Geophysical Journal International

Geophys. J. Int. (2011) doi: 10.1111/j.1365-246X.2011.04969.x

Inferring upper-mantle structure by full waveform tomography with the spectral element method

V. Lekić and B. Romanowicz

Berkeley Seismological Laboratory, 225 McCone Hall, University of California, Berkeley, CA 94720, *USA. E-mail: lekic@seismo.berkeley.edu*

Accepted 2011 January 24. Received 2011 January 24; in original form 2010 May 13

SUMMARY

Mapping the elastic and anelastic structure of the Earth's mantle is crucial for understanding the temperature, composition and dynamics of our planet. In the past quarter century, global tomography based on ray theory and first-order perturbation methods has imaged long-wavelength elastic velocity heterogeneities of the Earth's mantle. However, the approximate techniques upon which global tomographers have traditionally relied become inadequate when dealing with crustal structure, as well as short-wavelength or large amplitude mantle heterogeneity. The spectral element method, on the other hand, permits accurate calculation of wave propagation through highly heterogeneous structures, and is computationally economical when coupled with a normal mode solution and applied to a restricted region of the Earth such as the upper mantle (SEM). Importantly, SEM allows a dramatic improvement in accounting for the effects of crustal structure. Here, we develop and apply a new hybrid method of tomography, which allows us to leverage the accuracy of SEM to model fundamental and higher-mode long period (>60 s) waveforms. We then present the first global model of upper-mantle velocity and radial anisotropy developed using SEM. Our model, SEMum, confirms that the long-wavelength mantle structure imaged using approximate semi-analytic techniques is robust and representative of the Earth's true structure. Furthermore, it reveals structures in the upper mantle that were not clearly seen in previous global tomographic models. We show that SEMum favourably compares to and rivals the resolving power of continental-scale studies. This new hybrid approach to tomography can be applied to a larger and higher-frequency data set in order to gain new insights into the structure of the lower mantle and more robustly map seismic structure at the regional and smaller scales.

Key words: Inverse theory; Surface waves and free oscillations; Seismic anisotropy; Seismic tomography; Computational seismology; Wave propagation.

1 INTRODUCTION

Since the pioneering study of Dziewonski *et al.* (1977), seismic tomography has provided increasingly detailed images of the elastic structure of the Earth's deep interior. This progress was enabled by the proliferation of digital seismic data and the concomitant development of techniques for analysing them based on ray- and perturbation theory. At present, several tomographic models of global structure purport to resolve structures as small as 1000 km (e.g. Shapiro & Ritzwoller 2002; Ritsema *et al.* 2004; Panning & Romanowicz 2006; Simmons *et al.* 2006; Houser *et al.* 2008; Kustowski *et al.* 2008). Yet, only the long wavelength variations of isotropic shear wave speed appear to be robustly imaged on the global scale (Dziewonski 2005) and structures smaller than $\sim$2500 km correlate poorly across the available models (Becker & Boschi 2002). Discrepancies among models of variations of radial anisotropy (transverse isotropy) are present even at the longest wavelengths (e.g. Becker *et al.* 2007; Kustowski *et al.* 2008).

The discrepancies between global tomographic models of mantle elastic structure can arise from a combination of factors, including data utilization (e.g. traveltimes or waveforms), parametrization, regularization, theoretical limitations, and unmodelled crustal effects.

Forward modelling of wave propagation through a complex medium such as the Earth presents a particularly difficult challenge to the robust mapping of small scale heterogeneity. This is because ray theory, which underlies nearly all existing global tomographic models, is expected to breakdown as the lengthscale of the sought-after structure approaches that of the input waveforms (see e.g. Wang & Dahlen 1995; Spetzler *et al.* 2002). Even methods that include finite-frequency effects through single-scattering approximations (e.g. Dahlen *et al.* 2000; Zhou *et al.* 2006) are not accurate in modelling the effects of large anomalies (see Panning *et al.* 2009), which, due to the red spectrum of mantle heterogeneity (Su & Dziewonski 1991), are likely to dominate the observed waveforms.

Furthermore, traditional means of extracting information contained in seismic waveforms, such as phase-velocity and traveltime measurements of well-separated phases (e.g. Ritsema *et al.* 2004; Houser *et al.* 2008) present several drawbacks. First, they utilize only a small portion of the information contained in the seismogram, secondly, they discard the constraints encoded in wave amplitudes. Yet it is precisely the amplitude information that best constrains the gradients and short-wavelength variations in elastic properties (Romanowicz 1987). This is why Ferreira & Woodhouse (2006) found that a number of recent nominally high-resolution models of phase-velocity anomalies did not provide better fits to observed amplitudes than a spherically symmetric model. The wealth of information contained in amplitude measurements was illustrated by Dalton & Ekstrom (2006), who demonstrated that phase velocity maps can be successfully extracted from amplitude information alone.

Finally, long period seismic waves used for mapping mantle structure are sensitive to both crustal and mantle structure. Thus, unmodelled effects of crustal structure can complicate and, in the case of lateral variations of radial anisotropy, even obliterate the signal coming from mantle structure (e.g. Bozdağ. & Trampert 2008). Since long-period waveforms do not have the resolution required to jointly invert for crust and mantle structure, corrections based on an assumed crustal model are typically performed. Linear corrections have been shown to be inadequate in describing the effects of the crust on surface waveforms (e.g. Montagner & Jobert 1988). Even more accurate non-linear schemes (e.g. Kustowski *et al.* 2007; Marone & Romanowicz 2007) are liable to map inaccuracies in the assumed crustal structure, which, in the case of the most widely used CRUST2 model (Bassin & Masters 2000), can be substantial (e.g. Pasyanos 2005; Meier *et al.* 2007). Thus, eliminating the contamination of mantle images due to unmodelled crustal effects requires both the inclusion of higher-frequency data that provide better resolution of crustal structure and the use of forward modelling techniques capable of accurately predicting the effects of that structure on observed waveforms.

In this study, we have obtained a high resolution model of upper-mantle structure, based on the development and implementation of a new approach to waveform tomography, which exploits the accuracy of fully numerical wave propagation codes for forward modelling wave propagation through the Earth. The salient features of our approach include the following.

(i) Optimizing data utilization through the use of full waveform modelling.

(ii) Minimizing forward-modelling errors by using the spectral element method (SEM; e.g. Komatitsch & Vilotte 1998), which is also capable of accurately representing the effects of the oceans, topography/bathymetry, ellipticity, gravity, rotation and anelasticity (Komatitsch & Tromp 2002).

(iii) Minimizing crustal contamination by supplementing our data set of long period waveforms by higher frequency ($T > 25$ s) group velocity dispersion maps.

Computational costs are kept reasonable by (1) considering only long period waveforms, low pass filtered with a cut-off period of 60s, (2) implementing a smooth crustal model and (3) relying on approximate techniques for calculating partial derivatives that relate structure perturbations to waveform perturbations. The use of approximate partial derivatives decreases computational costs several-fold compared to adjoint methods (Tarantola 1984) applied recently on the local (Tape *et al.* 2009) and regional (Fichtner *et al.* 2009) scales. We stress that this study represents a break from traditional practice of tomography; for the first time, a global upper-mantle model is constrained in large part using a fully numerical wave propagation code that dispenses with the approximations and assumptions inherent in commonly used tomographic methods. To avoid introducing any bias in our 3-D model due to features of previous tomographic models, we choose a spherically symmetric 1-D model as a starting model in our inversion.

In what follows we successively discuss the starting model, model parametrization, implementation of the crust, forward and inverse modelling approach as well as the data set used in the inversion, and finally we present the 3-D radially anisotropic upper-mantle model obtained.

2 METHODS

Using seismic data to constrain the structure of the Earth's interior can be cast as a problem in which probabilities P are assigned to different possible interior structures given the available data. In this study, given a set of seismic waveforms and group velocity dispersion maps concatenated into the vector **d**, we infer the elastic parameters **m** describing the mantle, that is, P (**m** |**d**). In practice, calculating the probabilities requires us to:

(i) quantify data uncertainty;

(ii) incorporate *a priori* knowledge of correlations between elastic parameters in order to reduce the number of unknowns;

(iii) model propagation of seismic waves through heterogeneous mantle and crustal structures with minimal errors.

Waveforms of seismic waves that propagate through structure **m** are given by a non-linear function **g**(**m**). In practice, the computations and theory used to evaluate **g**(**m**) are inexact. This modelling uncertainty can be approximately summarized using a covariance matrix $\mathbf{C}_T$. We discuss the importance of this source of error in a separate section. If observational noise is close to Gaussian, we can also summarize the data uncertainty using a covariance matrix $\mathbf{C}_D$. We will summarize the *a priori* constraints on model parameters through a model covariance matrix $\mathbf{C}_M$ and a starting radially symmetric model $\mathbf{m}_0$.

Because **g**(**m**), the relation between Earth structure and seismic waveforms, is non-linear, inferring Earth structure from seismic data involves an iterative procedure. At the kth iteration, then, the partial derivatives of $\mathbf{g}(\mathbf{m}_k)$ with respect to model perturbations can be calculated, though they are only likely to be valid in the vicinity of the model $\mathbf{m}_k$ for which they are evaluated. Though a number of different techniques exist (see, for example Tarantola 2005), we opt for the quasi-Newton method, as it furnishes a compromise between keeping down computational costs while ensuring a fast convergence rate. At each iteration k, the model update $\delta\mathbf{m}_k$ is obtained by solving the linear system:

$$\left[\mathbf{I} + \mathbf{C}_M\mathbf{G}_k^T(\mathbf{C}_D + \mathbf{C}_T)^{-1}\mathbf{G}_k\right]\delta\mathbf{m}_k = \mathbf{C}_M\mathbf{G}_k^T(\mathbf{C}_D + \mathbf{C}_T)^{-1}[\mathbf{g}(\mathbf{m}_k) - \mathbf{d}] - \mathbf{m}_k + \mathbf{m}_0, \quad (1)$$

where $\mathbf{G}_k$ is the matrix of partial derivatives ($\partial d/\partial m$) relating model perturbations to data perturbations and evaluated for the current model $\mathbf{m}_k$. This expression is obtained by re-writing expression (25) in Tarantola & Valette (1982) to avoid taking the inverse of the $\mathbf{C}_M$ matrix. The mean of the Gaussian PDF that best approximates P (**m** |**d**) for iteration $k + 1$ is obtained by summing the model update $\delta\mathbf{m}_k$ and the model $\mathbf{m}_k$.

2.1 Model parametrization and *a priori* information

Propagation of seismic waves through an arbitrary Hookean medium depends on 21 parameters of the stiffness tensor, and inferring the values of all these parameters at all locations within the mantle is not feasible with available seismic data. However, by approximating the Earth as a transversely isotropic medium, we can drastically reduce the number of free parameters while capturing the first order observation that horizontally polarized surface waves travel, on average, faster than vertically polarized ones (e.g. Anderson 1961; McEvilly 1964). Such a medium can be described by introducing three anisotropic parameters in addition to the Voigt average isotropic velocities $V_{\rm Piso}$ and $V_{\rm Siso}$: $\xi = V_{SH}^2/V_{SV}^2$, $\phi = V_{PV}^2/V_{PH}^2$, and the parameter η which governs the variation of wave speed at directions intermediate to the horizontal and vertical. When η and ϕ are approximately equal to one, which is very likely the case in the mantle, we can approximately relate Voigt average velocities to those of vertically and horizontally polarized waves:

$$V_{\rm Piso}^2 = \frac{1}{5}\left(V_{PV}^2 + 4V_{PH}^2\right) \tag{2}$$

$$V_{\rm Siso}^2 = \frac{1}{3}\left(2V_{SV}^2 + V_{SH}^2\right) \tag{3}$$

as used by Panning & Romanowicz (2004). Because Love and Rayleigh waves are primarily sensitive to shear wave structure at periods longer than 60s (see, e.g. pp. 344–345 of Dahlen & Tromp 1998), we further decrease the number of parameters of interest by choosing not to invert for lateral variations in the poorly constrained $V_{\rm Piso}$, ϕ, ρ and η parameters. Instead, we parametrize the elastic structure of the mantle in terms of $V_{\rm Siso}$ and ξ and impose the following *a priori* correlations (which are fixed):

$$\delta \ln(\eta) = -2.5\delta \ln(\xi) \tag{4}$$

$$\delta \ln(V_{\rm Piso}) = 0.5\delta \ln(V_{\rm Siso}) \tag{5}$$

$$\delta \ln(\phi) = -1.5\delta \ln(\xi) \tag{6}$$

$$\delta \ln(\rho) = 0.3\delta \ln(V_{\rm Siso}). \tag{7}$$

Discussion of the reasons for this choice of physical parametrization can be found in appendix A of Panning & Romanowicz (2006).

In depth, the model is expressed on 21 cubic splines $\nu_q(r)$ defined in Mégnin & Romanowicz (2000), though we invert for structure only in the top 12 splines; deeper structure is fixed to that of SAW24B16 (Mégnin & Romanowicz 2000). The knot locations are at radii: 3480, 3600, 3775, 4000, 4275, 4550, 4850, 5150, 5375, 5575, 5750, 5900, 6050, 6100, 6150, 6200, 6250, 6300, 6346, 6361 km and the surface. Laterally, we parametrize our model spatially in terms of spherical splines $\beta_p(\theta, \phi)$ (Wang & Dahlen 1995). Thus, the value of a given model parameter m at any location in the Earth (θ, ϕ, r) can then be calculated from a set of spline coefficients m_{pq} by

$$m(\theta, \phi, r) = \sum_p \sum_q m_{pq}\beta_p(\theta, \phi)\nu_q(r). \tag{8}$$

The splines are a local basis, and thus help minimize the mapping of structure in one region into structure in distant regions, which can be an undesirable effect of global parametrizations such as spherical harmonics. By parametrizing our model, we put strict *a priori* constraints on the minimum length scale of structure allowed in our model. This truncation results in spectral leakage (aliasing) of short scale heterogeneity into longer length scales (Trampert & Snieder 1996), though the use of splines reduces this aliasing when compared to spherical harmonics or spherical pixels (Chiao & Kuo 2001). To further reduce the aliasing of retrieved structure, we allow structure to vary at shorter length-scales than those that we can reasonably expect to image and interpret (Spetzler & Trampert 2003).

Having chosen a parametrization for our upper-mantle model, we proceed to define a starting model for the inversion. We could have chosen a laterally heterogeneous starting model, which would have likely significantly accelerated the convergence of our iterative inversion scheme. However, we wanted to avoid biasing our results to any of the existing global tomographic models, all of which have been developed using approximate first-order perturbation techniques. By choosing as starting model a 1-D model, the model we have developed is independent of previous findings. Furthermore, we wanted to refer our 3-D model to a physically meaningful 1-D model, so that the 3-D perturbations could be more easily interpreted in terms of lateral variations in temperature and composition, given appropriate partial derivatives. Since we primarily focus on the top 400 km of the mantle, our reference and starting transversely isotropic velocity model has a spherically symmetric velocity profile which is identical to PREM (Dziewonski & Anderson 1981) below the 400 discontinuity. At depths shallower than 400 km, for the isotropic part of our 1-D starting model we consider a 1-D model obtained to fit long-period ($T > 60$ s) waveforms (Cammarano & Romanowicz 2007) starting from one of the physical reference models of Cammarano *et al.* (2005), which are calculated from a fixed composition (dry pyrolite) and a thermal profile using the elastic and anelastic properties of principal mantle minerals.

Because physical reference models of Cammarano *et al.* (2005) do not contain a 1-D profile of ξ, we obtain a reference model of transverse anisotropy ξ by carrying out a grid search in which we test several hundred candidate radial distributions of ξ against observed frequencies of spheroidal and toroidal modes, keeping fixed the elastic structure. To do this, we use the compilation of free oscillation frequencies from the reference earth model website (http://igppweb.ucsd.edu/ gabi/rem.html). We allow smoothly varying ξ to deviate from 1.0 (up to 1.2) at mantle depths shallower than 320 km, and do not allow values smaller than 1.0, which have been ruled out by numerous previous seismic studies (e.g. Dziewonski & Anderson 1981). The best-fitting profile of ξ is shown in Fig. 1, alongside the profile from PREM.

The *a priori* model covariance matrix $\mathbf{C}_M$, which specifies the expected deviation of true mantle structure from that specified by our starting model, is defined by the variance σ_0^2 (which are the diagonal entries) and the horizontal and vertical correlation lengths, h_0 and v_0, associated with each spline knot. Thus, the *a priori* model covariance for splines i and j whose average horizontal and vertical correlation lengths are h_0 and v_0 and that are separated by Δ_{ij} horizontally and d_{ij} vertically, is given by

$$c_M^{ij} = \text{const} \cdot \exp\left(\frac{\Delta_{ij} - 1}{h_0^2}\right) \exp\left(\frac{-2d_{ij}^2}{v_0^2}\right). \tag{9}$$

We choose vertical and horizontal lengths in line with the expected resolution of our data set and similar to those used in previous studies: ∼100 km is chosen to be the vertical correlation length for both V_S and ξ, while the horizontal correlation lengths are ∼800 km for V_S and ∼1200 km for ξ.

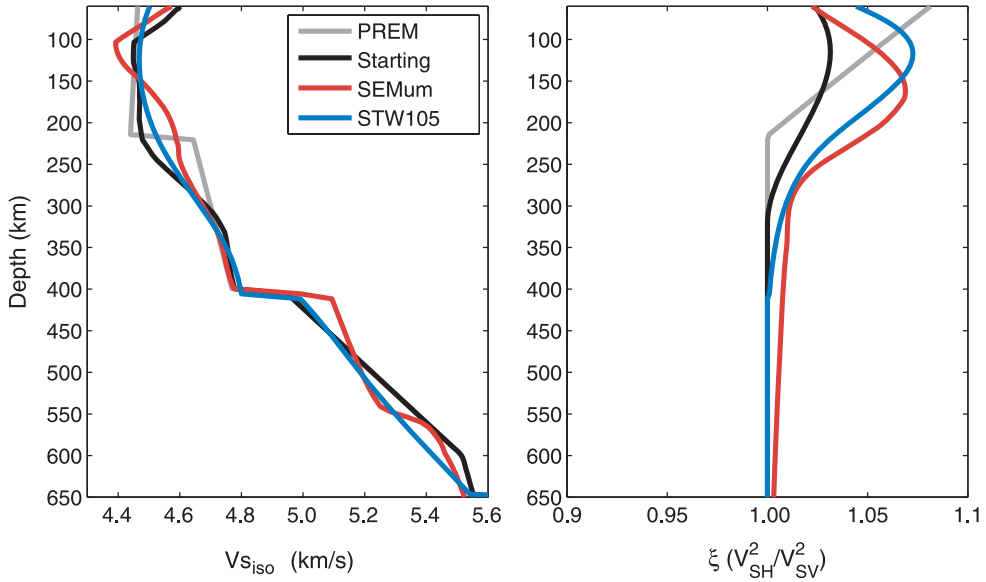

Figure 1. Left-hand panel: profiles of isotropic shear wave speed in our starting model, in SEMum, PREM and STW105 (Kustowski *et al*. 2008). Right-hand panel: profiles of ξ.

2.2 Modelling long period waveforms

Calculating the non-linear function g(**m**) that relates observed long period seismic waveforms to perturbations of isotropic shear wave speed and radial anisotropy commonly uses normal-mode summation approaches that rely on first order perturbation theory, asymptotic representations of Legendre polynomials and the stationary phase approximation (see Romanowicz *et al.* 2008). The most common of these approaches, the path average (great circle) approximation (PAVA: Woodhouse & Dziewonski 1984) further simplifies the calculations by neglecting heterogeneity-induced coupling between modes on different dispersion branches.

Despite the inaccuracies of this approach (see, e.g. Li & Romanowicz 1995; Romanowicz *et al.* 2008), PAVA allows efficient computation of both g(**m**) and $\mathbf{G}_k$, and was used, along with ray theory for body waves, to develop the most recent radially anisotropic global mantle model (S362ANI: Kustowski *et al.* 2008). An improvement was proposed by Li & Tanimoto (1993), who advocated considering coupling across mode branches. Li & Romanowicz (1995) implemented a related formalism for global tomography (NACT: non-linear asymptotic coupling theory), which introduced an additional term to PAVA that accounted for coupling across normal mode dispersion branches, bringing out the ray character of body waveforms. Several generations of global mantle elastic (Li & Romanowicz 1996; Mégnin & Romanowicz 2000) and anelastic (Gung & Romanowicz 2004) models have been developed using this approach. Most recently, Panning & Romanowicz (2006) and Panning *et al.* (2010) used NACT to develop a radially anisotropic model of the mantle (SAW642AN, SAW642ANb).

Fortunately, the development of computational techniques capable of fully modelling wave propagation through a complex, heterogeneous medium such as the Earth enables us to move away from these approximate techniques. In this study, instead of NACT seismograms, we use a version of the SEM that couples the 3-D mantle mesh to a 1-D normal-mode solution in the core, using a Dirichlet-to-Neumann operator (Capdeville *et al.* 2003). This reduces computational costs while preserving accuracy. Nevertheless, calculating synthetic waveforms for a single event (at all stations) requires ∼170 CPU hours; the nine iterations carried out in the development of the model we discuss here represent the result of ∼170 000 CPU hours (19 yr) of computation time. This figure does not include time for calculating partial derivatives, which is, in comparison, a negligible part of the overall computational time.

2.2.1 *Calculating g(**m**) and $\mathbf{C}_T$*

The use of the approximate techniques described above amounts to replacing the true relationship g(**m**) of eq. (1) with an approximate one, g'(**m**). Insofar as this modelling error can be described by Gaussian uncertainties, and does not systematically bias g(**m**), the use of approximate forward-modelling schemes introduces the additional covariance matrix $\mathbf{C}_T$ in eq. (1) (Tarantola 2005). Since variances are always positive, the additional variance arising from the use of such approximations will always increase the variances assigned to the observations. The use of approximate techniques can be thought of as the addition of systematic noise to the data. If the modelization error introduces a model-dependent bias on the predicted waveforms, then the situation is even worse, and convergence to the incorrect model is a strong possibility (see appendix).

Relative contributions of observation noise to modelling noise can be compared in order to quantify the importance of using an accurate theoretical framework for modelling wave propagation. Because of its sharp lateral gradients and its non-linear effect on surface waves (Montagner & Jobert 1988), crustal structure affects seismic waves in ways that are not readily captured by standard modelling approaches that rely on ray theory and first order perturbation theory. Bozdağ & Trampert (2008) compared the most common non-linear approach for dealing with crustal structure against reference synthetics calculated using the SEM and found that for long paths it resulted in errors larger than typical measurement error. Lekic *et al.* (2010) extended this analysis to waveforms and found the often-used linear approaches to calculating crustal corrections to be inadequate. Ferreira *et al.* (2010) have also shown that using different crustal models for crustal corrections can lead to different images of radial anisotropy. Even the effects of long-wavelength and smoothly-varying heterogeneities can be inaccurately captured by standard modelling techniques. Panning *et al.* (2009) find that

for realistic Earth structures, the use of Born theory can result in waveform modelling errors greater than measurement error.

Making the optimistic assumption that the modelization error is Gaussian and of the same magnitude as measurement error, then the use of inaccurate forward-modelling schemes is equivalent to doubling the uncertainty on the data. If data measurement error is also Gaussian, a data set analysed using accurate forward-modelling schemes carries the same uncertainty as a data set that is four times bigger but analysed with inaccurate forward-modelling. In fact, the more common situation is very much worse than this, since inaccuracies in forward-modelling are often correlated with Earth structure and are of different magnitude for different wavetypes. For example, inaccurately accounting for crustal structure affects Love waves more than Rayleigh waves, and can easily obliterate the anisotropic signal of the mantle (Lekic *et al.* 2010).

In this study, we minimize modelization error (rendering $\mathbf{C}_T$ negligible for our model parametrization) by using SEM to accurately calculate the propagation of waves through a complex and heterogeneous medium such as the Earth's mantle (Komatitsch & Tromp 2002). In the core, wave propagation is calculated using a 1-D normal mode summation approach and it is coupled to the SEM solution using a Dirichlet-to-Newman boundary-condition operator (Capdeville *et al.* 2003). Effects of the oceans, topography/bathymetry, ellipticity, gravity, rotation and anelasticity are all accounted for.

