GEOPHYSICS, VOL. 74, NO. 6 (NOVEMBER-DECEMBER 2009); P. WCC59–WCC68, 12 FIGS.
10.1190/1.3151869

An efficient multiscale method for time-domain waveform tomography

Chaiwoot Boonyasiriwat[1], Paul Valasek[2], Partha Routh[2], Weiping Cao[1], Gerard T. Schuster[1], and Brian Macy[2]

ABSTRACT

This efficient multiscale method for time-domain waveform tomography incorporates filters that are more efficient than Hamming-window filters. A strategy for choosing optimal frequency bands is proposed to achieve computational efficiency in the time domain. A staggered-grid, explicit finite-difference method with fourth-order accuracy in space and second-order accuracy in time is used for forward modeling and the adjoint calculation. The adjoint method is utilized in inverting for an efficient computation of the gradient directions. In the multiscale approach, multifrequency data and multiple grid sizes are used to overcome somewhat the severe local minima problem of waveform tomography. The method is applied successfully to 1D and 2D heterogeneous models; it can accurately recover low- and high-wavenumber components of the velocity models. The inversion result for the 2D model demonstrates that the multiscale method is computationally efficient and converges faster than a conventional, single-scale method.

INTRODUCTION

Traveltime tomography (Zhu and McMechan, 1989; Luo and Schuster, 1991; Pratt and Goulty, 1991; Schuster and Quintus-Bosz, 1993; Nemeth et al., 1997; Min and Shin, 2006) is a robust tool for estimating the subsurface velocity structure. By minimizing the difference between observed and calculated traveltimes, a velocity perturbation is calculated and the velocity model is updated iteratively. Although this method is fast and cost effective, the high-frequency approximation used by traveltime tomography violates the finite-frequency bandwidth of the seismic source. As a result, an estimate of the earth's velocity structure from traveltime tomography is suboptimal. Moreover, traveltime tomography only uses the traveltime of the first arrival and neglects other important information contained within the amplitude and phase of the recorded data.

To overcome the high-frequency assumption, Fresnel-volume or finite-frequency traveltime tomography (Červený and Soares, 1992; Woodward, 1992; Schuster and Quintus-Bosz, 1993; Vasco et al., 1995; Dahlen et al., 2000) and waveform tomography (Lailly, 1983; Tarantola, 1984, 1986; Mora, 1987) have been developed. Waveform tomography can be implemented in the frequency domain (Pratt, 1990; Liao and McMechan, 1996; Pratt et al., 1996; Pratt et al., 1998) or the time domain (Tarantola, 1984, 1986; Mora, 1987; Bunks et al., 1995; Zhou et al., 1995; Zhou et al., 1997). A velocity tomogram obtained using waveform tomography that inverts all arrivals has higher accuracy and resolution than one obtained by traveltime tomography (Sheng et al., 2006). However, the waveform misfit function is strongly nonlinear, and waveform tomography typically converges to a local minimum if the starting model is not in the vicinity of the global minimum (Gauthier et al., 1986).

Some efforts have been undertaken to overcome this local minima problem in the time domain (Bunks et al., 1995; Sheng et al., 2006) and the frequency domain (Sirgue and Pratt, 2004; Brenders and Pratt, 2007). In the time domain, Sheng et al. (2006) propose a method called early-arrival waveform tomography (EWT), in which a time window is applied to seismic data to allow for the inversion of only early-arrival wavefields. The misfit function for EWT is more linear with respect to slowness because fewer events need to be fitted than in conventional full waveform tomography. However, high-frequency data used in the inversion cause the misfit function to be highly nonlinear, and EWT can suffer from the local minima problem.

Alternatively, multiscale waveform tomography (MWT), proposed by Bunks et al. (1995) for time-domain waveform tomography, uses a finite-impulse response (FIR) Hamming-window filter for low-pass filtering of the seismic source wavelet and the data before inversion. This allows the inversion to proceed sequentially

Manuscript received by the Editor 27 December 2007; revised manuscript received 10 February 2009; published online 3 December 2009; corrected version published online 17 December 2009.
[1]University of Utah, Department of Geology and Geophysics, Salt Lake City, Utah, U.S.A. E-mail: chaiwoot@yahoo.com; wcao@mines.utah.edu; j.schuster@utah.edu.
[2]ConocoPhillips, Seismic Technology Development, Houston, Texas, U.S.A. E-mail: paul.a.valasek@conocophillips.com; partha.s.routh@conocophillips.com; brian.macy@conocophillips.com.

from low-frequency data to high-frequency data. Because the misfit function at low frequencies is more linear with respect to slowness than at high frequencies, MWT is more likely to reach the global minimum (Sirgue and Pratt, 2004).

The Hamming-window function used for low-pass filtering in Bunks et al. (1995) is not the most efficient filter for time-domain MWT. Furthermore, data with arbitrarily chosen frequency bands are used in the inversion method of Bunks et al. (1995). For each frequency band, data are inverted for some number of iterations, but the drawback is that too many frequency bands can result in a large amount of computation time. In the frequency domain, Sirgue and Pratt (2004) propose a strategy for choosing optimal frequencies. This can reduce the computational cost of waveform tomography greatly, but a modification of the method is required before it can be applied in the time domain.

The multiscale approach can be applied to frequency-domain waveform tomography because data already have been decomposed into separate frequency components using the Fourier transform. Combining the frequency selection strategy of Sirgue and Pratt (2004) with the multiscale approach, a frequency-domain waveform tomography method becomes much more efficient than a time-domain method, especially for 2D applications where forward modeling can be performed efficiently using LU decomposition, which is a direct linear solver. However, for 3D experiments, the LU decomposition used for solving the large, sparse linear system arising from frequency-domain forward modeling requires significant amounts of computer memory and thus limits the problem size and the maximum frequencies used for 3D applications (Operto et al., 2007). An iterative solver can avoid the memory issue (Warner et al., 2008), but the efficiency of an iterative solver relies strongly on a preconditioner. Without a good preconditioner, several iterations will be required, resulting in inefficient forward modeling.

Sirgue et al. (2007a, 2007b, 2008) propose a solution for 3D frequency-domain waveform inversion by using time-domain forward modeling and frequency-domain gradient computation. This hybrid method requires a discrete Fourier transform of the entire wavefield at each time step (Sirgue et al., 2008), which can be expensive to compute. In contrast, pure time-domain waveform tomography does not require such a large memory, preconditioner, and use of a discrete Fourier transform. Therefore, it is desirable to develop a more efficient multiscale waveform inversion method in the time domain.

In this paper, we improve the computational efficiency of the multiscale method of Bunks et al. (1995) by using a more efficient nonleaky low-pass filter and a time-domain strategy for choosing optimal frequency bands. Our filters are more efficient than the FIR Hamming-window low-pass filter used by Bunks et al. (1995) for time-domain multiscale waveform inversion. The strategy for selecting frequencies for frequency-domain waveform tomography proposed by Sirgue and Pratt (2004) is extended to time-domain waveform tomography. This revised strategy is verified by inversion results with 1D and 2D heterogeneous models.

