

# Refined crustal and uppermost mantle structure of southern California by ambient noise adjoint tomography

Kai Wang,<sup>1</sup> Yingjie Yang,<sup>1</sup> Piero Basini,<sup>2,\*</sup> Ping Tong,<sup>3</sup> Carl Tape<sup>4</sup> and Qinya Liu<sup>2,5</sup>

<sup>1</sup>ARC Centre of Excellence for Core to Crust Fluid System and GEMOC ARC National Key Centre, Earth and Planetary Sciences, Macquarie University, Sydney, 2109, Australia. E-mail: wangkaim8@126.com

<sup>2</sup>Department of Physics, University of Toronto, Toronto, Ontario, M5S 1A1, Canada

<sup>3</sup>Division of Mathematical Sciences, School of Physical and Mathematical Sciences and Asian School of the Environment, Nanyang Technological University, Singapore, 639798, Singapore

<sup>4</sup>Geophysical Institute and Department of Geosciences, University of Alaska Fairbanks, Fairbanks, AK 99775-7320, USA

<sup>5</sup>Department of Earth Sciences, University of Toronto, Toronto, Ontario, M5S 1A1, Canada

Accepted 2018 July 27. Received 2018 July 12; in original form 2018 March 07

## SUMMARY

We construct an improved shear wave velocity ( $V_s$ ) model of the southern Californian crust and uppermost mantle by performing an adjoint tomographic inversion using Rayleigh wave empirical Green's functions (EGFs) at 5–50 s periods from ambient noise cross correlations. Our initial model is the isotropic  $V_s$  model **M16** from Tape *et al.*, which was generated by three-component seismograms at 2–30 s periods from local earthquake data. Synthetic Green's functions (SGFs) from **M16** show good agreement with the EGFs at 5–10 and 10–20 s period bands, but they have an average 2.1 s time advance at 20–50 s. By minimizing the traveltimes differences between the EGFs and SGFs using a gradient-based algorithm, we successively refine the  $V_s$  model, and the total misfit is reduced by  $\sim 76.6$  per cent from 1.75 to 0.41 after five iterations. Relative to **M16**, our new  $V_s$  model reveals: (1) a lower crust (20–30 km) with the mean  $V_s$  about 6 per cent slower; (2) a faster  $V_s$  speed in the middle and lower crust at depths greater than 10 km in the regions beneath the Los Angeles Basin and Central Transverse Range; (3) higher  $V_s$  in the lower crust beneath the westernmost Peninsular Range Batholith (PRB); and an enhanced high-velocity zone in the middle crust beneath Salton Trough Basin. Our updated model also reveals refined lateral velocity gradients across PRB, Sierra Nevada Batholith and San Andreas Fault. Our study demonstrates the improvement of lateral coverage and depth sensitivity from using ambient noise instead of only earthquake data. The numerical spectral-element solver used in adjoint tomography provides accurate structural sensitivity kernels, and hence generates more robust images than those by traditional ambient noise tomography based on ray theory.

**Key words:** Computational seismology; Seismic noise; Waveform inversion; Seismic tomography; Seismic interferometry; Surface wave and free oscillation.

## 1 INTRODUCTION

Regional- and global-scale tomographic models have traditionally been constructed using measurements of body wave traveltimes or surface wave dispersions (including modes) extracted from teleseismic events (e.g. Dziewonski & Anderson 1981; Woodhouse & Dziewonski 1984; Romanowicz 2003; Thurber & Ritsema 2007). However, there are limitations from using teleseismic observations. Teleseismic body-wave tomography has limited resolution at lithospheric depths since ray paths from teleseismic body waves are

nearly vertical at shallow depths and thus hardly cross each other. Surface waves from teleseismic events often lack high-frequency information as a result of intrinsic attenuation and scattering. Meanwhile, local earthquake tomography affords higher resolution imaging, yet it is only feasible in seismically active regions with dense seismic networks, and source parameters may need to be estimated simultaneously (e.g. Thurber 1992; Zhao *et al.* 1992; Shearer 1997; Waldhauser & Ellsworth 2000; Zhang & Thurber 2003). All earthquake-based tomography methods are affected by uneven distributions of sources and receivers.

\* Now at: Total, 64000 Pau, France.

Ambient noise tomography (ANT) complements traditional techniques by offering high-resolution imaging in the absence (or presence) of earthquake data. ANT retrieves empirical Green's functions (EGFs) between a pair of receivers based on the assumption of a diffuse noise field (e.g. Lobkis & Weaver 2001; Snieder 2004; Roux *et al.* 2005; Wapenaar *et al.* 2010). Using this method, Shapiro *et al.* (2005) successfully performed the first ambient noise surface wave tomography, focusing on southern California, where they used Rayleigh waves extracted from cross correlations of ambient noise. Since then, ANT has been widely employed to image crustal and upper-mantle structures from regional scales (e.g. Yao *et al.* 2006; Lin *et al.* 2008; Yang *et al.* 2008) to continental scales (e.g. Yang *et al.* 2007; Bensen *et al.* 2009) and global scales (e.g. Nishida *et al.* 2009; Haned *et al.* 2015). Compared with traditional earthquake-based tomography, ANT is not affected by distributions of earthquakes and can be applied to any seismic array where seismic noises are readily recorded. Most ANT studies extract surface wave dispersions to invert for shear wave velocity structure using a two-step inversion method (Step 1: invert for phase velocity maps; and Step 2: invert for 1-D Vs profile using local dispersion curves extracted from the phase velocity maps constructed from Step 1; e.g. Stehly *et al.* 2009; Zheng *et al.* 2011; Shen *et al.* 2012; Xie *et al.* 2013; Jiang *et al.* 2014; Luo *et al.* 2015; Guo *et al.* 2016) in which ray theory is adopted to describe propagating ray paths of surface waves.

Benefiting from the developments of high-performance computing, it has become practical to simulate seismic wave propagation in complex 3-D earth models using numerical methods such as spectral-element method (SEM; e.g. Komatitsch & Vilotte 1998; Tromp *et al.* 2008; Peter *et al.* 2011). Utilizing forward computation capabilities, adjoint tomography based on SEM has been introduced to earthquake seismology (Tromp *et al.* 2005), realizing the pioneering efforts of the 1980s (Tarantola 1984; Gauthier *et al.* 1986). The tomographic technique has been used to image crustal structures of southern California (Tape *et al.* 2009, 2010; Lee *et al.* 2014) and upper-mantle structures of Australia, Europe, East Asia, North America (Fichtner *et al.* 2009, 2010; Zhu *et al.* 2012, 2015; Chen *et al.* 2015; Zhu *et al.* 2017), among others. The idea of adjoint tomography is to iteratively minimize a misfit function between observations and synthetics by using Fréchet derivatives, which are constructed from the interference between a forward wavefield and an adjoint wavefield (Liu & Tromp 2006). The use of full numerical solvers for both wavefields provides accurate Fréchet kernels for the current 2-D/3-D model at each iteration, and allows more data to be used successively as the model improves. The resulting images typically provide better data fits than those generated by classical tomography based on ray theory or finite-frequency theory using asymptotic or semi-analytical methods for 1-D reference models (Liu & Gu 2012). On the other hand, a hybrid method that interfaces 1-D analytical solution outside study area and 3-D high-resolution numerical solver inside the region (the so-called Box Tomography) has been developed to account for full wave propagation of teleseismic data in regional inversions (e.g. Tong *et al.* 2014a,b; Monteiller *et al.* 2015; Masson & Romanowicz 2017; Clouzet *et al.* 2018) and is a very promising technique for imaging high-resolution regional structure.

In recent years, surface waves extracted from ambient noise have been used in full waveform inversion through scattering-integral method based on finite-difference modelling (e.g. Gao & Shen 2014; Lee *et al.* 2014) and adjoint method based on SEM (e.g. Chen *et al.* 2014; Liu *et al.* 2017). For example, Gao & Shen (2014) used 7–200 s broad-band ambient noise to constrain the upper-mantle

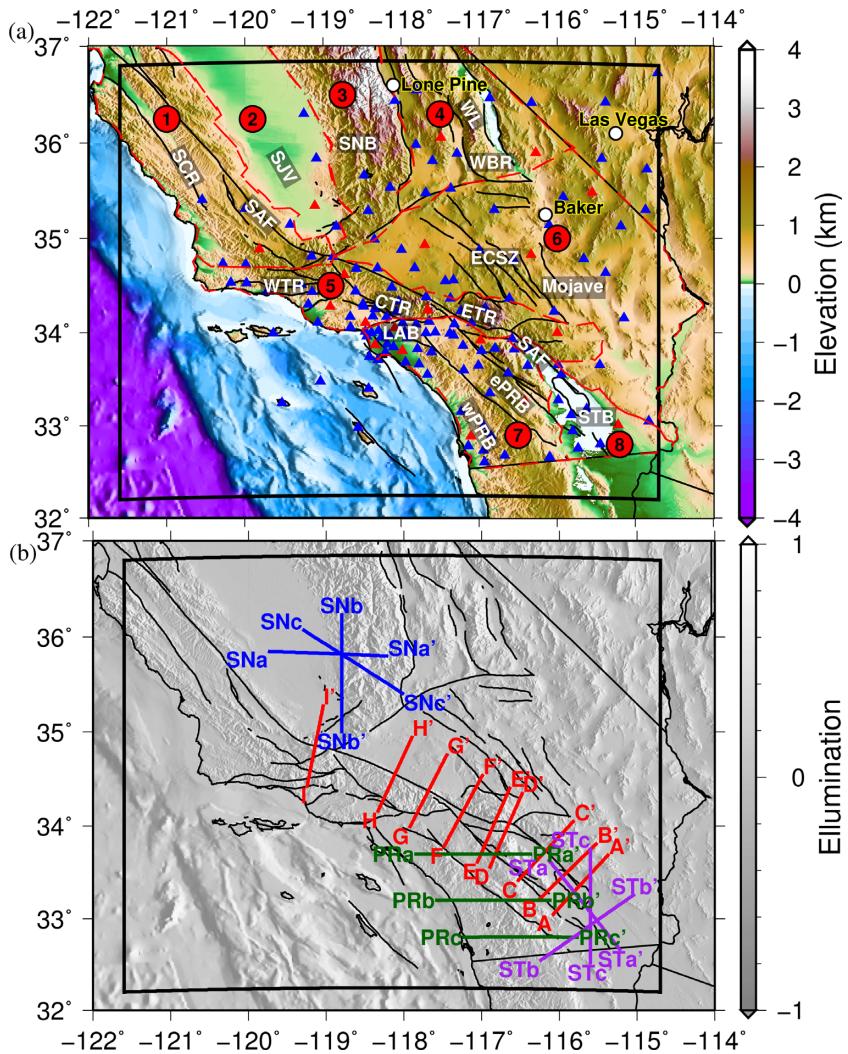
structures of the Cascadia subduction zone. Chen *et al.* (2014) applied adjoint tomography to short-period (10–40 s) ambient noise recorded at 25 stations to refine the crustal model in southeastern Tibet.

Southern California is a tectonically active region with a 200–300 km wide zone of transpressive deformation spanning the Pacific–North America plate boundary (Hauksson 2000). The complex tectonic history has left its marks in the form of several provinces such as the Transverse Ranges, the Peninsular Ranges, the Mojave Desert and the Salton Trough (Fig. 1a). Owing to abundant seismicity and one of the earliest regional permanent broad-band seismic networks, traditional seismic tomographic inversions have been performed based on *P* and *S* arrivals from local (e.g. Hauksson 2000) and teleseismic earthquake data (e.g. Humphreys & Clayton 1990; Schmandt & Humphreys 2010), and surface waves from earthquake data (e.g. Yang & Forsyth 2006) and ambient noise (e.g. Shapiro *et al.* 2005; Zigone *et al.* 2015; Jiang *et al.* 2018). Tape *et al.* (2009, 2010) generated a high-resolution 3-D crustal model M16 by applying adjoint tomography to a data set of waveforms recorded at 203 stations from 143 local earthquakes. Lee *et al.* (2014) applied a combination of scattering-integral and adjoint methods based on finite-difference modelling to both earthquake data and ambient noise at 5–50 s periods. In this study, we apply adjoint tomography to 5–50 s Rayleigh waves extracted from ambient noise to further improve the M16 model, especially structures in the lower crust. We use SEM to simulate wave propagation through M16 and successively update models.

In this paper, we first describe the construction of EGFs from ambient noise in Section 2. We then summarize the basic theory of adjoint tomography in Section 3, and describe the inversion procedures in Section 4. In Section 5, we investigate the characteristics of the initial model and present updated models from three stages of the inversion with increasing data from different period bands. Finally, we show the robustness of our inversion results by misfit evolution analysis, model validation from 143 local earthquakes, and comparison with two other published models.

## 2 AMBIENT NOISE DATA IN SOUTHERN CALIFORNIA

We first build a database of vertical–vertical component cross-correlations of ambient noise in southern California (Fig. 1a). We download vertical components of continuous long-period (LHZ) seismic data recorded at 148 stations between January 2006 and December 2012 from southern California Earthquake Data Center (SCEDC), and process them following the standard noise data processing procedures outlined in Bensen *et al.* (2007) to obtain cross-correlation functions at a period band of 5–50 s. We cut the raw continuous data into daily segments, then remove the trend, mean and instrument responses and bandpass filter them at a 5–50 s period band. Afterwards, we normalize the pre-processed seismograms using the running-absolute-mean normalization method in the time domain and whiten their spectra in the spectral domain to suppress the effects of earthquake signals and instrumental irregularities on cross-correlations. Finally, we compute the daily cross-correlations for each possible station pair and stack them linearly to produce the final cross-correlations. The positive and negative legs of each stacked cross-correlation are stacked further to obtain the symmetric cross-correlations. The resulting cross-correlations are dominated by Rayleigh wave signals as we only process vertical



**Figure 1.** (a) Map of southern California with topography, bathymetry and active faults. The solid black rectangle outlines the simulation region. The 148 stations used in this tomographic study are shown as triangles, out of which 19 are selected for line search shown in red. Faults are indicated by the bold black lines. Labels 1–8 denote the major eight geological provinces with their boundaries delineated by red dashed lines: 1. Coastal Ranges; 2. Great Central Valley; 3. Sierra Nevada; 4. Basin and Range; 5. Transverse Ranges; 6. Mojave Desert; 7. Peninsular Ranges; 8. Salton Trough. Geological features are labelled in bold white letters as references for subsequent figures: SCR, southern Coast Range; SAF, San Andreas Fault; SJV, San Joaquin Valley; SNB, Sierra Nevada Batholith; WL: Walker Lane; WBR: Western Basin and Range; WTR, CTR and ETR: western, central and eastern transverse range; ECSZ: Eastern California shear zone; LAB: Los Angeles Basin; ePRB and wPRB: east and west Peninsular Ranges Batholith; STB: Salton Trough Basin. (b) Locations of cross-sections across San Andreas Fault: AA'-II' (red); Peninsular Ranges: (PR)aa'-cc' (dark green); Sierra Nevada: (SN)aa'-cc' (blue); Salton Trough: (ST)aa'-cc' (purple).

data. In total, we obtain 10 862 cross-correlations with high-quality Rayleigh waves between these stations.

