Geophysical Journal International

Geophys. J. Int. (2019) **217,** 238–256
Advance Access publication 2019 January 16
GJI Seismology

doi: 10.1093/gji/ggz016

A model space search approach to finite-frequency SKS splitting intensity tomography in a reduced parameter space

Puskar Mondal and Maureen D. Long

Department of Geology and Geophysics, Yale University, New Haven, CT, USA. E-mail: puskar.mondal@yale.edu

Accepted 2019 January 8. Received 2018 December 21; in original form 2018 August 30

SUMMARY

Here we develop a theoretical and practical framework for the tomographic inversion of shear wave splitting intensity measurements for anisotropic structure in the upper mantle using a model space search approach. Treating the anisotropic scatterers as a first order perturbation to the background isotropic state, we implement the Born approximation to compute the integral sensitivity kernels in a finite frequency framework. We implement a parametrization of the anisotropy based on insights from olivine elasticity and fabric development that involves three parameters (corresponding to the azimuth and dip of the anisotropic symmetry axis, plus a strength parameter). Previous work on finite-frequency shear wave splitting tomography has implemented a linearization technique to invert splitting intensity data for the spatial distribution of anisotropic scatterers. The inverse problem, however, is strongly non-linear in terms of several of the involved parameters (those that describe the orientation of the symmetry axis), and their variation is not of first order. Therefore, in the case of a realistic upper mantle where anisotropic structure varies in a complicated manner, a linearization technique may not be adequate. To ameliorate these problems, we implement a model space search approach (specifically, a Markov chain Monte Carlo with Gibbs sampling algorithm) to the tomographic inversion of splitting intensity data. This approach allows for the visualization of posterior probability distributions for anisotropic parameters in the inversion. We perform a suite of synthetic resolution tests to demonstrate the reliability of our method, using a station distribution from an actual deployment of a dense seismic network. These resolution tests show that anisotropic structure may be resolved up to a length scale of roughly 50 km with teleseismic SKS waves for station spacing of 10–15 km.

Key words: Seismic anisotropy; Seismic tomography; Wave propagation.

1 INTRODUCTION

Relating shear wave splitting observations to anisotropic structure in the Earth's interior is a classical problem in seismology (e.g. Vinnik *et al.* 1984; Savage & Silver 1993; Silver 1996; Savage 1999). Shear wave splitting arises when shear waves travel through an anisotropic medium, and refers to the phase lead of the quasi-shear wave component polarized in the fast direction with respect to the component polarized in the slow direction. Observations of shear wave splitting (or other wave phenomena that reflect anisotropy) are important in understanding various geodynamic phenomena, such as overall patterns of mantle convection (e.g. Conrad & Behn 2010; Long & Becker 2010), subduction dynamics (e.g. Long & Silver 2008, 2009a; Faccenda & Capitanio 2012; Long & Wirth 2013), oceanic spreading centres and continental rift systems (e.g. Wolfe & Solomon 1998; Kendall *et al.* 2005; Bodmer *et al.* 2015), continental deformation (e.g. Hongsresawat *et al.* 2015; Long *et al.* 2016) and flow at the base of the mantle (e.g. Cottaar & Romanowicz 2013; Ford & Long 2015; Creasy *et al.* 2017).

A commonly used method to infer upper mantle anisotropy—and therefore upper mantle flow patterns—is to measure shear wave splitting parameters (fast direction, ϕ, and delay time, dt), typically for phases such as SKS that have been refracted through the outer core (e.g. Long & Silver 2009b). However, the interpretation of SKS splitting measurements is not straightforward, because they represent an integration of anisotropic structure along the ray path. Large delay times, for example, may yield several interpretations in terms of strength of anisotropy (a thin highly anisotropic layer versus a thick layer with weak anisotropy). Similarly, in the presence of multiple anisotropic layers, the apparent fast direction reflects a complicated (and non-commutative, e.g. Savage & Silver 1993; Silver & Long 2011) combination of layer parameters, and isolating the depth of different anisotropic regions is a major challenge.

To lessen the ambiguity of shear wave splitting interpretation, a tomographic approach that allows for the localization of anisotropy is desirable. However, tomographic inversion for anisotropic structure is a daunting task, since 21 parameters are needed to describe the most general form of anisotropy, and such a large parameter space does not allow for a tightly constrained tomographic inversion given limited data. Furthermore, the inverse problem for anisotropy is highly non-linear, as the sensitivities of splitting observations to anisotropy depend on the starting model used to compute them (e.g. Abt & Fischer 2008). Finally, shear wave splitting tomography is further complicated by the measurements themselves, which are challenging for noisy data (e.g. Monteiller & Chevrot 2010), resulting in much smaller data sets than traditional traveltime tomography. The measurement problem can be somewhat ameliorated by making measurements of the splitting intensity (e.g. Chevrot 2000) rather than the traditionally estimated splitting parameters (ϕ, dt), as this quantity can be more robustly estimated from noisy data (e.g. Long & van der Hilst 2005; Monteiller & Chevrot 2010) and has the advantage of being commutative along the ray path (e.g. Chevrot 2006, 2000; Silver & Long 2011).

Substantial effort has been devoted to developing theoretical frameworks and practical strategies for inverting shear wave splitting data for anisotropic structure, much of it in the context of finite-frequency sensitivity kernels that relate model perturbations to splitting observations (e.g. Favier & Chevrot 2003; Favier *et al.* 2004; Chevrot 2006; Long *et al.* 2008; Sieminski *et al.* 2008). However, there have only been a few published examples of the application of splitting tomography to actual data sets to study mantle anisotropy. Abt & Fischer (2008) applied a ray theoretical framework to invert for anisotropic structure in a subduction zone setting, taking advantage of seismicity within the subducting slab to provide ray coverage for local *S* phases; this approach was also applied by Calixto *et al.* (2014). Monteiller & Chevrot (2011) used data from dense seismic networks in southern California to apply finite-frequency SKS splitting tomography, producing a model of lateral and depth variations of anisotropy parameters. Lin *et al.* (2014a) applied a full-wave approach to the computation of sensitivities and a multiscale parametrization to the inversion of a similar data set from southern California, based on insights from full-wave modelling of SKS phases (Lin *et al.* 2014b).

While these tomographic approaches have enabled substantial progress in understanding the 3-D distribution of seismic anisotropy in well-sampled regions, they all share a significant limitation in that the inversion strategy is based on a linearization around a background model, which may result in bias in the model estimation. Sensitivity kernels for anisotropic structure depend strongly on the background model used to calculate them, and are substantially different for heterogeneous anisotropic starting models than for the homogeneous ones that are often used to calculate them (Long *et al.* 2008). Because of this limitation, the implementation of a model space search approach to shear wave splitting is desirable; such an approach, while computationally intensive, avoids the explicit linearization of a highly non-linear inverse problem, and yields constraints on the ranges of model parameters allowed by the observations (Wookey 2012).

Building on previous work on finite frequency shear wave splitting intensity tomography (Chevrot 2006; Long *et al.* 2008; Sieminski *et al.* 2008), here we develop a framework for tomographic inversion of SKS splitting intensity measurements from dense seismic networks, with a new parametrization of anisotropy and a model space search approach. We reparametrize the full elasticity tensor based on insights from olivine elasticity and fabric development, reducing the number of parameters to three (a measure of anisotropy strength, plus the dip and azimuth of the fast symmetry axis). We compute Born-approximation finite frequency sensitivity kernels relating model perturbations to splitting intensity, allowing us to sample the first few Fresnel zones of sensitivity instead of assuming sensitivity along an infinite frequency ray. This finite frequency approach allows us to obtain overlapping sensitivity kernels for observations at adjacent stations, even for nearly vertically propagating SKS waves. We implement a Markov chain Monte Carlo (McMC) technique to sample model space and obtain a most likely model for anisotropic structure, and sample the resulting probability distributions for various anisotropic parameters.

The structure of the paper is as follows. We begin with a development of the theoretical formulation necessary for the finite-frequency anisotropy tomography problem in a 3-D, Cartesian geometry, building on previous work by Chevrot (2006) and Long *et al.* (2008). We discuss a simplified parametrization for anisotropy based on elasticity and fabric development in olivine. We then test our forward formulation against various analytical solutions for the behaviour of splitting intensity observations for simple anisotropic models. Next, we develop and implement an McMC approach to probing parameter space, and perform inversion tests with synthetic data sets to evaluate computational cost and infer the resolving power of our algorithm for synthetic models with multiple anisotropic layers. Finally, we discuss the prospects for implementation of our tomography strategy to actual SKS splitting data sets.

2 THEORETICAL FORMULATION OF SENSITIVITY KERNELS

Equation of motion for the infinitesimal displacement along with the constitutive relation for linear elasticity lead to the wave equation

$$\rho \frac{\partial^2 \mathbf{u}}{\partial t^2} = \nabla \cdot (\mathbf{C} : \nabla \mathbf{u}), \tag{1}$$

where $\mathbf{u}$, $\mathbf{C}$ and ρ are the displacement, elasticity tensor and density, respectively. Given the initial displacement and velocity (i.e. the Cauchy data), eq. (1) has a unique time evolution [in $C^2(I \subset \mathbf{R}; W^{1,2}(V \subset \mathbf{R}^3)]$, which may be computed using Green's function (e.g. Graff 2012). Computation of the Green's function ($\mathbf{G}$) involves the solution of the following inhomogeneous partial differential equations (Aki & Richards 2002)

$$\rho \partial_t^2 \mathbf{G} = \nabla \cdot [\mathbf{C} : \nabla \mathbf{G}] + \mathbf{I}\delta(\mathbf{r} - \mathbf{r}_s)\delta(t - t_s), \tag{2}$$

where $\mathbf{I}$, $\mathbf{r}_s$ and t_s are the identity operator, seismic source location and source time, respectively.

To focus on anisotropic tomography, we express the wave related entities as first order perturbations from the background isotropic reference state:

$$\mathbf{u} = \mathbf{u}_0 + \epsilon \delta \mathbf{u}, \tag{3}$$

$$\Sigma = \Sigma_0 + \epsilon \sigma, \tag{4}$$

$$\mathbf{C} = \mathbf{C}_{Iso} + \epsilon \delta \mathbf{C}, \tag{5}$$

where $[\mathbf{u}_0, \Sigma_0, \mathbf{C}_{Iso}]$, $[\delta \mathbf{u}, \sigma, \delta \mathbf{C}]$ and ε represent the isotropic reference state, the perturbed state and the perturbation parameter, respectively. Subtracting the isotropic reference state and neglecting the advection term from the total derivative and higher order

terms $[o(\varepsilon^2)]$ leads to

$$\rho\frac{\partial^2\delta\mathbf{u}}{\partial t^2} = \nabla\cdot(\mathbf{C}_{Iso}:\nabla\delta\mathbf{u}) + \nabla\cdot(\delta\mathbf{C}:\nabla\mathbf{u}_0) \tag{6}$$

which resembles the equation of the motion of the unperturbed state, with $\nabla\cdot(\delta\mathbf{C}:\nabla\mathbf{u}_0)$ being the source term (forcing term). Therefore, we may now invoke the same Green's function in order to compute the perturbed displacement field. However, before doing so, we transform the entities to the frequency domain for convenience in computation

$$g(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} g(t')e^{-\iota\omega t'}\mathrm{d}t'\right)e^{\iota\omega t}\mathrm{d}\omega, \tag{7}$$

where ω is the angular frequency. Here, we assume that $g\in L^1(\mathbf{R})$, that is $\int_{-\infty}^{\infty}|g(t)|\mathrm{d}t<\infty$ with $\mathbf{R}$ being the real line in order to guarantee the existence of Fourier transform of Dirac's delta function (absolutely integrable, but not square integrable). Under such transformation, elliptic equations defining the perturbed state may be written as

$$-\rho\omega^2\delta\mathbf{u} = \nabla\cdot(\mathbf{C}_{Iso}:\nabla\delta\mathbf{u}) + \nabla\cdot(\delta\mathbf{C}:\nabla\mathbf{u}_0), \tag{8}$$

and existence of its solutions is well established (Agmon 1962). The eigenfrequency spectrum is unique and may be computed as follows. Taking dot product of both sides of eq. (8) by $\delta\mathbf{u}\in W^{1,2}(V)-\{0\}$ and integrating over the spatial domain and assuming that the perturbation vanishes on the boundary of the manifold (if it has a boundary), we obtain the following expression for ω:

$$\omega^2 = \frac{\int_V(\nabla\delta\mathbf{u}:\mathbf{C}_{Iso}:\nabla\delta\mathbf{u}+\nabla\delta\mathbf{u}:\delta\mathbf{C}:\nabla\mathbf{u}_0)\,\mathrm{d}V}{\int_V\rho|\delta\mathbf{u}|^2\mathrm{d}V}, \tag{9}$$

which guarantees the existence of the solution of eq. (8) given that the numerator is positive (which is true in the small data situation i.e., the perturbation to the elasticity tensor is small). We may write the solution in the frequency domain in terms of the Green's function as follows:

$$\delta\mathbf{u}(\omega) = \int_\Omega\mathbf{G}\cdot(\nabla\cdot(\delta\mathbf{C}:\nabla\mathbf{u}_0))\mathrm{d}V, \tag{10}$$

which may further be simplified

$$\begin{aligned}\delta\mathbf{u} &= \int_\Omega\nabla\cdot(\mathbf{G}\cdot(\delta\mathbf{C}:\nabla\mathbf{u}_0))\mathrm{d}V - \int_\Omega(\delta\mathbf{C}:\nabla\mathbf{u}_0):\nabla\mathbf{G}\mathrm{d}V, &(11)\\ &= \int_{\partial\Omega}(\mathbf{G}\cdot(\delta\mathbf{C}:\nabla\mathbf{u}_0))\cdot\mathbf{n}\mathrm{d}S - \int_\Omega(\delta\mathbf{C}:\nabla\mathbf{u}_0):\nabla\mathbf{G}\mathrm{d}V,\\ &= -\int_\Omega(\delta\mathbf{C}:\nabla\mathbf{u}_0):\nabla\mathbf{G}\mathrm{d}V, &(12)\end{aligned}$$

where we have assumed that the perturbation vanishes on the boundary. Now we move on to use the expressions of the useful entities to obtain the observables of seismic anisotropy.

In a general, 3-D elastic continuum, three body waves exist: one primary wave (P) and two shear waves (S). In an isotropic medium, $\mathbf{SO}(3)$ symmetry of the elasticity tensor forces the eigenspectrum of the wave operator (eq. 8) to be degenerate. In other words, both the shear waves (SV and SH) travel with the same speed. However, in the presence of anisotropic scatterer, $\mathbf{SO}(3)$ symmetry is broken and therefore, the degeneracy is lost; two shear waves have different speeds. Thus, a phase delay between two such shear wave phases provides an observational constraint on the anisotropic properties of the medium.

In this study, we consider the SKS phase. The SKS phase is generated from a P to S conversion at the core–mantle boundary and therefore, it is supposed to be radially polarized (that is, polarization in the source–receiver plane) in an isotropic medium. The presence of anisotropy in the mantle, however, can transfer energy from the radial part to the transverse part of the wave displacement field via scattering leading to a non-zero amplitude of the transverse component. Thus, we may use this scattered transverse amplitude as a measure to describe shear wave splitting. Following Chevrot (2000), shear wave splitting intensity (S) is defined as

$$S(\mathbf{r},\mathbf{r}_s) = 2\frac{Re\int\iota\omega\delta u^T(\mathbf{r},\mathbf{r}_s;\omega)u_0^{*R}(\mathbf{r},\mathbf{r}_s;\omega)\mathrm{d}\omega}{\int\omega^2|u_0^R(\mathbf{r},\mathbf{r}_s;\omega)|^2\mathrm{d}\omega}, \tag{13}$$

where δu^T and u_0^R are the transverse and radial components of the displacement, respectively. Using this particular measure of anisotropic characteristic of the SKS wave form, we construct an integral map between the anisotropic perturbation to the elasticity tensor and the physical observable such as the one we have just defined. Now we move on to compute the integral kernels relating the derived observable (splitting intensity) and the unknown anisotropic scatterers.

