Geophys. J. Int. (2004) doi: 10.1111/j.1365-246X.2004.02509.x

On sensitivity kernels for 'wave-equation' transmission tomography

Maarten V. de Hoop* and Robert D. van der Hilst

Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Rm 54-517A, *Cambridge,* MA 02139, *USA.*
E-mail: mdehoop@dix.mines.edu

Accepted 2004 October 15. Received 2004 October 8; in original form 2004 July 21

SUMMARY

We combine seismological scattering theory with the theory of distributions to study some properties of sensitivity kernels for finite frequency seismic delay times. The theory to be used for calculating the kernels depends on the way the measurements are made. For example, the sensitivity to the traveltime proper, that is the time to instantaneous onset of the phase arrival as defined in geometrical ray theory, has a non-zero value on the ray and zero elsewhere, whereas measurements of the smooth later part of the wave excitation can have more complicated kernels. Analysis based upon the Born approximation reveals that the behaviour of such kernels is determined by the regularization of the nth derivative of the Dirac delta supported on the unperturbed source-receiver ray, where n is the spatial dimension. Such regularization induces approximately the desired band-limitation and the associated Fresnel zone averaging. If the regularization is symmetric about the singular support of the delta its odd order derivatives vanish on this support, but the even orders render non-zero values. This explains why 3-D finite frequency kernels for body wave delay times seem—in specific, rather restrictive circumstances—to have zero sensitivity on the ray (as in the so called 'banana-doughnut kernels'). Indeed, a regularization of the kernel can have a zero on the unperturbed ray, in the absence of caustics (that is, the medium must be simple or quasi-homogeneous) and only if the exact source signature (e.g., source time function) is known and used; both conditions can be met in synthetic cases but not—in general—in applications to real data and heterogeneous media. In general, one will not know where the oscillatory kernels have zero values, which has implications for the use of such kernels for the linearization of transmission tomography. We briefly clarify these observations by invoking a multiresolution analysis of sensitivity kernels, connected with a time-frequency analysis of phase arrival times.

Key words: body wave traveltimes, sensitivity kernels, time correlation, wave-equation tomography.

1 INTRODUCTION

Traveltimes of seismic body waves play a key role in many seismological studies of the Earth's interior structure. For instance, tomographic inversions of vast amount of routinely processed travel time residuals from international data centres, such as the International Seismological Centre (ISC), have been used to delineate 3-D heterogeneity in rather spectacular detail (Inoue *et al.* 1990; Fukao *et al.* 1992; Su & Dziewonski 1997; Van der Hilst *et al.* 1997; Kennett *et al.* 1998; Bijwaard *et al.* 1998; Kárason & Van der Hilst 2000). These studies use geometrical ray theory (GRT) – asymptotically an infinitely high-frequency approximation—and project traveltime data on to a space of velocity models along infinitesimally narrow ray paths, in accordance with the (ray) geometrical sensitivity kernel. For these data, the use of GRT is justified because the instantaneous onset is picked at relatively short periods (about 1 Hz) and the seismic wavelengths are shorter than the scale lengths of heterogeneity in the resulting images of velocity.

Since the advent of routine digital recording, however, an increasing number of studies have begun to rely on traveltime information gleaned from the later parts of the wave excitation, for instance by measuring the maximum of the time-domain cross correlation of low- or bandpass filtered waveforms (VanDecar & Crosson 1991; Woodward & Masters 1991; Grand 1994). In some studies the lower frequency nature of such measurements is acknowledged by using GRT along with a parametrization and regularization that smoothes the model solution over larger spatial length scales. However, with techniques that do not measure the instantaneous onset of the phase arrival (i.e. the 'wave front' as implied in GRT) the applicability of (ray) geometrical kernels is no longer obvious.

For the proper interpretation of such data by seismic inversion one should consider 'finite-frequency' sensitivity kernels (Woodward

*On leave from: The Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA.

1992; Katzman *et al.* 1998; Dahlen *et al.* 2000; Zhao *et al.* 2000; Kárason & Van der Hilst 2001; Spetzler *et al.* 2001). Proper use of realistic sensitivity kernels in 'wave equation' tomography would (i) help extract more information from seismic records since measurements made at different frequencies can be interpreted correctly; (ii) allow internally consistent 'data fusion', that is, the integration and joint interpretation of data measured at different frequencies (such as dispersion of body- and multi-mode surface waves, direct and surface reflected body waves, short-period *P* and long-period *S* waves); and (iii) enable lower frequency data to constrain smoothly varying structures without preventing higher frequency data from resolving smaller-scale structure in regions of dense data coverage.

