

High-resolution seismic array imaging based on an SEM-FK hybrid method

Ping Tong,¹ Chin-wu Chen,² Dimitri Komatitsch,³ Piero Basini¹ and Qinya Liu¹

¹*Department of Physics, University of Toronto, Toronto, Ontario, Canada. E-mail: tongping85@gmail.com*

²*Institute of Oceanography, National Taiwan University, Taiwan*

³*LMA, CNRS UPR 7051, Aix-Marseille University, Centrale Marseille, F-13402 Marseille Cedex 20, France*

Accepted 2013 December 13. Received 2013 December 12; in original form 2013 July 3

SUMMARY

We demonstrate the feasibility of high-resolution seismic array imaging based on teleseismic recordings using full numerical wave simulations. We develop a hybrid method that interfaces a frequency–wavenumber (FK) calculation, which provides analytical solutions to 1-D layered background models with a spectral-element (SEM) numerical solver to calculate synthetic responses of local media to plane-wave incidence. This hybrid method accurately deals with local heterogeneities and discontinuity undulations, and represents an efficient tool for the forward modelling of teleseismic coda (including converted and scattered) waves. We benchmark the accuracy of the SEM-FK hybrid method against FK solutions for 1-D media. We then compute sensitivity kernels for teleseismic coda waves by interacting the forward teleseismic waves with an adjoint wavefield, produced by injecting coda waves as adjoint sources, based on adjoint techniques. These sensitivity kernels provide the basis for mapping variations in sub-surface discontinuities, density and velocity structures through non-linear conjugate-gradient methods. We illustrate various synthetic imaging experiments, including discontinuity characterization, volumetric structural inversion for the crust or subduction zones. These tests show that using pre-conditioners based upon the scaled product of sensitivity kernels for different phases, combining finite-frequency traveltime and waveform inversion, and/or adopting hierarchical inversions from long- to short-period waveforms could reduce the non-linearity of the seismic inverse problem and speed up its convergence. The encouraging results of these synthetic examples suggest that inversion of teleseismic coda phases based on the SEM-FK hybrid method and adjoint techniques is a promising tool for structural imaging beneath dense seismic arrays.

Key words: Numerical solutions; Coda waves; Seismic tomography; Wave scattering and diffraction; Wave propagation.

1 INTRODUCTION

Seismic tomography is a fundamental tool in revealing the 3-D structure of the Earth's interior at all scales (e.g. Romanowicz 1991, 2003; Rawlinson *et al.* 2010; Liu & Gu 2012). The resultant 3-D images provide crucial evidence for the tectonic evolution and internal geodynamic processes of our planet (e.g. Dziewonski 1984; Masters *et al.* 2000; Li *et al.* 2008; Fukao *et al.* 2009; Zhao *et al.* 2013). Driven by the continued demand for higher resolution images, and thanks to the proliferation and technological development of seismic instruments, seismic tomography has experienced rapid advances over the past three decades (e.g. Romanowicz 2008; Rawlinson *et al.* 2010). However, owing to the very uneven distribution of seismicity, regional tomography used to rely mainly on transmitted primary phases of global earthquakes (such as *P*; e.g. Aki *et al.* 1977), and often resolved a refined regional model embedded in a coarser global model (e.g. Spakman & Nolet 1988; Fukao *et al.* 1992; Li *et al.* 2008). In recent years, with the growing deployment of broad-band arrays and the rapid increase in data availability, traditional body-wave tomographic techniques have been adapted by measuring multichannel cross-correlation traveltimes of teleseismic waves at array stations (e.g. Vandecar & Crosson 1990), and attributing traveltime anomalies to structures beneath arrays (e.g. Sandoval *et al.* 2004; Jiang *et al.* 2009; Hung *et al.* 2011). Similarly, surface waves phase velocity maps for array regions are often inverted by regarding the incident fields as two- or multiphase waves (e.g. Friederich & Wielandt 1995; Forsyth *et al.* 1998; Yang & Forsyth 2006). Overall, array analysis has brought great vitality to the quest for regional structures around the globe (e.g. Nolet *et al.* 2005; Gu 2010; Chang & van der Lee 2011; Obrebski *et al.* 2011). On the other hand, coda waves such as converted, reflected and scattered waves that are

singly or multiply scattered by heterogeneities distributed beneath the surface, are prevalently used in the imaging of crust and upper-mantle structures beneath seismic arrays. As primary phases are affected mainly by long-wavelength structures (e.g. Wu & Toksoz 1987; Liu & Gu 2012), scattered phases are sensitive to short-wavelength structures, the use of scattered waves in seismic tomography, in principle, will offer substantially higher imaging resolution beneath seismic arrays (e.g. Bostock *et al.* 2001; Frederiksen & Revenaugh 2004; Rondenay 2009). Indeed, converted seismic waves have been used extensively to identify and characterize discontinuities of material properties in the subsurface for more than three decades since the advent of the powerful receiver function (RF) technique [e.g. Langston (1977); Vinnik (1977); Zhu & Kanamori (2000); Kind *et al.* (2012), see Rondenay (2009) for a review]. Different from point estimates of Moho depth through single station stacking (e.g. Zhu & Kanamori 2000; Yan & Clayton 2007), migration methods, prevalent in active-source exploration, map subsurface discontinuities and scatterers more generally through common conversion point (CCP) stacking, and produce 2- and 3-D stacked images of relative scattering potential beneath linear arrays (e.g. Revenaugh 1995; Sheehan *et al.* 2000). Bostock *et al.* (2001) introduced an inversion technique based on the Generalized Radon Transform to recover perturbations of elastic properties of the lithosphere beneath dense, passive seismic arrays from teleseismic coda waves. However, the principles of converted-/scattered-wave analysis are still rooted in the ray-theoretical treatment of the seismic waves equation (Rondenay 2009). These ray-based techniques may be appropriate for smoothly varying media, but are less accurate for imaging geologically and tectonically complex structures (e.g. Tape *et al.* 2009, 2010).

Shang *et al.* (2012) proposed a teleseismic reverse-time migration method to map interfaces in the crust and mantle beneath very dense seismic arrays. By making no assumptions on the presence or properties of interfaces, this wave-equation-based migration method numerically backpropagates transmitted waves from receivers and cross-correlates the *P* and *S* constituents to obtain an inverse scattering transform. Numerical solvers that capture complex wave phenomena and allow the use of smooth 3-D background models help produce more coherently migrated images compared to the RF techniques (such as CCP). However, the requirement that station spacing be less than half of the minimum horizontal apparent wavelength to avoid aliasing effect limits its applicability to the majority of current seismic array deployments.

Parallel to the development of array analysis for subsurface imaging, seismic tomography is also transiting from ray-based inversions using 1-D reference models towards adjoint tomography based on 3-D reference models (e.g. Tromp *et al.* 2005; Chen *et al.* 2007; Tromp *et al.* 2008; Fichtner *et al.* 2009; Tape *et al.* 2009), see Liu & Gu (2012) for a review. The latter approach takes advantages of full 3-D numerical simulations in forward modelling and sensitivity kernel calculation, and iteratively refines velocity models through non-linear optimization techniques. Adjoint tomography naturally and automatically includes as many fitted segments of seismograms as 3-D reference models allow over iterations, thus making identification of primary phases unnecessary (e.g. Maggi *et al.* 2009; Tape *et al.* 2010; Lee & Chen 2013). It significantly improves the seismic imaging of 3-D heterogeneous structures with strong velocity contrasts and has become the state-of-the-art technique for high-resolution imaging of various tectonically complex regions in the past few years (e.g. Chen *et al.* 2007; Fichtner *et al.* 2009; Tape *et al.* 2009; Zhu *et al.* 2012). Another notable feature of adjoint tomography is that the topography of interfaces can be determined in junction with volumetric material properties (Liu & Tromp 2008), prompting its use in mapping crustal and mantle discontinuities.

However, for regions with limited seismicity, seismic imaging relies mainly on teleseismic records. As numerical simulations of seismic waves below the period of 8 s at the scale of the globe are still computationally prohibitive, given standard cluster access (e.g. Komatitsch *et al.* 2005; Tromp *et al.* 2008), application of adjoint tomography to teleseismic waves at the frequencies relevant to regional high-resolution imaging (e.g. 1–2 s for *P* waves and 3–6 s for *S* waves as in RF and scattering imaging studies) is far from practical (Liu & Gu 2012; Monteiller *et al.* 2013). However, an adapted version of adjoint tomography, where the forward calculation is given by the interfacing of localized 2-/3-D numerical simulation with a fast computed 1-D solution on the boundary, will inherit all the advantages of adjoint tomography while drastically reducing the amount of computation involved. Of course, the assumption behind this hybrid approach is that for coda waves of teleseismic phases, only 2-/3-D effects inside the detailed computational domain matter, while 2-/3-D effects outside the domain are negligible, and only 1-D incident wavefield is computed. Indeed, interfacing of solutions to either 1-D layered model (Roecker *et al.* 2010; Liu & Chen 2011; Pageot *et al.* 2013) or 1-D spherical earth model (Chevrot *et al.* 2011; Monteiller *et al.* 2013) with spectral-element (SEM) solvers has been proposed to speed up forward simulations. Similar hybrid approaches have been used in previous studies of topography and basin effects for ground motion predictions (Bielak & Christiano 1984; Moczo *et al.* 1997; Bielak *et al.* 2003; Krishnan *et al.* 2006; Semblat *et al.* 2008). For example, two-step domain-reduction methods, which first calculate the ground motion for a background structure with localized geological features removed, and then simulate responses of a reduced region with the localized features, have been used to model earthquake ground motion in heterogeneous regions with large localized contrasts in wave speeds (Bielak & Christiano 1984; Bielak *et al.* 2003). To efficiently calculate seismic wavefields for localized 2-D near-surface structures embedded in a 1-D horizontally layered background medium, Zahradnik & Moczo (1996) combined the discrete-wavenumber (DW) solution of 1-D background model with the finite-difference (FD) simulations of local 2-D structures. Wen & Helmberger (1998) developed a *P*–*SV* hybrid method, which interfaces the generalized ray theory (GRT) solutions with FD calculations to model seismic wave propagation in the ultra-low velocity zone near the core–mantle boundary. Chen *et al.* (2005) used a similar hybrid method to study the lithospheric and upper-mantle structures. Zhao *et al.* (2008) further developed the GRT-FD hybrid method to calculate synthetic seismograms for 2-D localized heterogeneous and anisotropic media. To implement full waveform teleseismic tomography, Roecker *et al.* (2010) interfaced a 2-D spectral domain FD method with plane-wave propagation through a 2.5-D elastic medium. Although computationally efficient, the use of FD scheme defined on a regular grid still presents limitations on handling complex 2-D geological models with large free-surface or interface topography, as often encountered in practice (Zhou *et al.* 2012; Zheng *et al.* 2013). To overcome this limitation, a more geometrically flexible method, such as the finite-element method (FEM) or SEM may be used to conduct forward modelling. Following the work of Chevrot *et al.* (2004), Godinho *et al.* (2009) and Chevrot *et al.* (2011), Monteiller *et al.*

(2013) proposed a hybrid method that interfaces a direct solution method (DSM) for 1-D spherical earth with a SEM to compute short-period synthetics for teleseismic body waves in 3-D regional models.

As the wave fronts of teleseismic waves entering the upper mantle (above 410 km) beneath most seismic arrays can be safely assumed to be planar (Rondenay 2009), analytical methods such as frequency–wavenumber (FK) method can be used to precisely and efficiently compute plane-wave responses of 1-D medium (Zhu & Rivera 2002). A full numerical simulation can be then initiated to calculate the synthetic response of local 2-D/3-D media to plane-wave incidence, while at the boundaries, field values are exchanged with 1-D FK solutions, and scattered waves are absorbed by absorbing boundary conditions (Liu & Gu 2012). This hybrid method may efficiently model the high-frequency coda of main teleseismic phases, which consist largely of energy scattered off small heterogeneities in the receiver-side media. It combines the efficiency of analytical methods for 1-D background media with the accuracy of full numerical simulations for 2-D/3-D media, and accurately captures the interactions between teleseismic waves and heterogeneous structures beneath seismic arrays.

In this paper, we thus develop an SEM-FK hybrid method to calculate the synthetic response of local media to teleseismic plane-wave incidence for the modelling of teleseismic coda waves. The SEM accurately handles large velocity contrasts and undulating surfaces and discontinuities through a precisely designed finite-element mesh, and achieves exponential convergence properties of spectral methods (Komatitsch & Tromp 1999; Cohen 2002; De Basabe & Sen 2007; Seriani & Oliveira 2008). It has shown to be an efficient solver for effective modelling of complex media (Komatitsch *et al.* 2004; Stupazzini 2006; Lee *et al.* 2008; Peter *et al.* 2011). As numerical calculations are limited only to localized regions of interest, the SEM-FK hybrid method that we introduce can simulate the local propagation of teleseismic waves at very high frequencies. Hereafter, we first present the theory of the SEM-FK hybrid method, and then benchmark it against FK solutions for a 1-D layered model in Section 2. In Section 3, we give a brief introduction to adjoint tomography, which utilizes the forward SEM-FK calculations. We conduct synthetic tomographic tests for various models in Section 4 to demonstrate the promising resolving capabilities of adjoint tomography with teleseismic coda waves, for both subsurface interfaces and velocity anomalies.

2 SEM-FK HYBRID METHOD

2.1 FK method

In this study, we are interested in modelling responses of local heterogeneities to teleseismic body-wave incidence. For computational efficiency, we assume the wave fronts of teleseismic body waves to be planar when they arrive in the upper mantle beneath a seismic array with a limited aperture (Rondenay 2009), which can then be propagated into a local medium. As shown in Fig. 1, the plane waves impinge from the half-space below a stack of n layers with an incidence angle θ , and soon enter a target region (grey shaded box) which assumed to contain all local heterogeneities (Roecker *et al.* 2010). The interactions of plane waves with these heterogeneities will generate scattered waves, which can be recorded as coda (sometimes precursor) waves by seismic stations at the free surface.

We adapt the propagator matrix method (e.g. Haskell 1953; Takeuchi & Saito 1972) to compute the displacement response of a stack of layers to incident P and S plane waves from the half-space below. It uses the Thompson–Haskell propagator matrix technique to obtain the displacement field at any point within the n -layer over a half-space (Thomson 1950; Haskell 1962; Zhu *et al.* 2012). When free surface boundary condition and unit amplitude of incident waves are applied, the wavefields can be uniquely determined. We formulate the FK algorithm for plane-wave incidences and present its tenets in the Appendix.

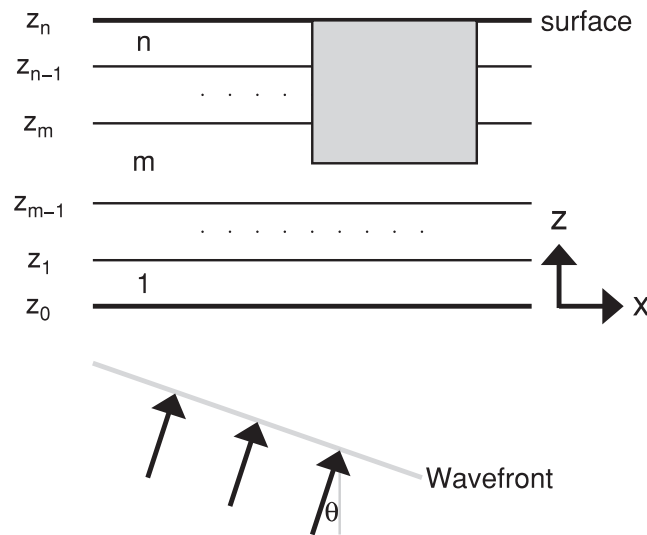


Figure 1. An n -layer over half-space model. The grey shaded region is the computational domain in which the spectral element method is applied. Teleseismic plane waves with an incidence angle θ enter from the bottom of the stacked layers. For simplicity, the interface between the bottom layer and the half-space is assumed to be located at $z_0 = 0$.

2.2 SEM

For 2- and 3-D heterogeneous media, the analytical techniques (such as the FK method discussed in the Appendix) can no longer adequately capture the complex wave propagation phenomena related to interactions with local heterogeneities, and numerical solvers have to be used. The SEM is a class of numerical methods that is currently widely used to simulate seismic wave propagation in heterogeneous media at both regional (e.g. Komatitsch *et al.* 2004; Peter *et al.* 2011) and global (e.g. Komatitsch & Tromp 2002a,b; Tromp *et al.* 2008) scales [see Chaljub *et al.* (2007) for a review]. It combines the flexibility of the finite-element method with the accuracy of spectral techniques and has good accuracy and convergence properties (Peter *et al.* 2011). With its flexible mesh design, the SEM readily incorporates surface topography and internal material interfaces, and with its high numerical accuracy it can tackle complex wave propagation in full detail.

To conduct feasibility studies of an SEM-FK hybrid method and its application to seismic imaging, we use the open-source software package SPECSEM2D, downloaded from Computational Infrastructure for Geodynamics (CIG) website (<http://www.geodynamics.org>). It is designed to simulate both forward and adjoint seismic wave propagation in 2-D acoustic, (an)elastic, poroelastic or coupled-(an)elastic-poroelastic media, although in this paper we focus mainly on the elastic properties of the media (i.e. density and P - and S -wave velocities). This package has the flexibility of handling both structured or unstructured meshes. For subhorizontally layered earth models, an internal mesher in the package can be used to generate quadrilateral mesh for later spectral-element simulations. For more complex 2-D models, external mesh generation tools need to be employed to honour all discontinuities and interfaces present in the models (e.g. Blacker *et al.* 1994; Casarotti *et al.* 2008; Peter *et al.* 2011), such as the CUBIT software package (<http://cubit.sandia.gov/>). The SPECSEM2D package also includes the capability of computing finite-frequency kernels for seismic inversion purposes (see Section 3.1). In this paper, we use CUBIT mesher for forward wave simulations in Section 2.3.2 and synthetic data calculations in Section 4.3. And for all other models as well as simulations in adjoint inversions, we apply the internal mesher.

2.3 SEM-FK hybrid method

When plane waves travelling upwards from the half-space through a stack of layers encounter local heterogeneities, which are encompassed by a domain drawn as a grey box in Fig. 1, reflected or scattered waves are generated within the domain through interactions between the incoming seismic waves and heterogeneous local structures. With displacement and traction of the 1-D background medium supplied by the FK method for the boundary, we can apply the SEM in the grey box to simulate the continued plane-wave propagation into the heterogeneous domain.

Some of the scattered waves generated by local heterogeneities may propagate outwards through the boundaries of the computational domain. In this case, the total wavefield $\mathbf{u}_{\text{total}}$ at the boundaries is the sum of the incident wavefield \mathbf{u}_{FK} calculated by the FK method and the scattered wavefield \mathbf{u}_{sc} generated by scatterers within the domain. At the boundaries, the outward scattered energy (and only that) needs to be absorbed. Following the approach of Komatitsch & Tromp (1999), inspired by Bielak & Christiano (1984) and also more recently used by Chevrot *et al.* (2004) and Godinho *et al.* (2009), we subtract the analytical expression of the incident wavefield from the total wavefield on the outer edges of the computational domain, and apply the absorbing boundary condition only to the remaining scattered wavefield. And in the case of the simple and approximate absorbing condition of Clayton & Engquist (1977), this leads to

$$(\mathbf{T}_{\text{total}} - \mathbf{T}_{\text{FK}}) \cdot \hat{\mathbf{n}} = \rho\alpha[\hat{\mathbf{n}} \cdot \partial_t(\mathbf{u}_{\text{total}} - \mathbf{u}_{\text{FK}})]\hat{\mathbf{n}} + \rho\beta[\hat{\mathbf{t}} \cdot \partial_t(\mathbf{u}_{\text{total}} - \mathbf{u}_{\text{FK}})]\hat{\mathbf{t}}, \quad (1)$$

on the boundaries of the SEM solver, where \mathbf{T} denotes the stress tensor, $\hat{\mathbf{n}}$ is the unit outward normal of the boundary and $\hat{\mathbf{t}}$ is the unit vector tangential to the boundary. This boundary treatment naturally combines SEM wavefield with FK solutions and is the key part of the SEM-FK hybrid implementation. Note that eq. (1) is only applied to the scattered waves (instead of the total wavefield), which is usually much weaker than the incoming teleseismic waves. Therefore the simple and approximate absorbing condition of Clayton & Engquist (1977) is sufficiently effective for our applications with plane-wave incidence despite its zeroth-order nature.