Let us denote the unit vectors in the radial and transverse directions by $\hat{\mathbf{r}}$ and $\hat{\mathbf{t}}$, respectively. Here $\hat{\mathbf{r}}$ is the particle displacement direction (not to be confused with the radial direction in spherical polar co-ordinate system) associated with the SKS waves polarized in the radial plane (source-station plane). Then the transverse component of the scattered wave field is $\delta\mathbf{u}\cdot\hat{\mathbf{t}}$, which may be written in the following form using eq. (12)

$$\begin{aligned}\delta u^T = \delta\mathbf{u}\cdot\hat{\mathbf{t}} &= -\left(\int_\Omega(\delta\mathbf{C}:\nabla\mathbf{u}_0):\nabla\mathbf{G}\mathrm{d}V\right)\cdot\hat{\mathbf{t}}, &(14)\\ &= -\int_\Omega((\delta\mathbf{C}:\nabla\mathbf{u}_0):\nabla\mathbf{G})\cdot\hat{\mathbf{t}}\mathrm{d}V, &(15)\\ &= -\int_\Omega\hat{\mathbf{t}}\cdot(\nabla\mathbf{u}_0\otimes\nabla\mathbf{G})::\delta\mathbf{C}\mathrm{d}V. &(16)\end{aligned}$$

We may neglect the perturbation to the radial component, as such perturbation has only second order effect on the splitting intensity. This could be seen immediately from eq. (13) if one were to consider the perturbation to the radial component, that is $u^R = u_0^R + \delta u^R$: the contribution of such perturbation takes the form $\delta u^T\delta u*^R$, which is clearly second order and thus may be neglected. Under such assumption, the radial component is written as

$$u_0^R = \mathbf{u}_0\cdot\hat{\mathbf{r}}. \tag{17}$$

Now that we have obtained each component required to compute the splitting intensity, we may substitute eqs (16) and (17) into eq. (13)

$$\begin{aligned}S(\mathbf{r},\mathbf{r}_s) &= -2\frac{Re\int\iota\omega\left(\int_\Omega\hat{\mathbf{t}}\cdot(\nabla\mathbf{u}_0\otimes\nabla\mathbf{G})::\delta\mathbf{C}\mathrm{d}V\right)(\mathbf{u}_0\cdot\hat{\mathbf{r}})^*\mathrm{d}\omega}{\int\omega^2|(\mathbf{u}_0\cdot\hat{\mathbf{r}})|^2\mathrm{d}\omega}, &(18)\\ &= -2\frac{\int_\Omega Re\left(\int\iota\omega(\mathbf{u}_0\cdot\hat{\mathbf{r}})^*\hat{\mathbf{t}}\cdot(\nabla\mathbf{u}_0\otimes\nabla\mathbf{G})\mathrm{d}\omega\right)::\delta\mathbf{C}\mathrm{d}V}{\int\omega^2|(\mathbf{u}_0\cdot\hat{\mathbf{r}})|^2\mathrm{d}\omega}, &(19)\\ &= \int_\Omega\mathbf{K}::\delta\mathbf{C}\mathrm{d}V, &(20)\end{aligned}$$

where the sensitivity kernel $\mathbf{K}$, which relates perturbations in $\delta\mathbf{c}$ to splitting observation S, is given as follows:

$$\mathbf{K} = -2\frac{Re\left(\int\iota\omega(\mathbf{u}_0\cdot\hat{\mathbf{r}})^*\hat{\mathbf{t}}\cdot(\nabla\mathbf{u}_0\otimes\nabla\mathbf{G})\mathrm{d}\omega\right)}{\int\omega^2|(\mathbf{u}_0\cdot\hat{\mathbf{r}})|^2\mathrm{d}\omega}. \tag{21}$$

Here, '$\otimes$' represents tensor product operation on vector bundles (tensors are treated as sections of vector bundles). In deriving the previous equation, we have considered the fact that in purely elastic medium, the elasticity tensor is independent of frequency.

In order to calculate the sensitivity kernels, we need the Green's function. For a homogeneous, isotropic wave speed model, the

Green's functions can be solved analytically (Aki & Richards 2002). However, for a velocity structure that varies with depth, a semi-analytical technique such as a propagator matrix method works well (e.g. Gilbert & Backus 1966; Franssens 1983; Schmidt & Tango 1986). In our case, we experimented with two different methods for approximating the Green's function in the kernel computation. First, we used analytical solutions for the Green's function for a single-layer model, using average isotropic P and S wave speeds for the upper mantle [$c_p = 7.9\,\mathrm{km\ s^{-1}}$ and $c_s = 4.4\,\mathrm{km\ s^{-1}}$ (Cammarano *et al.* 2003)]. Second, we implemented the semi-analytical propagator matrix technique for a depth-dependent vertical velocity profile (ak135) (Kennett *et al.* 1995). In the propagator matrix technique, the fields are expanded in the Fourier series in the horizontal direction (no horizontal variation of seismic wave speed), and the frequency domain wave equation is converted to a set of ordinary differential equations in z. Further details can be found in Aki & Richards (2002). In our kernel computations presented in the paper, we used the propagator matrix technique applied to the depth dependent velocity model ak135; because it is a semi-analytical technique, is very efficient and allows for exploration of the model space with reasonable number of iterations. We point out, however, that because the variations in isotropic wave speed within the upper mantle are not dramatic, the differences between our propagator matrix computations for the Green's functions in a depth-dependent velocity model and the analytical Green's function for a homogeneous, single-layer model are slight. We compared kernels using the two different methods of computing the Green's function, and found that they were nearly identical.

3 PARAMETRIZATION

Once we have derived the expression for the sensitivity kernels (eq. 24), in theory we may invert eq. (23) to obtain the full anisotropic perturbation of the elastic medium given the splitting intensity (S). However, such an inversion would involve solving for each independent elastic constant in a fourth-rank tensor (21 in the most general case), which represents an impractically large number of unknown parameters given the limitations of observational data sets. If the symmetry of anisotropy is known or assumed, then the number of independent parameters is reduced (for example, for an olivine crystal with orthorhombic symmetry, the number of independent elastic constants is 9; in practice, this is still too large a number for an anisotropic tomography problem).

In previous work, Chevrot (2006) suggested a vectorial tomography which employs a linear inversion technique in order to estimate three Thomsen's anisotropy parameters (ε, δ and γ). These three dimensionless parameters are defined as follows for hexagonal anisotropy (for the case in which anisotropy is weak, that is$||\delta\mathbf{c}||_{L^p} << ||\mathbf{C}_{iso}||_{L^p}$, for some $p \geq 1$, Mensch & Rasolofosaon 1997)

$$\epsilon = \frac{c_{11} - c_{33}}{2\rho\alpha^2}, \tag{22}$$

$$\delta = \frac{c_{13} - c_{33} + 2c_{44}}{\rho\alpha^2}, \tag{23}$$

$$\gamma = \frac{c_{66} - c_{44}}{2\rho\beta^2}, \tag{24}$$

where ρ, α and β are the density, isotropic P, and isotropic S wave speed, respectively. Together with the orientation of the symmetry axis in space (dip angle from the horizontal, θ, and azimuth from north, ϕ), in this formulation there are a total of five parameters to invert, and additional simplifications are needed to reduce this number. Because typical shear wave splitting data sets cannot recover all five parameters simultaneously without producing a bias towards some parameters, such additional simplifications are necessary for our problem.

Our approach is to reparametrize the strength of anisotropy using insights from mineral physics, specifically experimental and observational constraints on the development of crystallographic preferred orientation (CPO) in olivine aggregates. Seismic anisotropy of the upper mantle is thought to be mainly controlled by this mechanism (Karato *et al.* 2008); furthermore, olivine aggregates that have been deformed to high strains develop CPO such that the elasticity of textured aggregates has a geometry (though not a strength) that is generally similar to the elasticity of single-crystal olivine (e.g. Karato *et al.* 2008). We therefore assume that olivine aggregates in the upper mantle will exhibit behaviour characteristic of a linear mixture of two endmembers: an isotropic equivalent, in which the individual grains are randomly oriented, and a perfectly aligned aggregate, in which all grains exhibit the same (preferred) orientation and the elasticity is equivalent to a single crystal of olivine. In this scheme, the strength of anisotropy can be described with a single parameter that varies from 0 (random alignment) to 1 (complete alignment). We further assume that the elasticity of the aligned olivine endmember can be well described by a hexagonal approximation to the actual (orthorhombic) symmetry, an assumption that is justified by the elastic characteristics of single-crystal olivine, along with experimental results on deformation of polycrystalline olivine aggregates (e.g. Karato *et al.* 2008).

Our goal is to obtain a hexagonal equivalent of the orthorhombic olivine tensor. There are several ways of doing so; one way is to use the projection method described in Browaeys & Chevrot (2004). An alternative (but conceptually equivalent) method to obtain the hexagonal part of the orthorhombic tensor is to rotate about the [100] (fast) olivine axis and average the elements of the elasticity tensor describing the intermediate and slow velocities, as described below. We have tested both schemes and they yield virtually identical results; results shown in this paper were obtained using the latter approach. We carry out hexagonal averaging of the elasticity tensor from the lower (orthorhombic) symmetry class by taking the angular average of each of the three sets (c_{11}, c_{22}), (c_{44}, c_{55}) and (c_{13}, c_{23}). This averaging scheme utilizes the fact that elements of each set do not differ substantially from each other (that is, the orthorhombic symmetry is weak on the plane perpendicular to the c axis). The spherical harmonic averaging is done using

$$C_{11hexagonal} = \frac{\sum_{l=0}^{\infty}\int_0^{2\pi}(C_{11orthorhombic}(\Phi) + C_{22orthorhombic}(\Phi))P_l(\cos\Phi)\mathrm{d}\Phi}{2\sum_{l=0}^{\infty}\int_0^{2\pi}P_l(\cos\Phi)\mathrm{d}\Phi}, \tag{25}$$

where P_l's are the ordinary Legendre polynomials and Φ is angle of rotation about c axis. Similarly, other sets [(C_{44}, C_{55}) and (C_{13}, C_{23})] may also be averaged to obtain the hexagonal equivalence. Following homogeneous stress approximation (Reuss average; Reuss 1929), the elastic constants of the mixture C_{ij} may be expressed as

$$C_{ij} = \chi C_{ijsingle} + (1 - \chi)C_{ijIsotropic}, \tag{26}$$

where χ is the fraction of pure hexagonal crystal present in the mixture. This formulation uses the technique of Browaeys & Chevrot

(2004) in order to extract the isotropic part of the elasticity tensor from the full olivine anisotropic tensor. Using this particular approximation, we may compute the three Thomsen's parameters based on eqs (22)–(24) and note their evolution as a function of χ (Fig. 1). When the mixture consists of completely aligned crystals ($\chi = 1$), the Thomsen's parameters are that of (hexagonally averaged) olivine, and for the purely isotropic crystal ($\chi = 0$), they are identically zero.

We checked whether the Thomsen parameters that we computed for the hexagonally averaged crystal are close to those presented in Becker *et al.* (2006). This comparison is imperfect, however, because Becker *et al.* (2006) used an aggregate of olivine and enstatite, rather than the pure olivine aggregate that we use in our work. The values for the Thomsen parameters in our work are therefore not identical, but they are similar. Thomsen's parameters (ε, δ, γ) vary linearly with χ, as shown in Fig. 1(a). Figs 1(b)–(d) show the correlation between the Thomsen parameters. We also considered the correlations among different Thomsen parameters, both for our result and the one presented in Becker *et al.* (2006), and again the relationships were generally similar. Because in our approach we used a linear mixture between the isotropic and (hexagonal) anisotropic crystals, a linear correlation between the anisotropy parameters is expected, and this is indeed what is observed.

Using eqs (20), (21), (22)–(24) and (26) we may obtain the following simplified equation relating the splitting intensity and the 'strength' of anisotropy

$$S(\mathbf{r}) = \int_{\Omega} K_{\chi}(\mathbf{r}, \theta, \phi; \mathbf{r}_s)\chi(\mathbf{r}_s) d^3\mathbf{r}_s, \tag{27}$$

where K_{χ} is the corresponding sensitivity kernel, which is itself a non-linear function of the orientation parameters (θ, ϕ). One important thing is to note is that the eq. (27) is only valid under the assumptions made previously; for the case of relatively weak anisotropy due to CPO of olivine in the upper mantle, this equation yields a good approximation for the relationship between anisotropic model perturbations and the splitting intensity S. Under these conditions, we can represent the anisotropic geometry and strength with three parameters (θ, ϕ, χ), instead of the 21 independent parameters needed for the most general case of anisotropy.

4 EXAMPLES AND PROPERTIES OF SENSITIVITY KERNELS

The strong dependence of the sensitivity kernel on the background model is an important issue in shear wave splitting tomography (e.g. Long *et al.* 2008), and this dependence makes the linearization of the inverse problem difficult, as discussed in Section 5. In this section, we explore the behaviour of sensitivity kernels and provide some simple examples of their strong dependence on the background model. The kernel behaviour is illustrated via figures that show the variation of the sensitivity in the model space as functions of the background model properties (azimuth and dip of the symmetry axes of the anisotropic medium). In addition to representing linear maps between the model space and the splitting intensity observable, sensitivity kernels provide insights into the first-order behaviour of the wavefield in the anisotropic medium.

We choose the power spectrum of the incoming SKS wave to be similar to that of Long *et al.* (2008), that is

$$|\mathbf{u}_0^R(\omega)|^2 = \frac{\omega^2\tau^2}{4\pi} e^{-\frac{\omega^2\tau^2}{8\pi^2}}, \tag{28}$$

where τ is the characteristic period of the wave, which is taken to be 10 s for SKS phase. The integral over the frequency in eq. (21) is performed over the frequency range of 0.02−0.125 Hz which is the same as used by Long *et al.* (2008) and typical for SKS splitting studies (e.g. Long & Silver 2009b). We approximate the spectrum as the second derivative of a Gaussian; this is reasonable for SKS-type phases where the records have been deconvolved from the radial component waveform (Chevrot 2000; Long & van der Hilst 2005). With this choice of unperturbed wave spectrum, we compute the sensitivity kernels with different background anisotropic model. The kernels are computed on a fine grid [(5 km)^3 box] for two different cases: (i) variation of the dip of the symmetry axis and (ii) variation of the azimuth of the symmetry axis. The incoming polarization direction is kept constant in both of these cases. The width of the maximum sensitivity zone is calculated as $\approx\sqrt{\lambda L}$, where λ and L are the characteristic wavelength and the propagation distance, respectively. In our case, with the characteristic period of 10 s and approximate propagation distance of 400 km, we find the width of the maximum sensitivity zone to be ≈300 km. However, the nature of the sensitivity varies with each cases discussed below.

We compute a series of sensitivity kernels for different values of the azimuth of the symmetry axis (ϕ) while the dip ($\theta = 0^\circ$) and the incoming polarization direction ($\alpha = 0^\circ$) are kept fixed. The anisotropic perturbation is homogeneous in nature. The spatial distribution of the sensitivity is not substantially modified with the variation of azimuth (ϕ) of horizontal symmetry axis. Figs 2(B) and (C) show near symmetric sensitivity kernels corresponding to $\phi = 30^\circ$ and $\phi = 60^\circ$, respectively. However, the end member cases of $\phi = 0^\circ$ (polarization is parallel to the symmetry axis) and $\phi = 90^\circ$ (polarization is perpendicular to the symmetry axis) are quite interesting. While the latter corresponds to a trivial sensitivity kernel (Fig. 2D) leading to a zero splitting as expected, the former still has a non-trivial but perfectly antisymmetric nature (Fig. 2A), leading to a zero splitting intensity while integrated.