Several investigators have estimated 2-D and 3-D sensitivity kernels to describe the linear (integral) relationship between finite frequency observations and model perturbations, either by surface wave mode summation (Marquering *et al.* 1998), normal mode summation (Li & Romanowicz 1996; Katzman *et al.* 1998; Zhao *et al.* 2000; Kárason & Van der Hilst 2001), or ray theory (Dahlen *et al.* 2000). A general characteristic of the associated kernels is that the sensitivity is not confined to infinitesimally narrow paths but distributed over a finite volume, related to the Fresnel volume of the wave, the spatial extent of which typically increases with decreasing frequency. How exactly the sensitivity is distributed over these volumes has been debated, and some aspects have remained enigmatic and, perhaps, counterintuitive. One must realize that the precise behaviour of such kernels would largely be of academic interest if they are projected on sufficiently coarse basis functions and when severe damping (regularization) is needed to obtain stable solutions. Indeed, differences between results of global traveltime inversion based either on geometrical ray or on finite frequency theory have, so far, been small (Kárason, van der Hilst & Masters, private communication, 2001; Montelli *et al.* 2004) and—probably—less significant than the effects on the images of uneven spatial and spectral data coverage, data quality, parametrization, and regularization. However, the use of finite frequency kernels and wave equation tomography is likely to gain in importance because of the increased need for internally consistent data fusion, for the extraction and interpretation of new signal from broad-band data, and for a more accurate estimation of the magnitude of variations in elastic parameters.

Inspired by previous work (Woodward 1992; Li & Romanowicz 1996; Katzman *et al.* 1998; Dahlen *et al.* 2000; Zhao *et al.* 2000) we analyse fundamental properties of sensitivity kernels for wave equation transmission tomography. Generically, the sensitivity is oscillatory in cross-section perpendicular to the unperturbed source-receiver ray, and the value on this ray depends on several theoretical approximations and practical considerations. First, it is important to realize that different measurements give rise to different kernel behaviour, and consistency between measurement, mismatch criterion, and kernel calculation is essential. Secondly, the behaviour of the sensitivity kernel depends on source parameters (e.g. the moment rate function), geometrical amplitude factors, the dimensionality of the problem, the magnitude and scaling properties of heterogeneity (e.g. the presence of caustics), and the frequency bandwidth. Thirdly, whereas the cross correlation criterion may work very well for propagating delta pulses (which represent the high-frequency component), it can break down when more general wave fields are considered (Hörmann & de Hoop 2002). Also, from an observational point of view, the cross correlation of broad-band wavefields can be ambiguous, and—in general—the maximum of the cross correlation will not match the wave front arrival; the cross correlation criterion does not imply the match of waveforms either.

We show that the so called 'banana-doughnut'-like feature (Dahlen *et al.* 2000) – that is, the 'sensivity' being identical zero on the unperturbed source-receiver ray—can indeed appear in the regularization of 3-D sensitivity kernels but only if the measurement is derived from the maximum of the time-domain waveform cross correlation, if there are no caustics (i.e. the medium must be simple or quasi-homogeneous), and if the actual source signature (moment rate function) is exactly the same as the one used for modelling. The latter two conditions are typically only met in the framework of synthetic experiments. We will show, however, that the kernel itself does not have a zero on the unperturbed ray.

With finite frequency sensitivity kernels for wave equation transmission tomography one can, in principle, account for different scales in the data (frequency) and in the model (wavenumbers). On the one hand, advanced wave propagation theory and new data analysis can combine to create new abilities for extracting and interpreting more information from a broad band record, thus enhancing the effective spatial and spectral data coverage. On the other hand, the ability to account for finite frequency effects is of particular interest and importance for the joint interpretation of data measured at different frequencies (and which may, thus, probe different spatial length scales). As an alternative to the conventional technique of finding the optimum of a cross correlation, we propose the use of a multi resolution analysis of the cross correlation. A full analysis will be presented elsewhere (Douma *et al.* 2004, in preparation). Here, we briefly illustrate a multi resolution decomposition of the wave equation sensitivity kernel and how it relates to the traveltime sensitivity kernel.

