Geophysical Journal International

Geophys. J. Int. (2018) **215,** 1037–1063
Advance Access publication 2018 July 28
GJI Seismology

doi: 10.1093/gji/ggy310

Consequences of parametrization choices in surface wave inversion: insights from transdimensional Bayesian methods

Chao Gao and Vedran Lekić

Department of Geology, University of Maryland, College Park, MD 20740 USA. E-mail: cgao1@terpmail.umd.edu

Accepted 2018 July 26. Received 2018 July 10; in original form 2018 February 01

SUMMARY

Inversion of surface wave data for crustal and upper-mantle structure is a staple of passive seismology, particularly since the advent of techniques enabling surface wave dispersion (SWD) and Rayleigh wave ellipticity measurements from ambient noise. Recent development and application of transdimensional Bayesian (TB) seismic inversion offers an approach to quantify model parameter uncertainties and trade-offs with fewer assumptions than traditional methods. Using synthetic tests, we investigate choices in the implementation of TB for the inversion of SWD and Rayleigh wave ellipticity to constrain the structure of Earth's continental lithosphere. We focus on three aspects of the inversion: limitation of data sensitivity, assumed scaling among parameters (compressional wave speed, *Vp*, shear wave speed, *Vs*, density and radial anisotropy) and parametrization choices. We show that while surface wave data provide relatively strong constraints on the posterior distribution of *Vs* and, to a lesser extent, *Vp*, common parametrization choices can potentially bias structure estimates. This is particularly the case for radial anisotropy (ξ), due to the inability to distinguish variations of *Vp* and density from those of ξ. Inferred results therefore depend substantially on the parametrization and scaling choices. We illustrate how layered parametrizations can, in the TB framework, recover smoothly varying profiles, and quantify the number of layers recoverable at different levels of measurement uncertainty. Finally, we propose two types of model parametrization for TB inversion involving multiple types of parameters. We demonstrate that by implementing an independent parametrization for different physical quantities, we can avoid imposing identical model geometry across multiple types of model parameters, and obtain better model estimates with reduced trade-offs. We advocate for such a parametrization in TB inversion of radial anisotropy using surface wave data, and when targeting disparate *Vp* and *Vs* structures such as those associated with α-β quartz transtion.

Key words: Inverse theory; Crustal imaging; Seismic anisotropy; Surface waves and free oscillations.

1 INTRODUCTION

Seismic surface waves are strongly excited by shallow earthquakes, and more easily recorded at large epicentral distances compared to body waves due to lesser geometrical spreading. Because their sensitivity to structure depends on frequency, their propagation velocity does as well. Therefore, measurements of surface wave dispersion (SWD) provide constraints on crustal and upper-mantle structure with excellent global coverage and high lateral resolution (e.g. Romanowicz 2002). Indeed, seismic tomography based on SWD has been widely used to constrain the 3-D shear wave velocity in crust and upper mantle both on a global scale (Montagner & Tanimoto 1991; Trampert & Woodhouse 1996; Shapiro & Ritzwoller 2002; Ekstrom 2014; Pasyanos *et al.* 2014) and regional scale (e.g. Simons *et al.* 2002; Huang *et al.* 2003; Yao *et al.* 2006, 2008; Lin *et al.* 2008; Wagner *et al.* 2012). Differences between the dispersion of Rayleigh and Love waves led to the discovery of radial anisotropy in the upper mantle (Anderson 1961; Aki 1968), and are now routinely used to constrain profiles and lateral variations of radial anisotropy of Earth's crust and upper mantle (Panning & Romanowicz 2004; Kustowski *et al.* 2008; Ferreira *et al.* 2010). The particle motion of fundamental mode Rayleigh wave is elliptical. The measurement of the ratio of the vertical to horizontal amplitude of particle motion (ZH ratio) can be shown to be sensitive to elastic properties and density in the crust, and has also been used to constrain shallow Earth structure (Boore & Toksoz 1969; Tanimoto & Rivera 2008; Yano *et al.* 2009; Lin *et al.* 2014).

Inferring Earth properties from seismic data is a non-unique inverse problem because seismic observables provide only limited constraints (Franklin 1970). Love and Rayleigh waves depend on density and 13 independent elastic parameters (e.g. Chen & Tromp 2007), which can vary laterally and with depth. Therefore, inversions of SWD and ZH ratio data inherently involve significant non-uniqueness due to trade-offs among model parameters; when linearized and posed in terms of matrix inversion, the large number of model parameters yields singular matrices requiring regularization for a solution to be obtained. To best represent the Earth's structure given the available seismic observation, certain assumptions are often made to simplify the model. For example, the structure may be assumed to be layered or to vary smoothly with depth (e.g. Constable *et al.* 1987). Within each layer, the elastic properties might be assumed to be isotropic, so that they can be described with just three parameters: density (ρ), shear (V_S) and compressional (V_P) wave speed. Another common assumption is that of radial anisotropy (transverse isotropy), which involves three additional parameters: the squares of the ratios of wave speeds of horizontally and vertically-polarized waves, $\xi = (V_{SH}/V_{SV})^2$ and $\varphi = (V_{PV}/V_{PH})^2$, respectively, as well as a parameter, η, describing wave speeds at intermediate directions.

Even with these simplifying assumptions, constraints provided by SWD data are insufficient to reliably infer all the model parameters, particularly those to which the seismic observables are weakly sensitive—such as Vp, φ, η and ρ. Based on empirical trends, variations in these parameters are often scaled to variations in better-resolved parameters such as Vs (e.g. Brocker 2005) and ξ (e.g. Montagner & Anderson 1989). Indeed, φ and η are assumed and remain a workhorse of structural seismology. Recently, global tomographic models in which the scaling relationships are allowed to vary with depth or laterally have also been performed (e.g. Simmons *et al.* 2009; Moulik & Ekstrom 2016).

Standard inversion approaches are ill-suited for studying how parametrization choices and scaling assumptions affect the accuracy of seismic structures inferred from surface wave data. To start with, a certain amount of regularization—in the form of smoothing, damping, or *a priori* covariance among parameters—must be imposed to obtain a solution in the first place. Uncertainty analysis developed for linear(izable) problems can be applied (e.g. Backus & Gilbert 1967; Tarantola & Valette 1982) to study the trade-offs between inferences of Vs, Vp and density. However, SWD and ZH ratios depend on elastic properties in a non-linear way; their sensitivity to a parameter of interest can depend on the value of that and other parameters. Therefore, these linear approaches may not be appropriate. Even if they were appropriate, the analysis may depend on the parametrization—for example, for different choices of number and thickness of layers—limiting their generalizability. On the other hand, model space search methods do not require an inversion to be performed, and offer the potential to quantify the uncertainty of inferences even in highly non-linear problems (Mosegaard 1998). Yet, application of these approaches to overparametrized problems is stymied by the curse of dimensionality (e.g. Tarantola. 2005), and has, until recently, required the parametrization to be chosen prior to inversion.

To better represent the uncertainties from seismic imaging results and to incorporate complementary seismic observables with increasingly available measurements, seismic transdimensional Bayesian (TB) inversion has been developed (Malinverno 2002; Bodin & Sambridge 2009; Agostinetti & Malinverno 2010; Bodin *et al.* 2012). Under a Bayesian framework, all information is described in terms of probabilities. This allows for data uncertainties as well as prior assumptions about model parameters to be explicitly accounted for. Furthermore, since Bayesian inversion seeks an ensemble solution instead of a single best-fitting model, quantification of uncertainties of inferred model parameters and correlations between them is relatively straightforward. In contrast to traditional inversion methods, which treat the number of model parameters as a constant chosen prior to inversion, transdimensional inversion includes it as an unknown determined by the data (Sambridge *et al.* 2013). With a more flexible model parametrization, transdimensional inversion also more easily accommodates multiple data types with different, and therefore complementary, sensitivities to the seismic structure.

The TB method offers a new opportunity to quantify effects of parametrization choices and assumptions of scaling among parameters, enabling a reassessment of uncertainties in SWD and insight into outstanding questions, such as the origin of the relatively poor agreement among radially anisotropic global shear velocity models (Chang *et al.* 2015). Under a TB framework, we can eliminate scaling assumptions as well as assumptions concerning the number and thickness of structural layers, while simultaneously constraining multiple model parameters to various degrees.

In this paper, we use TB inversion to systematically explore the ability of SWD and ZH ratios to constrain profiles of Vs, Vp, ρ and ξ beneath a seismic station, under various model parametrization choices. Synthetic data of these two observables are inverted individually and jointly to investigate the complementarity of data sensitivity, the consequences of parametrization choices, and the influence of assumptions about scaling relationships between physical quantities.

2 METHOD

2.1 Seismic Bayesian inversion

Bayes' theorem (Bayes & Price 1763) relates the probability (p) of a model (m) conditional on a data set (d), written as $p(m|d)$, to the probability of observing the data set given a model, i.e. $p(d|m)$:

$$p\left(m|d\right) = \frac{p\left(d|m\right) \times p\left(m\right)}{p\left(d\right)} \tag{1}$$

Here, the model is represented by a vector quantity that includes all the model parameters of interest. Similarly, all observed data comprise a vector d. The aim of Bayesian inference is to quantify the posterior probability density $p(m|d)$, which is the probability density of the model parameter given the observed data (Smith 1991). The term $p(d)$ is called the evidence. Note that $p(d)$ is not a function of m, and should remain constant as we vary the model parameter under the same setting, allowing us to write:

$$p\,(m|d)\ \propto\ p\,(d|m) \times p\,(m) \tag{2}$$

Seismic inversion is the procedure of using the measurements made on seismic records (i.e. data, d) to infer a model (m) that quantitatively describes the Earth's, typically inaccessible, interior. In such case, the posterior is the probability of certain seismic structure given the observation and the prior. It is therefore proportional to the product of likelihood—the probability of observing the data given a seismic structure—and the prior probability on the model parameters. While the forward problem of predicting the outcome of some measurements given a complete description of the physical system has a unique solution, the inverse problem does not. This non-uniqueness arises both from data measurement errors and the insufficiency of information contained in the data. Unlike many common approaches to solving such inverse problems, which seek to reduce the non-uniqueness by introducing prior information in the form of smoothing or damping operators (e.g. Constable *et al.* 1987; Menke 2012), the Bayesian approach embraces the non-uniqueness and represents it in probabilistic terms.

2.1.1 *The prior*

In the Bayesian framework, the prior information $p(m)$ is used to describe our knowledge about the parameters that describe the model prior to introducing data (Sivia & Skilling 2006).

If the parameters that we are interested in inferring correspond to an unknown number n of physically non-overlapping regions, the prior can be separated into two terms:

$$p\,(m) = p\,(m_e, n) = p\,(m_e|n) \times p(n) \tag{3}$$

Here, m_e stands for the parameters describing the seismic structure (Vp, Vs, density and the physical location of the regions). We use a uniform distribution for n over the interval $I\ = \{n \in N \mid\ n_{\min} < n \leq n_{\max}\}$. Hence,

$$p(n) = \begin{cases} \frac{1}{\Delta n} & \text{if } n \in I \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

where $\Delta n = (n_{max} - n_{min})$.

In this study, the target model is parametrized in depth using Voronoi nuclei (Aurenhammer 1991); the region nearest to a given Voronoi nucleus is defines a layer of constant elastic parameters specified for that Voronoi nucleus. The boundary between adjacent layers is defined as the midpoint between two Voronoi nuclei (see Fig. 1). Since we are interested in the profiles of multiple types of parameters, we propose two different parametrization schemes, illustrated in Fig. 1.

For the first type of parametrization, we allow each Voronoi nucleus to specify all types of parameters. We call this type of parametrization 'attached'. Given a number of cells n, the probability distributions for the 4 $\times\, n$ parameters, 1-D Voronoi nucleus position (z), shear velocity (Vs), compressional velocity (Vp) and density (ρ) are assumed to be independent from each other, and so can be written as:

$$p\,(m_e|n) = p\,(z|n) \times p\,(v_s|n) \times p\,\left(v_p|n\right) \times p\,(\rho|n) \tag{5}$$

To minimize the amount of prior information introduced, we assume uniform distributions over specific intervals. For example, if we define $J_s = \{v_{s,i} \in \mathcal{R} \mid V_{\min} < v_i < V_{\max}\}$, we have:

$$p(v_{s,i}|n) = \begin{cases} \frac{1}{\Delta v_s} & \text{if } v_{s,i} \in J_s \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $\Delta v = (V_{\max} - V_{\min})$. Since the shear velocity in each Voronoi nucleus is assumed to be independent (i.e. no smoothing is imposed),

$$p\ \,(v_s|n) = \prod_{i=1}^{n} p(v_{s,i}|n) \tag{7}$$

Similarly, we can write:

$$p\ \,\left(v_p|n\right) = \prod_{i=1}^{n} p(v_{p,i}|n) \tag{8}$$

$$p\ \,(\rho|n) = \prod_{i=1}^{n} p(\rho_i|n) \tag{9}$$

For a 1-D-layered model, the possible positions of the Voronoi nuclei are distributed along depth. If we assume that there are N possible positions for n Voronoi nuclei, there are then $\frac{N!}{n!(N-n)!}$ possible configurations. Again, we assign an equal probability to each of the configurations, and can then write:

$$p\ \,(z|n) = \left[\frac{N!}{n!\,(N-n)!}\right]^{-1} \tag{10}$$

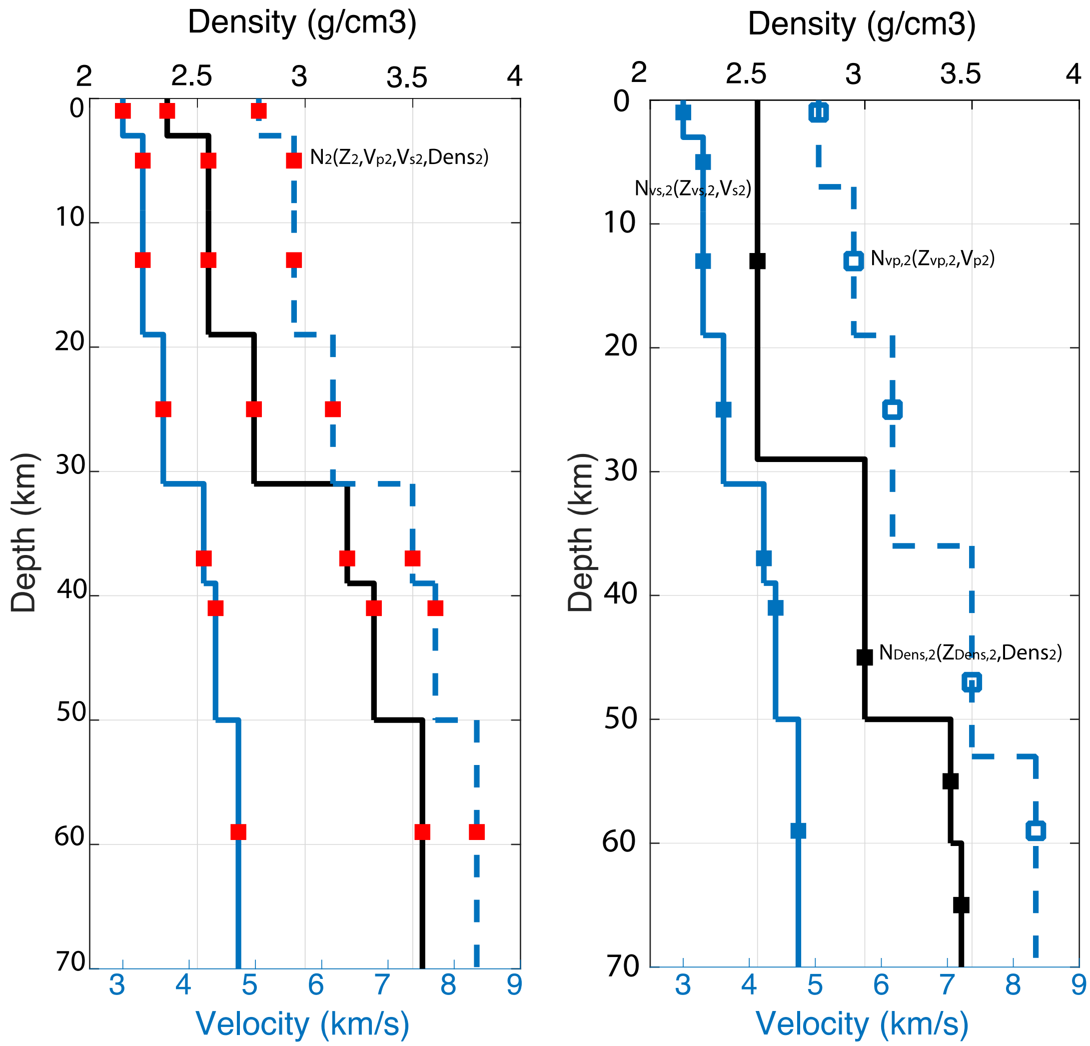

Figure 1. Two schemes for using Voronoi nuclei (squares) to characterize a 1-D isotropic seismic velocity model (*Vs*—blue, *Vp*—dashed and *ρ*—black). In the 'attached' scheme (left-hand panel), each Voronoi nucleus carries all three parameters (*Vs*, *Vp* and density). In the 'independent' scheme (right-hand panel), the three types of parameters are carried by independent sets of Voronoi nuclei. The second Voronoi nucleus is labelled in both panels to illustrate differences between the two parametrization schemes. The model shown in left-hand panel is the target model used to generate synthetic data inverted in later sections.

