Geophysical Journal International

Geophys. J. Int. (2016)
Advance Access publication 2016 April 4
GJI Seismology

doi: 10.1093/gji/ggw124

Imaging anisotropic layering with Bayesian inversion of multiple data types

T. Bodin,[1,2] J. Leiva,[1] B. Romanowicz,[1,3,4] V. Maupin[5] and H. Yuan[6]

[1] *Berkeley Seismological Laboratory,* 215 *McCone Hall, UC Berkeley, Berkeley CA* 94720-4760*, USA. E-mail: thomas.bodin@ens-lyon.fr*
[2] *Univ Lyon, Universite Lyon* 1*, Ens de Lyon, CNRS, UMR* 5276 *LGL-TPE, F-*69622 *Villeurbanne, France*
[3] *Insitut de Physique du Globe de Paris (IPGP),* 1 *Rue Jussieu F-*75005 *Paris, France*
[4] *College de France,* 11 *place Marcelin Berthelot, F-*75231 *Paris, France*
[5] *Department of Geosciences, University of Oslo, P.O. Box* 1047*, Blindern, Oslo* 0316*, Norway*
[6] *Department of Earth and Planetary Sciences, Macquarie University, New South Wales* 2109*, Australia*

Accepted 2016 March 31. Received 2016 March 31; in original form 2015 September 1

SUMMARY

Azimuthal anisotropy is a powerful tool to reveal information about both the present structure and past evolution of the mantle. Anisotropic images of the upper mantle are usually obtained by analysing various types of seismic observables, such as surface wave dispersion curves or waveforms, SKS splitting data, or receiver functions. These different data types sample different volumes of the earth, they are sensitive to different length scales, and hence are associated with different levels of uncertainties. They are traditionally interpreted separately, and often result in incompatible models. We present a Bayesian inversion approach to jointly invert these different data types. Seismograms for SKS and *P* phases are directly inverted using a cross-convolution approach, thus avoiding intermediate processing steps, such as numerical deconvolution or computation of splitting parameters. Probabilistic 1-D profiles are obtained with a transdimensional Markov chain Monte Carlo scheme, in which the number of layers, as well as the presence or absence of anisotropy in each layer, are treated as unknown parameters. In this way, seismic anisotropy is only introduced if required by the data. The algorithm is used to resolve both isotropic and anisotropic layering down to a depth of 350 km beneath two seismic stations in North America in two different tectonic settings: the stable Canadian shield (station FFC) and the tectonically active southern Basin and Range Province (station TA-214A). In both cases, the lithosphere–asthenosphere boundary is clearly visible, and marked by a change in direction of the fast axis of anisotropy. Our study confirms that azimuthal anisotropy is a powerful tool for detecting layering in the upper mantle.

Key words: Inverse theory; Body waves; Surface waves and free oscillations; Seismic anisotropy; North America.

1 INTRODUCTION

Seismic anisotropy in the crust and upper mantle can be produced by multiple physical processes at different spatial scales. In the mantle, plastic deformation of olivine aggregates results in a crystallographic preferential orientation (CPO) of minerals, and produces large-scale seismic anisotropy that can be observed seismologically. These observations are usually related to the strain field, and interpreted in terms of either present day flow, or 'frozen' flow from the geological past. Furthermore, the spatial distribution of cracks, fluid inclusions, or seismic discontinuities can induce apparent anisotropy, called shape-preferred orientation (SPO) anisotropy (Crampin & Booth 1985; Backus 1962). In this way, anisotropic properties of rocks are closely related to their geological history and present configuration, and hence reveal essential information about the Earth's structure and dynamics (e.g. Montagner & Guillot 2002).

Observations of seismic anisotropy depend on the 21 parameters of the full elastic tensor. However, all these parameters cannot be resolved independently at every location, and seismologists usually rely on simplified assumptions on the type of anisotropy, namely hexagonal symmetry. This type of anisotropy is defined by the five Love parameters of transverse isotropy (*A, C, F, L, N*) and two angles describing the direction of the axis of symmetry (Love 1927). Most seismological studies assume one of two types of anisotropy: (1) radial anisotropy, where the axis of hexagonal symmetry is vertical and with no azimuthal dependence and (2) azimuthal anisotropy, where the axis of hexagonal symmetry is horizontal with unknown direction. Retrieving the tilt of the hexagonal axis of symmetry is in principle possible (Montagner & Nataf 1988; Plomerová & Babuška

2010; Xie *et al.* 2015), but in practice difficult, due to limitations in the available azimuthal coverage, trade-offs with other competing factors, such as tilted layers in the case of body waves and non-uniqueness of the solution in the case of surface waves inversion.

Azimuthal anisotropy in the crust and upper mantle can be observed from different seismic measurements, sampling the Earth at different scales: surface wave observations, core-refracted shear wave (SKS) splitting measurements and receiver functions. The latter two methods rely on relatively high-frequency teleseismic body waves measurements, and therefore can provide good lateral resolution in those areas of continents where broad-band station coverage is dense, if good azimuthal coverage is available.

Receiver functions have the potential of resolving layered anisotropic structure locally. Large data sets from single seismic stations have been used to image both anisotropic and dipping structures primarily at crustal depths (e.g. Kosarev *et al.* 1984; Levin & Park 1997; Peng & Humphreys 1997; Savage 1998; Farra & Vinnik 2000; Frederiksen & Bostock 2000; Leidig & Zandt 2003; Vergne *et al.* 2003; Schulte-Pelkum & Mahan 2014; Audet 2015; Bianchi *et al.* 2015; Liu *et al.* 2015; Vinnik *et al.* 2016). Harmonic decomposition methods have been developed to distinguish the contributions from isotropic and anisotropic discontinuities, and dipping layers (Kosarev *et al.* 1984; Girardin & Farra 1998; Bianchi *et al.* 2010).

Shear wave splitting measurements in core-refracted phases (SKS and SKKS) provide constraints on the integrated effect of azimuthal anisotropy across the thickness of the mantle beneath a single station (e.g. Vinnik *et al.* 1984, 1989; Silver & Chan 1991; Vinnik *et al.* 1992; Silver 1996; Long & Silver 2009), but depth resolution is generally poor, even when considering finite-frequency kernels (Chevrot 2006), and there are trade-offs between the strength of anisotropy and the thickness of the anisotropic domain. Due to the lack of sufficient azimuthal coverage to distinguish more than one layer, shear wave splitting measurements are usually interpreted under the assumption of a single layer of anisotropy with a horizontal axis of symmetry. We note however several attempts to map multiple layers as well as a dipping fast axis (Silver & Savage 1994; Levin *et al.* 1999; Hartog & Schwartz 2000; Yuan *et al.* 2008).

Surface wave tomographic inversions provide constraints on both radial (Gung *et al.* 2003; Plomerová *et al.* 2002; Nettles & Dziewoński 2008; Fichtner *et al.* 2010) and azimuthal anisotropy at the regional (Forsyth 1975; Simons *et al.* 2002; Deschamps *et al.* 2008; Beghein *et al.* 2010; Fry *et al.* 2010; Adam & Lebedev 2012; Darbyshire *et al.* 2013; Zhu & Tromp 2013; Legendre *et al.* 2014; Köhler *et al.* 2015) and global scale (Tanimoto & Anderson 1985; Montagner & Nataf 1986; Trampert & van Heijst 2002; Trampert & Woodhouse 2003; Debayle *et al.* 2005; Beucler & Montagner 2006; Visser *et al.* 2008; Debayle & Ricard 2012, 2013; Yuan & Beghein 2013, 2014). Surface waves provide better vertical resolution than SKS data, but are limited in horizontal resolution due to the long wavelengths. Still, the depth range where vertical resolution is achieved depends on the frequency range considered (longer periods sample deeper depths), as well as the type of surface waves considered. Most studies are based on the analysis of fundamental mode surface wave dispersion up to about 200–250 s, which have good resolution down to lithospheric depths, although inclusion of surface wave overtones can improve resolution at depth (e.g. Yuan & Beghein 2014; Durand *et al.* 2015).

These three different data types are therefore characterized by different sensitivities to structure. They are modelled with different approximations of the wave equation, and associated with different noise levels. A well-known problem is that they often provide incompatible anisotropic models, and lead to contradictory interpretations. For example, surface waves and SKS waves sample different volumes in the earth, and SKS splitting measurements often disagree with predictions made from surface wave tomographic models (e.g. Montagner *et al.* 2000; Conrad *et al.* 2007; Becker *et al.* 2012; Wang & Tape 2014). This discrepancy can be explained by the progressive loss of resolution of fundamental mode surface waves below depths of 200–250 km. Furthermore, body waves and surface waves are measured in different frequency bands, and hence are sensitive to structure at different wavelengths. The sharp discontinuities that can be resolved by receiver functions are usually mapped into apparent radial anisotropy in smooth models constructed from surface waves (Capdeville *et al.* 2013; Bodin *et al.* 2015).

In order to improve resolution in anisotropy, several studies have proposed joint inversion algorithms combining body waves and surface waves. Marone & Romanowicz (2007), Yuan & Romanowicz (2010b) and Yuan *et al.* (2011) iteratively combined 3-D waveform tomography (including fundamental surface waves and overtones) with constraints from shear wave splitting data in North America. They showed that by incorporating body waves, the anisotropy strength significantly increases at the asthenospheric depth, while the directions remain largely unchanged. However, these models are obtained by linearized and damped inversions, where the produced seismic models strongly depend on choices made at the outset (reference model, regularization). This precludes propagation of uncertainties from observations to inverted models, and hence makes the interpretation difficult. In another approach, Vinnik *et al.* (2007), Obrebski *et al.* (2010) and Vinnik *et al.* (2014) performed a 1-D Monte Carlo joint inversion of SKS and receiver functions at individual broad-band stations, but long-wavelength information from surface waves was not used in this case. Therefore, two main challenges remain in anisotropic imaging:

(i) To our knowledge, azimuthal variations of surface wave dispersion measurements have never been inverted jointly with receiver functions.

(ii) It is difficult to jointly invert different data types, as inverted models strongly depend on the choice of parameters used to weigh the relative contribution of each data sets in the inversion.

In this work, we address these issues with a method for 1-D inversion under a seismic station. We jointly invert Rayleigh wave dispersion curves with their azimuthal variations, together with converted body waves and SKS data. For body waves, standard inversion procedures are usually based on secondary observables, such as deconvolved waveforms (receiver functions) or splitting parameters for SKS data. Here, we directly invert the different components of seismograms with a cross-convolution approach, as this allows us to better propagate uncertainties from recorded waveforms towards a velocity model (Menke & Levin 2003; Bodin *et al.* 2014). We cast the problem in a Bayesian framework, and explore the space of earth models with a Markov chain Monte Carlo algorithm. This allows us to deal with the non-linear and non-unique nature of the problem, and quantify uncertainties. The solution is a probabilistic 1-D profile describing shear wave velocity, strength of azimuthal anisotropy and fast axis direction, at each depth. We use a transdimensional formulation where the number of layers as well as the presence of anisotropy in each layer are treated as free variables.

2 METHODOLOGY

2.1 Model parametrization

The full elastic tensor of 21 parameters is usually described with the so-called Voigt notation 6×6 symmetric matrix C_{mn} (Maupin & Park 2007). An elastic medium with hexagonal (i.e. cylindrical) symmetry and horizontal axis of symmetry is called a horizontal transverse isotropic model (HTI), and is usually defined by the five Love parameters of transverse isotropy A, C, F, L, N (Love 1927):

$$\mathbf{C}_{mn} = \begin{pmatrix} A & F & (A-2N) & 0 & 0 & 0 \\ F & C & F & 0 & 0 & 0 \\ (A-2N) & F & A & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 \\ 0 & 0 & 0 & 0 & N & 0 \\ 0 & 0 & 0 & 0 & 0 & L \end{pmatrix} \tag{1}$$

Here, axis 3 is vertical and axis 2 is the horizontal axis of symmetry. A, C, N and L can be related to P- and S-wave velocities in different directions. If ψ_{fast} is the angle of the fast axis relative to North, the velocity of S waves propagating horizontally and polarized vertically (SV waves) is given by (Crampin 1984):

$$\rho V_{sv}^2(\psi) = \frac{(L+N)}{2} + \frac{(L-N)}{2}\cos(2(\psi - \psi_{\text{fast}})) \tag{2}$$

where ψ is the direction of propagation relative to North. The velocity of P waves and SH waves propagating in the horizontal plane are a bit more complicated as they contain $\cos(4\psi)$ terms. The corresponding expressions can be found in Crampin (1984).

In this work, instead of using elastic parameters, we follow the notation used in most body waves studies, and parametrize our model in terms of seismic velocities, where the isotropic component is given by the values of V_s and V_p, and the anisotropic component is defined in terms of 'peak to peak' level of anisotropy δV_p, δV_s (e.g. Farra *et al.* 1991; Romanowicz & Yuan 2012). These parameters are related to the five elastic parameters A, C, F, L, N by the following expressions:

$$\frac{C}{\rho} = \left(V_p + \frac{\delta V_p}{2}\right)^2, \qquad \frac{A}{\rho} = \left(V_p - \frac{\delta V_p}{2}\right)^2, \tag{3}$$

$$\frac{L}{\rho} = \left(V_s + \frac{\delta V_s}{2}\right)^2, \qquad \frac{N}{\rho} = \left(V_s - \frac{\delta V_s}{2}\right)^2, \tag{4}$$

The elastic parameter F controls the velocity along the direction intermediate between the fast and the slow directions. It is common to parametrize it with the fifth parameter $\eta = F/(A-2L)$, which we set to one (i.e. $F = A - 2L$) as in PREM (Dziewonski & Anderson 1981). Following Obrebski *et al.* (2010), we also impose $V_p/V_S = 1.7$ for the sake of simplicity. The density ρ is calculated through the empirical relation $\rho = 2.35 + 0.036(V_p - 3)^2$ as done in Tkalčić *et al.* (2006). In order to reduce the number of parameters, the ratio between the percentage of anisotropy for the compressional and shear waves $(\delta V_p/V_p)/(\delta V_s/V_s)$ is fixed at 1.5 based on the analysis of published data for the upper mantle (Obrebski *et al.* 2010). Here, we acknowledge that surface waves and normal modes are sensitive to parameter η, V_p and density (Beghein & Trampert 2004; Beghein *et al.* 2006; Kustowski *et al.* 2008), and that we could have easily treated these parameters as unknowns in the inversion. It has been demonstrated that η trade-offs with P-wave anisotropy (Beghein *et al.* 2006; Kustowski *et al.* 2008), implying that making assumptions on either one of these parameters will likely affect results and inferred model uncertainties. Although one could invert for the entire elastic tensor in each layer, this would be at increased computational cost. Here instead, we use these empirical scaling relations to determine the least constrained parameters.

As shown in Fig. 1, our model is parametrized in terms of a stack of layers with constant seismic velocity. In our transdimensional formalism, the number of unknowns is variable, as we want to explain our data sets with the least number of free parameters. Each layer can be either isotropic and described solely by its shear wave velocity V_s (in this case, $\delta V_s = 0$), or azimuthally anisotropic and described by three parameters: V_s, δV_s and ψ_{fast} the direction of the horizontal fast axis relative to the north. The layer thickness is also variable and the last layer is a half-space. The other parameters (ρ, V_p, δV_s) are given by the scaling relations mentioned above.

The number of layers k as well as the number of anisotropic layers $l \leq k$ are free parameters in the inversion (see Fig. 1). Therefore, the complete model to be inverted for is defined as

$$\mathbf{m} = [\mathbf{z}, \mathbf{V}_s, \delta\mathbf{V}_s, \mathbf{\Psi}_{\text{fast}}], \tag{5}$$

where the vector $\mathbf{z} = [z_1, \ldots, z_k]$ represents the depths of the k discontinuities, $\mathbf{V}_s$ is a vector of size k and $\delta\mathbf{V}_s$, $\mathbf{\Psi}_{\text{fast}}$ are vectors of size l. The total number of parameters in the problem (i.e. the dimension of vector $\mathbf{m}$) is therefore $2(k+l)$. We shall show how a Monte Carlo algorithm can explore different types of model parametrizations.

As in any data inference problem, it is clear that observations can always be better explained with more model parameters (with l and k large). However, we will see that in a Bayesian framework, overly complex models with a large number of parameters have a lower probability and are naturally penalized. Between a simple and a complex model that fit the data equally well, the simple one will be preferred. With this formulation, anisotropy will only be included into the model if required by the data.

When inverting long-period seismic waves, this flexible approach to parametrizing an elastic medium allowed us to quantify the trade-off between vertical heterogeneities (lots of small isotropic layers) and radial anisotropy (fewer anisotropic layers) (Bodin *et al.* 2015). This trade-off can be broken by adding higher frequency observations from body waves, thus allowing a consistent interpretation of different data types.

2.2 The data

For surface waves, we assume that some previous analysis (local or tomographic) provides us with the phase velocity dispersion at the station and its azimuthal variation. To first order, the phase velocity of surface waves in an anisotropic medium can be written as:

$$\mathbf{C}(T,\psi) = \mathbf{C}_0(T) + \mathbf{C}_1(T)\cos(2\psi) + \mathbf{C}_2(T)\sin(2\psi) + \mathbf{C}_3(T)\cos(4\psi) + \mathbf{C}_4(T)\sin(4\psi) \tag{6}$$

where T is the period and ψ is the direction of propagation relative to the north (Smith & Dahlen 1973). For fundamental mode Rayleigh waves, the 2ψ terms $\mathbf{C}_1$ and $\mathbf{C}_2$ are sensitive to depth variations of V_s, δV_s and ψ_{fast}, and the 4ψ terms are negligible, due to the low amplitude of sensitivity kernels (Montagner & Tanimoto 1991; Maupin & Park 2007). We will therefore only invert $\mathbf{C}_0(T)$, $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$, and ignore 4ψ terms.

For body waves, recorded seismograms of P and SKS phases for events coming from different backazimuths will be inverted. To reduce the level of noise in P waveforms, individual events coming from the same regions (i.e. within a small backazimuth–

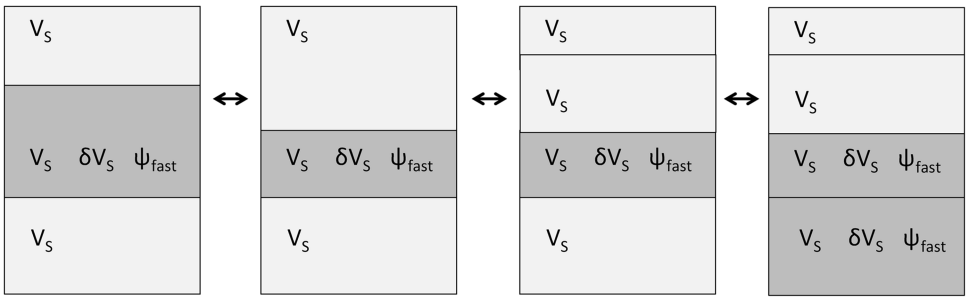

Figure 1. The 1-D model is parametrized with a variable number of layers, which can be either isotropic and described by one parameter V_s (light grey), or anisotropic and described by three parameters V_s, δV_s and ψ_{fast} (dark grey). As our Monte Carlo parameter search algorithm samples the space of possible models, different types of jumps (black arrows) are used to explore different geometries (change the depth of a discontinuity, add/remove an isotropic layer and add/remove anisotropy to an existing layer).

distance range) will be stacked (Kumar *et al.* 2010). This reduces the number of waveforms that need to be modelled by the inversion algorithm, and hence reduces computational cost. The range of ray parameters and backazimuths in each bin directly impacts the type of errors in the stacked seismograms. A large bin allows one to include more events resulting in ambient and instrumental noise reduction, although 3-D and moveout effects also become more significant in this case. A compromise needs to be found for each experiment when defining the range of incident rays.

