

Building good starting models for full-waveform inversion using adaptive matching filtering misfit

Hejun Zhu¹ and Sergey Fomel²

ABSTRACT

We have proposed a misfit function based on adaptive matching filtering (AMF) to tackle challenges associated with cycle skipping and local minima in full-waveform inversion (FWI). This AMF is designed to measure time-varying phase differences between observations and predictions. Compared with classical least-squares waveform differences, our misfit function behaves as a smooth, quadratic function with a broad basin of attraction. These characters are important because local gradient-based optimization approaches used in most FWI schemes cannot guarantee convergence toward true models if misfit functions include local minima or if the starting model is far away from the global minimum. The 1D and 2D synthetic experiments illustrate the advantages of the proposed misfit function compared with the classical least-squares waveform misfit. Furthermore, we have derived adjoint sources associated with the proposed misfit function and applied them in several 2D time-domain acoustic FWI experiments. Numerical results found that the proposed misfit function can provide good starting models for FWI, particularly when low-frequency signals are absent in recorded data.

INTRODUCTION

Full-waveform inversion (FWI) has become an important model building technique in exploration and global seismology over the past 10 years (Fichtner et al., 2009; Tape et al., 2009; Virieux and Operto, 2009; Zhu et al., 2012, 2013; Zhu and Tromp, 2013). Although its basic theory was proposed in the 1980s (Lailly, 1983; Tarantola, 1984; Gauthier et al., 1986), it has not been widely applied until the late 1990s (Pratt et al., 1998) because of limitations on computing power, numerical techniques, and data quality. Com-

pared with classical ray-based traveltime tomography, FWI enables us to obtain higher resolution subsurface images (Sirgue et al., 2010). The 2D and 3D results have shown FWI's capabilities to image structures in complicated subsurface environments (Ben-Hadj-Ali et al., 2008; Brossier et al., 2009). Meanwhile, multiparameter FWI has been developed to constrain different seismic parameters, such as velocity, density, anisotropy, and attenuation (Malinowski et al., 2011; Prieux et al., 2013; Warner et al., 2013).

Although recently FWI has been successfully applied in many synthetic and field data experiments, there are still many theoretical and practical challenges. Among these difficulties, cycle skipping and local minima are significant challenges because they directly determine whether FWI can converge toward a correct solution. One way to tackle these challenges is to design well-behaved misfit functions. The following criteria should be taken into account when we attempt to design new misfit functions. First, they should ideally behave as quadratic functions around the global minimum. Second, they should be smooth functions, avoiding abrupt jumps or discontinuities so that local gradient-based optimization approaches can converge robustly. Third, they should involve broad basins of attraction so that we are able to obtain robust inversion results.

Over the past several decades, a variety of misfit functions have been proposed for FWI. The most commonly used misfit is the least-squares (L2) waveform difference between observations and predictions. These waveform differences can be quantified in the time domain (Tarantola, 1984), frequency domain (Pratt, 1999), or Laplace domain (Shin and Cha, 1998). The second type of misfit functions separates traveltimes and amplitudes. The amplitudes are always difficult to work with because they are subject to a variety of effects besides the variations in subsurface model parameters, such as source magnitudes, radiation patterns, local site effects, etc. This type of misfit functions ignores amplitude discrepancies and focuses on traveltime or phase differences between observations and predictions (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010). These phase differences can be measured by crosscorrelation

Manuscript received by the Editor 2 November 2015; revised manuscript received 29 February 2016; published online 3 August 2016.

¹The University of Texas at Dallas, Department of Geosciences, Richardson, Texas, USA. E-mail: hejun.zhu@utdallas.edu.

²The University of Texas at Austin, Bureau of Economic Geology, Jackson School of Geosciences, Austin, Texas, USA. E-mail: sergey.fomel@beg.utexas.edu.

© 2016 Society of Exploration Geophysicists. All rights reserved.

or multitaper techniques (Zhou et al., 2004; Zhu et al., 2015). This approach has been successfully applied in transmission-type tomography. Instead of designing misfits in either time or frequency domains, some studies proposed misfit functions in the time-frequency domains, using instantaneous phases (Fichtner et al., 2008) or envelope differences (Wu et al., 2014). Recently, several approaches used in signal processing were adopted in FWI, such as dynamic warping (Ma and Hale, 2013), registration-guided waveform differences (Baek et al., 2013), and wavelet analysis (Yuan and Simons, 2014). They have shown advantages in comparison with L2 waveform misfits. Adaptive waveform inversion (Warner and Guasch, 2014) used matching filters, the Wiener filter, to quantify discrepancies between observations and predictions. Minimizing the filter coefficients, they were able to iteratively update velocity structures. Significant improvements have been shown to avoid cycle skipping and local minima, even in the cases without good starting models and low-frequency signals. Wasserstein metric has been proposed to measure transports between two time series (Engquist and Froese, 2014). It is a promising approach to tackle cycle-skipping problem in FWI. In addition, instead of quantifying misfit functions with L2 norms, some L1 and L1/L2 mixed norms, such as the Huber norm (Ha et al., 2009), were proposed in the waveform inversion.

In this paper, we propose a misfit function based on adaptive matching filtering (AMF). We first introduce the definition of AMF, which allows us to measure time-varying phase differences between two geophysical signals. Next, we propose a misfit function based on AMF. Then, we derive expressions for adjoint sources associated with the proposed misfit function. Several 1D and 2D synthetic experiments are used to illustrate the advantages of the proposed misfit function compared with L2 waveform misfits. Finally, we use several 2D numerical examples to show the potential of the proposed misfit function for building good starting models for FWI when low-frequency signals are absent in recorded data.

THEORY

Adaptive matching filtering

AMF is a method to measure nonstationary phase differences between two geophysical signals, which has been successfully used for adaptive subtraction of multiple reflections (Fomel, 2009). Given two time series, prediction $\mathbf{p}(t)$ and observation $\mathbf{d}(t)$, a sta-

tionary matching filter $\mathbf{f}(\gamma)$ may be defined as the solution to the following least-squares inverse problem:

$$\min \left\| \sum_{\gamma} \mathbf{d}(\gamma t) \mathbf{f}(\gamma) - \mathbf{p}(t) \right\|^2. \quad (1)$$

For the whole traces, it only gives us a single measurement $\mathbf{f}(\gamma)$ for a particular value of the stretching variable γ . However, most geophysical signals are nonstationary. Our goal here is to extract a time-varying matching filter for different values of the stretching variable γ . We can modify the stationary matching filter defined in equation 1 to a nonstationary AMF, denoted by $\mathbf{f}(\gamma, t)$. It is defined as the solution to the following modified least-squares inverse problem,

$$\min \left\| \sum_{\gamma} \mathbf{d}(\gamma t) \mathbf{f}(\gamma, t) - \mathbf{p}(t) \right\|^2. \quad (2)$$

Equation 2 is an underdetermined inverse problem. We have to add additional constraints, i.e., regularization, to solve this problem. Shaping regularization (Fomel, 2007) is one type of regularization approaches for solving ill-posed inversion problems. In this study, we use it to solve equation 2:

$$\mathbf{f} = [\mathbf{S}(\mathbf{D}^T \mathbf{D} - \lambda^2 \mathbf{I}) + \lambda^2 \mathbf{I}]^{-1} \mathbf{S} \mathbf{D}^T \mathbf{p}, \quad (3)$$

where \mathbf{D} is the data matrix composed of $\mathbf{d}(\gamma t)$ and \mathbf{S} denotes the shaping operator. We choose a smoothing operator in this case to impose the smoothness of filter coefficients. For all examples in this paper, we use a triangle smoothing operator as the shaping regularization operator. The AMF coefficients are solved trace by trace, and λ is chosen based on the L2-norm of observations, i.e., $\lambda = \|\mathbf{d}\|^2$. Following Fomel (2007), we can rewrite equation 3 as

$$\mathbf{f} = \mathbf{H}[\mathbf{H}^T(\mathbf{D}^T \mathbf{D} - \lambda^2 \mathbf{I})\mathbf{H} + \lambda^2 \mathbf{I}]^{-1} \mathbf{H}^T \mathbf{D}^T \mathbf{p}, \quad (4)$$

where the smoothing operator is $\mathbf{S} = \mathbf{H}\mathbf{H}^T$. The conjugate gradient method (Fletcher and Reeves, 1964) is used to iteratively solve equation 4. The method converges in practice in a small number of iterations. Compared to forward and adjoint calculations for solving wave equations, the computational cost for calculating AMF is relatively cheap. Moreover, the computation can be easily parallelized.

Misfit function based on AMF

Given two seismograms, AMF gives us a mapping, which is a function of time t and stretching variable γ . For instance, in Figure 1, we have two synthetic seismograms and each of them involves three events. These two seismograms have identical phases for the first event. However, for the second and third events, they have phase delays and advances, respectively. By solving the least-squares inverse problem in equation 2, we are able to calculate AMF as shown in Figure 2. There are three energy patches in the AMF image. Energies concentrating around the stretching variable of $\gamma = 1$ correspond to the first events with identical phases in the seismograms, whereas energies deviating away from $\gamma = 1$ correspond to the second and third events with nonstationary phase differences, i.e., phase advances and delays.