Combining together eqs (3)–(10), the full prior probability density function (PDF) can be written as:

$$p\,(m) = \begin{cases} \frac{n!(N-n)!}{N!(\Delta v_s)^n \left(\Delta v_p\right)^n (\Delta\rho)^n \Delta n} & if\ \left(n \in I \text{ and } \forall i \in [1, n],\ v_{si} \in J_s, v_{pi} \in J_p,\ \rho_i \in J_\rho\right) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

For the second type of parametrization, we assign independent sets of Voronoi nuclei to each type of parameter. Hence, we call this type of parametrization 'independent'. In this case, we have

$$p\ (m_e|n) = \prod_{j=1}^{3} p\left(z|n_j\right) \times p\left(m_{e,j}|n_j\right) \quad (12)$$

Here, $m_{e,\ j}$ stands for the elastic parameters (*Vs*, *Vp*, and *ρ*) the Voronoi nuclei carry. Unlike the first scheme, the number and the position of the Voronoi nuclei are independent from each other for different elastic parameters. In this way, we do not force all type of elastic parameters to be attached to a single Voronoi nucleus, which ideally will allow a more flexible parametrization. Similarly, we have

$$p\ \left(m_{e,j}\right) = \frac{n_j!\left(N-n_j\right)!}{N!\left(\Delta m_{e,j}\right)^{n_j}\Delta n_j} \quad if\ \left(n_j \in I \text{ and } \forall i \in \left[1, n_j\right],\ m_{e,j} \in J\right) \quad (13)$$

Each of the two types of parametrization has certain advantages for particular problems; we will further explore this in this paper in the joint inversion of SWD and ZH ratios in the discussion section. Bodin *et al.* (2016) described an alternative type of parametrization where additional parameters constraining anisotropy are proposed on existing isotropic shear velocity structures, and such proposed model is accepted based on the constraints from data only. Since all the anisotropic parameters are proposed attaching to the velocity layers, the geometry of the anisotropic structure will depend on the velocity structure to some degree. This kind of parametrization appears to lie between the two we proposed in terms of the dependence among different types of parameters.

2.1.2 Likelihood function

The likelihood $p(d|m)$ quantifies how likely we would be to observe the data if the actual structure were described by the set of parameters in vector m. We use a least-squares misfit function to describe the consistency between the predicted and observed data:

$$\Phi\ (m) = \left\| \frac{g\,(m) - \ d^2}{\sigma_d} \right\| \tag{14}$$

where $g(m)$ is the predicted data and σ_d^2 is the estimated variance describing the data uncertainties. This misfit function is appropriate for data with normally-distributed errors, and yields the following likelihood:

$$p\,(d|m) \propto \exp\left(\frac{-\Phi\,(m)}{2}\right) \tag{15}$$

In the rest of this paper, we do not explicitly contaminate our synthetic data with noise because our likelihood function takes into account the effect of noise if we assume it to be uncorrelated across different periods, as is commonly done in the literature.

2.2 Transdimensional sampling

Bayes' theorem quantifies how the posterior distribution is affected by the choice of the prior. The assumptions we make in formulating the inversion influence the outcome. In seismic inversion, assumptions about number of parameters are often made to fit the linearized inverse problem and to reduce non-uniqueness. These assumptions are often motivated by previous knowledge about the studied region. The risk in making these assumptions is that they could be biased or incorrect. The geophysical inversion literature abounds in examples in which the choice of the parametrization affects the inversion to different extents due to different degrees of correlation among model parameters. As an example, Trampert & Snieder (1996) showed how truncated expansions of basis functions could bias seismic tomography models. The motivation for applying a transdimensional sampling method into the inversion is to allow flexibility that does not require, but can nevertheless accommodate, strong prior assumptions about the model parametrization.

Allowing a flexible parametrization without any regulation may lead to another problem, where the model will contain complexities arising from attempting to fit details of the data as closely as possible. Since the data we measure contains error due to both instrumental and environmental noise sources, fitting the detailed data is ill-advised. The Bayesian formulation of model selection is naturally parsimonious (Malinverno 2000; Sivia & Skilling 2006); Malinverno (2002) showed that this is also the case with TB inversion. This means that if we have two competing models with different numbers of parameters that both fit the data equally well, the Bayesian formulation will favour the simpler model. Combining the transdimensional sampling method and the Bayesian framework, TB inversion therefore allows a more flexible parametrization with fewer assumptions made.

2.3 Reversible-jump Markov Chain Monte Carlo for multiparameter seismic structure

We apply a reversible-jump Markov Chain Monte Carlo (rjMCMC) algorithm to carry out the TB inversion. The MCMC is an iterative algorithm that draws random steps from a desired distribution; with sufficient number of iterations, the models are sampled proportional to their posterior probability, $p(m|d)$. The rjMCMC algorithm consists of two stages, proposing a new model (m') by perturbing the current model (m) and deciding accepting or rejecting it.

In a case of transdimensional sampling, the acceptance probability is:

$$\alpha = \min\left[1, \frac{p\,(m')\,p\,(d|m')\,q\,(n, m|n', m')}{p\,(m)\,p\,(d|m)\,q\,(n', m'|n, m)}\,|\boldsymbol{J}|\right] \tag{16}$$

Note that here the proposal ratio is different than in the fixed parametrization case. **J** is the Jacobian matrix of the transformation from m to m'. It is needed to account for the scale change only when there is a dimension change during the sampling process (Green 2003). In our case of discrete Voronoi positions, |**J**| equals to one (Bodin & Sambridge 2009). Therefore, the Jacobian is unity for each case of the rjMCMC sampling process and can be ignored.

An important part of designing an rjMCMC is choosing how to perturb the current model m into m' with some randomness, i.e. how to efficiently sample the parameter space. A schematic representation of our rjMCMC algorithm is shown in Fig. 2. Following the approach of Bodin & Sambridge (2009), we perturb the current model by randomly choosing one the four options with equal probability. However, since we propose two parametrization schemes for dealing with the multiparameter seismic structure, the Markov Chain could behave differently, especially when dimension changes are involved. We derive the relevant expressions in the Appendix A.

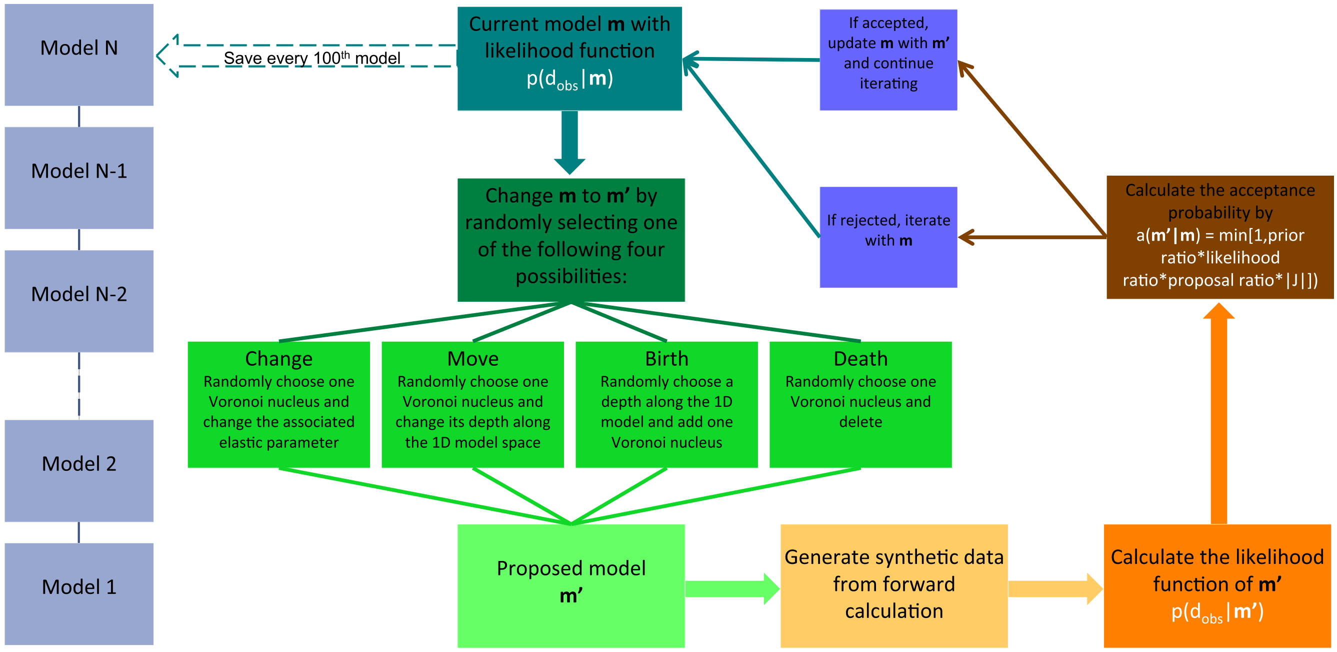

Figure 2. A schematic overview of the reversible-jump Markov Chain Monte Carlo algorithm used for inversion. Every 100th model is saved to the ensemble solution to increase the independence of sampled models.

2.4 Target model and forward problem

To test the performance of the TB joint inversion, we choose a realistic isotropic, layered target model (left-hand panel of Fig. 1) for the synthetic tests. The target model has a 3 km thick sedimentary layer, underlain by a two-layer crystalline crust with Moho at 31 km depth. The upper mantle shallower than 70 km is represented by three layers with increasing velocity. The velocities remain constant below 70 km. In the target model, *Vs*, *Vp* and density follow the empirical relations from Brocker (2005). The target model is designed in this way for the convenience of later discussion of scaling relationship effects. To predict SWD and ZH ratios, we use the reflectivity method (Hisada 1994; Aki & Richard 2002) to solve the eigenvalue problem for both Rayleigh wave and Love wave in an elastic, vertically heterogeneous medium, based on the implementation of Lai & Rix (1998). Later in the discussion about TB inversion of radial anisotropy, we modify the forward code to compute the SWD given elastic parameter A, C, N, L and F according to Harkrider & Anderson (1962) and Bhattacharya & Arora (1997). The code also takes into account the sphericity of the Earth based on the formulation of Bhattacharya (1996). We validate our implementation by comparing our predictions to those from MINEOS (Masters *et al.* 2011) for the upper 200 km of the Preliminary reference Earth model (PREM) (Dziewonski & Anderson 1981). Given the same period range, our implementation costs around 0.1 s to predict SWD, which is much faster than MINEOS. All of the software is written in MATLAB.

3 RESULTS

3.1 TB inversion of SWD

We perform TB inversion of both Rayleigh and Love wave dispersion data computed from synthetic input structures in the 5–100 s period range. This period range covers both the ambient noise data range and part of the teleseismic data range. For teleseismic earthquake data, the SWD data are usually measured between approximately 30 and 250 s (Laske & Masters 1996; Ekstrom *et al.* 1997; van Heijst & Woodhouse 1999; Boschi & Ekstrom 2002; Trampert & Woodhouse 2003; Ekstrom 2011; Ma *et al.* 2014). The dispersion data below 25 s period are relatively difficult to measure from teleseismic data due to scattering and potential for cycle skipping. Dispersion measurements made on ambient noise correlations are typically in the ∼5–40s range (e.g. Ekstrom 2014). The combination of these two period ranges comprehensively constrains *Vs* in the crust and upper mantle. We choose to not include dispersion data at periods larger than 100 s, because their primary sensitivity is below the depth range of interest in this manuscript (upper 70 km). The uncertainty of SWD measurements can be affected by data quality (e.g. signal-to-noise ratio), data coverage (e.g. distribution of earthquakes and stations), and measurement method (e.g. whether or not the smoothness of the dispersion curves is exploited). We assign a realistic 3 per cent uncertainty to the dispersion measurements at each period, and assume that measurements at different frequencies are uncorrelated. It should be noted that the assumption of uncorrelated data uncertainty, while ubiquitous in the literature, warrants further investigation.

In this test, we only invert for *Vs* due to the limited constraints SWD data have for *Vp* and density. We assign an uniform prior between 2.5 and 5.5 km s^{-1} for *Vs*. Bodin & Sambridge (2009) suggested that when the data constraints are strong enough, the choice of the broad prior

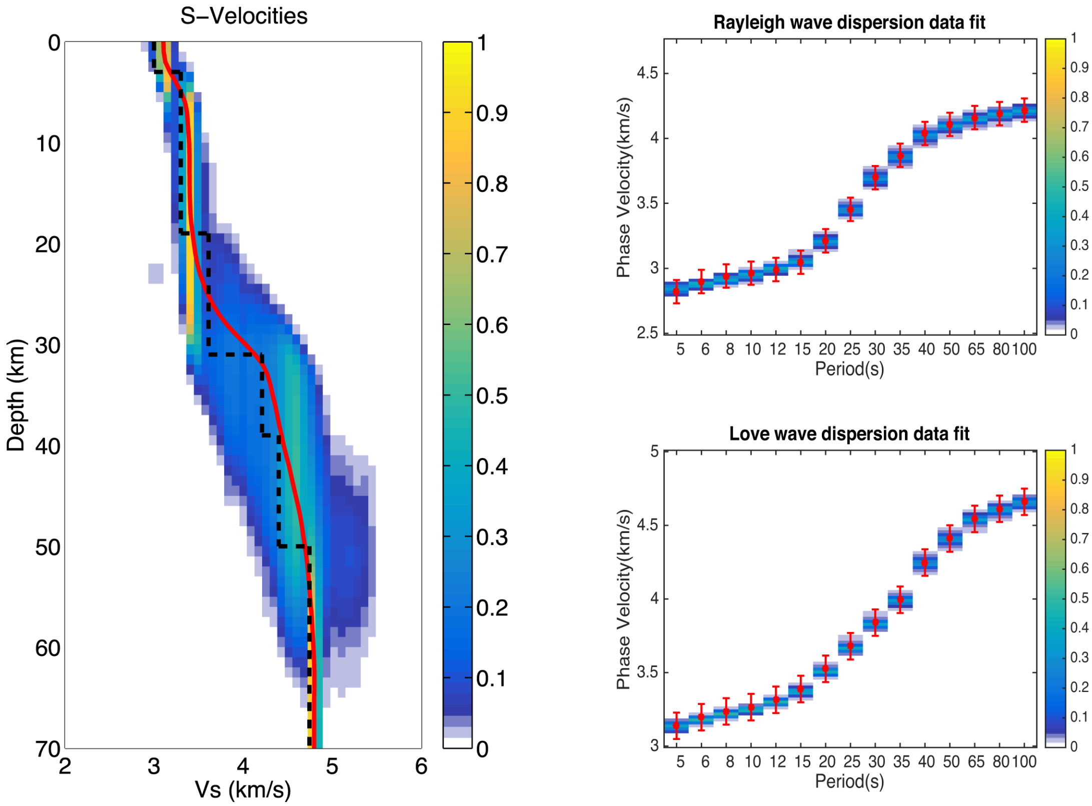

Figure 3. Left: *Vs* depth distributions retrieved using transdimensional Bayesian inversion of synthetic surface wave dispersion data. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) is used to generate synthetic data. Synthetic data (red) with specified 1σ uncertainties for Rayleigh (right top) and Love (right bottom) waves; data predicted by the ensemble solution plotted as probability density functions.

has little effect on the posterior. Our synthetic tests with different priors for *Vs* support such conclusion. Meanwhile, the choice of prior range for weakly constrained model parameters such as density during SWD inversion could significantly affect the posterior. For example, when inverting for *Vs* and density using SWD with independent parametrization, we find that broadening the prior range twice for density would resulting the preferred number of layer from the posterior to change from 5 to 3, while broadening the *Vs* prior range does not change the preferred number of layer from the posterior. During the inversion, the *Vp* and density are calculated using the empirical relations according to Brocker (2005), as used in the synthetic model (see Section 2.4 and Fig. 1). We want to point out that assuming a scaling relationship imposes additional prior information, requiring fewer model parameters to be inverted for. When the scaling relation is properly chosen, it will help reduce the variation in the posterior yielding tighter constraints on parameters of interest. Theoretically, an incorrect scaling relation on *Vp* and *Vs* will bias the estimates of both parameters. However, our synthetic tests suggest that for typical *Vp*/*Vs* ranges for crustal studies, imposing incorrect *Vp*/*Vs* only biases *Vp* estimates, while the *Vs* estimates remain indistinguishable from the posterior obtained with the correct *Vp*/*Vs*. We find this to be the case even when only Rayleigh wave dispersion is used in the inversion. We attribute this behaviour to the much greater sensitivity of *Vs* compared to *Vp* for SWD data. On the other hand, not assuming a scaling relation corresponds to a less informative prior; due to the naturally parsimonious nature of Bayesian inference, given the same observation, this will lead to a more simplified posterior. In later sections, we introduce more data types in joint inversions, allowing us to release *Vp* and density from the empirical scaling relations and allow them to vary independently.

The rjMCMC starts with a random initial structure. After a burn-in period during which the convergence is achieved, we save the accepted models into the ensemble solution, for a total of 5 million iterations. Due to the nature of the Markov Chain, each time we perturb the current model, only a small part of the proposed model is different from the current model. Therefore, consecutive models are highly correlated, even when the acceptance rate is optimal. To increase the independence of the model ensemble, we choose to save every 100th sampled model. We primarily rely on two approaches to assess the progress of the rjMCMC chain and to estimate the number of iterations needed to achieve convergence. First, we monitor how misfit evolves with iterations, making sure that it remains low. Second, we run several chains with different starting models and compare the statistical properties of the ensemble solutions obtained from each. For each chain's ensemble solution, we calculate the root-mean-square deviation (rmsd) to ensure that they are indistinguishable from one another after the burn-in period (see Figs B1 and B2).

