Simultaneous inversion of full data bandwidth by tomographic full waveform inversion (TFWI)

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 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. This linear modeling is based on a new linearization of the scalar wave equation where perturbation of the extended slowness-squared is convolved in time with the second time derivative of the background wavefield. The linearization is accurate for both reflected events and transmitted events. We show that it can effectively model both conventional reflection data as well as modern long-offset data containing diving waves. It also enables the simultaneous inversion of reflections and diving waves, even when the starting velocity model is far from being accurate. We solve the optimization problem related to the inversion with a nested algorithm. The inner iterations are based on the proposed linearization and on a mixing of scales between the short and long wavelength components of the velocity model. Numerical tests performed on synthetic data modeled on the Marmousi model and on the "Caspian Sea" portion of the well-known BP model demonstrate the global-convergence properties as well the high-resolution potential of the proposed method.

INTRODUCTION

Conventional seismic imaging relies on a separation of scales between the migration velocity model (long-wavelength components) and the 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

Submitted for publication to Geophysics

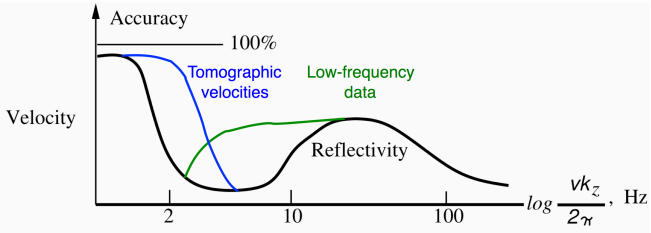

Figure 1: Simplified 1D graphical representation of the separation of scales in seismic imaging (black line) and how current industry trends are narrowing the gap between the estimation of long wavelengths and short wavelengths (blue and green lines). (Adapted from Jon Claerbout's *Imaging the Earth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

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 a 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) (Bamberger et al., 1982; Tarantola, 1984) could be mostly explained as an attempt to overcome the limitations imposed by the sequential imaging approach, as well as the availability of the 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 for imaging the high frequencies in the data under complex overburdens. FWI has been less successful in using the high-frequencies in the data to tomographically estimate the long-wavelengths in the model.

Since the 1980s it has been recognized that FWI has both a migration component

and a tomographic component (Mora, 1989). 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-wavelength perturbations cause substantial time shifts of the propagating wavefields that are poorly approximated by the linearization of the wave equation based on the 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 shifts 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 this idea to the extension of the propagation component of the velocity 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 ray-parameter axes. In this paper, we introduce an extension along the time lag axis (τ) because it is better suited to describe the 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 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 horizontal 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 is performed by solving the following optimization problem

$$\min_{\mathbf{s}^2} J_{\text{FWI}}\left(\mathbf{s}^2\right), \tag{1}$$

where:

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

$\mathbf{s} = s\left(\vec{x}\right)$ is the slowness vector, $\mathcal{L}$ is a wave-equation operator nonlinear with respect to slowness perturbations. The data vector $\mathbf{d}$ is the pressure field $\mathbf{P} = P\left(t, \vec{x}\right)$ measured at the surface through a sampling linear operator $\mathbf{S}$; such as $\mathbf{d} = \mathbf{SP}$

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

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

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.

The most 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 slowness perturbations $\delta\mathbf{s}^2$. This linear operator can be derived by perturbing equation 3 as follows

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

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

Equation 4 can be rewritten as the following coupled equations:

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

which represents a nonlinear relationship between the slowness perturbations and the 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:

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

The linear operator $\mathbf{L}$ is evaluated by recursively propagating the background wavefield $\mathbf{P_o}$ and the scattered wavefield $\delta\mathbf{P}$ by solving equations 7–8. Its adjoint operator $\mathbf{L}'$, which is needed to compute the gradient of the FWI objective function 1, is evaluated by backward propagating the scattered wavefield solving equation 8 using the data residuals as boundary conditions, and evaluating the zero time lag of the cross-correlation between the background and scattered wavefields.

Equations 7–8 define a linear relationship between $\delta\mathbf{P}$ and $\delta\mathbf{s}^2$; however, they cannot model large time shifts between the background wavefield and the scattered wavefield. These large time shifts are correctly modeled by the nonlinear equations 5–6 through multiple scattering, that is, by the accumulation of time shifts into the scattered wavefield which enter in the expression of the virtual sources injected by the right-hand-side of equation 6. When we drop the scattered wavefield from equation 6, we prevent this accumulation of large time shifts into the scattered wavefield.

These observations suggest that a simple method to improve the capability of the linearization to model large time shifts is to introduce time shifts directly into the slowness-perturbations term in the right-hand-side of equation 8. We extend the slowness model along the time lag axis τ and convolve its perturbations $\delta\tilde{\mathbf{s}}^2(\tau)$ with the second time derivative of the background wavefield; we rewrite equation 8 as:

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

where $\overset{\tau}{*}$ denotes convolution in τ. Onwards we use the tilde sign above operators and model vectors to denote their extension along the time-lag axis.

Equations 7 and 9 define the operator $\widetilde{\mathbf{L}}\left(\tilde{\mathbf{s}}(\tau=0)\right)$ which is linear with respect to $\delta\tilde{\mathbf{s}}^2(\tau)$, but nonlinear with respect to $\tilde{\mathbf{s}}^2(\tau=0)$. The combination of the wave equation operator $\mathcal{L}(\mathbf{s})$ and of $\widetilde{\mathbf{L}}$ defines the extended nonlinear operator

$$\tilde{\mathcal{L}}(\tilde{\mathbf{s}}) = \mathcal{L}\left(\tilde{\mathbf{s}}_o(\tau=0)\right) + \widetilde{\mathbf{L}}\left(\tilde{\mathbf{s}}_o(\tau=0)\right)\delta\tilde{\mathbf{s}}^2. \tag{10}$$

The modeling equation 10 is used to define the TFWI objective function as

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

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 slowness extension relaxes the constraints on the modeled data kinematics imposed by the data-fitting term (first term) in equation 11. This objective function can be minimized using a nested optimization algorithm with scale mixing, as discussed in the next section.

