Automating seismic waveform analysis for full 3-D waveform inversions

En-Jui Lee and Po Chen

Department of Geology and Geophysics, University of Wyoming, WY 82071, USA. E-mail: elee8@uwyo.edu

Accepted 2013 March 26. Received 2013 March 20; in original form 2012 July 12

SUMMARY

We present a semi-automatic seismic waveform selection algorithm that can be used in full 3-D waveform inversions for earthquake source parameters and/or earth structure models. The algorithm is applied on pairs of observed and synthetic seismograms. A pair of observed and synthetic seismograms are first segmented in the wavelet domain into a number of wave packets using a topological watershed algorithm. A set of user-adjustable criteria based on waveform similarities is then applied to match each wave packet obtained from the observed seismogram with the corresponding wave packet obtained from the synthetic seismogram. The selected wave packet pairs are then used for extracting frequency-dependent phase and amplitude misfit measurements, which can be used in seismic source and/or structural inversions. The algorithm takes advantage of time–frequency representations of seismograms and is able to separate seismic phases in both time and frequency domains. We demonstrate the flexibility of this algorithm using examples of full 3-D waveform inversions for earthquake centroid moment tensors and earth structure models at different geographic scales.

Key words: Time-series analysis; Wavelet transform; Earthquake source observations; Seismic tomography; Computational seismology.

1 INTRODUCTION

Recent advances in parallel computing technology and numerical methods (e.g. Olsen 1994; Graves 1996; Komatitsch & Tromp 2002) have made large-scale, 3-D simulations of the seismic wave-fields much more affordable, which has opened up the possibility of full 3-D waveform tomography (F3DT; e.g. Chen *et al.* 2007a; Fichtner *et al.* 2009; Tape *et al.* 2010; Chen 2011; Liu & Gu 2012) and earthquake source parameter inversions (e.g. Liu *et al.* 2004; Chen *et al.* 2005; Zhao *et al.* 2006; Hingee *et al.* 2011; Kim *et al.* 2011; Lee *et al.* 2011). In F3DT, both the starting model and the derived model perturbations are 3-D in space and the sensitivity (Fréchet) kernels are computed by solving the 3-D (visco)elastic seismic wave equation numerically (e.g. Zhao *et al.* 2005; Liu & Tromp 2006). The Fréchet kernels with respect to structural and/or source parameters can be constructed for an arbitrary segment on the seismogram using either the scattering-integral method (e.g. Zhao *et al.* 2005; Chen *et al.* 2007b) or the adjoint method (e.g. Tarantola 1988; Tromp *et al.* 2005; Liu & Tromp 2006). To fully take advantage of this new capability in waveform inversions, the waveform selection algorithm can no longer be limited to the nomenclature of conventional seismic phases and needs to be extended to accommodate waveforms resulting from complex wave propagation effects in 3-D structural models.

The motivation as well as the necessity for performing seismogram segmentation and waveform selection in waveform inversions

comes from a number of considerations. First, using waveforms of individual seismic phases rather than combinations of different phases can reduce possible non-linear effects due to the interference among multiple seismic arrivals. For example, the frequency-dependent phase delay measurements of a single seismic phase have quasi-linear relationship with respect to seismic velocity perturbations (i.e. the linearization involved in each iteration of the gradient-based non-linear inversion is the Rytov approximation, which is valid for large phase shifts as long as the phase perturbation *per wavelength* is small (e.g. Chernov 1960; Snieder & Lomax 1996), but if multiple seismic phases are considered at the same time, this quasi-linear relationship may break down (e.g. Keller 1969; Woodward 1989). Secondly, seismogram segmentation allows us to make incremental changes to our source and structure models. As demonstrated in Tape *et al.* (2010) and Fichtner *et al.* (2009), we can start the iterative inversion process by fitting segments of the observed seismograms that have high similarities with the synthetic seismograms and gradually improve our structure model and try to fit more observed waveforms through iterations. Previous experiences also suggest that this type of bootstrapping processes in time and/or frequency domain (e.g. Bunks *et al.* 1995) can also reduce the non-linearity effect in full-wave tomographic inversions. Thirdly, different types of waves may have different sensitivities to different types of model parameters. For example, in general the propagation paths of direct-arriving *P* and *S* waves can be quite similar, but their traveltimes can have very different sensitivities to

P- and *S*-wave velocity. Fourthly, different types of seismic phases (e.g. direct and reflected waves) may have very different spatial sampling pattern on structural models. Separating seismic phases allows the assignment of different weights to different phases to balance the sensitivities to different parts of the model in the inversion. For example, Lebedev *et al.* (2005) applied different weightings to different seismic phases to ensure that *S* and multiple-*S* phases have balanced contributions on imaging deeper structures.

Throughout the history of waveform inversions, a large number of waveform selection algorithms have been developed and many of them have been automated to some extent as the volume of high-quality digital seismograms keeps increasing. In surface-wave studies, automated algorithms have been developed to extract dispersion measurements from the fundamental mode (e.g. Trampert & Woodhouse 1995; Laske & Masters 1996; Levshin & Ritzwoller 2001). Algorithms based on time–frequency analysis have been applied to obtain measurements from higher modes (e.g. Debayle 1999; Lebedev *et al.* 2005; Visser *et al.* 2007). In body-wave studies and also in some surface wave studies or studies of joint inversions using both types of waves (e.g. Panning & Romanowicz 2006), automated waveform selection algorithms usually pick time windows around specific seismic phases defined by traveltimes predicted using ray-theoretic methods in 1-D structural models (e.g. Ritsema & van Heijst 2002; Sigloch & Nolet 2006; Lawrence & Shearer 2008). For full 3-D waveform inversions, we need a new type of waveform selection algorithm that is not tied to ray-theoretic traveltime predictions of known phases and can accommodate wave arrivals resulting from complex wave propagation effects in 3-D structural models.

For full 3-D waveform inversions, the difficulty in automating the waveform selection algorithm lies in two aspects. First, the synthetic and observed seismograms could be very complex, especially at local to regional distances, where complex 3-D wave effects caused by small-scale crustal heterogeneities are significant. Different types of waves recorded by a given sensor could overlap in the time and/or the frequency domains, which poses challenges to the design of automated seismogram segmentation algorithms. Secondly, the modelling assumptions, in particular, the reference structure and/or source models used in wave-propagation simulations, could be inadequate to describe many of the wave arrivals on the observed seismogram, which poses difficulties in automatically identifying robust features on the seismogram and contributing these features to particular aspects of the structure and/or source models.

To address some of the difficulties in the waveform selection algorithm used in full 3-D waveform inversions, Maggi *et al.* (2009) proposed an automated time-window selection algorithm based on the short-term average/long-term average ratio (STA/LTA) curve. A set of criteria based on similarities between the observed and synthetic waveforms is used to accept or reject time windows. The algorithm has been successfully applied to seismograms recorded at local, regional and global scales in different full 3-D waveform inversion and modelling studies (e.g. Tape *et al.* 2010; Zhu *et al.* 2012).

A purely time-domain waveform selection algorithm may have limitations in separating waves arriving at overlapping time windows. We have developed a new semi-automatic seismogram segmentation and waveform selection algorithm to address this issue. In our algorithm, seismogram segmentation is performed in the time–frequency domain through the continuous wavelet transform, which allows extra freedom in separating waves arriving at overlapping time windows but with disjoint frequency domain or time–frequency domain supports. We give an introduction of our algorithm in Section 2 and demonstrate its usefulness using examples

of full 3-D waveform inversions in Section 3. We summarize the results and discuss possible future developments of the algorithm in Section 4. The software we have developed is open-source and will be hosted by the Community Modeling Environment (CME) of the Southern California Earthquake Center (SCEC).

2 METHODOLOGY

The major steps in our seismogram segmentation and waveform selection algorithm are summarized in Fig. 1 and the parameters needed to adjust the behaviour of the algorithm are listed in Table 1. In the pre-processing step, criteria based on signal-to-noise ratios are used to reject noisy observed seismograms. In the seismogram

Figure 1. A flow chart of our seismogram analysis algorithm. There are three major components in the algorithm: seismogram pre-processing, seismogram segmentation and waveform selection, which are introduced in Sections 2.1, 2.2 and 2.3, respectively. The seismogram segmentation algorithm is based on the continuous wavelet transform (CWT), the topological watershed (TW) method, and the inverse continuous wavelet transform (ICWT).