When analysing azimuthal variations of receiver functions or SKS waveforms, a number of studies decompose waveforms into angular harmonics, in order to isolate π-periodic variations that can be explained by azimuthal anisotropy (e.g. Kosarev *et al.* 1984; Girardin & Farra 1998; Farra & Vinnik 2000; Bianchi *et al.* 2010; Audet 2015). However in this work, waveforms will not be filtered to isolate π-periodic azimuthal variations, and our 'raw' data will also contain azimuthal variations due to 3-D effects, such as dipping discontinuities (associated with 2π-periodic variations). These variations will be accounted for as data noise in our Bayesian formulation.

2.3 The forward calculation

We use a forward modelling approach, where at each step of a Monte Carlo sampler, a new model $\mathbf{m}$, as defined in Fig. 1, is tested, and synthetic data predicted from this model are compared to actual measurements.

For Rayleigh waves, dispersion curves and their azimuthal variations are computed with a normal mode formalism in a spherical Earth (Smith & Dahlen 1973). The term $\mathbf{C}_0(T)$ is computed in a fully non-linear fashion with a Runge–Kutta matrix integration (Saito 1967; Takeuchi & Saito 1972; Saito 1988). However, the relation linking the model parameters to the terms $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$ is linearized around the current model $\mathbf{m}$ averaged azimuthally, which is radially anisotropic (Maupin 1985; Montagner & Nataf 1986). A detailed description of the procedure is given in Appendix A. We acknowledge that we are limited to a linear approximation of the problem for azimuthal terms. Future work includes treating the problem fully non-linearly, and computing dispersion curves exactly as done in Thomson (1997).

For body waves, that is, *P* and SKS waveforms, the impulse response of the model $\mathbf{m}$ to an incoming planar wave with frequency ω and slowness p is computed with a reflectivity propagator-matrix method (Levin & Park 1998). The transmission response is calculated in the Fourier domain at a number of different frequencies. Particle motion at the surface is then obtained by an inverse Fourier transform. The algorithm is outlined in detail in Park (1996) and Levin & Park (1997). The computational cost of this algorithm varies linearly with the number of frequencies ω and the number of layers in $\mathbf{m}$.

We note here that the Rayleigh wave dispersion curves are computed in a spherical earth whereas body waves are predicted for a flat Earth, which may produce some inconsistencies. However, *P* and SKS waves propagate almost vertically under the station, and hence are only poorly sensitive to the earth sphericity.

2.4 Bayesian inference

We cast our inverse problem in a Bayesian framework, where information on model parameters is represented in probabilistic terms (Box & Tiao 1973; Smith 1991; Gelman *et al.* 1995). Geophysical applications of Bayesian inference are described in Tarantola & Valette (1982), Duijndam (1988a,b) and Mosegaard & Tarantola (1995). The solution is given by the *a posteriori* probability distribution (or posterior distribution) $p(\mathbf{m}|\mathbf{d}_{\text{obs}})$, which is the probability density of the model parameters $\mathbf{m}$, given the observed data $\mathbf{d}_{\text{obs}}$. The posterior is given by Bayes' theorem:

$$\text{posterior} \propto \text{likelihood} \times \text{prior} \tag{7}$$

$$p(\mathbf{m} \mid \mathbf{d}_{\text{obs}}) \propto p(\mathbf{d}_{\text{obs}} \mid \mathbf{m}) p(\mathbf{m}). \tag{8}$$

The term $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ is the likelihood function, which is the probability of observing the measured data given a particular model. $p(\mathbf{m})$ is the *a priori* probability density of $\mathbf{m}$, that is, what we know about the model $\mathbf{m}$ before measuring the data $\mathbf{d}_{\text{obs}}$.

In a transdimensional formulation, the number of unknowns (i.e. the dimension of $\mathbf{m}$) is not fixed in advance, and so the posterior is defined across spaces with different dimensions. Below we show how the likelihood and prior distributions are defined in our problem, and how a transdimensional Monte Carlo sampling scheme can be used to generate samples from the posterior distribution, that is, an ensemble of vectors $\mathbf{m}$ whose density reflects that of the posterior distribution.

2.5 The likelihood function

The likelihood function $p(\mathbf{d}_{\text{obs}}|\mathbf{m})$ quantifies how well a given model $\mathbf{m}$ can reproduce the observed data. Assuming that different data types are measured independently, we can write:

$$p(\mathbf{d}_{\text{obs}} \mid \mathbf{m}) = p(\mathbf{C}_0 \mid \mathbf{m}) p(\mathbf{C}_1 \mid \mathbf{m}) p(\mathbf{C}_2 \mid \mathbf{m}) p(\mathbf{d}_p \mid \mathbf{m}) p(\mathbf{d}_{\text{SKS}} \mid \mathbf{m}) \tag{9}$$

where $\mathbf{C}_0$, $\mathbf{C}_1$ and $\mathbf{C}_2$ are surface wave dispersion curves (see eq. 6), and where $\mathbf{d}_P$ and $\mathbf{d}_{\mathrm{SKS}}$ are seismograms observed for P and SKS waves.

2.5.1 *Surface wave measurements*

For Rayleigh wave dispersion curves ($\mathbf{C}_0(T)$, $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$), we assume that data errors (both observational and theoretical) are not correlated and are distributed according to a multivariate normal distribution with zero mean and variances σ_{C_0}, σ_{C_1} and σ_{C_2} respectively. For $\mathbf{C}_0(T)$, the likelihood probability distribution writes:

$$p(\mathbf{C}_0 \mid \mathbf{m}) = \frac{1}{(\sqrt{2\pi}\sigma_{C_0})^n} \times \exp\left\{\frac{-\|\mathbf{C}_0 - \mathbf{c}_0(\mathbf{m})\|^2}{2\sigma_{C_0}^2}\right\}, \tag{10}$$

where n is the number of data points, that is, the number of periods considered and $\mathbf{c}_0(\mathbf{m})$ is the dispersion curve predicted for model $\mathbf{m}$. In the same way, we define the likelihoods for 2ψ terms $p(\mathbf{C}_1|\mathbf{m})$ and $p(\mathbf{C}_2|\mathbf{m})$.

2.5.2 *A cross-convolution likelihood function for body waves*

In traditional receiver function analysis, the vertical component of a P waveform is deconvolved from the horizontal components, to remove source and distant path effects (Langston 1979). The resulting receiver function waveform can then be inverted for a 1-D seismic model, by minimizing the difference between observed and predicted receiver functions:

$$\phi(\mathbf{m}) = \left\| \frac{\mathbf{H}_{\mathrm{obs}}(t)}{\mathbf{V}_{\mathrm{obs}}(t)} - \frac{\mathbf{h}(t,\mathbf{m})}{\mathbf{v}(t,\mathbf{m})} \right\|, \tag{11}$$

where $\mathbf{V}_{\mathrm{obs}}(t)$ and $\mathbf{H}_{\mathrm{obs}}(t)$ are observed seismograms for vertical and radial components, and $\mathbf{v}(t, \mathbf{m})$ and $\mathbf{h}(t, \mathbf{m})$ are the vertical and radial impulse response functions of the near receiver structure, calculated for model $\mathbf{m}$. Here, the division sign represents a spectral division, or deconvolution. Although receiver function analysis has been extensively used for years, there are two well-known issues:

(i) The deconvolution is a numerical unstable procedure that needs to be stabilized (e.g. water level deconvolution; use of a low-pass filter). This results in a loss of resolution, which trade-offs with errors in the receiver function.

(ii) Uncertainties in receiver functions are therefore difficult to estimate.

These two issues have been well studied in the last decades (e.g. Park & Levin 2000; Kolb & Lekić 2014). Following Menke & Levin (2003), we propose a misfit function for inverting converted body waves without deconvolution, by defining a vector of residuals as follows (Bodin *et al.* 2014):

$$\mathbf{r}(\mathbf{m}, t) = \mathbf{v}(t, \mathbf{m}) * \mathbf{H}_{\mathrm{obs}}(t) - \mathbf{h}(t, \mathbf{m}) * \mathbf{V}_{\mathrm{obs}}(t), \tag{12}$$

where the sign $*$ represents a time-domain discrete convolution. The vector $\mathbf{r}$ is a function of observed and predicted data defined such that the unknown source function and distant path effects are accounted for in both terms giving $\mathbf{r} = 0$ for the true model parameters $\mathbf{m}$ and zero errors. The norm $\|\mathbf{r}(\mathbf{m})\|$ is used as a misfit function, and is equivalent to the distance between observed and predicted receiver functions in (11). However, (1) it does not involve any deconvolution and no damping parameters need to be chosen; (2) the probability density function for $\mathbf{r}(\mathbf{m}, t)$ can be estimated from errors statistics in observed seismograms. If we assume that errors in $\mathbf{V}_{\mathrm{obs}}(t)$ and $\mathbf{H}_{\mathrm{obs}}(t)$ are normally distributed and not correlated (Gaussian white noise), we have (see Appendix B for details):

$$p(\mathbf{r} \mid \mathbf{m}) = \frac{1}{(\sqrt{2\pi}\sigma_p)^n} \times \exp\left\{\frac{-\|\mathbf{r}(\mathbf{m})\|^2}{2\sigma_p^2}\right\}. \tag{13}$$

Table 1. Possible component pairs that can be used in an inversion based on the cross-convolution misfit function defined by Menke & Levin (2003). These four different pairs have complementary sensitivities to seismic discontinuities and anisotropy. The advantage of a cross-convolution misfit function is that these different data types can all be inverted in the same manner.

Conversions P_{SV}	R-Z components	Phase P
Conversions P_{SH}	T-Z components	Phase P
Conversions S_P	R-Z components	Phase S
SKS splitting	R-T components	Phases SKS and SKKS

For a given P waveform $\mathbf{d}_p = [\mathbf{V}_{\mathrm{obs}}(t), \mathbf{H}_{\mathrm{obs}}(t)]$, resulting from a stack of events coming from similar distances and backazimuths, we use the distribution in (13) as the likelihood function $p(\mathbf{d}_p|\mathbf{m})$ to quantify the level of agreement between observations and the predictions from a proposed earth model. Then, we combine a number of stacked waveforms measured at different backazimuths–distance bins by simply using the product of their likelihoods, thus resulting in a joint inversion of several waveforms with different incidence angles. A clear advantage is that we can use the same formalism to construct the likelihood function for SKS waveforms $p(\mathbf{d}_{\mathrm{SKS}}|\mathbf{m})$, as the vertical and radial components need simply be replaced by radial and transverse. The cross-convolution misfit function can also be used for incoming S waves, that is, S_P receiver functions, or transverse receiver functions, where the vertical component of a P waveform is deconvolved from its transverse component (see Table 1). In this way, we can integrate various data types in a consistent manner, with different sensitivities to the isotropic and anisotropic seismic structure beneath a station.

However, we acknowledge here that $p(\mathbf{r}|\mathbf{m})$ is not exactly a likelihood function *per se*, as it does not represent the probability distribution of data vectors $\mathbf{V}_{\mathrm{obs}}(t)$ and $\mathbf{H}_{\mathrm{obs}}(t)$, but rather the distribution of a vector of residuals conveniently defined. In a Bayesian framework, the vector of residuals is usually defined as a difference between observed data and predicted data: $\mathbf{r}(\mathbf{m}) = \mathbf{d}_{\mathrm{obs}} - \mathbf{d}_{\mathrm{est}}(\mathbf{m})$. In this case, the distribution of $\mathbf{r}$ for a given model $\mathbf{m}$ gives the distribution of the observed data ($p(\mathbf{r}|\mathbf{m}) = p(\mathbf{d}_{\mathrm{obs}}|\mathbf{m})$). However here, $p(\mathbf{r}|\mathbf{m})$ does not strictly represent the probability of observing the data, and hence cannot be strictly interpreted as a likelihood function. We note that this way of approximating the likelihood by the distribution of some residuals is also used by Stähler & Sigloch (2014), who proposed a Bayesian moment tensor inversion based on a cross-correlation misfit function. For a fully rigorous Bayesian approach to inversion of converted body waves, we refer the reader to Dettmer *et al.* (2015), who treated the source time function as an unknown in the problem.

2.6 Hierachical Bayes

The level of data errors for different data sets (σ_{C_0}, σ_{C_1}, σ_{C_2}, σ_p, σ_{SKS}, etc.) determines the width of the different Gaussian likelihood functions in (9), and hence the relative weight given to different data types in the inversion. Here, the level of noise also accounts for theoretical errors, that is, the part of the signal that we are not able to explain with our simplified 1-D parametrization and forward theory (Gouveia & Scales 1998; Duputel *et al.* 2014). For example, surface waves are sensitive to a larger volume around the station,

compared to higher frequency body waves arriving at the station with a near vertical incidence angle. Lateral inhomogeneities in the earth will then produce an incompatibility between these two types of observations, which here will be treated as data uncertainty.

In this work, we use a Hierarchical Bayes approach, and treat noise parameters as unknown in the inversion (Malinverno & Briggs 2004; Malinverno & Parker 2006). That is, each noise parameter is given a uniform prior distribution, and different values of noise (i.e. different weights) will be explored in the Monte Carlo parameter search. The range of possible noise parameters, that is, the width of the uniform prior distribution, is set large enough so that it does not affect final results (Bodin *et al.* 2012b). We then avoid the choice for arbitrary weights from the user, and the relative quantity of information brought by different data types is directly constrained by the data themselves.

2.7 The prior distribution

The Bayesian formulation enables one to account for prior knowledge, provided that this information can be expressed as a probability distribution $p(\mathbf{m})$ (Gouveia & Scales 1998). In a transdimensional case, the prior distribution prevents the algorithm from adding too many layers, as it naturally penalizes models with a large number of parameters $[l, k]$.

To illustrate this, let us look at the prior on the vector of isotropic velocity parameters $\mathbf{v}_s = [v_1, \ldots, v_k]$. We consider the velocity in each layer as *a priori* independent, that is, no smoothing constraint is applied, and then write:

$$p(\mathbf{v}_s \mid k) = \prod_{i=1}^{k} p(v_i). \tag{14}$$

For each parameter v_i, we use a uniform prior distribution over the range $[V_{\min} \quad V_{\max}]$. This uniform distribution integrates to one, and hence $p(v_i) = 1/\Delta V$, where $\Delta V = (V_{\max} - V_{\min})$. Therefore, for a given number of layers k we can write the prior on the vector $\mathbf{v}_s$ as:

$$p(\mathbf{v}_s \mid k) = \left(\frac{1}{\Delta V}\right)^k. \tag{15}$$

Here, the prior on velocity parameters decreases exponentially with k, and complex models with many layers are penalized. The complete mathematical form of our prior distribution including all model parameters is detailed in Appendix C.

In this way, the prior and likelihood distributions in our problem are in competition as complex models providing a good data fit (high likelihood) are simultaneously penalized with a low prior probability. This is an example of an implementation of the general principle of parsimony (or Occam's razor) that states that between two models (or theories) that predict the data equally well, the simplest should be preferred (see Malinverno 2002, for details). Although k is a free parameter that will be constrained by the data, the user still needs to choose the width of the prior distribution ΔV, which directly determines the volume of the model space, and hence the relative balance between the prior and the likelihood. The choice of ΔV therefore directly determines the number of layers in the solution models.

As expected, there is also a trade-off between the complexity of the model and the inferred value of data errors (σ_{C_0}, σ_{C_1}, σ_{C_2}, σ_p, σ_{SKS}, etc.). As the model complexity increases, the data can be better fit, and the inferred value of data errors decrease. However, this degree of trade-off is limited and the data clearly constrains the joint distribution of different parameters reasonably well (see Bodin *et al.* 2012b, for details).

2.8 Transdimensional sampling

Given the Bayesian framework described above, our goal is to generate a large number of 1-D profiles, the distribution of which approximates the posterior function. In our problem, the posterior distribution is defined in a space of variable dimension (i.e. transdimensional), and can be sampled with the reversible-jump Markov chain Monte Carlo (rj-McMC) sampler (Geyer & Møller 1994; Green 1995, 2003), which is a generalization of the well-known Metropolis–Hastings algorithm (Metropolis *et al.* 1953; Hastings 1970). A general review of transdimensional Markov chains is given by Sisson (2005).

The first use of these algorithms in the Geosciences was by Malinverno (2002) in the inversion of DC resistivity sounding data to infer 1-D depth profiles. Further applications of the rj-McMC have recently appeared in a variety of geophysical and geochemical data inference problems, including regression analysis (Gallagher *et al.* 2011; Bodin *et al.* 2012a; Choblet *et al.* 2014; Iaffaldano *et al.* 2014), geochemical mixing problems (Jasra *et al.* 2006), thermochronology (Stephenson *et al.* 2006; Fox *et al.* 2015b), geomorphology (Fox *et al.* 2015a), seismic tomography (Young *et al.* 2013a,b; Zulfakriza *et al.* 2014; Pilia *et al.* 2015), inversion of receiver functions (Piana Agostinetti & Malinverno 2010; Bodin *et al.* 2012b; Fontaine *et al.* 2015), geoacoustics (Dettmer *et al.* 2010, 2013; Dosso *et al.* 2014) and exploration geophysics (Malinverno & Leaney 2005; Ray *et al.* 2014). For an overview of the general methodology and its application to Earth Science problems, see also Sambridge *et al.* (2006), Gallagher *et al.* (2009) and Sambridge *et al.* (2013).

Here, we follow the implementation presented in Bodin *et al.* (2012b) for joint inversion of receiver functions and surface waves, but expand the parametrization to the case where a variable number of unknown parameters is associated to each layer, that is, where each layer can be either isotropic or anisotropic. In this section, we only briefly present the procedure, and give mathematical details of our particular implementation in Appendices C–E.

The algorithm produces a sequence of models, where each is a random perturbation of the last. The first sample is selected randomly (from the uniform distribution) and at each step, the perturbation is governed by the so-called proposal probability distribution which only depends on the current model. The procedure for a given iteration can be described as follows:

(i) Randomly perturb the current model $\mathbf{m}$, to produce a proposed model $\mathbf{m}'$, according to some chosen proposal distribution $q(\mathbf{m}'|\mathbf{m})$ (e.g. add/remove a layer, add/remove anisotropy to an existing layer, change the depth of a discontinuities, etc.). For details, see Appendix D.

(ii) Randomly accept or reject the proposed model (in terms of replacing the current model), according to the acceptance criterion ratio $\alpha(\mathbf{m}'|\mathbf{m})$. For details, see Appendix E.