2.3.1 Benchmark for 1-D models

Because of the flexibility of the SEM in dealing with subsurface discontinuities and heterogeneities, the SEM-FK hybrid method can be used to investigate the effect of local heterogeneous media on the coda waves of teleseismic main phases. Before proceeding further, let us first demonstrate the accuracy of SEM-FK by benchmarking it against FK solutions for purely 1-D media, that is, having no local lateral anomaly.

We consider the case of P plane waves with an incidence angle $\theta = 15^\circ$ travelling up in a two-layer crust-over-mantle model (Fig. 2, and the material properties of the two layers, typical of the crust and the mantle, are given in Table 1). Since recorded seismograms are affected by both source and structural variations, an accurate estimation of the source wavelet of the incident P or S plane waves is a key first step in successfully matching synthetics to observed scattered waves. In practice, the source-time function is estimated through a deconvolution procedure (e.g. Rondenay 2009; Kind *et al.* 2012). For the synthetic studies of this paper, we assume the incident P wave to follow a simple Gaussian source-time function

$$f(t) = \frac{f_0}{\sqrt{\pi}} \exp[-(f_0 t)^2], \quad (2)$$

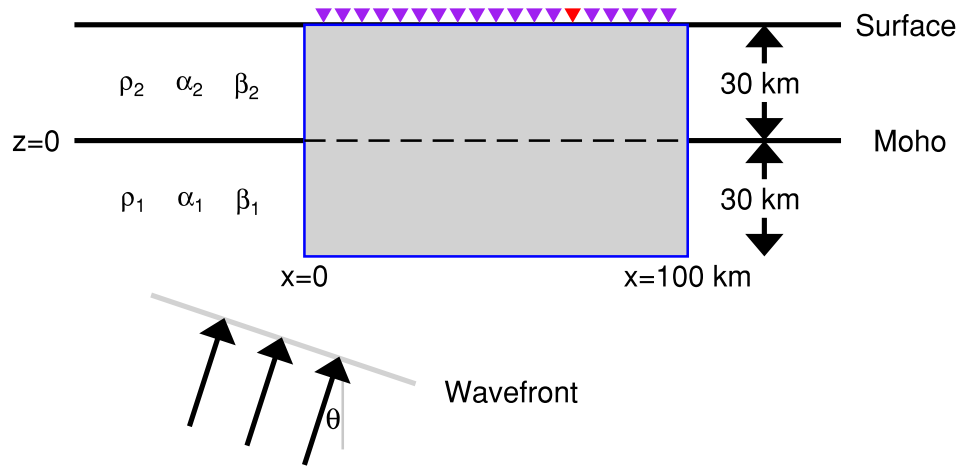


Figure 2. One layer over half-space model for SEM-FK benchmarking. The top and bottom layers have the typical velocities of the crust and mantle given in Table 1. The thickness of the crust is 30 km. The grey shaded region with a size of 100×60 km is the computational domain for the SEM, which may include undulated Moho (black dashed line) and heterogeneous density, V_p and V_s structures. Dense seismic arrays (inverse triangles) are placed along the free surface of the domain. The left boundary of this domain and the Moho discontinuity together define the local Cartesian coordinate system. This model also serves as the background model for various synthetic tests in Section 4.

Table 1. Material properties for the one-layer-over-half-space (e.g. crust-over-mantle) model in Fig. 2.

	Density ρ (kg m^{-3})	V_p (m s^{-1})	V_s (m s^{-1})
Crust	2600	5800	3198
Mantle	3380	8080	4485

where $f_0 = 2$ Hz is the cut-off frequency, and a maximum amplitude of 1 mm. We compute the displacement field for the two-layer model by both the analytical FK method and our SEM-FK hybrid method. A $100 \text{ km} \times 60 \text{ km}$ domain is used for the SEM calculations in the latter case (Fig. 2), and its mesh is built with the internal mesher in the SPECFEM2D package. The P plane waves start outside the simulation domain at initial time $t = 0$ (its wave front intersects the Moho at $x = -450$ km). At time $t = 10.8$ s, the planar wave front arrives at the bottom left corner of the SEM simulation domain. Fig. 3(a) shows the SEM-FK snapshots of the horizontal-component displacement field between 10 and 45 s. The P -wave incident from below is first partitioned into upgoing and downgoing P and S waves at the Moho interface. The upgoing waves are then reflected and converted at the surface. These reflected and converted waves may again transmit through the Moho to the mantle, or be reflected back to the crust. Because of the surface and of the Moho discontinuity, multiple secondary phases can be observed as reflected/transmitted/converted waves after the main P arrival (i.e. P coda waves) at seismic arrays deployed on the surface. For comparison, we also use the FK method itself to compute synthetic seismograms at surface stations. Figs 3(b) and (c) show the horizontal- and vertical-component velocity seismograms recorded by a station located at $x = 70$ km on the surface (red triangle in Fig. 2), including both the main P arrival and other secondary coda phases (identified in Figs 3b and c with their corresponding ray paths labelled in Fig. 3e). The results of the hybrid method (red traces) match very well with those of the FK method (blue traces) in Figs 3(b) and (c), their difference being less than 1 per cent of the amplitude of either the horizontal or vertical seismogram (Fig. 3d). This very minor difference may be partly due to the movement of interfaces on the boundaries that is unaccounted for by the hybrid method (Roecker *et al.* 2010). On the other hand, both FK and SEM are performed numerically, which is subjected to small numerical errors, therefore they are accurate but not exact.

2.3.2 Plane-wave incidence to a 2-D medium

To demonstrate the capability of the SEM-FK hybrid method in dealing with local heterogeneities, we show an example of the interaction of incoming P plane waves with a subducted slab. The computational domain, of size $200 \text{ km} \times 200 \text{ km}$, consists of the crust and uppermost mantle, with material properties given in Table 1. A subducted slab with a +4.0 per cent density and velocity perturbation is present in the mantle (Fig. 4j). As the FK method is only valid for layered models, we taper material properties at the left and right edges of the subducted slab. A P plane wave with a 15° incidence angle impinges on the bottom left corner of the computational domain at time $t = 5$ s. A cut-off frequency of $f_0 = 1$ Hz and maximum amplitude of 1 mm is used for this example. The mesh for this example is built with the CUBIT software package (Section 2.2) to accurately account for the geometry of the subducted slab. Figs 4(a)–(i) show snapshots of the horizontal displacement field computed by the SEM-FK hybrid method. The incident waves are reflected by or transmitted through the lower and upper slab boundaries (Figs 4c and d) as well as the Moho (Fig. 4e) and then reflected by the free surface (Fig. 4f). To further demonstrate the effect of the slab on seismic wave propagation, horizontal-component displacement recorded by 20 receivers at the surface is shown in Fig. 5. The arrivals of the main phase and coda phases labelled in Fig. 3(e) can be clearly identified in Fig. 5 (dashed lines). These coda waves will be critical in the detailed mapping of slab structures in the synthetic tests discussed in later sections.

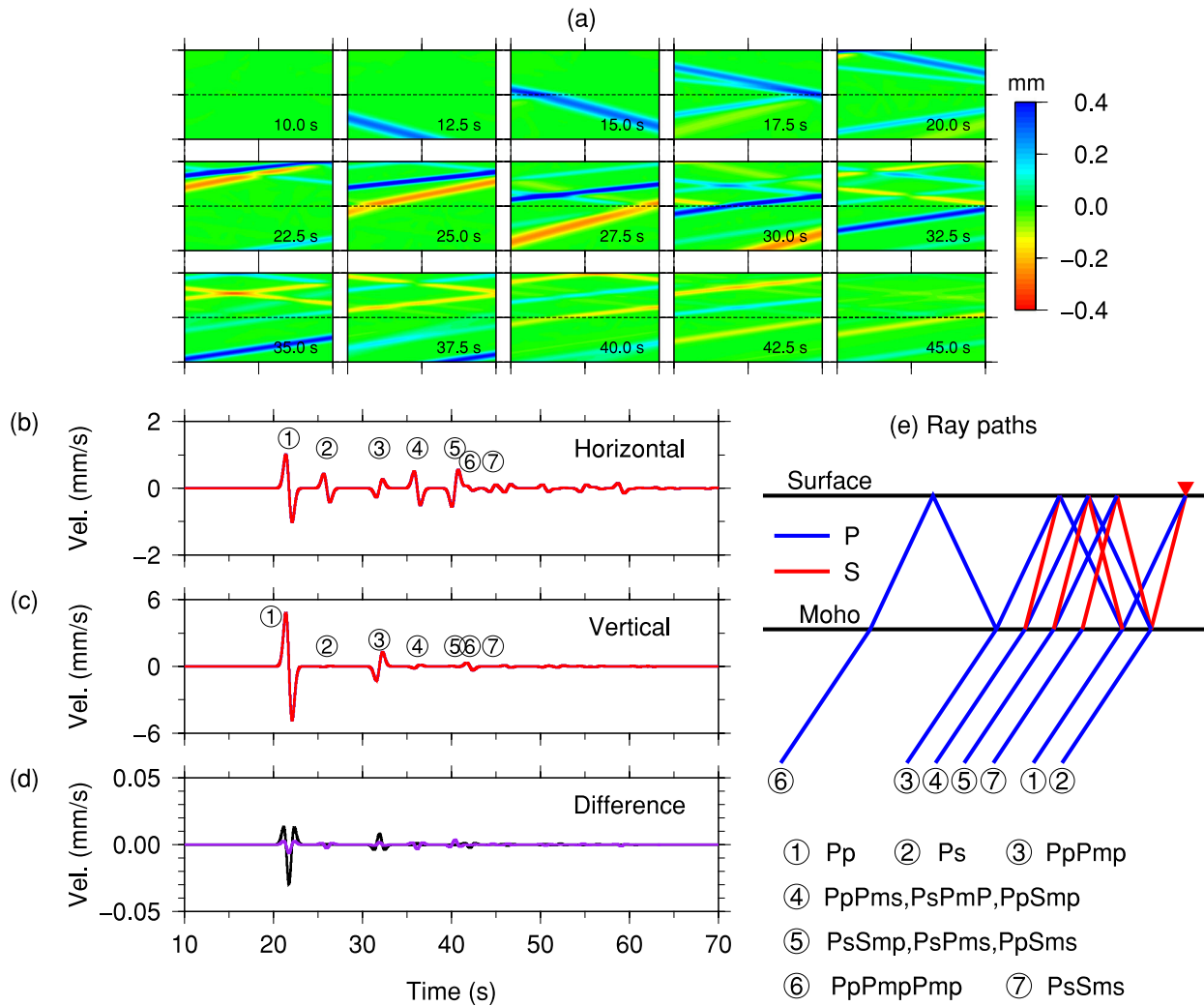


Figure 3. (a) Snapshots of horizontal-component displacement field computed by the SEM-FK hybrid method in a one-layer-over-half-space model (Fig. 2 and Table 1) due to a P plane wave with 15° incidence angle. (b) Horizontal- and (c) vertical-component velocity seismograms generated by the FK (blue) and SEM-FK hybrid (red) methods for a receiver located at $x = 70$ km on the surface (red triangle in Fig. 2) with various coda phases identified. (d) Differences between the FK and SEM-FK seismograms for the horizontal (purple) and vertical (black) components. Note that different vertical scales are used from (b) and (c). In (e) we have drawn the ray paths of various coda phases, with blue and red line segments showing the P and S legs, respectively.

3 BRIEF SUMMARY OF THE THEORY OF ADJOINT TOMOGRAPHY

Over the past three decades, discontinuity characterization using converted seismic waves and the determination of volumetric material properties by seismic tomography have formed the set of core methodologies used in seismic imaging of the solid Earth at regional and global scales. These two families of methods provide complementary means of mapping the internal structures of the Earth (Rondenay 2009). Actually, observed seismograms depend on variations in both material discontinuities and volumetric material properties. In the framework of adjoint tomography, we can determine simultaneously the internal discontinuity undulations and the variations of volumetric material properties (e.g. density, wave speeds). Here we only summarize the theoretical aspects of adjoint tomography in 2-D for later sections. More details can be found for instance in Tarantola (1984), Tromp *et al.* (2005), Liu & Tromp (2006), Tape *et al.* (2007), Liu & Tromp (2008), Tromp *et al.* (2008), Fichtner *et al.* (2009), Tape *et al.* (2009) and Liu & Gu (2012).

3.1 Fréchet kernels

In adjoint tomography, we seek to minimize a misfit function $\phi(\mathbf{m})$ between observed seismograms \mathbf{d} and the corresponding synthetics \mathbf{u} for a given earth model, for example, an isotropic 2-D model $\mathbf{m}(x, z) = \{\rho, \kappa, \mu, d\}$, where ρ , κ and μ denote the density, bulk and shear moduli, respectively, and d is the depth distribution of a set of discontinuities Γ . The misfit can be characterized in numerous forms, for example, cross-correlation traveltime and amplitude anomalies (Dahlen *et al.* 2000), multitaper phase and amplitude measurements (Zhou *et al.* 2004; Maggi *et al.* 2009) or straight waveform differences (Tarantola 1984; Tromp *et al.* 2008). As the choices of measurements only lead to

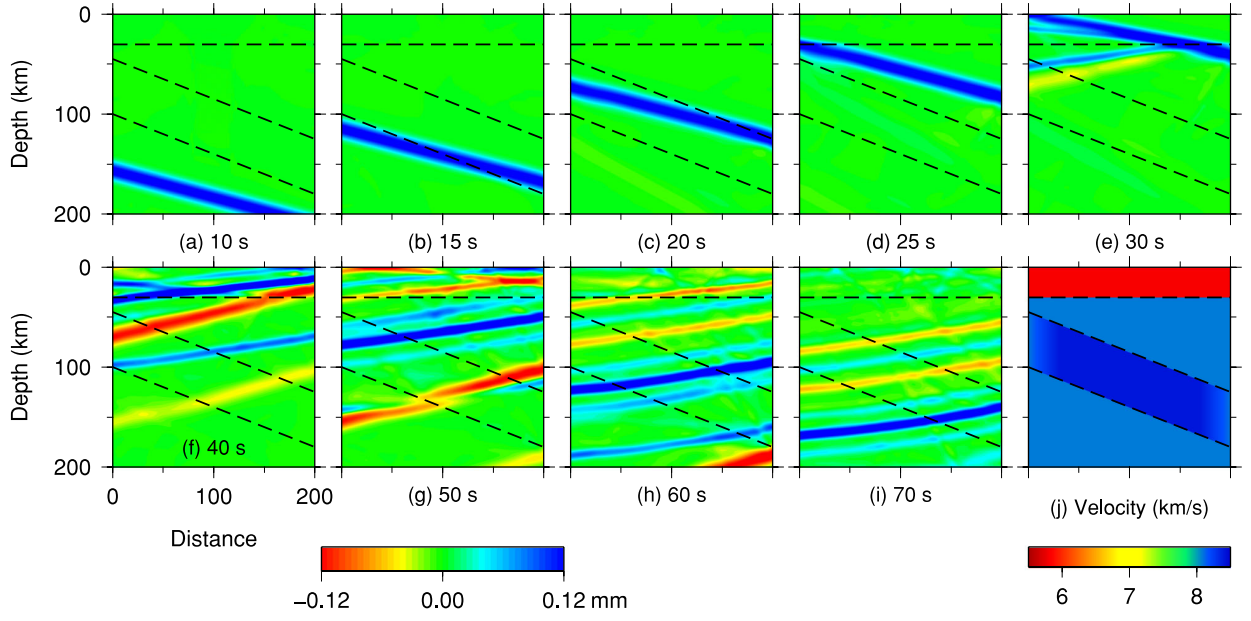


Figure 4. Snapshots of horizontal displacement (a–i) computed by our SEM-FK hybrid method in a heterogeneous model with a subducted slab (j). The computational domain, of size 200×200 km, consists of the crust and the uppermost mantle (Table 1), as well as a subducted slab with a +4.0 per cent density and velocity perturbation in the mantle.

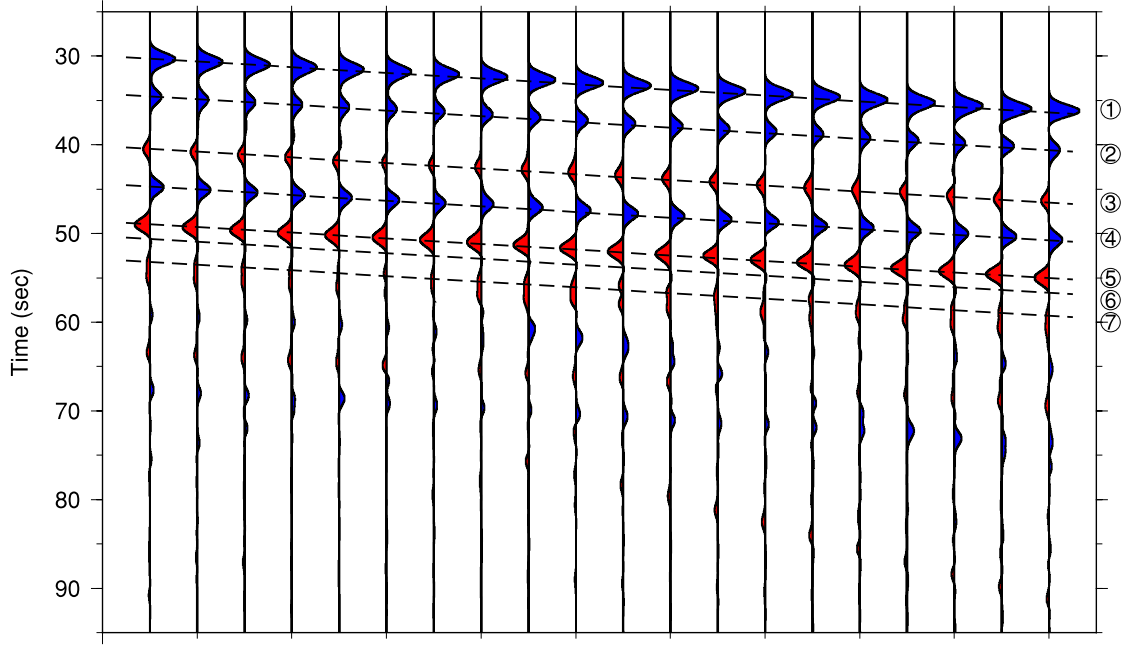


Figure 5. Horizontal-component displacement seismograms generated by the SEM-FK hybrid method for a subducted slab model (Fig. 4j). They are recorded by 20 evenly spaced receivers at the top surface of the computational domain. The seven dashed lines denote the predicted arrival times of the seven Moho related phases labelled in Fig. 3(e).

variations in the adjoint source-time functions (Liu & Gu 2012), here as an example, we use the total waveform misfit function calculated for N stations

$$\phi(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^N \int \left\| \mathbf{w}_i(t) [\mathbf{u}(\mathbf{x}_r^i, t; \mathbf{m}) - \mathbf{d}(\mathbf{x}_r^i, t)] \right\|^2 dt, \quad (3)$$

where $\mathbf{w}_i(t)$ is the windowing and weighting function assigned to both synthetic \mathbf{u} and data \mathbf{d} at receiver \mathbf{x}_r^i . Then, variations of ϕ may be formally expressed as

$$\delta\phi = \int_S (K_\rho \delta \ln \rho + K_\kappa \delta \ln \kappa + K_\mu \delta \ln \mu) d^2 \mathbf{x} + \int_\Gamma K_d \delta \ln d d\mathbf{x}, \quad (4)$$

where $\delta \ln \rho$, $\delta \ln \kappa$ and $\delta \ln \mu$ denote relative volumetric model perturbations, $\delta \ln d$ represents relative topographic variations of discontinuity Γ , and K_ρ , K_κ , K_μ and K_d are the corresponding Fréchet kernels, that is, formally

$$\mathbf{g}(\mathbf{m}) = \frac{\partial \phi}{\partial \ln \mathbf{m}}|_{\mathbf{m}} = (K_\rho, K_\kappa, K_\mu, K_d)|_{\mathbf{m}}. \quad (5)$$

In general, this non-linear optimization problem can be solved by iterative techniques, such as the conjugate-gradient (CG) or the Newton's methods. The gradient of the misfit function $\mathbf{g}(\mathbf{m})$ needs to be recomputed at each iteration for the newly updated model. In the case of Newton's methods, the second-order derivative of misfit with respect to model parameters $\frac{\partial^2 \phi}{\partial (\ln \mathbf{m})^2}$, that is, the so-called Hessian, also needs to be recomputed at each iteration. Convergence is assumed to be achieved when the misfit function $\phi(\mathbf{m})$ and/or model update fall below a preset threshold (Tape *et al.* 2009).