2.2.2 *Calculating G_k*

Due to the substantial increase in computational costs associated with the use of SEM, we rely on the approximate NACT approach to calculate the partial derivatives G_k. Even adjoint methods (e.g. Tarantola 1984; Tromp *et al.* 2005) which make possible efficient SEM-based calculation of G_k, would increase computational costs several fold, compared to the use of NACT. This is because separately weighting wavepackets according to their type, which allows fitting of overtone energies and equalizing sensitivity to horizontally and vertically polarized wavefields, would require separate calculation of adjoint kernels for each wavepacket type. Furthermore, while NACT kernels are indeed approximate, they do capture finite-frequency effects in the vertical plane defined by the great circle path, and thus enable meaningful representation of the sensitivities of body and overtone phases. They also capture the non-linearity associated with multiple forward scattering as does the PAVA approximation (Romanowicz *et al.* 2008). While we expect that inaccuracies of NACT kernels may slow down the convergence of our iterative procedure, we are confident that our accurate evaluation of the cost function at each step will ensure that a meaningful solution is obtained. Indeed, the only requirement on the kernels is that they capture the correct sign of the partial derivatives with respect to a given model parameter once the kernels for all available data points are summed.

In the NACT formalism, a model perturbation $\delta\mathbf{m}$ affects the seismic waveform $u(t)$ through coupling within a mode multiplet k and across multiplets k and k' within and across dispersion branches (Li & Romanowicz 1995):

$$u(t) = \Re e \left\{ \sum_k \left[(1 - it\widetilde{\omega}_{kk}) e^{i\widetilde{\omega}_{kk}t} \sum_m R_k^m S_k^m + \sum_{k' \geq k} \frac{e^{i\widetilde{\omega}_{kk}t} - e^{i\widetilde{\omega}_{k'k'}t}}{(\omega_k + \omega_{k'})(\widetilde{\omega}_{kk} - \widetilde{\omega}_{k'k'})} A_{kk'} \right] \right\}, \quad (10)$$

where k denotes a multiplet of radial order n and angular degree l, m is the azimuthal order of singlets within the multiplet, R_k^m and S_k^m are the source and receiver vectors defined in Woodhouse & Girnius (1982), ω_k is frequency of multiplet k, and

$$\widetilde{\omega}_{kk} = \omega_k + \frac{1}{\Delta} \int_S^R \delta\omega_{kk'} \delta_{kk'} \mathrm{d}s \quad (11)$$

is the new mode frequency shifted by coupling within the multiplet. Coupling across multiplets is contained in the $\mathrm{A}_{kk'}$ term

$$A_{kk'} = \frac{1}{2\pi} \left[Q_{kk'}^{(1)} \int_0^{2\pi} \delta\omega_{kk'}^2 \cos[(l' - l)\varphi] \, \mathrm{d}\varphi + Q_{kk'}^{(2)} \int_0^{2\pi} \delta\omega_{kk'}^2 \sin[(l' - l)\varphi] \, \mathrm{d}\varphi \right], \quad (12)$$

where the integrations are carried out on the great circle containing source and receiver and the expressions for $Q_{kk'}^{(1,2)}$ can be found in appendix A of Li & Romanowicz (1995). Finally, the mode frequency shifts due to heterogeneity-induced coupling are given by

$$\delta\omega_{kk'}(\theta, \phi) = \frac{1}{\omega_k + \omega_{k'}} \int_0^{R_\oplus} \delta\mathbf{m}(r, \theta, \phi) \mathbf{M}_{kk'}(r) r^2 \mathrm{d}r, \quad (13)$$

where $R_\oplus$ is the Earth's radius, and the kernels, $\mathbf{M}_{kk'}$ can be calculated according to expressions derived by Woodhouse & Dahlen (1978) in the case when $k = k'$ and Romanowicz (1987) when $k \neq k'$.

From these expressions, we derive the partial derivatives that make up $\mathbf{G}_k$ (for an explanation of how this is done, see Li & Romanowicz 1995). Effects of lateral heterogeneity $\delta\mathbf{m}$ on the seismic waveforms $u(t)$ are fully captured by considering the coupling-induced frequency shifts $\omega_{kk'}$ of normal modes. Symbolically, $\partial u(t)/\partial\delta\mathbf{m} = F(\delta\omega_{kk'})$, where F depends non linearly on the model through the exponential terms in eq. (10). Thus, unlike in a purely Born formalism, $\mathbf{G}_k$ depends on the iteration of the 3-D model. In fact, NACT waveform kernels can be thought of as weighted averages of individual mode frequency kernels $\mathbf{M}_{kk'}$, in which the weights depend on the seismic source characteristics, observation component, source-receiver distance and time. For the case of the fundamental mode dispersion branch ($n = 0$) which comprises Rayleigh and Love waves, it is sufficient to consider only along-branch coupling, and neglect modes for which $n' \neq 0$.

2.3 Implementing the crust

To accurately determine mantle structure, the effects of crustal structure on waveforms must be accurately accounted for. Our starting crustal model has average (harmonic mean) crustal velocities and thicknesses from CRUST2 (Bassin & Masters 2000) filtered by a 5.6° Gaussian filter to avoid spatial aliasing by the SEM mesh. Surface topography from ETOPO1 (Amante & Eakins 2008) and Moho topography of CRUST2 are similarly filtered. We deform the SEM mesh so that the Moho is always matched by an element boundary. This ensures that the sharp velocity jump of the Moho is accurately represented by SEM instead of being arbitrarily smoothed and aliased. However, ensuring accurate representation of crustal structure comes at a cost of very computationally expensive meshing of the thin oceanic crust.

Despite its widespread use, CRUST2 is inaccurate at both the global (Meier *et al.* 2007) and regional (e.g. Pasyanos & Nyblade 2007) scales. Furthermore, after four iterations, we found that CRUST2 did not allow us to simultaneously fit both Rayleigh

and Love waves. Improving the crustal model and better mapping shallow structure, however, requires higher frequency waveforms, which provide higher sensitivity to crustal structure. Therefore, we chose to supplement our waveform data set by shorter period Love and Rayleigh group velocity dispersion maps, and invert for crustal structure.

Inverting for a new model of crustal structure requires us to calculate kernels which capture the sensitivity of group velocities to perturbations of elastic structure. We explain how this is done in the next section. Here we wish to stress that since elastic properties of the crust vary substantially across the globe, the sensitivities of high frequency group velocities to elastic structure become themselves a function of the structure, that is, the non-linearities can no longer be neglected. Therefore, we must ensure that we use these kernels only in the valid, linear regime in which model perturbations are sufficiently small to be linearly related to group velocity perturbations. This is done by calculating kernels not just in a single reference earth model, but rather in a set of reference models which span a sufficiently broad range of profiles of crustal and mantle velocity structure to capture the heterogeneity present in the Earth. We accomplish this by taking a set of five profiles that span the variability present in a pre-existing model of upper-mantle and crustal-shear wave speed structure. To create a smooth model that will drastically reduce computational costs in SEM, while not biasing our modelling towards pre-existing models of crustal structure such as CRUST2.0, we conduct a grid search to develop a new starting model of crustal structure.

The smooth starting crustal model is obtained by generating 21 000 models of crustal structure in which we vary the model coefficients m_{pq} so that crustal V_S takes on values between 3 and 4.5 km s^{-1} in the oceans and 2–4 km s^{-1} in the continents. After a series of tests, we chose to keep apparent Moho depth fixed at 60 km and introduce crustal radial anisotropy (ξ) to compensate, allowing it to vary from 0.8 to 1.4. This is because the introduction of anisotropy allows a smooth model to have a similar response for long period waves as a model with thin layers (see Backus 1962; Capdeville & Marigo 2007). Having a deeper Moho avoids the need for meshing thin shallow layers, thereby reducing computational costs associated with the SEM three-fold. The group velocities for each of the candidate models are calculated by integrating the elasto-gravitational equations (Woodhouse 1998), and the model best predicting the observed Love and Rayleigh group velocity dispersion is selected at each point. Our crustal model, then, specifies a smoothed crustal structure beneath each point on the Earth that fits the group velocity dispersion data. Even though the best-fitting model is selected considering only fundamental mode dispersion, we confirm that it also provides adequate fits for overtones. This procedure is similar to the one used by Fichtner & Igel (2008). We then use this smooth crustal model alongside a long wavelength model of mantle structure to extract five reference models, within the vicinity of which the variations of group velocity lie in the linear regime. This is necessary to ensure that the perturbations of group velocity are always small enough to be linear and therefore accurately described by our kernels. These reference models are recalculated after each iteration of our inversion procedure.

2.3.1 Group velocity kernels

To include group velocity dispersion data to constrain shallow layers in our inversion, we need to develop expressions for group velocity kernels. Consider a wave whose speed of propagation depends on three interdependent variables: its frequency (ω), the elastic properties of the medium (m), and its wavenumber (κ). The cyclic chain rule relates the partial derivatives of ω, m and κ

$$\left(\frac{\partial\omega}{\partial\kappa}\right)_m \left(\frac{\partial\kappa}{\partial m}\right)_\omega = -\left(\frac{\partial\omega}{\partial m}\right)_\kappa . \tag{14}$$

Introducing the group velocity $U = \left(\frac{\partial\omega}{\partial\kappa}\right)_m$, and the wave speed $c = \omega/\kappa$, we can rearrange this expression to obtain

$$\frac{U}{c^2}\left(\frac{\partial c}{\partial m}\right)_\omega = \frac{1}{\omega}\left(\frac{\partial\omega}{\partial m}\right)_k . \tag{15}$$

Expression (15) can be used to calculate phase velocity kernels at a fixed period from eigenfrequency kernels calculated at fixed wavenumber. It is important that these partials are exactly the required ones, since we are keeping frequency constant, and phase (and group) velocity measurements are made at a specific frequency, rather than a particular wavenumber. If only coupling within a mode multiplet is considered, our waveform analysis is built upon kernels $\mathbf{M}_{kk}$ which represent the effect of a relative model perturbation $\delta\mathbf{m}/\mathbf{m}$ on the squared frequency ω^2, that is, $\mathbf{M}_{kk} = 2\omega m\left(\frac{\partial\omega}{\partial\mathbf{m}}\right)_\kappa$. Then, the logarithmic phase velocity kernel, $K^c = \frac{m}{c}\left(\frac{\partial c}{\partial m}\right)_\omega$ can be written as

$$K^c = \frac{c}{2U\omega^2}\mathbf{M}_{kk}. \tag{16}$$

To obtain the expressions for group velocity kernels, we start by expressing U in terms of c and $(\partial c/\partial\omega)_m$, and differentiate the expression with respect to **m**. Reorganizing, we obtain expressions for the group velocity kernels

$$\left(\frac{\partial U}{\partial m}\right)_\omega = \frac{U^2}{c^2}\left[\left(\frac{2c}{U}-1\right)\left(\frac{\partial c}{\partial m}\right)_\omega + \omega\left(\frac{\partial}{\partial\omega}\right)_m\left(\frac{\partial c}{\partial m}\right)_\omega\right]. \tag{17}$$

The second term of this expression involves taking the frequency derivative of the phase velocity kernels. This can be done numerically (Rodi *et al.* 1975) by differencing the phase kernels calculated at $\omega + \delta\omega$ and $\omega - \delta\omega$. In practice, we are concerned with group velocity dispersion measurements made on the fundamental mode branch, so in order to obtain group velocity kernel corresponding to the frequency of a mode with angular order $l = l_0$, we difference phase velocity kernels for $l = l_0 - 1$ and $l = l_0 + 1$, and divide by the difference in the eigenfrequencies $\Delta\omega = \omega_{l+1} - \omega_{l-1}$.

Casting eq. (17) in terms of K^c, defines a new group velocity kernel K^U which relates logarithmic perturbations in model parameters to logarithmic perturbations in group velocity

$$K^U = \frac{m}{U}\left(\frac{\partial U}{\partial m}\right)_\omega = K^c + \omega\frac{U}{c}\left(\frac{\partial}{\partial\omega}\right)_m K^c. \tag{18}$$

These kernels relate group velocity U at some point on the surface of the Earth (θ, ϕ) measured at frequency ω_j to the elastic structure beneath that point. Let the vector m_{pq} represent a set of coefficients that capture Earth structure parameters expressed in terms of spherical splines $\beta_p(\theta, \phi)$ and vertical cubic splines $\nu_q(r)$. The structure at point (r, θ, ϕ) is then given by eq. (8).

In general, the relationship between model vector m_{pq} and group velocity at a specified location $U_j(r, \theta, \phi)$ (where j indexes the frequency ω_j at which the group velocity is measured) is described by a non-linear function $\mathbf{g}(\mathbf{m})$. However, in the vicinity of a reference model ${}^i m_{pq}$, small changes in structure δm will not appreciably change the kernels ${}^i K_j^U(r)$; in this situation, deviations of group

velocity from the reference value ${}^iU_j$ will be linearly related to the perturbations of the model parameters from ${}^im_{pq}$

$$\sum_p \sum_q \frac{m_{pq} - {}^im_{pq}}{{}^im_{pq}} \beta_p(\theta,\phi) \int_0^a {}^iK_j^U(r')\nu_q(r')\mathrm{d}r' = \frac{U_j(r,\theta,\phi) - {}^iU_j}{{}^iU_j}, \quad (19)$$

where a is the radius of the Earth. By introducing ${}^iM_{j,q}^U$ as the radial integral of kernel ${}^iK_{j,q}^U(r)$ with vertical spline $\nu_q(r)$, we can rewrite the expression as

$$\sum_p \beta_p(\theta,\phi) \sum_q {}^iM_{j,q}^U \mathrm{d}\ln m_{pq} = \mathrm{d}\ln {}^iU_j \quad (20)$$

or in matrix notation

$$(\mathbf{M} \otimes \mathbf{B})\, \delta \ln \mathbf{m} = \mathbf{G}\, \delta \ln \mathbf{m} = \delta \ln \mathbf{U}, \quad (21)$$

where $\mathbf{B}$ is the matrix of spherical spline values at points of interest, and $\otimes$ denotes the Kronecker product. The matrix $\mathbf{G}^T\mathbf{G}$ will have the same dimension as that constructed from the waveform data set, and the set of linear equations that represent the constraints provided by group velocity maps can then be weighted and added to the set of equations furnished by the waveform data set. As we will see below, the group velocity data set is introduced in the inversion only after several iterations.

3 DATA AND NOISE

In this study, long period seismic waveforms and group velocity dispersion maps are used together in order to constrain the variations of crustal and upper-mantle shear wave speed and radial anisotropy. The group velocity dispersion data set is provided in the form of maps at 25 s, 30 s, 40 s, 45 s, 50 s, 60 s, 70 s, 80 s, 90 s, 100 s, 125 s and 150 s period by Ritzwoller (private communication, 2009). Shapiro & Ritzwoller (2002) explain the data and uncertainties associated with these dispersion maps. Group velocity dispersion measurements have the advantage of not being susceptible to cycle-skipping errors that beset phase measurements at high frequencies. In addition, at the same period, group velocity is sensitive to more shallow structure than is phase velocity.

Our waveform data set comprises fundamental mode Love and Rayleigh waves, which provide excellent coverage of the uppermost 300 km, long period overtones crucial to imaging the transition zone, and long period body waves which improve transition zone constraints while introducing some sensitivity to the lower mantle. In total, the waveform data set includes 21 491 Rayleigh wavepackets, 13 192 Love wavepackets, and 51 066 overtone wavepackets, of which ∼28 per cent are recorded on the transverse component. In addition, there are 12 932 wavepackets with mixed fundamental mode and overtone energy. Sensitivity tests show that lower mantle structure at most contributes a few percent to the misfit of the wavepackets that include body waves; nevertheless, we correct for lower mantle structure by using SAW24B16 model of Mégnin & Romanowicz (2000). Full waveform modelling of higher frequency waves can be computationally costly and prone to errors due to cycle-skipping or mis-mapping of multiply-reflected energy. In this study, our philosophy is to develop the waveform modelling starting at longer periods. In the future, we can extend this approach to progressively shorter periods.

We use three component long-period accelerograms bandpass filtered using a cosine-taper window with cut-offs at 60 and 400s and corners at 80 and 250s. To ensure high signal to noise level and limit the effects of possible complexity of the seismic moment-rate function, our data set is restricted to 203 earthquakes with moment magnitudes $6.0 \leq M_w \leq 6.9$. These are shown in Fig. 2. Moment tensors and source location are taken from the Harvard Centroid Moment Tensor project (www.globalcmt.org). The waveforms are recorded at broad-band stations of the global seismic network (GSN), GEOSCOPE, GEOFON and several regional networks.

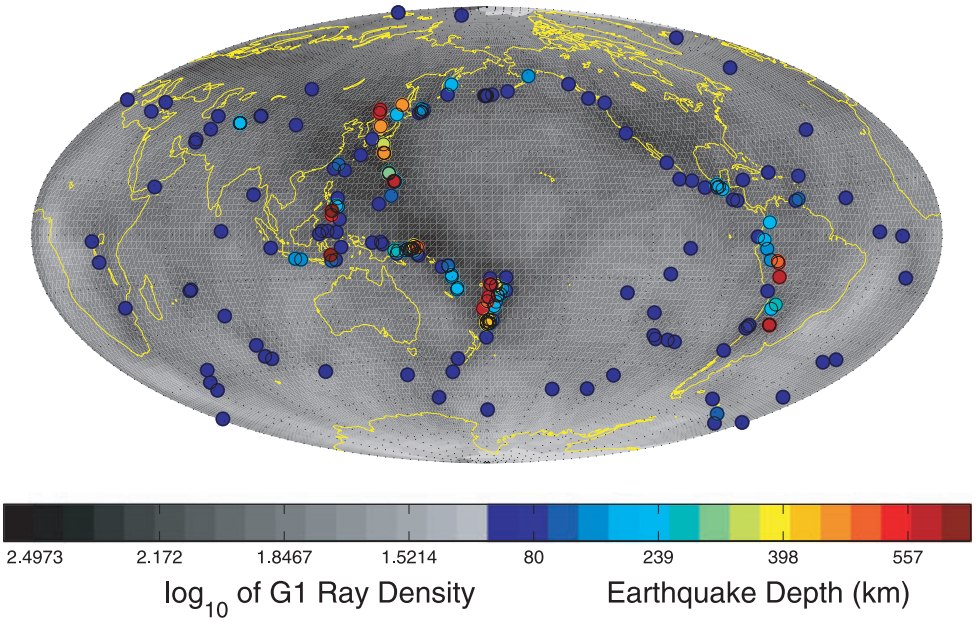

Figure 2. Map showing the 200 earthquakes used in our study, which are colour-coded according to centroid depth. The shading indicates the ray coverage number density on a log scale for minor-arc Love waves.

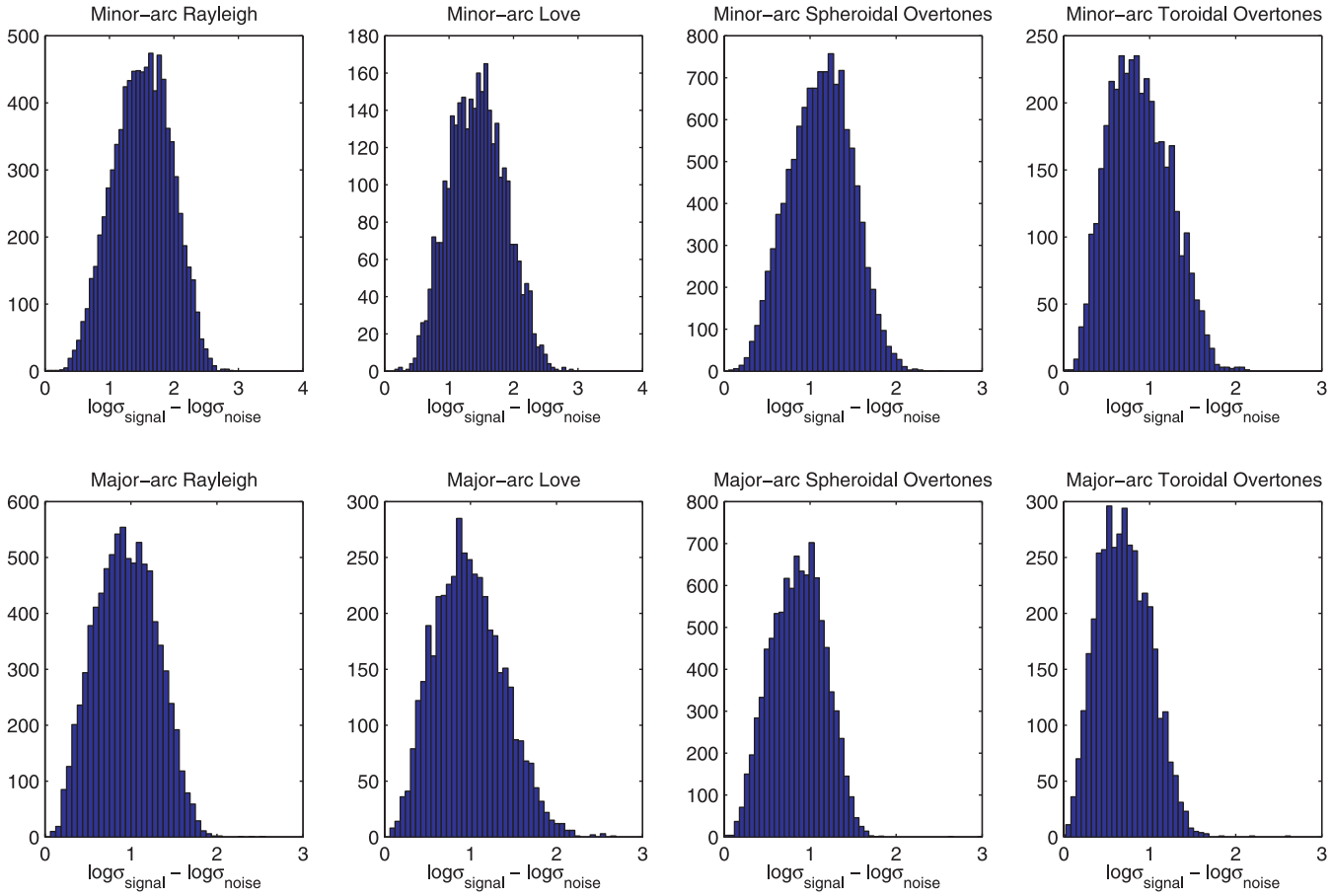

Figure 3. Histograms of the summary signal-to-noise ratios for each of the wavepacket types used in this study. The signal-to-noise ratios are approximated by taking the signal standard deviation (σ_{signal}) and dividing it by the noise standard deviation (σ_{signal}). We can see that even the least-well recorded wavepackets (second-orbit toroidal overtones) have noise levels below 20 per cent, while the minor-arc Rayleigh and Love waves have typical noise levels of only 3 per cent.

Each waveform is divided into wavepackets that isolate, in the time domain, the large amplitude fundamental-mode surface waves from smaller higher-mode waves. This allows separate weighting coefficients to be applied to the wavepackets, so that the large-amplitude signals are prevented from dominating the inversion. A detailed description of the scheme used for constructing wavepackets can be found in (Li & Romanowicz 1996, henceforth LR96). Our analysis includes both minor- and major-arc Love and Rayleigh waves and overtones since the major-arc phases provide complementary coverage to that afforded by the minor-arc phases. By including major-arc phases, we ensure much better coverage of the southern hemisphere in which there are many fewer broad-band stations compared to the northern hemisphere. Fig. 2 shows the density of ray coverage for the minor-arc Love waveform data set. The inclusion of overtones is crucial for resolving structure deeper than about 300 km, including the transition zone (e.g. Ritsema *et al.* 2004).

