JOURNAL OF GEOPHYSICAL RESEARCH: SOLID EARTH, VOL. 118, 1–20, doi:10.1002/jgrb.50391, 2013

# Global *P* wave tomography of Earth's lowermost mantle from partition modeling

M. K. Young,[1] H. Tkalčić,[1] T. Bodin,[2] and M. Sambridge[1]

Received 2 June 2013; revised 26 September 2013; accepted 27 September 2013.

[1] Determining the scale-length, magnitude, and distribution of heterogeneity in the lowermost mantle is crucial to understanding whole mantle dynamics, and yet it remains a much debated and ongoing challenge in geophysics. Common shortcomings of current seismically derived lowermost mantle models are incomplete raypath coverage, arbitrary model parameterization, inaccurate uncertainty estimates, and an ad hoc definition of the misfit function in the optimization framework. In response, we present a new approach to global tomography. Apart from improving the existing raypath coverage using only high-quality cross-correlated waveforms, the problem is addressed within a Bayesian framework where explicit regularization of model parameters is not required. We obtain high-resolution images, complete with uncertainty estimates, of the lowermost mantle *P* wave velocity structure using a hand-picked data set of *PKPab-df*, *PKPbc-df*, and *PcP-P* differential travel times. Most importantly, our results demonstrate that the root mean square of the *P* wave velocity variations in the lowermost mantle is approximately 0.87%, which is 3 times larger than previous global-scale estimates.

**Citation:** Young, M. K., H. Tkalčić, T. Bodin, and M. Sambridge (2013), Global *P* wave tomography of Earth's lowermost mantle from partition modeling, *J. Geophys. Res. Solid Earth*, *118*, doi:10.1002/jgrb.50391.

## 1. Introduction

[2] The lowermost mantle is one of the most intriguing and important layers of the Earth. Extreme contrasts in velocity, density, and viscosity are seen at the core mantle boundary, where the solid mantle meets the liquid core. Additionally, the lowermost mantle is heterogeneous in terms of viscosity, seismic velocity, density, chemistry, and temperature. The pattern of velocity heterogeneity has been studied at a variety of scale lengths; for example, global models clearly reveal long wavelength velocity patterns [e.g., *Wysession*, 1996; *Antolik et al.*, 2003; *Lei and Zhao*, 2006; *Li et al.*, 2008; *Ritsema et al.*, 2011]. In those studies, regions of higher wave speeds are typically attributed to cool subducted slabs while regions of slower wave speeds are regarded as the signature of hot, upwelling material. A pattern of large areas of low velocity under the southwest Pacific and southern Africa is well established in the deep Earth community [e.g., *Su et al.*, 1994; *Li and Romanowicz*, 1996; *Wysession*, 1996; *Grand et al.*, 1997; *Mégnin and Romanowicz*, 2000; *Gu et al.*, 2001; *Antolik et al.*, 2003; *Lei and Zhao*, 2006; *Houser et al.*, 2008; *Li et al.*, 2008; *Della Mora et al.*, 2011; *Soldati et al.*, 2012]. The African anomaly extends about 1000 km above the core mantle boundary [e.g., *Ritsema et al.*, 2011] and is likely both thermal and chemical in nature [e.g., *Ni and Helmberger*, 2001a; *Simmons et al.*, 2007]. There is increasing evidence that the Pacific anomaly is also at least partially chemical in origin [e.g., *Tkalčić and Romanowicz*, 2002; *Trampert et al.*, 2004; *Ishii and Tromp*, 2004; *He and Wen*, 2009]. Meanwhile, local studies of core phases and their precursors indicate the presence of short-scale heterogeneity [e.g., *Doornbos*, 1974; *Bataille and Flatte*, 1988; *Vidale and Hedlin*, 1998; *Bréger and Romanowicz*, 1998; *Cormier*, 1999; *Bréger et al.*, 2000; *Rost and Earle*, 2010; *Earle et al.*, 2011]. A goal of this study is to investigate the strength of the velocity perturbations, as there is evidence that this amplitude is far greater than the level implied by current global tomographic models [e.g., *Ritsema et al.*, 1998; *Garcia et al.*, 2009].

[3] There is increasing research to support a fine-scale and complex nature of heterogeneity at the base of the mantle. Developing a global model of lowermost mantle structure may prove vital to a more precise understanding of magnetic field generation, deep mantle flow, and other geodynamical processes. Common impediments to a reliable model include limited sampling due to the natural global seismicity pattern and seismic station availability, questionable quality of large data sets, influence by crustal and mid-upper mantle structure, artificially coarse model parameterizations, and inadequate quantification of data noise. Nonetheless, considerable advances in seismic tomography have been made since early global mantle studies [*Dziewonski et al.*, 1977; *Nakanishi and Anderson*, 1982; *Dziewonski*, 1984], which

---

[1] Research School of Earth Sciences, Australian National University, Canberra, Australian Capital Territory, Australia.

[2] Berkeley Seismological Laboratory, University of California, Berkeley, California, USA.

Corresponding author: M. K. Young, Research School of Earth Sciences, Australian National University, Bldg. 61 Mills Rd., Acton, Canberra, ACT 0200, Australia. (mallory.young@anu.edu.au)

2169-9313/13/10.1002/jgrb.50391

almost exclusively depended on absolute, *P* wave travel times and a simple block or spherical harmonic parameterization. Differential travel times are less sensitive to source mislocation and heterogeneities near the source and receiver. *Cormier and Choy* [1986] demonstrated this advantage and used differential *PKP* wave measurements to assess velocity heterogeneity in and near the inner core. *Woodward and Masters* [1991] then applied the differential travel time method to the mantle and used *PP-P* and *SS-S* measurements to map global upper mantle structure. It then became common to combine different data sets to improve spatial coverage and depth resolution. *Su et al.* [1994] used a synthesis of full waveform data, absolute *S* arrivals, and *SS-S* and *ScS-S* differential travel times to invert for a 3-D map of mantle shear velocity structure.

### 1.1. Advances in Data Quality

[4] Over the last couple of decades, many global models of *P* wave velocity in the mantle have been produced [e.g., *Inoue et al.*, 1990; *Vasco et al.*, 1995; *van der Hilst et al.*, 1997; *Vasco and Johnson*, 1998; *Antolik et al.*, 2003; *Fukao et al.*, 2003; *Montelli et al.*, 2004; *Lei and Zhao*, 2006; *Houser et al.*, 2008; *Li et al.*, 2008; *Zhao*, 2009; *Della Mora et al.*, 2011; *Zhao et al.*, 2012]. Nonetheless, *P* wave mantle models are still less common than *S* wave mantle models as a result of the relative difficulty in measuring *P* wave arrivals. The challenge of accruing a collection of *P* wave arrivals is due to several factors. First, the source mechanism radiation pattern of strike slip earthquakes does not produce much *P* wave energy down to the core mantle boundary. Also, the conversion from *P* waves to *S* waves is often more efficient than the conversion of *S* waves to *P* waves. Additionally, *S* waves can be viewed on both the transverse and radial component of the seismogram, whereas *P* waves are best observed on just the vertical component (at epicentral distances beyond 145° and often close to antipodal distances, the vertical and radial components of a seismogram are almost the same due to the steep incident angle of *PKP* waves). Finally, *S* wave models are typically derived from lower frequency data, which are more conducive to waveform modeling and inversion. *P* waves, especially *PcP*, on the other hand must be observed at higher frequencies, often leaving them buried in noise and susceptible to dispersion in the lowermost mantle.

[5] By inverting for the entire mantle structure instead of the lowermost mantle alone, mid-upper mantle heterogeneity is less likely to be erroneously mapped onto the bottommost layer. These whole mantle models are also useful for assessing heterogeneity at the base of the mantle in the sense that they reveal any vertical continuity of structure, such is especially pertinent when investigating the presence of plumes [*Zhao*, 2004]. The chief drawback of entire mantle models, however, is the quality of the data; it is far too time consuming to manually pick a data set for whole mantle imaging. Therefore, researchers generally use International Seismological Center (ISC) or *Engdahl et al.* [1998] (as in *Li et al.* [2008]) data sets, which are of questionable quality for many phases [*Inoue et al.*, 1990; *Vasco et al.*, 1995; *van der Hilst et al.*, 1997; *Vasco and Johnson*, 1998; *Boschi and Dziewonski*, 2000; *Fukao et al.*, 2003; *Zhao*, 2004; *Montelli et al.*, 2004; *Lei and Zhao*, 2006; *Zhao*, 2009; *Della Mora et al.*, 2011; *Zhao et al.*, 2012]. Despite the potentially enormous data sets, model resolution in the lowermost mantle is often limited due to the computational costs associated with the whole mantle model parameterization.

[6] The heterogeneity pattern above the core mantle boundary is frequently better retrieved when inverting for the lowermost mantle region alone. By reducing unwanted crustal and mid-upper mantle effects, differential travel times of body waves are especially helpful when deciphering the complexities of the mantle's bottommost layer. For example, *Sylvander and Souriau* [1996] used *PKPab-bc* differential travel times to retrieve *P* velocity structure in the lowermost mantle. A limitation of this approach, however, is incomplete raypath sampling due to the unfavorable configuration of earthquakes and land-based seismic stations. This situation can in part be ameliorated by using multiple seismic phases with different sampling patterns. For example, *Kárason and van der Hilst* [2001] used *P*, *pP*, and *pwP*, coupled with *PKPab-df*, *PKPab-bc*, and *PKPdf-Pdiff* differential travel times, and *Tkalčić et al.* [2002] used cross-correlated *PKPab-df* and *PcP-P* data to invert for lowermost mantle structure. In this study, we augment the data sets of *Tkalčić et al.* [2002] and *Tkalčić* [2010] by adding new *PKPab-df*, *PKPbc-df*, and *PcP-P* differential travel times with the intent of patching the spatial gaps in sampling.

### 1.2. Advances in Inverse Methods

[7] In addition to the steady improvement in data quality and quantity, in large part thanks to accumulated recordings of the Global Seismographic Network that the Incorporated Research Institutions for Seismology (IRIS) initiated in 1984, there has also been improvement in inversion methods. This progress is to a great degree a reflection of an increase in computing power. Irregular-sized block parameterization was first introduced by *Inoue et al.* [1990], which helped reduce over parameterization while at the same time allowing the resolution of small-scale structures where justified by the data. *Tkalčić et al.* [2002] also employed a variable size block inversion method and introduced an algorithm to restrict the aspect ratios of the blocks. *Antolik et al.* [2003] used a horizontal tessellation of spherical splines that provided the equivalent resolution as spherical harmonic degree 18 for a joint inversion of *P* wave and *S* wave velocity in the mantle. *Sambridge and Faletič* [2003] presented a self-adaptive inversion technique in which the initial parameterization, based on spherical triangles and Delaunay tetrahedra, is refined throughout the inversion. Recently, *Simons et al.* [2011] and *Chevrot et al.* [2012] proposed a spherical wavelet approach to global tomographic inversions whose multiresolution potential is ideal for spatially ill-distributed data sets. *Zhao et al.* [2012] adopted a flexible-grid parameterization for whole mantle tomography that was designed to better express mantle structure under the polar regions.

[8] In addition to increasingly sophisticated model parameterizations, inversion algorithms have also been improved. *Beghein et al.* [2002] used the neighborhood algorithm of *Sambridge* [1999] to invert for spherical harmonic degree 2 *P* and *S* wave velocity structure in the mantle. *Garcia et al.* [2009] approached lowermost mantle heterogeneity from a different perspective and used stochastic analysis to relate differential travel time measurements to the correlation function of velocity heterogeneity. Using *PKP* travel

times, they inverted for the statistical properties of velocity perturbations in the lowermost mantle and provided valuable constraints on the scale length and magnitude of *P* wave heterogeneity that are unbiased by smoothing, damping, or model parameterization. More modern tomography methods began to step away from classical ray theory when *Montelli et al.* [2004] implemented a finite frequency approach to travel time tomography and demonstrated that wavefront healing cannot be ignored when using long-period *P* and *PP* waves. Finite frequency effects were demonstrated to be especially important for *PKP* phases in the lowermost mantle [*Calvet and Chevrot*, 2005], although only recently were the first exact and complete finite frequency kernels for short period *PKP* waves computed [*Fuji et al.*, 2012]. *Garcia et al.* [2004, 2006] introduced another method known as Simulated Annealing Waveform Inversion of Body waves (SAWIB), which resolved the interference between direct *PKP* waves and their corresponding depth phases, thereby allowing the use of recordings from triplication distances and shallow earthquakes.

[9] If we are to accurately depict the scale-length of heterogeneity at the core mantle boundary, we must insure that the results are minimally influenced by arbitrary choices about model parameterization, smoothing, and damping. Such intervention is often the downfall of traditional linearized inversion methods, where models are artificially parameterized and either resolution is lost due to unnecessary smoothing or noise is misinterpreted as data complexity. In response to these pitfalls, we implement a fully nonlinear Bayesian partition-modeling technique for our tomographic inversion. This is an ensemble inference approach in which model space is sampled via Reversible Jump Markov Chain Monte Carlo sampling [*Green*, 1995, 2003]. Model parameters, including the level of data noise, are treated as unknowns in the problem and are represented by probability distribution functions rather than single values. Therefore, the complexity and amplitude of the velocity variations in our final model, which results from a complete ensemble of hundreds of thousands of sampled models, depend almost exclusively on the data itself, and there is no need to apply explicit smoothing or damping procedures [*Bodin et al.*, 2012a]. This method provides the statistical robustness of *Beghein et al.* [2002] and *Garcia et al.* [2009] and the multiscale resolution capabilities of *Simons et al.* [2011] and *Chevrot et al.* [2012] while introducing a novel approach to handling data noise and model parametrization.

[10] Conventional approaches to global tomography have two major drawbacks, both of which we seek to address in this study. The first is ad hoc data noise estimation and the subjective choice of smoothing and damping parameters. This can lead to the loss of valuable information, especially about model discontinuities, in an effort to avoid the overinterpretation of noise. We instead invert for the data uncertainty, and the level of complexity and strength of perturbations in the final model are appropriately limited by the noise content of the data [e.g., *Bodin et al.*, 2012a]. The second disadvantage is arbitrary model parameterization, often governed by equal-sized block cells [e.g., *Sylvander and Souriau*, 1996; *Tkalčić et al.*, 2002; *Houser et al.*, 2008] or limited-degree spherical harmonic expansions [e.g., *Tanaka*, 2010; *Ritsema et al.*, 2011]. Such rigid parameterizations prevent full utilization of data content while either oversimplifying and oversmoothing the model or adding unjustified complexity and artificial discontinuities. In this study, we partition the lowermost mantle into a mosaic of Voronoi polygons whose size, location, velocity, and number vary throughout the inversion according to the information content of the data. This probabilistic, partition modeling based approach enables us to present a global-scale model of *P* wave velocity variations in the lowermost mantle together with uncertainty estimates. Our choice of inversion method also allows valuable estimates of the scale length and strength of velocity heterogeneity in the lowermost mantle.