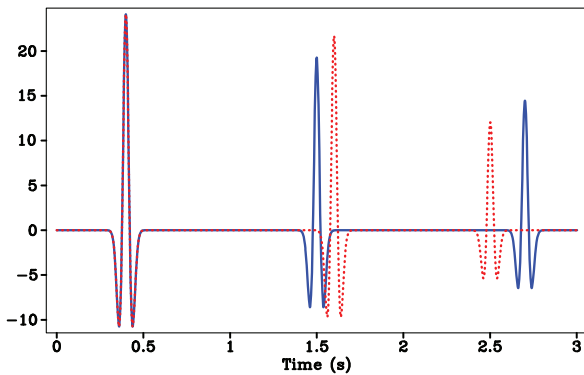


Figure 1. Comparisons of two seismograms with three events. For the first event, these two seismograms have identical phase. There are phase delay and advance for the second and third events.

With these properties, we propose a misfit function based on AMF. Our approach is inspired by the differential semblance optimization principle (Symes and Carazzone, 1991; Shen and Symes, 2008), adaptive waveform inversion (Warner and Guasch, 2014), and Wasserstein distance (Engquist and Froese, 2014). In the proposed misfit function, we use a penalty function $\gamma - 1$ to emphasize semblance energies deviating away from $\gamma = 1$,

$$J = \frac{1}{2} \frac{\iint [\mathbf{f}(\gamma, t)(\gamma - 1)]^2 d\gamma dt}{\iint [\mathbf{f}(\gamma, t)]^2 d\gamma dt}. \quad (5)$$

By minimizing the misfit function in equation 5, we are able to reduce semblance energies deviating away from $\gamma = 1$, which then aligns phases between the two seismograms.

Equation 5 can be rewritten in a discretized form as

$$J = \frac{1}{2} \frac{\|\mathbf{w}\mathbf{f}\|^2}{\|\mathbf{f}\|^2}, \quad (6)$$

where \mathbf{f} contains the coefficients of AMF and \mathbf{w} represents the weight function $\gamma - 1$.

Compared with crosscorrelation technique for quantifying phase discrepancies, this method allows us to measure phase differences without breaking the data into local windows. Therefore, we are able to avoid several nonintuitive input parameters, such as window sizes, overlaps between windows, taper functions, etc. In addition, the stretching and squeezing alignments used in AMF (equation 2) can better represent phase advances or delays due to velocity changes. This misfit function resembles the Wiener filter technique proposed by Warner and Guasch (2014). However, AMF allows us to measure nonstationary phase differences between two seismograms, whereas the Wiener filter used by Warner and Guasch (2014) is a long time invariant filter with lengths greater than the input seismograms. In addition, we use the stretch parameter γ instead of time shifts that are used as penalty functions in the misfit function based on the Wiener filter.

Adjoint sources for the proposed misfit function

A key step in FWI is the derivations of adjoint sources. In the adjoint methods, the adjoint sources can be calculated based on the partial derivative of misfit functions with respect to predictions (Plessix, 2006), i.e., $\partial J / \partial \mathbf{p}$. For the misfit function J defined in equations 5 and 6, the adjoint sources can be calculated as

$$\frac{\partial J}{\partial \mathbf{p}} = \left(\frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right)^T \frac{\mathbf{w}^T \mathbf{w} \mathbf{f} - 2J \mathbf{f}}{\mathbf{f}^T \mathbf{f}}, \quad (7)$$

where $(\partial \mathbf{f} / \partial \mathbf{p})^T$ can be computed based on equation 4

$$\left(\frac{\partial \mathbf{f}}{\partial \mathbf{p}} \right)^T = \mathbf{D} \mathbf{H} [\mathbf{H}^T (\mathbf{D}^T \mathbf{D} - \lambda^2 \mathbf{I}) \mathbf{H} + \lambda^2 \mathbf{I}]^{-1} \mathbf{H}^T. \quad (8)$$

We use the conjugate gradient method to iteratively solve equation 7. The derivations of equations 7 and 8 are similar to Warner and Guasch (2014) except that shaping regularization is used to solve nonstationary filter coefficients and adjoint sources. Once

we have expressions for the adjoint sources of the proposed misfit function, we can use them to drive adjoint wavefields. The correlations of forward and adjoint wavefields generate misfit gradients, which can be used to iteratively update model parameters by local gradient-based optimization approaches, such as the nonlinear conjugate gradient method (Fletcher and Reeves, 1964) or the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method (Nocedal, 1980).

NUMERICAL EXAMPLES

Behavior of misfit functions

In this section, we use 1D and 2D numerical examples to compare the behavior of misfit functions based on L2 waveform differences and AMF. First, we use a simple 1D analytical signal as an example,

$$S(t) = (1.0 + dA) \exp[-2(t - 3.5 + dt)^2] \cos[2\pi f_0(t - 3.5 + dt)]. \quad (9)$$

Equation 9 describes a cosine function modulated by a Gaussian envelope (Figure 3). Its amplitude and phase are controlled by two parameters dA and dt . The true model is set with $dA = 0$ and $dt = 0$. The central frequency is chosen as $f_0 = 3$ Hz.

We change the amplitude and phase parameters and compute misfit functions based on L2 waveform differences and AMF. In Figure 4, we observe that the L2 waveform misfit involves numerous local minima due to cycle skipping. This behavior would cause convergence troubles with local optimization approaches if the starting model is far away from the global minimum. On the other hand, the proposed misfit function based on AMF behaves as a

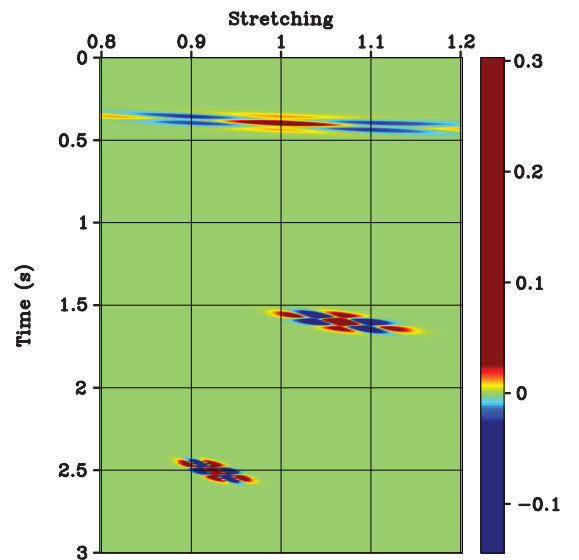


Figure 2. Magnitude of AMF for seismograms in Figure 1. The first energy patch concentrating around $\gamma = 1$ corresponds to the first event in which these two seismograms have identical phase. The second and third energy spots spreading around $\gamma = 1$ correspond to the second and third events in which these two seismograms have phase advance and delay.

smooth, quadratic function with a broad basin of attraction. Therefore, it would be much easier for local gradient-based optimization methods to work with the proposed misfit function.

We use a simple 2D three-layer model (Figure 5) in the second example. One source and 151 receivers are used in this experiment. Velocities in each layer are 2, 3, and 4 km/s in the true model. The depths of each layer are 1.3 and 2.6 km. We consider two cases. In the first case, two model parameters are velocity for the second layer and depth for the first layer. In the second case, the model parameters are the depths for the first and second layers. In all numerical examples in the following sections, we use a constant density, acoustic-wave equation to propagate wavefields, which is numeri-

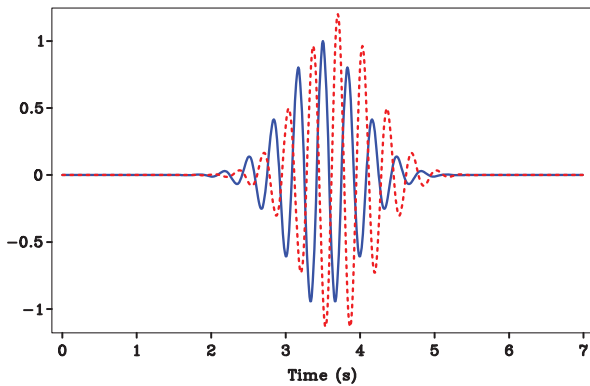


Figure 3. One-dimensional analytical signal used for comparing the behaviors of misfit functions. The blue solid line is original signal (equation 9). The red dashed line is a modified signal with $dt = 0.2$ s and $dA = 0.2$. In this case, time shift is greater than a half-period of the original signal.

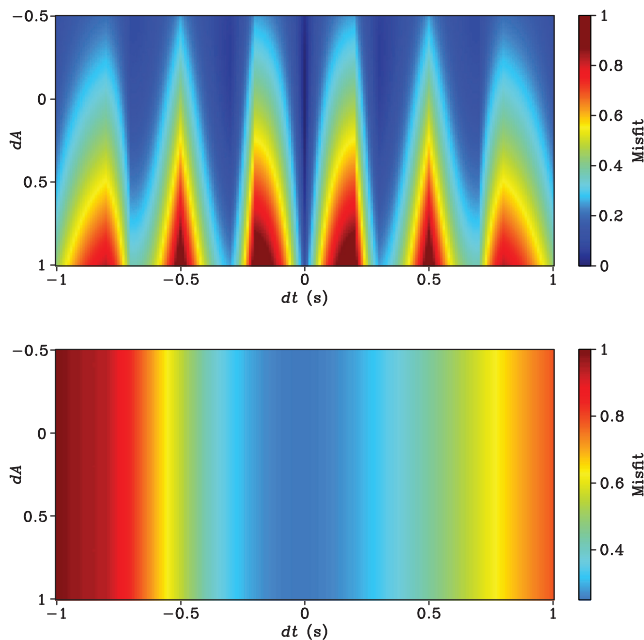


Figure 4. Behaviors of misfit functions for the analytical signal in Figure 3. From top to bottom are misfit functions based on L2 waveform differences and AMF.

cally solved by a finite-difference method. By changing these two model parameters and computing the misfits based on the comparisons of seismic data generated from baseline velocity from Figure 5 and perturbed models, we are able to compare the behavior of the misfit functions based on L2 waveform differences and AMF.