Models generated by the chain are asymptotically distributed according to the posterior probability distribution (for a detailed proof, see Green 1995, 2003). If the algorithm is run long enough, these samples should then provide a good approximation of the posterior distribution for the model parameters, that is, $p(\mathbf{m}|\mathbf{d}_{\mathrm{obs}})$. This ensemble solution contains many models with variable parametrizations, and inference can be carried out by plotting the histogram of

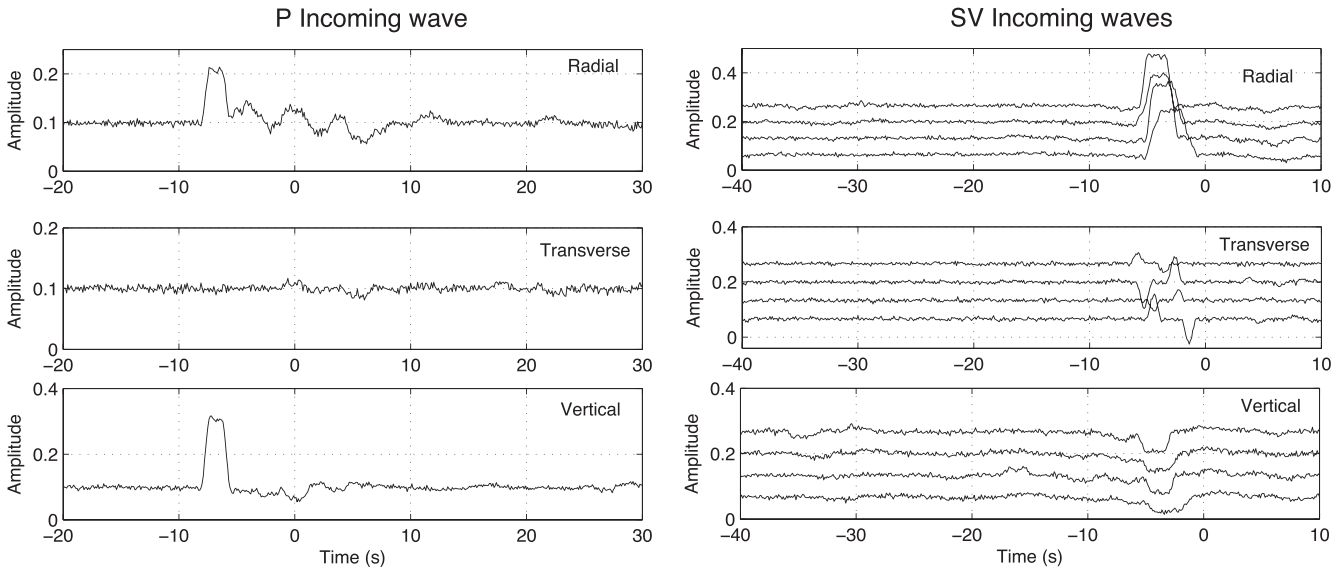

Figure 2. Synthetic body waves for the model shown in black in Fig. 3. Left: three component waveforms for an incoming *P* wave. Right: three component waveforms for four incoming *SV* waves arriving at different backazimuths (10°, 55°, 100°, 145°).

the parameter values (e.g. velocity at a given depth) in the ensemble solution.

3 SYNTHETIC TESTS

We first test our algorithm on synthetic data, and design an Earth model consisting of eight layers, among which only three are anisotropic (black line in Fig. 3). We use a reflectivity scheme (Levin & Park 1998) to propagate an incoming *P* wave, as well as four *SV* waves coming from different backazimuths (10°, 55°, 100°, 145°). There is only one *P* waveform here, and hence anisotropy will be constrained only from *S* waves in this experiment. Synthetic waveforms (Fig. 2) are created by convolving the Earth's impulse response (a Dirac comb), with a smoothed box car function. Then, some random Gaussian white noise is added to the waveforms. We acknowledge that these synthetic seismograms are far from being realistic, as for example observed *S* waves usually have a lower frequency content than *P* waveforms. The goal here is only to test the ability of the inversion procedure to integrate different data types. We also generate synthetic Rayleigh wave dispersion curves $\mathbf{C}_0(T)$, with 2ψ azimuthal terms $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$, for periods between 20 and 200 s, with added random noise (see Fig. 5).

The top panels of Fig. 3 show results when only Rayleigh wave dispersion measurements are inverted, that is, an ensemble of models distributed according to $p(\mathbf{m}|\mathbf{C}_0, \mathbf{C}_1, \mathbf{C}_2)$. Surface waves are long-period observations, and hence are only sensitive to the long-wavelength structure of the Earth. The sharp seismic discontinuities present in the true model (in black in Fig. 3A) cannot be resolved, and as expected, only a smooth averaged structure is recovered. In our method, there is no need for statistical tests or regularization procedures to choose the adequate model complexity or smoothness corresponding to a given degree of data uncertainty. Instead, the reversible jump technique automatically adjusts the underlying parametrization of the model to produce solutions with appropriate level of complexity to fit the data to statistically meaningful levels. This probabilistic scheme therefore allows us to quantify uncertainties in the solution, and level of constraints. For example, we observe that the direction of anisotropy in Fig. 3C is clearly better resolved than its amplitude in Fig. 3B.

Bottom panels of Fig. 3 show results for a joint inversion of surface waves and body waves. For body waves, we jointly invert four data types: P_{SV}, P_{SH}, S_P and SKS waveforms, given by all pairs of components described in Table 1. Here, both discontinuities and amplitude of anisotropy are better resolved, due to the complementary information brought by body waves, although we acknowledge that the distribution for the direction of anisotropy becomes bimodal below 250 km, certainly due to the lack of resolution at these depths.

Our Monte Carlo sampling of the model space allows us to treat the problem in a fully non-linear fashion (although we acknowledge that the function linking the model to $\mathbf{C}_1(T)$, and $\mathbf{C}_1(T)$ has been linearized around the isotropic average of the model). Contrary to linear or linearized inversions, here the solution is not simply described by a Gaussian posterior probability function, and can be multimodal. We illustrate this in Fig. 4 by showing the full distribution for V_s, δV_s and Ψ_{fast} at 150 km depth. The posterior distribution is shown in grey and the true model in red. This shows how adding body waves reduces the width of the posterior distribution as more information is added. Note that the distribution of the direction of anisotropy is multimodal, with two secondary peaks corresponding to directions of other anisotropic layers in the model (green and blue lines). We acknowledge that a multimodal distribution is hard to interpret, as in this case the mean and standard deviation of the distribution are meaningless.

Since the misfit function in eq. (12) is not a simple difference between observed and estimated data, it is difficult to get a visual idea of the level of data fit. Instead, in Fig. 5 we show the two terms of the misfit function, that is, $\mathbf{v}_p(t, \mathbf{m}) * \mathbf{H}(t)$ and $\mathbf{h}_p(t, \mathbf{m}) * \mathbf{V}(t)$ for the best-fitting model $\mathbf{m}$ in the ensemble solution. Although these two waveforms do not have any intuitive physical meaning, the misfit function has a minimum when these two vectors are equal, and plotting them together helps give a visual impression for the level of fit. Right-hand panels of Fig. 5 show observed and best-fitting data for surface wave observations $\mathbf{C}_0(T)$, $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$.

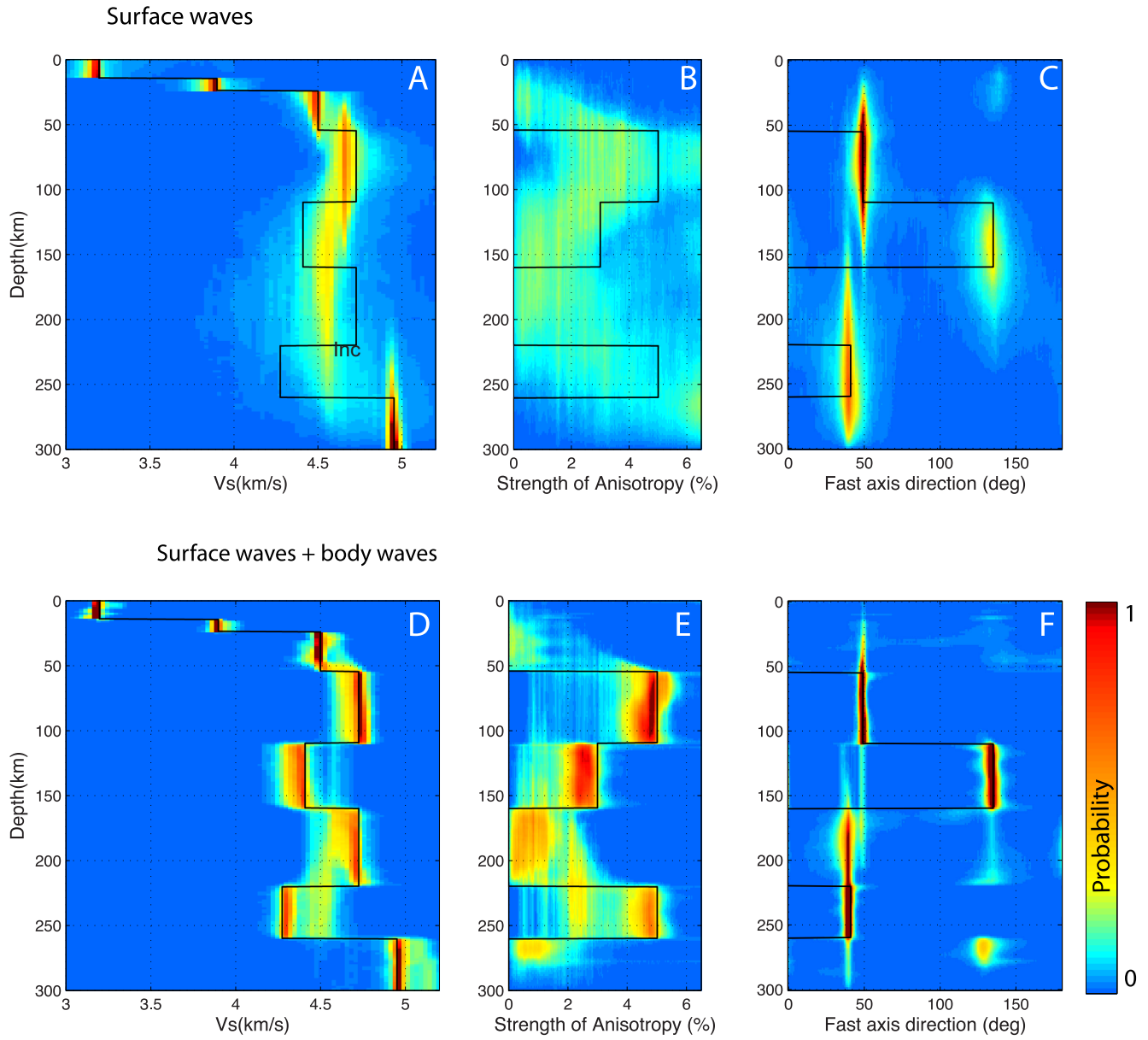

Figure 3. Transdimensional inversion of synthetic data shown in Fig. 2. Density plots show the probability of the model given the data for our three unknown parameters: V_s (left), δV_s (middle) and Ψ_{fast} (right). Black lines show the true model used to create noisy synthetic data. Top: inversion of surface wave dispersion only. Bottom: joint inversion of surface waves and body waves (i.e. P_{SV}, P_{SH}, S_P and SKS waveforms).

4 APPLICATION TO TWO DIFFERENT TECTONIC REGIONS IN NORTH AMERICA

We apply this method to seismic observations recorded at two different locations in North America. First, we invert data from station FFC (Canada), which is a permanent, reliable and well-studied station located at the core of the Slave Craton. Since a large number of studies have already been published about the structure under this station (e.g. Ramesh *et al.* 2002; Rychert & Shearer 2009; Miller & Eaton 2010; Yuan & Romanowicz 2010b), we view this as an opportunity to test and validate the proposed scheme.

In a second step, we invert seismic data recorded in Arizona at station TA-214A, of the US transportable array, which is a much noisier, recent, and less studied station, located in the southern Basin and Range Province, close to a diffuse plate boundary, where we expect more complex 3-D structure due to recent tectonic activity. Here, 3-D effects in our data would not be able to be accounted for by our 1-D model, and hence will be treated as data errors by our Bayesian scheme. The goal is to see how our inversion performs in a more difficult setting. The final results are summarized in Fig. 11, where velocity gradients observed under the two stations are interpreted in terms of well-known upper-mantle seismic discontinuities.

4.1 The North American craton

4.1.1 Tectonic setting

The North American craton comprises the stable portion of the continent, and differs from the more tectonically active Basin and

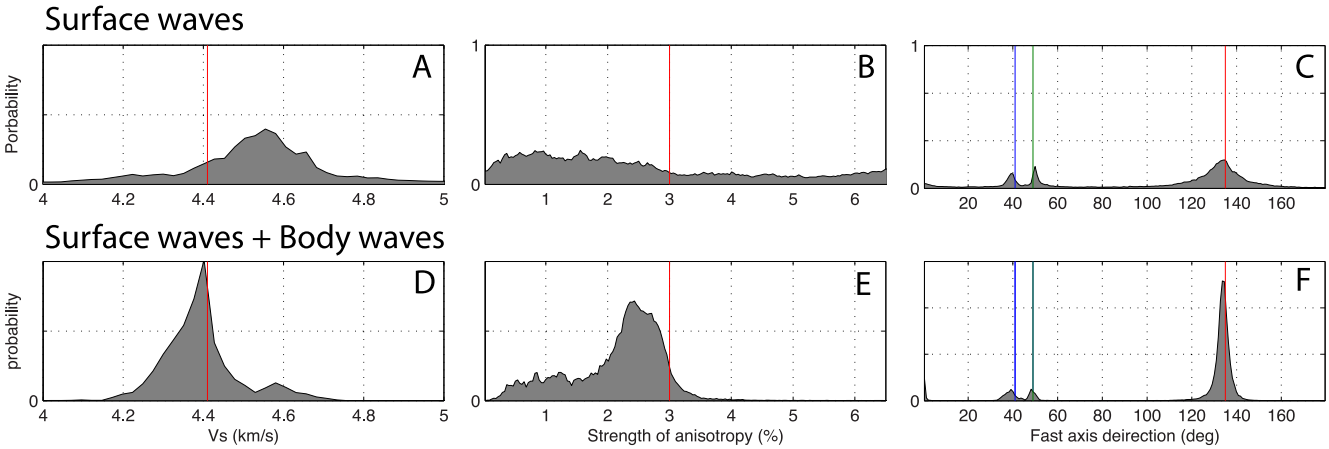

Figure 4. Synthetic test. Posterior marginal distribution for V_s (left), δV_s (middle) and Ψ_{fast} (right) at 150 km depth. Those are simply cross-sections of the density plots showed in Fig. 3. Red lines show the true model. In panels C and F, green and blue lines show the direction of anisotropy for the first and third layer in the true model.

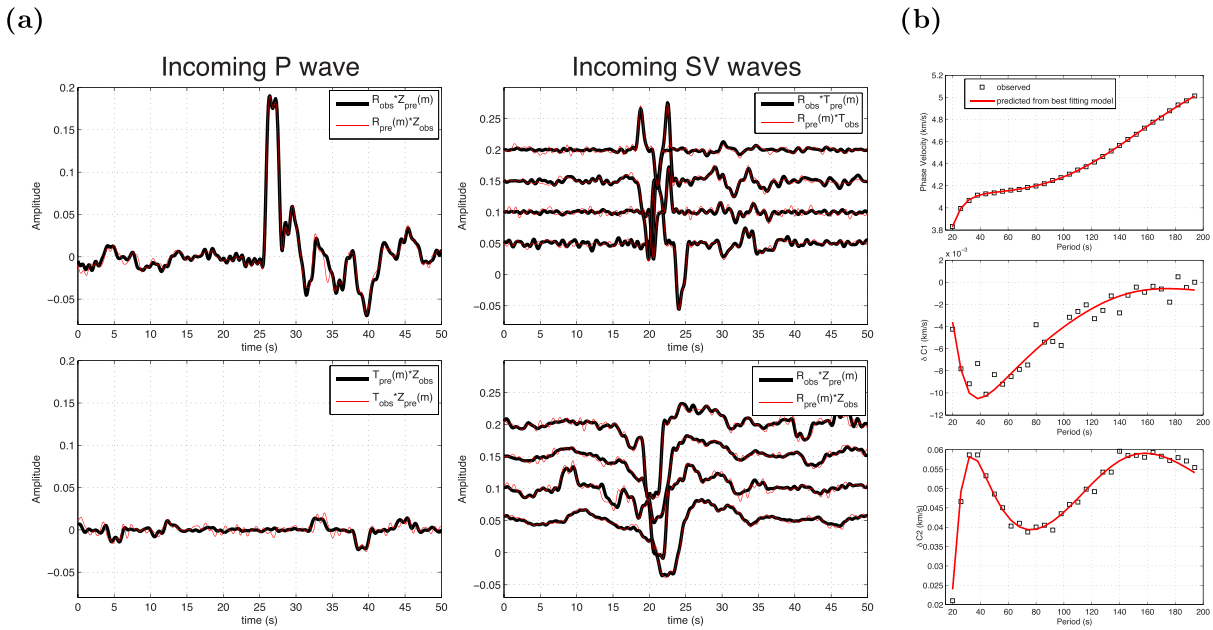

Figure 5. Synthetic data experiment. (a) Fit obtained by the cross-convolution modelling for the best-fitting model in the ensemble solution. (b) Fit to Rayleigh wave dispersion data for the best-fitting model.

Range province to the west. In general, cratonic regions represent areas of long-lived stability within the lithosphere that have remained compositionally unchanged, and have resisted destruction through subduction since as early as the Archean. Previous work in this region reveals anomalously high seismic velocities in the upper mantle. Numerous seismic tomography studies detect the base of the lithosphere at a depth between 150 and 300 km throughout the stable craton (e.g. Gung *et al.* 2003; Kustowski *et al.* 2008; Nettles & Dziewoński 2008; Romanowicz 2009; Ritsema *et al.* 2011; Pasyanos *et al.* 2014; Schaeffer & Lebedev 2014), but most receiver function studies fail to detect a corresponding drop in velocity at this depth.

Receiver functions studies do show, however, a decrease in velocity within the cratonic lithosphere, suggesting the potential existence of an intralithospheric discontinuity in this region (Abt *et al.* 2010; Miller & Eaton 2010; Kind *et al.* 2012; Hansen *et al.* 2015; Hopper & Fischer 2015). For recent reviews on studies of the mid-lithospheric discontinuity (MLD), see Rader *et al.* (2015), Karato *et al.* (2015) and Selway *et al.* (2015). Evidence for anisotropic layering within the cratonic lithosphere has also been previously shown (Yuan & Romanowicz 2010b; Wirth & Long 2014; Long *et al.* 2016).

The exact nature of the layered structure and composition of cratons, however, remains poorly understood. Competing hypotheses based on geochemical and petrologic constraints describe possible models for craton formation; these include underplating by hot mantle plumes and accretion by shallow subduction zones in continental or arc settings (Arndt *et al.* 2009).