Table 1. List of pre-set parameters and their meanings.

Seismogram pre-processing	
<i>SNRE1</i>	Energy signal-to-noise ratio for whole observed seismograms
<i>SNRA1</i>	Amplitude signal-to-noise ratio for whole observed seismograms
Waveform selection criteria	
<i>WIN_MIN</i>	Minimum accepted length of wave packets
<i>WIN_BEG</i>	Beginning time of wave packets
<i>WIN_END</i>	End time of wave packets
<i>SNRE2</i>	Energy signal-to-noise ratio for observed wave packets
<i>SNRA2</i>	Amplitude signal-to-noise ratio for observed wave packets
<i>MIN_NCCB</i>	Acceptance threshold for NCC of body wave phases
<i>MIN_NCCS</i>	Acceptance threshold for NCC of surface wave phases
<i>MAX_ΔTB</i>	Acceptance threshold for time-shift of body wave phases
<i>MAX_ΔTS</i>	Acceptance threshold for time-shift of surface wave phases
<i>MAX_ΔlnA</i>	Acceptance threshold for amplitude ratio
Measurement selection criteria	
<i>NCC_corr</i>	Acceptance threshold for NCC between observed and corrected synthetic wave packets
<i>SNR_f</i>	Acceptance threshold of amplitude ratio for measurements

segmentation step, the time-domain seismogram is converted into the time–frequency domain through the continuous wavelet transform, the scalogram (i.e. the squared modulus of the continuous wavelet transform) in the time–frequency domain is then segmented using a topological watershed algorithm and each portion of the segmented scalogram is transformed back to the time domain through the inverse continuous wavelet transform. Each wave packet obtained from segmenting the observed seismogram is then matched with a wave packet obtained from segmenting the corresponding synthetic seismogram based on criteria such as waveform similarity, the relative time-shift and the amplitude ratio. For each selected wave packet pair, we then extract frequency-dependent phase and amplitude misfit measurements, as well as the associated error estimates.

2.1 Seismogram pre-processing

In the pre-processing step, we reject noisy observed seismograms and prepare the selected observed and synthetic seismogram pairs for further analysis. Two types of signal-to-noise ratios are used to reject noisy observed seismograms, one is based on the energy ratio (SNR_E) and the other is based on the maximum amplitude ratio (SNR_A). The energy ratio and the maximum amplitude ratio are defined as

$$SNR_E = E_S/E_N = \frac{\int_{t_S^0}^{t_S^1} \bar{s}^2(t) dt / (t_S^1 - t_S^0)}{\int_{t_N^0}^{t_N^1} \bar{s}^2(t) dt / (t_N^1 - t_N^0)}, \quad (1)$$

$$SNR_A = A_S/A_N, \quad (2)$$

where $\bar{s}(t)$ is the observed seismogram, $[t_N^0, t_N^1]$ and $[t_S^0, t_S^1]$ are time windows containing the background noise and the seismic signal, A_N and A_S are the maximum amplitudes of the noise and the seismic signal. The noise window is usually selected to be a segment of the seismogram before the estimated P -wave arrival time. The specific definition of what exactly constitutes the seismic signal is problem dependent. In our local source parameter inversion and full 3-D waveform tomography applications, we use the segment between the estimated P -wave arrive time and the end of the minor-arc

Table 2. Values of the pre-set parameters for the local full 3-D tomography in Southern California and for a global full-wave tomography.

	Southern California	Global
<i>SNRE1</i>	3.0	3.0
<i>SNRA1</i>	3.0	3.0
<i>WIN_MIN</i>	5 s	50 s
<i>WIN_BEG</i>	First arrival – 5 s	First arrival – 15 s
<i>WIN_END</i>	Surface wave + 30 s	End time of synthetics
<i>SNRE2</i>	3.0	3.0
<i>SNRA2</i>	3.0	3.0
<i>MIN_NCCB</i>	0.75	0.65
<i>MIN_NCCS</i>	0.70	0.50
<i>MAX_ΔTB</i>	3.5 s	20.0 s
<i>MAX_ΔTS</i>	5.0 s	25.0 s
<i>MAX_ΔlnA</i>	2	2
<i>NCC_corr</i>	0.9	0.9
<i>SNR_f</i>	33 per cent	33 per cent

surface wave train as our signal. But in teleseismic applications, many useful phases may arrive after the minor-arc surface wave train. An observed seismogram with its SNR ratios below the pre-set thresholds is rejected. Examples of the threshold values for our local and teleseismic applications are listed in Table 2.

For a selected observed seismogram, we then remove the mean, the linear trend and apply a symmetric taper to each end of the seismogram. A filter is then applied on each pair of the observed and synthetic seismograms. The type and frequency bands of the filter depend upon the specific application.

2.2 Seismogram segmentation

A seismogram is composed of different wave arrivals. Our seismogram segmentation algorithm is based on the speculation that in the time–frequency domain different wave arrivals may correspond to distinct regions of high power in the scalogram. The process of our segmentation algorithm is composed of three major steps (Fig. 2), including: forward continuous wavelet transform of a time domain seismogram; topological watershed transform of the time–frequency domain scalogram; inverse continuous wavelet transform of the segmented waveforms back to the time domain.

Figure 2. An example of the seismogram segmentation process. First row: the time-domain seismogram after the pre-processing step; second row: the time-frequency domain scalogram obtained by applying CWT on the time-domain seismogram; third row: segmented scalogram obtained by applying the TW method on the scalogram; fourth row: the time-domain wave packets obtained by converting each portion of the segmented scalogram back into the time domain through ICWT.

2.2.1 Continuous wavelet transform

The continuous wavelet transform of a seismogram $s(t)$ with respect to a mother wavelet $g(t)$ is defined as (Holschneider 1995),

$$W_g s(b, a) = \int_{-\infty}^{+\infty} \frac{1}{a} g^* \left(\frac{t-b}{a} \right) s(t) dt, \quad (3)$$

where a is the dilation ($a > 0$), b is the translation and the symbol ‘*’ denotes complex conjugate. The scalogram used for segmentation is defined as the squared modulus of $W_g s(b, a)$. The corresponding inverse transform is then defined as

$$s(t) = \frac{2}{C_g} \operatorname{Re} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a} g \left(\frac{t-b}{a} \right) W_g s(b, a) \frac{db da}{a} \right], \quad (4)$$

where C_g is a normalization constant. In principle, our segmentation algorithm can be based on any type of time–frequency transform. The reason we are adopting the continuous wavelet transform in our algorithm is its linearity, which results in the absence of interfering cross-terms, and its dyadic pavement of the time–frequency space (e.g. Holschneider 1995; Kulesh *et al.* 2005), which allows efficient and high-resolution representation of the time–frequency content of the seismogram. In our approach, we did not adopt the discrete wavelet transform due to the lack of redundancy in the discrete wavelet bases, which could reduce the resolution of the resulting time–frequency domain image of the seismogram.

The appropriate choice of the mother wavelet is an important issue. We have selected the complex-valued Cauchy wavelet (Holschneider 1995), which is also known as the Paul wavelet, for our analysis. In frequency domain, the Cauchy wavelet is defined as

$$\hat{g}_m(\omega) = \begin{cases} 0, & \text{for } \omega < 0, \\ \omega^m e^{-\omega}, & \text{for } \omega \geq 0, \end{cases} \quad (5)$$

where ω is the angular frequency and m is the order of the wavelet. We have chosen the Cauchy wavelet because of its nice algebraic and analytic properties. First, the Cauchy wavelet is admissible, meaning that the seismogram can be reconstructed from its continuous wavelet transform without any loss of information. Secondly, the Cauchy wavelet has good time and frequency localization properties, which is important for improving the local resolution of the time–frequency domain image of the seismogram. Its time–frequency area approaches the smallest area allowable by the Heisenberg inequality asymptotically when its order tends towards infinity. Thirdly, the first- and second-order temporal derivatives of the Cauchy wavelet is still a Cauchy wavelet, which simplifies the application of our algorithm to velocity and acceleration seismograms. Fourthly, the Cauchy wavelet is progressive, meaning its Fourier transform has zero amplitude in negative frequency. This property allows us to separate the wavelet transform of the seismogram into prograde (i.e. the wavelet transform is zero for $a < 0$) and retrograde (i.e. the wavelet transform is zero for $a \geq 0$) components (Kulesh *et al.* 2008), which can be used for analysing the polarization properties of multicomponent seismograms and can potentially provide additional criteria for wave arrival separation (Kulesh *et al.* 2005).

