LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

LLNL-JRNL-426712

GyPSuM: A Detailed Tomographic Model of Mantle Density and Seismic Wave Speeds

N.A. Simmons, A.M. Forte, L. Boschi, S.P. Grand

March 31, 2010

Journal of Geophysical Research

Disclaimer

GyPSuM: A Detailed Tomographic Model of Mantle Density and Seismic Wave Speeds

Nathan A. Simmons[1], Alessandro M. Forte[2], Lapo Boschi[3], Stephen P. Grand[4]

[1] *Atmospheric, Earth & Energy Division, Lawrence Livermore National Laboratory, 7000 East Ave, L-046, Livermore CA, 94551, USA*
[2] *GEOTOP - Dépt. Sci. Terre & Atmosphère, Université du Québec à Montréal, CP 8888 succursale Centre-Ville, Montréal QC, H3C 3P8, Canada*
[3] *Eidgenössische Technische Hochschule Hönggerberg, Institute of Geophysics, Schafmattstr. 30, 8093 Zurich, Switzerland*
[4] *Jackson School of Geosciences, University of Texas at Austin, Austin TX, 78712, USA*

Abstract

GyPSuM is a tomographic model of mantle seismic shear wave (S) speeds, compressional wave (P) speeds and detailed density anomalies that drive mantle flow. The model is developed through simultaneous inversion of seismic body wave travel times (P and S) and geodynamic observations while considering realistic mineral physics parameters linking the relative behavior of mantle properties (wave speeds and density). Geodynamic observations include the (up to degree 16) global free-air gravity field, divergence of the tectonic plates, dynamic topography of the free surface, and the flow-induced excess ellipticity of the core-mantle boundary. GyPSuM is built with the philosophy that heterogeneity that most closely resembles thermal variations is the simplest possible solution. Models of the density field from Earth's free oscillations have provided great insight into the density configuration of the mantle; but are limited to very long-wavelength solutions. Alternatively, simply scaling higher resolution seismic images to density anomalies generates density fields that do not satisfy geodynamic observations. The current study provides detailed density structures in the mantle while directly satisfying geodynamic observations through a joint seismic-geodynamic inversion process. Notable density field observations include high-density piles at the base of superplume

structures, supporting the fundamental results of past normal mode studies. However, these features are more localized and lower amplitude than past studies would suggest. When we consider all seismic anomalies in GyPSuM, we find that P and S-wave speeds are strongly correlated throughout the mantle. However, correlations between the high-velocity S zones in the deep mantle (~2000 km depth) and corresponding P-wave anomalies are very low suggesting a systematic divergence from simplified thermal effects in ancient subducted slab anomalies. Nevertheless, we argue that temperature variations are the primary cause of P-wave, S-wave, and density anomalies in the mantle.

1. Introduction

Evaluating the relative behavior of various mantle properties is a powerful way to identify compositional variations and processes occurring in the mantle [e.g., *Robertson and Woodhouse* 1996; *Su and Dziewonski* 1997; *Kennett et al.* 1998; *van der Hilst and Kárason* 1999; *Ishii and Tromp* 1999, 2004; *Masters et al.* 2000; *Saltzer et al.* 2001; *Kennett and Gorbatov* 2004; *Trampert et al.* 2004]. It has been well established that correlations of shear and bulk sound speeds reduce in the deep mantle (lowest ~1000 km) and are in fact anti-correlated in some regions [e.g. *Su and Dziewonski* 1997; *Kennett et al.* 1998; *Masters et al.* 2000]. Moreover, studies incorporating normal mode splitting functions have provided evidence that high-density piles exist at the base of the mantle where 'superplumes' may originate thereby contradicting the usual assumption that deep mantle heterogeneities are driven solely by temperature variations [e.g., *Ishii and Tromp* 1999; *Trampert et al.* 2004]. Based upon very long-wavelength images of the mantle, it has been argued that chemical variations dominate thermal heterogeneity and, therefore, buoyancy forces in the deep mantle [*Trampert et al.* 2004].

However, density and wave speed heterogeneities derived from normal-mode splitting data, that are only sensitive to the longest length-scale heterogeneities, provide a limited understanding of the structure and dynamics of Earth's mantle. These data appear to be subject to considerable trade-offs and non-uniqueness [*Kuo and Romanowicz* 2002] and, moreover, the inferred density anomalies provide poor fits to fundamental long-wavelength surface geodynamic constraints [*Soldati et al.* 2009].

In order to understand the dynamics of the mantle, we must first acquire detailed density anomalies that drive mantle flow. One way to estimate detailed density anomalies in the mantle is through the translation of seismically derived models with mineral physics relationships. Most often, the resulting density models using this approach provide less than optimal fits to key geodynamic observations such as the global free-air gravity anomalies derived from satellite data [*Forte* 2007]. Thus, this approach requires the introduction of strong compositional effects leading to potential overestimates of non-thermal contributions to mantle heterogeneity [*Simmons et al.* 2009]. Although geodynamic observations are sensitive to global integrals of 3-D anomalies throughout the mantle, these observations alone cannot adequately resolve local density structure of the mantle. Therefore, independent seismic information is required to help define the local 3-D distribution of heterogeneity in the mantle. In our previous studies [*Simmons et al.* 2006, 2007, 2009], we combined geodynamic observations and S-wave travel times in global-scale simultaneous inversions for shear wave speed and density. In these previous studies, both types of information directly aided in the resolution of the 3-D structures as opposed to estimating density anomalies through the *a posteriori* scaling of a pure seismically derived model.

74 The benefit of joint inversion of multiple types of data is multifold. Firstly, joint
75 inversion directly accounts for the variable resolution of individual data sets that might produce
76 quite different distributions of mantle heterogeneities when considered alone. In addition,
77 simultaneous inversion of different forms of information allows for the determination of multiple
78 mantle properties that are most consistent with one another given some underlying hypothesis
79 such as the dominance of thermal effects. Moreover, we can more accurately evaluate the
80 relative behavior of mantle properties as determined through a direct joint inversion process
81 since the model parameterization and roughness levels are equivalent. In other words, the joint
82 inversion process removes a number of biases that would potentially lead to unnecessarily large
83 degrees of compositional influence to explain multiple observations simultaneously.
84 In this paper, we present the GyPSuM model (G=*Geodynamic*, P=*Compressional waves*,
85 S=*Shear waves*, M=*Mineral physics*) which represents the next step in the evolution of a detailed
86 multi-component mantle model presented in our previous studies [*Simmons et al.* 2006, 2007,
87 2009]. As in our previous work, GyPSuM is constructed through the simultaneous inversion of
88 seismic and geodynamic constraints using mineral physics relationships that relate the mantle
89 properties. Also similar to our previous model developments, we simultaneously consider
90 globally distributed S-wave arrival times, the global free-air gravity field, divergence of the
91 tectonic plates, dynamic topography of Earth's free surface, and the flow-driven excess ellipticity
92 of the core-mantle boundary (CMB). The most important improvement from our previous
93 studies is the incorporation of globally distributed P-wave travel time measurements, thus
94 allowing the generation of a detailed, 3-component mantle model (density, P-wave velocity, S-
95 wave velocity). We also perform non-linear inversions to gradually adjust heterogeneities and

96 mineral physics relationships rather than the pure linear approach employed in the
97 aforementioned previous studies.
98
99 **2. Seismic and Geodynamic Data**
100 The seismic data considered consist of globally distributed, teleseismic S-wave and P-
101 wave travel time observations (Table 1) that are not restricted to common source-receiver pairs.
102 The S-wave observations consist of ~46,000 travel time residuals derived from waveform signals
103 filtered to 0.01-0.07 Hz. The seismic phases include *S, ScS, sS, sScS, SKS* and *SKKS* phases
104 including surface-reflected multiples (e.g. *sSS*) and triplicated phases turning within the upper
105 mantle [*Grand* 1994, 2002; *Grand et al.* 1997; *Simmons et al.* 2006, 2007, 2009]. Travel time
106 residuals are relative to a 1-D model consisting of an average of the TNA/SNA models in the
107 upper mantle [*Grand and Helmberger* 1984] and PREM in the lower mantle [*Dziewonski and*
108 *Anderson* 1981]. Corrections for crustal structure and ellipticity are based upon the CRUST5.1
109 model of *Mooney et al.* [1998] and the techniques developed by *Dziewonski and Gilbert* [1976],
110 respectively. Earthquake locations were determined through an iterative process whereby
111 portions of the model space were determined followed by relocation in several steps [*Simmons et*
112 *al.* 2006]. We view these data as highly reliable given that they are based upon detailed analyses
113 of synthetic waveform correlations and, when necessary, detailed event modeling. Also, given
114 the diverse suite of phases reflecting and refracting through the mantle and our selection of
115 evaluated earthquakes, the coverage of the mantle is maximized. For more information regarding
116 these data (measurement procedures, coverage, etc.) see *Grand* [1994, 2002] and *Simmons et al.*
117 [2006, 2007].

118 The P-wave observations consist of ~626,000 summary travel time residuals computed in
119 *Antolik et al.* [2003]. The underlying data come from the recompilation and relocation of the
120 International Seismological Centre (ISC) direct P-wave arrivals performed by *Engdahl et al.*
121 [1998] (EHB). The seismic events were relocated on the basis of the S&P12/WM13 3-D model
122 [*Su and Dziewonski* 1993] and subsequently summarized on a 2x2 degree global grid and 50-100
123 km event depth increments [*Antolik et al.* 2003]. The original ISC arrivals are not based on the
124 more reliable waveform correlation techniques employed in the generation of the S-wave data
125 set; however, the level of scrutiny involved in the grooming of the P-wave data by *Engdahl et al.*
126 [1998], the summary-data analysis of *Antolik et al.* [2003], and the large amount of data yields a
127 valuable set of global P-wave observations. Similar to the corrections applied to the shear wave
128 data, these data were corrected on the basis of CRUST5.1 as well as the ellipticity of Earth.
129 The set of geodynamic constraints we employ consists of a suite of convection-related
130 observables sensitive to the viscosity of the mantle, the style of mantle flow, and 3-D density
131 variations. These surface observables include the free-air gravity field from the EGM96
132 geopotential model derived through a compilation of land-, air-, and space-based observations
133 [*Lemoine et al.* 1998]. In addition, we employ constraints on the motion of the tectonic plates in
134 the form of plate divergence calculated from the NUVEL-1 plate velocity model [*DeMets et al.*
135 1990]. The motion of tectonic plates are coupled to the underlying mantle flow driven by
136 density anomalies and are thus important constraints on 3-D density heterogeneity in the upper
137 and middle mantle. We also consider the dynamic topography of Earth's free surface estimated
138 by removal of the crustal isostatic topography signal [*Forte and Perry* 2000]. These data are the
139 least robust in the suite of geodynamic observations given the uncertainties of the global crustal
140 structure and rheologies; however, crust-corrected dynamic surface topography is a direct

141 mapping of vertical stresses in the mantle and thus provides important constraints on the range of
142 upper-mantle density configurations needed to explain the other geodynamic observations
143 considered. Large-scale mantle flow driven by density anomalies has a direct impact on the
144 overall shape of the CMB [see *Forte et al.* 1995]. Therefore, we also incorporate the excess
145 ellipticity of the CMB observed from Earth's free-core nutation processes as an additional
146 constraint on mantle density [*Herring et al.* 2002; *Mathews et al.* 2002]. These studies have
147 found that there is ~400 m of excess bulge of the CMB along the equator and we presume that
148 this bulge is in response to mantle flow above the boundary. See Table 1 for a summary of all of
149 the constraints employed in this study.
150
151 **3. Forward Model**
152 The forward model consists of a large and diverse set of linear equations to be solved
153 using inversion algorithms. The first step in developing these sets of linear equations is defining
154 the model space. We divided the mantle into blocks that are ~275x275 km in the lateral
155 dimension and 22 layers ranging from 75 to 240 km thick, providing a total 99,148 blocks
156 (model parameters). The seismic sensitivities to the model space were computed through 1-D
157 ray tracing (infinite frequency approximation) and compiled into matrices to form a set of
158 sensitivity (Fréchet) kernels for both the S and P data sets. The individual sets of kernels provide
159 the basis for the linear equations relating travel times and seismic slowness perturbations in the
160 typical way.
161 Geodynamic sensitivity kernels were developed through the theoretical linear relationship
162 between mantle density anomalies and each of the surface observables (free-air gravity field,
163 plate divergences, dynamic surface topography, and excess CMB ellipticity). This requires an

164 analytical description of the viscous flow response of the mantle to internal point sources of
165 density for each convection-related observation [*Richards and Hager* 1984; *Ricard et al.* 1984;
166 *Forte and Peltier* 1987]. We computed viscous flow responses for a compressible and
167 gravitationally consistent mantle whereby tectonic plate motions are dynamically coupled to the
168 underlying mantle flow [*Forte and Peltier* 1994; *Forte* 2007]. A combination of mixed free-slip
169 and no-slip surface boundary conditions were incorporated to calculate the responses assuming
170 the radially-symmetric viscosity profile derived from joint inversion of convection data and
171 glacial isostatic adjustment observations [*Mitrovica and Forte* 2004]. Based on the analysis of
172 *Simmons et al.* [2006], the responses were determined with the assumption that a whole-mantle
173 style of flow prevails. Therefore, no strict boundaries to vertical mass transport are incorporated
174 into building the responses with exception of the CMB and free surface. Each of the spherical
175 harmonic components of the geodynamic data fields has a unique sensitivity to the mantle
176 (Figure 1). Thus, simultaneously considering the entire suite of geodynamic observations
177 provides significant constraints on the 3-D arrangement of density anomalies that may exist in
178 the mantle.

179 Expressing the seismic and geodynamic observations as a single set of linear equations
180 yields a very large and complex system best described in the following matrix form:

181
$$\begin{bmatrix} \mathbf{L}_S \\ \lambda_P \mathbf{L}_P\left(R_{P/S}, V_S/V_P\right) \\ \lambda_G \mathbf{G}\left(R_{\rho/S}, -V_S\right) \\ \lambda_{CMB}\mathbf{c}\left(R_{\rho/S}\right) \\ \lambda_D \mathbf{D} \end{bmatrix} \Delta\mathbf{m}_S = \begin{bmatrix} \mathbf{r}_S \\ \lambda_P \mathbf{r}_P \\ \lambda_G \mathbf{g} \\ \lambda_{CMB} e \\ \mathbf{0} \end{bmatrix} \qquad (1)$$

182 where the notation is explained in Table 2. In Equation (1), the goal is to solve for a shear wave
183 slowness perturbation model ($\Delta\mathbf{m}_S$). Therefore, the P-wave and geodynamic sensitivities are

184 associated to shear wave slowness perturbations through relative heterogeneity ratios ($R_{\rho/S}$ and
185 $R_{P/S}$) obtained from mineral physics investigations. Additionally, conversions relating velocity
186 perturbations (ΔV) to slowness perturbations (Δs) must also be propagated through the
187 sensitivity kernel matrices using the approximation $\Delta V / V_0 \approx -V_0 \Delta s$ (where V_0 is the starting
188 model velocity).

