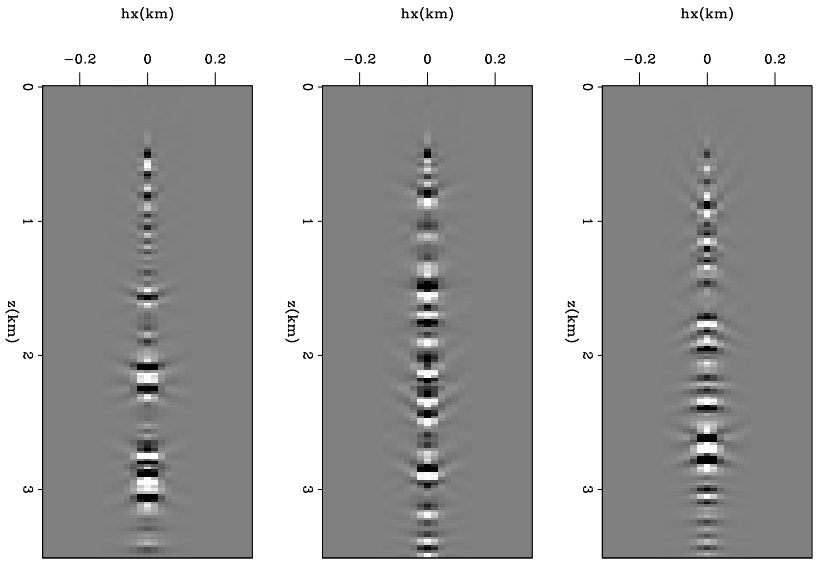


Figure 10: Three offset-domain common-image gathers of the inverted reflectivity model at x=2.5, 5, 7.5 km of marmousi model. [**CR**]

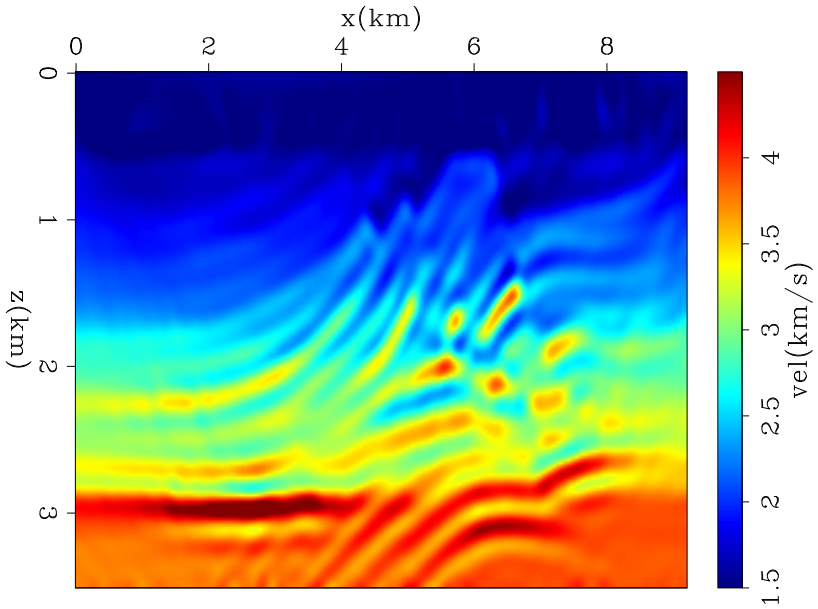


Figure 11: The inverted background of marmousi model. [**CR**]

