

Multiple-frequency *SH*-wave tomography of the western US upper mantle

Yue Tian, Karin Sigloch* and Guust Nolet†

Department of Geosciences, Princeton University, Princeton, NJ 08544, USA. E-mail: ytian@princeton.edu

Accepted 2009 April 22. Received 2009 April 21; in original form 2009 January 23

SUMMARY

We estimate the *SH*-wave velocity and attenuation structures of the western US upper mantle using the dense network of the USArray and new techniques: we observe a multiple-frequency data set of both traveltimes and amplitude anomalies, and interpret these with full 3-D finite-frequency sensitivity kernels. Amplitudes show stronger frequency dependence than traveltimes. We perform a joint inversion on the measured traveltimes and amplitude anomalies, interpreting them in terms of velocity and attenuation heterogeneities.

Aside from the expected clear division between the slow, tectonically active region in the west and the fast craton in the east, several interesting smaller velocity anomalies are observed. The subduction along the Cascades at 100–300 km depths shows lateral discontinuity, with a ‘slab hole’ (absence of fast anomalies) observed around 45°N. The delaminated Sierra Nevada Mountains root is observed to have sunk to 200 km depth. The Yellowstone plume seems to have an origin (weak slow velocity anomalies) near 1000 km depth, but the plume conduit seems to be interrupted by a fast anomaly, which is identified as a fragment of the Farallon slab. The *S*-velocity model shows a trench-perpendicular ‘slab gap’ (absence of fast anomalies) at almost the same location as in the *P* model recently published by Sigloch *et al.* (2008).

The methodological improvements described in the first paragraph have several benefits. Amplitude data help to sharpen the edges of narrow velocity heterogeneities in the shallow upper mantle. The focusing effect from velocity heterogeneity dominates over that of attenuation and must be considered when interpreting amplitude anomalies. In general, velocity and attenuation heterogeneities correlate positively, suggesting that temperature plays a major role in forming the anomalies.

Key words: Body waves; Seismic attenuation; Seismic tomography.

1 INTRODUCTION

A single large oceanic plate, the Farallon plate, has been subducting beneath the west coast of North America for over 150 Myr. Recent tectonic activity in the western US has been largely governed by this subduction history (e.g. Humphreys *et al.* 2003). The dip angle of the subducting plate decreased dramatically in the late Cretaceous, cooling the lithosphere above it. Volcanism in the Sierras halted and migrated eastward, followed about 80 Myr ago by the Laramide orogeny. At the end of this orogeny, about 50 Myr ago, volcanism flared up and started to migrate back towards the west. This is generally thought to reflect some form of detachment of the quasi-horizontal Farallon slab but details remain obscure. The geographical location of the magmatism shows two fronts that

converged from north and south onto an axis reaching from southern Nevada to northern Arizona. On the basis of this observation, Humphreys (1995) proposed a buckling scenario in which slab segments north and south of this axis started to sink. van der Lee & Nolet (1997) failed to observe the expected east–west trend in tomographic images of *S* velocity and favoured a more conventional roll-back towards the west, with the slab breaking up into fragments, allowing the asthenosphere to flow around the edges resulting in magmatism that converged with the observed east–west oriented symmetry. Resolution in that tomographic experiment was not sufficient to settle the issue, however.

More recently, with the advent of USArray and the use of new techniques (multiple-frequency inversion of both *P* delays and amplitudes), Sigloch *et al.* (2008) confirmed that the slab was heavily fragmented and brought unprecedented precision to the images, allowing the unambiguous identification of several tears in slab fragments that remain visible in the upper mantle. The use of multiple frequencies exploits the frequency-dependent sensitivity of traveltimes and amplitude data. If a delay is identical for high and low frequencies, the anomaly must have a size at least as large

*Now at: Department of Earth and Environmental Sciences, Ludwig-Maximilians-Universität, 80333 Munich, Germany.

†Now at: Géosciences Azur, Université de Nice/Antipolis, 06560 Sophia Antipolis, France.

as the widest Fresnel zone. If the anomaly is visible at high frequencies but disappears for low frequencies, it is only as large as the widest Fresnel zone for which the delay remains unaffected. Finite-frequency analysis also allows for the inversion of amplitude anomalies, which are not correctly modelled with ray theory even for higher frequencies because of the strong non-linearities introduced by the high-frequency approximation.

In this paper, we also use finite-frequency theory for delay times and for the effects of focusing on amplitudes of *SH* waves. For the effects of attenuation, a finite-frequency theory was derived by Nolet (2008, chapter 8), and a numerical implementation was given by Tian *et al.* (2007b). The effect of attenuation can be modelled by a small imaginary component to the seismic velocity, which gives sensitivity kernels that are similar in shape to those for traveltimes—including a zero sensitivity on the geometrical ray path. A first joint inversion of focusing and attenuation using finite-frequency theory was given by Sigloch *et al.* (2008) for *P* waves beneath North America. We extend the inversion formalism of Sigloch, and also allow for changes in attenuation to influence the body wave traveltime dispersion, although we found that attenuation has little effect on traveltimes, even for *S* waves.

We extract frequency-dependent delays and amplitudes from *SH* waveforms with high signal-to-noise ratio, using overlapping frequency bands and the matched filtering method developed by Sigloch & Nolet (2006). The dense coverage of the USArray, together with the multiple-frequency data, provides unprecedentedly high resolution beneath the western US. A similar multiple-frequency strategy for delay times was used by Hung *et al.* (2004) for imaging the Iceland plume, and by Yang *et al.* (2006) beneath Azores, though with much smaller data sets and fewer frequency bands than used in this study.

Models of intrinsic attenuation provide an important tool for studying the physical state of the Earth's interior, in particular because attenuation is a strong function of thermal structure (e.g. Artemieva *et al.* 2004; Yang & Forsyth 2008). Since variations in velocity and attenuation both affect wave amplitudes, a joint inversion for velocity and attenuation is called for. So far, few attenuation tomography studies are based on such a joint inversion. Attempts to eliminate the effects of focusing in surface waves were first made by Romanowicz (1990) and Durek *et al.* (1993), and were recently followed by successful attempts to invert for velocity and attenuation jointly (Billien *et al.* 2000; Dalton & Ekström

2006). However, previous attenuation tomography studies all used ray theory to model the effects of focusing and were almost fully confined to surface waves, for which the effects of focusing are not dominant for large-scale heterogeneities (Selby & Woodhouse 2002; Gung & Romanowicz 2004). For body waves, the effects of focusing are very non-linear when treated with ray theory, which gives singularities as amplitudes go to infinity over very small areas near focal lines or points. In addition, ray theory does not model the frequency dependence of the effects of focusing, which is observed in practice (Allen *et al.* 1999; Tibuleac *et al.* 2003; Sigloch & Nolet 2006). To first order, finite-frequency theory handles healing of amplitudes and therefore allows for a more stable modelling of focusing with frequency dependence and without excessive effects of non-linearity (Nolet 2008, chapter 8; Tian *et al.* 2007b). For body waves, Sigloch (2008) was the first to jointly invert for velocity and attenuation. She used finite-frequency theory and showed that the effects of focusing strongly dominate over intrinsic attenuation when inverting *P*-wave amplitudes. Due to the stronger attenuation of *S* waves, this conclusion need not necessarily be valid for the present study and we shall present an independent analysis of the issue for *S* waves.

2 THE DATA SET: MULTIPLE-FREQUENCY *SH*-WAVE TRAVELTIMES AND AMPLITUDES

2.1 Data set overview

We construct a new global data set of *SH*-wave traveltime and amplitude anomalies in multiple frequency bands by cross-correlating observed waveforms with synthetic waveforms. The synthetic waveforms are computed with the WKBJ method (Chapman 1978) for the IASP91 velocity model (Kennett & Engdahl 1991) extended with the *Q* model from PREM (Dziewonski & Anderson 1981), with attenuation and *Q*-related dispersion taken into account. In this study, we select from the global data set 177 events from 1995 to 2007 September (Fig. 1a) recorded by 968 stations in North America (Fig. 1b), most of which are USArray stations, within epicentral distances from 30° to 90°. Measurements were deemed acceptable by visual inspection of the source time function and the waveform fits after cross-correlation. We imposed a lower limit of 0.9 on the cross-correlation coefficient. Events with fewer than 10

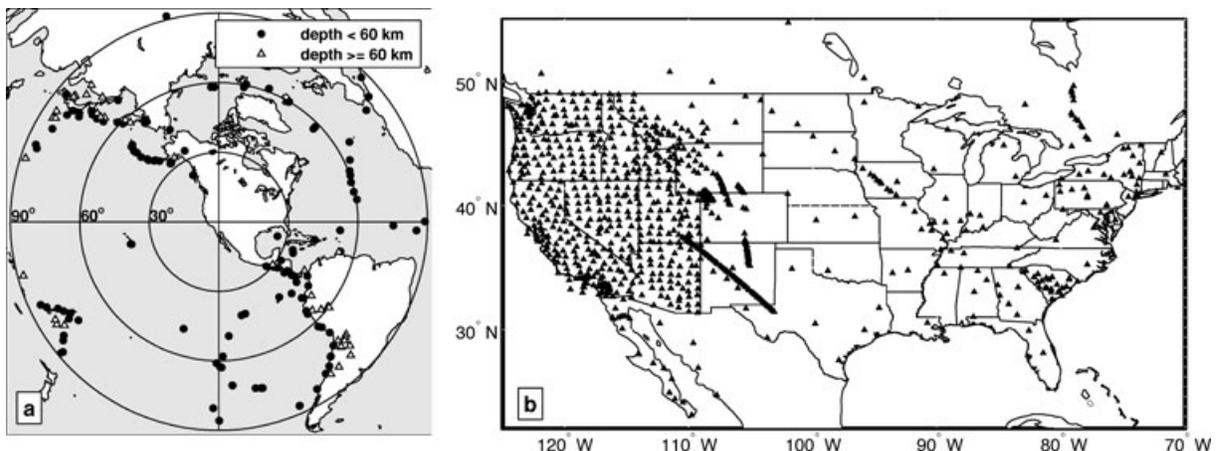


Figure 1. (a) Azimuthal map of the 177 events used in this study. The centre is at (36°N, 115°W). The three circles denote distances of 30°, 60° and 90°, respectively. Earthquake depths range from 2 to 678 km, with 134 events above 60 km and 43 events below 60 km. (b) Distribution of the 968 stations in North America used in this study.

acceptable traveltimes over a bandwidth spanning the full frequency range of the observed seismograms were rejected. The accepted events have moment magnitudes from 5.8 to 7.4. The preferred magnitude is in the range 6.2–6.5, small enough to keep the source time function simple while the signal-to-noise ratio is already high. The event depths range from 2 to 678 km. Counting all frequency bands, the data set contains 61 972 traveltimes anomalies and 45 914 amplitude anomalies.

2.2 Frequency dependence of the data

Traveltimes and amplitudes are measured both for broad-band signals and in five passbands with centre periods ranging from 2.5 to 40 s. We choose Gabor filters (Fig. 2), which are Gaussian in the log-frequency domain and have constant fractional bandwidth. The Gabor filters are preferred because they are good at separating the low-frequency bands which are very narrow and close to each other in the linear-frequency domain, and the finite-frequency effect is strongest for low-frequency data (Hung *et al.* 2000). Figs 3–4

show the histograms of traveltimes and amplitude anomalies in each band. The largest fraction of acceptable measurements is found for dominant periods between 10 and 20 s.

For each event, we obtain an initial estimate of the origin-time error by calculating the weighted average of broad-band traveltimes anomalies from the event’s global recordings. The weight for station *i* is inversely proportional to the number of acceptable broad-band traveltimes measurements in the unit area around station *i*. This weighted average intends to remove the effect of the uneven spatial distribution of stations, and is more likely to reflect the origin-time error. In Fig. 3, this weighted average is removed for each event. Therefore, the overall trend of late arrivals (positive traveltimes anomalies) cannot be completely explained by origin-time errors, and the delay is tentatively attributed to the slow mantle beneath the western US, where most stations are located. The median delay ranges from 1.4 s at long periods to 1.9 s at short periods (Fig. 3). Since the dispersion caused by attenuation is already taken into account in the synthetic waveform used for cross-correlation, such dispersion may be caused by residual attenuation not present in the

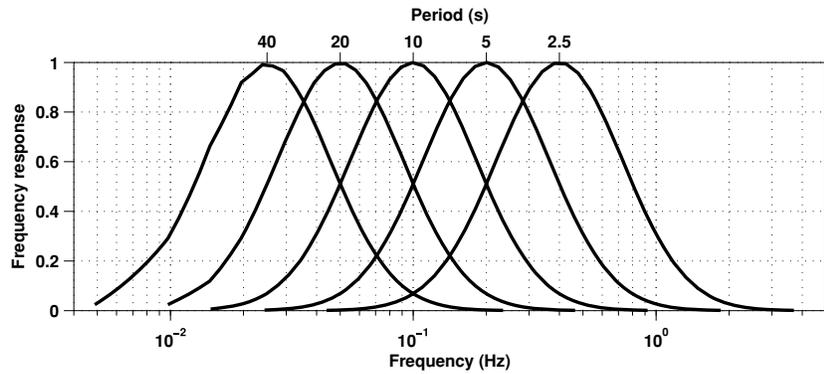


Figure 2. Frequency response of the Gabor filters used to filter *SH* waves. The centre periods of the five bands are 40, 20, 10, 5 and 2.5 s.