The posterior model density plot is shown in Fig. 3. At every 1 km, we evaluate all the seismic velocities from the ensemble and normalize them to compute the PDF. The PDF is represented so that warm colours correspond to higher probability and cool colours indicate lower probability. We want to point out that in such PDF plots, the absolute value of the probability is a function of the bin size used in

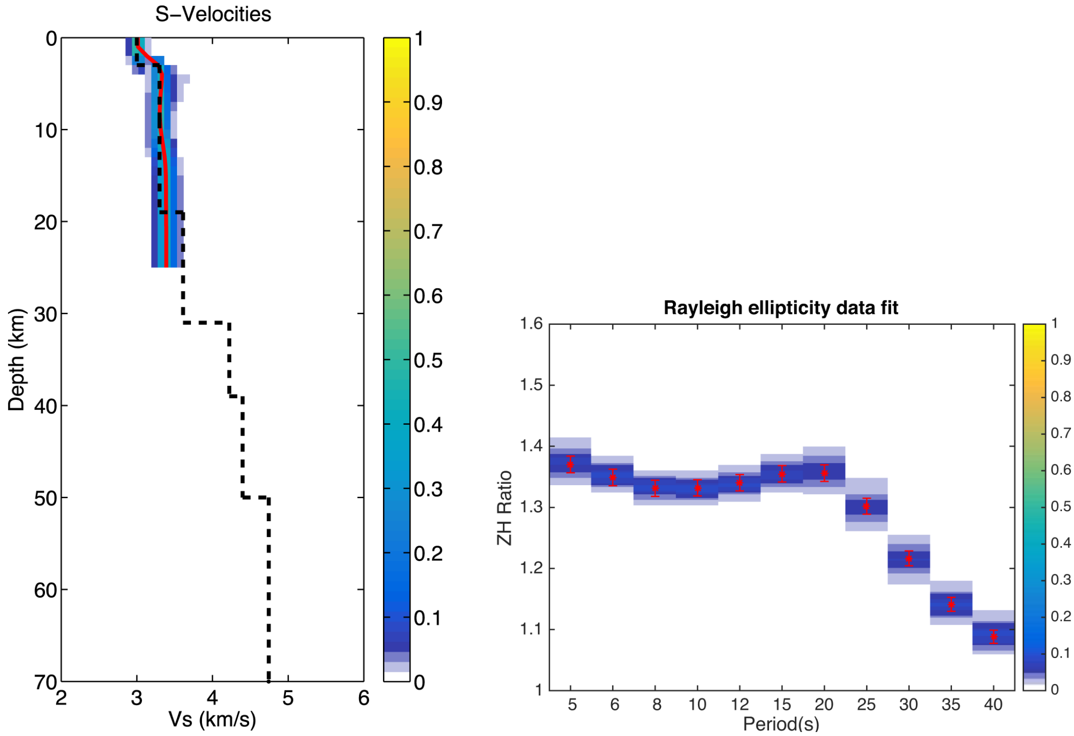

Figure 4. Left: *Vs* profile retrieved from transdimensional Bayesian inversion using synthetic ZH ratio data. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) is used to generate synthetic data. Right: data fit of ZH ratio inversion. The red stars with their associated uncertainties (1σ error bars) are the synthetic data used as an input of the TBI, while the colour tracks the probability density of the data predicted by the ensemble.

plotting. Therefore, in some cases, low absolute probability density does not necessarily represent poor resolution. The target model is plotted as a black dashed line for reference. It can be seen that we are able to resolve absolute *Vs* at different levels along the depth range allowed in the inversion. We find that *Vs* is constrained better at shallow compared to deeper depths in terms of posterior variance, mainly due to the sharper sensitivity kernel for shallow *Vs* structure from short-period SWD. It can also been seen that while the SWD is able to constrain the absolute value of *Vs*, it tends to retrieve the sharp velocity jump in the target model as a smooth transition. This is expected from the fact that SWD measurements depend on the integral of elastic properties across a range of depths. The uncertainties of *Vs* below 50 km seem to decrease with depth; we interpret this as a result of the fact that the period range of SWD we use here (up to 100 second) still has substantial sensitivity between 50 and 70 km and only one single layer is preferred by the TBI at this depth range.

3.2 TB inversion of ZH ratios

Having explored the ability of SWD constraining shear velocity structures, we turn our attention to ZH ratios, and perform a TB inversion of ZH ratios computed for the same synthetic model described in Section 2.4 (and shown in Fig. 1). Due to difficulty of reliably measuring ZH ratios at long periods (Ferreira & Woodhouse 2007), we restrict our attention to the 5–40 s period range, and assign 1 per cent uncorrelated uncertainty to the observation at each period. In reality, depending on whether standard deviation of the repeat measurements or the standard deviation of the mean of the repeated measurement is used, the measurement error for ZH ratios could be as large as 3–10 per cent (see Lin *et al.* 2012 and Lin *et al.* 2014); however, to illustrate the ability of ZH ratios to constrain elastic properties, we choose a relatively small value of uncertainty that might be achieved under ideal circumstances.

Our initial tests show that, compared to SWD, ZH ratios have limited potential for constraining elastic properties deeper than 20 km. When we invert for *Vs*, *Vp* and density using an attached-type parametrization, without assuming scaling relationship among parameters, the retrieved seismic structure shows large variations along depth and absolute *Vp*, *Vs* and density values are systematically biased at most depths. Even when we restrict the parameter space to the upper 25 km, the ensemble solutions show that we are unable to resolve the profiles of *Vs*, *Vp* and density simultaneously. Motivated by this finding, we invert only for *Vs*, and use empirical relations of Brocker (2005) to scale to *Vp* and density.

As with the SWD inversion, the total number of rjMCMC iterations is 5 million, of which the first 2.5 million are the burn-in period, in all ZH ratio individual inversions in this section. The convergence rate of ZH ratios individual inversion is similar to the SWD inversion. With different starting models, the Markov Chain is considered to have converged after about $\sim 2\ \times 10^5$ iterations. Therefore, we consider 2.5 million iteration to be a safe choice for burn-in period. As is shown in Fig. 4, the retrieved seismic structure is well constrained in terms

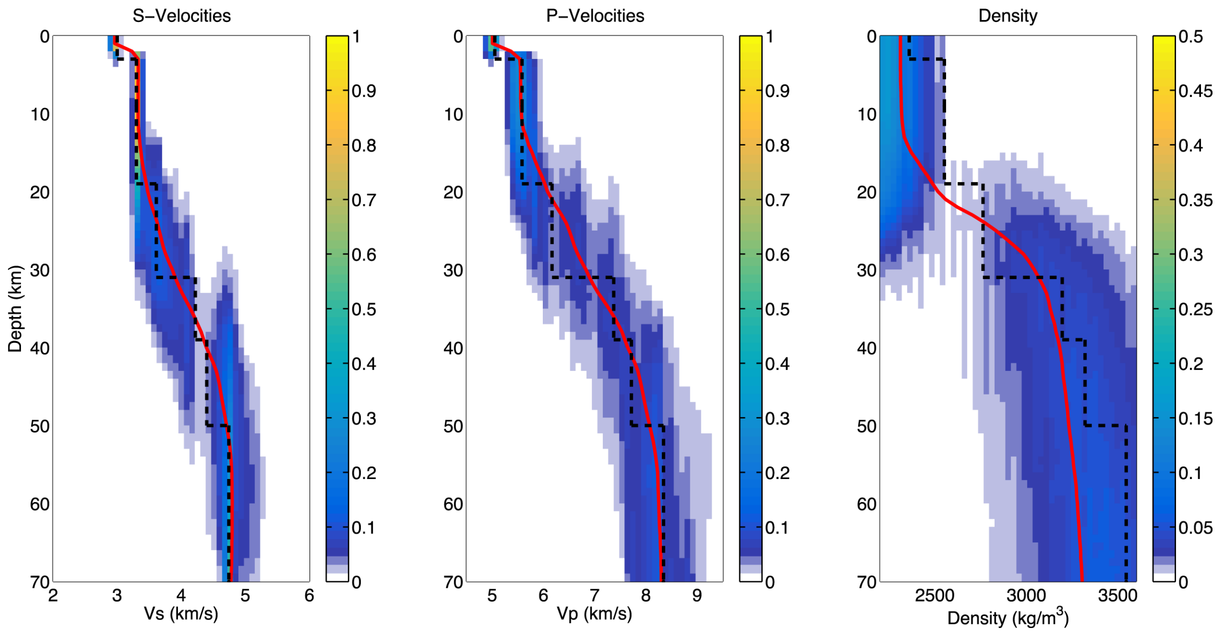

Figure 5. *Vs* (left), *Vp* (middle) and density (right) retrieved from transdimensional Bayesian joint inversion using synthetic SWD and ZH ratio data. In this test, all three types of parameters have independent parametrization. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dash lines) is used to generate synthetic data.

of *Vs*, and fit to data is excellent.

The TBI tests using ZH ratios are very informative. First, they show that compared to SWD, ZH ratios have limited ability to resolve structure below the crust. The shallow sensitivity of ZH ratios is well documented in the literature, with investigators usually using this data to constrain structure in the uppermost crust (e.g. Lin *et al.* 2012). Second, even though a strength of ZH ratios is their sensitivity to *Vp*, *Vs* and density, allowing a unconstrained inversion with all three types of parameters perturbed achieves very little resolution of structure due to nearly total trade-offs among parameters. When we reduce the number of parameters by fixing scaling relationships among them, we achieve better outcomes. Therefore, we should keep in mind that without additional constraints, it might not be practical to resolve an accurate and precise seismic structure from ZH ratios alone. These additional constraints could come either from the prior, such as by imposing scaling relationships among parameters appropriate for the geological setting of the inversion, or from the inclusion of other seismic observables to perform a joint inversion, such as SWD, which is the next topic we turn our attention to.

3.3 TB joint inversion of SWD and ZH ratios

In the previous sections, we showed that TB inversion is able to retrieve seismic structures with an adaptive parametrization using seismic observables one at a time. Here, we conduct a TB joint inversion by combining the Rayleigh wave dispersion, Love wave dispersion and ZH ratios. The motivation for doing joint inversion is to combine the strengths of different seismic data types to invert for a more comprehensive structure. The expected improvement in the ability to retrieve *Vp* and density structure also makes joint inversion a good example to illustrating the differences between the two types of parametrization proposed in this study.

We first invert for *Vp*, *Vs* and density using the independent-type parametrization. To better illustrate the data sensitivity, we do not impose any scaling factors between *Vs*, *Vp* and density. We want to point out that throughout this study, when *Vp* is inverted, it is actually parametrized as *Vp*/*Vs* ratio. Inverting *Vs* and *Vp*/*Vs* ratio is equivalent to inverting *Vs* and *Vp* if given the same prior. As is shown in Fig. 5, both *Vs* and *Vp* are well constrained above 20 km in terms of absolute value and variance compared to either of the individual inversion. When it comes to deeper structure, the variance of *Vp* increases significantly. Our calculation of normalized root-mean-square error (rmse) for *Vs* and *Vp* ensemble (Fig. C1) also shows that *Vs* is better constrained than *Vp* at most of the depths. The behaviour of the joint inversion is consistent with our expectations: SWD is able to constrain *Vs* in crust and upper mantle (with some sensitivity to *Vp*), while ZH ratios are able to constrain *Vp* and *Vs* in the crust. The preferred model from the density ensemble has two layers, while the *Vs* and *Vp* structures favour a five-layer model. This is because the much weaker data constraints on density yield density structures that are simpler than the actual target model or the retrieved *Vp* and *Vs* structures when inverted using an independent parametrization, in accordance with the lesser ability to resolve this parameter.

The attached-type parametrization is much more common in the seismic literature on the inversion of surface wave data than is the independent-type parametrization discussed above (e.g. Shapiro & Ritzwoller 2002; Yao *et al.* 2008; Chai *et al.* 2015; Shen & Ritzwoller 2016). In Fig. 6, we show the retrieved structures using an attached-type parametrization. The density ensemble shows a better fit to the

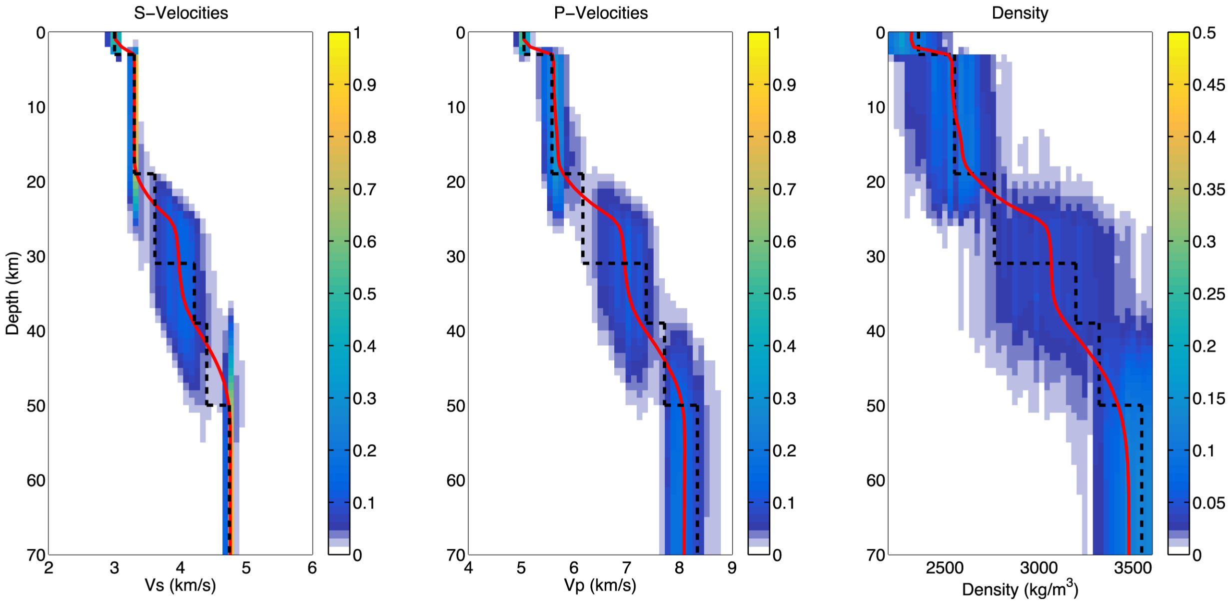

Figure 6. *Vs* (left), *Vp* (middle) and density (right) retrieved from transdimensional Bayesian joint inversion using synthetic SWD and ZH ratio data. In this test, all three types of parameters share the same geometry. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dash lines) is used to generate synthetic data.

true model, and contains more detailed structure. We should keep in mind that by using an attached-type parametrization, we impose a more informative prior that all types of parameters share the same geometry. When such prior information is valid, we could expect a better-resolved structure. This also suggests that prior assumptions about co-variance of parameters should be justified before being applied to actual inversion because their effects are significant, particularly for ill-resolved parameters like density. To illustrate the potential pitfalls of using attached-type parametrization, we show an example where synthetic data in computed for a structure in which the density does not share the same geometry with *Vs* and *Vp*. When we use the attached-type parametrization to perform a TB inversion of this data, the retrieved density structure exhibits artefacts that reflect the major features of *Vs* structure. This leads to a biased and misleading estimate of density (see Fig. D1).

4 DISCUSSION

4.1 Transdimensional versus fixed-parametrization inversion

In Section 3.1, we showed that TB inversion can recover a *Vs* profile from SWD measurements while treating the number of model parameters as an unknown. However, inversion of SWD data is most frequently done with a fixed parametrization (e.g. Hermann 2013). To gain insight into the relative advantages and disadvantages of a transdimensional inversion, we perform a Bayesian inversion with a fixed parametrization and compare our results to those obtained in Section 3.1. Using a starting model with the correct geometry, we only perturb the shear velocity during the MCMC. The retrieved ensemble structure is showed in Fig. 7. Like the transdimensional inversion, the fixed-parametrization inversion is able to recover the *Vs*. However, while both the absolute value and variance of *Vs* are well constrained at shallow depths (< 31 km), the variance at deeper depths increases significantly. We calculate the Kullback–Leibler divergence (KLD) for the posterior of both inversions with respect to their prior PDF. The KLD for discrete probability distributions is defined as:

$$D_{\mathrm{KL}}(P||Q) = \sum_i P(i)\ \log\frac{P(i)}{Q(i)} \tag{17}$$

The KLD from Q to P, denoted $\mathrm{D_{KL}}(P\|Q)$, can be interpreted as the information gained when one revises one's beliefs from the prior probability distribution Q to the posterior probability distribution P (Kullback & Leibler 1951). In Bayesian statistics, when P is the posterior and Q is the prior, KLD can be interpreted as a measure of the information gained from the data that transformed the prior distribution into the posterior distribution. The calculated KLD (Fig. 7, middle panel) shows that the fixed-parametrization inversion produced posterior solutions containing less information at deeper depths compared to the transdimensional inversion. A similar conclusion can also be drawn based on the comparison of rmse of these two tests (see Fig. 7, right-hand panel), which shows that, at deeper depths, TBI yields lower errors than the fixed parametrization inversion.