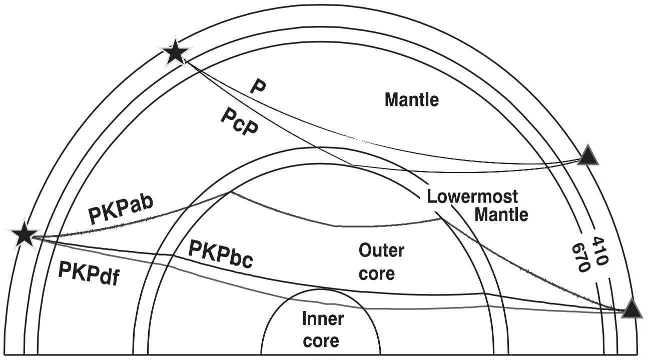


**Figure 1.** Raypaths of *PKPab*, *PKPbc*, *PKPdf*, *PcP*, and *P* body waves for the epicentral distance of 155° and 65°.

## 2. Data

[11] As a starting point, we use 1408 *PKPab-df*, 1068 *PKPbc-df*, and 399 *PcP-P* differential travel times taken from the hand-picked data sets of *Tkalčić et al.* [2002] and *Tkalčić* [2010]. Hand-picking allows the alignment of the onsets of the phases rather than the peaks or troughs of greatest amplitude, as is the tendency of automatic cross correlation. We have automated cross-correlation estimates for a large portion of the data set, and in most cases, the difference is less than 0.5 s. We do not use measurements from diffracted raypaths or epicentral distances less than 55° (in the case of *PcP-P*) to improve accuracy in forward modeling. The benefit of using differential travel times versus absolute travel times is that biases due to source mislocation and near-surface structure are greatly reduced due to the nearness of the two raypaths in the crust and upper mantle (Figure 1). To further reduce mantle effects, we experiment with correcting for mantle structure using current mantle models, as is further discussed in the next section. After assessing the gaps in the lowermost mantle sampling, we in part fill them by hand-picking an additional 463 *PKPab-df*, 224 *PKPbc-df*, and 281 *PcP-P* differential travel times. We use data recorded between the years 1965 and 2010 inclusive, mostly coming from large ($M_b$ >6) events with depths greater than 35 km. See Figure 2 for a demonstration of the improvement in raypath coverage. This 34% increase in travel time measurements greatly improves coverage in many areas, most notably Africa, the southwest Indian Ocean, the north Atlantic, and the north Australia/Indonesia.

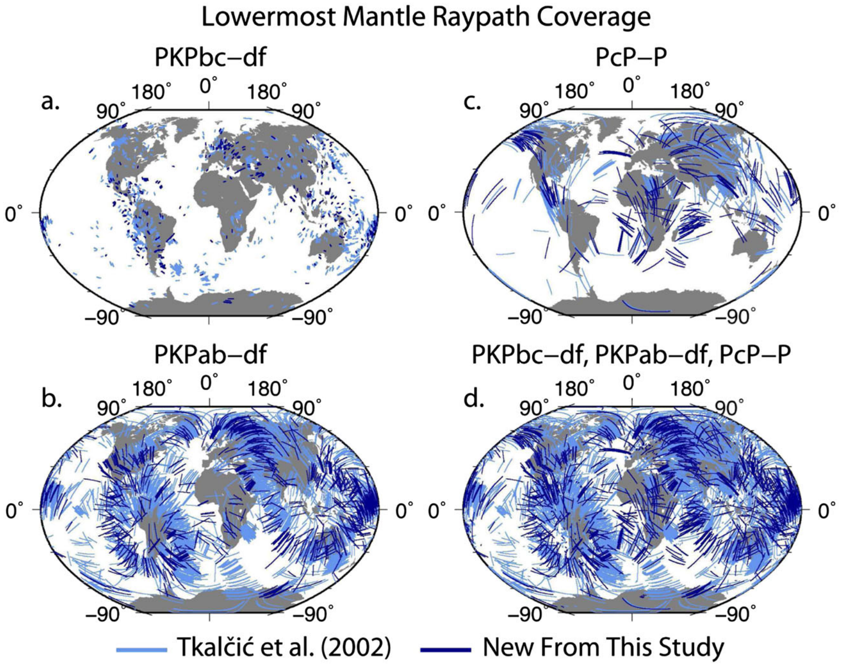


**Figure 2.** (a) *PKPbc* raypaths through the lowermost 300 km of the mantle associated with the *PKPbc-df* data set. (b) *PKPab* raypaths associated with the *PKPab-df* data set. (c) *PcP* raypaths associated with the *PcP-P* data set. (d) Raypaths associated with all three data sets. The raypaths from the data sets of *Tkalčić et al.* [2002] and *Tkalčić* [2010] are shown in light blue, and those of this study are shown in dark blue.

### 2.1. *PKPab-df* Data Set

[12] The differential *PKPab-df* travel time measurements are in part taken from the hand-picked data set of *Tkalčić et al.* [2002]. We have expanded this collection to improve spatial coverage (Figures 2b and 3b), although our measurement techniques remain much the same. First, we relocate earthquakes according to *Engdahl et al.* [1998], as a 10 km error in either source or receiver location will result in a differential travel time measurement error of between approximately 0.2 and 0.4 s. Then after a Hilbert transform is applied to the unfiltered *PKPab* waveform, the onsets of the seismic phase are aligned manually rather than by an automated cross-correlation technique (Figure 4a). Once precisely aligned, the travel time residual is calculated relative to the global reference model ak135 [*Kennett et al.*, 1995] and corrections are made to account for the Earth's ellipticity [*Kennett and Gudmundsson*, 1996]. Finally, corrected residuals range from –4 to +6 s (Figure 5b), although most of this range results from a small cluster of very anomalous paths of $\xi$ (angle between *PKPdf* leg and rotation axis of the Earth) between 20 and 30° originating in the South Sandwich Islands (SSI). The majority of residuals range between ±2 s when the SSI data is excluded (see section 3.6 for further details).

[13] Most of the data set (1871 individually measured differential travel times) comes from large (*Mb* >5.8) earthquakes of depths greater than 30 km recorded on the vertical component of primarily digital broadband instruments between 145 and 175° from the source. Although we use unfiltered data for measurement, *PKP* waves are in general best accentuated when filtered between 0.5 and 3.0 Hz. Data from large, deep earthquakes generally yield seismograms of good signal to noise ratios. Nonetheless, we only reserve measurements for the final inversion if the uncertainty associated with the temporal location of the phase onset is less than 0.5 s upon visual inspection. The *PKPdf* phase samples the inner core while the *PKPab* bottoms less deep and only samples the outer core (Figure 1). The two phases are very similar in the crust and upper mantle (less than ~7 km separation at 30 km depth), which helps to remove most of the unwanted effects of heterogeneity near the source or receiver. The differences in raypath geometry of the two phases is, however, very sensitive to heterogeneity in the lowermost mantle, as the path separation ranges from 12 to 47°, or 720 to 2820 km, at the core mantle boundary. Given a dominant frequency of 1.0 Hz and an epicentral distance of 150°, the Fresnel zone of a typical *PKPab* wave will be around 300 km wide in the lowermost mantle.

### 2.2. *PKPbc-df* Data Set

[14] *PKPbc-df* differential travel times are ideal for probing short-scale heterogeneity, as the separation of the two phases is less than 400 km at the core mantle boundary (Figure 1). Given a dominant frequency of 1.0 Hz and an epicentral distance of 150°, the width of the Fresnel zone of a typical *PKPbc* wave will be around 300 km in the lowermost mantle. The close proximity of the raypaths also means that mantle corrections are less critical than for the other data sets. We again focus on large, deep earthquakes, but this time from a epicentral distance range of 145 to 155°. Because of this more limited distance range, the *PKPbc-df* data set is smaller than the *PKPab-df* data set (1292 versus 1871). Less measurements and shorter raypaths in the lowermost

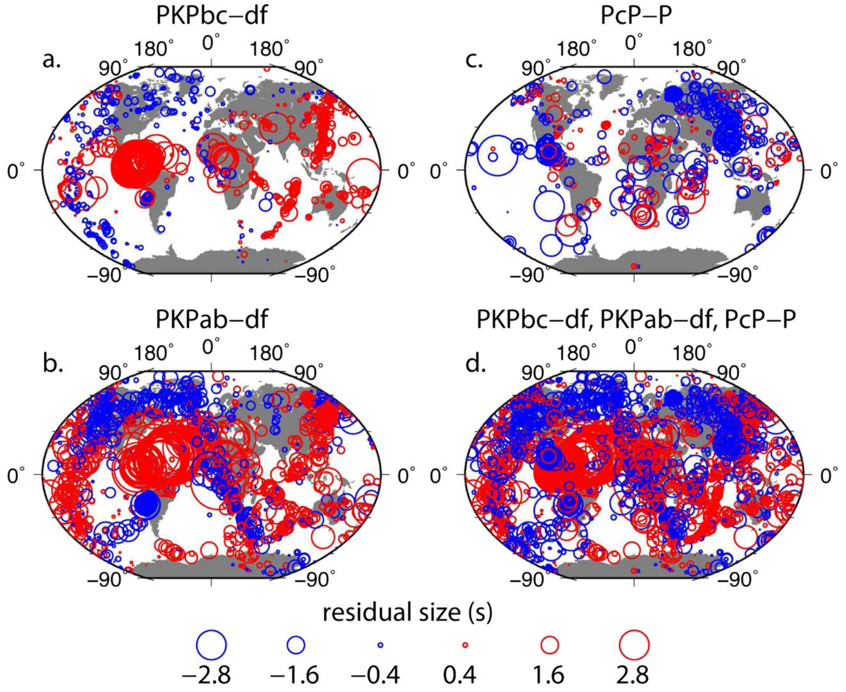


**Figure 3.** (a) *PKPbc-df* and (b) *PKPab-df* differential travel time residuals plotted at the surface projection of the bottoming point of the *PKPdf* raypath. (c) *PcP-P* differential travel time residuals plotted at the surface projection of the bottoming point of the *PcP* raypath. (d) Differential travel time residuals of all three data sets. Positive residuals are shown by red circles; negative residuals are shown by blue circles.

mantle makes the *PKPbc-df* sampling poor in comparison to the *PKPab-df* data set (Figures 2 and 3).

[15] Measurements are performed in much the same manner as with *PKPab-df* but without the application of the Hilbert transform. Relative phase offset times are carefully measured relative to ak135 after event relocation and ellipticity corrections are applied. The resulting residuals range from –1.5 to +5 s; however, like in the *PKPab-df* case, this larger range results from a small cluster of anomalous paths from the SSI; the majority of the residuals range between ±1 s (Figure 5a). When plotted with respect to $\xi$, it becomes clear that the spread in the *PKPbc-df* measurements is significantly less than that of the *PKPab-df* measurements.

### 2.3. *PcP-P* Data Set

[16] The inclusion of *PcP-P* differential travel times not only enhances spatial coverage (Figures 2 and 3) but also helps to resolve ambiguity about the location of heterogeneity on source or receiver sides. Moreover, *PcP* and *P* wave are not core phases; hence, they are useful for isolating inner core structure from the lowermost mantle when in combination with other data sets. Our data set of 680 differential travel times is derived from an epicentral distance range of 55 to 70°. This conservative limit helps to insure close raypath proximity within the mantle and crust, reducing unwanted effects of heterogeneity outside the lowermost mantle. At 300 km depth, raypath separation does not exceed 0.93°, or 98 km, for an earthquake with a surface depth of focus. Again, the same measurement procedures as described earlier are applied. Although we prefer to process the data without filtering, raw *PcP* arrivals are often buried in microseismic noise. To extract a measurable signal, some data required filtering between 1.0 and 3.0 Hz (Figure 4). At a dominant frequency of 1.0 Hz and an epicentral distance of 55°, the Fresnel zone of a typical *PcP* wave will be around 300 km in the lowermost mantle. The residuals range between ±3 s, exhibit no obvious correlation with epicentral distance, and have no anomalous clusters like in the *PKP* data sets.

## 3. Inversion Method

[17] In this paper, we present a new approach to global tomography. We use a Bayesian inversion technique [e.g., *Box and Tiao*, 1973; *Tarantola and Valette*, 1982; *Bernando and Smith*, 1994] to invert for lowermost mantle *P* wave velocity structure, as it is ideal for reliably representing model complexity and perturbation amplitude. This approach has been applied to a variety of seismological problems, including tomography [*Zollo et al.*, 2002; *Bodin and Sambridge*, 2009; *Khan et al.*, 2011; *Bodin et al.*, 2012a; *Mosca et al.*, 2012], receiver function inversion [*Piana Agostinetti and Malinverno*, 2010; *Bodin et al.*, 2012b], and seismic source parameter estimation [*Myers et al.*, 2007; *Monelli and Mai*, 2008]. For a more complete description of Bayesian analysis, refer to *Bodin and Sambridge* [2009], *Bodin et al.* [2012b, 2012a], and *Sambridge et al.* [2013].

### 3.1. Transdimensional Bayes

[18] The method relies on Bayes' theorem [*Bayes*, 1763], which provides the solution to the general inverse problem $\mathbf{d} = g(\mathbf{m})$, where $\mathbf{d}$ is the data vector and g maps

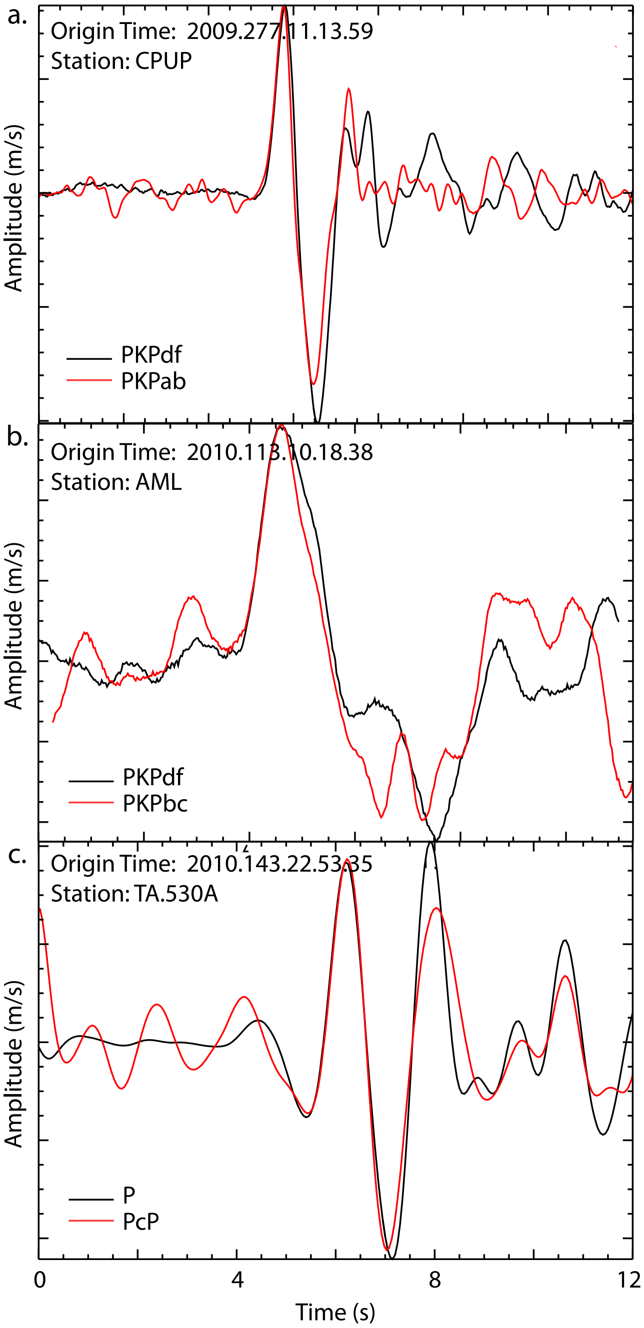


**Figure 4.** (a) Plot of the final alignment of *PKPdf* and *PKPab* phases used to measure the differential travel time *PKPab-df*. A Hilbert transform and polarity reversal have been applied to the *PKPab* phase. (b) Plot of the final alignment of *PKPdf* and *PKPbc* phases used to measure the differential travel time *PKPbc-df*. (c) Plot of the final alignment of *P* and *PcP* phases used to measure the differential travel time *PcP-P*.