The scalogram is a 2-D function of both translation (i.e. time) and dilation (i.e. frequency). An example of the scalogram computed

using the Cauchy wavelet for a seismogram used in our full 3-D waveform tomography project for Southern California is shown in the second row of Fig. 2. The dominant arrival on the seismogram is the surface wave, which corresponds to the global maximum on the scalogram. Other local maxima on the scalogram correspond to other seismic phases.

2.2.2 Topological watershed method

The segmentation is performed automatically on the time–frequency domain scalogram using the topological watershed method, a topological transformation designed to cluster all pixels that are connected to the same local extremum (e.g. Vincent & Soille 1991; Bertrand 2005).

The earliest watershed algorithm was introduced by Beucher & Lantuejoul (1979) as a tool for segmenting grey-scale images. Efficient implementations of the watershed algorithm based on immersion simulations were proposed by Vincent & Soille (1991) and Meyer (1992). In such immersion simulations, the 2-D grey-scale image is reversed and the local maxima become local minima, which are called the catchment basins. The catchment basins are flooded through inlets (seeds) pierced at those local minima. As the flooding progresses, some regions could start to mix and at this point a dam is built to keep the regions separated. As the flood reaches the top of the reversed scalogram, all the dams that have been built during the flooding process form the watershed of the scalogram. A conceptual representation of the immersion simulation using a 1-D topography is shown in Figs 3(a)–(c). The number and the locations of the seeds can be selected in advance to avoid oversegmentation when the signal-to-noise ratio is too low. Catchment basins without seeds can be flooded by water coming from a neighbouring catchment basin.

The topological approach to the watershed problem was introduced in Couprie & Bertrand (1997). The basic idea is to let the water gradually ‘erode’ the relief of the grey-scale image while preserving the connectivity of every lower cross-section. A conceptual comparison between the immersion simulation and the topological approach is shown in Figs 3(a–c) and (d–f). It was proved in Bertrand (2005) that unlike other types of watershed algorithms, this topological approach is ‘contrast-preserving’ (i.e. the height of the dam is the same as the height of the lowest mountain separating the catchment basins), which is a property useful for further processing (e.g. reconnection of oversegmented regions). The mathematical details of the topological watershed algorithm used in our segmentation processes are presented in Appendix A. The result of the topological watershed transform is a set of ‘mask files’ corresponding to the different portions on the segmented scalogram. Each mask file is composed of a sequence of ‘1’s and ‘0’s and each number corresponds to a single pixel on the scalogram. The value is ‘1’ for pixels within the region selected by the mask file and ‘0’ for the outside pixels.

The time–frequency domain window (i.e. the mask file), generated by applying the topological watershed transform to the scalogram $\|W_g s(b, a)\|$, is then multiplied onto the continuous wavelet transform $W_g s(b, a)$ and the result is then transformed back to the time-domain through the inverse continuous wavelet transform, as illustrated in Fig. 2.

An important application of our seismogram segmentation algorithm is full 3-D waveform tomography. In Appendix B, we show how full-wave sensitivity kernels can be computed for an arbitrary waveform selected in the time–frequency domain.

Figure 3. Conceptual representations of the immersion simulation used in conventional watershed algorithms (a)–(c) and the topological watershed transform (d)–(f) using a 1-D topography. In (a)–(c), different stages of flooding are shown in grey. As regions start to merge, a dam (black vertical bar) is built to separate the water. As the water level keeps increasing, more dams are built until the whole topography is fully flooded. In (d)–(f), the topological watershed transformation is computed by iteratively lowering W-destructible points until no such points exist and the dams are built. We note that the final altitudes of the dams in (c) are different from those in (f).

2.2.3 Comparison with Maggi *et al.* (2009)

The purpose for developing our seismogram segmentation and waveform selection algorithm is similar to that in Maggi *et al.* (2009). The authors have packaged their codes in the software ‘FLEXWIN’, which is open-source and hosted at the Computational Infrastructure for Geodynamics (CIG) website. In their algorithm, a target seismogram is separated into many small time windows based on the local minima in its short-term average/long-term average ratio (STA/LTA) curve. The local minima below the user-defined water level are ‘strong’ barriers and the local minima between strong barriers are ‘soft’ barriers. The time-windows that cross-strong barriers cannot be merged. The time-windows separated by soft barriers are allowed to merge with its neighbouring time-windows or to be rejected, depending on the local maxima of its neighbouring time-windows and the separation of distinct phases.

In our algorithm, the segmentation is performed in the time–frequency domain, which allows us to separate waves arriving at overlapping time windows but with disjoint frequency domain or time–frequency domain supports. Compared with conventional time–frequency domain filtering techniques, the topological watershed algorithm allows the boundaries separating different wave arrivals to naturally conform to the distribution of energy in the time–frequency domain.

An example is shown in Fig. 4. An explosive source is buried at 300 m depth and receivers with offset ranging from 2.5 to 5.3 km are evenly distributed on the left of the source on the surface. Synthetic seismograms were computed by solving the 3-D elastic seismic wave equation using a fourth-order staggered-grid finite-difference code (Olsen 1994). The vertical-component seismograms

for all receivers are shown in Fig. 4(a), from which we can see the direct-arriving *P* wave, the Rayleigh wave, a stronger reflected phase *PP* and a weaker reflected phase *pPP*. At around 4 km offset, the Rayleigh wave is interfering with the reflected *PP* and *pPP* arrivals and an example is shown in Fig. 4(c). We segmented all seismograms using our algorithm and plotted only the *PP* waves in Fig. 4(b), which shows a clean separation between *PP* and the Rayleigh wave for all offsets and the shapes of the *PP* waves are well preserved. The scalogram for the selected seismogram in Fig. 4(c) is shown in Fig. 4(d) and the segmented scalograms, as well as their corresponding time-domain waveforms, are shown in Figs 4(e)–(l). The seismogram shown in Fig. 4(c) was also segmented using the FLEXWIN code of Maggi *et al.* (2009) and the results are shown in Figs 4(m)–(n). It is possible to isolate the *PP* wave by manually adjusting the ‘internal minima’ parameter in FLEXWIN (Maggi *et al.* 2009), however, the shape of the isolated *PP* wave is still distorted by the Rayleigh wave as shown in Fig. 4(m).

We note that for phases that are well separated in the time domain, FLEXWIN and our algorithm gave very similar results. An example is the direct-arriving *P* wave as shown in Figs 4(f) and (m).

2.3 Waveform selection

The wave packets obtained by segmenting the synthetic seismogram need to be paired with those obtained by segmenting the corresponding observed seismogram before we can make frequency-dependent phase and amplitude misfit measurements. The process of our waveform selection algorithm has three major steps:

Step (1): We reject some of the wave packets based on their temporal lengths and their arrival times. We require the temporal

Figure 4. A comparison with FLEXWIN. (a) Synthetic seismograms of vertical-component velocity computed using a layer-over-half-space velocity model. Within the circled region, the PP reflection and the surface wave arrive in overlapping time windows. (b) The PP reflection isolated using our segmentation algorithm. (c) An example seismogram with PP, pPP and the surface wave arriving in overlapping time windows. (d) The scalogram of the seismogram shown in (c). (e–l) The segmented scalograms and their corresponding time-domain waveforms obtained using our segmentation algorithm. (m and n) The time windows selected by the FLEXWIN code, the boxes indicate the selected time windows and the dash line on the STA:LTA curve is the water level. The FLEXWIN parameters used for obtaining this result are $T_{0,1} = 0.0225, 10$; $w_E = 0.0365$; $c_0 = 1.0$; $c_1 = 1.0$; $c_2 = 0.0$; $c_{3a,b} = 2.0, 1.0$; $c_{4a,b} = 2.0, 6.0$; $w_{CC} = 1.0$; $w_{len} = 1.0$; $w_{nwin} = 1.0$.

length of the wave packet to be larger than the pre-set parameter WIN_MIN (Table 1), which is usually set to the shortest period of the pre-processed seismogram. We also require the arrival time of the wave packet to be later than WIN_BEG and earlier than WIN_END (Table 1). In our local waveform inversion examples, WIN_BEG was set at a few seconds before the first arrival and WIN_END was set based on the estimated surface-wave arrival time.