1D modeling example

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 wavefield difference 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 difficulties to converge, even though 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 value of 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. On the contrary, the linearized 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 the data residuals shown in Figure 6a, by the application of $\widetilde{\mathbf{L}}'$, 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 ap-

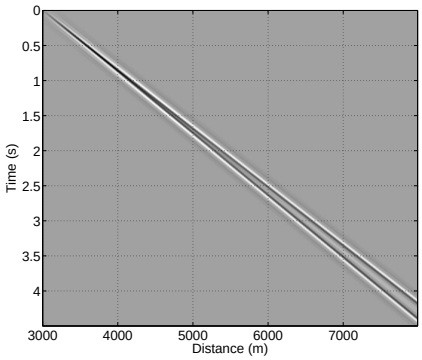

Figure 2: Difference between the background wavefield computed with the starting velocity (1.2 km/s) and the wavefield propagated with the true velocity (1.13 km/s). [**CR**]

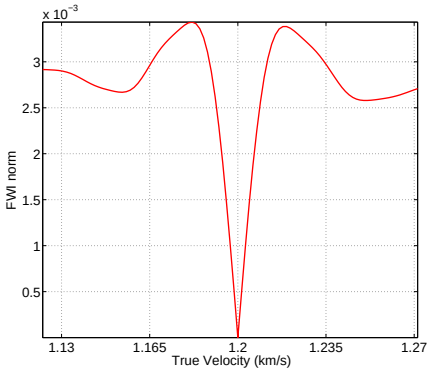

Figure 3: FWI norm as a function of the true velocity, when the starting velocity is equal to 1.2 km/s. [**CR**]

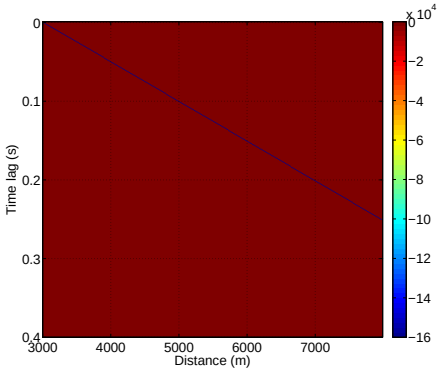

Figure 4: Extended velocity perturbation chosen to approximately model the wavefield difference shown in Figure 2. [**CR**]

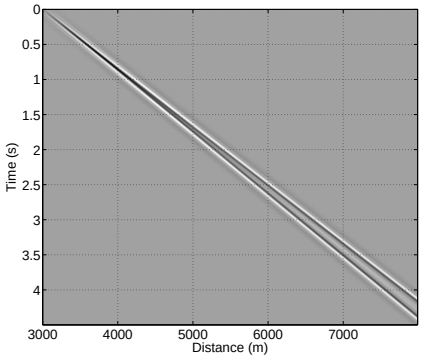

Figure 5: Perturbed wavefield computed by solving equation 9 with the model shown in Figure 4. [**CR**]

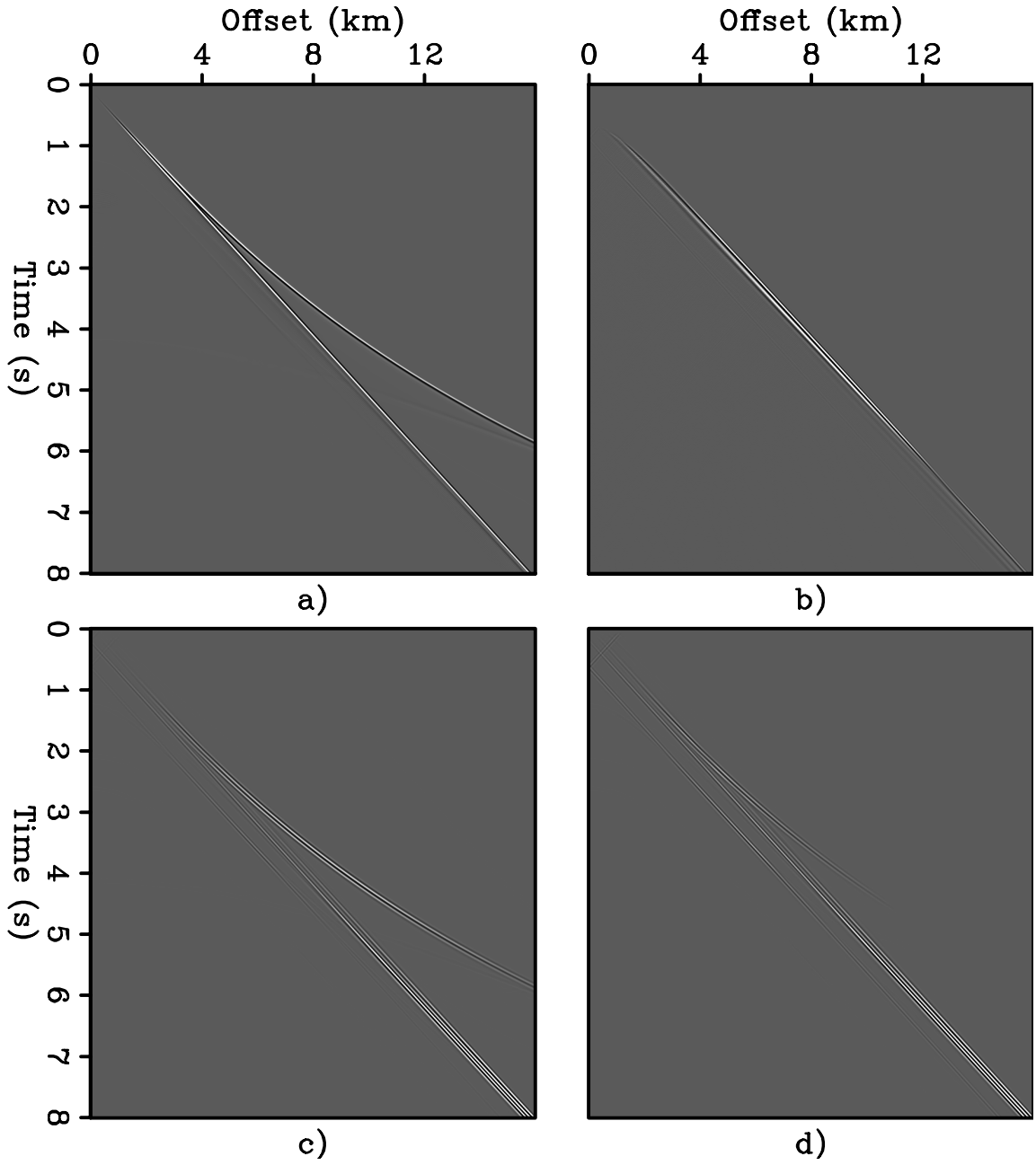

Figure 6: a) Difference between the background wavefield computed with the starting velocity and the wavefield propagated with the true velocity, b) data residual modeled from zero lag of the velocity perturbation cubes (front panel in cubes shown in both Figure 7 and Figure 8), c) data residual modeled 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**]

plication of $\widetilde{\mathbf{L}}$ to the extended velocity perturbation shown in Figure 7. Although a "squaring" of the wavelet is evident in Figure 6c, the kinematics of these modeled residuals are very close to the kinematics of the true residuals shown Figure 6a. In contrast, when we use only the zero time lag of the velocity perturbation (front panel of the cube shown in Figure 7) 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. A subsurface offset extension of the velocity has difficulties 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 the vertical subsurface offset in addition to the horizontal ones (Biondi and Symes, 2004), but the computational cost would increase accordingly. In 3D we would need to extend the velocity along 3 subsurface offsets, increasing further the dimensionality of the problem, and consequently its computational cost.

OPTIMIZATION METHOD

The minimization of the objective function in 11 is a challenging optimization problem because of the non linearities in the modeling operator $\tilde{\mathcal{L}}$ and the occasional contradictory search directions suggested by the gradients of the data fitting term and the focusing term. To overcome these challenges we devised and tested the nested optimization scheme with scale mixing described below.

