Geophysical Journal International

Geophys. J. Int. (2013) **194,** 417–449
Advance Access publication 2013 April 4

doi: 10.1093/gji/ggt095

Global shear speed structure of the upper mantle and transition zone

A. J. Schaeffer[1,2] and S. Lebedev[1]

[1] *Geophysics Section, School of Cosmic Physics, Dublin Institute for Advanced Studies, Dublin, Ireland. E-mail: aschaeff@cp.dias.ie*
[2] *School of Geological Sciences, University College Dublin, Dublin, Ireland*

Accepted 2013 March 6. Received 2013 February 15; in original form 2012 November 16

SUMMARY

The rapid expansion of broad-band seismic networks over the last decade has paved the way for a new generation of global tomographic models. Significantly improved resolution of global upper-mantle and crustal structure can now be achieved, provided that structural information is extracted effectively from both surface and body waves and that the effects of errors in the data are controlled and minimized. Here, we present a new global, vertically polarized shear speed model that yields considerable improvements in resolution, compared to previous ones, for a variety of features in the upper mantle and crust. The model, SL2013sv, is constrained by an unprecedentedly large set of waveform fits (∼3/4 of a million broad-band seismograms), computed in seismogram-dependent frequency bands, up to a maximum period range of 11–450 s. Automated multimode inversion of surface and *S*-wave forms was used to extract a set of linear equations with uncorrelated uncertainties from each seismogram. The equations described perturbations in elastic structure within approximate sensitivity volumes between sources and receivers. Going beyond ray theory, we calculated the phase of every mode at every frequency and its derivative with respect to *S*- and *P*-velocity perturbations by integration over a sensitivity area in a 3-D reference model; the (normally small) perturbations of the 3-D model required to fit the waveforms were then linearized using these accurate derivatives. The equations yielded by the waveform inversion of all the seismograms were simultaneously inverted for a 3-D model of shear and compressional speeds and azimuthal anisotropy within the crust and upper mantle. Elaborate outlier analysis was used to control the propagation of errors in the data (source parameters, timing at the stations, etc.). The selection of only the most mutually consistent equations exploited the data redundancy provided by our data set and strongly reduced the effect of the errors, increasing the resolution of the imaging.

Our new shear speed model is parametrized on a triangular grid with a ∼280 km spacing. In well-sampled continental domains, lateral resolution approaches or exceeds that of regional-scale studies. The close match of known surface expressions of deep structure with the distribution of anomalies in the model provides a useful benchmark. In oceanic regions, spreading ridges are very well resolved, with narrow anomalies in the shallow mantle closely confined near the ridge axis, and those deeper, down to 100–120 km, showing variability in their width and location with respect to the ridge. Major subduction zones worldwide are well captured, extending from shallow depths down to the transition zone. The large size of our waveform fit data set also provides a strong statistical foundation to re-examine the validity field of the JWKB approximation and surface wave ray theory. Our analysis shows that the approximations are likely to be valid within certain time–frequency portions of most seismograms with high signal-to-noise ratios, and these portions can be identified using a set of consistent criteria that we apply in the course of waveform fitting.

Key words: Inverse theory; Surface waves and free oscillations; Seismic tomography; Computational seismology; Dynamics of lithosphere and mantle.

1 INTRODUCTION

The resolving power of global seismic tomographic imaging has improved dramatically over the last several decades, beginning from the early work of Dziewónski *et al.* (1977). These improvements have been facilitated by the rapid growth of high-quality, broad-band, three-component seismic data recorded by global and regional seismic networks. This growth has accelerated particularly

© The Authors 2013. Published by Oxford University Press on behalf of The Royal Astronomical Society.

in the last few years, stemming from the emergence of large- and continental-scale, high-resolution seismic arrays (i.e. EarthScope USArray, VEBSN, etc.). Also important has been the development of computational infrastructure and advancements in semi- and fully automated data-processing techniques and modelling methodologies.

Together, these developments have now paved the way for a new generation of global tomographic models. They will provide substantially higher resolution of the structure of the lithosphere and underlying upper mantle compared to ones of only a few years ago.

The long-wavelength structure of the Earth's lithospheric mantle (down to 200 or 300 km depth) has been well resolved for a number of years, with strong correlation for scale lengths of several thousand kilometres between many global models (Becker & Boschi 2002). However, such wavelengths are too long for consistent comparisons with geological and geochemical evidence on regional-scale tectonics. In the deep upper mantle and in the mantle transition zone (410–660 km depths), existing global models show substantially weaker agreement, even at the long wavelengths of a few thousand kilometres.

Global models have been obtained using a variety of methods and data sets, including traveltimes (e.g. Grand *et al.* 1997; Bijwaard & Spakman 2000; Karason & van der Hilst 2000; Grand 2002; Amaru 2006; Simmons *et al.* 2006; Li *et al.* 2008), surface waves (e.g. Zhang & Tanimoto 1993; Shapiro & Ritzwoller 2002; Zhou *et al.* 2006; Nettles & Dziewónski 2008; Ekström 2011), surface waves and body waves (or fundamental and higher modes, e.g. Woodhouse & Dziewónski 1984; Mégnin & Romanowicz 2000; Debayle *et al.* 2005; Panning & Romanowicz 2006; Lebedev & van der Hilst 2008; Ferreira *et al.* 2010; Lekić & Romanowicz 2011; Debayle & Ricard 2012) and surface waves with traveltimes and normal modes often included as well (e.g. Su *et al.* 1994; Masters *et al.* 1996, 2000; Gu *et al.* 2001; Ritsema *et al.* 2004, 2011; Houser *et al.* 2008; Kustowski *et al.* 2008a).

Models at regional to subcontinental scales can take advantage of particularly dense data sampling within regions and target higher resolutions. Continental-scale models provide coverage across yet larger regions, but are still limited by their constrained dimensions. Some recent examples, grouped by continent, include: North America (Bedle & van der Lee 2009; Burdick *et al.* 2010; Tian *et al.* 2011; Yuan *et al.* 2011), South America (van der Lee *et al.* 2001; Schimmel *et al.* 2003; Feng *et al.* 2004; Heintz *et al.* 2005), Eurasia (Amaru 2006; Priestley *et al.* 2006; Kustowski *et al.* 2008b; Legendre *et al.* 2012; Panning *et al.* 2012; Zhu *et al.* 2012), Australia (Simons *et al.* 1999; Debayle & Kennett 2000; Yoshizawa 2004; Fishwick *et al.* 2008; Fichtner *et al.* 2010) and Africa (Sebai *et al.* 2006; Pasyanos & Nyblade 2007; Priestley *et al.* 2008; Fishwick 2010). Direct comparisons between different regional or continental models are not always straightforward (one continent to another, for example), due to differences in regularization, parametrization and the data sets themselves (Bijwaard *et al.* 1998; Nettles & Dziewónski 2008).

With the expansion of seismic networks, improvements in computational capabilities, and advancements in methodologies, higher resolution (a few hundreds of kilometre length scales) global models have now become a reality. These new models enable exploration of deep lithospheric processes at the fine scale of tectonic units across entire continental domains. At these shorter length scales, however, global models show greater variance than at longer wavelengths (Becker & Boschi 2002). Such differences can arise from a number of factors, including data set selection (i.e. earthquakes and stations), treatment of errors in the data, model parametrization and regularization, data type (wave type, waveforms, phase delays, etc.), methodological and theoretical limitations or treatment of the crust (Lekić & Romanowicz 2011).

A number of theoretical and computational approaches have been developed during the past decade-and-a-half to more accurately relate the seismic wavefield to seismic velocity structure. These techniques focused, in particular, on the frequency-dependent and 3-D nature of seismic wave sensitivity regions. Modelling finite-frequency effects using the first-order Born approximation was applied to body waves (Dahlen *et al.* 2000; Nolet & Dahlen 2000; Zhao *et al.* 2000; Montelli *et al.* 2004; Sigloch *et al.* 2008; Zaroli *et al.* 2010) and surface waves and multimode waveforms (Li & Romanowicz 1995, 1996; Marquering *et al.* 1996; Meier *et al.* 1997; Yoshizawa & Kennett 2002; Zhou *et al.* 2006). Recently, fully numerical wavefield simulations have also been applied, more and more, in waveform tomography (Chen *et al.* 2007; Fichtner *et al.* 2009, 2010; Tape *et al.* 2009; Lekić & Romanowicz 2011; Zhu *et al.* 2012). However, the improved precision and accuracy in modelling greater wavefield complexity trades off very steeply with increases in computational cost. Commonly, such models utilize only tens to hundreds of events, and several hundreds of stations.

Asymptotic and ray-based approaches (e.g. Debayle *et al.* 2005; Lebedev & van der Hilst 2008; Kustowski *et al.* 2008a; Ferreira *et al.* 2010; Ritsema *et al.* 2011; Debayle & Ricard 2012) are computationally inexpensive and can be used with significantly larger data sets, affording a much higher degree of data redundancy. This can play a critical role in minimizing the impact of errors common to different types of methodologies, the most significant being event mislocations and incorrect source mechanisms and station timing errors. Simple algorithms which leverage the data redundancy to identify and partition the affected seismograms can be implemented to reduce (and in some cases eliminate) their effect on the final inversion product. The increased redundancy can also enhance the validity of approximations themselves, thanks to a larger data set. In both these regards, the utility and relevance of asymptotic techniques currently remains very clear.

The comparative advantages of different approaches of seismic tomography are now a subject of scrutiny and debate, as the field is developing and improving methods to exploit the enormous—and growing—volumes of available broad-band data. Different approaches may work best for different targets. For example, to image a narrow plume in the deep mantle—a notoriously difficult target because wave front healing nearly erases the signal of such a structure in teleseismic travel times—accurate numerical modelling of seismic wave diffraction and scattering off the plume may be the most suitable approach, even when applied to a small number of seismograms (Rickers *et al.* 2012). A very different problem is presented by the imaging of the wave speed distribution in the lithosphere and upper mantle at a regional to global scale: asymptotic methods that are applicable to very large data sets and capable of effective extraction of structural information from both surface and body waves are likely to exploit the redundancy of the presently available data more effectively and, thus, can provide models with higher resolution and greater robustness.

Automated multimode inversion (AMI) of surface and *S*-wave forms (Lebedev & Nolet 2003; Lebedev *et al.* 2005; Lebedev & van der Hilst 2008) was developed on the basis of the partitioned waveform inversion (Nolet 1990), which splits a large-scale tomography problem into more tractable inversions of each seismogram individually, similar to the techniques of Cara & Lévêque (1987) and Gee & Jordan (1992). AMI enables efficient, automated, accurate processing of very large numbers of vertical- and horizontal-component seismograms. This is accomplished through an elaborate window-selection procedure isolating signals least likely to contain scattered arrivals, combined with appropriate weighting of windows containing waves of different amplitudes and types, while enforcing strict misfit criteria. AMI assumes the JWKB approximation (Dahlen & Tromp 1998); time–frequency portions of each seismogram are systematically selected to contain only signals that can be accurately modelled. Instead of relying on ray theory and the path-average approximation, the initial phase velocities and their derivatives with respect to *S* and *P* velocities are computed as integrals over approximate sensitivity areas between sources and stations, within a 3-D reference model.

In this study, we have used AMI to generate an unprecedentedly large data set of ~3/4 million vertical-component, multimode waveform fits, each yielding a set of linear equations constraining perturbations in Earth structure. These equations were inverted together for our new global, upper-mantle, shear velocity model. The model is constrained by substantially more waveforms compared to any previous ones and an order of magnitude more than in the previous global application of AMI (Lebedev & van der Hilst 2008). The improved data sampling and data redundancy enable finer global parametrization and global resolution, which serves to further close the resolution gap between global and regional mantle studies (Bijwaard *et al.* 1998).

In the sections below, we begin with an overview of the methods and the assembly and preparation of the data set prior to inversion. We then present our new model and discuss its major features, from the large-scale ones, already seen in past models, to the smaller scale ones that are now imaged much more clearly than previously. Finally, we utilize our new data set of waveform fits to provide a sound statistical sampling of the bulk fundamental- and higher mode Rayleigh wave dispersive properties of the crust and upper mantle, and also take the opportunity to re-examine the global validity of the JWKB approximation and surface wave ray theory. In the Appendices, we provide a further analysis of the data set as well as how the total frequency band of waveform fitting affects the resulting tomographic models.

2 INVERSION PROCEDURE

We have built a global, vertically polarized shear velocity model, extending from the crust to the base of the transition zone, through the application of a three-step waveform fitting and inversion procedure. We begin with the application of AMI to broad-band seismograms to generate sets of linear equations which constrain the sensitivity-volume average velocity perturbations between each source and receiver, with respect to a 3-D reference model. In the second step, we combine the equations from AMI into one large linear system and solve it for the 3-D distribution of *P*, *S* and azimuthal anisotropy perturbations, subject to regularization and smoothing, using the LSQR algorithm (Paige & Saunders 1982) and following the procedure of Lebedev & van der Hilst (2008). Finally, we perform an outlier analysis of the data set, and *a posteriori* select the most mutually consistent equations to be reinverted so as to constrain the final model.

2.1 Waveform inversion

AMI's numerical efficiency, the capacity to select signal for which theoretical approximations hold and then weight and balance the information derived from different portions of the wave train, enables accurate processing of very large numbers of waveforms, resulting in high-resolution models of the Earth's upper mantle and transition zone. The fully automated algorithm is built on the basis of the partitioned waveform inversion of Nolet (1990), as described in detail by Lebedev *et al.* (2005), with further advancements in Lebedev & van der Hilst (2008).

For each seismogram, AMI uses non-linear waveform fitting to derive a set of linear equations with uncorrelated uncertainties that describe finite-width sensitivity-volume-average *S*- and *P*-velocity perturbations [$\overline{\delta\beta(r)}$ and $\overline{\delta\alpha(r)}$], with respect to a 3-D reference model. Synthetic seismograms are computed in the frequency domain using the JWKB mode summation:

$$s(\omega) = \sum_m A_m(\omega) \exp\left(\mathrm{i}\omega\Delta\left[C_m^0(\omega) + \overline{\delta C_m(\omega)}\right]\right), \tag{1}$$

by summing over the modes, *m*, for the given source–receiver distance Δ. The initial phase velocities, $C_m^0(\omega)$, and their Fréchet derivatives are pre-computed for our 3-D reference model. For each source-station pair, they are averaged across approximate sensitivity kernels. The average phase-velocity perturbations, $\overline{\delta C_m(\omega)}$, are expressed as functions of the sensitivity-volume average perturbations in *P* and *S* velocity:

$$\overline{\delta C_m(\omega)} = \int_0^R \frac{\delta C_m^0(\omega)}{\delta\beta(r)}\, \overline{\delta\beta(r)} + \int_0^R \frac{\delta C_m^0(\omega)}{\delta\alpha(r)}\, \overline{\delta\alpha(r)}\, \mathrm{d}r, \tag{2}$$

where *R* is the radius of the Earth.

This scheme (Lebedev & van der Hilst 2008) goes significantly beyond ray theory and path-average approximations. The phase of every mode at every frequency and its derivatives with respect to seismic wave speeds are computed as integrals over the sensitivity area in the 3-D reference model. Only the perturbations to the phase velocities (small, in most cases) are linearized, and this is done using the accurate average derivatives. At no point do we use the cruder approximation of 1-D, path-average models. Generally, a 1-D model with the same phase velocities and their derivatives as the ones we compute by sensitivity-area integration is not likely to even exist (although, of course, there may exist models with dispersion properties that are similar).

Initially, $\overline{\delta\beta(r)}$ are parametrized using a set of 1-D basis functions $h_i(r)$ which span depths from 7 km within the crust to ~1600 km in the upper portion of the lower mantle (18 parameters for *S* and 10 parameters for *P*). Through diagonalization of the Hessian matrix, these are transformed into independent linear equations with uncorrelated uncertainties (Nolet 1990), with new parameters η_i corresponding to basis functions $g_i(r)$, with each $g_i(r)$ a linear combination of the original basis functions $h_i(r)$:

$$\overline{\delta\beta(r)} = \sum_{i=1}^{M} \eta_i g_i(r). \tag{3}$$

The strength of AMI is rooted in the fully automated selection and weighting of time–frequency windows, which enables reliable and accurate application to massive data sets. Computed normal-mode synthetics are matched with real seismograms in time–frequency windows that isolate the fundamental- and higher mode wave trains unaffected by scattered waves. Examples of this waveform fitting procedure are shown in Figs 1 and 2 for two paths, one shorter and one longer. In the following paragraphs, key elements of the technique are highlighted; for greater detail we refer the interested reader to Lebedev *et al.* (2005).

Slip along a fault generates a non-uniform radiation pattern, with the initial phase and amplitude of every mode varying as a function of azimuth and frequency. Source-station azimuths close to a node

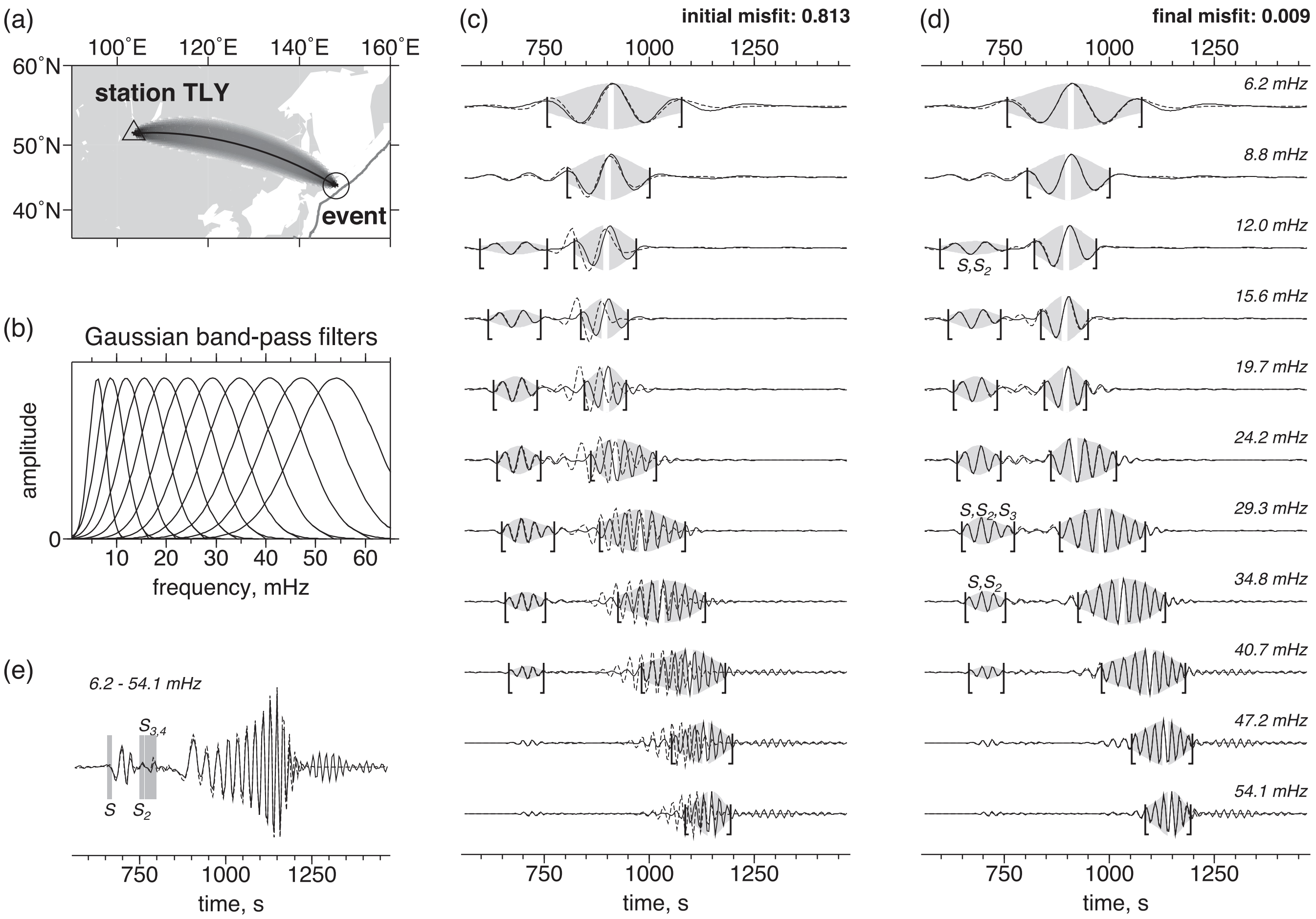

Figure 1. Automated multimode waveform inversion example. (a) An earthquake on 1997 February 28 (43°N, 148°E, 37 km depth, moment Magnitude $M_W = 5.8$) recorded on the vertical component at the broad-band station Talaya, Russia (TLY) of the Global Seismograph Network (GSN), operated IRIS/IDA; the source-station distance is 3380 km. The approximate sensitivity area is shaded grey, with darker colours indicating greater sensitivity. (b) 11 closely spaced Gaussian filters used in generating the different time–frequency windows. (c) The resulting waveforms (solid lines) are matched with synthetics (dashed lines) in 18 different time–frequency windows simultaneously. The time windows are indicated by half-brackets, with the signal envelope shaded. The fundamental-mode wave train is identified by vertical white bars at the maxima of the envelope. The initial fit is computed using our 3-D background model. (d) The misfit is minimized through non-linear inversion for the sensitivity-volume average perturbations $\overline{\delta\beta(r)}$ and $\overline{\delta\alpha(r)}$. Energy in the synthetic is equalized with that of the data in each window. All 18 selected time–frequency windows have final data-synthetic misfits less than 5 per cent. The average perturbations computed by waveform inversion constrain the *S*- and *P*-velocity perturbations within the sensitivity area shown in (a), used in tomographic inversion. (e) Final data-synthetic fit within a single, broad time–frequency window encompassing the entire frequency range of this waveform inversion. Arrival times of the *S*- and triplicated multiple-*S* waves predicted by AK135 (Kennett *et al.* 1995) are indicated by grey shading. The same phases are also indicated above their frequency windows in (d).

in the radiation pattern are more likely to contain relatively higher proportions of scattered energy within that portion of the seismogram, and should therefore be avoided. Prior to waveform fitting, the frequency- and azimuth-dependent nodal radiation patterns are computed for each seismogram. For each frequency, azimuth bands in which the amplitude of the predicted pattern are less than half the maximum across all azimuths at that frequency are determined, and discarded. If the given source-receiver geometry does not fall in any permitted azimuthal bands at any frequency, then that seismogram is discarded.

The Gaussian filter windows (B, Figs 1 and 2) are initially defined within the range selected by the frequency-dependent azimuthal nodal radiation pattern, for each given seismogram. This may be narrowed further through enforcement of the far-field and point-source approximations.

The far-field approximation ensures sufficient source–receiver distance to avoid complexities due to near-field wave propagation effects (e.g. evanescent waves); based on extrapolation from the work of Pollitz (2001), two fundamental-mode wavelengths are sufficient. The minimum frequency filter is then constructed such that the left-most tail frequency corresponding to an amplitude of 0.3× the filter central maximum contains exactly three fundamental-mode wavelengths between the source and receiver. This ensures the filter centre frequency (dominant frequency) contains more than three wavelengths. Effectively, it is the path length that controls the minimum filter frequency: the longer the paths, the lower the minimum frequency (i.e. the longer the maximum period) of the fundamental-mode waveforms.

