Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods

Qinya Liu* and Jeroen Tromp†

Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA. E-mail: lqy@gps.caltech.edu

SUMMARY

We determine adjoint equations and Fréchet kernels for global seismic wave propagation based upon a Lagrange multiplier method. We start from the equations of motion for a rotating, self-gravitating earth model initially in hydrostatic equilibrium, and derive the corresponding adjoint equations that involve motions on an earth model that rotates in the opposite direction. Variations in the misfit function χ then may be expressed as $\delta \chi = \int_V K_m \, \delta \ln m \, d^3\mathbf{r} + \int_\Sigma K_d \, \delta \ln d \, d^2\mathbf{r} + \int_{\Sigma_{FS}} \mathbf{K}_d \cdot \nabla^\Sigma \delta \ln d \, d^2\mathbf{r}$, where $\delta \ln m = \delta m/m$ denotes relative model perturbations in the volume V, $\delta \ln d$ denotes relative topographic variations on solid–solid or fluid–solid boundaries Σ, and $\nabla^\Sigma \delta \ln d$ denotes surface gradients in relative topographic variations on fluid–solid boundaries Σ_{FS}. The 3-D Fréchet kernel K_m determines the sensitivity to model perturbations $\delta \ln m$, and the 2-D kernels K_d and $\mathbf{K}_d$ determine the sensitivity to topographic variations $\delta \ln d$. We demonstrate also how anelasticity may be incorporated within the framework of adjoint methods. Finite-frequency sensitivity kernels are calculated by simultaneously computing the adjoint wavefield forward in time and reconstructing the regular wavefield backward in time. Both the forward and adjoint simulations are based upon a spectral-element method. We apply the adjoint technique to generate finite-frequency traveltime kernels for global seismic phases (*P*, *P*$_{\text{diff}}$, *PKP*, *S*, *SKS*, depth phases, surface-reflected phases, surface waves, etc.) in both 1-D and 3-D earth models. For 1-D models these adjoint-generated kernels generally agree well with results obtained from ray-based methods. However, adjoint methods do not have the same theoretical limitations as ray-based methods, and can produce sensitivity kernels for any given phase in any 3-D earth model. The Fréchet kernels presented in this paper illustrate the sensitivity of seismic observations to structural parameters and topography on internal discontinuities. These kernels form the basis of future 3-D tomographic inversions.

Key words: Inverse theory; Seismic tomography; Computational seismology; Theoretical seismology; Wave propagation.

1 INTRODUCTION

Global seismic tomography is transitioning from ray-based methods to finite-frequency methods that recognize effects associated with wave front healing. These advances in seismic imaging involve 'finite-frequency' or 'Fréchet' sensitivity kernels which, for 1-D models, may be obtained in a number of different ways, including surface wave Green's functions (Marquering *et al.* 1999), body-wave ray theory (the so-called 'banana–doughnut' kernels Dahlen *et al.* 2000; Hung *et al.* 2000), surface wave ray theory (Zhou *et al.* 2004), and normal-model theory (Zhao *et al.* 2000; Zhao & Jordan 2006). The first tomographic images based upon these new finite-frequency techniques have recently appeared (e.g. Montelli *et al.* 2004; Zhou *et al.* 2005).

Most finite-frequency techniques are based upon 1-D reference models, and the next important challenge is to develop finite-frequency techniques that involve fully 3-D starting models and corresponding 3-D Fréchet sensitivity kernels (e.g. Zhao *et al.* 2005). This goal must involve 3-D simulations of seismic wave propagation in 3-D global models, something that has finally become feasible in recent years

*Now at: Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093, USA.
†Now at: Department of Geosciences, Princeton University, Princeton, NJ 08540, USA.

265

(e.g. Komatitsch & Tromp 2002a,b; Chaljub *et al.* 2003). The challenge now is to harness these 3-D simulation capabilities to address the inverse problem. With this challenge in mind, Tromp *et al.* (2005) and Tape *et al.* (2007) deployed an adjoint formulation for the calculation of the Fréchet kernels, which involves the interaction between the regular forward wavefield and an adjoint wavefield produced by using time-reversed signal at the receivers as simultaneous, fictitious sources. Adjoint methods have been popular in the fields of meteorology and exploration geophysics for many years (e.g. Chavent *et al.* 1975; Tarantola 1984; Gauthier *et al.* 1986; Mora 1987, 1988; Talagrand & Courtier 1987; Luo & Schuster 1991).

The applicability of the adjoint method to the calculation of 3-D Fréchet sensitivity kernels for regional 3-D earth models was demonstrated in Liu & Tromp (2006). In this article we derive the complete set of adjoint equations and Fréchet kernels for global seismic wave propagation based upon a Lagrange multiplier method. We apply the adjoint method to compute 3-D traveltime sensitivity kernels for typical global seismic phases in both 1-D and 3-D global models, for example, *P, S, ScS, SS, sP*, etc.

From a computational point of view, one of the main advantages of the adjoint method is that for each earthquake only two numerical simulations are needed to compute the gradient of a misfit function. For a full 3-D tomographic inversion, an earthquake-specific 'event kernel' is computed simultaneously for all traveltime measurements made at all three-component stations. The sum of all event kernels, called the 'misfit' kernel, then represents the gradient of the misfit function. Because it is computationally costly to calculate the Hessian of the misfit function, a non-linear conjugate gradient algorithm is used to minimize the misfit function, thereby improving the earth model (Tape *et al.* 2007).

2 EQUATIONS OF MOTION

We begin by introducing the equations of motion for a rotating, self-gravitating earth model that is initially in hydrostatic equilibrium. In such an earth model, pressure gradients are balanced by gradients in the geopotential, which is the sum of the gravitational and centrifugal potentials. As a consequence, the equilibrium model is an oblate ellipsoid in which surfaces of constant pressure, density and the geopotential coincide. The results in this section are well known (e.g. Dahlen & Tromp 1998), but serve the purpose of introducing the necessary framework and the related notation.

The gravitational potential Φ is determined by Poisson's equation:

$$\nabla^2 \Phi = 4\pi G\rho, \tag{1}$$

where ρ denotes the density distribution and G the gravitational constant. For mathematical and notational convenience, the density distribution is defined to be zero outside of the earth model, where the gravitational potential satisfies Laplace's equation: $\nabla^2 \Phi = 0$. The earth model contains internal solid–solid discontinuities, collectively denoted by Σ_{SS}, and internal fluid–solid discontinuities, collectively denoted by Σ_{FS}. The collection of all internal solid–solid and fluid–solid discontinuities plus the surface of the model volume ∂V is denoted by $\Sigma = \partial V \cup \Sigma_{SS} \cup \Sigma_{FS}$. Poisson's eq. (1) must be solved subject to the boundary conditions

$$[\Phi]_-^+ = 0, \quad [\hat{\mathbf{n}} \cdot \nabla\Phi]_-^+ = 0 \quad \text{on } \Sigma, \tag{2}$$

where the notation $[\cdot]_-^+$ denotes the jump in the enclosed quantity when going from the outward (+) side to the inward (−) side of the discontinuity. The unit normal $\hat{\mathbf{n}}$ points outward from the '−' side to the '+' side of a discontinuity. The solution to the boundary value problem (1)–(2) is

$$\Phi(\mathbf{r}) = -G \int_V \frac{\rho(\mathbf{r}')}{||\mathbf{r} - \mathbf{r}'||} \, d^3\mathbf{r}', \tag{3}$$

where V denotes the volume of the earth model and $\mathbf{r}$ the position vector.

The centrifugal potential ψ is defined in terms of the Earth's angular rotation vector $\mathbf{\Omega}$ by

$$\psi = -\frac{1}{2}[\Omega^2 r^2 - (\mathbf{\Omega} \cdot \mathbf{r})^2], \tag{4}$$

where $r = |\mathbf{r}|$ denotes the radius and $\Omega = |\mathbf{\Omega}|$ is the angular rotation rate.

The displacement field $\mathbf{s}$ in a rotating, self-gravitating earth model that is initially in hydrostatic equilibrium is governed by the equation of motion (Dahlen & Tromp 1998, eq. 3.259)

$$\rho(\partial_t^2 \mathbf{s} + 2\mathbf{\Omega} \times \partial_t \mathbf{s}) - \nabla \cdot \mathbf{T} + \nabla[\rho \, \mathbf{s} \cdot \nabla(\Phi + \psi)] + \rho \nabla\phi - \nabla \cdot (\rho \, \mathbf{s})\nabla(\Phi + \psi) = \mathbf{f}. \tag{5}$$

In an anisotropic, elastic earth model the stress $\mathbf{T}$ is defined in terms of the displacement gradient by

$$\mathbf{T} = \mathbf{c} : \nabla \mathbf{s}, \tag{6}$$

where $\mathbf{c}$ denotes the elastic tensor. The complications associated with an anelastic earth model will be addressed in Section 4. On all boundaries the traction $\hat{\mathbf{n}} \cdot \mathbf{T}$ needs to be continuous:

$$[\hat{\mathbf{n}} \cdot \mathbf{T}]_-^+ = \mathbf{0} \quad \text{on } \Sigma. \tag{7}$$

For mathematical convenience, the earth model parameters ρ and $\mathbf{c}$ are defined to be zero outside of the model volume V, such that at the free surface ∂V the boundary condition (7) reduces to the vanishing of the traction: $\hat{\mathbf{n}} \cdot \mathbf{T} = \mathbf{0}$. On a fluid–solid boundary the traction is normal: $\hat{\mathbf{n}} \cdot \mathbf{T} = (\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. On all solid–solid boundaries the displacement $\mathbf{s}$ must be continuous

$$[\mathbf{s}]_-^+ = \mathbf{0} \quad \text{on } \Sigma_{SS}, \tag{8}$$

whereas on fluid–solid discontinuities there may be slip:

$$[\hat{\mathbf{n}} \cdot \mathbf{s}]_-^+ = 0 \quad \text{on } \Sigma_{\text{FS}}. \tag{9}$$

The displacement is subject to the initial conditions

$$\mathbf{s}(\mathbf{r}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}(\mathbf{r}, 0) = \mathbf{0}. \tag{10}$$

The perturbation in the gravitational potential ϕ induced by the displacement $\mathbf{s}$ is determined by

$$\nabla \cdot \boldsymbol{\xi} = 0, \tag{11}$$

$$[\phi]_-^+ = 0, \quad [\hat{\mathbf{n}} \cdot \boldsymbol{\xi}]_-^+ = 0 \quad \text{on } \Sigma, \tag{12}$$

where

$$\boldsymbol{\xi} = (4\pi G)^{-1} \nabla \phi + \rho \mathbf{s}. \tag{13}$$

The solution to the boundary-value problem (11)–(12) is

$$\phi(\mathbf{r}, t) = -G \int_V \rho(\mathbf{r}') \frac{\mathbf{s}(\mathbf{r}', t) \cdot (\mathbf{r} - \mathbf{r}')}{||\mathbf{r} - \mathbf{r}'||^3} \, d^3\mathbf{r}'. \tag{14}$$

3 ADJOINT EQUATIONS

In this section, we derive the adjoint equations which arise in the context of tomographic inverse problems. Our objective will be to minimize the difference between observed three-component seismograms and the corresponding synthetic seismograms. This may be accomplished based upon a wide variety of misfit criteria, for example, waveform or cross-correlation traveltime differences. In this section and in Section 4, we consider the minimization of a waveform misfit function χ subject to the constraint that the synthetic seismograms are determined by the equations of motion (5)–(12):

$$\begin{aligned} \chi = &\frac{1}{2} \sum_r \int_0^T ||\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)||^2 \, dt \\ &- \int_0^T \int_V \boldsymbol{\lambda} \cdot \left\{ \rho\left(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}\right) - \nabla \cdot \mathbf{T} + \nabla[\rho \, \mathbf{s} \cdot \nabla(\Phi + \psi)] + \rho \nabla \phi - \nabla \cdot (\rho \, \mathbf{s})\nabla(\Phi + \psi) - \mathbf{f} \right\} d^3\mathbf{r} \, dt \\ &+ \int_0^T \int_\bigcirc \mu(\nabla \cdot \boldsymbol{\xi}) \, d^3\mathbf{r} \, dt. \end{aligned} \tag{15}$$

Here $\bigcirc$ denotes all of space, and the vector Lagrange multiplier $\boldsymbol{\lambda}(\mathbf{r}, t)$ and the scalar Lagrange multiplier $\mu(\mathbf{r}, t)$ remain to be determined. The interval $[0, T]$ denotes the time-series of interest, $\mathbf{s}(\mathbf{r}_r, t)$ denotes the synthetic displacement field at receiver location $\mathbf{r}_r$, and $\mathbf{d}(\mathbf{r}_r, t)$ denotes the corresponding observed three-component displacement vector. In practice, both the data $\mathbf{d}$ and the synthetics $\mathbf{s}$ will be windowed, filtered, and possibly weighted on the time interval $[0, T]$. In what follows we will implicitly assume that such operations have been performed. We note that different measures of misfit simply give rise to different adjoint sources (Tromp *et al.* 2005).

Upon taking the variation of the action (15) we obtain

$$\begin{aligned} \delta\chi = &\int_0^T \int_V \sum_r [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)] \delta(\mathbf{r} - \mathbf{r}_r) \cdot \delta\mathbf{s}(\mathbf{r}, t) \, d^3\mathbf{r} \, dt \\ &- \int_0^T \int_V \boldsymbol{\lambda} \cdot \left\{ \rho\left(\partial_t^2 \delta\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \delta\mathbf{s}\right) + \delta\rho\left(\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}\right) - \nabla \cdot \delta\mathbf{T} + \nabla[\rho \, \delta\mathbf{s} \cdot \nabla(\Phi + \psi)] + \nabla[\delta\rho \, \mathbf{s} \cdot \nabla(\Phi + \psi)] \right. \\ &\left. + \nabla[\rho \, \mathbf{s} \cdot \nabla(\delta\Phi)] + \rho\nabla\delta\phi + \delta\rho\nabla\phi - \nabla \cdot (\rho \, \delta\mathbf{s})\nabla(\Phi + \psi) - \nabla \cdot (\delta\rho \, \mathbf{s})\nabla(\Phi + \psi) - \nabla \cdot (\rho \, \mathbf{s})\nabla(\delta\Phi) - \delta\mathbf{f} \right\} d^3\mathbf{r} \, dt \\ &+ \int_0^T \int_\bigcirc \mu(\nabla \cdot \delta\boldsymbol{\xi}) \, d^3\mathbf{r} \, dt. \end{aligned} \tag{16}$$

The perturbation in stress $\delta\mathbf{T}$ may be obtained by perturbing Hooke's law (6):

$$\delta\mathbf{T} = \mathbf{c} : \nabla\delta\mathbf{s} + \delta\mathbf{c} : \nabla\mathbf{s}, \tag{17}$$

and the perturbation $\delta\boldsymbol{\xi}$ is obtained by perturbing (13):

$$\delta\boldsymbol{\xi} = (4\pi G)^{-1} \nabla\delta\phi + \rho\delta\mathbf{s} + \delta\rho \, \mathbf{s}. \tag{18}$$

The perturbation in the gravitational potential $\delta\Phi$ associated with a perturbation in density $\delta\rho$ and a displacement δd of the discontinuities Σ is determined by (Dahlen & Tromp 1998)

$$\nabla^2(\delta\Phi) = 4\pi G\delta\rho, \tag{19}$$

$$[\delta\Phi]_-^+ = 0, \quad [\hat{\mathbf{n}} \cdot \nabla(\delta\Phi) + 4\pi G\rho\delta d]_-^+ = 0 \quad \text{on } \Sigma, \tag{20}$$

where the density perturbation $\delta\rho$ is defined to be zero outside of the volume of the earth model V. The solution to the boundary-value problem (19)–(20) is

$$\delta\Phi(\mathbf{r}) = -G \int_V \frac{\delta\rho(\mathbf{r}')}{||\mathbf{r}-\mathbf{r}'||}\, d^3\mathbf{r}' + G \int_\Sigma \frac{\delta d(\mathbf{r}')[\rho(\mathbf{r}')]^+_-}{||\mathbf{r}-\mathbf{r}'||}\, d^2\mathbf{r}', \tag{21}$$

that is, the perturbation in the gravitational potential $\delta\Phi$ is completely determined by the perturbations in density $\delta\rho$ and boundary topography δd.

