Journal of Geophysical Research: Solid Earth

Key Points:
- Imaged global variations in elastic properties and density using complementary data sets
- Several normal modes are incompatible with perfectly correlated Vs and density structure
- Results disfavor purely thermal contributions to heterogeneity in the lowermost mantle

Correspondence to:
P. Moulik,
moulik@umd.edu

The relationships between large-scale variations in shear velocity, density, and compressional velocity in the Earth's mantle

P. Moulik[1,2] and G. Ekström[1]

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York, USA, [2]Department of Geology, University of Maryland, College Park, Maryland, USA

Abstract A large data set of surface wave phase anomalies, body wave travel times, normal-mode splitting functions, and long-period waveforms is used to investigate the scaling between shear velocity, density, and compressional velocity in the Earth's mantle. We introduce a methodology that allows construction of joint models with various levels of scaling complexity ($\varrho = $ dln ρ/dln v_S, $\nu = $ dln v_S/dln v_P), in order to detect seismological signatures of chemical heterogeneity. We demonstrate that the data sets considered cannot be fit concurrently with a uniform ν or a positive and uniform ϱ throughout the mantle. The variance reductions to P wave travel times and v_P-sensitive modes are up to 40% higher with our preferred model of anisotropic shear and compressional velocity than the recent anisotropic shear velocity model S362ANI+M, which was constructed assuming a uniform ν throughout the mantle. Several features reported in earlier tomographic studies persist after the inclusion of new and larger data sets; anticorrelation between bulk sound and shear velocities in the lowermost mantle as well as an increase in ν with depth in the lower mantle are largely independent of the regularization scheme. When correlations between density and shear velocity variations are imposed in the lowermost mantle, variance reductions of several spheroidal and toroidal modes deteriorate by as much as 40%. Recent measurements of the splitting of $_0S_2$, in particular, are largely incompatible with perfectly correlated shear velocity and density heterogeneity throughout the mantle. A way to significantly improve the fits to various data sets is by allowing independent density perturbations in the lowermost mantle. Our preferred joint model consists of denser-than-average anomalies (∼1% peak to peak) at the base of the mantle roughly coincident with the low-velocity superplumes. The relative variation of shear velocity, density, and compressional velocity in our study disfavors a purely thermal contribution to heterogeneity in the lowermost mantle, with implications for the long-term stability and evolution of superplumes.

1. Introduction

Over the past 30 years, seismic tomography has been used extensively to make inferences about the physical and chemical state of the Earth's dynamic interior. Several studies have proposed interpretations of imaged velocity anomalies in terms of temperature [e.g., *Humler et al.*, 1993; *Faul and Jackson*, 2005; *Priestley and McKenzie*, 2013; *Dalton et al.*, 2014] or composition [e.g., *Lee*, 2003; *Trampert*, 2004]. Recent tomographic studies have focused on mapping the isotropic shear velocity structure alone [e.g., *Ritsema et al.*, 2011; *Debayle and Ricard*, 2012] or jointly with anisotropic variations in the mantle [e.g., *Kustowski et al.*, 2008; *Lekić and Romanowicz*, 2011; *Moulik and Ekström*, 2014; *French and Romanowicz*, 2014]. However, other physical properties like density [e.g., *Ishii and Tromp*, 1999] and compressional velocity [e.g., *Su and Dziewoński*, 1997; *Masters et al.*, 2000] provide complementary information useful for addressing several questions in the solid Earth sciences.

The interpretation of velocity anomalies in terms of thermal and chemical variations, in particular, has important implications for mantle geochemistry and geodynamics. The different isotopic signatures of mid-ocean ridge basalts (MORBs) and ocean-island basalts (OIBs) provide compelling evidence for geochemically distinct yet accessible mantle reservoirs [e.g., *Zindler and Hart*, 1986; *Kellogg et al.*, 1999; *Hofmann*, 2003]. While mantle convection is driven thermally by Earth's internal heat, its style and vigor could be influenced by large-scale chemical heterogeneities [e.g., *Tan and Gurnis*, 2007]. However, the length scale [*Coltice and Ricard*, 1999] and

distribution [e.g., *Helffrich and Wood*, 2001; *Davies et al.*, 2012] of chemical heterogeneities in the Earth's mantle is still a subject of debate. The lowermost mantle is a potential candidate for a geochemical reservoir either primordial in origin [e.g., *Labrosse et al.*, 2007], an outcome of the interactions with the iron-rich outer core [e.g., *Knittle and Jeanloz*, 1989], or as a possible "graveyard" of dense mineral assemblages sourced from the phase transformation of subducted oceanic crust at the base of the transition zone [e.g., *Christensen and Hofmann*, 1994]. This issue is related, but not equivalent, to the interpretation of slow-velocity superplumes in terms of purely thermal structures like "plume clusters" [e.g., *Schubert et al.*, 2004] versus thermochemical "piles" of denser material [e.g., *Hansen and Yuen*, 1988; *Tackley*, 1998].

The relative behavior of shear (v_S), bulk sound (v_ϕ), and compressional velocities (v_P) can be a diagnostic for chemical heterogeneity as we have a reasonable understanding of their expected behavior due to thermal [e.g., *Agnon and Bukowinski*, 1990; *Isaak et al.*, 1992; *Anderson and Isaak*, 1995] and associated anelastic effects [e.g., *Karato*, 1993; *Karato and Karki*, 2001; *Matas and Bukowinski*, 2007]. The variations in compressional velocity are sensitive almost equally to variations in shear velocity and bulk sound speed in the mantle [e.g., *Masters et al.*, 2000]. Several tomographic studies have attributed the following seismic features to regions of different bulk composition: (i) the superplumes have larger shear velocity anomalies and sharper edges than expected for an isochemical structure [e.g., *Ni et al.*, 2002; *Wang and Wen*, 2007], (ii) anomalies of shear velocity and bulk sound velocity are of the opposite sign (i.e., anticorrelated) in the lower mantle [e.g., *Su and Dziewoński*, 1997; *Masters et al.*, 2000; *Ishii and Tromp*, 2001], (iii) ratios of shear and compressional velocity anomalies ($v = $ dln v_S/dln v_P) are higher [e.g., *Robertson and Woodhouse*, 1996; *Masters et al.*, 2000; *Romanowicz*, 2001; *Brodholt et al.*, 2007], and the spread of their distributions wider [e.g., *Deschamps and Trampert*, 2003] in the lower mantle than expected for thermal heterogeneity alone. The interpretation of these anomalies may require better constraints in light of the discovery of a phase transition from perovskite to postperovskite in the lowermost ∼200 km of the mantle [e.g., *Murakami et al.*, 2004; *Oganov and Ono*, 2004; *Tsuchiya et al.*, 2004]. A few studies have argued that the high v in the lower mantle could be an artifact of "wavefront healing" in traveling waves and not due to compositional heterogeneity [e.g., *Nolet and Dahlen*, 2000; *Malcolm and Trampert*, 2011; *Davies et al.*, 2012]. The recent splitting measurements of v_P-sensitive modes [*Deuss et al.*, 2013] along with travel time data sets [e.g., *Masters et al.*, 2000] can provide robust constraints on the relative behavior of shear and compressional velocities, helping resolve the significance of compositional effects in the lowermost mantle.

In the absence of chemical variations, velocity anomalies should be proportional (i.e., correlated) to temperature anomalies and hence to density anomalies in the mantle. A compositionally distinct reservoir in the deep mantle may be expected to be intrinsically denser than the background material. The constituting dense material could result either from grain boundary diffusion of siderophile elements from outer core into the lower mantle [*Hayden and Watson*, 2007], accumulation of dense eclogitic components from the subducted oceanic crust [e.g., *Christensen and Hofmann*, 1994; *Brandenburg and van Keken*, 2007; *Tackley*, 2011] or from a dense primordial layer formed early during Earth's differentiation [e.g., *Solomatov and Stevenson*, 1993; *Abe*, 1997; *Labrosse et al.*, 2007]. The splitting observations of normal modes provide some of the best constraints on intrinsic density heterogeneity in the mantle. The gravest normal modes, in particular, have substantial sensitivity to density variations due to the effects of self-gravitation [e.g., *Dahlen and Tromp*, 1998]. Previous tomographic studies have arrived at contradictory results regarding density structure in the lowermost mantle. The presence of denser-than-average anomalies at the base of the mantle coincident with the low-velocity superplumes (i.e., anticorrelation) have been reported primarily based on normal-mode splitting data [*Ishii and Tromp*, 1999, 2001]. However, subsequent studies have argued that the data sets available at the time were of insufficient quality for the determination of a density model independently of velocity structure [e.g., *Masters et al.*, 2000]. Others have reported strong sensitivity of the inversions to the starting model [*Resovsky and Ritzwoller*, 1999] or trade-offs between density and velocity structure [*Kuo and Romanowicz*, 2002]. The v_S-ρ trade-offs have also been reported with sampling techniques applied to limited sets of surface wave and normal-mode data [e.g., *Resovsky and Trampert*, 2003]. The recent occurrence of several large ($M_W \geq 7.4$) as well as deep (>100 km) earthquakes has led to numerous measurements of splitting in normal modes [e.g., *Resovsky and Ritzwoller*, 1998; *Deuss et al.*, 2011, 2013], which could provide new constraints on the pattern of density heterogeneity in the lowermost mantle.

The development of models that constrain multiple variables has lagged behind that of shear velocity models due to data limitations, modeling approximations, and paucity of computational resources. The a priori scaling

relationships between elastic parameters, typically used in the construction of shear velocity models [e.g., *Kustowski et al.*, 2008; *Ritsema et al.*, 2011; *Moulik and Ekström*, 2014], are restrictive as they are consistent with only a thermal origin of seismic anomalies [e.g., *Karato*, 1993] and impose similar patterns of heterogeneity for all parameters [e.g., *Robertson and Woodhouse*, 1996]. Deviations from constant scaling factors have been suggested for the compressional velocities [e.g., *Su and Dziewoński*, 1997; *Masters et al.*, 2000] and density [e.g., *Ishii and Tromp*, 1999] in the deep mantle. Moreover, several studies have demonstrated that the isotropic shear velocity structure can contain artifacts if theoretical complexity of anisotropy is not taken into consideration [e.g., *Anderson and Dziewoński*, 1982; *Ekström*, 2011]. *Kustowski et al.* [2008] reported a severe trade-off between the radially anisotropic and isotropic variations in the lowermost mantle, even with the inclusion of several types of diverse data sets. Similar trade-offs have also been reported between density and velocity structure in tomographic inversions that use only normal-mode splitting data [e.g., *Kuo and Romanowicz*, 2002].

Joint analysis of complementary observations is a powerful way to diminish resolution limitations and trade-offs between parameters, allowing us to disentangle subtler signals from the data. We follow this approach to resolve large-scale elastic parameters and discontinuity topographies from various types of seismic data; any residual signal or peculiarities in splitting of the gravest normal modes can then potentially be attributed to density heterogeneities. Several earlier studies have used diverse types of data for joint inversions of shear and compressional velocities [e.g., *Dziewonski and Woodward*, 1992; *Masters et al.*, 2000; *Gu et al.*, 2001; *Ritsema et al.*, 2004]. This approach was also used by *Moulik and Ekström* [2014] who combined long-period full waveforms [e.g., *Woodhouse and Dziewoński*, 1984], phase velocities of fundamental-mode surface waves [e.g., *Ekström et al.*, 1997], body wave travel times [e.g., *Liu and Dziewonski*, 1998] and normal-mode splitting observations [e.g., *Resovsky and Ritzwoller*, 1998] to derive the model S362ANI+M. *Moulik and Ekström* [2014] also demonstrated the utility of normal-mode splitting data for reducing the trade-offs between isotropic velocity and anisotropy in the lowermost mantle. S362ANI+M was constructed without accounting for the additional sensitivity of the mode-splitting data to density structure. The approach of using constant v_P-v_S scaling relationships did not influence the patterns of anisotropic shear velocity in the mantle [*Kustowski et al.*, 2008; *Moulik and Ekström*, 2014; *French and Romanowicz*, 2014]. However, amplitudes of heterogeneity as well as their patterns may be affected by trade-offs when solving independently for density and elastic parameters in the mantle.

Here we extend the work of *Moulik and Ekström* [2014] by investigating the resolution of anisotropic velocities and density using an expanded data set that provides new constraints on these parameters in the Earth's mantle (section 2). We construct joint models of increasing complexity and examine the data fits a posteriori to arrive at our preferred model. This is the first tomographic study to exploit the sensitivities of several types of seismic data to constrain jointly the variations in shear and compressional velocities, radial anisotropy, discontinuity topographies, and density in the Earth's mantle. A detailed analysis of the resolution of radial anisotropy with current data sets was done by *Moulik and Ekström* [2014]; we focus here on mapping density and compressional velocity while building on our earlier conclusions on anisotropic structure. The various types of data used in this study, our joint inversion framework, and the new regularization scheme are outlined in section 3. In section 4, we describe our models of different complexity and show comparisons with other studies. We conclude, in sections 5 and 6, with a discussion of the results.

2. Data

We make use of a variety of data sets that provide sampling over the entire depth range of the mantle. These data sets contain measurements of body wave travel times, normal-mode splitting functions, surface wave dispersion, and long-period mantle and body wave seismograms (Table 1). We have expanded the compilation used in the construction of S362ANI+M [*Moulik and Ekström*, 2014] with absolute and differential travel times of *P* wave phases [e.g., *Liu and Dziewonski*, 1998; *Masters et al.*, 2000], new dispersion measurements at 35 – 250 s of fundamental-mode surface waves [*Ekström*, 2011] as well as the splitting functions of short-period (3 – 10 mHz) normal modes [*Deuss et al.*, 2013]. The forward modeling procedure and crustal correction scheme for the different data sets are outlined elsewhere [*Kustowski et al.*, 2007, 2008; *Moulik and Ekström*, 2014]. Here we briefly summarize the earlier data sets and focus primarily on updates to the modeling scheme and analysis of new data.

Table 1. Data Used in This Study[a]

Number of Surface Wave Phase Anomalies	
Love waves (G1) $T = 35-125$ s	**37,120–93,032**
Love waves (G1–G5) $T = 150-250$ s	**55,592–75,267**
Rayleigh waves (R1) $T = 35-125$ s	**187,090–311,516**
Rayleigh waves (R1–R5) $T = 150-250$ s	**139,394–145,183**
Number of Long-Period Waveforms	
Body waves $T > 50$ s	19,117–22,522
Mantle waves $T > 125$ s	16,440–24,101
Mantle waves $T > 200$ s	939–1,062
Number of Body Wave Travel Times	
S (HRV)	27,660
SS (HRV)	11,695
ScS (HRV)	4,397
ScSScS (HRV)	1,279
SS-S (HRV)	5,671
SS-S (SC)	16,180
ScS-S (HRV)	3,471
ScS-S (SC)	7,902
S-SKS (HRV)	3,671
SKKS-SKS (HRV)	2,232
SS-S410S (HRV)	18,677
SS-S650S (HRV)	18,670
S410S-S650S (HRV)	16,957
P (HRV)	**25,064**
PP (HRV)	**8,733**
PP-P (HRV)	**2,982**
PP-P (SC)	**20,267**
Source (Number of Modes, Splitting Coefficients)	
Spheroidal fundamental	R&R (18, 1191)
	D11 (18, 1278)
Spheroidal overtone	R&R (28, 714)
	D11 (30, 1555)
	D13 (72, 2964)
Toroidal fundamental	R&R (18, 522)
Toroidal overtone	R&R (11, 118)

[a]Body wave travel times indicated by HRV were measured at Harvard, those indicated by SC at Scripps. The sources of the splitting functions are abbreviated as follows: *Resovsky and Ritzwoller* [1998] as R&R, *Deuss et al.* [2011] as D11, and *Deuss et al.* [2013] as D13. The data sets that have been expanded with new measurements compared to the compilation from *Moulik and Ekström* [2014] are marked in bold.

2.1. Body Wave Travel Times

The compilation of body wave residuals in this study comprises multiple sets of *P* and *S* wave travel times measured from long-period waveforms with a dominant period of ∼20 s, collected by other authors and used in previous tomographic studies at Harvard and Scripps. The similar frequency content of the travel times in our compilation allows robust estimation of the relative variations in shear and compressional velocity. The first subset of travel time residuals was measured by cross-correlating the synthetic waveforms from PREM [*Dziewoński and Anderson*, 1981] with the observed seismograms of 1000 earthquakes ($M_W \geq 5.5$) of the period 1989–1994 [*Liu and Dziewonski*, 1998]. Both absolute travel times of phases P, PP, S, SS, ScS, and ScSScS as well
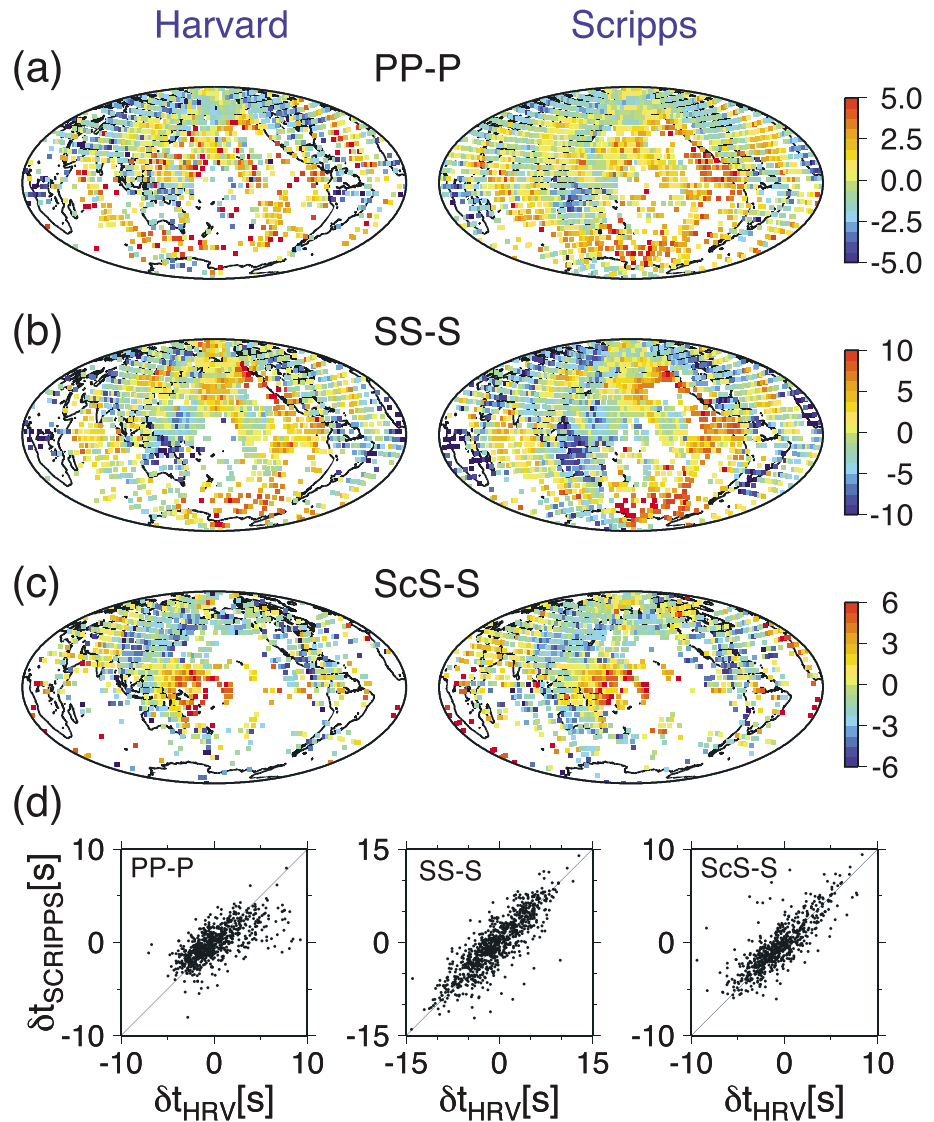
Figure 1. Travel time residuals averaged in 5° by 5° cells and plotted at the source-receiver midpoint. (a) All PP−P data. (b) All SS−S data. (c) All ScS−S data. (d) Scatter plots of the residuals from Harvard (HRV) and Scripps that have similar source-receiver configuration. A baseline shift of 1.1 s was added to the SS−S and PP−P residuals measured at Scripps in Figures 1a, 1b, and 1d, as discussed in section 2. All the travel time residuals are corrected for the effects of CRUST2.0 and Earth's ellipticity.

as differential travel times of two phases such as PP−P, S−SKS, and SKS−SKKS were collected. The second data set consists of SS−S$_{410}$S, SS−S$_{650}$S, and S$_{410}$S−S$_{650}$S differential travel times that provide constraints on the topography of the transition-zone discontinuities [*Gu and Dziewonski*, 2002; *Gu et al.*, 2003]. The third subset of arrival times, measured at Scripps by *Woodward and Masters* [1991] and *Bolton and Masters* [2001], includes differential ScS−S and SS−S travel times. This set also includes ∼20,000 differential PP−P travel times measured from long-period waveforms by *Masters et al.* [2000]. The Scripps data used a different technique involving cross correlation of the observed ScS or Hilbert-transformed SS and PP waveform with the *S* or *P* wave part of the seismogram. While the coverage provided for SS−S and ScS−S data is comparable between the two groups, Scripps's PP−P data are more extensive than Harvard's and have better coverage in the Southern Hemisphere (Figures 1a–1c).

We performed tests for consistency of the PP−P data sets to assess any systematic bias in the measurements; these experiments parallel those performed by *Kustowski et al.* [2008] for the *S* wave data. The ScS−S data sets from Scripps and Harvard were consistent with each other while a constant correction of 1.1 s needs to

be added to Scripps's SS−S data to account for a systematic difference of ∼1.1 s between the two sources of data [*Kustowski et al.*, 2008]. Our calculations of bin-averaged residuals from the PP−P data sets reveal a similar amount of discrepancy for the rays traversing overlapping paths. We therefore apply the same baseline correction of 1.1 s to Scripps's PP−P and SS−S data before including them in our inversions. The corrections remove any systematic bias in the various subsets of data as is evident from the lack of any curvature in scatter plots of the corrected residuals (Figure 1d). We have performed trial inversions to examine the impact of baseline corrections on our results. The uncorrected Scripps's PP−P residuals require faster (∼0.4%) compressional velocities than our reference model at their turning depths in the lower mantle and slower (∼0.2%) velocities in the upper mantle. The baseline correction of 1.1 s added to Scripps's PP−P data reduces dramatically the degree 0 v_P perturbations to within ±0.1% in most parts of the mantle.

