



Constraining seismic velocity and density for the mantle transition zone with reflected and transmitted waveforms

Jesse F. Lawrence and Peter M. Shearer

IGPP, Scripps Institution of Oceanography, 9500 Gilman Drive, IGPP 0225, La Jolla, California, USA
(jlawrence@ucsd.edu)

[1] We examine stacks of several seismic phases having different sensitivities to mantle transition zone structure. When analyzed separately, underside P and S reflections (PdP and SdS) are suggestive of very different structures despite similar raypaths and data coverage. By stacking the radial component of PdP rather than the vertical PdP , we show that this difference does not result from interference from other more steeply inclined phases such as PKP and $Ppdp_{diff}$. In general, stacks of P -to- S converted phases (Pds) appear to lack evidence of a 520-km discontinuity when examined without other phases. When these phases and stacked topside P reflections ($Ppdp$) are analyzed jointly using a nonlinear inversion method, consistent but nonunique, seismological models emerge. These models show that a discontinuity at ~ 653 km depth has smaller contrasts in density and velocity than found in most previous studies. A sub-660 gradient can account for the majority of this difference. A $1.6 \pm 0.5\%$ P -velocity contrast and a $2.2 \pm 0.3\%$ density contrast at ~ 518 km depth without a S -velocity contrast can explain the lack of a $P520s$, together with robust $Pp520p$ and $S520S$ phases. For models parameterized with a finite thickness for each discontinuity, the 410-km discontinuity is consistently ~ 3 times thicker than the 660-km discontinuity.

Components: 11643 words, 11 figures, 6 tables.

Keywords: mantle transition zone; discontinuity topography; phase transition; mantle structure; niching genetic algorithm.

Index Terms: 7203 Seismology: Body waves; 7208 Seismology: Mantle (1212, 1213, 8124); 8124 Tectonophysics: Earth's interior: composition and state (1212, 7207, 7208, 8105).

Received 18 April 2006; **Revised** 27 July 2006; **Accepted** 30 August 2006; **Published** 25 October 2006.

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

1. Introduction

[2] The mantle transition zone is an enigmatic region within the Earth where seismic velocity and density increase rapidly with depth. The transition zone, bounded by discontinuities at ~ 410 km and ~ 660 km depth (often referred to as the 410 and the 660), is thought to be of key importance for controlling mantle convection and the resulting heat transport within the Earth. These

discontinuities are likely due to phase changes of olivine and other minerals as pressure increases with depth [Ringwood, 1975; Jackson, 1983]. Over the past several decades our understanding of this region has vastly improved due to exciting progress in mineral physics and seismic analysis, and massive increases in data quantity and coverage. A variety of reflected and converted seismic waves are sensitive to the depth and impedance contrasts across the transition zone discontinuities [e.g., Vinnik, 1977; Revenaugh and Jordan, 1989,

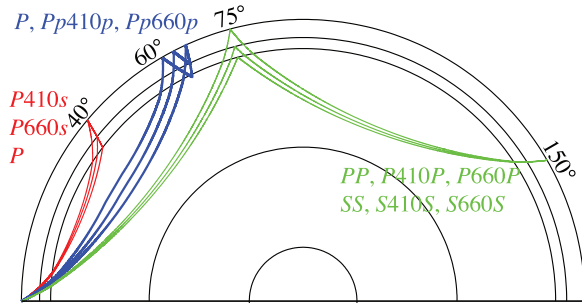


Figure 1. Raypaths of phases used in this study: *Pds*, *Ppdp*, *PdP*, and *SdS*. Note that all waves are symmetric from source to receiver except the receiver functions.

1991; Shearer, 1991, 1993]. Using a multitude of wave types and a variety of analyses, many general conclusions have been drawn regarding the nature of the transition zone.

[3] The discontinuities are shown to vary in topography on regional scales [e.g., Vidale and Benz, 1992; Wicks and Richards, 1993; Shen et al., 1996, 1998; Vinnik et al., 1996; Dueker and Sheehan, 1997; Flanagan and Shearer, 1998b; Li et al., 1998; Gilbert et al., 2003] and globally [e.g., Shearer and Masters, 1992; Shearer, 1993; Gossler and Kind, 1996; Gu et al., 1998, 2003; Flanagan and Shearer, 1998a; Chevrot et al., 1999; Gu and Dziewonski, 2002; Lawrence and Shearer, 2006] from studies using a variety of data types and processing techniques. The topographic variations of the 410- and 660-km discontinuities result in a thickening of the transition zone near subducted slabs and a thinning of the transition

zone elsewhere (especially beneath plumes [Li et al., 2003]). This thickening and thinning is likely the result of thermal variation in the mantle causing the respective phase changes to occur at different pressures (or depths) determined by Clapeyron slopes of opposite signs [e.g., Bina and Helffrich, 1994].

[4] The various types of data and processing techniques used to analyze topography and impedance contrasts often yield similar results, but frequently differ to some degree. Figure 1 shows a graphical summary of the different phases discussed in this study. While high-frequency *P*- and *S*-wave triplications are known to result from large velocity contrasts at ~660 km depth [e.g., Grand and Helmberger, 1984; Walck, 1984; Kennett, 1991; Ryberg et al., 1998] the *PP*-precursors (underside reflections off the discontinuities, *PdP*) indicate little to no impedance contrast at this depth range [e.g., Estabrook and Kind, 1996; Shearer and Flanagan, 1999].

[5] There are various problems with direct comparison among different types of data and analyses. One difficulty in relating results from different seismic phases is that each has different lateral coverage. While *SS*- and *PP*-precursors provide global coverage, triplication data are limited to seismic stations within ~33° of earthquakes. Figure 2 graphically represents the lateral data coverage for *Pds*, *PdP*, *SdS*, and *Ppdp*. Another difficulty arises from the frequency band used to study each type of wave. Longer-period waves,

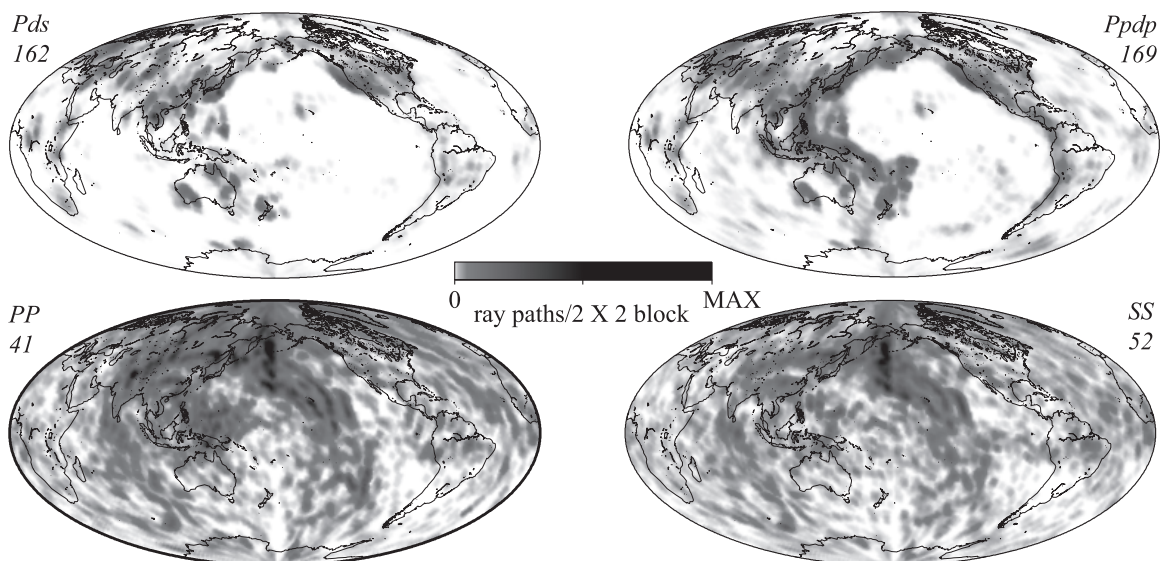


Figure 2. The number of stacked waves reflected or converted at the 660-km discontinuity for each 2° by 2° equal-area block. The coverage for the 410-km discontinuity is similar for each phase.

such as *SS*-precursors (or *SdS*), may view a sharp gradient as a discontinuity, whereas the shorter period *Pds* (*P*-to-*S* converted phases) can often differentiate between discontinuities and gradients.

[6] Additional problems arise due to different sensitivities in the Earth. These can be characterized in two ways. First, the waves are sensitive to varying scales of structure. For example, while *SdS* has a large X-shaped sensitivity kernel spanning more than 40° by 40° [Shearer, 1991; Shearer and Flanagan, 1999; Dahlen, 2005], *Pds* is only sensitive to a small region beneath each seismic station. Consequently, *Pds* can measure topography on the order of 50 km, while *SdS* studies typically average over 1000-km-wide regions. Nevertheless, the two data types yield roughly similar global patterns in transition zone thickness [Lawrence and Shearer, 2006]. Second, the different waves are sensitive to different elastic properties of the mantle. For example, the *SdS* is caused by reflectivity resulting from variations in shear velocity (V_S), and density (ρ), while *Pds* results from compressional to shear impedance, which is sensitive to compressional velocity (V_P), V_S , and ρ .

[7] In this study we characterize the transition zone by analyzing multiple data types: *SdS*, *PdP*, *Pds*, *Pdp* (Figure 1). The underside reflected *SdS* and *PdP* waves have similar coverage, but differing sensitivities to the elastic parameters. *S410S* and *S660S* are relatively large amplitude discontinuity phases that result in stable global stacks [e.g., Shearer, 1991, 1993; Flanagan and Shearer, 1998a; Gu et al., 1998]. The *S520S* is much weaker, but still robust once the effects of *S410S* and *S660S* are accounted for [Shearer, 1996; Ryberg et al., 1997]. Consequently, many of the constraints on transition zone structure come from *SdS* stacks. While *P410P* is a robust feature, *P660P* is surprisingly low amplitude [Estabrook and Kind, 1996]. The low amplitude of *P660P* appears even more remarkable when compared to that of the topside *P*-wave reflection: *Pdp*, which has large amplitudes for *Pp410p*, *Pp520p*, and *Pp660p* [Nguyen-Hai, 1963; Husebye and Madariaga, 1970; Davies et al., 1971; Gutowski and Kanasewich, 1974; Ward, 1978; Shearer, 1991]. *Pds* is known to have a similar global average and lateral variation in transition zone thickness to that of *SdS* [Lawrence and Shearer, 2006], which suggests that the two data types are compatible. However, the global stack of *Pds* lacks a *P520s* [Lawrence and Shearer, 2006], which is in contrast to the robust *S520S*.

[8] We test the hypothesis that the apparent discrepancies between the phases described here are largely reconcilable given the different sensitivities. To test this hypothesis we employ ray-theory waveform modeling and compare synthetic waveforms to observed stacks of each wave type. In doing so we solve for one model that best fits all of the data. Due to the complexity of this endeavor, linear inversions are impractical. Instead we use a mass-forward modeling technique, called the niching genetic algorithm (NGA), to locate the most optimal solution. While computationally expensive relative to linear inversions, the NGA uses an evolutionary paradigm to search the entire model space, and efficiently iterates toward the best solution. One advantage to this technique is that it allows for model comparison and trade-off analysis. We discuss the limitations of our model, data, and technique, and then draw conclusions from the robust features of the best-fitting model.