Step (2): For a wave packet obtained from the observed seismogram, we reject it if its signal-to-noise ratio is too low, that is, $SNR_A < SNRA2$ or $SNR_E < SNRE2$, where SNR_A and SNR_E are computed for the wave packet using eqs (1) and (2) and $SNRA2$ and $SNRE2$ are two pre-set parameters (Table 1), whose values are selected based on the specific application.

Step (3): We reject a pair of synthetic and observed wave packets if the waveform similarity between them is too low. In our algorithm, the waveform similarity is measured using three quantities: the normalized correlation coefficient (NCC), the cross-correlation traveltime shift ΔT and the natural log of the cross-correlation amplitude ratio $\Delta \ln A$. In our algorithm, the NCC is defined as

$$NCC = \frac{\int \tilde{u}(t - \Delta T)u(t) dt}{\sqrt{\int \tilde{u}^2(t - \Delta T) dt \int u^2(t) dt}}, \quad (6)$$

where $\tilde{u}(t)$ and $u(t)$ are the wave packets obtained from the observed and the corresponding synthetic seismograms, and ΔT is the amount of time-shift that maximizes the NCC value. The cross-correlation amplitude ratio is defined as (Ritsema *et al.* 2002),

$$A = \frac{\int \tilde{u}(t - \Delta T)u(t) dt}{\int u^2(t) dt}, \quad (7)$$

and we use the natural log of A to represent the amplitude discrepancy between the observed and synthetic waveforms. In our algorithm, we reject a wave packet pair if NCC is below a pre-set threshold or any one of the $|\Delta T|$ and $|\Delta \ln A|$ measurements is above the pre-set threshold. The criteria for rejecting body waves and surface waves may differ. In practice, the value of the NCC threshold for surface waves is usually smaller than that for body waves and the value of the $|\Delta T|$ and $|\Delta \ln A|$ thresholds for surface waves are larger than that for body waves.

2.4 Misfit measurement selection

For each accepted pair of observed and synthetic waveforms, we extract the frequency-dependent phase and amplitude misfit measurements using the multitaper technique (Thomson 1982; Laske & Masters 1996). It is also possible to use other types of methods, such as the generalized seismological data functionals (Gee & Jordan 1992; Chen *et al.* 2010) and some continuous-time techniques (e.g. Kristeková *et al.* 2006; Fichtner *et al.* 2008), to extract frequency-dependent phase and amplitude misfit measurements. As long as the corresponding sensitivity kernels are computed using the correct misfit measurement operators (Chen *et al.* 2007b), those different types of frequency-dependent measurements should work equally well in practice. In our algorithm, we use the first five of the 2.5π -prolate tapers (Laske & Masters 1996). The error estimates for the phase and amplitude misfit measurements can be obtained using a jack-knife procedure. It is possible to have cycle-skipping errors in the phase misfit measurements. We correct for possible cycle-skipping errors by bootstrapping the phase from the dominant frequency bilaterally towards lower and higher frequencies using the broad-band cross-correlation traveltime misfit ΔT as the anchor.

Using those frequency-dependent phase and amplitude misfit measurements, we can construct a transfer function, which can be used to correct the synthetic waveform to match the observed waveform. The NCC value between the observed waveform and the corrected synthetic waveform provides us a criterion (NCC_corr in Table 1) for rejecting erroneous measurements and in practice we usually assign NCC_corr a value larger than 0.9. At frequencies close to the lower and upper corner frequencies, the measurements may become unstable due to lower signal-to-noise ratios at those frequencies. We introduce another criterion, SNR_f , which is defined as the ratio between the amplitude of the waveform at the measurement frequency and the maximum amplitude of the waveform within the frequency band. We reject the misfit measurements if they are made at frequencies at which SNR_f for the synthetic or the observed waveform is below a threshold value. In Fig. 5, we illustrate the measurement process and the measurement selection process using two surface wave examples, one from a local data set and the other from a teleseismic data set.

Examples of the pre-set parameters used for defining the criteria in our measurement selection process are given in Table 1. It is important to note that the appropriate values for those parameters may differ for different data sets or for the same data set but at different stages of the iterative inversion process.

3 EXAMPLES

In this section, we demonstrate the usefulness of our algorithm using seismograms recorded at local and teleseismic distances. The local data set is collected in Southern California and includes local earthquake seismograms recorded by the California Integrated Seismic Network (CISN) and ambient-noise Green's functions for CISN station pairs. Our algorithm has been successfully applied in a waveform inversion for the centroid moment tensors (CMTs) of local earthquakes (Lee *et al.* 2011) and in a full 3-D waveform tomography for crustal structure in Southern California, which is still in progress (Lee *et al.* 2012).

3.1 CMT inversion for local earthquakes

Our CMT inversion algorithm is based on the use of receiver-side Green's tensors (RGTs), which are the spatial-temporal strain fields produced by three orthogonal unit impulsive forces acting at the receiver location (Zhao *et al.* 2006). We computed the RGTs for 219 CISN stations using a tomographically improved version of the 3-D SCEC Community Velocity Model (CVM4.0; Lee *et al.* 2012) and a staggered-grid finite difference code (Olsen 1994).

The initial source parameters used for our CMT inversions and the earthquake ID presented in the figures were obtained from the SCEC earthquake catalogue. Synthetic seismograms for any earthquakes in our modelling volume can be computed by extracting a small, source-centred volume from our RGT database and applying the reciprocity principle (Aki & Richards 2002). We have developed an automated algorithm that searches for optimal CMT parameters by minimizing the misfit between synthetic and observed waveforms. The details of our CMT inversion algorithm are documented in Lee *et al.* (2011). Here we use the 2008 July 29 M_w 5.4 Chino Hills earthquake as an example to illustrate our automated seismogram segmentation and waveform selection process in our CMT inversion algorithm (Fig. 6).

In the pre-processing step, we first rejected noisy observed seismograms with SNR below 3. After removing the mean, the linear

Figure 5. Examples of measuring frequency-dependent phase and amplitude misfits using seismograms recorded at local (left-hand panel) and teleseismic (right-hand panel) distances. (a and b) Observed (solid lines) and synthetic (dashed lines) seismograms, the vertical bars indicate the boundaries of the selected waveforms. (c) The top traces show the selected observed (solid line) and synthetic (dashed line) waveforms, the bottom traces show the fit between the observed waveform (solid line) and the corrected synthetic waveform (dashed line) obtained by applying the transfer function determined by the frequency-dependent phase and amplitude misfit measurements shown in (g) and (h). (d) The frequency-domain amplitude spectrum of the observed (solid line) and synthetic (dashed line) waveforms, the horizontal solid and dashed lines indicate 33 per cent of the peak amplitudes of the observed and synthetic waveforms, respectively. (e) Similar to (c). (f) Similar to (d). (g) Frequency-dependent phase-delay of the observed waveform relative to the synthetic waveform (solid line), horizontal dotted line indicates the broad-band cross-correlation time-shift, outside the two vertical dashed lines the amplitudes of the synthetic or observed waveforms fall below 33 per cent of the peak amplitudes. (h) Frequency-dependent amplitude misfit measurements, the horizontal dotted line indicates the natural log of the broad-band cross-correlation amplitude ratio, the two vertical dashed lines have the same meaning as in (g). (i) Similar to (g). (j) Similar to (h).

trend and tapering, we applied a fourth-order Butterworth low-pass filter with a corner period at 5 s to the observed seismograms and the synthetic Green's functions.

The selected observed seismograms were then segmented using the algorithm described in Section 2.2. At this point we rejected a set of wave packets obtained by segmenting the observed seismogram based on the first two criteria listed in Section 2.3. An example of this process is shown in Fig. 6(b). First, we rejected all the observed wave packets arriving 5 s before the estimated *P*-wave arrival time, which was provided by the seismic network, and all the observed wave packets arriving 30 s after the predicted surface-wave arrival time, which was estimated using the source-receiver distance and an estimated average velocity of 3 km s⁻¹ for surface waves in Southern California. We also rejected all the observed wave packets with temporal lengths less than the corner period, 5 s. Secondly, all the wave packets with SNR below 5 were also rejected. In our source parameter inversions, we introduced an additional criterion to select observed wave packets. The amplitude information is usually more sensitive to focal mechanisms than the phase information. So in the third step, among the remaining observed wave packets, we picked the ones with larger amplitudes, which usually have higher signal-to-noise ratios, and the first-arriving wave packet,

which carries polarity information. For our data set, the observed wave packets with larger amplitudes are usually *S* waves or surface waves. The first-arriving wave packets are usually *P* waves or *Pn* waves (Helmberger & Engen 1980).

