Imaging mantle transition zone thickness with *SdS-SS* finite-frequency sensitivity kernels

Jesse F. Lawrence[1,2]* and Peter M. Shearer[1]

[1]*IGPP, Scripps Institution of Oceanography, La Jolla, CA, USA. E-mail: jflawrence@stanford.edu*
[2]*Geophysics, Stanford University, Stanford, CA, USA*

SUMMARY

We invert differential *SdS-SS* traveltime residuals measured from stacked waveforms and finite-frequency sensitivity kernels for topography on the 410- and 660-km discontinuities. This approach yields higher resolution images of transition zone thickness than previous stacking methods, which simply average/smooth over topographic features. Apparent structure measured using simple stacking is highly dependent upon the bin size of each stack. By inverting for discontinuity topography with a variety of bin sizes, we can more accurately calculate the true structure. The inverted transition zone model is similar to simple stack models with an average thickness of 242 km, but the lateral variations in thickness are larger in amplitude and smaller in scale. Fast seismic velocities in 3-D mantle models such as SB4L18 correlate with areas of thicker transition zone. The elongated curvilinear regions of thickened transition zone that occur near subduction zones are narrow and high amplitude, which suggests relatively little lateral spreading and warming of subducted lithosphere within the transition zone. The anomalously thin transition zone regions are laterally narrow, and not broadly continuous. If these variations in transition zone thickness are interpreted as thermal in nature, then this model suggests significant temperature variations on small lateral scales.

Key words: Tomography; Phase transitions; Body waves; Interference waves; Computational seismology.

1 INTRODUCTION

The mantle transition zone is a dynamic region in the Earth between about 400 and 750 km depth where density and seismic velocity increase rapidly with depth (e.g. Dziewonski & Anderson 1981). Much of this increase in seismic velocity and density occurs over relatively sharp discontinuities at approximately 410, 520 and 660 km depth. Laboratory experiments demonstrate that these discontinuities likely represent pressure induced phase changes of α olivine to β-spinel structure at ~410 km, β-spinel to γ-spinel at ~520 km and γ-spinel to silicate perovskite and magnesiowüstite at ~660 km (Ringwood 1975; Jackson 1983; Ito & Takahashi 1989). The Clapeyron slopes of the 410- and 660-km discontinuities have opposite signs (e.g. Katsura & Ito 1989), which should cause the distance between the 410- and 660-km discontinuities to thin in warm regions and thicken in cold regions. For the remainder of this document we refer to the distance between the 410 and 660 as the transition zone thickness, or W_{TZ}. The ringwoodite to perovskite and magnesiowüstite phase transformation that occurs at the 660 is also accompanied by a majorite-garnet transformation between about 660 and 750 km depth (e.g. Ito & Takahashi 1989). Many seismic studies have confirmed that the transition zone thickens near subduction zones and thins elsewhere (e.g. Flanagan & Shearer 1998; Gu *et al.* 1998; Gu & Dziewonski 2002).

Resolving undulations on the discontinuities has been a key focus of seismology for the past several decades (e.g. Vinnik 1977; Shearer & Masters 1992; Li *et al.* 2003). Over the years many small-scale studies have produced detailed images of transition-zone topography and/or thickness using receiver functions for regions with sufficient data coverage (Vinnik 1977; Petersen *et al.* 1993; Bostock 1996; Shen *et al.* 1996; Vinnik *et al.* 1996; Dueker & Sheehan 1997; Gurrola & Minster 1998; Li *et al.* 1998, 2003; Shen *et al.* 1998; Gilbert *et al.* 2003; Lawrence and Shearer 2006a). Global coverage is best achieved using *SS* precursors, but these studies have been relatively long-wavelength, often averaging over areas as large as 20° in diameter (Shearer 1991, 1993; Shearer & Masters 1992; Gossler & Kind 1996; Lee & Grand 1996; Flanagan & Shearer 1998; Gu *et al.* 1998; Chevrot *et al.* 1999; Deuss & Woodhouse 2001, 2002; Gu & Dziewonski 2002).

One of the major obstacles to globally resolving small-scale discontinuity topography is that the discontinuity phases have low amplitudes due to the relatively low impedance contrast at each discontinuity. Consequently, these phases are usually at or below the amplitude of the ambient noise, so many waves must be stacked together to improve the signal-to-noise ratio (SNR). Another hindrance to obtaining high-resolution images of global transition zone structure is the limited spatial coverage of high signal-to-noise waveforms. *Pds* (*P*-to-*S* converted phases from an interface at depth *d*)
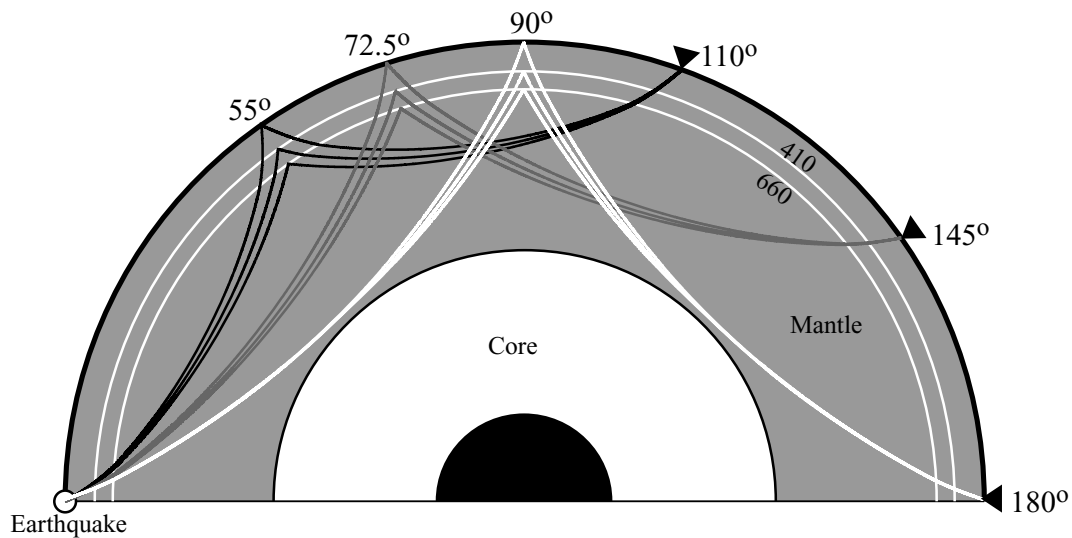
GJI Seismology

Figure 1. This figure depicts the ray paths for *SS*, *S410S* and *S660S* for three event-to-station distances. 110° (black), 145° (medium grey) and 180° (white).

and *Ppdp* (topside *P* reflections from the surface and an interface at depth *d*) are only measurable within a few degree region beneath a seismic station (Lawrence & Shearer 2006b). While global studies have been done with *Pds* (Chevrot *et al.* 1999; Lawrence & Shearer 2006a), the uneven lateral coverage of seismic stations limits the global resolution to harmonic degree 6 or less. *SdS* and *PdP* (underside *S* and *P* reflections at depth *d*, Fig. 1) provide much better global coverage of the transition zone, having sensitivity near the bounce point, which is about half way between the station and earthquake.

While *SdS* and *PdP* have similar paths, and can be examined with equivalent data sets, *PdP* typically has much lower amplitude than *SdS*. In particular *P660P* is at or below the level of the noise even after stacking thousands of waves (e.g. Estabrook & Kind 1996; Shearer & Flanagan 1999; Lawrence & Shearer 2006b). We focus on the *SdS* phase here because it provides the combination of good global coverage and a relatively high SNR.

Despite having higher amplitudes than other discontinuity phases, the *SdS* amplitudes are almost always too low to observe on individual seismograms, which impedes discontinuity topography studies. Previous studies stacked data into large circular bins (20° diameter) according to the location of the ray theoretical bounce point (e.g. Flanagan & Shearer 1998; Gu *et al.* 1998; Gu & Dziewonski 2002). This method, while varying slightly among studies, resulted in roughly similar long-wavelength structures. Stacking effectively averages the traces within a bin, which acts to damp amplitudes and smooth small-scale topography. Li *et al.* (2003) argued that the large radii stacks of *SdS* measurements naturally result in longer wavelength variations in 410-km discontinuity depths than the topography measured with smaller radii stacks of *Pds*. Choosing a smaller *SdS* bin radius would theoretically reduce the smoothing effect, but stacks with fewer data would be less reliable. Creating additional overlapping bins at distances shorter than half the bin radius would cause aliasing between short and long-wavelength structure.

Gu *et al.* (2003) attempted to solve for discontinuity topography using *SS-SdS* traveltimes that are interpolated from *SS-SdS* traveltime residuals of adjacent large circular bins. Gu *et al.* (2003) inverted these interpolated data using a damped least-squares method based on 1-D traveltime kernels for topography that are weighted by spherical splines. The interpolation of data from the stacked *SdS* traveltime residuals likely underestimates the variance in the true

data, resulting in a smooth solution with overly damped topography similar to that of previous stacking studies. Additionally, the geometry of a 1-D sensitivity kernel smoothed with a spherical spline has no theoretical basis and only has crude similarity to the true sensitivity kernel.

A single *SdS* wave is actually sensitive to a large area, not just the ray theoretical bounce point. This sensitivity extends in an X-shaped pattern (not a circular one) for 1000 s of kilometres (Fig. 2a), and is not limited to the bins of previous studies. The Fresnel zone (Fig. 2b) outlines the region within which *SdS* waves reflect with traveltimes within $\pm T/4$ of the ray theoretical traveltime, where *T* is the period. *SdS* energy following a ray path that bounces closer to the seismometer or earthquake will arrive earlier than the ray theory *SdS* wave. This is because more of the ray path is spent at greater depths, where seismic velocities are higher and smaller lateral distances traversed at depth correspond to greater distances at the surface. In contrast, *SdS* energy that reflects at points off the path from the event to the station will arrive later due to the greater distance it must travel. Along the arms of the X-shaped pattern, the effects of deeper paths and greater cross-path distances balance each other.

In order to observe short-wavelength features with *SdS* waves one must take these sensitivities into account. Neele & de Regt (1999) presented a method that inverts for discontinuity topography using *PdP* or *SdS* traveltimes and amplitudes. While the method is promising in theory, it fails in practice due to the fact that traveltimes and amplitudes are not reliable for individual *SdS* and *PdP* waves. As mentioned above, in order to obtain stable *SdS* amplitudes and traveltimes, it is advisable to stack many waves.

Using finite-frequency theory, Dahlen (2005) calculated the *SdS*-*SS* traveltime kernel for lateral variations in topography on an interface at depth. Here, we develop a method for employing these finite-frequency kernels in tandem with stacking to invert for finer structure than previous studies. We provide only a cursory discussion of the finite-frequency kernels themselves, referring readers to Dahlen (2005) for further detail. Instead we focus on describing how we employ the kernels in a stacking and inversion method for global *SdS* waveforms. We then present the results with application to real and synthetic data, and discuss the reliability of the method. Finally, we interpret our transition zone thickness maps in light of other mantle features, such as plumes and subducted slabs.
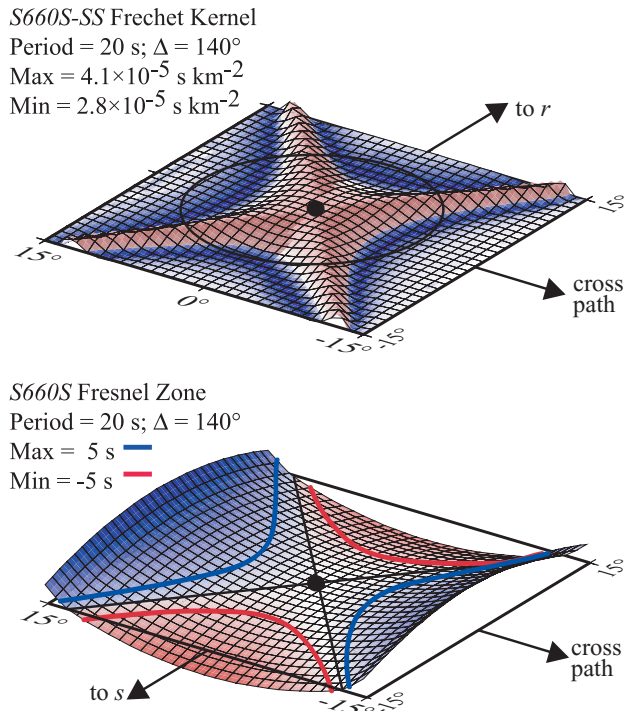
S660S-SS Frechet Kernel
Period = 20 s; Δ = 140°
Max = 4.1×10^{-5} s km^{-2}
Min = 2.8×10^{-5} s km^{-2}

S660S Fresnel Zone
Period = 20 s; Δ = 140°
Max = 5 s
Min = -5 s

Figure 2. (a) The *S660S-SS* differential traveltime Fréchet kernel, *K*, with sensitivity to topography for an event that is 140° from the station with a dominant period of 20 s. (b) The *S660S* Fresnel zone is defined by the region where traveltime variations that result from perturbations in *S660S* bounce location are less than *T*/4 from the ray theoretical value, where *T* is the dominant period. The directions towards the receiver, *r*, towards the source, *s* and the cross-path direction are indicated. The black dot at the centre identifies the ray theoretical bounce point.