189 One of the major issues when attempting to simultaneously invert multiple forms of data
190 is determining the relative weights to each set of observations. It might seem reasonable to
191 dramatically scale the P-wave data set down relative to the S-wave data set given the ~14-fold
192 increase in the number of data. Simply defining λ_P on the basis of number observations is not
193 appropriate since direct P-wave residual travel times are systematically smaller than the S-wave
194 residuals. This is especially true when considering S-wave phases that travel through the high-
195 amplitude anomalies in the upper mantle multiple times (e.g. *SSS*) producing very large residual
196 travel times (signals). Therefore, we chose these values on the basis of the relative data norms:

197 $$\lambda_P \equiv \frac{\|\mathbf{r}_S\|}{\|\mathbf{r}_P\|} \ . \qquad (2)$$

198 Equation (2) yields $\lambda_P \approx 0.7$ and provides an even balance of influence between these
199 independent seismic data sets based on our extensive testing. The weighting of the geodynamic
200 observations is more difficult since these data are represented as spherical harmonic coefficients
201 in constrast to travel time residuals. Based on our previous joint investigations [*Simmons et al.*
202 2006, 2007, 2009], the optimum weighting of the geodynamic observations relative to the shear
203 wave data (λ_G) was found to be ≈ 1000. In the current study, we chose $\lambda_G = 2 \times 1000 = 2000$ to
204 account for the addition of the P-wave observations providing an approximately equivalent

205 influence of geodynamic and seismic observations. λ_{CMB} was chosen to be sufficiently large in
206 order to fully match the observed CMB excess ellipticity similar to the treatment of *Simmons et*
207 *al.* [2009]. The following sections describe our multi-step approach to solving this complicated
208 system of equations.
209
210 **4. Density-P-S Coupling in a Thermal Scenario**
211 Adopting the philosophy of the joint seismic-geodynamic modeling approach in *Simmons*
212 *et al.* [2009], we chose to initially attribute mantle heterogeneities to temperature variations in
213 the mantle and subsequently relax this requirement. This approach involves first determining the
214 optimum scaling relationships ($R_{\rho/S}$ and $R_{P/S}$) that account for the relative behavior of mantle
215 properties when temperature variations are the dominant cause. One of the major issues is the
216 large range of uncertainty of these parameters even when only considering the effects of
217 temperature (see *Cammarano et al*. [2003] and *Karato and Karki* [2001]). To aid in the search
218 for the optimum scaling relationships, it is useful to evaluate how these ratios are related. Using
219 the representation of *Karato and Karki* [2001], the heterogeneity ratios linking 1) density to
220 shear speed variations, and 2) sound speed to shear speed variations can be written:

221
$$R_{\rho/S} = \frac{2}{(\Gamma - 1) + 2Q_S^{-1} X(\xi) C}$$
$$R_{\phi/S} = \frac{\delta_S - 1}{(\Gamma - 1) + 2Q_S^{-1} X(\xi) C} = \frac{\delta_S - 1}{2} R_{\rho/S} \qquad (3)$$

222 where ρ , S , and ϕ correspond to density, shear wave speed, and sound wave speed
223 respectively. The parameters Γ and δ_S are the Anderson-Grüneisen parameters [e.g. *Anderson*
224 1989] relating differential changes of the elastic moduli to density variations, and thus represent

the pure elastic effects to the ratios in Equation (3). Q_S is the shear wave quality factor and $X(\xi)C$ is a correction factor for anelasticity (*see Karato and Karki* [2001] for a thorough description). Therefore, $2Q_S^{-1}X(\xi)C$ represents the anelastic component to the relative heterogeneity ratios. If we also consider the relationship between the three types of seismic wave speeds, we can readily derive a relationship between $R_{\rho/S}$ and $R_{P/S}$:

$$\frac{\delta V_P}{V_P} = \gamma \frac{\delta V_S}{V_S} + (1-\gamma)\frac{\delta V_\phi}{V_\phi}, \qquad \gamma = \frac{4}{3}\frac{V_S^2}{V_P^2}$$
$$\Rightarrow R_{P/S} = \frac{(\delta_S - 1)}{2} R_{\rho/S}(1-\gamma) + \gamma \tag{4}$$

Note that this simple formulation is identical to *Karato and Karki* [2001] with the assumption that bulk attenuation is negligible, yielding $Q_S / Q_P = \gamma$. Equation (4) shows that for simple thermal variations in the mantle, $R_{\rho/S}$ scales to $R_{P/S}$ in a predictable way. Therefore, Equation (4) represents full $\rho - V_S - V_P$ coupling for thermally induced variations of iso-chemical mantle material. If we assume that δ_S is a fixed value at any given depth, and velocity variations are relatively small, we need to only determine one ratio ($R_{\rho/S}$ or $R_{P/S}$) and directly compute the other, thereby simplifying the optimization problem.

Simmons et al. [2009] determined that a simple 1-D representation of $R_{\rho/S}$ was not adequate to account for the relative behavior of density and shear wave velocity variations in the mantle. This conclusion was based on the inability to simultaneously explain geodynamic and seismic shear wave data with radially symmetric heterogeneity ratio profiles. There are multiple reasons for this, including: 1) cratonic keels are compositionally distinct, and 2) $R_{\rho/S}$ values in the upper mantle are affected by the background temperature conditions due the temperature-

244 dependence of Q_S. In order to account for these 1st-order effects, *Simmons et al.* [2009]
245 introduced scaling model correction derivatives of the form:

246 $$\kappa = \frac{\partial R_{\rho/S}}{\partial \ln V_S}. \qquad (5)$$

247 The search for the optimum κ values yielded highly negative values in the cratonic roots and
248 slightly positive numbers in the non-cratonic ('thermal') upper mantle. Therefore, $R_{\rho/S}$ values
249 in fast, cratonic zones were systematically reduced yielding less negative and sometimes positive
250 buoyancies in accord with the iron depletion hypothesized in these mantle regions [e.g. *Jordan*
251 1978]. Similarly, $R_{\rho/S}$ values in the low-velocity, non-cratonic upper mantle were significantly
252 reduced in agreement with the expected behavior due to the temperature-dependence of Q_S.
253
254 **5. Modeling Procedures**
255 In our joint modeling approach, we initially performed inversions with simple scaling
256 relationships thereby forcing the geographic patterns of the heterogeneity fields (density and
257 wave speeds) to be spatially correlated, which is likely an oversimplification [e.g. *Masters et al.*
258 2000]. These geographic patterns of the model property fields are thus allowed to deviate from
259 one another in a systematic manner thereby producing heterogeneity models most consistent in
260 pattern and relative amplitudes with some underlying hypothesis (e.g. thermal dominance).
261 Therefore, the scaling relationships between the model fields (velocities and density) evolve
262 from simplified 1-D profiles to fully three-dimensional relationships.
263 Before proceeding to the joint inversion, we tested the ability of simple seismic models
264 (derived entirely from seismic data alone) to match the other observations considered in this

265 study. We selected the optimum $R_{\rho/S}^{1D}$ profile (Figure 2) as well as the TX2008s shear wave
266 model from the work of *Simmons et al.* [2009] for evaluation. We note that, although a jointly
267 derived solution (TX2008j) was produced in *Simmons et al.* [2009], we chose to use the pure
268 seismically derived version to demonstrate the potential difficulties arising from the of shear
269 wave data alone. Equation (4) was used to calculate $R_{P/S}^{1D}$ assuming simple depth-dependent δ_S
270 and γ computed from the starting 1-D shear and compressional models (described previously).
271 In the upper mantle, bulk δ_S values were estimated by comparing the $R_{\rho/S}^{1D}$ profile from
272 *Simmons et al.* [2009] and the corresponding $R_{P/S}^{1D}$ profile computed in *Cammarano et al.* [2003]
273 yielding δ_S values between 2.2 and 3.8. In the lower mantle, we estimated δ_S to be linearly
274 decreasing from 2.7 (top of the lower mantle) to 1.3 (base of the mantle) based on the values for
275 $MgSiO_3$ perovskite presented in *Karato and Karki* [2001].
276 Applying the 1-D scaling profiles to the TX2008s shear wave model, we calculated the
277 variance reduction fit to all of the considered data fields (Table 3). The scaled TX2008s model is
278 incapable of satisfying the geodynamic constraints (with any 1-D scaling model) since it is
279 derived solely with shear wave constraints as pointed out in *Simmons et al.* [2009]. This scaled
280 shear wave model also provides a poor fit to the P-wave constraints employed in the current
281 study (~12% variance reduction). To gain perspective on this measure of P-wave data fit, we
282 performed inversions considering only the P-wave data and with the same parameterization and
283 inversion techniques used to develop TX2008s. The inversions were carried out using the
284 iterative LSQR algorithm [*Paige and Saunders* 1982] with a spectrum of smoothing weights.
285 The optimum P-wave model (henceforth referred to as P-only) provided ~32% variance
286 reduction fit to the P-wave data set. This relatively low degree of potential fit (compared to the

S-wave data set and model) is, in part, a product of the signal-to-noise ratio of the residual P-wave travel times and is comparable to studies using the same data (e.g. *Antolik et al.* [2003]). The P-only model, when scaled to shear velocities and densities (with $R_{\rho/S}^{ID}$ and $R_{P/S}^{ID}$), provides a poor level of fit to the shear wave data (~56%) and remarkably low degrees of fit to the geodynamic observations as well (Table 3). Clearly, we could likely find a better set of 1-D conversion factors to scale these independently produced models through a trial-and-error process. However, as shown in *Simmons et al.* [2009], selecting the best-fitting scaling model on the basis of a heterogeneity model derived from a single type of data does not necessarily lead to better models after a joint inversion is performed. These tests mainly serve to provide some insight into the potential fit that may be achieved with our model parameterization and they also provide reference, baseline fits for subsequent comparisons.

The initial step in the 3-component model construction is the joint inversion for density and wave speeds assuming the set of simple 1-D scaling models ($R_{\rho/S}^{ID}$ and $R_{P/S}^{ID}$) previously described. The primary reasons for initially solving the system in Equation (1) assuming 1-D scaling profiles are to: 1) develop an unbiased starting model for scaling model optimization, and 2) establish the appropriate level of model roughness when considering the combined data set. The joint solution with 1-D scaling models ('GyPSuM_1D') provides a more balanced level of fit to the seismic data sets than the independently produced models (variance reduction fit of 90% to the S-wave data and 28% to P-wave data set). In addition, the fits to the geodynamic observations are dramatically improved with the exception of the dynamic topography which is still poorly matched (Table 3). The problem of simultaneously reconciling the dynamic topography and gravity fields has been recognized for quite some time (e.g. *Forte et al.* 1993; *Le Stunff and Ricard* 1995). In the context of joint seismic-geodynamic inversion, this problem

310 persists when assuming simple 1-D scaling profiles that force the geographic pattern of wave
311 speeds and density to be identical. However, when directly considering the impact of iron
312 depletion of the cratons and the temperature dependence of Q_S through implementation of
313 scaling model corrections (i.e. Equation 5), the level of fit to the dynamic topography field can
314 be greatly improved while maintaining a good fit to the observed gravity field [*Simmons et al.*
315 2009]. The search for these scaling model correction terms as well as other free parameters is
316 described in the following section.
317
318 **5.1 Scaling Model Optimization (1.5-D *R* values)**
319 One major limitation of the *Simmons et al.* [2009] investigation was that the
320 determination of the velocity-based correction derivatives (κ terms) were based upon a fixed
321 (i.e. given) velocity model. In reality, this is a non-linear problem since the determination of
322 shear velocity structure in a joint modeling process is a function of the corrected $R_{\rho/S}$ values and
323 vice versa. In the current study, we wish to find the optimum 1-D scaling models ($R_{\rho/S}$) and
324 correction derivatives (e.g. Equation 5) while considering the non-linearity of the problem. We
325 are also incorporating P-wave constraints to yield a 3-component model of mantle heterogeneity.
326 Therefore, the selection of $R_{\rho/S}$ impacts the calculated values of $R_{P/S}$ (via Equation 4) and our
327 ability to simultaneously fit the P-wave data set as well.
328 Given these additional complexities, we have formulated the scaling model optimization
329 problem by defining a parametric form of $R_{\rho/S}$ in the following way:

330 $$R_{\rho/S} = aR_{\rho/S}^{1\text{-}D} + b + \kappa\delta \ln V_S \qquad (6)$$

where $R_{\rho/S}^{1\text{-}D}$ is a starting 1-D scaling model and κ is a velocity-based scaling correction derivative defined in Equation (5). The coefficients a and b represent the amplification and shifting of the starting 1-D scaling model, respectively. Therefore, a controls the overall shape of the 1-D portion of scaling factor profile, b controls the mean value, and κ adjusts the scaling factor model according to the underlying shear wave velocity structure. We will further refer to the dimension of such a scaling model as 1.5-D. Equation (6) presents the basic form of the optimization problem; however, the coefficients are regionally dependent. Specifically, we define two sets of coefficients ($[a_{um}, b_{um}]$ and $[a_{lm}, b_{lm}]$) that adjust the upper (subscript '*um*') and lower mantle (subscript '*lm*') scaling models independently. Additionally, we must define multiple κ terms to account for the cratonic versus the non-cratonic upper mantle zones as well as potential depth dependence of these values. In total, we consider 9 free parameters described in Table 4.

To appropriately test the validity of a single set of these 9 free parameters, we must perform a non-linear joint inversion involving the iterative updating of scaling models and global heterogeneity models (i.e. solve Equation 1). A grid search or Monte Carlo approach would be computationally daunting and we have therefore adapted and modified a very fast simulated annealing (VFSA) approach [*Ingber* 1989; *Jackson et al.* 2004] to identify the optimum set of scaling model adjustment parameters described in Equation (6) and Table 4. If we denote the i^{th} scaling model adjustment parameter as m_i and the iteration number as *k*, an updated model parameter can be written:

$$m_i^{new} = m_i^k + y_i(m_i^{\max} - m_i^{\min}). \qquad (7)$$

352 The update parameter, y_i, is drawn from a Cauchy distribution based on a randomly generated

353 number ($u_i \in [0,1]$) and a value analogous to temperature in an annealing process, T_k:

354 $$y_i = sign(u_i - 0.5) T_k \left[\left(1 + \frac{1}{T_k} \right)^{|2u_i - 1|} - 1 \right]. \quad (8)$$

355 The broadness of the Cauchy distribution is controlled by the temperature in that, at high

356 temperatures, the potential for a high-amplitude value to be drawn is greater than at lower

357 temperatures. After each iteration, the temperature drops according to a cooling schedule to

358 converge towards a solution. For our purposes, we chose a relatively rapid cooling schedule to

359 limit the number of full joint inversions required:

360 $$T_k = T_0 e^{-0.99 k^{1/2}}. \quad (9)$$

361 We limited the range of possible 1-D components of the upper mantle scaling factors to the range

362 calculated from a recent mineral physics compliation along the 1300°C adiabat [*Cammarano et*

363 *al.* 2003] and the lower mantle profile was limited to the range defined by a suite of lower mantle

364 profiles [*Karato and Karki* 2001; *Simmons et al.* 2009]. If the set of adjustment parameters

365 produced 1-D scaling profiles beyond this defined range, new random selections were made

366 before proceeding.

367 In order to determine whether to accept or reject a set of updated model parameters

368 ($\mathbf{m}^{new}$) we solved the joint system defined in Equation (1) in a non-linear sense by iteratively

369 updating the scaling model and global heterogeneity model using multiple applications of the

370 LSQR algorithm. The resulting data misfit was then calculated and compared to the misfit of the

371 previous model:

$$E\left(\mathbf{m}^{new}\right)=\frac{1}{6}\sum_{j=1}^{6}\frac{\left\|\mathbf{d}^{j}_{predicted}\left(\mathbf{m}^{new}\right)-\mathbf{d}^{j}_{observed}\right\|}{\left\|\mathbf{d}^{j}_{observed}\right\|},$$

372 (10)

$$\Delta E = E\left(\mathbf{m}^{new}\right)-E\left(\mathbf{m}^{k}\right)$$

373 where $\mathbf{d}^{j}_{predicted}$ and $\mathbf{d}^{j}_{observed}$ are the predicted and observed data vectors for the j^{th} data set listed
374 in Table 1. If the normalized average data misfit was lower than the previous updated model
375 ($\Delta E < 0$), the model was automatically accepted, the temperature was reduced, and the model
376 updating process was repeated. If the misfit was higher than in the previous step ($\Delta E > 0$), we
377 computed the probability function $P = \exp\left(-\Delta E / T_k\right)$ and compared it to a uniformly distributed
378 random number ($v \in [0,1]$). The model was accepted if $P > v$, otherwise the model was rejected
379 and the temperature remained constant for the next iteration. The process was terminated after
380 61 model updates due to the small changes of model parameter and misfit variations at this stage
381 (Figure 3).