The symmetric cross-correlation function  $C_{a,b}(t)$  between a station pair,  $a$  and  $b$ , is related to its EGF  $G_{a,b}(t)$  as

$$-\frac{\partial C_{a,b}(t)}{\partial t} = G_{a,b}(t), \quad (1)$$

when noise sources are homogenously distributed (e.g. Snieder 2004; Roux *et al.* 2005; Yao & Van Der Hilst 2009).

One potential concern of ANT is the influence of the uneven distribution of noise on the phases of surface wave signals emerging from cross-correlation functions (e.g. Tsai 2009; Yao & Van Der Hilst 2009; Froment *et al.* 2010; Basini *et al.* 2013; Wang *et al.* 2014; Ermert *et al.* 2015; Wang *et al.* 2016). For ambient noise studies at periods shorter than 40/50 s, it has been demonstrated that the use of long time-series of ambient noise data and standard pre-processing procedures (Bensen *et al.* 2007) is sufficient to make

the distribution of effective ambient noise energy nearly isotropic (e.g. Yao *et al.* 2008; Yang *et al.* 2008). At periods longer than 40/50 s, Yanovskaya *et al.* (2016) conclude that earthquake energy constitutes the main contribution to seismic noise, and thus may distort noise cross-correlation if the stacking time is less than 3 yr. In this study, we stack 6 yr of seismic noise data, which should be long enough to minimize such distortions.

In the context of full waveform modelling, Tromp *et al.* (2010) showed that how it is theoretically possible to account for the non-uniform character of the noise sources' distribution. More recently, Fichtner (2014) showed that unevenly distributed noise sources and seismic data processing can both affect cross-correlations of ambient noise and their sensitivities to structures. This implies that carrying out full waveform ANT without taking source heterogeneities and data processing schemes into account could introduce tomographic artefacts. Moreover, Fichtner (2014) also suggested that, compared with waveform-based measurements, narrow-band

phase traveltimes are only weakly affected by source heterogeneities and data processing schemes.

In order to choose proper period bands in measuring phase traveltimes of EGFs, we examine the dominant energy in the spectra of the noise cross-correlations. As seen in Figs 2(b)–(i), our noise cross-correlations generally exhibit two peaks in the spectra between 10–20 s and 5–10 s, which are referred to as primary and secondary microseism bands (Stehly *et al.* 2006). Therefore, in our inversion, we filter waveforms of EGFs in three narrow period bands, namely 5–10, 10–20 and 20–50 s. The noise bands we choose here are slightly wider than what Fichtner (2014) uses to demonstrate the robustness of the phase traveltime misfit, thus the possible effects of such slightly broader bands on the measurements of phase traveltime merit further investigation in the future. The selection of period bands and its effects on tomographic results will be further discussed in detail in the pre-processing steps of the inversion procedures (Section 4) and the Supporting Information.

### 3 BACKGROUND OF ADJOINT TOMOGRAPHY

The theory of adjoint tomography has been well documented in a number of review literatures (e.g. Tromp *et al.* 2005; Fichtner 2010; Chen 2011; Liu & Gu 2012; Luo 2012). Here, we outline the formal expressions for kernel calculations based on adjoint method, the basic concepts of event misfit kernel and gradient used in adjoint tomography.

If we assume  $\phi$  to be a scalar function that quantifies the misfit of measurements between observations and synthetic predictions based on a reference (or current) model  $m(\mathbf{x})$ , then the misfit perturbation can be linearized relative to the reference model.

$$\delta\phi = \int_V K(\mathbf{x}) \delta \ln m(\mathbf{x}) d^3\mathbf{x}, \quad (2)$$

where  $K(\mathbf{x})$  is the Fréchet kernels (also known as *sensitivity kernels*) defined for fractional perturbations of model parameters, such as density, shear and bulk moduli. Based on adjoint methods (Tromp *et al.* 2005; Liu & Tromp 2006), these kernels can be expressed in a general form:

$$K(\mathbf{x}) = \mathbf{u}(\mathbf{x}, t; \mathbf{x}_s) \otimes \mathbf{u}^\dagger(\mathbf{x}, T - t; \mathbf{x}_r), \quad (3)$$

where  $\mathbf{u}(\mathbf{x}, t; \mathbf{x}_s)$  is the forward wavefield emitted from the source,  $\mathbf{u}^\dagger(\mathbf{x}, T - t; \mathbf{x}_r)$  is the adjoint wavefield generated by the time-reversed adjoint sources injected at receivers. The  $\otimes$  operator represents general interactions of these two wavefields for different types of kernel which can be found in Tromp *et al.* (2005).

Adjoint tomography strikes a balance between storage demand and computational cost by only calculating and saving event misfit kernels (Tape *et al.* 2007), instead of individual kernel for each source-receiver pair. Assuming we have a total  $E$  events, each with associated event kernels, then we have in matrix form,

$$\Delta\phi = \mathbf{G}\Delta\mathbf{m}, \quad (4)$$

where  $\Delta\phi$  is the event misfit vector of size  $E \times 1$ ,  $\mathbf{G}$  is an  $E \times M$  matrix with discretized event kernels as row vectors and  $\Delta\mathbf{m}$  is the  $M \times 1$  model column vector. For most inversions,  $M$  is much larger than  $E$ , and eq. (4) is an underdetermined system that cannot provide a unique model update. There are numerous options to derive the model update using event kernels (Modrak & Tromp 2016). We choose an update that is linearly proportional to the sum of event kernels determined as follows. Considering the quadratic Taylor

expansion of misfit function (Tarantola 2005) :

$$\phi(\mathbf{m} + \delta\mathbf{m}) \approx \phi(\mathbf{m}) + \mathbf{g}^T \delta\mathbf{m} + \frac{1}{2} \delta\mathbf{m}^T \mathbf{H} \delta\mathbf{m} + O(\delta\mathbf{m}^3), \quad (5)$$

where the misfit gradient  $\mathbf{g}$  and Hessian  $\mathbf{H}$  are expressed as

$$\mathbf{g} = \left. \frac{\partial \phi}{\partial \mathbf{m}} \right|_{\mathbf{m}}, \mathbf{H} = \left. \frac{\partial^2 \phi}{\partial \mathbf{m} \partial \mathbf{m}} \right|_{\mathbf{m}}. \quad (6)$$

The total misfit can be reduced by either updating the model based on the negative gradient direction

$$\delta\mathbf{m} \propto -\mathbf{g} \quad (7)$$

or using the Hessian-based algorithm

$$\delta\mathbf{m} = -\mathbf{H}^{-1} \mathbf{g}, \quad (8)$$

where the gradient is ‘pre-conditioned’ by the generalized inverse of the Hessian. In adjoint tomography, we do not have access to individual kernels or the Hessian matrix. Nevertheless, a variety of pre-conditioners can be used to approximate the Hessian and applied to the gradient vector to improve convergence (see also Section 4.6).

### 4 INVERSION PROCEDURES

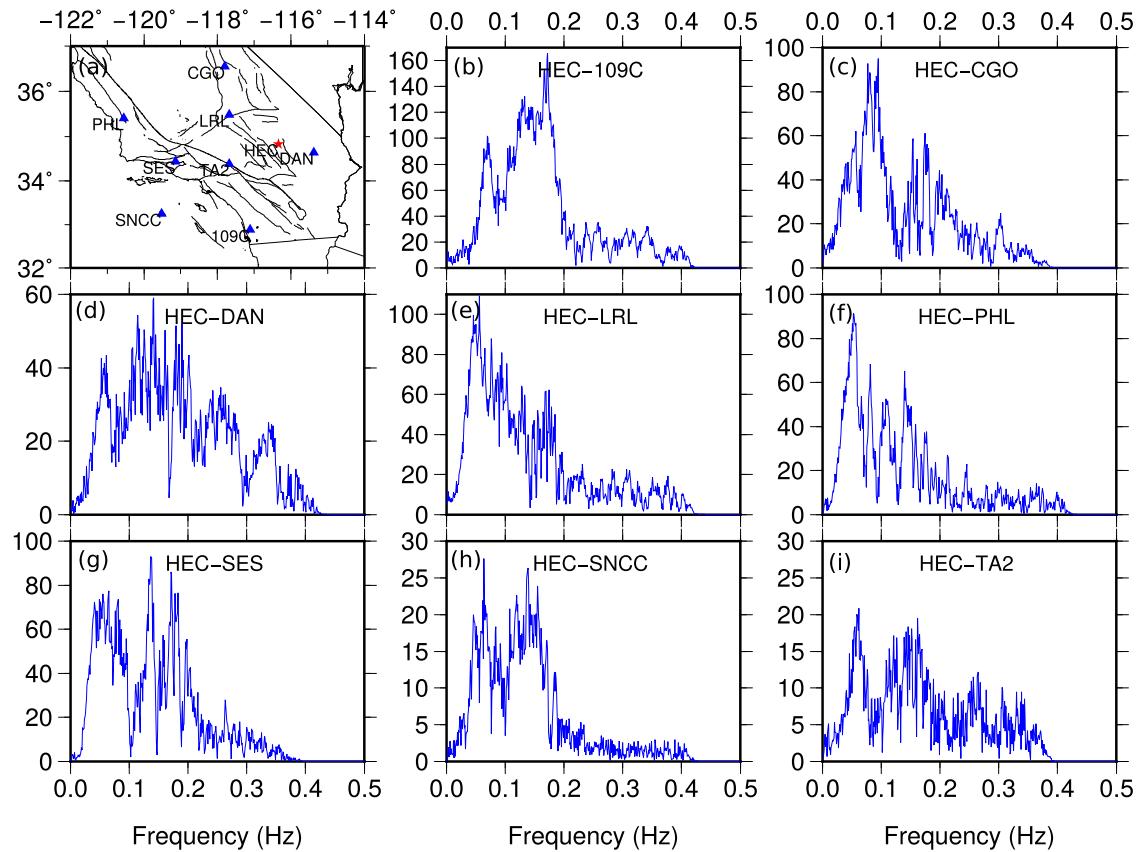
In this section, we outline the adjoint tomography procedures used to construct a new Vs model of southern California by minimizing the phase traveltime differences between the EGFs of station pairs and the synthetic Green’s functions (SGFs) computed based on numerical simulations. The amplitude differences between EGFs and SGFs are disregarded as the true amplitudes of the EGFs are lost during the data processing. As illustrated in Section 2, the EGFs are derived from negative time derivatives of cross-correlations of ambient noise using the empirical relation eq. (1). We simulate the set of corresponding SGFs, measure the phase traveltime misfits between EGFs and SGFs, and then successively update our model based on misfit gradient. All the forward and adjoint simulations in this study are performed using the open-source spectral-element package SPECFEM3D\_Cartesian (<https://geodynamics.org/cig/software/specfem3d/>).

#### 4.1 SEM mesh

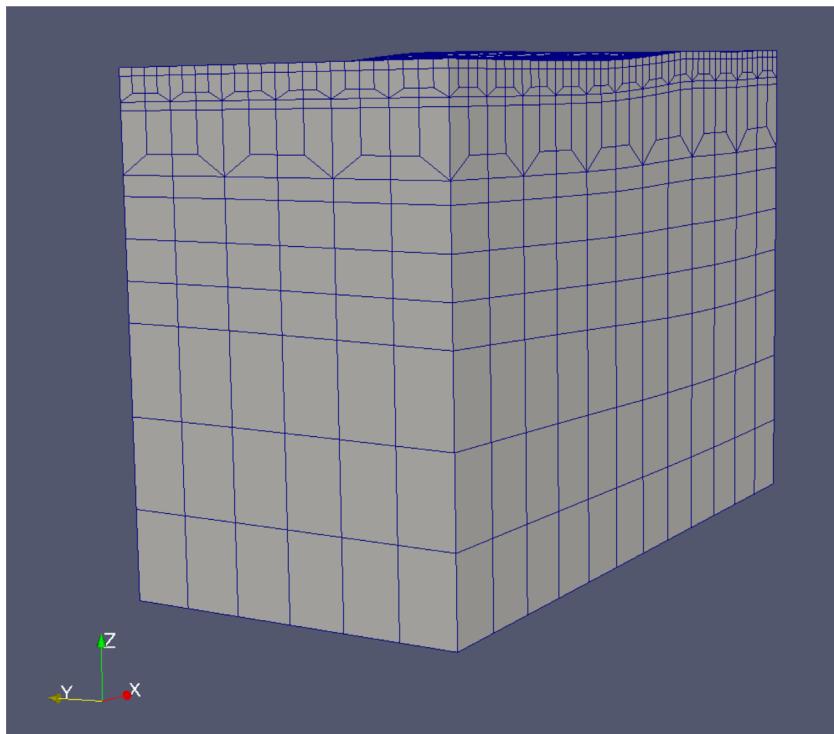
The mesh in this study is almost identical to that used in Tape *et al.* (2009) except that we use the new meshing tool in the latest version (v3.0) of SPECFEM3D\_Cartesian. It has 336 elements in longitude ( $\sim 639$  km), 288 elements in latitude ( $\sim 503$  km), and 11 layers in depth (60 km). Two doubling layers are used in the mesh in order to provide the smallest elements near the low-velocity regions at the surface. Fig. 3 shows a section of the mesh. The mesh yields a horizontal grid spacing of about 2 km at the surface and an increasing vertical spacing from  $\sim 1$  km at the top to  $\sim 10$  km at the bottom, which gives sufficient simulation accuracy at 5 s and longer periods. As our new mesh is slightly different from that of Tape (2009), the density, compression and shear wave velocities of the new mesh are assigned from the nearest neighbouring points in **M16** as the initial model.