The validity of the point-source approximation is ensured by setting a maximum frequency (minimum period) limit at $1/3\tau$, where τ is the earthquake source duration time, taken from Centroid Moment Tensor (CMT) catalogues. (If the period of the wave is comparable to the source duration time, then both its amplitude and phase will be affected by unmodelled complexity of the source). As a result, the earthquake magnitude controls, in part, the maximum frequency of fitting: the larger the magnitude, the longer the source

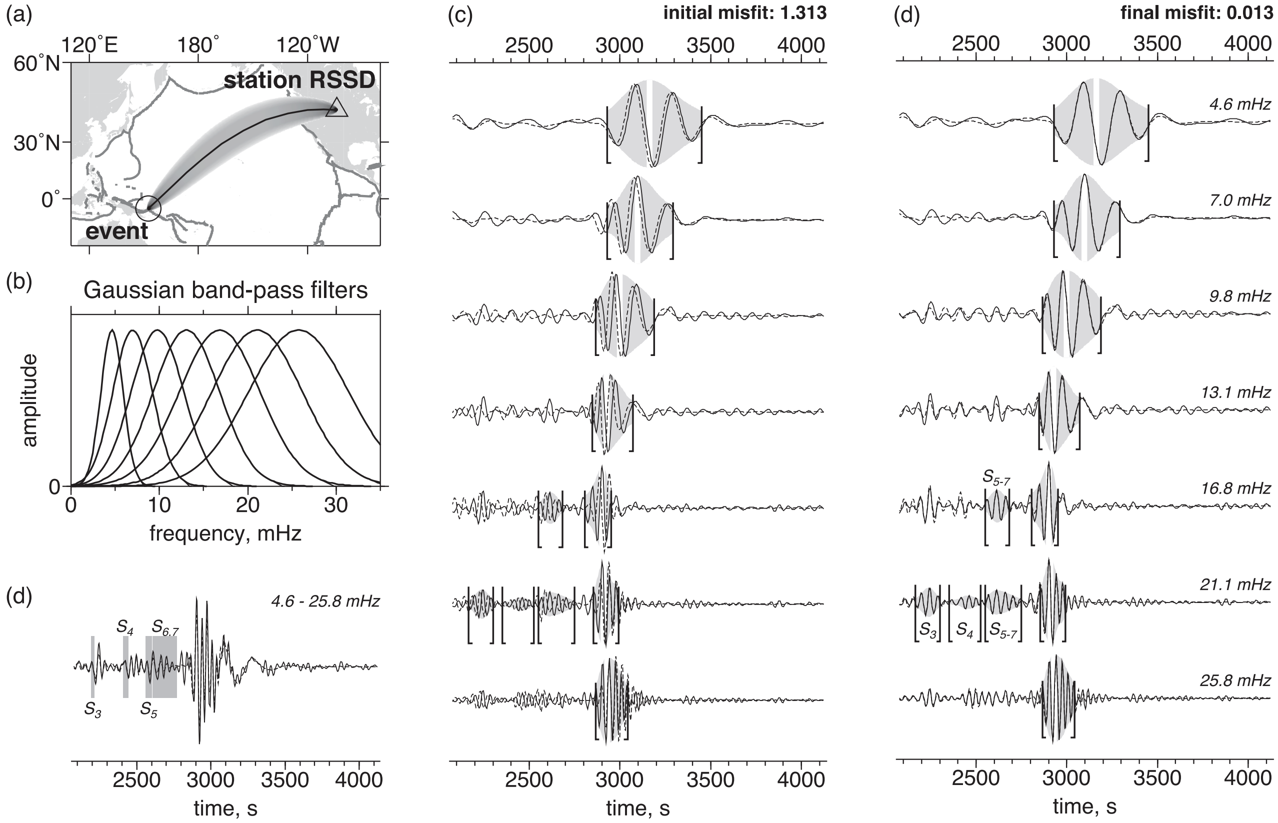

Figure 2. Automated multimode inversion example. Earthquake on 2003 June 7 (5.3°S, 152.6°E, 30 km depth, moment magnitude $M_{\mathrm{W}} = 6.6$) recorded on the vertical component of station Black Hills, South Dakota (RSSD of the GSN, operated by IRIS/USGS); the source-station distance is 11 485 km. Plots are the same as in Fig. 1, except with seven Gaussian filters, 11 time–frequency windows and arrivals of multiple $S_3 - S_7$ as indicated.

duration time, and the lower the filters' maximum frequency (the longer the shortest period).

The time–frequency windows are generated through application of boxcar time windows after bandpassing with the suite of Gaussian filters. The time windows contain individual wave trains or series of wave trains, and are selected such that their boundaries do not cut the middle of a wave packet: the signal at the nearest maximum of the envelope must be at least ∼3.5 times larger than at the window boundary. Lebedev *et al.* (2005) initially used a more conservative threshold of 4–5; however, further testing demonstrated this can safely be relaxed to a lower value, without detriment to waveform fits. The rightmost time-window limit immediately follows the fundamental-mode arrival at all filter frequencies, avoiding scattered waves in the coda. The leftmost time-window limit varies as a function of epicentral distance, with cut-offs set to eliminate *S* and multiple-*S* waves sampling the deep lower mantle. As a result, time–frequency windows at a given frequency may contain only a single fundamental-mode window, fundamental and higher modes or a group of windows containing fundamental and higher mode information (D, Figs 1 and 2).

Waveform fitting begins with the lowest frequency Gaussian filter (pre-determined such that the far-field approximation is valid and nodes in the radiation pattern are avoided) and widest possible time windows. The minimum centre frequency of the first (lowest frequency) filter is increased until the signal-to-noise ratio is sufficiently high, such that a low-noise window is found. By beginning at this lowest possible frequency, the likelihood of errors resulting from 2π phase ambiguities (cycle skipping) are minimized. For the same reason, the signal envelope is fit first in every time–frequency window, and the source-station distance is required to not exceed 20 wavelengths of the lowest frequency fundamental mode. Visual examination of thousands of waveform fits confirmed that these measures, together with our use of a 3-D reference model with realistic crust, are sufficient to rule out cycle skips. The only exceptions detected occurred due to very large (tens of seconds) timing errors at the stations. Waveform fits affected by such errors are removed at the later, outlier-removal stage.

After the minimum frequency filter is determined, waveform fitting proceeds with the iterative addition of higher frequency time–frequency windows, where individual wave trains in all time–frequency windows are inverted simultaneously, searching for $\overline{\delta\beta(r)}$ which minimizes the cumulative misfit across all windows (Lebedev *et al.* 2005). In addition, a fit is considered successful only if the data-synthetic misfit in each time–frequency window is less than 5 per cent. If the fit is acceptable only within portions of the window, it is iteratively narrowed or split to attain the target data-synthetic misfit in each new subwindow, while enforcing the requirement that each must contain a complete wave train. Fig. 2(d) at 16.8 and 21.1 mHz illustrates this window refinement procedure: what began as single windows spanning 2150–3000 s were iteratively split into two and four windows, respectively.

2.2 3-D reference model

To accurately relate the phase information in the waveform to perturbations in *S* and *P* velocity within its sensitivity area, we required as accurate as possible reference phase velocities $C_m^0(\omega)$ and their Fréchet derivatives $\delta C_m^0(\omega)/\,[\delta\beta(r),\ \delta\alpha(r)]$. Less accurate derivatives would result in more inconsistent equations, which manifest as

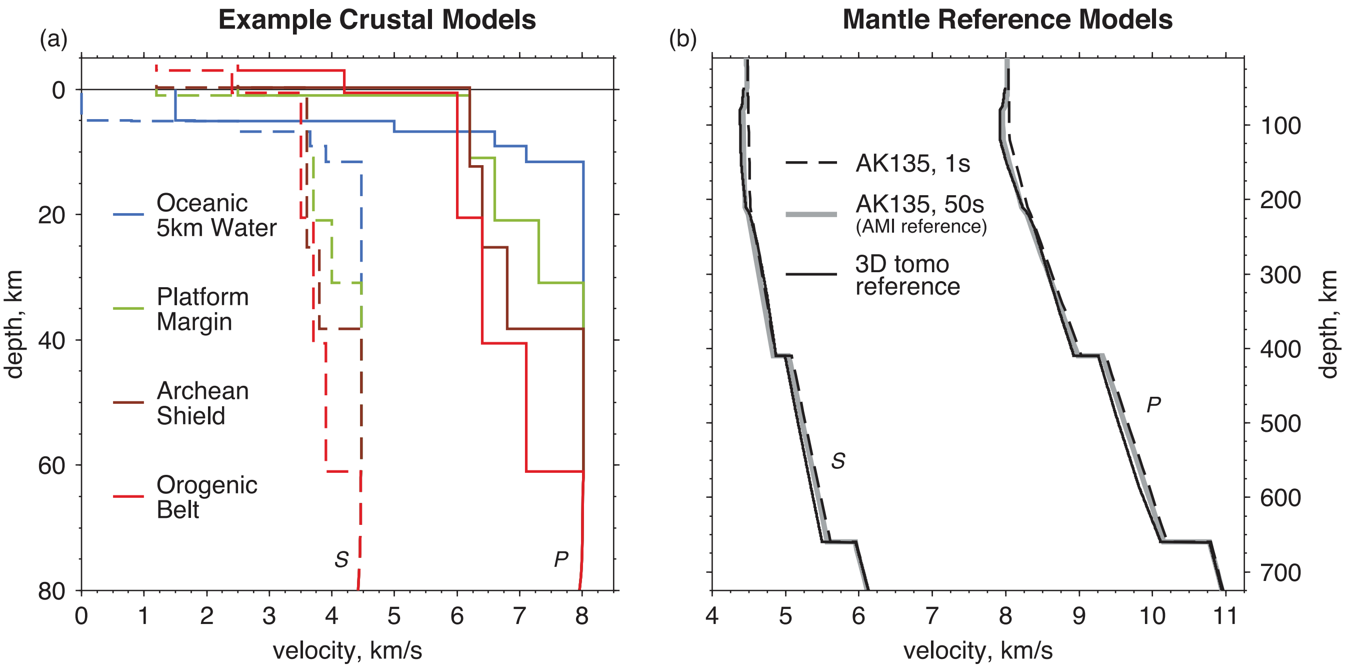

Figure 3. Reference velocity models used for both AMI and 3-D tomography. Left-hand panel (a) illustrates four example crustal models, based on CRUST2 (Bassin *et al.* 2000). In each crustal model, velocities below their respective Moho converge to the same mantle reference velocity profile (b). Topography is accounted for in each model, and is indicated by negative depth values. The oceanic model (blue) has a 5-km water layer and 100 m of sediments above a 'normal' oceanic crustal model (CRUST2 type *A0*); the platform margin model has 1 km of sediments with minimal topography (type *DG*); the Archean shield model (based on Canadian Shield) has 500 m of sediments and an elevation of 700 m (type *G2*); the orogenic model (based on Tibet/Himalaya) has 4.5 km of sediments and 4 km elevation (type *RC*). Right-hand panel (b) shows the different mantle reference models, including AK135 (Kennett *et al.* 1995). AMI uses a modified AK135, recomputed at a reference period of 50 s (*AK135_50*). The reference model for tomography is generated by adjusting *AK135_50* based on an initial global inversion by Lebedev & van der Hilst (2008).

noise in the tomographic inversion. Consequently, greater smoothing and norm damping are required, and the model resolution is decreased.

The high lateral heterogeneity in the Earth's crust gives rise to significant lateral variability of the Fréchet derivatives. When not adequately accounted for, this gives rise to artefacts in both the crustal and mantle parts of the tomographic model. Generally, crustal structure is often accounted for using 'crustal corrections', which involves computing period-dependent corrections based on an assumed crustal model, which are then applied to phase-velocity maps used in the inversion (e.g. Boschi & Ekström 2002; Gu *et al.* 2003; Chevrot & Zhao 2007; Kustowski *et al.* 2007; Marone & Romanowicz 2007; Bozdağ & Trampert 2008; Lekić *et al.* 2010). Instead of computing corrections, we construct a realistic 3-D reference model which includes *a priori* crustal structure, and then solve for velocity perturbations with respect to it. Thanks to this more accurate approach and to our fundamental-mode waveform fits at the relatively short periods of 15–25 s, obtained for shorter source-station distances and sensitive to crustal structure, we were able to resolve, typically, the average perturbations with respect to the 3-D reference model within the normal-continent crustal depth range, while also resolving intracrustal layering for thick continental crust. This increased the accuracy of our model in the upper mantle below as well.

We sampled the Earth's surface with a dense triangular grid of knots (Wang & Dahlen 1995a), over which the 3-D global crustal model CRUST2 (Bassin *et al.* 2000) was parametrized. The 360 type models were smoothed at the boundaries of the 2° × 2° cells and augmented with topographic and bathymetric databases to generate a larger suite of models encompassing greater variations in water, sediment and crustal thickness. The derivatives $\delta C_m(\omega)/[\delta\beta(r), \delta\alpha(r)]$ were compared and a subset of 664 exemplar models were selected and weighted across the triangular grid (average interknot spacing of ~28 km). Fig. 3(a) illustrates four example crustal models drawn from this set.

The mantle reference beneath the Moho is based on AK135 (Kennett *et al.* 1995), but recomputed at a reference period of 50 s to minimize errors due to lateral variations in attenuation, poorly resolved within the waveband of interest (hereon referred to as *AK135_50*). Fig. 3(b) illustrates the mantle reference models utilized by AMI and the tomographic inversion. As indicated, the thick grey line is the *AK135_50* reference used by AMI. For the tomographic inversion, this model has been modified slightly (solid black line) based on an initial global inversion (Lebedev & van der Hilst 2008), to bring the reference values closer to global averages.

Phase velocities and their Fréchet derivatives ($C_m(\omega)$ and $\delta C_m(\omega)/[\delta\beta(r), \delta\alpha(r)]$) are pre-computed for every lateral knot in the reference model, covering the broad frequency band 0.488–125 mHz (8–2048 s, extending beyond that used in waveform fitting). This enables efficient computation of $C_m^0(\omega)$ and $\delta C_m^0(\omega)/\,[\delta\beta(r), \delta\alpha(r)]$ for any source–receiver path simply by summing together weighted phase velocities and derivatives; the weights are based on an approximate sensitivity kernel $\overline{K(\theta,\phi)}$, averaged over the frequency band. The same approximate sensitivity areas are used in both waveform fitting and the tomographic inversion (see Section 2.3.1).

2.3 3-D inversion

The result from AMI for each seismogram consists of a number of equations on the orthogonal basis $g_i(r)$ with parameters η_i and uncorrelated uncertainties $\Delta\eta_i$ which describe the 1-D average *S*- and *P*-velocity perturbations within the sensitivity volume between the source and receiver, relative to the 3-D reference model (critically we note this is different than generating a 1-D path-average model). By combining together the equations obtained from all the successfully fit seismograms, a large linear system is constructed, from which the 3-D distribution of *P*, *S* and azimuthal anisotropy perturbations, from the 3-D reference model, are solved for with LSQR, subject to regularization and smoothing. The horizontal sensitivity of each seismogram is given by the same kernel, $\overline{K(\theta, \phi)}$, as in waveform fitting. The vertical structure of the kernels $g_i(r)$ differs for each seismogram. Only linear equations with corresponding eigenvalues exceeding a pre-determined threshold are incorporated into the inversion. On average, this results in ∼3.5 equations per path, more if *S* waves or broader band fundamental modes are included. Key elements of the inversion procedure are outlined in the following sections; for further detail, we refer to Lebedev & van der Hilst (2008).

2.3.1 *Gridding*

To build the linear system, we first generate two global, coregistered triangular grids of knots, using the method of Wang & Dahlen (1995a). The first is a dense integration grid with nominal interknot spacing of ∼28 km (same as the reference model). The second grid, with knot spacing of ∼280 km, is the model grid, on which perturbations of the isotropic-average shear and compressional speeds and shear velocity anisotropy are expanded and solved for. By design, the knots of the model grid are co-located with knots of the integration grid, enabling efficient transformations between the two. Fig. 4(a) illustrates the locations of the integration (black dots and yellow circles) and model (red and blue circles) grid nodes, and their relative sensitivities (yellow and blue circles) for the path in Fig. 1.

The same 'shell' of knots is used at all depths in the model. Vertically, *S*-velocity perturbations are parametrized on 18 'stem' nodes: 7, 20, 36, 56, 80, 110, 150, 200, 260, 330, 410-, 410+, 485, 585, 660-, 660+, 809 and 1007 km, whereas for *P* velocity there are only 10 parameters: 7, 20, 36, 60, 90, 150, 240, 350, 485 and 585 km. Anomalies between the knots of this 3-D grid are computed by trilinear interpolation. The 'stem' nodes are the same as the vertices of the triangular basis functions $h_i(r)$ used in the waveform inversion, prior to orthonormalization, as illustrated in Fig. 4(b). The transition zone discontinuities at 410 and 660 km are accommodated using pairs of half-triangles. The inclusion of the shallowest nodes ensures that globally there is at least one model node in the crust, and at times up to four. Therefore, perturbations from CRUST2 are solved for directly in the inversion, which helps to minimize the inaccuracies of CRUST2. As will be discussed in Section 5, the resulting model contains strong deviations from the crustal reference in many locations (e.g. across Tibet).

For a given seismogram fit by AMI, the sensitivity kernel $\overline{K(\theta, \phi)}$ around the corresponding path is evaluated on the integration grid, with the total weight for the *i*th knot being the product between the sensitivity $K(\theta_i, \phi_i)$ and the area $A_i(\theta_i, \phi_i)$ [defined by the hexagon (pentagon) that contains all points that are closer to this grid knot than to any other]. The sensitivity areas $\overline{K(\theta, \phi)}$ are similar to the 'influence zone' of Yoshizawa & Kennett (2002) and the traveltime kernels of Zhou *et al.* (2005), essentially encompassing the interior region bounded by the '$\pi/2$' Fresnel zone, computed at a single frequency in the middle of the fundamental-mode's frequency band. Weights are largest closest to the source and receiver; cross-sections reveal that at any point along the path, weights decrease with distance from the great-circle ray path, to a total width of ±δ, where δ is the width of the '$\pi/2$' Fresnel zone (yellow circles in Fig. 4a). The sensitivity kernels for each seismogram are then mapped onto the model grid through averaging of the integration grid knots, with the resulting weights applied to the parameters for that path in the inversion (blue circles in Fig. 4a).

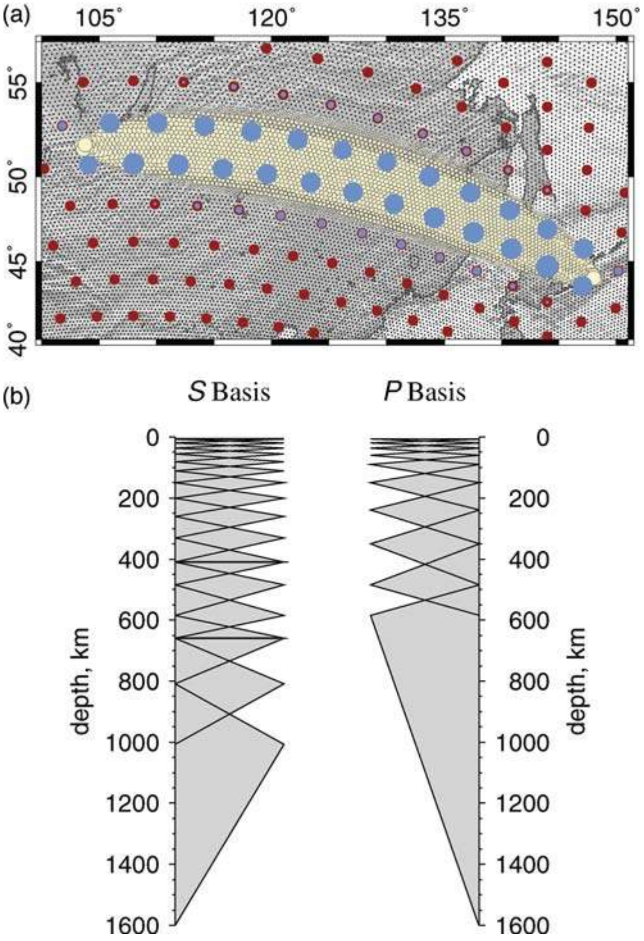

Figure 4. (a) Model and integration grids used in tomography shown by red circles and black dots, respectively. The source-station path illustrated is that from Fig. 1. Superimposed is the sensitivity area (kernel) $\overline{\kappa(\theta, \phi)}$ from the source-station path illustrated in Fig. 1, represented by the yellow (integration grid) and blue (model grid) circles. The circle sizes in the integration grid (yellow) scale with the weight of the knots in the sensitivity-area integral, whereas in the model (blue) grid their size indicates the contribution of each knot in inversion, for that path. (b) Triangular vertical basis functions used in parametrization for *S*- (left) and *P* velocity (right). Note that the discontinues at 410 and 660 km in the *S* parameters (absent in *P*) are generated using two half triangles, with one above and the other below the discontinuity.

2.3.2 *Path weighting*

The global distribution of seismometers and seismicity is not even, with station locations biased to continental regions and oceanic islands, and events clustering along plate boundaries. As a result, most large seismic data sets contain some sampling bias. It is clear from the stations and events shown in Fig. 5 that this is the case in our data set. To reduce the effect of common or 'bundled-rays,' a

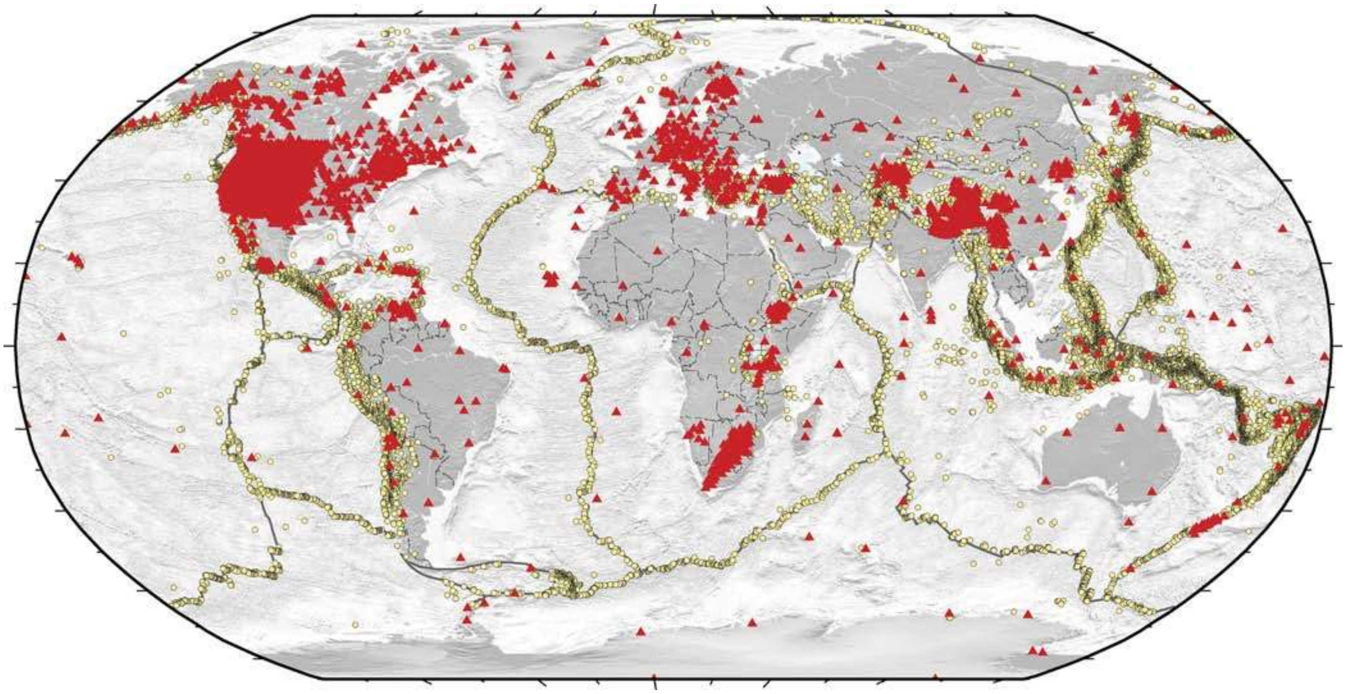

Figure 5. Map illustrating the distribution of stations and events in the data set. More than 5000 stations from International, National, Regional and temporary deployments are represented by red triangles. The ∼27 000 total events recorded at the different stations are shown as yellow circles.

reweighting of all paths is performed to produce a normalized path coverage.

For each path in the inversion, weights measuring cumulative similarity of a path to all other paths are computed and applied to the equations corresponding to this path. This results in relative downweighting of equations derived from commonly travelled paths, with more even distribution of relative sampling across the model space. The post-path reweighting coverage for several depths in the final model is illustrated in Fig. 8.

2.3.3 Parametrization and regularization

The inversion solves directly for perturbations in *P*, *S* and 2Ψ *S*-velocity azimuthal anisotropy with respect to our 3-D reference model. Lebedev & van der Hilst (2008) verified that the perturbations in *P* velocity could not be resolved independently using Rayleigh wave data. To avoid trade-offs, the difference between isotropic *P*- and *S*-velocity perturbations are damped, in the form $|\delta V_P(\mathrm{m\,s^{-1}}) - \delta V_S(\mathrm{m\,s^{-1}})|$. This offers greater freedom to the inversion, as opposed to forcing a rigid coupling; nevertheless, the resulting *P*- and *S*-velocity images are still quite similar.

