Title
Upper mantle radial anisotropy under the Indian Ocean from higher mode surface waves and a hierarchical transdismensional approach

Journal
Geophysical Journal International, 228(1)

ISSN
0956-540X

Authors
Weidner, Erik
Beghein, Caroline
Huang, Quancheng
et al.

Publication Date
2021-09-10

DOI
10.1093/gji/ggab340

Geophysical Journal International

Geophys. J. Int. (2022) **228,** 78–101
Advance Access publication 2021 August 20
GJI Seismology

https://doi.org/10.1093/gji/ggab340

Upper mantle radial anisotropy under the Indian Ocean from higher mode surface waves and a hierarchical transdismensional approach

Erik Weidner [1], Caroline Beghein [1], Quancheng Huang [2,3] and Nicholas Schmerr [2]

[1] *Department of Earth, Planetary, and Space Sciences, UCLA, Los Angeles,* 90095 *CA, USA. E-mail: eweidner@ucla.edu*
[2] *Department of Geology, University of Maryland, College Park, MD* 20742, *USA*
[3] *Department of Physics, New Mexico State University, Las Cruces, NM,* 88003, *USA*

Accepted 2021 August 17. Received 2021 July 6; in original form 2021 January 26

SUMMARY

We investigated the likelihood of radial anisotropy in the shallow and deep upper mantle, including the mantle transition zone (MTZ) under the Indian Ocean. Seismic anisotropy can be an indicator of mantle deformation through lattice preferred orientation of anisotropic crystals in the mantle. It has thus the potential to illuminate Earth's dynamic interior, but previous seismic tomography studies have not achieved consensus on the existence of radial anisotropy below ∼250 km depth. We developed a fully nonlinear transdimensional hierarchical Bayesian Markov Chain Monte Carlo approach to invert fundamental and higher mode surface wave dispersion data and applied it to a subset of a global Love and Rayleigh wave data set. We obtained posterior model parameter distributions for shear wave velocity (V_S) and radial anisotropy ξ under the Indian Ocean. These posterior model distributions were used to calculate the probability of having radial anisotropy at different depths. We demonstrated that separate inversions of Love and Rayleigh waves yield models compatible with the results of joint inversions within uncertainties. The obtained pattern of V_S anomalies agrees with most previous studies. They display negative anomalies along ridges in the uppermost mantle, but those are stronger than for regularized inversions. The Central Indian Ridge and the Southeastern Indian Ridge present velocity anomalies that extend to ∼200 km depth, whereas the Southwestern Indian Ridge seems to have a shallower origin. Weaker, laterally variable velocity perturbations were found at larger depths. The anisotropy models differ more strongly from regularized inversion results and their uncertainties were rather large. We found that anisotropy models from regularized inversions also depend on the chosen parametrization, which is consistent with the existence of a large model null-space. Apart from a fast horizontally polarized shear wave signal in the top 100 km, likely reflecting the horizontal plate motion due to asthenospheric deformation, no clear relation to surface geology was detected in the anisotropy models. We found that, although the anisotropy model uncertainties are rather large, and lateral variations are present, the data generally prefer at least 1 per cent anisotropy in the MTZ with fast vertically polarized shear waves, within errors. Incorporating group velocity data did not help better constrain deep structure by reducing parameter trade-offs. We also tested the effect of prior constraints on the 410- and 660-km topography and found that the undulations of these discontinuities had little effect on the resulting models in our study region.

Key words: Inverse theory; Probability distributions; Seismic anisotropy; Seismic tomography; Surface waves and free oscillations.

1 INTRODUCTION

Seismic anisotropy, which is the directional dependence of wave propagation, can provide clues to the properties and dynamics of the mantle. Radial anisotropy is the change in seismic wave velocity between the horizontal and vertical direction. Seismic anisotropy can occur via shape preferred orientation (SPO), which occurs where multiple isotropic mediums are stacked on top of each other, such as melt pockets and cracks (Kendall & Silver 1998; Faccenda *et al.* 2008; Sakamaki *et al.* 2013). Although the individual mediums are isotropic, the bulk medium has directional dependence. Seismic anisotropy can also develop due to the crystallographic or lattice preferred orientation (CPO or LPO) of intrinsically anisotropic crystals. If such crystals can deform by dislocation creep, they may

partially align within the strain field caused by mantle convection, resulting in bulk anisotropy that may be observable at the seismic scale. There are abundant observations of seismic anisotropy in the crust and upper mantle, both globally and regionally. Radial anisotropy was first proposed by Anderson (1962) to explain apparent discrepancies between shear wave velocity models derived from inversions of Love wave dispersion curves and those resulting from Rayleigh wave data. The Preliminary Reference Earth Model (PREM, Dziewonski & Anderson 1981) was the first global model to include this type of anisotropy in the top 220 km of the mantle. Since then, many global and regional studies have found evidence for its presence in the uppermost mantle (Montagner & Tanimoto 1991; Shapiro & Ritzwoller 2002; Visser *et al.* 2008a).

LPO of olivine is the prevailing explanation for the existence of seismic anisotropy at these depths due to the abundance of this mineral in the mantle and its high single crystal anisotropy (Mainprice *et al.* 2005). Mantle convection is thought to be the origin of LPO in the upper mantle, which is supported by observations of the seismically fast axes in the asthenosphere aligning with absolute plate motion (Becker *et al.* 2003; Debayle & Ricard 2013; Beghein *et al.* 2014; Burgos *et al.* 2014). Therefore, improving our constraints on seismic anisotropy is important to better understand the nature of mantle convection. In particular, determining whether it is present in the mantle transition zone (MTZ) is important since this region likely plays an important role in Earth's thermal evolution, but it is a challenging problem. Improving constraints on upper mantle anisotropy can also help determine the depth of the lithosphere–asthenosphere boundary (LAB), which directly relates to the fundamental concept of plate tectonics. Additionally, vertical changes in seismic anisotropy can indicate layering in the upper mantle. This, in turn, can provide insight on the LAB depth and can indicate the existence of compositional boundaries (Smith *et al.* 2004; Yuan & Romanowicz 2010; Beghein *et al.* 2014).

While the presence of seismic anisotropy in the upper 250 km of the mantle, in the lowermost mantle, and in the inner core is well established, its existence in the rest of the mantle is still debated. For decades, seismic anisotropy had not been observed in the deep upper mantle or in most of the lower mantle, which was interpreted as an indication that deformation at those depths did not produce LPO (Karato & Wu 1993). More recently, however, radial and azimuthal anisotropy observations below 250 km depth, including in and below the MTZ, have challenged this view (e.g. Gung *et al.* 2003; Panning & Romanowicz 2006; Yuan & Beghein 2013, 2014; Lynner & Long 2015; Huang *et al.* 2019). Constraining seismic anisotropy in the deep upper mantle is, however, not straightforward. Body waves do not generally have good vertical resolution as they travel quasi-vertically. Some shear wave splitting analyses using source-side measurements have been able to identify seismic anisotropy originating from large depths, though they cannot separate the contribution of the uppermost mantle from that of the transition zone and lower mantle (Lynner & Long 2014, 2015). Tomographic studies that include higher mode surface wave dispersion and/or long-period body waveforms generally result in models with better depth resolution, but lower lateral resolution. Several recent publications have reported seismic anisotropy of up to a few percents in the transition zone (e.g. Panning & Romanowicz 2006; Yuan & Beghein 2013; Auer *et al.* 2014; Moulik & Ekström 2014; Yuan & Beghein 2014; Chang *et al.* 2015; Ferreira *et al.* 2019), but there are large discrepancies between models at those depths (Schaeffer *et al.* 2016). Differences in the models can arise from differences in the data sets, but also from different inversion techniques and applied regularizations.

Tomographic inverse problems are typically solved by trying to minimize a cost function that measures the distance between data and model predictions in a least-squares sense. They are, however, generally not well posed and can thus be highly non-unique, that is, several solutions can explain the data equivalently well. Traditional inverse methods deal with this non-uniqueness by imposing subjective regularizations, but those can have significant impacts on the calculated solution. These regularizations include the choice of the reference model, the model parametrization, and/or the choice of *ad hoc* damping parameters that compromise between minimizing the misfit and the size of the model (Trampert 1998). In addition, regularized inversions can underestimate model uncertainties as the posterior model covariance is generally smaller than the prior by construction (Tarantola 2005). In addition, resolution analyses for regularized inverse methods depend on the regularization, which may give a false sense of certainty in some model parameters. It should also be noted that errors in tomographic models can be introduced by errors or approximations in the forward theory. One common example in seismic tomography is solving linearized equations instead of the full nonlinear problem.

In this study, in order to avoid some of these issues, we developed a Hierarchical Transdimensional Bayesian (TB) inverse method (Sambridge *et al.* 2005; Bodin *et al.* 2012, 2016; Gao & Lekić 2018), which uses a reversible jump Monte Carlo Markov Chain (rj-MCMC) algorithm (Green 1995, 2003; Gallagher *et al.* 2009) to solve the nonlinear forward problem associated with calculating higher mode surface wave dispersion from an interior model. With advances in computing, Monte Carlo methods, which repetitively solve the forward problem, have become more and more common in geophysics and seismic tomography in particular. An advantage of these techniques is that they can solve nonlinear problems directly instead of relying on a linearized approximation of the problem. In addition, besides the choice of the prior model parameter distribution, they do not require as much regularization as traditional inverse methods. By testing thousands or millions of possible models, they can provide an ensemble of solutions that fit the data instead of choosing a single one with a subjective regularization. This provides quantitative posterior model uncertainties and allows for an estimate of the probability of different values of the model parameters. In our study, we used the resulting parameter distributions to determine the likelihood of the presence of radial anisotropy at different depths in the upper mantle.

It should be noted that a different fully nonlinear approach to the same problem was previously taken by Visser *et al.* (2008a) to determine the likelihood of radial anisotropy in the top 1400 km of the mantle at the global scale. They had found that 1–2 per cent radial anisotropy with vertically polarized shear waves travelling faster than horizontally polarized shear waves was likely in the MTZ globally. Our technique has, however, several possible advantages compared to their work. First, it allows for the depth parametrization to change during the model space search, letting the data guide how complicated the model needs to be instead of imposing a fixed set of basis functions *a priori*. Second, the hierarchical nature of our MCMC algorithm enables us to include the data noise among the unknowns. Using incorrect data noise can negatively impact the accuracy of inversion results (Bodin *et al.* 2012). Instead, by inverting for data noise, we allow the algorithm to decide how much noise is needed in order to explain the measurements without overfitting them.

In this manuscript, we first present the data and method employed. Second, we demonstrate the dependence of radial anisotropy models on the regularization imposed when performing linearized inversions. Third, we apply our TB technique to a fundamental and higher mode surface wave data set sampling the Indian Ocean and compare the results with those of a regularized inversion. Fourth, we test the performance of the method with synthetic tests. Finally, we briefly discuss the results before drawing conclusions.

2 DATA

The measurements used in this study are the Rayleigh and Love wave fundamental and higher mode phase velocity maps measured by Visser *et al.* (2008b). The authors used a model space search to perform those measurements, which allowed them to also calculate quantitative data uncertainties. Both Rayleigh and Love waves are necessary to constrain radial anisotropy as they each are sensitive to separate shear wave velocity polarization directions. Rayleigh waves are primarily sensitive to vertically polarized shear wave speed (V_{SV}), and Love waves are primarily sensitive to horizontally polarized shear wave speed (V_{SH}). The Voigt average bulk shear wave speed (V_S) and radial anisotropy (ξ) can be calculated from V_{SV} and V_{SH} (Babuska & Cara 1991):

$$V_S = \sqrt{(2V_{SV}^2 + V_{SH}^2)/3} \tag{1}$$

$$\xi = V_{SH}^2 / V_{SV}^2 \tag{2}$$

Both wave types are also sensitive to density, ρ, and Rayleigh waves are additionally sensitive to elastic parameters relating to compressional wave velocities. However, due to the existence of parameter trade-offs, Rayleigh waves and Love waves can only resolve shear wave parameters. We thus inverted explicitly only for the two *S*-wave velocity parameters, V_{SH} and V_{SV}, and scaled the remaining parameters, as detailed in Section 3.

Visser *et al.* (2008b) measured phase velocity dispersion for branches up to the sixth overtone for Rayleigh waves and up to the fifth overtone for Love waves. Including higher mode data is critical to constrain elastic structure beyond the uppermost 200 km of the mantle because they are sensitive to deeper structure than fundamental modes at the same period. Sensitivity curves for V_S and ξ for this data set are displayed in Fig. 1 at periods between 35 and 175 s. They show that the higher modes are much more sensitive to structure below 200 km than the fundamental modes, and demonstrate that the data set used in this study is sensitive to structure well below the bottom of the MTZ. For this work, we employed the commonly used approach of discretizing the phase velocity maps into $5^\circ \times 5^\circ$ cells, with each cell being assigned velocities based on the values of the maps at the centre of the grid cell. The surface wave velocities were inverted at each of the resulting 238 grid cells separately using different methods, as described in Section 3, to obtain 1-D shear wave velocity depth profiles. The obtained individual 1-D velocity profiles were then used to construct shear wave velocity and anisotropy maps.

One caveat of discretizing phase velocity maps, which result from the linearized inversion of path-averaged measurements, is that it implicitly assume each grid cell is independent of the other. However, some trade-offs can be expected between structure along paths, especially where coverage is imperfect. In addition, the inherent differences in station-quake distribution for Love and Rayleigh can yield differences in the lateral resolution, but in our case it will mostly affect the small-scale features of the phase velocity maps. Visser *et al.* (2008b) indeed determined that their fundamental modes had generally good coverage and they estimated the lateral resolution of both their fundamental mode Love and Rayleigh wave phase velocity maps to correspond to spherical harmonic degree 25. So, overall the two types of data have similar resolution and it should not affect our discussion. A greater concern might be that the resolution of the fundamental mode phase velocity maps is greater than that of the overtones. Visser *et al.* (2008b) showed that the higher modes path coverage was very good in the Northern Hemisphere, but the Southern Hemisphere, where our study region is located, may suffer from poorer lateral resolution. In addition, the authors chose their regularization to insure that the relative model uncertainty remains constant for all modes. They explained that they opted for a compromise between aiming for a constant resolution for all modes and an increasing uncertainty for higher modes. A constant resolution would have been preferable for depth inversions but they opted against it due to the large difference in the number of data between modes. So, while we do not think these differences in resolution affect the discussion of the longer wavelength features we highlight in our paper, the reader should keep these caveats in mind.

3 METHOD

In this study, we compared results from regularized linear inversions of phase velocity maps with those obtained with a fully nonlinear Bayesian approach. Both methods are described below.

3.1 Regularized inversion

The relation between perturbations in phase velocity, c and perturbations in the parameters $\mathbf{m}$ describing the physical state of Earth's interior can be written as follows (Woodhouse & Dahlen 1978):

$$\delta d(T) = \int_0^R K_{\mathbf{m}}(T, r)\delta\mathbf{m}(r)\, dr + \sum_b \delta d_b\, K_b(T) \tag{3}$$

where R is the radius of the Earth, δd represents the phase or group velocity data, T is the period considered and b is the boundary radius. $K_{\mathbf{m}}(T, r)$ is a sensitivity kernel that maps perturbations to the elastic model $\mathbf{m}$ at depth r to velocity change. K_b is a boundary kernel for internal boundary b that relates perturbations in the depth d_b of the boundary to a perturbation in phase velocity. The vector $\mathbf{m}$ contains all the model properties, that is, density and elastic parameters, and a sum over each one of them is implicit. $K_{\mathbf{m}}$ are sensitivity kernels, or partial derivatives, for model perturbations $\delta\mathbf{m}$. An identical relation exists for group velocities.