#### 4.2 Forward simulation

In the forward simulation for Rayleigh-wave-type SGFs, a single vertical point force  $\mathbf{f} = (0, 0, f\hat{\mathbf{z}})$  is placed at the surface of a master



**Figure 2.** Examples of spectra of EGFs between master station HEC and a number of other stations shown in (b)–(i). The location of the master station HEC is shown as the red star, and other stations indicated by the blue triangles in (a).



**Figure 3.** A chunk of mesh used in this study. The 11 layers from the bottom to the surface are divided to four subregions: from the bottom of the model to the Moho (3 layers), the Moho to the basin basement (5 layers), the basement to the sediment (2 layers) and the sediment to the surface (1 layer). The depth of the sediment is set to one-fifth of that of the basement. Mesh doubling occurs right beneath the sediment (10th layer) and basement (8th layer).

station while all other stations are treated as receivers. A Gaussian function is used as the source time function of the point force:

$$g(t) = \frac{1}{\sqrt{\pi} \tau} e^{-\left(\frac{t}{\tau}\right)^2}, \quad (9)$$

where  $\tau$  is the half-duration of the source which we set to be 1.0 s since bandpass filters are applied in the processing later. The wavefield generated by this point-force vector source is recorded at all the other stations, and the vertical-component seismograms can be regarded as SGFs between the master station and each individual receiver. With the SEM mesh discussed in Section 4.1, we choose a time step of 0.01 s, smaller than the suggested maximum time step of 0.012 s based on Courant stability condition (Komatitsch & Vilotte 1998). The SEM calculations are performed in parallel and distributed to  $14 \times 12 = 168$  CPU cores (Intel Xeon E5540 at 2.53GHz). Each forward simulation typically takes about 25 min to generate seismograms of 240 s duration.

### 4.3 Preprocessing

A number of preprocessing steps are performed to construct the needed adjoint sources based on frequency-dependent traveltimes measurements between observed (EGFs) and synthetic (SGFs) seismograms.

First, the EGFs and SGFs are cut from 20 s before to 240 s after the initiation time, then filtered to multiple period bands. A 20-second window is added before the initiation time to better measure near-field Rayleigh waves. EGFs are bandpass filtered at different frequency bands and their amplitudes are normalized to the maximum absolute value of corresponding SGFs. The selection of multiple frequency bands is a multi-scale strategy employed to reduce the non-linearity in the inversion (Bunks *et al.* 1995; Akcelik *et al.* 2002). To this end, typically two different workflows can be used: 1) gradually introduce short-period data after the model has improved based on long-period data, as in some continental-scale inversions (Zhu *et al.* 2015); 2) alternatively, filter waveforms through multiple overlapped period bands, as commonly used in finite-frequency body wave tomography (Hung *et al.* 2004) and other regional-scale inversions (Chen *et al.* 2014). Here, we adopt the first option by first inverting long-period 20–50 s waveforms and gradually add 10–20 and 5–10 s data.

Second, a window selection procedure is performed to isolate Rayleigh waves from the filtered EGF and SGF pairs. Chen *et al.* (2014) used the automatic windowing algorithm FLEXWIN (Maggi *et al.* 2009) to isolate impulsive energy packets in both data and synthetics for misfit measurements. However, because EGFs from ambient noise are dominated by surface waves, it is sufficient to select windows based on the estimated arrivals of surface waves. For each pair of EGF and SGF, we define a time window around the predicted Rayleigh-wave group-velocity arrival time as  $\left[\frac{D}{U_{\max}} - \frac{T}{2}, \frac{D}{U_{\min}} + \frac{T}{2}\right]$ , where  $D$  represents the inter-station distance,  $T$  denotes the maximum period of data used in the current inversion,  $U_{\min}$  and  $U_{\max}$  are the typical minimum and maximum group velocities at these period bands. We choose  $U_{\min}$  and  $U_{\max}$  from a separate dispersion analysis, allowing the entire surface-wave package to be included in the misfit measurements.

### 4.4 Measurement

We measure the frequency-dependent phase traveltimes misfit over a single time window for a particular event  $e$  based on the multi-taper

technique as (Laske & Masters 1996; Zhou *et al.* 2004; Tape *et al.* 2010)

$$\phi_i^e(\mathbf{m}) = \int_{-\infty}^{+\infty} \frac{h_i(\omega)}{H_i} \left[ \frac{\Delta T_i(\omega, \mathbf{m})}{\sigma_i} \right]^2 d\omega, \quad (10)$$

where  $\mathbf{m}$  is the model vector,  $\Delta T_i(\omega, \mathbf{m}) = T_i^{\text{obs}}(\omega) - T_i(\omega, \mathbf{m})$  represents the frequency-dependent traveltime difference between the SGF and EGF over this time window  $i$  of a particular station with its uncertainty  $\sigma_i$ , and  $h_i(\omega)$  is a frequency-domain window to which we apply the normalization constant  $H_i = \int_{-\infty}^{+\infty} h_i(\omega) d\omega$ . The detailed expressions of the misfit and corresponding adjoint source can be found in appendix C of Tape (2009).

A set of quality control parameters are applied to only choose those windows with good fits between data and synthetics. This includes minimum/maximum values of time-shift ( $\Delta T$ ), amplitude difference ( $\Delta \ln A$ ), and minimum cross-correlation coefficient ( $CC_{\min}$ ). In Table 1, we list the values of these parameters used at each iteration. As we normalize the EGFs to the corresponding SGFs,  $\Delta \ln A$  is set to  $[-1, 1]$ . We use tighter quality control parameters as the inverted model improves iteratively, including smaller  $\Delta T$  for shorter period bands and larger  $CC_{\min}$ .

The total misfit for all measurement windows is then expressed as the average misfit for all individual windows in eq. (10):

$$\phi(\mathbf{m}) = \frac{1}{E} \frac{1}{N_e} \sum_{e=1}^E \sum_{i=1}^{N_e} \phi_i^e(\mathbf{m}), \quad (11)$$

where  $E$  and  $N_e$  are the number of events and measurements for  $e$ th event. If multiple period bands are used, we simply combine misfits or adjoint-sources of different period bands with the same weights.

### 4.5 Adjoint simulation

For a master station, after adjoint sources at all receivers are computed, we perform the adjoint simulation by putting time-reversed adjoint sources at the receivers simultaneously to generate the adjoint wavefield. Within the same simulation, we recover the forward wavefield from the last time step of displacement and velocity wavefield saved in the forward simulation. The multiplication of adjoint wavefield and backward reconstructed forward wavefield generate the event kernel for this master station.

These simulations are performed on the same 168 processors and it takes about 1.5 h to complete both the forward and adjoint simulation for a single master station. In our study, we intend to recover elastic structures based on sole traveltimes measurements. If attenuation is included in the seismic wave simulation, back reconstruction of forward wavefield becomes unstable and a checkpointing scheme, hence extra computer storage and I/O time, would be required to periodically reset the reconstructed wavefield (Komatitsch *et al.* 2016). We ignore the effect of attenuation, because we do not think its minor influence on our target periods ( $\geq 5$  s) to be worth the computational cost.

### 4.6 Pre-conditioner and kernel smoothing

As discussed in Section 3, adjoint tomography only provides the first-order derivative of total misfit function (i.e. gradient vector), not its second-order derivative (i.e. the Hessian). Nevertheless, a variety of pre-conditioners can be used to approximate the Hessian and applied to the gradient vector to improve convergence.

**Table 1.** Values of inversion parameters used over iterations.  $T$  is the period bands for filtered data used in each iteration;  $\Delta T$ ,  $\Delta \ln A$ ,  $CC_{\min}$  are time shift, amplitude difference and cross-correlation coefficient criteria for window selection; the two  $\sigma_{(h, v)}$  are the horizontal and vertical radii of the Gaussian function used to smooth the misfit gradient;  $\alpha$  represents the optimal step length chosen at each iteration.

Summary of misfit measurements and gradient smoothness parameters

	1st iteration	2nd–3rd iteration	4th–5th iteration		
$T(s)$	[20, 50]	[10, 20]	[20, 50]	[5, 10]	[10, 20]
$\Delta T(s)$	[−4.5, 4.5]	[−3.5, 3.5]	[−4.5, 4.5]	[−2.5, 2.5]	[−3.5, 3.5]
$\Delta \ln A(s)$	[−1.0, 1.0]	[−1.0, 1.0]	[−1.0, 1.0]	[−1.0, 1.0]	[−1.0, 1.0]
$CC_{\min}$	0.69	0.75	0.75	0.80	0.80
$\sigma_{(h, v)}(\text{km})$	(20.0, 10.0)	(15.0, 7.0)	(15.0, 7.0)	(10.0, 5.0)	(10.0, 5.0)
Optimal step length					
$\alpha$	1st iteration 0.12	2nd iteration 0.05	3rd iteration 0.03	4th iteration 0.02	5th iteration 0.02

We use the widely adopted pre-conditioner based on the second term of the diagonal Hessian (Zhu et al. 2015; Modrak & Tromp 2016), which is expressed as

$$P(x) = \int \partial_t^2 \mathbf{u}^\dagger(\mathbf{x}, T - t; \mathbf{x}_s) \cdot \partial_t^2 \mathbf{u}(\mathbf{x}, t; \mathbf{x}_s) dt \quad (12)$$

In SPECFEM3D\_Cartesian, this preconditioner is calculated by the vector dot product of the adjoint accelerations and forward acceleration.

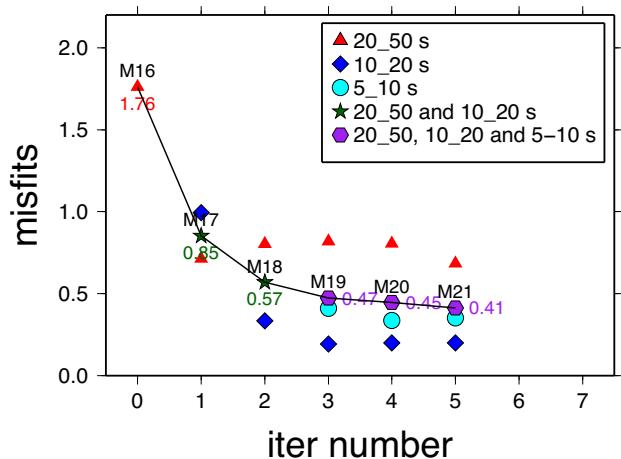
In this study, we use a 3-D Gaussian function to smooth the pre-conditioned misfit gradient. Generally, the smoothing radius should be at least one element size (to eliminate sensitivity singularity at sources and receivers) and close to the wavelength of resolvable structures by the current period band. See Table 1 ( $\sigma_{h, v}$ ) for the choices made in this study.

#### 4.7 Updating the model and reducing the seismogram misfit

We use a line search method in conjunction with a gradient-based optimization algorithm to decide the optimal step length of model updates in each iteration (see the Appendix). Then,  $V_p$  and  $V_s$  structures are updated based on their own misfit gradient. Surface waves from ambient noise data are not very sensitive to density variations, as indicated by the relatively small amplitudes of the misfit kernel for density. Therefore we update the density model based by a scaled version of the  $V_s$  perturbation,  $\delta\rho = 0.33 * \delta V_s$  (Montagner & Anderson 1989). In Fig. 4, we plot the total misfit reduction for the individual period band used in the various stages of the inversion. After the first iteration, we introduce measurements made at the 10–20 s band. It produces a slight increase in the misfit for the 20–50 s band which decreases in the last two iterations. Furthermore, the value of the averaged total misfit for the two bands continues to decrease during the second and third iterations. After introducing the third band of 5–10 s, we see that the overall misfit continues to decrease, albeit slower than previous iterations. After the fifth iteration, the averaged total misfit of these three bands over one iteration changes little and the model update relative to the previous iteration is less than 2.0 per cent, at which point we terminate our inversions.

## 5 RESULTS

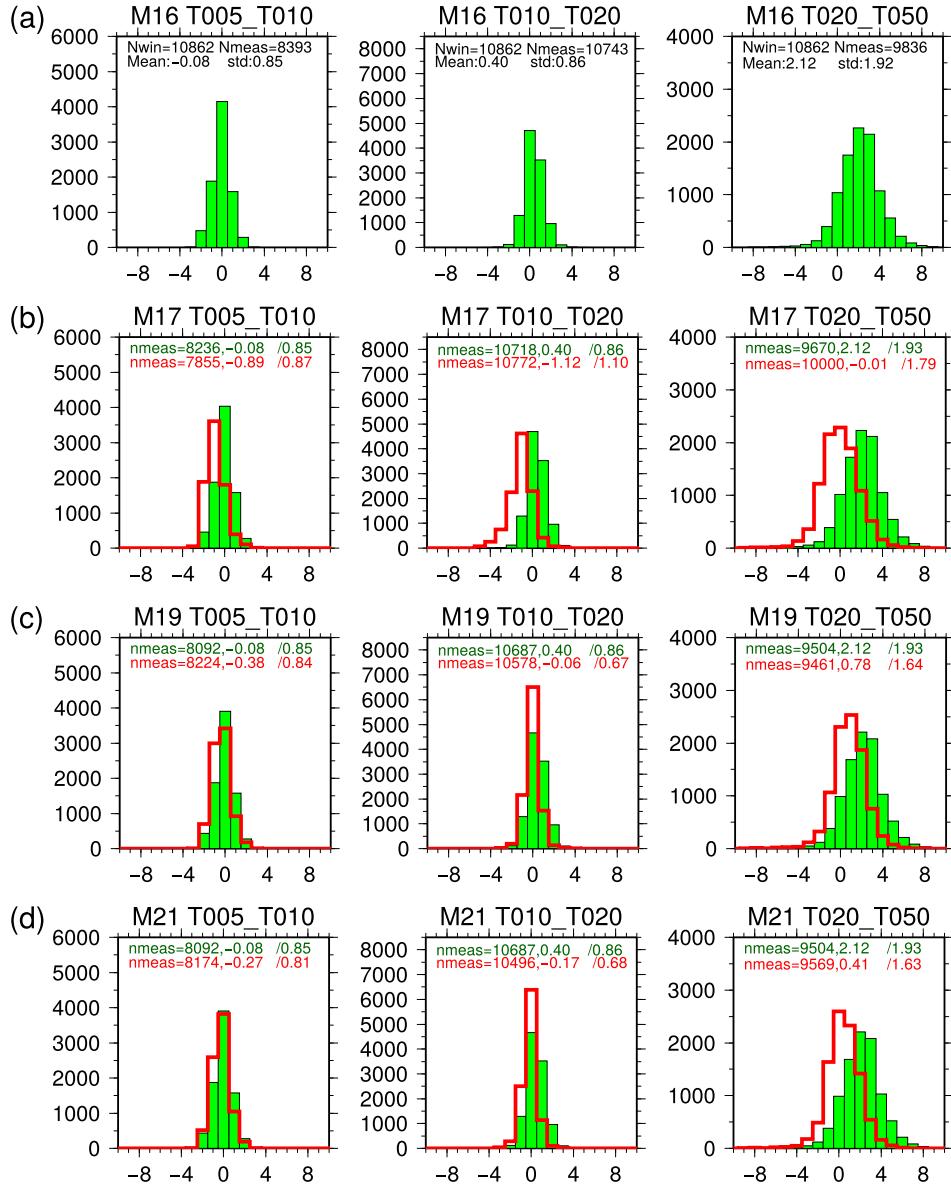
Here, we present and discuss the final shear wave speed model of southern California obtained after three stages of adjoint tomographic inversion from M16 to M17 (stage 1), M17 to M19 (stage 2) and M19 to M21 (stage 3).



