Comparison of ray- and adjoint-based sensitivity kernels for body-wave seismic tomography

D. Mercerat, G. Nolet

Comparison of ray- and adjoint-based sensitivity kernels for body-wave seismic tomography

E. D. Mercerat[1,2] and G. Nolet[1]

Received 11 April 2012; revised 11 May 2012; accepted 15 May 2012; published 16 June 2012.

[1] We compare finite-frequency sensitivity kernels computed from ray-theoretical wavefields ("banana-doughnut" kernels) and from full waveform computations (often called "adjoint" kernels) in order to evaluate resolution, accuracy and computational cost. We focus here on body-wave seismic tomography at regional and local scales. Our results show that: (1) for homogeneous reference media, ray-based and adjoint kernels agree except for the expected differences in the regions close to the source and the receiver, where near-field contributions are neglected in the ray-based kernels; (2) for a smooth 3D background velocity model, the differences in predicted delay times for the two methods are generally well below 10 % of the delay for P waves, though as much as 20 % for S-waves, suggesting that extra care should be taken when performing S-wave tomography with ray-based "banana doughnut" kernels. **Citation:** Mercerat, E. D., and G. Nolet (2012), Comparison of ray- and adjoint-based sensitivity kernels for body-wave seismic tomography, *Geophys. Res. Lett.*, *39*, L12301, doi:10.1029/2012GL052002.

1. Introduction

[2] In regional 3D tomographic studies, it is quite common to use just the onset times ('picks') of direct P and, when possible, direct S waves to image the Earth's upper mantle and crust. Complete seismograms are, in general, too complex to match all the individual wiggles in the coda of P and S direct waves. In recent years however, new tomographic techniques to image the Earth's upper mantle and crust have appeared using the adjoint method [*Chen et al.*, 2007a, 2007b; *Fichtner et al.*, 2009; *Tape et al.*, 2010]. These techniques do not require the identification of specific seismic phases, and therefore any portion of the seismogram can, in principle, be used in the inversion. On the other hand, adjoint techniques are much more computationally expensive than classical traveltime tomography, and still limited to low frequencies. A compromise is offered by ray-based finite-frequency kernels. These kernels take into account wavefront healing and are calculated via efficient dynamic ray tracing techniques where the adjoint field follows from reciprocity, but they are restricted to well-defined ray arrivals [*Dahlen et al.*, 2000; *Tian et al.*, 2007].

[3] Each seismic observable has its associated sensitivity kernel which relates it to the model parameters. Crosscorrelation based traveltime and amplitude measurements must be interpreted by means of the "banana-doughnut" sensitivity kernels originally derived by *Dahlen et al.* [2000]; *Dahlen and Baig* [2002] using a scattered wavefield computed with ray theory, or alternatively computed with a full waveform tool such as the Spectral Element Method (SEM) as discussed by *Tromp et al.* [2005]. From the comparison of surface wave kernels based on normal modes summation and adjoint techniques, the recent work of *Zhou et al.* [2011] shows that ray-based kernels calculated in 1D background velocity models are precise enough to carry out long period surface wave tomography at the global scale. Using seismograms computed by SEM [*Komatitsch and Vilotte*, 1998], we determine in this paper the accuracy of ray-based sensitivity kernels for computing body-wave traveltime delays at regional and local scales.

2. Linearized Inverse Problem

[4] The linearized seismic tomographic problem can be expressed as $\mathbf{Am} = \mathbf{d}$, where $\mathbf{m}$ is the vector of model perturbartions, $\mathbf{d}$ is the vector of observed crosscorrelation delay times δT and $\mathbf{A}$ is the sensitivity kernel matrix, in this case for P waves, defined by [*Dahlen et al.*, 2000; *Tromp et al.*, 2005]:

$$\delta T = -\frac{\int_{-\infty}^{\infty} \dot{u}(t')\delta u(t')\mathrm{d}t'}{\int_{-\infty}^{\infty} \ddot{u}(t')u(t')\mathrm{d}t'} = \int_V K_\alpha^P(\mathbf{r}_x)\frac{\delta\alpha}{\alpha}\mathbf{dr}_x + \int_V K_\beta^P(\mathbf{r}_x)\frac{\delta\beta}{\beta}\mathbf{dr}_x + \int_V K_\rho^P(\mathbf{r}_x)\frac{\delta\rho}{\rho}\mathbf{dr}_x, \quad (1)$$

where u denotes the unperturbed seismogram with the dot indicating the time derivative, δu the wave scattered by heterogeneities in P-wave velocity α, S-wave velocity β and density ρ, and $K_{[.]}^P$ are the finite-frequency sensitivity kernels of the P arrival. The integration volume V is the entire domain, but in practice, the integration is carried out only where the model perturbations affect the seismic observable δT (Fresnel volume). Though the kernels differ, the expression is the same for the delay of any arrival.