When constraining radial anisotropy, one needs to account for five elastic parameters in addition to density. These elastic parameters are : L and N, which relate to the speed of vertically and horizontally polarized shear waves, respectively ($V_{SV}^2 = L/\rho$ and $V_{SH}^2 = N/\rho$); C and A, which relate to P-wave speed for waves travelling vertically and horizontally, respectively ($V_{PV}^2 = C/\rho$ and $V_{PH}^2 = A/\rho$); and F, which describes waves travelling at intermediate angle. If we define $\eta = F/(A - 2L)$, we thus have (Montagner 1986):

$$\begin{aligned}\delta d(T) = \int_0^R & [K_{V_{SV}}(T, r)\delta V_{SV}(r) + K_{V_{SH}}(T, r)\delta V_{SH}(r) \\ & + K_{V_{PV}}(T, r)\delta V_{PV}(r) + K_{V_{PH}}(T, r)\delta V_{PH}(r) \\ & + K_\eta(T, r)\delta\eta(r) + K_\rho(T, r)\delta\rho(r)]\, dr \\ & + \sum_b \delta d_b\, K_b(T)\end{aligned} \tag{4}$$

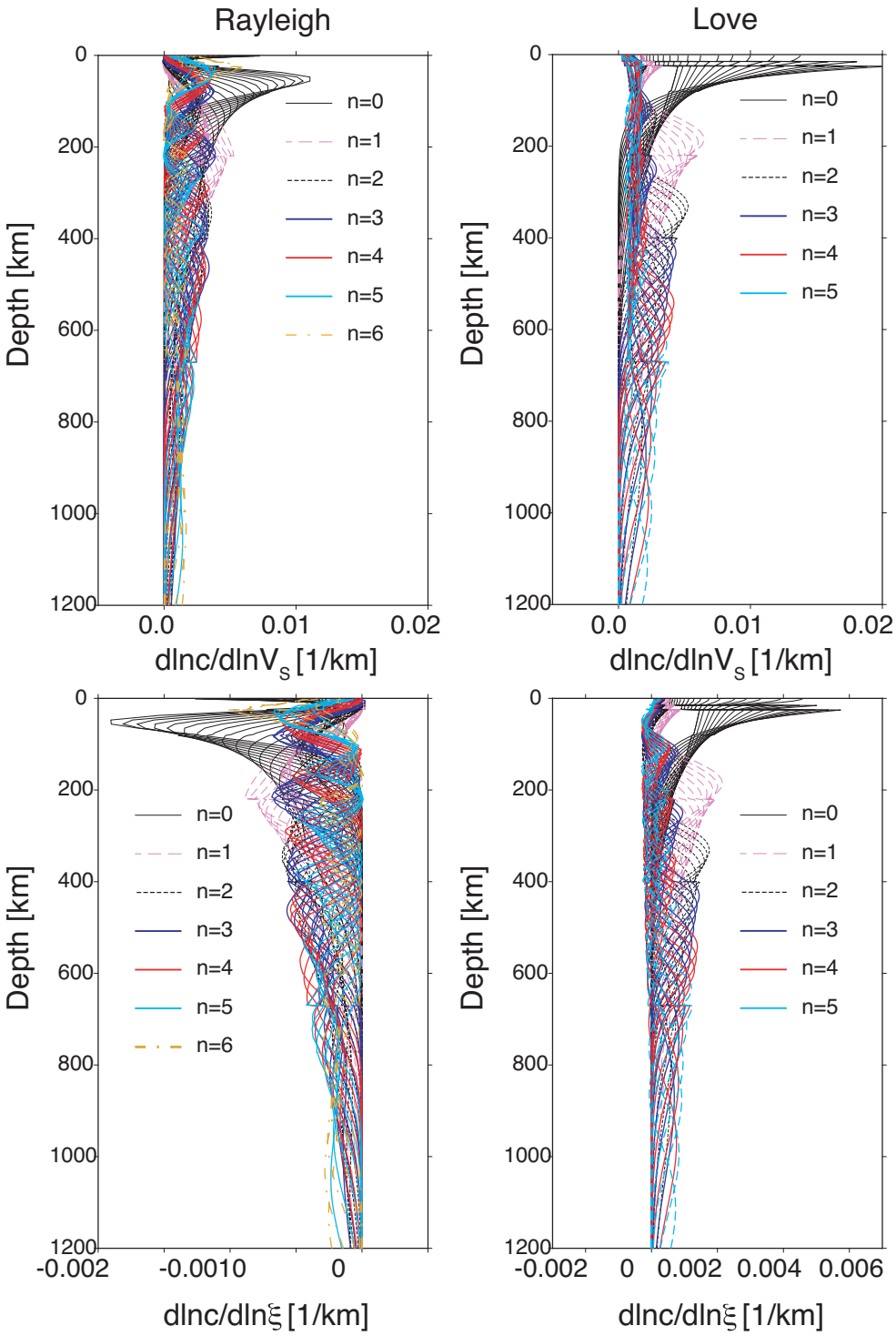

Figure 1. Rayleigh (left) and Love (right) wave phase velocity (c) sensitivity kernels for perturbations in V_S (top) and ξ (bottom) between 35 and 175 s period for fundamental ($n = 0$) and higher modes ($n > 0$).

Among these six parameters, only the two shear wave related parameters can be reasonably well resolved with surface waves (e.g. Beghein & Trampert 2004). It is customary to assume scaling relationships for the other parameters and to invert for the shear wave parameters only. Such relationships were derived by Montagner & Anderson (1989) for the uppermost 200 km of the mantle and are commonly used in surface wave inversions for radial anisotropy (e.g. Panning & Romanowicz 2006; Xing & Beghein 2015):

$$\frac{\delta \ln V_P}{\delta \ln V_S} = 0.5 \tag{5}$$

$$\frac{\delta \ln \rho}{\delta \ln V_S} = 0.33 \tag{6}$$

$$\frac{\delta \ln \phi}{\delta \ln \xi} = -2.5 \tag{7}$$

$$\frac{\delta \ln \eta}{\delta \ln \xi} = -1.5 \tag{8}$$

While these relations are strictly only valid for the shallow upper mantle, studies have shown that using them for the deeper mantle does not affect the shear wave velocity or anisotropy model significantly (Panning & Romanowicz 2006). We thus applied them to the entire depth range of interest. Eq. (3) then becomes:

$$\delta d(T) = \int_0^R \left[K'_{V_{SV}}(T,r)\delta V_{SV}(r) + K'_{V_{SH}}(T,r)\delta V_{SH}(r)\right] dr + \sum_b \delta d_b \, K_b(T) \tag{9}$$

where the K' kernels are linear combinations of the kernels in eq. (4). The problem can alternatively be described in terms of perturbations in bulk shear wave velocity and anisotropy:

$$\delta d(T) = \int_0^R \left[K_{V_S}(T,r)\delta V_S(r) + K_\xi(T,r)\delta\xi(r)\right] dr + \sum_b \delta d_b \, K_b(T) \tag{10}$$

where V_S and ξ are defined in eqs (1) and (2).

These equations can be discretized by introducing a basis function to parametrize the depth dependence of the model parameters. We chose a basis function in terms of J splines S_j:

$$\mathbf{m}(r) = \sum_{j=1}^{J} \mathbf{m}^j S_j(r) \tag{11}$$

eq. (3) then becomes:

$$\delta d(T) = \sum_{i=1}^{2}\sum_{j=1}^{J} m_i^j \int_0^R K_{m_i}(T,r) S_j(r)\, dr + \sum_b \delta d_b \, K_b(T) \tag{12}$$

where index i is for each one of the two elastic parameters we solve for. The forward problem can thus be written as:

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{13}$$

where $\mathbf{d}$ is the data vector. $\mathbf{G}$ is the kernel matrix that maps the model space into the data space:

$$G_{kj} = \sum_i \int_0^R K^k_{m_i}(r) S_j(r) dr \tag{14}$$

with G_{kj} an element of $\mathbf{G}$, and $K^k_{m_i}(r)$ the sensitivity kernel for the ith model parameter in $\mathbf{m}$ and the kth measurement in $\mathbf{d}$.

A generalized inverse matrix, $\mathbf{G}^g$, can be calculated via singular value decomposition (SVD) and can be used to determine a least-squares solution $\mathbf{m}$ (Jackson 1972; Lanczos 1961; Wiggins 1972). If $\mathbf{G}$ is a $n \times m$ matrix where n is the number of data points and m is the number of model parameters, it can be decomposed into:

$$\mathbf{G} = \mathbf{U}\Lambda\mathbf{V}^\mathbf{T} \tag{15}$$

where $\mathbf{U}$ is a $n \times n$ matrix of eigenvectors that span the data space, Λ is a $n \times m$ diagonal matrix whose columns are non-negative eigenvalues λ_i and $\mathbf{V}$ is a $m \times m$ matrix of eigenvectors that span the model space. The λ_i^2 are the singular values of $\mathbf{G}$. The generalized inverse of $\mathbf{G}$ is thus:

$$\mathbf{G}^g = \mathbf{V}_p \Lambda_p^{-1} \mathbf{U}_p^T \tag{16}$$

where p is the number of non-zero eigenvalues. The estimated model parameters $\mathbf{m}^{est}$ are then given by a sum limited to the non-zero eigenvalues to insure stability of the solution:

$$\mathbf{m}^{est} = \mathbf{G}^g\mathbf{d} \tag{17}$$

Here, we applied the method of Matsu'Ura & Hirata (1982) to determine p: we first normalize $\mathbf{G}$ by the data covariance matrix $\mathbf{C}_d$ and a prior model covariance matrix $\mathbf{C}_m$:

$$\mathbf{G}^\dagger = \mathbf{C}_d^{-1}\mathbf{G}\mathbf{C}_m \tag{18}$$

We then calculated its eigenvalues and determined the solution using all the eigenvalues smaller or equal to unity. For $\mathbf{C}_d$, we used a diagonal matrix and the uncertainties estimated by Visser *et al.* (2008b). The choice of model covariance matrix acts as implicit regularization as it yields a different cutoff in the number of eigenvalues (Snieder & Trampert 2000).

In this study, we inverted Love and Rayleigh waves jointly using the SVD technique described above, and compared models obtained with different prior values for $\mathbf{C}_m$ and with different model parametrizations (eqs 9 and 10). The goal was to illustrate the effect of explicit and implicit choices of regularization on the solution. The misfit (ϕ_d) was defined by:

$$\phi_d = \frac{1}{N}\sum_{i=1}^{N}(d_i^{\text{obs}} - d_i^{\text{pre}})^2, \tag{19}$$

where N is the total number of measurements, d_i^{obs} is the ith measured data and d_i^{pre} is the ith predicted data calculated with eq. (12). We followed Xing & Beghein (2015) and corrected for the effect of the crust by applying nonlinear crustal corrections based on the 3-D crustal model CRUST 1.0 (Laske *et al.* 2013). To calculate these nonlinear crustal corrections, we divided the phase velocity maps into $1^\circ \times 1^\circ$ cells and, at each grid cell, we created local 1-D models composed of CRUST1.0 and the PREM mantle. At each grid cell, we then used MINEOS (Masters *et al.* 2011) to predict phase velocities for the local 1-D models and for the pure PREM. The difference between the predicted phase velocities for those two models constitutes the contribution of CRUST1.0 to the phase velocities. Then, it was subtracted from the real measurements so that the corrected data contain information about mantle structure only. The corrected phase velocity maps were divided into $5^\circ \times 5^\circ$ cells. Inversions were performed at each grid cell using sensitivity kernels calculated for the local reference model.

3.2 Transdimensional Hierarchical Bayesian method

While solving the inverse problem using the regularized method described above is relatively simple and efficient, it yields solutions that depend on the regularizations imposed, including the choice of a reference model. In addition, posterior model uncertainties can be underestimated if the model null-space is large (Trampert 1998), and the nonlinear relation between phase velocity and structure are approximated by linearized equations. We thus developed instead a nonlinear model space search approach to find a range of plausible solutions instead of choosing one with a subjective regularization. Our technique is based on the hierarchical TB method of Bodin *et al.* (2012) to explore the model space with an rj-MCMC method, which we combined with Fortran code MINEOS (Masters *et al.* 2011) to solve the forward problem. At each iteration, the fully

nonlinear forward problem is solved by using MINEOS directly to calculate the dispersion, as opposed to using a linearized approach with sensitivity kernels as done in the SVD method. As stated before, some of the advantages of this method compared to traditional inversions include the fact that the depth parametrization can change at each iteration, allowing the data themselves to constrain how complicated the model needs to be. In addition, unknown data noise can be simultaneously inverted for instead of being fixed at a presumed level, which reduces the risk of mapping noise into the model. Another advantage is that model parameters are described by probability density functions (PDFs), which can be used to quantify their uncertainties.

In this study, noise was assumed to be Gaussian, and we chose to parametrize the models as a set of layers. For each layer, a Voronoi nuclei defines its elastic parameter, and the boundaries between layers are the midpoints between adjacent nuclei. The number of points and their depths are allowed to change at each iteration, meaning that the model depth parametrization is a variable of the inversion, in addition to the elastic parameters for each layer. Monte Carlo inversions are generally time-consuming, and this is especially true when solving the forward problem with MINEOS. Minimizing the number of unknowns to solve for is thus important. For this purpose, we inverted Love and Rayleigh waves separately for V_{SH} and V_{SV}, respectively. We verified that the models obtained were equivalent to joint inversion results (see Section 5.1). Scaling relations were imposed on the other elastic parameters as in Section 3.1 (Montagner & Anderson 1989). The resulting distributions of 1-D V_{SH} and V_{SV} profiles were then sampled and used to calculate V_S and ξ distributions (eqs 1 and 2).

With a hierarchical TB approach, model parameters or data noise level are perturbed iteratively, and each new model and data noise combination is either accepted or rejected based on an acceptance criteria as described below. If a change is accepted, the iteration repeats with the modified parameter, and a parameter is randomly selected to be perturbed again. If rejected, the iteration repeats with the model and noise used prior to the modification. Every 100 iterations, the current model is added to the ensemble of solutions. This process is repeated thousands of times in order for the chain to reach convergence, after which the saved ensemble of models gives a posterior for V_{SH} or V_{SV}. For our application, each chain had a burn-in period of 50 000 iterations before models were saved. We then ran enough chains and iterations to get 10 000 samples after the burn-in period.

Data noise perturbations consisted in random values generated from a prior Gaussian distribution that were added to the original noise following level. A model perturbation can consist in moving a Voronoi nucleus, inserting or removing a nucleus, or randomly modifying the elastic parameters at a nucleus using a prior Gaussian distribution. The acceptance or rejection criteria used depends on the likelihood and the prior probability as described in Bodin *et al.* (2012). According to Bayes' (1763) theorem, the posterior probability of having model $\mathbf{m}$ when a set of observed data $\mathbf{d}_{obs}$ is fixed was given by:

$$p(\mathbf{m}|\mathbf{d}_{\text{obs}}) \propto p(\mathbf{d}_{\text{obs}}|\mathbf{m})p(\mathbf{m}), \tag{20}$$

where $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is the likelihood function, which gives the probability of observing data given a particular model, and $p(\mathbf{m})$ is the *a priori* probability of model $\mathbf{m}$, which reflects our prior knowledge on the parameters before having the observed data. The acceptance criteria compares the posterior for the perturbed model to that of the original model. The larger the perturbed model posterior, the higher the likelihood the new model will be accepted. Here, the likelihood function is defined by:

$$p(\mathbf{d}_{\text{obs}}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi\sigma^2)^N}} e^{\frac{-(\mathbf{d}_{\text{pre}}-\mathbf{d}_{\text{obs}})^2}{2\sigma^2}} \tag{21}$$

where $\mathbf{d}_{\text{pre}}$ are the data predicted by model $\mathbf{m}$, N is the total number of measurements and σ is the total data noise level. The prior can be written as:

$$p(\mathbf{m}) = p(\mathbf{c}, \mathbf{v}|k)p(k)p(\mathbf{h}), \tag{22}$$

where $\mathbf{c}$ is the set of depths of the Voronoi nuclei associated with $\mathbf{m}$, $\mathbf{v}$ is the set of wave velocities, k is the number of Voronoi nuclei and $\mathbf{h}$ is the set of noise parameters.

In Bodin *et al.* (2012), the prior $p(\mathbf{c}, \mathbf{v})$ was a uniform distribution that was constant with depth. Thus, $p(\mathbf{c}, \mathbf{v}|k)$ can be written as the product of $p(\mathbf{c}|k)$ and $p(\mathbf{v}|k)$. The prior probabilities of models with the same number of Voronoi nuclei are equal, but they differ when layers are added and removed due to the prior on the velocities:

$$p(\mathbf{v}|k) = \prod_{i=1}^{k} p(v_i|k) \tag{23}$$

$$p(v_i|k) = \begin{cases} \frac{1}{\Delta v} & V_{\min} \leq v_i \leq V_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where v_i is the velocity of the ith Voronoi nuclei and Δv is the range of allowable velocities. In our case, the prior on V_{SV} or V_{SH} is a uniform distribution within 10 per cent of CRUST1.0 in the crust and within 10 per cent of PREM in the mantle. In contrast with our linearized inversions, the data are not corrected for the effect of the crust. Instead, the nonlinear inversions invert for structure in the crust as well as in the mantle. Since velocities are generally increasing with depth, δv increases with depth. This depth-dependent prior means that points at shallower depths will generally have a higher prior probability than points at larger depths.

4 LINEAR INVERSIONS RESULTS

4.1 Effect of explicit regularization and parametrization

Figs 2(a)–(h) and 3(a)–(h) illustrate how the choice of parametrization can affect the models in the case of a traditional inversion method using linearized equations. Figs S1 and S2 in the Supporting Information show the effect of explicit regularization. In one case, we parametrized the models in term of perturbations in V_S and ξ relative to PREM (eq. 10), and in the other case we tested a parametrization in terms of V_{SV} and V_{SH} perturbations (eq. 9), and calculated $d\ln V_S$ and ξ using eqs (1) and (2). We changed the regularization by modifying the prior model covariance matrix $\mathbf{C}_m$ (eq. 18), which resulted in different numbers of independent parameters. Increasing the prior model variance is equivalent to decreasing a damping factor (Snieder & Trampert 2000) and increases the number of independent parameters. The models displayed here were obtained with one iteration of the SVD method. We also tested whether a quasi-nonlinear inversion technique with iterative inversions and sensitivity kernels updated at each iteration would yield different results, but it did not change the models significantly.