2 WAVE PROPAGATION IN SMOOTHLY VARYING MEDIA

To provide a framework for a joint analysis of traveltime and wave-equation tomography, we introduce in this section the Green's functions and their Fréchet derivatives for high-frequency wave propagation.

2.1 High-frequency body waves

In seismic tomography one aims to estimate and interpret phase arrivals in the broad-band three-component wavefield. The full wave excitation by an earthquake can be written as a sum over normal modes. For body-wave traveltime tomography, however, there is no need to use the full mode expansion because traveltime information is contained in the singular (non-smooth) part of the wave excitation. This singular part allows a representation in terms of Maslov asymptotic theory (Kendall & Thomson 1993). Hence, we can use asymptotic methods to analyse and compute wave equation kernels.

With G the Green's tensor, such representation can be written in the form of an oscillatory integral for each mode (e.g. qP, S1, ...) of propagation (Taylor 1975):[1]

$$G(x,t,x_0) = \sum_i \int_{\mathbb{R}^{N^{(i)}}} a^{(i)}(x,t,x_0,\theta) \exp\left[\mathrm{i}\phi^{(i)}(x,x_0,t,\theta)\right] \mathrm{d}\theta, \tag{1}$$

where the $\phi^{(i)}$ are phase functions, the $a^{(i)}$ are amplitude functions, and $N^{(i)}$ denotes the required number of phase variables. Away from caustics, $N^{(i)} = 1$, $\theta = \omega$, and $\phi^{(i)}$ takes the form $\omega[T^{(i)}(x, x_0) - t]$ with $T^{(i)}(x, x_0)$ denoting the (possibly multi-valued) traveltime

[1]To simplify notation we suppress the factor $(2\pi)^{-\frac{N^{(i)}}{2}-\frac{2n+1}{4}}$; in fact, in (1) $a^{(i)}$ contains a factor $(2\pi)^{1/4}$.

between x and x_0 on branch i. Hereinafter we will suppress $\sum_i$ and focus on a particular branch. The oscillatory integral (1) is a kernel of a Fourier integral operator that solves the elastic wave equation. Away from caustics, the leading order amplitude induces the representation

$$G(x,t,x_0) \sim a_0(x,x_0)\mathcal{H}^\sigma \delta^{(m)}[t - T(x,x_0)], \sigma = \sigma(x,x_0), \tag{2}$$

where

$$\mathcal{H}^\sigma \delta^{(m)}(t) = \frac{1}{\pi}\mathrm{Re}\int_0^\infty (-\mathrm{i}\omega)^m \exp\left[-\mathrm{i}(\omega t - \tfrac{1}{2}\pi\sigma)\right] \mathrm{d}\omega. \tag{3}$$

In these expressions, $\mathcal{H}$ represents the Hilbert transform, which accounts for passage through a caustic ($\mathcal{H}^2 = -1$)), σ—referred to as the KMAH index—accounts for the number and types of caustics, $m = (n-3)/2$, and a_0 contains the dyadic product of polarization vectors evaluated at the endpoints of the ray that connects x with x_0 (these vectors define the radiation pattern, which for simplicity is suppressed in our notation Burridge *et al.* 1998, eqs 3.1, 3.12).

For our analysis we consider the location (in space and time) and orientation of wave fronts, or, formally speaking, the wave front set WF(G) of G. Given the oscillatory integral representation (1), the inclusion (Hörmander 1990, theorem 8.1.9)

$$\begin{aligned}\mathrm{WF}(G) \subseteq \{&[x,t;\partial_x\phi(x,x_0,t,\theta),\\ &\partial_t\phi(x,x_0,t,\theta)] \mid \partial_\theta\phi(x,x_0,t,\theta) = 0\}, \phi = \phi^{(i)}\end{aligned} \tag{4}$$

reveals the ray geometry associated with G. That is, there is a wave front at position x and time t if at that location the phase is stationary ($\partial_\theta\phi(x,x_0,t,\theta) = 0$); the direction of the ray at (x,t) on this wave front is determined by $[\partial_x\phi(x,x_0,t,\theta), \partial_t\phi(x,x_0,t,\theta)]$. Away from caustics, $\theta = \omega$. Then $\partial_\theta\phi = 0$ implies $t = T^{(i)}(x,x_0)$ while $\partial_t\phi = -\omega$ and $\omega^{-1}\partial_x\phi$ becomes the slowness vector $(\partial_x T^{(i)})(x,x_0)$ at x. In fact, $T^{(i)}$ solves the eikonal equation.

