Multiscale adjoint waveform tomography for surface and body waves

Yanhua O. Yuan[1], Frederik J. Simons[1], and Ebru Bozdağ[2]

ABSTRACT

We have developed a wavelet-multiscale adjoint scheme for the elastic full-waveform inversion of seismic data, including body waves (BWs) and surface waves (SWs). We start the inversion on the SW portion of the seismograms. To avoid cycle skipping and reduce the dependence on the initial model of these dispersive waves, we commence by minimizing an envelope-based misfit function. Subsequently, we proceed to the minimization of a waveform-difference (WD) metric applied to the SWs only. After that, we fit BWs and SWs indiscriminately using a WD misfit metric. In each of these three steps, we guide the iterative inversion through a sequence of nested subspace projections in a wavelet basis. SW analysis preserves a wealth of near-surface features that would be lost in conventional BW tomography. We used a toy model to illustrate the dispersive and cycle-skipping behavior of the SWs, and to introduce the two ways by which we combat the nonlinearity of waveform inversions involving SWs. The first is the wavelet-based multiscale character of the method, and the second the envelope-based misfit function. Next, we used an industry synthetic model to perform realistic numerical experiments to further develop a strategy for SW and joint SW as well as BW tomography. The effect of incorrect density information on wave-speed inversions was also evaluated. We ultimately formalize a flexible scheme for full-waveform inversion based on adjoint methods that includes BWs and SWs, and also considers P- and S-wave speeds, as well as density. Our method is applicable to waveform inversion in exploration geophysics, geotechnical engineering, regional, and global seismology.

INTRODUCTION

Near-surface heterogeneities are responsible for complex scattering and mode conversions. Characterizing near-surface heterogeneity is crucial for statics corrections and to analyze wave propagation in the deep structure. Rayleigh and Love waves account for the bulk of the energy in the seismic wavefield — two-thirds of the total energy input by a circular footing vibrating harmonically over a homogeneous isotropic elastic half-space (Miller and Pursey, 1955). The energy of surface waves (SWs) is dissipated proportionally to the distance from the source (Rayleigh, 1885), whereas the body-wave (BW) energy decay scales with the square of distance traveled in the whole space, and even faster near the free surface (Ewing et al., 1957; Richart et al., 1970). Thus, at some distance from the source, the seismic wavefield is essentially dominated by SWs. Unlike BWs, which may penetrate to great depths, SW propagation paths are concentrated to depths that are on the order of their wavelength (Dahlen and Tromp, 1998). Especially for S-wave velocities, SWs provide strong constraints on near-surface structure. Despite this, SWs ("ground roll") are most commonly removed in exploration-scale industry applications (Dobrin and Savit, 1988), which not only deprives the records of a certain amount of information, but also tends to introduce errors through transformation and filtering. In global seismology and mantle tomography, "crustal corrections" (Bozdağ and Trampert, 2008) *mutatis mutandis* play a role equivalent to SW removal. Involving the SWs in seismic tomography eliminates the burdensome step of their removal in preprocessing, and treats them for the signal that they are.

Surface waves: Applications

SW analysis has a long history in global seismology (Woodhouse, 1974) and exploration geophysics (McMechan and Yedlin, 1981), and it is now widely embraced as a valuable tool to conduct

subsurface characterization (Socco and Strobbia, 2004; O'Neill and Matsuoka, 2005; Socco et al., 2010) in different research fields, including geotechnical and geoenvironmental engineering (Foti, 2000; Rix et al., 2001). SW analysis may be more important even than other popular methods such as refraction or resistivity surveys or magnetic imaging (Crice, 2005). SW can characterize a medium at large (e.g., in regional and global seismology, to determine the structure of the earth's crust and upper mantle), intermediate (e.g., in exploration geophysics, to constrain near-surface structure, or to correct statics; or in geotechnical engineering, to infer the shear stiffness of the ground materials), and even at the smallest scales, for the nondestructive evaluation of engineering materials (e.g., using ultrasonic SWs to detect material defects; Thompson and Chimenti, 1997; Bagheri et al., 2014).

Known by their acronyms SWA, SASW, or MASW, surface-waves analysis and spectral or multichannel analysis (Park et al., 1999) of surface waves are now widely used methods for seismic site characterization in geotechnical engineering, which values their noninvasive nature, good resolution at shallow depth, and efficiency in time and cost. The methods have been developed greatly during the past few decades. Tokimatsu et al. (1992) recognized the effects of higher modes in some types of soil profiles, when the S-wave speed profile is not increasing regularly with depth. Foti (2000) used a multistation method for the robust determination of dispersion curves. Other variants of SW dispersion analysis include f-k (Capon, 1969), τ-p (Buland and Chapman, 1983), and spatial autocorrelation (Aki, 1957) methods. Multichannel methods (Miller et al., 1999; Pratt and Shipp, 1999) have been proposed to improve data quality control under the influence of noise. A variety of inversion algorithms (Yuan and Nazarian, 1993; Xia et al., 1999; Rix et al., 2001) have been developed for determining S-wave speed profiles from dispersion curves. In addition to the estimation of stiffness profiles, e.g., for pavement system evaluation and site characterization, SW analysis has been used to obtain in situ material damping ratio profiles for general site investigations either separately (Rix et al., 2000; Xia et al., 2002) or simultaneously (Lai and Rix, 1998; Foti, 2000) by considering the coupling between phase velocity and attenuation as part of two-station or multistation methods.

In regional and global seismology, the variation of the propagation speeds of long-period SWs has been observed and interpreted for the study of crustal and upper-mantle structure for decades (e.g., Woodhouse, 1974; Woodhouse and Dziewoński, 1984; Capdeville and Cance, 2015). By the 1980s, efforts to use SW dispersion and phase-velocity measurements at long periods to constrain regional and global-scale mantle S-wave structure, including its anisotropy, using tomographic techniques were very well established (Montagner and Nataf, 1988; Montagner and Jobert, 1988; Snieder, 1988a, 1988b). To name but a few additional examples: overtone SWs were used by Cara et al. (1984) and Lévêque and Cara (1985) to provide evidence for upper mantle anisotropy. Nolet et al. (1986) formalized a waveform-fitting approach using the conjugate gradient method for Love and Rayleigh waves. Stutzmann and Montagner (1993, 1994) extended the use of higher modes of SWs to study structure in the transition zone. Ekström et al. (1997) mapped phase velocity by minimizing dispersion residuals of fundamental Love and Rayleigh waveforms, isolated from interfering overtones via phase-matched filters. Van Heijst and Woodhouse (1999) measured global high-resolution phase velocity distributions of fundamental-mode and overtone SWs via a mode-branch strip-

ping technique (van Heijst and Woodhouse, 1997). Trampert and Woodhouse (2003) used fundamental-mode SWs to map anisotropic phase velocities. After S20RTS, Ritsema et al. (2011) developed S40RTS, a shear-velocity model of the mantle, using Rayleigh-wave dispersion, normal-mode splitting function, and S-wave traveltime measurements. Even as (high frequency) BW studies generally yield sharper tomographic images (Rawlinson and Sambridge, 2003; Romanowicz, 2003; Nolet, 2008), (low frequency) SWs have contributed greatly to our understanding of the long-wavelength internal structure of the earth. In some areas, such as beneath the ocean basins, they are crucial to increase vertical resolution, or to make up for insufficient sampling in regions with poor ray coverage due to lack of stations (Romanowicz and Giardini, 2001; Simons et al., 2006b).

In hydrocarbon exploration, seismic techniques have generally been based on BW propagation, in particular P-wave reflections. In contrast, SWs, despite their preponderance in the seismic record, are usually considered as coherent noise masking the reflections, hence to be removed by recording and processing procedures (e.g., Dobrin and Savit, 1988). In seismic exploration, McMechan and Yedlin (1981) extracted dispersion curves of Rayleigh waves from common-shot marine seismic profiles based on slant stacking followed by 1D Fourier transformation. Gabriels et al. (1987) determined S-wave velocities in sediments to a depth of 50 m by means of higher mode Rayleigh waves. The industry is increasingly recognizing the value of SWs for seismic inversions. Some examples: Ivanov et al. (2006) showed that a reference model derived from SW S-wave speed estimation reduces the nonuniqueness of the refraction inversion problem. Gouédard et al. (2012) combined SW eikonal tomography and crosscorrelation methods for phase arrival-time estimation for velocity analysis of a strongly heterogeneous and scattering medium in a hydrocarbon-exploration setting. Droujinine et al. (2012) developed an integrated workflow with dispersion curve and full-waveform inversion to retrieve complex shallow structure, as needed for the accurate imaging of deeper targets.

Surface waves: Methods and challenges

Most of the applications of SWs in different disciplines operate on the same principle, which is to estimate a set of dispersion curves from the data, and subsequently, to solve an inverse problem for elastic or anelastic parameters (Haskell, 1953; Stokoe et al., 2004; O'Neill and Matsuoka, 2005). SW analysis and its applications have their challenges. First, the success of SW dispersion-curve inversion depends on the clear separation of fundamental and higher modes, which can be realized only if very dense spatial sampling and long acquisition spreads are used (Socco and Strobbia, 2004). Many strategies in acquisition and processing have been developed to help identify and separate different modes (Lai and Rix, 1999; Beaty et al., 2002; O'Neill and Matsuoka, 2005); however, several authors (e.g., Zhang and Chan, 2003; Maraschini et al., 2010; Socco et al., 2010) show that mode misidentification is not easy to avoid and may produce significant errors. Second, SW dispersion-curve inversions usually do not consider the effect of 3D inhomogeneity in the medium, until many of them are combined to produce the lateral variations that are precisely the target also in exploration applications. In regional and global seismology, 1D profiles are jointly inverted to recover 2D or 3D lateral heterogeneities (Woodhouse, 1974; Nataf et al., 1986; Nolet, 1990). 1D

reference models, such as the Preliminary Reference Earth Model (Dziewoński and Anderson, 1981) are a good starting point for matching SW observations relatively easily.

As an alternative, full-waveform inversion of SWs considers the propagation of SWs in a realistically heterogeneous medium (Snieder, 1988b), in which case the separation of different modes and the prior estimation of dispersion curves are not necessary. Even in that case, due to their strongly dispersive nature, the complex interference of fundamental and higher modes and the potential for cycle skipping, particularly when an adequate initial estimate is not available, SWs are much more difficult to handle than BWs. In addition to these complications, the low-frequency ground motion may not be very well recorded in industry applications.

Cycle skipping causes local minima in the inversion (VanDecar and Crosson, 1990). Frequency-dependent phase measurements based on crosscorrelations of predicted and observed SWs, and layer-stripping approaches (Pratt et al., 1996) are designed to mitigate cycle-skipping problems of SW tomography (Lebedev et al., 2005; Sigloch and Nolet, 2006). Another possible solution to the nonlinearity problem is to use a multiscale approach (Bunks et al., 1995; Sirgue and Pratt, 2004; de Hoop et al., 2012; Yuan and Simons, 2014), proceeding from large (low frequencies) to small scales (high frequencies) in the seismograms, progressively involving higher (temporal) frequencies. Prieux et al. (2013a, 2013b) and Operto et al. (2013) discuss strategies to reduce the nonlinearity of the elastic multiparameter inversion with multicomponent data and to control the trade-off between parameters by hierarchically selecting the data components to invert and the parameter classes to update. Brossier et al. (2009) use a frequency-domain preconditioning scheme equivalent to time-domain damping to stabilize elastic inversion. Shin and Cha (2008) transform the wavefield to the Laplace domain to reduce the sensitivity on the initial model. Pérez Solano et al. (2014) use a windowed-amplitude waveform inversion method in the Fourier domain for near-surface imaging using SWs.

Surface waves: A new approach

Reducing the nonlinearity of waveform inversion problems involves a judicious choice of the objective function (Luo and Schuster, 1991; Gee and Jordan, 1992; Dahlen and Baig, 2002; Fichtner et al., 2008; Bozdağ et al., 2011; Rickers et al., 2012; Masoni et al., 2014). Bozdağ et al. (2011), in particular, discuss several misfit functions in seismic tomography. Among those, we develop the envelope-based objective function to measure SWs in this study. The extraction of envelopes from SWs through the Hilbert transform greatly reduces the nonlinearity of waveform inversion by the separation of phase and amplitude information.

In this paper, we explore the sensitivity of SWs in waveform inversion to estimate near-surface structure. To address the problems of cycle skipping in the waveform fitting, we develop a strategy based on a wavelet-multiscale approach, newly combined with an envelope-difference (ED) misfit functional (Yuan et al., 2014). We are motivated by the intrinsic multiresolution property of the wavelet transform (Mallat, 1989), which provides a simple and natural framework for interpreting signals at different levels of detail. Goals similar to those that we aim to achieve in this study could possibly be achieved via traditional, convolutional, or Fourier-domain time-frequency filtering approaches (e.g., Pratt, 1999; Sirgue and Pratt, 2004; Fichtner et al., 2013; Prieux et al., 2013a). In this study, we focus exclusively on wavelet-based time-scale analysis. Temporal scales of the seismo-

grams illuminate spatial scales in the subsurface structure, but of course the correspondence is not exact in any basis: specific subbands of the time-domain seismograms do not map to the same subbands of the space-domain structure (see, e.g., Beylkin, 1992; Ecoublet et al., 2002).

Our work should be considered as an extension to our previous paper (Yuan and Simons, 2014), in which we introduced a multiresolution technique to adjoint-based seismic tomography (e.g., Fichtner, 2011; Luo et al., 2014), but which excluded SWs (but we included all reflected, refracted, transmitted BW phases, and multiples). We used time-domain wavelet-based constructive approximation to progressively model smaller scale features in the seismogram. We treated SWs as noise and removed them using low pass and dip filtering in our preprocessing procedure before performing the inversions, as in common industry practice. In the present paper, we compare the sensitivities of SWs and BWs in full-waveform inversions, and we discuss strategies to combine both of them for elastic inversions. Furthermore, we also discuss the effects of incorrect density information on elastic parameter estimation. Finally, we formalize a flexible workflow to realize the goal of truly "full"-waveform inversion, in the form of an iterative adjoint tomography method that considers density- and elastic-wavespeed variations, and which treats SWs separately in a first step before conducting a joint analysis of SWs and BWs in the second stage.

MATHEMATICAL PRELIMINARIES

For the development of our method, we rely on two mathematical transforms, the wavelet transform and the Hilbert transform, which we briefly introduce here.

Wavelet transform

Readers not familiar with the wavelet transform may wish to consult the early work of Morlet et al. (1982a, 1982b), in this journal, or the textbooks by Daubechies (1992), Strang and Nguyen (1997), Mallat (2008), or Jensen and la Cour-Harbo (2001) before reading on.

For a given input signal $s(t)$, we first choose a particular wavelet basis (see Yuan and Simons, 2014). We consider two sets of scaling functions, $\tilde{\phi}_k^j$ for the analysis and ϕ_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 $j = 1, \ldots, J$ indicates the scale, J is the maximal decomposition depth, and k is a measure of the translation in time. The generic signal $s(t)$ can be represented as the sum over all translates k and scales $j = 1, \ldots, J$, in the expansion

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

and a partial reconstruction to scale j, leading to a constructive approximation, is

$$s_j(t) = \sum_k a_k^J \phi_k^J(t) + \sum_{j'=j+1}^{J} \sum_k d_k^{j'} \psi_k^{j'}(t) = \sum_k a_k^j \phi_k^j(t), \quad (2)$$

where the scaling (approximation, low frequency, and low pass) and wavelet (detail, high frequency, and high pass) coefficients, respectively, whatever the algorithmic implementation, are given by

$$a_k^J = \langle s, \tilde{\phi}_k^J \rangle, \quad \text{and} \quad d_k^j = \langle s, \tilde{\psi}_j^k \rangle. \tag{3}$$

Hilbert transform

Following Claerbout (1992), the analytic signal of a real-valued signal $s(t)$ can be expressed as

$$s_a(t) = s(t) + i\mathcal{H}\{s(t)\} = E(t)e^{i\phi(t)}, \tag{4}$$

where $\mathcal{H}\{s(t)\}$ is the Hilbert transform of the real signal $s(t)$ and $\phi(t)$ and $E(t)$ stand for the instantaneous phase and the instantaneous amplitude (or envelope) of the analytic signal, respectively,

$$\phi(t) = \arctan\frac{\mathcal{H}\{s(t)\}}{s(t)}, \tag{5}$$

$$E(t) = \sqrt{s^2(t) + \mathcal{H}^2\{s(t)\}}. \tag{6}$$

The separation of phase and amplitude information via the Hilbert transform can be carried out via the fast Fourier transform, suppressing negative frequencies. Additional details can be found in Mallat (2008). We now list two properties of the Hilbert transform that will be used later.

