

Multiscale adjoint waveform-difference tomography using wavelets

Yanhua O. Yuan¹ and Frederik J. Simons¹

ABSTRACT

Full-waveform seismic inversions based on minimizing the distance between observed and predicted seismograms are, in principle, able to yield better-resolved earth models than those minimizing misfits derived from traveltimes alone. Adjoint-based methods provide an efficient way of calculating the gradient of the misfit function via a sequence of forward-modeling steps, which, using spectral-element codes, can be carried out in realistically complex media. Convergence and stability of full-waveform-difference adjoint schemes are greatly improved when data and synthetics are progressively presented to the algorithms in a constructive multiscale approximation using a (bi)orthogonal wavelet transform. Wavelets provide the nonredundant spectral decomposition that paves the way for the inversion to proceed successively from long-wavelength fitting to detailed exploration of the phases in the seismogram. The choice of wavelet class and type, the initial depth of the multiscale decomposition, and the minimization algorithms used at every level continue to play crucial roles in our procedure, but adequate choices can be made that test successfully on 2C elastic seismograms generated in toy models, as well as in the industry-standard 2D Marmousi model. Although for simplicity our inversion ignored surface waves by prior tapering and filtered removal, those also appeared to be very well matched in the final model.

INTRODUCTION

This paper develops a strategy for waveform inversion, a tool that is related to seismic migration. The objective of prestack depth migration is to obtain an image of the subsurface reflectivity field; in most implementations, the image is not intended as a quantitative measure of the true subsurface impedances. The objective of seismic tomography is to find reliable models that minimize the

difference between seismograms recorded at stations and synthetically computed waveforms. The differences can be variously measured in terms of picked arrival times, crosscorrelation traveltimes, amplitude anomalies, or via direct waveform subtraction. The common goals in this branch of seismology are to obtain accurate, high-resolution velocity models, to accomplish this using efficient numerical solution strategies, and to appraise the uncertainties of the model results. Our study focuses on the first two objectives by developing an algorithm that regularizes the inversion via the use of wavelet-based constructive approximations applied to elastic waveform data, synthetic and observed, in a model that evolves as part of a gradient-based iterative scheme relying on forward and adjoint modeling carried out with a spectral-element method. We neither explicitly optimize the numerical efficiency nor do we spend much time on the a posteriori model evaluation or uncertainty analysis, but rather, we present and document the performance of a series of straightforward data processing steps that gently guide the often highly nonlinear waveform-difference adjoint tomography toward accurate and well-resolved models, even when starting from poor initial models, without much additional burden on computation time.

In the previous decade, seismic tomography has evolved from ray-based (Červený, 2001) to finite-frequency kernel-based methods (Marquering et al., 1999; Dahlen et al., 2000; Zhao et al., 2000, 2005; Dahlen and Baig, 2002). Under ray theory, which is an infinite-frequency approximation, traveltime changes are solely influenced by velocity perturbations along the geometric raypaths. In reality, phase and amplitude shifts are affected by velocity structure throughout the whole earth model. The main influence zone is centered along the geometric ray within an ellipsoid-like volume that takes the shape of a banana, but because the sensitivity is zero on the geometric path, there is a hole along the ray (that is, for cross-correlation traveltime measurements in simple background models) — hence the name *banana-doughnut kernels* (Marquering et al., 1999; Nolet et al., 2005). Under finite-frequency kernel theory, measurement anomalies can be expressed as a volume integral of sensitivity kernels against the corresponding model perturbations (Hung et al., 2000, 2001; Montelli et al., 2004; Nolet, 2008).

Manuscript received by the Editor 22 October 2013; revised manuscript received 30 January 2014; published online 22 May 2014.

¹Princeton University, Department of Geosciences, Princeton, New Jersey, USA. E-mail: yanhua@princeton.edu; fjsimons@alum.mit.edu.

© 2014 Society of Exploration Geophysicists. All rights reserved.

Fréchet sensitivity kernels can be calculated approximately using dynamic ray theory (Tian et al., 2007a, 2007b) or better yet, estimated numerically via the adjoint method by back-projecting residuals between observations and synthetics at seismic stations. These interact with the forward-propagating wavefield to yield an image of the discrepancies between the estimated and the true model (Tromp et al., 2005, 2010; Nissen-Meyer et al., 2007a; Tape et al., 2007). The number of simulations required scales with the number of earthquake sources or exploration shots, regardless of the number or type of measurements in use. Above all, the quality of inversions relies on accurate forward modeling. The advent of spectral-element methods, which combine the flexibility of finite-element schemes with the accuracy of pseudospectral methods, has enabled accurate forward and adjoint wave-propagation simulations in arbitrary earth-like models (Komatitsch et al., 2002, 2005; Nissen-Meyer et al., 2007b, 2008).

The choice of misfit functional is a crucial factor in adjoint-based optimization (Fichtner et al., 2008; Bozdağ et al., 2011; Rickers et al., 2012). Crosscorrelation traveltimes measurements (Luo and Schuster, 1991) are commonly used. These correspond to the offset that maximizes the crosscorrelation between predicted data and observations. Compared to waveform inversion, traveltimes anomalies are more linearly dependent on relative velocity perturbations and there are fewer local minima in the objective function, which render the inversion problem easier to implement (Pratt, 1999; Alkhalifah and Choi, 2012). However, crosscorrelation traveltimes inversion is insensitive to much of the information in the waveforms generated in complex earth models.

Efforts to seek globally optimal models by iterative optimization techniques were pioneered by Lailly (1983), Tarantola (1984, 1986), and Mora (1987, 1988). The growth in computer power has enabled full-waveform-difference inversions, which are able to yield higher resolution images — provided that the inversions converge (Gauthier et al., 1986; Mora, 1987; Bunks et al., 1995). However, the primary difficulty preventing the direct and wide application of iterative full-waveform inversion is the presence of numerous local minima in the objective function (Alkhalifah and Choi, 2012). The effects of nonlinearity are especially severe when the starting model is not in the neighborhood of the target or when the model contains details of great complexity.

Many careful schemes have been designed to avoid or alleviate the local-minima problem. One possible solution is to use hybrid approaches, which rely on traveltimes tomography to build good initial models for subsequent full-waveform inversion (Pratt and Shipp, 1999; Operto et al., 2004; Sirgue and Pratt, 2004). However, picking arrival times for inversions in smooth starting models is not an easy task.

Another promising strategy to combat nonlinearity in waveform inversion is to use a multiscale approach to waveform fitting. Working successively from long to short wavelengths is usually a powerful strategy to approach the global minimum (Nolet et al., 1986) — if indeed such a state exists in an absolute sense. For long-wavelength measurements, the number of local minima is greatly reduced and the inversion problem is more likely to converge faster to the global solution or to a local minimum in the neighborhood of the global one, than for the full problem (Bunks et al., 1995; Brossier et al., 2009). The updated solution from the quasilinear large-scale problem can serve as a starting point that is closer to the global target for subsequent inversions at smaller scales.

In seismic tomography, but also in other fields such as electrical-resistivity tomography (e.g., Plattner et al., 2012), multiscale methods in a formal, multiresolution sense using *wavelet* bases and frames have been formulated in data space (Pratt, 1999; Wu et al., 2007; Shin and Cha, 2008, 2009), in model space (Chiao and Kuo, 2001; Loris et al., 2007; Hung et al., 2010; Loris et al., 2010; Simons et al., 2011), or in the space of the imaging kernel (Wu and Yang, 1997; Chevrot and Zhao, 2007). Simons et al. (2011) and Charléty et al. (2013) apply a spherical wavelet transform in model space for global geophysical model representation, analysis, and construction by regularized inversion. Chevrot and Zhao (2007) and Chevrot et al. (2012) introduce a multiscale strategy that effectively compresses sensitivity kernels in global tomography.

Bunks et al. (1995) and Akçelik et al. (2003) prove the effectiveness of multiscale approaches in reducing problems with local minima. Bunks et al. (1995) develop a full-waveform inversion based on multigrid frequency decompositions applied to the 2D Marmousi model (Versteeg, 1993). Without elaborating, they write that scale decomposition by wavelets might be even more effective. Our paper picks up on their suggestion in developing a wavelet-based multiscale approach to waveform inversion. We refrain from making acoustic approximations and use a spectral-element method to solve the elastic wave equation, removing surface waves using tapered dip filtering prior to processing. We implement scale decompositions of the seismogram using wavelets rather than band-passing via the Fourier transform. The discrete wavelet transform provides a non-overlapping scale decomposition and is easily computable for 1D signal traces. In this paper, we apply our wavelet-based multiscale scheme exclusively in the data space. We do not regularize in the model space per se, although a plethora of approaches (e.g., Akçelik et al., 2002; Loris et al., 2010; Charléty et al., 2013) remains to be tried in combination with our methodology.

The spectral-element method of wavefield modeling is implemented on Gauss-Lobatto-Legendre grids (Komatitsch and Tromp, 2002a, 2002b). These grids provide a natural parameterization for all spatial fields, such as the velocity model and all kernels, where of course care must be taken to adequately represent their fine structure (e.g., Zhou et al., 2004; Boschi et al., 2006; Trampert and Spetzler, 2006). Although we do engage in some smoothing of the kernels for numerical stability (e.g., Tape et al., 2009; Zhu et al., 2012) as described below, we do not reparameterize the spatial fields, although various such strategies hold promise (Chevrot and Zhao, 2007; Fichtner et al., 2009; Simons et al., 2011; Chevrot et al., 2012), especially those using 2D and 3D wavelets.

Wavelets come with great flexibility in terms of the selection of the basis class (e.g., orthogonal, biorthogonal, bandlimited or not) or type (e.g., number of vanishing moments, separability), and there is a wealth of choice for the inversion-decomposition parameters (such as the maximum decomposition level, the progression through successive scales, and the number of iterations performed up to each scale). We illustrate the main concepts of our procedure using a toy model first, studying in particular the optimal basis set, decomposition level, and number of iterations per scale. Finally, we demonstrate our approach on a synthetic data set generated in the Marmousi model.

MULTISCALE WAVEFORM TOMOGRAPHY

In this section, we briefly present our method of using wavelet-based multiscale analysis in data space, as applied to waveform-

difference adjoint tomography. We give guidance on how to select suitable wavelets, the maximum decomposition level, and the number of iterations in each scale. We discuss how to implement an adjoint inversion within a single wavelet scale, and we show how to progress from scale to scale, developing an effective reconstruction workflow across the wavelet scales.

Choice of misfit

The flexibility of the adjoint method (Tromp et al., 2005) allows for great freedom in measurement because the Fréchet kernel expressions retain their form except for the necessary changes to the adjoint sources. Another advantage is that the bulk of the computational cost is due to the number of wavefield simulations, which solely depends on the number of sources (and sometimes not even that, e.g., Capdeville et al., 2005; Krebs et al., 2009; Li et al., 2012), and it is unaffected by increasing the numbers or types of measurements made. Constructing Fréchet derivatives for tomographic models usually requires one forward and one adjoint simulation for each source.

Traveltime measurements are widely used in global tomography owing to their relatively good inversion convergence behavior. Perturbations in traveltime are quasilinearly dependent on the relative perturbations in wave speeds, making the optimization easier and faster to converge (Pratt, 1999). Traveltime anomalies can be estimated by crosscorrelation of synthetic and observed arrivals identified using such software as FLEXWIN (Maggi et al., 2009). However, traveltime inversion ignores the effect of velocity anomalies (exactly) on the raypath result in amplitude changes but no time delay (Nolet et al., 2005). Furthermore, arrivals in the seismogram may not be distinguishable and separable, especially when the initial models are smooth. Traveltime inversion only uses one number (the measurement) for each (windowed and filtered) arrival. In contrast, waveform-difference tomography, as we understand it, solves the full elastic wave-propagation problem in heterogeneous media, and it attempts to explain all of the available recorded information; see Figure 1 for a cartoon illustration. If the inversions converge, waveform tomography has the potential to reveal more structural information than do traveltimes.

Waveform-difference adjoint tomography

The *waveform-difference misfit* function $\chi(\mathbf{m})$ computed in a model \mathbf{m} can be expressed as the sum of the squared residuals between the synthetic waveforms $\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and the corresponding observations $\mathbf{d}(\mathbf{x}_r, \mathbf{x}_s, t)$, over all sources s located at \mathbf{x}_s and receivers r at \mathbf{x}_r , over some time window T :

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int_0^T \|\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m}) - \mathbf{d}(\mathbf{x}_r, \mathbf{x}_s, t)\|^2 dt. \quad (1)$$

Gradient methods require the gradient of the misfit function, the Fréchet derivative of $\chi(\mathbf{m})$. We write

$$\delta\chi(\mathbf{m}) = \sum_{s,r} \int_0^T [\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m}) - \mathbf{d}(\mathbf{x}_r, \mathbf{x}_s, t)] \cdot \delta\mathbf{s}(\mathbf{m}) dt, \quad (2)$$

where $\delta\mathbf{s}(\mathbf{m})$ is the displacement-field perturbation due to model perturbations from \mathbf{m} , which can be approximated under the Born theory (Wu and Aki, 1985).