2. Data

[9] We employ an automated system for data selection and processing. First, all available digital long-period waveforms recorded from earthquakes having magnitude $M_b > 5$, between 1976 and 2003, are downloaded from the Incorporated Research Institutions for Seismology (IRIS) Data Management System (DMS). Then we correct for instrument response, rotate into vertical, tangential, and radial components, and calculate the signal-to-noise ratio (SNR) for *P*, *PP*, and *SS* waves. The maximum and minimum amplitudes of the signal, S , (from 10 seconds prior to 50 seconds following the theoretical travel time of each phase) are compared with the maximum and minimum of the noise, N , (the 60 seconds prior to the first arriving wave). Specifically, the signal-to-noise ratio is, $SNR = (S_{max} - S_{min}) / (N_{max} - N_{min})$. We only consider waves with SNR greater than 2. *Pds* and *Pdp* are considered for high SNR *P* waveforms recorded between 30° and 90° from the earthquake epicenter. These records are band pass filtered between 0.2 and 0.02 Hz. *SdS* and *PdP* are analyzed on all high SNR *SS* and *PP* waveforms recorded between 90 and 180 degrees. The *SS* and *PP* waves are band pass filtered between 0.1 and 0.01 Hz.

3. Stacks

[10] The stacking process of each wave is automated in a similar manner to Shearer [1991] in

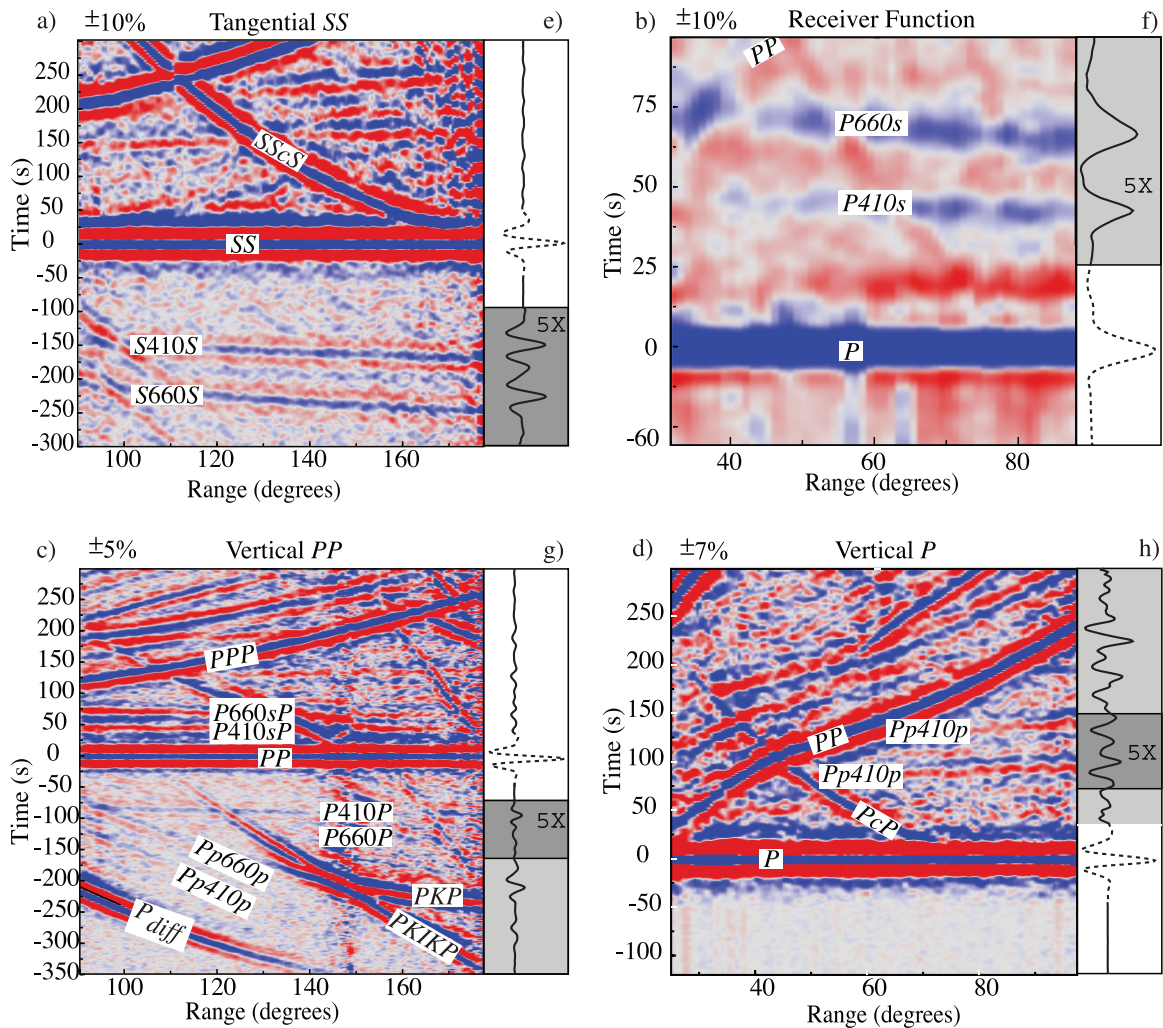
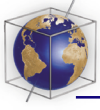


Figure 3. The (a–d) 2-D and (e–h) 1-D stacks for (a and e) *SdS*, (b and f) receiver functions, (c and g) *PdP*, and (d and h) *Ppdp*. The 2-D stacks represent relative amplitude with respect to a reference phase (shown at zero time) as functions of time and distance, where blue is positive and red is negative. *Ppdp* has a negative amplitude (red), and the rest have positive amplitudes (blue). The 2-D color plots are saturated at different levels to accentuate each phase, as marked. The 1-D stacks are obtained by stacking the 2-D results along the predicted travel time curves for discontinuities at different depths. The 1-D plots are amplified in the gray sections. The dark gray sections illuminate the phases of interest.

order to minimize bias. After performing quality control and preprocessing as described above, the waveforms are aligned on the *P*, *PP*, or *SS* reference phase. The time shift for each record is simply the time of maximum absolute amplitude within the “signal” window. The polarity is reversed for negative amplitude peaks. The amplitude of each wave is normalized such that the picked amplitude is set to the SNR, which emphasizes clean data and dampens noisy data. All amplitudes greater than \pm SNR (after the normalization) are capped at \pm SNR. Waveforms are binned and stacked according to event-to-station distance, with a bin size of 0.5° . A nine-point mean smoothing filter is applied to

the 2-D stack of amplitude plotted on time v. epicentral distance (Figure 3). The stacking for *Pds* varies from the others in only two respects. The vertical component is spectrally deconvolved from the radial component after alignment and prior to weighting. A water level of 0.02 and a Gaussian filter width of 0.4 are used to stabilize the spectrally deconvolved receiver functions [e.g., Ammon, 1991; C. J. Ammon, An overview of receiver-function analysis, <http://eqseis.geosc.psu.edu/~cammon/HTML/RftnDocs/rftn01.html>, 2006]. Additionally, the receiver functions are stacked into 1° bins rather than 0.5° bins.

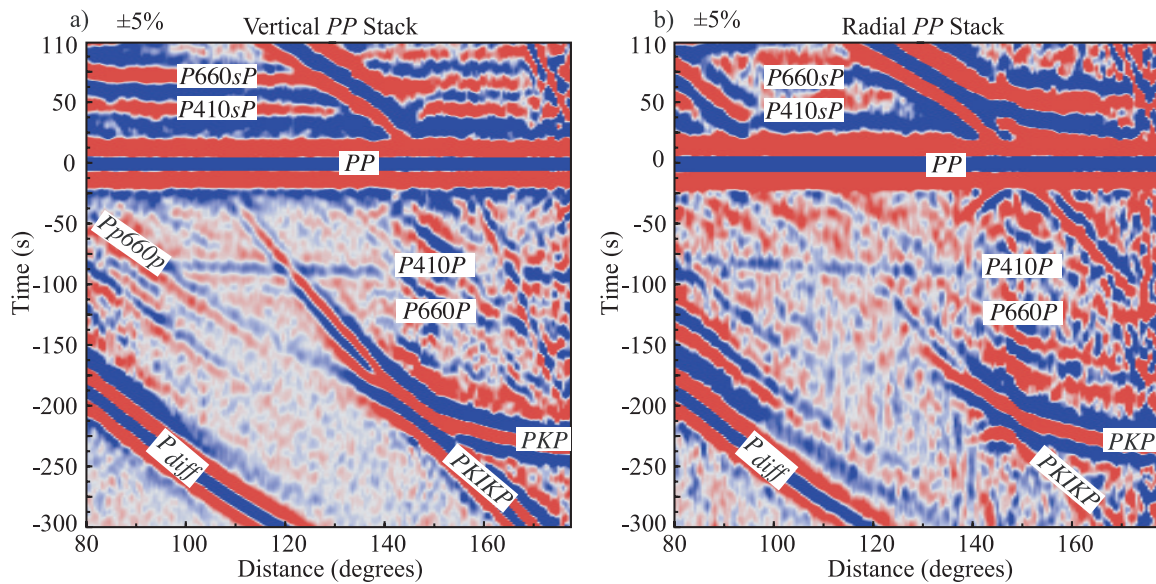


Figure 4. (a) Vertical and (b) radial stacks of *PdP* similar to Figure 3c. *P420P* is visible in both, and *P520P* is visible in the vertical stack. *P660P* is not visible in either stack. The high-amplitude *Ppdp* and *PKP* phases interfere with *PdP* in the vertical stack, but to a lesser degree in the horizontal stack.

[11] The stacked amplitudes are plotted on time v. distance maps with positive amplitudes in blue and negative amplitudes in red (Figures 3–5). The amplitudes are relative to the reference phase (*P*, *PP*, or *SS*), and saturation levels are noted in each subpanel. As observed by numerous other studies [e.g., Shearer, 1991; Shearer and Flanagan, 1999; Gu et al., 1998, 2003], the stacked *SdS* phases stand out several minutes prior to *SS* as having signal well above the noise (Figure 3a). The *P410P* is visible in the distance ranges 100°–118° and 130°–140° on the vertical *PP* stack (Figure 3c). The *P520P* and/or sidelobe of the *P410P* is visible between 106° and 123°. However, *P660P* is difficult to see in this distance range due to low amplitude, interfering waves, or both. On the radial *PP* stack (Figure 4), the interfering waves are damped more than the *PdP* phases because these waves arrive more steeply and therefore are recorded with lower amplitudes on the horizontal component (Figure 5). While the radial stack is less stable because the radial waveforms have lower SNR, the *P410P* is clearly visible without interference between 84° and 140° degrees. Clearly, if the *P520P* and *P660P* were of similar amplitude to *P410P*, these phases would also be visible on the radial stack, but they are not.

[12] The stacked *Ppdp* are clearly visible on the vertical *P*- and *PP*-wave stacks (Figures 3c and 3d) for the 410-, 520-, and 660-km discontinuities. The *Pp410p* is visible from 52° to 120°, while the

Pp660p is limited to between 70° and 120°. We do not analyze *Ppdp* beyond 90° to limit contamination due to core-mantle boundary phases and the heterogeneous lowermost mantle. The *Pp520p* and/or the sidelobes of the *Pp660p* and *Pp410p* is visible from 57° to 90°. The *Ppdp* all have negative amplitudes relative to the direct *P* wave, so they appear red. The *P660s* and *P410s* appear as robust phases on the receiver function plot (vertical deconvolved from radial; Figure 3b) from 40° to 90°, but the *P520s* is absent in this range.