However, when both the forward and adjoint wavefields are calculated by numerical techniques, it is in general computationally prohibitive to assemble the Hessian matrix, and thus Newton's methods are unusable for most practical problems. Methods that only rely on local gradients (e.g. non-linear CG methods) are thus favoured in the iterative procedures. The generic expression of the gradient of the misfit function $\mathbf{g}(\mathbf{m})$ can be derived from the Born approximation (Nolet 2008), which involves the interaction between the regular forward wavefield \mathbf{u} and an adjoint wavefield \mathbf{u}^\dagger produced by placing time-reversed adjoint source-time functions at receivers as simultaneous, virtual sources (e.g. Tromp *et al.* 2005, 2008; Tape *et al.* 2007). Liu & Tromp (2008) derived a complete set of adjoint equations and Fréchet kernels for global seismic wave propagation based upon a Lagrange multiplier method. For an isotropic earth model, various Fréchet kernels in eq. (5) are given by Tromp *et al.* (2005):

$$K_\rho(\mathbf{x}) = -\rho(\mathbf{x}) \int_0^T \mathbf{u}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{u}(\mathbf{x}, t) dt, \quad (6)$$

$$K_\kappa(\mathbf{x}) = -\kappa(\mathbf{x}) \int_0^T [\nabla \cdot \mathbf{u}^\dagger(\mathbf{x}, T-t)][\nabla \cdot \mathbf{u}(\mathbf{x}, t)] dt, \quad (7)$$

$$K_\mu(\mathbf{x}) = -2\mu(\mathbf{x}) \int_0^T \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) dt, \quad (8)$$

$$\begin{aligned} K_d(\mathbf{x}) = d(\mathbf{x}) \int_0^T \{ & \rho(\mathbf{x}) \mathbf{u}^\dagger(\mathbf{x}, T-t) \cdot \partial_t^2 \mathbf{u}(\mathbf{x}, t) \\ & + \kappa(\mathbf{x}) \nabla \cdot \mathbf{u}^\dagger(\mathbf{x}, T-t) \nabla \cdot \mathbf{u}(\mathbf{x}, t) + 2\mu(\mathbf{x}) \mathbf{D}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) \\ & - \kappa(\mathbf{x}) \hat{\mathbf{n}}(\mathbf{x}) \cdot [\partial_n \mathbf{u}^\dagger(\mathbf{x}, T-t) \nabla \cdot \mathbf{u}(\mathbf{x}, t) + \partial_n \mathbf{u}(\mathbf{x}, t) \nabla \cdot \mathbf{u}^\dagger(\mathbf{x}, T-t)] \\ & - 2\mu(\mathbf{x}) [\hat{\mathbf{n}}(\mathbf{x}) \partial_n \mathbf{u}^\dagger(\mathbf{x}, T-t) : \mathbf{D}(\mathbf{x}, t) + \hat{\mathbf{n}}(\mathbf{x}) \partial_n \mathbf{u}(\mathbf{x}, t) : \mathbf{D}^\dagger(\mathbf{x}, T-t)] \}^+ dt, \end{aligned} \quad (9)$$

where the adjoint wavefield \mathbf{u}^\dagger satisfies the adjoint wave equation

$$\rho \frac{\partial^2 \mathbf{u}^\dagger}{\partial t^2} = \nabla \cdot [\lambda(\nabla \cdot \mathbf{u}^\dagger) + \mu(\nabla \mathbf{u}^\dagger + (\nabla \mathbf{u}^\dagger)^T)] + \sum_{i=1}^N \mathbf{w}_i(t) [\mathbf{u}(\mathbf{x}_r^i, t; \mathbf{m}) - \mathbf{d}(\mathbf{x}_r^i, t)] \delta(\mathbf{x} - \mathbf{x}_r^i), \quad (10)$$

and \mathbf{D} and \mathbf{D}^\dagger are the traceless deviatoric strain tensors for the forward and adjoint fields. The notation $\{\cdot\}^+$ in eq. (9) denotes the jump in the enclosed quantity when going from the inward (−) side to the outward (+) side (i.e. in the unit normal $\hat{\mathbf{n}}$ direction) of discontinuity Γ .

For seismic inversions, we express the variation in the misfit function $\phi(\mathbf{m})$ as a function of relative perturbations in density $\delta \ln \rho$, P -wave velocity $\delta \ln \alpha$, S -wave velocity $\delta \ln \beta$ and discontinuity topography $\delta \ln d$

$$\delta \phi = \int_S (K_{\rho'} \delta \ln \rho + K_\alpha \delta \ln \alpha + K_\beta \delta \ln \beta) d^2 \mathbf{x} + \int_\Gamma K_d \delta \ln d d\mathbf{x}, \quad (11)$$

where the corresponding Fréchet kernels $K_{\rho'}$, K_α and K_β are given in terms of K_ρ , K_κ and K_μ

$$K_{\rho'} = K_\rho + K_\kappa + K_\mu, \quad K_\alpha = 2 \left(1 + \frac{4}{3} \frac{\mu}{\kappa} \right) K_\kappa, \quad K_\beta = 2 \left(K_\mu - \frac{4}{3} \frac{\mu}{\kappa} K_\kappa \right). \quad (12)$$

For seismic array imaging based on teleseismic coda waves in this paper, the SEM-FK hybrid method is first used to simulate the forward wavefield \mathbf{u} due to teleseismic plane wave incidence. Afterwards, the SEM is applied again to compute the adjoint wavefield \mathbf{u}^\dagger with properly calculated adjoint sources. The gradient $\mathbf{g}(\mathbf{m})$ of the misfit function $\phi(\mathbf{m})$ (i.e. Fréchet kernels) can then be obtained through the interaction of these two wavefields according to eqs (6)–(9) and (12).

Owing to the point-source representations in SEM packages, in practice, we apply a spatial smoothing to these Fréchet kernels in order to reduce the spurious amplitudes in the vicinity of receivers (Tape *et al.* 2007). The kernel value at a given point is obtained by averaging its neighbouring points through convolution with a smoothing function. We choose a 2-D Gaussian function,

$$G(x, z) = \frac{4}{\pi \sigma^2} e^{-4(x^2+z^2)/\sigma^2}, \quad (13)$$

where σ is the smoothing radius, and the smoothed kernel $\tilde{\mathbf{g}}(\mathbf{x}, z)$ is computed by

$$\tilde{\mathbf{g}}(\mathbf{x}, z) = \iint_S \mathbf{g}(\mathbf{x} - \mathbf{x}', z - z') G(\mathbf{x}', z') d\mathbf{x}' dz'. \quad (14)$$

Following Tape *et al.* (2007), we choose a smoothing radius that is approximately equal to the cut-off wavelength of the seismic waves.

3.2 Non-linear CG method

With the gradient of misfit \mathbf{g} readily calculated, velocity and discontinuity models can be updated by local gradient optimization techniques. For the sake of conciseness, in this section we assume that \mathbf{m} refers specifically to the model set $\{\delta \ln \rho, \delta \ln \alpha, \delta \ln \beta, \delta \ln d\}$. Since different types of model parameters exist in \mathbf{m} (such as topography of the discontinuities, density and velocity), it is important to non-dimensionalize or scale them by proper covariance matrices (Tape *et al.* 2007; Kim *et al.* 2011). Such a scaling approach guarantees that different model types contribute more or less equally to the magnitude of the gradient vector if no prior information is available (Kim *et al.* 2011). We assume a simple model covariance matrix of the form

$$\mathbf{C}_m = \begin{bmatrix} \sigma_\rho^2 & & & \\ & \sigma_\alpha^2 & & \\ & & \sigma_\beta^2 & \\ & & & \sigma_d^2 \end{bmatrix}, \quad (15)$$

where

$$\begin{aligned} \sigma_\rho^2 &= \left(\int_S K_{\rho'}(\mathbf{x}) K_{\rho'}(\mathbf{x}) d^2 \mathbf{x} \right)^{-1}, & \sigma_\alpha^2 &= \left(\int_S K_\alpha(\mathbf{x}) K_\alpha(\mathbf{x}) d^2 \mathbf{x} \right)^{-1}, \\ \sigma_\beta^2 &= \left(\int_S K_\beta(\mathbf{x}) K_\beta(\mathbf{x}) d^2 \mathbf{x} \right)^{-1}, & \sigma_d^2 &= \left(\int_\Gamma K_d(\mathbf{x}) K_d(\mathbf{x}) d\mathbf{x} \right)^{-1}, \end{aligned} \quad (16)$$

are the inverse of square integrations of various kernels. They are generated for the initial model \mathbf{m}^0 and fixed throughout iterations. Given the model variance matrix (15), we may non-dimensionalize the model parameters as

$$\hat{\mathbf{m}} = \left\{ \frac{\delta \ln \rho}{\sigma_\rho}, \frac{\delta \ln \alpha}{\sigma_\alpha}, \frac{\delta \ln \beta}{\sigma_\beta}, \frac{\delta \ln d}{\sigma_d} \right\}, \quad (17)$$

and the gradient $\mathbf{g}(\mathbf{m}) = \partial \phi / \partial \mathbf{m}$ is replaced by the scaled gradient

$$\hat{\mathbf{g}} = \partial \phi / \partial \hat{\mathbf{m}} = \mathbf{C}_m^{1/2} \mathbf{g}. \quad (18)$$

The model can then be updated based upon a non-linear CG algorithm, a step-by-step recipe of which can be found in Tape *et al.* (2007) and Kim *et al.* (2011).

As only the gradient $\mathbf{g}(\mathbf{m})$ is needed for model update, the non-linear CG method is characterized by low memory and computation requirements compared to Newton's methods. It has good local and global convergence properties (Hager & Zhang 2006), and provides successively improving approximations to the true solution, which may reach the tolerance threshold after relatively fewer number of iterations compared with the steepest descent method and linear CG method (Saad 2003). Nonetheless, properly designed pre-conditioners could help improve the convergence rate of the non-linear CG method (Hu *et al.* 2011).

4 SEISMIC ARRAY IMAGING

The SEM-FK hybrid method developed in Section 2 provides a tool to accurately calculate responses of local heterogeneities to incident plane waves. Based upon the Fréchet kernels computed by an adjoint method, we can exploit the sensitivities of various phases to both subsurface discontinuities and volumetric material properties. Furthermore, we can iteratively update the model by a non-linear CG method and achieve optimized models that are consistent with seismic data. In this section, we show the results of several synthetic tests and demonstrate the feasibility of high-resolution seismic array imaging based upon teleseismic converted/coda waves.

4.1 Discontinuity mapping

In the first example, a target model is built by making the Moho of a layer-over-half-space model (given in Table 1 and Fig. 2) vary with an undulated sinusoidal shape having a maximum variation of 3 km (black curves in Fig. 6). We then compute seismic wave propagation for eight teleseismic events with P plane-wave incidence angles of 4° , 12° , 20° and 28° (from both left and right) based upon our SEM-FK hybrid method. The incidence angles are restricted to the range of $[-30^\circ, 30^\circ]$ to reflect those of most teleseismic arrivals, where negative (or positive) angles represent plane-wave incidence from the lower left (or right) side of the model. All plane waves have a Gaussian source-time

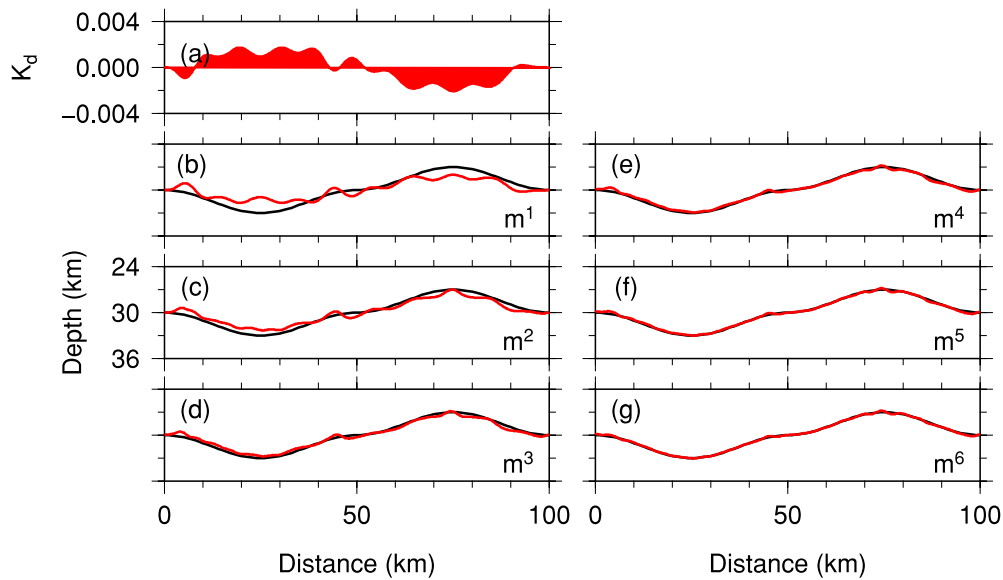


Figure 6. (a) Moho discontinuity kernel K_d (in the unit of 10^{-6} m s^{-1}) based on the starting two-layer model for the P_s converted waves with a source cut-off frequency of 2 Hz. (b)–(g) Updated models of Moho variations (red curves) over six iterations. The true Moho variation (black curves) follows a sinusoidal function with a maximum amplitude of 3.0 km. The vertical-to-horizontal exaggeration ratio is 2:1. Inversions are based on synthetic data recorded by 10 surface stations with an equal spacing of 10.0 km.

function with a cut-off frequency of 2 Hz, and 10 receivers with equal spacing of 10 km are deployed along the surface to record synthetic data.

In this case, we restrict inversions to Moho depth only and ignore other volumetric model parameters. As the arrivals of Moho-converted P_s waves are strongly affected by the Moho depth, we first invert the P_s waveforms in the synthetic data to recover Moho undulations. The starting model is chosen to be the 1-D background model with a flat Moho at 30 km depth (Fig. 2). Since converted P_s waves are most prominent on the horizontal components (Fig. 3), only horizontal components of P_s waveforms are identified and inverted.

For illustration purposes, the discontinuity kernel K_d computed based on eq. (9) for the P_s waves in the starting model is shown in Fig. 6(a). Even though the starting model has only a flat Moho, the depth kernel is informative on the necessary Moho variations to approach the ‘target’ Moho. The sign of the kernel is opposite to that of Moho topography update, while the amplitude of the kernel provides information on the magnitude of Moho variation needed in the first iteration. This and the consecutive versions of Moho kernel are smoothed using a radius of $\sigma = 1.5 \text{ km}$ in eq. (13).

Based upon the depth sensitivity kernels computed by the adjoint technique and the non-linear CG optimization method discussed in Section 3.2, we iteratively improve the Moho variations (Figs 6b–g). We remesh the model with the internal mesher at each iteration after the Moho depth has been updated. The Moho is fully recovered after six iterations, as demonstrated by the comparison between synthetic data (black traces) and synthetics (red traces) for the initial model, and models obtained at the third and sixth iterations (Figs 7a–c). Synthetics calculated for the sixth model match the synthetic data almost exactly (Fig. 7f). The misfit ϕ value is also reduced to less than 0.1 per cent of the initial misfit (Fig. 7g) after these six iterations.

This numerical example clearly demonstrates the feasibility of Moho discontinuity inversions based on converted P_s waveforms. Note that receiver spacing (10 km) is about six times the wavelength of the shear waves ($\sim 1.6 \text{ km}$). Posing the imaging problem as an iterative inverse problem avoids the formal anti-aliasing requirement (i.e. receiver spacing having to be less than half of the wavelength) in some other imaging techniques (Shang *et al.* 2012). Also note that images obtained from migrating P_s waveforms, known to be related to the inverted Moho from the first adjoint iteration (Fig. 6b) may still be quite far from the true Moho variations (Tarantola 1984). However, station spacing that is too wide may introduce artefacts in the inverted images, as shown by a similar numerical test with 22.5 km station spacing in Fig. 8. As expected, the inverted Moho has large unwanted oscillations and contains high frequency noise after six iterations. In practice, if other Moho-related phases (such as $PmPmp$ and $PmPms$) can be clearly identified (Fig. 3) in seismic recordings with high signal-to-noise ratio, P_s and these later phases can be inverted simultaneously to speed up the recovery of Moho geometry as they have complementary sensitivities to the Moho depth variations.

4.2 Volumetric structural inversion

4.2.1 P_s waveform inversion

In the second example, we first demonstrate the inversion of converted P_s waves for volumetric structures only. We design a target model that contains a 6.0 per cent slower mid-crustal V_s anomaly ($12 \text{ km} \times 8 \text{ km}$ white box in Fig. 9) in the two-layer background model (Table 1).

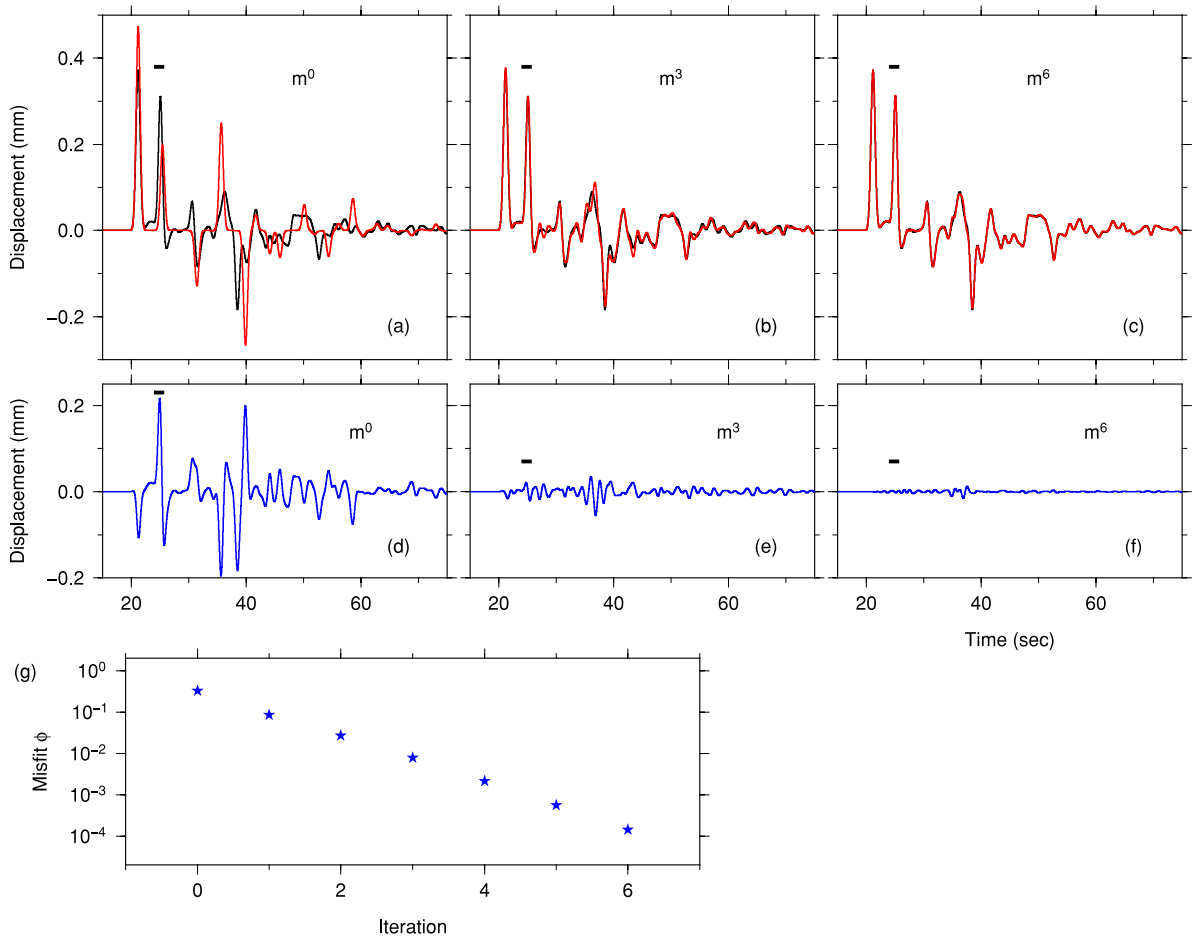


Figure 7. (a)–(c) Comparison of horizontal-component synthetic seismograms (red traces) recorded by the station located at $x = 65$ km for the initial model, and models obtained at the third and sixth non-linear CG iterations, with synthetic data (black traces). The time range of the P_s waves used in the inversion is indicated by black horizontal bars, and a Welch window function is applied over this time interval. (d)–(f) Differences between synthetic data and synthetic seismograms shown in (a)–(c). (g) Iterative reduction of the misfit function value. The unit of the misfit function ϕ is 10^{-6} s m^2 .