The next step in determining the adjoint equations involves manipulating the variation in the action (16) by performing integrations by parts over time, and repeatedly using Gauss' theorem for a volume with discontinuities:

$$\int_V \nabla \cdot \mathbf{v}\, d^3\mathbf{r} = -\int_\Sigma [\hat{\mathbf{n}} \cdot \mathbf{v}]^+_-\, d^2\mathbf{r}, \tag{22}$$

for any tensor $\mathbf{v}$. In an earth model without any internal discontinuities, or if $\hat{\mathbf{n}} \cdot \mathbf{v}$ is continuous on all internal discontinuities, this reduces to the familiar form

$$\int_V \nabla \cdot \mathbf{v}\, d^3\mathbf{r} = \int_{\partial V} \hat{\mathbf{n}} \cdot \mathbf{v}\, d^2\mathbf{r}. \tag{23}$$

After a significant amount of tedious algebra we can rewrite the variation of the action (16) in the form

$$\begin{aligned}
\delta\chi = & \int_0^T \int_V \sum_r [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)]\, \delta(\mathbf{r}-\mathbf{r}_r) \cdot \delta\mathbf{s}(\mathbf{r}, t)\, d^3\mathbf{r}\, dt \\
& - \int_0^T \int_V \delta\mathbf{s} \cdot \left\{ \rho(\partial_t^2\boldsymbol{\lambda} + 2\boldsymbol{\Omega} \times \partial_t\boldsymbol{\lambda}) - \nabla \cdot \boldsymbol{\sigma} + \nabla[\rho\,\boldsymbol{\lambda} \cdot \nabla(\Phi+\psi)] + \rho\nabla\mu - \nabla \cdot (\rho\,\boldsymbol{\lambda})\nabla(\Phi+\psi) \right\} d^3\mathbf{r}\, dt \\
& + \int_0^T \int_\bigcirc \delta\phi(\nabla \cdot \boldsymbol{\zeta})\, d^3\mathbf{r}\, dt \\
& - \int_V \rho\left[\boldsymbol{\lambda} \cdot \partial_t\delta\mathbf{s} - \partial_t\boldsymbol{\lambda} \cdot \delta\mathbf{s} + 2\boldsymbol{\lambda} \cdot (\boldsymbol{\Omega} \times \delta\mathbf{s})\right]_0^T d^3\mathbf{r} \\
& - \int_0^T \int_V \left\{ \delta\rho\left[\boldsymbol{\lambda} \cdot (\partial_t^2\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t\mathbf{s}) + \boldsymbol{\lambda} \cdot \nabla\phi + \mathbf{s} \cdot \nabla\mu + \mathbf{s} \cdot \nabla\nabla(\Phi+\psi) \cdot \boldsymbol{\lambda} + \nabla(\Phi+\psi) \cdot (\mathbf{s} \cdot \nabla\boldsymbol{\lambda} - \mathbf{s}\nabla \cdot \boldsymbol{\lambda})\right] \right. \\
& \left. + \nabla\boldsymbol{\lambda} : \delta\mathbf{c} : \nabla\mathbf{s} + \rho\mathbf{s} \cdot \nabla\nabla(\delta\Phi) \cdot \boldsymbol{\lambda} + \rho\nabla(\delta\Phi) \cdot (\mathbf{s} \cdot \nabla\boldsymbol{\lambda} - \mathbf{s}\nabla \cdot \boldsymbol{\lambda}) - \boldsymbol{\lambda} \cdot \delta\mathbf{f} \right\} d^3\mathbf{r}\, dt \\
& - \int_0^T \int_\Sigma [\hat{\mathbf{n}} \cdot \delta\mathbf{T} \cdot \boldsymbol{\lambda} - \hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \cdot \delta\mathbf{s} + \hat{\mathbf{n}} \cdot \delta\boldsymbol{\xi}\,\mu - \hat{\mathbf{n}} \cdot \boldsymbol{\zeta}\,\delta\phi]^+_-\, d^2\mathbf{r}\, dt,
\end{aligned} \tag{24}$$

where we have introduced the notation $[f]_0^T = f(T) - f(0)$, for any function f, and where we have defined

$$\boldsymbol{\sigma} = \mathbf{c} : \nabla\boldsymbol{\lambda}, \tag{25}$$

and

$$\boldsymbol{\zeta} = (4\pi G)^{-1}\nabla\mu + \rho\boldsymbol{\lambda}. \tag{26}$$

In the absence of model parameter perturbations $\delta\rho$, $\delta\mathbf{c}$, δd and $\delta\mathbf{f}$, the variation in the action (24) is stationary with respect to permissible (i.e. honouring the appropriate boundary and initial conditions) perturbations $\delta\mathbf{s}$ and $\delta\phi$ provided the vector Lagrange multiplier $\boldsymbol{\lambda}$ satisfies the equation

$$\rho(\partial_t^2\boldsymbol{\lambda} + 2\boldsymbol{\Omega} \times \partial_t\boldsymbol{\lambda}) - \nabla \cdot \boldsymbol{\sigma} + \nabla[\rho\,\boldsymbol{\lambda} \cdot \nabla(\Phi+\psi)] + \rho\nabla\mu - \nabla \cdot (\rho\,\boldsymbol{\lambda})\nabla(\Phi+\psi) = \sum_r [\mathbf{s}(\mathbf{r}_r, t) - \mathbf{d}(\mathbf{r}_r, t)]\,\delta(\mathbf{r}-\mathbf{r}_r), \tag{27}$$

subject to the boundary conditions

$$[\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}]^+_- = \mathbf{0} \quad \text{on } \Sigma, \tag{28}$$

$$[\boldsymbol{\lambda}]^+_- = \mathbf{0} \quad \text{on } \Sigma_{SS}, \tag{29}$$

$$[\hat{\mathbf{n}} \cdot \boldsymbol{\lambda}]^+_- = 0 \quad \text{on } \Sigma_{FS}, \tag{30}$$

and the end conditions

$$\boldsymbol{\lambda}(\mathbf{r}, T) = \mathbf{0}, \quad \partial_t\boldsymbol{\lambda}(\mathbf{r}, T) = \mathbf{0}. \tag{31}$$

On fluid–solid boundaries the quantity $\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}$ is normal: $\hat{\mathbf{n}} \cdot \boldsymbol{\sigma} = (\hat{\mathbf{n}} \cdot \boldsymbol{\sigma} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. The scalar Lagrange multiplier μ is determined by

$$\nabla \cdot \boldsymbol{\zeta} = 0, \tag{32}$$

subject to the boundary conditions

$$[\mu]^+_- = 0, \quad [\hat{\mathbf{n}} \cdot \boldsymbol{\zeta}]^+_- = 0 \quad \text{on } \Sigma. \tag{33}$$

At this point let us define the 'adjoint wavefield' $\mathbf{s}^\dagger$ in terms of the vector Lagrange multiplier wavefield $\boldsymbol{\lambda}$ by

$$\mathbf{s}^\dagger(\mathbf{r}, t) \equiv \boldsymbol{\lambda}(\mathbf{r}, T-t), \tag{34}$$

that is, the adjoint wavefield $\mathbf{s}^\dagger$ is the time-reversed Lagrange multiplier wavefield $\boldsymbol{\lambda}$. Similarly, we define the 'adjoint perturbed gravitational potential' $\phi^\dagger$ in terms of the scalar Lagrange multiplier μ by

$$\phi^\dagger(\mathbf{r}, t) \equiv \mu(\mathbf{r}, T - t). \tag{35}$$

Then the adjoint wavefield $\mathbf{s}^\dagger$ is determined by the set of equations

$$\rho\big(\partial_t^2 \mathbf{s}^\dagger - 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}^\dagger\big) - \nabla \cdot \mathbf{T}^\dagger + \nabla[\rho\, \mathbf{s}^\dagger \cdot \nabla(\Phi + \psi)] + \rho\nabla\phi^\dagger - \nabla \cdot (\rho\, \mathbf{s}^\dagger)\nabla(\Phi + \psi) = \sum_r [\mathbf{s}(\mathbf{r}_r, T - t) - \mathbf{d}(\mathbf{r}_r, T - t)]\, \delta(\mathbf{r} - \mathbf{r}_r), \tag{36}$$

where the adjoint stress is determined in terms of the adjoint displacement gradient by

$$\mathbf{T}^\dagger = \mathbf{c} : \nabla \mathbf{s}^\dagger. \tag{37}$$

On all boundaries the adjoint traction $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger$ needs to be continuous

$$[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger]_-^+ = \mathbf{0} \quad \text{on } \Sigma, \tag{38}$$

and on fluid–solid boundaries the adjoint traction is normal: $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger = (\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. On all solid–solid boundaries the adjoint displacement $\mathbf{s}^\dagger$ must be continuous

$$[\mathbf{s}^\dagger]_-^+ = \mathbf{0} \quad \text{on } \Sigma_{\mathrm{SS}}, \tag{39}$$

whereas on fluid–solid discontinuities

$$[\hat{\mathbf{n}} \cdot \mathbf{s}^\dagger]_-^+ = 0 \quad \text{on } \Sigma_{\mathrm{FS}}. \tag{40}$$

The adjoint displacement is subject to the initial conditions

$$\mathbf{s}^\dagger(\mathbf{r}, 0) = \mathbf{0}, \quad \partial_t \mathbf{s}^\dagger(\mathbf{r}, 0) = \mathbf{0}. \tag{41}$$

The adjoint perturbation in the gravitational potential $\phi^\dagger$ induced by the adjoint displacement $\mathbf{s}^\dagger$ is determined by the boundary-value problem

$$\nabla \cdot \boldsymbol{\xi}^\dagger = 0, \tag{42}$$

$$[\phi^\dagger]_-^+ = 0, \quad [\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger]_-^+ = 0 \quad \text{on } \Sigma, \tag{43}$$

where

$$\boldsymbol{\xi}^\dagger = (4\pi G)^{-1}\nabla\phi^\dagger + \rho \mathbf{s}^\dagger. \tag{44}$$

Upon comparing the equations for the regular wavefield (5)–(13) with those for the adjoint wavefield (36)–(44) we see that they are identical with two exceptions: first, the source term for the adjoint wave eq. (36) is $\sum_r[\mathbf{s}(\mathbf{r}_r, T-t) - \mathbf{d}(\mathbf{r}_r, T-t)]\delta(\mathbf{r} - \mathbf{r}_r)$ and involves the time-reversed differences between the synthetics and the data at all receivers, and, second, the adjoint wavefield must be calculated in an earth model with the opposite sense of rotation, as reflected in the change in the sign of the Coriolis term in (36). Thus the adjoint wavefield may be generated with the same software that generates the regular wavefield by reversing the sense of rotation of the earth model and by using a source that consists of the time-reversed differences between all available data and the corresponding synthetics. We note again that a different choice of misfit criterion, for example, based upon cross-correlation traveltime or multitaper phase and amplitude anomalies, simply leads to a different form of the adjoint source.

With the regular and adjoint wavefields now defined, the change in the misfit function (24) reduces to

$$\delta\chi = -\int_0^T \int_V \Big\{ \delta\rho\big[\mathbf{s}^\dagger \cdot (\partial_t^2 \mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t \mathbf{s}) + \mathbf{s}^\dagger \cdot \nabla\phi + \mathbf{s} \cdot \nabla\phi^\dagger + \mathbf{s} \cdot \nabla\nabla(\Phi + \psi) \cdot \mathbf{s}^\dagger + \nabla(\Phi + \psi) \cdot (\mathbf{s} \cdot \nabla \mathbf{s}^\dagger - \mathbf{s}\nabla \cdot \mathbf{s}^\dagger)\big]$$

$$+ \nabla \mathbf{s}^\dagger : \delta\mathbf{c} : \nabla \mathbf{s} + \rho \mathbf{s} \cdot \nabla\nabla(\delta\Phi) \cdot \mathbf{s}^\dagger + \rho\nabla(\delta\Phi) \cdot (\mathbf{s} \cdot \nabla \mathbf{s}^\dagger - \mathbf{s}\nabla \cdot \mathbf{s}^\dagger) - \mathbf{s}^\dagger \cdot \delta\mathbf{f} \Big\}\, d^3\mathbf{r}\, dt$$

$$- \int_0^T \int_\Sigma [\hat{\mathbf{n}} \cdot \delta\mathbf{T} \cdot \mathbf{s}^\dagger - \hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \delta\mathbf{s} + \hat{\mathbf{n}} \cdot \delta\boldsymbol{\xi}\, \phi^\dagger - \hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger\, \delta\phi]_-^+\, d^2\mathbf{r}\, dt, \tag{45}$$

where it is implied that when the regular wavefield $\mathbf{s}$ is evaluated at time t, the adjoint wavefield is evaluated at time $T - t$.

Because we are allowing perturbations δd in the location of all discontinuities Σ, we need to carefully consider the boundary integral in (45), because such perturbations render the variations $\delta\mathbf{s}$ and $\delta\phi$ inadmissible. Because the adjoint traction $\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger$ is continuous on Σ and normal on Σ_{FS} we have (Dahlen & Tromp 1998, eq. 13.35)

$$[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \delta\mathbf{s}]_-^+ = -\delta d[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \partial_n \mathbf{s}]_-^+ + \nabla^\Sigma(\delta d) \cdot [(\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}})\mathbf{s}]_-^+. \tag{46}$$

Here $\partial_n = \hat{\mathbf{n}} \cdot \nabla$ denotes the partial derivative in the direction of the outward normal $\hat{\mathbf{n}}$, and $\nabla^\Sigma = (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \nabla$ denotes the surface gradient on Σ, where $\mathbf{I}$ is the identity tensor. Thus we may express the gradient operator as $\nabla = \hat{\mathbf{n}}\partial_n + \nabla^\Sigma$. Note that on a solid–solid boundary both $\mathbf{s}$ and $\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}}$ are continuous, which means that the last term in (46) contributes only on a fluid–solid boundary. Next, using the fact that $[\phi^\dagger]_-^+ = 0$ (and Dahlen & Tromp 1998, eq. 13.62) we have

$$[\hat{\mathbf{n}} \cdot \delta\boldsymbol{\xi}\, \phi^\dagger]_-^+ = -\delta d[\phi^\dagger \hat{\mathbf{n}} \cdot \partial_n \boldsymbol{\xi}]_-^+ + \nabla^\Sigma(\delta d) \cdot [\phi^\dagger \boldsymbol{\xi}]_-^+ = \nabla^\Sigma \cdot [\delta d\, \phi^\dagger \boldsymbol{\xi}]_-^+ - \delta d[\boldsymbol{\xi} \cdot \nabla\phi^\dagger]_-^+ + \delta d\big[(\hat{\mathbf{n}} \cdot \boldsymbol{\xi})\partial_n\phi^\dagger\big]_-^+, \tag{47}$$

where we have used (11). Because $\nabla^\Sigma(\delta d)$ is a tangent vector to Σ, that is, $\hat{\mathbf{n}} \cdot \nabla^\Sigma(\delta d) = 0$, any vector dotted with $\nabla^\Sigma(\delta d)$ retains only its tangent components. Thus the integration by parts associated with the second equality involves only the tangent components of $\boldsymbol{\xi}$; this

becomes important when we use the 2-D version of Gauss' theorem in the surface integral in (45). Using the continuity $[\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger]_-^+$ (and Dahlen & Tromp 1998, eq. 13.36) we have

$$[\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger \, \delta\phi]_-^+ = -\delta d [\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger \partial_n \phi]_-^+. \tag{48}$$

Finally, on solid–solid boundaries $[\mathbf{s}^\dagger]_-^+ = \mathbf{0}$, which implies that (using Dahlen & Tromp 1998, eq. 13.61)

$$[\hat{\mathbf{n}} \cdot \delta\mathbf{T} \cdot \mathbf{s}^\dagger]_-^+ = -\delta d [\hat{\mathbf{n}} \cdot \partial_n \mathbf{T} \cdot \mathbf{s}^\dagger]_-^+ + \nabla^\Sigma(\delta d) \cdot [\mathbf{T} \cdot \mathbf{s}^\dagger]_-^+ = -\delta d [(\nabla \cdot \mathbf{T}) \cdot \mathbf{s}^\dagger]_-^+ + \nabla^\Sigma \cdot [\delta d \, \mathbf{T} \cdot \mathbf{s}^\dagger]_-^+ - \delta d [\mathbf{T}:\nabla\mathbf{s}^\dagger]_-^+ + \delta d [\mathbf{T}:\hat{\mathbf{n}}\partial_n\mathbf{s}^\dagger]_-^+, \tag{49}$$

where we note, again in anticipation of using the 2-D version of Gauss' theorem in (45), that in the second equality only the tangent part of $\mathbf{T} \cdot \mathbf{s}^\dagger$ is involved in the horizontal divergence. It may be shown that on a fluid–solid boundary (49) acquires one extra term:

$$[\hat{\mathbf{n}} \cdot \delta\mathbf{T} \cdot \mathbf{s}^\dagger]_-^+ = -\delta d [(\nabla \cdot \mathbf{T}) \cdot \mathbf{s}^\dagger]_-^+ + \nabla^\Sigma \cdot [\delta d \, \mathbf{T} \cdot \mathbf{s}^\dagger]_-^+ - \delta d [\mathbf{T}:\nabla\mathbf{s}^\dagger]_-^+ + \delta d [\mathbf{T}:\hat{\mathbf{n}}\partial_n\mathbf{s}^\dagger]_-^+ - \nabla^\Sigma(\delta d) \cdot [(\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}})\mathbf{s}^\dagger]_-^+. \tag{50}$$