**Figure 4.** Total misfit reduction over iterations. Different symbols with different colours, as shown in the legend box, are used to represent the different frequency bands used in measuring the total misfit. The solid curve connects the final total misfit reduction over the three inversion stages using different frequency bands: the first (20–50 s), the second (20–50 and 10–20 s) and the third (20–50, 10–20 and 5–10 s).

#### 5.1 Initial model

We start our inversion using the model M16 obtained by Tape et al. (2010) as described in Section 1. Following the procedures described in Section 4, we first perform a set of forward simulations using each individual station as the master station. Considering the energy distribution in the spectra of EGFs (Fig. 2), we filter waveforms at three different period bands, that is, 5–10, 10–20 and 20–50 s, and measure frequency-dependent traveltime misfits of SGF-EGF pairs. In Fig. 5(a), we present histograms of time-shift measurements made at these three bands in selected windows based on criteria outlined in Section 4.3. For M16, in the 5–10 s band (Fig. 5a, left), the histogram is symmetric around zero and most time-shifts are within the range of [−2, +2] s, indicating that M16, built solely on earthquake data, also produces good fits between SGFs and our EGFs at this period band. However, there is still room for improvements, for example, the number of windows that pass quality control criteria in misfit measurements (i.e. *good measurements*, thereafter) is small (8393 out of 10 862). Similar characteristics are also observed in the 10–20 s band (Fig. 5a, middle), with a slightly wider distribution of time-shift values in [−2, +3] s, a larger number of good measurements (10 743 out of 10 862), and a small skew towards positive time-shift. In the 20–50 s



**Figure 5.** Histograms of traveltime misfits obtained for models **M16** (a), **M17** (b), **M19** (c) and **M21** (d). The green solid bars represent misfits for the initial model **M16**, and the red bars in (b)–(d) are misfits for the updated models (**M17**, **M19** and **M21**). The total number of good measurements, average misfit and standard deviation are shown at the top of each histogram with the same colour specified above. Misfits are measured at all three period bands, 5–10 (left), 10–20 (middle) and 20–50 (right) s.  $CC_{min}$  in (a) are chosen as 0.80 for 5–10 s, 0.75 for 10–20 s, 0.69 for 20–50 s to show the misfits for the initial model. All the misfits are measured with  $CC_{min} = 0.75$  in (b) and  $CC_{min} = 0.80$  in (c) and (d) for comparison. Note that the measurements of 5–10 and 10–20 s bands in (a) are not used in the stage 1 inversion, and those of 5–10 s in (2) are not used in the stage 2 inversion.

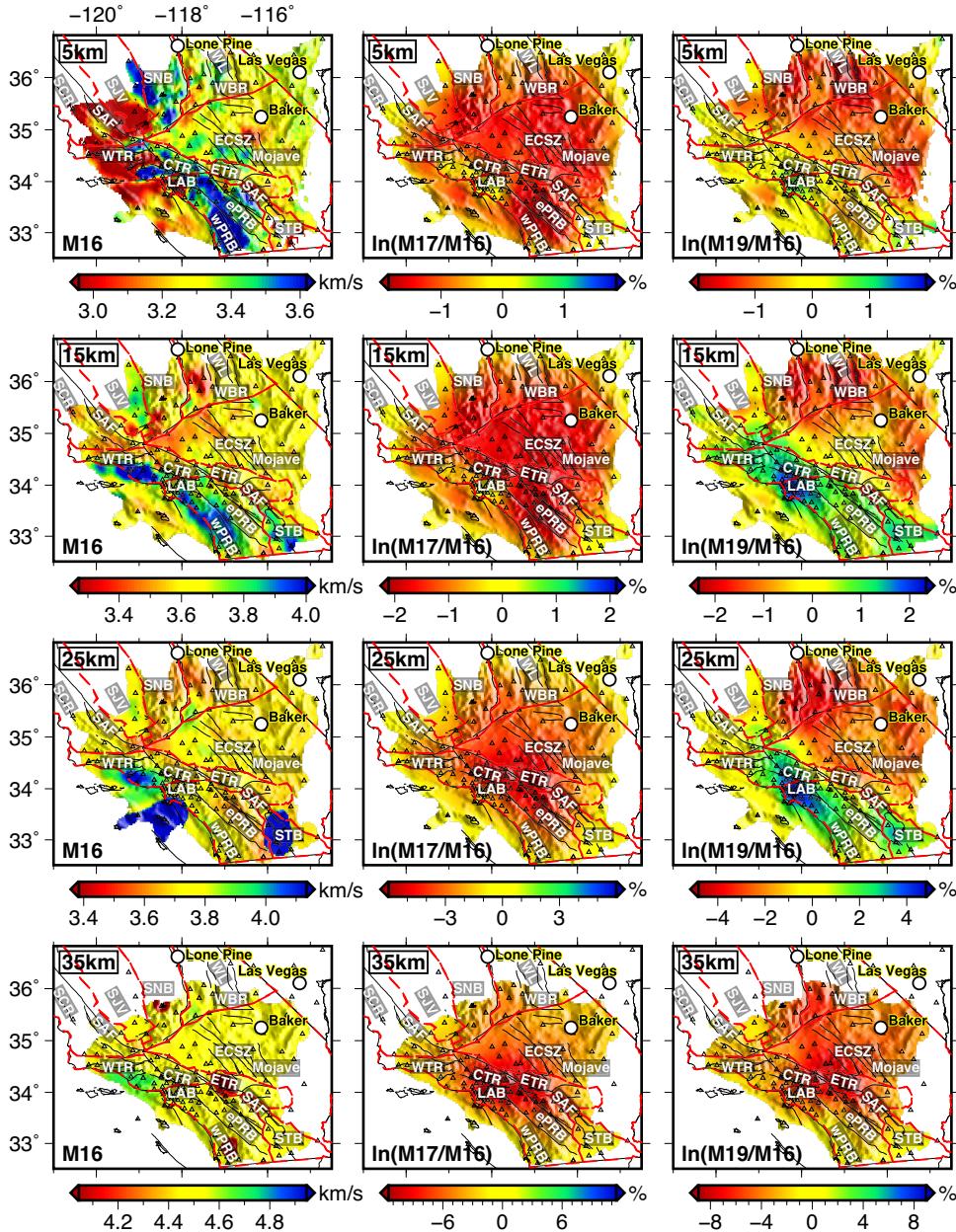
band (Fig. 5a, right), the distribution is clearly asymmetric and biased towards positive time-shift, which indicates that parts of the model may be significantly slower than **M16**. It is expected that our inversion will focus first on reducing the misfit in this 20–50 s band.

## 5.2 Updated models

We start our inversion with measurements only from the long-period band of 20–50 s, and progressively add measurements from the 10–20 and 5–10 s bands as the misfit reduction over iterations from existing band(s) becomes less evident. Alternative multi-scale strategy using broad bands (e.g. 5–50, 10–50 and 20–50 s) is also possible. We demonstrate that, in our case, this alternative strategy does not

alter the main features of the adjoint inversion (see Supporting Information Figs S1–S4).

In Fig. 6, we display the absolute shear wave speed of **M16** (left column) and cumulative variations in percentage with respect to **M16** for model **M17** (central column) obtained after the first stage inversion and model **M19** (right column) obtained after the second stage inversion. The models are masked based on amplitudes of volumetric coverage kernels shown in the Supporting Information (Fig. S6). After the first iteration using solely 20–50 s band measurements, we obtain model **M17**. As expected from the positive bias in the traveltime misfit distribution (Fig. 5a, right),  $V_s$  is broadly slowed down over the whole crust and uppermost mantle with local variations as much as –6 per cent in the lowest crust and –12 per cent in the uppermost mantle (35 km). The total misfit



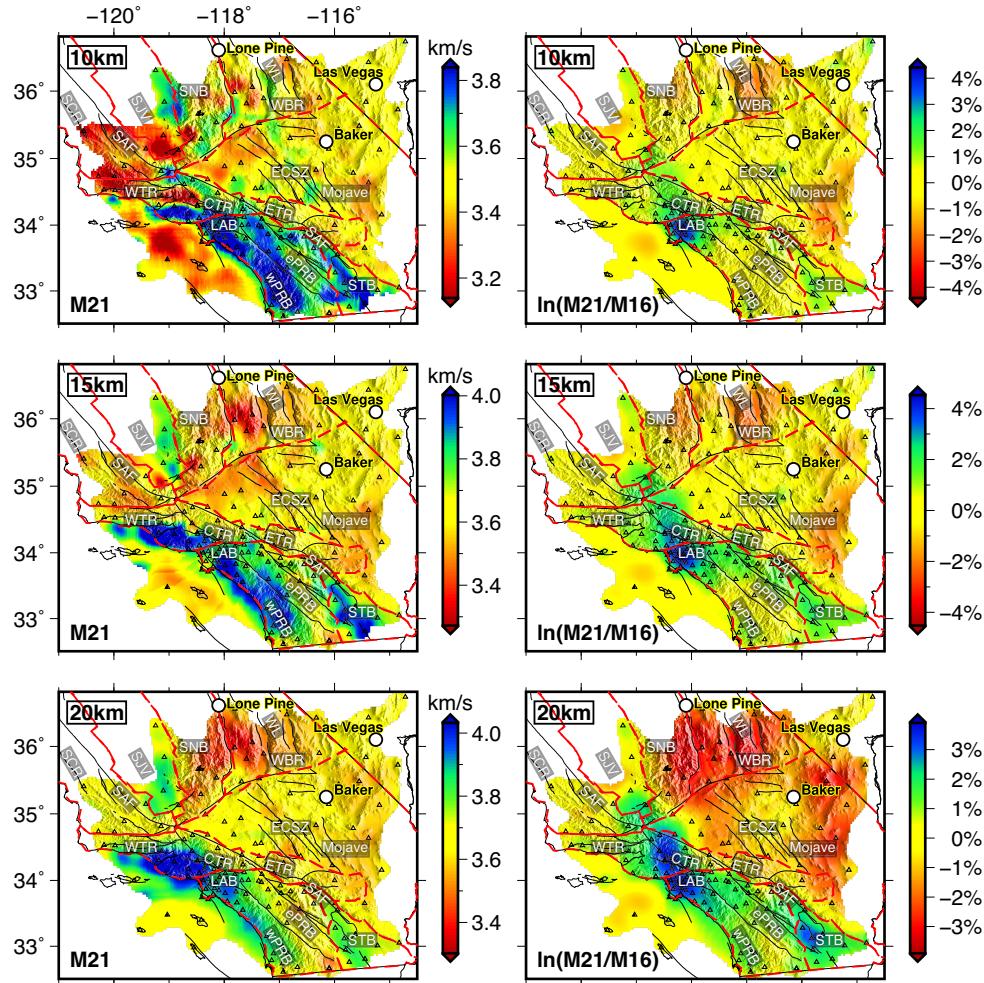
**Figure 6.** Horizontal cross-sections of shear velocity of model **M16** (left panel), **In(M17/M16)** (middle panel) and **In(M19/M16)** (right panel) at 5, 15, 25 and 35 km depths. In the vicinity of the Moho, the cross-sections reveal patches that represent wave speeds from the underlying mantle (or overlying crust).

for this band also drops significantly (Fig. 4). In the next iteration, misfits seem to change very little in the line search (Fig. A1b). It suggests that we already fit most measurements at this band in one iteration, prompting the inclusion of 10–20 s band data. We then perform two more iterations using joint 20–50 and 10–20 s band data, and obtain **M19**. With the additional data, **M19** shows an increase of  $V_s$  in the Central Transverse Range (CTR), Los Angeles Basin (LAB, with maximum +4 per cent perturbation), Peninsular Ranges Batholith (PRB), and Salton Trough Basin (STB) while further slowing down  $V_s$  in parts of the Mojave Desert, western edge of Basin and Ranges (WBR), and southern Sierra Nevada Batholith (SNB), particularly at the lower crustal depth. The emergence of these features coincides with the fact that traveltimemisfit distribution for the 10–20 s period band becomes fairly symmetric from Fig. 5(a) to Fig. 5(c).

### 5.3 Final model

We then perform another two iterations by including the measurements from the band of 5–10 s to obtain the final model **M21**. As stated in Section 2, EGFs in the secondary microseism band of 5–10 s have strong energy that can constrain the shallowest part of the model. As shown in Fig. 7, the inclusion of the 5–10 s measurements helps refine small-scale  $V_s$  structures in the shallow crust (10–15 km), with perturbations ranging between [−4.5, +4.5] per cent from the initial model.