This slow anomaly is expected to affect the arrival time of the P_s phase and produce scattered waves at nearby stations above the anomaly. No V_p or density perturbation is introduced such that P arrival times remain the same for all stations. The same eight events are used and the receiver spacing at the surface remains 10.0 km as in Section 4.1, but the cut-off frequency of incident plane waves is reduced to 1 Hz.

Assuming no prior knowledge of the location and magnitude of structural anomalies in the model, we simultaneously update density and velocity structures. Let us first show the unsmoothed density (Fig. 9a), V_p (Fig. 9b) and V_s (Fig. 9c) kernels calculated for the P_s waveform misfit in the initial model of a particular event (with a 12° incidence angle) recorded by a station located at $x = 65.0$ km (i.e. sensitivity kernels for one source–receiver pair). These individual kernels clearly delineate various isochrons for waves scattered off possible density, V_p and V_s heterogeneities arriving in the selected P_s time window (e.g. Bostock *et al.* 2001). The density kernel (Fig. 9a) is sensitive to backscattered waves such as $PpPp$ (i.e. Moho-transmitted and surface-reflected PpP waves scattered into P waves) and $PpPs$ (PpP scattered to S) waves. The V_p kernel (Fig. 9b) is mainly sensitive to P -to- P scattered waves (Wu & Aki 1985; Rondenay *et al.* 2008), such as Ppp (Moho-transmitted Pp waves scattered to P waves) and $PpPp$ waves. As V_s anomalies result in both P -to- P and P -to- S scattering, the V_s kernel in Fig. 9(c) is sensitive to all the scattered waves, including Ppp , $PpPp$, $PpPs$ as well as Pps (Pp scattered to S). Note that the Pps isochron is tangent to the Moho discontinuity at the position where the incoming P waves are transmitted and converted to S waves.

Even though individual kernels may not resemble the embedded anomaly, the misfit kernels (i.e. Fréchet kernels of the total misfit ϕ), computed by summing up individual kernels for all events and stations, should provide clues to possible locations of velocity anomalies. The smoothed kernels of total P_s waveform misfits for the initial model are shown in Figs 9(d)–(f). Indeed, the shear velocity misfit kernel (Fig. 9f) correctly identifies the slow anomaly in the white box, and provides model update for the next iteration. The misfit kernels for P -wave velocity (Fig. 9e) and density (Fig. 9d) also suggest probable changes in V_p or density distributions, although in this case, much weaker and mostly outside the white box. This is not surprising considering that the selected windows for the transmitted P_s phase may also include other scattered arrivals as indicated by the isochrons in Figs 9(a)–(c).

Although no density or V_p anomalies exist in the target model, we still simultaneously invert for density, V_p and V_s structures based on the non-linear CG algorithm, and we examine the ability of the inversion process to attribute the misfit to the correct parameters (in this case

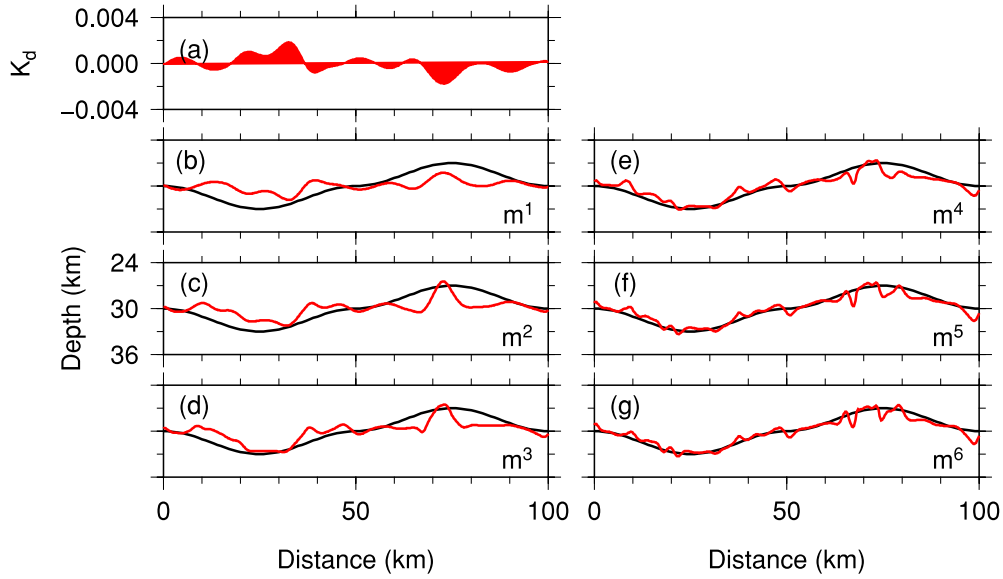


Figure 8. Similar to Fig. 6, but only horizontal P_s waveforms from five surface stations with a spacing of 22.5 km are used in the inversion.

V_s). The successively updated V_s models are shown in Figs 10(a)–(c). After nine iterations, no significant improvement in model or reduction in misfit ϕ is observed and the inversion is terminated (Fig. 10f). The final recovered slow V_s anomaly (Fig. 10c) coincides very well with the white box. On the other hand, no obvious structural anomalies can be identified in the final recovered density and P -wave models after nine iterations (Figs 10d and e), although some very weak artefacts may be visible near the surface. However, the recovered V_s anomaly has an average variation of 1.8 per cent, smaller than the presumed 6 per cent. This may be partly explained by the fact that out of the total 80 horizontal P_s waveforms used, only a few are sensitive to the small crustal anomaly and thus have limited recovery ability in the iterative process. To increase the recovered amplitude of the V_s anomaly and reduce the effect of artefacts, it may be beneficial to use more events, more stations, and include both the main P phase and/or longer coda phases (e.g. $PpPs$, $PsPs$) in the waveform inversion. Nonetheless, this numerical example verifies the capability of the SEM-FK method and adjoint tomography to map volumetric subsurface structures based on teleseismic converted/scattered waves.

4.2.2 P_s waveform inversion with pre-conditioner

To further improve the ability of seismic coda waves in imaging subsurface anomalies, we calculate scaled products of density, V_p or V_s kernels for different coda phases and use them to pre-condition the gradient vectors. For this example, misfit kernels for P_s converted waves, $PpPmp$ waves and all the remaining coda waves following $PpPmp$ (denoted by X for conciseness) are calculated and shown in Figs 11(a)–(i). Assuming density kernels for P_s , $PpPmp$ and the remaining coda waves X are $g_{\rho,1}(\mathbf{x})$, $g_{\rho,2}(\mathbf{x})$ and $g_{\rho,3}(\mathbf{x})$, we define the scaled product of these three density kernels as

$$g_{\rho}(\mathbf{x}) = \frac{1}{\eta} \prod_{i=1}^3 g_{\rho,i}(\mathbf{x}), \quad (19)$$

where η is the scaling parameter. The scaled product $g_{\alpha}(\mathbf{x})$ of V_p kernels or $g_{\beta}(\mathbf{x})$ of V_s kernels can be defined in a similar fashion. η is selected to be the maximum absolute value of all the products $g_{\rho}(\mathbf{x})$, $g_{\alpha}(\mathbf{x})$ and $g_{\beta}(\mathbf{x})$ before scaled, which in this case, is the L^{∞} norm of the product of the V_s kernels. The scaled product for V_s misfit kernels (Fig. 11l) clearly indicates a strong V_s anomaly mainly within the white box. In comparison, products of density (Fig. 11j) and V_p (Fig. 11k) kernels are very small, thus implying little model updates for density and V_p . This is consistent with the designed target model that contains neither density nor V_p anomalies, and only a strong V_s anomaly patch inside the white box. Unlike individual misfit kernels (Figs 11a–i), which do not necessarily resemble the true anomaly, the scaled products of kernels $g_{\rho}(\mathbf{x})$, $g_{\alpha}(\mathbf{x})$ and $g_{\beta}(\mathbf{x})$ are much more informative on the direction of model update. Hence, they provide a good pre-conditioner for the misfit kernels and help improve the model preferentially in regions that are simultaneously identified by multiple coda phases. This is very similar to the H - κ stacking techniques popular in RF analysis (Zhu & Kanamori 2000), in which the most probable crustal thickness and Poisson's ratio are determined by the intersections of prediction curves for different converted phases. To apply the pre-conditioner in the waveform inversion, we first define weight functions for density, V_p and V_s as

$$w_{\rho,\alpha,\beta}(\mathbf{x}) = \begin{cases} 1, & |g_{\rho,\alpha,\beta}(\mathbf{x})| \geq \xi_2; \\ (|g_{\rho,\alpha,\beta}(\mathbf{x})| - \xi_1)/(\xi_2 - \xi_1), & \xi_1 \leq |g_{\rho,\alpha,\beta}(\mathbf{x})| \leq \xi_2; \\ 0, & |g_{\rho,\alpha,\beta}(\mathbf{x})| \leq \xi_1 \end{cases} \quad (20)$$

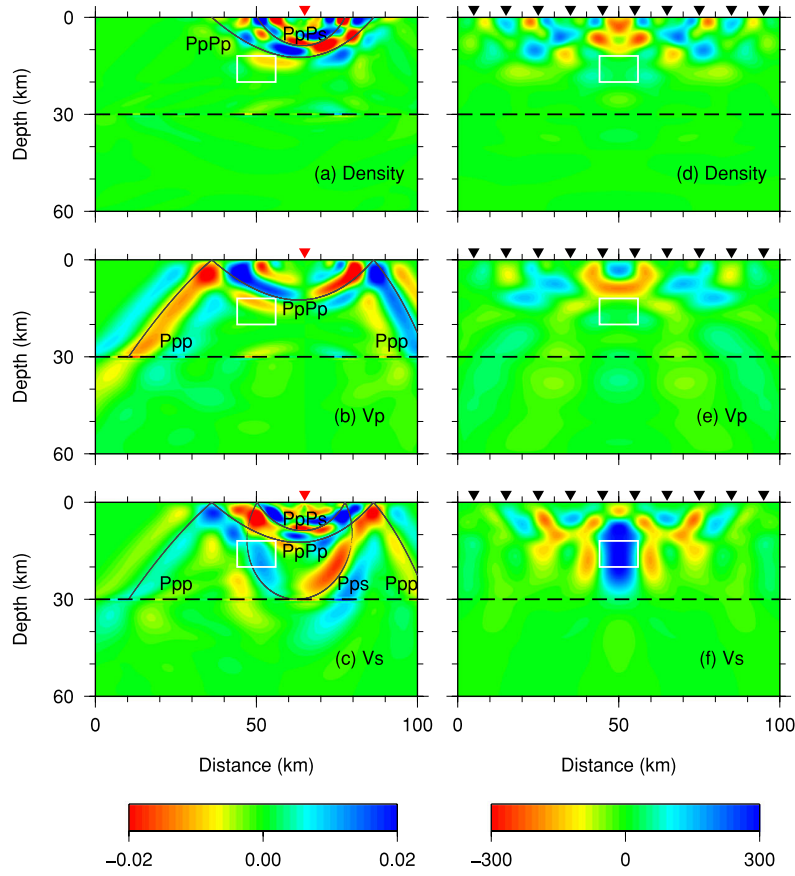


Figure 9. Density (a,d), V_p (b,e) and V_s (c,f) kernels of P_s converted waves computed for the starting model of Fig. 2. All these kernels have the unit of 10^{-14} s. (a)–(c) are the sensitivity kernels corresponding to a single event and station pair. The event has an incidence angle of 12° from the left and the station is located at $x = 65.0$ km on the surface. The grey curves denote the isochrons of forward-scattered P_{pp} , P_{ps} and backscattered P_{pPp} , P_{pPs} over the time window of the P_s arrival. (d)–(f) are the smoothed gradients of the misfit function ϕ for all events and receivers. Colour bars reflecting the amplitude of the kernels for (a)–(c) and (d)–(f) are shown at the bottom.

where ξ_1 and ξ_2 are threshold values, respectively, chosen to be 0.05 and 0.25 in this case. At each iteration, this pre-conditioner is applied to the gradient vector (i.e. misfit kernels $K_{p,\alpha,\beta}(\mathbf{x})$) before the computation of model update.

We perform a P_s waveform inversion with the above pre-conditioner and show the results of the first nine iterations in Fig. 12. The final model shows that the slow V_s variation within the white box is well recovered by inverting only the P_s waveforms, and thanks to the pre-conditioner, the amplitude of the slow V_s anomaly (-6 per cent) is much better recovered (with an average variation of 5.02 per cent) than in the example above without the pre-conditioner (Section 4.2.1 and Fig. 10). The scaled products of kernels for different seismic coda phases clearly provide very good guidance on the update of structural anomalies. In addition, the use of this pre-conditioner also significantly speeds up the convergence of the P_s waveform inversion as indicated by the faster reduction of misfit values shown in Fig. 12(f).

4.2.3 Combined traveltime and waveform inversion

Previous studies have shown that wave-equation based traveltime tomography can resolve velocity anomalies similar to the size of the first Fresnel zone, while reflected/scattered waveform inversion resolves structures of a size close to the dominant wavelength (e.g. Wu & Toksoz 1987; Virieux & Operto 2009; Liu & Gu 2012). This suggests that a combined traveltime and waveform inversion may be an efficient tool to recover subsurface structures at successively decreasing scales. As a synthetic example, we assume a target model with a 6.0 per cent slower V_p anomaly in the mid-crust, at the same location as the V_s anomaly in Sections 4.2.1 and 4.2.2. Note that V_p anomalies are more difficult to resolve from coda phases than V_s anomalies because they can only be imaged by P -to- P scatters (Wu & Aki 1985; Rondenay *et al.* 2008; Pageot *et al.* 2013). On the other hand, the arrival times of the direct P phase at some stations will be delayed by the slow V_p anomaly, thus complementing the coda waves in sensitivity. We start from a two-layer model and invert the traveltime misfits of the direct P phase to obtain a 2-D starting tomography model for later waveform inversion. The use of an improved 2-D starting model will reduce discrepancies of coda waves between data and synthetics, hence speeding up the convergence of the non-linear CG method.

The same set of event incidence angles and stations as in Section 4.2.1 is used in these inversions with the cut-off frequency of 1 Hz. Fig. 13 shows density, V_p and V_s kernels of the direct P -wave cross-correlation traveltime misfit for individual source–station pair (Figs 13a–c), as well as all sources and stations combined (Figs 13d–f). Clearly, individual density kernel (Figs 13a) is sensitive to impedance contrast

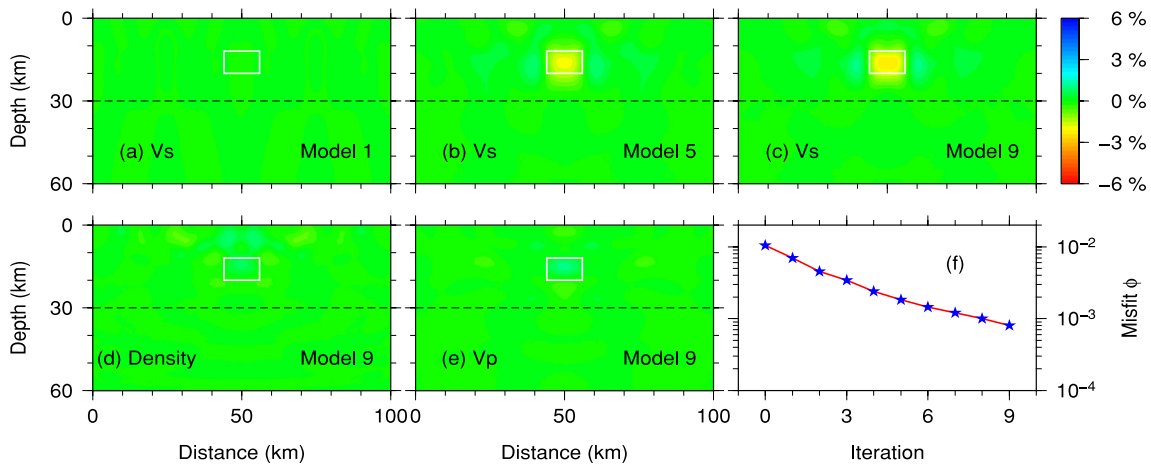


Figure 10. (a)–(c) Updated V_s structures for the first, fifth and ninth iterations of an inversion of P_s waveform only. A 6.0 per cent slower V_s anomaly is placed in the mid-crust as the target model. No variation in density or P velocity is assumed. The background is the two-layer model of Fig. 2 and Table 1, and the Moho is indicated by the dashed black line. (d) Density and (e) V_p perturbation structures after the ninth iteration. No obvious density or V_p anomaly can be identified. (f) Iterative reduction of the misfit function value. The unit of the misfit function ϕ is 10^{-6} s m^2 .

across the Moho and near the surface, while individual V_p (Fig. 13b) and V_s (Fig. 13c) kernels mainly cover the first Fresnel zones. Note that these kernels are computed similarly to those in Figs 9(a)–(c), although adjoint sources related to traveltime misfits of the direct P phase are used. The V_p kernel for traveltime misfit of all sources and receivers (Fig. 13e) exhibits positive values in the vicinity of the white box, which suggests a corresponding reduction in V_p . Figs 14(a)–(c) are the iteratively updated V_p models from inversion of only the direct P traveltime, which hints a slow V_p anomaly in the mid-crust after 16 iterations (Fig. 14c). We choose the V_p model from the 16-th iteration (Fig. 14c) of traveltime inversion as the starting model for the subsequent waveform inversion, while utilizing all coda waves following the direct P phase. After eight iterations, the 6 per cent slower V_p anomaly is almost fully covered (Fig. 14f), the average perturbation of the anomaly in the white box being -5.89 per cent.

This synthetic test nicely illustrates the power of combining traveltime inversion of teleseismic main arrivals (e.g. P phase) and waveform inversion of seismic coda waves (e.g. P_s , $PpPmp$), as the long-period structures determined by traveltime inversion reduce the non-linearity of the waveform misfit and serve as a natural initial model for higher resolution waveform inversions. This is similar to the frequency-domain full waveform inversion methods used in exploration seismology, in which transmission tomography at longer periods generates starting models for subsequent waveform inversions of reflected waves (e.g. Pratt *et al.* 1998; Pratt & Shipp 1999). For regions with seismic array deployments, 3-D traveltime tomography models are routinely inverted based on relative traveltime measurements (e.g. Vandecar & Crosson 1990; Hung *et al.* 2011). Instead of relying on simple smooth 1-D background models, coda-wave imaging may take advantage of these readily available 3-D models and speed up the convergence of waveform inversion.

4.3 Subduction zone models

Subduction zones are crucial components of plate tectonics and mantle geodynamics (e.g. Zhao 2012), and high-resolution images of subduction zones can improve our understanding of the subduction dynamics, arc magmatism and the nucleation of various types of earthquakes. Therefore, to explore the imaging capability of our SEM-FK hybrid technique for subduction zones, let us design two simplified subduction zone models and try to recover them based on teleseismic P and its coda waves.

