

# Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography

S. W. French<sup>1,\*</sup> and B. A. Romanowicz<sup>1,2,3</sup>

<sup>1</sup>Berkeley Seismological Laboratory, McCone Hall, University of California, Berkeley, CA 94720, USA. E-mail: [sfrench@seismo.berkeley.edu](mailto:sfrench@seismo.berkeley.edu)

<sup>2</sup>Institut de Physique du Globe de Paris, 1 Rue Jussieu, F-752382 Paris Cedex 05, France

<sup>3</sup>Collège de France, 11 Place Marcelin Berthelot, F-75005 Paris, France

Accepted 2014 August 29. Received 2014 August 19; in original form 2014 June 16

## SUMMARY

The radially anisotropic shear velocity structure of the Earth's mantle provides a critical window on the interior dynamics of the planet, with isotropic variations that are interpreted in terms of thermal and compositional heterogeneity and anisotropy in terms of flow. While significant progress has been made in the more than 30 yr since the advent of global seismic tomography, many open questions remain regarding the dual roles of temperature and composition in shaping mantle convection, as well as interactions between different dominant scales of convective phenomena. We believe that advanced seismic imaging techniques, such as waveform inversion using accurate numerical simulations of the seismic wavefield, represent a clear path forwards towards addressing these open questions through application to whole-mantle imaging. To this end, we employ a 'hybrid' waveform-inversion approach, which combines the accuracy and generality of the spectral finite element method (SEM) for *forward* modelling of the global wavefield, with non-linear asymptotic coupling theory for efficient *inverse* modelling. The resulting whole-mantle model (SEMUCB-WM1) builds on the earlier successful application of these techniques for global modelling at upper mantle and transition-zone depths ( $\leq 800$  km) which delivered the models SEMum and SEMum2. Indeed, SEMUCB-WM1 is the first whole-mantle model derived from fully numerical SEM-based forward modelling. Here, we detail the technical aspects of the development of our whole-mantle model, as well as provide a broad discussion of isotropic and radially anisotropic model structure. We also include an extensive discussion of model uncertainties, specifically focused on assessing our results at transition-zone and lower-mantle depths.

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

## 1 INTRODUCTION

Global seismic tomography has made considerable progress over the past 30 years in identifying robust, large-scale features in the seismic velocity structure of the Earth's mantle, including the large low shear velocity provinces (LLSVPs) in the deep mantle (e.g. Dziewonski *et al.* 1977; Lekić *et al.* 2012) and high-velocity anomalies associated with subducted slabs (e.g. van der Hilst *et al.* 1997), although the depth distribution of the latter is still debated (e.g. Fukao & Obayashi 2013). The earliest 3-D global models, such

as those of Dziewonski *et al.* (1977) and Woodhouse & Dziewonski (1984), focusing on the lower and upper mantle, respectively, were the first to obtain images of long-wavelength heterogeneity ( $> 5000$  km). Since then, several generations of models have been developed with steadily improving resolution. More recent global studies, including both whole-mantle (e.g. Panning & Romanowicz 2006; Simmons *et al.* 2006, 2010; Houser *et al.* 2008; Kustowski *et al.* 2008; Ritsema *et al.* 2011) and upper-mantle models (e.g. Lekić & Romanowicz 2011a; Debayle & Ricard 2012; Schaeffer & Lebedev 2013), with some claiming lateral resolution on the order of 1000 km, have extended the interpretable spectrum of model structure while largely confirming the long-wavelength features seen previously (e.g. Lekić *et al.* 2012). At the same time, shorter wavelength features ( $< 2500$  km) do not tend to correlate well across global models (Becker & Boschi 2002; Dziewonski 2005), as

\*Now at: The National Energy Research Scientific Computing Center (NERSC), Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

often noted in the context of intermodel comparisons (e.g. Kus-towski *et al.* 2008; Ritsema *et al.* 2011). These discrepancies also extend to anisotropic structure (e.g. Becker *et al.* 2008), which shows considerable variation across models at even long wave-lengths. Taken together, these observations imply significant uncer-tainty across published models, even at purportedly resolved scales.

Despite these uncertainties, seismic tomography is uniquely suited to help address open questions regarding the nature of mantle convection, including: the anticipated scales and depth extents of convective phenomena (e.g. Phipps-Morgan *et al.* 1995; Korenaga & Jordan 2004), the roles of the LLSVPs in shaping mantle con-vection patterns (e.g. Richards & Engebretson 1992; Davaille *et al.* 2005; Dziewonski *et al.* 2010), as well as their implications for com-positional heterogeneity (e.g. Davaille 1999; McNamara & Zhong 2005), and controls on the surface distribution of hotspots due to large-scale convection patterns (e.g. Husson & Conrad 2012). In conjunction with geodynamic studies implementing more realistic mantle convection simulations or laboratory analogue experiments (e.g. incorporating more complex rheological effects or stronger viscosity variations), higher resolution global-scale tomographic models, producing more detailed and realistic images of seismic structure while reducing and quantifying uncertainties, provide a key avenue for constraining which phenomena seen in simulation or the lab are likely to manifest in the mantle.

These goals should be realizable using advanced seismic imag-ing techniques, such as waveform inversion based on accurate num-erical wavefield simulations. The ‘hybrid’ waveform inversion approach, which uses the spectral finite element method (SEM; e.g. Komatitsch & Vilotte 1998) for wavefield modelling in con-junction with non-linear asymptotic coupling theory (NACT; Li & Romanowicz 1995) for efficient derivative estimation, is one such technique that has previously been applied to upper-mantle and transition-zone imaging—leading to the SEMum (Lekić & Romanowicz 2011a) and later SEMum2 (French *et al.* 2013) global models. Indeed, both of these models exhibit larger amplitudes of shear velocity heterogeneity than seen in previous models, as well as novel features not seen in earlier generations of global tomography based on approximate modelling techniques (Lekić & Romanowicz 2011b; French *et al.* 2013). Motivated by these successful applica-tions of the hybrid inversion approach for imaging structure above ~800-km depth, as well as a desire to apply these techniques to shed light on many of the open questions above, we here present the first whole-mantle tomographic model derived using SEM-based waveform tomography: SEMUCB-WM1.

In what follows, we present this new whole-mantle model, dis-cussing both its derivation and the structure of the model itself. First, in Section 2, we revisit the hybrid waveform tomographic technique used previously in developing the SEMum and SEMum2 models, focusing on (i) how this technique differs from other global modelling approaches and (ii) our work to extend this technique to whole-mantle imaging. Next, in Section 3, we discuss the wave-form data set used in our inversion, which now includes body wave seismograms down to 32-s period. In Section 4, we first lay out the specific steps involved in our inversion, with an emphasis on progressive incorporation of more data to shorter period, and then present SEMUCB-WM1 along with a discussion of model structure emphasizing general properties (with detailed interpretations of certain aspects of model structure appearing in a future publica-tion). Next, in Section 5, we evaluate the model, focused primarily on model performance and estimation of uncertainties. Finally, we conclude in Section 6, summarizing the specific contributions of this work and avenues for future study.

## 2 METHODOLOGY

### 2.1 Hybrid full-waveform inversion

We use the hybrid full-waveform inversion technique previously employed by our group in developing the SEMum (Lekić & Romanowicz 2011a) and SEMum2 (French *et al.* 2013) global models. This technique combines the accuracy and generality of the SEM (e.g. Komatitsch & Vilotte 1998) coupled to normal-mode computations in the core (Capdeville *et al.* 2003) for wavefield forward modelling, with efficient sensitivity kernel calculation using NACT (Li & Romanowicz 1995). We combine these two ap-proaches in the context of the generalized least-squares formalism of Tarantola & Valette (1982), in which we seek to iteratively minimize the non-linear least-squares misfit functional:

$$2\Phi(\mathbf{m}_k) = \|\mathbf{d} - \mathbf{g}(\mathbf{m}_k)\|_{\mathbf{C}_d^{-1}}^2 + \|\mathbf{m}_p - \mathbf{m}_k\|_{\mathbf{C}_m^{-1}}^2, \quad (1)$$

where  $\mathbf{m}_k$  represents the  $k$ th iterative model estimate,  $\mathbf{d}$  and  $\mathbf{g}(\mathbf{m}_k)$  are the observations and model predictions (the latter for waveform data are obtained using the SEM),  $\mathbf{m}_p$  is the model prior (generally taken to be zero perturbation) and  $\mathbf{C}_m$  and  $\mathbf{C}_d$  represent *a priori* model and data covariance operators. We discuss the state of prior information in greater detail in Section 2.3.

Using the SEM comes with numerous advantages, particularly in global-scale modelling (Komatitsch & Tromp 2002a,b; Chaljub *et al.* 2003), for instance: (1) accurate simulation of wave propaga-tion in complex anisotropic and anelastic media, including treatment of phenomena often ignored by approximate modelling techniques (such as wave front healing, e.g. Nolet & Dahlen 2000); (2) excellent numerical dispersion properties (Seriani & Oliveira 2008), critical for modelling surface waves at long propagation times (i.e. second-orbit fundamental modes) and (3) natural treatment of the free-surface boundary condition. Furthermore, unlike finite-difference methods, the SEM is able to mesh complex geometries character-ized by discontinuous material properties, such as the Earth’s crust (Komatitsch & Vilotte 1998).

NACT, based on normal-mode perturbation theory, yields finite-frequency waveform sensitivity kernels in the source–receiver great-circle plane, at little cost relative to SEM simulation (Li & Romanowicz 1995). This method includes coupling both along and across mode dispersion branches, necessary for accurate modelling of the sensitivity of body waves and overtone surface waves (Li & Romanowicz 1996; Mégnin & Romanowicz 1999). Unlike most approximate techniques for sensitivity kernel calculation, NACT is *non-linear*, evolving as the inversion progresses and the earth model iteratively evolves and includes the effect of multiple forward scat-tering (Romanowicz *et al.* 2008). For a more detailed discussion of normal-mode perturbation theory in waveform inversion, see Romanowicz *et al.* (2008).

One of the primary benefits of the hybrid approach over fully SEM-based techniques using first-order adjoint-state methods (e.g. Tarantola 1984; Tromp *et al.* 2005) is convergence rate. Direct access to the waveform Jacobian under NACT allows a rapidly con-verging Gauss–Newton optimization scheme to be used—namely, that arising naturally from the generalized least-squares formalism:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + (\mathbf{C}_m \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \mathbf{I})^{-1} \times (\mathbf{C}_m \mathbf{G}^T \mathbf{C}_d^{-1} [\mathbf{d} - \mathbf{g}(\mathbf{m}_k)] + \mathbf{m}_p - \mathbf{m}_k), \quad (2)$$

where we note that the model Jacobian matrix  $\mathbf{G}$  has been evaluated *at* the current model estimate  $\mathbf{m}_k$  (i.e. at each step, we linearize around  $\mathbf{m}_k$  in order to estimate  $\mathbf{m}_{k+1}$ ). This stands in contrast to the

gradient-based approaches required when using first-order adjoint. By converging on a model more quickly, thus reducing the number of inversion iterations, we are able to perform significantly fewer SEM simulations (which dominate the cost of the inversion). At the same time, many of the well-known limitations associated with waveform inversion—a strongly non-linear problem—still remain true of our approach. For example, care must be taken in selection of both the starting model as well as the data set, along with the manner in which the data set is expanded as the inversion progresses (typically by incrementally including shorter periods). Indeed, these factors have a significant influence on both the convergence rate of the inversion and the veracity of the resulting ‘optimal’ model. Furthermore, our Gauss–Newton scheme has computational drawbacks of its own, both in distributed construction of the approximate Hessian  $\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}$  via NACT and subsequent factorization. However, we have found that the overall savings due to enhanced convergence outweighs the complexity of addressing these latter issues.

We note that Gauss–Newton, and indeed full-Newton, schemes that are based purely on adjoint-state methods have been successfully implemented and applied to waveform inversion in exploration settings (e.g. Pratt *et al.* 1998). These approaches stand in contrast to the first-order adjoint techniques often employed in broad-band passive-source seismology at regional (and larger) scales, though considerable progress has been made in recent years towards incorporating Hessian information into the latter (e.g. Fichtner & Trampert 2011a). Quasi-Newton schemes such as the limited-memory Broyden–Fletcher–Goldfarb–Shanno algorithm (L-BFGS; Nocedal 1980), which build up an estimate of the Hessian based on the history of gradient vectors from previous inversion iterations (potentially obtained via adjoint computations), have also been used in exploration seismology (e.g. Brossier *et al.* 2009). However, for our particular application, where it is critically important to minimize the total number of expensive global wavefield simulations, it is not clear whether the resulting improvement in convergence would be sufficient to justify such a technique over our hybrid approach—particularly as it may take a number of gradient calculations, and thus many wavefield simulations, before the quasi-Newton approximate Hessian becomes sufficiently accurate.

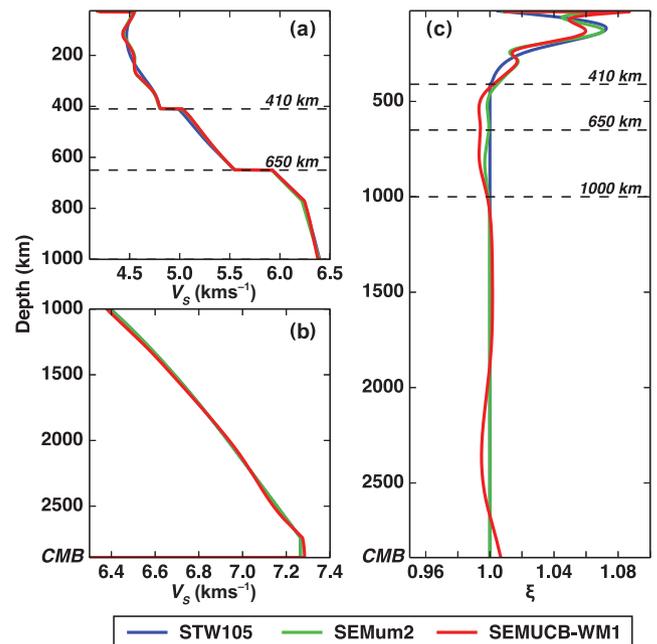
While waveform partial derivatives computed with NACT are approximate in nature, we note that the misfit functional (eq. 1) is at all times evaluated ‘exactly’ with the SEM. Indeed, it is clearly seen from eq. (2) that errors in the computation of  $\mathbf{g}(\mathbf{m}_k)$  will appear as first-order terms in the model update (i.e. propagate into  $\mathbf{m}_{k+1}$ ), while those in  $\mathbf{G}$  are second order (e.g. appendix A of Lekić & Romanowicz 2011a). Thus, presuming that we have kept the inverse problem well-posed, such as by progressively including shorter period waveform data and carefully selecting the starting model (which are also constraints imposed upon adjoint inversions), convergence of eq. (1) suggests that we are moving towards an optimal model as ‘viewed’ by the SEM. Furthermore, we note that our particular formulation, where derivative information is estimated on a trace-by-trace basis, allows us to take an informed approach to correcting for non-uniform data coverage and sensitivity (a condition that simply cannot be avoided in present-day global inversions). This stands in contrast to the *ad hoc* pre-conditioning and smoothing typically applied to gradients estimated using adjoint-state schemes (e.g. Fichtner *et al.* 2009).

Finally, we would like to emphasize that while the hybrid approach is an effective technique for our current application—namely, global-scale imaging at relatively long periods—it may not remain an appropriate choice for more advanced future inversions. For example, at significantly shorter periods, we expect that the particular formulation of our sensitivity kernels, based on

normal-mode perturbation theory, may no longer be sufficiently accurate and may grow significantly more expensive—the latter due to the  $O(\omega^4)$  dependence of the number of mode pairs to be coupled. In addition, non-linearity of the inverse problem may be sufficiently strong at shorter periods that the Gauss–Newton approximate Hessian no longer significantly aids convergence. Furthermore, high dimensionality of the earth model—required for commensurately higher resolution—may render storage and factorization of the approximate Hessian impractical, and encourage the adoption of matrix-free quasi-Newton (L-BFGS) or Newton–Krylov schemes (e.g. Epanomeritakis *et al.* 2008).

## 2.2 Model parametrization

SEMUCB-WM1 adopts the same parametrization as was used in SEMum2, which also served as our starting model above  $\sim 800$ -km depth (see Section 4.1). We invert for 3-D variations in Voigt-average isotropic shear wave velocity  $V_S$  and the radially anisotropic parameter  $\xi = V_{SH}^2 / V_{SV}^2$  with respect to a 1-D reference model that continuously evolves throughout the inversion; namely, by removing the degree-0 component from the 3-D structure at each iteration. While velocity structure in the 1-D reference model is allowed to change, we maintain a fixed 1-D attenuation model based on a smoothed version of the model QL6 of Durek & Ekström (1996). In Fig. 1, we show  $V_S$  and  $\xi$  structure from both the initial (SEMum2) and final-iteration SEMUCB-WM1 reference models, as well as the model STW105 of Kustowski *et al.* (2008) for comparison. We note that the initial and final-iteration models remain extremely similar with respect to upper mantle  $V_S$ —in fact, the inverted 3-D  $V_S$  models also remain very similar in this depth range (a point we return to in Section 4.2). In addition, all three 1-D  $V_S$  models remain quite similar throughout the lower mantle, which is not surprising given the amplitudes and spectral content of heterogeneity expected in this depth range. Conversely, the 1-D  $\xi$  profile associated with SEMUCB-WM1 does change notably relative to SEMum2 below



**Figure 1.** Voigt-average isotropic shear wave velocity  $V_S$  and the radially anisotropic parameter  $\xi = V_{SH}^2 / V_{SV}^2$  in the 1-D reference models associated with SEMum2 and SEMUCB-WM1, as well as STW105, the reference model for S362ANI of Kustowski *et al.* (2008), for comparison.

~400-km depth. Namely, the former exhibits weak  $V_{SV} > V_{SH}$  in both the transition zone and a ~500-km layer above the  $D''$  region, while also exhibiting  $V_{SH} > V_{SV}$  within  $D''$ . We will return to this observation in our discussion of anisotropic whole-mantle structure in Section 4.2.2.

From the perturbations to  $V_S$  and  $\xi$ , we use the empirical scaling relationships of Montagner & Anderson (1989) in order to estimate variations in the remaining four parameters of a radially anisotropic medium ( $V_P$ ,  $\phi = V_{PV}^2/V_{PH}^2$ ,  $\eta$  and  $\rho$ ), as in previous whole-mantle studies from our group (Panning & Romanowicz 2006). Such scalings are required, as our data set lacks sufficient sensitivity to invert for these parameters independently—even in the case of  $V_P$ , despite our addition of  $P$ - $SV$  waveform data down to 32 s (see Section 3). That said, these scalings are only technically valid in the upper mantle, and it is widely recognized that scaling relations for the deep mantle may differ significantly, such as the  $V_S$  to  $V_P$  perturbation scaling changing from ~2 in the upper and mid-mantle to ~3 or more in the lowermost mantle (e.g. Robertson & Woodhouse 1996; Romanowicz 2001). We return to this point in Section 5.2.1, where we discuss uncertainties in scaling factors and conclude that potential bias in the resulting mantle model is likely not significant (based on inversion experiments with plausible depth-variable scalings).

