

Waveform inversion of large data sets for radially anisotropic Earth structure

M. Witek,¹ S.-M. Lee,² S.-J. Chang² and S. van der Lee³

¹*Department of Earth and Environmental Sciences, Korea University, Seoul 02841, South Korea*

²*Department of Geophysics, Kangwon National University, Chuncheon 24341, South Korea. E-mail: sjchang@kangwon.ac.kr*

³*Department of Earth and Planetary Sciences, Northwestern University, Evanston, IL 60208, USA*

Accepted 2022 September 30. Received 2022 August 30; in original form 2022 March 28

SUMMARY

The amount of high-quality seismic data is expanding rapidly, and there is a need for algorithms that take advantage of classical methods to achieve high efficiency using widely available computing power. In this study, we develop a novel waveform inversion method to retrieve radially anisotropic Earth models that can be used to investigate deformation and flow in the mantle. Our method is comprised of two parts: (1) extraction and fitting of the fundamental mode and (2) fitting of the full synthetic waveform. The waveform inversion method results in path average model constraints with uniquely determined independent uncertainties. We demonstrate through synthetic testing that the method is able to retrieve radially anisotropic perturbations down to the mantle transition zone, and leakage effects due to ignoring *P*-wave anisotropy are minimal. We apply the method to ~16 000 waveforms generated by earthquakes occurring in the East Sea (Sea of Japan) region, and we demonstrate that the subsequent linear inversion of radially anisotropic path constraints produces models that are similar to those resulting from full waveform adjoint tomography methods. We validate our model by predicting waveforms for earthquakes not included in our inversion, and we show that our method is able to extract structural information. Our results indicate low-velocity anomalies and weak radial anisotropy in NE Japan, which may be due to competing influences from ascending fluids and/or melts and horizontal flow in the lower crust and upper mantle. In the southern East Sea, we image low velocities and relatively high radial anisotropy, which may reflect high temperatures, shallow dehydration and olivine LPO in the upper mantle.

Key words: Inverse theory; Waveform inversion; Seismic anisotropy; Seismic tomography; Surface waves and free oscillations; East Sea.

1 INTRODUCTION

The number of seismic data and their quality has been steadily increasing over the past several decades (e.g. Meltzer *et al.* 1999; Romanowicz & Giardini 2001; Simons *et al.* 2009), leading to an abundance of global and regional models of the Earth's crust and mantle (e.g. Romanowicz 2003). At the same time, advances in numerical methods and new developments in computing have made it possible to combine these new data sets with sophisticated tomographic techniques that model the physics of wave propagation to a level of detail not possible with analytical methods (e.g. Rawlinson *et al.* 2010; Liu & Gu 2012). For example, the adjoint-state method allows for the direct computation of the gradient of an objective function, saving time on the numerical calculation of Fréchet derivatives (e.g. Plessix 2006). Although formulated in the 1970s (e.g. Chavent 1974) and introduced to geophysics in the 1980s (e.g. Tarantola 1984), it was not until recently that the method started

gaining traction in the seismic tomography community (e.g. Tromp *et al.* 2005; Fichtner *et al.* 2006) with the advent of efficient parallel waveform simulations (e.g. Komatitsch *et al.* 2002; Fichtner *et al.* 2009).

However, full waveform inversion methods typically require access to high-performance computing resources. A recent global adjoint tomography study by Bozdağ *et al.* (2016) cited the use of 18 000 graphics processing units (GPUs) on the Oak Ridge Leadership Computing Facility Cray 'Titan' system. Such a significant computational effort may even involve challenges in achieving reasonable I/O throughput, necessitating the invention of new data formats and management strategies (e.g. Krischer *et al.* 2016; Lefebvre *et al.* 2017). Lei *et al.* (2020) extended the work of Bozdağ *et al.* (2016) by expanding the number of events from 253 to 1480 earthquakes, and they write that each iteration of their nonlinear optimization algorithm generates 'a few Petabytes of wavefield files.' These are impressive achievements pushing the boundaries of the field of

seismic tomography, but these kinds of numerical simulations at large supercomputing facilities may be out of reach for seismologists with limited computational resources. On the other hand, standard workstation GPU devices with thousands of cores continue to be developed, and careful extension of CPU algorithms can obtain significant increases in computational speed using just a single GPU device (e.g. Komatitsch *et al.* 2010), making it feasible to perform small-scale regional simulations. However, full waveform inversions are still dependent on an initial model to avoid the risk of converging to a local minimum (e.g. Chen *et al.* 2012; Bozdağ *et al.* 2016; Simutè *et al.* 2016; Tao *et al.* 2018). Therefore, there is a need to develop inversion methods that can feasibly run on regular hardware and generate solutions that can be used for further analysis.

The most popular and well-known method for forward calculations in seismic tomography is ray theory. Known for its simplicity and computational efficiency, a reference model is used to calculate the path and arrival time of a body wave in the case of traveltimes tomography (e.g. Crotwell *et al.* 1999). With publicly available catalogues of earthquake locations, focal mechanisms, and phase arrivals (e.g. Willemann & Storchak 2001; Ekström *et al.* 2012; Storchak *et al.* 2013), researchers have been able to construct increasingly detailed images of the Earth's mantle from measurements of P - and S -wave traveltimes (e.g. Inoue *et al.* 1990; Grand *et al.* 1997; van der Hilst *et al.* 1997; Bijwaard *et al.* 1998; Zhao 2004; Burdick *et al.* 2008). Complementary to body wave ray theory techniques are multimode waveform methods that use normal mode summation to create and compare synthetic waveforms with seismic records. By relating variations in the phases of surface and higher mode waveforms to averaged structural perturbations along a ray path, researchers have mapped the Earth's mantle in continental-scale (e.g. van der Lee & Nolet 1997; Lebedev & Nolet 2003; Maggi & Priestley 2005; Bedle & van der Lee 2009; Manaman *et al.* 2011) and global studies (e.g. Kustowski *et al.* 2008; Lebedev & van der Hilst 2008; Ho *et al.* 2016; Priestley *et al.* 2021).

In order to take advantage of the increasing number of high-quality broad-band seismic data, automated algorithms are becoming increasingly important. For example, while early studies performed manual analysis of $\sim 10^3$ seismograms (e.g. van der Lee & Nolet 1997), studies now automatically analyse $\sim 10^6$ seismograms (e.g. Schaeffer & Lebedev 2014), which would require significant time investments by multiple researchers in a manual study. Debayle (1999) automated the method of Cara & Lévêque (1987) to image the β_V structure of the upper mantle beneath Australia using 668 Rayleigh waveforms. Debayle & Ricard (2012) and Ho *et al.* (2016) later extended the method to image global β_V and β_H structure by inverting hundreds of thousands of Rayleigh and Love waveforms based on an initial 3-D reference model. Debayle *et al.* (2016) presented a model of azimuthally anisotropic β_V from an automated inversion of over 1.3 million Rayleigh waves, and now the model is continuously being updated as more waveforms are recorded. Lebedev & Nolet (2003) introduced an automated multimode inversion (AMI) method based on the partitioned waveform inversion (PWI) method (Nolet 1990), and later expanded AMI to include azimuthal anisotropy variations, P -velocity variations, and a reference 3-D model (Lebedev *et al.* 2005; Lebedev & van der Hilst 2008; Schaeffer & Lebedev 2013). Lebedev & van der Hilst (2008) and Schaeffer & Lebedev (2013) performed extensive testing of the AMI method using synthetic waveforms generated by the spectral element method (e.g. Komatitsch *et al.* 2002) and verified that the ray theory approximation is valid for the data set and sizes of heterogeneities considered.

In this study, we present a new waveform inversion algorithm based on PWI. Our method is also inspired by the AMI method of Lebedev *et al.* (2005), but differs from it in several aspects. First, whenever possible, we simultaneously fit three-component waveforms in order to better constrain path-averaged radial anisotropy, while AMI has focused on the vertical component (e.g. Schaeffer *et al.* 2016). Second, our fitting algorithm is composed of two steps where the fundamental mode is extracted and fit first to create an initial model for the subsequent step, in which the full synthetic mode sum is fit. Splitting the algorithm in this way can be helpful in avoiding cycle skipping. Third, we adopt a novel strategy to estimate the uncertainty in the linear path constraints that result from a successful waveform fit. Finally, our method can be easily applied to any 3-D reference model from which 1-D profiles can be extracted. The method is able to produce linearly independent constraints for a radially anisotropic Earth model that can be inverted either alone or as part of a joint inversion of various complementary data sets. For example, resolution testing by Chang *et al.* (2010) showed higher resolving power down to 1400 km depth by jointly inverting teleseismic arrival times, Rayleigh wave group velocities, and regional waveform fits than when inverting each data set separately. Joint inversions of millions of complementary data are computationally efficient and are now routinely performed (e.g. Feng *et al.* 2007; Kustowski *et al.* 2008; Schmid *et al.* 2008; Feng *et al.* 2010; Chang & van der Lee 2011; Ritsema *et al.* 2011; Auer *et al.* 2014; Chang *et al.* 2015), and, with denser data coverage, may approach the accuracy found in finite-frequency tomography (e.g. Sieminski *et al.* 2004; van der Lee & Frederiksen 2005; Boschi 2006).

Studies of seismic anisotropy can be helpful in determining the state of deformation in the crust and mantle, which can indicate present-day flow or past tectonic episodes. Here we focus on radial anisotropy, which can occur in layered isotropic media, anisotropic media, or any system displaying hexagonal symmetry (Anderson 1961). Such a medium will lead to an apparent discrepancy between observations of Rayleigh waves, which are mainly sensitive to perturbations in β_V , and Love waves, which are sensitive to variations in β_H . Efforts to resolve the discrepancy led to the incorporation of radial anisotropy in the Preliminary Reference Earth Model (PREM; Dziewoński & Anderson 1981), which remains one of the most widely used reference 1-D Earth models. Numerous studies around the globe have shown that isotropic models are unable to simultaneously fit Rayleigh and Love surface wave data, and radial anisotropy has been invoked and interpreted as evidence of current lower crustal flow (e.g. Shapiro *et al.* 2004), past crustal extension (e.g. Guo *et al.* 2016; Ai *et al.* 2020), ancient cratonic deformation (e.g. Priestley *et al.* 2021), and present asthenospheric flow (e.g. Panning & Romanowicz 2006).

However, while models of isotropic V_S structure largely agree (e.g. Lekić & Romanowicz 2011; Lekić *et al.* 2012), models of radial anisotropy show little coherency (e.g. Chang *et al.* 2014). Therefore, the primary motivation and focus of this study is to provide an automatic method that is able to retrieve additional constraints for radial anisotropy by simultaneously fitting multicomponent seismic waveforms, and the main benefit of the constraints is expected to be in regional-scale joint inversion studies (e.g. Schmid *et al.* 2008; Chang *et al.* 2010). In the following, we first give a short overview of the ray theory approximations we use. Then, we provide an explanation of our automatic algorithm. We apply our method to three-component S and surface waves recorded by seismic networks in the East Sea (Sea of Japan) region and show that the resulting model can achieve performance similar to other regional tomography studies.

2 METHODOLOGY

2.1 Background

We model displacement seismograms using the JWKB (after Jeffreys, Wentzel, Kramers and Brillouin) approximation (ray theory) as

$$u(\omega) = \sum_n A_n(\omega) \exp \left\{ i \int_{\Delta} k_n(s; \omega) ds \right\}, \quad (1)$$

where the sum is over all modes n along a great-circle minor arc from the source to the receiver. $A_n(\omega)$ is a complex term that includes receiver and source effects as well as attenuation effects due to geometric spreading and attenuation structure along the path. The path integral is done along the minor arc with Δ being the arc length and s representing a latitude, longitude position on the arc. We calculate source effects using moment tensor solutions from the Global Centroid Moment Tensor (CMT) catalogue (Ekström *et al.* 2012). We assume that most of the change in the seismogram comes from structural effects, so that $k_n(s; \omega) = k_n^0(s; \omega) + \delta k_n(s; \omega)$, where $k_n^0(s; \omega)$ is a local wavenumber found by solving the normal modes problem for a 1-D Earth with the same structure as a 1-D profile under the point s in a 3-D reference model $\mathbf{m}^0(s, r)$. The second term, $\delta k_n(s; \omega)$, is a perturbation to the local wavenumber due to first-order differences between the true Earth and the reference model. Fractional variations in the local wavenumbers, $\delta k_n(s; \omega)/k_n^0(s; \omega)$, can be represented as depth integrals of the associated fractional variations in the structural parameters, $\delta m_i(s, r)/m_i^0(s, r)$, weighted by local sensitivity kernels $K_{ni}(s, r; \omega)$,

$$\frac{\delta k_n(s; \omega)}{k_n^0(s; \omega)} = \int_0^a \sum_i K_{ni}(s, r; \omega) \frac{\delta m_i(s, r)}{m_i^0(s, r)} dr, \quad (2)$$

where a is the radius of the Earth, and $i = 1, 2, 3, \dots, N$ indexes the model parameters. For example, $i = 1$ could reference β_V , $i = 2$ could reference β_H , etc., depending on the chosen model parametrization. Combining eq. (2) with the path integral in eq. (1) yields

$$\int_{\Delta} k_n(s; \omega) ds = \int_{\Delta} \int_0^a k_n^0(s; \omega) \left[1 + K_{ni}(s, r; \omega) \frac{\delta m_i(s, r)}{m_i^0(s, r)} \right] dr ds. \quad (3)$$

Given $\mathbf{m}^0(s, r)$, the first term on the right-hand side of eq. (3) can be separately calculated to give a reference path-averaged wavenumber $\overline{k_n^0(\omega)}$. Next, we follow Nolet (1990) and make the assumption that these model differences can be adequately described by a 1-D path-average, defined as

$$\overline{\delta \ln m_i(r)} = \frac{1}{\Delta} \int_{\Delta} \frac{m_i(s, r) - m_i^0(s, r)}{m_i^0(s, r)} ds, \quad (4)$$

where $\mathbf{m}(s, r)$ is the true Earth model, and $\delta \mathbf{m}(s, r) = \mathbf{m}(s, r) - \mathbf{m}^0(s, r)$. We then expand the path-averaged structural variations onto radial basis functions $h_j(r)$,

$$\frac{\overline{\delta \ln m_i(r)}}{\sigma_i} = \sum_j \gamma_{ij} h_j(r), \quad (5)$$

where γ_{ij} are the radial basis function coefficients, $j = 1, \dots, M$ and M is the total number of depth knots defined at r_j . Here, we follow the method of van der Lee & Nolet (1997), and incorporate *a priori* estimates of the model parameter standard deviations σ_i . Typical choices for the $h_j(r)$ are cubic (e.g. Mégnin & Romanowicz 2000) or linear splines (e.g. van der Lee & Nolet 1997; Schaeffer &

Lebedev 2013). In this study, we use triangular linear splines with more details given in Section 2.2. The basis function expansion allows eq. (3) to now be expressed as

$$\int_{\Delta} k_n(s; \omega) ds = \overline{k_n^0(\omega)} \Delta + \sum_i \sum_j \gamma_{ij} \times \int_{\Delta} \int_0^a k_n^0(s; \omega) \sigma_i K_{ni}(s, r; \omega) h_j(r) dr ds. \quad (6)$$

In the original PWI method, Nolet (1990) suggested the use of different 1-D reference models for each path, which would greatly simplify the preceding expressions and reduce the amount of computation time since the wavenumbers and sensitivity kernels would only have to be calculated once per path. However, as noted by Lebedev & van der Hilst (2008), the choice of a representative 1-D model is not straightforward, particularly if the structure along the path is known to vary significantly. Furthermore, a 1-D reference model with the same sensitivity kernels as a path-average kernel calculated through a 3-D reference model is not guaranteed to exist. With the computational advances made since the study of Nolet (1990), however, solving the normal modes problem for a set of arbitrary 1-D profiles extracted from a 3-D reference model along a ray path is fairly straightforward.

The implementation of Lebedev & van der Hilst (2008) and Schaeffer & Lebedev (2013) relies on pre-calculating $K_{ni}(s, r; \omega)$ for a subset of representative models extracted from their 3-D reference model, which consists of smoothed version of CRUST2.0 (Bassin *et al.* 2000) overlying AK135 (Kennett *et al.* 1995). This allows for the rapid calculation of the second term in eq. (6) on a fine integration grid for any source–receiver path. In order to keep our method more general and applicable to any 3-D reference model, we calculate the second term in eq. (6) by discretizing the ray path into a series of points and extracting a 1-D profile from $\mathbf{m}^0(s, r)$ for each point. Details on the 3-D reference model used in this study are given in Section 2.3, but we note that any pre-existing 3-D model can be used to calculate synthetic seismograms as long as 1-D profiles can be extracted from it. Next, we calculate the local wavenumbers and sensitivity kernels for each point and project them onto the radial basis functions by first calculating the integral over depth. In practical terms, calculating the depth integral first avoids problems with varying crustal discontinuities along the ray path since the depth integration can be done piecewise, and the result is set of $M \times N$ scalar values for each point s . The integration over the ray path can now proceed trivially, and the double integral in eq. (6) can be collapsed into a matrix $\overline{K_{nij}(\omega)}$ that represents a path-average sensitivity kernel. Thus, the path integral in eq. (1) can be expressed as

$$\int_{\Delta} k_n(s; \omega) ds = \overline{k_n^0(\omega)} \Delta + \sum_i \sum_j \gamma_{ij} \overline{K_{nij}(\omega)}. \quad (7)$$

We will mention here that although we use the great-circle approximation in this study, the method can be extended to full ray theory by incorporating surface wave ray tracing through the reference model. The form of eq. (7) would remain unchanged, but the path integrals would depend on the ray paths for each mode and frequency considered. As a final practical note, we calculate the local modes using a heavily modified form of the surface wave codes of Nolet (2008), which are based on many of the original routines in the normal modes solver MINEOS (Woodhouse 1988).

Having derived a relationship between 1-D path-average model variations with respect to a 3-D reference model and a displacement seismogram, we will now detail a method to find an optimal set of

radial basis function coefficients γ_{ij} that minimizes the difference between a recorded and synthetic seismogram. First, we describe our 3-D model parametrization in Section 2.2 and the 3-D reference model in Section 2.3. Then, the definition of the waveform misfit and the general nonlinear optimization procedure for a three-component seismic waveform is given in Section 2.4. To facilitate automatic waveform fitting, we develop an algorithm that is split into two parts: (1) fundamental mode extraction, windowing, and fitting and (2) full synthetic waveform windowing and fitting. By attempting to fit the fundamental mode first, we can find an initial set of path average model perturbations that explain a significant portion of the waveform misfit. The fundamental mode portion of the algorithm is explained in Section 2.5. In the second portion, the model found by fitting the fundamental mode is used as a starting point to fit the full waveform sum. In this study, the mode sum contains modes $n = 0, 1, 2, \dots, 20$, which should be sufficient to reconstruct body wave phases such as S and S -wave multiples that result from higher mode interference. The initial perturbations (i.e. the γ_{ij}) used in this stage may also be zero if the fundamental mode is not sufficiently excited, as in the case of deep events. The full synthetic waveform fitting is described in Section 2.6. In Section 2.7 we give a description of how the linear constraints and their uncertainties are calculated following a successful nonlinear waveform fit. Finally, in Section 2.8 we describe how the set of linearly independent path constraints are gathered in order to perform a large linear inversion for 3-D model perturbations with respect to the reference model.

2.2 3-D Model parametrization

In this study, we focus on a radially anisotropic medium with a vertical symmetry axis. Such a medium can be defined by the density ρ and the five Love parameters $C = \rho\alpha_V^2$, $A = \rho\alpha_H^2$, $L = \rho\beta_V^2$, $N = \rho\beta_H^2$ and $F = \eta\rho(\alpha_H^2 - 2\beta_V^2)$ (Love 1927; Takeuchi & Saito 1972). The first two parameters, C and A , are related to vertically and horizontally propagating compressional waves. The next two parameters, L and N , give the speeds of vertically and horizontally polarized shear waves travelling in a horizontal direction. The final parameter, F , is related to the speeds of body waves travelling in intermediate directions via the parameter η . Here, we parametrize our model by defining the isotropic shear wave speed and shear wave anisotropy as follows:

$$V_S^2 = \frac{1}{2}(\beta_V^2 + \beta_H^2) = \frac{1}{2}(L + N), \quad (8)$$

$$\zeta_S = \frac{\beta_H^2 - \beta_V^2}{2V_S^2} = \frac{N - L}{L + N}. \quad (9)$$

Our sensitivity kernels were derived under the assumption that a model perturbation will produce the same shift in eigenfrequency for any parametrization (e.g. Panning & Romanowicz 2006), and so in this study we freely convert our results to different parametrizations. In order to make comparisons with other studies, we convert our model results to Voigt averaged $v_S = \sqrt{(2\beta_V^2 + \beta_H^2)}/3$ and the traditional anisotropy parameter $\xi = (\beta_H/\beta_V)^2$ when displaying tomographic images.

Previous studies have found that while surface waves are sensitive to variations in ρ and V_P , surface wave data alone are not enough to resolve them. For example, Tanimoto (1991) conducted a waveform inversion of long-period data and concluded that V_S variations are not strongly affected by the inclusion of ρ variations and that the resolution of ρ anomalies was low. Therefore, in this study we couple variations in density and isotropic P -wave velocity to variations

in isotropic S -wave velocity via Anderson *et al.* (1968)

$$\frac{d \ln \rho}{d \ln V_S} = 0.4, \quad (10)$$

and Robertson & Woodhouse (1995)

$$\frac{d \ln V_P}{d \ln V_S} = 0.5. \quad (11)$$

The sensitivities of the surface waves to other parameters and azimuthal anisotropy are ignored in this study. A study of the effects of including prior constraints for η and P -wave anisotropy was conducted by Beghein (2010), and they found that the patterns of upper mantle radial anisotropy were not strongly affected, but the amplitudes had larger uncertainties when prior constraints were not included. A study by Panning & Romanowicz (2006) tested the results of different scaling factors for the anisotropy and also performed a test where η , ξ , and P -wave anisotropy were simultaneously inverted. They found good agreement among all resulting ξ models, except for the region around the 670-km discontinuity where large trade-offs may exist. Since not much is known about η structure in the Earth, we will neglect η in this study but we will note that its effect on Rayleigh waves can potentially mask radial anisotropy. We discuss potential biases from ignoring P -wave anisotropy in Section 3.2.

Unaccounted for azimuthal anisotropy also has the potential to bias the results (e.g. Ekström & Dziewonski 1998; Ekström 2011; Auer *et al.* 2014), especially in regions where large anisotropy may exist in thin lithosphere (e.g. Lloyd & van der Lee 2008). However, sufficient waveform coverage may cause the effects of the azimuthal anisotropy to be averaged out (Montagner & Nataf 1988). Further discussion is given in Section 4.3.