2.2. Normal-Mode Splitting Functions

We update the long-period (<3 mHz) compilation used in the construction of S362ANI+M [*Moulik and Ekström*, 2014] with newer measurements of short-period (3–10 mHz) mode-splitting data. The splitting of a normal mode can be treated in isolation (self-coupling) or in resonance with other modes (cross-coupling). The self-coupled splitting functions included in this study provide constraints only on the even-degree components of the long-wavelength heterogeneity. While cross-coupling can provide sensitivity also to the odd degrees [e.g., *Resovsky and Ritzwoller*, 1995], we do not consider the small number of available cross-coupled splitting coefficient observations in our modeling.

The first subset of mode-splitting observations is that of long-period (<3 mHz) normal modes measured by other authors using the iterative spectral-fitting (ISF) technique [e.g., *Giardini et al.*, 1987]. *Moulik and Ekström* [2014] analyzed the long-period measurements from several studies for consistency and selected a subset that was used in the construction of S362ANI+M. This subset consists of splitting coefficient observations from *Deuss et al.* [2011] and *Resovsky and Ritzwoller* [1998] excluding the modes that have substantial sensitivity to the core ($_2S_3$, $_3S_1$, $_3S_2$, $_3S_8$, $_5S_2$, $_5S_3$, $_6S_1$, $_6S_3$, and $_8S_1$) and the ones with low level of consistency between various studies ($_4S_1$, $_4S_2$, $_0T_4$, and $_1T_1$). Several long-period fundamental spheroidal modes (e.g., $_0S_2$, $_0S_3$) and toroidal overtones are sensitive to density variations in the mantle. The modes $_0S_2$ and $_1T_3$, for example, have comparable or marginally higher sensitivity to density than to velocities in most of the lower mantle (Figure 2). Although the long-period modes are mostly sensitive to shear velocity and density, several modes in the fourth and fifth overtone branch ($_4S_3$, $_5S_4$, $_5S_5$, and $_5S_6$) are sensitive largely to compressional velocities in the mantle. The second subset of splitting coefficients consists of 72 short-period (3–10 mHz) spheroidal overtones that have been measured recently by *Deuss et al.* [2013] using the ISF technique. This set includes the splitting functions of 45 short-period v_P-sensitive modes that provide additional constraints on the even-degree compressional velocity structure.

2.3. Surface Wave Dispersion

This data set consists of minor arc and major arc arrivals (R1, R2; G1, G2) of Love and Rayleigh waves from the extensive compilation used in deriving the recent dispersion model GDM52 [*Ekström*, 2011]. The phase anomalies were measured on the seismograms from the years 2000 to 2009 of 3330 shallow ($h < 50$ km) earthquakes ($M_W \geq 5.5$) recorded on 258 stations of the Global Seismographic Network (GSN), Geoscope, the Canadian National Seismograph Network, and other regional networks. The measurement technique for the phase anomalies follows closely the method outlined in *Ekström et al.* [1997], which is based on phase-matched filtering and minimizes the residual dispersion between an observed seismogram and a synthetic fundamental-mode surface wave seismogram calculated using PREM. In this study, we have extended the procedure used by *Ekström* [2011] to measure dispersion of higher-orbit surface waves (R3–R5; G3–G5). While the assumption that rays travel along the great circle connecting the source and receiver may become less valid with increasing path length [e.g., *Woodhouse and Wong*, 1986], our experiments suggest that higher orbits are particularly useful for constraining even-degree structure in the transition zone. Although *Ekström* [2011] measured dispersion at periods down to 25 s, we restrict our analysis to the measurements between 35 and 250 s as the short-period (25–32 s) fundamental-mode surface waves are sensitive primarily to crustal structure. The final compilation consists of measurements of minor arc fundamental-mode Rayleigh and Love waves (R1, G1) at 11 periods between 35 and 250 s along with major arc (R2, G2) and higher orbit (R3–R5; G3–G5) measurements at 150, 200, and 250 s. Owing to the excellent coverage, the uppermost mantle is well sampled globally by these surface wave data and the inversions are largely independent of the regularization scheme.
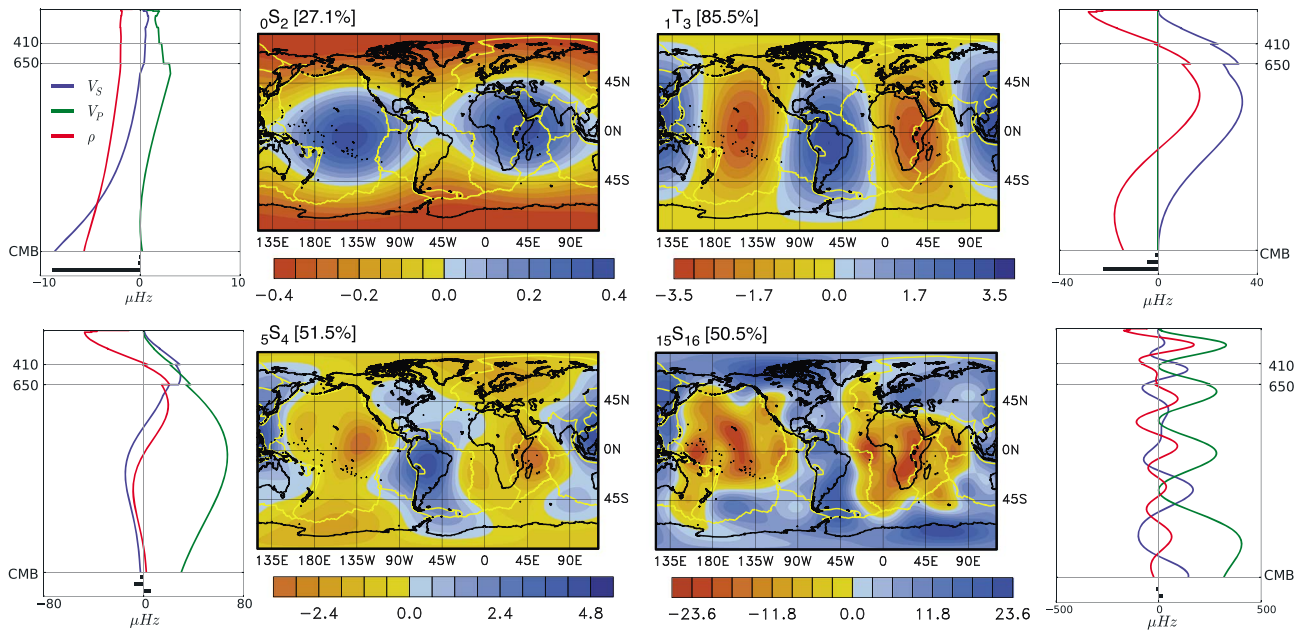
Figure 2. The observed splitting functions of four of the modes used in this study. The variance reductions provided by the recent anisotropic shear velocity model S362ANI+M are specified in the brackets [...]. The data for spheroidal modes $_0S_2$ and $_5S_4$ are from *Deuss et al.* [2011], $_{15}S_{16}$ from [*Deuss et al.*, 2013] while those for the toroidal mode $_1T_3$ are from *Resovsky and Ritzwoller* [1998]. The degree 2 sensitivity kernels, calculated using the 1-D reference model STW105 [*Kustowski et al.*, 2008], are plotted here for density and the Voigt average shear and compressional velocities. Note that anisotropic kernels are employed in the inversions (section 3.1) and the isotropic kernels plotted here are for illustration purposes only. The depth of the 410 km and 650 km discontinuities and the core-mantle boundary (CMB) are indicated by grey horizontal lines. The horizontal bars beneath the velocity kernels show, from top to bottom, the mode's sensitivity to topographic perturbations of the 410 and 650 km discontinuities and the CMB (K^2_{410}, K^2_{650}, and K^2_{CMB}). Note that the kernels are in units of μHz and correspond to variations in $\delta m_i/m_i$ or $\delta h/a$ of 1%, where $a = 6371$ km and that each graph is scaled independently.

Several corrections are applied to the observed phase anomalies ($\delta\phi_{obs}$) for a given source-receiver path before employing the corrected measurements ($\delta\phi_{corr}$) in our inversions. The corrected phase anomalies can be expressed as

$$\delta\Phi_{corr} = \delta\Phi_{obs} - (\delta\Phi_\zeta + \delta\Phi_{ref,ellip} + \delta\Phi_{crust}), \tag{1}$$

where $\delta\Phi_{ref,ellip}$ denotes correction due to different reference models and Earth's hydrostatic ellipticity, $\delta\Phi_{crust}$ denotes correction due to crustal structure, while $\delta\Phi_\zeta$ represents the correction for azimuthal variations in surface wave phase slowness (Appendix A). The corrected phase anomalies can then be attributed to variations in isotropic phase velocity (c) as

$$\delta\Phi_{corr} = -\frac{\omega}{c_0} \int\limits_{path} \frac{\delta c}{c_0}\, ds, \tag{2}$$

where the subscript "0" denotes the values from the reference model and is derived using the approximation $1/(1 + \delta c/c_0) \approx 1 - \delta c/c_0$ that is valid for small perturbations in phase velocity ($\delta c/c_0$).

2.4. Mantle and Body Wave Waveforms

The long-period waveforms provide one more type of constraint on mantle heterogeneity, especially due to the overtone information contained in the body wave part of the seismograms. The waveform data in this study are taken from the compilation of *Kustowski et al.* [2008] and include seismograms from the years 1994 to 2003 of 219 well-recorded earthquakes (M_W 6.5 – 8) and 10 great earthquakes (M_W > 8). The waveform fitting is restricted to very long periods (T > 200 s) for the great earthquakes. Good coverage is obtained by selecting globally either the best-recorded events or earthquakes with different depths or magnitudes. The synthetic waveforms are constructed by the superposition of normal modes while accounting for the sensitivity of local eigenfrequency of a mode to the underlying structure. We employ the path average approximation [*Woodhouse and Dziewoński*, 1984] in our inversions instead of the more sophisticated numerical [e.g., *Komatitsch and Tromp*, 2002a, 2002b; *Liu and Tromp*, 2008] and normal-mode approaches [e.g., *Li and Romanowicz*, 1995] as it allows for a very efficient calculation of synthetic seismograms in aspherical Earth models. The potential implications of this and other approximations for our results are discussed in section 5.3.

3. Methods

3.1. Parameterization

The lateral variations in our tomographic model are expressed in terms of perturbations to the 1-D reference model ($m_k^0(r)$) STW105 at various depths in the mantle, defined as

$$\frac{\delta m_k(r, \theta, \phi)}{m_k^0(r)} = \sum_i \sum_j c_{ij} B_i(r) S_j(\theta, \phi), \tag{3}$$

where m_k stands for four anisotropic parameters and density in the mantle while we also solve for the 2-D topography of the 410 km and 650 km discontinuities. Here c_{ij} is the coefficient of the model parameter corresponding to the ith radial basis function B_i and the jth lateral basis function $S_j(\theta, \phi)$. We describe the radial basis functions in terms of cubic B-splines [*Lancaster and Salkauskas*, 1986] and the lateral basis functions in terms of 362 evenly spaced spherical splines [e.g., *Wang and Dahlen*, 1995], as shown in Figure 3. A projection matrix is used to convert the spline coefficients in the model to spherical harmonic coefficients appropriate for use with the mode-splitting data [*Moulik and Ekström*, 2014, Appendix A].

We solve for independent variations in density $\left(\delta\rho/\rho^0\right)$ as well as radially anisotropic velocities in our inversions. The shear and compressional velocities are parameterized in terms of their isotropic (e.g., $\delta v_S/v_S^0 = 0.5 \times \left[\delta v_{SH}/v_{SH}^0 + \delta v_{SV}/v_{SV}^0\right]$) and anisotropic components (e.g., $\delta a_S = \delta v_{SH}/v_{SH}^0 - \delta v_{SV}/v_{SV}^0$). This parameterization scheme allows us to regularize independently the roughness of isotropic and anisotropic velocity variations. The model vector x for the joint inversions can be written as

$$x = \begin{bmatrix} m_{v_S} & m_{v_P} & m_{a_S} & m_{a_P} & m_\rho & m_{410} & m_{650} \end{bmatrix}^T \tag{4}$$

where m_k ($k = v_S, v_P, a_S, a_P, \rho, 410,$ and 650) represent vectors of spline coefficients c_{ij} describing the variations in isotropic shear velocity $\left(\delta v_S/v_S^0\right)$, isotropic compressional velocity $\left(\delta v_P/v_P^0\right)$, shear anisotropy ($\delta a_S$), compressional anisotropy (δa_P), density $\left(\delta\rho/\rho^0\right)$, and topography of the 410 km and 650 km discontinuities, respectively. We do not solve for lateral variations in the η parameter, which controls the variation of compressional and shear velocities at intermediate incidence angles [*Dziewoński and Anderson*, 1981]. Since most laboratory and tomographic studies of shear and compressional velocities do not account for radial anisotropy in their analysis [e.g., *Masters et al.*, 2000; *Karato and Karki*, 2001; *Matas and Bukowinski*, 2007], we derive Voigt average bulk sound (v_ϕ), shear ($v_{S_{iso}}$), and compressional ($v_{P_{iso}}$) velocities from our anisotropic models [e.g., *Babuska and Cara*, 1991]. These velocities are related to the equivalent isotropic bulk modulus (κ) and shear modulus (μ) by

$$\kappa = \rho v_\phi^2 = \rho\left(v_{P_{iso}}^2 - \frac{4}{3}v_{S_{iso}}^2\right), \quad \mu = \rho v_{S_{iso}}^2. \tag{5}$$

The relative behavior of heterogeneities is then described by the logarithmic ratios

$$\nu = \frac{\delta \ln v_{Siso}}{\delta \ln v_{Piso}}; \quad \varrho = \frac{\delta \ln \rho}{\delta \ln v_{Siso}}, \tag{6}$$

where ν and ϱ are the scaling ratios appropriate for the comparisons with other studies.

3.2. Regularization and Inversion

Since the solution vectors for the various parameters (equation (4)) contain model coefficients c_{ij} corresponding to a finite number of basis functions, we apply discrete inverse theory [e.g., *Menke*, 1989] to solve the joint inverse problem. The aim of a typical least squares inverse problem is to minimize the objective function χ^2 that represents the weighted (w_i) sum of squared misfits to all i types of data $\sum_i w_i \chi_i^2$, also expressed as $(Ax - d)^T(Ax - d)$ in matrix notation. Here, d corresponds to the data in section 2, A to their sensitivity to the perturbations in c_{ij}, and both are weighted by the uncertainties in the measurements. Due to imperfect data coverage and measurement errors, and in order to stabilize the inversion, we regularize the tomographic inversions by minimizing vertical ($\mathcal{R}_{v,k}^2$) and horizontal ($\mathcal{R}_{h,k}^2$) gradients for all k parameters. We choose not to minimize the norm of the solution in our inversions as the structure in models regularized by norm damping have a tendency to correlate with uneven data coverage [*Boschi and Dziewonski*, 1999]. Some modifications
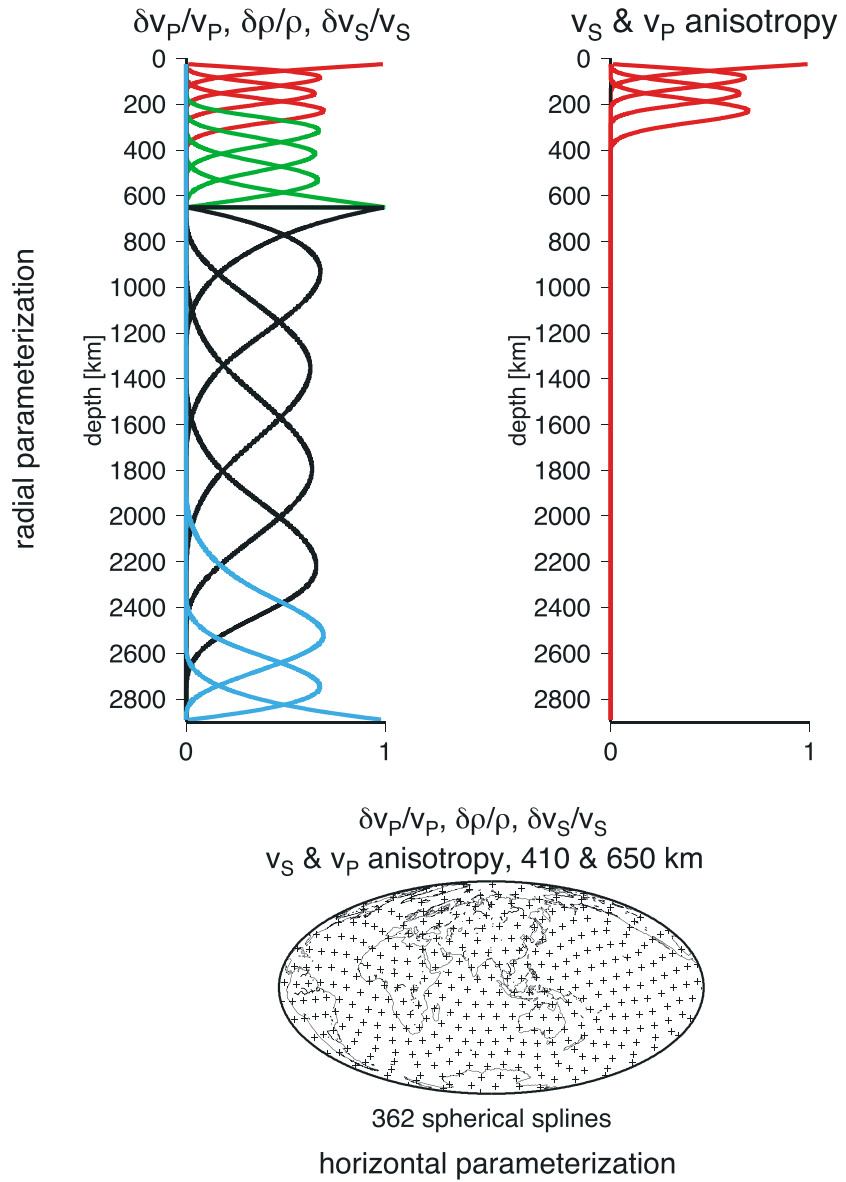
Figure 3. The three-dimensional parameterization of the mantle used in this study. Density and the isotropic shear and compressional velocities are expressed radially in terms of 16 cubic splines (B_i in equation (3)). The top four cubic splines (red) are used to describe the shear wave and compressional wave velocity as well as anisotropy in the uppermost mantle. The rest of the upper mantle including the transition zone is described radially in terms of four splines (green). The lower mantle is parameterized in terms of five splines in the central lower mantle (black) and three splines in the lowermost mantle (cyan). Pluses indicate knots of 362 spherical splines used to describe lateral variations in shear wave velocity, anisotropy, density, and the topographies of the 410 km and 650 km discontinuities (S_j in equation (3)). In case of topographies, value of the radial basis function (B_i) is set to 1.

are made to the modeling approach in *Moulik and Ekström* [2014] to solve independently for the lateral variations in shear and compressional velocity, anisotropy, and density in the mantle. Due to the limited and uneven coverage typically afforded by global data sets, resolution of the various physical parameters needs to be investigated. A new regularization scheme is adopted to impose a priori constraints on the relationship between variations in shear and compressional velocity or density. The objective function for the joint inverse problem is modified to

$$\tilde{\chi}^2 = \sum_i \boldsymbol{w}_i \chi_i^2 + \sum_k \left[\gamma_{h,k} \mathcal{R}_{h,k}^2 + \gamma_{v,k} \mathcal{R}_{v,k}^2 \right] + \gamma_P \mathcal{R}_{R_P}^2 + \gamma_\rho \mathcal{R}_{R_\rho}^2 + \gamma_a \mathcal{R}_{R_a}^2, \tag{7}$$

where the three additional terms $\mathcal{R}^2_{R_P}$, $\mathcal{R}^2_{R_\rho}$, and $\mathcal{R}^2_{R_a}$ impose scaling relationships between various pairs of parameters ($v_S \leftrightarrow v_P$, $v_S \leftrightarrow \rho$, and $a_S \leftrightarrow a_P$) with weights γ_P, γ_ρ, and γ_a, respectively. For example, the regularization term $\mathcal{R}^2_{R_P}$ imposes an a priori scaling factor R_P between variations in average shear (v_S) and compressional (v_P) velocities, given explicitly as

$$\mathcal{R}^2_{R_P} = \int_V \left[\frac{\delta v_P}{v_P^0} - R_P \frac{\delta v_S}{v_S^0} \right]^2 \mathrm{d}V, \tag{8}$$

where the volume integral is evaluated over a region such as the upper, lower, or whole mantle. The recent model S362ANI+M was constructed using constant values of R_P (0.55) and R_a (0.55) while not accounting for density variations in the mantle ($R_\rho = 0$). The equivalent outcome is obtained using the new parameterization and applying $R_P = R_a = 0.55$ with high weights γ_P, γ_ρ, and γ_a in our inversions. It is worth noting that the a priori scaling factors (e.g., R_P) used in the regularization scheme are different than the isotropic scaling ratios (e.g., ν, equation (6)) derived from our anisotropic models; ν would equal the inverse of R_P and ϱ would equal R_ρ in an isotropic medium.