Azimuthal anisotropy is described by Smith & Dahlen (1973) as a harmonic function of the form:

$$C(T,\psi) = C_0(T) + A_1(T)\cos(2\psi) + A_2(T)\sin(2\psi) + A_3(T)\cos(4\psi) + A_4(T)\sin(4\psi), \quad (4)$$

where T is period, ψ the azimuth, C the observed phase velocity, C_0 the isotropic phase velocity and A_i the anisotropy parameters. The strength of the 4ψ anisotropic terms is known from previous global and regional studies to be weak for long-period surface waves (Montagner & Tanimoto 1991; Trampert & Woodhouse 2003; Deschamps *et al.* 2008b; Darbyshire & Lebedev 2009; Adam & Lebedev 2012); therefore we disregard the 4ψ terms and focus only on the 2ψ components for vertically polarized Rayleigh waves. In this paper, we present the isotropic component of an azimuthally anisotropic model, which included highly smoothed 2ψ terms to reduce errors resulting from trade-offs between isotropic and anisotropic heterogeneity. The azimuthally anisotropic structure shall be the focus of subsequent work.

Regularization is carried out in the form of lateral and vertical smoothing and slight norm damping, to stabilize the mixed-determined inversion. Smoothing is the primary control, whereas the weak damping plays a secondary role. We apply two kinds of smoothing, with one penalizing the difference between the anomaly at a node and the average anomaly over this and the six (or five) nearest neighbour nodes and the other (gradient damping) penalizing the differences between pairs of neighbouring model knots. In both cases, the strength of the coefficients decreases as a function of increasing depth, to prevent oversmoothing at greater depths due to reduced sampling there (fundamental-mode sensitivity is lower at greater depths). Vertical gradient damping penalizes rapid changes with depth within the model.

Finally, norm damping penalizes the amplitude of the anomalies with respect to the 3-D reference model, with the strength of the damping, again, a function of depth. The data sampling (and hence the relative strength of regularization) is quantified using the column sums of the matrix **A** that relates the model vector $\overline{m}$ to the data vector d ($\mathbf{A}\overline{m} = \overline{d}$).

2.4 Outlier analysis

Outlier analysis and rejection (e.g. Lebedev & van der Hilst 2008) is critical for improving equation consistency, inversion convergence and the resolution of the imaging. Automated methods enable fast data processing and production of very large waveform data sets; it is important to assess the relative quality of the successfully fit seismograms utilized in the tomographic model. Manual analysis and examination of the waveform fits was carried out on a subset; however, this process is time-consuming and negates efficiencies gained by automation. To quantitatively assess the consistency and relative quality of hundreds of thousands of waveform fits, we perform an objective search for those equations which deviate significantly, most commonly due to source mislocations, errors in event origin times and source mechanisms, as well as station timing errors.

The basic outlier analysis procedure utilizes an initial tomographic inversion for the model $\overline{m_i}$, from which synthetic coefficients (data, $\overline{d_s}$) are generated through the matrix multiplication $\overline{d_s} = \mathbf{A}\overline{m_i}$. Then, the distribution of data-synthetic misfits is analysed, and those equations that lie in the tails of the distribution, well beyond the 2–3σ level (95–99 per cent), can be identified, examined and possibly rejected. In practice, eliminating just 1–2 per cent of outliers greatly improves the inversion convergence and the resulting model. However, the massive size of our new data set enables us to be more selective regarding which equations are retained.

We have undertaken a rigorous, conservative outlier analysis procedure, to select only the most mutually consistent equations for use in our final tomographic model. A series of smaller *a posteriori* outlier analyses were carried out. To identify the most consistent equations, we used as a benchmark the data set of global waveform fits from Lebedev & van der Hilst (2008). Subsets of 10–15 000 randomly selected seismograms from our data set were inverted together with the 51 004 waveform fits of the benchmark data set.

We elected to use this benchmarking method for several reasons. First, the model of Lebedev & van der Hilst (2008) accurately recovers the major S_V structure of the upper mantle and transition zone. Secondly, it has been shown by Becker *et al.* (2012) that the anisotropic component of this model correlates well with global *SKS* splitting measurements, indicating that its accuracy extends beyond the isotropic shear speed originally presented. Thirdly, several passes of outlier analysis were carried out on the benchmark data set, therefore its equations are highly mutually consistent. Finally, the use of such a benchmark inversion provides statistical constraints on our new data set, steering convergence towards a reasonable final model, while also leaving it the freedom to deviate if required by the data. Most importantly, this eliminates the fits affected by large errors in the data.

Fig. 6 provides an example of this procedure for one subset of 10 500 seismograms from stations of the Global Seismographic Network (GSN). Data residuals ($\overline{d} - \overline{d_s}$) are normalized by the estimated uncertainty of each datum (Nolet 1990; Lebedev & van der Hilst 2008). The misfits for the benchmark and new subset are separated, and plotted in different colours. Only those seismograms with corresponding misfits (blue points) inside the range of the benchmark data set (green), indicated by the red dashed lines, are retained. In this example, ~2–3 per cent of blue equations have residuals outside the accepted range; retaining only seismograms whose equations are within the limits, we discard (in this case) ~6–9 per cent of the seismograms. For a subset containing noisier data than that pictured here (e.g. some temporary and regional arrays), a larger per cent of seismograms may be removed (up to 15 per cent).

After this first set of outlier removals is carried out, a second pass is performed, where each reduced data set is reinverted. After this second pass, there is much less scatter in residuals, and fewer equations lie outside the misfit range defined by the benchmark. This second pass may result in a further reductions of 1–3 per cent of seismograms.

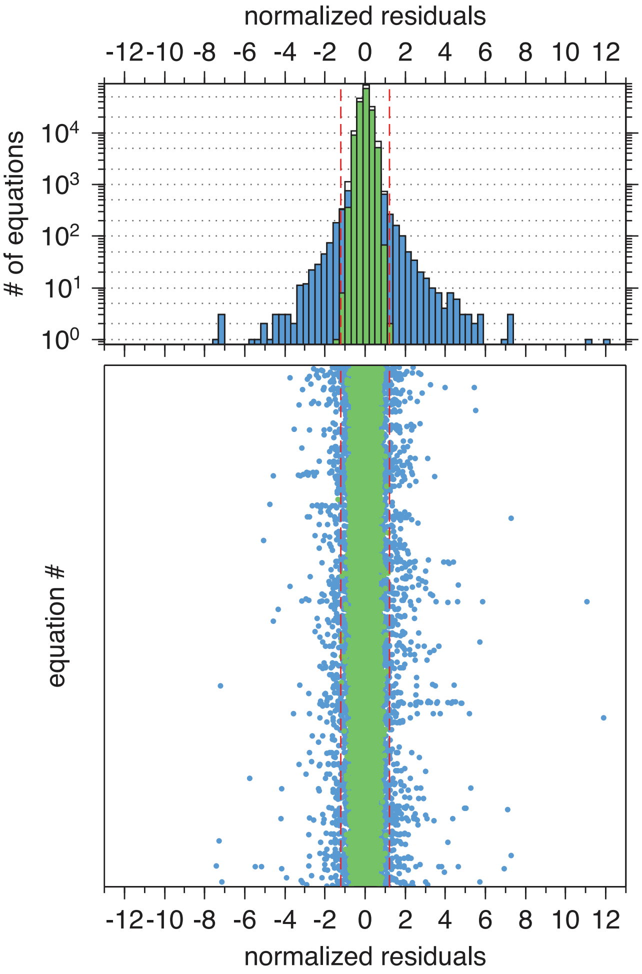

Figure 6. Example outlier analysis for a single data subset. Green represents the 51 004 seismograms (153 509 equations) that constrained the global model of Lebedev & van der Hilst (2008). Blue represents the subset of 10 500 GSN seismograms (~30 200 equations) randomly selected from data set B (Table A1). The outlier inversion is carried out on these 61 504 seismograms (~184 000 equations). The top panel illustrates the 'log'-scaled histogram of the data-model residuals. The bottom panel illustrates the raw residuals for each equation. It is clear the residuals from the data set of Lebedev & van der Hilst (2008) are substantially smaller than those in the subset, and restricted almost entirely to ±1 (range indicated by dashed red lines). The data subset from our new set of waveforms fits exhibits a much greater degree of scatter, and only those seismograms with misfits within the dashed red lines are retained.

3 DATA AND PROCESSING

3.1 Seismogram selection and preparation

Using our large new data set, we expect to improve resolution in the upper mantle and transition zone using the structural constraints extracted by AMI from surface, *S*, and multiple-*S* (up to at least S_7) waves over a broad range in periods spanning 11–450 s (note 11 s is the global minimum, more significant contributions begin at 20 s). We have assembled data from more than 120 international, national, regional and temporary seismic networks available from Incorporated Research Institutions for Seismology (IRIS), GFZ-Potsdam (GEOFON), Observatories and Research Facilities for European Seismology (ORFEUS) and Canadian National Seismic Network (CNSN) Data Centres; in total this includes data from more than 5000 stations.

The events used are those in the Global Centroid Moment Tensor (CMT) catalogue (e.g. Ekström *et al.* 2012), which contains more than 36 000 events since 1977. While we would like to obtain as much data as possible without prior discrimination, there are several criteria enforced to reduce the quantity of noisy data.

The primary criterion used for selecting seismograms is based on an empirical relationship between the epicentral distance, earthquake magnitude and the signal-to-noise ratio. For a seismogram to be requested, the earthquake magnitude must exceed a computed threshold, depending on the source–receiver distance. The minimum magnitude increases linearly with distance until ∼12 000 km, beyond which all earthquakes with $M_W \geq 5.7$ are requested. The parameters defining this empirical cut-off were selected based on examination of past waveform fitting results using AMI (e.g. Lebedev & van der Hilst 2008).

A number of pre-processing steps were undertaken to prepare the raw seismograms for input to AMI. First, seismograms were checked for segmentation due to clock drift, and subsequently merged if timing gaps were small ($\ll \Delta t$ where $\Delta t = 1$ s). Next, clipped seismograms and those with missing data were identified and removed. Finally, instrument response was removed, and the horizontal components were rotated into radial and transverse orientations. Seismograms not passing any of the checks were removed from the data set for follow up analysis. Finally, arrival times for the first arriving *P* wave are computed to estimate the signal-to-noise ratio prior to onset.

These rigorous checks resulted in a data set of more than 3.6 million vertical- and 2.9 million transverse-component broad-band seismograms recorded for events between 1981 January and 2010 March. The distribution of sources and receivers are shown in Fig. 5. The red triangles represent stations, the yellow circles indicate the events.

3.2 Waveform fitting

A single instance of AMI runs with a memory footprint of no more than 1 GB, used mostly in storing the 3-D reference model's phase velocities and derivatives. Therefore, with modern multicore desktop computers, the serial nature of waveform fitting is readily extended through parallelization using one of the available suites of tools. AMI determines and discards most unsuccessful fits in less than 1 s; successful fitting takes up to 2 min, depending on the number of time–frequency windows and higher mode content. Therefore, our data set of 3.6 million Rayleigh wave vertical-component seismograms can be processed using 3000–5000 CPU hr on a single 12-core high-performance server.

Given the computational efficiency of AMI, we elected to reprocess our full data set several times using different *a priori* settings, so as to examine their impact on the waveform fitting procedure and resulting tomographic models. We tested, first, the impact of near-nodal propagation on waveform fitting, and, secondly, the effect of the upper frequency cut-off (limiting the highest Gaussian filter) imposed in waveform fitting. A detailed description of these different tests is given in Appendix A, the results of which are summarized in Table A1.

In this work, we focus our attention on the vertically polarized shear speed structure, and therefore only utilize the vertical-component (Rayleigh wave) seismogram fits. Within windows selected for these 685 000–847 000 seismogram fits (depending on the constraints applied during fitting, see Appendix A), there are 2.9–3.6 million fundamental-mode and 226 000–409 000 higher mode successfully fit wave trains (time–frequency windows). We note that although the 'fundamental-mode windows' are dominated by the fundamental mode, they often include substantial higher mode energy as well, which adds resolving power to the signal within the windows. For simplicity during discussion, we will treat these windows as containing the fundamental mode only.

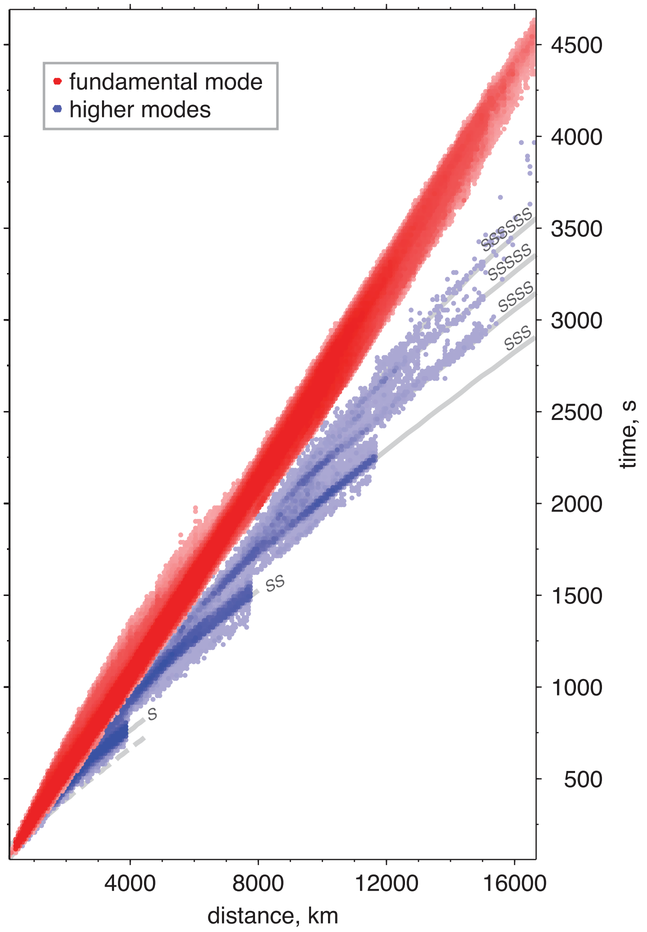

Figure 7. Estimated arrival times plotted as a function of epicentral distance for the 3.14 million fundamental- and 330 000 higher mode wave trains, all successfully fit using JWKB synthetics (data set B, Table A1), indicated by red and blue colours, respectively. The arrival times were measured at the maxima of the signal envelope within all the time–frequency windows. Darker shades indicate a greater density within a particular distance-time bin. Travel time curves of the *S* and multiple-*S* waves for a surface source in AK135 are plotted in solid grey lines, up to S_6. The traveltime curve for an *S* wave generated at a depth of 650 km is shown as a dashed line, to illustrate the range in arrival times expected due to source-depth variations.

In Fig. 7, the distribution of arrival times for all wave trains is plotted as a function of epicentral distance. Arrival times were computed at the envelope maximum within each time–frequency window, and were binned to illustrate their relative density. The fundamental-mode arrivals, shown in red, plot along a straight line across the range of epicentral distances. The higher mode arrivals plotted in blue tend to cluster around the predicted *S* and multiple-*S* (up to S_6 plotted) traveltime curves computed for AK135 (surface source). Although not clearly evident in this figure (due to the binning), there are a handful of seismograms with time–frequency windows corresponding to the S_7 branch. The sharp cut-offs seen for the direct *S* wave and the first two multiples (*SS* and *SSS*) were purposely enforced during fitting, to exclude body waves that bottom too deep in the lower mantle and have smaller sensitivity to upper-mantle structure, which is the focus of this study. The cut-offs are at 35° for *S*, 70° for *SS*, 105° for *SSS* and 140° for *SSSS*.

Based on the detailed analysis and discussion in Appendix A, the data set selected for inversion in the final tomographic model is made up of 521 705 vertical-component seismograms, whose waveform fits were computed without enforcing any upper frequency limit. These were selected from our master data set of more than 710 000 vertical-component waveform fits, using a rigorous process of outlier analysis (Section 2.4), including a final manual selection and removal (Appendix A3).

4 RESOLUTION ANALYSIS

It is common practice in tomographic studies to perform a series of resolution tests to assess the accuracy of the model. The tomographic resolution may depend on the data sampling, noise, regularization and *a priori* information (e.g. background model). However, conventional (checkerboard or spikes, for example) or resolution-kernel tests are intrinsically limited, as they are carried out assuming the same theoretical approximations as in the inversion of the data, and therefore do not examine methodological inaccuracies (Qin *et al.* 2008).

With the expansion of computing resources over the last decade, more exact methods are available to examine the full resolving power of models, including the methodological and theoretical foundations. One such method is the generation of benchmark seismic data sets. These consist of an arbitrarily complex synthetic model, through which seismograms are computed between synthetic sources and receivers. Recent techniques such as the spectral-element method (SEM; Komatitsch & Vilotte 1998; Chaljub *et al.* 2003) and the coupled spectral-element method (CSEM; Capdeville *et al.* 2003) are capable of simulating seismic wave propagation in heterogeneous 3-D anisotropic media in period ranges comparable to those used in waveform tomography (Qin *et al.* 2008). With access to large modern clusters, synthetic data sets with several thousands of seismograms can be generated and used to test and benchmark tomographic methods.

The resolving capability and accuracy of AMI has previously been examined through inversion of two benchmark data sets by Qin *et al.* (2006, 2008). The first benchmark inverted ∼3000 successfully fit CSEM synthetic seismograms, computed through a smooth global isotropic model (Lebedev & van der Hilst 2008). The results from this procedure confirmed the validity of the assumptions and approximations used by AMI and the subsequent inversion. At lithospheric depths, both the shape and amplitude of the anomalies were recovered, whereas at base of the transition zone amplitudes were underestimated by up to a factor of two. Lebedev & van der Hilst (2008) suggest that such underestimation could be resolved through explicit modelling of the sensitivity volumes (e.g. Meier *et al.* 1997). Alternatively, improvements in amplitude recovery may be achieved through the incorporation of more multiple-*S* higher mode constraints to improve sampling within the transition zone.

Several years later, Qin *et al.* (2008) constructed a more complex model containing a suite of ‘quasi’-realistic heterogeneities in velocity, radial and azimuthal anisotropy, attenuation and density, spanning a range in spatial scales. The authors then compared the results from two different global tomographic inversion techniques: AMI (simplified, for that application, to using only a 1-D reference model) and the more traditional three-step phase-velocity method. From this, it was clear that the use of an accurate 3-D crustal model was very important to prevent artefacts in the mantle due to smearing of unmodelled crustal structure. The higher modes were indispensable for retrieving deeper structures, at the base of the upper mantle and in the transition zone. The greater heterogeneity present in this new model further illustrated that the location of anomalies was often better constrained than the shape of the boundaries. This would be improved using a larger synthetic data set, with coverage more comparable to that of a real global data set.

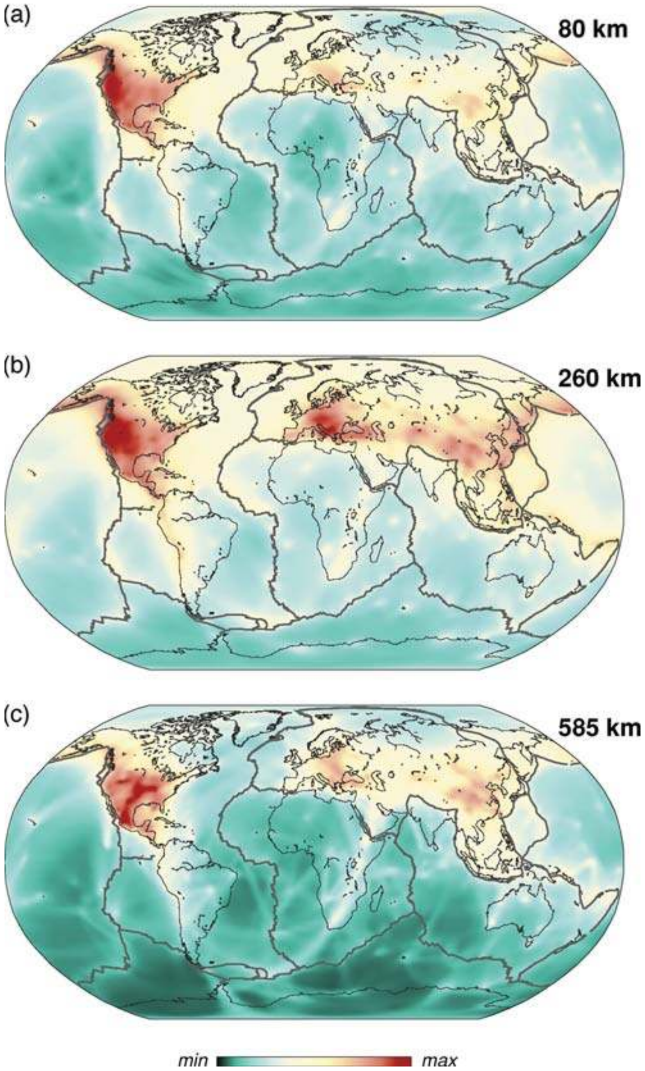

Figure 8. Relative lateral sampling at three different depths within the model. The colour scale is scaled to the min and max value for each depth. In no case is the sampling of any model node zero. Sampling is estimated using matrix column sums of the S_V parameter.

Having confirmed AMI’s ability to accurately invert for smooth and complex benchmark models, we now assess the resolving power of our current model and data set. We begin with the relative path coverage using reweighted matrix column-sums for the isotropic S_V parameter, plotted for depths of 80, 260 and 585 km in Fig. 8. Across all depths in the model, every shell node is sampled. Blue–green colours indicate lower sampling density, while red colours indicate the highest sampling density; the colour scale is normalized for each depth.

At shallow depths in the lithosphere (Fig. 8a), sampling is densest (red) beneath North America, Europe and eastern Asia. At 260 km depth (Fig. 8b), the relative sampling of Eurasia increases, becoming almost equivalent to that of North America. Within the transition zone (Fig. 8c), sampling remains strongest beneath the North American and Eurasian plates. Despite the efforts of normalization and re-weighting, several paths are evident in Fig. 8c, as bands of elevated sensitivity in lesser sampled regions. Several examples include paths to stations in the western Pacific and the South African Seismic Experiment (SASE), and paths from events along the Mid-Atlantic Ridge.

In the upper mantle, sampling is good across almost all the Northern Hemisphere continental regions and northern Atlantic Ocean (yellow colours), with weaker sampling in northern Africa. Oceanic regions and the Southern Hemisphere are relatively less well sampled. However, in all cases the sampling is non-zero and represents an increase over past global modelling efforts.

Finally, we discuss the results from a series of four different spike resolution tests illustrated in Fig. 9. The input models consist of varying width columnar perturbations of ± 300 m s^{-1}. Model A is simplest and consists of columns 6° in diameter spread around the globe. Model B has columns 10° in diameter between latitudes ±60° and 6.6° outside. Model C has columns 18° in diameter centred at latitudes of ±90°, ±45° and 0°; anomalies 12° in diameter are centred at ±60° latitude. The last model D has the largest columns, with diameters of: 30° at 0° latitude, 25° at ±30° latitude and 20° at ±60° and ±90° latitude.

Synthetic data were generated for each of the different models, and each inversion was run until convergence, with the same regularization as the real-data inversion. We plot the resulting models at three different depths, to illustrate the model's resolving capabilities. At 80 km depth, both model A and B anomalies are reliably recovered beneath the well-sampled continental regions, including north and central America, Europe and eastern and southeast Asia. The strength of the anomalies is reduced in regions with lesser sampling, for example, the Pacific Ocean. In the Pacific, some anomalies also exhibit a degree of smearing, more apparent for model A than B.

At 260 km depth, models B and C are illustrated. For model B, again the strength and shape of the anomalies are well recovered beneath regions of highest station density, as well as most continental regions. Anomalies in the mid-Pacific are somewhat underestimated. For model C, the amplitude and shape of anomalies are recovered properly, including those in the Pacific Ocean.

At 485 km depth, we show the results for models C and D. As with previous depths, the anomalies are accurately retrieved beneath densely sampled continental regions, particularly North and South America, most of Eurasia and southeast Asia. However, the anomalies are less well resolved beneath the Pacific Ocean, Africa and the Southern Hemisphere below 45°latitude, as would be expected based on the path sampling estimates shown in Fig. 8.

Overall, the results from the resolution tests indicate that the model is well resolved at a variety of length scales, in particular beneath continental regions. Features with dimensions of 6° are clearly recovered at lithospheric depths beneath North America, Eurasia and southeast Asia; in these densely sampled areas, smaller scale features would easily be recovered. Larger scale features remain accurately retrieved at depths into the transition zone. In more poorly sampled oceanic and some continental regions, although the anomalies are still recovered, their strength is underestimated and suffer from a degree of distortion (due to oversmoothing relative to the sparser data sampling).