3. Sensitivity Kernel Calculations

3.1. Ray-Based "Banana-Doughnut" Kernels

[5] The ray-based computation of sensitivity kernels uses a ray-theoretical expression for u and δu. The frequency dependence enters through the scattering coefficients, computed with first-order (Born) scattering theory. The kernel computation in 3D models is done in two steps. First, we calculate and store the traveltime and geometrical spreading fields from all sources as well as the reciprocal (adjoint) field from all receivers to every node in the 3D grid, in order to make them available at the second step when the integrals

[1]Géoazur, Observatoire de la Côte d'Azur, Université de Nice-Sophia Antipolis, CNRS UMR 6526, Valbonne, France.
[2]Now at Laboratoire Régional des Ponts et Chausées, CETE Méditerranée, Nice, France.

Corresponding author: E. D. Mercerat, Laboratoire Régional des Ponts et Chausées, CETE Méditerranée, 56, Boulevard de Stalingrad, F-06359 Nice CEDEX, France. (diego.mercerat@geoazur.unice.fr)

are calculated. For the first step, we use a ray bending method with a first ray trajectory provided by the shortest path method [*Moser*, 1991; *Nolet*, 2008]. The kernel $K_\alpha^P(\mathbf{r}_x)$, which gives the P-wave delay due to an anomaly in P-wave velocity at a scatterer point $\mathbf{r}_x$, depends on the geometrical spreading and the traveltime through the ray-theoretical expressions for $u(t)$ and $\delta u(t)$, but also on the source radiation pattern and the directivity of the scattering (for details see *Nolet* [2008, equation (7.28)]).

[6] The first step is the most time consuming and computing time strongly depends on the complexity of the velocity model. To obtain a fast calculation of $K_\alpha^P(\mathbf{r}_x)$ [*Tian et al.*, 2007], we assume that: 1) the far-field approximation is valid, 2) the scattered wave is close to the unperturbed ray path so that its directivity can be ignored, and 3) that variations in the source radiation pattern can be neglected [*Dahlen et al.*, 2000]. We also ignore cross-dependent sensitivity (e.g., an S-wave velocity anomaly affecting the P-wave arrival time). Cross-dependance has been studied by *Zhang and Shen* [2008] but its true importance has not yet been established in real applications. The neglect of directivity in the scattering coefficient is exact for P waves scattering to P waves from a P-wave velocity anomaly, but in all other cases the symmetry is broken, even for S-to-S scattering [*Nolet*, 2008].

3.2. Adjoint Based Kernels

[7] The adjoint based computation of sensitivity kernels also consists of two wave propagation simulations: one simulation of the regular field caused by an earthquake, and one simulation of the adjoint wavefield caused by an adjoint source at the receiver location [*Tromp et al.*, 2005]. We use the SPECFEM3D package (http://geodynamics.org) [*Peter et al.*, 2011]. As in the case of ray-based kernels, the adjoint source to be injected at the receiver position corresponds to the velocity seismogram of the forward simulation reversed in time [*Liu and Tromp*, 2006]. Contrary to waveform tomography, delay-time kernels do not depend on the observed data.

[8] From a computational point of view, the method requires one forward calculation, and one backward calculation of the adjoint source placed at the receiver location. In fact, in order to convolve the forward and backward fields, simultaneous access to both fields is needed. The forward field could be stored at each time step in the whole model space to be accessed at each time step, but this is difficult in 3D with current storage and I/O capabilities. Alternatively, at the time of adjoint field simulation the reconstruction of the forward field can be done backwards in time from the last snapshot of the forward field (and the field stored just at the boundaries of the model for all times). In the SPECFEM3D package, this second option is chosen, and therefore for a single kernel calculation (one earthquake - one station), a total of three wavefield computations are needed, although two of them are done at the same clock time (doubling memory requirements and floating point operations).