Tromp et al. (2005) show how the misfit perturbation of equation 2 in an elastic earth model can be expressed via the cumulative effect of sensitivity kernels K_ρ and $K_{c_{ijkl}}$ in the model parameter density ρ and elastic constants c_{ijkl} , respectively. With respect to their perturbations $\delta\rho$ and δc_{ijkl} , we rewrite the expressions of Tromp et al. (2005) and Zhu et al. (2009) by collecting the elastic kernels into a tensor \mathbf{K}_c to allow the following concise form:

$$\delta\chi = \int_V [K_\rho(\mathbf{x})\delta\rho(\mathbf{x}) + \mathbf{K}_c(\mathbf{x}) :: \delta\mathbf{c}(\mathbf{x})] d^3\mathbf{x}, \quad (3)$$

where the dependence on the model \mathbf{m} is implicit. The quadruple dot denotes tensor contraction over all four indices $\mathbf{K}_c(\mathbf{x}) :: \delta\mathbf{c}(\mathbf{x}) = K_{c_{ijkl}}(\mathbf{x})\delta c_{ijkl}(\mathbf{x})$ using Einstein's summation convention. The kernels for the density and the elastic-tensor components can be determined from the interaction between the forward (\mathbf{s}) and adjoint (\mathbf{s}^\dagger) wavefields. Again dropping the explicit notational dependence on the model \mathbf{m} , we can write

$$K_\rho(\mathbf{x}) = -\sum_{s,r} \int_0^T \mathbf{s}^\dagger(\mathbf{x}_r, \mathbf{x}, t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, \mathbf{x}_s, t) dt, \quad (4)$$

$$\mathbf{K}_c(\mathbf{x}) = -\sum_{s,r} \int_0^T [\nabla \mathbf{s}^\dagger(\mathbf{x}_r, \mathbf{x}, T-t)] [\nabla \mathbf{s}(\mathbf{x}, \mathbf{x}_s, t)] dt, \quad (5)$$

where $\mathbf{s}^\dagger(\mathbf{x}, \mathbf{x}_r, t)$ is the waveform adjoint wavefield obtained by back-projecting the time-reversed waveform residuals between predicted data and observations at receiver \mathbf{x}_r . The interaction of the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{x}_r, \mathbf{x}, t)$ with the forward wavefield $\mathbf{s}(\mathbf{x}, \mathbf{x}_s, t)$ yields *station sensitivity kernels* for each source-receiver pair $(\mathbf{x}_s, \mathbf{x}_r)$. The *event kernels* (Tape et al., 2007), or gradients of the misfit function corresponding to a single source at \mathbf{x}_s , are defined as the sum of all station sensitivity kernels, which can be efficiently obtained by back-projecting all waveform residual measurements from all receiver stations simultaneously. We refer to the sum of

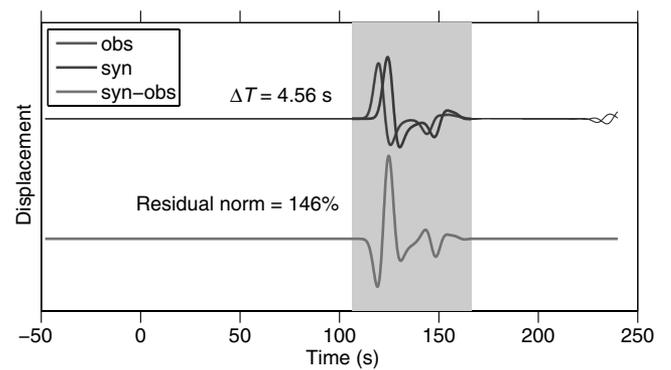


Figure 1. Comparison of crosscorrelation traveltime and waveform-difference measurements between observed (d , top) and predicted (s , top) waveforms. The norm of the waveform residual $s - d$, represented by the bottom line, is expressed relative to the norm of d (in percent, 146% in this example). The traveltime difference ($\Delta T = T_s - T_d = 4.56$ s in this example) is the time shift that maximizes the crosscorrelation between the seismic arrivals contained within the window of interest (shaded gray). Traveltime inversion aims to minimize the time shift, and waveform-difference inversion seeks to flatten the residual waveform by fitting all values within the window.

all event kernels as *misfit kernels* (Tromp et al., 2005), or gradients of the summed misfit over all sources and receivers in equation 1. We can reparametrize equations 4–5 into seismic-velocity and density kernels and express the misfit anomaly in equation 3 as a function of velocity and density perturbations (Tromp et al., 2005).

Multiscale approach based on wavelets

Compared to the limited applicability of traveltimes-based approaches, methods based on direct waveform subtraction would appear to be promising. However, the potential to explain all available information, including phase and amplitude simultaneously, exacerbates the nonlinearity of the inverse problem and causes waveform-difference adjoint tomography to suffer from convergence to secondary minima. The quasilinear dependence of traveltime anomalies on wave speed perturbations suggests that a hybrid method of traveltime fitting followed by amplitude inversion will provide better velocity resolution than explaining traveltime measurements alone (Wang and Houseman, 1995).

However, because long-wavelength measurements are more linearly related to wave speed perturbations, a multiscale approach might be a solution to the convergence problem even for waveform-difference inversions. Starting with low-frequency measurements, we may try to image the long-wavelength structures of the model before proceeding to explaining the higher frequencies in the seismogram, which should be able to map short-wavelength variations in the model with high resolution.

Compared to global representations via Fourier transforms, wavelet transforms admit a description of local features of a signal in the space and frequency domains. Wavelet transforms provide efficient spectral decompositions of the seismogram by extracting nonoverlapping phase-space information at different scales from local stretches of the time series. Another advantage of wavelet transforms lies in the flexibility to choose from a collection of orthogonal or biorthogonal bases to optimally represent the data under consideration. Wavelet transforms are invertible, and successive approximations over discrete ranges of scales do not display the truncation artifacts that typify Fourier-based filtering.

Multiscale waveform-difference inversion involves multiscale analysis of the seismograms $s(t)$, whereby we refer to either synthetic data or observations, which we distinguished in equation 1. We consider two sets of scaling functions ϕ_k^j for the analysis and $\tilde{\phi}_k^j$ for the synthesis as well as two sets of wavelet functions $\tilde{\psi}_k^j$ for analysis and ψ_k^j for synthesis, where j indicates a particular scale and k denotes translation in time (Beylkin, 1992; Daubechies, 1992; Mallat, 2008). In the case of orthogonal transforms, $\tilde{\phi}_k^j = \phi_k^j$ and $\tilde{\psi}_k^j = \psi_k^j$. Using these functions, the input signal $s(t)$ can be expanded in the synthesis (wavelet and scaling) functions using the inner products of the input signal with the analysis (wavelet and scaling) functions. The expansion coefficients

$$a_k^j = \langle s, \tilde{\phi}_k^j \rangle, \quad (6)$$

$$d_k^j = \langle s, \tilde{\psi}_k^j \rangle, \quad (7)$$

are termed *approximation (scaling) coefficients* and *detail (wavelet) coefficients*, respectively. Denoting the maximal scale for the breakdown of the input signal by J , the signal can be represented as the sum over all translates k and scales $j = 1, \dots, J$ as

$$s(t) = \sum_k a_k^j \phi_k^j + \sum_{j=1}^J \sum_k d_k^j \psi_k^j. \quad (8)$$

We denote the partial reconstruction of the signal up to scale j only as $s_j(t)$, which is thus given by

$$s_j(t) = \sum_k a_k^j \phi_k^j + \sum_{j'=j+1}^J \sum_k d_k^{j'} \psi_k^{j'}, \quad (9)$$

$$= \sum_k a_k^j \phi_k^j. \quad (10)$$

We call J the depth of the decomposition. The finest level of information about $s(t)$ is contained in the lowest-order wavelet coefficients d_k^1 . The coarsest level of information is obtained by the disregard of all wavelet coefficients, which leaves $s_j(t)$ as a reconstruction by scaling functions only, using the scaling coefficients a_k^j . In this notation, the original full-resolution seismogram is denoted $s(t) = s_0(t)$. On the dyadic grids typically used for the discrete wavelet transform, the physical scale lengths of d_k^1 , a_k^j , and all levels in between, will depend on the initial sampling rate of the signal and its total length (Strang and Nguyen, 1997; Jensen and la Cour-Harbo, 2001; Simons et al., 2006).

Choice of optimal wavelets

The main issue with the multiscale approach is how to select a collection of wavelets that are “optimally” suitable for the waveform inversion problem. The common ones are the orthogonal Daubechies (1992) wavelet families and the biorthogonal Cohen-Daubechies-Feauveau (CDF) wavelet sets (Cohen et al., 1992). Daubechies wavelets are asymmetric, but the synthesis and analysis transforms share the same filters, whereas CDF wavelets are symmetric with distinct synthesis and analysis functions. To be practical for seismogram decomposition in the context of waveform-difference tomography, three factors should be particularly considered:

- 1) The optimal wavelet basis will depend on the specific data under consideration. As a general rule, we choose wavelets with large numbers of vanishing moments, for smooth signals.
- 2) Wavelet expansions under which seismogram synthetics and observations have a high degree of similarity, measured by the mean-squared error of the residual waveforms, are preferred.
- 3) The computational cost of wavelet analysis and synthesis should be considered, especially for massive data processing. Decomposition using wavelets with larger numbers of vanishing moments is more expensive than with those with fewer vanishing moments, which involves fewer filtering coefficients.

Choice of decomposition parameters

Given optimal wavelets, the next crucial issue for the multiscale approach is the choice of the maximal decomposition scale and the successive reconstruction levels. The “best” decomposition depth should provide a good and easy starting point at which the synthetic seismograms are close to the corresponding observations in the

subspace defined by the wavelet basis at the maximal chosen scale level. Successive reconstructions can be terminated when they resemble the input signals according to a mean-squared misfit convergence criterion.

Choice of optimization method

Given optimal wavelets, maximum decomposition scale, and current analysis depth, observed and predicted seismograms can be wavelet transformed. We apply the spectral-element wavefield simulation and adjoint calculation of gradients combined with a conjugate-gradient algorithm to reduce misfit functions defined in the wavelet subspace. Other optimization methods may be substituted here (Brossier et al., 2009; van Leeuwen and Herrmann, 2013).

Choice of iterations at an individual scale

Misfits involving measurements down to a certain scale are decreased iteratively by following the conjugate-gradient directions. When the waveform residuals between the partially reconstructed synthetics and observations are smaller than a certain tolerance, when the model improvements are no longer obvious, or when the total number of iterations exceeds a certain predefined maximum, the iterations at that scale can be stopped. Switching to the next scale, cumulatively, over time will allow the adjoint procedure to fit more oscillatory waveform information, hence illuminating smaller structures in the model space more clearly.

Summary of the algorithm

Building on our prior discussion, we summarize the basic recipe for multiscale waveform-difference adjoint inversion as follows:

- 1) Based on the synthetics computed in the initial model, and on the observed seismograms, we define the multiresolution parameters, including the wavelet family, the number of vanishing moments, and the maximal decomposition scale J , i.e., the starting level.
- 2) We initialize $j = J$ (if coming from step 1), or we update to $j = j - 1$ (if coming from step 4).
- 3) Iterative minimization within scale level j . We make wavelet-transformed measurements on the observed and synthetic seismograms; we run forward and adjoint wavefield simulations to compute the gradient of the misfit; we update the current model in the search direction weighted by the step length estimated using a line search, as follows:
 - a) We conduct forward simulations in the current model $\mathbf{m}(\mathbf{x})$ with the spectral-element code, to obtain synthetic seismic data $\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ at all stations and for all receivers in the time window of interest.
 - b) We wavelet transform and partially reconstruct data and the synthetics in the current model up to level j to yield the subbands of seismograms $\mathbf{s}_j(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and $\mathbf{d}_j(\mathbf{x}_r, \mathbf{x}_s, t)$ in the notation of equations 1, 9, and 10.
 - c) We choose the specific time windows of length T to make the measurements of interest. Usually, we adopt longer windows for longer wavelength data up to higher scale levels and narrower windows for shorter wavelengths as the scale levels decrease.

- d) We calculate the difference between the windowed and partially reconstructed synthetics $\mathbf{s}_j(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and the observed seismograms $\mathbf{d}_j(\mathbf{x}_r, \mathbf{x}_s, t)$. Our measurements of the waveform-difference misfit become, in the notation of equation 1,

$$\chi_j(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int_0^T \|\mathbf{s}_j(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m}) - \mathbf{d}_j(\mathbf{x}_r, \mathbf{x}_s, t)\|^2 dt. \quad (11)$$

- e) We carry out an adjoint wavefield simulation with the spectral-element code, and we compute the interaction of the adjoint with the forward wavefields to obtain the gradient of the misfit as in equations 3–5, which we now denote \mathbf{g} , evaluated at model \mathbf{m} .
- f) We choose a search direction by conjugate gradients; at the k th iteration, $\mathbf{p}^k = -\mathbf{g}^k + \beta^k \mathbf{p}^{k-1}$, where $\mathbf{p}^1 = -\mathbf{g}^1$ for the first iteration, $k = 1$, and $\beta^k = \mathbf{g}^k \cdot (\mathbf{g}^k - \mathbf{g}^{k-1}) / (\mathbf{g}^{k-1} \cdot \mathbf{g}^{k-1})$ for $k > 1$.
- g) We perform a line search to obtain a scalar step size ν^k to minimize the misfit function χ_j along the direction defined by \mathbf{p}^k in the previous step.
- h) We update the current model in the search direction of step g weighted by the step size estimated in step f to compute the updated model $\mathbf{m}^{k+1} = \mathbf{m}^k + \nu^k \mathbf{p}^k$, and we return to step a.

- 4) The loop in step 3 is repeated until the partially reconstructed seismograms for scale j fail to add more information for misfit improvement, after which we return to step 2 to add shorter wavelength information by switching to the next scale of reconstruction $j - 1$.
- 5) We repeat the entire iteration scheme in steps 2–4 until convergence at the full-resolution scale $j = 0$.

NUMERICAL EXPERIMENTS — I

Geometry of the toy model

Our first test model measures 480×480 km, with 40 mesh nodes uniformly distributed in each dimension; thus, there are 1600 mesh elements in total. In each of those, there are 5×5 Gauss-Lobatto-Legendre integration points, for a total of $[40 \times (5 - 1) + 1]^2 = 25,921$ unique grid points, or $[40 \times 5]^2 = 40,000$ S-wave speeds. As shown in Figure 2, we designed the source-receiver geometry to be reminiscent of a vertical seismic profile exploration survey. We placed 12 virtual shots near the surface and 11 in a source “borehole” on the right of the study area. We aligned 30 virtual receivers near the surface and 30 in a receiver borehole on the left side of the study area. We sample the wavefield every 6×10^{-2} s for 4800 time steps. Sources and receivers were spaced 2 km from the edges to avoid boundary effects in the simulation. The target model for our inversion is shown in Figure 2 (left panel). It has a background S-wave speed $\beta = 3500$ m/s and two circular anomalous areas whose S-wave speeds are 3100 m/s and 3900 m/s, i.e., $\pm 11\%$ perturbations relative to the background velocity. Station spacing $\Delta x = 1.5$ km, and the shot interval $\Delta s = 4$ km, both of which satisfy the wavelength thresholds $\lambda/\Delta x > 2$ and $\lambda/\Delta s > 2$ to guarantee unaliased shot and receiver wavefields (Levander and Nolet,

2005) for wavelengths 300–500 km. The sampling interval for our simulations, $\Delta t = 6 \times 10^{-2}$ s, satisfies the Nyquist-Shannon and the Courant stability criteria (Virieux, 1986; Komatitsch and Tromp, 2002a). We used a first-derivative Gaussian source-time function.