The apparent superiority of TBI may be counterintuitive, since the fixed-parametrization inversion imposed a stronger and perfectly accurate prior (since the parametrization was fixed to that of the target model). Generally, the more restrictive the prior is, the less uncertain

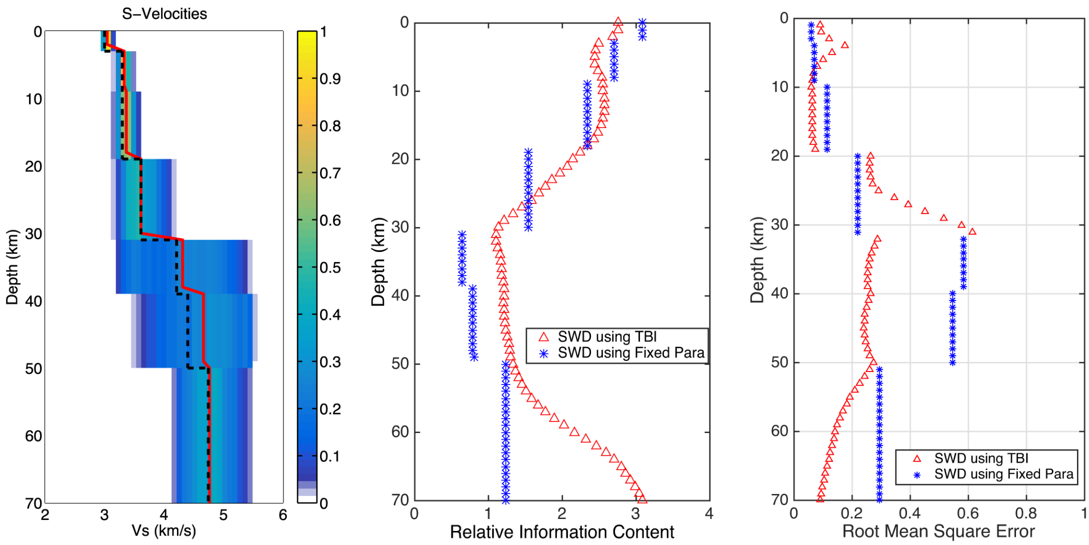

Figure 7. Left: *Vs* depth distributions retrieved using Bayesian inversion of synthetic surface wave dispersion data when parametrization is fixed to the same parametrization used to compute the synthetic dispersion data. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) is used to generate synthetic data. (middle) Kullback–Leibler divergence for TBI of surface wave dispersion (red triangle) and fixed-parametrization inversion of surface wave dispersion (blue star). Right: root-mean-square error for TBI of surface wave dispersion (red triangle) and fixed-parametrization inversion of surface wave dispersion (blue star)

the posterior should be. However, in this case, due to the trade-offs between shear velocities at different depths, the variance of the posterior is larger than the one using TBI. We stress that the relative advantage of TBI over traditional fixed-parametrization inversion would be even greater in the more realistic scenario using actual, rather than synthetic, SWD measurements, because in that scenario, the parametrization would not be known *a priori*. Moreover, the ensemble result from TBI allows statistical inferences of potential discontinuities along depths because of the less restrictive assumptions made on parametrization (Bodin *et al.* 2012).

4.2 Resolving gradual changing seismic structures using layered parametrizations

In our inversion for 1-D layered seismic structure, we assumed that elastic properties remain constant within each layer. While such assumption is widely made in SWD inversion, gradient-based model parametrization has also been employed in the literature (Gosselin *et al.* 2017). Since our inversions are inherently parsimonious, this assumption sets up an inconsistency between layered profiles preferred by the prior information (via parametrization), and gradual ones potentially preferred by realistic data. To test the implications of this assumption for resolving gradually changing velocity structures, we perform the following test. We compute synthetic Love and Rayleigh wave dispersion data in a target model with a 31 km thick crust, within which the *Vs* increases linearly from 3.0 to 3.6 km s^{-1}. The shear velocity jumps from 3.6 to 4.2 km s^{-1} at Moho, and then increases linearly to 4.74 km s^{-1} at the depth of 70 km. *Vp* and density are scaled to *Vs*, following the empirical relations from Brocker (2005) in both the target model and the later synthetic tests. We use both Rayleigh and Love wave dispersion data from 5 to 100 s with uncertainties of 3 per cent to invert for shear velocity, and obtain the *Vs* profiles shown in Fig. 8 (left-hand panel). We find that despite parametrizing the inversion with layers of constant properties, the ensemble solution partially resolves the gradually increasing velocity. The greatest exception occurs in the upper crust (above around 12 km), where the model ensemble shows a velocity jump overlaid by constant velocities. When SWD data between 2 and 5 s period are included, the model ensemble better resolves the gradually increasing velocity structure at the top of the crust (Fig. 8, right-hand panel). This test shows that the lack of data constraints in the uppermost crust is the main reason for the oversimplified structure. We can see a similar tendency of the inversion toward constant-velocity layers at the bottom of the model, where constraints from the data decrease once more. These tests suggest that throughout most of the crust, velocity gradients can be retrieved using TBI even with a layered parametrization; nevertheless, interpretations of gradients from inversions parametrized with constant-property layers should be cautious in areas where data constraints are less strong.

Aside from the period range of SWD, a large measurement error could be another cause for the limited data constraint, because greater measurement error degrades the amount of information contained in the data. Here, we test the effect of different measurement errors on the model complexity of the retrieved solution. We perform transdimensional inversion of 2–100 s Rayleigh and Love wave dispersion data with data measurement uncertainties ranging from 0.05 to 20 per cent. The target model used to generate synthetic data is the same gradually changing model shown in Fig. 8. For each test, we validate the convergence of the rjMCMC with the procedure described in Section 3.1, and then use the distribution of number of layers for the retrieved ensemble as a measure of model complexity. As is shown in Fig. 9, as data measurement uncertainty increases, the preferred number of layers for the retrieved ensemble decreases. Since the target model has a

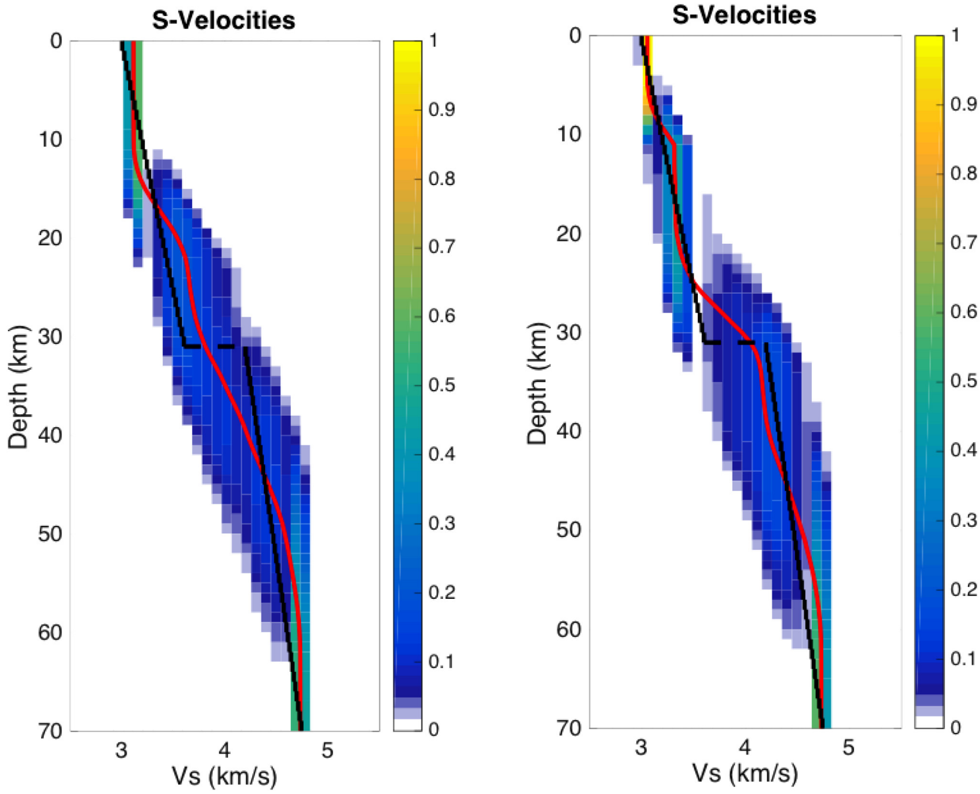

Figure 8. *Vs* depth distributions retrieved using transdimensional Bayesian inversion of synthetic surface wave dispersion data. The period range of SWD used is 5–100 s in the left-hand panel and 2–100 s in the right-hand panel. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) is used to generate synthetic data.

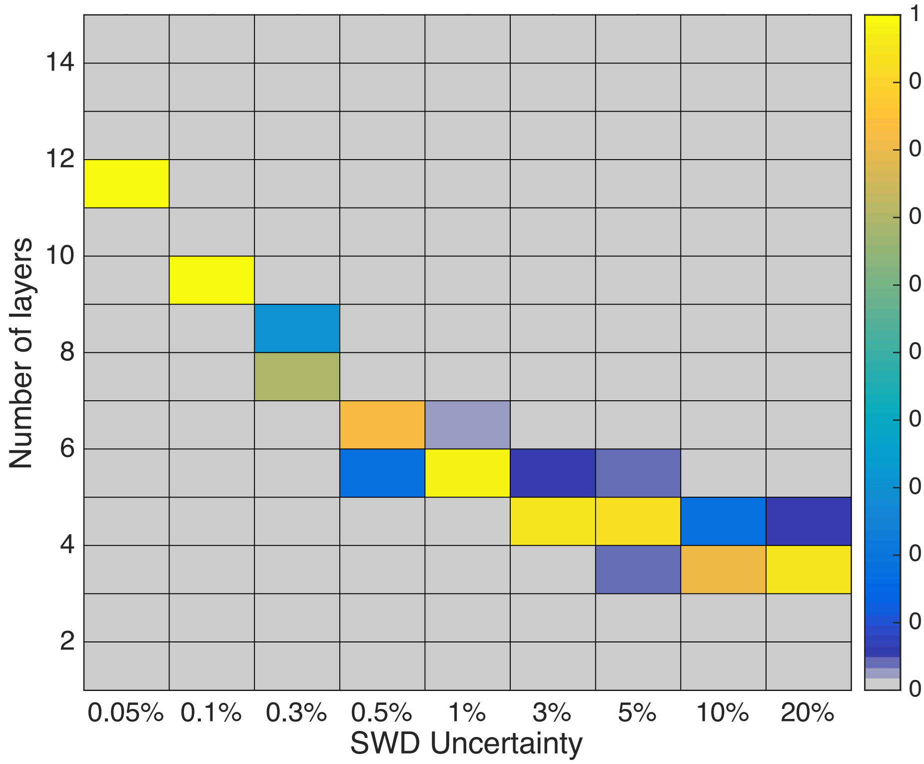

Figure 9. Posterior probability distribution of the number of layers in the retrieved ensemble for transdimensional inversion of surface wave dispersion measurements with different measurement uncertainties. The period range of both Rayleigh and Love wave dispersion used is 2–100 s.

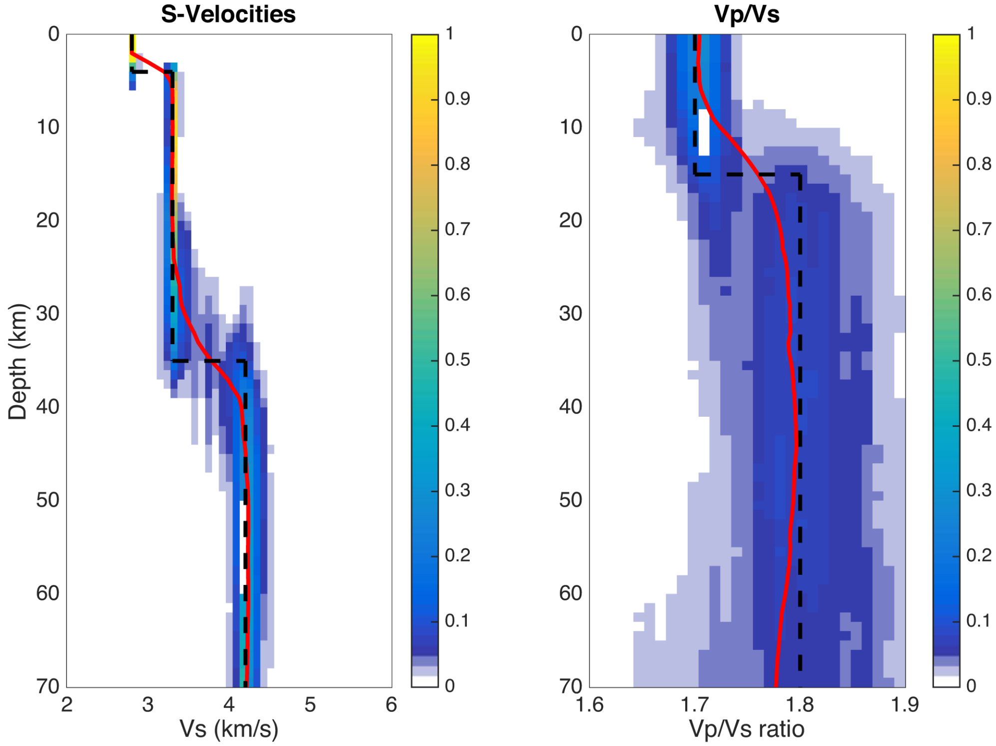

Figure 10. *Vs* (left) and *Vp*/*Vs* ratio (right) depth distribution retrieved using transdimensional Bayesian inversion of synthetic surface wave dispersion data. Independent parametrization is used for *Vs* and *Vp*/*Vs* ratio in this test. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) contains a jump in *Vp*/*Vs* due to the α–β quartz transition that is not accompanied by a change in *Vs* (Diaferia & Cammarano 2017).

gradually changing velocity with depth, a decrease in the preferred number of layers would eventually lead to an oversimplified structure that does not accurately reflect velocity gradients with depth.

4.3 Attached versus independent parametrizations

Despite the popularity of TB approaches, little discussion has concerned the parametrization of inversions involving multiple types of physical parameters. For model space search approaches—such as TB inversion—the number of model parameters is not limited by the number of measurements. This enables us to employ different parametrization schemes in the inversions, and quantify the effect of these choices on the posterior solution. Bodin *et al.* (2016) proposed a parametrization scheme for anisotropy inversion in which anisotropic parameters are proposed based on the existing isotropic structure. In this paper, we discussed two types of parametrization for dealing with multi-parameter problems: attached and independent.

In regions where the geotherm is sufficiently high, the α–β quartz transition is expected to occur in the middle–lower crust, resulting in a sharp *Vp*/*Vs* ratio increase that is not accompanied by a significant change in *Vs* (Kuo-Chen *et al.* 2012; Diaferia & Cammarano 2017). We show a synthetic test where the target model has a simplified one-layer crust on top of mantle *Vs* structure, and impose a mid-crustal *Vp*/*Vs* ratio increase from 1.7 to 1.8 representing the effects of the α–β quartz transition. This model is motivated by fig. 4 of Diaferia & Cammarano (2017). We perform TBI of SWD data (5–100 s range) using independent and attached parametrization of *Vs* and *Vp*/*Vs* ratio. The ensemble results show that when independent parametrization is applied (Fig. 10), the inversion resolves both the *Vs* and *Vp*/*Vs* structure accurately, despite the distinct geometries. However, when attached parametrization is applied (Fig. 11), the retrieved structure of *Vp*/*Vs* is strongly affected by the resolved geometry of Vs, to which the data are primarily sensitive.

This test illustrates how parametrization choices can be crucial to detecting complex structures with distinct geometries for different seismic parameters. Specifically, it shows that surface wave studies aiming to detect the α–β quartz transition in the middle–lower crust should employ a parametrization flexible enough to not suppress its detection.

4.4 Constraining radial anisotropy using TB inversion

In our previous synthetic tests, we assumed isotropic, layered structure, which may not always be an appropriate assumption to make, depending on the geological setting of the seismic station being analysed. Due to lattice-preferred orientation of anisotropic minerals or shape-preferred orientation of different rock layers or fractures, seismic wave velocities will depend on polarization and propagation directions (Crampin

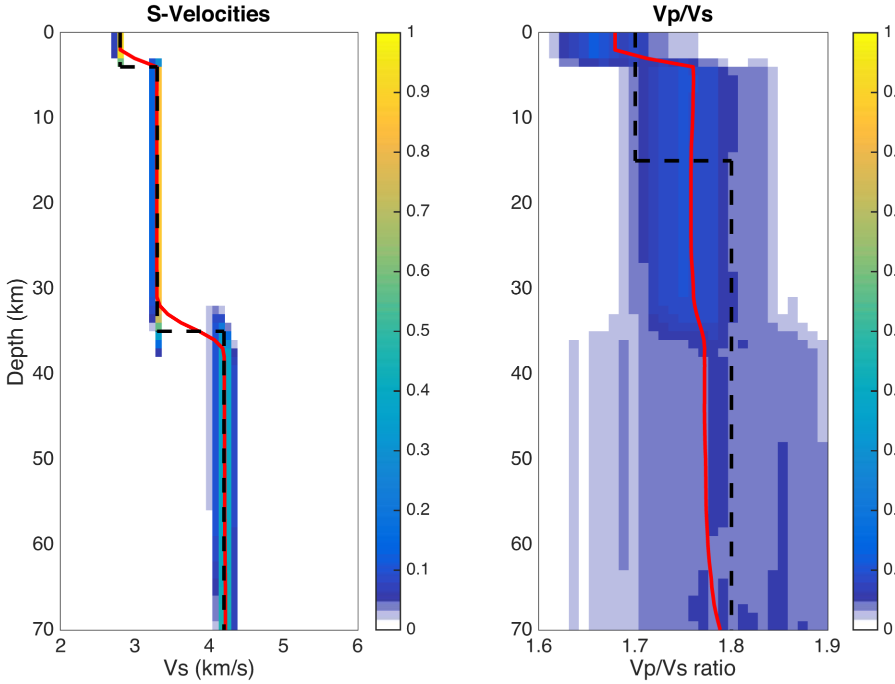

Figure 11. Vs (left) and Vp/Vs ratio (right) depth distribution retrieved using transdimensional Bayesian inversion of synthetic surface wave dispersion data. Attached parametrization is used for Vs and Vp/Vs ratio in this test. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) contains a jump in Vp/Vs due to the α–β quartz transition that is not accompanied by a change in Vs (Diaferia & Cammarano 2017).

et al. 1984). Studying seismic anisotropy in the crust and upper mantle can provide us insights into crust and mantle deformation (Kendall 2000; Becker *et al.* 2003), mantle composition (Montagner & Anderson 1989), lithosphere and asthenosphere coupling (Silver & Holt 2002; Becker 2006), and the net rotation of the lithosphere (Becker 2008). Here, we restrict our attention to radial anisotropy, in which the elastic properties of the medium can be described by five independent elastic coefficients (A, C, F, L and N; Love 1911) at each location. Seismic observations including surface waves and normal mode data are often used to constrain the radial anisotropy of the Earth (e.g. Ekstrom & Dziewonski 1998; Lekic & Romanowicz 2011; Chang *et al.* 2014; Moulik & Ekstrom 2014). Radial anisotropy in the Earth is often due to layering. However, recently studies also suggest that a large portion of anisotropy presented in the tomographic models may be due to unmapped discontinuities (Bodin *et al.* 2015).