Thus we find that we may rewrite the surface integral in (45) in the form

$$\int_0^T \int_\Sigma [\hat{\mathbf{n}} \cdot \delta\mathbf{T} \cdot \mathbf{s}^\dagger - \hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \delta\mathbf{s} + \hat{\mathbf{n}} \cdot \delta\boldsymbol{\xi}\, \phi^\dagger - \hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger \, \delta\phi]_-^+ \, d^2\mathbf{r} \, dt$$

$$= \int_0^T \int_\Sigma \delta d [\rho\{-(\partial_t^2\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t\mathbf{s}) \cdot \mathbf{s}^\dagger - \mathbf{s}^\dagger \cdot \nabla\phi - \mathbf{s} \cdot \nabla\phi^\dagger - \mathbf{s} \cdot \nabla\nabla(\Phi + \psi) \cdot \mathbf{s}^\dagger - \nabla(\Phi + \psi) \cdot (\mathbf{s}^\dagger \cdot \nabla\mathbf{s} - \mathbf{s}^\dagger\nabla \cdot \mathbf{s})\}$$

$$- \nabla\mathbf{s}^\dagger :\mathbf{c}:\nabla\mathbf{s} - (4\pi G)^{-1}(\nabla\phi \cdot \nabla\phi^\dagger) + (\hat{\mathbf{n}} \cdot \mathbf{T}) \cdot \partial_n\mathbf{s}^\dagger + (\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger) \cdot \partial_n\mathbf{s} + (\hat{\mathbf{n}} \cdot \boldsymbol{\xi})\partial_n\phi^\dagger + (\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger)\partial_n\phi]_-^+ \, d^2\mathbf{r} \, dt$$

$$- \int_0^T \int_{\Sigma_{\text{FS}}} \nabla^\Sigma(\delta d) \cdot [(\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}})\mathbf{s} + (\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}})\mathbf{s}^\dagger]_-^+ \, d^2\mathbf{r} \, dt, \tag{51}$$

where we have used the equation of motion (5) and the 2-D version of Gauss' theorem on a closed surface (Dahlen & Tromp 1998, eq. A.79):

$$\int_\Sigma \nabla^\Sigma \cdot \mathbf{v}^\Sigma \, d^2\mathbf{r} = 0, \tag{52}$$

for any tangent vector $\mathbf{v}^\Sigma$.

Upon combining (45) and (51) we may write the variation in the misfit function in the form

$$\delta\chi = \int_V \delta\oplus K_\oplus \, d^3\mathbf{r} + \int_\Sigma K_d \, \delta\ln d \, d^2\mathbf{r} + \int_{\Sigma_{\text{FS}}} \mathbf{K}_d \cdot \nabla^\Sigma \delta\ln d \, d^2\mathbf{r} + \int_0^T \int_V \mathbf{s}^\dagger(T - t) \cdot \delta\mathbf{f}(t) \, d^3\mathbf{r} \, dt, \tag{53}$$

where the symbol $\oplus$ denotes the earth model parameters ρ and $\mathbf{c}$, and the kernels $K_\oplus$, K_d and $\mathbf{K}_d$ are defined by

$$\delta\oplus K_\oplus = -\int_0^T \{\delta\rho\, [\mathbf{s}^\dagger \cdot (\partial_t^2\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t\mathbf{s}) + \mathbf{s}^\dagger \cdot \nabla\phi + \mathbf{s} \cdot \nabla\phi^\dagger + \mathbf{s} \cdot \nabla\nabla(\Phi + \psi) \cdot \mathbf{s}^\dagger + \nabla(\Phi + \psi) \cdot (\mathbf{s} \cdot \nabla\mathbf{s}^\dagger - \mathbf{s}\nabla \cdot \mathbf{s}^\dagger)]$$

$$+ \nabla\mathbf{s}^\dagger :\delta\mathbf{c}:\nabla\mathbf{s} + \rho\mathbf{s} \cdot \nabla\nabla(\delta\Phi) \cdot \mathbf{s}^\dagger + \rho\nabla(\delta\Phi) \cdot (\mathbf{s} \cdot \nabla\mathbf{s}^\dagger - \mathbf{s}\nabla \cdot \mathbf{s}^\dagger)\} \, dt, \tag{54}$$

$$K_d = -\int_0^T d \left[\rho\left\{-(\partial_t^2\mathbf{s} + 2\boldsymbol{\Omega} \times \partial_t\mathbf{s}) \cdot \mathbf{s}^\dagger - \mathbf{s}^\dagger \cdot \nabla\phi - \mathbf{s} \cdot \nabla\phi^\dagger - \mathbf{s} \cdot \nabla\nabla(\Phi + \psi) \cdot \mathbf{s}^\dagger - \nabla(\Phi + \psi) \cdot (\mathbf{s}^\dagger \cdot \nabla\mathbf{s} - \mathbf{s}^\dagger\nabla \cdot \mathbf{s})\right\}\right.$$

$$\left. - \nabla\mathbf{s}^\dagger :\mathbf{c}:\nabla\mathbf{s} - (4\pi G)^{-1}(\nabla\phi \cdot \nabla\phi^\dagger) + (\hat{\mathbf{n}} \cdot \mathbf{T}) \cdot \partial_n\mathbf{s}^\dagger + (\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger) \cdot \partial_n\mathbf{s} + (\hat{\mathbf{n}} \cdot \boldsymbol{\xi})\partial_n\phi^\dagger + (\hat{\mathbf{n}} \cdot \boldsymbol{\xi}^\dagger)\partial_n\phi\right]_-^+ \, dt, \tag{55}$$

$$\mathbf{K}_d = \int_0^T d \left[(\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger \cdot \hat{\mathbf{n}})\mathbf{s} + (\hat{\mathbf{n}} \cdot \mathbf{T} \cdot \hat{\mathbf{n}})\mathbf{s}^\dagger\right]_-^+ \, dt, \tag{56}$$

where it should be remembered that in these kernel expressions the adjoint field is evaluated at time $T - t$, whereas the regular wavefield is evaluated at time t.

If we ignore rotation and self-gravitation, the gradient (53) may be rewritten as

$$\delta\chi = \int_V (\delta\ln\rho K_\rho + \delta\mathbf{c}::\mathbf{K_c}) \, d^3\mathbf{r} + \int_\Sigma K_d \, \delta\ln d \, d^2\mathbf{r} + \int_{\Sigma_{\text{FS}}} \mathbf{K}_d \cdot \nabla^\Sigma \delta\ln d \, d^2\mathbf{r} + \int_0^T \int_V \mathbf{s}^\dagger(T - t) \cdot \delta\mathbf{f}(t) \, d^3\mathbf{r} \, dt, \tag{57}$$

where the kernels K_ρ, $\mathbf{K_c}$, K_d and $\mathbf{K}_d$ are defined by

$$K_\rho = -\int_0^T \rho\mathbf{s}^\dagger(T - t) \cdot \partial_t^2\mathbf{s}(t) \, dt, \tag{58}$$

$$\mathbf{K_c} = -\int_0^T \nabla\mathbf{s}^\dagger(T - t)\nabla\mathbf{s}(t) \, dt, \tag{59}$$

$$K_d = -\int_0^T d[-\rho\mathbf{s}^\dagger(T - t) \cdot \partial_t^2\mathbf{s}(t) - \nabla\mathbf{s}^\dagger(T - t):\mathbf{c}:\nabla\mathbf{s}(t) + \hat{\mathbf{n}} \cdot \mathbf{T}(t) \cdot \partial_n\mathbf{s}^\dagger(T - t) + \hat{\mathbf{n}} \cdot \mathbf{T}^\dagger(T - t) \cdot \partial_n\mathbf{s}(t)]_-^+ \, dt, \tag{60}$$

$$\mathbf{K}_d = \int_0^T d[\hat{\mathbf{n}} \cdot \mathbf{T}^\dagger(T - t) \cdot \hat{\mathbf{n}}\, \mathbf{s}(t) + \hat{\mathbf{n}} \cdot \mathbf{T}(t) \cdot \hat{\mathbf{n}}\, \mathbf{s}^\dagger(T - t)]_-^+ \, dt. \tag{61}$$

Sieminski *et al.* (2007a,b) explore the implications of full anisotropy for surface and body waves. In this paper we will restrict ourselves to isotropic kernels.

In an isotropic earth model we have $c_{jklm} = (\kappa - \frac{2}{3}\mu)\delta_{jk}\delta_{lm} + \mu(\delta_{jl}\delta_{km} + \delta_{jm}\delta_{kl})$, where μ and κ denote the shear and bulk moduli, respectively. Note that the shear modulus μ should not be confused with the scalar Lagrange multiplier used earlier. Thus we may write

$$\delta\mathbf{c}::\mathbf{K_c} = \delta\ln\mu\,K_\mu + \delta\ln\kappa\,K_\kappa, \tag{62}$$

where the isotropic kernels K_μ and K_κ represent Fréchet derivatives with respect to relative bulk and shear moduli perturbations $\delta\ln\kappa = \delta\kappa/\kappa$ and $\delta\ln\mu = \delta\mu/\mu$, respectively. These isotropic kernels are given by

$$K_\mu(\mathbf{r}) = -\int_0^T 2\mu(\mathbf{r})\mathbf{D}^\dagger(\mathbf{r}, T-t):\mathbf{D}(\mathbf{r},t)\,dt, \tag{63}$$

$$K_\kappa(\mathbf{r}) = -\int_0^T \kappa(\mathbf{r})[\boldsymbol{\nabla}\cdot\mathbf{s}^\dagger(\mathbf{r}, T-t)][\boldsymbol{\nabla}\cdot\mathbf{s}(\mathbf{r},t)]\,dt, \tag{64}$$

where $\mathbf{D}$ and $\mathbf{D}^\dagger$ denote the traceless deviatoric strain and its adjoint, respectively. Similarly, in an isotropic earth model (60) and (61) reduce to

$$\begin{aligned}
K_d(\mathbf{r}) = \int_0^T d(\mathbf{r})\big[&\rho(\mathbf{r})\mathbf{s}^\dagger(\mathbf{r}, T-t)\cdot\partial_t^2\mathbf{s}(\mathbf{r},t) + \kappa(\mathbf{r})\boldsymbol{\nabla}\cdot\mathbf{s}^\dagger(\mathbf{r}, T-t)\boldsymbol{\nabla}\cdot\mathbf{s}(\mathbf{r},t) + 2\mu(\mathbf{r})\mathbf{D}^\dagger(\mathbf{r}, T-t):\mathbf{D}(\mathbf{r},t) \\
&- \kappa(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})\cdot\partial_n\mathbf{s}^\dagger(\mathbf{r}, T-t)\boldsymbol{\nabla}\cdot\mathbf{s}(\mathbf{r},t) - 2\mu(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})\partial_n\mathbf{s}^\dagger(\mathbf{r}, T-t):\mathbf{D}(\mathbf{r},t) \\
&- \kappa(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})\cdot\partial_n\mathbf{s}(\mathbf{r},t)\boldsymbol{\nabla}\cdot\mathbf{s}^\dagger(\mathbf{r}, T-t) - 2\mu(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})\partial_n\mathbf{s}(\mathbf{r},t):\mathbf{D}^\dagger(\mathbf{r}, T-t)\big]_-^+\,dt,
\end{aligned} \tag{65}$$

$$\begin{aligned}
\mathbf{K}_d(\mathbf{r}) = \int_0^T d(\mathbf{r})\big[&\mathbf{s}^\dagger(\mathbf{r}, T-t)[\kappa(\mathbf{r})\boldsymbol{\nabla}\cdot\mathbf{s}(\mathbf{r},t) + 2\mu(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})\cdot\mathbf{D}(\mathbf{r},t)\cdot\hat{\mathbf{n}}(\mathbf{r})] \\
&+ \mathbf{s}(\mathbf{r},t)[\kappa(\mathbf{r})\boldsymbol{\nabla}\cdot\mathbf{s}^\dagger(\mathbf{r}, T-t) + 2\mu(\mathbf{r})\hat{\mathbf{n}}(\mathbf{r})\cdot\mathbf{D}^\dagger(\mathbf{r}, T-t)\cdot\hat{\mathbf{n}}(\mathbf{r})]\big]_-^+\,dt.
\end{aligned} \tag{66}$$

Finally, we may express the Fréchet derivatives in an isotropic earth model in terms of relative variations in density $\delta\ln\rho$, shear wave speed $\delta\ln\beta$, and compressional-wave speed $\delta\ln\alpha$ based upon the relationship

$$\delta\ln\rho\,K_\rho + \delta\ln\mu\,K_\mu + \delta\ln\kappa\,K_\kappa = \delta\ln\rho\,K_\rho' + \delta\ln\beta\,K_\beta + \delta\ln\alpha\,K_\alpha, \tag{67}$$

where

$$K_{\rho'} = K_\rho + K_\kappa + K_\mu, \qquad K_\beta = 2\left(K_\mu - \frac{4}{3}\frac{\mu}{\kappa}K_\kappa\right), \qquad K_\alpha = 2\left(\frac{\kappa + \frac{4}{3}\mu}{\kappa}\right)K_\kappa. \tag{68}$$

Tromp *et al.* (2005) demonstrated that the K_ρ', K_β and K_α kernels are more fundamental than the K_ρ, K_μ and K_κ kernels in the sense that the former are more relevant in the context of traveltime tomography.

4 ANELASTICITY

In an anelastic earth model, Hooke's law (6) needs to be replaced by (e.g. Aki & Richards 1980; Dahlen & Tromp 1998)

$$\mathbf{T}(t) = \int_0^t \mathbf{c}(t-t'):\partial_{t'}\boldsymbol{\nabla}\mathbf{s}(t')\,dt', \tag{69}$$

that is, the tensor $\mathbf{c}$ becomes time dependent and the stress at time t depends on the entire strain history. To avoid clutter we have suppressed the spatial dependence. In this section, we will determine the implications of this more general, anelastic constitutive relationship for the adjoint equations. We begin by manipulating the term of concern in (15):

$$\int_0^T\int_V \boldsymbol{\lambda}\cdot(\boldsymbol{\nabla}\cdot\mathbf{T})\,d^3\mathbf{r}\,dt = -\int_0^T\int_V \boldsymbol{\nabla}\boldsymbol{\lambda}:\mathbf{T}\,d^3\mathbf{r}\,dt = -\int_0^T\int_V \boldsymbol{\nabla}\boldsymbol{\lambda}(t):\left[\int_0^t \mathbf{c}(t-t'):\partial_{t'}\boldsymbol{\nabla}\mathbf{s}(t')\,dt'\right]d^3\mathbf{r}\,dt, \tag{70}$$

where we have used the boundary condition (7) and imposed the (anticipated) continuity conditions (29) and (30). Upon integrating by parts over t', using the initial condition $\mathbf{s}(0) = \mathbf{0}$ and the relationship $\partial_{t'}\mathbf{c}(t-t') = -\partial_t\mathbf{c}(t-t')$, changing the order of integration over t and t', integrating by parts over t, imposing the (anticipated) end condition (31), and the symmetries of the tensor $\mathbf{c}$, we obtain

$$\int_0^T\int_V \boldsymbol{\lambda}\cdot(\boldsymbol{\nabla}\cdot\mathbf{T})\,d^3\mathbf{r}\,dt = \int_V\int_0^T\left[\int_{t'}^T \mathbf{c}(t-t'):\partial_t\boldsymbol{\nabla}\boldsymbol{\lambda}(t)\,dt\right]:\boldsymbol{\nabla}\mathbf{s}(t')\,dt'\,d^3\mathbf{r}. \tag{71}$$

If we define the second-order tensor

$$\boldsymbol{\sigma}(t) = -\int_t^T \mathbf{c}(t'-t):\partial_{t'}\boldsymbol{\nabla}\boldsymbol{\lambda}(t')\,dt', \tag{72}$$

then it is straightforward to show based upon varying (71) that in an anelastic medium the Lagrange multiplier field $\boldsymbol{\lambda}$ is determined by the wave equation (27), provided we use the definition (72) for $\boldsymbol{\sigma}$, rather than the elastic form (25).

