AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Solid Earth

RESEARCH ARTICLE
10.1002/2015JB012641

Constraints on the heterogeneity spectrum of Earth's upper mantle

Nicholas Mancinelli[1], Peter Shearer[1], and Qinya Liu[2]

[1]Scripps Institution of Oceanography, University of California, San Diego, California, USA, [2]Department of Physics and Earth Sciences, University of Toronto, Toronto, Ontario, Canada

Key Points:
- Constraints on the upper mantle heterogeneity spectrum from seismic scattering observations
- New observations of long-period P coda envelopes constrain the intermediate-scale heterogeneity
- A self-similar mix of basalt and harzburgite may produce the small-scale portion of the spectrum

Correspondence to:
N. Mancinelli,
nicholas_mancinelli@brown.edu

Citation:
Mancinelli, N., P. Shearer, and Q. Liu (2016), Constraints on the heterogeneity spectrum of Earth's upper mantle, *J. Geophys. Res. Solid Earth*, *121*, doi:10.1002/2015JB012641.

Received 3 NOV 2015
Accepted 30 APR 2016
Accepted article online 13 MAY 2016

Abstract We constrain the heterogeneity spectrum of Earth's upper mantle at scales from a few kilometers to tens of thousands of kilometers using observations from high-frequency scattering, long-period scattering, and tomography. Tomography and high-frequency scattering constraints are drawn from previous studies, but constraints on mantle heterogeneity at intermediate scales (5–500 km) are lacking. To address this, we stack ~15,000 long-period *P* coda envelopes to characterize the globally averaged scattered wavefield at periods from 5 to 60 s and at ranges from 50 to 98°. To fit these observations, we consider models of random mantle heterogeneity and compute the corresponding global wavefield using both a ray theoretical "seismic particle" approach and full spectral element simulations. Von Kármán random media distributed throughout the uppermost 600 km of the mantle with $a = 2000$ km, $\epsilon = 10\%$, and $\kappa = 0.05$ provide a good fit to the time, range, and frequency dependence of the stacks, although there is a trade-off between ϵ and the thickness of the assumed scattering layer. This random media model also fits previously published 1 Hz stacks of *P* coda and agrees with constraints on long-wavelength structure from tomography. Finally, we explore geodynamically plausible scenarios that might be responsible for the RMS and falloff rate of the proposed spectrum, including a self-similar mixture of basalt and harzburgite.

1. Introduction

Seismic tomography has imaged the large-scale structure of the mantle and seismic scattering studies have constrained the strength of mantle heterogeneity at very short length scales, but the full heterogeneity spectrum of the mantle has not yet been resolved. In particular, there is a sizable gap in our knowledge of mantle heterogeneity at intermediate-scale lengths between about 5 and 500 km. Gaining a more complete understanding of mantle heterogeneity at all scales is essential to addressing key questions about Earth's history and dynamics, such as follows: What is the nature of mantle convection? Can geochemical reservoirs persist throughout geologic time? What is the ultimate fate of subducting slabs?

Over the past decades, global seismic tomography has made great progress toward constraining both the amplitude and location of large-scale lateral velocity perturbations in seismic velocities throughout Earth's mantle. Such efforts have resulted in a number of tomography models from different groups, all of which are based on regularized inversions of several types of seismic data, including (1) travel times of direct and reflected phases, (2) normal-mode splitting measurements, (3) surface wave dispersion measurements, and (4) full seismic waveforms. For structure within the uppermost 200 km of the mantle, most current generation shear wave models agree reasonably well to spherical harmonic degree 16, implying a global resolution limit of about 2400 km at these depths, although smaller features may be resolved in areas where favorable data coverage allows. In the deeper upper mantle (200–600 km) and the lowermost mantle (2500–2890 km), the models are well correlated out to degree 8. However, resolving large-scale middle mantle structure (600–2500 km) is more challenging; in this depth range models are often uncorrelated even at the lowest degrees. See *Becker and Boschi* [2002] and *Meschede and Romanowicz* [2015a] for detailed comparisons among many recent mantle models.

Regional tomography models can resolve heterogeneities as small as a few hundred kilometers, although the exact resolution of each model depends upon several factors including (1) the types of seismic data used (surface waves, body waves, noise cross correlations), (2) the inversion technique employed (matching travel

times versus fitting waveforms), and (3) the locations of the receivers and sources, which are unique to every region. Again, see *Meschede and Romanowicz* [2015a] for a survey of recently published regional models.

A rigorous comparison between regional and global tomography models was initially undertaken by *Passier and Snieder* [1995], who found that regional tomography models of Europe displayed 15–30 times as much power as global models at overlapping scales (about degree 30). Later, *Chevrot et al.* [1998] emphasized the difficulty of comparing global and regional models due to fundamental differences in parameterization—the former is parameterized in terms of spherical harmonic basis functions, and the latter is parameterized in terms of Fourier series—arguing that when the conversion from one basis to the other is properly done, global and regional tomography models are indeed consistent at overlapping length scales. Most recently, *Meschede and Romanowicz* [2015a] highlighted the importance of minimizing truncation effects when estimating spherical power spectra from regional models and used a multitaper approach developed by *Wieczorek and Simons* [2005] to confirm that regional and global tomography models generally agree at overlapping length scales.

Despite the efforts of regional tomography, constraints on the amplitude of mantle structure smaller than a few hundred kilometers remain equivocal. Residuals of *P* wave travel times suggest that about half of the heterogeneity power is at scales smaller than 300 km [*Gudmundsson et al.*, 1990]. *Su and Dziewonski* [1992], on the other hand, used *SS* residuals to argue that features larger than 2500–3500 km are dominant.

High-frequency scattering studies provide compelling evidence for significant small-scale (1–10 km) heterogeneity throughout the lithosphere and upper mantle [e.g., *Shearer and Earle*, 2004; *Sato et al.*, 2012; *Kennett and Furumura*, 2013]. The depth extent of this heterogeneity, however, is poorly resolved. Global observations of *PKP* precursors suggest that small-scale heterogeneity persists throughout the lower mantle [*Hedlin et al.*, 1997], but recent studies report that these perturbations are likely smaller than those in the lithosphere and upper mantle by an order of magnitude [*Margerin and Nolet*, 2003a; *Mancinelli and Shearer*, 2013].

Taken as a whole, these studies suggest Earth's upper mantle is seismically heterogeneous at length scales spanning at least 4 orders of magnitude, from a few kilometers to tens of thousands of kilometers. However, our knowledge of the complete mantle heterogeneity spectrum is limited in two important ways: (1) We lack key constraints on global mantle heterogeneity power at intermediate scales (5–500 km), and (2) the different parameterizations used in tomography and scattering studies hamper their direct comparison. Addressing the first point will require both new observations and accurate methods for modeling the effects of intermediate-scale structures on the global seismic wavefield at arbitrary periods. Energy-conserving particle-based methods [*Margerin and Nolet*, 2003b; *Shearer and Earle*, 2004] are ideal for modeling global short-period scattering observations because of their low computational cost. Such methods have been validated at high-frequencies by comparison with full-wavefield calculations in the simple case of a homogeneous background medium [*Przybilla et al.*, 2006; *Wegler et al.*, 2006], but it has not yet been shown that these approaches are valid at periods much longer than 1 s, as ray theoretical assumptions are expected to fail away from the infinite frequency limit. Addressing the second point requires relating the power per degree on the surface of a sphere (used in most global tomography studies) with the power spectral density of velocity perturbations in a volume (used in most scattering studies).

Here we attempt to address these issues. Specifically, we (1) model the time, range, and frequency dependence of globally averaged *P* coda amplitudes with a single upper mantle heterogeneity spectrum, (2) confirm the validity of a ray theoretical modeling approach with full-wavefield spectral element calculations, (3) relate our power spectral density function with large-scale constraints obtained from tomography models, and (4) suggest a geodynamically plausible scenario for the generation of the observed upper mantle heterogeneity spectrum.

2. Methods

Our study is motivated by the fact that long-period seismograms often display more energy between the main arrivals than is predicted by global tomography models (Figure 1). Our goal is to determine whether adding intermediate-scale (5–500 km) structure to the tomography models can produce scattering that can explain the excess observed energy. We begin by characterizing the average amplitude of scattered energy in a global data set of seismograms at teleseismic distances. We then attempt to fit these observations by considering models of random mantle heterogeneity and computing the corresponding global wavefield using both a ray theoretical "seismic particle" approach and full spectral element simulations.

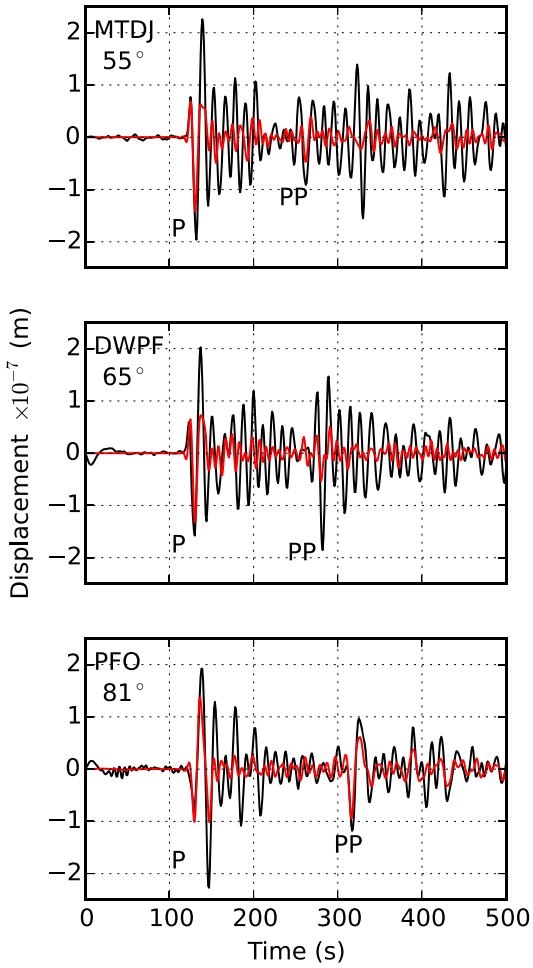

Figure 1. A comparison of recorded (black) and synthetic (red) seismograms for the 9 September 2010 M_w 6.2 earthquake near the coast of Central Chile. Synthetic seismograms were computed as part of the Global Shake Movie project [*Tromp et al.*, 2010] using 3-D velocity model S362ANI [*Kustowski et al.*, 2008]. Both the data and synthetics have been filtered from 20 to 40 s.