In case 1, we observe that the L2 waveform misfit involves a narrow and long valley. Away from the valley, there are “mountains” and “plateaus,” which lead to local minima (Figure 6a). If the starting model is far away from the global minimum, local gradient-based optimization approaches may converge toward local minima instead of the global minimum. In contrast, the proposed misfit function behaves as a smooth, quadratic function with a single global minimum. We expect that the local gradient-based optimization approaches should work efficiently on the proposed misfit function.

In case 2, two model parameters are the depths of the first two layers (Figure 7). We can draw a similar conclusion as in case 1. The L2 waveform misfit involves local minima, whereas the proposed misfit function exhibits a much better behavior.

Behavior of adjoint sources and misfit gradients

In this section, we compare the behavior of adjoint sources and gradients based on different misfit functions. The true velocity model is shown in Figure 8. The starting model is homogeneous with velocity equal to 4 km/s. We perform two experiments with different true models. The true model for the first case involves a Gaussian anomaly with a small amplitude perturbation (maximum 2.5%) in the middle of the model. For the second case, we put a Gaussian anomaly with a large amplitude perturbation (maximum 40%) in the middle of the model. Only one shot and one receiver are used in both cases. A Ricker wavelet with central frequency of 15 Hz is used.

Comparisons between data and synthetics are shown in Figure 9. For the first case with small velocity perturbations, differences between data and prediction are relatively small (within half-period of signal). The adjoint sources for both misfit functions only involve a single event (Figure 9a). However, for the experiment with large velocity perturbations, the delay of data with respect to prediction is noticeably greater than the half-period (Figure 9b). In this case, the adjoint source based on L2 waveform difference contains two events (one from data and one from prediction), whereas the adjoint source based on AMF has still only one event, which shows a behavior similar to that in the first experiment with small velocity perturbations.

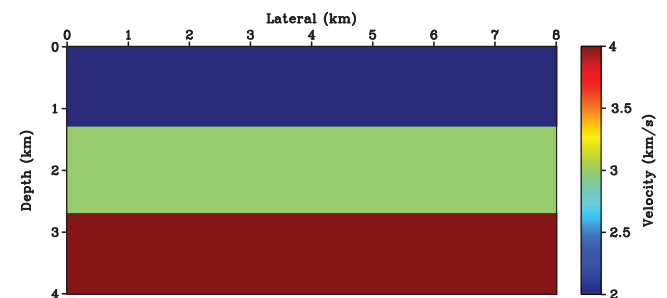


Figure 5. Three-layer model used for the comparisons of misfit functions. Shot is located at (4, 0), and receivers are equally distributed at earth surface.

We can back propagate these adjoint sources to generate adjoint wavefields, which are then correlated with forward wavefields to construct misfit gradients. Figure 10 compares misfit gradients based on different misfit functions for different cases. For the case with small velocity perturbations (Figure 10a), both misfit gradients

exhibit a similar behavior. The first Fresnel zone involves correct sensitivity, needed to reduce velocity. However, for the case with large velocity perturbations, the first Fresnel zone of the misfit gradient based on L2 waveform differences has a wrong sensitivity (Figure 10b) and does not carry any useful velocity information. On the other hand, the

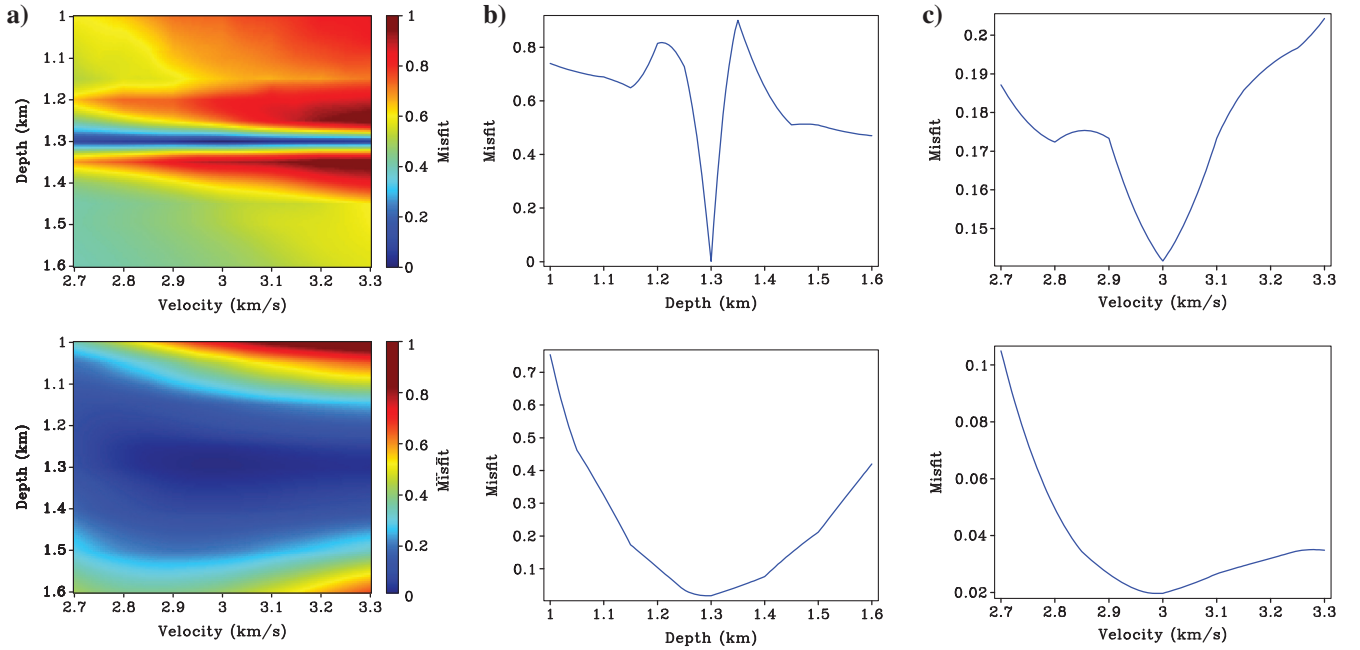


Figure 6. Comparisons of misfits functions for the three-layer model in Figure 5. Model parameters are the depth of the first layer d and the velocity of the second layer v . From top to bottom are misfit functions based on L2 waveform differences and AMF. (a) Shows 2D misfit functions, (b) shows a cross sections of misfits with $v = 3$ km/s, and (c) shows a cross sections of misfits with $d = 1.3$ km.

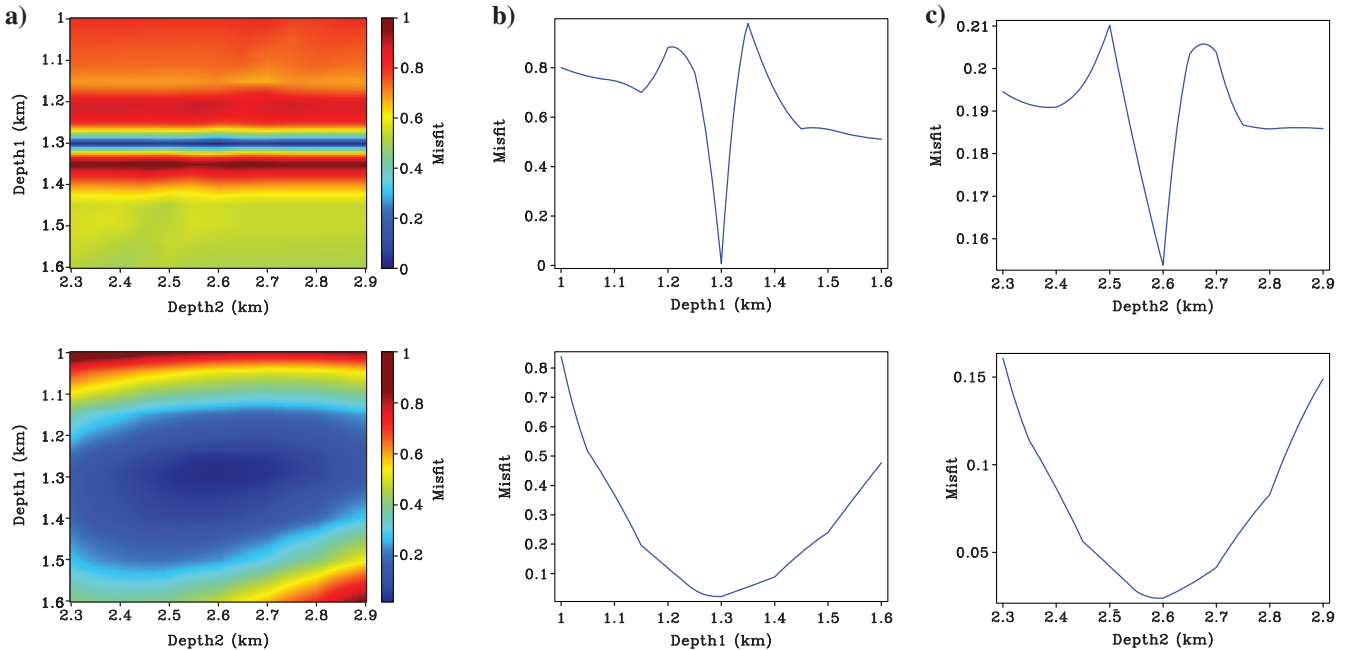


Figure 7. The same setting as Figure 6 except here: Model parameters are the depths of the first and second layers.

misfit function based on AMF still contains the correct velocity update information. These comparisons confirm the robustness of the proposed misfit function compared with L2 waveform differences.

