Tomographic full waveform inversion (TFWI) by extending the velocity model along the time-lag axis

Biondo Biondi and Ali Almomin

ABSTRACT

Convergence of full waveform inversion can be improved by extending the velocity model along either the subsurface-offset axis or the the time-lag axis. The extension of the velocity model along the time-lag axis enables us to linearly model large time shifts caused by velocity perturbations. The extension is based on a new linearization of the scalar wave equation where the extended-velocity perturbation is convolved in time with the Laplacian of the background wavefield. This linearization is accurate for both reflected events and transmitted events, and in particular for diving waves recorded at large offsets. The modeling capabilities of the proposed linearization enable the simultaneous inversion of reflections and diving waves even when the starting velocity model is far from being accurate. Numerical tests performed on synthetic data modeled on the "Caspian Sea" portion of the well-known BP model shows the global-convergence properties as well the high-resolution potential of the method.

INTRODUCTION

Conventional seismic imaging relies on a separation of scales between migration velocity model (long-wavelength components) and reflectivity (short-wavelength components). Figure 1 shows a simplified 1D graphical representation of the separation of scales concept. The black line represents the two disjointed wavelength ranges (mapped into corresponding temporal-frequency bands) and the consequent gap in information between long wavelengths and short wavelengths. This conceptual understanding leads naturally to a sequential approach for seismic imaging; the velocity model is estimated first, and then it is used as input to migration for imaging reflectivity. In current velocity-estimation practice, reflectivity is used only indirectly to measure the focusing power of the velocity model. The only important exceptions occur when migrated volumes are used to interpret boundaries of geobodies (e.g. salt bodies) and to estimate predominant dips in the geologic layering that are then used to constraint a tomographic velocity update.

The sequential imaging process is slowly being undermined by three long-standing trends in the industry: 1) acquisition of lower-frequency data, 2) imaging under

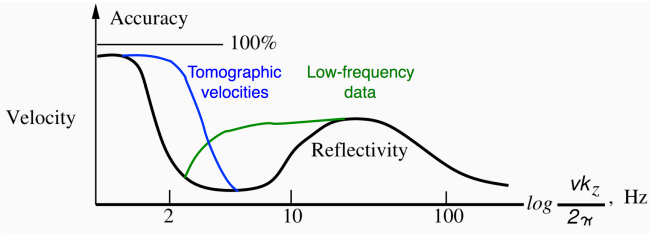

Figure 1: Simplified 1D graphical representation of the separation of scales in seismic imaging (black line) and how current industry trends are narrowing the gap between the estimation of long wavelengths and short wavelengths (blue and green lines). (Adapted from Jon Claerbout's *Imaging the Earth Interior*. [**NR**]

complex overburden which requires higher-resolution velocity models to focus and correctly position reflectors, and 3) acquisition of longer-offset data. As the industry strives to widen the data frequency band at both the low and high end, the reflectivity band is extended at the low end, as graphically represented by the green line in Figure 1. The high end of the velocity band is also pushed upward (blue line in Figure 1) by the application of sophisticated tomographic methods that enable the estimation of the high-resolution velocity models needed to focus reflectors located under complex overburden. However, tomography (either ray or wave-equation based) is a more challenging task than migration, and thus often it falls short of providing the accuracy and resolution necessary to satisfactorily image the high frequencies in the data. The acquisition of longer-offset data enables the recording of diving waves and refracted arrivals that provide a complementary illumination of the velocity components in the crucial scale gap and blur the distinction between migration and tomography since they contain forward-scattering perturbations to transmission events.

As the information gap narrows, imaging methods that simultaneously estimate the velocity and reflectivity model by taking advantage of all the information in the data are becoming more attractive. The renewed interest in full waveform inversion (FWI) can be explained as an attempt to overcome the limitations imposed by the sequential imaging approach, as well as the availability of computational power sufficient for practical applications of FWI. FWI has been the most successful when applied to the low frequencies in the data (green line in Figure 1) to improve the velocity-model estimation needed to image the data high frequencies under complex overburden. FWI has been less successful in using the high-frequencies in the data to tomographically estimate the long-wavelengths in the model.

Since the eighties (see for example (Mora, 1989)) it has been recognized that FWI has both a migration component and a tomographic component. However, to ensure convergence of the tomographic component the recorded and modeled data must be almost in phase with each other; the rule of thumb being that the residual time-shifts must be shorter than the half period of the dominant frequency in the data. Bootstrapping the inversion by starting from the low frequencies may ameliorate the convergence problems, but it still depends on conventional velocity estimation methods

to deliver starting models sufficiently accurate to satisfy the convergence criterion for the FWI tomographic component. It also undermines the goal of simultaneous estimation because the high frequencies in the data contain the high-resolution tomographic information that facilitates the estimation of the velocity components represented by the blue line in Figure 1.

To perform true simultaneous and synergistic inversion for all the model scales we must address the convergence problems of the tomographic term in FWI. These convergence problems are related to the non-linearity of the wave-equation with respect to perturbations in the long wavelengths of the velocity model. Long-wavelengths perturbations cause substantial time shifts of the propagating wavefields that are poorly approximated by the linearization of the wave equation based on first order Born approximation. In this paper we introduce a linearization of the wave equation based on an extension of the velocity model along the time lag axis (τ). This extension enables the linear modeling of large time shift in the propagating wavefields, and consequently in the data. Based on this extension we define an objective function that has a model-focusing term in addition to the conventional FWI data-fitting term. Numerical examples with realistically complex velocity models demonstrate that this objective function has excellent convergence behavior, although currently convergence is unsatisfactorily slow.