With the aid of the Green's tensor, the displacement u due to a body force f can be written in the form

$$u(x,t) = \int G(x,.,x_0) \overset{(t)}{*} f(x_0,.)\,\mathrm{d}V(x_0). \tag{5}$$

We assume a point body force located at s and with origin time $t_s = 0$: $f(x_0,t) = e\,\delta(x_0 - s)W_\varepsilon(t)$, where $W_\varepsilon(t)$ is the source signature and ε a parameter that controls the (frequency) band limitation, with $\varepsilon \to 0$ representing the broad-band (infinite bandwidth) limit. We discuss later the implications of this band limitation on the behaviour of the sensitivity kernels (for instance Sections 4.1–4.2). Furthermore, we characterize an earthquake by a moment tensor M of rank *2*, so that

$$f(x_0,t) = -M \cdot \partial_{x_0}\delta(x_0 - s)H(t),$$

where $H(t)$ is the Heaviside function (the idealized time-rise function). Substituting this equation into (5), and integrating by parts in x_0, produces again a leading order of $\delta^{(m)}$, while in (2) a_0 attains an additional slowness vector from $\partial_{x_0}T$ and a contraction with M; W_ε becomes the time derivative of the time-rise function.

With the obvious restriction that the wavefield u is recorded at Earth's surface, (5) defines a mapping from the source f to the data u. In practice, measurements are taken over some time interval (t_0, t_1) at discrete receiver locations (stations) x; hereinafter $u_r(t)$ denotes the time series at station position r. In fact, the transmission tomography problem discussed in this paper is naturally coupled to the inverse problem of estimating f (i.e. M and W_ε) from $u_r(t)$.

In combination, eqs (1) and (5) express how to generate an approximate solution of the wave equation from the solutions to the ray tracing equations. Thus, these equations provide the natural framework for a joint analysis of traveltime tomography and wave-equation tomography.

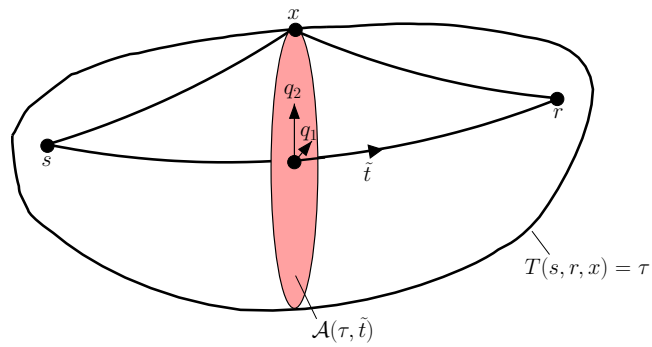

Figure 1. Isochrons, and support and singular support (unperturbed source-receiver ray) of sensitivity kernel. (The kernel is non-smooth on its singular support.) Also indicated is the notation for ray perturbation used in Appendix A. (The medium has a constant gradient.)

Throughout this paper we exclude rays with varying (shear-wave) multiplicities, but expanding the theory to include them is possible.

2.2 Fréchet derivatives

We follow (Dahlen *et al.* 2000) and investigate the Fréchet[2] or directional derivative of G or u with respect to medium parameters (stiffness tensor) $c = c_{ijk\ell}$ using the Born approximation. We note that this approximation is adequate in the context of smoothly varying medium parameters, but other approximations need to be adopted in the presence of sharp transitions and boundaries, for example to account for post-critical phenomena such as head waves (Červený & Ravindra 1971).

