A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements

Eric Debayle, Yanick Ricard

A global shear velocity model of the upper mantle from fundamental and higher Rayleigh mode measurements

Eric Debayle[1] and Yanick Ricard[1]

[1]   We present DR2012, a global SV-wave tomographic model of the upper mantle. We use an extension of the automated waveform inversion approach of Debayle (1999) which improves our mapping of the transition zone with extraction of fundamental and higher-mode information. The new approach is fully automated and has been successfully used to match approximately 375,000 Rayleigh waveforms. For each seismogram, we obtain a path average shear velocity and quality factor model, and a set of fundamental and higher-mode dispersion and attenuation curves. We incorporate the resulting set of path average shear velocity models into a tomographic inversion. In the uppermost 200 km of the mantle, SV wave heterogeneities correlate with surface tectonics. The high velocity signature of cratons is slightly shallower (≈200 km) than in other seismic models. Thicker continental roots are not required by our data, but can be produced by imposing *a priori* a smoother model in the vertical direction. Regions deeper than 200 km show no velocity contrasts larger than ±1% at large scale, except for high velocity slabs within the transition zone. Comparisons with other seismic models show that current surface wave datasets allow to build consistent models up to degrees 40 in the upper 200 km of the mantle. The agreement is poorer in the transition zone and confined to low harmonic degrees (≤10).

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

1. Introduction

[2]   The dramatic increase in the number of seismic stations in the last 25 years has stimulated the development of automated approaches for global imaging of Earth's upper mantle [e.g., *Trampert and Woodhouse*, 1995; *van Heijst and Woodhouse*, 1997; *Debayle*, 1999; *Beucler et al.*, 2003; *Lebedev et al.*, 2005; *Yoshizawa and Ekstrom*, 2010]. These approaches are based on the analysis of surface wave datasets [*Trampert and Woodhouse*, 1995] sometimes including higher-modes [*Debayle et al.*, 2005; *Beucler and Montagner*, 2006; *van Heijst and Woodhouse*, 1999].

[3]   Most resulting S-wave models have been built using ray theory. At first glance, these shear-wave models are very consistent in the upper 200 km of the mantle, where they all have a very strong correlation with surface tectonics. However, a further comparison shows that these seismic models differ, even for wavelengths which exceed 1500 km, where the most conservative estimate confirms that ray theory is valid [*Spetzler and Snieder*, 2001]. As an illustration we show in Figure 1 a comparison between two recent SV-wave models at 100 km depth, S40RTS by *Ritsema et al.* [2011] and DKP2005 by *Debayle et al.* [2005]. Both models are based on ray theory and use Rayleigh waves analyzed in the period range 50–200 s. Significant differences are present at wavelengths greater than 1500 km, for example beneath Tibet, Europe, Australia or in oceanic areas. Such differences cannot be attributed to the theory, which is the same for both models. They have to be related to the way data have been extracted from the seismograms, or to the strategy and practical details of the inversion. For example, DKP2005 first constructs 1D radial models for all the rays before combining them in a 3D model, while S40RTS starts from 2D phase dispersion maps, before a depth dependent inversion. The parameterization, the *a priori* model, the method of regularization and the data weighting, are also specific to each tomographic model.

[4]   This issue is important, because recent developments aim at improving the resolution of tomographic models using more sophisticated theories. In places where seismic models show differences over wavelengths at which ray theory is valid, there is few hopes that smaller scale features inferred from a more sophisticated theory will provide valuable new information.

[1]Laboratoire de Géologie de Lyon: Terre, Planètes et Environnement, CNRS, Université de Lyon 1, Ecole Normale Supérieure de Lyon, Villeurbanne, France.

Corresponding author: E. Debayle, Laboratoire de Géologie de Lyon: Terre, Planètes et Environnement, CNRS, Université de Lyon 1, Ecole Normale Supérieure de Lyon, 2 rue Raphaël Dubois, Bâtiment Géode, FR-69622 Villeurbanne CEDEX, France. (eric.debayle@ens-lyon.fr)
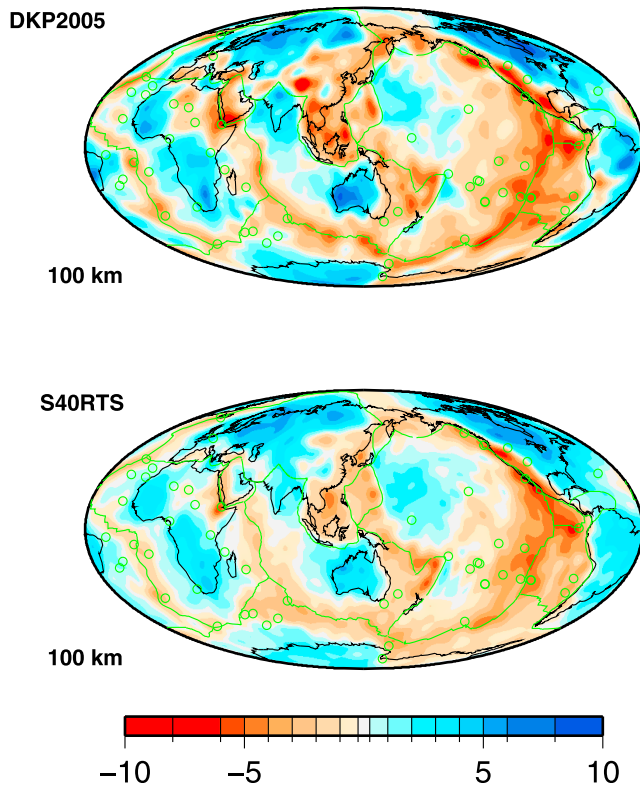
Figure 1. SV-wave perturbations relative to PREM at 100 km depth in two tomographic models: (top) DKP2005 by *Debayle et al.* [2005] and (bottom) S40RTS by *Ritsema et al.* [2011]. The color scale is in per cent.

[5] We believe therefore crucial to improve the way information is extracted from the data. One may hope that if seismologists agree on the robust dataset to be used in the inverse problem, a major cause of differences between current seismic models would disappear. This would provide a strong basis for future finite-frequency inversions.

[6] There are several ways to extract information from a surface wave seismogram. The most common approach is to derive path average phase or group velocity curves for the fundamental mode of surface waves [e.g., *Trampert and Woodhouse*, 1995; *Ekstrom et al.*, 1997; *Ritzwoller et al.*, 2002; *Nettles and Dziewonski*, 2008]. In the period range commonly used for global surface wave analysis (50–300 s), fundamental modes primarily constrain the upper 300 km of the mantle. Higher-mode information is sensitive to the deeper structure and can provide valuable information on the transition zone which is poorly sampled by body waves. However modeling higher-modes requires sophisticated approaches, because they propagate with similar group velocities and are thus difficult to separate in the time domain.

[7] Two approaches are commonly used to extract higher-mode information. The first involves clusters of stations [*Nolet*, 1975; *Cara*, 1979] or of events located at different depths but within a small epicentral area [*Stutzmann and Montagner*, 1993; *Beucler et al.*, 2003]. The second uses single seismograms [*Cara and Lévêque*, 1987; *Nolet*, 1990; *van Heijst and Woodhouse*, 1997; *Yoshizawa and Kennett*,

2002; *Visser et al.*, 2007] and is better suited to achieve dense ray coverages in tomographic studies.

[8] The mode branch stripping technique of *van Heijst and Woodhouse* [1997] was designed for long paths (>50°), typical of global tomography, for which higher-mode branches can be separated because their dispersion curves are sufficiently different.

[9] Waveform inversions [*Cara and Lévêque*, 1987; *Nolet*, 1990; *Yoshizawa and Kennett*, 2002] have been designed for short paths (<50°) typical of regional tomography. The idea is to find a path average 1D model which explains the waveform of a seismogram and takes into account all the information present in the different modes which interfere in this waveform. The resulting 1D-model can be seen either as a specific path-average structure [*Cara and Lévêque*, 1987; *Nolet*, 1990], or as a summary of the fundamental and higher-mode dispersion curves [*Yoshizawa and Kennett*, 2002; *Visser et al.*, 2008].

[10] *Yoshizawa and Kennett* [2002] and *Visser et al.* [2008] use non-linear waveform fitting with the neighborhood algorithm of *Sambridge* [1999] to obtain path specific 1D models which are then used to estimate multimode dispersion curves. *Nolet* [1990] uses waveform fitting of suitably filtered seismograms, starting from long period, where the seismogram is less sensitive to the strongest heterogeneities of the uppermost mantle, and shifting progressively to shorter periods. Because of the strong non linearity of the waveform inversion, this approach requires an accurate starting model in order to avoid solutions corresponding to secondary minima of the cost function.

[11] The approach of *Cara and Lévêque* [1987] is based on the definition of secondary observables, built up from the seismograms, having only a slightly non-linear dependence upon the model parameters. This minimizes the dependence on the starting model. *Debayle* [1999] extends this approach with an automated procedure able to match the waveforms of large volumes of individual records, starting with synthetic seismograms computed with a single upper mantle model. This automated approach has been used in many regional [e.g., *Debayle et al.*, 2001; *Sieminski et al.*, 2003; *Pilidou et al.*, 2004; *Heintz et al.*, 2005; *Maggi et al.*, 2006a, 2006b; *Priestley et al.*, 2008] or global [*Debayle et al.*, 2005] studies.

[12] A first goal of this paper is to extend *Debayle*'s [1999] approach in order to improve the extraction of information, especially higher-modes, from a surface wave seismogram. We summarize the original *Cara and Lévêque* [1987] approach in section 2 and the new automated scheme in section 3. The new method increases the computation time required to model a single waveform by a factor of 10, but a Beowulf computer makes it possible to process hundreds of thousands of seismograms in a few weeks.

[13] A second goal is to present DR2012, our new global 3D SV-model of the upper mantle. In section 4, we apply the new approach to a global dataset of fundamental and higher-mode Rayleigh waves that includes 374,897 waveforms. For each waveform we derive a path average 1D S-velocity and quality factor model, and a set of fundamental and higher-mode dispersion and attenuation curves compatible with the record. We then combine the set of 1D S-velocity models in a tomographic inversion to built DR2012. In section 5 we first show that the new approach extracts more information
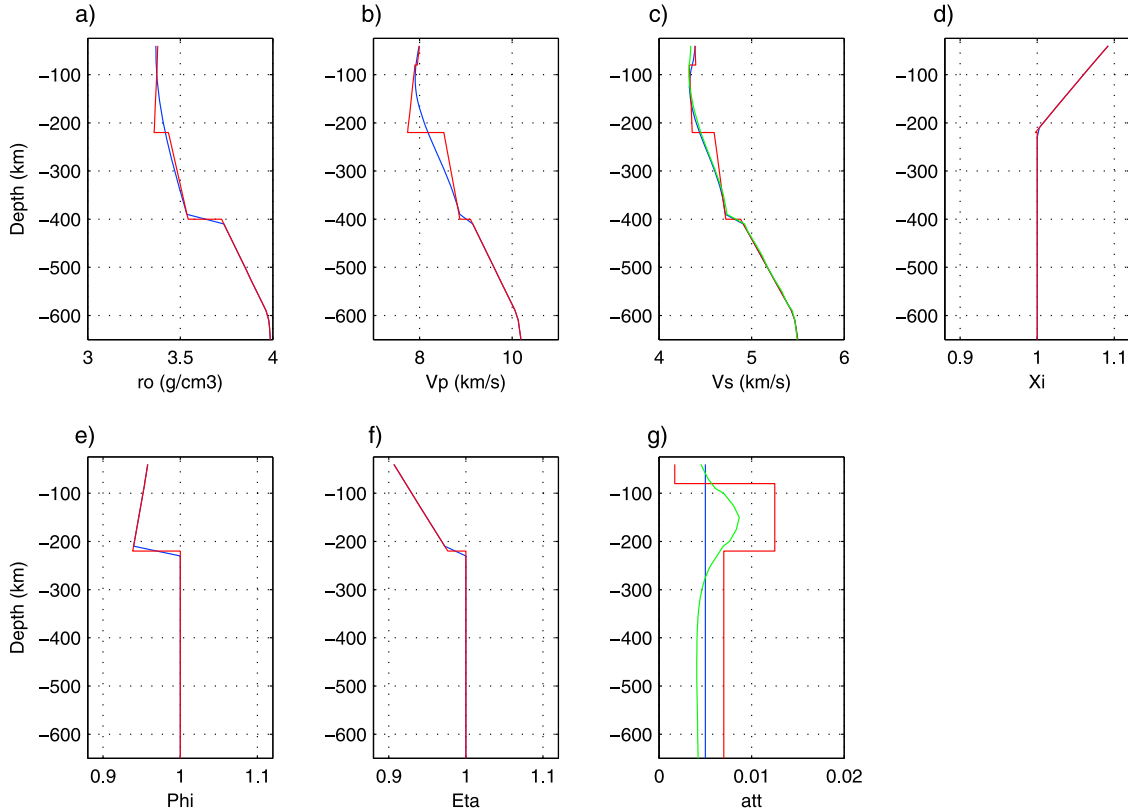
Figure 2. Our 1D starting model (blue) superimposed to PREM (red) at a reference period of 100 s. For the inverted parameters (Vs and the shear attenuation) we also plot in green the 1D models obtained after averaging the 374,897 inverted profiles. For Vs, this inverted model is very close to the 1D starting model in blue.

compared with *Debayle* [1999], and improves our mapping of heterogeneities, especially within the transition zone. Then we compare DR2012 with other seismic models.

2. Waveform Inversion

2.1. Synthetic Seismograms

[14] In *Cara and Lévêque* [1987], the surface wave is represented by a finite sum of pure-mode synthetics computed for a laterally homogeneous medium. The expression of a pure-mode synthetic $\hat{s}_p(t)$ is given by:

$$\hat{s}_p(t) = g(x) \int I(\omega) S_p(\omega) e^{-\alpha_p(\omega)x} e^{i[\omega t - k_p(\omega)x]} d\omega \qquad (1)$$

where p is the mode rank, x the epicentral distance, $g(x)$ the geometric expansion, ω the circular frequency, $I(\omega)$ the instrumental response, $S_p(\omega)$ the complex source excitation, $\alpha_p(\omega)$ the apparent attenuation factor and $k_p(\omega)$ the wavenumber function.

[15] The source excitation $S_p(\omega)$ is computed using a specific crustal and upper mantle model taken at the epicenter location. This 1D model is obtained by extracting density, seismic velocities and attenuation from 3SMAC [*Nataf and Ricard*, 1995] beneath the epicenter. $S_p(\omega)$ is then computed following *Cara* [1979], using the global CMT solution [*Dziewonski et al.*, 1981; *Ekstrom et al.*, 2012] issued at the Lamont-Doherty Earth Observatory of Columbia University.

[16] The attenuation $\alpha_p(\omega)$ and wavenumber $k_p(\omega)$ are computed following *Takeuchi and Saito* [1972] for a 1D model adapted for each ray. This model includes a path averaged crust structure estimated from 3SMAC [*Nataf and Ricard*, 1995]. The upper mantle part is radially anisotropic and very close to PREM at a reference period of 100 s [*Dziewonski and Anderson*, 1981] for density and elastic parameters, although the 220 km discontinuity has been smoothed out (see Figure 2). The attenuating layer located at 80 km depth in PREM represents a strong *a priori* choice which is not adapted to continental paths for which the attenuating layer is less pronounced or located deeper. We use therefore a uniform 1D quality factor $Q_\beta(z)$ of 200 as a starting upper mantle model (see Figure 2g). The wavenumber is corrected from physical dispersion using *Kanamori and Anderson* [1977] assuming a reference period of 100 s.

[17] This careful choice of the *a priori* information for the starting model in the source region and the average crustal structure along each epicenter-station path allows us to invert for the path-average upper mantle structure only, assuming the crustal structure and the source excitation are known. We tested the impact of the crustal model to the final tomographic maps in some of our previous papers [*Debayle and Kennett*, 2000; *Pilidou et al.*, 2004; *Priestley et al.*, 2008]. Crustal corrections done with 3SMAC or CRUST2 [*Bassin et al.*, 2000] have no effect under oceans where the average crust is much thinner (~7 km) than the depth of maximum sensibility of Rayleigh waves (≥70 km for

periods larger than 50 s). Even on continents where the crust is thick, the effects of crustal corrections, whether using 3SMAC or CRUST2, are undistinguishable at depths larger than 100 km depth and minor before [*Debayle and Kennett*, 2000; *Pilidou et al.*, 2004]. However, in order to avoid any possible contamination of our model by the crustal structure, we will only show tomographic maps for depths greater than or equal to 100 km.