the model parameter vector **m** into data, and can be stated mathematically as

$$p(\mathbf{m} \mid \mathbf{d}_{\text{obs}}) \propto p(\mathbf{d}_{\text{obs}} \mid \mathbf{m}) p(\mathbf{m}), \tag{1}$$

where $p(\mathbf{m} \mid \mathbf{d}_{\text{obs}})$ is the posterior probability distribution function of the unknown model parameters **m** solving the inverse problem given the data $\mathbf{d}_{\text{obs}}$. Solutions are described by a joint probability distribution function over all model parameters rather than a single optimal solution. Models that fit the data better will have a higher posterior probability, although models that offer a worse fit to the data will still be included in the posterior distribution. The term $p(\mathbf{d}_{\text{obs}} \mid \mathbf{m})$ is the likelihood function, which yields the probability of observing data $\mathbf{d}_{\text{obs}}$ given model **m**. This provides a measure of how well a particular model fits the data and depends on the misfit between the observed data and the synthetic travel times computed for a given model and also on the estimated variance of the data noise. For the case of Gaussian noise statistics, the likelihood function $p(\mathbf{d}_{\text{obs}} \mid \mathbf{m})$ can be expressed as

$$p(\mathbf{d} \mid \mathbf{m}) = \frac{1}{\sqrt{(2\pi)^N \mid \mathbf{C}_e \mid}} \times \exp \\ -\frac{1}{2}\{[g(\mathbf{m}) - \mathbf{d}]^T \mathbf{C}_e^{-1} [g(\mathbf{m}) - \mathbf{d}]\}. \tag{2}$$

[19] In our case, **d** is the vector of $N$ observed differential travel times, $g(\mathbf{m})$ is the vector of $N$ predicted differential travel times given the current lowermost mantle model **m**. The method is transdimensional in the sense that model **m** can have a variable number of defining parameters. In all inverse problems, the number of model parameters, and hence model complexity, is dependent on the level of data noise. Since the number of the model parameters can change, the required number of cells to fit the data can be expressed as a posterior probability distribution function. The term $\mathbf{C}_e$ is the data noise covariance matrix, and $|\mathbf{C}_e|$ is its determinant. If the estimated error is low, the required number of model parameters to fit the data will be relatively large, whereas high uncertainty will yield a simpler model of fewer parameters. By allowing a flexible number of unknowns, or a model of variable dimension, we appropriately leave model complexity to be determined by the data [*Bodin et al.*, 2012a]. Although it may seem that an overly complicated model would result in order to minimize data variance, the parsimonious nature of the Bayesian approach promotes the preservation of the simplest models that fit the data, thus preventing unjustified model complexity [*Malinverno*, 2002].

[20] Going back to equation (1), our prior information about model **m** is represented by the a priori probability distribution $p(\mathbf{m})$. This is a mathematical representation of what we think we know about the model prior to performing the inversion. In order to minimally affect the inversion outcome, we employ a uniform distribution between –5% and +5% perturbation from the global reference model ak135 [*Kennett et al.*, 1995] as part of our prior information. So $p(v_i) = 1/(v_{\max} - v_{\min})$ for values of $v_i$ between $v_{\min}$ and $v_{\max}$, and $p(v) = 0$ for all velocity values outside that range. Because the shape of any distribution multiplied by a uniform distribution will be unaffected, the prior information acts only as a lower and upper bound to the allowed velocity perturbations. Similarly, we impose prior assumptions about the number of Voronoi cells $n$ needed to represent the data. We allow between 4 and 5000 cells, so $p(n) = 1/(n_{\max} - n_{\min})$ for all values of $n$ between $n_{\min}$ and $n_{\max}$, and $p(n) = 0$ otherwise.

### 3.2. Hierarchical Bayes

[21] The level of data noise is an important unknown in modeling. In our case, "noise" is everything that contributes

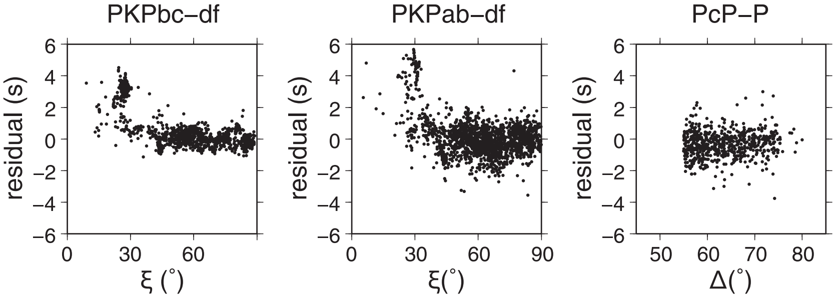


**Figure 5.** (left) The *PKPbc-df*, (middle) *PKPab-df*, and (right) *PcP-P* differential travel time data sets used in this study. Travel time residuals are plotted with respect to ξ(°) (left, middle) or epicentral distance *Δ* (°)(right).

to the difference between the observed and predicted differential travel time residuals [*Scales and Snieder*, 1998]; this includes either theory errors, which affect the predicted differential travel times, or measurement errors, which affect the observed differential travel times. An example of a theory error is the mapping of core-sensitive *PKP* phase differential travel times exclusively to the lowermost mantle. In a joint inversion that also includes *PcP-P* differential travel times, which do not sample the core, the inner core effects on core-sensitive phases will not be coherent and will hence be accounted for as "data noise" [*Bodin et al.*, 2012a]. Another source for theory error is assuming a uniform thickness of 300 km for the lowermost mantle. Additionally, raypath geometries were not iteratively updated and ray theory instead of finite frequency theory was utilized due to the prohibitive computational cost involved. For a typical *PcP-P* (epicentral distance of 55°), a 1% homogeneous increase in *P* wave velocity in the lowermost 300 km of the mantle would create a ∼0.12 s error if the raypath is not perturbed accordingly. This error may play a part in the differences between the *PcP* and *PKP* data sets (see section 4.2). For a *PKPab-df* raypath geometry with an epicentral distance range of 150°, the error is less, at around 0.02 s. For a *PKPbc-df* raypath geometry of the same epicentral distance, the error is even smaller, at around 0.001 s. The finite frequency effects are in part diminished by the fact that the application of various band-pass filters suggests the dominant period of the data sets is approximately 1 s. Additionally, we use differential travel times rather than absolute times. Finally, the reference model is derived from data in which the finite frequency approximation was used. A more significant source of error would likely stem from the use of catalog data, such as the ISC data sets employed by many previous workers, due to triplication effects, forgoing the Hilbert transform of the *PKPab* phase, and not accounting for the source time function (in the case of absolute arrival times). An example of a source of measurement error is event mislocation, even though we use the event catalog presented by *Engdahl et al.* [1998], which uses ISC relocation algorithms. These approximations contribute to the misfit and are also accounted for as data noise. It is not possible, however, to estimate the ratio of theory errors to measurement errors. Similarly, it is not possible to estimate the relative contribution of each error source, whether it be theory- or measurement-based. Noise essentially represents the difference between the true residuals and the residuals able to be explained by our model.

[22] We represent the data noise with the matrix $\mathbf{C}_e$, which for $N$ data is a symmetric $N \times N$ matrix. We assume invariant, uncorrelated Gaussian random noise that is dependent only on the data type (i.e., whether it is a *PKPab-df*, *PKPbc-df*, or *PcP-P* differential travel time measurement). Therefore, the noise correlation matrix $\mathbf{C}_e$ in equation (2) is diagonal and can be expressed as follows [*Gouveia and Scales*, 1998]:

$$\mathbf{C}_e = \begin{vmatrix} \sigma_i^2 & 0 & \dots & 0 \\ 0 & \sigma_i^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_i^2 \end{vmatrix}, \tag{3}$$

where $\sigma_i = \sigma_1$ for *PKPbc-df* data, $\sigma_i = \sigma_2$ for *PKPab-df* data, and $\sigma_i = \sigma_3$ for *PcP-P* data. Each data type is represented by a single hyperparameter $\sigma$ that is representative of the standard deviation of the data errors (measured in seconds). We invert for three different parameters to characterize the data uncertainty of the *PKPab-df*, *PKPbc-df*, and *PcP-P* data sets, as these data sets would not be expected to have the same level of noise; the raypaths, signal-to-noise ratios of the seismograms, and sensitivity to lowermost mantle structure and core mantle boundary topography are quite different. Although we predetermine the form of the matrix $\mathbf{C}_e$, the actual entry values are left as unknowns in the problem [*Malinverno and Briggs*, 2004; *Malinverno and Parker*, 2006]. We do, however, impose the prior assumption that the data noise values will range between 0.01 s and 2.0 s for each data set. The maximum of the posterior probability distribution functions of the $\sigma$ values is used as the estimated noise level for each data set. The relative uncertainty of the data sets determines their relative contribution to the inversion process [*Bodin et al.*, 2012a, 2012b], thereby removing any need for ad hoc weighting factors.

### 3.3. Model Parameterization

[23] The lowermost mantle is modeled as an ensemble of Voronoi cells of variable size and shape (Figure 6). The positions of these nonoverlapping polygons are defined by their Voronoi nuclei. Each point within a cell is closer to that cell's nucleus than to any other nucleus. Therefore, the cell walls are perpendicular bisectors of adjacent nuclei. For each cell, we have three unknowns: its position (defined by its latitude and longitude) and its velocity. The total number of these Voronoi cells is an unknown as well. A given model iteration is also parameterized by the $\sigma$ values within the data noise correlation matrix $\mathbf{C}_e$.

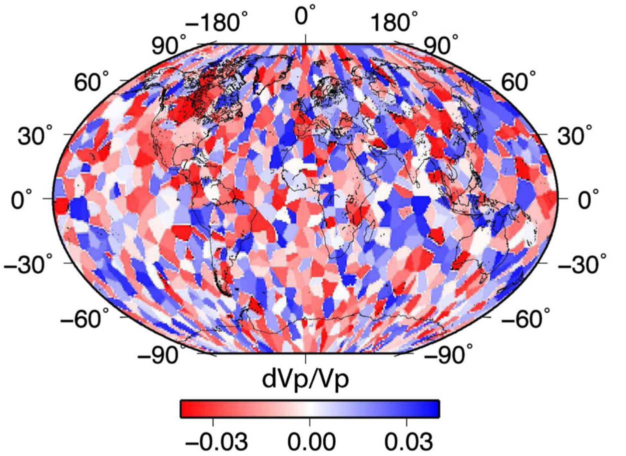


**Figure 6.** Example of a Voronoi *P* wave velocity model realization in the lowermost 300 km of the mantle, which consists of an irregular, interlocking set of polygons. Each Voronoi cell encompasses all the points within the 2-D space that are closer to its center than to any other Voronoi cell center.

[24] Previous global tomographic studies invert for velocity structure using models based on blocks [e.g., *Tkalčić et al.*, 2002; *Vasco et al.*, 2006; *Houser et al.*, 2008; *Soldati et al.*, 2012] or spherical harmonic expansions [e.g., *Su et al.*, 1994; *Antolik et al.*, 2003; *Ritsema et al.*, 2011]. The problem with block parameterization is that arbitrary and fictitious velocity boundaries arise between the blocks. Ad hoc error estimation and its consequent requirement for arbitrary smoothing and damping regularization are also issues, as the model is often overly complex in regions of poor sampling or high data noise. On the other hand, a spherical harmonic representation limits the resolution of fine-scale structure as might be expected of subducting slabs and mantle plumes. With a Bayesian approach, however, model parameterization is flexible and almost fully data-dependent. Large, small, smooth, and discontinuous structures can be imaged with appropriate levels of resolution.

[25] Since the Voronoi cells can occupy an infinite range of configurations, we can take the average of a large number of postconvergence realizations to generate a continuous final model. We do this by taking the spatial mean of the distribution of velocity values at each point (we sample at every 1° in latitude and longitude) across the 2-D region of interest. The grid of points for which we calculate an average velocity can be as fine as desired to enable visual evaluation of the model. This average model is both more complex and smoother than any one individual model. It is not, however, the most probable model or best fit model. Smoothing or damping procedures are unnecessary; for proper model resolution, complexity, and smoothness are innate to the averaged solution. We apply a similar procedure to approximate a map of the model error by calculating the standard deviation of the velocity distribution at each pixel. Areas of greater uncertainty, such as poorly sampled regions or the border between a sharp change in velocity, will undergo more changes in parameterization and will therefore have a higher velocity standard deviation.

[26] We model the core mantle boundary layer as being the last 300 km of the mantle in part based on the significant increase in the root-mean-square (RMS) of the velocity perturbations observed in this region [e.g., *Mégnin and Romanowicz*, 2000; *Antolik et al.*, 2003; *Simmons et al.*, 2009; *Zhao*, 2009]. We also performed a series of tomographic inversions with a layer thickness of 250, 300, and 350 km. The 300 km thickness scenario yielded the lowest noise estimates for the three data sets, indicating a better fit to the data. This thickness is in agreement with the statistical analysis of *Garcia et al.* [2009] who estimated the average thickness of the lowermost layer to be 350 ±50 km. If we model the layer as being too thin, valuable information from *P* raypaths with bottoming depths above the 300 km level may be lost. On the other hand, an erroneously thick lowermost layer can potentially result in underestimated heterogeneity. An extended Bayesian scheme where layer thickness is variable would quantify the trade-off between layer thickness and the amplitude of perturbations, and is the subject of future research. Nonetheless, for this study, the perceived noise in the data sets for the most part includes effects of this approximation.

[27] As discussed in section 3.1, velocity perturbations (relative to an average ak135 velocity of 13.61 km/s in the lowermost 300 km of the mantle) are allowed to range between ±5%. This conservative limit is necessary to properly assess the noise in the data. When the prior is too unrestrictive, the true variations in the data are misinterpreted as noise and a nearly laterally homogeneous model results.

[28] Because the Voronoi cell generation is based on a rectangular grid, the global model is flattened from 3-D to 2-D. To insure that the model agrees with itself along the edges when wrapped around to form a sphere once again, the data set is duplicated and shifted by 360°. The model space is made to range from –360 to +360° and when cropped back to –180 to +180° and wrapped around, the "seam" matches perfectly. Nonetheless, the models will remain somewhat "smeared" at the poles since the space is defined on a flat projection. Work to greatly improve this aspect of the model parameterization by implementing a spherical Voronoi cell grid is in progress.