Multiscale seismograms

As illustrated in Figure 3, we use a series of Daubechies (1988) wavelets with six vanishing moments to decompose the observed (thick solid lines) and initial-model synthetic (thin dashed lines) displacement seismograms over eight scales, and we successively reconstruct them from the top level down to 8, 7, and 5 scales (differentiated by color). For each of the observed-synthetic approximation pairs, we also show the waveform difference (thin solid lines) and quote the norm of this residual, expressed as a percentage of the norm of the observations. In the simple toy model of Figure 2, the multiresolution “measurements” shown in Figure 3 are simple, easily interpreted, and quickly made directly from the seismograms after one forward simulation.

The source-receiver pair for which the seismograms are being plotted in Figure 3 is the path highlighted by the line in Figure 2. Because the chosen trajectory samples the anomalously slow region in the target model, the synthetics calculated with a background S-wave speed of 3500 m/s in the featureless initial model arrive well before the observations, which is clearly visible by their relative time shift, especially at the lower scale numbers that represent higher frequency arrivals. The measurements are increasingly oscillatory in function of descending scale number, and their relative residual norm increases. The partial reconstruction up to scale 5 is able to represent almost all of the features of the original data (in black), thereby defining an effective resolution and compression criterion, which is, however, of no further significance to us here. Scale 0 formally captures all of the original signal in its entirety.

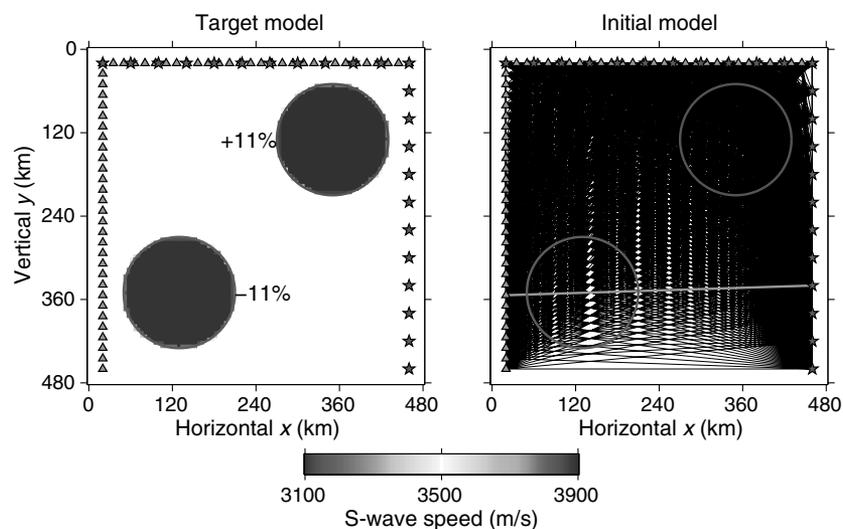


Figure 2. The 2D toy experiment for wavelet-based multiscale waveform-difference inversion. (a) Target S-wave speed model of a 3500 m/s background with two circular $\pm 11\%$ anomalies. The 23 shot points and 60 receivers are marked by stars and triangles, respectively. (b) Ray coverage in the initial model, which is homogeneous with the background wave speed. The outlines of the anomalies remain visible as circles. The line connecting one source-receiver pair will be used later to illustrate our analysis.

No noise was added to our data set. Uncorrelated, additive Gaussian noise would be easy to remove using thresholded wavelet-denoising techniques (e.g., Simons et al., 2009). On the other hand, jointly estimating and fully characterizing structural and signal-generated noise would be so involved that we continue to focus on the performance of the algorithm as applied to noise-free data for the sake of simplicity and brevity.

Multiscale waveform-difference station sensitivity kernels

How do the new wavelet multiscale waveform-difference measurements, introduced in the previous section and illustrated in Figure 3, sense the discrepancy between the target model and its initial, current, or final estimate? As discussed in the previous section, we can address this question by inspection of the station sensitivity kernels of the chosen waveform-difference misfit criterion defined for the corresponding source-station pair, with respect to model perturbations of the current model iterate. Their values reflect the spatial distribution of where S-wave speed perturbations to the model decrease the waveform difference measurements for the particular source-station pair, up to different wavelet-scale levels.

For the source-receiver pair connected by the line in Figure 2, we show the multiscale waveform-difference station sensitivity kernels in Figure 4. These are obtained via the zero-lag crosscorrelation of the forward wavefield from the source and the adjoint wavefield from the receiver, by back-projecting the time-reversed waveform residuals up to different scale levels from the station back to the source. For a fixed relative location of source and receiver, the width of the first Fresnel zone at the midpoint along the geometric raypath is proportional to the square root of the wavelength (Dahlen and Tromp, 1998; Baig et al., 2003). Figure 4 serves as a graphical illustration of this behavior.

The kernel up to wavelet scale 8 has a less oscillatory structure than those of the kernels corresponding to the shorter wavelength, lower scale measurements, shown up to scales 7 and 5. The positive values of the sensitivity kernel up to scale 8 reflect the need to negatively adjust the model in which the synthetics are being calculated to reduce the misfit with respect to the model in which the observations were made. As we know from Figure 2, the raypath shown directly samples the wave speed region that is anomalously slow when compared to the homogeneous initial-model background velocity. The steepest-descent model update is indeed the negative of the misfit kernel (Tape et al., 2007). In contrast, at shorter wavelengths, the complex rippled structures and overall narrower Fresnel zones are less easily related to a simple reliable direction of model improvement. This general behavior goes a long way toward explaining the issues plaguing stability and convergence in high-frequency seismic waveform modeling.

Iterative model evolution and final data fits

Through the multiscale representations of the recorded seismograms and their spectral-element synthetics, we can make waveform-difference

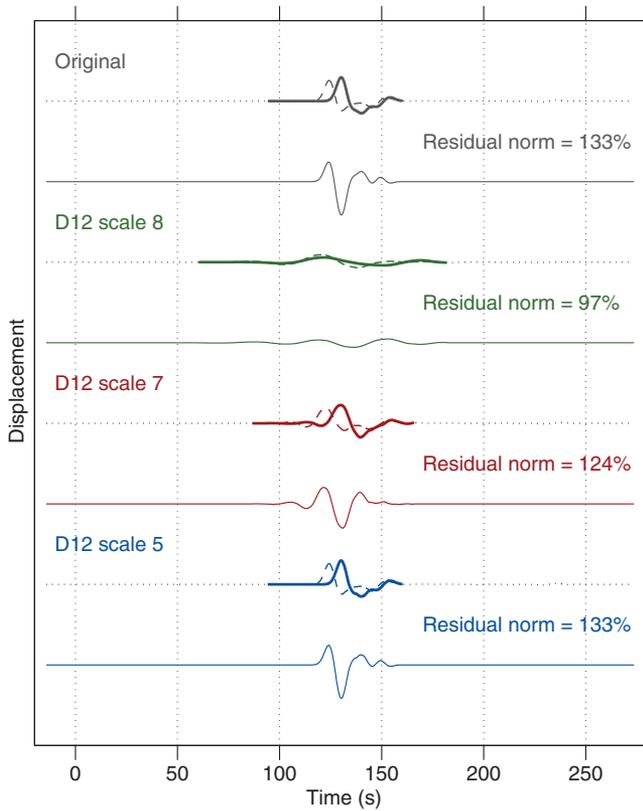


Figure 3. Initial data fits. Wavelet decomposition of observed (thick solid lines) and synthetic (thin dashed lines) seismograms using the target and initial models shown in Figure 2, where the particular source-receiver pair was highlighted. Colored curves represent the partial reconstructions of the original seismograms down to scales 8, 7, and 5, as labeled. The residual waveforms, obtained by subtraction of the observed from the synthetic waveforms, are shown by the thin solid lines, together with their norm, which is expressed as a percentage relative to the norm of the observations down to that scale.

measurements as the model evolves. Within each iteration, we use the appropriate misfit sensitivity kernels for all stations and all shots. The model is improved iteratively by following the conjugate-gradient search directions determined by the current misfit kernel and the previous update direction.

As can be seen in Figures 3 and 4, the seismograms up to the largest scales (for example, scale 8) correspond to long-wavelength (in the time domain) measurements. They correspond to a very blurry sensitivity to long-wavelength wave speed structure (in the space domain). Measurements and kernels made up to subsequent, lower scales gradually increase the level of detail in the data (time) domain and the model (space) domain. We run the adjoint inversion within a certain scale until no further obvious improvement of the solution can be obtained; see Figure 5. After that, we switch to the next, lower scale. At the end of this procedure, we have used all of the available data down to the lowest full-resolution scale (in our example, scale 5 contained virtually all of the information in the seismogram at this sampling rate) and have obtained a high-resolution final result. In our toy experiment, the synthetics generated in the final model match the observations extremely well, from long-wavelength trends down to short-wavelength details, as shown in Figure 6.

From the final model shown in Figure 5, we can see that the high-wave speed (fast) anomaly can be recovered nearly perfectly, but the low-wave speed (slow) area continues to lack some definition. We attribute the relative difference in model resolution first of all to the source-receiver geometry. Approximate connections between ray density, resolution, and a posteriori uncertainty have been formulated for the adjoint-modeling framework by Fichtner and Trampert (2011). As can be seen in Figure 2, the ray coverage is less dense in the bottom half of the model domain, which includes the slow anomaly. A secondary explanation is that slow anomalies are inherently more sensitive to wavefront healing (Williamson and Worthington, 1993; Nolet and Dahlen, 2000; Hung et al., 2001; Dahlen, 2004), which further degrades their potential resolution.

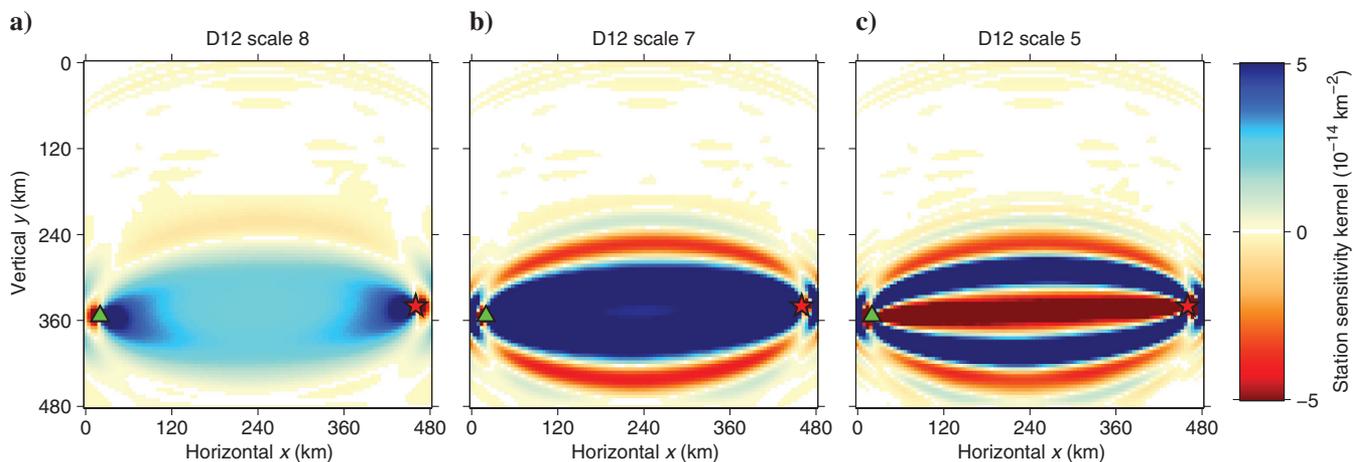


Figure 4. Multiscale station sensitivity kernels corresponding to waveform-difference measurements in Figure 3. The initial velocity model was too fast compared to the true wave speeds. In (a), we see that the coarse-scale, long-wavelength data sense larger scale anomalies. In (b) and (c), fine-scale, short-wavelength data give rise to more complicated rippled structures.

Comparison with classical misfit functions

The success of our new wavelet-multiscale waveform-difference method, which we introduced via the toy model, can be properly appreciated by comparison with more classical measurements and inversion methods that use the same adjoint modeling of the wave equation, but work either with crosscorrelation traveltimes, or with direct or filtered (but not multiscale) waveform-difference measures. To this end, we conducted experiments, for use with the classical methods, similar to the one described above, but with more challeng-

ing (i.e., farther from the target) starting models. Indeed, previous studies (Gauthier et al., 1986; Mora, 1987) showed that if the initial model is far from the target, gradient-based iterative algorithms might terminate into secondary solutions. In this subsection, we show the spectacular failure of waveform-difference inversions conducted at “full resolution” in such a case. We then show how regularization via conventional frequency filtering or, alternatively, using our wavelet-based multiscale-approach produces very well-resolved converged solutions, with the metrics favoring the wavelet method.

For a starting model, we chose a uniformly high wave speed, equal to the largest anomalous velocity in the target model (a +11% perturbation relative to the average velocity); see Figure 7a. All experiments were run to numerical convergence (except in the divergent case shown), with the number of iterations mentioned below.