Fig. 3 shows the variation of the sensitivity kernel with different values of the dip of the symmetry axis while the incoming polarization direction ($\alpha = 0^\circ$) and the azimuth of the symmetry ($\phi = 45^\circ$) are fixed. The presence of the inclined symmetry axis causes an uneven spatial distribution of the sensitivity, as expected (Figs 3B and C). The striking property of the sensitivity kernel is the non-triviality even when the propagation direction is parallel to the symmetry axis (Fig. 3D). Such a physical scenario should lead to a zero splitting intensity and the perfect antisymmetry of the sensitivity kernel (Fig. 3D) confirms this well known result (e.g. Silver 1996; Chevrot 2000; Long & Silver 2009b). This non-trivial antisymmetry of the sensitivity kernel is a remarkable property and it may help us to understand the underlying physics in more details.

The kernel examples shown in Figs 2 and 3, which show the sensitivity for a range of angles between the fast direction and the incoming polarization direction, illustrate some interesting and subtle features for the case in which the initial polarization (equivalent to the backazimuth for SKS waves) is aligned either parallel or perpendicular to the fast axis of the medium. For the case in which the wave propagation direction (i.e. the wave vector **k**) is parallel to the symmetry axis, we would observe a zero splitting intensity (e.g. Silver 1996; Chevrot 2000; Long & Silver 2009b). Zero splitting intensity may also be obtained in different scenarios, such as symmetry axis being parallel or perpendicular to the polarization direction. Therefore, each of these three cases predicts the same observation (zero splitting intensity). However, the sensitivity kernels in Fig. 2 illustrate that interactions of waves with

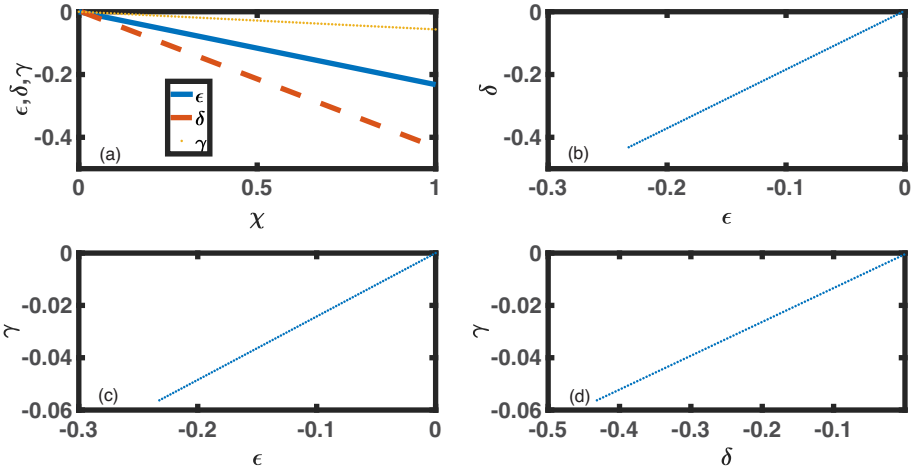

Figure 1. (a) Variation of the three dimensionless Thomsen parameters (ε, δ, γ) with the strength parameter χ, (b) Correlation between ε and δ, (c) Correlation between ε and γ, (d) Correlation between δ and γ. Our parametrization scheme uses χ instead of the Thomsen parameters to define the strength of upper mantle anisotropy, taking advantage of the linear relationships between χ and ε, δ and γ.

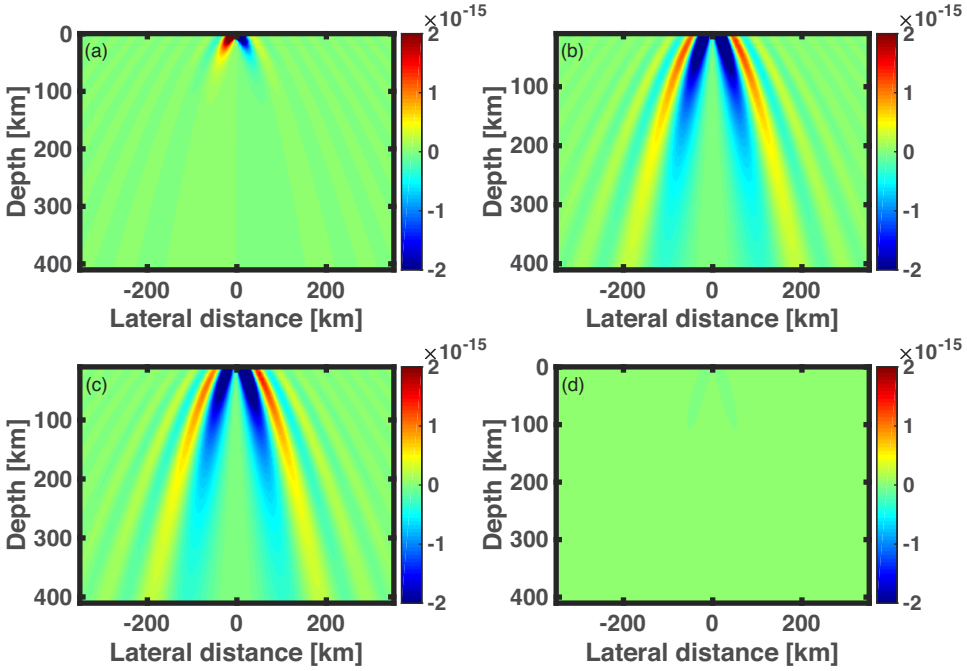

Figure 2. Sensitivity kernel examples: 2-D slice along the plane of propagation. Plots show behaviour of the sensitivity kernels at different values of the azimuth of the symmetry axis (ϕ) while the dip of the symmetry axis ($\Theta = 0^\circ$), and the incoming polarization direction ($\alpha = 0^\circ$) remain fixed. (a) symmetry axis is parallel to the polarization direction ($\phi = 0^\circ$): the kernel is perfectly antisymmetric leading to a zero splitting intensity for a homogeneous anisotropic model, (b) $\phi = 30^\circ$:, (c) $\phi = 60^\circ$: symmetric kernel and (d) symmetry axis is perpendicular to the polarization direction ($\phi = 90^\circ$): trivially zero sensitivity kernel yielding zero splitting intensity for arbitrary spatial distribution of anisotropic strength. Unit of the sensitivity is $\mathrm{m}^{-3}\mathrm{s}$.

the anisotropic perturbations corresponding to these three different cases are different. In case of a trivial sensitivity kernel (Fig. 2D), the constructive and destructive interferences cancel each other in either side of the vertical ray path. However, in the case of a perfectly antisymmetric kernel (Figs 3D and 2A), the opposite wave phases accumulate in both sides of the raypath with opposite polarity and thus, the sensitivity is non-vanishing. Therefore, in such special case, the finite frequency approach would be able to sense the anisotropic perturbation even if the observable is zero (corresponds to a non-trivial solution to the system of linear equations described by eq. 27). Such different interactions lead to ambiguous interpretations while ordinary ray theory is employed, because of the integration of the structure along the ray path. However, using a tomographic approach, we may be able to delineate the three distinct cases.

We may explain the importance of the non-trivial sensitivity kernel yielding zero sensitivity in tomography by invoking a simple concept of linear maps. We provide here a simplified description of the concept, and interested readers may consult Bourbaki (1966) for more involved concepts of linear maps and quotient spaces. We may

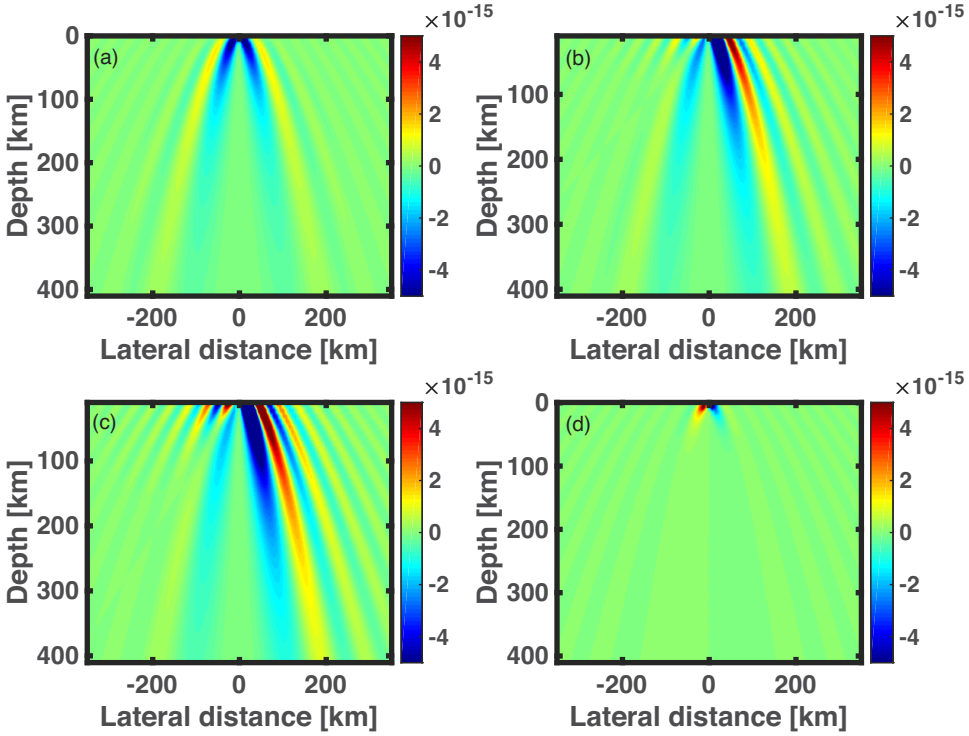

Figure 3. Sensitivity kernel examples: 2-D slice along the plane of propagation. Plots show behaviour of the sensitivity kernels at different values of the dip of the symmetry axis (Θ), while the azimuth of the symmetry axis ($\phi = 45^\circ$) and the polarization of the incoming wave ($\alpha = 0^\circ$) are kept fixed. (a) 2-D slice through a 3-D sensitivity kernel for a model with a horizontal symmetry axis ($\Theta = 0^\circ$): almost fully symmetric kernel and thus sensitivity is evenly distributed, (b) tilted symmetry axis ($\Theta = 30^\circ$), (c) $\Theta = 60^\circ$, and (d) vertical symmetry axis ($\Theta = 90^\circ$): perfectly antisymmetric sensitivity kernel yielding zero splitting intensity for homogeneous anisotropic model. Unit of the sensitivity is $\mathrm{m}^{-3}\mathrm{s}$.

write the eq. (27) as a linear map between the model space X (containing all possible anisotropic perturbations) and space of observable Y (all possible splitting intensity in this case), that is $\Psi: X \rightarrow Y$, where Ψ is the linear operator $\int_V dV K_X$. In case of a trivial sensitivity kernel $\Psi(x) = 0, \forall x \in X$, that is both sides of the eq. (27) are identically zero and thus, one cannot estimate the model parameters at all. However, zero splitting intensity may also be obtained by integrating the perfectly antisymmetric kernels with a heterogeneous strength of anisotropy (in other words, if the strength of the anisotropy is distributed in such a way that the integration of eq. (27) yields zero). Mathematically, in this case, we get a non-trivial anisotropic perturbation which consists the kernel of linear operator Ψ (kernel of a map is given as follows: $\Psi: X \rightarrow Y$, the kernel of Ψ is the subset of X for which Y vanishes i.e. kernel(Ψ)=$\{x \in X | \Psi(x) = 0\}$). In the splitting measurements, the null data (zero splitting intensity) are neglected. However, we do not know *a priori* the orientation of the symmetry axis (we wish to know this), and ray theory does not distinguish between these different situations in which the splitting intensities turn out to be zero. Thus, in ray theory, we eventually neglect a part of the model space or just resolve the quotient space $X/ker(\Psi)$. In our method, whenever sensitivity kernel is non-trivial, kernel of sensitivity kernel is estimated uniquely. Thus, the method overcomes the disadvantage of neglecting a portion of model space.

In order to benchmark our framework for computing sensitivity kernels, we compare our predictions for splitting intensity observations with known (ray theoretical) solutions for simple model cases, such as that of a single anisotropic layer with a horizontal symmetry axis. When the symmetry axis is horizontal, simple analytical solutions exist (Chevrot 2000) and we compare our solution with such limiting cases. Specifically, for the case of a single horizontal layer of anisotropy, the splitting intensity is predicted to vary as:

$$S = \delta t \sin[2(\alpha - \phi)], \tag{29}$$

where δt, α and ϕ are the delay time, incoming polarization backazimuth, and azimuth of the symmetry axis, respectively. We compute sensitivity kernels for a series of initial polarization directions for a simple, single layer model and predict splitting intensities by integrating over the model volume. Figs 4(A) and (B) show these modeled splitting intensities, along with predictions from eq. (29), demonstrating that our kernels capture the expected behaviour for the simplest models well. Specifically, we predict sinusoidal behaviour of the splitting intensity with backazimuth, as predicted by Chevrot (2000). We also compute predicted splitting intensities for homogeneous models with a non-horizontal horizontal symmetry axis to demonstrate the effect of a tilted symmetry axis (Figs 4C and D). These figures show that the splitting intensity maximum corresponds to a dip of 45°, consistent with the results of Chevrot & Van Der Hilst (2003).

5 INVERSION STRATEGY: MODEL SPACE SEARCH APPROACH

We have discussed in the previous section the non-linear relation between the splitting intensity and the anisotropy parameters (θ, ϕ) through the sensitivity kernel examples. Thus, any attempt in linearizing such parameters would severely bias the model estimation. The application of a non-linear search which does not rely on the starting model becomes desirable. Therefore, our inversion strategy invokes a model space search technique based on a Markov chain

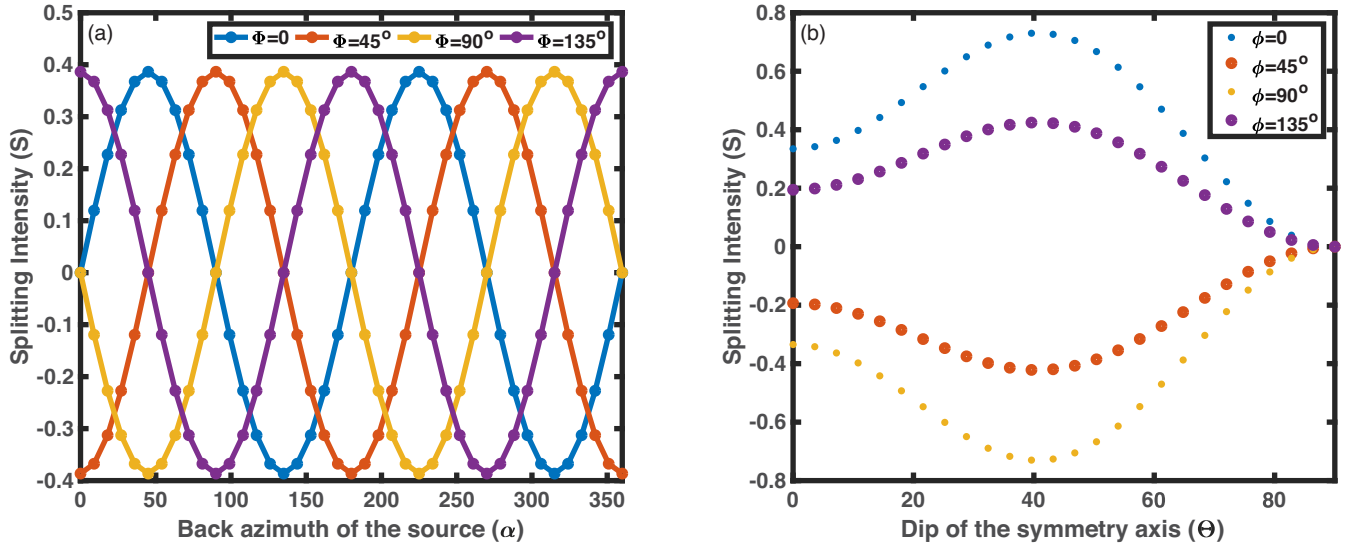

Figure 4. Behaviour of sensitivity kernels against analytical solutions. We compute splitting intensity by integrating the sensitivity kernels (27) for a series of event backazimuths for a model with single layer of horizontal anisotropy. Kernel predictions (dots) are plotted against the predicted splitting intensity from analytical solutions (line) with a horizontal symmetry axis (29) as a function of (a) incoming polarization back-azimuth (α) of the wave at different azimuth of the symmetry axis (ϕ). The variation of splitting intensity as a function of dip of the symmetry at (b) different azimuths of the symmetry axis. In (b), maximum splitting is obtained at $\theta \approx 45^\circ$ consistent with the result of Chevrot & Van Der Hilst (2003). The starting model used for these computations has Thomsen parameters of $\varepsilon = 0.03$, $\delta = 0.04$ and $\gamma = 0.01$ (corresponding to $\chi = 0.1$, Fig. 1) and with a layer thickness of 400 km.