The usefulness of extending the reflectivity model (as prestack images in the angle or offset domain) to manage the non-linearities in wave-equation velocity analysis was demonstrated in the context of differential semblance optimization (DSO) (Symes and Carazzone, 1991; Shen and Symes, 2008) and wave-equation migration velocity analysis (WEMVA) (Biondi and Sava, 1999; Sava and Biondi, 2004a,b). The generalization of extending the model to the propagation component of the model (long wavelength) was first introduced by Symes (2008) and more recently successfully applied by Sun and Symes (2012); Almomin and Biondi (2012); Biondi and Almomin (2012). These methods are based on a velocity extension along the subsurface-offset or plane-wave axes. We propose an extension along the time lag axis (τ) because it is better suited to describe large time shifts in wave propagation, that are at the root of FWI convergence problems. Furthermore, extending the velocity along the time lag axis can easily handle forward-scattered events recorded at long offsets as well as the reflections recorded at near and intermediate offsets. We thus believe that the time-lag formulation has the potential to deliver high-quality results also for modern long-offset data sets. Furthermore, a one-dimensional extension along time is computationally more efficient than a two-dimensional extension along subsurface offsets. Yang and Sava (2009, 2010) have discussed the use and the computational advantages of time-lag gathers to perform WEMVA for reflected events.

TOMOGRAPHIC FULL WAVEFORM INVERSION (TFWI)

Conventional full waveform inversion in the acoustic constant-density approximation is performed by solving the following optimization problem

$$\min_{\mathbf{v}^2} J_{\mathrm{FWI}}\left(\mathbf{v}^2\right) \tag{1}$$

where:

$$J_{\mathrm{FWI}}\left(\mathbf{v}^2\right) = \frac{1}{2}\left\|\mathcal{L}\left(\mathbf{v}^2\right) - \mathbf{d}\right\|_2^2, \tag{2}$$

$\mathbf{v} = v\left(\vec{x}\right)$ is the velocity vector, $\mathcal{L}$ is a wave-equation operator non linear with respect to velocity perturbations and the data vector $\mathbf{d}$ is the pressure field $\mathbf{P} = P\left(t, \vec{x}\right)$ measured at the surface.

The wave-equation operator is evaluated by recursively solving the following finite difference equation

$$\left[\mathbf{D_2} - \mathbf{v}^2\nabla^2\right]\mathbf{P} = \mathbf{f}, \tag{3}$$

where $\mathbf{D_2}$ is a finite-difference representation of the second derivative in time, ∇^2 is a finite-difference representation of the Laplacian, and $\mathbf{f}$ is the source function.

Efficient solution of the optimization problem expressed in equation 1 is performed by gradient-based methods, and thus requires the evaluation of the linear operator $\mathbf{L}$, which is the linearization of $\mathcal{L}$ with respect to velocity perturbations $\delta\mathbf{v}^2$. This linear operator can be derived by perturbing equation 3 as follows

$$\left[\mathbf{D_2} - \left(\mathbf{v_o}^2 + \delta\mathbf{v}^2\right)\nabla^2\right]\left(\mathbf{P_o} + \delta\mathbf{P}\right) = \mathbf{f}, \tag{4}$$

where $\mathbf{P_o}$ and $\mathbf{v_o}$ are the background wavefield and velocity, respectively, and $\delta\mathbf{P}$ is the scattered wavefield.

Equation 4 can be rewritten as the following two equations:

$$\begin{aligned}\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\mathbf{P_o} &= \mathbf{f}, &(5)\\ \left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\delta\mathbf{P} &= \delta\mathbf{v}^2\nabla^2\left(\mathbf{P_o} + \delta\mathbf{P}\right), &(6)\end{aligned}$$

which represent a nonlinear relationship between velocity perturbations and scattered wavefield. In conventional FWI, to linearize this relationship we drop the term multiplying the perturbations with each other; that is, we drop the scattered wavefield from the right hand side of equation 6 and obtain the following coupled equations:

$$\begin{aligned}\left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\mathbf{P_o} &= \mathbf{f}, &(7)\\ \left[\mathbf{D_2} - \mathbf{v_o}^2\nabla^2\right]\delta\mathbf{P} &= \delta\mathbf{v}^2\nabla^2\mathbf{P_o}. &(8)\end{aligned}$$

The linear operator $\mathbf{L}$ used to compute the gradient of the FWI objective function 2 is evaluated by recursively propagating the background wavefield $\mathbf{P_o}$ and the scattered wavefield $\delta\mathbf{P}$ by solving equations 7–8. The scattered wavefield $\delta\mathbf{P}$ is now a linear

function of the velocity perturbations $\delta\mathbf{v}^2$, equation 8 has the limitation that it takes into account only first-order scattering, and thus it is unsuited to model large time shifts between the background wavefield and the scattered wavefield.