An automated, but user-reviewed, picking scheme is used in order to select only well-recorded accelerograms (see appendix B of Panning & Romanowicz 2006). This is done to avoid noisy data and to identify other problems including reversals of polarity, timing errors, gaps, spikes and incorrect instrument response information. The data are then hand-reviewed and the data covariance matrix $\mathbf{C}_D$ is calculated. We assess the signal-to-noise level of our data set by taking the quietest 5 min interval within the time-period as a representative sample of underlying noise. The standard deviation of the signal is then divided by the standard deviation of the noise in order to obtain a signal-to-noise summary statistic for each wavepacket. The low-noise characteristics of the data summarized in Fig. 3 justifies our picking procedure. We use the scheme proposed by LR96 to approximate the data covariance matrix $\mathbf{C}_D$ by a diagonal matrix whose entries w_i are the product of three measures of data undesirability: (1) the signal rms level; (2) data content of each wavepacket and (3) path uniqueness. The final term is crucial since it homogenizes the data coverage across the globe.

Because surface waves are sensitive to variations in both azimuthal and radial anisotropy (e.g. Montagner & Jobert 1988), accurate retrieval of variations in radial anisotropy requires that the data provide broad sampling of azimuths, so that the azimuthal dependence can be averaged out and not contaminate the model of velocity or radial anisotropy. We verify that our data set provides sufficient azimuthal coverage by binning rays passing through $10^\circ \times 10^\circ$ bins by azimuth for each component of our data set and plot them in Fig. 4 on a rose diagram.

4 INVERSION AND FITS

We initialize our iterative inverse scheme with our starting 1-D model, CRUST2 crustal velocities and Mohorovicic topography, and first invert for long wavelength structure of the mantle V_{Siso}, which we accordingly parametrize with only 162 horizontal splines.

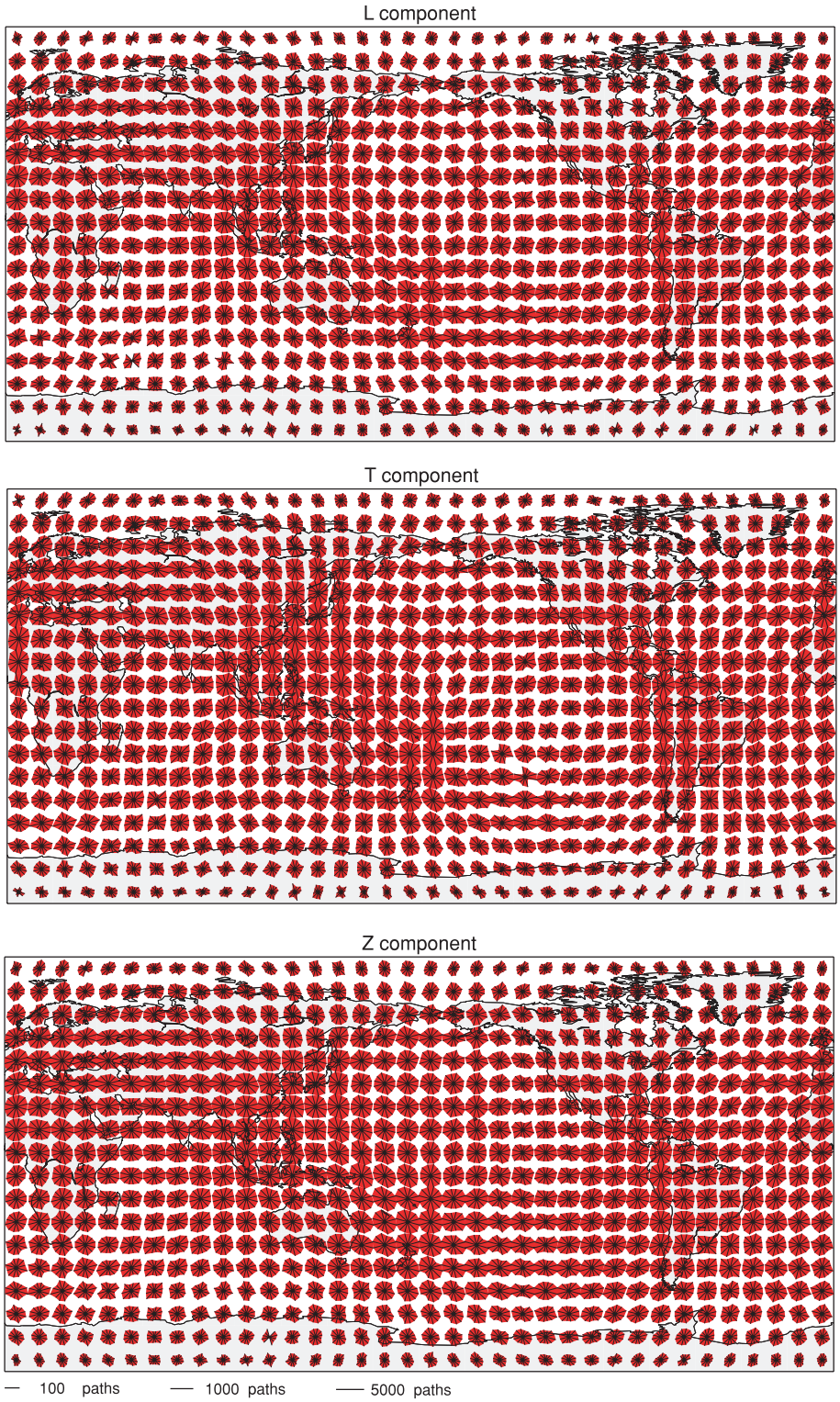

Figure 4. Rose diagrams showing the azimuthal distribution of ray paths passing through each 10° by 10° block. Note that the azimuthal coverage is good for the longitudinal (top panel), transverse (middle panel) and vertical components (bottom panel), indicating that we are unlikely to map azimuthal anisotropy into the variations of isotropic velocity and radial anisotropy.

In order to minimize computational costs, we begin the iterative scheme with a well-distributed subset (67) of the earthquakes in our data set. Once we retrieve the long-wavelength features of lateral heterogeneity, we refine our V_{Siso} horizontal parametrization to 642 horizontal splines, and expand the subset of earthquakes to 80. With subsequent iterations, we include a greater number of earthquakes until the entire 203 earthquake data set is used. Starting with the third iteration, we allow long-wavelength variations of radial anisotropy, parametrizing variations of ξ with 162 horizontal splines. We settle on a final parametrization with 2562 splines for V_{Siso} and 642 for

ξ. This corresponds to spherical harmonic expansions to degree ∼48 and 24, respectively. Thus, we enlarge our waveform data set and refine our parametrization, as we iteratively progress towards mapping smaller scale structures.

In addition to enlarging the subset of our waveform data set at each iteration in the inversion, the proportion of our waveform data set that is sufficiently similar to the synthetic waveforms and thus allowed into the inversion increases with each iteration. This is because we only use data that are sufficiently similar to the synthetic seismograms at each iteration, in order to avoid cycle-skipping problems to which waveform modelling in the time domain is susceptible. As we proceed through the iterative inversion, our model better captures the true structure of the Earth and fits to waveforms improve, thus allowing more of the waveforms to be included in the next iteration. We stress that fits improve systematically even for waveforms not included in the inversion. The fact that the number of acceptable waveforms increases with refinements to our model independently confirms the validity of our inversion scheme, our forward modelling approach, and the use of approximate sensitivity kernels G.

At each iteration, we calculate data misfits using SEM synthetic waveforms. We also recalculate the kernels for the partial derivatives matrix G in the updated 1-D model, and approximately account for the effects of 3-D structure on the partial derivatives by recalculating the frequency shifts $\omega_{kk'}$ of eq. (13). Doing this allows the NACT to at least partially capture the changes in the sensitivity of our waveforms due to changes in the structure along each path. In addition to being accounted for in SEM, crustal effects are also accounted for in the partial derivatives matrix. This is done by the introduction of additional normal mode frequency shifts $\omega_{kk'}$, calculated using the modified linear corrections approach developed by Lekic *et al.* (2010).

In the NACT formalism, the effect of 3-D structure on both g(m) and the partial derivatives matrix G is non-linear, because the frequency shifts appear in the exponent (see eq. 10). This allows us to introduce additional 'minor' iterations between SEM runs, with the goal of speeding up the convergence of the iterative scheme. Thus, in the first three iterations, which tend to produce large model updates δm, we introduce up to two 'minor' NACT iterations in which the waveform perturbation δu due only to the model update δm_k (not m_k the deviation of the current model from the 1-D profile) is added to the SEM synthetics for that iteration, and the residual $[\mathbf{g}(m_k + \delta m_k) - d]$ is approximated by $[\mathbf{g}(m_k) + \mathbf{g}'(\delta m_k) - d]$, where the NACT synthetic is primed. These approximate residuals are then inverted for another perturbation $\delta m'_k$ with an updated partial derivatives matrix. Thus, the effective model perturbation δm for kth 'major' iteration is the sum of the model updates: $\delta m = \delta m_k + \delta m'_k$. SEM synthetics are then used to calculate the exact residual for a model that incorporates this total model update, that is, $[\mathbf{g}(m_k + \delta m) - d]$.

Starting with the fifth iteration, we also invert for a smooth model of the crust. At this point, we supplement our waveform data set with group velocity maps and the associated kernels. To ensure that we use the most appropriate group velocity kernel for each location on the Earth, we use the current tomographic model at each iteration, and regionalize it into five representative profiles or radial structure (${}^i m_{p.q.}$). We then calculate the group velocity kernels for each of these canonical profiles and only use the kernel for the radial profile most similar to that beneath a given point when constructing the partial derivatives matrix G. Doing this helps to ensure that the group velocity perturbations remain in the linear regime in which the kernels are accurate.

We carried out a total of 10 iterations before our inversion appeared to converge, and misfits only marginally improved for two consecutive iterations. The final model, which we hereafter refer to as SEMum, provides >75 per cent variance reduction with respect to the starting model to the fundamental mode waveforms recorded on the longitudinal and vertical components, and 71 per cent improvement on the transverse component. For overtones, the final variance reduction is ∼40 per cent on the longitudinal component, but >50 per cent for transverse and vertical component. This needs to be considered together with the fact that the value of the final misfit for overtones is very similar to that of the fundamental mode, while the starting misfit in the latter is much larger, reflecting stronger heterogeneity in the shallow upper mantle and crust. Mixed, fundamental-overtone wavepackets had variance reductions of ∼70 per cent on all three components. Figs 5 and 6 show waveform fits before and after inversion for a typical event.Table I summarizes the final variance reductions obtained for different wavepackets. Note that they are significantly larger than for our previous waveform-based global models. Variance reduction for the group velocity data set is ∼60 per cent.

Since our waveform misfit function is affected by both amplitude and phase differences between data and synthetics, we separately analyse the contribution of phase alignment and amplitude similarity to the variance reduction for different wavepacket types and components. The results of this analysis are summarized by histograms in Fig. 7 for the vertical component, and in Fig. 8 for the transverse component. The rms waveform misfits between data and synthetics in both the starting model (grey) and SEMum (purple) are shown in the left-hand column of both figures. These are calculated by taking the square root of the variance of the residual seismogram between the synthetics and data, normalized by the variance of the data. We can see that for both components and all wavepacket types, misfit is reduced, though the reduction is more apparent for the minor-arc phases than the major-arc ones, and for surface waves than the somewhat noisier overtones.

The middle column of each figure shows histograms of the correlation coefficient between the synthetic and data waveforms. Correlation coefficients are only sensitive to phase alignment and are independent of amplitude misfits. Comparing the histograms for the starting model and SEMum synthetics, we see dramatic improvement in phase alignment for all wavepacket types, though, once again, we see poorer alignment for overtones and major-arc phases. To probe the improvement in amplitude fit, we calculate the envelopes of both data and synthetics and calculate the ratio of the ten largest data values divided by the ten largest values for the synthetics. The third column of both figures shows histograms of the natural logarithm of this ratio; a value of zero is perfect amplitude agreement, negative values indicate that synthetic waveforms are too large and positive values indicate that the synthetic waveforms are too small. SEMum synthetics clearly have more similar amplitudes to the observations than do synthetics in the starting model. This is particularly true for minor arc Love waves. This improvement in amplitude fit was obtained without allowing for lateral variations of seismic attenuation (Q), and indicates that SEMum is capable of at least partially accounting for the (de)focusing of seismic energy by gradients of elastic structure. Accounting for these purely elastic effects is crucial for the development of higher-resolution models of attenuation in the mantle.

To better understand the relative contributions of isotropic and anisotropic structure to the misfit reductions provided by SEMum, we have performed SEM simulations of 16 randomly chosen events through a model identical to SEMum in every way, except that

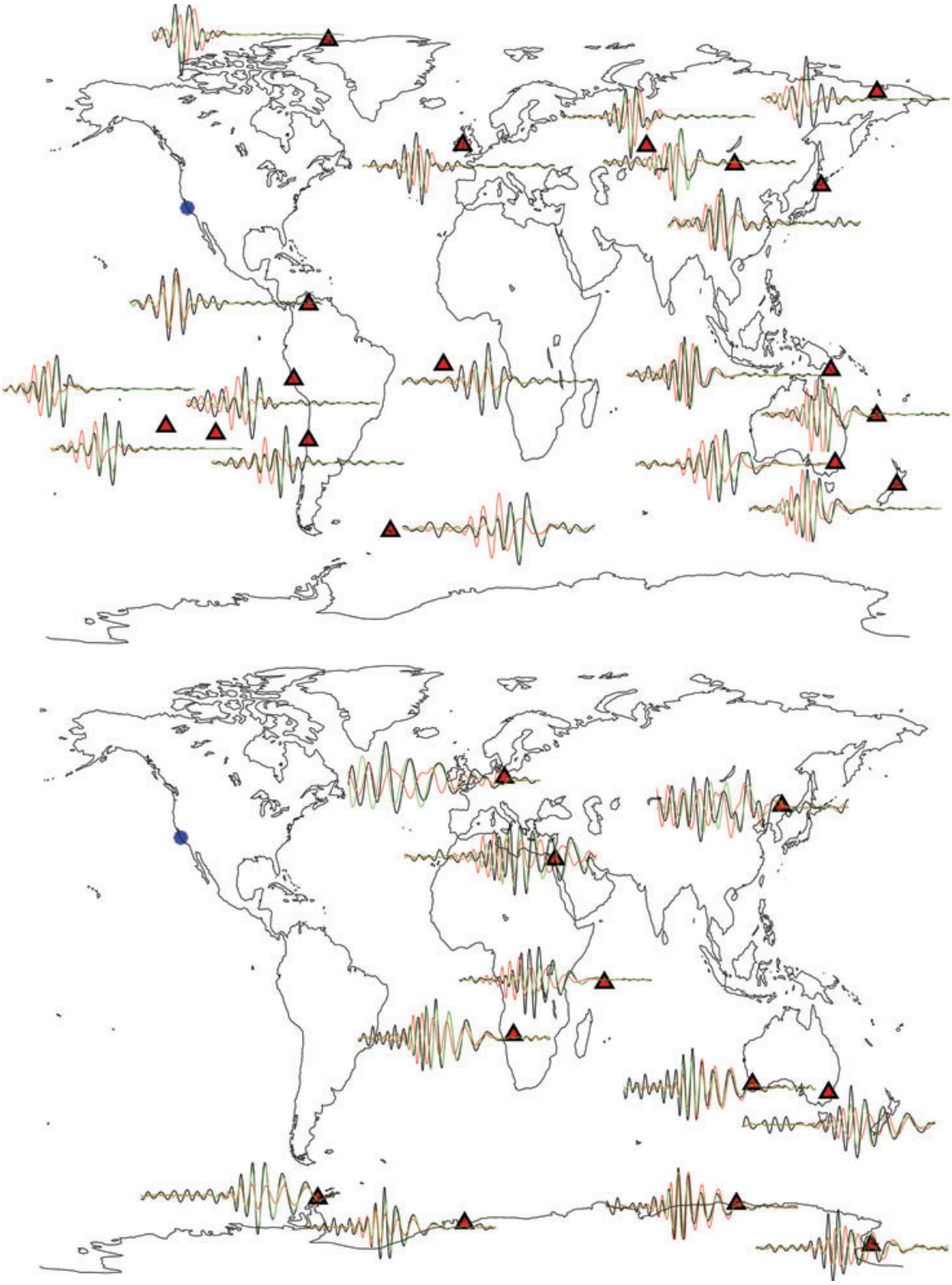

Figure 5. Observed minor arc (top panel) and major arc (bottom panel) Rayleigh waveforms (black) are compared to synthetic waveforms predicted by the starting model (red) and SEMum (green). The earthquake (blue) is the 2003 San Simeon earthquake and the station locations are marked by red triangles.

no variations of anisotropy from the 1-D component of SEMum are allowed in the mantle (depths >60 km). We find that removing lateral variations of ξ in the mantle increases the misfit of minor arc Rayleigh waves by 5.0 per cent and 6.0 per cent for L and Z components, respectively, and decreases by 2.6 per cent the misfit of minor arc Love waves. For major arc phases, misfit for Rayleigh waves increases by 3.9 per cent and 4.0 per cent for L and Z components, and decreases by 1.8 per cent for Love

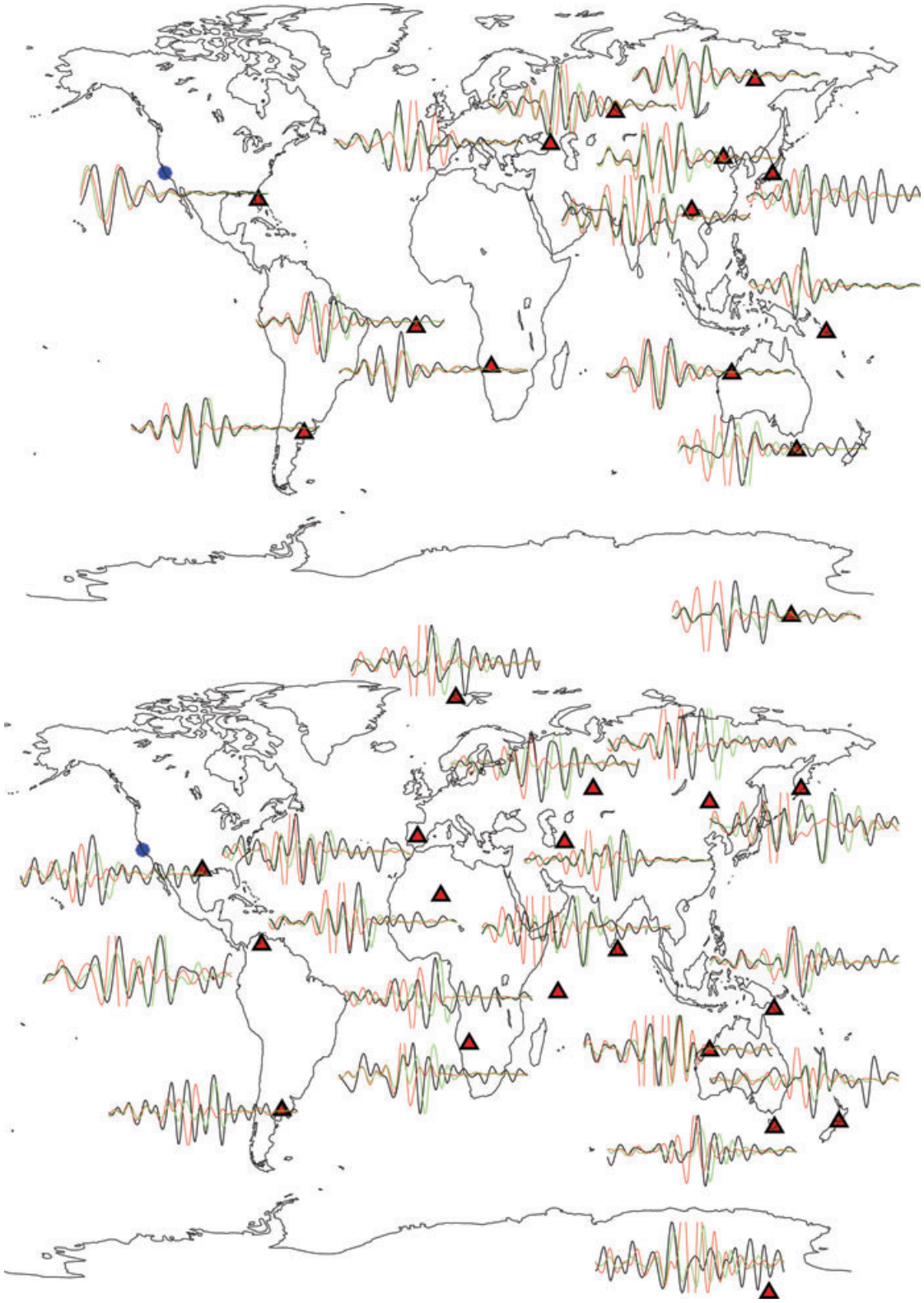

Figure 6. Observed minor arc (top panel) and major arc (bottom panel) Love waveforms (black) are compared to synthetic waveforms predicted by the starting model (red) and SEMum (green). The earthquake (blue) is the 2003 San Simeon earthquake and the station locations are marked by red triangles.

waves. Thus, weighing Love and Rayleigh waves equally—even though Rayleigh wavepackets are more numerous in our data set—we find that lateral variations in anisotropy account for a 2–3 per cent reduction in misfit. While we cannot rule out the possibility that the isotropic component of the model would not have been different if lateral variations of anisotropy were not allowed in the inversion, we believe this test confirms that the signal of lateral variations of anisotropy is approximately a few percent, in

Table 1. Final variance reduction as a function of wave and wavepacket type, and number of wavepackets used in the first and last iterations.

Wavepacket	L			T			Z		
	Per cent VR	no. start	No. end	Per cent VR	no. start	No. end	Per cent VR	no. start	No. end
Fundamental	75	4938	7968	71	7957	13192	78	8376	13523
Overtone	39	9151	14403	52	8853	14478	54	14007	22185
Mixed	69	1877	3423	70	2357	4579	70	2716	4930

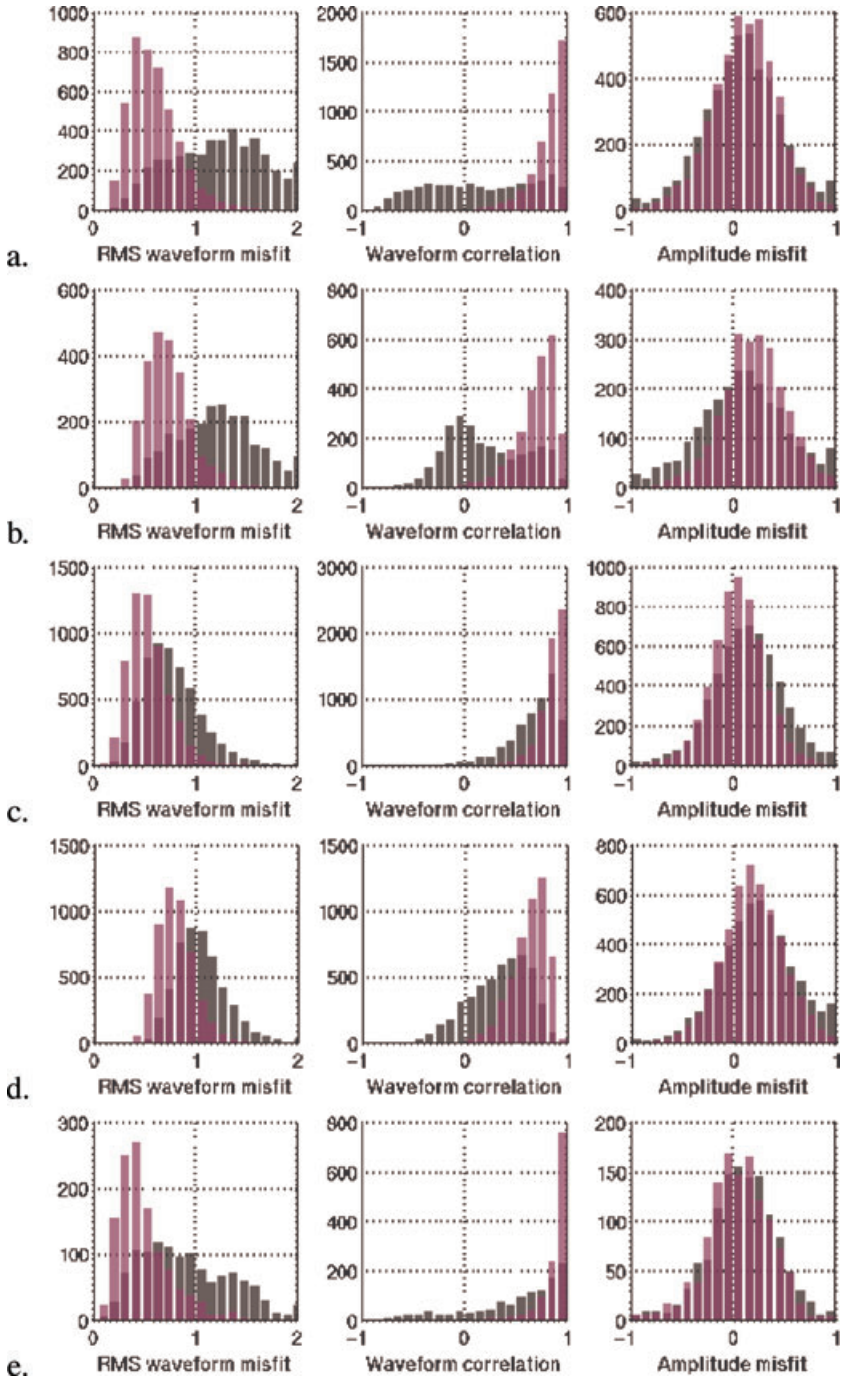

Figure 7. Measures of misfit between observed waveforms and those predicted by the starting model (grey) and SEMum (purple) for the vertical component. Left-hand panels show histograms of root-mean-squared misfits normalized by the observed waveforms. The centre panels show histograms of correlation coefficients between data and synthetics, which are only sensitive to phase alignment. The right-hand panels show histograms of the natural logarithm of amplitude ratios between the data and synthetics (0 = perfect fit). Different rows are for different wavepacket types: (a) minor-arc Rayleigh waves; (b) major-arc Rayleigh waves; (c) minor-arc overtones; (d) major-arc overtones and (e) mixed.