2.1. Characterizing Scattered Energy in Data

We use "Standing Order for Data" [*Owens et al.*, 2004] to obtain a global data set of broadband vertical waveforms downsampled to 0.5 Hz from an online archive maintained by the Incorporated Research Institutions for Seismology (IRIS) Data Management Center. Because we are primarily interested in globally averaged properties, we request data recorded at Global Seismic Network stations. We obtain waveforms from 661 normal and thrust events ($M_w \geq 5.7$, depth $\leq$50 km) occurring between 1 January 2010 and 17 July 2013. For reasons discussed later in the text, we exclude waveforms from events with strike-slip focal mechanisms.

In this work, we analyze the vertical component only, our justification being that the bulk of the ray path has to be in *P* or the arrival time would be delayed beyond the *PP* arrival. Although simulations of the wavefield by *Gaebler et al.* [2015] suggest that multiple-scattered shear waves dominate the teleseismic coda at high frequencies (>1 Hz) and long lapse times, in this study we focus on scattered energy at lower frequencies and shorter lapse times. We argue that most of the long-period scattered energy characterized is not diverted far away from its original direction of propagation, as evidenced by the large transport mean free paths, and thus it is unlikely that most of the long-period scattered energy is leaking into the horizontal components.

To measure globally averaged scattered energy as a function of time, range, and frequency, we construct global stacks after applying the following preprocessing steps. First, each waveform is filtered to isolate the frequency content of interest. Second, we assess the quality of each waveform by measuring the peak amplitude of the direct *P* wave pulse relative to the peak amplitude of the "noise" in a window 150 to 50 s prior to the predicted *P* arrival. Waveforms with poor signal to noise ($\leq$10) are rejected. Third, we compute the envelope function of the time series and remove preevent noise power from the entire trace. Fourth, we align the waveforms on the maximum amplitude occurring within $\pm$12 s of the predicted *P* arrival time. Finally, we normalize the entire trace by the peak *P* wave amplitude and stack each point in an appropriate time-distance bin.

Summing thousands of seismograms processed in this way produces a smoothly varying final stack that represents the globally averaged time and range dependence of the scattered amplitudes. We repeat this procedure for a set of bandpass filters (5–10 s, 10–20 s, 20–40 s, and 30–60 s) to assess how the scattered amplitudes depend upon frequency. Since the signal-to-noise check is applied separately at each frequency, the number of waveforms in each stack varies. The 5–10 s stack (near the microseism noise peak) has the fewest waveforms (6371), and the 20–40 s stack has the most waveforms (17,382). Because our focus in this study is on intermediate frequencies, we do not analyze data at periods shorter than 5 s. However, later we will compare our results with the 1 Hz scattering study of *Shearer and Earle* [2004].

The specific design of the filter does not significantly affect the main result, but filters with weak sidelobes in the time domain are preferred because they reduce contamination of the coda waves by energy leakage from the direct pulse. For consistency, it is important to use identical filters (i.e., the same sets of weights and the same algorithm) to process both the observed and synthetic waveforms.

Ideally, our global waveform stack represents Earth's globally averaged properties. However, our incomplete ray coverage, coupled with likely lateral variations in scattering strength, implies some uncertainty in this assumption. To assess this, we divide the observations into about 1600 subsets based upon similar source-receiver geometries and apply the stacking procedure to each subset. The mean of the regionally distinct substacks provides an estimate of the global average. Bootstrap resampling of the collection of substacks gives 2σ uncertainty bounds.

2.2. Modeling I—Monte Carlo Method

Particle-based Monte Carlo methods provide a powerful tool for modeling seismic amplitudes at high frequencies [*Gusev and Abubakirov*, 1987; *Abubakirov and Gusev*, 1990; *Hoshiba*, 1991; *Margerin et al.*, 2000; *Bal and Moscoso*, 2000; *Shearer and Earle*, 2004]. These methods, based on radiative transfer theory [*Chandrasekhar*, 1960], treat the seismic wavefield as a sum of discrete energy particles that individually experience random scattering events based on probabilities computed from the heterogeneity properties of the medium. These approaches naturally account for multiple scattering and geometrical spreading effects in an energy-conserving framework. In this work, we use the Monte Carlo phonon algorithm developed by *Shearer and Earle* [2004].

Initially, we divide the mantle into three concentric scattering volumes. The first volume extends from the Earth's surface to a depth of 200 km, the second from 200 km to 600 km, and the third from 600 km to the core-mantle boundary. The heterogeneity in each volume is statistically parameterized by a von Kármán autocorrelation function (ACF) [*Sato et al.*, 2012, p. 23],

$$R(r) = \frac{\epsilon^2 2^{1-\kappa}}{\Gamma(\kappa)} \left(\frac{r}{a}\right)^{\kappa} K_{\kappa}\left(\frac{r}{a}\right) \tag{1}$$

where r is the magnitude of the lag vector, a is the correlation length, and ϵ is the RMS velocity perturbation. Γ is the gamma function, and K_{κ} is the modified Bessel function of the second kind of order κ. This ACF applies to random media with isotropic properties; hence, it depends only upon the magnitude rather than the magnitude and direction, of the lag vector. The one-dimensional (1-D) Fourier transform of equation (1) gives the 1-D power spectral density function (PSDF)

$$P(m) = \frac{2\pi^{1/2}\Gamma(\kappa + 1/2)\,\epsilon^2 a}{\Gamma(\kappa)\,(1 + a^2 m^2)^{\kappa+1/2}} \tag{2}$$

where m is the magnitude of the wave number vector. Equation (2) resembles a power law with exponent $-2\kappa - 1$ for large wave numbers ($am \gg 1$). The role of κ then is to control the relative proportion of large- to small-scale heterogeneity in this limit.

Before moving forward, a few more assumptions must be made. To reduce the number of free parameters in these models, we assume that the fractional shear wave velocity perturbation field is equal to the fractional P wave perturbation field of the medium and that the fractional density perturbation field is proportional to the fractional velocity perturbation field scaled by a factor of 80%. We assume that the P velocity to S velocity ratio equals $\sqrt{3}$, which is true for a Poisson solid. Altering these parameters affects the directional scattering probabilities, as well as the probability of phase conversions between P and S. For example, increasing the relative density perturbations increases the amount of backscatter [*Sato et al.*, 2012, p. 142].

We assume an intrinsic attenuation model that is identical to the one favored by *Shearer and Earle* [2004], who noted that larger Q values provide better fits to high-frequency coda observations, possibly because Q models often include the effects of both scattering and intrinsic attenuation. Above a depth of 220 km, $Q_{\alpha} = 450$; beneath 220 km, $Q_{\alpha} = 2500$. For the attenuation of shear energy, we assume that $Q_{\beta} = (4/9)Q_{\alpha}$.

The code tracks individual seismic energy particles (termed *phonons*) sprayed over a discrete range of takeoff angles as they travel along trajectories obtained from a large table of precomputed ray paths for the 1-D reference model. Upon entering a scattering volume, a phonon is assigned a random path length—an instance of an exponentially distributed random variable whose mean is determined by the heterogeneity properties

of the medium. Once the phonon travels its assigned distance it scatters in a random direction, where the directional probability is also governed by the heterogeneity properties. At this point, another random path length is generated from the distribution and the phonon travels onward. When the phonon returns to the free surface, its energy is summed in a time-distance array, and the phonon is reflected downward. As this process is repeated for a large number of phonons, the time-distance array continues to populate, eventually converging to a smooth model of the seismic energy field. This method also handles *P*-to-*S* and *S*-to-*P* conversions, and reflection/transmission effects at interfaces using a probabilistic framework. In these models, the phonons are sprayed from an isotropic source at the free surface. The *S*/*P* energy ratio is held constant at 23.4, based on theoretical results for a double-couple source averaged over all directions [*Sato*, 1984]. For more details about this method, the reader is directed to *Shearer and Earle* [2004] and *Shearer and Earle* [2008].

This phonon method has been used previously to model scattered phases at frequencies near 1 Hz. In this study, however, we aim to model scattered waves at periods as long as 45 s. As a final step one must account for the broadening effects of the long-period filters used to process the waveforms. To this end, we compute the envelope function of the filter and convolve it, in power, with the output of the phonon code. In the following section, we validate our use of the phonon method at long periods by comparing its results against spectral element calculations.

2.3. Modeling II—Spectral Element Simulations

Recent advances in computing power have made it possible to accurately simulate seismic wave propagation in 3-D heterogeneous Earth models with full anisotropy, attenuation, and complex 3-D geometry [*Komatitsch and Tromp*, 2002a, 2002b; *Chaljub et al.*, 2007; *Tromp et al.*, 2008]. In particular, the SPECFEM3D_GLOBE (SEM, for short) software package [e.g., *Komatitsch et al.*, 2002] generates synthetic seismograms of global seismic waves accurate for periods from 17 to 500 s provided that one has access to a modest computer cluster [*Tromp et al.*, 2010]. In our experience, synthetics accurate to 17 s can be computed in about 2 h of computing time on 96 processors. We run simulations on parallel Xeon E5-2680 processors from the Texas Advanced Computing Center Stampede system. By contrast, the particle-based modeling approach often converges to a stable solution within a few hours of computing time on a single processor, but it sacrifices phase information, generating only seismic energy envelopes. SEM calculations, on the other hand, produce complete waveforms. Another disadvantage of the phonon approach is that it is based on geometrical ray theory and thus neglects finite frequency effects that become increasingly relevant at long periods.

In this section, we compare results of the phonon method and SEM. First, we explain how to generate random realizations of layered mantle heterogeneity in 3-D coordinates and discuss how to interpolate this realization to the SEM mesh. Second, we show our source-receiver distribution for the test. Finally, we compare stacks of SEM synthetics with output from the phonon code and discuss where the two models agree and disagree.

To generate random earth models, we begin in the wave number domain. For each element in a $501 \times 501 \times 501$ array, we set the Fourier amplitude to a value defined by the PSDF of choice and randomize the phase component. Transforming the array into the spatial domain, we obtain a cube of random media. Assuming that a side of this cube is equal to two times the radius of Earth, each voxel has a dimension of $25 \times 25 \times 25$ km. These models, then, can roughly capture heterogeneity as small as 50 km. See Appendix A for details regarding the generation of these blocks of random media.