We adopt the same parameterization for all seven variables in the model vector x (equation (4)) in order to examine the influence of regularization schemes on our results. When the same parameterization (equation (3)) is employed for v_P and v_S, the expression for $\mathcal{R}^2_{R_P}$ reduces to

$$\mathcal{R}^2_{R_P} = \left[m_{v_P} - R_P.m_{v_S} \right]^T \left[m_{v_P} - R_P.m_{v_S} \right], \tag{9}$$

in matrix notation. Similar expressions can be written for the regularization terms $\mathcal{R}^2_{R_\rho}$ and $\mathcal{R}^2_{R_a}$ that impose scaling relationships between other pairs of model parameters. The solution to the least squares problem can now be expressed as

$$x_{\mathrm{LS}} = \left[\sum_i w_i (A^T A)_i + \sum_j \gamma_j (D^T D)_j \right]^{-1} \left[\sum_i w_i (A^T d)_i \right], \tag{10}$$

where w_i and γ_j are the weights given to various types of data and damping, respectively. Here i corresponds to the body wave travel times, surface wave phase anomalies, body and mantle waveforms, and the normal-mode splitting coefficients. Also, j corresponds to scaling constraints defined above as well as the vertical and horizontal gradients whose damping matrices are calculated numerically [e.g., *Boschi and Dziewonski*, 1999; *Gu et al.*, 2001; *Kustowski et al.*, 2008]. We then calculate the weighted least squares solution to equation (10) using a standard Cholesky factorization for positive-definite matrices [e.g., *Trefethen and Bau*, 1997]. The weighting scheme for the long-period mode-splitting coefficients, S wave travel times, surface wave phase anomalies, and long-period waveforms is kept the same as the one used for determining S362ANI+M [*Moulik and Ekström*, 2014]. The relative weighting of the additional data sets of short-period mode-splitting coefficients and P wave travel times are decided in a way that maximizes their variance reductions without appreciably reducing the fit to the other types of data.

In order to prevent any bias due to features of previous tomographic models, we choose the spherically symmetric 1-D model STW105 [*Kustowski et al.*, 2008] as our starting model and adopt a computationally intensive two-step inversion scheme (Appendix B). Several results from previous studies are also used to simplify the inverse problem. We solve for anisotropic variations in the shallowest ~300 km of the mantle (Figure 3), as the bulk of the data sets do not require large-scale anisotropic variations in the deep mantle [*Kustowski et al.*, 2008; *Moulik and Ekström*, 2014]. We do not allow independent variations in shear wave and compressional wave anisotropy; high weights γ_a are adopted to ensure constant scaling and the same scaling factors are used for isotropic and anisotropic variations ($R_P = R_a$). The smoothness constraint for transition zone topographies is the same as the one used in the derivation of S362ANI+M. The smoothness damping is varied separately for velocity and density in the upper and lower mantle; this allows incorporation of data limitations such as the sensitivity of the mode-splitting data to only low even-degree variations in density structure (section 3.3).

The various data sets provide different depth-integrated sensitivity to velocity, anisotropy, and density in the Earth's mantle. To illustrate this, we compute the global average of the diagonal elements of the weighted inner-product matrix $A^T A$ as a proxy for the sensitivity at the depth of the spline [*Gu and Dziewonski*, 2002] and plot the values without any normalization [*Kustowski et al.*, 2008] in Figure 4. The shear velocities in the
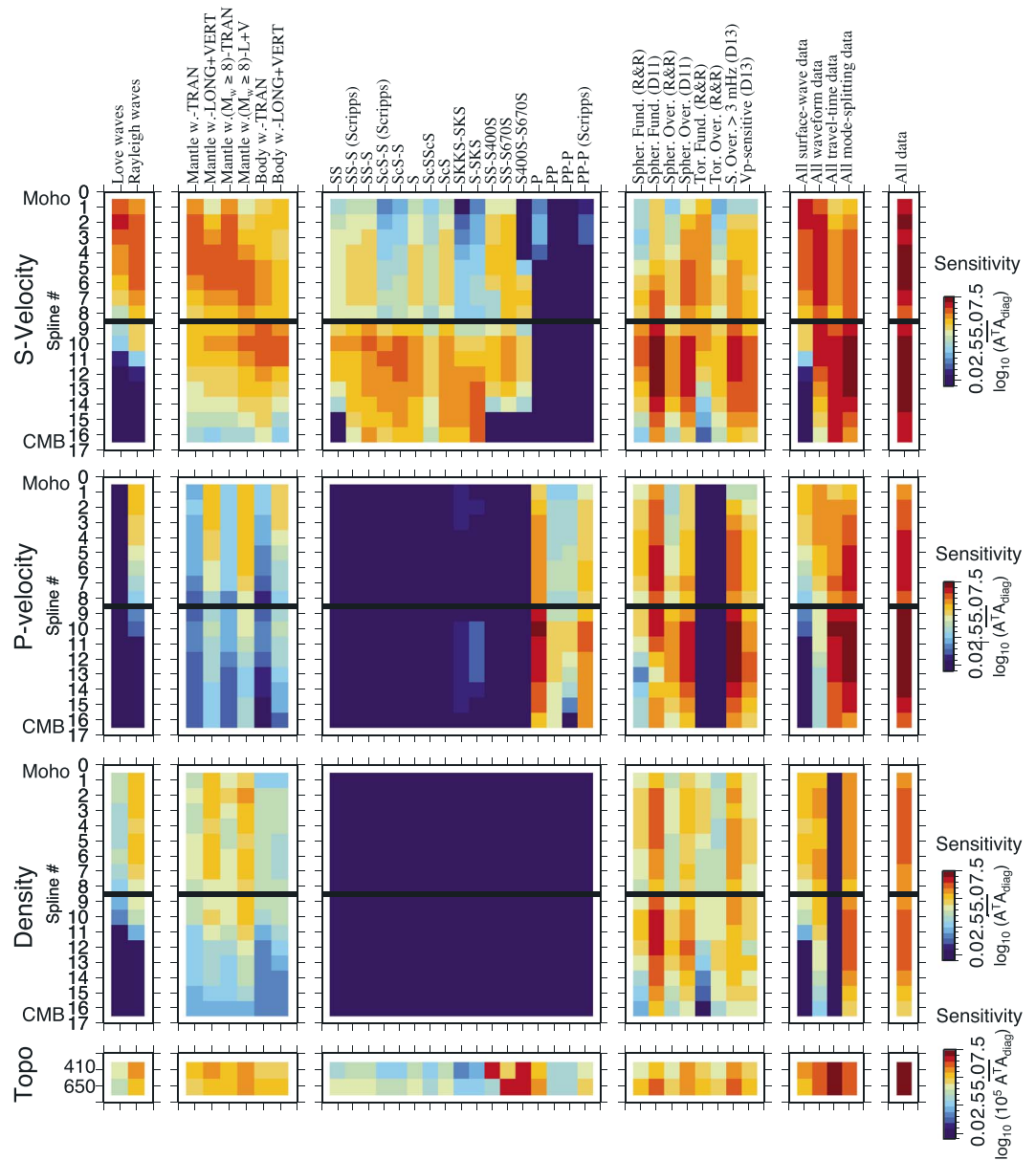
Figure 4. The sensitivity of different data sets is represented by the global average of the diagonal elements of the inner-product matrix $A^T A$ for every cubic spline and for the discontinuities in the transition zone. The $A^T A$ matrices are weighted in the same way as in the inversions for our preferred model (section 4.3.2). The panels on the right show cumulative sensitivities for the surface wave, waveform, travel time and normal-mode data, and for all data combined. LONG, TRAN, and VERT indicate longitudinal, transverse, and vertical components of a seismogram, and "w." denotes waveforms. The sources of the mode splitting functions are indicated as follows: "R&R" to *Resovsky and Ritzwoller* [1998] and "D11" to [*Deuss et al.*, 2011], while "Fund." and "Over." denote fundamental and overtone modes, respectively. Note the different scale for the topography sensitivities.

upper mantle at depths down to 300 km are best constrained by the fundamental-mode surface waves. The long-period Rayleigh waves provide constraints on compressional velocity and density although their peak sensitivity is shallower than that of shear velocity structure [e.g., *Dziewoński and Anderson*, 1981; *Antolik et al.*, 2003]. The sensitivity to the structure from the transition zone down to the midmantle is obtained from overtones, which are incorporated using the long-period waveforms and the overtone mode-splitting data. The sensitivity to density heterogeneity in the lower mantle is provided by the mode-splitting data, albeit only for the even-degree components. The shear velocity in the lowermost mantle is best sampled by horizontally polarized S waves diffracted at the core-mantle boundary, which are sensitive only to variations in v_{SH}.

Topography of the transition-zone discontinuities is determined primarily by travel times of SS precursors while some sensitivity is also afforded by the long-period waveforms.

3.3. Procedure for Adding Scaling Complexity

We obtain scaling relationships from three steps that incrementally add complexity to our models: (i) simple uniform scaling in the whole mantle (section 4.1), (ii) split uniform scaling in the upper and lower mantle (section 4.1), and (iii) depth-varying scaling derived from joint tomographic models (sections 4.2–4.3). The first two steps involve constructing anisotropic shear velocity models along with two (simple uniform scaling: $R_P^{UM} = R_P^{LM}$, $R_\rho^{UM} = R_\rho^{LM}$) or three (split uniform scaling: R_P^{UM}, R_P^{LM}, $R_\rho^{UM} = R_\rho^{LM}$) scaling factors that describe the relative variations of compressional velocity and density in the upper (UM) and lower (LM) mantle. We choose to minimize similarly the vertical and horizontal gradients of shear and compressional velocity ($\gamma_{h,v_P}\mathcal{R}^2_{h,v_P} = \gamma_{h,v_S}\mathcal{R}^2_{h,v_S}$, $\gamma_{v,v_P}\mathcal{R}^2_{v,v_P} = \gamma_{v,v_S}\mathcal{R}^2_{v,v_S}$) as well as anisotropy ($\gamma_{h,a_P}\mathcal{R}^2_{h,a_P} = \gamma_{h,a_S}\mathcal{R}^2_{h,a_S}$, $\gamma_{v,a_P}\mathcal{R}^2_{v,a_P} = \gamma_{v,a_S}\mathcal{R}^2_{v,a_S}$) to avoid the effects of smoothness regularization on their relative behavior in our models. Various combinations of velocity ($R_P = R_a \in [0, 1]$) and density scaling factors ($R_\rho \in [-1, 1]$) are imposed with very high weights γ_P, γ_ρ, and γ_a in our inversions, producing spatially correlated patterns of anomalies for different physical parameters and a laterally invariant ν and ϱ.

In the third step, we construct joint models of density as well as anisotropic shear and compressional velocities by relaxing constraints on uniform scaling ratios imposed during the earlier steps (i and ii) in our modeling. In section 4.2, we construct our preferred v_P-v_S model by relaxing the constraints on uniform ν in the upper and lower mantle. In practice, this involves varying the amount of scaling constraints (γ_P) while keeping smoothness for v_P and v_S the same in all regions of the mantle. This procedure is analogous to imposing similar spatial (i.e., radial and lateral) correlations in v_P and v_S structure by varying the amount of their a priori covariance in our inversions [e.g., *Tarantola and Valette*, 1982]. The scaling constraints are optimized in different regions to disfavor strong variations in ν with depth in the mantle (Appendix B). The procedure to obtain our preferred density model (section 4.3) involves two steps in succession: (1) varying the amount of scaling constraints (γ_ρ) while keeping smoothness for ρ and v_S the same in all regions of the mantle, (2) varying the smoothness of ρ variations ($\gamma_{h,k}$, $\gamma_{v,k}$) using the optimal amount of scaling constraints from step (1). A greater smoothness in ρ relative to v_S structure (10 times) is imposed in our preferred model due to the poorer ability of current data sets to resolve density heterogeneity; elements in the $A^T A$ matrix corresponding to the ρ variations are up to an order of magnitude smaller than those of the v_S variations.

4. Results

We employ the full data set to solve for models with different levels of scaling complexity, constraining first the optimal scaling factors followed by joint models of anisotropic velocities and density. Key questions are then asked to what extent do current data sets constrain scaling factors and how much structural complexity is required in joint tomographic models.

4.1. Optimal Scaling Factors

Figure 5 shows the data fits for anisotropic shear velocity models inverted with simple uniform scaling in the whole mantle. As outlined in section 3.3, various combinations of scaling factors (R_ρ, $R_P = R_a$ in Figure 5f) are imposed in joint inversions. We examine the improvement or deterioration in fits from our earlier anisotropic shear velocity model S362ANI+M ($R_\rho = 0$, $R_P = R_a = 0.55$). Similar patterns in fit for different sources of data (e.g., Figures 5a and 5b) suggest that the whole-mantle scaling factors are resolved well by current data sets. The parameter space of best fitting models (i.e., diameter and aspect ratio of white curves) is indicative of the sensitivity provided by individual data sets. Long-period Love waves (Figure 5g) provide limited constraints on scaling relationships (large diameter, small aspect ratio) while v_P-sensitive modes (Figure 5e) are more sensitive to ν than to ϱ in the whole mantle (small diameter, large aspect ratio). We select the optimal scaling factors based on a detailed analysis of fits to subsets of data as follows. The fits to the long-period normal modes do not improve when the additional complexity of scaling factors is introduced. However, we prefer a uniform velocity scaling R_P of ∼0.5 in the whole mantle so that the fits to spheroidal and toroidal overtones do not degrade from S362ANI+M (Figures 5c and 5d). The strongest sensitivity to compressional velocity is afforded by the use of v_P-sensitive modes and P wave travel times (Figure 5e). The models with the top 5% fits to PP–P travel times and the top 1% fits to v_P-sensitive modes fall in a narrow range of velocity scaling scenarios (Figures 5e–5f). While the short-period v_P-sensitive modes are fit reasonably well by S362ANI+M
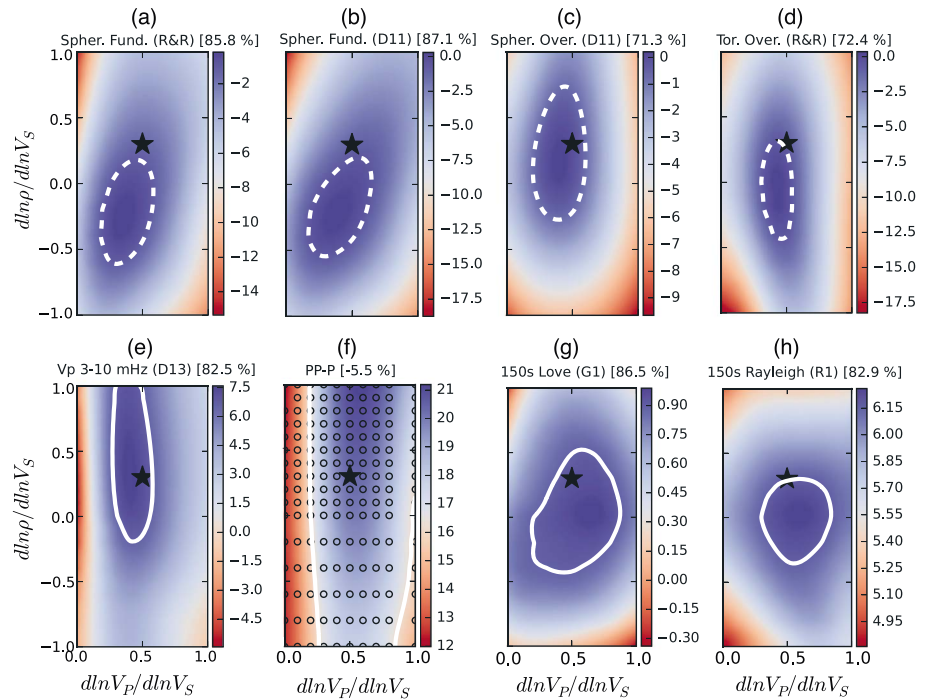
Figure 5. Variance reductions of joint models with two whole-mantle scaling factors. Values are with respect to the variance reductions using S362ANI+M, given in brackets [...]. The fits are calculated for models with different combinations of R_P and R_ρ, specified by the dots in (Figure 5f) and interpolated elsewhere. The white contours outline the best fitting models to mode-splitting data and signify either an improvement (solid) or deterioration (dashed) in fit of 1% relative to S362ANI+M. The white contour for PP−P delineates the models with the top 5% in fits to data. Since the fits to surface waves change marginally in these experiments, their white contours outline models with the top 0.1% in variance reduction (Figures 5g and 5h). The star denotes the preferred values of scaling factors determined by analyzing the overall fits to the data. Note that the color scale varies between different data sets and the sources of data are labeled in a similar way to Figure 4.

(R_P = 0.55) with variance reductions of 82.5%, a further improvement in fit of up to 7.5% is seen with R_P marginally lower than 0.5. The best fits to the PP−P data are observed at values of R_P slightly higher than 0.5, where improvements from S362ANI+M are as much as 21%. Since the long-period surface waves have limited sensitivity to compressional velocities, they are fit equally well by a wider range of R_P (0.4−0.6, Figures 5g and 5h).

Most data sets are fit equally well by a range of density scaling factors (R_ρ) due to the lower sensitivity to density than to velocity structure. The minor variations in fit to PP−P travel times with different R_ρ are primarily due to trade-offs between velocity and density structure (Figure 5f). However, several data sets provide direct sensitivity to density variations in the mantle. The short-period v_P-sensitive modes prefer joint models with positive R_ρ although a very wide range of values fit the data equally well (Figure 5e). The spheroidal overtones prefer positive whole-mantle R_ρ centered around a value of ∼0.3 (Figure 5c). The spheroidal fundamental modes are generally fit better with negative values of R_ρ, although some scenarios of positive values can provide comparable fits (Figures 5a and 5b). However, the bulk of best fitting models to spheroidal fundamental modes prefer anticorrelation between velocity and density in at least parts of the Earth's mantle. This pattern is similar for both sources of spheroidal-mode splitting data and is therefore independent of the subset of data used in our inversions (Figures 5a and 5b). The toroidal overtones also show a clear preference toward negative values of R_ρ, thereby disfavoring a positive and uniform R_ρ throughout the mantle (Figure 5d). Since the data sets have different depth sensitivities (Figure 4), the wide range of best fitting models could suggest a variation in R_ρ with depth in the mantle (section 4.3). As the majority of mode-splitting data are compatible with R_ρ of 0.3, we select it as the preferred value of density scaling for the whole mantle.

Several v_P-sensitive data sets such as the PP−P differential travel times with strong sensitivity at their turning depths in the lower mantle are not well fit by employing a single value of R_P in the mantle. Moreover, travel time data sets have different depth sensitivities to v_P than the mode-splitting and Rayleigh wave data sets.
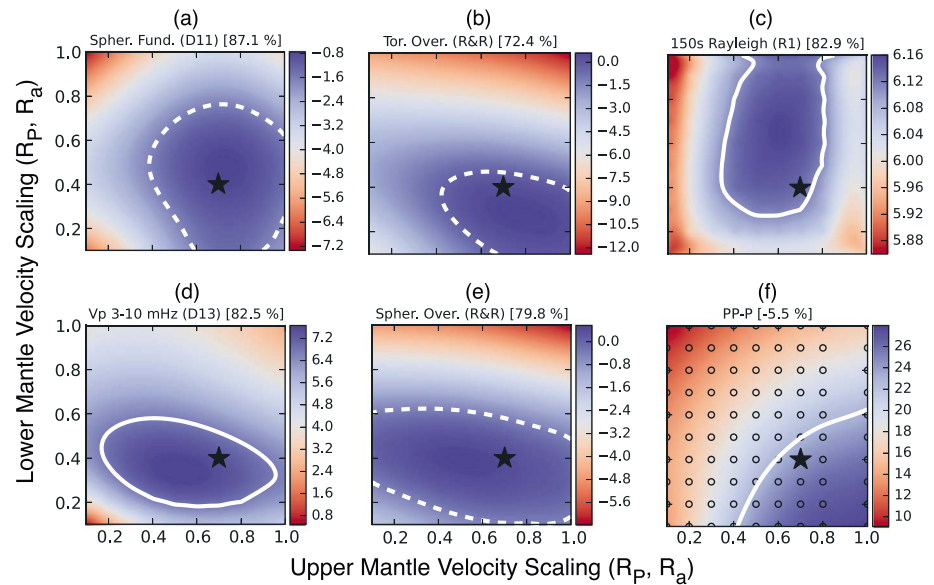
Figure 6. Data fits for joint models with a different upper and lower mantle velocity scaling factor ($R_P = R_a$). The whole-mantle density scaling factor (R_ρ) is constrained to be 0.3 based on the experiments in Figure 5. The variance reductions are calculated for models with different combinations of upper and lower mantle R_P, specified by the dots in (Figure 6f) and interpolated elsewhere. The values in brackets [...], the star, and the white contours are defined as in Figure 5.

A way to improve the fits is by examining split uniform scaling in the upper and lower mantle (section 3.3). We solve independently for a different upper (R_P^{UM}) and lower mantle velocity scaling ratio (R_P^{LM}) while imposing a whole mantle density scaling ratio (R_ρ) of 0.3 in our inversions.