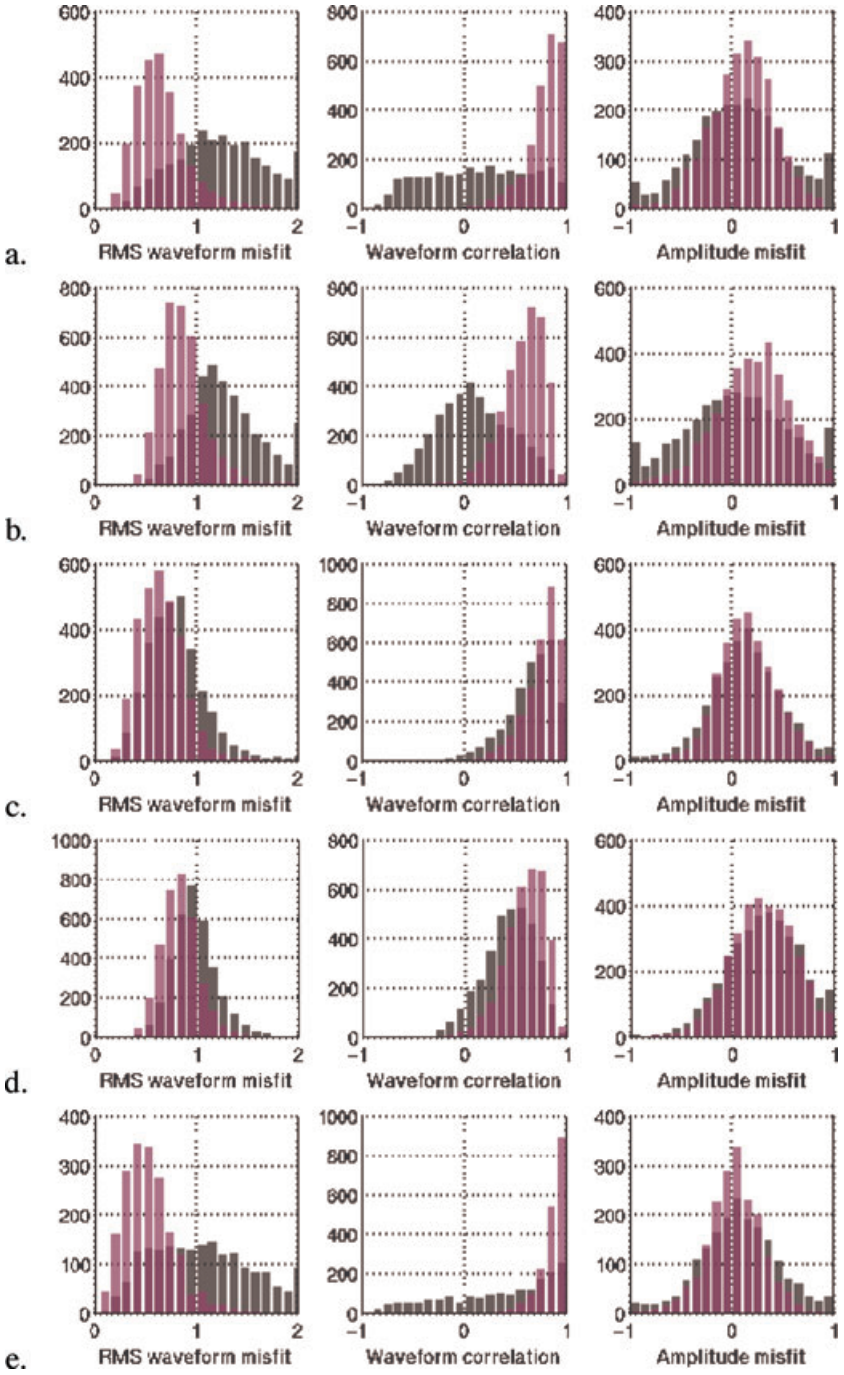

Figure 8. Measures of misfit between observed waveforms and those predicted by the starting model (grey) and SEMum (purple) for the transverse component. Left-hand panels show histograms of root-mean-squared misfits normalized by the observed waveforms. The centre panels show histograms of correlation coefficients between data and synthetics, which are only sensitive to phase alignment. The right-hand panels show histograms of the natural logarithm of amplitude ratios between the data and synthetics (0 = perfect fit). Different rows are for different wavepacket types: (a) minor-arc Love waves; (b) major-arc Love waves; (c) minor-arc overtones; (d) major-arc overtones and (e) mixed.

agreement with previously published studies (e.g. Montagner & Jobert 1988).

5 RESOLUTION TESTS

To ascertain the reliability of our model, we undertake a series of tests using the resolution matrix, calculated using the partial derivatives G of the final iteration. This analysis quantifies the resolving power of a model given the data distribution, sensitivity and noise, as well as the amount and character of *a priori* information used. However, resolution matrices are strictly only valid for linear problems, though they remain approximately valid for mildly non-linear problems (e.g. Tarantola 2005). Furthermore, they do not in any way account for inaccuracies due to theoretical and computational approximations. Because our hybrid method of tomography takes advantage of accurate SEM synthetics and thereby substantially reduces theoretical and computational errors, analysis of the resolving power of our data set based on the resolution matrix is more

appropriate in our case than for other tomographic inversions to which it is commonly applied. Using the cSEM to calculate synthetic seismograms for various input structures, and inverting those synthetics using our approximate partial derivatives matrices, would be a more accurate test; however, the computational costs of calculating the cSEM synthetics make such a test prohibitively expensive.

By applying the resolution matrix operator on a set of synthetic input models, we obtain output models which capture the ability of our data set to image the input structure. Before proceeding to explore the geographic resolving power of our data set, we conduct a set of tests that explores the expected amount of cross-contamination between isotropic and anisotropic structure in SEMum. The left-hand panel of Fig. 9 shows the retrieved V_S anomalies for an input model that contains only ξ structure, which is identical to the ξ structure of SEMum. We can see that variations of isotropic shear wave speed are not likely to be contaminated by anisotropy.

Starting in the fifth iteration, we inverted for crustal structure in addition to mantle structure. This increase in the number of model parameters has the potential for introducing trade-offs between crustal structure and mantle anisotropic structure. In order to investigate this possibility, we introduce a model with V_S structure identical to that in SEMum, but without any anisotropic structure. The right-hand panel of Fig. 9 shows the spurious ξ structure arising from the trade-offs between V_S and ξ parameters. We find that even with V_S perturbations as large as 30 per cent in the crust, contamination of ξ structure at depths of 70 km and deeper is smaller than 0.5 per cent. This test should not be misinterpreted to mean that crustal V_S structure does not have a large effect on the retrieval of mantle anisotropic structure; instead, it simply means that if the V_S crustal structure is modelled correctly—as we have done by using cSEM - the trade-offs in our combined waveform and group velocity data set are small.

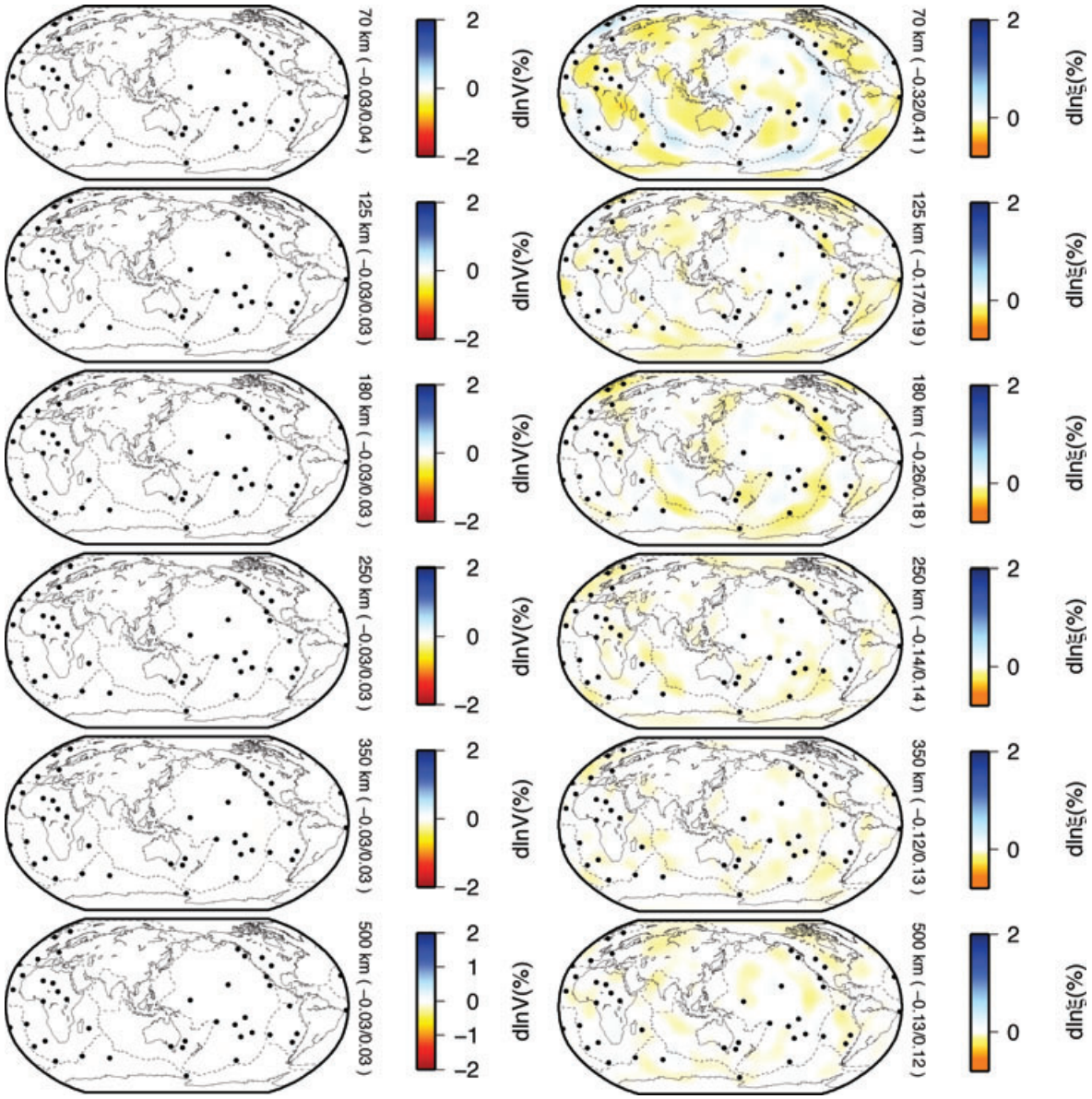

Figure 9. Left-hand panel: maps of output Voigt average shear wave speed variations with respect to the average velocity at each depth that are retrieved for an input model with no V_S variations and ξ structure identical to that of SEMum. No significant contamination of V_S by anisotropic structure is therefore expected in SEMum. Right-hand panel: maps of radial anisotropy parameter ξ that are retrieved for an input model with no ξ variations and V_S structure identical to that of SEMum. Once again, no significant contamination of ξ by V_S structure is expected in SEMum.

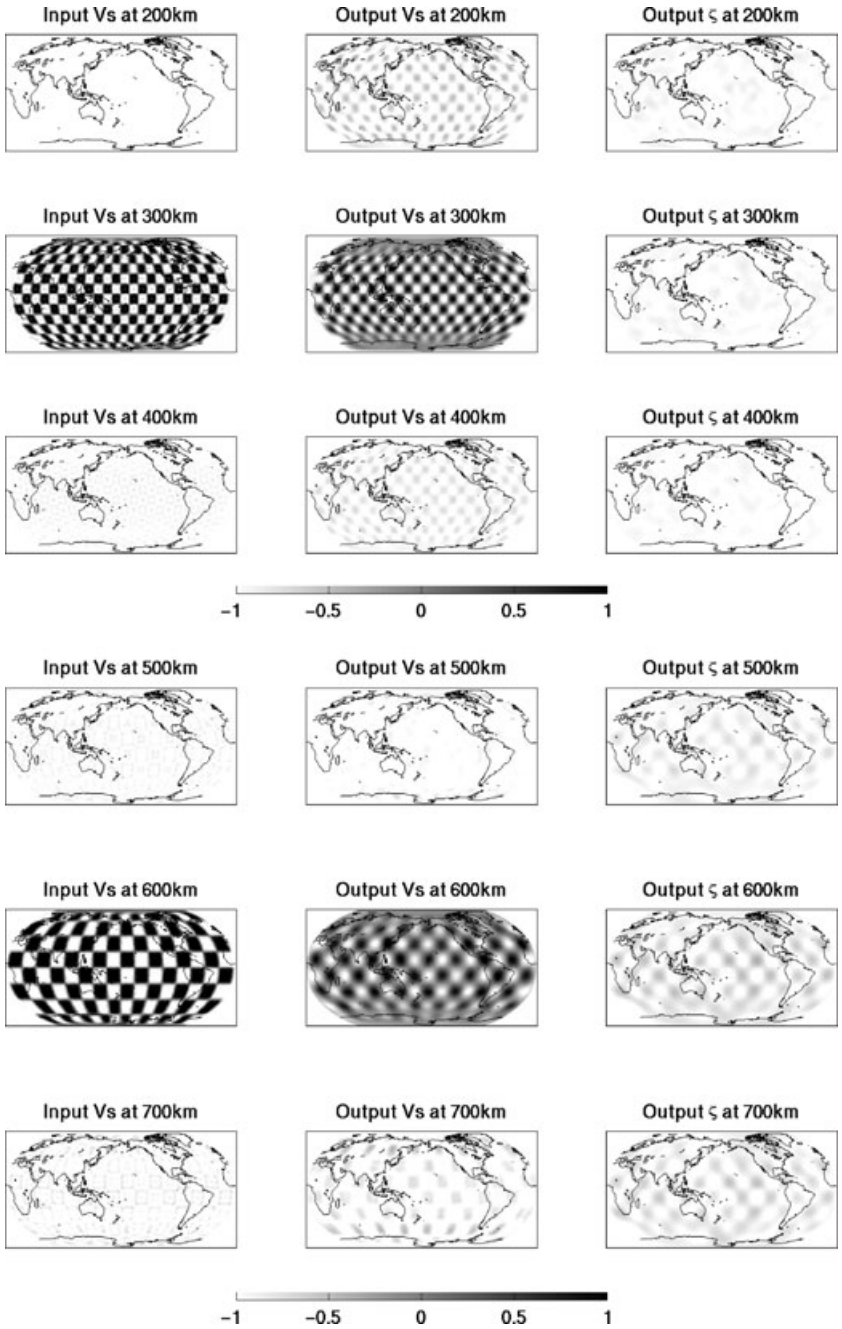

Figure 10. Tests of resolution of isotropic *Vs* structure. The input patterns are shown in the left-hand column, the retrieved *Vs* pattern is shown in the centre column, and the contamination of the anisotropic structure (ξ) is shown in the right-hand column. These tests indicate that we robustly resolve anomalies of ∼1500 km across at 300 km depth, and ∼2500 km across at 600 km depth. Resolution is better at shallower depths. Furthermore, there is very little depth-smearing of structure (<100 km) and negligible mapping of *Vs* structure into ξ.

We explore the resolving power of our data set at different depths by considering a set of input checkerboard patterns of various lengthscales. Fig. 10 shows checkerboard tests in which the input model contains only V_S variations; we show both V_S and ξ variations of the output model. At 300 km depth, we are able to robustly resolve both the amplitude and pattern of isotropic shear wave speed variations with lengthscales of ∼1500 km. Patterns with larger scale features are also robustly retrieved, and the smallest resolved lengthscale is even shorter at shallower depths. At a depth of 600 km, however, our resolution degrades, and we can only robustly retrieve V_S variations that are 2500 km across or bigger. Furthermore, whereas contamination of ξ structure was undetectable at 300 km depth, it is small but present in the transition zone. In particular, adding more intermediate depth and deep events to our data set and increasing the frequency range to include more body wave energy should help improve resolution in the transition zone.

Checkerboard resolution tests shown in Fig. 11 demonstrate that our resolving power for variations of ξ is weaker than for V_S. At 300 km depth, the minimum lengthscale of robustly imaged ξ structure is somewhat smaller than ∼2500 km. However, in the transition zone, we are only able to resolve anomalies 4000 km across. While no significant contamination of V_S structure by variations in ξ are seen

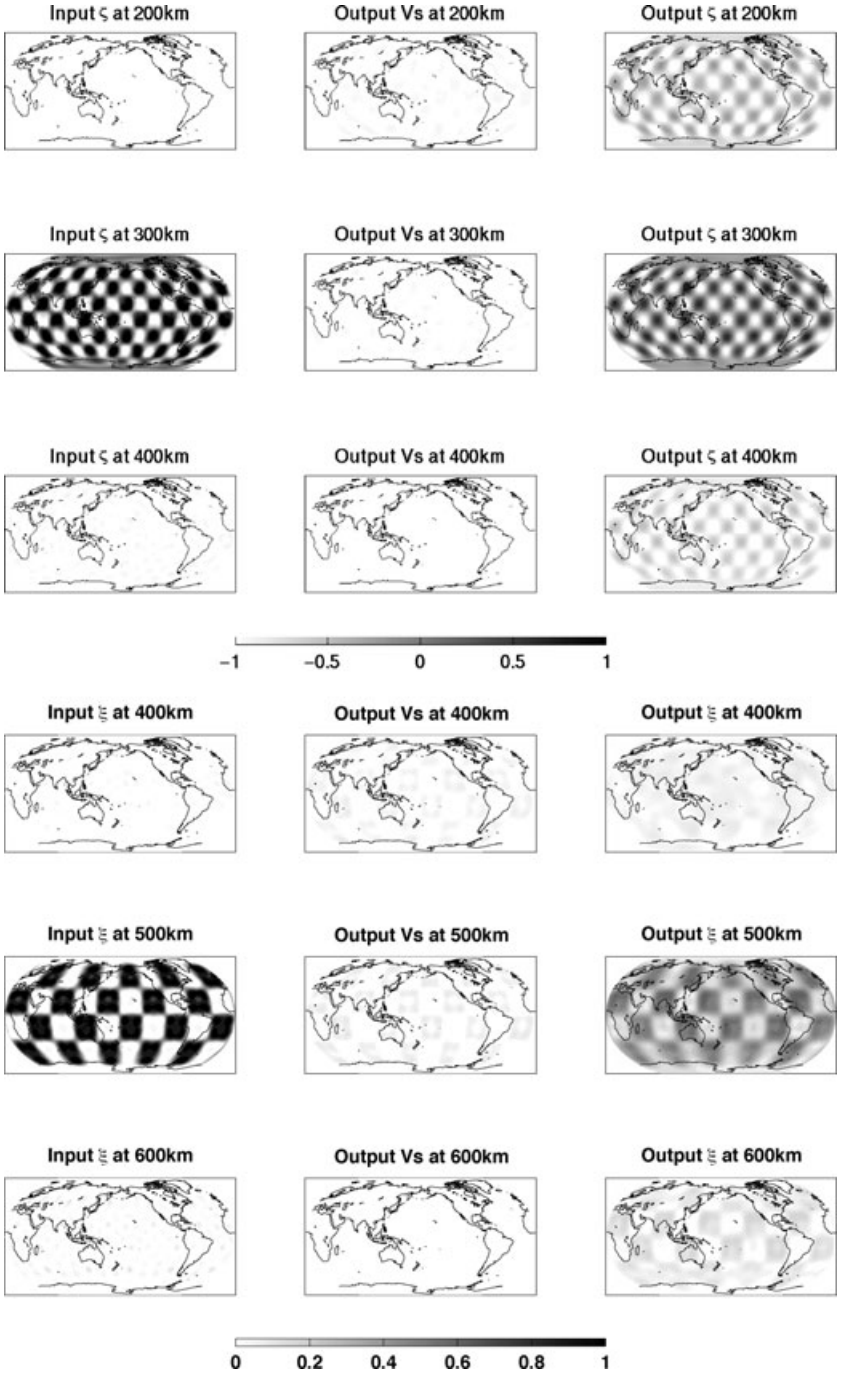

Figure 11. Tests of resolution of anisotropic parameter ξ. The input patterns are shown in the left-hand column, the retrieved ξ pattern is shown on the right-hand side, and the contamination of *Vs* structure is shown in the centre column. These tests indicate that we robustly resolve anomalies of ∼2500 km across at 300 km depth, and ∼4000 km across at 600 km depth. Resolution is better at shallower depths. While there is very little depth-smearing of structure (<100 km) and negligible mapping of ξ structure into *Vs* for well-resolved structures, both effects increase for shorter-lengthscale anomalies.

at either depth for the chosen checkerboard lengthscales, we note that smaller scale variations in ξ map strongly into V_S variations at 600 km depth. These tests show that our data set of overtone wavepackets needs to be expanded in order to provide resolution of anisotropic structures shorter than 4000 km in the transition zone. In a separate manuscript (Lekic & Romanowicz, 2010), we perform a clustering analysis of the velocity/depth profiles of our model at each geographical location, which allows us to objectively define reference shear velocity profiles for the main tectonic regions on the Earth, showing good agreement with regional studies, where they exist.