The 3-D model perturbations are parametrized on nested shell grids and the node locations are determined using the spherical tessellation method of Wang & Dahlen (1995). The shell grid is the same as that used by Witek *et al.* (2021) in order to facilitate future joint inversions, and here we give a brief description. The centre of the grid is placed at 37.5°N, 110.5°E and extends 50° in all directions. The node spacing is approximately 50 km on average at the surface. The model parameters are expanded onto linear triangle basis functions in the radial direction, with half triangles used for the first and last depth nodes. There are 27 depth nodes located at 0, 5, 10, 20, 35, 55, 75, 95, 120, 145, 170, 200, 230, 260, 290, 320, 350, 380, 410, 450, 490, 530, 610, 660, 720, 780 and 950 km depth. The model parameters are expanded onto the 3-D grid using

$$m_i(r, \theta, \phi) = \sum_{j,k} \mu_{ijk} f_j(\theta, \phi) h_k(r), \quad (12)$$

where the summation is done over the grid nodes j and depths k . The $f_j(\theta, \phi)$ are local spline coefficients derived through linear interpolation of the model in the lateral direction (van der Lee & Nolet 1997). The radial basis functions are represented by $h_k(r)$, and we are solving for the basis function coefficients μ_{ijk} . Here, $m_{i=1}$ represents fractional perturbations to V_S with respect to the reference 3-D model, and $m_{i=2}$ are ζ perturbations also with respect to the reference 3-D model. Additionally, we include perturbations to the Moho depth, which are expanded on a single lateral grid identical to the one described above. The Moho perturbations are treated in a similar manner as shown by van der Lee & Nolet (1997), but these are mainly included to limit crustal structure from leaking into the upper mantle (e.g. Chang *et al.* 2015).

2.3 Reference 3-D model

Typical studies of the upper mantle employ ‘crustal corrections’ to mitigate the nonlinear effect of the crust on seismic waveforms (e.g. Levshin & Ratnikova 1984; Woodhouse & Dziewonski 1984; Marone & Romanowicz 2007; Lekić *et al.* 2010; Panning *et al.* 2010). However, Bozdağ & Tampert (2008) found that improper crustal corrections can strongly bias upper mantle models of radial anisotropy, and suggested inverting for both crustal and mantle perturbations starting from a good 3-D reference model. Ferreira *et al.* (2010) studied the effects of different reference crustal models and found substantial differences in upper mantle radial anisotropy, suggesting that data able to resolve crustal structure should be included. Chang & Ferreira (2017) found that including short-period (<20 s) group velocity data in a joint inversion for global radially anisotropic mantle structure and crustal thickness helps to constrain thin oceanic crust and limits the effects of mantle structure contamination by unmodelled crustal effects.

Therefore, in this study we construct local sensitivity kernels with respect to a reference 3-D model to include crustal nonlinearity in our predicted waveforms, and we simultaneously invert for crustal and mantle perturbations using waveforms containing short periods (10 s for body waves, 16 s for the fundamental mode). The reference model is the same as that used in Witek *et al.* (2021). We use the CRUST1.0 model for the crust, which is a $1^\circ \times 1^\circ$ block model, each block having a unique eight-layer profile defined by V_P , V_S , and ρ (Laske *et al.* 2013). We assign the profiles to the centre of each block and use bilinear interpolation laterally in order to create a smoother reference model without significant vertical discontinuities. We use the AK135 model of Kennett *et al.* (1995) below the Moho. The sub-Moho model parameters in CRUST1.0 are taken from a modified version of the LLNL-G3Dv3 model (Simmons *et al.* 2012), which may cause undesirable ‘spikes’ if those parameters differ significantly from AK135. Therefore, we linearly interpolate model values from just below the Moho to 120 km depth in AK135.

2.4 Waveform misfit and nonlinear optimization

Given an observed record $d(t)$, we solve for the γ_{ij} in eq. (7) using a nonlinear optimization method. We minimize a misfit function defined as

$$F(\boldsymbol{\gamma}) = \begin{cases} \frac{1}{N_L} \sum_i F_i^L(\boldsymbol{\gamma}) & N_R = 0, N_L > 0 \\ \frac{1}{N_R} \sum_i F_i^R(\boldsymbol{\gamma}) & N_R > 0, N_L = 0, \\ \frac{1}{2N_L} \sum_i F_i^L(\boldsymbol{\gamma}) + \frac{1}{2N_R} \sum_i F_i^R(\boldsymbol{\gamma}) & N_R > 0, N_L > 0, \end{cases} \quad (13)$$

where the subscripts and superscripts L and R refer to Love and Rayleigh waves, respectively, N_L is the number of Love wave windows and N_R is the number of Rayleigh wave windows. The misfit may include either Rayleigh information, Love information or both depending on various factors (see Sections 2.5 and 2.6), so we attempt to equalize the contributions from the vertical and radial components, which carry Rayleigh wave information, and the transverse component, which carries Love wave information.

The observed waveform is separated into various time-frequency windows, so that the $F_i(\boldsymbol{\gamma})$ are misfits for each window defined as

$$F_i(\boldsymbol{\gamma}) = \frac{\int_{t_i} w_i(t) [d(t) - u(\boldsymbol{\gamma}, t)]^2 dt}{\int_{t_i} w_i(t) d^2(t) dt}. \quad (14)$$

Each time-frequency window is a bandpass filtered and time-windowed portion of a single component of the recorded or synthetic waveform that may contain the fundamental mode, higher modes, or both. The creation of the windows is discussed in Sections 2.5 and 2.6.

In each time-frequency window, the data and the synthetic are normalized by the energy of the signal inside the i th window,

$$d(t) \leftarrow \frac{d(t)}{\sqrt{\int_{t_i} d^2(t) dt}}, \quad u(\boldsymbol{\gamma}, t) \leftarrow \frac{u(\boldsymbol{\gamma}, t)}{\sqrt{\int_{t_i} u^2(t) dt}}. \quad (15)$$

This has the effect of making the misfits depend more on the waveform phase mismatch rather than the duration of the time window and removes sensitivity to amplitude differences resulting from uncertainties in attenuation structure or the source mechanism. To downweight the influence of large amplitude wave groups within a window, a weighting function $w_i(t)$ is introduced which weights the data by the inverse of the envelope, $e_i(t)$,

$$w_i(t) = \frac{1}{e_i(t) + \epsilon \cdot \max e_i(t)}, \quad (16)$$

where ϵ is a small number < 1 to prevent division by zero. We then normalize $w_i(t)$ such that $\int w_i(t) dt = 1$.

Depending on which time-frequency windows are added to the misfit equation, the fitting procedure may not be able to constrain all basis function coefficients, and we therefore regularize the inversion by imposing two additional constraints to the misfit equation. We penalize the norm of the basis function coefficients, i.e. $\frac{1}{MN} \sum_{ij} \gamma_{ij}^2$, in order to reduce the magnitude of γ_{ij} in cases where the data has no sensitivity. The factor of $1/MN$ normalizes the norm to one standard deviation in the model parameters. We also penalize the norm of simple first-order differences between adjacent γ_{ij} , that is $\sum_i \sum_{j=1}^{M-1} (\gamma_{ij} - \gamma_{i(j+1)})^2$, which effectively removes sharp gradients from the path-average model result unless required by the data. The total misfit function is therefore

$$F_{\text{total}}(\boldsymbol{\gamma}) = F(\boldsymbol{\gamma}) + \frac{\lambda_D^2}{MN} \sum_i \sum_j \gamma_{ij}^2 + \lambda_F^2 \sum_i \sum_{j=1}^{M-1} (\gamma_{ij} - \gamma_{i(j+1)})^2, \quad (17)$$

where λ_D and λ_F control the strength of the regularization for the model norm damping and first-order difference norm damping (flattening), respectively. To select appropriate values, we first set $\lambda_F = 0$ and we empirically choose a value for λ_D through trial and error by visually inspecting the path-average model results and selecting the value which gives the smallest coefficient norm that does not increase the waveform misfit significantly. Our experience has shown that one value is sufficient for a given data set, and in this study we set $\lambda_D = 1.5$ for all waveform inversions. After selecting λ_D , we set the value for λ_F after a similar process to be one-fourth of λ_D . Although we found these values in a rather arbitrary fashion to reduce computation time, this could be improved in the future by applying a more rigorous method that automatically balances the trade-off between the coefficient norm and waveform misfit for each individual waveform inversion.

Nonlinear optimization of eq. (17) is performed using the quasi-Newton Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (Nocedal & Wright 2006). Modrak & Tromp (2016) extensively tested a large-scale limited-memory version of the BFGS algorithm in a series of geophysical inversions and it was found to provide better performance over nonlinear conjugate gradient methods, which is similar to our experience.

2.5 Fundamental mode extraction, automatic windowing and fitting

In the first stage of waveform fitting, we attempt to fit only the fundamental mode Rayleigh and Love surface waves. Whether or not time-frequency windows will be created for all three waveform components depends on the radiation pattern of the seismic source and the quality of the data. The process described below is applied to each component separately, and at the end all windows that were created are gathered and the nonlinear optimization is applied.

2.5.1 Phase-matched filtering

The fundamental mode is extracted from the data in a series of operations that constitutes what we will refer to as a phase-matched filter. Dziewoński *et al.* (1972) may have been the first to apply the technique to the study of surface waves, although they termed the method ‘residual dispersion measurement’. Herrin & Goforth (1977) and Goforth & Herrin (1979) later applied phase-matched filters to the study of Rayleigh and Love waves, respectively, and they show through an iterative procedure that a phase-matched filter is able to retrieve fundamental mode surface waves in the presence of severe multipathing. The phase-matched filter technique has since been modified and incorporated into regular surface wave analysis (e.g. Russell *et al.* 1988; Levshin *et al.* 1989). For example, Ekström *et al.* (1997) applied phase-matched filtering to measure fundamental mode phase delays, and Levshin & Ritzwoller (2001) created an automated method to detect regional surface waves by constructing phase-matched filters from global maps of group velocity dispersion. Bensen *et al.* (2007) apply phase-matched filtering to increase the quality of group velocity dispersion curves measured from seismic ambient noise.

In our application, we use a reference 3-D model and information about the source from a published CMT solution (Ekström *et al.* 2012) to construct the spectrum of a synthetic reference fundamental mode waveform $P(\omega) = \exp \{i[k_0^0(\omega)\Delta + \phi_s(\omega)]\}$, where Δ is the epicentral distance, $k_0^0(\omega)$ is the path-averaged fundamental mode wavenumber calculated from the 3-D reference model, and $\phi_s(\omega)$ is the source phase calculated from the CMT solution. The next step in the phase-matched filter is to multiply the spectrum of the observed waveform by $P(\omega)$. Theoretically, this operation should ‘collapse’ the fundamental mode signal into a large amplitude delta function pulse in the time domain. However, since $k_0^0(\omega)$ will generally differ from the true wavenumber, the result is a short-duration signal. Any other signal not matching the dispersion characteristics of the fundamental mode will appear as noise.

Therefore, we calculate the envelope of the collapsed waveform, smooth it by using an 11-point moving average, and assume that the largest amplitude is the fundamental mode. Next, we apply a Gaussian window to the collapsed waveform centred on the maximum envelope value. The width of the Gaussian window is defined by finding where the envelope drops below the rms value of a 500 s region around the envelope maximum. The fundamental mode surface wave is finally extracted by multiplying the windowed collapsed waveform with the complex conjugate of $P(\omega)$. This process results in a signal where it is algorithmically simple to identify the fundamental mode and nearly guarantees that, at the very least, path model constraints from the fundamental mode will be retrieved. However, it is possible that higher-mode interference can subtly alter the extracted signal (e.g. Hariharan *et al.* 2022). Large amplitude noise may also be mistaken as the fundamental mode signal,

so some pre-processing must be done to ensure that the data has high enough signal quality before automatic fitting. For example, a simple method may be to bandpass filter the observed waveform such that the signal-to-noise ratio (SNR) rises to a sufficient level (e.g. see Section 4.1).

An example of phase-matched filtering is shown in Fig. 1. We calculate a synthetic transverse component waveform for a right-lateral strike-slip focal mechanism located at 30 km depth with a strike of 0°N. The receiver location is 16° to the west at an azimuth of 90°, which is a maximum of the Love wave fundamental mode radiation pattern. We used a modified PREM where we placed a +5 per cent Voigt V_S anomaly between the Moho and 74.4 km depth, or 50 km in thickness. Relative to PREM, this speeds up the Love wave fundamental mode group velocity to make it completely overlap the overtone group velocities between roughly 40–100 s period. To create the phase-matched filter, we use the unmodified PREM.

Fig. 1(a) shows the results for a noise-free waveform. Since the models used to create the phase-matched filter and the waveforms are different, the collapsed waveform shows some dispersion, but the extracted waveform closely matches the predicted waveform for the modified PREM. In Fig. 1(b), we add ground noise recorded at station IU.INCN on Oct. 8, 2021 to the synthetic waveform in Fig. 1(a). To demonstrate the ability of the phase-matched filter to extract the fundamental mode, we amplified the noise such that the SNR was ~ 5 . The extracted waveform also matches the predicted fundamental mode well, but small amplitude variations are visible due to noise. In Fig. 1(c), we add an interfering Love wave fundamental mode that travelled 17.5° along the same source–receiver path. This second fundamental mode arrival is visible in the collapsed waveform appearing with similar shape as the first arrival but with a smaller amplitude. Due to sufficient separation in time, the extracted waveform again matches the predicted fundamental mode. However, if interfering waves significantly overlap in time, the phase-matched filter is unable to resolve the desired fundamental mode.

2.5.2 Fundamental mode window creation and fitting

After extracting the fundamental mode, we next attempt to select a minimum fitting frequency that defines a set of overlapping Gaussian bandpass filters, $G(f) = \exp \{ -\alpha((f - f_0)/f_0)^2 \}$. Here, f_0 is the centre frequency of the filter, and α is a filter parameter chosen so that a filtered seismogram recorded at regional distances $< 50^\circ$ from the source will contain distinct wave groups (e.g. Lebedev *et al.* 2005). Time resolution trades off with the filter widths, and larger values for α will decrease the filter width and therefore create longer duration signals. After some trial and error, we found that the value for α does not need to be set with particular accuracy (e.g. Dziewoński *et al.* 1972; Levshin *et al.* 1989), and for this study it was sufficient to set a constant value of $\alpha = 12.5$. In the following, we will take filtering to mean the application of the Gaussian bandpass filter $G(f)$.

Since our eigenfunction calculations neglect the effects of gravity, we begin by setting an absolute minimum frequency of 3.33 mHz. Then, using the path-averaged reference phase velocity curve from the 3-D reference model, we find the lowest frequency such that at least three wavelengths fit between the source and receiver to enforce the far-field approximation. Next, we perform a series of quality control checks. We check the radiation pattern and avoid frequencies where the amplitude at the station azimuth drops below

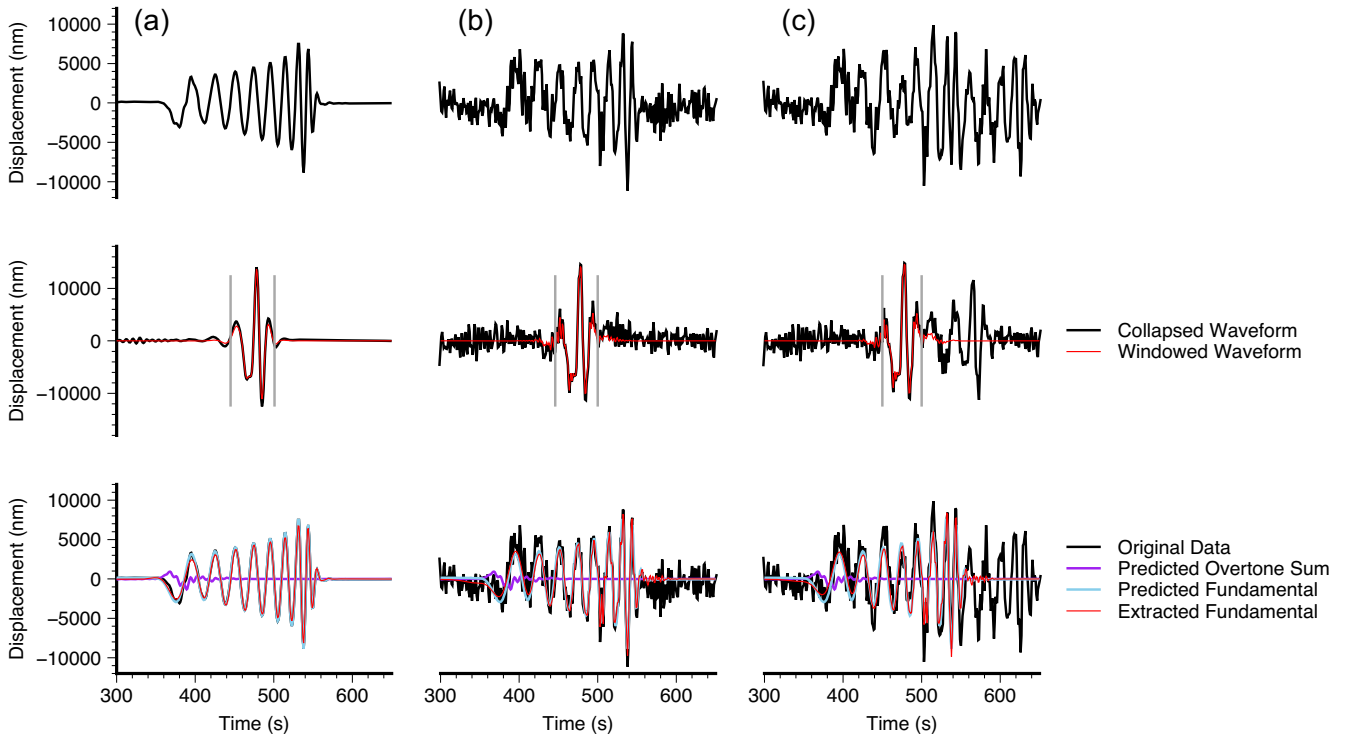


Figure 1. Three examples illustrating the phase-matched filtering process. Top row: original data. Middle row: collapsed waveforms in black after applying a phase-matched filter. The windowed waveform is shown in red after a Gaussian window is applied with a width shown by the vertical grey bars. Bottom row: extracted fundamental mode waveforms and comparison with the original data and predicted waveforms. Column (a) shows phase-matched filtering applied to a noise-free synthetic transverse component waveform calculated for a right-lateral strike-slip event at a depth of 30 km and a strike of 0°N . The model used was PREM with a 50 km thick +5 per cent Voigt V_S anomaly placed between 24.4 and 74.4 km depth, chosen so that the fundamental mode would overlap with the overtones. The epicentral distance between the source and receiver was 16° . Column (b) shows phase-matched filtering applied to the waveform in (a), but with real ground noise added. Column (c) shows phase-matched filtering applied to the waveform in (b), but an interfering Love wave fundamental mode is added. The interfering wave was calculated using the same source, model and take-off azimuth, but an epicentral distance of 17.5° was used.

50 per cent of the maximum (e.g. van der Lee *et al.* 2001; Rösler & van der Lee 2020). This criterion helps to avoid fitting scattered waves which are strongly sensitive to structure off the great-circle path (e.g. Meier *et al.* 1997; Lebedev *et al.* 2005) and near the source (e.g. van der Lee 1998). Stationary noise that is present in the observed waveform may be extracted along with the fundamental mode signal. To make sure that the extracted waveform is not significantly affected by noise (e.g. Fig. 1b), we check that the SNR is at least 10. We define the SNR as the ratio of the maximum amplitude of the filtered signal envelope to the rms of the pre- P -wave noise in the filtered original data. As a guard against a failed fundamental mode extraction, we check whether the arrival time of the filtered envelope maximum is within 20 per cent of the reference fundamental mode group arrival time. If any of these three checks fails, the minimum frequency is incremented a small amount and the process is repeated until a suitable minimum frequency is found. An upper limit is set on the minimum frequency by requiring that there are at most 12 wavelengths between the source and receiver, which prevents the phase from accumulating excessively and causing a cycle skip (Lebedev *et al.* 2005). If this limit is crossed, the waveform component under consideration will be discarded from the fundamental mode fit.

After the minimum frequency is found, the rest of the centre frequencies are calculated by requiring that the Gaussian filter for the subsequent centre frequency overlaps at 90 per cent of the filter amplitude. The maximum fundamental mode fitting frequency is set to 62.5 mHz. For each Gaussian filter, we calculate ‘left’ and

‘right’ frequency limits where the amplitude of the filter falls below 30 per cent of the maximum. If the ‘left’ limit is less than 3.33 mHz, that filter is removed. Likewise, if the ‘right’ limit is above 62.5 mHz, the filter is also removed. This step ensures that the frequency content of the signal energy being fit remains mostly within the minimum and maximum frequency limits.

The last step is to filter and window the extracted fundamental mode signal. Since the extracted signal usually only contains a single wave group, the waveforms can be windowed using a relatively simple algorithm. We first find the arrival time of the maximum envelope amplitude, and then we define the window boundaries to be where the envelope drops to a fourth of the maximum value. For each filtered window, we check the radiation pattern, the SNR, and the group arrival times as before. Additionally, we check that the amplitude ratio in each window between the synthetic and extracted fundamental modes is less than five, since errors in the published moment tensor solutions or unaccounted attenuation structure may result in significant differences in amplitude. A window failing any of these checks is removed. After defining all time windows, we collect the created windows for each waveform component. The nonlinear waveform fitting then proceeds starting with the lowest frequency window. After a successful fit, the next highest frequency is added to the global misfit function F_{total} (i.e. eq. (17)) until all fundamental mode windows are fit. An example three-component waveform fit is shown in Fig. 2 between station KS.DAG2 and an M_w 6.1 event occurring at 23:46:44 UTC on 28 August 2019. The initial model in this case produces a Rayleigh wave that is

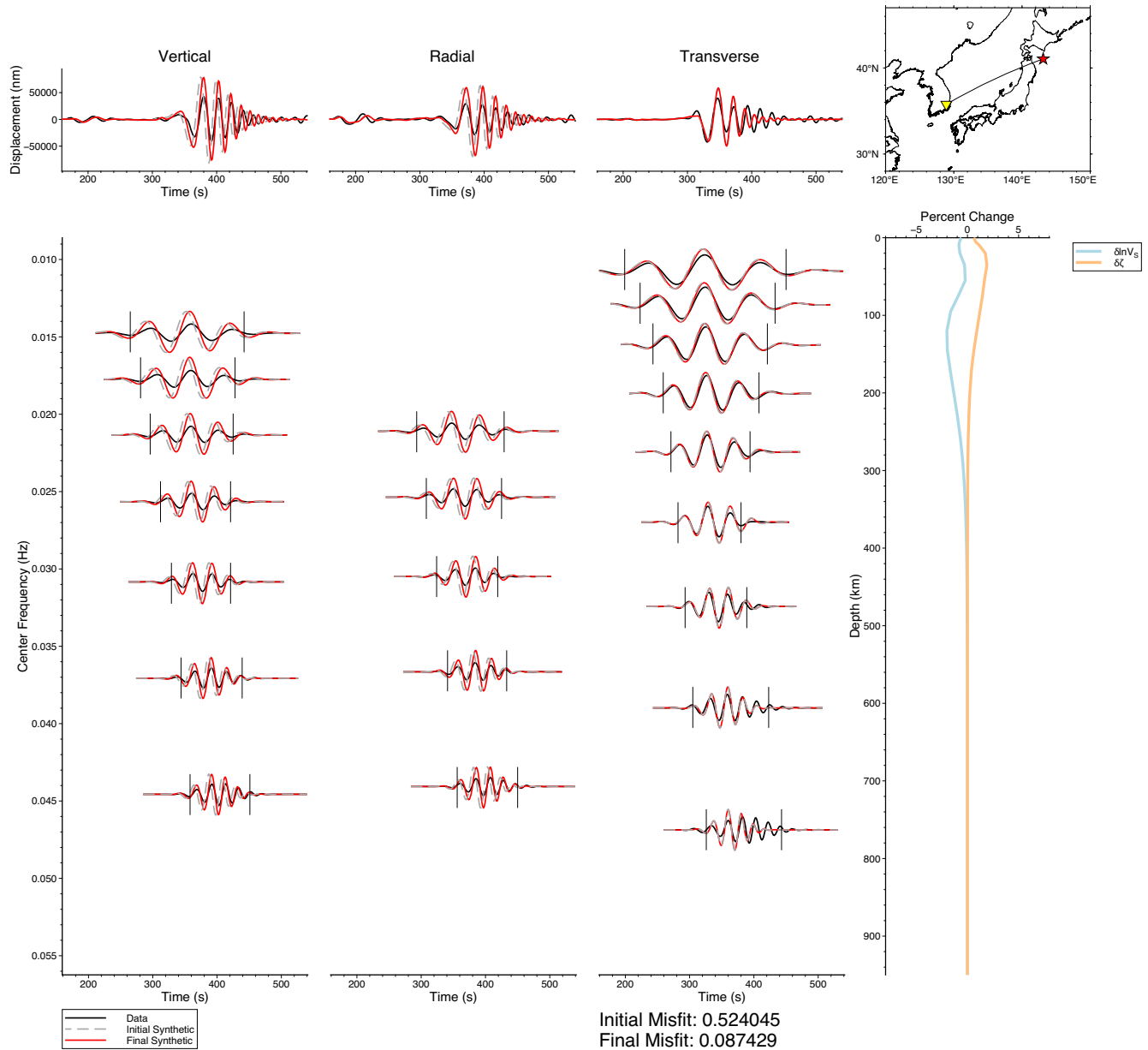


Figure 2. An example three-component fundamental waveform fit between station KS.DAG2 and an M_w 6.1 event occurring at 23:46:44 UTC on 28 August 2019. Top row shows the data in black, the initial waveform as a grey dashed line, and the final waveform in red for each component. The map in the upper right shows the locations of the event (red star) and station KS.DAG2 (yellow triangle). Below, the time-frequency windows are shown for each component, placed along the vertical axis according to the centre frequency of the Gaussian bandpass filter. For each time-frequency window, the data shows the fundamental mode after phase-matched filtering. Black vertical bars indicate the time window boundaries for fitting. The right shows the final path average model perturbations after fitting all fundamental mode windows. Initial and final misfits are given in the lower right.

slightly too fast compared to the data, while the initial Love wave fits the data fairly well. This causes the inversion to place a positive perturbation to radial anisotropy in the uppermost mantle, while a low V_S perturbation is found between 100–200 km depth.