The derivative of the Hilbert transform of a real signal $s(t)$ is the Hilbert transform of the derivative of $s(t)$:

$$\delta\mathcal{H}\{s(t)\} = \mathcal{H}[\delta\{s(t)\}]. \tag{7}$$

The Hilbert transform is an anti-self-adjoint operator. For two real signals $s(t)$ and $u(t)$, we have

$$\langle \mathcal{H}\{s(t)\}, u(t) \rangle = -\langle s(t), \mathcal{H}\{u(t)\} \rangle. \tag{8}$$

A MOTIVATING EXAMPLE

The inversion of waveforms that contain SWs greatly increases the nonlinearity of full-waveform inversion. Even a small perturbation, especially in S-wave speed structure, to a homogeneous background, will cause dispersion, turning a simple SW pulse into a complicated wavetrain.

We designed a simple 2D toy model to illustrate the dispersion and cycle-skipping behavior of SWs. Consider a homogeneous P-wave speed ($V_P = 2000$ m/s) and density ($\rho = 1200$ kg/m^3) model. The S-wave speed model shown in Figure 1a consists of a homogeneous background shear velocity $V_{S0} = 800$ m/s and an anomalous layer of $V_{S1} = 1000$ m/s. The model measures 400 m in the horizontal direction with 79 uniform mesh nodes (quadrangles in 2D), and it is 100 m in the vertical direction with 19 uniform nodes. A total of $(79 \times 4 + 1) \times (19 \times 4 + 1) = 24,409$ unique grid points and $79 \times 19 \times (4 + 1)^2 = 37,525$ S-wave speeds are used for 2D spectral element discretization with polynomial degree four (Komatitsch et al., 2005). The quadrangle mean spacing is 5.3 m.

We model horizontal and vertical components of the displacement seismograms using SPECFEM2D, a 2D spectral-element code (Komatitsch and Vilotte, 1998). A force is imposed normally to the surface with a 40-Hz Ricker wavelet, located at 50 m hori-

zontally from the left edge and at 0.5 m in depth. We use a total of 401 receivers located at 0.5 m depth and spaced 1 m apart. We also consider a homogeneous S-wave speed model, where $V_{S0} = V_{S1} = 900$ m/s, as an "initial model" for the purpose of our numerical evaluation, in which we implement SPECFEM2D to compute synthetics with the same source-receiver geometry.

Dispersion of surface waves

The dispersive nature of the SWs, waves of different wavelengths that travel with different speeds, is manifest by the drawn-out complex shapes of the waveform in the target model (Figure 1b and 1d) as compared with the waveforms in a homogeneous background (Figure 1c and 1e).

Horizontal components of the SWs are shown as recorded at a horizontal distance of 350 m from the left edge. The inclusion of the anomalous S-wave layer in the toy model causes the modeled SW to be dispersive. In contrast, the SW trace predicted in a homogeneous initial model is nondispersive. Figure 1b and 1c shows full-resolution seismograms and their envelopes, whereas Figure 1d and 1e shows their approximation after partial reconstruction to scale 8 using a Daubechies (1988) wavelet basis ("D12" or "db6" with six "vanishing moments," see Daubechies, 1992; Strang and Nguyen, 1997; Jensen and la Cour-Harbo, 2001). The number of vanishing moments is one more than the degree of polynomials that can be represented by the scaling functions without contributions from the wavelets.

The traditional analysis of SW would produce dispersion curves relating measured phase or group velocities to their dominant period and then invert those for an average 1D S-wave speed profile. In the presence of 2D or 3D lateral heterogeneity, such procedures would become necessarily very complex, although they have been a staple of waveform analysis for many decades.

Cycle skipping of surface waves

Waveform inversions starting from an inadequate initial model run the risk of convergence to a secondary minimum because the phase difference between observation and prediction may exceed half the period. Such cycle-skipping problems are more severe with a dispersive SW than with BW waveforms. To combat such nonlinearities, we previously developed a wavelet multiscale strategy for BW waveform inversion, in which the initial breakdown of the traces to the coarsest scales effectively linearized the inversion problem (Yuan and Simons, 2014). In this paper, we elaborate on this framework but, this time, we include the inversion of frequency-dependent SW waveforms.

Figure 1d and 1e shows the SW waveforms projected onto scale 8, our largest decomposition level in a basis using D12 wavelets. Although the discrepancy between the target and the initial SW waveforms (thick black lines) is clearly much smaller at this large-scale projection than at the original full resolution shown in Figure 1b and 1c, it remains too challenging to attempt waveform fitting with the mean-squared difference as a measure of distance to be minimized.

Thus, to further combat cycle skipping and other confounding effects caused by the highly oscillatory nature of SW phases even in this simple toy model, we propose to take the Hilbert transform of the waveforms and work with the amplitude information contained in their envelopes (dashed lines in Figure 1) before attempting to

reconstruct the full-phase information. Envelopes smooth out small-scale oscillations in the seismogram. It is also clear from Figure 1 that the envelopes of the target model and the homogeneous-model SWs are much closer to one another at the coarsest scale than they are at the full resolution, suggesting that they may be used in the initial steps of a full-waveform inversion.

ENVELOPE-BASED ADJOINT MODELING

Bozdağ et al. (2011) introduced a misfit function for adjoint-based elastic full-waveform tomography (see Appendix A for a brief descrip-

tion of the adjoint method) that relies on the squared logarithmic ratio of the envelopes of observed and synthetic seismograms (see Appendix B, equation B-1). Their adjoint-source expression contains the squared envelope of the synthetic seismogram in the denominator. The logarithmic metric is rather sensitive to small perturbations by scattering or weak reflections, especially in the coda (the tail of strongly scattered waves).

Here, we take the difference of the envelopes as an alternative measure of misfit to quantify the distance in the amplitudes. We define the ED misfit function between observed $d(\mathbf{x}_r, \mathbf{x}_s, t)$ and synthetic $s(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ data in the mean-squared sense over all sources s and receivers r as
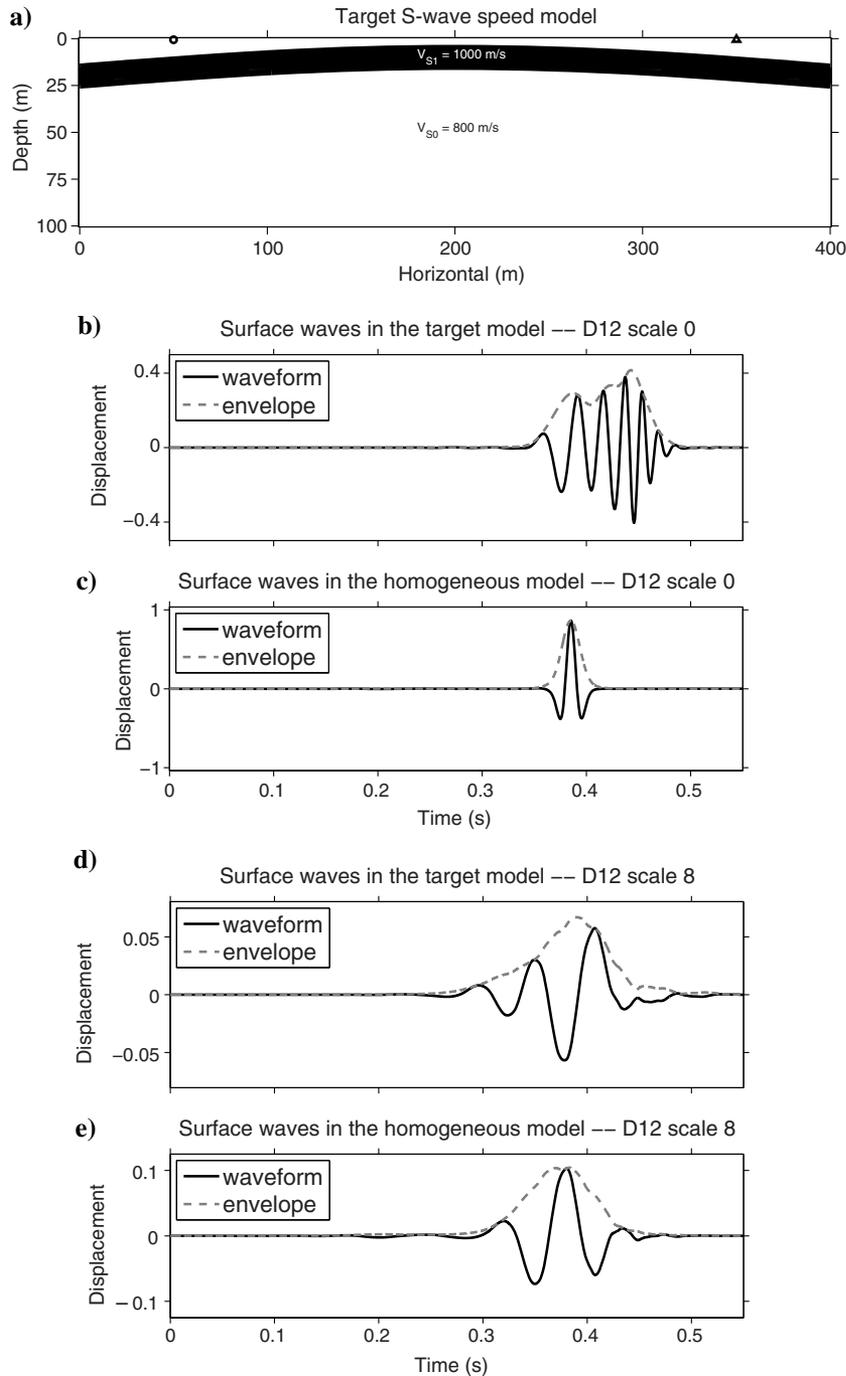
Figure 1. (a) A 2D elastic S-wave speed model which consists of a homogeneous $V_{S0} = 800$ m/s and an anomalous layer with $V_{S1} = 1000$ m/s. SW waveforms and their envelopes modeled in the target model for the source-receiver surface pair in the top panel, (b) at a scale 0 at full resolution and (d) at scale 8 of the D12 wavelet transform. SW waveforms and envelopes predicted at (c) full resolution and (e) at scale 8, calculated with the same P-wave speed and density model, but with a homogeneous S-wave speed of 900 m/s.

$$\chi_1(\mathbf{m}) = \frac{1}{2}\sum_{s,r}\int_0^T \|E_s(\mathbf{x}_r,\mathbf{x}_s,t;\mathbf{m}) - E_d(\mathbf{x}_r,\mathbf{x}_s,t)\|^2 dt, \quad (9)$$

where T is the window length. We use the symbol $\|\cdot\|$ throughout to denote the norm of single-component (s,d) or multicomponent $(\mathbf{s},\mathbf{d})$ seismograms, which we distinguish (by font weight) only in Appendix A. Because $E_s(t)$ and $E_d(t)$ are the envelopes of the synthetic $s(t)$ and the observed data $d(t)$, respectively, they are obtained via Hilbert transformation as

$$E_s(\mathbf{x}_r,\mathbf{x}_s,t;\mathbf{m}) = \sqrt{s^2(t) + \mathcal{H}^2\{s(t)\}}, \quad (10)$$

$$E_d(\mathbf{x}_r,\mathbf{x}_s,t) = \sqrt{d^2(t) + \mathcal{H}^2\{d(t)\}}. \quad (11)$$

Gradient-based methods (see Tarantola, 1984a, 1984b; Tromp et al., 2005) require the derivative with respect to the model parameters of the misfit function $\chi_1(\mathbf{m})$ in equation 9, which we express in terms of δE_s, the perturbation in the envelopes of the synthetics due to a perturbation of the current model $\delta\mathbf{m}$:

$$\delta\chi_1(\mathbf{m}) = \sum_{s,r}\int_0^T [E_s(\mathbf{x}_r,\mathbf{x}_s,t;\mathbf{m}) - E_d(\mathbf{x}_r,\mathbf{x}_s,t)]\delta E_s(\mathbf{x}_r,\mathbf{x}_s,t;\mathbf{m})dt.$$
$$(12)$$

To avoid notational clutter, we drop the dependence on the space, time, and model coordinates $\mathbf{x}_r$, $\mathbf{x}_s$, t, and $\mathbf{m}$, and write

$$\delta E_s = \frac{s\delta s + [\mathcal{H}\{s\}]\delta[\mathcal{H}\{s\}]}{E_s}, \quad (13)$$

where δs and $\delta[\mathcal{H}\{s\}]$ are the perturbations to the synthetic seismogram s and its Hilbert transform $\mathcal{H}\{s\}$ due to a model perturbation $\delta\mathbf{m}$.

We introduce the first of three possible envelope ratios (see Appendix B for the other two), E_1^{rat}, to capture the difference of the envelopes of the current predicted and the target seismograms relative to the envelope of the predicted seismograms:

$$E_1^{\text{rat}} = \frac{E_s - E_d}{E_s}, \quad (14)$$

so that after substituting equation 13 into equation 12, and using the differentiation rule (equation 7) and the anti-self-adjointness (equation 8) of the Hilbert transform, we can write

$$\delta\chi_1 = \sum_{s,r}\int_0^T E_1^{\text{rat}}(s\delta s + [\mathcal{H}\{s\}]\delta[\mathcal{H}\{s\}])dt,$$

$$= \sum_{s,r}\int_0^T (E_1^{\text{rat}}s\delta s + E_1^{\text{rat}}[\mathcal{H}\{s\}][\mathcal{H}\{\delta s\}])dt,$$

$$= \sum_{s,r}\int_0^T (E_1^{\text{rat}}s - \mathcal{H}\{E_1^{\text{rat}}[\mathcal{H}\{s\}]\})\delta s dt. \quad (15)$$

In short, the derivative of the misfit function is rewritten in the very compact form of equation 15 (see also Wu et al., 2014). The adjoint source associated with a single event $\mathbf{x}_s$ is given by (in time-reversed coordinates)

$$f^\dagger(\mathbf{x},t) = \sum_r (E_1^{\text{rat}}s - \mathcal{H}\{E_1^{\text{rat}}[\mathcal{H}\{s\}]\})\delta(\mathbf{x} - \mathbf{x}_r). \quad (16)$$

The adjoint source is retransmitted to generate an adjoint wavefield (see Tromp et al. [2005], their equation 11), which illuminates the discrepancy between the observed and the predicted envelopes corresponding to the event located at $\mathbf{x}_s$. The zero-lag crosscorrelation of the forward and the adjoint wavefields yields an event kernel (associated with one source, see Tape et al. [2007], their Section 5.1). The sum of all such kernels combines the contributions from all events to define a misfit kernel (see Tromp et al. [2005], their Section 4.2): the gradient of the ED misfit function (summed over all sources and receivers) with respect to the current model $\mathbf{m}$ (see also Yuan and Simons, 2014).

A STRATEGY FOR SURFACE-WAVE INVERSION

With the material developed until now, we identify four different types of ways (waveforms or envelopes, full-resolution, or multiscale analysis) to incorporate, specifically, SWs into a full-waveform inversion procedure that uses wavelets rather than traditional filtering approaches based on Fourier analysis (for an illustration of their difference, see Simons et al., 2006a). We can consider the waveforms at their native resolution or in a multiscale framework. Alternatively, we can work with the waveform envelopes at full resolution or in a multiscale approximation (see again Figure 1). When it comes to making measurements on (wavelet subspace) waveforms (or their envelopes), various options are available to us. Yuan and Simons (2014) focus on BWs with multiresolution waveform-difference (WD) measurements (their equations 1 and 11), which we restate here (for individual wavelet scales in the notation of our equation 2) as

$$\chi_j(\mathbf{m}) = \frac{1}{2}\sum_{s,r}\int_0^T \|s_j(\mathbf{x}_r,\mathbf{x}_s,t;\mathbf{m}) - d_j(\mathbf{x}_r,\mathbf{x}_s,t)\|^2 dt. \quad (17)$$

Now, to be able to incorporate SWs, we will discuss multiresolution waveform-envelope measurements, whereby for the specific measurement made on the envelopes of the seismograms constructively

approximated to scale j, we have the choice of misfits χ_1, per equation 9 (or, as found in Appendix B, χ_2 or χ_3, per equations B-1 and B-4). In this paper, only χ_1 is used for our numerical experiments that involve envelopes, but equations 9–11 acquire an index, identifying the wavelet scale of the reconstruction, with the envelopes calculated after the multiscale recomposition. Thus, we finally write the multiscale ED misfit function between the envelopes of the subspace projections of the observed $d_j(\mathbf{x}_r, \mathbf{x}_s, t)$ and the synthetic $s_j(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ data as

$$\chi_j^1(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int_0^T \|E_{s_j}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m}) - E_{d_j}(\mathbf{x}_r, \mathbf{x}_s, t)\|^2 dt,$$

(18)

where the envelopes of the partially reconstructed seismograms are, per equations 2, 10, and 11

$$E_{s_j} = \sqrt{s_j^2 + \mathcal{H}^2\{s_j\}}, \quad \text{and} \quad E_{d_j} = \sqrt{d_j^2 + \mathcal{H}^2\{d_j\}}.$$

(19)

To evaluate the behavior of the various possible measurements in SW inversions, we conduct a numerical experiment in which we consider four different objective functions. These are: (1) the WD at full resolution, i.e., the "traditional" metric; (2) the WD of seismograms progressively reconstructed in a wavelet multiscale fashion, i.e., χ_j of equation 17; (3) the ED at full resolution, i.e., χ_1 of equation 9, or (4) the difference of the envelopes of the seismograms reconstructed via wavelet multiscale analysis (i.e., χ_j^1 of equation 18).