Our CMT inversion algorithm is composed of two steps, a grid-search step and a gradient-based optimization step. In the grid-search step we search for the optimal focal mechanism (i.e. strike, dip, rake) and hypocentre location that maximize the *NCC* and minimize the $|\Delta T|$ and $|\Delta \ln A|$ measurements between observed and synthetic waveforms. In the gradient-based optimization step, we further refine the solution obtained in the grid-search step by minimizing the frequency-dependent phase and amplitude misfit measurements between synthetic and observed waveforms.

During the grid-search step, the segmentation algorithm is not applied to the synthetic seismogram. Instead, the time-frequency domain windows (i.e. the mask files) corresponding to the selected observed wave packets were applied to the synthetic seismogram directly. The resulting synthetic wave packets were then used for computing the *NCC* and the $|\Delta T|$ and $|\Delta \ln A|$ measurements with their corresponding observed wave packets. Errors in the seismic velocity model can cause a time-shift between observed and

Figure 6. (a) The best-fitting double-couple component of the optimal CMT solution obtained for the 2008 July 29 M_w 5.4 Chino Hills earthquake. (b) An example of the rejection process for the observed wave packets, the numbers (1), (2) and (3), correspond the three rejection steps explained in the text. (c) Examples of the observed and synthetic wave packets used in our gradient-based optimization step, black lines are observed seismograms and the red lines are synthetic seismograms calculated using the optimal CMT solution, the black horizontal bars indicate the selected waveforms.

synthetic waveforms. A common strategy for dealing with this problem is to allow a time-shift while minimizing waveform misfit (e.g. Zhao & Helmberger 1994; Zhu & Helmberger 1996). In our case, we allowed a maximum of ± 2.5 s time-shift (e.g. Liu *et al.* 2004; Lee *et al.* 2011) of the time–frequency domain windows applied on the synthetic seismogram while maximizing *NCC* in our grid-search step. The details of our grid-search algorithm are documented in Lee *et al.* (2011).

The CMT solution obtained in the grid-search step was used as the reference solution in the gradient-based optimization step. The details of our gradient-based optimization algorithm are documented in Zhao *et al.* (2006). In this step, the synthetic seismograms computed using the reference CMT solution were segmented and the synthetic and observed waveform pairs were selected according to the procedure discussed in Section 2.3. The frequency-dependent phase and amplitude misfit measurements were extracted for those selected waveform pairs. In Zhao *et al.* (2006) and Lee *et al.* (2011) we used the generalized seismological data functionals as our frequency-dependent misfit measurements. The solution usually converged after less than 10 iterations of a steepest descent algorithm. We did not resegment the synthetics in each iteration. An example of the obtained CMT solution (only the best-fitting double-couple component is shown), as well as examples of the synthetic and observed wave packets used in the gradient-based optimization for the Chino Hills earthquake is shown in Figs 6(a) and (c).

3.2 Full 3-D waveform tomography for local crustal structure

To demonstrate the usefulness of our seismogram segmentation and waveform selection algorithm in full 3-D waveform tomography, we use a local waveform data set collected in Southern California (Lee *et al.* 2012) as an example. Our data set is composed of both waveform recordings of local small to medium-sized earthquakes and ambient-noise Green's functions.

The ambient-noise Green's functions were provided by Marine Denolle (personal communication) and the technique for obtaining this data set is documented in Prieto *et al.* (2011). Ambient-noise Green's functions can be used in full 3-D waveform tomography just like earthquake recordings (e.g. Tromp *et al.* 2010; Xu *et al.* 2013). They allow us to take advantage of the density of a seismic network without waiting for earthquakes to occur and in practice they often complement earthquake recordings and can substantially improve the data coverage in tomographic inversions (e.g. Shapiro *et al.* 2005; Yang *et al.* 2007; Bensen *et al.* 2008). In our inversion, the target waveforms on the ambient-noise Green's functions are the vertical-component Rayleigh waves. Examples of the selected waveforms on the ambient-noise Green's functions and the corresponding synthetic Green's functions are shown in Fig. 8. The station 'RCT' is treated as the 'source' station. The interstation synthetic Green's functions were computed using a 3-D staggered-grid finite-difference code (Olsen 1994) by placing a unit impulsive vertical force at station 'RCT'. The reference earth structure model used in the calculation is a tomographically updated version of the 3-D SCEC CVM4.0 (Lee *et al.* 2012).

In the pre-processing step, we rejected noisy ambient-noise Green's functions if any of the SNR values is less than 3. Both the synthetic Green's functions and the ambient-noise Green's functions were band-pass filtered using a fourth-order Butterworth filter with corners at 9 and 5.5 s. After an ambient-noise Green's

function was segmented, we applied the first two criteria in Section 2.3 and rejected all wave packets arriving 15 s before or 30 s after the estimated surface wave arrival time. Among the remaining wave packets, we picked the one with the largest amplitude. The same procedure was also applied on the corresponding synthetic Green's function to obtain the synthetic surface wave packet. Frequency-dependent group-delay misfits were then measured at a set of frequencies evenly distributed across the frequency band. Fig. 7 shows the geographic distribution of the group-delay measurements at the dominant frequency. In this example, the measurements indicate that the reference shear-wave speed west of 'RCT' around the Coast Ranges and in the southern Great Valley is too fast and should be reduced while the reference shear-wave speed east of 'RCT' in the Sierra Nevada region is too slow and should be increased.

The way that we process earthquake recordings is different from the procedure used for processing ambient-noise Green's functions and also slightly different from the procedure used in our CMT inversions. Some examples from an earthquake (event ID: 9095528) are shown in Fig. 8. After the pre-processing step and segmenting both the observed and synthetic seismograms, we followed the three steps for waveform selection as discussed in Section 2.3. The parameters used in our procedure are listed in Table 2. In Fig. 8(a), the wave packets arriving 30 s after the estimated surface wave arrival time were rejected in step (1). The remaining two wave packets all satisfied the rest of the selection criteria. In Fig. 8(b), after rejecting wave packets arriving too late in the seismogram, there were three wave packets remaining. The wave packet arriving at around 60 s was rejected in the last step because the *NCC* value was below the threshold. In Fig. 8(c), the wave packet arriving at around 80 s was rejected in the last step because it did not meet any of the criteria for *NCC*, $|\Delta T|$ and $|\Delta \ln A|$. In Fig. 8(d), only one wave packet was selected because the rest of the wave packets did not meet the required SNR thresholds in step (2).

3.3 Global full 3-D waveform tomography

Our algorithm can also be applied to seismograms recorded at regional to teleseismic distances once the values for the pre-set parameters are adjusted accordingly (Table 2). Fig. 9 shows examples of the regional/teleseismic seismograms for the 2007 July 26 M_w 6.9 northern Molucca sea earthquake. The synthetic seismograms were computed using the SPECfEM3D_GLOBE code version 5.1 (Komatitsch & Tromp 2002). The 3-D earth structure model was obtained by combining the 3-D S40RTS mantle model (Ritsema *et al.* 2011), the 3-D CRUST2.0 crustal model (Bassin *et al.* 2000) and the 1-D PREM attenuation model (Dziewonski & Anderson 1981). The source parameters used in the calculations were obtained from the Global CMT catalogue (<http://www.globalcmt.org>). The synthetics were accurate to a shortest period of about 45.33 s.

In the pre-processing step, a fourth-order Butterworth band-pass filter with corners at 5 and 20 mHz was applied on both the observed and the synthetic seismograms. At teleseismic distances, some useful phases may arrive after the minor-arc surface wave; therefore in step (1) of our waveform selection process, we removed the criterion for rejecting wave packets arriving after the minor-arc surface waves. The minimum temporal length for accepting a wave packet was increased to 50 s, which was the corner period of the filter. Considering the accuracy of the starting model and the general fit between observed and synthetic waveforms, the minimum *NCC* value needed for accepting a wave packet is reduced to 0.65 for body

Figure 7. Examples of the ambient-noise Green's functions used in local full 3-D tomography. Grey star: the location of the 'source' station, RCT; coloured triangles: other 'receiver' stations for which the ambient-noise Green's functions have acceptable signal-to-noise ratios, the colour on those triangles indicates the group-delay measurements at the dominant frequency. Black lines: ambient-noise Green's functions; red lines: synthetic Green's functions; the black and red horizontal bars indicate the waveforms selected by our automated algorithm.