Nested Inversion

The proposed nested optimization scheme consists of an outer and an inner loop. In the outer loop we first compute the nonlinear data residual $\mathbf{\Delta d} = \mathbf{d} - \tilde{\mathcal{L}}\left(\tilde{\mathbf{s}}_o\right)$, where $\tilde{\mathbf{s}}_o$ is the current slowness model evaluated at $\tau = 0$. The nonlinear residual is used as the "observed" data for the inner loop. The output of the inner loop is a search direction $\Delta\tilde{\mathbf{s}}$. We then perform a nonlinear line search that estimates the optimal step length α that minimizes $\left\|\tilde{\mathcal{L}}\left(\tilde{\mathbf{s}}_o + \alpha\Delta\tilde{\mathbf{s}}\right) - \mathbf{d}\right\|_2$.

In the inner loop we formally separate the slowness model into a background model, $\mathbf{b}$, on which the operator $\widetilde{\mathbf{L}}$ depends non linearly, and a perturbation model

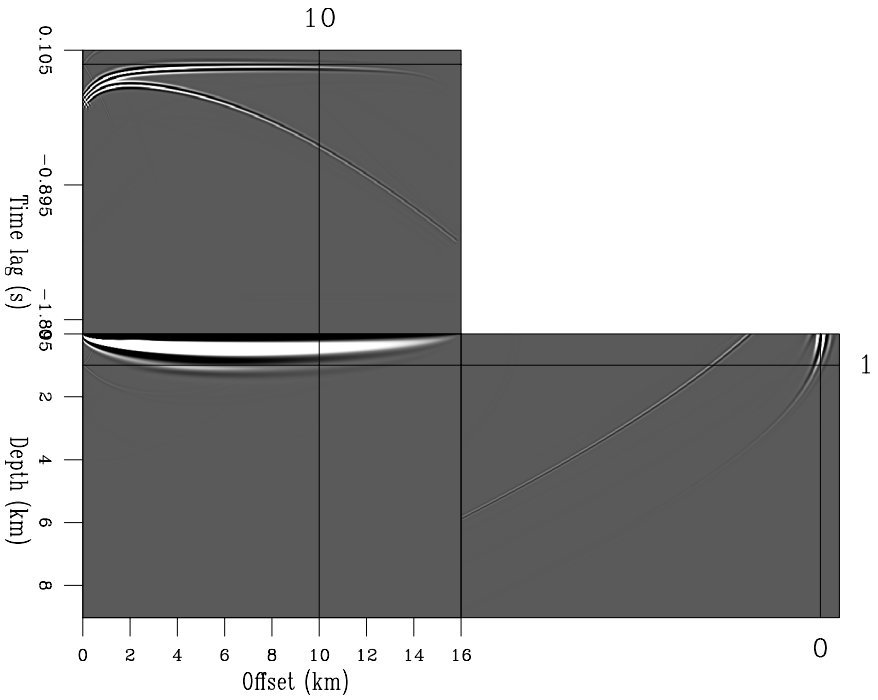

Figure 7: Velocity-perturbation cube extended along the time-lag axis and computed by backprojecting the data residuals shown Figure 6a. [**CR**]

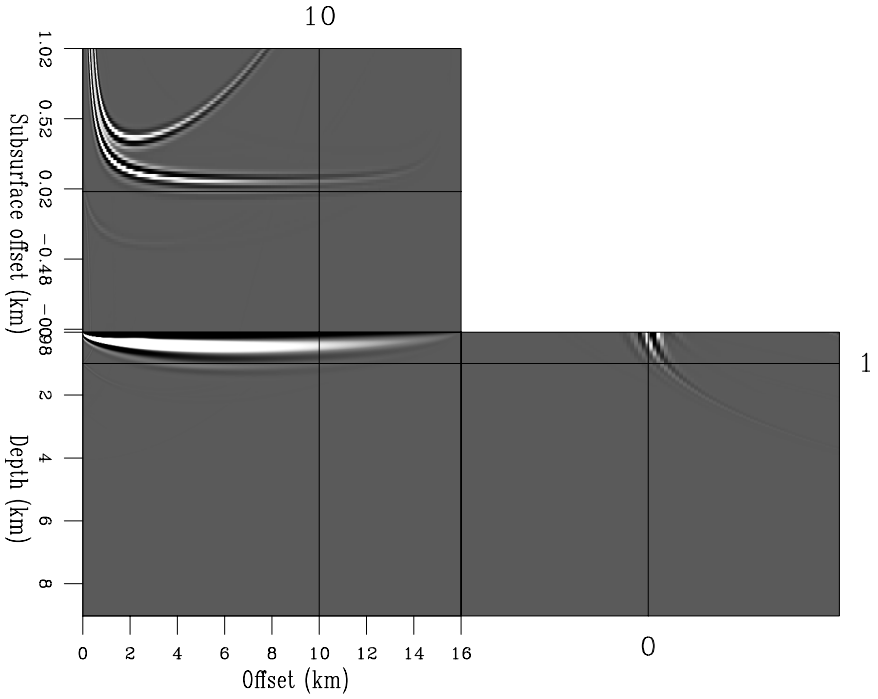

Figure 8: Velocity-perturbation cube extended along the horizontal subsurface-offset axis and computed by backprojecting the data residuals shown Figure 6a. [**CR**]

$\mathbf{p}^2(\tau)$, on which the operator $\widetilde{\mathbf{L}}$ depends linearly. The objective function minimized in the inner loop is

$$J_{\text{ITFWI}}(\mathbf{b},\mathbf{p}) = \frac{1}{2}\left\|\widetilde{\mathbf{L}}(\mathbf{b})\,\mathbf{p}^2 - \mathbf{\Delta d}\right\|_2^2 + \epsilon\left\||\tau|\,\mathbf{p}^2\right\|_2^2. \tag{12}$$

The starting models for the inner iterations are $\mathbf{b}_o = \tilde{\mathbf{s}}_o(\tau=0)$, and $\mathbf{p}_o = 0$. The output of the inner loop after N iterations is $\Delta\tilde{\mathbf{s}} = (\mathbf{b}_N - \mathbf{b}_o) + \mathbf{p}_N(\tau=0)$.

Unless the starting background slowness $\mathbf{b}_o$ is very close to the true slowness, no choice of $\mathbf{b}$ and $\mathbf{p}$ will simultaneously zero the two terms in the objective function 12. This happens because of the particular choice of the data residual $\mathbf{\Delta d}$, that is the difference between the recorded data and the data modeled with $\mathbf{b}_o$. However, the models $\mathbf{b}$ and $\mathbf{p}$ that minimize this objective function provide an effective search direction $\Delta\tilde{\mathbf{s}}$ for the outer loop of the nested optimization problem.