2 DATA PROCESSING

Owing to the large quantities of data necessary for this type of study we use an automated system for data selection and pre-processing. First, we download all available three-component long-period seismograms from the Incorporated Research Institutions for Seismology (IRIS) Data Management System (DMS) for Global Seismic Network (GSN) stations and some temporary network stations that recorded large magnitude (Mb > 5.8) earthquakes between 1976 and 2004. We remove the instrument responses, rotate the horizontal component records into the tangential direction (horizontal direction normal to the ray path), Parzan bandpass filter between 0.02 and 0.1 Hz, and measure the SNR. The SNR is measured as the maximum amplitude range (max–min) of the *SS* wave (in a window from 10 s prior to 50 s after the predicted *SS* arrival) relative to the maximum amplitude range of the noise (a 60 s window prior to the *SS* wave and its precursors). We limit the data to high SNR (>3) records with event-to-station distances between 110° and 175°. Within this distance range there are no strong phases that perceptibly interfere with tangential *SS* and its precursors. Only shallow earthquakes (depth < 30 km) are examined to ensure that the depth phases do not interfere with *SS* and *SdS*. The automated data selection reduces the number of records from more than 300 000 to 21 784 traces recorded at 619 stations.

The geographic locations of the seismic stations, recorded earthquakes, and ray theoretical *S660S* bounce points are shown in Fig. 3. The locations of *SS*, *S520S* and *S410S* bounce points are nearly identical to those of *S660S*, so they are not plotted. The source and
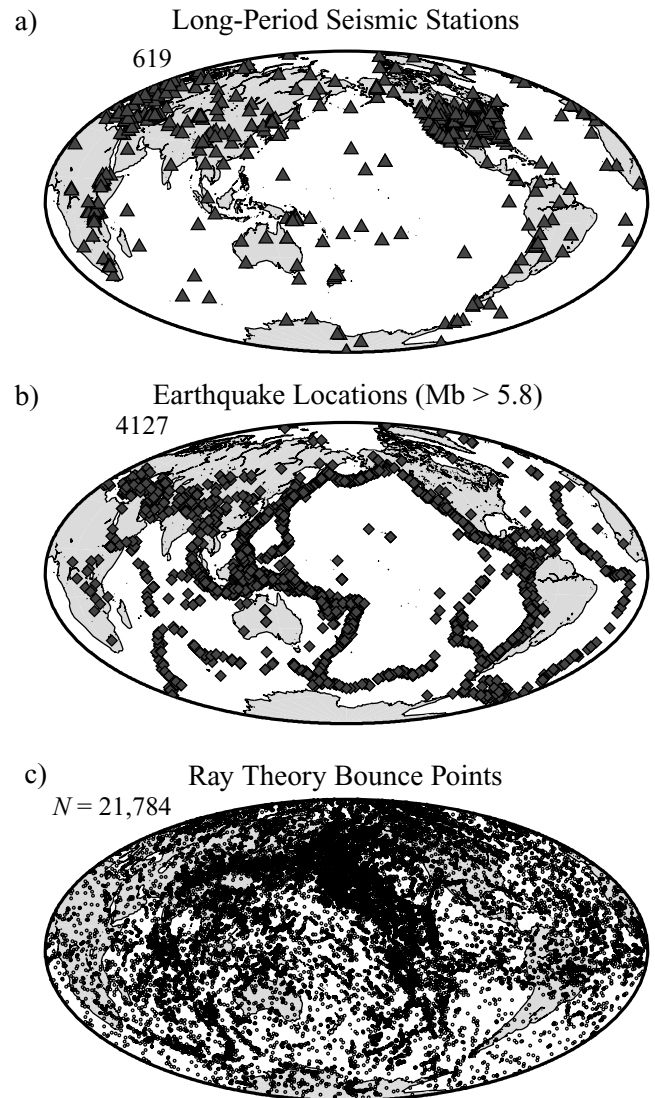
a) Long-Period Seismic Stations
619

b) Earthquake Locations (Mb > 5.8)
4127

c) Ray Theory Bounce Points
N = 21,784

Figure 3. The (a) long-period seismic stations, (b) seismic earthquake locations and (c) ray theoretical bounce points used in this study.

receiver distribution is non-uniform, with the stations distributed unevenly on land and most earthquakes occurring near plate boundaries. The distribution of ray theoretical *SS* bounce points is much better than that of the earthquakes and seismic stations. Nevertheless, more *SS* bounce points occur in the Pacific and Eastern Asia, and several locations have very low data coverage (e.g. southern South America).

3 METHOD

3.1 Mathematical background

In theory, an *SdS-SS* differential traveltime residual of the *i*th waveform, δt_i, is given by the sum of sensitivities, K_{ij}, multiplied by some finite topographic perturbations, Δz_j, on an interface at depth:

$$\delta t_i = \sum_{j}^{M} K_{ij} \Delta z_j. \tag{1}$$

Dahlen (2005) derived the *SdS* traveltime Fréchet kernel for topography at a surface point *x*,

$$K = \frac{1}{\pi} \frac{|\cos f|}{\beta} \sqrt{\frac{\cos^2 f |J_{rs}|}{|J_{xs} J_{xr}|}}$$
$$\frac{\int_0^\infty \omega^3 |\dot{m}(\omega)|^2 \cos \omega (T_{xs} + T_{xr} - T_{rs}) \, d\omega}{\int_0^\infty \omega^2 |\dot{m}(\omega)|^2 \, d\omega}, \tag{2}$$

where J and T are the ray theoretical Jacobian and traveltime to or from the source, s, or receiver, r, f is the angle from vertical of an incident or reflected *SdS* ray, β is the shear velocity and $|\dot{m}(\omega)|^2$ is the power spectrum of the *SdS* pulse. In practice the kernel is calculated by tracing each ray path from station to scatterer and scatterer to receiver for each location on the Earth's surface. This kernel neglects the near-zero sensitivity to the transmitted energy near the piercing points of the *SS* ray path through the discontinuity (2005). The *SdS-SS* sensitivity kernel is an X-shaped pattern with maximum amplitude near the ray theory bounce point and which decays in amplitude along the long arms of the X-shaped pattern (Fig. 2a). There is lessened sensitivity to topography with greater distance from the ray theory bounce point in the radial and tangential directions. Integrating the entire kernel yields the 1-D ray theoretical traveltime sensitivity kernel to the change in the discontinuity depth.

The power spectrum of the wave, $|\dot{m}(\omega)|^2$, controls the frequency content of the kernel. Therefore, using the correct power spectrum is important for determining accurate sensitivity kernels. Rather than using a theoretical power spectrum as in Dahlen (2005), we compute the power spectrum of the global stack of all aligned *SS* waves under the assumption that the stacked *SdS* pulse has the same power spectrum. We do not use distinct power spectra for individual kernels because the individual *SS* spectra are often less stable, providing semi-chaotic kernels.

In practice the differential traveltime residual of a small pulse with amplitudes at or below the ambient noise is only partially sensitive to structure, having significant contamination due to noise and/or interference. For this reason, we do not attempt to invert individual traveltime residuals for structure as suggested by Neele & de Regt (1999). Instead we develop an inverse method that employs differential traveltime residuals of stacked waves.

Stacking is simply a normalized summation that results in constructive interference of the signal and destructive interference of the noise. Stacking assumes that the signals in the trace are coherent and that the noise is largely random. This assumption becomes less valid as heterogeneity within the Earth causes greater systematic variations between the individual traces. Stacking based on the location of bounce points within the Earth (e.g. Flanagan & Shearer 1998; Gu & Dziewonski 2002) greatly increases the likelihood that signals are similar. By stacking, or averaging, waves in large bins, it is possible to obtain the long-wavelength topography of the transition zone discontinuities. Stacking effectively acts as an N_k-point mean, where N_k is the number of waves in the *k*th stack. The potentially incorrect assumption in previous *SdS* stacking studies is that the resulting traveltime residuals from a stack can be attributed to the geographical mean of the ray theory bounce points for all waves in the stack.

Here, we postulate that the *k*th differential traveltime residual, dt_{SdS-SS}, associated with a stacked waveform can be modelled as the normalized sum of sensitivities, K_{ij}, of the *i*th wave to the *j*th discrete topographic region,

$$dt_k = 1/N_k \sum_{i=1}^{N_k} \left[\sum_{j=1}^{M} K_{ij} \Delta z_j \right], \tag{3}$$

where there are N_k waves and M discrete topographic regions. This formulation requires that the individual signals within the stacked waveform are coherent. In practice, the difference between the majority of the differential traveltime residuals of the signals that compose the stack must be less than half the dominant period of the wave. Otherwise, in certain cases there could be more than one distinct pulse associated with the discontinuity, making a single mean traveltime residual inappropriate. As discussed later, we attempt to minimize the effect of such unstable stacks.

Because of the averaging effect of stacking, the summation order may be reversed, such that the traveltime residual is expressed as

$$dt_k = \sum_{j=1}^{M} \left[\sum_{i=1}^{N_k} K_{ij} / N_k \right] \Delta z_j. \tag{4}$$

In this form, the inner summation, or stack, becomes a single row matrix of stacked sensitivity values corresponding to M topographic regions. Because each individual stack, designated by k, results in a different stacked sensitivity, the inner summation can be written

$$S_{kj} = \left[\sum_{i=1}^{N_k} K_{ij} / N_k \right]. \tag{5}$$

Assuming that there are many differential *SdS-SS* traveltime residuals associated with many distinct stacks, these traveltime residuals can be calculated as

$$dt = S\Delta z. \tag{6}$$

Given this formulation, a topographic model Δz of interface depth perturbations can be determined from the stacked *SdS-SS* traveltime residuals d*t* and associated stacked sensitivity kernels, *S*. Before doing so, it is important to develop a consistent stacking methodology that maximizes resolution and does not violate the assumptions laid out above. In the following sections we present and apply such a method. First, we discuss the effect of discretizing the interface topography into nearly equal-area block structures. Then we present the effects of stack geometry, and the stability of both the sensitivities and the differential traveltime residuals.

3.2 Discrete sensitivity kernels

Partitioning the Earth into discrete cells for inversion purposes can affect the shape of a sensitivity kernel. Therefore, it is important to determine how discretization affects the stacked kernels and the resulting inversion. The size of the discrete areas has a large effect. At one limit, where the side length of each discrete quasi-square area is infinitesimal ($\delta x \rightarrow 0$), the discrete sensitivity kernel is equal to the continuous sensitivity kernel. At this limit there are an infinite number of semi-square areas, which is impractical for forward or inverse computations. At the other limit, where $\Delta x \rightarrow 360°$ and $\Delta y \rightarrow 180°$, there is one single kernel equal to the 1-D derivative of time with respect to reflection depth, $-\delta t/\delta z$. At this limit, there is no lateral resolution.

In order to determine the appropriate cell size that minimizes the number of topographic sections while maintaining an accurate representation of the Fréchet kernel, we test side lengths from 1° to 10° (Fig. 4). Each discrete kernel is the integral of the sensitivity kernel over its discrete area of that kernel. We approximate this integration by summing over 0.1° by 0.1 subareas. Side lengths of 2° or less approximate the continuous kernel with negligible distortion for waves with reasonably long dominant periods ($T_{max} > 10$ s). Larger discrete sections result in greater smoothing, which is most visible in the damping of the arms of the X-shaped pattern. Side-lengths
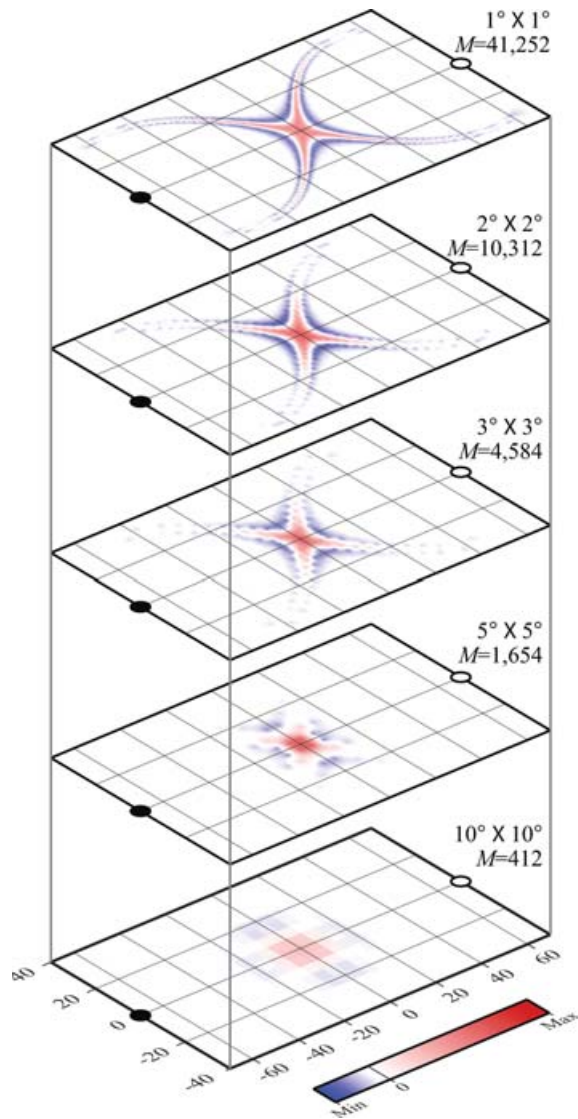
Figure 4. The discrete *S660S-SS* differential traveltime Fréchet kernel changes as a function of discrete block size. Here we present five kernels equivalent to Fig. 2 with discrete side-lengths from 1° to 10°. The 1°-side-length kernel approximates the continuous kernel. The 10°-side-length kernel approximates a 10° radius smoothing function.

of 2° provide a good compromise between the number of modelled topography parameters ($M = 10\,312$) in our inverse problem and the similarity between the continuous and discrete sensitivity kernels.