Next, we compare our final model **M21** with the initial model **M16** from Tape *et al.* (2010) by examining a series of horizontal (Figs 7 and 8) and vertical cross-sections (Figs 9–11), and discuss their differences introduced by the EGFs from ambient noise data.



**Figure 7.** Horizontal cross-sections of shear velocity of model **M21** (left panel) and **In(M21/M16)** (right panel) at 10, 15 and 20 km depths.

### 5.3.1 Horizontal cross-sections

In the uppermost crust ( $\leq 5$  km), we do not see significant improvements. This is not unexpected, considering the low number of measurements at short-period band (Fig. 5d, 5–10 s) in comparison to the large number of shorter-period band (2–30 s) measurements used to generate the initial model **M16** (Tape *et al.* 2010).

In the middle crust (10 – 15 km), the model update **In(M21/M16)** shown in Fig. 7 reveals three distinct regions with shear velocity reductions, including the southern SNB, the Walker Lane (WL) region, and the eastern Mojave Desert. These areas are not well resolved in the previous earthquake-data based adjoint inversion (Tape *et al.* 2010) due to the limited distribution of earthquakes. Small increases in shear velocity are observed for the LAB, CTR and STB area.

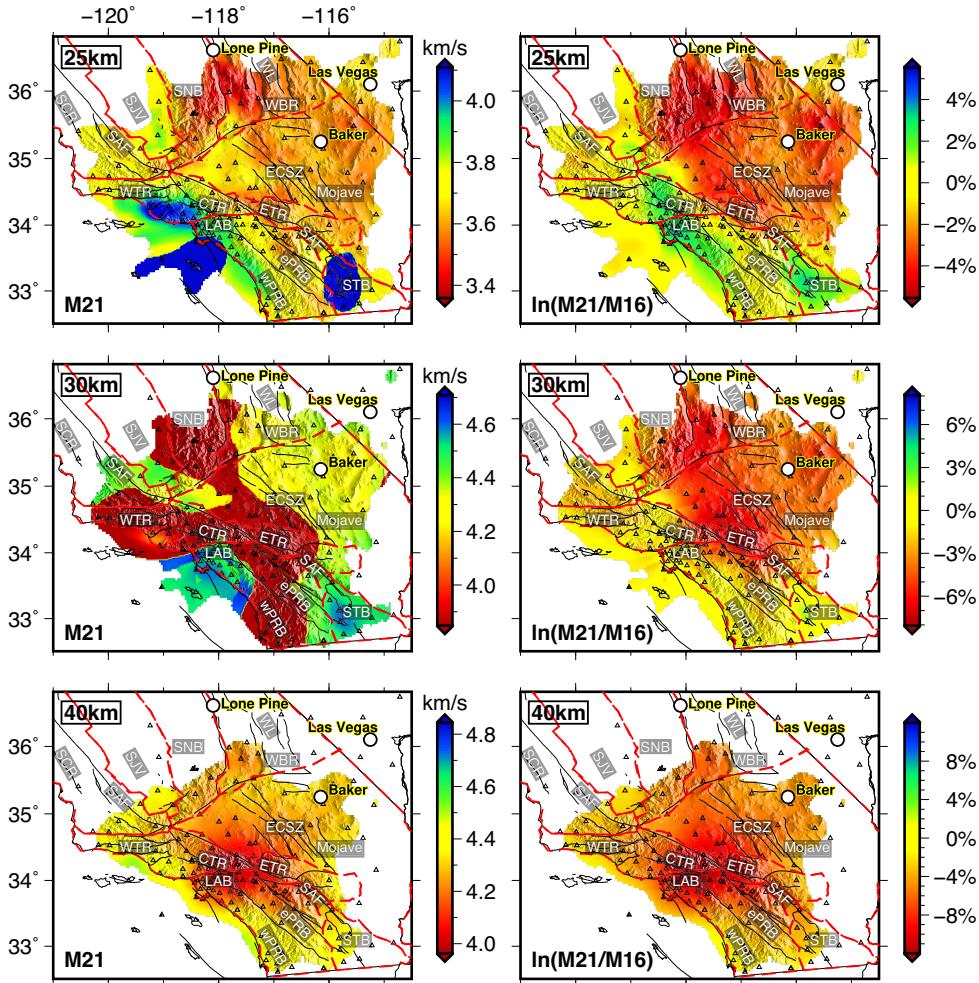
In the lower crust (20–30 km), high-velocity anomalies persist with changes up to +3 per cent in the CTR, STB, and the maximum reaching +4 per cent in the LAB. We also observe higher Vs beneath the westernmost PRB. Another distinct feature is that the lower crust in the SNB, WL, the eastern Mojave and WBR is significantly slowed down by 6 per cent. The whole uppermost mantle is slowed down by up to 12 per cent (Fig. 8).

### 5.3.2 Vertical cross-sections

We display several vertical profiles through SAF, PRB and STB (see Fig. 1 b for locations). These profiles are plotted in a way similar to those shown in Barak *et al.* (2015) as is further discussed in Section 6 (with some changes in width) for the purpose of comparison.

As shown in Fig. 9, both the initial model **M16** and our final model **M21** show shear velocity contrasts ( $\sim 3.8$  in the west and  $\sim 3.5 \text{ km s}^{-1}$  in the east) across SAF in the crust. At the southern part of SAF (profiles AA' to CC') through ePRB, ETR and STB, **M16** only shows this lateral velocity contrast across SAF down to the middle crust, while, in our model, it extends to the lower crust. Both models show that the high velocities in the west dip gradually to the northeast. For the central part of SAF (profiles DD' to FF') through ETR, our model exhibits a high velocity layer in the middle crust that dips slightly to the northeast, but this feature is not seen in the initial model. Towards the northwest (profiles GG' and HH') from ETR to SCR, our model reveals enhanced velocity contrasts that are not exactly collocated with the SAF.

In the PRB region (Fig. 10), a west–east velocity contrast is observed, dividing the seismic structure of PRB into western (wPRB) and eastern (ePRB) parts. Shear velocities reach  $\sim 3.9 \text{ km s}^{-1}$  in



**Figure 8.** Same as Fig. 7 but for 25, 30 and 40 km depths.

the middle and lower crust beneath wPRB, while remaining low ( $\leq 3.7 \text{ km s}^{-1}$ ) throughout the crust beneath ePRB. In the lower crust, high velocities are only observed at the westernmost part of wPRB, and they are significantly enhanced in our model relative to the initial model. The high-velocity zone beneath wPRB dips slightly to the northeast in the northeastern part, while the southern part shows a vertical-to-steep dipping angle.

From Fig. 11, it is observed that Vs in the STB region is enhanced in the middle and lower crust. At the southeastern edge of STB, a shallow thin sheet-like high-velocity structure is seen in the middle crust (10–20 km), about 30 km wide and surrounded by low shear velocities. Similar high velocities observed separately at the northwestern part in **M16** appear to be connected to the southeastern part ones in our new model (STaa' in Fig. 11).

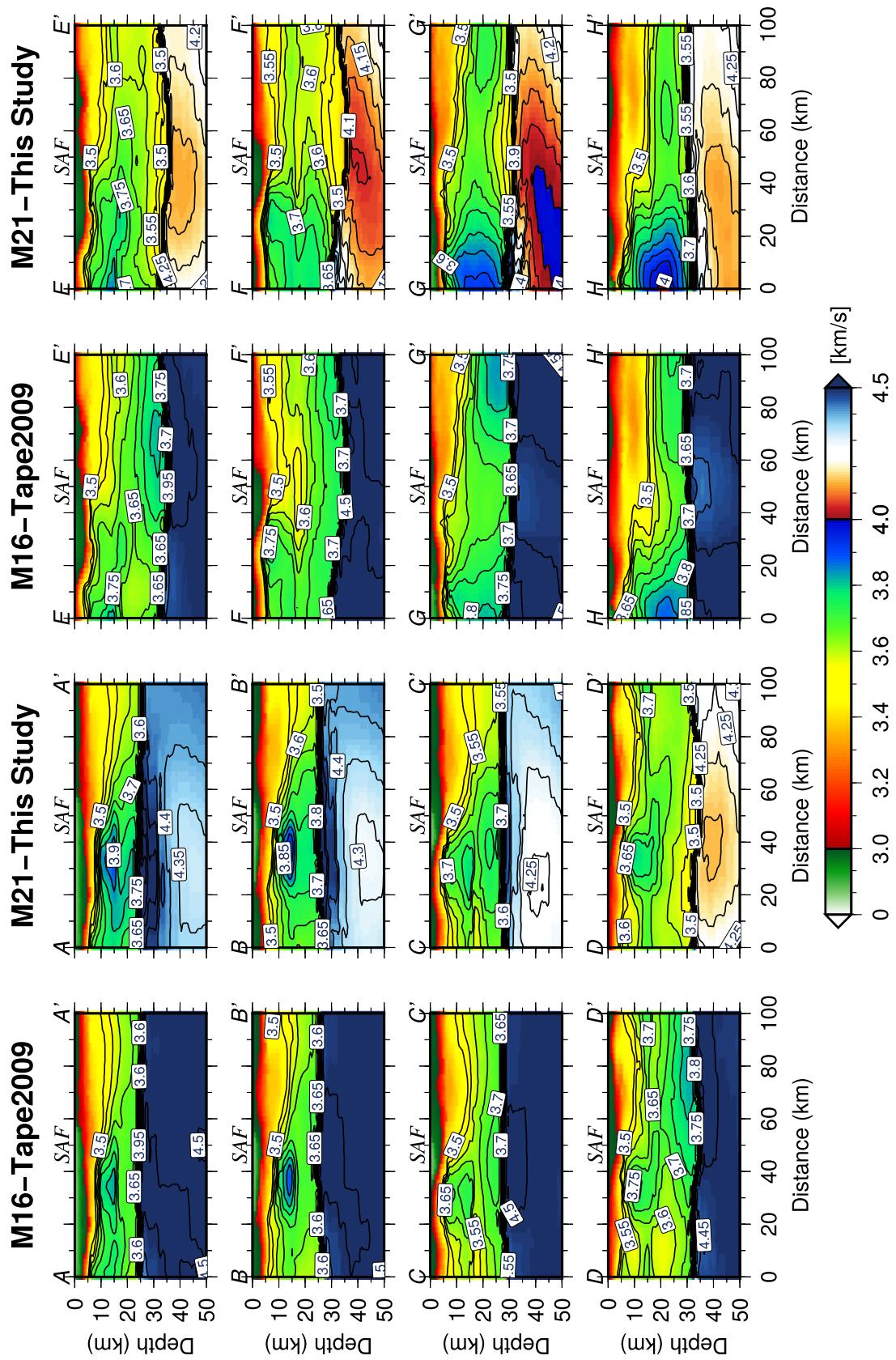
## 6 DISCUSSION

Here, we discuss the robustness of our inversion results through analysis of seismogram misfit, model validation by 143 local earthquakes and comparison of our 3-D model with three other published models. Additional information on waveform fitting, model uncertainties and resolution analysis are also provided in the Supporting Information (Figs S5–S7).

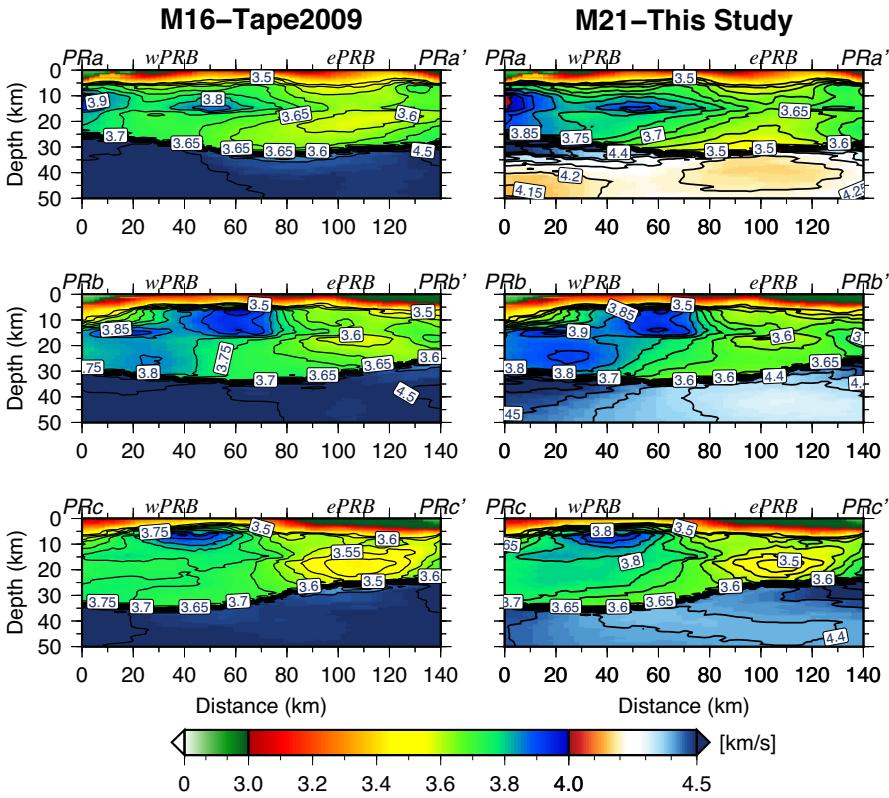
### 6.1 Misfit analysis

We first check the number of good measurements ( $N_{\text{meas}}$ ) obtained from the initial model and updated models to see whether the number is improved by our inversion procedure. As stated in Section 4, misfit measurements are strongly influenced by the  $\Delta T$ ,  $\Delta \ln A$  and  $CC_{\min}$  criteria used in window selections. The  $\Delta T$  and  $\Delta \ln A$  of the three bands are fixed in our inversion, while the  $CC_{\min}$  is improved from 0.69 (**M16**) to 0.75 (**M17** to **M18**), and to 0.80 (**M19** to **M21**). To make reliable comparisons, we additionally measure cross-correlation traveltimes of the three bands for **M16** using  $CC_{\min} = 0.75$  (Fig. 5b) when comparing the number between the stage 1 and 2 inversions and using  $CC_{\min} = 0.80$  (Figs 5c and d) between stage 1 and 3 inversions. For the 20–50 s band, the  $N_{\text{meas}}$  improves from 9670 to 10 000 from stage 1 to 2 (Fig. 5b, right) and 9504 to 9569 from stage 1 to 3 (Fig. 5d, right). The same effects are also observed with the 10–20 s band where  $N_{\text{meas}}$  improves from the initial 10 718 to 10 772 (Fig. 5b, middle). However, it slightly decreases during the last inversion stage. In the third inversion stage for the additional 5–10 s band, the  $N_{\text{meas}}$  improves from the initial 8092 to final 8174 (Fig. 5d, left).