4.1.2 The data

For *Ps* converted waveforms, we selected two regions with high seismicity (Aleutian islands and Guatemala) each defined by a small backazimuth and distance range (Fig. 6). For both regions, we

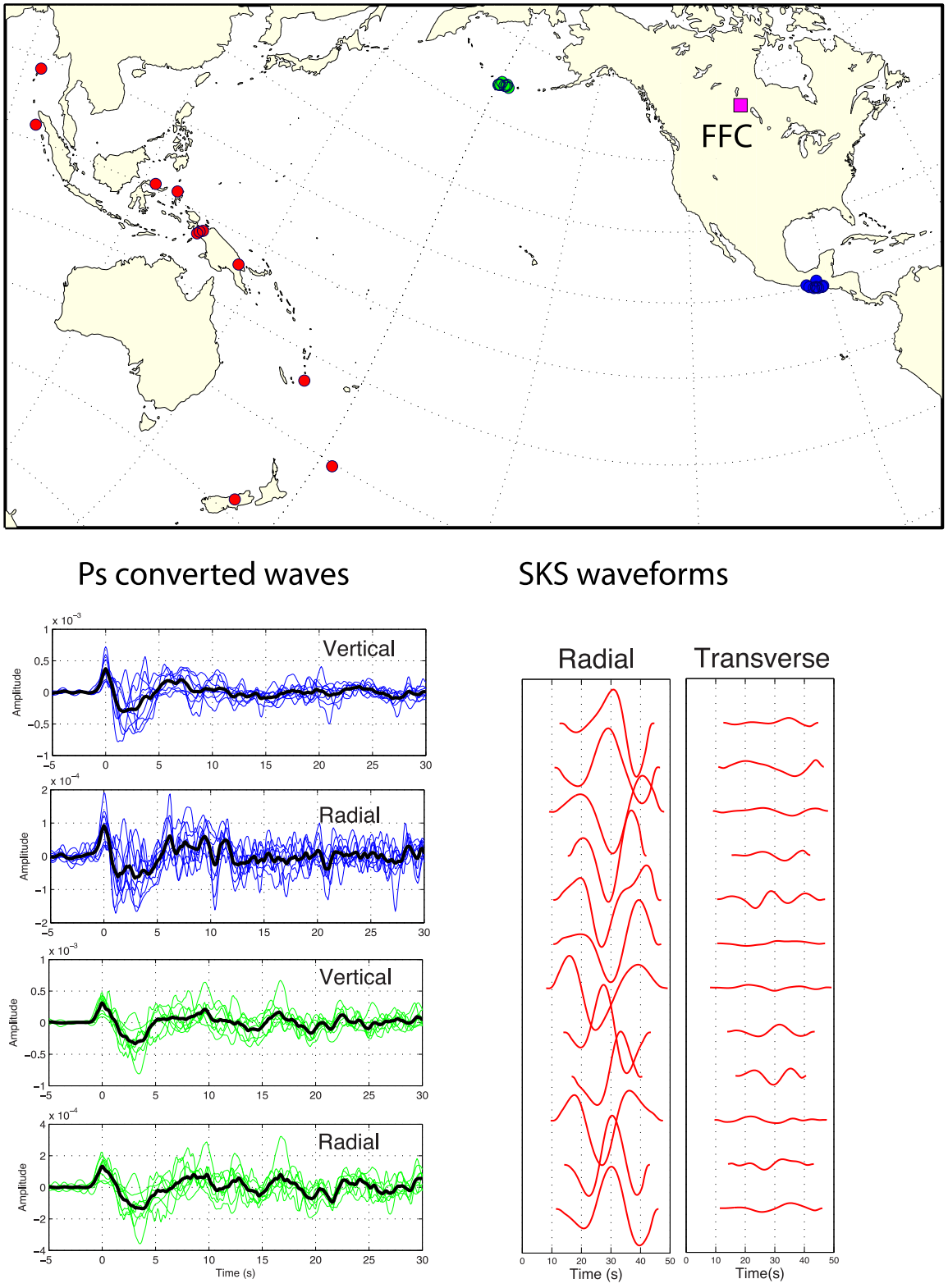

Figure 6. Body wave observations made at station FFC, Canada. For *P* waves, vertical and horizontal components are stacked over a set of events, at two different locations (blue and green). For SKS data, the waveform of 12 individual events are used (red). SKS waveforms are normalized to unit energy, and there is no amplitude information in the lower right-hand panel.

computed stacks of seismograms following the approach of Kumar *et al.* (2010) and described in Bodin *et al.* (2014). Waveforms of first *P* arrival are normalized to unit energy, aligned to maximum amplitude, and sign reversal is applied when the *P* arrival amplitude is negative. Moveout corrections are not needed here as stacked events have similar ray parameters. Both regions provide a pair of $\mathbf{V}_{\text{obs}}$ and $\mathbf{H}_{\text{obs}}$ stacked waveforms. Since we only use two backazimuths, receiver functions will not bring a lot of information about azimuthal anisotropy, which rather will be constrained from Rayleigh waves and SKS waveforms.

For shear wave splitting measurements, a number of individual SKS waveforms have been selected at different backazimuths (red circles in Fig. 6). The waveforms were manually picked based on small signal–noise ratio and large energy split onto the transverse component.

We also used fundamental mode Rayleigh wave phase velocity measurements (25–150 s) given by Ekström (2011) at this location. We recognize that these measurements are the result of a global tomographic inversion, and hence are not free from artefacts due to regularization and linearization. Better measurements could be obtained from local records obtained at small aperture arrays (e.g. Pedersen *et al.* 2006).

4.1.3 Results at station FFC

In Fig. 7, we show results obtained after three types of inversions with different data types at station FFC. The prior distribution is defined as a uniform distribution around a reference model consisting of a two-layered crust above a half-space. The structure of the crust is given by H–κ stacking method of receiver functions measured at this station (Zhu & Kanamori 2000).

Top panels show results for inversion of Rayleigh waves alone. The distribution of shear wave velocity shows a low-velocity zone in the range 150–300 km with no clear boundaries, as observed in some long-period tomographic models. The fast axis direction of azimuthal anisotropy is varying with depth, but again with no clear discontinuities.

Middle panels in Fig. 7 show results when receiver functions are added to the inversion. In this case, the isotropic velocity profile reaches very high values (4.9 km s^{-1}) in the upper part of the lithosphere, between 100 and 150 km depth, compatible with results from full waveform tomography (Yuan & Romanowicz 2010b). These high values are also observed in the Australian craton from multimode surface wave tomography (Yoshizawa & Kennett 2015). A sharp negative velocity jump appears at 150 km (also observed by Miller & Eaton 2010), that we shall interpret as an MLD, and defines the top of a low-velocity zone within the lithosphere between 150 and 180 km depth, as described by Thybo & Perchuć (1997) and Lekić & Romanowicz (2011). The sharp MLD can be interpreted in different ways, and a number of models can be invoked such as different hydration and melt effects, or metasomatism (Foster *et al.* 2014; Karato *et al.* 2015). At 250 km depth, we observe a small negative velocity drop, associated with a strong gradient in the direction of fast axis of anisotropy, going from 15° to 90°.

When SKS waveforms are added to the data set (lower panels in Fig. 7), the fast axis direction below 250 km reduces to 55°, and becomes aligned with the absolute motion of the North American plate in the hotspot reference frame (Gripp & Gordon 2002). This has two strong implications: (1) it demonstrates the sensitivity of SKS observations to structure below 250 km, poorly constrained by fundamental mode surface waves and (2) this allows us to interpret the discontinuity at 250 km as the lithopshere–asthenosphere boundary (LAB), below which the anisotropy would result from present day mantle flow associated with the motion of the North America plate. Above the LAB, the anisotropy in the lithosphere would be 'frozen-in' and related to past tectonic processes. This interpretation is depicted in the upper panels of Fig. 11. Fig. 8 shows the data fit for the best-fitting model in the ensemble solution. Overall, the results obtained here are quite compatible with the 3-D model from Yuan & Romanowicz (2010b) obtained by combining SKS splitting parameters and full waveform tomography.

4.2 The Southern Basin and Range

We apply now the method to station 214A of the Transportable Array, located in the South West, close to Organ Pipe National monument in Arizona, at the Mexican border.

4.2.1 Tectonic setting

The station is located at the northern end of the California Gulf extensional province, a dynamic boundary plate system. Here, the structure of the crust and upper-mantle results from the complex tectonic interaction between the Pacific, Farallon and North America plates. This region has been affected by a number of different major tectonic processes, such as the cessation of subduction, continental breakup and early stage of rifting (Obrebski & Castro 2008). An extensive review of the geology of the whole Golf of California region is given by Sedlock (2003).

At the regional scale, the station is located in the southern Basin and Range province, which underwent Cenozoic extensional deformation. Although an extension in the EW direction has been widely observed in numerous studies, the causes of the extension of the Basin and Range province are diverse and still debated: NA-Pacific plate interaction along San Andreas fault, gravity collapse of overthickened crust in early orogens, or in response to some upper-mantle upwelling (see Dickinson 2002, for a review).

From a seismological point of view, the site is at the southeastern corner of the intriguing 'circular' SKS pattern observed in the western US (Savage & Sheehan 2000; Liu 2009; Eakin *et al.* 2010; Yuan & Romanowicz 2010a).

4.2.2 The data

We use similar observations to those collected for station FFC. For *P*s converted waveforms, we averaged seismograms for a number of events from Japan and northern Chile (see Fig. 9). We also invert a single SKS waveform (red dot) and five SKKS waveforms. Rayleigh wave dispersion curves are extracted from global phase velocity tomographic maps given by Ekström (2011) at this location.

4.2.3 Results at station TA-214A

Results for station TA-214A are shown in Fig. 10 and a final interpretation is given in Fig. 11. The prior distribution is defined as a uniform distribution around a reference model consisting of a crust above a half-space. When only surface waves are inverted (top panels), a clear asthenospheric low-velocity zone is visible with a peak minimum at 120 km depth. As expected, no discontinuities

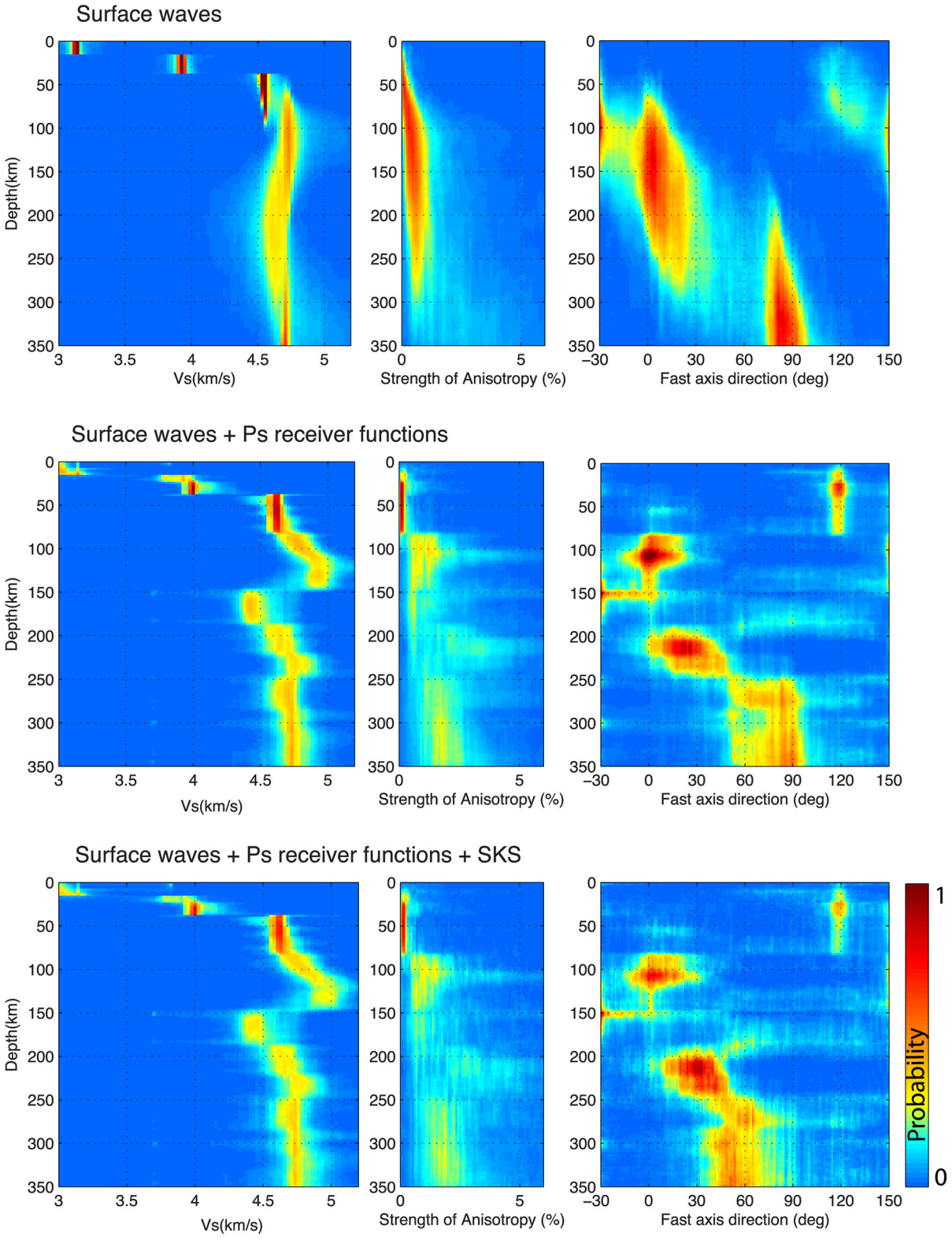

Figure 7. Inversion results at station FFC, located in the North American Craton. Density plots represent the ensemble of models sampled by the reversible jump algorithm, and represent the posterior probability function. The number of layers in individual models was allowed to vary between 2 and 60. For lower plots, the maximum of the posterior distribution on the number of layers is 41.

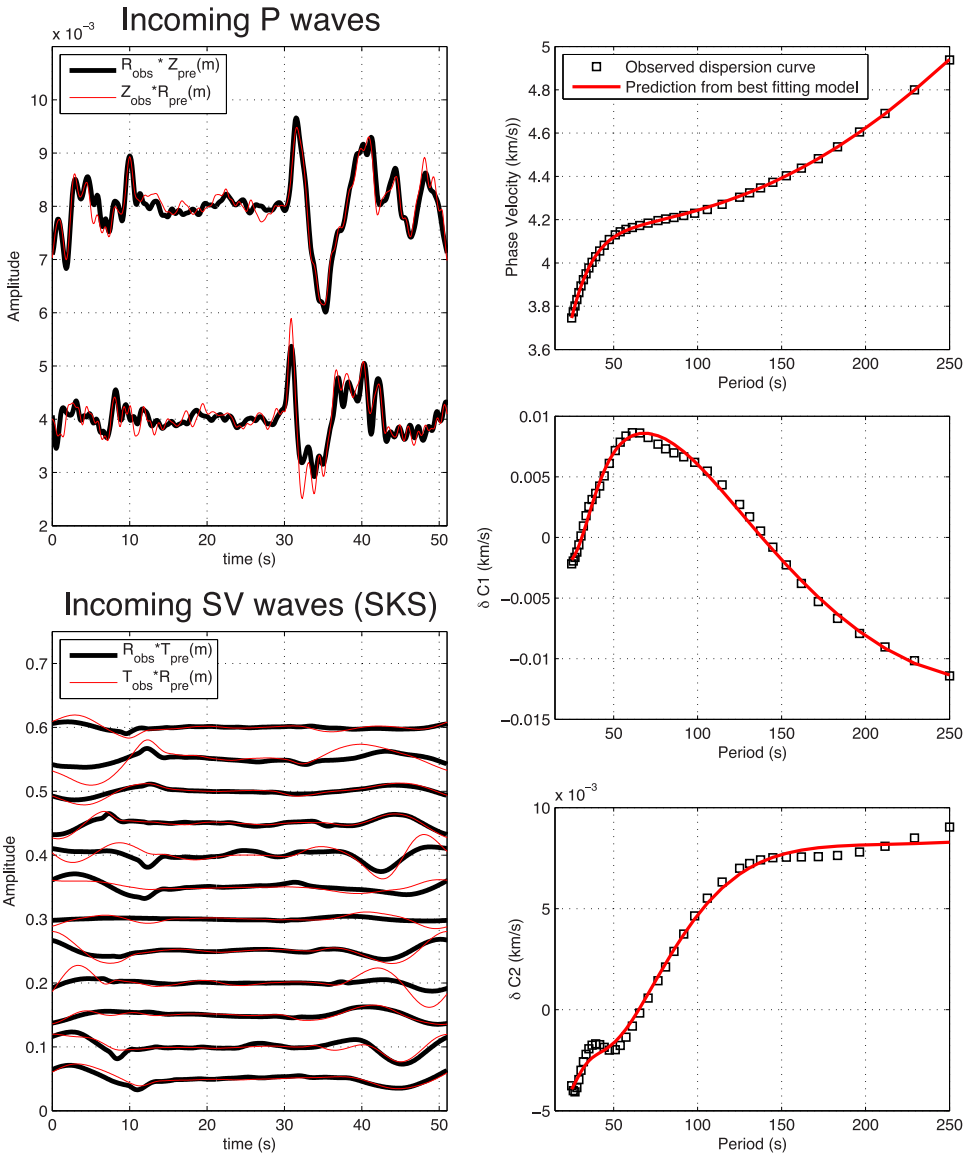

Figure 8. Station FFC (Canada). Data fit for best-fitting model collected by the Monte Carlo sampler. For body waves (left-hand panels), the cross-convolution misfit function is not constructed as a difference between observed and estimated data. Instead, we plot the two vectors $\mathbf{H}*\mathbf{v}(m)$ and $\mathbf{V}*\mathbf{h}(m)$, which difference we try to minimize.

in the upper mantle are visible, due to the lack of resolution of surface waves. Middle panels in Fig. 10 are obtained after adding converted *P* waves as constraints. As previously, seismic discontinuities are introduced. The bottom panels show results with SKS data, providing deeper constrains on anisotropy, below 200 km. Fig. 12 shows the data fit for the best-fitting model in the ensemble solution.

A clear negative discontinuity in *Vs* is visible at 100 km depth with a positive jump at 150 km, thus producing a 50 km thick low-velocity zone that could be interpreted as the asthenosphere. In this case, the shallow LAB at 100 km is compatible with a number of S_P receiver functions studies in the region (Levander & Miller 2012; Lekić & Fischer 2014). This low-velocity zone lying under a higher velocity 100 km thick lithospheric lid has been also observed in the shear wave tomographic model of Obrebski *et al.* (2011). Here, the sharp LAB discontinuity cannot be solely explained by a thermal gradient, and hence suggests the presence of partial melt in the asthenosphere in this region as proposed by Gao *et al.* (2004), Schmandt & Humphreys (2010) and Rau & Forsyth (2011).

The vertical distribution of fast axis direction (lower right-hand panel in Fig. 10) clearly shows three distinct domains:

(i) The lithospheric extension of the Basin and Range in the east–west direction (90°) is visible in the first 100 km. This direction of anisotropy in this depth range is also observed in surface wave (Zhang *et al.* 2007) or full waveform (Yuan *et al.* 2011) tomographic models. This E-W direction of fast axis is close to being perpendicular to the North America–Pacific plate boundary, and corresponds to the direction of opening of the Gulf of California; it is also similar to the direction of past subduction (Obrebski *et al.* 2006). We also note that the direction of fast axis in the lithosphere is gradually shifting to the North America absolute plate motion direction when approaching 100 km depth (75° in the hotspot frame).