The use of crosscorrelation traveltimes as measurements continues to result in a reasonable final model, as shown in Figure 7b. We attribute this to the pseudolinear misfit behavior of the measurements being sensitive enough to the great discrepancy in wave speed in this poorly chosen initial velocity model. In contrast, direct waveform-difference inversion of the seismograms, without any kind of filtering or multiscale decomposition, sends the inversion into an incorrect update direction from which it fails to recover. The result is a divergent final model that is unreasonably far from the target; see Figure 7c.

A hybrid approach (not shown), whereby the crosscorrelation traveltime inversion, to recover the overall background and large-scale structures, was followed by direct waveform-difference inversion of the remaining observational discrepancies, succeeded in providing more accurately detailed information on the wave speed anomalies compared to Figure 7b.

Figure 7d shows the results from a multifrequency inversion in the Fourier domain. In the first stage, synthetics and observations were low-pass filtered (Stockwell, 1999) with a cut-off frequency at 0.045 Hz. For the second stage, full-resolution measurements (with an approximate bandwidth of 0.084 Hz) were used in an iterative waveform inversion starting from the model obtained in the first stage. The end result reproduces the true model with great fidelity.

Finally, Figure 7e shows a final model obtained via the wavelet multiscale approach applied to the starting model in Figure 7a. By breaking down to eight wavelet scales the seismograms for synthetics and observations, and starting with the coarsest measurements, the inversion successfully recovers the overall background model. Adding the detail information back to the data, they begin to gradually image the anomalies, blurrily at first. The successive inclusion of all of the remaining structure in the seismograms approaches the target anomalies with good resolution. The resilient

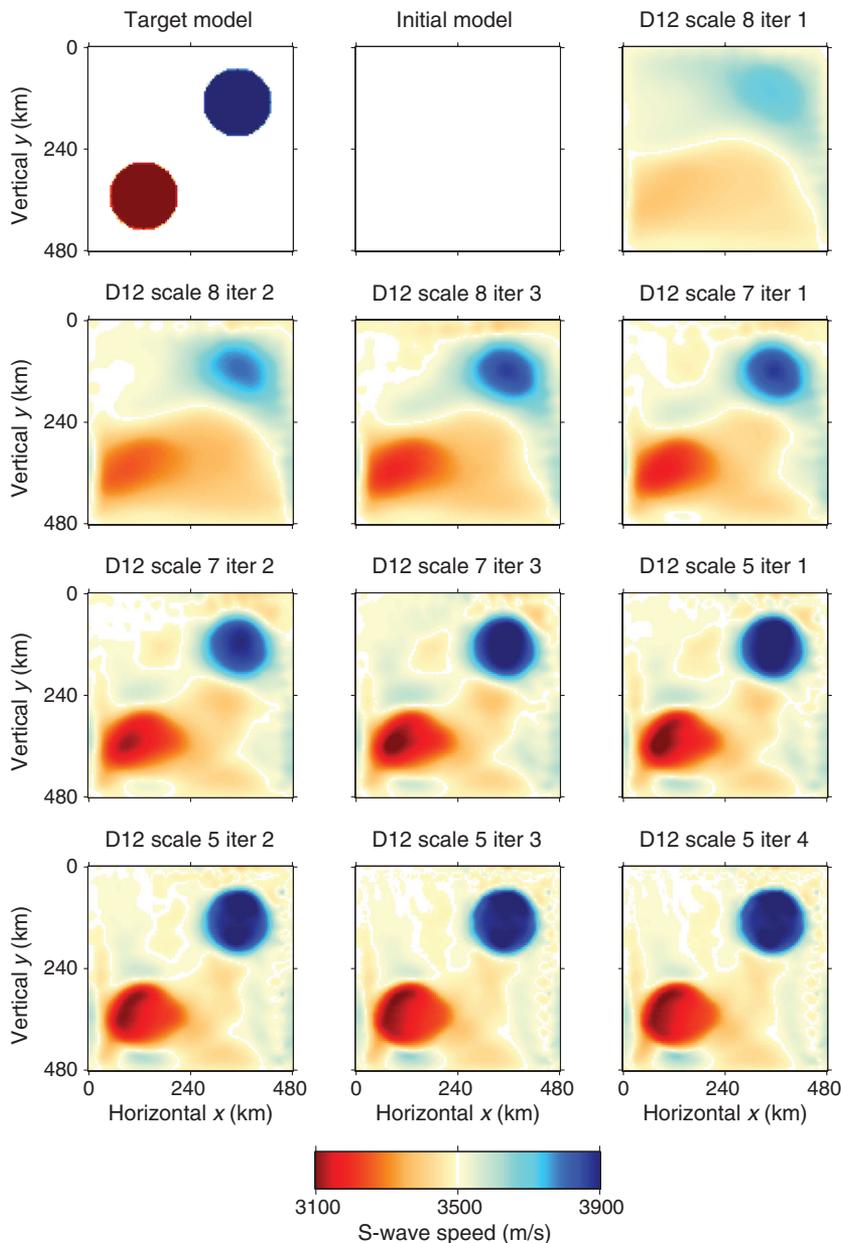


Figure 5. Inversion results for the toy model shown in Figure 2. Using multiscale waveform-difference adjoint modeling with time-domain wavelets, the model evolves through the iterations, as annotated, from the initial homogeneous, average background toward the target. In the final fits, the low-wave speed anomaly is less well recovered due to poorer ray coverage toward the bottom of the model domain. Some final fits to the data are shown in Figure 6.

border artifacts in Figure 7d–7e are due to the location of sources and receivers being near the edges of the model domain, which introduces some spurious structure in the kernels.

Adopting the codes TT for traveltimes shown in Figure 7b, WD for direct-waveform difference shown in Figure 7c, TH for the

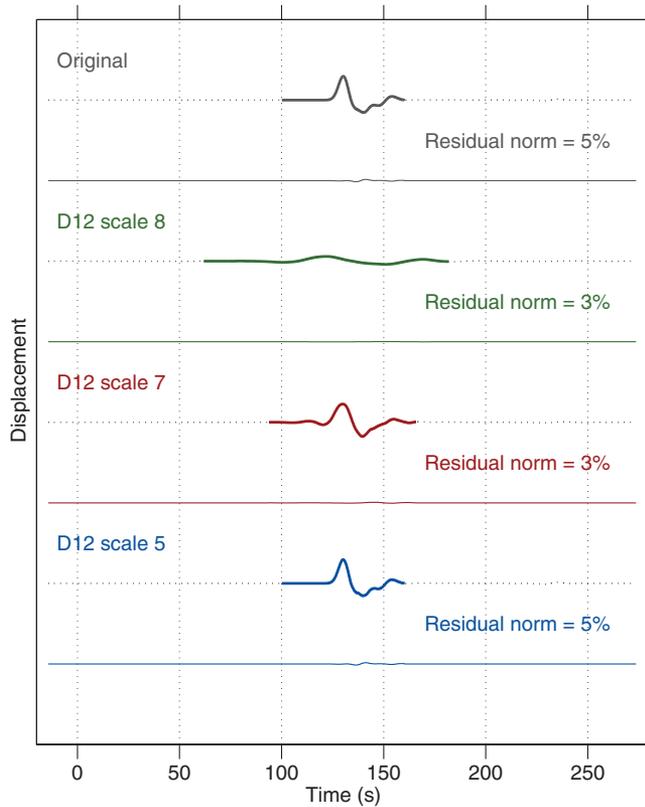


Figure 6. Final fits. Multiscale comparison of the synthetics in the final model shown in Figure 5, with the observations for the same source-receiver pair as used in Figure 3. For ease of comparison, the layout is identical to that of Figure 3.

hybrid method (not shown), MF for the multifrequency approach in Figure 7d, and MS for the wavelet multiscale result shown in Figure 7e, the metrics for the experiments discussed above are as follows: We computed the number of iterations (TT 15, WD 10, TH 30, MF 33, MS 25), the correlation coefficient of the final model with the true model (TT 0.73, WD -0.01 , TH 0.83, MF 0.74, MS 0.85), the root-mean squared (rms) misfit of the model relative to the norm of the true model, in percent (TT 3.4, WD 16, TH 2.7, MF 3.9, MS 2.7), the rms traveltime misfit relative to the initial misfit, in percent (TT 2.1, WD 107, TH 1.3, MF 1.0, MS 1.0), and the rms waveform misfit relative to the initial misfit, in percent (TT 13.5, WD 63.8, TH 7.5, MF 6.6, MS 5.5). Finally, we computed the relative CPU times, in percent (TT 68, WD 47, TH

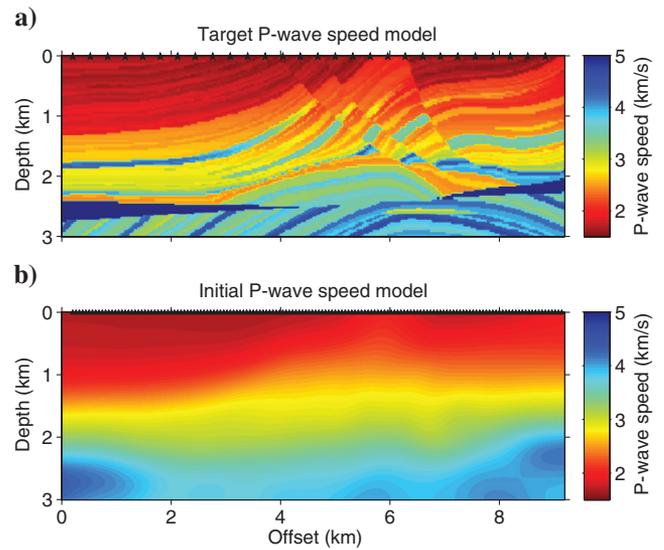


Figure 8. (a) The P-wave speed Marmousi model and (b) its isotropic Gaussian-kernel smoothed version that constitutes our starting model. The corresponding S-wave speed model (not shown) is derived from the compressional model assuming equality of the Lamé parameters as appropriate for a Poisson solid.

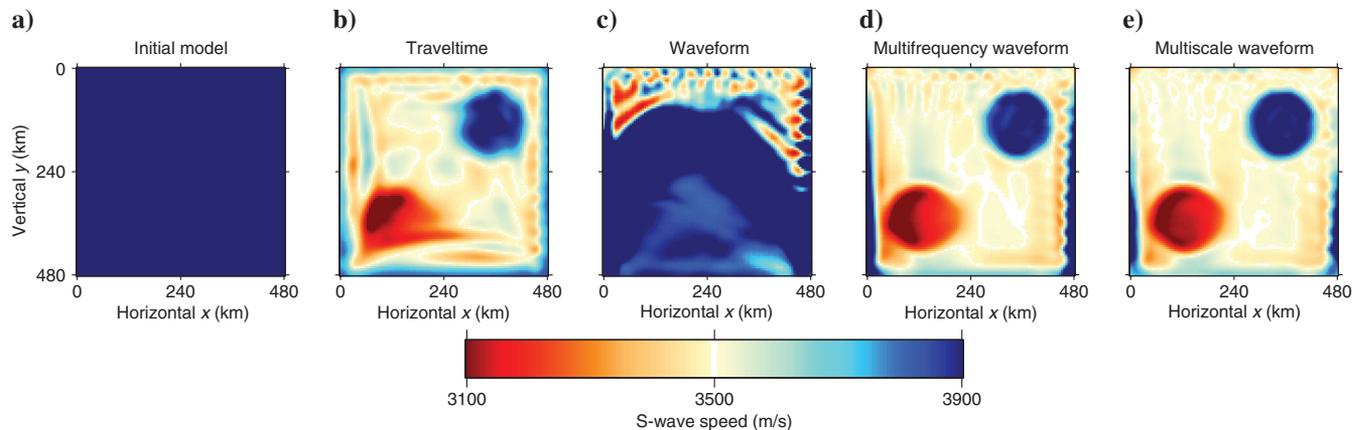


Figure 7. Comparison of inversion approaches for the toy model shown in Figure 2, but from a poor starting model, shown in (a). Final models, obtained using (b) crosscorrelation traveltimes, (c) direct waveform-differences, (d) multifrequency filtered waveform-differences, and finally, in (e) using the new wavelet multiscale waveform-difference measurements of this paper. Traveltime-based adjoint inversions converge to acceptable if smeared versions of the anomalies of interest. Waveform-difference modeling, by itself, succumbs to the severe nonlinearity of the misfit functions. The multifrequency and wavelet multiscale approaches yield well-resolved results.

120, MF 127, MS 100). All metrics favor the wavelet multiscale method (except of course the number of iterations for the divergent WD experiment).

In summary, our initial suite of experiments illustrates the success of our proposed multiscale waveform-difference approach in alleviating the convergence problem of waveform-difference based adjoint modeling techniques. The ability of wavelets to naturally break down and partially reconstruct the data as the iterations progress allows the imaging to proceed quite intuitively from the large-scale structures down to some of the finest details.

NUMERICAL EXPERIMENTS — II

Our experience with the toy model, for a variety of starting models and using a variety of measurement types, gives us confidence to apply the new wavelet-multiscale waveform-difference approach in a realistic setting relevant to seismic exploration problems. In this section, we illustrate the performance of our method as applied to data generated in the Marmousi model (Versteeg, 1993). To simulate fully elastic wavefields, we converted the acoustic model to an elastic one by assuming that the material is a Poisson solid; that is, $\beta = \alpha/\sqrt{3}$. Gauthier et al. (1986), among others, discuss how in gradient-based least-squares waveform inversion, initial models should be reasonably accurate to avoid getting trapped into local minima. Thus, to test the effectiveness of our wavelet multiscale approach, we have wanted to start from a model rather far from the target. To obtain our starting model, we generated a smooth version of the target model by homogenizing the complex geologic structures with an isotropic Gaussian kernel with a standard deviation of 300 m, equivalent to an e -folding width of 850 m. Intrinsic attenuation remains unaccounted for. The resulting models are shown in Figure 8. We applied free-surface conditions at the top of the model and perfectly matching layer absorbing boundary conditions (Festa and Nielsen, 2003) on the remaining three sides of the model domain. All data generated were considered to be noise free.