2.6 Full synthetic waveform windowing and fitting

In the second stage of the fitting algorithm, we fit the full synthetic mode sum. If the fundamental mode fit was successful, then the resulting model perturbations are used in this stage as an initial model. We note that although we do not exclude the fundamental

mode from the data or the synthetics (e.g. as in the mode-stripping technique of van Heijst & Woodhouse (1997)), we reason that by fitting the fundamental mode in the first stage, the misfit should be dominated by the overtones in this stage. Likewise, as explained below, we also expand the fitting frequency of the overtones to 100 mHz, where larger differences in the overtone and fundamental mode group velocities typically allow the overtones to be completely separated from the fundamental mode in time. In cases where the fundamental mode is not excited well, such as for deep events, the initial perturbations are zero and waveform fitting proceeds as normal, but steps are taken to avoid fitting windows that may contain fundamental mode energy. The mode sum contains modes $n = 0$,

1, 2, ..., 20 in order to ensure that the S - and multiple S -waves are sufficiently recreated. Note that this does not imply that all modes are equally constrained during the waveform fit.

Window selection for the full mode sum includes the same selection criterion for the fundamental mode portion as in the previous section. The only difference is that here we are not fitting an extracted waveform from a phase-matched filter, and instead we directly fit the observed waveform. We will show in our synthetic testing results that this allows us to fit overtones even when they interfere with the fundamental mode, since any remaining misfit after fitting the fundamental mode may be explained by the presence of overtones.

We set the minimum fitting frequency to be the same as that found for the fundamental mode. If the fundamental mode fitting failed, for example as in the case of a deep event that does not adequately excite the fundamental mode, then a minimum fitting frequency is set as before by requiring that at least three fundamental mode wavelengths fit between the source and receiver. Subsequent filter centre frequencies are defined by requiring overlap at 90 per cent of the Gaussian filter amplitude, as before. However, for the full synthetic waveform, we allow the maximum fitting frequency to reach 100 mHz for windows that do not contain the fundamental mode.

Window selection for a filtered full synthetic broadly works by finding all envelope maximums, iterating through them, and then deciding on time window boundaries based on various criteria. To start the window selection, we set an early time limit to be 10 per cent before the S -wave arrival time if the distance is less than 35° , or 10 per cent before the SS -wave arrival otherwise. In this study we used *ak135* to predict body-wave arrival times, but in principle any model could be used for this purpose. The distance criterion is to avoid fitting waves that bottom in the lower mantle. If an envelope maximum within a time window occurs before this early time limit, it is discarded. We also discard envelope maximums that are within 5 per cent of the P -wave arrival (which may occur for shorter regional paths) to avoid fitting P -wave energy since we neglect P -wave velocity variations in this study. We set a late time limit to be 20 per cent past the fundamental mode arrival predicted for the centre frequency of the filter, and we discard any envelope maximums arriving past this point. With these initial checks, we can narrow our window search to a smaller set of envelope maximums that may contain our signals of interest.

Next, we construct time window boundaries around each remaining envelope maximum. We start at each maximum and expand the time boundaries to earlier and later times until we either hit the early or late time limits, or the envelope value drops to one fourth of the current maximum value under consideration. If the time boundary search cannot find a sufficient decrease in the envelope value and instead passes a greater maximum value, the current maximum value is updated and the one-fourth threshold is changed to reflect the new value in the search direction. If the initial fundamental mode fitting in the previous section failed, then the time window boundary is not expanded to include the fundamental mode.

Now we perform quality control on the created windows and merge any overlaps. For windows before the predicted fundamental mode arrival, we check the radiation pattern for the overtone sum (i.e. $n = 1, 2, \dots, 20$) and discard windows where the radiation pattern amplitude is less than 50 per cent of the maximum. Similarly, if the window contains the fundamental mode, then we also check the fundamental mode radiation pattern. Next we check the SNR of the windows. For windows only containing overtone signals, we check that the SNR is at least 5. We set this lower than the fundamental

mode SNR check of 10 since the overtones generally have lower amplitudes. To ensure that an overtone window contains overtone energy, we check that the synthetic SNR (i.e. the ratio of the synthetic overtone sum envelope maximum within the current window to the noise rms) is also at least 5. In order to avoid large errors in the moment tensor solution, we make sure that the amplitude ratios between the data and the synthetics for each window are not greater than 5. Finally, all windows are checked for overlaps and any overlapping windows are merged.

Fitting proceeds in the same way as for the fundamental mode. Fitting begins at the lowest frequency window, and subsequent windows are added to the total misfit function until all windows are fit. An example is shown in Fig. 3 for the same event and station pair as in Fig. 2. Here, since we use fundamental waveform fitting result as an initial model, the observed waveform and the initial synthetic match closely, particularly in the vertical and radial components. Therefore, additional perturbations to the path average model appear correlated in V_S and ζ in order to fit the overtones and fundamental mode on the transverse component.

2.7 Linear constraints and uncertainties

One strength of the PWI method over the method of Cara & L ev eque (1987) and Debayle (1999) is the ability to retrieve uncorrelated linear constraints (Nolet 1990). After finding an optimal model $\boldsymbol{\gamma}^{\text{opt}}$ that minimizes F_{total} , the Hessian matrix H of the total misfit can be diagonalized, i.e. $H = S\Lambda S^T$, to give linearly independent combinations of the path average model parameters that are constrained by the waveform fit. Here, S is the matrix of eigenvectors and Λ is the diagonal matrix of eigenvalues, and the linear constraints are given by $\boldsymbol{\eta} = S^T \boldsymbol{\gamma}^{\text{opt}}$. In the vicinity of $\boldsymbol{\gamma}^{\text{opt}}$, the second-order term in the Taylor expansion of F_{total} can be used to define a confidence region $\Delta\boldsymbol{\gamma} = \boldsymbol{\gamma} - \boldsymbol{\gamma}^{\text{opt}}$ such that $\frac{1}{2} \Delta\boldsymbol{\gamma}^T H \Delta\boldsymbol{\gamma} < \epsilon$, where ϵ is a small amount that F_{total} is allowed to deviate. After a suitable ϵ is found, the uncertainties in $\boldsymbol{\eta}$ are given by $\Delta\eta_i = \sqrt{2\epsilon/\lambda_i}$.

Nolet (1990) writes that the choice of ϵ is made by generating synthetics for small deviations $\Delta\boldsymbol{\gamma}$, observing the resulting waveform misfit, and then deciding which ϵ to use in an ad hoc manner. Here, we automate the process by creating a set of models $\boldsymbol{\gamma}^{(k)}$ using the linear constraints $\boldsymbol{\eta}$ and the eigenvectors S . Since H is a symmetric matrix, the eigenvectors are orthonormal, that is $S^T S = S S^T = I$, and $\boldsymbol{\gamma}^{\text{opt}} = S\boldsymbol{\eta}$. Then we define $\boldsymbol{\gamma}^{(k)}$ as

$$\gamma_i^{(k)} = \sum_{j=1}^k S_{ij} \eta_j, \quad (18)$$

where $k = 1, 2, \dots, MN, MN + 1$, and $\boldsymbol{\gamma}^{(k=MN+1)} = \boldsymbol{\gamma}^{\text{opt}}$. Note that the indices i and j used here do not correspond to the same indices as introduced in eq. (5). Instead, the basis function coefficient matrix in eq. (5) is converted into a 1-D vector, such that $i = 1, 2, \dots, M$ are the basis function coefficients for the first model parameter, $i = M + 1, M + 2, \dots, 2M$ are the basis function coefficients for the second model parameter, etc., and therefore the total number of coefficients for N model parameters and M depth knots is $MN + 1$, where the extra one is for the Moho depth parameter. Next we calculate a misfit value for each $\boldsymbol{\gamma}^{(k)}$, which represents a path average model constructed from a truncated set of linear constraints. We choose a cut-off k_{cut} either at the ‘corner’ of the misfit curve, or at 10 per cent above the minimum misfit value, whichever is lower. We find that this heuristic choice results in synthetic waveforms that are reasonably perturbed away from the optimal waveform fit.

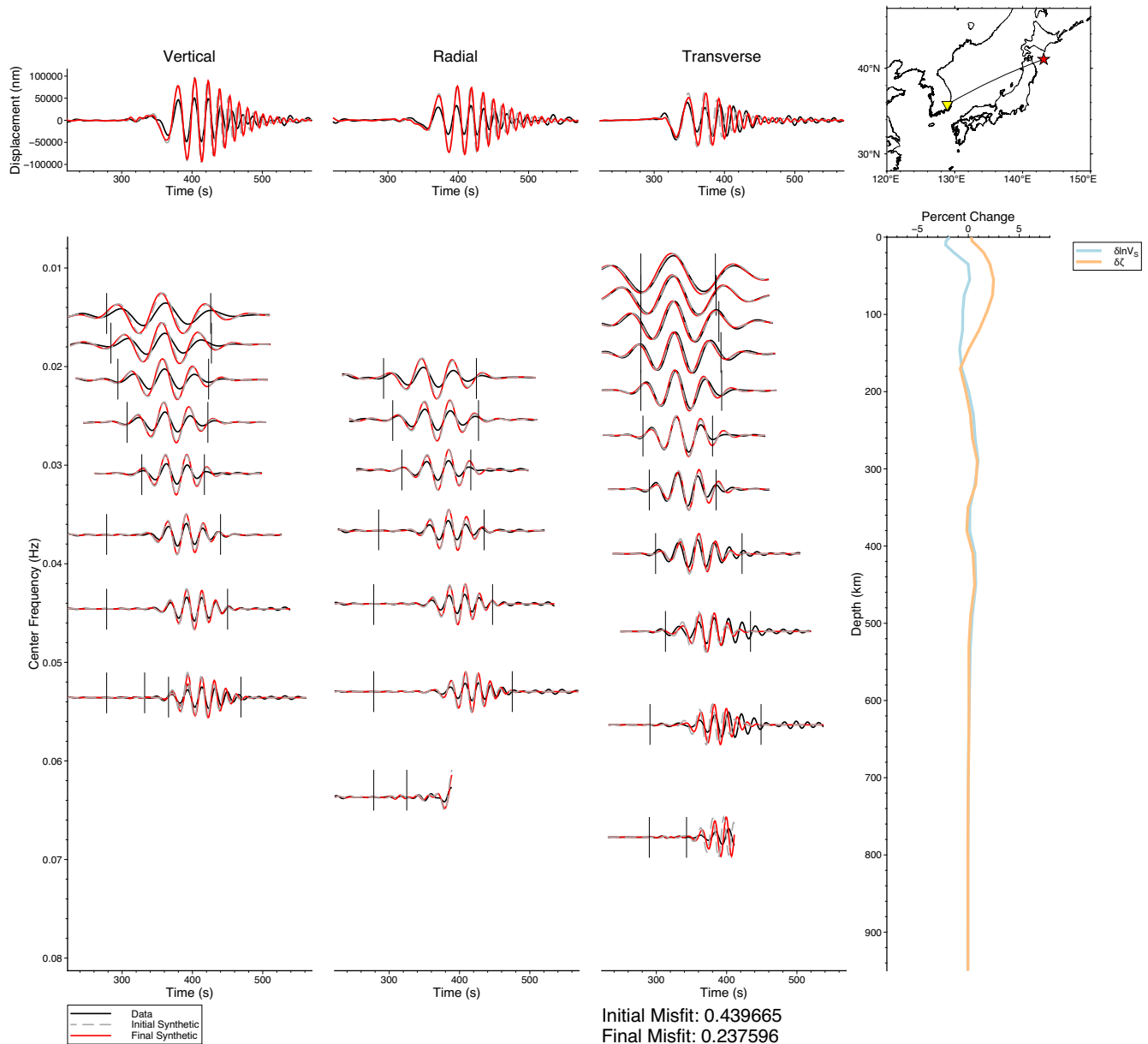


Figure 3. Same as for Fig. 2, except for the full synthetic waveform. Top row shows the data in black, the initial waveform as a grey dashed line, and the final waveform in red for each component. The map in the upper right shows the locations of the event (red star) and station KS.DAG2 (yellow triangle). Below, the time-frequency windows are shown for each component, placed along the vertical axis according to the centre frequency of the Gaussian bandpass filter. Black vertical bars indicate the time window boundaries for fitting. The right shows the final path average model perturbations after fitting all windows. Initial and final misfits are given in the lower right.

Using this cut-off allows us to define $\Delta \boldsymbol{y} = \boldsymbol{y}^{(k_{cut})} - \boldsymbol{y}^{opt}$ and to calculate the value of ϵ and therefore the uncertainties $\Delta \boldsymbol{\eta}$.

Fig. 4 shows one example determination of ϵ . In this example, we show the results of a waveform fit for a M_w 6 event that occurred on 3 April 2004 at 23:02:05 UTC and recorded at station IU.INCN. For this particular event, the station azimuth was in the nodal plane of the Love wave radiation pattern, and therefore only the vertical and radial components were fit. Since these components are mostly sensitive to perturbations in β_V , the V_S and ζ perturbations appear anti-correlated. Path average results such as this may bias the large-scale inversion for 3-D structure (see Section 2.8) if transverse component fits from other ray paths traversing a similar area are not available.

In Fig. 4(a), the misfit curve shows a rapid decrease in the misfit with the first few linear constraints followed by a long plateau. Then, another significant drop in the misfit occurs around $k = 30$, and our heuristic choice selected $k = 35$ for the cut-off point, which represents a misfit reduction of 87 per cent relative to the initial misfit. Fig. 4(b) shows that the remaining misfit lies in the higher frequency portion of the fundamental mode, which we deem to be reasonable since that part may be more affected by unmodelled effects such as scattering. The optimal path average model and the model constructed from the first 35 linear constraints are shown in Fig. 4(c), and the difference between these two models allows us to calculate ϵ and the uncertainties $\Delta \boldsymbol{\eta}$. We note that k_{cut} is only used in the determination of the uncertainties, and all linear constraints

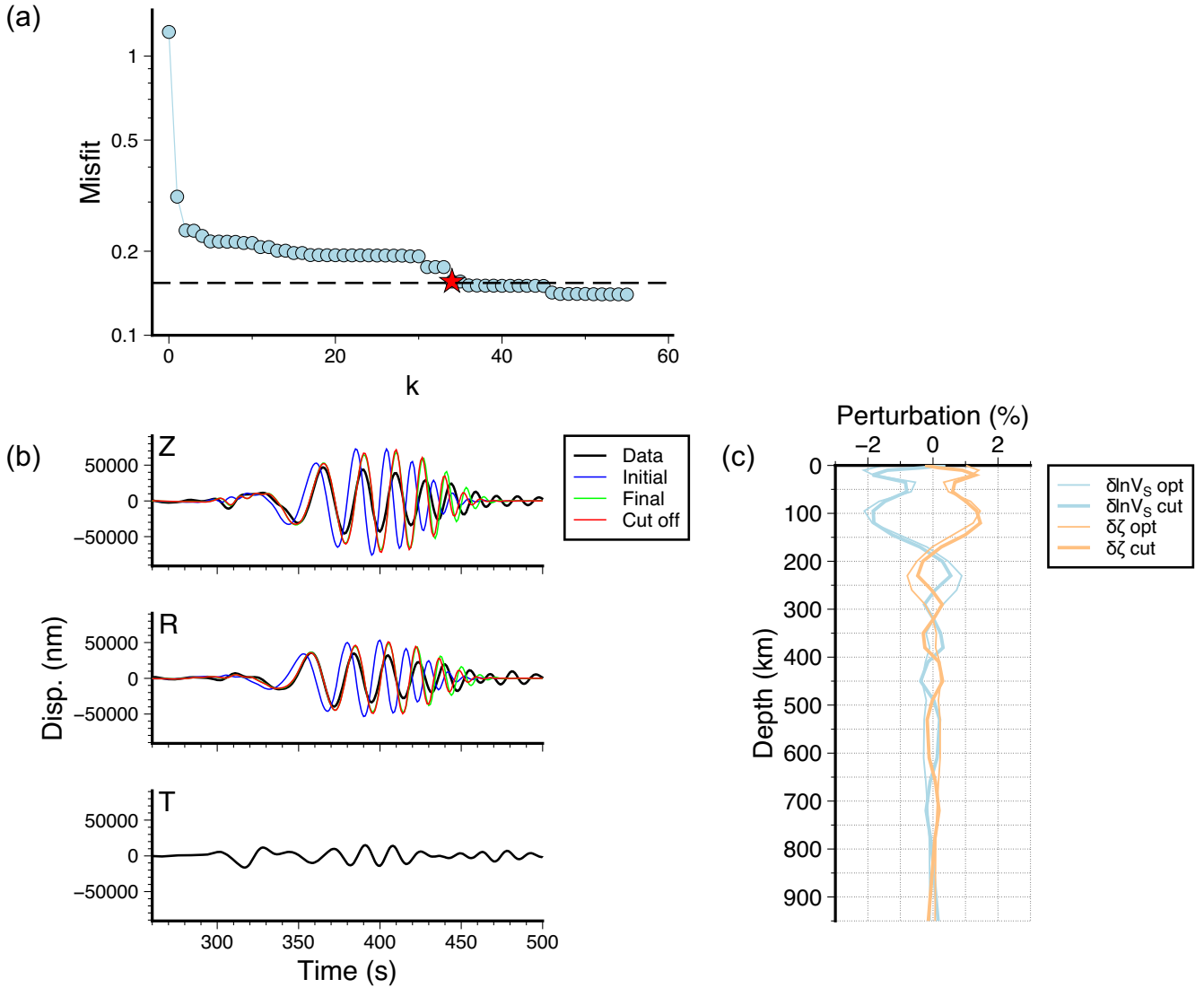


Figure 4. Example determination of ϵ . (a) Misfit as a function of the number of linear constraints used to create $\gamma^{(k)}$. To highlight relative changes in the misfit, a log scale is used on the vertical axis. Red star indicates the cut-off point at $k = 35$, and the dashed line indicates the 10 per cent level above the minimum misfit. (b) Three-component displacement waveform recorded at station IU.INCN. The waveforms are bandpassed between 12–80 mHz, which was the frequency range determined by the automatic windowing algorithm. Black lines indicate the observed waveforms, blue lines are the predicted waveforms from our 3-D reference model, green lines are the final waveforms after nonlinear optimization, and the red lines show waveforms for the model $\gamma^{(k=35)}$. (c) Path average models for the optimal fit (thin lines) and for the cut-off model $\gamma^{(k=35)}$ (thick lines). Isotropic V_S variations are shown in blue, and radially anisotropic variations are shown in orange. The difference between γ^{opt} and $\gamma^{(k=35)}$ determines ϵ and therefore the uncertainty in the linear constraints.

η associated with eigenvalues above a small threshold are used for the large-scale linear inversion.

2.8 Linear inversion

After applying the automatic waveform fitting procedure to all ray paths in the data set, we gather all independent constraints η into a single large data vector \mathbf{d} , and we create a linear constraints matrix G by projecting the eigenvectors of the Hessian from each waveform fit onto the 3-D model's basis functions. We then optimize a misfit equation defined as

$$S(\boldsymbol{\mu}) = (\mathbf{G}\boldsymbol{\mu} - \mathbf{d})^T C_e^{-1} (\mathbf{G}\boldsymbol{\mu} - \mathbf{d}) + \lambda_D^2 |\mathbf{I}\boldsymbol{\mu}|^2 + \lambda_F^2 |\mathbf{F}\boldsymbol{\mu}|^2, \quad (19)$$

where $\boldsymbol{\mu}$ is the vector of basis function coefficients and C_e is a diagonal data covariance matrix. We include model norm damping

and horizontal gradient norm damping (flattening), with \mathbf{I} being the identity matrix and \mathbf{F} representing the discrete horizontal gradient operator acting on $\boldsymbol{\mu}$. The strength of the damping is controlled by λ_D and λ_F . Optimal values may be calculated by estimating *a priori* covariance matrices for distributions of the model parameters and their gradients (e.g. Tarantola 2005), but for the purposes of this study we opt to perform a trade-off-curve analysis. In order to down-weight data in clusters, we count the number of nearby ray paths N_r for each waveform fit by searching within 2 per cent of the path length. The diagonal elements of C_e are thus set to $\Delta\eta^2 N_r$, where $\Delta\eta$ is an uncorrelated uncertainty resulting from diagonalizing each Hessian matrix. The misfit equation is optimized using the LSQR algorithm of Paige & Saunders (1982). After conducting an initial trade-off-curve analysis, we choose a model close to the point of maximum curvature and calculate a residual vector. We then identify data more than three standard deviations away from the mean

as potential outliers and remove them. Typically 2–3 per cent of the data are removed this way. We conduct another round of inversions to do a final trade-off-curve analysis and choose a model that is slightly rougher than the model at the point of maximum curvature.