[18] At periods greater than 50 s, a combination of 6 pure-mode synthetics $\hat{s}_p(t)$ is sufficient for an accurate description of the waveform in the group velocity window associated with surface waves. The complete synthetic seismogram $\hat{s}(t)$ is thus obtained by summing pure-mode synthetics $\hat{s}_p(t)$ for the fundamental mode and the first five overtones:

$$\hat{s}(t) = \sum_{p=1}^{6} \hat{s}_p(t). \tag{2}$$

2.2. Secondary Observables

[19] The secondary observables introduced by *Cara and Lévêque* [1987] are the starting data for the waveform inversion. They are extracted from the observed and synthetic seismograms, $s(t)$ and $\hat{s}(t)$ by cross-correlation with synthetic seismograms for individual modes $\hat{s}_p(t)$ computed for a reference model. The resulting cross-correlograms are then band-pass filtered around different frequencies. The combination of band-pass filtering and cross-correlation can be represented as:

$$g_p(\omega_q, t) = h(\omega_q, t) * s(t) * \check{s}_p(-t), \tag{3}$$

for an observed cross-correlogram, and:

$$\hat{g}_p(\omega_q, t) = h(\omega_q, t) * \hat{s}(t) * \check{s}_p(-t), \tag{4}$$

for a synthetic cross-correlogram. Here $*$ denotes a convolution, $h(\omega_q, t)$ is the impulse response of a band-pass filter centered on the circular frequency ω_q, and $\check{s}_p$ denotes the complex conjugate of $\hat{s}_p$. The actual secondary observables used by *Cara and Lévêque* [1987] are defined by sampling three values taken at different time lags on the observed envelope of the modal cross-correlograms $g_p(\omega_q, t)$, with a visual inspection of the envelope. One value is taken at the appropriate maximum of the envelope and two others on either side of this position. The inversion minimizes the difference between these observables and those computed for a complete synthetic seismogram $\hat{s}(t)$. The instantaneous phase of the cross-correlogram, taken at the time where the envelope reaches its maximum is generally also included to adjust the waveform fit.

2.3. 1D Model Inversion

[20] A 1D model is then derived to explain each seismogram. Following *Debayle* [1999], this inverted model $\hat{\mathbf{m}}$ includes the shear wave velocity $\beta_v(z)$, the attenuation parameterized by $\log(Q_\beta(z))$ and the scalar moment through $\log(M_0)$. The inversion of $\log(Q_\beta(z))$ accounts for frequency-dependent amplitude differences between the synthetic and recorded waveforms. As waveform modeling means both phase and amplitude modeling, the inverted $\beta_v(z)$ profiles might be biased if attenuation is not inverted for and if the starting attenuation model is not accurate enough.

[21] We could have used some *a priori* information to correlate changes in shear velocity to those in compressional velocity, density and radial anisotropy. However, our experience in agreement with previous studies [e.g., *Nishimura and Forsyth*, 1989], shows that this kind of *a priori* coupling does not change the results for the best resolved parameters $\beta_v(z)$ and $\log(Q_\beta(z))$, while the others are essentially constrained by the *a priori* information. Therefore, the radial anisotropy, density and Vp velocity profiles remain fixed to their initial values in the inversion.

[22] In this paper, we do not discuss in details the inverted $\log(Q_\beta(z))$ model. Unambiguous interpretation of $\log(Q_\beta(z))$ requires a good control of the source parameters (there is a strong trade-off between them [*Lévêque et al.*, 1991]) and corrections of focusing-defocusing effects. However, we show in Figure 2 the 1D $\beta_v(z)$ and $\log(Q_\beta(z))$ models obtained after averaging our whole dataset. The average velocity model (Figure 2c, green curve), is very close to the starting model. The average attenuation model remains globally less attenuating than PREM (Figure 2g, green curve). It has an attenuating layer located between 100 and 200 km, similar to PREM. This attenuating layer is not present in the starting model, and is therefore required by our data.

[23] The secondary observables depend upon the model parameters through weakly non-linear relations [*Cara and Lévêque*, 1987] which are inverted using few iterations [*Lévêque et al.*, 1991; *Tarantola and Valette*, 1982]. The inverted model $\hat{\mathbf{m}}_{k+1}$ at iteration $k + 1$ is given by:

$$\hat{\mathbf{m}}_{k+1} = \mathbf{m}_0 + \mathbf{C}_{m0}\mathbf{G}_k^t(\mathbf{G}_k\mathbf{C}_{m0}\mathbf{G}_k^t + \mathbf{C}_{d0})^{-1} \\ [\mathbf{d} - \mathbf{d}(\hat{\mathbf{m}}_k) + \mathbf{G}_k(\hat{\mathbf{m}}_k - \mathbf{m}_0)] \tag{5}$$

In equation (5), $\mathbf{m}_0$ is the *a priori* model, $\mathbf{t}$ denotes the transpose, the matrix $\mathbf{G}_k$ contains the partial derivatives of the secondary observables with respect to the model parameters, calculated following *Cara and Lévêque* [1987], $\mathbf{C}_{m0}$ and $\mathbf{C}_{d0}$ are the *a priori* covariance matrices for model and data.

[24] The matrix $\mathbf{C}_{d0}$ is assumed to be diagonal. Its diagonal terms are the variances of the secondary observables and describe the errors made on the data measurements. We use a standard deviation of 10% of the value of the envelope data and 5% of 2π radians for the phase data.

[25] The *a priori* covariance function C_{m0} is composed of three sub-matrices expressing separately the covariances on velocity, attenuation and seismic moment. No *a priori* cross-covariances exist between these sub-matrices. For velocity and attenuation, the covariance function between two depths z_1 and z_2 is defined after *Lévêque et al.* [1991]:

$$C_{m0}(z_1, z_2) = \sigma_1 \sigma_2 \exp\left(\frac{-(z_1 - z_2)^2}{2L^2}\right). \tag{6}$$

The standard deviation σ_i controls the amplitude of a component of the model perturbation allowed at a given depth z_i. We use constant values of 0.05 km s^{-1} and 0.25 at all depths for those of $\beta_v(z)$ and $\log(Q_\beta(z))$. The correlation length L controls the vertical smoothness of the model and we use $L = 50$ km. A very large standard deviation value of 0.5 for $\log(M_0)$ accounts for amplitude differences between synthetic and actual waveforms. There is a clear trade off between seismic moment and attenuation. To reduce suspicion on the inverted $\log(Q_\beta(z))$ models, we reject the paths for which
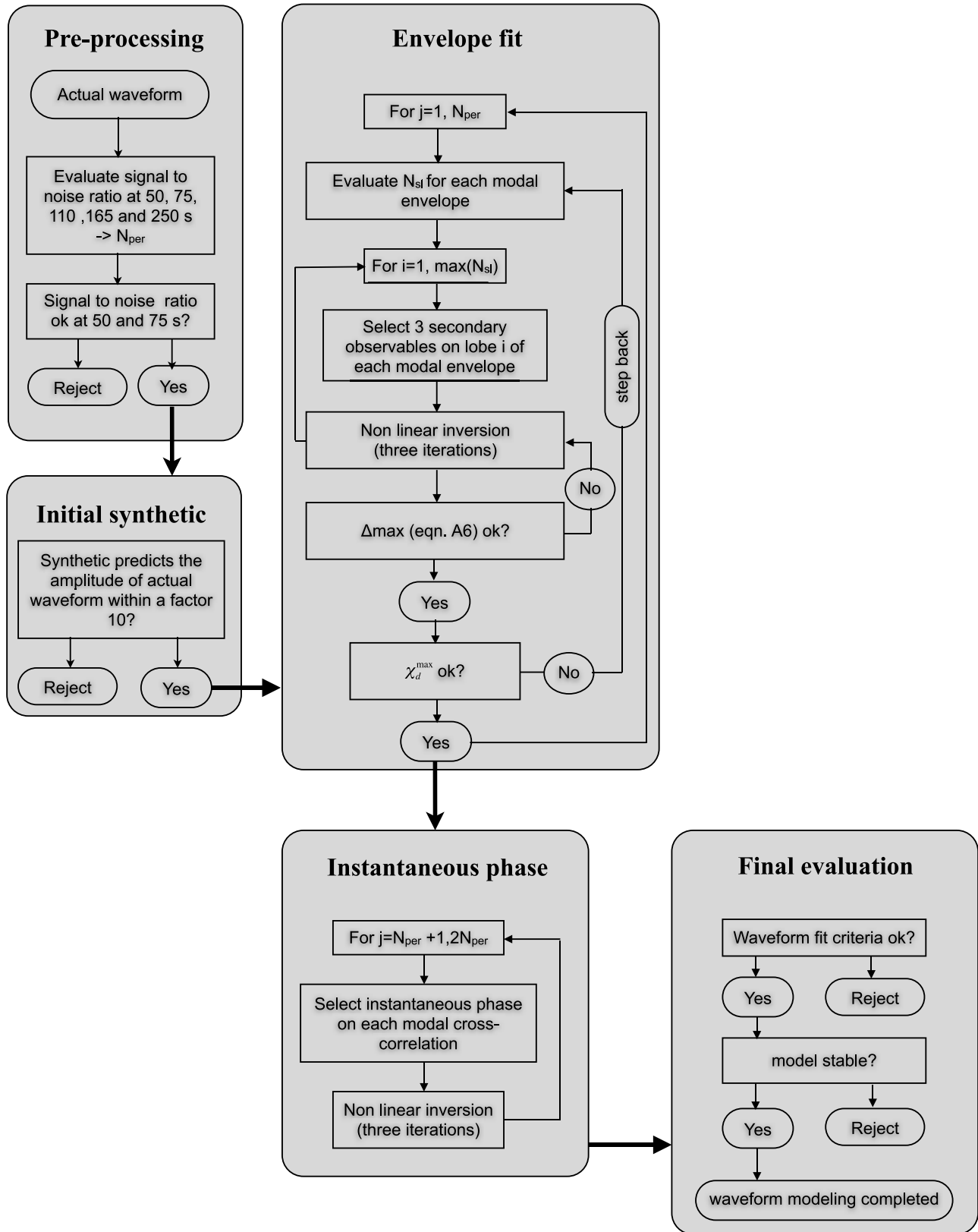
Figure 3. Flow-chart of the automated scheme. N_{per} is defined in section 3, N_{sl} and χ^d_{max} are defined in Appendix A.

the amplitude of the initial synthetic seismogram differs by a factor greater than 10 from the actual waveform.

[26] *Tarantola and Valette* [1982] show that when the non-linearity is weak, the *a posteriori* covariance $\mathbf{C_m}$ and resolution $\mathbf{R}$ matrices can be approximated by the formulae obtained in the linear case:

$$\mathbf{C_m} = \mathbf{C_{m0}} - \mathbf{C_{m0}} \mathbf{G^t} \left(\mathbf{G C_{m0} G^t} + \mathbf{C_{d0}} \right)^{-1} \mathbf{G C_{m0}}, \qquad (7)$$

and

$$\mathbf{R} = \mathbf{C_{m0}} \mathbf{G^t} \left(\mathbf{G C_{m0} G^t} + \mathbf{C_{d0}} \right)^{-1} \mathbf{G}, \qquad (8)$$

where the partial derivatives of $\mathbf{G}$ are computed at the final iteration. We use equation (7) to estimate the *a posteriori* variance on the inverted parameters. From equation (8), we compute the trace of $\mathbf{R}$ which corresponds to the number of independent pieces of information which are extracted from the data.

3. New Automated Scheme

[27] A first automation of the *Cara and Lévêque* [1987] approach was proposed by *Debayle* [1999]. The idea was to mimic the choices of a manual user: selection of a suitable period range of inversion for each waveform, decomposition of the waveform inversion into long and short period parts, selection of secondary observables on the envelope of the modal cross-correlogram functions $g_p(\omega_q, t)$.

[28] Here, we revisit the automation and further divide the waveform modeling into a larger number of elementary steps. Such a division gives more flexibility to the code, allowing a better selection of the robust information and a better higher-mode extraction. As in *Debayle* [1999], the automated code can be divided into a pre-processing step and a waveform inversion step which includes the automatic selection of the secondary observables.

3.1. Pre-processing

[29] During this step, we prepare the data for the waveform inversion and we choose the central frequencies ω_q of the band pass filter $g_p(\omega_q, t)$. We first check for each record that all the necessary information (instrument response, global CMT determination [*Dziewonski et al.*, 1981; *Ekstrom et al.*, 2012]) is available. Then, we set the header records to the centroid origin time and location provided by the Lamont-Doherty Earth Observatory of Columbia University and we select a group velocity window (between 2.5 and 7 km s^{-1}) around the surface wave part of the seismogram.

[30] For each record, we use a maximum of five band-pass filters, with the following sequence of central periods: 250, 165, 110, 75 and 50 s. These filters are described in details in Appendix A.

[31] Starting from the band pass filter at 250 s, we compare the amplitude of the envelope of the signal A_s (part of the record arriving with a group velocity larger than 3 km s^{-1}) to the amplitude of the envelope of the noise A_n (part of the record arriving with a group velocity smaller than 3 km s^{-1}). Only when A_s/A_n is larger than a threshold value (3 in this study), is the data kept at this period. Then, we repeat the process to the next period, 165 s, until the last period of 50 s. All the records for which the signal to noise criterion $A_s/A_n > 3$

is not reached at least at 50 and 75 s are rejected. The use of filtered cross-correlogram in the period range 50–75 s is sufficient to constrain the SV-wave velocity up to a depth greater than 250 km, even when only the fundamental mode is taken into account in the inversion [*Lévêque et al.*, 1991]. At the end of the preprocessing step, an appropriate number N_{per} of band-pass filters has been selected for each record, with N_{per} ranging between 2 and 5. For the best records, $N_{per} = 5$ corresponding to robust records around the periods 250-165-110-75-50 s. The lowest quality records have $N_{per} = 2$ at periods 75–50 s.

3.2. Step-by-Step Inversion

[32] For each path, the recorded waveform is fitted though a complex series of steps, each step requiring a new level of accuracy in the quality of the fit.

[33] First, we match the envelopes of the cross-correlograms. For a seismogram for which N_{per} band-pass filters have been selected, the inversion starts with the longest period. Once the filtered modal envelope has been matched at some period, the automated code adds the next shorter period. In the best cases, the fit of the cross-correlograms envelopes are therefore refined $N_{per} = 5$ times (at 250 s only, then simultaneously at 250 s and 165 s, until the five periods 250-165-110-75-50 s are included). In the cases $N_{per} = 2$, only two refinements are made (75 s, then 75 s and 50 s).

[34] Second, for the same record, we also match the instantaneous phase of the filtered cross-correlograms. Here again we proceed period by period, starting from the longest period and adding successively the details of the shorter periods. The inversion of the envelope for the whole frequency range explains the group velocity of the different mode branches of the seismogram. Adding the instantaneous phase further increases the precision of the fit. The strategy of incremental fit from long to short periods, avoids phase skips of 2π. This approach is close to that followed by *van Heijst and Woodhouse* [1997] but the inversion is not separated mode by mode. At a given frequency, all the modes which contribute more than 1% of the total energy of the initial synthetic seismogram are considered in the inversion.

[35] The new scheme allows us to analyze the fundamental and up to 5 overtones in the period range 50–250 s. Figure 3 provides a flow-chart of the new scheme and a complete description of its implementation is provided in Appendix A.

4. Tomographic Inversion

4.1. Dataset

[36] We have applied our automated scheme to a global dataset of 960,364 Rayleigh wave seismograms recorded between 1977 and 2009. Most seismograms have been recorded at permanent stations of the Global Seismographic Network (GSN) and the International Federation of Digital Seismographic Networks (FDSN). We also use data from French temporary experiments in the Horn of Africa, French Polynesia [*Barruol et al.*, 2002] and Aegean-Anatolia region after the SIMBAAD experiment [*Salaün et al.*, 2012], from PASSCAL experiments in Africa, Tibet and New-Zealand and from the SKIPPY temporary deployment in Australia. Finally, we included data of USArray Transportable Array
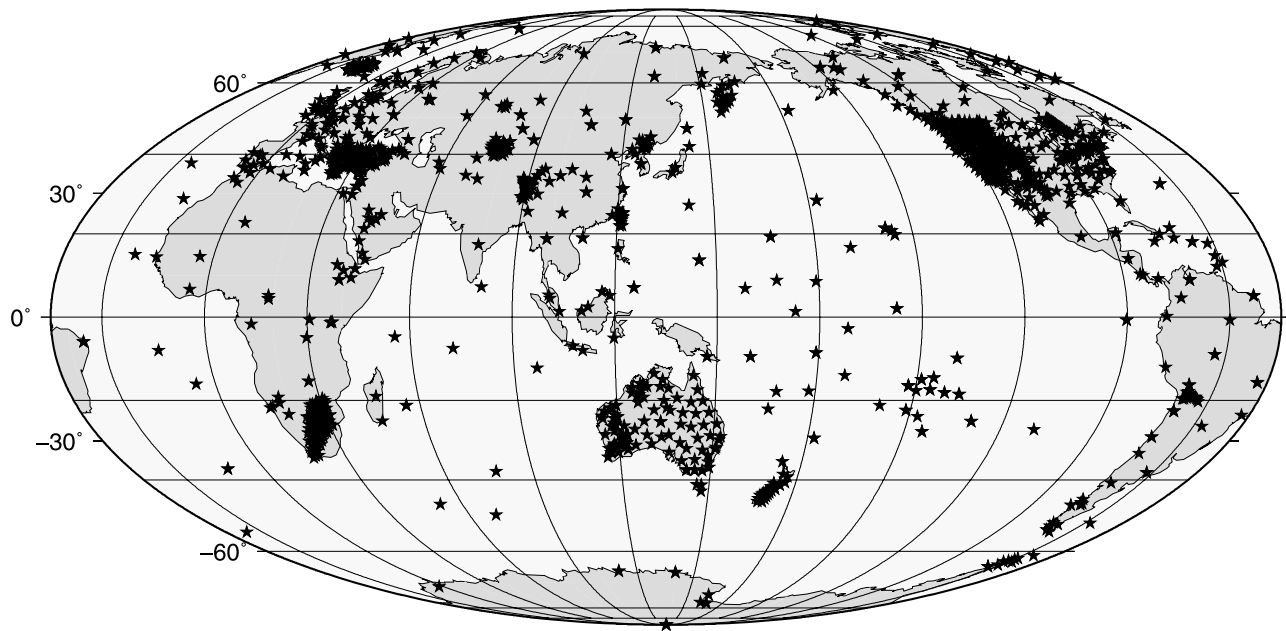
Figure 4. Location of stations (stars) used in this study.

and Reference Network Stations. The distribution of stations used in this study is shown in Figure 4.