4. Results

4.1. Homogeneous Medium

[9] An homogeneous medium of 200 km × 200 km × 50 km with P-wave velocity (α) = 6 km/s, S-wave velocity (β) = 3.47 km/s, and density (ρ) = 2500 kg/cm^3 is used for a first series of simulations. A horizontal point force in the X direction, placed at X = 50 km, Y = 50 km simultaneously generates a pure P wave which propagates in the X direction, and a pure SH wave propagating in the Y direction. We placed receivers at X = 150 km, Y = 50 km for the P-wave kernel K_α^P and at X = 50 km, Y = 150 km for the S-wave kernel K_β^S, i.e., the kernel that links the S-wave delay to anomalies in S-wave velocity.

[10] Comparisons between ray-based and adjoint SEM kernels are shown in Figure 1. They agree for all practical purposes for K_α^P, except for the expected differences close to the source and the receiver. For K_β^S the differences are slightly higher and they are found within the whole kernel volume. This is probably due to the neglect of S-wave scattering directivity, which should perhaps be abandoned at the cost of slightly more involved computations. Vertical cross-sections midway the ray-paths are plotted in Figure 2 to better quantify the differences. While they are negligible for K_α^P, they become noticeable for K_β^S (though less than 10 %). In particular, the central portion ("doughnut hole") of the adjoint based kernel is not exactly zero on the geometrical ray path as it is case in the ray-based kernel.

4.2. Smooth 3D Background Medium

[11] We now turn to kernel calculations in a smoothly varying 3D background medium. The focusing/defocusing caused by velocity anomalies will affect the amplitude (geometrical spreading of the incoming wavefield), and thus the sensitivity. *Gautier et al.* [2008] show that the nonlinearity associated with this is not serious, and that convergence within a few iterations of a tomographic inversion can be obtained even for strongly heterogeneous models.

[12] We use a 3D tomographic model of the central Chile subduction zone of *Deshayes* [2008]. The model measures 660 km × 660 km and reaches 220 km depth (Figure 3). We select an earthquake at 110 km depth, recorded by a station near 200 km epicentral distance. In order to limit as much as possible any factors that complicate the interpretation, we use an explosive source for the P-wave kernel and a double couple mechanism with vertical fault plane oriented in the azimuth of the station to generate the S-wave kernel. The source time function has a Gaussian shape with 0.25 Hz central frequency. For the adjoint sources, we use the vertical component and the transverse component of the velocity field recorded at the station for K_α^P and K_β^S, respectively. The results shown in Figure 3 are in accordance with the ones obtained in the homogeneous medium: the visual differences between ray-based and adjoint based sensitivity kernels to α velocity of the P-wave arrival remain small, while the sensitivity kernels to β velocity of the S-wave arrival differ slightly.

[13] In order to quantify the effect on delay times, we include several model perturbations and calculate, via numerical integration of the kernels, the expected delay times using each methodology. We place spherical anomalies of 5% in P- and S-wave velocities in the middle of the ray path between the earthquake and the station (in the center of the "doughnut hole", see Figure 3), and vary the size of the sphere in order to capture different aspects of the sensitivity kernels.

[14] Figure 4 shows the predicted delay times for different anomaly radii. For P-wave time delays, both approaches give similar results, despite the approximations in the ray-based "banana doughnuts" kernel calculations. The "doughnut hole" effect is clearly visible for a radius lower than 30 km, where wavefront healing causes the delay to remain close to