4.3.1 Fast subducted slab model

In Section 2, we have investigated the interaction of an incident P plane wave with a simple subduction zone model (Fig. 4j) which included a 51-km-thick and 21.8° dipping subducted slab with 4.0 per cent density, fast V_p and V_s perturbations. In this section, we try to recover this subducted anomaly starting from the two-layer (crust-over-mantle) background model. 10 teleseismic events with incidence angles that evenly sample $[-27^\circ, 27^\circ]$ with an interval of 6° are used to generate synthetic data for the target model. 20 receivers are placed along the surface with an even spacing of 10 km, which is slightly larger than the wavelength of 2.5 s S waves at the surface. To constrain the slab structure, the P wave and all of its coda waves are used in the waveform inversion. In this case, only volumetric model parameters (density, V_p and V_s) are inverted and no Moho variations are included in the inversion. As waveform misfit is highly non-linear, models may be trapped in local minima and never reach the global minimum (i.e. the true model) in the iterative optimization procedure. To avoid local minima and reduce the non-linearity of waveform misfit function, we adopt a hierarchical strategy in which we successively invert waveforms from long period (10 s, $f_0 = 0.1$ Hz) to short period (2.5 s, $f_0 = 0.4$ Hz; Virieux & Operto 2009).

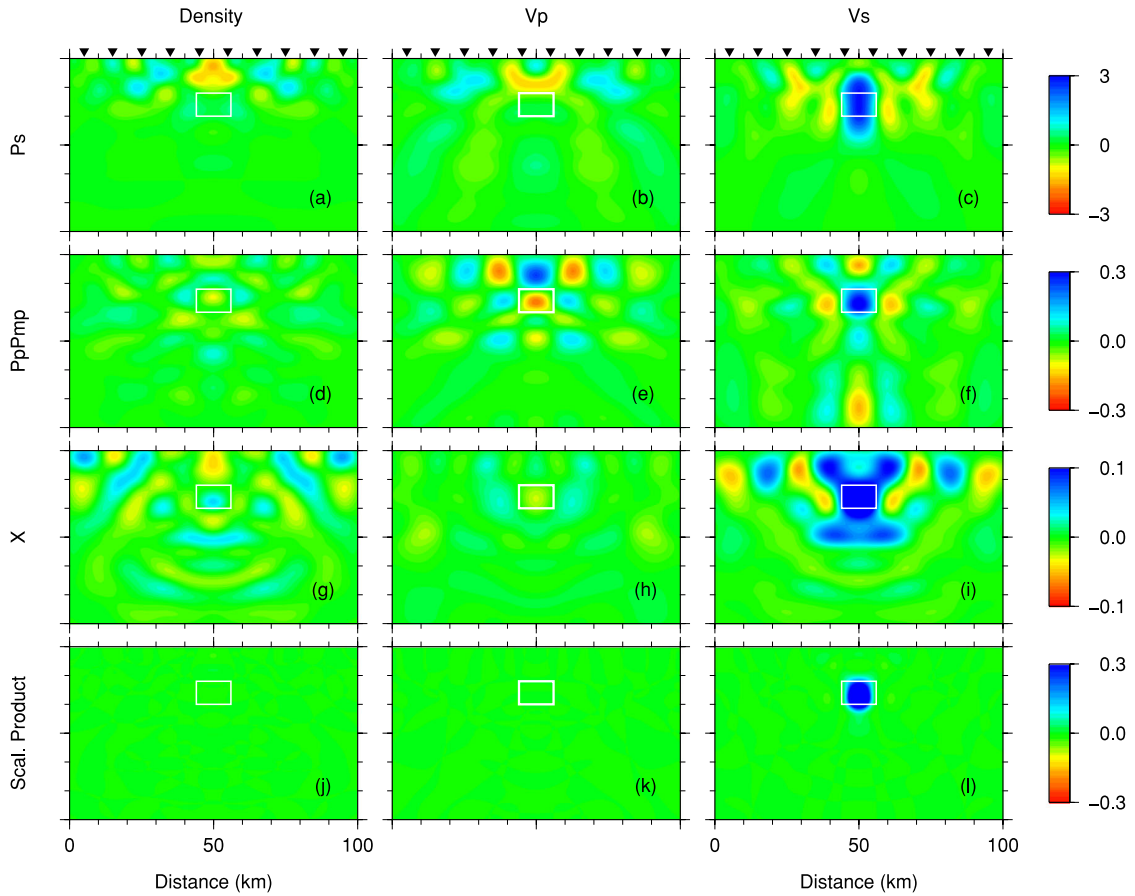


Figure 11. Smoothed kernels of the misfit function ϕ with respect to density (a,d,g), V_p (b,e,h) and V_s (c,f,i) computed for the starting model (Fig. 2) for different phases. All the kernels in (a)–(i) have the unit of 10^{-12} s. (a)–(c) Kernels of P_s converted waves, (d)–(f) kernels of $PpPmp$ waves and (g)–(i) kernels of all the remaining coda waves following the $PpPmp$ phase (shortened as X). (j) Scaled product of density kernels in (a,d,g). (k) Scaled product of V_p kernels in (b,e,h). (l) Scaled product of V_s kernels in (c,f,i). The target model contains a 6.0 per cent slower V_s perturbation in the white box.

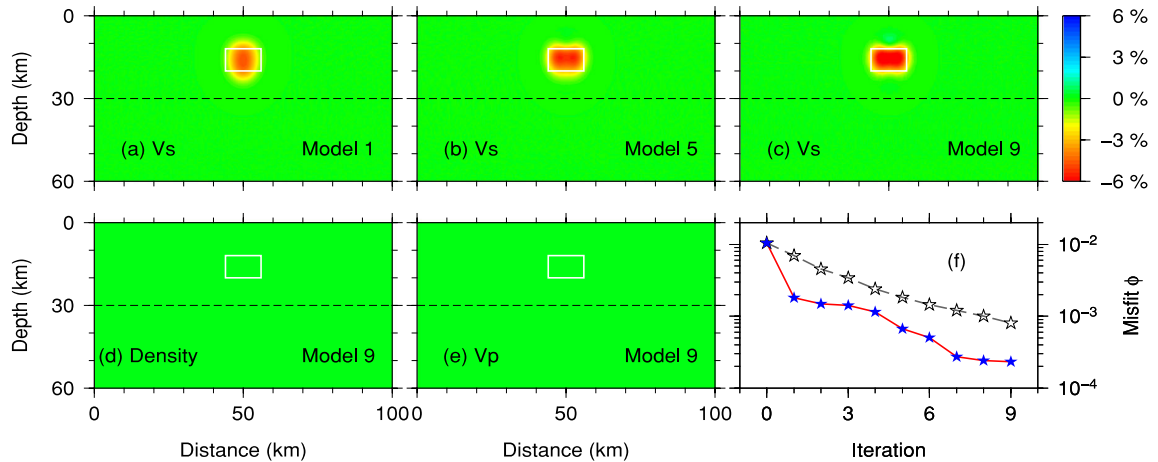


Figure 12. (a)–(e) Same as Fig. 10, but a pre-conditioner from the scaled products of kernels for different coda waves is used in the waveform inversion. (f) Iterative reduction of the misfit function values for inversions with a pre-conditioner (denoted by blue stars) and without a pre-conditioner (denoted by empty stars, also shown in Fig. 10f).

We first start waveform inversions at 10 s, the period at which the wavelength of S waves in the mantle is comparable to the thickness of the slab. The iteratively updated density, V_p and V_s models are shown in Fig. 15(a). The inversion is stopped at the eighth iteration when no more significant reduction in misfit value is observed (Fig. 15c). From model 8 in Figs 15(a), it is clear that most of the slab can be recovered by the 10 s synthetic data. Actually, the contour of the anomalous slab emerges even after one iteration. However, the amplitude of the slab anomaly (4 per cent) are much better until the eighth iteration. This agrees with the discussion in Section 4.2.3 that waveform inversions are capable of resolving anomalous structures of a size similar to the wavelength of the seismic waves used.

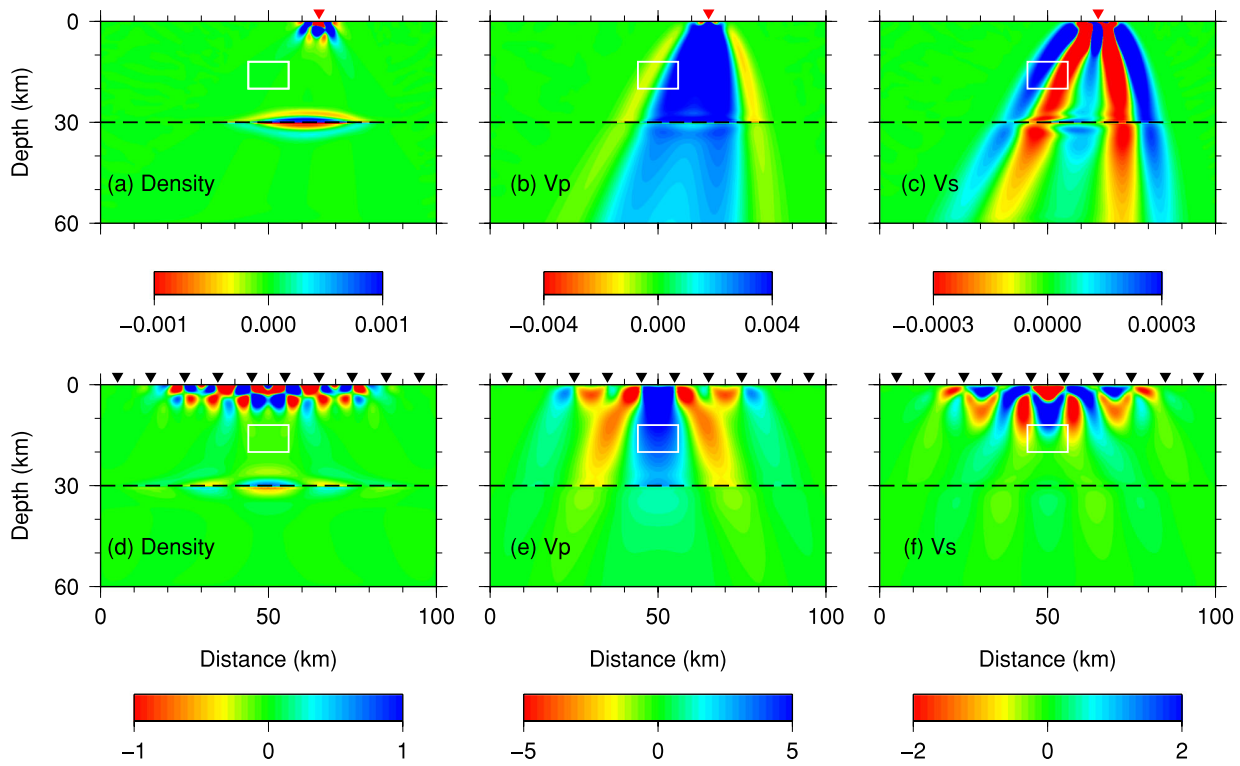


Figure 13. Traveltime sensitivity kernels of density (a), V_p (b) and V_s velocity (c) for the direct P wave of a single event and station pair in the starting model of Fig. 2. The station is placed at $x = 65.0$ km and the event has an incidence angle of 12° . (d)–(f) Unsmoothed gradients of traveltime misfit function for all events and receivers. All these traveltime kernels have the unit of $10^{-8} \text{ s}^2 \text{ m}^{-2}$.

However, the sharp top and bottom boundaries of the slab are not very well determined due to the long-period nature of the data. To further refine the inversion results, we restart the waveform inversion with 2.5 s synthetic data, using the final (eighth) model of the 10 s waveform inversion as the starting model. Six iterations are performed in this case (Fig. 15b), and the 4 per cent density and velocity anomaly in the subducted slab is then almost fully recovered. Slab boundaries are more clearly delineated and match well with the target model, especially for V_p model (Fig. 15b). However, artefacts of low-amplitude slow wave speed or negative density anomalies still exist outside the slab in the final models (Model 6 in Fig. 15b). They appear mostly on the lower side of the slab for V_s model, while in both the crust and the mantle wedge for density model. This is not surprising, as the V_p structures are well recovered from the inversion of combined direct P waves and P -to- P scattering in the coda waves, while density structures can only rely on the backscattered waves recorded in the coda. The V_s model is better inverted than density due to its sensitivity to all coda phases (Section 4.2.1). The existence of artefacts below the slabs (Fig. 15b) in the V_s model compared to the V_p model may be a result of the lack of direct S waves in the inversions.

4.3.2 Two-layer subducted slab model

In the last synthetic example, we consider a more realistic and complex subduction zone model inspired by the Alaska subduction zone model imaged by Rondenay *et al.* (2008) using the Generalized Radon Transform technique. The subducted slab consists of two layers, a slower layer (−6.0 per cent variation in density, V_p and V_s) of subducted oceanic crust atop a fast (4.0 per cent perturbation in all three parameters) subducted lithospheric mantle (Fig. 16b). The anomalies of the slab near the left and right boundaries of the computational domain are tapered to match the 1-D model used in the FK calculation (Fig. 16b). In addition, continental Moho contains a depression of maximal size 10.0 km from its two flat ends. The receiver locations are the same as in the previous example (Section 4.3.1). However, 16 teleseismic events with incidence angles that evenly span $[-30^\circ, 30^\circ]$ at a 4° interval are used to improve image resolution. The teleseismic P wave and all of its coda waves are used to invert for both volumetric anomalies and Moho variations from a two-layer reference model (Fig. 16a). We compute Fréchet kernels for the Moho depth, density, V_p and V_s , and simultaneously invert for both volumetric structural changes and discontinuity variations.

Again, a hierarchical inversion strategy is employed from long to short period progressively through 10, 6.25, 5, 4 and 2.5 s waveform data. Inversions proceed to a shorter period when no significant improvement in misfit value is observed in inversion of longer period data, while the final model of the longer period inversion is used as the initial model (i.e. Model 0) for the ensuing shorter-period inversion.

The Moho depth (Fig. 16c), density (Fig. 16d), V_p (Fig. 16e) and V_s (Fig. 16f) misfit kernels for the 10 s waveform data of direct P and coda waves in the starting model (Fig. 16a) clearly highlight the necessary model update in the first iteration. The positive Moho misfit kernel (Fig. 16c) implies a deepening of the central segment of the Moho. The observed large positive values of the density, V_p and V_s kernels in

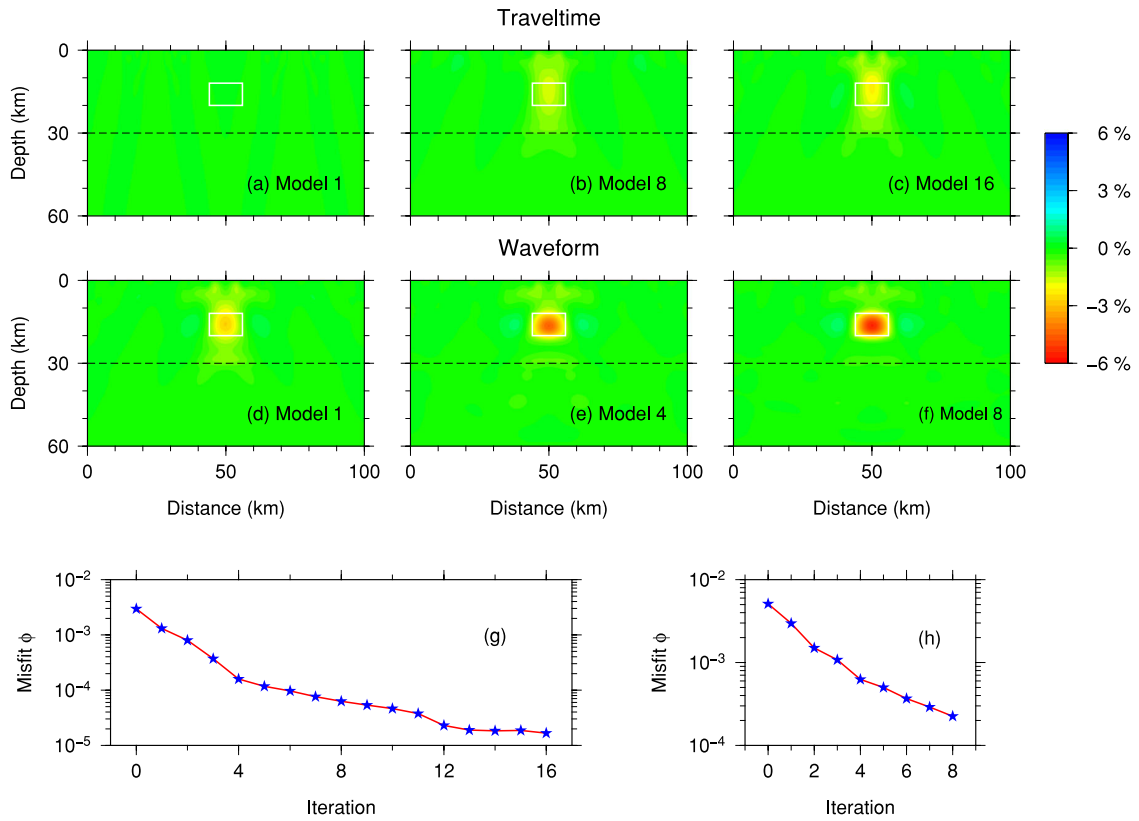


Figure 14. Iteratively updated V_p models in a combined traveltime and waveform inversion. For the target model, a -6.0 per cent V_p perturbation with respect to the surrounding background two-layer model is placed in the white box. The dashed black curve denotes the Moho. (a)–(c) Iteratively updated V_p models by inverting traveltime anomalies of the direct P wave. The starting model \mathbf{m}_0 is the two-layer model of Fig. 2 and Table 1. (d)–(f) Updated models by waveform inversion of coda waves following the direct P phase. The 2-D model in (c) is used as the starting model. (g) and (h) Iterative reduction of the misfit function value of traveltime inversion (g) and coda waveform inversion (h). The units of the misfit functions in (g) and (h) are s^2 and 10^{-6} s^2 , respectively.

the vicinity of the Moho suggest that a local decrease in the density, V_p and V_s may also reduce waveform misfit, which is another indication of a deeper central segment of the Moho. Kernel amplitudes in the vicinity of the subducted slab are much weaker than those around the Moho, mainly due to the fact that density/velocity perturbations in the subducted slab is either -6.0 per cent or 4.0 per cent, much smaller than the material property contrast between the crust and the mantle. It can thus be predicted that the first model update will be predominantly corrections to the Moho topography, and density/velocity adjustments in the slab will follow in later iterations. Nevertheless, some patches of negative kernel values can be observed in the bottom layer of the subducted slab (Figs 16d and f), indicating a necessary local increase in density and V_s in the first model update. As the discontinuity kernel has a different dimension from density and velocity kernels, we have non-dimensionalized all kernels in the inversion process based on eq. (18).

Fig. 17 shows the iteratively updated Moho discontinuity topography and V_p models through inversions of successively shorter period waveform data. Indeed, Moho topography is almost fully recovered by the 10 s and 6.25 s data (see the inverted Moho represented as the white curve and the actual Moho as the black curve in Model 5 of Fig. 17b). However, the two-layer slab, especially its thin low-velocity top layer, is much less well imaged by data up to 6.25 s because the thickness of the top layer, 14 km , is much shorter than the wavelength of 6.25 s P and S waves in the mantle ($\sim 50 \text{ km}$ and 28 km). Only a very weak fast-velocity anomaly (~ 3 per cent) in the bottom layer emerges after 6.25 s inversions (Model 5 in Fig. 17b), probably thanks to the 37 km thickness of the faster bottom layer, which is more comparable to the wavelength of 6.25 s S waves in the mantle (28 km). Short-period data are obviously critical in refining structures and obtaining true amplitudes of the slab anomalies.

The subsequent inversions of 5 and 4 s data gradually reveal the slow top layer of the slab (Figs 17c and d), as the wavelength of 4 s S wave in the mantle (18 km) approaches the thickness of the top layer (14 km). However, artefacts around the Moho and in the upper crust (Model 8 in Fig. 17d) still persist, maybe owing to the trade-off between variations in Moho depth and volumetric parameters. To refine slab structures and reduce these artefacts, 2.5 s waveform data are further inverted. After another eight iterations, the artefacts fade away, and the slow anomalies on the top and fast anomalies on the bottom (Model 8 in Fig. 17e) layer of the slab are well constrained within the assumed regions. We take the density, V_p , V_s and Moho depth model of the last iteration of 2.5 s waveforms as our final model (Fig. 18). The amplitudes of the recovered density, V_p and V_s anomalies are close to -6 per cent and 4 per cent, which are the values that were used to construct the top and the bottom layers of the slab. Minor weak artefacts in the vicinity of the Moho are still visible in the final density (Fig. 18a) and V_s (Fig. 18c) models, the reduction of which may require further inversions of shorter period data, and/or more events for wider illumination angles.