5 MODEL

Our new isotropic global upper mantle and transition zone S_V velocity model, SL2013sv, is computed on a ~280 km (minimum 250 km, maximum 296 km) triangular grid using 521 705 vertical-component broad-band seismograms selected from our master data set of almost 3/4 million. The resulting inverse problem consisted of 1.55 million data equations and 1.47 million smoothing and damping constraints to solve for 501 888 unknown model parameters (7842 shell nodes × 18×3 S parameters and 10 P parameters). The final model has a variance reduction of 90 per cent with respect to our 3-D reference model. The increased quantity of data provides the ability to decrease grid spacing, targeting higher resolution compared to past models.

In Figs 10–13, we plot horizontal slices globally at 12 different depths through the model: 36, 56 and 80 km (Fig. 10); 110, 150 and 200 km (Fig. 11); 260, 330 410 km (Fig. 12) and 485, 585, 660 km (Fig. 13). In Figs 14 and 15, 12 vertical cross-sections slice through various parts of the model, with inset maps indicating their locations. White circles overplotted indicate seismicity within 40 km laterally from the profile. Slices A–C cross Africa and western Eurasia; D crosses eastern Eurasia and the western Pacific; E and F slice the Mid and south Pacific from west to east, G–J focus on the North American continent; and finally K and L cross South America. For both the horizontal and vertical images, perturbations in shear velocity are plotted with respect to our reference model. In the horizontal slices, the indicated reference velocities are extracted from the 1-D mantle reference model (solid black line, Fig. 3). At depths greater than the Moho, perturbations are in per cent from the reference. It is important to note that at depths shallower than the Moho, model perturbations are in m s^{-1} relative to the 3-D crustal model (depth slices at 36 and 56 km indicate the range in m s^{-1} in addition to percentage). Although this makes interpretation of the magnitude of velocity perturbations in crustal regions more complex, relative variations are still readily interpreted in terms of structure. For all vertical cross-sections, perturbations are in m s^{-1}, with values indicated in the captions.

Long-wavelength, lithospheric-depth features in our new model are in agreement with observations from past models that use different methodological approaches and parametrizations, as well as differing types and sizes of data sets (e.g. Debayle *et al.* 2005; Panning & Romanowicz 2006; Simmons *et al.* 2006; Zhou *et al.* 2006; Houser *et al.* 2008; Kustowski *et al.* 2008a; Lebedev & van der Hilst 2008; Nettles & Dziewóński 2008; Ferreira *et al.* 2010; Lekić & Romanowicz 2011; Ritsema *et al.* 2011; Debayle & Ricard 2012). However, at greater depths (e.g. in the transition zone) variations between models are large even at long wavelengths (Ritsema *et al.* 2011). In Section 6.1, we will examine the differences between our new model and five other global tomographic models.

In our model, we observe improvements in the resolution of fine-scale regional structures. The prominent features in our model display deep expressions of regional tectonic structures and processes. We observe sharp velocity contrasts across many tectonic boundaries, for example, subduction systems and associated backarc volcanics, actively deforming regions and continental orogens. The strongest velocity anomalies in the model are associated with stable continental cratons (positive), mid-ocean ridges (MORs) and rift systems (negative) and backarcs and active orogens (negative).

In the continental crust, strong perturbations of more than 350 m s^{-1} (from the 3-D reference) are observed beneath the Himalaya and Tibet, the Hangai Dome (western Mongolia), the Afar Depression, the Pamirs, southern Alaska and the Yukon and

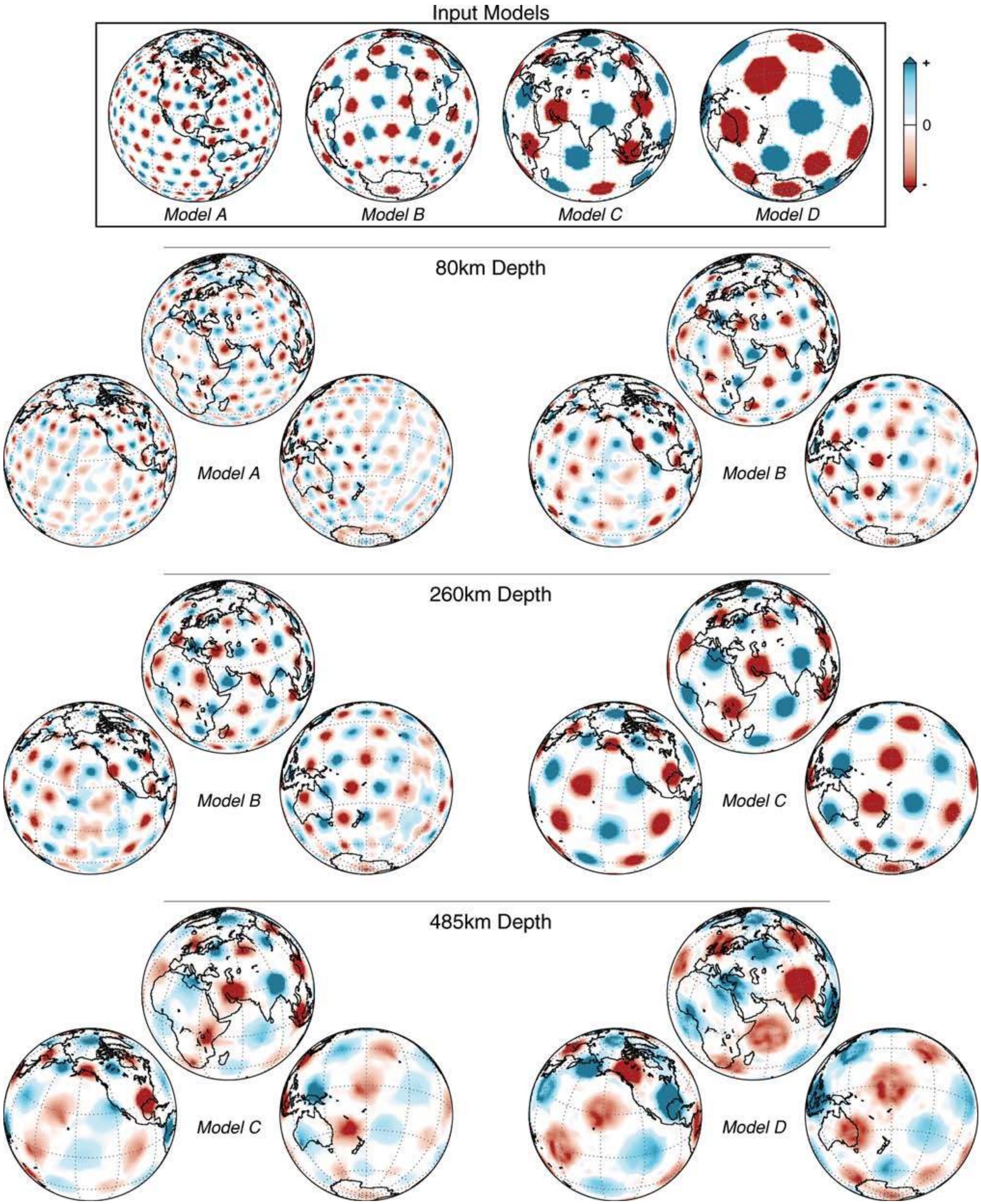
Input Models
Model A
Model B
Model C
Model D
+
0
-
80km Depth
Model A
Model B
260km Depth
Model B
Model C
485km Depth
Model C
Model D

Figure 9. Synthetic resolution tests illustrating the sensitivity of the final model. The top panel shows the four different input models A, B, C and D. Each consists of columnar perturbations of $\pm 300\,\mathrm{m\,s^{-1}}$ with varying dimensions. Model A has columns 6° in diameter; model B has columns 10° in diameter between ±60° and 6.6° outside ±60°. Model C has columns 18° in diameter centred at ±90°, ±45°, 0° latitude and 12° anomalies centred at ±60° latitude; model D consists of 30° diameter columns at 0° latitude, 25° diameter at ±30° and 20° diameters at ±60° and ±90°. The three lower panels show the resulting inversions at 80 km depth for models A and B, 260 km depth for models B and C and 485 km depth for models C and D.

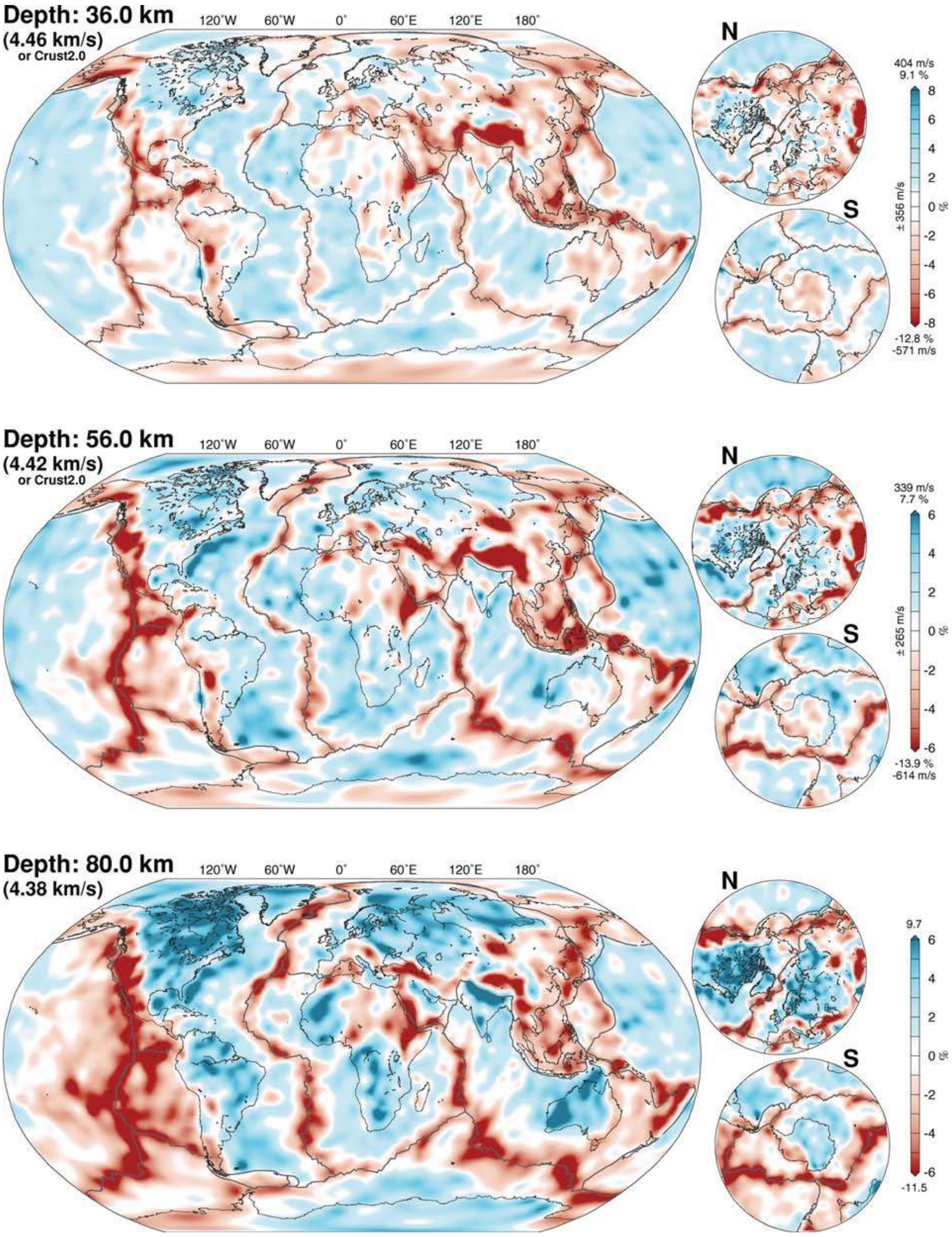
Depth: 36.0 km
(4.46 km/s)
or Crust2.0
120°W
60°W
0°
60°E
120°E
180°
N
S
404 m/s
9.1 %
8
6
4
2
0
-2
-4
-6
-8
± 356 m/s
%
-12.8 %
-571 m/s
Depth: 56.0 km
(4.42 km/s)
or Crust2.0
120°W
60°W
0°
60°E
120°E
180°
N
S
339 m/s
7.7 %
6
4
2
0
-2
-4
-6
± 265 m/s
%
-13.9 %
-614 m/s
Depth: 80.0 km
(4.38 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
9.7
6
4
2
0
-2
-4
-6
%
-11.5

Figure 10. Horizontal cross-sections through the tomographic model SL2013sv at three depths in the shallow upper mantle (and crust in some continental regions). Approximate plate boundaries are indicated. The reference S_V velocity values (at a reference period of 50 s) are indicated. Perturbations from the reference are indicated in percentage, with the absolute minimum (maximum) indicated below (above) the colour bar. Note that at 36 and 56 km, some continental regions are still in the crust, therefore perturbations are indicated in m s^{-1} (colour scale range and absolute minimum and maximum are labelled), relative to the 3-D reference model. North and south polar views are labelled at right.

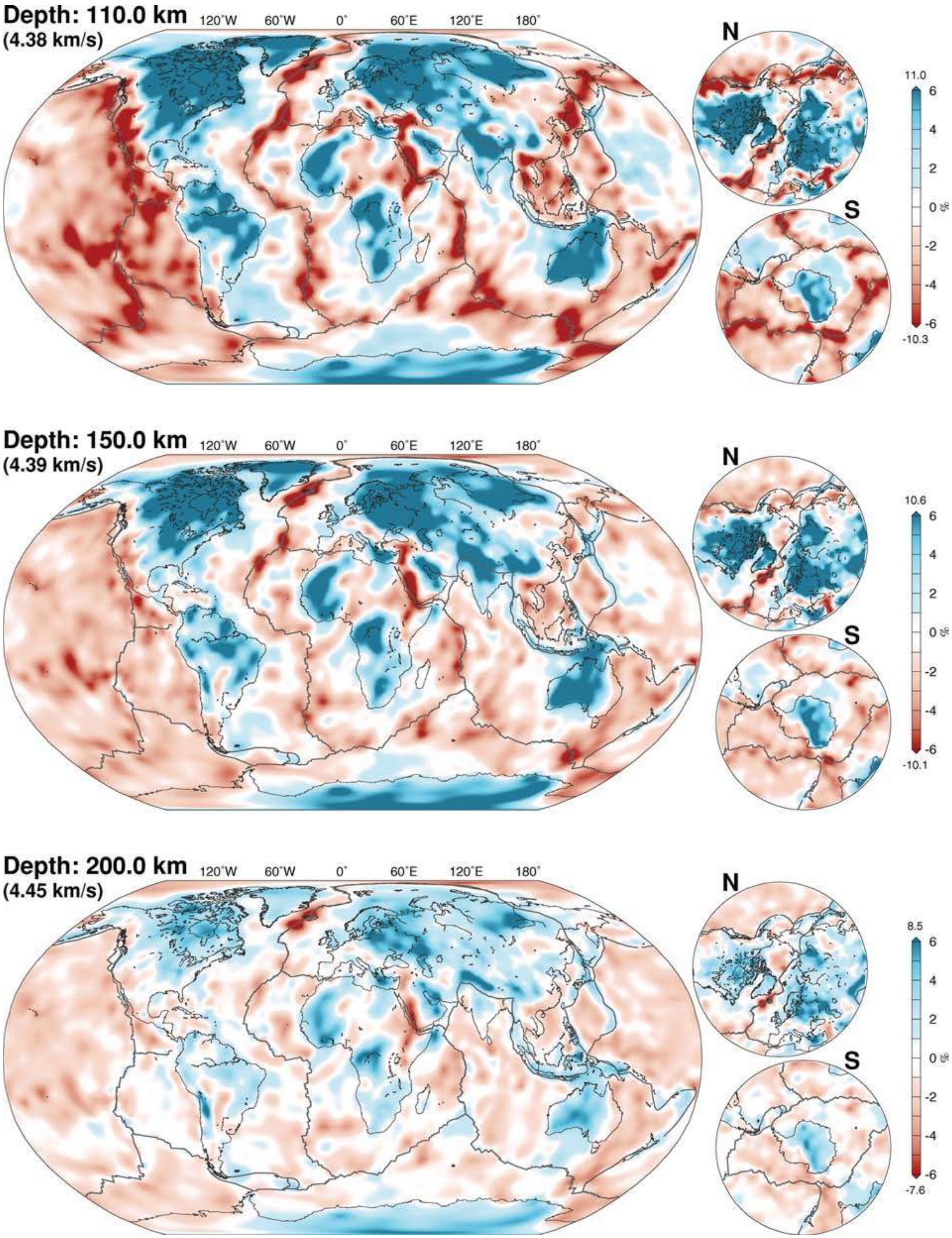
Depth: 110.0 km
(4.38 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
11.0
6
4
2
0 %
-2
-4
-6
-10.3
Depth: 150.0 km
(4.39 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
10.6
6
4
2
0 %
-2
-4
-6
-10.1
Depth: 200.0 km
(4.45 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
8.5
6
4
2
0 %
-2
-4
-6
-7.6

Figure 11. Horizontal cross-sections through SL2013sv at three depths in the lithospheric mantle. Plate boundaries and reference velocities follow as in the previous figure.

western United States. The largest crustal and shallow-mantle anomalies beneath oceans are associated with backarcs of the western Pacific and spreading ridges (most notably in the east Pacific).

5.1 Oceanic regions

In the upper ∼120 km beneath oceans (Figs 10 and 11), the most apparent feature of the model is the clear signature of the low-velocity

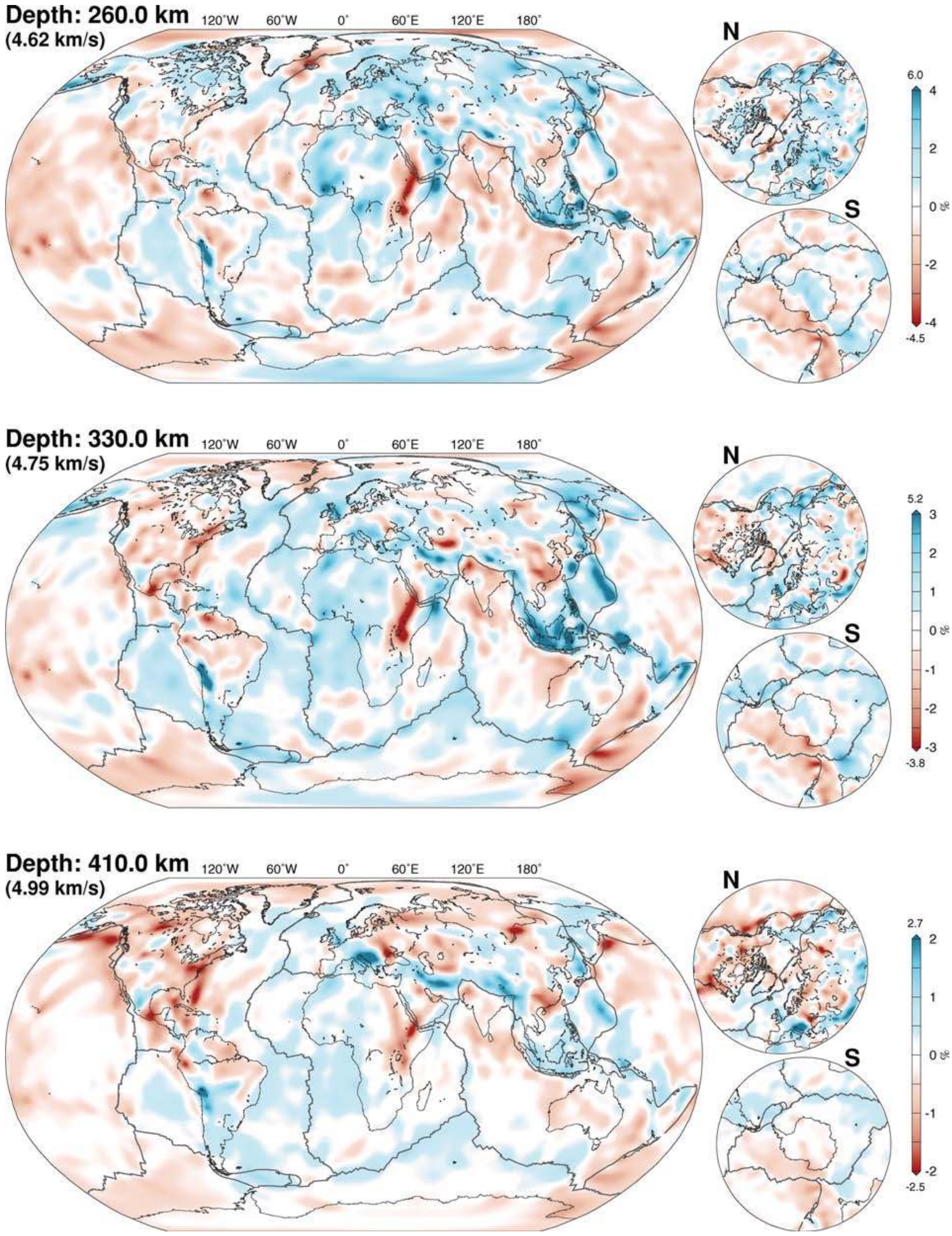
Depth: 260.0 km
(4.62 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
6.0
4
2
0 %
-2
-4
-4.5
Depth: 330.0 km
(4.75 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
5.2
3
2
1
0 %
-1
-2
-3
-3.8
Depth: 410.0 km
(4.99 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
2.7
2
1
0 %
-1
-2
-2.5

Figure 12. Horizontal cross-sections through SL2013sv at three depths in the lower upper mantle and top of the transition zone.

anomalies associated with spreading at the MORs. Their width increases as a function of depth (South Atlantic Ridge, slices *KK*′ and *LL*′ in Fig. 15), as would be expected based on a simple triangular decompression melting model. The more rapidly spreading East Pacific Rise ridge system is wider than others (i.e. the Mid-Atlantic or southwest Indian ridges).