### 3.4. Misfit Evaluation

[29] The Markov chain algorithm requires calculation of the likelihood for every model iteration. The likelihood function $p(\mathbf{d}_{\mathrm{obs}} \mid \mathbf{m})$ describes quantitatively to what level the current model $\mathbf{m}$ can reproduce the observed data, and as in equation (2) is related to the least squares misfit of the predicted ($g(\mathbf{m})$) and observed ($\mathbf{d}_{\mathrm{obs}}$) data as follows:

$$p(\mathbf{d}_{\mathrm{obs}} \mid \mathbf{m}) \propto \exp\frac{-\boldsymbol{\phi}(\mathbf{m})}{2}, \tag{4}$$

where

$$\boldsymbol{\phi}(\mathbf{m}) = \| \frac{g(\mathbf{m}) - \mathbf{d}_{\mathrm{obs}_i}}{\boldsymbol{\sigma}_i} \|^2. \tag{5}$$

[30] In our case, $\mathbf{d}_{\mathrm{obs}_i}$ refers to the observed differential travel time residuals ($\Delta\delta\mathbf{t}$), which are equal to the difference between the observed differential travel time ($\delta\mathbf{t}_{\mathrm{obs}}$) between the first and second phase and the predicted differential travel time between the two phases according to ak135 ($\delta\mathbf{t}_{\mathrm{ak135}}$) [*Kennett et al.*, 1995]. Here $\mathbf{d}_{\mathrm{obs}_i} = \mathbf{d}_{\mathrm{obs}_1}$ and $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_1$

for *PKPbc-df* data, $\mathbf{d}_{\text{obs}_i} = \mathbf{d}_{\text{obs}_2}$ and $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_2$ for *PKPab-df* data, and $\mathbf{d}_{\text{obs}_i} = \mathbf{d}_{\text{obs}_3}$ and $\boldsymbol{\sigma}_i = \boldsymbol{\sigma}_3$ for *PcP-P* data. A typical value for $\|g(\mathbf{m}) - d_{\text{obs}}\|$ prior to inversion is ~0.5 s and postinversion is ~0.2 s, so the difference between the observed and predicted differential travel times generally decreases by more than half. If $\delta\mathbf{t}_m$ is the predicted differential travel time between the first and second phase given model **m**,

$$\boldsymbol{\phi}(\mathbf{m}) = \{\|(\delta\mathbf{t}_m - \delta\mathbf{t}_{\text{ak135}}) - (\Delta\delta\mathbf{t}_{\text{obs}})\|/\boldsymbol{\sigma}_d\}^2. \quad (6)$$

[31] Using the simple relationship $\mathbf{d} = \mathbf{vt}$ and the fact that we use fixed raypaths according to ak135,

$$\boldsymbol{\phi}(\mathbf{m}) = \{\| \frac{\delta\mathbf{d}(\mathbf{m})}{v_{\text{ak135}}[1 + \mathbf{p}(\mathbf{m})]} - \frac{\delta\mathbf{d}(\mathbf{m})}{v_{\text{ak135}}} - (\Delta\delta\mathbf{t}_{\text{obs}})\|/\boldsymbol{\sigma}_d\}^2, \quad (7)$$

where $\mathbf{p}(\mathbf{m})$ is equal to the velocity perturbation relative to the average velocity of the reference model ak135 in the lowermost 300 km of the mantle (13.61 km/s). The vector $\delta\mathbf{d}(\mathbf{m})$ is the difference in raypath distance between the first and second phase across each Voronoi cell of model **m**. We integrate $\boldsymbol{\phi}(\mathbf{m})$ over all Voronoi cells. Each path length difference is calculated by sampling the raypaths at increments on the order of 10 km and determining to which cell the midpoint of the ray segment belongs. Segment lengths belonging to the same cell are then added together prior to integrating along the raypath. As explained in more detail in section 3.2, $\sigma_d$ is dependent on the phase type and is therefore represented by one of three unknowns, $\sigma_1$ for *PKPbc-df*, $\sigma_2$ for *PKPab-df*, and $\sigma_3$ for *PcP-P*.

### 3.5. Sampling of the Model Space

[32] The Bayesian class of inversion is based on ensemble inference, meaning that a large number of models are generated, each with varying parameters, according to the generalized version of Markov chain Monte Carlo sampling called Reversible Jump MCMC [*Green*, 1995, 2003], which is based on the Metropolis-Hasting algorithm of *Metropolis et al.* [1953] and *Hastings* [1970]. Prior to producing the first model, which is randomly created by selecting values from the prior distributions, raypaths are determined according to ak135 [*Kennett et al.*, 1995] using TauP [*Crotwell et al.*, 1999]. At each step in the inversion, a proposed model is created from a random perturbation of the current model [*Mosegaard and Tarantola*, 1995]. The perturbation can result from one of five possible changes. First, a new set of $\sigma$ values can be selected to represent the data noise; this change is applied at every iteration. At every odd iteration, the second type of change is performed, which is that the velocity value of one randomly selected Voronoi cell is changed. This new velocity value $v'$ is chosen such that

$$v' = v + u\lambda, \quad (8)$$

where $v$ is the original velocity of the cell, $u$ is a random number selected from a normal distribution between –1 and 1, and $\lambda$ is the standard deviation of the proposal. The $\lambda$ value is determined by the user and will affect the rate of convergence, but not the end result of the inversion. At every even iteration, one of the three remaining types of change is chosen at random to occur: (1) the position of one randomly selected Voronoi cell is perturbed according to a 2-D Gaussian proposal probability density centered at the current position, (2) a new Voronoi cell is added at a randomly selected location with a velocity chosen from a Gaussian proposal probability centered on the current velocity value where the "birth" takes place (same form as equation (8)), or (3) a randomly selected Voronoi cell is deleted.

[33] After the model is updated, the differential travel times are recomputed and compared with the observed differential travel times. The new model is then either accepted or rejected according to certain acceptance criteria [*Bodin and Sambridge*, 2009]. The probability of accepting a proposed model $\mathbf{m}'$ given the current model $\mathbf{m}$ is $\alpha(\mathbf{m}' \mid \mathbf{m})$, which *Green* [1995, 2003] equates to the following in order to ensure convergence of the models toward the posterior distribution:

$$\alpha(\mathbf{m}' \mid \mathbf{m}) = \min\{1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \times \frac{p(\mathbf{d}_{\text{obs}} \mid \mathbf{m}')}{p(\mathbf{d}_{\text{obs}} \mid \mathbf{m})} \times \frac{q(\mathbf{m} \mid \mathbf{m}')}{q(\mathbf{m}' \mid \mathbf{m})} \times |\,\mathbf{J}\,|\}, \quad (9)$$

where $p(\mathbf{m}')$ and $p(\mathbf{m})$ are the prior for the proposed and current model, $p(\mathbf{d} \mid \mathbf{m}')$ and $p(\mathbf{d} \mid \mathbf{m})$ are the likelihood functions (equation (2)) for the proposed and current model, and $q(\mathbf{m} \mid \mathbf{m}')$ and $q(\mathbf{m}' \mid \mathbf{m})$ are the forward and reverse proposal functions. The term $|\,\mathbf{J}\,|$ is the Jacobian, which enables a transformation between models of different dimension or parameterization. In the case of changes that do not affect model dimension (velocity change or cell move), $q(\mathbf{m}' \mid \mathbf{m})$ is equal to $q(\mathbf{m} \mid \mathbf{m}')$, meaning that it is equally probable that we generate a proposed model $\mathbf{m}'$ if we start with model **m** as it is that we generate model **m** if we start with model $\mathbf{m}'$. In the case of a birth or death of a cell, however, the model dimension changes and the proposal functions encourage changes to the model. So in the event of a cell birth, the proposed cell's velocity is encouraged to deviate from the velocity of the cell that was in that position in the original model. Likewise, a cell death is favored if the velocity of the deleted cell varies considerably from the velocity of the cell that will replace its position upon removal. These trends are of course subdued once combined with the likelihood and prior ratios of equation (9). One end result is that uneven raypath sampling is reflected in the relative size of Voronoi cells of an individual model. Larger cells will persist or form (via cell death) in regions of sparser sampling while much smaller cells can be justified and created (via cell birth) in regions of densest sampling [*Bodin and Sambridge*, 2009].

[34] To determine if the proposed model $\mathbf{m}'$ is accepted or rejected, we generate a uniformly distributed random number $u$ between zero and one. If $u(\mathbf{m}')$ is greater than $\alpha(\mathbf{m}')$, then the model $\mathbf{m}'$ is rejected and model **m** is retained for the next step in the chain. Otherwise, model $\mathbf{m}'$ is accepted and forms the new basis to which the next perturbation will occur. According to equation (9), if model $\mathbf{m}'$ fits the data better than model **m**, it will always be accepted. Models that fit the data almost as well as the previous model will be accepted most of the time, and models that fit very poorly in comparison to the previous model will be accepted only very rarely. In this manner, the chain's path will be guided toward parameter space of high target density; the sampling distribution will mimic the target posterior distribution.

[35] The first portion of unstable, "burn-in" iterations (usually around 2 million for the inversions considered here) is discarded, while the remaining, postconvergence portion is deemed representative of the posterior distribution [*Green*, 1995, 2003]. We define postconvergence as the point at which average data misfit, average number of cells, and all

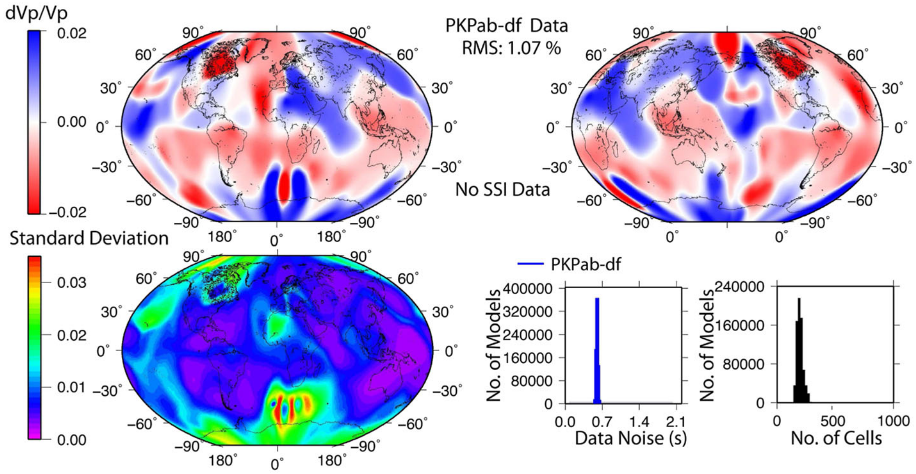


**Figure 7.** (top) Model of *P* wave velocity variations in the lowermost mantle using the *PKPab-df* differential travel time data set. Mantle corrections are applied according to *Della Mora et al.* [2011]. Perturbations are shown relative to the ak135 average velocity of the layer (13.61 km/s). (bottom left) Map of the standard deviation of the velocity perturbation. (bottom right) Posterior probability distributions of the *PKPab-df* data noise and number of cells.

velocity parameters cease to fluctuate beyond that of a normal white noise process and the solution map stabilizes. Once we acquire a large ensemble of independent models, we can extract useful properties, such as the average, median, or best model [*Smith*, 1991]. To insure model independence, we "thin" this second part of the chain by only retaining every 50th or 100th model for the final ensemble from which we calculate a pointwise spatial average model, so usually around 800,000 postconvergence models are generated.

### 3.6. Measurement Corrections

[36] *Creager* [1999] suggested that the inner core is strongly cylindrically anisotropic throughout most of the western hemisphere with the fast direction aligned with or near the spin axis. Although a simple model of cylindrical inner core anisotropy with the fast axis aligned with the Earth's rotation axis has been disputed [e.g., *Tkalčić*, 2010], here we test whether such a scenario may be affecting our lowermost mantle model inversion results by inverting the *PKPab-df*, *PKPbc-df*, and *PcP-P* differential travel times for two cases: (1) all raypaths (Figure 9) and (2) all raypaths excepting those recorded from events in the South Sandwich Islands (SSI), which are strongly polar (Figure 8). The total number of *PKPab-df* data points decreases from 1871 to 1786, and the number of *PKPbc-df* data points decreases from 1291 to 1042. The exclusion of these very polar paths (angle $\xi < 30°$ ) does not significantly affect the pattern of the velocity variations of *P* wave velocity model of the lowermost mantle; however, the RMS of the velocity perturbations does change significantly. By removing SSI data, the RMS reduces from 1.00% to 0.87%. In addition, when excluding the SSI data, the noise estimates for each data set decrease by approximately 5%, indicating a higher quality data set is achieved by removing these anomalous paths. When removing the relatively few polar paths not associated with the SSI, similar effects on data noise estimates and the strength of perturbations are not seen. Therefore, our preferred average model (Figure 8) includes polar paths except for those coming from the SSI, which may be influenced by strong mantle heterogeneity related to slabs or fragments [*Tkalčić*, 2010]. Consequently, we also exclude SSI data from all inversions using the *PKPbc-df* and *PKPab-df* data sets (Figures 7, 8, 10, and 11).

[37] Despite our preventative measures of removing SSI data, there will still be some mapping of core structure to the lowermost mantle. We further mitigate this effect by including *PcP-P* measurements, which do not sample the core. These differential travel times help to remove the ambiguity between lowermost mantle and inner core contributions to core phase arrival times. We also use a hierarchical Bayesian model, which allows everything that cannot be explained by our model to be treated as data noise. If there are inner core effects, they will imply incoherency between different data types and will therefore be treated as "theory errors." The overall agreement between areas of adequate coverage by all three data sets helps to confirm that core structure can only have a minor effect on the differential travel times (Figures 10 and 12).

[38] Mantle effects are corrected for using the 3-D mantle model of *Della Mora et al.* [2011], which results from the inversion of direct *P* wave arrival times using a regularized least squares framework. This model is based on ~620,000 measurements from the ISC bulletin and is undoubtedly biased by inconsistent measurement quality, exclusive use of direct arrivals, smoothing and damping, and block parametrization. Nonetheless, it provides a reasonable approximation of upper mantle *P* wave velocity perturbations, which is where the power of heterogeneity is greatest and therefore most important for us to consider. *Della Mora et al.* [2011] importantly correct for crustal structure and relocate sources according to *Antolik et al.* [2003].