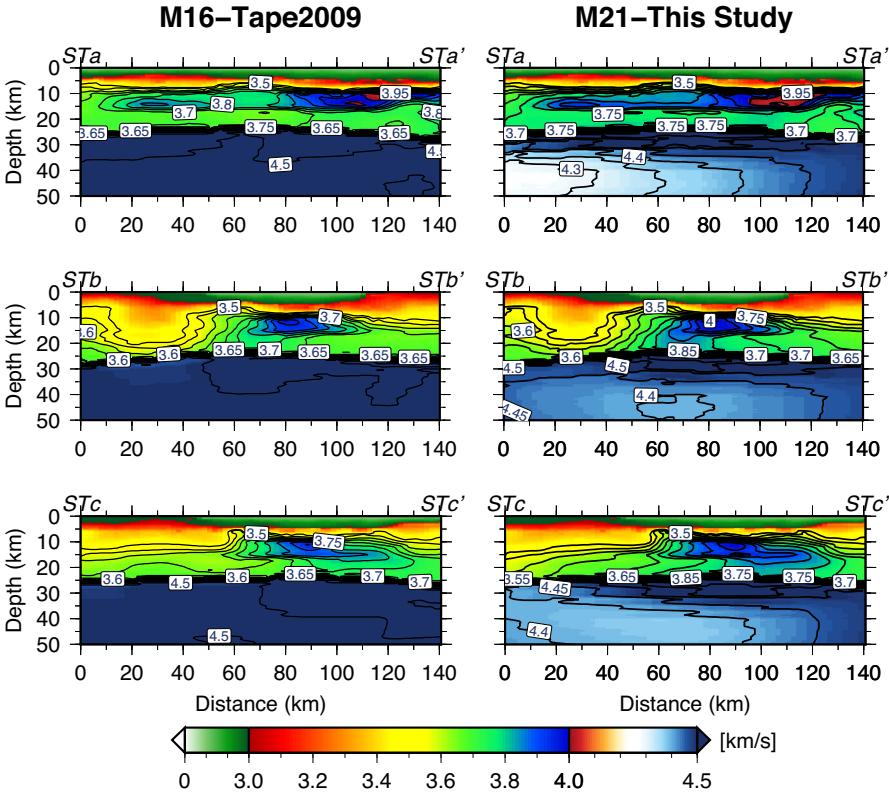
We investigate the changes of traveltimes misfits for the three updated models (**M17**, **M19**, **M21**) relative to those misfits for the initial model **M16**. The most significant improvements are obtained in the long-period band of 20–50 s for which traveltimes misfits are



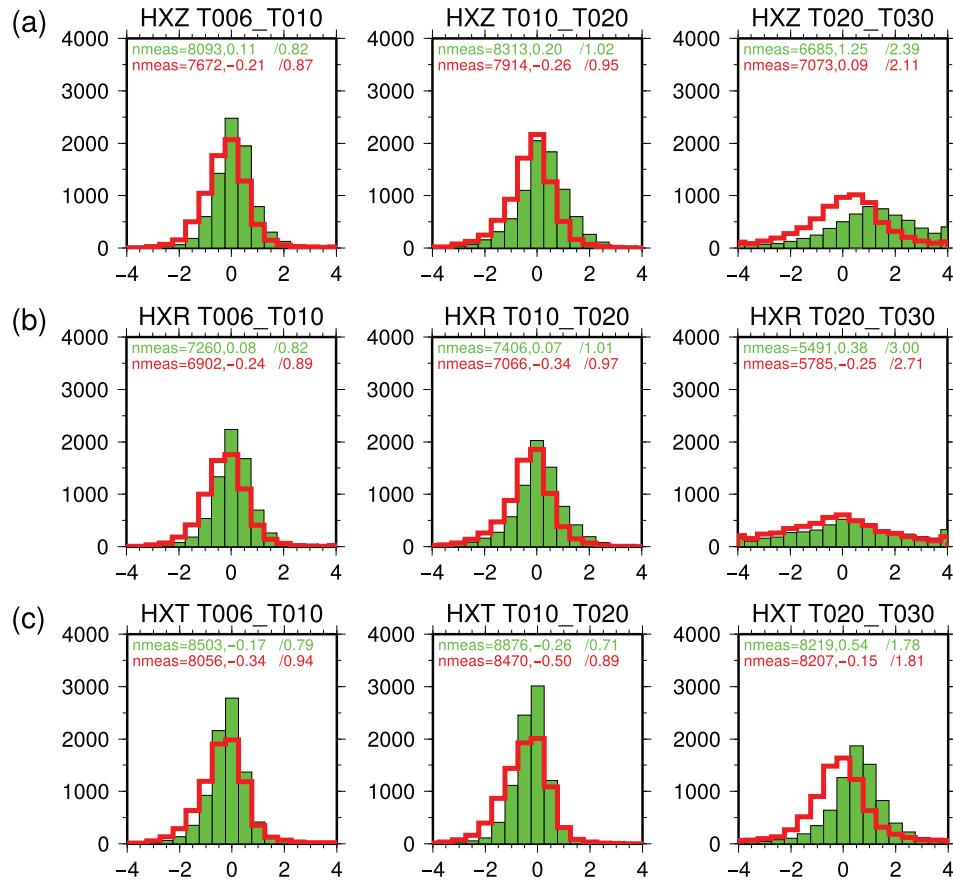
**Figure 9.** Cross-sections of shear velocity of model M16 and M21 across the San Andreas Faults with map locations of these profiles shown in Fig. 1(b).



**Figure 10.** Cross-sections of shear velocity of model M16 (left panel) and M21 (right panel) across the Peninsular Range Batholith (PRaa'-PRcc') with the locations of these profiles shown in Fig. 1(b).



**Figure 11.** Cross-sections of shear velocity of M16 (left panel) and M21 across the Salton Through (STaa'-STcc') as the profiles shown in Fig. 1(b).



**Figure 12.** Comparison of traveltime misfits between **M16** (green) and **M21** (red) for the 143 local earthquakes, measured at period bands of 6–10, 10–20 and 20–30 s on the vertical (a), radial (b) and transverse (c) components.

significantly reduced with the mean and standard deviation (STD) values changing from the initial  $2.12 \pm 1.93$  s to  $0.41 \pm 1.63$  s (Fig. 5d, right). For the 10–20 s band, the traveltime misfit reductions are relatively small, with the mean and STD values reduced from  $0.40 \pm 0.86$  s to  $-0.17 \pm 0.68$  s (Fig. 5d, middle). The histogram for the 5–10 s band shows almost no change in the time-shift distribution. Furthermore, we note that the number of good measurements with  $|\Delta T| \leq 1.0$  s increases in all three bands.

## 6.2 Model validation

To validate our new model, we perform forward simulations for the 143 selected local earthquakes used in Tape (2009) using our final **M21** model. We measure the cross-correlation traveltime differences between observed seismograms and synthetics filtered at 6–10, 10–20 and 20–30 s. Misfit measurements are made on the filtered three-component seismograms of both surface and body waves, and the corresponding misfit histograms are shown in Fig. 12. At the long-period band of 20–30 s, misfit distribution for both the vertical and horizontal components displays a positive bias in **M16**, while this distribution is nearly zero-centred in our new model **M21**. At the other two period bands, our new model does not lead to much improvements in the misfit, which is reasonable because the initial model already fits earthquake data at short period band very well. In addition, our model is actually a Vsv model constructed by using only Rayleigh waves without considering radial anisotropy. In the future, joint inversion of three-component earthquake and ambient

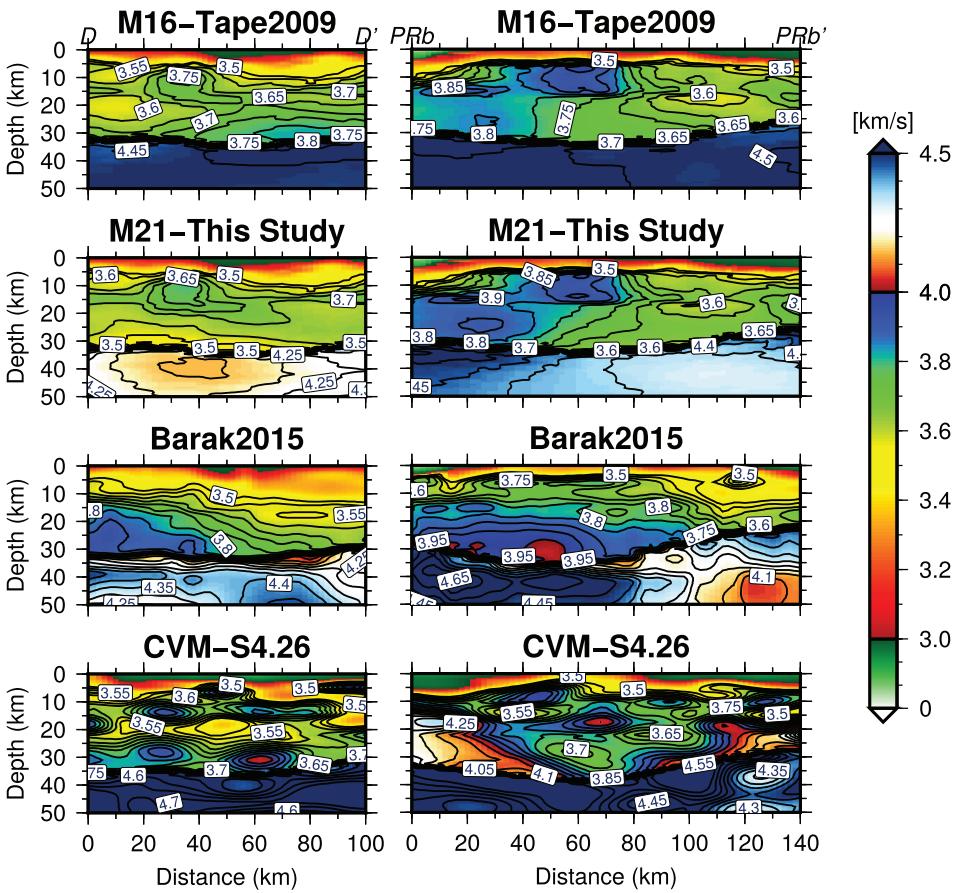
noise data are required to obtain an anisotropic model in southern California.

## 6.3 Model comparison

We now compare the initial model **M16** and our final model **M21** with two other models. The first is CVM-S4.26 from Lee *et al.* (2014), which used both earthquake data and ambient noise at 5–50 s, finite-difference forward modelling, and a combination of scattering-integral and adjoint methods for the tomographic inversion. The second model is Barak2015, constructed by traditional ANT (Barak *et al.* 2015). It was built by first obtaining group velocity maps at 4–40 s periods from 2-D surface wave tomography, and then averaging the median result of 13 inversions with different starting models that are partly based on **M16**. We compare our final model with these models by examining a series of vertical cross-sections in the Supporting Information (Figs S8–S25).

In general, the CVM-S4.26 model contains more small-scale features than the other three models and has much higher velocities, perhaps due to the larger number of data used and many iterations performed. Based on the similar initial model and data set, our final model shows velocity variations that are generally in agreement with those of Barak2015: (1) lower-crust and upper-mantle wave speeds are significantly lower by global standards; (2) higher shear wave velocities are observed beneath the westernmost PRB and southeastern STB in the lower crust.

Our final model differs from the Barak2015 model in several aspects. In Fig. 13, we choose two vertical cross-sections in which



**Figure 13.** Comparison of model **M16**, **M21**, Barak2015 and CVM-S4.26 along profile DD' across SAF and profile PRbb' across PRB. See more results in the Supporting Information.

distinct differences are observed, demonstrating the effects of using 3-D sensitivity kernels in adjoint tomography on the final tomographic results compared with results based on traditional ANT method. At the central part of SAF (such as profiles DD'), our model exhibits a high velocity zone in the middle crust that dips to the northeast slightly, while the Barak2015 model shows that it dips into the lower crust. Moreover, the high velocity zone in the lower crust of wPRB in our model is not as evident as that shown in Barak2015 model and is only observed in the westernmost part (profile PRbb'). Our model shows that the southern boundary of wPRB and ePRB dips steeply towards the northeast into the lower crust, while it only dips slightly towards the northeast in the middle crust (PRbb') in Barak2015. In addition, our model shows increased velocities beneath LAB and CTR (profiles GG'-HH'), Supporting Information Figs S14–S15), which are not identified in the Barak2015 model.

#### 6.4 Inversion strategies

Differences among these four models (**M16**, **M21**, CVM-S4.26, Barak2015) are expected, given that different data and tomographic inversion techniques were used. Given the major efforts involved in simulation-based inversions (as in the first three models above), here we offer some perspectives for future efforts in seismic tomography.

The five iterations we conduct using ambient noise data in this study can be regarded as an extension to the earthquake-based

adjoint tomography by Tape (2009). Using the same underlying technique as used to build **M16**—spectral-element and adjoint methods—we demonstrate that surface waves from ambient noise data are necessary for resolving lower crustal and uppermost mantle structures.

The CVM-S4.26 model of Lee *et al.* (2014) began its inversion with ambient noise data, then added shorter-period earthquake data after several iterations. This sequence has the advantage that earthquake sources can be introduced into the inversion after the velocity structure has been improved. If all techniques and data are available, we would advocate the following steps sequentially: (1) classical tomography using ambient noise, (2) adjoint-based ANT, (3) earthquake-based adjoint tomography. To our knowledge, this sequence has still not been performed by the same research group at any scale.