[13] The 2-D stacks are collapsed into 1-D stacks by summing the stacks for each distance and depth along the appropriate set of moveouts associated with the desired phase (Figures 3e–3h). This summation is conducted prior to smoothing the 2-D stacks to avoid unnecessary pulse broadening/damping. This is achieved with minimal waveform distortion by first interpolating from travel time to bounce depth or conversion depth using the correct distance range for each bin, and then interpolating depth back to time using a single distance. In this manner, the 2-D stack is correctly collapsed for all times and distances, and bias is reduced by not collapsing the stacks along a single moveout associated with a particular depth. Times associated with negative depths are simply stacked with the moveout of the reference phase of the initial stack (*P*, *PP*, or *SS*). Some features appear more clearly in these 1-D stacks than in the 3-D stacks. For example, *P520P* and *P660P* are visible in the 1-D

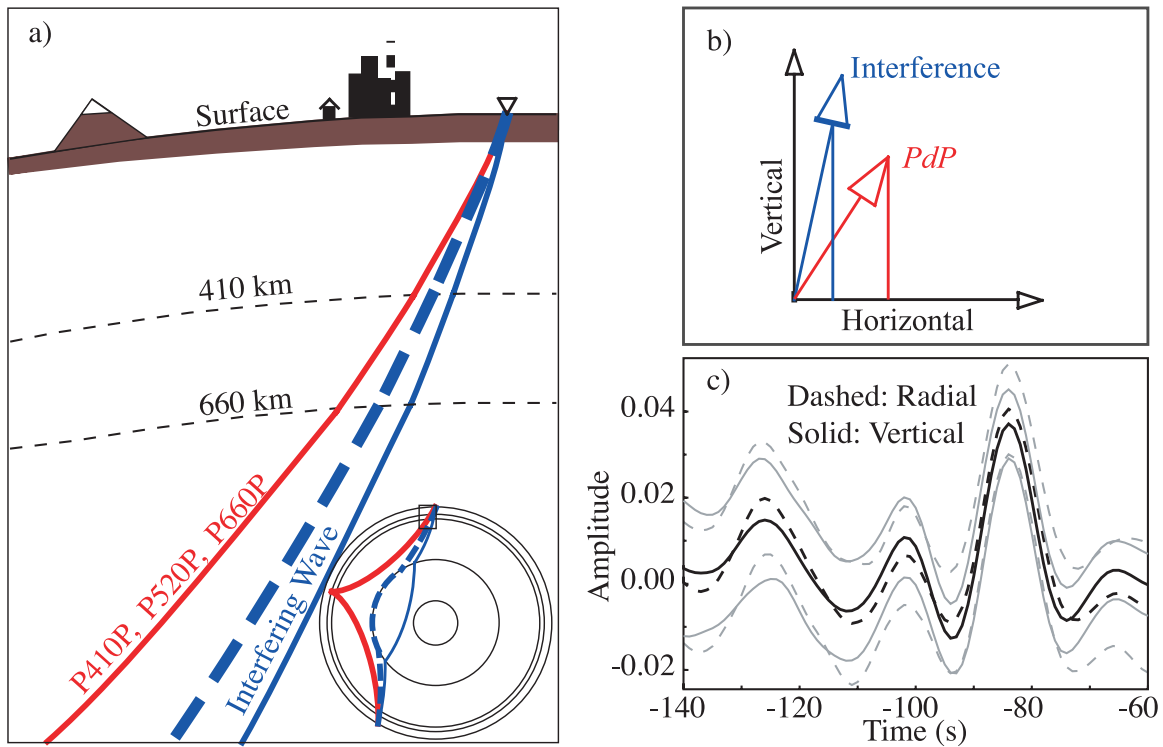


Figure 5. *PP*-precursors (*PdP*) are imaged on vertical and radial components. (a) Interfering waves such as *PKP* (blue solid) and *Ppdp_{diff}* (blue dashed) follow steeper raypaths than *PdP*. (b) By looking at the radial waveform, the amplitudes of the *PdP* phase are accentuated relative to interfering waves. However, the *PdP* amplitudes are also lower relative to ambient noise. (c) The 1-D stacks of *PdP* are similar for vertical and radial component waveforms, suggesting that the stacks are reliable and that undesired phases are not influencing the 1-D stack.

stack, whereas they were below the noise in the 2-D stacks. Error bounds for each stacked waveform are determined using a bootstrap method.

4. Waveform Modeling

[14] In this section we describe the creation of synthetic waveforms and the technique we use to compare transition zone structure with observed waveforms. We employ the following ray-theoretical method because it is fast and computationally efficient. As discussed below, we calculate hundreds of thousands of synthetic waves in a mass-forward modeling technique, so computational efficiency is crucial. As with *Shearer* [1996] we start with the assumption that the source function is given by the reference phase of each stack (*P*, *PP*, or *SS*), and that only the 1-D elastic structure in the transition zone is important. These assumptions are fair considering that the reference phase waveforms are stacked from thousands of traces. Consequently, the reference phases do approximate source time functions [e.g., *Shearer*, 1991], and the laterally varying structure is effec-

tively averaged into 1-D. Therefore the waveform is easily calculated by convolving the transition zone's elastic response function with the reference phase.

[15] We follow the subsequent steps in the calculation of each synthetic waveform. Figure 6 graphically illustrates the steps involved in the computation for a synthetic *SS*-precursor, but the technique is equivalent for each wave type. First, we construct a discrete 1-D velocity and density model as a function of depth, $V(z)$. Then, given an event-to-station distance, we calculate the ray theoretical travel time of a reflection or *P*-to-*S* conversion at each depth, and map the velocity structure into time, $V(t)$. The amplitude of a reflected or transmitted phase is calculated from the reflection or transmission coefficient [*Aki and Richards*, 1980] associated with each depth or time, $R(t)$. The amplitudes of this Earth response function are corrected, $s(t)$, by scaling according to the change in amplitude, $A(t)$, due to geometric spreading and anelastic decay: $s(t) = A(t)/A(t_0)R(t)$, where t_0 is the time of the reference phase. We only illustrate $S(t)$ here and not $R(t)$ because the geo-

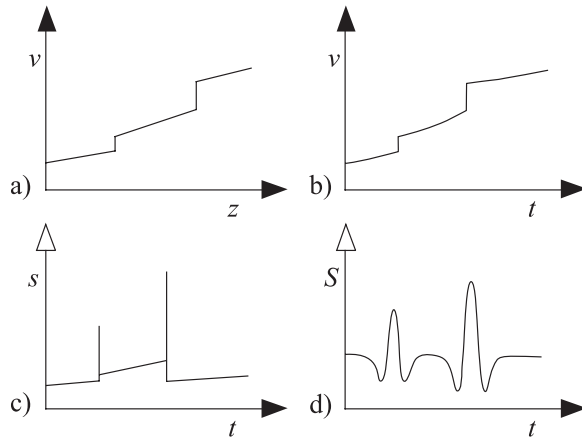


Figure 6. A graphical representation of the synthetic waveform calculation for SdS . We convert (a) the velocity-depth model into (b) a velocity-time model from the ray theoretical travel time of reflected or converted from each depth. (c) The reflection or transmission coefficient is calculated for each wave and adjusted by a geometric spreading factor before it is convolved with the source function to determine (d) the synthetic SdS . Synthetics of other waves are calculated in this same manner. This figure is after Shearer [1996].

metric spreading factor is typically near unity for raypaths that are similar to the reference phase. Therefore $R(t)$ and $S(t)$ are very similar. For geometric spreading we use $[r_0/r(t)]^2$, where r is the distance traveled by the phase reflected/transmitted at a depth corresponding to time t . For anelastic decay we employ the quality factor model, QL6 [Durek and Ekstrom, 1996]. The final step is to convolve the reference pulse, $Ref(t)$, with the amplitude corrected Earth response function, $S(t) = Ref(t)*s(t)$. The calculation of a synthetic receiver function, $S(t)$ varies in that the vertical synthetic, $S_Z(t)$, is deconvolved from the horizontal synthetic, $S_R(t)$, and that multiple phases are modeled simultaneously (Pds and $Ppdp$). For consistency, the spectral division used in calculation of the synthetic receiver functions is stabilized with a water level of 0.02 and Gaussian filter width of 0.4, in the same manner as the observed receiver functions.

[16] Under the assumption that the transition zone discontinuities have laterally varying topography, and that this topography causes pulse broadening in the observed stacks, we apply pulse broadening to our synthetic waveforms. The distributions of discontinuity topography from Flanagan and Shearer [1998a] are assumed to represent the true distribution of topography. Standard deviations of

21.8, 28.0, and 33.8 km are calculated for the 410-, 520-, and 660-km discontinuities respectively. We divide the discontinuity among the depths within 2 standard deviations of the modeled depth giving each depth a velocity increase proportional to modeled discontinuity jump multiplied by the Gaussian operator associated with that depth. This has the effect of smoothing the discontinuities resulting in pulse broadening.

5. Upper Mantle Corrections

[17] One-dimensional models produce artifacts due to three-dimensional heterogeneities in crustal and mantle structure. One problem with heterogeneity is that the average crust and mantle structure, as sampled by the various phases, is different from the geographically averaged crustal structure predicted by the model, AK135(-F) [Montagner and Kennett, 1996; Kennett et al., 1995; Engdahl et al., 1998]. Three-dimensional heterogeneity in the upper mantle has differing effects on the global stacks of the various discontinuity phases due to different data coverage. Therefore, in order to model the 1-D transition zone structure accurately, we must account for the effects of three-dimensional crust and upper-mantle structure. Theoretical travel time residuals between a reference phase and discontinuity phases (e.g., $SS-SdS$) are determined by tracing the appropriate rays through 3-D velocity models of the mantle (SB10L18 [Masters et al., 2000]) and the crust (CRUST 2.0 [Bassin et al., 2000]). The model SB10L18 was chosen because it accounts for both P - and S -velocity anomalies equally. We calculate three (one for each discontinuity) theoretical travel time residuals for each waveform that goes into each stack. These travel time residuals are migrated back to the reference distance (or ray parameter) for each 1-D stack. The average travel time residual for each stacked waveform is nonzero because the data coverage is uneven. In order to facilitate comparison between the observed and synthetic waveforms, we shift the synthetic waveforms by the theoretical travel time residual calculated for each phase. We apply the theoretical travel time residual as a single average time shift with opposite sign for the whole synthetic waveform rather than distorting the waveform by shifting each phase independently. These time shifts are 0.4 s for Pds , 0.2 s for $Ppdp$, 0.5 s for PdP , and 0.9 s for SS . The net result of these corrections is to deepen the transition zone interfaces by ~ 3 km.

[18] Shearer [1991] and Flanagan and Shearer [1998a] tested for a systematic offset in SS travel

Table 1. Parameterization 1

Discontinuity	V_S , %	V_B , %	ρ , %	Depth, km
410	0–12	0–12	0–6	385–435
520	0–12	0–12	0–6	495–545
660	0–12	0–12	0–6	635–685
Sub-660 Peg-depth	-	-	-	700–820

times due to upper mantle-structure by computing the SS - S travel time residual, δt , relative to the reference model. This was accomplished by cross-correlation of the Hilbert transform of the S wave with the SS wave for the stacked waves with distances from 65° to 95° degrees. In practice, the correction is ill-constrained, varying as a function of distance, the reference quality factor model, and whether the reference stack is S or SS . Beyond a distance of $\sim 90^\circ$ the S wave interacts with the highly heterogeneous lowermost mantle and the core-mantle boundary, which contaminates the δt value. The value of δt varies between 0.03 s and 1 s for different subsets of stacked waveforms from 65° to 85° using the reference model, QL6 [Durek and Ekstrom, 1996] with a reference stack of S . The value ranges from -0.2 s to 0.7 s for an equivalent stack referenced to SS . The δt values change by ± 0.6 s when using different quality factor models (PREM [Dziewonski and Anderson, 1981], PAR3P [Okal and Jo, 1990], QM1 [Widmer et al., 1991], AK135(-F) [Montagner and Kennett, 1996] and QLM9 [Lawrence and Wysession, 2006]). The best estimate we have of the SS - S correction is 0.4 ± 0.7 s. The equivalent estimate for PP - P is -0.2 ± 0.5 s. Because of the small value and large error bars of these corrections, we choose to ignore them.