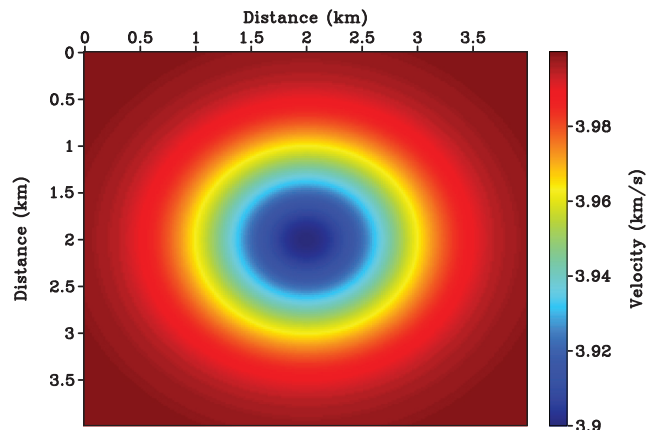


Figure 8. Velocity model used to compare misfit gradients based on different misfit functions. This is the case with small Gaussian perturbation (maximum 2.5%). The shot is located at (2, 0.5), and the receiver is located at (1, 3.5).

Transmission waveform tomography experiments

Once we have misfit gradients, we can integrate them with gradient-based optimization approaches to build velocity models. In this section, we consider two transmission tomography experiments.

The first experiment involves four Gaussian anomalies with alternate signs in the middle of the model. The starting model is homogeneous with velocity equal to 4 km/s. The true model involves Gaussian anomalies with maximum velocity perturbations equal to 12.5% (Figure 11a). This is a case with large velocity perturbations that cause data and predictions to be out of phase. We use a Ricker wavelet with the central frequency of 15 Hz. Eleven shots are equally distributed at $z = 0.5$ km, and 151 receivers are distributed at $z = 3.5$ km. No multiscale strategy (Bunks et al., 1995; Sirgue and Pratt, 2004) is used in this experiment. We performed 10 conjugate gradient iterations for misfit functions based on L2 waveform differences and AMF.

For the case with L2 waveform differences, we observe that the inversion converges toward a wrong velocity model. Maximum velocity perturbation in the final model reaches up to 65% (Figure 11b). Evidently, in this case, the inversion suffers from cycle skipping and local minima problems. In contrast, for the misfit function based on AMF, the inversion converges toward the global minimum. We are able to recover the shapes and amplitudes of these

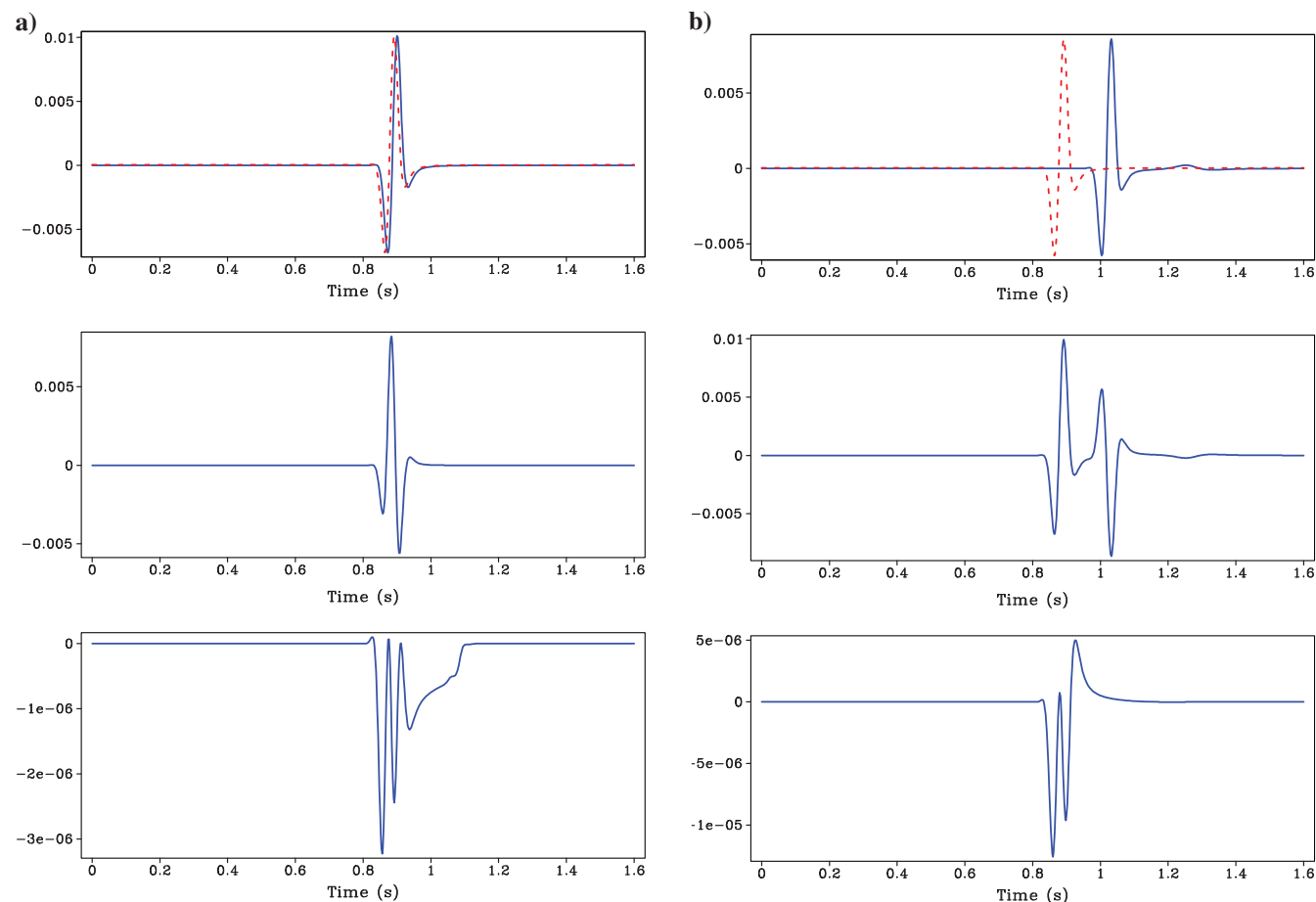


Figure 9. Comparisons of data and predictions as well as adjoint sources for different misfit functions. Left panel (a) for the case with small Gaussian perturbation and right panel (b) for the case with large Gaussian perturbation. From top to bottom are comparisons between data (blue) and predictions (red), adjoint sources based on L2 waveform differences and AMF, respectively.

four Gaussian anomalies (Figure 12). The elongation of the recovered velocity perturbations along vertical direction is caused by the limited illumination of the survey.

For the second experiment, we test the proposed misfit function for a crosswell tomography problem. The starting velocity model is laterally homogeneous. A sequence of true models is generated by adding relatively complicated velocity perturbation to the starting model (Figure 13). We tried two cases in this experiment. For the first case, scaled small perturbations (maximum 0.2 km/s) are added to the $V(z)$ starting model. For the second case, relatively large perturbations (maximum 0.4 km/s) are added to the starting model. For this crosswell experiment, we use a Ricker wavelet with the central frequency of 10 Hz. Ten shots are equally distributed along the left side of the model, and 31 receivers are equally distributed along the right side of the model. Only 10 conjugate-gradient iterations are performed in both cases.

For the first case with small velocity perturbations, the inversion based on L2 waveform differences converges toward the true model. We are able to recover true velocity perturbations. Data and model misfits are smoothly reduced (Figure 14). However, for the case with large velocity perturbations, the inversion based on L2 waveform differences does not converge to the right answer. Although data misfits are reduced to less than 0.2, model misfits diverge (Figure 15). In contrast, in the case of the misfit function based on AMF, the inversion still converges toward the true model even with large velocity perturbations. Data and model misfits are gradually reduced (Figure 16). Thus, in these transmission waveform tomography experiments, the proposed AMF misfit function appears more robust for cases with poor starting models and high-frequency signals.