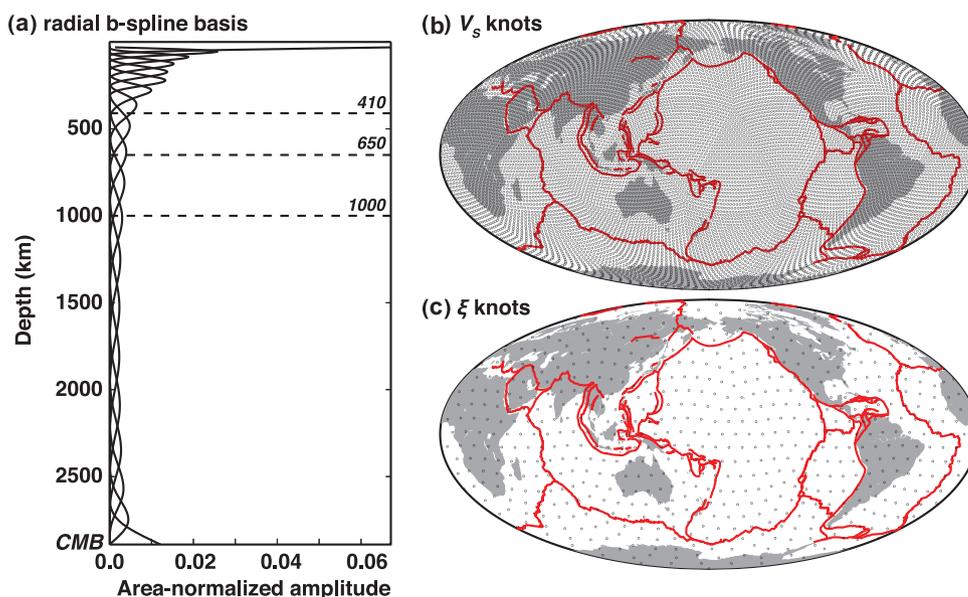
We express perturbations to mantle  $V_S$  and  $\xi$  in cubic b-splines radially (e.g. Mégnin & Romanowicz 2000) and in spherical splines laterally (Wang & Dahlen 1995). This particular combination of basis functions has been used widely in Berkeley global and regional-scale tomographic models since the work of Panning & Romanowicz (2006). The knots associated with our spherical-spline basis exhibit average lateral spacings of less than  $2^\circ$  for  $V_S$  and  $8^\circ$  for  $\xi$ , while our radial b-spline basis uses 20 knots with variable spacing between the core–mantle boundary (CMB) and 30-km depth (the shallowest Moho depth associated with our smooth crustal model; see Section 2.4). The latter are clustered more tightly near the top and bottom of the mantle model, where better radial resolution is expected (Fig. 2). Together, our radially anisotropic whole-mantle

model is comprised of a set of  $2.2 \times 10^5$  discrete spline coefficients, denoted  $\mathbf{m}$  in eqs (1) and (2).

One of the issues we encountered in developing our whole-mantle inversion was the size of the NACT-based Hessian matrix, which has the same (square) dimension as  $\mathbf{m}$  and thus requires nearly 100 GB at single precision to store just the upper or lower triangular part (as the Hessian is symmetric). Linear algebra operations requiring the Hessian to be stored in full form double this storage requirement. While neither I/O (easily tuned to take advantage of collective buffering on high-performance parallel filesystems) nor factorization of the Hessian (e.g. using ScaLAPACK; Blackford *et al.* 1997), proved problematic at this new problem size, assembly of the Hessian did. In particular, as the full Hessian can no longer fit in memory on a typical shared-memory compute node, it must now be distributed among some number of nodes, while at the same time supporting performant and safe (i.e. atomic) updates from large numbers of parallel NACT computations. To this end, we designed a high-performance distributed-memory solution based on the partitioned global address space (PGAS) model of parallel computations—namely, using the functionality exposed by UPC++, a set of PGAS extensions to C++ (Zheng *et al.* 2014). We found that this particular approach enables considerably improved scaling behaviour over analogous implementations based on message passing (MPI) and allows us to tackle inversions characterized by problem sizes well beyond our previous capabilities.

### 2.3 Prior information

Furthermore, we include prior information in order to keep the inverse problem well-posed—namely, through the data and model covariance operators  $\mathbf{C}_d$  and  $\mathbf{C}_m$  appearing in the generalized least-squares misfit functional (eq. 1). The data covariance matrix  $\mathbf{C}_d$  is diagonal, with entries reflecting data quality and uniqueness; see appendix A of Li & Romanowicz (1996) for a detailed discussion of its particular form. The model covariance operator  $\mathbf{C}_m$  is used to limit the permissible length-scales of heterogeneity in our



**Figure 2.** Parametrization of our whole-mantle model, SEMUCB-WM1. (Panel a): 20 radial b-spline basis functions (e.g. Mégnin & Romanowicz 2000) distributed between the CMB and the shallowest Moho in our smooth crustal model (30 km) according to expected resolution. (Panels b and c): the distribution of knots supporting our spherical spline (Wang & Dahlen 1995) basis for  $V_S$  and  $\xi$  (with average spacing of  $<2^\circ$  and  $<8^\circ$ , corresponding approximately to spherical-harmonic degrees 96 and 24, respectively).

model to those that should be well resolved by the data. Namely,  $\mathbf{C}_m$  is based upon prescribed model-parameter variances and correlation lengths, with the latter spatially tuned according to data-coverage density and quality. As in Lekić & Romanowicz (2011a) and later in French *et al.* (2013), the element of  $\mathbf{C}_m$  jointly describing the state of *a priori* information on model coefficients  $i$  and  $j$  is given by

$$(\mathbf{C}_m)_{ij} = \sigma_{ij} \exp \left[ \frac{\cos \Delta_{ij} - 1}{h_{ij}^2} \right] \exp \left[ -\frac{2r_{ij}}{v_{ij}^2} \right], \quad (3)$$

where  $\Delta_{ij}$  is the minor-arc distance separating  $i$  and  $j$ ,  $r_{ij}$  their radial separation,  $v_{ij}$  and  $h_{ij}$  the averages of their associated radial and lateral correlation lengths ( $h_{ij}$  is normalized to the range  $0 \leq |\cos \Delta_{ij} - 1| \leq 1$ ) and  $\sigma_{ij}$  is a constant near-unit variance pre-factor assigned to each of  $V_S$  and  $\xi$  (which are not assumed *a priori* to covary). Note that, because of the locality properties of the spline basis functions, we have made the approximation that model coefficients may be treated directly in the above formulation, not model values.

We prescribe radial correlation lengths with a fixed depth dependence, consistent with the expected resolving power of our data set: ranging from 50 km at the shallowest upper-mantle depths up to 300 km in the mid-mantle. As in SEMum and SEMum2, we adopt a scheme for adaptive lateral correlation-length selection, based on inferred data sensitivity (measured from the diagonal of the NACT-based Hessian). In particular, we take the diagonal entries of the Gauss–Newton Hessian  $\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}$  associated with the data misfit term in eq. (1) to represent aggregate sensitivity of the data to each model coefficient. Here, contributions from different data are weighted according to the quality and uncertainty estimates appearing in  $\mathbf{C}_d$ . These values are, in turn, used to select ‘local’ correlation-length estimates by scaling to the interval between prescribed minimum and maximum values for a given model parameter ( $V_S$ ,  $\xi$ ). By analogy with our earlier upper-mantle inversion for SEMum2 ( $\leq 800$ -km depth), which limited these ranges of correlation lengths to 400–800 and 1200–2400 km for  $V_S$  and  $\xi$ , respectively, we selected 400–1200 and 1200–3600 km for the whole-mantle inversion. Based on the sensitivity profile associated with our data, this configuration leads to a similar distribution of correlation lengths as before in the SEMum2 depth range, while permitting relatively long correlation in the mid-mantle where aggregate sensitivity is poorest, and again allowing shorter wavelength structure at the base of the mantle where sensitivity improves. Given the form of eq. (3), we expect significant correlation between neighbouring model coefficients at up to twice these distances.

## 2.4 Crustal model

To reduce the cost of our SEM simulations, in addition to our efforts to reduce the total number thereof as discussed in Section 2.1, we use a smooth anisotropic crustal layer in lieu of a more geologically plausible thinly layered model (e.g. Crust2.0; Bassin *et al.* 2000). This approach enables us to relax constraints on SEM time stability due to the crustal model (namely, in conjunction with the coupled SEM; Capdeville *et al.* 2003), allowing larger and fewer time-integration steps. Here, we use the same technique for parametrization and calibration of the crustal layer as in SEMum2, which is characterized by a minimum thickness of 30 km and is based on matching surface wave dispersion measurements down to 25-s period. As discussed extensively in French *et al.* (2013), this approach leads to no significant loss of accuracy for long-period

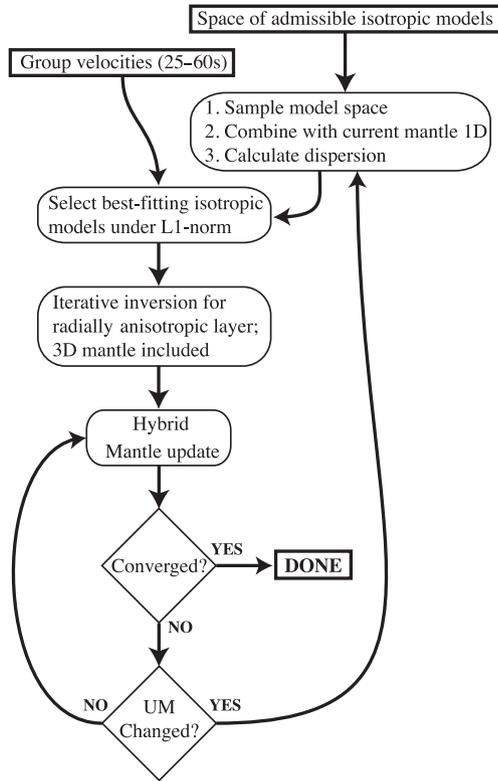
waveform modelling. Indeed, we note that this is still well below the minimum period considered in this study (see discussion of data in Section 3), particularly for our surface wave data set, which exhibits the greatest sensitivity to crustal structure. Here, we provide a more thorough overview of the crustal modelling procedure than was presented in French *et al.* (2013), where many of the details were omitted in the interest of brevity.

### 2.4.1 Motivation

In the spectral element method, solution accuracy is strongly controlled by fidelity to: (1) the shortest wavelengths present in the wavefield; and (2) spatial variation of the underlying earth model (*cf.* Komatitsch & Tromp 2002a, 2002b; Chaljub *et al.* 2003). The latter concern requires that discontinuities in material properties or their depth-derivatives must be coincident with element faces. Using an explicit time-integration scheme, stability of the SEM is determined by the Courant–Friedrichs–Lewy (CFL) condition—characterizing the maximum stable time step in terms of the minimum ratio between spatial discretization and wave speed. The time step size in turn determines the overall cost of time integration, given a desired synthetic-record duration (e.g. 10 000 s in order to capture most second-orbit fundamental-mode Rayleigh waves). In a pure global SEM, such as SPECSEM3D\_GLOBE (Tromp *et al.* 2008), the CFL condition is often dominated by high compressional velocities in the core. If the core is replaced with a pseudo-analytical modal solution such as used here (Capdeville *et al.* 2003), then small spatial discretization in the thin oceanic crust dominates instead.

The fact that crustal thickness controls our CFL condition motivates us to implement a homogenization scheme (e.g. Capdeville & Marigo 2007), in which a set of thin discontinuous layers are replaced with a single, smoothly varying anisotropic layer that has an equivalent seismic response at long periods [i.e. the zeroth-order scheme of Backus (1962), in the terminology of Capdeville & Marigo (2007)]. Like its predecessors, SEMUCB-WM1 was developed using a crustal model based on a ‘hidden’ homogenization, designed to fit *observed* surface wave group-velocity dispersion directly, instead of homogenization of an intermediate or *a priori* crustal model (i.e. ‘hidden’ in that we have no knowledge of the true crust, only its effect on the wavefield measured through surface wave dispersion). The rationale for this choice is that we prefer to match the effect of the crust on the wavefield directly, as opposed to that predicted by currently available global crustal models (e.g. Crust2.0), which rely on extrapolation from geologically analogous regions in areas of poor sampling (where the uncertainties associated with this approach are not well quantified).

The model SEMum of Lekić & Romanowicz (2011a) was the first global tomographic model from the Berkeley group to use such a scheme—enabling large time steps in the SEM by using a uniform 60-km thickness for the crustal layer. In the model SEMum2, as well as this study, we adopt a more geologically plausible laterally varying thickness, at the expense of a more restrictive CFL condition. Starting from Crust2.0 Moho depth, we restrict crustal thickness to the interval between 30 and 60 km and filter this fictitious Moho surface to twice the lateral resolution of our SEM mesh. The filtering step, valid at long periods, prevents spatial aliasing of Moho topography in the SEM. Given this prescribed geometry, we use a two-step model-space sampling approach to find a radially anisotropic crustal-layer structure that fits the dispersion data. Crustal structure is parametrized in depth using degree-4 Lagrange polynomials with Gauss–Lobatto–Legendre interpolation points (hereafter GLL) of the same type used in the SEM—a similar approach to that taken



**Figure 3.** A high-level overview of the development of the smooth crustal layer (Section 2.4), including its overall context in the mantle-model inversion (Section 4.1).

by Fichtner & Igel (2008) in their alternative implementation of a smooth crustal layer.

#### 2.4.2 Development

The crustal model is generated following a two-step procedure. A high-level overview of this process, as well as its role in the overall mantle inversion, appears in Fig. 3. As shown, our crustal model implementation is *iterative* in nature—receiving periodic updates in order to remain consistent with evolution of the mantle model below.

##### Step 1: Crustal-layer initialization

We begin by defining a space of admissible *isotropic* GLL-parametrized 1-D crustal models, characterized by uniform distributions over

- (i) GLL-nodal shear velocity:  $3.0 \leq V_S \leq 4.6 \text{ km s}^{-1}$ ;
- (ii) Restricted Moho depth:  $30 \leq h_m \leq 60 \text{ km}$ ;
- (iii) Realistic bathymetry:  $0 \leq h_s \leq 6 \text{ km}$ .

The above range of  $V_S$  nodal values was chosen based upon examination of realistic (e.g. Crust2.0) crustal layers overlying typical upper-mantle structure at depths  $\leq 60 \text{ km}$ , which have in turn been vertically smoothed with a  $\sim 10\text{-km}$  sliding window.  $V_P$  and  $\rho$  nodal values for the crustal layer are scaled from  $V_S$  following the relations of Brocher (2005), leading to a family of 1-D models of the crustal layer parametrized as

$$V_S(r) = \sum_{n=0}^4 \hat{V}_S^n \ell^n [x(r)], \quad (4)$$

$$V_P(r) = \sum_{n=0}^4 \hat{V}_P^n \ell^n [x(r)], \quad (5)$$

$$\rho(r) = \sum_{n=0}^4 \hat{\rho}^n \ell^n [x(r)], \quad (6)$$

where  $\{\hat{V}_S^n, \hat{V}_P^n, \hat{\rho}^n : n = 0, \dots, 4\}$  represent nodal values,  $r$  is Earth radius within the crustal layer,  $\ell^n(x)$  are the GLL–Lagrange basis functions and  $x(r)$  maps crustal radii onto the reference interval  $x \in [-1, 1]$ . Crustal  $Q$  is taken from QL6 (Durek & Ekström 1996), while  $\rho$ ,  $V_P$  and  $V_S$  are held constant in the water layer, which is not parametrized using the polynomial basis—consistent with the fact that the ocean is modelled as a load in the SEM and is not internally meshed (e.g. Komatitsch & Tromp 2002b). In a grid-search fashion, we draw approximately  $2 \times 10^5$  realizations of smooth structure from these *a priori* distributions, and splice the resulting crustal models to our 1-D-reference mantle structure below the Moho. For each model realization, we calculate fundamental-mode Love and Rayleigh wave group-velocity dispersion.

Next, group-velocity dispersion maps (25–60 s period) are resampled on the same set of 10 242 nodes against which the  $V_S$  spherical splines are registered (equivalent to  $\sim 2^\circ$  lateral spacing). This particular set of nodes forms the mesh over the surface of the Earth on which the smooth crustal model is constructed. The dispersion data are the same as those used for crustal model calibration in the earlier SEMum and SEMum2 models (M. Ritzwoller, personal communication, 2009; Ritzwoller *et al.* 2002). At each crustal-model node, we select the 1-D model realization that best fits the resampled Rayleigh and Love wave dispersion simultaneously, while honouring local crustal thickness from the filtered Moho surface introduced above and the bathymetry from a similarly filtered ETOPO2 (<http://www.ngdc.noaa.gov>). We are able to perform this matching step because of the manner in which we parametrized the above sampling. In particular, for each realization of *volumetric* crustal-layer structure (i.e. velocity and density), letting  $\mathbf{c}_i$  denote the  $i$ th realization, we have calculated synthetic dispersion curves over a regular grid of Moho depth  $h_m$  and bathymetry  $h_s$ —thus defining a discrete group-velocity ‘surface’  $U_{L,R}^{\text{syn}}[\mathbf{c}_i](h_m, h_s)$  at any modelled period. We may therefore obtain an accurate estimate of  $U_{L,R}^{\text{syn}}[\mathbf{c}_i]$  at arbitrary  $(h_m, h_s)$ , within the bounds of the sampling, using bicubic interpolation. Once all crustal-layer realizations are interpolated to exactly the  $(h_m, h_s)$  point associated with a given model node location, we can easily select the best-fitting realization—a process that is repeated for all 10 242 nodes. Fit between synthetic and observed group velocities is measured in the  $L_1$ -norm to reduce sensitivity to outlier measurements common at short periods, with additional weighting by estimated measurement uncertainty (Shapiro & Ritzwoller 2002); for example,

$$\Phi(\mathbf{c}_i) = \frac{N_p(\sigma_R + \sigma_L)}{\sigma_R \sigma_L} \sum_{c \in \{R,L\}} \sum_{j=1}^{N_p} \frac{|U_c^{\text{syn}}[\mathbf{c}_i](T_j) - U_c^{\text{obs}}(T_j)|}{\sigma_c}, \quad (7)$$

where  $\mathbf{c}_i$  again represents the  $i$ th crustal-layer realization,  $\{T_j : j = 1, \dots, N_p\}$  is the range of periods considered,  $U_{R,L}^{\text{syn}}[\mathbf{c}_i](T_j)$  and  $U_{R,L}^{\text{obs}}(T_j)$  are synthetic and observed Rayleigh and Love wave group velocities and  $\sigma_R$  and  $\sigma_L$  are measurement uncertainties assigned to Rayleigh and Love wave data (40 and  $50 \text{ m s}^{-1}$ , respectively—see Shapiro & Ritzwoller 2002). The resulting isotropic initial model is thus parametrized radially in GLL–Lagrange interpolants and laterally by linear barycentric interpolation on simplices defined by triples of neighbouring model nodes.