We use true source wavelets in both experiments. In the 2D model experiment, we show that MWT can mitigate the local minima problem of waveform tomography and yield an accurate estimate of the velocity model, but conventional, single-scale waveform tomography converges to a local minimum, resulting in an inaccurate estimate of the true velocity model.

METHODS FOR TIME-DOMAIN WAVEFORM INVERSION

Acoustic waveform tomography

In this section, we review the theory of time-domain waveform tomography. The acoustic wave equation is used as our forward model, given by

$$\frac{1}{c^2(\mathbf{r})}\frac{\partial^2 p(\mathbf{r},t|\mathbf{r}_s)}{\partial t^2} - \nabla^2 p(\mathbf{r},t|\mathbf{r}_s) = s(\mathbf{r},t|\mathbf{r}_s), \tag{1}$$

where $p(\mathbf{r},t|\mathbf{r}_s)$ is a pressure field at position $\mathbf{r}$ at time t from a source at $\mathbf{r}_s$, $c(\mathbf{r})$ is the velocity model, and $s(\mathbf{r},t|\mathbf{r}_s)$ is the source function. The forward solution $p(\mathbf{r},t;\mathbf{r}_s)$ of the two-way wave equation 1 is computed by a staggered-grid, explicit finite-difference method with fourth-order accuracy in space and second-order accuracy in time (Levander, 1988; Sheng et al., 2006). The free-surface boundary condition is applied to the top boundary of the model, and the perfectly matched layer (PML) boundary conditions (Berenger, 1994; Chew and Liu, 1996; Zeng et al., 2001; Festa and Nielson, 2003) are used at the other boundaries. The solution to equation 1 can be written as

$$p(\mathbf{r},t|\mathbf{r}_s) = \int G(\mathbf{r},t|\mathbf{r}',0) * s(\mathbf{r}',t|\mathbf{r}_s)d\mathbf{r}', \tag{2}$$

where $G(\mathbf{r},t|\mathbf{r}',0)$ is the Green's function and the symbol $*$ represents temporal convolution.

Our inversion scheme is based on the adjoint method proposed by Tarantola (1984). The data residuals are defined as

$$\delta p(\mathbf{r}_g,t|\mathbf{r}_s) = p_{\text{obs}}(\mathbf{r}_g,t|\mathbf{r}_s) - p_{\text{calc}}(\mathbf{r}_g,t|\mathbf{r}_s), \tag{3}$$

where $p_{\text{obs}}(\mathbf{r}_g,t|\mathbf{r}_s)$ and $p_{\text{calc}}(\mathbf{r}_g,t|\mathbf{r}_s)$ are, respectively, the observed and calculated data. The velocity model is updated by minimizing the misfit function, defined as the L_2-norm of the data residuals:

$$E = \frac{1}{2}\sum_s \sum_g \int (\delta p(\mathbf{r}_g,t|\mathbf{r}_s))^2 dt. \tag{4}$$

A nonlinear preconditioned conjugate-gradient method (Luo and Schuster, 1991; Sheng et al., 2006) is used to minimize the misfit function. The gradient of the misfit function with respect to the velocity model is computed by the zero-lag correlation between the forward-propagated wavefields and back-projected wavefield residuals (Tarantola, 1984; Luo and Schuster, 1991; Zhou et al., 1995; Zhou et al., 1997; Sheng et al., 2006):

$$g(\mathbf{r}) = \frac{2}{c(\mathbf{r})}\sum_s \int \dot{p}(\mathbf{r},t|\mathbf{r}_s)\dot{p}'(\mathbf{r},t|\mathbf{r}_s)dt, \tag{5}$$

where $\dot{p}$ denotes the time derivative of p, $p(\mathbf{r},t|\mathbf{r}_s)$ represents the forward-propagated wavefields, and $p'(\mathbf{r},t|\mathbf{r}_s)$ represents the back-projected wavefield residuals, given by

$$p'(\mathbf{r},t|\mathbf{r}_s) = \int G(\mathbf{r},-t|\mathbf{r}',0) * \delta s(\mathbf{r}',t|\mathbf{r}_s)d\mathbf{r}' \tag{6}$$

and

$$\delta s(\mathbf{r}',t|\mathbf{r}_s) = \sum_g \delta(\mathbf{r}' - \mathbf{r}_g)\delta p(\mathbf{r}_g,t|\mathbf{r}_s). \quad (7)$$

The velocity model is updated iteratively along the conjugate directions defined by

$$\mathbf{d}_k = -\mathbf{P}_k\mathbf{g}_k + \beta_k\mathbf{d}_{k-1}, \quad (8)$$

where iterations $k = 1,2,\ldots,k_{\max}$, $\mathbf{g} = g(\mathbf{r})$ for all image points $\mathbf{r}$ in the model, and $\mathbf{P}$ is the conventional geometric spreading preconditioner (Causse et al., 1999).

At the first iteration, $\mathbf{d}_0 = -\mathbf{g}_0$. The parameter β_k is obtained using the Polak-Ribiére formula (Nocedal and Wright, 1999):

$$\beta_k = \frac{\mathbf{g}_k^T\cdot(\mathbf{P}_k\mathbf{g}_k - \mathbf{P}_{k-1}\mathbf{g}_{k-1})}{\mathbf{g}_{k-1}^T\cdot\mathbf{P}_{k-1}\mathbf{g}_{k-1}}. \quad (9)$$

The velocity model is updated by

$$c_{k+1}(\mathbf{r}) = c_k(\mathbf{r}) + \lambda_k d_k(\mathbf{r}), \quad (10)$$

where λ_k is the step length, determined by a quadratic line-search method (Nocedal and Wright, 1999), and $d_k(\mathbf{r})$ is the component of the conjugate-direction vector $\mathbf{d}_k$ at position $\mathbf{r}$. At each iteration, one forward propagation and one back projection are needed to compute the gradient direction. Additional forward modeling is required for the line search.

In this work, the starting model $c_0(\mathbf{r})$ is obtained using traveltime tomography with dynamic smoothing filters (Nemeth et al., 1997). Forward modeling and inversion are parallelized by shot numbers and computed on a Linux cluster of AMD Opteron 1-GHz processor. Equations 8–10 are applied iteratively until a stopping criterion is satisfied, which is when the maximum number of iterations is reached or when there is no further decrease in data residuals. In practice, the rms error or L_2-norm of the data residuals may be used as a stopping criterion, e.g., the inversion stops when the error is less than a previously chosen tolerance value.

Efficient low-pass filtering for MWT

Conventional time-domain waveform tomography (Tarantola, 1984) using one frequency band of the data and one finite-difference grid can lead to the local minima problem. By using several frequency bands of the data and varying grid sizes, the multiscale method of Bunks et al. (1995) successfully inverts for the complex Marmousi model. Low-pass filtering is crucial in making multiscale waveform inversion possible in the time domain.