Away from caustics the high-frequency Born approximation for the perturbed Green's tensor δG yields for a particular mode pair[3]

$$\begin{aligned}\delta G = -\int\int &A(s,r,t,x,\omega)\Omega(s,r,x)\delta c(x)(-\omega^2)\\ &\times \exp\Big[\mathrm{i}\omega(\underbrace{T(r,x) + T(x,s)}_{T(s,r,x)} - t)\Big]\,\mathrm{d}\omega\,\mathrm{d}V(x),\end{aligned} \tag{6}$$

where $-\omega^2$ enters via the second-order time derivative in the wave operator (see, e.g. De Hoop *et al.* 1999; this equation corresponds to Dahlen *et al.* 2000, eqs 37 and 56). The ray geometry associated with (6) is illustrated in Fig. 1. The medium perturbation δc is density normalized, smooth, and supported (non-zero) in a ball of finite radius. The radiation patterns $\Omega(s,r,x)$ are a dyadic product of incoming (s to x) and outgoing (x to r) slowness and polarization vectors. $\Omega(s,r,x)\,\delta c$ denotes a tensor contraction of radiation patterns and the components of δc (Burridge *et al.* 1998, eqs 3.29 and 3.30). The leading order can be written in the form

$$\begin{aligned}\delta G \sim -\int &\underbrace{A_0(s,r,x)}_{\rho(x)a_0(r,x)a_0(s,x)}\ \Omega(s,r,x)\mathcal{H}^\sigma\delta^{(2m+2)}[t - T(s,r,x)]\\ &\times \delta c(x)\,\mathrm{d}V(x), \sigma = \sigma(r,x) + \sigma(x,s),\end{aligned}$$

with $2m + 2 = n - 1$, and where ρ denotes mass density.

[2]In the mathematical literature this derivative is known as the Gateaux derivative—named after Gaston Gateaux.

[3]To simplify notation we suppress the factor $(2\pi)^{-\frac{3n-1}{4}-\frac{1}{2}}$; in fact, in (6) A contains a factor $(2\pi)^{-(n-1)/4}$.

3 TRAVELTIME DATA FROM WAVEFORM CROSS CORRELATION

From an inverse problems point of view, the objective of seismic tomography is to find an Earth model whose predictions (i.e. modelled data) match—within acceptable error—the observations considered. In order words, a solution is acceptable if the observed data are contained in the range of the modelling operator, that is, the mapping from f to u, mentioned at the end of Section 2.1. In seismic reflection tomography this principle underlies the use of differential semblance (in the absence of caustics) or data annihilators (allowing the presence of caustics) criteria (Symes & Carazzone 1991; Stolk & De Hoop 2004).

In traveltime transmission tomography one aims to match wave front sets, that is, the location and orientation of wave fronts, in modelled and observed wavefields. A common tool used in the development of a mismatch criterion is the cross correlation of the observed and modelled data. This development can be carried out in several ways, each having ramifications for the sensitivity kernels that are to be used for inverting the resulting measurements. For instance, using broad-band waveform data, a wavelet transform applied to the cross correlation would reveal the offset between observed and modelled traveltimes. Alternatively, for narrow-band data one has resorted to using the maximum of the cross correlation as an estimate for offset in 'finite-frequency' (i.e. narrow-band) traveltimes. It should be noted, however, that the latter approach does not necessarily lead to matching either the full waveforms or the wave front sets; this implies that the information yielded by the narrow-band maximum criterion is fundamentally different from the traveltime offset implied in ray theory (where the criterion is the phase of the instantaneous onset, or wave front, of the arrival).

In order to better understand the meaning and behaviour of sensitivity kernels we will explore in detail some properties of correlations and mismatch criteria.

3.1 The time-correlation function

We let u represent an unperturbed and v a perturbed wavefield, the perturbation being generated by a perturbation of the elastic properties (such as density normalized stiffness) of the Earth's interior. Assuming that the cross correlation between u and v can be taken,[4] we let

$$C(t) = \int_{\mathbb{R}} u_r(t')v_r(t+t')\mathrm{d}t'.$$

In the framework of traveltime tomography, one aims to develop a procedure that would determine from such a correlation δt, the projection on to the time axis of the offset between wave front sets of u and v at r.

3.2 The maximum of the correlation

Luo & Schuster (1991) describe a traveltime inversion that uses optimal fitting of traveltimes from synthetic seismograms according to wave equation solutions of (elastic) model perturbations. The fitting criterion is based upon the time domain cross-correlation of the observed (u^{obs}) and the synthetic (u) seismic data. The validity of this approach is usually taken for granted, but it is important to realize that it is only justified in certain circumstances (Hörmann & de Hoop 2002).