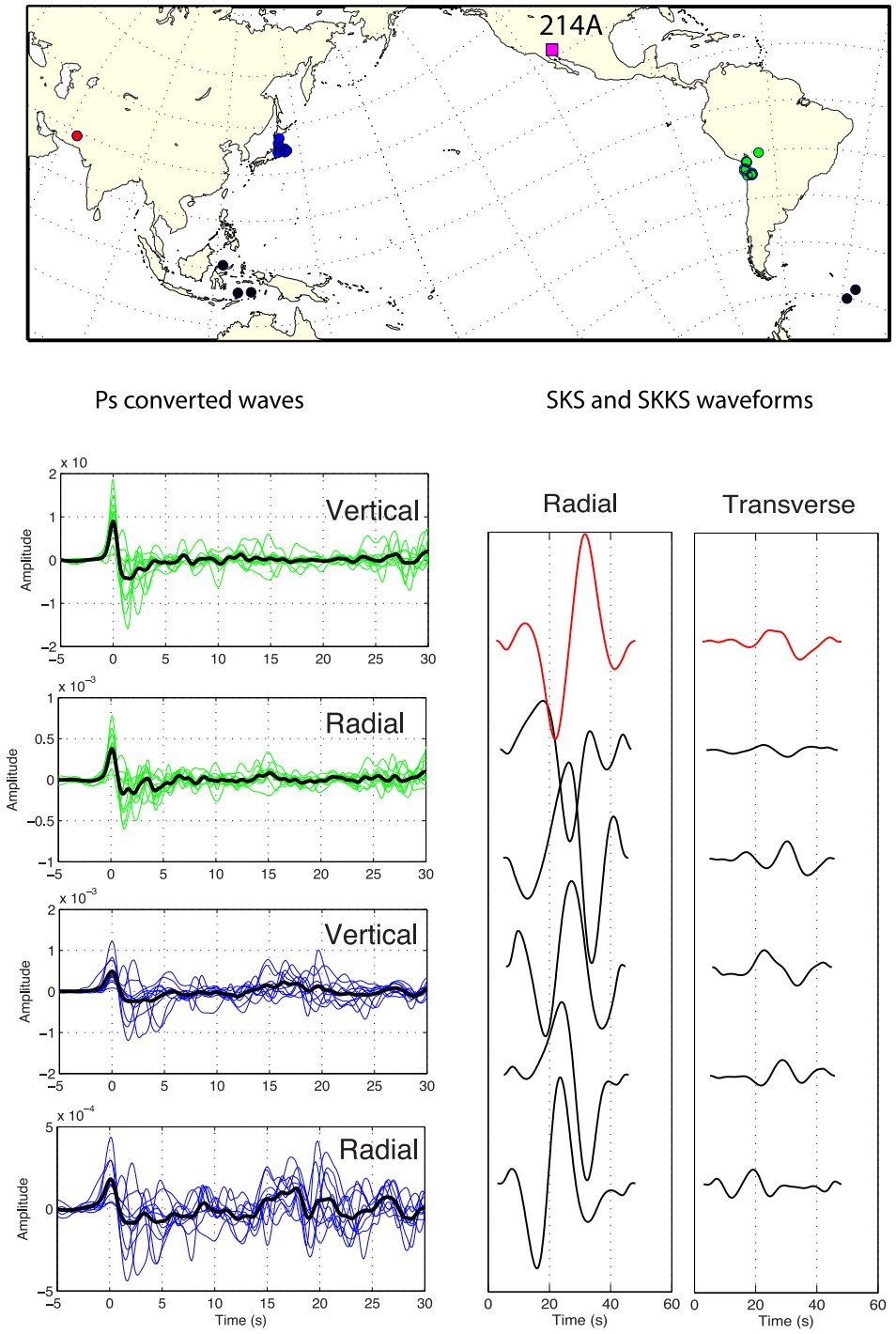

Figure 9. Body wave observations used for the 1-D inversion under station 214A, located in the Basin and Range province. We use two stacks of *P* wave seismograms, from Japan (blue) and North Chile (green), as well as five SKKS individual waveforms (black) and one SKS waveform (red). SKKS and SKS waveforms are normalized to unit energy, and there is no amplitude information in the lower right-hand panel.

(ii) There is a sharp change of direction of anisotropy at 100 km depth, which confirms the interpretation of the negative discontinuity as the LAB. A distinct layer between 100 and 180 km is clearly visible with a direction of 150°, that is, parallel to the absolute plate motion of the Pacific Plate, and in agreement with tomographic inversions combining surface waveforms and SKS splitting data (Yuan & Romanowicz 2010b). Also in agreement with the latter study, anisotropy strength decreases beneath 200 km depth. This direction is also compatible with shear wave splitting observations obtained in the Mexican side of the southern Basin and Range province (Obrebski *et al.* 2006). Interestingly, this Pacific APM parallel direction continues down to 180–200 km, that is, a bit below the bottom of the low-velocity zone as defined from the isotropic plot.

(iii) The jump at 180–200 km in the direction of anisotropy to about 60° is a very interesting feature which seems associated with

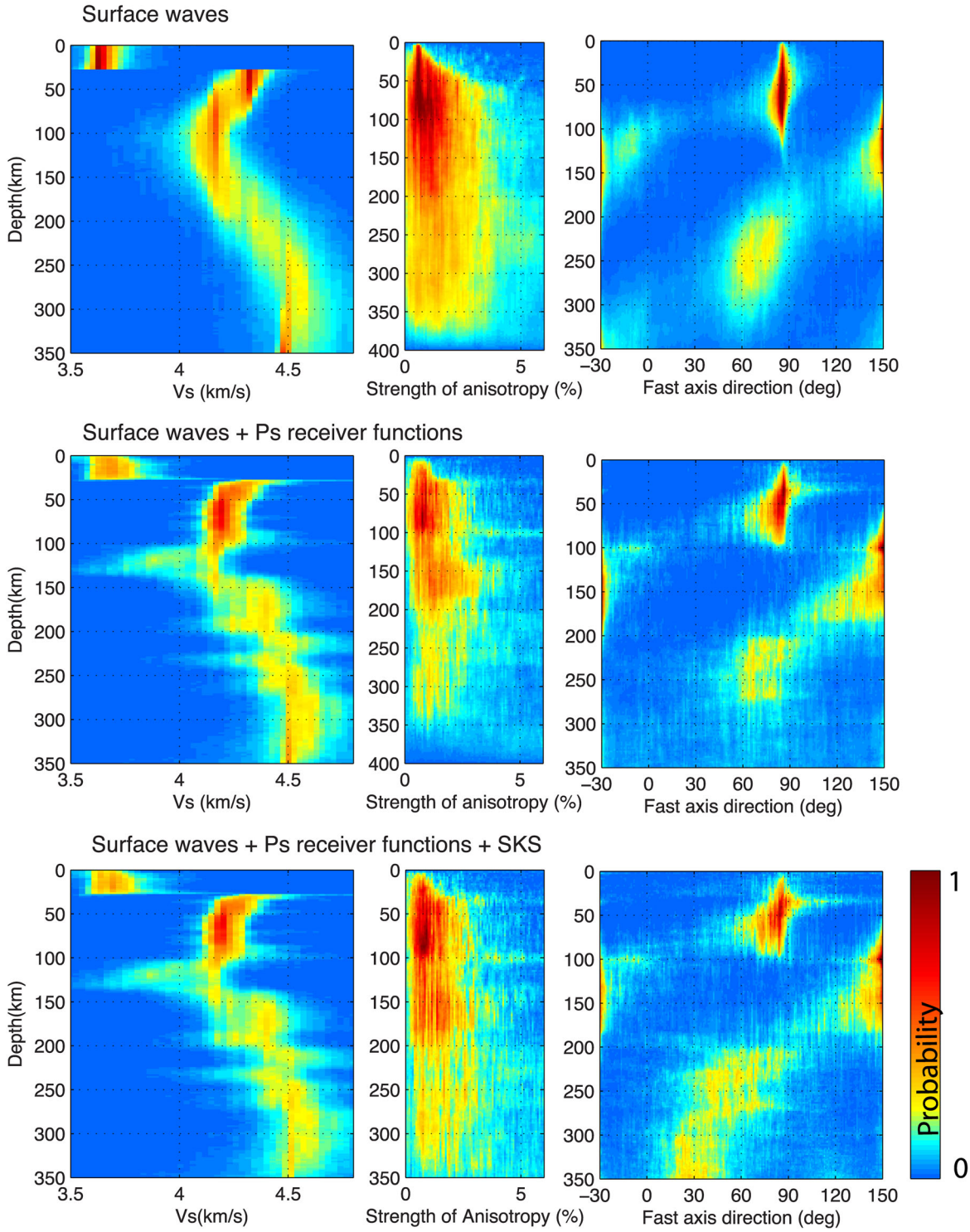

Figure 10. Inversion results at station TA-214A, located in the southern Basin and Range Province. Density plots represent the ensemble of models sampled by the reversible jump algorithm, and represent the posterior probability function. The number of layers in individual models was allowed to vary between 2 and 80. For lower plots, the maximum of the posterior distribution on the number of layers is 55.

a positive step in the velocity, and could be the 'Lehmann' discontinuity (Gung *et al.* 2003). In that case, either Lehmann is not the base of the asthenosphere, or the asthenosphere extends to $\sim$ 200 km, but consists of two levels. The 60° direction between 200 and 350 km might reflect some secondary scale convection/dynamics in this depth range. However, the anisotropy signal is much weaker, or more diffuse below 200 km, and one should not over interpret results at these depths.

As expected, here the structure is clearly less well resolved than for station FFC, and in particular the amplitude and direction of anisotropy below 200 km. This may be due to higher noise levels at

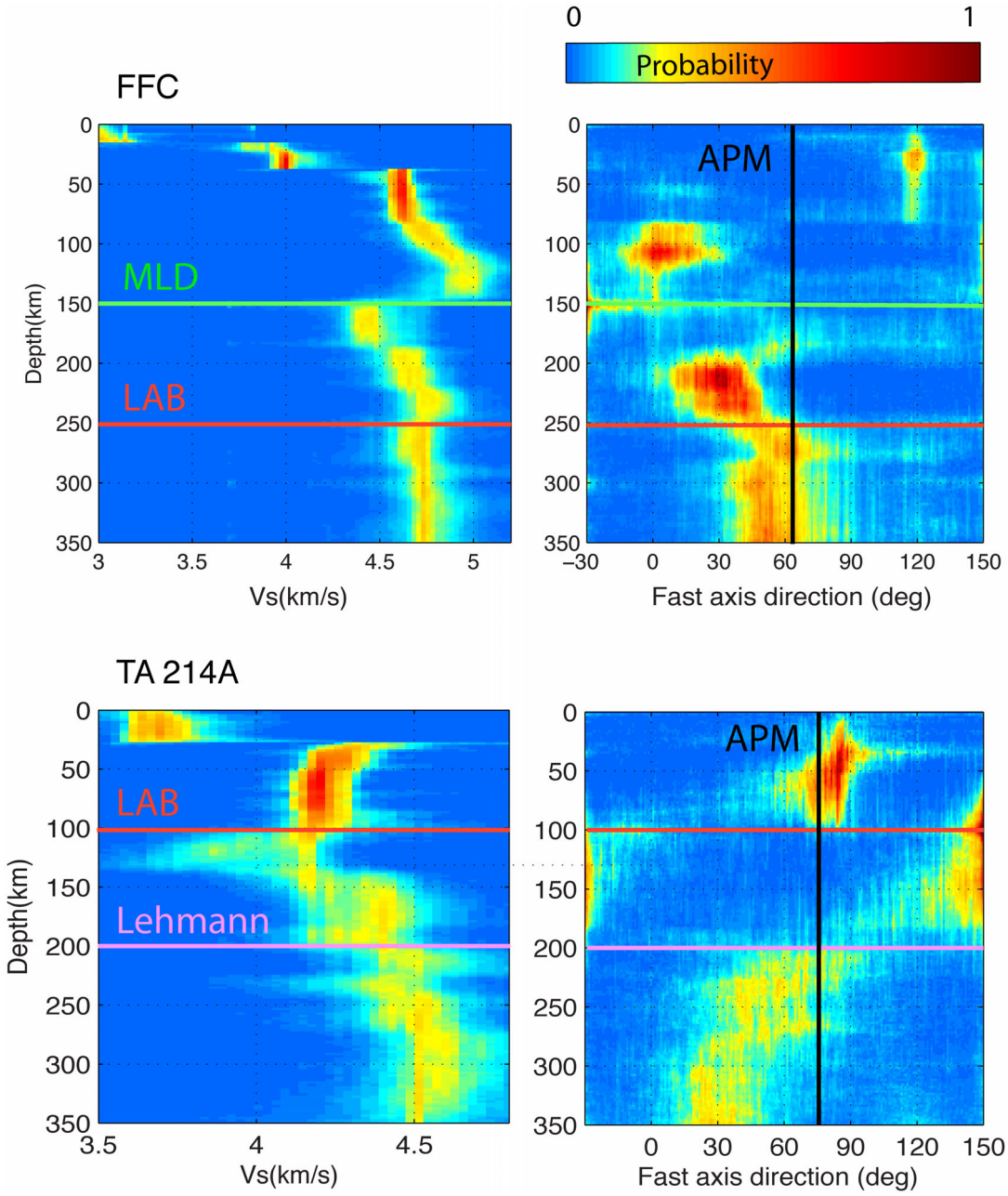

Figure 11. Interpretation of results for both stations. Top: results at FFC (North American Craton) for joint inversion of surface waves, *P* and SKS waveforms. Bottom: results at TA-214A for joint inversion of surface waves, *P* and SKS waveforms. Vertical black lines represent the direction of the absolute motion of the North American plate in the hotspot reference frame (Gripp & Gordon 2002).

this temporary station, or because the structure is more complex and the 1-D assumption less appropriate. In a complex 3-D setting, the fact that the data see different volumes results in incompatibilities, and here in wider posterior distributions.

5 CONCLUSIONS

We have presented a 1-D Bayesian Monte Carlo approach to constrain depth variations in azimuthal anisotropy, by simultaneously inverting body and Rayleigh wave phase velocity measurements observed at individual stations. We use a flexible parametrization where the number of layers, as well as the presence or absence of anisotropy in each layer, are treated as unknown parameters, and are directly constrained by the data. This adaptive parametrization turns out to be particularly useful, as the different types of data involved are sensitive to different volumes and length scales in the Earth. The level of noise in each data type (i.e. the required level of fit) is also treated as an unknown to be inferred by the data. In this manner, both observational and theoretical data errors (effect of 3-D structure and dipping layers) are accounted for in the inversion, without need to choose weights to balance different data sets.

For the first time, azimuthal variations of dispersion curves were jointly inverted with receiver functions and SKS data, for both crust and upper-mantle structure. The procedure was applied to

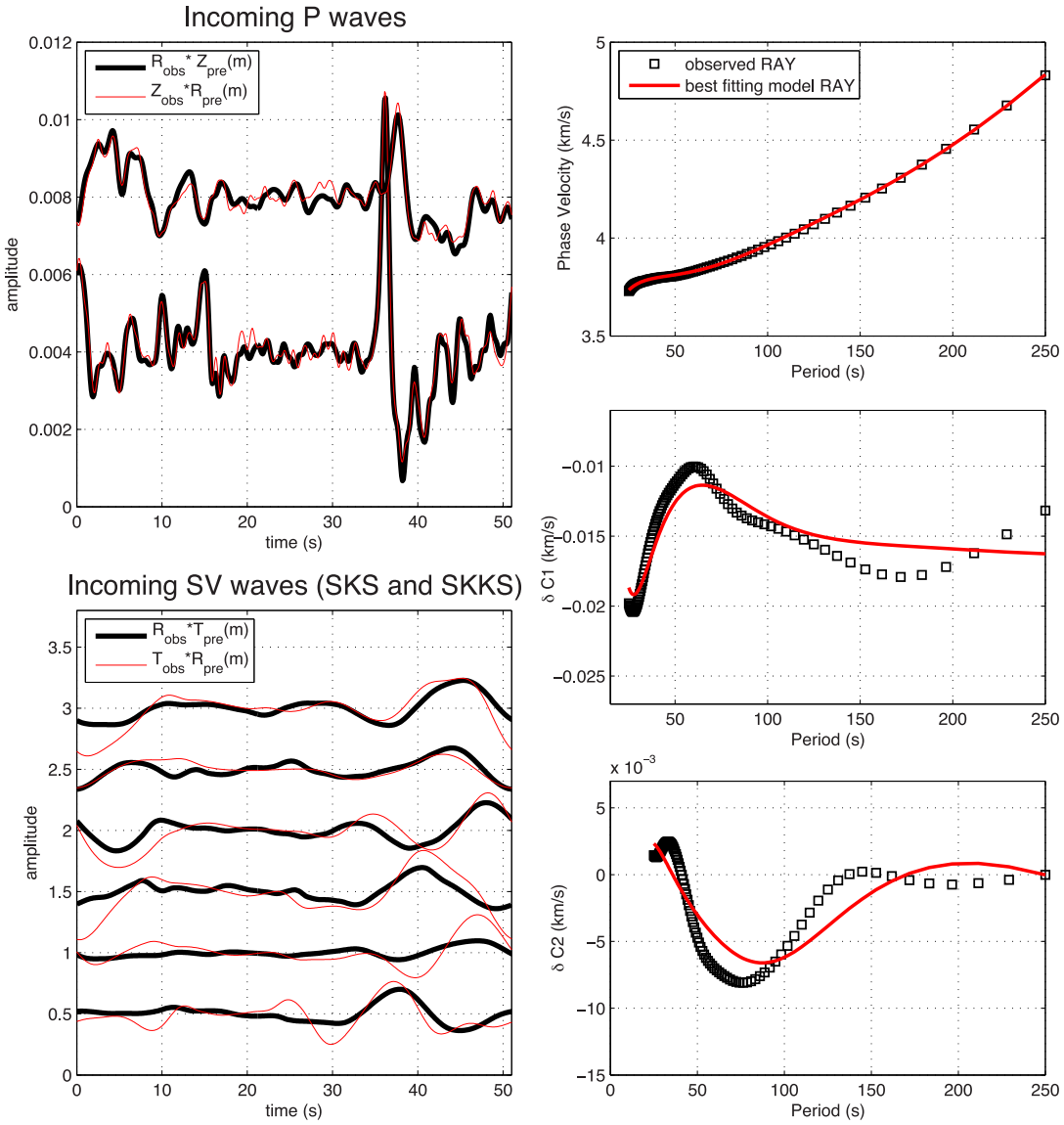

Figure 12. Station TA-214A (Arizona). Data fit for best-fitting model collected by the Monte Carlo sampler. For body waves (left-hand panels), the cross-convolution misfit function is not constructed as a difference between observed and estimated data. Instead, we plot the two vectors $\mathbf{H}*\mathbf{v}(m)$ and $\mathbf{V}*\mathbf{h}(m)$, the difference between which we try to minimize.

data recorded at two different stations in North America, in two different tectonic regimes. In both cases, results are compatible with previous studies, and allow us to better image anisotropic layering. In both cases, we observed a LAB characterized by both isotropic and anisotropic sharp discontinuities in the mantle, thus implying that the LAB cannot be defined as a simple thermal transition, but also reflects changes in composition and rheology.

ACKNOWLEDGEMENTS

TB wishes to acknowledge support from the Miller Institute for Basic Research at the University of California, Berkeley. This work was partially supported by a Labfees research collaborative grant from the U.C.O.P. (12-LR-236345) and by NSF Earthscope grant EAR-1460205.

REFERENCES

Abt, D.L., Fischer, K.M., French, S.W., Ford, H.A., Yuan, H. & Romanowicz, B., 2010. North American lithospheric discontinuity structure imaged by *Ps* and *Sp* receiver functions, *J. geophys. Res.,* **115,** B09301, doi:10.1029/2009JB006914.

Adam, J.M.-C. & Lebedev, S., 2012. Azimuthal anisotropy beneath Southern Africa from very broad-band surface-wave dispersion measurements, *Geophys. J. Int.,* **191**(1), 155–174.