At 110 km depth, the strongest anomalies beneath the ridges begin to become more localized, and by 150 km depth, the signature of

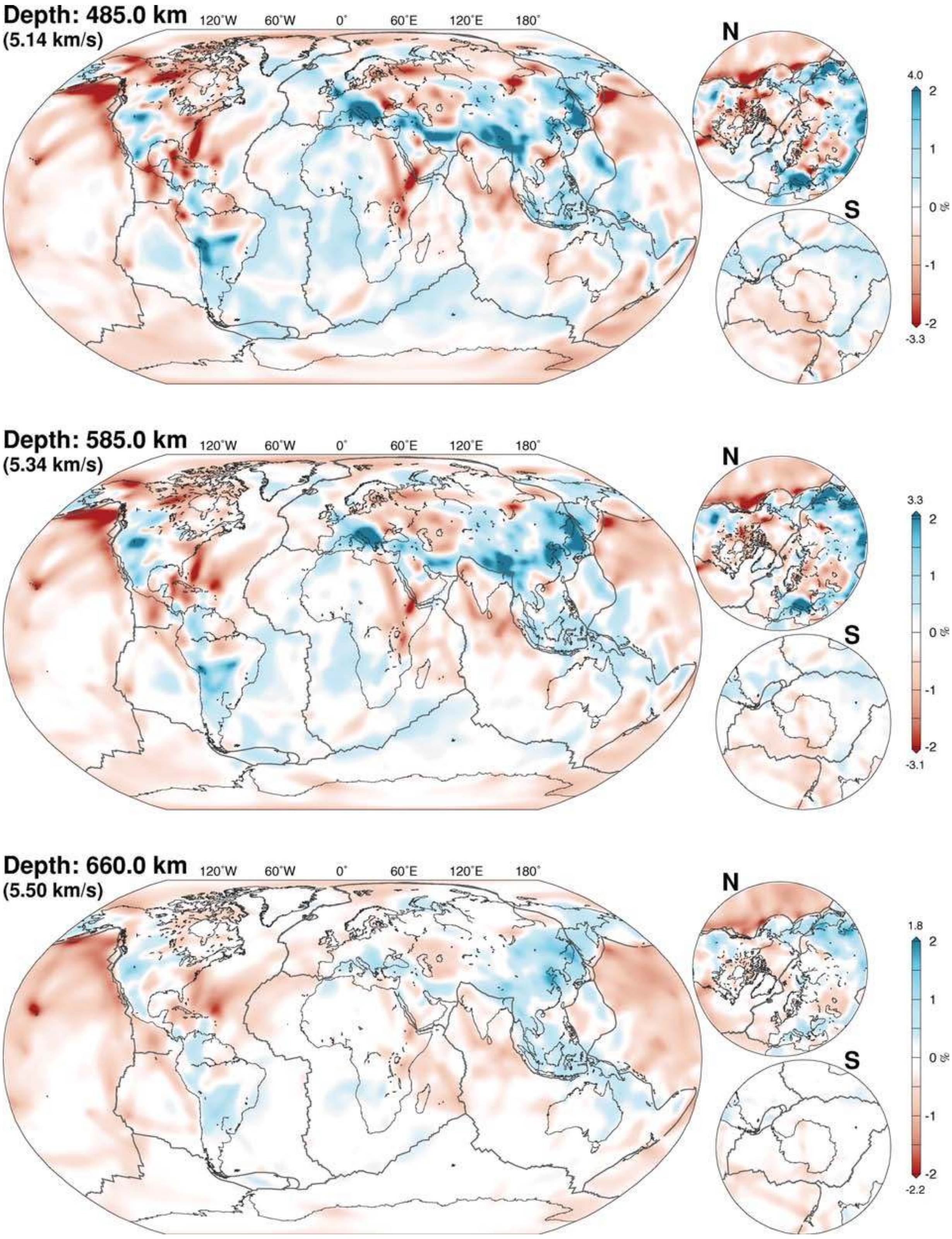
Depth: 485.0 km
(5.14 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
4.0
2
1
0 %
-1
-2
-3.3
Depth: 585.0 km
(5.34 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
3.3
2
1
0 %
-1
-2
-3.1
Depth: 660.0 km
(5.50 km/s)
120°W
60°W
0°
60°E
120°E
180°
N
S
1.8
2
1
0 %
-1
-2
-2.2

Figure 13. Horizontal cross-sections through SL2013sv at three depths within the transition zone.

most MORs no longer stands out from the lower velocities observed across the rest of the ocean basins. Therefore, we conclude that in most cases, significant partial melting beneath MORs is confined to depths less than ∼120 km, with lower degree melting at greater depths no longer visible in vertically polarized shear velocity. This is in agreement with some past studies (e.g. Zhang & Tanimoto 1992; Forsyth *et al.* 1998), but does not confirm inferences from others regarding MOR anomalies and processes extending into the

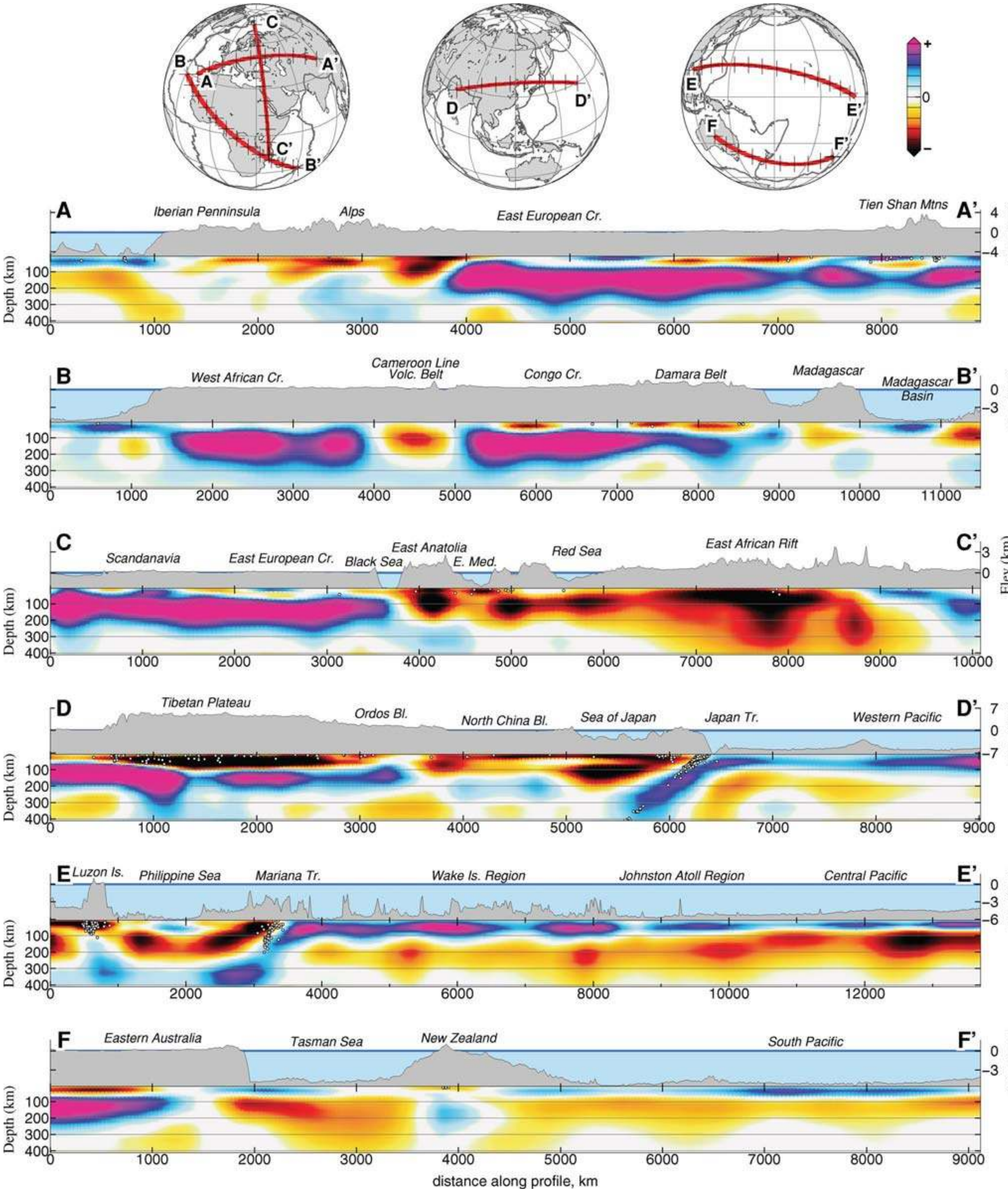

Figure 14. Vertical cross-sections of six profiles through SL2013sv. The location of each section is indicated in the maps at the top. Model is plotted from the shallowest model node (7 km) to a depth of 410 km. Elevation/bathymetry is indicated at right, and is smoothed from ETOPO1 (Amante & Eakins 2009). Velocity perturbations for each section are: A ±240, B ±240, C ±240, D ±240, E ±180 and F ±240 m s^{-1}.

deep upper mantle (e.g. Su *et al.* 1992). Where slow anomalies do remain below depths of 150 km, often they are coincident with oceanic islands.

Our new model has made significant improvements in the lateral definition of the MOR anomalies. The central low-velocity anomalies, due to decompression melting associated with hot upwelling mantle, are narrowly confined beneath the ridge spreading centre. Sharp lateral boundaries between these anomalies and the smoothly varying shear speeds in surrounding oceanic lithosphere and asthenosphere are marked contrasts.

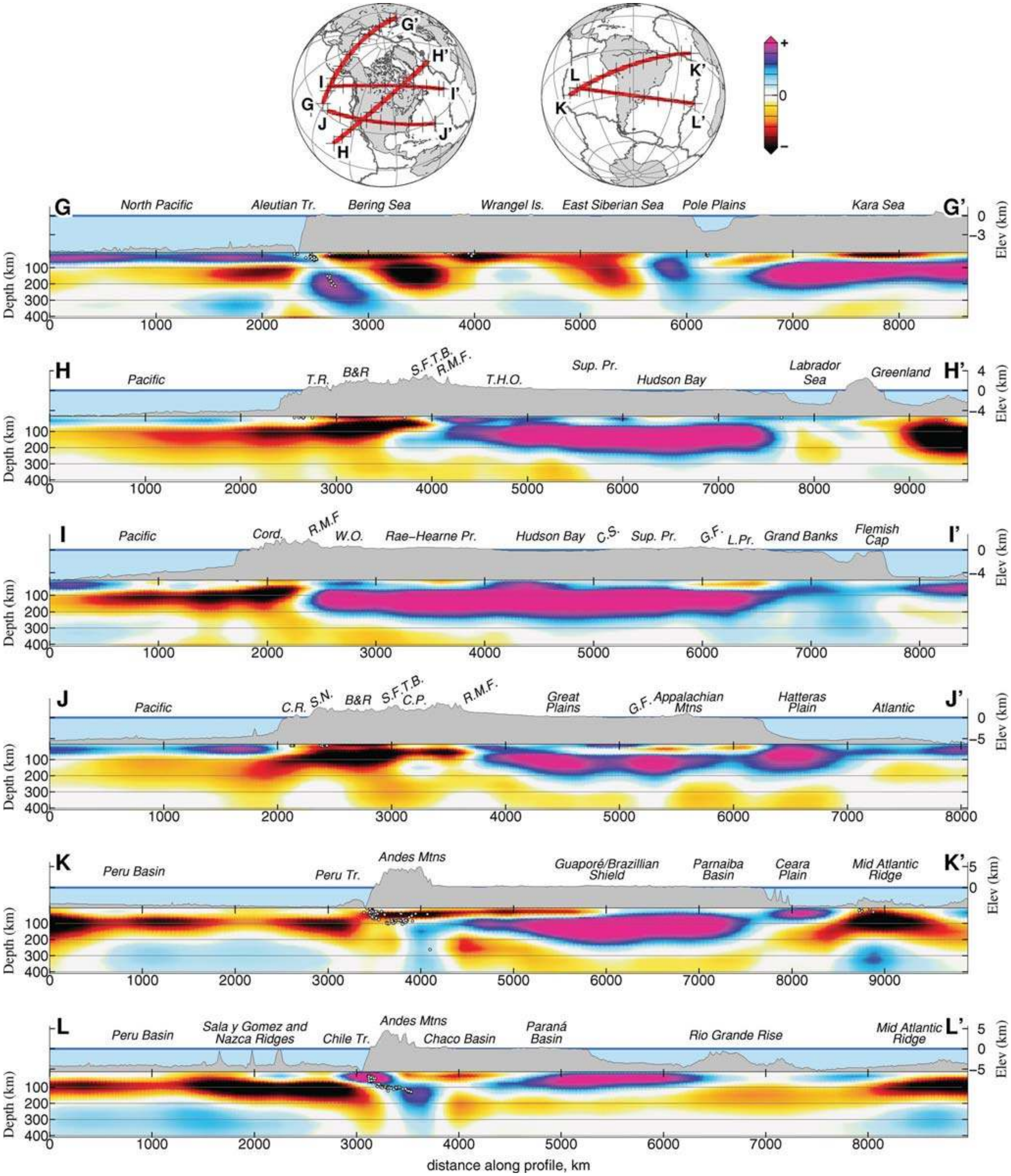

Figure 15. Vertical cross-sections of six additional profiles through SL2013sv. As in previous, the location of each profile is indicated in the maps at the top. Velocity perturbations are: G ±180, H ±240, I ±240, J ±240, K ±180 and L ±180 $\mathrm{m\,s^{-1}}$. Labels are: B&R, Basin and Range; Cord., Cordillera; C.P., Colorado Plateau; C.R., Coast Ranges; C.S., Cape Smith Belt; G.F., Grenville Front; L.Pr., Labradonian Province.; R.M.F., Rocky Mountain Front; S.F.T.B., Sevier Fold and Thrust Belt; S.N., Sierra Nevada; Sup. Pr., Superior Province; T.H.O., Trans-Hudson Orogen; T.R., Transverse Ranges; W.O., Wopmay Orogen.

Away from the MORs in the oceans, we observe relatively high velocities, with older regions remaining fast to greater depths, consistent with cooling-induced thickening of the oceanic lithosphere. This can be seen clearly seen across the Pacific Basin, and is qualitatively similar to the observations of Maggi *et al.* (2006). The leading eastward edge of oceanic lithosphere (transition from red to blue moving west from the Pacific MOR system) progresses westwards across the Pacific with increasing depth (from 36 to 150 km).

At 150 km depth we observe the deepest fast anomaly associated with the ancient western Pacific, immediately east of the Marianas trench; by 200 km, this anomaly is gone. This age progression of the lithospheric thickness is clear in the vertical cross-section *EE*′ (Fig. 14) through the Pacific, with the lithosphere thinning eastwards from the trench.

The backarc regions near ocean-0ocean and ocean–continent convergent boundaries are characterized by low-velocity anomalies, albeit relatively weaker than those beneath MORs. The most prominent backarc anomalies at shallower depths (∼80 km) include Tonga–Kermadec, New Hebrides and Indonesia–Sumatra–Java; they fade gradually with depth down to around 150 km and are weak or absent at greater depths (200 km and below). The anomalies beneath the Mariana, Izu-Bonin, Japan, Kuriles and Aleutians volcanic arcs are similar in strength at 110 km. A number of these anomalies are visible in the vertical cross-sections (*DD*′, *EE*′, *GG*′, *KK*′ and *LL*′, Figs 14 and 15). Due to the much higher saturations used here to display deeper structures (where perturbations are much smaller) clearly, the volcanic-arc low-velocity anomalies appear to extend deeper, and along the subducting plate interface.

At 150 km depth, a system of prominent high-velocity anomalies indicates subducting oceanic lithospheric slabs. This is seen in much of the western Pacific, from the Aleutians (*GG*′, Fig. 15) through the Kuriles, Japan (*DD*′, Fig. 14), Izu-Bonin, Mariana (*EE*′, Fig. 14), to Indonesia-Sumatra, as well as the Hikurangi (*FF*′, Fig. 14) and portions of the Andean (*LL*′ and *KK*′, Fig. 15) . At 200 km depth and below, these subducting slabs are even clearer, with smaller anomalies associated with the subduction at the Cascadia, Lesser Antilles, Scotia and other arcs also apparent.

5.2 Continental regions

At depths less than 200–260 km beneath continents (Figs 10 and 11), the strongest low-velocity anomalies are associated with tectonically active regions undergoing deformation. One of the strongest of the anomalies is beneath the Himalayas and the Tibetan Plateau. The anomaly boundaries at crustal depths closely match the boundaries of the plateau at the surface; the very low velocities within the mid-lower crust are consistent with pervasive partial melting in it (Nelson *et al.* 1996). In the mantle beneath, high-velocity anomalies beneath much of the plateau probably indicate the underthrusting and subduction of the Indian lithosphere (80–200 km depth and below), the nature of which varies with position along the thrust.

The region of shallow low velocities underlying Tibet appears to be part of a much broader zone of convergence and deformation, which originates beneath Burma to the South, and extends westwards, almost continuously, through the Zagros Mountains, the Anatolian Plateau, into the Aegean Sea, and northwards towards the Alps and Pannonian Basin.

Other prominent low-velocity anomalies include the Cameroon Line volcanic belt, which bisects the African continent to depths of 200–260 km (centred at ∼4500 km along profile *BB*′, Fig. 14), as well as the signatures of the Red Sea and East African rifts. We image the East African Rift extending through the upper mantle approaching the transition zone (profile *CC*′, Fig. 14).

Finally, we also image the structure of a pervasive low-velocity anomaly underlying the western margin of North America, which is much younger than the rest of the continent and is undergoing active deformation. The high station density (USArray) means this portion of the model is well resolved, and the structures observed are robust. We clearly image this transition of low to high velocities (west to east), the Rocky Mountain Front, which separates the juvenile western margin of North America from the ancient continental backstop (Fig. 15, *HH*′, *II*′ and *JJ*′).

Notable low-velocity features in this region (Figs 10 and 11 at depths of 56–150 km) include the Snake River Plain volcanic belt (as imaged previously in, e.g. Tian *et al.* 2009) and extensional Basin-and-Range province (additionally in *HH*′ and *JJ*′, Fig. 15). At depths greater than 150 km, neither feature stands out; however, the western margin does remain distinct (low velocity) from continental North America, through depths to the base of the continental lithosphere. The base of the main low-velocity anomalies appears to terminate more sharply and at shallower depths than in most past surface wave models (e.g. Lebedev & van der Hilst 2008; Kustowski *et al.* 2008a; Nettles & Dziewónski 2008; Lekić & Romanowicz 2011; Ritsema *et al.* 2011).

Large-scale high-velocity anomalies in the uppermost 250 km beneath continental regions have been recognized as the signatures of ancient continental cratons in global surface wave tomographic models for more than 25 yr (e.g. Woodhouse & Dziewónski 1984). These anomalies are the dominant high-velocity features in our new model. The difference between our model and other recent global tomography models is in the relative roughness of the craton margins and fine structure within the cratons, which we resolve particularly well in densely sampled regions such as North America, Europe and eastern Asia.

By 260 km depth, there are very few high-velocity seismic anomalies remaining beneath cratons. We can therefore conclude that the thickness of the high-velocity lithospheric roots beneath cratons is unlikely to exceed ∼200–220 km depth in most cases (Lebedev & van der Hilst 2008; Debayle & Ricard 2012). Several vertical cross-sections in Figs 14 and 15 bisect such ancient Archean cratons, including in eastern Europe (*AA*′ and *CC*′), central Australia (*FF*′), North America (*HH*′, *II*′ and *JJ*′) and South America (*KK*′ and *LL*′).

5.3 Sublithospheric mantle and transition zone

In the depth range from 260 km to 660 km (Figs 12 and 13), the most prominent high-velocity anomalies in the model are beneath areas of past and current subduction. The most conspicuous of these are the various subduction zones in the western Pacific (*DD*′ and *EE*′), which are seen from shallower depths (Section 5.1) to the transition zone. Also more clearly evident is subduction of the south–central Nazca Plate beneath South America (*KK*′ and *LL*′), where the highest velocities are located beneath Peru, Bolivia, Chile and Argentina. With increasing depth into the TZ, the signature of the plate spreads out laterally across much of central and southern South America; similar observations are noted in northern South America, where a clear signature of the Nazca Plate is imaged beneath Columbia, Venezuela and northern Brazil.

Beneath North America, the distribution of high-velocity anomalies reflects the complex history of subduction. At upper-mantle depths (200–410 km), the Juan de Fuca Plate is imaged subducting beneath the Cascades (British Columbia, Oregon and Washington). Within the transition zone itself, we image fragments of both the Juan du Fuca and Farallon plates subducted over the last 150 Myr (e.g. Sigloch *et al.* 2008; Tian *et al.* 2009). Towards the base of the transition zone, we image the signature of the Farallon Plate extending across much of the continental US, as far east as the Great Lakes (660 km depth, Fig. 13).

In addition to the Farallon Plate beneath the US, we also image a similar high-velocity feature beneath western and central Canada (mostly British Columbia, Alberta and Saskatchewan). This is imaged more clearly and detached from the slabs to the south than in past models of North America (e.g. Frederiksen *et al.* 2001; van der Lee & Frederiksen 2005; Nettles & Dziewónski 2008; Bedle & van der Lee 2009). Previous high-resolution models of subduction beneath north America (e.g. Sigloch *et al.* 2008; Burdick *et al.* 2010) focused only on the western US, and did not extend northwards into Canada.

Although not as high velocity as the signals associated with the subducting plates, the central and south–central Atlantic ocean maintains a small-to-moderate positive velocity anomaly from 260 to 585 km depth. Such a feature was documented in S20RTS (Ritsema *et al.* 2004), and was speculated by King & Ritsema (2000) to be the signature of edge-driven convection. A similar feature can be observed in a number of more recent models (e.g. Lekić & Romanowicz 2011; Ritsema *et al.* 2011; Debayle & Ricard 2012), with some variations in its depth and horizontal location. The presence of this anomaly in models using differing parametrizations and data sets suggests that it is a robust, though low-amplitude, structure.

A strong, continuous band of high velocities is observed in the transition zone, stretching from western and central Europe eastwards, through Anatolia and into the Tibetan Plateau. It is the strongest in the mid-transition zone, becoming more diffuse at the base. In comparison with other surface wave models, such fast velocities have been observed, though not as continuously (Kustowski *et al.* 2008a; Lekić & Romanowicz 2011; Debayle & Ricard 2012). We conclude that this high-velocity material likely represents the final fragments of ocean basins, continental lithospheres and portions of continental margins subducted after the closure of the Tethys Ocean. The subducted oceanic Tethyan lithosphere itself is already almost entirely well within the lower mantle, as has been previously imaged in teleseismic *P*-wave traveltime tomography (Bijwaard *et al.* 1998; Van der Voo *et al.* 1999; Amaru 2006; Hafkenscheid *et al.* 2006).

6 DISCUSSION

In the following sections, we examine our new data set and model from a number of different perspectives. First, we present a comparison of SL2013sv with five recent, published global models. In the next section, we examine the bulk dispersive properties of the Earth's heterogeneous upper mantle and crust as sampled by our data set of more than 700 000 fundamental- and 475 000 higher mode group- and phase-velocity curves. Finally, we leverage the superior statistical sampling of this data set to re-examine the validity field of the JWKB approximation and the overall success rate of waveform fitting using AMI.

6.1 Comparison with other global models

We have compared our new model SL2013sv with five other global shear velocity models in Fig. 16: CUB (Shapiro & Ritzwoller 2002), DR2012 (Debayle & Ricard 2012), SEMum (Lekić & Romanowicz 2011), S362ANI (Kustowski *et al.* 2008a) and S40RTS (Ritsema *et al.* 2011). Each of these models is computed with different data sets and modelling methodologies. The mean was removed at each depth and model perturbations were plotted in per cent from this value. The limits of the (saturated) colour scales are indicated at the left of each row (e.g. −8 to +8 per cent at 100 and 150 km depth); beneath each map the total range is indicated. The models are ordered from left to right by decreasing peak-to-peak perturbations at 100 and 150 km depth.

Model DR2012 is an upper-mantle S_V-wave model constrained by multimode Rayleigh wave seismograms, using an approach similar to that used in generating SL2013sv. The CUB model (specifically CU_SDT1.0) is a crust and upper-most mantle isotropic shear velocity and radial anisotropy model computed from fundamental-mode Rayleigh and Love group and phase measurements. SEMum is a global upper-mantle Voigt-average shear speed and radially anisotropic model derived from long-period seismic waveforms (multimode Rayleigh and Love waves and long-period body waves) and group-velocity dispersion maps. Model S362ANI is a whole-mantle Voigt-average isotropic shear velocity model generated using surface wave dispersion measurements, mantle and body wave waveforms and body wave traveltimes. Finally, S40RTS is an isotropic shear velocity model of the mantle constrained by three data sets: minor and major arc Rayleigh wave dispersion (fundamental and first four overtones), teleseismic body wave traveltimes and spheroidal mode splitting functions. For each model, we plot the S_V component.

In the uppermost mantle, from 50 km (partially crust) to 150 km depth, the long-wavelength (several thousands of kilometres) features are consistent across the models. For example, all show low-velocity anomalies in the eastern Pacific and higher velocities in the ancient Western Pacific. High-velocity anomalies representing the continental cratonic roots are clear in each, although the amplitude and clarity does vary (e.g. the Southern Hemisphere cratons in S362ANI and S40RTS).

At shorter length scales, there are much greater differences between the models. SL2013sv displays the highest resolution, particularly at lithospheric and crustal depths. One key difference between the models is how the crust is treated. In SL2013sv, crustal perturbations, with respect to our 3-D reference model, are solved for directly in the inversion. In continental regions, this often includes three vertical crustal knots (7, 20 and 36 km), whereas in the oceans there is commonly only one (7 km). As a result, not only are deeper mantle artefacts due to unaccounted for or assumed crustal structure prevented, but also a high-resolution crustal model is generated (depths from as shallow as 7 km). The CUB model is most similar to SL2013sv at these depths, due to the inclusion of crustal parameters in the inversion; in the other models, crustal structure tends to be much smoother, both laterally and vertically.

At 50 km depth (top row Fig. 16), the dominant features are the signature of spreading ridges, backarc basins and regions of continental deformation. Although these can be observed in each model, SL2013sv obtains the highest definition. For example, the spreading ridges are much narrower and with large perturbations (more continuous red and black colours) tightly confined near the ridge axes. Although DR2012, CUB and SEMum all show well-defined ridges, the highest anomalies are not as continuous along the spreading centres. A second feature in common between those models is the low-velocity anomaly associated with the partially molten Tibetan crust. Clearly, the structure resolved in SL2013sv is better correlated with surface tectonic boundaries, including the low velocities in the Hindu Kush and Pamirs to the west of the Himalaya, high velocities of the Tarim Basin north of the Altyn Tagh Fault, and a clear extension of the partially molten Tibetan crust southeastwards around the eastern syntaxis of the India–Eurasia collision.