382 The 1-D component of the resulting $R^{1.5\text{-}D}_{\rho/S}$ relationship (Figure 2) in the upper mantle
383 converged to the lowest allowable values in the upper mantle similar to the results of *Simmons et*
384 *al.* [2009]. This bias towards the lowest acceptable values in the upper mantle (based on the
385 range for the 1300°C adiabat calculations in [*Cammarano et al.* 2003]) is most likely a product
386 of the combined effects of cratons and high-temperature zones that both require lowered scaling
387 values to account for iron depletion and the temperature-dependence of Q_S, respectively. The
388 strongest upper mantle velocity signatures are 'atypical' in that they either deviate from a simple
389 pyrolitic composition or are very high-temperature, yet they occupy a large portion of the
390 shallow upper mantle and therefore greatly influence the 1-D scaling model solution.

391 In the lower mantle however, we find that the 1-D component differs from the results of
392 *Simmons et al.* [2009] who strictly tested a limited number of possible profiles. In the current
393 case, the 1-D component of the lower mantle $R_{\rho/S}^{1.5-D}$ relationship is lower amplitude in the middle
394 of the mantle and higher amplitude near the base. This 'straightening' of the scaling model
395 possibly reflects the lessening of the contributions of anelasticity relative to the starting solution
396 from *Karato and Karki* [2001]. The profiles in *Karato and Karki* [2001] might in fact
397 exaggerate the bulk contributions of anelasticity to mantle heterogeneity as recently reported in
398 some mineral physics studies [e.g. *Brodholt et al.* 2007; *Matas and Bukowinski* 2007] and our
399 results suggest a similar conclusion. Since the $R_{P/S}^{1.5-D}$ relationship is tied to the $R_{\rho/S}^{1.5\text{-}D}$
400 relationship for thermally induced variations according to Equation (4), the 1-D component of
401 the $R_{P/S}^{1.5-D}$ relationship is nearly identical to the starting model in the upper mantle. The only
402 notable change from the starting model is near the top of the lower mantle (Figure 2). The very
403 minor divergence from the starting model in the deep mantle is due to the systematic decrease of
404 δ_S with depth in the lower mantle, approaching a minimum of 1.3 assigned to the D'' layer. In
405 addition, the thermally induced $R_{\rho/S}$ values systematically decrease with depth, further
406 diminishing the influence of variations of $R_{\rho/S}$ on the resulting $R_{P/S}$ values in the deepest parts of
407 the mantle (see Equation 4).
408 We find velocity-based scaling model correction terms that are highly negative (Table 4)
409 in the cratonic roots reflecting the mass deficiency observed in numerous previous studies. In
410 the non-cratonic upper mantle, we find positive κ terms which is in general agreement with the
411 expected behavior due to the temperature dependence of the shear wave attenuation parameter

412 (often denoted Q_S) in the upper mantle [*Cammarano et al.* 2003]. With these scaling model
413 improvements ($R_{\rho/S}^{1-D} \rightarrow R_{\rho/S}^{1.5-D}$ and $R_{P/S}^{1-D} \rightarrow R_{P/S}^{1.5-D}$), we can increase the level of fit to all data sets
414 considered after joint inversion ('GyPSuM_1.5D'; Table 3). The most notable improvement is
415 to the dynamic surface topography field since we are directly accounting for the 1st-order effects
416 of cratons and other scaling model complexities in the upper mantle; thereby allowing for a
417 greater reconciliation of the free-air gravity and dynamic surface topography simultaneously.
418
419 **5.2 Compositional Decoupling (3-D *R* values)**
420 As shown in the previous section, 1.5-D scaling models including the 1st-order effects of
421 compositional variability in the cratons and temperature dependence of Q_S improves the level of
422 fit to the combined set of observations. However, these corrected relationships ($R_{\rho/S}^{1.5\text{-}D}$ and
423 $R_{P/S}^{1.5-D}$) do not completely describe the relative behavior of density and wave speeds since we
424 still are not able to satisfy the seismic observations to similar levels of fit obtainable by
425 independently produced models (e.g. TX2008s and P-only in Table 3). The presence of lateral
426 compositional variations requires 3-D $R_{\rho/S}$ and $R_{P/S}$ values to fully relate mantle heterogeneity
427 fields. Therefore, the scaling relationships must diverge from values representing thermal effects
428 in an iso-chemical mantle layer. We have thus developed 3-D scaling models through a process
429 called 'compositional decoupling' since the scaling models are allowed to deviate from the
430 estimated purely thermal values.
431 In this iterative process, we allow for slow divergence of the scaling relationships from
432 the 1.5-D values and subsequently update the heterogeneity model through non-linear joint

433 inversion. We first update $R_{\rho/S}^{3-D}$ through inversion of the geodynamic constraints while

434 assuming a fixed shear wave slowness perturbation model determined in the previous step.

435 Similarly, we update $R_{P/S}^{3-D}$ through inversion of P-wave information, disregarding the shear

436 wave and geodynamic constraints since our goal is to improve the level of fit to the P-wave data

437 given a shear wave model. The linear systems to be inverted can be represented using the

438 variables defined in Table 2:

439

440

$$\begin{bmatrix} \mathbf{G}\left(\Delta\mathbf{m}_S^k\right) \\ \mathbf{c}\left(\Delta\mathbf{m}_S^k\right) \end{bmatrix} R_{\rho/S}^{3-D,k} = \begin{bmatrix} \mathbf{g} \\ e \end{bmatrix} \xrightarrow{inversion} R_{\rho/S}^{3-D,k}$$

$$\left[\mathbf{L}_P\left(\Delta\mathbf{m}_S^k\right)\right] R_{P/S}^{3-D,k} = \left[\mathbf{r}_P\right] \xrightarrow{inversion} R_{P/S}^{3-D,k+1} \qquad (11)$$

441

442

443 where we have omitted regularization, weighting and slowness-velocity conversion terms for

444 simplicity. At each cycle, *k*, we performed a limited number of LSQR iterations (we chose 4) to

445 restrict the divergence of scaling models from the previous model state. There is no formal basis

446 for the selection of the number of LSQR iterations (4), but through trial-and-error testing we

447 found that 4 iterations did not allow the scaling models to move incredibly far from the previous

448 state. To complete the k^{th} inversion cycle, we inverted the full system of equations (consisting of

449 all seismic and geodynamic information) for mantle structure assuming the updated $R_{\rho/S}^{3-D}$ and

450 $R_{P/S}^{3-D}$ models, again with a limited number of LSQR iterations. The k^{th} shear wave slowness

451 model ($\Delta\mathbf{m}_S^k$) then formed the basis for the subsequent cycle of 3-D scaling model inversions.

452 The process was repeated until we observed no significant improvement in the level of fit to all

453 of the data fields considered. Convergence occurred after 64 inversion cycles and the process
454 was thus terminated.

455 The resulting 3-D scaling model distributions as a function of depth are illustrated in
456 Figure 4. The distribution of $R_{\rho/S}^{3-D}$ values in the shallow upper mantle is broad and skewed
457 towards zero, with a substantial number of negative values. This distribution demonstrates
458 lowered scaling factors due to 1) iron depletion in the cold cratonic roots, producing positive
459 buoyancy in some cases, and 2) the temperature-dependence of Q_S in the hot non-cratonic
460 regions such as the mid-ocean ridges and rift zones, reducing the apparent buoyancy. A similar
461 distribution is found in the deepest mantle likely demonstrating significant compositional
462 variations, including within the 'superplume' structures as evidenced in numerous studies [e.g.
463 *Ishii and Tromp* 1999; *Ritsema et al.* 1999; *van der Hilst and Kárason* 1999; *Masters et al.* 2000;
464 *Wen* 2001; *Ni et al.* 2002; *Ni and Helmberger* 2003; *Trampert et al.* 2004; *Simmons et al.* 2007].
465 Specifically, the intrinsically high density material in the superplume structures competes with
466 the thermally induced density anomalies thereby reducing the overall $R_{\rho/S}$ values [*Simmons et*
467 *al.* 2007]. The basal portion of the superplume structures could also be intrinsically slow, further
468 lowering $R_{\rho/S}$ values. Aside from the shallow and very deep mantle, we find relatively compact
469 $R_{\rho/S}^{3-D}$ distributions with modes similar to the simple 1-D scaling relationship that assumes
470 thermal-dominance. We note that, although we find anti-correlations (i.e. negative scaling) of
471 density and shear speed in the lower mantle, they are not the dominant signature even in the
472 deepest mantle. Since density and shear wave variations are highly sensitive to temperature
473 variations, it may be argued that the non-cratonic mantle shear wave speeds and density

474 variations are primarily products of temperature variations [*Schuberth et al.* 2009*; Simmons et al.*
475 2009].

476 The relative heterogeneity ratio (or correlation parameter) $R_{P/S}$ is most often referred to by
477 the inverse ($R_{P/S}^{-1} = R_{S/P}$) in the literature and a wide range of values have been reported [*Vasco et*
478 *al.* 1994; *Robertson and Woodhouse* 1996; *Su and Dziewonski* 1997; *Kennett et al.* 1998;
479 *Masters et al.* 2000; *Saltzer et al.* 2001; *Ritsema and van Heijst* 2002; *Antolik et al.* 2003]. It has
480 been well established that there are numerous regions in the deep mantle with significant S-wave
481 heterogeneities corresponding to very small P-wave anomalies. This situation generates a large
482 quantity of nearly unbounded $R_{S/P}$ values and may contribute to the large range of average
483 values reported in the literature. This may in turn lead to discounting thermal variations as a
484 major contributor to seismic anomalies in the lower mantle since the average values often fall
485 well outside the expected thermal bounds [e.g. *Karato* 1993; *Masters et al.* 2000]. For these
486 reasons, we have solved for $R_{P/S}^{3-D}$ in the current study.

487 Our results show that modes of the distributions of $R_{P/S}^{3-D}$ tend to follow the 1-D
488 component of the VFSA solution ($R_{P/S}^{1.5-D}$) throughout most of the mantle (Figure 4). If we
489 compare these distributions to those generated by models produced solely with the individual
490 seismic data sets (TX2008s and P-only), we find that the joint inversion procedure clearly
491 produces simultaneous P and S-wave models more consistent with pure thermal effects. This
492 conclusion is based upon the ability to dramatically collapse the distributions of $R_{P/S}^{3-D}$ about the
493 1-D thermal profile with respect to the unconstrained, independent solutions (Figure 4).
494 However, there are still some broadly ranging values of scaling values even though attempts

were made to produce an entirely thermally generated heterogeneity model in the previous stages. This is most obvious in the deepest mantle where the distributions of $R_{P/S}^{3-D}$ become broad and skewed in some layers. The broad distributions likely, in part, reflect compositional anomalies in the deep mantle as concluded in numerous previous studies of the relative behavior of mantle properties [e.g. *Ishii and Tromp* 1999; *van der Hilst and Kárason* 1999; *Masters et al.* 2000; *Saltzer et al.* 2001; *Trampert et al.* 2004].

Relative seismic heterogeneity in the D'' layer appears to be least consistent with thermal variations since a nearly bimodal distribution of $R_{P/S}^{3-D}$ is detected (Figure 4). The primary mode of the distribution (centered near 0.5) could potentially be explained by temperature variations; however the secondary mode in the distribution (centered near 0), clearly violates the expected thermal behavior in an iso-chemical mantle [*Karato and Karki* 2001]. The decreased values of $R_{P/S}$ in the D'' layer might be indicative of pressure-induced phase changes to post-perovskite (pPv) creating a partial de-correlation of P and S-wave heterogeneity [*Murakami et al.* 2004; *Oganov and Ono* 2004; *Iitaka et al.* 2004; *Tsuchiya et al.* 2004; *Wookey et al.* 2005; *Hirose* 2006; *Hernlund and Houser* 2008; *Hutko et al.* 2008]. This hypothesis is based upon mineral physics analyses that indicate that shear modulus and density increase while bulk modulus is relatively unchanged due to the transition to the pPv phase [e.g. *Tsuchiya et al.* 2004]. Moreover, the large spread of relative heterogeneity values near zero may be due to patchy occurrences of pPv (and perhaps lens-like seams) modulated by the background temperature [*Hernlund et al.* 2005; *Lay et al.* 2006; *Hernlund and Houser* 2008]. The discrepancy between the distributions of P- and S-wave anomalies in the deep mantle is robust [*Hernlund and Houser* 2008] and persists after performing a joint inversion. However, we point out that the joint inversion process performed in this study leads to seismic anomalies more consistent with thermal variations as manifested through the distribution of $R_{P/S}$ (Figure 4).

6. GyPSuM Model and Correlations

The final model, GyPSuM, is illustrated in Figures 5-7. The inversion procedures described within this report allow for the estimation of what we refer to 'thermal' and 'non-thermal' contributions to mantle heterogeneity. The 'thermal' contributions to each of the model fields (velocities and density) are estimated from the joint inversion results obtained by employing the temperature-induced part of the optimum 1.5-D scaling model. The 'non-thermal' contributions are computed by subtracting the thermal components from the final solution developed in the previous section. Therefore, the non-thermal heterogeneity field may be thought of as a residual field required in addition to the thermal field in order to fit the data. It should be noted that there is a significant chance that some of the signatures we refer to as 'non-thermal' might actually be due to effects of laterally varying Q_S since, in the development of the 1.5-D scaling model, we only considered 1st-order thermal scaling corrections in the upper mantle and no corrections in the lower mantle (see Section 5.1).

In the upper mantle, we find relatively minor non-thermal contributions to the shear wave heterogeneity field (Figure 5). Some of the most prominent non-thermal contributions to the shear wave heterogeneity field are found in the deep mantle within the superplume structures (most notably beneath Africa). These structures are found to be intrinsically slow in terms of shear wave speeds lending support to the idea that these features are partially produced by the sweeping of compositionally distinct material into regions of large-scale upwelling mantle [*McNamara and Zhong* 2005]. However, we note that increased temperatures contribute greatly to their overall observed low-velocity signatures (Figure 5). Moreover, it is possible that some of the estimated non-thermal signals within the superplume structures are products of laterally

varying Q_S due to a temperature increase. Specifically, near 2280 km depth (Figure 5), we find significant low-velocity shear wave signatures in the 'non-thermal' field with the African superplume structure. This apparent non-thermal signature manifests through the modeled decrease of $R_{\rho/S}$ found through the 3-D scaling model inversion process relative to the 1-D scaling profile (i.e. thermal profile). However, based on the work of *Karato and Karki* [2001], a similar effect may be expected due to increased temperatures and the impact to Q_S. Alternatively, our inferred non-thermal (or compositional) signature within these superplumes (notably under Africa) may be explained in terms of iron-enrichment since this will produce an reduction in shear wave speed in addition to the reductions expected for increased temperature [e.g. *Forte and Mitrovica* 2001].

Although there are apparent non-thermal influences on the shear velocity field, temperature variations appear to be the dominant factor. This is demonstrated more quantitatively through the calculation of the root-mean-squared (RMS) amplitudes of the individual thermal/non-thermal heterogeneity fields (Figure 8). We find that the RMS amplitude of the thermally induced shear wave heterogeneity is the largest contributor to the overall field throughout the entire mantle. This result is in direct agreement with *Quéré and Forte* [2006] and *Schuberth et al.* [2009] who found that steep lateral temperature gradients could mostly account for the large-scale, low-velocity superplume structures beneath Africa and the Pacific Ocean.