Modified Marmousi model

In this section, we use a modified Marmousi model as our final experiment. Figure 17 shows the starting $V(z)$ model and true model used in this experiment. In total, 12 shots and 181 receivers are used here, and each shot is recorded by all receivers. The shots

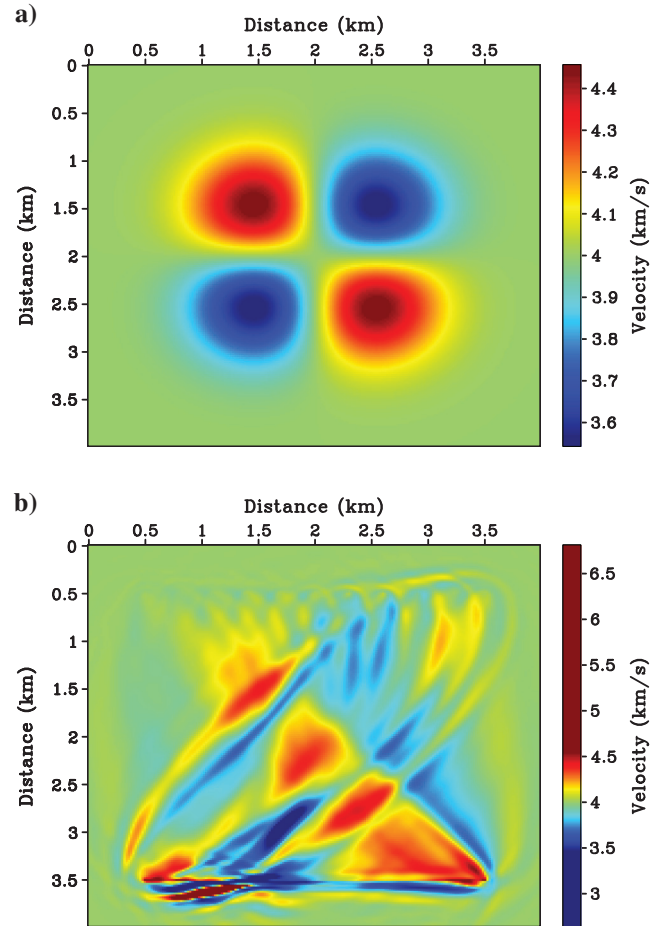


Figure 11. Inversion results based on L2 misfit function for Gaussian model. (a) Shows true velocity model with four Gaussian anomalies. Maximum velocity perturbation is 12.5%. (b) Recovered model based on L2 waveform differences.

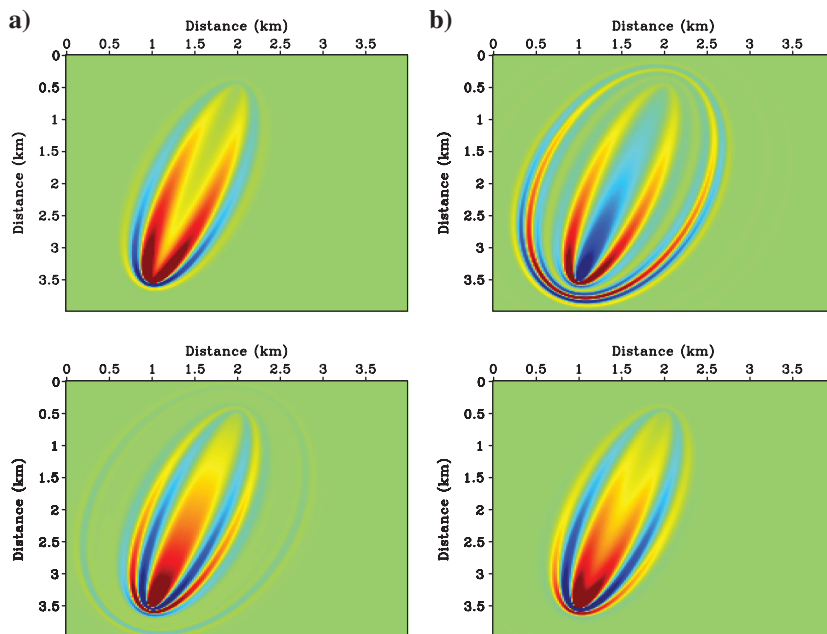


Figure 10. Comparisons of misfit gradients based on different misfit functions. Panel (a) for the case with small Gaussian perturbation and panel (b) for the case with large Gaussian perturbation. From top to bottom are misfit gradients based on L2 waveform differences and AMF, respectively.

and receivers are located at a depth of 5 m. The spacing between shots and receivers is 0.64 and 0.048 km, respectively. Maximum velocity perturbation in this experiment is greater than 30%. A Ricker wavelet with 10 Hz central frequency is used.

First, we run 10 iterations based on L2 waveform misfit function. Figure 18a shows the recovered velocity perturbations. The inversion appears to diverge away from the global minimum and reaches a local minimum. Data and model misfits increase significantly, indicating cycle skipping and local minima issues.

For the second case, we first perform six iterations with AMF misfit function. For the calculations of AMF, the stretching variable γ is allowed to vary from 0.9 to 1.1. We use triangle smoothing as the shaping operator and smoothing lengths are five grid points in both directions. The recovered perturbation and model are shown in Figure 19a and 19b, respectively. We observe that the AMF misfit function allows us to recover long-wavelength velocity structures even without low-frequency signals. We then use this updated model as the starting model and perform an additional 10 iterations with the L2 waveform misfit. The final recovered perturbations and model are displayed in Figure 19c and 19d, respectively. The inversion appears to converge toward the true model (compared with Figures 17 and 18).

Figure 20 shows that data and model misfits are steadily reduced for this inversion strategy. Because of the broad basins of the AMF misfit function, the reduction in data misfit is relatively slow for the first six iterations. For the final 10 iterations, due to the high resolution of the L2 waveform misfit, the reductions in data and model misfits are significant.

Finally, we compare predictions and data for two different shots in Figures 21 and 22. Because of a poor starting model, the predictions and data are obviously out of phase at the beginning. After six iterations with AMF misfits, we observe that the refraction and diving waves are corrected to fit the input data. After 10 additional iterations with L2 waveform misfit, we are able to fit diving waves and reflected waves. These comparisons again confirm the robustness of the proposed inversion strategy.

DISCUSSION

Choosing a well-behaved misfit function is crucial in FWI because of the following reasons. The FWI is an ill-posed inverse problem due

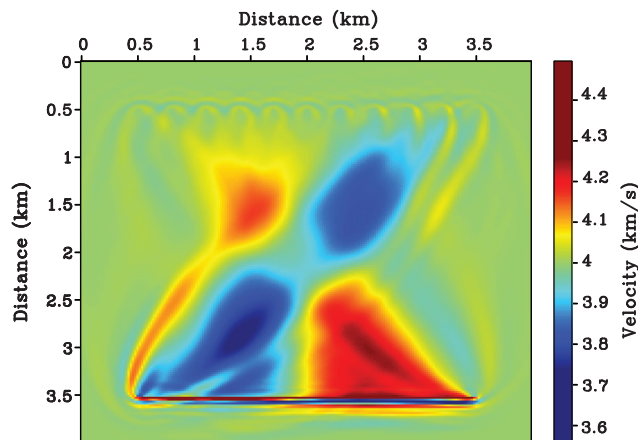


Figure 12. Inversion result based on AMF misfit function for Gaussian model.

to inaccurate, insufficient, and inconsistent observations (Jackson, 1972). It is also a highly nonlinear problem because of the strong nonlinear relations between waveform observations and seismic parameters. In addition, seismograms are oscillatory time series, which are subject to cycle skipping when phase differences between observations and predictions are greater than a half-period of signals

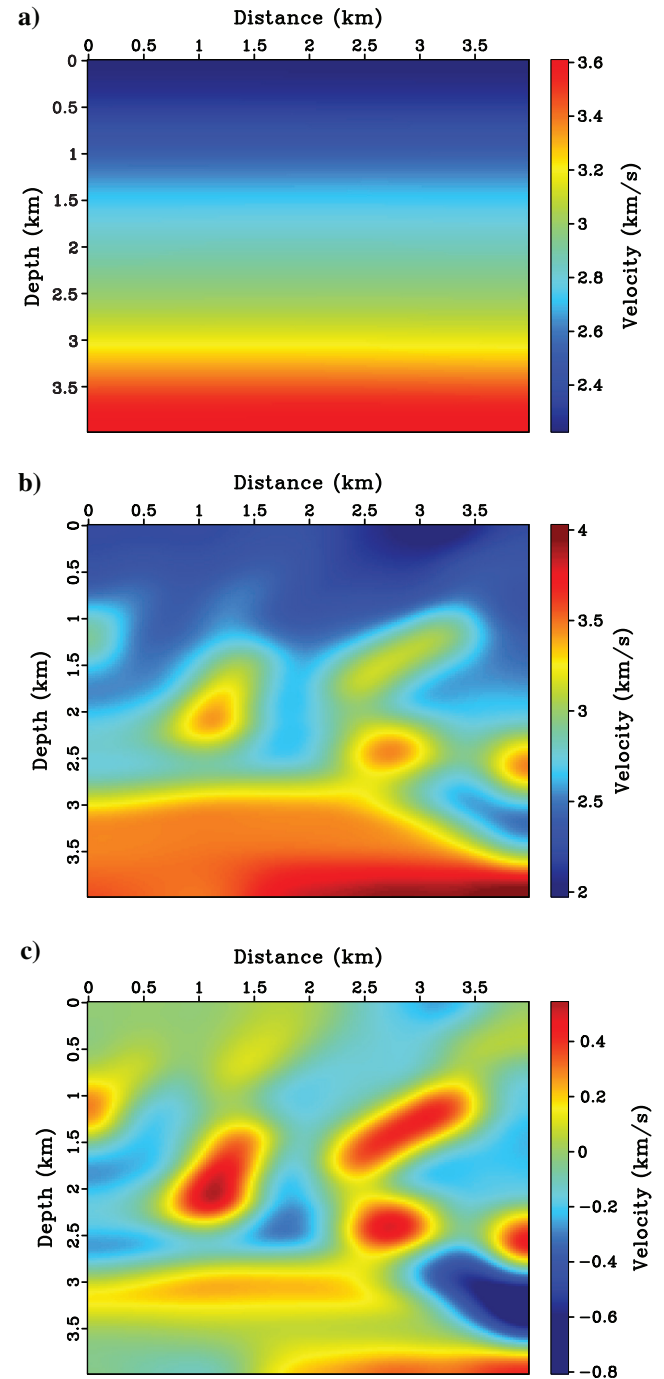


Figure 13. True and starting models for crosswell tomography. (a) Starting $V(z)$ model. (b) True velocity model and (c) perturbation between the true and starting models. For the experiment with small velocity perturbations, we add 50% maximum perturbations of (c) on the starting model (a) to generate the true model (see Figure 14).