At depths of 100 km, the spreading ridge anomalies are still clearly observed in each of the models, though still more narrowly

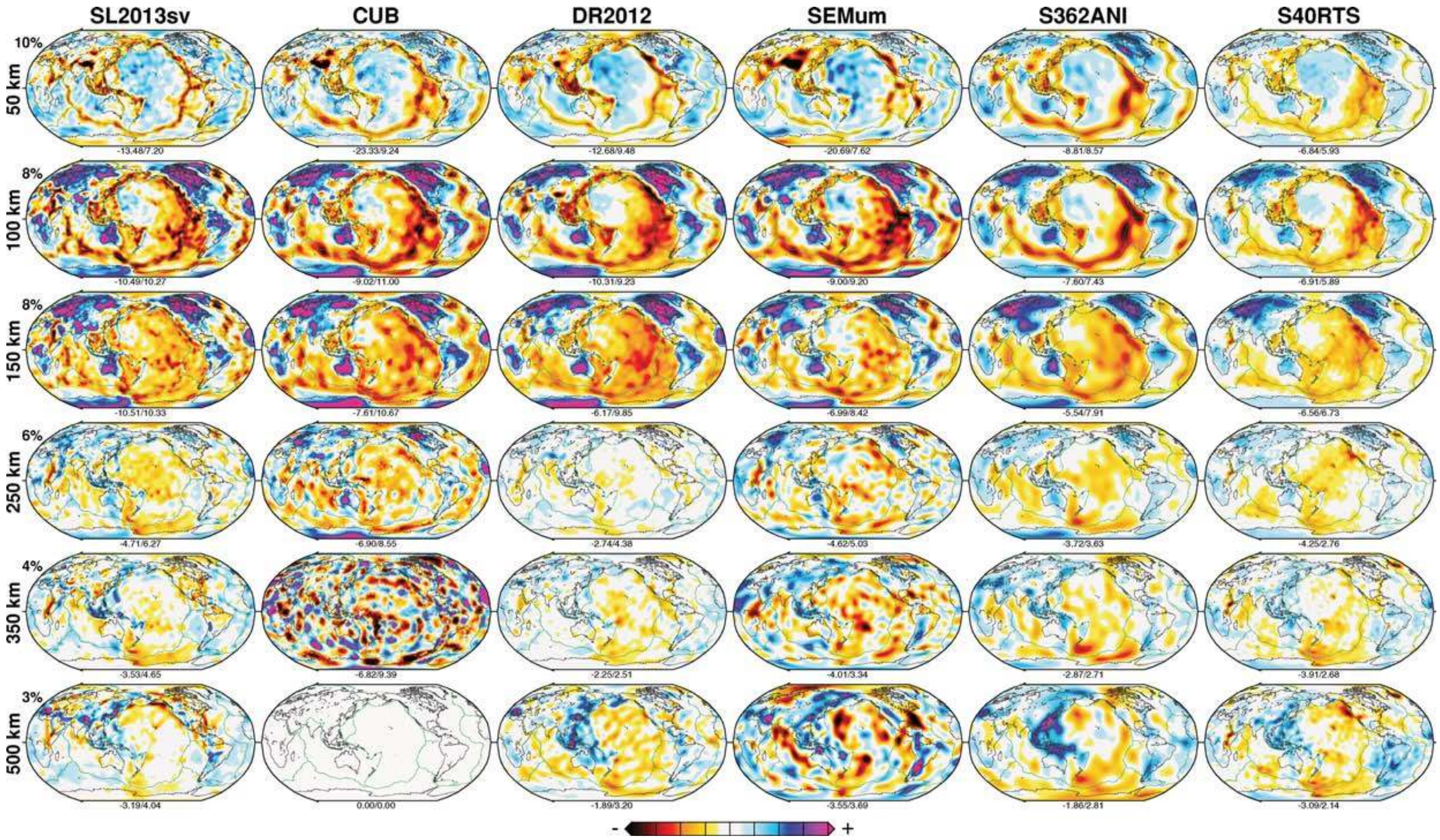
SL2013sv
CUB
DR2012
SEMum
S362ANI
S40RTS
10%
50 km
-13.48/7.20
-23.33/9.24
-12.68/9.48
-20.69/7.62
-8.81/8.57
-6.84/5.93
8%
100 km
-10.49/10.27
-9.02/11.00
-10.31/9.23
-9.00/9.20
-7.60/7.43
-6.91/5.89
8%
150 km
-10.51/10.33
-7.61/10.67
-6.17/9.85
-6.99/8.42
-5.54/7.91
-6.56/6.73
6%
250 km
-4.71/6.27
-6.90/8.55
-2.74/4.38
-4.62/5.03
-3.72/3.63
-4.25/2.76
4%
350 km
-3.53/4.65
-6.82/9.39
-2.25/2.51
-4.01/3.34
-2.87/2.71
-3.91/2.68
3%
500 km
-3.19/4.04
0.00/0.00
-1.89/3.20
-3.55/3.69
-1.86/2.81
-3.09/2.14
-
+

Figure 16. Comparison of SL2013sv with five recent global tomographic models: CUB (Shapiro & Ritzwoller 2002), DR2012 (Debayle & Ricard 2012), SEMum (Lekić & Romanowicz 2011), S362ANI (Kustowski *et al.* 2008a) and S40RTS (Ritsema *et al.* 2011). At each of five depths in the upper mantle (top to bottom 50, 100, 150, 250, 350 and 500 km), perturbations are plotted in per cent with respect to the mean value for that model. The minimum and maximum values are indicated underneath each map, and the same linear colour scale spans from negative to positive saturation values indicated for each depth (at left). Models are ordered left-to-right by decreasing peak-to-peak variations at 100 and 150 km depths.

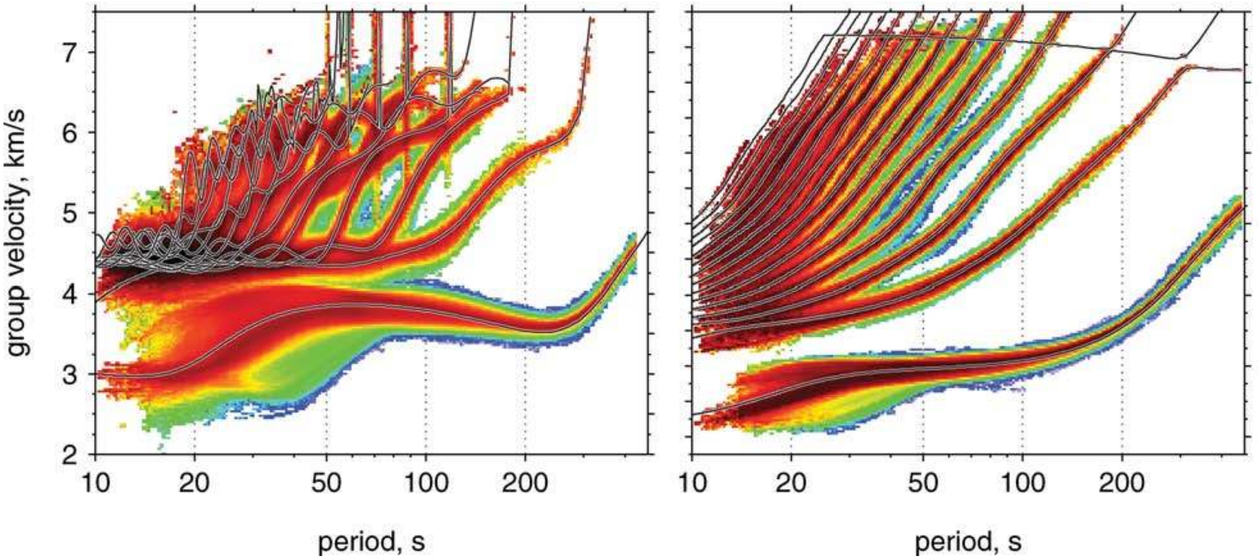

Figure 17. Empirical distribution of the group (left-hand panel) and phase velocities (right-hand panel) of the fundamental and higher modes. All multimode dispersion curves measured from the model subset of waveform fits (E, Table A1) were binned together; blue colours indicate minimum density, reds through black indicate maximum density. The fundamental mode and overtone dispersion curves calculated for *AK135_50* are plotted as white/black lines. For group velocity, the first 12 overtones are plotted, for phase velocity the first 14 are plotted.

confined and of higher amplitude in SL2013sv. By 150 km depth, the dominant ridge anomalies are gone in all models. At both these depths, high-velocity anomalies associated with continental lithosphere are evident in each of the models. As previously mentioned, over long wavelengths (thousands of kilometres), the cratons are quite similar; however, at shorter wavelengths (500–1000 km and shorter), there are strong differences across models. In SL2013sv, the structural boundaries within the high-velocity continental-cores are more finely resolved, and individual cratons are more readily observed. For example, the different cratons in South America, the cratonic blocks in southern Africa, the structural details along the boundaries of stable North America, and the clear linear signature of the Indian lithosphere deepening beneath the Himalaya and Tibet. In addition, very narrow high-velocity subducting oceanic lithosphere is imaged along most of the western Pacific subduction zones. Longer wavelength equivalents of these anomalies are imaged in the other models, but with reduced correspondence with the plate boundaries (green lines).

At depths corresponding to the base of the continental lithosphere and in the sublithospheric mantle (250 and 350 km), differences between models continue to increase, even at longer wavelengths. Subduction zones are evident in most models, in particularly in the western Pacific and South America; however, the shape of the subducting slabs are very different. They are imaged most clearly in SL2013sv, as finely localized near the plate boundaries, and of higher amplitudes. Across the rest of the Pacific Ocean, all models show a predominance of low velocities (yellow–orange colours), but at length scales less than ∼5000 km, their amplitudes and shapes vary strongly. In addition, each model has low-amplitude fast anomalies beneath most continents; CUB and SEMum show the highest amplitude high-velocity anomalies extending to greater depths. As we noted in the previous section, SL2013sv does not require continental roots to extend to depths much beyond 200 km.

At depths below 300–350 km, the sensitivity of the fundamental mode decreases rapidly (note that this is beyond the depth range of the CUB model). The inclusion of higher mode surface waves and teleseismic body waves become critical to resolve structures in the transition zone. Regardless of the methods used, the sampling of the sublithospheric mantle and transition zone is relatively poorer than in the lithospheric mantle for all the models, and, due to this reduced sampling, a wider range in structures is observed. This is clear from the 500 km depth maps. Each model images high-velocity anomalies in the western Pacific and beneath eastern Eurasia. However, there are large differences in the amplitude and location of these slabs even at long wavelengths (>3000 km). The large contribution of multiple-*S* body waves (higher modes) in SL2013sv has enabled relatively sharp images of the subducted slabs in the transition zone, particularly beneath North and South America, eastern Eurasia and through the Tethys suture towards the Mediterranean.

6.2 Multimode phase-velocity measurements

Following successful waveform fitting of a seismogram, AMI can measure phase velocities of the fundamental and higher modes, for those modes the velocities of which are constrained by the waveform fit within the set of time–frequency windows. The tomographic inversion in this study used only the linear equations yielded by the fitting, not phase-velocity measurements. We did, however, measure >700 000 fundamental-mode and >475 000 higher mode, Rayleigh-wave, phase-velocity curves. These are well-suited for incorporation into a variety of other imaging studies, for example, array-based, teleseismic interferometry (Meier *et al.* 2004; Lebedev *et al.* 2006; Deschamps *et al.* 2008a,b; Darbyshire & Lebedev 2009; Zhang *et al.* 2009; Endrun *et al.* 2011; Adam & Lebedev 2012). In this work, we restrict ourselves to simply examining their variability and, thus, the bulk dispersive properties of the Earth's crust, upper mantle and transition zone.

Fig. 17 displays the binned multimode group- and phase-velocity curves measured from the ∼521 000 vertical-component seismogram fits used in the final tomographic model (E, Table A1). The group-velocity curves (left-hand panel) are not independently measured, but computed from phase velocities (right-hand panel) using:

$$C(\omega) = \frac{c\,(\omega)}{1 - \left(\frac{\omega}{c(\omega)}\right)\left(\frac{\mathrm{d}c}{\mathrm{d}\omega}\right)}, \qquad (5)$$

where $\omega = 2\pi/T$ is the angular frequency, T the period and $C(\omega)$ and $c(\omega)$ are the group and phase speeds, respectively. Blue colours in Fig. 17 indicate the lowest density bins, whereas red through black colours indicate increasing density. Group- and phase-velocity curves for *AK135_50* are superimposed. For group velocity, the first 12 overtones were plotted, whereas for phase velocity the first 14 were plotted.

In the fundamental-mode group-velocity curves (left-hand panel), the greatest variability is seen at periods less than 45 s, whereas at longer periods of 100–450 s the range in group velocity at each period is much smaller. In the transitional band at 50–80 s, the spread in velocity at longer periods is a factor of 2 less than at shorter periods. The increasing group velocity at periods above 200 s is due to the higher S velocities in the lithospheric and sublithospheric mantle, to which long periods are more sensitive.

The variability in group velocity at shorter periods ($\leq$40–50 s) results from sampling of more heterogeneous shallow structure. In continental regions, this period band is most sensitive to the crust. Where the Moho is deeper (mainly, beneath orogens), low velocities extend to greater periods, manifesting as the thick green band at 30–70 s. In oceanic regions, however, group-velocity samples the uppermost mantle at periods of 15–40 s, and therefore plots faster than *AK135_50*.

Although the depth sensitivity functions of higher mode group velocities are more complex than for the fundamental mode, the same reduction in the spread of group velocities for each mode at increasing periods is observed. As with the fundamental mode, this results from sensitivity to a broader and deeper depth range. In addition, the 'ray-mode duality' is clearly seen in the overtone branches superimposing beginning at periods $\leq$60 s, and converging towards a group velocity of $\sim$4.3 km s^{-1} with decreasing period; this represents an S wave travelling in the upper mantle.

The phase velocities for the fundamental- and first 14 higher modes are shown in the right-hand panel of Fig. 17. Given that the sensitivity of phase- and group velocity differ substantially (Lebedev *et al.* 2013), it is no surprise that the phase-velocity curves are different in character. Unlike in group velocity, the 'average' fundamental-mode phase velocity monotonically increases as a function of period. In addition, beyond 50 s the spread in phase velocity varies minimally. At periods shorter than 50 s, however, the variation increases (in a manner similar to group velocity at periods $\leq$60 s) due to the increasing sensitivity to more heterogeneous shallow structure.

The higher mode phase-velocity curves can be distinctly identified, particularly up to modes seven or eight. Those higher than nine are more closely spaced and, especially at short periods, are more difficult to distinguish on the plot. At high phase velocities, $\sim$8.5 km s^{-1}, sudden jumps in overtone branches related to core–mantle-boundary Stoneley-modes are clearly recovered for at least six of the overtones (e.g. Dahlen & Tromp 1998).

The different character between fundamental-mode oceanic and continental dispersion curves is clearly observed in both the group ($<$80 s) and phase ($<$50 s) velocity images; curves from faster, oceanic paths lie above *AK135_50*, while those from slower paths (across continents or backarcs) lie below. Interestingly, the bin density is lower for the former (i.e. oceanic paths), rather than the latter. This is contrary to expectation, as oceanic crust accounts for more than 50 per cent of the Earth's surface. As has already been discussed, however, the sources and receivers are not evenly distributed, and therefore impart a 'sampling filter' on the results.

Based on an analysis of the distribution of path lengths and minimum-filter centre periods, we observe that the shortest period phase-velocity curves result from paths sampling dominantly continental regions and backarcs. From this, we can conclude that the source–receiver distribution is affecting the relative sampling density at shorter periods (which sample the heterogeneous lithospheric mantle and crust), resulting in a relative oversampling of low velocities (compared to *AK135_50*). The effects of the biased sampling is reduced at longer periods, where the range of path lengths is wider and phase velocities are sensitive to the more homogeneous deeper structure.

6.3 Validity of the JWKB approximation

In 2005, Lebedev *et al.* used a data set of 4038 vertical-component seismogram fits computed for the western Pacific and southeast Asia to examine the validity field of the JWKB approximation. Our new data set (B, Table A1) of almost 3/4 of a million waveform fits provides a useful opportunity to revisit the stability fields of the assumptions utilized by AMI, with a more substantial sampling. In the following sections, we first quantify AMI's success rate of waveform selection and fitting and then expand on the work of Lebedev *et al.* (2005), examining the validity field of the JWKB approximation, as implemented in AMI. Although surface wave ray theory was not, strictly speaking, applied in the waveform fitting in this study (we integrated across approximate sensitivity areas), the frequency-dependent success rates of fitting would be similar if we used rays instead of Frésnel zones. The results in this section thus apply to the validity of surface wave ray theory as well.

6.3.1 Success rate of AMI

Using this new data set of waveform fits, we have further verified AMI's ability to successfully process large volumes of seismic data. The accuracy of processing depends both on the approximations and on successfully discriminating between true signal and noise. In the top panel of Fig. 18, a black dot is plotted at the distance magnitude for each successfully fit seismogram (one per seismogram, not each time–frequency window), mapping out the AMI fitting field. In this case, we have selected only seismograms from high-quality, long-term stations of the GSN to reduce the effects of *in situ* and instrument noise on the fitting statistics of AMI. As would be expected, the pattern obtained when including noisier stations is similar, albeit with a reduced overall success rate.

The shape of the successfully fit region results from both the source-station geometry, as well as the approximations and conditions enforced by AMI. The white regions around the perimeter of the plot represent source-station configurations for which AMI does not fit a seismogram. At the bottom (low magnitudes), the grey line indicates the distance versus magnitude threshold employed when selecting seismograms. The sharp vertical boundary at long path lengths ($\Delta > 16\,500$ km) is enforced to avoid source-stations configurations nearing the antipode, where interference of major and minor arc phases results in greater complexity and large amplitude variations of the arrivals.

The distribution of points in Fig. 18 (top panel) was binned to generate a quantitative measure of AMI's success rate (bottom panel). White indicates a success ratio of 0 per cent, blue through red an increasing fit success rate, and black represents 100 per cent successful fitting. Three smoothed contours are superimposed, the interior of which indicates all seismograms are fit successfully at least 20, 50 and 70 per cent of the time; below the grey line no fits are attempted.

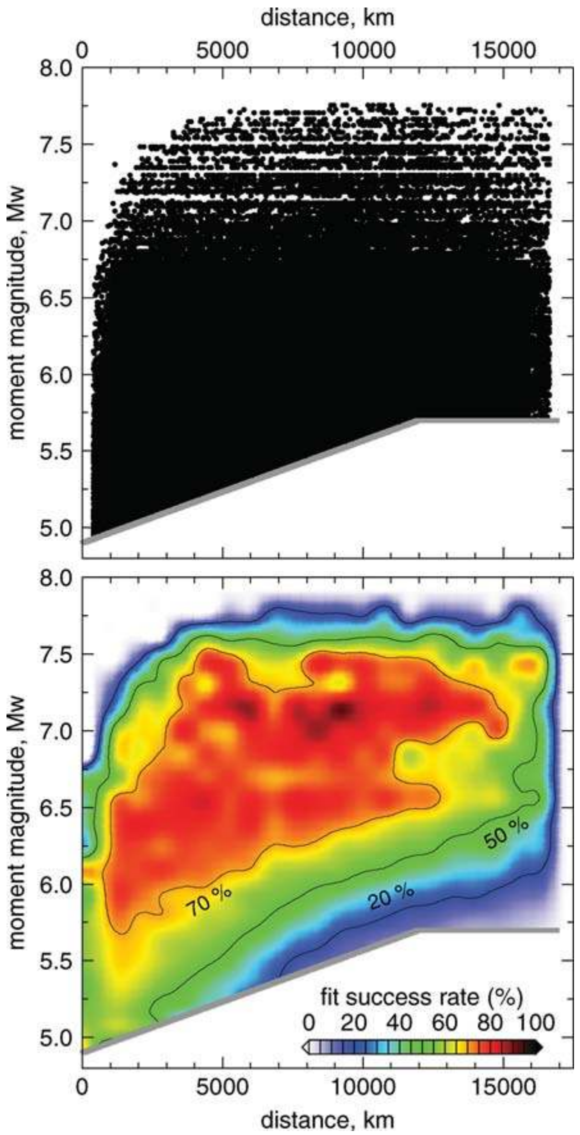

Figure 18. Success of AMI waveform fitting in the epicentral distance–earthquake magnitude plane. In the top panel, we plot the raw data. Each black dot represents a successfully fit seismogram. In the bottom panel, we have computed the success of AMI waveform fitting. Colour indicates the percentage of successful fits in each bin. White regions indicate areas of the depth-magnitude plane where no fits are obtained; this may be due to noise or scattering or both. The grey line at the bottom is the empirical distance-magnitude threshold utilized in selecting data.

At magnitudes greater than 6.5 M_W, success rates of >70 per cent extend across the distance axis, indicating that for long periods, the JWKB approximation is successful at fitting in most cases. The different behaviour at lower magnitudes is due to lower signal-to-noise ratios. This is clearly observed in Fig. 18 (bottom panel): for a given magnitude, for example, 5.75 M_W, fitting success is inversely proportional to epicentral distance. The 50 per cent contour illustrates this decay of signal-to-noise ratio; with increasing distance, the minimum magnitude required to achieve 50 per cent success increases.

It is clear that AMI is not only successfully fitting seismograms across a large area of the distance-magnitude plane (red areas representing >70 per cent success rate), but is also effectively identifying and discarding noisy seismograms. Although the minimum magnitude of requested data is restricted, the highest concentration of seismograms lies at these lowest magnitudes, even though most of them are too noisy to yield useful fits. In examining the success rate of AMI, it is encouraging to observe that at these low magnitudes, very few seismograms are fit. Although this may result in the impression of AMI underperforming (e.g. ~750 000 fits out of 3.6 million seismograms, an overall success rate of 20 per cent), this is certainly not the case. As the large majority of the 2.85 million seismograms not fit lie at these low magnitudes, AMI has effectively discarded noisy seismograms while simultaneously obtaining high success rates across a large space of the distance-magnitude plane (e.g. the >50 and >70 per cent successfully fit regions).

6.3.2 Empirical bounds for the JWKB approximation

The validity of the surface wave JWKB theory—in this context neglecting the effects of scattering while incorporating finite width sensitivity regions—is, in general, only warranted for waves travelling through regions of smooth lateral heterogeneity (Kennett & Nolet 1990; Wang & Dahlen 1995b; Dahlen & Tromp 1998; Lebedev *et al.* 2005). In many regions of the Earth, particularly the crust and upper mantle, heterogeneity sampled by surface waves is rough compared to Fresnel-zone widths at the periods of interest (Wang & Dahlen 1995b); therefore, in many cases the validity of JWKB theory is not warranted, meaning that it may or may not be valid for any given time–frequency portion of a particular seismogram.

In Lebedev *et al.* (2005) examined the validity field of surface wave ray theory using a data set of 4038 vertical-component seismograms fit by AMI. Using this data set, the authors concluded that AMI's case-by-case selection of the time–frequency portions of seismograms that can successfully be modelled using eqs (1) and (2) is well suited to ensure the validity of the approximations. Our new data set of more than 175× the number of seismograms (and, similarly, the number of fundamental- and higher mode time–frequency windows) is well suited to further explore the empirical validity field of the JWKB approximation.

In the top panels of Fig. 19, a single black dot is plotted for each successfully fit time–frequency window for the fundamental mode (left-hand side) and higher modes (right-hand side); each point is mapped based on its Gaussian filter centre period and epicentral distance. The lower left corners are devoid empty, reflecting the far-field approximation implemented in AMI. The white areas in the top right corners are regions where the JWKB approximation is never valid, as no matter how many attempts are made, no time–frequency windows are successfully fit; this region represents the scattering regime.

The distributions shown in the top panels of Fig. 19 provide an empirical estimate for the boundaries of the validity field of the JWKB approximation as implemented by AMI. By cumulatively binning the point clouds for each distance, from minimum to maximum period, more quantitative empirical validity field estimates are presented in the bottom panels of Fig. 19. Colours going from blue towards red and then black indicate increasing density of successfully fit time–frequency windows.