Density heterogeneity in the shallow upper mantle is produced by both temperature and compositional variations (Figure 7, 8). As expected, the thermal signature in the cratonic regions is high-density due to decreased temperatures while the compositional signature is low-density due to depletion of basaltic components. Due to the temperature-dependence of Q_S, the thermal

563 signatures are higher amplitude in seismically fast (cold) regions than in slow (hot) regions.
564 However, the competing thermal/compositional effects in the cratonic regions produce far lower
565 amplitude densities in the cratons overall, relative to the purely thermally induced density field
566 (Figure 7). Compositional density anomalies are small relative to thermally driven density field
567 throughout most of the non-cratonic mantle based on the RMS amplitude calculations (Figure 8).
568 However, in the bottom ~400 km of the mantle, the compositional influence increases
569 dramatically. The major contributions to the compositional density fields are within the African
570 and Pacific superplume structures. The central cores of these features are found to be
571 intrinsically dense. The positive intrinsic density of these structures severely offsets the
572 thermally driven density in portions of the superplume structures thereby reducing the overall
573 buoyancy. These results are similar to the result of *Simmons et al.* [2007; 2009]. However, the
574 magnitude of the positive intrinsic density found in the current study generates slightly positive
575 density overall in localized portions of the African superplume structure, in particular in the D''
576 layer (Figure 7). Studies utilizing normal mode splitting functions [e.g. *Ishii and Tromp* 1999;
577 *Trampert et al.* 2004] have found overall high-densities in these deep-mantle structures. Our
578 results differ in that the high-density material observed in these features is far more localized and
579 much lower amplitude. It is unclear whether the discrepancy is a product of 1) the differing
580 wavelengths of heterogeneity modeled in each case, 2) the different types of data used, 3) the
581 starting model and built-in assumptions, or 4) a combination of all of these factors. We can say,
582 however, that the data considered in this study do not require large-scale zones of high-amplitude
583 positive density anomalies in the superplume structures to simultaneously explain all the
584 geodynamic and seismic constraints we have employed.

585 Compressional wave heterogeneity is shown to have significantly more influence from
586 compositional variations relative to shear wave velocity (Figures 6, 8). These signatures are
587 most notable in the cratonic regions which are known to be compositionally heterogeneous. We
588 also detect significant compositional anomalies throughout the lower mantle that are typically
589 half of the amplitude of the thermally induced anomalies. One of the more notable non-thermal
590 P-wave velocity signatures in the deep mantle are negative values associated with the subducted
591 Farallon and Tethys slabs below ~1500 km depth (see Figure 6 at 1830 km depth). These
592 negative non-thermal signatures oppose the thermal signatures (fast) and act to mask the overall
593 P-wave signatures of the deepest extent of the ancient subducted slabs. It is therefore clear that
594 compositional variations contribute significantly to P-wave heterogeneity; yet, the RMS
595 amplitudes of the individual thermal/non-thermal structures suggest that temperature variations
596 are most often the largest contributor.

597 We have examined this result further by computing bulk sound speed variations from the
598 modeled P and S-wave anomalies, and similarly computed the RMS amplitudes as a function of
599 depth (Figure 8). We find that the primary contributor to sound speed heterogeneity is variations
600 in composition, rather than temperature. This finding provides strong support to the initial
601 results by *Forte and Mitrovica* [2001] that showed that bulk sound speed anomalies provide an
602 effective mapping of compositional heterogeneity in the deep mantle. This is especially evident
603 in deep mantle where the sensitivity of sound speed to temperature is weak due to δ_S values
604 approaching 1 (see Equation 3). However, the considerable effect of temperature on the relative
605 behavior of shear modulus and density produces P-wave anomalies largely controlled by thermal
606 variations. Thus, elucidating the reason for well-correlated S- and P-wave anomalies while

607 sound speed is often poorly correlated with S-wave speeds [*Su and Dziewonski* 1997; *Masters et*
608 *al.* 2000; *Kennett and Gorbatov* 2004].

609 Based on the RMS amplitudes of wave speeds and density heterogeneities (Figure 8), we
610 may conclude that temperature variations are the primary cause of mantle heterogeneity. We
611 also find that the correlation of shear wave velocity and the other modeled fields are highly
612 positive in the bulk of the mantle (Figure 9). In particular, the correlation between density and
613 S-wave speeds approaches 1 in a large portion of the mantle. This correlation drops off in the
614 deepest mantle owing to the compositional anomalies associated with the superplume structures.
615 Similarly, P- and S-wave heterogeneities are highly correlated when all anomalies (estimated
616 thermal and non-thermal contributions) are considered. This result, along with the systematic
617 de-correlation of sound and shear speeds with depth, is generally consistent with past results [e.g.
618 *Masters et al.* 2000; *Saltzer et al.* 2001]. Moreover, sound and shear speeds become anti-
619 correlated in the deep mantle with a negative peak just below ~2000 km depth; a result that is
620 very similar to the results presented in *Saltzer et al.* [2001]. However, just above the D'' layer,
621 the correlation jumps to positive values and then quickly returns to negative within the D'' layer
622 unlike the aforementioned study. Both *Saltzer et al.* [2001] and *Masters et al.* [2000] show hints
623 of this cyclical correlation behavior, but the amplitude of the correlation jumps presented in the
624 current study are more dramatic.

625 Although the modeled P/S-wave speeds and density are mostly correlated when all
626 anomalies are considered in the calculation, separating the fields according to fast and slow shear
627 wave anomalies reveals more complicated results (Figure 9). In particular, the correlation
628 between the slow S-wave structures and density show much smaller (and sometimes negative)
629 correlations in the deep mantle. The primary source of this de-correlation is easily recognized to

630 be the effects of the opposing thermal/compositional density signatures associated with the
631 superplume structures (Figure 7). The correlation between slow S-wave anomalies and the
632 corresponding P-wave anomalies is high, similar to the total correlation. However, if we
633 compute the correlation of high shear velocity zones with the corresponding P-wave anomalies,
634 we find a dramatically different result (Figure 9). Specifically, P-wave anomalies in regions with
635 high S-wave velocities systematically de-correlate with depth beginning at ~1500 km and
636 peaking to slightly negative values at ~2100 km depth. Thus, the deep-mantle high shear
637 velocity zones, that may be attributed to ancient subducted slab remnants, have significantly
638 different geographic patterns than the P-wave anomalies due to strong anti-correlations of shear
639 and sound speeds. In extreme cases, the high-velocity shear zones correspond to low-velocity P-
640 wave anomalies. Therefore, the thermal P-wave signatures are significantly countered by the
641 'non-thermal' component (see Figure 6; 1830 km depth) producing muted or absent total P-wave
642 structure where S-wave velocities are fast. As pointed out in *Boschi et al.* [2007], deep lower
643 mantle anomalies are dominated by negative velocities at low spatial frequencies. On the other
644 hand, high-velocity anomalies in the lower mantle are typically restricted to high spatial
645 frequencies that may not be well-resolved in some cases. Correlation properties of fast
646 anomalies in the lower mantle, based on independently derived P- and S-wave models, would
647 therefore be dubious. However, in the development of the GyPSuM model, both P- and S-wave
648 data were modeled simultaneously with the same parameterization and regularization mitigating
649 a number of issues including the varying resolution of each data set.

650 Due to the relative temperature sensitivity of the bulk and shear moduli, it is possible to
651 have anti-correlated shear and sound speeds if sound speeds anomalies are small [*Kennett and*
652 *Gorbatov* 2004]. Similar to the study of *Kennett and Gorbatov* [2004], we find that inferred

subducted slab features below North America and southern Asia follow this pattern down to ~1500 km depth. However, we find strongly anti-correlated sound and shear speeds below ~1500 km depth with significant sound speed variations when we consider only structures with fast shear speeds (Figure 9). It is unclear if the deep-mantle high shear velocity blobs are remnant slab materials. But it is evident that these features have very different P-wave signatures that do not adhere to the expected thermal behavior, differing from zones with low shear velocity in the lower half of the mantle.

7. Conclusions and Discussion

We have constructed a tomographic model of mantle shear wave speeds, compressional wave speeds and detailed density anomalies through the simultaneous inversion of seismic and geodynamic observations. The mantle model (labeled GyPSuM) was constructed with the underlying hypothesis that temperature variations are the dominant cause of mantle heterogeneity via the integration of mineral physics parameters that describe the relative behavior of mantle properties due to thermal effects. In a bulk sense, we find that P-wave, S-wave, and density anomalies in the non-cratonic mantle may be primarily attributed to variations in temperature. Thermal dominance of heterogeneity is best demonstrated through the calculated RMS amplitudes of the estimated thermal and non-thermal contributions to the model fields (wave speeds and density) shown in Figure 8.

Wave speeds and density are highly correlated throughout most of the mantle when all anomalies are considered (Figure 9). However, high shear velocity zones and the corresponding P-wave anomalies systematically de-correlate in the ~1500-2500 km depth range suggesting another mechanism besides simple thermal variations. The combined thermal and other possible

676 mechanism(s) mute P-wave velocity signatures in zones commonly interpreted as subducted
677 slabs in the deep mantle on the basis of S-wave tomographic solutions that show persistent fast
678 anomalies at these depths (compare the 'Total' V_S and V_P fields in Figures 5- 6 at 1830 km
679 depth). A possible explanation for the P- and S-wave discrepancies is the effects of electronic
680 spin transitions in iron-bearing minerals. Studies of the elastic effects of spin transition suggest
681 that, at mid-mantle depths, the transition will generate negative seismic velocity anomalies over
682 a broad depth range [e.g. *Crowhurst et al.* 2009; *Wentzcovitch et al.* 2009]. The effects of this
683 mechanism would then oppose the thermally induced fast velocity signatures within the
684 subducted slab material. Moreover, if the temperature decrease is small enough, the effects of
685 spin transitions could overwhelm the thermally induced high P-wave velocities thereby muting
686 out the structure entirely. The S-wave signatures in the possible subducted slab remnants could
687 remain fast given the relatively increased sensitivity of S-wave velocity to thermal variations. In
688 such a scenario, the high-temperature zones (producing low-velocity signatures) would remain
689 slow since the combined effects of increased temperature and spin transition would be
690 constructive. Therefore the correlation of low shear velocity zones with the corresponding P-
691 wave values would remain large and positive.
692 One potential problem with the hypothesis stated above is the fast, non-thermal shear
693 wave speeds we estimate in these zones (Figure 5; 1830 km depth). If the previous scenario
694 were true, we would expect an opposite (slow) non-thermal V_S signature. A possible explanation
695 for this apparent fast non-thermal anomaly is the underestimation of the thermal contributions to
696 the shear wave velocity field in this depth range. The joint inversion with thermal scaling
697 relationships (1-D in the lower mantle) incorporated both S- and P-wave data that conflict when
698 scaled in this simplified way as evidenced by our results. Thus the thermal S-wave heterogeneity

699 solution could have been corrupted by the P-wave data that require nearly no anomalies in these
700 subducted slab remnants. Nonetheless, we find that zones with fast shear wave speeds (centered
701 at ~2000 km depth) correlate poorly with the corresponding P-wave anomalies in the *total*
702 velocity fields (Figure 9) and are often muted and/or absent (Figures 5-6). Spin transitions
703 potentially explain our observations, but the actual effects that electronic spin transitions have on
704 mantle materials is up for debate [see for example *Badro et al.* 2003, 2004; *Hofmeister* 2006; *Lin*
705 *et al.* 2007, 2008; *Speziale et al.* 2007; *Stackhouse et al.* 2007; *McCammon et al.* 2008;
706 *Crowhurst et al.* 2009; *Wentzcovitch et al.* 2009].

707 Aside from the discrepancies of wave speeds in the ancient subducted slab remnants, the
708 superplume structures beneath Africa and the Pacific Ocean possess properties that cannot be
709 explained by temperature variations alone. Most notably, portions of the superplume structures
710 have significant positive non-thermal density signatures that are relatively broad in the D'' layer
711 and extend upward through the mid-mantle with a narrower lateral extent (Figure 7). These
712 density signatures are interpreted as intrinsically dense material that is partially entrained within
713 the upwelling superplumes [*Simmons et al.* 2007]. The intrinsic density of this material counters
714 the thermally induced density, thereby reducing the overall buoyancy of the upwellings [see
715 *Simmons et al.* 2007, 2009; *Forte et al.* 2010]. The amplitude of the non-thermal high-density
716 signatures in the superplumes are larger than our previous tomography results [*Simmons et al.*
717 2007, 2009] owing to the addition of the P-wave constraints that help limit the range of possible
718 configurations of density heterogeneity in the joint inversion process.

719 Combining the thermal and non-thermal components of the density field, we find overall
720 positive density anomalies within the South Africa superplume structure. Studies incorporating
721 normal mode splitting functions have similarly modeled high density signatures in the low-

velocity superplume structures [e.g. *Ishii and Tromp* 1999; *Trampert et al.* 2004] suggesting a dominant compositional influence on the heterogeneity in these regions. However, our results dramatically differ in that the overall high-density zones appear to be far more localized and lower amplitude. Specifically, we find only slightly positive density anomalies beneath the extreme southern tip of Africa that is mostly confined to the D'' layer. Beneath the Pacific Ocean, we find no significant positive density anomalies in the low-velocity structures. However, a localized portion of the Pacific superplume density structure is strongly affected by the positive intrinsic density of the material, also severely muting the temperature-induced buoyancy.

Even with these aforementioned complexities, we argue that temperature variations dominate heterogeneity throughout the majority of the non-cratonic mantle. The incorporation of additional data types (not considered herein) in detailed joint tomographic studies may disprove this argument in the near future. Nonetheless, we believe that the initial assumption that thermal variations dominate mantle heterogeneity should be applied in future investigations to more accurately determine the necessary compositional influence.

Acknowledgements

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. LLNL-?????

References

Anderson, D.L. (1989), *Theory of the Earth*, Blackwell Sci., Malden, Mass.

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

Badro, J., G. Fiquet, F. Guyot, J.-P. Rueff, V.V. Struzhkin, G. Vankó and G. Monaco (2003), Iron partitioning in Earth's mantle: Toward a deep lower mantle discontinuity, *Science*, *300*, 789-791.

Badro, J., J.-P. Rueff, G. Vankó, G. Monaco, G. Fiquet and F. Guyot (2004), Electronic transitions in perovskite: Possible nonconvecting layers in the lower mantle, *Science*, *305*, 383-386.

Boschi, L., T.W. Becker and B. Steinberger (2007), Mantle plumes: Dynamic models and seismic images, *Geochem. Geophys. Geosys.*, *8*(10), doi: 10.1029/2007GC001733.

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

Cammarano, F., S. Goes, P. Vacher & D. Giardini (2003), Inferring upper-mantle temperatures from seismic velocities, *Phys. Earth Planet. Inter.*, *138*(3-4), 197-222.

Crowhurst, J.C., J.M. Brown, A.F. Goncharov and S.D. Jacobsen (2009), Elasticity of (Mg,Fe)O through the spin transition of iron in the lower mantle, *Science*, *319*, 451-453.

DeMets, C.R., Gordon, R.G., Argus, D.F. & S. Stein (1990), Current plate motions, *Geophys. J. Int.*, *101*(2), 425-478.

Dziewonski, A.M. & F. Gilbert (1976), The effect of small, aspherical perturbations on travel times and a re-examination of the corrections for ellipticity, *Geophys. J. R. Astron. Soc. 44*, 7-17.

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

Engdahl, E., R.D. van der Hilst and R. Buland (1998), Global teleseismic earthquake relocation with improved travel times and procedures for depth discrimination, *Bull. Seis. Soc. Amer.*, *88*, 722-743.