We give a brief schematic description of this fitting strategy. Let u and u^{obs} represent the modelled and observed wavefields, respectively. Here, $u = u[\gamma]$ is calculated from a model that is parametrized by the variable $\gamma(x)$; one can think of $\gamma(x)$ as a point on a curve in the space of density normalized stiffness tensor models $c(x)$. The correlation function is then dependent on both t and γ, which we indicate in the notation

$$C(t)[\gamma] = \int_{\mathbb{R}} u_r[\gamma](t')u_r^{\mathrm{obs}}(t+t')\,\mathrm{d}t',$$

where $(\ldots)$ denotes the scalar and $[\ldots]$ the functional argument of C. One would expect that the exact traveltime shift induced by a perturbation between u_r^{obs} and u_r yields an optimum match (overlap) of their corresponding seismogram windows and, therefore, the cross correlation should be maximal.

We hence introduce the derivative

$$F(t)[\gamma] = \partial_t C(t)[\gamma]. \tag{7}$$

Leaving possible maxima at time interval boundaries aside, we search a (γ, t) relation for which the cross correlation is stationary, i.e.

$$F(\Delta t)[\gamma] = 0, \tag{8}$$

with the condition of matching,

$$\Delta t = \Delta t[\gamma](u^{\mathrm{obs}}) = 0.$$

We can linearize the above relationship by considering $F(t)[\gamma]$ to be an implicit definition of a functional relationship between γ and t. Under the condition that $\partial_t F = \partial_t^2 C \neq 0$ we would thus solve eq. (8) locally for δt, the estimate of δt, as a function of γ and find a quasi-explicit representation by

$$\partial_\gamma \Delta t = -\frac{\partial_\gamma F(0)}{\partial_t F(0)}, \quad \text{or} \quad \delta t = -\frac{\delta F}{\partial_t F}. \tag{9}$$

The behaviour of sensitivity kernels as discussed by Dahlen *et al.* (2000) is a direct consequence of eq. (9)—or Dahlen *et al.* (2000, eq. 66)—which is assumed to follow from eq. (7).

3.3 The wave front set of the correlation

The reasoning followed in the previous subsection is justified if u and u^{obs} can be represented as propagating delta waves, as in eq. (2). Delta waves are representative of the high frequency (short period) part of the data. Indeed, eq. (7) aims at detecting the singular support (the projection of the wave front set of C on the time axis) of the correlation C. With techniques from microlocal analysis, however, one can design a scaled smoothing operator that enables the detection of the sought time shift from more general wavefields or broad-band data without the need to use eq. (9). We explain this next.

We let φ be the Gaussian in one dimension and define

$$\psi_{p,t}(q) = \frac{1}{p}\varphi\left(\frac{q-t}{p}\right). \tag{10}$$

Here, p plays the role of scaling parameter, much like the parameter ε introduced below eq. (5) and used below in eq. (17). Let $\mathcal{F}$ denote the Fourier transform and $\mathcal{F}C$ the Fourier transform of C. We introduce

[4]For $\tau \in \mathbb{R}$, T_τ denotes a translation by τ on $\mathbb{R}$. If the product $w_{r,t} = u_r \cdot \overline{T_t^* v_r}$, where $T_t^* v_r = v_r \circ T_t$ ($\circ$ stands for composition), yields an integrable distribution (Horváth 1966, Section 4.5) we define the value of the correlation function at t by $C[u_r, v_r](t) = \langle u_r \cdot \overline{T_t^* v_r}, 1\rangle = \langle w_{r,t}, 1\rangle$. In essence, this representation reveals whether the correlation can be taken or not. Whenever there is no ambiguity about the distributions u and v and the point r under consideration we will denote the correlation by $C(t)$.

$\mathcal{W}_{\psi_{p,t}} C(\omega) = \mathcal{F}\psi_{p,t} \overset{(\omega)}{*} \mathcal{F}C$ as a convolution in the Fourier domain for $\omega = \pm 1$, which acts as a smoothing pseudo-differential operator on $\mathcal{F}C$. The growth properties of $\mathcal{W}_{\psi_{p,t}} C$ reveal the wave front set of C at t in the direction $\omega = \pm 1$. These growth properties can be detected directly in the graph of $\mathcal{W}_{\psi_{p,t}} C(\pm 1)$ as a function of p (as $p \downarrow 0$) for each t. The actual testing of these properties is as follows: $(t, \pm 1) \notin \mathrm{WF}(C)$ if for any $N \in \mathbb{N}$,

$$|\mathcal{W}_{\psi_{p,t}} C(\pm 1)| \leq c_N p^N \quad \text{for} \quad p \in (0, 1] \tag{11}$$

for some constant c_N that may change with the power N. Effectively, this leads to a scanning procedure over t: Whenever condition (11) is not satisfied, t belongs to the projection of $\mathrm{WF}(C)$ on the time axis, and yields δt. In particular, this applies if $|\mathcal{W}_{\psi_{p,t}} C(\pm 1)| \approx p^M$ for some fixed M.