As in the elastic case, we define the adjoint wavefield $\mathbf{s}^\dagger$ in terms of the Lagrange multiplier field $\boldsymbol{\lambda}$ by (34). By analogy, let us define the anelastic adjoint stress $\mathbf{T}^\dagger$ in terms of the the second-order tensor $\boldsymbol{\sigma}$ by

$$\mathbf{T}^\dagger(\mathbf{r}, t) = \boldsymbol{\sigma}(\mathbf{r}, T - t). \tag{73}$$

Based upon (72) and (73) it is now straightforward to show that in an anelastic medium the adjoint wavefield $\mathbf{s}^\dagger$ is determined by (36), provided the anelastic adjoint stress $\mathbf{T}^\dagger$ is defined in terms of the gradient of the adjoint displacement by

$$\mathbf{T}^\dagger(t) = \int_0^t \mathbf{c}(t - t') : \partial_{t'} \boldsymbol{\nabla} \mathbf{s}^\dagger(t') \, dt'. \tag{74}$$

We see that in an anelastic earth model, just as in the elastic case, the adjoint wavefield is determined by exactly the same equations as the regular wavefield, with the exception of the source term and the sense of rotation.

In the context of the tomographic inverse problem, the gradient of the misfit function becomes

$$\delta\chi = \int_V \left[\delta \ln \rho \, K_\rho - \int_0^T \delta\mathbf{T}^\dagger(T - t) : \boldsymbol{\nabla}\mathbf{s}(t) \, dt \right] d^3\mathbf{r} + \int_\Sigma K_d \, \delta \ln d \, d^2\mathbf{r} + \int_{\Sigma_{\mathrm{FS}}} \mathbf{K}_d \cdot \boldsymbol{\nabla}^\Sigma \delta \ln d \, d^2\mathbf{r} + \int_0^T \int_V \mathbf{s}^\dagger(T - t) \cdot \delta\mathbf{f}(t) \, d^3\mathbf{r} \, dt, \tag{75}$$

where K_ρ is defined by (58) and where we have defined

$$\delta\mathbf{T}^\dagger(t) = \int_0^t \delta\mathbf{c}(t - t') : \partial_{t'} \boldsymbol{\nabla} \mathbf{s}^\dagger(t') \, dt'. \tag{76}$$

Note that for purely elastic perturbations, when the perturbation $\delta\mathbf{c}$ does not depend on time, the gradient (75) reduces to (57). This implies that a tomographic inversion for elastic perturbations superimposed upon an anelastic reference model may be accomplished based upon the gradient (57), provided the regular and adjoint wavefields are calculated in the anelastic reference model.

5 GLOBAL SEISMIC WAVE PROPAGATION BASED UPON THE SPECTRAL-ELEMENT METHOD

As demonstrated in Section 3, the expressions for the Fréchet derivatives of volumetric model parameters, for example, (58), (59), (63) and (64), and topographic variations on internal discontinuities, for example (60) and (61), involve simultaneous access to the forward wavefield $\mathbf{s}(\mathbf{r}, t)$ and the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{r}, T - t)$. We calculate both the forward and adjoint wavefields based upon the spectral-element method (SEM), which has been successfully applied to simulate wave propagation on both global and regional scales (e.g. Komatitsch & Vilotte 1998; Komatitsch & Tromp 1999, 2002a,b; Chaljub *et al.* 2003; Komatitsch *et al.* 2004). The method combines the geometrical flexibility of the finite-element method with an accurate representation of the wavefield in terms of high-order Lagrange polynomials. The global spectral-element mesh is designed to honour topography and bathymetry on the Earth's surface and wave speed discontinuities inside the model, such as the Moho, the upper mantle discontinuities, the core–mantle boundary (CMB), and the inner-core boundary (ICB). The choice of Lagrange interpolants in combination with Gauss–Lobatto–Legendre quadrature produces an exactly diagonal mass matrix, which makes it straightforward to implement the method on parallel computers. A simple predictor-corrector scheme is used to time-march the wavefield. The detailed numerical implementation and numerous benchmarks are discussed in Komatitsch & Tromp (2002a,b). In general, use of the SEM to simulate wave propagation in 3-D global earth models results in better traveltime predictions and waveform fits compared to synthetics generated based upon semi-analytical methods for 1-D earth models (e.g. Ritsema *et al.* 2002; Tsuboi *et al.* 2003; Ji *et al.* 2005). A typical 3-D global forward simulation for 40 minutes long synthetic seismograms accurate to a shortest period of 18 s takes approximately 4 hr on 19 nodes of the Dell cluster CITerra (http://citerra.gps.caltech.edu/). Each node has 12 GB of memory and two Intel Quad Core 2.33 GHz Clovertown processors.

6 NUMERICAL IMPLEMENTATION OF THE ADJOINT METHOD

Although it is straightforward to compute the forward and adjoint wavefields by solving the regular wave equation (5) and the adjoint wave equation (36) using the SEM, evaluation of the kernel expressions (58), (59), (60) and (61) involves simultaneous access to the forward wavefield $\mathbf{s}(\mathbf{r}, t)$ and the adjoint wavefield in reverse time $\mathbf{s}^\dagger(\mathbf{r}, T - t)$. This makes it difficult to carry both the forward and adjoint wavefield simultaneously in the SEM, because at any given time t we would only be able to access the forward wavefield $\mathbf{s}(\mathbf{r}, t)$ and the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{r}, t)$, while the time integration involved in constructing the kernels requires access to the time-reversed adjoint field $\mathbf{s}^\dagger(\mathbf{r}, T - t)$. One apparent solution is to run the forward simulation and save the entire regular wavefield $\mathbf{s}(\mathbf{r}, t)$ as a function of both space and time, followed by an adjoint simulation during which at any given time t we read back the forward wavefield at time $T - t$. This scheme is easy to understand and theoretically simple to implement, however, realistic global SEM simulations accurate down to 18 s generally require approximately 12 GB of storage space per time step to save the forward wavefield, and a typical simulation for 40 minutes long seismograms involves a number of time steps on the order of 10^4. Therefore, the total storage requirements involve hundreds of terabytes of disk space, and the related I/O communications may significantly increase the total simulation time. One possible solution may be to employ an efficient data compression scheme to reduce the storage requirements and overhead time associated with accessing the time slices of the forward wavefield (Tu *et al.* 2003).

In this paper, we employ an approach first introduced by Liu & Tromp (2006) for regional wave propagation problems, in which, ignoring anelastic effects, we reconstruct the forward wavefield by solving the equation of motion (5) backwards in time:

$$\rho\left(\partial_t^2\mathbf{s} + 2\mathbf{\Omega}\times\partial_t\mathbf{s}\right) - \nabla\cdot\mathbf{T} + \nabla[\rho\,\mathbf{s}\cdot\nabla(\Phi+\psi)] + \rho\nabla\phi - \nabla\cdot(\rho\,\mathbf{s})\nabla(\Phi+\psi) = \mathbf{f}, \tag{77}$$

$$[\hat{\mathbf{n}}\cdot\mathbf{T}]_-^+ = \mathbf{0} \quad \text{on } \Sigma, \tag{78}$$

$$[\mathbf{s}]_-^+ = \mathbf{0} \quad \text{on } \Sigma_{\text{SS}}, \qquad [\hat{\mathbf{n}}\cdot\mathbf{s}]_-^+ = 0 \quad \text{on } \Sigma_{\text{FS}}, \tag{79}$$

$$\mathbf{s}(\mathbf{r}, T) \quad \text{and} \quad \partial_t\mathbf{s}(\mathbf{r}, T) \quad \text{given}. \tag{80}$$

This 'initial' and boundary value problem can be solved to reconstruct $\mathbf{s}(\mathbf{r}, t)$ for $T \geq t \geq 0$ the same way the forward wave equation (5) is solved subject to the boundary conditions (7), (8) and (9), and the initial conditions (10). Actually, we simply need to replace the time step parameter Δt in the forward simulation by $-\Delta t$ in the backward simulation, while reintroducing the source term $\mathbf{f}$ in reverse time. Note that we need to have the 'ending conditions' $\mathbf{s}(\mathbf{r}, T)$ and $\partial_t\mathbf{s}(\mathbf{r}, T)$ to start the reconstruction of $\mathbf{s}(\mathbf{r}, t)$, and therefore we need to save on disk the final displacement and velocity fields of a previous forward simulation. By carrying the backward simulation (77) and the adjoint simulation (36) simultaneously, at any given time t we have access to both the regular wavefield $\mathbf{s}(\mathbf{r}, T - t)$ and the adjoint wavefield $\mathbf{s}^\dagger(\mathbf{r}, t)$, making it feasible to compute on-the-fly the contributions to the time integrations involved in the construction of the finite-frequency kernels (58), (59), (60) and (61).

In summary, sensitivity kernels may be calculated based upon just two spectral-element simulations. One forward simulation provides the synthetic seismograms that may be used for a cross-correlation traveltime measurement. During the second simulation we calculate the adjoint wavefield while simultaneously reconstructing the forward wavefield based upon the last snapshots from the first simulation, thereby enabling the on-the-fly calculation of the sensitivity kernels. The second simulation requires about twice the memory and run time of the first simulation, but avoids overwhelming storage requirements and I/O demands.

As discussed by Liu & Tromp (2006), in regional spectral-element simulations one needs to save the wavefield absorbed on the fictitious boundaries of the model domain as well as a snapshot of the entire final wavefield. Since the global simulations involve no absorbing boundaries, only the final snapshot needs to be saved. This requires very little storage space compared to the absorbing contributions that need to be saved at every time step during the regional forward simulations. We note, however, that our open source global spectral-element software can be used for regional applications by using only one or two 'chunks' of the cubed-sphere mesh, in which case we do store the wavefield absorbed on the artificial boundaries of the domain.

Although anelasticity can be naturally incorporated in adjoint methods, as discussed in Section 4, the procedure for calculating kernels for an anelastic reference model is more involved. The irreversibility of the anelastic wave equation makes it numerically unstable to solve the backward eqs (77)–(80), and therefore makes it impossible to perform the kernel calculations on-the-fly. One solution to this dilemma is to resort to the generic approach of storing the whole forward wavefield as a function of space and time. A compromise—an approach generally known as 'checkpointing' (Restrepo & Griewank 1998; Griewank & Walther 2000; Hinze *et al.* 2005)—is to solve the backward equations with attenuation on, but only for a number of time steps that is within the numerically stable regime, and then 'resetting' the wavefield based upon a snapshot from the forward solution. Using this approach we can reconstruct the entire forward wavefield using a limited number of snapshots from the forward simulation to regularly 'reset' the backward construction, thereby significantly reducing the storage requirements compared to the alternative approach of saving the entire forward wavefield. For this study we focus on probing the elastic structure of the Earth, and therefore we ignore attenuation in the kernel calculations presented in what follows. We should note that kernels calculated by ray-based methods (e.g. Dahlen *et al.* 2000; Hung *et al.* 2000) also do not accommodate the effects of attenuation.

7 APPLICATION TO SPHERICALLY SYMMETRIC MODELS

From the formulation of the Fréchet kernels, it is apparent that the adjoint approach can be used to compute sensitivity kernels for any 3-D reference models at no extra cost. However, to date most research on finite-frequency sensitivity kernels has been based primarily on 1-D reference models for which synthetic seismograms can be generated semi-analytically (Marquering *et al.* 1999; Dahlen *et al.* 2000; Hung *et al.* 2000; Zhou *et al.* 2004; Zhao *et al.* 2000; Zhao & Jordan 2006). Therefore, for comparison, in this section we calculate cross-correlation traveltime sensitivity kernels for typical global arrivals in the 1-D isotropic Preliminary Reference Earth Model (PREM) (Dziewonski & Anderson 1981). Also note that although we have presented the adjoint eq. (36) in its most general form, incorporating the effects of self-gravitation and rotation, these complications are usually ignored in the context of global tomographic studies. We also ignore them for the calculation of sensitivity kernels discussed in Sections 7 and 8.

For a given component of the seismogram, we may express the cross-correlation traveltime perturbation as (Luo & Schuster 1991; Marquering *et al.* 1999; Dahlen *et al.* 2000)

$$\delta T = \int_V \delta\oplus K_\oplus\,\mathrm{d}^3\mathbf{r} + \int_\Sigma K_d\,\delta\ln d\,\mathrm{d}^2\mathbf{r} + \int_{\Sigma_{\text{FS}}}\mathbf{K}_d\cdot\nabla^\Sigma\delta\ln d\,\mathrm{d}^2\mathbf{r}. \tag{81}$$

The expressions for the volumetric kernels $K_\oplus$ and the boundary kernels K_d and $\mathbf{K}_d$ are the same as equations (54), (55) and (56), except that for a cross-correlation traveltime measurement the associated adjoint wavefield $\mathbf{s}^\dagger$ is generated by the adjoint source (Tromp *et al.*

2005)

$$\mathbf{f}^{\dagger} = \frac{1}{M_r} w_r(T - t)\partial_t \mathbf{s}(\mathbf{r}_r, T - t)\delta(\mathbf{r} - \mathbf{r}_r), \tag{82}$$

where $\mathbf{r}_r$ denotes the receiver location, $w(t)$ is the cross-correlation window, and the normalization factor M is given by

$$M = \int_0^T w(t)s(\mathbf{r}_r, t)\partial_t^2 s(\mathbf{r}_r, t)\, dt. \tag{83}$$

Notice that the adjoint source $\mathbf{f}^{\dagger}$ and thus the adjoint wavefield $\mathbf{s}^{\dagger}$ are determined by the (three-component) synthetic velocity seismogram $\partial_t \mathbf{s}(t)$, and therefore the sensitivity kernels can be computed based upon only this information.

We simulate three-component seismograms for the 1994 June 9 Bolivia Earthquake (except in Sections 7.9 and 7.10.1). This earthquake occurred at a depth of 647 km and is one of the largest deep events in modern recording history. We use the event location and the centroid-moment tensor (CMT) solution from the Global CMT catalogue (http://www.globalcmt.org) for our simulations, and we modify the half duration of the CMT solution to be compatible with the resolution of the spectral-element mesh. We follow the procedures outlined in Section 6 to calculate sensitivity kernels for typical global phases at various distances. Most of the kernels are generated for a hypothetical station at a source azimuth of 65°, unless otherwise noted. We will consider the sensitivity kernels of the following phases: direct P and S at an epicentral distance of 60°, the core-reflected phase ScS at a distance of 60°, the core phases $PKPab$ and $PKPdf$ at a distance of 170° and SKS at a distance of 114°, the core-diffracted phase P_{diff} at a distance of 103°, the depth phases sP at a distance of 40° and sS at a distance of 90°, as well as the surface-reflected phases PP at a distance of 114° and SS at a distance of 90°. These arrivals are chosen both because they are readily identified in the seismograms in this particular azimuth at these specific distances, and because of their wide-spread use in global seismology. Fig. 1 illustrates the ray geometry of these arrivals for the PREM model.

One of the advantages of our fully-numerical approach based upon the adjoint method is that we need not be able to 'label' a particular phase in the seismogram, that is, we need not have any knowledge of the ray path associated with a particular phase. By performing the forward and adjoint simulations, and combining the forward and adjoint wavefields during the construction of the kernels, we automatically obtain the 3-D sensitivity associated with this particular phase in the seismogram. Sometimes the kernels may be readily identified with a specific geometrical ray path, but frequently the sensitivity kernels are much richer. Another advantage of the adjoint approach is that it may be used for fully 3-D reference earth models, as we discuss further in Sections 8 and 9.