To improve the capability of the linearization to model large time shifts, we extend the velocity model along the time lag axis τ and convolve its perturbations $\delta\tilde{\mathbf{v}}^2(\tau)$ with the Laplacian of the background wavefield; we rewrite equation 6 as:

$$\left[\mathbf{D_2} - \tilde{\mathbf{v}}^2(\tau=0)\nabla^2\right]\delta\mathbf{P} = \delta\tilde{\mathbf{v}}^2(\tau) \overset{\tau}{*} \nabla^2\mathbf{P_o}, \tag{9}$$

which defines the linear operator $\widehat{\mathbf{L}}$ and where $\overset{\tau}{*}$ denotes convolution in τ. The combination of the wave equation non-linear operator $\mathcal{L}(\mathbf{v})$ and of the linear operator $\widehat{\mathbf{L}}$ defines the extended non-linear operator

$$\tilde{\mathcal{L}}(\tilde{\mathbf{v}}) = \mathcal{L}(\tilde{\mathbf{v}}(\tau=0)) + \widehat{\mathbf{L}}(\tilde{\mathbf{v}}(\tau=0))\,\delta\tilde{\mathbf{v}}^2, \tag{10}$$

that can be used to define the TFWI objective function as

$$J_{\text{TFWI}}(\tilde{\mathbf{v}}) = \frac{1}{2}\left\|\tilde{\mathcal{L}}(\tilde{\mathbf{v}}) - \mathbf{d}\right\|_2^2 + \epsilon\,\||\tau|\,\tilde{\mathbf{v}}\|_2^2. \tag{11}$$

The second term in the equation 11 rewards focusing of the data around zero time lag. It introduces a strong tomographic component, which is necessary to constrain the optimization problem because the velocity extension relaxes the constraints on the modeled data kinematics imposed by the data fitting term in equation 11. This objective function can be minimized using the nested optimization algorithm with scale mixing that we discussed in Almomin and Biondi (2013).

We will use a simple 1D numerical example to analyze some of the characteristics of the TFWI method we introduced above. Figure 2 shows the difference between the background wavefield propagated with $\mathbf{v}$=1.2 km/s and the wavefield propagated with the true velocity of $\mathbf{v}$=1.13 km/s. The source function is a zero-phase wavelet bandlimited between 5 and 20 Hz. The difference wavefield is displayed as a function of propagation distance and traveltime. The velocity error is sufficiently high that the wavefields are completely out of phase after propagating for a couple of kilometers. This is therefore a situation like the ones described where the first order Born linearization (equation 8) would fail to to model the data residuals and conventional FWI would have troubles to converge, even if the problem is extremely simple.

Figure 3 shows the conventional FWI objective functions when the data are recorded with a single receiver located at 7 km for a total of 4 km offset from the source. The plot shows the initial value of the objective function for several 1D transmission problems sharing the same starting velocity (1.2 km/s) and with different true velocities. If the true velocity is lower than $\approx$ 1.18 km/s or larger than $\approx$ 1.22 km/s a gradient-based method starting from a velocity of 1.2 km/s will not converge to the right solution. An explanation for the lack of convergence of a gradient-based method is that conventional first-order Born approximation cannot model the data residuals when the true velocity is outside this range. On the contrary, the linearized

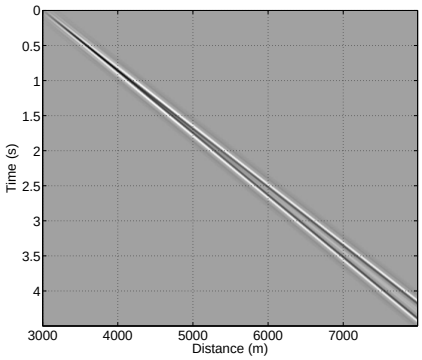

Figure 2: Difference between background wavefield computed with the starting velocity (1.2 km/s) and the wavefield propagated with the true velocity (1.13 km/s). [**CR**]

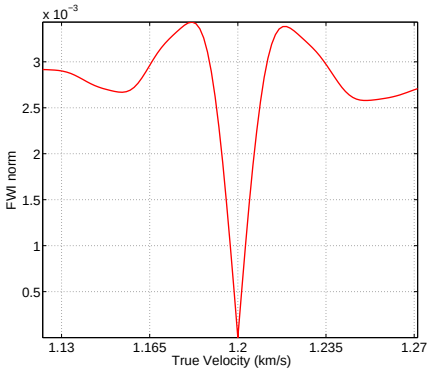

Figure 3: FWI norm as a function of the true velocity, when the starting velocity is equal to 1.2 km/s. [**CR**]

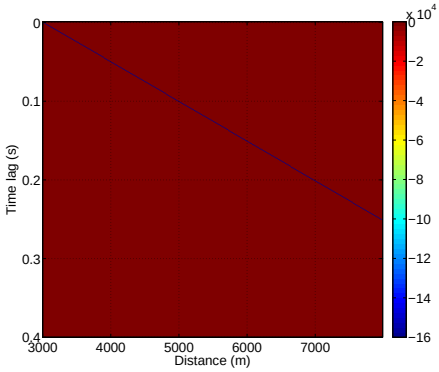

Figure 4: Extended velocity perturbation chosen to approximately model the wavefield difference shown in Figure 2. [**CR**]

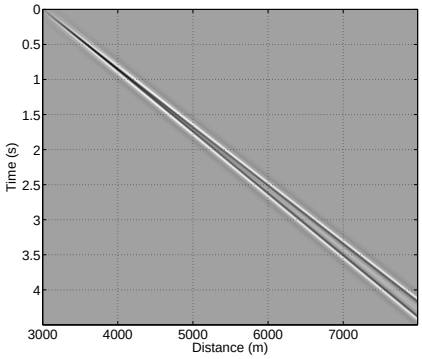

Figure 5: Perturbed wavefield computed by solving equation 9 with the model shown in Figure 4. [**CR**]

modeling equation defined in equation 9 would have no troubles to model the data residual.