The modeling operator $\widetilde{\mathbf{L}}$ is linear with respect to perturbation $\mathbf{p}$, but nonlinear with respect to the background component $\mathbf{b}$. Therefore, another linearization around the "background" background is required to compute the gradient. The Born approximation is used (again) to linearize the $\widetilde{\mathbf{L}}$ operator with respect to the background resulting in a data-space tomographic operator, $\mathbf{T}$. Appendix A describes the derivation of this new data-space tomographic operator and how to numerically evaluate it and its adjoint. The expression of the two gradients at the inner iteration i is the following:

$$\nabla_{\mathbf{b}^2} = \mathbf{T}'(\mathbf{b}_i,\mathbf{p}_i)\left[\widetilde{\mathbf{L}}(\mathbf{b}_i)\,\mathbf{p}_i^2 - \mathbf{\Delta d}\right], \tag{13}$$

$$\nabla_{\mathbf{p}^2} = \widetilde{\mathbf{L}}'(\mathbf{b}_i)\left[\widetilde{\mathbf{L}}(\mathbf{b}_i)\,\mathbf{p}_i^2 - \mathbf{\Delta d}\right] + \epsilon\tau^2\mathbf{p}_i^2. \tag{14}$$

Scale Mixing

In the inner loop, a straightforward use of the gradients is used to update their corresponding models directly. However, this would hinders the simultaneous inversion of different wavelengths of the model. This problem becomes apparent when we examine the result of the two operators in the inner loop. At the first iteration, the application of $\widetilde{\mathbf{L}}'$ to the data residual $\mathbf{\Delta d}$ is equivalent to "migrating" the data and it could give a tomographic update which manifests as a low wavenumber update. In a conventional migration, this low wavenumber component is considered noise and filtered out. However, it is actually a tomographic component that should feed into the background model. The opposite argument is also true for the tomographic operator creating short-wavelengths perturbations. Therefore, to improve our inversion results, we first mix the two gradients $(\nabla_{\mathbf{b}^2}, \nabla_{\mathbf{p}^2})$ and then separate them in the Fourier domain to get the update of each model as follows:

$$\Delta\mathbf{b}^2(\vec{x}) = -\mathbf{C}_{\mathbf{b}^2}\left(\nabla_{\mathbf{b}^2}(\vec{x}) + \nabla_{\mathbf{p}^2}(\vec{x},\tau=0)\right), \tag{15}$$

where $\Delta \mathbf{b}^2(\vec{x})$ is the search direction of the background model and $\mathbf{C_b}$ is a low-pass filter along the space coordinates $\vec{x}$. Similarly, we can compute the update of the perturbation model as:

$$\Delta \mathbf{p}^2(\vec{x},\tau) = -\mathbf{C}_{\mathbf{p}^2}\left(\nabla_{\mathbf{b}^2}(\vec{x}) + \nabla_{\mathbf{p}^2}(\vec{x},\tau)\right), \tag{16}$$

where $\Delta \mathbf{p}^2(\vec{x},\tau)$ is the search direction of the perturbation model and $\mathbf{C_p}$ is a high-pass filter along the space coordinates $\vec{x}$. 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.

In the examples of 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 slowness of the initial model. The two filters were designed such that they always sum to one at all wavelength to maintain the energy of the gradients. It is possible to design a more accurate filter that varies with frequency and slowness, but it is not necessary because both models will eventually be added to the slowness. This relative insensitivity is another benefit of applying the nested scheme we presented, in comparison with inverting the models separately and combining them only at the end. If the latter approach were implemented, the final results would be more sensitive to the choice of the scale-separation parameters.

This nested scheme has many benefits compared to a more conventional way of solving a tomographic inversion followed by an imaging inversion or even doing them simultaneously. The first benefit of this scheme is that the limitation of the linearized, first-order Born operator will not prevent the optimization from inverting higher order scatterings, such as multiple reflections and prismatic events. In this setup, the Born operator will attempt to match the first-order scattering from the background, which is not necessarily primary data since the background itself can generate many orders of scatterings if it contains sharp boundaries (and it will in later iterations). In other words, the linearized operator will account for one additional order of scattering at a time. Higher and higher scattering orders will be introduced with outer-loop iterations, until we eventually invert all the multiple-scattered events that are present in the data. This also means that higher-order scattered energy will initially be wrongly positioned in the model, but later iterations will correct for this mispositioning. The second benefit of this scheme is that it produces only one slowness model, because we keep pushing both background and perturbation into the slowness model, in contrast to the method presented in Almomin and Biondi (2012). Therefore, our goal is to drive the perturbation to a minimum and have the background explain the data.

TFWI OF MARMOUSI DATA

A modified Marmousi model is used for the first synthetic examples where 500m of water layer is added to the top. Figure 9 shows the true velocity model. The thicker

water layer reduces the amount of refracted energy being recorded in the data. Therefore, this dataset tests the capabilities of the inversion method to converge in presence of almost exclusively reflected events. In contrast, the example presented in the next section tests the capabilities of the method to deal with mixed reflected/refracted arrivals.

There are 461 fixed receivers with a spacing of 20 m and 93 sources with a spacing of 100 m. We use a bandpassed wavelet with a frequency range between 5 Hz to 10 Hz and a small taper on both ends. The purpose of using this wavelet is to completely eliminate unrealistically low frequencies in the data. The initial model is shown in Figure 10 which is obtained by strongly smoothing the true model laterally after clipping the salt anomalies out. The inversion results after 900 outer-loop iterations, each of them with 10 inner iterations, are shown in Figure 11. The inversion shows remarkable results in reconstructing most features in the velocity model. The results are most accurate in the top and middle than at the bottom because the have the best illumination and coverage.

Figure 12 and 13 provide a clear illustration of the substantial improvements in the accuracy with which the final velocity model describes the data kinematics as compared with the initial one. Figure 12 shows the migrated image corresponding to the initial velocity; most of the reflectors are out of focus and mispositioned. In contrast, Figure 13 shows the migrated image corresponding to the final velocity. Reflectors are well focused and the structures are well imaged. In the middle of the section even the deepest reflectors are well focused and not grossly distorted.

TFWI OF LONG-OFFSET DATA

To verify the capabilities of the TFWI method based on a 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 14. 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 a bandpassed wavelet between 5 and 10 Hz. No energy was present below 3 Hz.

Figure 15 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 16 shows the shot gather located at the same location as the one shown in Figure 15, but modeled with the starting model, which is shown in Figure 17. As a