7.1 P kernel

We generate 3-D compressional-wave speed sensitivity kernels K_α for the P phase at an epicentral distance of 60°. Figs 2(a)–(c) show source–receiver cross-sections of the kernels computed based upon the adjoint method for simulations accurate down to periods of 27, 18 and 9 s, respectively. The corresponding vertical-component velocity seismograms are shown in Fig. 2(d). All kernels exhibit very nice banana–doughnut shapes, in good agreement with ray-based finite-frequency kernels (Dahlen *et al.* 2000; Hung *et al.* 2000; Zhou *et al.* 2004) as well as kernels calculated based upon normal-mode methods (Zhao *et al.* 2005; Zhao & Jordan 2006). Notice that the size of the doughnut holes decreases with increasing resolution, in accordance with the scaling relation width $\sim \sqrt{\lambda L}$, where L denotes the length of the ray path and λ the wavelength. At very high frequencies the sensitivity kernel will asymptotically collapse onto the P ray path shown in Fig. 1(a). Notice that the absolute amplitude of the P kernel at 9 s is about five times larger than the amplitude of the P kernel at 27 s, which is not unexpected since the P wave traveltime anomaly $\delta T_p = \int K_\alpha \delta \ln \alpha \, d^3\mathbf{x}$ is roughly frequency independent, and the wider the kernel, the larger the effective integration volume, and thus the smaller the kernel amplitude in order to keep the traveltime anomaly the same. We also note that the source radiation pattern is clearly imprinted on the kernel near the hypocentre (denoted by a white circle in Figs 2a–c).
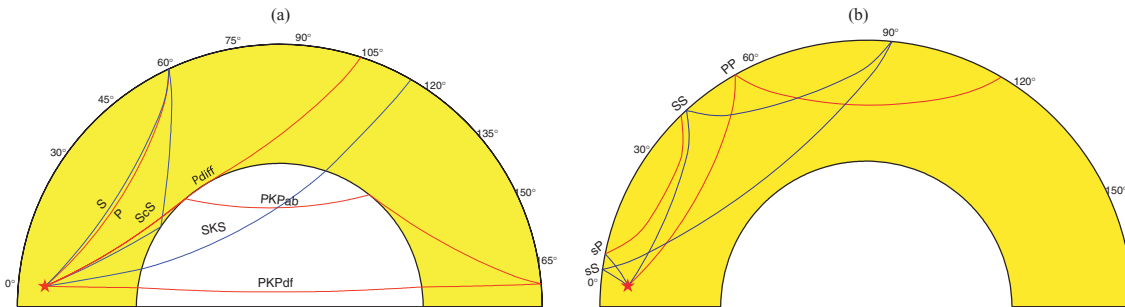
Figure 1. (a) PREM ray geometry for the following arrivals: P, S, ScS, P_{diff}, SKS, $PKPab$ and $PKPdf$. The S legs of the ray paths are coloured blue, while the P legs of the ray paths are coloured red. The hypocentre location of the 1994 June 9, Bolivia earthquake at a depth of 647 km is denoted by a red star. (b) PREM ray geometry for surface-reflected phases such as PP, SS and the depth-phases sS and sP.
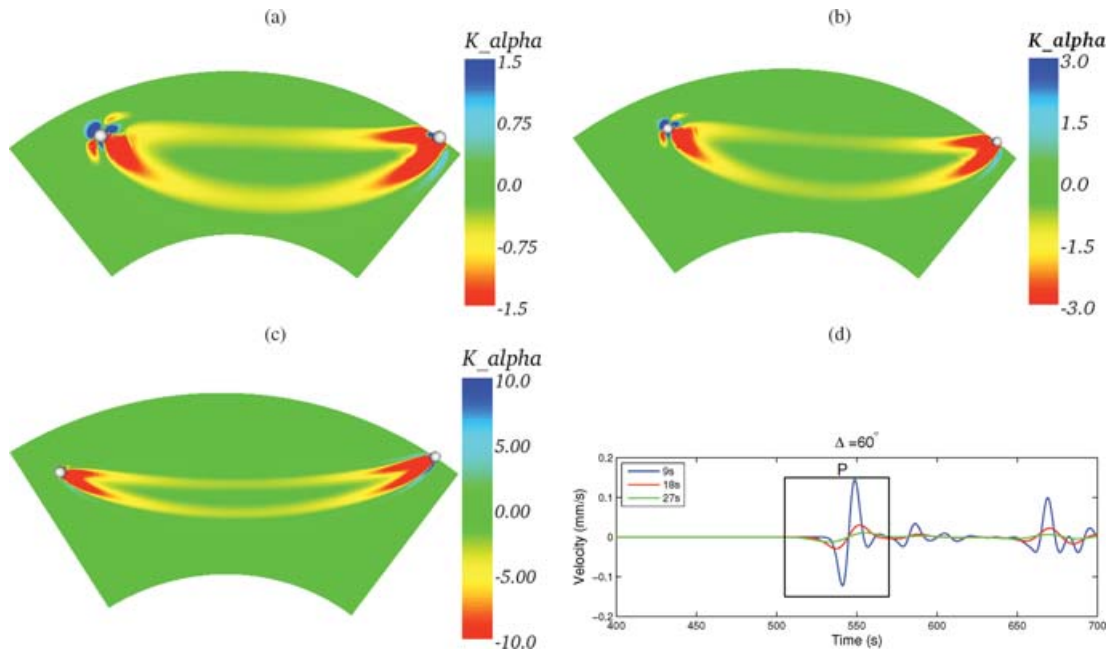
Figure 2. (a) Source–receiver cross-section of the K_α kernel for 27-s-and-longer P waves recorded at an epicentral distance of $60°$. The source and receiver locations are denoted by two small white spheres. The units of the volumetric sensitivity kernels (as opposed to the boundary kernels discussed in Section 7.10) are 10^{-7} s km^{-3} throughout this paper. The reference model for these seismic wave simulations is PREM throughout Section 7. We will also refer to the 27/18/9 s and longer kernels as the '27/18/9 s kernels' for brevity. (b) 18 s K_α kernel for P waves recorded at an epicentral distance of $60°$. (c) 9 s K_α kernel for P waves recorded at an epicentral distance of $60°$. (d) Corresponding vertical-component synthetic velocity seismograms recorded at an epicentral distance of $60°$ for simulations accurate down to periods of 27, 18 and 9 s, respectively. The P phases are identified by the black box.

7.2 P_{diff} kernel

Because of the sudden drop in compressional-wave speed across the CMB towards the outer core, a shadow zone exists beyond an epicentral distance of approximately $103°$ for the direct P wave. P-wave energy enters the shadow zone in the form of waves diffracted along the CMB. The ray path for the P_{diff} phase at an epicentral distance of $103°$ is shown in Fig. 1(a). Because of its diffracted nature, the P_{diff} phase generally has a lower frequency content compared to the direct P wave. Fig. 3(c) shows a vertical-component velocity seismogram recorded at an epicentral distance of $103°$, and Fig. 3(a) shows the source–receiver cross-section of the corresponding K_α kernel. It is similar to a regular banana–doughnut kernel except that it has been cut through at the bottom by the CMB. Fig. 3(b) shows a map view of the sensitivity of the P_{diff} phase on the mantle side of the CMB, which has an elliptical shape with a major axis that is comparable in length to the diffracted portion of the P_{diff} ray path along the CMB.

Note the asymmetry of the elliptical pattern on the CMB above and below the source–receiver great circle in Fig. 3(b). To confirm that this asymmetry is due to the radiation pattern of the earthquake, we compute P_{diff} sensitivity kernels for several stations at the same distance ($\Delta = 120°$) but in different azimuths (Fig. 4). For reference, we also compute the P_{diff} sensitivity kernel for a P_{diff} phase generated by an explosive source at the same distance (in the blue box in Fig. 4). Clearly the radiation pattern has a noticeable effect on the kernels. This is readily understandable from the kernel expressions, for example, eqs (58) and (59), because a kernel calculation involves the interaction between the forward wavefield, which is affected by the radiation pattern of the earthquake, and the adjoint wavefield, which involves the radiation pattern generated by a point force exerted at the receiver.

The P_{diff} phase is frequently assumed to be mainly sensitive to the diffracted segment of the ray path along the CMB. Note, however, that the diffracted portion of the kernel indicates significant sensitivity away from the ray path (Fig. 1a), that is, to regions more than 500 km above the CMB. Finally, we note that the calculation of P_{diff} kernels based upon ray-based methods is complicated due to the diffracted nature of the waveform, which leads to a fractional frequency dependence of its phase.

7.3 PKP kernel

For stations beyond the shadow zone, the first arrival is usually the P_{diff} phase. However, at epicentral distance larger than approximately $130°$, the PKP phases that arrive about 60–100 s after the diffracted P phase are usually the first clearly identifiable arrivals in vertical component seismograms. The most noticeable PKP arrival branches are $PKPab$ and $PKPdf$, whose ray paths are shown in Fig. 1(a). Notice that the $PKPab$ phase generally samples the shallow part of the outer core, while the $PKPdf$ phase samples the inner core. Fig. 5(c) shows a vertical velocity seismogram recorded at an epicentral distance of $170°$ for a simulation accurate down to 9 s. The $PKPdf$ and $PKPab$ arrivals can be clearly identified in the seismogram. We pick an epicentral distance of $170°$ because the traveltime difference between the $PKPdf$ and
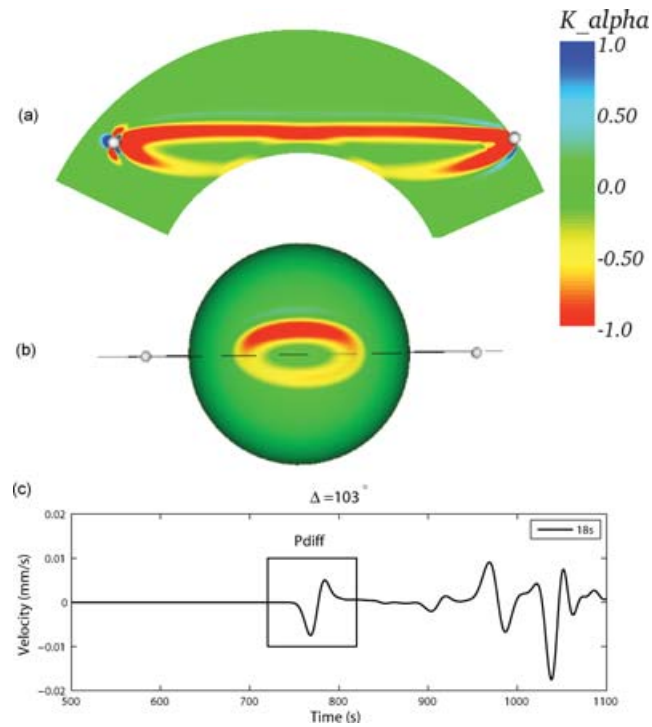
Figure 3. (a) Source–receiver cross-section of the 18 s K_α kernel for the P_{diff} phase recorded at an epicentral distance of 103°. (b) Map view on the CMB of the same kernel as in (a). (c) Corresponding vertical-component synthetic velocity seismogram recorded at an epicentral distance of 103°. The P_{diff} phase is identified by the black box.
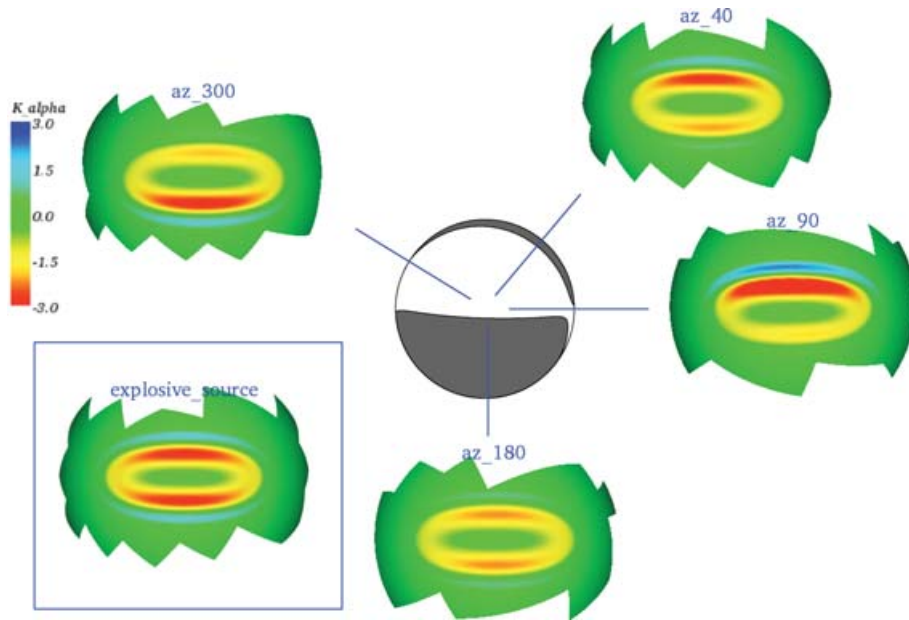
Figure 4. CMB map views of 18 s P_{diff} kernels at a distance of $\Delta = 120°$ and in various azimuths relative to the source mechanism denoted by the beach ball in the centre. Note the change in the kernels with azimuth, illustrating the effect of the radiation pattern on the kernel. A symmetric P_{diff} kernel generated by an explosive source is shown in the blue box for comparison.

PKPab branches can be as large as 70 s at this distance, which makes it very easy to identify and separate them. Figs 5(a) and (b) show the finite-frequency traveltime sensitivity kernels K_α for these two *PKP* branches. The *PKPdf* kernel clearly follows the ray path shown in Fig. 1(a) while exhibiting the typical banana–doughnut shape. The upper part of the kernel has relatively larger sensitivity compared to the lower part, which is partly a result of the source radiation pattern. The *PKPab* kernel not only follows the conventional ray path, but also shows sensitivity along a path that goes from the source to the receiver in the major-arc direction. This is a result of the finite-frequency nature of cross-correlation traveltime measurements: points along the unconventional path represent scatterers that produce arrivals within the *PKPab*
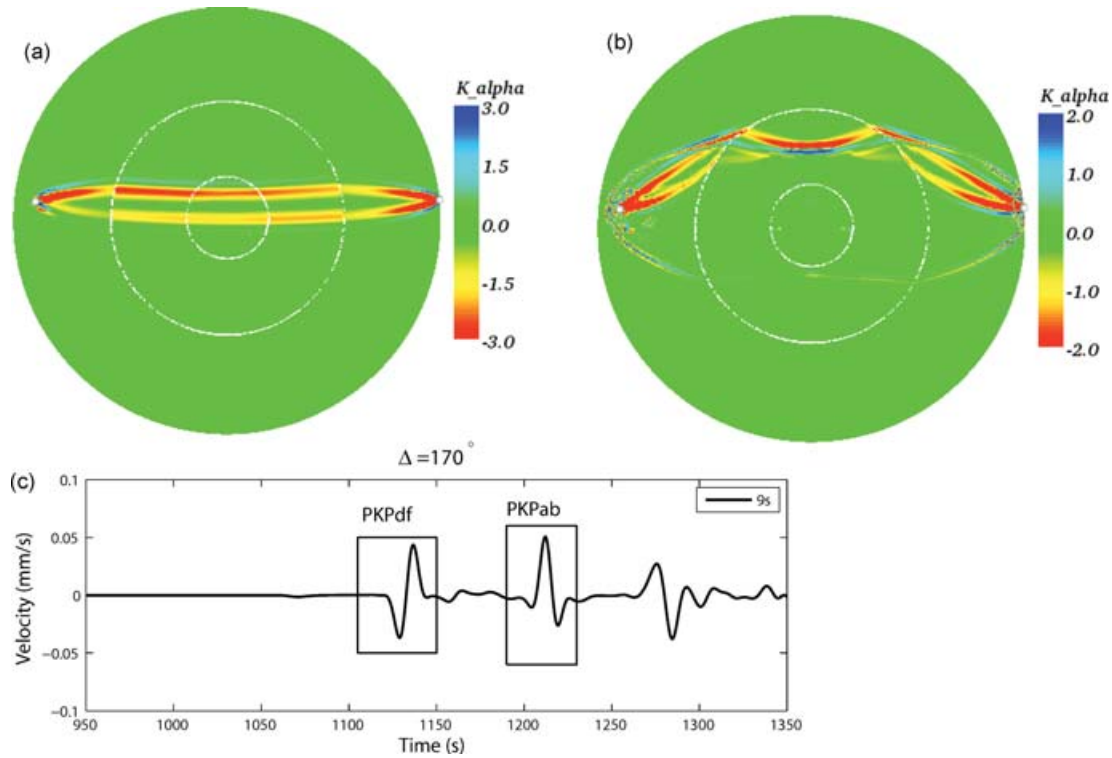
Figure 5. (a) 9 s K_α kernel for a *PKPdf* phase at 170°; (b) 9 s K_α kernel for a *PKPab* phase at 170°. (c) Corresponding vertical-component synthetic velocity seismogram recorded at $\Delta = 170°$, with the *PKPdf* and *PKPab* phases labelled.

cross-correlation traveltime window. Note, however, that the oscillatory nature of the kernel along this major-arc path tends to average out longer wavelength heterogeneity.