In general, multiresolution analysis of the (cross) correlation allows us to obtain a more precise estimate of traveltime offset δt. The analysis will require the use of multiple scales in the data (i.e. the data need to be broad band). Once a multiresolution analysis is invoked, however, the possibility of a time-frequency decomposition of the correlation appears. Incorporating this in a mismatch criterion would account for finite frequency 'bands' in a consistent way, and lead to multiresolution sensitivity kernels through linearization, see eq. (16) below.

4 SENSITIVITY KERNELS

4.1 Linearization procedure

Eq. (9) defines the sensitivity kernel for the travel time residual δt (at receiver location r, from a source at s), which in the Born approximation can be written as

$$\delta t = \frac{\int_{\mathbb{R}} \partial_t u_r(t) \delta u_r^{\mathrm{obs}}(t)\, \mathrm{d}t}{\int_{\mathbb{R}} \partial_t^2 u_r(t) u_r^{\mathrm{obs}}(t)\, \mathrm{d}t}, \text{ subject to } u_r^{\mathrm{obs}} = u_r. \tag{12}$$

In this expression, the denominator can be identified as an 'energy',

$$N_r = \int_{\mathbb{R}} |\partial_t u_r(t)|^2\, \mathrm{d}t$$

which is strictly positive. We take some caution with respect to the substitution $u_r^{\mathrm{obs}} = u_r$ in the Born approximation above. In fact, we will depart from this substitution, but only as far as the source signature is concerned.

We analyse the numerator with the aid of Green's tensors substituted for u_r (observed) and δu_r (modelled) displacement fields,

$$\int_{\mathbb{R}} \partial_t G_r(t, s) \delta G_r(t, s)\, \mathrm{d}t.$$

In this expression, $G_r(t, s)$ is $G(r, t, s)$ represented by (2) subject to the substitution $x_0 = s$ and $x = r$, and δG is represented by (6): To leading order the numerator is thus given by

$$(-1)^n \int a_0(r, s) A_0(s, r, x) \Omega(s, r, x) \delta c(x) \mathcal{H}^\sigma \delta^{(n)}[T(s, r, x) - T(r, s)]\, \mathrm{d}V(x),\ \sigma = \sigma(r, x) + \sigma(x, s) - \sigma(r, s). \tag{13}$$

Fig. 1 illustrates the geometry underlying this expression (x is an intersection point of the rays taking off at s and r, in the support of δc). Assuming a point source—see the remarks following (5)—we obtain an oscillatory integral representation for (12),

$$\delta t = -(-1)^n \int \Big\{ N_r^{-1} a_0(r, s) 2\mathrm{Re} \int_0^\infty A_0(s, r, x) \Omega(s, r, x) (-\mathrm{i}\omega)^n (\mathcal{F}\chi_\varepsilon)(\omega) \times \exp\big[-\mathrm{i}(\omega[(T(s, r, x) - T(r, s)] - \tfrac{1}{2}\pi\sigma)\big] \mathrm{d}\omega \Big\} \delta c(x)\, \mathrm{d}V(x), \tag{14}$$

where

$$\chi_\varepsilon(t) = C[W_d, W_m](t), \tag{15}$$

reminiscent of the cross correlation of observed and modelled data. Here, W_d is the band limited source signature of the measured data (associated with u_r) and W_m is the band limited source signature of the modelled data (associated with δu_r). The expression in between braces in (14)—subsequently referred to as 'the kernel'—is a function of $[(s, r), x]$ that maps $\delta c(x)$ to $\delta t(s, r)$.