Figure 8. Examples of the waveform selection process as discussed in Section 2.4. The seismograms are from an earthquake (event ID: 9095528) and are used in our full 3-D waveform tomography for crustal structure in Southern California. In each example, the first vertical dash line indicates the time 5 s before the estimated P arrival time and the second vertical dash line indicates the time 30 s after the estimated surface wave arrival time. The numbers, (1), (2) and (3), correspond to the three different steps discussed in Section 2.4.

waves and to 0.5 for surface waves. The allowed maximum cross-correlation time-shift was increased to 20 s for body waves and to 25 s for surface waves. In general, the number of waveforms picked automatically by our algorithm increases with source–receiver distance. At station CTAO, which is 29° from the epicentre, the number of picked waveforms is 8; while at station DRLN, which is 128° from the epicentre, the number of picked waveforms is 24 (Fig. 9).

4 SUMMARY AND DISCUSSION

The purpose of this study is to develop a semi-automatic seismic waveform analysis tool that can be used in full 3-D waveform inversions and other types of seismological studies that may require waveform separation and selection. Our algorithm consists of three major components: seismogram pre-processing, seismogram segmentation and waveform selection. Our seismogram

Figure 9. Examples of teleseismic earthquake recordings. Yellow star: epicentre location of the 2007 July 26 M_w 6.9 northern Molucca sea earthquake; red triangles: stations that have usable waveforms; blue lines: the source-station pairs whose seismograms are shown. Three histograms show the distribution of NCC , broad-band cross-correlation time-shift and amplitude ratio for all selected waveforms. Black lines: observed seismograms; red lines: synthetic seismograms; black horizontal bars: selected observed waveform; red horizontal bars: selected synthetic waveform. Dash-line boxes indicate the portions of the seismograms that have been amplified to show the waveform details.

segmentation algorithm is based on segmenting the scalogram in the time–frequency domain using a topological watershed method. A set of criteria is proposed to select observed and synthetic wave packets that can be used for extracting waveform misfit measurements. We have demonstrated the usefulness of our algorithm using earthquake seismograms and ambient-noise Green’s functions at local, regional and teleseismic distances in waveform inversions for CMT parameters and earth structure models. Using those examples, we also gave some suggested values for the parameters that control the behaviour of our algorithm for different types of seismic data.

The performance of our seismogram segmentation algorithm depends upon both the seismogram being segmented and the resolution of the time–frequency domain scalogram. An important limitation of our algorithm is that it would not be able to separate different phases if they have significant overlap in the time–frequency domain. If that is the case, we may need to explore additional conditions that can be used for wave arrival separation. The resolution of the time–frequency domain scalogram is determined by the wavelet transform and the mother wavelet. Other types of time–frequency transforms such as the short-time Fourier transform can also be adopted. Depending upon how well the different phases are separated in time and in frequency, one can improve the resolution in time or in frequency by adjusting the mother wavelet. However, any improvement of resolution in time (frequency) can reduce the resolution in frequency (time). For wave arrivals that are well separated in the time domain, we do not expect our algorithm to provide results that are significantly different from FLEXWIN.

At the current stage, the appropriate values for the pre-set parameters listed in Table 1 still need to be assigned or fine-tuned manually through a trial-and-error process for different data sets or different applications. In recent years, the artificial neural network (ANN) has been adopted to perform certain seismic data processing tasks with different degrees of success (e.g. Shimshoni & Intrator 1998; Del Pezzo *et al.* 2003; Scarpetta *et al.* 2005; Dai & MacBeth 2007; Valentine & Woodhouse 2010). In Diersen *et al.* (2011), we applied an importance-aided neural network (IANN) to the waveform selection process and tested it against a set of manual picks using the ambient-noise Green’s functions collected in Southern California. A set of 15 criteria, including those listed in Section 2.4, was built into the IANN with different weights. This IANN was trained using 1250 manual picks (about one tenth of our total ambient-noise Green’s function data set) and then tested using 504 examples previously unseen by the IANN. Out of 504 test examples, the IANN gave zero false negatives and two false positives (i.e. accepted two wave packets that were rejected in the manual pick), a success rate of about 99.6 per cent. More experiments on different types of data sets still need to be conducted to fully test the robustness of the IANN implementation.

ACKNOWLEDGEMENTS

The continuous wavelet transform code used in this study was obtained from the Geophysical Wavelet Library at <http://users.math.uni-potsdam.de/~gwl/> and the topological watershed code used in this study was obtained from <http://www.esiee.fr/~info/tw/index.html>. The finite-difference wave-propagation simulations used in this study were performed at the Argonne Leadership Computing Facility (ALCF), which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357. We thank Marine Denolle and Gregory C. Beroza for providing us the ambient-noise

Green’s functions in Southern California. The first author of this paper is supported by the School of Energy Resources at University of Wyoming. We appreciate the Editor, Dr. Andrea Morelli, and four anonymous reviews’ careful reviews and helpful comments.

REFERENCES

- Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, University Science Books, Sausalito, California.
- Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.*, **81**, F897.
- Bensen, G.D., Ritzwoller, M.H. & Shapiro, N.M., 2008. Broad-band ambient noise surface wave tomography across the United States, *J. geophys. Res.*, **113**, B05306, doi:10.1029/2007JB005248.
- Bertrand, G., 2005. On topological watersheds, *J. Math. Imag. Vis.*, **22**, 217–230.
- Beucher, B. & Lantuejoul, Ch., 1979. Use of watersheds in contour detection, in *Proceedings of International Workshop on Image Processing, Real-Time Edge and Motion Detection/Estimation*, Rennes, France.
- Bunks, C., Saleck, F., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**, 1457–1473.
- Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances, *Pure appl. Geophys.*, **168**(10), 1527–1552.
- Chen, P., Jordan, T. & Zhao, L., 2005. Finite-moment tensor of the 3 September 2002 Yorba Linda earthquake, *Bull. seism. Soc. Am.*, **95**(3), 1170–1180.
- Chen, P., Zhao, L. & Jordan, T.H., 2007a. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.*, **97**(4), 1094–1120.
- Chen, P., Jordan, T.H. & Zhao, L., 2007b. Full three-dimensional tomography: a comparison between the scattering-integral and adjoint-wavefield methods, *Geophys. J. Int.*, **170**, 175–181.
- Chen, P., Jordan, T. & Lee, E., 2010. Perturbation kernels for generalized seismological data functionals (GSDF), *Geophys. J. Int.*, **183**(2), 869–883.
- Chernov, L.A., 1960. *Wave Propagation in a Random Medium*, McGraw-Hill, New York, 168pp.
- Coupré, M. & Bertrand, G., 1997. Topological gray-scale watershed transformation, in *SPIE Vision Geometry VI Proceedings*, San Diego, CA, Vol. 3168, doi:10.1117/12.292778.
- Coupré, M., Najman, L. & Bertrand, G., 2005. Quasi-linear algorithms for the topological watershed, *J. Math. Imag. Vis.*, **22**, 231–249.
- Dai, H. & MacBeth, C., 2007. Automatic picking of seismic arrivals in local earthquake data using an artificial neural network, *Geophys. J. Int.*, **120**(3), 758–774.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton University Press, New Jersey.
- Debayle, E., 1999. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion, *Geophys. J. Int.*, **137**(3), 747–754.
- Del Pezzo, E., Esposito, A., Giudicepietro, F., Marinaro, M., Martini, M. & Scarpetta, S., 2003. Discrimination of earthquakes and underwater explosions using neural networks, *Bull. seism. Soc. Am.*, **93**(1), 215–223.
- Diersen, S., Lee, E.-J., Spears, D., Chen, P. & Wang, L., 2011. Classification of Seismic windows using artificial neural networks, in *Proceedings of the International Workshop on Data Mining in Earth System Science (DMESS 2011), in Conjunction with International Conference on Computational Science (ICCS)*, Tsukuba, Japan, Elsevier.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.
- Fichtner, A., Kennett, B., Igel, H. & Bunge, H., 2008. Theoretical background for continental and global-scale full-waveform inversion in the time–frequency domain, *Geophys. J. Int.*, **175**, 666–685.
- Fichtner, A., Kennett, B., Igel, H. & Bunge, H., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**, 1703–1725.