[37] The automated waveform inversion is distributed over the nodes of a Beowulf cluster. Using a 45 nodes Beowulf cluster, each node with two quad-core processors, the whole dataset is processed in about 10 days. About 39% of the records pass successfully all stages of the waveform inversion, corresponding to 374,897 accepted waveforms through a total of ~20 million secondary observables fitted.

4.2. Regionalization

[38] The waveform inversion yields about 375,000 path average $\beta_v(z)$ and $\log(Q_\beta(z))$ models. It also yields a set of multimode dispersion and attenuation curves compatible with each surface wave record. These multimode dispersion curves provide a unique global database for future finite frequency inversions. In this paper, we focus on the tomographic inversion of the velocity and we leave the inversion of attenuation to a future paper (notice that this effective quality factor includes effects related to the focusing-defocusing of the wave energy by the seismic velocity anomalies, in addition to real dissipation).

[39] The tomographic inversion is based on the continuous regionalization formalism of *Montagner* [1986], which is derived from the approach developed by *Tarantola and Valette* [1982]. A smooth model is obtained by imposing correlations between neighboring points using a Gaussian *a priori* covariance function on the form of equation (6), but where L is now a horizontal correlation length which controls the lateral degree of smoothing of the inverted model, while the *a priori* model standard deviation σ controls its amplitude.

[40] *Montagner* [1986] inverted path average group or phase velocity curves for lateral variations of velocity and azimuthal anisotropy. However, his approach can also be used to retrieve the local distribution of shear velocity and azimuthal anisotropy from a set of path average $\beta_v(z)$ models [*Lévêque et al.*, 1998]. In this paper, we use a tomographic scheme for massive surface wave inversion developed by *Debayle and Sambridge* [2004] which incorporates various sophisticated geometrical algorithms which dramatically increase the computational efficiency and render possible the inversion of massive datasets.

[41] Our tomographic scheme is based on the "great circle ray theory" assumption that surface waves propagate along the source-station great circles, and that they are only sensitive to the structure along a zero-width ray. Although more sophisticated finite frequency theories exist [*Marquering et al.*, 1996; *Yoshizawa and Kennett*, 2002; *Spetzler et al.*, 2002], several authors [e.g., *Sieminski et al.*, 2004; *Trampert and Spetzler*, 2006] have shown that finite-frequency effects can be accounted for by a physically-based regularization of the inversion. *Sieminski et al.* [2004] performed a series of synthetic tests, comparing *Debayle and Sambridge* [2004] tomographic scheme with a finite frequency inversion based on linearized scattering theory [*Spetzler et al.*, 2002]. They show that results obtained using *Debayle and Sambridge* [2004] scheme are consistent with those of finite-frequency inversion, provided the path coverage is dense enough. For these reasons, we believe that a finite frequency approach would be computationally too expensive and would not allow significant progress compared with our tomographic scheme, which incorporates a sophisticated a priori information in the inverse problem.

[42] Following *Lévêque et al.* [1998], we apply the *Debayle and Sambridge* [2004] code directly to the $\beta_v(z)$ path average models and we invert for the local distribution of shear velocity including or not azimuthal anisotropy. We use a horizontal correlation length of $L = 400$ km, and a standard deviation $\sigma = 0.05$ km s^{-1} for the isotropic component of

the shear wave velocity. We have undertaken inversions with or without azimuthal anisotropy. When inverting for azimuthal anisotropy, we use an *a priori* standard deviation of 0.005 km s^{-1} for the anisotropic components of the shear wave velocity. Increasing L leads to smoother tomographic images with higher amplitude anomalies, but the overall pattern of anomalies remains unchanged. For comparison with models expressed in spherical harmonics, our model uses spherical harmonics up to the degree $l \approx (2\pi R)/L \approx 45$, where R is Earth's radius. In this paper we only discuss the isotropic part of the model and leave the anisotropic part to a future paper.

4.3. Path Clustering

[43] We follow *Debayle and Sambridge* [2004] to combine the selected $\beta_v(z)$ models in the tomographic inversion, as described in section 4.2. At this stage, we use the *a posteriori* error on $\beta_v(z)$ models to weight the data. The *a posteriori* covariance matrix is related to the *a priori* covariance matrix by:

$$\mathbf{C_m} = (\mathbf{I} - \mathbf{R})\mathbf{C_{m0}} \qquad (9)$$

Therefore, at a given depth, the *a posteriori* error is large and close to the *a priori* error when the actual waveform contains little information on the $\beta_v(z)$ structure (i.e. when $\mathbf{R}$ is close to zero), while it decreases when $\mathbf{R}$ approaches identity. For this reason, we keep only the $\beta_v(z)$ models for which the *a posteriori* error is smaller than 80% of the *a priori* error at a given depth. By this way, only well resolved path average $\beta_v(z)$ model are used in the tomographic inversion at all depths.

[44] To decrease the computing time and memory requirement, we cluster at each depth the selected path-average $\beta_v(z)$ models associated with close epicenters recorded at a given station. We use a cluster radius of 200 km (i.e. smaller than the correlation length) and consider that the waves of a given cluster follow the same path and see the same average Earth structure between the epicenters and the station. At each depth, we select from each cluster the "best resolved" $\beta_v(z)$ model. The best resolved $\beta_v(z)$ model has the largest trace of $\mathbf{R}$.

[45] After this clustering, our $\sim$375,000 $\beta_v(z)$ models reduce to $\sim$125,000 clusters corresponding to "independent" well-resolved paths. At depths greater than 200 km, the number of independent paths decreases because only intermediate-to-deep events located in subduction zones provide well excited higher-modes and well resolved shear velocities. At 400 km depth, the number of independent paths is close to $\sim$30,000. It exceeds 20,000 paths down to 575 km, while 11,007 paths still remain at 700 km depth. We show in section 4.4 that this provides a global coverage of the Earth within the upper mantle and transition zone.

[46] Note that our data selection strategy differs from previous applications of the *Debayle* [1999] algorithm by *Debayle et al.* [2005] and *Maggi et al.* [2006a]. These authors use the *a posteriori* error on each individual $\beta_v(z)$ model as a guide to discard data which provide little resolution in a given depth range. Then, they cluster the remaining data and extract from each cluster the mean shear wave velocity

$\bar{\beta}_v(z)$ and its standard deviation $\bar{\sigma}(z)$. They use $\bar{\beta}_v(z)$ and $\bar{\sigma}(z)$ along the average paths for the tomographic inversion.

[47] Our choice of extracting the best resolved $\beta_v(z)$ model of each cluster preserves the number of independent paths but avoids loss of resolution through averaging with less resolved models. Furthermore, it allows us to use the *a posteriori* error issued from the waveform modeling in the tomographic inversion. By this way, we carry in the tomographic inversion the information related to the depth sensitivity of the actual dataset, an information which is lost when using $\bar{\beta}_v(z)$ and its standard deviation.

4.4. Data Coverage and Errors

[48] Figure 5 shows the number of rays crossing a 400 × 400 km surface area at various depths within the upper mantle. These maps are typical of global seismology with strong bias towards the continents of the northern hemisphere. The transition from fundamental to higher mode coverage is observed between 200 and 450 km (Figure 5). However, even in the less covered areas of the transition zone, Figure 5 shows that our seismic model includes higher-mode constraints from 10 to 100 different paths in each 400 × 400 km surface area. Therefore, our higher-mode dataset provides global coverage of the transition zone with a large number of redundant data.

[49] At this stage, it is worth noting that errors in the source parameters may bias the path-average shear velocity estimation for some of the 1D models. However, the paths for which this bias might be strong are rejected by applying the selection criteria (see section A2). For the other paths, these effects have no reason to be coherent since they are related to earthquakes with different focal mechanisms or different focal depths. Therefore, they should average out in our tomographic inversion which involves a large number of paths with different azimuths.

4.5. Tomographic Maps

[50] The SV-velocity distribution at different depths for an isotropic inversion is depicted in Figure 6. Figure 7 shows the equivalent result for an inversion including azimuthal anisotropy. Comparison of Figures 6 and 7 proves that extracting azimuthal anisotropy does not alter significantly the isotropic component. The main difference is a slightly larger amplitude of SV-wave perturbation in the isotropic inversion by up to 2% at 150 km depth, reducing to less than 0.6% at depths greater than 200 km. At 150 km depth, the most noticeable differences occur in the Pacific, in agreement with *Ekstrom* [2011]. Our preferred SV-velocity model is the isotropic component of the anisotropic inversion as shown in Figure 7, because it is corrected for the small bias due to azimuthal anisotropy. In the next figures of this paper, we will always discuss and show the isotropic part of an anisotropic model.

[51] In the upper 200 km of the mantle, DR2012 (Figure 7) shows a very close correlation with surface tectonics. Mid oceanic ridges have a strong slow velocity signature in the first 100 km depth under all oceans younger than $\sim$40 Myrs. This localized slow signature of ridges is rather shallow and vanishes rapidly between 100 and 150 km depths after which all oceans have a rather homogenous slow velocity.
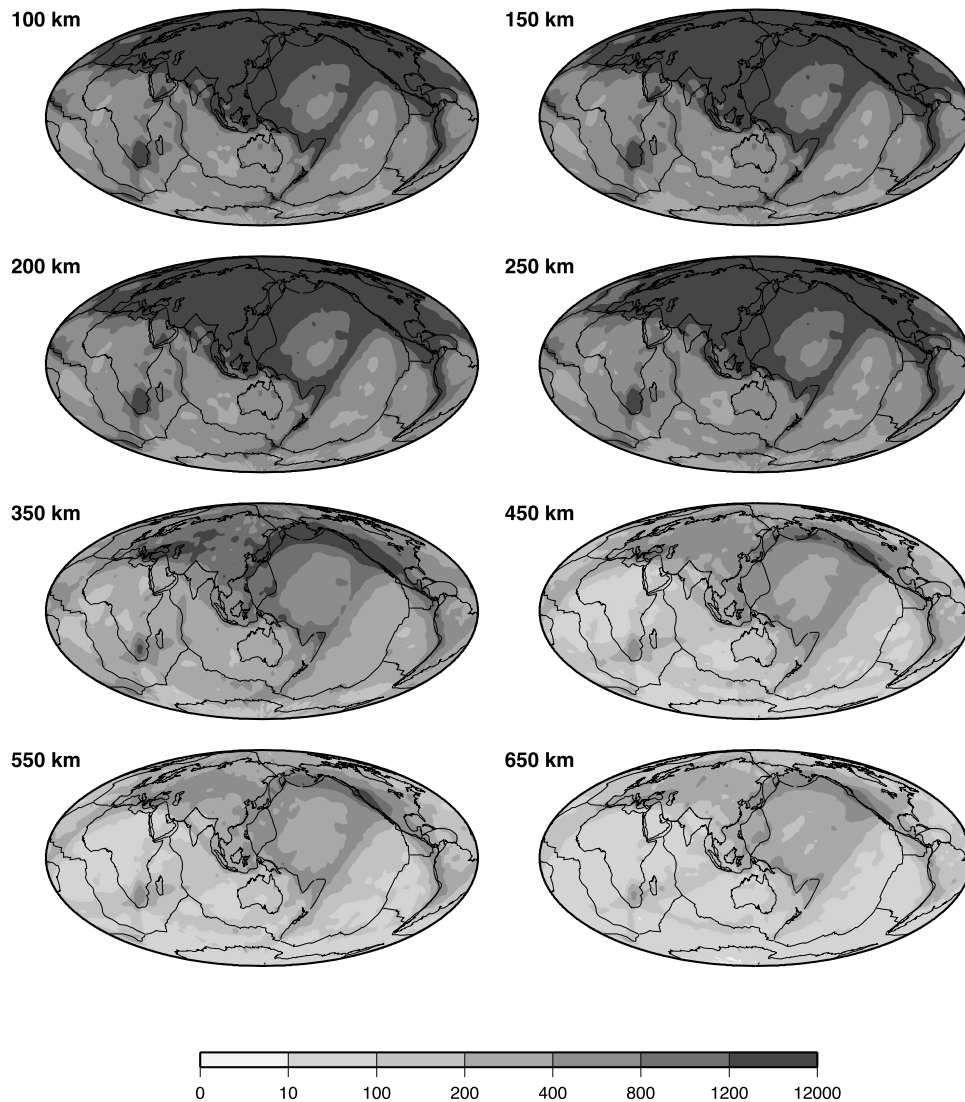
Figure 5. Ray density maps at different depths. Black and white scale indicates the number of rays normalized over a 4 by 4 degrees area.

[52] The map at 150 km depth shows high seismic velocities associated with the thick lithosphere of Archean and Proterozoic cratons and low seismic velocities beneath oceanic regions and most Phanerozoic continents. In these regions, low seismic velocities correspond to the low velocity layer associated with the asthenosphere. Exceptions are observed in the Andes where active subduction is present, and beneath Tibet, where continental collision of Asia and southern Eurasia favors underthrusting by a high velocity Indian mantle [*Priestley et al.*, 2006]. The signature of continental roots fades away around 200 km depth suggesting a seismic lithosphere on the low end of what has been estimated for the thermal cratonic lithosphere (thicknesses of 200–300 km are usually proposed [e.g., *Jaupart and Mareschal*, 1999]), but in agreement with what is deduced from gravity models [*Lestunff and Ricard*, 1995].

[53] Between 250 km depth and down to the top of the transition zone, the correlation with surface tectonics is lost (Figure 7). However, within the transition zone, a pattern of high seismic velocities associated with subductions around Indonesia and the Pacific, and in the Mediterranean region is retrieved. This suggests some "ponding" of the slabs within the transition zone, producing a broad scale high velocity signature picked-up by long period surface waves.