To include depth-varying heterogeneity in our models, we generate multiple cubes of random media, each with different heterogeneity spectra. We patch together concentric spherical shells from the different cubes to obtain a model that has depth-varying properties. To avoid the sharp velocity changes created by the boundary between two different random heterogeneity models, the transition areas between the spherical shells are smoothed by removing the power above a cutoff wave number in the Fourier domain (Figure 2). Figure 3 shows a slice through the equator of an example random 3-D Earth model with a von Kármán PSDF. We use a trilinear interpolation scheme to assign values from our evenly spaced random model to points on the SEM mesh, whose spacing generally varies with depth.

The setup for the SEM simulation is illustrated by Figure 4. To obtain many independent measurements per simulation, we place 26,484 virtual sensors at distances from 50 to 130°from the source, at all azimuths. We place a virtual source at the equator and experiment with different source radiation patterns. After each simulation, the synthetics are processed and stacked using the same code that processes the real data.

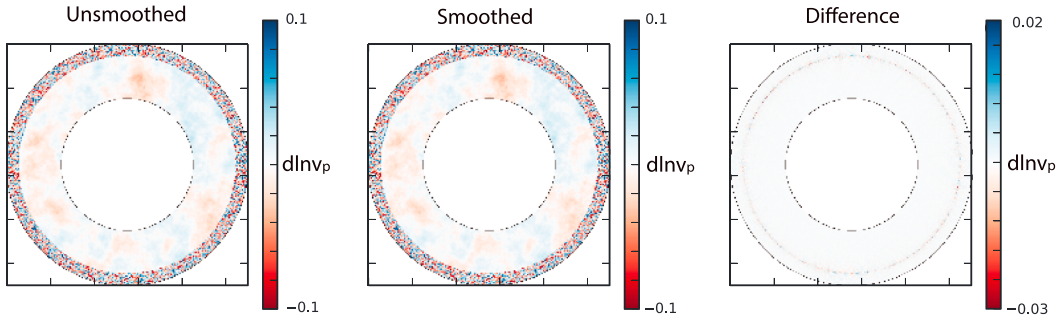

Figure 2. Illustration of the smoothing procedure. This random heterogeneity model has strong small-scale heterogeneity in the upper mantle and weak large-scale heterogeneity in the lower mantle. The transition between the two layers has been smoothed in the wave number domain by suppressing all structure with a wavelength smaller than 60 km.

The SEM mesh contains 256 spectral elements along each side of the six "chunks" in the cubed sphere. Benchmarks against semianalytical normal-mode synthetic seismograms show that this mesh is accurate to periods of ~17s [*Komatitsch and Tromp*, 2002a, 2002b].

Output from the two methods is compared in Figure 5. We find good agreement between the two methods for the scattered energy between *P* and *PP* at ranges from 50 to 96°. The agreement is best for isotropic, normal, and thrust sources. For strike-slip sources, however, the SEM synthetics have codas which are much more energetic than those modeled by the phonon code. This is likely due to the radiation pattern. For strike-slip events, the steep takeoff angles of teleseismic *P* waves means that they originate near a node in the *P* wave radiation pattern. Scattering processes redirect energy radiated from other parts of the focal sphere into the *P* nodal plane; this results in high-amplitude codas relative to the direct phase. In previous works [e.g., *Shearer and Earle*, 2004], it was implicitly assumed that stacking seismograms from many events would result in a net isotropic source, but here we find that including strike-slip events may bias results. Thus, we exclude strike-slip events from this study.

Although the methods agree regarding the energy between *P* and *PP*, there was a marked disagreement on the amplitude of the *PKP* phase (not shown) in the 20–40 s band. The source of the discrepancy is likely due to destructive interference between PKP_{cd} and PKP_{df}. If this effect persists at higher frequencies, lower mantle

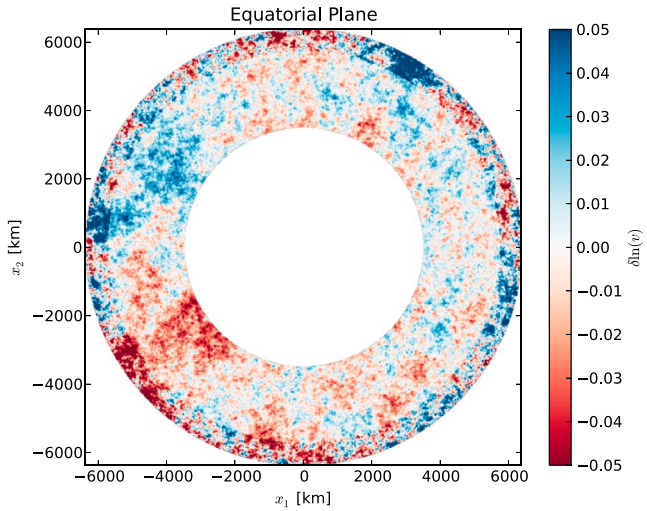

Figure 3. Cross section through an example 3-D mantle model with random heterogeneity. The mantle is divided into two layers. The upper mantle—600 km thick—has stronger heterogeneity than the lower mantle. The heterogeneity in both the upper and lower mantle layers are random realizations of a Von Kármán autocorrelation function. Note that this figure should be viewed merely as an illustration of our method; comparisons with data suggest that lower mantle heterogeneity is much weaker than is shown here.

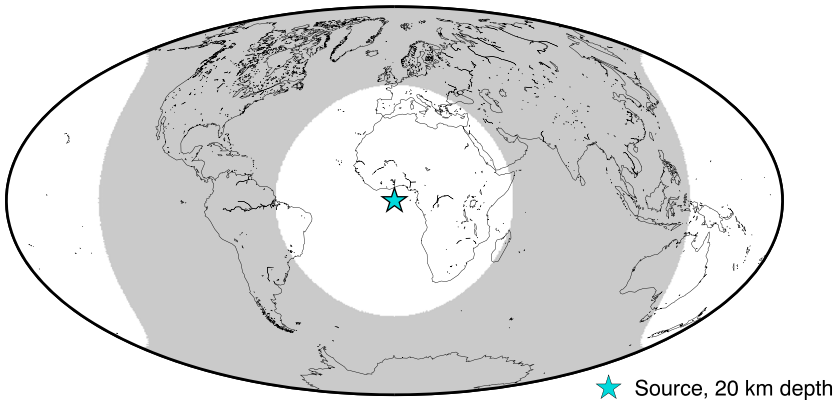

Figure 4. SEM model setup. The 26,484 receivers are located in the gray region (50–130° range) at ~1° spacing. The source is placed at the intersection of the equator and the prime meridian, at a depth of 20 km.

heterogeneity constraints from *PKP* precursors should be revisited. However, such an effort would require SEM simulations accurate to 1 Hz, which is well beyond the scope of this study.

3. Results

Here we describe the main features of the data stacks, and the predictions of our preferred model, as plotted in Figure 6. In addition to the direct *P* arrival, surface multiples *PP* and *PPP* are clearly visible. At long ranges, the relative amplitude of *PP* begins to increase. This is mainly due to the decrease in the amplitude of *P* approaching the onset of the shadow zone at 98°. At longer periods, the pulse width of the main arrival broadens since the width of the time domain pulse is proportional to the reciprocal of the bandwidth (in Hertz). Our global stacks show significant energy between the main phases, most notably between *P* and *PP*, which we interpret to be scattered from volumetric heterogeneity within the mantle. This scattered energy is most notable at high frequencies, but is distinctly nonzero even at the longest observed period (30–60 s).

Using the phonon method, we test the sensitivity of the synthetics to heterogeneity at various depths within the mantle. Ultimately, we find that our data do not possess the resolution to uniquely place heterogeneity in either the lithosphere or upper mantle, so henceforth, we treat the top 600 km as a single layer, although we will discuss how the assumed thickness of this layer affects results at the end of this section. We also find that scattered energy from the lower mantle is not needed to fit these data. In fact, if one makes the lower mantle as heterogeneous as the upper mantle, our synthetics predict that more scattered energy would be seen at long ranges than is observed. This is consistent with the notion that the lower mantle is much less heterogeneous than the upper mantle and lithosphere. Adding the weak amount of lower mantle scattering

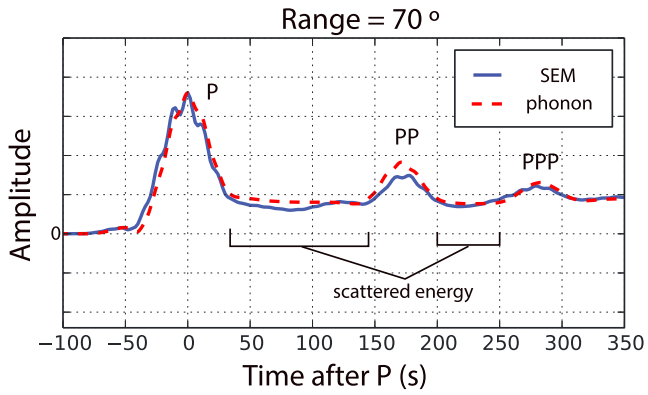

Figure 5. A comparison between the phonon and SEM methods for *P* coda. The SEM synthetics were computed for a thrust source. They were filtered from 20 to 40 s before stacking. The phonon code was run at 30 s. The output was convolved, in power, with the filter wavelet used to postprocess the SEM synthetics.