- Graves, R., 1996. Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences, *Bull. seism. Soc. Am.*, **86**, 1091–1106.
- Gee, L. & Jordan, T., 1992. Generalized seismological data functionals, *Geophys. J. Int.*, **111**(2), 363–390.
- Helmberger, D.V. & Engen, G.R., 1980. Modeling the long-period body waves from shallow earthquakes at regional ranges, *Bull. seism. Soc. Am.*, **70**, 1699–1714.
- Hingee, M., Tkalcic, H. & Fichtner, A., 2011. Seismic moment tensor inversion using a 3-D structural model: applications for the Australian region, *Geophys. J. Int.*, **184**, 949–964.
- Holschneider, M., 1995. *Wavelets: An Analysis Tool*, Univ. Press, Oxford.
- Keller, J.B., 1969. Accuracy and validity of the Born and Rytov approximations, *J. Opt. Soc. Am.*, **59**, 1003–1004.
- Kim, Y.H., Liu, Q. & Tromp, J., 2011. Adjoint centroid-moment tensor inversions, *Geophys. J. Int.*, **186**, 264–278.
- Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.*, **49**, 390–412.
- Kristeková, M., Kristek, J., Moczo, P. & Day, S.M., 2006. Misfit criteria for quantitative comparison of seismograms, *Bull. seism. Soc. Am.*, **96**(5), 1836–1850.
- Kulesh, M., Diallo, M. & Holschneider, M., 2005. Wavelet analysis of ellipticity, dispersion, and dissipation properties of Rayleigh waves, *Acoust. Phys.*, **51**(4), 425–434.
- Kulesh, M., Holschneider, M. & Diallo, M., 2008. Geophysical wavelet library: applications of the continuous wavelet transform to the polarization and dispersion analysis of signals, *Comp. Geosci.*, **34**(12), 1732–1752.
- Laske, G. & Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data, *J. geophys. Res.*, **101**(B7), 16 059–16 075.
- Lawrence, J.F. & Shearer, P.M., 2008. Imaging mantle transition zone thickness with SdS-SS finite-frequency sensitivity kernels, *Geophys. J. Int.*, **174**, 143–158.
- Lebedev, S., Nolet, G., Meier, T. & van der Hilst, R., 2005. Automated multimode inversion of surface and S waveforms, *Geophys. J. Int.*, **162**, 951–964.
- Lee, E., Chen, P., Jordan, T. & Wang, L., 2011. Rapid full-wave centroid moment tensor (CMT) inversion in a three-dimensional earth structure model for earthquakes in Southern California, *Geophys. J. Int.*, **186**, 311–330.
- Lee, E., Chen, P., Jordan, T.H., Maechling, P.J., Denolle, M. & Beroza, G.C., 2012. Full-3D waveform tomography for Southern California, in *Proceedings of the 2012 AGU Fall Meeting Abstract*, AGU.
- Levshin, A.L. & Ritzwoller, M.H., 2001. Automated detection, extraction, and measurement of regional surface waves, *Pure appl. Geophys.*, **158**, 1531–1545.
- Liu, Q. & Gu, Y.J., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics*, **566–567**, 31–66.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.*, **96**(6), 2383–2397.
- Liu, Q., Polet, J., Komatitsch, D. & Tromp, J., 2004. Spectral-element moment tensor inversions for earthquakes in Southern California, *Bull. seism. Soc. Am.*, **94**(5), 1748–1761.
- Maggi, A., Tape, C., Chen, M. & Chao, D., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**, 257–281.
- Meyer, F., 1992. Un algorithme optimal de ligne de partage des eaux, in *Proceedings of the 8th Conferences on Reconnaissance des FReconnaisance des Formes et Intelligence Artificielle*, Vol. 2, pp. 847–859, AFCET Ed., Lyon.
- Olsen, K., 1994. *Simulation of three-dimensional wave propagation in the Salt Lake Basin*, University of Utah.
- Panning, M. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.*, **167**, 361–379.
- Prieto, G.A., Denolle, M., Lawrence, J.F. & Beroza, G.C., 2011. On amplitude information carried by the ambient seismic field, *Compt. Rend. Geosci.*, **343**, 600–614.
- Ritsema, J. & van Heijst, H.-J., 2002. Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKP traveltimes, *Geophys. J. Int.*, **149**(2), 482–489.
- Ritsema, J., Deuss, A., Van Heijst, H.J. & Woodhouse, J.H., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.*, **184**, 1223–1236.
- Ritsema, J., Rivera, L., Komatitsch, D., Tromp, J. & van Heijst, H.-J., 2002. Effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios, *Geophys. Res. Lett.*, **29**, doi:10.1029/2001GL013831.
- Scarpetta, S., Giudicepietro, F., Ezin, E.C., Petrosino, S., Del Pezzo, E., Martini, M. & Marinaro, M., 2005. Automatic classification of seismic signals at Mt. Vesuvius volcano, Italy, using neural networks, *Bull. seism. Soc. Am.*, **95**(1), 185–196.
- Shapiro, N., Campillo, M., Stehly, L. & Ritzwoller, M., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.
- Shimshoni, Y. & Intrator, N., 1998. Classification of seismic signals by integrating ensembles of neural networks, *IEEE Trans. Signal Process.*, **46**(5), 1194–1201.
- Sigloch, K. & Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes, *Geophys. J. Int.*, **167**, 217–287.
- Snieder, R. & Lomax, A., 1996. Wavefield smoothing and the effect of rough velocity perturbations on arrival times and amplitudes, *Geophys. J. Int.*, **125**, 796–812.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2010. Seismic tomography of the southern California crust based on spectral-element and adjoint methods, *Geophys. J. Int.*, **180**, 433–462.
- Tarantola, A., 1988. Theoretical background for the inversion of seismic waveforms including elasticity and attenuation, *Pure appl. Geophys.*, **128**, 365–399.
- Thomson, D.J., 1982. Spectrum estimation and harmonic analysis, *Proc. IEEE*, **70**(9), 1055–1096.
- Trampert, J. & Woodhouse, J.H., 1995. Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.*, **122**(2), 675–690.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Tromp, J., Luo, Y. & Hanasoge, S., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.*, **183**, 791–819.
- Valentine, A.P. & Woodhouse, J.H., 2010. Approaches to automated data selection for global seismic tomography, *Geophys. J. Int.*, **183**, 1001–1012.
- Vincent, L. & Soille, P., 1991. Watersheds in digital spaces: an efficient algorithm based on immersion simulations, *IEEE Trans. Pattern Anal. Mach. Intell.*, **13**, 583–598.
- Visser, K., Lebedev, S., Trampert, J. & Kennett, B.L.N., 2007. Global love wave overtone measurements, *Geophys. Res. Lett.*, **34**, L03302, doi:10.1029/2006GL028671.
- Woodward, M., 1989. A qualitative comparison of the first order Born and Rytov approximations, SEP-60: Stanford Exploration Project, pp. 203–214.
- Xu, Z., Chen, P. & Chen, Y., 2013. Sensitivity kernel for the weighted norm of the frequency-dependent phase correlation, *Pure appl. Geophys.*, **170**, 353–371.
- Yang, Y., Ritzwoller, M. & Levshin, A., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.*, **168**, 259–274.
- Zhao, L. & Helmberger, D., 1994. Source estimation from broadband regional seismograms, *Bull. seism. Soc. Am.*, **84**, 91–104.
- Zhao, L., Jordan, T., Olsen, K. & Chen, P., 2005. Frechet kernels for imaging regional earth structure based on three-dimensional reference models, *Bull. seism. Soc. Am.*, **95**(6), 2066–2080.
- Zhao, L., Chen, P. & Jordan, T., 2006. Strain Green's tensors, reciprocity, and their applications to seismic source and structure studies, *Bull. seism. Soc. Am.*, **96**, 1753–1763.

- Zhu, H., Bozdog, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.*, **5**(7), 493–498.
- Zhu, L. & Helmberger, D., 1996. Advancement in source estimation techniques using broadband regional seismograms, *Bull. seism. Soc. Am.*, **86**, 1634–1641.
- Zhou, Y., Liu, Q. & Tromp, J., 2011. Surface wave sensitivity: mode summation versus adjoint SEM, *Geophys. J. Int.*, **187**(3), 1560–1576.