To study shear wave radial anisotropy, we use the Voigt average isotropic shear wave velocity Vs and radial anisotropy parameter $\xi = (Vsh/Vsv)^2$ instead of isotropic shear velocity alone to represent the seismic structure. With limited studies discussing the relationship between isotropic Vs and ξ geometries, we propose to use independent parametrization to represent such ignorance.

For the synthetic test, we set the vertical shear wave velocity Vsv for the target model to be same as the Vs value from the isotropic target model we used in previous sections. We set the horizontal shear wave velocity Vsh to be different from the Vsv structure so that the radial anisotropy parameter $\xi = (Vsh/Vsv)^2$ has a value of 1.149 between 19 and 50 km and a value of 1.000 at other depths (Fig. 12, black dashed line). Vp and density in thebr target model are derived from isotropic Vs using the empirical relationship from Brocker (2005). To systematically investigate the effects of parameter trade-offs and data uncertainties on the retrieved structures, we perform a series of synthetic tests with different combinations of parameter types and data uncertainties.

We start by inverting SWD data, assuming a 2 per cent measurement error uncorrelated between periods. Here, we invert for Vs, ξ and Vp without assuming scaling relationships between any of them. In order to study the effect of trade-offs between different model parameters in a systemic way, we scale density to Vs, using expressions in Brocker (2005). The prior on ξ is set to be a uniform distribution between 0.81 and 1.21. The retrieved structure is shown in Fig. 12. Compared to the velocity structures, the ensemble of ξ spreads widely across the prior space. Despite the large variance, the inversion is able to resolve an anisotropic layer between 19 and 50 km. In Fig. 12, we plot the trimmed mean and the mode of the posterior PDF to better illustrate the inversion result. Both the trimmed mean value and the mode are overestimated between 10 and 19 km, as can be seen by comparing them against the target structure shown in black.

In Fig. 13, we show the scatter plots coloured by their density from the posterior PDF of Vs, Vp and ξ at three different depths. The scatter plots between Vp and ξ show that there are trade-offs between these two parameters. These trade-offs result from the fact that in a radially anisotropic medium, Rayleigh waves, whose sensitivity is primarily to Vsv, are also sensitive to Vp, while Love waves, whose primary sensitivity is to Vsh, are not sensitive to Vp. This trade-off between Vp and ξ limits our ability to retrieve radial anisotropy given the limited constraints provided by SWD data. Similarly, we test the trade-off between density and ξ by performing an inversion for Vs, ξ and density without assuming a fixed scaling relationship between them, while fixing the Vp scaling to Vs, using the expressions in Brocker (2005).

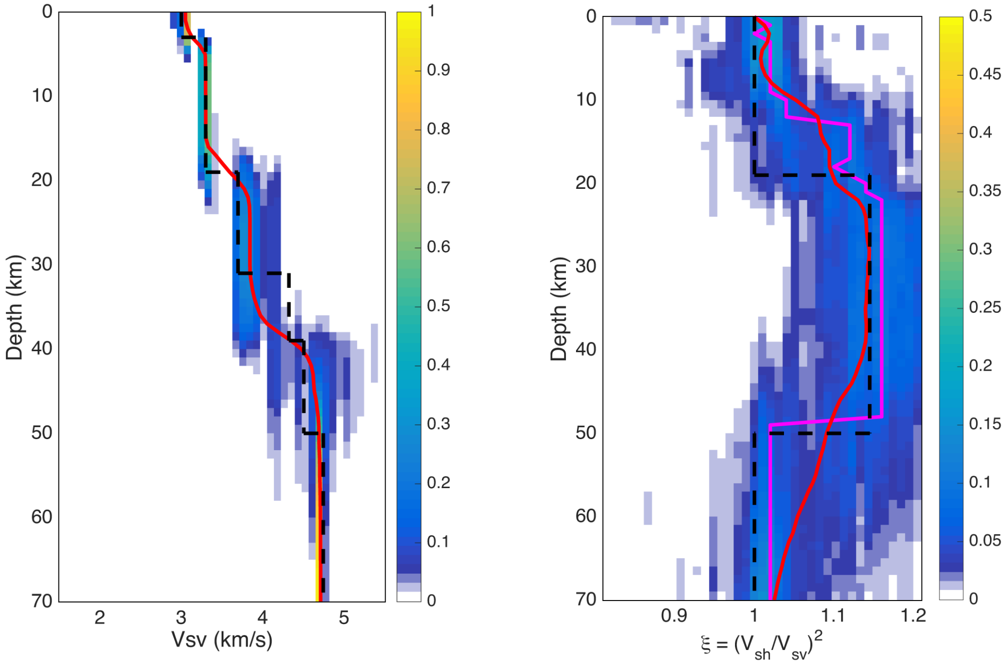

Figure 12. Anisotropy inversion with independent parametrization and no scaling on Vp. The ensemble solutions (Vs on the left and ξ on the right) are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, the solid red line denoting the 5 per cent trimmed mean of the posterior and the magenta line denoting its mode. The target model (black dashed lines) is used to generate synthetic data.

The target model is same as the one in Fig. 12. The retrieved structure of ξ is shown in Fig. 14 (left). While the ensemble results resolve an anisotropic structure approximately between 19 and 50 km, both the trimmed mean and the mode are underestimated within this depths range. Moreover, the thickness of the anisotropic layer is not well constrained. We interpret this as a result of trade-off between density and ξ limiting our ability of resolving radial anisotropy giving the limited constraints provided by SWD data.

To validate the effect of the trade-off on the inference of ξ, we apply the same empirical relationship from Brocker (2005) used in the target model to derive both Vp and density from Vs in our synthetic test to reduce the trade-off between Vp, density and ξ. We keep the rest of the inversion set-up the same as in the previous two tests. The retrieved ξ structure is shown in Fig. 14 (right). Compared to the tests with no scaling applied to Vp or density, the radial anisotropy structure is better constrained at several depths. The estimated ξ from the mode of posterior PDF is closer to the true value between 24 and 48 km. The mode of the ensemble in this test also better tracks the thickness of the anisotropic layer. We plot the marginal posterior PDF in Fig. 15 at several depths to better illustrate the retrieved ξ against the actual structure. In Fig. 15, we also show the coloured scatter plot between Vp and ξ, which illustrated the reduced Vp– ξ trade-off compared to the results shown in Fig. 13.

The three tests together show that the trade-offs between Vp and ξ as well as between density and ξ affect the inversion result for ξ. This means that without introducing additional data, applying a proper scaling between Vp, density and ξ helps resolve the radial anisotropy; however, on the other hand, Vp–Vs–density scaling assumptions need to be carefully made since an unsuitable scaling could bias the estimate of ξ. As a comparison, in Fig. 16, we show the retrieved anisotropy structure where we apply a constant Vp/Vs ratio of 1.68 as the scaling during the inversion. As expected, the incorrect scaling biases our estimate of Vp. The retrieved anisotropy structure also deviates from the true value at several depths. We estimate an anisotropic layer between 24 and 48 km for the correct scaling case, while for the fixed Vp/Vs case, the anisotropic layer is only recovered between 23 and 38 km. In Fig. 16, we calculated the rmse of the ensembles for the three radial anisotropy tests against the true value from the input model. It can be seen that when assumptions about Vp–Vs scaling are incorrect, the disagreement between the posterior and the true value is the largest between 20 and 50 km, which covers most of the anisotropic layer. The rmse in the incorrect scaling case can be as large as 0.16, while the largest rmse in the correct scaling is 0.12.

Roy & Romanowicz (2017) investigated the effect of assuming a fixed Vp/Vs on the inversion of SWD and converted body wave (P-to-S) data for Vs radial anisotropy. By comparing results obtained fixing Vp/Vs to those obtained treating Vp/Vs as unknown, they concluded that the slight difference in the choice of Vp/Vs would not affect the retrieved structure. While the inclusion of body wave data does not provide direct constraint on radial anisotropy, it is expected to improve the constraints on the depths and impedance contrasts across discontinuities in the velocity structures, and indirectly lead to better resolution of radial anisotropy. Therefore, the work of Roy & Romanowicz (2017) complements the analysis presented here, in which we explore the effects of scaling assumptions on inversion of SWD alone, without including observables such as receiver function. Furthermore, unlike this study, Roy & Romanowicz (2017) assume the geometry of variations in isotropic Vs and radial anisotropy to be the same. As is discussed in Section 3.3, when such assumption is justified, it helps resolve the posterior; our study provides a different, complementary perspective on how different parametrization choices could affect the resolution of radial anisotropy.

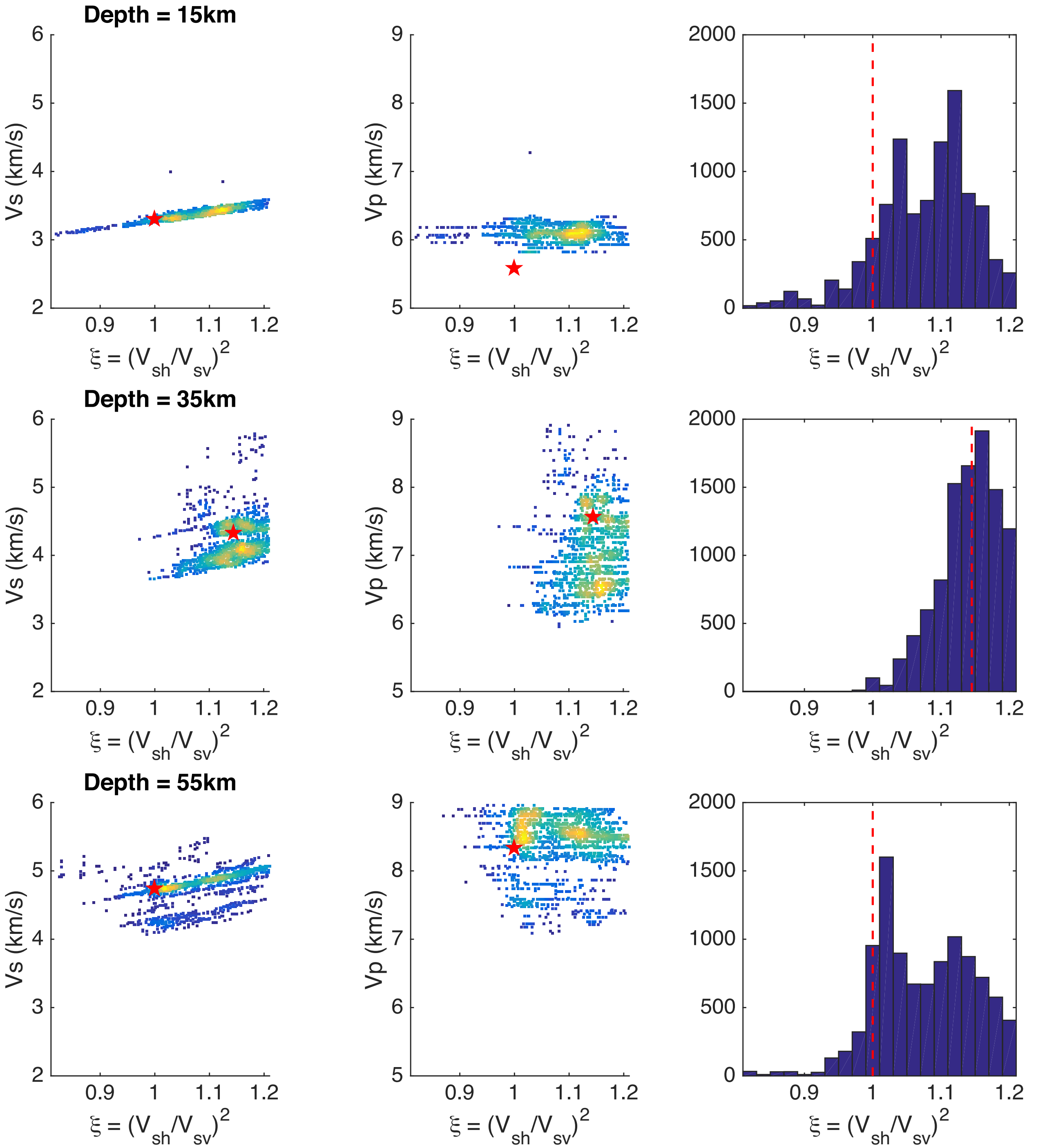

Figure 13. Parameter trade-offs and marginal posterior PDF at 15 km (upper), 35 km (middle) and 55 km (bottom) from the radial anisotropy inversion with no scaling relation assumed between Vs and Vp. Left-hand panels show the scatter plot of Vs and ξ values from the 10 000 models in the ensemble solution. The scatter plot is coloured by the density of points to better reveal the parameter trade-off. Warm coluors denote higher probabilities and cool colours denote lower probabilities. Middle panels show coloured scatter plots but for Vp and ξ values in the ensemble. The right-hand panels are the marginal posterior PDF of ξ. The true value at that depth is plotted as red dashed line.

5 CONCLUSION

TB inversion has recently gained increasing attention in the area of geophysics. Its applications to various topics including seismic tomography (Young *et al.* 2013; Calo *et al.* 2016; Petrescu *et al.* 2016; Burdick & Lekic 2017; Olugboji *et al.* 2017), earthquake source inversion (Dettmer *et al.* 2014), receiver function estimation (Kolb & Lekic 2014), controlled source exploration geophysics (Ray *et al.* 2014; Gehrmann *et al.* 2015), geoacoustic inversion (Dettmer *et al.* 2010) and viscosity inversion (Rudolph *et al.* 2015), show the utility of flexible yet naturally

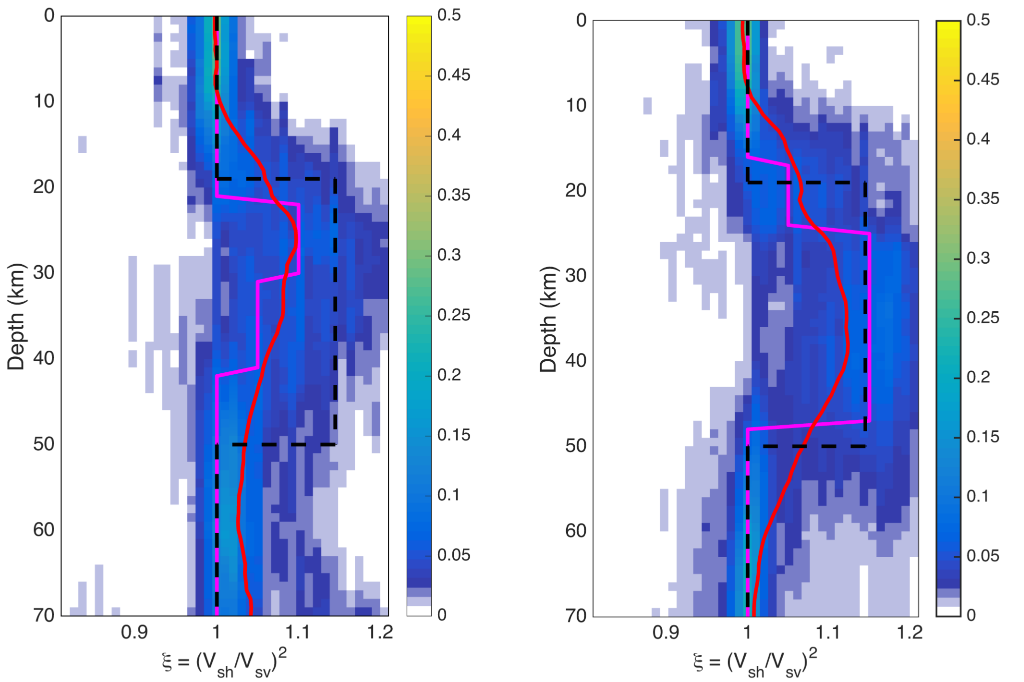

Figure 14. Anisotropy inversions with independent parametrization and no scaling on density (left)/correct scaling on *Vp* and density (right). The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, the solid red line denoting the 5 per cent trimmed mean of the posterior and the magenta line denoting its mode. The target model (black dashed lines) is used to generate synthetic data.

parsimonious parametrization in geophysical inversion as well as the capability of uncertainty quantification of inversion results that better represent the data sensitivity.

In this paper, we first show that we are able to retrieve seismic structures using SWD and Rayleigh wave ZH ratio individually and together with less restrictive assumptions. Our synthetic tests suggest that although SWD and the ZH ratios are sensitive to *Vp*, *Vs* and density to various degrees, neither data set is individually able to resolve a comprehensive structure. However, because of the flexibility of the transdimensional inversion, we can easily combine the ZH ratio data with SWD in a joint inversion. We show that TB inversion can take advantage of the complementary sensitivity of the two data types to simultaneously constrain the continental lithosphere *Vs* structure, as well as the crustal *Vp* structure.