For $d\ln V_S$ at 100 km depth, increasing the number of independent parameters does not significantly affect the strength of the anomalies or their overall distribution across the study region (Fig. S1, Supporting Information). In both cases shown, oceanic ridges are associated with negative velocity anomalies and Southeastern Africa and the southern tip of India are characterized by positive velocity anomalies. We found, however, that changing the parametrization (Figs S1

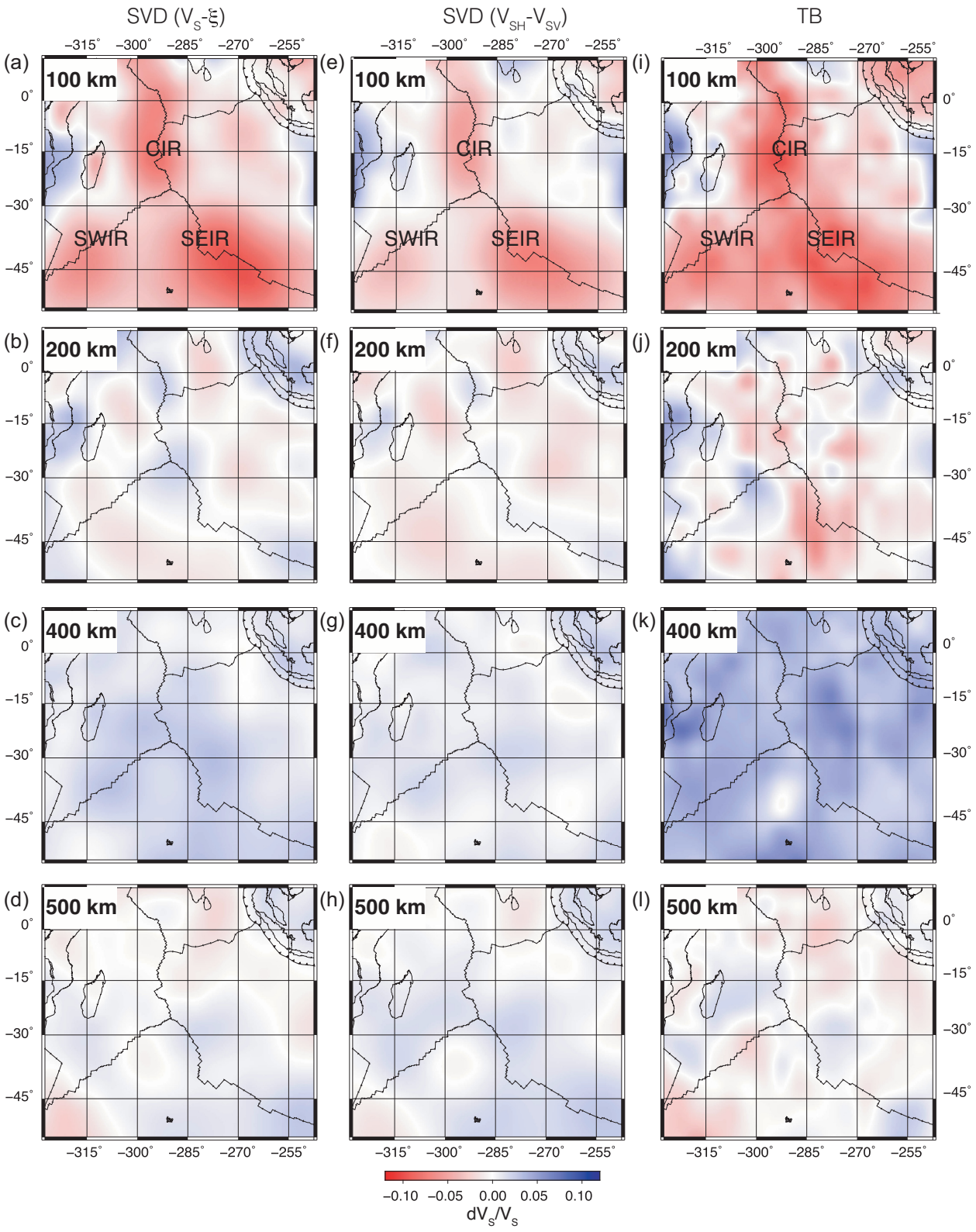

Figure 2. (a)−(h) Bulk shear wave velocity perturbations estimated from the SVD inversion and (i)–(l) from the mean of the posterior distribution obtained with the nonlinear TB inversion. Panels (a)-(d) were obtained using a parametrization in terms of V_S and ξ perturbations. Panels (e)–(h) were obtained using a parametrization in terms of V_{SV} and V_{SH} perturbations. The CIR, SWIR and SEIR branches are labelled in the top panels for reference.

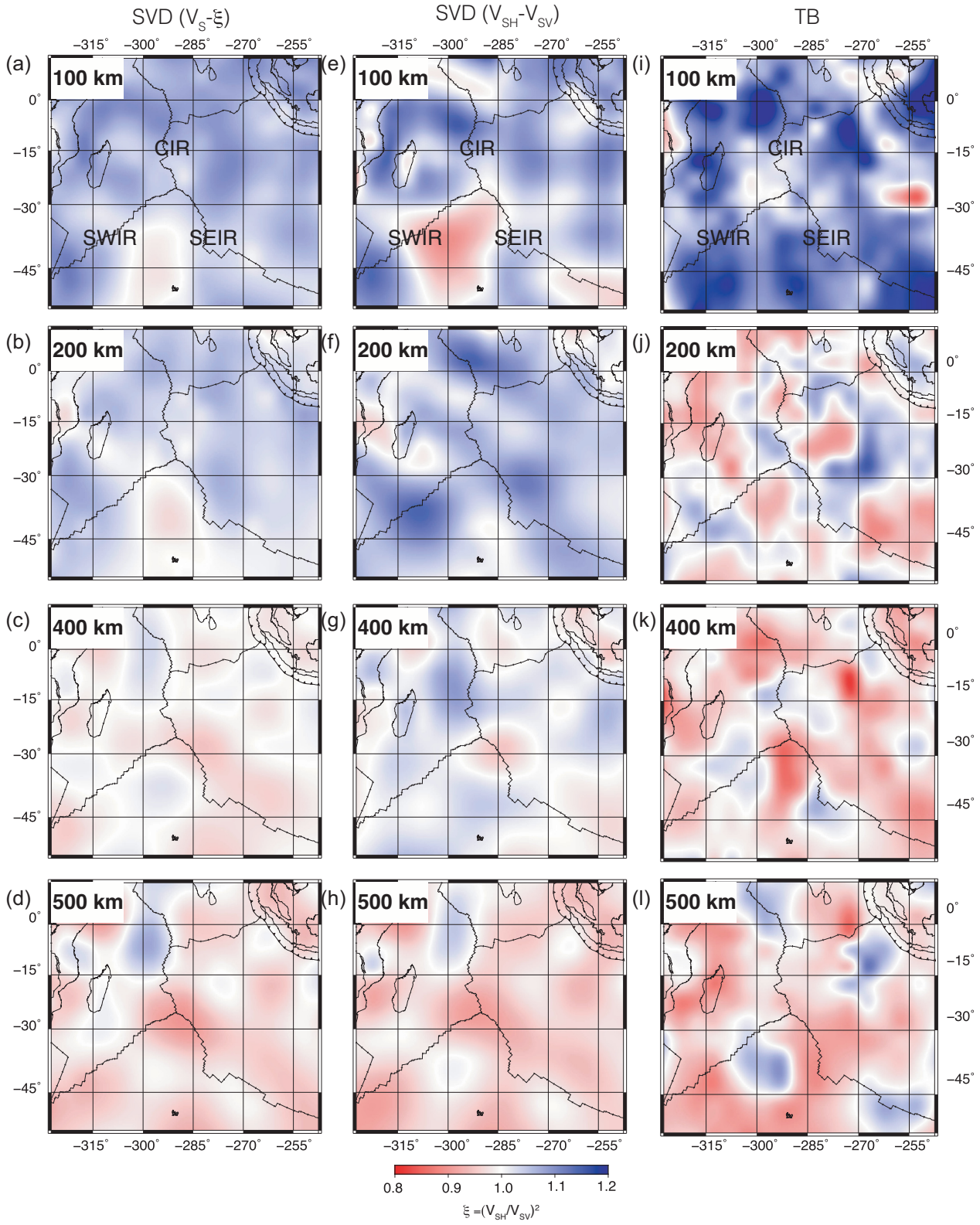

Figure 3. Same as Fig. 2, but for shear wave anisotropy.

and 2, Supporting Information) affects the strength of the $d\ln V_S < 0$ anomalies located under and near mid-ocean ridges: for the same number of independent parameters, the parametrization in terms of ξ and V_S shows stronger negative $d\ln V_S$ than the parametrization in terms of V_{SV} and V_{SH} at 100 km depth (Fig. S1, Supporting Information). In addition, the change in parametrization produces different results under Madagascar where inversions for V_{SH} and V_{SV} yield $d\ln V_S > 0$ whereas inversions for V_S and ξ yield $d\ln V_S < 0$. This

difference is important as it can lead to different interpretations of the results in terms of lithospheric thickness under Madagascar or in terms of temperature under the ridge, for instance.

For anisotropy, we observe a similar dependence of the model on the parametrization and less so on the explicit regularization. Fig. 3 and Fig. S2 in the Supporting Information show that for the V_{SV} and V_{SH} inversions at 100 km depth, both regularizations give $\xi < 1$ (i.e. $V_{SV} > V_{SH}$) at several locations, including under Madagascar, eastern Africa, and along the Central Indian Ridge (CIR). On the contrary, with the ξ and V_S parametrization, most of the study area is characterized by $\xi < 1$ (i.e. $V_{SV} < V_{SH}$). This dependence of the anisotropy model on parametrization is an indication that the model null-space and thus the model uncertainties are likely large. Our tests also]illustrate how the prior choice of parametrization can affect the results. This is a well-known caveat of regularized inversions since the choice of parametrization consists in a form of implicit regularization (Trampert 1998), and this can affect the solution in the existence of parameter trade-offs as one can expect in joint Love–Rayleigh inversions (Crampin 1976; Kirkwood 1978).

4.2 Effect of mantle transition zone topography

In theory, topography of the MTZ boundaries can affect higher mode phase velocities (eq. 12). Higher modes have thus the potential to contain information about the MTZ thickness, which could be used to constrain its thermochemical state (Meier *et al.* 2009). Here, we tested the effect of MTZ topography on our results by first inverting the dispersion data provided by Visser *et al.* (2008b) and corrected for the effect of the crust as described above, and then comparing the models with those obtained by inverting the same data set corrected with a prior model of MTZ topography. For this, we used the 410 and 670 km topography values obtained by Huang *et al.* (2019) with SS precursors and calculated their contributions to Rayleigh and Love wave phase velocities (eq. 12). Fig. S3 in the Supporting Information displays examples of MTZ topography corrections for MTZ sensitive modes and shows that they are at least an order of magnitude smaller than the measurements in the study region, and are thus unlikely to have an effect on the models. These contributions were nevertheless removed from the dispersion measurements and the remaining signal was inverted for at several grid cells.

Fig. S4 in the Supporting Information shows the 1-D V_S and ξ profiles obtained at two different locations within the study area using data corrected for the crust and data corrected both for the crust and for deviations in the depth of the 440 and 660 km boundaries. We see that the effect of the boundary topography on the model velocities is not significant as both sets of inversions produced models that were nearly identical. This is in agreement with the study of Meier *et al.* (2009) who inverted these same higher mode dispersion measurements using neural networks and found that the 410- and 660-km topographies were not well constrained with respect to the prior PPDFs. We thus proceeded with uncorrected data for the rest of the study.

5 NONLINEAR INVERSION RESULTS

5.1 Joint versus separate inversions

Nonlinear inversions with Monte Carlo methods are computationally prohibitive. This is especially true for joint Love and Rayleigh wave inversions as they double the number of unknowns compared to separate inversions. Previous studies have shown that Love and Rayleigh wave velocities are sufficiently independent of each other to justify separate inversions since Love waves are mostly sensitive to V_{SH} and Rayleigh waves are mostly sensitive to V_{SV}. However, most of those studies were based on regularized linear inversions (e.g. Ekström & Dziewonski 1998) and may not apply to our study. An exception is the work of Visser *et al.* (2008b) who used a fully nonlinear Monte Carlo approach to invert for radial anisotropy at the global scale. By separating the anisotropic PREM model into a horizontally polarized model and a vertically polarized model, the authors verified that dispersion curves calculated for Love and Rayleigh waves separately assuming isotropy did not differ significantly from those calculated for anisotropic PREM. The differences were within data errors. They did not, however, compare joint and separate nonlinear inversions of real dispersion data. For completeness, we thus decided to verify whether this assumption is valid for our method using real data.

For this purpose, we inverted Love and Rayleigh wave fundamental and higher mode dispersion data separately at a few grid cells, and compared the models with those obtained from joint Love and Rayleigh wave TB inversions (Fig. 4). An important aspect of these joint inversions is that they were performed using independent depth parametrizations for V_{SH} and for V_{SV}. As demonstrated by Gao & Lekić (2018), inversions performed in a TB framework that use identical choices of depth parametrizations for the different unknowns can bias structure estimates. This 'attached-type' parametrization is very common in seismology and has the advantage of potentially reducing the number of parameters, but it also constitutes a form of implicit regularization that imposes the same geometry to all parameters. It often results in a model geometry that is determined by the best-resolved parameters but that does not necessarily appropriate for the others. We verified that this is true for our data set as well. Fig. 4 compares results of joint inversions with independent parametrizations for V_{SH} and for V_{SV} and results of separate Love and Rayleigh wave inversions using our nonlinear TB inversion approach. It displays the mean models and their 90 per cent confidence intervals on the same plots for the two types of inversions. We can see that there are few differences between the mean V_{SV} models and between the mean reconstructed V_S models. Some differences are, however, visible between the mean V_{SH} models and, consequently, between the mean ξ models, especially in the top 300 km. This may be in part related to the shallower sensitivity of Love waves compared to Rayleigh waves at the same period, including sensitivity to the crust and shallowmost mantle for shorter period waves. This could imply that solving for crustal and shallow mantle anisotropy benefits from jointly inverting Rayleigh and Love wave data, though we were not able to confirm this with synthetic tests (Fig. S5, Supporting Information). Nevertheless, we can also conclude that separate inversions yield equivalent results to joint inversions as long as model uncertainties are accounted for. We thus proceeded with separate inversions for the rest of our study.

5.2 Effect of noise parametrization

As explained above, one advantage of the TB method is that it can invert for unknown data noise jointly with other parameters to prevent the mapping of data uncertainties into the model. While the phase velocity maps we used here (Visser *et al.* 2008b) were published with quantitative uncertainties at each period, these uncertainties resulted from averaging the errors on each measured earthquake-station pair. They therefore do not reflect the lateral variations in the uncertainties that can be caused by differences in ray coverage,

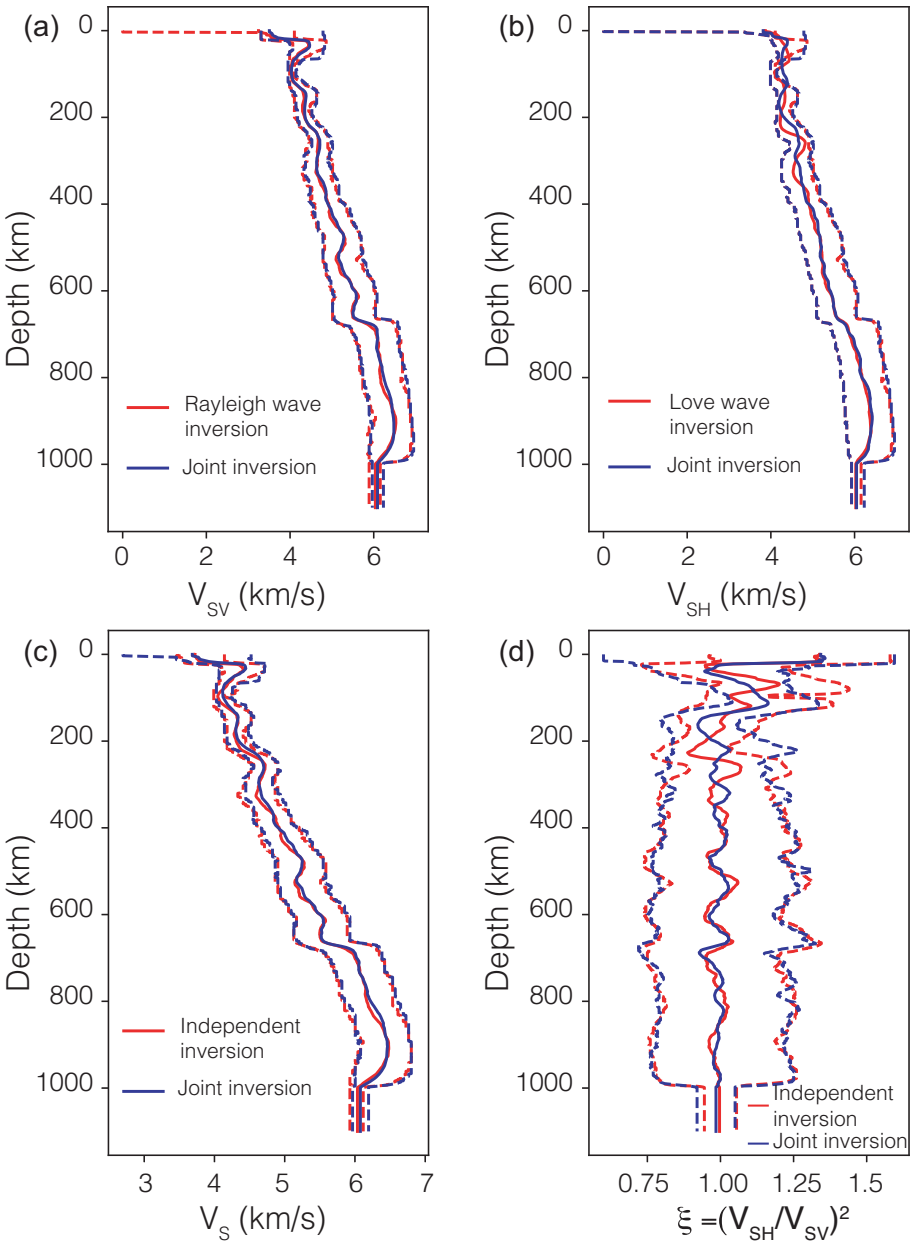

Figure 4. Comparison of V_{SV} (a) and V_{SH} (b) models obtained by independent and joint Love and Rayleigh wave inversions using our TB method at (38°S, 79°E). The solid lines represent the mean model and the dashed lines represent their 90 per cent confidence interval. (c) and (d) represent the resulting V_S and ξ models, respectively, and their 90 per cent confidence interval.

for instance. It is thus reasonable for us to invert for data noise in addition to V_S and ξ instead of taking the values estimated by Visser *et al.* (2008b) at face value. However, assumptions have to be made about the prior noise distribution. On one hand, assuming that all the measurements have the same Gaussian noise adds only one unknown to the inverse problem and does not significantly increase the computational cost. This assumption may not, however, be valid, in which case any unmapped data noise can get mapped into the posterior distribution. On the other hand, assuming different noise levels for the different types of data can increase the computational cost of an already time-consuming technique and could increase trade-offs between resolving data noise parameters and model parameters.