APPENDIX A

The topological watershed algorithm proceeds by iteratively lowering the altitude of some points satisfying a destructibility condition until no such points exist. Suppose $G = (E, \Gamma, F)$ is a weighted graph, where E is the set of vertices (points), Γ corresponds to an adjacency relation between the points in E and F is a function representing a grey-scale mapping, that is, for any point $x \in E$, $F(x)$ is the grey level (altitude) of the point x . Two points $p_0, p_k \in E$ are said to be connected if there exists an ordered family of points (i.e. a path) $(p_0, p_1, \dots, p_k)$ such that $p_{i+1} \in \Gamma(p_i)$. A subset $X \subseteq E$ is said to be a connected component of E if any two points in X are connected. A level k cross-section of F is defined as $F_k = \{p \in E, F(p) \geq k\}$. A point x is said to be W-destructible for F if its altitude can be lowered by one unit without changing the number of connected components of the level-set $\bar{F}_k = \{p \in E, F(p) < k\}$, with $k = F(x)$. A W-thinning of F is a map obtained by iteratively selecting a W-destructible point and lowering its altitude by one unit. A topological watershed of F is a W-thinning of F that contains no W-destructible points. A fast implementation can be obtained if a W-destructible point is not lowered by only one altitude unit at a time, but by an amount as large as possible. The lowest value to which a W-destructible point can be lowered can be computed by searching for the lowest common ancestor of neighbouring components in a component tree (i.e. a hierarchical representation of the connected components at cross-sections of increasing altitude; Couprie *et al.* 2005).

APPENDIX B

Full 3-D sensitivity (Fréchet) kernels with respect to structural parameters can be computed using the scattering-integral (SI) method (e.g. Zhao *et al.* 2005) or the adjoint wavefield (AW) method (e.g. Tromp *et al.* 2005; Liu & Tromp 2006) for an arbitrary segment on the time-domain seismogram. The connection between the SI and AW methods, as well as the differences in their computational requirements, were documented in Chen *et al.* (2007b). In the following we show how the Fréchet kernels can be constructed for an arbitrary target waveform obtained through a filtering operation in the wavelet domain. The notations follow Chen *et al.* (2007b).

For a certain data functional d_{in}^{sr} that measures the misfit between the observed and the synthetic waveforms, we introduce the ‘seismogram perturbation kernel’ $J_{in}^{sr}(t)$, which is the integration kernel for the Fréchet derivative of d_{in}^{sr} with respect to the target waveform $u_i^s(\mathbf{x}_r, t)$,

$$\delta d_{in}^{sr} = \int_{-\infty}^{+\infty} J_{in}^{sr}(t) \delta u_i^s(\mathbf{x}_r, t) dt, \quad (\text{B1})$$

where s, r, i, n are indices for source, receiver, component and measurement. Examples of data functionals and their corresponding seismogram perturbation kernels were given in Chen *et al.* (2007b). The derivation of the adjoint source, which is very close to our

definition of the seismogram perturbation kernel, for multitaper measurements can be found in Zhou *et al.* (2011). Suppose the target waveform $u_i^s(\mathbf{x}_r, t)$ can be obtained from the complete displacement seismogram $s_i^s(\mathbf{x}_r, t)$ through a filtering operation in the wavelet domain using a weighting function $W(a, b)$, which can be obtained using the topological watershed algorithm. Considering the forward and inverse continuous wavelet transform in eqs (3)–(4), we have

$$\begin{aligned} u_i^s(\mathbf{x}_r, t) &= \frac{2}{C_g} \text{Re} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a} g\left(\frac{t-b}{a}\right) W(a, b) W_g s_i^s(\mathbf{x}_r, b, a) \frac{db da}{a} \right] \\ &= \frac{2}{C_g} \text{Re} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a} g\left(\frac{t-b}{a}\right) W(a, b) \right. \\ &\quad \times \left. \int_{-\infty}^{+\infty} \frac{1}{a} g^*\left(\frac{t'-b}{a}\right) s_i^s(\mathbf{x}_r, t') dt' \frac{db da}{a} \right]. \end{aligned} \quad (\text{B2})$$

Bring eq. (B2) into (B1) and exchange the integration over t and the integration over t' , we obtain,

$$\begin{aligned} \delta d_{in}^{sr} &= \int_{-\infty}^{+\infty} J_{in}^{sr}(t) \delta u_i^s(\mathbf{x}_r, t) dt \\ &= \int_{-\infty}^{+\infty} dt J_{in}^{sr}(t) \frac{2}{C_g} \text{Re} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{a} g\left(\frac{t-b}{a}\right) W(a, b) \right. \\ &\quad \times \left. \int_{-\infty}^{+\infty} \frac{1}{a} g^*\left(\frac{t'-b}{a}\right) \delta s_i^s(\mathbf{x}_r, t') dt' \frac{db da}{a} \right] \\ &= \int_{-\infty}^{+\infty} \delta s_i^s(\mathbf{x}_r, t') dt' \frac{2}{C_g} \text{Re} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} W(a, b) \frac{1}{a} g^*\left(\frac{t'-b}{a}\right) \right. \\ &\quad \times \left. \int_{-\infty}^{+\infty} \frac{1}{a} g\left(\frac{t-b}{a}\right) J_{in}^{sr}(t) dt \frac{db da}{a} \right] \\ &= \int_{-\infty}^{+\infty} L_{in}^{sr}(t) \delta s_i^s(\mathbf{x}_r, t) dt, \end{aligned} \quad (\text{B3})$$

where, in the last step, we have changed the variables $t \leftrightarrow t'$ and the integration kernel $L_{in}^{sr}(t)$ is defined as,

$$\begin{aligned} L_{in}^{sr}(t) &\equiv \frac{2}{C_g} \text{Re} \left[\int_0^{+\infty} \int_{-\infty}^{+\infty} W(a, b) \frac{1}{a} g^*\left(\frac{t-b}{a}\right) \right. \\ &\quad \times \left. \int_{-\infty}^{+\infty} \frac{1}{a} g\left(\frac{t'-b}{a}\right) J_{in}^{sr}(t') dt' \frac{db da}{a} \right]. \end{aligned} \quad (\text{B4})$$

The perturbation of the displacement seismogram $\delta s_i^s(\mathbf{x}_r, t')$ can be related to the structural parameters, density $\rho(\mathbf{x})$ and elastic moduli $c_{jklm}(\mathbf{x})$, through the Born approximation (Dahlen & Tromp 1998). Using the SI method, the Fréchet kernels with respect to density and

Figure B1. Full 3-D P -velocity sensitivity kernels of the broad-band cross-correlation traveltimes for the waveforms shown in Figs 4(h)–(l). Cross-section views perpendicular to the source-receiver plane are shown. Grey star: the explosive source location; grey triangle: receiver location. Warm colours represent negative sensitivity (i.e. an increase in P -velocity will lead to a decrease in the cross-correlation delay time) and cool colours represent positive sensitivity (i.e. an increase in P -velocity will lead to an increase in the cross-correlation delay time). (a) The kernel for the waveform within the time window 1.45–1.95 s, which is a combination of PP, pPP and the Rayleigh wave; (b) the kernel for the PP phase; (c) the kernel for the pPP phase; (d) the kernel for the Rayleigh wave.

elastic moduli are given by

$$K_{d_{in}^{sr}}^{\rho}(\mathbf{x}) = - \int dt \int d\tau L_{in}^{sr}(t) \sum_j G_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r) \partial_{\tau}^2 s_j^s(\mathbf{x}, \tau), \quad (\text{B5})$$

$$K_{d_{in}^{sr}}^{c_{ijklm}}(\mathbf{x}) = - \int dt \int d\tau L_{in}^{sr}(t) \partial_k G_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r) \partial_l s_m^s(\mathbf{x}, \tau), \quad (\text{B6})$$

where $G_{ji}(\mathbf{x}, t - \tau; \mathbf{x}_r)$ is the receiver-side Green's tensor (Zhao *et al.* 2005; Zhao *et al.* 2006). If the adjoint method is used for computing the kernels, the adjoint source (e.g. Tromp *et al.* 2005) is given by the misfit-weighted, time-reversed $L_{in}^{sr}(t)$. Examples of full 3-D Fréchet kernels of wide-band cross-correlation traveltimes with respect to P -wave speed for the Rayleigh wave, PP and pPP waves shown in Figs 4(g)–(l) are plotted in Fig. B1. The kernels were computed using the adjoint method.