While the group velocity maps that we use to supplement our waveform data set are primarily sensitive to structure in the upper 60 km, they may also contribute to the mantle structures of SEMum. In order to quantify how much this data set contributes to the resolution of structure as a function of depth, we have calculated the *a posteriori* model covariance matrices using the waveform data set alone, and again with the combined data set. The diagonals of

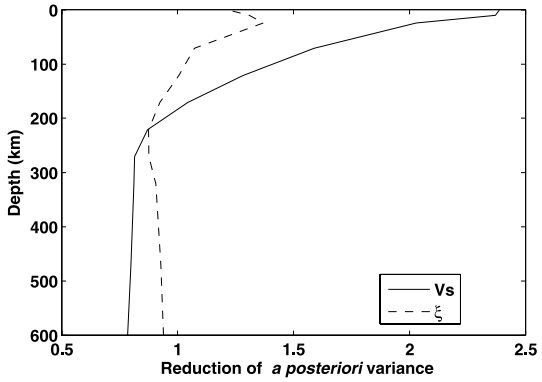

Figure 12. Ratio between the *a posteriori* model covariance of the combined data set used in creating SEMum and of a waveform only data set that is not supplemented by group velocity maps. The covariance is averaged for each depth spline, and larger numbers indicate that the addition of the group velocity maps improves the resolution of the V_S (solid line) and ξ (dashed line) structure at that depth. The group velocity maps improve the retrieval of V_S structure in the top 125 km (top 4 splines) and ξ structure within the crust.

these matrices represent the *a posteriori* variance of the model parameters of SEMum, that is, they quantify how well constrained the values of the parameters are. The ratio of the diagonals of the two covariance matrices represents the reduction in variance accomplished by the introduction of the group velocity data set. We plot this ratio in Fig. 12, which shows that the addition of the group velocity data set reduces *a posteriori* model covariance of V_S in the upper ∼125 km and ξ within the crust; deeper structure is constrained nearly entirely by the waveform data set.

6 RADIAL PROFILES OF V_S AND ξ

Fig. 1 shows the global average isotropic shear wave speed and radial anisotropy of SEMum, compared to those of PREM, our starting model, and the latest 1-D reference model developed by the Harvard group (STW105: Kustowski *et al.* 2008). While the models show very good agreement at depths greater than 300 km, substantial differences exist at asthenospheric depths.

The V_S profile of SEMum is characterized by a rather narrow (<100 km) low velocity zone (LVZ) centred at a depth of ∼100 km, with slowest velocities of 4.4 km s^{-1}. The LVZ is bounded below by a rather steep velocity gradient, with velocities increasing by ∼12.5 m s^{-1} km^{-1} down to ∼200 km depth. This velocity structure is not present in STW105 or our starting model. In PREM, the very large velocity jump associated with the 220 discontinuity, which is not thought to be a global feature, may well obscure a steep gradient that we observe. Indeed, the TNA model of Grand & Helmberger (1984), obtained by forward-modelling of waveforms that traverse the western United States is characterized by a very similar LVZ to that in SEMum, albeit with lower minimum velocities as to be expected in a tectonically active region.

We leave for future work the intepretation of the radial velocity profile of SEMum in terms of thermal and compositional variations with depth. In particular, the inclusion of constraints from mineral physics (e.g. Cammarano *et al.* 2005; Xu *et al.* 2008; Cammarano *et al.* 2009) can shed light on whether the narrow asthenospheric LVZ of SEMum can be explained with temperature alone. A separate question is whether the large velocity gradients we find at the base of the LVZ are consistent with a purely thermal origin. Finally, is our velocity profile below 300 km consistent with a pyrolitic composition, or does it require enrichment in garnet-rich components as proposed by Cammarano & Romanowicz (2007).

We validate radial profiles of V_S and ξ of SEMum against measurements of frequencies of toroidal and spheroidal free oscillations on the first four overtone branches. Because we did not use any free oscillation frequencies in the inversion of SEMum, this represents an independent test of our model's predictive power. Fig. 13 shows the predicted frequencies of free oscillations for SEMum and PREM calculated using a modified MINEOS code (Woodhouse 1998). On average, our model fits measured frequencies better than PREM, even though these were used in constructing PREM. The most dramatic improvement is in the fundamental spheroidal modes, which we match almost within measurement uncertainty at frequencies higher than 5 mHz, though this comes at the expense of slightly degrading the fits at longer periods (still, we are always with 0.5 per cent of the observed frequencies). Fits to the first five toroidal overtone branches are systematically improved. For spheroidal overtones, the fits are similar to those of PREM, though they are degraded for high frequency modes of the third-overtone branch.

The discrepancies between existing 1-D profiles of ξ can be due to a number of factors, including bias due to the use of different starting models, approximate treatment of kernels in a radially anisotropic medium, use of regional kernels, different approaches to performing corrections for crustal structure, as well as different regularization schemes and data sets used. We believe that our retrieved profile of radial anisotropy is likely to more closely represent the true variation of ξ in the mantle because we: (1) reduce bias by starting from a model found by a grid search to fit measured free oscillation periods and (2) reduce crustal contamination and inaccuracies inherent in approximate techniques by using the SEM for calculating wave propagation.

No consensus exists concerning the radial profile of ξ in the upper mantle. The ξ profile of the model SAW642AN (Panning & Romanowicz 2006) obtained by long-period waveform modelling using NACT mirrors that of PREM, peaking at the top of the LVZ (below the fast lid associated with the lithosphere), and decreases down to unity by ∼220 km. Recent models obtained by the Harvard group (ND08: Nettles & Dziewoński 2008; S362ANI: Kustowski *et al.* 2008), on the other hand, find anisotropy peaking at ∼120 km, decreasing above and below that depth, and nearly disappearing by ∼250 km. The ξ profile of SEMum is very different from that in PREM, showing peak values of ξ at a depth of 150 km, which is significantly deeper than the peaks in S362ANI and ND08. Like all of these models, we do not find that V_{SH} is substantially faster than V_{SV} on average at depths below 250 km.

Independent information on expected radial anisotropy profiles can be gleaned from theoretical work. Becker *et al.* (2007) constructed models of radial anisotropy resulting from formation of lattice preferred orientation (LPO) due to mantle flow driven by prescribed plate velocities and by density differences scaled from variations of shear wave velocity. They found that inclusion of lateral viscosity variations through a pressure, temperature and strain-rate dependent olivine creep law (assuming A-type slip systems, see Karato *et al.* 2008), significantly improved the fit to the seismic models. Whether or not the authors restricted LPO formation to dislocation creep or both dislocation and diffusion creep, radial anisotropy peaked at 150 km depth, deeper than that in S362ANI and ND08. This prediction, however, agrees with the depth of largest values of ξ in SEMum, providing further indication that we successfully characterize the profile of upper-mantle anisotropy compared to previous studies.

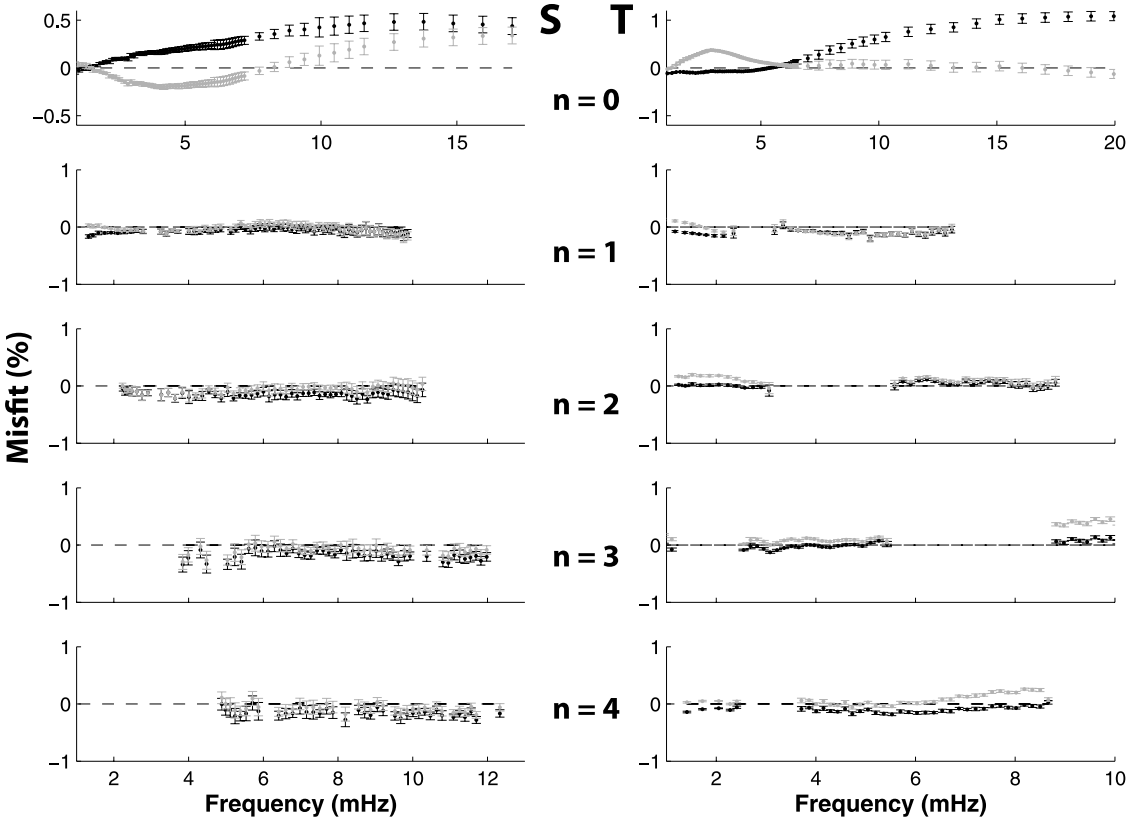

Figure 13. Predictions of toroidal (left-hand column) and spheroidal (right-hand column) eigenfrequencies of free oscillation for the fundamental branch (top panel), and first through fourth overtones. The *y*-axis denotes percent difference between observed frequencies and predictions of PREM (black) and SEMum1-D (grey).

Next, we describe the laterally varying characteristics of our upper-mantle anisotropic model SEMum. We analyse the model in the spatial (map) as well as the wavenumber domain, and consider separately the Voigt average shear velocity component and variations of radial anisotropy ξ.

7 ISOTROPIC VELOCITY VARIATIONS

Fig. 14 shows the isotropic shear wave speed variations of SEMum with respect to the average velocity at each depth. The model confirms the long-wavelength upper-mantle structures imaged previously with more approximate techniques. The most prominent slow anomalies underlie the mid ocean ridge (MOR) system down to a depth of less than 200 km. This confirms the findings of Zhang & Tanimoto (1992) but is inconsistent with the study of Su *et al.* (1992). The width of the LVZs associated with all the MORs widen with depth in the upper 150 km, though the widening is far greater beneath the faster-spreading East Pacific Rise system than it is under more slowly spreading Mid-Atlantic Ridge.

The backarcs of all major ocean–ocean convergent boundaries are also characterized by slow velocities in the uppermost 200 km, though their signature is considerably weaker than that of the MORs. The backarc of the Marianas subduction zone shows the most anomalously slow velocities at shallower depths while the low velocities associated with backarc spreading in the Tonga-Kermadec subduction zone increase in amplitude with depth and become dominant at 180 km. In contrast, subduction beneath South America shows no clear signature of a slow mantle wedge.

Finally, a number of localized low velocity features not clearly resolved in previous global shear wave speed models can be seen in the continents. At a depth of 70 km, a continuous band of low velocities can be seen running from the Tibetan plateau in the east, through the Hindu Kush, the Zagros Mountains and terminating on the west beyond the Anatolian Plateau. At similar depths, we also image a low velocity channel running from the St. Helena hotspot underneath the Cameroon Volcanic Line and terminating in a broader LVZ underlying the Hoggar, Tibesti and Darfur hotpots. Also, we find that the low velocities associated with Red Sea/East Africa rifting extend northward all the way to the Anatolian collision zone between 100 and 200 km depth.

Large-scale fast anomalies in the uppermost 200 km can be interpreted as signatures of either continental cratons and platforms or thickening oceanic lithosphere. Away from mid-ocean ridges, the ocean basins appear as seismically fast anomalies in the upper 100 km, with faster velocities persisting to greater depths with increasing age, consistent with cooling-induced lithospheric thickening (see, for example Shapiro & Ritzwoller 2002). Seismically fast keels beneath stable cratonic regions were apparent in global tomographic models a quarter century ago (e.g. Woodhouse & Dziewonski 1984), and remain one of the most prominent features of our tomographic model. Indeed, the largest difference between our model and other recent global tomographic studies is that the amplitude of the fast anomalies we observe beneath cratons is larger: up to 9 per cent faster at 125 km depth. Despite their stronger amplitudes, however, we find that the signature of the cratonic keels weakens considerably below 200 km and disappears altogether around 250 km depth. This

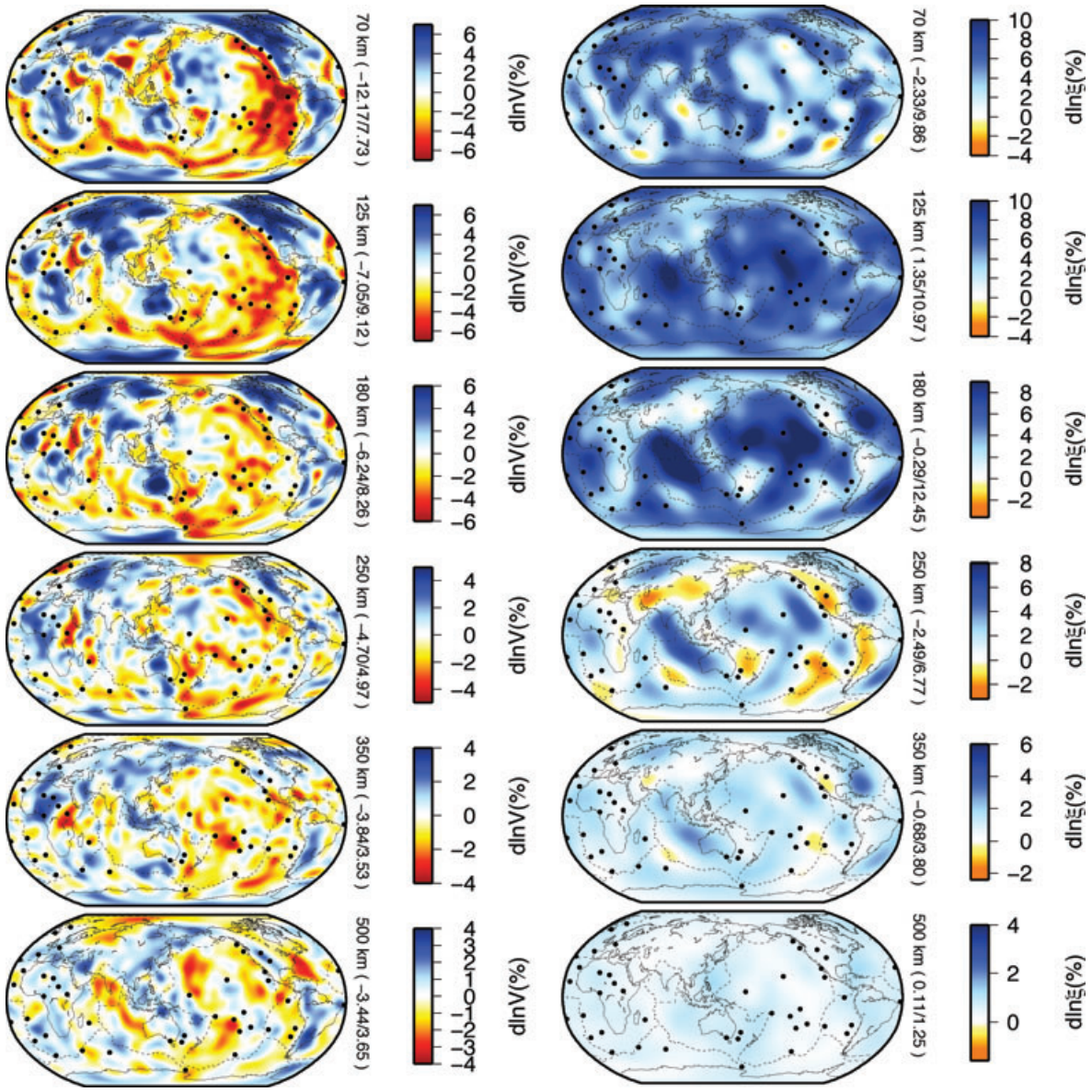

Figure 14. Left-hand panel: maps of the Voigt average shear wave speed variations with respect to the average velocity at each depth. Note that the limits of colour scales change with depth and that the colours saturate in certain regions. Right-hand panel: maps of radial anisotropy parameter ξ, showing regions in which horizontally polarized waves are faster (blue) and slower (orange) than vertically polarized waves. Note the asymmetry of the colourscale. Black circles indicate locations of hotspots from Steinberger (2000).

is consistent with the findings of Gung *et al.* (2003) and models based on heatflow measurements (e.g. Artemieva 2006) and xenoliths (e.g. Rudnick *et al.* 1998).

The spectral character of the velocity anomalies in the upper 200 km is shown in the left-hand panel of Fig. 15. In this depth range, the power peaks at degree 5, corresponding to the signature of the continent-ocean function, falling off rapidly past degree 6 or 7. This confirms that the red spectrum of mantle heterogeneities noted by Su & Dziewonski (1991) is a robust feature of the Earth and not an artefact due to the use of approximate forward modelling techniques. Power, including that at degree 5, decreases rapidly at depths below 200 km, consistent with the disappearance of the seismically fast continental keels and slow MORs. These features of the spectrum of upper-mantle velocity anomalies are also found in the models of Kustowski *et al.* (2008) and Panning & Romanowicz (2006).

Seismic structure in the 250–400 km depth range is weaker in amplitude and has a decidedly whiter spectral character than more shallow structure. It is also uncorrelated with overlying structure, as can be seen in the radial correlation function in panel A of Fig. 16. Unlike Panning & Romanowicz (2006), we do not find structures at this depth range to be anticorrelated with overlying structures. The most prominent fast anomalies appear to be associated with subduction of the Nazca slab beneath South America, the Australian–Indian Plate beneath Java, and the Pacific Plate beneath the Aleutians, Kuriles and Japan. Fast anomalies are also seen beneath Western Africa, though they are rather weak and more diffuse than the overlying signature of the West African craton. Finally, fast anomalies are present in a few locations beneath the ridges encircling Antarctica, with the most prominent one being associated with the Australian–Antarctic discordance. In this depth range, strong (∼−3.5 per cent) low velocities appear to concentrate in

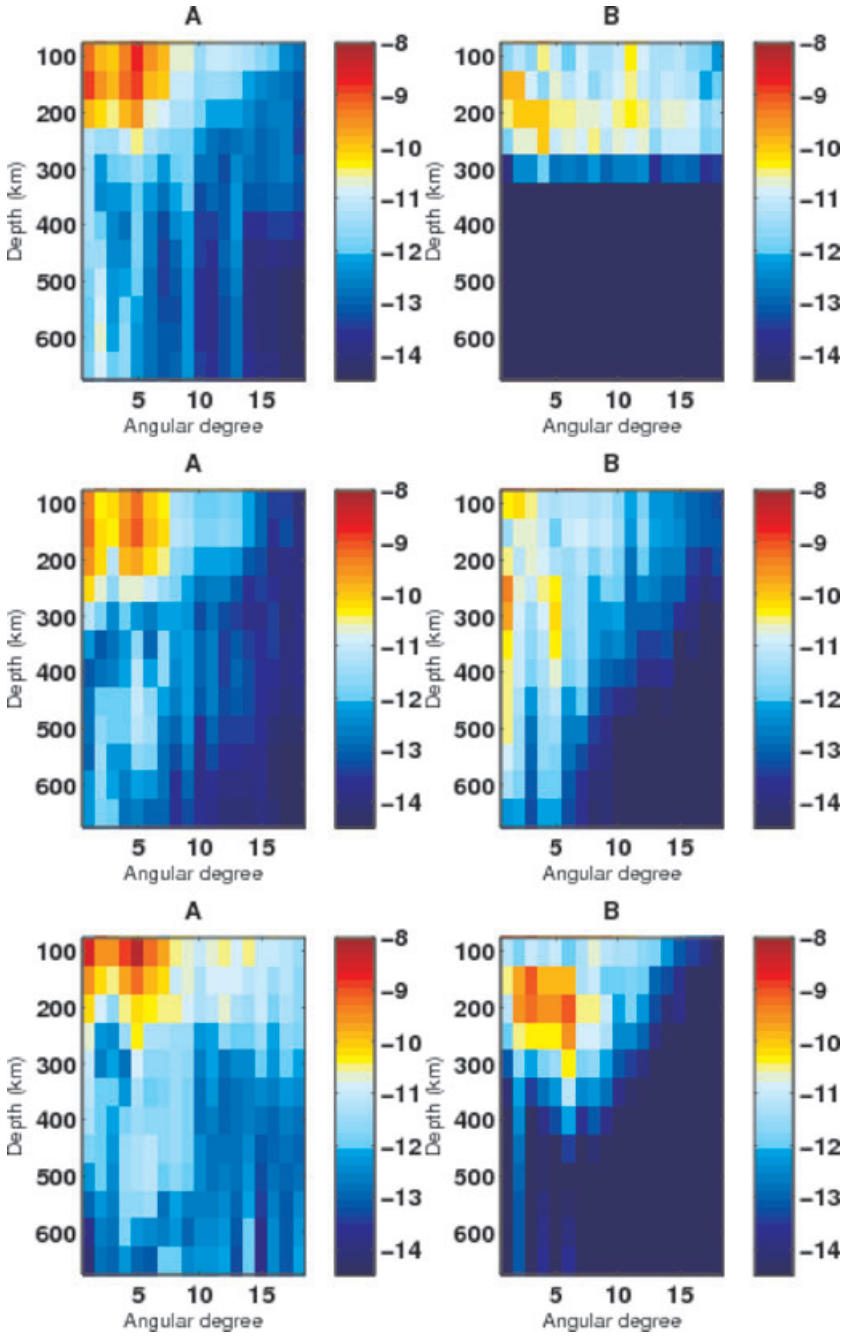

Figure 15. Power of the V_{Siso} (left-hand panel) and ξ (right-hand panel) model as a function of depth and angular degree (wavenumber). The colourscale is logarithmic. The top row is for S362ANI, middle is SAW642AN, and bottom is this study SEMum.

two regions: one centred in the southcentral Pacific in the triangle formed by the Tahiti, Macdonald and Samoa hotspots and another centred beneath the Tanzanian segment of the East African Rift. Weaker anomalies are generally seen beneath the Pacific, and, to a lesser extent, the Indian Ocean.

Fast velocity anomalies within the transition zone are dominated by the signature of subduction in the Western Pacific. These form a fast band running from Kamchatka in the northeast, over to Java in the west and beneath Fiji in the southwest. Additional strong fast velocities are seen beneath South America, associated with the subduction of the Nazca slab, and beneath the North American Cordillera, where they are likely to be associated with subduction of the Farallon slab. We image prominent slow anomalies in four broad locations of the transition zone. The first of these may be a continuation of the slow anomaly centred between Samoa and Tahiti. The second is a slow anomaly eastward of the Marianas/Japan/Kurile trenches, while a third stretches along the western margin of the Sumatra-Andaman/Java trench system. The fourth slow anomaly can be seen beneath the northwestern Atlantic abutting the North American shelf.