Monte Carlo sampling algorithm. It is an efficient model space search technique based on Bayes' theorem and Markov process. Bayes' theorem describes how to update the probabilities of hypotheses given the evidence, and follows simply from the axioms of conditional probability. Let us assume that we have a data vector d which is given or observed, and several models can reproduce this data within different error limits. Therefore, we may construct a conditional probability: given the data d, what is the probability that a particular model m would satisfy it? The probability of the assumed model m satisfying the given data d is the conditional probability $P(m/d)$. Clearly, if this conditional probability is high enough, we may assert that the model fit is reasonably good. Now, from Bayes' theorem, we may write the following identity

$$P(m/d) = \frac{P(d/m)P(m)}{\sum_m P(d/m)P(m)}, \tag{30}$$

where $P(d/m)$ is the probability of achieving the given data with the chosen model parameter m, and $P(m)$ is probability of the model parameter m to exist. The denominator describes the total conditional probability of observing this particular data. Clearly, if we choose the model parameters from the physically plausible set, then $P(m)$ is identically 1, and 0 if it is chosen from outside of the set. Bayes' theorem is applied in conjunction with the Markov process, which states that in a sequence of N events, the outcome of ith event depends only on the outcome of $(i-1)$th event. Following these two basic notions, several sampling algorithms have been developed, including the Metropolis-Hastings algorithm, the Gibbs sampling algorithm, and several adaptive algorithms. We implement the Gibbs sampling technique (Liu 2008), which allows us to sample the models with higher probability efficiently. Such technique does not involve the concept of a starting model, as the model space is sampled randomly following a particular probability distribution. The details of the Gibbs algorithm are not presented here, as it has been extensively applied in computational physics (e.g. Foulkes *et al.* 2001; de Saavedra & Kalos 2003), computational chemistry (e.g. Hammond *et al.* 1994), and geophysics (e.g. Mosegaard & Sambridge 2002; Korenaga & Karato 2008; Wirth *et al.* 2016). The misfit is chosen to be the ordinary Euclidean distance between the observed data and synthetic data produced by the chosen model sets weighted with the squared data error vector:

$$E^2 = \sum_{i=1}^{N} \frac{[d_i - f_i(\mathbf{m})]^2}{2\sigma_i^2}, \tag{31}$$

where σ_i is the data error vector. The Bayesian likelihood function $P(d/m)$ is just the Gaussian distribution (adopted in accordance with the central limit theorem):

$$P(d/m) = \exp(-E^2). \tag{32}$$

In our approach, we avoid the Metropolisn–Hastings algorithm, as the convergence of such algorithm is extremely sensitive to the step size of the random jump. Gibbs sampling, on the other hand, has an advantage of sampling all misfit wells (or, correspondingly, probability hills, e.g. Korenaga & Karato 2008). However, in problems such as tomographic inversions, it is extremely difficult to resolve small scale features as the number of parameters increases. In such cases, inversion becomes computationally expensive (for example, three parameters defined at each spatial point leads to a total of 300 parameters with a grid size of 50 km × 50 km × 50 in a 500 km × 500 km × 50 km domain) and thus the Markov chain requires larger number of iteration to converge. The power of our approach, and the number of parameters that can reasonably be resolved, thus depend on the available computational resources. We will discuss the advantages and limitations of model space search technique in the following sections using the results of resolution tests.

6 SYNTHETIC TESTS

In this section, we provide several examples of synthetic tests to show how well our method can handle different anisotropic scenarios. We start with the simplest model (a single homogeneous layer of anisotropy) and increase the complexity thereby increasing size of

the parameter space. This sequential increase in complexity in our resolution tests helps us to illustrate the possible length scales and size of the parameter space which can be resolved with confidence. In this study we restrict ourselves to a grid based parametrization (discrete). There are other possible parametrizations as well. One such possibility would be representing the anisotropic strength by linear combination of basis functions in proper function space (e.g. L^2 orthonormal basis) and solving for the coefficients. Such alternative parametrization will be explored in future work. A summary of the specifications for each resolution test presented in this paper can be found in Table 1.

6.1 Single layer case

A single layer model is our simplest scenario, but it is important in order to understand the underlying complexity of the data-model parameter relation. We need to estimate three parameters: the strength of anisotropy (χ), the dip (θ) and the azimuth (ϕ) of the symmetry axis. We choose a physical domain with size $\approx$800 km $\times$ 800 km (in the horizontal dimensions) $\times$400 km (in the vertical dimension, Fig. 5 A), designed to be typical of a regional anisotropy inversion. We generate a set of 25 randomly distributed station locations (Fig. 5), and compute a synthetic data set of splitting intensities using an azimuthally random distribution of 20 sources and a starting model with parameters [$\chi = 0.1$, $\theta = 45^{\circ}(0.7854\ rad)$, $\phi = 30^{\circ}(0.5236\ rad)$]. We use this set of synthetic splitting intensities (see example in Fig. 5D) as the input for our inverse model [i.e. to invert eq. 27 for (χ, θ, ϕ)]. Using this input model, we ran a model space search for a total of 8000 iterations (length of the Markov chain); this number of iterations was chosen after a series of tests to investigate how many iterations were needed for the algorithm to converge. For this single-layer test in which all three parameters were allowed to vary, the sensitivity kernels were recomputed at each step for each candidate starting model; the inversion thus comes at a non-trivial computational cost, despite the small number of parameters. This particular initial test ran for 6 hours on a laptop computer with 8 cores.

The results of our initial single-layer recovery test are shown in Fig. 6, which shows the misfit function for a fixed χ (projection of the 3-D error hyper-surface onto $\chi = constant$ plane is used in order to represent graphically) and the points sampled by the MCMC algorithm. The shape of the 'valley' of misfit indicates that multiple choices of θ (dip of the symmetry axis) explain the data equally well. The posterior distribution also shows that the maximally sampled points may not correspond to the actual dip (θ) (Fig. 6C). However, the anisotropy strength (χ) and the azimuth of the symmetry axis (ϕ) are extremely well resolved (Figs 6B and C) for the single-layer case.

Despite the fact that splitting intensity depends on the strength of anisotropy (χ) in a linear fashion (eq. 27), the non-zero covariance among parameters makes a non-linear search crucial, even for a hypothetical inversion cases in which χ is the only parameter of interest. These parameter covariances are demonstrated in Fig. 7(B), which shows a power law relation $\theta \approx 0.1\chi^{-1.8}$ between χ and θ when χ is small (within a few standard deviations from the original value of 0.1).This demonstrates that the presence of a dipping symmetry axis impedes the convergence of the model space search algorithm, as well as presenting a source of non-uniqueness in the inversion. We emphasize that our inability to constrain the dip of the symmetry axis in our synthetic test is not unique to our study and is not a consequence of our model space search approach; rather, it is due to our use of nearly vertically propagating SKS waves. On the other hand, the resolved azimuth (ϕ) is insensitive to (that is, does not trade off with) the strength (χ) and the dip of the symmetry axis (θ) (Figs 7C and D). This implies that SKS splitting observations can distinguish variability in ϕ from variability in other parameters, and furthermore suggests that strong constraints on ϕ can be achieved very fast through the McMC technique. In many geophysical studies of anisotropy, a horizontal symmetry axis is either suggested by the data or (more commonly) assumed. Therefore, we suggest a strategy that assumes a horizontal symmetry axis the most appropriate configuration for many problems; if this is assumed, our approach can effectively provide a robust estimation of the strength of the anisotropy and the fast axis direction for simple models.

The test shown in Fig. 6, like most of the test cases we present in our paper, was carried out with noise-free synthetic data. In order to investigate the effect of random noise on our inversion scheme, we carried out a second version of the single-layer test with the same parameters describing the input model, but with Gaussian-distributed noise (standard deviation of 0.1) applied to the synthetic splitting intensity observations. For noisy data, it is necessary to run the McMC algorithm for a larger number of iterations to achieve convergence; for the example shown in Fig. 8, we ran the algorithm for 12 000 iterations (12 000 versus 8000 for Fig. 6). Even with a larger number of iterations, this test demonstrates that the McMC algorithm demonstrates output with a greater uncertainty than that of noise-free data, as expected. For this test, we found that the most likely model for the strength parameter χ is somewhat different (0.129), from the input strength (0.1), with larger uncertainty than for the corresponding noise-free case. This test demonstrates that even for modest amounts of noise, the convergence of the McMC algorithm will be slower, which can represent a practical limitation for real data.

6.2 Fully heterogeneous case with 3-D structure

Next we perform a series of synthetic inversion tests with increasing model complexity, including lateral heterogeneities in the anisotropic strength parameter χ. For the first set of these tests, we focus on input models that allow for lateral variations in χ, but not in ϕ; in all of our subsequent tests, the dip parameter is fixed at zero, such that the symmetry axis is horizontal. Initially, we choose an input model with four distinct blocks within the model domain ($2 \times 2 \times 1$) with alternating strength χ (i.e. two subsequent blocks have different strength). Each block of this input model is of size $\approx 350 \times 350 \times 400\ km^3$. Using this initial model, we compute a series of splitting intensities with the station-source distribution same as that used in our one-layer test (Section 6.1). For this series of tests, we parametrize the model space with a grid size of $\approx 200 \times 200 \times 400\ km^3$, that is 16 blocks model ($4 \times 4 \times 1$).

Fig. 9(A) shows the retrieved anisotropic strength and the posterior distribution corresponding to the best and worst resolved grid for this checkerboard test (with 4 input blocks). Our result suggests that the distribution of stations greatly affects the resolution. In this particular test, the average station spacing of $\approx$100 km yields a fairly good resolution for the chosen grid size, but there is variability: grid blocks that happen to have good station sampling are well resolved (see 'best resolved' histogram in Fig. 9A), while grid blocks with few stations that happen to be located along the edges of the block are poorly resolved. This is demonstrated by a similar test that has the same input model but a different (and much denser) station distribution. In addition to our tests with a random spatial distribution

Table 1. A summary of the specifications for each resolution tests presented in this paper.

Test number	Type of variability	Station configuration	Parameters varied	Grid size	Type/size of input anomalies	Figure number
1	one layer	random	χ, θ, ϕ	$\approx 800^2 \times 400\,\text{km}^3$	single layer	6
2	2D (lateral)	random	χ	$\approx 200^2 \times 400\,\text{km}^3$	$\approx$400 km checkerboard	9(A)
3	2D (lateral)	HLP	χ	$\approx 100^2 \times 400\,\text{km}^3$	$\approx$400 km checkerboard	10(A)
4	2D (lateral)	HLP	χ, ϕ	$\approx 200^2 \times 400\,\text{km}^3$	$\approx$400 km checkerboard	12(A)
5	2D (lateral)	HLP	χ	$\approx 50^2 \times 400\,\text{km}^3$	$\approx$100 km spike	14(A)
6	2D (vertical)	HLP	χ	$\approx 50^2 \times 400\,\text{km}^3$	$\approx$50 km checkerboard	14(B)
7	Fully 3D	random	χ	$\approx 200^2 \times 200\,\text{km}^3$	$\approx 400^2 \times 200\,\text{km}^3$ checkerboard	9(B),(C),(D)
8	Fully 3D	HLP	χ	$\approx 200^2 \times 100\,\text{km}^3$	$\approx 400^2 \times 200\,\text{km}^3$ checkerboard	10(B),(C),(D)
9	Fully 3D	HLP	χ, ϕ	$\approx 200^2 \times 100\,\text{km}^3$	$\approx 400^2 \times 200\,\text{km}^3$ checkerboard	12(B)

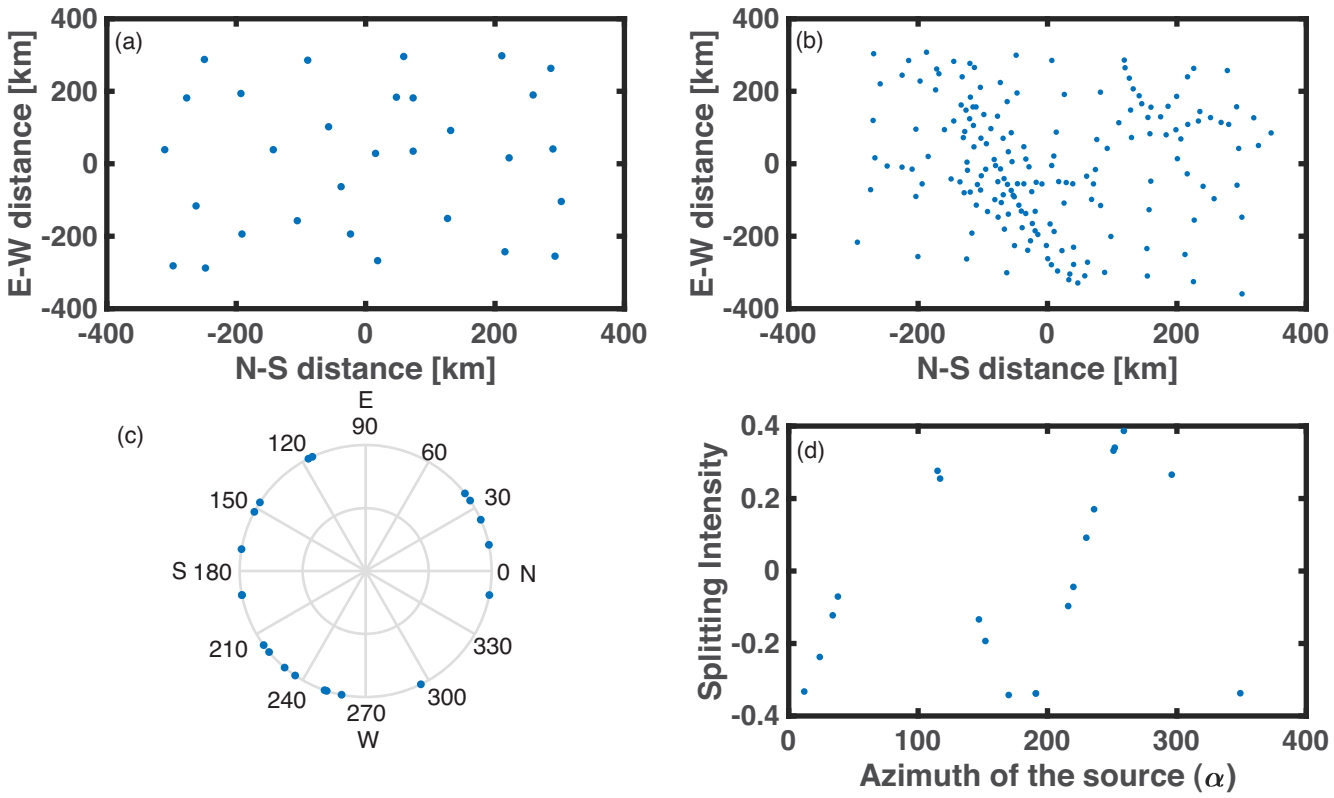

Figure 5. (a) The station distribution is shown on the horizontal plane (blue dots); the average station spacing is roughly 100 km. (b) The station distribution of the dense high lava plain seismic network. (c) Azimuthal distribution of the randomly chosen 20 seismic events. (d) Synthetic splitting intensity as a function of azimuth of the events for a particular station.

of 25 stations, we also carried out resolution tests using the actual station locations of the dense High Lava Plains seismic network (Long *et al.* 2009; Wagner & Long 2013), which operated Oregon, Nevada and Idaho (USA) between 2006 and 2009. Specifically, we used the identical station configuration as Long *et al.* (2009), which incorporated HLP stations as well as surrounding USArray Transportable Array stations and the nearby Wallowa Mountains array. The station distribution for this synthetic experiment (Fig. 5B) is irregular, with station spacing as fine as 10–15 km along the dense HLP lines (oriented WNW–ENE and N–S) but as coarse as 70 km in the portions of the model domain that are sampled mainly by TA stations. We carried out a test with the same input model as that shown in Fig. 9(A), but with a finer grid parametrization (8 × 8 × 1). The results of this test are shown in Fig. 10(A), which presents the input–output pattern along with the station distribution and the posterior distribution corresponding to the best and worst resolved grid. As with the random station distribution, this test makes it clear that regions of the model with poor station sampling yield highly uncertain parameter estimates, while model blocks with good station sampling are well resolved (compare best-resolved and worst-resolved posterior distributions in Fig. 10A).