6. Forward Modeling Approach

[19] We determine the model that best explains the data by searching the whole model space with an automatic algorithm rather than searching the model space by hand. While trial and error forward modeling is a reasonable approach to fit a single discontinuity phase [e.g., Shearer, 1996], it is difficult to account for tradeoffs between the model and multiple waveforms by hand. Linear inversion is not always stable in an environment where tradeoffs exist between parameters (e.g., V_S , V_B , and ρ , and depth). Consequently, we choose a parameterization that requires as few free variables as possible so that mass forward modeling is made feasible. Given the parameterization described below, with only 13 search

variables (3 ΔV_S , 3 ΔV_B , 3 $\Delta \rho$, 3 discontinuity depths, and 1 peg depth), a simple two-stage grid search would require the creation of more than 2×10^{13} models with associated synthetic waveforms to locate a reasonable model without severely limiting the viable multidimensional parameter space. The creation of so many models is not practical. As discussed below, even a more sophisticated search algorithm requires the computation of 10^4 – 10^6 models and associated synthetic waveforms to ensure that the whole model space is searched, even with a simplified parameterization.

[20] We model three discontinuities with 0–12% contrasts in V_S , V_B and ρ at 410 ± 25 , 520 ± 25 , and 660 ± 25 km depth. Above and below each discontinuity, the velocities and density increase with the same gradients as AK135(-F) [Montagner and Kennett, 1996]. AK135(-F) is a modified version of AK135 [Kennett et al., 1995; Engdahl et al., 1998] that includes a 1-D density profile of the Earth that is constrained by normal modes in addition to fitting the ISC travel times of P and S waves. Hereafter, AK135(-F) is referred to as AK135. Additionally, we model a steep velocity gradient beneath the 660-km discontinuity with a single parameter describing a peg-depth (700–820 km), below which the model is set to AK135. This sub-660 gradient is defined by the modeled bottom-side V_S , V_B and ρ at the 660-km discontinuity and the AK135 velocity and density at the peg depth. Above and below each discontinuity, the velocities increase with the same gradient as AK135. This parameterization allows us to model V_S , V_B and ρ from 300 km to 850 km with only the 13 parameters described in Table 1. More complex parameterizations with 16 and 17 variable parameters are also discussed below.

7. Niching Genetic Algorithm (NGA)

[21] Rather than applying a simple grid search we use a niching genetic algorithm (NGA) [Mahfoud, 1995; Koper et al., 1999], which uses an evolutionary paradigm, to search the model space efficiently for locally optimal and globally optimal solutions. We refer readers to Koper et al. [1999] for a general overview of NGA and its application to geophysical problems, and only provide a cursory description here. A standard genetic algorithm operates by first creating a population of random models and comparing the forward model with the observed data. Models associated with high misfit, or cost, are removed from the population of models. Models with low cost continue on to the next

generation, where new models are constructed from random perturbation and cross-breeding between the best models. In this manner, after several generations, only models associated with low cost are retained, and the population converges toward an optimal solution. The NGA is a compound version of the genetic algorithm, where multiple genetic algorithms, each controlling a subpopulation, compete for a portion of the model space. The competition is imposed by applying an artificially high cost to any model in a lower-order subpopulation that is sufficiently similar to the best models of the higher-order subpopulation.

[22] The cost is calculated as the normalized sum of squared differences between the observed and synthetic waveforms. For each stacked waveform at each discrete time step, i , we have a standard deviation (σ^2) calculated using the bootstrap method. We therefore normalize the difference by the variance (σ) before summing the squared difference. The sum of squared differences normalizes the misfit by the number of points, M , so the NGA does not favor one waveform over another. The total misfit cost is described by equation (1):

$$Cost = \sum_{n=1}^{\#waves} \sqrt{\sum_{i=1}^M [(syn_i - obs_i)/\sigma_i]^2} / M + A. \quad (1)$$

[23] We place several a priori constraints, A , on the forward modeling process to ensure that it converges toward a realistic solution. Under the assumption that the mass of the Earth is well resolved, we impose a cost associated with excess or shortage of mass relative to AK135; the cost is equal to the difference in the sums of density divided by the number of layers (500). We do not directly impose moment of inertia constraints because we only model small density differences from AK135, so the redistribution of mass is minor. Rather than modeling the waveform at one distance, we model the waves at two distances, which imposes the general constraints of AK135 on the model because the stacked waveforms are summed along the moveout prescribed by AK135. These distances are 120° and 140° for PP and SS , and 70° and 85° for Pds and $Ppdp$. For graphical purposes in the following figures (Figures 711–13) we only present waveform fits associated with the greater distance for each waveform. Just as a secondary precaution we add another cost equal to the difference in vertical travel times from AK135 for P and S waves. Therefore the resultant model should not violate the data that were used in the creation of AK135.

[24] There are several advantages of using a niching genetic algorithm [e.g., *Koper et al.*, 1999]. First, the NGA is faster and more computationally efficient than a standard grid search. Second, it does not depend upon a starting model. Third, it allows the user to examine a suit of models, rather than just one. This is important for understanding tradeoff and placing error bounds on the best estimate. Fourth, the NGA locates local minima with each subpopulation so that if alternate solutions exist they will be located.

8. Results

8.1. Parameterization 1: Step Function Discontinuities

[25] In this study the NGA searches the model space with 10 subpopulations having 20 models each for 100 generations. Consequently, 20,000 models and associated synthetic waveforms are created and compared with each stacked waveform for each simulated inversion. With five NGA runs using different randomizing seeds, 100,000 models are compared with the stacked waveforms. The improvement in cost declines approximately as an exponential decay, such that there is little improvement after 30–40 generations (Figure 7a). We run the NGA five times with different randomizing seeds to ensure that the most optimal model is a robust solution. As a rule of thumb, reasonable models have lower cost than the worst model in the best subpopulation of the last generation. The population density functions (PDF) of these reasonable models are shown in Figures 7c and 7d. The most-optimal model occupies the center of this PDF, which indicates that the most-optimal model is representative of the suite of models that best fit the observed waveforms. In this case all five most optimal models obtained with different randomization seeds fit within the same PDF. While not shown here, the most-optimal model does not occupy the center of the PDF associated with all models in all subpopulations and all generations because other subpopulations represented other lesser-optimal local cost minima.

[26] The fit to the observed data may be seen in Figures 7h–7k. Clearly the fit for the best model is better than that of AK135. The model AK135 (Figures 7d–7g), having no structure at 520 km depth, does not provide a good fit to any of the 520-km depth phases. In general, the 410-km depth phases are well modeled by AK135. AK135 greatly overestimates the amplitude of the 660-km compress-

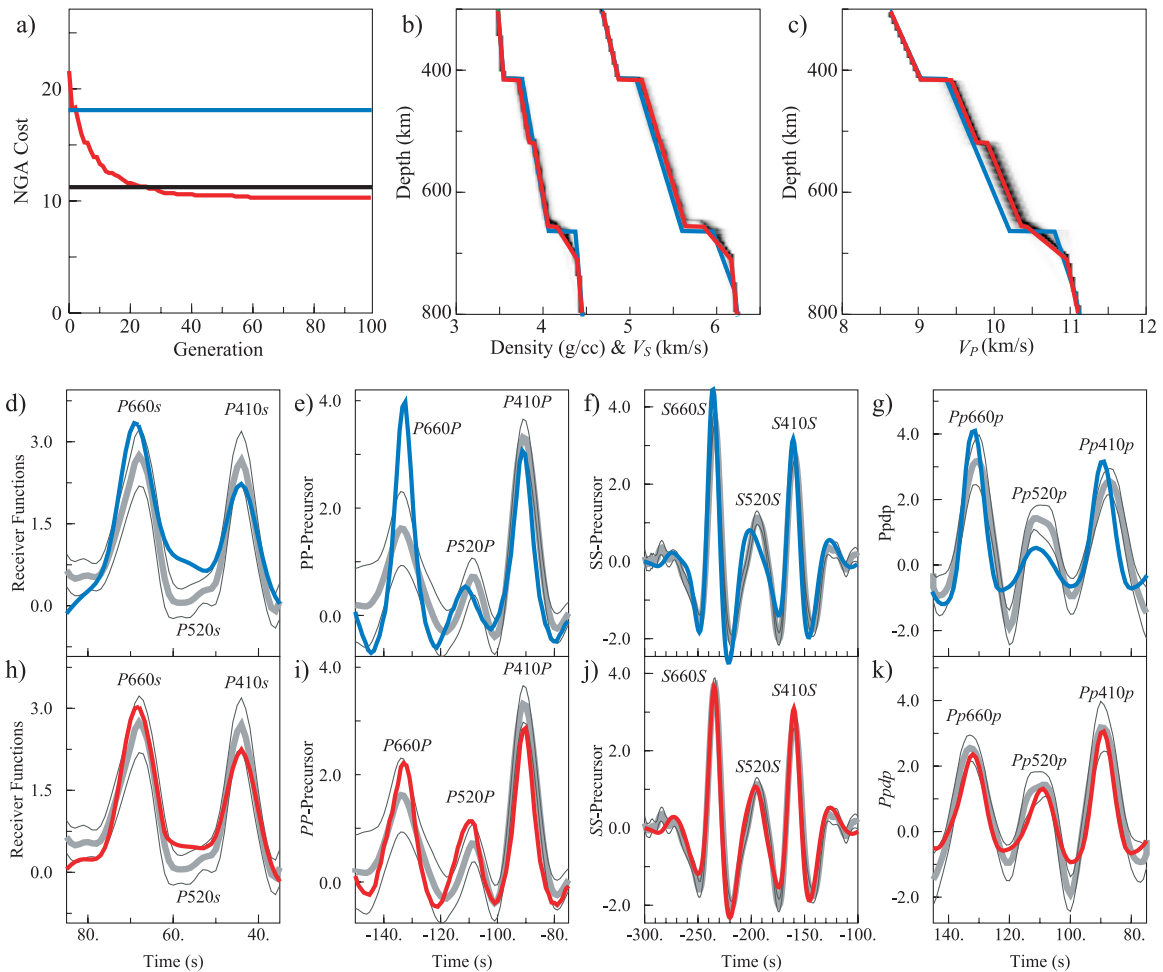


Figure 7. After (a) 100 generations of optimization with the niching genetic algorithm (NGA) the best fit model (red) of (b) density, shear velocity, and (c) compressional velocity has a better fit (Figures 7h–7k) to the data (gray) than does AK135 (blue, Figures 7d–7g). Each phase is labeled. The black line in Figure 7a represents the maximum error in the best subpopulation at the end of the NGA search. The probability density functions (PDF) of the models having cost less than the black line are plotted under the best-fit models to show the possible tradeoff. Black indicates high probability, and white indicates low.

sional phases ($P660P$ and $Pp660p$). The best-fit model does a much better job at fitting the 520 and the 660 phases than AK135. Nevertheless, there is still significant misfit between the stacked waveforms and the synthetic waveforms associated with the most optimal model, which stems from the simplicity of the parameterization. A more complex model parameterization can match the data better at the cost of having greater trade-off between parameters.

[27] The best-fit model is listed in Table 2, and has the following characteristics. The depths of the interfaces are well constrained at 413, 515, and 654 km depth, indicating a 241 km thick transition zone. Note that two other less optimal subpopulations consistently found that the 660-km disconti-

nuity could be modeled at greater or lesser depths with little additional misfit to the observed data. The density contrasts at each interface are $\Delta\rho_{410} = 4.8\%$, $\Delta\rho_{520} = 2.7\%$, and $\Delta\rho_{660} = 4.4\%$, which indicates that the 660 has less of a density contrast, and the 520 has a larger contrast than previously estimated [Shearer, 2000]. Yet, the density contrast of the 660 and the sub-660 density gradient accounts for a density increase greater than both the 410- and 520-km discontinuities. The large shear and compressional impedance contrasts necessary for $S660S$ and $Pp660p$ are maintained by the large gradient beneath the 660.