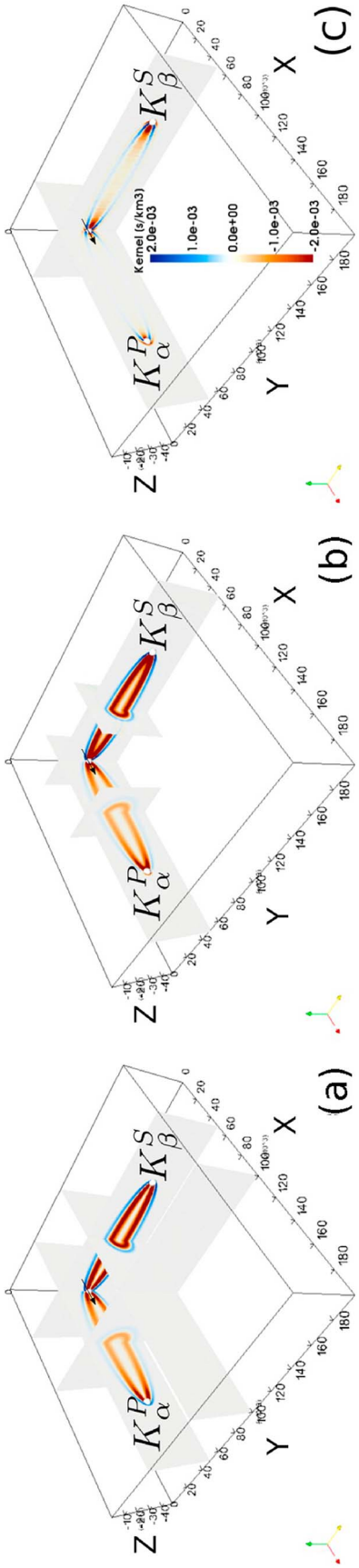

Figure 1. Traveltime kernels K_α^P and K_β^S calculated by (a) ray-based method and (b) adjoint based method (SEM). (c) Difference between Figures 1a and 1b. The same color scale is used in the three plots. Axes in km. The horizontal point force at X = 50 km and Y = 50 km is marked with a black arrow.

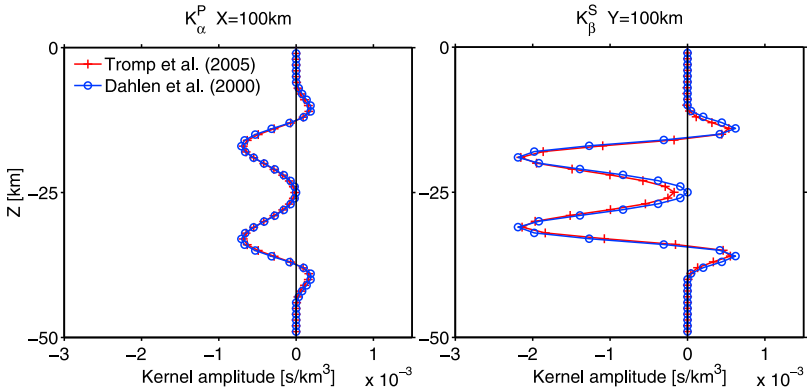

Figure 2. Vertical sections of traveltime kernels K^P_α at X = 100 km and K^S_β at Y = 100 km calculated by ray-based [*Dahlen et al.*, 2000] and adjoint SEM [*Tromp et al.*, 2005].

zero. The difference in the predicted delay time remains below 0.05 s for P waves. For the direct S-wave, time delays show differences higher than 0.1 s for anomalies with radius larger than 30 km, and, surprisingly, do not converge for a large radius, despite the fact the Fresnel zone for S waves is narrower than for P waves. All other approximations being present in both kernels, we conclude that this is a consequence of the directivity in the scattering of S waves.

5. Discussion and Conclusion

[15] Adjoint based as well as ray-based kernels can be computed in 3D media, and one can in principle iterate the linearized inverse problem until convergence is obtained [*Fichtner et al.*, 2009; *Tape et al.*, 2010], though the ray-based approach remains limited to smooth 3D background models.

[16] Computing kernels using ray-theoretical wavefields leads to savings in computation time of almost three orders of magnitude. For the 3D tomographic model interpolated in the hexahedral mesh of 1.07 10^6 elements of ~5 km side, one forward and one backward wavefield simulation (plus the forward reconstruction) with the SPECFEM3D solver takes almost 10 hours on a InfiniBand cluster of 64 AMD Opteron(TM) 1.1 GHz processors. Doing the same for the ray-based kernel calculation (one forward and one backward traveltime and geometrical spreading field computation) takes less than one hour on just one single processor for the 3D tomographic model sampled at voxels of 5 km side (~5 10^5 nodes).