In order to determine the best approach for our problem, we compared TB inversion results for different noise parametrizations. In the first case we used one noise parameter for the Love wave inversions and one for the Rayleigh wave inversions. In the second case, each of the data subsets (Love and Rayleigh) were divided into two groups: one group contains mode branches $n = 0$–2 and the other group contains the mode branches $n = 3$ and higher. Each of these groups had its own noise level. In the third case, each mode branch within the two subsets had its own noise level. Results are shown for V_{SV} and V_{SH} in Fig. 5. Figs S6–S8 in the Supporting Information display results for the reconstructed V_S and ξ and the noise parameter PPDFs. The models are not significantly different, though the cases with additional noise parameters yielded less well-resolved V_S and ξ. This is expected when increasing the number of unknowns in the inversion if trade-offs are present. However, the same general trends still exist across these different model distributions. We also

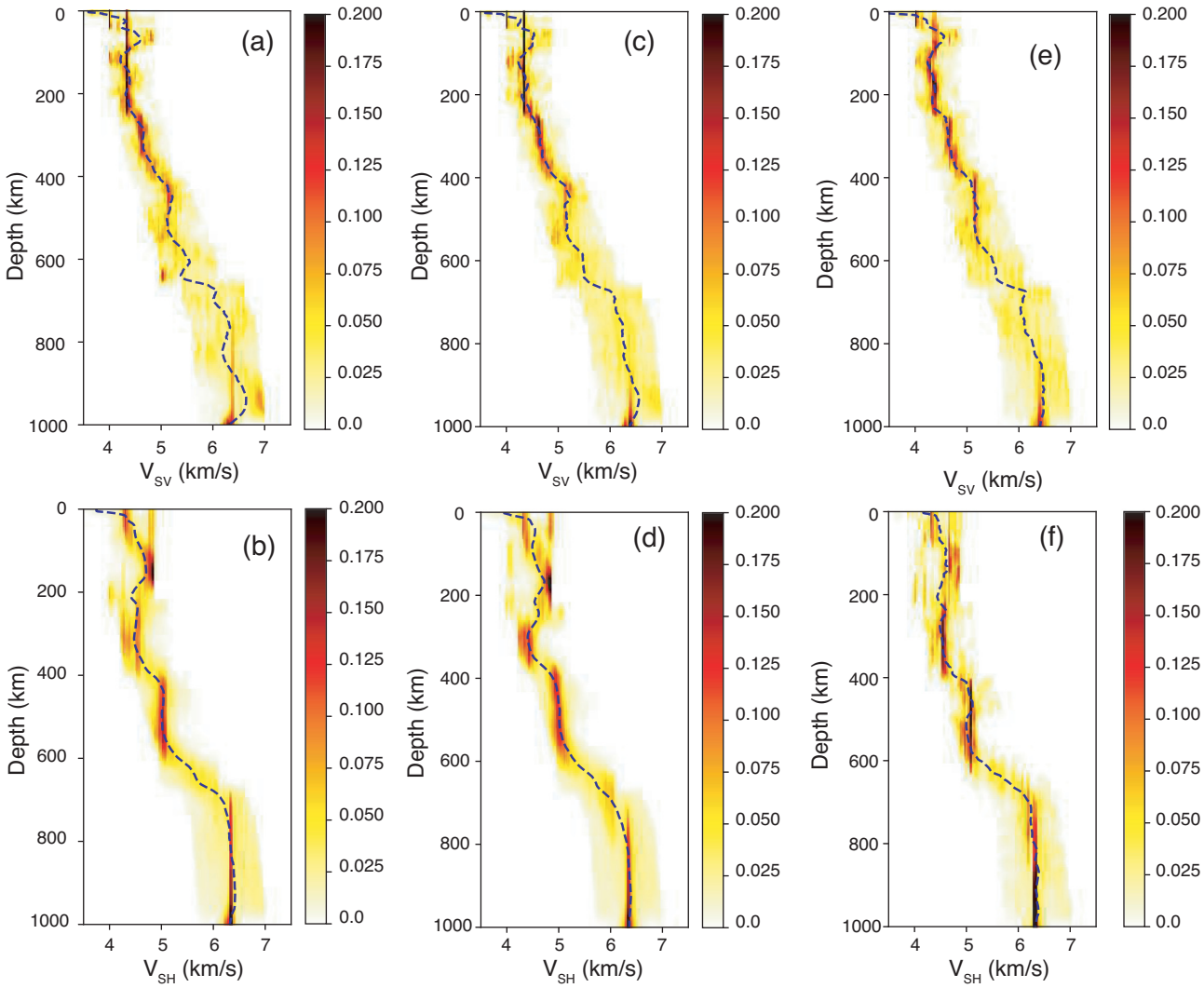

Figure 5. Comparison of inversion of Rayleigh (top) and Love (bottom) phase velocity data using a single noise parameter for all the modes ((a) and (b)), using one noise parameter for $n = 0$–2 and another one for $n \geq 3$ ((c) and (d)), and using one noise level per mode branch ((e) and (f)). The dashed lines represent the mean models. The models shown here are for a grid cell at (3°S, 79°E).

note that the posterior noise parameters are generally larger for Love waves than for Rayleigh waves, indicating that Love wave dispersion data are noisier. Additionally, the noise parameter distributions we obtained generally indicate smaller values than or consistent with the average values of Visser *et al.* (2008b). The exception is for the sixth Rayleigh wave overtone for which we found larger noise levels than estimated by Visser *et al.* (2008b). Considering how little effect the number of noise parameters has on the velocity and anisotropy models, we proceeded using only one for each data set inversion.

5.3 Models

Results from both the linear and the TB inversions are shown in Figs 2((i)–(l)) and 3((i)–(l)) for anomalies relative to PREM of V_S and ξ, respectively. The TB results are the mean of the posterior model distributions at each grid cell. The data were not corrected for topography at the 410 and 660 km boundaries since we showed that it does not have a significant effect on the models (Figs S3 and S4, Supporting Information).

We see that at 100 km both methods result in negative V_S perturbations under the CIR, the Southwestern Indian Ridge (SWIR), and Southeastern Indian Ridge (SEIR) branches. Under Madagascar, while the sign of the anomalies at that depth in the SVD results depends on parametrization, the TB results show possible weak positive and negative anomalies. Between the CIR and the Indonesian subduction zone the TB results display generally negative velocity anomalies, in closer agreement with the SVD inversion that employed the V_S and ξ parametrization than with the V_{SV} and V_{SH} parametrization. At larger depths, we see a strong reduction in the strength of the velocity anomalies, as often found in tomographic mantle models, for the TB and the SVD models. At 200 km, the TB results present evidence for $d\ln V_S < 0$ along the CIR and SEIR of stronger amplitude (3–5 per cent) than in the SVD models. On the contrary, at that depth, the TB model shows $d\ln V_S > 0$ under the SWIR, whereas the SVD models still display some amount of negative anomalies. At 400 km, the two methods are in better agreement with a ubiquitous $d\ln V_S > 0$ across the region but with stronger anomalies in the TB model. In the upper MTZ, at 500 km, both techniques show weak positive $d\ln V_S$ around Indonesia, east of the SEIR, and west of the CIR, as well as negative $d\ln V_S$ near India.

For radial anisotropy, the methods have generally less agreement than for $d\ln V_S$ (Fig. 3). At 100 km depth, both methods show a general horizontally fast direction ($\xi > 1$). However, the total anisotropy is greater for the TB inversion ($\xi = 1.15$–1.20) than for the SVD inversions ($\xi \sim 1.10$) and the $\xi < 1$ anomaly seen in the southern part of the study area with the SVD method is not found with the TB method. In addition, a strong vertically fast region south of the Indonesian subduction zone is visible in the TB results but not in the SVD models. At 200 km, the linearized results show a horizontally fast zone throughout the region. On the contrary, the TB results show a heterogeneous mixture of horizontally fast and vertically fast regions. At 400 km, the results for both methods are heterogeneous and the anisotropy is clearly stronger in the TB model, but the TB results point toward a more widespread $\xi < 1$ over the region. This suggests that the TB inversion tends to favour vertically fast shear waves at that depth, which is in slightly better agreement with the SVD inversions using the V_S and ξ parametrization than those using the V_{SV} and V_{SH} parametrization. In the MTZ, while details in the pattern of the anisotropy differ between the two techniques, both inversion methods favour mostly vertically fast shear waves. Both models also suggest horizontally fast waves under the northern part of the CIR.

The TB method does not only offer a way to invert data with less assumptions than traditional methods, it also provides a way to quantify model parameter uncertainties. Representing uncertainties is nevertheless not simple as the Gaussian assumption one typically makes when plotting uncertainty maps may not be valid. While one can always try to fit a Gaussian through the posterior model distribution, it may give a false sense of uncertainty or certainty as it does not convey any skew, multiple peaks, or other non-Gaussian features. We thus decided instead to follow previous studies (Beghein & Trampert 2006; Visser *et al.* 2008a) and use the posteriors to calculate the likelihood of the presence of positive or negative anomalies in the model parameters. These results are shown in Figs 6 and 7 for $d\ln V_S$ relative to PREM and in Figs 8 and 9 for radial anisotropy.

In Figs 6 and 7, we see that at 100 km depth, the probability P that V_S anomalies are negative underneath oceanic ridges is greater or equal to 0.75, and that the anomalies are of at least 1 per cent in strength. In addition, while the negative anomalies beneath the SWIR are likely to be between 1 per cent and 5 per cent, the CIR and SEIR are likely ($P \geq 0.75$) characterized by stronger anomalies of at least -5 per cent at that depth. These relatively low velocity zones largely disappear at 200 km depth, with the exception of a likely $d\ln V_S < -1$ per cent underneath part of the SEIR. At this depth, positive relative velocity perturbations of at least 1 per cent likely exist under East Africa and the SWIR, but they are unlikely to be >5 per cent. At 400 km, an almost ubiquitious $d\ln V_s > 1$ per cent is found with a high likelihood. This anomaly is, however, probably lower than 5 per cent for much of the region, with the exceptions of the region between Southern Africa and Madagascar and east of the CIR. At 500 km, relative negative or positive velocity perturbations of 5 per cent or more are very unlikely to exist. At those depths, positive anomalies of at least 1 per cent may exist near La Réunion Island and the SEIR, and negative anomalies of at least 1 per cent may exist on the SWIR and east of the CIR.

Figs 8 and 9 reveal that at 100 km depth, there is a high likelihood ($P \geq 0.75$) for $V_{SH} > V_{SV}$ with strength of at least 5 per cent underneath much of the Indian ocean. The exceptions are the Rodrigues Triple Junction (RTJ), where the CIR, SEIR, and SWIR intersect, and the region around the Indonesian subduction zone where the likelihood of $\xi > 1$ is low (Fig. 8). At the RTJ, the likelihood of $\xi < 1$ is also weak (Fig. 9), suggesting the area may not have any radial anisotropy. A small pocket of likely $V_{SH} < V_{SV}$ with strength of at least 5 per cent is also visible south of the Indonesian slab around 30°S, 105°E. At 200 km, the likelihood of anisotropy anomalies is more heterogeneous and we see both horizontally fast and vertically fast anomalies of about 1 per cent, indicating a likely reduction in the amplitude of the anisotropy with depth. There is also a pocket east of the RTJ with likely horizontally fast velocities with at least 5 per cent anisotropy. At 400 km depth, the likelihood of $V_{SH} > V_{SV}$ decreases further and a few locations are characterized by a probability of 0.7 or greater of having $V_{SH} < V_{SV}$ with about 1 per cent anisotropy ($\xi \simeq 0.99$). At that depth, a zone of likely vertically fast velocities with at least 5 per cent anisotropy is found south of the RTJ and in between the the CIR and Indonesian subduction zone. At 500 km, we see that much of the region has likely $V_{SH} < V_{SV}$ with $0.95 < \xi < 0.99$.

5.4 TB inversions synthetic tests

In order to test the robustness of the above results, we ran synthetic tests using some of the mean models displayed in Figs 2 and 3 as input to calculate the synthetic data. The prior and the noise parametrization were defined in the same way as in the real data inversions. Fig. 10 shows the result of one such test. For V_{SV} and the reconstructed V_S, we find that the posterior model distribution encompasses the input model and that the mean of this distribution is representative of the input model, especially above 400 km depth. At greater depths, the uncertainties increase. For V_{SH}, the peaks of the posterior distribution are offset compared to the input model, showing that the TB inversion tends to overestimate the model amplitudes. However, the mean model follows the input model at most depths except between 400 and 550 km where rapid vertical changes in the input model are not accurately captured by the mean of the posterior distribution. This is likely due to limitations in the vertical resolution of the data. Additionally, the 50–150 km depth range is characterized by multiple peaks in the posterior distribution that indicate trade-offs: the data cannot distinguish between a high velocity zone above a low velocity layer and a constant velocity structure at these depths. Since ξ is the square of the ratio of V_{SH} to V_{SV}, these differences between input and mean model are amplified in the reconstructed ξ profile. We see that the general trend of the posterior distribution roughly follows that of the input ξ model, but the amplitudes of the input model and the mean or the peak of the posterior distribution differ. Nevertheless, it should be noted that the posterior ξ distribution encompasses the input model, which highlights the importance of accounting for model uncertainties when interpreting results.