3 SYNTHETIC TESTING

3.1 Radial anisotropy retrieval

To assess the ability of our method to extract information on radially anisotropic structure, we conduct a series of synthetic tests. We use an M_w 5.2, 45° dipping thrust fault moment tensor with a strike of 0°N. We place receivers spaced 2.5° apart at an azimuth of 45° and we set a maximum distance of 50°. This test set-up ensures that both Love and Rayleigh waves will be sufficiently excited to demonstrate the ability to fit three-component waveforms. The synthetic waveforms have the same ground noise added as in Fig. 1, except we do not amplify the noise in these tests. We conduct tests with events placed at 10, 50, and 300 km depth to observe the effects of varying fundamental mode and overtone excitations. We create a number of input models by adding ± 5 per cent ξ perturbations to isotropic PREM that are 200 km in thickness, starting at the Moho (i.e. 24.4 km depth) and then incremented by 100 km depth until the bottom of the anomaly reaches 424 km depth. For each test, we use isotropic PREM as the reference model. Since PREM contains an ocean layer and a relatively thin crust, the surface wave overtones will typically overlap with the fundamental mode. This series of tests will therefore determine the ability of our method to extract path average model information while dealing with the problem of overtone interference.

The results are shown in Fig. 5. Instead of showing the full seismogram, we separate the fundamental mode and the overtone sum to provide a direct comparison of the final waveform fits with the input model waveforms. In general, we find fairly good retrieval of the input anomalies. The poorest retrieval is found for short paths and deep perturbations. The requirement that at least three path lengths exist between the source and receiver puts a lower limit on the frequency range, and since lower frequency fundamental mode waveforms are generally sensitive to deeper structures, we lose resolution at depth. In Fig. 5(d), the input anomalies between 124.4 and 324.4 km depth do not significantly affect the fundamental mode waveform above ~ 20 mHz and the overtones are not significantly excited by an event at 50 km depth. However, in Fig. 5(g), placing an event at 300 km depth excites the overtones and allows us to sense a perturbation to ξ structure without fitting the fundamental mode. At an intermediate distance of $\sim 15^\circ$, we observe that the overtones completely interfere with the long-period fundamental mode in Figs 5(b), (e) and (h). We also observe overtone interference with the Love wave on the transverse component in nearly all tests. Despite this, the test results show that the inversion is able to fit both the fundamental mode and the overtones. We attribute this to the fact that we allow the overtones to be fit up to a frequency of 100 mHz, where greater differences between the overtone and fundamental mode group velocities exist. This causes the slower moving, high-frequency fundamental mode to be separated in time from the overtones, and therefore allows the overtones to be windowed and fit. In Figs 5(h) and (i), we show that we are able to retrieve deep anomalies by primarily fitting the overtone sum.

3.2 Effects of P -wave anisotropy

In this study we neglect the effects of P -wave anisotropy, and we explore potential biases in this section. In a redefinition of the

η parameter, Kawakatsu (2016) shows that fundamental mode and overtone Rayleigh waves are sensitive to shallow P -wave anisotropy anomalies compared to deeper sensitivity to other parameters at the same period. A major source of P -wave anisotropy in the upper mantle is the lattice preferred orientation (LPO) of olivine, which forms under simple shear during oceanic crust formation (e.g. Hess 1964; Karato *et al.* 2008). In our first test, we use elastic constants calculated from an average of 110 olivine aggregate samples to calculate $\phi = (\alpha_V/\alpha_H)^2 \approx 0.959$ (Ismail & Mainprice 1998). We then apply this as a -4 per cent ϕ perturbation to isotropic PREM in a 50 km thick layer underneath the Moho, which represents an approximate lithosphere thickness in the East Sea region (Kumar & Kawakatsu 2011). Since the volume of olivine is about 60 per cent in the upper mantle, the amplitude of -4 per cent is quite large in nature. Figs 6(a)–(c) show the results for a thrust fault moment tensor (e.g. Section 3.1) at 30 km depth at 10°, 20° and 30° distance, respectively. We find that the primary effect of the ϕ perturbation is to slow down the fundamental mode Rayleigh wave relative to the isotropic model. Since the Love waves are unaffected by P -wave anisotropy, this manifests itself as a $+1$ per cent ξ perturbation in the inversion results.

Now we test the effects of P -wave anisotropy in the crust. Amphibolites are considered to be a major constituent of the middle to lower crust (e.g. Christensen & Mooney 1995; Rudnick & Fountain 1995) and are known to form several types of LPO depending on the conditions under which deformation occurs (e.g. Ko & Jung 2015). In most regions of the crust, Type I amphibole is expected to form under conditions of low stress and low temperature, and a hornblende aggregate with Type I fabric reported by Ko & Jung (2015) may have $\phi = 0.81$ and $\xi = 1.18$. Therefore, amphibolite facies may cause a large fraction of the reported seismic anisotropy in the crust. The relative abundance of amphibolites may range from 35–40 per cent in the middle to lower crust (Christensen & Mooney 1995), so to simulate the effects of amphibole in an extended region of the middle to lower crust, we placed a -5 per cent ϕ perturbation from 15 to 24.4 km depth in isotropic PREM. The results are shown in Figs 6(d)–(f), using the same moment tensor and distances as before. The perturbation is placed at a shallower depth, which results in higher frequencies being slowed down in the fundamental mode Rayleigh wave compared to the input model. We find that a -5 per cent perturbation in ϕ can result in a $+2$ per cent ξ perturbation and a -1 per cent Voigt V_S perturbation at roughly the same depths as the ϕ perturbation.

Next, we test the effects of P -wave anisotropy due to aligned fractures or structures in the upper crust. Confining pressures at lower crustal depths close cracks, but in the upper crust fluid-filled cracks may align with the regional stress-field to cause P - and shear wave anisotropy (e.g. Anderson *et al.* 1974; Crampin 1987). As vertical dykes and vertically dipping faults are found throughout the East Sea (e.g. Kim *et al.* 2011), we place a $+5$ per cent ϕ perturbation in the top 8 km of isotropic PREM, and we use the same moment tensor and receiver locations as before. The results are shown in Figs 6(g)–(i). We find that the effects are very minor and leakage into ξ structure is <0.5 per cent, indicating that the frequency range that we fit is not sensitive to P -wave anisotropy in the upper crust.

These tests did not assume any scaling between ϕ and ξ , that is $\delta \ln \phi / \delta \ln \xi = 0$, while previous studies have used an anti-correlated scaling factor (e.g. $\delta \ln \phi / \delta \ln \xi = -1.5$) derived from petrologic constraints (e.g. Montagner & Anderson 1989; Panning & Romanowicz 2006; Beghein 2010). In order to see what potential bias may occur due to ignoring ϕ - ξ scaling, we repeat the previous tests but set the

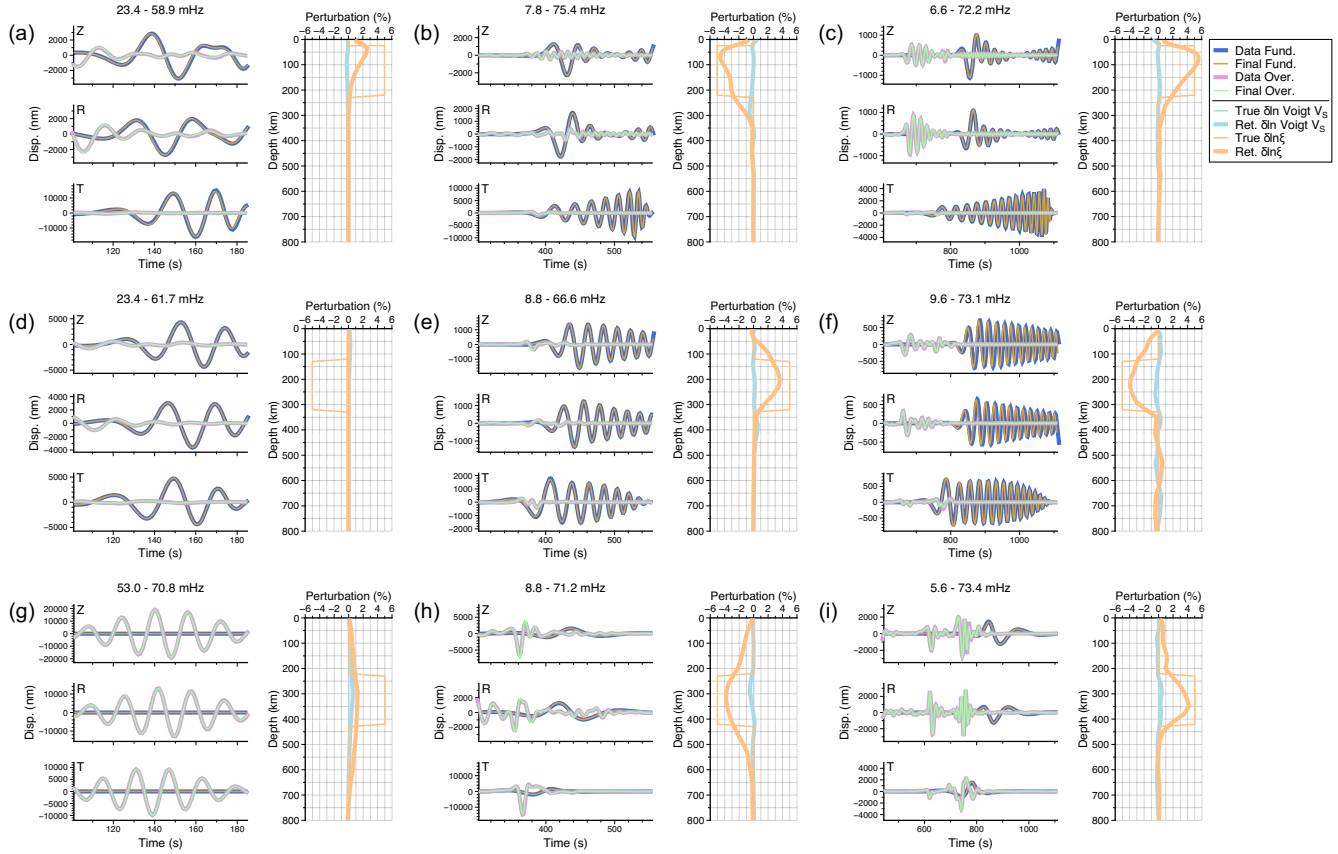


Figure 5. Results of 1-D synthetic testing to retrieve path-average radially anisotropic model perturbations. Top row shows results for an event at 10 km depth and ± 5 per cent ξ perturbations placed between 24.4 and 224.4 km depth, middle row shows results for 50 km depth and perturbations between 124.4 and 324.4 km depth, and the bottom row shows results for 300 km depth and perturbations between 224.4 and 424.4 km depth. Leftmost column is for a station at $\sim 5^\circ$ distance, middle column shows the station at $\sim 15^\circ$ distance, and the rightmost column shows results for $\sim 30^\circ$ distance. Above each subpanel is the maximum frequency range that was used for each inversion. Each subpanel shows a comparison between the final three-component waveform after inversion and the input model waveform. Waveform amplitudes are in nanometres of displacement, and note the changes in scale for each waveform. To the right of each waveform comparison is a comparison between the input model perturbations and the inversion result.

scaling factor to be either -1.5 or 1.5 , and the results are shown in Supporting Information Figs S1 and S2, respectively. When an anti-correlated scaling is used, the largest effect is to decrease the apparent ξ at shallow depths, while apparent ξ in the mantle is not significantly affected. On the other hand, if we assume that ξ and ϕ perturbations are positively correlated, then we find that the apparent ξ shifts to shallower regions relative to the true locations of the ϕ anomalies. Since Rayleigh waves are sensitive to shallow P -wave anisotropy (Kawakatsu 2016), using a positive scaling factor can cause the amplitude of the kernel sum $K_\xi + \frac{\delta \ln \phi}{\delta \ln \xi} K_\phi$ to become positive at shallow depths. Because a negative perturbation to ϕ decreases β_V , a positive kernel sum means that a negative Rayleigh wave phase velocity anomaly can be resolved with a positive ξ perturbation at shallow depths as well, and this is what we observe in Supporting Information Fig. S2.

Our final series of tests attempts to see whether ξ anomalies can be retrieved in the presence of ϕ anomalies. The previous tests consider only ϕ anomalies in isolation, but the reason scaling is invoked in the first place is that it may be more natural to expect ϕ and ξ anomalies occurring together. In Supporting Information Figs S3–S5, we consider coupled ξ anomalies in the mantle as in our previous test where we placed a 50-km-thick -4 per cent ϕ anomaly below the Moho. For each figure, subpanels (d)–(f) show the anti-correlated case, and subpanels (g)–(i) show the correlated

case. Supporting Information Fig. S3 shows results when no scaling is used during waveform inversion, Supporting Information Fig. S4 shows results with $\delta \ln \phi / \delta \ln \xi = -1.5$ used during the inversion, and Supporting Information Fig. S5 shows results with $\delta \ln \phi / \delta \ln \xi = 1.5$ used during the inversion. Our findings mimic those previously shown. Using an anti-correlated scaling factor reduces the apparent ξ in shallow regions, while a correlated scaling factor increases it. Using the anti-correlated scaling factor does not seem to significantly affect the amplitude of the retrieved $+\xi$ anomaly. The only case where a correlated scaling factor seems to help is when the true anomalies are also positively correlated, which is expected. However, when no scaling factor is used, the $-\xi$ anomaly is somewhat better retrieved than when an anti-correlated scaling factor is used, which may be important in regions where compositional changes may affect P -wave anisotropy without changing S -wave anisotropy (e.g. Mainprice *et al.* 2000; Beghein 2010).

If we assume that an anticorrelation dominates between ξ and ϕ , the results in this section imply that the main effect of neglecting P -wave anisotropy is to increase the amplitude of the retrieved ξ structure; a negative ϕ perturbation will result in a positive perturbation to ξ , and vice versa. Although leakage to the crust is minimized when a scaling factor of -1.5 is used, leakage to depth ranges outside of the perturbation is not significant when no scaling is used, and so we do not expect to have significant bias from leakage

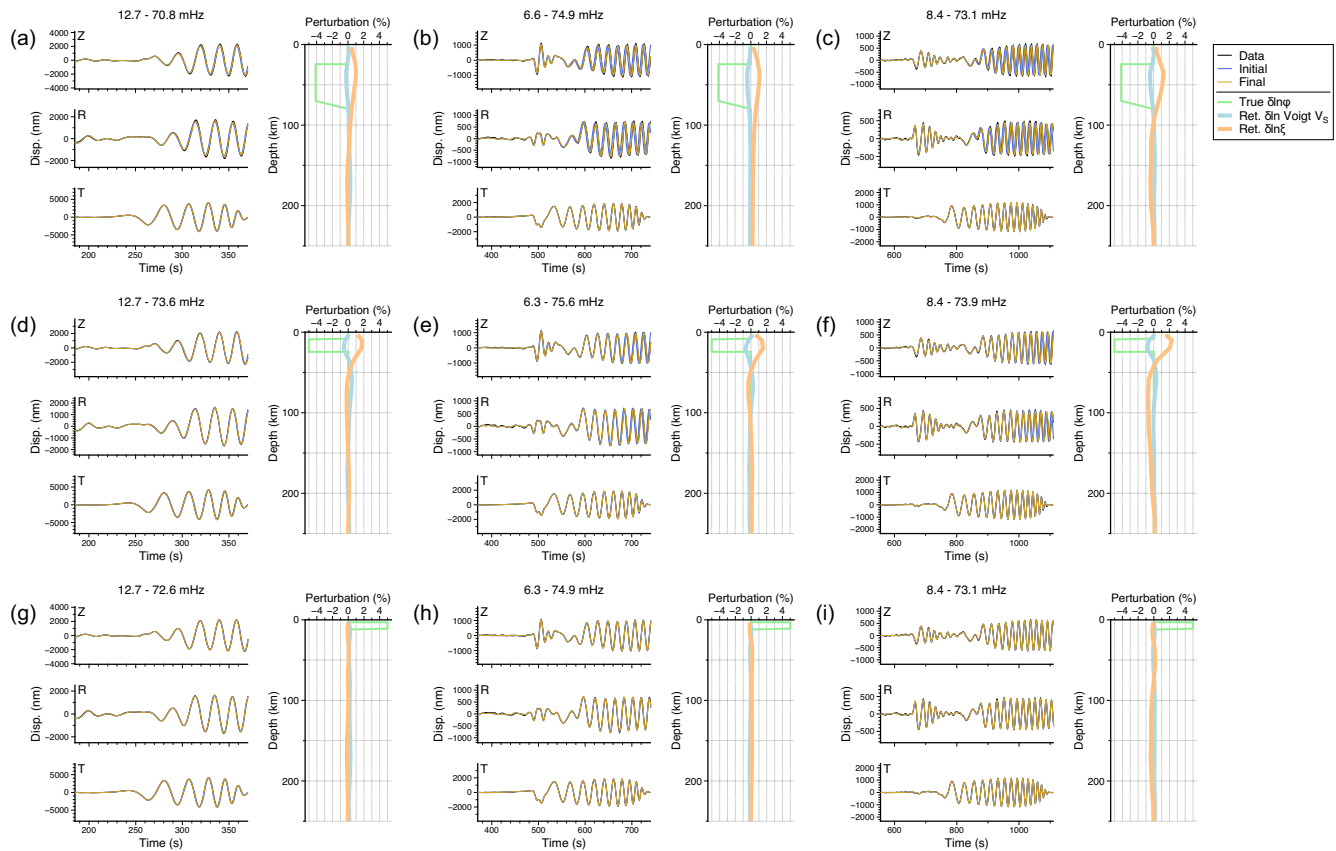


Figure 6. Results of 1-D synthetic testing on the effects of neglecting ϕ . (a)–(c) Results for a 50 km thick -4 per cent ϕ perturbation that may result from olivine LPO arising from the formation of oceanic crust. (d)–(f) Results for a -5 per cent ϕ perturbation in the lower crust of isotropic PREM simulating radial anisotropy due to extended amphibolites. (g)–(i) Results for a $+5$ per cent ϕ perturbation in the upper crust that may occur in regions of vertical faulting or volcanic dykes. The waveforms presented here are all calculated for an M_w 5.2 thrust fault event located at 30 km depth. Left column shows results for a station located at 10° distance, middle column shows 20° , and the right column shows results for 30° . Above each subpanel is the frequency range that was determined by the waveform fitting algorithm. Each subpanel shows the three-component displacement waveform at left. The ‘data,’ calculated using the true model, are shown in black, the initial waveforms are shown in blue, and the final waveforms after fitting are shown in orange. At right are the true and retrieved model perturbations. The true ϕ perturbation is shown in green, while the retrieved Voigt V_S and ξ perturbations are shown in blue and orange, respectively.

of crustal ϕ structure into the mantle. Our main concern is therefore the contribution of olivine P -wave anisotropy into ξ . Different olivine fabrics can form under varying stress, temperature, and water content conditions (Karato *et al.* 2008), and a database of natural olivine fabrics by Michibayashi *et al.* (2016) shows that peridotite samples from NE Japan show mainly A-, D-, and E-type olivine. All three types will produce maximum P -wave speeds roughly in the direction of shear (Michibayashi *et al.* 2016), and therefore we may expect our retrieved ξ model to have a slight amplitude bias towards higher values wherever radial anisotropy exists.

4 APPLICATION TO SEISMIC DATA

4.1 Data

In this study, we apply the method to waveforms generated by earthquakes occurring in the East Sea (Sea of Japan) region (Fig. 7). We gather data recorded by regional (Korea Meteorological Administration (KMA), Korea Institute of Geoscience and Mineral Resources (KIGAM), Full Range Seismic Network of Japan (F-net), Northeast China Extended Seismic Array (NECESSArray), Japan Meteorological Agency (JMA), China Earthquake Administration

(CEA), New China Digital Seismograph Network (NCDSN), Ocean Hemisphere Network (OHP)) and global [French Global Network of Seismic Stations (GEOSCOPE), Global Seismograph Network (GSN)] seismic networks during years 2004–2019. We limit the earthquakes we use to $5 < M_w < 8$ to ensure a good SNR at the low end and to avoid problems with source finiteness at the high end. Our initial data set consisted of 1384 earthquakes and 337 stations.

Before waveform fitting, we remove the instrument response for each waveform and change the units to displacement. We then apply some pre-processing to ensure the input waveforms will have a high enough SNR. This entails iteratively applying a high-pass filter, which begins with a corner period of 300 s. The corner period is increased a small amount until the SNR of the waveform reaches a value of at least 7. After applying this process to all waveforms, the total count was 16 362 vertical component waveforms, 15 552 radial component waveforms, and 16 097 transverse component waveforms. Greater noise levels generally result in a lower amount of waveform fits for the horizontal components (Wolin *et al.* 2015). After applying our method, we successfully fit 10 017 vertical component waveforms, 9245 radial component waveforms, and 8688 transverse component waveforms, resulting in a total of 12 635 linearly independent path constraints. Of these path constraints, 3937 ray paths contained only Rayleigh wave constraints,

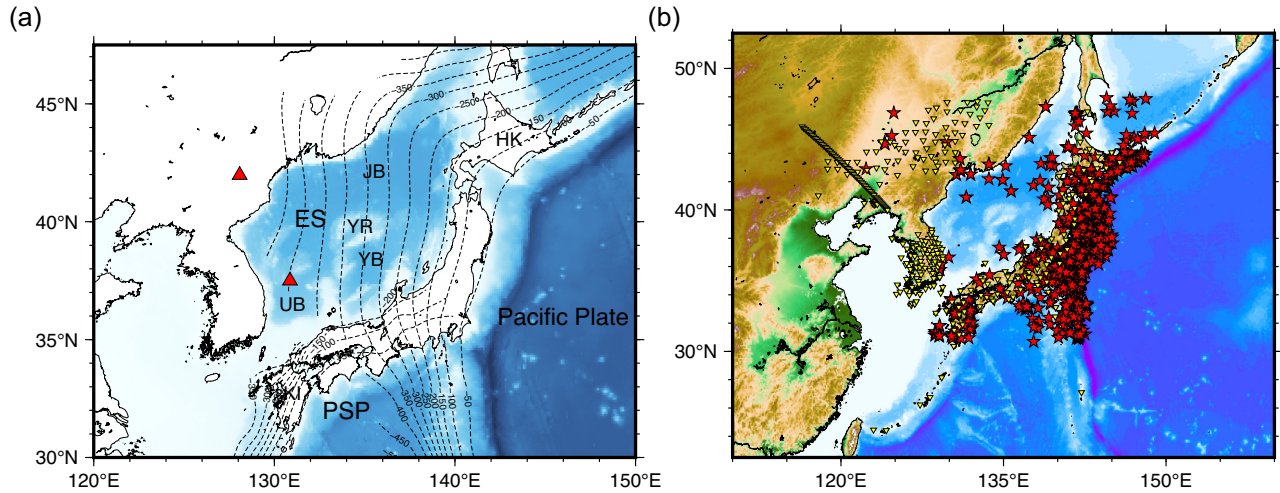


Figure 7. (a) Bathymetry of the East Sea and major geologic locations. Contours show depths to the tops of slabs (Hayes *et al.* 2018). Red triangles show the locations of intraplate volcanoes Mt. Baekdu (north) and Ulleung-do (south). ES: East Sea, HK: Hokkaido, JB: Japan Basin, KI: Kyushu Island, YB: Yamato Basin, YR: Yamato Rise, UB: Ulleung Basin, PSP: Philippine Sea Plate. (b) Distribution of stations (yellow triangles) and earthquakes (red stars) used in this study.