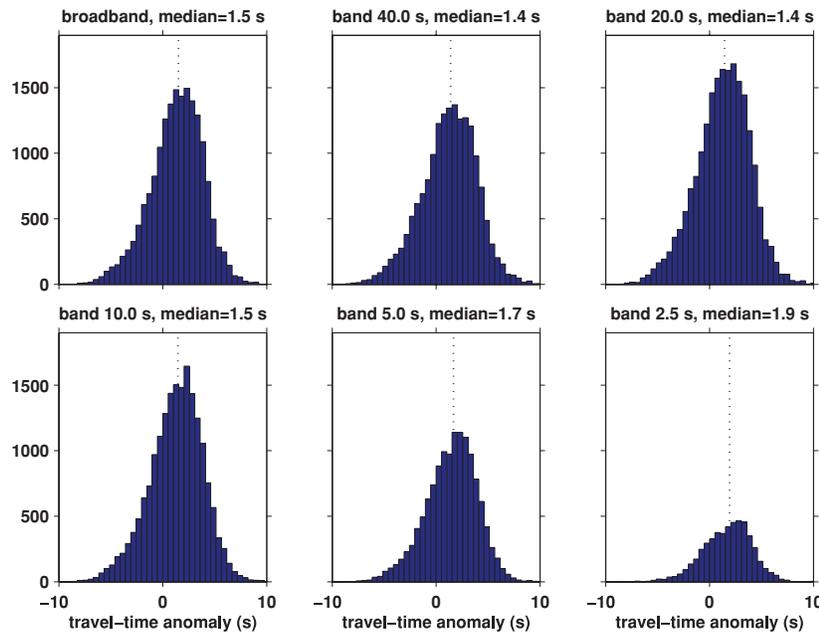


Figure 3. Histograms of traveltimes anomalies in broadband and the five passbands. The title indicates the centre period of the band and the median of the data in that band. The median value is also shown by the dotted line. For each event, we remove the estimated origin-time error, which is the weighted average of broad-band traveltimes anomalies from the global recordings. The weight for station *i* is inversely proportional to the number of acceptable broad-band traveltimes measurements in the unit area around station *i*.

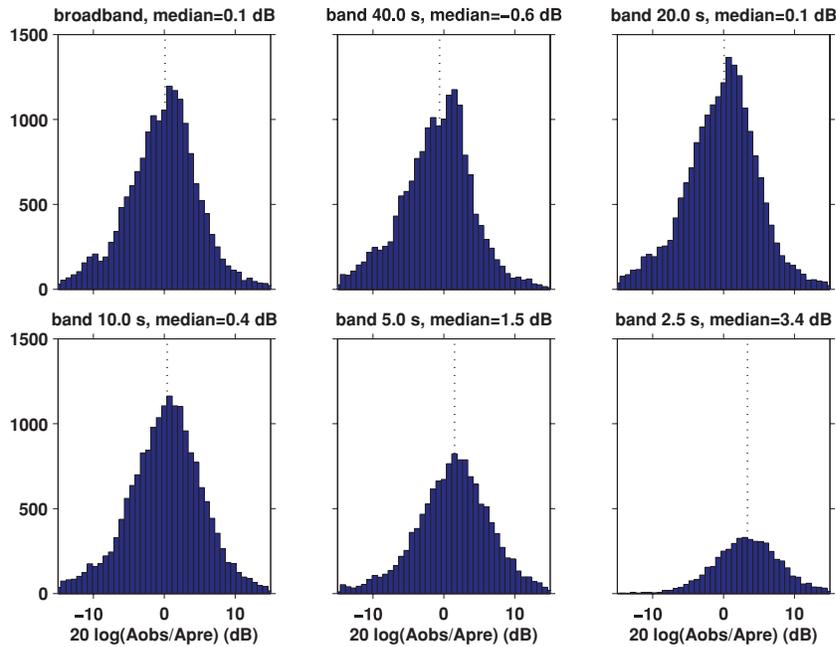


Figure 4. Histograms of amplitude anomalies (ratio between observed and predicted amplitudes in dB) in broadband and the five passbands. Titles indicate the centre period and the median of the data in that band. The median value is also shown by the dotted line. No ‘de-mean’ operation is done on amplitudes.

PREM Q -model, by wave front healing effects being dominated by small slow anomalies, or a combination of both.

Amplitudes show stronger frequency dependence than travel-times, especially in the two bands of shortest period (Fig. 4). The Fréchet kernels for amplitudes do not show the zero sensitivity hole on the ray (Dahlen & Baig 2002) that characterizes the sensitivity of body wave delay times, so we would expect a somewhat reduced ‘finite-frequency’ effect in case an anomaly is smaller than the Fresnel zone. We suspect therefore that this dispersion is at least partially due to surface sediment effects. Zhou *et al.* (2003) observed that amplification effects due to thick sediment are frequency-dependent for periods less than about 8 s, and that these cannot be modelled by a frequency-independent station amplitude correction term. With increasing frequency, amplitude anomalies increase in magnitude and become more scattered. We decide therefore not to use amplitude measurements in the bands with centre periods of 2.5 and 5 s. We do not ‘de-mean’ amplitudes, since we are not certain our global data set should have zero mean. However, as explained in Section 3 we use amplitude correction terms in the tomographic inversion to compensate for possible station amplitude bias.

2.3 Spatial distribution of the data

Since the Fréchet kernels narrow down near source and receiver, sensitivity is strongly enhanced for small anomalies near the surface, and measured anomalies often reflect shallow structures. We plot the average traveltime and amplitude anomalies at each station (in broadband), which reflect anomalies in the shallow subsurface beneath the stations. The traveltime anomalies (Fig. 5) show a clear division between early arrivals (fast velocity) in the eastern Craton region and late arrivals (slow velocity) in the western US. Smaller-scale structures are also expressed by these delay-time averages: for example, early arrivals for the northern Cascade Mountains and northern Rocky Mountains, late arrivals in the Rio Grande Rift along the La Ristra Array, and early arrivals in the southern Sierra Nevada.

However, the amplitudes show a more complex pattern (Fig. 6), less correlated with tectonic structures. This lack of correlation reflects a more complicated relationship between the Earth and the observed amplitudes, which are simultaneously affected by changes in velocity gradients, by attenuation, and by surface sediment effects. For example, if amplitude anomalies are attributed to attenuation only, late arrivals (red in Fig. 5) should correspond to small amplitudes (red in Fig. 6). Instead, we observe co-existence of late arrivals and large amplitudes for many stations, for example, along the west coast (especially in the northern tip and southern California), in the Snake River Plain, along the Montana BB Array and the Laramie Telemetered Array in Wyoming. Similar patterns were observed for P waves by Butler (1983). This shows that attenuation cannot always be the dominant effect on amplitudes, and that other factors such as focusing/defocusing and surface sediment amplifications have to be considered.

3 THE JOINT LINEAR SYSTEM UNDER FINITE-FREQUENCY SENSITIVITY THEORY

In finite-frequency tomography, the general form of the linear inverse problem is

$$d_i = \int_V K_i(\mathbf{r})m(\mathbf{r})d^3\mathbf{r}. \quad (1)$$

The i th datum d_i is the deviation of an observed quantity from its prediction for the background model—either a delay time $\delta T_i = T_i^{\text{obs}} - T_i^{\text{pred}}$ or an amplitude anomaly $\delta \ln A_i = (A_i^{\text{obs}} - A_i^{\text{pred}})/A_i^{\text{pred}}$, both in a particular frequency band. The model parameter m represents velocity heterogeneities $\delta \ln V_S$ or attenuation heterogeneities $\delta \ln Q_S^{-1}$. The finite-frequency sensitivity kernel for traveltimes is given by Dahlen *et al.* (2000), for focusing by Dahlen & Baig (2002), and for attenuation by Tian *et al.* (2007b). See also Nolet (2008), chapters 7–8.

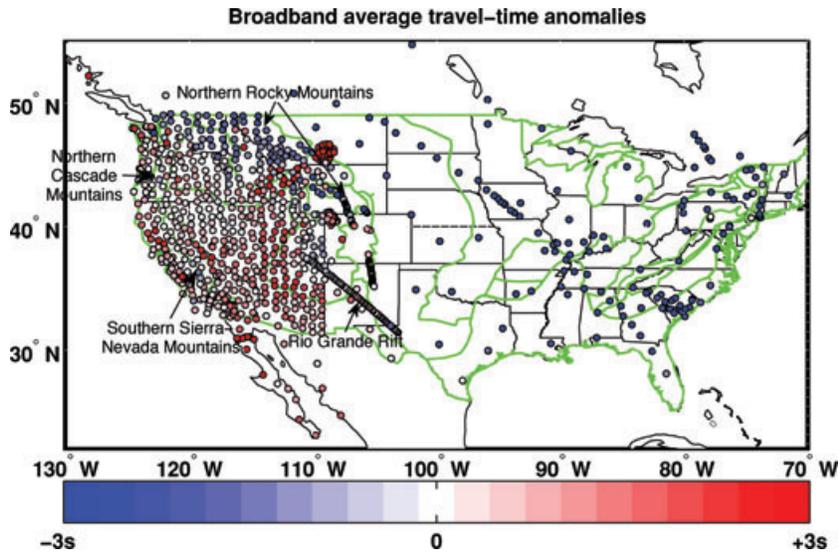


Figure 5. Broad-band traveltimes anomalies averaged at each station. Positive values reflect slow velocity anomalies in the vicinity of the station, and negative values indicate fast velocity anomalies. Examples of local structures expressed by traveltimes anomalies are labelled. Green lines delineate the geological boundaries in the US.

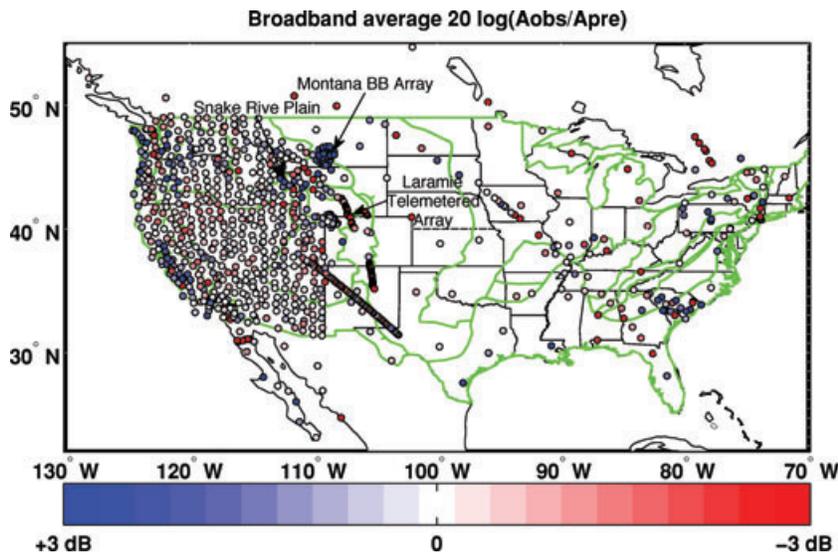


Figure 6. Broad-band average amplitude anomalies (ratio between observed and predicted amplitudes in dB) at each station. Green lines delineate the geological boundaries in the US. Notice that blue represents large amplitudes and red represents small amplitudes. Labeled are examples of regions with co-existence of late arrivals and large amplitudes.

We use the same local model parametrization in the form of a tetrahedral mesh with linear interpolation in between grid nodes as Sigloch *et al.* (2008). The grid spacing of the mesh is about 66 km beneath the United States, and increases to roughly 200 km at 660 km depth. The mesh is produced with the Matlab software developed by Persson & Strang (2004). With source and receiver correction terms and regularization, the complete discrete linear system of the tomographic problem can be written as

$$\begin{pmatrix} K_V^T & K_Q^T & K_C^T & \mathbf{0} \\ K_V^A & K_Q^A & \mathbf{0} & K_C^A \\ \epsilon_1 I & \epsilon_1 I & \epsilon_1 I & \epsilon_1 I \\ \epsilon_2 R & \epsilon_2 R & \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \delta \ln V_S \\ \delta \ln Q_S^{-1} \\ C^T \\ C^A \end{pmatrix} = \begin{pmatrix} \delta T \\ \delta \ln A \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix}. \quad (2)$$

The vector on the right-hand side contains data, with traveltimes anomaly δT already corrected for ellipticity, topography, and crustal

structure using model CRUST2.0 (Bassin *et al.* 2000). On the left-hand side the model vector includes C^T with hypocentre and origin-time corrections, and C^A which represents receiver amplitude corrections accounting for sediment effects and problems with the instrument response, source amplitude corrections accounting for the scalar moment errors and the uncertainty of source time function scaling in the measurement process (Sigloch & Nolet 2006). A similar approach of corrections for surface wave studies was pioneered by Dalton & Ekström (2006).

The submatrices $K_V^T, K_Q^T, K_V^A, K_Q^A$ contain finite-frequency sensitivity kernels, computed using the software by Tian *et al.* (2007a,b). K_Q^T is proportional to K_V^T through the dispersion relation

$$K_Q^T(\mathbf{r}) = \frac{\ln \bar{\omega} - \ln \omega_0}{\pi Q_S} K_V^T(\mathbf{r}),$$

where Q_S is for the background model, the reference frequency ω_0 is taken as 2π rad s^{-1} , and $\bar{\omega}$ is the average frequency of the band with frequency response $\dot{m}(\omega)$

$$\bar{\omega} = \frac{\int_0^\infty \omega |\dot{m}(\omega)|^2 d\omega}{\int_0^\infty |\dot{m}(\omega)|^2 d\omega}.$$

$\mathbf{K}_C^T, \mathbf{K}_C^A$ are *ad hoc* matrices for corrections $\mathbf{C}^T, \mathbf{C}^A$. \mathbf{I} is the identity matrix and ϵ_1 is the norm damping parameter. \mathbf{R} is the Laplacian roughening operator (see Nolet 2008, section 14.5) and ϵ_2 is the smoothing parameter.