For a given frequency band, a finite-difference grid size and a time step can be determined by the numerical dispersion and stability conditions associated with the chosen explicit finite-difference method. The numerical dispersion condition for the finite-difference scheme used in this paper requires at least five grid points per minimum wavelength (Levander, 1988); a square grid ($dx = dz$) is used in our finite-difference scheme. Therefore, the grid size used in forward modeling is determined by

$$dx \le \frac{\lambda_{\min}}{5} \le \frac{c_{\min}}{5f_{\max}}, \quad (11)$$

where dx is the grid size, $\lambda_{\min}$ is the minimum wavelength, $c_{\min}$ is the minimum velocity, and $f_{\max}$ is the maximum frequency of the band. Once the grid size is determined, the time step dt is determined by the 2D numerical stability condition (Bunks et al., 1995):

$$dt < \frac{dx}{\sqrt{2}c_{\max}}, \quad (12)$$

where $c_{\max}$ is the maximum velocity.

In the original work of Bunks et al. (1995), a Hamming-window function (Rabiner and Gold, 1975) is used for low-pass filtering the source wavelet and the data. The source wavelet is presumably known or estimated prior to the inversion. Several methods for estimating a seismic source wavelet have been proposed (Oldenburg et al., 1981; Lazear, 1993; Walden and White, 1998; Behura, 2007).

The Hamming-window function is a leaky low-pass filter as shown in Figure 1d by the solid line. Figure 1d shows that the leaked high-frequency components are weak compared to large-amplitude frequency components, e.g., a dominant-frequency component. However, these leaked high-frequency components can cause the maximum frequency $f_{\max}$ to be a large number; therefore, a small grid size is required to satisfy equation 11, resulting in computational inefficiency. If a large grid size is used, numerical dispersion can

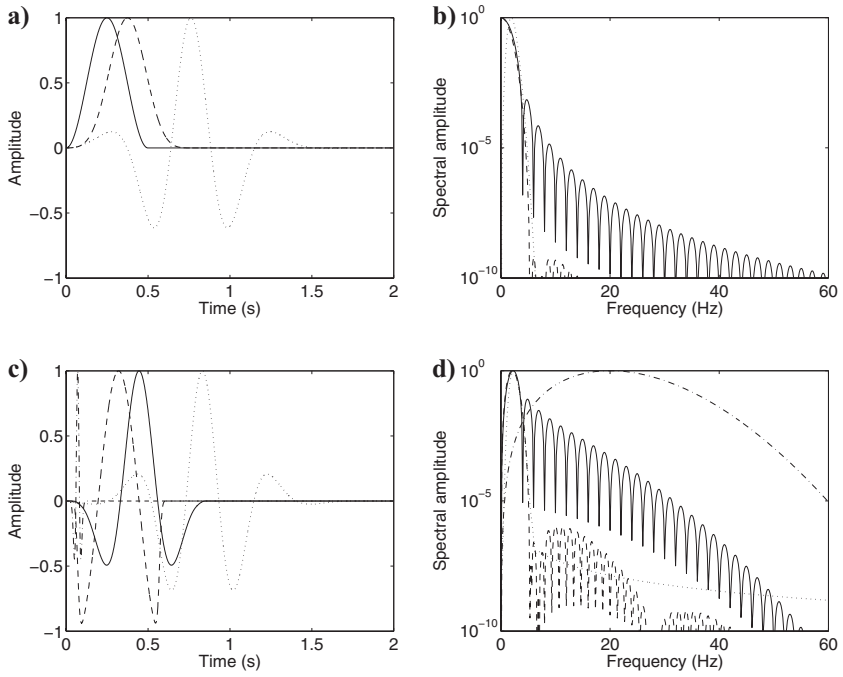

Figure 1. (a) Low-pass filters in the time domain. (b) Amplitude spectra of the low-pass filters in (a). (c) Original and filtered wavelets. The dash-dotted line is the original 20-Hz Ricker wavelet, low-pass filtered such that the peak frequency is about 2 Hz. (d) Amplitude spectra of the original and filtered wavelets in (c). In all views, the solid line is a Hamming window, the dashed line is a Blackman-Harris window, and the dotted line is a Wiener filter in the time domain.

occur in forward modeling and may create artifacts in inversion results or cause convergence problems. It is therefore desirable to implement an alternative low-pass filter for MWT that minimizes leaking of the high-frequency components.

Several suitable filters are considered for MWT, including Wiener filter, Blackman-Harris window function (Harris, 1978), Nuttall window function (Nuttall, 1981), and Ricker function. The filtering results of the first two filters in Figure 1 show they are more efficient than the Hamming-window function. The amplitudes of leaked high-frequency components from these filters are several orders of magnitude smaller than those from the Hamming filter (Figure 1d). Nuttall window and Ricker functions provide similar results but are not shown here. The Wiener filter usually is applied in the frequency domain; the other filters are applied in the time domain.

One merit of the Wiener filter is that it can filter one signal to closely match another target signal. A low-pass Wiener filter can be computed by

$$f_{\text{Wiener}}(\omega) = \frac{W_{\text{target}}(\omega) W^{\dagger}_{\text{original}}(\omega)}{|W_{\text{original}}(\omega)|^2 + \epsilon^2}, \tag{13}$$

where f_{Wiener} is the Wiener filter, W_{original} is the original wavelet, W_{target} is the low-frequency target wavelet, ω is the angular frequency, ϵ is a small parameter that prevents numerical overflow, and † denotes the complex conjugate. In Figure 1, the original and target wavelets are 20- and 2-Hz Ricker wavelets, respectively. The filtered wavelet obtained by Wiener filtering is not exactly the same as the target wavelet. However, the main purpose of our filtering is to obtain a wavelet with a low-frequency band and minimal spectral leakage.

Figure 1 shows the results of low-pass filtering a source wavelet. In time-domain MWT, low-pass filtering also is applied such that the filtered source wavelet and the data have roughly the same frequency range. In this paper, we use the Wiener filter, although other filters can be chosen.

Strategy for choosing optimal frequency bands

The strategy proposed by Sirgue and Pratt (2004) for choosing optimal frequencies for frequency-domain waveform tomography is extended to the time domain to reduce redundancy in the recovered wavenumber spectrum of a subsurface velocity structure. At a single frequency, the contribution of a single source-receiver pair has only a single wavenumber component. A range of vertical wavenumber components of the velocity model can be updated by using a range of source-receiver pairs (Sirgue and Pratt, 2004).