The tests shown in Figs 9(A) and 10(A) only involved variability in the strength parameter in the input model; however, it is desirable to carry out inversions for χ and ϕ simultaneously. In Fig. 12(A), we show the results of such a synthetic test, in which we vary both χ and ϕ, keeping $\theta = 0$, with a 4 × 4 × 1 block model; the total number of parameters is thus $16 \times 2 = 32$. The station distribution is again taken to be the that of the HLP experiment and surrounding stations, following Long *et al.* (2009). The input–output model is presented in Fig. 12(A). As in other laterally heterogeneous cases (Figs 9 A and A), grid blocks with a dense station distribution (number of station ≥4) are resolved fairly well, while grid blocks

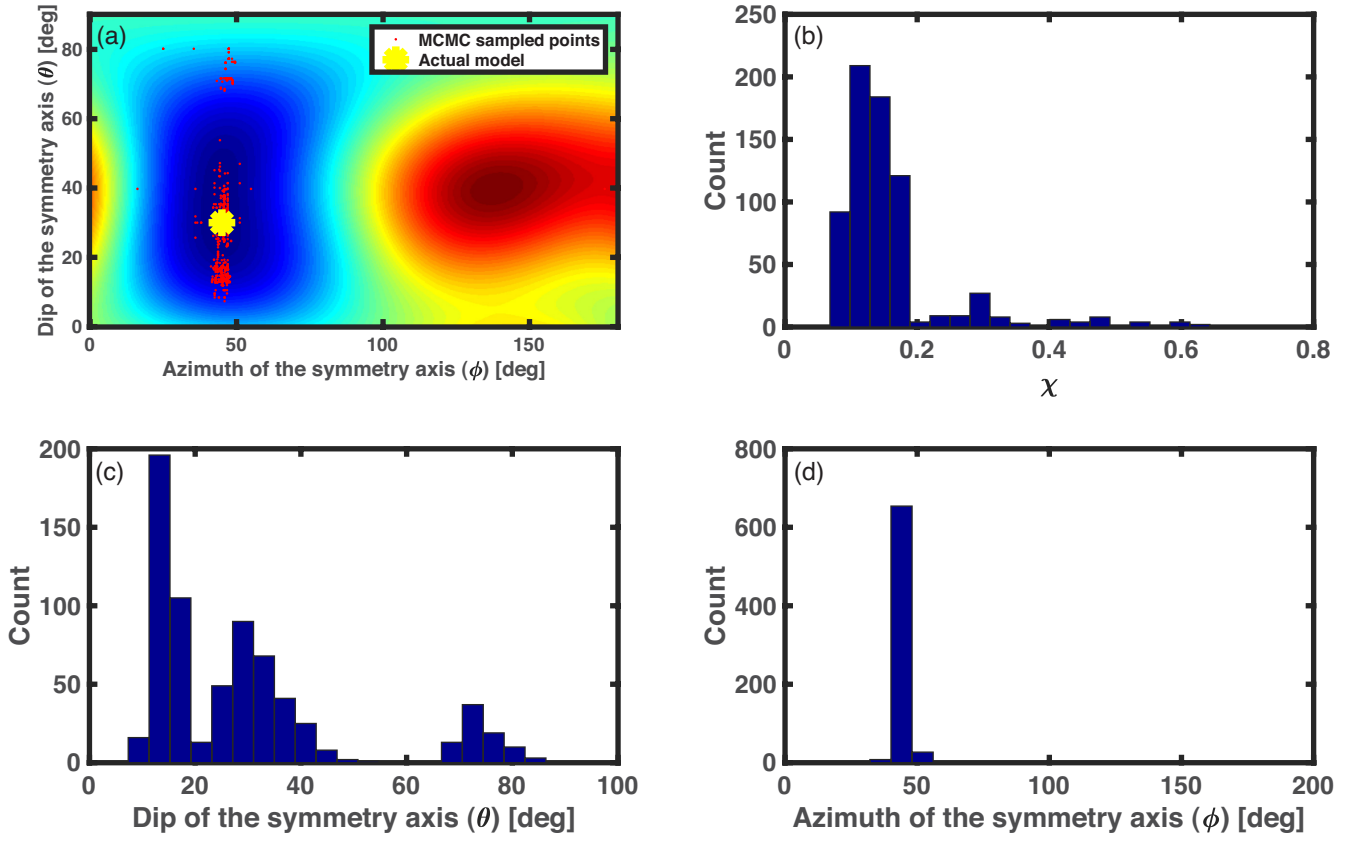

Figure 6. Single layer inversion result. The input model consists of a single layer of anisotropy with $\chi = 0.1$, $\theta = 0.52$ rad (30°) and $\phi = 0.78$ rad (45°). (a) 2-D section of the normalized squared misfit (colours) and the points (red dots) sampled by the MCMC inversion algorithm. The yellow dot corresponds to the location of the true input model. (b) The posterior distribution of the strength parameter χ with a mean ($\bar{\chi}$) of 0.1075 and standard deviation (σ) of 0.0277. (c) The posterior distribution of the dip of the symmetry axis (θ) with a mean of 0.4419 (25.4°) and standard deviation of 0.1658 (9.5°). (d) The posterior distribution of the azimuth of the symmetry axis (ϕ) with a mean of 0.7933 (45.4527°) and standard deviation (σ) of 0.1152 (6.6°).

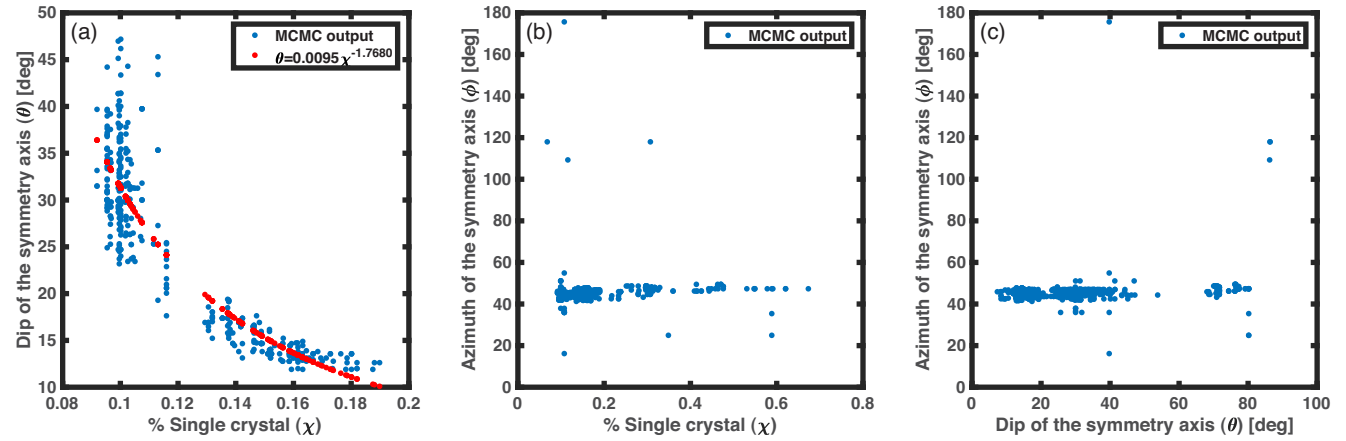

Figure 7. Estimation of trade-offs among parameters. (a) McMC sampled dip (θ) is plotted against the MCMC sampled strength (χ), clearly showing a correlation and thus a parameter trade-off. (b) The nearly horizontal trend between the MCMC sampled azimuth (ϕ) and the MCMC sampled strength (χ) depicts that the former is insensitive to the later and (c) similar trend is observed between MCMC sampled azimuth (ϕ) and the MCMC sampled dip (ϕ).

devoid of stations do not produce any constrained result. This is shown in detail in Fig. 13, which presents the posterior distribution of the symmetry axis azimuth (ϕ) for four different grid blocks in the inversion, some of which are well constrained and others of which are poorly constrained.

Our last depth-independent recovery test consists of a spike test, in which two distinct highly anisotropic anomalies of dimension 150 km × 200 km inserted in the model space, and we explore how well the algorithm can recover them. For this test, θ and ϕ are kept constant, with $\theta = 0$ and $\phi = 45°$), and only χ is varied in the input

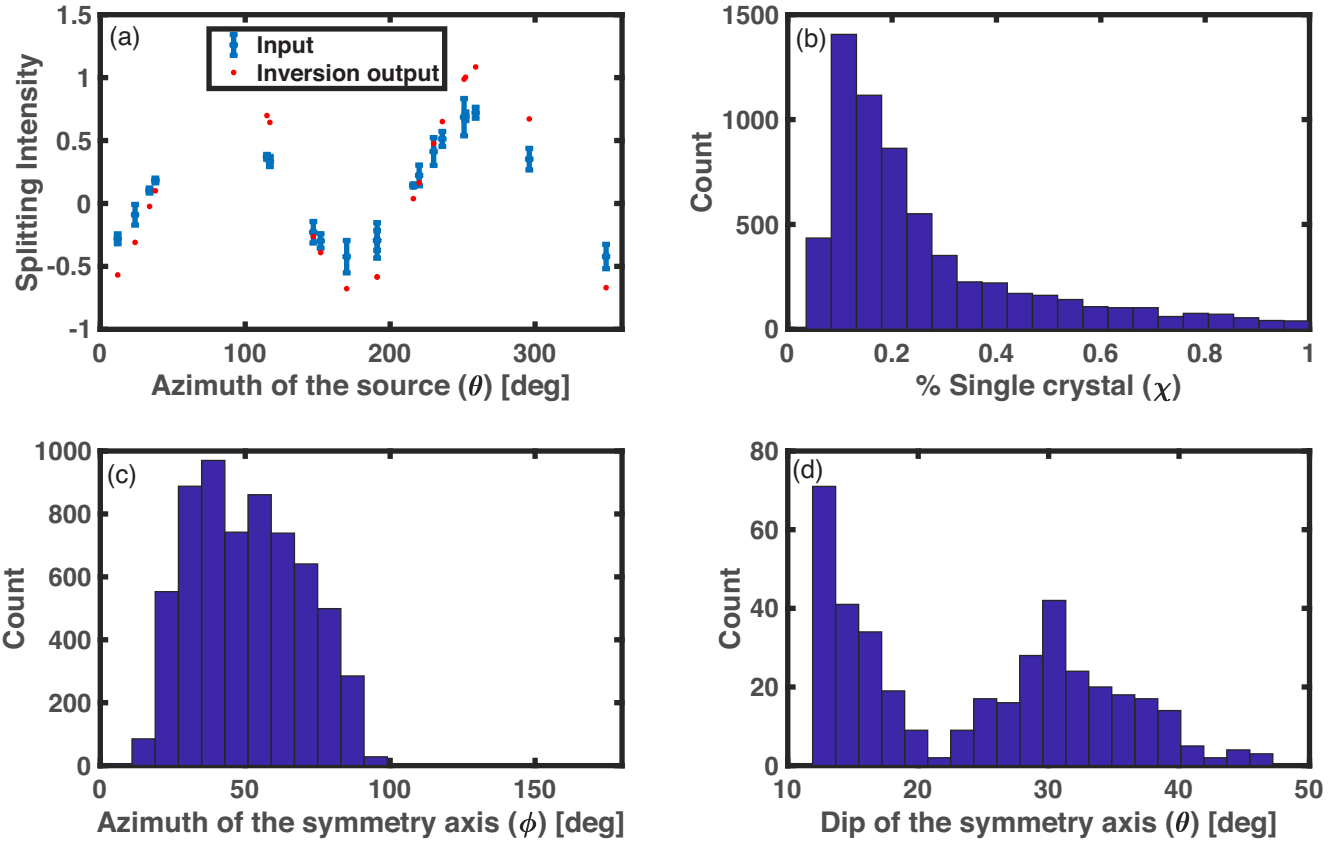

Figure 8. Single layer inversion result with Gaussian-distributed (standard deviation of 0.1) noisy input data. The input model consists of a single layer of anisotropy with $\chi = 0.1$, $\theta = 30^\circ$ and $\phi = 45^\circ$. (a) Synthetic splitting intensity as a function of azimuth of the events for a particular station (blue dots with error bar) and the splitting intensity computed using the most likely model obtained from McMC algorithm (red dots). (b) The posterior distribution of the strength parameter χ with a mean ($\bar{\chi}$) of 0.1291 and standard deviation (σ) of 0.1086. (c) The posterior distribution of the azimuth of the symmetry axis (ϕ) with a mean of 46° and standard deviation (σ) of 9°. (d) The posterior bimodal distribution for the dip of the symmetry axis (θ).

model. The HLP station configuration is used. Fig. 14(A) presents the input-output model with a $8 \times 8 \times 1$ model grid parametrization (note that we plot the most likely model result given by the McMC algorithm after smoothing). Because both 'spike' anomalies were inserted in the model in regions with dense station spacing, the algorithm correctly identifies the input model parameters. This test shows that a dense seismic network can easily resolve single-layered structures with lateral dimension of 150 km, and the finer grid resolution suggests that structures on the length scale of ≈50 km can be resolved.

6.3 Totally heterogeneous case

In this section, we present synthetic tests with input models that include vertical heterogeneity in addition to the lateral heterogeneity. These fully heterogeneous cases illustrate a key advantage of moving to a tomographic approach in SKS splitting studies. In traditional shear wave splitting studies, which measure and interpret the fast splitting direction and splitting delay time (ϕ, dt), there is little or no depth resolution, as splitting is a path-integrated measurement and SKS phases propagate nearly vertically. In our approach, however, the finite width of the sensitivity kernels enables us to obtain depth constraints, even with nearly vertical SKS phases, as long as the station configuration is dense enough to yield overlapping sensitivity kernels. For these tests with depth-dependent starting models, we implement a series of two-layer input models, involving the same number of horizontal blocks as in previous tests. First, we implement a test with a total number of 32 blocks (i.e. $4 \times 4 \times 2$) with the randomly generated station distribution; we vary the strength parameter (χ) while keeping θ (0) and ϕ (45°) fixed. Figs 9(B) and (C) presents the results of this test in map view (showing the horizontal resolution) at two different depths ($z = 100$ and 300 km), while Fig. 9(D) presents the input–output model of a vertical section, illustrating the vertical resolution. We find that the horizontal resolution deteriorates with increasing depth (compare Figs 9B and C); this is a consequence of the decaying strength of sensitivity kernel with depth ($\approx \frac{1}{r^a}$ for some $a > 0$), as shown in Figs 2 and 3.