At this point, it should be stated that traditionally, seismic waves are considered 'insensitive' to velocity anomalies with scales much smaller than the dominant wavelength of the wave because short-wavelength features generally average out and the scale of short-wavelength features are ambiguous given a single wave. However, this is different from stating that long-period waves lack sensitivity to small-scale features. While it is difficult to distinguish between the sharpness of contrasts with single waves when the scale is shorter than the wavelength, it is still possible to invert scales shorter than the dominant wavelength with sufficient data. Otherwise long-period surface waves could be used to not invert for anything but continental-scale features, and small-scale features such as those observed by Ni *et al.* (2003) could not be observed.

3.3 Adaptive stacking

At the time of this study we find ~21 784 *SS* records that have suitably high signal-to-noise ratios (SNR > 3) to justify their use in our study. As previously noted, the individual *SdS* phases are too weak to yield accurate *SdS-SS* traveltimes, necessitating some form of stacking. Stacking *N* traces into one waveform removes any information that might be present about differences in the *SdS* times among different traces in the stack. However, the information loss can be reduced by stacking overlapping sets of waves in different configurations such that different traveltime residuals and sensitivities result from the different stacks.

Stacking data into overlapping bins by bounce point and allowing the bins to vary in size is a simple way to create numerous stacks with varying subsets of data that provide moderately different stacked *SdS-SS* traveltime residuals and stacked sensitivity kernels. This also tends to organize similar traces into common midpoint bins because the heterogeneity increases waveform differences as the ray paths diverge more. There is a trade-off between having large bins that produce more stable stacks by including more traces and having smaller, less stable stacks that yield more distinct traveltimes and sensitivity kernels. For this reason we employ an adaptive binning technique which is illustrated by the dark grey section in Fig. 5.

The adaptive stacking algorithm creates *M* bin foci located at the centres of the 2° topographic subareas and 10 bin radii for each bin focus. This ensures that the inversion is overdetermined. For each of the bin foci the first stack initially includes the 100 traces that have the ray theory bounce points closest to the bin focus. For *S410S* and *S660S* the automated algorithm identifies the time of maximum amplitude relative to the time of the *SS* peak amplitude. The mechanics of the slant stack summation and automated picking algorithm are described in more detail below. For each stacked differential *SdS-SS* traveltime, an equivalent ray theoretical depth is estimated from the 1-D velocity model, IASP91 (Kennett & Engdahl 1991; Kennett *et al.* 1995). An automated bootstrap method estimates the standard error of the *SdS* time/depth measurement. The stack is considered unstable if the standard error is greater than 5 km, in which case we increase the number of traces in increments of five and repeat the stacking process until a stable stack is produced for the first bin. Stacked traveltime residuals and sensitivity kernels are stored for each stable stack.

We progressively increase the radii for nine colocated bins by sequentially adding the 75 traces with the next closest bounce points to the bin focus. The bin radius, r_k, is defined by the maximum distance from the bin centre to the farthest bounce point in the stack. In bins with radii larger than ~20° the individual *SdS* sensitivity kernels often do not overlap, so the stacked kernels tend to be less coherent. For this reason, we reject stacks with $r_k > 20°$, which results in a ~17 per cent reduction of the data set. Fig. 6 maps the smallest bin radius for each of the bin centres. The bin radii decrease in regions with good data coverage (e.g. North Pacific) and increase in regions with reduced data coverage (e.g. southern South America). The larger stacks tend to damp the results and overlap more with other bins so the input traces and resultant stacks are more similar, which provides sensitivity to long-wavelengths. Bins with smaller radii have more distinct subsets of traces, yielding more dissimilar stacks and providing shorter wavelength resolution.

Each stack is created using the following technique. After each trace is pre-processed, we adjust the trace such that the maximum absolute amplitude of the *SS* pulse occurs at time zero and unit

Adaptive Stacking Flow Chart

```
┌─────────────────────────┐      ┌─────────────────────────┐
│ Store all long-period   │─────▶│ Find event-station      │
│ tangential data:        │      │ paths with              │
│ 1986-2003               │      │ 110°<Δ<175°             │
└─────────────────────────┘      └─────────────────────────┘

┌─────────────────────────┐      ┌─────────────────────────┐
│ Quality Control:        │─────▶│ Find bounce points &    │
│ SS S:N > 3              │      │ 3D time corrections     │
│                         │      │ for each trace          │
└─────────────────────────┘      └─────────────────────────┘

┌─────────────────────────┐      ┌─────────────────────────┐
│ Grid TZ into equal-area │─────▶│ Calculate one finite-   │
│ topographic sections    │      │ frequency kernel for    │
│                         │      │ each trace              │
└─────────────────────────┘      └─────────────────────────┘

┌─────────────────────────┐      ┌─────────────────────────┐
│ Create a series of      │─────▶│ Select 100 traces with  │
│ bin foci (1 per equal-  │      │ the closest SdS bounce  │
│ area section)           │      │ points to each node     │
└─────────────────────────┘      └─────────────────────────┘

┌─────────────────────────┐      ┌─────────────────────────┐
│ Stack selected traces & │◀─────│ Add 5 traces with the   │
│ sensitivity kernels     │      │ next closest SdS        │
│                         │      │ bounce points to the    │
│                         │      │ node                    │
└─────────────────────────┘      └─────────────────────────┘

If Nk < 1000 and rk < 20°                 Unstable stack

┌─────────────────────────┐      ┌─────────────────────────┐
│ Find time/depth for     │─────▶│ Bootstrap stack (25X).  │
│ S410S, S520S, & S660S   │      │ Unstable if σ > 5 km    │
└─────────────────────────┘      └─────────────────────────┘

                                          Stable stack

┌─────────────────────────┐      ┌─────────────────────────┐
│ Add 75 traces           │◀─────│ Store stacked time &    │
│ Go to next bin focus    │      │ stacked sensitivity     │
│ if Nk > 1000            │      │ kernel                  │
│ & Rk > 20°.             │      │                         │
└─────────────────────────┘      └─────────────────────────┘
        Once all nodes are done

┌─────────────────────────┐      ┌─────────────────────────┐
│ Invert stacked travel   │─────▶│ Output topography model │
│ times with stacked      │      │                         │
│ sensitivity kernels     │      │                         │
└─────────────────────────┘      └─────────────────────────┘
```
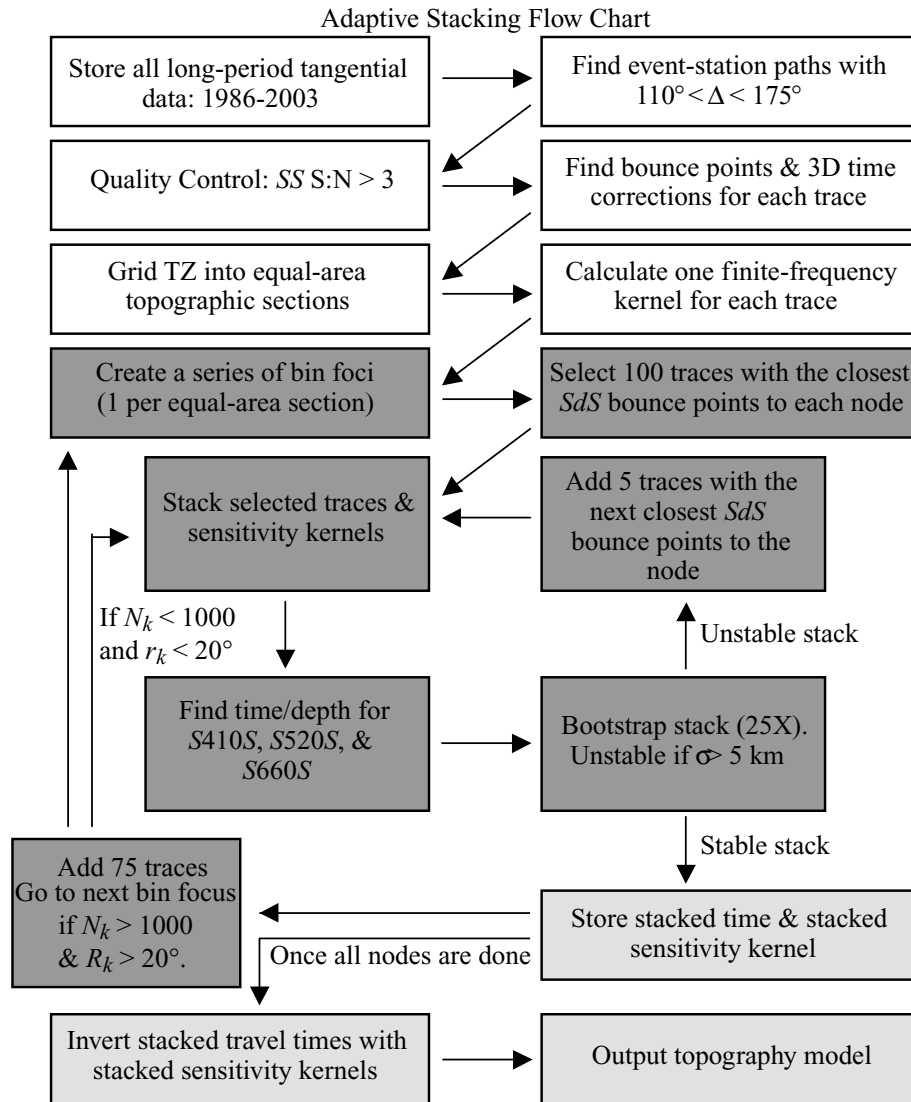
Figure 5. This flow chart describes the pre-processing (white), the adaptive stacking (dark grey) and the inversion process (light grey).

amplitude. For each time in the *SdS* to *SS* record the amplitude, $A(t)$, is mapped to the equivalent 140° event-to-station distance moveout by first calculating the depth associated with a *SdS* wave reflecting at that time and distance and then calculating the time associated with the same depth but at 140° distance. This mapping method transfers all *SdS* energy to the depth associated with a 140° Epicentral distance underside reflection rather than assuming a single moveout correction. There is minor waveform distortion in time as a result of this method, but this is preferable to mapping the distance correction with a physical distortion. Each 140° corrected trace is then interpolated so that aliasing is avoided and then added to the stack.

The sensitivity kernels for every trace must also be corrected to reflect the mapping of moveout to a 140° event-to-station distance. This is achieved by scaling the amplitude of the finite-frequency Fréchet kernel by the ratio of the 140° 1-D sensitivity kernel to the 1-D kernel for the true distance [$K_{ij,140} = K_{ij,\Delta} (\partial t/\partial z_{140})/(\partial t/\partial z_{\Delta})$]. The resulting individual kernel has the same pattern of sensitivity as that of the true distance, but the summation provides the correct value for the stacked 140° time–depth sensitivity.

3.4 3-D Heterogeneity corrections

3-D heterogeneity in seismic velocity and surface topography can change the stacked *SS* and *SdS-SS* traveltimes by several seconds. Near the receiver and earthquake hypocenter, the *SS* and *SdS* sensitivities to seismic velocity and topography are nearly identical, so differential traveltime anomalies largely represent 3-D structure near the *SS* and *SdS* bounce points. Applying corrections for 3-D structure is complicated in the case of stacked waveforms with differing ray paths, and there are many different ways in which these corrections could be applied. We have tested several techniques of pre-stack, costack, and post-stack corrections, and find that their results are generally similar, with variations much smaller than those obtained from changes in the assumed 3-D velocity model.