Forte, A.M. & W.R. Peltier (1987), Plate tectonics and aspherical Earth structure; the importance of poloidal-toroidal coupling, *J. Geophys. Res. 92*, 3645-3679.

786 Forte, A.M., W.R. Peltier, A.M. Dziewonski and R.L. Woodward (1993), Dynamic surface
787 topography: A new interpretation based upon mantle flow models derived from seismic
788 tomography, *Geophys. Res. Lett*., *20*, 225-228.
789
790 Forte, A.M., J.X. Mitrovica and R.L. Woodward (1995), Seismic-geodynamic determination of
791 the origin of excess ellipticity of the core-mantle boundary, *Geophys. Res. Lett*., *22*,
792 1013-1016.
793
794 Forte, A.M. & H.K.C. Perry (2000), Geodynamic evidence for a chemically depleted continental
795 tectosphere, *Science*, *290*, 1940-1944.
796
797 Forte, A.M. & J.X. Mitrovica (2001), Deep-mantle high-viscosity flow and thermochemical
798 structure inferred from seismic and geodynamic data, *Nature*, *410*(6832), 1049-1056.
799
800 Forte, A.M. (2007), Constraints on seismic models from other disciplines-implications for
801 mantle dynamics and composition, in *Volume 1 of Treatise of Geophysics*, edited by B.
802 Romanowicz and A.M. Dziewonski, Elsevier, pp. 805–858.
803
804 Forte, A.M, S. Quéré, R. Moucha, N.A. Simmons, S.P. Grand, J.X. Mitrovica and D.B. Rowley
805 (2010), Joint seismic-geodynamic-mineral physical modeling of African geodynamics: A
806 reconciliation of deep-mantle convection with surface geophysical constraints, *Earth*
807 *Planet. Sci. Lett.* (accepted).
808
809 Grand, S.P. & D.V. Helmberger (1984), Upper mantle shear structure of North America,
810 *Geophys. J. R. Astron. Soc*. 76, 399-438.
811
812 Grand, S.P. (1994), Mantle shear structure beneath the Americas and surrounding oceans, *J.*
813 *Geophys. Res.*, *99*, 11591-11621.
814
815 Grand, S.P., R.D. Van der Hilst & S. Widiyantoro (1997), Global seismic tomography: a
816 snapshot of convection in the Earth, *GSA Today*, *7*, 1-7.
817
818 Grand, S.P. (2002), Mantle shear-wave tomography and the fate of subducted slabs, *Phil. Trans.*
819 *R. Soc. Lond. A*, *360*(1800), 2475-2491.
820
821 Hernlund, J.W., C. Thomas and P.J. Tackley (2005), A doubling of the post-perovskite phase
822 boundary and structure of the Earth's lowermost mantle, *Nature*, *434*, 882–886.
823
824 Hernlund, J.W. and C. Houser (2008), On the statistical distribution of seismic velocities in
825 Earth's deep mantle, *Earth Planet. Sci. Lett*., *265*, 423-437.
826
827 Herring, T.A., P.M. Mathews & B.A. Buffett (2002), Modeling of nutation-precession; very long
828 baseline interferometry results, *J. Geophys. Res. Solid Earth*, *107*(B4), doi:
829 10.1029/2001JB000165.
830

831 Hirose, K. (2006), Postperovskite phase transition and its geophysical implications, *Rev.*
832 *Geophys*., *44*(RG3001), doi:10.1029/2005RG000186.
833
834 Hofmeister, A.M. (2006), Is low-spin Fe^{2+} present in Earth's mantle?, *Earth Planet. Sci. Lett.*,
835 *243*, 44-52.
836
837 Hutko, A.R., T. Lay, J. Revenaugh and E.J. Garnero (2008), Anticorrelated seismic velocity
838 anomalies from post-perovskite in the lowermost mantle, *Science*, *320*, 1070-1074.
839
840 Iitaka, T., K. Hirose, K. Kawamura, and M. Murakami (2004), The elasticity of the $MgSiO_3$
841 post-perovskite phase in the Earth's lowermost mantle, *Nature*, *430*, 442-445.
842
843 Ingber, L. (1989), Very fast simulated re-annealing. *Math. Comput. Modell., 12***,** 967–973.
844
845 Ishii, M. & J. Tromp (1999), Normal-mode and free-air gravity constraints on lateral variations
846 in velocity and density of Earth's mantle, *Science*, *285*(5431), 1231-1236.
847
848 Ishii, M. & J. Tromp (2004), Constraining large-scale mantle heterogeneity using mantle and
849 inner-core sensitive normal modes, *Phys. Earth Planet. Inter.*, *146*(1-2), 113-124.
850
851 Jackson, C., M.K. Sen and P.L. Stoffa (2004), An efficient stochastic Bayesian approach to
852 optimal parameter and uncertainty estimation for climate model predictions, *J. Climate*,
853 *17* 2828-2841.
854
855 Jordan, T.H. (1978), Composition and development of the continental tectosphere, *Nature*, *274*,
856 544-548.
857
858 Karato, S.I. (1993), Importance of anelasticity in the interpretation of seismic tomography,
859 *Geophys. Res. Lett.*, *20*, 1623-1626.
860
861 Karato, S.I. & B.B. Karki (2001), Origin of lateral variation of seismic wave velocities and
862 density in the deep mantle, *J. Geophys. Res. Solid Earth*, *106*(B10), 21771-21783.
863
864 Kennett, B.L.N., S. Widiyantoro and R.D. van der Hilst (1998), Joint seismic tomography for
865 bulk sound and shear wave speed in the Earth's mantle, *J. Geophys. Res*., *103*, 12469-
866 12493.
867
868 Kennett, B.L.N. & A. Gorbatov (2004), Seismic heterogeneity in the mantle—strong shear wave
869 signature of slabs from joint tomography, *Phys. Earth Planet. Inter.*, *146*, 87-100.
870
871 Kuo, C. & B. Romanowicz (2002), On the resolution of density anomalies in the Earth's mantle
872 using spectral fitting of normal-mode data, *Geophys. J. Int.*, *150*, 162-179.
873
874 Lay, T., J. Hernlund, E.J. Garnero and M.S. Thorne (2006), A postperovskite lens and D″ heat
875 flux beneath the central Pacific, *Science*, *314*, 1272–1276.

Lemoine, F. G., et al. (1998), The development of the joint NASA GSFC and NIMA Geopotential Model EGM96, NASA Tech. Rep., 1998-206861.

Le Stunff, Y. and Y. Ricard (1995), Topography and geoid due to lithospheric mass anomalies, *Geophys. J. Int.*, *122*, 982-990.

Lin, J.-F., G. Vankó, S.D. Jacobsen, V. Iota, V.V. Struzhkin, V.V. Prakapenka, A. Kuznetsov, C.-S. Yoo (2007), Spin transition zone in Earth's lower mantle, *Science*, *317*, 1740-1743.

Lin, J.-F., H. Watson, G. Vankó, E.E. Alp, V.V. Prakapenka, P. Dera, V.V. Struzhkin, A. Kubo, J. Zhao, C. McCammon and W.J. Evans (2008), Intermediate-spin ferrous iron in lowermost mantle post-perovskite and perovskite, *Nature Geosci.*, *1*, 688-691.

Masters, G., G. Laske, H. Bolton & A.M. Dziewonski (2000), The relative behavior of shear velocity, bulk sound speed, and compressional velocity in the mantle: implications for chemical and thermal structure, *in Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale*, edited by S.-I. Karato et al., pp. 63-87, AGU, Washington, DC.

Matas, J. & M. 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., T.A. Herring & B.A. Buffett (2002), Modeling of nutation and precession: new nutation series for nonrigid Earth and insights into the Earth's interior, *J. Geophys. Res. Solid Earth*, *107*(B4), doi: 10.1029/2001JB000390.

McCammon, C., I. Kantor, O. Narygina, J. Rouquette, U. Ponkratz, I. Sergueev, M. Mezouar, V. Prakapenka and L. Dubrovinsky (2008), Stable intermediate-spin ferrous iron in lower-mantle perovskite, *Nature Geosci.*, *1*, 684-687.

McNamara, A.K. & S. Zhong (2005), Thermochemical structures beneath Africa and the Pacific Ocean, *Nature*, *437*(7062), 1136-1139.

Mitrovica, J.X. & A.M. Forte (2004), A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data, *Earth Planet. Sci. Lett.*, *225*(1-2), 177-189.

Mooney, W.D., G. Laske & G. Masters (1998), CRUST 5.1: a global crustal model at 5 x 5 degrees, *J. Geophys. Res. Solid Earth*, *103*(B1), 727-747.

Murakami, M., K. Hirose, K. Kawamura, N. Sata and Y. Ohishi (2004), Post-perovskite phase transition in $MgSiO_3$, *Science*, *304*, 855-858.

Ni, S.D., E. Tan, M. Gurnis & D.V. Helmberger (2002), Sharp sides to the African superplume, *Science*, *296*(5574), 1850-1852.

Ni, S.D. & D.V. Helmberger (2003), Seismological constraints on the South African superplume; could be the oldest distinct structure on Earth, *Earth Planet. Sci. Lett.*, *206*(1-2), 119-131.

Oganov, A.R. and S. Ono (2004), Theoretical and experimental evidence for a post-perovskite phase of $MgSiO_3$ in Earth's D'' layer, *Nature*, *430*, 445-448.

Paige, C.C. & M.A. Saunders (1982), LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Transactions on Mathematical Software*, *8*, 43-71.

Quéré, S. & A.M. Forte (2006), Influence of past and present-day plate motions on spherical models of mantle convection: implications for mantle plumes and hotspots, *Geophys. J. Int*., *165*, 1041-1057.

Ricard, Y., L. Fleitout & C. Froidevaux (1984), Geoid heights and lithospheric stresses for a dynamic Earth, Ann*. Geophys. 2*, 267-286.

Richards, M.A. & B.H. Hager (1984), Geoid anomalies in a dynamic earth, *J. Geophys. Res. Solid Earth*, *89*(B7), 5987-6002.

Ritsema, J., H.J. Van Heijst & J.H. Woodhouse (1999), Complex shear wave velocity structure imaged beneath Africa and Iceland, *Science*, *286*(5446) 1925-1928.

Ritsema, J. and H.H. van Heijst (2002), Constraints on the correlation of P- and S-wave velocity heterogeneity in the mantle from P, PP, PPP and PKPab traveltimes*, Geophys. J. Int*., *149*, 482-489.

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

Saltzer, R.L., R.D.Van der Hilst & H. Kárason (2001), Comparing P and S wave heterogeneity in the mantle, *Geophys. Res. Lett.*, *28*(7) 1335-1338.

Schuberth, B.S.A., Bunge, H.P. and J. Ritsema (2009), Tomographic filtering of high-resolution mantle circulation models: Can seismic heterogeneity be explained by temperature alone?, *Geochem. Geophys. Geosys*., *10*, DOI: 10.1029/2009GC002401.

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

Simmons,N.A., Forte, A.M. and S.P. Grand (2007), Thermochemical structure and dynamics of the African superplume, *Geophys. Res. Lett., 34*(2), L02301, doi: 10.1029/2006GL028009.

Simmons, N., Forte, A.M. and S.P. Grand (2009), Joint seismic, geodynamic and mineral physical constraints on three-dimensional mantle heterogeneity: implications for the relative importance of thermal versus compositional heterogeneity, *Geophys. J. Int*. *177*(5), 1284–1304.

Soldati, G., L. Boschi, F. Deschamps and D. Giardini (2009), Inferring radial models of mantle viscosity from gravity (GRACE) data and an evolutionary algorithm, *Phys. Earth Planet Inter.*, *176*, 19-32.

Speziale, S., V.E. Lee, S.M. Clark, J.-F. Lin, M.P. Pasternak and R. Jeanloz (2007), Effects of Fe spin transition on the elasticity of (Mg, Fe)O magnesiowüstites and implications for the seismological properties of the Earth's lower mantle, *J. Geophys. Res*., *112*, doi:10.1029/2006JB004730.

Stackhouse, S., J.P. Brodholt and G.D. Price (2007), Electronic spin transitions in iron-bearing $MgSiO_3$ perovskite, *Earth Planet. Sci. Lett*., *253*, 282-290.

Su, W. & A.M. Dziewonski (1997), Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth Planet. Inter.*, *100*(1-4), 135-156.

Trampert, J., F. Deschamps, J. Resovsky & D. Yuen (2004), Probabilistic tomography maps chemical heterogeneities throughout the lower mantle, *Science*, *306*(5697), 853-856.

Tsuchiya, T., J. Tsuchiya, K. Umemoto and R.M. Wetzcovitch (2004), Elasticity of post-perovskite $MgSiO_3$, *Geophys.Res. Lett*., *31*, L14603.

Van der Hilst, R.D. & H. Kárason (1999), Compositional heterogeneity in the bottom 1000 kilometers of the Earth's mantle: Toward a hybrid convection model, *Science*, *283*, 1885-1888.

Vasco, D.W., L.R. Johnson, R.J. Pulliam and P.S. Earle (1994), Robust inversion of IASP91 travel-time residuals for mantle P and S velocity structure, earthquake mislocations, and station corrections, *J. Geophys. Res*., *99*, 13727-13755.

Wen, L.X. (2001), Seismic evidence for a rapidly varying compositional anomaly at the base of the Earth's mantle beneath the Indian Ocean, *Earth Planet. Sci. Lett.*, *194*(1-2), 83-95.

Wentzcovitch, R.M., J.F. Justo, Z. Wu, C.R.S. da Silva, D.A. Yuen and D. Kohlstedt (2009), Anomalous compressibility of ferropericlase throughout the iron spin cross-over, *Proc. Natl. Acad. Sci*., *106*(21), 8447-8452.

Wookey, J., S. Stackhouse, J.-M. Kendall, J. Brodholt and G.D. Price (2005), Efficacy of the post-perovskite phase as an explanation of lowermost mantle seismic properties, *Nature*, *438*, 1004–1008.

1010 **Figure Captions**

1011 **Figure 1.** Radially-symmetric viscosity profile and associated geodynamic sensitivity kernels.
1012 (a) Viscosity profile from *Mitrovica and Forte* [2004]. (b) Tectonic plate divergence kernels. (c)
1013 Free-air gravity kernels for the free-slip tectonic plate boundary condition. (d) Free-air gravity
1014 kernels for the no-slip boundary condition. See *Forte and Peltier* [1994] for further description
1015 of the dynamic coupling of the plate motions. (e) Dynamic surface topography sensitivity kernels
1016 for the free-slip boundary conditions (f) and no-slip boundary conditions. The CMB topography
1017 kernels are similarly shown (g, h).

1018

1019 **Figure 2.** Optimized relative heterogeneity ratios found through a very fast simulated annealing
1020 (VFSA) process. Only the 1-D part of the 1.5-D scaling models are shown. (Left) Density-shear
1021 scaling relationships where the blue line is the starting solution from *Simmons et al.* [2009], the
1022 red line is the solution found in the current study, and gray lines are the individual VFSA
1023 iterations. The possible 1-D solutions were bounded based upon the work of *Cammarano et al.*
1024 [2003] and *Karato and Karki* [2001]. (Right) Optimized 1-D compressional-shear velocity
1025 profile based on the fully coupled density-shear-compressional fields and joint inversions (see
1026 text).