[54] Figure 8 depicts the root mean square (RMS) of velocity perturbations as a function of depth. The strongest RMS perturbations are between 100 and 150 km depths, where DR2012 correlates well with surface tectonics. These large RMS perturbations are due to the strong high velocity signature of cratons contrasting the strong low velocity signature of oceanic and Phanerozoic asthenosphere. The RMS of velocity perturbations decreases by a factor of 3.5 between 150 and 250 km and stabilizes at greater depths. This abrupt decrease is also observed in other seismic models [*Kustowski et al.*, 2008]. It likely marks the base of the continental lithosphere.
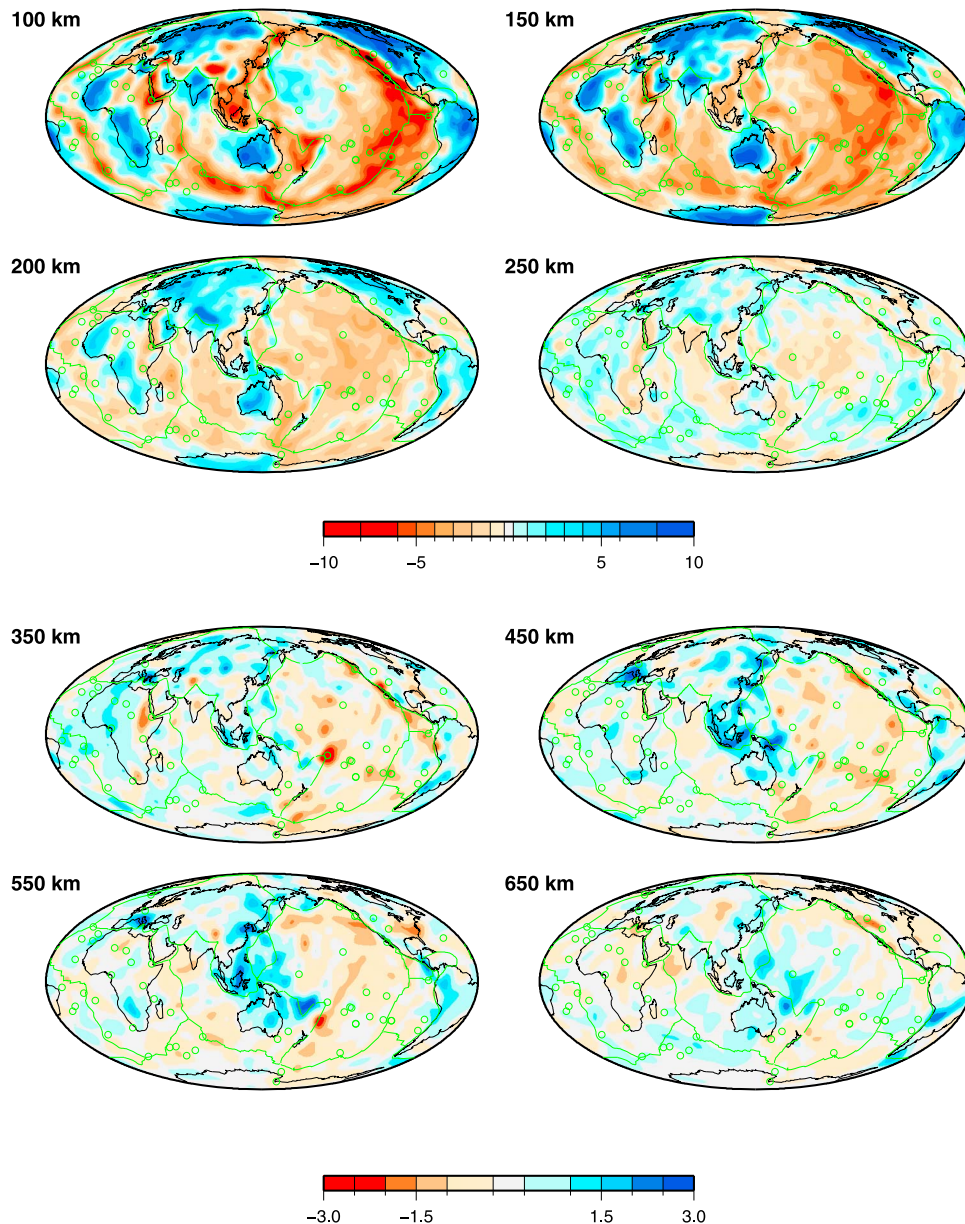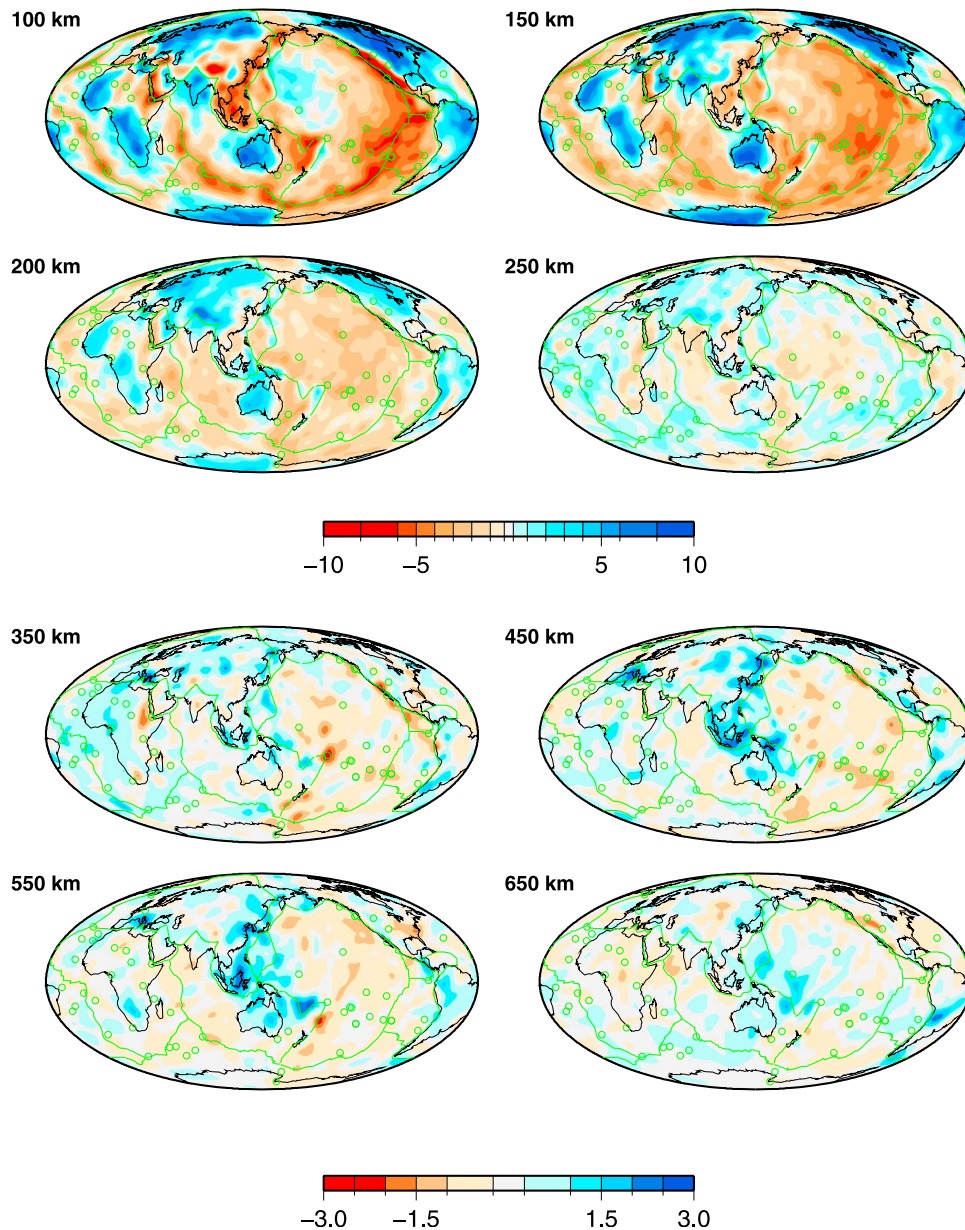
Figure 6. SV velocity distribution at different depths for an isotropic inversion. Perturbations from the reference velocity in percent are displayed by color coding. The velocity varies from −10% to +10% from the average value in the uppermost 250 km. At greater depths, shear velocity perturbations are between −3% to +3% to emphasize smaller contrasts.

[55] In Figure 9 we plot the age-dependent average cross-section of DR2012. The oceanic part of Figure 9 was created by taking sliding window averages of our SV-velocity distribution along the *Müller et al.* [2008] isochrons. SV-velocity contours deepen progressively with age and follow approximately the trend predicted by the square root of age cooling model [*Turcotte and Schubert*, 2002]. This observation confirms with an up-to-date global dataset a result suggested by most previous surface wave studies of lithospheric cooling [e.g., *Forsyth*, 1977; *Zhang and Tanimoto*, 1991; *Zhang and Lay*, 1999; *Maggi et al.*, 2006a]. However, for ages larger than 100 Myrs, some flattening of the isotherms may be seen as predicted by models where a

constant heat flux is provided at depth [*Doin and Fleitout*, 1996; *Goutorbe*, 2010].

[56] The continental part of Figure 9 consists in three average 1D profiles of SV-velocity, for Phanerozoic, Proterozoic and Archean tectonic provinces. Tectonic provinces are defined after the 3SMAC model of *Nataf and Ricard* [1995]. These 1D profiles are shown as absolute velocities (Figure 9, right) and as SV-velocity perturbations (Figure 9, left). Continental profiles are only shown at depths greater or equal to 100 km, where the effect of crustal correction is negligible. Phanerozoic provinces regroup heterogeneous regions including active orogens (e.g. Tibet and the Andean Cordillera) and stable Paleozoic lithosphere. For this reason,

Figure 7. Same as Figure 6 but for an inversion including azimuthal anisotropy. Comparison with Figure 6 shows that the bias due to the presence of azimuthal anisotropy is small. This SV velocity distribution corrected from this small bias, is our new seismic model DR2012.

the 1D Phanerozoic profile is difficult to interpret. On average, Archean lithosphere is faster and slightly thicker than Proterozoic lithosphere. Figure 9 confirms that the seismic signature of cratons has entirely vanished at 250 km depth.

5. Discussion

5.1. Improvements Due to the New Automated Scheme

[57] In this section, we compare the new scheme with the automated waveform modeling of *Debayle* [1999]. 63,514 waveforms have been matched successfully using both approaches. The resolutions matrix **R** is computed after each waveform inversion using equation (8). The trace of **R** gives the number of independent parameters extracted

by each inversion. Table 1 reports the trace of **R** averaged for events belonging to different depth ranges. For all depth ranges, the new scheme extracts more information. This additional information is due to a wider period range (up to 250 s), the inclusion of 5 higher-modes instead of 4 in *Debayle* [1999], and the selection of more secondary observables on each modal envelope. The smallest improvement (0.32) is obtained for shallow events (0–50 km) which concentrate their energy in the fundamental mode of surface waves. For deep events (200–700 km) which have most of their energy in the higher-modes, 0.63 additional degree of freedoms are extracted. The largest improvement (0.96) is obtained for intermediate depth events (50–200 km), for which both fundamental and higher-modes are well excited.
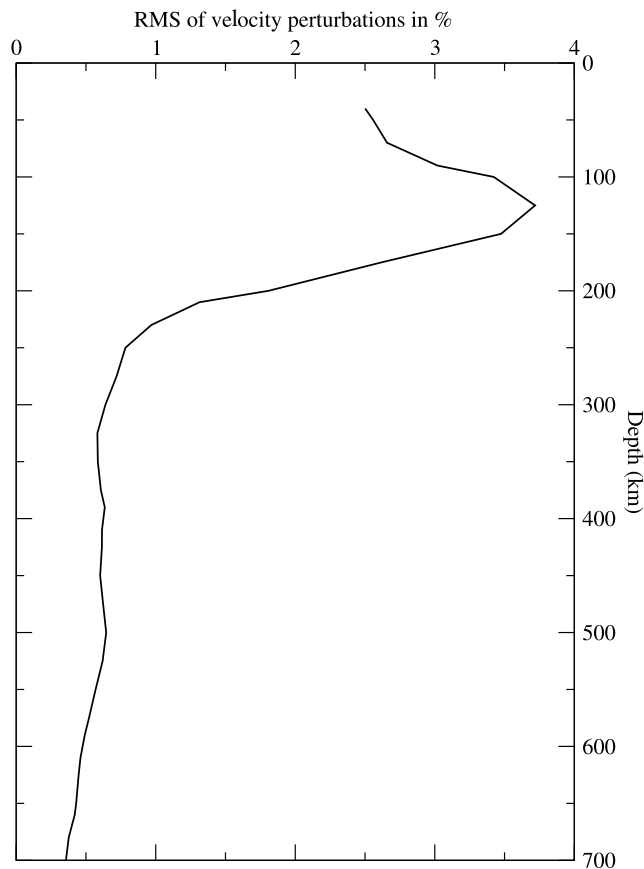
Figure 8. RMS of velocity perturbation in % computed within DR2012 as a function of depth.

Table 1. Average Amount of Independent Information Extracted With the New Scheme Against the *Debayle* [1999] Algorithm

|	Event's Depth Range			
	0–50 km	50–200 km	200–700 km	0–700 km
Number of waveforms	43,868	16,559	3087	63,514
Debayle [1999]	4.43	4.93	5.86	4.60
New scheme	4.75	5.89	6.49	5.13
Improvement	0.32	0.96	0.63	0.53

[58] According to equation (6), two points z_1 and z_2 distant by L and $2L$ are *a priori* correlated with values of 0.60 and 0.36 respectively. If we consider, as a rule of thumb, that two depths distant by $2L$ are "independent", extracting one more independent parameter means that we constrain one additional layer with an approximate width of $2L$. The width of this layer would be 100 km for $L = 50$ km.

[59] Figure 10 compares DKP2005 [*Debayle et al.*, 2005] and DR2012 at 150 and 500 km depth. Both models use the same vertical ($L = 50$ km) and horizontal ($L = 400$ km) smoothings. DKP2005 was built from 100,777 waveforms analyzed using *Debayle*'s [1999] approach. This model was only published for its shallowest 400 km part and the map at 500 km has been only produced for the purpose of the present study.

[60] Figure 10 shows that the new scheme improves mapping of the transition zone. At 500 km depths, DR2012 displays high seismic velocities associated with subductions, around the Pacific plate and in Europe. This pattern is in good agreement with others seismic models (see section 5.2). In contrast, DKP2005 is patchy at 500 km. This proves that the new approach better extracts the long wavelength seismic heterogeneities at depth, due to a more efficient selection of
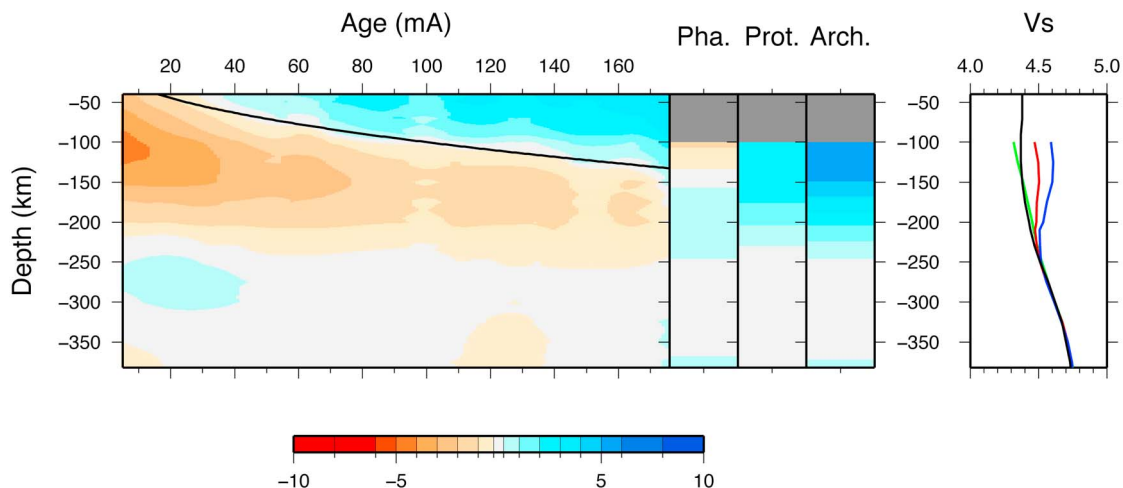
Figure 9. (left) Cross-section with respect to age for oceanic regions and Phanerozoic (labeled Pha.), Proterozoic (labeled Prot.) and Archean (labeled Arch.) continental provinces. Color coding shows perturbations in percent from the reference SV velocity profile displayed in black in Figure 9 (right). For oceanic regions, a smoothed image was created by averaging SV velocity along the *Müller et al.* [2008] isochrons, using a sliding age window of 10 Ma width. The continuous black line indicates the position of the thermal boundary layer for the half-space cooling model as defined in, e.g., *Turcotte and Schubert* [2002]. (right) Reference (black), Paleozoic (green), Proterozoic (red) and Archean (blue) absolute SV velocity profiles.
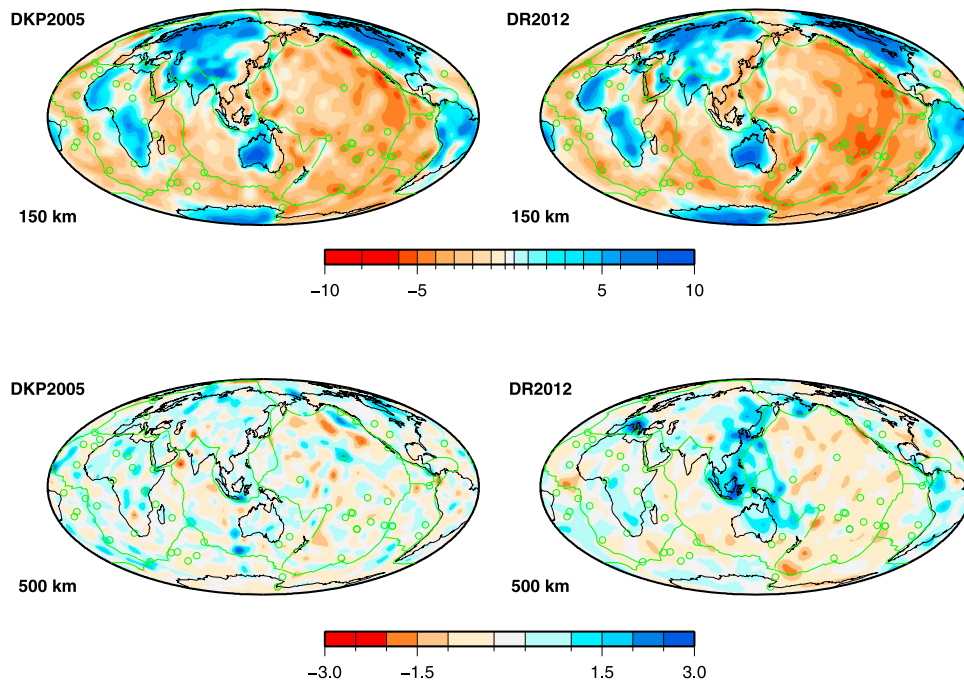
Figure 10. SV-wave velocity distribution at (top) 150 km and (bottom) 500 km depth in DKP2005 by *Debayle et al.* [2005] compared with maps at the same depths in DR2012. SV-wave perturbations are in per cent relative to PREM.

robust secondary observables. The greater benefit is obtained for higher-modes, which often have a lower signal to noise ratio compared to the fundamental mode. At 150 km depth, DKP2005 and DR2012 are very consistent although DR2012 has a slightly larger amplitude. Here both approaches extract very well the fundamental mode of surface waves. However, some new features in DR2012 show a better agreement with surface tectonics (e.g. the slow velocities beneath Hawaii [*Wolfe et al.*, 2011] or the broad region of slow velocities beneath the South Pacific [*Tanaka et al.*, 2009]).