In this study the 3-D velocity corrections are applied by altering the Fréchet kernel. When we trace the 1-D ray theoretical *SS* and *SdS* ray paths through IASP91 from the source to each possible scatterer and then to the receiver, we correct the traveltimes and velocities in eq. (2) by the theoretical traveltime anomaly and velocities of CRUST2.0 (model available
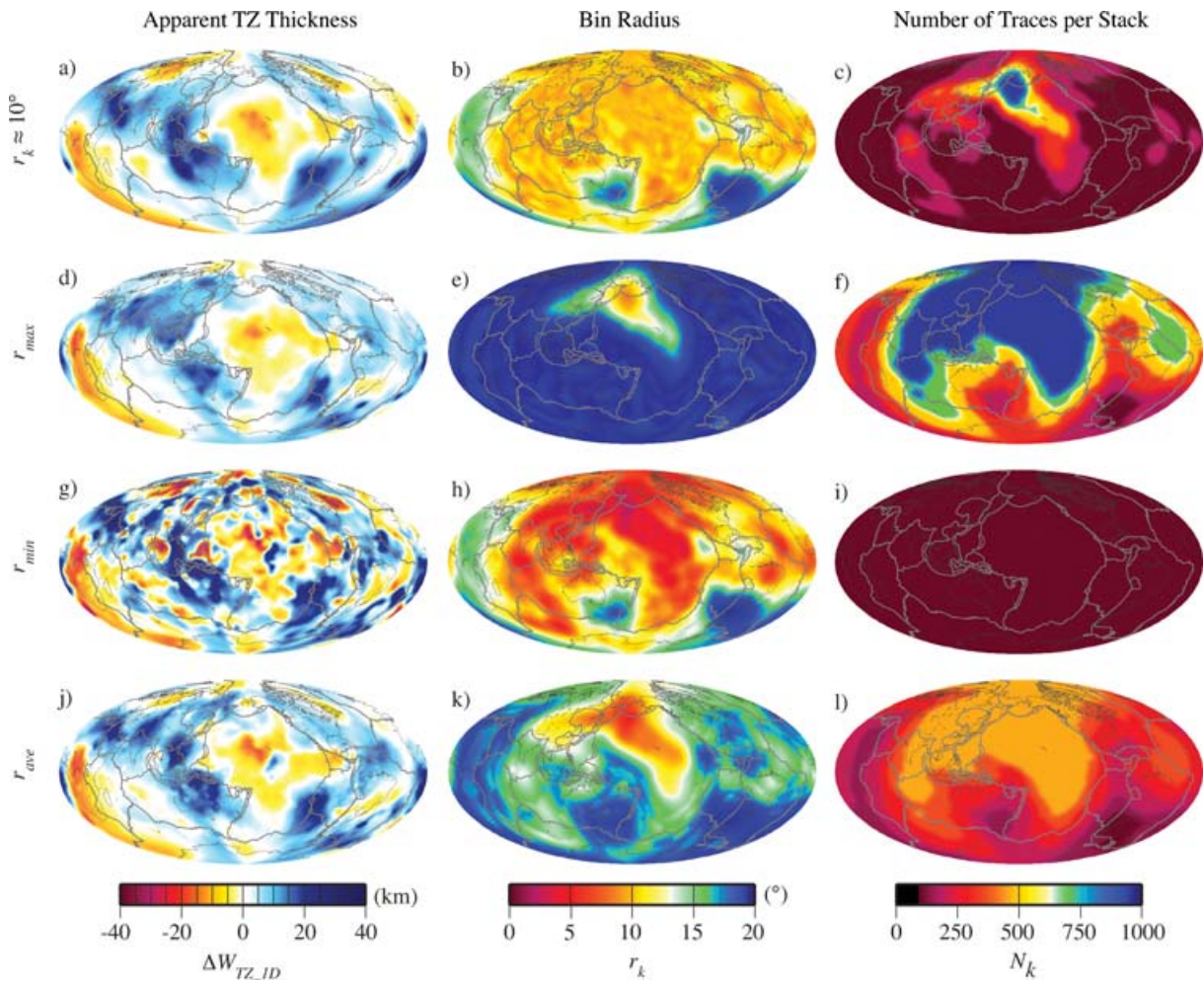
Figure 6. The variation in transition zone thickness, ΔW_{TZ_1-D}, estimated from 1-D time-thickness derivatives for bins with different radii. The upper row of maps describes (a) ΔW_{TZ_1-D}, (b) the bin radius r_k and (c) the number of traces per bin N_k for bins with radii closest to $10°$ for each bin focus. Maps of (d) ΔW_{TZ_1-D}, (e) r_k, and (f) N_k for the bins that have the largest radii per bin focus. Maps of (g) ΔW_{TZ_1-D}, (h) r_k, and (i) N_k for the bins that have the smallest radii per bin focus. Maps of (j) ΔW_{TZ_1-D} and (k) r_k, and (l) N_k for the bins that have the median radii per bin focus. Red indicates anomalously thick and blue indicates anomalously thin.

through http://mahi.ucsd.edu/Gabi/rem.html), SB4L18 (Masters *et al.* 2000), and an ellipticity correction (e.g. Dziewonski & Gilbert 1976). The smooth quality of most 3-D shear velocity models tends to result in similar corrected Fréchet kernels regardless of the assumed velocity model. Furthermore, because of the lateral averaging effect of stacking, the stacked sensitivity kernels also tend to be very similar. Perhaps in the future we will invert for seismic velocity variations and transition zone topography simultaneously as in Gu *et al.* (2003). Typically, in such inversions, the seismic velocity perturbations are largely dependent upon model parametrization and more standard body waves such as *S*, *ScS* and *SS*, and are generally unaffected by the discontinuity phases (Houser *et al.* 2008). Here, we prefer to focus on the topography given a well-known 3-D seismic velocity model.

4 STACKED DATA

For each of the $L = 103\,120$ possible bins (10 per bin focus) we add $100 \leq N_k \leq 1000$ waveforms into distinct bins with radii r_k. In Fig. 6, we present the transition zone thickness perturbations as

imaged with different sets of bins relative to a mean W_{TZ} of 242 km. For each case we present the apparent transition zone thickness perturbation (ΔW_{TZ_1-D}) estimated from 1-D time-thickness derivatives for one bin radius at each bin focus. Fig. 6(a) maps the apparent transition zone thickness given the stacked traveltimes of bins with radii closest to $10°$ for each bin focus. Note that at this point we are not yet inverting for the topography using the stacked sensitivity kernels. Rather we assign the value to each $2°$ cell simply from the stack of all the waveforms with bounce points within a specified radius. We focus on transition zone thickness rather than interface depth because the uncertainty in thickness is much less than the uncertainty in the absolute depth to either discontinuity, owing to the fact that the 3-D velocity models produce traveltime corrections for structure above 410 km that are much greater than the corrections for structure between 410 and 660 km. The imaged structure and method are equivalent to the $20°$ diameter stacking of Flanagan & Shearer (1998) and Gu *et al.* (1998). The number of traces stacked into each bin varies from 100 to 1000 with an average of 168. Note that some regions only have larger bin radii ($r_k > 10$), so these regions present the greatest deviations from previous studies.

In Fig. 6(b) the apparent $\Delta W_{\mathrm{TZ1-D}}$ variations are plotted for stacks with the largest bin radii for each bin focus. The long-wavelength pattern of $\Delta W_{\mathrm{TZ\_1-D}}$ is similar to Fig. 6(a). For all but the northwest Pacific stacks the bin radii are approximately $20°$. Each of these stacks has more data, with an average of 598 traces per bin. Fig. 6(c) shows the thickness perturbations as imaged by the smallest cap size for each bin centre. While these smaller bins sum at least 100 traces, the average trace count per bin is only 112. These stacks have higher amplitude and shorter wavelength variations in apparent transition zone thickness than the other stacks. The average bin size per bin focus yields a structure (Fig. 6d) that strikes a compromise between the structures imaged by the maximum and minimum bin radii.

The long-wavelength pattern of apparent transition zone thickness remains the same regardless of bin size and trace count. The transition zone is thicker beneath subduction zones and thinner beneath the Pacific and African plates. The smaller bin sizes isolate larger variations in transition zone thickness on shorter wavelengths. While some of this variability may be due to the smaller number of traces per bin, all stacks have bootstrap resampling estimates of standard error that are less than 5 km for the 410 and 660 apparent depths. The average standard error of $\Delta W_{\mathrm{TZ\_1-D}}$ for the smallest bins of each bin centre is less than 2.7 km. Nevertheless, a significant amount of aliasing compromises the reliability of the short-wavelength structure, especially for regions with fewer SdS bounce points.

The stacked sensitivity kernels typically have different characteristics than the individual sensitivity kernels. The binning process tends to focus the largest sensitivity to near the centre of the bin with a moat of lesser sensitivity outside the radius of the bin. Bins with larger radii typically have less coherent stacked sensitivity kernels because the individual kernels for each trace are less similar. Fig. 7 shows examples of four stacked sensitivity kernels with bins ranging from $5°$ to $20°$ in radius. The decreased coherence due to a larger bin radius leads to broader, lower amplitude stacked kernels. These broadened kernels indicate lower sensitivity to topography in poorly sampled regions. The small bin radii limit the event and station geometries more, so the stacked sensitivity kernels of smaller bins tend to include more similar individual sensitivity kernels. The resulting summed sensitivities more closely resemble the sensitivity kernels of the dominant event-to-station geometry.

Clearly, different sized bins for the same bin focus (e.g. Figs 7a and c) are sensitive to different topographic structures. The larger bin radii result in averaging over broader regions, which should cause the traveltimes to differ if the structure is different on long- and short-wavelengths.

5 INVERSION

We invert the stacked SdS-SS traveltime residuals and the stacked sensitivity kernels of eq. (6) for discontinuity topography using a damped conjugant gradient method (LSQR) (e.g. Paige & Saunders 1982; Nolet 1987). The sensitivity matrix constructed here has a higher percentage of non-zero elements than typical ray theory tomography inversions because of the large size of the individual sensitivity kernels. Each node of the inversion is weighted by the root-mean-squared sum of all the sensitivities for all the stacks for that node. Therefore, nodes with low data coverage are downweighted to reduce spurious, unresolved structure. The inversion is regularized with a first derivative smoothing constraint and is
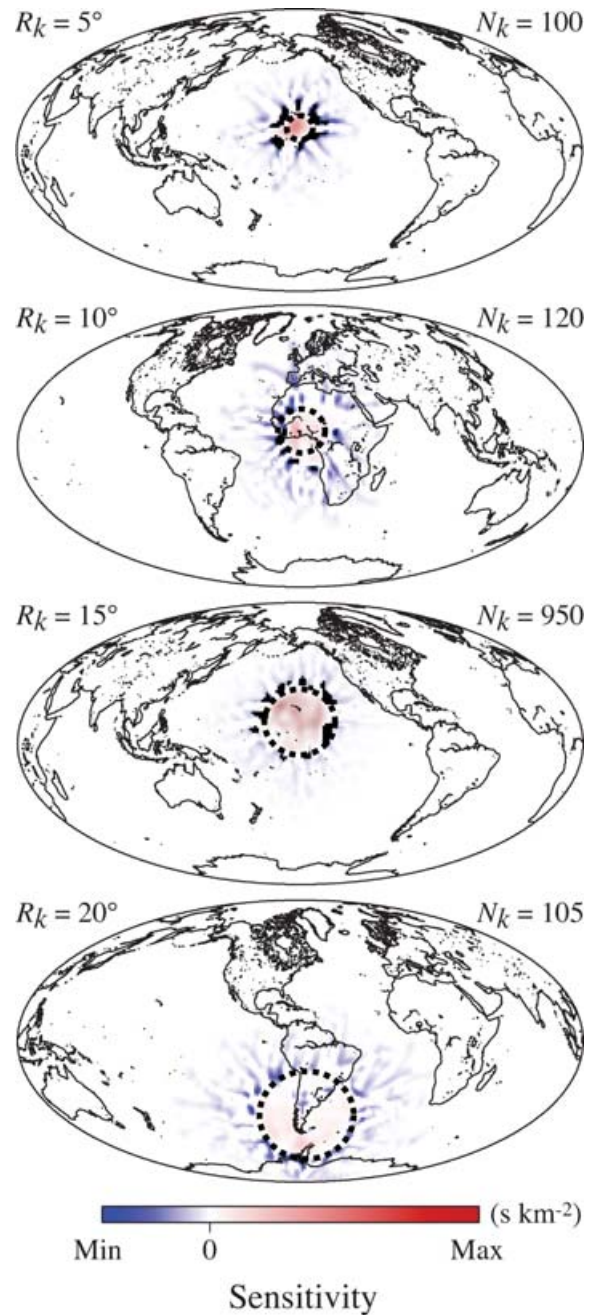

Figure 7. The stacked sensitivity kernels for four bins with (a) $r_k = 5°$ beneath the Pacific, (b) $r_k = 10°$ beneath Africa, (c) $r_k = 15°$ beneath the Pacific and (d) $r_k = 20°$ beneath South America. The stacked kernels in (a) and (c) have the same bin focus but different bin radii.

weighted by the standard errors of the stacked SdS-SS traveltime residuals.

5.1 Roughness-Misfit trade-off

As with many geophysical inversions, the results are highly dependent upon the prescribed regularization constraints. We perform 11 inversions with various smoothness weights. The resulting 'L-curve' (Fig. 8) illustrates the trade-off between misfit (1 – variance reduction) and the average roughness per topographic section, $\bar{\xi}$. The roughness associated with a single cell is defined here as the
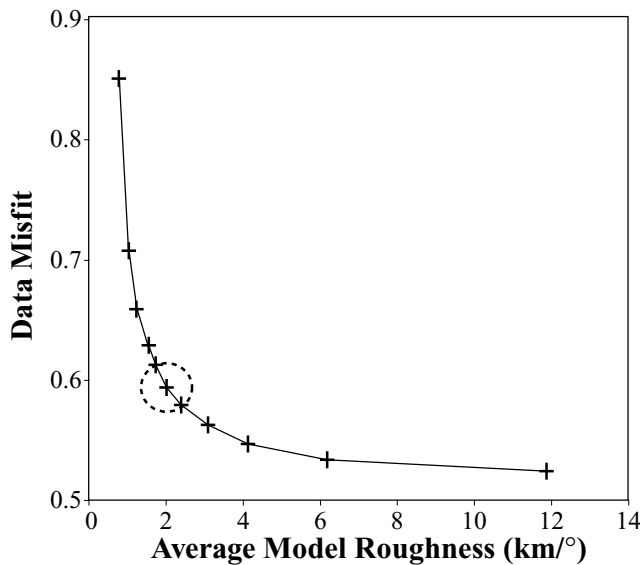
Figure 8. The '*L*-curve' shows the trade-off between data misfit (1 – variance reduction) and average model roughness for various inversions with different regularization constraints. An average model roughness of ~2 km deg^{-1} provides a balance between low data misfit without yielding an unnecessarily rough model.

deviation between its topography and the average topography of the eight adjacent cells (i.e. proportional to the estimated Laplacian), $\zeta_j = 1/8 \sum_{k=1}^{8} (\Delta z_k - \Delta z_j)/\Delta x$, where Δx is the side length of the cell. The average roughness is simply $\bar{\zeta} = 1/N \sum_{J=1}^{N} \zeta_j$. Smoother models ($\bar{\zeta} < 1.5$ km deg^{-1}) have difficulty describing the stacked traveltimes, resulting in a large rise in data misfit for additional smoothness. Models with high roughness ($\bar{\zeta} > 3$ km deg^{-1}) yield only minor improvements to the data misfit even for large increases in roughness. Based on this *L*-curve we choose to invert for transition zone thickness with a moderate roughness ($\bar{\zeta} = 2.0$ km deg^{-1}).