With reference to Figure 1 and again using SPECFEM2D, we calculate the relevant misfit values and plot them as contoured 2D surfaces in the variables V_{S0} and V_{S1}, the background S-wave speed and the perturbed S-wave speed in the curved layer, respectively. The true values are $V_{S0} = 800$ m/s and $V_{S1} = 1000$ m/s and the calculations range from 600 to 1400 m/s for both of them. To evaluate the misfits, we use one shot gather of a 40 Hz Ricker wavelet source imposed normally to the surface, located at 50 m horizontally from the left edge and at 0.5 m in depth. We consider the vertical and the horizontal components of displacement recorded by a total of 401 receivers spaced 1 m apart and at a depth of 0.5 m below the surface. The total misfit is the sum of the misfits for each component.

The full-resolution ("D12 scale 0") WD misfit surface is shown in Figure 2a, and the full-resolution ED misfit surface is shown in Figure 2b. The WD misfit displays numerous local minima caused by the cycle skipping of SWs, which ultimately will prevent inversions from converging to the target solution (which lies at the intersection of the white lines) when starting from a homogeneous model ($V_{S0} = V_{S1} = 900$ m/s,

denoted by the filled red circle). On the other hand, the ED misfit is characterized by a relatively wide basin from which convergence to the target solution can be expected. Figure 2c and 2d shows profiles through the misfit surfaces taken along the white lines plotted in Figure 2a and 2b, at constant $V_{S1} = 1000$ m/s, in Figure 2c, and at constant $V_{S0} = 800$ m/s, in Figure 2d. Making ED measurements results in a much more navigable misfit valley, as is clearly seen.

For this same experiment, Figure 3a and 3b shows the misfit kernels, the gradients of the WD (Figure 3a) and the ED (Figure 3b) misfit functions with respect to the S-wave speed, computed at $V_{S0} = V_{S1} = 900$ m/s. To evaluate the misfit kernels, we use the same receiver geometry as for the misfit surface calculations, but for a total of 39 sources with 40 Hz wavelet imposed normally to the surface, located between 10 and 390 m horizontally from the left edge with 10 m equal spacing and at a depth of 0.5 m. For all the numerical experiments, Gaussian smoothing operators were applied to suppress numerical artifacts and spurious energy in the vicinity of sources and receivers (see also Favier et al., 2004; Tian et al., 2007). Figure 3c and 3d shows models arrived at by line searching in the direction opposite to the gradients shown in the top row. The results of one iteration in the adjoint scheme, from a starting point marked by the red filled circles in Figure 2a and 2b, the solutions shown are marked by the white filled circles in the top panels of Figure 2a and 2b, and connected to the starting model by a white line. Neither by the WD metric nor by the ED metric do we reach the global minimum of the target solution after just one iteration, but using the ED measurement leads to a reasonably well-defined anomaly — but still with the wrong sign. Nevertheless, the first-iteration ED model
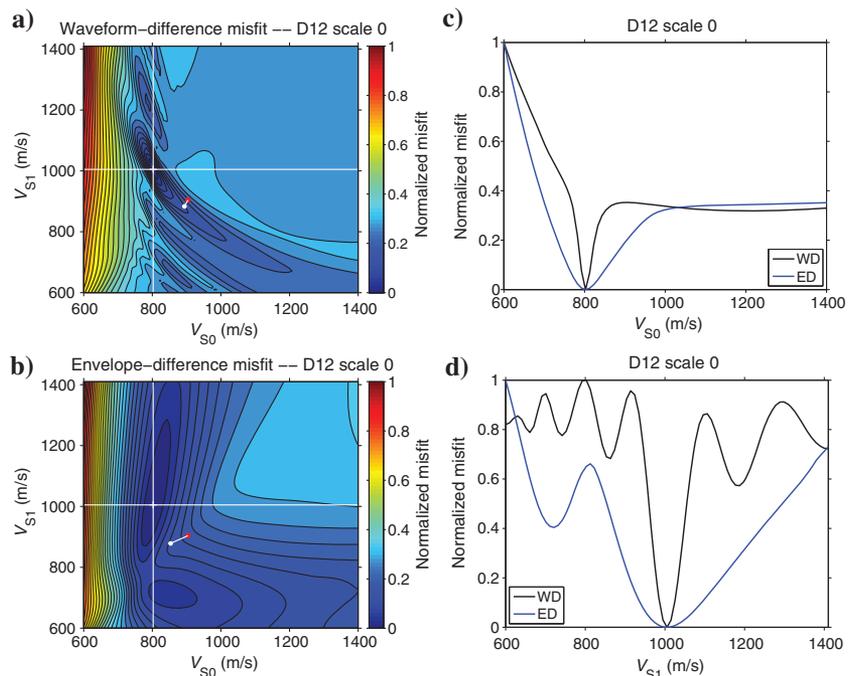
Figure 2. Misfit surfaces (normalized by their maxima) of (a) WD and (b) ED measurements made on the toy model in Figure 1, at the full resolution (scale 0), with respect to the background (V_{S0}) and the anomalous (V_{S1}) S-wave speeds. Profiles through the misfit surfaces taken (c) at constant $V_{S1} = 1000$ m/s and (d) at constant $V_{S0} = 800$ m/s.

in Figure 3d is a more promising approximation to the target shown in Figure 1a than the WD model shown in Figure 3c.

The misfit surface of the WD measured at scale 8 (Figure 4a) is devoid of the many secondary minima that were present at the full resolution, at scale 0 (Figure 2a), hinting at an effective remediation of cycle-skipping effects by coarse-scale wavelet approximation of the seismograms. Such a finding is consistent with the BW inversion experiments of Yuan and Simons (2014), who prove that constructive approximation by wavelet multiscale analysis reduces the number of local minima even when poor initial models are taken as a starting reference. However, unlike with BW waveforms, WD measurements made on the coarsest-scale wavelet representation of SW, when used for gradient computations, still insufficiently reduce the distance between the modeled and the predicted SW waveforms in the inversion. It is not difficult to encounter examples in which convergence to the global optimum remains out of reach due to the severe nonlinearity of SW inversion caused by cycle skipping of these dispersive and highly oscillatory phases.

By ignoring phase variations, ED measurements made at the coarsest level of a multiscale analysis usefully reduce the nonlinearity of SW inversions. Figure 4b shows that the ED misfit
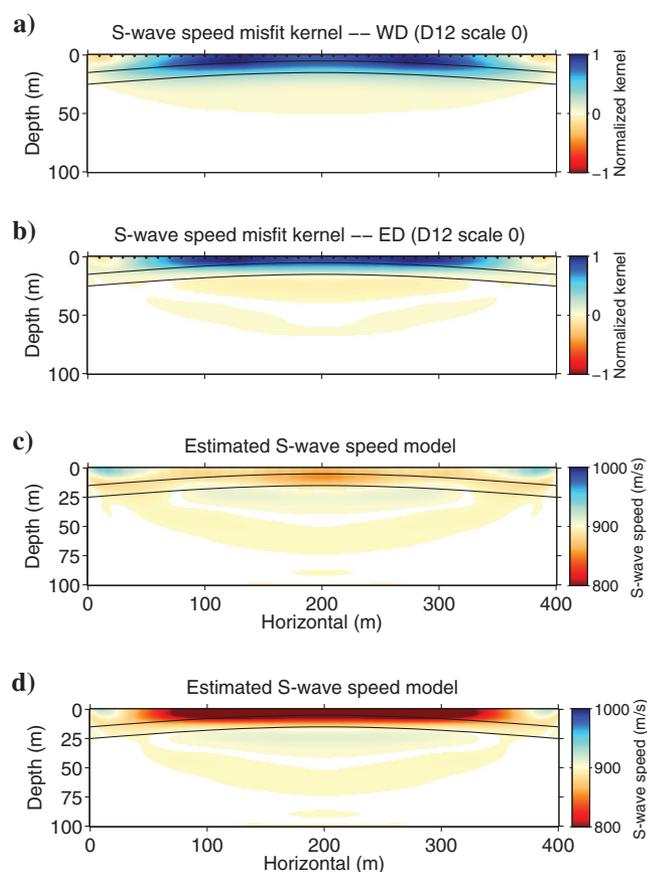
Figure 3. Misfit kernels (gradients with respect to S-wave speed perturbations) for (a) WD and (b) ED measurements made in the toy model of Figure 1a, at full resolution (scale 0), calculated in a homogeneous background model with S-wave speed 900 m/s. Models obtained after one iteration in the direction opposite to the gradient using (c) waveforms and (d) envelopes of the waveforms at full resolution.

function at scale 8 has a wider convergence basin than the corresponding WD misfit at the same scale shown in Figure 4a. The update path after one iteration away from the homogeneous starting model of $V_S = 900$ m/s points almost directly at the target solution. The profiles through the misfit function shown in Figure 4c and 4d also show that coarse-resolution projection via wavelet analysis tends to widen the convergence troughs, compared with the profiles at full resolution shown in Figure 2c and 2d. Furthermore, comparison of the ED misfit profiles (blue lines) with the WD profiles (black lines) in Figures 2 and 4 illustrates the general widening of the minimization surfaces for envelopes compared with waveforms.

Figure 5a and 5b shows the misfit kernels for the WD (Figure 5a) and ED (Figure 5b) SW waveform measurements made at scale 8. Figure 5c and 5d displays the models updated using the kernels as the corresponding gradients, after one iteration. One clearly sees how the multiscale WD measurements successfully correct the background S-wave speeds but, thus far, fail to distinguish the layer of anomalous velocity (Figure 5c). On the other hand, multiscale ED measurements yield single-iteration updates that well resolve the target strip of fast S-wave speeds (Figure 5d), while illuminating the slower S-wave speeds above and below the fast layer (whose upper and lower boundaries are drawn with black lines).

As far as a toy model can be our guide, the above experiments lead us to conclude that ED measurements made within a wavelet-multiscale analysis will greatly reduce the nonlinearity of SW tomography, and reduce the critical dependence of waveform inversions on the quality of the initial estimates. In the following section, we will elaborate on this initial strategy and develop it further to also include BWs in an adjoint-based elastic and truly full-waveform inversion.

A STRATEGY FOR BODY- AND SURFACE-WAVE INVERSION

In the previous section, we used a toy model and only considered variations in V_S to introduce alternative misfit measures dedicated to the incorporation of SW in full-waveform inversion. The use of a multiscale representation and the application of the ED measurement lessen the nonlinearity of SW inversion problems.

In this section, we switch to more realistic synthetic experiments, and we consider perturbations (relative to the evolving earth model) in ρ, V_S and V_P, and we use BWs as well as SWs. We subdivide the section into two parts: one that considers the density known and one where density is also inverted for (see Appendix A for the expressions of the kernels used in either case). In the first part, assuming the exact density is known, we consider SWs and BWs separately first, and then simultaneously. In the second part, we use an incorrect starting density structure and discuss elastic inversions with and without density updates. Our ultimate goal is to formulate a flexible procedure to solve for elastic parameters and density using BWs and SWs, and this section will demonstrate considerable progress in our ability to do so.

Figure 6 shows the target model (modified after Pérez Solano et al., 2014) for density and the elastic P- and S-wave speeds that we will use for the numerical experiments in this section. With this model, we use 73 and 14 uniform mesh nodes in the horizontal and the vertical directions, respectively; thus, there are a total of $(73 \times 4 + 1) \times (14 \times 4 + 1) = 16{,}701$ unique grid points, as well

as $73 \times 14 \times 25 = 25{,}550$ P- and S-wave speeds, as well as density parameters to characterize the model. The quadrangle mean spacing is 10 m.

We place 42 shots (black circles in Figure 6a) equally spaced with 15-m separation, between the horizontal range of 10–625 m at a depth 0.5 m below the model surface. The source is a vertical elastic force with a dominant 10-Hz Ricker wavelet. In total, 106 receivers are equally distributed between 3 and 633 m, at the depth of the source. In this 2D model, we calculated the horizontal and vertical components of the wavefield using the spectral-element code SPECFEM2D implemented on a uniform mesh. The sampling interval is 600 μs, and the record length is 2.1 s, which is sufficient to include SW trains at the largest offsets.

Inversions with known density

In the first synthetic inversion test, we assume an initial model that has the exact density information, a homogeneous $V_P = 1000$ m/s, and a homogeneous $V_S = 500$ m/s. The inversion is for the 2D P- and S-wave speeds. We isolated SW in the time-space domain on both components of the data, between the group velocities of 500 and 300 m/s.

Choice of misfit function for surface-wave inversions

We tested three different approaches to measuring misfit: the multiscale WD scheme introduced by Yuan and Simons (2014), the multiscale ED metric introduced in this paper, and a hybrid multiscale envelope-waveform strategy that we will be discussing further down.

Figure 7 shows the inverted P- (Figure 7a) and S-wave (Figure 7b) speed models obtained by iteratively minimizing the WD misfit of SWs defined in a multiresolution framework (via our equation 17). Replicating the procedure and settings recommended by Yuan and Simons (2014), we use Daubechies (1988) "12-tap" wavelets (D12) to decompose the windowed SWs up to the maximum wavelet level (scale 8), within which iterative WD inversions are performed, before scaling down to scales 7, 6, and 5 (when, for our sampling rate, the waveforms are almost identical to the full-resolution versions, scale 0). As Figure 7 shows, neither the P- nor the S-wave speed model converges to a model near the target shown in Figure 6. The multiscale WD inversion of Yuan and Simons (2014) was not designed, and as the example amply illustrates is generally unsuited, for the analysis of SWs.

To combat the nonlinearities of SW inversion due to cycle skipping, we next replace waveforms with their envelopes in the objective function: In other words, we minimize the discrepancy defined via the envelope criterion in equation 18. It bears repeating that we are making wavelet-multiresolution constructive approximations to the seismograms, and then calculating their envelopes, and not the other way around (working with wavelet-multiresolution constructive approximations to the envelopes of the full-resolution seismograms).

We start the inversions of the ED between SWs predicted in the initial model and modeled in the exact model at scale 8, and progressively involve envelopes of smaller-scale seismograms, for descending scale levels, until the ED of full-resolution SW waveforms is being minimized at scale 0.

Figure 8 shows the V_P and V_S models that result. Working only with the envelopes, but in a multiresolution framework, leads to reasonably successful recovery of the V_S model. The S-wave velocity structure is resolved albeit with a low resolution, and only the very top portion of the P-wave speed model is showing signs of being updated in the desirable direction. Multiscale ED inversions appear to converge at the price of losing resolution by discarding phase information.

Last, we propose a hybrid, envelope-then-waveform method for SWs in the same wavelet multiresolution framework. We first minimize the ED metric at a certain high scale (here, scale 8), and then we switch to the WD metric. We repeat this envelope-then-waveform optimization at each descending scale until the envelopes and phased waveforms of SWs are explained at the full resolution of the acquired data.