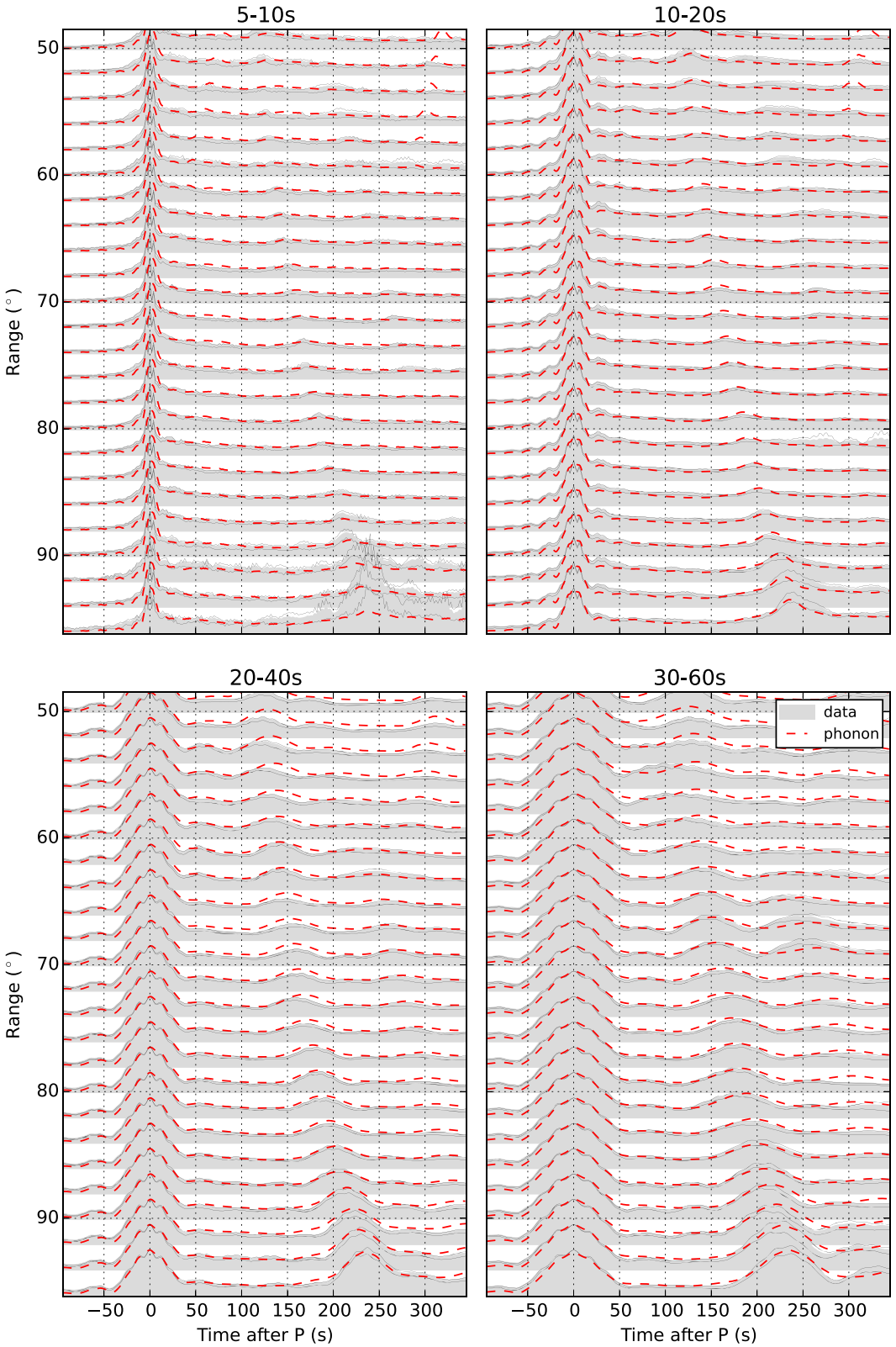

Figure 6. Time and range dependence of *P* coda filtered from (top left) 5 to 10 s, (top right) 10 to 20 s, (bottom left) 20 to 40 s, (bottom right) and 30 to 60 s. Data are plotted in gray; 95% confidence bounds obtained by regional bootstrapping are shown by thin black curves. The amplitudes predicted by the phonon code for the preferred random heterogeneity model are shown by dashed red curves. The phonon models were run at 7.5 s (Figure 6, top left), 15 s (Figure 6, top right), 30 s (Figure 6, bottom left), and 45 s (Figure 6, bottom right).

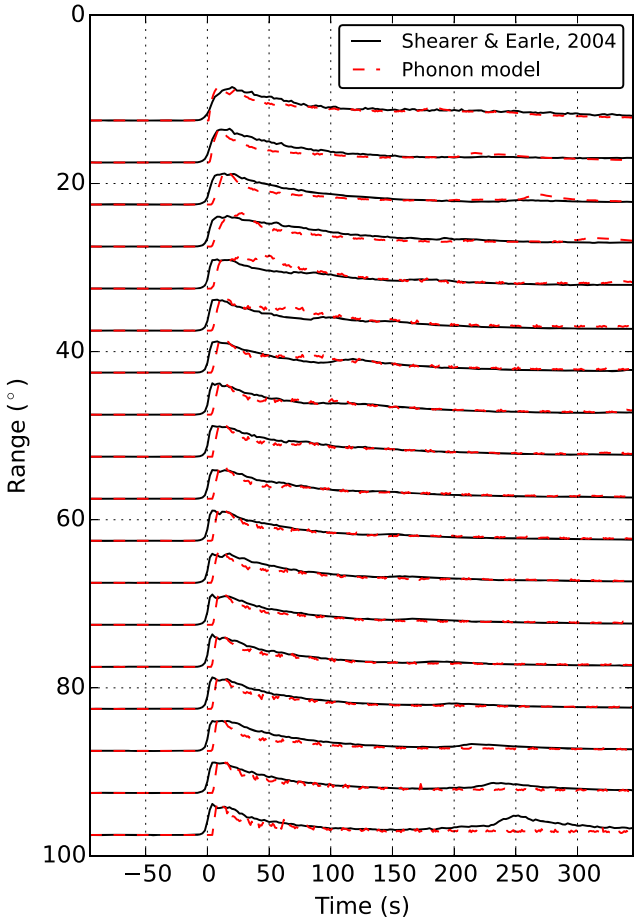

Figure 7. Fits of our preferred model (dashed red) to the 1 Hz data stacks published by *Shearer and Earle* [2004]. The phonon code is run at 1 s.

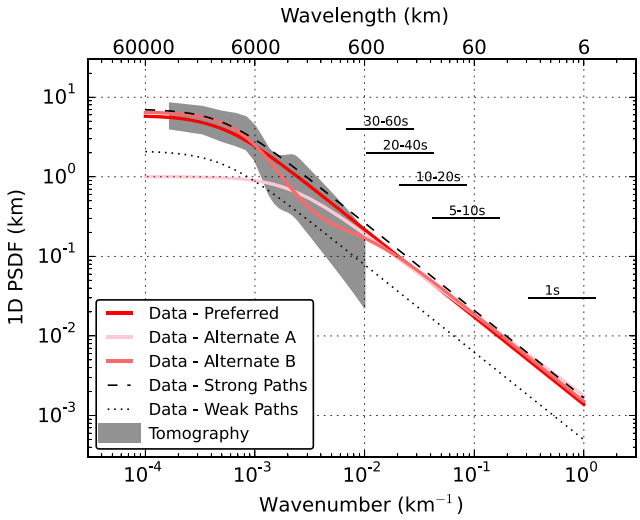

Figure 8. The preferred heterogeneity spectrum based on fits to scattering observations and tomography results. Two alternate spectra are plotted that fit the scattering observations almost as well. Bounds for the preferred spectrum are shown by black dashed lines. Each period of observation constrains the spectrum for a range of wave numbers; these are shown by the black horizontal bars. The bounds on the tomography results are derived from the variability in tomography spectra as published by *Meschede and Romanowicz* [2015a].

Table 1. Mean Free Paths (in km) for *P* Energy (l_P) and *S* Energy (l_S) for the Preferred Heterogeneity Model[a]

	Lithosphere (0–200 km)				Upper Mantle (200–600 km)			
Period (s)	l_P	l_S	l_P^t	l_S^t	l_P	l_S	l_P^t	l_S^t
1	11	7.2	370	330	12	7.8	410	360
7.5	67	44	2,300	2,000	77	48	2,500	2,200
15	130	83	4,300	3,700	150	91	4,600	4,100
30	270	160	8,000	7,000	310	180	8,700	7,600
45	450	240	12,000	10,000	520	270	13,000	11,000

[a]The superscript *t* denotes the transport mean free path.

required by *PKP* precursor results [*Margerin and Nolet*, 2003a; *Mancinelli and Shearer*, 2013] has a negligible effect on our results; thus, for simplicity in this study we set the RMS perturbation of the lower mantle to zero and focus on constraining the upper mantle spectrum.

We find that a von Kármán PSDF (equation (2)) with a correlation length of 2000 km, an RMS perturbation of 10%, and $\kappa = 0.05$ provides a good first-order fit to the observations at all periods. Moreover, this PSDF provides a reasonable fit to the 1 Hz *P* coda stacks previously published by *Shearer and Earle* [2004] (Figure 7) and overlaps with long-wavelength constraints from tomography (Figure 8). The tomography constraints are derived from spherical harmonic spectra of the uppermost 250 km of the mantle published by *Meschede and Romanowicz* [2015a]. This derivation can be found in Appendix B.

Here are a few observations regarding the fitting process. Increasing κ increases the relative proportion of large- to small-scale heterogeneity, thereby affecting the frequency dependence of the synthetics. Values of κ close to zero fit the data well, but increasing κ much beyond 0.05 impedes our ability to fit the frequency dependence. Our preferred model has $\kappa = 0.05$ (rather than $\kappa = 0.01$) because the increased slope provides better agreement with the tomography results. Adjusting the correlation length has little effect on the synthetics, provided it remains greater than about 500 km. If the correlation length drops below this value, fits at the longest periods begin to degrade. Increasing the RMS perturbation increases the amplitude of scattered energy but not necessarily in a linear fashion. These models generally involve multiple scattering, having mean free paths ranging from about 7 to 500 km for our preferred spectrum (exact values depend on frequency and depth, see Table 1). Due to our large choice of a (which is required to match the large-scale structure in tomography), most scattering events do not redirect energy far from the original direction of propagation. In media where forward scattering is dominant, a more appropriate quantity for comparison is the transport (i.e., momentum transfer) mean free path [e.g., *Przybilla et al.*, 2009], which corrects for anisotropic scattering probabilities. Our preferred phonon models have transport mean free paths for *P* energy in the upper mantle of about 2500, 4600, 8700, and 13,000 km for periods of 7.5, 15, 30, and 45 s, respectively. At 1 Hz, the transport mean free path is ~410 km.

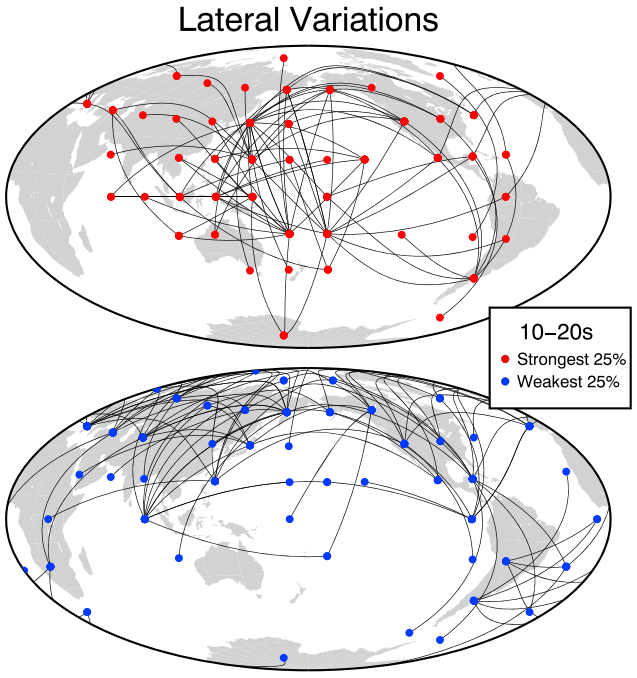

Figure 9. Lateral variations in *P* coda amplitude when filtered from 10 to 20 s. Data are grouped based on their ray endpoints (both source and receiver sides). For each populated "bin" we show the associated source cell and receiver cell (colored circles) connected by a black curve. The statistical significance of these variations is confirmed by bootstrap resampling the measurements in each bin.