5.2 Resolution test 1

Before presenting the results using real data we demonstrate the usefulness of the inversion technique by conducting resolution tests with synthetic data. First, we create an arbitrary model of transition zone thickness that resembles a checkerboard pattern with cells of ~10° × 10°. We then compute one theoretical *SdS-SS* traveltime residual for each real trace with equation 1. We randomly swap out 25 per cent of these synthetic traveltimes with traveltimes associated with other ray paths to simulate low signal to noise. Synthetic stacked sensitivity kernels and traveltimes are calculated with eqs (3) and (5), where the inner summation of 25 per cent of the traces is randomly swapped out. The input and output transition zone thickness models are shown in Fig. 9. The pattern of the input structure is recovered in the output model even though significant amounts of noise were added. As expected, the amplitudes of transition zone thickness are damped in the southern hemisphere, where the data coverage is lower (particularly beneath South America and the Southern Atlantic Ocean).

We repeat the checkerboard test with noise added at $N = 4$ different approximate block sizes (5°, 10°, 15° and 20°) with both sine and cosine patterns (eight in total). We then calculate the normalized sum squared misfit between the input and output results and
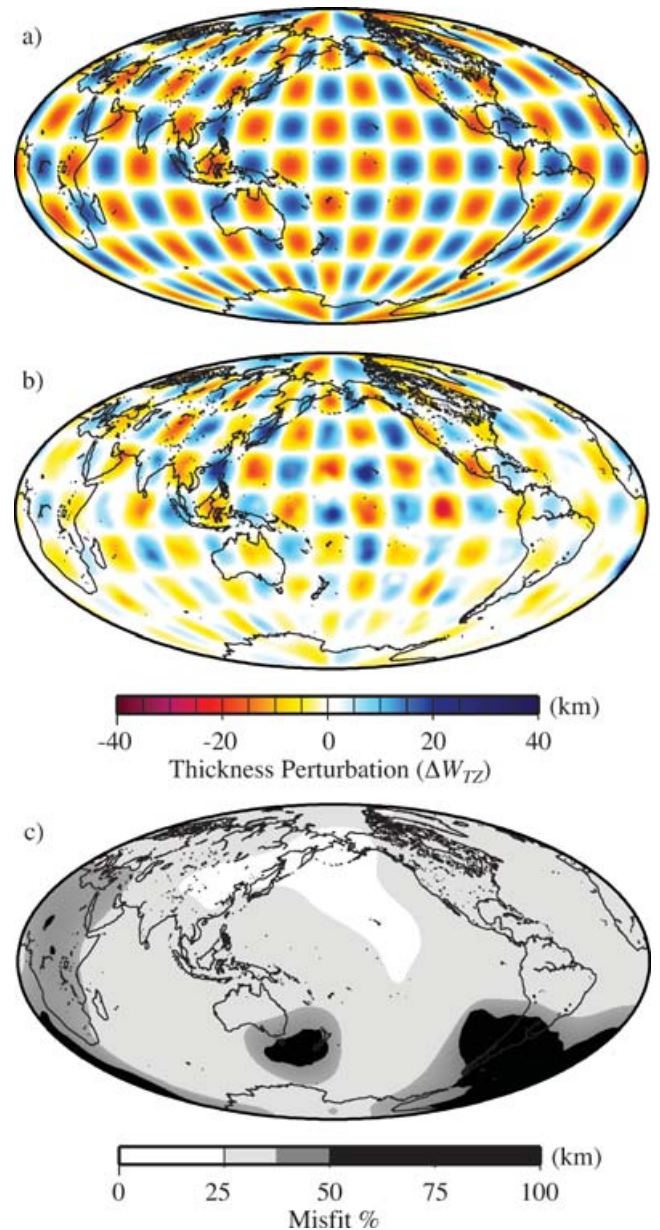

Figure 9. The resolution of the data and inversion method can be visually determined by comparing the (a) input and (b) output checkerboard pattern of transition zone thickness. 25 per cent of the individual synthetic traveltimes (eq. 1) are randomly swapped prior to calculating stacked traveltimes (eq. 3) and inverting to recover the structure. Red indicates anomalously thin and blue indicates anomalously thick. Panel (c) shows the normalized sum squared misfit between input and output checkerboard patterns for four different approximate block sizes (5°, 10°, 15° and 20°). White indicates little or no misfit, black is significant or total misfit.

divide by four $[1/N \sum_{i=1}^{N} (\sin + \cos) = 1]$, which roughly shows where the model is well resolved (Fig. 9c). The normalized misfit is highly correlated to the bouncepoint density and minimum bin radius (Fig. 6), which suggests that the resolution is highly dependent upon data coverage. In proceeding figures we black out regions with greater than 50 per cent normalized sum squared misfit where resolution is known to be poor, and contour the surrounding regions at ~33.3 per cent misfit where resolution is suspect.
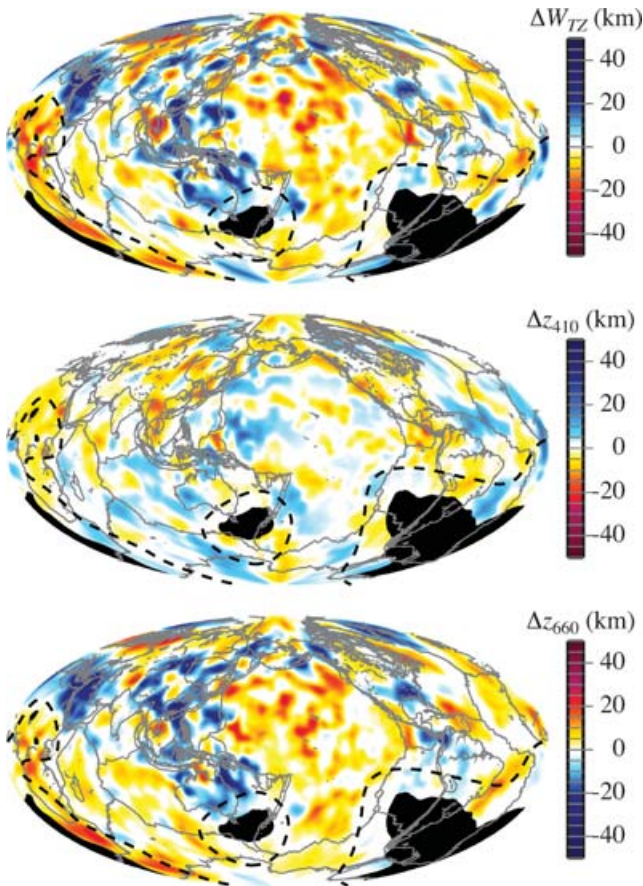
Figure 10. Fig. 10 illustrates the inverted (a) transition zone thickness perturbations (ΔW_{TZ}), (b) 410-km discontinuity topography (Δz_{410}) and (c) 660-km discontinuity topography (Δz_{660}). Red indicates anomalously thin transition zone, depressed 410 topography, and elevated 660 topography. Blue anomalies indicate thick transition zone, elevated 410 topography, and depressed 660 topography. The blackened regions have low resolution, the regions bound by the dashed lines have marginally acceptable resolution.

5.3 Inverted results

The inverted transition zone thickness map (Fig. 10) has similar, but shorter wavelength variations than the maps obtained from simple waveform binning (Figs 6 a–d). As described in Section 5.3, regions with insufficient resolution are blacked out. Many of the anomalies in the inverted ΔW_{TZ} map are smaller in amplitude and

scale, whereas others are larger in amplitude. For example, the transition zone beneath the southern portion of the East Pacific Rise appears thick when imaged using simple bouncepoint caps, but the inversion reveals that this region has relatively average transition zone thickness.

The thick transition zone anomalies beneath the western Pacific subduction zones form a nearly continuous laterally narrow curvilinear pattern in the inverted image. These are the thickest areas in the entire model, with multiple sections exceeding 277 km ($\Delta W_{TZ} > +35$ km). Another prominent thick region extends from the Red Sea to the Aegean, where the African plate subducted beneath Eurasia. The thinnest regions in the model include anomalies beneath and northwest of Hawaii and beneath Central Africa with thickness values less than 212 km ($\Delta W_{TZ} < -35$ km). The thin anomaly beneath Thailand is nearly completely surrounded by thick transition zone anomalies associated with subduction zones, providing the sharpest transition from thick to thin transition zone in the model.

While we do not concentrate on the details of the individual 410- and 660-km discontinuity topographies in this study, it is important to present them and provide a cursory description. The small-scale topography changes significantly (>5 km) in some regions depending on which 3-D seismic model is used to correct for 3-D heterogeneity. The long-wavelength pattern remains nearly identical regardless of this correction. The majority of the transition zone thinning and thickening occurs due to topography on the 660 (Fig. 10), which is consistent with previous studies (Flanagan & Shearer 1998; Gu *et al.* 2003). The angular power per harmonic degree (Fig. 11) clearly demonstrates that more topographic variation occurs on the 660 than the 410. The 660 deepens beneath subduction zone regions where the transition zone is thick, and becomes more shallow in regions where the transition zone thins (e.g. beneath Hawaii).

The average transition zone thickness is 242.3 km, which is consistent with previous studies (Gu *et al.* 1998, 2003; Flanagan & Shearer 1998; Gu & Dziewonski 2002; Lawrence & Shearer 2006a,b). The distribution of the transition zone thickness is roughly Gaussian with a standard deviation of $\sigma(W_{TZ}) = 8.6$ km. The standard deviation of the inverted thickness is less than the standard deviations of the long-wavelength transition zone thickness estimates derived from simple bouncepoint stacks of this study [$\sigma(W_{TZ\_1-D}) = 14.6 \pm 0.7$ km] and that of Flanagan & Shearer (1998) [$\sigma(W_{TZ1-D}) = 12.3$ km]. The lower standard deviation and greater maximum deviation of the inverted thickness model reflects its greater focusing of the largest anomalies into more compact regions and the tendency of the simple undamped stacking method to produce large-scale structure in poorly constrained regions with sparse data coverage.
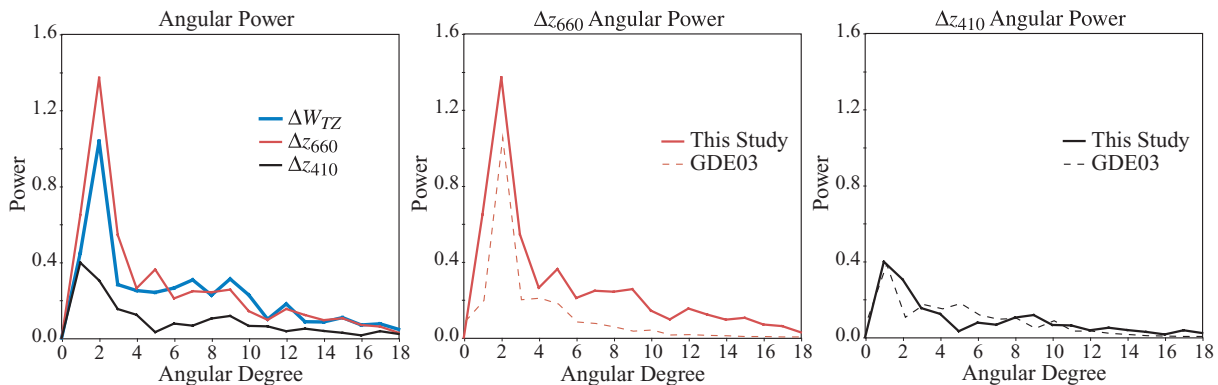

Figure 11. The angular power per spherical harmonic degree of (a) the inverted model, (b) the inverted Δz_{660} compared with Gu *et al.* (2003) (GDE03) and (c) the inverted Δz_{410} compared with GDE03. The angular powers of ΔW_{TZ} and Δz_{660} are much greater than for Δz_{410}.
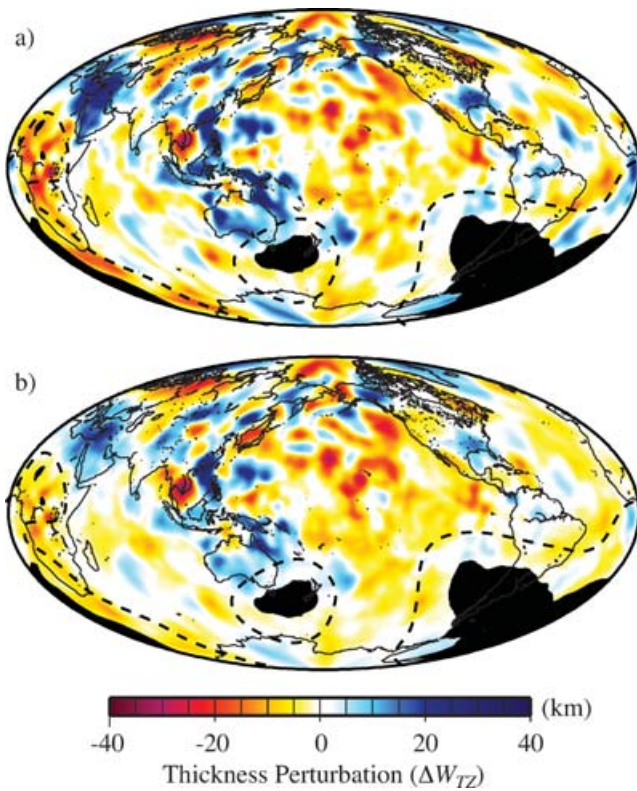
Figure 12. This resolution test is identical to that of Fig. 9 except that the (a) input transition zone thickness is the inverted model of transition zone thickness in Fig. 10, which is recovered well in (b). The blackened regions have low resolution, the regions bound by the dashed lines have marginally acceptable resolution.