(Virieux and Operto, 2009). This cycle-skipping problem usually leads to local minima in the misfit functions. Furthermore, most current FWI studies depend on local gradient-based optimization approaches, such as conjugate gradients (Fletcher and Reeves, 1964) or L-BFGS (Nocedal, 1980). If a misfit function involves numerous local minima, and the starting model is far away from the true model, these local gradient-based optimization approaches are doomed to fail. Because of these issues, FWI based on L2 waveform differences cannot always guarantee convergence toward the global minimum.

This problem becomes more serious when velocity perturbations are large, for instance exceeding 10%, which are ubiquitous at ex-

ploration scales. These difficulties might be mitigated by a multi-scale inversion strategy (Bunks et al., 1995; Sirgue and Pratt, 2004), which starts with low-frequency signals to constrain long-wavelength structures and gradually incorporates high-frequency signals to constrain short-wavelength features. However, field data rarely have enough low-frequency components, making this multiscale strategy nearly infeasible in practice.

In this paper, we propose a misfit function to directly tackle cycle skipping and local minima problems. This misfit function is based on AMF, which measures time-varying phase differences between observations and predictions. Because no amplitude information is

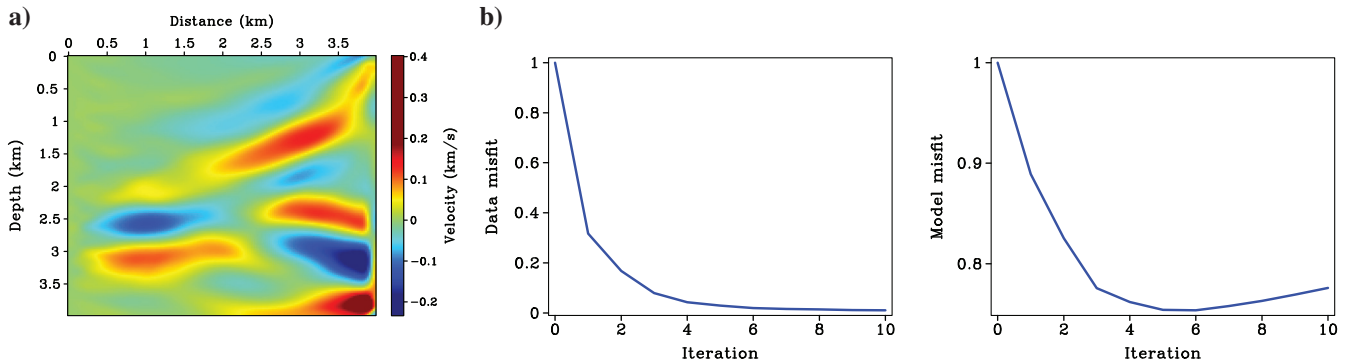


Figure 14. Inversion result for small velocity perturbations with L2 waveform differences. (a) The recovered model after 10 iterations. (b) Shows the evolution of data (left) and model (right) misfits.

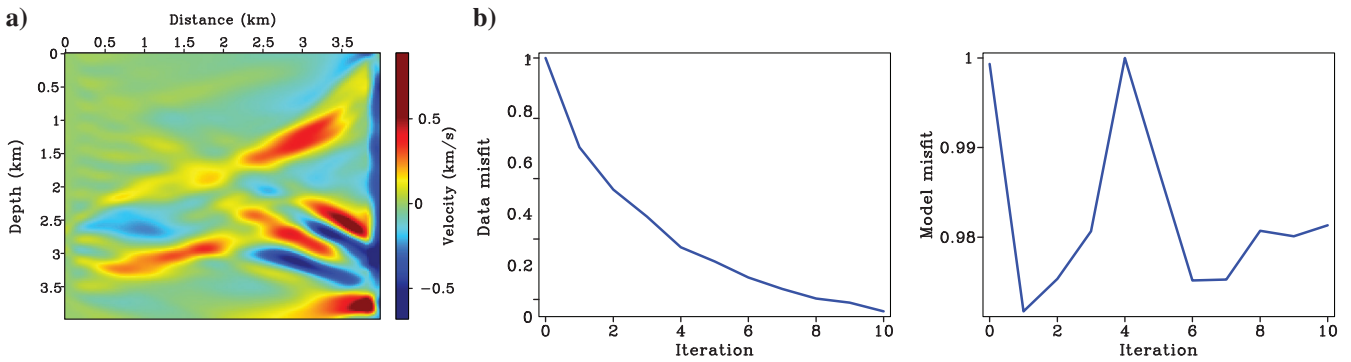


Figure 15. The same setting as Figure 14, except here large velocity perturbations are used. The inversion diverges toward local minima.

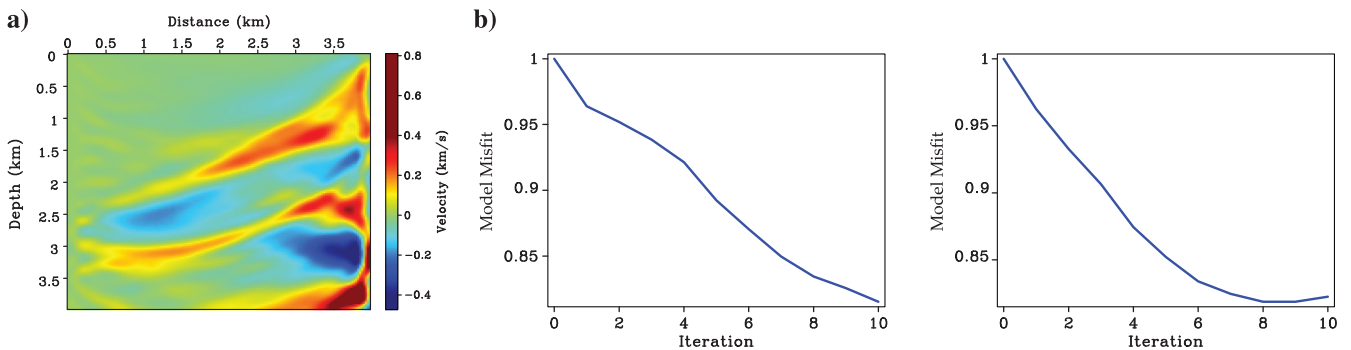


Figure 16. The same setting as Figure 15, except here we use AMF misfit function. The inversion converges toward correct velocity model (see Figure 14).

considered in this measurements, the proposed misfit is more robust than direct waveform differences. Based on numerical comparisons with L2 waveform misfit, we observe that the proposed misfit function behaves as a smooth, quadratic function. It also involves a broad basin of attraction, which can help when the starting model is far away from the true model. However, it may have slower convergence rates compared to the more high-resolution L2 waveform misfit. Therefore, our strategy is to use AMF misfit function to build a good starting model in cases we have relatively poor knowledge about the model. Once the iterations with AMF help us reach valleys near the

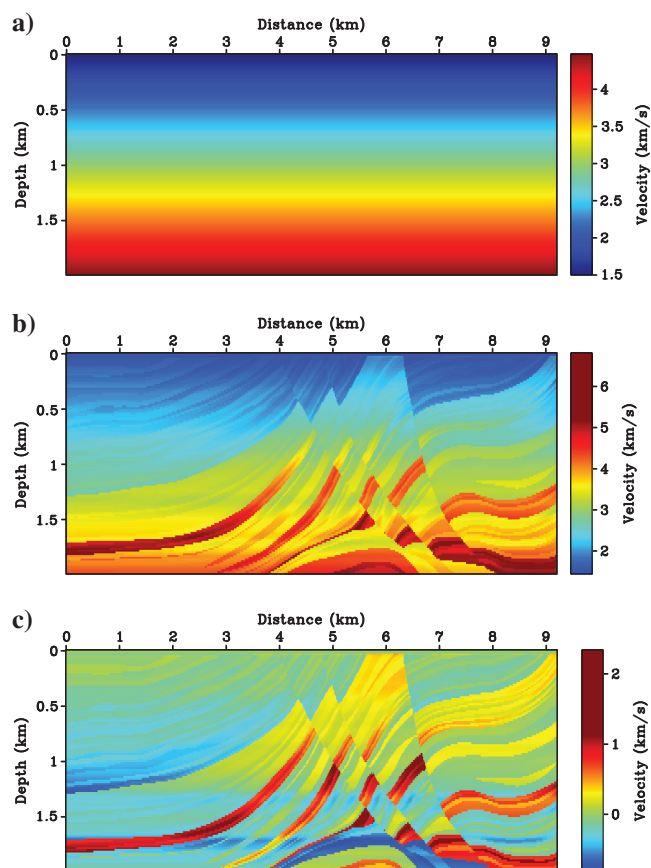


Figure 17. (a) Starting model, (b) true model, and (c) perturbations between the starting and true models for the modified Marmousi experiment.