### Step 2: Iterative inversion for the anisotropic crustal layer

We next perform an iterative inversion for ‘radially anisotropic’ crustal-layer structure in a neighbourhood near the 3-D initial model. The introduction of anisotropy is required in order to simultaneously fit the Rayleigh and Love wave dispersion more consistently than is possible with the purely isotropic initial model (e.g. Backus 1962). Importantly, the non-linear, model-space sampling approach of the first step considerably reduces the required number of inversion iterations relative to a linearized scheme alone. We follow a generalized least-squares approach similar to that employed in the upper-mantle inversion (eq. 2). Lateral smoothness is again enforced through the *a priori* model covariance matrix  $\mathbf{C}_m$ —here assigning fixed 500 and 1000 km correlation lengths to crustal  $V_S$  and  $\xi$ , respectively. The data covariance matrix  $\mathbf{C}_d$  is assumed diagonal, with elements equal to the corresponding  $\sigma_R$  and  $\sigma_L$  values introduced above. As in the first step, group velocities between 25 and 60 s are considered.

Unique group-velocity partial derivatives are used at each model node, which are calculated in—and therefore appropriate for—the current crustal *and* mantle model estimates. Partial derivatives are calculated following a straightforward finite-difference approach, conceptually similar to the ‘brute force’ scheme employed by Rodi *et al.* (1975) to validate their technique for deriving group-velocity partial derivatives from those for phase velocity. The crustal model at node  $n$  may be represented by a vector  $\mathbf{c}_n$  containing GLL-nodal  $V_S$  and  $\xi$  values. The elements of  $\mathbf{c}_n$  are successively perturbed by small quantities  $\epsilon$  and forward-modelled, yielding finite-difference approximations to the desired partial derivatives:

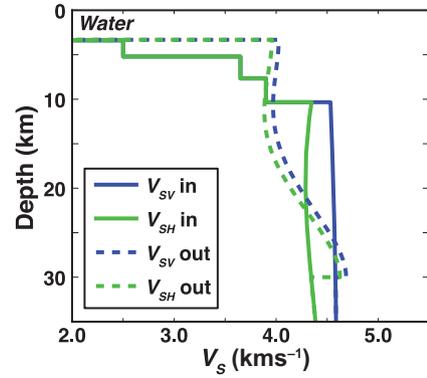
$$\left(\frac{\partial U_c}{\partial \mathbf{c}_n}\right)_{ij} \simeq \frac{U_c^{\text{syn}}[\mathbf{c}_n + \epsilon \mathbf{e}_j](T_i) - U_c^{\text{syn}}[\mathbf{c}_n](T_i)}{\epsilon}, \quad (8)$$

where  $c$  again denotes evaluation for either Rayleigh or Love waves,  $i$  is an index over measurement periods and  $\mathbf{e}_j$  denotes the unit vector in the direction of the  $j$ th model-vector component. The underlying mantle structure (i.e.  $V_S$  and  $\xi$  perturbations evaluated beneath node  $n$ ) is held fixed and, as mentioned earlier, included in the computation of  $U_c^{\text{syn}}[\cdot]$ . Given a local crustal model of  $N$  nodal-valued components, the associated group-velocity partial derivatives may be recovered in  $2(N + 1)$  dispersion calculations (including both Rayleigh and Love waves and all considered periods), which can be evaluated independently in parallel.

We find that after three iterative crustal updates, the mean absolute misfit over all periods and model nodes falls into the target range of  $<50 \text{ m s}^{-1}$ . A new crustal model is derived each time the underlying upper-mantle structure is updated (i.e. at each inversion iteration for SEMum2, but only at the first iteration for SEMUCB-WM1; see Section 4.1). Furthermore, a similar finite-difference approach may be used to estimate group-velocity partial derivatives with respect to the mantle model, in which case, the crustal model is held fixed, while the elements of  $\mathbf{m}$  are perturbed and the range of  $T_i$  considered is extended to longer periods (up to 150 s in practice). As noted in French *et al.* (2013), group-velocity data may be incorporated into the mantle model inversion in order to enforce consistency between the crustal model and mantle  $V_S$  adjacent to the Moho—an approach also taken in the first iteration of this study (Section 4.1).

#### 2.4.3 A synthetic example

In Figs 4 and 5, we present an illustrative synthetic example of the crustal modelling scheme described above, similar to that used to demonstrate this technique in the supporting material of French

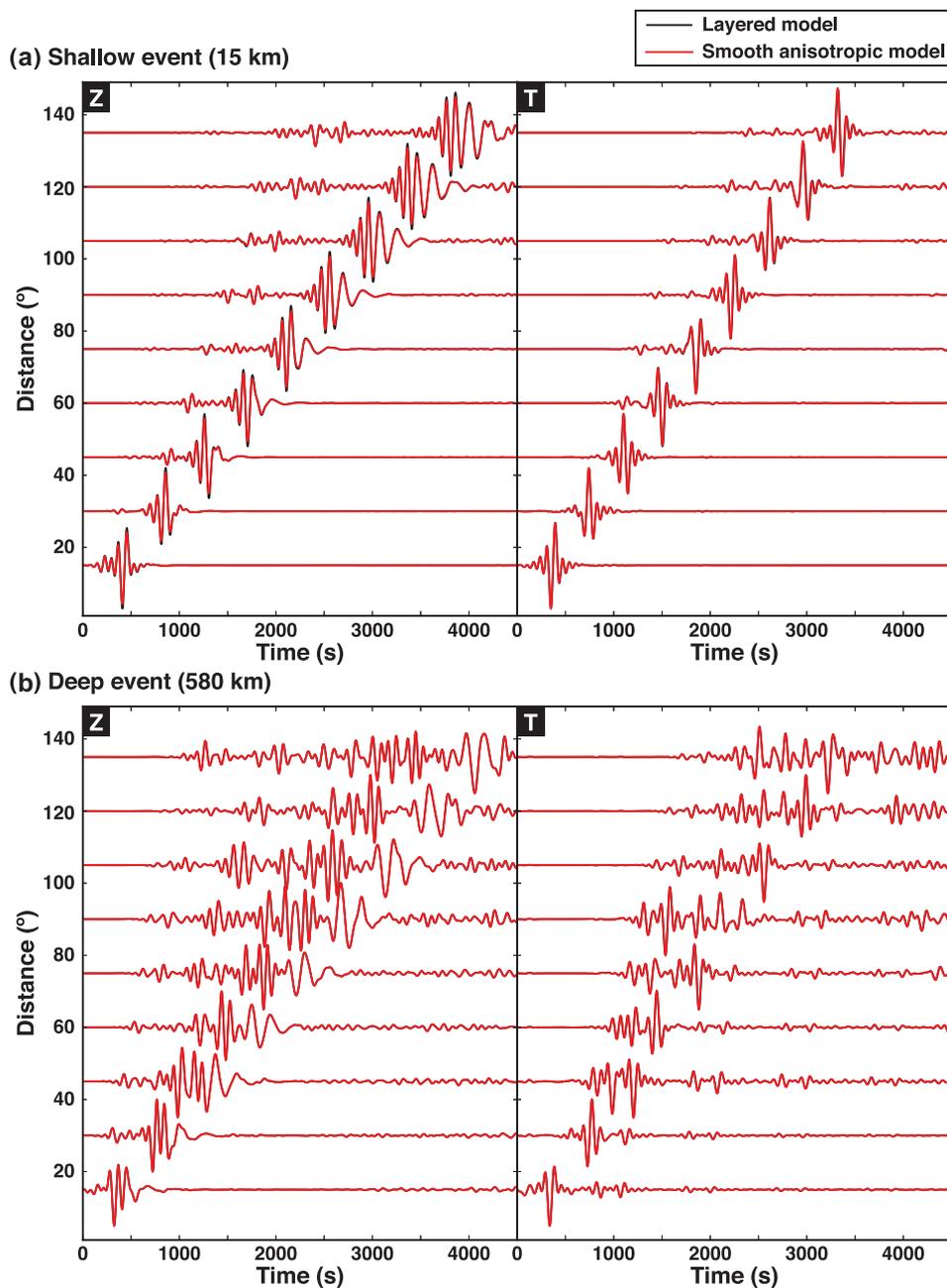


**Figure 4.** A synthetic example of the crustal-layer calibration scheme: an ‘oceanic’ layered input model and a corresponding smooth anisotropic output model, with the latter derived from the former using the calibration scheme described in the text (Section 2.4).

*et al.* (2013). Here, we focus specifically on oceanic settings, where a 30-km Moho depth most clearly deviates from reality. In Fig. 4, we show 1-D shear velocity profiles from two models: (1) a hypothetical layered input model that is typical of ‘average’ oceanic crust; and (2) the resulting smooth anisotropic layer (with a Moho at 30-km depth) that has been calibrated using the two-step scheme in Section 2.4.2 by matching the group-velocity dispersion predicted from the former, which is treated as synthetic data. The layered crustal model is a representative example drawn from an ensemble of samples taken from Crust2.0 (with additional random perturbations to layer thicknesses) and superimposed on top of randomly selected profiles of oceanic mantle structure from our starting model, SEMum2 (mantle structure was taken to be known *a priori* during the crustal-layer calibration). Associated with this, in Fig. 5, we compare full long-period waveforms (computed in our surface wave passband) obtained in both models for shallow and deep sources on the vertical and transverse components. Importantly, we note that both the overtone and fundamental-mode surface waves, the latter being the most crust-sensitive subset of our data, fit quite well between the layered input and smooth output models, even at long propagation times. This example serves to illustrate that our particular choice of calibration scheme and associated parametrization of the smooth crustal layer (i.e. basis functions,  $\rho$  and  $V_P$  scalings, etc.) should not be expected to introduce artefacts in the wavefield capable of biasing our mantle model in regions where the prescribed 30-km Moho depth is not realistic.

## 2.5 Crustal structure in NACT

Although the SEM may accurately treat arbitrary 3-D crustal structure, care must be taken to properly account for crustal effects in our NACT-based sensitivity kernel calculations. Because the SEMum model was derived using a uniform-thickness crustal layer, crust-induced path effects in NACT were easily handled by accounting for volumetric structure alone (perturbations to  $V_S$  and  $\xi$ ). In other words, the strongly non-linear effect of a laterally varying crustal thickness—which is generally more difficult to accurately model with perturbation theory—was not an issue. The variable Moho depth employed in both the SEMum2 and SEMUCB-WM1 models, on the other hand, necessitates accurately treating the non-linear effects of crustal structure on the NACT-based kernels. We adopt a two-part approach to ensure accurate NACT-based modelling: (1) specially formulated corrections for crust-induced path effects



**Figure 5.** Long-period ( $T \geq 60$  s) teleseismic waveforms computed in the synthetic example layered input (black) and smooth anisotropic output (red) models shown in Fig. 4, for shallow (a) and deep (b) sources recorded on the vertical (left-hand side) and transverse (right-hand side) components.

that mimic the non-linear response of the crust considerably better than ‘standard’ linear corrections (Lekić *et al.* 2010); and (2) mantle-structure sensitivity kernels modified to reflect overlying crustal thickness variations. The detailed description of both schemes is given in the Appendix.

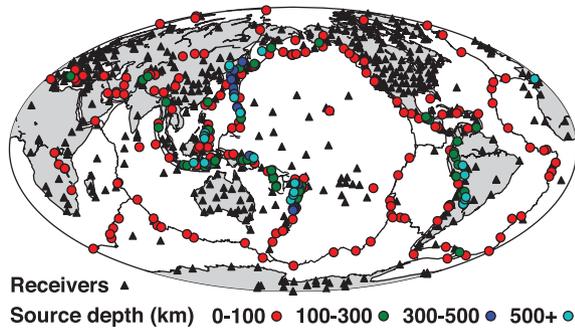
### 3 DATA SET

While previous SEM-based global tomographic models from our group (Lekić & Romanowicz 2011a; French *et al.* 2013) focused primarily on upper-mantle and transition-zone structure, and thus included only fundamental and overtone mode surface wave waveforms at long periods ( $T \geq 60$  s), our current whole-mantle modelling must now include body wave data. To this end, we use a

data set comprised of full, three-component teleseismic waveforms, filtered in multiple passbands, allowing us to incrementally incorporate higher frequency body wave data as the inversion progresses:

- (i) Surface wave passband: cut-off at 400 and 60 s (corners at 250 and 80 s);
- (ii) Body wave passband (Filter I): cut-off at 300 and 36 s (corners at 180 and 45 s);
- (iii) Body wave passband (Filter II): cut-off at 300 and 32 s (corners at 180 and 38 s).

We note that this surface wave passband is identical to that used in constructing SEMum and SEMum2. In addition, we expand our data set as the inversion progresses by incorporating additional events: starting from the same 203 events used in developing the latter two



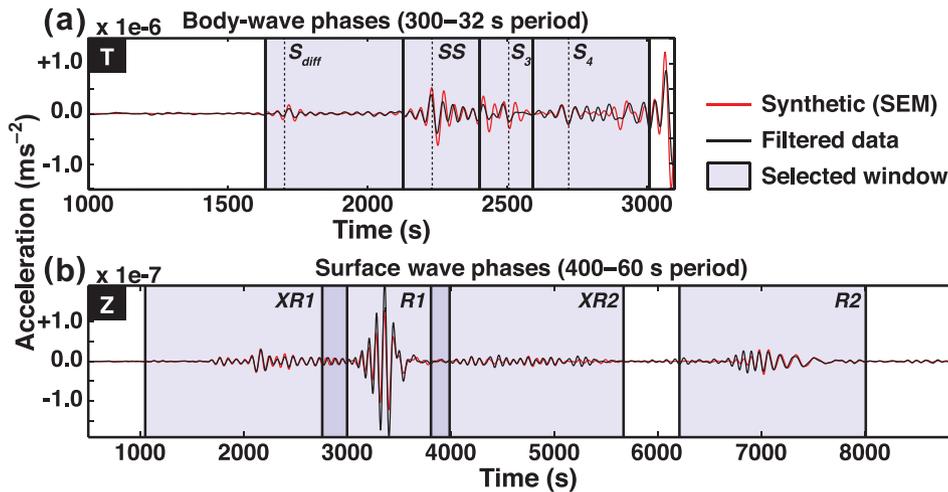
**Figure 6.** Global distribution of the 273 seismic sources (circles, colour-coded by centroid depth) and over 500 seismic receivers (black triangles) used in this study.

models, and increasing to 273 events total. The particular set of new events were chosen to be spatially distributed in a complementary manner to the earlier set of 203 (to maximize independent constraint on structure). The complete set of sources and receivers used in our inversion is shown in Fig. 6.

Waveform data are processed using an automated (but human-verified) windowing approach (Li & Romanowicz 1996; Panning & Romanowicz 2006) that groups trains of phases by amplitude and constraint on Earth structure (Fig. 7). Using this technique, we are able to invert long-period waveforms in the time domain—making use of both phase and amplitude information—while preventing large-amplitude phases (e.g. the  $SS$  body wave phase, sensitive to upper and mid-mantle structure) from dominating comparatively low-amplitude ones (e.g.  $S_{diff}$ , critical to resolving lowermost man-

tle structure). Furthermore, these windows may in turn be weighted according to data uncertainty and/or noise estimates, as well as redundancy in sensitivity—see Li & Romanowicz (1996, appendix A) for a more detailed discussion of the weighting scheme. Indeed, these weights are the basis for construction of the data covariance operator referred to in Section 2.3. Window selection is performed based on similarity to SEM synthetic seismograms, computed in the previous iteration of the 3-D mantle model—an operation that is performed at iteration of our inversion. The windowing technique is applied *each* to waveform data out to 10 000 s post-origin time for the surface wave data set (thereby including phases out through second-orbit Rayleigh waves) and out to 5500 s for body wave data (capturing multiple- $ScS$  phases valuable for constraining lower most mantle structure). By the end of the inversion, our data set is comprised of over 447 800 waveform windows, corresponding to nearly 15.5M datapoints sampled at their corresponding Nyquist rates (Table 1).

Finally, as noted in Section 2.4.2, in addition to our large waveform data set, we employ global surface wave group-velocity dispersion maps at periods between 25 and 150 s in our inversion (M. Ritzwoller, personal communication, 2009). Namely, the dispersion maps are those of Ritzwoller *et al.* (2002) and a discussion of uncertainties associated with the underlying measurements may be found in Shapiro & Ritzwoller (2002). We use group velocities between 25 and 60 s to constrain the structure of our smooth crustal layer, while including periods up to 150 s in our mantle inversion in order to enforce consistency between the crustal layer and mantle structure adjacent to the Moho when necessary. We note that the resolving power of the dispersion data is rather weak at depths below roughly 100 km and that upper-mantle structure is overwhelmingly



**Figure 7.** Example windowed body wave (a) and surface wave (b) waveforms using the processing scheme described in the text (Li & Romanowicz 1996). Data windows are selected based on similarity to SEM synthetics computed in the most recent iteration of the mantle model (here, synthetics are from the final model, SEMUCB-WM1). Thus, the data set is reprocessed at each iteration, and typically expands as the inversion progresses and the model improves.

**Table 1.** Final-iteration waveform-window counts  $N_{wp}$  and aggregate data set sizes (number of datapoints  $N_d$ , sampled at the corresponding Nyquist rate), organized by receiver component and waveform data type/passband.

	L		T		Z	
	$N_{wp}$	$N_d$	$N_{wp}$	$N_d$	$N_{wp}$	$N_d$
Body (32 s)	83 374	1 928 538	62 619	959 878	106 497	2 527 759
Fundamental (60 s)	17 688	921 443	26 643	990 836	30 974	1 635 960
Overtone (60 s)	25 932	1 603 959	25 452	883 297	42 297	2 502 012
Mixed (60 s)	6703	422 053	9411	469 645	10 253	645 579

constrained by the waveform data set (as confirmed by our earlier tests during the construction of the SEMum2 model).

## 4 RESULTS

### 4.1 Inversion and fits

The starting model for our whole-mantle inversion consists of SEMum2 in the upper mantle and transition zone ( $\leq 800$ -km depth) and the model SAW24B16 of Mégnin & Romanowicz (2000) below, smoothly combined over radial length scales on the order of  $\geq 200$  km (lower-mantle structure was previously fixed to SAW24B16 during the SEMum and SEMum2 inversions). Throughout our iterative inversion, we assume seismic source parameters for each event consistent with those reported by the Global CMT project (<http://www.globalcmt.org>)—we return to the question of seismic source *inversion* in Section 5.1.2. The structural inversion is comprised of three phases, allowing us to incrementally enlarge our data set and incorporate shorter period waveform data:

(i) Phase I: In the first phase, we performed one iteration of inversion for whole-mantle structure using the 60-s surface wave and 36-s body wave (Filter I) data sets, picked from the 203 events used in SEMum and SEMum2. Because this iteration inverted for mantle structure including that adjacent to the Moho, we also included constraints from surface wave group-velocity maps between 25 and 150 s as discussed in Section 3. We found that the upper-mantle structure in the model changed very little, aside from slightly larger amplitudes following the introduction of the body wave data. Motivated by this, we chose only to invert for structure at  $\geq 300$ -km depth in the remaining iterations of the inversion. However, since upper-mantle amplitudes had changed slightly, we first performed one last recalibration of the crustal model using the two-step inversion scheme described in Section 2.4.2.