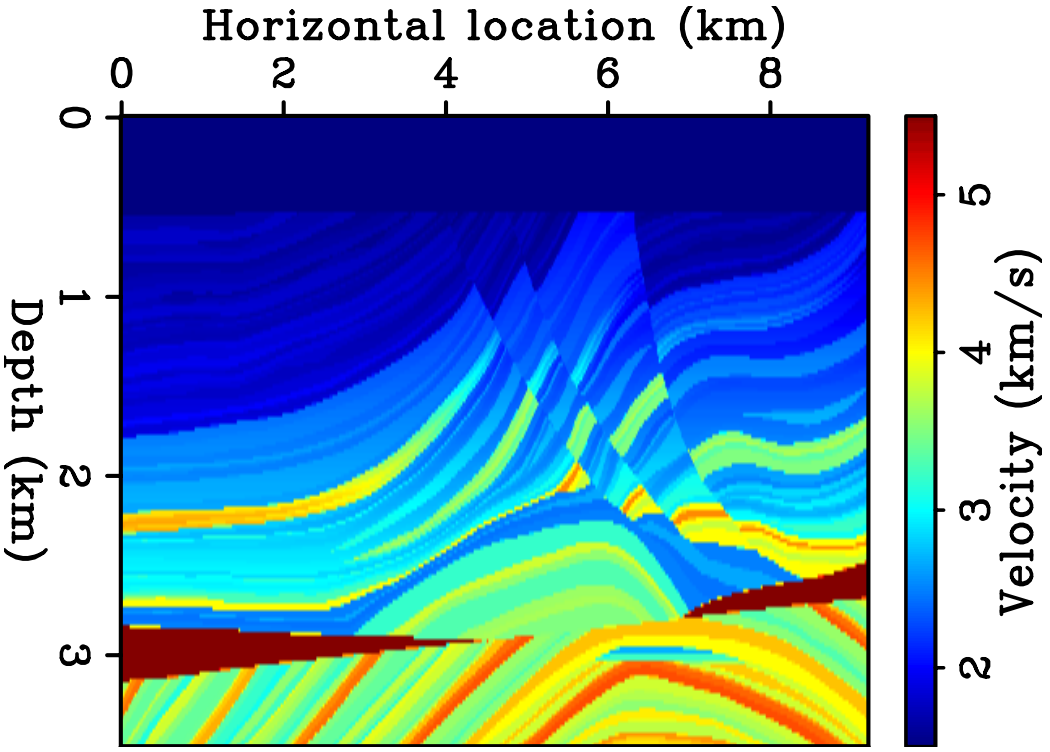

Figure 9: The true velocity of the Marmousi example. [**ER**]

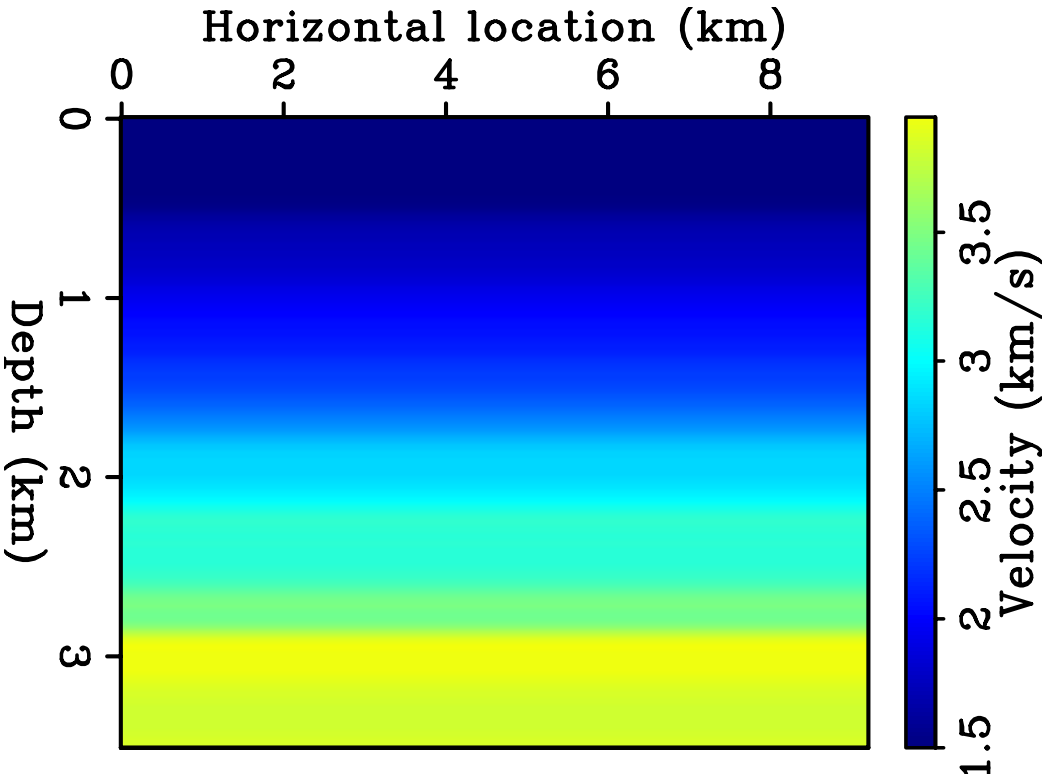

Figure 10: The initial velocity of the Marmousi example. [**ER**]

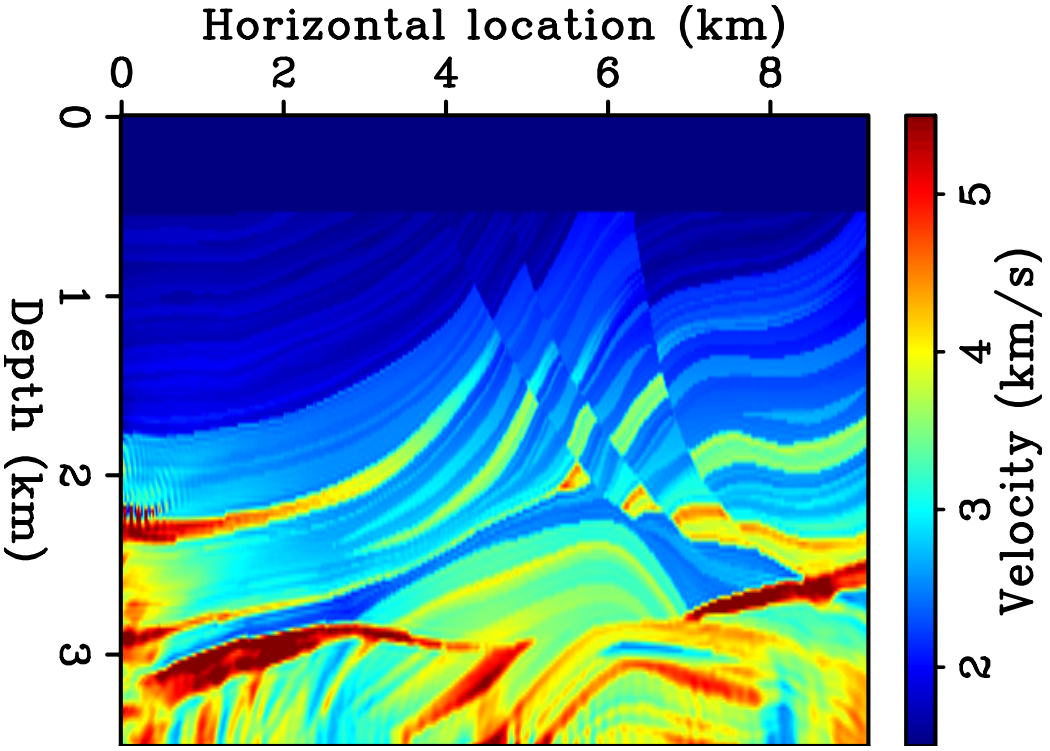

Figure 11: The inverted velocity of the Marmousi example. [**CR**]