We expect a high resolution beneath the dense USArray, however have meshed the entire globe in order to absorb delays or focusing effects acquired along wave paths outside of North America. No unique solution exists and the linear system (2) is necessarily regularized. We aim for a model, which interprets the data within their margin of uncertainties (satisfying the first two rows of eq. 2), while staying close to the reference model (the third row of eq. 2), and remaining as smooth as possible (the fourth row of eq. 2). This trade-off between fitting the data and regularizing the model is controlled by ϵ_1 and ϵ_2 . Eq. (2) shows that velocity and attenuation have coupled effects on both traveltimes and amplitudes, and our inversion interprets these data simultaneously.

We scale the data with their standard deviation and the model parameters with their ‘prior uncertainties’. The estimated data uncertainties are 0.6–1.0 s for δT and 0.1–0.25 for $\delta \ln A$. The prior parameter ‘uncertainty’ in the model is 0.02 for $\delta \ln V_S$ and 0.4 for $\delta \ln Q_S^{-1}$. The data uncertainty is estimated by comparing data from close earthquake pairs and varying cross-correlation window lengths. Model parameter uncertainties are first estimated from the prior knowledge of the magnitude of V_S and Q_S heterogeneities, origin–time error, etc. Both estimates carry a subjective uncertainty and have in effect been slightly adjusted within their margin of error as we gained experience about data compatibilities during early inversion runs. Changes in these uncertainties would force us to change the damping to keep χ^2/N fixed near a value of 1. Fig. 8 shows this dependence on the damping. The scaled linear system is solved by LSQR (Paige & Saunders 1982; Nolet 1987).

As a measure of how strongly a particular tetrahedral volume in the Earth is sensed by the combined set of kernels, we compute kernel ‘column densities’, defined as

$$D_j = \frac{\sum_i K_{ij}}{\max_j \sum_i K_{ij}}, \quad (3)$$

where K stands for the tomographic matrices ($\mathbf{K}_V^T, \mathbf{K}_Q^T, \mathbf{K}_V^A, \mathbf{K}_Q^A$ in eq. 2). Fig. 7 shows $\log(D_j)$ for velocity V_S and attenuation Q_S , respectively. The colour scaling for the two kernels differs, and is adapted to $\max_j \sum_i K_{ij}$ of 4381 s for the V_S kernel and 552 for the Q_S kernel. Thus, the sensitivity to attenuation is an order of magnitude weaker than that to velocity. The dense coverage of the USArray is reflected in the large values of D_j in the western US. D_j increases with depth because of the larger size of tetrahedra at depth. At depth, the wide kernels from different wave paths overlap and sum to a large density, even though the magnitude of individual kernels decreases.

4 INVERSION RESULTS

The models of velocity and attenuation perturbations presented in this paper are with respect to the starting 1-D model, that is, the IASP91 velocity model (Kennett & Engdahl 1991) extended with CRUST2.0 (Bassin *et al.* 2000) and the PREM Q model

(Dziewonski & Anderson 1981). Resolution tests (Appendix B1) show that velocity is significantly better resolved than attenuation, so we rely initially on the velocity structure to analyse the results and study the western US upper mantle. We also investigate quantitatively the relative importance of velocity and attenuation in interpreting the amplitude data.

4.1 Trade-off curves

As described in Section 3, there is trade-off between fitting the data and the degree of regularization of the model. We use χ^2/N as a measure for the misfit of the observed data to the values predicted by finite-frequency theory

$$\frac{\chi^2}{N} = \frac{1}{N} \sum_{i=1}^N \left(\frac{d_i - \sum_j K'_{ij} m_j}{\sigma_i} \right)^2, \quad (4)$$

where the sum over i is over the first two matrix rows in (2), K'_{ij} is the submatrix formed by those rows, σ_i is the estimated uncertainty of datum d_i , and N is the total number of data. If σ_i is correctly estimated, χ^2/N should be close to 1 for a model that fits the data however which is not forced to fit the data closer than the (1σ) uncertainty bars on average.

Fig. 8 shows the trade-off curves for $\delta \ln V_S$ and $\delta \ln Q_S^{-1}$, with a fixed ratio $\epsilon_1/\epsilon_2 = 1$. The preferred model has $\chi^2/N = 0.966$. After the scaling as described in Section 3, the norms (roughness) of $\delta \ln V_S$ and $\delta \ln Q_S^{-1}$ are of the same order, suggesting that the estimates for prior uncertainties in velocity and attenuation are reasonable. Fig. 8(b) shows that in the mantle the variation of Q_S is smoother than that of V_S , which may reflect a lack of resolution rather than a physical cause (see Appendix B1).

4.2 S velocity structure of the western US upper mantle

A catalog of the western US $\delta \ln V_S$ structure from 100 to 800 km depth can be found in Figs A1–A2 (centre panels). In this section, we focus on some identified features of the velocity model.

Fig. 9 shows $\delta \ln V_S$ at 100, 200 and 300 km depths, and Fig. 10 displays 6 cross-sections with depth. The divide between slow mantle in the western part and deep fast cratonic lithosphere in the east is apparent along the Rocky Mountains. The craton root extends below 200 km. It is most pronounced beneath the Wyoming Basin, with a strong fast anomaly reaching deeper than anywhere else beneath the Rocky Mountains, as imaged by Sigloch *et al.* (2008). The deepest part is near the northern tip of the Wyoming Basin, apparently deeper than 300 km (sections BB', FF'). The large heterogeneity revealed by the high-resolution models seems to complicate the otherwise attractive explanation of the apparent large craton depth through radial anisotropy (Gung *et al.* 2003; Marone & Romanowicz 2007). The western limit of the cratonic root is visible beneath the Wyoming Basin, and there is a sharp boundary between the craton and the mantle west of it (section BB').

In Fig. 9, the subduction of the Juan de Fuca Plate along the Cascades is expressed by a fast velocity anomaly in the shallow mantle. The subduction extends as far south as the latitude of the Mendocino Triple Junction (40°N), where the transform boundary (south of 40°N) between the Pacific Plate and the North American Plate switches to convergent boundary (north of 40°N) between the Juan de Fuca Plate and the North American Plate. At 200–300 km depth a discontinuity in subduction is visible as an absence of fast velocity anomalies around 45°N. Section CC' shows that this ‘slab

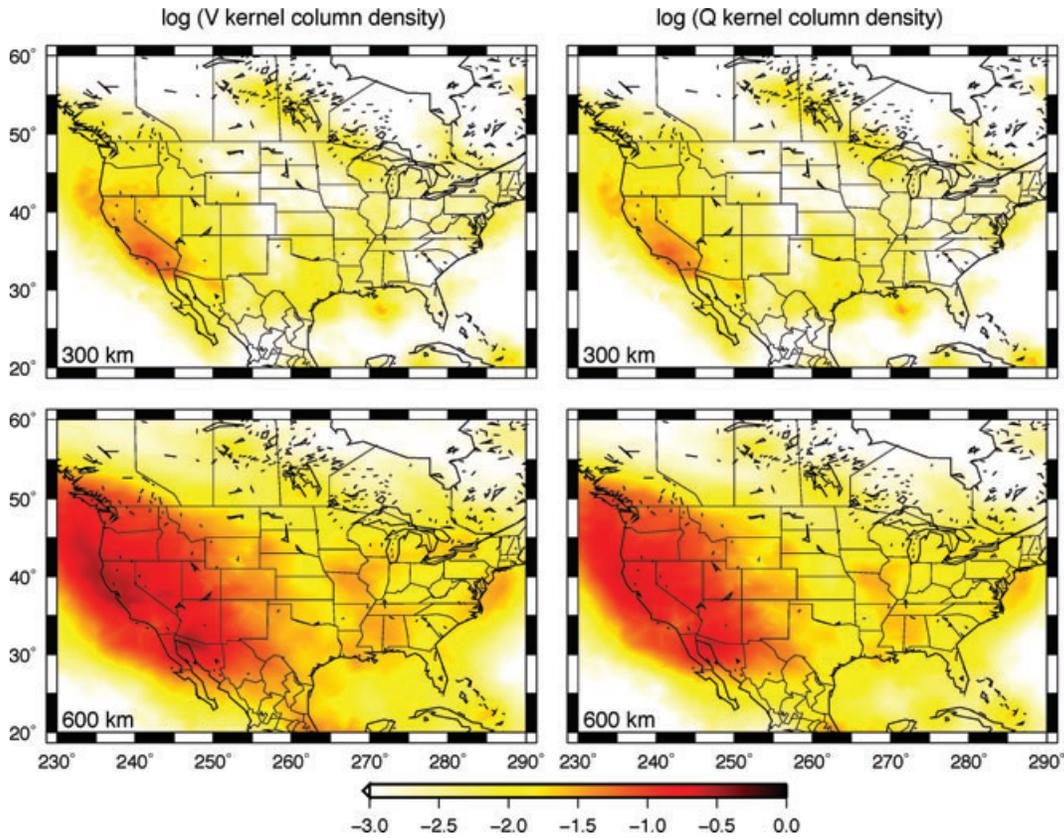


Figure 7. Sensitivity of the model parameters V_S and Q_S to the data. The maps show the logarithm (base 10) of the kernel column density (eq. 3) for V_S and Q_S at 300 km and 600 km depths, with the absolute level of $\max_j \sum_i K_{ij} = 4381$ s for the V_S kernel and $\max_j \sum_i K_{ij} = 552$ for the Q_S kernel. Zero means maximum density.

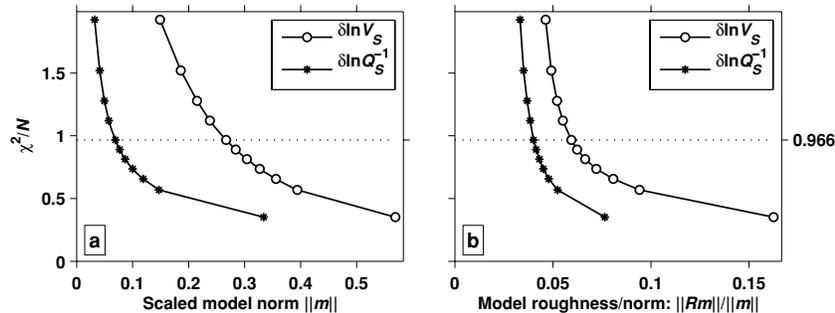


Figure 8. Trade-off curves for the model $m(\delta \ln V_S$ or $\delta \ln Q_S^{-1})$ with fixed ratio $\epsilon_1/\epsilon_2 = 1$ (see eq. 2). Different points on the same curve correspond to different absolute values of ϵ_1 (or ϵ_2). χ^2/N is defined by eq. (4). (a) Trade-off between fitting the data and norm damping. (b) Trade-off between fitting the data and smoothing the model. R is the Laplacian roughening operator. The dotted line indicates the preferred model with $\chi^2/N = 0.966$.

hole’ extends from 10 to 300 km depth, and indicates a recent break of the slab beneath the Cascades.

A back-of-the envelope calculation allows us to assign an approximate time at which this hole was formed. If the current subduction rate is 40 mm yr^{-1} (Gripp & Gordon 2002), the 100 km length of the slab measured from the surface suggests that the slab break occurred within the last 2.5 Myr. Meanwhile, the slab break should occur early enough to allow the 200-km-high ‘slab hole’ to develop. A ‘slab hole’ around 45°N is also observed by Burdick *et al.* (2008), but at 400–600 km, and by Roth *et al.* (2008) at 200–400 km. Roth *et al.* (2008) suggested that this ‘hole’ is an inversion artefact as a consequence of limited ray path coverage. However, our resolution test suggests that the ‘slab hole’ is real (see Appendix B2 and Fig. B4), at least as far as our data set is concerned.

We find two strong fast anomalies at 100 km depth in southern California (Fig. 9). The northern one has been observed in other tomographic studies and was explained by Zandt *et al.* (2004) as the dense, ultramafic root of the southern Sierra Nevada Mountains, which was destabilized and removed 3–10 Myr ago, and sank to the west of the southern Sierras. The location of the fast anomaly observed in our model is consistent with their study, and our model suggests that the mountain root has sunk to 200 km depth (Fig. 9b). The southern fast anomaly is beneath the Transverse Range and extends to ~ 150 km depth. Humphreys & Hager (1990) explained it as the convergence and sinking of the sub-crustal lithosphere; Prindle & Tanimoto (2006) explained its western part as a remnant of old oceanic plate which rotated to its current place.