## 4. Results

### 4.1. Resolution Tests

[39] Each data set uniquely contributes to the retrieval of a lowermost mantle *P* wave velocity model due to the different

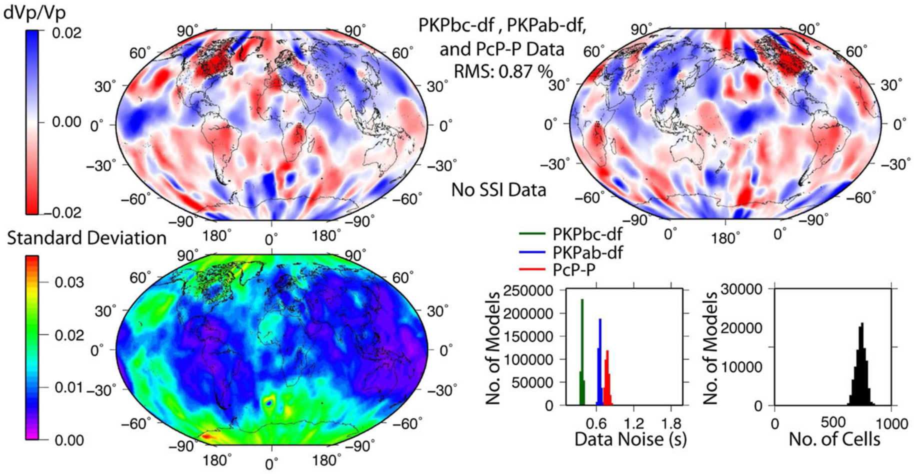


**Figure 8.** As for Figure 7 but for the *PKPbc-df*, *PKPab-df*, and *PcP-P* data sets combined. The South Sandwich Islands earthquakes are excluded from the data set prior to inversion.

sampling patterns of the raypaths. This is illustrated through two sets of resolution tests (Figures 13 and 14). For each test in the first set, synthetic data is calculated according to the raypath geometry of the actual data set for a known velocity model consisting of 20° squares alternating between –2% and +2% velocity variation perturbations (relative to ak135). Gaussian random noise with a standard deviation of 0.5 s is added to the synthetic data. The synthetic travel times are inverted using the same Bayesian technique as is used for the real data, and the retrieved model is compared to the actual model to assess the resolution potential given the data and method. Figure 13f shows the actual model used to produce the synthetic travel times. The retrieved model of the first test (Figure 13a) is a visual representation of the resolution potential of the *PKPab-df* data set alone. As expected, the best model recovery is obtained in regions of best sampling, in particular Africa, the Atlantic, and east Asia/Indonesia. The resolution test is performed for the *PKPbc-df* and *PcP-P* data sets as well (Figures 13b and 13c). From these results, it is clear that the *PKPbc-df* data alone cannot provide reliable information about the velocity structure for most of the globe; for this reason, we only use *PKPbc-df* measurements when in combination with other data sets.

[40] Finally, we join all three data sets. Now the *PcP-P* raypaths geometries help attribute the *PKPab-df* and *PKPbc-df* residuals to either the source or receiver side of the raypaths and much improvement is seen in the recovered model (Figure 13d). These tests show that in general, structure on the order of 20° in diameter is easily resolved given the method and data sampling. Areas where some smearing and damping of actual perturbation amplitude should be expected, however, include the mid-Pacific, the north Atlantic, the south Indian ocean, and the poles. The poorly resolved areas coincide with spatial gaps in raypath coverage. In these areas, there is either no sampling, or given the noise in the data, insufficient sampling to recover the true model.

[41] Figure 13e demonstrates the effect of neglecting mantle effects in the data corrections. Here we display the recovered model using all three data sets with additional noise added that is equal to the travel time corrections associated with the *Della Mora et al.* [2011] mantle model. These noise values range from –0.7 s to 1.2 s depending on the ray path geometry. From this figure it is evident that mantle corrections have very little effect on the final model, as the recovered models in Figures 13d and 13e

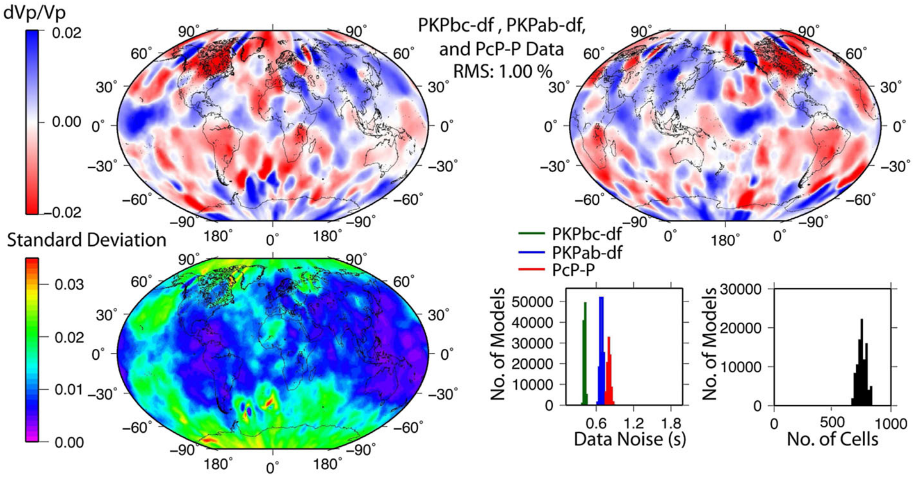


**Figure 9.** As for Figure 8 but with South Sandwich Islands data included.

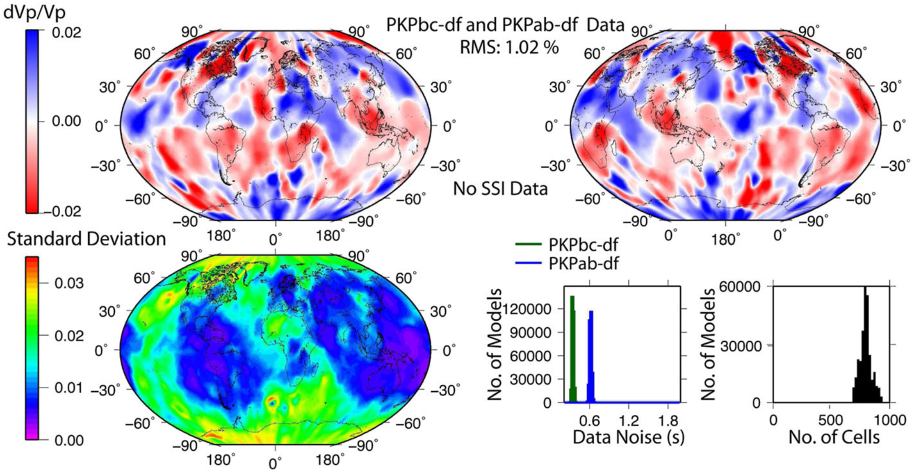


**Figure 10.** As for Figure 7 but for the *PKPab-df* and *PKPbc-df* data sets combined.

are nearly indistinguishable. When mantle corrections are considered (Figure 13d), however, the RMS of the velocity variations is slightly more accurately recovered (1.38% versus 1.32%).

[42] The second set of synthetic tests is designed to compare our Bayesian inversion approach to that of a more traditional linearized inversion method, namely one that uses a lower triangular-upper triangular decomposition algorithm [e.g., *Tkalčić et al.*, 2002]. The true model of this second test has structure of varying size, shape, and orientation (Figure 14) as to test the limits of the resolution potential of the complete data set (*PKPab-df*, *PKPbc-df*, and *PcP-P*) given each inversion method. The positive anomalies are 2% faster than the ak135 average velocity and the negative anomalies are 2% slower. This time, no noise is added to the synthetics. For the Bayesian inversion, we allow the perturbations to range between ±5%. In areas of good raypath coverage, the method is able to retrieve the pattern and strength of the velocity pattern with reasonable accuracy (Figure 14). The RMS of the velocity perturbations is 1.44% (actual RMS is 2%). The method fails only to retrieve the structures of the finest scale (small circles of 6° (∼360 km) diameter). The Bayesian approach is clearly able to retrieve sharp velocity contrasts and both small- and large-scale features in areas of adequate raypath sampling. The method is also shown to be fully capable of recovering curved discontinuities despite the fact that Voronoi cells have only straight edges; after averaging many differently positioned straight-sided polygons, curved lines can be retrieved.

[43] For the linear inversion, we use the same data set but now parameterize the model into 5°×5° squares. We apply a damping factor of $5{\times}10^5$, but no smoothing regularization, as was applied to the favored model of *Tkalčić et al.* [2002]. Blocks of inadequate sampling are omitted (colored black). Although we do not intend to provide an exhaustive comparison between methods, the linear inversion technique using this set of tuning parameters clearly struggles to recover the complex velocity pattern compared to the Bayesian inversion approach (Figure 14). Only in regions of best sampling (southwest Pacific, west Asia, and the Middle East, Central America) is the true model moderately recovered. Even then, the amplitude of the perturbations is significantly underestimated. The RMS of the linear inversion result is only 0.8%. Although the computational cost of a Bayesian inversion is in general 3 to 4 orders of magnitude greater than that of the linear approach of *Tkalčić et al.* [2002], the improved ability to recover perturbation amplitudes and both velocity discontinuities and gradations and the provision of uncertainty and data error estimates merit the additional time and resources. Nonetheless, the immense computational cost of ensemble inference approaches currently prohibits the Bayesian inversion for whole mantle structure.

### 4.2. Models Using Different Data Subsets

[44] An image of *P* wave velocity variations in the lowermost mantle resulting from the *PKPab-df* data set alone is shown in Figure 7. There were 3.5 million model iterations produced on each of 60 CPU processors after running for 200 h. The initial 2.5 million burn-in iterations were discarded, and the model shown is the result of averaging every 50th model of the subsequent iterations. The pattern of velocity perturbations is complex, with the average number of cells used to parameterize the model being ∼200. The uncertainty in the data is treated as a single hyperparameter, which after inversion peaks at ∼0.60 s. The RMS of the perturbations is 1.07%, although the maximum perturbation is 4.74%. This figure is several times larger than previous estimates [e.g., *Tkalčić et al.*, 2002; *Antolik et al.*, 2003; *Lei and Zhao*, 2006; *Houser et al.*, 2008; *Li et al.*, 2008; *Della Mora et al.*, 2011; *Soldati et al.*, 2012], which were driven by subjective choices for damping. Because our method does not require explicit damping of the final model and such strong velocity perturbations are required by the data, we are able to justify a large variance reduction. The final RMS of the residuals is 0.58 s, which is 59% less than the RMS of the residuals resulting from a homogeneous lowermost mantle layer with a velocity of 13.61 km/s.

[45] Next, we invert the *PcP-P* data set alone (Figure 12). The resulting average model is similar to the results of the *PKPab-df* inversion in that, in general, Asia, Central America, and the Middle East are fast, while the north Atlantic, southwest Pacific, and North America are in general slow. Areas of discrepancy include South America, Australia, and Africa, where *PcP-P* coverage is poor. The

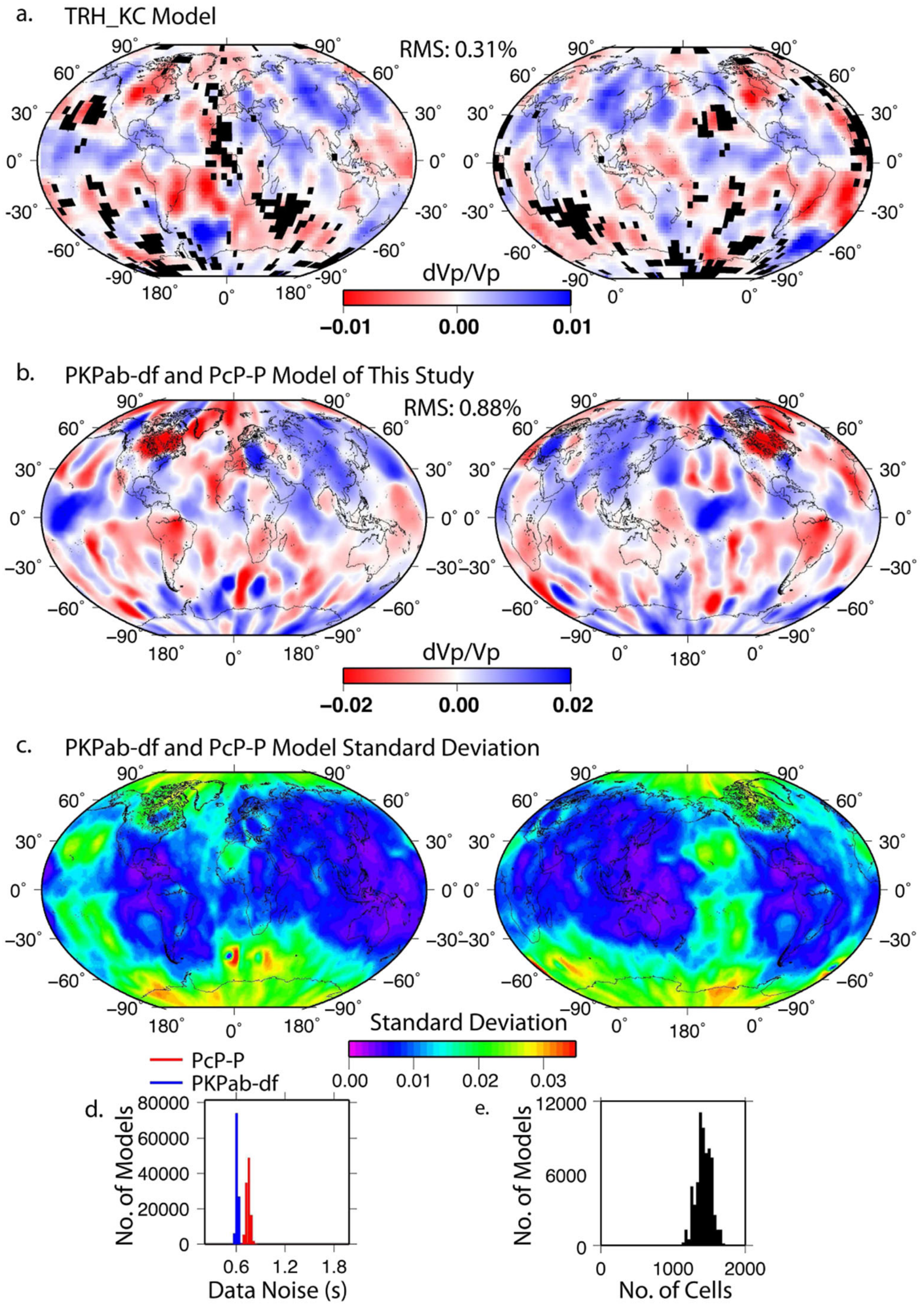


**Figure 11.** (a) The result (model TRH_KC) of the previous high-quality data set of *PKPab-df* and *PcP-P* differential travel times from *Tkalčić et al.* [2002] who performed a linear inversion for the *P* wave velocity structure of the lowermost mantle. Areas of insufficient data coverage are shown by black squares. (b) Bayesian inversion results for the *PKPab-df* and *PcP-P* data sets of this study. Note the difference between the color scale ranges and the root-mean-square values of the velocity perturbations between the two models. (c) A map of the standard deviations of the velocity distributions for each latitude/longitude pair of the model in Figure 11b. (d) Posterior probability distribution of the *PKPab-df* and *PcP-P* noise and (e) posterior probability distribution of the number of cells used to describe the model.