As is expected, we observe a decrease in the likelihood of the validity of the JWKB approximation with increasing distance and decreasing period (increasing frequency), both for the fundamental and higher modes. It is important to note that, by selection, only *S* and multiple-*S* waves propagating primarily within the upper mantle and transition zone are included in the higher mode parts of the waveforms that are fitted. Therefore, they sample greater

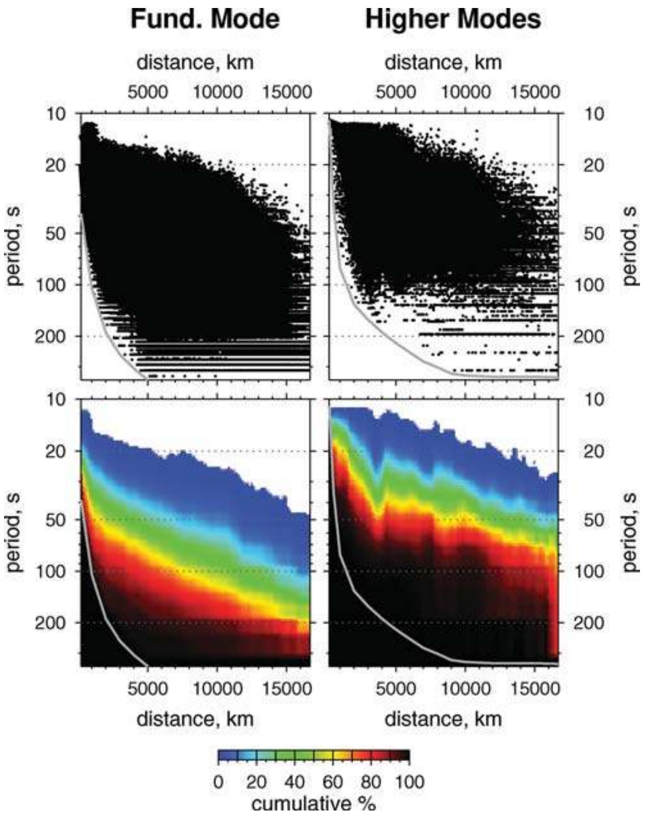

Figure 19. Empirical validity field for the JWKB approximation for the fundamental- and higher modes, left-hand and right-hand panels, respectively. In the top panels, each successfully fit time–frequency window is plotted as a black dot in distance-period plane. These dot-plots are converted into cumulatively-binned point densities in the bottom panels. For each distance (column spanning entire period range), colours indicate the cumulative successful fits starting from 0 per cent (white) at the smallest periods and increasing to 100 per cent (black) at greater periods. White regions indicate the scattering regime (where no fits were computed successfully) whereas coloured regions indicate where the JWKB approximation can be valid (fits were successfully computed). Note that all period axes are logarithmically scaled.

heterogeneity than teleseismic *S* waves travelling in the deep lower mantle.

In the past, theoretical and empirical estimates were made on the bounds of validity of the JWKB approximation. Kennett & Nolet (1990) modelled wave propagation in realistic upper-mantle models to infer a validity threshold of 50 s (20 mHz), for propagation distances of 3350 km. Lebedev *et al.* (2005) found that this threshold was conservative, as their hit-count only begins to decrease at this point. Wang & Dahlen (1995b) point out that the misfit of the surface wave ray approximation depends on diffraction and other finite-frequency effects ignored by JWKB theory, and is dependent on the quantity $\bar{s}/\sqrt{4\pi l}$ (where $\bar{s}$ is the root-mean-square degree of the phase-velocity perturbation δc and l is the degree of the equivalent mode ${}_nS_l$ or ${}_nT_l$). From Fig. 19, it is clear the validity field can be extended to much shorter periods (higher frequencies) than past conservative estimates, as long as portions of the seismogram affected by unmodelled wave propagation effects (scattering) can be identified and avoided, as is automatically carried out by AMI.

At an epicentral distance of 3350 km, the fundamental-mode validity field (in Fig. 19) extends to periods as short as ∼20 s (50 mHz), almost half (double) the past estimates. The empirical cut-off of ∼10 per cent (a bin density of ∼3), used by Lebedev *et al.* (2005, Fig. 10), gives a transition frequency of ∼33 mHz at 3350 km, which is very similar to the results obtained in this study when employing the same 10 per cent cut-off. The similarity verifies the accuracy of the past estimate, but with a more robust statistical sampling.

7 CONCLUSIONS

We have generated an unprecedentedly large data set of ∼3/4 of a million vertical-component waveform fits. We use this new data set to validate key aspects of the multimode waveform inversion, to assess the bulk dispersive properties of the upper mantle, and to re-examine the validity field of the JWKB approximation, the surface wave ray-theoretical foundation underpinning most of the past global models, on which much of our current understanding of large-scale mantle structure and dynamics is based.

Through recomputing our full data set of waveform fits four separate times (see Appendix A), we examined and compared the frequency-dependent sensitivity of the derived tomographic models. Whereas Lebedev & van der Hilst (2008) imposed an upper frequency cut-off of ∼60 mHz (∼44 mHz filter centre frequency) to avoid breakdowns of the path-average approximation for shorter period surface waves sampling the Earth's heterogeneous crust, we show, using tests with our much larger data set, that the negative effect of this on the model is very limited, smaller than the probable effects of the errors in the source locations and mechanisms, station timing and of unmodelled diffraction. The net effect of expanding the frequency band is, instead, positive, thanks to the extra structural information from now-included higher frequency *S* and multiple-*S* wave (higher mode) data. Our preferred data set (B) was thus generated without any imposed upper frequency limit.

The large number of waveform fits offers new insight into the validity of the basic approximations that are used extensively in upper-mantle imaging. We were able to confirm both the consistency of AMI in detecting noisy seismograms (which should not be fit) and correctly identifying and fitting large numbers of seismograms for which the approximations are valid, across much of the earthquake magnitude–epicentral distance plane. We have re-examined the empirical validity field for the JWKB approximation and demonstrated that it is valid for a large proportion of the data, particularly at shorter periods than previously theorized. Importantly, the time–frequency portions of the signal for which the approximation is valid can be consistently identified on a case-by-case basis, for use in the imaging.

Our new global, upper-mantle, vertically polarized shear speed model, SL2013sv is constrained by more than half-a-million of the most mutually consistent waveform fits, selected using a rigorous outlier analysis procedure. This new model is capable of resolving features smaller than 6° laterally globally, and certainly much finer in well-sampled continental regions. In oceanic regions, we have captured striking images of spreading ridge anomalies which are more localized near the ridge axis in the uppermost mantle than in the past models. In continental regions, we conclude that the high-velocity, cold cratonic roots are not required to extend far beyond 200 km depth. Between 150 km and the base of the transition zone, we obtain clear images of most of the major subduction zones, including many in the western Pacific, Cascadia and the South American margin (Andean). Finally, in the transition zone we see clear evidence for the lithosphere subducted during the closure of the Tethys Ocean and subsequent continental collisions, stretching almost continuously from the Mediterranean to southeast Asia. Observed agreement of the deep-crustal and upper-mantle structure resolved by the model with regional-scale surface tectonics and the clear recovery of transition zone structures show that the significantly increased data coverage and data redundancy, coupled with AMI's accurate extraction of structural information from surface and *S* waveforms, translate into significantly improved resolution in the lithosphere and upper mantle.

ACKNOWLEDGEMENTS

Waveform data were obtained from the facilities of the IRIS, the ORFEUS, the GFZ Seismological Data Archive and the CNSN. We are grateful to the operators and Principal Investigators of the many networks used in this study. This work was supported by Science Foundation Ireland (grant 09/RFP/GEO2550). All figures were generated using Generic Mapping Tools (GMT; Wessel & Smith 1995). We appreciate the insightful comments from Jeroen Ritsema, an anonymous reviewer, and the Associate Editor Andrea Morelli, which improved this manuscript. Our tomographic model, SL2013sv, can be obtained from http://www.dias.ie/∼aschaeff/SL2013sv.html.

REFERENCES

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

Amante, C. & Eakins, B., 2009. ETOPO1 1 Arc-Minute Global Relief Model: procedures, data sources and analysis, NOAA Technical Memorandum, NESDIS NGD, 1–19.

Amaru, M.L., 2006. Global travel time tomography with 3-D reference models, *PhD thesis,* Universiteit Utrech.

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,** F897.

Becker, T.W. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosys.,* **3**(1), 1003, doi:10.1029/2011GC000168.

Becker, T.W., Lebedev, S. & Long, M.D., 2012. On the relationship between azimuthal anisotropy from shear wave splitting and surface wave tomography, *J. geophys. Res.,* **117**(B1), 1–17.

Bedle, H. & van der Lee, S., 2009. *S* velocity variations beneath North America, *J. geophys. Res.,* **114**(B7), doi:10.1029/2008JB005949.

Bijwaard, H. & Spakman, W., 2000. Nonlinear global *P*-wave tomography by iterated linearized inversion, *Geophys. J. Int.,* **141,** 71–82.

Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.,* **103**(B12), 30055–30078.

Boschi, L. & Ekström, G., 2002. New images of the Earths upper mantle from measurements of surface wave phase velocity anomalies, *J. geophys. Res.,* **107**(B4), 2059, doi:10.1029/2000JB000059.

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

Burdick, S. *et al.*, 2010. Model Update 2010: upper mantle heterogeneity beneath North America from traveltime tomography with global and USArray Transportable Array data, *Seism. Res. Lett.,* **81**(5), 689–693.

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.

Cara, M. & Lévêque, J.-J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.,* **14,** 1046–1049.

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**(2), 457–491.

Chen, P., Zhao, L. & Jordan, T.H., 2007. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97**(4), 1094–1120.

Chevrot, S. & Zhao, L., 2007. Multiscale finite-frequency Rayleigh wave tomography of the Kaapvaal craton, *Geophys. J. Int.,* **169,** 201–215.

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

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

Darbyshire, F.A. & Lebedev, S., 2009. Rayleigh wave phase-velocity heterogeneity and multilayered azimuthal anisotropy of the Superior Craton, Ontario, *Geophys. J. Int.,* **176,** 215–234.

Debayle, E. & Kennett, B.L.N., 2000. The Australian continental upper mantle: structure and deformation inferred from surface waves, *J. geophys. Res.,* **105**(11), 25 423–25 450.

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), 1–24.

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

Deschamps, F., Lebedev, S., Meier, T. & Trampert, J., 2008a. Stratified seismic anisotropy reveals past and present deformation beneath the east-central United States, *Earth planet. Sci. Lett.,* **274**(3–4), 489–498.

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

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

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

Ekström, G., Nettles, M. & Dziewónski, A.M., 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes, *Phys. Earth planet. Inter.,* **200–201,** 1–9.

Endrun, B., Lebedev, S., Meier, T., Tirel, C. & Friederich, W., 2011. Complex layered deformation within the Aegean crust and mantle revealed by seismic anisotropy, *Nature Geosci.,* **4**(3), 203–207.

Feng, M., Assumpção, M. & van der Lee, S., 2004. Group-velocity tomography and lithospheric S-velocity structure of the South American continent, *Phys. Earth planet. Inter.,* **147,** 315–331.

Ferreira, A.M.G., Woodhouse, J.H., Visser, K. & Trampert, J., 2010. On the robustness of global radially anisotropic surface wave tomography, *J. geophys. Res.,* **115**(B4), 1–16.

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

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

Fishwick, S., 2010. Surface wave tomography: imaging of the lithosphere–asthenosphere boundary beneath central and southern Africa? *Lithos,* **120,** 63–73.

Fishwick, S., Heintz, M., Kennett, B.L.N., Reading, A.M. & Yoshizawa, K., 2008. Steps in lithospheric thickness within eastern Australia, evidence from surface wave tomography, *Tectonics,* **27,** 1–17.

Forsyth, D.W., Scheirer, D. & Webb, S., the MELT Seismic Team, 1998. Imaging the deep seismic structure beneath a mid-ocean ridge: the MELT experiment, *Science,* **280**(5367), 1215–1218.

Frederiksen, A.W., Bostock, M.G. & Cassidy, J.F., 2001. S-wave velocity structure of the Canadian upper mantle, *Phys. Earth planet. Inter.,* **124,** 175–191.

Gee, L.S. & Jordan, T.H., 1992. Generalised seismological data functionals, *Geophys. J. Int.,* **111,** 363–390.

Grand, S.P., 2002. Mantle shear-wave tomography and the fate of subducted slabs., *Philos. Trans. R. Soc. Lond.,* **360,** 2475–2491.

Grand, S.P., van der Hilst, R.D. & Widiyantoro, S., 1997. Global seismic tomography: a snapshot of convection in the Earth, *GSA Today,* **7**(4), 1–7.

Gu, Y.J., Dziewónski, A.M., Su, W. & Ekström, G., 2001. Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities, *J. geophys. Res.,* **106**(B6), 11 169–11 199.

Gu, Y.J., Dziewónski, A.M. & Ekström, G., 2003. Simultaneous inversion for mantle shear velocity and topography of transition zone discontinuities, *Geophys. J. Int.,* **154,** 559–583.

Hafkenscheid, E., Wortel, M.J.R. & Spakman, W., 2006. Subduction history of the Tethyan region derived from seismic tomography and tectonic reconstructions, *J. geophys. Res.,* **111**(B8), 1–26.

Heintz, M., Debayle, E. & Vauchez, A., 2005. Upper mantle structure of the South American continent and neighboring oceans from surface wave tomography, *Tectonophysics,* **406,** 115–139.

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

Karason, H. & van der Hilst, R.D., 2000. Constraints on mantle convection from seismic tomography, in *The History and Dynamics of Global Plate Motions,* pp. 277–288, AGU Geophysical Monograph Series.

Kennett, B.L.N. & Nolet, G., 1990. The interaction of the S-wavefield with upper mantle heterogeneity, *Geophys. J. Int.,* **101,** 751–762.

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

King, S. & Ritsema, J., 2000. African hot spot volcanism: small-scale convection in the upper mantle beneath cratons, *Science,* **290,** 1137–1140.

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**(2), 368–392.

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

Kustowski, B., Ekström, G. & Dziewónski, A.M., 2008a. Anisotropic shear-wave velocity structure of the Earth's mantle: a global model, *J. geophys. Res.,* **113**(B6), 1–23.

Kustowski, B., Ekström, G. & Dziewónski, A.M., 2008b. The shear-wave velocity structure in the upper mantle beneath Eurasia, *Geophys. J. Int.,* **174,** 978–992.

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

Lebedev, S. & van der Hilst, R.D., 2008. Global upper-mantle tomography with the automated multimode inversion of surface and *S*-wave forms, *Geophys. J. Int.,* **173,** 505–518.

Lebedev, S., Adam, J. & Meier, T., 2013. Mapping the Moho with seismic surface waves: a review, resolution analysis, and recommended inversion strategies, in *Tectonophysics,* "Moho" special issue, doi:10.1016/j.tecto.2012.12.030.

Lebedev, S., Meier, T. & van der Hilst, R.D., 2006. Asthenospheric flow and origin of volcanism in the Baikal Rift area, *Earth planet. Sci. Lett.,* **249,** 415–424.

Lebedev, S., Nolet, G. & van der Hilst, R.D., 1997. The upper mantle beneath the Philippine Sea region from waveform inversions, *Geophys. Res. Lett.,* **24**(15), 1851–1854.

Lebedev, S., Nolet, G., Meier, T. & van der Hilst, R.D., 2005. Automated multimode inversion of surface and *S* waveforms, *Geophys. J. Int.,* **162,** 951–964.

Legendre, C.P., Meier, T., Lebedev, S., Friederich, W. & Viereck-Götte, L., 2012. A shear wave velocity model of the European upper mantle from automated inversion of seismic shear and surface waveforms, *Geophys. J. Int.,* **191,** 282–304.

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

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

Li, C., van der Hilst, R.D., Engdahl, E.R. & Burdick, S., 2008. A new global model for *P* wave speed variations in Earth's mantle, *Geochem. Geophys. Geosys.,* **9**(5), Q05018, doi:10.1029/2007GC001806.

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

Li, X.-D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory, *J. geophys. Res.,* **101**(B10), 22 245–22 272.

Maggi, A., Debayle, E., Priestley, K. & Barruol, G., 2006. Multimode surface waveform tomography of the Pacific Ocean: a closer look at the lithospheric cooling signature, *Geophys. J. Int.,* **166**(3), 1384–1397.

Marone, F. & Romanowicz, B., 2007. Non-linear crustal corrections in high-resolution regional waveform seismic tomography, *Geophys. J. Int.,* **170**(1), 460–467.

Marquering, H., Snieder, R. & Nolet, G., 1996. Waveform inversions and the significance of surface wave mode coupling, *Geophys. J. Int.,* **125**(I), 258–278.

Masters, G., Johnson, S., Laske, G. & Bolton, H., 1996. A shear-velocity model of the mantle, *Philos. Trans. R. Soc. Lond.,* **354,** 1385–1411.

Masters, G., Laske, G., Bolton, H. & Dziewónski, A.M., 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,* eds Karato, S., Forte, A., Liebermann, R., Masters, G. & Stixrude, L., AGU Monograph 117, Washington, DC.

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**(3), 709–728.

Meier, T., Lebedev, S., Nolet, G. & Dahlen, F.A., 1997. Diffraction tomography using multimode surface waves, *J. geophys. Res.,* **102**(B4), 8255–8267.

Meier, T., Dietrich, K., Stockhert, B. & Harjes, H.-P., 2004. One-dimensional models of shear wave velocity for the eastern Mediterranean obtained from the inversion of Rayleigh wave phase velocities and tectonic implications, *Geophys. J. Int.,* **156**(1), 45–58.

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

Montelli, R., Nolet, G., Masters, G., Dahlen, F.A. & Hung, S.-H., 2004. Global *P* and *PP* traveltime tomography: rays versus waves, *Geophys. J. Int.,* **158**(2), 637–654.

Nelson, K. *et al.*, 1996. Partially molten middle crust beneath southern Tibet: synthesis of project INDEPTH results, *Science,* **274**(December), 1684–1688.

Nettles, M. & Dziewónski, A.M., 2008. Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America, *J. geophys. Res.,* **113**(B2), 1–27.

Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs, *J. geophys. Res.,* **95**(B6), 8499–8512.

Nolet, G. & Dahlen, F.A., 2000. Wave front healing and the evolution of seismic delay times, *J. geophys. Res.,* **105**(B8), 19 043–19 054.

Paige, C. & Saunders, M., 1982. LSQR: an algorithm for sparse linear equations and sparse least squares, *Trans. Math. Softw.,* **8,** 43–71.

Panning, M.P. & Romanowicz, B., 2006. A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.,* **167**(1), 361–379.

Panning, M.P., Cao, A., Kim, A. & Romanowicz, B., 2012. Non-linear 3D Born shear waveform tomography in southeastern Asia, *Geophys. J. Int.,* **190,** 463–475.

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

Pollitz, F.F., 2001. Remarks on the travelling wave decomposition, *Geophys. J. Int.,* **144**(2), 233–246.

Priestley, K., Debayle, E., McKenzie, D. & Pilidou, S., 2006. Upper mantle structure of eastern Asia from multimode surface waveform tomography, *J. geophys. Res.,* **111**(10), 1–20.

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

Qin, Y., Capdeville, Y., Maupin, V. & Montagner, J.-P., 2006. Synthetic dataset to benchmark global tomographic methods, *EOS, Trans. Am. geophys. Un.,* **87**(46), 512, doi:10.1029/2006EO460004.

Qin, Y., Capdeville, Y., Maupin, V., Montagner, J.-P., Lebedev, S. & Beucler, E., 2008. SPICE benchmark for global tomographic methods, *Geophys. J. Int.,* **175**(2), 598–616.

Rickers, F., Fichtner, A. & Trampert, J., 2012. Imaging mantle plumes with instantaneous phase measurements of diffracted waves, *Geophys. J. Int.,* **190,** 650–664, doi:10.111/j.1365-246X.2012.05515.x.

Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 2004. Global transition zone tomography, *J. geophys. Res.,* **109**(B2), doi:10.1029/2003JB002610.

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

Schimmel, M., Assumpção, M. & VanDecar, J.C., 2003. Seismic velocity anomalies beneath SE Brazil from *P* and *S* wave travel time inversions, *J. geophys. Res.,* **108**(B4), 2191, doi:10.1029/2001JB000187.

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

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

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

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

Simons, F.J., Zielhuis, A. & van der Hilst, R.D., 1999. The deep structure of the Australian continent from surface wave tomography, *Lithos,* **48**(1–4), 17–43.

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

Su, W.-J., Woodward, R. & Dziewónski, A.M., 1992. Deep origin of mid-ocean-ridge seismic velocity anomalies, *Nature,* **360,** 149–152.

Su, W.-j., Woodward, R. & Dziewónski, A.M., 1994. Degree 12 model of shear velocity heterogeneity in the mantle, *J. geophys. Res.,* **99**(B4), 6945–6980.

Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science,* **325,** 988–992.

Tian, Y., Sigloch, K. & Nolet, G., 2009. Multiple-frequency *SH* -wave tomography of the western US upper mantle, *Geophys. J. Int.,* **178**(3), 1384–1402.

Tian, Y., Zhou, Y., Sigloch, K., Nolet, G. & Laske, G., 2011. Structure of North American mantle constrained by simultaneous inversion of multiple-frequency *SH*, *SS*, and Love waves, *J. geophys. Res.,* **116**(B2), 1–18.

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

van der Lee, S. & Frederiksen, A.W., 2005. Surface wave tomography applied to the North American upper mantle, in *Seismic Earth: Array Analysis of Broadband Seismograms,* Vol. 157, pp. 67–80, eds Levander, A. & Nolet, G., AGU Geophysical Monograph Series, Washington, DC.

van der Lee, S., James, D.E. & Silver, P.G., 2001. Upper mantle *S* velocity structure of central and western South America, *J. geophys. Res.,* **106**(12), 30 821–30 835.

Van der Voo, R., Spakman, W. & Bijwaard, H., 1999. Tethyan subducted slabs under India, *Earth planet. Sci. Lett.,* **171**(1), 7–20.

Visser, K., Lebedev, S., Trampert, J. & Kennett, B.L.N., 2007. Global Love wave overtone measurements, *Geophys. Res. Lett.,* **34**(3), 1–6.

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

Wang, Z. & Dahlen, F.A., 1995a. Spherical-spline parameterization of three-dimensional Earth models, *Geophys. Res. Lett.,* **22**(22), 3099–3102.

Wang, Z. & Dahlen, F.A., 1995b. Validity of surface-wave ray theory on a laterally heterogeneous earth, *Geophys. J. Int.,* **123**(3), 757–773.

Wessel, P. & Smith, W., 1995. New version of the Generic Mapping Tools released, *EOS, Trans. Am. geophys. Un.,* **76,** 329.

Woodhouse, J.H. & Dziewónski, A.M., 1984. Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. geophys. Res.,* **89**(B7), 5953–5986.

Yoshizawa, K., 2004. Multimode surface wave tomography for the Australian region using a three-stage approach incorporating finite frequency effects, *J. geophys. Res.,* **109**(B2), B02310, doi:10.1029/2002JB002254.

Yoshizawa, K. & Kennett, B.L.N., 2002. Determination of the influence zone for surface wave paths, *Geophys. J. Int.,* **149**(2), 440–453.

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

Zaroli, C., Debayle, E. & Sambridge, M., 2010. Frequency-dependent effects on global *S*-wave traveltimes: wavefront-healing, scattering and attenuation, *Geophys. J. Int.,* **182**(2), 1025–1042.

Zhang, X., Paulssen, H., Lebedev, S. & Meier, T., 2009. 3D shear velocity structure beneath the Gulf of California from Rayleigh wave dispersion, *Earth planet. Sci. Lett.,* **279**(3–4), 255–262.

Zhang, Y. & Tanimoto, T., 1992. Ridges, hotspots and their interaction as observed in seismic velocity maps, *Nature,* **355,** 45–49.

Zhang, Y. & Tanimoto, T., 1993. High-resolution global upper mantle structure and plate tectonics, *J. geophys. Res.,* **98**(B6), 9793–9823.

Zhao, L., Jordan, T.H. & Chapman, C., 2000. Three–dimensional Fréchet differential kernels for seismic delay times, *Geophys. J. Int.,* **141,** 558–576.

Zhou, Y., Dahlen, F.A., Nolet, G. & Laske, G., 2005. Finite-frequency effects in global surface-wave tomography, *Geophys. J. Int.,* **163**(3), 1087–1111.

Zhou, Y., Nolet, G., Dahlen, F.A. & Laske, G., 2006. Global upper-mantle structure from finite-frequency surface-wave tomography, *J. geophys. Res.,* **111**(B4), 1–24.

Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nature Geosci.,* **5**(7), 493–498.

APPENDIX A: SELECTING A WAVEFORM DATA SET

Outlined are the additional procedures used to select the data set used in generating the final tomographic model. Tests were carried out to analyse the effects of near-nodal radiation and the impact of the upper frequency limit used in waveform inversion on both the fits and the tomographic models. Finally, the treatment of remaining errors is discussed. The four main data sets examined are summarized in Table A1.

To start, waveform fits are generated with no enforced upper frequency limit and no consideration of near-nodal radiation. The resulting data set A (NR, Table A1) consists of 846 360 vertical-component waveform fits. Data set B (RAD, Table A1) excludes time–frequency portions of seismograms at near-nodal azimuths; this data set comprises 712 077 waveform fits. The final two data sets, C (60 mHz) and D (43 mHz), were computed with progressively lower upper frequency cut-offs imposed during waveform fitting and include 692 540 and 685 146 fits, respectively. The values of 60 and 43 mHz indicate the frequencies at which the high-frequency tail of the highest frequency Gaussian filter decreases to an amplitude 0.3× the filter's central amplitude. As a result, the centre frequencies of the respective highest frequency filters are ∼48 (∼20 s) and ∼35 mHz (∼29 s).