A question that may still arise is whether we might be mapping anisotropy into an isotropic structure. We thus ran another synthetic test with an isotropic input model below 200 km (Fig. 11). As in Fig. 10, we find a close match between the input model and the mean of the posterior for V_{SV} and V_S down to about 400 km depth. At greater depths, the posterior becomes wider, and the mean model oscillates around the input model. For V_{SH}, the same multimodal peaks found in Fig. 10 are visible between 50 and 150 km, and deviations are found between the input and mean posterior model around 200 km, which marks the abrupt transition between anisotropy and isotropy in the input model. At larger depths, the mean V_{SH} model follows the input model relatively well, though with larger uncertainties. For radial anisotropy, the mean model shows a strong $\xi < 1$ near 200 km depth that peaks about 50 km deeper than the corresponding anomaly of the input model. At larger depths, the

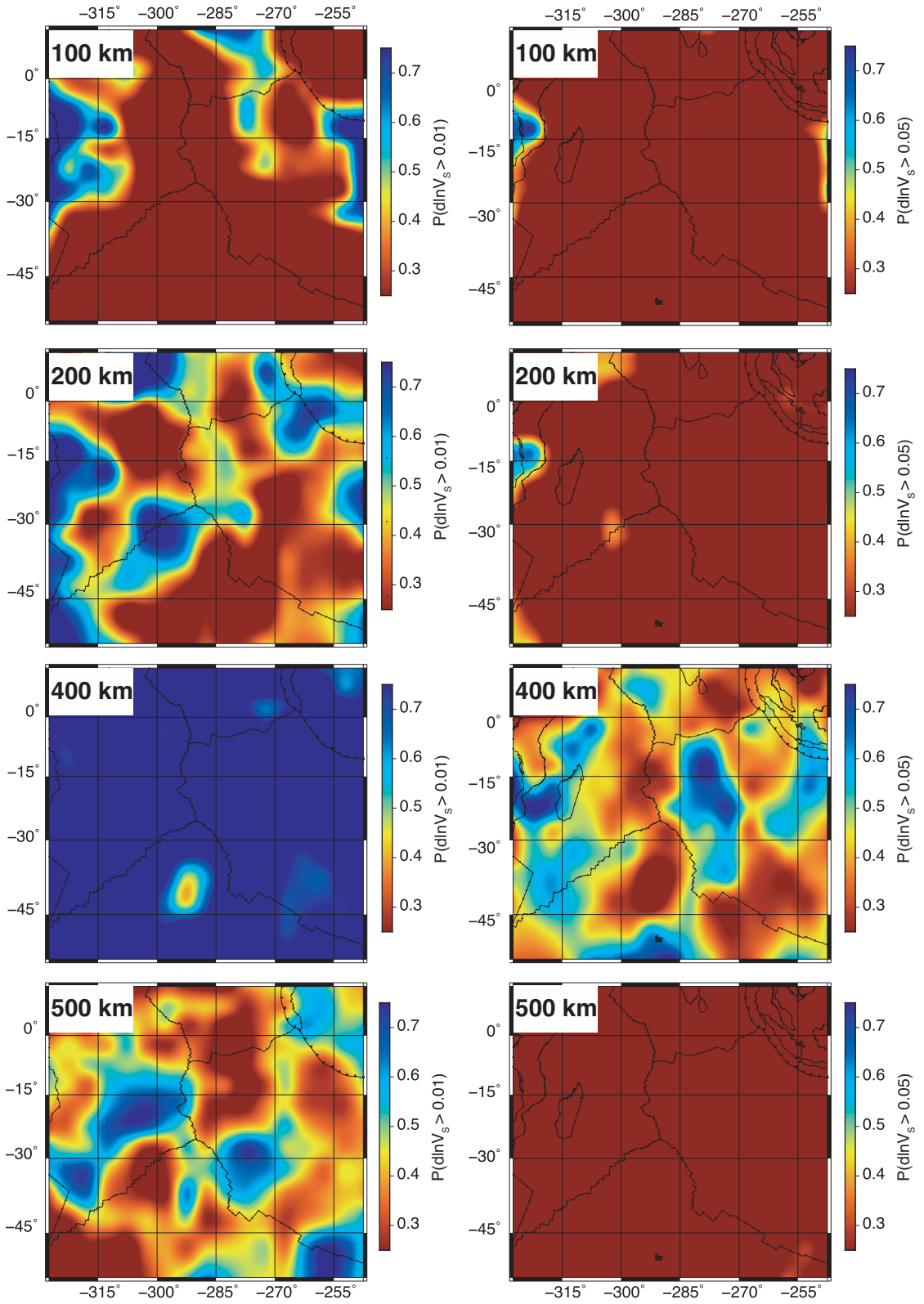

Figure 6. Probability of $d\ln V_S > 1$ per cent (left) and $d\ln V_S > 5$ per cent (right) at different depths.

mean model oscillates around the isotropic input model, with peak anisotropy approaching 5 per cent at some depths. If one were to interpret these results based on the mean anisotropy model alone, one would conclude that there is a risk of finding deep anisotropy with our method that is not constrained by the data. However, we note that the posterior ξ distribution below 200 km is wide and allows a wide range of values of both $\xi > 1$ and < 1 at these depths. This deep anisotropy found in the mean model is therefore not resolved, but

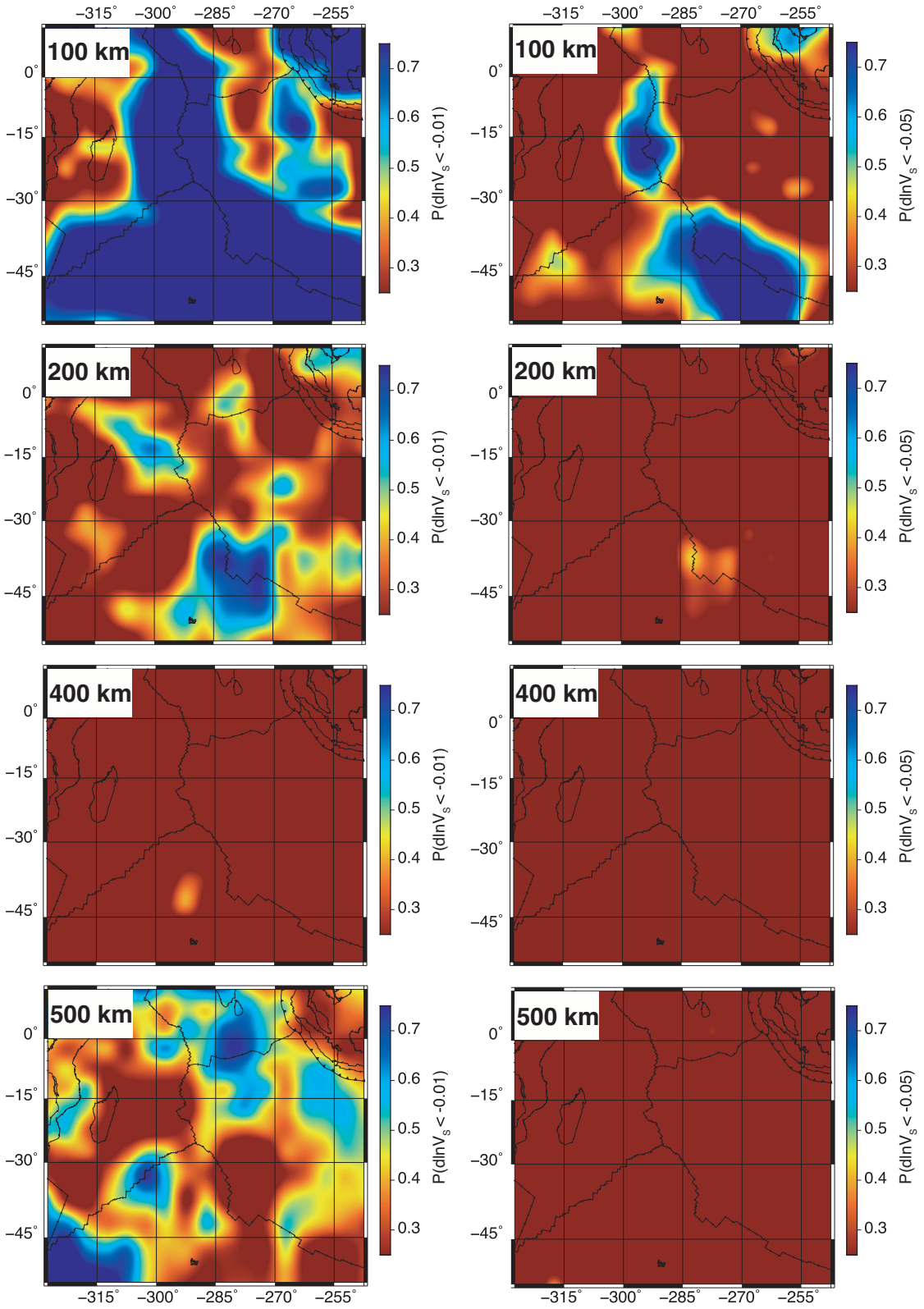

Figure 7. Probability of $d\ln V_S < -1$ per cent (left) and $d\ln V_S < -5$ per cent (right) at different depths.

this again highlights how essential it is to determine the posterior model uncertainties before interpreting the anisotropy signal.

In order to determine the level of confidence with which we can interpret the likelihood map shown in Figs 8 and 9, we used the posterior ξ distribution displayed in Fig. 11 to calculate the likelihood of finding 1 per cent and 5 per cent anisotropy for both $V_{SV} > V_{SH}$ and $V_{SH} > V_{SV}$. The results are shown in Fig. 12. We see a strong likelihood ($P = 0.8$) of finding $\xi < 0.95$ between 200

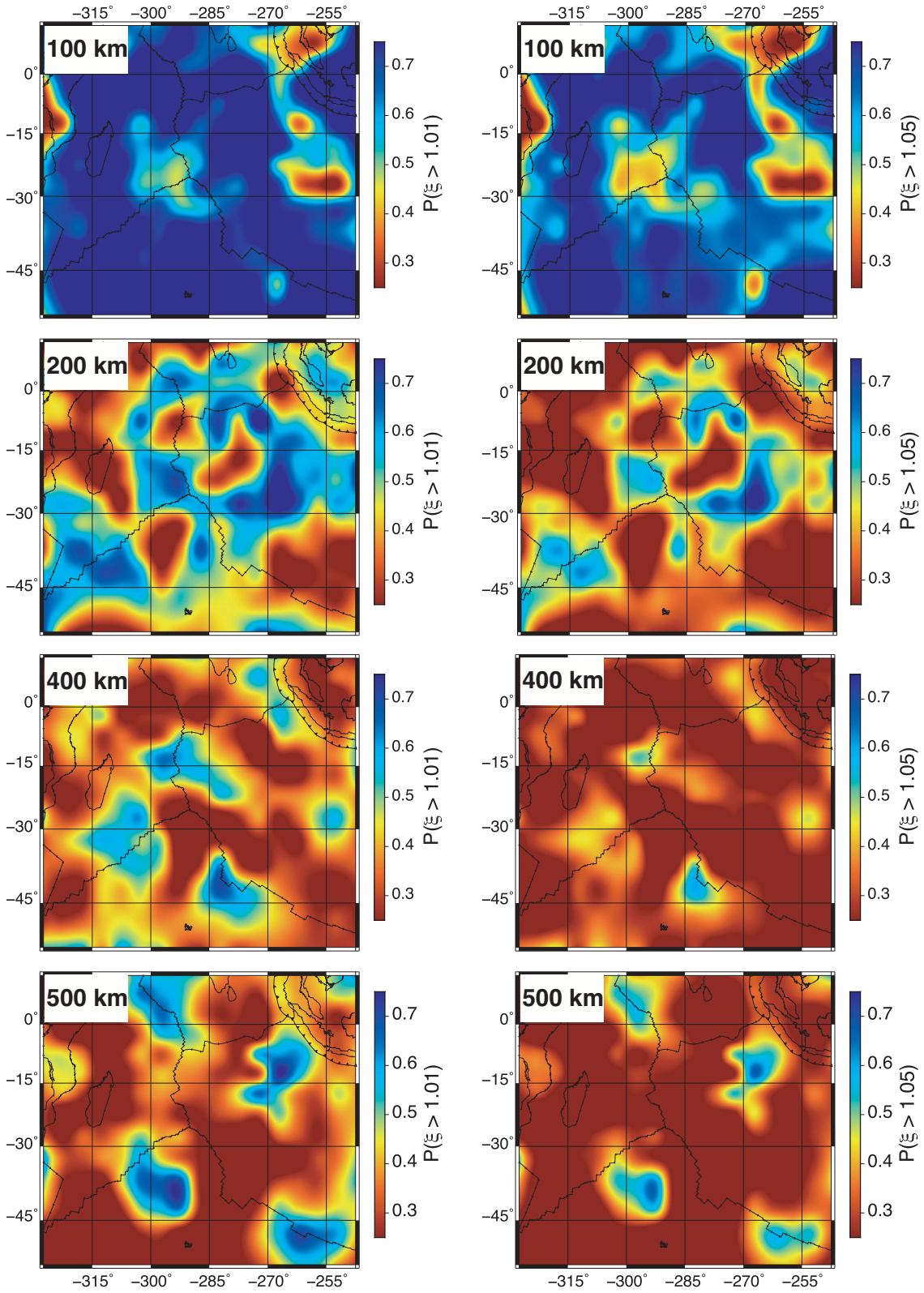

Figure 8. Probability of having $V_{SH} > V_{SV}$ (i.e. $\xi > 1$) at different depths with at least 1 per cent (left) and 5 per cent (right) anisotropy.

and ∼230 km, where the input model is isotropic. We attribute this to the inability of the surface waves to recover a sharp velocity or anisotropy vertical change. Most noteworthy, however, is the fact that below ∼250 km depth, the probability that we might find an anisotropy signal not constrained by the data is generally less than 0.6. This means that, when discussing our results (Figs 8 and 9), we can confidently conclude that the deep anisotropy signal in our models is constrained by the data as long as we only interpret anisotropy associated with a likelihood higher than 0.6.

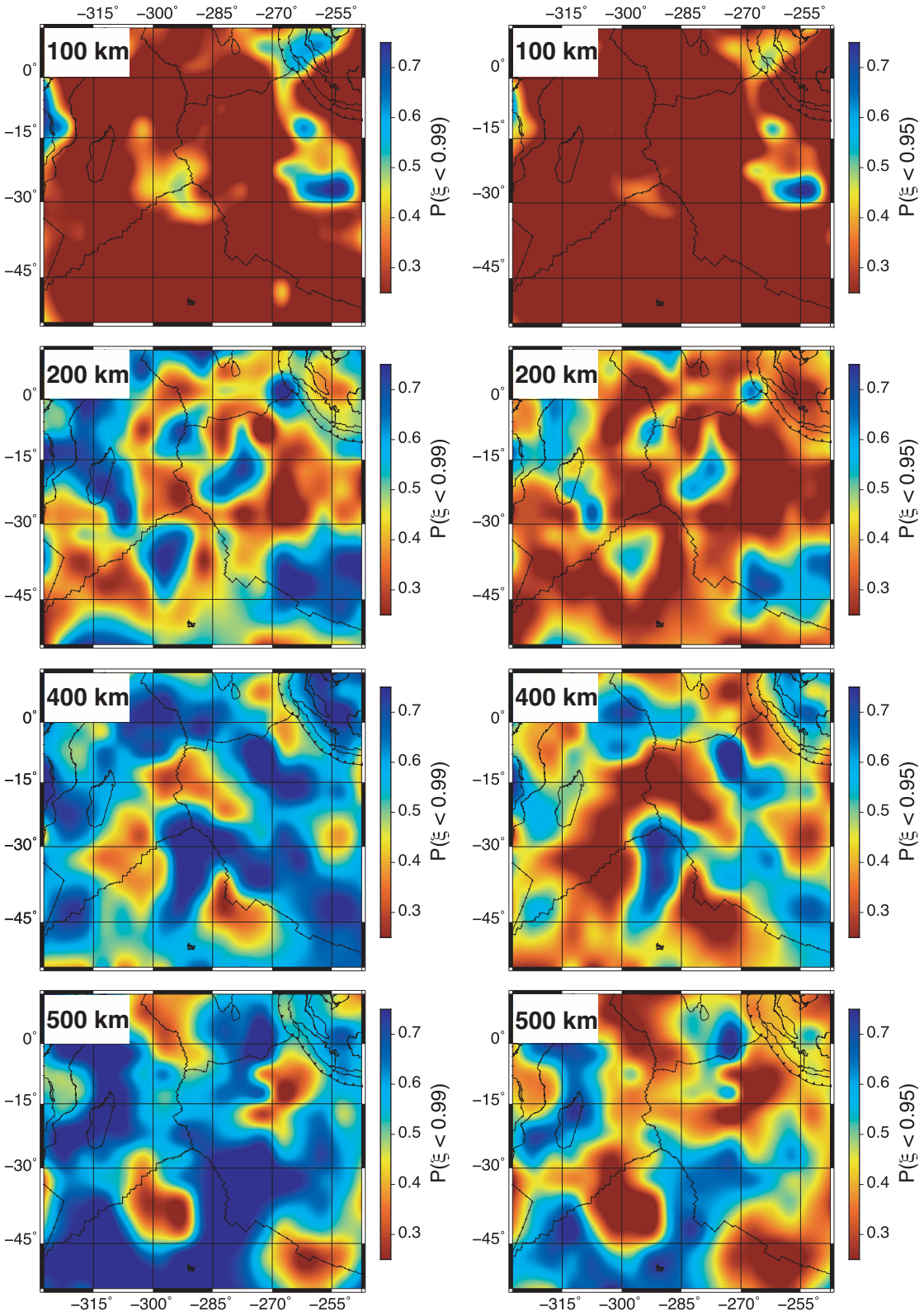

Figure 9. Probability of having $V_{SH} < V_{SV}$ (i.e. $\xi < 1$) at different depths with at least 1 per cent (left) and 5 per cent (right) anisotropy.