[28] The most-optimal model has several key differences from AK135. First, we observe a global 520-km discontinuity with significant density

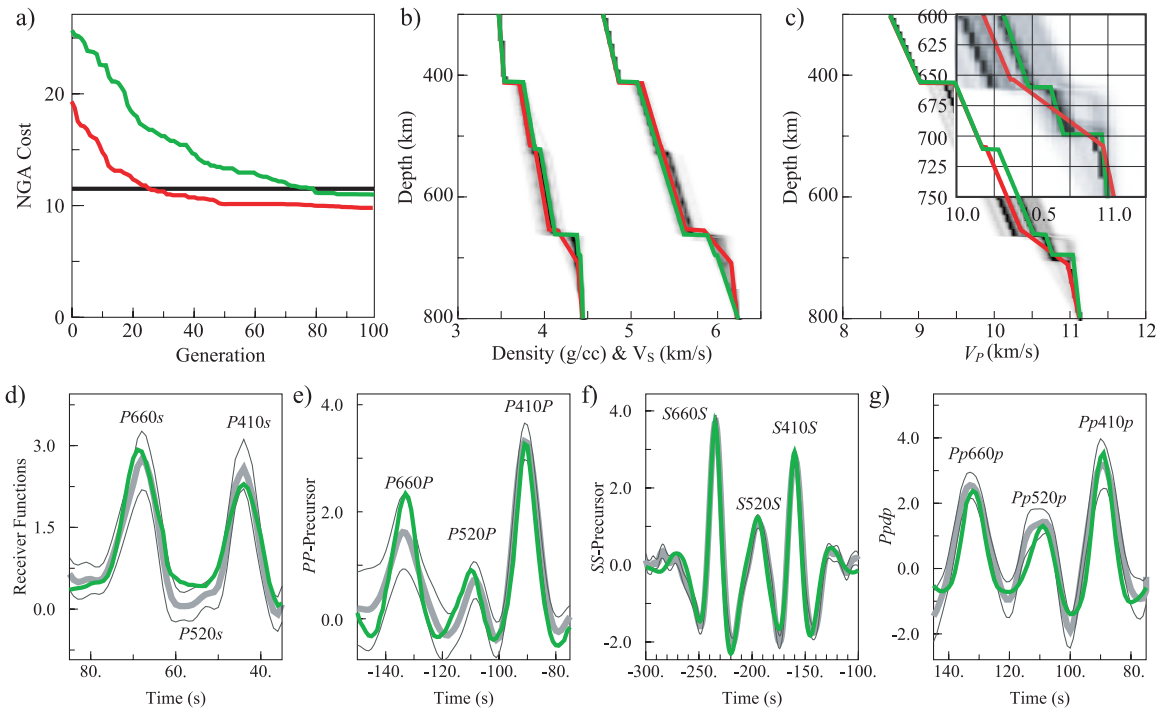
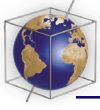


Figure 8. Similar to Figure 7, but for a model having two discontinuities near 660 km depth. While the red model has lower cost, it is very similar to the model represented in Figure 7 and as such adds no great insight into transition zone structure. The NGA also located an alternate subpopulation (green) that had two distinct interfaces near 660-km depth and had cost less than the maximum of the best subpopulation. This model has no steep gradients below the 660 km depth. The PDF in Figures 8b and 8c (high probability is black, and low is white) is limited to all models but those in the best subpopulation to indicate likelihood of model excluding models similar to Figure 7 or the best-fit model (red).

and compressional velocity contrasts. Second, the 660-km discontinuity is shallower than its 660-km depth in AK135, at approximately 654 km depth. Third, the sub-660 gradients are steeper and account for larger portions of the total contrast near 660 km depth than in AK135.

8.2. Parameterization 2: A Dual 660 Interface

[29] Noting the limitations of our forward modeling approach, we attempt to construct alternate models using different parameterizations. It is possible for multiple phase transformations to occur near 660 km depth; in a pyrolytic composition a transition to garnet accompanies the β - γ phase change [e.g., Weidner and Wang, 2000; Deuss *et al.*, 2006]. Depending on temperature, composition, and water concentration these phase transformations may cause multiple seismic discontinuities or steep gradients in elastic properties rather than discontinuities. We attempt to model two interfaces for the 660-km discontinuity because it is the only discontinuity in which

multiple distinct depths were localized by different subpopulations of the model search described above. This adds four additional parameters: one for the vertical distance to the lower interface, and the V_S , V_P and ρ contrasts at the lower interface (Table 3). We allow the lower interface to vary from 1 to 60 km beneath the upper 660-km interface. In order to model twinning, or interference effects, we multiplied the Gaussian widths by a factor of $\zeta = 0.25$, so that the pulse broadening effect during the synthetic calculation does not effectively smooth over the two interfaces. The same Gaussian width is used for both the 660-km discontinuity and the lower interface. We find that the most optimal models for five NGA runs with different randomizing seeds are very similar to that described above (Figure 8). In each case the two interfaces are less than 10 km apart, or the lower interface has negligible contrasts. Because of the high similarity with the model discussed previously, we do not focus on the most-optimal solution. A distinctly different locally optimal model for four out of the five NGA runs had nearly equally low misfit, suggesting an alternate possible model type.

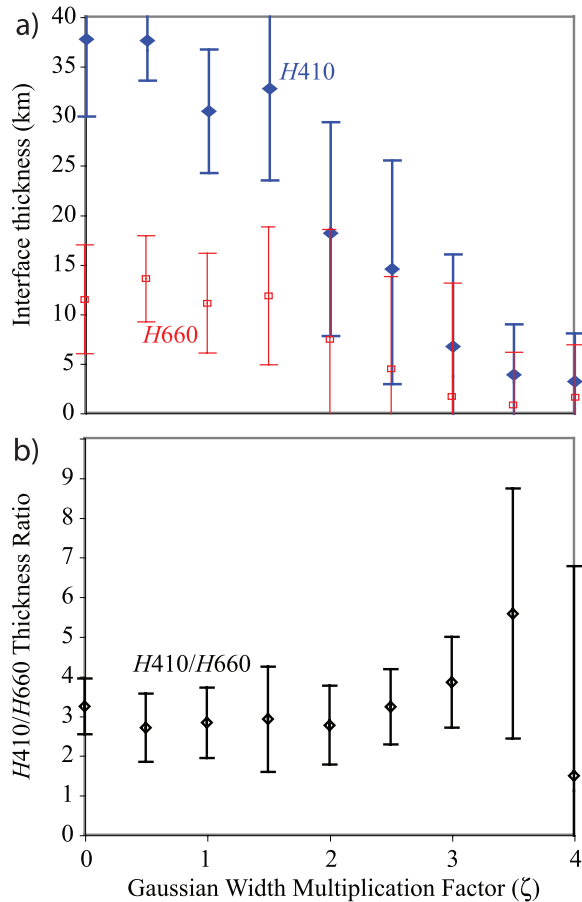


Figure 9. The most optimal models from NGA simulated inversions parameterized with interfaces having finite thickness vary as a function of the Gaussian filter used to approximate pulse broadening resulting from 3-D structure. (a) The thickness of the 410-km (*H410*) and 660-km (*H660*) discontinuities decreases as the factor (ζ) (by which Gaussian widths are multiplied) increases. (b) The ratio between the thicknesses (*H410/H660*) remains constant at ~ 3 for most values of ζ .

In particular models having one interface at ~ 664 km depth and another at ~ 699 km depth with the V_P jump distributed nearly evenly between the two were successful at modeling the various phases (Table 4). In these models, the deeper interface had only small ($<0.8\%$) V_S and ρ jumps and the steep sub-660 gradient is not present. The dual-discontinuity causes twinning, or wave interference, which dampens the amplitude of the combined waveforms.

8.3. Parameterization 3: Interface Thickness

[30] Another parameter that we explore is the thickness of each interface. The assumption here

is that the discontinuities may have some finite width. In this parameterization we add three variables: one width for each discontinuity (Table 5). In each case we allow the width (*H410*, *H520*, and *H660*) to vary between 0 and 40 km. The Gaussian pulse broadening is applied for all layers within the finite thickness of the interface. The V_S , V_P and ρ perturbations account for the elastic parameter jump from the top to the bottom of the interface. In a finite width interface the velocity and density contrasts across the interface have both adiabatic and nonadiabatic components. We convert the parameterized contrasts to nonadiabatic contrasts associated with each discontinuity by removing the average AK135 gradient over the 420-600 km depth range. This correction is typically small (<0.1 km/s). Because of the inherent uncertainty of the amount of broadening one should expect from topography and seismic velocity heterogeneity, the thickness estimate is highly nonunique. For this reason, we solve for structures using nine distinct multiplication factors ($\zeta = [0, 0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0]$), with which we multiply the Gaussian filter. A factor of zero models a homogeneous Earth. A factor of 4.0 models an Earth that is 4 times more heterogeneous than modeled by *Flanagan and Shearer [1998a]*. Reasonable values of ζ are likely between 0.5 and 2.0. For each of the 9 values of multiplication factor we run the NGA five times with different randomization seeds to ensure that we obtain robust results. Figure 9a shows that as the Gaussian filter width increases the median thickness of the five resultant most optimal models decreases. While the actual thicknesses of the interfaces varies as a function of pulse broadening, the ratio between the thicknesses for the 410- and 660-km discontinuities (*H410/H660*) is consistently ~ 3 for a wide range of ζ (Figure 9b). For high multiplication factors, the long-period waves are insensitive to small changes in interface thickness, so the ratio (*H410/H660*) becomes unstable. For high multiplication factors ($\zeta > 4.0$), the pulse broadening is too large, resulting in higher misfit.

[31] We present the most-optimal model with the median 410-km interface thickness solved with a multiplication factor of $\zeta = 1.0$. This low multiplication factor likely exaggerates the interface thickness, better emphasizing the effect of interface thickness. Table 6 and Figure 10 characterize this best-fit model. Here, the amplitudes of the elastic perturbations are greater than for the zero thickness interface method described above. This is due to the tradeoff between the impedance contrast and

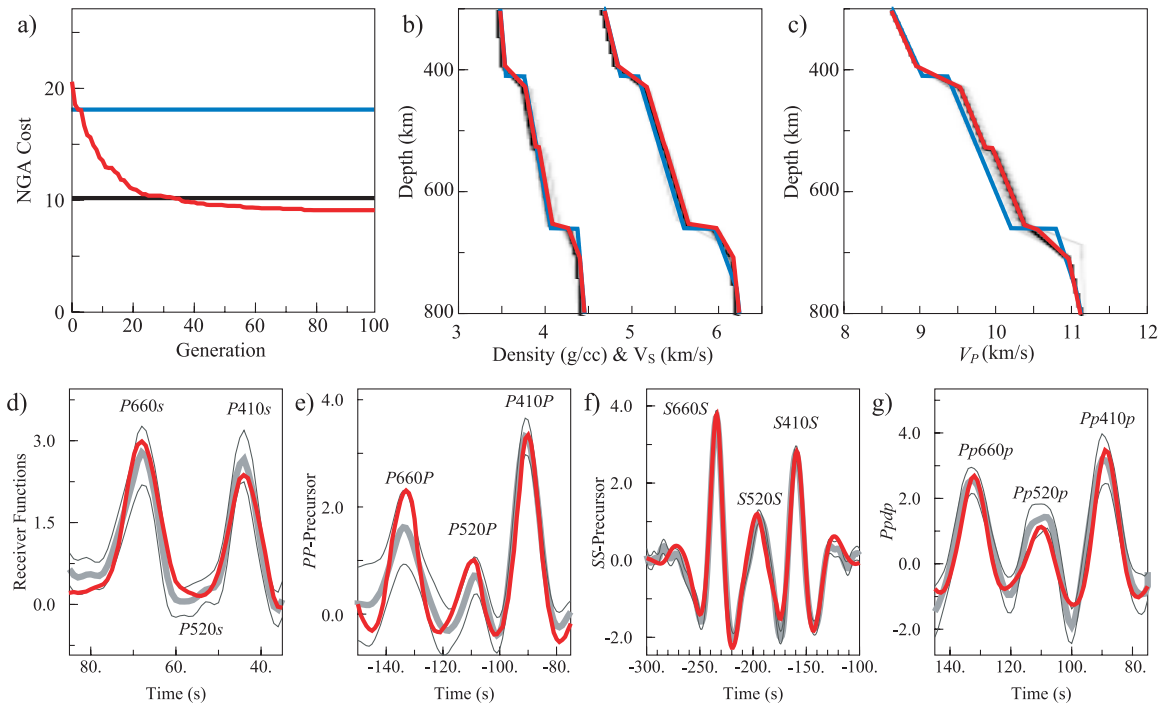
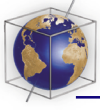


Figure 10. Similar to Figure 7, but for models with finite width interfaces rather than sharp discontinuities.