We repeated this multiple-layer model test a four-layer input model; in this test the total number of model parameters is 64 ($4 \times 4 \times 4$). For this increased amount of model complexity, we use the dense HLP station distribution. The results of this test are shown in Figs 10(B) and (C), which show the horizontal resolution at two different depths, and in Fig. 10(D), which presents the resolution of the vertical section. This test demonstrates that for a dense station configuration, finite-frequency SKS splitting tomography is capable of resolving depth-dependent anisotropic structure throughout the upper mantle, and the vertical resolution length scale is roughly 100 km (the thickness of the layers in the four-layer input model). This increased vertical resolution is a crucial advantage to SKS splitting tomography, as opposed to traditional SKS splitting analysis.

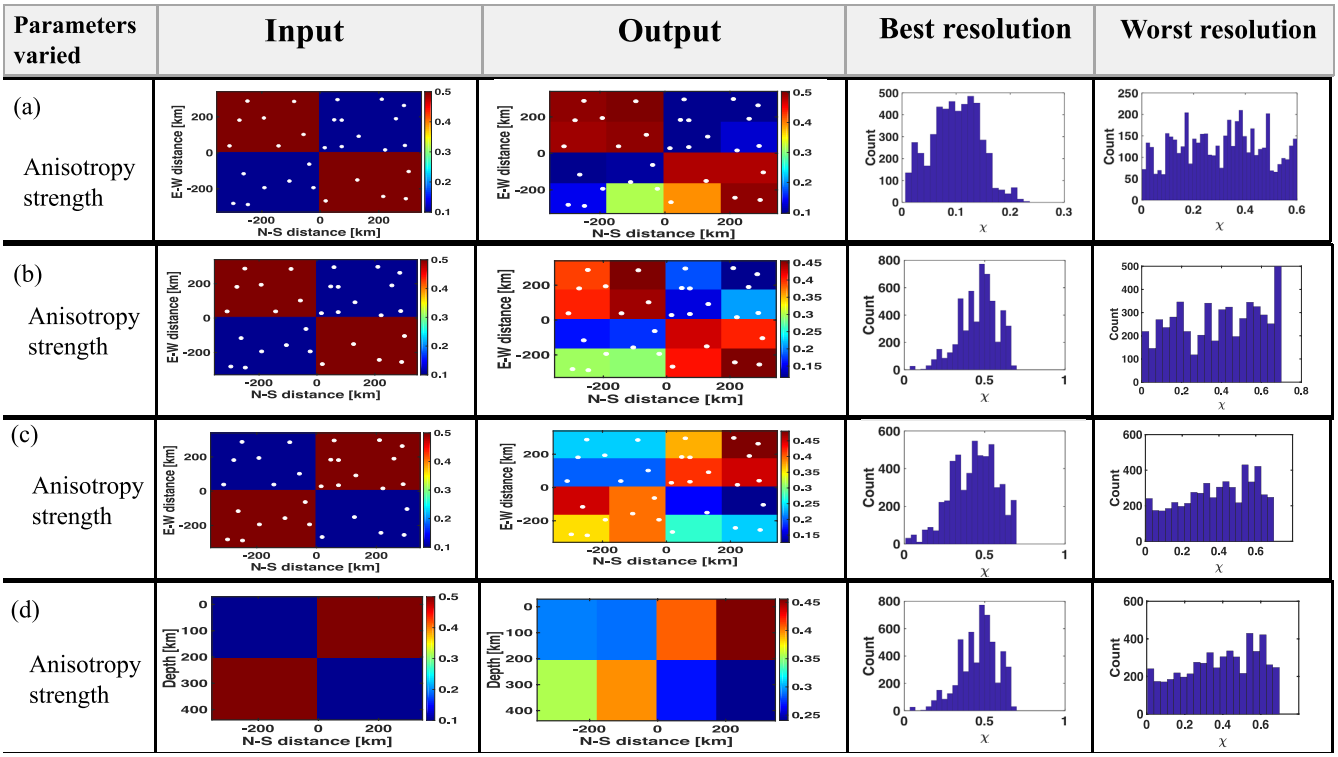

Figure 9. Results of checkerboard resolution tests with several different input model configurations for a random station configuration. The station locations are shown by white dots. In each of these tests, the strength parameter (χ) is varied while keeping $\theta = 0$ and $\phi = 45°$. (a) Horizontal slice of input–output model and the posterior distributions of the $4 \times 4 \times 1$ checkerboard test in a model domain of $\approx 700 \times 700 \times 400$ km^3, (b) horizontal slice of Input–output model at $z = 100$ km and the posterior distributions of the $4 \times 4 \times 2$ checkerboard test in a model domain of $\approx 700 \times 700 \times 400$ km^3, (c) horizontal slice of Input–output model at $z = 300$ km and the posterior distributions of the $4 \times 4 \times 2$ checkerboard test in a model domain of $\approx 700 \times 700 \times 400$ km^3 and (d) vertical slice of the input output model at $y = 100$ km and the posterior distributions of the $4 \times 4 \times 2$ checkerboard test in a model domain of $\approx 700 \times 700 \times 400$ km^3. The best and worst resolutions correspond to the grids for which the inversion algorithm yields outcomes closest to the input model.

In our final depth-dependent case, we consider two cases involving a 2-D inversion for depth-dependent anisotropic structure along the WNW–ENE dense line of stations of the HLP experiment (station spacing along this line is roughly 10 km on average). Because of the reduction in dimensions (of physical domain) from three to two, we choose a finer input model for this test (16×16). The number of parameters is therefore $16 \times 16 = 256$, with input anomalies with length scale of 50 km both laterally and with depth. Our pattern of input anomalies involves a checkerboard pattern in χ, with the other parameters fixed ($\theta = 0$ and $\phi = 45°$). Fig. 14(B) shows the input and output models, demonstrating the potential for a dense line of stations such as that included in the HLP experiment to resolve depth-dependent anisotropic structure in the upper mantle. We also performed a similar test with simultaneous variation of χ and ϕ at $\theta = 0$ (i.e. horizontal axis of symmetry) in the input model. However, due to the increased number of unknown parameters, we used a coarser grid in our model space parametrization (total number of blocks 4×4; total number of parameters 16×2). Fig. 12(B) shows the input–output pattern of ϕ for this 2-D, depth dependent inversion. Taken together, our two synthetic 2-D inversions for depth-dependent structure show that structure is generally well-resolved to a depth of approximately 100–150 km. This may be explained by the properties of the sensitivity kernels; specifically, the strength of the kernel decays with the depth and the maximum power of the kernels is confined within the first 100–150 km depth range. As a consequence, we expect that the resolution is best within the region where kernel holds its maximum power. However, the observations contain at least some information about anisotropic structure down to the base of the mantle.

6.4 A note on model agreement and number of parameters

The results of our synthetic tests, shown in Figs 9–14, show the input and output models as well as selected views of the posterior distribution for some parameters, helping us to understand the quality of the resolution for different cases. A direct comparison of input–output models and posterior distributions among different synthetic tests is difficult, however, because the inverse problem is configured differently across different tests. A different view of the inversion scheme's success comes from a comparison between the (synthetic) splitting intensity observations computed using the input model, and the set of predicted splitting intensities generated from the output of the inversion (that is, from the most likely model). Fig. 11 shows the actual synthetic splitting intensity (using input model) and the predicted splitting intensity (using output model of inversion) for a suite of resolution tests of increasing complexity. This demonstrates a reasonable correspondence between the predicted and actual values of the model outputs for the aforementioned calculations, showing that our inversion scheme is capable of identifying models that are consistent with the synthetic 'observations', even for fairly complex cases. However, the deviation from the one to one correspondence depicts the extent of failure to fit the model exactly. Fig. 11 shows that the deviation from one-to-one

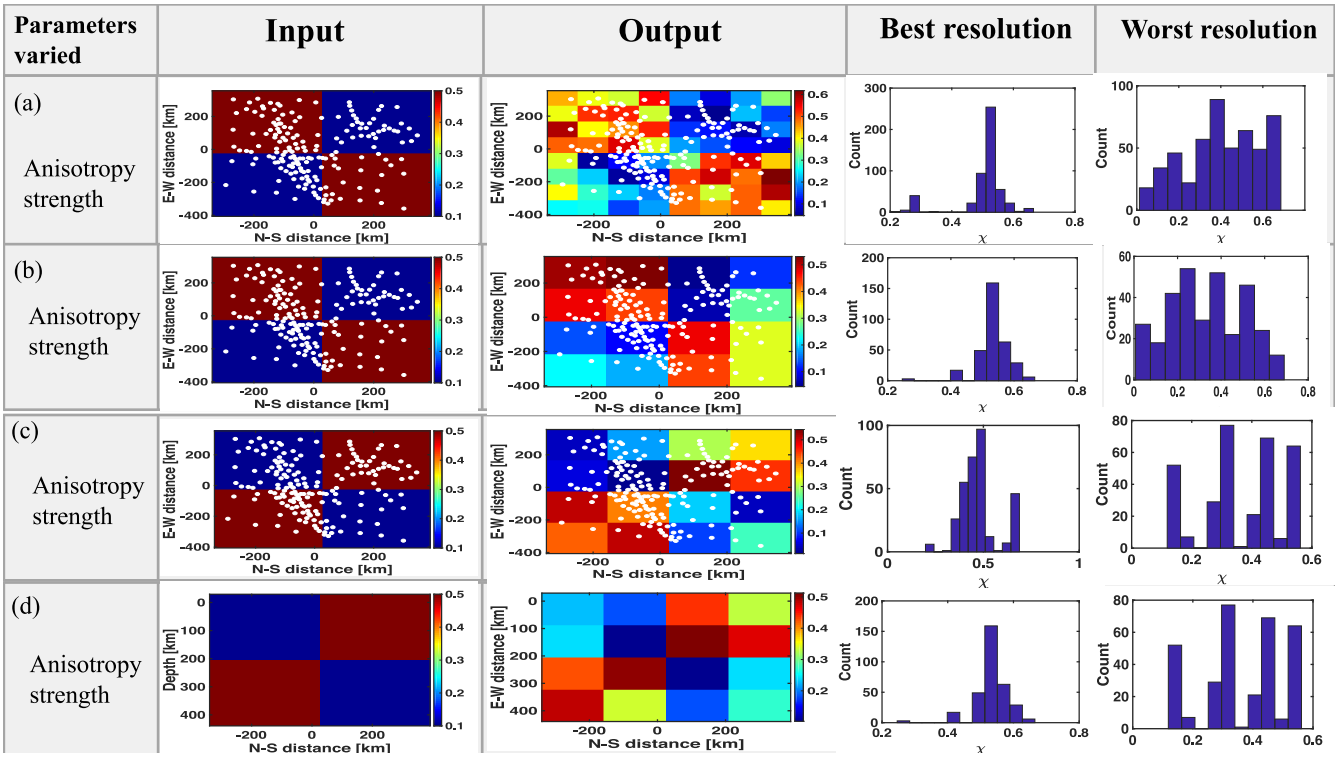

Figure 10. Results of the checkerboard tests with several different input model configurations for the HLP station geometry. The station locations are shown by white dots. In all of these tests, the strength parameter (χ) is varied while keeping $\theta = 0$ and $\phi = 45°$. (a) Horizontal slice of Input–output model and the posterior distributions of the 8 × 8 × 1 checkerboard test in a model domain of ≈700 × 700 × 400 km^3, (b) horizontal slice of Input–output model at $z =$ 50 km and the posterior distributions of the 4 × 4 × 4 checkerboard test in a model domain of ≈700 × 700 × 400 km^3, (c) horizontal slice of Input–output model at $z = 250$ km and the posterior distributions of the 4 × 4 × 4 checkerboard test in a model domain of ≈700 × 700 × 400 km^3 and (d) vertical slice of the input output model at $y = 100$ km and the posterior distributions of the 4 × 4 × 4 checkerboard test in a model domain of ≈700 × 700 × 400 km^3. The best and worst resolutions correspond to the grids for which the inversion algorithm yields outcomes closest to the input model.

correspondence increases with increasing number of unknown parameters in the inversion. This highlights a limitation of the model space search approach in the case of large number of parameters or a highly complex model.

7 DISCUSSION

The theoretical and practical framework for the forward and inverse problem associated with anisotropy tomography with SKS observations developed here highlights both the promise and the challenge of SKS splitting tomography. In our implementation, the size of the parameter space (in general 21 parameters for a homogeneous, general anisotropic model) is reduced to three based on the results from mineral physics regarding the elasticity and fabric development in olivine, the dominant mineral in the upper mantle. This parametrization scheme addresses one of the major challenges in shear wave splitting tomography, namely the unrealistically large number of parameters needed to describe anisotropy. Our implementation of finite-frequency theory enables us to address another of the major challenges, namely how to achieve the overlapping sensitivity for nearly vertically propagating SKS waves (see also Chevrot 2006). In our implementation, we achieve satisfactory depth resolution shallower than roughly 100–150 km depth, which lessens the ambiguity of traditional shear wave splitting interpretation.

Our implementation also addresses a third great challenge of anisotropy tomography, namely that the non-linear relation between the observed splitting intensity and the anisotropy parameters (χ, θ, ϕ) of the underlying medium results in a bias in the model estimation while implementing an inversion strategy based on a linearization around a background model. Sensitivity kernels for anisotropic structure depend strongly on the background model used to calculate them, and are substantially different for heterogeneous anisotropic starting models than for the homogeneous ones that are often used to calculate them (Long *et al.* 2008). Our technique based on a model space search accounts for this non-linear nature of the sensitivity kernels, and is able to resolve two parameters (namely the strength of anisotropy, χ, and the azimuth of the symmetry axis, ϕ) with generally high confidence. In contrast, the uncertainty in estimating the dip parameter (θ) is much higher, due to the use of only (nearly) vertically propagating SKS waves. This challenge may be ameliorated by using teleseismic direct S waves (Long & van der Hilst 2005) in conjunction with SKS/SKKS phases; in this case, however, contamination from anisotropy near the earthquake source is a concern (e.g. Foley & Long 2011).