An alternate spectrum (Alternate A) with $a = 500$, $\epsilon = 18\%$, and $\kappa = 0.01$ fits the scattering observations just as well but does not overlap with the tomography constraints at small wave numbers. More complicated spectra (e.g., Alternate B) fit the scattering observations well and plot within the

Table 2. Mean Free Paths (in km) for *P* Energy (l_P) and *S* Energy (l_S) for the Heterogeneity Model That Fits the Subset of Data With the Strongest 25% of Coda Amplitudes[a]

	Lithosphere (0–200 km)				Upper Mantle (200–600 km)			
Period (s)	l_P	l_S	l_P^t	l_S^t	l_P	l_S	l_P^t	l_S^t
1	9	5.9	310	270	10	6.5	330	290
7.5	58	36	1,900	1,700	63	40	2,100	1,800
15	110	69	3,500	3,100	120	75	3,800	3,400
30	230	130	6,600	5,800	250	150	7,200	6,300
45	370	200	9,600	8,300	430	220	10,000	9,100

[a]The superscript *t* denotes the transport mean free path.

bounds allowed by tomography. At present, it is difficult to distinguish among these more complicated heterogeneity spectra, but comparisons with phase velocity maps may improve constraints at long wavelengths. It may seem surprising that ϵ changes quite dramatically for seemingly small changes in κ. As κ approaches zero increasingly more heterogeneity is added at infinitesimal scales. Despite their differing ϵ values, Alternate A and the preferred spectrum plot almost on top of each other in Figure 8.

We investigate the possibility of lateral variations in scattering strength in our observations by employing the following procedure. For each data trace, we measure the mean amplitude of the *P* coda in a window from 50 to 100 s after the peak *P* amplitude. The measurements are then divided into 1637 groups based on ray-path endpoints. In each regional group, bootstrap resampling with replacement is applied to determine the mean and variance of the coda amplitude. After discarding the groups with few (<5) or wildly variable (error bars wider than 15% of the reference *P* wave amplitude) measurements, we sort the remaining ~350 groups from smallest to largest mean amplitude, and plot their spatial distribution.

We find significant lateral variations of the scattered energy at all periods of observation except the longest (30–60 s). Paths traversing the continents tend to result in weaker-than-average codas, whereas paths traversing the oceans (in particular subduction zones) tend to display the opposite effect. This pattern is shown, for example, in the 10–20 s band of observation (Figure 9). Similar patterns are observed for the 5–10 s and 20–40 s bands.

To fit the subset of data from the 25% of the paths with the smallest scattered amplitudes, ϵ of the preferred spectrum needs to be reduced from 10% to 6%. Likewise, fitting the subset of data from the largest 25%, ϵ needs to be increased only slightly, from 10% to 11%. Spectra that match the regional substacks displaying either the strongest or weakest coda amplitudes are plotted in Figure 8, and the mean free paths for these models are given in Tables 2 and 3.

Before moving on, we should note the linear trade-off between ϵ and the assumed thickness of the upper mantle scattering layer. That is, halving the assumed thickness of the scattering layer requires a doubling of ϵ in order to achieve similar fits. This is only true down to a certain thickness. Below a thickness of 100 km, the synthetics are unable to match the observed time dependence for any ϵ. Therefore, we argue that the observed scattering cannot come from crustal heterogeneities alone.

4. Discussion

In the previous section we proposed a single mantle heterogeneity spectrum that simultaneously explains, to first order, (1) the long-period constraints presented in this study, (2) the high-frequency constraints reported previously [*Shearer and Earle*, 2004], (3) and constraints from global and regional tomography. Here we discuss geodynamically plausible scenarios for the generation of the heterogeneity observed at scales from 5 to 800 km.

Table 3. Mean Free Paths (in km) for *P* Energy (l_P) and *S* Energy (l_S) for the Heterogeneity Model That Fits the Subset of Data With the Weakest 25% of Coda Amplitudes[a]

	Lithosphere (0–200 km)				Upper Mantle (200–600 km)			
Period (s)	l_P	l_S	l_P^t	l_S^t	l_P	l_S	l_P^t	l_S^t
1	31	20	1,000	900	34	22	1,100	990
7.5	190	120	6,300	5,500	210	130	6,900	6,100
15	370	230	12,000	10,000	410	250	13,000	11,000
30	760	450	22,000	19,000	850	490	24,000	21,000
45	1,300	680	32,000	28,000	1,400	760	35,000	31,000

[a]The superscript *t* denotes the transport mean free path.

Nearly 30 years ago, it was suggested that the upper mantle might contain elongated strips of subducted lithosphere that have been stretched and folded to small scales by convective strains [*Allègre and Turcotte*, 1986].

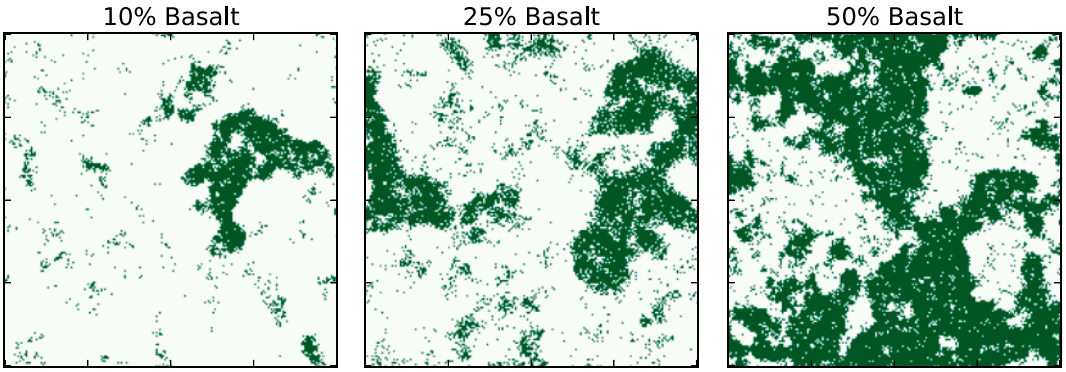

Figure 10. Illustrations of random phase two-component mixtures of basalt (dark green) and harzburgite (light green). The mixtures are self-similar so any linear length-scale may be applied to the axes. These were generated from a starting model by randomly swapping pixels until the desired self-similar spectrum was achieved.

This so-called marble cake mantle contains heterogeneity at many scales, likely possessing a power law spectrum with an exponent of −1 (i.e., $\kappa \rightarrow 0$ in equation (2)) [*Batchelor*, 1959; *Antonsen and Ott*, 1991; *Agranier et al.*, 2005; *Ricard et al.*, 2014]. The two primary components of this mixture would be basalt and harzburgite. (Basalt undergoes a phase change to eclogite at depth. For the purposes of this paper, "basalt" refers to both of these phases.) Although the seismic velocities of these compositions vary with depth, they differ from each other by about 10% when averaged over the uppermost 800 km of the mantle [*Xu et al.*, 2008]. Figure 10 illustrates the textures associated with self-similar (i.e., $\kappa \rightarrow 0$ in equation (2)) mixtures of basalt and harzburgite.

In Figure 11, we plot spectra for various basalt-harzburgite mixtures from the relation

$$P(m) = \frac{\pi \left(d \ln v\right)^2 f_b \left(1 - f_b\right)}{\ln\left(\lambda_2/\lambda_1\right) m} \tag{3}$$

where $d \ln v$ is the depth-averaged fractional velocity difference between the two compositions, f_b is the basalt fraction, and wavelengths λ_1 and λ_2 give the lower and upper bounds on the scale range of mixing. First, assuming that mixing occurs only at wavelengths between 5 and 800 km, we vary the basalt fraction of the mantle (10%, 25%, and 50%). Unsurprisingly, the 50–50 mixture produces the greatest RMS scattering, but it is not large enough to match the preferred spectrum for the paths with the strongest codas. Second, we vary the minimum wavelength at which mixing occurs, assuming that the mantle is 50% basalt. We find that mixing down to the centimeter or meter scale significantly reduces the RMS of the spectrum. If a basalt-harzburgite mixture is indeed responsible for the scattered energy in *P* coda, this result implies that the mixing does not occur at scales much smaller than a kilometer. This is contrary to the notion that a marble cake texture persists to the centimeter scale, at which point chemical diffusion homogenizes the mantle [*Kellogg and Turcotte*, 1987].

Another factor that may affect the velocity contrast between basalt and harzburgite is the presence of water. If basalt and harzburgite are fully hydrated, the contrast may increase from 10% to 20% or even 30% in the uppermost 185 km of the mantle in the vicinity of low-temperature subduction zones [*Hacker*, 2003]. Such an effect may contribute to the lateral variations in scattering strength discussed in the previous section.

Earlier we noted a trade-off between the thickness of the assumed scattering layer and the value of ϵ required to match the amplitude of the scattered energy. Figure 12 shows the seismic velocity contrast between two components that would be required to match the observations. A scattering layer 100 km thick, for example, would require a velocity contrast of 45% between compositions. More reasonable velocity contrasts are compatible with thicker scattering layers, such as the 600 km thick one we have assumed throughout much of this study.

Of course, it is possible that scattering is stronger in the crust and lithosphere than in the underlying upper mantle due to compositional heterogeneity with stronger velocity contrasts than our assumed 10%. The presence of such a strong shallow scattering zone would reduce the scattering required at deeper depths in our model, but unfortunately, we lack the depth resolution to test this idea.