1027

1028 **Figure 3.** VFSA cooling schedule and misfits for 1.5-D scaling model optimization iterations
1029 shown in Figure 2. (Top) We chose a rapid cooling schedule since, with each possible set of
1030 scaling model parameters, a full-scale non-linear joint inversion of all seismic and geodynamic
1031 data was performed. (Bottom) Normalized average data misfit to all data fields using scaling

1032 models at each VFSA iteration. The process was stopped after 61 models were accepted due to
1033 the very minor misfit variations at this stage.
1034
1035 **Figure 4.** Modeled 3-D scaling model distributions (dark shaded regions) compared to the 1-D
1036 components (blue lines) found through optimization procedures. (Left) Density-shear wave
1037 speed relationship distributions tend to have modes aligning with the inferred 1-D thermal
1038 values. The major exception is in the shallow upper mantle where cratons reside. Additionally,
1039 distributions skewing towards zero are present near the base of the upper mantle and the deepest
1040 mantle. (Right) Compressional-shear speeds relative heterogeneity ratios also tend to have
1041 distribution modes falling along the inferred 1-D thermal values. The distributions tend to be
1042 significantly broader than the density-shear values and are bi-modal in the D'' layer. For
1043 reference, the red and green lines correspond to $R_{P/S}^{-1} = R_{S/P}$ values of 2.7 and 1.7, respectively.
1044 The distributions of $R_{P/S}$ values were computed for the pure seismic models (TX2008s and P-
1045 only) for comparison (light gray shaded areas) to demonstrate the impact of forcing the solution
1046 to be more consistent with thermal variations. Note that there are a number of values near zero
1047 in the deep mantle. Therefore, the inverse, $R_{P/S}^{-1} = R_{S/P}$, is unbounded.
1048
1049 **Figure 5.** GyPSuM model shear wave heterogeneity field. The 'non-thermal' part of the field is
1050 computed through subtracting the result assuming the thermal scaling profile (left column) from
1051 the final result (right column) based on non-linear scaling and heterogeneity model inversions
1052 described in the text. Note that the color scales change for each panel according to the amplitude
1053 ('X') printed in the lower left corner of each panel. The old cratonic regions clearly show up as

1054 fast anomalies and the mid-ocean ridges and rifts are slow. Note that the 'non-thermal'
1055 heterogeneity levels are significantly smaller than the thermal signatures in the upper half of the
1056 mantle.
1057
1058 **Figure 5 (continued)** Linear fast zones attributed to past subduction beneath North America and
1059 south Asia clearly persist beyond 1800 km depth. Low-velocity zones dominate much of the
1060 deep mantle shear wave heterogeneity and are most evident in the African and Pacific
1061 superplume structures. High-velocity zones in the D'' layer generally occur beneath regions
1062 where subduction has occurred since the past supercontinent cycle [*Anderson* 2002] suggesting
1063 that slabs eventually reach the base of the mantle.
1064
1065 **Figure 6.** GyPSuM model compressional wave heterogeneity field. The depths and other details
1066 are the same as in Figure 5. Compositional variations play a major role in the development of
1067 compressional wave heterogeneity in the shallow upper mantle. This is most notable in cratonic
1068 regions, but other scattered non-thermal anomalies are present as well. We note that the P-wave
1069 information in this study consists of direct teleseismic arrivals and thus the resolvability of the
1070 these features is in terms of P-wave heterogeneity is very limited.
1071
1072 **Figure 6 (continued)** Non-thermal influences on P-wave heterogeneity are significant in the
1073 lower half of the mantle. Most notably, the thermally-induced high-velocity zones in the deep
1074 mantle are generally offset by non-thermal low-velocity signatures that mute the total high-
1075 velocity structures. However, the patterns of the low-velocity regions in the total field generally
1076 match the thermally-induced field.

1077

1078 **Figure 7.** GyPSuM model density heterogeneity field. The depths and other details are the same
1079 as in Figures 5 and 6. The thermally-induced high-density signatures in the cratonic roots are
1080 offset by the non-thermal contributions severely reducing the amplitude of the overall negative
1081 buoyancies. Outside of the cratons, upper mantle density structure is dominated by the effects of
1082 temperature variations.

1083

1084 **Figure 7 (continued)** The primary non-thermal contributions to the density field are within the
1085 superplume structures beneath Africa and the Pacific Ocean. The central cores of these features
1086 are found to be intrinsically dense, offsetting the effects of temperature. In some localized zones
1087 within the superplumes, we find near neutral buoyancies and mild negative buoyancies which
1088 may account for the stabilization of the deep part of these structures over long time scales.

1089

1090 **Figure 8.** Root-mean-squared (RMS) amplitudes of wave speeds and density. Each field is
1091 separated into thermal and non-thermal contributions and RMS amplitudes are subsequently
1092 computed. Shear wave speeds are dominantly produced by variations in temperature throughout
1093 the mantle. Compressional wave heterogeneity is also largely controlled by variations in
1094 temperature, but to a lesser extent than shear waves given the strong compositional influence on
1095 bulk sound speed. Thermally induced density anomalies are typically larger than
1096 compositionally induced structures. However, the mass depletion in the cratons and the intrinsic
1097 high density observed in parts of the superplume structures produces opposing compositional
1098 anomalies that rival the thermally induced structures. Sound speed variations are dominated by
1099 compositional variations throughout most of the mantle. Of particular interest is the

compositionally induced amplitudes that increase with depth from ~1500 km depth to the CMB. The low amplitude of the thermally-induced sound speed field in the deep mantle is a product of δ_S values approaching 1. Therefore, the relative behaviors of the bulk modulus and density are very similar in the presence of temperature variations in the lower mantle providing very small variations without compositional anomalies.

Figure 9. Depth-dependent correlation of shear wave heterogeneity with P-wave speeds, sound speeds and density. (Left) Correlations considering all structure. Shear waves are generally highly correlated with P-wave and density heterogeneities throughout the mantle. However, sound speed is often de-correlated or mildly anti-correlated with shear wave heterogeneity. (Center) Correlations based on low shear velocity regionalization. When we consider only the low shear velocity structures in the correlation measurement, we find that P-wave anomalies corresponding to these zones remain fairly well correlated. However, density heterogeneity systematically de-correlates with shear speed in the deepest mantle primarily owing to the intrinsic high-density signatures of the superplume structures. (Right) Correlations based on high shear velocity regionalization. Fast shear velocity structure is highly correlated with density throughout the lower mantle. P-wave heterogeneity in regions where shear velocity is fast systematically de-correlates with S beginning at ~1500 km depth demonstrating the distinctly different properties of the high shear velocity blobs relative to the low velocity structures that appear to be primarily products of temperature variations.

1123

1124

1125

1126

1127

1128

1129

Table 1. Constraints on mantle heterogeneity employed in this study.

Constraint	Description	Quantity	Source
S-waves	Residual travel times for globally distributed *S*, *ScS*, *sS*, *sScS*, *SKS*, *SKKS* phases. Data also includes surface reflected multiples (e.g. *SSS, ScS3*, etc.) and triplicated phases turning within the upper mantle.	~46,000	*Grand* [2002] *Simmons et al.* [2007]
P-waves	Residual travel times for teleseismic P-waves based on a recompilation of International Seismic Centre (ISC) data (EHB). Arrivals are condensed using a summary ray approach over 2 x 2 degree bins.	~626,000	*Engdahl et al.* [1998] (EHB) *Antolik et al.* [2003]
Gravity	Global free-air gravity field from the EGM96 geopotential model derived from a compilation of land-, air-, and space-based observations.	spherical harmonic degrees 2-16	*Lemoine et al.* [1998]
Plate Motions	Divergence of all tectonic plates calculated from the NUVEL-1 plate velocity model.	spherical harmonic degrees 1-16	*DeMets et al.* [1990]
Dynamic Topography	Non-isostatic topography of Earth's free surface. Estimated from removal of the crustal isostatic topography signal on the basis of CRUST2.0.	spherical harmonic degrees 1-16	*Bassin et al.* [2000] *Forte and Perry* [2000]
CMB Ellipticity	Excess ellipticity of the core-mantle boundary (CMB) relative to Earth's surface. The CMB is elevated by ~400 m along the equator according to studies of Earth's free-core nutation processes.	zonal harmonic degree 2	*Herring et al.* [2002] *Mathews et al.* [2002]
Mineral Physics Parameters	Bounds on the relative heterogeneity ratios ($R_{\rho/S} = d\ln\rho / d\ln V_S$ and $R_{P/S} = d\ln V_P / d\ln V_S$) due to pure thermal variations. Anharmonic and anelastic effects included.	N/A	*Karato and Karki* [2001] *Cammarano et al.* [2003]

Table 2. Forward model variables.

Variable	Description	Details
$\mathbf{L}_S$	Shear wave sensitivity kernels	1-D ray path lengths 46,000 x 99,148 sparse elements
$\mathbf{L}_P$	Compressional wave sensitivity kernels	1-D ray path lengths 620,000 x 99,148 sparse elements
$\mathbf{G}$	Geodynamic data viscous flow response kernels	Spatial sensitivities of spherical harmonic components up to degree 16; free-air gravity, plate divergences, dynamic topography 861 x 99,148 full elements
$\mathbf{c}$	CMB ellipticity kernel	Spatial sensitivity of the degree 2 zonal harmonic 1 x 99,148 full elements (row vector)
$\mathbf{D}$	Regularization operator	2nd-order digital smoothing filter
$\lambda_P, \lambda_G, \lambda_{CMB}, \lambda_D$	Weights	Relative weighting between data sets ($\lambda_P, \lambda_G, \lambda_{CMB}$); regularization weight (λ_D)
V_S , V_P , ρ	Wave speeds and density	Starting model velocities and density from PREM and TNA/SNA models (see text)
$R_{P/S}$	Compressional/shear wave relative heterogeneity ratio	$R_{P/S} = d \ln V_P / d \ln V_S$
$R_{\rho/S}$	Density/shear wave relative heterogeneity ratio	$R_{\rho/S} = d \ln \rho / d \ln V_S$
$\mathbf{r}_S$	Shear wave travel time residual measurements	~46,000 residuals (column vector)
$\mathbf{r}_P$	Compressional wave travel time residual measurements	~626,000 residuals (column vector)
$\mathbf{g}$	Geodynamic data	Spherical harmonic coefficients up to degree 16 of the free-air gravity, plate divergences, dynamic topography; 861 elements (column vector)
e	CMB ellipticity	Degree 2 zonal harmonic coefficient representing 400m of excess ellipticity on the core-mantle boundary
$\Delta\mathbf{m}_S$	Shear wave slowness perturbation model	Heterogeneity model to be determined through inversion

1131

1132

Table 3. Model fits to all data considered.

Model Name	Scaling Models $R_{\rho/S}$, $R_{P/S}$	S-wave arrivals (%)	P-wave arrivals (%)	Free-air gravity (%)	Tectonic plate divergence (%)	Dynamic surface topography (%)	CMB excess ellipticity (km)
TX2008s[a]	1-D[b]	94	12	1	48	-44	1.3
P-only[c]	1-D[b]	56	32	-57	30	-164	1.3
TX2008j[a]	3-D[b]	93	12	91	99	80	0.4
GyPSuM_1D[d]	1-D[b]	90	28	69	97	-70	0.4
GyPSuM_1.5D[d]	1.5D[d]	90	29	73	97	32	0.4
GyPSuM[e]	3-D[e]	93	31	88	99	72	0.4

Fits are expressed as variance reduction (in percent) except for excess CMB ellipticity expressed in km (0.4 km is the expected value).

[a]Shear wave models produced with seismic data alone (TX2008s) and joint inversion of shear wave travel times and geodynamic information (TX2008j) from *Simmons et al.* [2009].

[b]Optimum 1-D and 3-D density-shear scaling models found on the basis of joint seismic-geodynamic inversion [*Simmons et al.* 2009] and computed compressional-shear scaling (see text).

[c]Compressional velocity model produced using only P-wave data described in the text. TX2008s and P-only have similar model roughness. Models are for fit comparison only.

[d]Joint P-S-geodynamic models developed using 1-D scaling (GyPSuM_1D) and a modified scaling model that accounts for 1st- order effects of cratonic mass depletion and the temperature dependence of Q (GyPSuM_1.5D). Scaling model 1.5D was found through the simulated annealing process described in the text.

[e]Final P-S-density and scaling models developed in this study.

1133

1134

Table 4. Scaling model optimization parameters.

General form: $R_{\rho/S} = aR_{\rho/S}^{1\text{-}D} + b + \kappa\delta \ln V_S$

Parameter	Description	VFSA solution
$R_{\rho/S}^{1\text{-}D}$	Starting density-shear velocity relative heterogeneity ratio from *Simmons et al.* [2009].	N/A
$\delta \ln V_S$	Shear wave velocity heterogeneity	N/A
κ	Velocity-based scaling model correction derivatives of the form $\partial R_{\rho/S} / \partial \ln V_S$	N/A
a_{um}	Amplification factor applied to the starting 1-D upper mantle density-velocity scaling profile.	0.916
b_{um}	Additive term that adjusts the mean of the 1-D upper mantle scaling profile.	0.0157
a_{lm}	Amplification factor applied to the starting 1-D lower mantle density-velocity scaling profile.	0.651
b_{lm}	Additive term that adjusts the mean of the 1-D lower mantle scaling profile.	0.0619
κ_{craton}^{100}	Scaling model correction derivative defined only in the cratonic roots in the upper 100 km.	-5.92
κ_{craton}^{250}	Scaling model correction derivative defined only in the cratonic roots from 100 to 250 km depth.	-3.13
$\kappa_{thermal}^{100}$	Scaling model correction derivative in the non-cratonic upper 100 km.	1.26
$\kappa_{thermal}^{250}$	Scaling model correction derivative in the non-cratonic mantle in the 100-250 km depth range.	1.59
$\kappa_{thermal}^{650}$	Scaling model correction derivative in the non-cratonic mantle in the 250-650 km depth range.	0.341