By comparing the results from the transdimensional inversion and inversion with fixed but correct parametrization, we illustrate that a fixed parametrization with strong prior information could bias the estimate of model parameters. One might argue that given a fixed but correct parametrization, the inversion results should represent the true uncertainties of the model parameters. However, the estimates of model parameters include not just the value of elastic parameters but also their distribution along depth (i.e. layering). A fixed parametrization is equivalent to using a prior that assumes no uncertainty for the latter, which affects the estimates of the former due to model parameter trade-offs.

While a 1-D layered model with constant elastic value within each layer is assumed in our TB inversion, our synthetic tests suggest that given robust data constraints, this layered structure parametrization is still able to resolve structures in which elastic properties change gradually with depth. We show that the limited SWD resolution and large measurement error could both result in gradient structures appearing oversimplified when using parametrizations based on constant-velocity layers.

We then explore the effects of choices of parametrization on the retrieval of isotropic and anisotropic structure from surface wave inversion. Specifically, we propose and contrast two distinct parametrization choices: attached, in which all parameters of interest share the same geometry; and, independent, in which the geometry of different physical parameters can vary.

Using synthetic tests, we show that the attached-type scheme tends to yield results whose geometry is mainly determined by the parameter that is best constrained by the data at hand. When other parameters share the same geometry with the best-resolved parameter, the use of attached-type parametrization is advised (see Fig. 7). On the other hand, when parameters do not share the same geometry, the estimate of less well constrained could be biased due to trade-offs (see Fig. D1). The attached-type parametrization we discuss in this study is similar to the scheme proposed by Bodin *et al.* (2016), since parameters of different type share the same geometry. It differs from the Bodin *et al.* scheme, in that since the anisotropic structure is sampled by adding/removing anisotropic parameters from an existing layer, our attached-type parametrization assumes uniform prior on the additional parameters and samples them together.

The independent-type parametrization we introduce offers a more flexible parametrization containing less prior information. By assuming no correlation between the geometry of different parameters, we are able to detect potentially complex structures with distinct geometries of

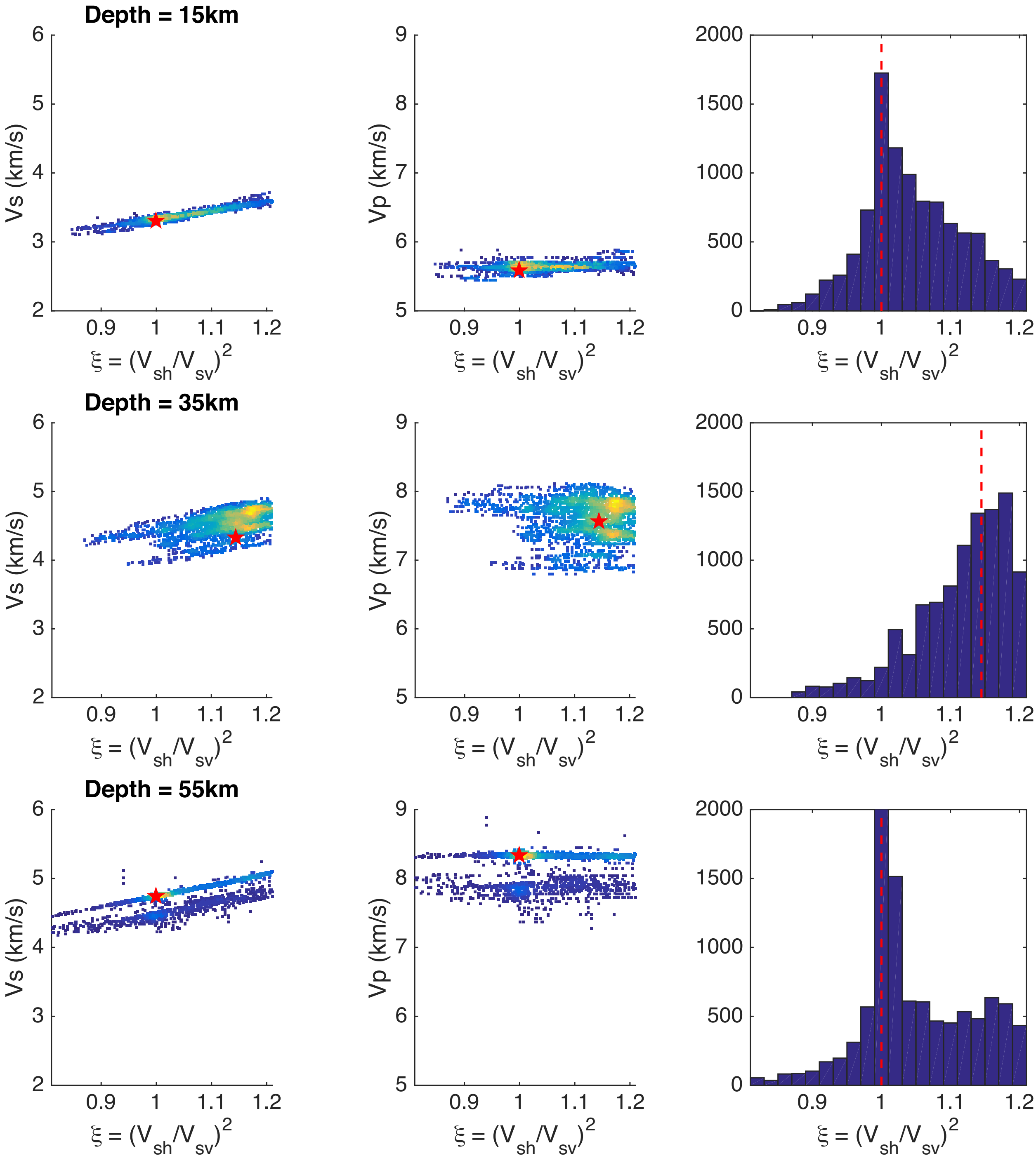

Figure 15. Parameter trade-offs and marginal posterior PDF from the radial anisotropy inversion with correct scaling relation between *Vp* and *Vs* at 15 km (upper), 35 km (middle) and 55 km (bottom). Left-hand panels show the scatter plot of *Vs* and ξ values from the 10 000 models in the ensemble solution. Middle panels are the coloured scatter plots but for *Vp* and ξ. The right-hand panels are the marginal posterior PDF of ξ. The true value at that depth is plotted as red dashed line.

different parameters while using an optimally parsimonious number of parameters. This situation would accompany the α–β quartz transition as shown in Fig. 10 and discussed in Section 4.3. Additionally, it would be expected in the presence of a layer of partial melt where *Vs* drops dramatically but *Vp* does not, which might be associated with volcanic regions, regions with elevated temperatures in the lower crust, or even glacial firn aquifers recently seismically characterized in Greenland (Montgomery *et al.* 2017).

Seismic Bayesian inversion has been used to investigate the radial anisotropy of the Earth (Shapiro & Ritzwoller 2002; Beghein & Trampert 2004; Beghein *et al.* 2014; Calo *et al.* 2016). In particular, Calo *et al.* (2016) applied a transdimensional inversion using SWD as well as receiver function data, and relied on empirical scaling laws between *Vs* and *Vp* to reduce parameter trade-offs. Here, we show

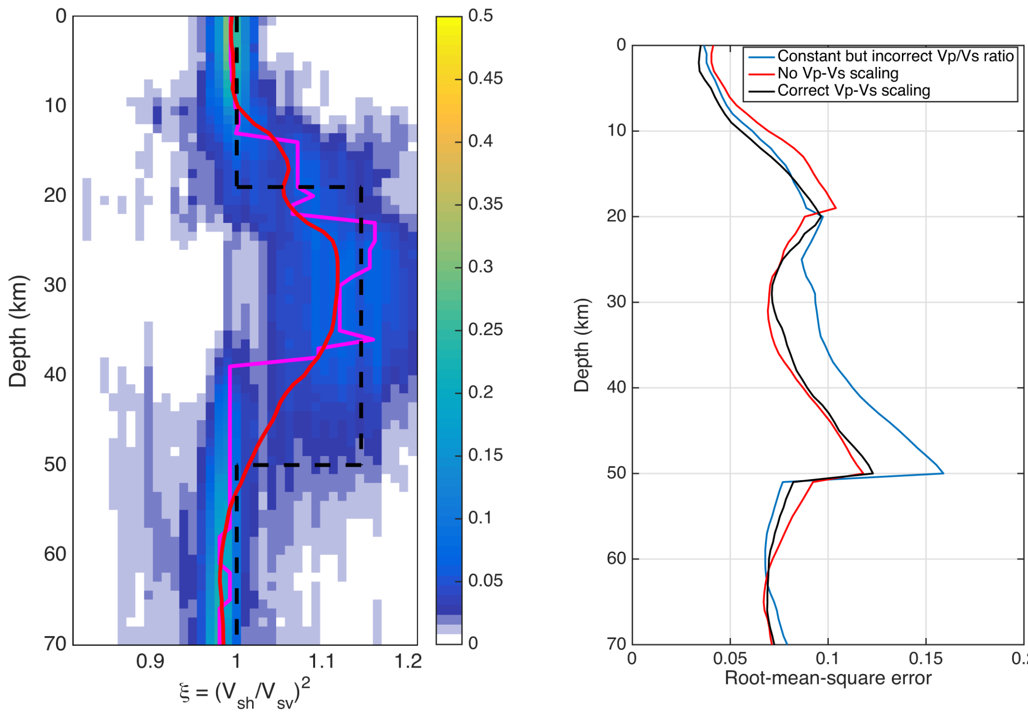

Figure 16. Left: anisotropy inversion with independent parametrization and a fixed *Vp*/*Vs* ratio of 1.68. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, the solid red line denoting the 5 per cent trimmed mean of the posterior and the magenta line denoting its mode. The target model (black dashed lines) is used to generate synthetic data. Right: root-mean-square error of radial anisotropy inversion ensemble against the input model. The blue curve is the TB inversion with fixed $Vp/Vs = 1.68$ using Rayleigh and Love wave dispersions. The red curve is the same inversion except using no *Vp*–*Vs* scaling. The black curve uses the correct *Vp*–*Vs* scaling as the input model, while the rest of the setting is same as the previous two.

that inferences of radial anisotropy from SWD are affected by prior information imposed on the inversion process. Specifically, applying proper scaling relations between *Vp*, *Vs* and density helps improve the constraints on radial anisotropy, while inaccurate assumptions about *Vp*–*Vs*–density scaling can bias estimates of radial anisotropy.

Previous studies have highlighted the potential of unmodelled crustal structure to bias inferences of upper-mantle radial anisotropy (e.g. Bozdag & Trampert 2008; Ferreira *et al.* 2010; Lekic *et al.* 2010), but the effect of *Vp*–*Vs*–density scaling assumptions has not garnered equal attention. Therefore, we stress that careful choices must be made to prevent the estimates of radial anisotropy from being biased due to unmodelled Vp and density structure. We find that trade-offs between *Vp* and radial anisotropy can increase rmse by 33 per cent in estimates of radial anisotropy.

Our inversion for radial anisotropy adopts an independent parametrization in which the geometries of isotropic *Vs* and ξ are not assumed to be the same. With fewer assumptions made to avoid potential bias in the inversion results, we are able to resolve the main anisotropic feature in the synthetic model. The independent-type parametrization scheme is particularly appropriate since radial anisotropy need not share the same geometry as isotropic wave speeds, as pointed out by Montagner (2002).

In this study, we considered three sources of uncertainty: limitation of data sensitivity; assumed scaling among parameters; and the choice of parametrization, including both the number of parameters and attached vesuss independent parametrizations for multiparameter problems. The discussion of different sources of uncertainties presented herein should help inform choices for inversions on surface wave measurements on their own and in combination with other, complementary data types (e.g. receiver functions). While better constraints on seismic velocity profiles are expected when combining multiple seismic observables, it is necessary to attribute the influence of certain parametrization choices. This becomes particularly important for inferences of parameters such as radial anisotropy and density—which are less well constrained by available data—since our findings suggests that model parametrization can significantly bias them.

ACKNOWLEDGEMENTS

5 Acknowledgments

The authors thank Scott Burdick and Michael N. Evans for suggestions and two anonymous reviews that have improved this manuscript. The authors also thank Glenn J. Rix and Carlo G. Lai for making their Rayleigh wave dispersion code open to public. VL acknowledges support from NSF grants EAR1352214 and EAR1361325, as well as from the Packard Fellowship for Science and Technology.

REFERENCE

Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181**(2), 858–872.

Aki, K., 1968. Seismological evidences for the existence of soft thin layers in the upper mantle under Japan, *J. geophys. Res.,* **73**(2), 585–594.

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology,* 2nd edn, University Science Books, Sausalito, CA.

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

Aurenhammer, F., 1991. Voronoi diagrams—a survey of a fundamental geometric data structure, *ACM Comput. Surv.,* **23**(3), 345–405.

Backus, G.E. & Gilbert, J.F., 1967. Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. Int.,* **13**(1–3), 247–276.

Bayes, M., & Price, M., 1763. An essay towards solving a problem in the doctrine of chances. by the late Rev. Mr. Bayes, F.R.S. communicated by Mr. price, in a letter to John Canton, A.M.F.R.S., *Phil. Trans. (1683-1775),* **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**(2), 696–714.

Becker, T.W., 2006. On the effect of temperature and strain-rate dependent viscosity on global mantle flow, net rotation, and plate-driving forces, *Geophys. J. Int.,* **167**(2), 943–957.

Becker, T.W., 2008. Azimuthal seismic anisotropy constrains net rotation of the lithosphere, *Geophys. Res. Lett.,* **35**(5), L05303.

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**(1), 151–162.

Beghein, C., Yuan, K., Schmerr, N. & Xing, Z., 2014. Changes in seismic anisotropy shed light on the nature of the Gutenberg discontinuity, *Science,* **343**(6176), 1237–1240.

Bhattacharya, S.N., 1996. Earth-flattening transformation for P-SV waves, *Bull. seism. Soc. Am.,* **86**(6), 1979–1982.

Bhattacharya, S.N. & Arora, S., 1997. A flattening transformation for P-SV waves in a transversely isotropic Earth, *Bull. seism. Soc. Am.,* **87**(5), 1297–1304.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178**(3), 1411–1436.

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.,* **117**(B2), B02301.

Bodin, T., Capdeville, Y., Romanowicz, B. & Montagner, J.P., 2015. Interpreting radial anisotropy in global and regional tomographic models, In *The Earth's Heterogeneous Mantle,* pp. 105–144. Springer International Publishing.

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**(1), 605–629.

Boore, D.M. & Toksöz, M.N., 1969. Rayleigh wave particle motion and crustal structure, *Bull. seism. Soc. Am.,* **59**(1), 331–346.

Boschi, L. & Ekström, G., 2002. New images of the Earth's upper mantle from measurements of surface wave phase velocity anomalies, *J. geophys. Res.,* **107**(B4), 2059.

Bozdağ, E. & Trampert, J., 2008. On crustal corrections in surface wave tomography, *Geophys. J. Int.,* **172**(3), 1066–1082.

Brocker, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. seism. Soc. Am.,* **95**(6), 2081–2092.

Burdick, S. & Lekić, V., 2017. Velocity variations and uncertainty from transdimensional P-wave tomography of North America, *Geophys. J. Int.,* **209**(2), 1337–1351.

Calò, M., Bodin, T. & Romanowicz, B., 2016. Layered structure in the upper mantle across North America from joint inversion of long and short period seismic data, *Earth planet. Sci. Lett.,* **449,** 164–175.

Chai, C., Ammon, C.J., Maceira, M. & Herrmann, R.B., 2015. Inverting interpolated receiver functions with surface wave dispersion and gravity: application to the western US and adjacent Canada and Mexico, *Geophys. Res. Lett.,* **42**(11), 4359–4366.

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

Chang, S.J., Ferreira, A.M., Ritsema, J., Heijst, H.J. & Woodhouse, J.H., 2015. Joint inversion for global isotropic and radially anisotropic mantle structure including crustal thickness perturbations, *J. geophys. Res.,* **120**(6), 4278–4300.

Chen, M. & Tromp, J., 2007. Theoretical and numerical investigations of global and regional seismic wave propagation in weakly anisotropic earth models, *Geophys. J. Int.,* **168**(3), 1130–1152.

Constable, S.C., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: a practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics,* **52**(3), 289–300.

Crampin, S., Chesnokov, E.M. & Hipkin, R.G., 1984. Seismic anisotropy—the state of the art: II, *Geophys. J. Int.,* **76**(1), 1–16.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128**(6), 3393–3405.

Dettmer, J., Benavente, R., Cummins, P.R. & Sambridge, M., 2014. Trans-dimensional finite-fault inversion, *Geophys. J. Int.,* **199**(2), 735–751.

Diaferia, G. & Cammarano, F., 2017. Seismic signature of the continental crust: what thermodynamics says. An example from the Italian Peninsula, *Tectonics,* **36,** 3192–3208.

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25**(4), 297–356.

Ekström, G., Tromp, J. & Larson, E.W., 1997. Measurements and global models of surface wave propagation, *J. geophys. Res.,* **102**(B4), 8137–8157.

Ekström, G. & Dziewonski, A.M., 1998. The unique anisotropy of the Pacific upper mantle, *Nature,* **394**(6689), 168–172.

Ekström, G., 2011. A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25–250 s, *Geophys. J. Int.,* **187**(3), 1668–1686.

Ekström, G., 2014. Love and rayleigh phase-velocity maps, 5–40 s, of the western and central USA from USArray data, *Earth planet. Sci. Lett.,* **402,** 42–49.

Ferreira, A.M. & Woodhouse, J.H., 2007. Observations of long period Rayleigh wave ellipticity, *Geophys. J. Int.,* **169**(1), 161–169.

Ferreira, A.M.G., Woodhouse, J.H., Visser, K. & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *J. geophys. Res.,* **115**(B4), B04313.