In the wavenumber domain, the combined signature of the seismic anomalies within the transition zone presents itself as an increase in power at degrees 4–8 (see Fig. 15), which is different from the dominantly degree 2 character of the anomalies inferred by Kustowski *et al.* (2008). Furthermore, unlike Kustowski *et al.* (2008), we do not observe a dramatic broadening of the radial correlation function

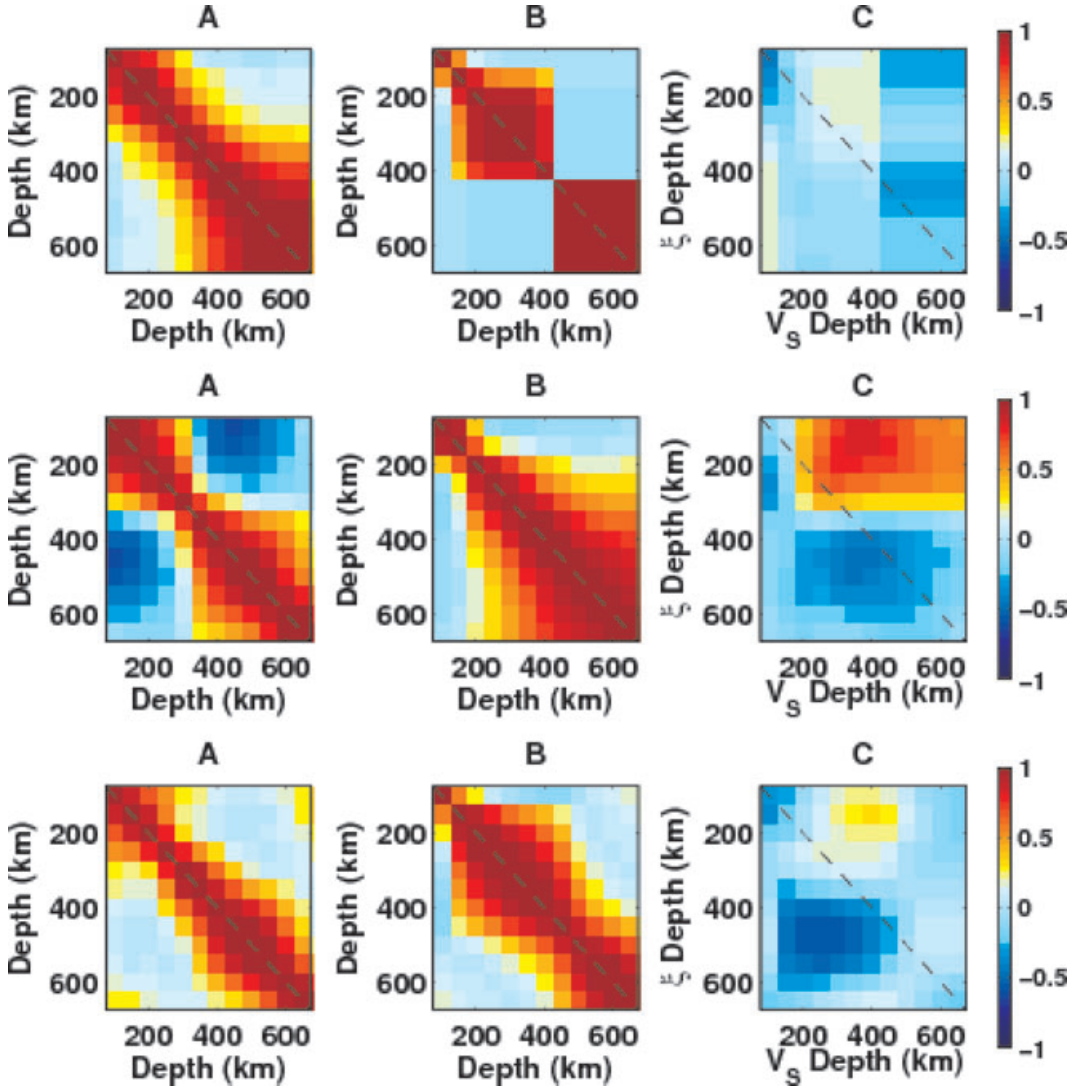

Figure 16. (A) Radial correlation function of the V_{Siso} anomalies. (B) Radial correlation function for ξ anomalies. (C) Cross-correlation between the variations of V_{Siso} and ξ. The top row is for S362ANI, middle is SAW642AN and bottom is this study SEMum.

within the transition zone. This may indicate that we image features in the transition zone resulting from flow that is not only vertical, but has a significant lateral component.

Our model of V_S and ξ variations in the crust is an 'equivalent' model that mimics the effect of the true crust on our data set. Therefore, its features cannot be interpreted directly, and we have decided against discussing them here. However, independent waveform modelling carried out to shorter periods than used in the creation of SEMum ($T > 40$ s), show that the crustal model of SEMum is superior than CRUST2, at least for paths traversing North America (H. Yuan, private communication, 2011).

8 VARIATIONS OF RADIAL ANISOTROPY

Fig. 14 (right-hand panels) shows the variations of the anisotropic parameter ξ with respect to isotropy at a variety of depths. Regions where $\xi > 1.0$ (shown in blue hues) are ones in which horizontally polarized waves travel more rapidly than vertically polarized ones, that is, $V_{SH} > V_{SV}$, and ones with $\xi < 1.0$ (shown in orange hues) have $V_{SV} > V_{SH}$. If this seismic anisotropy is due to LPO of olivine crystals induced by flow-driven deformation, then blue regions of Fig. 14 are ones in which the direction of the time-integrated longest finite strain ellipsoid is in the horizontal plane (e.g. see Ribe 1989, 1992). However, because the dominant slip systems that give rise to LPO are themselves sensitive to temperature, pressure, strain-rate and volatile-content, a variety of slip systems might be operative in the upper mantle, complicating the interpretation of anisotropy (see Karato *et al.* 2008).

Before proceeding to describe and discuss the spatial characteristics of variations in ξ, it is interesting to consider the spectral character of the model and compare it with that of the isotropic velocity variations. The right-hand panel of Fig. 15 shows the power of the anisotropic model as a function of angular degree, and coloured on a logarithmic scale. At a depth of 100 km, the spectrum is rather white, and is markedly different from the red spectrum of isotropic velocity variations. Below about 125 km, almost the entire power of the anisotropic model is contained in degrees 2–6, even though the model parametrization allows for structure up to degree 24. Finally, very little power is present at depths greater than 300 km, confirming previous results of Panning & Romanowicz (2006) and

Kustowski *et al.* (2008) that lateral variations of ξ are not strongly required by the data at these depths.

It is immediately apparent that the uppermost ∼200 km are characterized by $V_{SH} > V_{SV}$, as seen in the radial profiles of ξ, presented earlier. This is consistent with the dominantly horizontal deformation induced by the motion of lithospheric plates over the asthenosphere. Indeed, our model does not show any large regions with $V_{SV} > V_{SH}$ until below 200 km depth. That is not to say that the model in the upper 200 km is featureless. In fact, substantial differences in the anisotropic signature of continents and oceans are clearly present in this depth range.

First, continental regions appear to have larger values of ξ in the uppermost 100 km than do oceanic regions, which are essentially isotropic away from the MORs. This observation is complicated somewhat by our smooth parametrization of crustal structure, which can only match the seismic response to that of a layered crust with the introduction of spurious anisotropy. However, we believe that this effect is not dominant at a depth of 100 km. A possible explanation is that since seismic anisotropy depends not on the present but rather the time-integrated finite strain, the strength of anisotropy in the shallow continental lithosphere is the result of it having been subjected to more deformation over its considerably older age than has the oceanic lithosphere.

The second feature of interest that can be seen in the 70 km map of Fig. 14 is that the mantle wedges of the Western Pacific have decidedly greater values of ξ than do the surrounding oceans. This is also the case in the S362ANI model of Kustowski *et al.* (2008). It is not immediately apparent why the mantle wedges should have ξ larger than 1.0 when the opposite sense of anisotropy is predicted by Becker *et al.* (2007) based on A-type slip in olivine (alignment of fast axis with the direction of flow). This prediction is based on the preponderance of vertical deformation associated with subduction. One possibility is that the A-type fabric might not be dominant in subduction zones, and instead the B-type or C-type fabrics dominate, aligning the fast axis perpendicular to the vertical flow. This may be a plausible explanation, since mantle wedges have high water content (e.g. Hirschmann 2006) favouring B- and C-type fabric formation (Katayama & Karato 2006).

Mid ocean ridges at depths shallower than 100 km appear to have somewhat larger ξ values than the ocean basins, though their signature is less strong than that associated with the subduction zones. This character of MORs is also seen in S362ANI, and is also seen in the modelling of Becker *et al.* (2007). It results from A-type olivine fabric formation within a dominantly horizontal flow induced in the vicinity of spreading centres by the motion of the overriding oceanic lithosphere. However, it is surprising that the strength of the MOR ξ anomalies appears to be comparable across all the MORs, regardless of the spreading rate, which is predicted to be strongly correlated with ξ by Becker *et al.* (2007).

Finally, a band of anomalously high ξ and trending northwest–southeast across central Pacific can be seen in the 70 km map of Fig. 14. We do not have any ready explanation for this feature, and note that it has not been previously reported. However, we note that it may be associated with the strong $\xi > 1.07$ anomaly centred beneath Hawaii.

At 125 km, the ocean basins become the locus of highest values of ξ, while the continents appear more isotropic than at shallower depths. Greatest anisotropy is seen under the Pacific, centred beneath Hawaii. This anomaly was previously imaged by Montagner & Tanimoto (1991) and Ekstrom & Dziewonski (1998), and is present in models of both Kustowski *et al.* (2008) and Panning & Romanowicz (2006). Like Montagner & Tanimoto (1991), we also observe a second maximum beneath the Indian Ocean, centred south of India on the equator. This strong $V_{SH} > V_{SV}$ anomaly is clearly imaged by Gung *et al.* (2003), but is less strong in both Kustowski *et al.* (2008) and Panning & Romanowicz (2006). At this depth, the MORs and subduction zones are not easily distinguished, and are characterized by ξ values in the 1.04–1.07 range.

By 180 km, the continents appear to be nearly radially isotropic, while the ξ values underneath the oceans increase further, reaching a maximum of ∼1.12 beneath both the Pacific and the Indian Ocean, and somewhat lower values beneath the Atlantic Ocean. The most notable feature of the variations in radial anisotropy in this depth range is the emergence of three nearly isotropic regions: one beneath the backarc associated with subduction beneath Tonga-Kermadec, a second one near the western edge of the Southeast Indian Ridge, and a third one in the general vicinity of the triple junction between the East Pacific Rise, the Pacific-Antarctic Ridge and the Juan Fernández Ridge.

These three isotropic regions become more anomalous with increasing depth and by 250 km show clear evidence of $\xi < 1.0$. Other regions with $\xi < 1.0$ can also be seen at a depth of 250 km: a band running along the western margin of both North and South America from the Yukon in the north to central Chile in the south, and another, east-west trending band stretching from Iran in the west through China, Mongolia and Manchuria in the east. All of these regions appear to be associated with either spreading or subduction, and it is likely that their anisotropic signature is indicative of the prevalence of vertical flow. This can be seen in another way by looking at the cross-correlation between the isotropic and anisotropic structure shown Fig. 16(C): anisotropic structure below 200 km depth is moderately well correlated with seismic structure in the upper 200 km, because the regions of anomalous $V_{SV} > V_{SH}$ anisotropy are preferentially located in regions associated with either spreading centres or subduction/convergence zones which are characterized by shallow low isotropic velocity anomalies. Beneath the MORs, we expect this flow to be upward, while it is reasonable to expect flow to be downward in regions of convergence/subduction. We note that these regions are broadly consistent with the models of Gung *et al.* (2003) and Panning & Romanowicz (2006), and to a lesser extent that of Kustowski *et al.* (2008).

At this depth, the character of anisotropy beneath the oceans also changes substantially; whereas the mantle beneath Hawaii hosted largest ξ anomalies at 150 km, now it is conspicuously isotropic, separating broad swaths with larger ξ values to the east and the west. Furthermore, large values of ξ appear to persist to greater depth beneath the Indian Ocean and the western margin of the North Atlantic, than they do beneath the Pacific Ocean. The differences in the ξ model between the upper 200 km and deeper structure is clearly seen in the radial correlation functions shown Fig. 16(B). No substantial lateral variations of radial anisotropy are found below ∼300 km.

Gung *et al.* (2003) reported significant positive radial anisotropy under most cratons in the depth range 250–400 km, consistent with findings of Plomerova *et al.* (2002). This signal was interpreted as being related to flow in the subcontinenal asthenosphere. SEMum does not appear to show the same structure. While we have yet to identify the single reason for the differences between the lateral variations of radial anisotropy in SEMum and in the model of Gung *et al.* (2003), we identify a number of potential reasons that may explain the difference. First, the forward modelling scheme used in our study is more accurate, especially in the treatment of crustal structure. Secondly, our starting 1-D profile of radial anisotropy is different from that of PREM, which was used by Gung *et al.*

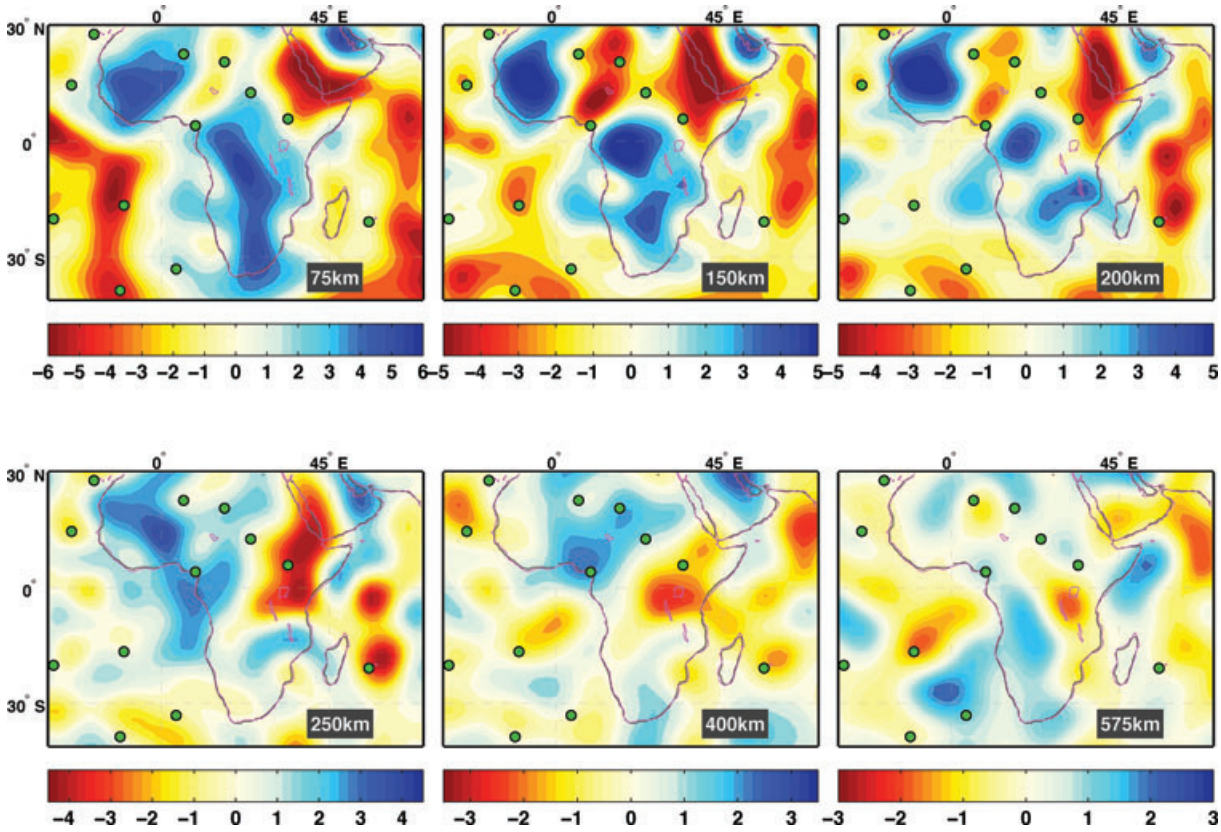

Figure 17. Left-hand panel: maps of the Voigt average shear wave speed variations in Africa and surrounding oceans with respect to the average velocity at each depth. Note that the limits of colour scales change with depth and that the colours saturate in certain regions.

Thirdly, Gung *et al.* did not use group velocity dispersion maps to help constrain shallower structure. That said, without further testing, we do not feel comfortable making any claims based on the present global model as to whether strong positive radial anisotropy is or is not present beneath the continental lithosphere.

9 COMPARISON WITH REGIONAL MODELS

9.1 Africa

Africa is the site of four main cratons, several hotspots and active continental rifting. As such, the upper-mantle structure beneath Africa has been re-examined in the last few years by a number of continental-scale tomographic studies (e.g. Sebai *et al.* 2006; Pasyanos & Nyblade 2007; Priestley *et al.* 2008). We compare our findings with inferences made in these studies and focus our attention on three salient tomographic features: (1) the differences in depth extent of seismically fast keels that underly cratons; (2) the depth extent and morphology of seismically slow anomalies beneath the East African Rift, and (3) the relationship between upper-mantle velocity and Africa's hotspots.

Even though they were first imaged a quarter century ago (Woodhouse & Dziewonski 1984), controversy still brews concerning the depth extent of the seismically fast keels beneath the West African, Congo, Tanzanian and Kalahari cratons. Based on waveform inversion of long period Rayleigh waves, Priestley *et al.* (2008) argue that the fast roots extend to depths of 225–250 km beneath all but the Kalahari craton, below which they retrieve fast anomalies only down to ∼170 km. This finding is in conflict with the study of Sebai *et al.* (2006), which found fast anomalies beneath the Tanzanian craton to be of anomalously shallow extent (∼180 km), in agreement with earlier findings by Weeraratne *et al.* (2003) whose study was focused on Tanzania. Finally, Pasyanos *et al.* per cent (2007) use a very large data set of group velocity dispersion measurements to image both crustal and upper-mantle structure beneath Africa; they find that the Congo craton is the anomalous one, with a weak signature in the upper mantle.

Fig. 17 shows corresponding map views of our model at six depths. At 150 km depth, all four African cratons are clearly seen to be underlain by fast anomalies. However, by 200 km, the signature of the Tanzanian craton is gone, and the fastest anomalies have shifted northeastward into Mozambique. This is consistent with the findings of Pasyanos *et al.* (2007) and Priestley *et al.* (2008) concerning the Kalahari craton, and confirms the shallow extent of the Tanzanian craton, as found by Weeraratne *et al.* (2003) and later Sebai *et al.* (2006). However, contrary to the findings of Pasyanos *et al.* (2007), we see a robust signature of the Congo craton extending down to ∼220 km.

The most pronounced slow anomalies shown in Fig. 17 are associated with the Red Sea and the East African Rift. At depths shallower than 150 km, these trend northwest–southeast and are concentrated beneath the Red Sea and the Ethiopian segment of the East African Rift. Starting at ∼200 km, however, they assume a north–south trend and move progressively southward with depth, extending into Tanzania, where Weeraratne *et al.* (2003) found evidence for the presence of a mantle plume. This behaviour is also seen by Sebai *et al.* (2006) and Pasyanos *et al.* (2007), but is not present in the model of Priestley *et al.* (2008), where the southern

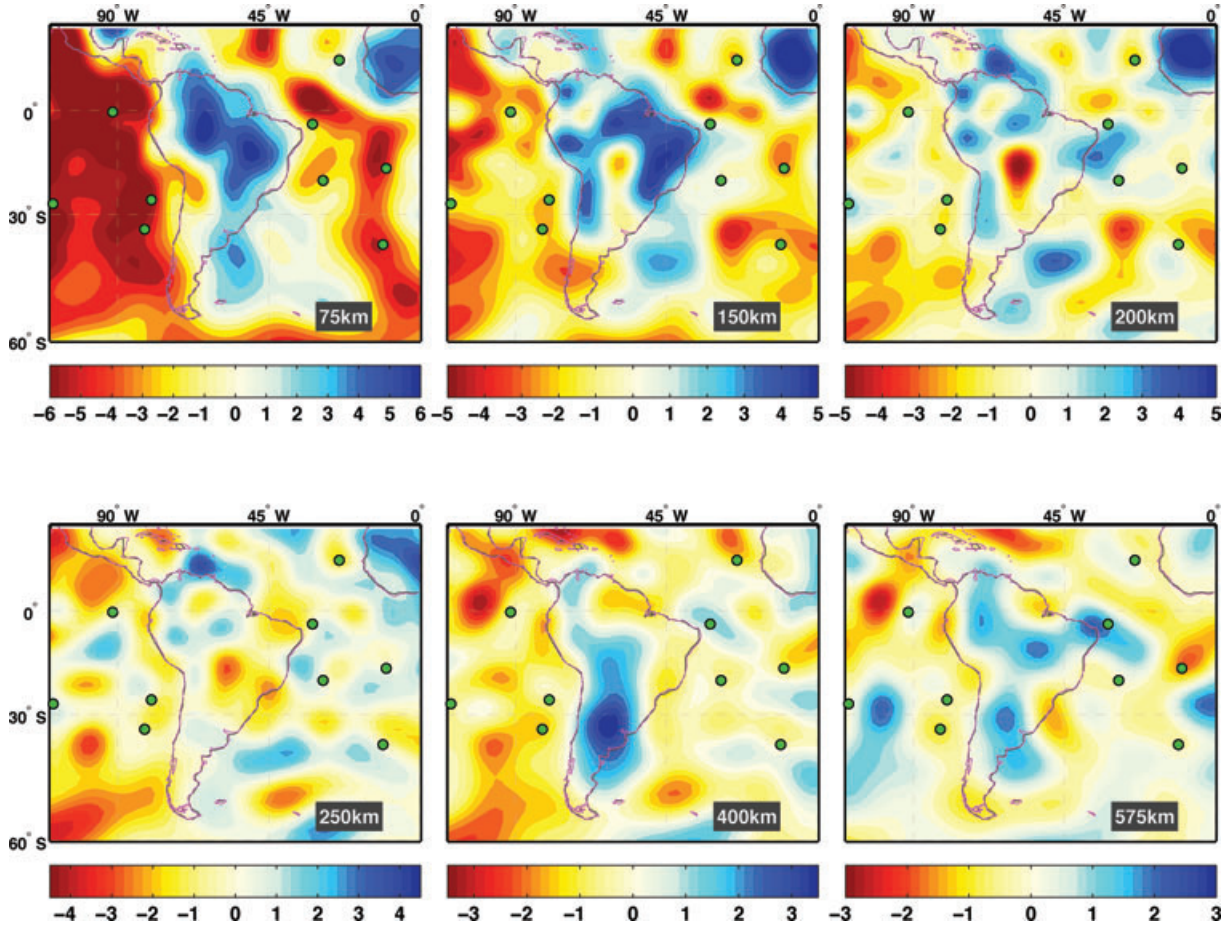

Figure 18. Maps of the Voigt average shear wave speed variations in South America and surrounding oceans with respect to the average velocity at each depth. Note that the limits of colour scales change with depth and that the colours saturate in certain regions. Green circles indicate locations of hotspots from Steinberger (2000).

East African Rift is underlain by fast velocities at depths below 200 km. In the transition zone, we find slowest velocities beneath Tanzania, where they assume a circular morphology consistent with the presence of a deep plume.

We observe secondary slow anomalies trending from St. Helena hotspot, through Mt. Cameroon and the Tibesti hotspot. These slow anomalies separate the fast keels of the West African and Congo cratons, and are also present beneath the Darfur and Hoggar hotspots. The upper-mantle signature of the African hotspots is present in both the model of Priestley *et al.* (2008) and that of Pasyanos *et al.* (2007), but is absent in the tomography of Sebai *et al.* (2006).

9.2 South America

The South American continent comprises two main cratons. The Amazonian craton which stretches from southeastern Venezuela down to northeastern Bolivia, and the Sao Francisco craton in eastern Brazil. The Amazonian craton is itself separated by Amazonian rifting into a northern Guyana and southern Guapore shields. Further south, the Parana basin is the site of a major Large Igneous Province (LIP). Active subduction of the Nazca Plate dominates the tectonics of the western margin of the continent forming the Andean Cordillera. The strike of this subduction changes dramatically between Chile and Peru, and is associated with a change in the morphology of the Wadati-Benioff zone (Lekic 2004).

Fig. 18 shows map views of our model at 6 depths. In the uppermost mantle, we find slowest velocities beneath the East Pacific Rise, and along the Carnegie and Cocos Ridges, which meet at the Galapagos hotspot. Other slow velocities are observed in the vicinity of the San Felix and Juan Fernandez hotspots, though these cease to be anomalously slow between 150 and 200 km depth. The Mid Atlantic Ridge appears to be characterized by moderately slow velocities to a depth of less than 200 km. At 75 km depth, all of South America, except the Altiplano, is underlain by seismically fast anomalies, which, by 150 km depth, appear to be centred beneath the Amazonian and Sao Francisco cratons. Unlike the regional study of Heintz *et al.* (2005), we do not image a less fast band along the Amazonian rift separating the Guyana and Guapore shields. The seismic signature of both cratonic keels narrows and shifts to the east with increasing depth, and disappears altogether deeper than ∼200 km.