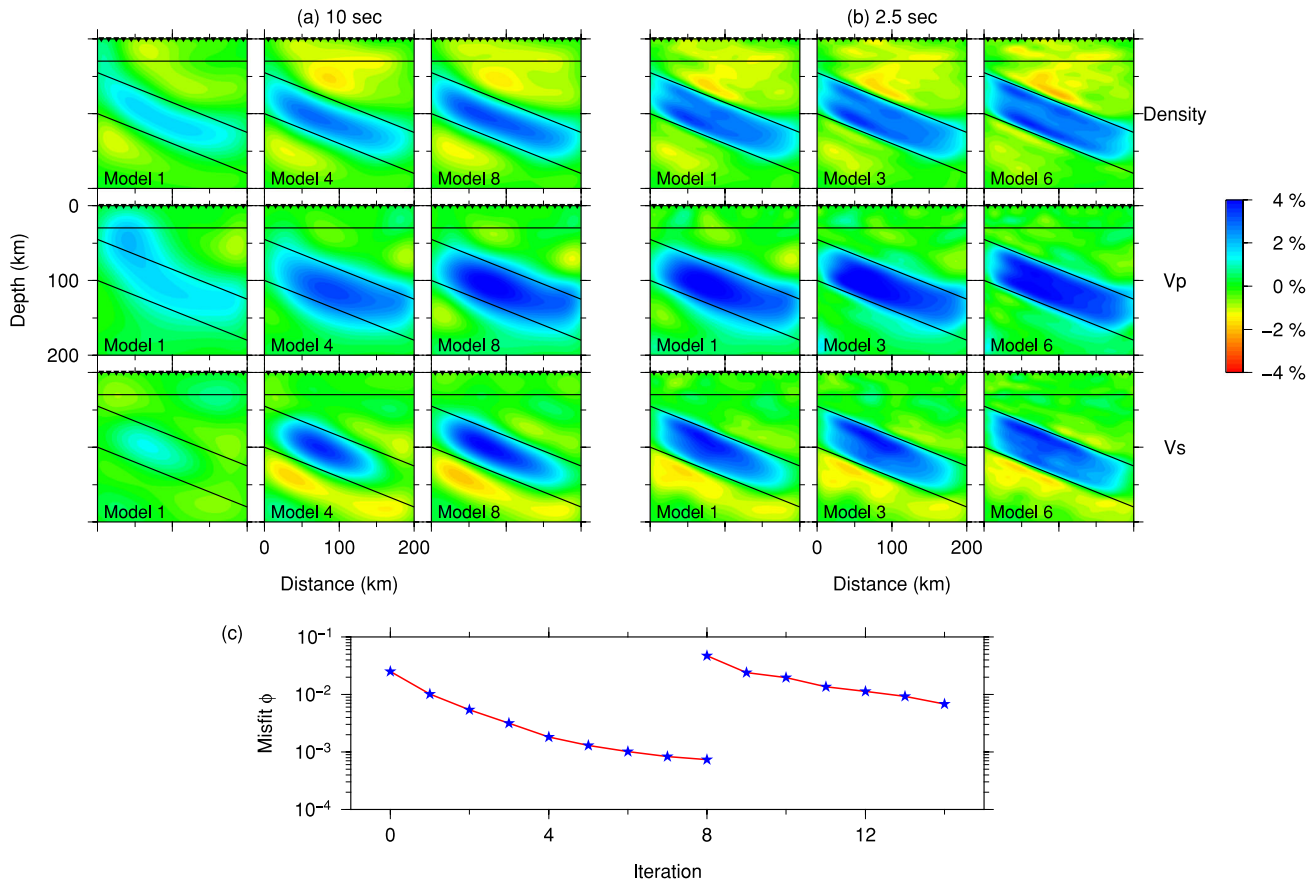


Figure 15. Iteratively updated density, V_p and V_s models from waveform inversions of the direct P phase and all of its coda waves. The target model contains a fast slab in a two-layer background model (Fig. 4j). (a) Density, P - and S -wave velocity models obtained by waveform inversions of 10 s synthetic data at the first, fourth and eighth iteration. (b) Similar to (a), but generated by inversion of 2.5 s synthetic data with ‘Model 8’ in (a) as the starting model. (c) Iterative reduction of the misfit function value for waveform inversions of first 10 s and then 2.5 s data. The unit of the misfit function is 10^{-6} s m^2 .

Overall, a total of 32 iterations are performed in the whole inversion process for various period bands. Figs 19(a)–(c) and (g)–(i) show the horizontal and vertical displacement seismograms of an event with an incidence angle of 18° for the target model, the initial two-layer model \mathbf{m}_0 , model \mathbf{m}_{23} from the 23-rd iteration (the final model of 4 s inversion, i.e. Model 8 in Fig. 17d), and model \mathbf{m}_{31} from the 31-st iteration (the final model of 2.5 s inversion, i.e. Model 8 in Fig. 17e). The station is located at $x = 95 \text{ km}$, and the cut-off period of these seismograms is 2.5 s. The differences between the seismograms for the target model and three other models are displayed in Figs 19(d)–(f) and (j)–(l). Clearly, waveform misfits for successive models (\mathbf{m}_0 , \mathbf{m}_{23} and \mathbf{m}_{31}) are significantly reduced, both for the direct P waves and coda phases. The synthetics for the final model match those of the target model almost exactly. The evolution of the misfit value in the hierarchical waveform inversion process is shown in Fig. 19(m). Significant reduction is achieved in each iterative inversion process of a particular period-band data.

For a complicated subduction zone model that includes a depression in the Moho and a lower velocity subducted oceanic crust atop a fast subducted lithospheric mantle, more sophisticated hierarchical inversion techniques have to be employed to ensure the approximate linearity of waveform misfits. Therefore, this synthetic example not only demonstrates a successful application of high-resolution array imaging by the teleseismic main phase and coda waves based on an SEM-FK hybrid method but also serves as another reminder of the high non-linearity of the inverse problem itself and the importance of carefully designed inversion strategies (e.g. Liu & Gu 2012).

5 DISCUSSIONS AND FUTURE WORKS

The SEM-FK hybrid method developed in this study provides a tool for the application of adjoint tomography to teleseismic converted/scattered wave imaging of crust and upper-mantle structures beneath seismic arrays. The interfacing of analytical wavefields for 1-D background models and numerically calculated full wavefield for 2-/3-D inhomogeneous media alleviates us from the currently prohibitively expensive computation of global seismic wave propagation at frequencies relevant to high-resolution array imaging. This cost limitation will of course become less stringent with the future evolution of supercomputer technology, but there is no hope of being able to use full 3-D simulations globally for inverse problems at such high frequencies before at least a decade, maybe even two. The response of local anomalies to incoming plane waves is accurately captured by the SEM-FK hybrid method that we have introduced. The planar incident wavefront assumption is valid at regional scales of several hundreds of kilometres (e.g. Rondenay *et al.* 2008), hence applicable to most seismic array deployments.

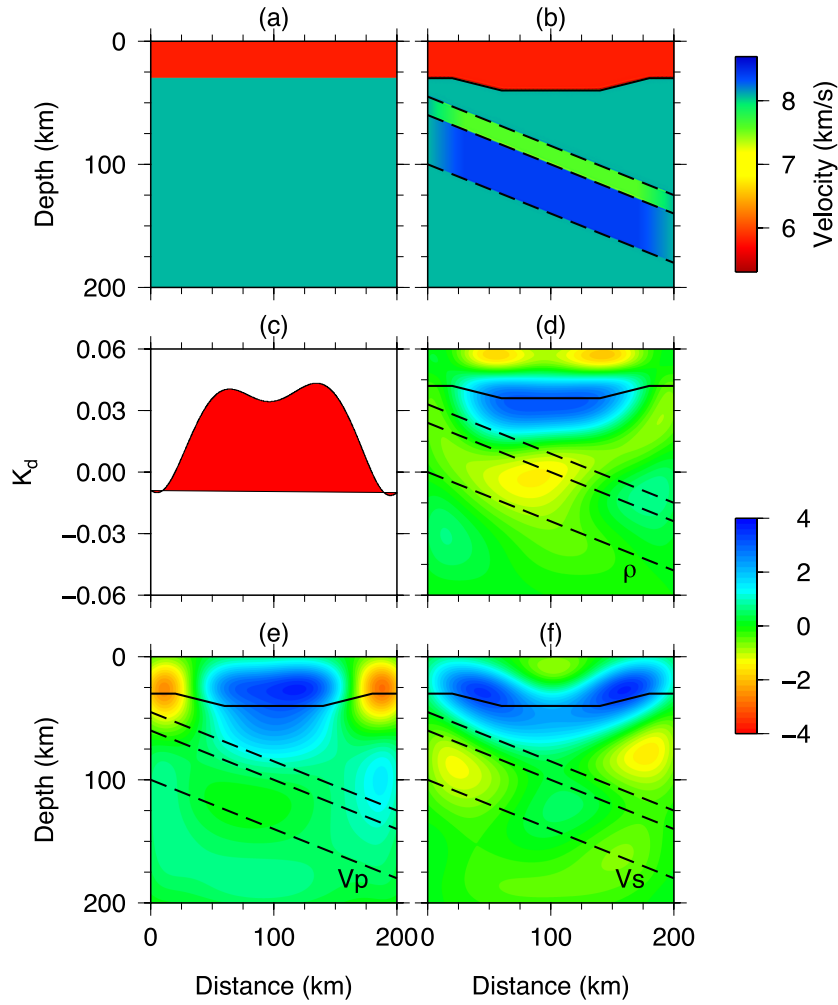


Figure 16. The starting (a) and target (b) V_p model for waveform inversion of a two-layer subducted slab model. Moho discontinuity (c), density (d), V_p (e) and V_s (f) kernels for teleseismic P wave and its coda waves in the starting model. The cut-off period of the incoming P waves is 10 s. The unit of the discontinuity kernel in (c) is 10^{-6} m·s, and the unit of the volumetric kernels in (d)–(f) is 10^{-14} s.

Adjoint tomography based on the SEM-FK hybrid method is advantageous over traditional RF type of techniques (including migration methods such as CCP stacking) in several aspects. First, RF studies that obtain point estimates of the Moho depth assume a flat Moho (Zhu & Kanamori 2000) and may produce biased estimates for regions with large variations in Moho topography (Rondenay 2009), while CCP stacking methods generally assume a smoothly varying 1-D background model and are only capable of mapping out discontinuities. In comparison, adjoint tomography based on an SEM-FK hybrid method computes the forward and adjoint wavefields numerically, thus avoiding ray approximation and making no limiting assumption on the Moho geometry and background model. Therefore, it potentially allows for the use of 2-/3-D heterogeneous models, determined by traditional body- and surface wave methods, as initial models, thus reducing the non-linearity of waveform misfit and the trade-off between Moho variations and velocity perturbations. Secondly, adjoint tomography accurately calculates sensitivity kernels for converted and scattered waves that map inaccuracies of density, V_p , V_s and Moho topography in the current model, without the necessity of individual phase identification. The formal iterative optimization process sharpens models successively to fit the traveltime and waveform of both the main phase (e.g. P) and coda waves. In the two-layer subducted slab example (Section 4.3.2), discontinuity topography, density, V_p and V_s are updated simultaneously. Both structures and true amplitudes of the slab anomalies are well recovered in 32 iterations.

The use of inverse scattering techniques (such as the Generalized Random Transform) has made it possible to produce more physical images of density, V_p and V_s variations compared to reflectivity images from RF migration methods (Beylkin 1985; Bostock & Rondenay 1999; Rondenay *et al.* 2008). However, the sparse and uneven distribution of sources (or incident angles) and receivers degrades the accuracy of the amplitudes of the density, V_p and V_s images. The wave-equation pre-stack depth migration method developed by Shang *et al.* (2012) cross-correlates the reverse-migrated P - and S -wave components, where reverse-time migration is performed by a FD calculation and allows heterogeneous background model and variations of interfaces, hence superior to CCP stacking for complex geological environments. However, it is effectively a one-step inversion, and the use of correlation between only P and S components places stringent requirements on receiver spacing ($\sim 1/5$ of the wavelength of the S wave), thus limiting its applicability in the case of most seismic arrays in practice (Shang *et al.*

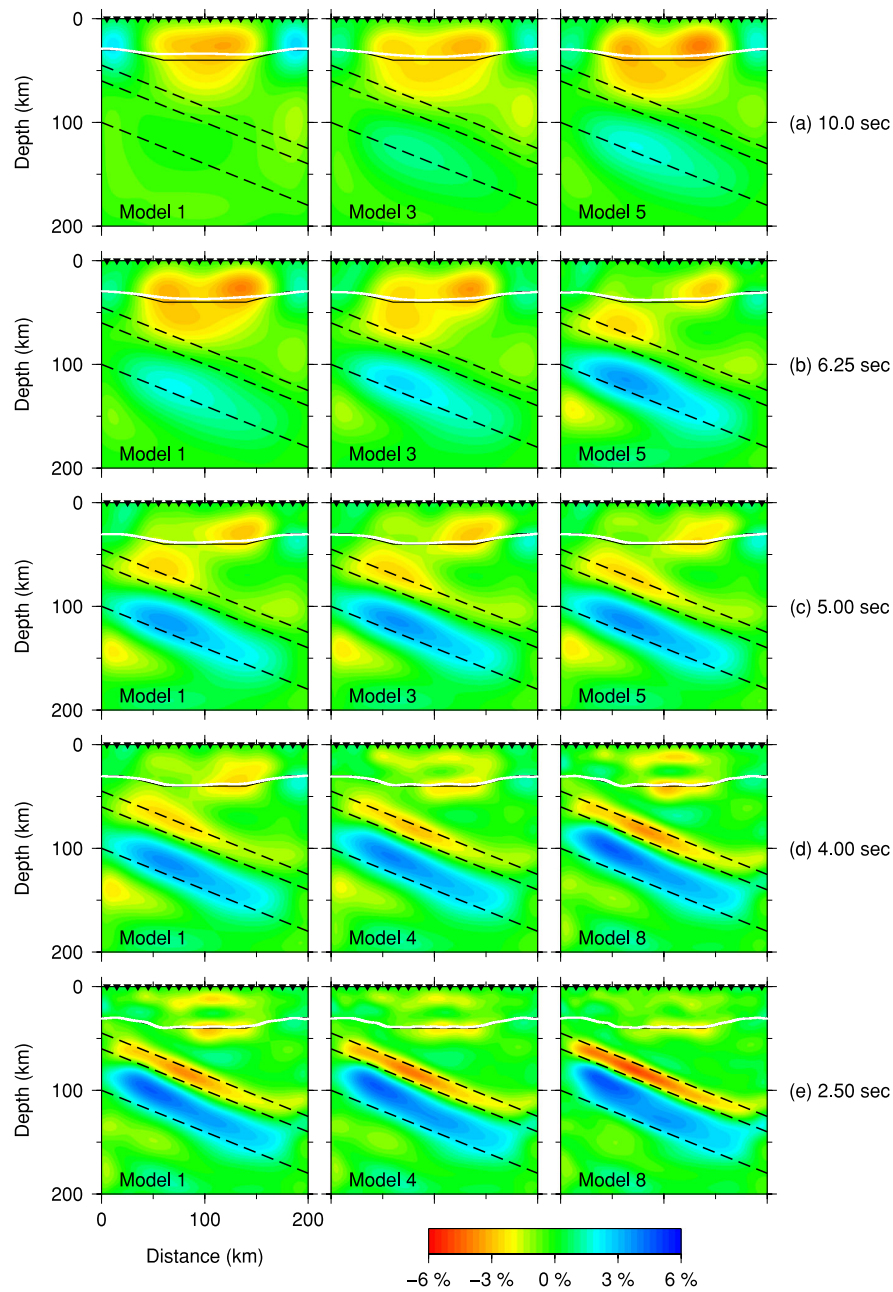


Figure 17. Iteratively updated Moho and V_p structures obtained by successive inversions of (a) 10 s, (b) 6.25 s, (c) 5 s, (d) 4 s and (e) 2.5 s waveforms. The white curves represent the inverted Moho, while the solid black curves denote the true Moho topography. The two-layer subducted slab is outlined by black dashed lines.

2012). Examples in our study show that with an SEM-FK hybrid method and adjoint techniques, receiver spacing may range from one to six times the dominant S -wave length to achieve waveform convergence, depending on the complexity of target models. Therefore, adjoint tomography based on an SEM-FK hybrid method can be readily used for imaging in most seismic arrays.

Adjoint tomography based on SEM-FK methods for converted/scattered waves, similar to all waveform inversion problems, is prone to entrapment in local minima, and designing proper inversion strategies that ensures that the process reaches the global minimum remains challenging. Our study shows that scaled product of sensitivity kernels for different coda phases provide a good pre-conditioner to preferentially update regions indicated by all phases. The combination of finite-frequency traveltimes inversion of the main phase and waveform inversion of coda waves helps determine both long- and short-wavelength structures. Hierarchical inversions from long- to short-period waveforms reduce non-linearity at each stage, and by speeding up convergence, significantly reduce the numerical cost of 2-/3-D inversions. Other pre-conditioners and practices that are popular in the field of full-waveform inversion (e.g. Virieux & Operto 2009) could be exploited as well.

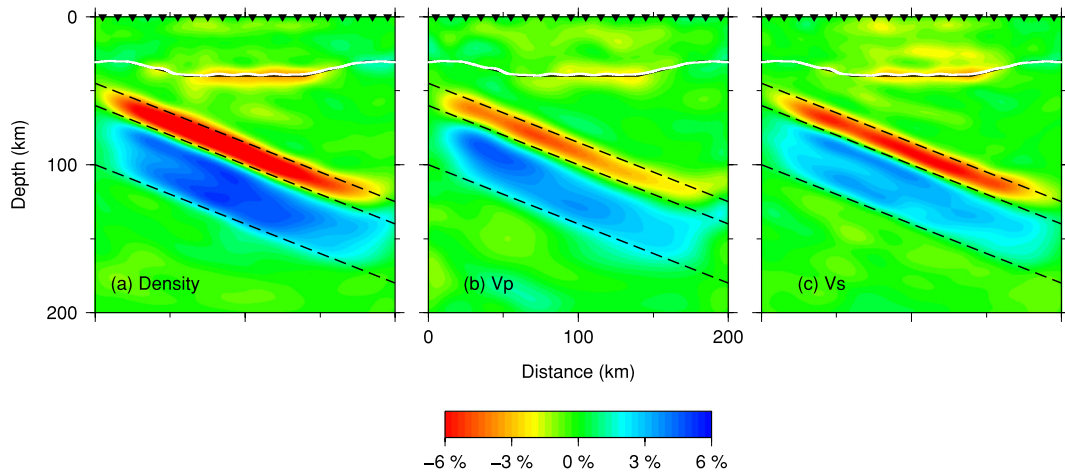


Figure 18. Final density (a), V_p (b) and V_s (c) models after a total of 32 iterations. The white curves denote the location of the inverted Moho, while the black solid curves denote the true Moho topography. The two-layer subducted slab is outlined by the black dashed lines.

Although seismic waves propagation at local scales requires either 2.5-D (Roecker *et al.* 2010) or 3-D (Komatitsch *et al.* 2004) modelling, and seismic array are rarely precisely linear, we believe that our exhaustive 2-D synthetic tests in this study could provide useful guidance for future applications of more realistic 2.5- or 3-D modelling of teleseismic data. Our future work will focus on (1) application of 2-D SEM-FK adjoint tomography to realistic seismic array data in the 2.5-D framework (Roecker *et al.* 2010; Zhou *et al.* 2012; Xiong *et al.* 2013; Pageot *et al.* 2013), (2) extension of implementation and application to the 3-D case and (3) inclusion of seismic anisotropic structural parameters that provide possible clues to tectonic fabric and mantle flow patterns. Meanwhile, challenges posed by the estimation of source-time functions of the incident field, as well as the effects of noise in recorded scattered waves need to be carefully addressed in real applications. All these efforts should make adjoint tomography based on the SEM-FK hybrid method a useful tool in high-resolution imaging of tectonically and geodynamically complex regions beneath dense seismic arrays in the near future.