1135

1136

Figures

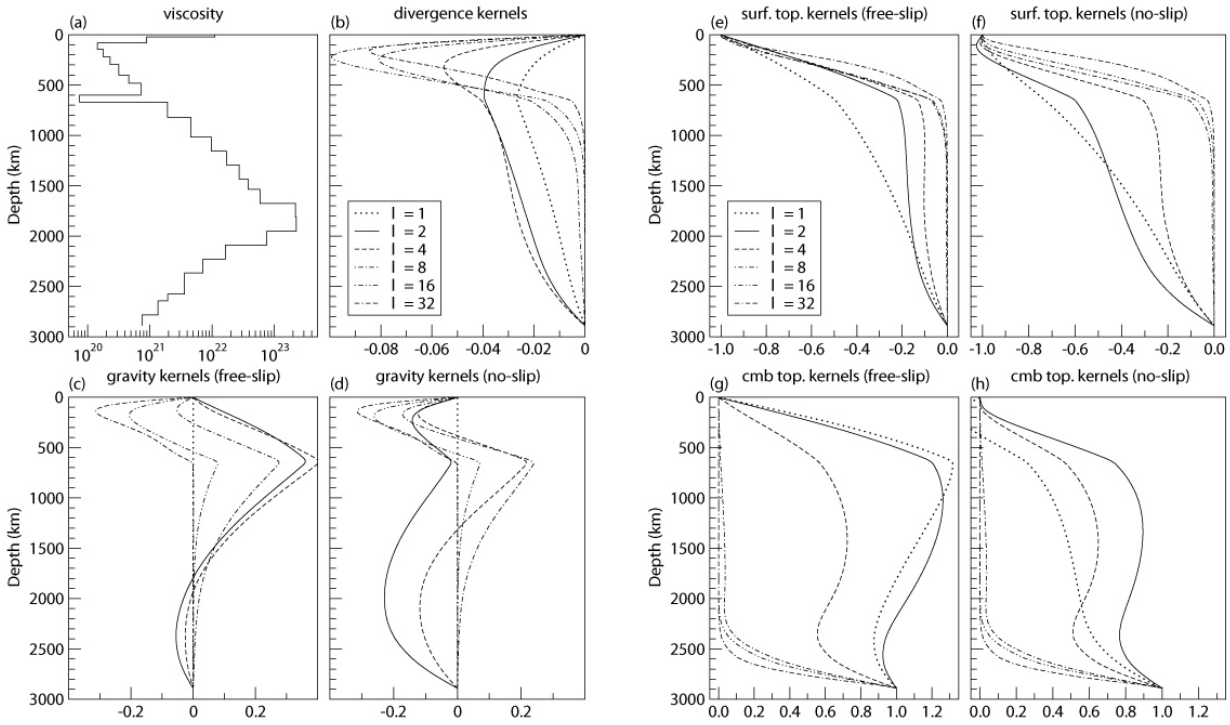
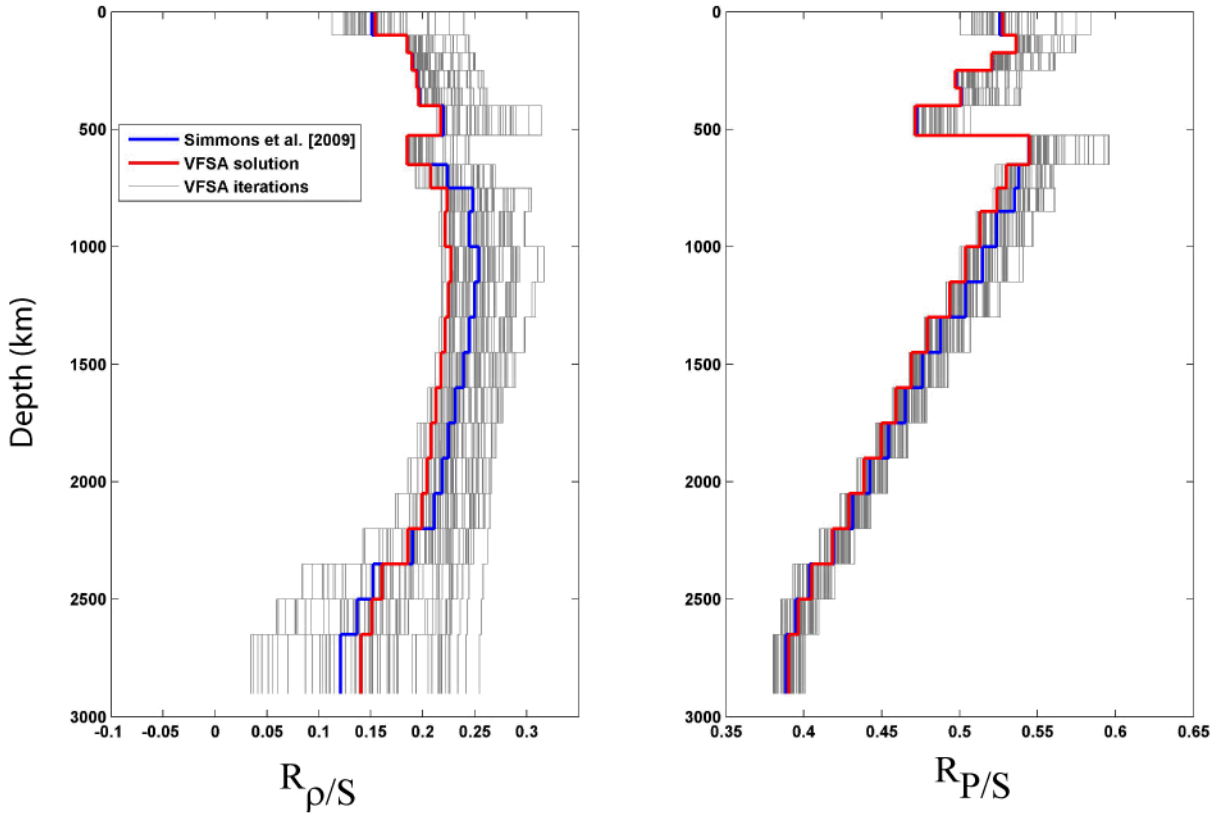


Figure 1



1141

1142

1143 **Figure 2**

1144

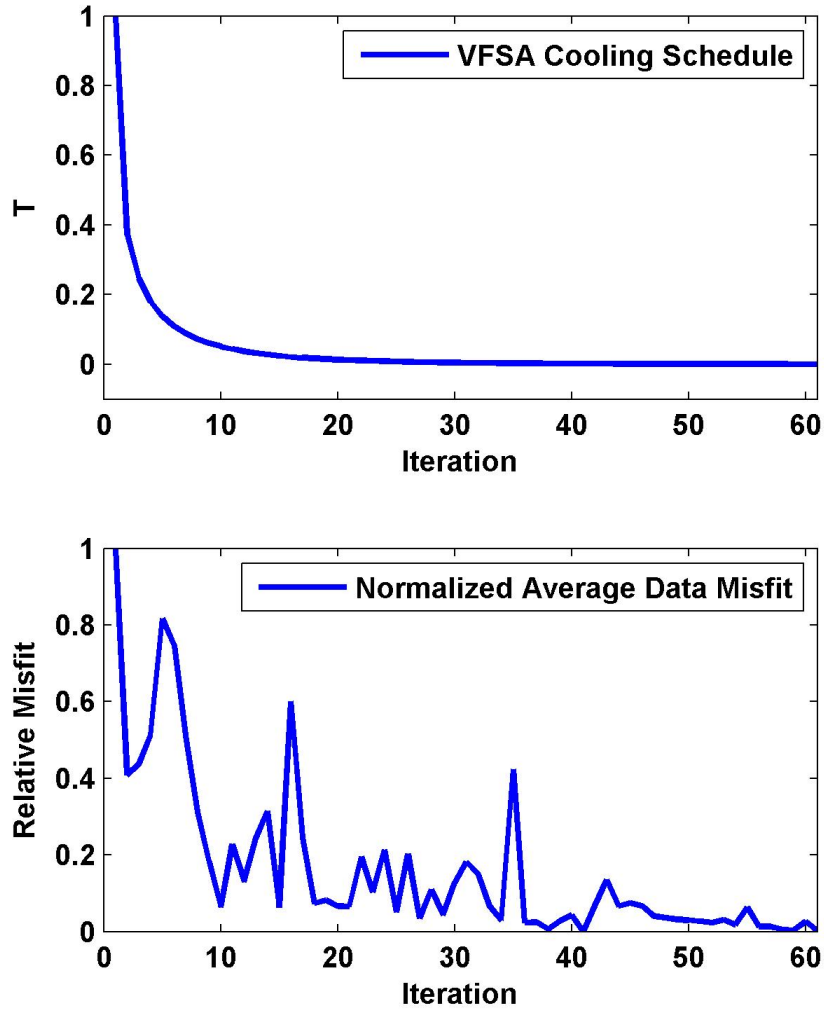

VFSA Cooling Schedule
T
Iteration
Normalized Average Data Misfit
Relative Misfit
Iteration

Figure 3

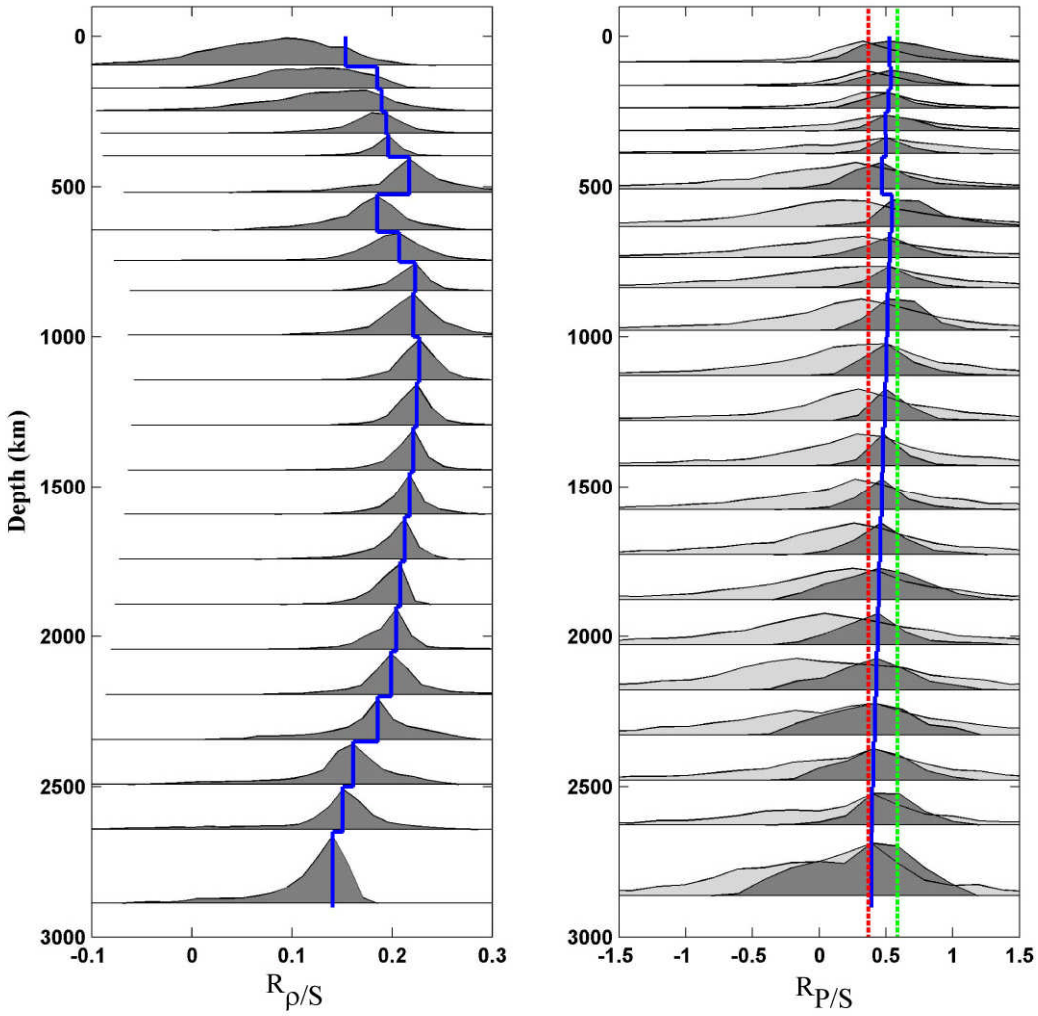
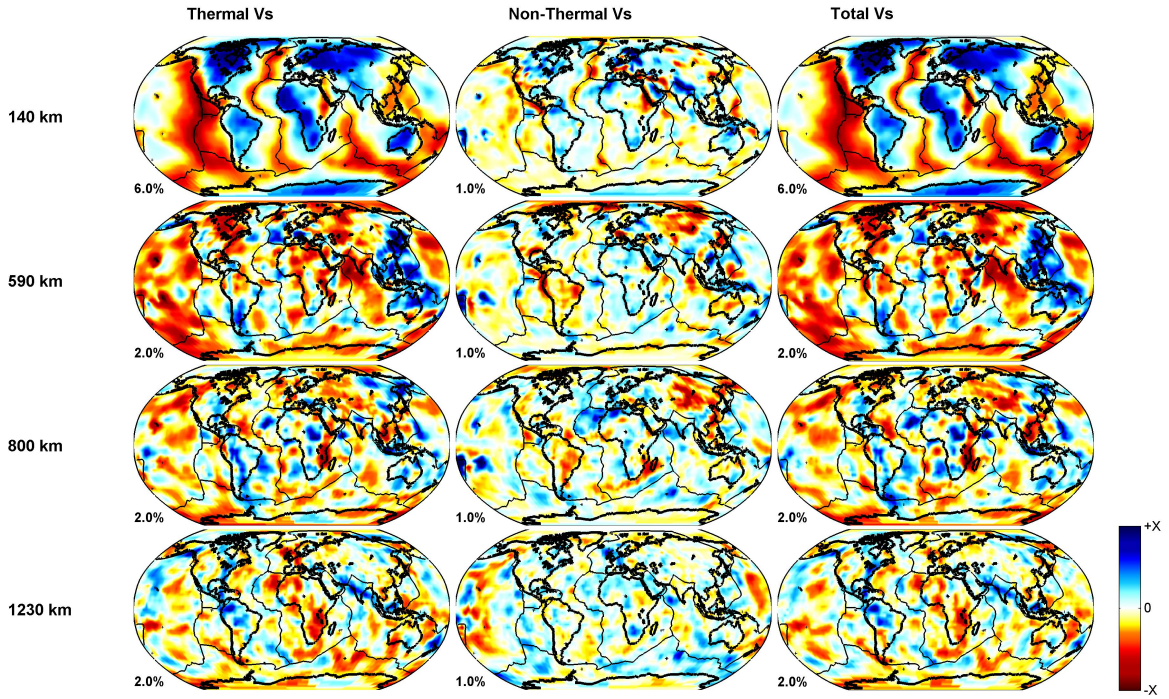
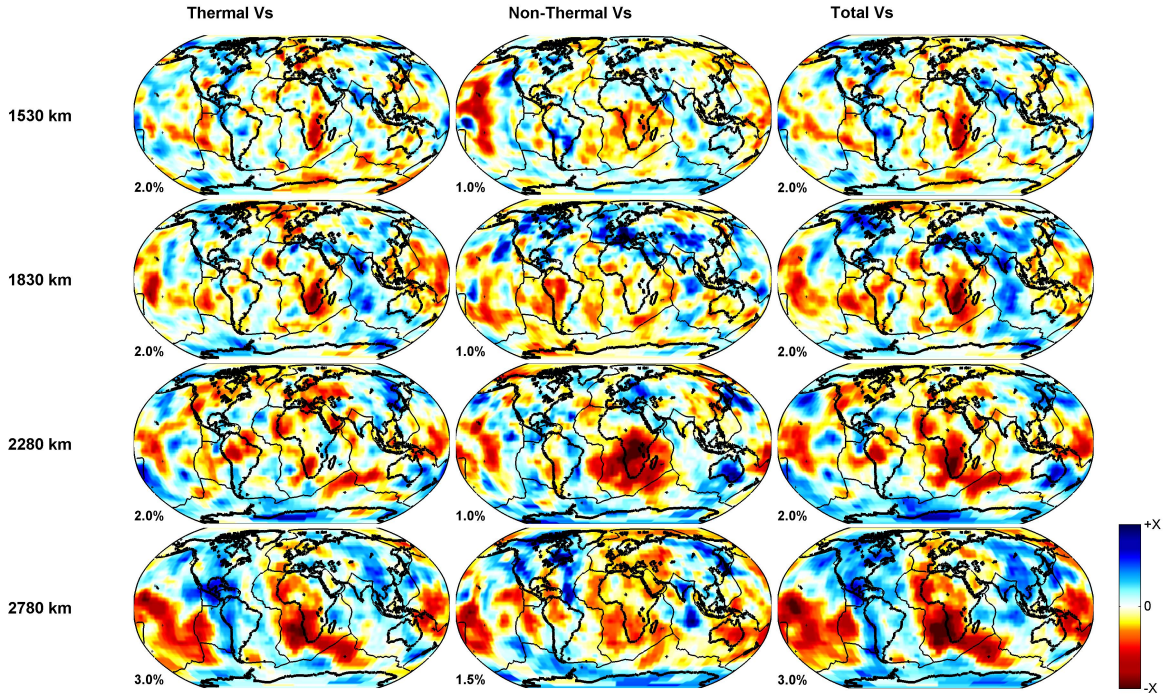

Depth (km)
0
500
1000
1500
2000
2500
3000
-0.1
0
0.1
0.2
0.3
$R_{\rho/S}$
-1.5
-1
-0.5
0
0.5
1
1.5
$R_{P/S}$