[61] We have performed various tests indicating a systematic improvement of the resolution of DR2012 with respect to DKP2005 due to both an increase of the number of analyzed seismic paths by a factor 3.5 and a more accurate waveform modeling. We already published synthetic experiments for subsets of our data set for regions including Horn of Africa [*Debayle et al.*, 2001], Antarctica [*Sieminski et al.*, 2003], North Atlantic and surrounding regions [*Pilidou et al.*, 2005], Australia [*Debayle et al.*, 2005], Asia [*Priestley et al.*, 2006] and Africa [*Priestley et al.*, 2008]. These tests confirm that we are able to isolate upper mantle seismic velocity anomalies located at depths larger than 250 km. Rather than repeating here these tedious synthetics, we find it more interesting to compare DR2012 with other seismic models obtained using different modeling approaches and different datasets. We show in the next section that the pattern of seismic heterogeneities in DR2012 is consistent with other S-wave models up to degree 40 within the uppermost 200 km of the mantle and up to degree 10 within the transition zone.

5.2. Correlation With Other Seismic Models

[62] We compute correlations between DR2012 and three other seismic models: S40RTS by *Ritsema et al.* [2011],

S362ANI by *Kustowski et al.* [2008] and HMSL by *Houser et al.* [2008].

[63] S40RTS is an isotropic S-wave model based on three data sets: fundamental and higher mode Rayleigh waves, spheroidal normal modes and teleseismic travel times of long period body waves. Teleseismic travel times are mostly SH-type waves measured on the transverse components (except for SKS phases). These waves, which generally have steep incidence in the upper mantle, are essentially affected by the lower mantle structure. The upper mantle is largely constrained by Rayleigh wave fundamental and higher modes and spheroidal normal modes, which are sensitive to SV. However, some SH-type body wave phases like SS, SSS and SSSS have a significant portion of their paths in the upper mantle. In a transversely isotropic medium with a vertical axis of symmetry like PREM, SH-polarized waves traveling nearly vertically experience no splitting and are sensitive to SV. However, for typical S-wave wavelengths of few tens of kilometers, the preferred orientation of anisotropic minerals generally produces azimuthal anisotropy which is better represented by a transversely isotropic medium with an horizontal symmetry axis. Although this azimuthal anisotropy is averaged out in a global model like PREM, most SH-type waves traveling nearly vertically in the upper mantle may provide a different information than Rayleigh waves and spheroidal modes. We assume that the difference in parameter sensibilities of these two data sets was accounted for in S40RTS and we consider S40RTS as essentially an SV model in the upper mantle.

[64] S362ANI includes surface wave phase anomalies, long period waveforms and body wave travel times. The upper mantle part of S362ANI is primarily constrained by surface waves and their higher-modes. Higher-modes are
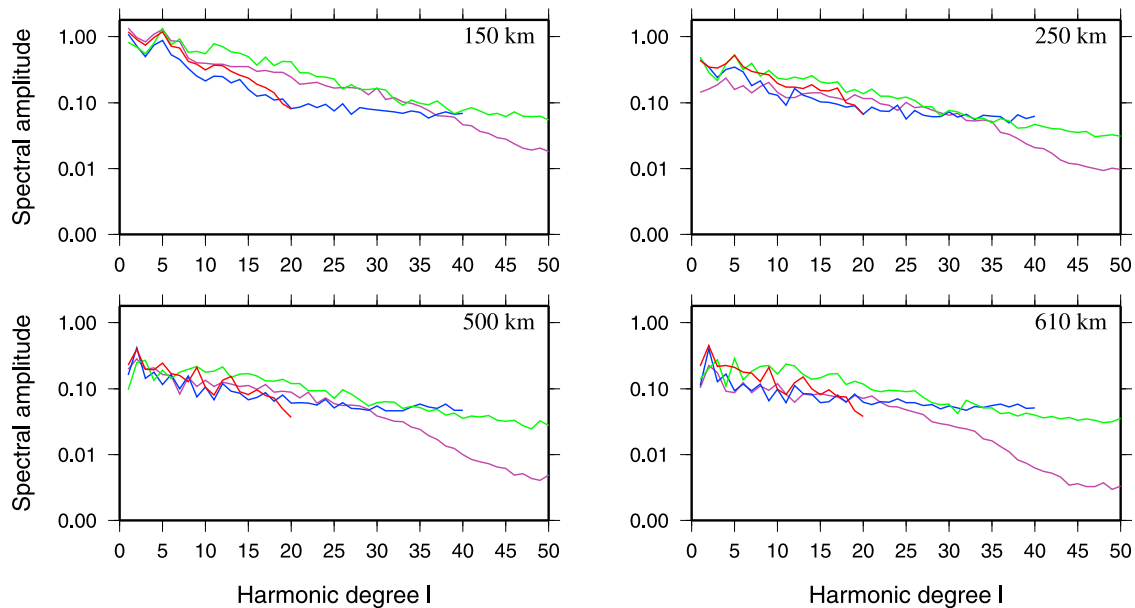
Figure 11. Spectral amplitude as a function of degree for DR2012 (purple line), S40RTS [*Ritsema et al.*, 2011] (blue line), HMSL [*Houser et al.*, 2008] (green line) and S362ANI [*Kustowski et al.*, 2008] (red line) at depths of 150, 250, 500 and 610 km.

included in S362ANI through the inversion of long period mantle (>125 s) and body (>50 s) waveforms. S362ANI includes measurements from transverse, longitudinal and radial components and allows for the presence of radial anisotropy in the upper 400 km of the mantle. We compare our SV model with the SV-wave part of S362ANI.

[65] HMSL is an isotropic S and P velocity model based on surface waves, long period body wave travel times and differential body wave travel times. The upper mantle S-wave velocity structure of HMSL is mostly constrained by Rayleigh and Love phase velocities with minor contribution from body waves. However, HMSL does not incorporate radial anisotropy, a difference with other models, which are either based on the analysis of Rayleigh waves only (S40RTS and DR2012), or incorporate radial anisotropy to account for the Love-Rayleigh discrepancy (S362ANI). This may explain some differences with other models at the top of the upper mantle, although the use of longer period surface waves (66–250 s) in HMSL mitigates its sensitivity to shallow radial anisotropy.

[66] S40RTS, S362ANI and HMSL span the range of possible data (surface waves fundamental and higher modes, body waves, normal modes) which can be used for mapping upper mantle S-wave heterogeneities. They also span a wide range of possible model parameterizations (spherical harmonics, splines, blocs, Gaussian correlations) including radial anisotropy or not. Other recent seismic models agree well with at least one of these models. For example, SAW642AN by *Panning and Romanowicz* [2006] is based on a data set comparable to that of S362ANI. This model correlates well with S362ANI and S40RTS, both for the depth extent of seismic heterogeneities and for their long wavelength pattern [*Kustowski et al.*, 2008; *Ritsema et al.*, 2011]. We believe that S40RTS, S362ANI and HMSL summarize our current knowledge of the large scale S-wave upper mantle structure.

In addition, although these models incorporate datasets which are not present in DR2012, their upper mantle part is in all case largely constrained by surface waves.

[67] Figure 11 shows the spectra $S_A^2(l) = \sum_{m=-l}^{l} A_l^m A_l^{m*}$ of these tomographic models at 150 km, 250 km, 500 km and 610 km (A_l^m are the spherical harmonic coefficients at degree l and azimuthal order m of the tomographic model A, and $*$ denotes the complex conjugate). S362ANI is clearly the smoothest model. It uses 362 spherical splines, equivalent to a spherical-harmonic degree 18 expansion. For this reason, we do not show spectra or correlations for $l > 20$ for this model. S40RTS is only provided up to degree 40. DR2012 and HMSL show a gentle decrease of spectral amplitude. This decrease is more pronounced at high degrees in the case of DR2012 which does not include body waves. S40RTS seems to favor the even degree anomalies at depth, which is likely due to the contribution of the normal modes in the model.

[68] We use the following relationship to compute the correlation between two seismic models A and B:

$$C(l) = \frac{\sum_{m=-l}^{l} A_l^m B_l^{m*}}{S_A(l) S_B(l)} \quad (10)$$

Figure 12 displays the correlations $C(l)$ computed at 150 km, 250 km, 500 km and 610 km depths between DR2012 and the three other seismic models. At 150 km depth, correlations are representative of those obtained in the uppermost 200 km. At 250 km depth they are representative of those obtained between 200 and 400 km, while at 500 and 610 km depths, they represent correlations in the upper and lower parts of the transition zone.

[69] At 150 km depth, the correlation between DR2012 and S362ANI is above the 95% confidence level up to degree 20.
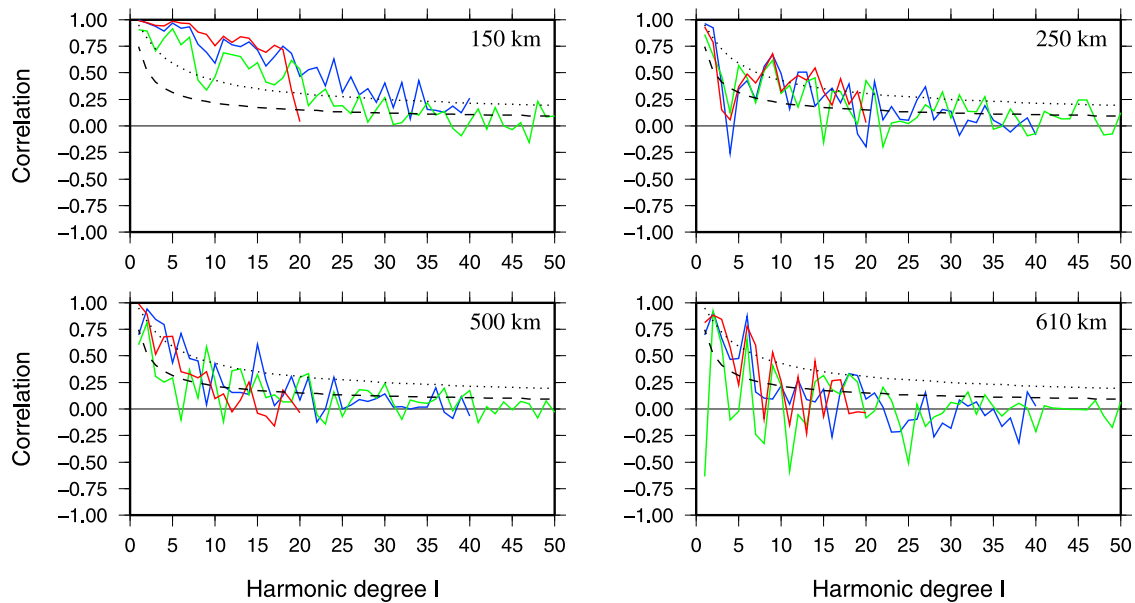
Figure 12. Correlations as a function of degree between DR2012 and S40RTS by *Ritsema et al.* [2011] (blue line); HMSL by *Houser et al.* [2008] (green line); S362ANI by *Kustowski et al.* [2008] (red line). Correlations are computed at 150, 250, 500 and 610 km. In each plot the dotted and dashed lines indicate the 95% and 66% significance levels respectively.

The agreement between DR2012 and S362ANI is therefore very good over the range of wavelengths resolved by S362ANI. At 150 km depth, the correlation with HMSL is above the 95% confidence level up to degree 22, and stays above the 66% confidence level up to degree 28. This is a fair result, considering that HMSL is based on a smaller number of surface wave data (42,000 Rayleigh and 14,000 Love waves) analyzed at longer periods (>66 s for HMSL compared to >50 s in DR2012). S40RTS provides the best agreement with DR2012 at 150 km depth. This is not surprising as the data sets used to constrain the uppermost mantle in both models are the most similar (several hundred of thousands of Rayleigh waveforms analyzed in the period range 50–250 s). In addition both inversion approaches, although different, allow for seismic heterogeneities up to wavelengths of 1000 km. S40RTS and DR2012 correlate above the 95% confidence level up to degrees 30–35 and remain above the 66% confidence levels up to degree 40. We consider this as a very encouraging result. It suggests that in spite of different approaches to extract information from the seismograms and different strategies of inversion, it is now possible to build global tomographic models consistent up to degree 40 in the upper 200 km of the mantle.

[70] At 250 km depth within the region just below the lithosphere, DR2012 shows a weak correlation with other models even at low degree (Figure 12). This correlation stays above the 66% level only for a limited range of harmonic degrees (1–2 and 8–15 or 8–20 depending on the seismic model). For degrees 3 to 7, DR2012 is uncorrelated with the other seismic models. Therefore the pattern of seismic anomalies in DR2012 differs from the other models, even at long wavelengths. We show in the next section that DR2012 can be reconciled with other models by using a larger vertical degree of smoothing.

[71] Within the transition zone, correlations between DR2012 and the other seismic models are only significant at very long wavelengths (Figure 12). At a depth of 500 km, DR2012 correlates with S40RTS at the 95% confidence level up to degree 8, falls below the 66% level for degree 9, then increases and remains above the 66% confidence up to degree 17. The correlation with S362ANI stays above the 66% confidence up to degree 9 and the correlation with HMSL is poor. At the depth of 610 km, the correlations curves show strong oscillations, likely due to oscillations present in the spectra at degrees smaller than 20. These oscillations, like those of the amplitudes (see Figure 11), suggest that the different models have a different degree-dependent sensibility at depth. S362ANI and S40RTS correlate with DR2012 above the 66% confidence level up to degree 8. Between degrees 8 and 15, the correlations oscillate around the 66% level curve. Correlation curves with HMSL have strong oscillations and indicate a general poor agreement.

[72] For completion, we discuss in Appendix B correlations $C(l)$ obtained using successively S362ANI, S40RTS and HMSL as a reference model.

5.3. Effect of Vertical Smoothing

[73] Figures 13 (top) and 13 (middle) depict DR2012 and S40RTS at 250 km depth. Although DR2012 displays a strong correlation with surface tectonics at shallower depths (Figure 7), this correlation is already lost at 250 km depth. This correlation still holds for S40RTS, where high seismic velocities beneath stable continents contrasts with low seismic velocities beneath most oceanic regions. At 250 km depth, HMSL and S362ANI (not shown) are similar to S40RTS and display the same correlation with surface tectonics. In DR2012, vertical smoothing is controlled by a

DR2012 Lc=50 km

S40RTS

DR2012 Lc=100 km
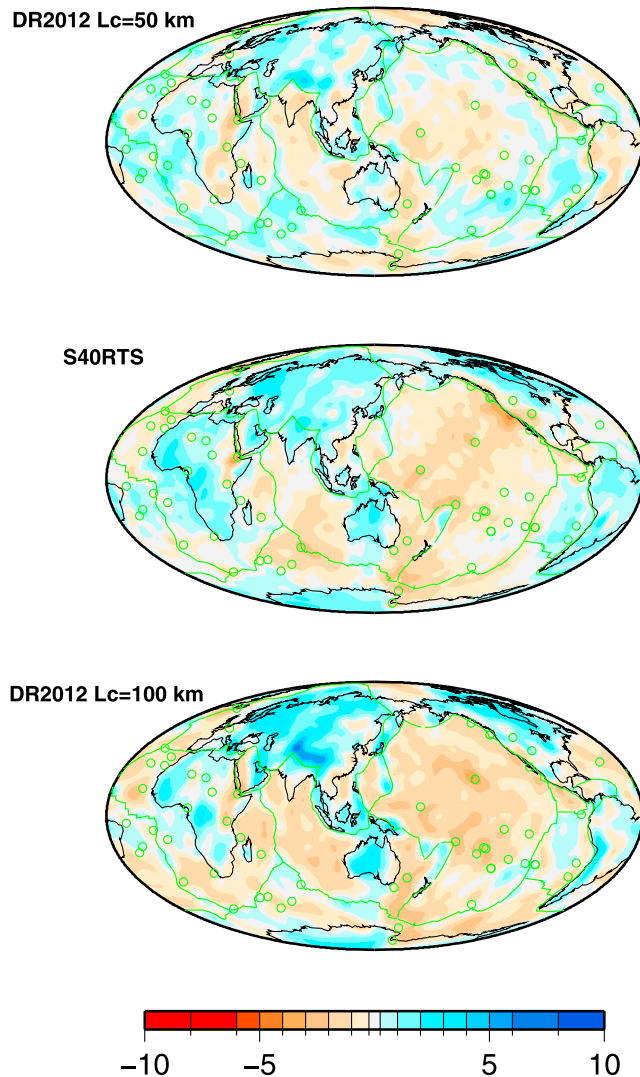
Figure 13. SV-wave velocity distribution at 250 km depth (in per cent relative to PREM) obtained using a vertical correlation length of (top) 50 km and (bottom) 100 km as compared with (middle) the SV wave distribution at the same depth in S40RTS by *Ritsema et al.* [2011]. A longer correlation length enforces the similarity between S40RTS and DR2012.

vertical correlation length (equation (6)) set to $L = 50$ km in the waveform inversion. We tried a new waveform inversion of our entire dataset using a value of $L = 100$ km. All other a priori parameters and criteria to accept or reject a waveform remain the same. With $L = 100$ km, the number of records which pass successfully all stages of the waveform inversion decreases to 348,995 (374,897 records were successful using $L = 50$ km). Therefore, smoother path average models are statistically less able to fulfill the same criteria of waveform fit and of convergence towards a unique 1D model. Figure 13 (bottom) shows the pattern of seismic heterogeneities at 250 km depth obtained using $L = 100$ km which is now much more similar to S40RTS (Figure 13, middle) than with $L = 50$ km (Figure 13, top). At shallower depths, the $L = 100$ km model is very close to

DR2012. It is clear that imposing *a priori* a smoother vertical model is a simple way to reconcile our seismic model with other models. However, this is done at the price of a 7% reduction in the number of waveforms that can be explained. This is not dramatic, but we argue that this does not favor the smoothest model. Furthermore, the trace of the resolution matrix of the waveform inversion (see Table 1, last column) is consistently larger than 5 with only a weak dependence to L. Our data set is therefore able to be inverted for more than five layers in the shallowest upper mantle. A correlation length $L = 50$ km is thus a fair choice to characterize the first $2L \times tr(R) \geq 500$ km of the mantle. We therefore consider DR2012, built using $L = 50$ km as our preferred model, and the difference with other models around 250 km depth is mostly due to the slightly thinner cratons of our model.