A1 Effects of near-nodal radiation

Automated multimode inversion (AMI) is implemented with a frequency- and azimuth-dependent threshold, set up to avoid fitting seismograms with source–receiver azimuths near a node in the

Table A1. Table summarizing the four different data sets of waveform fits generated using AMI (A–D). For each case, the same set of ∼3.6 million seismograms was used. The final data set, E, is that used for generating the final tomographic model, and is derived from data set B. Column one indicates the name of the data set. The second column indicates what conditions were imposed during waveform fitting. The third and fourth columns indicate the 'minimum' and 'maximum' Gaussian filter centre frequencies for each data set. The fifth column indicates the total number of fits for that data set. The sixth and seventh columns list the number of fundamental and higher mode time–frequency windows.

		Gaussian filters (mHz)			No. time–frequency windows	
Data set name	AMI conditions	Min.	Max.	No. fits	Fund. mode	Higher modes
A) NR	Nodal radiation pattern disregarded; No frequency limits	2.93	100.7	846 360	3 611 349	408 566
B) RAD	Nodal radiation pattern accounted for; No frequency limits	2.93	100.7	712 077	3 137 154	330 322
C) 60 mHz	Nodal radiation pattern accounted for; Upper frequency limit at 60 mHz	2.93	50.0	692 540	3 055 588	293 827
D) 43 mHz	Nodal radiation pattern accounted for; Upper frequency limit at 43 mHz	2.93	34.6	685 146	2 897 027	226 194
E) Model	Same as B	2.93	87.1	521 705	2 302 157	171 260

seismic wave radiation pattern. There are several reasons to avoid fitting such seismograms. First, there is a greater likelihood that these portions of the seismogram will be dominated by scattered waves, and that the synthetic seismograms used to model them will be the most affected by phase errors due to source-mechanism uncertainties. Finally, waveform sensitivity kernels increase in complexity near to nodes in the radiation pattern, and as a result sample a larger volume lying off the source–receiver great-circle path (Meier *et al.* 1997; Lebedev *et al.* 2005).

The initial data set of waveforms fits (A) was generated without accounting for the radiation pattern, and resulted in 846 360 successfully fit seismograms. Reprocessing the same seismograms accounting for near-nodal radiation resulted in 712 077 waveform fits (data set B). Avoidance of nodes in the radiation pattern thus cost ∼135 000 seismograms, ∼19 per cent of the total in data set B. We computed the waveform fits both with and without considering the radial notes for completeness, obtaining an estimate of how many waveform fits in the data set A were 'near-nodal'. For tomography, we shall proceed using data set (B), with near-nodal signals excluded.

A2 High-frequency cut-offs

The validity of the path-average (sensitivity-area average) approximation (1, 2) depends upon the smallness of differences between the sensitivity-area averages of phase-velocity derivatives $[\delta C_m^0(\omega)/\delta\beta(r)]$ and the derivatives at every point $[\delta C_m(\omega, \theta, \phi)/\delta\beta(r)]$ within the sensitivity area (Lebedev & van der Hilst 2008). Essentially, this means that for paths crossing strong lateral heterogeneities, the approximation may no longer be valid; $\delta C_m(\omega, \theta, \phi)/\delta\beta(r)$ at some points $\langle\theta_i, \psi_i\rangle$ within the kernel may deviate substantially from $\delta C_m^0(\omega)/\delta\beta(r)$.

Lebedev & van der Hilst (2008) performed a series of tests to quantitatively investigate the effect of 3-D heterogeneity in the sensitivity areas, using the misfit of synthetic seismograms as a measure. They observed that enforcement of tighter misfit limits resulted in most fundamental-mode time–frequency windows at higher frequencies being rejected, due to the lateral heterogeneity of the crust. Importantly, however, their upper-mantle images were not affected significantly. Therefore, they imposed a uniform high-frequency limit of ∼44 mHz (23 s) for the centre frequency of their Gaussian filters for their final processed data set used in the tomographic model.

We explored the effect the upper frequency limit has on the waveform fitting procedure and resulting tomographic models using our much larger data set of waveform fits. The starting data set of 3.6 million seismograms has been reprocessed using two different cut-offs (both account for nodal radiation patterns), and are referred to as data sets C and D. Data set C was processed imposing a similar 60 mHz maximum (∼48 mHz average filter centre) frequency cut-off as that used by Lebedev & van der Hilst (2008), whereas data set D uses a stronger cut-off of 43 mHz (∼33 mHz filter centre). As expected, a lower cut-off decreases the quantity of successful waveform fits: only 692 540 fits for 60 mHz and 685 146 for 43 mHz.

The effects of the upper frequency limit as it pertains to the resulting tomographic inversions were of particular interest. To this end, three inversions were performed using a common set of ∼540 000 successful fits. Use of the same seismograms in each test enables a more consistent appraisal of the effects that frequency band has on the inversion and resulting model.

Minor variations in frequency content of each data set results in small differences in the dimensions of the inverse problem: data set B uses 1 635 342 equations and data set D 1 586 951 equations. As data set C uses a cut-off intermediate to B and D, and, as expected the results lie within the range given by B and D; therefore attention is paid solely to these end-members. Each inversion was run for 3000 iterations, yielding variance reductions of ∼90 per cent, and model norms within 5 per cent of each other.

Fig. A1 shows three maps at 20, 36 and 585 km depth (left to right) through three models generated using different data sets. The top panel shows results for data set B and the middle panel for D. The largest differences are expected in the crust and transition zone. Shallow structure is sensitive to higher frequencies, therefore restricting the maximum frequency limits resolution. In the transition zone, higher modes are critical for resolving structure. However, reducing the upper frequency limit reduces the higher mode content and therefore decreases resolving power below 250–300 km depth.

Relatively little difference in apparent resolution is observed at the shallow depths (20 and 36 km) between models B and D (top and

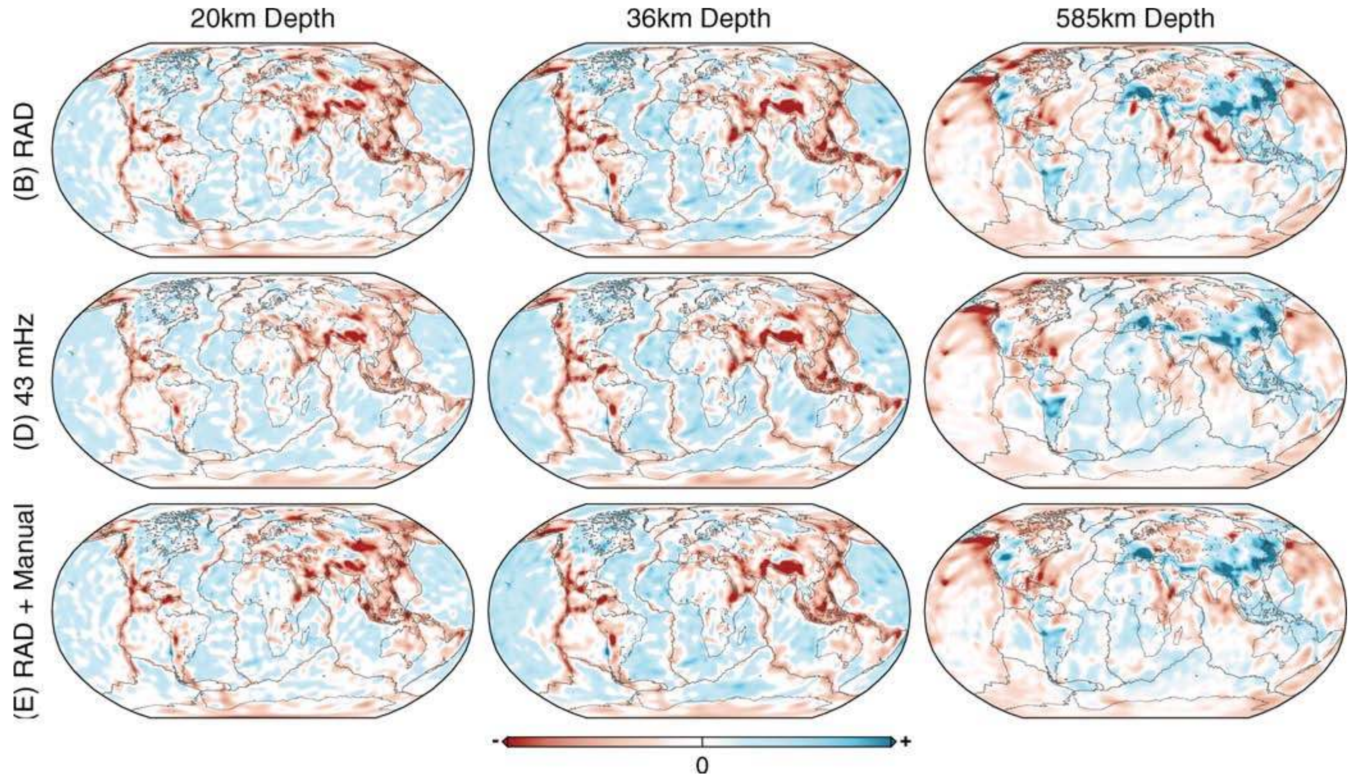

Figure A1. Comparison of tomographic results from three different inversions using a common set of ∼540 000 seismograms. The top panel illustrates the model generated using the fits drawn from data set B (Table A1, no upper frequency limit), whereas the middle panel shows the model generated drawing the same set of seismograms from data set C (cut-off of ∼43 mHz). Three depths at 20 km (left), 36 km (centre) and 585 km (right) are illustrated for each. Perturbations are saturated at ±360 m s^{-1} for both 20 and 36 km, and ±107 m s^{-1} at 585 km depth. Note that at depths 20 and 36 km the perturbations are with respect to CRUST2 when within the crust and to mantle reference when in the mantle, whereas at 585 km depth the perturbations are with respect to the mantle reference model (equivalent to variations of ±2 per cent). The bottom model (E, Table A1) is generated using the same initial data set as the top panel (∼540 000 best fits from data set B) as a starting point, however an additional ∼20 000 paths were manually selected and removed to reduce artefacts in the transition zone.

middle panels). Amplitudes of the largest anomalies are reduced by several tens of metres per second, however, the general structure remains largely unchanged.

At lithospheric mantle depths, both models are equally well constrained, and exhibit few differences. The maximum amplitudes are reduced by at most 35 m s^{-1} (∼0.7 per cent), and the shape of anomalies remains the same. A portion of the amplitude reduction results from slightly higher effective damping for data set D, as ∼3 per cent fewer equations are incorporated (with the same regularization coefficients).

The largest differences are at depths ≥250 km, where reduction in higher modes reduces resolving power notably (e.g. Fig. A1, 585 km). Overall, the model norm is reduced ∼20–30 per cent in data set D, and therefore features appear smoother, and in many cases with a lower amplitude.

The primary reason for reducing maximum frequency during waveform fitting was to enforce more strict validity criteria, reducing errors due to assumptions of constant Fréchet derivatives across the sensitivity areas. It is possible that such errors may propagate into the model and result in artefacts. As is suggested by Lebedev & van der Hilst (2008) and observed in our tests here, however, errors due to the assumption of constant phase-velocity derivatives are small.

An examination of the two models presented in Fig. A1 (top two panels) reveals that, although some artefacts are reduced, they are not eliminated. Therefore, they are likely to be due to other errors.

A3 Treatment of remaining errors

The main sources of remaining errors are event location errors, incorrect origin times and source parameters, station timing errors and unmodelled diffraction of surface and body waves. The impact of errors in event locations and origin times on tomography was examined by Lebedev *et al.* (1997) and found to be limited. Using a much smaller data set in the Philippine Sea region, two individual inversions were performed: the first used locations and origin times derived from short-period body wave arrivals (NEIC catalogue), whereas the other used the Harvard CMT catalogue. Anomalies in the resulting tomographic models did not differ substantially, despite large systematic differences in source parameters. The effect of unmodelled diffraction on AMI tomography was tested in 'spectral-element' resolution tests (Lebedev & van der Hilst 2008; Qin *et al.* 2008), which showed that the sensitivity-area-average approximation was adequate for the recovery of anomalies that were sufficiently well sampled by crossing rays. Although these previous tests suggest that the effect of errors in the data is overall limited if the data sampling is dense, isolated artefacts often remain in the tomographic models. Based on the analysis described earlier, we have chosen to retain data set B (no frequency limit), and perform a manual analysis to identify and remove additional seismograms potentially contaminated by errors.

The final data set, E (Table A1), includes the best ∼540 000 seismograms selected (using outlier analysis) from data set B. This was further reduced, beginning by removing all seismograms for events prior to 1994, as the moment-tensor solutions and source locations for these older events commonly have larger errors and are less well constrained. Next, the locations of suspected artefacts were compared with the locations of stations and events. As the sensitivity kernels have the largest values near their endpoints, errors may concentrate in these regions and result in a corresponding increase in anomaly amplitude. Seismograms recorded at stations or originating from events in close proximity to apparent artefacts were identified, examined and discarded if deemed suspicious.

This process of manual analysis identified $\sim$20 000 additional seismograms for removal. The results from the inversion of this data set are plotted in the bottom panel of Fig. A1, for comparison with the previous two inversions. In this case, the amplitudes in the crust and shallow mantle obtained in the top panel are maintained, as are the amplitudes in the transition zone. In addition, the reduction of transition zone artefacts previously achieved by limiting the maximum frequency during fitting (middle panel) has been reproduced. Therefore, this subset of 521 705 (data set E, Table A1) was selected for our final tomographic model.

APPENDIX B: ANALYSIS OF WAVEFORM FITTING RESULTS

In this section, we explore the properties of the waveform fits generated by AMI as a function of the different constraints applied during waveform fitting. As discussed previously in Appendix A, the result of numerous frequency-limit tests and outlier removal was 521 705 of the most mutually consistent vertical-component waveform fits used to generate the final tomographic model. In the following plots, we examine differences in the properties of four different full data sets and the final, ‘model’ data set (E, Table A1). This comparison offers insight into the statistical nature of the effects of the frequency limits during fitting, as well as what constitutes an ‘outlier’.

Fig. B1 illustrates the path-length distribution of each full AMI data set (A–D) and model subset (E, orange). The top panel represents seismograms, whereas the middle and bottom panels show histograms for the fundamental- and higher mode time–frequency windows, respectively. In all three panels, the distributions are bimodal. The secondary lobe centred at 10 000 km results from numerous USArray stations sampling seismicity in the western Pacific. Such a double-lobe distribution has also been observed in regional-scale modelling, where the local seismicity dominated over several large-distance events, included to help constrain the structure at the model domain boundaries (e.g. Legendre *et al.* 2012).

The different constraints applied during waveform fitting affect the distributions. Accounting for the radiation pattern reduces the number of fits relatively evenly across all distance bins (white compared to green, all panels). However, the restriction of the upper frequency limit during waveform fitting reduces the number of the shorter paths more than the number of longer paths (green compared to purple and blue); this effect is particularly clear in the higher mode panel (bottom), where a significant proportion of paths lie in the range 1200–3500 km. The *a posteriori* outlier removal appears to have a similar effect: shorter path lengths are preferentially removed. This is expected, as the fits at short distances are affected more by source mislocations and timing errors than those for longer paths.

Fig. B2 plots the distribution of Gaussian filter centre periods for each data sets. Both fundamental and higher mode time–frequency windows sample the broad period range from 10 to $\sim$320 s. However, since the histogram represents only filter centre periods, the

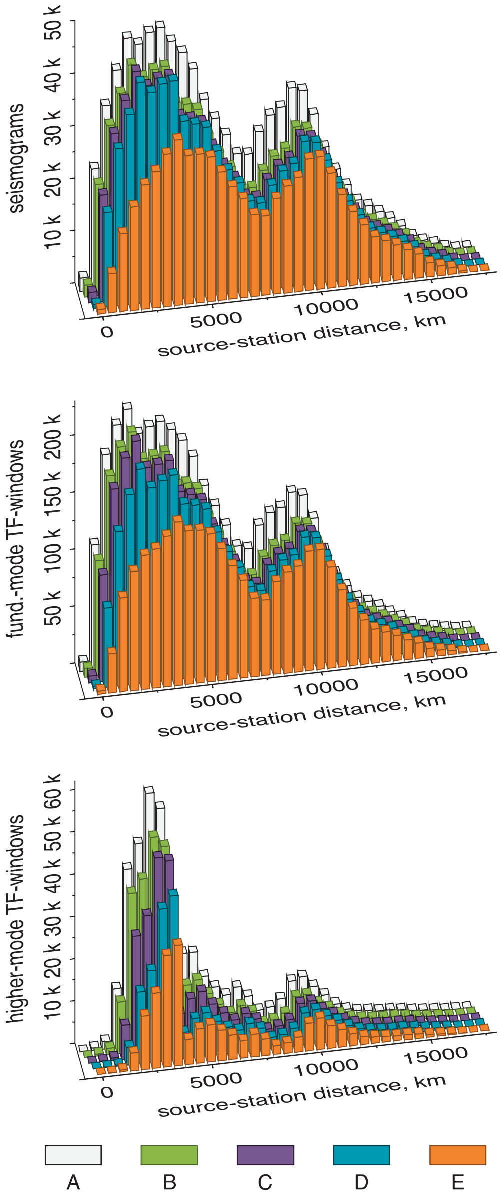

Figure B1. Path-length distribution of successfully fit seismograms. The five different colours represent the different data sets fit using AMI (Table A1): data set A (white) with no cut-offs and ignoring the radiation pattern; data set B (green) accounts for the radiation pattern; data set C (purple) additionally imposes a 60 mHz high-frequency cut-off; data set D (blue) imposes a 43 mHz high-frequency cut-off and E (orange) is the subset of 521 705 waveform fits selected from data set B used to compute the final tomographic model. Top panel is the histogram of the number of seismograms, the middle panel is the histogram for the number of fundamental-mode wave trains ($\geq$2 time–frequency windows for each seismogram) and the bottom panel shows the number of distinct higher mode wave trains (one count per time–frequency window with $\geq$1 higher mode fit). The largest contribution comes from paths between 2500 and 7500 km. The secondary lobe centred at $\sim$10 000 km results from a large number of circum-Pacific paths between stations of the USArray TA and western Pacific seismicity. Higher modes (bottom panel) are clearly dominated by shorter path lengths ($\leq$4000 km). *Y*-axis is linearly scaled and indicates the number of seismograms or windows, in thousands (e.g. 55 k $\equiv$ 55 000).

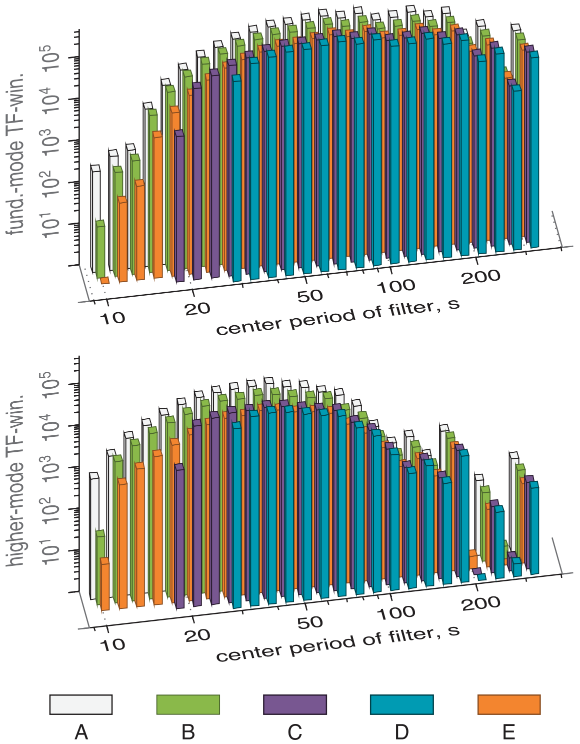

Figure B2. Histogram of the Gaussian filter centre periods for the fundamental (top panel) and higher (bottom panel) modes. As in the previous figure, colours represent the different data sets in Table A1. For the fundamental mode, one count indicates an arrival of this mode within a single time–frequency window for a successfully fit seismogram. For higher modes, more than one mode usually contributes to each wave train or time–frequency window. See Fig. B3 for the contribution of the individual modes. Note that in the axis labels, *TF-win* indicates 'time–frequency windows'.

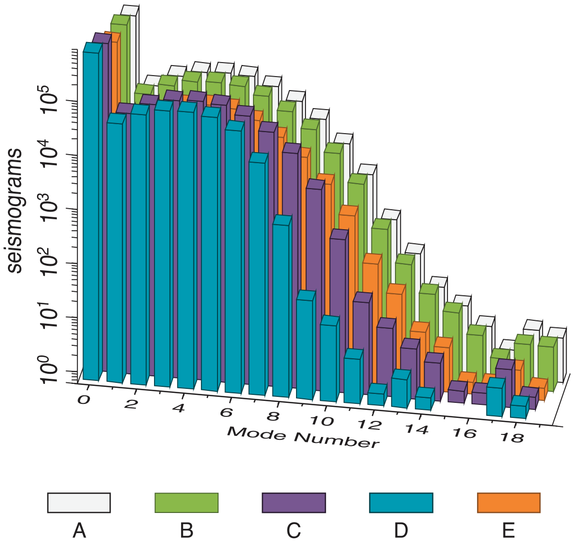

Figure B3. Contributions of the fundamental (mode 0) and higher modes in each of the five data sets in Table A1, colour coded as in the previous two figures. The number of the fundamental-mode curves is an order of magnitude greater than that of the higher modes, in part because the fundamental mode was required to be included for a waveform fit to be accepted. There are fewer higher mode curves because *S* waves are not included in all waveform fits. The effect of the cut-off frequency used during waveform fitting is clearly visible, and results in a decrease of the number of higher mode phase-velocity curves with increasing minimum period (decreasing maximum frequency). Note that the *Y*-axis is logarithmically scaled.

full finite width of the Gaussian filters broadens the complete period range, in particular extending to longer periods. Therefore, the full range spans 10–455 s (observed in Fig. 17). In Fig. B2, the effect of the 60 mHz (C, 16 s) and 43 mHz (D, 23 s) upper frequency cut-offs (purple and blue, respectively) is clear. The 16 s (purple) cut-off results in a minimum filter centre period of ∼19 s, with no time–frequency windows at shorter periods. For the 23 s (blue) cut-off, the minimum filter centre period is ∼29 s. At long periods, the number of time–frequency windows are similar for all the data sets.

The fundamental-mode waveform fits in data set E (model) sample the range 35–200 s almost uniformly, with a drop-off (approximately half an order of magnitude) at periods from 200–350 s. For higher modes, sampling is strongest in the period range 20–100 s, and decreases at longer periods (100–350 s). It is important to note that the counting of higher mode time–frequency windows is incremented only once for each successfully fit window, not for each higher mode in the window. Since a higher mode wave train is generated through interference of a number of modes, this distribution does not reflect the number of individual higher modes at each period.

The distribution of fundamental- and higher modes, measured by AMI after waveform fitting, are plotted in Fig. B3, for each data set. AMI employs conservative criteria for the selection of frequency ranges in which a given mode has a sufficiently strong contribution to the waveform fit and can, thus be measured. Therefore, the higher mode content represents a conservative, lower limit estimate of the actual higher mode contributions.

The fundamental mode, indicated by mode number 0, has one phase-velocity curve for every successfully fit seismogram; this is one of AMI's criteria for accepting a waveform fit. The different constraints imposed during waveform fitting are clear. The inclusion of nodal radiation patterns has a minimal effect across the higher modes (green compared to white). The frequency cut-offs however, have a stronger influence, with greater reduction for the narrower frequency bands (D versus C). The 60 mHz (purple) and 43 mHz (blue) data sets show progressively fewer higher modes, which, as discussed in Section A2, reduce data redundancy and therefore resolution and recovered amplitude in the transition zone.

The first seven higher modes contribute most significantly in data set E (model), with ∼10 000–80 000 (modes three through five only) dispersion curves. For modes 8–10, thousands of phase-velocity curves are measured, and account for ∼5 per cent of the total overtones. At the highest mode numbers (11–18) less than 500 curves are measured, and contribute only ∼0.05 per cent. We can compare the number of higher modes we obtain with other studies, for example, Visser *et al.* (2008, their table 2). Although they obtain more first overtone measurements, our new data set contains much more measurements at higher modes, by a factor of two for modes 4–6. The selection criteria for our new data set are also more strict (compared to that outlined in Visser *et al.* 2007), with much closer data-synthetic fits required.