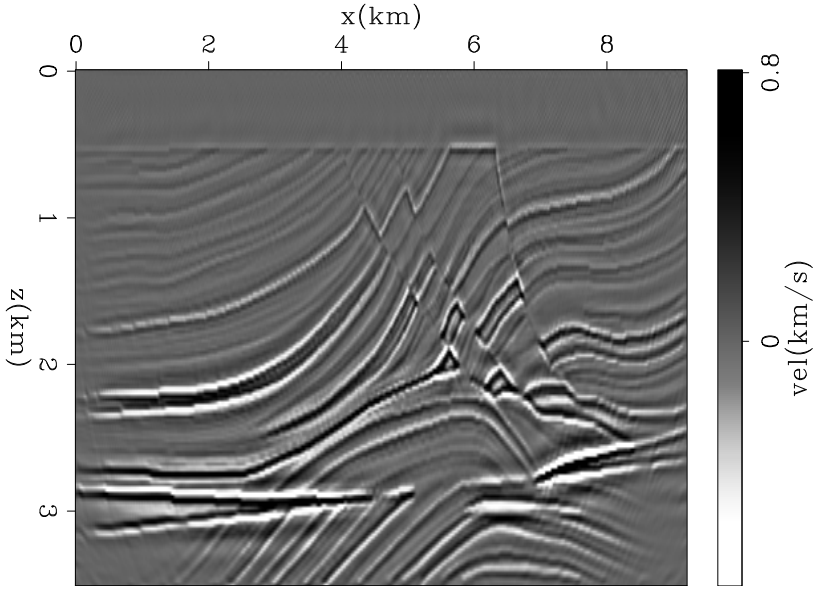


Figure 12: The inverted reflectivity at h=0 of marmousi model. [**CR**]

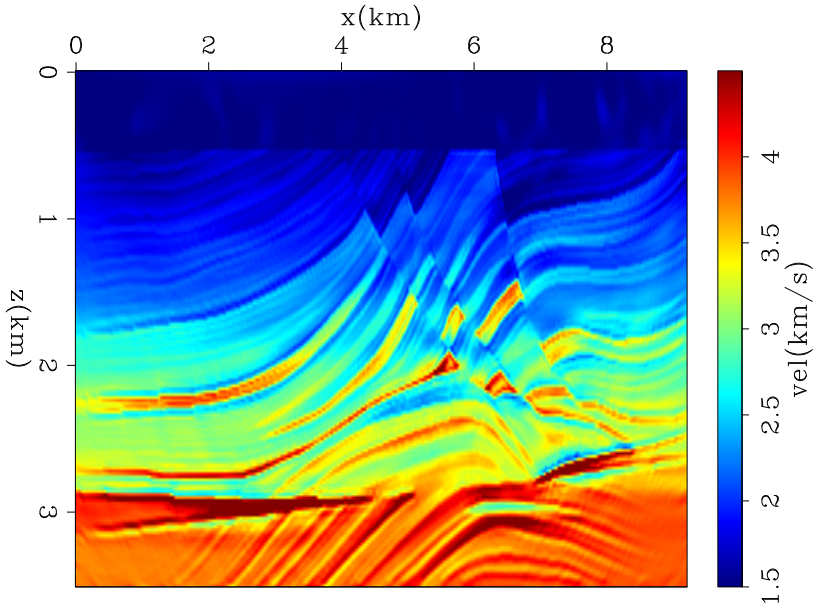


Figure 13: The total inversion results of marmousi model. [**CR**]

CONCLUSIONS

We presented a practical approach to tomographic full-waveform inversion where the computational cost is significantly reduced. This was achieved by first breaking the model into a background component and a perturbation component, and then by restricting the offset axis of the background component to zero subsurface offset only. Breaking the model into two components assumes the data contain primary only. However, we managed to maintain the simultaneous inversion of different wavelengths of the model by mixing the gradients of the two components in Fourier domain using a high-pass and a low-pass filters. The synthetic examples show remarkable results even when the initial model had large errors. More sophisticated mixing schemes need to be further investigated.

ACKNOWLEDGMENTS

We would like to thank the Stanford Exploration Project affiliate companies for financial support. Almomin would like the thank Saudi Aramco for supporting his Ph.D. studies at Stanford University.

REFERENCES

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEP-Report, **147**, 1–12.

Biondi, B. and P. Sava, 1999, Wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, **18**, 1723–1726.

Clapp, M. L., 2005, Imaging under salt: Illumination compensation by regularized inversion: PhD thesis, Stanford University.

Luo, Y. and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

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

Shen, P., 2004, Wave equation migration velocity analysis by differential semblance optimization: PhD thesis, Rice University.