Figure 6 shows the data fits from the different combinations of upper and lower mantle velocity scaling ratios in the range of [0, 1]. The spheroidal fundamental modes prefer high values of R_P^{UM} ($\sim$0.7) but are largely insensitive to scaling factor in the lower mantle (Figure 6a). The long-period surface waves also prefer a high R_P^{UM} with a narrow range of possible values centered around 0.6 (Figure 6c). Several data sets provide sensitivity to the velocity scaling in the lower mantle. For example, both long-period and the shorter-period v_P-sensitive spheroidal overtone modes prefer a lower R_P of 0.4 in the lower mantle (Figures 6d and 6e). The PP–P travel times and toroidal overtones also prefer a low R_P^{LM} along with a high R_P^{UM} though the patterns allow a slightly wider range of possible scenarios (Figures 6b and 6f). Our choice of adding complexity based on data fits and depth sensitivities results in an anisotropic shear velocity model with the associated scaling factors that describe the relative behavior in density ($R_\rho = 0.3$) and compressional velocity ($R_P^{UM} = 0.7$, $R_P^{LM} = 0.4$). The improvements in fit from S362ANI+M with the inclusion of three additional parameters are

Table 2. Suite of v_P-v_S Models Described in Appendix C[a]

Region	Feature	Depth (km)	γ_P-Dependence	△VR
Upper mantle	$R(v_P, v_S)$	24.4–410	⊗	★★
Upper mantle	ν	24.4–410	⊗⊗	★★
Midmantle	ν	650–2000	⊗ ⊗ ⊗	★
Lowermost mantle	$R(v_P, v_S)$	2500–2891	⊗	★★
Lowermost mantle	ν	2500–2891	⊗⊗	★★
Lowermost mantle	$R(v_\phi, v_S)$	2500–2891	⊗	★ ★ ★

[a]R indicates correlation between physical parameters and ν is the scaling ratio defined in equation (6). Weak, medium, and strong dependence of a feature on scaling regularization (γ_P) are indicated by the number of crossed circles. Small ($\leq$2%), medium, and large ($\geq$10%) improvements in variance reduction (△VR) to any subset of data are indicated by the number of stars. A feature is, therefore, well constrained if it is weakly dependent on γ_P (one crossed circle) and leads to large improvements in fit (three stars).
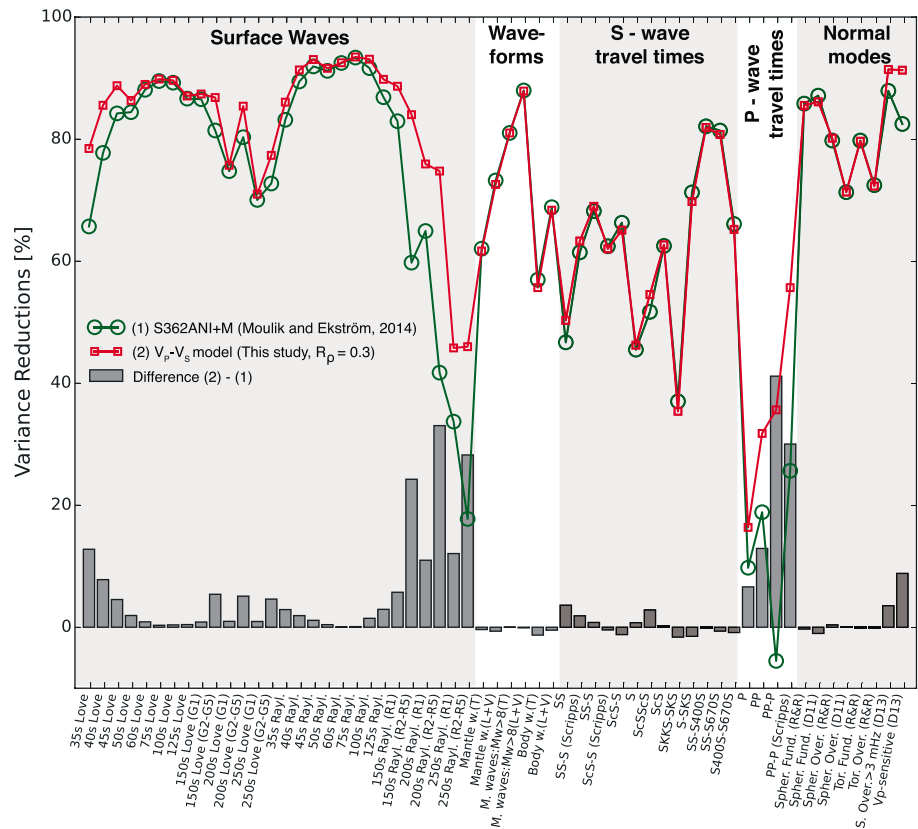
Figure 7. The variance reductions from the joint v_P-v_S model in this study and S362ANI+M [*Moulik and Ekström*, 2014] for different subsets of the data. Variance reductions are calculated based on the difference between the residual variance and the variance in mode-splitting data. Different subsets of the long-period (<3 mHz) mode-splitting data [*Resovsky and Ritzwoller*, 1998; *Deuss et al.*, 2011] are plotted separately; short-period (3–10 mHz) spheroidal overtone data [*Deuss et al.*, 2013] are categorized into two categories based on their sensitivity to v_P structure. The splitting coefficients from both catalogs are used while computing the variance reduction for a common mode. The predictions include the crustal contribution to the splitting functions from CRUST2.0.

as much as 25% for both sources of PP–P travel times (Harvard: −5.5 → 18.8%; Scripps: 25.6 → 50.2%), while the short-period v_P-sensitive modes also show a modest improvement of ∼7.5%. Additional complexity in the form of three scaling factors improves substantially the fit to several v_P-sensitive data sets without deteriorating the fit (>1%) to most other types of data. The SKKS-SKS, S-SKS, and Scripps's SS-S data are slightly worse fit (1–2.5%) with the inclusion of scaling factors, which may be due to trade-offs between shear and compressional velocity in the lower mantle. The overall fits to spheroidal fundamental modes and toroidal overtones also deteriorate by ∼1% from S362ANI+M. Variance reductions for several modes deteriorate dramatically by more than 5% when the three scaling factors are included ($_0S_2$: 43.3%, $_0S_3$: 7.8%, $_0S_9$: 5.2%, $_0S_{10}$: 7.8%, $_0S_{11}$: 5.3%, $_0T_7$: 6.4%, $_0T_8$: 9%, $_0T_9$: 7.6%, $_2T_2$: 13%). The misfit characteristics (Figure 6) vary between different types of data owing to their different depth sensitivities to structural heterogeneity. A constant scaling ratio would, therefore, be mainly representative of the region that is well sampled by the specific type of seismic data.

4.2. Joint v_P-v_S Model
In order to derive depth-varying scaling relationships, we perform a large suite of joint inversions for radially anisotropic shear and compressional velocity structure. As outlined in section 3.3, we use a constant scaling factor for density variations ($R_\rho = 0.3$) in the whole mantle and relax this assumption in the derivation of our v_P-v_S−p model (Section 4.3). The correlation and scaling ratios between v_P and v_S variations in the upper and lowermost mantle are robust (Table 2), while constraints in the midmantle improve when waveform data sets are included (Appendix C).

4.2.1. Fits to Data
Our joint v_P-v_S model fits most of the data sets substantially better than our earlier anisotropic shear velocity model S362ANI+M (Figure 7). Several v_P-sensitive modes are observed to be more weakly split with
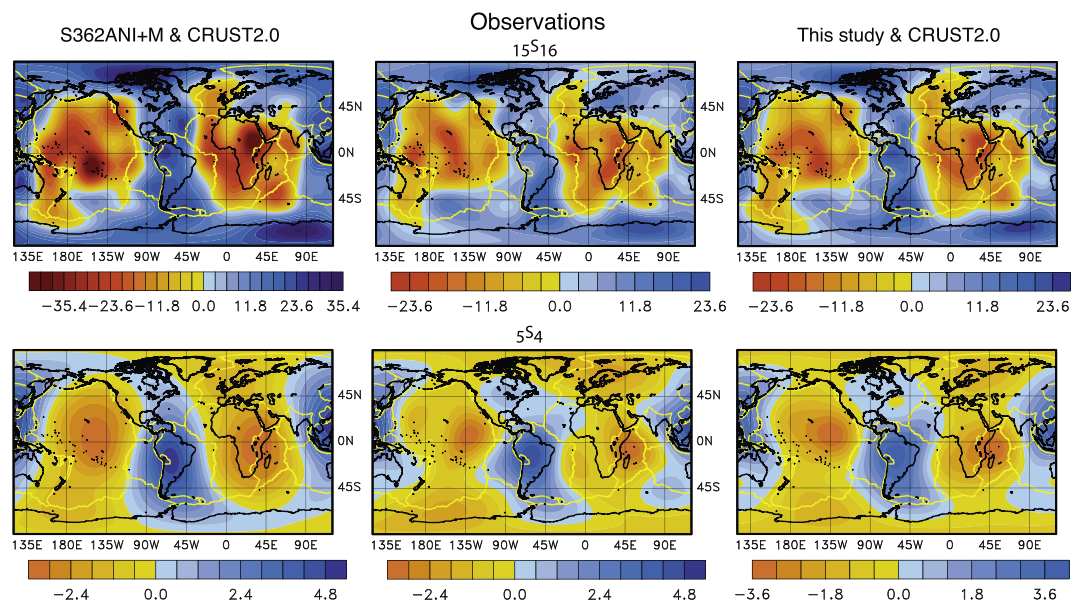
Figure 8. Measurements of splitting functions compared with the predictions by S362ANI+M and the joint model in this study. The measured splitting functions of two v_P-sensitive spheroidal modes [*Deuss et al.*, 2013] are plotted in the middle column. The right column contains prediction from our joint v_P-v_S model and the crustal model CRUST2.0 while the left column contains that of S362ANI+M and CRUST2.0. The crustal splitting function is calculated using a scheme that accounts for all crustal variations and is consistent with the corrections applied to the waveforms [*Moulik and Ekström*, 2014]. All the splitting functions are in units of μHz.

amplitudes of splitting functions substantially lower (~50%) than predicted by S362ANI+M (Figure 8). For example, S362ANI+M predicts splitting that is stronger by up to ~12 μHz for the mode $_{15}S_{16}$, which is sensitive to v_P in the whole mantle (Figure 2). Our analysis suggests that a uniform whole-mantle R_P (0.55) is insufficient to describe the scaling complexity in the mantle; reduced amplitude of v_P heterogeneity in the lower mantle is required by the new mode-splitting data.

Our joint velocity model also gives substantial improvement in fits to several types of v_P-sensitive seismic data compared to S362ANI+M (Figure 7). Scripps's PP–P differential travel times show variance reductions of ~60%, an improvement in fit of as much as 35%. Absolute travel times of the phase PP also show substantial improvements in fit in the range of 25–40%. The variance reduction for the direct P phase is lower (~20%) than that of other data sets; similar fits are given by other low-resolution (degree ~18) models of v_P heterogeneity [e.g., *Antolik et al.*, 2003, *Houser et al.*, 2008]. The short-period (3–10 mHz) mode splitting data show the most substantial improvements in fit of as much as 40% with our joint v_P-v_S model (Figure 9a). Both long-($_4S_3$, $_5S_4$), and short-period ($_{11}S_{10}$, $_{12}S_7$, $_{15}S_{15}$, $_{15}S_{16}$), normal modes that exhibit the largest improvements in fit (≥30%) are sensitive primarily to v_P variations in the mantle (Figure 9). While several long-period v_P-sensitive modes like $_5S_4$ were used in the construction of S362ANI+M, these measurements were fit poorly with variance reductions of less than 60% [*Moulik and Ekström*, 2014]. Better fits to long-period v_P-sensitive modes are likely due to the independent v_P variations in our model and the related reduction in trade-offs between v_P and v_S structure.

The improvements in fit to the v_P-sensitive data does not accompany a substantial deterioration in fit to the remaining types of data that are sensitive primarily to shear velocity variations in the mantle. The S wave travel times (Figure 7) and most of the toroidal fundamental modes (Figure 9a) are fit similarly by both our joint v_P-v_S model and S362ANI+M. The largest improvements in variance reduction (≥20%) are obtained for the major arc and higher-orbit components of long-period Rayleigh waves (R2–R5). The short-period (3–10 mHz) spheroidal overtones with their strong sensitivity to v_S variations are also fit better by our joint model (Figure 9b). These improvements suggest that the relative v_P and v_S variations are adequately captured by our modeling scheme. However, several low-order fundamental modes ($_0S_2$, $_0S_3$, $_0T_8$) and toroidal overtones ($_1T_7$, $_2T_2$, $_2T_4$) show substantial deterioration (5–50%) in fit with our new model (Figure 9a). The mode $_0S_2$ shows the largest reduction in fit of around 50% with our joint model compared to S362ANI+M.
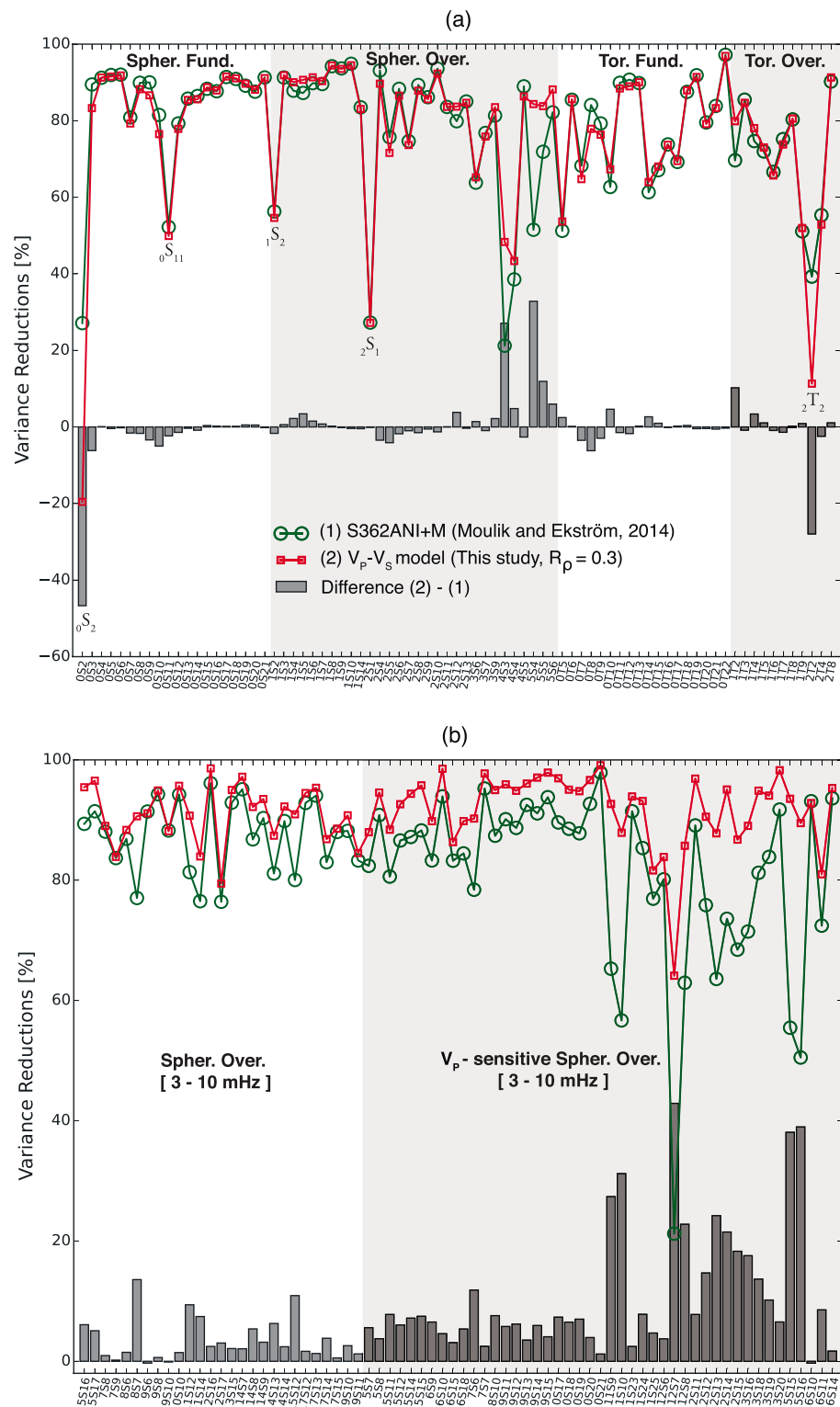
Figure 9. The variance reductions from the joint v_P-v_S model in this study and S362ANI+M [*Moulik and Ekström*, 2014] to the splitting functions of individual modes: (a) long-period (<3 mHz) normal modes [*Resovsky and Ritzwoller*, 1998; *Deuss et al.*, 2011]; (b) short-period (3–10 mHz) spheroidal overtone data [*Deuss et al.*, 2013].
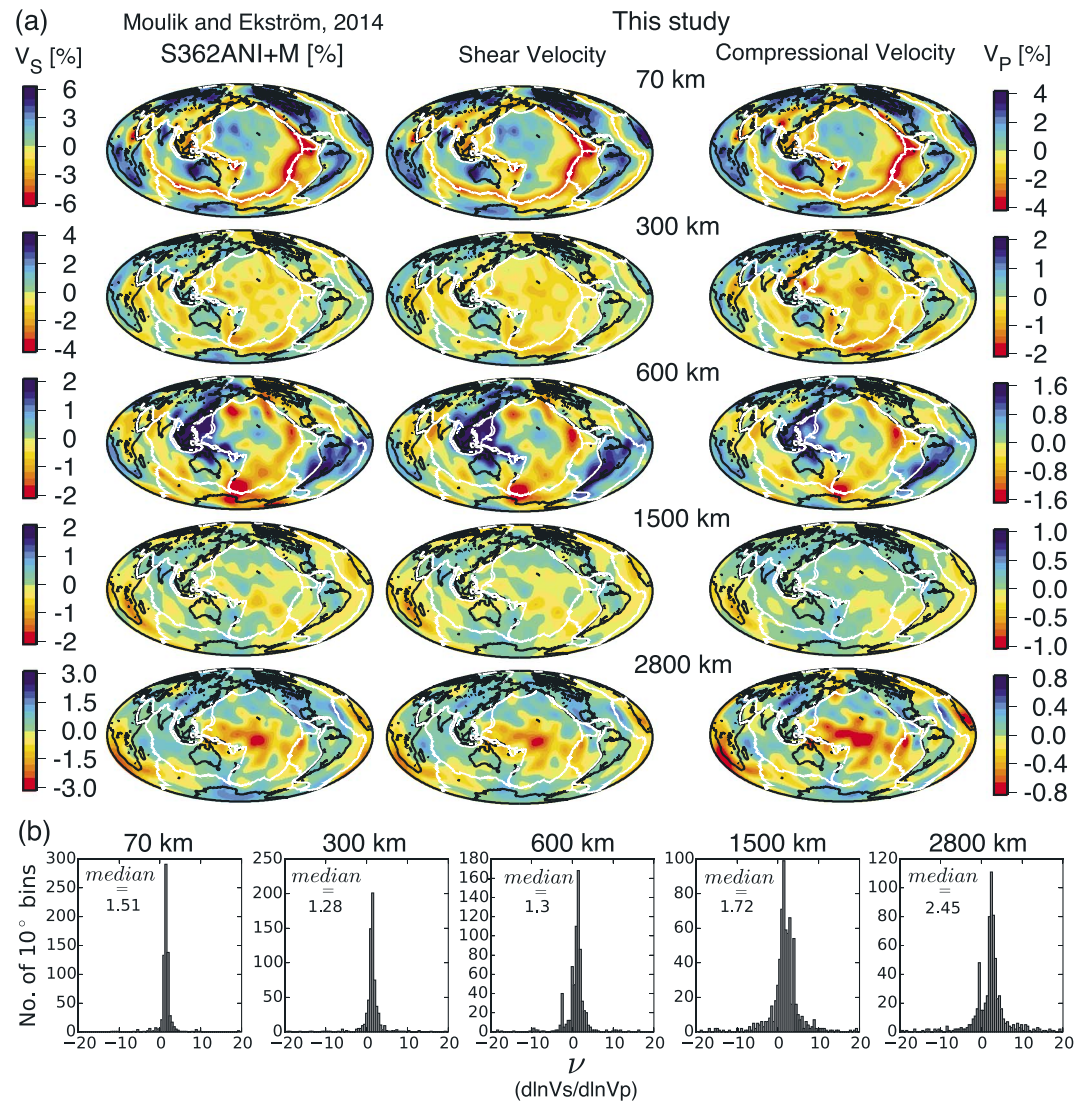
Figure 10. Features of our joint v_P-v_S model: (a) perturbations to the Voigt-averaged isotropic shear wave velocities ($\frac{\delta v_S}{v_S^0}$) at various depths in the mantle from S362ANI+M and our joint v_P-v_S model. Also shown are the perturbations in compressional wave velocity as described in section 4.2.2. Note that the color scale varies between the maps of shear and compressional velocities. (b) Histograms and the median values of ν, calculated from every 10° by 10° pixel with nonzero values of velocity perturbations.

A shared characteristic of all of these modes is their strong sensitivity to density variations in the lower mantle (Figure 2). We created an extensive suite of v_P-v_S models that included scenarios with anisotropic variations throughout the mantle; the anisotropic anomalies in the lowermost mantle obtained from such inversions [*Moulik and Ekström*, 2014] do not improve appreciably the fits to any density-sensitive mode.

4.2.2. Characteristics of v_P-v_S Variations
Figure 10 shows the variations in shear and compressional velocities from our preferred joint v_P-v_S model compared with the shear velocity perturbations in S362ANI+M. The shear velocity variations in the shallowest parts of the mantle are very similar for the models both in their amplitude and pattern ($R \geq 0.9$) of anomalies. The strong v_S perturbations in the shallowest mantle reach amplitudes of over 6% in cratonic regions and mid-ocean ridges. Most of the differences between the shear velocities of the models arise at depths below 300 km. The heterogeneity in the transition zone (410–650 km) is dominated by the faster-than-average anomalies associated with major subduction zones. The shear velocity anomalies in our joint model geographically east of the South American subduction zone are of slightly reduced amplitude and lateral extent than those in S362ANI+M. These differences could be due to the trade-offs between shear and compressional
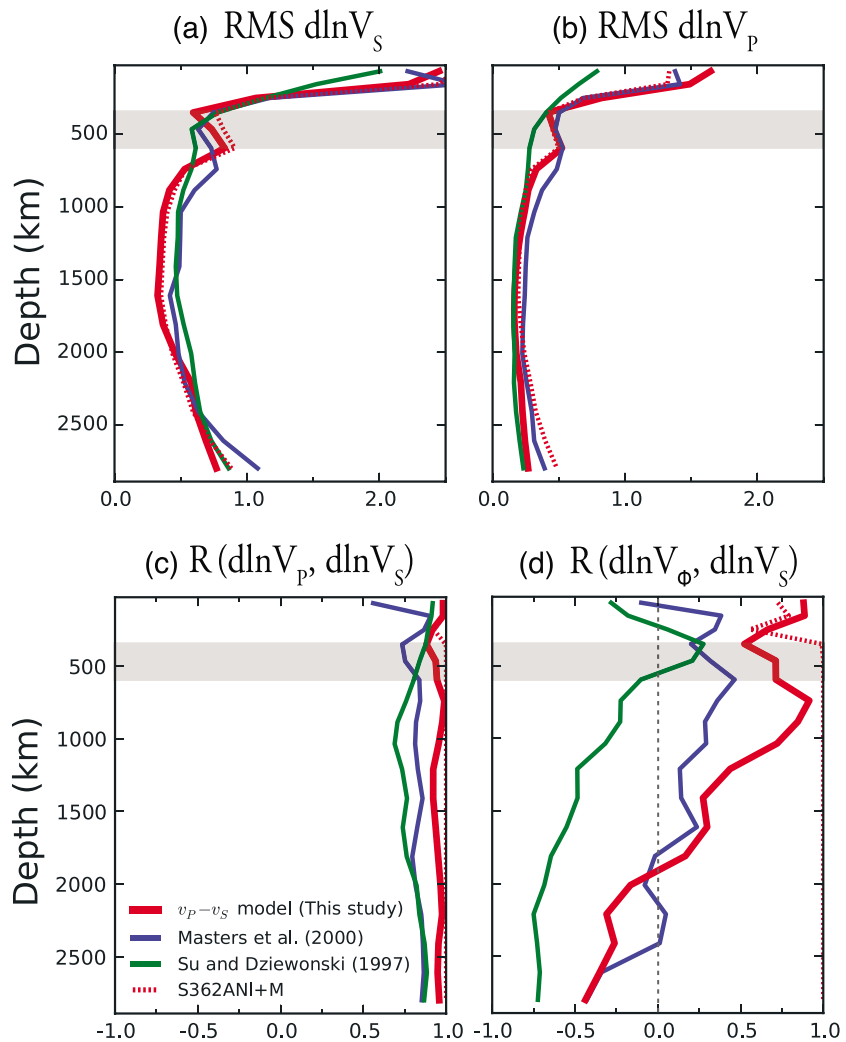
Figure 11. The characteristics of joint v_P-v_S models: (a, b) root-mean-square values of the Voigt average isotropic anomalies; (c, d) correlations between v_P-v_S and v_ϕ-v_S variations. Note that the correlations, denoted here by R, are distinct from the scaling factors R_P, R_P, and R_ρ. The values from the preferred v_P-v_S joint model (section 4.2.2) in red curves are compared to that of S362ANI+M in dotted red, *Su and Dziewoński* [1997] in green, and *Masters et al.* [2000] in blue curves.

velocity structure that are reduced in this study by the use of v_P-sensitive data and solving independently for the v_P and v_S variations. The variations in the midmantle are uniformly low and comparable between the two models. The strength of heterogeneity increases in the lowermost mantle but is weaker than those in S362ANI+M. For example, the amplitude of the low-velocity superplume beneath the Pacific is weaker (∼0.3%) by approximately 10% of the total strength of heterogeneity in S362ANI+M.