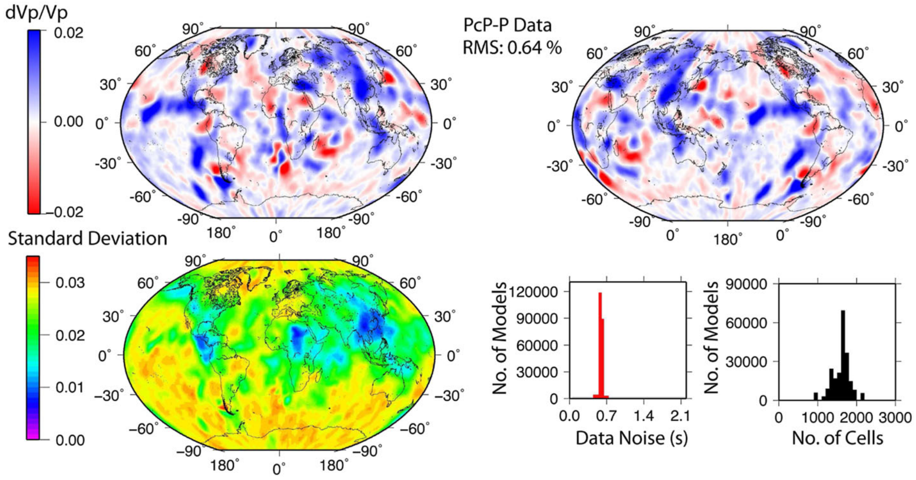


**Figure 12.** As for Figure 7 but for the *PcP-P* data set.

basic congruency between areas of adequate coverage, however, suggests that the inner and outer core structure have no more than a minor effect on the differential travel times. The estimated noise in the *PcP-P* data set (0.57 s) is comparable to that of the *PKPab-df* data set and much greater than that of the *PKPbc-df* data set. The noise is likely due to a combination of measurement errors, core mantle boundary topography effects on the *PcP* travel times, mantle structure, and forward modeling approximations. Given the poorer sampling and much greater sensitivity to mantle structure, the *PcP-P* data set is less coherent than the *PKP* data sets. Moreover, the *PcP* data only sample the lowermost mantle once, whereas *PKPbc* and *PKPab* rays sample it on both the source and receiver side. The resulting lack of flexibility in redistribution of the residual times to either the source or receiver side combined with the aforesaid poorer coherency means that many Voronoi cells are required to achieve the 55% decrease in residual variance. The hierarchical Bayes approach of this study accounts for this effect when jointly inverting different data types, as less weight is naturally given to the more inconsistent data sets.

[46] When the *PKPab-df* and *PKPbc-df* data sets are combined (Figure 10), the resolution improves relative to when *PKPab-df* is used alone, and the average number of cells used to parameterize the models is ~800. This general increase in the required number of cells is reflected in the obvious increase in model resolution (Figure 13), although the main features of *PKPab-df*-only and the *PKPab-df* + *PKPbc-df* models are in excellent agreement, indicating a strong compatibility between the two data sets. The resulting average model explains 51% of the *PKPab-df* residuals and 49% of the *PKPbc-df* residuals, and the RMS of the perturbations is 1.02%. The estimated noise of the *PKPab-df* data set is significantly greater than that of the *PKPbc-df* data set (0.62 versus 0.34 s). This difference is likely a reflection of multiple factors, one of which is the decrease in scatter of the *PKPbc-df* residual data compared to the *PKPab-df* residual data (Figure 5). Another consideration is that *PKPbc* waves sample much shallower in the inner core, which means they are less attenuated, and therefore more easily and accurately measured than the deeper-sampling *PKPab* waves. Additionally, one must apply a Hilbert transform to

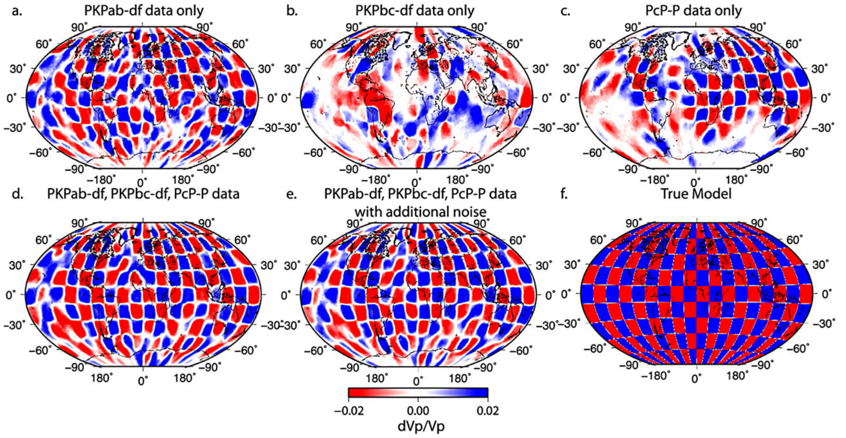


**Figure 13.** The recovered model using (a) the *PKPab-df* data set only, (b) the *PKPbc-df* data set only, (c) the *PcP-P* data set only, (d) all three data sets, and (e) all three data sets with additional noise added. Also shown is (f) the true model used in the checkerboard resolution tests to produce synthetic data. This input model (Figure 13f) has alternating 20° squares of ±2% velocity perturbations.

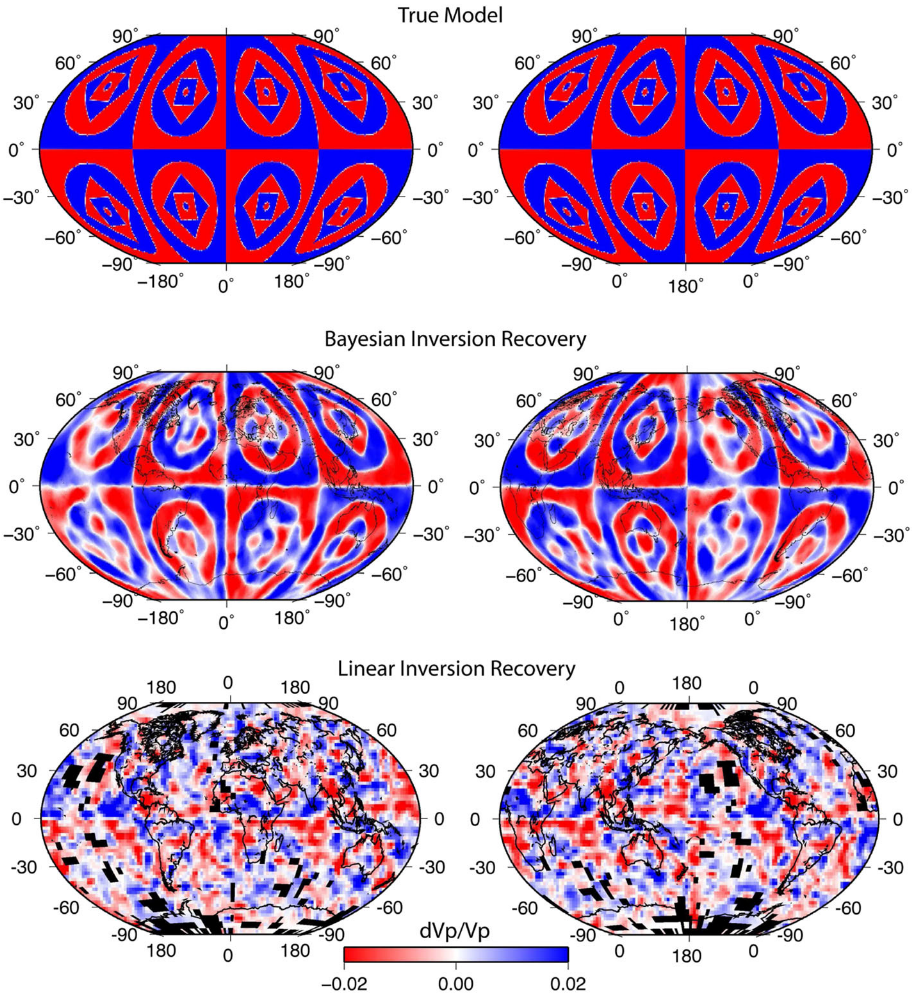


**Figure 14.** Synthetic resolution test using (top) an irregular true model. The positive anomalies are 2% faster than the ak135 average velocity and the negative anomalies are 2% slower. (middle) The recovered model using a Bayesian inversion of the *PKPab-df*, *PKPbc-df*, and *PcP-P* data sets. (bottom) The recovered model using a more traditional linearized inversion method from *Tkalčić et al.* [2002].

the *PKPab* waveforms prior to comparison with the *PKPdf* waveform; this procedure can result in approximations and errors in the measurements. Finally, the *PKPbc* and *PKPdf* raypaths travel much closer together throughout the mantle than the *PKPab* and *PKPdf* raypaths, meaning that unwanted mantle effects will be greatly diminished. Nonetheless, the *PKPab-df* data set is critical to the inversion as it provides significantly better spatial coverage than the smaller *PKPbc-df* data set. The noise estimates of the two data sets act as relative weights in the inversion. Even though the *PKPbc-df* data set is smaller, it has less associated error and will therefore have a similar impact on the final model as the *PKPab-df* data set.

[47] Next, we add the *PcP-P* differential travel times to the *PKPab-df* times (Figure 11b). The two data sets have similar noise estimates (0.76 s for *PcP-P* and 0.61 s for *PKPab-df*) and so have comparable weight in the inversion. Figure 11a shows the tomographic model TRH_KC from the work of *Tkalčić et al.* [2002] for comparison. In this case, a linear inversion of *PKPab-df* and *PcP-P* travel time residuals was performed on a grid of variable size blocks. Blocks of inadequate sampling are colored black. Since damping and smoothing procedures were applied, the amplitude variations of model TRH_KC (RMS of 0.31%) are about a third as strong as the amplitudes retrieved by our modeling (RMS of 0.88%). The absence of block parameterization and smoothing procedures and the increase in spatial sampling all contribute to the increased resolution of our final model. Despite the vast differences in inversion method, the large-scale features of the two models are in good agreement. For example, Canada, the southwest Pacific, most of South America, and the south Atlantic are slow, while Asia, Central America, Antarctica, and the Middle East are fast. Even some of the finer-scale features are congruous, such as the sharp transition from fast to slow velocities at the eastern Alaskan border. Besides improvements in resolution, our model is noteworthy in that we are able to retrieve the strength of perturbations as well as the model uncertainty (Figure 11c). Model uncertainty is expressed visually by plotting the standard deviation of the velocity distribution at each pixel. Another appealing feature is that, because the variable nature of both the

parameterization and the data noise, the resolution of the final model is automatically controlled by the information content of the data.

### 4.3. Final P Wave Velocity Model

[48] When we invert all three data sets (Figure 8), we obtain a high-resolution *P* wave velocity model of the lowermost mantle as evidenced by the resolution test in Figure 13f. Convergence requires over 500 CPU hours on each of 60 processors. The *PKPbc-df* data set has the lowest estimated noise level of 0.37 s; the *PcP-P* and *PKPab-df* noise estimates are again very close (0.78 s and 0.66 s). The RMS of the velocity perturbations is 0.87% with a maximum perturbation of 4.74%, which is significantly stronger than previous estimates [e.g., *Tkalčić et al.*, 2002; *Antolik et al.*, 2003; *Lei and Zhao*, 2006; *Houser et al.*, 2008; *Li et al.*, 2008; *Della Mora et al.*, 2011; *Soldati et al.*, 2012].

[49] The resulting model reduces the differential travel time variance relative to a homogeneous model of 13.61 km/s by 45% for the *PKPbc-df* data set, 49% for the *PKPab-df* data set, and 23% for the *PcP-P* data set, which suggests a general compatibility of the three data sets. The noise, sampling, and size of each data set prevents further reduction of data misfit. Traditional linear inversions seek to produce a single model that minimizes the variance of the data; the Bayesian method, however, produces an ensemble of solutions whose complexity is reflected by the interpreted noise in the data. The data uncertainty determines how accurately the measurements are fit. Consequently, one should not consider the variance reductions given in the study in quite the same light as would be done for a linearized inversion, as data noise and unmodeled effects are also accounted for. Furthermore, the average model is only one measure of the ensemble of solutions, and it is often possible to select individual models that yield greater reductions in data variance [e.g., *Shapiro and Campillo*, 2004; *Moschetti et al.*, 2010; *Behr et al.*, 2010].

[50] The large-scale patterns inferred here are consistently seen in other travel time tomography models [e.g., *van der Hilst and Kárason*, 1999; *Tkalčić et al.*, 2002; *Antolik et al.*, 2003; *Vasco et al.*, 2006; *Zhao*, 2004; *Lei and Zhao*, 2006; *Li et al.*, 2008; *Houser et al.*, 2008; *Zhao*, 2009; *Soldati et al.*, 2012], as there is agreement about the presence of fast velocities beneath Central America and East Asia and slow velocities beneath south Africa and the southwest Pacific. Common areas of discrepancy include North America, Australia, and Europe, which could be a result of poorer sampling, systematic errors in travel time picks, inversion method, and/or data noise. The chief advantage of our model over previous models is that because there is no arbitrary smoothing, damping, or grid-spacing, the scale-size and amplitude of the velocity heterogeneity are controlled directly by the data.

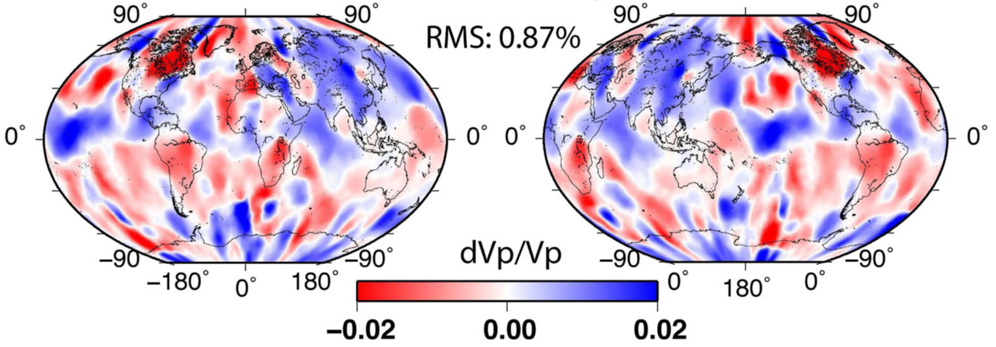
## 5. Discussion

[51] The primary goal of this work is to invert for the *P* wave velocity heterogeneity in the lowermost mantle as can be reliably retrieved from body wave differential travel times. The scale length of the heterogeneity depends on the data set(s) used, but invariably structure ranging from wavelengths of hundreds to thousands of kilometers is revealed. Lowermost mantle heterogeneity at even smaller scales is almost certainly present, however, [e.g., *Cormier*, 1999; *Helffrich*, 2002; *Margerin and Nolet*, 2003; *Garcia et al.*, 2009], but its resolution is not justified by the inferred noise levels of our data sets. It is important to note that no single data set, or even pair of data sets, produces as high resolution a map as does the combination of all three data sets (Figure 13). Each data set provides unique and invaluable information that is critical to revealing the complexities of the lowermost layer. The *PKPbc-df* data set is crucial to identifying the exact location of anomaly edges, as the *PKPbc* and *PKPdf* raypaths are very close within the lowermost mantle (raypath separation < 3° ). The *PKPab-df* data set is the largest and provides the best spatial coverage. However, the *PKPab* waves sample deeper in the inner core, making them more attenuated and more difficult to measure that the more shallow-sampling *PKPbc* waves. The *PKPab-df* data set also suffers from small errors resulting from the application of the Hilbert transform prior to alignment of the *PKPab* waveform with the *PKPdf* waveform. Finally, the *PcP-P* data set is important because it is not subject to inner core effects and because of its ability to resolve ambiguity about the location of heterogeneity on source or receiver sides. The capabilities of our inversion approach combined with an exceptionally high-quality data set enables the inversion for a global velocity model of the lowermost mantle with unprecedented reliability. The difference between our results and previous models is likely due to our evasion of block parameterization, truncation of spherical harmonic expansions, and smoothing and damping regularization and our improvements to spatial coverage and data quality.