In the 2D synthetic setting, 112 shots are deployed at a 10 m depth with an 80 m horizontal interval and 361 2C (vertical: Z-component and horizontal: X-component) receivers at 5 m depth with 25 m horizontal spacing. The sampling interval was taken to be 5×10^{-4} s, and the recording duration was set to 6 s. The maximum frequency modeled is 25 Hz, using a Ricker-wavelet (second derivative of the Gaussian) source. To resolve at least five points for the shortest wavelength, we use 120 elements in the horizontal and 40 elements in the vertical direction. As in the toy model, the elements are distributed uniformly. The mesh was not designed to honor known discontinuities (Zhu et al., 2009; Capdeville et al., 2010), nor are the positions of the interfaces model parameters. During the course of the algorithm, we run all simulations on the same numerical grid, despite the fact that we approach the tomographic inverse problem from the coarse to the fine scales in data space, and by extension in the space of the kernels and, ultimately, the model space. Scale-adaptive mesh refinement (e.g., Akçelik et al., 2002) or even wavelet-decomposed operator approaches (e.g., Le Bras et al., 1992; Le Bras and Mellman, 1994; Wu and Yang, 1997) might lead to computational speedup if further upscaling of our methods was to be sought, e.g., for large-volume 3D exploration or global settings.

Surface waves are naturally present with the free-surface boundary conditions via the constructive interference of P-waves and vertically polarized S-waves. Nevertheless, adhering to industry

practice, we performed band-pass filtering with cut-off frequencies at 3 and 30 Hz, and dip-filtering with a cut-off slope at 0.0011 s/m remove to the predictions from the initial model and the synthetic “observations” in the evolving target model, to remove low-frequency and low-velocity surface waves. Our processing scheme was carried out using Seismic Un*x and also incorporated time-domain tapering to circumvent Fourier-domain edge artifacts and antialiasing techniques to remove the distortions caused by spatial and temporal undersampling (Stockwell, 1999). Examples for one shot gather (the 20th shot located 1.8 km from the left model boundary) are shown in Figure 9 for the target model. We duly note that the synthetic seismograms in the initial model are devoid of many of the complex features that would be generated by reflectors, which are absent in the starting model.

Choice of wavelets

Although we have shown before how to use wavelets to perform a multiscale decomposition of the seismograms before delivering waveform-difference measurements made on partially reexpanded observations to the spectral-element based adjoint method, we have not advocated any particular class or type of wavelets. To realize the gains of the multiscale method, the selection of a basis of wavelets suitable to the break down the seismograms in the target and starting model is guided by three criteria.

First, our wavelet basis should be able to naturally represent the oscillatory properties of the seismograms. Wavelets with large numbers of vanishing moments are generally adept at encoding smooth and oscillatory features in the seismograms, whereas wavelets with fewer vanishing moments capture rough and sudden variations (Tae-Kyung and Kennett, 2002). Second, the wavelets are chosen so that the similarity between the synthetic and observed waveforms is initially great enough to result in relatively small residuals in the first iteration. Third, on the grounds of efficiency, we are interested in choosing wavelets that lead to particularly sparse representations of the seismograms.

Figure 10 illustrates the effect of our wavelet choices by plotting the mean-squared error between the seismograms in the initial and target models, at a particular scale as indicated in the legend, and for a variety of orthogonal Daubechies (1988) and biorthogonal Cohen et al. (1992) wavelet types. The orthogonal wavelet bases (D) outperform the biorthogonal ones (CDF) with respect to the misfit metric.

Using a somewhat loose interpretation of the three criteria outlined above, we settled on the orthogonal Daubechies 12-tap wavelets (D12 with six vanishing moments) as the optimal basis functions to conduct multiscale waveform-difference adjoint inversions in the Marmousi model. Using these wavelets for the (partial) reconstructions of the seismograms gives rise to adequate and sparse representations of the data, and it leads to small discrepancies between the initial synthetics and the observations.

Multiscale seismograms

All seismograms were decomposed into nine scales for the successive reconstruction and inversion. Our waveform-difference-based adjoint algorithm first fits the longest-wavelength measurements, at scale 9. We switch to a lower scale as the residuals in the inversion within a certain scale level off. Overall, the agreement between the seismograms in the evolving model and the target

model improves, when measured at the lowest scale taken into consideration. Figure 11a shows the relatively high degree of similarity between the initial (green) and the target (black) traces at the highest scale, for a selection of traces from the twentieth shot gather. After 301 iterations conducted using, ultimately, all of the scales, the modeled seismograms (red) are a close but not a perfect match for the seismograms at this elevated scale. Focusing on scale 8 next, Figure 11b shows the mismatch of the initial and the final models at that scale. The final model is a good but not a perfect match for the observations, in terms of arrival times and amplitudes. As shown in Figure 11c, after 301 iterations, the match between the data and the synthetics in the final model is very good at the lowest, most detailed scale and our objective of fitting the observations as faithfully as possibly is satisfied.

We reiterate that when a smooth initial model is used, most traveltimes-based inversions are not practical because in the synthetics

(green lines in Figure 11c), there are no obvious reflected arrivals to pick and compare with the observations (black lines in Figure 11c). In addition, direct waveform-difference inversion of the original measurements (i.e., up to scale 0 in Figure 11c) is not a good strategy due to our smooth starting model being too far from the target: We would be comparing fairly featureless synthetics with much more oscillatory observations.

Surface waves, though not incorporated in the inversion, are remarkably well matched in the final model, as is apparent from Figure 12.

Multiscale misfit kernels

By themselves, the multiscale misfit kernels paint a complete picture of how the seismic waveform-residuals, for all shots and all

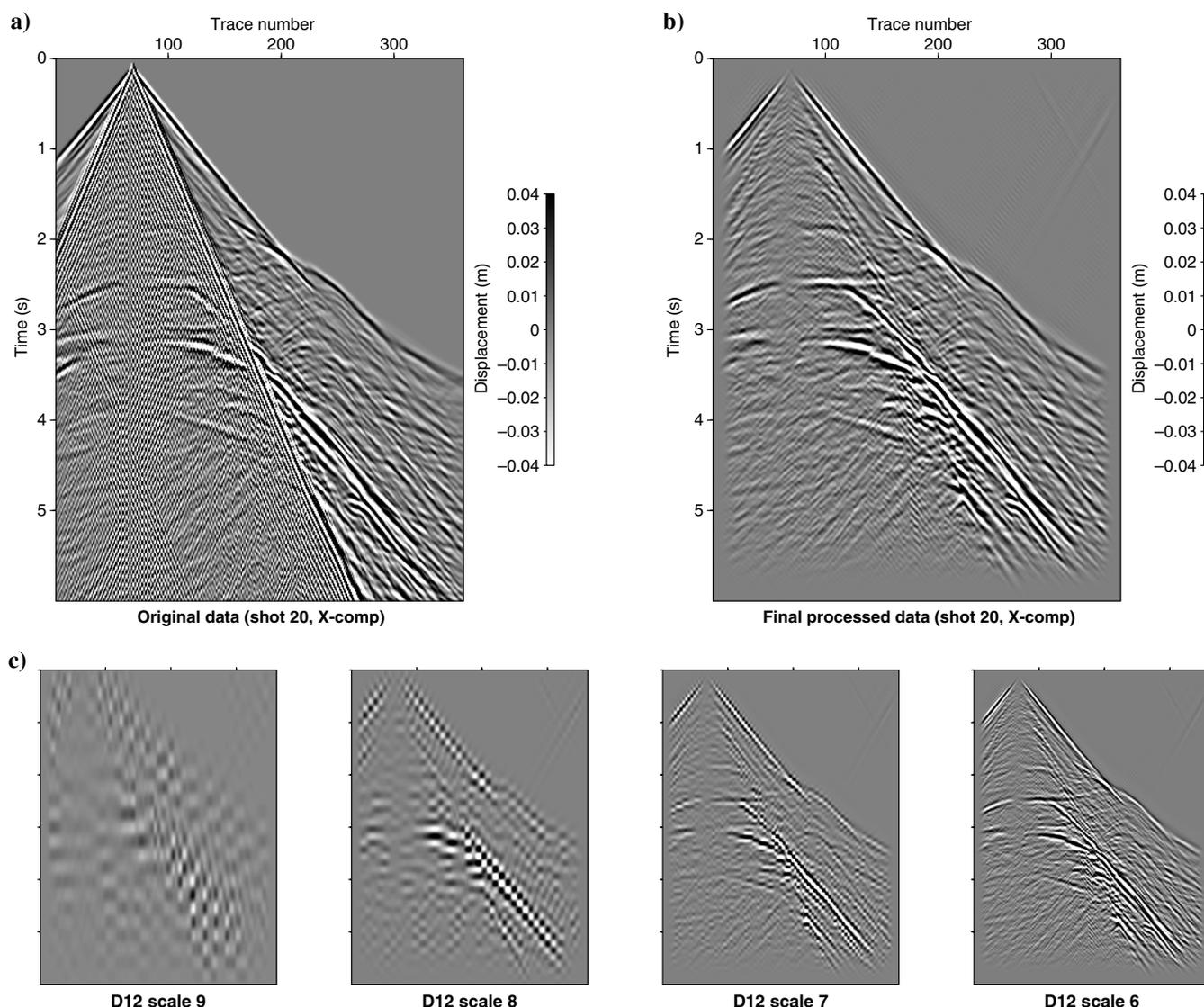


Figure 9. Data processing for the removal of surface waves. (a) X-component displacement seismograms from shot 20. Many of the reflections that could be used to image subsurface structure are severely contaminated by surface waves. After processing by tapering, low-pass and dip-filtering, and antialiasing, undesired surface waves are successfully removed, with only a hint of residual artifacts (b). (c) Subspace representations of the processed shot gather up to different scale levels using Daubechies (1988) D12 wavelets with six vanishing moments.

stations, capture the discrepancy between the current model and the target at each of the wavelet scales in the seismograms. Because 2C seismograms are recorded, the misfit functions make the sum of the squared waveform differences for each component.

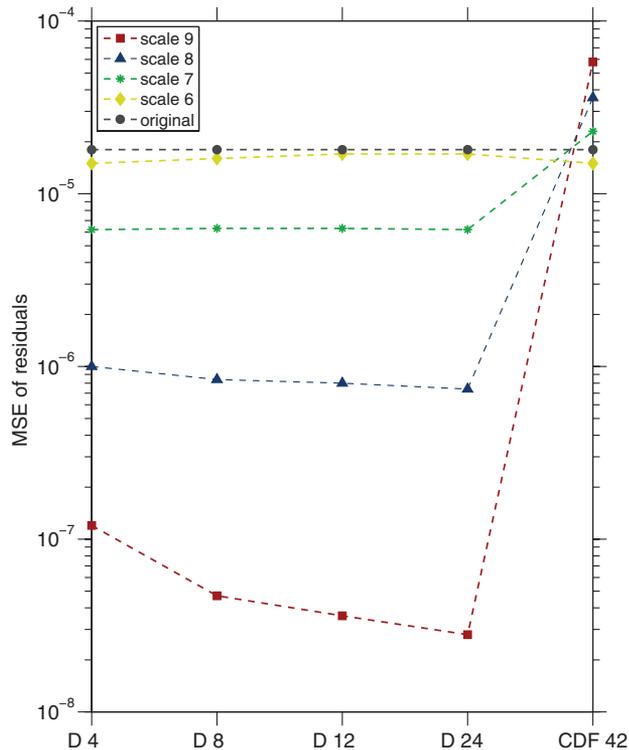


Figure 10. Choice of wavelets. Mean-squared residuals between data in the starting and target models, at different scale levels, for several types of wavelets. We accepted a trade-off between signal representation and computational efficiency in choosing the Daubechies (1988) D12 basis for this class of seismograms.

To suppress short-wavelength artifacts introduced by the numerical computation and spuriously elevated kernel amplitudes in the vicinity of sources and receivers, we applied a Gaussian smoothing operator to the misfit kernels. The scale length of the smoothing depends on the smallest wavelengths that can be resolved. To safeguard the resolution of structural heterogeneities in iterative inversions, more Gaussian smoothing is applied to the misfit kernels for larger scale wavelengths, but smoothing lengths are decreased to embrace finer structures when measurements at smaller scales are added.

At the start of the iterations (iteration 0), and for the largest scale 9, the waveform-difference measurements are sensitive to the largest-scale anomalies. As shown by Figure 13a, the P- and S-wave speed kernels K_α and K_β do not contain much detail of the structure to be imaged, but rather they are sensitive to the wholesale adjustment of the initial model necessary to bring the model closer to reproducing the observed seismograms. The energy in the S-wave kernels is much stronger than that for the P-waves. After 21 iterations carried out at scale 9, the residual kernels are subdued, as shown in Figure 13b and the model improvement has bottomed out for this scale. Figure 13c shows the kernels at the same point in the iteration sequence, for scale 8, which is the point of departure for the next suite of adjoint optimization steps.

A comparison of the initial-model kernels at scales 8 and 7, shown in Figure 14a and 14b, illustrates why the comparatively simpler kernels at scale 9, which were shown in Figure 13a, were a suitable point of departure for the iterations. The more detailed the kernels, the greater the potential for nonlinear effects to steer the inversion in poorly resolved directions. Finally, the full-resolution kernels, shown at iteration 119 and for scale 0, in Figure 14c are the result of having accounted for the residuals down to scale 6 as best we could. The S-wave speed kernel K_β shows that much of the structure has been accounted for, at least in the top half of the model, whereas the P-wave speed kernel K_α reveals the intricate detail that is still necessary to image the short-wavelength variations in the target model. These last kernels are the starting point for the remainder of the iterations, which in our experiment number 301 in total.