[17] Yet the kernels are in good agreement in the most relevant portions of the kernel volume, and yield delay time predictions with an accuracy that may be sufficient when compared to the errors generally obtained in the observed delay times. Although there are significant differences in regions near the source and the receiver due to the far-field approximation in the ray-based approach, it is common practice in regional tomographic studies [*Tape et al.*, 2010] to apply smoothing operators to the kernels in order to remove large but integrable amplitudes in the vicinity of singularities at focal points, sources and receivers. Some care should be taken while inverting S-wave time delays, where the radiation pattern and the scattering coefficient effects in the kernel's calculation are non-negligible.

[18] One could pose the question: if one goes to the computational trouble of using a full wavefield modeling tool, such as SEM, why just use selected body wave phases and not revert to full waveform inversion and extract much more information? It is clear that the best approach to full waveform tomography, either in the time domain [*Chen et al.*, 2007a, 2007b; *Fichtner et al.*, 2009; *Tape et al.*, 2010], or in the frequency domain [*Pratt*, 1999; *Sirgue and Pratt*, 2004], is to calculate sensitivity kernels using wavefields with most of the physics taken into account (e.g., near and intermediate fields effects, diffractions, wave conversions, surface waves).The fact that the adjoint method for waveforms only computes the gradient of the misfit function, instead of the full kernel matrix, makes no fundamental difference: the gradient search that forms the basis of the 'adjoint method', is just a memory-efficient way to solve the linearized inverse problem [*Chen et al.*, 2007a, 2007b].

[19] But the use of delay times significantly reduces the data volume, even if delays are measured in several frequency bands to capture some of the 'waveform' information in the dispersion of body waves. Thus, the full matrix is of smaller dimension and can often be computed explicitly. Another advantage of using the full matrix in the tomographic inverse problem (rather than a gradient search) is that many regularization techniques and numerical solvers are available [*Loris et al.*, 2010;*Hung et al.*, 2011], and the Hessian matrix can be approximated efficiently. Most importantly, by using cross-correlation delay times rather than waveform mismatches, we avoid the non-linearity associated with waveforms caused by the fact that phase delays ΔT enters in the exponential expression $\exp(i\omega\Delta T)$ for the perturbed wavefield δu. In smoothly varying background media, the traveltime itself is accurately predicted by ray-theory. The Fréchet kernel is a derivative independent of the magnitude of $\delta\alpha$, and therefore linearity holds much better for the delays themselves than for the waveforms, where one has to use the linearization $\exp(i\omega\Delta T) \approx 1 + i\omega\Delta T$. Some of the waveform information -at least that in body waves- is captured by measuring crosscorrelation delays in different frequency bands (multi-frequency tomography [*Sigloch et al.*, 2008]).

[20] Not surprisingly then, delay times remain the data of choice for seismic tomography at the regional [*Chen et al.*, 2007a, 2007b; *Sigloch et al.*, 2008; *Hung et al.*, 2011; *Obrebski et al.*, 2011] as well as the global scale [*Obayashi et al.*, 2004; *Montelli et al.*, 2004, 2006; *Ritsema et al.*, 2009; *Zaroli et al.*, 2010].