5.5 Joint group and phase velocity inversion

Some of the results presented above show multiple peaks in the posterior distributions, indicating that trade-offs exist between the different depth parameters (Fig. 4). In addition, it is clear that Love wave phase velocity data alone cannot resolve V_{SH} as well as Rayleigh wave phase velocities can resolve V_{SV}, resulting in large uncertainties in ξ. Group velocity dispersion data are generally sensitive to shallower depths than phase velocities at the

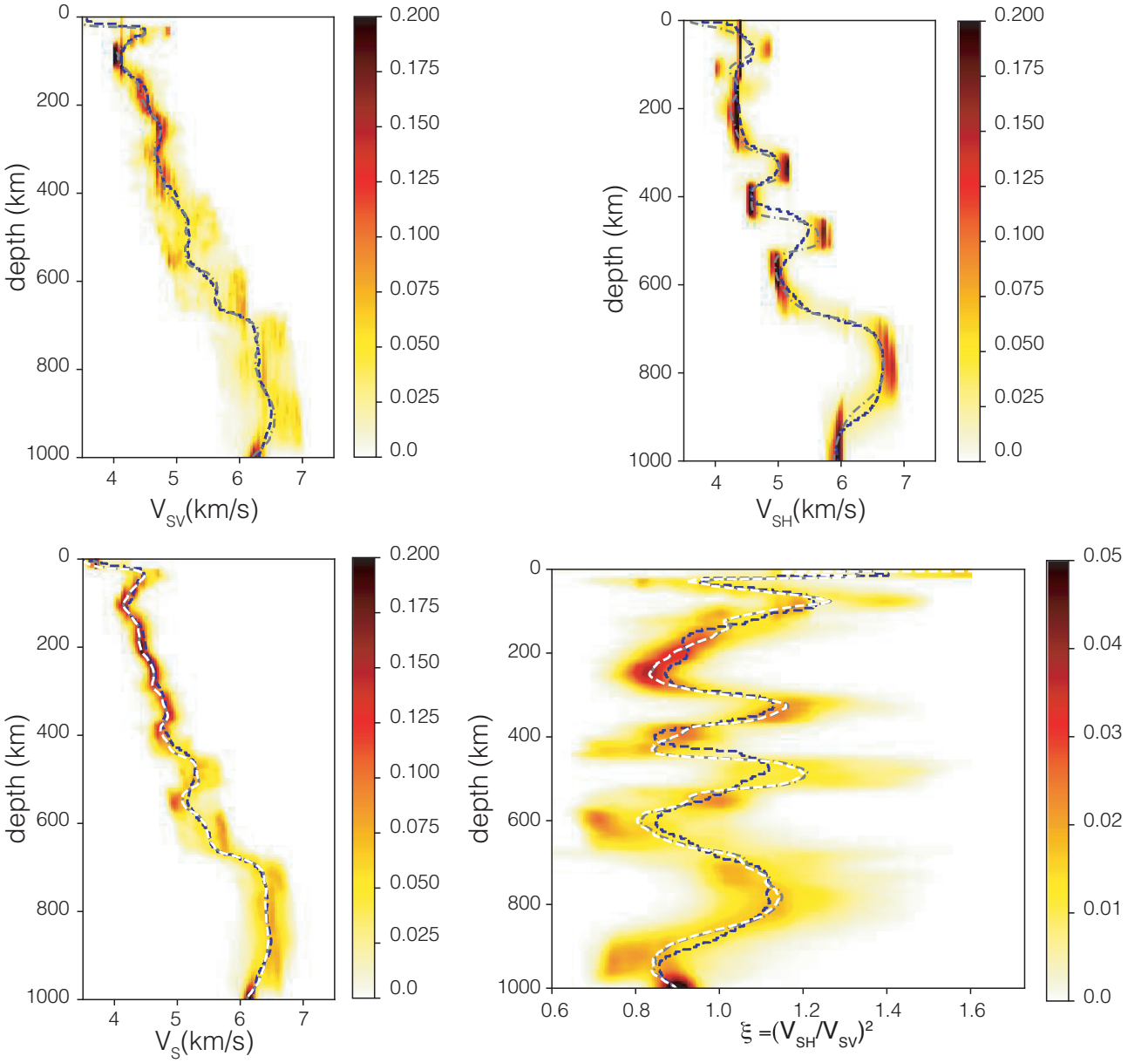

Figure 10. Example of a synthetic test. Top: posterior V_{SV} and V_{SH} distributions resulting from the separate Rayleigh and Love wave synthetic dispersion data; and bottom: corresponding V_S and ξ posterior distributions. The colour scale represents the probability that a model parameter has a specific value at a given depth. The input model, which was the mean model from results at 38°S, 69°E, is shown by the dashed blue line. The dashed–dotted grey line is the mean of the posterior distribution. The dashed white line in the V_S and ξ profiles calculated using the mean V_{SV} and mean V_{SH} results.

same periods (Fig. 13) and may therefore help break trade-offs between shallow and deep structure. In this section, we tested whether adding the group velocity data measured by Ma *et al.* (2014) could reduce trade-offs and help better constrain the deep anisotropy signal.

Since the group and phase velocity data came from two different studies, we included independent noise parameters for each data type in our inversions. Fig. S9 in the Supporting Information represents the posterior noise distributions. It shows that the Love wave data are generally noisier than the Rayleigh wave data and that the group velocity data also contain a larger amount of noise than the phase velocities. Fig. 14 compares the posterior model distributions for an inversion with both phase and group velocity measurements and an inversion with phase velocity measurements only. Both inversions use data at the same grid cell: (3°S, 79°E). We find that the multimodal aspect of the V_{SH}, and to some extent V_{SV}, PPDFs at shallow depths is less pronounced when group velocities are included in the data set. This leads to narrower distributions for V_S and ξ in the top 50–100 km, showing that the addition of group velocities improves constraints on the shallow structure. The amplitude of the mean anisotropy model is also found to be slightly lower in the top 50 km when group velocities are included but the general trend does not change significantly. Importantly, we find that the posterior distributions do not change significantly below ∼100 km. This suggests that, although the group velocity data can help constrain the shallow structure better, it does not change the model resolution in the rest

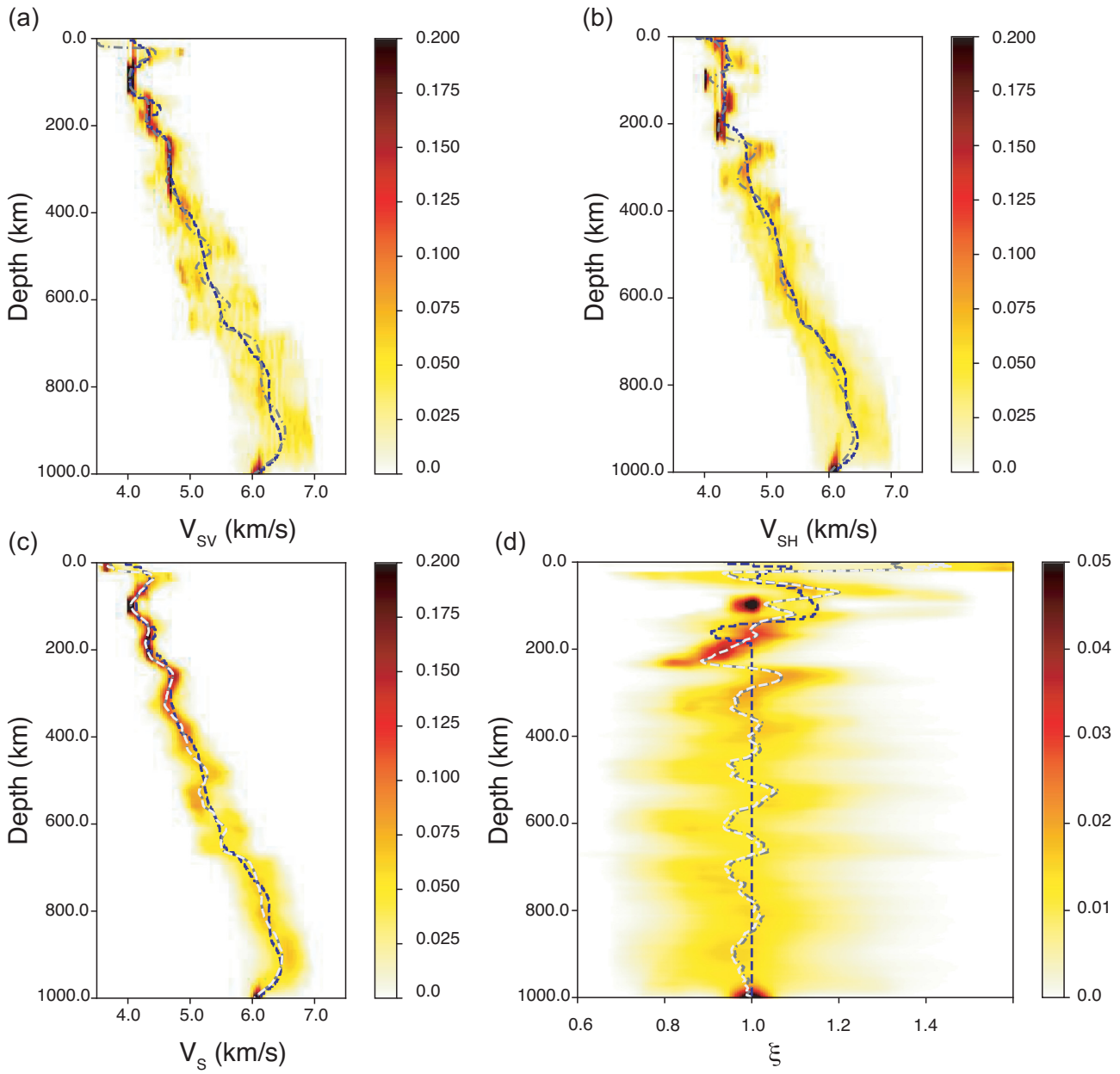

Figure 11. Synthetic test using an input model that is isotropic below 220 km depth. The colour scale indicates the probability that the model parameter has a specific value at a given depth. The input model is shown as a dashed blue line. The dashed–dotted grey line is the mean model of the posterior distribution. The dashed white line in the V_S and ξ plots was obtained from the mean V_{SV} and mean V_{SH} results.

of the upper mantle and does not affect our interpretation of the results.

6 DISCUSSION

The surface wave tomography of Debayle & Lévêque (1997) revealed that the lowest shear wave velocity anomalies at 100 km depth were located along the ridges, with the CIR and SEIR showing stronger anomalies than the SWIR. Low velocities along the CIR are also seen in the fundamental Rayleigh wave phase and group velocity study of Mazzullo *et al.* (2017). This is in agreement with our mean model (Fig. 2) and is confirmed by our likelihood estimates (Fig. 7) and model distributions (Fig. 15). Figs 7 and 15 indeed show that both the CIR and SEIR are likely to have $d\ln V_S \leq -0.05$ while the SWIR is characterized by a high likelihood of $d\ln V_S \leq -0.01$ but not $d\ln V_S \leq -0.05$. Interestingly, our results suggest that the low velocity anomalies seen under the CIR and SEIR persist down to about 200 km as in Debayle & Lévêque (1997), but that the SWIR has a shallower origin. Debayle & Lévêque (1997) found that these anomalies largely disappear at around 200 km. This is qualitatively consistent with our results, which show a decrease in the amplitude of the anomalies. At 400 km, our results favour a positive velocity perturbation of at least 1 per cent and likely less than 5 per cent across the whole study area, which suggests a relatively cold mantle. This differs from several other studies (Kustowski *et al.* 2008; Ritsema *et al.* 2011; French & Romanowicz 2014), though exceptions include the work of Panning & Romanowicz (2006) Panning *et al.* (2010) and Auer *et al.* (2014), which displayed $d\ln V_S > 0$

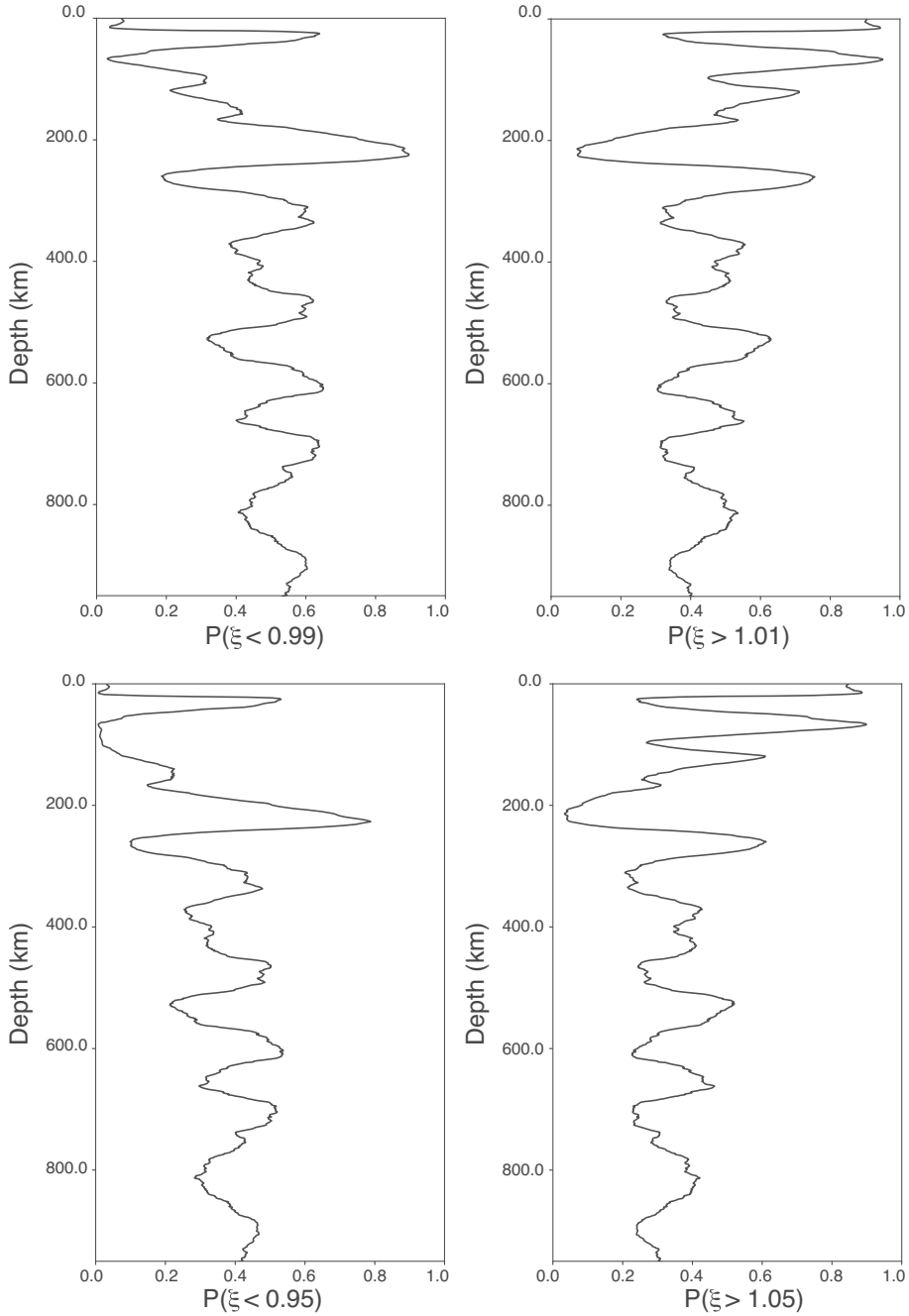

Figure 12. Depth profile of likelihood of ξ anomalies relative to isotropy ($\xi = 1$) based on posterior distributions shown in Fig. 11.

along ridges at similar depths. In the MTZ, our models present weak positive and negative velocity anomalies of about 1 per cent, qualitatively similar to other studies. However, we do not find a broad high-velocity zone in the MTZ south of the SEIR that could be interpreted as a subducting slab as found in Simmons *et al.* (2015). We do not find evidence for a broad $d\ln V_S > 0$ in the MTZ south of the SEIR as found in Simmons *et al.* (2015), which they interpreted as a subducting slab.