(ii) Phase II: Next, we introduced 70 new events with moment-magnitudes between 5.8 and 7.3, selected so as to be maximally complementary to the distribution of the 203 earlier events and primarily motivated by enlarging our body wave data set. We picked the new-event data and reprocessed the older event data using the SEM synthetics from the previous iteration. Thereafter, we performed another inversion iteration, again using the 60-s and 36-s filter passbands, but now for structure below 300-km depth only. We chose to include both overtone and fundamental-mode waveform data, as their sensitivity is non-negligible in the depth range considered (though clearly small for the fundamental modes).

(iii) Phase III: In the final phase, we reprocessed the data from the 273 events using a new shorter period body wave passband (Filter II). We then inverted one last time for structure below 300-km depth using the 60 s and now 32-s data passbands.

In Table 2, we summarize the final-iteration waveform fits for the different filter passbands and receiver components used in our in-

**Table 2.** Final-iteration waveform variance reduction, defined as one minus the squared 2-norm of the waveform residual normalized by that of the data, expressed in percent as  $VR = 100 \times [1 - \|\mathbf{d} - \mathbf{g}(\mathbf{m})\|_2^2 / \|\mathbf{d}\|_2^2]$  and organized by component and data type/passband.

	L (per cent)	T (per cent)	Z (per cent)
Body (32 s)	54.6	56.8	51.5
Fundamental (60 s)	69.1	76.6	71.9
Overtone (60 s)	78.2	68.2	79.4
Mixed (60 s)	75.4	79.6	78.9

version. To ensure that our inversion was converging, we inspected at each iteration whether more waveform windows were selected in the next data reprocessing round than would have been using synthetics from the previous model. By the final iteration, we found only small gains in the numbers of selected windows, indicative that the inversion had likely converged (for the particular passbands considered). This assessment of convergence may also be supplemented by testing fits to held-out data—namely, waveforms from events not included in the inversion—to ensure we are not overfitting the inversion data set (i.e. fitting noise). For example, this latter approach was used to confirm that our final iteration was indeed warranted (see Section 5.1.1).

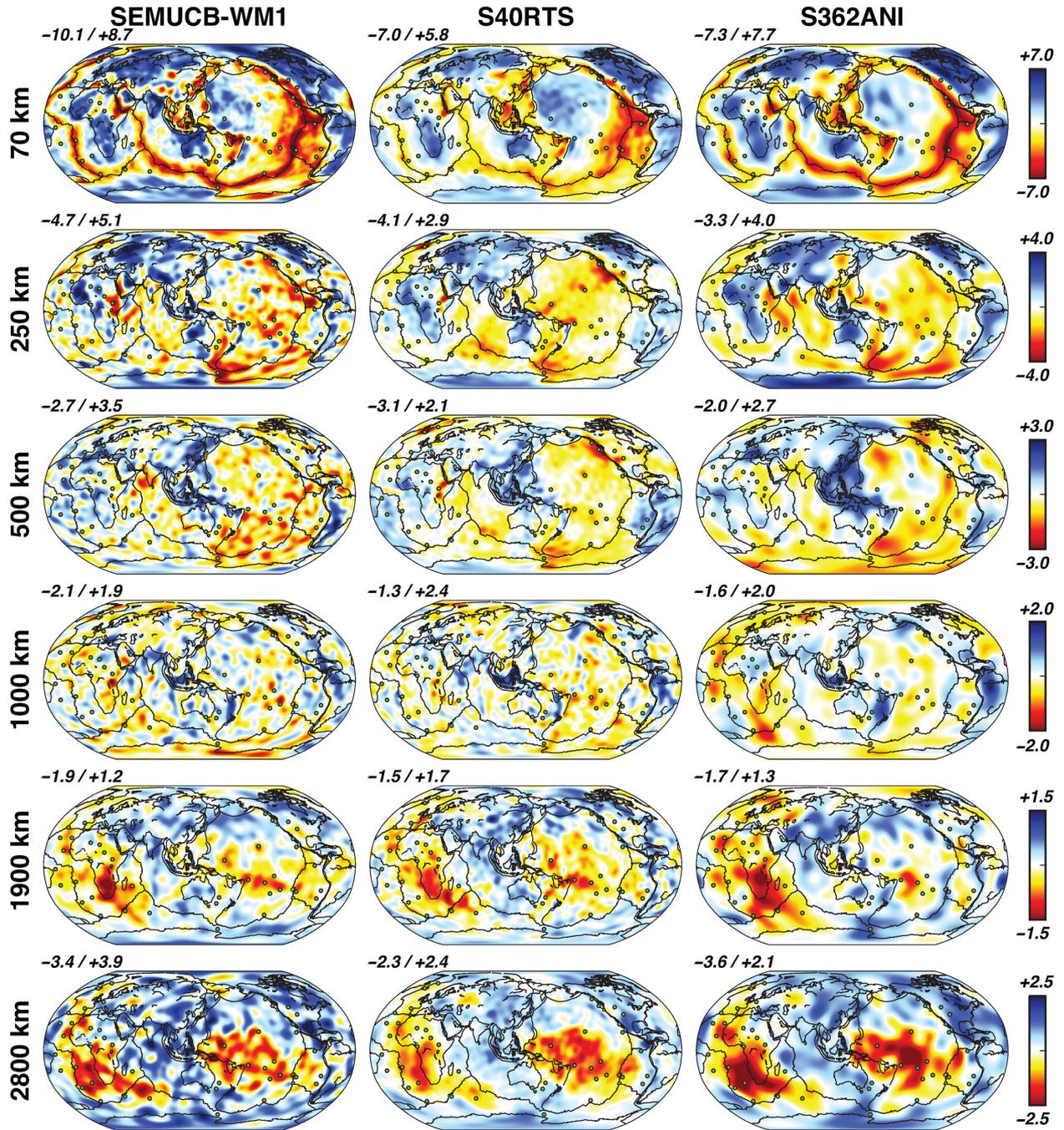
Surface wave variance reduction (VR) values in Table 2 are quite similar to those obtained for SEMum2, which is not surprising given that upper-mantle and transition-zone structure has remained largely the same, while body wave VR is in general higher than that reported for previous Berkeley whole-mantle models—typically  $< 50$  per cent (e.g. Mégnin & Romanowicz 2000; Panning & Romanowicz 2006). The latter comparison is not straightforward, however, as VR for these earlier models was evaluated using approximate waveform modelling techniques (i.e. not SEM). Unexplained variance may be attributable to a number of factors, including: unmodelled structure (for example, below the limit of our parametrization), noise or minor instrument errors (though data are rejected when either is particularly severe) and uncertainty in seismic source parameters (we will return to this latter point in Section 5.1.2).

Overall, four rounds of SEM simulation were required to complete these three phases of inversion (one for the earlier 203-event data set, and three for the enlarged 273-event data set) in addition to ancillary simulations needed for validation (e.g. ensuring convergence against held-out data, numerical experiments for tuning the weights assigned to our waveform data sets, etc.). Omitting our NACT-based Hessian-estimation calculations, as well as parallel matrix factorization for computing model updates under the Gauss–Newton scheme, this study required approximately 3M CPU hr for SEM simulations alone. These latter simulations were performed on *Hopper*, a Cray XE6 at the National Energy Research Scientific Computing Center (NERSC), while the NACT-based Hessian estimation and Gauss–Newton model update computations were performed on NERSC *Edison*, a Cray XC30.

### 4.2 Whole-mantle shear velocity structure

#### 4.2.1 Global isotropic structure

In Fig. 8, we show global maps of isotropic  $V_S$  variations at a range of depths in our model. For comparison, we also show  $V_S$  structure from two other recent whole-mantle studies: S40RTS of Ritsema *et al.* (2011) and S362ANI of Kustowski *et al.* (2008). From these maps, it is immediately clear that all three models tend to agree well at long wavelengths—consistent with our earlier observations for the SEMum2 model at upper-mantle and transition-zone depths (French *et al.* 2013). In Fig. 9, we examine intermodel correlations of  $V_S$  structure as a function of depth between SEMUCB-WM1 and these same two models, as well as both our starting model (SEMum2 + SAW24B16) and SAW642ANb of Panning *et al.* (2010) [a revised version of SAW642AN from Panning & Romanowicz (2006), using a new treatment of the crust]. Here, we confirm quantitatively that long-wavelength  $V_S$  variations in SEMUCB-WM1 do indeed correlate well with all four models. Up to spherical-harmonic degree 12, we find that each correlates with SEMUCB-WM1 at over 0.6 throughout nearly the entire mantle (except portions of the

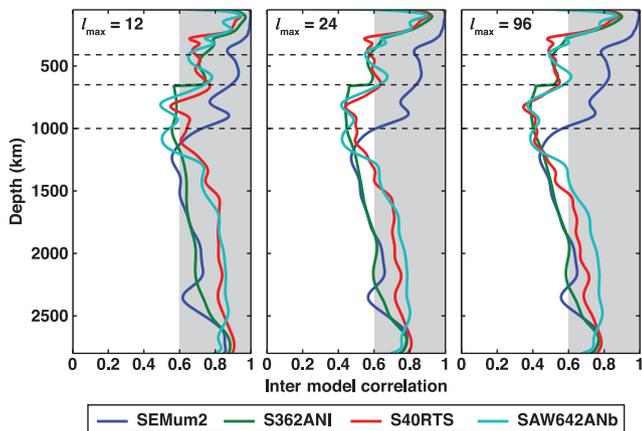


**Figure 8.** Map views of global  $V_S$  variations at a range of depths throughout the mantle for the model obtained in this study (SEMUCB-WM1), as well as recent whole-mantle models S40RTS of Ritsema *et al.* (2011) and S362ANI of Kustowski *et al.* (2008). Variations are plotted in percent with respect to the global mean at each depth, with the exception of 2800 km, plotted with respect to the 1-D model PREM (Dziewonski & Anderson 1981). Inset values (upper-left corner of each panel) represent maximum peak-to-peak variation for each model at the corresponding depth. Circles denote hotspot locations from Steinberger (2000).

uppermost 500 km of the lower mantle), with the highest correlations seen in the uppermost ( $>0.9$ ) and lowermost ( $>0.8$ ) mantle where data coverage is also highest. By degree 24, which includes the full spectral range of models S362ANI and SAW642ANb, we see the same radial dependence of intermodel correlation, with values again remaining quite high in both the uppermost ( $>0.8$ ) and lowermost ( $>0.7$ ) mantle. Finally, by degree 96, corresponding to the full spectral range of the SEMum2 and SEMUCB-WM1  $V_S$  basis, we can see that our model remains very close to SEMum2 in the transition zone and above—correlating at approximately 0.8

and higher in this depth range. Furthermore, we see that in all spectral passbands, structure at the base of the mantle in SEMUCB-WM1 correlates most strongly with S40RTS, not the starting model (SAW24B16 at this depth, but labelled as SEMum2 in Fig. 9).

Returning to Fig. 8, one of the clear differences across models is again amplitudes: peak-to-peak variation in  $V_S$  is larger in our model, particularly within the top and bottom boundary layers of the mantle where heterogeneity is expected to be strong, while comparable variation is seen in the mid-mantle. This effect is especially pronounced at shallow depth, with our model showing nearly 19

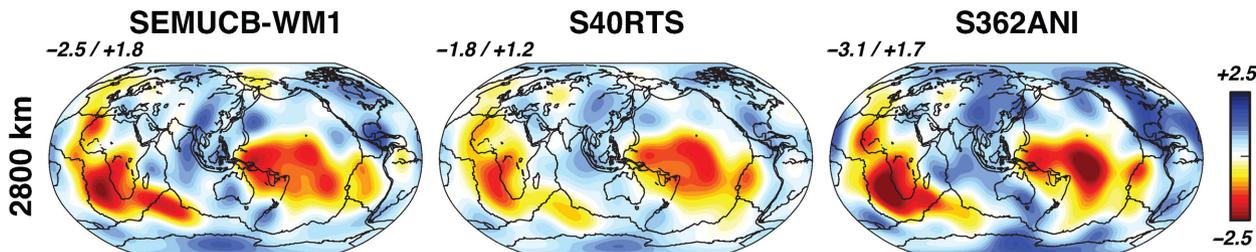


**Figure 9.** Intermodel  $V_S$  correlations as a function of depth for  $l_{\max} \in \{12, 24, 96\}$  (96 corresponds to the full spectral range of the SEMUCB-WM1  $V_S$  basis) between SEMUCB-WM1 and models: (1) SEMum2 (French *et al.* 2013) with SAW24B16 (Mégnin & Romanowicz 2000) below  $\sim 800$  km; (2) S362ANI (Kustowski *et al.* 2008); (3) S40RTS (Ritsema *et al.* 2011) and (4) SAW642ANb (Panning *et al.* 2010) [a revised version of SAW642AN of Panning & Romanowicz (2006), obtained using a different treatment of crustal structure]. The shaded region corresponds to correlations above 0.6.

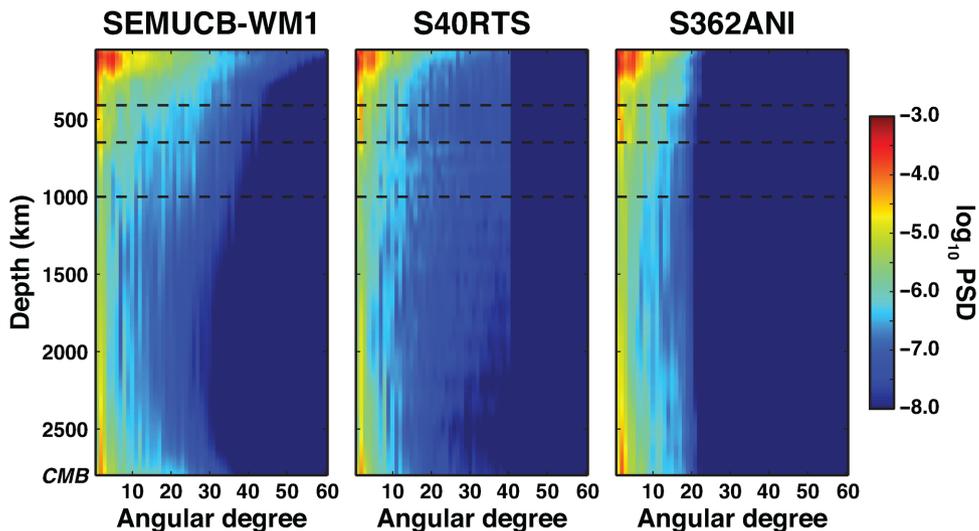
percent peak-to-peak variation at 70 km. Indeed, similar relative amplitudes were seen for SEMum2 and our upper-mantle structure has not changed significantly (as noted in Section 4.1). As noted previously by Lekić & Romanowicz (2011a) and French *et al.* (2013),

a likely candidate for this difference in retrieved amplitudes is the use of SEM-based forward modelling, which easily accounts for wave-propagation phenomena that can otherwise lead to systematic underestimation of amplitudes if neglected, such as wave front healing (e.g. Nolet & Dahlen 2000).

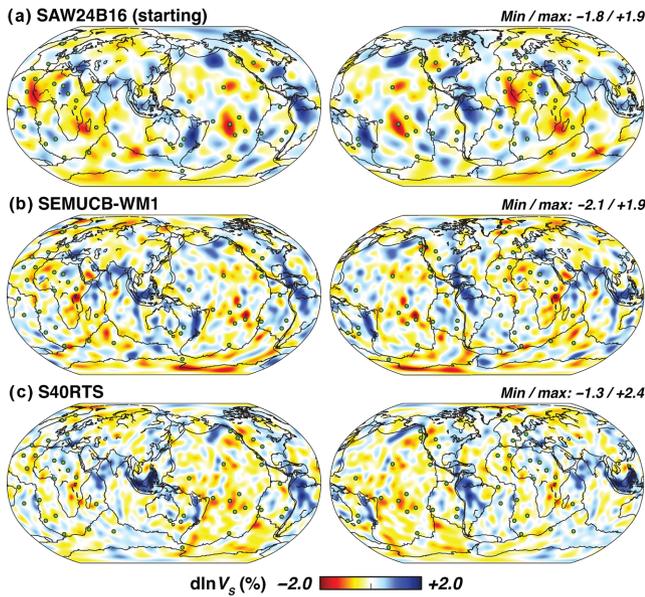
Another intriguing observation from Fig. 8 is the clear difference in appearance between the three models shown near the base of the mantle. Namely, the LLSVPs in SEMUCB-WM1 appear to contain more concentrated and distinct low-velocity anomalies than seen in either of the other two models, which exhibit LLSVPs that are in general more laterally continuous. Indeed, this latter observation may seem at odds with our earlier discussion of long-wavelength correlation between models in the context of Fig. 9. However, looking now to Fig. 10, where we show the same three models at 2800-km depth filtered to low degrees (namely,  $l_{\max} = 12$ ), it is immediately apparent that all three contain very similar long-wavelength structure, especially SEMUCB-WM1 and S40RTS (although SEMUCB-WM1 appears to contain somewhat stronger variations at these low degrees). We continue to examine the spectral properties of our model in Fig. 11, where we compare spherical-harmonic power spectral density across the same three models shown in Fig. 8. There, we find that the well-known degree-2 signal, found both at the base of the mantle and in the transition zone, remains quite prominent in SEMUCB-WM1, though in both depth ranges our model also has significant power out to higher degrees than either S40RTS or S362ANI. Together, these observations lend us confidence that our inversion is able to recover the robust and well-known long-wavelength background structure at the base of the



**Figure 10.** Global maps of  $V_S$  variations at 2800 km for the three models shown in Fig. 8, but now filtered to spherical-harmonic degree  $l_{\max} = 12$ . As in Fig. 8, inset values (upper-left corner of each panel) represent maximum peak-to-peak variation for each filtered model.



**Figure 11.** Spherical-harmonic power spectral densities, computed for the three global models shown in Fig. 8 as a function of depth (note: colour-scale is logarithmic). Horizontal dashed lines correspond to 410, 650 and 1000-km depth (from top to bottom).



**Figure 12.** A comparison between starting and final-iteration  $V_S$  structure at 1000-km depth, as well as model S40RTS of Ritsema *et al.* (2011), plotted as per cent variation with respect to the global mean at that depth. Left-hand panels: Pacific-centred view; Right-hand panels: Atlantic-centred view. At 1000 km, the starting model is SAW24B16 of Mégnin & Romanowicz (2000). Inset values (upper-right corner of each pair of panels) represent maximum peak-to-peak variation for each model at 1000 km. Circles denote hotspot locations from Steinberger (2000).

mantle (dominated by the LLSVPs), while also imaging more concentrated low-velocity anomalies embedded therein. We will return to the question of whether the concentrations of anomalies could instead be the result of irregular sensitivity in Section 5.2.2, where our resolution analysis confirms that large-scale anomalies should be recovered without small-scale artefacts. One additional important point to note is that there do exist smaller scale features at the base of the mantle—indeed, exterior to the LLSVPs—which consistently appear in all three models, such as the low-velocity anomaly approximately located beneath Perm, Russia (Lekić *et al.* 2012).