7.4 *S* and *ScS* kernels

In this section we calculate the shear wave speed traveltime sensitivity kernels K_β for the *S* and *ScS* phases at 60° for the Bolivia earthquake. These two phases are readily identified on transverse velocity seismograms as illustrated in Fig. 6(e) for synthetics with a shortest period of 27 and 18 s, respectively. The K_β Fréchet kernels corresponding to these phases are compared in Figs 6(a)–(d). Notice that they all nicely follow their respective theoretical ray paths (Fig. 1a) and exhibit clear banana–doughnut shapes, except near the core-reflection point of the *ScS* phase, where the two *S* legs come together, producing a complicated pattern on and right above the CMB. The kernels corresponding to an accuracy of 18 s and longer are generally sharper and have larger amplitudes than the 27 s kernels, in agreement with the results obtained for the *P* kernels in Section 7.1.

7.5 *SKS* kernel

Next, we generate traveltime sensitivity kernels for an 18 s *SKS* phase at an epicentral distance of 114°. Because the outer core is liquid, the *SKS* phase can only be observed on the vertical and radial components of PREM seismograms. In Fig. 7(b) we identify the *SKS* phase on the radial component. Its mantle K_β kernel and outer core K_α kernel are jointly plotted in Fig. 7(a). Overall the sensitivity kernel follows the *SKS* ray path (Fig. 1a), and due to the large contrast between the *S*-wave speed above the CMB and the *P*-wave speed below the CMB, the *K* waves propagate in the outer core with a large incidence angle, causing the sensitivity of the upper part of the banana–doughnut kernel to be much stronger than the lower part of the kernel. We speculate that the same wave speed contrast introduces a head-wave type of *K* wave gliding along the CMB, which is perhaps responsible for the notable portion of the kernel near the incident points on the CMB. Because the *SKS* phase is not the first arrival in the seismogram, mantle reverberated waves caused by reflections and refractions from both the Earth's surface and internal discontinuities may also arrive in the *SKS* cross-correlation time window, and hence the sensitivity kernel has more structure than kernels for a simple direct arrival, such as *P* or P_{diff}. This is true in general for later arriving phases, especially when observed on the radial or vertical components, which involve many *SV*-to-*P* and *P*-to-*SV* conversions. As remarked earlier, the associated rapid oscillations in the amplitude of the kernel provide no constraints on relatively smooth structure.
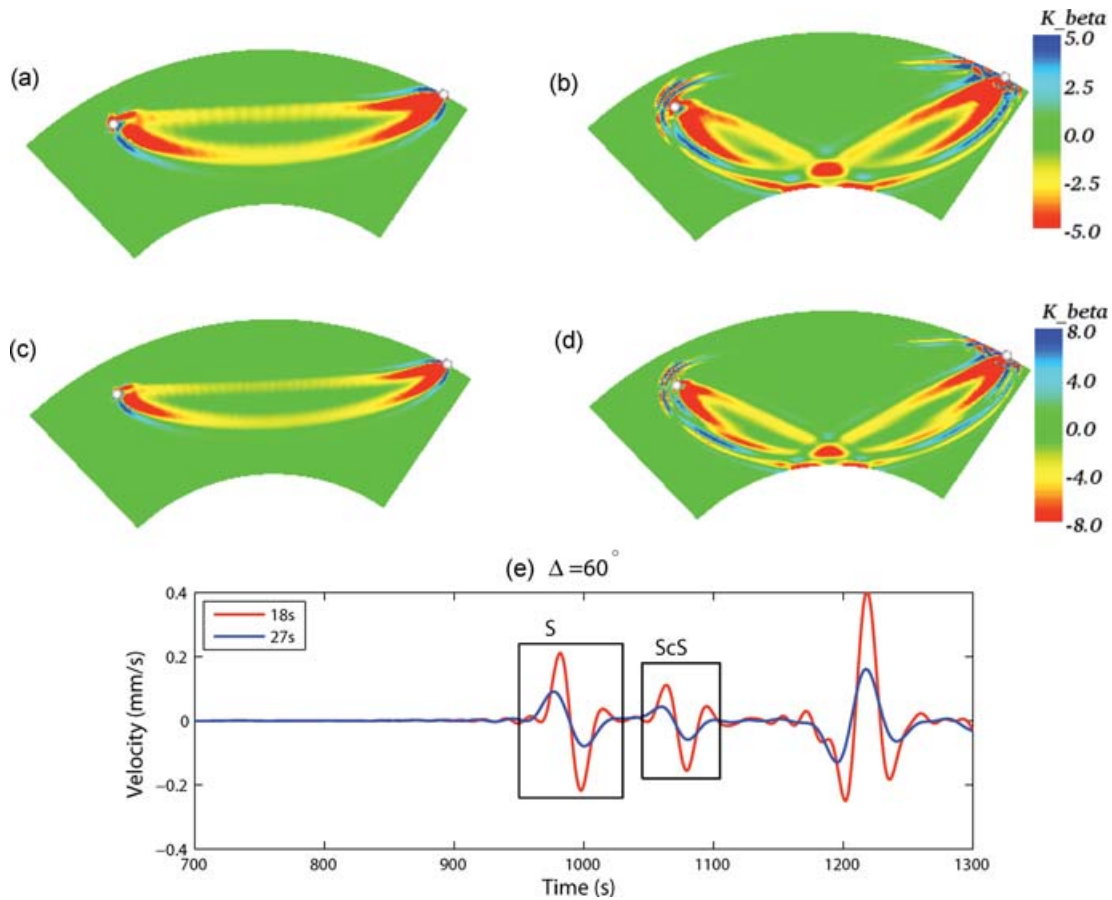
Figure 6. (a) 27 s K_β kernel for *S* waves at $\Delta = 60°$. (b) 27 s K_β kernel for *ScS* waves at $\Delta = 60°$. (a) and (b) share the same colour bar on the right. (c) 18 s K_β kernel for *S* waves at $\Delta = 60°$. (d) 18 s K_β kernel for *ScS* waves at $\Delta = 60°$. (c) and (d) also share the same colour bar. (e) Corresponding transverse-component velocity seismograms computed based upon spectral-element simulations accurate down to periods of 18 and 27 s, respectively. The *S* and *ScS* phases are labelled.

7.6 Depth phases

Since the Bolivia earthquake occurred at a depth of 647 km, we can clearly identify the depth phases in three-component seismograms. In this section, we pick the *sP* phase from a vertical velocity seismogram recorded at a distance of 40° (Fig. 8c) and the *sS* phase in a transverse velocity seismogram recorded at a distance of 90° (Fig. 8f). Fig. 8a shows the K_β sensitivity kernel for the *sP* phase, while Fig. 8(b) shows the corresponding K_α sensitivity kernel. Clearly the K_α kernel reflects the down-going *P* leg of the *sP* phase, while the K_β kernel represents the up-going *S* leg of the *sP* phase. Note that the amplitude of the K_β kernel is much larger than the K_α kernel. Fig. 8(d) shows the K_β sensitivity kernel for the *sS* arrival, which clearly follows the geometrical ray path and exhibits the characteristic banana–doughnut shape along the down-going *S* leg. Notice that although the ray path of the *sS* arrival is at least 500 km above the CMB, as shown in Fig. 1(b), the lower part of the sensitivity kernel actually touches the CMB and leaves a small elliptical-shaped region of sensitivity on the CMB, as shown in Fig. 8(e).

7.7 Kernels for surface-reflected phases

Surface-reflected phases generally become distinct at epicentral distances greater than 40° and dominate the body-wave signals at distances of approximately 100°–110°. These arrivals are mostly sensitive to wave speed structure in the upper mantle, in particular in the transition zone. We compute the K_α kernel (Fig. 9a) for the *PP* phase recorded on the vertical component at a distance of 114° (Fig. 9b), and the K_β sensitivity kernels (Fig. 9c) for the *SS* phase recorded on the transverse component at a distance of 90° (Fig. 9d). The K_β kernel for the *SS* phase follows the ray path as shown in Fig. 1(b), and displays a complicated folded pattern near the surface reflection point. Similar patterns are also observed in the K_α kernel for the *PP* phase.

7.8 *P′P′* kernel

The adjoint approach can also be used to look at more exotic arrivals. As an example, we generate a 9 s K_α sensitivity kernel for the *P′P′* arrival (also known as *PKPPKP*): a *PKP* phase reflected off the Earth's surface. This phase, which arrives around 2300 s, is nearly unidentifiable in
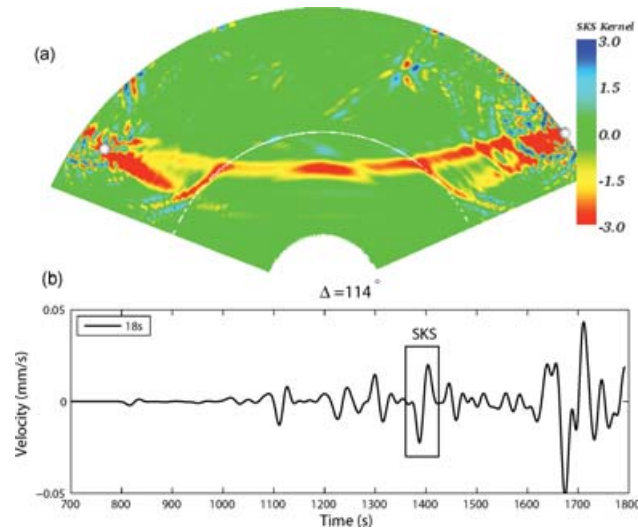
Figure 7. (a) Source–receiver cross-section of the 18 s sensitivity kernel for the *SKS* phase at $\Delta = 114°$. Shown is the K_β kernel in the mantle and the K_α kernel in the outer core. (b) Corresponding radial-component velocity seismogram for a spectral-element simulation accurate at periods of 18 s and longer. The prominent *SKS* phase is labelled.
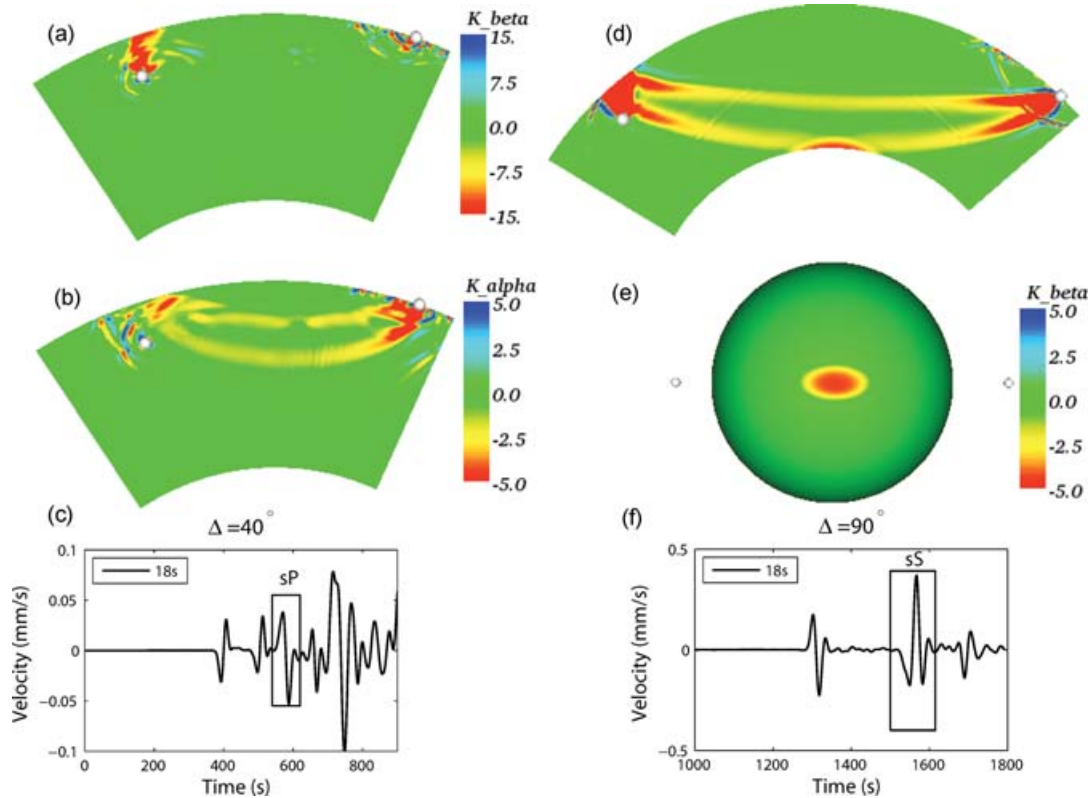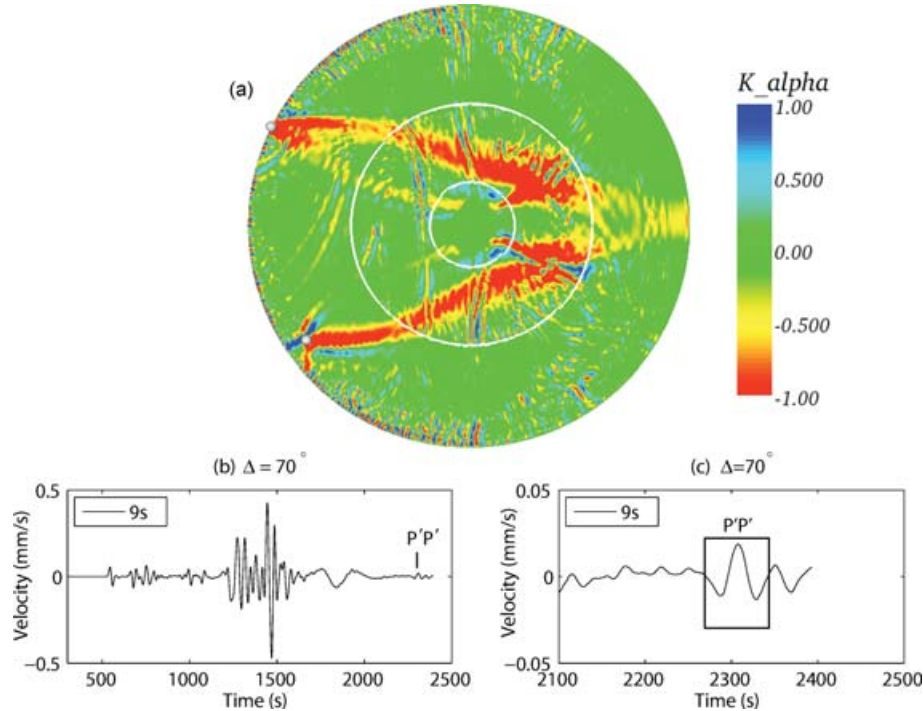

Figure 8. (a) 18 s K_β kernel for the *sP* phase recorded at $\Delta = 40°$. (b) 18 s K_α kernel for the *sP* phase recorded at $\Delta = 40°$. (c) Corresponding vertical-component velocity seismograms for a spectral-element simulation accurate at periods of 18 s and longer. The *sP* phase is labelled. (d) Source–receiver cross-section of the 18 s K_β kernel for the sS phase recorded at $\Delta = 90°$. (e) Map view on the CMB of the same kernel as in (d). (f) Corresponding transverse-component velocity seismograms for a spectral-element simulation accurate down to periods of 18 s. The prominent depth phase *sS* is labelled.

the complete vertical-component seismogram, as shown in Fig. 10(b). However, when the last few hundred seconds of the seismogram are enhanced by a factor of 10, as in Fig. 10(c), we can identify the *P′P′* phase clearly at the tail of the surface waves. Its corresponding K_α kernel is shown in Fig. 10(a), which clearly follows the *PKPPKP* ray path. As discussed in previous sections, since this phase arrives very late in the seismogram, the kernel is relatively 'noisy' due to mantle and core reverberations. Note also the *PKIKPPKIKP* contribution associated with the inner core.
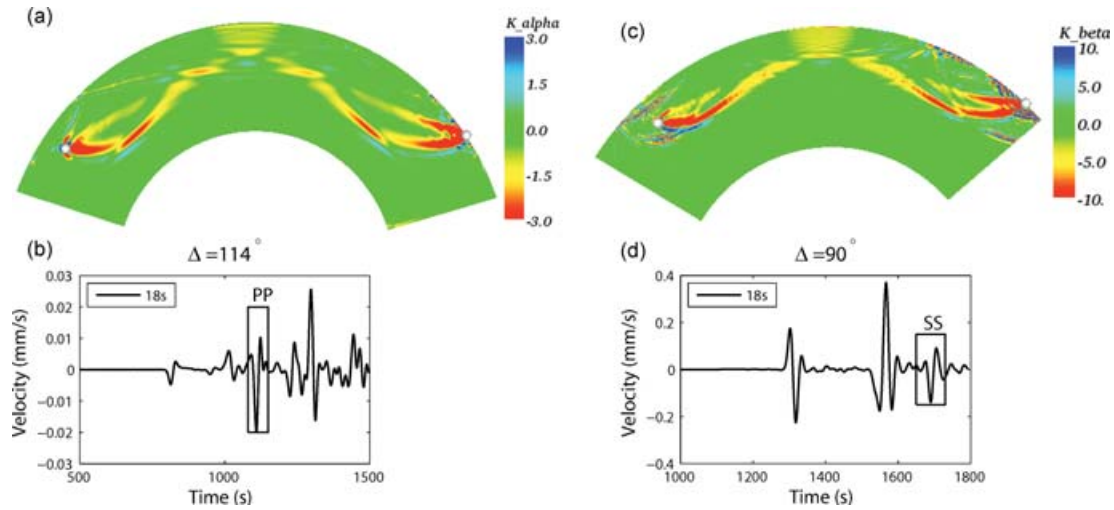
Figure 9. (a) 18 s K_α kernel for the PP phase at a distance of 114°. (b) Corresponding vertical-component velocity seismogram for a spectral-element simulation accurate at periods of 18 s and longer. The *PP* arrival is labelled. (c) 18 s K_β kernel for the *SS* phase at a distance of 90°. (d) Corresponding transverse-component velocity seismogram with the *SS* phase labelled.