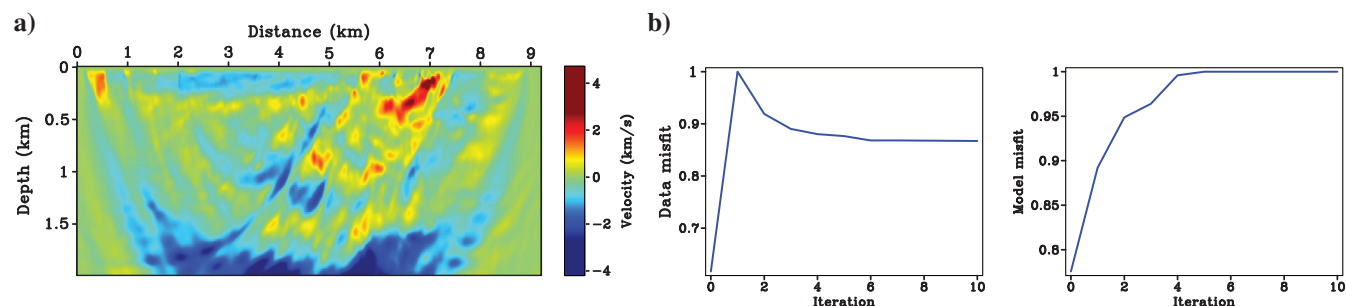


Figure 18. (a) Recovered perturbation and (b) the evolutions of data (left) and model (right) misfits based on L2 waveform misfit function for the modified Marmousi experiment.

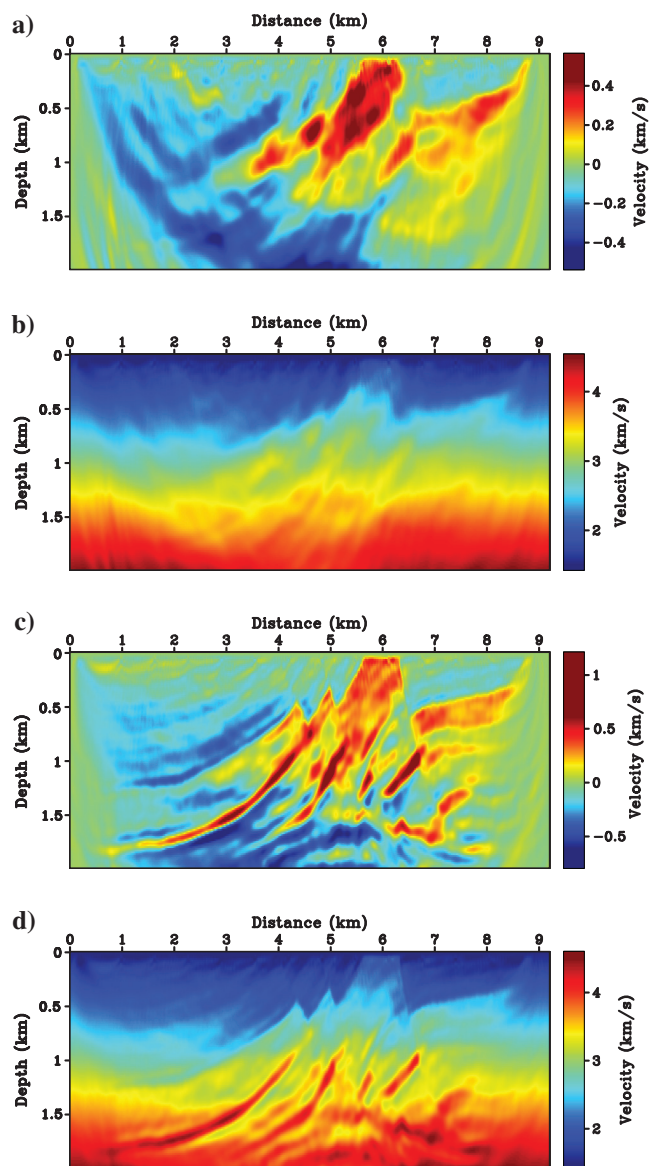


Figure 19. (a) Recovered perturbation and (b) model after six iterations based on AMF misfit. This recovered model is used as starting model for FWI based on L2 waveform differences. (c) Recovered perturbation and (d) model after additional 10 iterations based on L2 waveform differences.

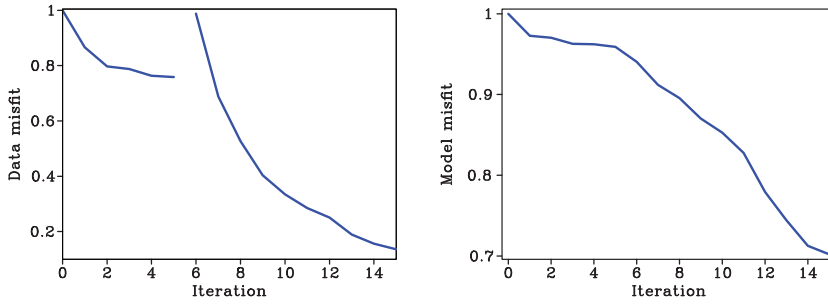


Figure 20. Evolution of data (left) and model (right) misfits for the modified Marmousi experiment. The first six iterations use AMF misfit and the final 10 iterations use L2 waveform misfit.

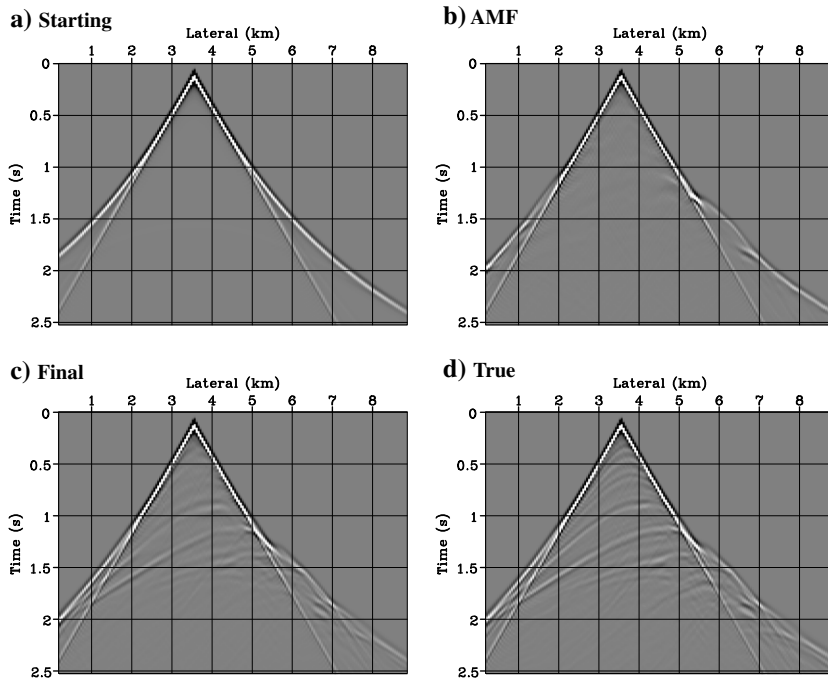


Figure 21. Comparisons between predictions and data for the modified Marmousi experiment. (a-d) Common-shot gathers for a shot at 2 km. (a) Predictions for the starting model, (b) predictions after six iterations with AMF misfit function, (c) predictions based on final recovered model, and (d) original data.

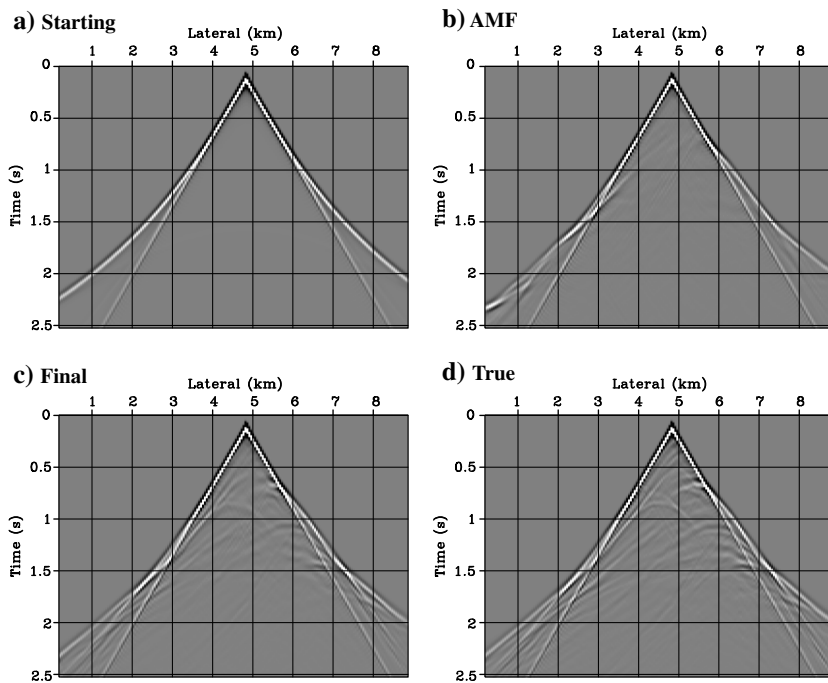


Figure 22. The same setting as Figure 21 except for a shot at 4 km.

global minimum, L2 waveform misfit can be used to speed up the inversion toward the global minimum. In this strategy, our initial model can be far away from the true model, and we do not require low-frequency components in the data to constrain long-wavelength structures. At the current stage, velocity updates in our inversion mainly rely on diving/transmitted waves. However, the AMF misfit function can be extended for reflection FWI in future research.