Franklin, J.N., 1970. Well-posed stochastic extensions of ill-posed linear problems, *J. Math. Anal. Appl.,* **31**(3), 682–716.

Gehrmann, R.A., Dettmer, J., Schwalenberg, K., Engels, M., Dosso, S.E. & Özmaral, A., 2015. Trans-dimensional Bayesian inversion of controlled-source electromagnetic data in the German North Sea, *Geophys. Prospect.,* **63**(6), 1314–1333.

Green, P.J., 2003. Trans-dimensional markov chain monte carlo, in *Highly Structured Stochastic Systems ,* pp. 179–198. Oxford Univ. Press, Oxford.

Gosselin, J.M., Dosso, S.E., Cassidy, J.F., Quijano, J.E., Molnar, S. & Dettmer, J., 2017. A gradient-based model parametrization using Bernstein polynomials in Bayesian inversion of surface wave dispersion, *Geophys. J. Int.,* **211**(1), 528–540.

Harkrider, D.G. & Anderson, D.L., 1962. Computation of surface wave dispersion for multilayered anisotropic media, *Bull. seism. Soc. Am.,* **52**(2), 321–332.

Herrmann, R.B., 2013. Computer programs in seismology: An evolving tool for instruction and research, *Seismol. Res. Lett.,* **84**(6), 1081–1088.

Hisada, Y., 1994. An efficient method for computing Green's functions for a layered half-space with sources and receivers at close depths, *Bull. seism. Soc. Am.,* **84**(5), 1456–1472.

Huang, Z., Su, W., Peng, Y., Zheng, Y. & Li, H., 2003. Rayleigh wave tomography of China and adjacent regions, *J. geophys. Res.,* **108**(B2), 2073.

Kendall, J.M., 2000. Seismic anisotropy in the boundary layers of the mantle, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale,* pp. 133–159. American Geophysical Union, Washington DC.

Kolb, J.M. & Lekić, V., 2014. Receiver function deconvolution using trans-dimensional hierarchical Bayesian inference, *Geophys. J. Int.,* **197**(3), 1719–1735.

Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Anna. Math. Stat.,* **22**(1), 79–86.

Kuo-Chen, H., Wu, F.T., Jenkins, D.M., Mechie, J., Roecker, S.W., Wang, C.Y. & Huang, B.S., 2012. Seismic evidence for the α-β quartz transition beneath Taiwan from Vp/Vs tomography, *Geophys. Res. Lett.,* **39**(22), L22302, doi:10.1029/2012GL053649.

Kustowski, B., Ekström, G. & Dziewoński, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res. Earth,* **113**(B6), B06306.

Lai, C.G. & Rix, G.J., 1998. *Simultaneous Inversion of Rayleigh Phase Velocity and Attenuation for Near-Surface Site Characterization,* p. 258, School of Civil and Environmental Engineering, Georgia Institute of Technology, Georgia.

Laske, G. & Masters, G., 1996. Constraints on global phase velocity maps from long-period polarization data, *J. geophys. Res.,* **101**(B7), 16059–16075.

Lekić, V., Panning, M. & Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography, *Geophys. J. Int.,* **182**(1), 265–278.

Lekić, V. & Romanowicz, B., 2011. Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.,* **185**(2), 799–831.

Love, A.E.H., 1911. *Some Problems of Geodynamics,* Cambridge University Press, London.

Lin, F.C., Moschetti, M.P. & Ritzwoller, M.H., 2008. Surface wave tomography of the western United States from ambient seismic noise: Rayleigh and Love wave phase velocity maps, *Geophys. J. Int.,* **173**(1), 281–298.

Lin, F.C., Schmandt, B. & Tsai, V.C., 2012. Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: constraining velocity and density structure in the upper crust, *Geophys. Res. Lett.,* **39**(12), L12303.

Lin, F.C., Tsai, V.C. & Schmandt, B., 2014. 3-D crustal structure of the western United States: application of Rayleigh-wave ellipticity extracted from noise cross-correlations, *Geophys. J. Int.,* **198**(2), 656–670.

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**(1), 113–135.

Malinverno, A., 2000. A Bayesian criterion for simplicity in inverse problem parametrization, *Geophys. J. Int.,* **140**(2), 267–285.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151**(3), 675–688.

Masters, G., Woodhouse, J.H. & Freeman, G., 2011. Mineos v1.0.2 [software], Computational Infrastructure for Geodynamics, tempMi3. Available at: https://geodynamics.org/cig/software/mineos/, last accessed 1 Feb 2018.

Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory,* Vol. **45,** Academic Press.

Montagner, J.P. & Anderson, D.L., 1989. Constrained reference mantle model, *Phys. Earth planet. Inter.,* **58**(2), 205–227.

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

Montagner, J.P., 2002. Upper mantle low anisotropy channels below the pacific plate, *Earth planet. Sci. Lett.,* **202**(2), 263–274.

Montgomery, L.N. *et al.*, 2017. Investigation of firn aquifer structure in southeastern Greenland using active source seismology, *Front. Earth Sci.,* **5,** 10.

Mosegaard, K., 1998. Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse Prob.,* **14**(3), 405–426.

Moulik, P. & Ekström, G., 2016. The relationships between large-scale variations in shear velocity, density, and compressional velocity in the Earth's mantle, *J. geophys. Res. Earth,* **121**(4), 2737–2771

Olugboji, T.M., Lekic, V. & McDonough, W., 2017. A statistical assessment of seismic models of the US continental crust using Bayesian inversion of ambient noise surface wave dispersion data, *Tectonics,* **36,** 1232–1253.

Panning, M. & Romanowicz, B., 2004. Inferences on flow at the base of Earth's mantle based on seismic anisotropy, *Science,* **303**(5656), 351–353.

Pasyanos, M.E., Masters, T.G., Laske, G. & Ma, Z., 2014. LITHO1. 0: An updated crust and lithospheric model of the Earth, *J. geophys. Res.,* **119**(3), 2153–2173.

Petrescu, L., Bastow, I.D., Darbyshire, F.A., Gilligan, A., Bodin, T., Menke, W. & Levin, V., 2016. Three billion years of crustal evolution in eastern Canada: constraints from receiver functions, *J. geophys. Res.* , **121**(2), 788–811.

Ray, A., Key, K., Bodin, T., Myer, D. & Constable, S., 2014. Bayesian inversion of marine CSEM data from the Scarborough gas field using a transdimensional 2-D parametrization, *Geophys. J. Int.,* **199**(3), 1847–1860.

Romanowicz, B., 2002. Inversion of surface waves: a review, *Int. Geophys. Ser.,* **81**(A), 149–174.

Roy, C. & Romanowicz, B.A., 2017. On the implications of a priori constraints in transdimensional bayesian inversion for continental lithospheric layering, *J. geophys. Res.* , **122,** 10,118–10,131.

Rudolph, M.L., Lekić, V. & Lithgow-Bertelloni, C., 2015. Viscosity jump in Earth's mid-mantle, *Science,* **350**(6266), 1349–1352.

Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A,* **371**(1984), 20110547.

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**(1), 88–105.

Silver, P.G. & Holt, W.E., 2002. The mantle flow field beneath western North America, *Science,* **295**(5557), 1054–1057.

Sivia, D. & Skilling, J., 2006. *Data Analysis: A Bayesian Tutorial,* Oxford Univ. Press, Oxford.

Shen, W. & Ritzwoller, M.H., 2016. Crustal and uppermost mantle structure beneath the United States, *J. geophys. Res.,* **121**(6), 4306–4342.

Simmons, N.A., Forte, A.M. & Grand, S.P., 2009. Joint seismic, geodynamic and mineral physical constraints on three-dimensional mantle heterogeneity: implications for the relative importance of thermal versus compositional heterogeneity, *Geophys. J. Int.,* **177**(5), 1284–1304.

Simons, F.J., Van Der Hilst, R.D., Montagner, J.P. & Zielhuis, A., 2002. Multimode Rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle, *Geophys. J. Int.,* **151**(3), 738–754.

Smith, A.F., 1991. Bayesian computational methods, *Phil. Trans. R. Soc. Lond. A: Math. Phys. Eng. Sci.,* **337**(1647), 369–386.

Tanimoto, T. & Rivera, L., 2008. The ZH ratio method for long-period seismic data: sensitivity kernels and observational techniques, *Geophys. J. Int.,* **172**(1), 187–198.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.,* **20**(2), 219–232.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* Society for Industrial and Applied Mathematics.

Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: possible artifacts in seismic tomography, *Science,* **271**(5253), 1257–1260.

Trampert, J. & Woodhouse, J.H., 1996. High resolution global phase velocity distributions, *Geophys. Res. Lett.,* **23**(1), 21–24.

Trampert, J. & Woodhouse, J.H., 2003. Global anisotropic phase velocity maps for fundamental mode surface waves between 40 and 150 s, *Geophys. J. Int.,* **154**(1), 154–165.

van Heijst, H.J. & Woodhouse, J., 1999. Global high-resolution phase velocity distributions of overtone and fundamental-mode surface waves determined by mode branch stripping, *Geophys. J. Int.,* **137**(3), 601–620.

Wagner, L.S., Fouch, M.J., James, D.E. & Hanson-Hedgecock, S., 2012. Crust and upper mantle structure beneath the Pacific Northwest from joint inversions of ambient noise and earthquake data, *Geochem. Geophys. Geosyst.,* **13**(12).

Yano, T., Tanimoto, T. & Rivera, L., 2009. The ZH ratio method for long-period seismic data: inversion for S-wave velocity structure, *Geophys. J. Int.*, **179**(1), 413–424.
Yao, H., van Der Hilst, R.D. & Maarten, V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis—I. Phase velocity maps, *Geophys. J. Int.*, **166**(2), 732–744.
Yao, H., Beghein, C. & Van Der Hilst, R.D., 2008. Surface wave array tomography in SE Tibet from ambient seismic noise and two-station analysis—II. Crustal and upper-mantle structure, *Geophys. J. Int.*, **173**(1), 205–219.
Young, M.K., Tkalcić, H., Bodin, T. & Sambridge, M., 2013. Global P wave tomography of Earth's lowermost mantle from partition modeling, *J. geophys. Res.*, **118**(10), 5467–5486.

APPENDIX A: ATTACHED-TYPE PARAMETRIZATION

Proposal probabilities

For the 'attached' type parametrization, all three elastic parameters are assigned to one Voronoi nucleus. Below, we specify the proposal probabilities corresponding to each of the four possible steps in the Markov Chain: change, move, birth and death.

(1) *Change.* Randomly choose one of the Voronoi nuclei and randomly change all three parameters (Vp, Vs and ρ) based on a normal distribution with specified standard deviation. The proposal ratio is given by:

$$q_{vs}\left(v_s'|v_s\right) = \frac{1}{\sigma_{vs}\sqrt{2\pi}} \exp\left\{-\frac{\left(v_s' - v_s\right)^2}{2\sigma_{vs}^2}\right\} \tag{A1}$$

The proposed value of Vs is denoted by a prime, and is obtained by:

$$v_s' = v_s + u\sigma_{vs} \tag{A2}$$

where σ_{vs} is the standard deviation for perturbations to Vs, and u is a random number generated from the standard normal distribution. Expressions for Vp and density are analogous:

$$q_{vp}\left(v_p'|v_p\right) = \frac{1}{\sigma_{vp}\sqrt{2\pi}} \exp\left\{-\frac{\left(v_p' - v_p\right)^2}{2\sigma_{vp}^2}\right\} \tag{A3}$$

$$v_p' = v_p + u\sigma_{vp} \tag{A4}$$

$$q_{\rho}\left(\rho'|\rho\right) = \frac{1}{\sigma_{\rho}\sqrt{2\pi}} \exp\left\{-\frac{(\rho' - \rho)^2}{2\sigma_{\rho}^2}\right\} \tag{A5}$$

$$\rho' = \rho + u\sigma_{\rho} \tag{A6}$$

In this type of model perturbation, the expressions for the proposal ratios can be used to compute the final acceptance probability. Eqs (A1), (A3) and (A5) clearly satisfy detailed balance conditions:

$$q_{vs}\left({v'}_s|v_s\right) = q_{vs}\left(v_s|v_s'\right) \tag{A7}$$

$$q_{vp}\left({v'}_p|v_p\right) = q_{vp}\left(v_p|v_p'\right) \tag{A8}$$

$$q_{\rho}\left(\rho'|\rho\right) = q_{\rho}\left(\rho|\rho'\right) \tag{A9}$$

Since the three parameters are perturbed at the same time, we have

$$q_{vs}\left({v'}_s|v_s\right) q_{vp}\left({v'}_p|v_p\right) q_{\rho}\left(\rho'|\rho\right) = q_{vs}\left(v_s|v_s'\right) q_{vp}\left(v_p|v_p'\right) q_{\rho}\left(\rho|\rho'\right) \tag{A10}$$

Therefore,

$$\frac{q(m|m')}{q(m'|m)} = 1 \tag{A11}$$

(2) *Move.* In a move step, the depth of a randomly chosen Voronoi nucleus is perturbed based on a normal distribution with specified standard deviation.

$$q_c\left(c_i'|c_i\right) = \frac{1}{\sigma_c\sqrt{2\pi}} \exp\left\{-\frac{\left(c_i' - c_i\right)^2}{2\sigma_c^2}\right\} \tag{A12}$$

Even though the model parametrization changes, the number of Voronoi nuclei does not change, so no additional parameter is added in this step. When considering the proposal ratio of changing from m to m', eq. (A12) clearly satisfies detailed balance:

$$q_c\left({c'}_i|c_i\right) = q_c\left(c_i|c_i'\right) \tag{A13}$$

Therefore,

$$\frac{q(m|m')}{q(m'|m)} = 1 \tag{A14}$$

(3) *Birth.* In the birth step, we randomly choose a depth defined by the uniform prior, and create a new Voronoi nucleus. The velocity and density values corresponding to the newly created nucleus are drawn from normal distributions centred on the current values of Vp, Vs and density at the chosen depth. The standard deviations of the normal distributions (σ_{vp2}, σ_{vs2}, $\sigma_{\rho 2}$) can differ from, but in this study are fixed to be the same as, ones used in the Change step.

The proposal probability for assigning a velocity v_s' to the new Voronoi nucleus (denoted by subscript $n+1$) is given by:

$$q\ \left(v_s'|m\right) = \frac{1}{\sigma_{vs2}\sqrt{2\pi}} \exp\left\{-\frac{\left(v_{s,n+1}' - v_{s,i}\right)^2}{2\sigma_{vs2}^2}\right\} \tag{A15}$$

where $v_{s,i}$ is the velocity of the Voronoi nucleus closest to the depth of the newly born nucleus.

Analogous expressions can be written for Vp and density:

$$q\ \left(v_p'|m\right) = \frac{1}{\sigma_{vp2}\sqrt{2\pi}} \exp\left\{-\frac{\left(v_{p,n+1}' - v_{p,i}\right)^2}{2\sigma_{vp2}^2}\right\} \tag{A16}$$

$$q\ \left(\rho'|m\right) = \frac{1}{\sigma_{\rho 2}\sqrt{2\pi}} \exp\left\{-\frac{\left(\rho_{n+1}' - \rho_i\right)^2}{2\sigma_{\rho 2}^2}\right\} \tag{A17}$$

We now turn our attention to defining the probabilities related to the depth of the Voronoi nucleus. Assume that we have n Voronoi nuclei in the current model m, and there are N possible positions in total for placing a Voronoi nucleus. The probability of having a $(n+1)$th Voronoi nucleus in the rest of the available positions will be:

$$q\ \left(c'|m\right) = \frac{1}{N-n} \tag{A18}$$

The reverse process is to delete the added Voronoi nucleus from m'. The probability of deleting one Voronoi nucleus from $(n+1)$ nuclei is

$$q\ \left(c|m'\right) = \frac{1}{n+1} \tag{A19}$$

The associated probabilities of removing the elastic parameters when their associated Voronoi nucleus is deleted are:

$$q\ \left(v_s|m'\right) = 1 \tag{A20}$$

$$q\ \left(v_p|m'\right) = 1 \tag{A21}$$

$$q\ \left(\rho|m'\right) = 1 \tag{A22}$$

We write the proposal ratio as:

$$\frac{q(m|m')}{q(m'|m)} = \frac{q(c|m')}{q(c'|m)}\frac{q(v_s|m')}{q(v_s'|m)}\frac{q(v_p|m')}{q(v_p'|m)}\frac{q(\rho|m')}{q(\rho'|m)} \tag{A23}$$

Substituting expressions (A15)–(A22) into eq. (A23), we obtain:

$$\left(\frac{q\ (m|m')}{q\ (m'|m)}\right)_{\text{birth}} = \frac{(2\pi)^{3/2}\sigma_{vs2}\sigma_{vp2}\sigma_{\rho 2}\ (N-n)}{(n+1)} \exp\left[\frac{\left(v_{s,n+1}' - v_{s,i}\right)^2}{2\sigma_{vs2}^2} + \frac{\left(v_{p,n+1}' - v_{p,i}\right)^2}{2\sigma_{vp2}^2} + \frac{\left(\rho_{n+1}' - \rho_i\right)^2}{2\sigma_{\rho 2}^2}\right] \tag{A24}$$

(4) *Death.* The death step is the exact reverse of the birth step, since one of the existing n Voronoi nuclei, denoted by index j, is randomly chosen and deleted to create a model m' with $(n-1)$ nuclei. For the death step, since it is supposed to be the exact reverse step of birth, here we consider the situation of changing from n to $(n-1)$ Voronoi nuclei, we have

$$\left(\frac{q\ (m|m')}{q\ (m'|m)}\right)_{\text{death}} = \frac{n}{\sigma_{vs2}\sigma_{vp2}\sigma_{\rho 2}(2\pi)^{3/2}\ (N-n+1)} \exp -\left[\frac{\left(v_{s,j}' - v_{s,i}\right)^2}{2\sigma_{vs2}^2} + \frac{\left(v_{p,j}' - v_{p,i}\right)^2}{2\sigma_{vp2}^2} + \frac{\left(\rho_j' - \rho_i\right)^2}{2\sigma_{\rho 2}^2}\right] \tag{A25}$$

where index i denotes the Voronoi nucleus closest to the deleted nucleus.