As the upper mantle in SEMUCB-WM1 remains very similar to that of SEMum2, which has been discussed in publication previously (French *et al.* 2013), we chose to focus on depths in and below the transition zone for the majority of this discussion, as this is where our model has evolved the most. We examine this point further—namely, the evolution of our model—in Fig. 12, where we show  $V_S$  structure at 1000-km depth in both our final and starting model, with the latter based on model SAW24B16 (Mégnin & Romanowicz 2000) at this depth, as well as model S40RTS (Ritsema *et al.* 2011) for comparison. We note that  $V_S$  structure is considerably sharper in our new whole-mantle model relative to SAW24B16, including narrower *distinct* low-velocity features beneath the South-Pacific Superswell (McNutt & Fischer 1987) in the vicinity of major hotspots, as well as more concentrated fast anomalies associated with subducted slabs (e.g. Tonga-Kermadec, South/Central America). Narrow low-velocity anomalies have also now appeared beneath Tanzania and Iceland, which were not present in our starting model. A number of large, isolated high-velocity anomalies previously appearing at this depth in SAW24B16 are now missing—for example, in the Southeast Pacific. Overall, the contrast between our starting and final models, achieved with only three inversion iterations, also serves as an illustration of the power of our hybrid inversion approach, which allows us to take large steps

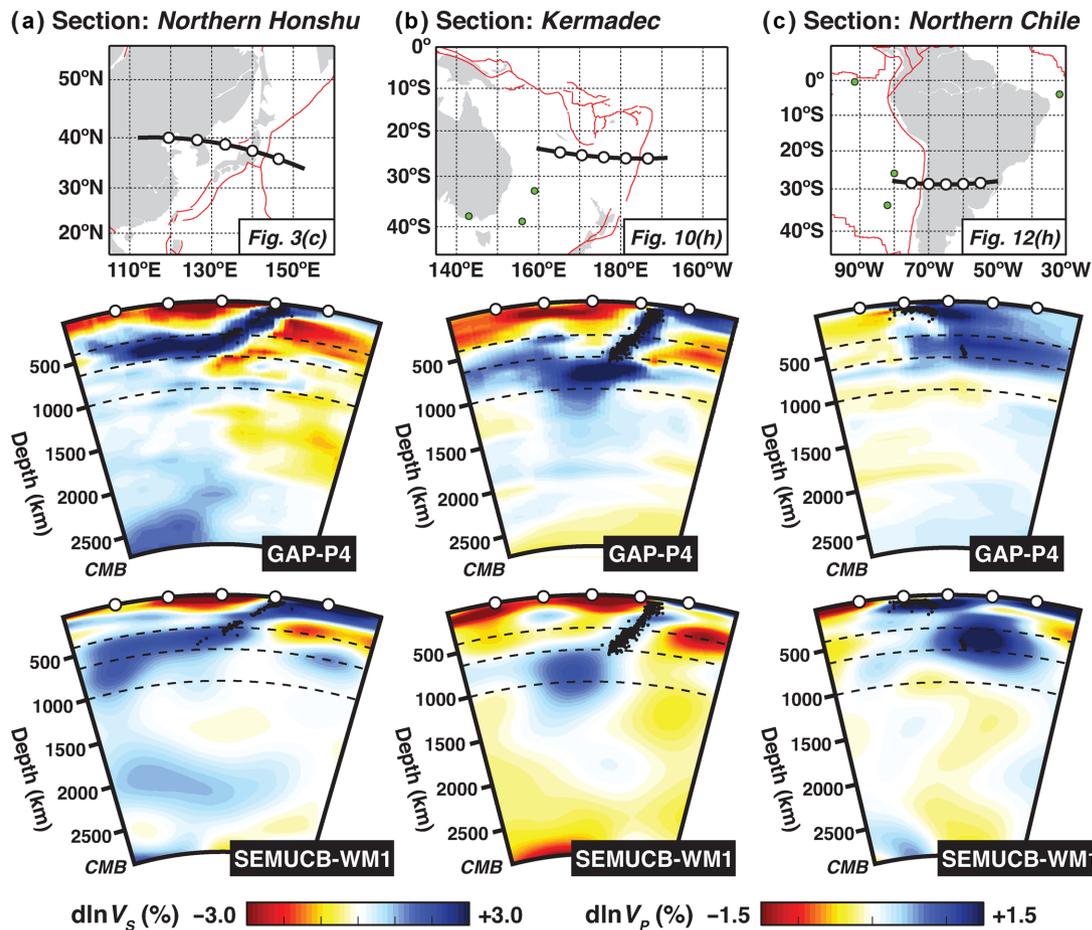
in the model space and rapidly converge on our final result. Turning our attention to Fig. 12(c), it appears that both SEMUCB-WM1 and S40RTS exhibit similar dominant length scales of heterogeneity at this depth, and are in general broadly compatible, particularly in their distribution of high-velocity anomalies. At the same time, these two models clearly differ in detail, especially for many of the isolated low-velocity anomalies seen in SEMUCB-WM1 approximately collocated with major hotspots. We will examine many of these smaller scale structures imaged in SEMUCB-WM1 in greater detail in a forthcoming publication.

Now that our  $V_S$  model extends throughout the whole mantle, and indeed exhibits a notable improvement in resolution relative to our starting model (e.g. Fig. 12), one aspect of Earth structure that we may now examine in greater detail is subduction zones. The behaviour of subducted slabs, particularly their distribution with depth—whether and where they stagnate while descending into the lower mantle—has long been a subject of debate with significant geodynamic implications. Fukao & Obayashi (2013) recently presented a global survey of subduction zones, specifically focused on cataloguing the behaviour of downgoing slabs, using the new GAP-P4 global finite-frequency  $V_P$  model of Obayashi *et al.* (2013). Indeed, GAP-P4 was constructed specifically with improved resolution of subducted slabs in mind. In Fig. 13, we show three whole-mantle cross-sections in both the SEMUCB-WM1  $V_S$  and GAP-P4  $V_P$  models focused on subduction-zone structure, featuring the NW Pacific (Japan), SW Pacific (Kermadec) and South America (Chile). These particular lines of section, as noted in Fig. 13 (inset), were intentionally chosen to match three of those used in the survey of Fukao & Obayashi (2013). Overall, we observe close agreement in both lateral distribution (and associated correlation with seismicity) as well as apparent stagnation depth between the high-velocity anomalies associated with subducted slabs in both models. We also see good agreement in the distribution of low-velocity anomalies found in the lowermost portion of the upper mantle and the transition zone between SEMUCB-WM1 and GAP-P4. Indeed, these comparisons provide independent confirmation of the anomalies imaged in our model, which lends us additional confidence that we are properly resolving the structure of subducted slabs in the upper  $\geq 1000$  km of the mantle.

#### 4.2.2 Global anisotropic structure

Next, in Fig. 14, we present maps of global  $\xi$  structure (expressed in perturbation away from isotropy, i.e.  $\xi = 1$ ) in our model, SAW642ANb of Panning *et al.* (2010) and S362WMANI of Kustowski *et al.* (2008, the whole-mantle radially anisotropic variant of S362ANI). Examining our model first, we see the expected pattern of strong radial anisotropy at the top and bottom boundary layers of the mantle, where lateral shear is expected to be strongest, and weak anisotropy at mid-mantle depths. Indeed, this observation mirrors the radial distribution of 1-D  $\xi$  structure that developed in our 1-D reference model during the course of the inversion, as previously noted in Section 2.2. In the upper 250 km, we see strong  $\xi > 1$  beneath the ocean basins, consistent with strong horizontal shear (i.e. flow), accompanied by  $\xi \leq 1$  beneath the mid-ocean ridge system, typically interpreted as a transition to dominantly vertical flow.

At the base of the mantle, we see structure decomposed into contiguous domains of strong  $\xi > 1$  and  $\xi < 1$ . Intriguingly, these patterns match the long-wavelength background (i.e. degree-2) pattern seen in  $V_S$ —with  $\xi > 1$  in high-velocity regions often hypothesized to be dominated by horizontal shear due to subducted

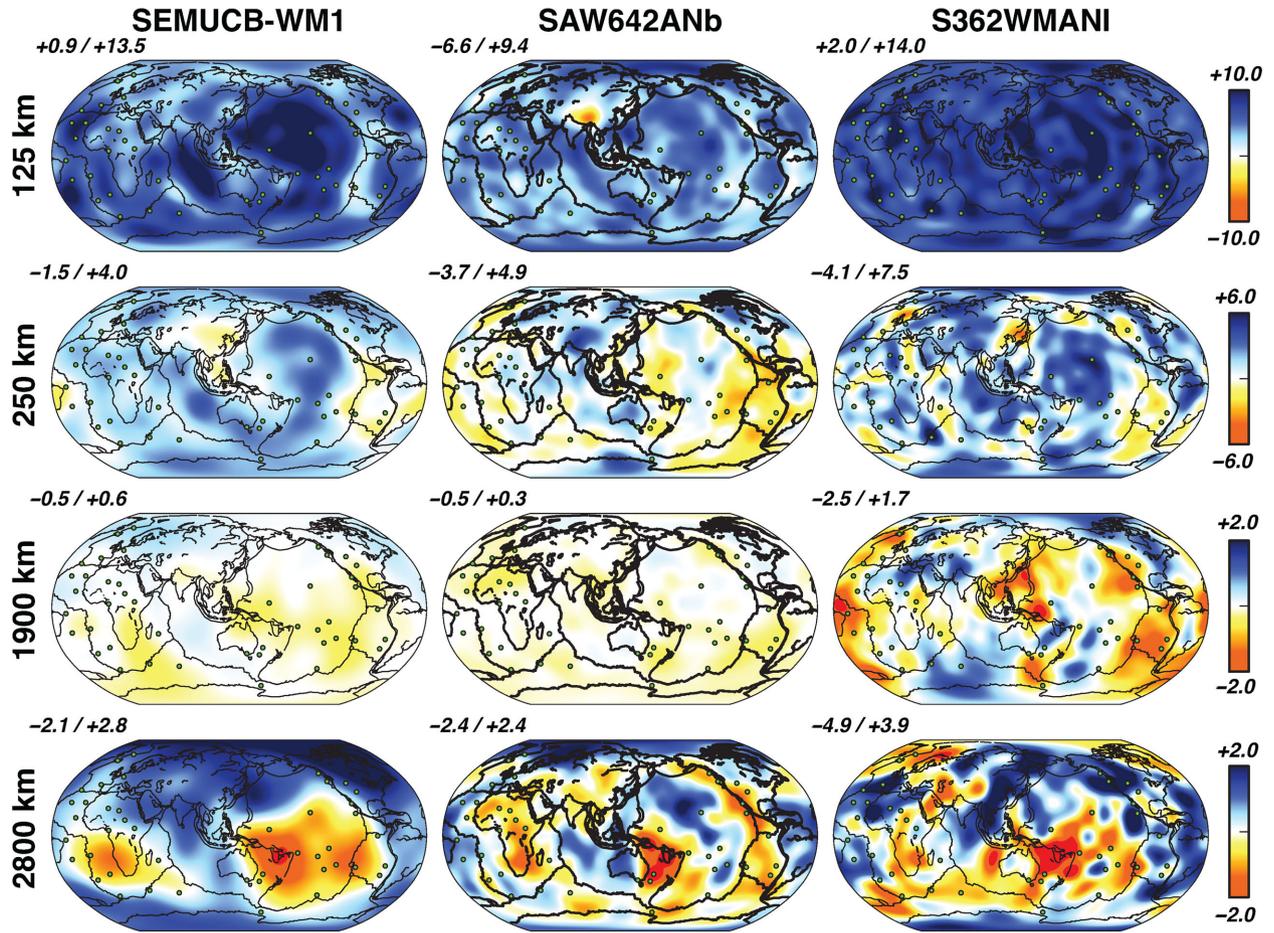


**Figure 13.** Three whole-mantle cross-sections focused on subduction zones, comparing SEMUCB-WM1  $V_S$  structure with the GAP-P4  $V_P$  model of Obayashi *et al.* (2013). Map panels indicate corresponding lines of section. Figure numbers shown inset within maps identify specific lines of section examined by Fukao & Obayashi (2013), who used GAP-P4 to survey the global behaviour of descending slabs. Deep earthquakes are shown overlain on velocity structure (black circles). Dashed lines correspond to 410, 650 and 1000-km depth (top to bottom). Note the different saturation for  $V_S$  and  $V_P$  anomalies.

slabs impinging upon the CMB, and  $\xi < 1$  associated with the low-velocity LLSVPs, potentially areas characterized by broad upwelling. Of course, the exact mechanism for the development of these anisotropic textures is not immediately clear—particularly the roles of shape versus lattice preferred orientation (e.g. McNamara *et al.* 2002), as well as compositional controls on anisotropy development, such as the presence of post-perovskite in the D'' region (e.g. Murakami *et al.* 2004). In addition, we cannot rule out that the observed anisotropic pattern may be affected by our data coverage: namely, differential lateral distribution in sensitivity of the  $SH$  and  $P-SV$  portions of our body wave data set (a potential issue we return to in discussing the results of our resolution analysis; see Section 5.2.2).

In comparison with the  $\xi$  structure of S362WMANI, we find that our model is in general smoother and weaker than the latter—with the exception of 125-km depth, where peak-to-peak variation is slightly stronger in our model. Indeed, mid-mantle radial anisotropy (for example, at 1900 km) is considerably stronger in S362WMANI than in our model, and corresponds to a depth range long expected to be nearly isotropic (e.g. Dziewonski & Anderson 1981). Conversely, we find that the amplitudes of  $\xi$  structure in SAW642ANb exhibit similar depth variation to our model, with the exception of 125 km where peak-to-peak heterogeneity in the former is stronger. Furthermore, we find that our  $\xi$  model is again smoother than SAW642ANb in general. While lateral variation in  $\xi$  agrees relatively well between

all three models at the longest wavelengths, we find that local relative variation is in some locations anticorrelated across models. For example, there is the local  $\xi$  maximum seen in S362WMANI at 250-km beneath Hawaii, which corresponds to a local minimum in our model. Similarly, the strong  $\xi$  minimum beneath Tibet in SAW642ANb at 125 km is missing from our model, which exhibits locally high  $\xi$ . In addition, there are numerous small-scale variations seen at the base of the mantle which are not consistent between models. Indeed, the level of disagreement between  $\xi$  structure seen in our model and the others is not surprising—whole-mantle radial anisotropy is notoriously difficult to constrain, typically explaining only a small fraction of data variance and is in general poorly correlated across published models (e.g. Becker *et al.* 2008). In light of this, and given limits imposed by the sensitivity of their data set, Kustowski *et al.* (2008) chose to feature S362ANI, which only inverted for anisotropy above 400-km depth. More broadly, we note that these three models were obtained using different data sets, regularization and resolution of parametrization—complicating detailed quantitative comparison across models, while at the same time suggesting qualitative explanations for observed behaviour. For example, the observation that SEMUCB-WM1 has stronger  $V_S$  but weaker  $\xi$  heterogeneity than S362WMANI may not be surprising, as heterogeneity in these two parameters can be expected to trade-off (with respect to explained data variance) under different choices of regularization and parametrization.



**Figure 14.** Map views of global  $\xi$  variations, expressed in per cent relative to isotropy ( $\xi = 1$ ), at a range of depths throughout the mantle. We show  $\xi$  structure from model SEMUCB-WM1 obtained in this study (left-hand side), as well as that from the models SAW642ANb (middle) of Panning *et al.* (2010) and S362WMANI (right-hand side) of Kustowski *et al.* (2008). SAW642ANb is a revised version of SAW642AN of Panning & Romanowicz (2006), obtained using a different treatment of crustal structure, while S362WMANI is the whole-mantle anisotropic variant of S362ANI (from the same study). Inset values (upper-left corner of each panel) represent maximum peak-to-peak variation for each model at the corresponding depth. Circles denote hotspot locations from Steinberger (2000).

## 5 MODEL PERFORMANCE AND UNCERTAINTIES

### 5.1 Performance

#### 5.1.1 Held-out waveform data

One key metric for evaluation of tomographic models is waveform VR for held-out data, that is, waveforms not included in the inversion. We previously used this technique to validate the SEMum2 model by examining SEM-predicted waveforms for held-out events (notably, at shorter periods than used in developing the model, as discussed in the supporting material of French *et al.* 2013). Since upper-mantle structure has not changed significantly from that of SEMum2, we choose to focus primarily on evaluation of deeper (transition-zone and lower-mantle) structure using body wave data. Indeed, as noted above (Section 4.1), our tests using held-out events at the 32-s body wave inversion passband demonstrated that our model progressively improved fits to these data (and was thus unlikely to be overfitting the inversion data set). This was quantified by the number of waveform windows selected by our automated picking procedure, which was applied to the held-out data, but without incorporating it into our inversion data set. For example, we found that  $\sim 100$  new lower-mantle sensitive windows (e.g.  $S_{\text{diff}}$ ) were

**Table 3.** Final-iteration waveform variance reduction for 10 held-out events (i.e. not included in the inversion), defined as in Table 2 and again organized by component and data type/passband.

	L (per cent)	T (per cent)	Z (per cent)
Body (32 s)	52.2	66.6	54.7
Fundamental (60 s)	67.5	67.7	69.8
Overtone (60 s)	69.6	60.6	79.3
Mixed (60 s)	82.0	83.0	86.4

selected from a set of 10 held-out test events when using SEM synthetics computed in our final model relative to the previous inversion iteration.

In Table 3, we summarize waveform VR for these 10 held-out events using SEM synthetics computed in the final-iteration model. We see that body wave VR is quite similar to that reported for the inversion data set in Table 2, albeit with some stochastic variation, as expected given that the held-out set is over an order of magnitude smaller. For held-out surface wave data, we tend to find slightly larger variation in VR relative the values shown in Table 2, though all values remain within 10 per cent of those reported for the inversion set. Importantly, given our focus on lower-mantle structure, we find overall body wave VR for *both* the held-out and inversion data sets

to be nearly the same, at  $\sim 54$  per cent. This observation provides strong evidence that SEMUCB-WM1 does not overfit the inversion data and should indeed generalize well.