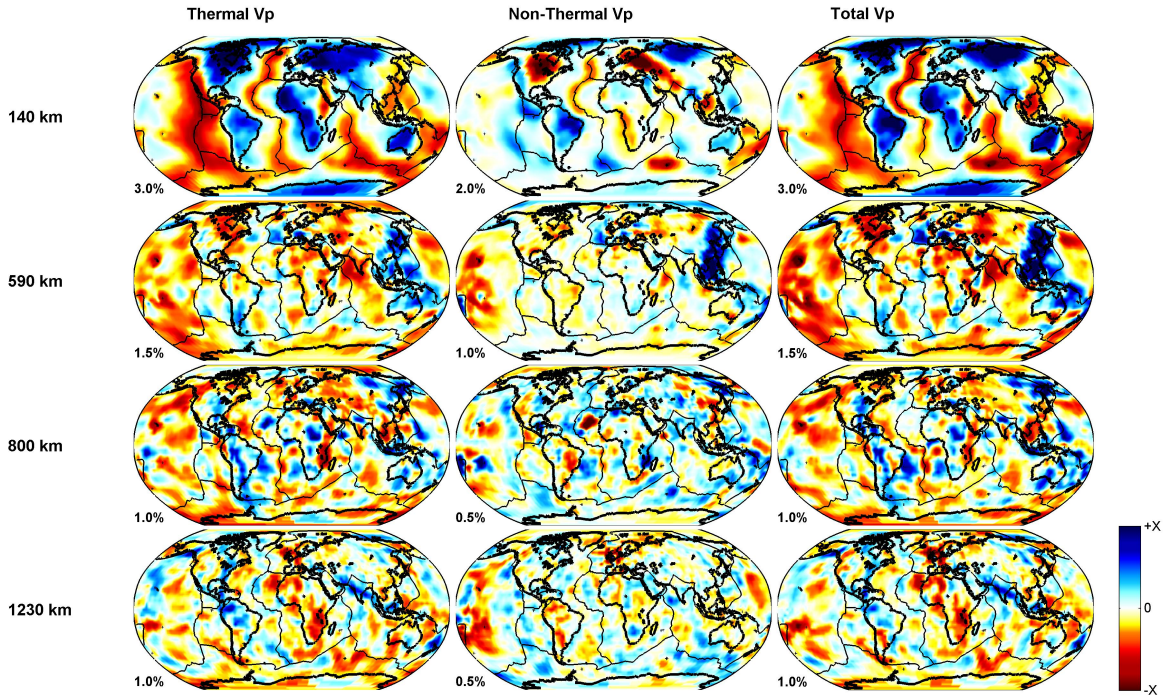
Figure 4



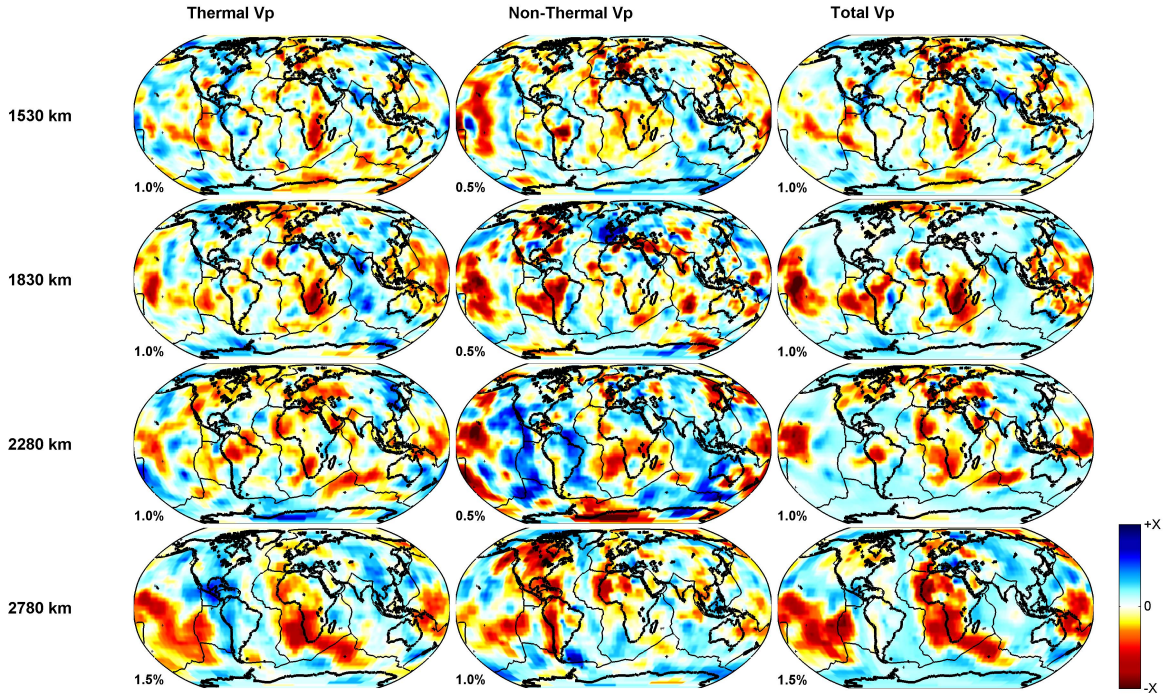
1154



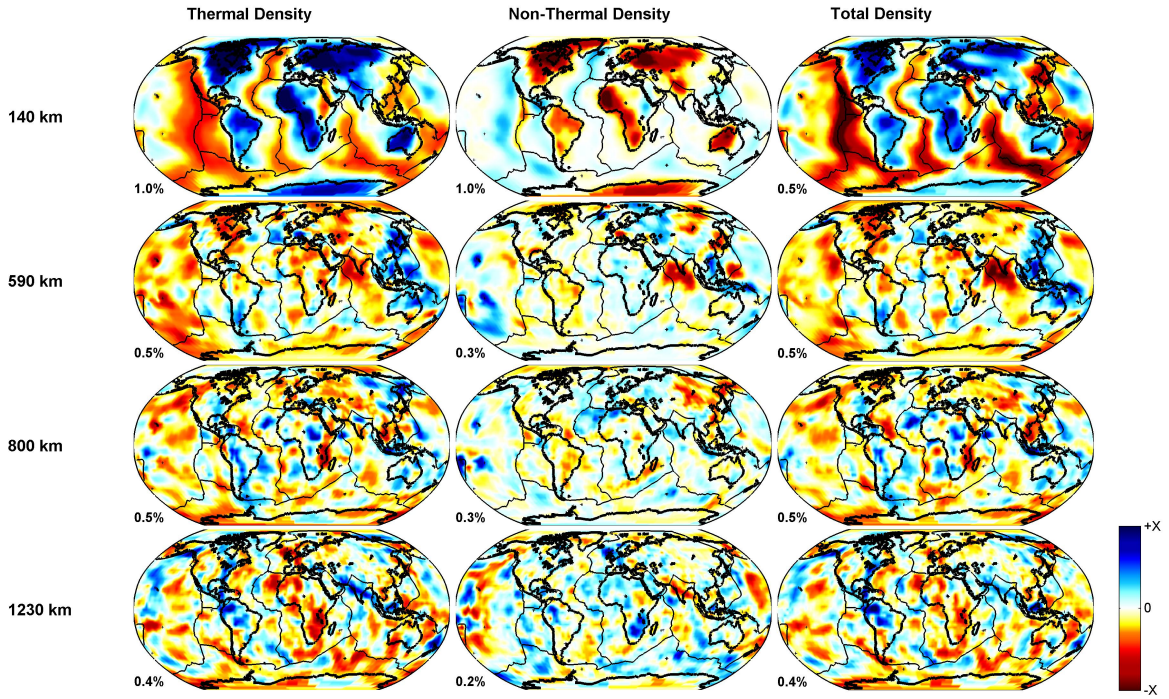
1155
1156 **Figure 5**



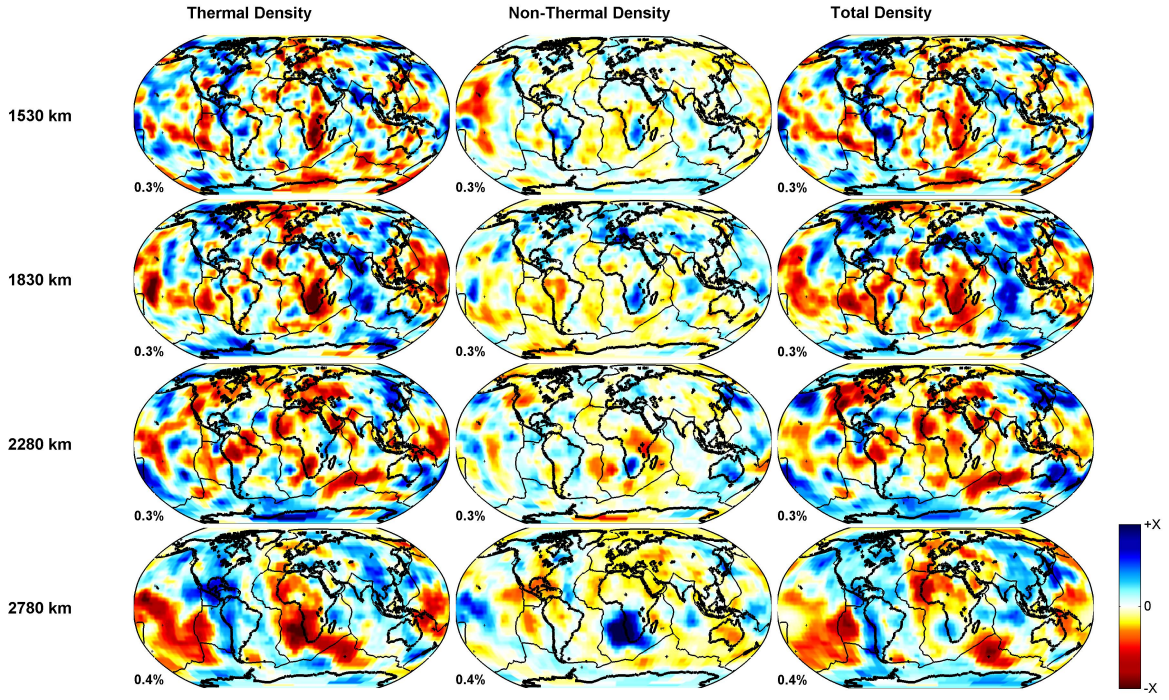
1157



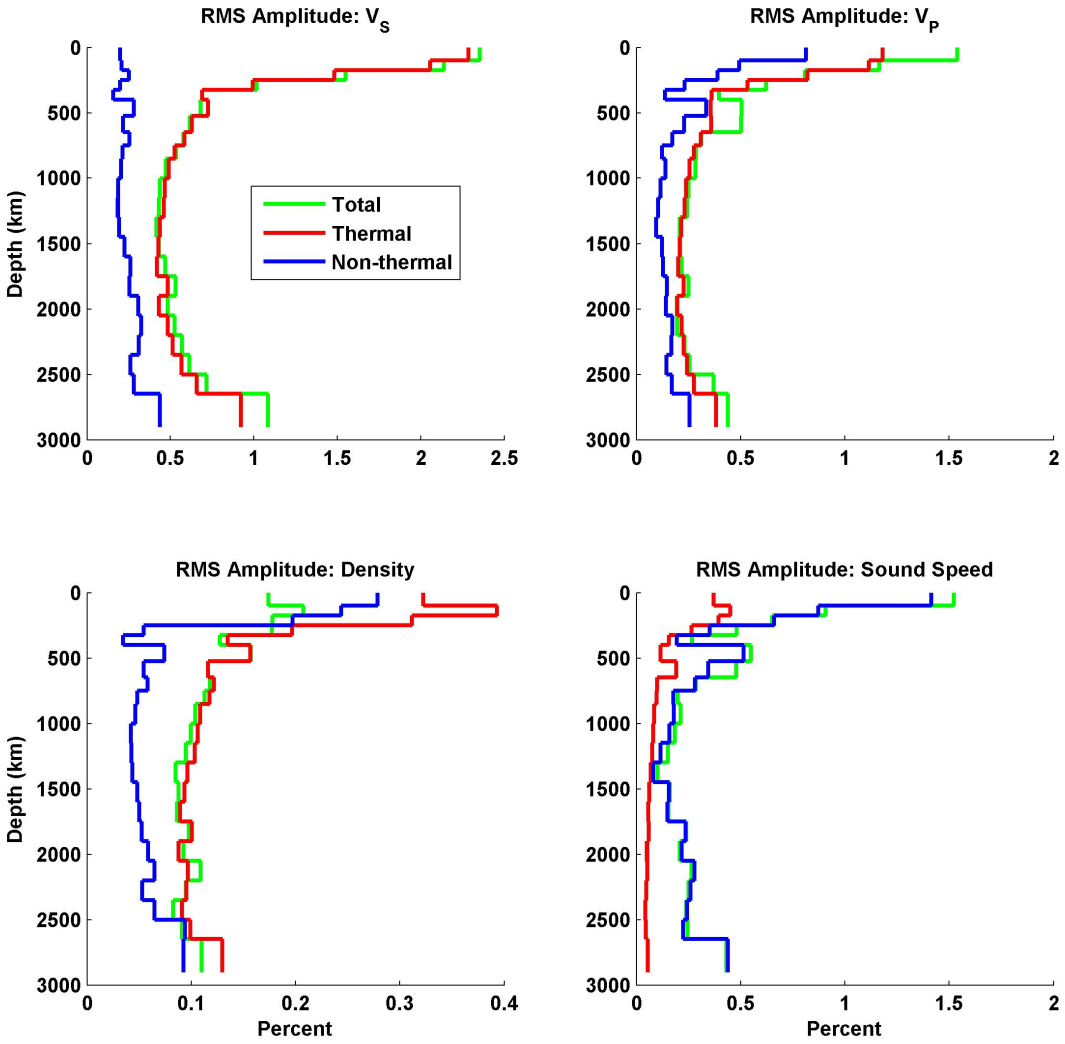
1158
1159 **Figure 6**



1160



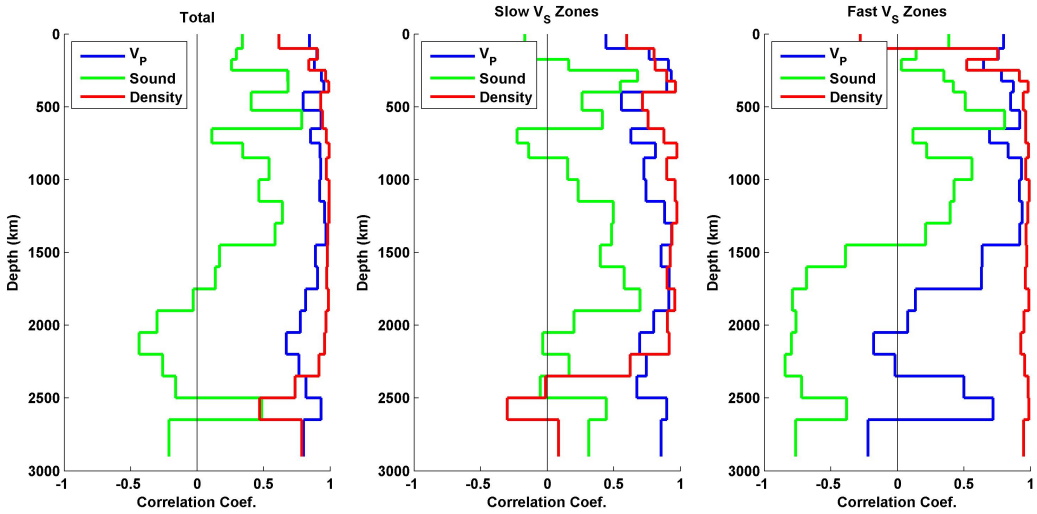
1161
1162 **Figure 7**



1163

1164 **Figure 8**

1165



1166

1167 **Figure 9**