Acceptance probabilities

In the move and change step, the number of model parameters does not change, and the proposal ratios are unity. The acceptance probability can be written as:

$$\alpha\ \left(m'|m\right) = \min\left[1, \frac{p\ (m')}{p\ (m)} \cdot \frac{p\ (d|m')}{p\ (d|m)}\right] \tag{A26}$$

Since the dimension of the model does not change, and the priors on all the parameters are uniform, the prior ratio will be either zero or unity, and the acceptance probability can be simplified to:

$$\alpha\ \left(m'|m\right) = \begin{cases} \min\left[1,\ \frac{p\left(d|m'\right)}{p\left(d|m\right)}\right] & \text{if } \forall i \in [1, n],\ v_{si} \in J_s,\ v_{pi} \in J_p,\ \rho_i \in J_\rho \\ 0 & \text{otherwise} \end{cases} \tag{A27}$$

For a birth step, according to eq. (11), the prior ratio takes the form

$$\left(\frac{p\left(m'\right)}{p\left(m\right)}\right)_{\text{birth}} = \begin{cases} \frac{n+1}{(N-n)\Delta v_s \Delta v_p \Delta\rho} & \text{if } (n+1) \in I,\ \text{and } v'_{s,n+1} \in J_s,\ v'_{p,n+1} \in J_p,\ \rho'_{n+1} \in J_\rho \\ 0 & \text{otherwise} \end{cases} \tag{A28}$$

Substituting eqs (15), (A24) and (A28) into eq. (16), the acceptance probability becomes

$$\alpha\ \left(m'|m\right)_{\text{birth}} = \begin{cases} \frac{\left(\sigma_{vs2}\sigma_{vp2}\sigma_{\rho 2}\right)(2\pi)^{3/2}}{\Delta v_s \Delta v_p \Delta\rho} \exp\left\{\frac{\left(v'_{s,n+1}-v_{s,i}\right)^2}{2\sigma^2_{vs2}} + \frac{\left(v'_{p,n+1}-v_{p,i}\right)^2}{2\sigma^2_{vp2}} + \frac{\left(\rho'_{n+1}-\rho_i\right)^2}{2\sigma^2_{\rho 2}} - \frac{\Phi\left(m'\right)-\Phi(m)}{2}\right\} \\ \qquad \text{if } (n+1) \in I,\ \text{and } v'_{s,n+1} \in J_s,\ v'_{p,n+1} \in J_p,\ \rho'_{n+1} \in J_\rho \\ 0 \qquad \text{otherwise} \end{cases} \tag{A29}$$

For the death step, the prior ratio should be the inverse of eq. (A28), except we change from n nuclei to $(n-1)$ nuclei. Therefore, substituting eqs (15) and (A25) into eq. (16), we have:

$$\alpha\ \left(m'|m\right)_{\text{death}} = \begin{cases} \frac{\Delta v_s \Delta v_p \Delta\rho}{\sigma_{vs2}\sigma_{vp2}\sigma_{\rho 2}(2\pi)^{3/2}} \exp\left\{-\frac{\left(v'_{s,j}-v_{s,i}\right)^2}{2\sigma^2_{vs2}} - \frac{\left(v'_{p,j}-v_{p,i}\right)^2}{2\sigma^2_{vp2}} - \frac{\left(\rho'_j-\rho_i\right)^2}{2\sigma^2_{\rho 2}} - \frac{\Phi\left(m'\right)-\Phi(m)}{2}\right\} & \text{if } (n-1) \in I \\ 0 & \text{otherwise} \end{cases} \tag{A30}$$

Independent-type parametrization

In the 'independent' type parametrization, we have three independent sets of Voronoi nuclei, each specifying *Vs*, *Vp* or density. When changing the current model in the Markov Chain, we first randomly choose one out of the three types of parameters (*Vs*, *Vp* and density) with equal probability. Once a specific type of parameter is chosen, the rest of the process and the proposal probabilities are identical to that described in Bodin & Sambridge (2009).

Once the type of parameter to perturb is chosen, it is straightforward to derive the acceptance probability. We take perturbing density as an example. When choosing to change the density value or to move one of the Voronoi nuclei (denoted by index i), the dimension of the model does not change. Therefore, we have:

$$\alpha\ \left(m'|m\right) = \begin{cases} \min\left[1,\ \frac{p\left(d|m'\right)}{p\left(d|m\right)}\right] & \text{if } \forall i \in [1, n],\ \rho_i \in J_\rho \\ 0 & \text{otherwise} \end{cases} \tag{A31}$$

For a birth step:

$$\alpha\ \left(m'|m\right)_{\text{birth}} = \begin{cases} \frac{\left(\sigma_{\rho 2}\right)(2\pi)^{1/2}}{\Delta\rho} \exp\left\{\frac{\left(\rho'_{n+1}-\rho_i\right)^2}{2\sigma^2_{\rho 2}} - \frac{\Phi\left(m'\right)-\Phi(m)}{2}\right\} & \text{if } (n+1) \in I,\ \text{and } \rho'_{n+1} \in J_\rho \\ 0 & \text{otherwise} \end{cases} \tag{A32}$$

where n is the number of Voronoi nuclei defining the density structure.

Finally, for the death step in which Voronoi nucleus j is deleted, and Voronoi nucleus i is the remaining nucleus closest to the deleted one, we have:

$$\alpha\ \left(m'|m\right)_{\text{death}} = \begin{cases} \frac{\Delta\rho}{\sigma_{\rho 2}(2\pi)^{1/2}} \exp\left\{-\frac{\left(\rho'_j-\rho_i\right)^2}{2\sigma^2_{\rho 2}} - \frac{\Phi\left(m'\right)-\Phi(m)}{2}\right\} & \text{if } (n-1) \in I \\ 0 & \text{otherwise} \end{cases} \tag{A33}$$

APPENDIX B: CONVERGENCE ANALYSIS FOR TB INVERSION OF SWD

$$\text{rmsd}_{vs}\ (z) = \sqrt{\frac{\sum_{i=1}^{n(z)}\left(y_i\left(z\right) - \overline{y_i\left(z\right)}\right)^2}{n\left(z\right)}} \tag{B1}$$

Here, z stands for the depth; n stands for the number of bins used to discretize the PDF; $y_i(z)$ is the value of the PDF at the depth z; $\overline{y_i(z)}$ is the mean value of the PDF across four ensembles. The rmsds of the four ensembles are close to each other and have low absolute values across all depths. This confirms that after the chosen burn-in period, the ensemble results from different starting models are indistinguishable from each other.

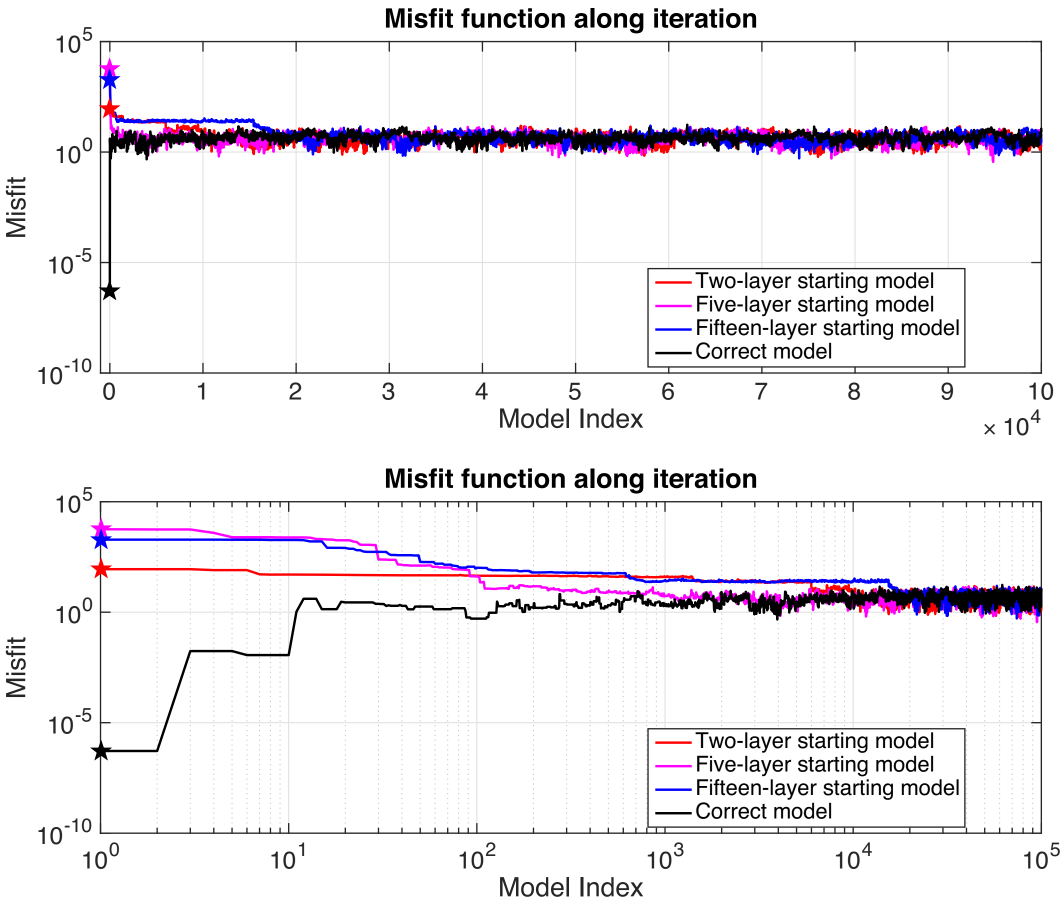

Figure B1. The evolution of misfit along the rjMCMC with different starting models. Four starting models with different initial numbers of layers are used here to invert surface wave dispersion data between 5 and 100 s. Every 100th model from a total of 10 million iterations is plotted on a linear (top) and logarithmic (bottom) x-axis. After about 2×10^5 iterations, all four rjMCMC remain at a low misfit. We choose a burn-in period of 2.5×10^6 iterations to ensure the rjMCMC have converge before sampling the ensemble solution.

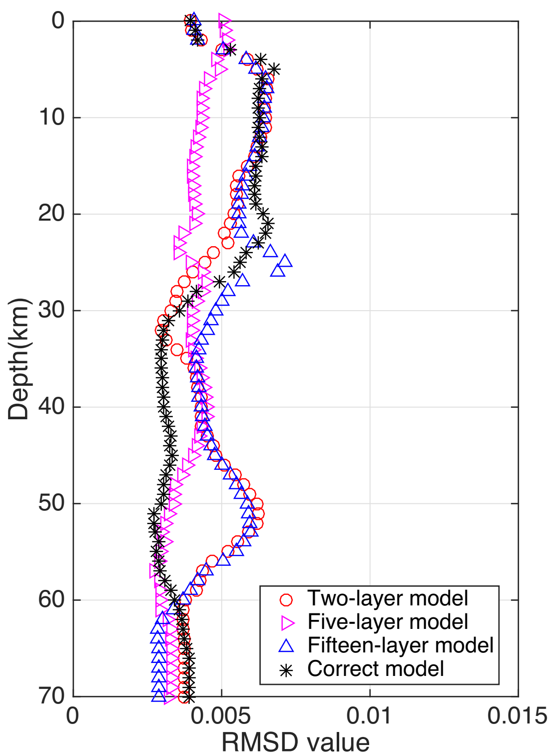

Figure B2. Root-mean-square deviation of *Vs* posterior probability density functions from TB inversion of surface wave dispersion using four different starting models. We use the following equation to calculate to root-mean-square deviation (rmsd) for the four ensembles.

APPENDIX C: NORMALIZED ROOT-MEAN-SQUARE ERROR OF *Vs* AND *Vp* ENSEMBLE FROM Fig. 5

$$\mathrm{nrmse}_{Vs}\ (i) = \frac{\mathrm{rmse}_{Vs}\,(i)}{\overline{V_s}\,(i)}$$

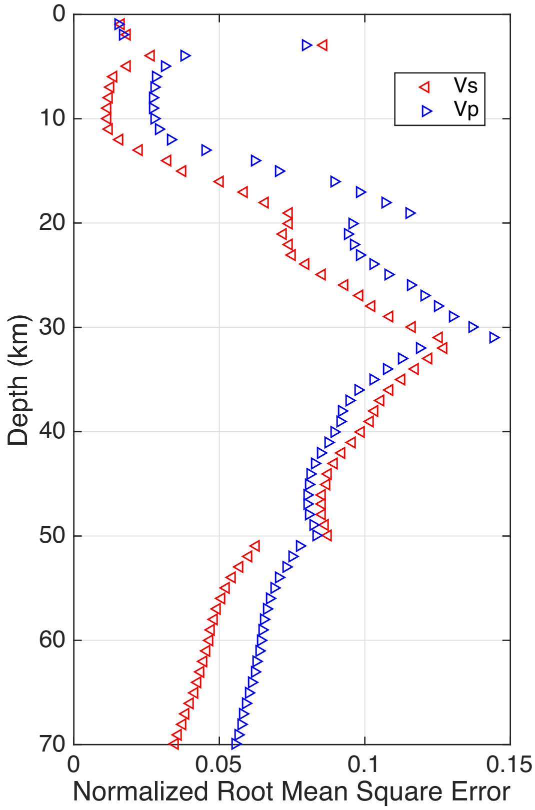

Figure C1. Normalized root-mean-square error of *Vs* (red) and *Vp* (blue) inversion ensemble against the input model. The ensemble is taken from test shown in Fig. 5 where *Vs*, *Vp* and density are inverted using SWD and ZH ratio. The rmse of *Vs* and *Vp* are normalized by their mean at that given depth.

Take *Vs* as an example; here i stands for the depth, $\overline{V_s}(i)$ is the mean value of *Vs* from the ensemble at that depth, $\mathrm{rmse}_{Vs}(i)$ is the root-mean-square error of *Vs* against the target at that depth.

APPENDIX D: TB JOINT INVERSION OF SWD AND ZH RATIOS FOR TARGET MODEL IN WHICH THE DENSITY GEOMETRY IS DIFFERENT FROM THAT OF *Vs*

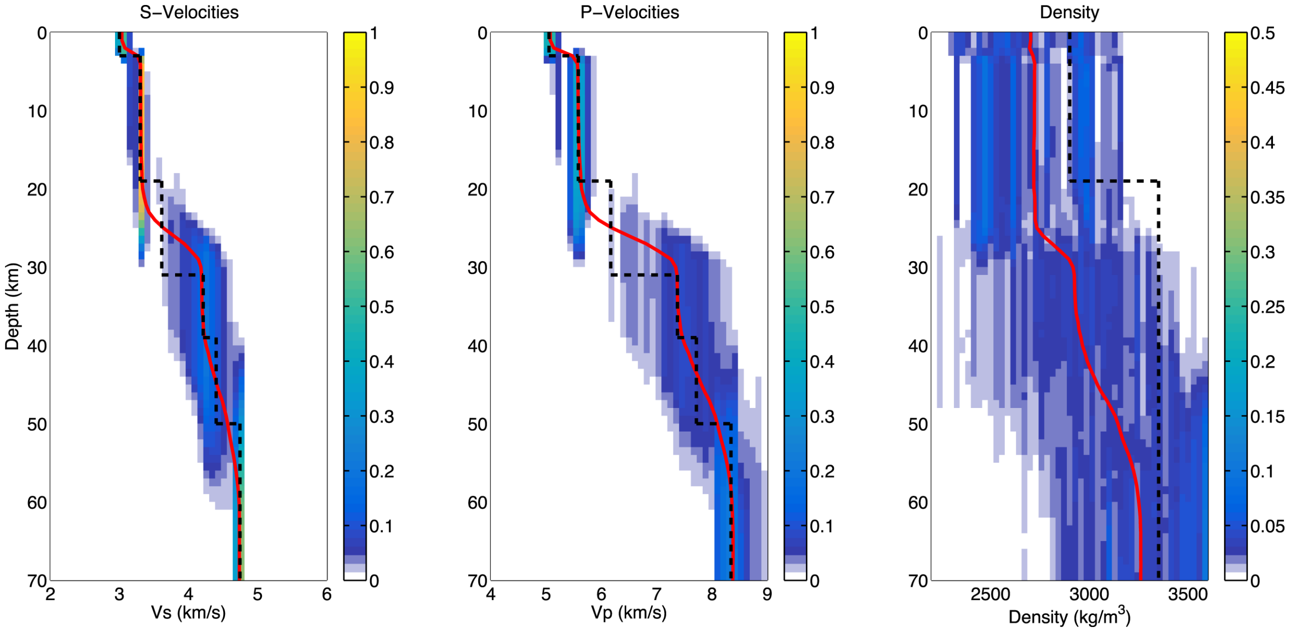

Figure D1. *Vs* (left), *Vp* (middle) and density (right) posterior probability density functions obtained by TB joint inversion using synthetic SWD and ZH ratio data. In this test, the attached-type parametrization is used. The ensemble solutions are displayed as probability density functions at each depth, with warmer colours corresponding to higher posterior probabilities, and the solid red line denoting the 5 per cent trimmed mean of the posterior. The target model (black dashed lines) is used to generate synthetic data. The density structure of the target model has a single abrupt change at 19 km depth, unlike the velocity structure, where the Moho is at 31 km. The retrieved density structure does not resolve the density increase at 19 km, and instead shows an increase around 31 km, imposed by the geometry of velocity structure through the attached-type parametrization used for inversion.