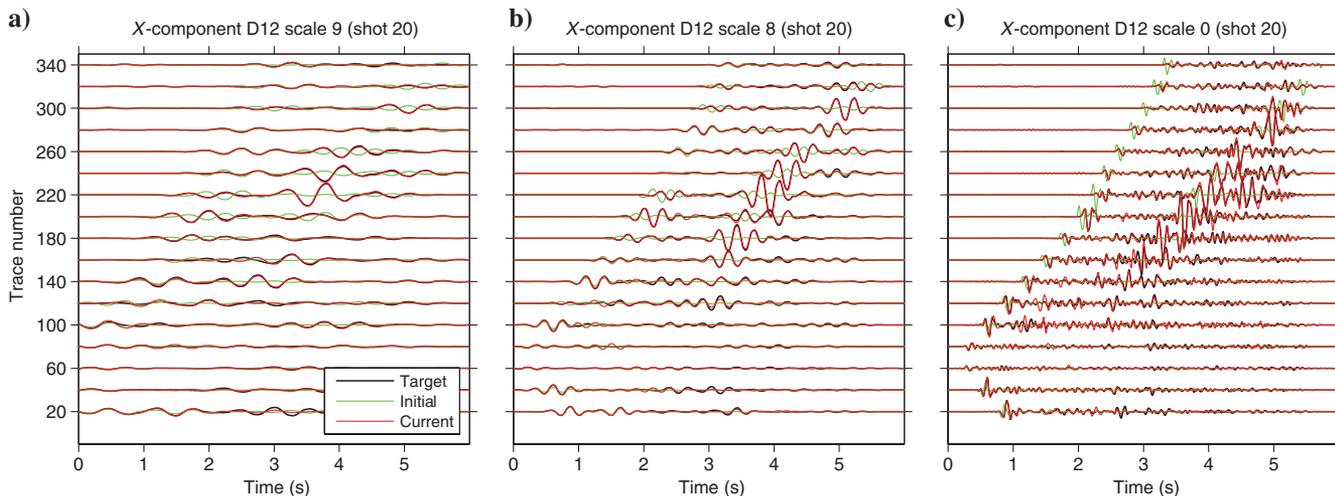


Figure 11. Seismograms (after surface-wave suppression) for the target (black), initial (green), and current final (red) model after 301 iterations using the adjoint waveform-difference method across the wavelet scales. The final model provides a very good match for the full-resolution seismograms at scales 9, 8, and all of the intermediate scales down to scale 0 (full resolution), as shown.

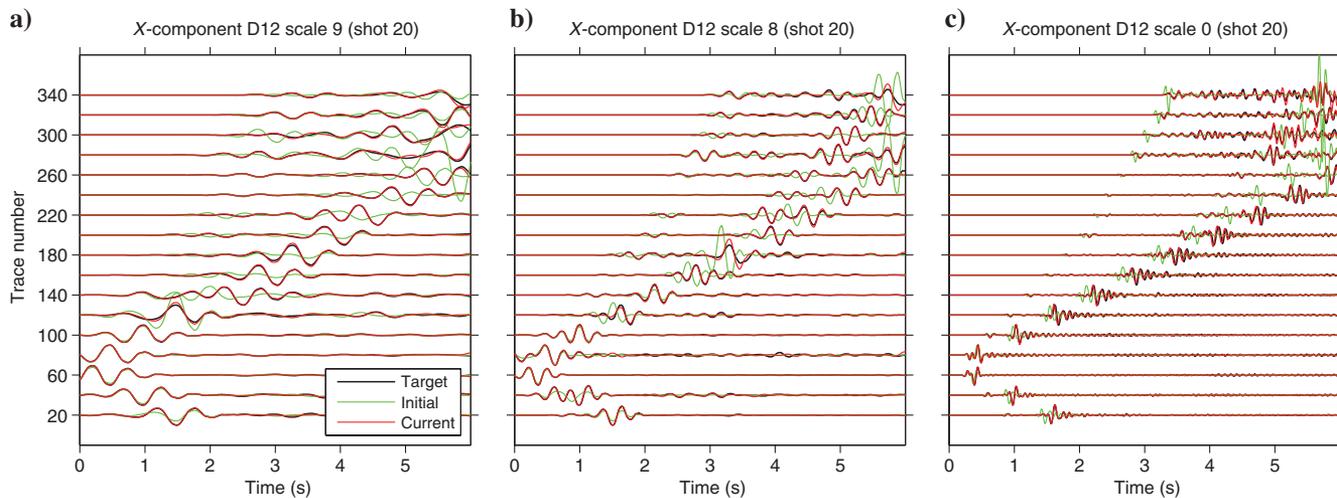


Figure 12. Seismograms (without surface-wave suppression) in an identical layout as in Figure 11. The dominant features in this figure are the surface waves, which, though not the target of the waveform-difference inversion, are also very well matched in the final model, shown after 301 iterations.

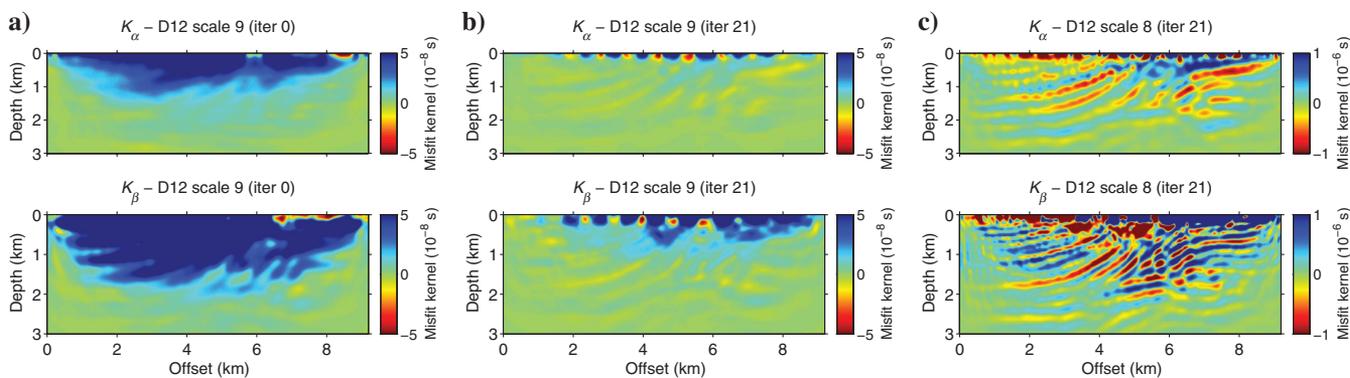


Figure 13. Multiscale misfit kernels for the model updates at different iteration and scale levels. At the start of the inversion, at scale 9, we mainly update larger scale structures near the surface. After 21 iterations at scale 9, most of the structure at that scale has been imaged. At that point, the kernels at scale 8 reveal additional detailed structures remaining to be imaged.

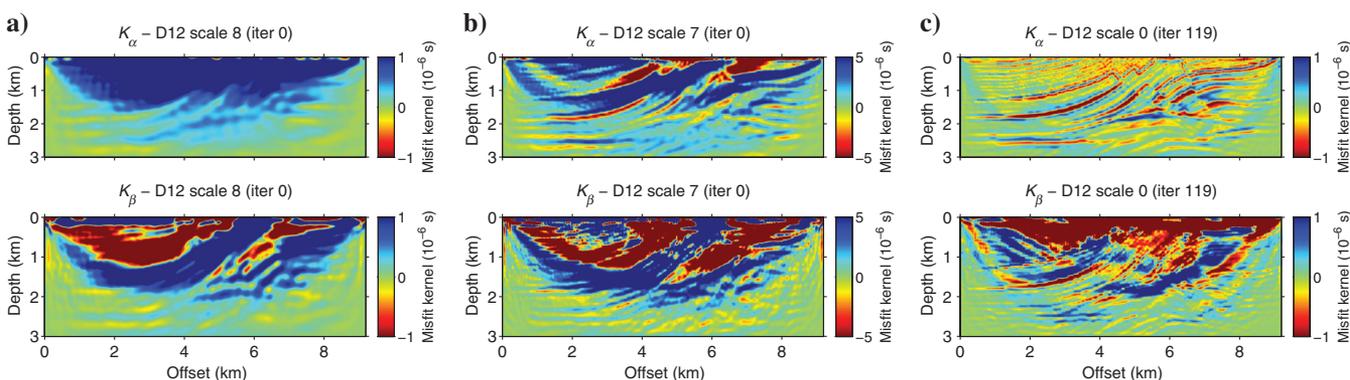


Figure 14. Multiscale misfit kernels at different iteration steps for different scale levels. The kernels at scales 8 and 7 contain much detail that renders the inversion potentially unstable, compared to the initial-model kernels at scale 9 that were shown in Figure 13a. After 119 iterations, the full-resolution (scale 0) kernels reveal the fine-grained structure that is the starting point for several more adjoint modeling steps until the end point, at 301 iterations, for this experiment.

Multiscale misfit evolution

Figure 15 shows the detailed sequence of data residual and model norms reached as the algorithm progresses. There are three parts that enter the evaluation. The first part, shown in Figure 15a, shows the overall rms misfit within the scale levels of the approximation for the iterations, on a logarithmic scale for clarity. The norm is represented relative to the norm of the metric in the starting model, as a dimensionless ratio starting at one at iteration zero for scale 9. This residual norm is decreased by the adjoint modeling until iteration 21, when we switch to reconstructing the seismogram down to scale 8. Once again, the residual norm is relative to the zeroth iteration. As a consequence, it is already much lower than one for iteration 21, but compared to that residual norm at scale 9, the inclusion of extra detail in the seismogram leads to a higher residual at scale 8 at the start of the iterations that also consider this next lower scale. We switch to including scales 7 and 6 after a total of 45 and 71 iterations. After scale 6, we cut directly to full resolution by involving all scale levels because our sampling rate is much higher than needed to fully represent the signals, and scales smaller than five contain almost no information. The point at 119 iterations is the start for a final suite of optimization steps until at iteration 301, when our experiment is stopped.

The black line in Figure 15b shows the evolution of the rms misfit for all scales without surface waves, but now the normalization is such that the normalized point of departure is the residual of the initial model for the full-resolution seismograms after removing surface waves. As we deduce from Figure 15a, the residual norms normalized per scale tend to flatten out with diminishing returns — our cue to bring in additional, lower wavelet scales, where they jump up as new information in the seismogram is brought into the inversion. On the other hand, the residuals normalized to the initial norm of the full-resolution seismograms without considering surface waves, for all scales down to 0, show a mostly monotonic decrease. Evidently, the progressive inclusion of increasingly detailed structure in the seismograms leads to an overall decline of the entire misfit, with downward steps every time an additional

scale is added into the mix. This is the desired behavior, by which minimization within a certain scale tends to efficiently guide the algorithm to reduce the overall data misfit over the course of the iterations. The red line in Figure 15b shows the residual norms of the full-resolution seismograms, now including surface waves, relative to the zeroth iteration residual norm of all scales containing surface waves. Because surface waves were not considered in our inversions, the behavior of this curve is much more erratic throughout the iterations. Ultimately, though, the full-resolution seismograms including surface waves define a misfit threshold that has dropped to a very low level. In the final model, surface waves are also well matched, even though they were never the explicit target of our optimization procedure.

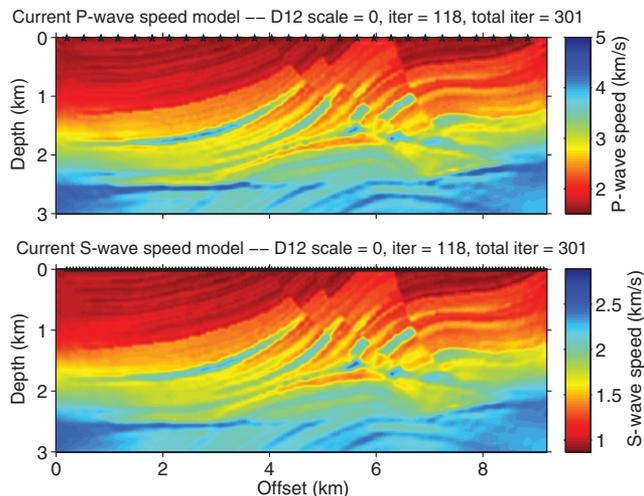


Figure 16. Final V_p and V_s models obtained after total 301 multiscale waveform-difference adjoint inversion steps, using all the available multiscale information in the seismograms. Compare with Figure 8.

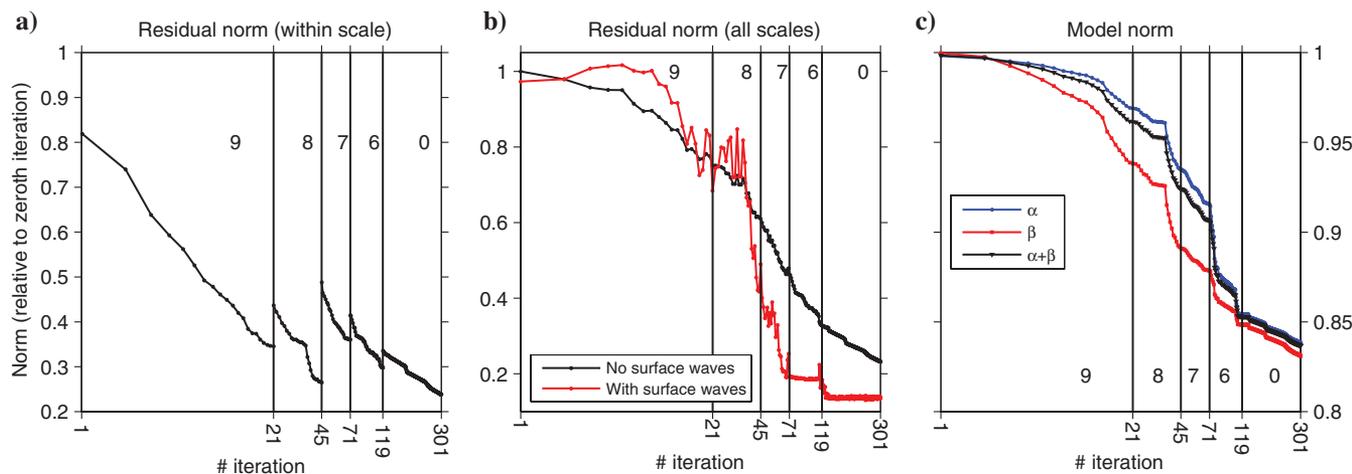


Figure 15. Residual norm evolution as a function of the running iteration number, with the scales within which the adjoint optimization is being conducted noted. At the marked points, additional (lower) scales of the seismograms are being fed to the algorithm. The norms are shown in the data space, normalized relative to the initial-model residual norm, (a) within the scale of the approximation, (b) considering all scales of the full-resolution seismograms with and without surface waves, and (c) in the model space, separately for the P- and S-wave speed portions of the Marmousi model and their combination. The abscissas are logarithmic to facilitate inspection of the convergence behavior.