For example, we can easily reproduce the wavefield difference shown in Figure 2 by setting the extended-velocity perturbation to be a delta function along the τ axis, where the shift of the delta function linearly increases with the distance from the origin. This linear shift is computed by integrating the difference in slowness between the background model and the true model. The extended-velocity perturbation is shown in Figure 4. Figure 5 shows the result of solving equation 9 with the model shown in Figure 4. The approximation of the scattered wavefield $\Delta\mathbf{P}$ is almost identical to the wavefield difference shown in Figure 2.

DIVING-WAVE MODELING EXAMPLE

One of the advantages of extending the velocity model along the time-lag axis τ is the capability to model with a linear operator large time shifts in the diving waves recorded by modern long-offset acquisition geometries. The capability of modeling time shifts in these events enables robust convergence of the inversion even when the starting velocity model is far from the correct one. To show these modeling capabilities we use one long-offset shot profile recorded over a half space with a vertical velocity gradient. The starting velocity model is assumed to be uniform and equal to the velocity at the surface. Figure 6a shows the data residual; both the recorded diving wave as well as the data modeled with the starting velocity are clearly visible.

The backprojection of data residuals shown Figure 6a by the application of $\widehat{\mathbf{L}}'$ generates the velocity perturbation cube shown in Figure 7. The front panel of the cube shown in Figure 7 displays the zero time lag of the velocity perturbations. A substantial amount of the energy in the residual has been backprojected away from the zero time-lag panel.

Figure 6c displays the result of forward modeling the data residuals by the application of $\widehat{\mathbf{L}}$ to the extended velocity perturbation shown in Figure 7. The kinematics of these modeled residuals are very close to the kinematics of the true residuals shown Figure 6a. In contrast, when only the zero time lag of the velocity perturbation (front panel of the cube shown in Figure 7) is used to model the data residuals, we obtain the seismograms displayed in Figure 6b. The diving wave is totally missing from these modeled residuals because the background wavefield propagates with constant velocity along the horizontal direction.

Figure 8 shows the velocity perturbation cube when the velocity is extended along the horizontal subsurface offset axis. The front panel of the cube displays the zero subsurface offset, and thus it is identical to the front panel of the cube shown in Figure 7. Figure 6d displays the result of forward modeling the data residuals starting from the extended velocity perturbation shown in Figure 8. The diving wave event is present in these modeled residuals. However, it dies out at larger offsets, starting at about 8 kilometers offset. A subsurface offset extension of the velocity has difficulties