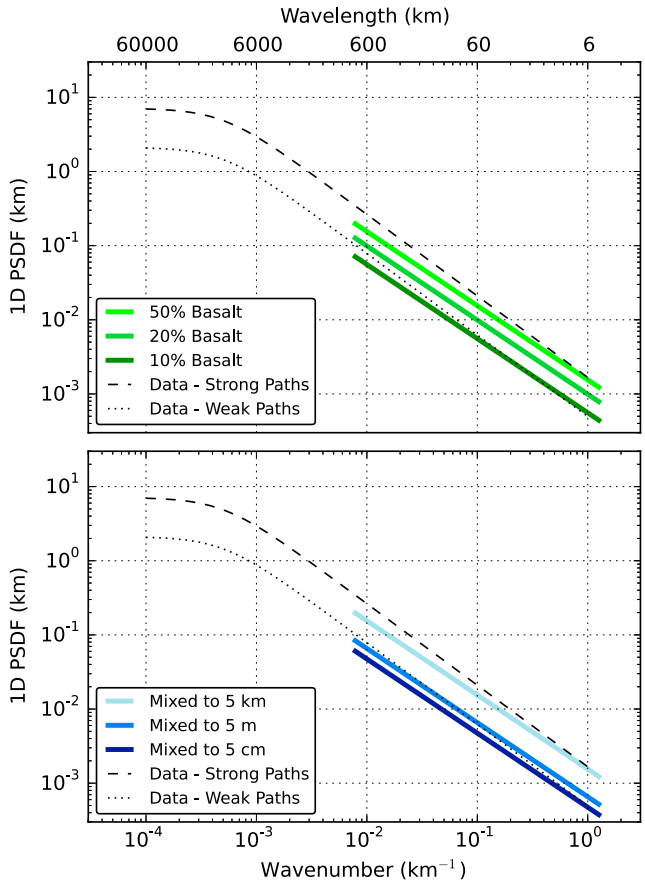

Figure 11. (top) Predicted spectra for a marble cake mixture of basalt and harzburgite. The net power of the spectrum is affected by changing the basalt fraction of the mixture. These models assume mixing occurs primarily at wavelengths between 5 and 800 km. (bottom) Predicted spectra for a marble cake mixture of basalt (50%) and harzburgite (50%). The power of the spectrum over the range of observations (5–800 km) is reduced by allowing mixing to occur at smaller scales.

We can, however, place this result within the context of the numerous studies that have used high-frequency *S* wave envelopes to estimate the scattering properties of the lithosphere. At 1 Hz, regional studies typically suggest lithospheric mean free paths around 150 km [e.g., Japan, *Hoshiba*, 1993], although depending on the region this value can be as low as 28 km (northern Greece [*Hatzidimitriou*, 1994]) and as high as 1000 km (Germany [*Sens-Schönfelder and Wegler*, 2006]). Our preferred spectrum has a transport mean free path of 330 km for 1 Hz *S* energy in the lithosphere, which is within the range of previously reported values. Our spectrum also predicts an increase in g_0 (reciprocal of the mean free path) with increasing frequency between 1 and 10 Hz. While some regional *S* coda studies [e.g., *Fehler et al.*, 1992; *Przybilla et al.*, 2009] agree with this trend, a majority show a constant or decreasing g_0 with frequency [e.g., *Hoshiba*, 1993; *Giampiccolo et al.*, 2006].

Another important question is whether scattering in an anisotropic random medium plays an important role in this problem. *Kawakatsu et al.* [2009] proposed horizontal melt-rich bands beneath the lithosphere-asthenosphere boundary; whereas *Kennett and Furumura* [2013] invoked horizontally elongated heterogeneities throughout the lithosphere and asthenosphere to model the codas of *Po* and *So*. The presence of intrinsic anisotropy in seismic wave speed may also be relevant. Varying orientations of olivine might provide a viable way to scatter enough energy, as the velocity difference between the fast and slow axes of olivine is in the order of 10%.

Thus far, we have focused on how compositional variations might shape the observed spectrum, but lateral variations in temperature also play a role, especially at long wavelengths. Large-scale variations in the uppermost mantle are dominated by the thermal differences between continents and oceans, i.e., warm, seismically

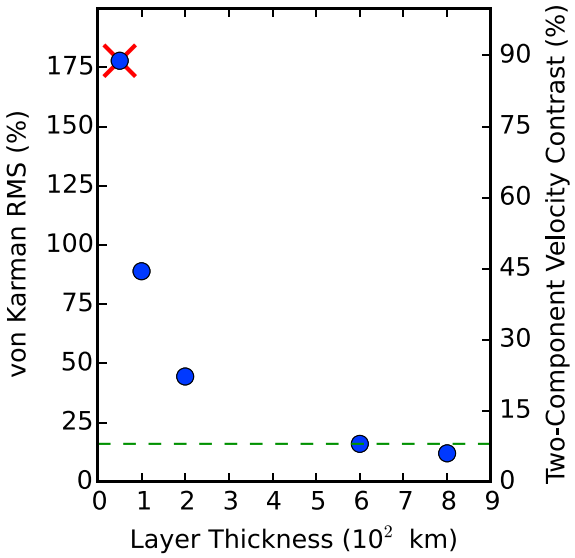

Figure 12. Trade-offs between the thickness of the scattering layer and the required RMS velocity perturbation. The thinnest layer tested (denoted by a red *x*) did not fit the data as well as the thicker layers. Assuming a two-component mixture with self-similar structure from 5 to 800 km, the equivalent velocity contrast between the two components can be calculated. The dashed green line represents the average velocity contrast between basalt and harzburgite throughout the upper mantle.

slow spreading ridges versus cold, seismically fast cratons. At scales of ~10 km, the heterogeneity is most likely compositional because small-scale temperature perturbations quickly diffuse away. Thus, the intermediate scales may represent a transition region where both thermal and compositional variations give significant contributions to observed perturbations in seismic velocity. This, perhaps, is an overly simplistic view, as it has been argued that large-scale variations in both temperature and composition are required to satisfy gravity observations [*Deschamps et al.*, 2002].

In the future, extending these results to shorter ranges may help resolve the depth extent of scattering. Modeling scattered regional phases such as *Pn* coda [e.g., *Tittgemeyer et al.*, 1996] may facilitate comparisons with local and regional coda studies and improve constraints on heterogeneity in the uppermost mantle.

Scattering of surface wave energy may also contribute to our understanding of intermediate-scale structure in the upper mantle. Incoherent long-period (~100 s) energy following the fundamental-mode surface waves can be modeled by invoking self-similar small-scale velocity perturbations throughout the mantle [*Meschede and Romanowicz*, 2015b]. More events should be analyzed in order to obtain quantitative global constraints, as these will be important in confirming or refining our proposed spectrum.

Appendix A: Generating Cubes of Random Media

We take the following approach (following *Hoshiba* [2000]) to generate global random media models. The fractional velocity fluctuation of the random medium $\xi(\mathbf{x})$ by definition has zero mean and RMS perturbation ϵ. Let $n(\mathbf{x})$ be the discretized version of the fractional velocity fluctuation with discretization intervals Δ_x, Δ_y, and Δ_z along the respective *x*, *y*, and *z* axes, and N_x, N_y, and N_z are the total number of discretization points in these orthogonal directions.

The random fluctuations $n(\mathbf{x})$ can be generated by taking the three-dimensional inverse fast Fourier transform (FFT) of a function whose amplitude spectrum is the square root of the discretized PSDF and whose phase $\theta(\mathbf{m})$ is randomly generated from a uniform probability density in $[0, 2\pi)$ and satisfies $\theta(-\mathbf{m}) = -\theta(\mathbf{m})$:

$$n(\mathbf{x}) = F^{-1}\left[P(|\mathbf{m}|)^{1/2}\left(N_x\Delta_x N_y\Delta_y N_z\Delta_z\right)^{1/2} e^{i\theta(\mathbf{m})}\right]. \tag{A1}$$

Limitations imposed by computer memory make it difficult to span many scale lengths in these models. For models shown in this work, Δ_x is typically around 25 km, meaning that power at wavelengths smaller than about 50 km cannot be represented in the discretized form. The RMS perturbations in these models will always be less than or equal to the ϵ used to parameterize the continuous PSDF defined on $(-\infty, \infty)$. If the PSDF decays away rapidly at high wave numbers (as is the case for a Gaussian PSDF), then

$$\epsilon^2 = \frac{1}{N_x N_y N_z}\sum_{i,j,k} n_{ijk}^2 \tag{A2}$$

holds true. We use this relation to verify the accuracy of our FFT routine and normalization factors for the case of Gaussian random media. We can then be confident that our algorithm is correct for arbitrary random media parameterizations, including those that have a considerable fraction of heterogeneity power at subgrid length scales.

Appendix B: Comparison With Tomography

We wish to relate our scattering results to long-wavelength heterogeneity power that has been constrained by global and regional inversions of seismic data. *Meschede and Romanowicz* [2015a] conduct a comprehensive analysis of global and regional tomography models, providing a constraint on heterogeneity power in the upper mantle down to scales of about 300 km (i.e., up to degree 128). *Meschede and Romanowicz* [2015a] present their results in units of logarithmic density, or power per octave, $s(l)$, which is a function of spherical harmonic degree, l. As it is not immediately obvious how to relate $s(l)$ with the power spectral density function that is constrained by our scattering observations, we dedicate this section to defining the different measures of heterogeneity power so that we can reliably compare one metric with another.

B1. Spherical Spectra

Following the notation used by *Meschede and Romanowicz* [2015a], any square integrable function on a sphere, $f(\mathbf{r})$, can be represented by a weighted sum of the spherical harmonic functions $Y_l^m(\mathbf{r})$, with degree l and angular order m:

$$f(\mathbf{r}) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} c_{lm} Y_l^m(\mathbf{r}) \tag{B1}$$

with real valued c_{lm} coefficients. The spherical harmonic functions are given by

$$Y_l^m(\mathbf{r}) = \sqrt{(2l-1)\frac{(l-m)!}{(l+m)!}} P_{lm}(\cos\theta) e^{im\phi} \tag{B2}$$

where P_{lm} is the associated Legendre function of degree l and order m, θ is the polar angle, and ϕ is the azimuthal angle. These basis functions, in addition to being orthogonal and complete, posses unit power normalization:

$$\frac{1}{4\pi} \int_{\Omega} Y_l^m(\mathbf{r}) Y_n^{k*}(\mathbf{r})\, \mathrm{d}\Omega = \delta_{ln}\delta_{mk}. \tag{B3}$$

Here Ω denotes the surface of the unit sphere. The mean square velocity perturbation of the model can be obtained directly from the c_{lm} coefficients via Parseval's Theorem

$$\frac{1}{4\pi} \int_{\Omega} f(\mathbf{r})^2\, \mathrm{d}\Omega = \sum_l \sum_m |c_{lm}|^2 = \sum_l S_l \tag{B4}$$

where $S_l = \sum_m |c_{lm}|^2$ is the power per degree. *Meschede and Romanowicz* [2015a] present their results in power per octave, $s(l)$, given by

$$s(l) = \log_{10}(2) l S_l \tag{B5}$$

which has the desirable property that for self-similar media $s(l)$ plots as a constant function. Converting from power per octave to power per degree is straightforward using equation (B5), but before relating power per degree to power spectral density, we will precisely define the power spectral density function and discuss how the choice of dimensionality affects results.