As in Visser *et al.* (2008a), we find a high likelihood ($P \geq 0.75$) of $V_{SH} > V_{SV}$ at 100 km depth throughout most of the study region. This is consistent with previously published radial anisotropy models of our study area (Lévèque *et al.* 1998) as well as in other regions and globally (Ekström & Dziewonski 1998; Gung *et al.* 2003; Panning & Romanowicz 2006; Kustowski *et al.* 2008; Panning *et al.* 2010; Auer *et al.* 2014; French & Romanowicz 2014; Chang *et al.* 2015) and is often interpreted as the signature of horizontal asthenospheric motion. At 200 km, our models present likely ($P \geq 0.75$) $V_{SV} > V_{SH}$ and $V_{SH} > V_{SV}$ anomalies with 1–5 per cent anisotropy scattered across the region, which differs somewhat from Visser *et al.* (2008a) who found a generally low likelihood of $V_{SH} > V_{SV}$ at that depth. No clear anisotropy pattern or relation to geological features emerges from our results below 100 km depth except for a vertically fast signal at 200 km along the Indonesian subduction zone (Fig. 3), though it is likely to be of small amplitude (Fig. 9). Contrary to the global models of Gung *et al.* (2003), Panning & Romanowicz (2006) and Zhou *et al.* (2006), we do not see any clear anisotropy signal $V_{SV} > V_{SH}$

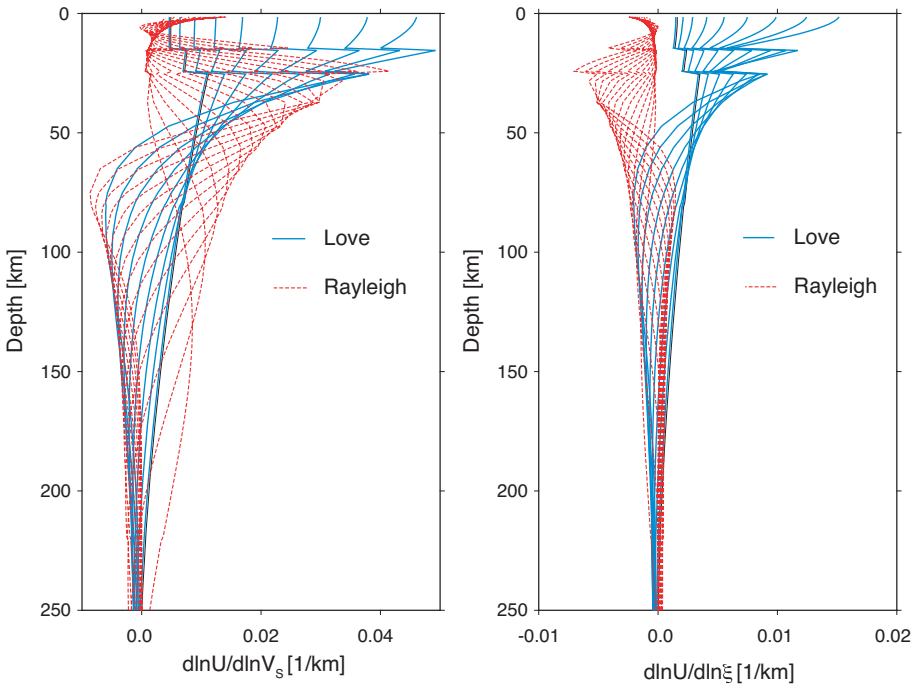

Figure 13. Fundamental mode group velocity sensitivity kernels for perturbations in V_S (left) and ξ (right) between 10 s and 30 s for Love waves and between 7.5 and 40 s for Rayleigh waves.

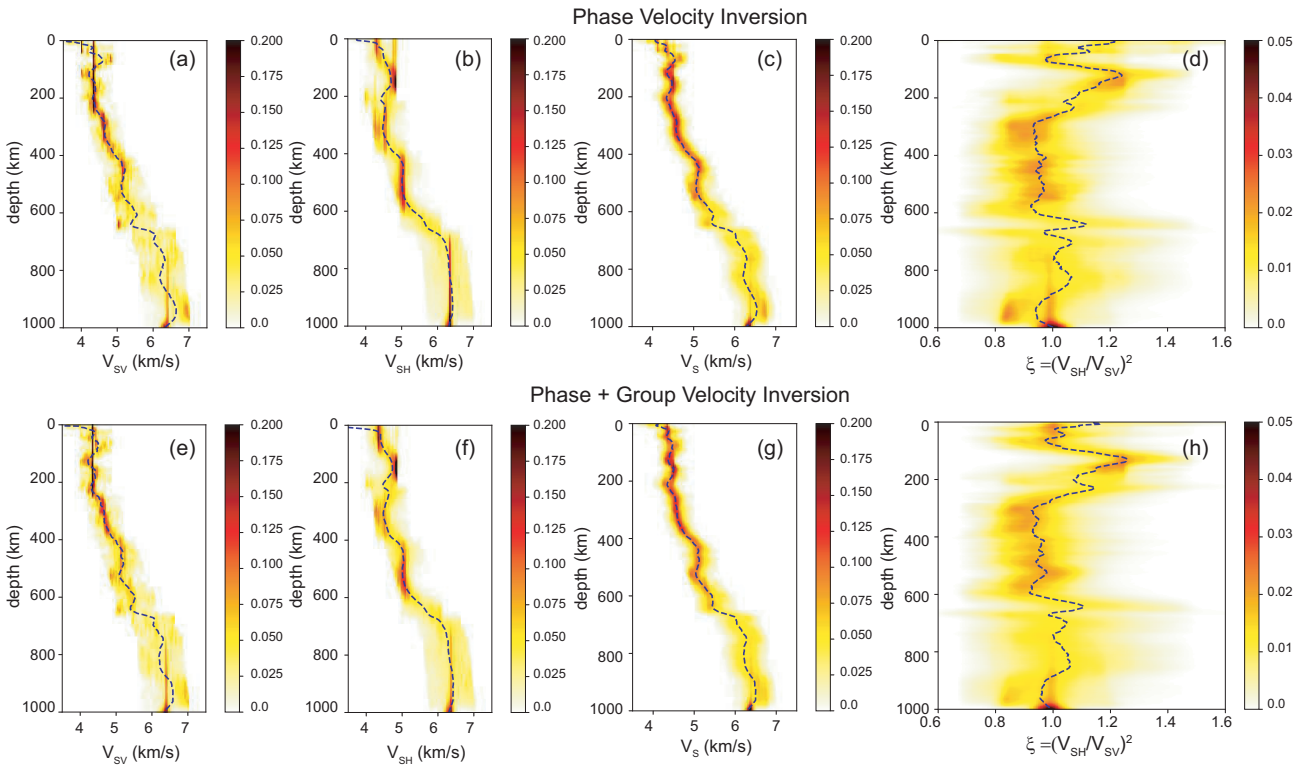

Figure 14. Comparison of inversion of phase velocity data only (top) and joint inversion of phase and group velocity data (bottom) at 3°S, 79°E. Inversions were performed separately for V_{SV} ((a) and (e)) and for V_{SH} ((b) and (f)). Panels (c) and (g) represent the reconstructed posterior V_S model distribution and (d) and (h) represent the reconstructed ξ model distribution. The dashed lines correspond to the mean of the distributions.

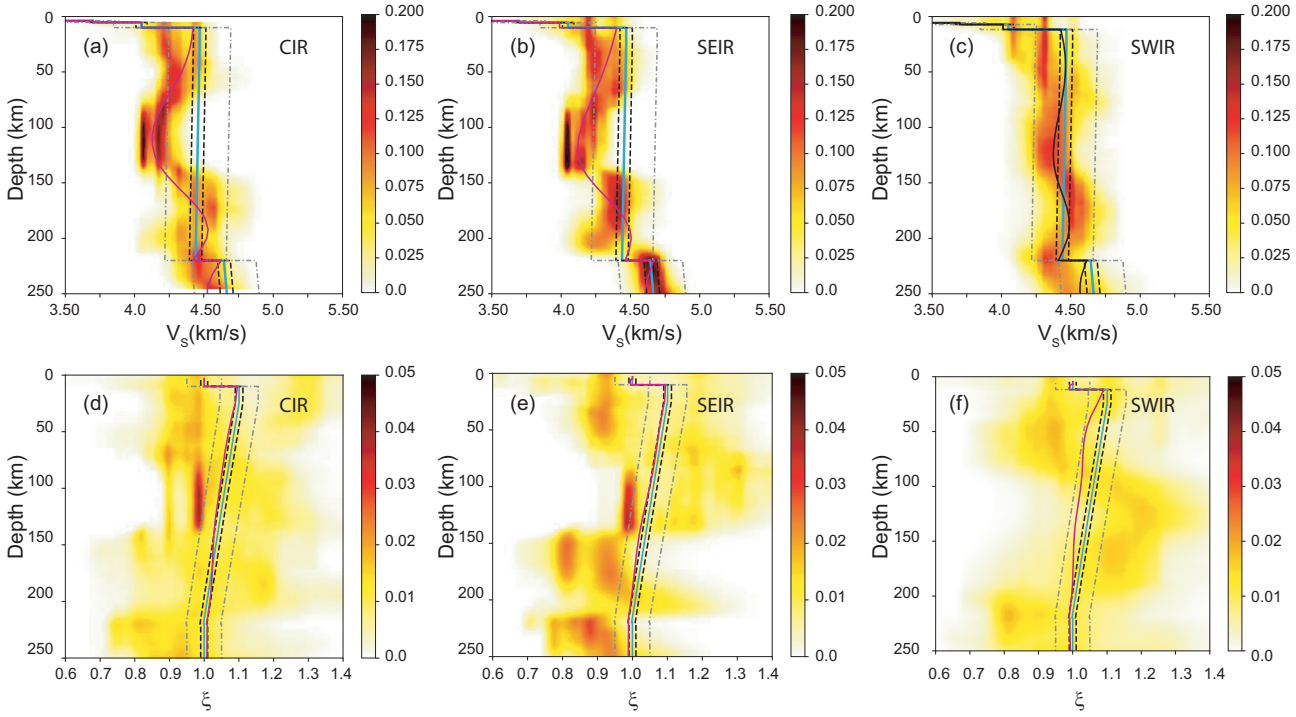

Figure 15. Distributions of V_S (top) and ξ (bottom) models obtained at the CIR, SEIR and SWIR. The colour scale represents the likelihood, the magenta solid line represents the mean model and the dashed lines indicate 1 per cent and 5 per cent perturbations from PREM.

underneath ocean ridges below 100 km. Overall, the tendency of the higher mode data to favour $V_{SV} > V_{SH}$ in the deep mantle, including the MTZ, found by Visser *et al.* (2008a) is similar to what we found, though their results seem to suggest a more ubiquitous $\xi < 1$ signal than ours. Other models have previously shown small, laterally variable radial anisotropy in the MTZ (Auer *et al.* 2014; Moulik & Ekström 2014) though they do not necessarily share the same patterns.

It is generally accepted that the [100](010) slip system is dominant in the upper mantle and produces A-type olivine LPO for which the fast seismic direction aligns with the direction of mantle flow (Karato 2012). It should, however, be noted that the relationship between fast seismic direction and mantle deformation is not straightforward at greater depths and that we should not directly conclude from our results that $V_{SV} > V_{SH}$ implies vertical flow. Unlike for the shallow mantle, the anisotropy fast axis may not be a good proxy for flow direction as deformation could occur with a different slip system than at shallower depths. A transition from dominant [100] to [001] slip at high pressure can occur for olivine between 250 and 400 km depth, which could lead to $V_{SV} > V_{SH}$ in horizontal flow (Mainprice *et al.* 2005). Similarly, there are many uncertainties regarding the origin of observed seismic anisotropy in the MTZ. Seismic anisotropy in the upper MTZ (410–520 km depth) may be attributed to the LPO of wadsleyite as it can reach up to 14 per cent single-crystal V_S anisotropy (e.g., Zha *et al.* 1997). However, the dominant slip system is still unclear. Kawazoe *et al.* (2013) found that the [001](010) slip system is dominant in hydrous wadsleyite, and interpreted fast vertically polarized shear waves in the MTZ as indicating the preferred orientation of wadsleyite in horizontal mantle flow. Other potential slip system have also been proposed (e.g. Thurel *et al.* 2003; Farla *et al.* 2015; Ohuchi *et al.* 2015) and LPO patterns may depend on the water content (Demouchy *et al.* 2011). Ringwoodite, which can be found in the lower part of the MTZ (between 520 and 670 km depth) is generally considered isotropic due its cubic structure (Kiefer *et al.* 1997) and therefore not a candidate to explain MTZ anisotropy. Garnet and clinopyroxene do not have sufficient single crystal anisotropy to yield any observable seismic anisotropy (Mainprice *et al.* 2005) Finally, SPO related to layering inside subducting slabs could also result in seismic anisotropy (Faccenda *et al.* 2019).

7 CONCLUSION

We developed a nonlinear inversion method based on Markov Chain Monte Carlo within a transdismensional Bayesian framework to constrain Earth's interior structure with fundamental and higher mode surface wave dispersion measurements. Besides the ability to solve the nonlinear problem, the advantages of this approach include the fact that the choice of the depth parametrization is iteratively driven by the data. The results are therefore not influenced by an arbitrary prior choice of the number of model parameters. Another advantage lies in the ability to invert the noise level jointly with the model parameters, which reduces the risk of mapping unknown data noise into the models. In addition, the use of a Monte Carlo approach enables us to sample all possible models and to represent the ensemble of solutions with a distribution of parameters instead of choosing one model among many possibilities. This allowed us to account for model uncertainties by calculating a likelihood for every parameter of interest.

We applied our technique on a subset of the global phase velocity maps obtained by Visser *et al.* (2008b) for Love and Rayleigh waves up to the fifth and sixth overtones, respectively. We focused on the Indian Ocean and determined how well these data can constrain radial anisotropy in the deep upper mantle and transition zone. We found that correcting the measurements for the effect of perturbations in the depth of the MTZ boundaries did not impact the

resulting models significantly, despite the higher mode data having sensitivity to these depths. We also demonstrated that separate Love and Rayleigh wave inversions produced results equivalent within model uncertainties to those of joint inversions, as long as the depth parametrizations for V_{SV} and V_{SH} are allowed to differ in the joint inversions.

The velocity models resulting from the mean of the posterior model distribution were in general agreement with results from linearized inversions, though with stronger negative anomalies along ridges in the upper 100–200 km. Stronger deviations between the models were found in the MTZ. This suggests that using a fully nonlinear TB method to constrain shear wave velocities with higher mode surface waves may not result in significant changes in the model compared to regularized inverse methods at least in the upper mantle above the MTZ. Our results show, however, that using such as a technique and quantifying model uncertainties is essential when discussing models of radial anisotropy. Large discrepancies were indeed found between the two methods for ξ, even at 100 km depth where the data sensitivity was high, and model uncertainties were larger below ∼300 km. Despite these large posterior model uncertainties, we were able to calculate the likelihood of anisotropy at different depths and determine how well-resolved deep anisotropy signal is. Overall, while the dominant signal at 100 km is $V_{SH} > V_{SV}$ as expected for seismic anisotropy resulting from horizontal mantle motion, we found likely lateral variations in ξ at greater depths, with a tendency for the data to favour a small amount of anisotropy with $V_{SV} > V_{SH}$ across most of the region in the deep upper mantle and MTZ. The synthetic tests we performed also demonstrated that the anisotropy signal found at these depths is resolved by the higher mode data and does not result from leakage. We also found that the inclusion of group velocity dispersion data may help reduce parameter trade-offs in the top 100 km of the mantle but does not improve the model resolution in the deep upper mantle or MTZ.

For our TB inversions, we chose a prior that would not significantly influence the results. Thus, the posterior distribution is mostly a function of the likelihood of a model given the misfit of the predicted data and the observations. This results in models that can explain the data used, but may not necessarily be in agreement with other data sets. This is reflected in some of the very large deviations of ξ from 1 (isotropy), with the posterior distributions sometimes including anisotropy of 25 per cent or more. Additional information could improve constraints on model parameters, guiding the posterior distributions to more realistic values. This could take the form of complementary types of data that reduce trade-offs from the current data set. This could also take the form of a narrower prior, which could be based on mineral physics. This will be implemented in future work.

ACKNOWLEDGEMENTS

E. W. and C. B.were supported by National Science Foundation (NSF) Grant EAR-1446978. Q. H. and N. S. were supported by NSF grant NSF EAR-1447041. The authors thank Mr. and Mrs. Michels for their generous donation to CB's research group, which helped fund this manuscript. We also thank Dr. Joshua Russell and an anonymous reviewer for their helpful comments. The authors do not have any conflict of interest.

DATA AVAILABILITY

The data used in this study were downloaded from http://www.geo.uu.nl/~jeannot/JTweb/downloads.html for the phase velocity maps, and from https://igppweb.ucsd.edu/~gabi/litho1.0.html#surfcodes for the group velocity data. The CRUTST1.0 model employed was downloaded from the IRIS DMC at https://doi.org/10.17611/DP/EMC.1. The data underlying this article are available at https://faculty.epss.ucla.edu/~cbeghein/research/models-download/.

REFERENCES

Anderson, D.L., 1962. Love wave dispersion in heterogeneous anisotropic media, *Geophysics,* **27,** 445–454.

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. Solid Earth,* **119,** 3006–3034.

Babuska, V. & Cara, M., 1991. *Seismic Anisotropy in the Earth,* vol. **10,** Springer Science & Business Media.

Bayes, T., 1763. An essay towards solving a problem in the doctrine of chances, *Philos. Trans. R. Soc.,* **53,** 370–418.

Becker, T.W., Kellogg, J.B., Ekström, G. & O'Connell, R.J., 2003. Comparison of azimuthal seismic anisotropy from surface waves and finite strain from global mantle-circulation models, *Geophys. J. Int.,* **155,** 696–714.

Beghein, C. & Trampert, J., 2004. Probability density functions for radial anisotropy: implications for the upper 1200 km of the mantle, *Earth planet. Sci. Lett.,* **217,** 151–162.

Beghein, C. & Trampert, J., 2006. Radial anisotropy in seismic reference models of the mantle, *J. geophys. Res.,* **111**(B02303), doi:10.1029/2005JB003728.

Beghein, C., Yuan, K., Schmerr, N. & Xing, Z., 2014. Changes in seismic anisotropy shed light on the nature of the Gutenberg discontinuity, *Science,* **343,** 1237–1240.

Bodin, T., Leiva, J., Romanowicz, B., Maupin, V. & Yuan, H., 2016. Imaging anisotropic layering with bayesian inversion of multiple data types, *Geophys. J Int.,* **206,** 605–629.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res. Solid Earth,* **117**(B2).