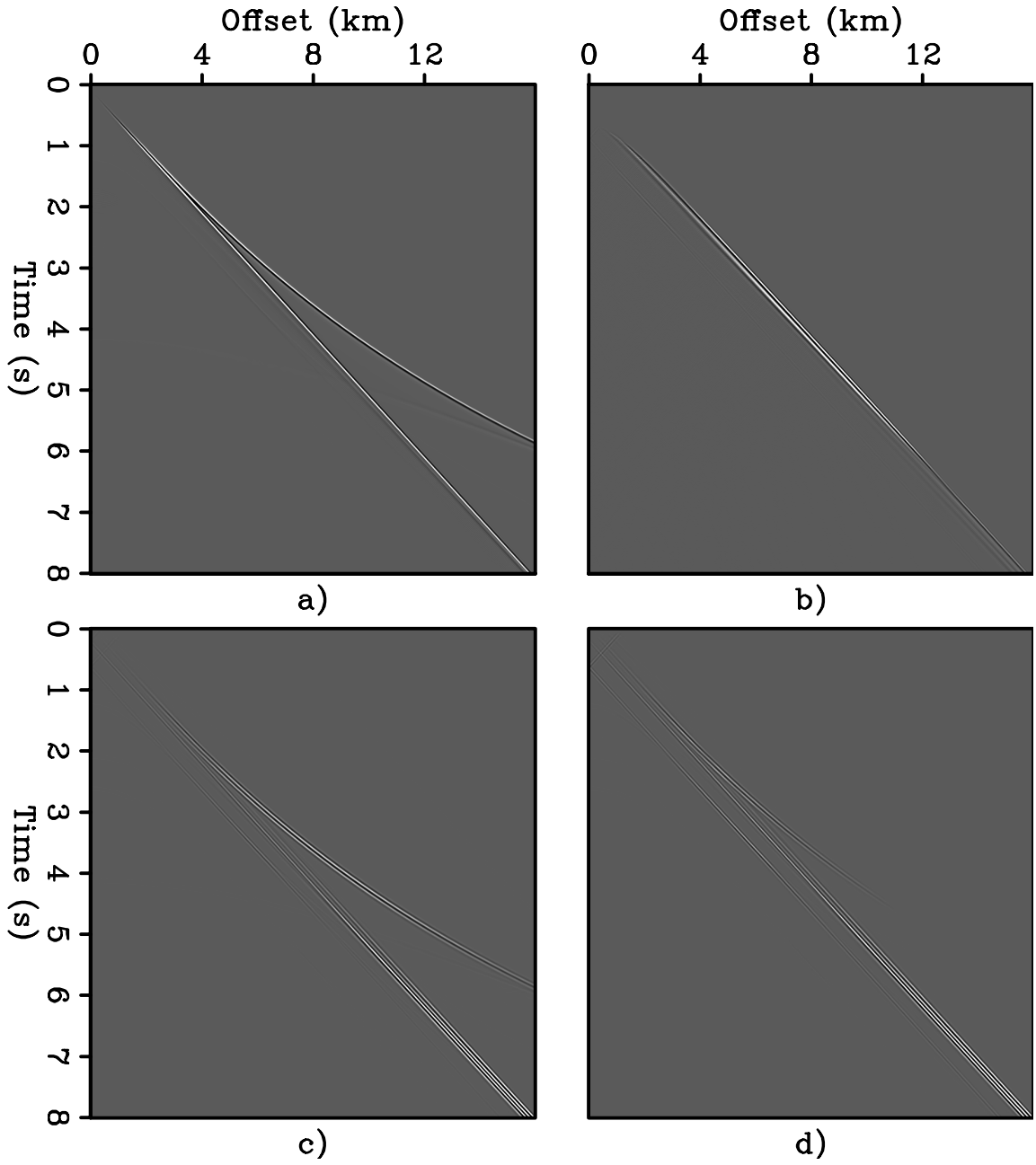

Figure 6: a) Difference between background wavefield computed with the starting velocity and the wavefield propagated with the true velocity, b) data residual modeled from zero lag of the velocity perturbation cubes (front panel in cubes shown in both Figure 7 and Figure 8), c) data residual modeled from the velocity perturbations extended along the time-lag axis, (Figure 7), and d) data residual modeled from the velocity perturbations extended along the horizontal subsurface offset axis, (Figure 8). [**CR**]

in modeling large time shifts in transmitted events propagating in directions orthogonal to the subsurface offset axes. Consequently, as the propagation paths of the diving waves have longer vertical components, the less accurate the modeled residuals are. To address this limitation we could use vertical subsurface offset in addition to the horizontal ones (Biondi and Symes, 2004), but the computational cost would increase accordingly.

TFWI OF LONG-OFFSET DATA

To verify the capabilities of the TFWI method based on time-lag extension of the velocity model we tested the method on a synthetic data set recorded with long offsets. The data were generated over the "Caspian Sea" portion of the well-known BP velocity model, as shown in Figure 9. The receiver array was assumed to be fixed at the surface, and thus data with more than 20 kilometers long offsets were recorded. The source was bandpassed wavelet between 5 and 10 Hz. No energy was present below 3 Hz.

Figure 10 shows the data recorded for the leftmost shot location. Strong and complex diving waves and refracted arrivals are visible in the data starting from approximately 8 kilometers offsets. These events carry useful information on the velocity, in particular in the shallow part of the section. In this data set, they are extremely useful to define the low-velocity anomalies present around the depth of two kilometers.

The starting model for the TFWI inversion was obtained by a strong horizontal smoothing of the true model, after the low and high velocity anomalies were removed. Figure 11 shows the shot gather located at the same location as the one shown in Figure 10, but modeled with the starting model, which is shown in Figure 12. As a direct comparison of Figure 10 with Figure 11 demonstrates, the differences between the true and starting models cause large time shifts in the diving-waves arrivals. The inaccuracies of the starting model, together with the lack of low frequencies in the data, prevent conventional FWI from converging to any useful model.

The straightforward application of an optimization algorithm designed for non-quadratic problems to the TFWI minimization problem defined by equation 11 would be very expensive. We therefore solved the problem by a nested optimization algorithm as described in Almomin and Biondi (2013). The nested inversion converged towards the accurate model shown in Figure 13. The main features of the true model are accurately reconstructed. Some edge artifacts are present; they are caused by the finite span of the receiver array and the finite range of source locations. No sources were activated outside the displayed model, and the receiver array was fixed and covering the whole model.

Although the resolution of the model and the robustness of convergence are extremely attractive, the rate of convergence is slower than ideal; 500 iterations of the outer loop of the nested TFWI inversion were required to estimate the model shown