Arndt, N., Coltice, N., Helmstaedt, H. & Gregoire, M., 2009. Origin of archean subcontinental lithospheric mantle: some petrological constraints, *Lithos,* **109**(1), 61–71.

Audet, P., 2015. Layered crustal anisotropy around the San Andreas Fault near Parkfield, California: crustal anisotropy around San Andreas, *J. geophys. Res.,* **120,** 3527–3543.

Backus, G.E., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.,* **67**(11), 4427–4440.

Becker, T.W., Lebedev, S. & Long, M.D., 2012. On the relationship between azimuthal anisotropy from shear wave splitting and surface wave tomography, *J. geophys. Res.,* **117**(B1), B01306, doi:10.1029/2011JB008705.

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., Trampert, J. & van Heijst, H.J., 2006. Radial anisotropy in seismic reference models of the mantle, *J. geophys. Res.,* **111,** B02303, doi:10.1029/2005JB003728.

Beghein, C., Snoke, J.A. & Fouch, M.J., 2010. Depth constraints on azimuthal anisotropy in the Great Basin from Rayleigh-wave phase velocity maps, *Earth planet. Sci. Lett.,* **289**(3), 467–478.

Beucler, É. & Montagner, J.-P., 2006. Computation of large anisotropic seismic heterogeneities (CLASH), *Geophys. J. Int.,* **165**(2), 447–468.

Bianchi, I., Park, J., Agostinetti, N.P. & Levin, V., 2010. Mapping seismic anisotropy using harmonic decomposition of receiver functions: an application to Northern Apennines, Italy, *J. geophys. Res.,* **115**(B12), B12317, doi:10.1029/2009JB007061.

Bianchi, I., Bokelmann, G. & Shiomi, K., 2015. Crustal anisotropy across northern Japan from receiver functions, *J. geophys. Res.,* **120**(7), 4998–5012.

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

Bodin, T., Salmon, M., Kennett, B.L.N. & Sambridge, M., 2012a. Probabilistic surface reconstruction from multiple data sets: an example for the Australian Moho, *J. geophys. Res.,* **117**(B10), B10307, doi: 10.1029/2012JB009547.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012b. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117,** B02301, doi:10.1029/2011JB008560.

Bodin, T., Yuan, H. & Romanowicz, B., 2014. Inversion of receiver functions without deconvolution—application to the Indian craton, *Geophys. J. Int.,* **196**(2), 1025–1033.

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, eds Khan, A. & Deschamps, F., Springer.

Box, G.E. & Tiao, G.C., 1973. *Bayesian Inference in Statistical Inference,* Addison-Wesley.

Capdeville, Y., Stutzmann, É., Wang, N. & Montagner, J.-P., 2013. Residual homogenization for seismic forward and inverse problems in layered media, *Geophys. J. Int.,* **194**(1), 470–487.

Chevrot, S., 2006. Finite-frequency vectorial tomography: a new method for high-resolution imaging of upper mantle anisotropy, *Geophys. J. Int.,* **165**(2), 641–657.

Choblet, G., Husson, L. & Bodin, T., 2014. Probabilistic surface reconstruction of coastal sea level rise during the twentieth century, *J. geophys. Res.,* **119**(12), 9206–9236.

Conrad, C.P., Behn, M.D. & Silver, P.G., 2007. Global mantle flow and the development of seismic anisotropy: differences between the oceanic and continental upper mantle, *J. geophys. Res.,* **112**(B7), B07317, doi:10.1029/2006JB004608.

Crampin, S., 1984. An introduction to wave propagation in anisotropic media, *Geophys. J. Int.,* **76**(1), 17–28.

Crampin, S. & Booth, D.C., 1985. Shear-wave polarizations near the North Anatolian Fault—II. Interpretation in terms of crack-induced anisotropy, *Geophys. J. Int.,* **83**(1), 75–92.

Darbyshire, F.A., Eaton, D.W. & Bastow, I.D., 2013. Seismic imaging of the lithosphere beneath Hudson Bay: episodic growth of the Laurentian mantle keel, *Earth planet. Sci. Lett.,* **373,** 179–193.

Debayle, E. & Ricard, Y., 2012. A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements, *J. geophys. Res.,* **117**(B10), B10308, doi:10.1029/2012JB009288.

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

Debayle, E., Kennett, B. & Priestley, K., 2005. Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia, *Nature,* **433**(7025), 509–512.

Deschamps, F., Lebedev, S., Meier, T. & Trampert, J., 2008. Azimuthal anisotropy of Rayleigh-wave phase velocities in the east-central United States, *Geophys. J. Int.,* **173**(3), 827–843.

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

Dettmer, J., Holland, C.W. & Dosso, S.E., 2013. Transdimensional uncertainty estimation for dispersive seabed sediments, *Geophysics,* **78**(3), WB63–WB76.

Dettmer, J., Dosso, S.E., Bodin, T., Stipčević, J. & Cummins, P.R., 2015. Direct-seismogram inversion for receiver-side structure with uncertain source–time functions, *Geophys. J. Int.,* **203**(2), 1373–1387.

Dickinson, W.R., 2002. The basin and range province as a composite extensional domain, *Int. Geol. Rev.,* **44**(1), 1–38.

Dosso, S.E., Dettmer, J., Steininger, G. & Holland, C.W., 2014. Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation, *Inverse Probl.,* **30**(11), 114018, doi:10.1088/0266-5611/30/11/114018.

Duijndam, A., 1988a. Bayesian estimation in seismic inversion. Part I: Principles, *Geophys. Prospect.,* **36**(8), 878–898.

Duijndam, A., 1988b. Bayesian estimation in seismic inversion. Part II: uncertainty analysis, , *Geophys. Prospect.,* **36,** 899–918.

Duputel, Z., Agram, P.S., Simons, M., Minson, S.E. & Beck, J.L., 2014. Accounting for prediction uncertainty when inferring subsurface fault slip, *Geophys. J. Int.,* **197**(1), 464–482.

Durand, S., Debayle, E. & Ricard, Y., 2015. Rayleigh wave phase velocity and error maps up to the fifth overtone, *Geophys. Res. Lett.,* **42**(9), 3266–3272.

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

Eakin, C.M., Obrebski, M., Allen, R.M., Boyarko, D.C., Brudzinski, M.R. & Porritt, R., 2010. Seismic anisotropy beneath Cascadia and the Mendocino triple junction: interaction of the subducting slab with mantle flow, *Earth planet. Sci. Lett.,* **297**(3), 627–632.

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.

Farra, V. & Vinnik, L., 2000. Upper mantle stratification by *P* and *S* receiver functions, *Geophys. J. Int.,* **141**(3), 699–712.

Farra, V., Vinnik, L., Romanowicz, B., Kosarev, G. & Kind, R., 1991. Inversion of teleseismic s particle motion for azimuthal anisotropy in the upper mantle: a feasibility study, *Geophys. J. Int.,* **106**(2), 421–431.

Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.-P., 2010. Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle, *Earth planet. Sci. Lett.,* **290**(3), 270–280.

Fontaine, F.R., Barruol, G., Tkalčić, H., Wölbern, I., Rümpker, G., Bodin, T. & Haugmard, M., 2015. Crustal and uppermost mantle structure variation beneath La Réunion hotspot track, *Geophys. J. Int.,* **203**(1), 107–126.

Forsyth, D.W., 1975. The early structural evolution and anisotropy of the oceanic upper mantle, *Geophys. J. Int.,* **43**(1), 103–162.

Foster, K., Dueker, K., Schmandt, B. & Yuan, H., 2014. A sharp cratonic lithosphere–asthenosphere boundary beneath the American midwest and its relation to mantle flow, *Earth planet. Sci. Lett.,* **402,** 82–89.

Fox, M., Bodin, T. & Shuster, D.L., 2015a. Abrupt changes in the rate of andean plateau uplift from reversible jump Markov chain Monte Carlo inversion of river profiles, *Geomorphology,* **238,** 1–14.

Fox, M., Leith, K., Bodin, T., Balco, G. & Shuster, D.L., 2015b. Rate of fluvial incision in the Central Alps constrained through joint inversion of detrital ^{10}Be and thermochronometric data, *Earth planet. Sci. Lett.,* **411,** 27–36.

Frederiksen, A. & Bostock, M., 2000. Modelling teleseismic waves in dipping anisotropic structures, *Geophys. J. Int.,* **141**(2), 401–412.

Fry, B., Deschamps, F., Kissling, E., Stehly, L. & Giardini, D., 2010. Layered azimuthal anisotropy of Rayleigh wave phase velocities in the European alpine lithosphere inferred from ambient noise, *Earth planet. Sci. Lett.,* **297**(1), 95–102.

Gallagher, K., Charvin, K., Nielsen, S., Sambridge, M. & Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to

determine optimal models, model resolution and model choice for Earth Science problems, *Mar. Petrol. Geol.,* **26**(4), 525–535.

Gallagher, K., Bodin, T., Sambridge, M., Weiss, D., Kylander, M. & Large, D., 2011. Inference of abrupt changes in noisy geochemical records using transdimensional changepoint models, *Earth planet. Sci. Lett.,* **311,** 182–194.

Gao, W., Grand, S.P., Baldridge, W.S., Wilson, D., West, M., Ni, J.F. & Aster, R., 2004. Upper mantle convection beneath the central Rio Grande rift imaged by *P* and *S* wave tomography, *J. geophys. Res.,* **109**(B3), B03305, doi:10.1029/2003JB002743.

Gelman, A., Carlin, J., Stern, H. & Rubin, D., 1995. *Bayesian Data Analysis,* Chapman & Hall.

Geyer, C. & Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Stat.,* **21**(4), 359–373.

Girardin, N. & Farra, V., 1998. Azimuthal anisotropy in the upper mantle from observations of *P*-to-*S* converted phases: application to southeast Australia, *Geophys. J. Int.,* **133**(3), 615–629.

Gouveia, W. & Scales, J., 1998. Bayesian seismic waveform inversion—parameter estimation and uncertainty analysis, *J. geophys. Res.,* **103**(B2), 2759–2780.

Green, P., 1995. Reversible jump MCMC computation and Bayesian model selection, *Biometrika,* **82,** 711–732.

Green, P., 2003. Trans-dimensional Markov chain Monte Carlo, *Highly Struct. Stoch. Syst.,* **27,** 179–198.

Gripp, A.E. & Gordon, R.G., 2002. Young tracks of hotspots and current plate velocities, *Geophys. J. Int.,* **150**(2), 321–361.

Gung, Y., Panning, M. & Romanowicz, B., 2003. Global anisotropy and the thickness of continents, *Nature,* **422**(6933), 707–711.

Hansen, S.M., Dueker, K. & Schmandt, B., 2015. Thermal classification of lithospheric discontinuities beneath USArray, *Earth planet. Sci. Lett.,* **431,** 36–47.

Hartog, R. & Schwartz, S.Y., 2000. Subduction-induced strain in the upper mantle east of the Mendocino triple junction, California, *J. geophys. Res.,* **105**(B4), 7909–7930.

Hastings, W., 1970. Monte Carlo simulation methods using Markov chains and their applications, *Biometrika,* **57,** 97–109.

Hopper, E. & Fischer, K.M., 2015. The meaning of midlithospheric discontinuities: a case study in the northern U.S. craton, *Geochem. Geophys. Geosyst.,* **16**(12), 4057–4083.

Iaffaldano, G., Hawkins, R., Bodin, T. & Sambridge, M., 2014. Redback: open-source software for efficient noise-reduction in plate kinematic reconstructions, *Geochem. Geophys. Geosyst.,* **15**(4), 1663–1670.

Jasra, A., Stephens, D., Gallagher, K. & Holmes, C., 2006. Bayesian mixture modelling in geochronology via Markov chain Monte Carlo, *Math. Geol.,* **38**(3), 269–300.

Karato, S.I., Olugboji, T. & Park, J., 2015. Mechanisms and geologic significance of the mid-lithosphere discontinuity in the continents, *Nature Geosci.,* **8**(7), 509–514.

Kind, R., Yuan, X. & Kumar, P., 2012. Seismic receiver functions and the lithosphere–asthenosphere boundary, *Tectonophysics,* **536,** 25–43.

Köhler, A., Maupin, V. & Balling, N., 2015. Surface wave tomography across the Sorgenfrei–Tornquist Zone, SW Scandinavia, using ambient noise and earthquake data, *Geophys. J. Int.,* **203**(1), 284–311.

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

Kosarev, G., Makeyeva, L. & Vinnik, L., 1984. Anisotropy of the mantle inferred from observations of *P* to *S* converted waves, *Geophys. J. Int.,* **76**(1), 209–220.

Kumar, P., Kind, R. & Yuan, X., 2010. Receiver function summation without deconvolution, *Geophys. J. Int.,* **180**(3), 1223–1230.

Kustowski, B., Ekström, G. & Dziewoński, A.M., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res.,* **113**(B6), B06306, doi:10.1029/2007JB005169.

Langston, C., 1979. Structure under Mount Rainier, Washington, inferred from teleseismic body waves, *J. geophys. Res.,* **84**(B9), 4749–4762.

Legendre, C., Deschamps, F., Zhao, L., Lebedev, S. & Chen, Q.-F., 2014. Anisotropic Rayleigh wave phase velocity maps of Eastern China, *J. geophys. Res.,* **119**(6), 4802–4820.

Leidig, M. & Zandt, G., 2003. Modeling of highly anisotropic crust and application to the Altiplano-Puna volcanic complex of the central Andes, *J. geophys. Res.,* **108**(B1), ESE 5-1–ESE 5-15.

Lekić, V. & Fischer, K.M., 2014. Contrasting lithospheric signatures across the Western United States revealed by Sp receiver functions, *Earth planet. Sci. Lett.,* **402,** 90–98.

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.

Levander, A. & Miller, M.S., 2012. Evolutionary aspects of lithosphere discontinuity structure in the western U.S., *Geochem. Geophys. Geosyst.,* **13**, Q0AK07, doi:10.1029/2012GC004056.

Levin, V. & Park, J., 1997. Crustal anisotropy in the ural mountains foredeep from teleseismic receiver functions, *Geophys. Res. Lett.,* **24**(11), 1283–1286.

Levin, V. & Park, J., 1998. *P-SH* conversions in layered media with hexagonally symmetric anisotropy: a cookbook, in *Geodynamics of Lithosphere and Earth's Mantle,* pp. 669–697, eds Plomerová, J., Liebermann, R.C. & Babuška, V., Springer.

Levin, V., Menke, W. & Park, J., 1999. Shear wave splitting in the Appalachians and the Urals: a case for multilayered anisotropy, *J. geophys. Res.,* **104**(B8), 17 975–17 993.

Liu, K.H., 2009. NA-SWS-1.1: a uniform database of teleseismic shear wave splitting measurements for North America, *Geochem. Geophys. Geosyst.,* **10**(5), doi:10.1029/2009GC002440.

Liu, Z., Park, J. & Rye, D.M., 2015. Crustal anisotropy in northeastern Tibetan Plateau inferred from receiver functions: rock textures caused by metamorphic fluids and lower crust flow?, *Tectonophysics,* **661,** 66–80.

Long, M.D. & Silver, P.G., 2009. Shear wave splitting and mantle anisotropy: measurements, interpretations, and new directions, *Surv. Geophys.,* **30** (4-5), 407–461.

Long, M.D., Jackson, K.G. & McNamara, J.F., 2016. *SKS* splitting beneath Transportable Array stations in eastern North America and the signature of past lithospheric deformation, *Geochem. Geophys. Geosyst,* **17**(1), 2–15.

Love, A.E.H., 1927. *A Treatise on the Theory of Elasticity,* 4th edn, Vol. 15, Cambridge Univ. Press.

MacKay, D.J., 2003. *Information Theory, Inference, and Learning Algorithms,* Cambridge Univ. Press.

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

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and Empirical Bayes, *Geophysics,* **69**(4), 1005–1016.

Malinverno, A. & Leaney, W., 2005. Monte-Carlo Bayesian look-ahead inversion of walkaway vertical seismic profiles, *Geophys. Prospect.,* **53**(5), 689–703.

Malinverno, A. & Parker, R.L., 2006. Two ways to quantify uncertainty in geophysical inverse problems, *Geophysics,* **71**(3), W15–W27.

Marone, F. & Romanowicz, B., 2007. The depth distribution of azimuthal anisotropy in the continental upper mantle, *Nature,* **447**(7141), 198–201.

Maupin, V., 1985. Partial derivatives of surface wave phase velocities for flat anisotropic models, *Geophys. J. Int.,* **83**(2), 379–398.

Maupin, V. & Park, J., 2007. Theory and observations – wave propagation in anisotropic media, in *Treatise on Geophysics,* Vol. 1, pp. 289–321, ed. Schubert, G., Elsevier.

Menke, W. & Levin, V., 2003. The cross-convolution method for interpreting SKS splitting observations, with application to one and two-layer anisotropic Earth models, *Geophys. J. Int.,* **154**(2), 379–392.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computationing machines, *J. Chem. Phys.,* **21**(6), 1087–1092.

Miller, M.S. & Eaton, D.W., 2010. Formation of cratonic mantle keels by arc accretion: evidence from S receiver functions, *Geophys. Res. Lett.,* **37**(18), L18305, doi:10.1029/2010GL044366.

Montagner, J.-P. & Guillot, L., 2002. Seismic anisotropy and global geodynamics, *Rev. Mineral. Geochem.,* **51**(1), 353–385.

Montagner, J.-P. & Nataf, H.-C., 1986. A simple method for inverting the azimuthal anisotropy of surface waves, *J. geophys. Res.,* **91**(B1), 511–520.

Montagner, J.-P. & Nataf, H.-C., 1988. Vectorial tomography. I. Theory, *Geophys. J. Int.,* **94**(2), 295–307.

Montagner, J.-P. & Tanimoto, T., 1991. Global upper mantle tomography of seismic velocities and anisotropies, *J. geophys. Res.,* **96**(B12), 20 337–20 351.

Montagner, J.-P., Griot-Pommera, D.-A. & Lavé, J., 2000. How to relate body wave and surface wave anisotropy?, *J. geophys. Res.,* **105**(B8), 19 015–19 027.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Nettles, M. & Dziewoński, A.M., 2008. Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America, *J. geophys. Res.,* **113**(B2), B02303, doi:10.1029/2006JB004819.

Obrebski, M. & Castro, R.R., 2008. Seismic anisotropy in northern and central gulf of California region, Mexico, from teleseismic receiver functions and new evidence of possible plate capture, *J. geophys. Res.,* **113**(B3), B03301, doi:10.1029/2007JB005156.

Obrebski, M., Castro, R.R., Valenzuela, R.W., van Benthem, S. & Rebollar, C.J., 2006. Shear-wave splitting observations at the regions of northern Baja California and southern Basin and Range in Mexico, *Geophys. Res. Lett.,* **33**(5), L05302, doi:10.1029/2005GL024720.

Obrebski, M., Kiselev, S., Vinnik, L. & Montagner, J.-P., 2010. Anisotropic stratification beneath Africa from joint inversion of SKS and P receiver functions, *J. geophys. Res.,* **115**(B9), B09313, doi:10.1029/2009JB006923.

Obrebski, M., Allen, R.M., Pollitz, F. & Hung, S.-H., 2011. Lithosphere–asthenosphere interaction beneath the Western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities, *Geophys. J. Int.,* **185**(2), 1003–1021.