Figure 9 shows the results of adopting the hybrid philosophy. Most of the S-wave velocity structure is well resolved; however, only the very top portion of the P-wave speed model is acceptably imaged. These results are consistent with the theoretical expectation (e.g., from SW mode summation kernels, Simons and van der Hilst, 2003) that SWs should be most sensitive to S-wave speed structure. Comparing the hybrid inversion result in Figure 9 with that of Figure 8, which involved solely envelopes, it is seen that the inclusion
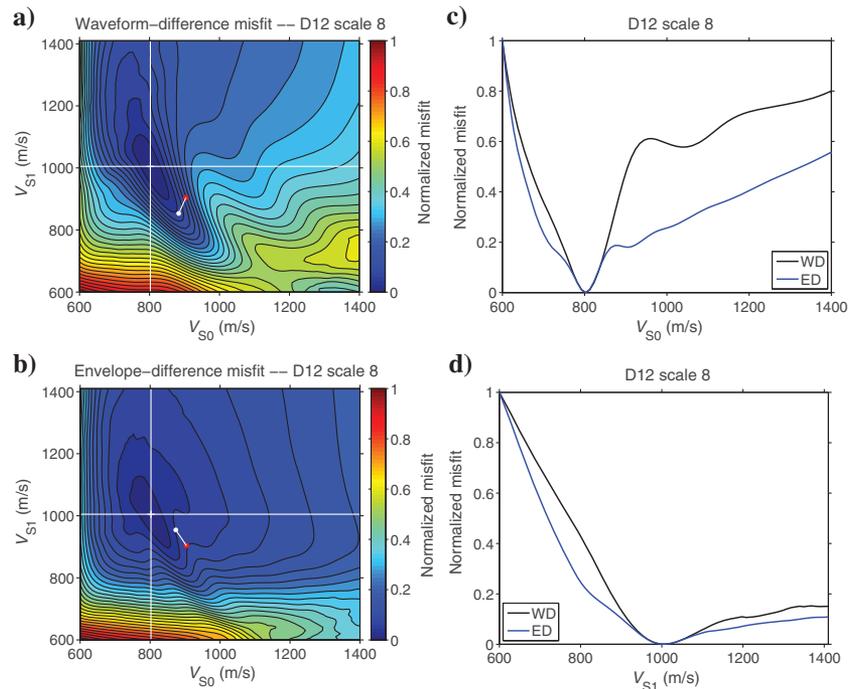
Figure 4. Normalized misfit surfaces (as in Figure 2) of (a) WD and (b) ED measurements made on the toy model in Figure 1, at the maximum D12 wavelet decomposition depth (scale 8), with respect to the background (V_{S0}) and the anomalous (V_{S1}) S-wave speeds. Profiles through the misfit surfaces taken (c) at constant $V_{S1} = 1000$ m/s and (d) at constant $V_{S0} = 800$ m/s.

a)

S-wave speed misfit kernel –– WD (D12 scale 8)

b)

S-wave speed misfit kernel –– ED (D12 scale 8)

c)

Estimated S-wave speed model

d)

Estimated S-wave speed model
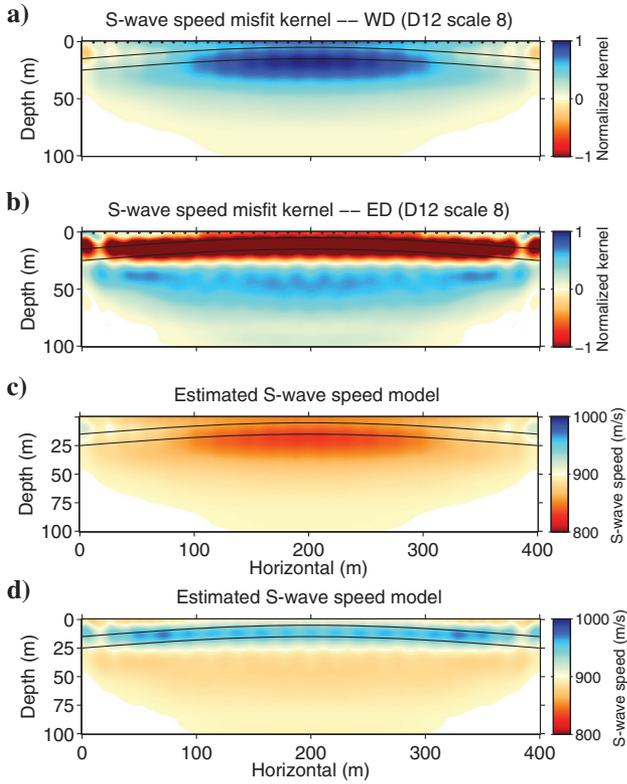
Figure 5. Misfit kernels (as in Figure 3) for (a) WD and (b) ED measurements made in the toy model of Figure 1, at the maximum D12 wavelet decomposition depth (scale 8), calculated in a homogeneous background model with S-wave speed 900 m/s. Models obtained after one iteration in the direction opposite to the gradient using (c) waveforms and (d) envelopes of the waveforms at this coarsest scale.

a)

Density: target model

b)

P-wave speed: target model

c)

S-wave speed: target model
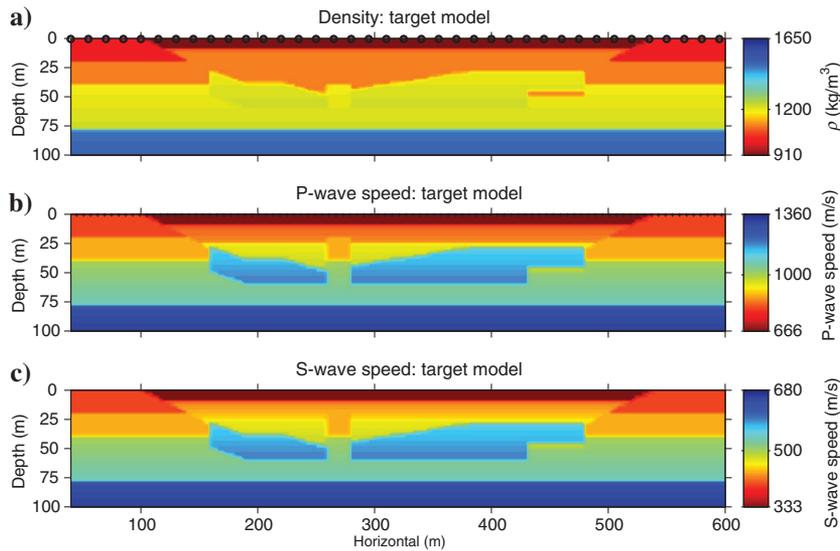
Figure 6. The modified Pérez Solano et al. (2014) synthetic test model, specifying (a) density, (b) P-, and (c) S-wave speeds. The black circles in the top panel show the sources used in our inversion. The black crosses in the middle panel are receiver locations, at the same depth as the sources.

of phase information yields a higher resolution S-wave speed model and provides additional structure to the P-wave speed model.

Comparison of body- and surface-wave inversions

The results in this section, until now, have been obtained using SWs exclusively. To understand more fully the relative sensitivities of SWs and BWs in adjoint inversions, we perform another elastic inversion test, using only BWs. We follow the multiscale WD method introduced by Yuan and Simons (2014), removing SWs by low-pass and dip filtering. Once again, we use the D12 wavelet basis up to scale 8, and we iteratively minimize the WD misfit obtained at each level, successively involving finer scales until convergence at scale 0. The S- and P-wave velocity models, thus obtained are shown in Figure 10.

Compared with the envelope-waveform hybrid but SW only inversion result shown in Figure 9, the BW-only tomography in Figure 10 is characterized by slightly inferior resolution as far as the S-wave speed is concerned, but with somewhat better resolution in terms of the P-wave speeds. These results are to be understood on the basis of BW sensitivities that take into consideration radiation-pattern, scattering, and mode-conversion effects (e.g., Prieux et al., 2013b). We conduct a quantitative analysis of the relative merits of the models by calculating the root-mean-squared (rms) error between the estimated model and the known target. The error is computed at all model nodes in a 10-m depth interval and is assigned to the middle of each depth range.

Figure 11 shows this rms error, normalized by the rms value of the target model. From left to right, we compare results for V_P, V_S, and their combination, and within each panel, we plot the rms error for the initial model (green lines), and then for the model of Figure 9, obtained from the hybrid-misfit SW procedure (red lines), and for the model of Figure 10 obtained from waveform-misfit BWs only (blue lines). For V_P, SWs resolve the upper half of the model domain very well, compared with the deeper structure, whereas BWs have more uniformly high-quality resolution throughout depth, but do not to recover the near-surface structure as well as SWs. For V_S, SWs and BWs lead to well-resolved models, with the SW inversion slightly superior near the top and bottom, as well as in the fast anomalies in the middle of the model domain. It is to be understood that targeted, high-frequency, possibly acoustic, full-waveform inversion approaches might recover V_P models of a far higher quality (in an absolute sense), but in this experiment, we only test the relative merits of BWs versus SWs over the same range of (altogether low) frequencies.

In Figures 12–15, we show how well the waveform data (with and without SW) are fit under SW and BW tomography as described in this paper. The top rows show the horizontal (x) displacement components, and the bottom rows show the vertical (z) components for a selection of traces from one shot gather. From left to right, we display the seismograms reconstructed at scales 8, 7, and 0 (full resolution) in the wavelet basis D12. The thick black lines represent the target data, thin blue lines represent the initial predictions, and thick gray lines re-

present the final models, by which we mean the model shown in Figure 9 for Figures 12 and 14, and the model in Figure 10 for Figures 13 and 15. In each of these cases, when referring to the *final model*, we mean the one obtained at the end of the fitting procedure, after all scales have been fit, but considering seismograms only reconstructed up to the relevant scale. Below each trace, we list the rms error of the initial and final fits, as a percentage of the rms of the data. For example, at scale 8 in Figure 12d, trace number 101 has an initial misfit of 91%, which gets reduced to 18% after the inversion, and thus the trace receives the annotation "91 → 18." To get a sense of the relative amplitude of the SWs compared with the BWs in the signals shown at each scale level, and between scales, the amplifications of the traces in Figures 12–15 are listed in the top left of every panel. For example, the true amplitudes in Figure 12c are 23/2.5 times the size of those in Figure 12a, and those of Figure 14c are 8/23 times the size of those in Figure 12c.

In Figures 12 and 13, we show the complete seismograms, containing also BW but of course dominated in amplitude by the SW. SW tomography (Figure 12) using SWs windowed in the time-space domain between 300 and 500 m/s group-velocity slopes (dashed lines in Figures 12 and 13) correctly fits most of the features of all traces. Using BW, tomography (Figure 13) leaves some scope for improvement in the fit to the SWs.

In Figures 14 and 15, we remove the SWs using low pass and dip filtering for a clean look at the BWs and how they are fit. Thus, we show the BW seismograms processed via SW removal. Figure 14 illustrates that when SWs are used in the adjoint inversion, there is still much room to improve the fit to the BWs, even though the SWs may have been matched to almost-perfect agreement with the target traces, as was shown by Figure 12. Conversely, Figure 15 shows that by involving only BWs in the inversion, we are able to match the BWs reasonably well, but the SW fits to continue to show inaccurate matches to the target traces, as was seen in Figure 13.

Combining body and surface waves

From the comparison of separate BW and SW inversions, we see how BWs sample deeper structure, whereas SWs uncover a wealth of high-resolution near-surface structure that conventional BW tomography fails to resolve. With both wave types, the wavelet-multiscale approach provides the gentle guidance for the adjoint method to converge toward global minima. We thus propose to combine these approaches. We first extract SWs, and, to the scale of the
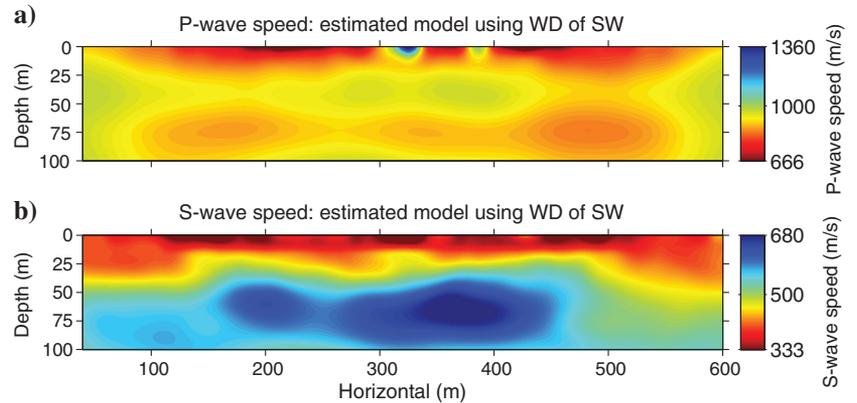
Figure 7. (a) P- and (b) S-wave speed models (after 36 iterations) that result from minimizing multiscale WD measures of windowed SWs, starting from a homogeneous $V_P = 1000$ m/s and $V_S = 500$ m/s and the correct density model. The inversion has not converged, and is not likely to converge, to the correct target solution shown in Figure 6.
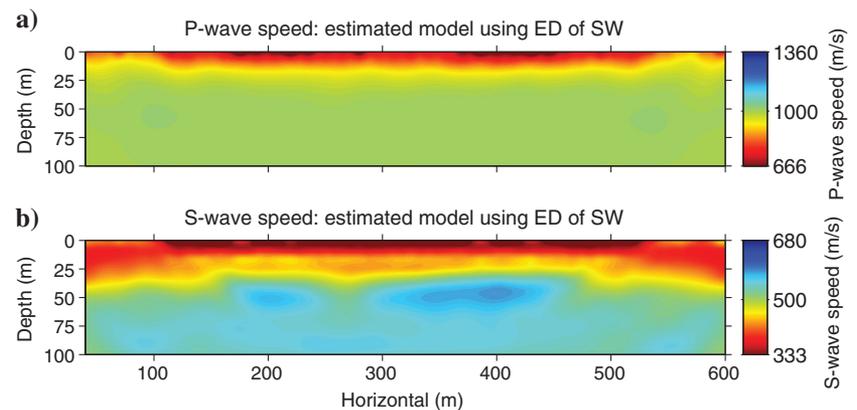
Figure 8. (a) P- and (b) S-wave speed models (after 41 iterations) that result from minimizing multiscale ED measures of windowed SW, starting from a homogeneous $V_P = 1000$ m/s and $V_S = 500$ m/s and the correct density model. Only the top part of the P-wave speed model shows updates in the desirable direction. The resulting S-wave speed model resolves most of the target structure, but the tomographic resolution is not very high.
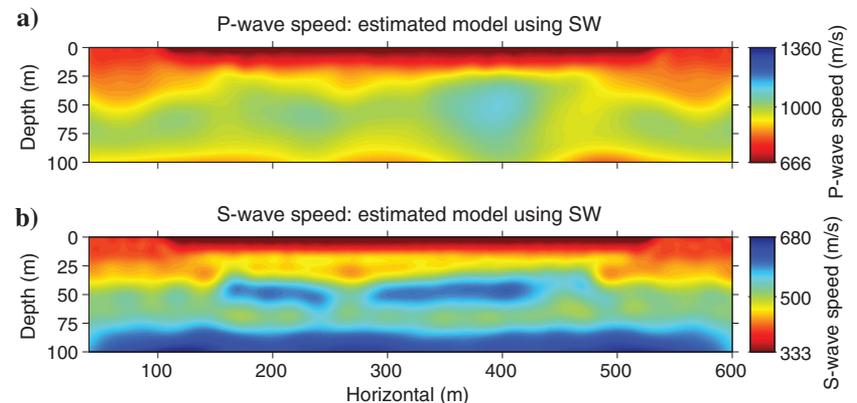
Figure 9. (a) P- and (b) S-wave speed models (after 74 iterations) resulting from the hybrid multiscale envelope-then-waveform inversion of the SWs, starting from a homogeneous $V_P = 1000$ m/s and $V_S = 500$ m/s, and the correct density model. The S-wave speed model is well resolved, but the deep structure in the P-wave speed model remains smeared.

wavelet subspace, we use them in hybrid envelope-waveform inversions to constrain the primarily near-surface structure. After the SWs have been well explained to this scale, we add the BWs back into the seismograms, and we use the WD inversion to illuminate the deeper structure. The prior alignment of the SW traces linearizes the problem to the point where WD inversions function without the need for a clear separation nor any identification of BWs and SWs. The outer loop is over the wavelet scales, from coarse to fine.

Figure 16 shows the final inverted V_P and V_S models obtained in this way, which indeed retrieve most of the features in the target model (compare with Figure 6).

Inversions with unknown density

In our numerical experiments so far, we have used the exact density model in our inversions. In real applications, however, we usually do not have correct density information. In this section, we will take into consideration the influence of inadequate density structure

on elastic inversions. For this purpose, we use a homogeneous P-wave speed $V_P = 1000$ m/s, a homogeneous S-wave speed $V_S = 500$ m/s, and a homogeneous density $\rho = 1200$ kg/m³ as our starting model for an inversion that ultimately also solves for the density perturbation.

In the first scenario, we use the approach discussed above to update V_S and V_P using surface, and then BWs and SWs, iteratively implementing the hybrid multiscale envelope-waveform scheme. The final result is shown in the top three panels of Figure 17. Because density is not updated, the final density model is still the homogeneous initial model. Comparing these results with Figure 16, we observe the detrimental effects on the wave speed models of incorrect density information. The P-wave speed model is more affected than the S-wave velocities.

In the second scenario, we also consider density corrections in the inversion. For successive approximations to each wavelet scale, we first perform hybrid envelope-waveform optimizations using windowed SWs, then waveform optimizations using BWs and SWs.