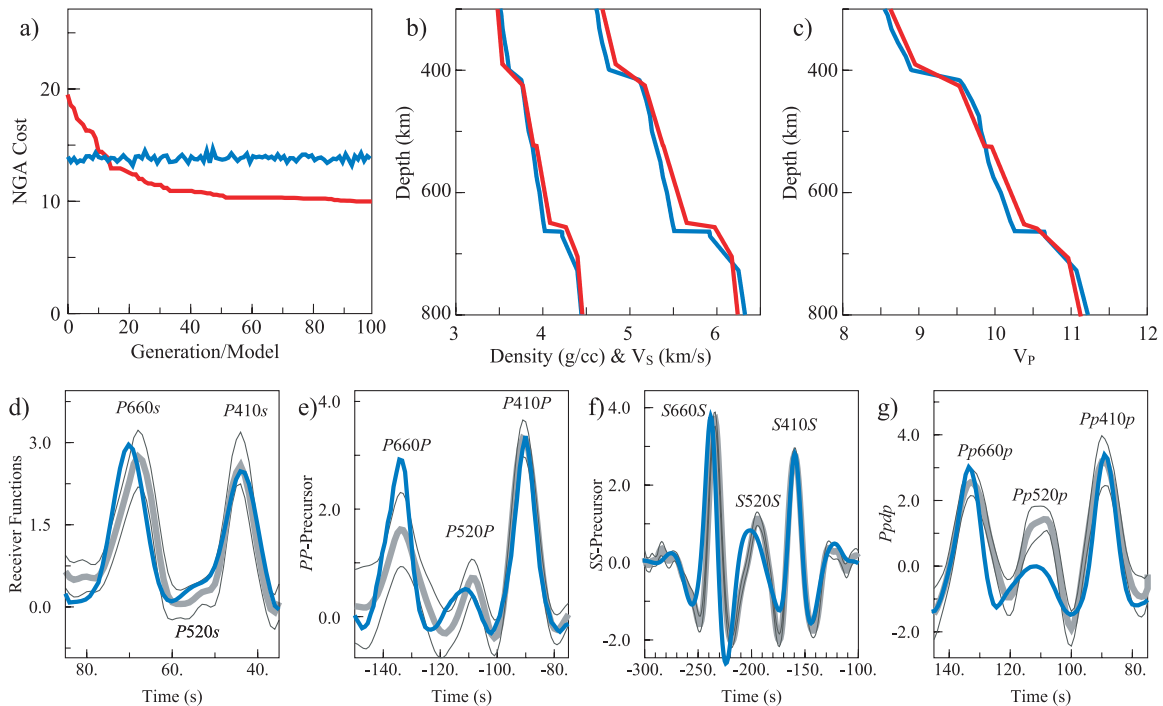


Figure 11. The results shown in Figure 9 (red) compared with the best-fit PREF model of *Cammarano et al.* [2005] (blue). (a) The cost associated with each PREF model compared with the cost associated with the best-fit model after each generation. While the fit to the data (Figures 11a and 11d–11g) is not as good for the PREF synthetics, the best-fit PREF model (Figures 11b and 11c) is very similar to the best-fit model found here.

Table 2. Best Fit Model for Parameterization 1

Discontinuity	V_S , %	V_P , %	ρ , %	Depth, km
410	5.1 ± 0.4	4.8 ± 0.1	4.8 ± 0.2	413.1 ± 1.3
520	0.3 ± 0.3	1.4 ± 0.2	2.7 ± 0.7	515.5 ± 1.4
660	4.2 ± 0.3	0.7 ± 0.8	4.4 ± 0.7	654.2 ± 0.8
Sub-660	-	-	-	715.1 ± 3.7
Peg-depth				

interface thickness. The greater the interface thickness, the lower amplitude and more diffuse the pulse becomes. This parameterization, with $\zeta = 1.0$, results in a 16% reduction in NGA cost due to the greater flexibility provided by the three additional parameters. However, the solution is less certain because of greater parameter tradeoff and bias due to the unknown degree of pulse broadening.

[32] In all of three parameterizations described above several model features remain robust for the most optimal models. (1) The average depths of the 410- and 660-km discontinuities remain constant at 413 ± 1 km and 654 ± 1 km. (2) The sub-660 velocity gradient is steep, extending to a depth of 716 ± 2 km. (3) The shear velocity contrast at the 520-km discontinuity is indistinguishable from zero. (4) The 410-km discontinuity has a larger density contrast than the 660-km discontinuity. In the second parameterization the lesser optimal solution found that the 660-km discontinuity is 10 km deeper, and the steep sub-660 gradient can be replaced by a second interface at ~ 699 km depth. This second interface is largely a V_P contrast, with only minor V_S and density contrasts.

9. Comparison to PREF

[33] The gradual increase in V_S , V_P and ρ through the transition zone is assumed to be largely due to an adiabatic increase in temperature and pressure. The discontinuities likely represent temperature and pressure conditions where phase changes occur.

Table 3. Parameterization 2: Dual 660 Interface

Discontinuity	V_S , %	V_P , %	ρ , %	Depth, km
410	0–12	0–12	0–6	385–435
520	0–12	0–12	0–6	495–545
1st 660	0–6	0–6	0–6	635–685
2nd 660	0–7	0–6	0–6	1st 660 + 1–60
Sub-660	-	-	-	700–820
Peg-depth				

Table 4. Locally Optimal Model for Parameterization 2

Discontinuity	V_S , %	V_P , %	ρ , %	Depth, km
410	4.6 ± 0.4	5.0 ± 0.1	5.8 ± 0.1	414.2 ± 0.7
520	0.5 ± 0.9	2.1 ± 0.3	2.1 ± 0.7	518.9 ± 2.8
1st 660	5.0 ± 0.4	1.3 ± 0.7	5.8 ± 0.6	663.5 ± 3.4
2nd 660	0.4 ± 0.2	2.5 ± 1.3	0.5 ± 0.2	698.7 ± 10.9
Sub-660	-	-	-	794.6 ± 39.9
Peg-depth				

High-pressure mineral physics experiments show that phase changes can produce rapid changes in elastic moduli and density [e.g., *Ringwood, 1975*]. In this section we compare the most optimal seismic models described above with seismic models constrained by geochemistry and mineral physics.

[34] *Cammarano et al. [2005]* provides 99 seismic models constrained by pyrolytic composition that fit ISC P - and S -wave travel times and fundamental spheroidal and toroidal modes as satisfactorily as AK135. These 99 physical reference models (PREF) fit the travel times and fundamental modes best from 100,000 models used in a Monte Carlo type inversion that varied 70 mineral physics parameters. In general these models have lower velocities above the transition zone, a larger jump near the 410-km discontinuity, lower gradients in the transition zone, and high gradients below the 660-km discontinuity than AK135. We compare these 99 seismic models with the stacked waveforms and our lowest misfit model (Figure 11). The pyrolytic model fits the stacked waveforms better than AK135, but worse than our most-optimal models. Because the pyrolytic models lack a 520-km discontinuity, phases associated with this discontinuity are missing in the synthetic waveforms. For all waves, the 410-km discontinuity phases are matched well. While the amplitudes of the 660-km discontinuity phases are better than those of AK135, the times of these phases are off because the depth of the interface is greater (~ 665 km).

[35] There are necessary differences between the pyrolytic models and those presented here due to their respective constraints. Our models are not

Table 5. Parameterization 3: Finite Width Interfaces

Discontinuity	V_S , %	V_P , %	ρ , %	Depth, km	Width, km
410	0–12	0–12	0–6	385–435	0–40
520	0–12	0–12	0–6	495–545	0–40
660	0–12	0–12	0–6	635–685	0–40
Sub-660	-	-	-	700–820	
Peg-depth					

Table 6. Best Fit Model for Parameterization 3

Discontinuity	V_S , %	V_P , %	ρ , %	Depth, km	Width, km
410	7.0 ± 0.2	6.1 ± 0.4	6.9 ± 0.6	412.9 ± 4.1	30.7 ± 8.2
520	0.2 ± 0.3	1.5 ± 0.3	1.9 ± 0.4	521.1 ± 3.1	4.3 ± 5.9
660	5.7 ± 0.4	2.1 ± 1.2	4.7 ± 0.5	653.4 ± 2.1	11.5 ± 6.8
Sub-660 Peg-depth	-	-	-	717.2 ± 12.3	

constrained by fundamental modes, and are only constrained to P - and S -wave travel times through similarity to AK135. Additionally, the gradients above, within, and below the transition zone are set to AK135, so the pyrolitic models are outside the model space explored here. The pyrolitic models are limited by their pyrolitic composition and poorly constrained mineral physics parameters. The suite of models presented by *Cammarano et al.* [2005] have no 520-km discontinuity, which limits their appropriateness.

[36] Despite differences between the PREF models and our models, there are several significant similarities. Both PREF and our models have relatively shallow gradients for the 410 and relatively steep gradients beneath the 660 compared to AK135. Both models indicate that the 410 has greater density and velocity contrasts than the 660, and the 660 is sharp relative to the 410. The fact that mineral physics predicts these observations on the basis of a pyrolitic composition suggests that these may be robust features. The PREF models lack parameterization of a second interface related to the garnet phase transformation, so it is difficult to compare the lesser optimal solution with two interfaces with PREF.

10. Discussion

[37] The forward modeling and simulated inversion techniques used here make some key assumptions. First, the stacks are assumed to be representative of the spherically averaged global structure. However, as shown in Figure 2, the lateral coverage for each phase is quite different, and is generally uneven. Additionally, the stacks are collapsed to two distances for each wave, but the same data are used for each stack, and the amplitudes change with distance for the synthetic waveform calculation. So the inversion process must compromise between the two distances. It is possible that this causes some bias; however our results change only slightly if only one or different distances are chosen. We take lateral variations in discontinuity topography into account using the probability density function of the long-wavelength

model of *Flanagan and Shearer* [1998a] but the transition zone discontinuities are known to vary by different amounts on short-wavelength scales [e.g., *Li et al.*, 2000, 2003]. Consequently, there may be some differences in pulse broadening.

[38] The data used here are all long-period seismograms sampled at 1 Hz. The waveforms were band-limited between 0.2 and 0.02 Hz or 0.1 and 0.01 Hz, which reduces the sensitivity to the sharpness of the discontinuities. At low frequencies the synthetics do not distinguish between steep gradients and sharp interfaces. At higher frequencies, only sharp discontinuities are observed. So future experiments may benefit from similar broadband analyses or multiple band-limited analyses. However, complications due to the frequency dependence of anelasticity, and the lack of signal coherence at higher frequencies may impede such experiments.