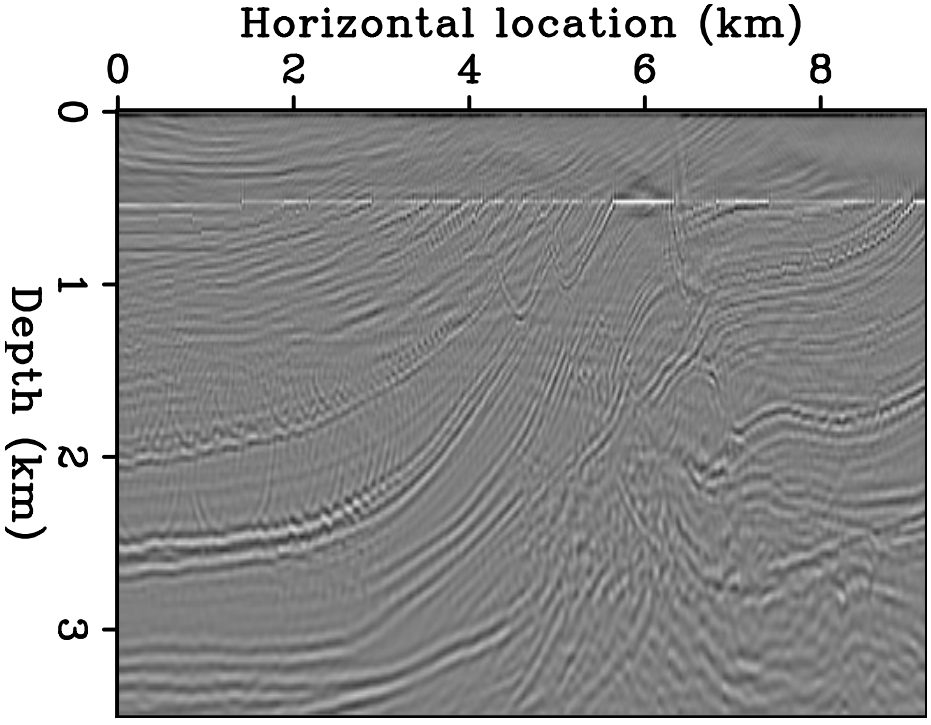

Figure 12: Migrated image using the initial velocity for the Marmousi example. [**CR**]

direct comparison of Figure 15 with Figure 16 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.

We solved the problem by a nested optimization algorithm as described earlier. The nested inversion converged towards the accurate model shown in Figure 18. 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 covered 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 iterations, each of them with 10 inner iterations, were required to estimate the model shown in Figure 18.

DISCUSSIONS AND CONCLUSIONS

The integration of FWI and WEMVA into TFWI has the potential of yielding a waveform-inversion method that robustly converges to high-resolution models using the whole bandwidth of the seismic data simultaneously. 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 demonstrate by two numerical examples based on simple 1D and layered models.

To minimize the TFWI objective function we propose a specialized nested inversion scheme. In the inner loop of this scheme the extended velocity model is separated into its background and perturbation components. The inversion scheme performs simultaneous inversion of different model scales by mixing the gradients of the two components and then separating them in the Fourier domain.

The results of the inversion of the Marmousi model demonstrate the strong convergence properties of the new method for typical reflection data. The results of the inversion of a long-offset data set recorded over the BP "Caspian Sea" demonstrate that the inversion method converges when both reflections and diving waves are recorded in the data, and inaccuracies of the starting velocity model create large errors in the kinematics of the diving waves.

The proposed optimization is effective to demonstrate TFWI convergence properties on realistic 2D data sets. However, the number of iterations required to produce the results shown in the paper are large: 900 for the Marmousi example and 500 for the "Caspian Sea" example. This suggests that further optimization of the inversion algorithm is needed for applications to 3D field data.

ACKNOWLEDGMENTS

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

REFERENCES

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

Bamberger, A., G. Chavent, C. Hemon, and P. Lailly, 1982, Inversion of normal incidence seismograms: Geophysics, **47**, 757–770.

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.

Sava, P. and I. Vlad, 2008, Numeric implementation of wave-equation migration velocity analysis operators: Geophysics, **73**, VE145–VE159.

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.

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

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.

APPENDIX A

The minimization of the objective function 12 requires the computation of the gradients with respect to both $\mathbf{b}$ and $\mathbf{p}$. These gradients can be computed by a perturbation analysis of the modeling operator $\widetilde{\mathbf{L}}$. As discussed in the main text, $\widetilde{\mathbf{L}}$ is evaluated by solving equations 7 and 9. Rewriting these equations in terms of $\mathbf{b}$ and $\mathbf{p}$, and the incident wavefield, $\mathbf{P_i}$ and the scattered wavefield, $\mathbf{P_s}$, we obtain

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} = \mathbf{f}, \tag{A-1}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_s} = \mathbf{p}_i^2(\tau) \overset{\tau}{*} \mathbf{D_2}\mathbf{P_i}. \tag{A-2}$$

Introducing the perturbations, $\delta\mathbf{b}^2$ and $\delta\mathbf{p}^2$, in the two model variables into equations A-1–A-2, and introducing the corresponding perturbations in the wavefields, $\delta\mathbf{P_i}$ and $\delta\mathbf{P_s}$ yields the following perturbed system:

$$\left[\left(\mathbf{b}_i^2 + \delta\mathbf{b}^2\right)\mathbf{D_2} - \nabla^2\right](\mathbf{P_i} + \delta\mathbf{P_i}) = \mathbf{f}, \tag{A-3}$$

$$\left[\left(\mathbf{b}_i^2 + \delta\mathbf{b}^2\right)\mathbf{D_2} - \nabla^2\right](\mathbf{P_s} + \delta\mathbf{P_s}) = \left[\mathbf{p}_i^2(\tau) + \delta\mathbf{p}^2(\tau)\right] \overset{\tau}{*} \mathbf{D_2}(\mathbf{P_i} + \delta\mathbf{P_i}). \tag{A-4}$$

By setting $\delta\mathbf{b}^2 = 0$ in equations A-3–A-4, we derive the following system of equations:

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} = \mathbf{f}, \tag{A-5}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_s} = \delta\mathbf{p}^2(\tau) \overset{\tau}{*} \mathbf{D_2}\mathbf{P_i}, \tag{A-6}$$

which defines again $\widetilde{\mathbf{L}}$ and can be used to evaluate the perturbations in the scattered wavefield $\delta\mathbf{P_s}$, and consequently in the recorded data $\delta\mathbf{d} = \mathbf{S}\delta\mathbf{P_s}$, caused by perturbations $\delta\mathbf{p}^2$.

We can derive the data-space tomographic operator, $\mathbf{T}$, which relates perturbations in the scattered wavefield, $\delta\mathbf{P_s}$, to perturbations in the background model, $\delta\mathbf{b}^2$, by setting $\delta\mathbf{p}^2 = 0$ and neglecting the higher-order terms in $\delta\mathbf{b}^2$ in equations A-3–A-4.