In each iteration step, we simultaneously update density ρ, P-wave speed V_P, and S-wave speed V_S, using the horizontal and vertical components of the seismograms. The bottom three panels in Figure 17 show the final density and P- and S-wave speed models obtained under this scenario. The rms error comparisons of these results with the first scenario (homogeneous fixed density structure), and with the previous experiment (starting from the exact density model, leading to the results shown in Figure 16), are shown in Figure 18. By simultaneously updating ρ-V_P-V_S (thick solid black lines), V_P and V_S have been considerably improved compared with the inversion results that held the density fixed (thick dashed gray lines). The improvements are especially clear in the top and bottom portions of the V_P model (Figure 18a), and the top and the middle segments of the V_S model (Figure 18b). The misfit for the ρ model is found in Figure 18c.
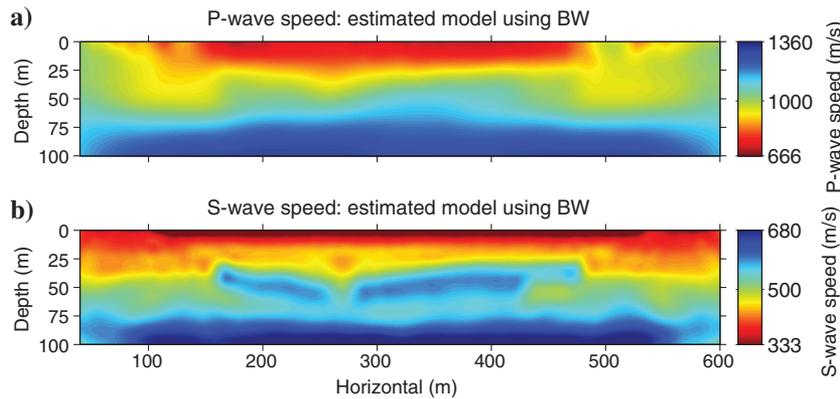
Figure 10. (a) P- and (b) S-wave speed models (after 31 iterations) resulting from multiscale WD inversion of the BWs, from a homogeneous $V_P = 1000$ m/s and $V_S = 500$ m/s starting model and the correct density. Compared with Figure 9, there is an improvement in V_P speed resolution but a slight degradation of the quality of the V_S model.
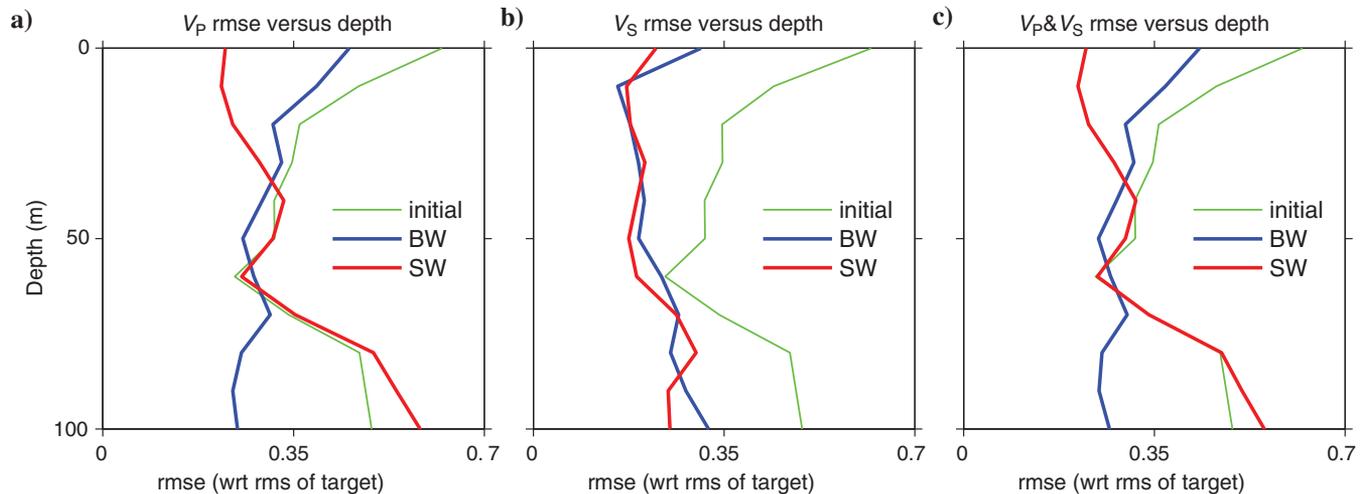
Figure 11. Normalized rms error between the homogeneous initial model, the inversion results of Figures 9 and 10, and the target model of Figure 6. From left to right are shown the rms error in terms of V_P, V_S, and their combination. The green lines stand for the initial model, blue lines represent the model obtained using BWs only, and red lines denote the model obtained from only SWs.

Even though the exact density structure is not retrieved (noting that progress toward this goal would require additional constraints, e.g., from gravity data), by simultaneously updating ρ-V_P-V_S starting from incorrect density information, we obtain V_P-V_S results that are comparable with those obtained from inversions that use the exact density information.

SUMMARY OF PROCEDURE

Summarizing the above discussion and our synthetic experiments, we now formalize a flexible scheme for full-waveform inversion. First, SWs are analyzed in a wavelet multiresolution framework. For successive approximations at a certain scale, thus for partially reconstructed, constructively approximated SW waveforms, we separate phase and amplitude information. That is, we attempt to fit envelope (amplitude) information first, and after most of the arrivals appear to be in phase (within half of the wavelength), we add phase information back into the problem. After that, we combine BWs and SWs together in the full seismogram, and iteratively minimize the WD in the subspace. If density information is

unknown, we perform a density inversion together with the wave-speed inversions in each scale. Subsequently, we progress through the scales until the last inversions are run at the full resolution of the data collected.

1) Based on the synthetics computed using SPECFEM2D for the source-receiver geometry in the initial model, and given the target seismograms, we choose the multiscale analysis parameters, i.e., the wavelet family and number of vanishing moments and the maximum decomposition depth J.
2) Iterative optimization to current scale j ($j = J$ if progressing from step 1, $j = j - 1$ if arriving from step 3). We first run the SW tomography, and subsequently combine BWs and SWs to simultaneously update wave speeds and density structure. In our inversions, we do not separate the P- and S-wave contributions to the seismograms. All components and phases are used together to explain discrepancies in all model parameters.

   a) For the current model $\mathbf{m}(\mathbf{x})$, we calculate the synthetics $s(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ using the spectral-element code.
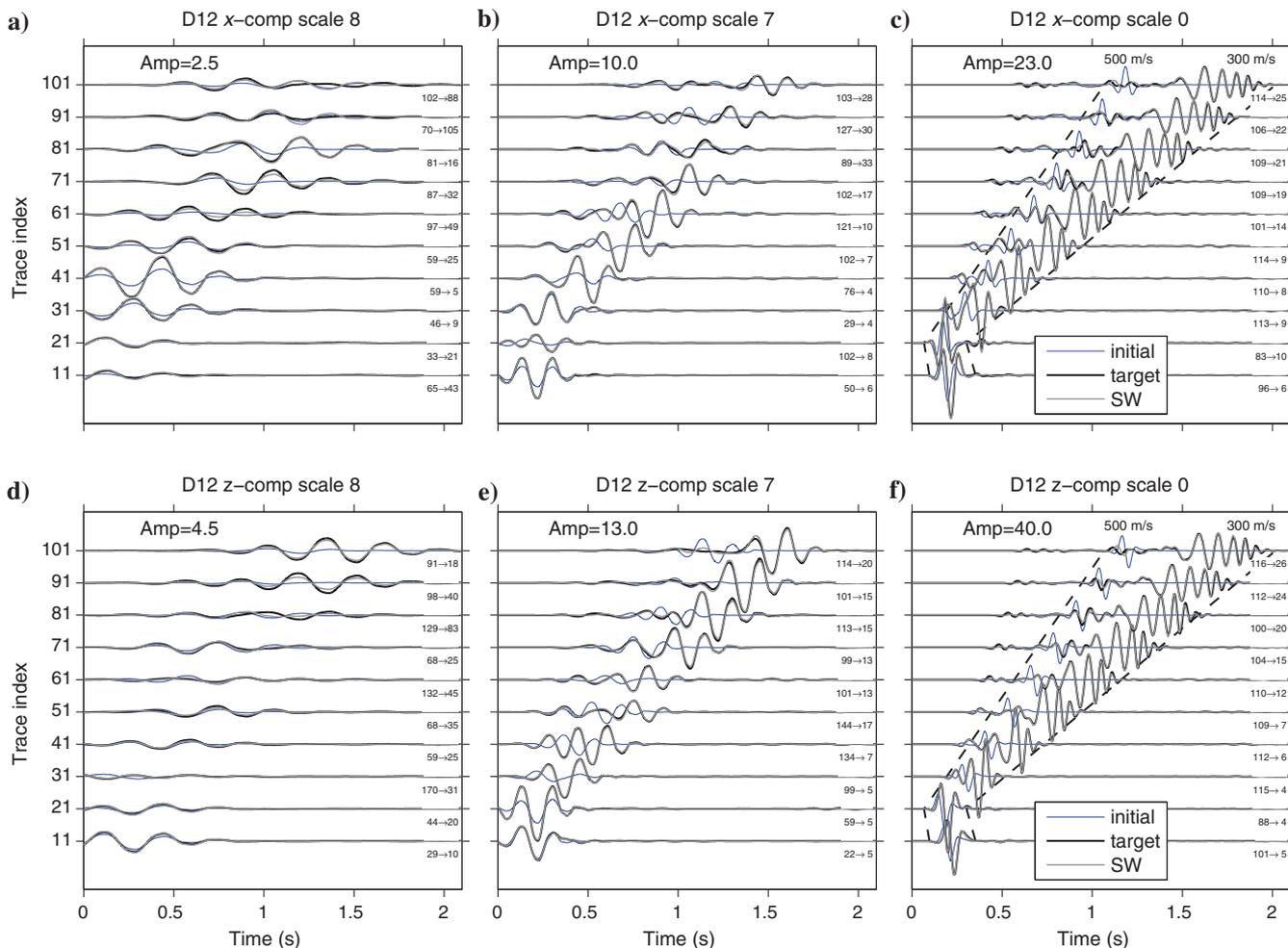
Figure 12. Whole seismograms (including BWs and SWs) shown at (a and d) scale 8, (b and e) scale 7, and (c and f) the full-resolution scale 0. (Top) Horizontal components and (Bottom) vertical components. Only SW within the group-velocity windows (dashed) were used for model updates. Thick black lines are the target seismograms, thin blue lines the initial ones, and the thick gray lines labeled SW are seismograms computed in the final model shown in Figure 9.

b) We window all components of the predicted $s(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and the observed $d(\mathbf{x}_r, \mathbf{x}_s, t)$ seismograms using group-velocity bounds in the time-space domain to yield SW waveforms $\hat{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and $\hat{d}(\mathbf{x}_r, \mathbf{x}_s, t)$.

c) We apply the wavelet transform to the windowed SWs, and we partially reconstruct predicted and target SW waveforms to scale j, to obtain $\hat{s}_j(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and $\hat{d}_j(\mathbf{x}_r, \mathbf{x}_s, t)$.

d) We take the Hilbert transform of the partially reconstructed SW to extract their envelopes $E_{\hat{s}_j}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ and $E_{\hat{d}_j}(\mathbf{x}_r, \mathbf{x}_s, t)$, and we choose the ED as our objective function to be minimized:

$$\chi_j^1(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int_0^T \| E_{\hat{s}_j} - E_{\hat{d}_j} \|^2 dt. \qquad (20)$$

e) We carry out the adjoint simulations and calculate the interaction of the forward and adjoint wavefields to obtain the numerical kernel or gradient $\mathbf{g}$ associated with the misfit function $\chi_j^1(\mathbf{m})$.

f) We define our search direction by the conjugate gradient method (the Hessian is neither computed nor approximated). For the kth iteration, we have $\mathbf{p}^k = -\mathbf{g}^k + \beta^k \mathbf{p}^{k-1}$, where $\beta^k = \mathbf{g}^k \cdot (\mathbf{g}^k - \mathbf{g}^{k-1})/(\mathbf{g}^{k-1} \cdot \mathbf{g}^{k-1})$ for $k > 1$, and $\beta = 0$ if $k = 1$ or when the calculated β is negative.

g) We find the "optimal" step length ν^k to minimize the misfit value of $\chi_j^1(\mathbf{m} + \nu^k \mathbf{p}^k)$ with a line search method by generating a limited number of trial step lengths until we find one that approximates the minimum of the misfit function $\chi_j^1(\mathbf{m})$ in the search direction of $\mathbf{p}^k$ defined in step (2f). More sophisticated line search algorithms can be substituted (e.g., Nocedal and Wright, 2006).

h) We update current model $\mathbf{m}^k$ using the step length ν^k and the search direction $\mathbf{p}^k$ to get a new model $\mathbf{m}^{k+1}$ for the next iteration: $\mathbf{m}^{k+1} = \mathbf{m}^k + \nu^k \mathbf{p}^k$, and return to step (2a) for optimization starting from current model $\mathbf{m} = \mathbf{m}^{k+1}$.

i) The loop from step (2a) to step (2h) is repeated by iteratively updating V_P, V_S (and ρ) until the defined SW ED misfit is minimized at current scale j. After that, we add phase infor-
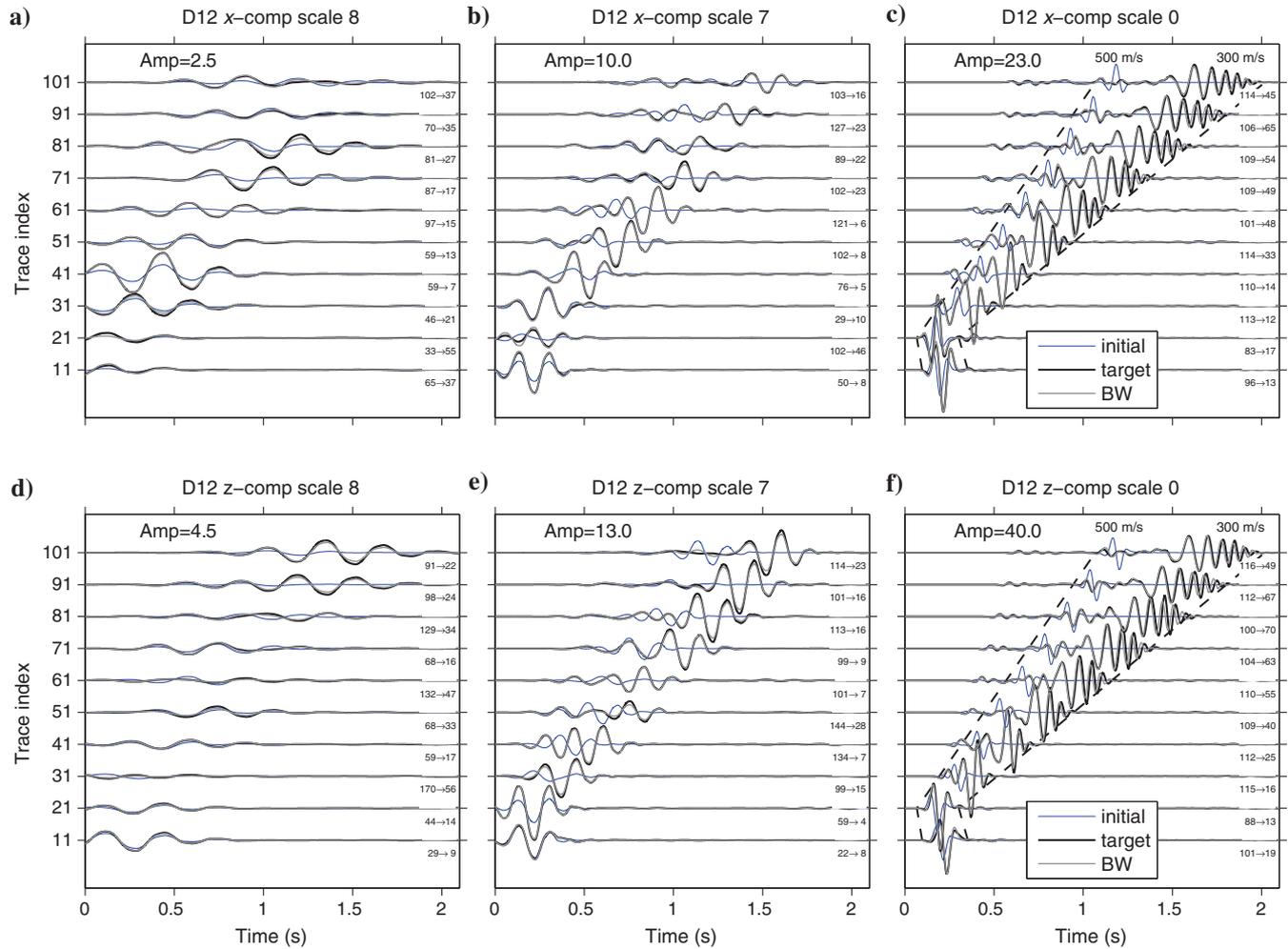
Figure 13. Whole seismograms at scales 8, 7, and 0. Layout is as in Figure 12. Only BWs were used to make model updates. The seismograms labeled BW, drawn by thick gray lines, were calculated in the final model shown in Figure 10. There remains significant room for improvement.

mation and perform WD inversions as Yuan and Simons (2014) do, for the SW. We iteratively minimize

$$\chi_j^2(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int_0^T \|\hat{s}_j - \hat{d}_j\|^2 dt. \qquad (21)$$

j) After the SW WD misfit $\chi_j^2(\mathbf{m})$ has converged, we combine BWs and SWs to perform WD inversions, as Yuan and Simons (2014) do, but now for the entire seismograms. At the current scale j, we minimize

$$\chi_j^3(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int_0^T \|s_j - d_j\|^2 dt. \qquad (22)$$

k) We simultaneously update V_P, V_S (and ρ) to fit the BW and SW waveforms until convergence.