We image the Nazca slab at 150 km depth, though at a depth of 200 km one of the most prominent features is not the slab itself, but, rather, a slow anomaly centred immediately to the east of the bend in the trench. This slow anomaly is also present in the model of Heintz *et al.* (2005), and might obscure the fast anomalies associated with the slab. At greater depths, this anomaly spreads to the southeast, where it underlies the Parana LIP. Heintz *et al.* (2005) also observe slow velocities, though in a more restricted region, that they interpret as a mantle signature of the Parana LIP. In the transition zone, a broad, fast, north–south oriented feature is seen, probably due to

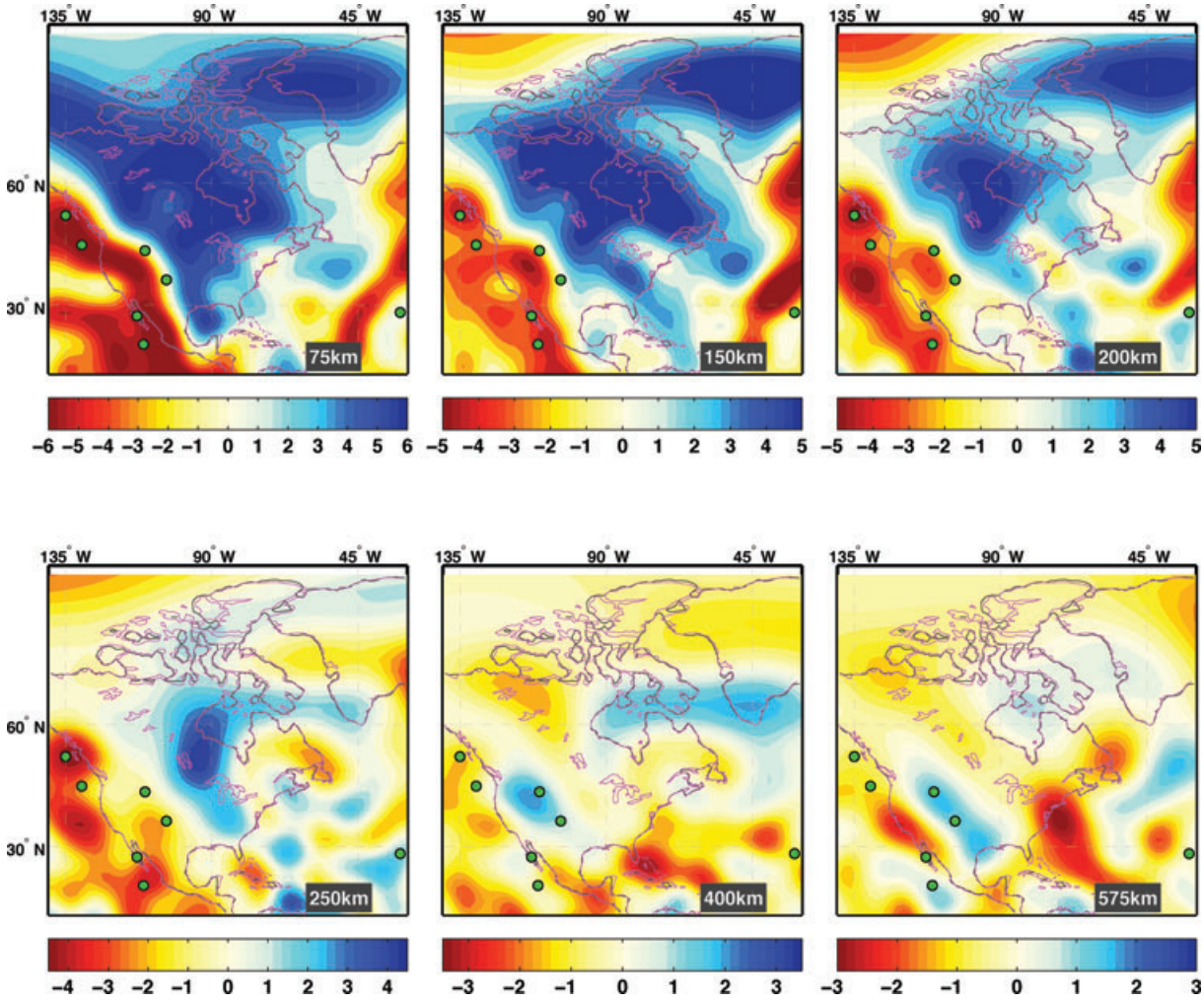

Figure 19. Maps of the Voigt average shear wave speed variations in North America and surrounding oceans with respect to the average velocity at each depth. Note that the limits of colour scales change with depth and that the colours saturate in certain regions. Green circles indicate locations of hotspots from Steinberger (2000).

the presence of the Nazca slab; deep seismicity is seen throughout the region covered by the fast anomaly. At depths below 500 km, a slow anomaly is present beneath the eastern edge of the Parana LIP, in agreement with *P* and *S*-wave regional traveltime tomography of Schimmel *et al.* (2003).

9.3 North America

North American upper mantle has been mapped by a number of recent surface wave studies (Godey *et al.* 2004; Marone *et al.* 2007; Nettles & Dziewoński 2008; Bedle & van der Lee 2009; Yuan & Romanowicz 2010). Fig. 19 shows corresponding maps of the isotropic shear wave speed variations of SEMum. The most prominent seismic feature in the upper 200 km beneath North America, and one that is imaged by all of the recent tomographic studies, and also present in the earliest studies (e.g. Romanowicz 1979), is the sharp contrast between the tectonically active and seismically slow western region and the seismically fast, stable continental platform to the east. However, the details of velocity variations within each region differ between models.

At 75 km, our model shows two regions of especially fast velocities beneath the stable continent: a northwestern one in the vicinity of the Slave craton, and a larger, faster one centred on the southern shore of Hudson Bay in the location of the Superior craton. We image a third craton beneath northwest Greenland. The craton locations are broadly consistent with the morphology of fast anomalies imaged in the aforementioned regional studies. Two 'tongues' of fast anomalies appear to extend from these cratonic regions into the Atlantic Ocean. By 150 km, the fastest anomalies appear to merge, shifting somewhat northward, directly beneath Hudson Bay. At 200 km, the fastest velocities are seen in a circular region centred on the western shore of Hudson Bay, and persist until ∼250 km before becoming indistinguishable from ambient mantle. The Greenland craton loses its fast signature between 200 and 250 km depth.

A number of smaller-scale features can be seen in the seismically slow western portion of the continent. The most striking of these is a less-slow band at 75 km which stretches from the California coast towards the Pacific. We see a sharp drop of velocities across the Mendocino Transform Fault that separates the Pacific Plate from the Juan the Fuca Plate to the north. The southern edge of this band occurs at the tip of active rifting occurring in northern Gulf of California. Because this feature appears to be confined to the strike-slip San Andreas Fault plate boundary, and its signature disappears below 150 km depth, we interpret this feature as the manifestation of colder oceanic lithosphere that is no longer subject to active spreading occurring to the north and south.

In the east, slow velocities are seen in a narrow band around the Mid Atlantic Ridge. A small, circular low velocity anomaly is imaged in the vicinity of Bermuda. This anomaly may be associated with a weak, northwest–southeast trending band of slow anomalies that splits the domain of fast anomalies running from northern Quebec to south of the Great Lakes, before petering out near Lake Erie. Though this feature appears to persist until a depth of 200 km, it is not clearly seen in any of the regional models.

The slow anomalies seen beneath the Basin and Range disappear between 200 and 250 km, which is somewhat shallower than the signature of the slow anomalies further to the west and south. Nevertheless, our model shows that western North America is clearly anomalously slow to a depth of 200–250 km, which is also found by Nettles & Dziewoński (2008) and Bedle & van der Lee (2009) but is opposite to the maps of Godey *et al.* (2004).

In the transition zone, we image a northwest-southeast trending fast anomaly that stretches from the Cascadia subduction zone down to the Gulf of Mexico. We interpret this to be a signature of the Farallon slab. The location of this fast anomaly is roughly consistent with the images of the slab-related fast anomalies imaged using the finite-frequency, teleseismic *P*-wave traveltime model of Sigloch *et al.* (2008). Two strong slow anomalies are also seen in this depth range: one beneath the central segment of the East Coast of North America, stretching from Massachusetts in the north, down to southern Virginia, and a second, smaller anomaly beneath western/central California.

9.4 Australia

A favourable distribution of earthquakes that occur at a large range of depths along the Tonga-Kermadec and Vanuatu subduction zones to the east and the Solomon Islands, Papua New Guinea, Banda Sea and Java subduction zones to the north, has aided the development of tomographic models of the mantle structure beneath Australia. We will compare our inferred velocity structure beneath Australia with three recent surface wave based tomographic studies of the continent's upper-mantle structure (Simons *et al.* 2002; Fishwick *et al.* 2005; Fichtner *et al.* 2009). All three of these studies use only vertical component seismograms, and are thus models of vertically polarized shear wave speed variations. The model of Fichtner *et al.* (2009, henceforth FAU) is, like our model, developed using the SEM, though there are a number of important differences between our approaches: (1) we use three component data, whereas FAU uses only vertical component seismograms; (2) we initialize our inversion with a 1-D model, whereas FAU start from a 3-D model that shares much of the features of their final model; (3) we use approximate finite frequency kernels calculated using NACT as opposed to the adjoint kernels used by FAU and (4) our misfit function is a waveform difference calculated point-by-point in the time domain, whereas FAU use a more complicated technique that calculates time-frequency misfits.

Fig. 20 shows map views of our model at a variety of depths. At 75 km depth, we see very low velocities associated with spreading occurring along the Pacific-Antarctic and Southeast Indian Ridges, as well as the Tonga-Kermadec backarc. All of Australia is characterized by faster-than-average velocities, except the easternmost margin and the southeast region near Tasmania. Simons *et al.* (2002, henceforth SAU) and Fishwick *et al.* (2005, henceforth FSW) both find low velocities beneath Tasmania at this depth, though FAU does not. The fast anomalies in the bulk of the continent show a less-fast central region, flanked by fast anomalies to the north, east and west (but not south), consistent with findings of FAU and FSW but not SAU, whose model appears more or less-uniformly fast in the entire region west of the Tasman Line. FSW point out that these lower velocities in the central portion of Australia are confirmed by body wave data.

At 150 km, central and western Australia (west of the Tasman Line) is seismically fast, with the fastest velocities concentrated in an east–west elongated region. This fast anomaly has a similar shape and amplitude in all of the regional studies. At this depth, we also start to image the subducting slabs beneath Java, the Banda Sea and Vanuatu, though the Tonga slab is not seen to be anomalously fast. This may be due to the strength of the low velocities associated with backarc spreading, whose amplitude increases with depth, peaking between ∼150 and 200 km depth. Of the three regional studies, only the model of FAU extends sufficiently far east to cover the Vanuatu subduction zone; however, they do not image any increased velocities corresponding to subducting slabs. The slow anomalies seen in the MORs south of Australia cease to be continuous in this depth range. In fact, by 200 km, only a narrow sliver of low velocities persists along the northern edge of the spreading centre.

By 200 km depth, the fast anomalies beneath central Australia have somewhat shrunk in their eastern reach, and only the central region appears anomalously fast at ∼250 km depth. All three regional models find fastest anomalies at 250 km depth to be in north-central Australia, consistent with the location of the fast anomaly present in our model. However, we are unable to resolve fast velocities in the southwestern corner of Australia, which are especially prominent in SAU and FSW, and somewhat weaker in FAU; this may be due to contamination by small-scale variations of radial anisotropy. At 250 km, two fast anomalies appear, one at each end of the Australian–Antarctic discordance, which is a site of unusual topography, unique geochemistry (Christie *et al.* 1998) and anomalous seismic upper-mantle structure (Forsyth *et al.* 1987; Ritzwoller *et al.* 2003). While at 250 km, the eastern anomaly appears to be stronger than the western one, the western one becomes dominant by 350 km depth, and both disappear in the transition zone.

The greatest differences among the regional models and the results of our study are apparent at depths below 300 km. Aside from the fast anomalies associated with the Australian–Antarctic discordance, the only prominent fast velocities in our model at these depths are the images of the subducting slabs beneath Java, Banda Sea and Papua New Guinea. Aside from a strong low velocity anomaly beneath the southern tip of the southern island of New Zeland, the map is rather bland. This is broadly consistent with the results of FSW. However, FAU finds that almost the entire region is seismically fast at these depths, and interprets these fast anomalies as the northward extension of North Australian craton. Our model presents no evidence that would warrant such a conclusion.

9.5 Eurasia

Eurasia is the site of active continental collision (Tibet and the Mediterranean), active rifting (Lake Baikal), and its southern and eastern margin host significant shallow and deep seismicity. Nevertheless, continent-scale shear wave speed tomography is made difficult by the fact that most of the continental interior is aseismic, and seismic station coverage is sparse in Russia and the Central Asian republics. However, when a global data set is used, surface wave and overtone coverage across Asia is excellent, allowing for higher-resolution parametrization to be used within Asia (as done by Kustowski *et al.* 2008), or for smaller-scale features to be

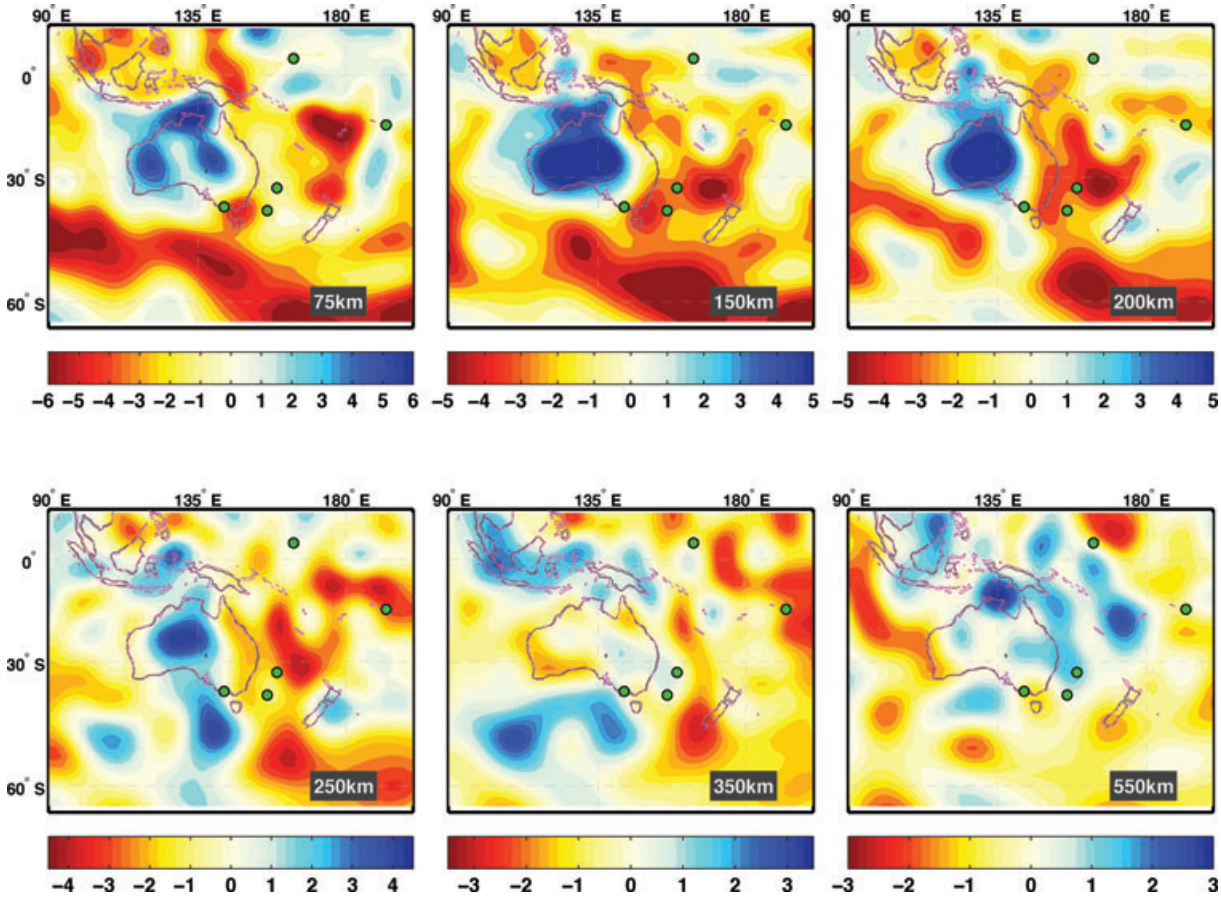

Figure 20. Maps of the Voigt average shear wave speed variations in Australia and surrounding oceans with respect to the average velocity at each depth. Note that the limits of colour scales change with depth and that the colours saturate in certain regions. Green circles indicate locations of hotspots from Steinberger (2000).

robustly imaged within a more-densely parametrized global model (as is the case in our study). Furthermore, the last decade saw the development of a number of large-scale regional studies of vertically polarized shear wave speed variations (e.g. Friederich 2003; Lebedev & Nolet 2003; Boschi *et al.* 2004; Priestley *et al.* 2006).

Fig. 21 shows map views of our model at a variety of depths. The structure of the uppermost mantle at 75 km depth beneath the northern part of the continent shows a large domain of fast velocities stretching from eastern Siberia all the way to the western margin of the East European craton. A band of somewhat slow ~ -2 per cent anomalies that extend from Tibet in the east to the Anatolian Convergence Zone in the west separate the fast velocities in the north from smaller but prominent fast anomalies that can be seen beneath the stable part of Saudi Arabia and India. This structure is clearly seen in the model of Kustowski *et al.* (2008), and the slow anomalies beneath Anatolia are seen in the model of Boschi *et al.* (2004). Small amplitude ($\sim$2 per cent) fast anomalies are imaged beneath the Tarim and Sichuan basins, bounding the low velocities of Tibet to the north and east, respectively. These small features are also imaged by Priestley *et al.* (2006) and Friederich (2003). Like Kustowski *et al.* (2008) and Priestley *et al.* (2006), we also image a prominent slow anomaly beneath the Altai Mountains of Mongolia at this depth, though this anomaly is not clearly seen in the model of Friederich (2003). Slow velocities are also seen in the mantle wedges of all the subduction zones in the east of the continent.

At a depth of 150 km, Tibet is seen to be underlain by very fast velocities, which is consistent with all the aforementioned studies. Anomalously fast mantle is once again imaged beneath the Tarim and Sichuan basins, India, and Arabia. In the north, the fast anomalies are clearly strongest beneath the East European and Siberian cratons, and are separated by a band of somewhat less fast velocities. This clear separation of the two largest Asian cratons is not obvious in either the Priestley or Kustowski tomography, but is consistent with the location of the Siberian Traps. The slow velocities that are present beneath the Altai Mountains have shifted northeastward with depth, so that they are now centred to the east of Lake Baikal. This is seen in Kustowski and Priestley tomography, but is a bit west of the structure imaged by Friederich, who found slowest velocities at this depth to be precisely beneath Lake Baikal. In the west, a notable, fast anomaly appears to be associated with the Hellenic Arc, consistent with the results of Boschi *et al.* (2004).

By 250 km depth, we see a weakening of seismic signature beneath all the cratons, with the substantial fast anomalies only persisting beneath the East European Craton. Nevertheless, smaller-amplitude fast anomalies are still seen beneath the Siberian and Arabian cratons, though their shape is considerably altered: fragmented beneath Siberia and elongated in the north–south direction under Arabia. Remarkably, the remaining small-scale fast anomalies beneath Siberia are found at identical locations by Priestley *et al.* (2006). Fast velocities are also seen beneath Tibet, in

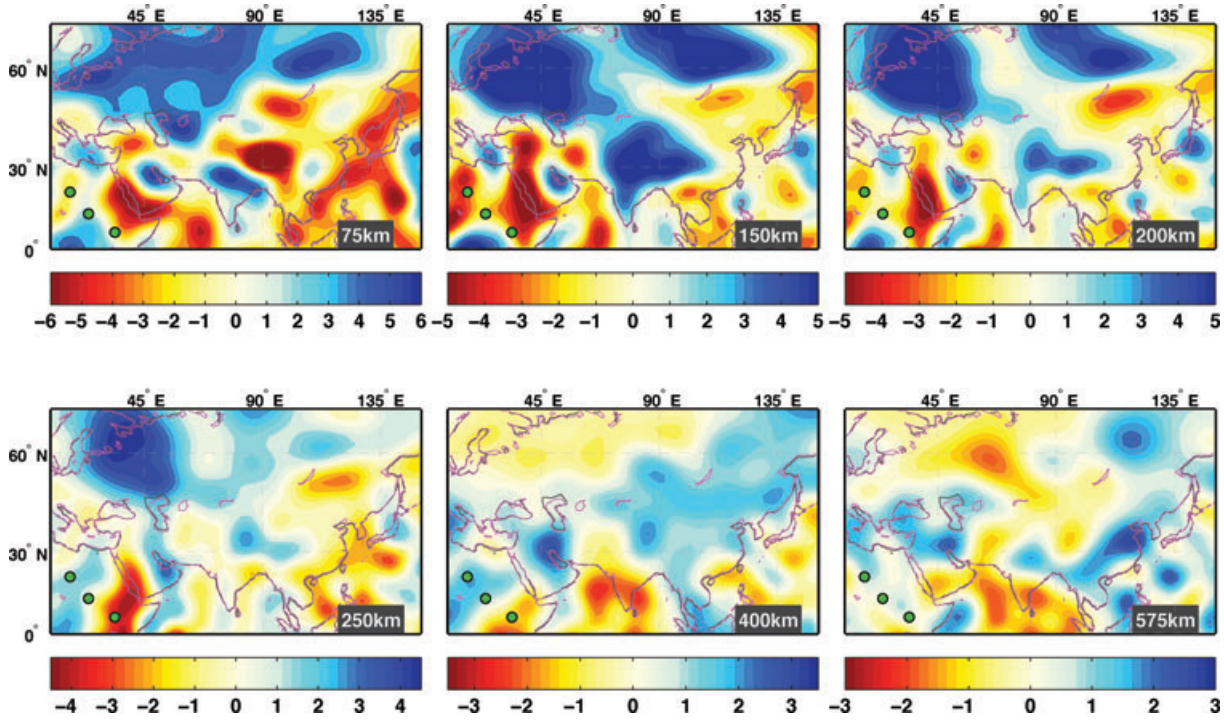

Figure 21. Maps of the Voigt average shear wave speed variations in Asia and surrounding oceans with respect to the average velocity at each depth. Note that the limits of colour scales change with depth and that the colours saturate in certain regions. Green circles indicate locations of hotspots from Steinberger (2000).

agreement with all the regional studies. Finally, the low velocities to the west of Lake Baikal persist at this depth.

The pattern of seismic anomalies changes drastically by 350 km depth. No signature of fast cratonic keels is seen at this depth, and the most prominent structure is a broad zone of fast velocities extending from the Himalayan front northward into central Siberia. Unlike Kustowski *et al.* (2008), we do not image slow velocities beneath Tibet at this depth. Furthermore, unlike Friederich (2003), who traces anomalously low velocities beneath Lake Baikal into the transition zone, we cease to resolve a clear LVZ associated with the Baikal by 350 km depth.

Within the transition zone, we image a band of fast velocities stretching from Italy into Iran, which was seen by Kustowski *et al.* (2008), and interpreted to be associated with cold, subducted material, which also elevated the 400 km discontinuity. In the east, fast velocities are seen along the entire continental margin, which is probably a signature of subduction of oceanic lithosphere. These fast velocity anomalies persist to the base of the transition zone. In this depth range, low velocities appear to underly most of central and western Russia, as well as southern India and Arabia. This is broadly consistent with the transition zone images of Kustowski and Friederich, though significant differences in details can be seen.

10 CONCLUSIONS

We developed and applied a new waveform tomography approach, which allowed us to leverage an accurate, fully numerical wave propagation modelling technique to image the anisotropic structure of the Earth's mantle. This new method reduces the contamination of mantle structure that besets widely used approximate methods, in particular due to inaccurate treatment of crustal effects. Our tomographic model is by no means an end in and of itself. Instead, its construction is important for three distinct reasons:

(i) We have developed and validated a new way of tomographically mapping the Earth's interior using the SEM and a waveform approach that allows us to include all phases interacting within a seismogram. This 'hybrid' approach to tomography can now be applied to a bigger and higher-frequency data set in order to not only better image the upper mantle, but also gain new insights into the structure of the lower mantle and make more robust regional and small-scale models of elastic structure.