This tomographic operator is equal to the sum of two operators, $\mathbf{T_i}$ and $\mathbf{T_s}$. The first operator, $\mathbf{T_i}$, models perturbations in the scattered wavefield caused by perturbations in the propagation of the incident wavefield:

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} = \mathbf{f}, \tag{A-7}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_i} = \delta\mathbf{b}^2\mathbf{D_2}\mathbf{P_i}, \tag{A-8}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_s} = \mathbf{p}_i^2(\tau) \overset{\tau}{*} \mathbf{D_2}\delta\mathbf{P_i}, \tag{A-9}$$

The second operator, $\mathbf{T_s}$, models perturbations in the scattered wavefield caused by perturbations in the propagation of the scattered wavefield itself:

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_i} = \mathbf{f}, \tag{A-10}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\mathbf{P_s} = \mathbf{p}_i^2(\tau) \overset{\tau}{*} \mathbf{D_2}\mathbf{P_i}, \tag{A-11}$$

$$\left[\mathbf{b}_i^2\mathbf{D_2} - \nabla^2\right]\delta\mathbf{P_s} = \delta\mathbf{b}^2\mathbf{D_2}\mathbf{P_s}, \tag{A-12}$$

Both of these tomographic operators depend nonlinearly on $\mathbf{b}_i$ and linearly on $\mathbf{p}_i^2$. They have zero output when $\mathbf{p}_i^2$ is equal to zero; that is, at the first iteration of the inner loop because we set $\mathbf{p}_o^2 = 0$ as a starting model of the inner iterations. However, as we update the linearization at each iteration, starting from the second iteration the output of $\mathbf{T}$ becomes different from zero.

The data-space tomographic operator represented by equations A-7–A-9 and equations A-10–A-12 is analogous to the WEMVA operator (Biondi and Sava, 1999; Sava and Vlad, 2008) except that in the WEMVA operator, we keep the data fixed and vary the image; it is the other way around in this tomographic operator.

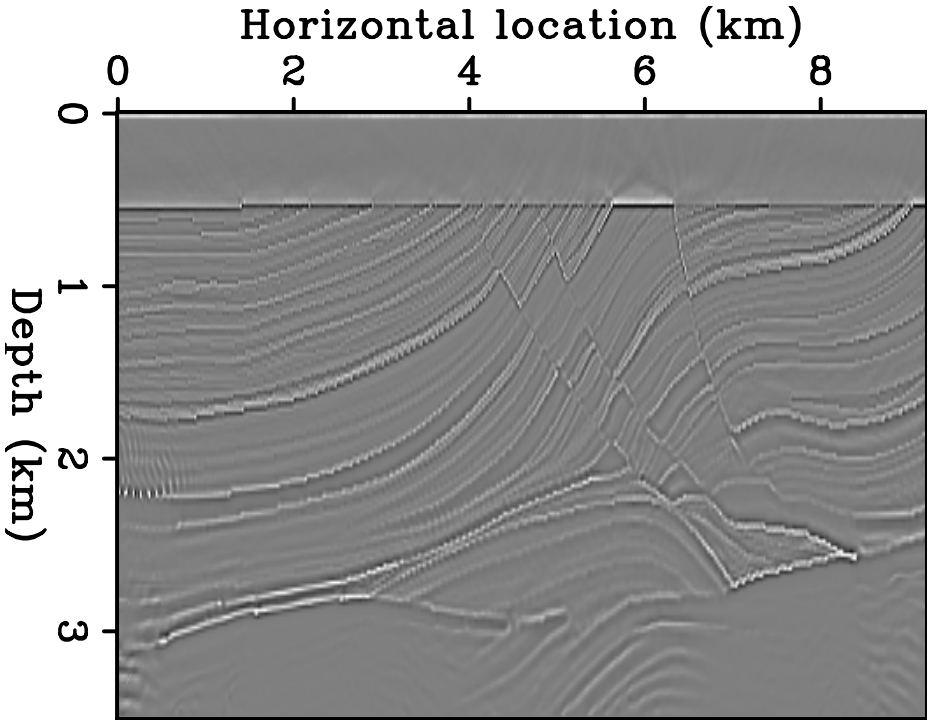

Figure 13: Migrated image using the final velocity for the Marmousi example. [**CR**]

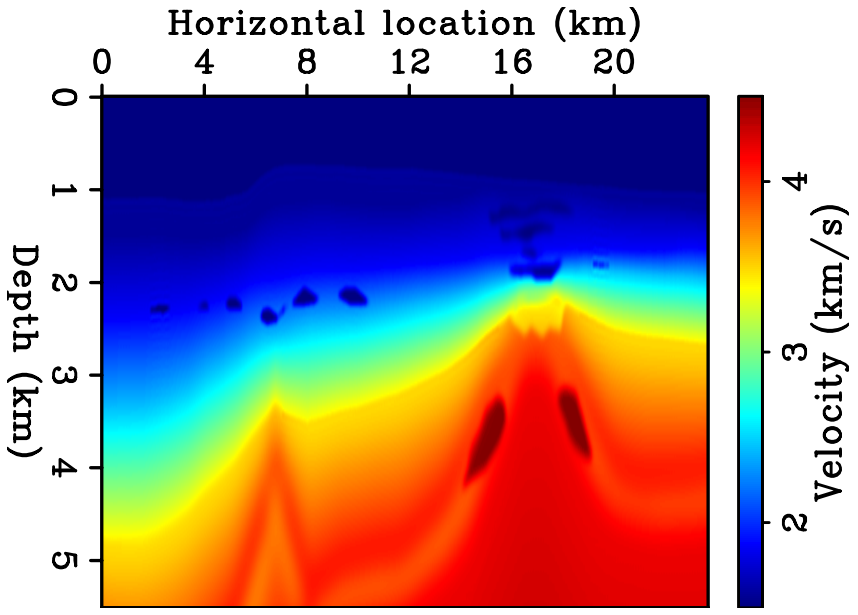

Figure 14: Portion of the BP velocity model used for the numerical test of the proposed TFWI method. The model contains both low-velocity anomalies (shallow gas) as well as high-velocity anomaly on the flanks of the mud volcano. [**ER**]

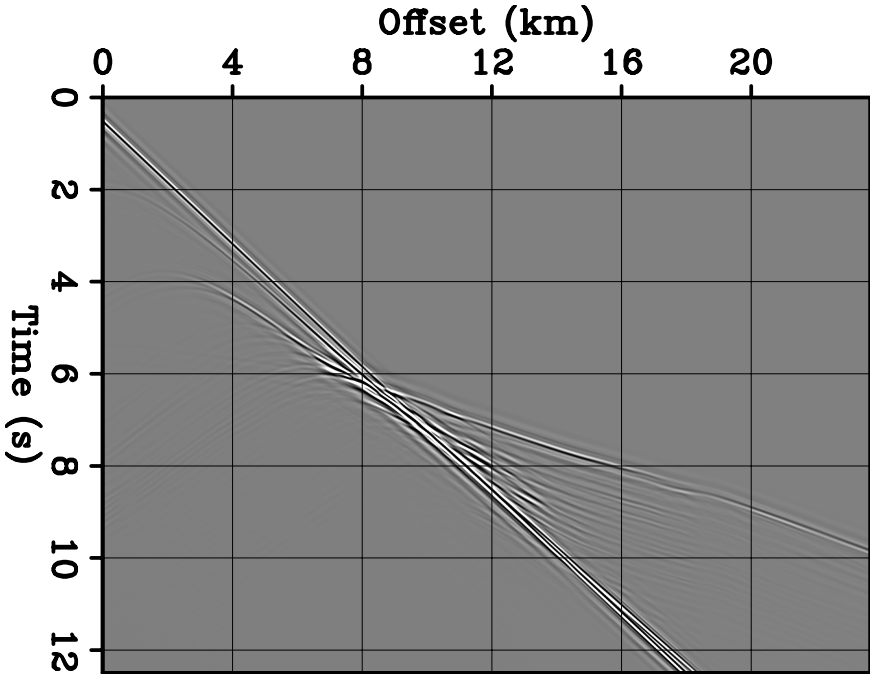

Figure 15: Leftmost shot profile recorded on the model shown in Figure 14. Notice several diving waves and refractions present in the data at offset larger than 8 kilometers. These events carry useful information for the estimation of the velocity anomalies present in the model. [**CR**]