1976 ray paths contained only Love wave constraints, and the remaining 6722 ray paths contained constraints from simultaneously fitting both Rayleigh and Love waves.

In Supporting Information Fig. S6, we show the distribution of centre frequencies used for waveform fitting as a function of ray path length, and we can clearly see the effects of data culling based on the three-wavelength requirement at the low-frequency end, while at the high-frequency end we see the effects of the maximum frequency limits for the fundamental mode and full synthetic fits. We also notice that a majority of the centre frequencies appear to fall in discrete bins, which is a result of our pre-processing method described above that ensures that the waveforms have adequate SNR before fitting. Centre frequencies falling outside of the discrete bins are a result of waveform data lacking adequate SNR at lower frequencies, thus necessitating higher starting frequencies.

Relative to the maximum possible amount of data, the quality control steps taken during waveform fitting resulted in a rejection rate of 39, 41 and 46 per cent for the vertical, radial and transverse component data, respectively. In comparison with other studies, this is lower than the ~ 67 per cent rejection rate for the vertical component after pre-processing reported by Debayle *et al.* (2016) and the ~ 80 per cent rejection rate reported by Schaeffer & Lebedev (2013). Those studies, however, began with a much larger global data set of seismic waveforms recorded for all events listed in the CMT catalogue since 1976. As a result, most waveforms not fit by those methods are for low SNR recordings from low-magnitude events, while the shorter scale of our regional study helps to ensure that more low-magnitude waveforms will be kept.

Finally, we collected all path model constraints resulting from the nonlinear waveform fitting and applied the linear inversion method outlined in Section 2.8. The final model had a variance reduction of 68.8 per cent with respect to the 3-D reference model.

4.2 Phase velocities

We can use the result of each waveform fit to calculate perturbations to the 3-D reference model phase velocities using the path-averaged sensitivity kernels in eq. (7) and the fact that $\delta k/k = -\delta c/c$. The path

average phase velocities that result could potentially be used in a variety of other imaging studies, but here we restrict our discussion to the observed variations for each mode branch. We show a comparison between the initial and final phase velocities for the Rayleigh and Love waves in Fig. 8, where the velocities are placed into phase velocity and frequency bins. The fundamental mode and overtone branches $n = 1 \dots 20$ are clearly observed, and a horizontal line at $\sim 8.5 \text{ km s}^{-1}$ related to the Stonely modes (Dahlen & Tromp 1998) is also recovered in the Rayleigh wave phase velocities. Variations in the fundamental mode phase velocities are strongest at higher frequencies that are sensitive to shallow, more heterogeneous structure. Compared to the Rayleigh waves, more variation is seen in the Love wave phase velocities, which may be an indication of radial anisotropy in the region more strongly affecting β_H structure compared to β_V . Both Rayleigh and Love wave overtones show an increase in the amount of variation after fitting, showing that body waves sensitive to deep structure are being fit. Vertical lines seen at roughly 42, 63 and 83 mHz are an artefact due to differing frequency spectrum sampling rates for each waveform fit; waveforms for shorter ray paths are shorter in duration and therefore may have a coarser frequency sampling to calculate the frequency spectrum in eq. (1). Therefore, some bins may be overrepresented due to phase velocities with frequencies lying close together.

4.3 Resolution testing

To check the resolution and reliability of our model, we perform sparse checkerboard resolution testing. Normal (dense) checkerboard tests may give misleading results due to the superposition of smeared anomalies in close proximity (Rawlinson & Spakman 2016). We create input models with 3-D 5 per cent sinusoidal variations in V_S and ζ (i.e. eqs 8 and 9). The synthetic data are created by taking the product of the linear constraints matrix G and the input model. To check leakage of isotropic V_S structure into anisotropic structure, we create input models that only have V_S perturbations and zero ζ perturbations. We perform the opposite test to check leakage of anisotropic structure into V_S structure, and we also perform tests where perturbations in both model parameters are present. For all resolution tests, we add Gaussian noise to the synthetic data with a

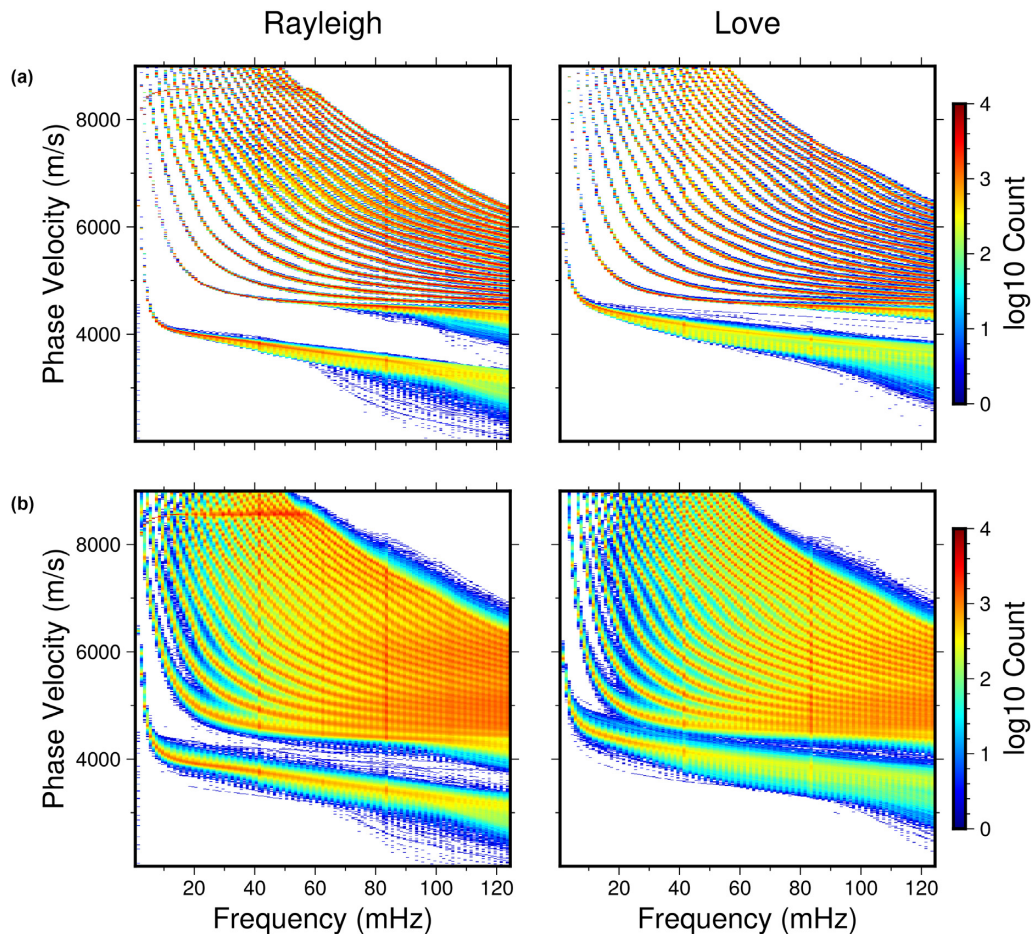


Figure 8. 2-D histograms of the (a) initial phase velocities for the 3-D reference model and the (b) final phase velocities after waveform fitting. Colours indicate amount of measurements in a bin, shown on a log scale. Colour scale is chosen to highlight the average phase velocities, with red showing bins with the largest amount of measurements.

noise level proportional to the ratio of the observed data norm and the norm of the uncorrelated uncertainties.

Results of the checkerboard resolution tests are shown in Fig. 9. We show the ability of the model to resolve 1.5° anomalies in Figs 9(a) and (c). The anomalies have a thickness of 50 km and their centres are placed at 60 and 260 km. We find that leakage effects are minimal and generally less than 1 per cent. Anomaly amplitudes are best retrieved in the southern East Sea and around Japan, while the northwestern portion of the study region typically shows poorer anomaly retrieval and smearing in an NW–SE direction. Amplitude retrieval for ζ is typically lower compared to V_S , and smearing effects are larger. Next, we show 3° anomalies centred at 300 km depth with additional anomalies separated by intervals of 150 km in Figs 9(b) and (d). These anomalies have a thickness of 200 km. Anomaly amplitudes are generally well retrieved in the south around the Japanese island. Similar to before, V_S anomalies are better retrieved compared to ζ , and we find more severe smearing effects in the northwest. At greater depths, we begin to completely lose resolution in the north.

Next we perform a structural resolution test where we set the input model to be the full waveform inversion model FWEA18 (Tao *et al.* 2018) by converting it to per cent perturbations with respect to our reference 3-D model. We calculate synthetic data using the linear coefficient matrix G (i.e. Section 2.8), and we add Gaussian noise as before. The results of the inversion are shown in Fig. 10. We find

that the subducting Pacific slab from the input model is retrieved well down to 200 km depth. At greater depths, however, the input amplitudes become reduced, and while we can retrieve similar slab contours, the inversion test reveals a large amount of lateral smearing in the north that would make it more difficult to distinguish the slab in a real data inversion. Small-scale anomalies, such the positive V_S anomalies seen in the input model at 200, 300, and 400 km merge together or are overwhelmed by the surrounding larger-scale anomalies. Similar results are seen for ζ . At 100 km depth we retrieve the large-scale positive anisotropy in the input model, but small-scale isotropic regions disappear and the anisotropy becomes slightly positive. At 200 km depth, we generally retrieve the pattern of negative and positive anisotropy, but again we see smearing in the north that removes the negative anisotropy there, and we lose the continuous structure linking the negative anisotropy in the Japan Basin to the Korea Peninsula. Below 200 km depth FWEA18 becomes isotropic, which allows us to observe leakage of isotropic V_S into ζ , and we find that the maximum level of leakage is on the order of 1 per cent, which is similar to the checkerboard resolution results.

Our synthetic waveform inversion tests (i.e. Section 3.1) show that deep anomalies are better retrieved when longer paths and deep events are used, while shorter paths (which are limited to higher frequencies during waveform fitting) provide constraints for shallow anomalies. Fig. 11(a) shows the path length distribution

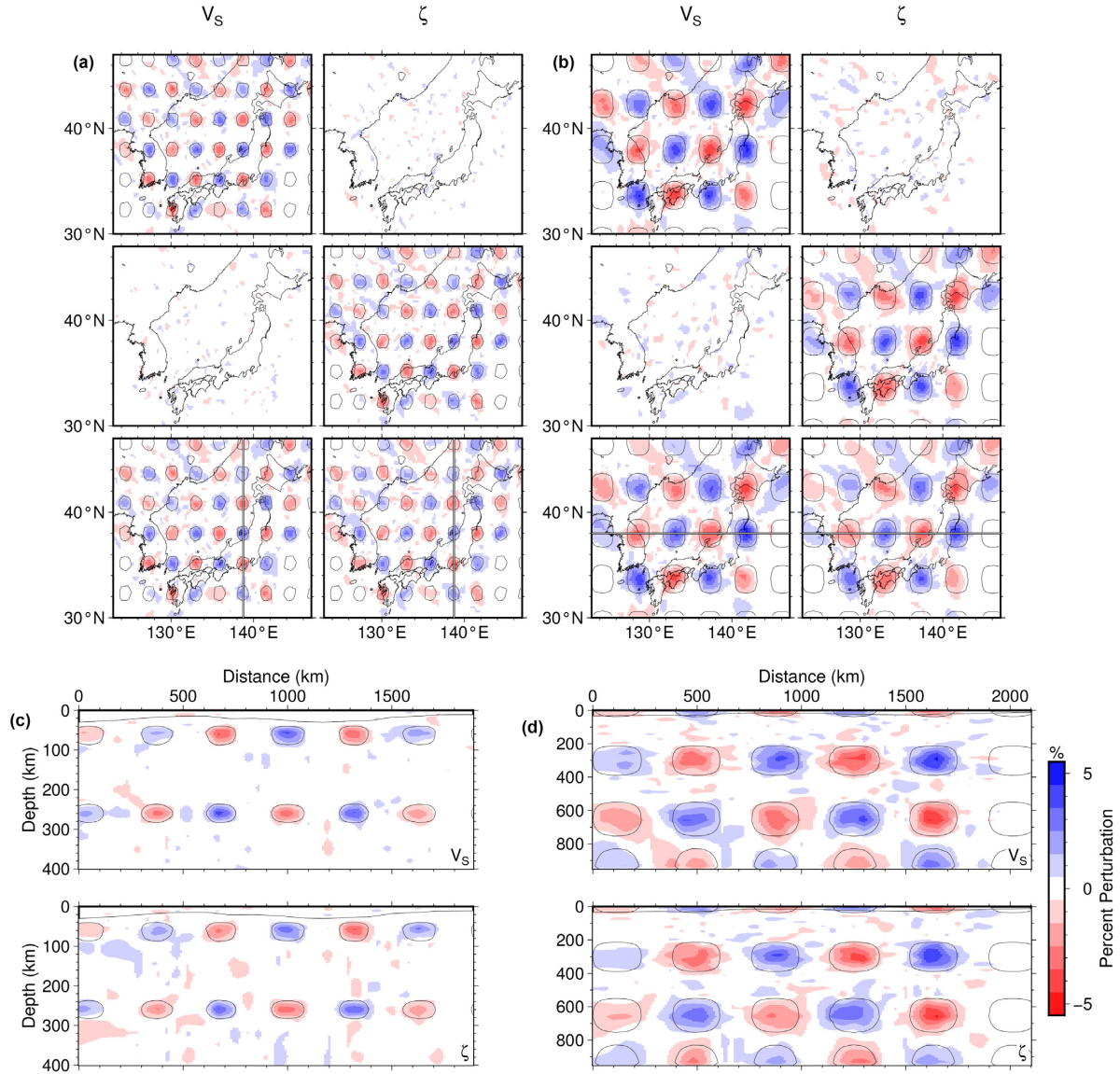


Figure 9. Sparse checkerboard resolution test results. (a) Results for 1.5° anomalies placed at 60 km depth with 50 km thickness and 150 km separation. Left column shows V_s anomalies, and the right column shows ζ anomalies. Contours in each subfigure indicate the ± 1 per cent level for the input model anomalies. Top row shows results for the test with only V_s input anomalies, middle row shows only ζ input anomalies, and the bottom row shows results with both anomalies present in the input model. Vertical lines in the bottom row indicate the location of the vertical slice in (c). (b) Same as for (a), except for 3° anomalies placed at 300 km depth with 200 km thickness and 150 km separation. Horizontal lines in the bottom row indicate the location of the vertical slice in (d). (c) Vertical slice through the output model shown in the bottom row of (a). The origin is located at the north and distances increase going south. (d) Vertical slice through the output model shown in the bottom row of (b), and the origin is located at west with distance increasing going east. Note the depth scale change from (c). Anomalies appear at 950 km depth due to the inability of our model depth parametrization (i.e. knot locations) to distinguish anomalies with a thickness of 200 km. The colour scale was chosen to only show anomalies with amplitudes above 0.5 per cent.

for this study, and Figs 11(b) and (c) show hit counts for ray paths under and over 10° length, respectively. The concentration of short paths around Japan is not surprising, but it helps explain the shallow checkerboard test results in Figs 9(a) and (c). Longer paths greater than 10° also tend to concentrate in NE Japan, but we can see that the longer paths traverse the entire region which explains why we have good depth resolution in the resolution tests.

Additionally, we show the azimuthal distribution of the data for different grid sizes in Fig. 12. Since we generally expect lower anisotropy resolution, checking the azimuthal distributions for different size grids allows us to see the length scales at which the distributions become uniform. We neglect the effects of azimuthal

anisotropy in this study, so that robust retrieval of radial anisotropy relies on the ray paths in the data set having uniform azimuthal distributions, thereby averaging out azimuthally anisotropic structure (e.g. Montagner & Nataf 1988). Rayleigh and Love waves are sensitive to different kinds of azimuthal anisotropy (e.g. Larson *et al.* 1998; Sieminski *et al.* 2007), and therefore we extract ray paths that contain path constraints from the vertical and radial components and the transverse component separately. The azimuthal distributions are quantified using a function proposed by Barmin *et al.* (2001),

$$\chi = \frac{\frac{1}{n} \sum_{i=1}^n f_i}{\max_i f_i}. \quad (20)$$

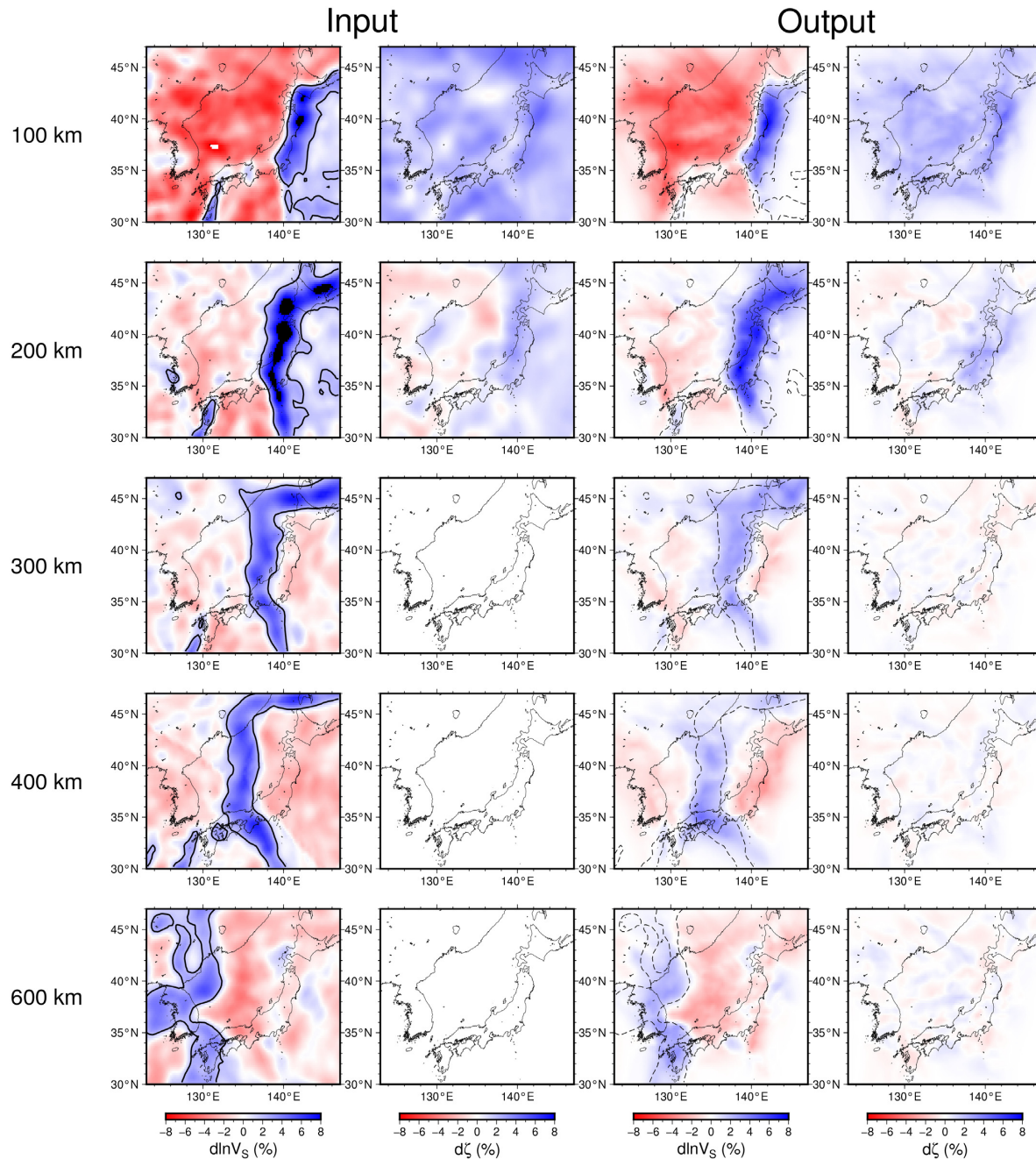


Figure 10. Results from the structural resolution test. Left two columns show the input model, and the contours indicate +2 per cent V_S anomalies. The right two columns show the output model, with light dashed contours indicating the +2 per cent level of the input model. The depth for each row is indicated to the left.

Here, f_i is the number of paths falling into the i th azimuth bin, and n is the total number of bins. When the azimuthal distribution at that grid node is uniform, then the f_i will be roughly equal, so that $\chi \approx 1$. However, when the distribution is dominated by a small range of azimuths, then most f_i will be close to zero and the sum will be approximately equal to the maximum f_i , so that $\chi \approx 1/n$. We set the number of bins to $n = 10$, and we calculate χ for grids with average spacings of 1° , 2° , and 3° . We find that the azimuthal distributions tend to uniformity around Japan and the southern East Sea, which may explain why the highest ξ resolution occurs in Japan in Fig. 9. As the grid spacing increases, we expect to see more uniform azimuthal distributions across the study region, and

this is what we observe in Fig. 12. Therefore, we may expect the most robust retrieval of radial anisotropy around Japan and the southern East Sea, but robust retrieval in the rest of the study region is limited to larger wavelength structures.

However, previous studies using less data over a wider region have shown that the biasing effects of azimuthal anisotropy may be insignificant. For example, Marone *et al.* (2004) conducted a traditional PWI inversion in the Eurasia-Africa plate boundary region, and they found that the potential bias from observed azimuthal anisotropy in the region was not significant, even in an extreme test case where the anisotropy was concentrated in a shallow layer. Kendall *et al.* (2021) conduct a study of radial anisotropy beneath the

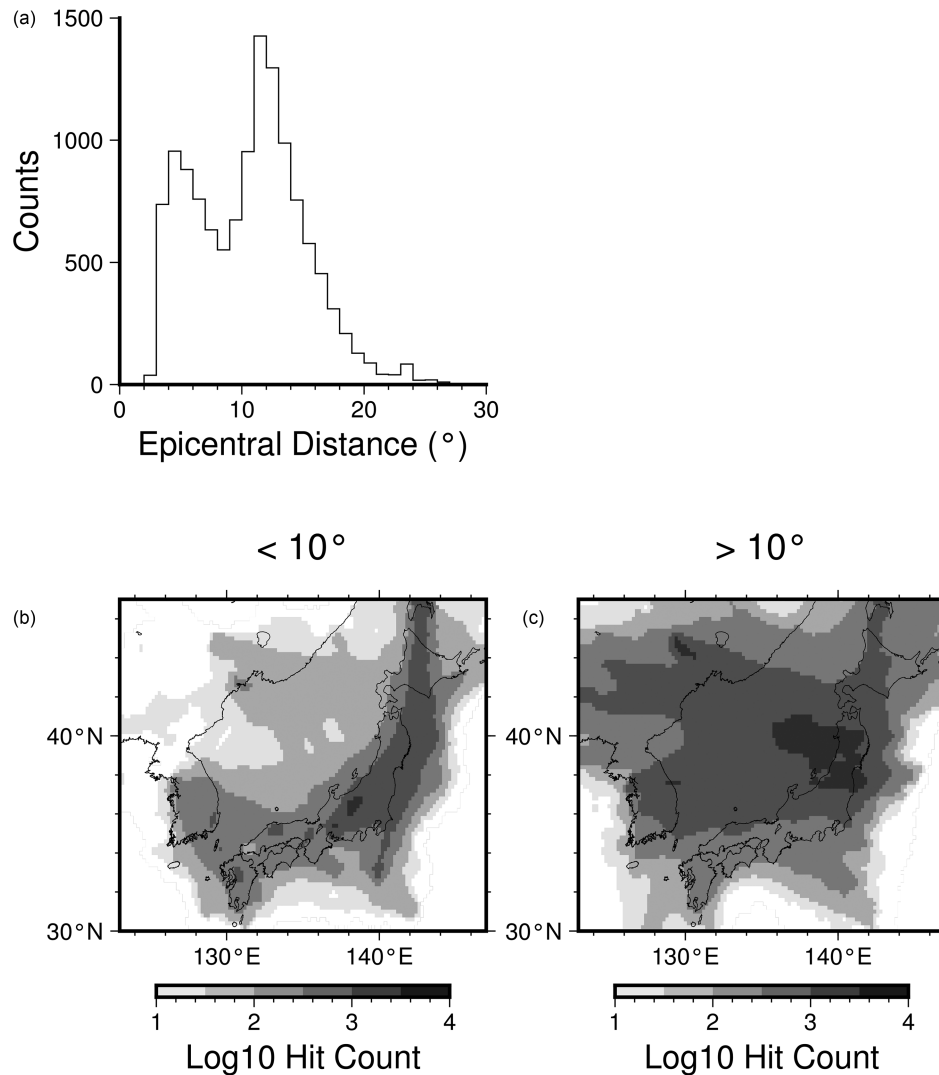


Figure 11. (a) Path length distribution for the ray paths used in this study. (b) Hit counts for ray paths under 10° length. (c) Hit counts for ray paths over 10° length.