Burgos, G., Montagner, J.-P., Beucler, E., Capdeville, Y., Mocquet, A. & Drilleau, M., 2014. Oceanic lithosphere-asthenosphere boundary from surface wave dispersion data, *J. geophys. Res. Solid Earth,* **119,** 1079–1093.

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

Crampin, S., 1976. A comment on 'The Early Structural Evolution and Anisotropy of the Oceanic Upper Mantle', *Geophys. J. R. astr. Soc.,* **46**(1), 193–197.

Debayle, E. & Lévêque, J.J., 1997. Upper mantle heterogeneities in the Indian Ocean from waveform inversion, *Geophys. Res. Lett.,* **24,** 245–248.

Debayle, E. & Ricard, Y., 2013. Seismic observations of large-scale deformation at the bottom of fast-moving plates, *Earth planet. Sci. Lett.,* **376,** 165–177.

Demouchy, S., Mainprice, D., Tommasi, A., Couvy, H., Barou, F., Frost, D.J. & Cordier, P., 2011. Forsterite to wadsleyite phase transformation under shear stress and consequences for the Earth's mantle transition zone, *Phys. Earth planet. Inter.,* **184**(1–2), 91–104.

Dziewonski, A.M. & Anderson, D.L., 1981. Premlinary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Ekström, G. & Dziewonski, A.M., 1998. The unique anisotropy of the Pacific upper mantle, *Nature,* **394,** 168–172.

Faccenda, M., Burlini, L., Gerya, T.V. & Mainprice, D., 2008. Fault-induced seismic anisotropy by hydration in subducting oceanic plates, *Nature,* **455**(7216), 1097–1100.

Faccenda, M., Ferreira, A.M., Tisato, N., Lithgow-Bertelloni, C., Stixrude, L. & Pennacchioni, G., 2019. Extrinsic elastic anisotropy in a compositionally heterogeneous Earth's mantle, *J. geophys. Res.: Solid Earth,* **124**(2), 1671–1687.

Farla, R., Amulele, G., Girard, J., Miyajima, N. & Karato, S.-I., 2015. High-pressure and high-temperature deformation experiments on polycrystalline wadsleyite using the rotational Drickamer apparatus, *Phys. Chem. Miner.,* **42**(7), 541–558.

Ferreira, A.M.G., Faccenda, M., Sturgeon, W., Chang, S.-J. & Schardong, L., 2019. Ubiquitous lower-mantle anisotropy beneath subduction zones, *Nat. Geosci.,* **12,** 301–306.

French, S.W. & Romanowicz, B., 2014. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography, *Geophys. J Int.,* **199,** 1303–1327.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M. & Stephenson, J., 2009. Markov chain monte carlo (mcmc) sampling methods to determine optimal models, model resolution and model choice for earth science problems, *Mar. Pet. Geol.,* **26,** 525–535.

Gao, C. & Lekić, V., 2018. Consequences of parameterization choices in surface wave inversion: insights from transdimensional Bayesian methods, *Geophys. J. Int.,* **215,** 1037–1063.

Green, P.J., 1995. Reversible jump markov chain monte carlo computation and bayesian model determination, *Biometrika,* **82,** 711–732.

Green, P.J., 2003. Trans-dimensional mcmc, in *Highly Structured Stochastic Systems,* chap. 6, pp. 179–196, eds Green, P.J., Hjort, N. & Richardson, S., Oxford Statistical Science Series.

Gung, Y., Panning, M. & Romanowicz, B., 2003. Global anisotropy and thickness of continents, *Nature,* **422,** 707–711.

Huang, Q., Schmerr, N., Waszek, L. & Beghein, C., 2019. Constraints on seismic anisotropy in the mantle transition zone from long-period SS precursors, *J. geophys. Res. Solid Earth,* **124,** 6779–6800.

Jackson, D.D., 1972. Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. astr. Soc.,* **28**(2), 97–109.

Karato, S.-I., 2012. On the origin of the asthenosphere, *Earth planet. Sci. Lett.,* **321-322,** 95–103.

Karato, S.-I. & Wu, P., 1993. Rheology of the upper mantle—a Synthesis, *Science,* **260**(5109), 771–778.

Kawazoe, T., Ohuchi, T., Nishihara, Y., Nishiyama, N., Fujino, K. & Irifune, T., 2013. Seismic anisotropy in the mantle transition zone induced by shear deformation of wadsleyite, *Phys. Earth planet. Inter.,* **216,** 91–98.

Kendall, J.-M. & Silver, P.G., 1998. Investigating Causes of D" Anisotropy, in *The Core-Mantle Boundary Region, vol. 18 of* ***Geodynamics****,* pp. 97–116, American Geophysical Union.

Kiefer, B., Stixrude, L. & Wentzcovitch, R.M., 1997. Calculated elastic constants and anisotropy of Mg2SiO4 spinel at high pressure, *Geophys. Res. Lett.,* **24**(22), 2841–2844.

Kirkwood, S., 1978. The significance of isotropic inversion of anisotropic surface-wave dispersion, *Geophys. J. R. astr. Soc.,* **55,** 131–142.

Kustowski, B., Ekström, G. & Dziewonski, 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.

Lanczos, C., 1961. *Linear Differential Operators,* D. Van Nostrand Reinhold Co.

Laske, G., Masters, G., Ma, Z. & Pasyanos, M., 2013. Update on CRUST1.0—A 1-degree Global Model of Earth's Crust, *Geophys. Res. Abstracts,* **15,** Abstract EGU2013–2658.

Lévèque, J., Debayle, E. & Maupin, V., 1998. Anisotropy in the Indian Ocean upper mantle from Rayleigh- and Love-waveform inversion, *Geophys. J. Int.,* **133,** 529–540.

Lynner, C. & Long, M.D., 2014. Lowermost mantle anisotropy and deformation along the boundary of the African LLSVP, *Geophys. Res. Lett.,* **41,** 3447–3454.

Lynner, C. & Long, M.D., 2015. Heterogeneous seismic anisotropy in the transition zone and uppermost lower mantle: evidence from South America, Izu-Bonin and Japan, *Geophys. J. Int.,* **201,** 1545–1552.

Ma, Z., Masters, G., Laske, G. & Pasyanos, M., 2014. A comprehensive dispersion model of surface wave phase and group velocity for the globe, *Geophys. J Int.,* **199,** 113–135.

Mainprice, D., Tommasi, A., Couvy, H., Cordier, P. & Frost, D.J., 2005. Pressure sensitivity of olivine slip systems and seismic anisotropy of Earth's upper mantle, *Nature,* **433**(7027), 731–733.

Masters, G., Woodhouse, J.H. & Freeman, G., 2011. Mineos v1.0.2 [software], Computational Infrastructure for Geodynamics, url: https://geodynamics.org/cig/software/mineos/.

Matsu'Ura, M. & Hirata, N., 1982. Generalized least-squares solutions to quasi-linear inverse problems with a priori information., *J. Phys. Earth,* **30**(6), 451–468.

Mazzullo, A., Stutzmann, E., Montagner, J.-P., Kiselev, S., Maurya, S., Barruol, G. & Sigloch, K., 2017. Anisotropic tomography around La Réunion Island from rayleigh waves, *J. geophys. Res. Solid Earth,* **122,** 9132–9148.

Meier, U., Trampert, J. & Curtis, A., 2009. Global variations of temperature and water content in the mantle transition zone from higher mode surface waves, *Earth planet. Sci. Lett.,* **282,** 91–101.

Montagner, J.-P., 1986. Regional three-dimensional structures using long-period surface waves, *Ann. Geophys,* **4,** 283–294.

Montagner, J.-P. & Anderson, D.L., 1989. Petrological Constraints on Seismic Anisotropy, *Phys. Earth planet. Inter.,* **54,** 82–105.

Montagner, J.-P. & Tanimoto, T., 1991. Global upper mantle tomography of seismic velocities and anisotropies, *J. geophys. Res.,* **96**(B12), 20337–20351.

Moulik, P. & Ekström, G., 2014. An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophys. J Int.,* **199,** 1713–1738.

Ohuchi, T. *et al.*, 2015. In situ observation of crystallographic preferred orientation of deforming olivine at high pressure and high temperature, *Phys. Earth planet. Inter.,* **243**(C), 1–21.

Panning, M. P., Lekić, V. & Romanowicz, B., 2010. Importance of crustal corrections in the development of a new global model of radial anisotropy, *J. geophys. Res.,* **115,** doi:10.1029/2010JB007520.

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.

Ritsema, J., Deuss, A., van Heijst, H.-J. & Woodhouse, J.H., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. R. astr. Soc.,* **184,** 1223–1236.

Sakamaki, T. *et al.*, 2013. Ponded melt at the boundary between the lithosphere and asthenosphere, *Nat. Geosci.,* **6,** 1041–1044.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2005. Trans-dimensional inverse problems, model comparison and the evidence, *Geophs. J. Int.,* **167,** 528–542.

Schaeffer, A., Lebedev, S. & Becker, T., 2016. Azimuthal seismic anisotropy in the Earth's upper mantle and the thickness of tectonic plates, *Geophys. J. Int.,* **207,** 901–933.

Shapiro, N.M. & Ritzwoller, M.H., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J Int.,* **151,** 88–105.

Simmons, N.A., Myers, S.C., Johannesson, G., Matzel, E. & Grand, S.P., 2015. Evidence for long-lived subdiction of an ancient tectonic plate beneath the southern indian ocean, *Geophys. Res. Lett.,* **42,** 9270–9278.

Smith, D.B., Ritzwoller, M.H. & Shapiro, N.M., 2004. Stratification of anisotropy in the Pacific upper mantle, *J. geophys. Res.,* **109,** doi:10.1029/2004JB003200.

Snieder, R. & Trampert, J., 2000. Linear and nonlinear inverse problems, in *Geomatic Method for the Analysis of Data in the Earth Sciences,* pp. 93–164, eds Dermanis, A., Grün, A. & Sansó, F., Springer Berlin Heidelberg, Berlin, Heidelberg.

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation, Society for Industrial and Applied Mathematics*.

Thurel, E., Douin, J. & Cordier, P., 2003. Plastic deformation of wadsleyite: III. Interpretation of dislocations and slip systems, *Physics and Chemistry of Minerals,* **30**(5), 271–279.

Trampert, J., 1998. Global seismic tomography: the inverse problem and beyond, *Inverse Probl.,* **14,** 371–385.

Visser, K., Trampert, J. & Kennett, B.L.N., 2008b. Global anisotropic phase velocity maps for higher mode Love and Rayleigh waves, *Geophys. J. Int.,* **172,** 1016–1032.

Visser, K., Trampert, J., Lebedev, S. & Kennett, B., 2008a. Probability of radial anisotropy in the deep mantle, *Earth planet. Sci. Lett.,* **270,** 241–250.

Wiggins, R.A., 1972. The general linear inverse problem: Implication of surface waves and free oscillations for Earth structure, *Rev. Geophys.,* **10,** 251–285.

Woodhouse, J.H. & Dahlen, F.A., 1978. The effect of a general aspherical perturbation on the free oscillations of the Earth, *Geophys. J. R. astr. Soc.,* **53,** 335–354.

Xing, Z. & Beghein, C., 2015. A Bayesian approach to assess the importance of crustal corrections in global anisotropic surface wave tomography, *Geophys. J. Int.,* **203**(3), 1832–1846.

Yuan, H. & Romanowicz, B., 2010. Lithospheric layering in the North American craton, *Nature,* **466,** 1063–1068.

Yuan, K. & Beghein, C., 2013. Seismic anisotropy changes across upper mantle phase transitions, *Earth planet. Sci. Lett.,* **374,** 132–144.

Yuan, K. & Beghein, C., 2014. Three-dimensional variations in Love and Rayleigh wave azimuthal anisotropy for the upper 800 km of the mantle, *J. geophys. Res. Solid Earth,* **119**(4), 3232–3255.

Zha, C.-S., Duffy, T.S., Mao, H.-k., Downs, R.T., Hemley, R.J. & Weidner, D.J., 1997. Single-crystal elasticity of $\beta -Mg_2\ SiO_4$ to the pressure of the 410 km seismic discontinuity in the Earth's mantle, *Earth planet. Sci. Lett.,* **147**(1–4).

Zhou, Y., Nolet, G., Dahlen, F.A. & Laske, G., 2006. Global upper mantle structure from finite-frequency surface-wave tomography, *J. geophys. Res.,* **111,** B04304.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. $d\ln V_S$ model at 100 km depth resulting from linearized inversions with 15 (left) and 18 (right) independent parameters. Top: $d\ln V_S$ calculated from inversions parametrized in terms of V_{SV} and V_{SH} perturbations. Bottom: $d\ln V_S$ resulting from inversions parametrized in terms of $d\ln V_S$ and $d\ln\xi$.

Figure S2: ξ model at 100 km depth resulting from linearized inversions with 15 (left) and 18 (right) independent parameters. Top: ξ calculated from inversions parametrized in terms of V_{SV} and VSH perturbations. Bottom: resulting from inversions parametrized in terms of $d\ln V_S$ and $d\ln\xi$.

Figure S3: Comparison between overtone phase velocity maps measurements ((a) and (b) and calculated contributions to the phase velocities due to MTZ topography ((c) and (d)). The phase velocity data are from Visser *et al.* (2008a) and the MTZ model is from Huang *et al.* (2019). Data are for the third overtone Rayleigh wave at a period of 78 s (top) and for the third overtone Love wave at 69 s (bottom). Both are sensitive to structure at MTZ depths. The calculated effect of the topography MTZ on the phase velocities is more than an order of magnitude smaller than the measured anomalies.

Figure S4: Comparisons of SVD inversion results with and without corrections for the effect of MTZ boundary topography. The dashed lines indicate results from inversions of data corrected for MTZ boundary topography and the solid lines indicate inversions of data with no MTZ boundary topography correction. (a) is for the grid cell at 3°S, 79°E and (b) is for 18°S, 89°E. All models shown have 18 independent parameters.

Figure S5: Synthetic test comparing (a) V_{SV} and (b) V_{SH} models obtained by independent and joint Love and Rayleigh wave inversions using our TB method. The solid black line is the input model, the solid red and blue lines represent the mean model, and the dashed lines represent their 90 per cent confidence interval. (c) and (d) represent the resulting V_S and ξ models, respectively, and their 90 per cent confidence interval

Figure S6: Posterior distribution of noise parameters for inversion of (a) all Rayleigh wave phase velocities and (b) all Love wave phase velocities at 3°S, 79°E using our nonlinear TB inversion. The noise parameter is the standard deviation of the Gaussian distribution that represents the uncertainty around the measured velocity. The vertical dashed line in (a) and (b) represent the noise level estimated by Visser *et al.* (2008a). The TB method finds that the Rayleigh wave data are well explained with a lower level of noise than the Love wave data. (c) and (d) represent the distribution of V_S and ξ models obtained, respectively. The colour scale represents the likelihood and the dashed black line is the mean model.

Figure S7: (a)–(d) Posterior distribution of noise parameters for inversion of Rayleigh wave phase velocities and Love wave phase velocities at 3°S, 79°E using our TB technique. The noise parameter is the standard deviation of the Gaussian distribution that represents the uncertainty around the measured velocity. The vertical dashed line in (a) and (b) represent the noise level estimated by Visser *et al.* (2008a). The Rayleigh wave data were split into two groups: one group with fundamental modes and the first two overtones ($n \geq 2$) and the other group with the remaining higher modes ($n \leq 3$). The same was done for the Love wave data set. Each group has their own noise parameters, as shown above. (e) and (f) represent the distribution of V_S and ξ models obtained, respectively. The colour scale represents the likelihood and the dashed black line is the mean model.

Figure S8: Posterior distribution of noise parameters for inversion of (a)–(g) Rayleigh wave phase velocities and (h)–(m) Love wave phase velocities at 3°S, 79°E using our TB method. The noise parameter is the standard deviation of the Gaussian distribution that represents the uncertainty around the measured velocity. The vertical dashed line in (a) and (b) represent the noise level estimated by Visser *et al.* (2008a). Each mode branch has its own noise parameter, as shown above. For Love waves, the fundamental mode is found to be the least noisy data while the first or fourth overtone is found to be the noisiest data set. For Rayleigh waves, each of the branches that have a corresponding Love wave branch ($n = 0$–5) is less noisy than its Love wave counterpart. (n) and (o) represent the distribution of V_S and ξ models obtained, respectively. The colour scale represents the likelihood and the dashed black line is the mean model

Figure S9: Posterior distribution of noise parameters for inversion of (a) Rayleigh wave phase and (b) group velocities and (c) Love wave phase and (d) group velocities at 3°S, 79°E using the nonlinear TB. The noise parameter is the standard deviation of the Gaussian distribution that represents the uncertainty around the measured velocity. Rayleigh wave phase velocity data are found to be less noisy than Love wave phase velocity data, and this relationship remains true for Rayleigh wave group velocity and Love wave group velocity data. Group velocity data are also noisier than the phase velocity counterpart.

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.