5.4 Resolution test 2

While the checkerboard test illustrates the level of resolvable structure, it does not indicate how well specific structures are resolved. Here we use the observed transition zone structure as the input for a second resolution test. We calculate synthetic sensitivity kernels and traveltimes with eqs (3) and (5) and then invert these times (with 25 per cent of the times randomly switched) in an attempt to regain the input structure. The input and output structures are presented in Fig. 12. The output of the resolution test is highly similar to the input, suggesting that the observed structure is well resolved.

6 DISCUSSION

Neele & de Regt (1999) showed that small-scale (<4°), large amplitude topography (>50 km) on the 410 and 660 can yield large artefacts in apparent structure of the transition zone if proper depth-time sensitivities are not taken into account with underside reflections. This is of particular concern for subduction zone regions, where a cold downwelling slab might provide large-amplitude short-wavelength anomalies. Previous *SdS* studies avoided this concern by (1) using longer-period data (> ~30 s) than modelled by Neele & de Regt (1999), (2) noting that the model predicted that anti-symmetric variations occur for the *S410S* and *S660S* whereas the observed subduction zone anomalies are dominated by variations in the *S660S* and (3) only examining very long-wavelength topography (*l* <12) (Flanagan and Shearer 1998; Gu *et al.* 1998; Gu and Dziewonski 2002; Gu *et al.* 2003). Furthermore, Shearer *et al.* (1999) showed that the apparent topography calculated from large-bin radii *SdS-SS* stacks of long-period Kirchoff synthetic waveforms remain gen-

erally insensitive to short-wavelength topography (Shearer *et al.* 1999).

While Neele & de Regt (1999) were on the right track for developing a method that could resolve finer structure in the transition zone through inversion, their method did not account for the need to stack low-amplitude transition zone phases. Consequently they only performed resolution tests given individual noiseless synthetic waveforms of *PdP*. In practice *P660P* has such small amplitude that it is hard to observe even with thousands of stacked waveforms, let alone from a single waveform (e.g. Estabrook & Kind 1995; Shearer & Flanagan 1999; Lawrence & Shearer 2006b).

In this study, we use appropriate sensitivities and consistent stacking and inversion methods with real *SS* precursor data. The apparent structures from simple stacking (Fig. 6) differ from the inverted structures (Fig. 10) and from each other when different bin radii are used. These differences are largest for short-wavelength features and relatively small for long-wavelength features. This suggests that the stacking method of Flanagan & Shearer (1998) and Gu *et al.* (1998) was a reasonably accurate approach for determining the long-wavelength structure of the transition zone. To test the applicability of using the *SdS* in place of the *SdS-SS* kernel (as recommended by Dahlen 2005), we apply the same inversion technique with *SdS* kernels having marginally different frequency content to mimic small differences that might arise due to including the negligible sensitivities to the transmitted energy near the piercing points of the *SS* ray path through the discontinuities. As expected, the differences were negligible, so we do not present the results here.

While this study represents a significant improvement in resolution of global transition zone thickness, further work may yield even higher resolution images. Finite-frequency Fréchet kernels change as a function of the frequency content of the waveforms used to obtain the measurement. A similar study with multiple data sets having different frequency content might yield even higher resolution. A regional study beneath Hawaii by Schmerr & Garnero (2006) indicates that the apparent topography/thickness variations change as a function of frequency content. Such an inversion would require much larger computations and perhaps more sophisticated stacking algorithms. Stacking at higher frequencies becomes more difficult because high-frequency waveforms are generally less coherent. The assumption that anomalous topographies result in a single wave (and therefore a single stacked traveltime residual) may not be appropriate for higher frequencies. Further problems may result from correcting the traveltimes with a single seismic velocity model for all frequency bands.

Dahlen (2005) developed finite-frequency sensitivity kernels for other phases, which yield different coverage of the transition zone. By inverting multiple data sets simultaneously with the method described here it may be possible to improve global resolution. This would be particularly important for regions such as South America where there is little *SdS* data coverage. However, the other data types are less evenly distributed than *SdS*, so such a study would have to be extremely careful with regularization of the inversion. Additionally, phases such as *Pds* are sensitive to much smaller-scale seismic velocity heterogeneity, so the dependence on assumed velocity structure may be much more important for a combined analysis.

One potential problem associated with this technique is that it treats each interface as a sharp discontinuity. It is likely that the 410 is actually a gradient with some finite thickness and the 660 is underlain by a relatively steep gradient and in some places an additional interface that can contribute to the *S660S* waveform (Shearer 1996; Cammarano *et al.* 2005; Lawrence & Shearer 2006; Wang *et al.* 2006). The slope of the sub-660 shear velocity gradient is a
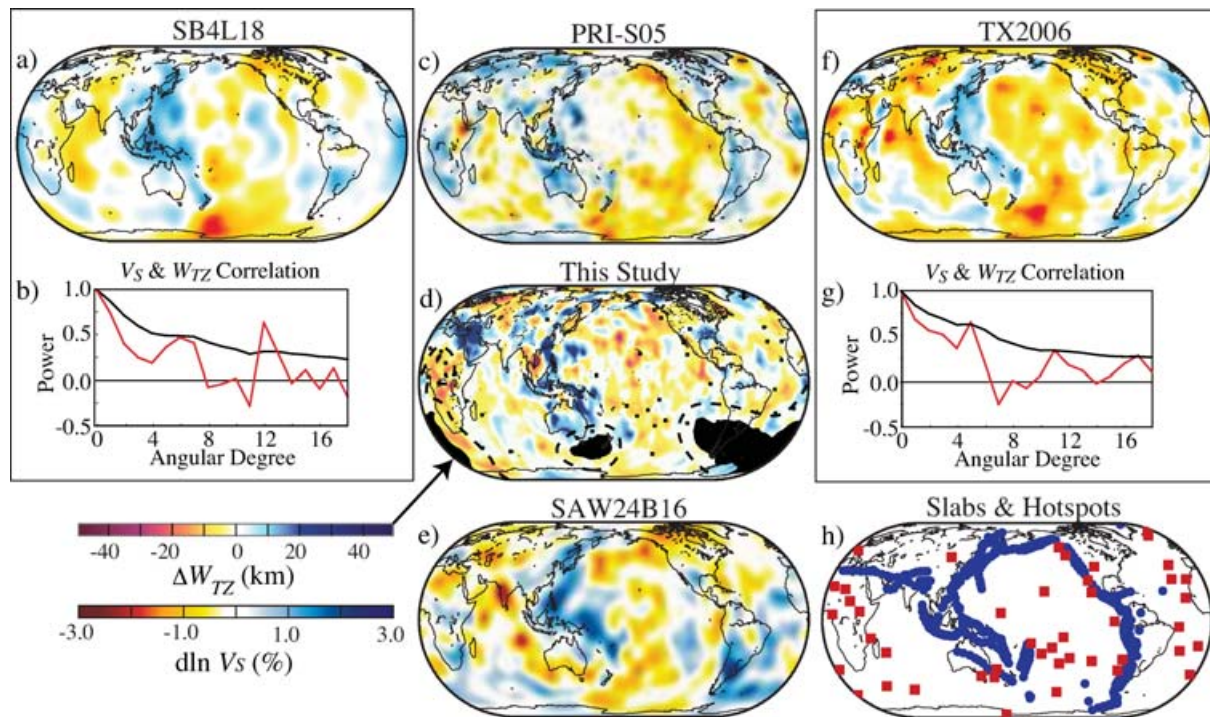
Figure 13. This figure graphically compares the lateral variation depth-averaged transition zone seismic velocity for models (a) SB4L18 (Masters *et al.* 2000), (c) PRI-S05 (Montelli *et al.* 2006), (e) SAW24B16 (Megnin & Romanowicz 2000) and (f) TX2006 (updated from Simmons *et al.* 2006) with (d) the inverted transition zone thickness of this study. The spherical harmonic cross-model correlations between the inverted ΔW_{TZ} and (b) SB4L18 and (g) TX2006 indicate positive correlation between high velocity and thickened transition zone for low angular degrees (red) and cumulatively (black) up to degree 18. The thickened and thinned transition zone regions correspond, respectively to (h) the (blue) potentially dense regions modelled through subduction history (Lithgow-Bertelloni & Richards 1998) and (red) hotspots (modified from Steinberger & O'Connell 1998).

temperature and chemically dependent result of a gradual transition from majorite to garnet (e.g. Ito & Takahashi 1989; Wang *et al.* 2006). A sharper sub-660 gradient or interface with less contrast across the 660 has the tendency to make the 660 appear deeper. The method presented here is only valid for a single interface causing each phase. Nevertheless, it is generally accepted that the majority of impedance contrast for the *S410S* and *S660S* are accrue at depths associated with the respective olivine phase transformations.

6.1 Comparison with seismic velocity

The global seismology community has produced ever-improving tomographic images of seismic wave speed for the whole mantle (e.g. Masters *et al.* 2000; Megnin & Romanowicz 2000; Simmons *et al.* 2006; Montelli *et al.* 2006; Houser *et al.* 2008). The seismic wave speed of transition zone material is dependent upon the same thermochemical variations that effect transition zone thickness. Fig. 13 compares the lateral variation in depth-averaged transition zone shear velocity for several global models with our inverted transition zone thickness model. While these seismic velocity models were calculated with different data and inversion techniques, the results are mostly similar. Regions of thick transition zone generally overlap regions with seismically fast material in the transition zone.

Fig. 13 also shows spherical harmonic cross-model correlations (e.g. Becker & Boschi 2002) between the inverted transition zone thickness model and the seismic shear velocity models of Masters *et al.* (2000) and Simmons *et al.* (2006) at 600 km depth. While the correlations for individual harmonic degrees may differ, the cumulative correlation out to degree 18 is reasonably good (>3.5).

6.2 Interpretation as temperature

The scale and amplitude of transition zone thickness anomalies are likely coupled with the thermo-chemical state of the transition zone. The 410- and 660-km discontinuities are commonly idealized as pressure and temperature dependant phase changes that are defined by Clapeyron slopes of opposing sign. While quench experiments (Ito & Takahashi 1989) and calorimetric studies (Akaogi & Ito 1993) estimate that the γ-spinel to perovskite and (Mg,Fe)O ferropericlase Clapeyron slope is -3 MPa $°K^{-1}$, recent *in situ* X-ray diffraction studies suggest smaller values between -0.4 and -2.0 MPa $°K^{-1}$ (Katsura *et al.* 2003; Fei *et al.* 2004; Litasov *et al.* 2005). Recent *in situ* X-ray diffraction studies indicate that the $\alpha-\beta$ transition has a positive Clapeyron slope of 3.6 to 4.0 MPa $°K^{-1}$ (Katsura *et al.* 2004; Morishima *et al.* 1994) while quenching experiments indicate values of $\sim$2.5 MPa $°K^{-1}$ (Katsura & Ito 1989).