Figure 10. (a) 9 s K_α kernel for the $P'P'$ phase recorded at a distance of 70°. (b) Corresponding vertical-component velocity seismogram with the $P'P'$ phase identified. (c) Last 300 s of the seismogram shown in (b) enhanced by a factor of 10.

7.9 Surface wave kernels

We may readily compute finite-frequency sensitivity kernels for surface waves. Since surface waves are more prominent for shallow earthquakes, we selected the 1999 November 26, Vanuatu event, which occurred at a depth of 15 km. Fig. 11(e) shows a vertical-component velocity seismogram with a dominant period of 60 s recorded at an epicentral distance of 120° and an azimuth of 290°, in which the Rayleigh wave is labelled. The corresponding compressional- and shear wave sensitivity kernels are shown in Figs 11(a)–(d). The kernels follow the minor-arc of the source–receiver great circle path on the Earth's surface and are confined to shallow depths in the cross-sections. Also note that the K_β kernel samples much deeper than the K_α kernel for Rayleigh waves. See Sieminski *et al.* (2007a) for an extensive discussion of surface wave finite-frequency sensitivity kernels in isotropic and anisotropic earth models.
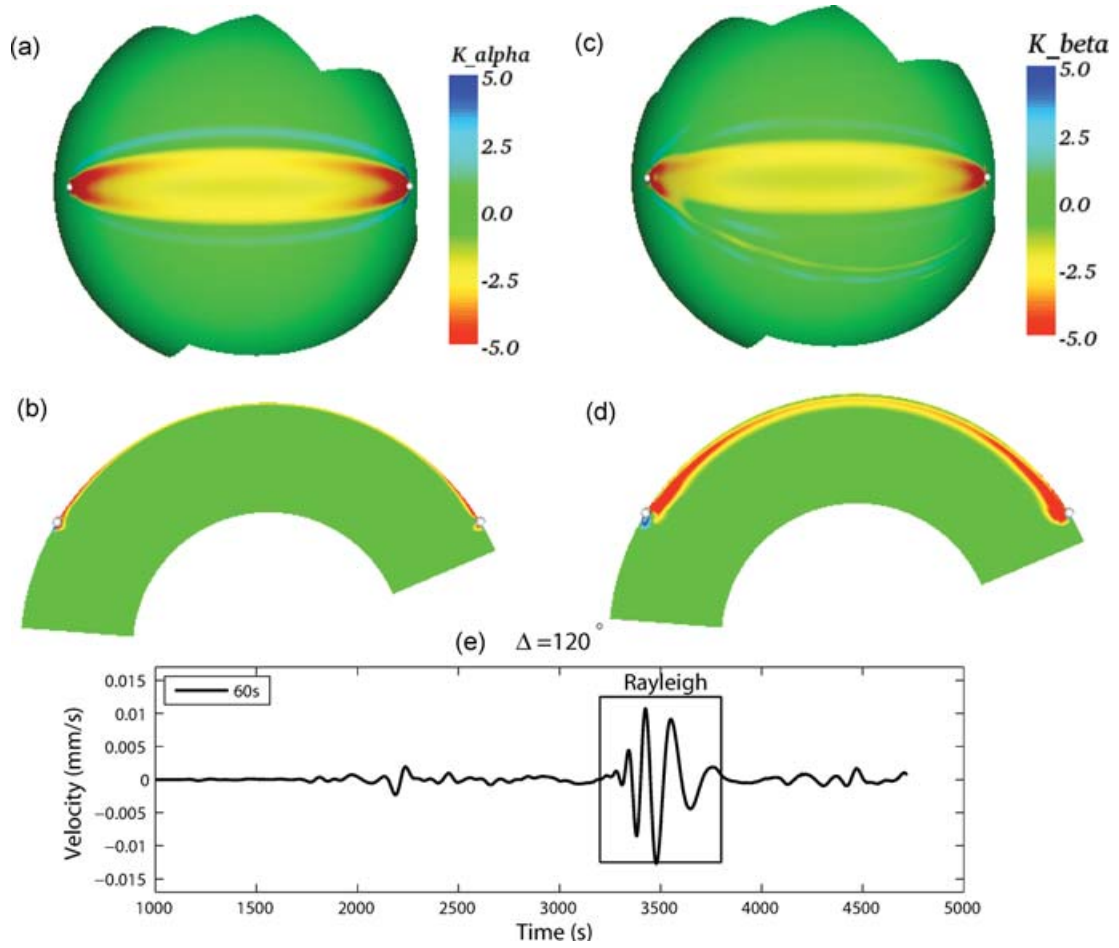
Figure 11. (a) Surface map view of a 60 s K_α kernel for a Rayleigh wave recorded at a distance of 120°. (b) Source–receiver cross-section of the same kernel as in (a). (c) Surface view of the corresponding K_β kernel. (d) Source–receiver cross-section of the same kernel as in (c). (e) Corresponding vertical-component velocity seismogram in which the Rayleigh wave is labelled.

7.10 Boundary kernels

As discussed in Section 3, we can also compute boundary sensitivity kernels, which associate a perturbation in the topography on any internal discontinuity with a corresponding perturbation in the traveltime for a particular phase that interacts with the discontinuity. Following Dahlen (2005), we rewrite the fluid–solid boundary part of expression (57) as

$$\int_{\Sigma_{FS}} \mathbf{K}_d \cdot \boldsymbol{\nabla}^\Sigma \delta \ln d \, \mathrm{d}^2 \mathbf{r} = -\int_{\Sigma_{FS}} (\boldsymbol{\nabla}^\Sigma \cdot \mathbf{K}_d) \delta \ln d \, \mathrm{d}^2 \mathbf{r}, \tag{84}$$

where we have used the 2-D version of the Gauss's theorem (52) and the fact that $[\hat{\mathbf{n}} \cdot \mathbf{s}]_-^+ = 0$ and $[\hat{\mathbf{n}} \cdot \mathbf{s}^\dagger]_-^+ = 0$. Therefore, in the absence of any volumetric model perturbations, the boundary kernel for a fluid–solid boundary (such as the CMB) is given by

$$K_d' = K_d - \boldsymbol{\nabla}^\Sigma \cdot \mathbf{K}_d. \tag{85}$$

The boundary kernels in this section have units of 10^{-4} s km^{-2} when the topographic variations are measured in kilometres.

7.10.1 CMB boundary kernel for ScS

We generate a CMB boundary kernel (Fig. 12c) for the same 18 s *ScS* wave as shown in Fig. 6(e). This boundary kernel exhibits an elliptical 'doughnut' shape around the reflection point and is elongated in the source–receiver direction, with a negative ring of larger amplitudes enclosed by a positive ring of relatively smaller amplitudes (Fig. 12b), as also observed by Dahlen (2005) for the PcP phase. Note that the footprint of the K_β kernel on the CMB (Fig. 12b) is distinctly different from the CMB sensitivity kernel (Fig. 12c), indicating that the traveltime anomaly of boundary-reflected phases is not entirely due to wave speed contrasts across the discontinuity.
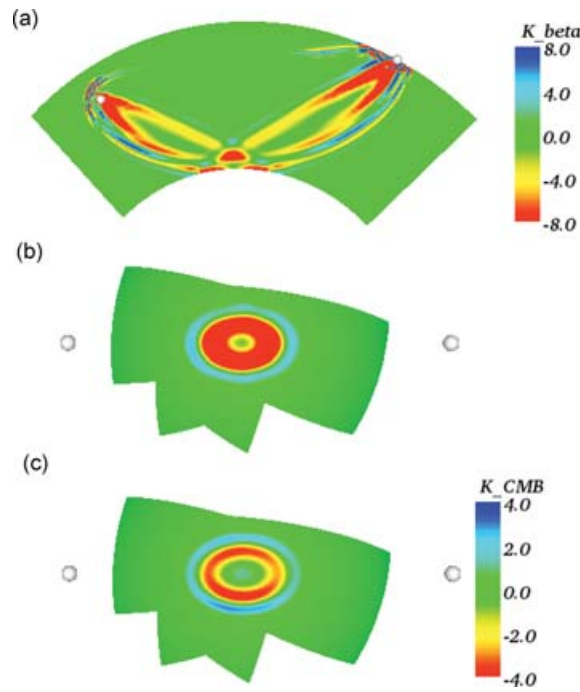
Figure 12. (a) The same 18 s K_β kernel for the *ScS* phase as shown in Fig. 6(b). (b) Map view of the volumetric K_β kernel on the mantle side of the CMB. (c) The CMB boundary kernel of the same *ScS* phase. Red colours indicate that elevated topography on the CMB causes an advance in the arrival time of the *ScS* phase.

7.10.2 670 km discontinuity boundary kernel for S670S

We generate the solid–solid boundary sensitivity kernel (Fig. 13c) for the S670S phase identified in a 9 s transverse velocity record at an epicentral distance of 140° (Fig. 13a). All other *SS* precursors associated with upper-mantle discontinuities in the 1-D PREM model are also clearly visible in the seismogram. The boundary kernel shows the characteristic 'X' shape on the 670-km discontinuity, with a maximum sensitivity at the reflection point, in accordance with the mini-max nature of this arrival (Dahlen 2005).

8 APPLICATIONS TO 3-D MODELS

To study the effects of a 3-D initial model on finite-frequency sensitivity kernels, we use global tomographic model S20RTS (Ritsema *et al.* 1999; Ritsema & Van Heijst 2000) and a source–receiver combination which produces an *S* ray path that samples a significant portion of the African superplume (Fig. 14). To investigate the effects of stronger lateral heterogeneity, we also consider a model obtained by scaling the S20RTS shear-wave speed heterogeneity by a factor of three. The corresponding 9-second-and-longer transverse velocity seismograms for the three models (labelled 1D, 3D and 3D_3x) at an epicentral distance of 95° are shown in Fig. 15(b). Notice the delay of the *S*-wave arrivals for the 3-D models with respect to the 1-D PREM model. The corresponding K_β sensitivity kernels are shown in Figs 15(a), (c) and (e). These kernels look very similar to each other, and their differences when shown on colour scales that are enhanced by a factor of 10 (Fig. 15d) and 3 (Fig. 15f), respectively, are still very minor. Most of the differences concentrate along the outer rim of the original 1-D kernel. Note that S20RTS produces an asymmetric pattern in the 3D_diff and 3D_3x_diff kernels, with the receiver side being a little bit stronger than the source side. Also, as expected, the 3D_3x_diff kernel is about three times stronger than the 3-D_diff kernel, illustrating the variation of the 3-D sensitivity kernels with respect to the initial model.

9 PROSPECTS FOR ADJOINT TOMOGRAPHY

The Fréchet kernels discussed in Sections 7 and 8 are the so-called 'banana–doughnut kernels' (Dahlen *et al.* 2000; Hung *et al.* 2000; Zhou *et al.* 2004). One such kernel relates the traveltime delay of one particular arrival to perturbation in one particular model parameters. For a finite-frequency traveltime inversion, we seek to minimize the following misfit function:

$$\chi(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{N} \left[T_i^{\text{obs}} - T_i(\mathbf{m}) \right]^2, \tag{86}$$

where T_i^{obs} denotes the observed traveltime for the *i*th source–receiver combination, $T_i(\mathbf{m})$ the predicted traveltime based upon the current model m, and *N* the number of traveltime measurements. If we define the traveltime anomaly $\Delta T_i = T_i^{\text{obs}} - T_i(\mathbf{m})$, then the variation of the
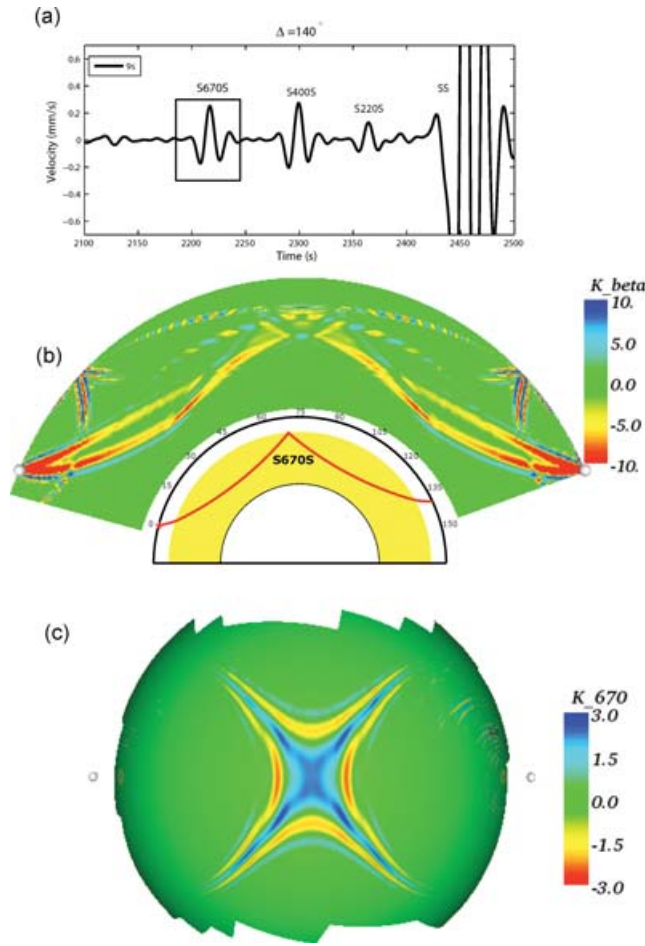
Figure 13. (a) 9-s-and-longer transverse-component velocity seismogram recorded at a distance $\Delta = 140°$ in which the *SS* arrival and the major *SS* precursors are identified. (b) The corresponding K_β kernel for the S670S phase, with its ray path shown for reference in the centre. (c) 670-km boundary sensitivity kernel for the S670S phase, portraying the characteristic 'X' shape.

misfit function (86) becomes

$$\delta\chi = -\sum_{i=1}^{N} \Delta T_i \, \delta T_i. \tag{87}$$

Upon ignoring the boundary kernels in (53) for brevity, we can relate perturbations in traveltime to fractional perturbations in the volumetric model parameters by

$$\delta T_i = \int_V K_i \delta \ln m \, \mathrm{d}^3\mathbf{r}. \tag{88}$$

Now if we define the misfit kernel as

$$K(\mathbf{r}) = -\sum_{i=1}^{N} \Delta T_i \, K_i(\mathbf{r}), \tag{89}$$

and expand our fractional phase-speed perturbations, $\delta \ln m(\mathbf{r})$, into basis functions $B_k(\mathbf{r})$,