Pacific and find that the effect of azimuthal anisotropy corrections for Rayleigh wave data on the retrieved structure is small, which may indicate that sufficient ray path coverage with surface wave data can effectively retrieve radial anisotropy. Furthermore, Legendre *et al.* (2016) measured Rayleigh wave azimuthal anisotropy in the East Sea region in the period range 30–80 s period and found 2ψ anisotropy up to 1 per cent in amplitude, while 4ψ anisotropy was not found necessary to explain their data. At 30–40 s period, they show an NE–SW fast propagation direction in the East Sea, and a roughly E–W pattern in Japan. At longer periods, the fast propagation direction switches to a predominantly NE–SW direction across our entire study region. Liu & Zhao (2016) also perform a Rayleigh wave azimuthal anisotropy study in the period range 20–150 s. They show fairly similar results as Legendre *et al.* (2016) around Japan at short periods, while at longer periods >80 s the amplitude of the anisotropy around Japan and the East Sea are ≤ 1 per cent. Since our data set contains ray paths both parallel and perpendicular to the fast velocity directions shown in these studies, we do not expect the azimuthal anisotropy to significantly bias the radial anisotropy results. Nevertheless, azimuthal anisotropy should be included and

we plan on expanding our method to include its effects in future work.

4.4 3-D model results

The focus of this study is to provide regional constraints for radially anisotropic structure from an automatic waveform fitting method. In this section, we compare our results with two other high-resolution models of the East Sea region to highlight that our method is able to produce similar features while extending the depth range of retrievable radially anisotropic structure. FWEA18 (Tao *et al.* 2018) is a model of East Asia constructed from a set of 95 earthquakes recorded at 1590 seismic stations located across Taiwan, China, South Korea, Japan, and Mongolia, and their final data set included $\sim 1\,350\,000$ body wave windows filtered between 8–100 s period and $\sim 72\,000$ surface wave windows filtered between 40–100 s period. KEA20 (Witek *et al.* 2021) is a radially anisotropic model of East Asia derived from $\sim 175\,000$ Love wave and $\sim 500\,000$ Rayleigh wave group and phase velocity dispersion curves in the

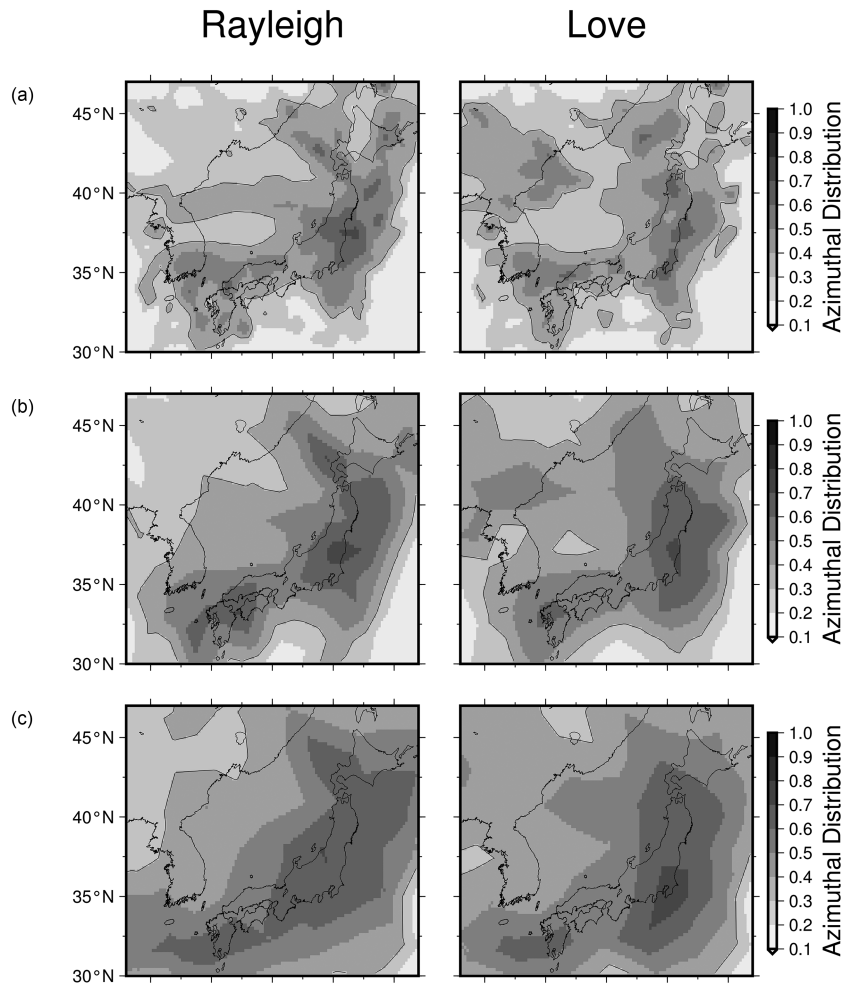


Figure 12. Azimuthal distributions for grids with average spacings of (a) 1° , (b) 2° and (c) 3° . Column headings denote wave types: Rayleigh for vertical and radial component contributions and Love for the transverse component. A contour is drawn at $\chi = 0.3$.

period range 5–375 s period that accounts for crustal nonlinearity by using a fully 3-D reference model and local sensitivity kernels to invert for structural perturbations in the crust. Figs 13 and 14 compare horizontal depth slices of Voigt v_s for our model and models FWEA18 and KEA20, and Figs 15 and 16 compare the radial anisotropy ξ . We mask our model using the summed squares of the columns of the weighted constraints matrix $C_e^{-1/2}G$. Regions with high sums roughly correspond to areas in our model with high ray path density and high resolution.

4.4.1 Isotropic shear wave speeds

In general, all three models show a great deal of similarity at shallow depths. At 40–60 km depth, we observe low velocity anomalies in NE Japan, Hokkaido, and Kyushu Island, which can be related to ongoing volcanism in those regions. For example, high attenuation (Kita *et al.* 2014) and low V_S (Koulakov *et al.* 2015) are imaged underneath volcanoes in Hokkaido at crustal depths and in the mantle wedge. Similar low-velocity anomalies are imaged in NE Japan under active arc volcanoes in regions of low-frequency microearthquakes associated with fluid magma movement (e.g. Xia *et al.* 2007). A magnetotelluric study by Hata *et al.* (2015) imaged high conductivity under active volcanoes on Kyushu indicating fluid upwelling. We image relatively low V_S anomalies along

the eastern margin of the Korean Peninsula, which may be due to a small fraction of partial melt in the subcontinental lithosphere (e.g. Song *et al.* 2020). Differences between our model and KEA20 may be explained by the latter containing shorter period dispersion data, which are sensitive to shallower crustal structures and may therefore improve resolution in the uppermost mantle by removing ambiguity in the longer period data that are also sensitive to shallow structures. The low velocity anomalies observed in our model along the eastern margin of the East Sea at 40 km depth may therefore be the result of vertical smearing from deeper anomalies at 60–80 km depth (e.g. Fig. 9). We also note that in the original study of Tao *et al.* (2018), they did not discuss features in FWEA18 above 80 km depth. At 80 and 100 km depth, the East Sea, the Korean Peninsula, NE China, and Far East Russia transition to a wide region of low velocity anomalies. This has been imaged in previous studies (e.g. Zheng *et al.* 2011; Chen *et al.* 2012; Shen *et al.* 2016; Simutè *et al.* 2016), and has been interpreted as a warm and hydrated upwelling caused by shallow and deep slab dehydration reactions (e.g. Zhao *et al.* 2007).

At 60 km depth, we begin to see stronger high-velocity anomalies in the eastern offshore region of NE Japan, indicating the Pacific plate. The Philippine Sea plate can be seen as a high-velocity anomaly to the south of SW Japan. As depth increases, we can clearly see the Pacific slab anomaly moving westward, although smearing effects (e.g. Figs 9 and 10) prevent us from obtaining

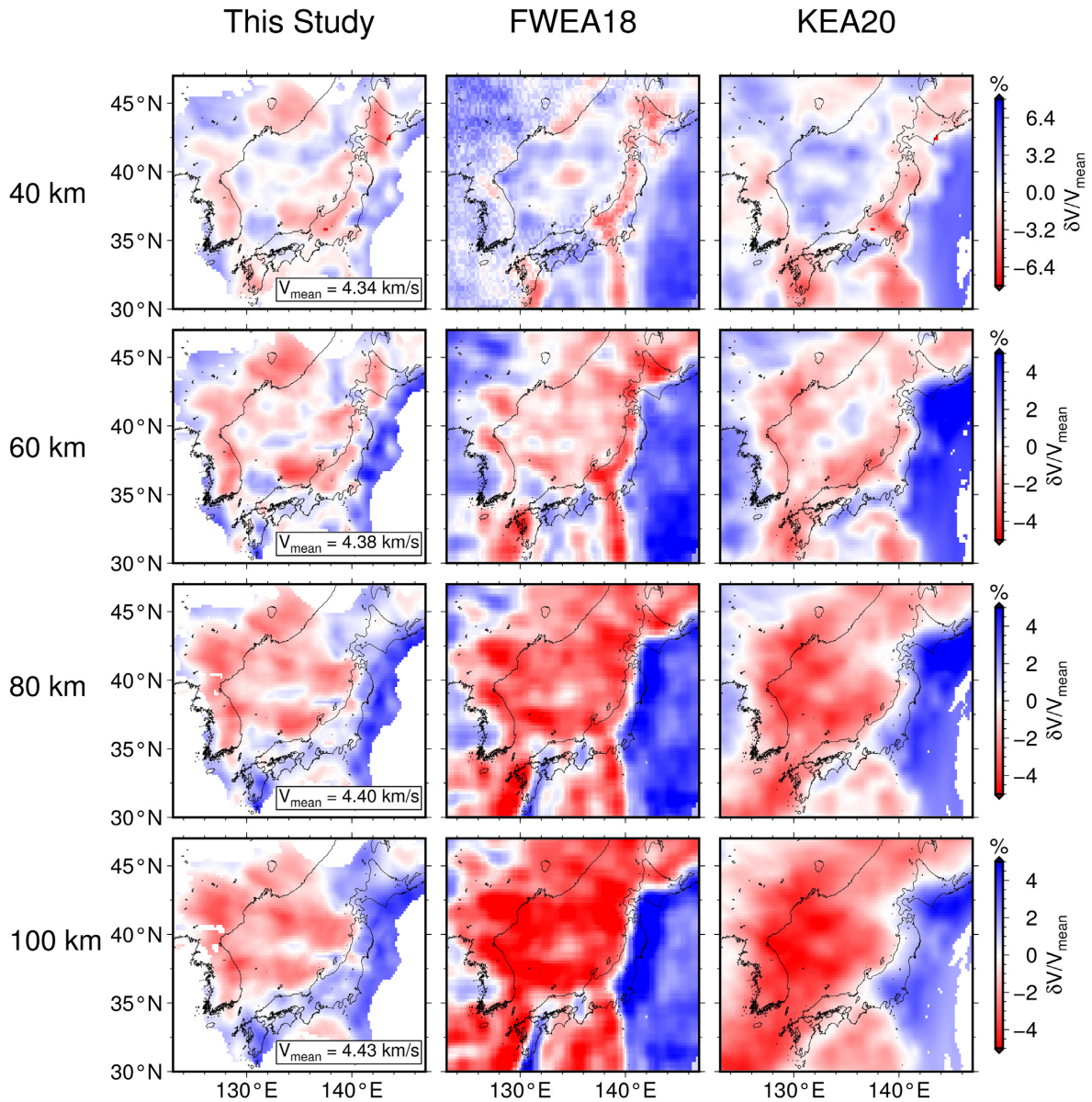


Figure 13. Horizontal depth slices of Voigt v_s through our model (left column) and models FWEA18 (middle column, Tao *et al.* 2018) and KEA20 (right column, Witek *et al.* 2021) at 40, 60, 80 and 100 km depths. Depths are shown in the bottom left corner in the leftmost column. Colour scale shows perturbations with respect to the mean Voigt v_s in our model, shown in the bottom right in the leftmost column.

clear slab boundaries as in FWEA18. KEA20 also shows the same general trend, but loses resolution at depths greater than 200 km due to a paucity of overtone data, which shows the importance of fitting body wave phases in our method. Our model also shows poorer resolution below 200 km depth compared to FWEA18, and this may be due to the short ray paths considered in this regional study (e.g. Fig. 11).

4.4.2 Radial anisotropy

The three models of radial anisotropy show more differences compared to isotropic shear wave speeds. At shallow depths between 40–80 km, our model appears more consistent with KEA20 than with FWEA18. For example, FWEA18 images a region of positive radial anisotropy (i.e. $\beta_H > \beta_V$) in the Japan basin between

40–60 km depth, whereas our model and KEA20 show isotropic to weakly negative anisotropy (i.e. $\beta_V > \beta_H$). Our model also shows negative anisotropy in the Ulleung basin, whereas FWEA18 shows weakly positive anisotropy and KEA20 has a broad region of strong positive anisotropy. Since this study contains a larger combination of short-period Love and Rayleigh wave paths with short lengths (e.g. Fig. 11) compared to KEA20 and FWEA18, we believe that our results may be more robust. The negative anisotropy seen in the Japan and Ulleung basins occurs with high V_S anomalies (e.g. Fig. 13 at 40 km), and may represent regions of foundering lower crust or frozen vertical upwellings that formed during the formation of the East Sea.

Strong positive anisotropy and low V_S is imaged off the western coast of NE Japan near the Yamato basin, which may indicate return flow of mantle material back towards the mantle wedge. The strong positive anisotropy transitions to an area of weak positive

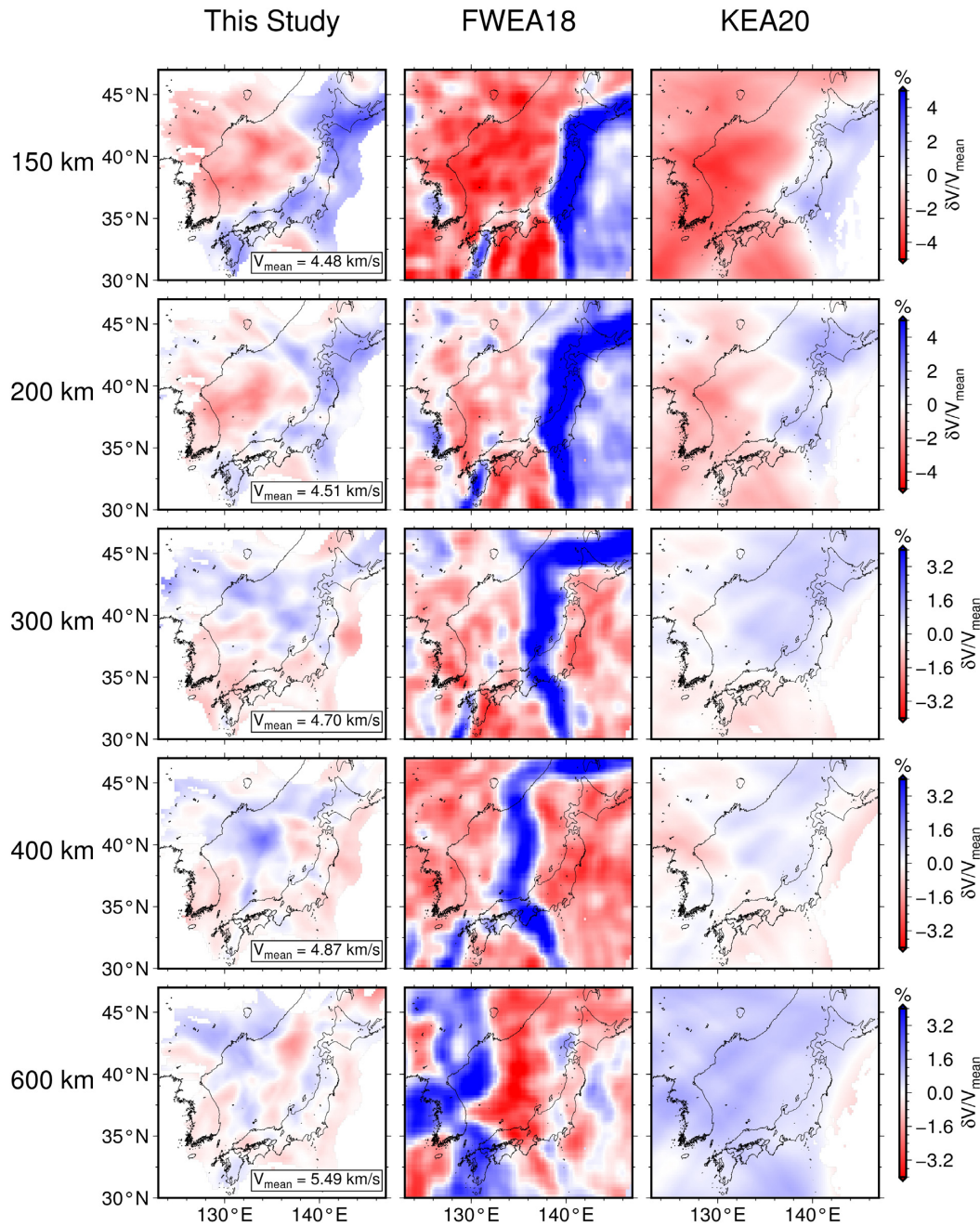


Figure 14. Same as Fig. 13, but for depths 150, 200, 300, 400 and 600 km.

anisotropy in the mantle wedge underneath NE Japan. A recent study of P -wave anisotropy by Wang & Zhao (2021), who considered hexagonal symmetry with the symmetry axis orientation as a free parameter, shows a complex mantle flow field underneath NE Japan. Since we assume a vertical symmetry axis in this study, the reduced strength of radial anisotropy we observe in the upper mantle may be due to 3-D mantle flow aligning mineral fabrics in a subhorizontal orientation. A study of P -wave anisotropy by Liu & Zhao (2017) indicates that E-type olivine fabric may dominate the mantle wedge underneath NE Japan, which would also cause weak positive ξ under horizontal shear (Karato *et al.* 2008).

In contrast to NE Japan, subduction underneath SW Japan is thought to take place at temperatures that are 300–500 °C warmer,

which may result in shallower eclogite transformation and slab dehydration reactions (e.g. Peacock & Wang 1999). Consequently, the combination of low stresses, higher temperatures, and larger water contents in the shallower mantle wedge may cause olivine fabrics to change from A-type to E-type and then C-type olivine (e.g. Karato *et al.* 2008). In SW Japan, Liu & Zhao (2017) suggested that C-type olivine may dominate the mantle wedge, although there may be E-type olivine around Kyushu Island. A warm, hydrated upwelling from the subducting Philippine Sea plate may cause C-type olivine to show weakly positive anisotropy due to vertical shear and negative anisotropy if there is horizontal flow (Karato *et al.* 2008). At 40–80 km depth where we may expect horizontal flow back towards the trench, our model and FWEA18 both show positive radial

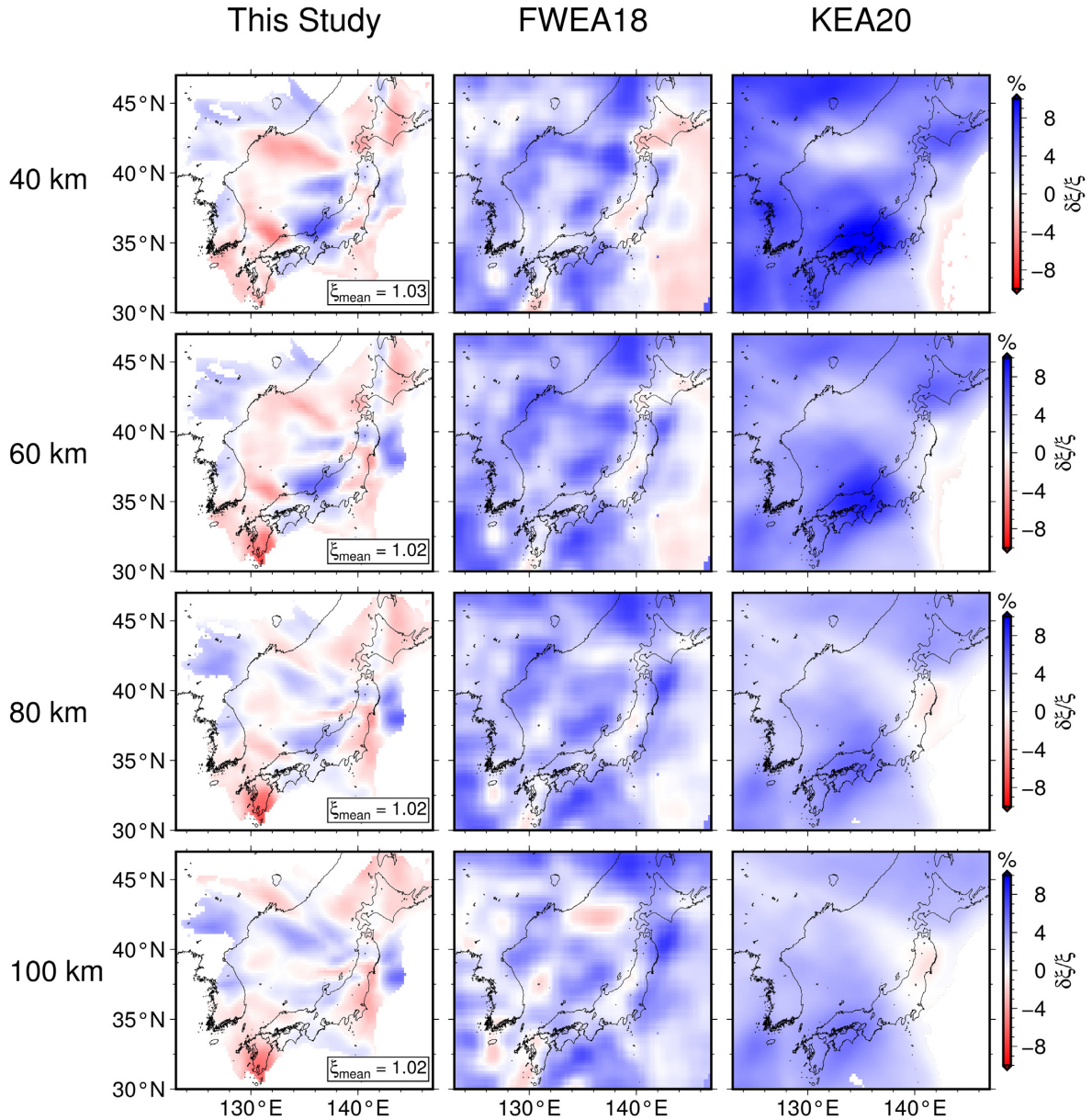


Figure 15. Horizontal depth slices of ξ through our model (left column) and models FWEA18 (middle column, Tao *et al.* 2018) and KEA20 (right column, Witek *et al.* 2021) at 40, 60, 80 and 100 km depths. Depths are shown in the bottom left corner in the leftmost column. Colour scale shows per cent perturbations with respect to the mean value in our model at the given depth.

anisotropy anomalies above the slab. Our results may be more consistent with A- or E-type olivine under SW Japan, depending on the amount of water present. C-type olivine under horizontal shear could produce $\xi \approx 0.90\text{--}0.95$, which may be consistent with our results at Kyushu, with a possibility of more water content there.