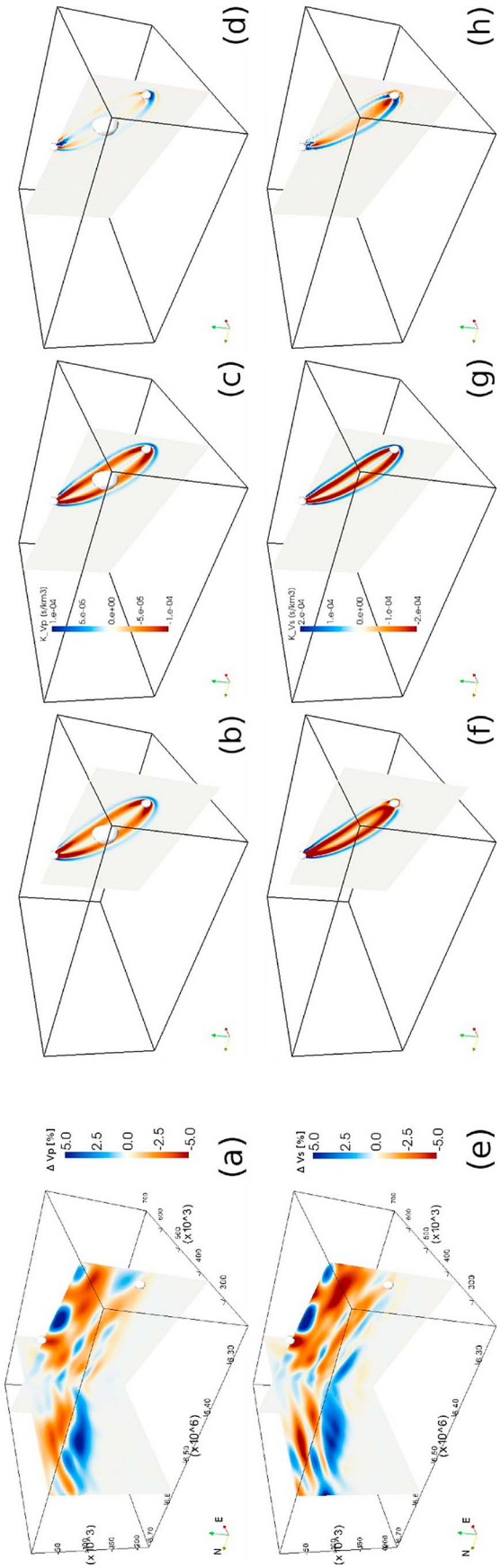

Figure 3. Ray-based and adjoint SEM kernels for P and S direct wave arrivals calculated in the 3D tomographic model of central Chile. (a) P-velocity and (e) S-velocity perturbations to the 1D model. (b) Adjoint based K_α^P kernel calculated in the 3D model. (c) Ray-based "banana doughnut" P-wave kernel for the same source-station pair. (d) Differences between Figures 3b and 3c using a common color scale. (f) Adjoint based K_β^S kernel calculated in the 3D model. (g) Ray-based "banana doughnut" S wave kernel for the same source-station pair. (h) Differences between Figures 3f and 3g using a common color scale. In the first row, the spherical anomaly of 25 km radius is plotted.

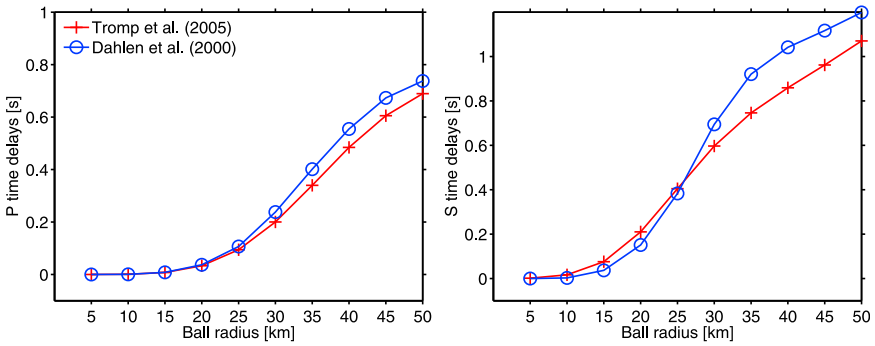

Figure 4. Traveltime delays for P and S waves predicted by ray-based kernel [*Dahlen et al.*, 2000] and adjoint kernel [*Tromp et al.*, 2005] for different anomaly radii.

[21] **Acknowledgments.** We thank Perrine Deshayes and Tony Monfret for providing the 3D tomographic model of central Chile. We also acknowledge the support from the European Research Council (ERC) Advanced grant 226837. This work has benefited from expeditious and thoughtful reviews by Jeroen Ritsema, Jeroen Tromp and the editor Michael Wysession.

[22] The Editor also thanks the reviewers for assisting in the evaluation of this paper.

References

Chen, P., L. Zhao, and T. H. Jordan (2007a), Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. Seismol. Soc. Am.*, *97*(4), 1094–1120.

Chen, P., T. Jordan, and L. Zhao (2007b), Full three-dimensional tomography: A comparison between the scattering-integral and adjoint-wavefield methods, *Geophys. J. Int.*, *170*, 175–181.

Dahlen, F., and A. Baig (2002), Fréchet kernels for body-wave amplitudes, *Geophys. J. Int.*, *150*, 440–466.

Dahlen, F., S.-H. Hung, and G. Nolet (2000), Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.*, *141*, 157–174.