3) When the procedure at scale j has converged, we add smaller-scale information to the seismogram by switching to the next,

lower, scale level of reconstruction, $j = j - 1$, and we return to step (2).

4) We loop over all defined scale levels from step (2) to step (3) until convergence at the nominal full resolution $j = 0$.

DISCUSSION

Many published techniques have addressed the problem of non-linearity in full-waveform inversion. The key to success, especially when SWs are being considered also, is the progressive widening of the frequency band of the seismic records under consideration. Traditionally, such approaches have relied on conventional filtering. In this paper, as in our previous contribution which focused exclusively on the analysis of BWs (Yuan and Simons, 2014), we have given the notion of "multiscale analysis," the strict interpretation of a wavelet-based (Daubechies, 1992), multiresolution (Mallat, 2008) decomposition of the time-domain seismograms, as a means to steer the characterization of subsurface velocity and density structure from the coarser to the smaller spatial scales.
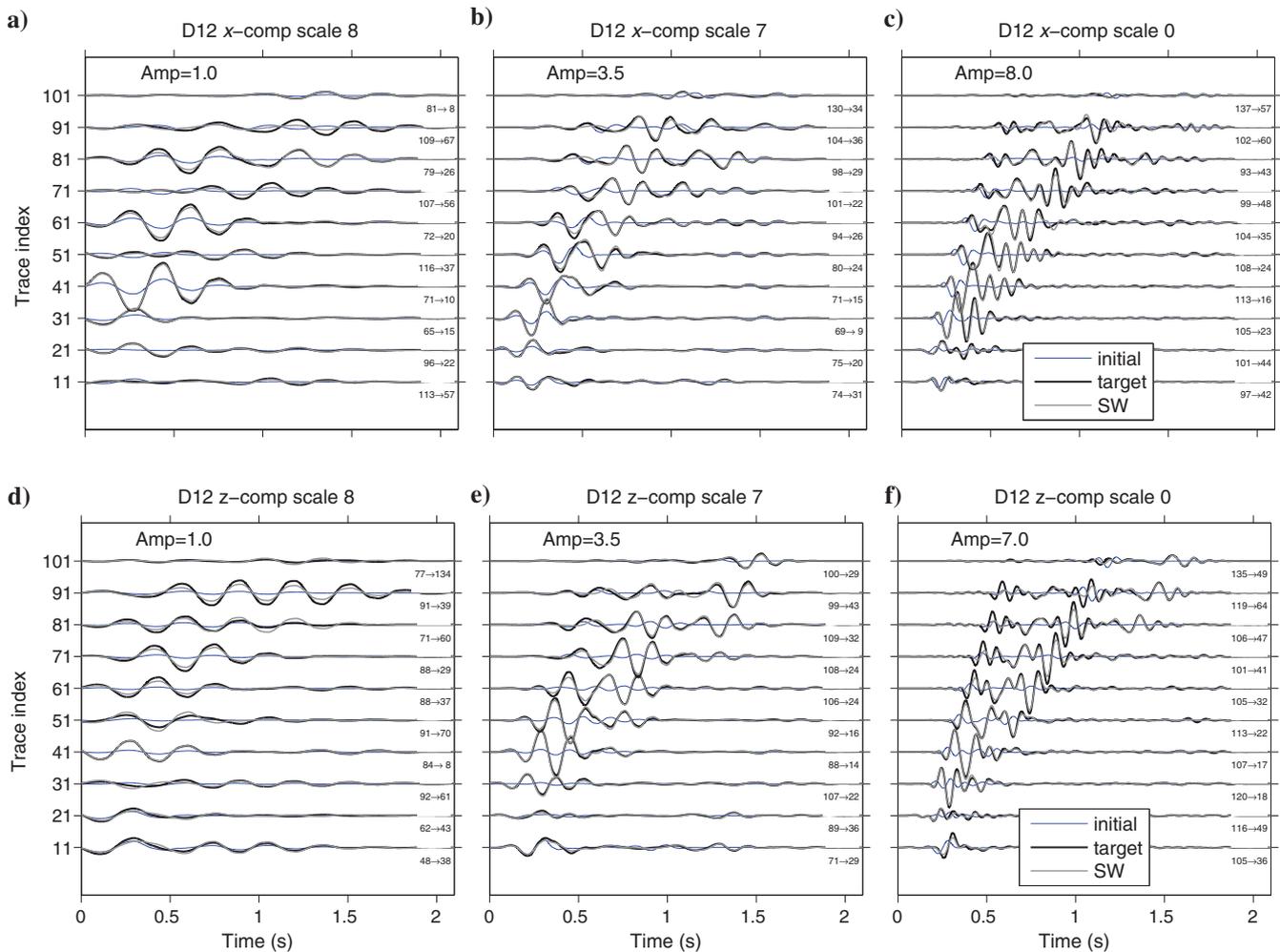
Figure 14. BW seismograms (SWs removed) at scales 8, 7, and 0, before and after fitting by the hybrid wavelet multiscale adjoint SW tomography method as described in the text. Thick gray seismograms labeled SW are computed in the final model shown in Figure 9. There remains significant room for improvement.

The wavelet-based analysis of seismic signals is theoretically attractive and has been developing ever since the seminal work by Morlet et al. (1982a, 1982b) in this very journal. Similarly, adjoint-based tomography methods have a long history, pioneered by Tarantola (1984a, 1984b), also in this journal. However, the idea of using successive wavelet-based approximations as part of an adjoint-based full-waveform inversion scheme had not been worked out in the detail that it requires, until now. As could perhaps be expected, but is now convincingly shown, issues with cycle skipping that complicate seismic inversions, especially in multiparameter elastic inversions using multicomponent data, can be successfully addressed.
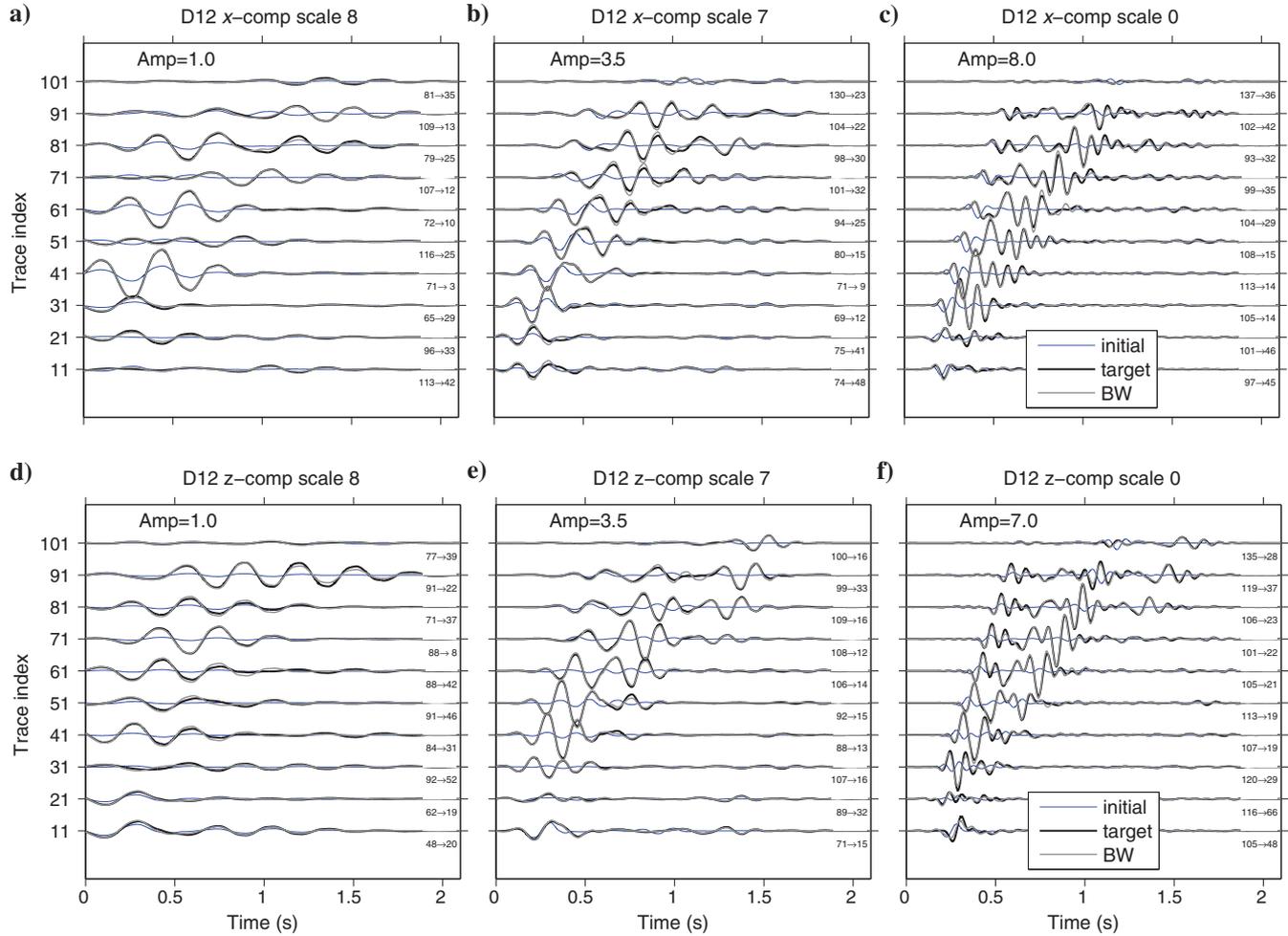
Figure 15. BW seismograms (SWs removed) at scales 8, 7, and 0 before and after fitting by wavelet multiscale adjoint BW tomography as described in the text. Seismograms labeled BW drawn by thick gray lines are computed in the final model of Figure 10.

Figure 16. Inversion results (after 145 iterations) using SW and BW measurements, made in that order, starting from a homogeneous $V_P = 1000$ m/s and $V_S = 500$ m/s and with the exact density model.
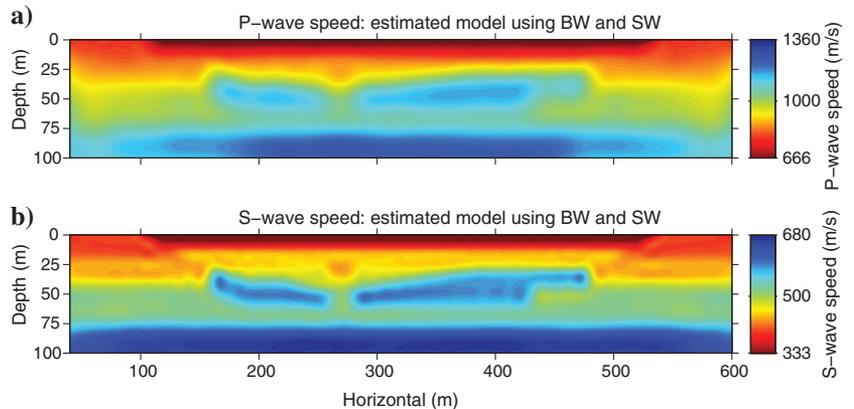
Wavelet-based analysis is powerful in seismic full-waveform inversion problems because of its flexibility. There are numerous wavelet constructions, implementations, and algorithms to choose from, e.g., continuous or discrete, complex or real, orthogonal or biorthogonal, and various wavelet types defined by different parameters (Strang and Nguyen, 1997). Algorithmically, especially when computed using the lifting approach (Sweldens, 1996), there are few complications: The methodology can easily be inserted into generally accepted workflows, by incorporating it in the preprocessing procedure before making measurements of any kind. The intrinsic time-domain nature of the procedure is a benefit when forward and adjoint wavefield modeling is carried out using software such as the open-source spectral-element package SPECFEM (Komatitsch et al., 2005). Finally, as may become more important in the future, when field data will be analyzed, wavelets are an excellent nonlinear denoising tool (Donoho and Johnstone, 1994; Simons et al., 2009).

It goes without saying that our synthetic examples in this paper only represent the early successes of what will require continued development prior to being applied in real-world situations. Among some of the remaining complications are the difficulties associated with generating and recording SWs in the first place, which is chiefly a problem of ensuring that sufficiently low frequencies are available for the initial modeling steps. To yield high-quality tomographic reconstruction of our target structure, acquisition geometry should be optimally designed for resolution control.

Our numerical experiments were designed to make sure the frequencies we modeled are within the scope of the recording spectrum in industry, and our proposed strategy of envelope and phase separation also helps to reduce the dependence on low frequencies (Wu et al., 2014). Nevertheless, challenges remain. Having access to a good initial model will continue to be beneficial, although we showed how using wavelets and using the strategy of envelope-based functionals are reasonably robust to poor initial models. In all our synthetic experiments, we used the exact source time function. In real applications, the input source wavelet can either be inverted from field data (Weglein and Secrest, 1990), or more directly be recorded in marine acquisitions. However, there will be uncertainties associated with imperfect knowledge of the source wavelet.

Encouraged by the results presented in this paper, we have now moved on to treating industry data sets (in a marine setting with
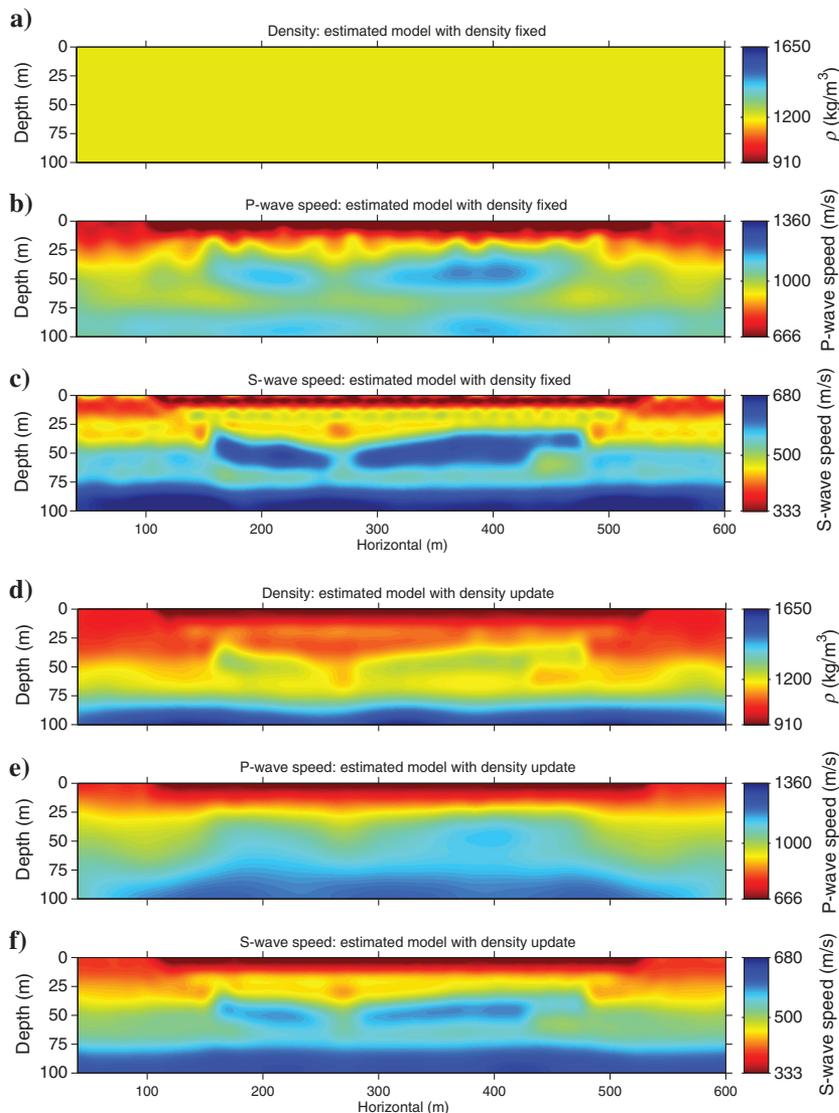
Figure 17. (a-c) Inversions for V_P and V_S from a homogeneous density starting model: without density updates (after 204 iterations), and (d-f) for the full system V_P-V_S-ρ, with density updates (after 99 iterations) using BW and SW measurements. The starting model has $V_P = 1000$ m/s, $V_S = 500$ m/s, and $\rho = 1200$ kg/m^3.