At greater depths below 150 km, we see a transition to large regions showing negative anisotropy. At 200 km depth, we image negative anisotropy in the Japan basin, Russia, and NE China, which is similar to FWEA18. Below 150 km depth, the resolution of KEA20 drops significantly, resulting in a mostly smooth model, while FWEA18 limited the existence of radial anisotropy to the top 200 km of their model. While we show anisotropy below 200 km depth in Fig. 16 for completeness, we do not attempt to interpret the model since the resolution there is more limited (e.g. Fig. 9).

4.5 Model validation

In order to judge the performance of our model, we calculate three-component synthetic waveforms for events not used in the inversions for our model, FWEA18, and KEA20, and we compare them with the observed waveforms as well as with synthetic waveforms calculated for our 3-D reference model. The first event we show is an M_w 6 thrust fault occurring at 34 km depth off the coast of NE Japan at 2:44:11 UTC on 12 September 2020. We choose two stations for this event: IU.YSS, which is 8.2° away almost due north of the event, and IC.MDJ, which is located 11.1° to the northwest. The results are shown in Fig. 17. The waveforms are bandpassed between 10–30 mHz, and we calculate window misfits according to eq. (14) between 245–390 s for station IC.MDJ, and 180–350 s for station IU.YSS. In both cases, the reference model waveforms are reasonably close to the observations, which is an indication that the

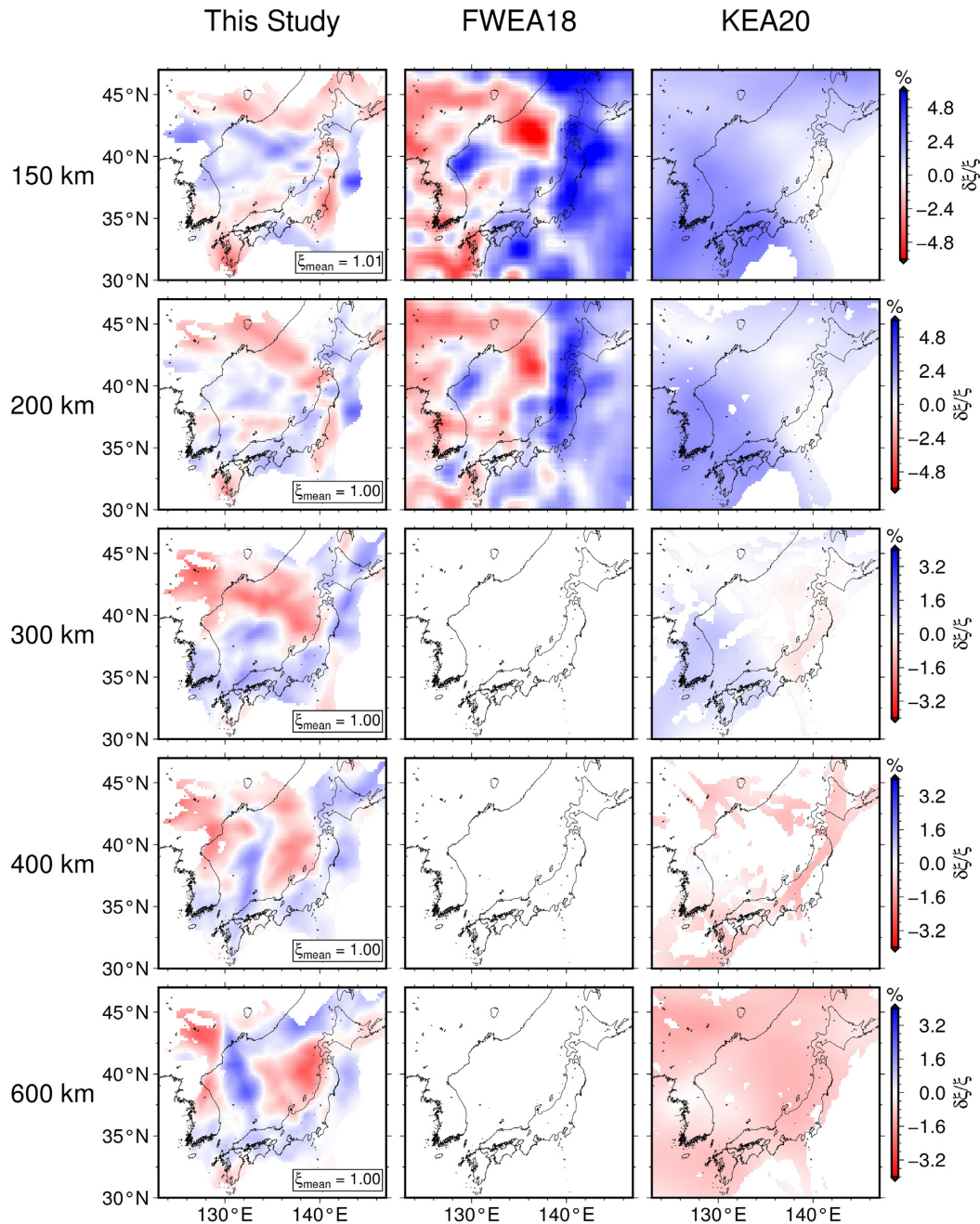


Figure 16. Same as Fig. 15, for but depths 150, 200, 300, 400 and 600 km.

linearization in eq. (2) is justified by our choice of reference 3-D model. For the waveforms recorded at station IC.MDJ, our model performs better than the reference model, although the synthetic vertical and radial component waveforms are slightly slower than the observations, and FWEA18 shows higher levels of misfit compared to our model and KEA20. For the waveforms recorded at station IU.YSS, our model performs similar to the reference model on the vertical and radial components, but we see a much higher misfit on the transverse component. Similar results are seen for KEA20, but this may be expected since the event-receiver path for station IU.YSS is on the edge of the study region for both models. FWEA18, on the other hand, shows slightly poorer results on the vertical and radial components, while the transverse component sees a good fit to the observations.

In Fig. 18, we show three-component waveform comparisons for a deep M_w 6 event occurring at 214 km depth at 2:39:43 UTC on 12 January 2021. We chose two stations, G.INU and IU.INCN, whose ray paths lie within the well-resolved region in our model. Similar to before, we bandpass the waveforms between 10–30 mHz and we calculate misfits between 150–350 s for station G.INU, and between 250–450 s for station IU.INCN. Compared to the reference model, our model reduces the misfit quite well for the vertical and radial components, while the transverse component only sees a mild improvement in the misfit. Since this is a deep event, the waveforms are primarily composed of excited overtone signals that are sensitive to deeper structures compared to the fundamental mode. The reduction in misfit that we see is a good indication that the deep structures in our model are reliable. It is interesting to note that

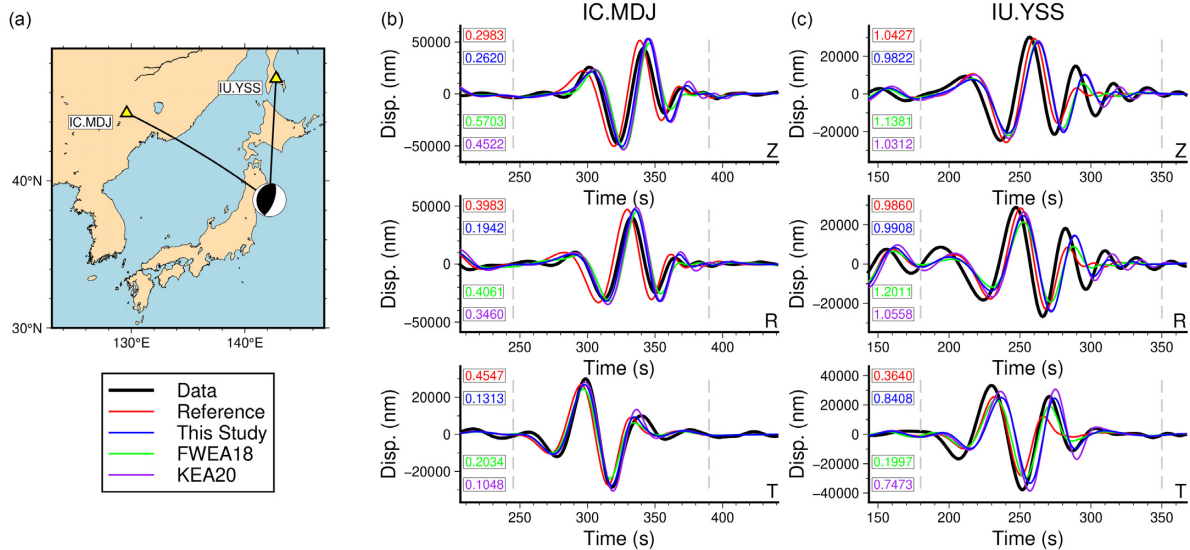


Figure 17. Waveform comparisons for an M_w 6 event occurring off the coast of NE Japan at 2:44:11 UTC on 12 September 2020 at a depth of 34 km. (a) Map showing the locations of the event and stations IC.MDJ and IU.YSS. (b) Three-component waveform comparison between the data for station IC.MDJ and synthetics calculated for the reference 3-D model, the model shown in this study, FWEA18, and KEA20. Legend is shown below (a). Coloured text shows misfits calculated for a window between 245–390 s, with colours matching those shown in the legend. (c) Same as (b), but for station IU.YSS. Misfits are calculated between 180–350 s. All waveforms are bandpassed between 10–30 mHz. Grey dashed vertical lines denote time windows used.

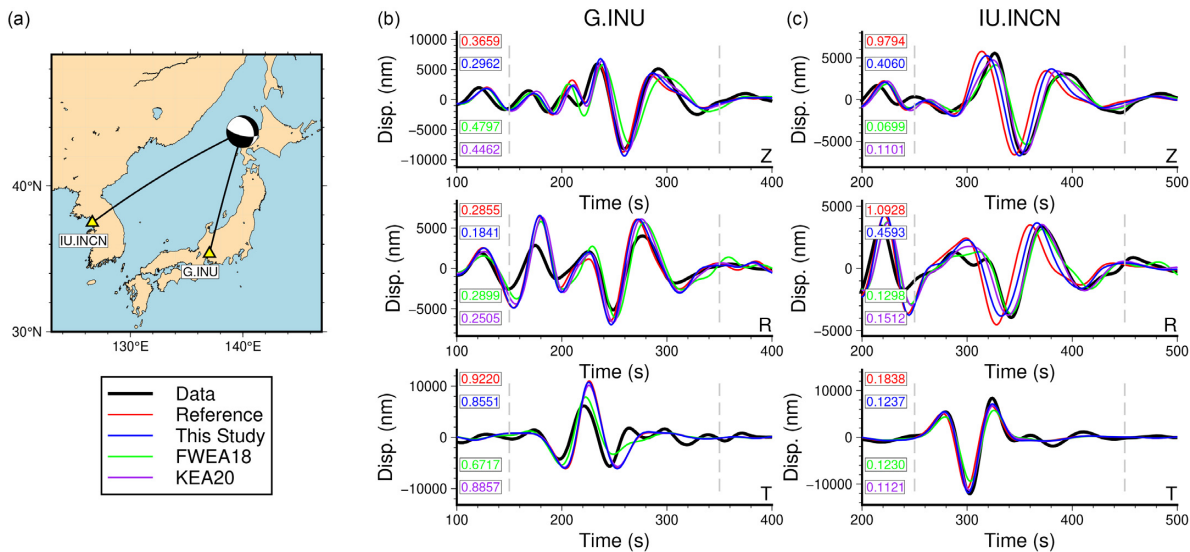


Figure 18. Waveform comparisons for an M_w 6 event occurring west of Hokkaido at 2:39:43 UTC on 12 January 2021 at a depth of 214 km. (a) Map showing the locations of the event and stations G.INU and IU.INCN. (b) Three-component waveform comparison for station G.INU, with the legend shown below (a). Coloured text shows misfits calculated between 150–350 s. (c) Same as (b), but for station IU.INCN. Misfits are calculated between 250–450 s. All waveforms are bandpassed between 10–30 mHz. Grey dashed vertical lines denote time windows used.

KEA20 and FWEA18 show lower misfits for the waveforms observed at IU.INCN, while our model shows better misfit reduction at G.INU. This could be because our model has better resolution in the region closer to NE Japan compared to the northern East Sea, as we showed in our resolution testing (e.g. Section 4.3).

5 CONCLUSIONS

In this study, we present a novel S and surface waveform inversion method that is based on surface wave ray theory and the PWI method (Nolet 1990). We expand and modify PWI to retrieve

linear constraints on radially anisotropic Earth structures by using phase-matched filtering to extract and fit the fundamental mode signal, and to obtain an optimal reference model and kernel for each path. Afterwards, we fit the full normal mode sum composed of the fundamental mode and overtones up to $n = 20$, and we show that, depending on the ray path lengths used, we are able to extract information on radial anisotropy down to the mantle transition zone even when the overtones significantly interfere with the fundamental mode. We also present a novel method to automatically estimate the level of uncertainty in the resulting path average model constraints, which in previous applications was done in an ad hoc manner for each waveform. Previous studies using the PWI method

have so far focused on Rayleigh waves (e.g. Schaeffer *et al.* 2016), while other multimode waveform fitting studies consider Love and Rayleigh waves in separate path-average model inversions (e.g. Ho *et al.* 2016; Priestley *et al.* 2019).

We apply our method to ~16 000 waveforms propagating through the East Sea and recorded at seismic networks in China, South Korea, and Japan. The resulting linear constraints and uncertainties were used in an inversion for a radially anisotropic model of the crust and upper mantle of the East Sea. Our model shows features that are generally consistent with two recent models, FWEA18 (Tao *et al.* 2018), which was derived using a full-waveform inversion method, and KEA20 (Witek *et al.* 2021), which has derived from an inversion of surface wave dispersion data. We validate our model by forward calculating waveforms for earthquakes that were not used in our inversion. The waveforms for our model, FWEA18, and KEA20 match the data well, showing that our automatic waveform inversion method extracts structural information and can produce models that perform as well as previous studies.

ACKNOWLEDGMENTS

This work was financially supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2019R1A2C208506111), by the Basic Science Research Program through the NRF funded by the Ministry of Education (No. 2019R1A6A1A03033167), by the Ministry of the Interior and Safety as Human Resource Development Project in Disaster Management and by the Korea Meteorological Administration Research and Development Program under grants KMI 2018–02910 and KMI 2022-00810. Figures were created using the Generic Mapping Tools software (Wessel *et al.* 2019).

DATA AVAILABILITY

All earthquake waveform data used in this study are publicly available. KIGAM data are available at <https://www.kigam.re.kr/quake> (English: <https://www.kigam.re.kr/english>). KMA data are available at <http://necis.kma.go.kr> (English: http://www.kma.go.kr/eng/weather/kma_service/observation.jsp). KMA and KIGAM waveform data can also be found at the IRIS Data Management Center (DMC) under FDSN codes KS and KG, respectively. F-net data can be found at <https://www.fnet.bosai.go.jp>. All other waveforms may be accessed via the IRIS DMC [Institut de physique du globe de Paris (IPGP) & École et Observatoire des Sciences de la Terre de Strasbourg (EOST) 1982; Scripps Institution of Oceanography 1986; Albuquerque Seismological Laboratory (ASL)/USGS 1988; Grand & Ni 2009; Albuquerque Seismological Laboratory (ASL)/USGS 1992; Institute of Geophysics China Earthquake Administration (IGPCEA) 2000]. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048.

REFERENCES

Ai, S., Zheng, Y. & Wang, S., 2020. Crustal deformations of the central North China Craton constrained by radial anisotropy, *J. geophys. Res.*, **125**, e2019JB018374, doi:10.1029/2019JB018374.

Albuquerque Seismological Laboratory (ASL)/USGS, 2014. *Global Seismograph Network - IRIS/USGS [Data set]*, International Federation of Digital Seismograph Networks, doi:10.7914/SN/IU.

Albuquerque Seismological Laboratory (ASL)/USGS, 1992. *New China Digital Seismograph Network [Data set]*, International Federation of Digital Seismograph Networks, doi:10.7914/SN/IC.

Anderson, D.L., 1961. Elastic wave propagation in layered anisotropic media, *J. geophys. Res.*, **66**(9), 2953–2963.

Anderson, D.L., Minster, B. & Cole, D., 1974. Effect of oriented cracks on seismic velocities, *J. geophys. Res.*, **79**(26), 4011–4015.

Anderson, O.L., Schreiber, E., Liebermann, R.C. & Soga, N., 1968. Some elastic constant data on minerals relevant to geophysics, *Rev. Geophys.*, **6**(4), 491–524.

Auer, L., Boschi, L., Becker, T.W., Nissen-Meyer, T. & Giardini, D., 2014. Savani: a variable resolution whole-mantle model of anisotropic shear velocity variations based on multiple data sets, *J. geophys. Res.*, **119**, 3006–3034.

Barmin, M.P., Ritzwoller, M.H. & Levshin, A.L., 2001. A fast and reliable method for surface wave tomography, *Pure appl. Geophys.*, **158**, 1351–1375.

Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.*, **81**, F897.

Bedle, H. & van der Lee, S., 2009. S velocity variations beneath North America, *J. geophys. Res.*, **114**, B07308, doi:10.1029/2008JB005949.

Beghein, C., 2010. Radial anisotropy and prior petrological constraints: a comparative study, *J. geophys. Res.*, **115**, B03303, doi:10.1029/2008JB005842.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**(3), 1239–1260.

Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.*, **103**(B12), 30055–30078.

Boschi, L., 2006. Global multiresolution models of surface wave propagation: comparing equivalently regularized Born and ray theoretical solutions, *Geophys. J. Int.*, **167**(1), 238–252.

Bozdağ, E. & Tappert, J., 2008. On crustal corrections in surface wave tomography, *Geophys. J. Int.*, **172**, 1066–1082.

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: First-generation model, *Geophys. J. Int.*, **207**, 1739–1766.

Burdick, S. *et al.*, 2008. Upper mantle heterogeneity beneath North America from travel time tomography with global and USArray Transportable Array data, *Seismol. Res. Lett.*, **79**(3), 384–392.

Cara, M. & Lévêque, J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.*, **14**, 1046–1049.

Chang, S.-J. & Ferreira, A. M.G., 2017. Improving global radial anisotropy tomography: the importance of simultaneously inverting for crustal and mantle structure, *Bull. seism. Soc. Am.*, **107**(2), 624–638.

Chang, S.-J. & van der Lee, S., 2011. Mantle plumes and associated flow beneath Arabia and east Africa, *Earth planet. Sci. Lett.*, **302**(3–4), 448–454.

Chang, S.-J. *et al.*, 2010. Joint inversion for three-dimensional S velocity mantle structure along the Tethyan margin, *J. geophys. Res.*, **115**, B08309, doi:10.1029/2009JB007204.

Chang, S.-J., Ferreira, A. M.G., Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 2014. Global radially anisotropic mantle structure from multiple datasets: a review, current challenges, and outlook, *Tectonophysics*, **617**, 1–19.

Chang, S.-J., Ferreira, A. M.G., Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 2015. Joint inversion for global isotropic and radially anisotropic mantle structure including crustal thickness perturbations, *J. geophys. Res.*, **120**, 4278–4300.

Chavent, G., 1974. Identification of functional parameters in partial differential equations, in *Joint Automatic Control Conference*, Vol. 12, pp. 155–156, Austin, Texas, USA.

Chen, M., Niu, F., Liu, Q., Tromp, J. & Zheng, X., 2012. Multiparameter adjoint tomography of the crust and upper mantle beneath East Asia: 1. Model construction and comparisons, *J. geophys. Res.*, **120**, 1762–1786.