[74] Figure 14 displays the correlation at 250 km between our smooth model, $L = 100$ km, and the three other seismic models. This figure confirms that increasing the vertical smoothing is a simple way to reconcile our model with other models. Our smoother model now correlates above the 95% confidence level up to degree 20 with S362ANI and S40RTS and up to degree 10 with HMSL.

5.4. Possible Effects of Radial Anisotropy on the Thickness of Cratons

[75] DKP2005, our previous global tomographic model [*Debayle et al.*, 2005], was built using an isotropic version of PREM as starting model. In DR2012, the initial synthetic seismogram includes the radial anisotropy of PREM. The depth extent of cratons is similar in DKP2005 and DR2012, suggesting that accounting for the radial anisotropy of PREM with Rayleigh waves only has little effect.

[76] Inverting for radial anisotropy in the uppermost mantle would require a simultaneous analysis of Love and Rayleigh waves. From our experience, this would not change significantly the thickness of the SV-wave high velocity lid. In *Debayle and Kennett* [2000] we inverted simultaneously a set of Love and Rayleigh waves and found significant radial anisotropy with SH faster than SV down to
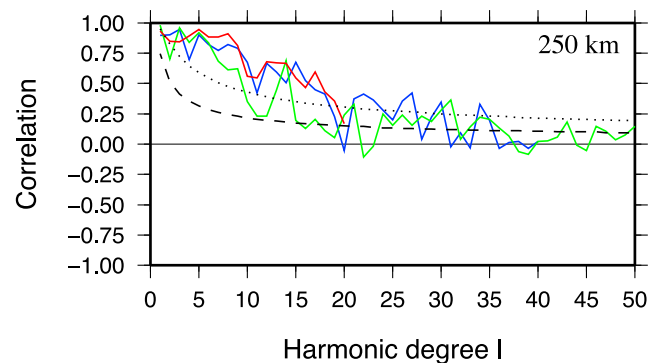
Figure 14. Correlations at a depth of 250 km between a "smoothed" version of DR2012 (using a vertical correlation length of 100 km in the waveform modeling) and S40RTS by *Ritsema et al.* [2011] (blue line), S362ANI by *Kustowski et al.* [2008] (red line) and HMSL by *Houser et al.* [2008] (green line).
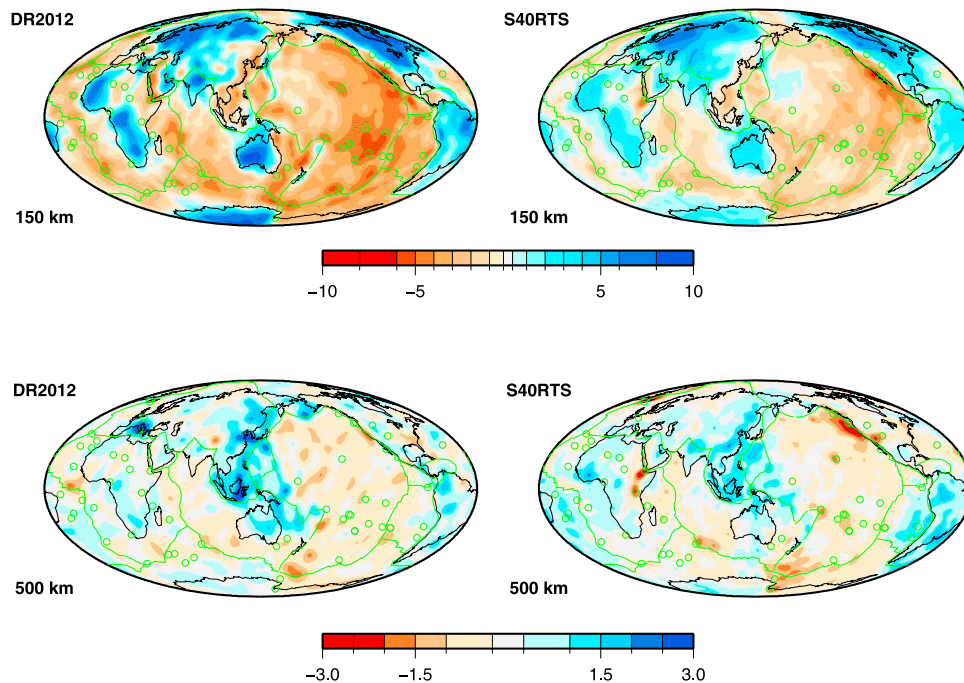
Figure 15. SV-wave perturbations (in per cent relative to PREM) at (top) 150 km and (bottom) 500 km depths in DR2012, compared with maps at the same depth in S40RTS by *Ritsema et al.* [2011].

about 200 km beneath Australia, but no clear difference in the thickness of the SV-wave high velocity lid compared to an isotropic inversion.

[77] Finally, notice that models S40RTS, HMSL and S362ANI all show a thicker cratonic lid than DR2012, although S40RTS results from isotropic inversion of Rayleigh waves, HMSL from isotropic inversion of both Love and Rayleigh waves and S362ANI from anisotropic inversion of Love and Rayleigh waves. We conclude that the limited depth extent of the cratonic lithosphere in our model is not a consequence of non inverting for the radial anisotropy.

6. Concluding Remarks

[78] Our extension of the *Debayle* [1999] automated inversion algorithm further divides the waveform modeling into elementary steps. This gives more flexibility to the code, allowing a better selection of the robust information and a better extraction of higher-modes.

[79] Using the new scheme, we matched successfully 374,897 fundamental and higher-modes Rayleigh waveforms which provide a global coverage of the upper mantle. For each waveform, the inversion extracts a path average elastic and anelastic model and the corresponding fundamental and higher mode dispersion and attenuation curves. The dispersion and attenuation curves provide a dataset for future finite-frequency inversion.

[80] Compared to DKP2005, our new 3D SV-wave velocity model DR2012 represents a significant improvement in the transition zone (see Figure 10). In the uppermost 200 km of the mantle, some new features of DR2012 show a better agreement with surface tectonics, suggesting that the new scheme and the increased number of data, both

contribute to the improvement of SV-wave mapping at shallow depths.

[81] A comparison of DR2012 with three other recent seismic models (S40RTS, S362ANI and HMSL), emphasizes that in the uppermost 200 km of the mantle, all models correlate above the 95% confidence level up to harmonic degree 20. DR2012 and S40RTS which are the most similar in terms of datasets and of horizontal wavelengths allowed in the inversion, are also the most similar models. These two models correlate above the 95% and 66% confidence levels up to degree 35 and 40 respectively. This is an encouraging result, proving that in spite of different approaches to extract information from the seismograms and to invert for this information, it is now possible to build global tomographic models consistent up to degree 40 in the uppermost 200 km of the mantle.

[82] The region between 200 km depth and the top of the transition zone is an intermediate region where the very strong heterogeneities present in the upper 200 km progressively vanish. Continental roots disappear between 200 and 250 km in DR2012, at a shallowest depth than in other models. It is possible to obtain thicker continental roots in DR2012 by imposing *a priori* smoother vertical variations. This does not seem required by the resolution of the waveform inversion and leads to a less successful waveform inversion.

[83] Within the transition zone, correlations between DR2012 and other models are significant only up to degree 10. In this depth range, all models image a broad scale high velocity signature around Indonesia and the Pacific, suggesting some "ponding" of the slabs.

[84] To illustrate the agreement between two well correlated models, we show in Figure 15 DR2012 and S40RTS
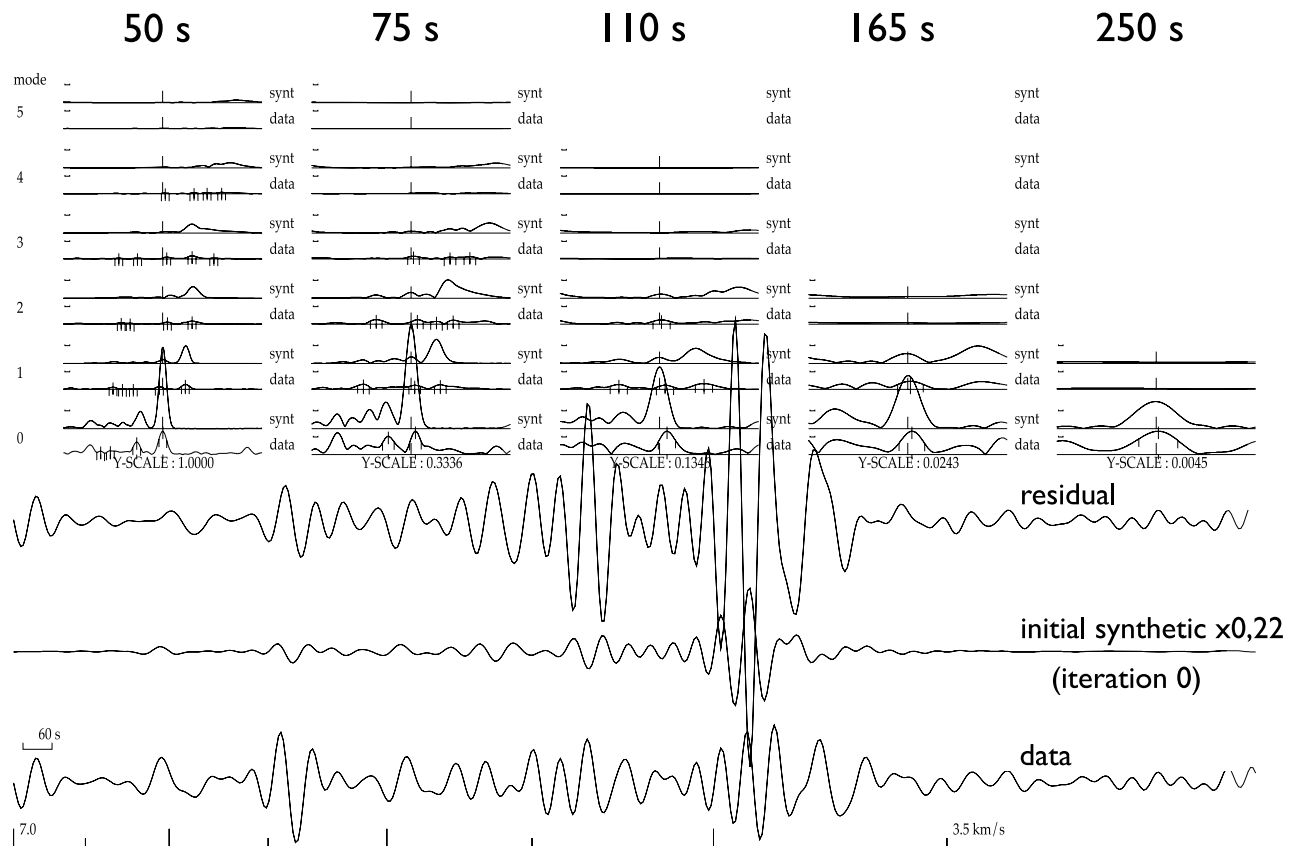
Figure A1. Situation before waveform inversion for the February 13, 2008 event recorded at station CAN (Canberra, Australia). The lower part of the figure shows: the observed seismogram (labeled "data"), the synthetic seismogram computed for the initial model ("initial") and their difference ("residual"). The upper part shows the envelope of the filtered cross-correlogram functions $g_p(\omega_q, t)$ for modes ranging from the fundamental mode (0) up to the fifth overtones and for filters centered from 50 (left) to 250 s (right) periods. For each mode and each period the lower functions are the envelope of the actual cross-correlogram functions (labeled "data") while the upper functions are the envelope of the synthetic cross-correlogram functions (labeled "synt"). The vertical marks on the envelopes indicate where the actual envelopes are sampled for the inversion.

at 150 and 500 km depth. Although at 150 km depth, the two models agree above the 95% confidence level up to degree 35, differences of few percents persist over wavelengths greater than 1200 km both in continental (e.g. East Asia) and oceanic (e.g. South Pacific) regions. Within the transition zone, small scale heterogeneities are generally not correlated. By applying more sophisticated finite frequency theory on the datasets which have been used to build the current seismic models, we may add more details in the models, but it is unlikely that we will remove their differences at long wavelengths where the ray theory remains an excellent approximation.

[85] For these reason, we believe that we first need to improve current seismic models. For DR2012, a first step is to complete our surface wave dataset with available long period S-wave datasets [e.g., from *Zaroli et al.*, 2010] and with normal mode measurements in order to map the entire mantle. Then, by exploiting our new set of dispersion and attenuation curves, we hope to built consistent 3D shear velocity and quality factor models. This will contribute to

disentangle the thermal and compositional structures of the mantle.

Appendix A: Description of the New Automated Scheme

A1. Automatic Selection of the Secondary Observables

[86] In this section we describe how we sample the modal envelopes at a given period to improve higher-mode extraction.

[87] An example is presented in Figure A1 for the event of February 13, 2008, located in Oaxaca, Mexico (16.35°S; −94.51°E) and recorded at station CAN in Canberra, Australia. For this intermediate depth event (87.1 km), a dominant fundamental mode is observed between 3.5 and 4 km s^{-1} on the actual signal (bottom part of the figure), while the less energetic overtones have higher group velocities. The initial synthetic seismogram is more energetic (it has been multiplied by a factor of 0.22 in Figure A1) and clearly the residual is huge.
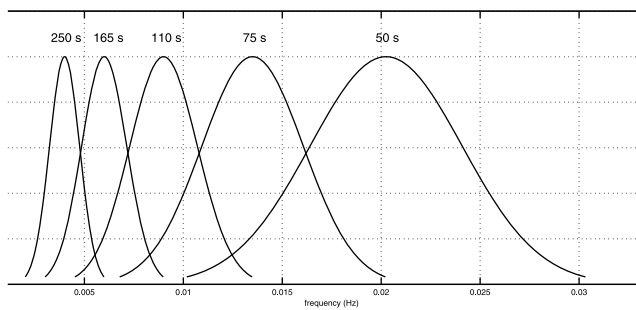
Figure A2. Five gaussian filters $h(\omega_q, t)$ used for the waveform inversion. The width and central period $(1/\omega_q)$ are chosen to cover the period range 50–250 s.

[88] In the top part of Figure A1, the modal envelopes of the synthetic (labeled "synth") and actual (labeled "data") cross-correlograms are shown for the 5 periods corresponding to the central periods of the gaussian filters $h(\omega_q, t)$ and for the number of modes N_{mode} considered at each period (two modes are enough for an accurate description of the signal at 250 s, six modes are needed at 75 and 50 s). The actual envelopes result from the cross-correlation between the actual seismogram and each mode of a synthetic seismogram computed for a reference, initial model (see equation (3)). The filters $h(\omega_q, t)$ are gaussian in frequency and their width at −30 dB (3% of the maximum) is equal to their central frequency ω_q. The spacing of the central frequencies is equal to half the gaussian width at −30 dB, so that the following sequence of central periods is chosen: 250, 165, 110, 75 and 50 s. This choice allows us to cover the period range 50–250 s with weakly correlated secondary observables (Figure A2), so that the data covariance matrix can be considered as diagonal.

[89] For mode 0, the actual seismogram is cross-correlated with the fundamental mode of the reference synthetic seismogram. At 250 s period, higher modes are poorly excited. The actual envelope has a single lobe associated with the most energetic fundamental mode. At shorter periods, the actual envelopes have several maxima corresponding to different alignments of the cross-correlated signals. The largest maximum on each actual envelope is produced by the cross-correlation of the dominant fundamental mode of the actual and synthetic seismograms. It is found in the center of the diagram, close to the reference time $t_0 = 0$ for which there is no delay between the cross-correlated signals. This is due to the small difference in arrival time between the fundamental mode of the synthetic and actual signals. The less energetic overtones of the actual seismogram produce secondary lobes with a delay time corresponding to the delay between the overtones of the actual seismogram and the fundamental mode of the reference synthetic.

[90] For modes $p = 1$ to 5, the actual seismogram is cross-correlated with mode p of the reference synthetic seismogram. The energetic fundamental mode of the actual seismogram still dominates the amplitude of the envelope, but the corresponding lobe is shifted. It is found at a delay time corresponding to the delay between the fundamental mode of the actual seismogram and the overtones of the synthetic. The less energetic overtones of the actual seismogram produce secondary lobes close to the time $t_0 = 0$

because the reference model gives a fair prediction of the arrival time of the overtones.