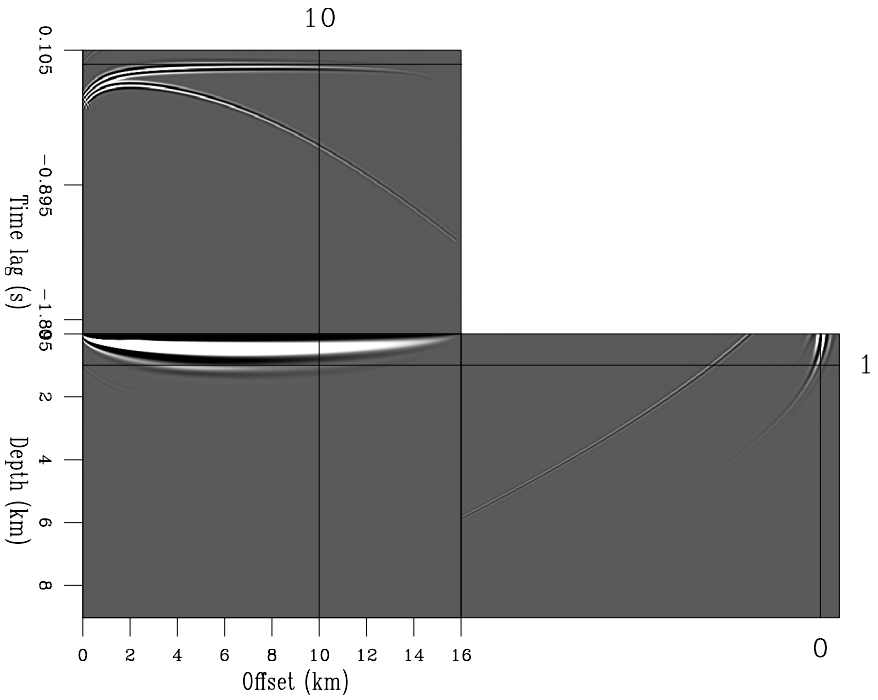

Figure 7: Velocity-perturbation cube extended along the time-lag axis and computed by backprojecting the data residuals shown Figure 6a. [**CR**]

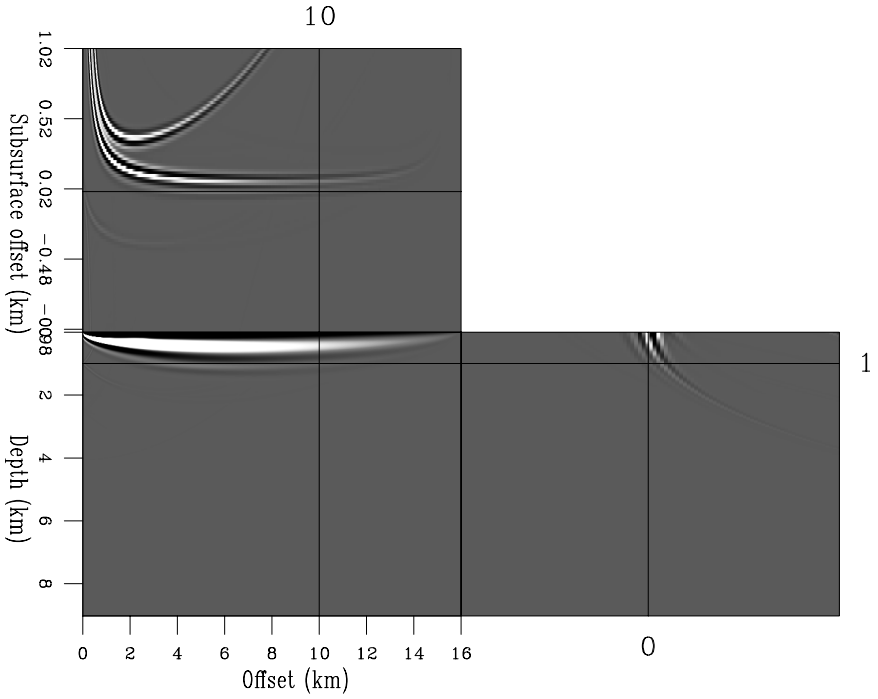

Figure 8: Velocity-perturbation cube extended along the horizontal subsurface-offset axis and computed by backprojecting the data residuals shown Figure 6a. [**CR**]

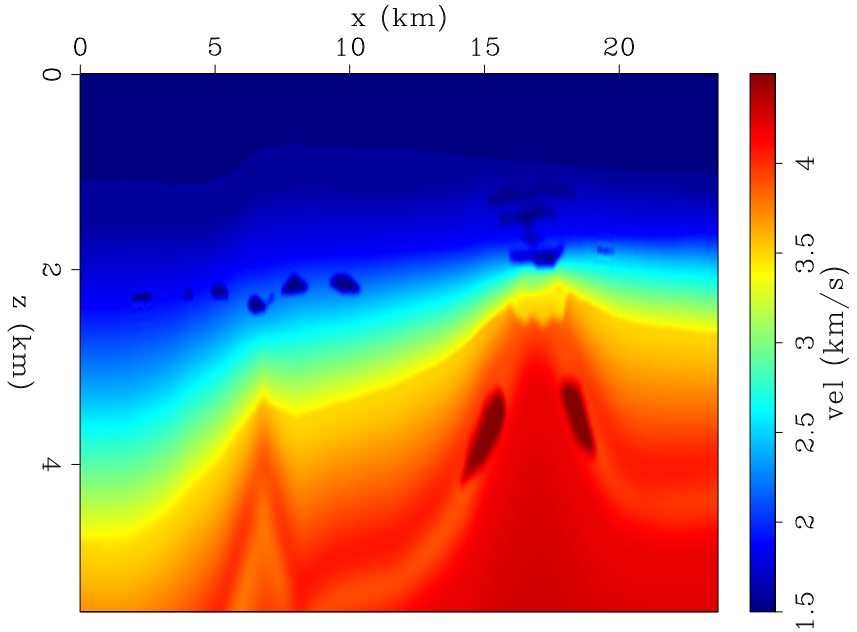

Figure 9: Portion of the BP velocity model used for the numerical test of the proposed TFWI method. The model contains both low-velocity anomalies (shallow gas) as well as high-velocity anomaly on the flanks of the mud volcano. [**ER**]

in Figure 13. We believe that several numerical techniques could be applied to speed up the convergence, but further experiments are needed.