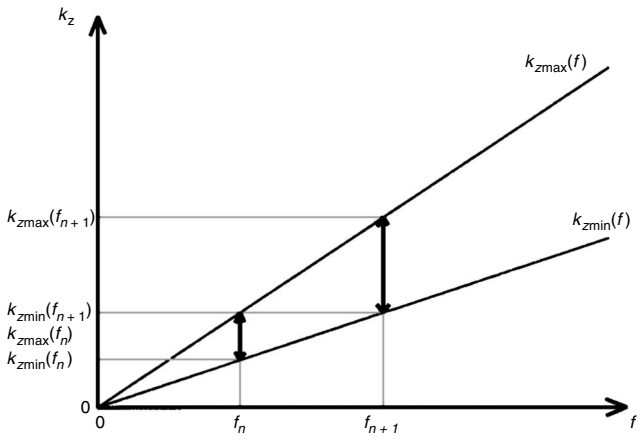

Figure 2. Strategy for choosing frequencies for frequency-domain waveform tomography. A range of vertical wavenumbers can be recovered by a single frequency component of the data by using a range of source-receiver pairs. A continuous coverage in vertical wavenumbers is the key for choosing the next frequency. (after Sirgue and Pratt, 2004)

Time-domain waveform tomography uses multiple frequencies simultaneously during the inversion. Therefore, it can update a much wider range of wavenumbers than using only a single frequency at a time, as in frequency-domain waveform tomography. However, the band-limited nature of seismic data, i.e., each frequency component of the data has a different amplitude or strength, results in a band-limited range of recovered wavenumber components of the velocity model without sharp boundaries. Strong frequency components of the data strongly contribute to the wavenumber update, whereas weak frequency components (especially at the low and high ends of a frequency band) have weak contributions and are ignored in our strategy.

The formula for choosing frequencies proposed by Sirgue and Pratt (2004) is

$$f_{n+1} = \frac{f_n}{\alpha_{\min}}, \tag{14}$$

where f_n is the current frequency, f_{n+1} is the next frequency to be chosen, and $\alpha_{\min} = z/\sqrt{h^2 + z^2}$ is the parameter that depends on the maximum half-offset h and the maximum depth to be imaged z. At f_n, the vertical wavenumber range $[k_{z\,\min}(f_n), k_{z\,\max}(f_n)]$ is updated and can be determined by

$$k_{z\,\min}(f_n) = \frac{4\pi f_n \alpha_{\min}}{c_o},$$
$$k_{z\,\max}(f_n) = \frac{4\pi f_n}{c_o}, \tag{15}$$

where c_o is the homogeneous background velocity. Equation 14 ensures that at the next frequency f_{n+1}, the lowest wavenumber to be updated is equal to the highest wavenumber at the current frequency,

$$k_{z\,\min}(f_{n+1}) = k_{z\,\max}(f_n), \tag{16}$$

as illustrated in Figure 2. Equations 14–16 are after Sirgue and Pratt (2004).

In time-domain inversion, multiple frequencies for a given bandwidth are used simultaneously. Therefore, a criterion is needed to determine the minimum and maximum frequencies of the band, which are used to calculate the recovered wavenumber range. Our criterion is to use the frequencies at half the maximum spectral amplitude of the frequency band.

This is illustrated in Figure 3b, which shows the amplitude spectrum of a 15-Hz Ricker wavelet (Figure 3a). At half the maximum amplitude, the lower frequency is set as the minimum frequency of the band $f_{\min}$, and the higher frequency is set as the maximum frequency $f_{\max}$. The frequency components whose amplitudes are smaller than half the maximum amplitude are considered to have insignificant contributions to the wavenumber recovery. Nonetheless, the contributions from these frequency components can be considered as some overlapped region between two recovered wavenumber ranges from two consecutive frequency bands.

The criterion for choosing minimum and maximum frequencies of a frequency band can be varied and depends on how much we

want two recovered wavenumber ranges to overlap. Sirgue and Pratt (2004) show that more frequency components are required for inverting noisy data. Similarly, larger overlapped regions of recovered wavenumber components are required in our case and result in more frequency bands in the inversion of noisy data.

In the nth frequency band, the minimum and maximum frequencies of the band $f_{\min}(n)$ and $f_{\max}(n)$, respectively, determine the wavenumber range:

$$k_{z\,\min}(n) = \frac{4\pi f_{\min}(n)\alpha_{\min}}{c_o},$$
$$k_{z\,\max}(n) = \frac{4\pi f_{\max}(n)}{c_o}. \quad (17)$$

In a manner similar to equation 16, at the next frequency band $n+1$, the lowest wavenumber to be updated is equal to the highest wavenumber at the current frequency band n:

$$k_{z\,\min}(n+1) = k_{z\,\max}(n). \quad (18)$$

This principle is illustrated in Figure 4.

In summary, the computational efficiency of time-domain acoustic waveform tomography increases by using a more efficient, nonleaky low-pass filter and a strategy for choosing optimal frequency bands. Regardless of computational efficiency, our method and the method of Bunks et al. (1995) can yield about the same result. Our method repeatedly uses low-frequency components of the data in the inversion, similar to the approach of Bunks et al. (1995).

The starting frequency bands used in this paper are chosen arbitrarily. A good starting frequency band depends on model complexity. We believe, for a specific model, there is a range of starting frequency bands to be used for successful inversion. However, our goal here is not to present a criterion for choosing a starting frequency band. In field data cases, a reasonable criterion is to use a frequency band with the lowest peak frequency allowed as the starting frequency band.

NUMERICAL RESULTS

We now apply MWT with optimally chosen frequency bands to two synthetic data sets generated from 1D and 2D models. The strategy for choosing frequencies is based on the 1D assumption in a homogeneous medium, originally developed for application in the frequency domain (Sirgue and Pratt, 2004). Thus, we need to validate our strategy in the time domain by applying it to a 1D model. Our strategy is then applied to a 2D model to show that it also applies to more realistic, geologic settings. The 2D model experiment is designed to show the advantages of MWT compared to conventional, single-scale waveform tomography (SWT). In both experiments, true source wavelets are used and data contain low-frequency components, generally unrealizable using current acquisition technology.

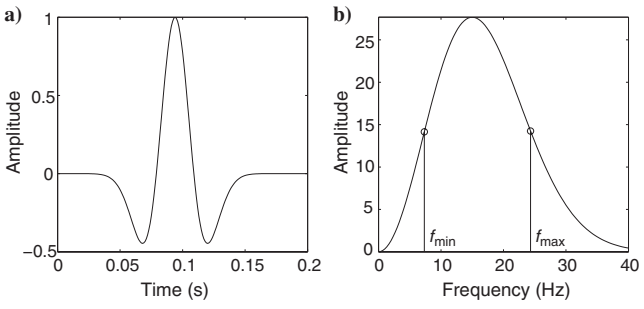

Figure 3. (a) A 15-Hz Ricker wavelet. (b) Its corresponding amplitude spectrum. The frequency at half the maximum spectral amplitude at the lower end is set as the minimum frequency of the band $f_{\min}$; the frequency at half the maximum spectral amplitude at the higher end is set as the maximum frequency $f_{\max}$.

1D model

Our 1D velocity model (Figure 5a) is similar to the model used by Sirgue and Pratt (2004), whose strategy is applied successfully to 1D and 2D heterogeneous models. Our experiment includes 81 sources and 401 receivers per shot; the maximum offset is 4 km. The sources and receivers are distributed evenly along the surface, with spatial intervals of 50 and 10 m, respectively. The thin layer located at 1.15 km (the bottom of the layer) is the target to be imaged. Figure 5b shows a shot gather from a source at $x = 0$ m.