The synthetic tests presented in this paper demonstrate that the quality of the resolution is greatly affected by the station configuration. This limitation is a general aspect of the seismic tomography inverse problem; however, given the complexity associated with anisotropic tomography, including the large dimensionality of the parameter space and the non-linearity issue, the station geometry is particularly crucial. Specifically, the presence of overlapping sensitivity kernels in the tomography problem is essential to provide confidence on the estimated model parameters. Our sensitivity tests

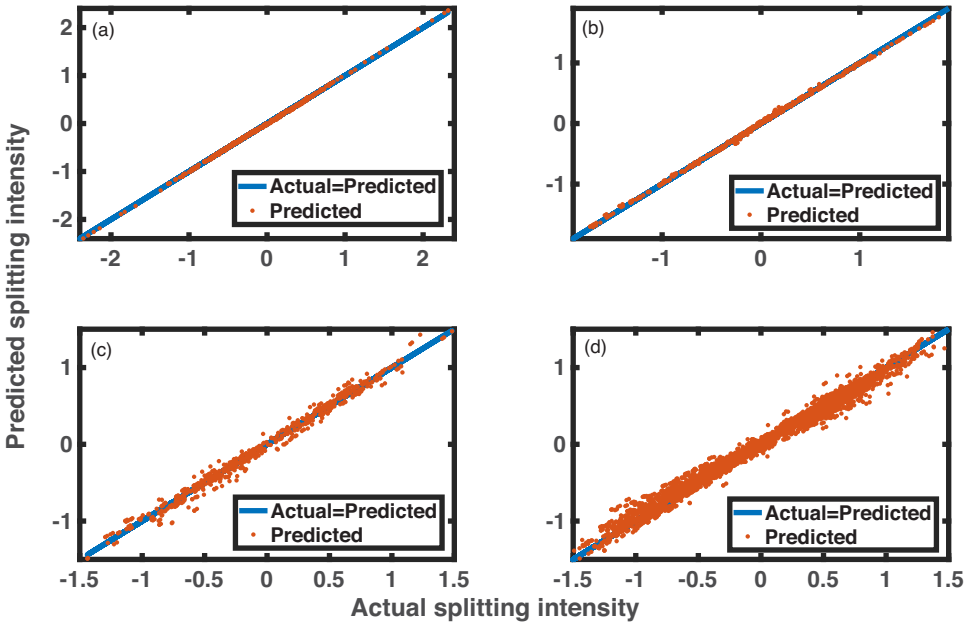

Figure 11. Actual (synthetic) and predicted observations for selected sensitivity tests. Correspondence between the predicted (using the most likely model from the output of MCMC algorithm) and the actual (using the chosen input model) splitting intensities for (a) single layer model; number of parameters being 3, (b) model with $4 \times 4 \times 1$ blocks in a model domain of $\approx 700 \times 700 \times 400$ km^3; number of parameters being $4 \times 4 = 16$, (c) model with $4 \times 4 \times 2$ blocks in a model domain of $\approx 700 \times 700 \times 400$ km^3; number of parameters being $4 \times 4 \times 2 = 32$ and (d)model with $4 \times 4 \times 4$ blocks in a model domain of $\approx 700 \times 700 \times 400$ km^3; number of parameters being $4 \times 4 \times 4 = 64$. The solid line represents the exact one-one correspondence between the actual and predicted splitting intensities, that is exact model fit. An increase in the number of model parameters (A through D) increases the deviation of the predicted data from the one-one correspondence.

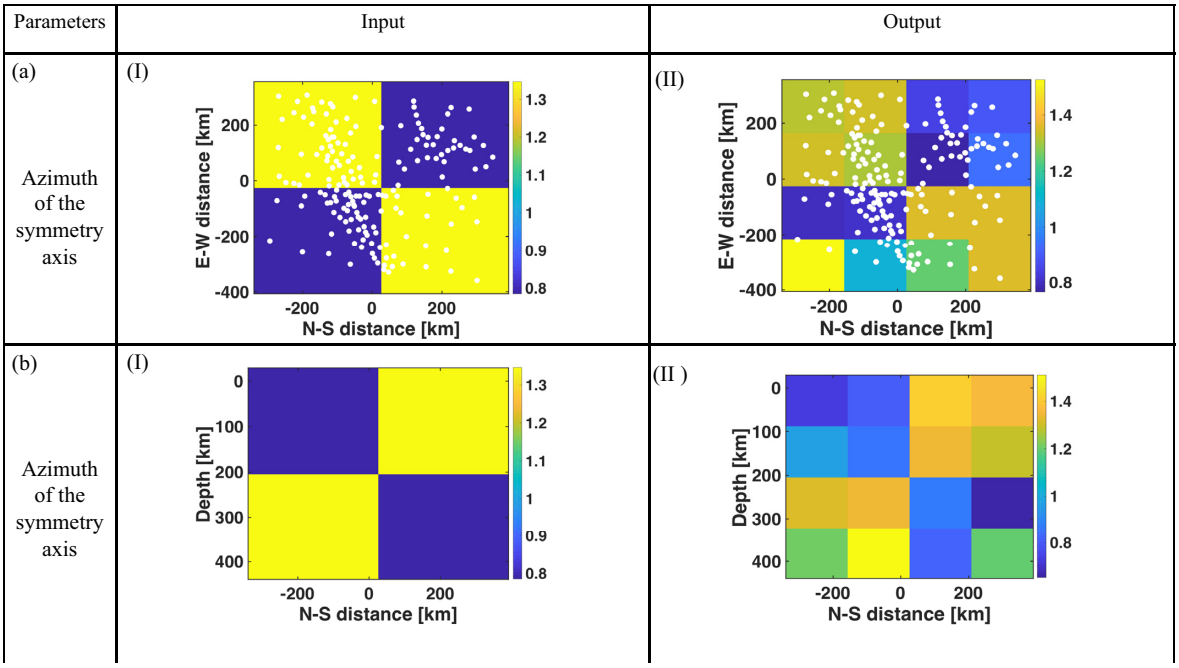

Figure 12. Results of checkerboard recovery tests for input models that vary the fast axis orientation. (a) (I) Spatial distribution of the azimuth of the symmetry axis used as input while only lateral variation is considered in an one layer model of $700 \times 700 \times 400$ km^3. High: 80° or 1.35 rad , Low: 45° or 0.78 rad, (a)(II) recovered spatial distribution of the azimuth with a $4 \times 4 \times 1$ checkerboard test. Unit of the angle is given in radian here. The red circles denote the station locations. (b)(I) Spatial distribution of the azimuth of the symmetry axis used as input while only vertical variation is considered in a 2-D section below densely distributed stations. High: 80° or 1.35 rad, Low: 45° or 0.78 rad, (b)(II) recovered spatial distribution of the azimuth. Unit of the angle is given in radian here.

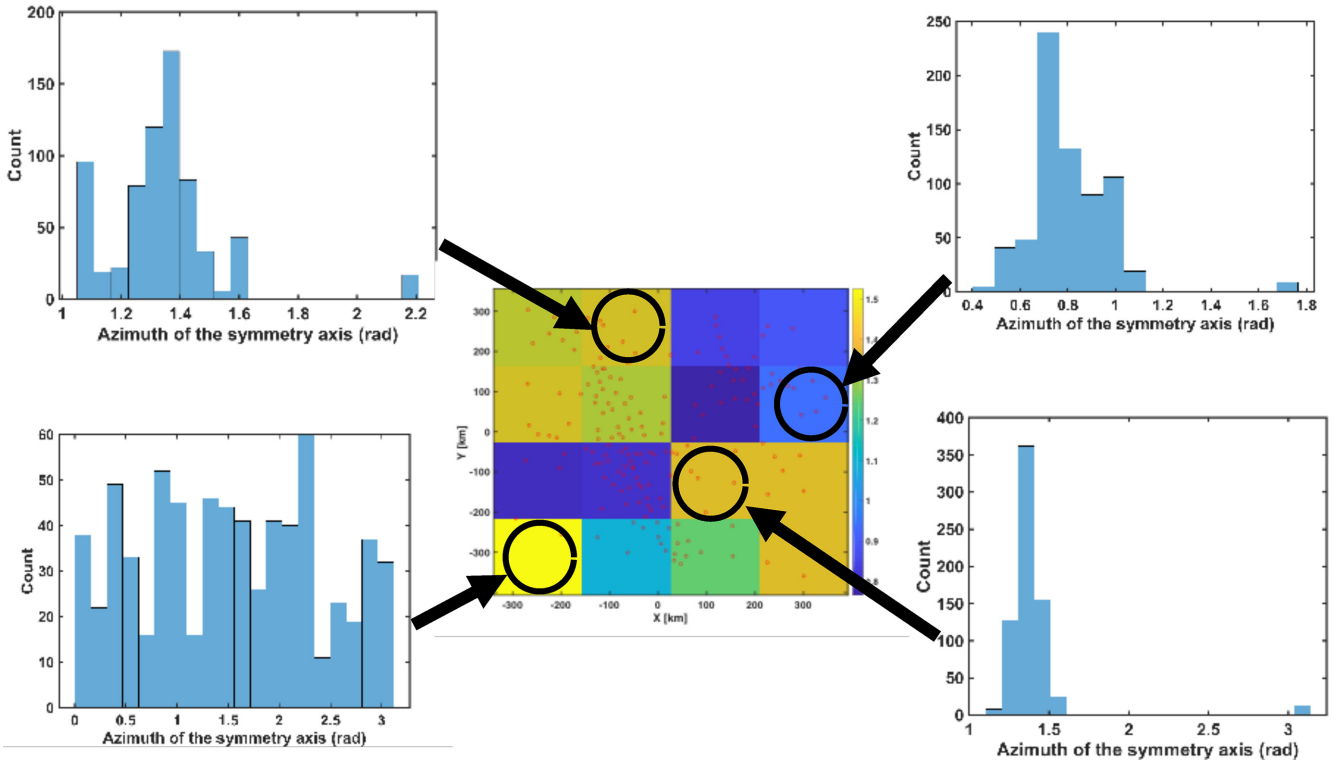

Figure 13. Posterior distributions of the recovered azimuth (ϕ) of the symmetry axis. This corresponds to $4 \times 4 \times 1$ checkerboard test in a model domain of $700 \times 700 \times 400$ km^3 and χ and ϕ are varied simultaneously; total number of parameters being $4 \times 4 \times 2 = 32$ with $\theta = 0$. Red dots denote the station locations. This test demonstrates that the input is well resolved where station coverage is good, and poorly resolved where it is sparse.

demonstrate that with a dense station configuration, such as that obtained by the High Lava Plains (HLP) experiment (average station spacing 10–15 km in its densest portions), anisotropic structure in the upper mantle may be resolved at a length scale of roughly 50 km with teleseismic SKS waves in a 2-D inversion (Fig. 14 B).

Our analysis also underscores the importance of a deep understanding of links between seismic anisotropy and mantle geodynamic processes. The orientation of the symmetry axis is closely related to the flow velocity field, but the relationship depends on the type of olivine fabric (i.e. A, B, C, D and E) as well as the strain geometry. For example, under the application of simple shear, the [100] axis of A type olivine is parallel to the shear direction (Zhang & Karato 1995; Karato *et al.* 2008); however, the details of fabric development are different for other fabric types. Therefore, the interpretation of the seismic anisotropy to infer mantle geodynamic processes critically hinges not only on the results from experimental mineral physics but also on the ability to resolve the details of seismic anisotropy at depth with a tomographic approach.

The results of the resolution tests for depth-dependent models presented in this study highlight the potential for constraining the depth distribution of anisotropy in the upper mantle, a particularly exciting avenue for future application of SKS splitting intensity tomography to actual data. We illustrate this potential by discussing results from the HLP SKS splitting data set, which was presented in Long *et al.* (2009) and Wagner & Long (2013). These authors documented striking variability in SKS splitting delay times along the HLP dense line (see fig. 6 of Wagner & Long 2013), with an increase in measured delay times from 1.5 to 2.5 s across a lateral distance of 200 km. This lateral variability in delay times was interpreted to represent small-scale lateral variability in the thickness of the anisotropic layer, the strength of olivine LPO in the upper mantle, or the presence of partial melt (in a shape preferred orientation, or SPO) in the shallowest upper mantle. Each of these possible scenarios would have important implications for the structure and dynamics of the HLP region, and for the origin of anomalous, young volcanism in the Cascades backarc (Long *et al.* 2009). However, because SKS splitting observations themselves do not constrain the depth distribution of the anisotropy, Long *et al.* (2009) were unable to distinguish among the three different possible models for the lateral variations in delay times. The resolution test shown in Fig. 12(B), which was constructed using the station distribution of the HLP experiment, points to a resolution of this outstanding problem, as it demonstrates that finite-frequency SKS splitting tomography applied to the HLP data set can resolve depth-dependent anisotropic structure in the upper mantle beneath this region.

8 SUMMARY AND OUTLOOK

Here we have developed and presented a theoretical and practical framework for the application of finite-frequency SKS splitting tomography with a sharply reduced number of parameters. Our approach relies on a McMC with Gibbs sampling algorithm to explore parameter space. Our strategies for parametrizing the problem, calculating sensitivity kernels and searching parameter space allow us to ameliorate some of the well-known challenges inherent in the inversion of shear wave splitting observations for anisotropic structure. These challenges include the large number of parameters needed to describe anisotropy, the difficulty of obtaining overlapping sensitivity zones (crossing rays, for the ray theoretical case) for nearly vertically propagating SKS phases, and the highly nonlinear nature of the inversion, in which the sensitivity kernels strongly

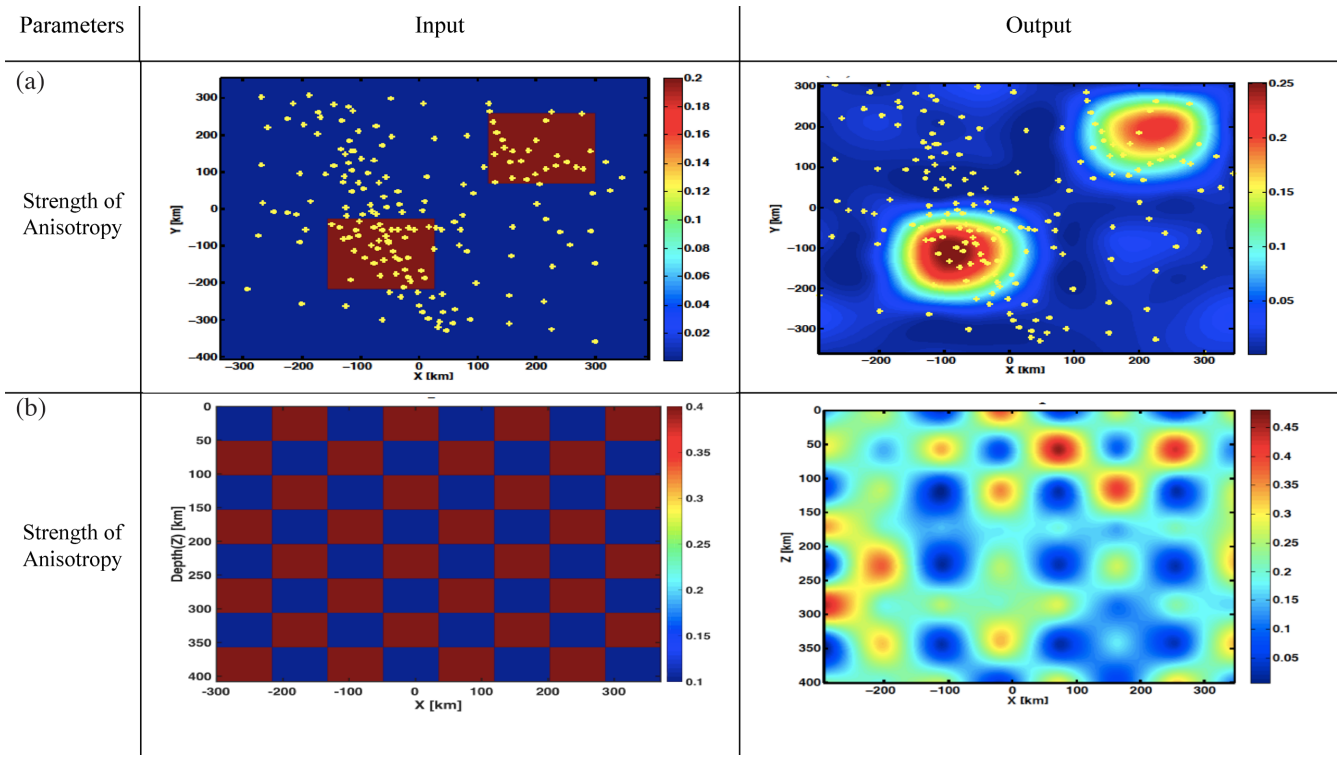

Figure 14. Results of the sensitivity tests using the HLP station configuration, with two different input model configurations. In both of these tests, the strength parameter (χ) is varied while keeping $\theta = 0$, and $\phi = 45^{\circ}$. (a) Horizontal slice of input-output model of the two block spike test in a physical domain of ≈800 × 800 × 400 km^3. The inversion is performed on 8 × 8 horizontal grids and output is smoothed. The station locations are denoted by yellow dots. (b) 2-D input–output model of the 8 × 8 checkerboard test in a physical domain of ≈700 × 400 km^2 along the dense linear array of the HLP experiment. This test demonstrates the promise of 2-D splitting intensity tomography for dense linear arrays (station spacing of 10–15 km in this example). X and Y denote N–S and E–W distance, respectively. Z denotes depth.

depend on the starting model. We have benchmarked our sensitivity kernel computations against well-known analytical solutions for simple models, demonstrating that our parametrization and kernel computations are valid. We carried out a suite of synthetic tests for input models of varying complexity to establish the effectiveness of our approach, to explore tradeoffs among different parameters, and investigate the length scale of structure that can reasonably be resolved by our approach. We find that SKS phases are not well suited to constraining the dip of the symmetry axis, consistent with previous work, but that the strength and azimuth of the anisotropy can be reliably constrained for a series of increasingly complex input models. A test of resolution for a 2-D inversion along a dense linear array of stations (using the station configuration from the HLP experiment in the Pacific Northwest USA) shows that structure in the upper mantle (depths up to 150 km) can be reliably retrieved at length scales of 50 km.