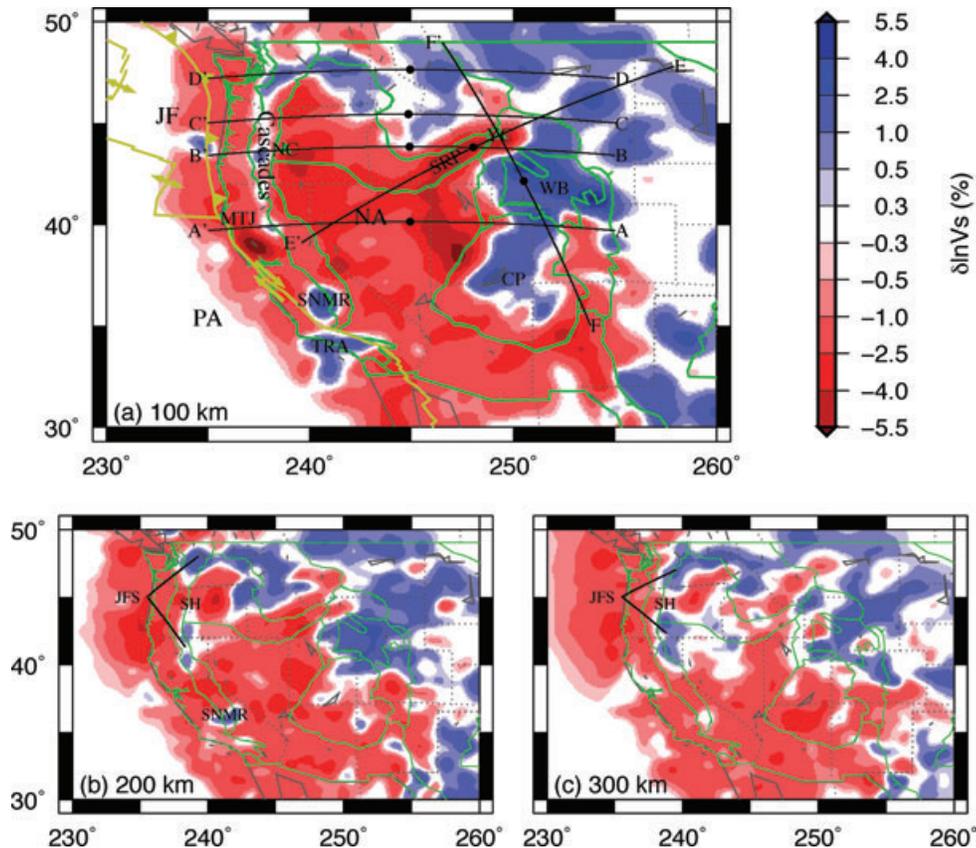


Figure 9. (a) Map view of $\delta \ln V_S$ at 100 km depth. Yellow lines delineate the plate boundaries. PA: Pacific Plate, NA: North American Plate, JF: Juan de Fuca Plate. The intersection of the three plates is the Mendocino Triple Junction (MTJ). Green lines delineate geological features and dotted lines are state boundaries. Cascades: Cascade Mountains, SRP: Snake River Plain, WB: Wyoming Basin, CP: Colorado Plateau. Green triangles represent the Newberry Caldera (NC) and the Yellowstone Caldera (YC). The two fast anomalies in southern California are the southern Sierra Nevada Mountains root (SNMR) and the Transverse Range Anomaly (TRA). Solid black lines are great circle arcs corresponding to the cross-sections in Fig. 10, and the black dot represents the centre of the line. (b) and (c) Map views of $\delta \ln V_S$ at 200 and 300 km depths. JFS: Juan de Fuca slab, SH: slab hole, SNMR: southern Sierra Nevada Mountains root.

Several smaller details of interest can be observed. Various deep sources of volcanism can be recognized in the tomographic model. In section BB', a very slow anomaly as large as -5.5 per cent is observed beneath the Newberry Caldera (NC) and right above the subducting Juan de Fuca slab. The lack of deeper origin of this slow anomaly suggests that it could be caused by volatile release from the subducting slab. A more detailed study of the Newberry Hotspot Track by Xue & Allen (2006) suggests that the migration pattern of hot material to the surface is caused by lithosphere-controlled processes. Strong slow anomalies are also observed along the Snake River Plain (Fig. 9a and section EE'), which are probably reflecting the plate motion over the Yellowstone hotspot. A shallow thermal anomaly is currently located beneath the Yellowstone Caldera, visible as a very slow anomaly of as much as -8.0 per cent (sections EE' and FF').

We observe a possible deep trace of a Yellowstone plume as a vertically oriented, weak slow anomaly down to 1000 km depth, different from the images of plumes dipping northwest obtained by local tomographic studies (Yuan & Dueker 2005; Waite *et al.* 2006). Though section EE' shows no clear connection between the slow anomalies beneath the surface and below the 410-km discontinuity, a shallow low velocity anomaly north of the current caldera is connected to a continuous conduit visible down to 1000 km depth in section FF'. Is a second (older) conduit extending southward to 500 km depth? The situation is extremely complicated, with a

fast anomaly between the plume head and the deeper origin that is identified as a fragment of the Farallon slab in both our model and the *P* model by Sigloch *et al.* (2008). Lack of agreement between various studies argues for caution in interpretation. More definite images may become available when USArray has moved further east.

4.3 Comparison of *S*- and *P*-velocity models

In their high-resolution *P*-velocity model, Sigloch *et al.* (2008) observe heavy fragmentations of the younger part of the subducted Farallon slab. We focus the comparison of *S*- and *P*-velocity models on the subduction history of the Farallon Plate.

Sigloch *et al.* (2008) observe a 'slab gap' parallel to the direction of plate subduction, expressed as the absence of fast *P*-velocity anomalies from near the trench in Oregon to southern Saskatchewan in Canada, shown as the black line in Fig. 11(a). It divides the Farallon slab into northern parts (N1 and N2) and southern parts (S1 and S2). In Fig. 11(b), the 'slab gap' at almost the same location is plotted on top of the *S* velocity model. It is clear that the *S* model has the 'slab gap' and Farallon slab segments (S1, N1, S2, and N2) coinciding well with similar features in the *P* model. Limited by the data coverage, the *S* model only reveals the top parts of S2 and N2, compared with the *P* model by Sigloch *et al.* (2008). Note that in Fig. 11(b), the fast anomaly beneath the eastern tip of the 'slab gap'

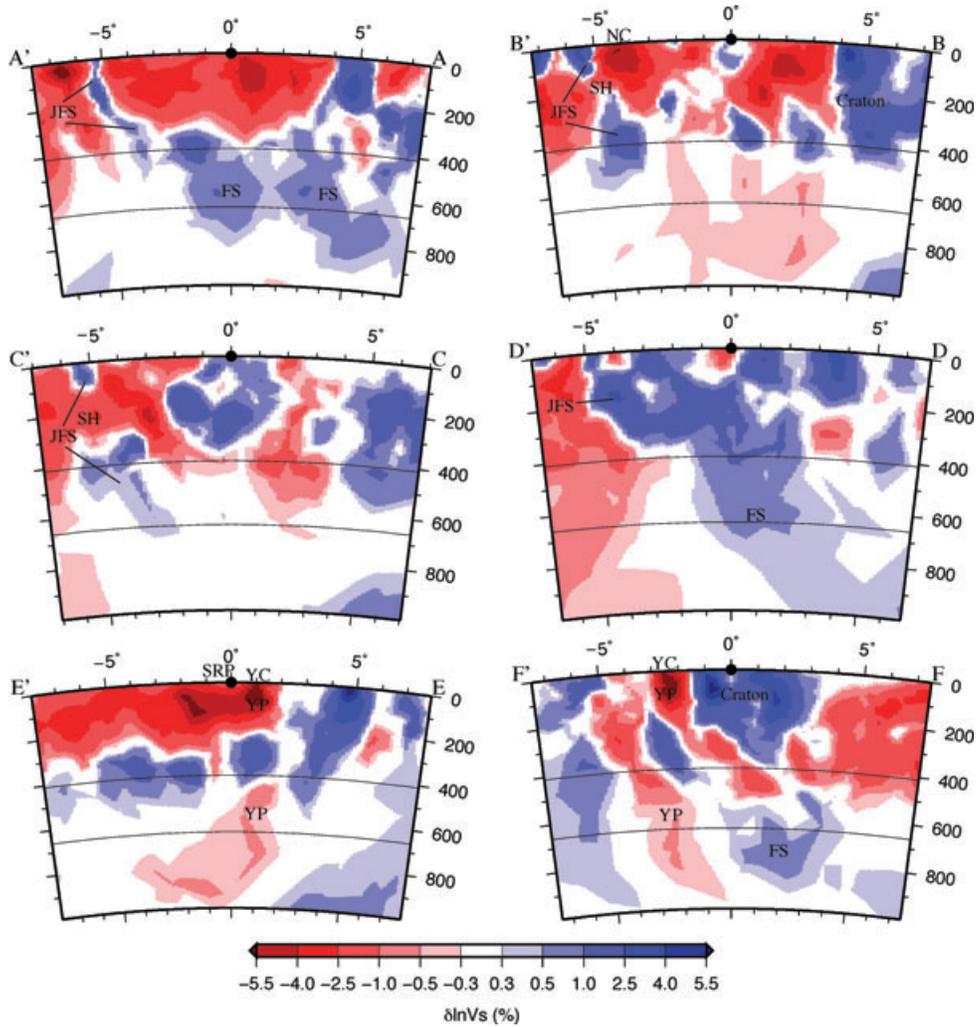


Figure 10. Great-circle cross-sections of $\delta \ln V_S$ till 1000 km depth, corresponding to the arcs (black solid lines) in Fig. 9. The black dot and 0° represent the midpoint of the great circle arc, corresponding to the black dot in Fig. 9. The 410- and 660-km discontinuities are indicated. JFS: Juan de Fuca slab, FS: Farallon slab, SH: slab hole, NC: Newberry Caldera, SRP: Snake River Plain, YC: Yellowstone Caldera, YP: Yellowstone plume.

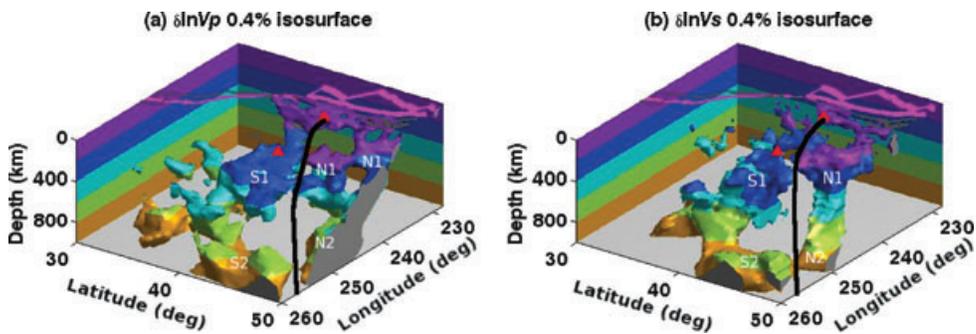


Figure 11. Three-dimensional subduction systems of the western US down to 1000 km depth from (a) P velocity model (Sigloch *et al.* 2008) and (b) S velocity model. The $+0.4$ per cent isosurface of velocity perturbation is plotted, with shallow, fast structures not related to subduction (e.g. craton, Colorado Plateau root) removed. The depth is coloured. The bird's eye view is from northeast. The view and the not displayed parts of shallow fast anomalies are the same for P and S models. The pink line represents plate boundaries. The two red triangles represent the Newberry Caldera (west) and the Yellowstone Caldera (east). The black line indicates the 'slab gap', a long continuous path in absence of fast velocity anomalies. It is the same in (a) and (b), except that at 100–300 km depths, the 'slab gap' in (b) is about 1° south to that in (a). S1, N1, S2 and N2 represent the segments of the Farallon slab identified by Sigloch *et al.* (2008).

more likely belongs to the Juan de Fuca slab. The 'slab gap' can also be seen in Fig. A2 (centre and right-hand panels), as the division between two major fast velocity fragments on the north and south. In addition, the 'slab gap' shows in the lateral cross sections AA' to

DD'. At the southernmost section AA' as well as section DD' in the north, there are large volumes of fast anomalies in and below the transition zone in the central and eastern parts (i.e. far away from the trench). On the other hand, such fast anomalies are not observed

in sections BB' and CC', because they cut through the 'slab gap'. The 'slab gap' is the longest break of the Farallon slab beneath the western US observed by Sigloch *et al.* (2008). They indicated that the trench-perpendicular break of the Farallon slab may have modified subduction dynamics by changing a slab's aspect ratio. It could be responsible for the complex space-time evolution of the post-Laramide magmatic events, and for the breakup of the Juan de Fuca Plate in the last 10 Myr. The observation of similar slab break from independent S velocity models further supports their argument.

The biggest differences between P and S models, as far as subduction is concerned, are in the images of the Juan de Fuca slab. In the P model, the most recently subducted plate material beneath the Cascades seems to break up at around 200 km depth, whereas in the S model it seems to remain intact down to 400 km.

A complete comparison of P and S velocity models is shown in Figs A1–A2 (centre and right-hand panels). Similar structures can be identified in the two models. For example, above 300 km depth, the Juan de Fuca slab, the craton root beneath the Wyoming Basin, the southern Sierra Nevada Mountains root, and the Transverse Range Anomaly; beneath 400 km depth, the 'slab gap' and the two fast velocity fragments divided by it, as discussed above. The 3-D correlation coefficient between these two models in the region of 30°–50°N, 230°–260°E and 0–1000 km depth is 0.673, which we consider high. The correlation coefficient is influenced by many factors and may be low even when the human eye can pick out 'structures' that correlate very well. Many tomographers, including us, therefore prefer figures over numbers to study the correlation between V_P and V_S models. The positive correlation coefficient is visually confirmed in Figs A1–A2.

Meanwhile it is worthy noticing that compared with the S -wave data set, the P -wave data set contains more high-frequency (around 2.5 s) information, and provides much better coverage in the eastern US (59 801 source–receiver pairs for P waves versus 22043 for S waves). As a result, the P velocity model has a vertical resolving length of ~ 230 km in the upper mantle, and resolves the eastern US upper mantle fairly well (Sigloch 2008), while the S velocity model has a vertical resolving length of ~ 350 km in the western US, and can poorly resolve the eastern US upper mantle.