6 CONCLUSIONS

We have developed and implemented an SEM-FK hybrid method that interfaces FK solutions for 1-D background media with accurate SEM simulations of plane-wave propagation in 2-D local media. It takes into account complex wave phenomena associated with interactions of incoming teleseismic wavefield with local heterogeneities. The much smaller local computational domain permits the application of SEM for high-frequency body waves (~ 1 s) with modest computational resources. The Fréchet kernels constructed based on the hybrid method naturally take into account the finite-frequency effect of seismic wave. Adjoint tomography that employs SEM-FK methods for forward calculations provides a powerful high-resolution imaging tool for teleseismic converted/scattered waves beneath seismic arrays. A non-linear CG method is used for the adjoint tomography to avoid the costly computation of the Hessian matrix. Our 2-D synthetic examples show promising results of this technique in determining density, V_p and V_s anomalies of the size of the dominant wavelength of S waves and with a magnitude of up to ± 6 per cent, as well as large Moho variations. Combination of traveltime and waveform inversions is ideal in recovering both the long- and short-wavelength structures while keeping the inverse problem in the quasi-linear regime. For tectonically complex regions (e.g. subduction zones), we have used a hierarchical strategy from long to short periods to deal with the very non-linear character of waveform misfits, and we have successfully recovered both volumetric anomalies and Moho variations. These effective inversion strategies, together with the SEM-FK hybrid method, pave the way for future high-resolution regional imaging based on realistic seismic array recordings.

ACKNOWLEDGEMENTS

This research was supported by the G8 Research Councils Initiative on Multilateral Research Grant and the Discovery Grants of the Natural Sciences and Engineering Research Council of Canada (NSERC). Computations for this study were performed on hardware purchased through the combined funding of Canada Foundation for Innovation (CFI), Ontario Research Fund (ORF) and University of Toronto Startup Fund, and partly hosted by the SciNet HPC Consortium. SciNet is funded by CFI under the auspices of Compute Canada, the Government of Ontario, ORF-Research Excellence and the University of Toronto. We thank the developers of the SPECfem2D package for their continued community support. We also thank Stéphane Rondenay, Jeroen Tromp and Shu-huei Hung for their helpful comments and suggestion. We are grateful to Dr Andrea Morelli and two anonymous reviewers for their critical comments and suggestions that have improved the original manuscript. Many figures were made with the Generic Mapping Tools (GMT; Wessel & Smith 1991).

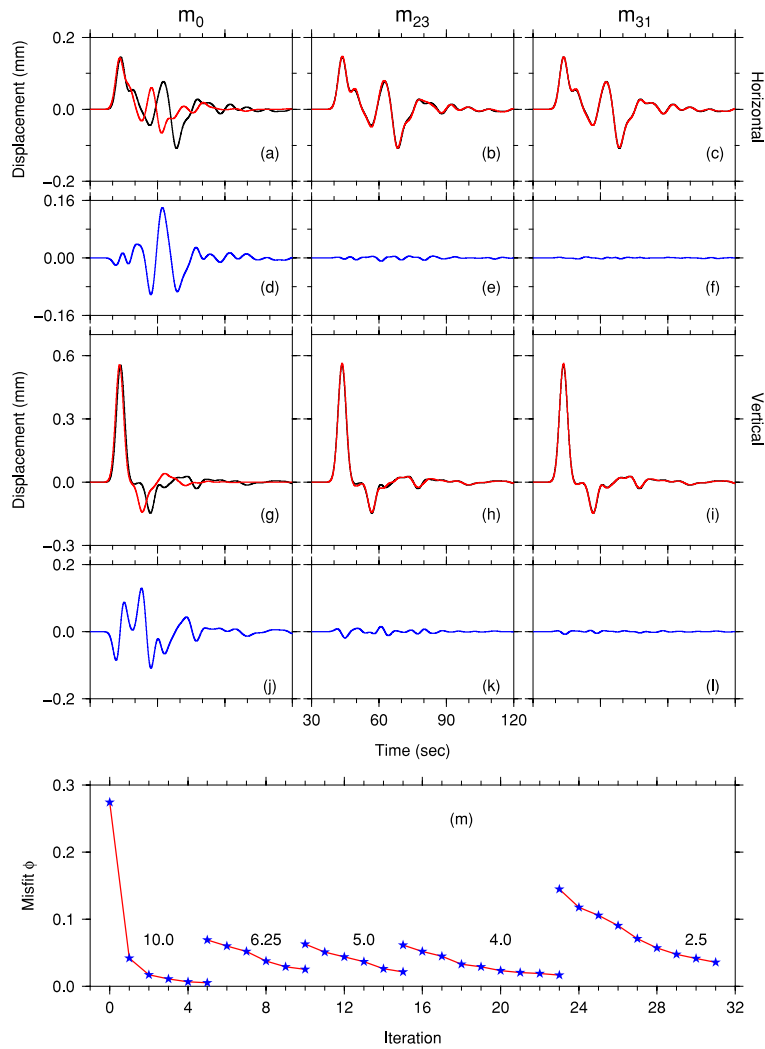


Figure 19. Comparison of horizontal (a–f) and vertical (g–l) component displacement seismograms in the starting model, m_0 (left-hand panel), the final model of the 4 s waveform inversions, m_{23} (middle panel) and the final model, m_{31} (right-hand panel). The differences between the synthetic data for the target model (black traces in a–c, g–i) and the synthetic seismograms for iteratively updated models (red traces) are shown as blue traces (d–f, j–l). These seismograms are generated for an event with a P incidence angle of 18° and recorded at $x = 95.0$ km on the surface. (m) Iterative reduction of the misfit function value for 10, 6.25, 5.0, 4.0 and 2.5 s waveform inversions. The unit of the misfit function is 10^{-6} s m^2 .

REFERENCES

- Aki, K., Christofferson, A. & Husebye, E., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.*, **82**, 277–296.
- Beylkin, G., 1985. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform, *J. Math. Phys.*, **26**, 99–108.
- Bielak, J. & Christiano, P., 1984. On the effective seismic input for non-linear soil-structure interaction systems, *Earthq. Eng. Struct. Dyn.*, **12**, 107–119.
- Bielak, J., Loukakis, K., Hisada, Y. & Yoshimura, C., 2003. Domain reduction method for three-dimensional earthquake modeling in localized regions, part I: theory, *Bull. seism. Soc. Am.*, **93**(2), 817–824.
- Blacker, T.D., Bohnho, W.J. & Edwards, T.L., 1994. *CUBIT Mesh Generation Environment. Volume 1: Users Manual*, Sandia National Labs.
- Bostock, M.G. & Rondenay, S., 1999. Migration of scattered teleseismic body waves, *Geophys. J. Int.*, **137**, 732–746.
- Bostock, M.G., Rondenay, S. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 1. Theory for oblique incidence, *J. geophys. Res.*, **106**(12), 30 771–30 782.
- Casarotti, E., Stupazzini, M., Lee, S., Komatitsch, D., Piersanti, A. & Tromp, J., 2008. Cubit and seismic wave propagation based upon the spectral-element method: an advanced unstructured mesh for complex 3D geological media, in *Proceedings of the 16th International Meshing Roundtable*, Vol. **5B.4**, pp. 579–597, Seattle, WA, USA.
- Chaljub, E., Komatitsch, D., Vilotte, J.P., Capdeville, Y., Valette, B. & Festa, G., 2007. Spectral-element analysis in seismology, *Adv. Geophys.*, **48**, 365–418.
- Chang, S.J. & van der Lee, S., 2011. Mantle plumes and associated flow beneath Arabia and East Africa, *Earth planet. Sci. Lett.*, **302**, 448–454.
- Chen, L., Wen, L. & Zheng, T., 2005. A wave equation migration method for receiver function imaging: 2. Application to the Japan subduction zone, *J. geophys. Res.*, **110**, B11310, doi:10.1029/2005JB003666.
- Chen, M., Tromp, J., Helmberger, D.V. & Kanamori, H., 2007. Waveform modeling of the slab beneath Japan, *J. geophys. Res.*, **112**, B02305, doi:10.1029/2006JB004394.
- Chevrot, S., Favier, N. & Komatitsch, D., 2004. Shear wave splitting in three-dimensional anisotropic media, *Geophys. J. Int.*, **159**(2), 711–720.
- Chevrot, S., Monteiller, V., Komatitsch, D., Fuji, N. & Martin, R., 2011. A hybrid technique for 3-D waveform modeling and inversion of high frequency teleseismic body waves, Abstract S11D-05, in *Proceedings of 2011 Fall Meeting, AGU*.
- Clayton, R. & Engquist, B., 1977. Absorbing boundary conditions for acoustic and elastic wave equations, *Bull. seism. Soc. Am.*, **67**, 1529–1540.

- Cohen, G., 2002. *Higher-Order Numerical Methods for Transient Wave Equations*, Springer-Verlag.
- Dahlen, F., Nolet, G. & Hung, S., 2000. Fréchet kernels for finite-frequency traveltimes: I. Theory, *Geophys. J. Int.*, **141**, 157–174.
- De Basabe, J.D. & Sen, M.K., 2007. Grid dispersion and stability criteria of some common finite-element methods for acoustic and elastic wave equations, *Geophysics*, **72**(6), T81–T95.
- Dziewonski, A., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.*, **89**, 5929–5952.
- Fichtner, A., Brian, L.N.K., Igel, H. & Bunge, H.P., 2009. Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**, 1703–1725.
- Forsyth, D.W., Webb, S.C., Dorman, L.M. & Shen, Y., 1998. Phase velocities of Rayleigh waves in the MELT experiment on the east Pacific rise, *Science*, **280**, 1235–1238.
- Frederiksen, A.W. & Revenaugh, J., 2004. Lithospheric imaging via teleseismic scattering tomography, *Geophys. J. Int.*, **159**, 978–990.
- Friederich, W. & Wielandt, E., 1995. Interpretation of seismic surface waves in regional networks: joint estimation of wavefield geometry and local phase velocity. Method and numerical tests, *Geophys. J. Int.*, **120**, 731–744.
- Fukao, Y., Obayashi, M., Inoue, H. & Nenbai, M., 1992. Subducting slabs stagnant in the mantle transition zone, *J. geophys. Res.*, **97**, 4809–4822.
- Fukao, Y., Obayashi, M. & Nakakuki, T., 2009. Stagnant slab: a review, *Annu. Rev. Earth planet. Sci.*, **37**, 19–46.
- Godinho, L., Amado, P.A., Tadeu, A., Cadena-Isaza, A., Smerzini, C., Sanchez-Sesma, F.J., Madec, R. & Komatitsch, D., 2009. Numerical simulation of ground rotations along 2D topographical profiles under the incidence of elastic plane waves, *Bull. seism. Soc. Am.*, **99**, 1147–1161.
- Gu, Y.J., 2010. *Arrays and Array Methods in Global Seismology*, Springer-Verlag.
- Hager, W.W. & Zhang, H., 2006. A survey of nonlinear conjugate gradient methods, *Pac. J. Optim.*, **2**(1), 35–58.
- Haskell, N., 1962. Crustal reflection of plane P and SV waves, *J. geophys. Res.*, **67**(12), 4751–4767.
- Haskell, N.B., 1953. The dispersion of surface waves on multilayered media, *Bull. seism. Soc. Am.*, **43**(1), 17–34.
- Hu, W., Abubakar, A., Habashy, T.M. & Liu, J., 2011. Preconditioned nonlinear conjugate gradient method for frequency domain full-waveform seismic inversion, *Geophys. Prospect.*, **59**(3), 477–491.
- Hung, S.H., Chen, W.-P. & Chiao, L.-Y., 2011. A data-adaptive, multiscale approach of finite-frequency, traveltime tomography with special reference to P and S wave data from central Tibet, *J. geophys. Res.*, **116**, B06307, doi:10.1029/2010JB008190.
- Jiang, G., Zhao, D. & Zhang, G., 2009. Seismic tomography of the Pacific slab edge under Kamchatka, *Tectonophysics*, **465**, 190–203.
- Kim, Y., Liu, Q. & Tromp, J., 2011. Adjoint centroid-moment tensor inversions, *Geophys. J. Int.*, **186**, 264–278.
- Kind, R. & Yuan, X., 2010. Seismic images of the biggest crash on earth, *Science*, **329**(5998), 1479–1480.
- Kind, R., Yuan, X. & Kumar, P., 2012. Seismic receiver functions and the lithosphere-asthenosphere boundary, *Tectonophysics*, **536–537**, 25–43.
- Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral element method for three-dimensional seismic wave propagation, *Geophys. J. Int.*, **139**, 806–822.
- Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation-I. Validation, *Geophys. J. Int.*, **149**(2), 390–412.
- Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation-II. 3-D models, oceans, rotation, and self-gravitation, *Geophys. J. Int.*, **150**(1), 303–318.
- Komatitsch, D., Liu, Q., Tromp, J., Suss, M.P., Stidham, C. & Shaw, J.H., 2004. Simulations of ground motion in the Los Angeles Basin based upon the spectral-element method, *Bull. seism. Soc. Am.*, **94**, 187–206.
- Komatitsch, D., Tsuboi, S. & Tromp, J., 2005. The spectral-element method in seismology, in *Seismic Earth: Array Analysis of Broadband Seismograms: Geophysical Monograph. Vol. 157*, pp. 205–227, eds Levander, A. & Nolet, G., AGU.
- Krishnan, S., Ji, C., Komatitsch, D. & Tromp, J., 2006. Case studies of damage to tall steel moment-frame buildings in Southern California during large San Andreas earthquakes, *Bull. seism. Soc. Am.*, **96**(4A), 1523–1537.
- Langston, C.A., 1977. Corvallis, Oregon, crustal and upper mantle receiver structure from teleseismic P and S waves, *Bull. seism. Soc. Am.*, **67**, 713–724.
- Lee, E.-J. & Chen, P., 2013. Automating seismic waveform analysis for full 3-D waveform inversions, *Geophys. J. Int.*, **194**, 572–589.
- Lee, S.J., Chen, H.W., Liu, Q., Komatitsch, D., Tromp, J. & Huang, B.S., 2008. Three-dimensional simulations of seismic-wave propagation in the Taipei basin with realistic topography based upon the spectral-element method, *Bull. seism. Soc. Am.*, **98**, 253–264.
- Li, C., van der Hilst, R.D., Engdahl, E.R. & Burdick, S., 2008. A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, **9**, Q05018, doi:10.1029/2007GC001806.
- Liu, Q. & Chen, C.W., 2011. High-resolution array imaging using teleseismic converted waves based on adjoint methods, Abstract S13C-03, in *Proceedings of 2011 Fall Meeting*, AGU.
- 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**, 2283–2297.
- Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, **174**, 265–286.
- Maggi, A., Tape, C.H., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**, 257–281.
- Masters, G., Laske, G., Bolton, H. & Dziewonski, A., 2000. The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale: Geophysical Monograph*, Vol. 117, pp. 63–87, AGU.
- Moczo, P., Bystricky, E., Kristek, J., Carcione, J.M. & Bouchon, M., 1997. Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures, *Bull. seism. Soc. Am.*, **87**(5), 1305–1323.
- Monteiller, V., Chevrot, S., Komatitsch, D. & Fuji, N., 2013. A hybrid method to compute short period synthetic seismograms of teleseismic body waves in a 3-D regional model, *Geophys. J. Int.*, **192**, 230–247.
- Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*, Cambridge Univ. Press.
- Nolet, G., Dahlen, F.A. & Montelli, R., 2005. Traveltimes and amplitudes of seismic waves: a re-assessment, in *Seismic Earth: Array Analysis of Broadband Seismograms*, pp. 37–47, eds Levander, A. & Nolet, G., AGU Monograph.
- Obrebski, M., Allen, R.M., Pollitz, F. & Hung, S.H., 2011. Lithosphere-asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities, *Geophys. J. Int.*, **185**, 1003–1021.
- Pageot, D., Operto, S., Vallee, M., Brossier, R. & Virieux, J., 2013. A parametric analysis of two-dimensional elastic full waveform inversion of teleseismic data for lithospheric imaging, *Geophys. J. Int.*, **193**, 1479–1505.
- Peter, D. et al., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.*, **186**, 721–739.
- Pratt, R.G. & Shipp, R.M., 1999. Seismic waveform inversion in the frequency domain, part 2: fault delineation in sediments using crosshole data, *Geophysics*, **64**, 902–914.
- Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, **133**(2), 341–362.
- Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep Earth, *Phys. Earth planet. Inter.*, **178**, 101–135.
- Revenaugh, J., 1995. A scattered-wave image of subduction beneath the transverse ranges, *Science*, **268**(5219), 1888–1892.

- Roecker, S., Baker, B. & McLaughlin, J., 2010. A finite-difference algorithm for full waveform teleseismic tomography, *Geophys. J. Int.*, **181**, 1017–1040.
- Romanowicz, B., 1991. Seismic tomography of the Earth's mantle, *Annu. Rev. Earth planet. Sci.*, **19**, 77–99.
- Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years, *Annu. Rev. Earth planet. Sci.*, **31**, 303–328.
- Romanowicz, B., 2008. Using seismic waves to image Earth's internal structure, *Nature*, **451**, 266–268.
- Rondenay, S., 2009. Upper mantle imaging with array recordings of converted and scattered teleseismic waves, *Surv. Geophys.*, **30**, 377–405.
- Rondenay, S., Abers, G.A. & van Keken, P.E., 2008. Seismic imaging of subduction zone metamorphism, *Geology*, **36**, 275–278.
- Saad, Y., 2003. *Iterative Methods for Sparse Linear Systems*, 2nd edn, Society for Industrial and Applied Mathematics.
- Sandoval, S., Kissling, E., Ansorge, J. & the SVEKALAPKO Seismic Tomography Working Group, 2004. High-resolution body wave tomography beneath the SVEKALAPKO array: II. Anomalous upper mantle structure beneath the central Baltic Shield, *Geophys. J. Int.*, **157**, 200–214.
- Semblat, J.-F., Kham, M. & Bard, P.-Y., 2008. Seismic-wave propagation in alluvial basins and influence of site-city interaction, *Bull. seism. Soc. Am.*, **98**(6), 2665–2678.
- Seriani, G. & Oliveira, S.P., 2008. Dispersion analysis of spectral-element methods for elastic wave propagation, *Wave Motion*, **45**, 729–744.
- Shang, X., de Hoop, M.V. & van der Hilst, D., 2012. Beyond receiver functions: passive source reverse time migration and inverse scattering of converted waves, *Geophys. Res. Lett.*, **39**, L15308, doi:10.1029/2012GL052289.
- Sheehan, A.F., Shearer, P.M., Gilbert, H.J. & Dueker, K.G., 2000. Seismic migration processing of P-SV converted phases for mantle discontinuity structure beneath the Snake River Plain, western United States, *J. geophys. Res.*, **105**(B8), 19 055–19 065.
- Spakman, W. & Nolet, G., 1988. Imaging algorithms, accuracy and resolution in delay time tomography, in *Mathematical Geophysics: A Survey of Recent Developments in Seismology and Geodynamics, Modern Approaches in Geophysics*, pp. 155–187, eds Vlaar, N.J., Nolet, G., Wortel, M.J.R. & Cloetingh, S., Springer.
- Stupazzini, M., 2006. 3D ground motion simulation of the Grenoble Valley by GeoELSE, in *Proceedings of the 3rd International Symposium on the Effects of Surface Geology on Seismic Motion (ESG)*, Grenoble, France.
- Takeuchi, H. & Saito, M., 1972. Seismic surface waves, in *Methods in Computational Physics*, Vol. **11**, no. 1966, pp. 217–295, Academic Press.
- Tape, C., Liu, Q. & Tromp, J., 2007. Finite-frequency tomography using adjoint methods: methodology and examples using membrane surface waves, *Geophys. J. Int.*, **168**, 1105–1129.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**, 988–992.
- 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., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.
- Thomson, W., 1950. Transmission of elastic waves through a stratified solid medium, *J. appl. Phys.*, **21**, 89–93.
- 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., Komatitsch, J. & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Comm. Comput. Phys.*, **3**, 1–32.
- Vandecar, J.C. & Crosson, R.S., 1990. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least-squares, *Bull. seism. Soc. Am.*, **80**, 150–169.
- Vinnik, L.P., 1977. Detection of waves converted from P to SV in the mantle, *Phys. Earth planet. Inter.*, **15**, 39–45.
- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**, WCC1–WCC26.
- Wen, L.X. & Helmberger, D.V., 1998. A two-dimensional P-SV hybrid method and its application to modeling localized structures near the core-mantle boundary, *J. geophys. Res.*, **103**(B8), 17 901–17 918.
- Wessel, P. & Smith, W.H.F., 1991. Free software helps map and display data, *EOS, Tran. Am. geophys. Un.*, **72**(1), 441–445.
- Wu, R.S. & Aki, K., 1985. Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, **50**(4), 582–595.
- Wu, R.S. & Toksoz, M.N., 1987. Diffraction tomography and multisource holography applied to seismic imaging, *Geophysics*, **52**, 11–25.
- Xiong, J.L., Lin, Y., Abubakar, A. & Habashy, T.M., 2013. 2.5-D forward and inverse modelling of full-waveform elastic seismic survey, *Geophys. J. Int.*, **193**, 938–948.
- Yan, Z. & Clayton, R.W., 2007. Regional mapping of the crustal structure in southern California from receiver functions, *J. geophys. Res.*, **112**, B05311, doi:10.1029/2006JB004622.
- Yang, Y. & Forsyth, D.W., 2006. Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern California, *J. geophys. Res.*, **111**, B07306, doi:10.1029/2005JB004180.
- Zahradnik, J. & Moczo, P., 1996. Hybrid seismic modeling based on discrete-wave number and finite-difference methods, *Pure appl. Geophys.*, **148**(1/2), 21–38.
- Zhao, D., 2012. Tomography and dynamics of western-Pacific subduction zones, *Monogr. Environ. Earth Planets*, **1**(1), 1–70.
- Zhao, L., Wen, L.X., Chen, L. & Zheng, T., 2008. A two-dimensional hybrid method for modeling seismic wave propagation in anisotropic media, *J. geophys. Res.*, **113**, B12307, doi:10.1029/2008JB005733.
- Zhao, D., Yamamoto, Y. & Yanada, T., 2013. Global mantle heterogeneity and its influence on teleseismic regional tomography, *Gondwana Res.*, **23**, 595–616.
- Zheng, L., Zhao, Q., Milkereit, B., Grasselli, G. & Liu, Q., 2013. Spectral-element simulations of elastic wave propagation in exploration and geotechnical applications, *Earthq. Sci.*, in press.
- Zhou, Y., Nolet, G. & Dahlen, F.A., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**, 142–168.
- Zhou, B., Greenhalgh, S. & Greenhalgh, M., 2012. Wavenumber sampling strategies for 2.5-D frequency-domain seismic wave modelling in general anisotropic media, *Geophys. J. Int.*, **188**, 223–238.
- Zhu, L. & Kanamori, H., 2000. Moho depth variation in southern California from teleseismic receiver functions, *J. geophys. Res.*, **105**, 2969–2980.
- Zhu, L. & Rivera, L.A., 2002. A note on the dynamic and static displacements from a point source in multilayered media, *Geophys. J. Int.*, **148**, 619–627.
- Zhu, H., Bozdog, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nature Geosci.*, **5**, 493–498.