Southern California provides the challenge and opportunity for using additional data sets to improve the seismic velocity models. For example, significant improvements to basin models and the Moho model were made following the completion of **M16** (CVM-H 15.1, Shaw *et al.* 2015). These detailed improvements impact earthquake and ambient noise measurements, and therefore further tomographic iterations are warranted.

Looking to the future, the prevalence of adjoint tomography will likely increase with the improvement of scientific workflows (Bozdağ *et al.* 2016; Krischer *et al.* 2015; Modrak & Tromp 2016) and the wider availability of computational resources for these problems.

## 7 CONCLUSIONS

We apply adjoint tomography (Tromp *et al.* 2005) to EGFs constructed from ambient noise data in southern California in order to iteratively improve the 3-D tomographic model **M16** constructed based on earthquake data (Tape *et al.* 2010). EGFs from ambient noise are compared with SGFs from numerical simulation base on SEM. Traveltime misfits between EGFs and SGFs are measured at three period bands: 5–10, 10–20 and 20–50 s. The EGFs from ambient noise show a good agreement with the SGFs from **M16** at 5–10 and 10–20 s, while exhibiting an average 2.12 s time advance at 20–50 s. By minimizing the traveltimes differences between EGFs and SGFs through local-gradient-based optimization algorithm, we obtain a final model **M21** with better data fits compared with the initial model **M16**.

We start our inversion with only data in the 20–50 s band and progressively introduce the other two shorter-period bands. After five iterations, the averaged total misfit combining the three bands is reduced from 1.75 to 0.41. The new Vs model reveals several new features in the middle and lower crust, including: (1) the mean speed of lower crust slowed down by about 6 per cent; (2) higher Vs anomalies (up to +4 per cent) observed in the LAB and CTR throughout the middle and lower crust; (3) higher Vs anomalies seen in the lower crust beneath the westernmost PRB; (4) an enhanced shallow high velocity zone in the middle crust is observed beneath STB. Our model also shows refined lateral velocity gradients across PRB, SNB, SAF, which provides constraints on the west-east compositional boundary of PRB, SNB, as well as the dip angle and the depth extent of SAF.

As ambient noise cross-correlations can be obtained between any station pairs, the new tomographic model provides a better resolution relative to the initial one in areas not well-covered by event-station paths. Also owing to the longer-period features of ambient noise than the seismic data from local earthquakes, the lower-crust is better illuminated. The total misfit reduction and a series of point-source resolution tests (Supporting Information Fig. S7) all support the robustness of our new tomographic model.

## ACKNOWLEDGEMENTS

The seismic data used were downloaded from the SCEDC (<http://service.scedc.caltech.edu/>). The authors would like to thank the developers of the open-source software SPECFEM3D\_Cartesian for their continued community work. We thank the two reviewers Laura Ermert and Yder Masson for their helpful suggestions which have improved the manuscript. KW benefits from discussion with Chuangxin Lin and Xing Song about SEM methods and also appreciates the help of Qi Zhao and Changyi Yu during his multiple visits to University of Toronto. KW is supported by an international Macquarie University Research Excellence Scholarship. QL is supported by the NSERC Discovery Grant 487237. PB and PT were supported by the G8 Research Councils Initiative on Multi-lateral Research Grant between 2013 and 2014. Computations for this study were performed on hardware acquired 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. YY is supported by Australian Research Council Future Fellowship (FT130101220). This is contribution 1190 from the ARC Centre of Excellence for Core to Crust Fluid Systems (<http://www.cccfs.mq.edu.au>) and 1242 in the GEMOC Key Centre (<http://www.gemoc.mq.edu.au>).

## REFERENCES

- Akcelik, V., Biros, G. & Ghattas, O., 2002. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, in *Supercomputing, ACM/IEEE 2002 Conference*, pp. 41–41, IEEE.
- Barak, S., Klemperer, S.L. & Lawrence, J.F., 2015. San Andreas Fault dip, Peninsular Ranges mafic lower crust and partial melt in the Salton Trough, Southern California, from ambient-noise tomography, *Geochem. Geophys. Geosyst.*, **16**(11), 3946–3972.
- Basini, P., Nissen-Meyer, T., Boschi, L., Casarotti, E., Verbeke, J., Schenk, O. & Giardini, D., 2013. The influence of nonuniform ambient noise on crustal tomography in Europe, *Geochem. Geophys. Geosyst.*, **14**(5), 1471–1492.
- Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A., Lin, F., Moschetti, M., Shapiro, N. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**(3), 1239–1260.
- Bensen, G., Ritzwoller, M. & Yang, Y., 2009. A 3-D shear velocity model of the crust and uppermost mantle beneath the United States from ambient seismic noise, *Geophys. J. Int.*, **177**(3), 1177–1196.
- Bozdağ, E., Peter, D., Lefebvre, M., Komatsitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. Suppl. Mon. Not. R. astr. Soc.*, **207**(3), 1739–1766.
- Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.
- Chen, M., Huang, H., Yao, H., Hilst, R. & Niu, F., 2014. Low wave speed zones in the crust beneath SE Tibet revealed by ambient noise adjoint tomography, *Geophys. Res. Lett.*, **41**(2), 334–340.
- Chen, M., Niu, F., Liu, Q., Tromp, J. & Zheng, X., 2015. Multiparameter adjoint tomography of the crust and upper mantle beneath East Asia: 1. Model construction and comparisons, *J. geophys. Res.*, **120**(3), 1762–1786.
- Chen, P., 2011. Full-wave seismic data assimilation: theoretical background and recent advances, *Pure appl. Geophys.*, **168**(10), 1527–1552.
- Clouzet, P., Masson, Y. & Romanowicz, B., 2018. Box Tomography: first application to the imaging of upper-mantle shear velocity and radial anisotropy structure beneath the North American continent, *Geophys. J. Int.*, **213**(3), 1849–1875.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.
- Ermert, L., Villaseñor, A. & Fichtner, A., 2015. Cross-correlation imaging of ambient noise sources, *Geophys. J. Int.*, **204**(1), 347–364.
- Fichtner, A., 2010. *Full Seismic Waveform Modelling and Inversion*, Springer Science & Business Media.
- Fichtner, A., 2014. Source and processing effects on noise correlations, *Geophys. J. Int.*, **197**(3), 1527–1531.
- Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.-P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**(3), 1703–1725.
- Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.-P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle, *Earth planet. Sci. Lett.*, **290**(3), 270–280.
- Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A. & Weaver, R.L., 2010. Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations, *Geophysics*, **75**(5), SA85–SA93.
- Gao, H. & Shen, Y., 2014. Upper mantle structure of the Cascades from full-wave ambient noise tomography: Evidence for 3D mantle upwelling in the back-arc, *Earth planet. Sci. Lett.*, **390**, 222–233.
- Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: Numerical results, *Geophysics*, **51**(7), 1387–1403.
- Guo, Z., Chen, Y.J., Ning, J., Yang, Y., Afonso, J.C. & Tang, Y., 2016. Seismic evidence of on-going sublithosphere upper mantle convection for intra-plate volcanism in northeast China, *Earth planet. Sci. Lett.*, **433**, 31–43.

- Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S., Davaille, A. & Yelles-Chaouche, A., 2015. Global tomography using seismic hum, *Geophys. J. Int.*, **204**(2), 1222–1236.
- Hauksson, E., 2000. Crustal structure and seismicity distribution adjacent to the Pacific and North America plate boundary in southern California, *J. geophys. Res.*, **105**(B6), 13 875–13 903.
- Humphreys, E.D. & Clayton, R.W., 1990. Tomographic image of the southern California mantle, *J. geophys. Res.*, **95**(B12), 19 725–19 746.
- Hung, S.-H., Shen, Y. & Chiao, L.-Y., 2004. Imaging seismic velocity structure beneath the Iceland hot spot: A finite frequency approach, *J. geophys. Res.*, **109**(B08305).
- Jiang, C., Yang, Y. & Zheng, Y., 2014. Penetration of mid-crustal low velocity zone across the Kunlun fault in the NE Tibetan plateau revealed by ambient noise tomography, *Earth planet. Sci. Lett.*, **406**, 81–92.
- Jiang, C., Schmandt, B., Hansen, S.M., Dougherty, S.L., Clayton, R.W., Farrell, J. & Lin, F.-C., 2018. Rayleigh and s wave tomography constraints on subduction termination and lithospheric foundering in central California, *Earth planet. Sci. Lett.*, **488**, 14–26.
- Komatitsch, D. & Vilotte, J.-P., 1998. The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.*, **88**(2), 368–392.
- Komatitsch, D., Xie, Z., Bozdağ, E., Sales de Andrade, E., Peter, D., Liu, Q. & Tromp, J., 2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion, *Geophys. J. Int.*, **206**(3), 1467–1478.
- Krischer, L., Fichtner, A., Zukauskaite, S. & Igel, H., 2015. Large-scale seismic inversion framework, *Seismol. Res. Lett.*, **86**, 1198–1207.
- Laske, G. & Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data, *J. geophys. Res.*, **101**(B7), 16 059–16 075.
- Lee, E.-J., Chen, P., Jordan, T.H., Maeckling, P.B., Denolle, M.A. & Beroza, G.C., 2014. Full-3-D tomography for crustal structure in southern California based on the scattering-integral and the adjoint-wavefield methods, *J. geophys. Res.*, **119**(8), 6421–6451.
- Lin, F.-C., Moschetti, M.P. & Ritzwoller, M.H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophys. J. Int.*, **173**(1), 281–298.
- Liu, Q. & Gu, Y., 2012. Seismic imaging: From classical to adjoint tomography, *Tectonophysics*, **566**, 31–66.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.*, **96**(6), 2383–2397.
- Liu, Y., Niu, F., Chen, M. & Yang, W., 2017. 3-D crustal and uppermost mantle structure beneath NE China revealed by ambient noise adjoint tomography, *Earth planet. Sci. Lett.*, **461**, 20–29.
- Lobkis, O.I. & Weaver, R.L., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. acoust. Soc. Am.*, **110**(6), 3011–3017.
- Luo, Y., 2012. *Seismic imaging and inversion based on spectral-element and adjoint methods*, PhD thesis, Princeton University.
- Luo, Y., Yang, Y., Xu, Y., Xu, H., Zhao, K. & Wang, K., 2015. On the limitations of interstation distances in ambient noise tomography, *Geophys. J. Int.*, **201**(2), 652–661.
- Maggi, A., Tape, C., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.*, **178**(1), 257–281.
- Masson, Y. & Romanowicz, B., 2017. Box tomography: localized imaging of remote targets buried in an unknown medium, a step forward for understanding key structures in the deep earth, *Geophys. J. Int.*, **211**(1), 141–163.
- Modrak, R. & Tromp, J., 2016. Seismic waveform inversion best practices: regional, global and exploration test cases, *Geophys. J. Int.*, **206**(3), 1864–1889.
- Montagner, J.-P. & Anderson, D.L., 1989. Petrological constraints on seismic anisotropy, *Phys. Earth planet. Inter.*, **54**(1–2), 82–105.
- Monteiller, V., Chevrot, S., Komatitsch, D. & Wang, Y., 2015. Three-dimensional full waveform inversion of short-period teleseismic wavefields based upon the SEM–DSM hybrid method, *Geophys. J. Int.*, **202**(2), 811–827.
- Nishida, K., Montagner, J.-P. & Kawakatsu, H., 2009. Global surface wave tomography using seismic hum, *Science*, **326**(5949), 112–112.
- Peter, D. et al., 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.*, **186**(2), 721–739.
- Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years, *Annu. Rev. Earth Planet. Sci.*, **31**(1), 303–328.
- Roux, P., Sabra, K.G., Kuperman, W.A. & Roux, A., 2005. Ambient noise cross correlation in free space: theoretical approach, *J. acoust. Soc. Am.*, **117**(1), 79–84.
- Schmandt, B. & Humphreys, E., 2010. Seismic heterogeneity and small-scale convection in the southern California upper mantle, *Geochem. Geophys. Geosyst.*, **11**(5).
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.
- Shaw, J.H. et al., 2015. Unified structural representation of the southern California crust and upper mantle, *Earth planet. Sci. Lett.*, **415**, 1–15.
- Shearer, P.M., 1997. Improving local earthquake locations using the l1 norm and waveform cross correlation: application to the Whittier Narrows, California, aftershock sequence, *J. geophys. Res.*, **102**(B4), 8269–8283.
- Shen, W., Ritzwoller, M.H., Schulte-Pelkum, V. & Lin, F.-C., 2012. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophys. J. Int.*, **192**(2), 807–836.
- Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E*, **69**(4), 046610.
- Stehly, L., Campillo, M. & Shapiro, N., 2006. A study of the seismic noise from its long-range correlation properties, *J. geophys. Res.*, **111**(B10).
- Stehly, L., Fry, B., Campillo, M., Shapiro, N., Guilbert, J., Boschi, L. & Giardini, D., 2009. Tomography of the alpine region from observations of seismic ambient noise, *Geophys. J. Int.*, **178**(1), 338–350.
- Tape, C., 2009. *Seismic tomography of southern California using adjoint methods*, PhD thesis, California Institute of Technology.
- Tape, C., Liu, Q. & Tromp, J., 2007. Finite-frequency tomography using adjoint methods—Methodology and examples using membrane surface waves, *Geophys. J. Int.*, **168**(3), 1105–1129.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**(5943), 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**(1), 433–462.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
- Thurber, C. & Ritsema, J., 2007. Theory and observations – seismic tomography and inverse methods, in *Seismology and Structure of the Earth, Vol. 1 of Treatise on Geophysics*, pp. 323–360, eds Romanowicz, B. & Dziewonski, A., Elsevier.
- Thurber, C.H., 1992. Hypocenter-velocity structure coupling in local earthquake tomography, *Phys. Earth planet. Inter.*, **75**(1–3), 55–62.
- Tong, P., Chen, C.-W., Komatitsch, D., Basini, P. & Liu, Q., 2014a. High-resolution seismic array imaging based on an SEM-FK hybrid method, *Geophys. J. Int.*, **197**(1), 369–395.
- Tong, P., Komatitsch, D., Tseng, T.-L., Hung, S.-H., Chen, C.-W., Basini, P. & Liu, Q., 2014b. A 3-D spectral-element and frequency-wave number hybrid method for high-resolution seismic array imaging, *Geophys. Res. Lett.*, **41**(20), 7025–7034.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**(1), 195–216.
- Tromp, J., Komatitsch, D. & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Commun. Comput. Phys.*, **3**(1), 1–32.