The variations in compressional velocities exhibit patterns similar to those of shear velocity in most regions of the mantle. Although the scaling constraints enforce correlations between v_P and v_S heterogeneity, our suite of inversions suggest that the data sets do not require substantial differences in the patterns of their anomalies. The absolute amplitudes of the anomalies in compressional velocity are a factor of ∼2–3 lower than that of shear velocity throughout the mantle. The peak variations in v_P reach values of up to ±4% in the uppermost mantle while the variations in the transition zone are much smaller with values in the range of ±1.6%. The strength of v_P heterogeneity at midmantle depths are the weakest with perturbations of less than ∼0.3%. The v_P anomalies in the lowermost mantle are down to ∼0.8% slower than average in the super-plumes. There is, however, some spatial variability between the locations of the slow perturbations in shear and compressional velocity in the lowermost mantle. The ring of high velocities surrounding the Pacific basin in the lowermost mantle is also more prominent in the v_P variations than in the v_S model.

Figure 11 shows the root-mean-square (RMS) amplitudes of, and correlations between, the variations in isotropic shear and compressional velocities from our preferred v_P-v_S model. The RMS variations in shear velocities are largely independent of the scaling constraints used in the inversion (Figure 11a). Our preferred model has slightly reduced (∼0.1%) shear velocity RMS variations in the transition zone and the lowermost mantle compared to S362ANI+M. However, the strength of shear velocity heterogeneity in the transition zone is still stronger than in the earlier model S362ANI [*Kustowski et al.*, 2008], especially at the bottom of the transition zone [*Moulik and Ekström*, 2014]. The reduction in the amplitudes of shear velocity variations is accompanied by deviations from S362ANI+M of v_P heterogeneity throughout the mantle (Figure 11b). The constant scaling factor R_P (0.55) used in the construction of S362ANI+M results in a stronger v_P heterogeneity in the lowermost mantle than is compatible with our v_P-sensitive data sets. The RMS variations of compressional velocities are ∼0.25% in the lowermost mantle, almost half of that predicted by the scaling factor in S362ANI+M. The RMS strength of the variations in compressional velocity changes less with depth in the mantle compared to the RMS variations in shear velocity. However, the v_P heterogeneity is in fact stronger in the transition zone and uppermost mantle compared to S362ANI+M. Strong v_P heterogeneity in the uppermost mantle is expected from the low values (∼0.7) of the upper mantle scaling factor R_P that are preferred by the long-period surface wave data (Figure 6c). The correlations between the variations in isotropic velocities show two features that are largely independent of the regularization scheme: high correlation between variations in shear and compressional velocity (Figure 11c) and positive correlation followed by anticorrelation in the upper and lower mantle, respectively, between the variations in shear and bulk sound velocity (Figure 11d). The v_S-v_ϕ anticorrelation in our model exists only in the lowermost 800–900 km of the mantle, roughly consistent with the results from *Masters et al.* [2000].

4.3. Density Variations
4.3.1. Sign of Density Anomalies in the Lowermost Mantle
A simple uniform scaling for density in the whole mantle ($R_\rho = 0.3$) may be incompatible with the fundamental spheroidal mode data in general (section 4.1) and with several long-period modes, in particular, (section 4.2.1). A way to improve fits to splitting functions of the gravest normal modes is to add the complexity of independent variations of density in parts of the Earth's mantle. The structure in the lowermost mantle, in particular, could deviate from a constant R_ρ and help explain low fits to some of these modes. We perform an experiment to evaluate whether the current data sets can discriminate between correlation ($R \geq 0$), decorrelation ($R \sim 0$) or anticorrelation ($R \leq 0$) between shear velocity and density in the lowermost mantle. We vary the scaling constraints between shear velocity and density γ_ρ while keeping smoothness for ρ and v_S the same in all regions of the mantle. The regularization for the v_P variations is kept the same as the optimal scheme found from our analysis in section 4.2.2. We tested different regularization schemes for the v_P variations (γ_P) and obtained similar results on v_S-ρ correlation described below.

An unconstrained inversion ($\gamma_\rho = 0$) for density variations in the lowermost mantle leads to a model with strong denser-than-average anomalies roughly coincident with the large-scale slow-velocity superplumes beneath the Pacific and Africa (Figure 12b). The denser-than-average anomalies in our inverted models are as much as 2% in the Central Pacific. The fit to the $_0S_2$ splitting function deteriorates substantially when the models are constrained to have positive v_S-ρ correlations with increasing amount of damping γ_ρ (Figure 12a). The variance reduction worsens by as much as 40% compared to the unconstrained model ($\gamma_\rho = 0$) when a constant correlation ($R_\rho = 0.3$) is imposed using a high γ_ρ. Several other modes show similar reduction in fit with increasing correlation between the variations in density and shear velocity (Figure 12d). These results favor anticorrelation between the patterns of shear velocity and density in the lowermost mantle and are in broad agreement with the conclusions of *Ishii and Tromp* [1999]. Discriminating between v_S-ρ decorrelation and strong anticorrelation or evaluating the amplitude of density heterogeneity requires a more detailed set of inversions summarized below.

4.3.2. Preferred Density Model
We construct a model that allows independent variations of density in parts of the mantle in order to provide acceptable fits to splitting functions of the gravest normal modes. Our experiments suggest that the current data sets do not require independent variations in upper-mantle density structure; we choose to impose the constant scaling factor R_ρ of 0.3 in the upper mantle with high weights γ_ρ and focus on exploring density variations in the lower mantle. High values of the scaling weight γ_ρ are applied in the midmantle to avoid strong lateral variations in density scaling ratio ϱ, which is not expected in this depth range due to the low RMS amplitudes of both anisotropic velocities and density. We progressively reduce the scaling constraint γ_ρ
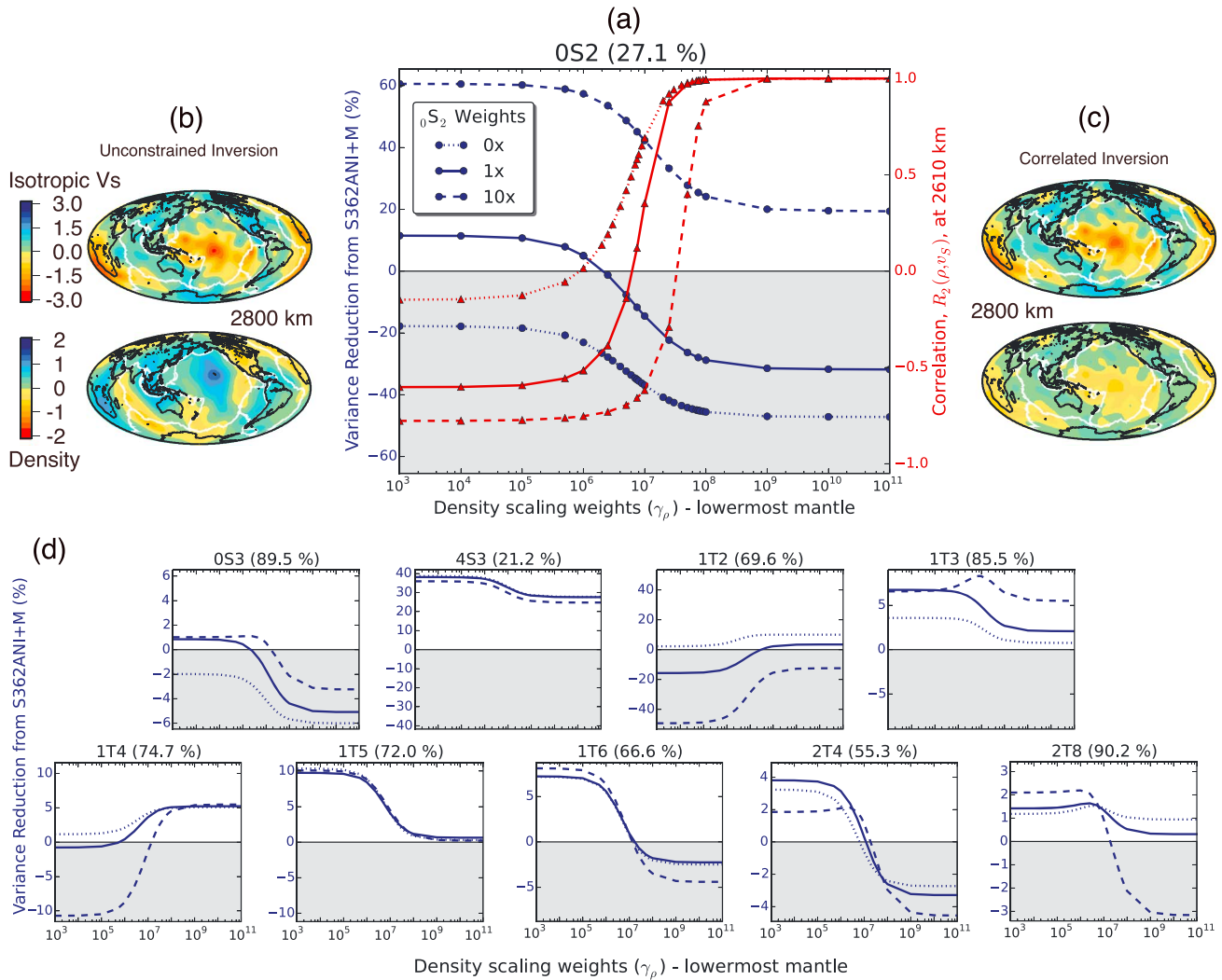
Figure 12. (a) Variation in fits to the $_0S_2$ splitting function for models inverting using different weights γ_ρ for the scaling constraints between shear velocities and density ($R_\rho = 0.3$) in the lowermost mantle (bottom three splines in Figure 3). The curves corresponding to our preferred weights for $_0S_2$ data (solid lines) are compared against the ones that exclude $_0S_2$ data (dotted) or give it 10 times our preferred weight (dashed). The corresponding curves for the degree 2 correlations between shear velocity and density in the lowermost mantle (2610 km) are plotted in red. (b, c) The end-member joint models for shear velocity and density correspond to unconstrained ($\gamma_\rho = 0$) and constrained ($\gamma_\rho = 10^{11}$) inversions. (d) The majority of other modes show a substantial reduction in fit ($_0S_3$, $_4S_3$, $_1T_3$, $_1T_5$, $_1T_6$, $_2T_4$, $_2T_8$) while a couple of them show a slight improvement in fit ($_1T_2$, $_1T_4$) when positive correlation is imposed in the lowermost mantle. Only the modes whose variance reductions change by more than 5% during our experiments in section 4.3.1 are plotted.

with depth in the lower mantle to allow independent variations in regions where we have strong sensitivity to density structure from the gravest normal modes. Since the resolution of density from mode-splitting data is limited to the low and even degrees, we experiment with different amounts of smoothness damping ($\gamma_{v,\rho}\mathcal{R}_{v,\rho}^2$, $\gamma_{h,\rho}\mathcal{R}_{v,\rho}^2$) in the lower mantle to select our preferred density model. In order to glean the density signal from the data, we used higher weights for the ρ-sensitive mode $_0S_2$ that is not fit well by our v_P-v_S model (section 4.2.1); this choice does not affect appreciably the pattern of density anomalies obtained from our inversions (section 5.3).

Figure 13 summarizes the fits obtained for the long-period mode-splitting data using our joint models and S362ANI+M, the preferred model from *Moulik and Ekström* [2014]. The reduction of v_S-ρ correlation in the lower mantle leads to better fits to the modes that were not fit well by our joint v_P-v_S model. Our preferred density model fits several spheroidal fundamental modes and toroidal overtones up to ~50% better than S362ANI+M, thereby reversing the trend of worse fits with our v_P-v_S model (section 4.2.1). The fits to other long-period modes either improve slightly in our density model or remain comparable ($\pm 5\%$) to that of our joint velocity model. The only exception is the toroidal-mode overtone $_1T_2$ that is fit better by our
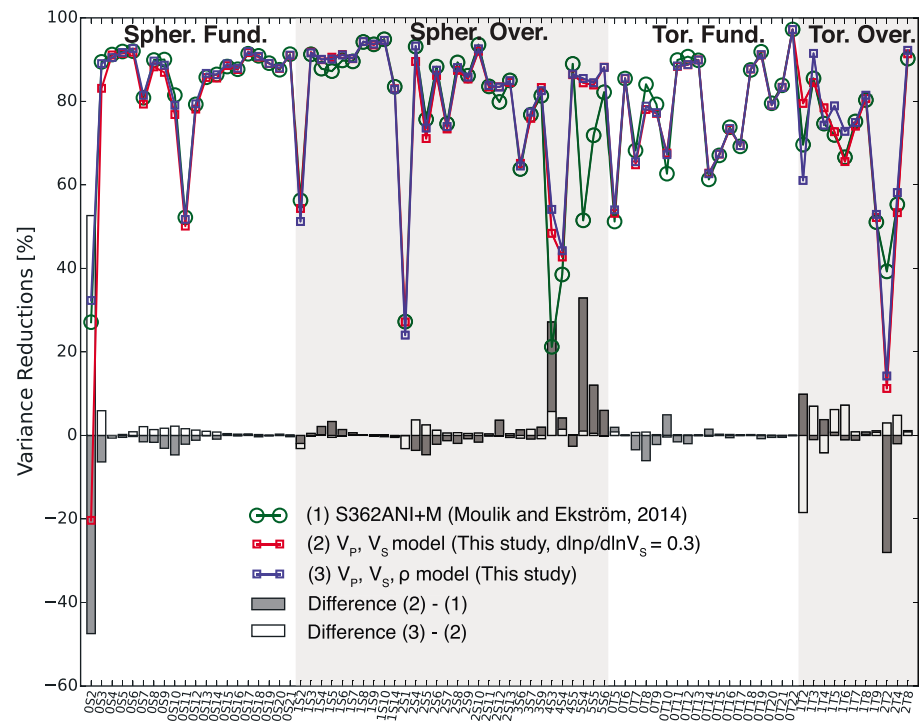
Figure 13. The variance reductions from S362ANI+M [*Moulik and Ekström*, 2014] and the joint v_P-v_S and v_P-v_S-ρ model from this study for the long-period (<3 mHz) modes. The difference in variance reductions are shown with the bars plotted underneath: grey bars represent the improvements with our joint v_P-v_S model relative to S362ANI+M while the white bars represent the additional improvement with our v_P-v_S-ρ model.

v_P-v_S model. Although not shown here, the fits to the other data sets of surface waves, long-period waveforms, and body wave travel times are very similar between our joint v_P-v_S and v_P-v_S-ρ model. For example, the variance reductions to the splitting functions do not vary by more than 2% for any short-period (3–10 mHz) mode when the additional complexity of density variations in the lower mantle is included in our inversions. When very high weights are applied to fit the $_0S_2$ splitting function perfectly, strong anticorrelation with shear velocity ($R \leq -0.5$) and high RMS variations are observed for density structure in the lowermost mantle. Moreover, variance reductions to the other types of data do not worsen by more than 5% in models with strong v_S-ρ anticorrelation in the lowermost mantle. We have applied moderate weights for the $_0S_2$ data to achieve a reasonable amount of fit ($\geq$30%) that is marginally better than that afforded by S362ANI+M (28%) and substantially better than our joint v_P-v_S model (−50%). The analysis of data fits indicates that the even-degree long-wavelength density structure can be resolved independently of velocity variations in the lowermost mantle.

Figure 14 describes the density variations from our preferred joint model of anisotropic velocities and density in the mantle. The perturbations in anisotropic shear and compressional velocity in this model are very similar in pattern ($R \geq 0.95$) to our joint v_P-v_S model (section 4.2.2) and are not described here. Due to the scaling constraints imposed in our inversions in the upper mantle, the density variations also have a pattern similar to that of the velocity heterogeneity. In the lowermost mantle, the preferred v_P-v_S-ρ model exhibits denser-than-average anomalies roughly coincident with the slower-than-average superplumes. The amplitude of lower mantle density perturbations are somewhat lower in our preferred model than in the unconstrained inversions in section 4.3.1 due to the additional smoothness applied to density than to velocity variations. The positive variations in density are in the range of 0.5–0.6% underneath the Pacific and Africa. The power of density heterogeneity in the lowermost 300 km of the mantle is dominated ($\geq$80%) by their degree 2 components.

The large-scale patterns of density and shear velocity gradually become dissimilar with depth in the lower mantle, especially in their long-wavelength components. The correlations between degree 6 variations decrease in the lower mantle and become anticorrelated in the lowermost mantle (Figure 14c).
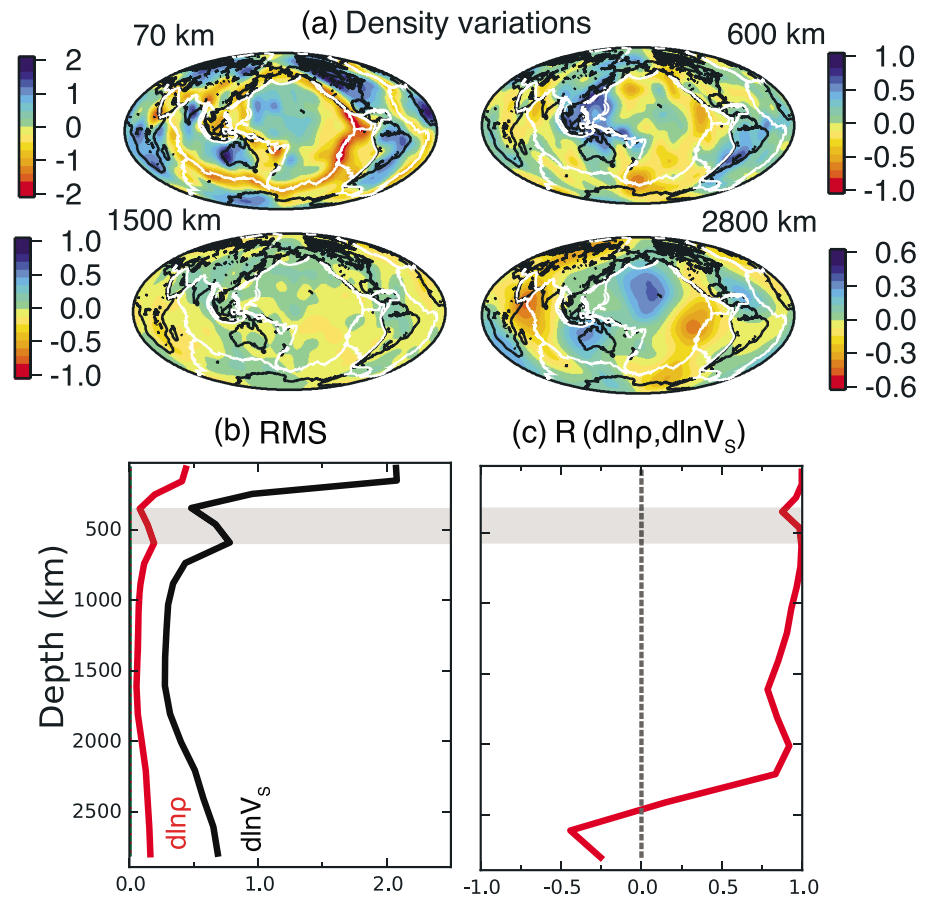
Figure 14. Features of our preferred joint v_P-v_S-ρ model: (a) density perturbations ($\frac{\delta\rho}{\rho^0}$) at 50, 600, 1500, and 2800 km; (b) the root-mean-square (RMS) values of the variations in shear velocity (degrees 1–18) and density (degrees 2,4, and 6); (c) the even-degree correlations (degrees 2, 4, and 6) between the isotropic shear velocity and density variations.

Our preferred model shows modest v_S-ρ anticorrelation ($R_2 \sim -0.46$ to -0.25) in the lowermost mantle. This model is inverted using a depth-varying scaling constraint (γ_ρ) that stabilizes the inversion and provides a smooth variation in degree 6 v_S-ρ correlation. The reduction in correlation is a feature that is independent of the type and amount of scaling constraint employed in our inversions; a variety of scaling scenarios, not shown here, give a similar trend in our inverted models. We have also experimented with lateral variations of radial anisotropy throughout the mantle [e.g., *Moulik and Ekström*, 2014]; such complexities in elastic structure give v_S-ρ correlations similar to those in our preferred model. The variation in correlation is required by both the low-order spheroidal fundamental modes and toroidal overtones; inversions excluding either one of these subsets of data display similar patterns in their v_S-ρ correlations. Based on this suite of inverted models, the data strongly disfavors perfect or even positive correlation ($R_2 \geq 0$) between shear velocity and density in the lowermost mantle.