The thickest regions of the transition zone ($W_{TZ} > 272$ km) would require the transition zone to be 180–310 $°K$ cooler than average for the Clapeyron slopes mentioned above. This inferred low temperature is consistent with the hypothesis that cold subducted oceanic lithosphere has been transported to the transition zone (e.g. Shearer 1993). Conversely, the thinnest regions of transition zone ($W_{TZ} < 212$ km) necessitate temperatures between $180°$ and $300 °K$ warmer than average. Any amount of temperature anomaly should produce 120 per cent more topography on the 660 than the 410 given the quenching experiment Clapeyron slopes and assuming equal temperature anomalies at the two depths. Alternatively, the *in situ* X-ray diffraction studies Clapeyron slopes would predict 2–10 times more topography on the 410 than on the 660. While the topographies
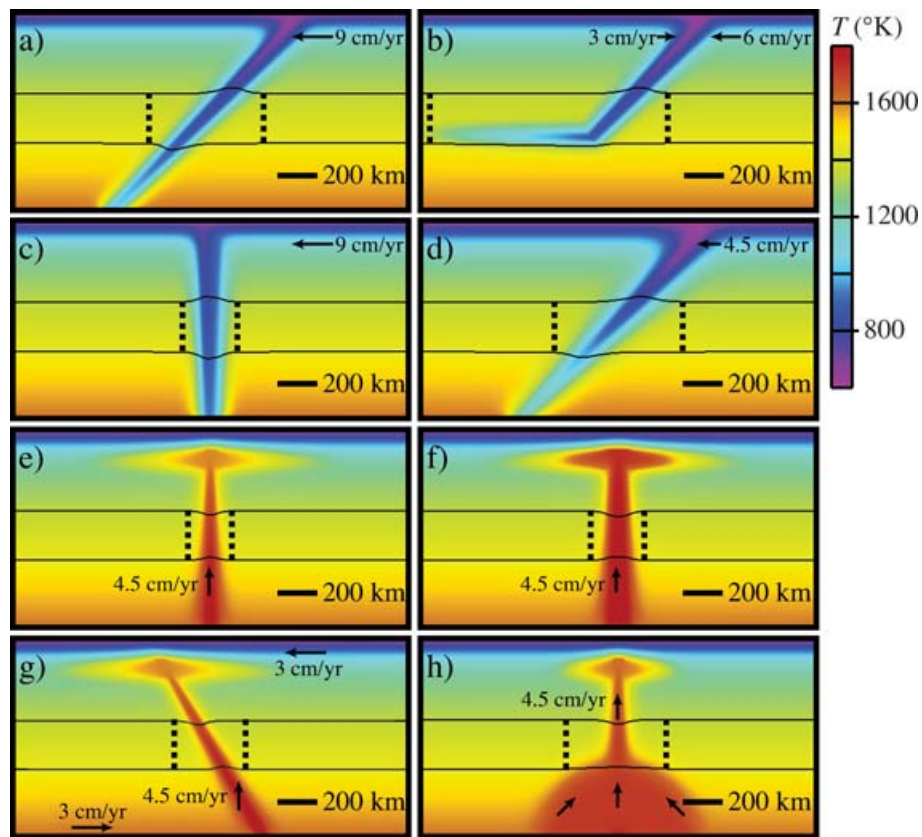
Figure 14. 2-D (2000 × 1000 km^2) schematic thermal conduction models with a thermal diffusivity of $\kappa = 10^{-6}$ m^{2}s^{-1}. For subduction models we prescribe slab motions at 9 cm yr^{-1} for (a) a slab that penetrates the 660, (b) a slab that deflects at the 660, (c) a vertical slab, and (d) a slower (4.5 cm yr^{-1}) slab. For plume models we prescribe upwelling at 4.5 cm yr^{-1} for (e) a 100-km diameter plume tail, (f) a 200-km diameter plume tail, (g) an advecting plume, and (h) a plume rising from a lower-mantle thermochemical pile. The white lines indicate α–β olivine and γ-olivine to perovskite and magnesiowüstite phase changes. The vertical dashed lines indicate the lateral widths of topographic features associated with each thermal anomaly. The blackened regions have low resolution, the regions bound by the dashed lines have marginally acceptable resolution.

of the 410 and 660 are not uniquely constrained because of trade-offs with assumed upper mantle velocity structures, the inverted topography on the 660 is much greater than on the 410. If the inverted transition zone structure observed here is accurate, then the temperature anomaly associated with subduction zones is approximately two times that of previous models.

One possible failure in the interpretation of transition zone thickness variations as a result of a single temperature variation is that both boundaries need not have equal or even correlated temperature anomalies. It is possible that the 660 could be heated from below while the 410 is being cooled from above, or vise versa. A large difference between temperature anomalies across the 410 and 660 discontinuities could result from variations in flow patterns due to changes in viscosity between upper and lower mantle (e.g. Forte & Mitrovica 1996). Another likely problem associated with the temperature interpretation is that the interface depths are also chemically dependant (Wang *et al.* 2006).

6.3 Slabs in the transition zone

The regions of thickened transition zone correspond to the approximate locations of subducted slabs inferred from the history of subduction and plate motion (Fig. 13h) (Lithgow-Bertelloni & Richards 1998; Steinberger 2000). If the thickened regions of transition zone near subduction zones are caused by cool temperatures within descending oceanic lithosphere, then the inverted model implies cooler

slabs than previous models, and that less cool slab material is currently in the transition zone. If the variations in transition zone thickness are interpreted as thermal, then these anomalies are 100–300 °K cooler than the average mantle. The smaller lateral scale of these curvilinear anomalies equates to a smaller volume of cold slab material within the transition zone.

In order to interpret our inverted transition zone thickness model, we compute several kinematic finite-difference conductive heating models with prescribed mass transport. These models solve the 2-D diffusion equation for conductive heating/cooling at time steps of 10^4 yr with a thermal diffusivity of $\kappa = 10^{-6}$ m^{2}s^{-1} (e.g. Creager & Boyd 1991). The beginning temperature increases linearly from 273 °K at the surface to 1000 °K at 150 km depth. From 150 to 1000 km the temperature increases linearly to 1800 °K. The surface is always 273 °K. The top and bottom boundary conditions maintain 273 °K at the surface and 1800 °K at 1000 km depth. The side boundary temperature conditions assume zero temperature gradient/conductive heat flow. Material can be transported across the bottom and side boundaries. Laterally entering material has average temperature for each depth. Material may enter across the bottom boundary at 1800 °K. The first model (Fig. 14a) assumes a 9 cm yr^{-1} rate of subduction for a 100 km thick lithosphere that descends with a dip of 45°. The models represent a 2000 km wide by 1000 km deep region gridded into 2 × 2 km blocks. We then calculate the theoretical transition zone thickness given Clapeyron slopes of 3 and –3 MPa °K^{-1} for the 410 and 660, respectively. These models are intended
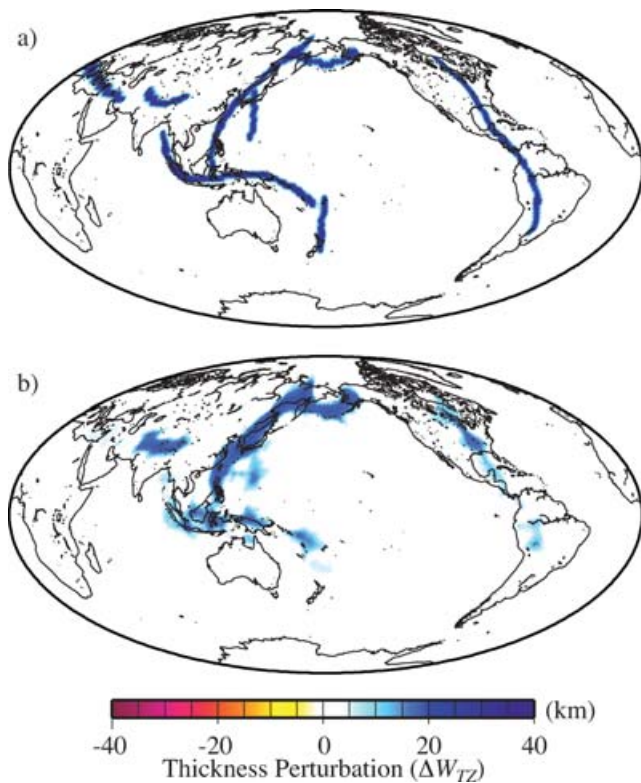
Figure 15. This resolution test is identical to that of Fig. 9 except that the (a) input transition zone thickness is defined by an unperturbed 410 and a 660 that is depressed by ~50 km in 200 km wide curvilinear regions located in regions where we expect subducted slabs. The (b) recovered model is laterally smoothed compared to the input model with lower amplitudes and greater width.

to be purely schematic, not representative of true convection, so the interpretation will be limited to qualitative differences.

After 20 Myr of mass transport and thermal conduction the thermal signature of a slab that descends into the lower mantle would cause a much narrower and higher amplitude ΔW_{TZ} anomaly than an equivalent slab that remains in the transition zone (Fig. 14). More steeply descending slabs cause even narrower and higher amplitude thermal and ΔW_{TZ} anomalies (Fig. 14c). Conversely, shallow dipping slabs could lose their thermal signature prior to reaching the transition zone, which would yield little to no transition zone thickening. Lower amplitude and broader ΔW_{TZ} variations result from using either higher thermal conductivity or slower subduction rates (Fig. 14d).

Stagnant slabs within the transition zone create broad anomalies like those observed by previous *SdS* stacking studies. Some thickened anomalies along the western Pacific margin are laterally narrow in our inverted model, which may indicate that slabs penetrate through the 660. The thickness and breadth of these anomalies vary, which may represent either variable resolution or different convection regimes. On the other hand, a nearly vertical slab may be at or smaller than the resolution of the inverted model. Fig. 15 demonstrates the potential to resolve a series of 200 km wide, 50 km deep curvilinear troughs depressed into the 660. Clearly, small-scale (<200 km wide) anomalies are laterally smeared by a factor of ~2 and vertically damped by a factor of ~0.5. Consequently, variable slab dip angles may cause the anomaly to drop below or rise above the resolution.

Within quickly descending cold slabs it may be possible for a thin meta-stable olivine wedge to exist at depths greater than ~410 km (Rubie & Ross 1994). While the phase change from α- to-β olivine is only kinematically sluggish within the coldest core of the slab, it may be possible for reflections from a meta-stable wedge to cause some *S410S* energy to arrive earlier. Consequently, it is possible that a small-scale metastable wedge could effectively mask an upwarping of the 410 due to cold temperatures. This may explain why the inverted undulations are far greater on the 660 than the 410 in subduction zone regions. Alternatively, downwelling slabs may contain stable hydrous silicate minerals such as phases B and D (Shieh *et al.* 1998) that may become entrained in the hydrophilic transition zone minerals (Bercovici & Karato 2003; Karato 2006). It may be possible for this H_2O to concentrate near the 410 due to buoyancy differences between aquios and dry densities, creating a low-velocity layer (e.g. Song & Helmberger 2004). It may be possible for reflective phases from a water-induced low-velocity layer could interfere with the *S410S* and the resultant estimated transition zone topography.

Subduction along the Eastern edge of the Pacific tends to occur at much shallower angles than along the Western margin. Consequently, these slabs may lose a significant amount of their thermal signature prior to reaching the transition zone. This would explain the general lack of thick transition zone beneath the Americas relative to Eastern Asia. However, the resolution beneath South America is particularly poor for this study, so thickened regions may exist there as suggested by receiver function studies (e.g. Chevrot *et al.* 1999; Lawrence & Shearer 2006a).

6.4 Hotspots

If plumes originating within or beneath the transition zone were the source of hotspots, then we would expect narrow (~100 km) anomalously thin regions beneath hotspots. However, plumes could migrate laterally while ascending so there is some uncertainty in knowing where to look for thinned transition zone relative to the hotspots. For this reason we compare the difference in average and minimum transition zone thickness beneath hotspot locations (Fig. 13h) (Steinberger & O'Connell 2000) with equivalent values beneath randomly sampled locations. By repeating the random resampling 25 times we obtain a more reliable result and a measure of uncertainty in the result. The average transition zone thickness within 2° radii of the hotspots (excluding regions with insufficient resolution) is 3.9 ± 1.1 km thinner than the randomly sampled locations. The minimum transition zone thickness within 2° radii of the hotspots (excluding regions with insufficient resolution) is 5.2 ± 1.7 km thinner than the randomly sampled locations. A temperature increase between 25° and 40°K would result in ~4 km thinning.

The likelihood of imaging a plume tail rising through the transition zone depends on the radius of the upwelling portion. Imagining small plume tails with upwelling radii < 50 km is unlikely, given the resolution of the inverted model. In Fig. 14, we present several finite-difference conduction models that depict thermal anomalies for prescribed upwelling. As with the forced slab conduction models described above, we prescribe the mass transfer and allow conduction to heat/cool the material. In each model we assume the thermally buoyant plume material rises at 4 cm yr^{-1}. Again, these models are intended to be purely schematic, so only qualitative differences between the models will be discussed here.

Calculations with larger radii defining the upwelling regions result in laterally broader and vertically thinner transition zone anomalies (Figs 14e and f). Plumes with small radii (<100 km) may have

topographic signatures that are smaller than the resolution of this model. A plume that is laterally displaced by mantle currents could result in lateral broadening of the thinned anomaly (Fig. 14g). Alternatively, laterally broad and vertically thin anomalies could result from large-scale convection forming large thermochemical piles (McNamara & Zhong 2005). If such large-scale structures were to extend up to the 660, they would cause very broad regions of thinned transition zone. This might explain the extensive regions of thinned transition zone beneath the Pacific and Africa, where large low-velocity structures have been seismically imaged extending from the core-mantle boundary to the upper mantle and the top of the lower mantle, respectively (e.g. Berger & Romanowicz 1998; Ritsema *et al.* 1999; Masters *et al.* 2000; Ni *et al.* 2002).

7 CONCLUSION

In this paper we develop and apply a new adaptive stacking technique in tandem with finite-frequency theory to invert stacked *SdS-SS* traveltimes and stacked sensitivity kernels for transition zone topography and thickness. This technique yields a high-resolution image of the transition zone that isolates higher amplitude and smaller-scale anomalies than have been previously imaged. The curvilinear regions of thickened transition zone generally correlate with seismically fast regions in tomographic models, which is consistent with cold temperatures in and near downwelling slabs in the mantle. The inverted topography on the 410 is much smaller amplitude than on the 660, suggesting that either the Clapeyron slope of the α olivine to β-spinel phase transition is shallow or that reflections from a deeper meta-stable wedge may counteract the elevated topography within slabs at ∼410 km depth. Comparative analysis between multiple kinematic finite-difference conductive heating models of subduction suggests that the short wavelengths and the large amplitudes of the thickened anomalies are best explained by slabs that penetrate through the 660. On average, hotspots are located above or near discontinuous regions of thinned transition zone. Further kinematic thermal modelling suggests that the large amplitudes of the thinned anomalies likely result from warm upwellings that originate from higher temperature sources much deeper in the mantle.

REFERENCES

Akaogi, M. & Ito, E., 1993. Heat capacity of $MgSiO_3$ perovskite, *Geophys. Res. Lett.,* **20,** 105–108.

Becker, T.W. & Boschi, L., 2002. A comparison of tomographic and geodynamic mantle models, *G-cubed,* **3,** dio:10.1029/2001GC000168.