[39] The most optimal models for all three parameterizations of the simulated inversion indicate average discontinuity depths of 413 ± 1 km, 517 ± 4 km, and 654 ± 1 km. If the 660 is actually composed of two distinct discontinuities [e.g., *Deuss et al.*, 2006] and no underlying steep gradient, as found by the locally optimal model described in Figure 8 and Table 4, then the discontinuity depths are likely 414 ± 1 km, 519 ± 3 km, 663 ± 3 km, and 699 ± 11 km depth. In this parameterization the transition zone is 249 km thick. Although the absolute depths for the 410, 520, and 660 may be biased by the use of AK135 above and below the transition zone, the relative values should not be significantly biased. While we attempt to correct for 3-D heterogeneity, it is never possible to be certain that the 3-D corrections are completely correct. Because the 3-D corrections must account for P - and S -wave velocity, few 3-D models of mantle velocity are capable of producing both corrections accurately. The P - and S -wave velocity model used here, SB10L18 [*Masters et al.*, 2000], has lower resolution (10°) compared to individual P - [e.g., *Montelli et al.*, 2003] or S -wave models (e.g., SB4L18 [*Masters et al.*, 2000]). However, the resolution is equivalent for both P - and S -wave velocity in SB10L18, so the correc-

tions do not add bias due to uneven resolution. Consequently, the observation of a $\sim 241 \pm 2$ km thick transition zone is robust and consistent with previous stacking studies [Shearer, 1996; Flanagan and Shearer, 1998a; Gu et al., 1998, 2003; Lawrence and Shearer, 2006].

[40] The agreement between the depths of the seismic discontinuities (at approximately 410, 520, and 660 km) and the corresponding pressures of experimentally determined phase changes suggests that the two are directly linked. There is a wide range of reported depths of each discontinuity [Shearer, 2000], which is partially due to lateral variation in topography. The best estimates for the average depths of each discontinuity likely come from the stacking of long-period SS-precursors. Flanagan and Shearer [1998a] found average depths of 418, 515, and 660 km. Gu and Dziewonski [2002] observed averages of 411 and 654 km depth. Gu et al. [2003] inverted for gradual velocity heterogeneity and transition zone discontinuity topography simultaneously, resulting in average discontinuity depths of 409 and 649 km. The differences between these studies likely stems from differences in travel time corrections due to upper-mantle and crustal structure. While the absolute depths are muddled by different corrections from 1-D and 3-D velocity structures, Flanagan and Shearer [1998a], Gu et al. [1998], and Gu et al. [2003] agree on transition zone thickness, 242 ± 2 km. Recently, Lawrence and Shearer [2006] used receiver functions to demonstrate a similar spherically averaged transition zone thickness (242 km).

[41] The amplitudes of the velocity and density contrasts at each discontinuity are much less certain than the topography due to ambiguities in seismic modeling and greater scatter in the amplitude data than in travel time data. Yet constraints on density and velocity contrasts for each discontinuity are important for mineral physics constraints on geodynamic modeling of mass and heat transport between the upper and lower mantle. Therefore these seismic constraints have large implications for global convection. Previous seismic studies provide a range of velocity and density contrasts for the 410 ($\Delta V_P(410) = 5.5 \pm 2.5\%$, $\Delta V_S(410) = 4.9 \pm 0.8\%$) and the 660 ($\Delta V_P(660) = 4.5 \pm 2.5\%$, $\Delta V_S(660) = 6.8 \pm 0.5\%$) [Shearer, 2000]. The elastic contrasts of the 520-km discontinuity are much less certain. Most studies model the density and velocity contrasts at 410, 520 and 660 km with single first order discontinuities, rather than as a gradient, a discontinuity underlain by a

steep gradient, or two distinct interfaces, so these estimates are biased by the assumed geometries.

[42] In this study we obtain velocity and density contrasts for each interface as described by Tables 2, 4, and 6. Because we model multiple seismic phases rather than just one, we have more constraints, which reduces the tradeoff between the velocity and density contrasts. Nevertheless, some tradeoffs still exist. The best-fit model having finite width discontinuities has 410-km discontinuity contrasts of $\Delta V_P = 6.1\%$, $\Delta V_S = 7.0\%$, and $\Delta \rho = 6.9\%$. These values are more similar to the theoretical contrasts for a pyrolitic composition ($\Delta V_P = 5 \pm 0.5\%$, $\Delta V_S = 8 \pm 1\%$, and $\Delta \rho = 4.3 \pm 0.3\%$) [Weidner and Wang, 2000] than those of previous works [Shearer, 2000]. However, the theoretical results can vary significantly for small changes in temperature and composition. Alternatively, the best-fit model having a first-order 410-km discontinuity (Table 2) has lower contrasts for all values ($\Delta V_P = 4.8\%$, $\Delta V_S = 5.1\%$, and $\Delta \rho = 4.8\%$), which are more in line with previous studies [Shearer, 2000]. This study cannot differentiate between a thin and thick 410-km interface due to the long wavelength of the data used here, so nonuniqueness exists. The transition zone likely has 410-km contrasts between those presented in Tables 2 and 6.

[43] The 660-km discontinuity is more difficult to compare with other results due to more complex phase changes associated with both olivine and garnet [Simmons and Gurrola, 2000]. There is likely a steep gradient or curvilinear increase in velocity and density beneath the 660 [Shearer, 1996; Weidner and Wang, 2000], which makes interpretation of contrasts difficult to quantify. Again, theoretical seismic profiles of pyrolitic composition are highly dependent upon temperature and vary even with slight compositional variations [e.g., Cammarano et al., 2005]. Nevertheless, the theoretical seismic profiles of Weidner and Wang [2000] indicate that the discontinuity depth is closer to 650 km than 670 km depth (as seen in PREM [Dziewonski and Anderson, 1981]), and that the rate of change as a function of depth decays toward adiabatic by 700 ± 20 km depth. The results presented here indicate that the V_P contrast at the 660 is smaller than predicted by previous experiments, having the majority of the velocity increase accommodated by a sub-660 gradient. This is most consistent with higher temperature (1900–2100 K) pyrolitic composition at 660-km depth [Weidner and Wang, 2000]. By decreasing the aluminum content from 5% to 3%

and lowering the temperature to 1700 K, the 660 can gain a second interface at ~ 660 km depth [Weidner and Wang, 2000], which lends credence to the double interface model (Figure 8, Table 4) [Deuss *et al.*, 2006]. While this is possible, the mineral physics calculations show that the deeper (garnet) interface should have a larger density contrast than P -wave velocity contrast, compared to the upper interface, which is inconsistent with our observations. Indeed, the lower interface observed here has small contrasts for both V_S and ρ . Additionally, the deeper interface observed here with several less-optimal models is at ~ 700 km, which is not consistent with the Weidner and Wang [2000] result (~ 660 km). Of course, the transition zone varies laterally in both temperature and composition, so the spherically averaged models presented here likely represent an average of all plausible theoretical profiles, not the average condition profile.

[44] The double interface for the 660 presented here agrees marginally with the results of [Deuss *et al.*, 2006], insofar as to show that it is possible to model one interface at ~ 660 km depth and another near ~ 700 km depth. However, having two interfaces seems less likely as a global feature because (1) the distinct double interface model provides worse waveform misfit than the single interface model, (2) the double interface model is only plausible if the already smoothed models of Flanagan and Shearer [1998a] and Masters *et al.* [2000] are damped even further with a low multiplication factor of $\zeta < 0.25$, and (3) the observation of the double interface has only been identified in isolated regions [Deuss *et al.*, 2006]. Therefore we suggest that two distinct interfaces are not likely as global features.

[45] While the 410- and 660-km discontinuities are routinely observed with refraction experiments, these experiments often fail to observe a 520-km discontinuity [e.g., Cummins *et al.*, 1992; Jones *et al.*, 1992]. Stronger support for the observations of a global 520-km discontinuity comes from observations of long-period reflected phases [e.g., Shearer, 1991, 1996; Revenaugh and Jordan, 1991; Gu *et al.*, 1998; Deuss and Woodhouse, 2001]. Because the 520 is more pronounced in reflected phases than in the refraction of seismic waves, it has been proposed that the bulk of the impedance change occurs in density rather than shear velocity [Shearer, 1996]. This is supported by mineral physics experiments studying the elastic properties of the β - to γ -olivine phase change, where changes on the order of $\Delta V_P = 1\text{--}2\%$, $\Delta V_S = 0.8\text{--}1.5\%$, and $\Delta \rho = 2.5\text{--}3\%$ over as much

as 50 km depth range. In this study we observe a 520-km discontinuity having, $\Delta V_P = 1.2 \pm 0.2\%$, $\Delta V_S = 0.4 \pm 0.4\%$, and $\Delta \rho = 2.1 \pm 0.8\%$, which is roughly consistent with mineral physics results considering that the discontinuity is modeled as a thin interface rather than a 50-km thick nonlinear phase transition.

[46] The widths of the 410, 520, and 660 presented in Table 6 are likely upper bounds. To reduce these bounds one simply needs to increase the Gaussian filter width used to model transition zone topography and 3-D velocity heterogeneity. Short period (1 Hz) underside reflections from $P'P'$ precursors are observed for the 410 and 660, but not for the 520 [e.g., Benz and Vidale, 1993; Xu *et al.*, 1998, 2003], which signifies that the 410 and the 660 may be sharp (< 5 km) interfaces while the 520 is more gradual. However the 410 is observed less consistently with 1 Hz $P'P'$ precursors than is the 660 [Xu *et al.*, 2003], which suggests that the 410 may be more gradual. While we cannot constrain the actual thickness of either interface due to unknown broadening resulting from 3-D heterogeneity, the 410 appears to be ~ 3 times thicker than the 660. Additionally, the larger 410 thickness presented here is similar to that expected for a pyrolitic composition [Helffrich and Bina, 1994; Strixrude, 1997; Cammarano *et al.*, 2005]. It is possible that lateral variations in temperature, composition, and water concentration change the sharpness and shape of the 410 such that there are both gradients and sharp interfaces with different strengths in different locations [Xu *et al.*, 1998, 2003]. The lack of $P'P'$ precursor observations associated with the 520 may simply reflect the low impedance contrast and general difficulty in observing the 520.

[47] Future application of the methods presented here to regional subsets of the data will likely increase our resolution on the shape, sharpness, and contrasts of the mantle transition zone discontinuities. Rather than stacking highly variable structures, such a study would require less severe Gaussian filters to account for 3-D velocity heterogeneity and discontinuity topography. Individual regions may be shown to have different shapes to their respective elastic profiles. If this proves to be true, complex spherically averaged models such as those presented here may not be appropriate.

11. Conclusions

[48] We have shown that it is possible to explain a variety of apparent discrepancies among transition

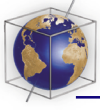
zone phases (*PdP*, *SdS*, *Ppdp*, and *Pds*) with a single seismic model of density, *P* velocity, and *S* velocity. While such models are nonunique depending on a priori constraints, some features are robust. The most optimal models have larger velocity and density contrasts at ~ 413 km than at ~ 654 km. However, a steep sub-660 gradient combined with the 660-km discontinuity provides a greater total density and velocity increase than the 410-km discontinuity. This model has many features in common with PREF models calculated using mineral physics estimates of pyrolitic composition. The 410-km discontinuity likely has a finite thickness that is ~ 3 times greater than of the 660-km discontinuity (not including the sub-660 gradient). The 520-km discontinuity likely has little or no *S*-velocity jump but does possess *P*-velocity and density jumps, which results in a small negative pulse for receiver functions. The zero to negative amplitude of the *P520s* pulse may not be directly observed due to very small amplitude and the interfering effects of the positive *P410s* and *P660s*, but it is required by the data. Consequently, many previous null observations of the *P520s* may have resulted from mistakenly looking for a positive pulse.

Acknowledgments

[49] The data were made available by IRIS DMS and preprocessed using codes written by Guy Masters. We thank the reviewers for thorough reviews. PREF models were downloaded from Fabio Cammarano's Web site: <http://seismo.berkeley.edu/~fabio/PREFum.html>. This research was funded under NSF grant EAR02-29323.