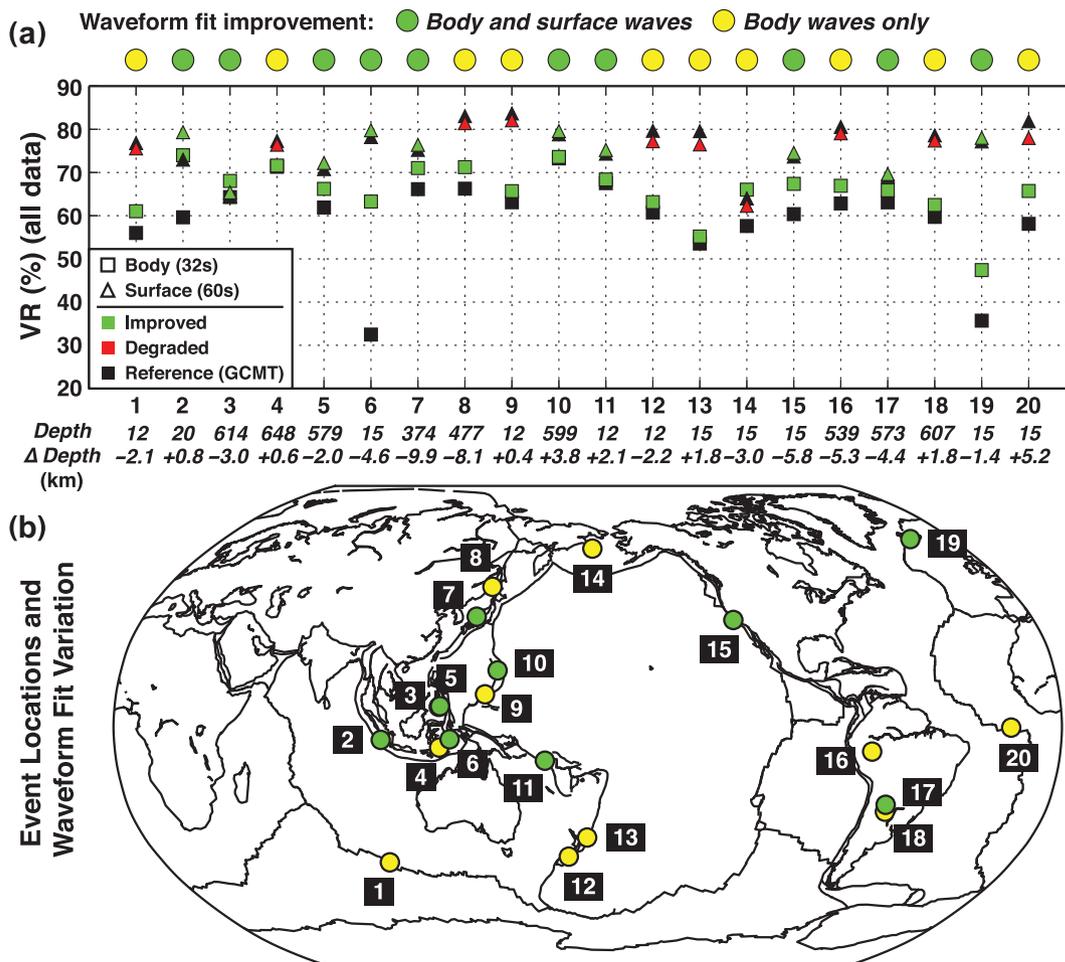
### 5.1.2 Limitations due to source errors

Before beginning our whole-mantle inversion, preliminary tests involving source inversion at the 60-s surface wave passband and using the combined SEMum2 + SAW24B16 starting model to estimate mantle structure, determined that sensitivity to source parameters was quite small. These earlier experiments led only to very slight changes in source-depth and moment-tensor elements (parameters to which our surface wave data set should be most sensitive) relative to parameters published by the Global CMT project and used in our whole-mantle inversion. These findings implied that shorter period modelling (including body waves) would be necessary to yield meaningful inverted sources.

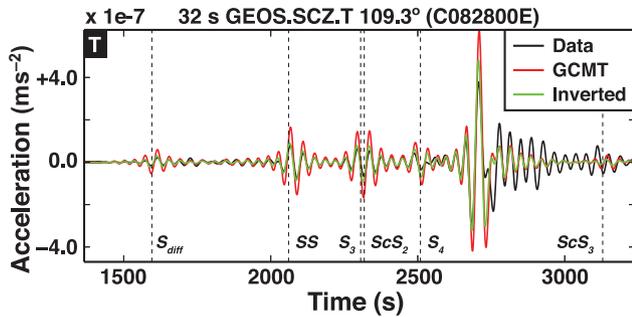
Having inverted for whole-mantle structure, we may now re-examine source inversion while including body wave data, in hopes of: (i) attaining further waveform VR (thereby estimating the degree to which source errors limit our achievable VR) and (ii) retrieving improved estimates of source parameters for use in future studies. Namely, the structural inversion was a necessary first step in order

to achieve sufficient waveform *phase* fits at the 32-s body wave passband, so that we can include these data when inverting for source parameters (where misfits will predominantly manifest in waveform *amplitudes*, especially along paths that are nearly nodal). To this end, we employed an approach similar to that of Liu *et al.* (2004) to compute SEM-based 3-D Green's functions and location partial derivatives in SEMUCB-WM1 and estimate updated sources for 20 inversion events (using both 60-s surface wave and 32-s body wave waveform data). For each event, we performed a single iteration of damped least-squares inversion (in effect, a Gauss–Newton scheme) for source location and moment-tensor elements. Perturbations to the Global CMT moment tensor were additionally required to have no isotropic part. Waveform processing was performed in a manner analogous to the structural inversion, with waveform redundancy now measured for each event individually with respect to source–receiver azimuth.

The results of our experiment are summarized in Fig. 15. We find that all inverted events show improvements in body wave fits relative to their associated Global CMT solutions, while approximately half also show improvements for surface waves. Changes in surface wave fits are in general quite small, particularly in cases where fits degrade slightly (perhaps indicative that the relative weights assigned to body and surface wave data could be further optimized, or that additional source-inversion iterations are needed). Only one



**Figure 15.** Results of the source-inversion experiment (see Section 5.1.2 for details) including 20 events used previously in the structural inversion. Panel (a): waveform variance reduction (VR) for surface wave (triangles) and body wave (squares) data, computed using both the Global CMT (black) and inverted (green, red) source parameters. GCMT source depth and inverted depth perturbation also shown. Panel (b): geographic distribution of inverted sources, classified by waveform-fit improvement (yellow: body wave improvement only; green: improvement for both data types). Classification also appears above Panel (a).



**Figure 16.** Example waveforms for the 2000 August  $M_w$  6.8 Banda Sea event referred to in the text (Section 5.1.2). The clear amplitude misfit between the observed data (black) and SEM synthetics computed using the Global CMT source parameters (red) is easily remedied in a single iteration of the source inversion (green). All SEM synthetics are computed in the SEMUCB-WM1 model.

of the 20 events exhibits changes in body wave VR in excess of 20 percent—the 2000 August  $M_w$  6.8 Banda Sea event featured in Fig. 16, where significant body wave amplitude misfits can be seen for the original Global CMT source. Overall, we found that waveform data in our particular choice of passbands did not require significant lateral perturbations to centroid location, instead leading only to changes in depth and moment-tensor elements. One intriguing result from our experiment is a slight systematic reduction in  $M_w$  seen for all sources inverted using this method (median reduction: 0.05 magnitude units). This observation may indicate that our 1-D Q model, a smoothed version of model QL6 from Durek & Ekström (1996), is not sufficiently attenuating. Namely, if Q is too high, then our SEM simulations will tend to overpredict waveform amplitudes. In order to reconcile these mispredictions with the (comparatively lower amplitude) waveform data, the source inversion may systematically reduce the moment magnitude of the inverted sources. We note that this observation should be revisited if, for example, SEMUCB-WM1 is to be used as a starting model in inversions for anelastic structure. In addition, we do not see systematic changes in waveform VR or centroid depth that appear to correlate with tectonic setting. This observation is somewhat surprising in light of the synthetic inversion tests performed by Hjörleifsdóttir & Ekström (2010), who did see variations in source-parameter uncertainties that correlated to some degree with tectonics. Of course, our sample of earthquakes is considerably smaller than the latter study, such that systematic behaviour may not be apparent.

## 5.2 Model uncertainties

### 5.2.1 Scaling factors

As noted in Section 2.2, the scaling factors that we use to estimate model perturbations to the remaining four parameters of a radially anisotropic medium are only strictly valid in the upper mantle (Montagner & Anderson 1989). While our sensitivity to  $\rho$  is anticipated to be quite small throughout the mantle, and anisotropy is expected to be largely limited to the boundary layers, our sensitivity to Voigt-average isotropic  $V_p$  may be non-negligible (particularly as we include shorter period waveform data). Depth dependence of the scaling factor between  $V_S$  and  $V_p$  is well known, and potentially quite strong in the lowermost mantle (e.g. Su & Dziewonski 1997; Masters *et al.* 2000; Romanowicz 2001), leading some studies to incorporate depth-dependent scaling factors into their inversions (e.g. Ritsema *et al.* 2011).

Motivated by these observations, we sought to assess the effect of ‘realistic’ (i.e. as implied by previous studies) variation in low-ermost mantle  $V_S$  to  $V_p$  perturbation scaling. Namely, we repeated our final-iteration inversion for two different configurations of the  $\delta \ln V_p / \delta \ln V_S$  scaling factor: using a ratio of 0.5 throughout the mantle, except for the bottom  $\sim 300$  km, where we used ratios of either 0.33 or 0.25. We found that the two models obtained were nearly identical to our preferred model, which used 0.5 everywhere (to within a fraction of a per cent  $V_S$  perturbation). A similar result was obtained by Kustowski *et al.* (2008) in their tests of scaling-factor sensitivity. One implication of this result is that our aggregate sensitivity to  $V_p$ , even at 32-s period, remains quite small. In future work incorporating even shorter period waveform data, however, these scaling factors will have to be re-examined in greater detail.

### 5.2.2 Linear resolution analysis

It is well known that the standard resolution analysis employed in many tomographic studies is strictly valid only for linear problems, though such an approach may be warranted for non-linear problems near the global optimum (Tarantola 2005). Furthermore, even in the linear case, counter-intuitive results may be obtained under common implementations of the analysis (Lévêque *et al.* 1993) and care must be taken in assessing model quality with this approach. Fully numerical alternatives to the linear analysis, relying upon second-order adjoint state methods for assessment of posterior uncertainty (Fichtner & Trampert 2011b), have only recently seen use in regional studies (e.g. Fichtner *et al.* 2013) and remain prohibitively costly for application to global-scale inversions. Thus, while acknowledging the above qualifications, the linear analysis still provides one of the best currently available tools for probing certain characteristics of our model that result from data quality/coverage, limitations of the chosen model basis and the influence of *a priori* information.

Under the standard analysis, the resolution operator  $\mathbf{R} = \mathbf{I} - \mathbf{C}'_m \mathbf{C}_m^{-1}$ , with the posterior covariance operator  $\mathbf{C}'_m = (\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \mathbf{C}_m^{-1})^{-1}$ , is applied to a model perturbation  $\mathbf{m}'$ , such as the commonly used checkerboard pattern. The difference between  $\mathbf{m}'$  and  $\mathbf{Rm}'$  in some sense characterizes the quality of model recovery for the problem considered. Intuitively, it is clear that if the posterior state of uncertainty is equal to the prior state, meaning the data supplies no constraint on the model, then  $\mathbf{C}_m = \mathbf{C}'_m$  and  $\mathbf{R} = \mathbf{0}$ , that is, nothing is resolved. Further examination of the expressions above makes clear that the resolution operator depends solely upon: (1) the data and model prior covariance operators; and (2) the model Jacobian  $\mathbf{G}$ . Therefore,  $\mathbf{R}$  reflects only the *a priori* estimates of data noise and model uncertainty/smoothness, as well as spatial variation in data coverage characterized by  $\mathbf{G}$ . Note that any reflection upon the accuracy of the modelling theory is conspicuously absent. The resolution operator may thus be viewed as a spatially variable smoothing operator: a filter applied to a hypothetical model perturbation in a way that gives guidance as to what scales of solution structure *might* be interpretable, and highlights pathological conditions such as strongly non-uniform data coverage (resulting in smearing). Furthermore, the model basis itself also plays an important role in determining whether a hypothetical test structure is recoverable under this analysis—thereby acting as an additional *implicit* filter. Namely, poor fidelity to a chosen test structure upon projection onto the model basis clearly implies that structure of similar sharpness or scale cannot be imaged in either the test or the actual inversion.

We further note that the underlying linearization of the forward problem, inherent in development of the analysis, is not consistent with our inversion methodology. In particular, the approximation:

$$\begin{aligned} \mathbf{d} - \mathbf{g}(\mathbf{m}_k + \mathbf{m}') & \\ & \simeq \mathbf{d} - \mathbf{g}(\mathbf{m}_k) - \mathbf{G}\mathbf{m}' \\ & \simeq -\mathbf{G}\mathbf{m}' \text{ as } \|\mathbf{d} - \mathbf{g}(\mathbf{m}_k)\| \ll \|\mathbf{G}\mathbf{m}'\| \text{ near the solution} \end{aligned} \quad (9)$$

is not compatible with our ‘exact’ spectral-element implementation of  $\mathbf{g}(\cdot)$ . However, SEM-based resolution analysis—in which we directly model  $\mathbf{g}(\mathbf{m}_k + \mathbf{m}')$  for the entire data set—is prohibitively expensive. Thus, while the linear analysis is useful for validating recovery of the morphology of the test structure (i.e. against smearing), the amplitudes of model recovery obtained under the analysis should underpredict those expected for the hybrid scheme.

Keeping these limitations in mind, we perform the standard analysis for a suite of test patterns in order to determine rough bounds on the scales of interpretable structure in SEMUCB-WM1. These estimates may in turn be used to validate the *a priori* correlation lengths introduced in Section 2.3. In Figs 17 and 18, we present checkerboard resolution tests at a range of transition-zone and lower-mantle depths. These test patterns use blocks of azimuthal width (i.e. in longitude) ranging from  $45^\circ$  to  $11.25^\circ$  in both equatorial and polar orientations. The use of multiple orientations is desirable because the often-used equatorial checkerboard breaks down at the poles, leading many authors simply to ignore these regions when supplementary test patterns are not used.

As can be seen in Fig. 17(a), recovery of the morphology of  $V_S$  input structure is quite good, showing no evidence of systematic ray-like smearing or gaps in sensitivity, even at the poles. As expected, amplitude recovery is more modest, with output structure typically at  $\sim 50$  per cent of the input and in some cases even weaker, such as at 2800-km depth for the smallest scale of test structure (where the maximum anomaly width is  $< 700$  km). We will discuss the implications of these tests, in the context of placing approximate *lower* bounds on the scales of well-resolved  $V_S$  structure in SEMUCB-WM1, later in this section. In the meantime, however, we would like to raise an important point regarding the recovery of long-wavelength structure at the base of the mantle. Namely, at 2800-km depth, the largest scale of input structure in Fig. 17(a) ( $45^\circ$  azimuthal-width blocks) is clearly recovered as *continuous* anomalies, with smaller scale variations similar to those seen at the base of the mantle in SEMUCB-WM1 (Fig. 8) markedly absent. Indeed, if similar variations were retrieved in this particular resolution test, they could only have been attributable to uneven sensitivity. This observation provides further evidence that the smaller scale, concentrated low-velocity structures embedded within the LLSVPs in SEMUCB-WM1, previously discussed Section 4.2.1, are not an artefact of our inversion. In Fig. 17(b), we show cross-contamination with  $\xi$  structure at the four greatest depths and the largest scale of input structure shown in (a) (where trade-offs between the two is expected to be strongest due to both data coverage and model parametrization). Here, we find that cross-contamination of  $\xi$  from  $V_S$  is quite small, except at the base of the mantle, where structure is primarily constrained by  $S_{\text{diff}}$  and multiple-ScS phases, both of which are more prominent on the transverse component (potentially leading to an imbalance between  $V_{SH}$  and  $V_{SV}$  sensitivity).

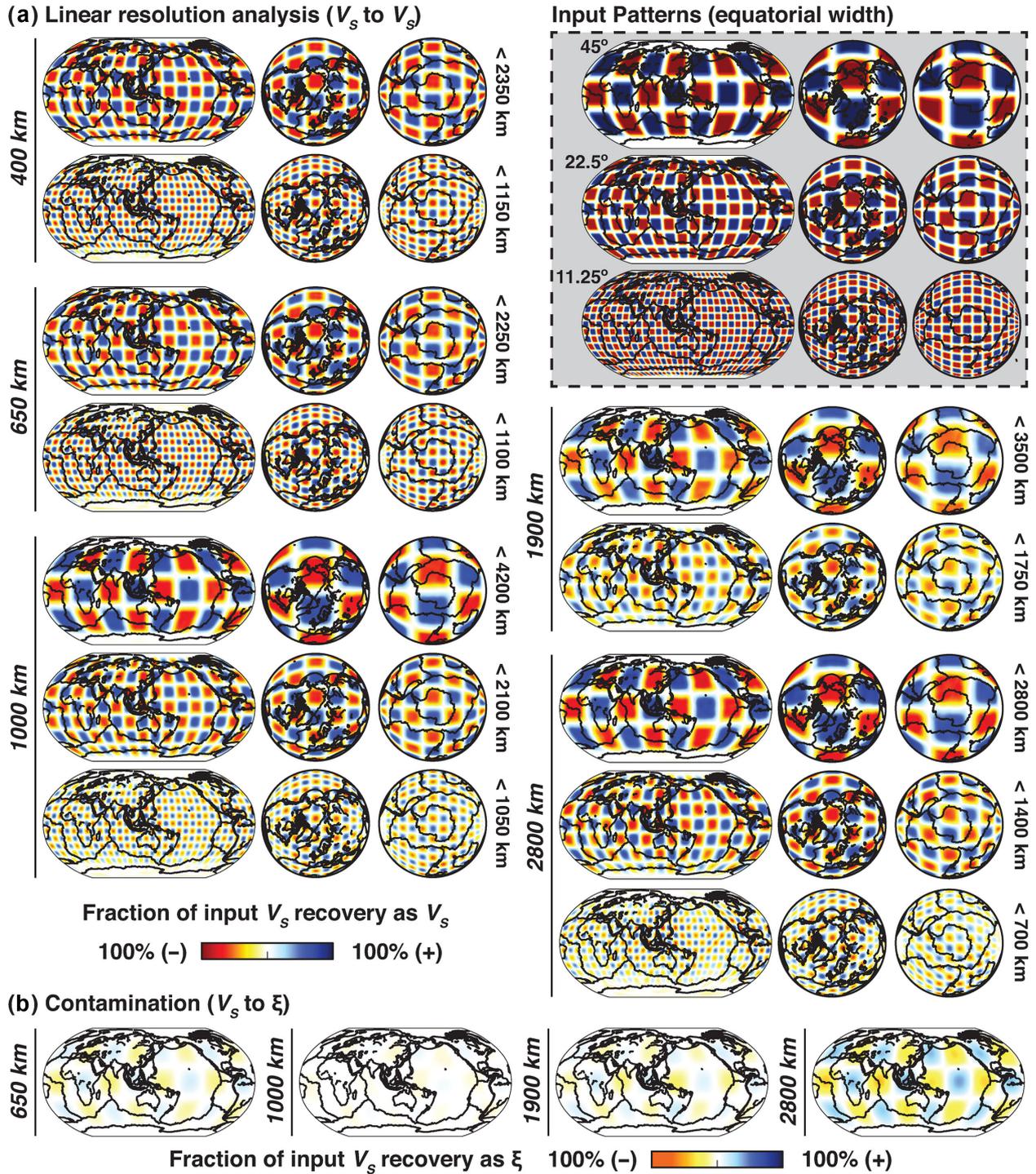
Turning our attention to Fig. 18(a), it can be seen that model recovery for  $\xi$  is less impressive, and we are unable to satisfactorily recover the  $22.5^\circ$  test pattern by mid-mantle depths (where maximum block widths are on the order of 2000 km). The morphology of the largest scale test pattern ( $45^\circ$ ) is recovered fairly well throughout

the mantle, including the mid- and lower mantle where maximum block widths vary from  $< 3000$  to  $\sim 4000$  km, though amplitude recovery is much weaker than seen in the  $V_S$  analysis. In Fig. 18(b), we see that  $\xi$  to  $V_S$  cross-contamination is characterized by similar depth variation as seen in the  $V_S$  to  $\xi$  case (namely, stronger at base of mantle), and again possibly due to differential sensitivity between  $V_{SH}$  and  $V_{SV}$ . In addition,  $\xi$  to  $V_S$  contamination appears weaker than for  $V_S$  to  $\xi$ , which is consistent with the expectation that  $V_S$  is better constrained by the data.