4.4. Statistical Significance of the Modeled Complexities

A challenge in seismic tomography is testing the significance of additional model complexities in contrast to a simpler model that provides a reasonably good fit to the data. The number of basis functions in tomographic models (m = length of x_{LS}, section 3.2) provides a rough indication of model complexity; 7964 basis coefficients were inverted during the construction of S362ANI+M while 20,996 basis coefficients have been inverted throughout this study (Figure 3). However, the complexity in our inverse problem is controlled both by m and the a priori regularization [e.g., *Buja et al.*, 1989; *Hastie and Tibshirani*, 1990]. The amount of information extracted from observations or the "effective degrees of freedom for model" (df) can be approximated as

$$\text{df} = \text{tr}(H) = \text{tr}(R),\tag{11}$$

Table 3. Statistical Significance of the Improvements in Data Fits With Our Preferred v_P-v_S-ρ Model[a]

			Models		
			1. v_P-v_S	2. v_P-v_S-ρ	
Subsets of Data	n	Test	$df_1 = 3465.8$	$df_2 = 3576.3$	Significance
NM	8342	χ_1^2/χ_2^2		1.0784	Greater than 99% ($F_{0.01} = 1.0693$)
		$\triangle AIC_c$		37200.4	Low relative likelihood of Model 1
		$\triangle BIC$		36864.6	Evidence against Model 1 very strong
NM + TT	196906	χ_1^2/χ_2^2		1.0287	Greater than 99% ($F_{0.01} = 1.0106$)
		$\triangle AIC_c$		59013.0	Low relative likelihood of Model 1
		$\triangle BIC$		57895.1	Evidence against Model 1 very strong
NM + TT + SW	3150297	χ_1^2/χ_2^2		1.0081	Greater than 99% ($F_{0.01}=1.0019$)
		$\triangle AIC_c$		63062.6	Low relative likelihood of Model 1
		$\triangle BIC$		61630.7	Evidence against Model 1 very strong

[a]Various combinations of normal mode (NM), surface wave (SW), and body wave (TT) data are tested while considering each of the n number of measurements independent and the data noise as known a priori. The improvements in χ^2 fits ($\sum_i w_i \chi_i^2$ from section 3.2) are significant above the 99% confidence level ($\alpha = 0.01$) according to an F test criterion [*Menke*, 1989]. The thresholds for satisfying $F_\alpha(n - df_1, n - df_2)$ are calculated for various subsets of data, where df is the effective degree of freedom in a model from equation (11). Bayesian Information criteria (BIC $= \chi^2 + \ln(n) \cdot df + c_1$) typically penalizes greater degrees of freedom more severely than the Akaike Information Criterion ($AIC_c = \chi^2 + 2 \cdot df + [2 \cdot df \cdot (df + 1)/(n - df - 1) + c_2]$). The constant terms c_1 and c_2 cancel out when comparisons ($\triangle AIC_c$, $\triangle BIC$) are made between candidate models using the same subsets of data. Evidence against v_P-v_S model with the higher BIC is very strong ($\triangle BIC >10$) [*Raftery*, 1995] and its relative likelihood $\exp\left[-\triangle^{AIC_c}/2\right]$ is low [*Burnham and Anderson*, 2002], compared to our preferred v_P-v_S-ρ model.

where H is the "hat" matrix used in a variety of data mining applications and $tr(\cdot)$ denotes its trace [e.g., *Cardinali et al.*, 2004; *Ye*, 2012; *Ruppert*, 2012]. In least squares inverse problems that are the subject of this study, H is equivalent to the resolution matrix R [*Menke*, 1989]. Due to the strong regularization imposed in our inversions (section 3.2), the effective degrees of freedom increases only marginally from the v_P-v_S model in section 4.2 (df = 3465.8) to our preferred v_P-v_S-ρ model in section 4.3.2 (df = 3576.3).

When we allow additional degrees of freedom with v_P and ρ variations, variance is reduced and the reduction is systematic and large (often exceeding 50%) for the types of data sensitive to a given depth range and physical parameter. In the interest of brevity, we only report statistical significance of the improvements in fit with our v_P-v_S-ρ model compared to the v_P-v_S model (Table 3). The additional 110.5 effective degrees of freedom reduce the χ^2 fits to $_0S_2$ by 2.68 times and to all normal-mode splitting observations by 1.078 times, which is significant above the 99% confidence level according to an F test criterion [*Menke*, 1989]. It is difficult, however, to determine with F tests whether an improvement in fit is statistically significant when the number of observations is large [e.g., *Panning and Romanowicz*, 2006]. We adopt model selection methods that are typically used to identify preferred Markov Chain Monte Carlo models, such as the Akaike information criterion (AIC) [*Akaike*, 1974] and the Bayesian information criterion (BIC) [*Schwarz*, 1978]. BIC strongly penalizes model complexity in global tomography as it accounts explicitly for the large number of seismic observations, which are typically considered independent. Nevertheless, the relative likelihood of the simpler v_P-v_S model is low while the strength of evidence against it is very strong (Table 3). The improvements in data fit from the additional complexity of density variations in the lowermost mantle are, therefore, statistically very significant.

5. Discussion

5.1. Scaling Relationships

One of the applications of joint tomographic models is to use estimates from mineral physics in order to make thermochemical inferences from the seismically derived elastic parameters and density. A common practice in interdisciplinary comparisons is to assume proportionality between the lateral variations in two parameters, such as shear velocity (dln v_S) and density (dln ρ), and focus on the scaling factor (ϱ) as a function of depth. Figure 15a shows the variation of the radial scaling factors ν and ϱ from our joint v_P-v_S-ρ model. Although the scaling factors derived here are based solely on seismological observables, radial scaling factors in the upper mantle are consistent with those predicted from purely thermal effects based on experimentally
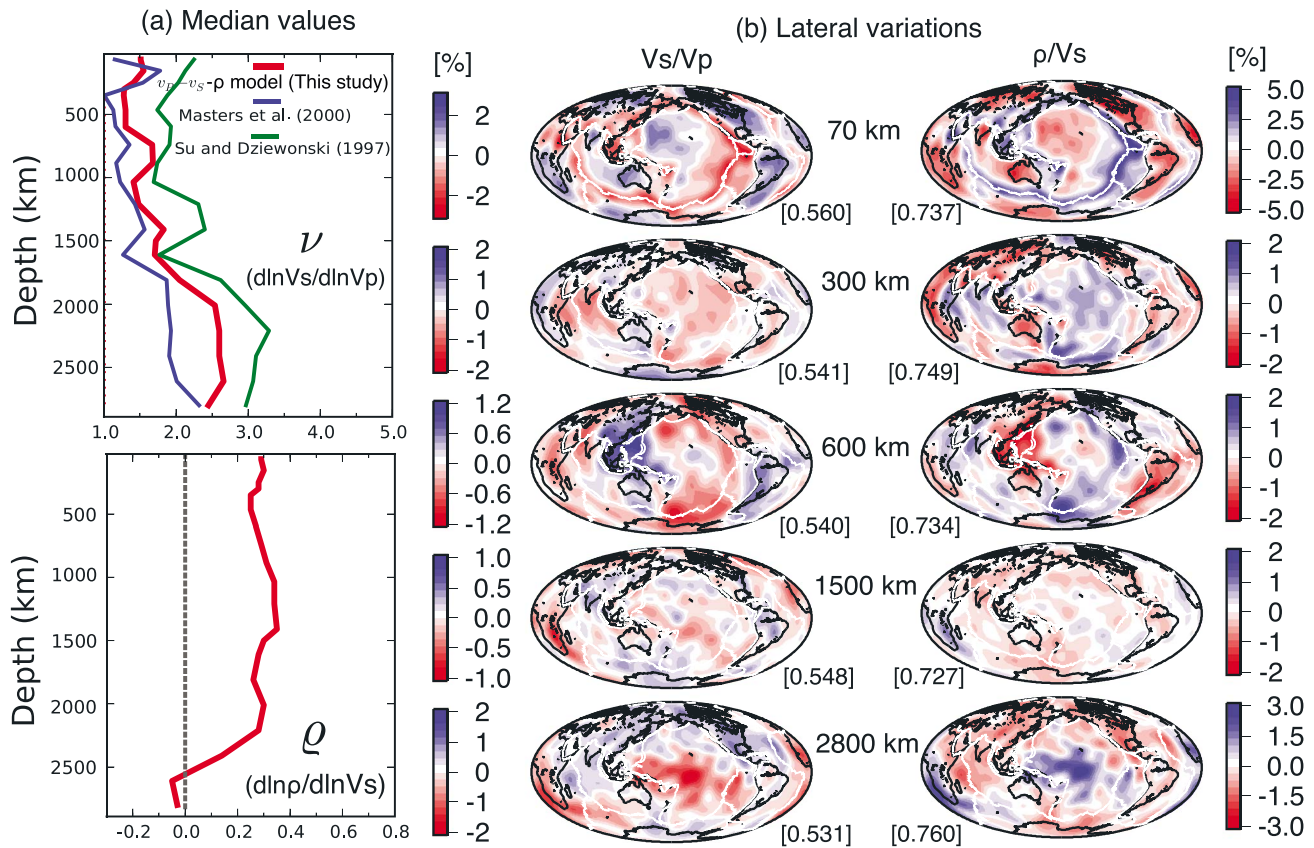
Figure 15. (a) Global estimates of the median ν and ϱ from our preferred v_P-v_S-ρ joint model (section 4.3.2) in red curves compared to that of *Su and Dziewoński* [1997] in green and *Masters et al.* [2000] in blue curves. Note that the median values are calculated using the procedure described in Figure 10b. (b) Lateral variations in scaling relationships are estimated from our joint v_P-v_S-ρ model and reference model STW105; perturbations are relative to the spherically averaged values, which are given in brackets [...].

derived scaling relationships [*Karato*, 1993; *Karato and Karki*, 2001; *Matas and Bukowinski*, 2007] and petrological constraints [*Montagner and Anderson*, 1989]. The median values of ν are also in broad agreement with the results from earlier tomographic studies of shear and compressional wave velocities in the mantle [e.g., *Robertson and Woodhouse*, 1996; *Masters et al.*, 2000].

The radial scaling ratios are adequate in describing the relative behavior of velocities and density only when their variations are perfectly correlated throughout the mantle. We have shown that this assumption, typically used in several geophysical [e.g., *Ritsema et al.*, 2011] and geochemical studies, may not be valid in the lowermost mantle, especially for the density variations. Comparisons between mineralogical and seismological estimates of radial scaling ratios may not be adequate in regions with strong lateral variations in composition; mineralogical estimates are typically based on isolated geologic samples of a uniform composition such as olivine, pyrolite, or piclogite [e.g., *Estey and Douglas*, 1986; *Montagner and Nataf*, 1986]. Figure 15b shows the lateral variations in scaling ratios (v_S/v_P, ρ/v_S) from our joint three-dimensional model. The laterally varying v_S/v_P ratio shows a pattern similar to that of the S velocity model. The ratio of density to v_S is generally dominated by the pattern of the v_S anomaly since lateral variations in density are weaker than those in v_S (Figure 14b). The strong lateral variations (up to 3%) along with the low ρ-v_S correlations suggest that very limited inferences can be drawn based on a depth-averaged scaling factor in the lowermost mantle. Lateral variation in scaling ratios from our model could help deconvolve the thermal and chemical contributions to heterogeneity in the lowermost mantle.

5.2. Thermochemical Interpretations of the Lowermost Mantle
Large-scale variations of the elastic properties modeled in this study disfavor a purely thermal contribution to heterogeneity in the lowermost mantle. The anticorrelation between shear and bulk sound velocities in parts of the lower mantle [*Su and Dziewoński*, 1997; *Masters et al.*, 2000] is found to persist even with the inclusion

of new mode-splitting data (Figure 11d). The velocity scaling factor v increases gradually and exceeds 2.5 in the lowermost mantle; such values are too large to be explained by purely thermal variations [*Karato*, 1993; *Karato and Karki*, 2001; *Matas and Bukowinski*, 2007]. The experiments using various subsets of the data set and amounts of scaling regularization (e.g., Appendix C) suggest that the variation of v is resolved well in the lowermost mantle. Ray theory predicts that the body waves tend to diffract toward high-velocity regions in order to find minimum-time trajectories [e.g., *Wielandt*, 1987]; this effect may lead to smaller or "healed" travel time residuals and therefore faster estimated velocities in tomographic inversions. Some studies focused on body wave travel times have attributed large values of v to the stronger wavefront healing in compressional waves than shear waves of the same frequency [*Malcolm and Trampert*, 2011; *Davies et al.*, 2012; *Schuberth et al.*, 2012]. The effects of wavefront healing are less severe in surface waves [*Nolet and Dahlen*, 2000] and especially in long-period normal modes as they sample the integrated effects of volumetric heterogeneity. Since the new splitting data set of v_P-sensitive modes [*Deuss et al.*, 2013] also implies large values of v in the lowermost mantle, this feature in our model is unlikely to be a consequence of differential wavefront healing of shear and compressional waves. The proposition that travel times of a few body wave phases may be explained by purely thermal anomalies, therefore, cannot be extrapolated to other types of data and to structure in the lowermost mantle.

Our results on the relative behavior of v_P and v_S variations are within the depth range where perovskite (Pv) could transform to postperovskite (pPv) in the lowermost mantle [e.g., *Murakami et al.*, 2004, 2004; *Oganov and Ono*, 2004; *Tsuchiya et al.*, 2004]. Relative to perovskite, postperovskite has higher shear wave velocities, lower compressional velocity and, therefore, a higher v [e.g., *Wookey et al.*, 2005; *Ammann et al.*, 2010; *Stixrude and Lithgow-Bertelloni*, 2011]. However, the pressure and temperature conditions of the postperovskite phase boundary remain a subject of ongoing debate. The effects of chemical composition [e.g., *Catalli et al.*, 2009; *Grocholski et al.*, 2012] as well as the uncertainties in absolute temperatures at the core-mantle boundary (CMB) [e.g., *Asanuma et al.*, 2010; *Kamada et al.*, 2010], in experimental measurements [e.g., *Anderson et al.*, 1989; *Hirose et al.*, 2008], and in ab initio calculations [e.g., *Oganov et al.*, 2002; *Oganov and Ono*, 2004; *Cococcioni and de Gironcoli*, 2005] can translate into the Pv-pPv phase transition occurring at pressures corresponding to various depths in the lower mantle or even below the core-mantle boundary. Postperovskite is more likely to be present in localized, faster-velocity regions in the deepest ∼200 km of the mantle [e.g., *Houser*, 2007; *Hernlund and Houser*, 2008; *Cobden et al.*, 2012; *Mosca et al.*, 2012]. On the contrary, the highest values of v in our model are in regions where shear wave speed is lowest, i.e., in the lowermost mantle beneath southern Africa and the central and western Pacific. The lowermost mantle is also the region that exhibits a wide distribution of v (Figure 10b), a feature that is diagnostic of chemical heterogeneity in the mantle [*Deschamps and Trampert*, 2003]. Based on the lateral distribution of v in our models and available constraints from mineral physics, the variations of v in the lowermost mantle cannot easily be attributed to a Pv-pPv phase transition, unless the postperovskite stability field is found to be pervasive in the lowermost mantle.

The anticorrelation between variations in shear velocity and density in our models also indicate a substantial contribution from chemical heterogeneity in the lowermost mantle, broadly consistent with earlier studies based on mode-splitting and free-air gravity data [*Ishii and Tromp*, 1999, 2001]. An intrinsic density contrast of ∼1% (peak to peak) roughly coincident with the slow-velocity superplumes provides a way to explain the splitting of gravest modes (section 4.3.2). The large-scale denser-than-average anomalies are indicative of geochemically distinct material, potentially sourced from oceanic crust that remains denser than the surrounding mantle during subduction [e.g., *Hirose et al.*, 2005; *Ricolleau et al.*, 2010]. The Pv-pPv phase transition can lead to a density increase of 1–1.5% in the lowermost mantle [e.g., *Murakami et al.*, 2004; *Wookey et al.*, 2005]. However, the denser anomalies in our model are almost coincident with low shear velocity superplumes, regions less likely to be in the pPv stability field [e.g., *Houser*, 2007; *Hernlund and Houser*, 2008]. While higher-velocity regions surrounding the superplumes may contain pPv-bearing mineral assemblages, high densities and slower velocities in superplumes are more consistent with a chemically distinct dense material. The contribution from chemical variations is also indicated by other geophysical constraints on lower mantle heterogeneity. Plumes arising from the top of a dense, compositionally stratified layer in the lowermost mantle will exhibit smaller excess temperatures relative to the ambient mantle with values comparable to the estimates from hotspots [e.g., *Albers and Christensen*, 1996; *Farnetani*, 1997; *Dannberg and Sobolev*, 2015]. In the absence of a dense layer, geodynamic calculations tend to predict too strong CMB relief and excess ellipticity in the Earth [e.g., *Steinberger and Calderwood*, 2006; *Lassak et al.*, 2007; *Steinberger and Holme*, 2008].

The amplitude and patterns of density variations have important implications for the age and fixity of the low-velocity superplumes as well as the current stage of their entrainment by mantle convection. Several laboratory, analytical, and numerical studies agree that a chemical density contrast of ~1–2% could permit a dense layer to be stable over the age of the Earth [e.g., *Gonnermann et al.*, 2002; *Zhong and Hager*, 2003; *Lin and van Keken*, 2006; *Tackley*, 2012; *Mulyukova et al.*, 2015]. Since entrainment rates depend also on the viscosity contrast, vigor of convection and physical parameters like the thermal expansivity in the lowermost mantle [*Tackley*, 2007], the density contrast required for long-lived superplumes remains an outstanding question in global geodynamics. Our tomographic estimates of peak-to-peak amplitudes in density structure can only provide a snapshot of dynamic processes and are limited by the resolution of current data sets. Therefore, amplitude of density heterogeneity in our models cannot be readily used to discriminate between stability of superplumes over time scales of billions of years [e.g., *Dziewonski et al.*, 2010] or a few hundred million years [e.g., *Tackley*, 2012]. While we have imposed strong smoothness for density variations in light of the low- and even-degree constraints afforded by mode-splitting data, patterns of density variations in the lowermost mantle (section 4.3.2) are largely independent of the regularization scheme. Based on the v_S-ρ anticorrelations in our preferred model and geodynamic considerations, the superplumes may be long-lived ($\geq$100 Ma) thermochemical features rather than transient thermal upwellings arising from the core-mantle boundary.

5.3. Modeling Limitations

Our estimates of density heterogeneity in the lowermost mantle are largely based on splitting observations of the gravest normal modes. While splitting coefficients reported by recent studies can be directly attributed to lateral heterogeneity, the additional splitting due to Earth's rotation and ellipticity is in fact dominant at low angular degrees and frequencies [e.g., *Backus and Gilbert*, 1961; *Luh*, 1974]. Several fundamental spheroidal- and toroidal-mode multiplets comprising the largest signals in long-period seismograms can be strongly coupled in the 1.5–3 mHz range ($_0S_l$–$_0T_{l+1}$), primarily due to the Coriolis force [e.g., *Park*, 1986]. The details of corrections applied to the spectra can, therefore, influence the splitting coefficient measurements and our imaged density structure. While $_0S_2$ splitting coefficients from *Deuss et al.* [2011] are broadly consistent with independent constraints from superconducting gravimeter data [e.g., *Rosat et al.*, 2005; *Häfner and Widmer-Schnidrig*, 2013], we perform additional experiments to check the robustness of our density model. We obtain similar variations to our preferred density model (section 4.3.2) even when the $_0S_2$-splitting data are excluded from our inversions. All correlated v_S-ρ models exhibit too strong amplitudes of $_0S_2$ splitting function due to the same sign and comparable sensitivity to density and shear velocity structure (Figure 2); low amplitude of the measured $_0S_2$ splitting function, therefore, limits the amount of positive v_S-ρ correlation compatible with the data. Fitting perfectly the current $_0S_2$-splitting data would necessitate strong v_S-ρ anticorrelation ($R_2 \leq -0.5$); we prefer moderate weights to $_0S_2$ that result in mild anticorrelation in the lowermost ~300 km of the mantle (section 4.3.2). We obtain similar results in other experiments that examine the sign of density anomalies in the lowermost mantle (section 4.3.1); all ρ-sensitive modes show reduced fits with high v_S-ρ scaling constraints, even when $_0S_2$ data are excluded from our inversions (Figure 12, dotted curves). Our experiments indicate that the pattern of density anomalies in the lowermost mantle are robust and compatible with different subsets of the density-sensitive mode-splitting data.

Recent theoretical developments on the splitting of normal modes may provide new constraints on large-scale scaling relationships in the mantle. The splitting of several modes used in this study were measured while accounting for the coupling between pairs or groups of modes [*Resovsky and Ritzwoller*, 1998; *Deuss et al.*, 2011]. Recent studies have suggested that the so-called "full-coupling" calculations may be necessary to obtain sufficiently accurate synthetic spectra employed in the measurements of normal-mode splitting [e.g., *Deuss and Woodhouse*, 2001; *Al-Attar et al.*, 2012]. It is not clear how the scaling relationships derived in this study may be influenced by accounting explicitly for the theoretical interactions between modes over a wide frequency band. The coupling due to structural heterogeneity is expected to be low for the isolated modes like $_0S_2$ and $_0S_3$; it seems unlikely that low- and even-degree v_P-ρ anticorrelation in the lowermost mantle preferred by these data (section 4.3.1) will be removed with theoretical advancements. However, the more complete full-coupling calculations could provide constraints on odd- and high-degree variations in scaling relationships that were not modeled in this study.