4.4 Relationship between velocity and attenuation heterogeneities

4.4.1 Relative importance of $\delta \ln V_S$ and $\delta \ln Q_S$ in explaining the data

Sigloch (2008) concludes that focusing dominates the P wave amplitudes so strongly that little information about attenuation can be extracted from the data at the present signal-to-noise ratio. Since attenuation is stronger for S waves, we cannot *a priori* assume that such a conclusion is also valid for S waves. One way to compare the influence of each factor is to compare the relative weight of the attenuation and the velocity component to the model in contributing to the data fit. To this end we introduce the following ratios between the product of the kernel times the model perturbation and the corresponding data norm, both in a rms sense (see also eq. 2)

$$c_{TV} = \frac{|\mathbf{K}_V^T \delta \ln V_S|}{|\delta T|},$$

$$c_{TQ} = \frac{|\mathbf{K}_Q^T \delta \ln Q_S^{-1}|}{|\delta T|},$$

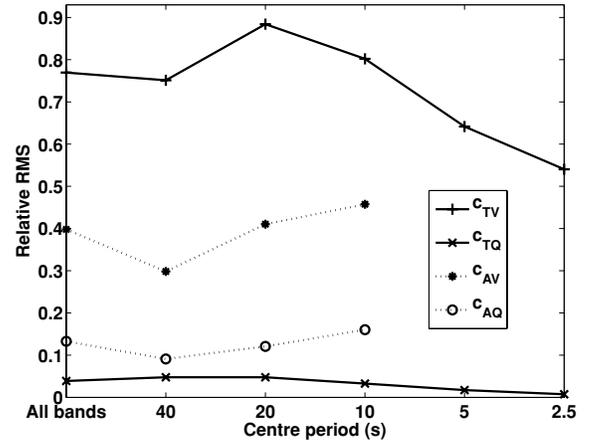


Figure 12. The relative weight of $\delta \ln V_S$ and $\delta \ln Q_S$ in contributing to the data fit. It is measured as the rms ratio of the predictions relative to the observations. The observations are δT in the case of c_{TV} , c_{TQ} and $\delta \ln A$ in the case of c_{AV} , c_{AQ} (see eq. 5).

$$c_{AV} = \frac{|\mathbf{K}_V^A \delta \ln V_S|}{|\delta \ln A|},$$

$$c_{AQ} = \frac{|\mathbf{K}_Q^A \delta \ln Q_S^{-1}|}{|\delta \ln A|}. \quad (5)$$

Here c_{TV} and c_{TQ} measure the relative contribution of velocity and attenuation in interpreting delay times, respectively, while c_{AV} and c_{AQ} measure the relative contribution of velocity and attenuation in interpreting amplitude anomalies, respectively.

Fig. 12 shows these contributions and their variation with frequency. A vector sum of the contributions should sum to 1 (within the uncertainty bounds) if the correction terms in eq. (2) can be ignored. The discrepancy between 1 and the sum of absolute values thus gives a lower limit to the influence of the corrections. With this in mind we observe the following.

- (i) The velocity model dominates over the attenuation model in explaining the data since for all frequencies c_{TV} and c_{AV} are well above c_{TQ} and c_{AQ} .
- (ii) Since $c_{TV} \gg c_{TQ}$, the velocity component explains most of the observed delay times, and the traveltime dispersion due to Q is not important.
- (iii) Since $c_{AV} + c_{AQ} \ll 1$, a significant portion of the amplitude observations is explained by source and receiver amplitude corrections.
- (iv) With decreasing period, more of the delay time information is explained by the correction factors.

The fact that $\delta \ln V_S$ dominates over $\delta \ln Q_S$ in interpreting the data is also supported by inversions with different parameters.

(i) We experimented with velocity-only inversion, that is, removing $\delta \ln Q_S^{-1}$ and related matrices in eq. (2). The resulting velocity model is almost identical to the model obtained from the joint inversion. With the same velocity magnitude and roughness, χ^2/N (see eq. 4) is 0.994 for the velocity-only inversion, only slightly larger than for the joint inversion ($\chi^2/N = 0.966$). The c_{AV} (see eq. 5) for the velocity-only inversion is 0.30, which is 75 per cent of the c_{AV} for the joint inversion (0.40 as shown in Fig. 12). The above observations indicate that a velocity model alone can explain both traveltime and amplitude deviations, whereas attenuation has only a second-order effect.

(ii) To confirm the importance of focusing to explain the amplitude variations, we inverted the amplitude data separately for $\delta \ln Q_S^{-1}$ only, that is, removing $\delta T, \delta \ln V_S, C^T$ and related matrices in eq. (2). The attenuation model so obtained shows that high attenuation (low Q_S) anomalies and fast velocity anomalies coexist in the southern Sierra Nevada Mountains root and beneath the Transverse Range at 100 km depth. This coincidence can be explained as an artefact of the neglect of focusing: the fast velocity anomalies in fact defocus the seismic waves and produce small amplitudes at seismic stations. If we attribute the observed small amplitudes entirely to attenuation anomalies, high attenuation anomalies are needed.

We conclude that focusing is dominant even for S waves and must be taken into consideration when interpreting S wave amplitudes; inversion with attenuation as the only model parameter gives wrong results.

4.4.2 Correlation between $\delta \ln V_S$ and $\delta \ln Q_S$

Further insight into the effect of attenuation on amplitudes, and possibly into the cause of attenuation, comes from a comparison between velocity and attenuation anomalies. We plot a comparison of $\delta \ln V_S$ and $\delta \ln Q_S$ at evenly spaced points in the western US mantle (0–800 km depth) in Fig. 13(a). The point spacing is 60 km, about the minimum spacing of the tetrahedral grid, and thus the following analysis is in ‘broadband’, that is, containing information of both short and long wave lengths. The anomalies of V_S and Q_S are positively correlated, with a correlation coefficient of 0.604, though the scatter is obviously large. The 95 per cent confidence interval of the correlation coefficient is [0.588, 0.621], and the p -value for testing the hypothesis of no correlation is smaller than 0.001. Both suggest that the correlation is significant. A major-

ity of the points fall into the region of low velocity (-3 per cent $\leq \delta \ln V_S \leq 0$ per cent) and high attenuation (-20 per cent $\leq \delta \ln Q_S \leq 0$ per cent), which is the characteristic of the shallow mantle (see Fig. A1 left-hand and centre panels). Such positive correlation is also observed in global models derived from surface waves (e.g. Romanowicz 1990; Artemieva *et al.* 2004; Dalton & Ekström 2006). The model given by Dalton & Ekström (2006) shows a much stronger correlation of 0.78 through degree 12 at 75-s period. Note that the linear relationship (i.e. the slope) varies with depth due to the variation of rock properties with depth, and it contributes to the relatively low average correlation coefficient in Fig. 13(a). This leads us to examine the correlation at each depth.

In Fig. 13(b), at each depth, the fact that the p -value for testing the hypothesis of no correlation is smaller than 0.05 and the 95 per cent confidence interval suggest that the correlation is significant. Correlation coefficients are above 0.5 at all depths. This suggests that thermal effects, which are shown to produce such positive correlation by experiments on olivine-rich rocks (Jackson *et al.* 2002), play a major role in forming velocity and attenuation anomalies, although volatiles cannot be ruled out either. On the other hand, the correlation coefficients are not close to 1.0 at most depths. A possible cause of the limited correlation is the trade-off of attenuation model against velocity model. Note that $\delta \ln V_S$ and $\delta \ln Q_S$ are coupled in affecting $\delta \ln A$ (see eq. 2). For example, a positive $\delta \ln A$ can be explained by a strong negative $\delta \ln V_S$ and a weak negative $\delta \ln Q_S$, but can also be explained by an intermediate negative $\delta \ln V_S$ and an intermediate positive $\delta \ln Q_S$. The correlation coefficients increase rapidly throughout the transition zone, where the heterogeneities have larger scale. It is largest below the transition zone, where there are less heterogeneities, especially for $\delta \ln Q_S$ (see Fig. A2 left-hand and centre panels).

On average, we find that $\delta \ln Q_S = 13.6 \delta \ln V_S$ (Fig. 13a), with the slope determined by the major axis direction of the error ellipse. The

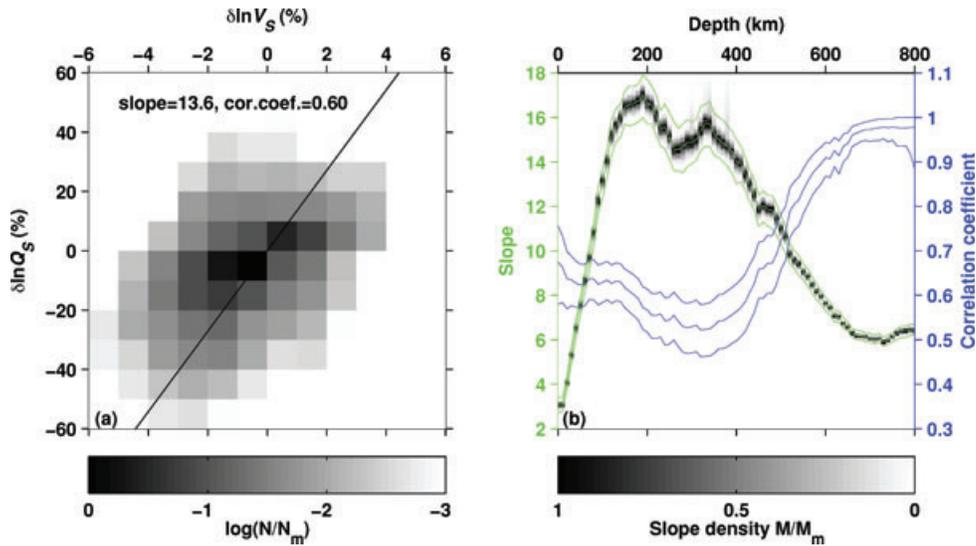


Figure 13. (a) Histogram of $(\delta \ln Q_S, \delta \ln V_S)$ pairs beneath the western US (0 to 800 km depth). The data are sampled at evenly spaced points. Note the logarithmic scale of the colour bar, where N is the number of points falling into each block and N_m is the maximum value of N . Here $N_m = 1578$, and $\log(N/N_m) = -3$ means $N = 1.6$. The solid line has a slope determined by the direction of the major axis of the error ellipse: $\delta \ln Q_S = 13.6 \delta \ln V_S$. The overall correlation coefficient is 0.60. (b) Variation of the correlation coefficient (blue) and the slope (green) with depth. The middle blue line represents the estimated correlation coefficient, and the upper and lower blue lines represent the 95 per cent confidence interval. At each depth, the slope is estimated as the direction of the major axis of the error ellipse. The slope uncertainty is estimated with a jackknife-based resampling technique. The three green lines (from top to bottom) give the upper bound, mean, and lower bound of the slope estimate. At each depth, the grey scale colour describes the distribution of the slope estimate, where M is the number of points falling into each slope interval, and M_m is the maximum value of M . Points with $|\delta \ln V_S| < 0.3$ per cent or $|\delta \ln Q_S| < 3$ per cent are not used in producing (a) and (b).

slope is very close to that obtained by Dalton & Ekström (2006), which is 14.5 for 150-s Rayleigh waves. Because the temperature sensitivity of Q_S (exponential dependence) is larger than that of V_S (linear dependence), the large slopes at 100–420 km depths in Fig. 13(b) indicate that thermal effects may be most dominant in the upper mantle beneath the western US (except in the top 100 km). To obtain an idea of the uncertainty of the slope estimate, we use a resampling technique based on the jackknifing philosophy. At each depth, we estimate 1000 slopes from 1000 subsets of the complete set of $(\delta \ln V_S, \delta \ln Q_S)$ pairs. Each subset contains 90 per cent of the complete set from random resampling. Then the 1000 slopes are used to estimate the slope uncertainty, which is smaller than 2 as shown in Fig. 13(b) (grey scale colour and green lines). The largest slope uncertainties are at 150–500 km depths, corresponding to relatively low correlation coefficients. If the size of the subset is reduced to 70 per cent, the uncertainty is about doubled. We apply the same jackknifing technique (90 per cent subset) to estimate the uncertainty of the overall slope in Fig. 13(a). The standard deviation of the slope estimate is only 0.1.

Fig. 13(a) shows that there are regions where velocity and attenuation are anti-correlated (non-zero histogram in quadrants II and IV). Fig. A1 (left-hand and centre panels) demonstrates that the strongest negative correlation occurs at 100–300 km depths along the Cascades and in the Central Valley. In these regions, low velocity coincides with low attenuation. Experiments on olivine at 1200° C (about 100 km depth) suggest $\delta \ln Q_S = 0.26 \delta \ln d$ (d is the grain size), and little grain-size-sensitivity of the shear modulus for the grain size on the order of 10 μm (Jackson *et al.* 2002). This may be responsible for part of the negative correlation observed, assuming the extrapolation to the representative upper mantle pressure and grain size is correct. Li & Weidner (2008) found that, compared with a single phase, the coexisting phases at a phase transition re-

gion can significantly reduce P -wave velocity while having little effect on the bulk attenuation (Q_k). Two-phase mixture at 100–300 km depths may exist in regions of partial melting, which is likely to occur above the subducting slab due to water release. Although experiments on shear modulus and attenuation are not available, the P -wave results suggest a possible explanation to account for the anticorrelation. On the other hand, smearing of the attenuation model due to its poor resolution (see Appendix B1) may partially contribute to the anticorrelation.