Park, J., 1996. Surface waves in layered anisotropic structures, *Geophys. J. Int.,* **126**(1), 173–183.

Park, J. & Levin, V., 2000. Receiver functions from multiple-taper spectral correlation estimates, *Bull. seism. Soc. Am.,* **90**(6), 1507–1520.

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.

Pedersen, H.A., Bruneton, M., Maupin, V., & SVEKALAPKO Seismic Tomography Working Group, 2006. Lithospheric and sublithospheric anisotropy beneath the Baltic shield from surface-wave array analysis, *Earth planet. Sci. Lett.,* **244**(3), 590–605.

Peng, X. & Humphreys, E.D., 1997. Moho dip and crustal anisotropy in northwestern Nevada from teleseismic receiver functions, *Bull. seism. Soc. Am.,* **87**(3), 745–754.

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

Pilia, S., Rawlinson, N., Cayley, R.A., Bodin, T., Musgrave, R., Reading, A.M., Direen, N.G. & Young, M.K., 2015. Evidence of micro-continent entrainment during crustal accretion, *Sci. Rep.,* **5**, doi:10.1038/srep08218.

Plomerová, J. & Babuška, V., 2010. Long memory of mantle lithosphere fabric—European LAB constrained from seismic anisotropy, *Lithos,* **120**(1), 131–143.

Plomerová, J., Kouba, D. & Babuška, V., 2002. Mapping the lithosphere–asthenosphere boundary through changes in surface-wave anisotropy, *Tectonophysics,* **358**(1), 175–185.

Rader, E., Emry, E., Schmerr, N., Frost, D., Cheng, C., Menard, J., Yu, C.-Q. & Geist, D., 2015. Characterization and petrological constraints of the midlithospheric discontinuity, *Geochem. Geophys. Geosyst.,* **16**(10), 3484–3504.

Ramesh, D., Kind, R. & Yuan, X., 2002. Receiver function analysis of the North American crust and upper mantle, *Geophys. J. Int.,* **150**(1), 91–108.

Rau, C.J. & Forsyth, D.W., 2011. Melt in the mantle beneath the amagmatic zone, Southern Nevada, *Geology,* **39**(10), 975–978.

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.

Ritsema, J., Deuss, A., van Heijst, H.J. & Woodhouse, J.H., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.,* **184**(3), 1223–1236.

Romanowicz, B., 2009. The thickness of tectonic plates, *Science,* **324**(5926), 474–476.

Romanowicz, B. & Yuan, H., 2012. On the interpretation of SKS splitting measurements in the presence of several layers of anisotropy, *Geophys. J. Int.,* **188**(3), 1129–1140.

Rosenthal, J., 2000. Parallel computing and Monte Carlo algorithms, *Far East J. Theor. Stat.,* **4**(2), 207–236.

Rychert, C.A. & Shearer, P.M., 2009. A global view of the lithosphere-asthenosphere boundary, *Science,* **324**(5926), 495–498.

Saito, M., 1967. Excitation of free oscillations and surface waves by a point source in a vertically heterogeneous earth, *J. geophys. Res.,* **72**(14), 3689–3699.

Saito, M.A.S.A.N.O.R.I., 1988. DISPER80: a subroutine package for the calculation of seismic normal-mode solutions, in *Seismological Algorithms,* pp. 293–319, ed. Doornbos, D., Elsevier.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Transdimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167**(2), 528–542.

Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. Lond. A: Math. Phys. Eng. Sci.,* **371**(1984), 20110547, doi:10.1098/rsta.2011.0547.

Savage, M. & Sheehan, A., 2000. Seismic anisotropy and mantle flow from the Great Basin to the Great Plains, western United States, *J. geophys. Res.,* **105**(B6), 13 715–13 734.

Savage, M.K., 1998. Lower crustal anisotropy or dipping boundaries? Effects on receiver functions and a case study in New Zealand, *J. geophys. Res.,* **103**(B7), 15 069–15 087.

Schaeffer, A. & Lebedev, S., 2014. Imaging the North American continent using waveform inversion of global and USArray data, *Earth planet. Sci. Lett.,* **402,** 26–41.

Schmandt, B. & Humphreys, E., 2010. Complex subduction and small-scale convection revealed by body-wave tomography of the western United States upper mantle, *Earth planet. Sci. Lett.,* **297**(3), 435–445.

Schulte-Pelkum, V. & Mahan, K.H., 2014. A method for mapping crustal deformation and anisotropy with receiver functions and first results from USArray, *Earth planet. Sci. Lett.,* **402,** 221–233.

Sedlock, R.L., 2003. Geology and tectonics of the Baja California Peninsula and adjacent areas, in *Geological Society of America Special Papers, 374,* pp. 1–42.

Selway, K., Ford, H. & Kelemen, P., 2015. The seismic mid-lithosphere discontinuity, *Earth planet. Sci. Lett.,* **414,** 45–57.

Silver, P.G., 1996. Seismic anisotropy beneath the continents: probing the depths of geology, *Annu. Rev. Earth Planet. Sci.,* **24,** 385–432.

Silver, P.G. & Chan, W.W., 1991. Shear wave splitting and subcontinental mantle deformation, *J. geophys. Res.,* **96**(B10), 16 429–16 454.

Silver, P.G. & Savage, M.K., 1994. The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers, *Geophys. J. Int.,* **119**(3), 949–963.

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.

Sisson, S., 2005. Transdimensional Markov chains: a decade of progress and future perspectives, *J. Am. Stat. Assoc.,* **100**(471), 1077–1090.

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

Smith, M.L. & Dahlen, F., 1973. The azimuthal dependence of Love and Rayleigh wave propagation in a slightly anisotropic medium, *J. geophys. Res.,* **78**(17), 3321–3333.

Stähler, S. & Sigloch, K., 2014. Fully probabilistic seismic source inversion—Part 1: Efficient parameterisation, *Solid Earth,* **5**(2), 1055–1069.

Stephenson, J., Gallagher, K. & Holmes, C., 2006. Low temperature thermochronology and strategies for multiple samples 2: Partition modelling for 2D/3D distributions with discontinuities, *Earth planet. Sci. Lett.,* **241**(3–4), 557–570.

Takeuchi, H. & Saito, M., 1972. Seismic surface waves, *Methods Comput. Phys.,* **11,** 217–295.

Tanimoto, T. & Anderson, D.L., 1985. Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 s, *J. geophys. Res.,* **90**(B2), 1842–1858.

Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *J. Geophys.,* **50**(3), 150–170.

Thomson, C., 1997. Modelling surface waves in anisotropic structures I. Theory, *Phys. Earth planet. Inter.,* **103**(3), 195–206.

Thybo, H. & Perchuć, E., 1997. The seismic 8° discontinuity and partial melting in continental mantle, *Science,* **275**(5306), 1626–1629.

Tkalčić, H., Pasyanos, M.E., Rodgers, A.J., Gök, W.R., Walter, W. & Al-Amri, A., 2006. A multistep approach for joint modeling of surface wave dispersion and teleseismic receiver functions: implications for lithospheric structure of the Arabian Peninsula, *J. geophys. Res.,* **111**(B11), B11311, doi:10.1029/2005JB004130.

Trampert, J. & van Heijst, H.J., 2002. Global azimuthal anisotropy in the transition zone, *Science,* **296**(5571), 1297–1299.

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.

Vergne, J., Wittlinger, G., Farra, V. & Su, H., 2003. Evidence for upper crustal anisotropy in the Songpan-Ganze (northeastern Tibet) terrane, *Geophys. Res. Lett.,* **30**(11), 1552, doi:10.1029/2002GL016847.

Vinnik, L., Kosarev, G. & Makeeva, L., 1984. Lithosphere anisotropy from the observation of SKS and SKKS waves, *Dokl. Akad. Nauk SSSR,* **278**(6), 1335–1339.

Vinnik, L., Makeyeva, L., Milev, A. & Usenko, A.Y., 1992. Global patterns of azimuthal anisotropy and deformations in the continental mantle, *Geophys. J. Int.,* **111**(3), 433–447.

Vinnik, L., Aleshin, I., Kiselev, S., Kosarev, G. & Makeyeva, L., 2007. Depth localized azimuthal anisotropy from *SKS* and *P* receiver functions: the Tien Shan, *Geophys. J. Int.,* **169**(3), 1289–1299.

Vinnik, L., Oreshin, S., Makeyeva, L., Peregoudov, D., Kozlovskaya, E. & POLENET/LAPNET Working Group, 2014. Anisotropic lithosphere under the Fennoscandian shield from *P* receiver functions and SKS waveforms of the POLENET/LAPNET array, *Tectonophysics,* **628,** 45–54.

Vinnik, L., Kozlovskaya, E., Oreshin, S., Kosarev, G., Piiponen, K. & Silvennoinen, H., 2016. The lithosphere, LAB, LVZ and Lehmann discontinuity under central Fennoscandia from receiver functions, *Tectonophysics,* **667,** 189–198.

Vinnik, L.P., Farra, V. & Romanowicz, B., 1989. Azimuthal anisotropy in the earth from observations of SKS at geoscope and nars broadband stations, *Bull. seism. Soc. Am.,* **79**(5), 1542–1558.

Visser, K., Trampert, J. & Kennett, B., 2008. Global anisotropic phase velocity maps for higher mode Love and Rayleigh waves, *Geophys. J. Int.,* **172**(3), 1016–1032.

Wang, Y. & Tape, C., 2014. Seismic velocity structure and anisotropy of the Alaska subduction zone based on surface wave tomography, *J. geophys. Res.,* **119**(12), 8845–8865.

Wirth, E.A. & Long, M.D., 2014. A contrast in anisotropy across mid-lithospheric discontinuities beneath the Central United States—a relic of craton formation, *Geology,* **42**(10), 851–854.

Xie, J., Ritzwoller, M.H., Brownlee, S. & Hacker, B., 2015. Inferring the oriented elastic tensor from surface wave observations: preliminary application across the western United States, *Geophys. J. Int.,* **201**(2), 996–1019.

Yoshizawa, K. & Kennett, B.L.N., 2015. The lithosphere-asthenosphere transition and radial anisotropy beneath the Australian continent, *Geophys. Res. Lett.,* **42**(10), 3839–3846.

Young, M., Tkalcić, H., Bodin, T. & Sambridge, M., 2013a. Global *P* wave tomography of Earth's lowermost mantle from partition modeling, *J. geophys. Res.,* **118**(10), 5467–5486.

Young, M.K., Rawlinson, N. & Bodin, T., 2013b . Transdimensional inversion of ambient seismic noise for 3D shear velocity structure of the Tasmanian crust, *Geophysics,* **78**(3), WB49–WB62.

Yuan, H. & Romanowicz, B., 2010a. Depth dependent azimuthal anisotropy in the western US upper mantle, *Earth planet. Sci. Lett.,* **300**(3), 385–394.

Yuan, H. & Romanowicz, B., 2010b. Lithospheric layering in the North American craton, *Nature,* **466**(7310), 1063–1068.

Yuan, H., Dueker, K. & Schutt, D.L., 2008. Testing five of the simplest upper mantle anisotropic velocity parameterizations using teleseismic S and SKS data from the Billings, Montana PASSCAL array, *J. geophys. Res.,* **113**(B3), B03304, doi:10.1029/2007JB005092.

Yuan, H., Romanowicz, B., Fischer, K.M. & Abt, D., 2011. 3-D shear wave radially and azimuthally anisotropic velocity model of the North American upper mantle, *Geophys. J. Int.,* **184**(3), 1237–1260.

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

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

Zhang, X., Paulssen, H., Lebedev, S. & Meier, T., 2007. Surface wave tomography of the Gulf of California, *Geophys. Res. Lett.,* **34,** L15305, doi:10.1029/2007GL030631.

Zhu, H. & Tromp, J., 2013. Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean, *Science,* **341**(6148), 871–875.

Zhu, L. & Kanamori, H., 2000. Moho depth variation in Southern California from teleseismic receiver functions, *J. geophys. Res.,* **105**(B2), 2969–2980.

Zulfakriza, Z., Saygin, E., Cummins, P., Widiyantoro, S., Nugraha, A., Lühr, B.-G. & Bodin, T., 2014. Upper crustal structure of central java, Indonesia, from transdimensional seismic ambient noise tomography, *Geophys. J. Int.,* **197**(1), 630–635.

APPENDIX A: COMPUTING DISPERSION CURVES

Our model is parametrized in terms of a stack of layers with constant velocity $\mathbf{V}_s$, $\delta\mathbf{V}_s$, $\mathbf{V}_p$, $\delta\mathbf{V}_p$ and $\mathbf{\Psi}_{\text{fast}}$. The full elastic tensor of HTI is given by

$$\mathbf{C}_{mn} = \begin{pmatrix} A & A-2L & (A-2N) & 0 & 0 & 0 \\ A-2L & C & A-2L & 0 & 0 & 0 \\ (A-2N) & A-2L & A & 0 & 0 & 0 \\ 0 & 0 & 0 & L & 0 & 0 \\ 0 & 0 & 0 & 0 & N & 0 \\ 0 & 0 & 0 & 0 & 0 & L \end{pmatrix} \tag{A1}$$

where axis 3 is vertical (1 and 2 are horizontal) and axis 2 is the axis of symmetry. Elastic parameters can be derived from velocity parameters from the following equations (Farra *et al.* 1991):

$$C = \rho\left(\mathbf{V}_p + \frac{\delta\mathbf{V}_p}{2}\right)^2, \qquad A = \rho\left(\mathbf{V}_p - \frac{\delta\mathbf{V}_p}{2}\right)^2,$$

$$L = \rho\left(\mathbf{V}_s + \frac{\delta\mathbf{V}_s}{2}\right)^2, \qquad N = \rho\left(\mathbf{V}_s - \frac{\delta\mathbf{V}_s}{2}\right)^2.$$

Theoretical expressions relating the anisotropic elastic tensor $\mathbf{C}_{mn}$ to surface wave observations have been developed in the asymptotic

limit by Smith & Dahlen (1973). Ignoring 4ψ terms, the phase velocity of Rayleigh waves is given by:

$$\mathbf{C}(T, \psi) = \mathbf{C}_0(T) + \mathbf{C}_1(T)\cos(2\psi) + \mathbf{C}_2(T)\sin(2\psi) \tag{A2}$$

where T is the period. $\mathbf{C}_0(T)$ does not depend on azimuthal terms, it is equal to the dispersion curve of phase velocities predicted from for the equivalent vertical transverse isotropic (VTI) medium, that is, after averaging (A1) over all azimuths. The equivalent VTI tensor is given by:

$$\mathbf{C}_{mn} = \begin{pmatrix} \mathbf{A}_0 & (\mathbf{A}_0 - 2\mathbf{N}_0) & \mathbf{F}_0 & 0 & 0 & 0 \\ (\mathbf{A}_0 - 2\mathbf{N}_0) & \mathbf{A}_0 & \mathbf{F}_0 & 0 & 0 & 0 \\ \mathbf{F}_0 & \mathbf{F}_0 & \mathbf{C}_0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \mathbf{L}_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{L}_0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{N}_0 \end{pmatrix}$$

Now, the vertical axis (axis 3) is the axis of hexagonal symmetry. The VTI parameters are given by the following relation (Montagner & Nataf 1986):

$$\mathbf{A}_0 = \left(\frac{5}{8}A + \frac{3}{8}C\right)$$

$$\mathbf{C}_0 = A$$

$$\mathbf{F}_0 = A - N - L$$

$$\mathbf{L}_0 = (L + N)/2$$

$$\mathbf{N}_0 = \left(\frac{1}{8}C - \frac{1}{8}A + L\right)$$

For both Love and Rayleigh waves, the isotropic phase velocity $\mathbf{C}_0(T)$ can be easily computed from $\mathbf{A}_0$, $\mathbf{C}_0$, $\mathbf{F}_0$, $\mathbf{L}_0$, $\mathbf{N}_0$ by normal mode summation in a fully non-linear fashion (Saito 1988).

The other terms $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$ are given by the following linearized relation (Maupin 1985; Montagner & Nataf 1986):

$$\mathbf{C}_1(T) = \int_{z=0}^{\infty} \left(\frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{A}_0(z)} \mathbf{B}_c(z) + \frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{F}_0(z)} \mathbf{H}_c(z) + \frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{L}_0(z)} \mathbf{G}_c(z) \right) dz \tag{A3}$$

$$\mathbf{C}_2(T) = \int_{z=0}^{\infty} \left(\frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{A}_0(z)} \mathbf{B}_s(z) + \frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{F}_0(z)} \mathbf{H}_s(z) + \frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{L}_0(z)} \mathbf{G}_s(z) \right) dz \tag{A4}$$

where the partial derivatives indicates the sensitivity kernels of $\mathbf{C}_0$ with regards to elastic parameters $\mathbf{A}_0$, $\mathbf{F}_0$ and $\mathbf{L}_0$ for the equivalent VTI medium. These kernels can be easily computed by normal mode summation (Saito 1988). Since $\mathbf{C}_0$ is poorly dependent on elastic parameter $\mathbf{F}_0$, we set $\frac{\partial \mathbf{C}_0(T)}{\partial \mathbf{F}_0(z)} = 0$. The elastic parameters $\mathbf{G}_c$, $\mathbf{G}_s$, $\mathbf{B}_c$ and $\mathbf{B}_s$ are given by:

$$\mathbf{G}_c = \mathbf{G}\cos(2\mathbf{\Psi}_{\text{fast}})$$

$$\mathbf{G}_s = \mathbf{G}\sin(2\mathbf{\Psi}_{\text{fast}})$$

$$\mathbf{B}_c = \mathbf{B}\cos(2\mathbf{\Psi}_{\text{fast}})$$

$$\mathbf{B}_s = \mathbf{B}\sin(2\mathbf{\Psi}_{\text{fast}})$$

where

$$\mathbf{G} = (L - N)/2 \tag{A5}$$

$$\mathbf{B} = (C - A)/2 \tag{A6}$$

In this way, for each period T, $\mathbf{C}_0(T)$, $\mathbf{C}_1(T)$ and $\mathbf{C}_2(T)$ can be computed from $\mathbf{V}_s(z)$, $\delta\mathbf{V}_s(z)$ and $\mathbf{\Psi}_{\text{fast}}(z)$.

APPENDIX B: A CROSS-CONVOLUTION LIKELIHOOD FUNCTION

For a given earth model $\mathbf{m}$, consider that the observed seismograms can be written as:

$$\mathbf{V}_{\text{obs}}(t) = \mathbf{s}(t) * \mathbf{v}(t, \mathbf{m}) + \boldsymbol{\varepsilon}_v(t) \tag{B1}$$

$$\mathbf{H}_{\text{obs}}(t) = \mathbf{s}(t) * \mathbf{h}(t, \mathbf{m}) + \boldsymbol{\varepsilon}_h(t) \tag{B2}$$

where $\mathbf{s}(t)$ is the source time function, $\mathbf{v}(t, \mathbf{m})$ and $\mathbf{h}(t, \mathbf{m})$ are the vertical and radial impulse response functions of the near receiver structure, calculated for model $\mathbf{m}$, and $\boldsymbol{\varepsilon}_h(t)$ and $\boldsymbol{\varepsilon}_v(t)$ are random errors.