CONCLUSIONS

The integration of FWI and WEMVA into TFWI promises to enable robust convergence to high-resolution models. We introduced a TFWI method based on the extension of the velocity along the τ axis. This extension is based on a linear operator capable of correctly modeling transmitted events with large time shifts, as we demonstrated by a simple numerical example. The results of the inversion of a long-offset data set recorded over the BP "Caspian Sea" demonstrate the strong convergence properties of new method when both reflections and diving waves are present.

ACKNOWLEDGMENTS

We would like to thank BP for making publicly available the velocity model of which we used a portion. Almomin would like the thank Saudi Aramco for supporting his Ph.D. studies at Stanford.

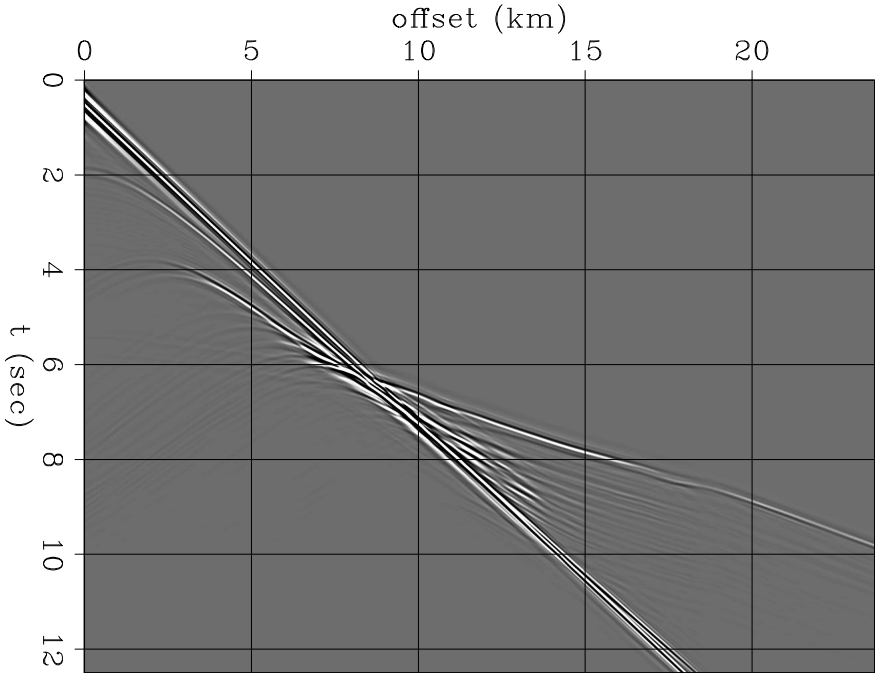

Figure 10: Leftmost shot profile recorded on the model shown in Figure 9. Notice several diving waves and refractions present in the data at offset larger than 8 kilometers. These events carry useful information for the estimation of the velocity anomalies present in the model. [**CR**]

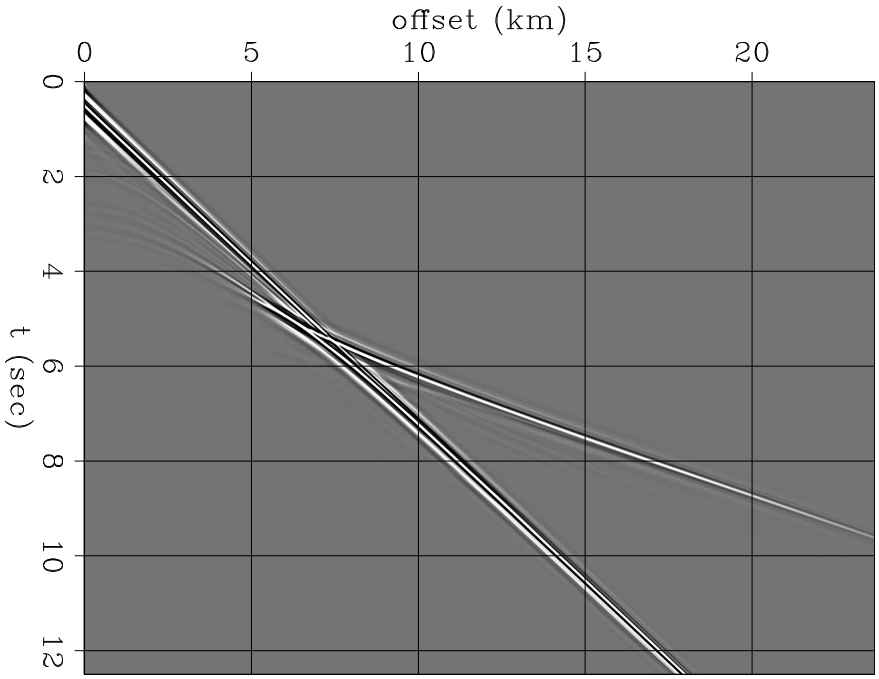

Figure 11: Shot gather modeled assuming the starting model shown in Figure 12 at the same shot location as the data shown in Figure 10. Notice the large time shifts between the diving-wave arrivals in this gather with the one shown in Figure 10. [**CR**]

REFERENCES

Almomin, A. and B. Biondi, 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: SEG Technical Program Expanded Abstracts, **31**, 500–505.