The synthetic experiments and proof of concept presented in this paper pave the way for future applications of our method to actual SKS splitting data sets, which have only been used in tomographic inversions in a few previous studies that have focused on Southern California (Monteiller & Chevrot 2010; Lin *et al.* 2014). Based on our results, the implementation of 2-D inversions along dense linear arrays is particularly likely to be successful, given their relatively modest computational requirements. Given the substantial increase in availability of data from dense broad-band seismic deployments since the mid-2000s, when the early papers developing the theory underpinning finite-frequency SKS splitting tomography were published (e.g. Favier & Chevrot 2003; Chevrot 2006; Long *et al.* 2008; Sieminski *et al.* 2008), the time is right for substantial progress on the application of SKS splitting tomography to real data. We are currently implementing our approach to data from the HLP SKS splitting data set (Long *et al.* 2009; Wagner & Long 2013), and other recently deployed experiments that include dense linear arrays represent obvious candidates for future work (e.g. the PULSE and CAUGHT experiments in South America; Eakin *et al.* 2015; Long *et al.* 2016). Broad-band seismic experiments with dense station spacing within North America that were enabled by the recently completed EarthScope USArray data gathering effort also represent prime targets for future application of our technique. Examples include the SPREE (Ola *et al.* 2016), OIINK (Yang *et al.* 2017), and NELE (Nyamwandha & Powell 2016) experiments in the mid-continent, as well as the SESAME (Hopper *et al.* 2017), MAGIC (Aragon *et al.* 2017) and QM-III (Chen *et al.* 2018) experiments in eastern North America.

ACKNOWLEDGEMENTS

We are grateful to Shun Karato for helpful discussions on olivine elasticity and texturing, and Maarten de Hoop and Rob van der Hilst for helpful discussions on the splitting intensity tomography problem. Thoughtful reviews by Li Zhao and Sébastien Chevrot are gratefully acknowledged. This work was supported by Yale University.

REFERENCES

Abt, D.L. & Fischer, K.M., 2008. Resolving three-dimensional anisotropic structure with shear wave splitting tomography, *Geophys. J. Int.,* **173**(3), 859–886.

Agmon, S., 1962. On the eigenfunctions and on the eigenvalues of general elliptic boundary value problems, *Commun. Pure appl. Math.,* **15**(2), 119–147.

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology,* University Science Books.

Aragon, J.C., Long, M.D. & Benoit, M.H., 2017. Lateral variations in sks splitting across the MAGIC array, central appalachians, *Geochem., Geophys., Geosyst.,* **18**(11), 4136–4155.

Becker, T.W., Chevrot, S., Schulte-Pelkum, V. & Blackman, D.K., 2006. Statistical properties of seismic anisotropy predicted by upper mantle geodynamic models, *J. geophys. Res.: Solid Earth,* **111**(B8).

Bodmer, M., Toomey, D.R., Hooft, E.E., Nábělek, J. & Braunmiller, J., 2015. Seismic anisotropy beneath the juan de fuca plate system: Evidence for heterogeneous mantle flow, *Geology,* **43**(12), 1095–1098.

Bourbaki, N., 1966. *Elements of Mathematics: General Topology. Part 1,* **Chapters 1-4.,** Hermann.

Browaeys, J.T. & Chevrot, S., 2004. Decomposition of the elastic tensor and geophysical applications, *Geophys. J. Int.,* **159**(2), 667–678.

Calixto, F.J. *et al.*, 2014. Shear wave splitting and shear wave splitting tomography of the southern Puna plateau, *Geophys. J. Int.,* **199**(2), 688–699.

Cammarano, F., Goes, S., Vacher, P. & Giardini, D., 2003. Inferring upper-mantle temperatures from seismic velocities, *Phys. Earth planet. Inter.,* **138**(3-4), 197–222.

Chen, X., Li, Y. & Levin, V., 2018. Shear wave splitting beneath Eastern North American continent: evidence for a multi-layered and laterally variable anisotropic structure, *Geochem., Geophys., Geosyst.,* **19**(8), 2857–2871 .

Chevrot, S., 2000. Multichannel analysis of shear wave splitting, *J. geophys. Res.: Solid Earth,* **105**(B9), 21579–21590.

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.

Chevrot, S. & Van Der Hilst, R.D., 2003. On the effects of a dipping axis of symmetry on shear wave splitting measurements in a transversely isotropic medium, *Geophys. J. Int.,* **152**(2), 497–505.

Conrad, C.P. & Behn, M.D., 2010. Constraints on lithosphere net rotation and asthenospheric viscosity from global mantle flow models and seismic anisotropy, *Geochem., Geophys., Geosyst.,* **11**(5), doi:10.1029/2009GC002970.

Cottaar, S. & Romanowicz, B., 2013. Observations of changing anisotropy across the southern margin of the African LLSVP, *Geophys. J. Int.,* **195**(2), 1184–1195.

Creasy, N., Long, M.D. & Ford, H.A., 2017. Deformation in the lowermost mantle beneath australia from observations and models of seismic anisotropy, *J. geophys. Res.: Solid Earth,* **122**(7), 5243–5267.

de Saavedra, F.A. & Kalos, M., 2003. Bilinear diffusion quantum monte carlo methods, *Phys. Rev. E,* **67**(2), 026708.

Eakin, C.M., Long, M.D., Wagner, L.S., Beck, S.L. & Tavera, H., 2015. Upper mantle anisotropy beneath Peru from SKS splitting: constraints on flat slab dynamics and interaction with the Nazca ridge, *Earth planet. Sci. Lett.,* **412,** 152–162.

Faccenda, M. & Capitanio, F., 2012. Development of mantle seismic anisotropy during subduction-induced 3-d flow, *Geophys. Res. Lett.,* **39**(11).

Favier, N. & Chevrot, S., 2003. Sensitivity kernels for shear wave splitting in transverse isotropic media, *Geophys. J. Int.,* **153**(1), 213–228.

Favier, N., Chevrot, S. & Komatitsch, D., 2004. Near-field influence on shear wave splitting and traveltime sensitivity kernels, *Geophys. J. Int.,* **156**(3), 467–482.

Foley, B.J. & Long, M.D., 2011. Upper and mid-mantle anisotropy beneath the Tonga slab, *Geophys. Res. Lett.,* **38**(2).

Ford, H.A. & Long, M.D., 2015. A regional test of global models for flow, rheology, and seismic anisotropy at the base of the mantle, *Phys. Earth planet. Inter.,* **245,** 71–75.

Foulkes, W., Mitas, L., Needs, R. & Rajagopal, G., 2001. Quantum Monte Carlo simulations of solids, *Rev. Modern Phys.,* **73**(1), 33.

Franssens, G.R., 1983. Calculation of the elasto-dynamic Green's function in layered media by means of a modified propagator matrix method, *Geophys. J. Int.,* **75**(3), 669–691.

Gilbert, F. & Backus, G.E., 1966. Propagator matrices in elastic wave and vibration problems, *Geophysics,* **31**(2), 326–332.

Graff, K.F., 2012. *Wave Motion in Elastic Solids,* Courier Corporation.

Hammond, B.L., Lester, W.A. & Reynolds, P.J., 1994. *Monte Carlo Methods in Ab Initio Quantum Chemistry,* Vol. **1,** World Scientific.

Hongsresawat, S., Panning, M.P., Russo, R.M., Foster, D.A., Monteiller, V. & Chevrot, S., 2015. Usarray shear wave splitting shows seismic anisotropy from both lithosphere and asthenosphere, *Geology,* **43**(8), 667–670.

Hopper, E., Fischer, K.M., Wagner, L.S. & Hawman, R.B., 2017. Reconstructing the end of the Appalachian orogeny, *Geology,* **45**(1), 15–18.

Karato, S.-I., Jung, H., Katayama, I. & Skemer, P., 2008. Geodynamic significance of seismic anisotropy of the upper mantle: new insights from laboratory studies, *Annu. Rev. Earth planet. Sci.,* **36,** 59–95.

Kendall, J.-M., Stuart, G., Ebinger, C., Bastow, I. & Keir, D., 2005. Magma-assisted rifting in Ethiopia, *Nature,* **433**(7022), 146.

Kennett, B., Engdahl, E. & Buland, R., 1995. Constraints on seismic velocities in the earth from traveltimes, *Geophys. J. Int.,* **122**(1), 108–124.

Korenaga, J. & Karato, S.-I., 2008. A new analysis of experimental data on olivine rheology, *J. geophys. Res.: Solid Earth,* **113**(B2).

Lin, Y.-P., Zhao, L. & Hung, S.-H., 2014a. Full-wave effects on shear wave splitting, *Geophys. Res. Lett.,* **41**(3), 799–804.

Lin, Y.P., Zhao, L. & Hung, S.H., 2014b. Full wave multiscale anistropy tomography in Southern California, *Geophys. Res. Lett.,* **41**(24), 8809–8817.

Liu, J.S., 2008. *Monte Carlo Strategies in Scientific Computing,* Springer Science & Business Media.

Long, M.D. & Becker, T.W., 2010. Mantle dynamics and seismic anisotropy, *Earth planet. Sci. Lett.,* **297**(3-4), 341–354.

Long, M.D., Gao, H., Klaus, A., Wagner, L.S., Fouch, M.J., James, D.E. & Humphreys, E., 2009. Shear wave splitting and the pattern of mantle flow beneath eastern Oregon, *Earth planet. Sci. Lett.,* **288**(3-4), 359–369.

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, *Geochemistry, Geophysics, Geosystems,* **17**(1), 2–15.

Long, M.D., Maarten, V. & Van Der Hilst, R.D., 2008. Wave-equation shear wave splitting tomography, *Geophys. J. Int.,* **172**(1), 311–330.

Long, M.D. & Silver, P.G., 2008. The subduction zone flow field from seismic anisotropy: a global view, *Science,* **319**(5861), 315–318.

Long, M.D. & Silver, P.G., 2009a. Mantle flow in subduction systems: the subslab flow field and implications for mantle dynamics, *J. geophys. Res.: Solid Earth,* **114**(B10).

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

Long, M.D. & van der Hilst, R.D., 2005. Upper mantle anisotropy beneath Japan from shear wave splitting, *Phys. Earth planet. Inter.,* **151**(3-4), 206–222.

Long, M.D. & Wirth, E.A., 2013. Mantle flow in subduction systems: the mantle wedge flow field and implications for wedge processes, *J. geophys. Res.: Solid Earth,* **118**(2), 583–606.

Mensch, T. & Rasolofosaon, P., 1997. Elastic-wave velocities in anisotropic media of arbitrary symmetry? Generalization of thomsen's parameters ε, δ and γ, *Geophys. J. Int.,* **128**(1), 43–64.

Monteiller, V. & Chevrot, S., 2010. How to make robust splitting measurements for single-station analysis and three-dimensional imaging of seismic anisotropy, *Geophys. J. Int.,* **182**(1), 311–328.

Monteiller, V. & Chevrot, S., 2011. High-resolution imaging of the deep anisotropic structure of the San Andreas Fault system beneath southern California, *Geophys. J. Int.,* **186**(2), 418–446.

Mosegaard, K. & Sambridge, M., 2002. Monte carlo analysis of inverse problems, *Inverse Problems,* **18**(3), R29.

Nyamwandha, C.A. & Powell, C.A., 2016. Seismic anisotropy beneath the Mississippi Embayment and the New Madrid seismic zone: a study of shear wave splitting, *J. geophys. Res.: Solid Earth,* **121**(11), 8239–8253.

Ola, O. *et al.*, 2016. Anisotropic zonation in the lithosphere of central north america: Influence of a strong cratonic lithosphere on the mid-continent rift, *Tectonophysics,* **683,** 367–381.

Reuss, A., 1929. Berechnung der fließgrenze von mischkristallen auf grund der plastizitätsbedingung für einkristalle., *J. Appl. Math. Mech./Zeitschrift für Angewandte Mathematik und Mechanik,* **9**(1), 49–58.

Savage, M., 1999. Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting? *Rev. Geophys.,* **37**(1), 65–106.

Savage, M.K. & Silver, P.G., 1993. Mantle deformation and tectonics: constraints from seismic anisotropy in the western united states, *Phys. Earth planet. Inter.,* **78**(3-4), 207–227.

Schmidt, H. & Tango, G., 1986. Efficient global matrix approach to the computation of synthetic seismograms, *Geophys. J. R. astr. Soc.,* **84**(2), 331–359.

Sieminski, A., Paulssen, H., Trampert, J. & Tromp, J., 2008. Finite-frequency SKS splitting: measurement and sensitivity kernels, *Bull. seism. Soc. Am.,* **98**(4), 1797–1810.

Silver, P.G., 1996. Seismic anisotropy beneath the continents: probing the depths of geology, *Ann. Rev. Earth planet. Sci.,* **24**(1), 385–432.

Silver, P.G. & Long, M.D., 2011. The non-commutivity of shear wave splitting operators at low frequencies and implications for anisotropy tomography, *Geophys. J. Int.,* **184**(3), 1415–1427.

Vinnik, L., Kosarev, G. & Makeeva, L., 1984. Lithosphere anisotropy from the observation of sks and skks waves, *Doklady Akademii Nauk SSSR,* **278**(6), 1335–1339.

Wagner, L.S. & Long, M.D., 2013. Distinctive upper mantle anisotropy beneath the High Lava Plains and Eastern Snake River Plain, Pacific Northwest, USA, *Geochem., Geophys., Geosyst.,* **14**(10), 4647–4666.

Wirth, E.A., Long, M.D. & Moriarty, J.C., 2016. A Markov chain Monte Carlo with Gibbs sampling approach to anisotropic receiver function forward modeling, *Geophys. J. Int.,* **208**(1), 10–23.

Wolfe, C.J. & Solomon, S.C., 1998. Shear-wave splitting and implications for mantle flow beneath the melt region of the east pacific rise, *Science,* **280**(5367), 1230–1232.

Wookey, J., 2012. Direct probabilistic inversion of shear wave data for seismic anisotropy, *Geophys. J. Int.,* **189**(2), 1025–1037.

Yang, X. *et al.*, 2017. Detailed crustal thickness variations beneath the illinois basin area: implications for crustal evolution of the midcontinent, *J. geophys. Res.: Solid Earth,* **122**(8), 6323–6345.

Zhang, S. & Karato, S.-I., 1995. Lattice preferred orientation of olivine aggregates deformed in simple shear, *Nature,* **375**(6534), 774.