The vector of residuals is defined as follows:

$$\mathbf{r}(\mathbf{m}, t) = \mathbf{v}(t, \mathbf{m}) * \mathbf{H}_{\text{obs}}(t) - \mathbf{h}(t, \mathbf{m}) * \mathbf{V}_{\text{obs}}(t) \tag{B3}$$

Hence, we have:

$$\mathbf{r} = \mathbf{v} * \boldsymbol{\varepsilon}_h - \mathbf{h} * \boldsymbol{\varepsilon}_v. \tag{B4}$$

Since each term of (B4) is a discrete time-series, we can write:

$$\mathbf{r}[i] = \sum_{k=1}^{n} \mathbf{v}[i-k]\boldsymbol{\varepsilon}_h[k] - \sum_{k=1}^{n} \mathbf{h}[i-k]\boldsymbol{\varepsilon}_v[k], \tag{B5}$$

where n is the number of samples in the seismograms. In this way, each component of $\mathbf{r}$ is a linear combination of random variables. With simple algebra of random variables, the multivariate probability density function for vector $\mathbf{r}$ can easily be computed from the distribution of $\boldsymbol{\varepsilon}_h(t)$ and $\boldsymbol{\varepsilon}_v(t)$. In matrix form, we have:

$$\mathbf{r} = \mathbf{M}_v\boldsymbol{\varepsilon}_h - \mathbf{M}_h\boldsymbol{\varepsilon}_v. \tag{B6}$$

where $\mathbf{M}_v[i, j] = \mathbf{v}[i - j]$ and $\mathbf{M}_h[i, j] = \mathbf{h}[i - j]$ are matrices representing a linear transformation. If $\boldsymbol{\varepsilon}_h(t)$ and $\boldsymbol{\varepsilon}_v(t)$ are normally distributed with covariance matrices $\mathbf{C}_v$ and $\mathbf{C}_h$, each component of $\mathbf{r}$ is therefore also normally distributed, and the covariance matrix for $\mathbf{r}$ is given by:

$$\mathbf{C}_r = \mathbf{M}_v\mathbf{C}_v\mathbf{M}_v^T + \mathbf{M}_h\mathbf{C}_h\mathbf{M}_h^T \tag{B7}$$

The probability distribution of residuals $\mathbf{r}$ then writes:

$$p(\mathbf{r} \mid \mathbf{m}) = \frac{1}{\sqrt{2\pi^n|\mathbf{C}_r|}} \times \exp\left\{ -\frac{1}{2}\left(\mathbf{r}^T\mathbf{C}_r^{-1}\mathbf{r}\right)\right\}. \tag{B8}$$

For simplicity, we assume that $\boldsymbol{\varepsilon}_h(t)$ and $\boldsymbol{\varepsilon}_v(t)$ are white noise with standard deviation σ_p (i.e. $\mathbf{C}_h = \mathbf{C}_v = \sigma_p^2\mathbf{I}_n$). For seismograms normalized to unit energy ($\sum_{i=1}^{n} \mathbf{h}_i^2 + \mathbf{v}_i^2 = 1$) and after ignoring off-diagonal terms in $\mathbf{C}_r$, we have:

$$\mathbf{C}_r = \sigma_p^2\mathbf{I}_n \tag{B9}$$

and hence, we can approximate the distribution of residuals with the following expression:

$$p(\mathbf{r} \mid \mathbf{m}) = \frac{1}{(\sqrt{2\pi}\sigma_p)^n} \times \exp\left\{ \frac{-\|\mathbf{r}(\mathbf{m})\|^2}{2\sigma_p^2} \right\}, \tag{B10}$$

which is used as a likelihood probability density function in our formulation of the problem.

APPENDIX C: THE PRIOR

We use a prior distribution following Bodin *et al.* (2012b), where the number of layers k was variable. However in this study, among the k layers, we also allow for a variable number l of anisotropic layers, and have modified the prior accordingly. The prior can be separated into two terms:

$$p(\mathbf{m}) = p(k)p(\mathbf{m}|k) \tag{C1}$$

where $p(k)$ is the prior for the number of layers. We use a uniform distribution over the range $[k_{\min} \quad k_{\max}]$ and hence $p(k) = 1/(k_{\max} - k_{\min}) = 1/\Delta k$. Among these k layers, l layers are anisotropic (with $0 \leq l < k$). The prior $p(\mathbf{m}|k)$ can then be also separated into two terms:

$$p(\mathbf{m}|k) = p(l|k)p(\mathbf{m}|l, k). \tag{C2}$$

Given k layers, we let the number of anisotropic layers l be uniformly distributed over the range $[0 \quad k-1]$ (the last layer is a half-space and cannot be anisotropic), and set $p(l|k) = 1/k$.

For a given pair of k, l, the prior probability distributions for the model parameters are independent from each other, and so can be written in separable form

$$p(\mathbf{m}|l, k) = p(\mathbf{z}, \mathbf{V}_s, \delta\mathbf{V}_s, \mathbf{\Psi}_{\text{fast}} \mid k, l) \tag{C3}$$

$$p(\mathbf{m}|l, k) = p(\mathbf{z} \mid k)p(\mathbf{V}_s \mid k)p(\delta\mathbf{V}_s \mid l, k)p(\mathbf{\Psi}_{\text{fast}} \mid l, k). \tag{C4}$$

Each of the terms in (C4) are a uniform distribution over the range of possible values, and integrate to one.

As in Bodin & Sambridge (2009), for mathematical convenience, let us for the moment assume that the depths of discontinuities $\mathbf{z}$ can only take place on an underlying grid of finite points defined by N possible depths. For k discontinuities, there are $C_N^k = (N!)/(k!(N-k)!)$ possible configurations on the N possible depths of the underlying grid. We give equal probability to each of these configurations, and hence

$$p(\mathbf{z} \mid k) = (C_N^k)^{-1}. \tag{C5}$$

For velocity, the prior for each component v_i of the vector $\mathbf{V}_s$ is also uniform over a defined interval $[V_{\min} \quad V_{\max}]$. Hence, we have

$$p(v_i \mid k) = 1/(\Delta V_s) \tag{C6}$$

where $\Delta V = (V_{\max} - V_{\min})$. If v_i is outside the interval, the prior is zero. Since the velocity in each layer is considered independent *a priori*,

$$p(\mathbf{V}_s \mid k) = \prod_{i=1}^{k} p(v_i \mid k) = \left(\frac{1}{\Delta V}\right)^k. \tag{C7}$$

Similarly, for anisotropic layers we shall also account for the fact that there are a number of possible configurations. There are $C_k^l = (k!)/(l!(k-l)!)$ different ways to choose l anisotropic layers among k. Once the locations of anisotropic layers is set, the two extra anisotropic parameters can vary over the range $[\Psi_{\max} \quad \Psi_{\min}]$ and $[\delta V_{\max} \quad \delta V_{\min}]$, and the prior probability for those parameters writes:

$$p(\delta\mathbf{V}_s \mid l, k)p(\mathbf{\Psi}_{\text{fast}} \mid l, k) = (C_k^l)^{-1}\left(\frac{1}{\Delta\delta V \Delta\Psi}\right)^l \tag{C8}$$

where $\Delta\Psi = (\Psi_{\max} - \Psi_{\min})$ and $\Delta\delta V = (\delta V_{\max} - \delta V_{\min})$. Combining all terms, the full prior distribution writes:

$$p(\mathbf{m}) = \frac{1}{kC_N^k C_k^l \Delta k (\Delta V)^k (\Delta\delta V)^l (\Delta\Psi)^l} \tag{C9}$$

when model parameters are inside defined intervals and zero otherwise. Here, it is clear how the prior probability decreases as the dimension of the space l and k increase, or as the volume $\delta V \Delta\Psi \Delta\delta V$ increases.

APPENDIX D: PROPOSAL DISTRIBUTIONS

Having randomly initialized the model parameters by drawing values from the prior distribution of each parameter, the algorithm proceeds iteratively. At each iteration of the chain, we propose a new model by drawing from a probability distribution $q(\mathbf{m}'|\mathbf{m})$ such that the new proposed model $\mathbf{m}'$ is conditional only on the current model $\mathbf{m}$. At each iteration of the reversible jump algorithm, one type of move is uniformly randomly selected from the five following possibilities:

(i) *Change the depth of a discontinuity.* Randomly pick a layer i from a uniform distribution and randomly change the position of its discontinuity according to

$$q_z(z_i' \mid z_i) = \frac{1}{\theta_z\sqrt{2\pi}} \exp\left\{-\frac{(z_i' - z_i)^2}{2\theta_z^2}\right\}. \tag{D1}$$

The variance θ_z^2 of the proposal distribution is defined by the user.

(ii) *Change the velocity in one layer.* Randomly select a layer and randomly propose a new value v_i' using a Gaussian probability distribution centred at the current value v_i:

$$q_{v1}(v_i' \mid v_i) = \frac{1}{\theta_{v1}\sqrt{2\pi}} \exp\left\{-\frac{(v_i' - v_i)^2}{2\theta_{v1}^2}\right\}. \tag{D2}$$

Again, the variance θ_{v1}^2 of the Gaussian function is a parameter to be chosen. Hence, we have

$$v_i' = v_i + u \tag{D3}$$

where u is a random deviate from a normal distribution $N(0, \theta_{v1})$. All the other model parameters are kept constant, and hence this proposal does not involve a change in dimension.

(iii) *Birth of an isotropic layer.* Create a new isotropic layer. Add a new discontinuity with the position z_{k+1} found by choosing uniformly randomly a point from the underlying grid that is not already occupied. There are $(N-k)$ discrete points available. Then, a new velocity value v_{k+1}' needs to be assigned to the new layer. This is drawn from a Gaussian proposal probability density with the same form as (D2)

$$q_{v2}(v_{k+1}' \mid v_i) = \frac{1}{\theta_{v2}\sqrt{2\pi}} \exp\left\{-\frac{(v_{k+1}' - v_i)^2}{2\theta_{v2}^2}\right\} \tag{D4}$$

where v_i is the current velocity value at this depth where the birth takes place. The variance θ_{v2}^2 of the Gaussian function is a parameter to be chosen.

(iv) *Death of an isotropic layer.* Remove at random one isotropic layer by drawing a number from a uniform distribution over the range $[1, k-l]$. The velocity of the neighboring layers remain unchanged.

(v) *Birth of an anisotropic layer.* Create a new anisotropic layer by simply adding two extra parameters to one of the already existing $k-l$ isotropic layers. The number of layers k is unchanged. A new δv and ψ parameter must be assigned to this layer. There are drawn from a uniform distribution over the ranges $[\Psi_{\max} \quad \Psi_{\min}]$ and $[\delta V_{\max} \quad \delta V_{\min}]$.

(vi) *Death of an anisotropic layer.* Choose at random one anisotropic layer by drawing a number from a uniform distribution

over the range $[1, l]$. The anisotropic parameters δv for this layer is set to zero, and the layer becomes isotropic. The isotropic parameter v stays unchanged.

(vii) *Change the estimated data noise.* Randomly select a data type and perturb its noise amplitude (e.g. $\sigma_{C_0}, \sigma_{C_1}, \sigma_{C_2}, \sigma_p, \sigma_{\mathrm{SKS}}$). As before, the new value is proposed by drawing a number according to a Gaussian distribution around the current value. This type of perturbation changes the weight given to the chosen data type in the joint inversion.

The standard deviations $(\theta_z, \theta_{v1}, \theta_{v2}, \theta_{\sigma_i})$ of the Gaussian proposal functions are parameters to be fixed by the user. As shown in MacKay (2003), the magnitude of perturbations does not affect the solution but rather the sampling efficiency of the algorithm. Thus, the width of proposal distributions are 'tuned' by trial-and-error in order to have an acceptance rate as close to 44 per cent for each type of perturbation (Rosenthal 2000).

APPENDIX E: ACCEPTING OR REJECTING THE PROPOSED MODEL

Once a proposed model has been drawn from the distribution $q(\mathbf{m}'|\mathbf{m})$, the new model is then accepted with a probability $\alpha(\mathbf{m}'|\mathbf{m})$, that is, a uniform random deviate, r, is generated between 0 and 1. If $r \leq \alpha$, the move is accepted, the current model $\mathbf{m}$ is replaced with $\mathbf{m}'$ and the chain moves to the next step. If $r > \alpha$, the move is rejected and the current model is retained for the next step of the chain where the process is repeated. The acceptance probability, $\alpha(\mathbf{m}'|\mathbf{m})$, is the key to ensuring that the samples will be generated according to the target density $p(\mathbf{m}|\mathbf{d}_{\mathrm{obs}})$. It can be shown (Green 1995, 2003) that the chain of sampled models will converge to the transdimensional posterior distribution, $p(\mathbf{m}|\mathbf{d}_{\mathrm{obs}})$, if

$$\alpha(\mathbf{m}' \mid \mathbf{m}) = \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \cdot \frac{p(\mathbf{d}_{\mathrm{obs}} \mid \mathbf{m}')}{p(\mathbf{d}_{\mathrm{obs}} \mid \mathbf{m})} \cdot \frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})} \cdot |\mathbf{J}|\right] \tag{E1}$$

where the matrix $\mathbf{J}$ is the Jacobian of the transformation from $\mathbf{m}$ to $\mathbf{m}'$ (Green 2003). As shown in Bodin *et al.* (2012b), the Jacobian is one for our problem and can therefore be ignored.

For the proposal types that do not involve a change of dimension (i.e. perturbations in velocity, depth of discontinuities and noise parameters) distributions are symmetrical. That is, the probability to go from $\mathbf{m}$ to $\mathbf{m}'$ is equal to the probability to go from $\mathbf{m}'$ to $\mathbf{m}$, and hence $q(\mathbf{m}|\mathbf{m}') = q(\mathbf{m}'|\mathbf{m})$. Furthermore, since the dimension is unchanged, we have $k = k'$ and $l = l'$, and according to (C9), the prior ratio is also one ($p(\mathbf{m}) = p(\mathbf{m}')$).

In this way, for perturbations not involving a change of dimension (proposals 1, 2 and 7), the probability of acceptance in (E2) reduces to the classical metropolis rule, where only the likelihood ratio need to be computed:

$$\alpha(\mathbf{m}' \mid \mathbf{m}) = \min\left[1, \frac{p(\mathbf{d}_{\mathrm{obs}} \mid \mathbf{m}')}{p(\mathbf{d}_{\mathrm{obs}} \mid \mathbf{m})}\right] \tag{E2}$$

Below we detail the form of the acceptance rule for transdimensional moves (proposals 3–6).

E1 Give birth/death to an isotropic layer

When a new isotropic layer is added to the model, the algorithm jumps between a model $\mathbf{m}$ with k layers and l anisotropic layers to a model $\mathbf{m}'$ with $(k + 1)$ layers and l anisotropic layers. A new discontinuity is added, among the $(N - k)$ discrete points available. The probability of a birth at position z'_{k+1} is then given by

$$q(\mathbf{z}' \mid \mathbf{m}) = 1/(N - k). \tag{E3}$$

A new velocity value v'_{k+1} is given to the newborn layer, with probability given in (D4):

$$q(\mathbf{V}'_S \mid \mathbf{m}) = q_{v2}(v'_{k+1}|v_i) \tag{E4}$$

For the reverse step, where the discontinuity at z'_{k+1} is removed from $\mathbf{m}'$, the probability of deleting the interface at position z'_{k+1} is

$$q(\mathbf{z} \mid \mathbf{m}') = 1/(k + 1 - l) \tag{E5}$$

and the probability of removing velocity v'_{k+1} when cell $k + 1$ is deleted is:

$$q(\mathbf{V}_s \mid \mathbf{m}') = 1. \tag{E6}$$

We have therefore

$$\left(\frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})}\right)_{\mathrm{birth}} = \frac{(N - k)}{(k + 1 - l) q_{v2}(v'_{k+1}|v_i)}. \tag{E7}$$

For the prior ratio, according to (C9), we have:

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{k C_N^k C_k^l \Delta k (\Delta V_s)^k (\Delta \delta V)^l (\Delta \Psi)^l}{(k + 1) C_N^{k+1} C_{k+1}^l \Delta k (\Delta V_s)^{k+1} (\Delta \delta V)^l (\Delta \Psi)^l} \tag{E8}$$

which reduces to

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{k(k - l + 1)}{\Delta V_s (N - k)(k + 1)} \tag{E9}$$

After combining (E7) and (E9) , the full acceptance term is then given by

$$\alpha(\mathbf{m} \mid \mathbf{m}')_{\mathrm{birth}} = \min\left[1, \frac{1}{q_{v2}(v'_{k+1}|v_i)\Delta V_s} \left(\frac{k}{k+1}\right) \frac{p(d|\mathbf{m}')}{p(d|\mathbf{m})}\right]. \tag{E10}$$

Here, note that if instead of a normal distribution for proposing the new velocity v'_{k+1}, we use a uniform proposal distribution q_{v2} equal to the prior, we have $q_{v2}(v'_{k+1}|v_i)\Delta V_s = 1$.

For a death move, where the Markov chain jumps between a model with k layers to a model with $k - 1$ layers, that is, where one of the $k - l$ isotropic layers is removed, we can similarly show that

$$\alpha(\mathbf{m} \mid \mathbf{m}')_{\mathrm{death}} = \min\left[1, q_{v2}(v_i|v'_j)\Delta V_s \left(\frac{k}{k - 1}\right) \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}\right] \tag{E11}$$

where v_i is velocity of the layer that gets deleted, and v'_j is the velocity in the proposed model at depth z_i, where the discontinuity was deleted.

E2 Give birth/death to an anisotropic layer

In the case where we give birth to an anisotropic layer, we randomly choose one of the $k - l$ isotropic layers, and add two parameters to this layer: δv and ψ, which values are drawn from a uniform distributions over the ranges $\Delta\delta V$ and $\Delta\Psi$. In this case, the total number of layers k is unchanged. The probability of proposing $\mathbf{m}'$ is therefore given by:

$$q(\mathbf{m}' \mid \mathbf{m}) = \frac{1}{\Delta\delta V \Delta\Psi} \frac{1}{(k - l)}. \tag{E12}$$

For the reverse step, we remove one anisotropic layer among the $(l + 1)$ possibilities and

$$q(\mathbf{m} \mid \mathbf{m}') = \frac{1}{l + 1}. \tag{E13}$$

The proposal ratio is then given by

$$\left(\frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})}\right)_{\text{birth}} = \Delta\delta V \Delta\Psi \frac{(k-l)}{(l+1)}. \tag{E14}$$

The prior ratio is given by:

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{kC_N^k C_k^l (\Delta V_s)^k (\Delta\delta V)^l (\Delta\Psi)^l}{kC_N^k C_k^{l+1} (\Delta V_s)^k (\Delta\delta V)^{l+1} (\Delta\Psi)^{l+1}} \tag{E15}$$

which reduces to

$$\frac{p(\mathbf{m}')}{p(\mathbf{m})} = \frac{1}{\Delta\delta V \Delta\Psi} \frac{(l+1)}{(k-l)}. \tag{E16}$$

And hence, the full acceptance term in (E2) simply reduces to

$$\alpha(\mathbf{m} \mid \mathbf{m}')_{\text{birth}} = \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}\right]. \tag{E17}$$

For a death of an anisotropic layer, the Markov chain jumps from a model with l anisotropic layers to a model with $(l-1)$ anisotropic layers. The number of layers k is unchanged. In this case, we can similarly show that the probability of accepting $\mathbf{m}'$ is

$$\alpha(\mathbf{m} \mid \mathbf{m}')_{\text{death}} = \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}\right] \tag{E18}$$

Therefore, the birth and death acceptance terms for anisotropic parameters are identical to fixed dimension moves.