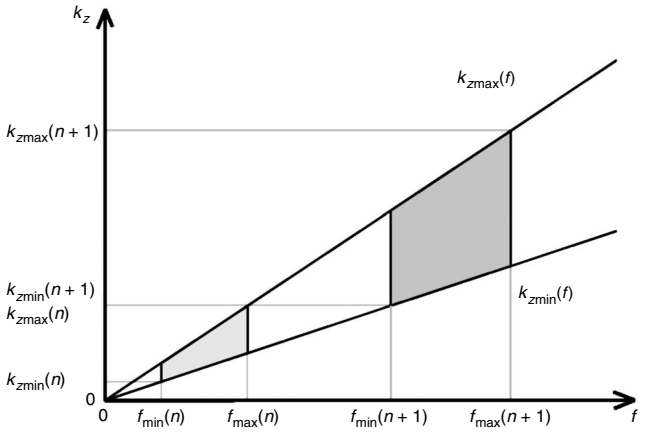

Figure 4. Strategy for choosing optimal frequency bands for time-domain waveform tomography. Data within a limited range of frequencies are inverted, allowing recovery of a range of vertical wavenumbers. Continuous coverage within the vertical wavenumber spectrum is the key for choosing the next frequency band.

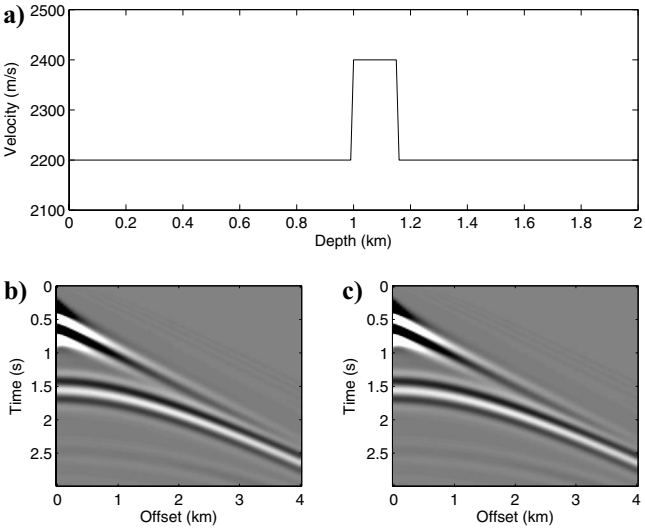

Figure 5. (a) A 1D velocity model. The target reflector is located at 1.15 km depth. (b) Shot gather of the 1D model generated from a 10-Hz Ricker source wavelet. (c) The shot gather low-pass filtered by the Wiener filter to the frequency band of a 2.5-Hz Ricker wavelet.

The original data were generated by using a 10-Hz Ricker wavelet, shown as the dashed line in Figure 6a, with a total record length of 3 s and a grid size of 10 m. To satisfy the numerical stability condition, the time-sampling interval is set to 2 ms. The source and recorded data are low-pass filtered using the Wiener filter to the bandwidth of a 2.5-Hz Ricker wavelet (solid line, Figure 6a). Figure 5b shows a shot gather after low-pass filtering. At this low-frequency band, we can use a larger grid size, $dx = 50$ m, and a larger time sampling interval, $dt = 10$ ms. Therefore, the inversion at this low-frequency band will be more efficient computationally, allowing us to compute many iterations to recover the low-wavenumber components of the velocity model as accurately as possible. The total number of model parameters to be estimated at this band is $81 \times 41 = 3321$. The filtered data have a total record length of 3 s, so 300 time steps are required to generate the calculated data P_{calc} for comparison with the observed data P_{obs}.

We now apply our strategy for choosing the frequency bands for the 1D experiment. The low-frequency band (solid line Figure 6b) is the first band used in the inversion, $n = 1$; the minimum and maximum frequencies at half the peak of this band are denoted as $f_{min}(1)$ and $f_{max}(1)$. The wavenumber range that can be recovered by this low-frequency band is shown in the solid area in Figure 6c. The next frequency band that satisfies equation 17 is for the 10-Hz Ricker wavelet, the original source wavelet. Therefore, the grid size and time step used for this high-frequency band are the same as the parameters used to generate the observed data originally. The total number of model parameters to be estimated with this band is $401 \times 201 = 80{,}601$, and the original data, with 1501 time samples per trace, are used in the inversion. The minimum and maximum frequencies of the second frequency band, $f_{min}(2)$ and $f_{max}(2)$, are shown in Figure 6b; the corresponding wavenumber range is shown in Figure 6c (solid area in the high-frequency range).

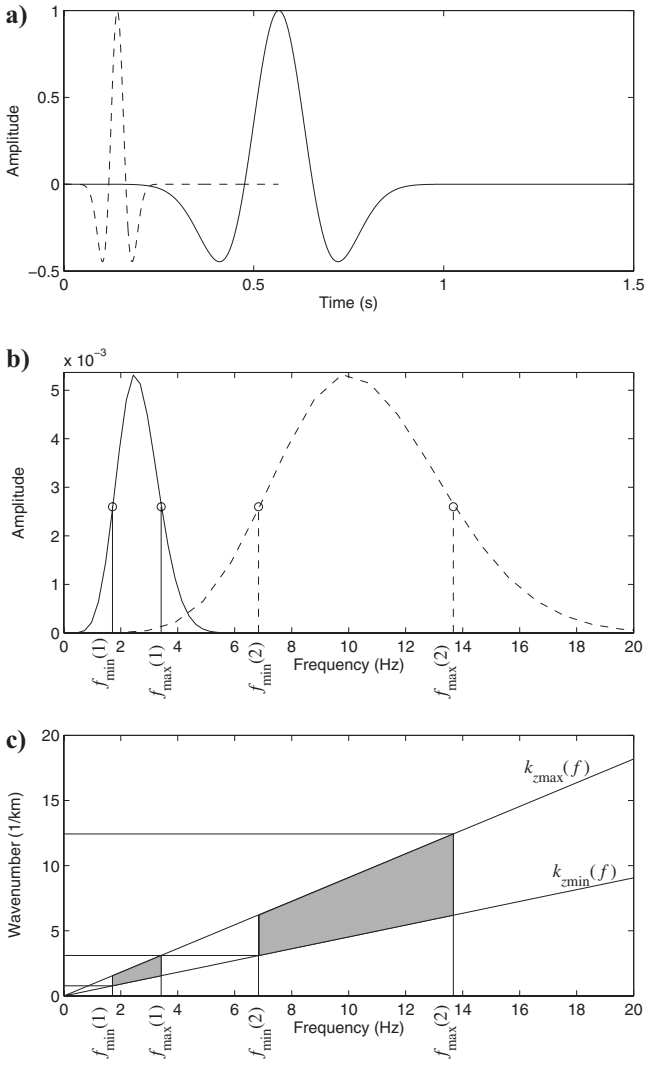

Figure 6. (a) Source (Ricker) wavelets used in the 1D model. (b) Amplitude spectra of the wavelets in (a) and sequential frequency ranges used to calculate recovered vertical wavenumber ranges. (c) Sequence of frequency bands generated by the strategy for choosing frequency range for the 1D model.