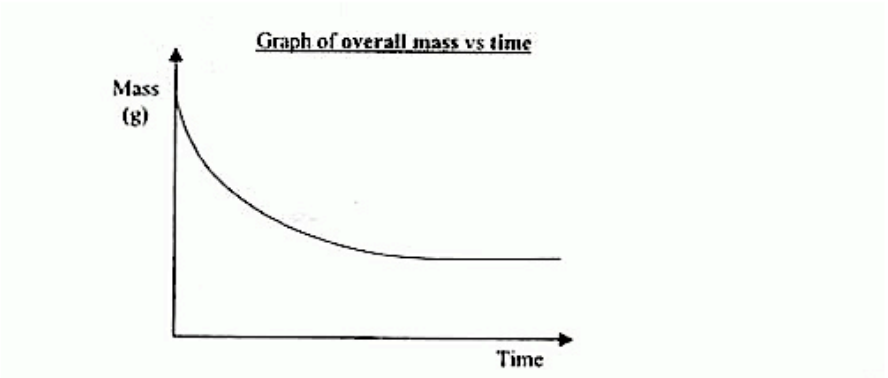
Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Symes, W. W. and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Tang, Y., C. Guerra, and B. Biondi, 2008, Image-space wave-equation tomography in the generalized source domain: SEP-Report, **136**, 1–22.

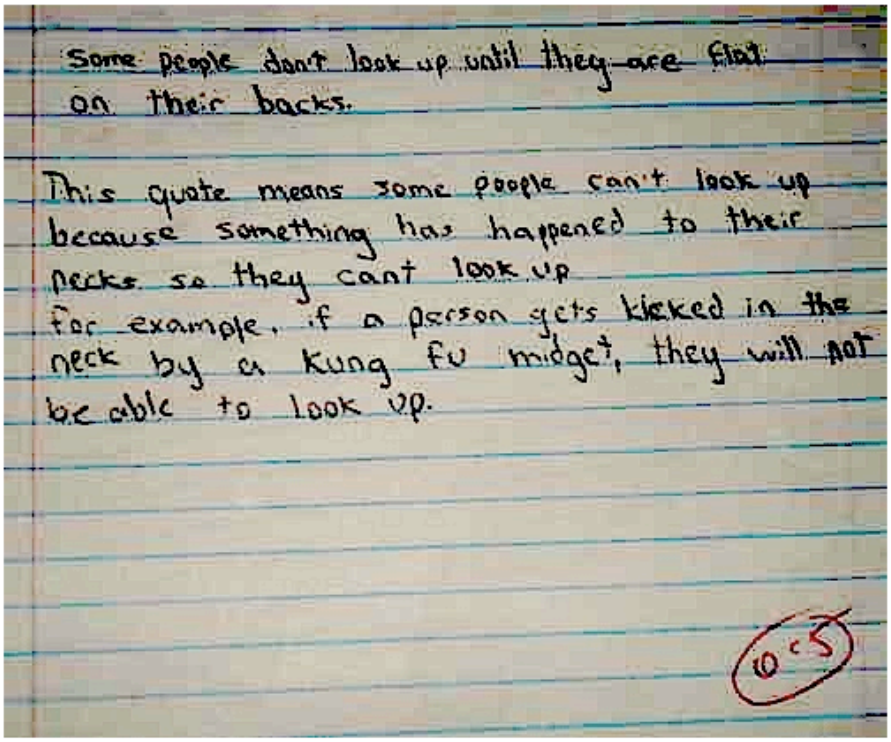
Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Zhang, Y., B. Biondi, and Y. Tang, 2012, Residual moveout-based wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, **31**, *submitted for publication.*



1. **Explain the shape of the graph.**

It's curvy, with a higher bit at the end and a rather aesthetically pleasing slope downards towards a pretty flat strait bit. The actual graph itself consists of 2 strait lines meeting at the lower left hand corner of the graph and moving away at a 90° angle. Each line has an arrow head on the end.



Some people don't look up until they are flat on their backs.

This quote means some people can't look up because something has happened to their necks so they cant look up for example, if a person gets kicked in the neck by a Kung fu midget, they will not be able to look up.

10/5