We model long-period waveforms in the context of the path average approximation [e.g., *Woodhouse and Dziewoński*, 1984] while modern numerical techniques, such as the spectral element method [e.g., *Komatitsch et al.*, 2002], can theoretically attribute every portion of the seismogram to heterogeneity using adjoint methods [e.g., *Tarantola*, 1984; *Tromp et al.*, 2005; *Liu and Tromp*, 2006, 2008]. Although several studies have shown

that retrieved structure is influenced less by theory than choices of parameterization, regularization, and data coverage [e.g., *Spetzler et al.*, 2002; *Sieminski et al.*, 2004; *van der Hilst and de Hoop*, 2005; *Boschi et al.*, 2006; *Trampert and Spetzler*, 2006], accurate numerical modeling of the P-SV portions of body wave waveforms may be useful for constraining small-scale variations of ν in the mantle. Another theoretical complexity is the potential trade-off between source location and heterogeneity in tomographic inversions [e.g., *Thurber*, 1992; *Roecker et al.*, 2006]. We account partially for the errors in source location in our inversions by solving iteratively for the centroid moment tensor (CMT) locations with our waveform data. Similar procedures need to be adopted for the remaining data sets; P wave travel times, in particular, can contain comparable contributions from location errors and velocity heterogeneity [e.g., *Masters et al.*, 1996]. While our choice of applying higher weights to body wave waveforms than P wave travel times potentially reduce source-structure trade-offs, details of the v_P heterogeneity need to be interpreted with caution till the relocation step is implemented for our travel-time data. We focus on deriving scaling relationships and a comprehensive model of v_P structure that could be used for applications like earthquake location [e.g., *Kennett et al.*, 1995] and nuclear test monitoring [e.g., *Antolik et al.*, 2003] is beyond the scope of this work.

The inclusion of additional data sets could also improve the resolution of scaling relationships in the mantle. The small-scale and odd-degree variations, in particular, are not constrained by the self-coupling splitting data. The long-period surface waves and P wave travel times alleviate somewhat the problem of limited sensitivity of mode-splitting data to odd-degree v_P variations in the mantle. In this study, our comparisons between velocity and density structure is limited to the low and even ($\leq6°$) (section 5.1); odd-degree density variations are primarily due to damping and should be interpreted with caution. Free-air gravity constraints can provide odd-degree sensitivity to density structure [e.g., *Simmons et al.*, 2010] and are potentially useful when combined with mode-splitting data [*Ishii and Tromp*, 1999, 2001]. While the increasing number of cross-coupling splitting data should provide new constraints on odd-degree density variations in the mantle, our results on even-degree structure would remain largely unaffected by these data. A way to probe scaling relationships in a specific region like the lowermost mantle is by including additional measurements of diffracted body waves (P_{diff}, S_{diff}) and incorporating CMB-Stoneley modes that are strongly sensitive to structure at these depths [e.g., *Ritsema et al.*, 2011; *Koelemeijer et al.*, 2016]. However, our experiments with the splitting coefficients of CMB-Stoneley modes reported by a single study [*Koelemeijer et al.*, 2013] reveal that their data are fit well by our models with overall variance reductions in excess of 92%. The CMB-Stoneley modes are difficult to measure due to the limited displacements of their eigenfunctions on the Earth's surface. They also have limited sensitivity to density variations in the lowermost mantle and our models with a variety of density scaling scenarios give comparable fits to the reported data. While we have included several diffracted arrivals in our inversions, the resolution of ν will improve with an expanded data set of body wave travel times and newer data sets available from the seismological community.

The exclusion of certain parameters in our inversions could potentially trade off with scaling relationships in the mantle. Several low-order spheroidal modes used in this study are also sensitive to heterogeneity in the outer core [e.g., *Romanowicz and Bréger*, 2000], a structural complexity that has not been evaluated in this study. The core-reflected body wave phases (e.g., ScS − S) and long-period mode-splitting data are also sensitive to the undulations of the core-mantle boundary (CMB) but do not have the resolving power to constrain its topography independently of the transition-zone discontinuity topographies and volumetric heterogeneity in the mantle. We can use constraints from earlier studies to place a priori bounds on CMB topography and examine potential trade-offs with other parameters in our model. The CMB topographic relief, defined as deviation from the hydrostatic ellipsoidal shape caused by Earth's rotation, is debated in seismological studies with peak-to-valley values ranging from ±1.5 km to more than 6 km [e.g., *Morelli and Dziewonski*, 1987; *Rodgers and Wahr*, 1993; *Obayashi and Fukao*, 1997; *Sze and van der Hilst*, 2003]. Some of these estimates exceed the values inferred from other geophysical techniques like space geodesy [e.g., *Gwinn et al.*, 1986, 0.4 km] and geoid modeling [e.g., *Hager et al.*, 1985, ≤1.5 km]. We perform trial inversions where the smoothness damping is adjusted such that the Y_2^0 component of CMB topography is ~0.5 km, consistent with values inferred from very long baseline interferometry (VLBI) observations of Earth's forced nutations [e.g., *Herring et al.*, 1986; *Mathews and Shapiro*, 1992]. The rest of the regularization and data weighting scheme is kept the same as used in the construction of our preferred joint model (section 4.3.2). Our models inverted with and without CMB topography as an additional parameter in the inversions exhibit similar variations ($R \geq 0.9$) in density and other structural heterogeneities in the mantle. We infer that the current bounds on the long-wavelength CMB topography from other data sets preclude substantial trade-offs with other parameters in our inversions.

6. Conclusions

We have devised a new method to solve for density, anisotropic shear, and compressional velocities as well as discontinuity topographies in the Earth's mantle using various seismological observations. Our method involves three steps that result in models with different levels of scaling complexity: (i) simple uniform scaling ratio for density ($R_\rho = 0.3$) and split uniform scaling ratios for velocity in the upper ($R_P^{UM} = 0.7$) and lower mantle ($R_P^{LM} = 0.4$), (ii) relaxing constraints on velocity scaling to derive a joint v_P-v_S model, and finally, (iii) relaxing constraints on v_S-ρ scaling to arrive at our preferred v_P-v_S-ρ model. The fits to the v_P-sensitive modes and body wave travel times improve substantially when the split uniform scaling is used for the upper and lower mantle. Our v_P-v_S model reveals several robust features consistent with earlier studies of mantle heterogeneity: (1) negative correlation between shear and bulk sound velocities in the lowermost 800–900 km of the mantle and (2) a gradual increase in the velocity scaling ratio ν to values in the range of 2–2.5. These features cannot be attributed solely to the effects of wavefront healing in body wave travel times since they persist with the inclusion of mode splitting and other long-period observations.

We demonstrate that the multitude of data sets considered in this study cannot be fit concurrently with a constant and positive R_ρ throughout the mantle. The fits to several ρ-sensitive modes are reduced with the use of uniform scaling factors and do not improve even when velocity scaling constraints are relaxed in our v_P-v_S model. We have found no scenarios of large-scale elastic heterogeneity that can provide good fits to some of the density-sensitive modes. Several spheroidal and toroidal modes strongly disfavor a positive correlation between v_S and density variations in the lowermost mantle. Relaxing constraints on v_S-ρ scaling in the lowermost mantle leads to statistically significant improvements in fit to almost all observations in this study. The choice between v_S-ρ decorrelation and anticorrelation is more subtle and depends on the fits to individual observations. Our preferred model exhibits large-scale denser-than-average anomalies mildly anticorrelated with low-v_S superplumes in the lowermost mantle. The pattern of density anomalies in the lowermost mantle is largely independent of the regularization scheme and whether individual ρ-sensitive modes like $_0S_2$ are included in our inversions. The modeled relationships between shear velocity, density and compressional velocity cannot be reconciled with a purely thermal source of large-scale heterogeneities in the lowermost mantle.

Appendix A: Corrections to Surface Wave Phase Anomalies

Several corrections are applied to the surface wave phase anomalies before including them in our tomographic inversions (equation (1)). The crustal correction involves a nonlinear scheme that computes local phase velocities for the reference model overlain by CRUST2.0 structure and then calculates the integrated phase anomaly along each source-receiver path ($\delta\Phi_{crust}$). In contrast to our earlier studies based on minor arc data [*Kustowski et al.*, 2008; *Moulik and Ekström*, 2014], we correct all minor arc, major arc, and higher-orbit phase anomalies for the effects of Earth's ellipticity and account for azimuthally varying phase slowness in Rayleigh waves. The correction for Earth's ellipticity and reference model can be expressed as

$$\delta\Phi_{ref,ellip} = -\frac{\omega}{c_0} \int_{path} \left[\frac{c_{ref} - c_{PREM}}{c_{PREM}} Y_0^0(\theta, \phi) + c^e Y_2^0(\theta, \phi) \right] ds, \tag{A1}$$

where c^e represents the theoretical contribution of hydrostatic ellipticity [*Dahlen and Tromp*, 1998] while c_{PREM} and c_{ref} are the phase velocities from PREM and the reference model employed in our tomographic inversion (section 3.1), respectively. The azimuthal variations in Love and Rayleigh wave phase velocity (or slowness) can be described by patterns having twofold or fourfold azimuthal symmetry with respect to the propagation azimuth from the North ζ [*Smith and Dahlen*, 1973]. While most authors consider only the twofold (2ζ) anisotropy of Rayleigh waves in their analysis [e.g., *Montagner and Nataf*, 1986; *Nishimura and Forsyth*, 1989; *Maggi et al.*, 2006], some studies have argued that other types of variations may be resolved in both Love and Rayleigh wave data [e.g., *Trampert and Woodhouse*, 2003; *Visser and Trampert*, 2008]. *Ekström* [2011] has demonstrated that only the 2ζ azimuthal variations in Rayleigh waves are required by the bulk of our data, and the fits do not improve substantially when other azimuthally varying terms are included. The dominant role of the 2ζ variations of Rayleigh waves is also predicted by petrological studies [e.g., *Montagner and Nataf*, 1986; *Montagner and Anderson*, 1989] and has been linked to fabric-generating geodynamic processes like the

lattice-preferred orientation of olivine resulting from mantle flow [e.g., *Nishimura and Forsyth*, 1989; *Gaboret et al.*, 2003; *Becker et al.*, 2003, 2014]. We estimate the phase anomaly due to azimuthal variations in slowness as

$$\delta\Phi_\zeta = \frac{\omega}{c_0} \int\limits_{path} [A(\theta, \phi) \cos 2\zeta + B(\theta, \phi) \sin 2\zeta] \, ds, \tag{A2}$$

where the laterally varying coefficients $A(\theta, \phi)$ and $B(\theta, \phi)$ describe the 2ζ azimuthal variations in slowness with respect to the reference isotropic slowness. The coefficients are expanded in terms of 1442 lateral splines and equation (A2) is slightly modified to allow a smoothly varying direction in regions near the poles [*Ekström*, 2006]. The spline coefficients for the azimuthal corrections to Rayleigh waves are taken from the global dispersion model GDM52 [*Ekström*, 2011], which was derived using the minor arc and major arc subsets of our data.

Appendix B: Details of the Two-Step Inversion Scheme

Since the sensitivity kernels for the long-period waveforms require a 3-D starting model, we first invert the other data sets using STW105 [*Kustowski et al.*, 2008] as our 1-D starting model. Several experiments on the regularization scheme are conducted (section 3.3) to arrive at the preferred range of damping parameters. The scaling constraints are also varied separately in the four different regions of the mantle: (1) upper mantle, (2) transition zone, (3) central lower mantle, and (4) the lowermost mantle (Figure 3). The variable weighting scheme allows us to test the resolution of scaling relationships from current data sets in various regions of the mantle. The model inverted using normal-mode splitting functions, body wave travel times, and surface wave phase anomalies is optimized to explain most of the variance in these data. The final model is constructed by incorporating the waveform data sets in our joint modeling. In the first iteration of the waveform inversion, we calculate synthetic seismograms using the Global CMT solutions [e.g., *Ekström et al.*, 2005] and the 3-D model derived without the waveform data sets. We then accumulate all the data and regularization inner product matrices and invert them jointly for the velocity structure and discontinuity topographies. The new model is then used to determine new CMT solutions, the structural kernels are updated and accumulated, and several iterations of the joint inversions are done till we achieve convergence in fits to all data sets. The inversion procedure is computationally intensive since five iterations with updated 3-D starting models, comprising six subiterations that solve for source parameters (i.e., CMTs) of 229 events are required to reach convergence in fit to our waveform data sets.

Appendix C: Robustness of the v_P-v_S Variations

The v_P-v_S correlations and ν in the upper mantle are largely independent of the scaling constraints imposed in our inversions; values of scaling weights γ_P up to 1000 times the preferred value (γ_P^{pref}) give similar variations. Adopting low γ_P (e.g., $0.1 \times \gamma_P^{pref}$) results in reduced values of v_P-v_S correlations with ν below 1 at depths around 300 km; this complexity in structure does not appreciably improve ($\leq$1.5%) the fit to any data. We found that the anticorrelation between bulk sound and shear velocity in the lowermost mantle is largely independent of the scaling regularization (γ_P) in the lowermost mantle. However, the value of v_P-v_S correlations and ν in the lowermost mantle is mildly dependent on γ_P. Our data favors a slightly negative gradient in ν in the bottom 300 km of the mantle with stronger gradients when the scaling constraints are relaxed ($0.05-0.1 \times \gamma_P^{pref}$). Imposing v_P-v_S correlations in the lowermost mantle using a very high γ_P, on the other hand, leads to a substantial reduction in fit to the P wave travel times ($\geq$5%). The variation of ν at midmantle depths ($650-2000$ km) is the most strongly affected by the choice of regularization scheme in the absence of waveform data sets. Harvard's P and PP data sets favor low ν ($\leq$1) in the midmantle with strong power in the small-scale v_P heterogeneity while the mode-splitting data give comparable fits for a variety of scaling constraints. This observation suggests that criteria other than the fits to individual data sets are needed in evaluating whether the inversion results are meaningful. We choose to not give additional weights to Harvard's P and PP data, and our scaling weights γ for midmantle splines (Figure 3) are adjusted to obtain $\nu \sim$1.5, a value consistent with experimentally derived estimates from purely thermal variations at these depths in the mantle [e.g., *Karato and Karki*, 2001].

The v_P-v_S model constructed without the waveforms is largely consistent with the waveform data; median variance reductions of waveforms from the 229 earthquakes are within 4% of the values afforded by S362ANI+M. Nevertheless, waveform data sets are crucial to our modeling as they provide additional constraints on the variation of ν in the lower mantle. Most of the constraints on ν between 650 and 2000 km are obtained from waveform data sets and their inclusion reduces the importance of strong damping toward experimentally derived values imposed in their absence. Several data sets sensitive to v_P-v_S relationships in the midmantle tend to trade off with each other; fits to several body wave phases bottoming at these depths (SS, P, and PP) are reduced by ∼4% due to the new constraints from waveform data. Such reductions in fit are expected from the low weights of P and PP data, a choice that is informed by our limitations in modeling absolute travel times (section 5.3).

References

Abe, Y. (1997), Thermal and chemical evolution of the terrestrial magma ocean, *Phys. Earth Planet. Inter.*, *100*, 27–39.

Agnon, A., and M. S. T. Bukowinski (1990), δ_s at high pressure and $d\ln V_s/d\ln V_p$ in the lower mantle, *Geophys. Res. Lett.*, *17*(8), 1149–1152.

Akaike, H. (1974), A new look at the statistical model identification, *IEEE Trans. Autom.Control*, *19*(6), 716–723.

Al-Attar, D., J. H. Woodhouse, and A. Deuss (2012), Calculation of normal mode spectra in laterally heterogeneous Earth models using an iterative direct solution method, *Geophys. J. Int.*, *189*(2), 1038–1046.

Albers, M., and U. R. Christensen (1996), The excess temperature of plumes rising from the core-mantle boundary, *Geophys. Res. Lett.*, *23*(24), 3567–3570.

Ammann, M. W., J. P. Brodholt, J. Wookey, and D. P. Dobson (2010), First-principles constraints on diffusion in lower-mantle minerals and a weak D″ layer, *Nature*, *465*, 462–465.

Anderson, D. L., and A. M. Dziewoński (1982), Upper mantle anisotropy: Evidence from free oscillations, *Geophys. J. Int.*, *69*(2), 383–404.

Anderson, O., and D. Isaak (1995), Elastic constants of mantle minerals at high temperature, in *Mineral Physics and Crystallography*, edited by T. Ahrens, pp. 64–97, AGU, Washington, D. C.

Anderson, O. L., D. G. Isaak, and S. Yamamoto (1989), Anharmonicity and the equation of state for gold, *J. Appl. Phys.*, *65*(4), 1534–1543.

Antolik, M., Y. J. Gu, G. Ekström, and A. M. Dziewonski (2003), J362D28: A new joint model of compressional and shear velocity in the Earth's mantle, *Geophys. J. Int.*, *153*(2), 443–466.

Asanuma, H., E. Ohtani, T. Sakai, H. Terasaki, S. Kamada, T. Kondo, and T. Kikegawa (2010), Melting of iron-silicon alloy up to the core-mantle boundary pressure: Implications to the thermal structure of the Earth's core, *Phys. Chem. Miner.*, *37*(6), 353–359.

Babuska, V., and M. Cara (1991), *Seismic Anisotropy in the Earth*, Kluwer Academic Press, Boston, Mass.

Backus, G., and F. Gilbert (1961), The rotational splitting of the free oscillations of the Earth, *Proc. Natl. Acad. Sci.*, *47*, 362–371.

Becker, T. W., J. B. Kellogg, G. Ekström, and R. J. O'Connell (2003), Comparison of azimuthal seismic anisotropy from surface waves and finite strain from global mantle-circulation models, *Geophys. J. Int.*, *155*, 696–714.

Becker, T. W., C. P. Conrad, A. J. Schaeffer, and S. Lebedev (2014), Origin of azimuthal seismic anisotropy in oceanic plates and mantle, *Earth Planet. Sci. Lett.*, *401*, 236–250.

Bolton, H., and G. Masters (2001), Travel times of *P* and *S* from the global digital seismic networks: Implications for the relative variation of *P* and *S* velocity in the mantle, *J. Geophys. Res.*, *106*, 13527–13540.

Boschi, L., and A. M. Dziewonski (1999), High- and low-resolution images of the Earth's mantle: Implications of different approaches to tomographic modeling, *J. Geophys. Res.*, *104*, 25,567–25,594.

Boschi, L., T. W. Becker, G. Soldati, and A. M. Dziewonski (2006), On the relevance of Born theory in global seismic tomography, *Geophys. Res. Lett.*, *33*, L06302, doi:10.1029/2005GL025063.

Brandenburg, J. P., and P. E. van Keken (2007), Deep storage of oceanic crust in a vigorously convecting mantle, *J. Geophys. Res.*, *112*, B06403, doi:10.1029/2006JB004813.

Brodholt, J. P., G. Helffrich, and J. Trampert (2007), Chemical versus thermal heterogeneity in the lower mantle: The most likely role of anelasticity, *Earth Planet. Sci. Lett.*, *262*, 429–437.

Buja, A., T. Hastie, and R. Tibshirani (1989), Linear smoothers and additive models on JSTOR, *Ann. Stat.*, *17*, 453–510.

Burnham, K. P., and D. R. Anderson (2002), *Model selection and multimodel inference: A practical information-theoretic approach*, 2nd ed., Springer-Verlag, New York.

Cardinali, C., S. Pezzulli, and E. Andersson (2004), Influence-matrix diagnostic of a data assimilation system, *Q. J. R. Meteorol. Soc.*, *130*, 2767–2786.

Catalli, K., S.-H. Shim, and V. Prakapenka (2009), Thickness and Clapeyron slope of the post-perovskite boundary, *Nature*, *462*, 782–785.

Christensen, U. R., and A. W. Hofmann (1994), Segregation of subducted oceanic crust in the convecting mantle, *J. Geophys. Res.*, *99*, 19,867–19,884.

Cobden, L., I. Mosca, J. Trampert, and J. Ritsema (2012), On the likelihood of post-perovskite near the core-mantle boundary: A statistical interpretation of seismic observations, *Phys. Earth Planet. Inter.*, *210-211*, 21–35.

Cococcioni, M., and S. de Gironcoli (2005), Linear response approach to the calculation of the effective interaction parameters in the LDA+U method, *Phys. Rev. B*, *71*, 35105.

Coltice, N., and Y. Ricard (1999), Geochemical observations and one layer mantle convection, *Earth Planet. Sci. Lett.*, *174*, 125–137.

Dahlen, F. A., and J. Tromp (1998), *Theoretical Global Seismology*, Princeton Univ. Press, Princeton, N. J.

Dalton, C., C. H. Langmuir, and A. Gale (2014), Geophysical and geochemical evidence for deep temperature variations beneath mid-ocean ridges, *Science*, *344*, 80–83.

Dannberg, J., and S. V. Sobolev (2015), Low-buoyancy thermochemical plumes resolve controversy of classical mantle plume concept, *Nat. Commun.*, *6*, 6960, doi:10.1038/ncomms7960.

Davies, D. R., S. Goes, J. H. Davies, B. S. A. Schuberth, H. P. Bunge, and J. Ritsema (2012), Reconciling dynamic and seismic models of Earth's lower mantle: The dominant role of thermal heterogeneity, *Earth Planet. Sci. Lett.*, *353-354*, 253–269.

Debayle, E., and Y. Ricard (2012), A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements, *J. Geophys. Res.*, *117*, B10308, doi:10.1029/2012JB009288.

Deschamps, F., and J. Trampert (2003), Mantle tomography and its relation to temperature and composition, *Phys. Earth Planet. Inter.*, *140*(4), 277–291.

Deuss, A., and J. H. Woodhouse (2001), Theoretical free-oscillation spectra: The importance of wide band coupling, *Geophys. J. Int.*, *146*, 833–842.

Deuss, A., J. Ritsema, and H. van Heijst (2011), Splitting function measurements for Earth's longest period normal modes using recent large earthquakes, *Geophys. Res. Lett.*, *38*, L04303, doi:10.1029/2010GL046115.

Deuss, A., J. Ritsema, and H. van Heijst (2013), A new catalogue of normal-mode splitting function measurements up to 10 mHz, *Geophys. J. Int.*, *193*, 920–937.

Dziewoński, A. M., and D. L. Anderson (1981), Preliminary reference Earth model, *Phys. Earth Planet. Inter.*, *25*, 297–356.

Dziewonski, A. M., and R. Woodward (1992), Acoustic imaging at the planetary scale, in *Acoustical Imaging*, vol. 19, edited by A. Dziewonski and R. Woodward, pp. 785–797, Plenum Press, New York.