4.5 Contribution of amplitudes to constraining velocity structure

Since we have seen that velocity structure has the most important effect on amplitudes, it is interesting to turn this around, and question how much extra information the amplitude measurements give to constrain the velocity structure. Amplitude perturbations are mostly sensitive to the second-order spatial derivative of velocity, and thus could help to sharpen the edges of narrow velocity heterogeneities. To investigate this aspect, we compare our velocity model with one obtained by inverting traveltimes only, ignoring the amplitude data (only keeping $\delta T, \delta \ln V_S, C^T$ and corresponding matrices in eq. 2). Comparison of the two models (Fig. 14) shows that the addition of amplitude information did change the small-scale velocity structures. While it is difficult to compare models not only with the same data fit but also the same model norm or roughness, because we deal with a different collection of model parameters, we can say something about the sharpness of the heterogeneity edges from these two models. We notice that the jointly inverted model is not significantly rougher than the model obtained from traveltimes only, which attests to the reliability of the amplitude data and our estimation of their uncertainty. Meanwhile, the jointly inverted model

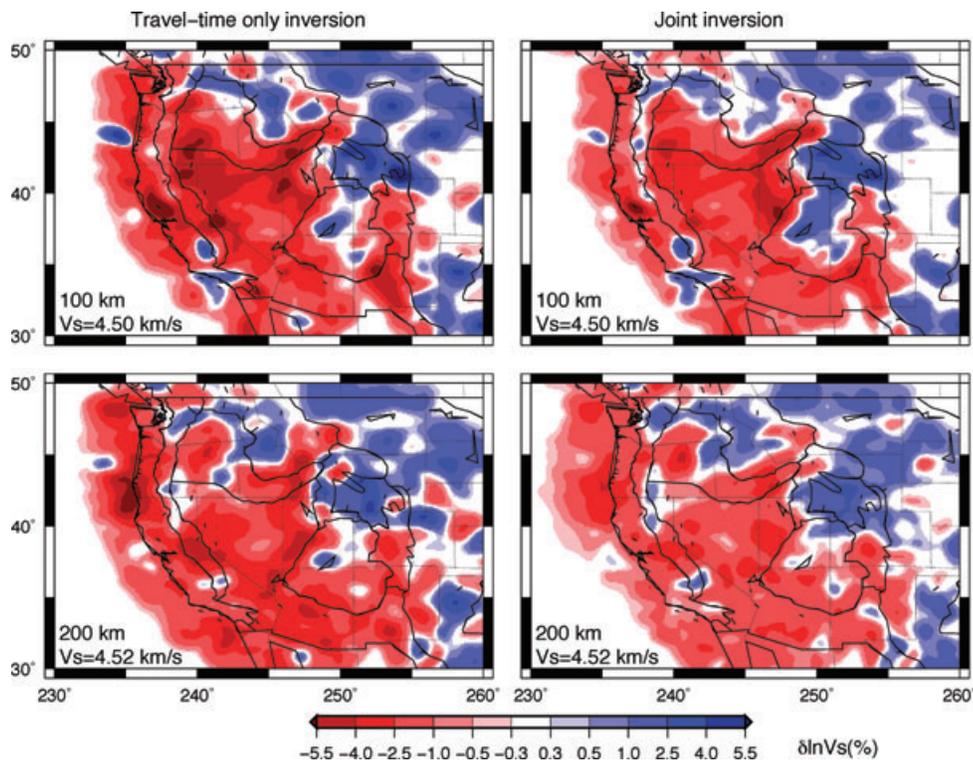


Figure 14. Comparison of shallow velocity structures from the inversion with δT and $\delta \ln V_S$ only (left-hand panel) and from the joint inversion (right-hand panel). Geological and political boundaries are plotted. Depth and the reference model value are shown at the lower left-hand corner of each map.

does show sharper heterogeneity edges for the following features: the fast anomalies along the Cascades (100 and 200 km depths), the fast anomaly beneath the southern Sierra Nevada Mountains (100 and 200 km depths), the fast anomaly beneath the Transverse Range (100 km depth), the boundary between the fast region beneath the Wyoming Craton and the Colorado Plateau and the slow region on the west (100 km depth). We conclude that amplitudes do provide extra constraints independent of traveltimes on velocity structures, and that amplitudes are especially helpful in sharpening the edges of narrow velocity heterogeneities.

5 CONCLUSIONS AND DISCUSSION

We estimate *SH*-wave traveltimes and amplitudes for five frequency bands, and obtain *S* velocity and attenuation structures of the western US upper mantle.

We obtain a high-resolution *S* velocity model beneath the western US that provides supporting evidence of the slab fragmentation in the western US. The *S* model reveals a 2000-km-long absence of fast anomalies, parallel to the direction of plate subduction, which coincides with the ‘slab gap’ observed by Sigloch *et al.* (2008), the longest break of the slab in the western US. Difference between *P* and *S* models for smaller-scale slab tears exists, however. The trench parallel tear at $\sim 255^\circ\text{E}$ (‘T’ in Sigloch *et al.* 2008) in the *S* model is not as long as in the *P* model. Above 700 km depth, both *P* and *S* models show no subducted slab south of 37°N , the eastern continuation of the Mendocino Transform Fault (‘Me’ in Sigloch *et al.* 2008). However, the separation of the deeper slab fragments south of ‘Me’ from the major Farallon slab north of ‘Me’ in the *S* model is not as clear as in the *P* model.

In addition, we reach the following major conclusions.

- (i) Traveltimes have a weaker frequency dependence and show a closer correlation with shallow tectonic structures than amplitudes.
- (ii) The *S* velocity image shows a ‘slab hole’ (lack of fast anomalies) around 45°N in the subducting Juan de Fuca slab. The Yellowstone plume seems to have an origin of weak slow velocity anomalies from ~ 1000 km depth and seems interrupted by a fast anomaly.
- (iii) The effect of $\delta \ln V_S$ dominates over the effect of $\delta \ln Q_S$. The $\delta \ln Q_S$ contribution to δT is about 5 per cent of the $\delta \ln V_S$ contribution. The $\delta \ln Q_S$ contribution to $\delta \ln A$ is about 30 per cent of the $\delta \ln V_S$ contribution.
- (iv) In general, $\delta \ln V_S$ and $\delta \ln Q_S$ are positively correlated, suggesting a common cause, most likely the thermal effect.
- (v) Velocity heterogeneities must be considered when interpreting amplitude anomalies, otherwise unreliable results may be obtained.
- (vi) Including the amplitude data has the effect of sharpening edges of narrow velocity heterogeneities in the shallow upper mantle.

Using multiple-frequency data that take full advantage of finite-frequency theory, in a joint inversion of both traveltimes and amplitudes, and the application of this method to USArray, one of the densest networks in the world, yields a very high resolution in the tomographic images.

ACKNOWLEDGMENTS

We thank Frederik Simons, Mark Panning, Nadine McQuarrie, Tom Duffy and Li Li for helpful discussions. This project was supported

by the US National Science Foundation under grant numbers EAR-0309298 and EAR-0105387. Part of the maps are produced with GMT (Wessel & Smith 1998).

REFERENCES

- Allen, R.M. *et al.*, 1999. The thin hot plume beneath Iceland, *Geophys. J. Int.*, **137**, 51–63.
- Artemieva, I.M., Billien, M., L  v  que, J.-J. & Mooney, W.D., 2004. Shear wave velocity, seismic attenuation, and thermal structure of the continental upper mantle, *Geophys. J. Int.*, **157**, 607–628.
- Bassin, C., Laske, G. & Masters, G., 2000. The current limits of resolution for surface wave tomography in {North America}, *EOS, Trans. Am. geophys. Un.*, **81**, F897.
- Billien, M., L  v  que, J.-J. & Trampert, J., 2000. Global maps of rayleigh wave attenuation for periods between 40 and 150 seconds, *Geophys. Res. Lett.*, **27**, 3619–3622.
- Burdick, S. *et al.*, 2008. Upper mantle heterogeneity beneath North America from travel time tomography with global and USArray transportable array data, *Seism. Res. Lett.*, **79**, 384–392.
- Butler, R., 1983. *P*-wave travel time and amplitude in western North America, *Nature*, **306**, 677–678.
- Chapman, C.H., 1978. A new method for computing synthetic seismograms, *Geophys. J. R. astr. Soc.*, **54**, 481–518.
- Dahlen, F.A. & Baig, A.M., 2002. Fr  chet kernels for body-wave amplitudes, *Geophys. J. Int.*, **150**, 440–466.
- Dahlen, F.A., Hung, S.-H. & Nolet, G., 2000. Fr  chet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.*, **141**, 157–174.
- Dalton, C.A. & Ekstr  m, G., 2006. Global models of surface wave attenuation, *J. geophys. Res.*, **111**, B05317.
- Durek, J.J., Ritzwoller, M.H. & Woodhouse, J.H., 1993. Constraining upper mantle anelasticity using surface wave amplitude anomalies, *Geophys. J. Int.*, **114**, 249–272.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.*, **25**, 297–356.
- Gripp, A.E. & Gordon, R.G., 2002. Young tracks of hotspots and current plate velocities, *Geophys. J. Int.*, **150**, 321–361.
- Gung, Y. & Romanowicz, B., 2004. {*Q*} tomography of the upper mantle using three-component long-period waveforms, *Geophys. J. Int.*, **157**, 813–830.
- Gung, Y., Panning, M. & Romanowicz, B., 2003. Global anisotropy and the thickness of continents, *Nature*, **422**, 707–710.
- Humphreys, E.D., 1995. Post-Laramide removal of the Farallon slab, western United States, *Geology*, **23**, 987–990.
- Humphreys, E.D. & Hager, B.H., 1990. A kinematic model for the late Cenozoic development of southern California crust and upper mantle, *J. geophys. Res.*, **95**, 19 747–19 762.
- Humphreys, E., Hessler, E., Dueker, K., Erslev, E., Farmer, G. & Atwater, T., 2003. How Laramide-age hydration of North America by the Farallon slab controlled subsequent activity in the western U.S., *Int. Geol. Rev.*, **45**, 575–595.
- Hung, S.-H., Dahlen, F.A. & Nolet, G., 2000. Fr  chet kernels for finite-frequency traveltimes—II. Examples, *Geophys. J. Int.*, **141**, 175–203.
- Hung, S.-H., Shen, Y. & Chiao, L.-Y., 2004. Imaging seismic velocity beneath the Iceland hot spot: a finite-frequency approach, *J. geophys. Res.*, **109**, B08305.
- Jackson, I., Gerald, J.D.F., Faul, U.H. & Tan, B.H., 2002. Grain-size-sensitive seismic wave attenuation in polycrystalline olivine, *J. geophys. Res.*, **107**(B12), 2360, doi:10.1029/2001JB001225.
- Kennett, B.L.N. & Engdahl, E.R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429–465.
- Li, L. & Weidner, D.J., 2008. Effect of phase transitions on compressional-wave velocities in the Earth’s mantle, *Nature*, **454**, 984–986.
- Marone, F. & Romanowicz, B., 2007. The depth distribution of azimuthal anisotropy in the continental upper mantle, *Nature*, **447**, 198–201.

- Nolet, G., 1987. Seismic wave propagation and seismic tomography, in *Seismic Tomography*, pp. 1–23, ed. Nolet, G., Reidel, Dordrecht.
- Nolet, G., 2008. *A Breviary of Seismic Tomography*, Cambridge University Press, Cambridge, UK.
- Paige, C. & Saunders, M., 1982. LSQR: an algorithm for sparse, linear equations and sparse least squares, *A.C.M. Trans. Math. Softw.*, **8**, 43–71.
- Persson, P.-O. & Strang, G., 2004. A simple mesh generator in Matlab, *SIAM Rev.*, **46**, 329–345.
- Prindle, K. & Tanimoto, T., 2006. Teleseismic surface wave study for S-wave velocity structure under an array: Southern California, *Geophys. J. Int.*, **166**, 601–621.
- Romanowicz, B., 1990. The upper mantle degree 2: constraints and inferences from global mantle wave attenuation measurements, *J. geophys. Res.*, **95**, 11 051–11 071.
- Roth, J.B., Fouch, M.J., James, D.E. & Carlson, R.W., 2008. Three-dimensional seismic velocity structure of the northwestern United States, *Geophys. Res. Lett.*, **35**, L15304, doi:10.1029/2008GL034669.
- Selby, N.D. & Woodhouse, J.H., 2002. The Q structure of the upper mantle: constraints from Rayleigh wave amplitudes, *J. geophys. Res.*, **107**(B5), 2097.
- Sigloch, K., 2008. Multiple-frequency body-wave tomography, *PhD thesis*, Princeton University, 249 pp.
- Sigloch, K. & Nolet, G., 2006. Measuring finite-frequency body-wave amplitudes and traveltimes, *Geophys. J. Int.*, **167**, 271–287.
- Sigloch, K., McQuarrie, N. & Nolet, G., 2008. Two-stage subduction history under North America inferred from finite-frequency tomography, *Nature GEO*, **1**, 458–462.
- Tian, Y., Hung, S.-H., Nolet, G., Montelli, R. & Dahlen, F.A., 2007a. Dynamic ray tracing and traveltimes corrections for global seismic tomography, *J. Comput. Phys.*, **226**, 672–687.
- Tian, Y., Montelli, R., Nolet, G. & Dahlen, F.A., 2007b. Computing travel-time and amplitude sensitivity kernels in finite-frequency tomography, *J. Comput. Phys.*, **226**, 2271–2288.
- Tibuleac, I.M., Nolet, G., Michaelson, C. & Koulakov, I., 2003. P wave amplitudes in a 3-D Earth, *Geophys. J. Int.*, **155**, 1–10.
- van der Lee, S. & Nolet, G., 1997. Seismic image of the subducted trailing fragments of the Farallon plate, *Nature*, **386**, 266–269.
- Waite, G.P., Smith, R.B. & Allen, R.M., 2006. V_P and V_S structure of the Yellowstone hot spot from teleseismic tomography: evidence for an upper mantle plume, *J. geophys. Res.*, **111**, B04303, doi:10.1029/2005JB003867.
- Wessel, P. & Smith, W.H.F., 1998. New, improved version of the Generic Mapping Tools released, *EOS, Trans. Am. geophys. Un.*, **79**, 579.
- Xue, M. & Allen, R.M., 2006. Origin of the Newberry Hotspot Track: evidence from shear-wave splitting, *Earth planet. Sci. Lett.*, **244**, 315–322.
- Yang, Y. & Forsyth, D.W., 2008. Attenuation in the upper mantle beneath Southern California: physical state of the lithosphere and asthenosphere, *J. geophys. Res.*, **113**, B03308, doi:10.1029/2007JB005118.
- Yang, T., Shen, Y., van der Lee, S., Solomon, S.C. & Hung, S.-H., 2006. Upper mantle structure beneath the Azores hot spot from finite-frequency seismic tomography, *Earth planet. Sci. Lett.*, **250**, 11–26.
- Yuan, H. & Dueker, K., 2005. Teleseismic P -wave tomogram of the Yellowstone plume, *Geophys. Res. Lett.*, **32**, L07304, doi:10.1029/2004GL022056.
- Zandt, G., Gilbert, H., Owens, T.J., Ducea, M., Saleeby, J. & Jones, C.H., 2004. Active foundering of a continental arc root beneath the southern Sierra Nevada in California, *Nature*, **431**, 41–46.
- Zhou, Y., Nolet, G. & Dahlen, F.A., 2003. Surface sediment effects on teleseismic P wave amplitude, *J. geophys. Res.*, **108**(B9), 2417, doi:10.1029/2003JB002428.