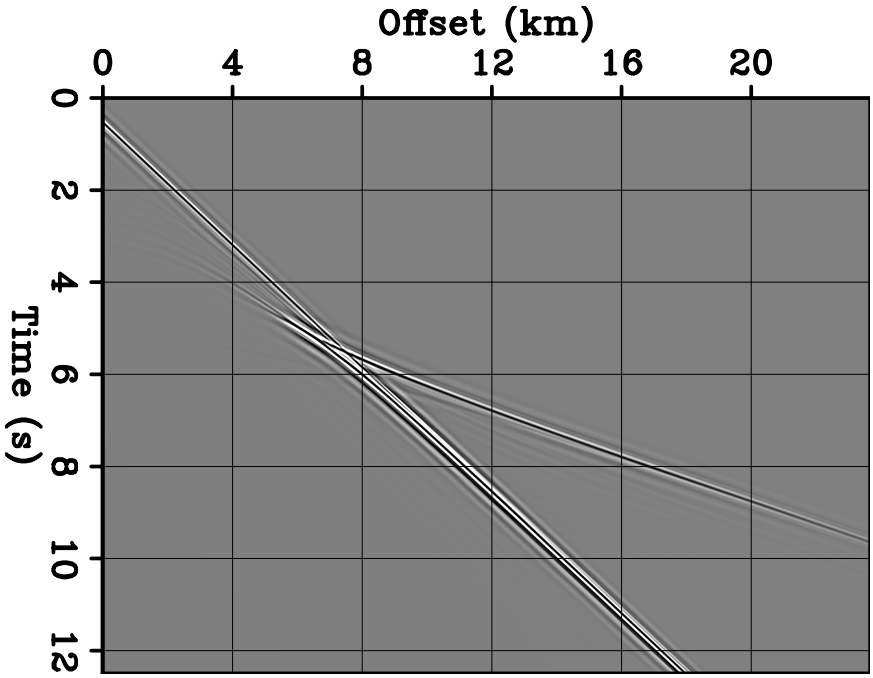

Figure 16: Shot gather modeled assuming the starting model shown in Figure 17 at the same shot location as the data shown in Figure 15. Notice the large time shifts between the diving-wave arrivals in this gather with the one shown in Figure 15. [**CR**]

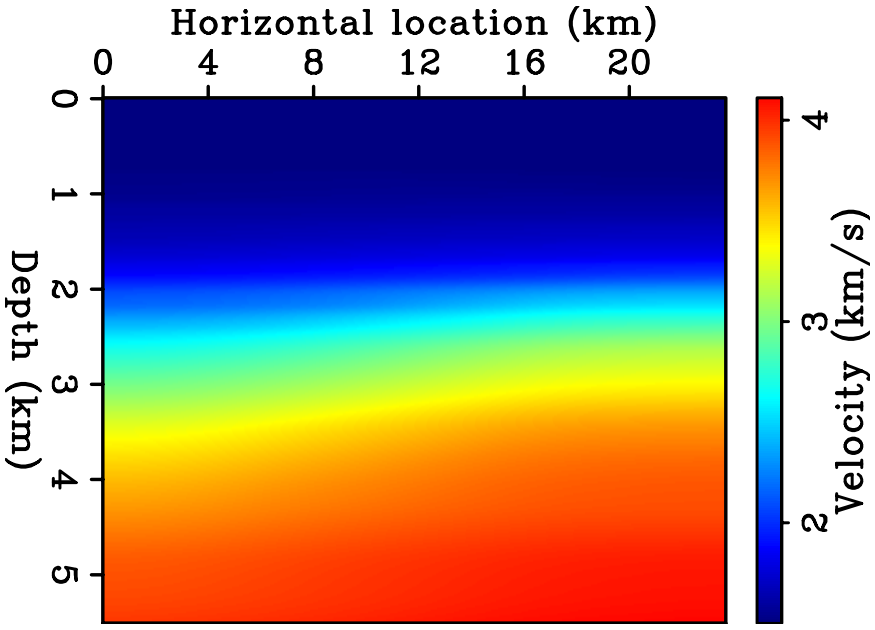

Figure 17: Starting model for the TFWI inversion. This model was obtained by strong horizontal smoothing of the model shown in Figure 14, after the low and high velocity anomalies were removed. The lack of low frequencies in the data makes this model inappropriate for starting a conventional FWI inversion. [**ER**]

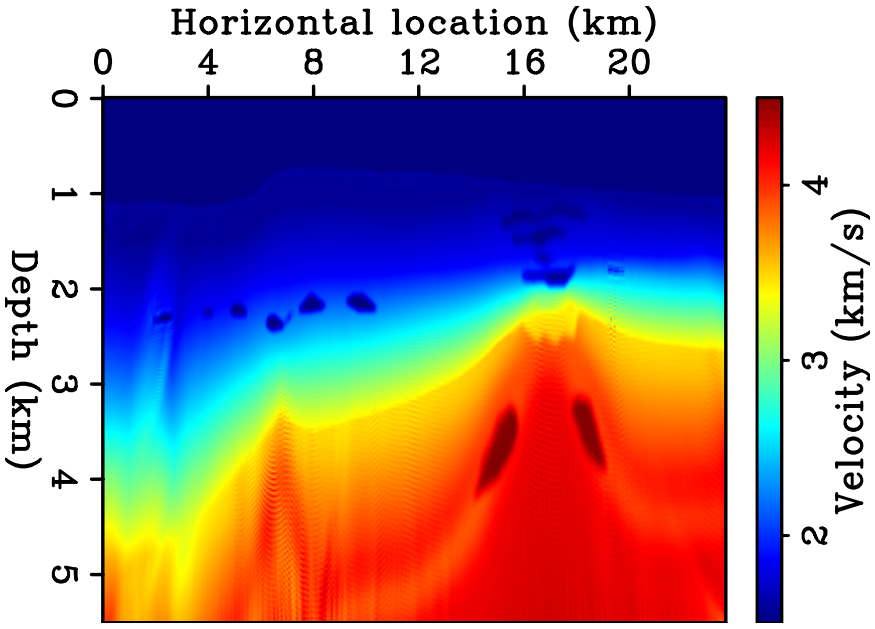

Figure 18: Estimated model after 500 iterations of the outer loop of the TFWI inversion based on time-lag extension of the velocity model. The main features of the true model are accurately reconstructed. Some edge artifacts are present; they are caused by the finite span of the receiver array and the finite range of source locations. No sources were activated outside the displayed model, and the receiver array was fixed and covered the whole model. [**CR**]