Dziewonski, A. M., V. Lekic, and B. A. Romanowicz (2010), Mantle Anchor structure: An argument for bottom up tectonics, *Earth Planet. Sci. Lett.*, *299*, 69–79.

Ekström, G. (2006), A simple method of representing azimuthal anisotropy on a sphere, *Geophys. J. Int.*, *165*, 668–671.

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

Ekström, G., J. Tromp, and E. Larson (1997), Measurements and global models of surface wave propagation, *J. Geophys. Res.*, *102*, 8137–8157.

Ekström, G., A. M. Dziewoński, N. N. Maternovskaya, and M. Nettles (2005), Global seismicity of 2003: Centroid-moment-tensor solutions for 1087 earthquakes, *Phys. Earth Planet. Inter.*, *148*, 327–351.

Estey, L. H., and B. J. Douglas (1986), Upper mantle anisotropy: A preliminary model, *J. Geophys. Res.*, *91*(B11), 11,393–11,406.

Farnetani, C. G. (1997), Excess temperature of mantle plumes: The role of chemical stratification across D″, *Geophys. Res. Lett.*, *24*, 1583–1586.

Faul, U. H., and I. Jackson (2005), The seismological signature of temperature and grain size variations in the upper mantle, *Earth Planet. Sci. Lett.*, *234*, 119–134.

French, S. W., and B. A. Romanowicz (2014), Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography, *Geophys. J. Int.*, *199*, 1303–1327.

Gaboret, C., A. M. Forte, and J. Montagner (2003), The unique dynamics of the Pacific Hemisphere mantle and its signature on seismic anisotropy, *Earth Planet. Sci. Lett.*, *208*(3), 219–233.

Giardini, D., X.-D. Li, and J. H. Woodhouse (1987), Three-dimensional structure of the Earth from splitting in free-oscillation spectra, *Nature*, *325*, 405–411.

Gonnermann, H. M., M. Manga, and A. Mark Jellinek (2002), Dynamics and longevity of an initially stratified mantle, *Geophys. Res. Lett.*, *29*, 1399, doi:10.1029/2002GL014851.

Grocholski, B., K. Catalli, S.-H. Shim, and V. Prakapenka (2012), Mineralogical effects on the detectability of the postperovskite boundary, *Proc. Natl. Acad. Sci.*, *109*, 2275–2279.

Gu, Y. J., and A. M. Dziewonski (2002), Global variability of transition zone thickness, *J. Geophys. Res.*, *107*, 2135, doi:10.1029/2001JB000489.

Gu, Y. J., A. M. Dziewonski, W. Su, and G. Ekström (2001), Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities, *J. Geophys. Res.*, *106*, 11169–11199.

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

Gwinn, C. R., T. A. Herring, and I. I. Shapiro (1986), Geodesy by radio interferometry: Studies of the forced nutations of the Earth: 2. Interpretation, *J. Geophys. Res.*, *91*, 4755–4765.

Häfner, R., and R. Widmer-Schnidrig (2013), Signature of 3-D density structure in spectra of the spheroidal free oscillation 0S2, *Geophys. J. Int.*, *192*, 285–294.

Hager, B., R. Clayton, M. Richards, and R. Comer (1985), Lower mantle heterogeneity, dynamic topography and the geoid, *Nature*, *313*, 541–545.

Hansen, U., and D. A. Yuen (1988), Numerical simulations of thermal-chemical instabilities at the core-mantle boundary, *Nature*, *334*, 237–240.

Hastie, T., and R. Tibshirani (1990), *Generalized Additive Models*, Chapman and Hall, London.

Hayden, L. A., and E. B. Watson (2007), A diffusion mechanism for core-mantle interaction, *Nature*, *450*, 709–711.

Helffrich, G. R., and B. J. Wood (2001), The Earth's mantle, *Nature*, *412*, 501–507.

Hernlund, J. W., and C. Houser (2008), On the statistical distribution of seismic velocities in Earth's deep mantle, *Earth Planet. Sci. Lett.*, *265*, 423–437.

Herring, T. A., C. R. Gwinn, and I. I. Shapiro (1986), Geodesy by radio interferometry: Studies of the forced nutations of the Earth: 1. Data analysis, *J. Geophys. Res.*, *91*, 4745–4754.

Hirose, K., N. Takafuji, N. Sata, and Y. Ohishi (2005), Phase transition and density of subducted MORB crust in the lower mantle, *Earth Planet. Sci. Lett.*, *237*, 239–251.

Hirose, K., N. Sata, T. Komabayashi, and Y. Ohishi (2008), Simultaneous volume measurements of Au and MgO to 140 GPa and thermal equation of state of Au based on the MgO pressure scale, *Phys. Earth Planet. Inter.*, *167*, 149–154.

Hofmann, A. W. (2003), Sampling mantle heterogeneity through oceanic basalts: Isotopes and trace elements, in *Treatise on Geochemistry: The Mantle and Core*, edited by R. W. Carlson, H. D. Holland, and K. K. Turekian, pp. 61–91, Elsevier, New York.

Houser, C. (2007), Constraints on the presence or absence of post-perovskite in the lowermost mantle from long-period seismology, in *The Last Mantle Phase Transition, Geophys. Monogr. Ser.*, vol. 17, edited by K. Hirose et al., pp. 191–216, AGU, Washington, D. C.

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

Humler, E., J. Louis Thirot, and J. Paul Montagner (1993), Global correlations of mid-ocean-ridge basalt chemistry with seismic tomographic images, *Nature*, *364*, 225–228.

Isaak, D. G., O. L. Anderson, and R. E. Cohen (1992), The relationship between shear and compressional velocities at high pressures: Reconciliation of seismic tomography and mineral physics, *Geophys. Res. Lett.*, *19*, 741–744.

Ishii, M., and J. Tromp (1999), Normal-mode and free-air gravity constraints on lateral variations in velocity and density of Earth's mantle, *Science*, *285*, 1231–1236.

Ishii, M., and J. Tromp (2001), Even-degree lateral variations in the Earth's mantle constrained by free oscillations and the free-air gravity anomaly, *Geophys. J. Int.*, *145*, 77–96.

Kamada, S., H. Terasaki, E. Ohtani, T. Sakai, T. Kikegawa, Y. Ohishi, N. Hirao, N. Sata, and T. Kondo (2010), Phase relationships of the Fe-FeS system in conditions up to the Earth's outer core, *Earth Planet. Sci. Lett.*, *294*, 94–100.

Karato, S.-i (1993), Importance of anelasticity in the interpretation of seismic tomography, *Geophys. Res. Lett.*, *20*, 1623–1626.

Karato, S.-i., and B. B. Karki (2001), Origin of lateral variation of seismic wave velocities and density in the deep mantle, *J. Geophys. Res.*, *106*, 21,771–21,783.

Kellogg, L., B. Hager, and R. D. van der Hilst (1999), Compositional stratification in the deep mantle, *Science*, *283*(5409), 1881–1884.

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

Knittle, E., and R. Jeanloz (1989), Simulating the core-mantle boundary: An experimental study of high-pressure reactions between silicates and liquid iron, *Geophys. Res. Lett.*, *16*(7), 609–612.

Koelemeijer, P., A. Deuss, and J. Ritsema (2013), Observations of core-mantle boundary Stoneley modes, *Geophys. Res. Lett.*, *40*, 2557–2561, doi:10.1002/grl.50514.

Koelemeijer, P., J. Ritsema, A. Deuss, and H. J. van Heijst (2016), SP12RTS: A degree-12 model of shear- and compressional-wave velocity for Earth's mantle, *Geophys. J. Int.*, *204*, 1024–1039.

Komatitsch, D., and J. Tromp (2002a), Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.*, *149*, 390–412.

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

Komatitsch, D., J. Ritsema, and J. Tromp (2002), The spectral-element method, Beowulf computing, and global seismology, *Science*, *298*, 1737–1742.

Kuo, C., and B. Romanowicz (2002), On the resolution of density anomalies in the Earth's mantle using spectral fitting of normal-mode data, *Geophys. J. Int.*, *150*, 162–179.

Kustowski, B., G. Ekström, and A. M. Dziewoński (2007), Nonlinear crustal corrections for normal-mode seismograms, *Bull. Seismol. Soc. Am.*, *97*, 1756–1762.

Kustowski, B., G. Ekström, and A. M. Dziewonski (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *J. Geophys. Res.*, *113*, B06306, doi:10.1029/2007JB005169.

Labrosse, S., J. W. Hernlund, and N. Coltice (2007), A crystallizing dense magma ocean at the base of the Earth's mantle, *Nature*, *450*, 866–869.

Lancaster, P., and K. Salkauskas (1986), *Curve and Surface Fitting: An Introduction*, Academic Press, London.

Lassak, T. M., A. K. McNamara, and S. Zhong (2007), Influence of thermochemical piles on topography at Earth's core-mantle boundary, *Earth Planet. Sci. Lett.*, *261*, 443–455.

Lee, C.-T. A. (2003), Compositional variation of density and seismic velocities in natural peridotites at STP conditions: Implications for seismic imaging of compositional heterogeneities in the upper mantle, *J. Geophys. Res.*, *108*, 2441, doi:10.1029/2003JB002413.

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

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

Lin, S. C., and P. E. van Keken (2006), Dynamics of thermochemical plumes: 1. Plume formation and entrainment of a dense layer, *Geochem. Geophys. Geosyst.*, *7*, Q02006, doi:10.1029/2005GC001071.

Liu, Q., and J. Tromp (2006), Finite-frequency Kernels based on adjoint methods, *Bull. Seismol. Soc. Am.*, *96*, 2383–2397.

Liu, Q., and J. Tromp (2008), Finite frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.*, *174*, 265–286.

Liu, X., and A. Dziewonski (1998), Global analysis of shear wave velocity anomalies in the lower-most mantle, in *The Core-Mantle Boundary Region, Geodyn. Ser.*, vol. 28, edited by M. Gurnis et al., pp. 21–36, AGU, Washington, D. C.

Luh, P. C. (1974), Normal modes of a rotating, self-gravitating inhomogeneous Earth, *Geophys. J. Int.*, *38*, 187–224.

Maggi, A., E. Debayle, K. Priestley, and G. Barruol (2006), Azimuthal anisotropy of the Pacific region, *Earth Planet. Sci. Lett.*, *250*, 53–71.

Malcolm, A. E., and J. Trampert (2011), Tomographic errors from wave front healing: More than just a fast bias, *Geophys. J. Int.*, *185*, 385–402.

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

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 Atlantic to the Global Scale, Geophys. Monogr. Ser.*, vol. 117, edited by S. Karato et al., pp. 63–87, AGU, Washington, D. C.

Matas, J., and M. S. T. Bukowinski (2007), On the anelastic contribution to the temperature dependence of lower mantle seismic velocities, *Earth Planet. Sci. Lett.*, *259*, 51–65.

Mathews, P. M., and I. I. Shapiro (1992), Nutations of the Earth, *Annu. Rev. Earth Planet. Sci.*, *20*, 469–500.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, San Diego, Calif.

Montagner, J., and D. L. Anderson (1989), Petrological constraints on seismic anisotropy, *Phys. Earth Planet. Inter.*, *54*, 82–105.

Montagner, J.-P., and H.-C. Nataf (1986), A simple method for inverting the azimuthal anisotropy of surface waves, *J. Geophys. Res.*, *91*(B1), 511–520.

Morelli, A., and A. M. Dziewonski (1987), Topography of the core-mantle boundary and lateral homogeneity of the liquid core, *Nature*, *325*, 678–683.

Mosca, I., L. Cobden, A. Deuss, J. Ritsema, and J. Trampert (2012), Seismic and mineralogical structures of the lower mantle from probabilistic tomography, *J. Geophys. Res.*, *117*, B06304, doi:10.1029/2011JB008851.

Moulik, P., and G. Ekström (2014), An anisotropic shear velocity model of the Earth's mantle using normal modes, body waves, surface waves and long-period waveforms, *Geophys. J. Int.*, *199*, 1713–1738.

Mulyukova, E., B. Steinberger, M. Dabrowski, and S. V. Sobolev (2015), Survival of LLSVPs for billions of years in a vigorously convecting mantle: Replenishment and destruction of chemical anomaly, *J. Geophys. Res. Solid Earth*, *120*, 3824–3847, doi:10.1002/2014JB011688.

Murakami, M., K. Hirose, K. Kawamura, N. Sata, and Y. Ohishi (2004), Post-perovskite phase transition in MgSiO3, *Science*, *304*, 855–858.

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

Nishimura, C. E., and D. W. Forsyth (1989), The anisotropic structure of the upper mantle in the Pacific, *Geophys. J. Int.*, *96*, 203–229.

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

Obayashi, M., and Y. Fukao (1997), P and PcP travel time tomography for the core-mantle boundary, *J. Geophys. Res.*, *102*, 17,825–17,841.

Oganov, A. R., and S. Ono (2004), Theoretical and experimental evidence for a post-perovskite phase of MgSiO3 in Earth's D″ layer, *Nature*, *430*, 445–448.

Oganov, A. R., J. P. Brodholt, and G. D. Price (2002), Ab initio theory of phase transitions and thermoelasticity of minerals, in *Energy Modelling in Minerals*, vol. 4, edited by C. M. Gramaccioli, pp. 83–170, Eötvös Univ. Press, Budapest.

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

Park, J. (1986), Synthetic seismograms from coupled free oscillations: Effects of lateral structure and rotation, *J. Geophys. Res.*, *91*, 6441–6464.

Priestley, K., and D. McKenzie (2013), The relationship between shear wave velocity, temperature, attenuation and viscosity in the shallow part of the mantle, *Earth Planet. Sci. Lett.*, *381*, 78–91.

Raftery, A. E. (1995), Bayesian model selection in social research, *Sociol. Methodol.*, *25*, 111–164.

Resovsky, J., and M. Ritzwoller (1995), Constraining odd-degree Earth structure with coupled free-oscillations, *Geophys. Res. Lett.*, *22*, 2301–2304.

Resovsky, J., and M. Ritzwoller (1998), New and refined constraints on three-dimensional Earth structure from normal modes below 3 mHz, *J. Geophys. Res.*, *103*, 783–810.

Resovsky, J., and M. Ritzwoller (1999), Regularization uncertainty in density models estimated from normal mode data, *Geophys. Res. Lett.*, *26*, 2319–2322.

Resovsky, J., and J. Trampert (2003), Using probabilistic seismic tomography to test mantle velocity-density relationships, *Earth Planet. Sci. Lett.*, *215*, 121–134.

Ricolleau, A., J. P. Perrillat, G. Fiquet, I. Daniel, J. Matas, A. Addad, N. Menguy, H. Cardon, M. Mezouar, and N. Guignot (2010), Phase relations and equation of state of a natural MORB: Implications for the density profile of subducted oceanic crust in the Earth's lower mantle, *J. Geophys. Res.*, *115*, B08202, doi:10.1029/2009JB006709.

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

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

Robertson, G. S., and J. H. Woodhouse (1996), Ratio of relative S to P velocity heterogeneity in the lower mantle, *J. Geophys. Res.*, *101*, 20041–20052.

Rodgers, A., and J. Wahr (1993), Inference of core-mantle boundary topography from ISC PcP and PKP travel-times, *Geophys. J. Int.*, *115*, 991–1011.

Roecker, S., C. Thurber, K. Roberts, and L. Powell (2006), Refining the image of the San Andreas Fault near Parkfield, California using a finite difference travel time computation technique, *Tectonophysics*, *426*, 189–205.

Romanowicz, B. (2001), Can we resolve 3D density heterogeneity in the lower mantle, *Geophys. Res. Lett.*, *28*(6), 1107–1110.

Romanowicz, B., and L. Bréger (2000), Anomalous splitting of free oscillations—A reevaluation of possible interpretations, *J. Geophys. Res.*, *105*, 21,559–21,578.

Rosat, S., T. Sato, Y. Imanishi, J. Hinderer, Y. Tamura, H. McQueen, and M. Ohashi (2005), High-resolution analysis of the gravest seismic normal modes after the 2004 $M_W = 9$ Sumatra earthquake using superconducting gravimeter data, *Geophys. Res. Lett.*, *32*, L13304, doi:10.1029/2005GL023128.

Ruppert, D. (2012), Selecting the number of knots for penalized splines, *J. Comput. Graphical Stat.*, *11*, 735–757.

Schubert, G., G. Masters, P. Olson, and P. Tackley (2004), Superplumes or plume clusters?, *Phys. Earth Planet. Inter.*, *146*, 147–162.

Schuberth, B. S. A., C. Zaroli, and G. Nolet (2012), Synthetic seismograms for a synthetic Earth: Long-period P- and S-wave traveltime variations can be explained by temperature alone, *Geophys. J. Int.*, *188*, 1393–1412.

Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*, 461–464.

Sieminski, A., J. J. Lévêque, and E. Debayle (2004), Can finite-frequency effects be accounted for in ray theory surface wave tomography?, *Geophys. Res. Lett.*, *31*, L24614, doi:10.1029/2004GL021402.

Simmons, N. A., A. M. Forte, L. Boschi, and S. P. Grand (2010), GyPSuM: A joint tomographic model of mantle density and seismic wave speeds, *J. Geophys. Res.*, *115*, B12310, doi:10.1029/2010JB007631.

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

Solomatov, V. S., and D. J. Stevenson (1993), Suspension in convective layers and style of differentiation of a terrestrial magma ocean, *J. Geophys. Res.*, *98*, 5375–5390.

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

Steinberger, B., and A. R. Calderwood (2006), Models of large-scale viscous flow in the Earth's mantle with constraints from mineral physics and surface observations, *Geophys. J. Int.*, *167*, 1461–1481.

Steinberger, B., and R. Holme (2008), Mantle flow models with core-mantle boundary constraints and chemical heterogeneities in the lowermost mantle, *J. Geophys. Res.*, *113*, B05403, doi:10.1029/2007JB005080.

Stixrude, L., and C. Lithgow-Bertelloni (2011), Thermodynamics of mantle minerals—II. Phase equilibria, *Geophys. J. Int.*, *184*, 1180–1213.

Su, W., and A. M. Dziewoński (1997), Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth Planet. Inter.*, *100*, 135–156.

Sze, E. K. M., and R. D. van der Hilst (2003), Core mantle boundary topography from short period PcP, PKP, and PKKP data, *Phys. Earth Planet. Inter.*, *135*, 27–46.

Tackley, P. J. (1998), Three-dimensional simulations of mantle convection with a thermo-chemical basal boundary layer: D″?, in *The Core-Mantle Boundary Region*, pp. 231–253, AGU, Washington, D. C.

Tackley, P. J. (2007), Mantle geochemical geodynamics, in *Treatise of Geophysics: Mantle Dynamics*, vol. 7, edited by P. J. Tackley, pp. 437–505, Elsevier, Amsterdam.

Tackley, P. J. (2011), Living dead slabs in 3-D: The dynamics of compositionally-stratified slabs entering a "graveyard" above the core-mantle boundary, *Phys. Earth Planet. Inter.*, *188*(3-4), 150–162.

Tackley, P. J. (2012), Dynamics and evolution of the deep mantle resulting from thermal, chemical, phase and melting effects, *Earth Sci. Rev.*, *110*, 1–25.

Tan, E., and M. Gurnis (2007), Compressible thermochemical convection and application to lower mantle structures, *J. Geophys. Res.*, *112*, B06304, doi:10.1029/2006JB004505.

Tarantola, A. (1984), Linearized inversion of seismic reflection data, *Geophys. Prospect.*, *32*, 998–1015.

Tarantola, A., and B. Valette (1982), Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.*, *20*, 219–232.

Thurber, C. H. (1992), Hypocenter-velocity structure coupling in local earthquake tomography, *Phys. Earth Planet. Inter.*, *75*, 55–62.

Trampert, J. (2004), Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science*, *306*, 853–856.

Trampert, J., and J. Spetzler (2006), Surface wave tomography: Finite-frequency effects lost in the null space, *Geophys. J. Int.*, *164*, 394–400.

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

Trefethen, L., and D. Bau (1997), *Numerical Linear Algebra*, Soc. Ind. Appl. Math., Philadelphia, Pa.

Tromp, J., C. Tape, and Q. Liu (2005), Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, *160*, 195–216.

Tsuchiya, T., J. Tsuchiya, K. Umemoto, and R. M. Wentzcovitch (2004), Phase transition in MgSiO3 perovskite in the Earth's lower mantle, *Earth Planet. Sci. Lett.*, *224*, 241–248.

van der Hilst, R. D., and M. V. de Hoop (2005), Banana-doughnut kernels and mantle tomography, *Geophys. J. Int.*, *163*, 956–961.

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

Wang, Y., and L. Wen (2007), Geometry and P and S velocity structure of the "African Anomaly", *J. Geophys. Res.*, *112*, B05313, doi:10.1029/2006JB004483.

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

Wielandt, E. (1987), On the validity of the ray approximation for interpreting delay times, in *Seismic Tomography*, edited by G. Nolet, pp. 85–98, D. Reidel, Dordrecht, Netherlands.

Woodhouse, J. H., and A. M. Dziewoński (1984), Mapping the upper mantle: Three-dimensional modeling of Earth structure by inversion of seismic waveforms, *J. Geophys. Res.*, *89*, 5953–5986.

Woodhouse, J. H., and Y. K. Wong (1986), Amplitude, phase and path anomalies of mantle waves, *Geophys. J. Int.*, *87*, 753–773.

Woodward, R. L., and G. Masters (1991), Global upper mantle structure from long-period differential travel times, *J. Geophys. Res.*, *96*, 6351–6377.

Wookey, J., S. Stackhouse, J. M. Kendall, J. Brodholt, and G. D. Price (2005), Efficacy of the post-perovskite phase as an explanation for lowermost-mantle seismic properties, *Nature*, *438*, 1004–1007.

Ye, J. (2012), On measuring and correcting the effects of data mining and model selection, *J. Am. Stat. Assoc.*, *93*, 120–131.

Zhong, S., and B. H. Hager (2003), Entrainment of a dense layer by thermal plumes, *Geophys. J. Int.*, *154*, 666–676.

Zindler, A., and S. Hart (1986), Chemical geodynamics, *Annu. Rev. Earth Planet. Sci.*, *14*, 493–571.