The inversion results of this experiment are shown in Figure 7. With the low-frequency-band data, the inversion was iterated for 50 iterations, starting from an initial homogeneous velocity model ($c_o = 2200$ m/s). Twenty iterations were carried out with the high-frequency-band data. The vertical profile of the final results at $x = 2$ km is shown in Figure 7a; its corresponding wavenumber components are shown in Figure 7b. Figure 7c and d shows the contributions for each frequency band to the final image. The low-wavenumber components of the velocity model can be recovered quite accurately using the low-frequency band, whereas the high-frequency band recovers the moderate- and high-wavenumber components.

2D model

The 2D model used in this experiment consists of several small-scale velocity anomalies embedded within a background of slowly varying macrolayers (Figure 8). These small-scale anomalies are very difficult to recover using conventional traveltime tomography. The seismic data for this experiment consist of 201 sources and 201 receivers per shot; the maximum offset is 4 km. The sources and receivers are distributed evenly along the surface, with spatial intervals of 20 m. Figure 9a shows a shot gather from a source at $x = 0$ m. The source signature used to generate the data is a 20-Hz Ricker wavelet. The grid size of 5 m and the time-sampling interval of 0.5 ms are used to generate the observed data, with a total record length of 2.5 s.

Both SWT and MWT were applied to this data set. In SWT, the original source and data were used in the inversion; in MWT, multiple frequency bands of the source and the data were used. In this experiment, the starting frequency band has a peak frequency of 5 Hz. By applying our strategy for choosing frequency bands to this experiment, only the frequency bands of the 5- and 20-Hz Ricker wavelets, i.e., the original source wavelet, are needed.

Figure 9b shows a shot gather after low-pass filtering. With the low-frequency band, we can use a grid size of 20 m and a time-sampling interval of 2 ms without risk of numerical dispersion and instability. The total number of model parameters to be estimated at this band is $201 \times 51 = 10251$, and the filtered data with the same total record length have 1249 time samples per trace. With the high-frequency band, the number of model parameters to be estimated increases to $801 \times 201 = 161{,}001$, and the original data with 5000 time samples per trace are used.

The initial velocity model used by SWT and MWT was obtained by the traveltime tomography method of Nemeth et al. (1997) with a 50-m square smoothing filter. One hundred iterations were carried out to obtain the traveltime velocity tomogram, which is incapable

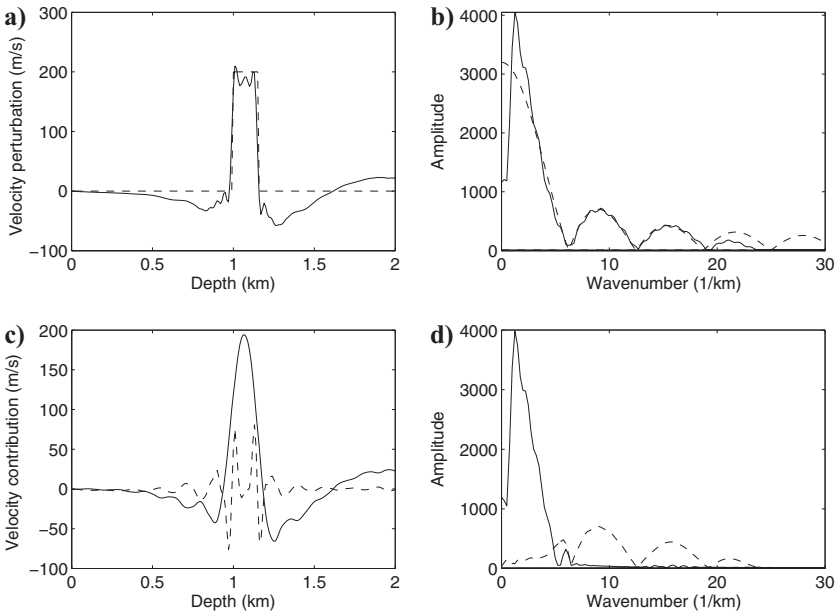

Figure 7. Time-domain multiscale waveform inversion results for the 1D model. True velocity perturbations in the (a) space and (b) wavenumber domains, are shown as a dashed line. The final estimated velocity perturbations from the background are a solid line. Contributions in the (c) space and (d) wavenumber domains from the two frequency bands. The solid lines are the contribution from low-frequency-band data (peak frequency = 2.5 Hz). The dashed lines are from high-frequency-band data (peak frequency = 10 Hz).

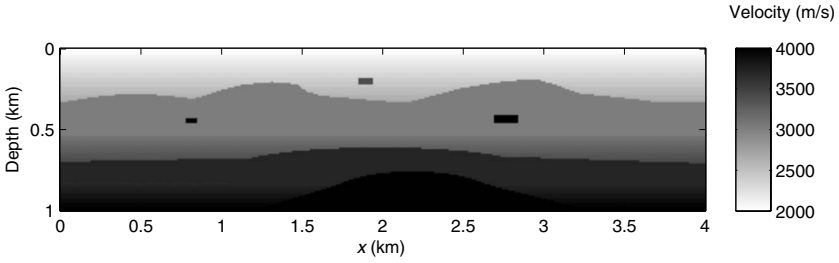

Figure 8. The 2D velocity model used to test the applicability of the strategy for choosing optimal frequency bands in a 2D heterogeneous case.

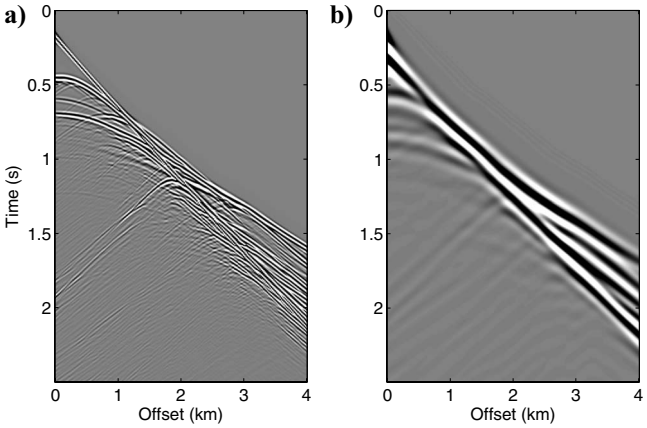

Figure 9. (a) Original shot gather of the 2D model generated from a 20-Hz Ricker source wavelet. (b) Shot gather low-pass filtered by the Wiener filter to the frequency band of a 5-Hz Ricker wavelet.

of resolving the small-scale anomalies (Figure 10a). The ray-density diagram from traveltime tomography (Figure 10b) can be used to determine reliable regions in the traveltime tomogram.