From the luxury of having access to the “true” target Marmousi model, we can also compute a model-space norm. Its evolution is shown in Figure 15c, separately for the P-wave speed ($\alpha = V_p$), the S-wave speed ($\beta = V_s$), and their combination.

Final models

After reaching a final number of 301 iterations in the multiscale waveform-difference adjoint modeling (Figure 16), the upper part of the Marmousi model has been very well recovered. The lower part continues to suffer from lower resolution, likely resulting from insufficient ray coverage in our experiment as could be clearly seen also in the misfit kernels. Overall, the S-wave speed model is imaged earlier in the procedure, starting with scale 9. This is due to the larger amplitudes of the shear waveforms, and this effect was already apparent from the relative energies of the misfit kernels shown in Figure 13a, where the S-wave speed kernels K_β showed a clear dominance. After most of the shear waveforms are fitted, the compressional model starts to improve, which can be verified from the model norm evolution at scale 6 in Figure 15c.

Although there remains room for model improvement, the gradual conditioning of the adjoint inversion problem, using only differences of the waveforms as measurements, has reached a point where the full-resolution waveform data can be used for additional inversion steps using the standard adjoint waveform-difference modeling, and these should converge to a final model that is only limited by the available data coverage.

CONCLUSIONS

Waveform tomography is an intrinsically nonlinear procedure, in which the complexity of the multiparameter misfit surface dictates the ease with which any inversion method can navigate it to a, hopefully, global minimum. Various strategies have been proposed to mitigate the difficulties with full-resolution waveform inversion, whether operating in the data space, on the compression or complexity reduction of the sensitivity kernels, or in the model domain. Of course, numerous factors contribute to the success of tomographic modeling, broadly speaking. The influence of structural or signal-generated noise, the choice of misfit functionals, the meshing for the wavefield calculations, the parameterization of the model domain, the chosen form of any additional model regularization, the optimization algorithm, and real-world issues with actual data collected in the field all play a role. By ignoring all these subtleties, in this paper we proposed and tested a data-space procedure that uses multiscale constructive approximations of the seismograms and sequentially feeds them to a conjugate-gradient-based inverse modeling procedure using a fully elastic spectral-element formalism for the computation of the forward and adjoint wavefields. More sophisticated optimization algorithms, e.g., preconditioned conjugate-gradient or quasi-Newton methods, can probably be taken advantage of to speed up convergence rate even further.

From our numerical tests, we conclude that multiscale wavelet decomposition of the seismograms helps solve the convergence problem of waveform-difference tomography. Starting from a coarse representation of the seismograms, we make waveform-difference measurements and apply the usual procedure of adjoint tomography to improve the initial model for several iterations, until data fitting at that scale level no longer improves. Subsequently, we add finer scale information back to the seismograms and proceed

with more iterations. We repeat the procedure until the data are well matched by the final model. Using wavelets to implement a truly multiscale analysis puts the usual practice of incorporating shorter period information as the iterations proceed on a more solid and objective footing, improves the overall ability of adjoint waveform-difference tomography, and yields good final models — even when the initial model is far from the target.

Surface waves are generally regarded as interference in most exploration-scale seismic inversions. Although never explicitly targeting them, but in fact removing them from the seismograms throughout the inversion, our obtained models nevertheless match observed surface waves extremely well. In further work, we will attempt to incorporate surface waves back into the inversion, fully embracing their power to more robustly constrain shallow subsurface structure and possibly speed up convergence rates of wavelet-based multiscale full-waveform tomography.

ACKNOWLEDGMENTS

We thank the United States National Science Foundation, which supported this work under grants EAR-1150145 to F. J. Simons and DMS-1025418 to I. C. Daubechies. We thank the “When Dense Meets Sparse” collaborative at Princeton University, the Free University of Brussels (I. Loris), the University of Nice-Sophia Antipolis (G. Nolet), and Duke University (I. Daubechies) for a pleasant and open exchange of knowledge, ideas, and computer code over the years and J. Tromp for comments on a preliminary version of the manuscript. Y. Yuan thanks E. Bozdağ, R. Modrak, Y. Luo, and H. Zhu for advice on running the SPEC-FEM adjoint modeling package, M. Lefebvre and H. Rusmanugroho for aid in processing and computation, and J. Tromp for making high-performance computing facilities available through the Princeton Institute for Computational Science and Engineering (PICSciE). Reviewers A. Fichtner and C. Tape and the associate editor I. Vasconcelos are sincerely thanked for their enormously helpful and detailed comments on the submitted manuscript, which greatly improved the final paper on many levels.

REFERENCES

- Akçelik, V., J. Bielak, G. Biros, I. Epanomeritakis, A. Fernández, O. Ghattas, E. J. Kim, J. López, D. O’Hallaron, T. Tu, and J. Urbanic, 2003, High resolution forward and inverse earthquake modeling on terascale computers: Proceedings of the ACM/IEEE Conference on Supercomputing, 52.
- Akçelik, V., G. Biros, and O. Ghattas, 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: Proceedings of the ACM/IEEE Conference on Supercomputing, 41.
- Alkhalifah, T., and Y. Choi, 2012, Taming waveform inversion non-linearity through phase unwrapping of the model and objective functions: *Geophysical Journal International*, **191**, 1171–1178, doi: [10.1111/j.1365-246X.2012.05699.x](https://doi.org/10.1111/j.1365-246X.2012.05699.x).
- Baig, A. M., F. A. Dahlen, and S.-H. Hung, 2003, Traveltimes of waves in three-dimensional random media: *Geophysical Journal International*, **153**, 467–482, doi: [10.1046/j.1365-246X.2003.01905.x](https://doi.org/10.1046/j.1365-246X.2003.01905.x).
- Beylkin, G., 1992, On the representation of operators in bases of compactly supported wavelets: *SIAM Journal on Numerical Analysis*, **6**, 1715–1740.
- Boschi, L., T. W. Becker, G. Soldati, and A. M. Dziewoński, 2006, On the relevance of Born theory in global seismic tomography: *Geophysical Research Letters*, **33**, L06302, doi: [10.1029/2005GL025063](https://doi.org/10.1029/2005GL025063).
- Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: *Geophysical Journal International*, **185**, 845–870, doi: [10.1111/j.1365-246X.2011.04970.x](https://doi.org/10.1111/j.1365-246X.2011.04970.x).

- 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).
- Capdeville, Y., L. Guillot, and J.-J. Marigo, 2010, 2-D non-periodic homogenization to upscale elastic media for P-SV waves: *Geophysical Journal International*, **182**, 903–922, doi: [10.1111/j.1365-246X.2010.04636.x](https://doi.org/10.1111/j.1365-246X.2010.04636.x).
- Capdeville, Y., Y. Gung, and B. Romanowicz, 2005, Towards global earth tomography using the spectral element method: A technique based on source stacking: *Geophysical Journal International*, **162**, 541–554, doi: [10.1111/j.1365-246X.2005.02689.x](https://doi.org/10.1111/j.1365-246X.2005.02689.x).
- Červený, V., 2001, *Seismic ray theory*: Cambridge University Press.
- Charléty, J., S. Voronin, G. Nolet, I. Loris, F. J. Simons, K. Sigloch, and I. C. Daubechies, 2013, Global seismic tomography with sparsity constraints: Comparison with smoothing and damping regularization: *Journal of Geophysical Research*, **118**, 1–13, doi: [10.1029/2012JD018174](https://doi.org/10.1029/2012JD018174).
- Chevrot, S., R. Martin, and D. Komatitsch, 2012, Optimized discrete wavelet transforms in the cubed sphere with the lifting scheme — Implications for global finite-frequency tomography: *Geophysical Journal International*, **191**, 1391–1402, doi: [10.1111/j.1365-246X.2012.05686.x](https://doi.org/10.1111/j.1365-246X.2012.05686.x).
- Chevrot, S., and L. Zhao, 2007, Multiscale finite-frequency Rayleigh wave tomography of the Kaapvaal craton: *Geophysical Journal International*, **169**, 201–215, doi: [10.1111/j.1365-246X.2006.03289.x](https://doi.org/10.1111/j.1365-246X.2006.03289.x).
- Chiao, L.-Y., and B.-Y. Kuo, 2001, Multiscale seismic tomography: *Geophysical Journal International*, **145**, 517–527, doi: [10.1046/j.0956-540X.2001.01403.x](https://doi.org/10.1046/j.0956-540X.2001.01403.x).
- Cohen, A., I. Daubechies, and J. Feauveau, 1992, Biorthogonal bases of compactly supported wavelets: *Communications on Pure and Applied Mathematics*, **45**, 485–560, doi: [10.1002/cpa.3160450502](https://doi.org/10.1002/cpa.3160450502).
- Dahlen, F. A., 2004, Resolution limit of traveltimes tomography: *Geophysical Journal International*, **157**, 315–331, doi: [10.1111/j.1365-246X.2004.02214.x](https://doi.org/10.1111/j.1365-246X.2004.02214.x).
- Dahlen, F. A., and A. Baig, 2002, Fréchet kernels for body-wave amplitudes: *Geophysical Journal International*, **150**, 440–466, doi: [10.1046/j.1365-246X.2002.01718.x](https://doi.org/10.1046/j.1365-246X.2002.01718.x).
- Dahlen, F. A., S.-H. Hung, and G. Nolet, 2000, Fréchet kernels for finite-frequency traveltimes — I: Theory: *Geophysical Journal International*, **141**, 157–174, doi: [10.1046/j.1365-246X.2000.00070.x](https://doi.org/10.1046/j.1365-246X.2000.00070.x).
- Dahlen, F. A., and J. Tromp, 1998, *Theoretical global seismology*: Princeton University Press.
- Daubechies, I., 1988, Orthonormal bases of compactly supported wavelets: *Communications on Pure and Applied Mathematics*, **41**, 909–996, doi: [10.1002/cpa.3160410705](https://doi.org/10.1002/cpa.3160410705).
- Daubechies, I., 1992, Ten lectures on wavelets: SIAM, **61**, of CBMS-NSF Regional Conference Series in Applied Mathematics.
- Festa, G., and S. Nielsen, 2003, PML absorbing boundaries: *Bulletin of the Seismological Society of America*, **93**, 891–903, doi: [10.1785/0120020098](https://doi.org/10.1785/0120020098).
- Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge, 2008, Theoretical background for continental and global-scale full-waveform inversion in the time-frequency domain: *Geophysical Journal International*, **175**, 665–685, doi: [10.1111/j.1365-246X.2008.03923.x](https://doi.org/10.1111/j.1365-246X.2008.03923.x).
- Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. 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/j.1365-246X.2009.04368.x](https://doi.org/10.1111/j.1365-246X.2009.04368.x).
- Fichtner, A., and J. Trampert, 2011, Resolution analysis in full waveform inversion: *Geophysical Journal International*, **187**, 1604–1624, doi: [10.1111/j.1365-246X.2011.05218.x](https://doi.org/10.1111/j.1365-246X.2011.05218.x).
- Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic wave-forms: Numerical results: *Geophysics*, **51**, 1387–1403, doi: [10.1190/1.1442188](https://doi.org/10.1190/1.1442188).
- Hung, S.-H., W.-P. Chen, L.-Y. Chiao, and T.-L. Tseng, 2010, First multi-scale, finite-frequency tomography illuminates 3D anatomy of the Tibetan Plateau: *Geophysical Research Letters*, **37**, L06304, doi: [10.1029/2009GL041875](https://doi.org/10.1029/2009GL041875).
- Hung, S.-H., F. A. Dahlen, and G. Nolet, 2000, Fréchet kernels for finite-frequency traveltimes — II: Examples: *Geophysical Journal International*, **141**, 175–203, doi: [10.1046/j.1365-246X.2000.00072.x](https://doi.org/10.1046/j.1365-246X.2000.00072.x).
- Hung, S.-H., F. A. Dahlen, and G. Nolet, 2001, Wavefront healing: A banana-doughnut perspective: *Geophysical Journal International*, **146**, 289–312, doi: [10.1046/j.1365-246X.2001.01466.x](https://doi.org/10.1046/j.1365-246X.2001.01466.x).
- Jensen, A., and A. la Cour-Harbo, 2001, *Ripples in mathematics*: Springer.
- Komatitsch, D., J. Ritsema, and J. Tromp, 2002, The spectral-element method, Beowulf computing, and global seismology: *Science*, **298**, 1737–1742, doi: [10.1126/science.1076024](https://doi.org/10.1126/science.1076024).
- Komatitsch, D., and J. Tromp, 2002a, Spectral-element simulations of global seismic wave propagation — I: Validation: *Geophysical Journal International*, **149**, 390–412, doi: [10.1046/j.1365-246X.2002.01653.x](https://doi.org/10.1046/j.1365-246X.2002.01653.x).
- Komatitsch, D., and J. Tromp, 2002b, Spectral-element simulations of global seismic wave propagation — II: Three-dimensional models, oceans, rotation and self-gravitation: *Geophysical Journal International*, **150**, 303–318, doi: [10.1046/j.1365-246X.2002.01716.x](https://doi.org/10.1046/j.1365-246X.2002.01716.x).
- Komatitsch, D., S. Tsuboi, and J. Tromp, 2005, The spectral-element method in seismology, in A. Levander, and G. Nolet, eds., *Seismic earth: Array analysis of broadband seismograms*: AGU, **157**, of *Geophysical Monograph*, 205–227.
- Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: *Geophysics*, **74**, no. 6, WCC177–WCC188, doi: [10.1190/1.3230502](https://doi.org/10.1190/1.3230502).
- Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations, in J. B. Bednar, R. Redner, E. Robinson, and A. Weglein, eds., *Conference on inverse scattering: Theory and application*: SIAM, 206–220.
- Le Bras, R., and G. Mellman, 1994, Wavelet transforms and downward continuation, in J. S. Byrnes, J. L. Byrnes, K. A. Hargreaves, and K. Berry, eds., *Wavelets and their applications*: Kluwer, **442**, of NATO ASI Series C: Mathematical and Physical Sciences, 291–296.
- Le Bras, R., G. Mellman, and M. Peters, 1992, Wavelet transform method for downward continuation: 62nd Annual International Meeting, SEG, Expanded Abstracts, 889–892.
- Levander, A., and G. Nolet, 2005, Perspectives on array seismology and USArray, in A. Levander, and G. Nolet, eds., *Seismic earth: Array analysis of broadband seismograms*: AGU, **157**, of *Geophysical Monograph*, 1–6.
- Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: *Geophysics*, **77**, no. 3, A13–A17, doi: [10.1190/geo2011-0410.1](https://doi.org/10.1190/geo2011-0410.1).
- Loris, I., H. Douma, G. Nolet, I. Daubechies, and C. Regone, 2010, Non-linear regularization techniques for seismic tomography: *Journal of Computational Physics*, **229**, 890–905, doi: [10.1016/j.jcp.2009.10.020](https://doi.org/10.1016/j.jcp.2009.10.020).
- Loris, I., G. Nolet, I. Daubechies, and F. A. Dahlen, 2007, Tomographic inversion using ℓ_1 -norm regularization of wavelet coefficients: *Geophysical Journal International*, **170**, 359–370, doi: [10.1111/j.1365-246X.2007.03409.x](https://doi.org/10.1111/j.1365-246X.2007.03409.x).
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltimes inversion: *Geophysics*, **56**, 645–653, doi: [10.1190/1.1443081](https://doi.org/10.1190/1.1443081).
- Maggi, A., C. Tape, M. Chen, D. Chao, and J. Tromp, 2009, An automated time-window selection algorithm for seismic tomography: *Geophysical Journal International*, **178**, 257–281, doi: [10.1111/j.1365-246X.2009.04099.x](https://doi.org/10.1111/j.1365-246X.2009.04099.x).
- Mallat, S., 2008, *A wavelet tour of signal processing, the sparse way*, 3rd ed.: Academic Press.
- Marquering, H., F. A. Dahlen, and G. Nolet, 1999, Three-dimensional sensitivity kernels for finite-frequency travel times: The banana-doughnut paradox: *Geophysical Journal International*, **137**, 805–815, doi: [10.1046/j.1365-246X.1999.00837.x](https://doi.org/10.1046/j.1365-246X.1999.00837.x).
- Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung, 2004, Global P and PP traveltimes tomography: Rays versus waves: *Geophysical Journal International*, **158**, 637–654, doi: [10.1111/j.1365-246X.2004.02346.x](https://doi.org/10.1111/j.1365-246X.2004.02346.x).
- Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: *Geophysics*, **52**, 1211–1228, doi: [10.1190/1.1442384](https://doi.org/10.1190/1.1442384).
- Mora, P., 1988, Elastic wave-field inversion of reflection and transmission data: *Geophysics*, **53**, 750–759, doi: [10.1190/1.1442510](https://doi.org/10.1190/1.1442510).
- Nissen-Meyer, T., F. A. Dahlen, and A. Fournier, 2007a, Spherical-earth Fréchet sensitivity kernels: *Geophysical Journal International*, **168**, 1051–1066, doi: [10.1111/j.1365-246X.2006.03123.x](https://doi.org/10.1111/j.1365-246X.2006.03123.x).
- Nissen-Meyer, T., A. Fournier, and F. A. Dahlen, 2007b, A two-dimensional spectral-element method for computing spherical-earth seismograms — I: Moment-tensor source: *Geophysical Journal International*, **168**, 1067–1092, doi: [10.1111/j.1365-246X.2006.03121.x](https://doi.org/10.1111/j.1365-246X.2006.03121.x).
- Nissen-Meyer, T., A. Fournier, and F. A. Dahlen, 2008, A two-dimensional spectral-element method for computing spherical-earth seismograms — II: Waves in solid-fluid media: *Geophysical Journal International*, **174**, 873–888, doi: [10.1111/j.1365-246X.2008.03813.x](https://doi.org/10.1111/j.1365-246X.2008.03813.x).
- Nolet, G., 2008, *A breviary for seismic tomography*: Cambridge University Press.
- Nolet, G., and F. A. Dahlen, 2000, Wave front healing and the evolution of seismic delay times: *Journal of Geophysical Research*, **105**, 19043–19054, doi: [10.1029/2000JB900161](https://doi.org/10.1029/2000JB900161).
- Nolet, G., F. A. Dahlen, and R. Montelli, 2005, Traveltimes and amplitudes of seismic waves: A reassessment, in A. Levander, and G. Nolet, eds., *Seismic earth: Array analysis of broadband seismograms*: AGU, **157**, of *Geophysical Monograph*, 37–47.
- Nolet, G., J. van Trier, and R. Huisman, 1986, A formalism for nonlinear inversion of seismic surface waves: *Geophysical Research Letters*, **13**, 26–29, doi: [10.1029/GL013i001p00026](https://doi.org/10.1029/GL013i001p00026).
- Operto, S., C. Ravaut, L. Impropa, J. Virieux, A. Herrero, and P. D. Aversana, 2004, Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltimes and waveform inversions:

- A case study: Geophysical Prospecting, **52**, 625–651, doi: [10.1111/j.1365-2478.2004.00452.x](https://doi.org/10.1111/j.1365-2478.2004.00452.x).
- Plattner, A., H.-R. Maurer, J. Vorloeper, and M. Blome, 2012, 3-D electrical resistivity tomography using adaptive wavelet parameter grids: Geophysical Journal International, **189**, 317–330, doi: [10.1111/j.1365-246X.2012.05374.x](https://doi.org/10.1111/j.1365-246X.2012.05374.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).
- Pratt, R. G., and R. M. Shipp, 1999, Seismic waveform inversion in the frequency domain, Part 2: Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914, doi: [10.1190/1.1444598](https://doi.org/10.1190/1.1444598).
- Rickers, F., A. Fichtner, and J. Trampert, 2012, Imaging mantle plumes with instantaneous phase measurements of diffracted waves: Geophysical Journal International, **190**, 650–664, doi: [10.1111/j.1365-246X.2012.05515.x](https://doi.org/10.1111/j.1365-246X.2012.05515.x).
- Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931, doi: [10.1111/j.1365-246X.2008.03768.x](https://doi.org/10.1111/j.1365-246X.2008.03768.x).
- Shin, C., and Y. H. Cha, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, **177**, 1067–1079, doi: [10.1111/j.1365-246X.2009.04102.x](https://doi.org/10.1111/j.1365-246X.2009.04102.x).
- Simons, F. J., B. D. E. Dando, and R. M. Allen, 2006, Automatic detection and rapid determination of earthquake magnitude by wavelet multiscale analysis of the primary arrival: Earth and Planetary Science Letters, **250**, 214–223, doi: [10.1016/j.epsl.2006.07.039](https://doi.org/10.1016/j.epsl.2006.07.039).
- Simons, F. J., I. Loris, G. Nolet, I. C. Daubechies, S. Voronin, J. S. Judd, P. A. Vetter, J. Charlety, and C. Vonesch, 2011, Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity: Geophysical Journal International, **187**, 969–988, doi: [10.1111/j.1365-246X.2011.05190.x](https://doi.org/10.1111/j.1365-246X.2011.05190.x).
- Simons, F. J., G. Nolet, P. Georgief, J. M. Babcock, L. A. Regier, and R. E. Davis, 2009, On the potential of recording earthquakes for global seismic tomography by low-cost autonomous instruments in the oceans: Journal of Geophysical Research, **114**, B05307, doi: [10.1029/2008JB006088](https://doi.org/10.1029/2008JB006088).
- Sirgue, L., and R. 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).
- Stockwell, J. W., 1999, The CWP/SU: Seismic Un*x package: Computers & Geosciences, **25**, 415–419, doi: [10.1016/S0098-3004\(98\)00145-9](https://doi.org/10.1016/S0098-3004(98)00145-9).
- Strang, G., and T. Nguyen, 1997, Wavelets and filter banks, 2nd ed.: Wellesley-Cambridge Press.
- Tae-Kyung, H., and B. L. N. Kennett, 2002, A wavelet-based method for simulation of two-dimensional elastic wave propagation: Geophysical Journal International, **150**, 610–638, doi: [10.1046/j.1365-246X.2002.01714.x](https://doi.org/10.1046/j.1365-246X.2002.01714.x).
- 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).
- Tape, C., Q. Liu, and J. Tromp, 2007, Finite-frequency tomography using adjoint methods — Methodology and examples using membrane surface waves: Geophysical Journal International, **168**, 1105–1129, doi: [10.1111/j.1365-246X.2006.03191.x](https://doi.org/10.1111/j.1365-246X.2006.03191.x).
- 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).
- Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903, doi: [10.1190/1.1442046](https://doi.org/10.1190/1.1442046).
- Tian, Y., S.-H. Hung, G. Nolet, R. Montelli, and F. A. Dahlen, 2007a, Dynamic ray tracing and traveltimes corrections for global seismic tomography: Journal of Computational Physics, **226**, 672–687, doi: [10.1016/j.jcp.2007.04.025](https://doi.org/10.1016/j.jcp.2007.04.025).
- Tian, Y., R. Montelli, G. Nolet, and F. A. Dahlen, 2007b, Computing traveltimes and amplitude sensitivity kernels in finite-frequency tomography: Journal of Computational Physics, **226**, 2271–2288, doi: [10.1016/j.jcp.2007.07.004](https://doi.org/10.1016/j.jcp.2007.07.004).
- Trampert, J., and J. Spetzler, 2006, Surface wave tomography: Finite-frequency effects lost in the null space: Geophysical Journal International, **164**, 394–400, doi: [10.1111/j.1365-246X.2006.02864.x](https://doi.org/10.1111/j.1365-246X.2006.02864.x).
- Tromp, J., D. Komatitsch, V. Hjørleifsdóttir, Q. Liu, H. Zhu, D. Peter, E. Bozdag, D. McRitchie, P. Friberg, C. Trabant, and A. Hutko, 2010, Near real-time simulations of global CMT earthquakes: Geophysical Journal International, **183**, 381–389, doi: [10.1111/j.1365-246X.2010.04734.x](https://doi.org/10.1111/j.1365-246X.2010.04734.x).
- Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216, doi: [10.1111/j.1365-246X.2004.02453.x](https://doi.org/10.1111/j.1365-246X.2004.02453.x).
- van Leeuwen, T., and F. Herrmann, 2013, Fast waveform inversion without source-encoding: Geophysical Prospecting, **61**, 10–19, doi: [10.1111/j.1365-2478.2012.01096.x](https://doi.org/10.1111/j.1365-2478.2012.01096.x).
- Versteeg, R. J., 1993, Sensitivity of prestack depth migration to the velocity model: Geophysics, **58**, 873–882, doi: [10.1190/1.1443471](https://doi.org/10.1190/1.1443471).
- Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901, doi: [10.1190/1.1442147](https://doi.org/10.1190/1.1442147).
- Wang, Y., and G. A. Houseman, 1995, Tomographic inversion of reflection seismic amplitude data for velocity variation: Geophysical Journal International, **123**, 355–372, doi: [10.1111/j.1365-246X.1995.tb06859.x](https://doi.org/10.1111/j.1365-246X.1995.tb06859.x).
- Williamson, P. R., and M. H. Worthington, 1993, Resolution limits in ray tomography due to wave behavior: Numerical experiments: Geophysics, **58**, 727–735, doi: [10.1190/1.1443457](https://doi.org/10.1190/1.1443457).
- Wu, Q., Y. Li, R. Zhang, and R. Zeng, 2007, Wavelet modelling of broadband receiver functions: Geophysical Journal International, **170**, 534–544, doi: [10.1111/j.1365-246X.2007.03467.x](https://doi.org/10.1111/j.1365-246X.2007.03467.x).
- Wu, R., and K. Aki, 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595, doi: [10.1190/1.1441934](https://doi.org/10.1190/1.1441934).
- Wu, R.-S., and F. Yang, 1997, Seismic imaging in wavelet domain: Decomposition and compression of imaging operator, in A. Aldroubi, A. F. Laine, and M. A. Unser, eds., Wavelet applications in signal and image processing V, **3169**: SPIE, 148–162.
- Zhao, L., T. H. Jordan, and C. H. Chapman, 2000, Three-dimensional Fréchet differential kernels for seismic delay times: Geophysical Journal International, **141**, 558–576, doi: [10.1046/j.1365-246X.2000.00085.x](https://doi.org/10.1046/j.1365-246X.2000.00085.x).
- Zhou, Y., F. A. Dahlen, and G. Nolet, 2004, Three-dimensional sensitivity kernels for surface wave observables: Geophysical Journal International, **158**, 142–168, doi: [10.1111/j.1365-246X.2004.02324.x](https://doi.org/10.1111/j.1365-246X.2004.02324.x).
- Zhou, Y., F. A. Dahlen, and G. Nolet, 2005, Finite-frequency effects in global surface-wave tomography: Geophysical Journal International, **163**, 1087–1111, doi: [10.1111/j.1365-246X.2005.02780.x](https://doi.org/10.1111/j.1365-246X.2005.02780.x).
- Zhu, H., E. Bozdag, 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., Y. Luo, T. Nissen-Meyer, C. Morency, and J. Tromp, 2009, Elastic imaging and time-lapse migration based on adjoint methods: Geophysics, **74**, no. 6, WCA167–WCA177, doi: [10.1190/1.3261747](https://doi.org/10.1190/1.3261747).