B2. Power Spectral Density

For functions defined on $\mathbb{R}^N$, the power spectral density function (PSD or PSDF) is the common measure of how a function's power is distributed across a range of frequencies. The PSD of a time series with units U, to take a common example from geophysics, has units of power per Hertz (U^2 s). In this work, we are concerned with variations of an elastic medium in space, so the units of these spectra will be in either power per inverse distance (U^2 km) or power per inverse area (U^2 km^2) or power per inverse volume (U^2 km^3). When concerned with fractional fluctuations of a property, as is our case, U is dimensionless.

We assume that the *P* velocity α and the shear velocity β exhibit equal fractional fluctuations ξ. The fluctuation in density, ρ, is assumed to be proportional to ξ by a scaling factor ν:

$$\xi = d \ln \alpha = d \ln \beta = \nu \, d \ln \rho. \tag{B6}$$

Birch's Law suggests that $\nu = 0.8$ is an appropriate value for the lithosphere, but smaller values may be appropriate for the deeper mantle. The autocorrelation function of our random medium is given by

$$R(\mathbf{x}) = \langle \xi(\mathbf{y}) \xi(\mathbf{y} + \mathbf{x}) \rangle \tag{B7}$$

where the brackets denote an ensemble average. This function depends on only the lag vector **x** if the statistics are stationary with respect to space. Moreover, if the statistics are isotropic (i.e., the perturbations have an average aspect ratio of unity), *R* depends only on the magnitude of the lag vector, $x \equiv ||\mathbf{x}||$.

Sato et al. [2012, p. 20] define the power spectral density function (PSDF) as the three-dimensional Fourier transform of a function $R(x)$:

$$P(\mathbf{m}) = P(m) = \iiint_{-\infty}^{\infty} R(x) e^{-i\mathbf{m}\cdot\mathbf{x}} \, \mathrm{d}\mathbf{x} \tag{B8}$$

where **m** is the (angular) wave number vector and $m \equiv ||\mathbf{m}||$. This function is important because it directly relates to the scattering properties of the medium. To our knowledge, there is no current theory that relates heterogeneity parameterized in terms of spherical harmonics to scattering properties.

In this work, we are dealing only with isotropic, stationary media, so we will be presenting power spectral densities computed for one dimension. One can convert an *N*-dimensional PSDF to an *M*-dimensional PSDF by using the following relations:

$$R(x) = \frac{1}{(2\pi)^N} \int P_{ND}(m) e^{i\mathbf{m}\cdot\mathbf{x}} \, \mathrm{d}^N\mathbf{m} \tag{B9}$$

$$P_{MD}(m) = \int R(x) e^{-i\mathbf{m}'\cdot\mathbf{x}} \, \mathrm{d}^M\mathbf{x} \tag{B10}$$

where the integrals are over $\mathbb{R}^N$ and $\mathbb{R}^M$, respectively. In the second relation, **m**′ is also a wave number vector, but it may differ in dimensionality from **m**.

As a note of caution, the choice of dimensionality affects the shape and units of the power spectrum. For example, the statistics of a self-similar random medium are described by a three-dimensional PSDF of power law form with an exponent −3 which has units of distance cubed. The equivalent one-dimensional PSDF computed for a random line drawn through that same volume has a power law exponent of −1 and has units of distance. This is an important point when comparing spectral decay rates from different studies. Similarly, spherical spectra are subject to arbitrary choices in normalization (e.g., power per degree versus power per octave) which affect the apparent amplitude and falloff rate.

B3. Direct Comparisons Between Power Per Degree and Power Spectral Density

Now that power per degree and power spectral density have been explicitly defined, we move to explain how to directly compare the power per octave data of *Meschede and Romanowicz* [2015a] and our constrained 1-D PSDF. In other words, how do we take spherical harmonic spectra, S_l, and plot them on the same axes as the PSDF?

First, we obtain the spherical autocorrelation function, $\breve{R}(\theta)$, defined over the interval $[-\pi, \pi]$ for the model at a given radius, *r*, where θ represents the angle of separation at the origin between two points on the sphere. This can be done by taking a weighted sum of the Legendre polynomials, P_l:

$$\breve{R}(\theta) = \sum_l S_l P_l(\cos\theta). \tag{B11}$$

We provide a brief proof of equation (B11) in the following section.

After obtaining $\breve{R}(\theta)$, we can estimate what the corresponding ACF, $R(x)$, would be by unwrapping and flattening $\breve{R}(\theta)$ onto the Cartesian x axis. This is done by scaling θ by r, which yields a function defined on the interval $[-\pi r,\ \pi r]$

$$R(x) = \begin{cases} \breve{R}(x/r) & \text{if } x \in [-\pi r, \pi r] \\ \text{undefined} & \text{otherwise.} \end{cases}$$

Since we have no interest in wavelengths longer than $2\pi r$, we need not worry about the finite domain of this function.

Third, we estimate $P(m)$ numerically by discretizing $R(x)$ into n elements with indices $j = 0, 1, \ldots, n-1$ evenly spaced by Δx. We take the scaled digital Fourier transform of R_j to obtain a discretized PSDF:

$$P_k = \sum_{j=0}^{n-1} R_j \exp\left(-2\pi i \frac{jk}{n}\right) \Delta x \qquad k = 0, \ldots, n-1 \tag{B12}$$

with corresponding angular wave numbers

$$m_k = \frac{2\pi k}{n\Delta x} \qquad k = 0, \ldots, n-1. \tag{B13}$$

To perform this operation, we use built-in real fast Fourier transform and functions provided by Scipy (www.scipy.org).

B4. Proof

To prove $\breve{R}(\cos\theta) = \sum_l S_l P_l(\cos\theta)$, we begin by defining the spherical autocorrelation function as

$$\breve{R}(\hat{r}\cdot\hat{s}) = \langle f(\hat{r}) f(\hat{s})\rangle = \langle f(\hat{r}) f(\hat{s})^*\rangle. \tag{B14}$$

The second step is justified because we are concerned with real functions on a sphere. The model $f(\hat{r})$ can be written as a weighted sum of the spherical harmonic functions:

$$f(\hat{r}) = \sum_{l,m} c_{lm} Y_l^m(\hat{r}) \tag{B15}$$

so equation (B14) becomes

$$\breve{R}(\hat{r}\cdot\hat{s}) = \left\langle \sum_{l,m} c_{lm} Y_l^m(\hat{r}) \sum_{n,k} c_{nk}^* Y_n^k(\hat{s})^* \right\rangle = \sum_{l,m} \langle |c_{lm}|^2\rangle Y_l^m(\hat{r}) Y_l^m(\hat{s})^* \tag{B16}$$

where the second step is justified because the c_{lm} coefficients are uncorrelated for a random stationary process on a sphere. Using the spherical harmonic addition theorem, we obtain

$$\breve{R}(\hat{r}\cdot\hat{s}) = \sum_l \langle |c_{lm}|^2\rangle (2l+1) P_l\left(\hat{r}\cdot\hat{s}\right) = \sum_l S_l P_l\left(\hat{r}\cdot\hat{s}\right) \tag{B17}$$

where $S_l = \sum_m |c_{lm}|^2$ or equivalent $(2l+1)\langle |c_{lm}|^2\rangle$.

The addition theorem is given by

$$(2l+1) P_l(\hat{s}\cdot\hat{r}) = \sum_{m=-l}^{l} Y_l^m(\hat{s}) Y_l^{m*}(\hat{r}). \tag{B18}$$

Acknowledgments

This research was supported by National Science Foundation grant EAR-111111 and the National Science Foundation Graduate Research Fellowship Program. Qinya Liu was supported by the Discovery Grants of the Natural Sciences and Engineering Research Council of Canada (NSERC, 487237). The facilities of IRIS Data Services, and specifically the IRIS Data Management Center, were used for access to waveforms and related metadata used in this study. The facilities of IRIS Data Services, and specifically the IRIS Data Management Center (https://ds.iris.edu/ds/nodes/dmc/), were used for access to waveforms and related metadata used in this study. The data set for this paper is available by contacting the corresponding author at nicholas_mancinelli@brown.edu. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience and EarthScope (SAGE) Proposal of the National Science Foundation under Cooperative Agreement EAR-1261681. Computing resources for the SPECFEM3D_GLOBE simulations were provided by the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant ACI-1053575. The authors would like to acknowledge Zhitu Ma for generating normal mode synthetics which helped us diagnose the cause of the discrepant *PKP* amplitudes. This paper benefited from constructive comments provided by two anonymous reviewers.

References

Abubakirov, I., and A. Gusev (1990), Estimation of scattering properties of lithosphere of Kamchatka based on Monte-Carlo simulation of record envelope of a near earthquake, *Phys. Earth Planet. Inter.*, *64*(1), 52–67, doi:10.1016/0031-9201(90)90005-I.

Agranier, A., J. Blichert-Toft, D. Graham, V. Debaille, P. Schiano, and F. Albarède (2005), The spectra of isotopic heterogeneities along the mid-Atlantic Ridge, *Earth Planet. Sci. Lett.*, *238*, 96–109, doi:10.1016/j.epsl.2005.07.011.

Allègre, C. J., and D. L. Turcotte (1986), Implications of a two-component marble-cake mantle, *Nature*, *323*, 123–127, doi:10.1038/323123a0.

Antonsen, T. M., and E. Ott (1991), Multifractal power spectra of passive scalars convected by chaotic fluid flows, *Phys. Rev. A*, *44*(2), 851–857.

Bal, G., and M. Moscoso (2000), Polarization effects of seismic waves on the basis of radiative transport theory, *Geophys. J. Int.*, *142*, 571–585.

Batchelor, G. K. (1959), Small-scale variation of convected quantities like temperature in turbulent fluid: Part 1. General discussion and the case of small conductivity, *J. Fluid Mech.*, *5*(01), 113–133, doi:10.1017/S002211205900009X.