a) V_P rms error versus depth
b) V_S rms error versus depth
c) density rms error versus depth
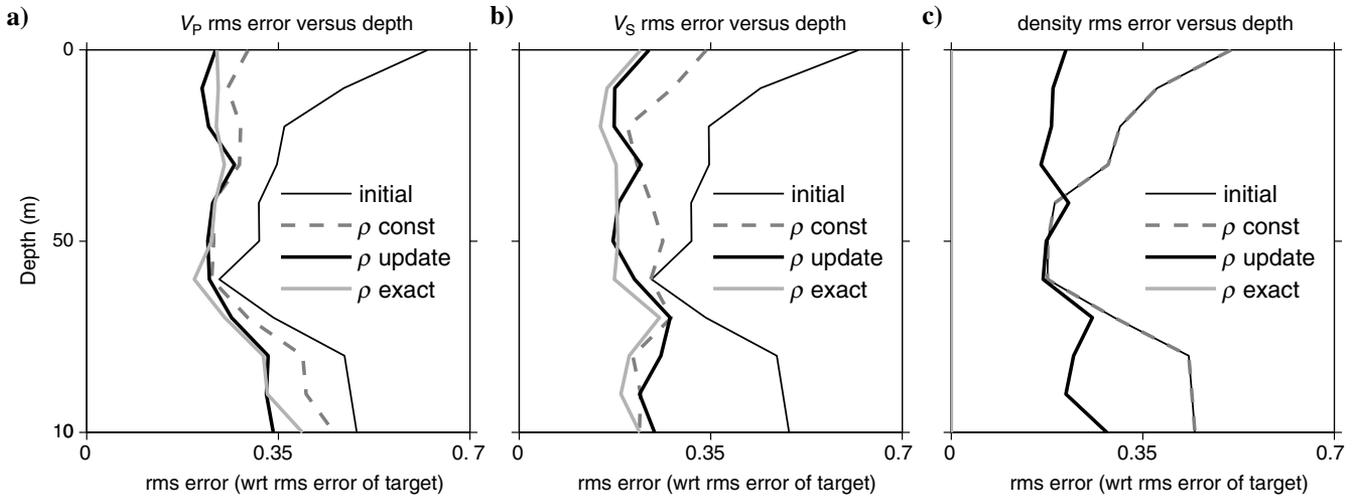
Figure 18. Normalized rms error between the target model and a variety of models including: the initial model with homogeneous density structure (black, thin solid line), (top three panels) the model of Figure 17 with uniform density, held fixed (gray, thick dashed line), (bottom three panels) the model of Figure 17 estimated by simultaneously updating ρ, V_P, and V_S (black, thick solid line), and the model of Figure 16, estimated with the exact density held fixed (gray, thick solid line). From left to right are shown the rms error in terms of V_P, V_S, and ρ. We note that in panel (c), the line for constant ρ overlaps with the initial line, and that the density rms error line with exact ρ is zero everywhere.

well-expressed Scholte waves, and a land-based survey with well-recorded Rayleigh and Love waves), and we look forward to reporting the results soon.

CONCLUSIONS

Two of the grandest challenges for full-waveform inversions in exploration seismology are to be fully elastic (rather than making acoustic approximations), and to use the information contained in the SW records (rather than treating them as undesirable noise). We present a multiscale scheme for full-waveform adjoint tomography based on a (bi)orthogonal wavelet transform. Convergence and stability of the inversions are greatly improved when data and synthetics are progressively presented to the algorithms in a constructive multiscale approximation. Within the industry-standard elastic Marmousi model, we previously applied the multiscale approach successfully to the BWs generated. We started with coarse representations of the seismogram to correct a large-scale background model, and subsequently explained the residuals in the fine scales of the seismogram to map heterogeneities with great complexity. In this paper, we explore the sensitivity of SW in WD tomography. The incorporation of SWs escalates the cycle-skipping problems compared with inversions considering BWs only. However, envelope-based misfit functions applied in the same multiscale framework, on the SW portions of the seismograms, rid the misfit surface of the numerous local minima otherwise present, as we show by spectral-element based adjoint modeling in a synthetic model with challenging near-surface heterogeneity. From an initial set of envelope-based SW inversions, we proceed to a set of WD inversions of the SW, and ultimately conduct a set of WD inversions on the entire seismogram containing SWs and BWs. Throughout this sequence, we maintain the multiscale approach of presenting increasingly fine-resolution wavelet subspaces of the multicomponent seismogram. We investigated the effect of incorrect density information on elastic inversions, and conducted trial inversions

for which the density was also an unknown, by simultaneously updating V_P, V_S, and ρ. Based on our analyses and numerical experiments, we formalize a flexible scheme for full-waveform inversion including BWs and SWs, considering density, P-, and S-wave speeds. Our scheme can be applied to exploration problems, but we expect that it will also be useful for global-scale tomography, as well as for smaller-scale inversion problems in geoengineering.

ACKNOWLEDGMENTS

APPENDIX A

THE ADJOINT METHOD

For completeness, we restate some of the main theoretical concepts as reported by Bozdağ et al. (2011). For more complete and detailed derivations, suggested references include Liu and Tromp (2006), Tape et al. (2007), and Zhu et al. (2009). For a review of the adjoint-state method, recommended references are Tarantola and Nercessian (1984), Tromp et al. (2005), Plessix (2006), and Fichtner (2011).

Suppose $g(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ is any metric to measure the distance between the observations $\mathbf{d}(\mathbf{x}_r, \mathbf{x}_s, t)$ and the model predictions $\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$, for model $\mathbf{m}$, then the misfit function by combining all sources s and all receivers r is defined by

$$\chi(\mathbf{m}) = \sum_{s,r} \int_0^T g(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m}) dt. \qquad \text{(A-1)}$$

The gradient of the misfit function is

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

where $\delta\mathbf{s}$ denotes the perturbations in the synthetic wavefield $\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ due to perturbations to model $\delta\mathbf{m}$, and $\partial_{\mathbf{s}} g(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ denotes the partial derivative of $g(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ over the modeled wavefield $\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$. Suppose we parameterize our model $\mathbf{m}$ in terms of the mass density ρ and the elastic tensor c_{ijkl}.

Under the Born approximation (Wu and Aki, 1985), the ith component of the perturbed wavefield $\delta\mathbf{s}$ can be expressed using the components of the Green's function G_{ij} (the displacement response in the i-direction to a unit point-force impulse in the j-direction) as

$$\delta s_i(\mathbf{x}_r, \mathbf{x}_s, t; \rho, c_{jklm}) = -\int_0^t \int_V [\delta\rho(\mathbf{x}')G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t')\partial_{t'}^2 s_j(\mathbf{x}', t')$$
$$+ \delta c_{jklm}(\mathbf{x}')\partial_k' G_{ij}(\mathbf{x}_r, \mathbf{x}'; t-t')$$
$$\times \partial_l' s_m(\mathbf{x}', t')] d^3\mathbf{x}' dt'. \qquad \text{(A-3)}$$

We introduce an adjoint source given by

$$f_i^\dagger(\mathbf{x}, t) = \sum_r \partial_{s_i} g(\mathbf{x}_r, \mathbf{x}_s, T-t; \mathbf{m})\delta(\mathbf{x} - \mathbf{x}_r). \qquad \text{(A-4)}$$

The adjoint source involves the time-reversed derivative of the defined metric $g(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ over the synthetic wavefield $\mathbf{s}(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ located at the station $\mathbf{x}_r$. We simultaneously back-project $f_i^\dagger$ corresponding to one single source $\mathbf{x}_s$ and all receivers to get the adjoint wavefield components

$$s_j^\dagger(\mathbf{x}', t') = \int_0^{t'} \int_V G_{ji}(\mathbf{x}', \mathbf{x}; t'-t) f_i^\dagger(\mathbf{x}, t) d^3\mathbf{x} dt. \quad \text{(A-5)}$$

For an isotropic material, the gradient of the misfit function in equation A-2 can be expressed in terms of the model perturbations $\delta\rho$, $\delta\kappa$, and $\delta\mu$,

$$\delta\chi = \int_V [K_\rho(\mathbf{x})\delta\rho(\mathbf{x}) + K_\kappa(\mathbf{x})\delta\kappa(\mathbf{x}) + K_\mu(\mathbf{x})\delta\mu(\mathbf{x})] d^3\mathbf{x},$$

$$\text{(A-6)}$$

where K_ρ, K_κ, and K_μ are Fréchet derivatives with respect to density ρ, bulk modulus κ, and shear modulus μ, with the other two parameters held fixed:

$$K_\rho(\mathbf{x}) = -\sum_s \int_0^T \mathbf{s}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{s}(\mathbf{x}, t) dt, \qquad \text{(A-7)}$$

$$K_\kappa(\mathbf{x}) = -\sum_s \int_0^T [\nabla \cdot \mathbf{s}(\mathbf{x}, T-t)][\nabla \cdot \mathbf{s}(\mathbf{x}, t)] dt, \quad \text{(A-8)}$$

$$K_\mu(\mathbf{x}) = -\sum_s \int_0^T 2\mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) dt, \qquad \text{(A-9)}$$

where $\mathbf{D}$ and $\mathbf{D}^\dagger$ are the traceless strain deviator and its adjoint, respectively. The above expressions are termed *misfit kernels* because they involve the combined effects of making measurements for all events and over all stations. In practice, we work with relative, not absolute, and model perturbations.

For each source, the computation of kernels involves calculations of one regular wavefield $\mathbf{s}$ and one adjoint wavefield $\mathbf{s}^\dagger$. The adjoint wavefield (equation A-5) relies on the form of the adjoint source (equation A-4), which in turn depends on our choice of the metric $g(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})$ that defines the data misfit (equation A-1). Changing the objective function only entails modifying the adjoint source.

For our inversions, we use the kernels K_ρ', the impedance kernel (see also Prieux et al., 2013a), K_α, and K_β, representing sensitivity to density, P-wave speed, and S-wave speed perturbations, respectively (Zhu et al., 2009). These are combinations of the expressions A-7–A-9, namely,

$$K_\rho'(\mathbf{x}) = K_\rho + K_\kappa + K_\mu, \qquad \text{(A-10)}$$

$$K_\alpha(\mathbf{x}) = 2\left(1 + \frac{4}{3}\frac{\mu}{\kappa}\right)K_\kappa, \qquad \text{(A-11)}$$

$$K_\beta(\mathbf{x}) = 2\left(K_\mu - \frac{4}{3}\frac{\mu}{\kappa}K_\kappa\right). \qquad \text{(A-12)}$$

In the inversions considering the density as a model parameter, we used the above three expressions. In the cases with known exact (density), or when the density was held fixed, we used K_α and K_β but we set $K_\rho' = 0$. Again, to ensure adequate scaling of the gradients, in practice, we work with relative model perturbations and their associated kernels.

APPENDIX B

ALTERNATIVE ENVELOPE MISFITS

We discuss alternatives to equation 9, namely two different objective functions that are based on envelopes of the seismic waveform, proposed by Bozdağ et al. (2011) but not used anywhere else in this paper.

The squared logarithmic envelope-ratio misfit function is

$$\chi_2(\mathbf{m}) = \frac{1}{2}\sum_{s,r} \int_0^T \left[\ln \frac{E_d(\mathbf{x}_r, \mathbf{x}_s, t)}{E_s(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})}\right]^2 dt. \qquad \text{(B-1)}$$

Its associated adjoint source takes the same form as equation 16,

$$f^{\dagger}(\mathbf{x}, t) = \sum_{r} (E_2^{\text{rat}} s - \mathcal{H}\{E_2^{\text{rat}}[\mathcal{H}\{s\}]\})\delta(\mathbf{x} - \mathbf{x}_r), \quad \text{(B-2)}$$

but with a newly defined ratio E_2^{rat} taking the place of E_1^{rat} in equation 16, namely,

$$E_2^{\text{rat}} = \frac{\ln(E_s) - \ln(E_d)}{(E_s)^2}. \quad \text{(B-3)}$$

The weighting of small residuals by the ED misfit function E_1^{rat} is very different than that by the squared logarithmic envelope ratio misfit function E_2^{rat}. In both cases, care has to be taken to avoid division by small numbers.

A third alternative envelope-based misfit function is to measure the crosscorrelation envelope-traveltime misfit ΔT^E, determined as the time shift that maximizes the crosscorrelation (Luo and Schuster, 1991) of the envelopes of synthetic E_s and observed data E_d:

$$\chi_3(\mathbf{m}) = \frac{1}{2} \sum_{s,r} [\Delta T^E(\mathbf{x}_r, \mathbf{x}_s, t; \mathbf{m})]^2. \quad \text{(B-4)}$$

The corresponding adjoint source has the same expression as equation B-2, but with the ratio E_3^{rat} replacing E_2^{rat},

$$E_3^{\text{rat}} = M^{tt} w \frac{\partial_t E_s}{E_s}, \quad \text{(B-5)}$$

where w is a window function (e.g., from software such as FLEX-WIN; Maggi et al., 2009) applied to the records before taking crosscorrelations, and M^{tt} contains the traveltime shift between the envelopes of synthetics and observations determined by crosscorrelation, normalized by the temporal integral of the windowed product of the synthetic envelope and its second time derivative:

$$M^{tt} = \frac{\Delta T^E}{\int_0^T w(t)\partial_{tt}E_s(t)E_s(t)dt}. \quad \text{(B-6)}$$

The derivations are contained in the work by Luo and Schuster (1991), Marquering et al. (1999), and Dahlen et al. (2000).

REFERENCES

Aki, K., 1957, Space and time spectra of stationary stochastic waves, with special reference to microtremors: Bulletin of the Earthquake Research Institute, **35**, 415–456.

Bagheri, A., E. Pistone, and P. Rizzo, 2014, Guided ultrasonic wave imaging for immersed plates based on wavelet transform and probabilistic analysis: Research in Nondestructive Evaluation, **25**, 63–81, doi: 10.1080/09349847.2013.837212.

Beaty, K. S., D. R. Schmitt, and M. Sacchi, 2002, Simulated annealing inversion of multimode Rayleigh wave dispersion curves for geological structure: Geophysical Journal International, **151**, 622–631, doi: 10.1046/j.1365-246X.2002.01809.x.

Beylkin, G., 1992, On the representation of operators in bases of compactly supported wavelets: SIAM Journal on Numerical Analysis, **29**, 1716–1740, doi: 10.1137/0729097.

Bozdağ, E., and J. Trampert, 2008, On crustal corrections in surface wave tomography: Geophysical Journal International, **172**, 1066–1082, doi: 10.1111/j.1365-246X.2007.03690.x.

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.

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.

Buland, R., and C. H. Chapman, 1983, The computation of seismic travel times: Bulletin of the Seismological Society of America, **73**, 1271–1302.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473, doi: 10.1190/1.1443880.

Capdeville, Y., and P. Cance, 2015, Residual homogenization for elastic wave propagation in complex media: Geophysical Journal International, **200**, 984–997, doi: 10.1093/gji/ggu452.

Capon, J., 1969, High-resolution frequency-wavenumber spectrum analysis: Proceedings of the IEEE, **57**, 1408–1418, doi: 10.1109/PROC.1969.7278.

Cara, M., J.-J. Lévêque, and V. Maupin, 1984, Density-versus-depth models form multimode surface waves: Geophysical Research Letters, **11**, 633–636, doi: 10.1029/GL011i007p00633.

Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell.

Crice, D., 2005, MASW, the wave of the future — Editorial: Journal of Environmental and Engineering Geophysics, **10**, 77–79, doi: 10.2113/JEEG10.2.77.

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.

Dahlen, F. A., S.-H. Hung, and G. Nolet, 2000, Fréchet kernels for finite-frequency traveltimes. Part I: Theory: Geophysical Journal International, **141**, 157–174, doi: 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.

Daubechies, I., 1992, Ten lectures on wavelets, CBMS-NSF Regional Conference Series in Applied Mathematics: SIAM.

de Hoop, M. V., L. Qiu, and O. Scherzer, 2012, Local analysis of inverse problems: Holder stability and iterative reconstruction: Inverse Problems, **28**, 045001, doi: 10.1088/0266-5611/28/4/045001.

Dobrin, M. B., and C. H. Savit, 1988, Introduction to geophysical prospecting, 4th ed.: McGraw-Hill.

Donoho, D. L., and I. M. Johnstone, 1994, Ideal spatial adaptation by wavelet shrinkage: Biometrika, **81**, 425–455, doi: 10.1093/biomet/81.3.425.

Droujinine, A., R.-E. Plessix, and F. Ernst, 2012, Integration of dispersion curve and full waveform inversion techniques for onshore velocity model building — Inner Mongolia study: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20149893.

Dziewoński, A. M., and D. L. Anderson, 1981, Preliminary reference earth model: Physics of the Earth and Planetary Interiors, **25**, 297–356, doi: 10.1016/0031-9201(81)90046-7.

Ecoublet, P. E., S. C. Singh, C. H. Chapman, and G. M. Jackson, 2002, Bent-ray traveltime tomography and migration without ray tracing: Geophysical Journal International, **149**, 633–645, doi: 10.1046/j.1365-246X.2002.01665.x.

Ekström, G., J. Tromp, and E. W. F. Larson, 1997, Measurements and global models of surface wave propagation: Journal of Geophysical Research, **102**, 8137–8157, doi: 10.1029/96JB03729.