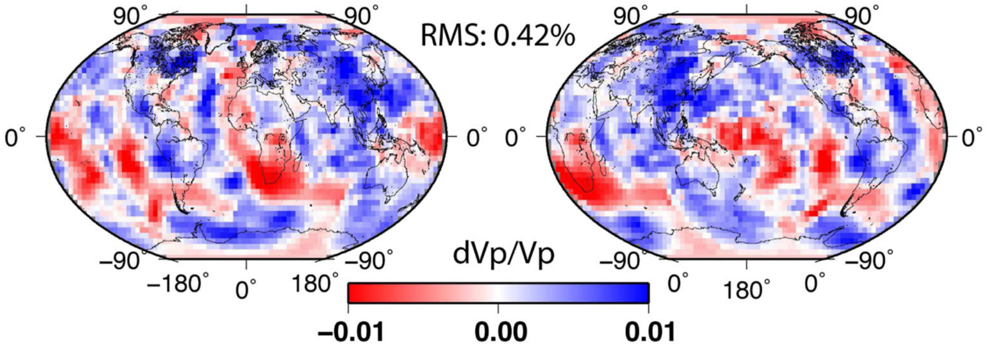
[52] The RMS heterogeneity level of 0.87% from our final tomographic model is significantly larger than the majority of previous estimates obtained from body wave travel time analysis [*Tkalčić et al.*, 2002; *Zhao*, 2004; *Lei and Zhao*, 2006; *Zhao*, 2009; *Houser et al.*, 2008; *Soldati et al.*, 2012]. *Garcia et al.* [2009], who performed a statistical analysis of *P* wave heterogeneity in the lowermost mantle also obtained a higher estimate (1.2±0.3%). This can be explained by the fact that like in our study, *Garcia et al.* [2009] did not use the damping procedures typically employed in other inversion approaches. There is evidence, however, that *P* wave velocity variations in the lowermost mantle can reach even greater extremes than is revealed in our study. For example, in ultralow velocity zones (ULVZs) *P* wave anomalies can be up to 10% [*Ni and Helmberger*, 2001b; *Garnero et al.*, 1998]. We cannot expect to reveal such features, however, for the thickness of ULVZs is on the order of tens of kilometers [*Wen and Helmberger*, 1998; *Garnero et al.*, 1998], whereas our model presents a depth-averaged image of the lowermost 300 km.

[53] Distinguishing between the possible origins of the imaged *P* wave velocity anomalies is difficult, but our maps indicate that the cause(s) must correspond to lateral dimensions on the order of hundreds to a few thousand kilometers. This is an important constraint which supports the findings of both large-scale tomographic models [e.g., *Sylvander et al.*, 1997; *Zhao*, 2004] and work on localized regions [e.g., *Garnero and Lay*, 2003]. The large-scale features of our final model agree well with the results of *Tkalčić et al.* [2002]. For example, Canada, the southwest

a. P-wave Model Result of This Study



b. P-wave Model from Houser et al. 2008



c. P-wave Model from Soldati et al. 2012

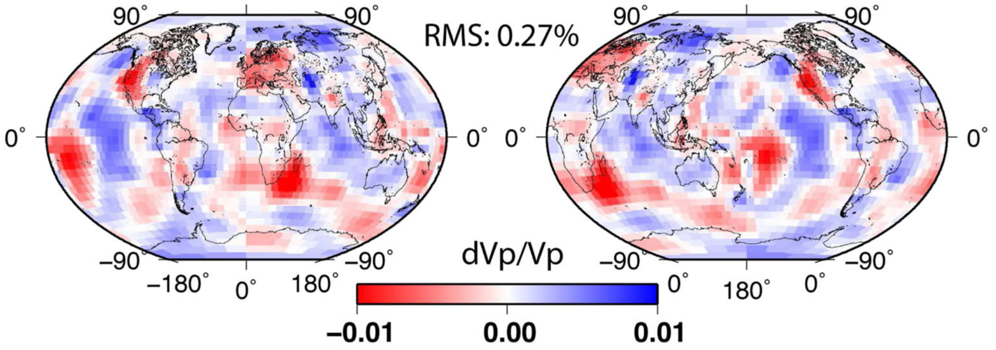


**Figure 15.** (a) The *P* wave model of the lowermost 300 km of the mantle from this study. (b) The *P* wave model of the lowermost 200 km of the mantle from *Houser et al.* [2008]. (c) The *P* wave model of the lowermost 200 km of the mantle from *Soldati et al.* [2012]. Note the differences in the color scale ranges for the different models. The root-mean-square (RMS) value of each model is noted.

Pacific, most of South America, and the south Atlantic are slow, while Asia, Central America, Antarctica, and the Middle East are fast. Even some of the finer-scale features are congruous, such as the sharp transition from fast to slow velocities at the eastern Alaskan border. Besides improvements in resolution, our model is noteworthy in that we are able to retrieve the strength of perturbations as well as the model uncertainty (Figure 11c).

[54] Figure 15 shows a comparison of our results with those of *Houser et al.* [2008] and *Soldati et al.* [2012]. The *P* wave model from *Houser et al.* [2008] is of the lowermost 200 km of the mantle. Over 290,000 *P*, *PP*, *PP-P*, and Rayleigh wave travel times were used in a least-squares inversion on a grid of equal area blocks (4°×4° at the equator). The applied smoothing and damping is the likely cause of the 0.42% RMS of the velocity perturbations, which is significantly less than that of our model. Whether in regions of good or poor raypath coverage, the RMS of the perturbations in our model is consistently higher than that of the *Houser et al.* [2008] model. For example, the RMS of only the well-sampled Asia region (from 30 to 120° in longitude and 30 and 80° in latitude) for our model is 0.98%, while it is only 0.53% in the *Houser et al.* [2008] model. The RMS of the more poorly sampled southeast Pacific region (from –180 to –100° in longitude and 0 to –60° in latitude) is 0.64 and 0.49, respectively. The areas of best coverage in the *Houser et al.* [2008] model include Asia, the northwest Pacific, and the north Atlantic. These areas agree with our results, and it is only in areas of poor resolution (as determined by the checkerboard resolution tests of *Houser et al.* [2008]) that we see significant discrepancies. Our model is slow beneath Canada,

south Australia, and mid-South America and fast beneath the mid-east Pacific whereas the opposite holds true in the *Houser et al.* [2008] model. Because very few *P*, *PP*, and *PP-P* raypath geometries are sensitive to lowermost mantle structure, the resolving power in the lowermost mantle of the *Houser et al.* [2008] study is much weaker than that of our study. Since we employ a variety of seismic phase measurements that sample the lowermost mantle and use differential travel times, our resolution is such that most areas of the globe in the lowermost mantle layer can be reliably retrieved.

[55] The *P* wave model of the lowermost 200 km of the mantle from *Soldati et al.* [2012] is shown in Figure 15c. Over 880,000 ISC *PcP*, *PKPbc*, *PKPdf*, and *P* travel times along with a viscosity profile of the mantle were used in a least squares inversion on a grid of equal area blocks ($5^\circ \times 5^\circ$ at the equator). Here the RMS of the perturbations is even lower (0.27% overall, 0.32% in the well-sampled Asia region, and 0.36% in the poorly sampled southeast Pacific region), but we see more agreement with the distribution of perturbations. This is likely due to the overall increase in data number and data type. There are strikingly similar features in Africa and Asia; however, there are also differences elsewhere, such as the mid-Pacific, Australia, and North America. Again, this is attributable to differences in data quality, spatial coverage, and handling of data noise. For example, there are likely to be systematic biases in the *PKP* catalog data due to the fact that no Hilbert transform is applied when calculating differential travel times. The cross-correlation coefficient between the *Houser et al.* [2008] and the *Soldati et al.* [2012] models is 0.224; that between ours and that of *Houser et al.* [2008] is 0.149, and that between ours and that of *Soldati et al.* [2012] is 0.146. This suggests a reasonable level of agreement with our model, especially considering that we do not apply any smoothing regularization.

[56] The lowermost mantle has long been considered the graveyard of subducted slabs [*Richards and Engebretson*, 1992; *Grand*, 2002] and the birth place of mantle plumes [*Yuen and Peltier*, 1980; *Stacey and Loper*, 1983]. One possible explanation for the large-scale anomalously fast zones at the bottom of the mantle is the penetration of slab material into the lowermost mantle and its subsequent repose at the base of the mantle. In this manner, the thermal and chemical heterogeneity of a “slab graveyard” would account for some of the lowermost mantle velocity anomalies [*Grand*, 2002; *Garnero and Lay*, 2003]. There has been some evidence in support of the connectivity of slabs from the surface down to the bottom of the mantle. For example, the Caribbean is consistently shown to have high-velocity structures extending down to the core mantle boundary [*Kito et al.*, 2008]. There is geochemical evidence as well that crustal material can descend to the lowermost mantle. *Hirose et al.* [1999] argue that former basaltic crust with perovskite lithology would gravitationally sink to the deep mantle.

[57] Whether chemical or thermal in nature, the causes of the velocity heterogeneity in the lowermost mantle must be considered to have a larger impact on *P* wave velocity than previously suggested by global tomographic models, as evidenced by the significant increase in the RMS of the velocity anomalies predicted by our results. This is in line with the strong lateral velocity gradients across compositionally varying domains discovered previously from a direct comparison of *PcP-P* and *ScS-S* data [*Tkalčić and Romanowicz*, 2002]. Furthermore, velocity variations up to 4.74%, as seen in our study, is unlikely merely an effect of the core mantle boundary topography, and we instead favor lateral variations in temperature and/or chemistry as explanation for the observed compressional wave velocity variations. A joint inversion for velocity structure and topography is needed to test this interpretation.

## 6. Conclusions

[58] We present a new approach to global tomography using a hand-picked data set of *PKPab-df*, *PKPbc-df*, and *PcP-P* differential travel times. We use a probabilistic, fully nonlinear Bayesian inversion scheme to invert for lowermost mantle structure and obtain a new model of the distribution and amplitude of the *P* wave velocity heterogeneity in the lowermost mantle. Model parameters, including the level of data noise, are treated as unknowns in the inversion problem and are therefore driven by the information content of the data. The resulting *P* wave velocity model reveals heterogeneity on a range of scale lengths and provides an important bridge between the long-wavelength images produced from previous global models and the very short-scale mapping of localized scattering studies. The root-mean-square of the velocity perturbations in our final tomographic model is 0.87%, which is significantly larger than previous estimates obtained from a global-scale analysis of body wave travel times. Importantly, model uncertainty is also retrieved, which is a major step forward for global-scale tomographic inversions.

[59] Our results provide a unique view of the lowermost mantle, as the resolving capability is better than that of previous global models, yet is not limited spatially to a local or regional context as are current high-resolution images based on scattering or array seismology methods. The most dominant features of our preferred model include fast velocities beneath Central America and east Asia and slow velocities beneath southern Africa and the southwest Pacific. These large-scale patterns agree with other travel time tomography models [e.g., *van der Hilst and Kárason*, 1999; *Tkalčić et al.*, 2002; *Antolik et al.*, 2003; *Vasco et al.*, 2006; *Zhao*, 2004; *Lei and Zhao*, 2006; *Li et al.*, 2008; *Houser et al.*, 2008; *Zhao*, 2009; *Soldati et al.*, 2012]. Our model also includes new insights on the *P* wave velocity structure of more difficult to image regions such as Africa, Canada, South America, and Australia. Most of these regions are slower than average in our model; however, the opposite holds true in the models of some previous studies [e.g., *Houser et al.*, 2008; *Soldati et al.*, 2012].

[60] A further consideration includes accounting for the possible effects of topography on the travel time residuals, which will involve a joint inversion for topography and velocity. This inclusion will likely decrease the data variance and increase the reliability of our velocity model. In this paper, we made the assumption that the lowermost mantle can be modeled as a single layer. Forthcoming work also includes allowing for multiple layers in the lowermost mantle model, where the thickness of these layers is an unknown. Given the continual increase in available computing power and data records, the Bayesian approach to

inversion should enable an ever-improving understanding of lowermost mantle heterogeneity and its geodynamical implications.

[61] **Acknowledgments.** We thank Christine Houser and Lapo Boschi for providing their mantle models. Calculations were performed on the Terrawulf cluster, a computational facility supported through the AuScope Australian Geophysical Observing System (AGOS). AuScope is funded under the National Collaborative Research Infrastructure Strategy (NCRIS), and the Education investment Fund (EIF3) both Australian Commonwealth Government Programs. We appreciate Caroline Bartlett for her contributions to the *PcP-P* data set. Mallory Young was supported by an Australian National University Research Scholarship. We also thank two anonymous reviewers.

## References

Antolik, M., Y. Gu, G. Elström, and A. Dziewonski (2003), J362D28: A new joint model of compressional and shear velocity in the Earth's mantle, *Geophys. J. Int.*, *153*, 443–446.

Bataille, K., and S. Flatte (1988), Inhomogeneities near the core-mantle boundary inferred from short-period scattered *PKP* waves recorded at the Global Digital Seismograph Network, *J. Geophys. Res.*, *93*, 15,057–15,064.

Bayes, T. (1763), An essay towards solving a problem in the doctrine of chances, *Phil. Trans. R. Soc.*, *53*, 370–418.

Beghein, C., J. S. Resovsky, and J. Trampert (2002), *P* and *S* tomography using normal-mode and surface waves data with a neighbourhood algorithm, *Geophys. J. Int.*, *149*, 646–658.

Behr, Y., J. Townend, S. Bannister, and M. Savage (2010), Shear velocity structure of the Northland Penninsula, New Zealand, inferred from ambient noise correlations, *J. Geophys. Res.*, *115*, B05309, doi:10.1029/2009JB006737.

Bernando, J., and A. Smith (1994), *Bayesian Theory*, John Wiley, New York.

Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, *173*(3), 1411–1436.

Bodin, T., M. Sambridge, N. Rawlinson, and P. Arroucau (2012a), Transdimensional tomography with unknown data noise, *Geophys. J. Int.*, *189*, 1536–1556.

Bodin, T., M. Sambridge, H. Tkalčić, P. Arroucau, K. Gallagher, and N. Rawlinson (2012b), Transdimensional inversion of receiver functions and surface wave dispersion, *J. Geophys. Res.*, *117*, B02301, doi:10.1029/2011JB008560.

Boschi, L., and A. M. Dziewonski (2000), Whole Earth tomography from delay times of *P*, *PcP*, and *PKP* phases: Lateral heterogeneities in the outer core or radial anisotropy in the mantle? *J. Geophys. Res.*, *105*, 13,675–13,696.

Box, G., and G. Tiao (1973), *Bayesian Inference in Statistical Analysis*, John Wiley, New York.

Bréger, L., and B. Romanowicz (1998), Three-dimensional structure at the base of the mantle beneath the central Pacific, *Science*, *282*, 718–720.