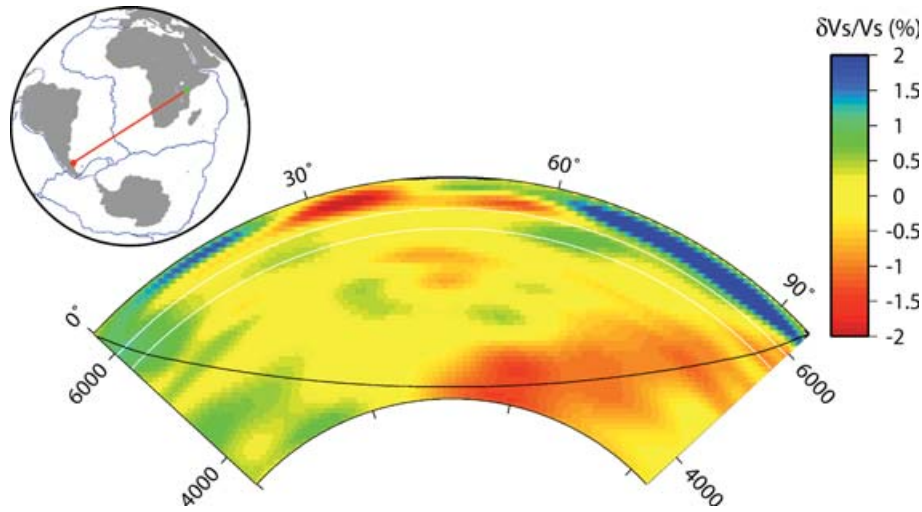$$\delta \ln m(\mathbf{r}) = \sum_{k=1}^{M} \delta m_k \, B_k(\mathbf{r}), \tag{90}$$

where δm_k, $k = 1, \ldots, M$, represent the perturbed model coefficients, then the perturbation in the misfit function can be rewritten as

$$\delta\chi = \sum_{k=1}^{M} g_k \delta m_k. \tag{91}$$

The gradient of the misfit function χ with respect to the kth discrete model parameter m_k is defined as

$$g_k = \frac{\partial \chi}{\partial m_k} = \int_V K \, B_k \, \mathrm{d}^3\mathbf{r}, \qquad k = 1, \ldots, M. \tag{92}$$

Figure 14. Source–receiver pair chosen to generate the *S* sensitivity kernel for 3-D reference model S20RTS. The source and the receiver are indicated by red and green dots in the top-left figure, which also shows the outlines of the continents and major plate boundaries. A source–receiver cross-section through model S20RTS with the geometrical *S* ray path is shown in the middle.
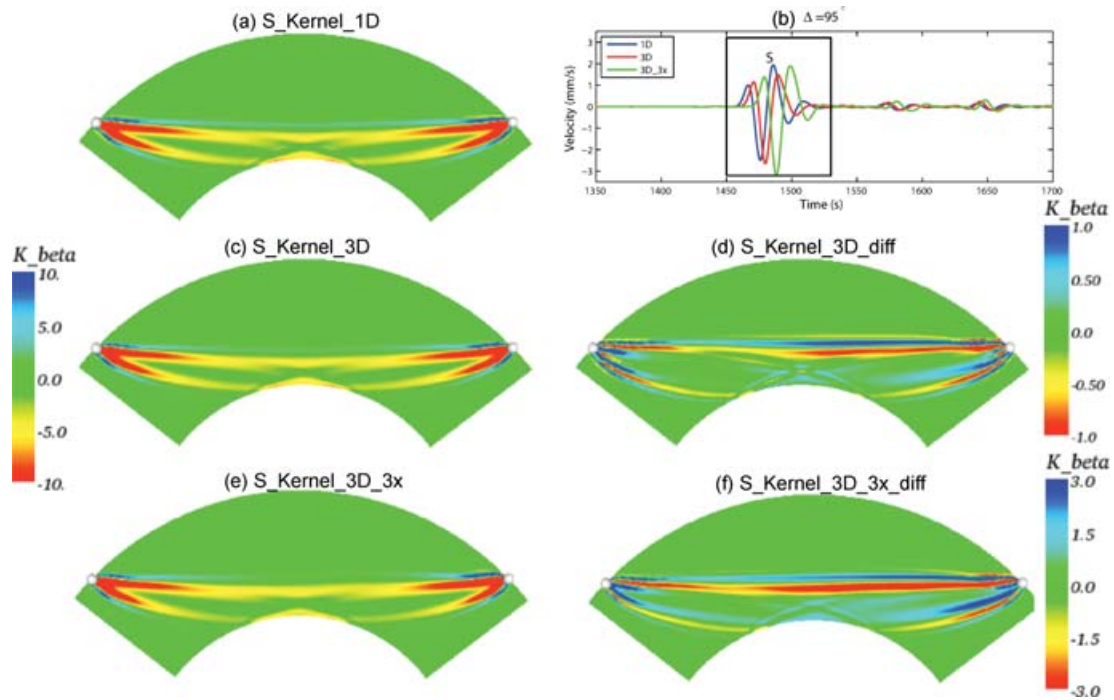
Figure 15. (a) 9 s K_β kernel for the *S* phase at a distance $\Delta = 95°$ for the 1-D PREM model. (b) Transverse-component velocity seismograms for the source–receiver pair described in Fig. 14 for various reference models: 1-D (PREM), 3-D (S20RTS) and 3D_3x (S20RTS enhanced by a factor of 3). (c) Same as (a), but for the 3-D model. (d) The difference between the kernels in (c) and (a). (e) Same as (a), but for the 3D_3x model. (f) The difference between the kernels in (e) and (a). Note that the colour bars for figures (d) and (f) are different from the colour bar for figures (a), (c) and (e).

Notice that to obtain g_k we only need to compute the misfit kernel $K(\mathbf{r})$, not the individual banana–doughnut kernels $K_i(\mathbf{r})$. Therefore, for each earthquake we generate the event kernel by using the adjoint source

$$\mathbf{f}^\dagger(\mathbf{r}, t) = -\sum_{r=1}^{N_r} \Delta T_r \frac{1}{M_r} w_r(T - t) \partial_t \mathbf{s}(\mathbf{r}_r, T - t) \delta(\mathbf{r} - \mathbf{r}_r), \tag{93}$$

where N_r denotes the number of receivers, M_r the normalization factor (83), w_r the time window that contains the arrival of interest, and $\mathbf{r}_r$ the receiver location. Note now, unlike the banana–doughnut adjoint source (82), all receivers simultaneously transmit their predicted time-reversed (three-component) velocity seismogram $\partial_t \mathbf{s}(\mathbf{r}_r, T - t)$ weighted by the traveltime anomaly ΔT_r observed at that particular receiver. This reduces the number of spectral-element simulations required to compute the gradient of the misfit function (53) to two simulations per event: one forward simulation for the current 3-D model, and a second 'adjoint' simulation based upon the adjoint source (93).

Although an expression for the Hessian of the misfit function can be obtained in a similar manner, as shown by Tape *et al.* (2007), individual banana–doughnut kernels $G_{ik} = \int_V K_i B_k \, \mathrm{d}^3\mathbf{r}$ have to be generated, meaning every K_i has to be computed and stored separately, which makes this calculation computationally prohibitive when large numbers of events, source–receiver pairs, and traveltime measurements are involved. Thus the challenge lies in improving the model without access to the Hessian, that is, solely based upon the value of the misfit function and its gradient. Conjugate gradient methods are the natural choice for these types of inverse problems (e.g. Tarantola 2005; Mora 1987, 1988; Luo & Schuster 1991), and have been successfully applied to both source and structural inversions in synthetic experiments (Tromp *et al.* 2005; Tape *et al.* 2007).

10 CONCLUSIONS

The sensitivity of a seismic observable, such as a cross-correlation traveltime measurement, to variations in wave speeds, density and topography on internal discontinuities may be expressed in terms of finite-frequency kernels. The 3-D volumetric kernels for structural model parameters and the 2-D boundary kernels for topographic variations are calculated based upon an adjoint method. We show that for a general 3-D reference model, these kernels may be obtained based upon the interaction between the regular forward wavefield that travels from the source to the receiver, and an 'adjoint' wavefield that travels from the receiver back to the source. The equations that govern the propagation of the adjoint wavefield are obtained based upon a Lagrange multiplier method and involve an earth model with the opposite sense of rotation. The effects of anelasticity are readily accounted for. Construction of the sensitivity kernels requires simultaneous access to the regular forward and adjoint wavefields, which is accomplished by reconstructing the regular forward wavefield backward in time during the calculation of the adjoint wavefield. Thus, there is no need to store the regular wavefield as a function of space and time on disk, because both the regular and adjoint wavefields are carried in computer memory. Chen *et al.* (2007b,a) demonstrate that for a small region one can afford to numerically store the wavefield and construct individual banana–doughnut kernels based upon a scattering-integral method rather than an adjoint method. The storage requirements for global problems currently prohibit such an approach.

We have illustrated the method for various global seismic phases, such as *P, S, ScS*, P_{diff}, *SKS, PKP, sP, sS, PP, SS* and surface waves. The kernels presented in this paper are valid for isotropic model perturbations. Recent work by Sieminski *et al.* (2007a,b) extends the calculation of adjoint sensitivity kernels to anisotropic surface and body waves.

The Fréchet kernels presented in this paper will form the basis of future 3-D 'adjoint' tomographic inversions. In adjoint tomography we have access to the value of the misfit function and its derivative, but not its second derivative or Hessian. In classical traveltime tomography one generally has access to both the gradient and the Hessian of the misfit function, and thus one can use a so-called Newton method to solve the inverse problem. When only the gradient of the misfit function is available, one needs to resort to a gradient method, for example, steepest decent or conjugate gradient methods (Mora 1987, 1988; Luo & Schuster 1991; Akcelik *et al.* 2002, 2003; Tarantola 2005).

The main advantages of our adjoint approach are: (1) the kernels are calculated on-the-fly by carrying the adjoint and regular forward wavefields in memory at the same time; (2) the kernels may be calculated for fully 3-D reference models; (3) the approach scales linearly with the number of earthquakes but is independent of the number of receivers and the number of arrivals that are used in the inversion; (4) any time segment where the data and the synthetics have significant amplitudes and match reasonably well is suitable for a measurement, and one does not need to be able to label the phase because the adjoint simulation will reveal how this particular measurements 'sees' the earth model and (5) the cost of the simulation is independent of the number of model parameters, that is, one can consider fully anisotropic earth models with 21 elastic parameters for practically the same numerical cost as an isotropic simulation involving just two parameters.

The software used to perform the spectral-element simulations shown in this article is freely available via http://www.geodynamics.org/.

REFERENCES

Akcelik, V., Biros, G. & Ghattas, O., 2002. Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation, in *Proceedings of the ACM/IEEE SC2002 Conference*, Baltimore, MD, USA, Nov 2002.

Akcelik, V. *et al.*, 2003. High-resolution forward and inverse earthquake modeling on terascale computers, in *Proceedings of the ACM/IEEE SC2003 Conference*, Phoenix, AZ, USA, Nov 2003.

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology, Theory and Methods*, W.H. Freeman, San Francisco, ISBN:0716710595.

Chaljub, E., Capdeville, Y. & Vilotte, J.P., 2003. Solving elastodynamics in a fluid–solid heterogeneous sphere: a parallel spectral element approximation on non-conforming grids, *J. Comput. Phys.,* **187,** 457–491.

Chavent, G., Dupuy, M. & Lenmonier, C., 1975. History matching by use of optimal control theory, *Soc. Petrol. Eng. J.,* **15**(1), 74–86.

Chen, P., Jordan, T.H. & Zhao, L., 2007a. Full 3D waveform tomography: a comparison between the scattering-integral and adjoint-wavefield methods, *Geophys. J. Int.,* **170,** 175–181.

Chen, P., Zhao, L. & Jordan, T.H., 2007b. Full 3D tomography for crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97,** 1094–1120.

Dahlen, F.A., 2005. Finite-frequency sensitivity kernels for boundary topography perturbations, *Geophys. J. Int.,* **162,** 525–540.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton University Press, Princeton.

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

Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference Earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51,** 1387–1403.

Griewank, A. & Walther, A., 2000. Revolve: an implementation of checkpointing for the reverse or adjoint mode of differentiation, *ACM Trans. Math. Software,* **26**(1), 19–45.

Hinze, M., Walther, A. & Sternberg, J., 2005. An optimal memory-reduced procedure for calculating adjoints of the instationary Navier-Stokes equations, *Opt. Contl. Appl. Methods,* **27**(1), 19–40.

Hung, S.-H., Dahlen, F.A. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes – II. Examples, *Geophys. J. Int.,* **141,** 175–203.

Ji, C., Tsuboi, S., Komatitsch, D. & Tromp, J., 2005. Rayleigh-wave multipathing along the West coast of North America, *Bull. seism. Soc. Am.,* **95,** 2115–2124.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.,* **139,** 806–822.

Komatitsch, D. & Tromp, J., 2002a. Spectral-element simulations of global seismic wave propagation – I. Validation, *Geophys. J. Int.,* **149,** 390–412.

Komatitsch, D. & Tromp, J., 2002b. Spectral-element simulations of global seismic wave propagation – II. 3-D models, oceans, rotation, and self-gravitation, *Geophys. J. Int.,* **150,** 303–318.

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

Komatitsch, D., Liu, Q., Tromp, J., Süss, P., Stidham, C. & Shaw, J.H., 2004. Simulations of ground motion in the Los Angeles basin based upon the Spectral-Element method, *Bull. seism. Soc. Am.,* **94,** 187–206.

Liu, Q. & Tromp, J., 2006. Finite-frequency Kernels based upon adjoint method, *Bull. seism. Soc. Am.,* **96**(6), 2383–2397.

Luo, Y. & Schuster, G., 1991. Wave-equation traveltime inversion, *Geophysics,* **56,** 645–653.

Marquering, H., Dahlen, F. & Nolet, G., 1999. Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.,* **137,** 805–815.

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

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, **52**(9), 1211–1228.

Mora, P., 1988. Elastic wave-field inversion of reflection and transmission data, *Geophysics,* **53**(6), 750–759.

Restrepo, J.M., Leaf, G.K. & Griewank, A., 1998. Circumventing storage limitations in variational data assimilation studies, *SIAM J. Sci. Comput.,* **19**(5), 1586–1605.

Ritsema, J. & Van Heijst, H.J., 2000. Seismic imaging of structural heterogeneity in Earth's mantle: evidence for large-scale mantle flow, *Sci. Prog.,* **83,** 243–259.

Ritsema, J., Van Heijst, H.J. & Woodhouse, J.H., 1999. Complex shear velocity structure imaged beneath Africa and Iceland, *Science,* **286,** 1925–1928.

Ritsema, J., Rivera, L.A., Komatitsch, D., Tromp, J. & van Heijst, H.J., 2002. The effects of crust and mantle heterogeneity on PP/P and SS/S amplitude ratios, *Geophys. Res. Lett.,* **29,** 72-1–72-4.

Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007a. Finite-frequency sensitivity of surface waves to anisotropy based upon adjoint methods, *Geophys. J. Int.,* **168,** 1153–1174.

Sieminski, A., Liu, Q., Trampert, J. & Tromp, J., 2007b. Finite-frequency sensitivity of body waves to anisotropy based upon adjoint methods, *Geophys. J. Int.,* **171,** 368–389.

Talagrand, O. & Courtier, P., 1987. Variational assimilation of meteorological observations with the adjoint vorticity equatuation. I: theory, *Q. J. R. Meteorol. Soc.,* **113,** 1311–1328.

Tape, C., Liu, Q. & Tromp, J., 2007. Finite-frequency tomography using adjoint methods—methodology and examples using membrane surface waves, *Geophys. J. Int.,* **168,** 1105–1129.

Tarantola, 1984. Linearized inversion of seismic reflection data, *Geophys. Prospect.,* **32,** 998–1015.

Tarantola, A., 2005. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation,* Society for Industrial and Applied Mathematics (SIAM), Philadelphia.

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

Tsuboi, S., Komatitsch, D., Ji, C. & Tromp, J., 2003. Spectral-element simulations of the November 3, 2002, Denali, Alaska earthquake on the Earth simulator, *Phys. Earth planet. Inter.,* **139,** 305–313.

Tu, T., O'Hallaron, D.R. & Lopez, J., 2003. The Etree library: a System for manipulating large Octrees on disk, *Technical Report CMU-CS-03-174.*

Zhao, L. & Jordan, T.H., 2006. Structure sensitivities of finite-frequency seismic waves: a full-wave approach, *Geophys. J. Int.,* **165,** 981–990.

Zhao, L., Jordan, T.H. & Chapman, C.H., 2000. Three-dimensional Frèchet differential kernels for seismic delay times, *Geophys. J. Int.,* **141,** 558–576.

Zhao, L., Jordan, T.H., Olsen, K.B. & Chen, P., 2005. Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Geophys. J. Int.,* **95,** 2066–2080.

Zhou, Y., Dahlen, F.A. & Nolet, G., 2004. 3-D sensitivity kernels for surface-wave observables, *Geophys. J. Int.,* **158,** 142–168.

Zhou, Y., Nolet, G., Dahlen, F.A. & Laske, G., 2005. Global upper-mantle structure from finite-frequency surface-wave tomography, *J. geophys. Res.,* **111,** doi:10.1029/2005JB003677.