APPENDIX A: A CATALOGUE OF MODEL MAPS

A comparison of the preferred attenuation and S -velocity models, and the P -velocity model by Sigloch *et al.* (2008) is shown in Figs A1–A2 from 100 to 800 km depth.

APPENDIX B: RESOLUTION TESTS

B1 Resolution tests with regularly spaced Gaussian balls

Gaussian ball resolution tests are performed to provide some sense of the reliability of the model. The input velocity and attenuation anomalies are equally spaced balls, in which the anomaly amplitude decays with increasing radius following a Gaussian curve. The spacing between adjacent Gaussian balls is large enough to avoid overlapping of the balls. The finite-frequency sensitivity kernels are used to produce the theoretical data. Added to the theoretical data is randomly produced noise, which has a normal distribution. The joint inversion is implemented on the perturbed theoretical data to produce the output model.

The dense USArray network, plus the multiple-frequency data, provides very high lateral resolution. We measure the resolving length by the $1/e$ diameter of the Gaussian ball, where the anomaly amplitude is $1/e$ of the peak amplitude at the centre. For $\delta \ln V_S$ (Fig. B1), the resolving length is 150 km above 400 km depth. The resolution gets worse towards deeper region, and the resolving length becomes 300 km below 600 km depth. In Fig. B1, it is very clear that the resolution deteriorates rapidly out of the USArray-covered region (west of 235°E, and east of 255°E, see Fig. 1b). The resolution out of the USArray-covered region is better at 600 and 800 km than at shallow depths, because the finite-frequency sensitivity kernels become broader towards deeper depth and thus sample a larger region (see Fig. 7).

Compared with velocity, attenuation is more poorly resolved (Fig. B2), both in terms of resolving length and amplitude recovery. Above 400 km depth, the resolving length for $\delta \ln Q_S$ is 300 km. The resolution becomes worse towards deeper depth, and the resolving length becomes 500 km below 600 km depth. Again, the resolution is best in the USArray-covered region (235°–255°E). According to the above results, the smallest blobs of $\delta \ln Q_S$ in Figs A1–A2 are not reliable. The resolution of both velocity and attenuation is independent of the signs of $\delta \ln V_S$ and $\delta \ln Q_S$ in the input model, that is, it is independent of whether the $\delta \ln V_S$ and $\delta \ln Q_S$ Gaussian balls at the same location have the same or opposite signs.

The depth resolution of $\delta \ln V_S$ is shown in Fig. B4. Along the vertical direction, our data set can resolve velocity anomalies with a size of at least 350 km, with some leaking. The resolving length is less than 200 km right beneath the surface. Thus, the slabs and hotspots observed in Fig. 10 are likely reliable. As observed in the lateral resolution test, the resolution becomes worse towards deeper region, especially the amplitude recovery. In addition, the resolution deteriorates out of the USArray-covered region (western and eastern ends of the cross-sections). The resolution of the eastern part can be improved by incorporating the latest USArray data.

B2 Resolution test of the ‘slab hole’

In order to investigate whether or not the ‘slab hole’ (discussed in Section 4.2 and shown in Figs 9 and 10 section CC’) is real, we design a resolution test as shown in Fig. B4. If there is a fast velocity anomaly sitting at the position of the ‘slab hole’, our data set allows us to resolve this fast anomaly with some vertical leaking. In our tomographic model (the ‘output’ model), no fast anomaly is observed in the ‘slab hole’. This suggests that in the real mantle (the ‘input’ model), there should be no fast anomaly at this position either. Therefore, the ‘slab hole’ we observed is likely to be real instead of an inversion artefact, which goes against the conclusion by Roth *et al.* (2008), for our data set.

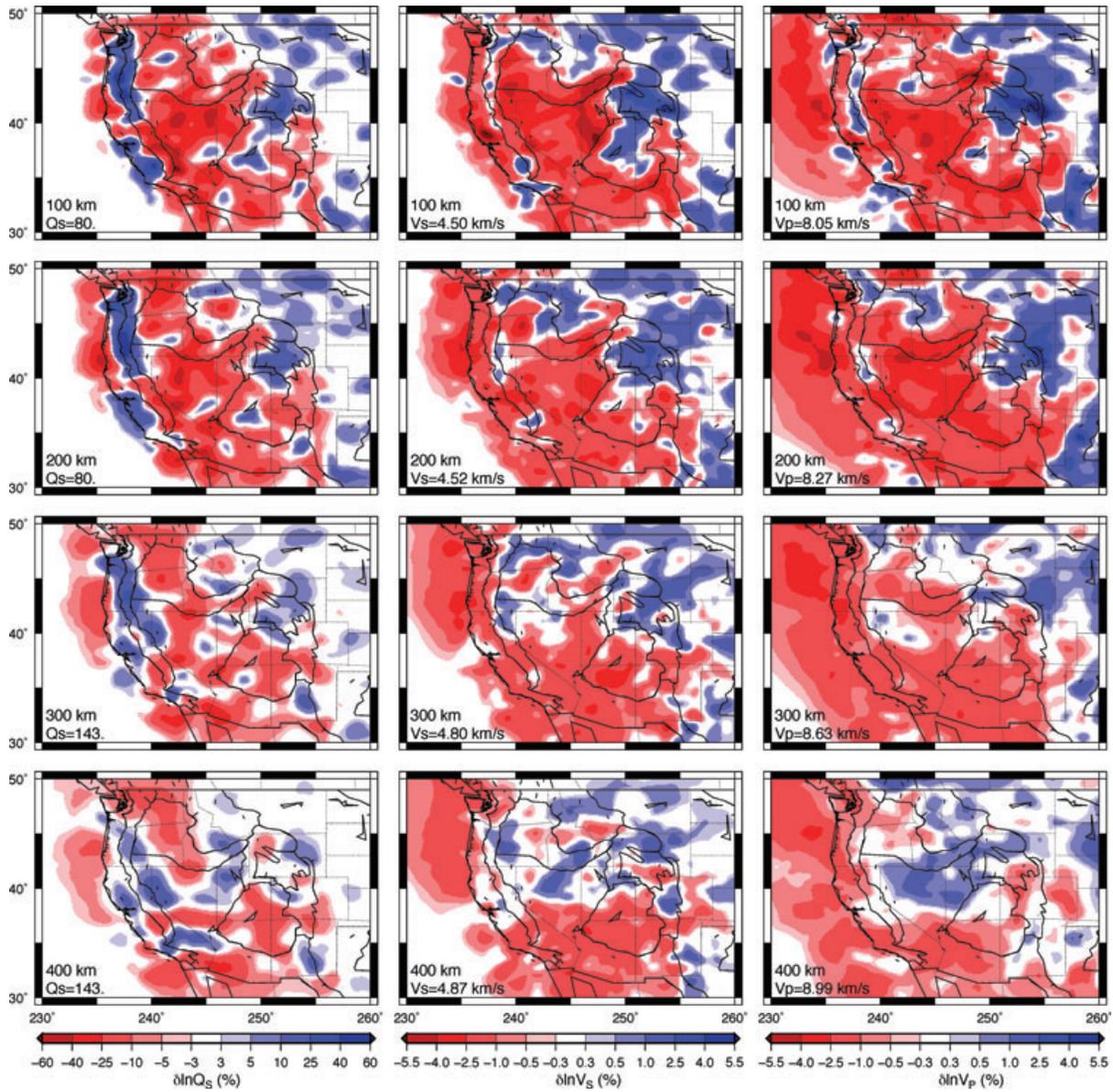


Figure A1. Map views of the preferred models of $\delta \ln Q_s$ (left-hand panel) and $\delta \ln V_s$ (centre panel), and the $\delta \ln V_p$ model (right-hand panel) by Sigloch *et al.* (2008), at 100, 200, 300 and 400 km depths in the western US. Geological and political boundaries are plotted. Depth and the reference model value are shown at the lower left-hand corner of each map.

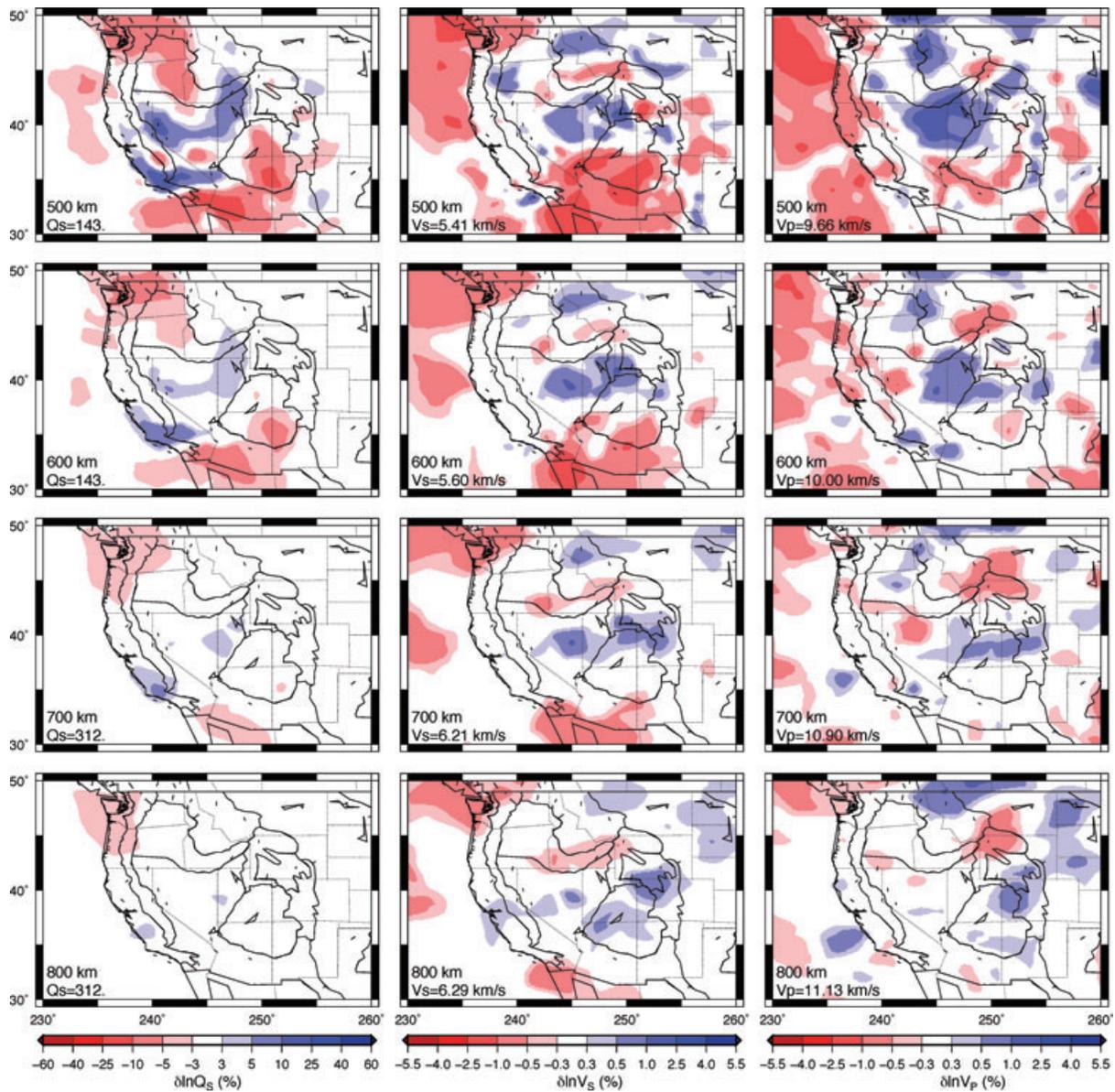


Figure A2. Map views of the preferred models of $\delta \ln Q_S$ (left-hand panel) and $\delta \ln V_S$ (centre panel), and the $\delta \ln V_P$ model (right-hand panel) by Sigloch *et al.* (2008), at 500, 600, 700 and 800 km depths in the western US. Geological and political boundaries are plotted. Depth and the reference model value are shown at the lower left-hand corner of each map.