Bercovici, E. & Karato, S.-I., 2003. Whole-mantle convection and the transition–zone water filter, *Nature,* **425,** 39–44.

Bréger, L. & Romanowicz, B., 1998. Thermal and chemical 3-D heterogeneity in D", *Science,* **282,** 718–720.

Boschi, L., Becker, T.W., Soldati, G. & Dziewonski, A.M., 2006. On the relevance of Born theory in global seismic tomography, *Geophys. Res. Lett.,* **33,** doi:10.1029/2005GL025063.

Bostock, M.G., 1996. Ps conversions from the upper mantle transition zone beneath the Canadian landmass, *J. geophys. Res.,* **101,** 8393–8402.

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

Chevrot, S., Vinnik, L.P. & Montagner, J.-P., 1999. Global-scale analysis of the mantle Pds phases, *J. geophys. Res,* **104,** 20203–20219.

Creager, K.C. & Boyd, T.M., 1991. The geometry of Aleutian subduction: three-dimensional kinematic flow modeling, *J. geophys. Res.,* **96,** 2293–2307.

Dahlen, F.A., 2005. Finite-frequency sensitivity kernels for boundary topography perturbations, *Geophys. J. Int.,* **162,** 525–540.

Dahlen, F.A. & Nolet, G., 2005. Comment on 'On sensitivity kernels for 'wave-equation' transmission tomography' by de Hoop and van der Hilst, *Geophys. J. Int.,* **163,** 949–951.

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

Deuss, A.J. & Woodhouse, J.H., 2001. Seismic observations of splitting of the mid-transition zone discontinuity in Earth's mantle, *Science,* **294,** 354–357.

Deuss, A.J. & Woodhouse, J.H., 2002. A systematic search for mantle discontinuities using SS-precursors, *Geophys. Res. Lett.,* **20,** dio:10.1029/2002GL014768.

Dueker, K.G. & Sheehan, A.F., 1997. Mantle discontinuity structure from midpoint stacks of converted P to S waves across the Yellowstone hot spot track, *J. Geophys. Res.,* **102,** 8313–8327.

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

Dziewonski, A.M. & Gilbert, F., 1976. The effect of small, aspherical perturbations on traveltimes and a re-examination of the corrections for ellipticity, *Geophys. J. R. Astron. Soc.,* **44,** 7–16.

Estabrook, C.H. & Kind, R., 1996. The nature of the 660-km upper-mantle discontinuity from precursors to the PP phase, *Science,* **274,** 1179–1182.

Fei, Y. *et al.*, 2004. *J. gesophys. Res.,* **109,** doi:10.1029/2003JB002562.

Flanagan, M.P. & Shearer, P.M., 1998. Global mapping of topography on transition zone velocity discontinuities by stacking *SS* precursors, *J. geophys. Res.,* **103,** 2673–2692.

Forte M. & Mitrovica, J.X., 1996. New inferences of mantle viscosity from joint inversion of long-wavelength mantle convection and post-glacial rebound data, *Geophys. Res. Lett.,* **23,** 1147–1150.

Gilbert, H.J., Sheehan, A.F., Dueker, K.G. & Molnar, P., 2003. Receiver functions in the western United States, with implications for upper mantle structure and dynamics, *J. geophys Res.,* **108,** doi:10.1029/2001JB001194.

Gossler, J. & Kind, R., 1996. Seismic evidence fo very deep roots of continents, *Earth Planet. Sci. Lett.,* **138,** 1–13.

Gu, Y.J. & Dziewonski, A.M., 2002. Global variability of transition zone thickness, *J. geophys. Res.,* **107,** dio:10.1029/2001JB000489.

Gu, Y.J., Dziewonski, A.M. & Agee, C.B., 1998. Global de-correlation of the topography of transition zone discontinuities, *Earth Planet. Sci. Lett.,* **157,** 57–67.

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

Gurrola, H. & Minster, J.B., 1998. Thickness estimates of the upper-mantle transition zone from bootstrapped velocity spectrum stacks of receiver functions, *Geophys. J. Int.,* **133,** 31–43.

Houser, C., Masters, G., Shearer, P. & Laske, G., 2008. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms, *Geophys. J. Int.,* doi:10.1111/j.1365-246X.2008.03763.x.

Jackson, I., 1983. Some geophysical constraints on the chemical composition of the Earth's lower mantle, *Earth Planet. Sci. Lett.,* **62,** 143–164.

Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.,* **105,** 113–130.

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

Karato, S., 2006. Remote sensing of hydrogen in Earth's mantle, in *Water in Nominally Anhydrous Minerals*, *Rev. Min. & Geochem.*, Vol. 62, pp. 343–375.

Katsura, T. & Ito, E., 1989. The system Mg_2SiO_4-Fe_2SiO_4 at high-pressures and temperatures—precise determination of stabilities of olivine, modified spinel, and spinel, *J. geophys. Res.,* **94** 15 663–15 670.

Katsura, T. *et al.*, 2003. Post-spinel transition in Mg_2SiO_4 determined by high P-T in situ X-ray diffractometry, *Phys. Earth Planet. Inter.,* **136,** 11–24.

Katsura, Yamada, H., Nishikawa, O., Song, M., Kubo, A., Ito, E. & Funakoshi, K., 2004. Olivine-wadsleyite transition in the system $(Mg,Fe)_2SiO_4$, *J. Geophys. Lett.,* **109,** B02209, doi:10.1029/2003JB002438.

Ito, E. & Takahashi, E., 1989. Post-spinel transformations in the system Mg_2SiO_4-Fe_2SiO_4 and some geophysical implications, *J. geophys. Res.,* **94,** 10 637–10 646.

Lawrence, J.F. & Shearer, P.M., 2006a. A global study of transition zone: thickness using receiver functions, *J. geophys. Res.,* **111,** dio:10.1029/2005JB003973.

Lawrence, J.F. & Shearer, P.M., 2006b. Constraining seismic velocity and density for the mantle transition zone with reflected and transmitted waveforms, *Geochem. Geophys. Geosyst.,* **7,** Q10012, doi:10.1029/2006GC001339.

Lee, D.-K. & Grand, S., 1996. Depth of upper mantle discontinuities beneath the East Pacific Rise, *Geophys. Res. Lett.,* **23,** 3369–3372.

Li, A., Fischer, K.M., Wysession, M.E. & Clarke, T.J., 1998. Mantle discontinuities and temperature under the North American continental keel, *Nature,* **395,** 160–163.

Li, X., Kind, R., Yuan, X., Sobolev, S.V., Hanka, W., Ramesh,D.S., Gu, J. & Dziewonski, A.M., 2003. Seismic observation of narrow plumes in the oceanic upper mantle, *Geophys. Res. Lett.,* **30,** doi:10.1029/2002GL015411.

Litasov, K.D., Ohtani, E., Sano, A., Suzuki, A. & Funakoshi, K., 2005. Wet subduction versus cold subduction, *Geophys. Res. Lett.,* **32,** doi:10.1029/2005GL022921.

Lithgow-Bertelloni, C. & Richards, M., 1998. The dynamics of Cenozoic and Mesozoic plate motions, *Rev. Geophys.,* **36,** 27–78.

Masters G., Laske, G., Bolton, H. & Dziewonski, A., 2000. The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, in *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale, AGU Geophysical Monograph 117,* AGU, Washington, DC.

Mégnin, C. & Romanowicz, B., 2000. The shear velocity structure of the mantle from the inversion of body, surface and higher modes waveforms, *Geophys. J. Int.,* **143**(3), 709–728.

McNamara, A.K. & Zhong, S., 2005. Thermochemical structures beneath Africa and the Pacific Ocean, *Nature,* **437,** 1136–1139.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., Engdahl, E.R. & Hung, S.-H., 2004a. Finite frequency tomography reveals a variety of plumes in the mantle, *Science,* **303,** 338–343.

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

Montelli, R., Nolet, G. & Dahlen, F. A., 2006. Comment on 'Banana-doughnut kernels and mantle tomography' by van der Hilst and de Hoop, *Geophys. J. Int.,* **167,** 1204–1210.

Morishima H,, Kato, T., Suto,M., Ohtani, E., Urakawa, S., Utsumi, W., Shimomura, O. & Kikegawa, T., 1994. The phase boundary between a- and b-Mg_2SiO_4 determined by in situ X-ray observation, *Science,* **265,** 1202–3.

Neele, F. & de Regt, H., 1999. Imaging upper-mantle discontinuity topography using underside-reflection data, *Geophys. J. Int.,* **137,** 91–106.

Ni, S., Tan, E., Gurnis M. & Helmberger, D., 2002. Sharp sides to the African superplume, *Science,* **296,** 1850–1852.

Ni, S., Sidao, Cormier, V.F. & Helmberger, D.V., 2003. "A comparison of synthetic seismograms for 2D structures: semianalytical versus numerical, *Bull. seism. Soc. Am.,* **93,** 2752–2757.

Nolet, G., 1987. Seismic wave propagation and seismic tomography, in *Seismic Tomography,* pp. 1–23, G. Nolet, ed., Reidel, Dordrecht.

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

Petersen, N., Vinnik, L.P., Kosarev, G., Kind, R., Oreshin, S. & Stammler, K., 1993. Sharpness of the mantle discontinuities, *Geophys. Res. Lett.,* **20,** 859–862.

Ringwood, A.E., 1975. Composition and petrology of the Earth's mantle, McGraw-Hill, New York, pp. 618.

Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 1999, Global transition zone tomography, *Science,* **286,** 1925–1925.

Ritsema, J., van Heijst, H.J. & Woodhouse, J.H., 2004. Complex shear velocity structure imaged beneath Africa and Iceland, *J. geophys. Res.,* **109,** doi:10.1029/2003JB002610.

Rubie, D. C. & Ross II, C. R., 1994. Kinetics of the olivine-spinel transformation in subducting lithosphere: experimental constraints and implications for deep slab processes, *Phys. Earth Planet. Inter.,* **86,** 223–241.

Schmerr, N. & Garnero, E., 2006. Investigation of upper mantle discontinuity structure beneath the central Pacific using *SS* precursors, *J. geophys. Res.,* **111,** doi:10.1029/2005JB04197.

Shearer, P.M., 1991. Constrains on upper mantle discontinuities from observations of long-period reflected and converted phases, *J. geophys. Res.,* **96,** 18147–18182.

Shearer, P.M., 1993. Global mapping of upper mantle reflectors from long period SS precursors, *Geophys. J. Int.,* **115,** 878–904.

Shearer, P.M., 1996. Transition zone velocity gradients and the 520-km discontinuity, *J. geophys. Res.,* **101,** 3053–3066.

Shearer, P.M. & Masters, G., 1992. Global mapping of topography on the 660-km discontinuity, *Nature,* **355,** 791–796.

Shearer, P.M., Flanagan, M.P. & Hedlin, M.A.H., 1999. Experiments in migration processing of SS precursor data to image upper mantle discontinuity structure, *J. geophys. Res.,* **104,** 7229–7242.

Shen, Y., Solomon, S.C., Bjarnason, I.T. & Purdy, G.M., 1996. Hot mantle transition zone beneath Iceland and the adjacent Mid-Atlantic Ridge inferred from P-to-S conversions and the 410- and 660-km discontinuities, *Geophys. Res. Lett.,* **23,** 3527–2530.

Shen, Y., Sheehan, A.F., Dueker, K.G., de Groot-Hedlin, C. & Gilbert, F., 1998. Mantle discontinuity structure beneath the southern East Pacific Rise from P-to-S converted phases, *Science,* **280,** 1232–1235.

Simmons, N.A., Forte, A.M. & Grand, S. P., 2006. Constraining mantle from with seismic and geodynamic data: a joint approach, *Earth Planet. Sci. Lett.,* **246,** 109–124.

Stammler, K., Kind, R., Petersen, N., Vinnik, L.P. & Qiyuan, L., 1992. The upper mantle discontinuities correlated or anticorrelated?, *Geophys. Res. Lett.,* **19,** 1563–1566.

Steinberger, B.M., 2000. Slabs in the lower mantle-results of dynamic modelling compared with tomographic images and geoid, *J. geophys. Res.,* **78,** 3321–3333.

Steinberger, B. & O'Connell, R.J., 1998. Advection of plumes in mantle flow; implications on hotspot motion, mantle viscosity and plume distribution, *Geophys. J. Int.,* **132,** 412–434, doi:10.1046/j.1365–246x.00447.x, 1998.

Steinberger, B. & O'Connell, R. J., 2000. Effects of mantle flow on hotspot motion, in *The History and Dynamics of Global Plate Motions, Geophys. Mon. Ser., 121,* pp. 377–398, ed. Richards, M. A., Gordon, R. G. & van der Hilst, R. D., AGU, Washington, DC.

Van Der Hilst, R. D. & de Hoop, M. V., 2005. Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.,* **163,** 956–961, doi:10.1111/j.1365–246X.2005.02817.

Vinnik, L.P., 1977. Detection of waves converted from P to SV in the mantle, *Phys. Earth Planet. Int.,* **15,** 39–45.

Vinnik, L.P., Kosarev, G. & Petersen, N., 1996. Mantle transition zone beneath Eurasia, *Geophys. Res. Lett.,* **23,** 1485–1488.

Wang, Y., Wen, L., Weidner, D. & He, Y., 2006. SH velocity and compositional models near the 660-km discontinuity beneath South America and northeast Asia, *J. geophys. Res.,* **111,** B07305, doi:10.1029/2005JB003849.