Based on these results, it appears that  $V_S$  structure in SEMUCB-WM1 is well resolved down to  $\sim 1000$ -km scales in the transition zone and into the upper portion of the mid-mantle. This bound then increases to at most  $\sim 1800$ -km scales (and possibly less) in the remainder of the mid-mantle. At the base of the mantle, where sensitivity again improves, the lower bound on resolved scales decreases again: structure at  $\leq 1400$ -km scales is clearly quite well resolved, while recovery of structure at  $\leq 700$ -km scales is questionable. These observations bracket the corresponding bound on resolved structure, which must fall somewhere between—perhaps on the order of 1000–1200 km (as a conservative estimate). For  $\xi$  structure, we find that large-scale variation ( $\sim 4500$  km) in the transition zone is clearly well constrained, with structure down to  $\sim 2400$ -km scales likely interpretable. In the mid-mantle, we find that  $\xi$  structure down to  $\sim 4200$ -km scales is clearly robust at 1000-km depth, though the lower bound is likely much less (though larger than the shallower  $> 2400$ -km estimate), increasing to  $\sim 3500$ -km scales deeper in the mid-mantle (where amplitude recovery is quite low) and finally decreasing again to  $< 3000$  km at the base of mantle. Returning to the discussion of *a priori* correlation lengths in Section 2.3, we find that these scales of resolved structure are broadly compatible with the latter. In particular, we expect these (data sensitivity) adaptive correlation-length values to range from 800 to 1200 km for  $V_S$  and 2400 to 3600 km for  $\xi$  at transition zone and greater depths, with significant correlation at up to twice these lengths given the form of eq. (3). Thus, in some sense, our resolution analysis validates *a posteriori* our earlier choice of correlation lengths, and is likely indicative that our selected prior is reasonable given the resolving power of our data set. Namely, while we cannot rule out that our attainable resolution may in fact be better than is permitted by our prior in some regions, the chosen correlation lengths impose an appropriate lower bound on the permissible resolution: one that is sufficiently conservative such that we do not observe ray-like streaks or gaps in sensitivity for the scales of structure reported above. Of course, these observations are accompanied by the caveats enumerated at the beginning of this section regarding the limitations of the linear analysis.

### 5.2.3 Resampling techniques

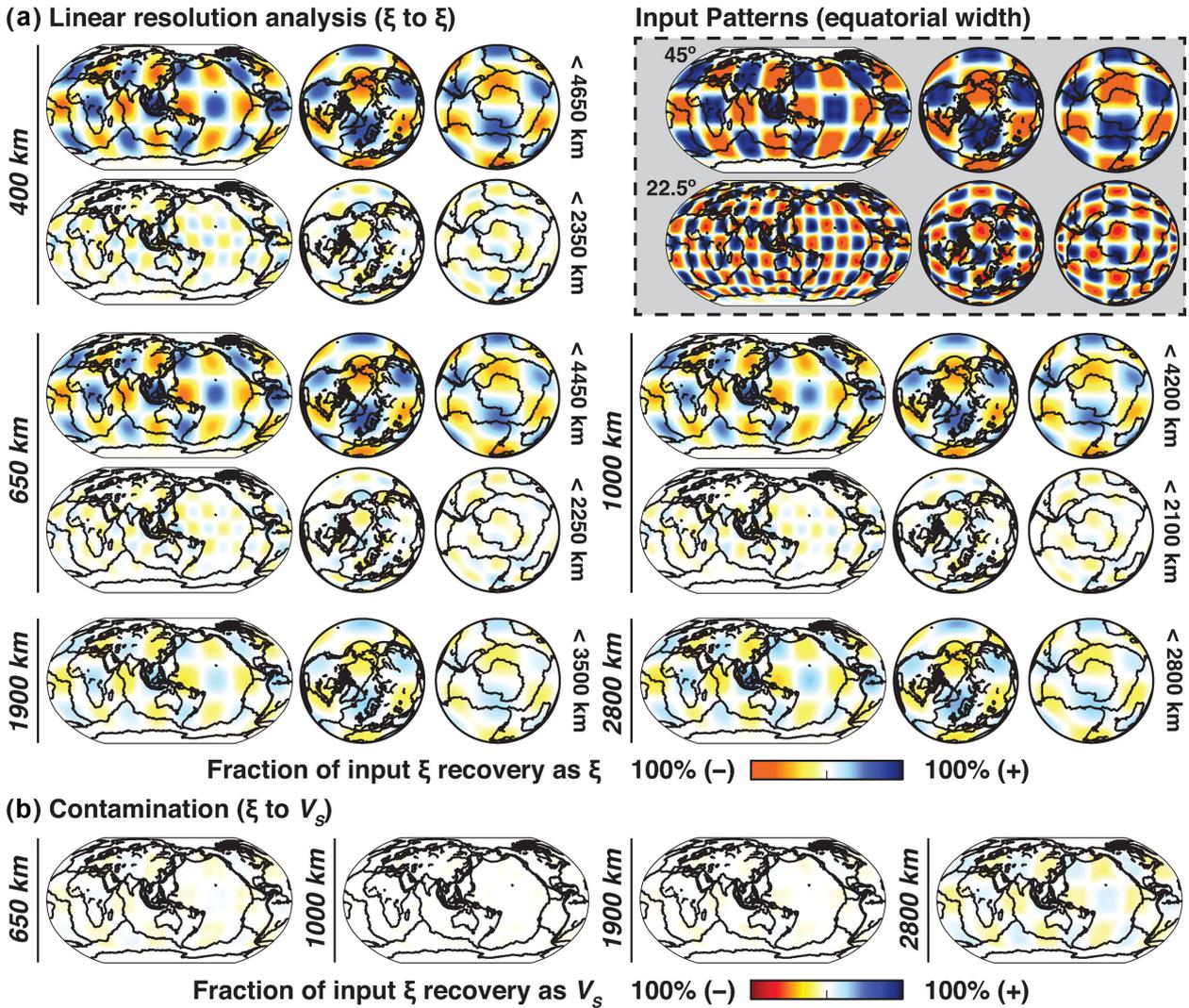
Another way to probe uncertainty in the resulting model is through statistical resampling techniques, such as the deleted jackknife (e.g. Efron & Stein 1981) or bootstrap (e.g. Efron & Tibshirani 1991). For example, we found bootstrap resampling to be a valuable tool for uncertainty estimation in our supplementary analysis of the SEMUM2 model (French *et al.* 2013). Here, we take a similar resampling approach, based on repeating our final inversion iteration for many different realizations of our waveform data set (again owing to the large changes to the model that are possible in a single step of the Gauss–Newton scheme, should the data warrant it). Previously, Panning & Romanowicz (2006) used both the bootstrap and jackknife techniques in the analysis of their global model SAW642AN, and found that the jackknife rapidly converged to the same uncertainty



**Figure 17.** Linear resolution analysis for global  $V_s$  structure, using a range of scales and orientations of input models (see Section 5.2.2) centred at 400, 650, 1000, 1900 and 2800-km depths (labelled along the left-hand margin of each panel). Input models are shown in the grey-shaded panel appearing in the upper right-hand corner, along with the associated equatorial block widths ( $^\circ$ ). For each depth, the corresponding maximum block width of each scale of input structure is shown (converted to km) along the right-hand margin of the panel. See Section 5.2.2 for detailed discussion. Panel (a): recovery of  $V_s$  input model structure as  $V_s$ . Panel (b): recovery of long-wavelength  $V_s$  input structure as  $\xi$  (i.e. cross-contamination).

estimate as the bootstrap, but required significantly fewer samples. This latter point is an advantageous property, given that each sample requires a significant amount of I/O (reading the Gauss–Newton Hessian) and subsequent factorization, and drove us specifically to consider the jackknife technique.

Here, we follow a similar approach to that in Panning & Romanowicz (2006). Namely, we separate our waveform data set into 12 bins by month of earthquake occurrence, and use these bins as aggregate data (i.e. our data set contains 12 observations). We then, in turn, use the deleted jackknife technique to derive standard-error



**Figure 18.** Linear resolution analysis for global  $\xi$  structure, using a range of scales and orientations of input model (again appearing in the upper right-hand grey-shaded panel). Here, we use a layout and labelling scheme similar to that in Fig. 17, although we have omitted the shortest wavelengths of test model appearing in the latter due to the limits of our  $\xi$  parametrization. Again, see Section 5.2.2 for detailed discussion. Panel (a): recovery of  $\xi$  input model structure as  $\xi$ . Panel (b): recovery of long-wavelength  $\xi$  input structure as  $V_s$ .

estimates  $\hat{\sigma}_{\text{jackknife}}$  for our model via

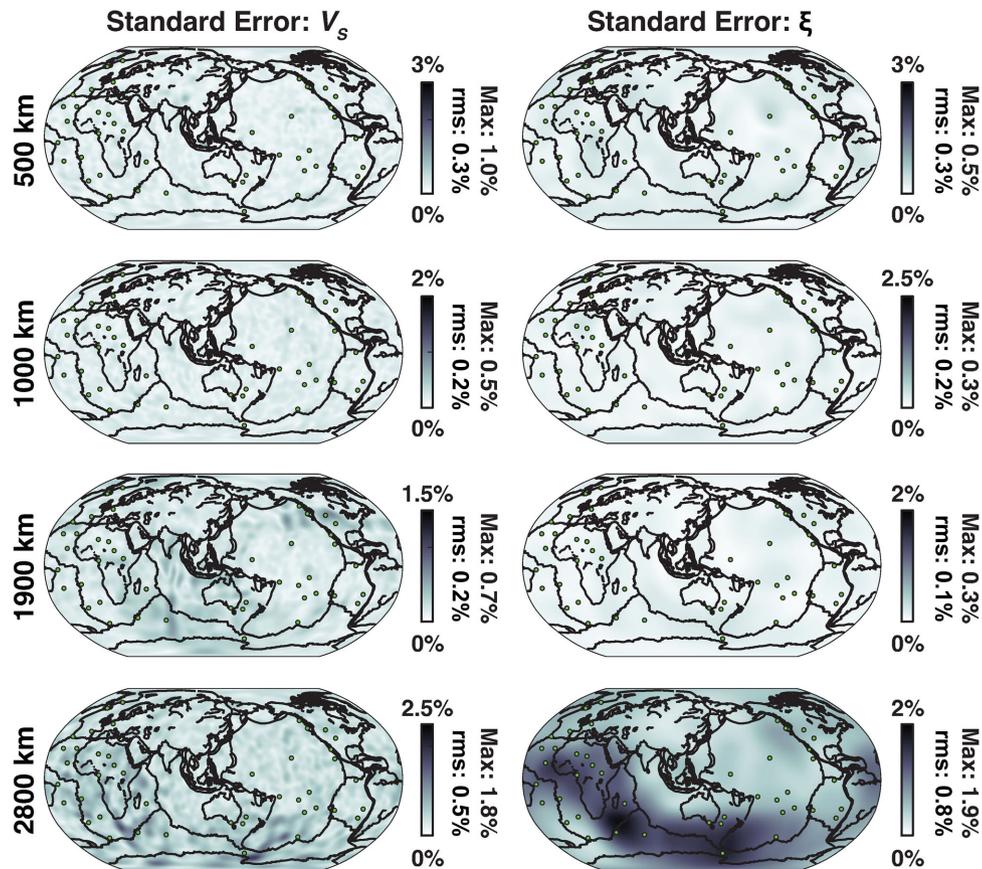
$$\hat{\sigma}_{\text{jackknife}} = \sqrt{\frac{n-d}{d \cdot C} \sum_{i=1}^C [\theta_{(i)} - \hat{\theta}]^2},$$

$$\text{where } \hat{\theta} = \frac{1}{C} \sum_{i=1}^C \theta_{(i)} \text{ and } C = \frac{n!}{d!(n-d)!}, \quad (10)$$

where  $\theta_{(i)}$  is the model estimate produced by the  $i$ th deleted jackknife realization of our data set, the number of data  $n$  is 12 and the number of deletions  $d$  per data set realization is 1 or 2. For the  $d = 2$  case, this approach requires only 66 solves (each producing a model sample), while the month-based binning scheme requires pre-computation and storage of only 12 Hessian matrices which may be read in different combinations at solution time (occupying approximately 2.2 TB if stored in full form, chiefly to simplify collective I/O prior to factorization or 1.1 TB if storing only the upper triangular part). We further note that the uncertainty estimates recovered in this manner will be *conservative overestimates* thereof,

in that we do not reoptimize the waveform-window weights for each jackknife sample (which our inversion does in practice prior to solution, allowing us to account to some extent for uneven data coverage).

In Fig. 19, we show standard-error estimates obtained at four transition-zone and lower-mantle depths with  $d = 2$  (the  $d = 1$  case is very similar) for both  $V_s$  and  $\xi$  in map view, along with rms values thereof at each depth. We find that isotropic  $V_s$  error estimates are in general quite low, though slightly elevated at the base of the mantle (2800-km depth). Furthermore, while error estimates appear geographically unbiased at most depths, they are slightly elevated in the Southern Hemisphere at 2800 km, though they remain quite small relative to the corresponding amplitudes of model structure seen in Fig. 8 (which are plotted at the same saturation level). This distribution is not surprising, as the acquisition geometry imposed by the natural distribution of seismic sources, along with the geographic bias of receivers towards land masses in the Northern Hemisphere, makes the Southern Hemisphere notoriously difficult to illuminate seismically. Turning our attention to  $\xi$  structure, we again find standard-error estimates that are quite low relative to the



**Figure 19.** Standard-error estimates  $\hat{\sigma}_{\text{jackknife}}$  for  $V_S$  and  $\xi$  structure in SEMUCB-WM1 derived using a jackknife resampling analysis (see Section 5.2.3), plotted at a range of transition-zone and lower-mantle depths. Estimates are plotted in model units of per cent perturbation with respect to the 1-D reference model (i.e. not per cent variation relative to the model SEMUCB-WM1), and saturation levels match those used in Figs 8 and 14. Maximum and rms standard-error estimate values listed for both parameters at each depth.

model amplitudes seen in Fig. 14, though also with elevated estimates at the base of the mantle. Indeed, we see a more pronounced geographic bias in the  $\xi$  error estimates at 2800-km depth than seen for  $V_S$ , with a pronounced band extending from the Southwest Pacific to South Africa. Importantly, however, we note that this pattern does not correlate with the lateral distribution of  $\xi$  structure seen at 2800 km in Fig. 14—thereby giving us some degree of confidence that the imaged structure is realistic.

Finally, much like the linear resolution analysis, we note that these error estimates also suffer from the limitation that they do not reflect the accuracy of the underlying theoretical formalism used for inversion. Instead, resampling analyses are valuable for assessing sensitivity to random errors in the data and suboptimal illumination of the Earth's interior. In addition, as also noted by Panning & Romanowicz (2006), these analyses help to characterize the inherent trade-off in the inversion between theoretical resolution and error, which is in turn controlled by the role of prior information (Section 2.3). Thus, resampling techniques provide an assessment of model quality that is complementary to the resolution estimates presented above in Section 5.2.2.

## 6 CONCLUSIONS

Here, we presented the model SEMUCB-WM1, the first radially anisotropic shear wave velocity model of the Earth's whole mantle derived from SEM-based forward waveform modelling. Examining

the model, we found that long-wavelength  $V_S$  variations throughout the mantle correlate quite well with other recent global studies (Kustowski *et al.* 2008; Panning *et al.* 2010; Ritsema *et al.* 2011), particularly in the upper- and lowermost mantle where data coverage is invariably better, while shorter wavelength structure tends to differ more clearly across models.  $V_S$  structure in the transition zone and upper mantle remains highly correlated at all scales with that of SEMum2 (our starting model in this depth range, derived using a similar inversion technique). At the same time, mid and lowermost mantle structures differ considerably from SAW24B16 (our lower-mantle starting model), sharpening significantly and revealing more concentrated fast anomalies associated with subducted slabs, as well as narrow low-velocity anomalies approximately collocated with major hotspots (Section 4.2.1). The strongest *anisotropic* structure in SEMUCB-WM1 is found in the uppermost and lowermost mantle, boundary layers where lateral shear is typically expected to be strongest, while weak anisotropy is seen at mid-mantle depths in agreement with previous studies (e.g. Panning & Romanowicz 2006). The common pattern of strong  $\xi > 1$  in the uppermost mantle beneath the ocean basins and  $\xi < 1$  beneath the mid-ocean ridge system is observed in our model, though there also exist long-wavelength variations in  $\xi$  structure at these depths that clearly differ in detail from previous studies (Section 4.2.2). Furthermore, we observe a pattern of  $\xi$  structure with a strong degree-2 component in the lowermost mantle—the morphology of which does not appear correlated with the associated uncertainty estimates (Section 5.2.3), though we note that  $\xi$  may trade off with  $V_S$  structure to some

extent in this depth range (Section 5.2.2). Detailed analysis of certain aspects of the SEMUCB-WM1 model, including low-velocity structures in the lower mantle and their relationships with features previously imaged in SEMum2, will appear in a forthcoming publication.

Our approach in this study has been a progressive one: starting from our earlier SEMum2 model and iteratively increasing both our data set size and the complexity of our modelling—allowing us to use available computational resources wisely and learn something about the Earth in each phase of the inversion. This consideration motivated us to carefully select the frequency bands used in our work: first to only 60-s period for upper-mantle and transition-zone imaging, and later down to 32 s for our whole-mantle model. In this vein, the logical next step in the continuation of this work is extension to shorter period modelling. Not only will higher frequency seismic data enable future inversions to attain higher resolution, it will also allow them to constrain  $V_p$  structure. Importantly, the ability to invert independently for  $V_p$  may enable inferences regarding the presence of compositional heterogeneity (based upon the manner in which  $V_p$  and  $V_s$  covary, or fail to). This will further alleviate some of the theoretical shortcomings of using a scaling-based approach to account for  $V_p$  sensitivity (although, our experiments did not indicate they should have a significant impact at 32 s; see Section 5.2.1). In addition, we are confident that the infrastructure on which the hybrid inversion is based—that which enables the quickly converging Gauss–Newton model optimization scheme—should readily scale to the larger problem sizes necessary to support higher resolution modelling. Finally, in conjunction with the structural inversion, further investigation of inversion for seismic source parameters will become necessary as future work extends to shorter periods. Based on the source-inversion experiments presented here (Section 5.1.2), we are confident that the 3-D Green's function-based approach adapted from Liu *et al.* (2004) could be successfully applied in this context.