- Tromp, J., Luo, Y., Hanasoge, S. & Peter, D., 2010. Noise cross-correlation sensitivity kernels, *Geophys. J. Int.*, **183**(2), 791–819.
- Tsai, V.C., 2009. On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophys. J. Int.*, **178**(3), 1555–1564.
- Waldhauser, F. & Ellsworth, W.L., 2000. A double-difference earthquake location algorithm: method and application to the northern Hayward fault, California, *Bull. seism. Soc. Am.*, **90**(6), 1353–1368.
- Wang, K., Luo, Y. & Li, H., 2014. The nature of ambient noise over a field in western Junggar basin near Karamay, China, in *The 6th International Conference on Environmental and Engineering Geophysics*, pp. 92–97, Xian, China.
- Wang, K., Luo, Y. & Yang, Y., 2016. Correction of phase velocity bias caused by strong directional noise sources in high-frequency ambient noise tomography: a case study in Karamay, China, *Geophys. J. Int.*, **205**(2), 715–727.
- Wapenaar, K., Draganov, D., Snieder, R., Campman, X. & Verdel, A., 2010. Tutorial on seismic interferometry: Part 1-Basic principles and applications, *Geophysics*, **75**(5), 75A195–75A209.
- Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: Three-dimensional modeling of earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**(B7), 5953–5986.
- Xie, J., Ritzwoller, M.H., Shen, W., Yang, Y., Zheng, Y. & Zhou, L., 2013. Crustal radial anisotropy across eastern Tibet and the western Yangtze craton, *J. geophys. Res.*, **118**(8), 4226–4252.
- Yang, Y. & Forsyth, D.W., 2006. Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern California, *J. geophys. Res.: Solid Earth*, **111**(B07306).
- Yang, Y., Ritzwoller, M.H., Levshin, A.L. & Shapiro, N.M., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.*, **168**(1), 259–274.
- Yang, Y., Ritzwoller, M.H., Lin, F.-C., Moschetti, M. & Shapiro, N.M., 2008. Structure of the crust and uppermost mantle beneath the western United States revealed by ambient noise and earthquake tomography, *J. geophys. Res.*, **113**(B12310).
- Yanovskaya, T., Koroleva, T. & Lyskova, E., 2016. Effect of earthquakes on ambient noise surface wave tomography in upper-mantle studies, *Geophys. J. Int.*, **205**(2), 1208–1220.
- Yao, H. & Van Der Hilst, R.D., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet, *Geophys. J. Int.*, **179**(2), 1113–1132.
- Yao, H., van Der Hilst, R.D. & Maarten, V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis-I. Phase velocity maps, *Geophys. J. Int.*, **166**(2), 732–744.
- Yao, H., Beghein, C. & Van Der Hilst, R.D., 2008. Surface wave array tomography in SE Tibet from ambient seismic noise and two-station analysis-II. Crustal and upper-mantle structure, *Geophys. J. Int.*, **173**(1), 205–219.
- Zhang, H. & Thurber, C.H., 2003. Double-difference tomography: The method and its application to the Hayward fault, California, *Bull. seism. Soc. Am.*, **93**(5), 1875–1889.
- Zhao, D., Hasegawa, A. & Horiuchi, S., 1992. Tomographic imaging of p and s wave velocity structure beneath northeastern Japan, *J. geophys. Res.*, **97**(B13), 19 909–19 928.
- Zheng, Y., Shen, W., Zhou, L., Yang, Y., Xie, Z. & Ritzwoller, M.H., 2011. Crust and uppermost mantle beneath the north China craton, northeastern China, and the sea of Japan from ambient noise tomography, *J. geophys. Res.*, **116**(B12312).
- Zhou, Y., Dahmen, F. & Nolet, G., 2004. Three-dimensional sensitivity kernels for surface wave observables, *Geophys. J. Int.*, **158**(1), 142–168.
- Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.*, **5**(7), 493–498.
- Zhu, H., Bozdağ, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.*, **201**(1), 18–52.
- Zhu, H., Komatsitsch, D. & Tromp, J., 2017. Radial anisotropy of the North American upper mantle based on adjoint tomography with USArray, *Geophys. J. Int.*, **211**(1), 349–377.
- Zigone, D., Ben-Zion, Y., Campillo, M. & Roux, P., 2015. Seismic tomography of the Southern California plate boundary region from noise-based Rayleigh and Love waves, *Pure appl. Geophys.*, **172**(5), 1007–1032.

## SUPPORTING INFORMATION

Supplementary data are available at [GJI](#) online.

**Figure S1.** (a–e) Line search results at each iteration showing the variation of misfit function values as a function of step length for various period bands and the total misfit (also connected by solid curves); (f) Total misfit reduction over iterations for broad-band data.

**Figure S2.** Histograms of traveltimes misfits obtained for model **M16** (a), **M17** (b), **M19** (c) and **M21** (d) with the use of broadband data. The green solid bars represent misfits for the initial model, and the red bars in (b–d) are misfits for updated models through iterations. The total number of good measurements, average misfit and standard deviation are shown at the top of each histogram with the same colour specified above. Misfits are measured at all three period bands, 5–50 (left), 10–50 (middle) and 20–50 (right) s.  $CC_{\min}$  in (a) are specially chosen as 0.80 (5–50 s), 0.75 (10–50 s), 0.69 (20–50 s) to show the misfit of the initial model. All the misfits are measured with  $CC_{\min} = 0.75$  in (b) and  $CC_{\min} = 0.80$  in (c) and (d) for comparison. Note that 5–50 and 10–50 s bands of **M16** and 5–50 s band of **M17** are not used in the inversion.

**Figure S3.** Horizontal slices of shear velocity for model **M21** (left panel) using broad-band data, **M21.narrowband** (middle panel) using narrowband data, and their differences (right panel) in percentage at 5, 10, 15 and 20 km depths.

**Figure S4.** Same as Fig. S3 but for 25, 30, 35, 40 km depths.

**Figure S5.** (a) Ray paths of all station pairs displayed; (b–d) waveforms of station pairs filtered at 5–50 (first column), 10–50 (second column) and 20–50 (third column) s. The EGFs are displayed as black solid lines, and SGFs are plotted as green solid lines for **M16** and red solid lines for **M21**. Cross-correlation traveltimes and uncertainties are also shown with corresponding colours.

**Figure S6.** (a–d) Volumetric sensitivities for Vs tomographic models shown at depths of 10, 15, 20 and 25 km. (e–h) Masked volumetric sensitivities determined by a subjective threshold value of  $K = 4.0 \times 10^{-13} \text{ m}^{-3}$ .

**Figure S7.** Model resolution tests. The left panels show the locations of the fast/slow shear velocity anomalies at the depth of 15 km: (a) one with a 5 km radius put at the LA basin and (b) the other with a 10 km radius at northeastern part of the model region. The right panels are the corresponding Hessian kernels (Fichtner & Trampert 2011; Zhu *et al.* 2015) showing the degree of blurring in recovering the model perturbations.

**Figure S8.** Comparison of model **M16**, **M21**, Barak2015 and CVM-S4.26 along profile AA' across SAF.

**Figure S9.** Comparison of model **M16**, **M21**, Barak2015 and CVM-S4.26 along profile BB' across SAF.

**Figure S10.** Comparison of model **M16**, **M21**, Barak2015 and CVM-S4.26 along profile CC' across SAF.

**Figure S11.** Comparison of model **M16**, **M21**, Barak2015 and CVM-S4.26 along profile DD' across SAF.

**Figure S12.** Comparison of model **M16**, **M21**, Barak2015 and CVM-S4.26 along profile EE' across SAF.

**Figure S13.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile FF' across SAF.

**Figure S14.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile GG' across SAF.

**Figure S15.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile HH' across SAF.

**Figure S16.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile II' across SAF.

**Figure S17.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile PRaa' across PRB.

**Figure S18.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile PRbb' across PRB.

**Figure S19.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile PRcc' across PRB.

**Figure S20.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile SNaa' across SNB.

**Figure S21.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile SNbb' across SNB.

**Figure S22.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile SNcc' across SNB.

**Figure S23.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile STaa' across ST.

**Figure S24.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile STbb' across ST.

**Figure S25.** Comparison of model **M16,M21**, Barak2015 and CVM-S4.26 along profile STcc' across ST.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

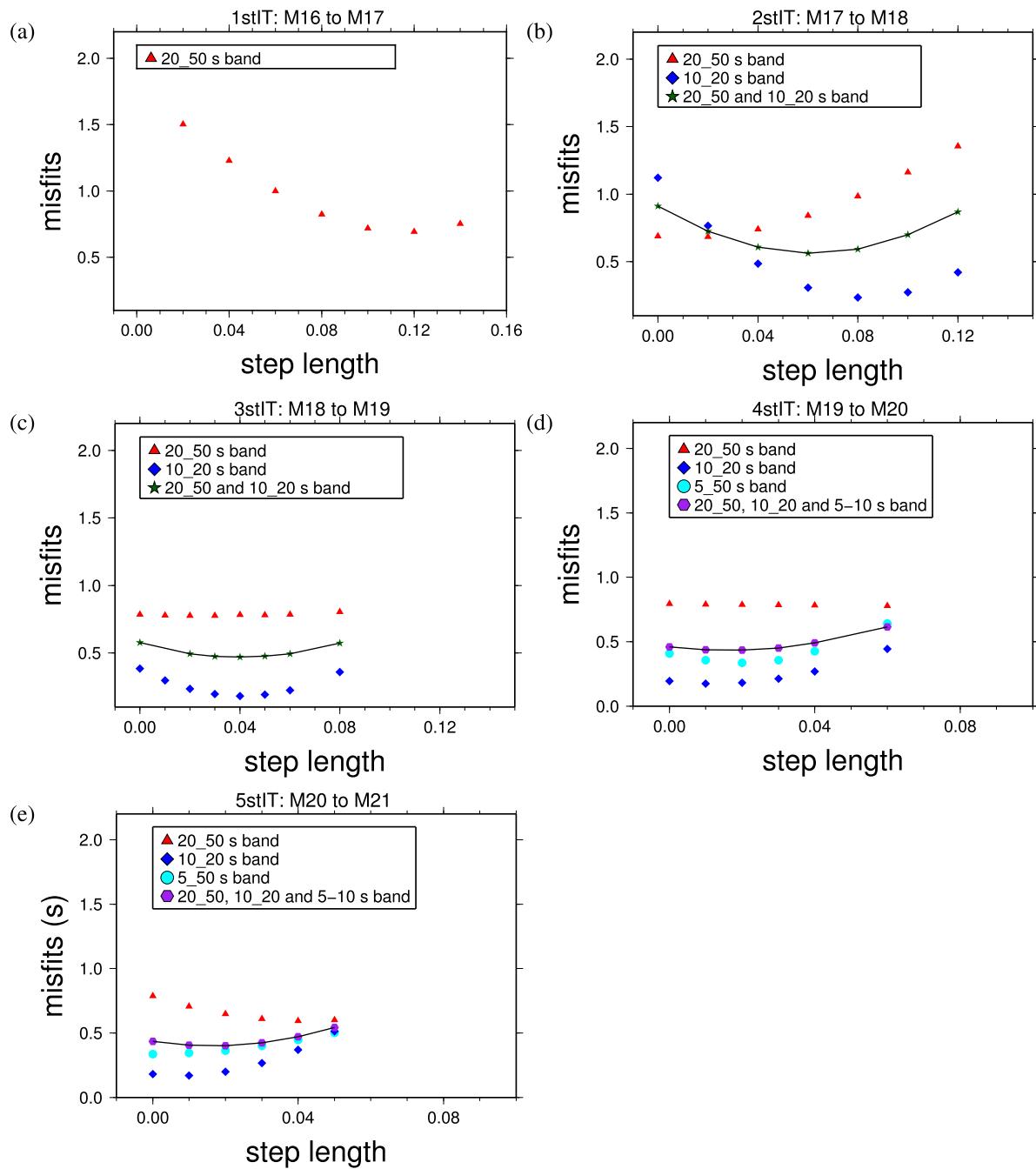
## APPENDIX A: LINE SEARCH AND MODEL UPDATE

We update the model  $\delta m = (\delta\alpha, \delta\beta, \delta\rho)$  based on a search direction  $\mathbf{d}$  defined as the negative preconditioned and smoothed misfit gradient

$$\Delta\mathbf{m} = \alpha\mathbf{d} = \alpha(-\mathbf{P}^{-1}\mathbf{g}), \quad (\text{A1})$$

where  $\alpha$  represents the step length,  $\mathbf{g}$  denotes the misfit gradient, and  $\mathbf{P}$  is the pre-conditioner. A few trial models can be built for a series of  $\alpha$  values, and their corresponding total misfits are calculated through forward simulations from which the optimal step length can be determined. However, this involves  $N_s$  number of forward simulations for each  $\alpha$  value and therefore is computationally very expensive. Instead, a representative subset of 19 master stations that evenly cover the surface region (red triangles in Fig. 1) are selected to do the line search. We have tested some other subsets of stations and found that including a number of stations in the LAB area helps better represent the total misfit.

Figs A1(a)–(e) show the line search results at each iteration, with the corresponding optimal step lengths listed in Table 1. We add 10–20 s and 5–10 s bands at the second (2ndIT) and fourth (4thIT) iteration when the averaged total misfit at longer-period bands changes little. When multiple frequency bands are included, we usually choose the optimal step length from the minimum of the averaged total misfit curve (solid lines in Figs A1 a–e). This sometimes results in slight increase of misfit for the band(s) that have already been reduced to local minimum in previous iterations. For example, at the second iteration, the total misfit for 20–50 s band increases slightly when we introduce 10–20 s band (Fig. A1b). At the third iteration, we choose a more conservative step length of 0.03 instead of the optimal one at 0.05 from averaged total misfit curve to ensure the total misfit for 10–20 s drops.



**Figure A1.** (a–e) Line search results at each iteration, showing the variation of misfit function values as a function of step length for various period bands and the total misfit (also connected by solid curves).