The inversion results of this experiment are shown in Figure 11. The data were inverted for 50 iterations with the low-frequency band, followed by 20 iterations with the high-frequency band. Inversion of low-frequency data was completed in 2 hours on our Linux cluster with 100 CPUs; the high-frequency inversion was completed in 12 hours. The velocity tomogram obtained after the low-frequency inversion is shown in Figure 11a. The tomogram obtained after the low- and high-frequency inversions, our final inversion result from MWT, is shown in Figure 11b. Figure 11c illustrates the final inversion result of SWT after 45 iterations. Although the maximum number of iterations was set to 70, the inversion was stopped at iteration 45 as a result of no further decrease in misfit. This indicates local minima problems encountered by SWT.

The final model from SWT in Figure 11c shows an effect of the local minima problem; the recovered model is much inferior to that obtained using MWT. The total run time of SWT is 23 hours, 64% longer than MWT, without additional benefit to the inversion result, as evident by comparing the SWT result to the MWT tomogram. The residual plot shown in Figure 12 shows that MWT residual is lower than SWT residual by 16%. Comparing the tomograms to the true model suggests that MWT converged to the global minimum but SWT did not. By applying MWT successfully to a 2D model and comparing the results to conventional single-scale waveform to-

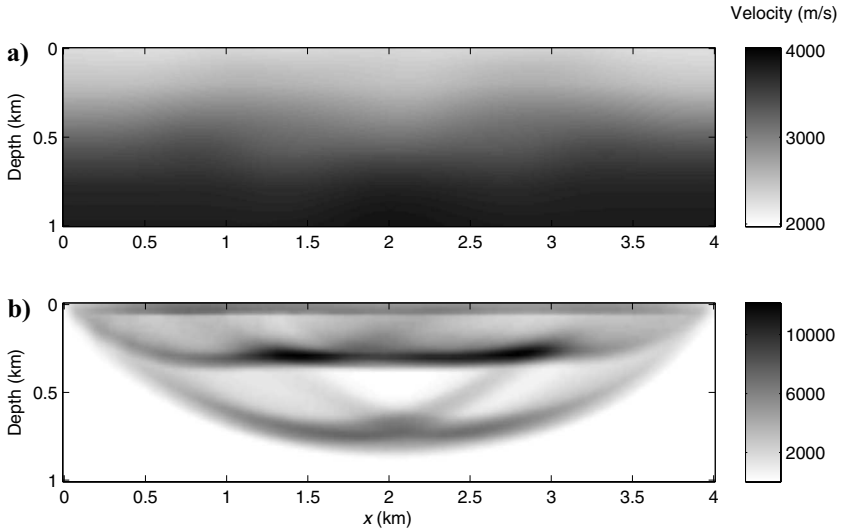

Figure 10. (a) Velocity tomogram from traveltime tomography, the initial velocity model for waveform tomography methods used in the 2D model experiment. (b) Ray density diagram from traveltime tomography. Note the number of rays passing through a model block. Also note the areas where waveform tomography will improve the result using refraction energies.

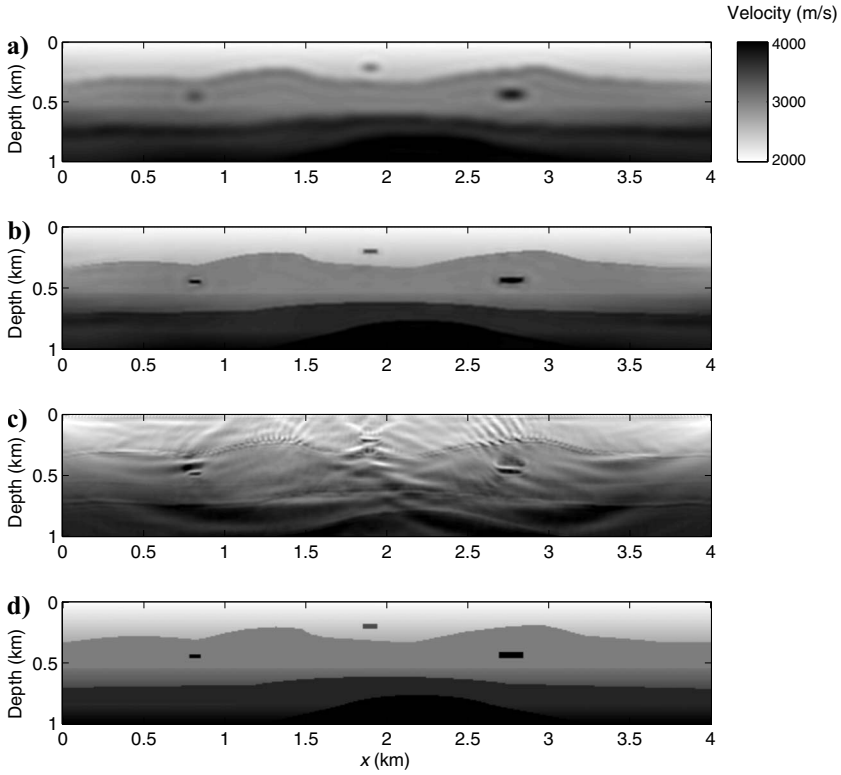

Figure 11. Time-domain multiscale waveform inversion results for the 2D model: (a) MWT velocity tomogram obtained after inversion using 5-Hz peak-frequency data, (b) MWT velocity tomogram obtained after inversions using 5- and 20-Hz peak-frequency data, (c) SWT velocity tomogram obtained after inversion using 20-Hz peak-frequency data, and (d) the true velocity model.

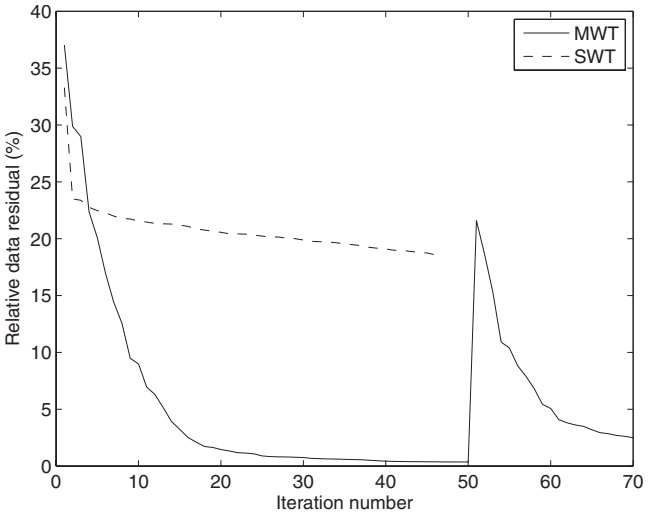

Figure 12. Relative data-residual plot. The solid line shows the relative data residuals of MWT versus the iteration number. The residuals at the first 50 iterations are from the low-frequency inversion using data with a peak frequency of 5 Hz. The last 20 residuals are from the high-frequency inversion using data with a peak frequency of 20 Hz. The dashed line shows the relative data residuals of SWT.

mography, it is clear that MWT has more computational efficiency and a faster convergence rate than single-scale waveform tomography.