Bréger, L., H. Tkalčić, and B. Romanowicz (2000), The effect of D" on *PKP(AB-DF)* travel time residuals and possible implications for inner core structure, *Earth Planet. Sci. Lett.*, *175*, 133–143.

Calvet, M., and S. Chevrot (2005), Traveltime sensitivity kernels for *PKP* phases in the mantle, *Phys. Earth Planet. Inter.*, *153*, 21–31.

Chevrot, S., R. Martin, and D. Komatitsch (2012), Optimized discrete wavelet sphere with the lifting scheme—Implications for global finite-frequency tomography, *Geophys. J. Int.*, *191*, 1391–1402.

Cormier, V. (1999), Anisotropy of heterogeneity scale lengths in the lower mantle from *PKIKP* precursors, *Geophys. J. Int.*, *136*, 373–384.

Cormier, V., and G. L. Choy (1986), A search for lateral heterogeneity in the inner core from differential travel times near *PKP-D* and *PKP-C*, *Geophys. Res. Lett.*, *13*, 1553–1556.

Creager, K. (1999), Large-scale variations in inner core anisotropy, *J. Geophys. Res.*, *104*, 23,127–23,139.

Crotwell, H. P., T. J. Owens, and J. Ritsema (1999), The TauP Toolkit: Flixible seismic travel-time and ray-path utilities, *Seismol. Res. Lett.*, *70*, 154–160.

Della Mora, L., L. Boschi, P. J. Tackley, T. Nakagawa, and D. Giardini (2011), Low seismic resolution cannot explain *S*/*P* decorrelation in the lower mantle, *Geophys. Res. Lett.*, *38*, L12303, doi:10.1029/2011GL047559.

Doornbos, D. (1974), Seismic wave scattering near caustics: Observations of *PKKP* precursors, *Nature*, *247*, 352–353

Dziewonski, A. (1984), Mapping the lower mantle: Determination of lateral heterogeneity in *P* wave velocity up to degree and order 6, *J. Geophys. Res.*, *89*(7), 5929–5952.

Dziewonski, A., B. H. Hager, and R. J. O'Connel (1977), Large-scale heterogeneities in the lower mantle, *J. Geophys. Res.*, *82*(2), 239–255.

Earle, P. S., S. Rost, P. M. Shearer, and C. Thomas (2011), Scattered *P'P'* waves observed at short distances, *B. Seismol. Soc. Am.*, *101*, 2843–2854.

Engdahl, E., R. D. Van der Hilst, and R. Buland (1998), Global teleseismic earthquake relocation with improved travel times and procedures for depth determination, *B. Seismol. Soc. Am.*, *88*, 722–743.

Fuji, N., S. Chevrot, L. Zhao, R. J. Geller, and K. Kawai (2012), Finite-frequency structural sensitivities of short-period compressional body waves, *Geophys. J. Int.*, *190*, 522–540.

Fukao, Y., A. To, and M. Obayashi (2003), Whole mantle *P* wave tomography using *P* and *PP-P* data, *J. Geophys. Res.*, *108*(B1), 2021, doi:10.1029/2001JB000989.

Garcia, R., J. Revenaugh, and M. Weber (2004), Nonlinear waveform and delay time analysis of triplicated core phases, *Geophys. J. Int.*, *109*, doi:10.1029/2003JB002429.

Garcia, R., H. Tkalčić, and S. Chevrot (2006), A new global *PKP* dataset to study Earth's core and deep mantle, *Phys. Earth Planet. Inter.*, *159*, 15–31.

Garcia, R., S. Chevrot, and M. Calvet (2009), Statistical study of seismic heterogeneities at the base of the mantle from *PKP* differential traveltimes, *Geophys. J. Int.*, *179*, 1607–1616.

Garnero, E., and T. Lay (2003), D" shear velocity heterogeneity, anisotropy and discontinuity structure beneath the Caribbean and Central America, *Phys. Earth Planet. Inter.*, *140*, 219–242.

Garnero, E., J. Revenaugh, Q. Williams, T. Lay, and L. Kellogg (1998), Ultralow velocity zone at the core-mantle boundary, *Geodynamics*, *28*, 319–334.

Gouveia, W., and J. Scales (1998), Bayesian seismic waveform inversion—Parameter estimation and uncertainty analysis, *J. Geophys. Res.*, *103*, 2759–2780.

Grand, S. (2002), Mantle shear-wave tomography and the fate of subducted slabs, *Phil. Trans. R. Soc. A*, *360*, 2475–2491.

Grand, S., R. D. van der Hilst, and S. Widiyantoro (1997), High resolution global tomography: A snapshot of convection in the Earth, *Geol. Soc. Am. TODAY*, *7*, 1–7.

Green, P. J. (1995), Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika*, *82*, 711–732.

Green, P. J. (2003), Transdimensional Markov chain Monte Carlo, in *Highly Structured Stochastic Systems*, edited by P. J. Green, N. Hjort, and S. Richardson, pp. 179–196, Oxford Statistical Sciences Series, Oxford, U. K.

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

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*, 97–109.

He, Y., and L. Wen (2009), Structural features and shear-velocity structure of the "Pacific Anomaly", *J. Geophys. Res.*, *114*, B02309, doi:10.1029/2008JB005814.

Helffrich, G. (2002), Chemical and seismological constraints on mantle heterogeneity, *Phil. Trans. R. Soc. Lond. A*, *360*, 2493–2505.

Hirose, K., Y. Fei, Y. Ma, and H. Mao (1999), The fate of subducted basaltic crust in the Earth's lower mantle, *Nature*, *397*, 53–56.

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

Inoue, H., Y. Fukao, K. Tanabe, and Y. Ogata (1990), Whole mantle *P*-wave travel time tomography, *Phys. Earth Planet. Inter.*, *59*, 294–328.

Ishii, M., and J. Tromp (2004), Constraining large-scale mantle heterogeneity using mantle and inner-core sensitive normal modes, *Phys. Earth Planet. Inter.*, *146*, 113–124.

Kárason, H., and R. D. van der Hilst (2001), Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (*PKP*, *Pdiff*), *J. Geophys. Res.*, *106*(B4), 6569–6587.

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

Kennett, B. L. N., and O. Gudmundsson (1996), Ellipticity corrections for seismic phases, *Geophys. J. Int.*, *127*, 40–48.

Khan, A., A. Zunino, and F. Deschamps (2011), The thermo-chemical and physical structure beneath the North American continent from Bayesian inversion of surface-wave phase velocities, *J. Geophys. Res.*, *116*, B09304, doi:10.1029/2011JB008380.

Kito, T., C. Thomas, A. Rietbrock, E. Garnero, S. Nippress, and A. Heath (2008), Mantle shear-wave tomography and the fate of subducted slabs, *Geophys. J. Int.*, *174*, 1019–1028.

Lei, J., and D. Zhao (2006), Global *P*-wave tomography: On the effect of various mantle and core phases, *Phys. Earth Planet. Inter.*, *154*, 44–69.

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

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

Malinverno, A. (2002), Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, *151*(3), 675–688.

Malinverno, A., and V. Briggs (2004), Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics*, *69*, 1005–1016.

Malinverno, A., and R. Parker (2006), Two ways to quantify uncertainty in geophysical inverse problems, *Geophysics*, *71*, W15–W27.

Margerin, L., and G. Nolet (2003), Multiple scattering of high-frequency seismic waves in the deep Earth: *PKP* precursor analysis and inversion for mantle granularity, *J. Geophys. Res.*, *108*, 2514, doi:10.1029/2003JB002455.

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

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1093.

Monelli, D., and P. M. Mai (2008), Bayesian inference of kinematic earthquake rupture paramters through fitting of strong motion data, *Geophys. J. Int.*, *173*, 220–232.

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

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

Moschetti, M. P., M. H. Ritzwoller, F.-C. Lin, and Y. Yang (2010), Crustal shear wave velocity structure of the western United States inferred from ambient seismic noise and earthquake data, *J. Geophys. Res.*, *115*, B10306, doi:10.1029/2010JB007448.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*(B7), 12,431–12,447.

Myers, S. C., G. Johannesson, and W. Hanley (2007), A Bayesian hierarchical method for multiple-event seismic location, *Geophys. J. Int.*, *171*, 1049–1063.

Nakanishi, I., and D. L. Anderson (1982), Worldwide distribution of group velocity of mantle Rayleigh waves as determined by spherical harmonic inversion, *B. Seismol. Soc. Am.*, *72*(4), 1185–1194.

Ni, S., and D. V. Helmberger (2001a), Horizontal transition from fast to slow structures at the core-mantle boundary; south Atlantic, *Earth Planet. Sci. Lett.*, *187*(3-4), 301–310.

Ni, S., and D. V. Helmberger (2001b), Probing an ultra-low velocity zone at the core mantle boundary with *P* and *S* waves, *Geophys. Res. Lett.*, *28*(12), 2345–2348.

Piana Agostinetti, N., and A. Malinverno (2010), Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.*, *181*, 858–872.

Richards, M. A., and D. C. Engebretson (1992), Large-scale mantle convection and the history of subduction, *Nature*, *355*, 437–440.

Ritsema, J., S. Ni, D. Helmberger, and H. Crotwell (1998), Evidence for strong shear velocity reductions and velocity gradients in the lower mantle beneath Africa, *Geophys. Res. Lett.*, *24*, 4245–4248.

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

Rost, S., and P. S. Earle (2010), Identifying regions of strong scatering at the core-mantle boundary from analysis of *PKKP* precursor energy, *Earth Planet. Sci. Lett.*, *297*, 616–626.

Sambridge, M. (1999), Geophysical inversion with a neighbourhood algorithm I. Searching a parameter space, *Geophys. J. Int.*, *103*, 4839–4878.

Sambridge, M., and R. Faletič (2003), Adaptive whole Earth tomography, *Geochem. Geophys. Geosyst.*, *4*(3), 1022, doi:10.1029/2001GC000213.

Sambridge, M., T. Bodin, K. Gallagher, and H. Tkalčić (2013), Transdimensional inference in the geosciences, *Philos. Trans. R. Soc. A*, *371* (20110547), doi:10.1098/rsta.2011.0547.

Scales, J., and R. Snieder (1998), What is noise? *Geophysics*, *63*, 1122–1124.

Shapiro, N. M., and M. Campillo (2004), Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, *31*, L07614, doi:10.1029/2004GL019491.

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

Simmons, N., A. Forte, and S. 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*, 1284–1304.

Simons, F. J., I. Loris, G. Nolet, I. C. Daubehies, S. Voronin, J. S. Judd, P. A. Vetter, J. Charlety, and C. Vonesch (2011), Solving or resolving global tomographic models with spherical wavelets, and the scale and spasity of seismic heterogeneity, *Geophys. J. Int.*, *187*(2), 969–988.

Smith, A. (1991), Bayesian computational methods, *Philos. Trans. R. Soc. A*, *337*, 369–386.

Soldati, G., L. Boschi, and A. Forte (2012), Tomography of core-mantle boundary and lowermost mantle coupled by geodynamics, *Geophys. J. Int.*, *189*, 730–746.

Stacey, F., and D. Loper (1983), The thermal boundary interpretation of D and its role as a plume source, *Phys. Earth Planet. Inter.*, *33*, 45–55.

Su, W., R. Woodward, and A. Dziewonski (1994), Degree 12 model of shear velocity heterogeneity in the mantle, *J. Geophys. Res.*, *99*, 6945–6980.

Sylvander, M., and A. Souriau (1996), *P*-velocity structure of the core-mantle boundary region inferred from *PKP(A B)-PKP(BC)* differential travel times, *Geophys. Res. Lett.*, *23*, 853–856.

Sylvander, M., B. Ponce, and A. Souriau (1997), Seismic velocities at the core-mantle boundary inferred from *P* waves diffracted around the core, *Phys. Earth Planet. Inter.*, *101*, 189–202.

Tanaka, S. (2010), Constraints on the core-mantle boundary topography from *P4KP-PcP* differential travel times, *J. Geophys. Res.*, *115*, B04310, doi:10.1029/2009JB006563.

Tarantola, T., and B. Valette (1982), Inverse problems = Quest for information, *J. Geophys.*, *50*, 159–170.

Tkalčić, H. (2010), Large variations in travel times of mantle-sensitive seismic waves from the South Sandwich Islands: Is the Earth's inner core a conglomerate of anisotropic domains? *Geophys. Res. Lett.*, *37*, L14312, doi:10.1029/2010GL043841.

Tkalčić, H., and B. Romanowicz (2002), Short scale heterogeneity in the lowermost mantle: Insights from *PcP-P* and *ScS-S* data, *Earth Planet. Sci. Lett.*, *201*, 57–68.

Tkalčić, H., B. Romanowicz, and N. Houy (2002), Constraints on D" structure using *PKP(AB-DF)*, *PKP(BC-DF)* and *PcP-P* traveltime data from broad-band records, *Geophys. J. Int.*, *148*, 599–616.

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

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

van der Hilst, R. S., S. Widiyantoro, and E. R. Engdahl (1997), Evidence for deep mantle circulation from global tomography, *Nature*, *386*, 578–584.

Vasco, D., and L. Johnson (1998), Whole Earth structure estimated from seismic arrival times, *J. Geophys. Res.*, *103*, 2633–2671.

Vasco, D., L. Johnson, and J. Pulliam (1995), Lateral variations in mantle velocity structure and discontinuities determined from *P*, *PP*, *S*, *SS*, and *SS-SdS* travel time residuals, *J. Geophys. Res.*, *100*, 24,037–24,059.

Vasco, D., L. Johnson, and O. Marques (2006), Resolution, uncertainty, and whole Earth tomography, *J. Geophys. Res.*, *108*, 2022, doi:10.1029/2001JB000412.

Vidale, J. E., and M. Hedlin (1998), Evidence for partial melt at the core-mantle boundary north of Tonga from the strong scattering of seismic waves, *Nature*, *391*, 682–685.

Wen, L., and D. V. Helmberger (1998), Ultra-low velocity zones near the core-mantle boundary from broadband *PKP* precursors, *Science*, *279*, 1701–1703.

Woodward, R. L., and G. Masters (1991), Global upper mantle structure from long-period differential travel times, *J. Geophys. Res.*, *96*(B4), 6351–6377.

Wysession, M. E. (1996), Large-scale structure at the core-mantle boundary from diffracted waves, *Nature*, *382*, 244–248.

Yuen, D. A., and W. R. Peltier (1980), Mantle plumes and thermal stability of the D layer, *Geophys. Res. Lett.*, *7*, 625–628.

Zhao, D. (2004), Global tomographic images of mantle plumes and subducting slabs: Insight into deep Earth dynamics, *Phys. Earth Planet. Inter.*, *146*, 3–34.

Zhao, D. (2009), Multiscale seismic tomography and mantle dynamics, *Gondwana Res.*, *15*, 297–323.

Zhao, D., Y. Yamamoto, and Y. Yanada (2012), Global mantle heterogeneity and its influence on teleseismic regional tomography, *Gondwana Res.*, *23*, 595–616.

Zollo, A., L. D'Auria, R. D. Matteis, A. Herrero, H. Virieux, and P. Gasparini (2002), Bayesian estimation of 2-D *P*-velocity models from active seismic arrival time data: Imaging of the shallow structure of Mt. Vesuvius (Southern Italy), *Geophys. J. Int.*, *151*, 566–582.