(ii) We have demonstrated that the long-wavelength mantle structure imaged using approximate semi-analytic techniques is robust and validated by highly accurate forward modelling wave propagation codes.

(iii) We have demonstrated excellent agreement between our global tomographic model and images from smaller-scale tomographic studies, thus replicating on a global scale the recovery of shapes and amplitudes of lateral heterogeneity previously only furnished by these smaller-scale studies. In particular, clustering analysis conducted on the velocity profiles of our model indicates improved constraints on the amplitudes of lateral variations in shear velocity at the global scale (Lekic & Romanowicz, 2010), providing more rigorous constraints on the temperature, composition as well as flow in the mantle than those previously accessible from global modelling.

One of the main goals of seismic tomography is to image the interior structure of the Earth so as to improve our knowledge of Earth's temperature, composition and dynamics. Variations of shear wave speed that we have mapped within the upper mantle arise from variations in composition and temperature. Constraints from mineral physics can inform interpretations of observed velocities in terms of temperature and abundances of major mantle mineral phases. The average profile of shear wave speed of SEMum is characterized by a more prominent LVZ which is bounded by steeper velocity

gradients with depth than those present in other 1-D models of the Earth (e.g. Dziewonski & Anderson 1981; Montagner & Kennett 1996; Kustowski *et al.* 2008). Furthermore, we retrieve stronger anomalies than previous global tomographic models; these amplitudes are in better agreement with results from regional and local studies. This is especially true of low-velocity anomalies, which are particularly challenging for approximate techniques, but whose effects are accurately predicted by SEM. We stress that these features of our isotropic velocity model hold important implications for thermochemical interpretations based on mineral physics. SEMum can be obtained from http://seismo.berkeley.edu/∼lekic/SEMum.html.

ACKNOWLEDGMENTS

The authors would like to thank Yann Capdeville for providing us the cSEM code, and Scott French and Huaiyu Yuan for discussions and assistance. This work benefited from thorough and thoughtful reviews by Ana Ferreira and Qinya Liu. Support for VL was provided in part by a National Science Foundation Graduate Fellowship. We acknowledge support from NSF (grant EAR-0738284). This is Berkeley Seismological Laboratory contribution 11-01.

REFERENCES

Amante, C. & Eakins, B.W., 2008. Etopo1 1 arc-minute global relief model: procedures, data sources and analysis, National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce, Boulder, CO.

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

Artemieva, I., 2006. Global 1× 1 thermal model TC1 for the continental lithosphere: implications for lithosphere secular evolution, *Tectonophysics,* **416**(1-4), 245–277.

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

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

Becker, T. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst,* **3**(1), 1003, doi:10.1029/2001GC000168.

Becker, T., Kustowski, B. & Ekstrom, G., 2007. Radial seismic anisotropy as a constraint for upper mantle rheology, *Earth planet. Sci. Lett.,* **267,** 213–227.

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

Boschi, L., Ekstrom, G. & Kustowski, B., 2004. Multiple resolution surface wave tomography: the Mediterranean basin, *Geophys. J. Int.,* **157**(1), 293–304.

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

Cammarano, F. & Romanowicz, B., 2007. Insights into the nature of the transition zone from physically constrained inversion of long-period seismic data, *Proc. Natl. Acad. Sci.,* **104**(22), 9139–9144.

Cammarano, F., Deuss, A., Goes, S. & Giardini, D., 2005. One-dimensional physical reference models for the upper mantle and transition zone: combining seismic and mineral physics constraints, *J. geophys. Res.,* **110,** B01306, doi:10.1029/2004JB003272.

Cammarano, F., Romanowicz, B., Stixrude, L., Lithgow-Bertelloni, C. & Xu, W., 2009. Inferring the thermochemical structure of the upper mantle from seismic data, *Geophys. J. Int.,* **179**(2), 1169–1185.

Capdeville, Y. & Marigo, J., 2007. Second order homogenization of the elastic wave equation for non-periodic layered media, *Geophys. J. Int.,* **170**(2), 823–838.

Capdeville, Y., Chaljub, E., Vilotte, J.P. & Montagner, J.P., 2003. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models, *Geophys. J. Int.,* **152,** 34–67.

Chiao, L. & Kuo, B., 2001. Multiscale seismic tomography, *Geophys. J. Int.,* **145**(2), 517–527.

Christie, D.M., West, B.P., Pyle, D.G.&B., H.B., 1998. Chaotic topography, mantle flow and mantle migration in the Australian-Antarctic discordance, *Nature,* **394,** 637–644.

Dahlen, F. & Tromp, J., 1998. *Theoretical Global Seismology*, Princeton University Press, Princeton.

Dahlen, F., Hung, S. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes-I. Theory, *Geophys. J. Int.,* **141**(1), 157–174.

Dalton, C. & Ekstrom, G., 2006. Constraints on global maps of phase velocity from surface-wave amplitudes, *Geophys. J. Int.,* **167**(2), 820–826.

Dziewonski, A., 2005. The robust aspects of global seismic tomography, *Plates, Plumes, and Paradigms*, Geol. Soc. Am. Spec. Pub. 388, p. 147, eds Foulger, G.R., Natland, J.H., Presnall, D.C. & Anderson, D.L., Geol. Soc. America, Boulder, CO.

Dziewonski, A., Hager, B. & O'Connell, R., 1977. Large-scale heterogeneities in the lower mantle, *J. geophys. Res.,* **82,** 239–255.

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

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

Ferreira, A. & Woodhouse, J., 2006. Source, path and receiver effects on seismic surface waves, *Geophys. J. Int.,* **168**(1), 109–132.

Ferreira, A.M.G., Woodhouse, J.H., Visser, K. & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *Geophys. J. Int.,* **115,** doi:10.1029/2009JB006716.

Fichtner, A. & Igel, H., 2008. Efficient numerical surface wave propagation through the optimization of discrete crustal models-a technique based on non-linear dispersion curve matching (DCM), *Geophys. J. Int.,* **173**(2), 519–533.

Fichtner, A., Kennett, B., Igel, H. & Bunge, H., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179**(3), 1703–1725.

Fishwick, S., Kennett, B. & Reading, A., 2005. Contrasts in lithospheric structure within the Australian craton 1/2 insights from surface wave tomography, *Earth planet. Sci. Lett.,* **231**(3-4), 163–176.

Forsyth, D., Ehrenbard, R. & Chapin, S., 1987. Anomalous upper mantle beneath the Australian-Antarctic Discordance, *Earth planet. Sci. Lett.,* **84,** 471–478.

Friederich, W., 2003. The S-velocity structure of the East Asian mantle from inversion of shear and surface waveforms, *Geophys. J. Int.,* **153**(1), 88–102.

Godey, S., Deschamps, F., Trampert, J. & Snieder, R., 2004. Thermal and compositional anomalies beneath the North American continent, *J geophys. Res.,* **109,** doi:10.1029/2002JB002263.

Grand, S. & Helmberger, D., 1984. Upper mantle shear structure of North America, *Geophys. J. Int,* **76,** 399–438.

Gung, Y. & Romanowicz, B., 2004. Q tomography of the upper mantle using three-component long-period waveforms, *Geophys. J. Int.,* **157,** 813–830, doi: 10.1111/j.1365–246X.2004.02265.x.

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

Heintz, M., Debayle, E. & Vauchez, A., 2005. Upper mantle structure of the South American continent and neighbouring oceans from surface wave tomography, *Tectonophysics,* **405,** 115–139.

Hirschmann, M., 2006. Water, melting, and the deep Earth H_2O cycle, *Ann. Rev. Earth planet. Sci.,* **34,** 629–653.

Houser, C., Masters, G., Shearer, P. & Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.,* **174**(1), 195–212.

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

Katayama, I. & Karato, S., 2006. Effect of temperature on the B-to C-type olivine fabric transition and implication for flow pattern in subduction zones, *Phys. Earth planet. Inter.,* **157**(1-2), 33–45.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation – II. Three-dimensional models, oceans, rotation and self-gravitation, *Geophys. J. Int.,* **150,** 303–318.

Komatitsch, D. & Vilotte, J., 1998. The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Kustowski, B., Dziewonski, A. & Ekstrom, G., 2007. Nonlinear crustal corrections for normal-mode seismograms, *Bull. seism. Soc. Am.,* **97**(5), 1756–1762.

Kustowski, B., Ekstrom, G. & Dziewonski, A., 2008. The shear-wave velocity structure in the upper mantle beneath Eurasia, *Geophys. J. Int.,* **174**(3), 978–992.

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

Lebedev, S. & Nolet, G., 2003. Upper mantle beneath Southeast Asia from S velocity tomography, *J geophys. Res.-Solid Earth,* **108**(B1), 2048, doi:doi:10.1029/2000JB000073.

Lekic, V., 2004. Distribution of stresses in descending lithospheric slabs from a global survey of Harvard CMT solutions for subduction-zone earthquakes, *Honors thesis*, Harvard University.

Lekic, V. & Romanowicz, B., 2010. Tectonic regionalization without a priori information: a cluster analysis of tomography, *Earth planet. Sci. Lett.,* submitted.

Lekic, V., Panning, M. & Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography, *Geophys. J. Int.,* **182**(1), 265–278, doi:10.1111/j.1365–246X.2010.04602.x.

Li, X. & Tanimoto, T., 1993. Waveforms of long-period body waves in a slightly aspherical earth model., *Geophys. J. Int.,* **112,** 92–102.

Li, X.-D. & Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.,* **121,** 695–709.

Li, X.-D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.,* **101,** 22 245–22 272.

Marone, F. & Romanowicz, B., 2007. Non-linear crustal corrections in high-resolution regional waveform seismic tomography, *Geophys. J. Int.,* **170,** 460–467.

Marone, F., Gung, Y. & Romanowicz, B., 2007. Three-dimensional radial anisotropic structure of the North American upper mantle from inversion of surface waveform data, *Geophys. J. Int.,* **171**(1), 206–222.

McEvilly, T., 1964. Central US crust-upper mantle structure from Love and Rayleigh wave phase velocity inversion, *Bull. seism. Soc. Am,* **54,** 1997–2015.

Mégnin, C. & Romanowicz, B., 2000. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms, *Geophys. J. Int.,* **143,** 709–728.

Meier, U., Curtis, A. & Trampert, J., 2007. Fully nonlinear inversion of fundamental mode surface waves for a global crustal model, *Geophys. Res. Lett.,* **34,** L16304, doi:10.1029/2007GL030989.

Montagner, J.-P. & Jobert, N., 1988. Vectorial tomography. II: application to the indian ocean, *Geophys. J.,* **94,** 309–344.

Montagner, J.-P. & Kennett, B., 1996. How to reconcile body-wave and normal-mode reference earth models, *Geophys. J. Int.,* **125,** 229–248.

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

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

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

Panning, M. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.,* **167,** 361–379.

Panning, M., Capdeville, Y. & Romanowicz, B., 2009. Seismic waveform modelling in a 3-D Earth using the Born approximation: potential shortcomings and a remedy, *Geophys. J. Int.,* **177**(1), 161–178.

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

Pasyanos, M. & Nyblade, A., 2007. A top to bottom lithospheric study of Africa and Arabia, *Tectonophysics,* **444**(1-4), 27–44.

Pasyanos, M.E., 2005. A variable resolution surface wave dispersion study of Eurasia, North Africa, and surrounding regions, *J. geophys. Res.,* **110,** B12301, doi:10.1029/2005JB003749.

Plomerova, J., Kouba, D. & Babuska, V., 2002. Mapping the lithosphere-asthenosphere boundary through changes in surface-wave anisotropy, *Tectonophysics,* **358,** 175–185.

Priestley, K., Debayle, E., McKenzie, D. & Pilidou, S., 2006. Upper mantle structure of eastern Asia from multimode surface waveform tomography, *J. geophys. Res.,* **111,** B10304, doi:10.1029/2005JB004082.

Priestley, K., McKenzie, D., Debayle, E. & Pilidou, S., 2008. The African upper mantle and its relationship to tectonics and surface geology, *Geophys. J. Int.,* **175**(3), 1108–1126.

Ribe, N., 1989. Seismic anisotropy and mantle flow, *J. geophys. Res,* **94,** 4213–4223.

Ribe, N., 1992. On the relation between seismic anisotropy and finite strain, *J. geophys. Res.,* **97**(B6), 8737–8747.

Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 2004. Global transition zone tomography, *J. geophys. Res.,* **109**(B02302), doi:10.1029/2003JB002610.

Ritzwoller, M., Shapiro, N. & Leahy, G., 2003. A resolved mantle anomaly as the cause of the Australian-Antarctic Discordance, *J. geophys. Res,* **108**(B12), 2559, doi:10.1029/2003JB002522.

Rodi, W., Glover, P., Li, T. & Alexander, S., 1975. A fast, accurate method for computing group-velocity partial derivatives for Rayleigh and Love modes, *Bull. seism. Soc. Am.,* **65**(5), 1105–1114.

Romanowicz, B., 1987. Multiplet-multiplet coupling due to lateral heterogeneity: asymptotic effects on the amplitude and frequency of the Earth's normal modes, *Geophy. J. Int.,* **90,** 75–100.

Romanowicz, B., Panning, M., Gung, Y. & Capdeville, Y., 2008. On the computation of long period seismograms in a 3-D Earth using normal mode based approximations, *Geophys. J. Int.,* **175**(2), 520–536.

Romanowicz, B.A., 1979. Seismic structure of the upper mantle beneath the United States by three-dimensional inversion of body wave arrival times, *Geophys. J. R. astr. Soc.,* **57**(2), 479–506.

Rudnick, R., McDonough, W. & O'Connell, R., 1998. Thermal structure, thickness and composition of continental lithosphere, *Chem. Geol.,* **145,** 395–411.

Schimmel, M., Assumpçao, M. & VanDecar, J., 2003. Seismic velocity anomalies beneath SE Brazil from P and S wave travel time inversions, *J. geophys. Res,* **108,** 2191, doi:10.1029/2001JB000187.

Sebai, A., Stutzmann, E., Montagner, J., Sicilia, D. & Beucler, E., 2006. Anisotropic structure of the African upper mantle from Rayleigh and Love wave tomography, *Phys. Earth planet. Int.,* **155**(1-2), 48–62.

Shapiro, N. & Ritzwoller, M., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.,* **151**(1), 88–105.

Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.,* **1**(7), 458–462.

Simmons, N., Forte, A. & Grand, S., 2006. Constraining mantle flow with seismic and geodynamic data: a joint approach, *Earth planet. Sci. Lett.,* **246**(1-2), 109–124.

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

Spetzler, J. & Trampert, J., 2003. Implementing spectral leakage corrections in global surface wave tomography, *Geophys. J. Int.,* **155**(2), 532–538.

Spetzler, J., Trampert, J. & Snieder, R., 2002. The effect of scattering in surface wave tomography, *Geophys. J. Int.,* **149**(3), 755–767.

Steinberger, B., 2000. Plumes in a convecting mantle: models and observations for individual hotspots, *J. geophys. Res.,* **105**(B5), 11 127–11 152.

Su, W. & Dziewonski, A., 1991. Predominance of long-wavelength heterogeneity in the mantle, *Nature,* **352,** 121–126.

Su, W., Woodward, R. & Dziewonski, A., 1992. Deep origin of mid-ocean-ridge seismic velocity anomalies, *Nature,* **360,** 149–152.
Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science,* **325**(5943), 988.
Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.
Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial Mathematics, Philadelphia, PA.
Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. geophys. Space Phys.,* **20,** 219–232.
Trampert, J. & Snieder, R., 1996. Model estimations biased by truncated expansions: possible artifacts in seismic tomography, *Science,* **271**(5253), 1257–1260.
Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160**(1), 195–216.
Wang, Z. & Dahlen, F., 1995. Validity of surface-wave ray theory on a laterally heterogeneous Earth, *Geophys. J. Int.,* **123**(3), 757–773.
Wang, Z. & Dahlen, F.A., 1995. Spherical-spline parametrization of three-dimensional earth models, *Geophys. Res. Lett.,* **22,** 3099–3102.
Weeraratne, D.S., Forsyth, D.W., Fischer, K.M. & Nyblade, A.A., 2003. Evidence for an upper mantle plume beneath the Tanzanian craton from Rayleigh wave tomography, *J. geophys. Res.,* **108,** 2427, doi:10.1029/2002JB002273.
Woodhouse, J.H., 1998. The calculation of eigenfrequencies and eigenfunctions of the free oscillations of the Earth and the Sun, in *Seismological Algorithms*, pp. 321–370, ed. Doornbos, D.J., Elsevier, New York.
Woodhouse, J.H. & Dahlen, F.A., 1978. The effect of a general aspherical perturbation on the free oscillations of the Earth, *Geophys. J. R. astr. Soc.,* **53,** 335–354.
Woodhouse, J.H. & Dziewonski, A.M., 1984. Mapping the upper mantle: three dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.,* **89,** 5953–5986.
Woodhouse, J.H. & Girnius, T.P., 1982. Surface waves and free oscillations in a regionalized Earth model, *Geophys. J. R. astr. Soc.,* **68,** 653–673.
Xu, W., Lithgow-Bertelloni, C., Stixrude, L. & Ritsema, J., 2008. The effect of bulk composition and temperature on mantle seismic structure, *Earth planet. Sci. Lett.,* **275**(1-2), 70–79.
Yuan, H. & Romanowicz, B., 2010. Lithospheric layering in the North American Craton, *Nature,* **466,** 1063–1068.
Zhang, Y. & Tanimoto, T., 1992. Ridges, hotspots and their interaction as observed in seismic velocity maps, *Nature,* **355,** 45–49.
Zhou, Y., Nolet, G., Dahlen, F. & Laske, G., 2006. Global upper-mantle structure from finite-frequency surface-wave tomography, *J. geophys. Res.-Solid Earth,* **111**(B04304), doi:10.1029/2005JB003677.

APPENDIX A

In this appendix, we compare the relative importance of inaccuracies in computing the misfit function versus the partial derivatives in the non-linear inverse problem, such as that of waveform tomography. We show that the former is of first order, while the latter is of second order importance. We begin with eq. (25) of Tarantola & Valette (1982):

$$\delta m_k = \left(G_k^T C_D^{-1} G_k + C_M^{-1}\right)^{-1} \times \left\{G_k^T C_D^{-1}[d - g(m_k)] - C_M^{-1}(m_k - m_0)\right\}, \quad \text{(A1)}$$

where δm_k is the model update at the kth iteration, C_D is the data covariance matrix, C_M is the *a priori* model covariance matrix, G_k is the matrix of partial derivatives relating model and data perturbations, d is the data, and $g(m_k)$ are the synthetic data predicted by the current model m_k. We introduce ϵ, a perturbation to G_k, which represents error in the calculation of partial derivatives of data with respect to the model parameters, and a perturbation η to $g(m_k)$, which represents the error in calculating the forward problem. Eq. (A1) then becomes

$$\delta m_k = \left[(G_k + \epsilon)^T C_D^{-1}(G_k + \epsilon) + C_M^{-1}\right]^{-1} \times \left\{(G_k + \epsilon)^T C_D^{-1}[d - g(m_k) + \eta] - C_M^{-1}(m_k - m_0)\right\}. \quad \text{(A2)}$$

Rewriting by introducing matrices $A = G_k^T C_D^{-1} G_k + C_M^{-1}$ and $B = G_k^T C_D^{-1} + C_D^{-1} G_k$, and keeping only the terms that are first order

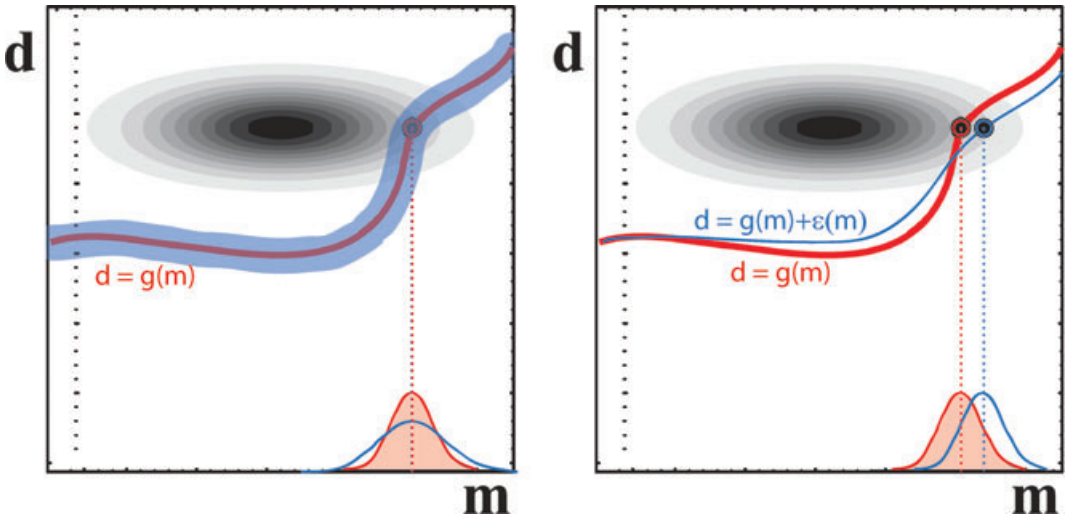

Figure A1. Cartoon depicting a non-linear problem that can be solved using an iterative scheme. The grey contours represent the information we have on the data and *a priori* information on the model parameters. The red curve denotes the true non-linear relationship between data and model. Left-hand panel: if the error introduced by the forward modelling scheme is normally distributed about the true value, then this modelling error increases the *a posteriori* variance of the retrieved model about the maximum likelihood model (indicated by the blue shading). Right-hand panel: if the forward modelling error introduces bias ϵ (m) and defines a different non-linear relationship between model and data (blue curve), then the retrieved model can be displaced from the true maximum likelihood model (dashed blue line).

in the small quantities ϵ and η, we have

$$\delta m_k = (I - \epsilon A^{-1}B)A^{-1}\left\{G_k^T C_D^{-1}[d - g(m_k) + \eta] - C_M^{-1}(m_k - m_0) + \epsilon C_D^{-1}[d - g(m_k)]\right\}. \quad \text{(A3)}$$

Finally,

$$\begin{aligned}\delta m_k &= A^{-1}G_k^T C_D^{-1}\eta + A^{-1}\left\{G_k^T C_D^{-1}[d - g(m_k)] - C_M^{-1}(m_k - m_0)\right\} \\ &\quad + A^{-1}\epsilon C_D^{-1}[d - g(m_k)] - \epsilon A^{-1}BA^{-1}\left\{G_k^T C_D^{-1}[d - g(m_k)] \right. \\ &\quad \left. - C_M^{-1}(m_k - m_0)\right\}\end{aligned} \quad \text{(A4)}$$

from which we see that whereas η has a first-order effect on the model update, ϵ terms always appear multiplied by small quantities $[d - g(m_k)]$ or C_M^{-1} $(m_k - m_0)$, and have an accordingly smaller effect on the retrieved model.

Fig. A1 is a cartoon illustrating the effect that forward modelling errors can have on the retrieved model and its *a posteriori* covariance. The illustration is adapted from fig. 3.2 of Tarantola (2005), and depicts a non-linear problem that cannot be linearized from the starting model, but that can be solved using an iterative scheme. The grey contours represent the information we have on the data and *a priori* information on the model parameters. The red curve denotes the true non-linear relationship between data and model. If the error introduced by the forward modelling scheme is normally distributed about the true value, then this modelling error increases the *a posteriori* variance of the retrieved model about the maximum likelihood model, in he same way as if the data was noisier (left-hand panel). However, if the forward modelling error introduces bias and defines a different non-linear relationship between model and data (blue curve), then the retrieved model can be displaced from the true maximum likelihood model (indicated by the dashed red and blue lines in the right-hand panel).