[91] Figure A1 shows that for multimode seismograms, the modal cross-correlogram functions have a complex shape, with several maxima related to the different modes present in the signal. In addition, when a mode j is clearly dominant in the signal (e.g. the fundamental mode for a shallow event) it often contaminates the $g_p(\omega_q, t)$ functions and provides the largest maximum of the envelope even when $p \neq j$. Therefore, by sampling three points on the largest maximum of each actual envelope, there is a risk of extracting redundant information on the dominant mode, while loosing the information on the other modes.

[92] To overcome this problem, *Debayle* [1999] over-parameterizes the cross-correlation information. When the shape of the envelope has several maxima, *Debayle* [1999] selects the two best lobes, using the ratio $A_{max}/|t_{max} - t_0|$ as a criterion, where A_{max} is the amplitude of the maximum located at time t_{max} and t_0 the reference time. This criterion favors the largest lobes close to t_0, where the current mode is expected. By selecting two lobes, *Debayle* [1999] reduces the risk of extracting redundant information on the dominant mode. The new scheme selects up to N_{mode} lobes, corresponding to the N_{mode} possible modes which interfere on each envelope.

[93] During steps 1 to N_{per} (where N_{per} is the number of selected band-pass filters as defined in section 3.1), the automated code matches the envelopes of the cross correlograms. At a given period, the code first extracts a number $N_{sl} \leq N_{mode}$ of significant maxima for each modal envelope. A maximum is considered to be significant when its amplitude is larger than $R_{max} = 6$ times the average of the minima of the envelope. The selected maxima are then ranked according a decreasing $A_{max}/|t_{max} - t_0|$ criterion. The inversion starts with all the first lobes of the modal envelopes. Three iterations are allowed to deal with the weak non-linearity of the secondary observables. Then a second lobe is added on each modal envelopes for which it is found to be significant and three new iterations are allowed to simultaneously match the first and second lobes. This process is repeated until all the significant lobes have been selected and inverted.

[94] At this stage, the automated code decides, based on the criteria given in the next section, whether it:

[95] 1. adds the next shorter period

[96] 2. adds iterations and follow the inversion of the current dataset to complete convergence (a maximum of 18 iterations is allowed to match the current period)

[97] 3. steps back and redo the inversion of the current period using a more severe criterion to select the envelopes maxima (the criterion R_{max} is incremented by 1).

[98] During steps $N_{per} + 1$ to $2N_{per}$ the code matches the instantaneous phase, taken at the time where the highest $A_{max}/|t_{max} - t_0|$ value has been found (i.e. when the first lobe of each envelope reaches its maximum). The algorithm has three iterations to match the phase at each period, so that $3 \times N_{per}$ iterations are required to match all phase data. Once all phase data have been matched, the waveform analysis is completed. The top part of Figure A3 shows the automated selection of the envelope secondary observables achieved at the end of step N_{per} for the 2008 February 13 event recorded at station CAN. The vertical marks on the actual envelopes
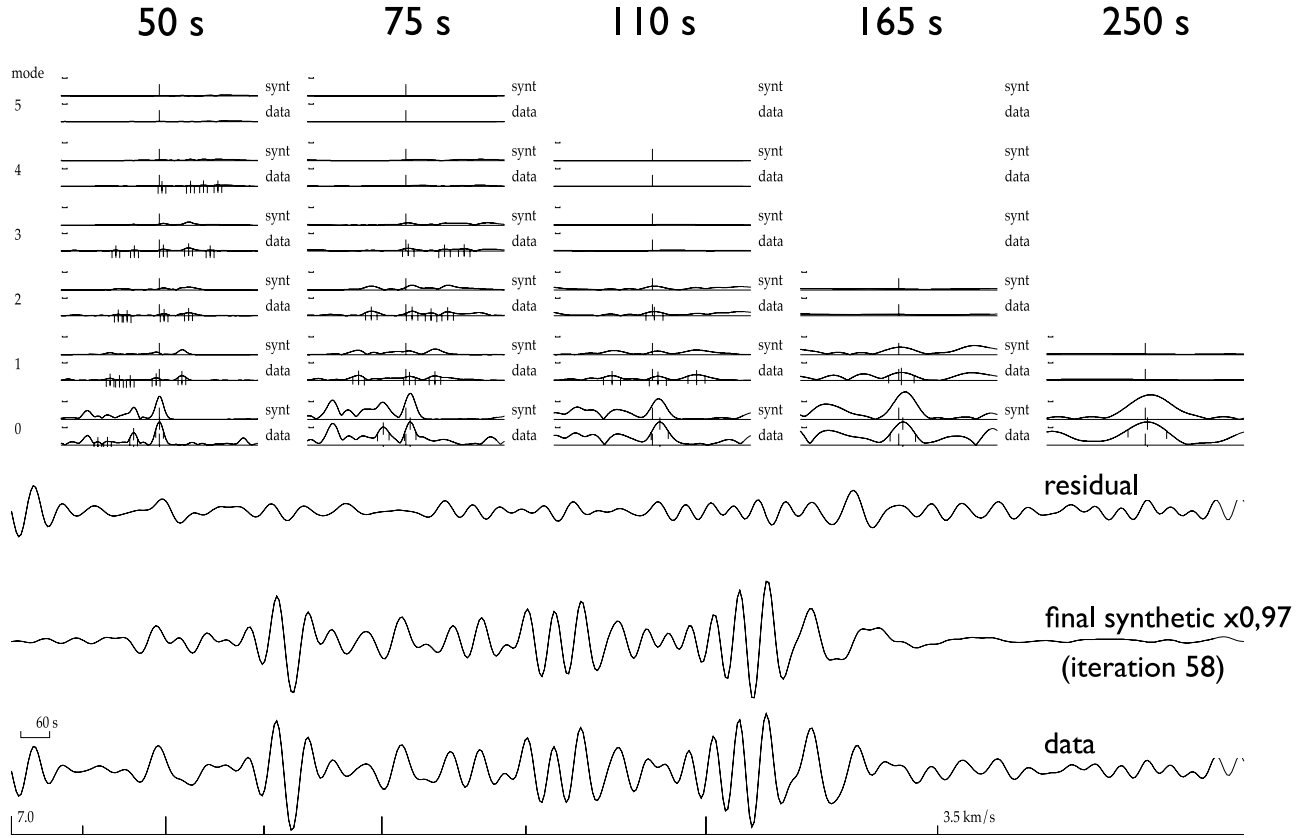
Figure A3. Same as Figure A1 but after the waveform inversion.

labeled "data" indicate the points where the cross-correlograms have been sampled. The bottom part of Figure A3 shows the observed, synthetic and residual seismograms at the last iteration of the waveform inversion process, once the phase of all cross-correlograms has been incorporated. All together, the inversion of the seismogram depicted in Figure A3 matches 42 peak maxima (each of one described by 3 values) and 15 peak phases (not shown), which add up to a total of 141 secondary variables. Our 3D seismic model DR2012, incorporates about 375,000 waveforms corresponding to a total of about 20 million observations.

A2. Evaluation of the Waveform Inversion

[99] We use several criteria to evaluate the quality of a waveform inversion:

[100] 1. A misfit criterion for the data:

$$\chi_d = \sqrt{\frac{1}{n}\sum_i^n \left(\frac{\hat{d}_i - d_{ci}}{\sigma_{di}}\right)^2}, \qquad (A1)$$

where $\hat{d}_i$ are the data (secondary observables) in number n, σ_{di} their standard deviation, and d_{ci} the predicted data.

[101] 2. A misfit criterion for the model:

$$\chi_m = \sqrt{\frac{1}{p}\sum_i^p \left(\frac{\hat{m}_i - m_{0i}}{\sigma_{mi}}\right)^2}, \qquad (A2)$$

where $\hat{m}_i$ are the model parameters in number p, σ_{mi} their a priori standard deviations, and m_{0i} the parameters describing the a priori model $\mathbf{m_0}$.

[102] 3. The ratio of energy between the residual and the actual seismogram:

$$R_1 = \frac{E_{res}^{fin}}{E_{actual}} \qquad (A3)$$

where E_{res}^{fin} is the energy of the residual signal at the final iteration and E_{actual} is the energy of the actual signal.

[103] 4. An energy reduction parameter:

$$R_2 = 1 - \frac{E_{res}^{fin}}{E_{res}^{init}} \qquad (A4)$$

where E_{res}^{init} is the energy of the residual signal between the observed and initial synthetic seismogram.

[104] Several conditions have to be fulfilled in order to progress through the different steps of the waveform inversion.

[105] 1. A seismogram is considered for waveform modeling if the initial synthetic predicts the amplitude of the real seismogram within a factor of 10. A poorer amplitude prediction should indicate inappropriate source parameters.

[106] 2. From steps 1 to N_{per}, we compute the misfit criterion χ_d at the end of each step for different subsets of the data. We compute χ_d^{tot} using all the secondary observables
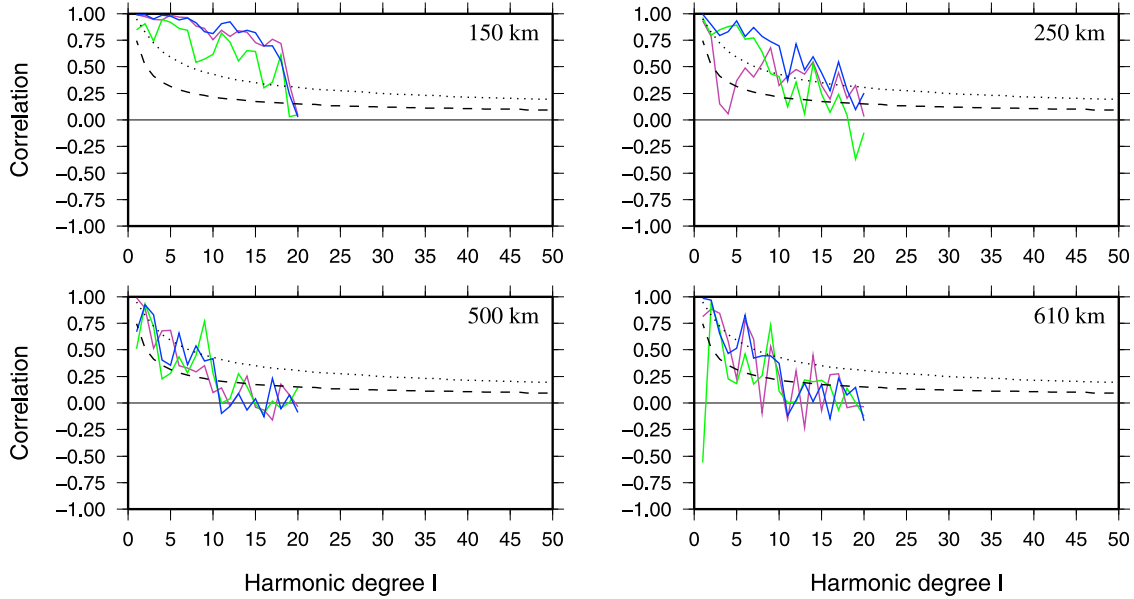
Figure B1. Same as Figure 12 but for correlations between S362ANI by *Kustowski et al.* [2008] and S40RTS by *Ritsema et al.* [2011] (blue line); HMSL by *Houser et al.* [2008] (green line); DR2012 (purple line).

from periods 1 to the current period and χ_d^{sp} using all the secondary observables of the last period processed (the shortest period) only. We also compute an individual misfit χ_d^{so}, for each secondary observables of the shortest period. We monitor χ_d^{max}, the largest value of χ_d^{so}. The process is allowed to start the following step only if the following misfit criteria is verified

$$\chi_d^{max} \le T_\chi \tag{A5}$$

and when convergence is achieved

$$\Delta \max\left(\chi_d^{tot}, \chi_d^{sp}\right) \le T_\Delta \tag{A6}$$

where ΔX is the change of X between the two last iterations. In this study we use threshold values $T_\chi = 3$ and $T_\Delta = 0.5$.

[107] If $\chi_d^{max} > T_\chi$, we assume that some secondary observable cannot be fitted properly, suggesting that it should not have been chosen. The automated code steps back, increases R_{max} by 1 and restarts the automatic selection of the secondary observables at the current period. Increasing R_{max} reduces the number of selected lobes at a given period. The selection and inversion of secondary observables is refined through this process, until the criteria $\chi_d^{max} \le T_\chi$ is reached.

[108] Note that the criterion associated with T_Δ, ensures that the data fit would not significantly improve by adding further iterations. If the misfit criteria (equations (A5) and (A6)) are not reached after a maximum of 18 iterations (without counting erased iterations when the algorithm steps back) at a given period, the data is rejected.

[109] 3. From steps $N_{per} + 1$ to $2N_{per}$ the code uses 3 iterations per period to match the instantaneous phase of the modal cross-correlograms. The instantaneous phase is always taken at the time corresponding to the maximum of the "first" lobe, according to the ranking described in section A1. A quality control on the phase adjustment is done through the waveform fit criteria computed at the end of the last

iteration. A phase shift that remains between the actual and synthetic cross-correlograms produces a phase shift between the actual and synthetic waveforms. If a significant phase shift remains between the synthetic and observed waveforms, the seismogram does not pass the waveform fit criteria described below and is therefore discarded.

[110] 4. At the end of the last iteration we compute the two energy reduction criteria R_1 and R_2 over 3 group velocity windows. The first one, between 3.5 and 6 km s^{-1}, covers the fundamental mode and the higher-modes, the second one, between 3.5 and 4.2 km s^{-1}, covers mostly the fundamental mode, while the third one, between 4.2 and 6 km s^{-1} covers mostly the higher-modes. We also use these group velocity windows to evaluate the ratio $E_{h/f}$ of the higher-mode energy over the fundamental mode energy for the actual seismogram.

[111] A waveform inversion is considered to be successful if:

[112] 1. The final model predicts a synthetic seismogram in agreement with the recorded waveform. We use the following waveform fit criteria:

[113] (i) R_1 must be smaller than 0.3 or R_2 must be larger than 90% over the group velocity window 3.5–6 km s^{-1}.

[114] (ii) when higher-modes, over the group velocity window of 4.2–6 km s^{-1}, dominate the actual signal ($E_{h/f} > 4$), in addition to (i), we also require that restricted over this specific velocity window, R_1 must be smaller than 0.3 or R_2 must be larger than 90%

[115] (iii) when the fundamental mode clearly dominates the actual signal ($E_{h/f} < 0.25$) in addition to (i), R_1 must be smaller than 0.3 or R_2 must be larger than 90% over the group velocity windows 3.5–4.2 km s^{-1}.

[116] 2. The iterative process has converged toward a unique model. Our experience shows that this condition is verified when χ_m does not change by more than 0.05 between the two last iterations.
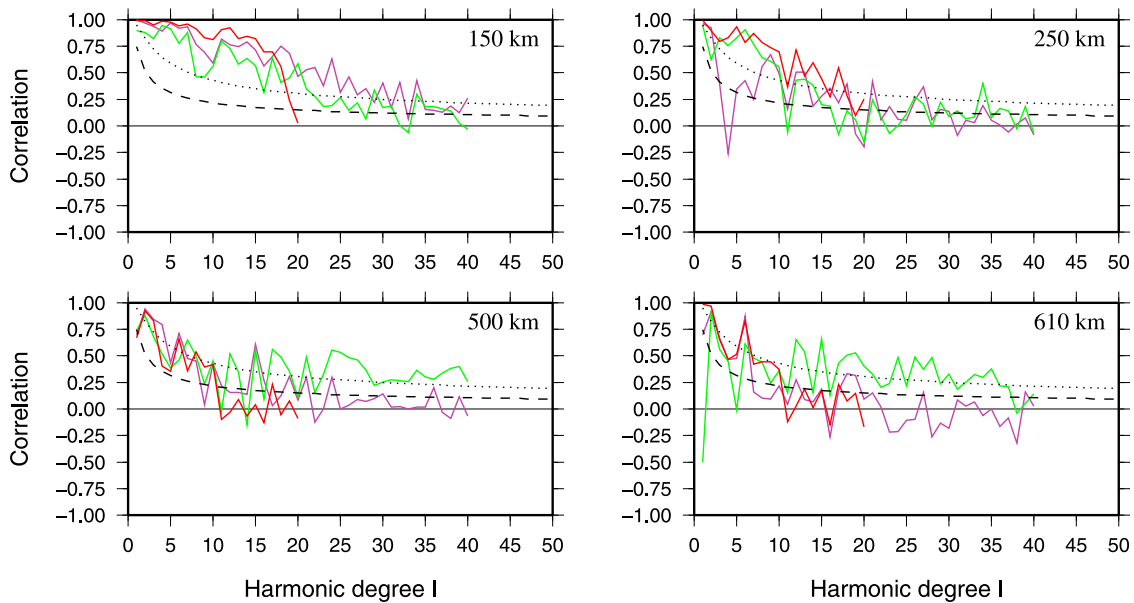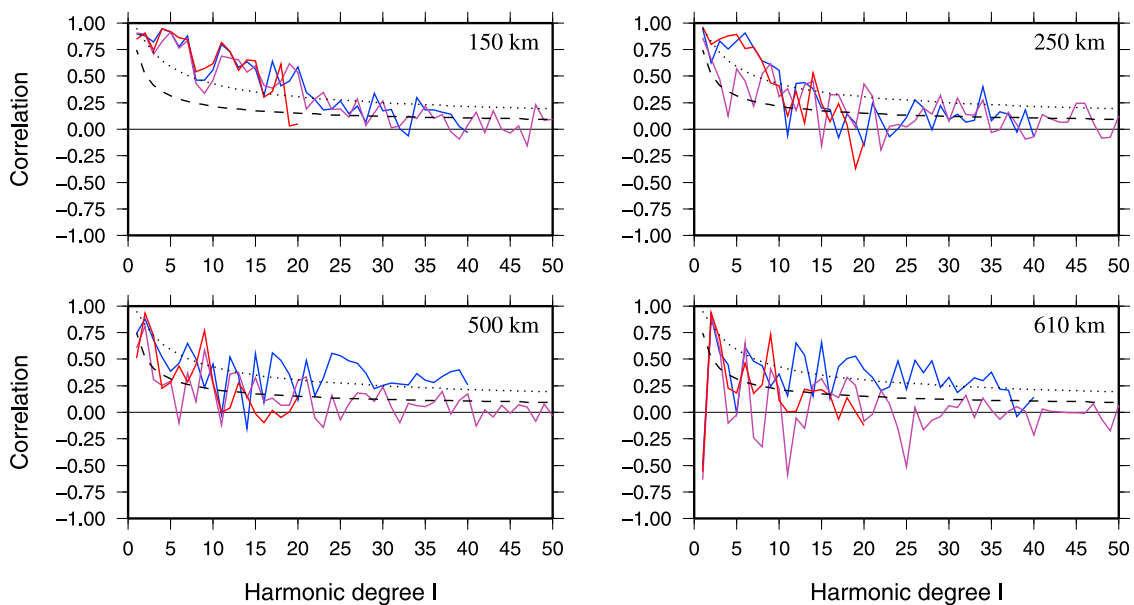
Figure B2. Same as Figure 12 but for correlations between S40RTS by *Ritsema et al.* [2011] and HMSL by *Houser et al.* [2008] (green line); S362ANI by *Kustowski et al.* [2008] (red line); DR2012 (purple line).