———, 2013, Tomographic full waveform inversion (TFWI) by successive linearizations and scale separations: SEP-Report, **149**, 51–58.

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, **31**, 547–552.

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

Biondi, B. and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, 1575–1586.

Sava, P. and B. Biondi, 2004a, Wave-equation migration velocity analysis—I: Theory: Geophysical Prospecting, **52**, 593–623.

———, 2004b, Wave-equation migration velocity analysis—II: Examples: Geophysical Prospecting, **52**, 607–623.

Shen, P. and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–VE59.

Sun, D. and W. Symes, 2012, Waveform inversion via nonlinear differential semblance optimization: SEG Technical Program Expanded Abstracts, **31**, 497–502.

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.

Yang, T. and P. Sava, 2009, Wave-equation migration velocity analysis using extended images: SEG Technical Program Expanded Abstracts, **28**, 3715–3719.

———, 2010, Moveout analysis of wave-equation extended images: Geophysics, **75**, S151–S161.

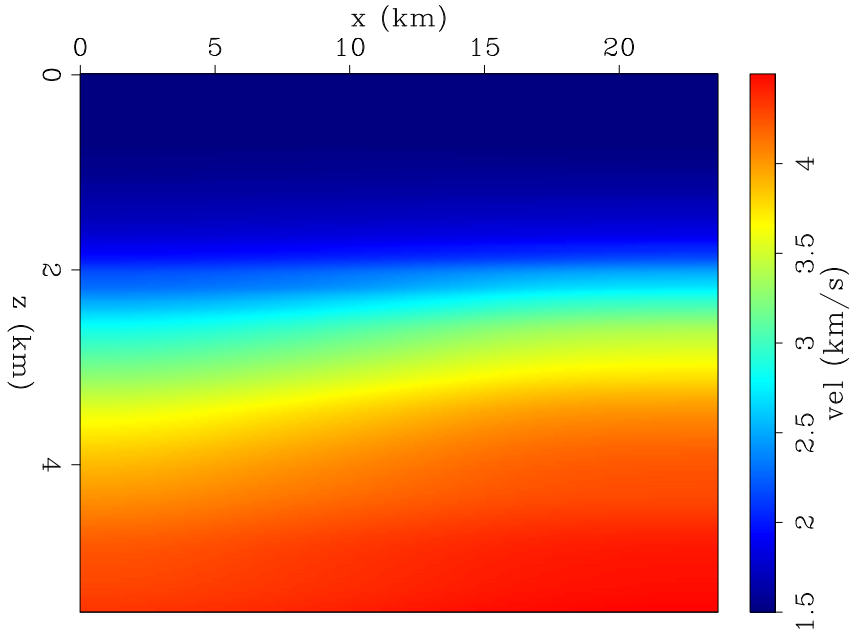

Figure 12: Starting model for the TFWI inversion. This model was obtained by strong horizontal smoothing of the model shown in Figure 9, after the low and high velocity anomalies were removed. The lack of low frequencies in the data makes this model inappropriate for starting a conventional FWI inversion. [**ER**]

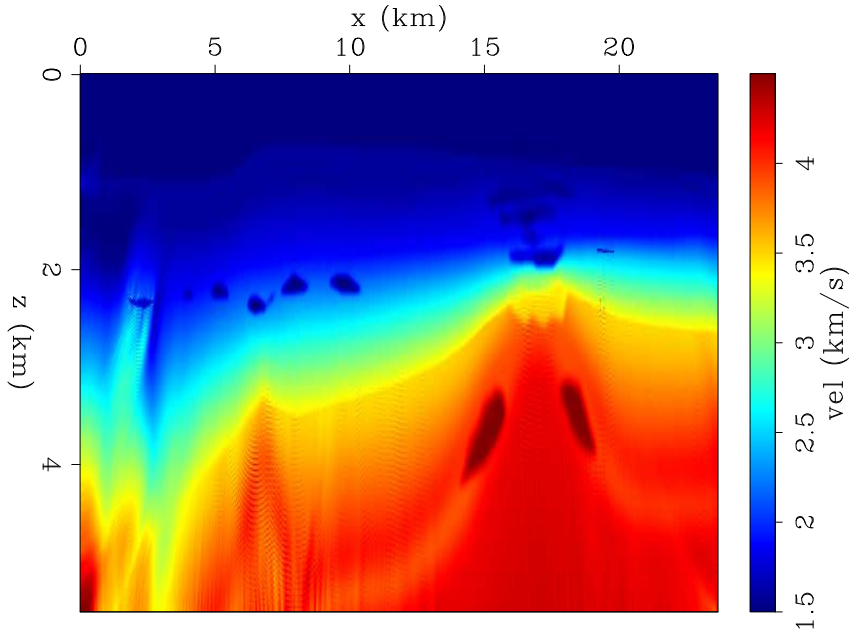

Figure 13: Estimated model after 500 iterations of the outer loop of the TFWI inversion based on time-lag extension of the velocity model. The main features of the true model are accurately reconstructed. Some edge artifacts are present; they are caused by the finite span of the receiver array and the finite range of source locations. No sources were activated outside the displayed model, and the receiver array was fixed and covering the whole model. [**CR**]