## ACKNOWLEDGEMENTS

The authors would like to thank Andreas Fichtner and one anonymous reviewer for their constructive comments on the manuscript, from which the latter benefitted greatly. The authors would also like to thank Vedran Lekić for valuable discussions and Huaiyu Yuan for assistance in collecting the supplementary waveform data set. In addition, the authors would like to thank Masayuki Obayashi for kindly providing access to the GAP-P4 model in electronic form. The authors acknowledge support from the National Science Foundation (grant EAR-0738284) and SWF acknowledges additional support from the NSF Graduate Research Fellowship Programme. All waveform data used in this study are available from the Incorporated Research Institutions for Seismology (<http://www.iris.edu>). Computations were performed at the National Energy Research Scientific Computing Center, supported by the US Department of Energy Office of Science (Contract No. DE-AC02-05CH11231).

## REFERENCES

- Backus, G., 1962. Long-wave elastic anisotropy produced by horizontal layering, *J. geophys. Res.*, **67**(11), 4427–4440.
- Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in North America, *EOS, Trans. Am. geophys. Un.*, **81**, 1351–1375.
- Becker, T. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, **3**(1), 1351–1375.
- Becker, T., Kustowski, B. & Ekström, G., 2008. Radial seismic anisotropy as a constraint for upper mantle rheology, *Earth planet. Sci. Lett.*, **267**(1), 213–227.
- Blackford, L.S. *et al.*, 1997. *ScaLAPACK Users' Guide*, Society for Industrial and Applied Mathematics.
- Brocher, T., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust, *Bull. seism. Soc. Am.*, **95**, 2081–2092.
- Brossier, R., Operto, S. & Virieux, J., 2009. Seismic imaging of complex on-shore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.
- Capdeville, Y. & Marigo, J.J., 2007. Second order homogenization of the elastic wave equation for non-periodic layered media, *Geophys. J. Int.*, **170**, 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.
- Chaljub, E., Capdeville, Y. & Vilotte, J.P., 2003. Solving elastodynamics in a fluid-solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids, *J. Comp. Phys.*, **187**, 457–491.
- Davaille, A., 1999. Simultaneous generation of hotspots and superswells by convection in a heterogeneous planetary mantle, *Nature*, **402**(6763), 756–760.
- Davaille, A., Stutzmann, E., Silveira, G., Besse, J. & Courtillot, V., 2005. Convective patterns under the Indo-Atlantic 'box', *Earth planet. Sci. Lett.*, **239**, 233–252.
- Debayle, E. & Ricard, Y., 2012. A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements, *J. geophys. Res.*, **117**(B10), doi:10.1029/2012JB009288.
- Durek, J.J. & Ekström, G., 1996. A radial model of anelasticity consistent with long-period surface-wave attenuation, *Bull. seism. Soc. Am.*, **86**, 144–158.
- Dziewonski, A., 2005. The robust aspects of global seismic tomography, in *Plates, Plumes and Paradigms*, Geol. Soc. Am. Spec. Pub. 388, pp. 147–154, eds Foulger, G.R., Natland, J.H., Presnall, D.C. & Anderson, D.L., Geological Society America, Boulder, CO.
- Dziewonski, A. & Anderson, D., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**, 297–356.
- Dziewonski, A., Hager, B. & O'Connell, R., 1977. Large-scale heterogeneities in the lower mantle, *J. geophys. Res.*, **82**, 239–255.
- Dziewonski, A.M., Lekić, V. & Romanowicz, B.A., 2010. Mantle anchor structure: an argument for bottom up tectonics, *Earth planet. Sci. Lett.*, **299**(1), 69–79.
- Efron, B. & Stein, C., 1981. The jack knife estimate of variance, *Ann. Stat.*, **9**(3), 586–596.
- Efron, B. & Tibshirani, R.J., 1991. *An Introduction to Bootstrap*, Chapman and Hall.
- Epanomeritakis, I., Akcelik, V., Ghattas, O. & Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, *Inverse Probl.*, **24**, 034015, doi:10.1088/0266-5611/24/3/034015.
- 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**, 519–533.
- Fichtner, A. & Trampert, J., 2011a. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.*, **185**, 775–798.
- Fichtner, A. & Trampert, J., 2011b. Resolution analysis in full waveform inversion, *Geophys. J. Int.*, **187**, 1604–1624.
- 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**, 1703–1725.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villaseñor, A., 2013. Multiscale full waveform inversion, *Geophys. J. Int.*, **194**(1), 534–556.
- French, S., Lekić, V. & Romanowicz, B., 2013. Waveform tomography reveals channeled flow at the base of the oceanic asthenosphere, *Science*, **342**, 227–230.

- Fukao, Y. & Obayashi, M., 2013. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity, *J. geophys. Res.*, **118**(11), 5920–5938.
- Hjörleifsdóttir, V. & Ekström, G., 2010. Effects of three-dimensional Earth structure on CMT earthquake parameters, *Phys. Earth planet. Inter.*, **179**(3), 178–190.
- 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**, 195–212.
- Husson, L. & Conrad, C., 2012. On the location of hotspots in the framework of mantle convection, *Geophys. Res. Lett.*, **39**(17), doi:10.1029/2012GL052866.
- Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.*, **149**, 390–412.
- Komatitsch, D. & Tromp, J., 2002b. 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.P., 1998. The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.*, **88**, 368–392.
- Korenaga, J. & Jordan, T., 2004. Physics of multiscale convection in earth's mantle: evolution of sublithospheric convection, *J. geophys. Res.*, **109**(B01), doi:10.1029/2003JB002464.
- Kustowski, B., Ekström, G. & Dziewonski, A., 2008. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res.*, **113**, 303–318.
- Lekić, V. & Romanowicz, B., 2011a. Inferring upper-mantle structure by full waveform tomography with the spectral element method, *Geophys. J. Int.*, **185**, 799–831.
- Lekić, V. & Romanowicz, B., 2011b. Tectonic regionalization without a priori information: a cluster analysis of upper mantle tomography, *Earth planet. Sci. Lett.*, **308**, 151–160.
- Lekić, V., Panning, M. & Romanowicz, B., 2010. A simple method for improving crustal corrections in waveform tomography, *Geophys. J. Int.*, **182**, 265–278.
- Lekić, V., Cottaar, S., Dziewonski, A. & Romanowicz, B., 2012. Cluster analysis of global lower mantle heterogeneity: a new class of structure and implications for chemical heterogeneity, *Earth planet. Sci. Lett.*, **182**, 265–278.
- Lévéque, J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**, 313–318.
- Li, X. & Romanowicz, B., 1995. Comparison of global waveform inversions with and without considering cross-branch modal coupling, *Geophys. J. Int.*, **121**, 695–709.
- Li, X. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.*, **101**, 22 245–22 272.
- Liu, Q., Polet, J., Komatitsch, D. & Tromp, J., 2004. Spectral-element moment tensor inversions for earthquakes in southern California, *Bull. seism. Soc. Am.*, **94**(5), 1748–1761.
- Marone, F. & Romanowicz, B., 2007. Non-linear crustal corrections in high-resolution regional waveform seismic tomography, *Geophys. J. Int.*, **170**, 460–467.
- Masters, G., Laske, G., Bolton, H. & Dziewonski, A., 2000. The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*, pp. 63–87, eds Karato, S., Forte, A., Liebermann, R., Masters, G. & Stixrude, L., American Geophysical Union.
- McNamara, A. & Zhong, S., 2005. Thermochemical structures beneath Africa and the Pacific Ocean, *Nature*, **437**(7062), 1136–1139.
- McNamara, A.K., van Keken, P.E. & Karato, S.I., 2002. Development of anisotropic structure in the Earth's lower mantle by solid-state convection, *Nature*, **416**(6878), 310–314.
- McNutt, M.K. & Fischer, K.M., 1987. The South Pacific superswell, in *Islands and Atolls*, Geophys. Monogr., pp. 25–34, ed. Keating, B.H., AGU.
- Mégnin, C. & Romanowicz, B., 1999. The effects of the theoretical formalism and data selection on mantle models derived from waveform tomography, *Geophys. J. Int.*, **138**, 366–380.
- 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.
- Montagner, J. & Anderson, D., 1989. Petrological constraints on seismic anisotropy, *Phys. Earth planet. Inter.*, **54**, 82–105.
- Murakami, M., Hirose, K., Kawamura, K., Sata, N. & Ohishi, Y., 2004. Post-perovskite phase transition in MgSiO<sub>3</sub>, *Science*, **304**(5672), 855–858.
- Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage, *Math. Comput.*, **35**(151), 773–782.
- Nolet, G. & Dahlen, F., 2000. Wave front healing and the evolution of seismic delay times, *J. geophys. Res.*, **105**, 19 043–19 054.
- Obayashi, M., Yoshimitsu, J., Nolet, G., Fukao, Y., Shiobara, H., Sugioka, H., Miyamachi, H. & Gao, Y., 2013. Finite frequency whole mantle p wave tomography: improvement of subducted slab images, *Geophys. Res. Lett.*, **40**(21), 5652–5657.
- 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., Lekić, V. & Romanowicz, B., 2010. Importance of crustal corrections in the development of a new global model of radial anisotropy, *J. geophys. Res.*, **115**, 361–379.
- Phipps-Morgan, J., Morgan, W.J., Zhang, Y.S. & Smith, W.H.F., 1995. Observational hints for a plume-fed, suboceanic asthenosphere and its role in mantle convection, *J. geophys. Res.*, **100**(B7), 12 753–12 767.
- Pratt, R.G., Shin, C. & Hick, G., 1998. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.*, **133**(2), 341–362.
- Richards, M. & Engebretson, D., 1992. Large-scale mantle convection and the history of subduction, *Nature*, **355**, 437–440.
- Ritsema, J., Deuss, A., van Heijst, H. & Woodhouse, J., 2011. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements, *Geophys. J. Int.*, **184**, 1223–1236.
- Ritzwoller, M., Shapiro, N., Barmin, M. & Levshin, A., 2002. Global surface wave diffraction tomography, *J. geophys. Res.*, **107**(B12), ESE 4-1–4-13.
- Robertson, G. & Woodhouse, J., 1996. Ratio of relative S to P velocity heterogeneity in the lower mantle, *J. geophys. Res.*, **101**(B9), 20 041–20 052.
- 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**, 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, *Geophys. J. Int.*, **90**(1), 75–100.
- Romanowicz, B., 2001. Can we resolve 3D density heterogeneity in the lower mantle? *Geophys. Res. Lett.*, **28**(6), 1107–1110.
- 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**, 520–536.
- Schaeffer, A. & Lebedev, S., 2013. Global shear speed structure of the upper mantle and transition zone, *Geophys. J. Int.*, **194**(1), 417–449.
- Seriani, G. & Oliveira, S., 2008. Dispersion analysis of spectral element methods for elastic wave propagation, *Wave Motion*, **45**(6), 729–744.
- Shapiro, N. & Ritzwoller, M., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.*, **151**, 88–105.
- Simmons, N., Forte, A. & Grand, S., 2006. Constraining mantle flow with seismic and geodynamic data: a joint approach, *Earth planet. Sci. Lett.*, **246**, 109–124.
- Simmons, N., Forte, A., Boschi, L. & Grand, S., 2010. GyPSuM: a joint tomographic model of mantle density and seismic wave speeds, *J. geophys. Res.*, **115**, 109–124.
- Steinberger, B., 2000. Plumes in a convecting mantle: models and observations for individual hotspots, *J. geophys. Res.*, **105**, 11 127–11 152.

- Su, W.J. & Dziewonski, A., 1997. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth. planet. Inter.*, **100**(1), 135–156.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**, 1259–1266.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Society for Industrial and Applied Mathematics.
- Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.*, **20**, 219–232.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana–doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Tromp, J., Komatitsch, D. & Liu, Q., 2008. Spectral-element and adjoint methods in seismology, *Commun. Comput. Phys.*, **3**, 1–32.
- van der Hilst, R., Widiyantoro, S. & Engdahl, E., 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**(6625), 578–584.
- Wang, Z. & Dahlen, F.A., 1995. Spherical-spline parameterization of three-dimensional Earth models, *Geophys. Res. Lett.*, **22**, 3099–3102.
- Woodhouse, J., 1980. The coupling and attenuation of nearly resonant multiplets in the Earth’s free oscillation spectrum, *Geophys. J. R. astr. Soc.*, **61**, 261–283.
- Woodhouse, J. & Dziewonski, A., 1984. Mapping the upper mantle: three dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.*, **89**, 5953–5986.
- Zheng, Y., Kamil, A., Driscoll, M., Shan, H. & Yelick, K., 2014. UPC++: a PGAS extension for C++, in *Proceedings of 28th IEEE International Parallel and Distributed Processing Symposium (IPDPS)*.

## APPENDIX: CRUSTAL STRUCTURE IN NACT

### A1 Background: non-linear asymptotic coupling theory (NACT)-based modelling

NACT is a method for waveform and sensitivity kernel calculation using normal-mode perturbation theory, and we refer the reader to Li & Romanowicz (1995) for a detailed introduction. Under NACT, along-branch mode coupling is treated in a non-linear fashion, thus including the effect of multiple forward scattering, while cross-branch coupling allows for accurate modelling of overtone surface waves (Romanowicz *et al.* 2008). To avoid coupling calculations over the whole sphere, sensitivity kernels are limited to the 2-D plane containing the source and receiver. Volumetric model perturbations

$$\mathbf{f}(\mathbf{m}; r, \theta, \phi) = \sum_p \sum_q \mathbf{m}_{pq} \beta_p(\theta, \phi) \nu_q(r), \quad (\text{A1})$$

where  $\beta$  and  $\nu$  represent the spherical-spline and b-spline model basis functions, respectively (Section 2.2), and perturbations to discontinuity topography  $\{h_d: d = 1, \dots, N_d\}$ , enter into the NACT-predicted accelerogram  $u(t)$  (and sensitivity kernels  $\partial u / \partial \mathbf{m}$ ) through source–receiver minor-arc and great-circle path integrations over the local frequency shift  $\delta\omega_{kk'}(\theta, \phi)$  (Woodhouse 1980). This quantity corresponds to the shift in average degenerate eigenfrequency of the multiplet pair  $(k, k')$  due to coupling induced by structural heterogeneity beneath point  $(\theta, \phi)$ :

$$\delta\omega_{kk'}(\theta, \phi) = \frac{1}{\omega_k + \omega_{k'}} \int_0^a \mathbf{f}(\mathbf{m}; r, \theta, \phi) \cdot \mathbf{M}_{kk'}(r) r^2 dr - \sum_d^{N_d} r_d^2 h_d(\theta, \phi) H_{kk'}^d, \quad (\text{A2})$$

where  $a$  is Earth radius,  $r_d$  is the unperturbed radius of discontinuity  $d$  and  $\mathbf{M}_{kk'}(r)$  and  $H_{kk'}^d$  are eigenfrequency partial derivatives with respect to volumetric and discontinuity-topography perturbations (e.g. Romanowicz 1987).

Given the (linearized) form of eq. (A2), two particular concerns arise regarding treatment of the smooth crustal layer (Section 2.4): (1) fidelity to the true non-linear response of  $\delta\omega_{kk'}$  to crustal thickness variation; and (2) recovery of sensitivity kernels reflecting the split parametrization between crust and mantle—that is,  $\mathbf{f}(\mathbf{m}; r, \theta, \phi)$  is only strictly defined at mantle depths. Point (2) may easily be treated by ensuring that the upper limit of integration on the radial inner product in eq. (A2) honours the local Moho depth  $h_M(\theta, \phi)$ , that is,

$$\int_0^{h_M(\theta, \phi)} \mathbf{f}(\mathbf{m}; r, \theta, \phi) \cdot \mathbf{M}_{kk'}(r) r^2 dr. \quad (\text{A3})$$

However, treatment of point (1) is more involved, and we now discuss our solution at greater length.

### A2 Non-regionalized modified linear crustal corrections (MLCs)

We adopt an extended implementation of the MLC scheme of Lekić *et al.* (2010), extended in that we do not limit ourselves to a small set of ‘regionalized’ crustal models. Instead, we obtain continuously parametrized corrections, directly reflecting lateral variation of the crustal model.

Given a 1-D reference earth model  $m_{\text{ref}}$  and structure associated with each crustal-model node  $\{\mathbf{c}_n: n = 1, \dots, N_k\}$ , we calculate normal-mode eigenfrequencies for both  $m_{\text{ref}}$  itself and a set of  $N_k$  composite 1-D models, denoted  $m_{\text{ref}}^n$ , that combine crustal structure from  $\mathbf{c}_n$  with mantle structure from  $m_{\text{ref}}$ . For a given normal-mode multiplet  $k$ , we refer to

$$\delta\omega_k^n = \omega_k [m_{\text{ref}}^n] - \omega_k [m_{\text{ref}}] \quad (\text{A4})$$

as the *non-linear* local eigenfrequency shift induced by crustal structure at node  $n$ —non-linear because it is recovered through direct evaluation of eigenfrequency dispersion in the perturbed model, not through the linearized approach in eq. (A2) above. Furthermore, given  $m_{\text{ref}}$ , we calculate eigenfrequency partial derivatives due to self-coupling ( $k = k'$ ) for perturbations in Moho and seafloor topography, as well as average crustal shear wave velocity: here denoted  $H_{kk}^M$ ,  $H_{kk}^S$  and  $\bar{M}_{kk}^c$ , respectively. For each crustal-model node  $n$ , we seek linear perturbations to Moho depth ( $h_M$ ), topography/bathymetry ( $h_S$ ) and crustal average  $V_S$  ( $\bar{V}_{S, \text{iso}}^c$ ):

$$\delta\mathbf{x}_n = (\delta h_M, \delta h_S, \delta \ln \bar{V}_{S, \text{iso}}^c)^T \quad (\text{A5})$$

such that the overdetermined linear system

$$\begin{pmatrix} H_{k_1 k_1}^M & H_{k_1 k_1}^S & \bar{M}_{k_1 k_1}^c \\ H_{k_2 k_2}^M & H_{k_2 k_2}^S & \bar{M}_{k_2 k_2}^c \\ \vdots & \vdots & \vdots \end{pmatrix} \delta\mathbf{x}_n = \begin{pmatrix} \delta\omega_{k_1}^n \\ \delta\omega_{k_2}^n \\ \vdots \end{pmatrix} \quad (\text{A6})$$

is satisfied in a least-squares sense for multiplets  $\{k_1, k_2, \dots\}$  within the period range of interest (e.g. 60–400 s). The set of three-factor correction coefficients  $\{\delta\mathbf{x}_n: n = 1, \dots, N_k\}$ , mimicking non-linearity induced by local crustal structure, are expanded laterally in spherical splines and used to calculate path-dependent corrections to integrations over  $\delta\omega_{kk'}$  in NACT.

Note that, in deriving our correction factors, we have limited our attention to the self-coupling case only. As noted by Lekić *et al.* (2010), satisfying the self-coupling case alone should be sufficient for the phases most affected by crustal structure at long periods (e.g. fundamental-mode surface waves). The resulting three-factor MLCs are sought for spheroidal and toroidal, and fundamental and overtone modes separately. This scheme allows us to use a single set of eigenfrequency partial deriva-

tives (those for  $m_{\text{ref}}$ ) in our NACT calculations, reducing file I/O and memory overhead relative to previous regionalized non-linear schemes (Marone & Romanowicz 2007), while achieving similar fidelity to crustal effects on the long-period wavefield. The accuracy of this approach relative to both ‘traditional’ crustal corrections and SEM benchmarks has previously been demonstrated by Lekić *et al.* (2010), and we refer the reader to their results.