APPENDIX A: FK METHOD FOR 1-D LAYERED MODELS

For an isotropic earth model, seismic wave propagation outside the source region is governed by the homogeneous elastodynamic equation (in the frequency domain):

$$-\rho\omega^2\mathbf{u} = \nabla \cdot [\lambda(\nabla \cdot \mathbf{u}) + \mu(\nabla\mathbf{u} + \nabla\mathbf{u}^T)], \quad (\text{A1})$$

where \mathbf{u} represents the displacement vector, ρ , λ , μ give the spatial distribution of density and Lamé parameters in the medium and T denotes the matrix transpose operation. For a layered medium, the displacement field \mathbf{u} can be assumed to depend only on horizontal and vertical

coordinates as $\mathbf{u} = \mathbf{u}(x, z; \omega)$ even though the wavefield $\mathbf{u} = (u_x, u_y, u_z)$ can still have motion in all three components. By applying a Fourier transform of x to horizontal wavenumber k and time t to angular frequency ω , and defining

$$\mathbf{u} = [-iy_1(z, k; \omega), y_2(z, k; \omega), y_3(z, k; \omega)], \quad (\text{A2})$$

eq. (A1) can be reduced to two sets of ordinary differential equations (ODEs; e.g. Takeuchi & Saito 1972; Zhu & Rivera 2002):

$$\frac{d}{dz} \begin{bmatrix} y_1 \\ y_3 \\ y_4 \\ y_6 \end{bmatrix} = \begin{bmatrix} 0 & -k & \frac{1}{\mu} & 0 \\ (1-2\xi)k & 0 & 0 & \frac{\xi}{\mu} \\ 4k^2\mu(1-\xi) - \rho\omega^2 & 0 & 0 & k(2\xi-1) \\ 0 & -\rho\omega^2 & k & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_3 \\ y_4 \\ y_6 \end{bmatrix} = \mathbf{M}(\xi, \mu, \rho; \omega, k) \begin{bmatrix} y_1 \\ y_3 \\ y_4 \\ y_6 \end{bmatrix}, \quad (\text{A3})$$

and

$$\frac{d}{dz} \begin{bmatrix} y_2 \\ y_5 \end{bmatrix} = \begin{bmatrix} 0 & \frac{1}{\mu} \\ k^2\mu - \rho\omega^2 & 0 \end{bmatrix} \begin{bmatrix} y_2 \\ y_5 \end{bmatrix} = \mathbf{M}(\mu, \rho; \omega, k) \begin{bmatrix} y_2 \\ y_5 \end{bmatrix}, \quad (\text{A4})$$

where

$$\xi = \frac{\mu}{\lambda + 2\mu}, \quad y_4 = k\mu y_3 + \mu y_{1,z}, \quad y_5 = \mu y_{2,z}, \quad y_6 = -\lambda k y_1 + (\lambda + 2\mu) y_{3,z}.$$

The first ODE set (A3) provides solutions to P - SV (or Rayleigh) waves, which clearly decouple from the SH (or Love) waves from the second ODE set (A4) for a layered model.

By computing the eigenvalues and eigenvectors for matrix \mathbf{M} in eqs (A3) and (A4), we obtain the following general solutions for P - SV (Rayleigh) and SH (Love) waves in a homogeneous layer:

$$\begin{bmatrix} y_1 \\ y_3 \\ y_4 \\ y_6 \end{bmatrix} = \begin{bmatrix} \frac{-iv_s}{k} & i\frac{v_s}{k} & 1 & 1 \\ 1 & 1 & -i\frac{v_p}{k} & i\frac{v_p}{k} \\ 2k\mu\gamma_1 & 2k\mu\gamma_1 & -2i\mu v_p & 2i\mu v_p \\ -2i\mu v_s & 2i\mu v_s & 2k\mu\gamma_1 & 2k\mu\gamma_1 \end{bmatrix} \begin{bmatrix} e^{-iv_s z} & 0 & 0 & 0 \\ 0 & e^{iv_s z} & 0 & 0 \\ 0 & 0 & e^{-iv_p z} & 0 \\ 0 & 0 & 0 & e^{iv_p z} \end{bmatrix} \begin{bmatrix} C_1 \\ C_3 \\ C_4 \\ C_6 \end{bmatrix} = \mathbf{E}(\mu, \xi; \omega, p) \Gamma(z; \omega, p) \mathbf{C}_R \quad (\text{A5})$$

and

$$\begin{bmatrix} y_2 \\ y_5 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -i\mu v_s & -i\mu v_s \end{bmatrix} \begin{bmatrix} e^{-iv_s z} & 0 \\ 0 & e^{iv_s z} \end{bmatrix} \begin{bmatrix} C_2 \\ C_5 \end{bmatrix} = \mathbf{E}(\mu; \omega, p) \Gamma(z; \omega, p) \mathbf{C}_L, \quad (\text{A6})$$

where C_1, C_3, C_4 and C_6 (i.e. \mathbf{C}_R) correspond to the amplitudes of upgoing SV , downgoing SV , upgoing P and downgoing P waves in this layer while C_2 and C_5 (i.e. \mathbf{C}_L) correspond to the amplitudes of upgoing SH and downgoing SH waves in the layer. In the above expressions, α and β denote the P - and S -wave velocity of the layer (they are also referred to as V_p and V_s in later sections), $p = \frac{k}{\omega}$ is the ray parameter (i.e. horizontal slowness, or reciprocal of phase velocity for surface waves),

$$v_p = \omega \left(\frac{1}{\alpha^2} - p^2 \right)^{1/2}, \quad v_s = \omega \left(\frac{1}{\beta^2} - p^2 \right)^{1/2},$$

are the vertical wavenumbers for P and S waves, and

$$\gamma = 2p^2\beta^2, \quad \gamma_1 = 1 - \frac{1}{\gamma}$$

are auxiliary variables.

We seek to compute the displacement field induced by plane waves incident from below a stack of n -layers over a half-space (Fig. 1). Following Zhu & Rivera (2002), for the m -th layer ($m \in \{1, 2, \dots, n\}$), the variable set \mathbf{y} at the upper layer boundary $z = z_m$ and the lower layer boundary z_{m-1} can be related by the propagator matrix \mathbf{P}_m as

$$\mathbf{y}_m = \mathbf{P}_m \mathbf{y}_{m-1}, \quad (\text{A7})$$

where the propagator matrix for P - SV (Rayleigh) waves \mathbf{P}_m^R and for SH (Love) waves \mathbf{P}_m^L can be written as

$$\mathbf{P}_m^R = \gamma \begin{bmatrix} C_\alpha - \gamma_1 C_\beta & X_\beta - \gamma_1 Y_\alpha & (Y_\alpha - X_\beta)/(2k\mu) & (C_\beta - C_\alpha)/(2k\mu) \\ X_\alpha - \gamma_1 Y_\beta & C_\beta - \gamma_1 C_\alpha & (C_\alpha - C_\beta)/(2k\mu) & (Y_\beta - X_\alpha)/(2k\mu) \\ 2k\mu(X_\alpha - \gamma_1^2 Y_\beta) & 2k\mu\gamma_1(C_\beta - C_\alpha) & C_\alpha - \gamma_1 C_\beta & \gamma_1 Y_\beta - X_\alpha \\ 2k\mu\gamma_1(C_\alpha - C_\beta) & 2k\mu(X_\beta - \gamma_1^2 Y_\alpha) & \gamma_1 Y_\alpha - X_\beta & C_\beta - \gamma_1 C_\alpha \end{bmatrix}, \quad (\text{A8})$$

$$\mathbf{P}_m^L = \begin{bmatrix} C_\beta & Y_\beta/(k\mu) \\ \mu k X_\beta & C_\beta \end{bmatrix}, \quad (\text{A9})$$

where the thickness of the m -th layer is $h = z_m - z_{m-1}$, and

$$C_{\alpha,\beta} = \cos(v_{p,s}h), \quad S_{\alpha,\beta} = -\sin(v_{p,s}h), \quad X_{\alpha,\beta} = \frac{-i v_{p,s} S_{\alpha,\beta}}{\omega p}, \quad Y_{\alpha,\beta} = \frac{i \omega p S_{\alpha,\beta}}{v_{p,s}}.$$

A1 Displacement fields for P plane-wave incidence

For the case of P plane-wave incidence, if we set the bottom interface at $z = 0$ (Fig. 1) and assume that the propagator matrices from the bottom layer $i = 1$ to the top layer $i = n$ are given by $\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n$, then we can propagate the wave solution in the half-space layer at interface $z = 0$ according to eq. (A5),

$$\begin{bmatrix} y_1 \\ y_3 \\ y_4 \\ y_6 \end{bmatrix}_{z=0} = \begin{bmatrix} \frac{-i v_s}{k} & i \frac{v_s}{k} & 1 & 1 \\ 1 & 1 & -i \frac{v_p}{k} & i \frac{v_p}{k} \\ 2k\mu\gamma_1 & 2k\mu\gamma_1 & -2i\mu v_p & 2i\mu v_p \\ -2i\mu v_s & 2i\mu v_s & 2k\mu\gamma_1 & 2k\mu\gamma_1 \end{bmatrix}_0 \begin{bmatrix} C_1 \\ C_3 \\ C_4 \\ C_6 \end{bmatrix} = \mathbf{E}_0(\mu, \xi; \omega, p) \begin{bmatrix} C_1 \\ C_3 \\ C_4 \\ C_6 \end{bmatrix}, \quad (\text{A10})$$

to the surface as

$$\begin{bmatrix} y_1 \\ y_3 \\ y_4 \\ y_6 \end{bmatrix}_{z=z_n} = \mathbf{P}_n \cdots \mathbf{P}_1 \mathbf{E}_0 \begin{bmatrix} C_1 \\ C_3 \\ C_4 \\ C_6 \end{bmatrix} = \mathbf{N} \begin{bmatrix} C_1 \\ C_3 \\ C_4 \\ C_6 \end{bmatrix}, \quad (\text{A11})$$

where

$$\mathbf{N} = [N_{ij}]_{4 \times 4} = \mathbf{P}_n \cdots \mathbf{P}_1 \mathbf{E}_0. \quad (\text{A12})$$

Subscript 0 in \mathbf{E}_0 indicates that material properties (i.e. Lamé parameters) in the half-space below $z = 0$ will be used in its calculation, while subscript i in \mathbf{P}_i indicates that material properties in layer i will be used to compute \mathbf{P}_i .

Coefficients C_i ($i = 1, 3, 4, 6$) in eq. (A11) represent the amplitudes of incident and reflected P and S waves in the half-space, and can be determined by the free surface boundary conditions. As only P plane-wave incidence (corresponding to eigenvector for $e^{-i v_p z}$) is considered, both reflected P and $S(V)$ waves should be observed (corresponding to eigenvectors for $e^{i v_p z}$ and $e^{i v_s z}$) in the half-space, while no incident $S(V)$ waves (corresponding to eigenvectors for $e^{-i v_s z}$) should exist, that is, $C_1 = 0$. If we further assume that the incident P wave

$$\begin{bmatrix} u_x \\ u_z \end{bmatrix}_{z=0} = \begin{bmatrix} -i \\ 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_3 \end{bmatrix}_{z=0} = \begin{bmatrix} -i \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ -i \frac{v_p}{k} \end{bmatrix} C_4 = -i \begin{bmatrix} 1 \\ \frac{v_p}{k} \end{bmatrix} C_4, \quad (\text{A13})$$

has unit amplitude, then

$$C_4 = \pm i \sin \theta, \quad (\text{A14})$$

where θ is the plane-wave incident angle (defined as the angle between the wave propagation direction and the vertical direction shown in Fig. 1). On the other hand, traction at the free surface is given by the last two rows of eq. (A11),

$$\begin{bmatrix} y_4 \\ y_6 \end{bmatrix}_{z=z_n} = \begin{bmatrix} N_{31} & N_{32} & N_{33} & N_{34} \\ N_{41} & N_{42} & N_{43} & N_{44} \end{bmatrix} \begin{bmatrix} 0 \\ C_3 \\ C_4 \\ C_6 \end{bmatrix}, \quad (\text{A15})$$

which when set to zero, helps determine the amplitudes of reflected P - and S -wave in the bottom half-space in terms of C_4 ,

$$\begin{bmatrix} C_3 \\ C_6 \end{bmatrix} = - \begin{bmatrix} N_{32} & N_{34} \\ N_{42} & N_{44} \end{bmatrix}^{-1} \begin{bmatrix} N_{33} \\ N_{43} \end{bmatrix} C_4. \quad (\text{A16})$$

With all the values of C_i ($i = 1, 3, 4, 6$) known, the displacement response at the free surface to an incident P -wave with unit amplitude from below a stack of n -layers can be easily computed based on eq. (A11). The wavefield at any intermediate depth z can be similarly computed with the proper propagator matrices. For the case of an SV or SH plane-wave incidence, the derivation follows similarly (see the Appendix for details).

A2 *SV*-wave case

If *SV* plane-wave incidence (corresponding to eigenvector for $e^{-i\nu_s z}$ in eq. A5) is considered, both reflected *P* and *SV* waves should be observed (corresponding to eigenvectors for $e^{i\nu_p z}$ and $e^{i\nu_s z}$) in the bottom half-space. In this case, C_4 in eq. (A11), which is the coefficient of upgoing *P* waves (corresponding to eigenvector for $e^{-\nu_p z}$), should be set to zero. If the incident *SV* wave, that is,

$$\begin{bmatrix} u_x \\ u_z \end{bmatrix}_{z=0} = \begin{bmatrix} -i \\ 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_3 \end{bmatrix}_{z=0} = \begin{bmatrix} -i \\ 1 \end{bmatrix} \begin{bmatrix} -i\nu_s/k \\ 1 \end{bmatrix} C_1 = \begin{bmatrix} -\nu_s/k \\ 1 \end{bmatrix} C_1, \quad (\text{A17})$$

has unit amplitude, then

$$C_1 = \pm \sin \theta. \quad (\text{A18})$$

As traction at the free surface expressed as

$$\begin{bmatrix} y_4 \\ y_6 \end{bmatrix}_{z=z_n} = \begin{bmatrix} N_{31} & N_{32} & N_{33} & N_{34} \\ N_{41} & N_{42} & N_{43} & N_{44} \end{bmatrix} \begin{bmatrix} C_1 \\ C_3 \\ 0 \\ C_6 \end{bmatrix}, \quad (\text{A19})$$

vanishes by definition, C_3 and C_6 can be calculated from C_1 by

$$\begin{bmatrix} C_3 \\ C_6 \end{bmatrix} = - \begin{bmatrix} N_{32} & N_{34} \\ N_{42} & N_{44} \end{bmatrix}^{-1} \begin{bmatrix} N_{31} \\ N_{41} \end{bmatrix} C_1. \quad (\text{A20})$$

By substituting the values of C_i ($i = 1, 3, 4, 6$) into eq. (A11), the response at the free surface to a unit *SV* plane wave incident from below a stack of n -layers can be obtained, and similarly at any location inside the medium. When interfaced with SEM calculations, it can help model the *S* precursors (e.g. *sP* RFs) for 2-/3-D heterogeneous media (Kind & Yuan 2010; Kind *et al.* 2012).

A3 *SH*-wave case

We can also compute synthetic responses of 1-D local media to *SH* plane-wave incidence from below a stack of n -layers, which can be used to model *SH* coda waves. Setting the bottom interface at $z = 0$ and assuming that the propagator matrix in layer i is given by Q_i , $\{i = 1, \dots, n\}$, we can then propagate the *SH*-wave solution from the bottom layer interface $z = 0$ (based on eq. A6)

$$\begin{bmatrix} y_2 \\ y_5 \end{bmatrix}_{z=0} = \begin{bmatrix} 1 & -1 \\ -i\mu\nu_s & -i\mu\nu_s \end{bmatrix}_{z=0} \begin{bmatrix} C_2 \\ C_5 \end{bmatrix} = \mathbf{E}(\mu_{z=0}; \omega, p) \begin{bmatrix} C_2 \\ C_5 \end{bmatrix} \quad (\text{A21})$$

to the free surface as

$$\begin{bmatrix} y_2 \\ y_5 \end{bmatrix}_{z=z_n} = Q_n \cdots Q_1 \mathbf{E}_0 \begin{bmatrix} C_2 \\ C_5 \end{bmatrix} = \mathbf{L} \begin{bmatrix} C_2 \\ C_5 \end{bmatrix}, \quad (\text{A22})$$

where

$$\mathbf{L} = (L_{ij})_{2 \times 2} = Q_n \cdots Q_1 \mathbf{E}_0, \quad (\text{A23})$$

and subscript 0 in \mathbf{E}_0 indicates that its computation is conducted at $z = 0$ and with the material properties of the bottom half-space. If the incident *SH* wave has unit amplitude, then

$$u_y|_{z=0} = y_2|_{z=0} = C_2 = \pm 1. \quad (\text{A24})$$

Applying the free surface condition, that is, using the fact that the traction at the top interface calculated by

$$y_5|_{z=z_n} = \begin{bmatrix} L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} C_2 \\ C_5 \end{bmatrix} \quad (\text{A25})$$

vanishes, we obtain

$$C_5 = -C_2 L_{21} / L_{22}. \quad (\text{A26})$$

Substituting the values of C_2 and C_5 into eq. (A22), we can compute the response at the free surface to a unit *SH* plane wave incident from below a stack of n -layers, and similarly at any location inside the medium.