CONCLUSIONS

The goal of this study is to design well-behaved misfit functions for FWI to tackle challenges associated with cycle skipping and local minima. We propose a misfit function based on AMF, which measures time varying phase differences between observations and predictions. Based on numerical experiments, the proposed misfit function behaves as a smooth, quadratic function with a broad basin of attraction. It can increase probabilities for local gradient-based optimization approaches to reach the global minimum. We have applied the proposed misfit function to several 2D acoustic time-domain FWI experiments. Numerical results indicate that the proposed misfit function can help us build good starting models for FWI even when low-frequency signals are absent in the recorded data.

ACKNOWLEDGMENTS

This work was supported by the Jackson School Distinguished Postdoctoral Fellowship and the Texas Consortium for Computational Seismology. We appreciate discussions with H. Baek, B. Engquist, B. Froese, S. Treitel, and M. Warner. We thank the associate editor K. Innanen and three anonymous reviewers for comments and suggestions that helped to improve the manuscript. We thank the Texas Advanced Computing Center for providing computational resources.

REFERENCES

- Baek, H., H. Calandra, and L. Demanet, 2013, Velocity estimation via registration-guided least-squares inversion: *Geophysics*, **79**, no. 2, R79–R89, doi: [10.1190/geo2013-0146.1](https://doi.org/10.1190/geo2013-0146.1).
- Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2008, Velocity model building by 3D frequency domain, full-waveform inversion of wide-aperture seismic data: *Geophysics*, **73**, no. 5, VE101–VE117, doi: [10.1190/1.2957948](https://doi.org/10.1190/1.2957948).
- Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: *Geophysics*, **74**, no. 6, WCC105–WCC118, doi: [10.1190/1.3215771](https://doi.org/10.1190/1.3215771).
- Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **60**, 1457–1473, doi: [10.1190/1.1443880](https://doi.org/10.1190/1.1443880).
- Engquist, B., and B. Froese, 2014, Application of the Wasserstein metric to seismic signals: *Communications in Mathematical Sciences*, **12**, 979–988, doi: [10.4310/CMS.2014.v12.n5.a7](https://doi.org/10.4310/CMS.2014.v12.n5.a7).
- Fichtner, A., B. Kennett, H. Bunge, and H. Igel, 2008, Theoretical background for continent and global scale full-waveform inversion in the time-frequency domain: *Geophysical Journal International*, **175**, 665–685, doi: [10.1111/gji.2008.175.issue-2](https://doi.org/10.1111/gji.2008.175.issue-2).
- Fichtner, A., B. Kennett, H. Igel, and H. Bunge, 2009, Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods: *Geophysical Journal International*, **179**, 1703–1725, doi: [10.1111/gji.2009.179.issue-3](https://doi.org/10.1111/gji.2009.179.issue-3).
- Fletcher, R., and C. Reeves, 1964, Function minimization by conjugate gradients: *Computer Journal*, **7**, 149–154, doi: [10.1093/comjnl/7.2.149](https://doi.org/10.1093/comjnl/7.2.149).
- Fomel, S., 2007, Shaping regularization in geophysical-estimation problems: *Geophysics*, **72**, no. 2, R29–R36, doi: [10.1190/1.2433716](https://doi.org/10.1190/1.2433716).
- Fomel, S., 2009, Adaptive multiple subtraction using regularized nonstationary regression: *Geophysics*, **74**, no. 1, V25–V33, doi: [10.1190/1.3043447](https://doi.org/10.1190/1.3043447).
- Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: numerical results: *Geophysics*, **51**, 1387–1403, doi: [10.1190/1.1442188](https://doi.org/10.1190/1.1442188).
- Ha, T., W. Chung, and C. Shin, 2009, Waveform inversion using a back-propagation algorithm and a Huber function norm: *Geophysics*, **74**, no. 3, R15–R24, doi: [10.1190/1.3112572](https://doi.org/10.1190/1.3112572).
- Jackson, D., 1972, Interpretation of inaccurate, insufficient, inconsistent data: *Geophysical Journal of the Royal Astronomical Society*, **28**, 97–109, doi: [10.1111/j.1365-246X.1972.tb06115.x](https://doi.org/10.1111/j.1365-246X.1972.tb06115.x).
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration, in J. Bednar, ed., *Conference on inverse scattering: Theory and application*: SIAM, 206–220.
- Luo, Y., and G. Schuster, 1991, Wave-equation traveltime inversion: *Geophysics*, **56**, 645–653, doi: [10.1190/1.1443081](https://doi.org/10.1190/1.1443081).
- Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: *Geophysics*, **78**, no. 6, R223–R233, doi: [10.1190/geo2013-0004.1](https://doi.org/10.1190/geo2013-0004.1).
- Malinowski, M., S. Operto, and A. Bibodetti, 2011, High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full waveform inversion: *Geophysical Journal International*, **186**, 1179–1204, doi: [10.1111/j.1365-246X.2011.05098.x](https://doi.org/10.1111/j.1365-246X.2011.05098.x).
- Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: *Mathematics of Computation*, **35**, 773–773, doi: [10.1090/S0025-5718-1980-0572855-7](https://doi.org/10.1090/S0025-5718-1980-0572855-7).
- Plessix, R., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503, doi: [10.1111/gji.2006.167.issue-2](https://doi.org/10.1111/gji.2006.167.issue-2).
- Pratt, R., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi: [10.1046/j.1365-246X.1998.00498.x](https://doi.org/10.1046/j.1365-246X.1998.00498.x).
- 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, doi: [10.1190/1.1444597](https://doi.org/10.1190/1.1444597).
- Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field, Part 1: Imaging compressional wave speed, density and attenuation: *Geophysical Journal International*, **194**, 1640–1664, doi: [10.1093/gjg/ggt177](https://doi.org/10.1093/gjg/ggt177).
- Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: *Geophysics*, **73**, no. 5, VE49–VE59, doi: [10.1190/1.2972021](https://doi.org/10.1190/1.2972021).
- Shin, C., and Y. Cha, 1998, Waveform inversion in the Laplace domain: *Geophysical Journal International*, **173**, 922–931, doi: [10.1111/gji.2008.173.issue-3](https://doi.org/10.1111/gji.2008.173.issue-3).
- Sirgue, L., O. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: *First Break*, **28**, 65–70, doi: [10.3997/1365-2397.2010012](https://doi.org/10.3997/1365-2397.2010012).
- Sirgue, L., and G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248, doi: [10.1190/1.1649391](https://doi.org/10.1190/1.1649391).
- Symes, W., and J. Carazzone, 1991, Velocity inversion by differential semblance optimization: *Geophysics*, **56**, 654–663, doi: [10.1190/1.1443082](https://doi.org/10.1190/1.1443082).
- Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2009, Adjoint tomography of the southern California crust: *Science*, **325**, 988–992, doi: [10.1126/science.1175298](https://doi.org/10.1126/science.1175298).
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: [10.1190/1.1441754](https://doi.org/10.1190/1.1441754).
- van Leeuwen, T., and M. A. Mulder, 2010, A correlation-based misfit criterion from wave equation traveltime tomography: *Geophysical Journal International*, **182**, 1383–1394, doi: [10.1111/j.1365-246X.2010.04681.x](https://doi.org/10.1111/j.1365-246X.2010.04681.x).
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1092.
- Warner, M., A. Ratcliffe, and T. Nangoo, 2013, Anisotropic 3D full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80, doi: [10.1190/geo2012-0338.1](https://doi.org/10.1190/geo2012-0338.1).
- Wu, R., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: *Geophysics*, **79**, no. 3, WA13–WA24, doi: [10.1190/geo2013-0294.1](https://doi.org/10.1190/geo2013-0294.1).
- Yuan, Y., and F. Simons, 2014, Multiscale adjoint waveform-difference tomography using wavelets: *Geophysics*, **79**, no. 3, W79–W95, doi: [10.1190/geo2013-0383.1](https://doi.org/10.1190/geo2013-0383.1).
- Zhou, Y., F. Dahlen, and G. Nolet, 2004, Three-dimensional sensitivity kernels for surface wave observables: *Geophysical Journal International*, **158**, 142–168, doi: [10.1111/gji.2004.158.issue-1](https://doi.org/10.1111/gji.2004.158.issue-1).
- Zhu, H., E. Bozdog, T. Duffy, and J. Tromp, 2013, Seismic attenuation beneath Europe and the North Atlantic: Implications for water in the mantle: *Earth and Planetary Science Letters*, **381**, 1–11, doi: [10.1016/j.epsl.2013.08.030](https://doi.org/10.1016/j.epsl.2013.08.030).
- Zhu, H., E. Bozdog, D. Peter, and J. Tromp, 2012, Structure of the European upper mantle revealed by adjoint tomography: *Nature Geoscience*, **5**, 493–498, doi: [10.1038/ngeo1501](https://doi.org/10.1038/ngeo1501).
- Zhu, H., E. Bozdog, and J. Tromp, 2015, Seismic structure of the European upper mantle based on adjoint tomography: *Geophysical Journal International*, **201**, 18–52, doi: [10.1093/gjg/ggu492](https://doi.org/10.1093/gjg/ggu492).
- Zhu, H., and J. Tromp, 2013, Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean: *Science*, **341**, 871–875, doi: [10.1126/science.1241335](https://doi.org/10.1126/science.1241335).