CONCLUSIONS

We have developed a multiscale method that is more efficient than a time-domain waveform tomography method. Increased computational efficiency is achieved by using the Wiener filter and the strategy for choosing optimal frequency bands. We propose the Wiener filter for efficient low-pass filtering for MWT in the time domain. By applying the strategy for choosing optimal frequency bands, fewer frequency bands are needed and unnecessary computational costs from using too many bands are avoided. At low frequencies, forward modeling and inversion are very efficient because larger grid sizes and time steps can be used than at higher frequencies. By progressively recovering higher wavenumber components of the velocity structure, in a low to high fashion, the multiscale approach can improve the convergence property of waveform tomography and partially overcome the problem of encountering local minima in the misfit function.

The frequency-band selection strategy was validated in an experiment with a 1D model. By using the two frequency bands calculated by our strategy, low- and high-wavenumber components of the model were recovered accurately. This result agrees with published work for waveform inversion in the frequency domain. Results from an experiment with a 2D model suggest that our multiscale method, when applied for a 2D heterogeneous case, can increase computational efficiency and convergence rate when compared to a conventional waveform tomography method.

In the field data cases with missing low-frequency components, Wiener filtering using a Ricker wavelet as a target low-frequency band can be inefficient. The other proposed filters can be used instead. The criteria for selecting minimum and maximum frequencies in a frequency band can be varied, depending on the data set, and several frequency bands may be required.

ACKNOWLEDGMENTS

We are grateful for support from the members of the University of Utah Tomography and Modeling/Migration Consortium. We thank associate editor Bill Harlan, reviewer Andrew J. Brenders, and three anonymous reviewers for many useful comments and helpful suggestions. C. Boonyasiriwat thanks the seismic technology group at ConocoPhillips for their help and valuable discussions. Finally, we thank Kris Sikorski and Sherif Hanafy for their useful suggestions.

REFERENCES

Behura, J., 2007, Virtual real source: 77th Annual International Meeting, SEG, Expanded Abstracts, 2693–2696.
Berenger, J., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200.
Brenders, A. J., and R. G. Pratt, 2007, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.
Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.
Causse, E., R. Mittet, and B. Ursin, 1999, Preconditioning for full-waveform inversion in viscoacoustic media: Geophysics, **64**, 130–145.
Červený, V., and E. Soares, 1992, Fresnel volume ray tracing: Geophysics, **57**, 902–915.
Chew, W., and Q. Liu, 1996, Perfectly matched layers for elastrodynamics: Journal of Computational Acoustics, **4**, 341–359.
Dahlen, F. A., S. H. Hung, and G. Nolet, 2000, Fréchet kernels for finite-frequency travel-times — I. Theory: Geophysical Journal International, **141**, 157–174.
Festa, G., and S. Nielson, 2003, PML absorbing boundaries: Bulletin of the Seismological Society of America, **93**, 891–903.
Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403.
Harris F. J., 1978, On the use of windows for harmonic analysis with the discrete Fourier transform: Proceedings of the IEEE, **66**, 51–84.
Lailly, P., 1983, The seismic inverse problems as a sequence of before stack migration: Conference on Inverse Scattering Theory and Application, Society of Industrial and Applied Mathematics, Proceedings, 206–220.
Lazear, G. D., 1993, Mixed-phase wavelet estimation using fourth-order cumulants: Geophysics, **58**, 1042–1051.
Levander, A., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1437.
Liao, Q., and G. A. McMechan, 1996, Multifrequency viscoacoustic modeling and inversion: Geophysics, **61**, 1371–1378.
Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.
Min, D., and C. Shin, 2006, Refraction tomography using a waveform-inversion back-propagation technique: Geophysics, **71**, no. 3, R21–R30.
Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228.
Nemeth, T., E. Normark, and F. Qin, 1997, Dynamic smoothing in crosswell traveltime tomography: Geophysics, **62**, 168–176.
Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer series in operations research and financial engineering: Springer-Verlag New York, Inc.
Nuttall, A. H., 1981, Some windows with very good sidelobe behavior: IEEE Transactions on Acoustics, Speech, and Signal Processing, **29**, 84–91.
Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: Geophysics, **46**, 1528–1542.
Operto, S., J. Virieux, P. Amestoy, J. L'Excellent, L. Giraud, and J. B. H. Ali, 2007, 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, no. 5, SM195–SM211.
Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography — Part I: Acoustic wave-equation method: Geophysical Prospecting, **39**, 287–310.
Pratt, R. G., and N. R. Goulty, 1991, Combining wave-equation imaging with traveltime tomography to form high-resolution images from crosshole data: Geophysics, **56**, 208–224.
Pratt, R. G., 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.
Pratt, R. G., Z. M. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340.

Rabiner, L. R., and B. Gold, 1975, Theory and application of digital signal processing: Prentice-Hall, Inc.
Schuster, G. T., and A. Quintus-Bosz, 1993, Wavepath eikonal traveltime inversion: Theory: Geophysics, **58**, 1314–1323.
Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**, no. 4, U47–U57.
Sirgue, L., J. T. Etgen, and U. Albertin, 2007a, 3D full-waveform inversion: Wide-versus narrow-azimuth acquisitions: 77th Annual International Meeting, SEG, Expanded Abstracts, 1760–1764.
——–, 2008, 3D frequency-domain waveform inversion using time-domain finite-difference methods: 70th Annual International Meeting, EAGE, Extended Abstracts.
Sirgue, L., J. T. Etgen, U. Albertin, and S. Brandsberg-Dahl, 2007b, 3D frequency-domain waveform inversion based on 3D time-domain forward modeling: U. S. Patent 20 070 282 535 A1.
Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.
Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.
——–, 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903.
Vasco, D. W., J. E. Peterson, and E. L. Majer, 1995, Beyond ray tomography: Wavepaths and Fresnel volumes: Geophysics, **60**, 1790–1804.
Walden, A. T., and R. E. White, 1998, Seismic wavelet estimation: A frequency domain solution to a geophysical noisy input-output problem: IEEE Transactions on Geoscience and Remote Sensing, **36**, 287–297.
Warner, M., I. Stekl, A. Umpleby, J. Morgan, C. Pain, and Y. Wang, 2008, 3D wavefield tomography: Problems, opportunities and future directions: 70th Annual International Meeting, EAGE, Extended Abstracts.
Woodward, M. J., 1992, Wave-equation tomography: Geophysics, **57**, 15–26.
Zeng, Y., J. Q. He, and Q. Liu, 2001, The application of the perfectly matched layer in numerical modeling of wave propagation in poroelastic media: Geophysics, **66**, 1258–1266.
Zhou, C., W. Cai, Y. Luo, G. T. Schuster, and S. Hassanzadeh, 1995, Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data: Geophysics, **60**, 765–773.
Zhou, C., G. T. Schuster, S. Hassanzadeh, and J. M. Harris, 1997, Elastic wave-equation traveltime and waveform inversion of crosshole seismic data: Geophysics, **62**, 853–868.
Zhu, X., and G. A. McMechan, 1989, Estimation of a two-dimensional seismic compressional-wave velocity distribution by iterative tomographic imaging: International Journal of Imaging Systems and Technology, **1**, 13–17.