[117] The main steps of our automated scheme are summarized in Figure 3.

Appendix B: Correlation Using S362ANI, S40RTS and HMSL as Reference Models

[118] We depict in Figures B1–B3 the results obtained using successively S362ANI, S40RTS and HMSL as a reference model for the computation of the correlations C(l).

In the first 200 km DR2012 and S40RTS are the two models that correlate best at all degree, but up to l = 20 the four models perform equally. In the transition zone DR2012, S362ANI and S40RTS are in agreement up to degree 10 only and HMSL might be less resolved. Notice that the fact that S362ANI, S40RTS and HMSL include normal modes and/or body waves while our model does not, does not seem to increase their mutual correlation at depth. Clearly DR2012 is different from the three other models around 250 km depth.

Figure B3. Same as Figure 12 but for correlations between HMSL by *Houser et al.* [2008] and S362ANI by *Kustowski et al.* [2008] (red line); S40RTS by *Ritsema et al.* [2011] (blue line); DR2012 (purple line).

References

Barruol, G., D. Bosch, V. Clouard, E. Debayle, M. Doin, M. Godard, F. Fontaine, F. Masson, A. Tommasi, and C. Thoraval (2002), Polynesian lithosphere and upper mantle experiment (plume): The South Pacific superswell under investigation, *Eos Trans. AGU*, *83*(45), 510–514.

Bassin, C., G. Laske, and G. Master (2000), The current limits of resolution for surface wave tomography in north america, *Eos Trans. AGU*, *81*(48), Fall Meet. Suppl., Abstract S12A-03.

Beucler, E., and J. Montagner (2006), Computation of large anisotropic seismic heterogeneities (CLASH), *Geophys. J. Int.*, *165*(2), 447–468, doi:10.1111/j.1365-246X.2005.02813.x.

Beucler, E., E. Stutzmann, and J. Montagner (2003), Surface wave higher-mode phase velocity measurements using a roller-coaster-type algorithm, *Geophys. J. Int.*, *155*(1), 289–307.

Cara, M. (1979), Lateral variations of S-velocity in the upper mantle from higher Rayleigh modes, *Geophys. J. R. Astron. Soc.*, *57*(3), 649–670.

Cara, M., and J. Lévêque (1987), Waveform inversion using secondary observables, *Geophys. Res. Lett.*, *14*, 1046–1049.

Debayle, E. (1999), SV-wave azimuthal anisotropy in the Australian upper mantle: Preliminary results from automated Rayleigh waveform inversion, *Geophys. J. Int.*, *137*(3), 747–754.

Debayle, E., and B. Kennett (2000), Anisotropy in the Australasian upper mantle from Love and Rayleigh waveform inversion, *Earth Planet. Sci. Lett.*, *184*(1), 339–351.

Debayle, E., and M. Sambridge (2004), Inversion of massive surface wave data sets: Model construction and resolution assessment, *J. Geophys. Res.*, *109*, B02316, doi:10.1029/2003JB002652.

Debayle, E., J. Lévêque, and M. Cara (2001), Seismic evidence for a deeply rooted low-velocity anomaly in the upper mantle beneath the northeastern Afro/Arabian continent, *Earth Planet. Sci. Lett.*, *193*(3–4), 423–436.

Debayle, E., B. Kennett, and K. Priestley (2005), Global azimuthal seismic anisotropy and the unique plate-motion deformation of Australia, *Nature*, *433*(7025), 509–512, doi:10.1038/nature03247.

Doin, M., and L. Fleitout (1996), Thermal evolution of the oceanic lithosphere: An alternative view, *Earth Planet. Sci. Lett.*, *142*(1–2), 121–136, doi:10.1016/0012-821X(96)00082-9.

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

Dziewonski, A., T. Chou, and J. Woodhouse (1981), Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. Geophys. Res.*, *86*(B4), 2825–2852, doi:10.1029/JB086iB04p02825.

Ekstrom, G. (2011), A global model of Love and Rayleigh surface wave dispersion and anisotropy, 25–250 s, *Geophys. J. Int.*, *187*(3), 1668–1686, doi:10.1111/j.1365-246X.2011.05225.x.

Ekstrom, G., J. Tromp, and E. Larson (1997), Measurements and global models of surface wave propagation, *J. Geophys. Res.*, *102*(B4), 8137–8157.

Ekstrom, G., M. Nettles, and A. M. Dziewonski (2012), The global CMT project 2004–2010: Centroid-moment tensors for 13,017 earthquakes, *Phys. Earth Planet. Inter.*, *200*, 1–9, doi:10.1016/j.pepi.2012.04.002.

Forsyth, D. (1977), Evolution of upper mantle beneath mid-ocean ridges, *Tectonophysics*, *38*(1–2), 89–118, doi:10.1016/0040-1951(77)90202-5.

Goutorbe, B. (2010), Combining seismically derived temperature with heat flow and bathymetry to constrain the thermal structure of oceanic lithosphere, *Earth Planet. Sci. Lett.*, *295*(3–4), 390–400, doi:10.1016/j.epsl.2010.04.013.

Heintz, M., E. Debayle, and A. Vauchez (2005), Upper mantle structure of the South American continent and neighboring oceans from surface wave tomography, *Tectonophysics*, *406*(1–2), 115–139, doi:10.1016/j.tecto.2005.05.006.

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*(1), 195–212, doi:10.1111/j.1365-246X.2008.03763.x.

Jaupart, C., and J. Mareschal (1999), The thermal structure and thickness of continental roots, *Lithos*, *48*(1–4), 93–114, doi:10.1016/S0024-4937(99)00023-7.

Kanamori, H., and D. Anderson (1977), Importance of physical dispersion in surface-wave and free oscillation problems: Review, *Rev. Geophys.*, *15*(1), 105–112, doi:10.1029/RG015i001p00105.

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

Lebedev, S., G. Nolet, T. Meier, R. van der Hilst, and G. Masters (2005), Automated multimode inversion of surface and S waveforms, *Geophys. J. Int.*, *162*(3), 951–964, doi:10.1111/j.1365-246X.2005.02708.x.

Lestunff, Y., and Y. Ricard (1995), Topography and geoid due to lithospheric mass anomalies, *Geophys. J. Int.*, *122*(3), 982–990, doi:10.1111/j.1365-246X.1995.tb06850.x.

Lévêque, J., M. Cara, and D. Rouland (1991), Waveform inversion of surface wave data: Test of a new tool for systematic investigation of upper mantle structures, *Geophys. J. Int.*, *104*, 561–581.

Lévêque, J., E. Debayle, and V. Maupin (1998), Anisotropy in the Indian Ocean upper mantle from Rayleigh and Love waveform inversion, *Geophys. J. Int.*, *133*(3), 529–540.

Maggi, A., E. Debayle, K. Priestley, and G. Barruol (2006a), Multimode surface waveform tomography of the Pacific Ocean: A closer look at the lithospheric cooling signature, *Geophys. J. Int.*, *166*(3), 1384–1397, doi:10.1111/j.1365-246X.2006.03037.x.

Maggi, A., E. Debayle, K. Priestley, and G. Barruol (2006b), Azimuthal anisotropy of the Pacific region, *Earth Planet. Sci. Lett.*, *250*(1–2), 53–71, doi:10.1016/j.epsl.2006.07.010.

Marquering, H., R. Snieder, and G. Nolet (1996), Waveform inversion and the significance of surface-wave mode coupling, *Geophys. J. Int.*, *124*, 258–278.

Montagner, J. (1986), Regional three-dimensional structures using long-period surface waves, *Ann. Geophys.*, *4*, 283–294.

Müller, R., M. Sdrolias, C. Gaina, and W. Roest (2008), Age, spreading rates and spreading symmetry of the world's ocean crust, *Geochem. Geophys. Geosyst.*, *9*, Q04006, doi:10.1029/2007GC001743.

Nataf, H., and Y. Ricard (1995), 3SMAC: An *a priori* tomographic model of the upper mantle based on geophysical modeling, *Phys. Earth Planet. Inter.*, *95*, 101–122.

Nettles, M., and A. M. Dziewonski (2008), Radially anisotropic shear velocity structure of the upper mantle globally and beneath North America, *J. Geophys. Res.*, *113*, B02303, doi:10.1029/2006JB004819.

Nishimura, C., and D. Forsyth (1989), The anisotropic structure of the upper mantle in the Pacific, *Geophys. J.*, *96*(2), 203–229, doi:10.1111/j.1365-246X.1989.tb04446.x.

Nolet, G. (1975), Higher Rayleigh modes in western-Europe, *Geophys. Res. Lett.*, *2*(2), 60–62.

Nolet, G. (1990), Partitioned waveform inversion and two dimensional structure under the network of autonomously recording seismographs, *J. Geophys. Res.*, *95*, 8499–8512.

Panning, M., and B. Romanowicz (2006), A three-dimensional radially anisotropic model of shear velocity in the whole mantle, *Geophys. J. Int.*, *167*(1), 361–379, doi:10.1111/j.1365-246X.2006.03100.x.

Pilidou, S., K. Priestley, O. Gudmundsson, and E. Debayle (2004), Upper mantle S-wave speed heterogeneity and anisotropy beneath the North Atlantic from regional surface wave tomography: The Iceland and Azores plumes, *Geophys. J. Int.*, *159*(3), 1057–1076, doi:10.1111/j.1365-246X.2004.02462.x.

Pilidou, S., K. Priestley, E. Debayle, and O. Gudmundsson (2005), Rayleigh wave tomography in the North Atlantic: High resolution images of the Iceland, Azores and Eifel mantle plumes, *Lithos*, *79*(3–4), 453–474, doi:10.1016/j.lithos.2004.09.012.

Priestley, K., E. Debayle, D. McKenzie, and S. Pilidou (2006), Upper mantle structure of eastern Asia from multimode surface waveform tomography, *J. Geophys. Res.*, *111*, B10304, doi:10.1029/2005JB004082.

Priestley, K., D. McKenzie, E. Debayle, and S. Pilidou (2008), The African upper mantle and its relationship to tectonics and surface geology, *Geophys. J. Int.*, *175*(3), 1108–1126, doi:10.1111/j.1365-246X.2008.03951.x.

Ritsema, J., A. Deuss, H. J. van Heijst, and J. H. 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*(3), 1223–1236, doi:10.1111/j.1365-246X.2010.04884.x.

Ritzwoller, M. H., N. M. Shapiro, M. P. Barmin, and A. L. Levshin (2002), Global surface wave diffraction tomography, *J. Geophys. Res.*, *107*(B12), 2335, doi:10.1029/2002JB001777.

Salaün, G., H. A. Pedersen, A. Paul, V. Farra, H. Karabulut, D. Hatzfeld, C. Papazachos, D. M. Childs, C. Pequegnat, and SIMBAAD Team (2012), High-resolution surface wave tomography beneath the Aegean-Anatolia

region: Constraints on upper-mantle structure, *Geophys. J. Int.*, *190*(1), 406–420, doi:10.1111/j.1365-246X.2012.05483.x.

Sambridge, M. (1999), Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space, *Geophys. J. Int.*, *138*(2), 479–494.

Sieminski, A., E. Debayle, and J. Lévêque (2003), Seismic evidence for deep low-velocity anomalies in the transition zone beneath West Antarctica, *Earth Planet. Sci. Lett.*, *216*(4), 645–661, doi:10.1016/S0012-821X(03)00518-1.

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

Spetzler, J., and R. Snieder (2001), The effect of small-scale heterogeneity on the arrival time of waves, *Geophys. J. Int.*, *145*, 786–796.

Spetzler, J., J. Trampert, and R. Snieder (2002), The effect of scattering in surface wave tomography, *Geophys. J. Int.*, *149*(3), 755–767, doi:10.1046/j.1365-246X.2002.01683.x.

Stutzmann, E., and J. Montagner (1993), An inverse technique for retrieving higher mode phase-velocity and mantle structure, *Geophys. J. Int.*, *113*(3), 669–683.

Takeuchi, H., and M. Saito (1972), Seismic surface waves, in *Methods in Computational Physics*, vol. 11, edited by B. A. Bolt, pp. 217–295, Academic, New York.

Tanaka, S., D. Suetsugu, H. Shiobara, H. Sugioka, T. Kanazawa, Y. Fukao, G. Barruol, and D. Reymond (2009), On the vertical extent of the large low shear velocity province beneath the South Pacific Superswell, *Geophys. Res. Lett.*, *36*, L07305, doi:10.1029/2009GL037568.

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

Trampert, J., and J. Spetzler (2006), Surface wave tomography: Finite-frequency effects lost in the null space, *Geophys. J. Int.*, *164*(2), 394–400, doi:10.1111/j.1365-246X.2006.02864.x.

Trampert, J., and J. Woodhouse (1995), Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds, *Geophys. J. Int.*, *122*, 675–690.

Turcotte, D., and G. Schubert (2002), *Geodynamics*, 2nd ed., Cambridge Univ. Press, Cambridge, U. K.

van Heijst, H., and J. Woodhouse (1997), Measuring surface-wave overtone phase velocities using a mode branch stripping technique, *Geophys. J. Int.*, *131*, 209–230.

van Heijst, H., and J. Woodhouse (1999), Global high-resolution phase velocity distributions of overtone and fundamental-mode surface waves determined by mode branch stripping, *Geophys. J. Int.*, *137*, 601–620.

Visser, K., S. Lebedev, J. Trampert, and B. L. N. Kennett (2007), Global Love wave overtone measurements, *Geophys. Res. Lett.*, *34*, L03302, doi:10.1029/2006GL028671.

Visser, K., J. Trampert, and B. L. N. Kennett (2008), Global anisotropic phase velocity maps for higher mode Love and Rayleigh waves, *Geophys. J. Int.*, *172*(3), 1016–1032, doi:10.1111/j.1365-246X.2007.03685.x.

Wolfe, C. J., S. C. Solomon, G. Laske, J. A. Collins, R. S. Detrick, J. A. Orcutt, D. Bercovici, and E. H. Hauri (2011), Mantle P-wave velocity structure beneath the Hawaiian hotspot, *Earth Planet. Sci. Lett.*, *303*(3–4), 267–280, doi:10.1016/j.epsl.2011.01.004.

Yoshizawa, K., and G. Ekstrom (2010), Automated multimode phase speed measurements for high-resolution regional-scale tomography: Application to North America, *Geophys. J. Int.*, *183*(3), 1538–1558, doi:10.1111/j.1365-246X.2010.04814.x.

Yoshizawa, K., and B. Kennett (2002), Determination of the influence zone for surface wave paths, *Geophys. J. Int.*, *149*, 441–454.

Zaroli, C., E. Debayle, and M. Sambridge (2010), Frequency-dependent effects on global S-wave traveltimes: Wavefront-healing, scattering and attenuation, *Geophys. J. Int.*, *182*(2), 1025–1042, doi:10.1111/j.1365-246X.2010.04667.x.

Zhang, Y., and T. Lay (1999), Evolution of oceanic upper mantle structure, *Phys. Earth Planet. Inter.*, *114*(1–2), 71–80, doi:10.1016/S0031-9201(99)00047-3.

Zhang, Y., and T. Tanimoto (1991), Global Love wave phase-velocity variation and its significance to plate-tectonics, *Phys. Earth Planet. Inter.*, *66*(3–4), 160–202.