Deshayes, P. (2008), Seismic velocity and attenuation tomography of the central Chili subduction zone—Western Argentine (29°S–34°S) from local earthquake data, PhD thesis, Univ. of Nice-Sophia Antipolis, Valbonne, France.

Fichtner, A., B. L. N. Kennett, H. Igel, and H.-P. Bunge (2009), Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, *179*(3), 1703–1725.

Gautier, S., G. Nolet, and J. Virieux (2008), Finite-frequency tomography in a crustal environment: Application to the western part of the Gulf of Corinth, *Geophys. Prospect.*, *56*, 493–503.

Hung, S.-H., W.-P. Chen, and L.-Y. Chiao (2011), A data-adaptive, multiscale approach of finite-frequency, traveltime tomography with special reference to P and S wave data from central Tibet, *J. Geophys. Res.*, *116*, B06307, doi:10.1029/2010JB008190.

Komatitsch, D., and J. P. Vilotte (1998), The spectral-element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. Seismol. Soc. Am.*, *88*(2), 368–392.

Liu, Q., and J. Tromp (2006), Finite-frequency kernels based on adjoint methods, *Bull. Seismol. Soc. Am.*, *96*, 2383–2397.

Loris, I., H. Douma, G. Nolet, I. Daubechies, and C. Regone (2010), Nonlinear regularization techniques for seismic tomography, *J. Comput. Phys.*, *229*(3), 890–905.

Montelli, R., G. Nolet, G. Masters, F. Dahlen, and S.-H. Hung (2004), Global P and PP traveltime tomography: Rays versus waves, *Geophys. J. Int.*, *158*, 637–654.

Montelli, R., G. Nolet, F. A. Dahlen, and G. Masters (2006), A catalogue of deep mantle plumes: New results from finite-frequency tomography, *Geochem. Geophys. Geosyst.*, *7*, Q11007, doi:10.1029/2006GC001248.

Moser, T. J. (1991), Shortest path calculations of seismic rays, *Geophysics*, *56*, 59–67.

Nolet, G. (2008), *A Breviary of Seismic Tomography*, 344 pp., Cambridge Univ. Press, Cambridge, U. K.

Obayashi, M., D. Suetsugu, and Y. Fukao (2004), PP-P differential traveltime measurement with crustal correction, *Geophys. J. Int.*, *157*, 1152–1162.

Obrebski, M., R. M. Allen, F. Pollitz, and S.-H. Hung (2011), Lithosphere-asthenosphere interaction beneath the western United States from the joint inversion of body-wave traveltimes and surface-wave phase velocities, *Geophys. J. Int.*, *185*(2), 1003–1021.

Peter, D., et al. (2011), Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.*, *186*(2), 721–739.

Pratt, R. (1999), Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physical scale model, *Geophysics*, *64*, 888–901.

Ritsema, J., H. van Heijst, J. Woodhouse, and A. Deuss (2009), Long-period body wave traveltimes through the crust: Implication for crustal corrections and seismic tomography, *Geophys. J. Int.*, *179*, 1255–1261.

Sigloch, K., N. McQuarrie, and G. Nolet (2008), Two-stage subduction history under North America inferred from multiple-frequency tomography, *Nat. Geosci.*, *1*, 458–462.

Sirgue, L., and R. G. Pratt (2004), Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics*, *69*(1), 231–248.

Tape, C., Q. Liu, A. Maggi, and J. Tromp (2010), Seismic tomography of the Southern California crust based upon spectral-element and adjoint methods, *Geophys. J. Int.*, *180*, 433–462.

Tian, Y., S.-H. Hung, G. Nolet, R. Montelli, and F. Dahlen (2007), Dynamic ray tracing and traveltime corrections for global seismic tomography, *J. Comput. Phys.*, *226*(1), 672–687.

Tromp, J., C. Tape, and Q. Liu (2005), Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, *160*, 195–216.

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*, 1025–1042.

Zhang, Z., and Y. Shen (2008), Cross-dependence of finite-frequency compressional waveforms to shear seismic wave speeds, *Geophys. J. Int.*, *174*, 941–948.

Zhou, Y., Q. Liu, and J. Tromp (2011), Surface wave sensitivity: Mode summation versus adjoint SEM, *Geophys. J. Int.*, *187*, 1560–1576.