- Christensen, N.I. & Mooney, W.D., 1995. Seismic velocity structure and composition of the continental crust: a global view, *J. geophys. Res.*, **100**(B7), 9761–9788.
- Crampton, S., 1987. Geological and industrial implications of extensive-dilatancy anisotropy, *Nature*, **328**, 491–496.
- Crotwell, H.P., Owens, T.J. & Ritsema, J., 1999. The TauP toolkit: flexible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.*, **70**(2), 154–160.
- Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton University Press.
- Debayle, E., 1999. SV-wave azimuthal anisotropy in the Australian upper mantle: preliminary results from automated Rayleigh waveform inversion, *Geophys. J. Int.*, **137**(3), 747–754.
- Debayle, E. & Ricard, Y., 2012. A global shear velocity model of the upper mantle from fundamental and higher rayleigh mode measurements, *J. geophys. Res.*, **117**, B10308, doi:10.1029/2012JB009288.
- Debayle, E., Dubuffet, F. & Durand, S., 2016. An automatically updated S-wave model of the upper mantle and the depth extent of azimuthal anisotropy, *Geophys. Res. Lett.*, **43**, 674–682.
- Dziewoński, A., Mills, J. & Bloch, S., 1972. Residual dispersion measurement - a new method of surface-wave analysis, *Bull. seism. Soc. Am.*, **62**(1), 129–139.
- Dziewoński, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**, 297–356.
- Ekström, G., 2011. A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25–250 s, *Geophys. J. Int.*, **187**, 1668–1686.
- Ekström, G. & Dziewoński, A.M., 1998. The unique anisotropy of the Pacific upper mantle, *Nature*, **394**, 168–172.
- Ekström, G., Tromp, J. & Larson, E. W.F., 1997. Measurements and global models of surface wave propagation, *J. geophys. Res.*, **102**(B4), 8137–8157.
- Ekström, G., Nettles, M. & Dziewoński, A.M., 2012. The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.*, **200–201**, 1–9.
- Feng, M., van der Lee, S. & Assumpção, M., 2007. Upper mantle structure of south America from joint inversion of waveforms and fundamental mode group velocities of Rayleigh waves, *J. Geophys. Res.*, **112**(B4), doi:10.1029/2006jb004449.
- Feng, M., van der Lee, S., An, M. & Zhao, Y., 2010. Lithospheric thickness, thinning, subduction, and interaction with the asthenosphere beneath china from the joint inversion of seismic S-wave train fits and Rayleigh-wave dispersion curves, *Lithos*, **120**(1–2), 116–130.
- Ferreira, A. M.G., Woodhouse, J.H., Visser, K. & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *J. geophys. Res.*, **115**(B04313), doi:10.1029/2009JB006716.
- Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology I. Theory, *Phys. Earth planet. Inter.*, **157**, 86–104.
- Fichtner, A., Kennett, B. L.N., 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.
- Goforth, T. & Herrin, E., 1979. Phase-matched filters: Application to the study of Love waves, *Bull. seism. Soc. Am.*, **69**(1), 27–44.
- Grand, S.P., van der Hilst, R.D. & Widiyantoro, S., 1997. High resolution global tomography: a snapshot of convection in the Earth, *Geol. Soc. Am. Today*, **7**(4), 1–7.
- Grand, S. & Ni, J., 2009. *Collaborative research: Northeast China Extended Seismic Array: Deep subduction, mantle dynamics and lithospheric evolution beneath Northeast China [Data set]*, International Federation of Digital Seismograph Networks, doi:10.7914/SN/YP_2009.
- Guo, Z., Yang, Y. & Chen, Y.J., 2016. Crustal radial anisotropy in Northeast China and its implications for the regional tectonic extension, *Geophys. J. Int.*, **207**, 197–208.
- Hariharan, A., Dalton, C.A., Babikoff, J. & Ekström, G., 2022. Controls on surface wave overtone interference, *Geophys. J. Int.*, **228**, 1665–1683.
- Hata, M., Oshiman, N., Yoshimura, R., Tanaka, Y. & Uyeshima, M., 2015. Three-dimensional electromagnetic imaging of upwelling fluids in the Kyushu subduction zone, Japan, *J. geophys. Res.*, **120**, 1–17.
- Hayes, G.P., Moore, G.L., Porter, D.E., Hearne, M., Flamme, H., Furtney, M. & Smoczyk, G.M., 2018. Slab2, a comprehensive subduction zone geometry model, *Science*, **362**, 58–61.
- Herrin, E. & Goforth, T., 1977. Phase-matched filters: Application to the study of Rayleigh waves, *Bull. seism. Soc. Am.*, **67**(5), 1259–1275.
- Hess, H.H., 1964. Seismic anisotropy of the uppermost mantle under oceans, *Nature*, **203**, 629–631.
- Ho, T., Priestley, K. & Debayle, E., 2016. A global horizontal shear velocity model of the upper mantle from multimode Love wave measurements, *Geophys. J. Int.*, **207**, 542–561.
- Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y., 1990. Whole mantle P-wave travel time tomography, *Phys. Earth planet. Inter.*, **59**, 294–328.
- Institut de physique du globe de Paris (IPGP), École et Observatoire des Sciences de la Terre de Strasbourg (EOST), 1982. *GEOSCOPE, French Global Network of broad band seismic stations*, Institut de physique du globe de Paris (IPGP), Université de Paris, doi:10.18715/GEOSCOPE.G.
- Institute of Geophysics China Earthquake Administration (IGPCEA), 2000. *China National Seismic Network, Data Management Centre of China National Seismic Network at Institute of Geophysics, CEA [Data set]*, International Federation of Digital Seismograph Networks, doi:10.7914/SN/CB.
- Ismail, W.B. & Mainprice, D., 1998. An olivine fabric database: an overview of upper mantle fabrics and seismic anisotropy, *Tectonophysics*, **296**, 145–157.
- Karato, S., Jung, H., Katayama, I. & Skemer, P., 2008. Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies, *Annu. Rev. Earth Planet. Sci.*, **36**, 59–95.
- Kawakatsu, H., 2016. A new fifth parameter for transverse isotropy II: partial derivatives, *Geophys. J. Int.*, **206**, 360–367.
- Kendall, E., Ferreira, A. M.G., Chang, S.-J., Witek, M. & Peter, D., 2021. Constraints on the upper mantle structure beneath the Pacific from 3-D anisotropic waveform modeling, *J. geophys. Res.*, **126**, e2020JB020003, doi:10.1029/2020JB020003.
- Kennett, B. L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int.*, **122**, 108–124.
- Kim, G., Yoon, S., Chough, S., Kwon, Y. & Ryu, B., 2011. Seismic reflection study of the acoustic basement in the South Korea Plateau, the Ulleung Interplain Gap, and the northern Ulleung Basin: Volcano-tectonic implications for Tertiary back-arc evolution in the southern East Sea, *Tectonophysics*, **504**, 43–56.
- Kita, S., Nakajima, J., Hasegawa, A., Okada, T., Katsumata, K., Asano, Y. & Kimura, T., 2014. Detailed seismic attenuation structure beneath Hokkaido, northeastern Japan: Arc-arc collision process, arc magmatism, and seismotectonics, *J. geophys. Res.*, **119**, 6486–6511.
- Ko, B. & Jung, H., 2015. Crystal preferred orientation of an amphibole experimentally deformed by simple shear, *Nat. Commun.*, **6**(6586), 1–10.
- Komatitsch, D., Ritsema, J. & Tromp, J., 2002. The spectral-element method, beowulf computing, and global seismology, *Science*, **298**(5599), 1737–1742.
- Komatitsch, D., Erlebacher, G., Göddeke, D. & Michéa, D., 2010. High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster, *J. Comput. Phys.*, **229**(20), 7692–7714.
- Koulakov, I., Kukarina, E., Fathi, I.H., Khrepy, S.E. & Al-Arifi, N., 2015. Anisotropic tomography of Hokkaido reveals delamination-induced flow above a subducting slab, *J. geophys. Res.*, **120**, 3219–3239.
- Krischer, L. et al., 2016. An adaptable seismic data format, *Geophys. J. Int.*, **207**, 1003–1011.
- Kumar, P. & Kawakatsu, H., 2011. Imaging the seismic lithosphere-asthenosphere boundary of the oceanic plate, *Geochem. Geophys. Geosyst.*, **12**(1), Q01006, doi:10.1029/2010GC003358.
- Kustowski, B., Ekström, G. & Dziewoński, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res.*, **113**, B06306, doi:10.1029/2007JB005169.
- Larson, E. W.F., Tromp, J. & Ekström, G., 1998. Effects of slight anisotropy on surface waves, *Geophys. J. Int.*, **132**, 654–666.
- Laske, G., Masters, G., Ma, Z. & Pasyanos, M., 2013. Update on CRUST1.0—a 1-degree global model of Earth's crust, *Geophys. Res. Abstr.*, **15**, Abstract EGU2013–2658.

- Lebedev, S. & Nolet, G., 2003. Upper mantle beneath southeast Asia from S-velocity tomography, *J. geophys. Res.*, **108**(B1), doi:10.1029/2000jb00073.
- Lebedev, S. & van der Hilst, R.D., 2008. Global upper-mantle tomography with the automated multimode inversion of surface and S-wave forms, *Geophys. J. Int.*, **173**, 505–518.
- Lebedev, S., Nolet, G., Meier, T. & van der Hilst, R.D., 2005. Automated multimode inversion of surface and S waveforms, *Geophys. J. Int.*, **162**, 951–964.
- Lefebvre, M. et al., 2017. Data and workflow management for exascale global adjoint tomography, in *Exascale Scientific Applications: Scalability and Performance Portability*, pp. 279–306, eds. Straatsma, T.P., Antypas, K.B., Williams, T.J., Chapman and Hall/CRC, New York.
- Legendre, C.P., Zhao, L., Deschamps, F. & Chen, Q.-F., 2016. Layered anisotropy within the crust and lithospheric mantle beneath the Sea of Japan, *J. Asian Earth Sci.*, **128**, 181–195.
- Lei, W. et al., 2020. Global adjoint tomography–model GLAD-M25, *Geophys. J. Int.*, **223**(1), 1–21.
- Lekić, V. & Romanowicz, B., 2011. Tectonic regionalization without a priori information: a cluster analysis of upper mantle tomography, *Earth planet. Sci. Lett.*, **308**, 151–160.
- Lekić, V., Panning, M. & Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography, *Geophys. J. Int.*, **182**, 265–278.
- Lekić, V., Cottaar, S., Dziewoński, A.M. & Romanowicz, B., 2012. Cluster analysis of global lower mantle tomography: a new class of structure and implications for chemical heterogeneity, *Earth planet. Sci. Lett.*, **357–358**, 68–77.
- Levshin, A. & Ratnikova, L., 1984. Apparent anisotropy in inhomogeneous media, *Geophys. J. Int.*, **76**(1), 65–69.
- Levshin, A.L. & Ritzwoller, M.H., 2001. Automated detection, extraction, and measurement of regional surface waves, *Pure appl. Geophys.*, **158**(8), 1531–1545.
- Levshin, A.L., Yanovskaya, T.B., Lander, A.V., Bukchin, B.G., Barmin, M.P., Ratnikova, L.I. & Its, E.N., 1989. *Seismic Surface Waves in a Laterally Inhomogeneous Earth*, Kluwer Academic.
- Liu, Q. & Gu, Y., 2012. Seismic imaging: From classical to adjoint tomography, *Tectonophysics*, **566–567**, 31–66.
- Liu, X. & Zhao, D., 2016. Backarc spreading and mantle wedge flow beneath the Japan Sea: insight from Rayleigh-wave anisotropic tomography, *Geophys. J. Int.*, **207**, 357–373.
- Liu, X. & Zhao, D., 2017. P-wave anisotropy, mantle wedge flow and olivine fabrics beneath Japan, *Geophys. J. Int.*, **210**, 1410–1431.
- Lloyd, S.M. & van der Lee, S., 2008. Influence of observed mantle anisotropy on isotropic tomographic models, *Geochem. Geophys. Geosyst.*, **9**(7), Q07007, doi:10.1029/2008GC001997.
- Love, A. E.H., 1927. *A Treatise on the Theory of Elasticity*, Cambridge Univ. Press.
- Maggi, A. & Priestley, K., 2005. Surface waveform tomography of the Turkish-Iranian plateau, *Geophys. J. Int.*, **160**(3), 1068–1080.
- Mainprice, D., Barruol, G. & Ismail, W.B., 2000. *The seismic anisotropy of the Earth's mantle: from single crystal to polycrystal*, pp. 237–264, in: *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale*, eds. Karato, S.-I., Forte, A., Liebermann, R., Masters, G., Stixrude, L., American Geophysical Union.
- Manaman, N.S., Shomali, H. & Koyi, H., 2011. New constraints on upper-mantle S-velocity structure and crustal thickness of the Iranian plateau using partitioned waveform inversion, *Geophys. J. Int.*, **184**(1), 247–267.
- Marone, F. & Romanowicz, B., 2007. Non-linear crustal corrections in high-resolution regional waveform seismic tomography, *Geophys. J. Int.*, **170**, 460–467.
- Marone, F., van der Lee, S. & Giardini, D., 2004. Three-dimensional upper-mantle S-velocity model for the Eurasia-Africa plate boundary region, *Geophys. J. Int.*, **158**, 109–130.
- Mégnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface, and higher-mode waveforms, *Geophys. J. Int.*, **143**, 709–728.
- Meier, T., Lebedev, S., Nolet, G. & Dahlen, F.A., 1997. Diffraction tomography using multimode surface waves, *J. geophys. Res.*, **102**(B4), 8255–8267.
- Meltzer, A. et al., 1999. USArray Initiative, *GSA Today*, **9**, 8–10.
- Michibayashi, K. et al., 2016. Natural olivine crystal-fabrics in the western Pacific convergence region: A new method to identify fabric type, *Earth planet. Sci. Lett.*, **443**, 70–80.
- Modrak, R. & Tromp, J., 2016. Seismic waveform inversion best practices: regional, global and exploration test cases, *Geophys. J. Int.*, **206**, 1864–1889.
- Montagner, J.-P. & Anderson, D.L., 1989. Petrological constraints on seismic anisotropy, *Phys. Earth planet. Inter.*, **54**(1–2), 82–105.
- Montagner, J.-P. & Nataf, H.-C., 1988. Vectorial tomography - I. Theory, *Geophys. J. Int.*, **94**(2), 295–307.
- Nocedal, J. & Wright, S.J., 2006. *Numerical Optimization*, Springer.
- Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs, *J. geophys. Res.*, **95**(B6), 8499–8512.
- Nolet, G., 2008. *A Breviary of Seismic Tomography*, Cambridge Univ. Press.
- Paige, C.C. & Saunders, M.A., 1982. Algorithm 583: LSQR: sparse linear equations and least squares problems, *ACM Trans. Math. Softw.*, **8**(2), 195–209.
- Panning, M. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.*, **167**(1), 361–379.
- Panning, M.P., Lekić, V. & Romanowicz, B.A., 2010. Importance of crustal corrections in the development of a new global model of radial anisotropy, *J. geophys. Res.*, **115**, B12325, doi:10.1029/2010JB007520.
- Peacock, S.M. & Wang, K., 1999. Seismic consequences of warm versus cool subduction metamorphism: examples from southwest and northeast Japan, *Science*, **286**, 937–939.
- Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**, 495–503.
- Priestley, K., MacKenzie, D. & Ho, T., 2019. *A lithosphere-asthenosphere boundary—a global model derived from multimode surface-wave tomography and petrology*, chap. 6, pp. 111–123, in: *Lithospheric Discontinuities*, eds. Yuan, H. and Romanowicz, B., American Geophysical Union.
- Priestley, K., Ho, T. & McKenzie, D., 2021. The formation of continental roots, *Geology*, **49**(2), 190–194.
- Rawlinson, N. & Spakman, W., 2016. On the use of sensitivity tests in seismic tomography, *Geophys. J. Int.*, **205**, 1221–1243.
- Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep Earth, *Phys. Earth planet. Inter.*, **178**, 101–135.
- Ritsema, J., Deuss, A., van Heijst, H.J. & Woodhouse, J.H., 2011. S4ORTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltimes and normal-mode splitting function measurements, *Geophys. J. Int.*, **184**, 1223–1236.
- Robertson, G.S. & Woodhouse, J.H., 1995. Evidence for proportionality of P and S heterogeneity in the lower mantle, *Geophys. J. Int.*, **123**, 85–116.
- Romanowicz, B., 2003. Global mantle tomography: Progress status in the past 10 years, *Annu. Rev. Earth planet. Sci.*, **31**, 303–328.
- Romanowicz, B. & Giardini, D., 2001. The future of permanent seismic networks, *Science*, **293**(5537), 2000–2001.
- Rösler, B. & van der Lee, S., 2020. Using seismic source parameters to model frequency-dependent surface-wave radiation patterns, *Seismol. Res. Lett.*, **91**(2A), 992–1002.
- Rudnick, R.L. & Fountain, D.M., 1995. Nature and composition of the continental crust: a lower crustal perspective, *Rev. Geophys.*, **33**(3), 267–309.
- Russell, D.R., Herrmann, R.B. & Hwang, H.-J., 1988. Application of frequency variable filters to surface-wave amplitude analysis, *Bull. seism. Soc. Am.*, **78**(1), 339–354.
- Schaeffer, A.J. & Lebedev, S., 2013. Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.*, **194**, 417–449.
- Schaeffer, A.J. & Lebedev, S., 2014. Imaging the North American continent using waveform inversion of global and USArray data, *Earth planet. Sci. Lett.*, **403**, 26–41.

- Schaeffer, A.J., Lebedev, S. & Becker, T.W., 2016. Azimuthal seismic anisotropy in the Earth's upper mantle and the thickness of tectonic plates, *Geophys. J. Int.*, **207**, 901–933.
- Schmid, C., van der Lee, S., VanDecar, J.C., Engdahl, E.R. & Giardini, D., 2008. Three-dimensional S velocity of the mantle in the Africa-Eurasia plate boundary region from phase arrival times and regional waveforms, *J. geophys. Res.*, **113**(B3), doi:10.1029/2005jb004193.
- Scripps Institution of Oceanography, 1986. *Global Seismograph Network - IRIS/IDA [Data set]*, International Federation of Digital Seismograph Networks, doi:10.7914/SN/II.
- Shapiro, N.M., Ritzwoller, M.H., Molnar, P. & Levin, V., 2004. Thinning and flow of Tibetan crust constrained by seismic anisotropy, *Science*, **305**(5681), 233–236.
- Shen, W. *et al.*, 2016. A seismic reference model for the crust and uppermost mantle beneath China from surface wave dispersion, *Geophys. J. Int.*, **2016**, 954–979.
- Sieminski, A., L  v  que, J.-J. & Debayle, E., 2004. Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.*, **31**, L24614, doi:10.1029/2004GL021402.
- Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.*, **168**, 1153–1174.
- Simmons, N.A., Myers, S.C., Johannesson, G. & Matzel, E., 2012. LLNL-G3Dv3: Global P wave tomography model for improved regional and teleseismic travel time prediction, *J. geophys. Res.*, **117**(B10302), 1–28.
- Simons, F.J., Nolet, G., Georgieff, P., Babcock, J.M., Regier, L.A. & Davis, R.E., 2009. On the potential of recording earthquakes for global seismic tomography by low-cost autonomous instruments in the oceans, *J. geophys. Res.*, **114**(B5), doi:10.1029/2008JB006088.
- Simut  , S., Steptoe, H., Cobden, L., Gokhberg, A. & Fichtner, A., 2016. Full-waveform inversion of the Japanese Islands region, *J. geophys. Res.*, **121**, 3722–3741.
- Song, J.-H., Kim, S. & Rhie, J., 2020. Heterogeneous modification and reactivation of a craton margin beneath the Korean Peninsula from teleseismic travel time tomography, *Gondwana Res.*, **81**, 475–489.
- Storchak, D.A., Giacomo, D.D., Bond  r, I., Engdahl, E.R., Harris, J., Lee, W. H.K., nor, A.V. & Bormann, P., 2013. Public release of the ISC-GEM Global Instrumental Earthquake Catalogue, *Seismol. Res. Lett.*, **84**(5), 810–815.
- Takeuchi, H. & Saito, M., 1972. Seismic surface waves, *Methods Comput. Phys.*, **11**, 217–295.
- Tanimoto, T., 1991. Waveform inversion for three-dimensional density and S wave structure, *J. geophys. Res.*, **96**(B5), 8167–8189.
- Tao, K., Grand, S.P. & Niu, F., 2018. Seismic structure of the upper mantle beneath eastern Asia from full waveform seismic tomography, *Geochem. Geophys. Geosyst.*, **19**, 2732–2763.
- 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 Estimates*, SIAM.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- van der Hilst, R.D., Widiyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**, 578–584.
- van der Lee, S., 1998. Observations and origins of Rayleigh-wave amplitude anomalies, *Geophys. J. Int.*, **135**, 691–699.
- van der Lee, S. & Frederiksen, A., 2005. *Surface wave tomography applied to the North American upper mantle*, pp. 67–80, in: *Seismic Earth: Array Analysis of Broadband Seismograms*, eds. Levander, A. and Nolet, G., American Geophysical Union (AGU).
- van der Lee, S. & Nolet, G., 1997. Upper mantle S velocity structure of North America, *J. geophys. Res.*, **102**(B10), 22815–22838.
- van der Lee, S., James, D. & Silver, P., 2001. Upper-mantle S-velocity structure of central and western South America, *J. geophys. Res.*, **106**(12), 30821–30835.
- van Heijst, H.J. & Woodhouse, J., 1997. Measuring surface-wave overtone phase velocities using a mode-branch stripping technique, *Geophys. J. Int.*, **131**, 209–230.
- Wang, Z. & Dahlen, F.A., 1995. Spherical-spline parameterization of three-dimensional earth models, *Geophys. Res. Lett.*, **22**(22), 3099–3102.
- Wang, Z. & Zhao, D., 2021. 3D anisotropic structure of the Japan subduction zone, *Sci. Adv.*, **7**(4), eabc9620, doi:10.1126/sciadv.abc9620.
- Wessel, P., Luis, J.F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H.F. & Tian, D., 2019. The Generic Mapping Tools version 6, *Geochem. Geophys. Geosyst.*, **20**, 5556–5564.
- Willemann, R.J. & Storchak, D.A., 2001. Data collection at the International Seismological Centre, *Seismol. Res. Lett.*, **72**(4), 440–453.
- Witek, M., Chang, S.-J., Lim, D., Ning, S. & Ning, J., 2021. Radial anisotropy in East Asia from multimode surface wave tomography, *J. geophys. Res.*, **126**, doi:10.1029/2020JB021201.
- Wolin, E., van der Lee, S., Bollmann, T.A., Wiens, D.A., Revenaugh, J., Darbyshire, F.A., Frederiksen, A.W., Stein, S. & Wysession, M.E., 2015. Seasonal and diurnal variations in long-period noise at SPREE stations: the influence of soil characteristics on shallow stations' performance, *Bull. seism. Soc. Am.*, **105**(5), 2433–2452.
- Woodhouse, J.H., 1988. *The calculation of eigenfrequencies and eigenfunctions of the free oscillations of the Earth and the Sun*, pp. 321–370, in: *Seismological Algorithms*, ed. Doornbos D.J., Academic Press.
- 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.
- Xia, S., Zhao, D., Qui, X., Nakajima, J., Matsuzawa, T. & Hasegawa, A., 2007. Mapping the crustal structure under active volcanoes in central Tohoku, Japan using P and PmP data, *Geophys. Res. Lett.*, **34**, L10309, doi:10.1029/2007GL030026.
- Zhao, D., 2004. Global tomographic images of mantle plumes and subducting slabs: insight into deep Earth dynamics, *Phys. Earth planet. Inter.*, **146**(1–2), 3–34.
- Zhao, D., Maruyama, S. & Omori, S., 2007. Mantle dynamics of Western Pacific and East Asia: Insight from seismic tomography and mineral physics, *Gondwana Res.*, **11**, 120–131.
- Zheng, Y., Shen, W., Zhou, L., Yang, Y., Xie, Z. & Ritzwoller, M.H., 2011. Crust and uppermost mantle beneath the North China Craton, north-eastern China, and the Sea of Japan from ambient noise tomography, *J. geophys. Res.*, **116**, B12312, doi:10.1029/2011JB008637.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Same as Figure 6, except $\delta \ln \phi / \delta \ln \xi = -1.5$ during waveform inversion.

Figure S2. Same as Figure 6, except $\delta \ln \phi / \delta \ln \xi = 1.5$ during waveform inversion.

Figure S3. Effects of including ξ perturbations that are correlated with ϕ perturbations, $\delta \ln \phi / \delta \ln \xi = 0$ during waveform inversion.

Figure S4. Same as Figure S3, but $\delta \ln \phi / \delta \ln \xi = -1.5$ during waveform inversion.

Figure S5. Same as Figure S3, but $\delta \ln \phi / \delta \ln \xi = 1.5$ during waveform inversion.

Figure S6. Distribution of center frequencies used in waveform fitting as a function of distance.

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 paper.