Ewing, W. M., W. S. Jardetzky, and F. Press, 1957, Elastic waves in layered media, International Series in the Earth Sciences: McGraw-Hill.

Favier, N., S. Chevrot, and D. Komatitsch, 2004, Near-field influence on shear wave splitting and traveltime sensitivity kernels: Geophysical Journal International, **156**, 467–482, doi: 10.1111/j.1365-246X.2004.02178.x.

Fichtner, A., 2011, Full seismic waveform modelling and inversion: Springer.

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.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. V. nor, 2013, Multiscale full waveform inversion: Geophysical Journal International, **194**, 534–556, doi: 10.1093/gji/ggt118.

Foti, S., 2000, Multistation methods for geotechnical characterization using surface waves: Ph.D. thesis, Politecnico di Torino.

Gabriels, P., R. Snieder, and G. Nolet, 1987, In situ measurements of shear-wave velocity in sediments with higher-mode Rayleigh waves: Geophysical Prospecting, **35**, 187–196, doi: 10.1111/j.1365-2478.1987.tb00812.x.

Gee, L. S., and T. H. Jordan, 1992, Generalized seismological data functionals: Geophysical Journal International, **111**, 363–390, doi: 10.1111/j.1365-246X.1992.tb00584.x.

Gouédard, P., H. Yao, F. Ernst, and R. D. van der Hilst, 2012, Surface wave eikonal tomography in heterogeneous media using exploration data: Geophysical Journal International, **191**, 781–788, doi: 10.1111/j.1365-246X.2012.05652.x.

Haskell, N. A., 1953, The dispersion of surface waves on multilayered media: Bulletin of the Seismological Society of America, **43**, 18–34.

Ivanov, J., R. D. Miller, J. Xia, D. Steeples, and C. B. Park, 2006, Joint analysis of refractions with surface waves: An inverse solution to the refraction-traveltime problem: Geophysics, **71**, no. 6, R131–R138, doi: 10.1190/1.2360226.

Jensen, A., and A. la Cour-Harbo, 2001, Ripples in mathematics: Springer.

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: American Geophysical Union, Geophysical Monograph 157, 205–227, doi: 10.1029/157GM13.

Komatitsch, D., and J. P. Vilotte, 1998, The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of the Seismological Society of America, **88**, 368–392.

Lai, C. G., and G. J. Rix, 1998, Simultaneous inversion of Rayleigh phase velocity and attenuation for near-surface site characterization: School of Civil and Environmental Engineering, Georgia Institute of Technology, Technical report, http://geosystems.ce.gatech.edu/soil_dynamics/docs/Lai_Rix_98.pdf.

Lai, C. G., and G. J. Rix, 1999, Inversion of multi-mode effective dispersion curves, *in* M. Jamiolkowski, R. Lancellotta, and D. Lo Presti, eds., Proceedings of the Second International Symposium on pre-failure deformation characteristics of geomaterials: Balkema, 411–418.

Lebedev, S., T. Meier, G. Nolet, and R. D. van der Hilst, 2005, Automated multi-mode inversion of surface and S waveforms: Geophysical Journal International, **162**, 961–964, doi: 10.1111/j.1365-246X.2005.02708.x.

Lévêque, J.-J., and M. Cara, 1985, Inversion of multimode surface wave data: Evidence for sub-lithospheric anisotropy: Geophysical Journal of the Royal Astronomical Society, **83**, 753–773, doi: 10.1111/j.1365-246X.1985.tb04336.x.

Liu, Q., and J. Tromp, 2006, Finite-frequency sensitivity kernels based upon adjoint methods: Bulletin of the Seismological Society of America, **96**, 2383–2397, doi: 10.1785/0120060041.

Luo, Y., R. Modrak, and J. Tromp, 2014, Strategies in adjoint tomography, *in* W. Freeden, M. Zuhair Nashed, and T. Sonar, eds., Handbook of Geomathematics, 2nd ed.: Springer.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653, doi: 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.

Mallat, S., 2008, A wavelet tour of signal processing: The sparse way, 3rd ed.: Academic Press.

Mallat, S. G., 1989, Multifrequency channel decompositions of images and wavelet models: IEEE Transactions on Acoustics, Speech, and Signal Processing, **37**, 2091–2110, doi: 10.1109/29.45554.

Maraschini, M., F. Ernst, S. Foti, and L. V. Socco, 2010, A new misfit function for multimodal inversion of surface waves: Geophysics, **75**, no. 4, G31–G43, doi: 10.1190/1.3436539.

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.

Masoni, I., R. Brossier, J. Virieux, and J. L. Boelle, 2014, Robust full waveform inversion of surface waves: 84th Annual International Meeting, SEG, Expanded Abstracts, 1126–1130.

McMechan, G. A., and M. J. Yedlin, 1981, Analysis of dispersive waves by wave field transformation: Geophysics, **46**, 869–874, doi: 10.1190/1.1441225.

Miller, G. F., and H. Pursey, 1955, On the partition of energy between elastic waves in a semi-infinite solid: Proceedings of the Royal Society of London, Series A, **233**, 55–69, doi: 10.1098/rspa.1955.0245.

Miller, R. D., J. Xia, C. B. Park, and J. M. Ivanov, 1999, Multichannel analysis of surface waves to map bedrock: The Leading Edge, **18**, 1392–1396, doi: 10.1190/1.1438226.

Montagner, J.-P., and N. Jobert, 1988, Vectorial tomography. II: Application to the Indian Ocean: Geophysical Journal International, **94**, 309–344, doi: 10.1111/j.1365-246X.1988.tb05904.x.

Montagner, J.-P., and H.-C. Nataf, 1988, Vectorial tomography. Part I: Theory: Geophysical Journal International, **94**, 295–307, doi: 10.1111/j.1365-246X.1988.tb05903.x.

Morlet, J., G. Arens, E. Fourgeau, and D. Giard, 1982a, Wave propagation and sampling theory. Part I: Complex signal and scattering in multilayered media: Geophysics, **47**, 203–221, doi: 10.1190/1.1441328.

Morlet, J., G. Arens, E. Fourgeau, and D. Giard, 1982b, Wave propagation and sampling theory. Part II: Sampling theory and complex waves: Geophysics, **47**, 222–236, doi: 10.1190/1.1441329.

Nataf, H.-C., I. Nakanishi, and D. L. Anderson, 1986, Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy. Part 3: Inversion: Journal of Geophysical Research, **91**, 7261–7307, doi: 10.1029/JB091iB07p07261.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization: Springer.

Nolet, G., 1990, Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs: Journal of Geophysical Research, **95**, 8499–8512, doi: 10.1029/JB095iB06p08499.

Nolet, G., 2008, A breviary for seismic tomography: Cambridge University Press.

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.

O'Neill, A., and T. Matsuoka, 2005, Dominant higher surface-wave modes and possible inversion pitfalls: Journal of Environmental and Engineering Geophysics, **10**, 185–201, doi: 10.2113/JEEG10.2.185.

Operto, S., Y. Gholami, R. Brossier, L. Métivier, V. Prieux, A. Ribodetti, and J. Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, 1040–1054, doi: 10.1190/tle32091040.1.

Park, C. B., R. D. Miller, and J. Xia, 1999, Multichannel analysis of surface waves: Geophysics, **64**, 800–808, doi: 10.1190/1.1444590.

Pérez Solano, C. A., D. Donno, and H. Chauris, 2014, Alternative waveform inversion for surface wave analysis in 2-D media: Geophysical Journal International, **198**, 1359–1372, doi: 10.1093/gji/ggu211.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503, doi: 10.1111/j.1365-246X.2006.02978.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.

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.

Pratt, R. G., Z.-M. Song, P. Williamson, and M. Warner, 1996, Two-dimensional velocity models from wide-angle seismic data by wavefield inversion: Geophysical Journal International, **124**, 323–340, doi: 10.1111/j.1365-246X.1996.tb07023.x.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013a, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 1: Imaging compressional wave speed, density and attenuation: Geophysical Journal International, **194**, 1640–1664, doi: 10.1093/gji/ggt177.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013b, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the Valhall field. Part 2: Imaging compressive-wave and shear-wave velocities: Geophysical Journal International, **194**, 1665–1681, doi: 10.1093/gji/ggt178.

Rawlinson, N., and M. Sambridge, 2003, Seismic traveltime tomography of the crust and lithosphere: Advances in Geophysics, **46**, 81–198, doi: 10.1016/S0065-2687(03)46002-0.

Rayleigh, J. W. S., 1885, On waves propagated along the plane surface of an elastic solid: Proceedings of the London Mathematical Society, **17**, 4–11, doi: 10.1112/plms/s1-17.1.4.

Richart, F. E., J. R. Hall, and R. D. Woods, 1970, Vibrations of soils and foundations: Prentice-Hall.

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.

Ritsema, J., A. A. Deuss, H. J. van Heijst, and J. H. Woodhouse, 2011, S40RTS: A degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements: Geophysical Journal International, **184**, 1223–1236, doi: 10.1111/j.1365-246X.2010.04884.x.

Rix, G. J., C. G. Lai, and A. W. S. Jr, 2000, In situ measurement of damping ratio using surface waves: Journal of Geotechnical and Geoenvironmental Engineering, **126**, 472–480, doi: 10.1061/(ASCE)1090-0241(2000)126:5(472).

Rix, G. J., C. G. Lai, M. C. Orozco, G. L. Hebeler, and V. Roma, 2001, Recent advances in surface wave methods for geotechnical site characterization: Presented at XV International Conference on Soil Mechanics and Geotechnical Engineering, 499–502.

Romanowicz, B., 2003, Global mantle tomography: Progress status in the last 10 years: Annual Reviews of Geophysics and Space Physics, **31**, 303–328.

Romanowicz, B., and D. Giardini, 2001, The future of permanent seismic networks: Science, **293**, 2000–2001, doi: 10.1126/science.1061771.

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.

Sigloch, K., and G. Nolet, 2006, Measuring finite-frequency body-wave amplitudes and traveltimes: Geophysical Journal International, **167**, 271–287, doi: 10.1111/j.1365-246X.2006.03116.x.

Simons, F. J., B. D. E. Dando, and R. M. Allen, 2006a, 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.

Simons, F. J., G. Nolet, J. M. Babcock, R. E. Davis, and J. A. Orcutt, 2006b, A future for drifting seismic networks: Eos — Transactions of the American Geophysical Union, **87**, 305–307, doi: 10.1029/2006EO310002.

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.

Simons, F. J., and R. D. van der Hilst, 2003, Seismic and mechanical anisotropy and the past and present deformation of the Australian lithosphere: Earth and Planetary Science Letters, **211**, 271–286, doi: 10.1016/S0012-821X(03)00198-5.

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.

Snieder, R., 1988a, Large-scale waveform inversions of surface waves for lateral heterogeneity. Part 1: Theory and numerical examples: Journal of Geophysical Research, **93**, 12055–12065, doi: 10.1029/JB093iB10p12055.

Snieder, R., 1988b, Large-scale waveform inversions of surface waves for lateral heterogeneity. Part 2: Application to surface waves in Europe and the Mediterranean: Journal of Geophysical Research, **93**, 12067–12080, doi: 10.1029/JB093iB10p12067.

Socco, L. V., S. Foti, and D. Boiero, 2010, Surface-wave analysis for building near-surface velocity models — Established approaches and new perspectives: Geophysics, **75**, no. 5, 75A83–75A102, doi: 10.1190/1.3479491.

Socco, L. V., and C. Strobbia, 2004, Surface-wave method for near-surface characterization: A tutorial: Near Surface Geophysics, **2**, 165–185.

Stokoe, K. H., J. Sung-Ho, and R. D. Woods, 2004, Some contributions of in situ geophysical measurements to solving geotechnical engineering problems: Geotechnical and geophysical site characterization: Proceedings of the 2nd International Conference on Site Characterization, International Society of Soil Mechanics and Geotechnical Engineering, 97–132.

Strang, G., and T. Nguyen, 1997, Wavelets and filter banks, 2nd ed.: Wellesley-Cambridge Press.

Stutzmann, E., and J.-P. Montagner, 1993, An inverse technique for retrieving higher mode phase velocity and mantle structure: Geophysical Journal International, **113**, 669–683, doi: 10.1111/j.1365-246X.1993.tb04659.x.

Stutzmann, E., and J.-P. Montagner, 1994, Tomography of the transition zone from the inversion of higher-mode surface-waves: Physics of the Earth and Planetary Interiors, **86**, 99–115, doi: 10.1016/0031-9201(94)05064-3.

Sweldens, W., 1996, The lifting scheme: A custom-design construction of biorthogonal wavelets: Applied and Computational Harmonic Analysis, **3**, 186–200, doi: 10.1006/acha.1996.0015.

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.

Tarantola, A., 1984a, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.

Tarantola, A., 1984b, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015, doi: 10.1111/j.1365-2478.1984.tb00751.x.

Tarantola, A., and A. Nercessian, 1984, Three-dimensional inversion without blocks: Geophysical Journal of the Royal Astronomical Society, **76**, 299–306, doi: 10.1111/j.1365-246X.1984.tb05047.x.

Thompson, D. O., and D. E. Chimenti, 1997, Review of progress in quantitative nondestructive evaluation 16th ed.: Springer.

Tian, Y., R. Montelli, G. Nolet, and F. A. Dahlen, 2007, Computing traveltime and amplitude sensitivity kernels in finite-frequency tomography: Journal of Computational Physics, **226**, 2271–2288, doi: 10.1016/j.jcp.2007.07.004.

Tokimatsu, K., S. Tamura, and H. Kojima, 1992, Effects of multiple modes on Rayleigh wave dispersion characteristics: Journal of Geotechnical Engineering, **118**, 1529–1543, doi: 10.1061/(ASCE)0733-9410(1992)118:10(1529).

Trampert, J., and J. H. Woodhouse, 2003, Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s: Geophysical Journal International, **154**, 154–165, doi: 10.1046/j.1365-246X.2003.01952.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.

VanDecar, J. C., and R. S. Crosson, 1990, Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares: Bulletin of the Seismological Society of America, **80**, 150–169.

van Heijst, H. J., and J. H. Woodhouse, 1997, Measuring surface-wave overtone phase velocities using a mode-branch stripping technique: Geophysical Journal International, **131**, 209–230, doi: 10.1111/j.1365-246X.1997.tb01217.x.

Van Heijst, H. J., and J. H. Woodhouse, 1999, Global high-resolution phase velocity distributions of overtone and fundamental-mode surface waves determined by mode branch stripping: Geophysical Journal International, **137**, 601–620, doi: 10.1046/j.1365-246x.1999.00825.x.

Weglein, A. B., and B. G. Secrest, 1990, Wavelet estimation for a multidimensional acoustic or elastic earth: Geophysics, **55**, 902–913, doi: 10.1190/1.1442905.

Woodhouse, J. H., 1974, Surface waves in a laterally varying layered structure: Geophysical Journal of the Royal Astronomical Society, **37**, 461–490, doi: 10.1111/j.1365-246X.1974.tb04098.x.

Woodhouse, J. H., and A. M. Dziewoński, 1984, Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms: Journal of Geophysical Research, **89**, 5953–5986, doi: 10.1029/JB089iB07p05953.

Wu, R.-S., and K. Aki, 1985, Scattering characteristics of elastic waves by an elastic heterogeneity: Geophysics, **50**, 582–595, doi: 10.1190/1.1441934.

Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, no. 3, WA13–WA24, doi: 10.1190/geo2013-0294.1.

Xia, J., R. D. Miller, and C. B. Park, 1999, Estimation of near-surface shear-wave velocity by inversion of Rayleigh waves: Geophysics, **64**, 691–700, doi: 10.1190/1.1444578.

Xia, J., R. D. Miller, C. B. Park, and G. Tian, 2002, Determining Q of near-surface materials from Rayleigh waves: Journal of Applied Geophysics, **51**, 121–129, doi: 10.1016/S0926-9851(02)00228-8.

Yuan, D., and S. Nazarian, 1993, Automated surface wave method: Inversion technique: Journal of Geotechnical Engineering, **119**, 1112–1126, doi: 10.1061/(ASCE)0733-9410(1993)119:7(1112).

Yuan, Y. O., and F. J. Simons, 2014, Multiscale adjoint waveform-difference tomography using wavelets: Geophysics, **79**, no. 3, WA79–WA95, doi: 10.1190/geo2013-0383.1.

Yuan, Y. O., F. J. Simons, and E. Bozdağ, 2014, Full-waveform adjoint tomography in a multiscale perspective: 84th Annual International Meeting, SEG, Expanded Abstracts, 1194–1199.

Zhang, S. X., and L. S. Chan, 2003, Possible effects of misidentified mode number on Rayleigh wave inversion: Journal of Applied Geophysics, **53**, 17–29, doi: 10.1016/S0926-9851(03)00014-4.

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.