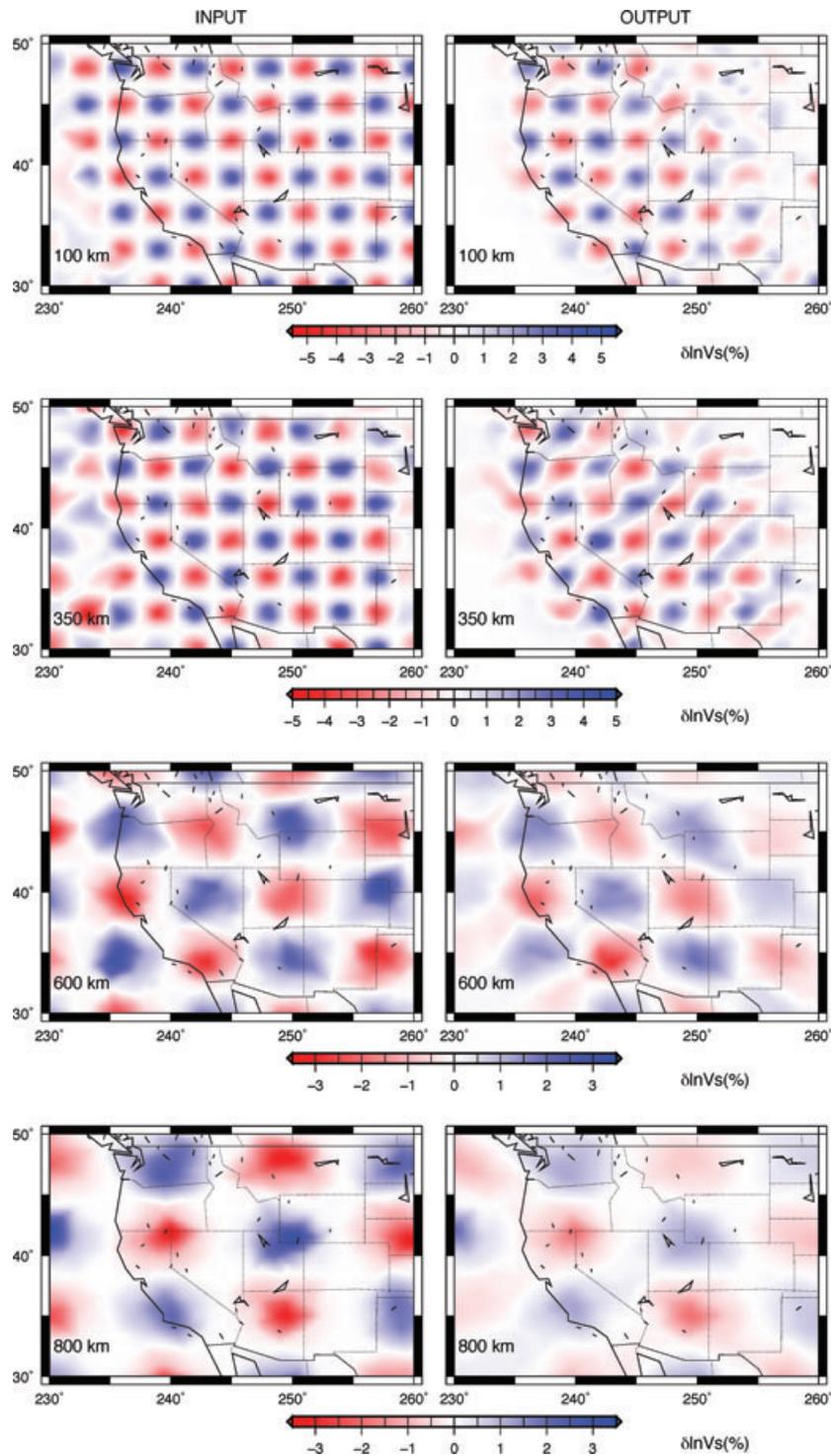


Figure B1. Lateral resolution for $\delta \ln V_s$. The depth of each map is indicated at the lower left-hand corner. The input anomalies are equally spaced balls, in which the anomaly amplitude decays with increasing radius following a Gaussian curve. At 100 and 350 km depths, the $1/e$ width of the Gaussian curve is 150 km. At 600 and 800 km depths, the $1/e$ width of the Gaussian curve is 300 km.

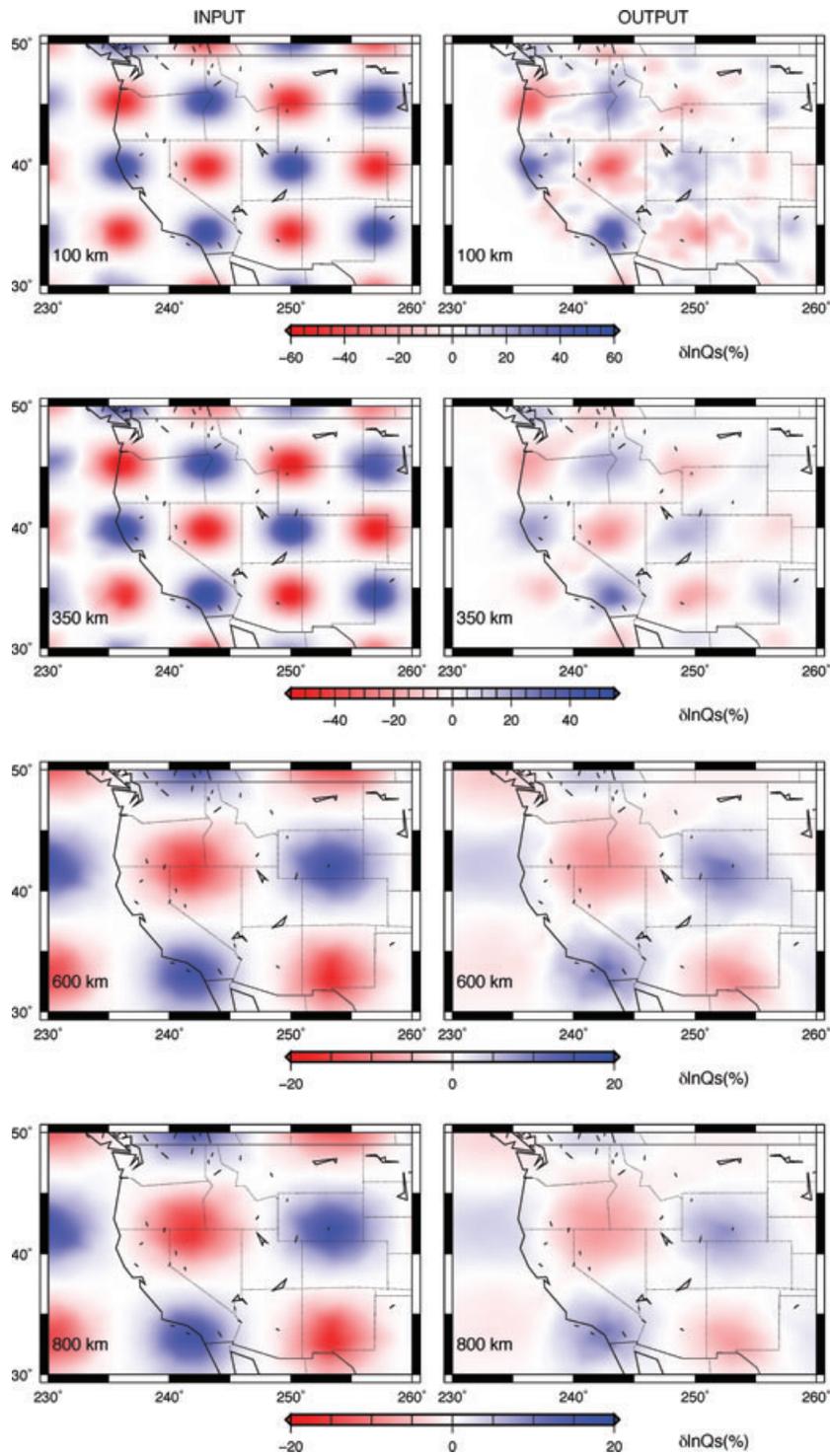


Figure B2. Lateral resolution for $\delta \ln Q_s$. The depth of each map is indicated at the lower left-hand corner. The input anomalies are equally spaced balls, in which the anomaly amplitude decays with increasing radius following a Gaussian curve. At 100 and 350 km depths, the $1/e$ width of the Gaussian curve is 300 km. At 600 and 800 km depths, the $1/e$ width of the Gaussian curve is 500 km.

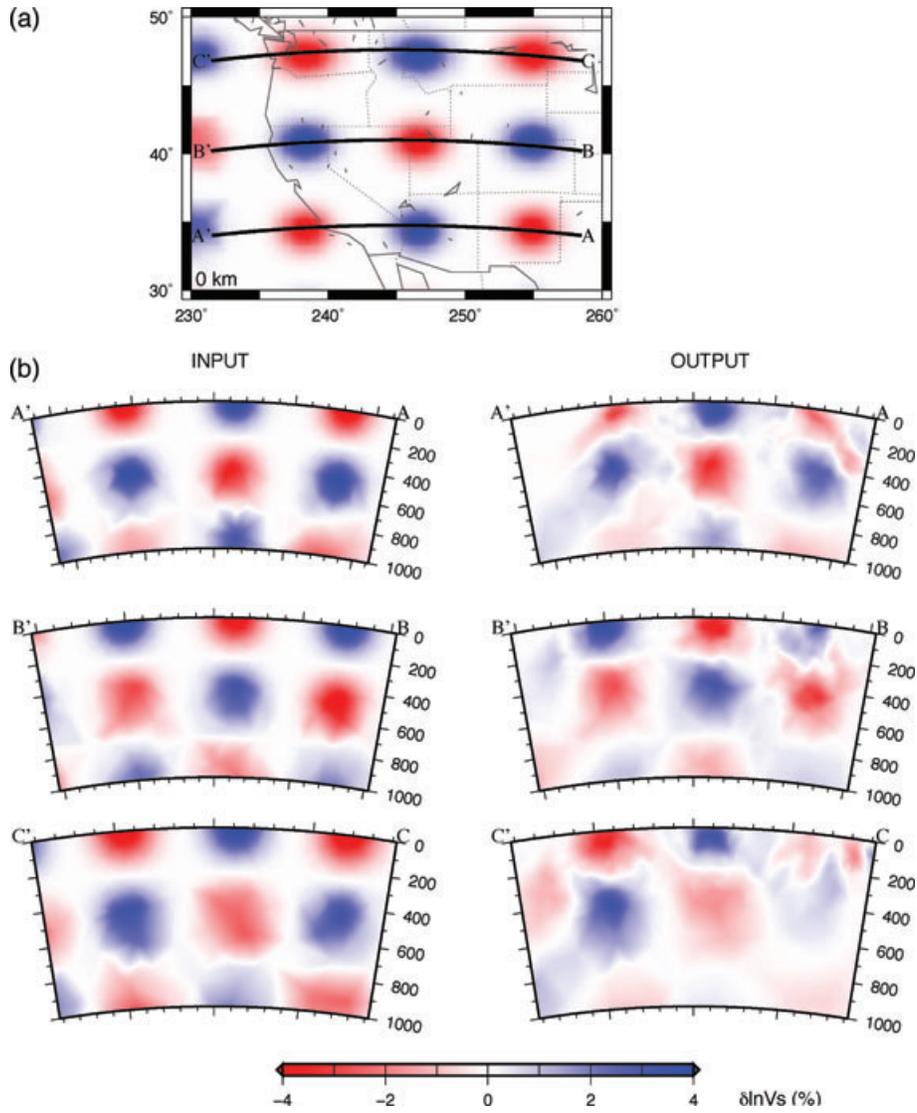


Figure B3. (a) Map view showing the locations of the cross-sections. (b) Vertical resolution for $\delta \ln V_S$. The input anomalies are equally spaced balls, in which the anomaly amplitude decays with increasing radius following a Gaussian curve. The $1/e$ width of the Gaussian curve is 300 km.

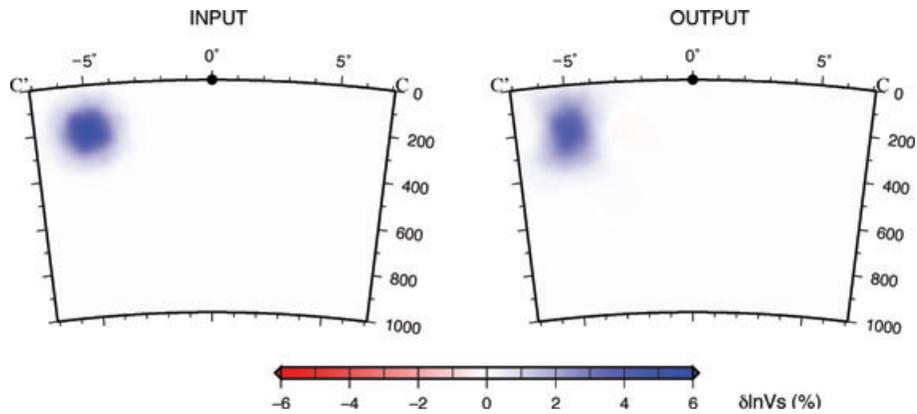


Figure B4. Resolution test of the 'slab hole'. The input model is a ball in which the anomaly amplitude decays with increasing radius following a Gaussian curve. The $1/e$ width of the Gaussian curve is 200 km. The ball is centred at (45.2°N, 238°E) and at 200 km depth, approximately the same location as the 'slab hole' shown in Fig. 10 section CC'. The cross-section is the same as the cross-section CC' in Figs 9(a) and 10.