References

- Aki, K., and P. G. Richards (1980), *Quantitative Seismology*, W. H. Freeman, New York.
- Ammon, C. J. (1991), The isolation of receiver effects from teleseismic P waveforms, *Bull. Seismol. Soc. Am.*, *81*(6), 2504–2510.
- Bassin, C., G. Laske, and G. Masters (2000), The current limits of resolution for surface wave tomography in North America, *Eos Trans. AGU*, *81*(48), Fall Meet. Suppl., S12A-03.
- Benz, H. M., and J. E. Vidale (1993), Sharpness of upper-mantle discontinuities determined from high-frequency reflections, *Nature*, *365*, 147–150.
- Bina, C. R., and G. Helffrich (1994), Phase transition Clapeyron slopes and transition zone seismic discontinuity topography, *J. Geophys. Res.*, *99*, 15,853–15,860.
- Chevrot, S., L. P. Vinnik, and J.-P. Montagner (1999), Global-scale analysis of the mantle *Pds* phases, *J. Geophys. Res.*, *104*, 20,203–20,219.
- Cammarano, F., A. Deuss, S. Goes, and D. Giardini (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.
- Cummins, P. R., B. L. N. Kennett, J. R. Bowman, and M. G. Bostock (1992), The 520-km discontinuity, *Bull. Seismol. Soc. Am.*, *82*, 323–336.
- Dahlen, F. A. (2005), Finite-frequency sensitivity kernels for boundary tomography perturbations, *Geophys. J. Int.*, *162*, 525–540.
- Davies, D., E. J. Kelly, and J. R. Filson (1971), Vespa process for analysis of seismic signals, *Nature*, *232*, 8–13.
- Deuss, A. J., and J. H. Woodhouse (2001), Seismic observations of splitting of the mid-transition zone discontinuity in Earth's mantle, *Science*, *294*, 354–357.
- Deuss, A. J., S. A. T. Redfern, K. Chambers, and J. H. Woodhouse (2006), The nature of the 660-kilometer discontinuity in Earth's mantle from global seismic observations of PP precursors, *Nature*, *311*, 198–201.
- Dueker, K. G., and A. F. Sheehan (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.
- Durek, J. J., and G. Ekstrom (1996), A radial model of anelasticity consistent with long period surface wave attenuation, *Bull. Seismol. Soc. Am.*, *86*, 144–158.
- Dziewonski, A. M., and D. Anderson (1981), Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356.
- Engdahl, E. R., R. D. van der Hilst, and R. P. Buland (1998), Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *Bull. Seismol. Soc. Am.*, *88*, 722–743.
- Estabrook, C. H., and R. Kind (1996), The nature of the 660-km upper-mantle discontinuity from precursors to the PP phase, *Science*, *274*, 1179–1182.
- Flanagan, M. P., and P. M. Shearer (1998a), Global mapping of topography on transition zone velocity discontinuities by stacking *SS* precursors, *J. Geophys. Res.*, *103*, 2673–2692.
- Flanagan, M. P., and P. M. Shearer (1998b), Topography on the 410-km seismic velocity discontinuity near subduction zones from stacking of *sS*, *sP*, and *pP* precursors, *J. Geophys. Res.*, *103*, 2673–2692.
- Gilbert, H. J., A. F. Sheehan, K. G. Dueker, and P. Molnar (2003), Receiver functions in the western United States, with implications for upper mantle structure and dynamics, *J. Geophys. Res.*, *108*(B5), 2229, doi:10.1029/2001JB001194.
- Gossler, J., and R. Kind (1996), Seismic evidence for very deep roots of continents, *Earth Planet. Sci. Lett.*, *138*, 1–13.
- Grand, S. P., and D. V. Helmberger (1984), Upper mantle cross section from Tonga to Newfoundland, *J. Geophys. Res.*, *76*, 399–438.
- Gu, Y. J., and A. M. Dziewonski (2002), Global variability of transition zone thickness, *J. Geophys. Res.*, *107*(B7), 2135, doi:10.1029/2001JB000489.
- Gu, Y. J., A. M. Dziewonski, and C. B. Agee (1998), Global de-correlation of the topography of transition zone discontinuities, *Earth Planet. Sci. Lett.*, *157*, 57–67.
- Gu, Y. J., A. M. Dziewonski, and G. Ekstrom (2003), Simultaneous inversion for mantle shear velocity and topography of transition zone discontinuities, *Geophys. J. Int.*, *154*, 559–583.
- Gutowski, P. R., and E. R. Kanasewich (1974), Velocity spectral evidence of upper mantle discontinuities, *Geophys. J.R. Astron. Soc.*, *36*, 21–32.
- Helffrich, G., and C. R. Bina (1994), Frequency dependence of the visibility and depths of mantle seismic discontinuities, *Geophys. Res. Lett.*, *21*, 2613–2616.
- Husebye, E., and R. Madariaga (1970), The origins of precursors to core waves, *Bull. Seismol. Soc. Am.*, *60*, 939–952.



- Jackson, I. (1983), Some geophysical constraints on the chemical composition of the Earth's lower mantle, *Earth Planet. Sci. Lett.*, *62*, 143–164.
- Jones, L. E., J. Mori, and D. V. Helmberger (1992), Short-period constraints on the proposed transition zone discontinuity, *J. Geophys. Res.*, *97*, 8765–8774.
- Kennett, B. L. N., (Ed.) (1991), *IASPEI 1991 Seismological Tables*, Res. School of Earth Sci., Aust. Natl. Univ., Canberra.
- Kennett, B. L. N., E. R. Engdahl, and R. P. Buland (1995), Constraints on seismic velocities in the Earth from travel times, *Geophys. J. Int.*, *122*, 108–124.
- Koper, K. D., M. E. Wyssession, and D. A. Wiens (1999), Multimodal function optimization with a niching genetic algorithm: A seismological example, *Bull. Seismol. Soc. Am.*, *89*, 978–988.
- Lawrence, J. F., and P. M. Shearer (2006), A global study of transition zone thickness using receiver functions, *J. Geophys. Res.*, *111*, B06307, doi:10.1029/2005JB003973.
- Lawrence, J. F., and M. E. Wyssession (2006), QLM9: A new radial quality factor model for the lower mantle, *Earth Planet. Sci. Lett.*, *241*, 962–971.
- Li, A., K. M. Fischer, M. E. Wyssession, and T. J. Clarke (1998), Mantle discontinuities and temperature under the North American continental keel, *Nature*, *395*, 160–163.
- Li, X., S. V. Sobolev, R. Kind, X. Yuan, and C. H. Estabrook (2000), A detailed receiver function image of the upper mantle discontinuities in the Japan subduction zone, *Earth Planet. Sci. Lett.*, *183*, 527–541.
- Li, X., R. Kind, X. Yuan, S. V. Sobolev, W. Hanka, D. S. Ramesh, Y. Gu, and A. M. Dziewonski (2003), Seismic observation of narrow plumes in the oceanic upper mantle, *Geophys. Res. Lett.*, *30*(6), 1334, doi:10.1029/2002GL015411.
- Mahfoud, S. W. (1995), Niching methods for genetic algorithms, Ph.D. thesis, Univ. of Ill., Champagne.
- Masters, G., G. Laske, H. Bolton, and A. Dziewonski (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*, *Geophys. Monogr. Ser.*, vol. 117, edited by S. Karato et al., pp. 63–86, AGU, Washington, D. C.
- Montagner, J. P., and B. L. N. Kennett (1996), How to reconcile body-wave and normal-mode reference Earth models?, *Geophys. J. Int.*, *125*, 229–248.
- Montelli, R., G. Nolet, F. A. Dahlen, G. Masters, E. R. Engdahl, and S.-H. Hung (2003), Finite-frequency tomography reveals a variety of plumes in the mantle, *Science*, *303*, 338–343.
- Nguyen-Hai, (1963), Propagation des ondes longitudinales dans le noyau terrestre, *Ann. Geophys.*, *15*, 285–346.
- Okal, E. A., and B.-G. Jo (1990), Q measurements for phase x overtones, *Pure Appl. Geophys.*, *132*, 331–362.
- Revenaugh, J., and T. H. Jordan (1989), A study of mantle layering beneath the western Pacific, *J. Geophys. Res.*, *94*, 5787–5813.
- Revenaugh, J., and T. H. Jordan (1991), Mantle layering from ScS reverberations: 2. The transition zone, *J. Geophys. Res.*, *96*, 19,763–19,780.
- Ringwood, A. E. (1975), *Composition and Petrology of the Earth's Mantle*, pp. 618, McGraw-Hill, New York.
- Ryberg, T., F. Wenzel, A. V. Egorkin, and L. Solodilov (1997), Short-period observation of the 520 km discontinuity in northern Eurasia, *J. Geophys. Res.*, *102*, 5413–5422.
- Ryberg, T., F. Wenzel, A. V. Egorkin, and L. Solodilov (1998), Properties of the mantle transition zone in northern Eurasia, *J. Geophys. Res.*, *103*, 811–822.
- Shearer, P. M. (1991), Constraints on upper mantle discontinuities from observations of long-period reflected and converted phases, *J. Geophys. Res.*, *96*, 18,147–18,182.
- 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. (2000), Upper mantle seismic discontinuities, in *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale*, *Geophys. Monogr. Ser.*, vol. 117, edited by S. Karato et al., pp. 115–131, AGU, Washington, D. C.
- Shearer, P. M., and M. P. Flanagan (1999), Seismic velocity and density jumps across the 410- and 660-kilometer discontinuities, *Science*, *285*, 1545–1548.
- Shearer, P. M., and G. Masters (1992), Global mapping of topography on the 660-km discontinuity, *Nature*, *355*, 791–796.
- Shen, Y., S. C. Solomon, I. T. Bjarnason, and G. M. Purdy (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*, 2530–3527.
- Shen, Y., A. F. Sheehan, K. G. Dueker, C. de Groot-Hedlin, and F. Gilbert (1998), Mantle discontinuity structure beneath the southern East Pacific Rise from P-to-S converted phases, *Science*, *280*, 1232–1235.
- Simmons, N. A., and H. Gurrola (2000), Multiple seismic discontinuities near the base of the transition zone in the Earth's mantle, *Nature*, *405*, 559–562.
- Strixrude, L. (1997), Structure and sharpness of phase transitions and mantle discontinuities, *J. Geophys. Res.*, *102*, 14,835–14,852.
- Vidale, J. E., and H. M. Benz (1992), Upper-mantle seismic discontinuities and the thermal structure of subduction zones, *Nature*, *356*, 678–683.
- Vinnik, L. P. (1977), Detection of waves converted from P to SV in the mantle, *Phys. Earth Planet. Inter.*, *15*, 39–45.
- Vinnik, L. P., G. Kosarev, and N. Petersen (1996), Mantle transition zone beneath Eurasia, *Geophys. Res. Lett.*, *23*, 1485–1488.
- Walck, M. C. (1984), The P -wave upper mantle structure beneath an active spreading centre: The gulf of California, *Geophys. J. R. Astron. Soc.*, *76*, 697–723.
- Ward, S. N. (1978), Long-period reflected and converted upper mantle phases, *Bull. Seismol. Soc. Am.*, *68*, 133–153.
- Weidner, D. J., and Y. Wang (2000), Phase transformations: Implications for mantle structure, in *Earth's Deep Interior: Mineral Physics and Tomography From the Atomic to the Global Scale*, *Geophys. Monogr. Ser.*, vol. 117, edited by S. Karato et al., pp. 215–235, AGU, Washington, D. C.
- Wicks, C. W., and M. A. Richards (1993), A detailed map of the 660-kilometer discontinuity beneath the Izu-Bonin subduction zone, *Science*, *261*, 1424–1427.
- Widmer, R., G. Masters, and F. Gilbert (1991), Spherically symmetric attenuation within the Earth from normal mode data, *Geophys. J. Int.*, *104*, 541–553.
- Xu, F., J. E. Vidale, P. S. Earle, and H. M. Benz (1998), Mantle discontinuities under southern Africa from precursors to $P'P'_{df}$, *Geophys. Res. Lett.*, *25*, 571–574.
- Xu, F., J. E. Vidale, and P. S. Earle (2003), Survey of precursors to PP' : Fine structure of mantle discontinuities, *J. Geophys. Res.*, *108*(B1), 2024, doi:10.1029/2001JB000817.