Becker, T. W., and L. Boschi (2002), A comparison of tomographic and geodynamic mantle models, *Geochem. Geophys. Geosyst.*, *3*, 1003, doi:10.1029/2001GC000168.

Chaljub, E., D. Komatitsch, J.-P. Vilotte, Y. Capdeville, B. Valette, and G. Festa (2007), Spectral-element analysis in seismology, *Adv. Geophys.*, *48*, 365–419.
Chandrasekhar, S. (1960), *Radiative Transfer*, Dover Publ., Inc., New York.
Chevrot, S., J. P. Montagner, and R. Sneider (1998), The spectrum of tomographic Earth models, *Geophys. J. Int.*, *133*, 783–788.
Deschamps, F., J. Trampert, and R. K. Snieder (2002), Anomalies of temperature and iron in the uppermost mantle inferred from gravity data and tomographic models, *Phys. Earth Planet. Inter.*, *129*(3–4), 245–264, doi:10.1016/S0031-9201(01)00294-1.
Fehler, M., M. Hoshiba, H. Sato, and K. Obara (1992), Separation of scattering and intrinsic attenuation for the Kanto-Tokai region, Japan, using measurements of *S*-wave energy versus hypocentral distance, *Geophys. J. Int.*, *108*(3), 787–800, doi:10.1111/j.1365-246X.1992.tb03470.x.
Gaebler, P. J., C. Sens-Schönfelder, and M. Korn (2015), The influence of crustal scattering on translational and rotational motions in regional and teleseismic coda waves, *Geophys. J. Int.*, *201*(1), 355–371, doi:10.1093/gji/ggv00.
Giampiccolo, E., T. Tuvè, S. Gresta, and D. Patanè (2006), *S*-waves attenuation and separation of scattering and intrinsic absorption of seismic energy in southeastern Sicily (Italy), *Geophys. J. Int.*, *165*, 211–222, doi:10.1111/j.1365-246X.2006.02881.x.
Gudmundsson, O., J. H. Davies, and R. W. Clayton (1990), Stochastic analysis of global travel time data: Mantle heterogeneity and random errors in the ISC data, *Geophys. J. Int.*, *102*, 25–43, doi:10.1111/j.1365-246X.1990.tb00528.x.
Gusev, A. A., and I. R. Abubakirov (1987), Monte-Carlo simulation of record envelope of a near earthquake, *Phys. Earth Planet. Inter.*, *49*(1–2), 30–6, doi:10.1016/0031-9201(87)90130-0.
Hacker, B. R. (2003), Subduction factory 1. Theoretical mineralogy, densities, seismic wave speeds, and H_2O contents, *J. Geophys. Res.*, *108*, 1–26, doi:10.1029/2001JB001127.
Hatzidimitriou, P. M. (1994), Scattering and anelastic attenuation of seismic energy in northern Greece, *Pure Appl. Geophys.*, *143*(4), 557–601.
Hedlin, M. A. H., P. M. Shearer, and P. S. Earle (1997), Seismic evidence for small-scale heterogeneity throughout the Earth's mantle, *Nature*, *387*(6629), 145–150, doi:10.1038/387145a0.
Hoshiba, M. (1991), Simulation of multiple-scattered coda wave excitation based on the energy conservation law, *Phys. Earth Planet. Inter.*, *67*(1–2), 123–136, doi:10.1016/0031-9201(91)90066-Q.
Hoshiba, M. (1993), Separation of scattering attenuation and intrinsic absorption in Japan using the multiple lapse time window analysis of full seismogram envelope, *J. Geophys. Res.*, *98*(B9), 15,809–15,824, doi:10.1029/93JB00347.
Hoshiba, M. (2000), Large fluctuation of wave amplitude produced by small fluctuation of velocity structure, *Phys. Earth Planet. Inter.*, *120*(3), 201–217, doi:10.1016/S0031-9201(99)00165-X.
Kawakatsu, H., P. Kumar, Y. Takei, M. Shinohara, T. Kanazawa, E. Araki, and K. Suyehiro (2009), Seismic evidence for sharp boundaries of oceanic plates, *Science*, *324*, 499–502.
Kellogg, L., and D. Turcotte (1987), Homogenization of the mantle by convective mixing and diffusion, *Earth Planet. Sci. Lett.*, *81*(4), 371–378, doi:10.1016/0012-821X(87)90124-5.
Kennett, B. L. N., and T. Furumura (2013), High-frequency *Po*/*So* guided waves in the oceanic lithosphere: I—long-distance propagation, *Geophys. J. Int.*, *195*(3), 1862–1877, doi:10.1093/gji/ggt344.
Komatitsch, D., and J. Tromp (2002a), Spectral-element simulations of global seismic wave propagation—I. Validation, *Geophys. J. Int.*, *149*(2), 390–412, doi:10.1046/j.1365-246X.2002.01653.x.
Komatitsch, D., and J. Tromp (2002b), Spectral-element simulations of global seismic wave propagation—II. Three-dimensional models, oceans, rotation, and self-gravitation, *Geophys. J. Int.*, *149*(2), 390–412, doi:10.1046/j.1365-246X.2002.01653.x.
Komatitsch, D., J. Ritsema, and J. Tromp (2002), The spectral-element method, Beowulf computing, and global seismology, *Science*, *298*(5599), 1737–1742, doi:10.1126/science.1076024.
Kustowski, B., G. Ekström, and A. M. Dziewoński (2008), Anisotropic shear-wave velocity structure of the Earth's mantle: A global model, *J. Geophys. Res.*, *113*(B6306), doi:10.1029/2007JB005169.
Mancinelli, N. J., and P. M. Shearer (2013), Reconciling discrepancies among estimates of small-scale mantle heterogeneity from *PKP* precursors, *Geophys. J. Int.*, *195*(3), 1721–1729, doi:10.1093/gji/ggt319.
Margerin, L., and G. Nolet (2003a), 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.
Margerin, L., and G. Nolet (2003b), Multiple scattering of high-frequency seismic waves in the deep Earth: Modeling and numerical examples, *J. Geophys. Res.*, *108*(2234), doi:10.1029/2002JB001974.
Margerin, L., M. Campillo, and B. Van Tiggelen (2000), Monte Carlo simulation of multiple scattering of elastic waves, *J. Geophys. Res.*, *105*(B4), 7873–7892, doi:10.1029/1999JB900359.
Meschede, M., and B. Romanowicz (2015a), Lateral heterogeneity scales in regional and global upper mantle shear velocity models, *Geophys. J. Int.*, *200*(2), 1076–1093, doi:10.1093/gji/ggu424.
Meschede, M., and B. Romanowicz (2015b), Non-stationary spherical random media and their effect on long-period mantle waves, *Geophys. J. Int.*, *203*(3), 1605–1625, doi:10.1093/gji/ggv356.
Owens, T. J., H. P. Crotwell, C. Groves, and P. Oliver-Paul (2004), SOD: Standing Order for Data, *Seismol. Res. Lett.*, *75*(4), 515–520.
Passier, M. L., and R. K. Snieder (1995), On the presence of intermediate-scale heterogeneity in the upper mantle, *Planet. Space Sci.*, *123*, 817–837, doi:10.1016/0032-0633(75)90189-0.
Przybilla, J., M. Korn, and U. Wegler (2006), Radiative transfer of elastic waves versus finite difference simulations in two-dimensional random media, *J. Geophys. Res.*, *111*(4), 1–13, doi:10.1029/2005JB003952.
Przybilla, J., U. Wegler, and M. Korn (2009), Estimation of crustal scattering parameters with elastic radiative transfer theory, *Geophys. J. Int.*, *178*(2), 1105–1111, doi:10.1111/j.1365-246X.2009.04204.x.
Ricard, Y., S. Durand, J.-P. Montagner, and F. Chambat (2014), Is there seismic attenuation in the mantle?, *Earth Planet. Sci. Lett.*, *388*(0), 257–264, doi:10.1016/j.epsl.2013.12.008.
Sato, H. (1984), Attenuation and envelope formation of three-component seismograms of small local earthquakes in randomly inhomogeneous lithosphere, *J. Geophys. Res.*, *89*(B2), 1221–1241.
Sato, H., M. C. Fehler, and T. Maeda (2012), *Seismic Wave Propagation and Scattering in the Heterogeneous Earth: Second Edition*, Springer, Berlin.
Sens-Schönfelder, C., and U. Wegler (2006), Radiative transfer theory for estimation of the seismic moment, *Geophys. J. Int.*, *167*(3), 1363–1372, doi:10.1111/j.1365-246X.2006.03139.x.
Shearer, P. M., and P. S. Earle (2004), The global short-period wavefield modelled with a Monte Carlo seismic phonon method, *Geophys. J. Int.*, *158*(3), 1103–1117, doi:10.1111/j.1365-246X.2004.02378.x.
Shearer, P. M., and P. S. Earle (2008), Observing and modeling elastic scattering in the deep Earth, *Adv. Geophys.*, *50*, 167–193.

Su, W.-j., and A. M. Dziewonski (1992), On the scale of mantle heterogeneity, *Phys. Earth Planet. Inter.*, *74*(1–2), 29–54, doi:10.1016/0031-9201(92)90066-5.

Tittgemeyer, M., F. Wenzel, K. Fuchs, and T. Ryberg (1996), Wave propagation in a multiple-scattering upper mantle—Observations and modelling, *Geophys. J. Int.*, *127*(2), 492–502.

Tromp, J., D. Komatitsch, and Q. Liu (2008), Spectral-element and adjoint methods in seismology, *Commun. Comput. Phys.*, *3*(1), 1–32.

Tromp, J., et al. (2010), Near real-time simulations of global CMT earthquakes, *Geophys. J. Int.*, *183*(1), 381–389, doi:10.1111/j.1365-246X.2010.04734.x.

Wegler, U., M. Korn, and J. Przybilla (2006), Modeling full seismogram envelopes using radiative transfer theory with born scattering coefficients, *Pure Appl. Geophys.*, *163*(2–3), 503–531, doi:10.1007/s00024-005-0027-5.

Wieczorek, M. A., and F. J. Simons (2005), Localized spectral analysis on the sphere, *Geophys. J. Int.*, *162*(3), 655–675, doi:10.1111/j.1365-246X.2005.02687.x.

Xu, W., C. Lithgow-Bertelloni, L. Stixrude, and J. Ritsema (2008), The effect of bulk composition and temperature on mantle seismic structure, *Earth Planet. Sci. Lett.*, *275*, 70–79, doi:10.1016/j.epsl.2008.08.012.