Finally, with the expansion of δc into basis functions $\gamma_k(x)$,

$$\delta c(x) = \sum_k \langle \gamma_k, \delta c \rangle \gamma_k(x),$$

for each γ_k, eq. (14) can be written in the form

$$\int \left\{ \mathcal{A}(s, r, x) \mathcal{H}^\sigma (\chi_\varepsilon * (-1)^n \delta^{(n)})[\phi_{s,r}(x)] \right\} \gamma_k(x) \mathrm{d}V(x), \tag{16}$$

where $\mathcal{A}(s, r, x) = -N_r^{-1} a_0(r, s) A_0(s, r, x) \Omega(s, r, x)$ represents a smooth amplitude function, $\phi_{s,r}(x) = T(s, r, x) - T(r, s)$ is a phase function, and $\sigma = \sigma(r, x) + \sigma(x, s) - \sigma(r, s)$. Eq. (16) has mathematical meaning in the following sense: (i) The expression in between braces is the distribution in x that represents the kernel for wave-equation tomography; the notation emphasizes the fact that the essence of kernel behaviour is controlled by the singular (delta) part $\mathcal{H}^\sigma(\chi_\varepsilon * (-1)^n \delta^{(n)})$; (ii) χ_ε is a smooth function with properies elaborated in the next subsection, which is the cross correlation of the actual and the modelled source signatures (*cf.* eq. 15); (iii) $\chi_\varepsilon * \delta(t) = \chi_\varepsilon(t)$ can then be considered as a regularization of $\delta(t)$ and represents (approximately) the frequency band limitation; in the limit of infinite bandwidth ($\varepsilon \to 0$) $\chi_\varepsilon * \delta(t)$ tends to $\delta(t)$; (iv) $\delta^{(n)}(t)$ is the nth derivative of the delta, with n the spatial dimension;[5] and (v) the basis function $\gamma_k(x)$ serves as a test function on which the kernel operates.

Appendix A shows that the kernel for traveltime tomography can be obtained as a distributional broad-band limit of the kernel for 'wave-equation' tomography. Indeed, expression (16) reveals that the 'wave-equation' tomography and traveltime tomography share the same sensitivity kernel. On the other hand, to leading order, bandpass filtering of the data generates the net of regularizations of the kernel predicted by ray perturbation theory.[6]

Whether the regularization of the sensitivity kernel shows a zero on the unperturbed ray [where $T(s, r, x) = T(r, s)$ or $\phi_{s,r}(x) = 0$] depends on the presence of caustics (the value of σ), on χ, and

[5] $\delta[\phi_{s,r}(x)]$ is the pull back of δ under $\phi_{s,r}$, that is $\delta \circ \phi_{s,r} = \phi_{s,r}^* \delta$).

[6] Using the Rytov approximation as the point of departure, leads to an analysis of the sensitivity kernel of the type (Zhao *et al.* 2000, eq. 10) $\int \mathrm{Re}\{\delta G_r(\omega, s)/(-\mathrm{i}\omega) G_r(\omega, s)\}\, \mathrm{d}\omega$, in which $G_r(\omega, s)$, $\delta G_r(\omega, s)$ are the Fourier transforms of $G_r(t, s)$, $\delta G_r(t, s)$, respectively. Then $\delta^{(n)}[T(s, r, x) - T(r, s)]$ in the above needs to be replaced by $\delta^{(m+1)}[T(s, r, x) - T(r, s)]$ with $m + 1 = (n - 1)/2$ in the sense of (2). Observe that for $n = 3$, $m + 1 = 1$ yielding again an odd order derivative.

on the spatial dimension n. The behaviour of the sensitivity kernel, however, should not be derived from a particular regularization. Indeed, the notion that in the broad-band limit the wave-equation kernel (16) coincides with the ray-geometrical kernel must, in fact, mean that sensitivity is finite on the ray. To see this in a more general sense we must realize that kernels are distributions and that their action on test functions, here $\gamma_k(x)$, must be considered. The action of the kernel on an arbitrary test function would only be zero if that function spans precisely the volume on which $\phi_{s,r}(x) = 0$. But this is infinitesimally narrow, and for test functions of finite dimension the action of the kernel on it will always render finite values (even though—in the absence of caustics—they can be small).[7] In Section 4.3 this is illustrated both for even and odd test functions.

4.2 Regularization and band limitation

The generic kernel behaviour revealed by (16) is most conveniently explained in ray centred—or Fermi—coordinates in the vicinity of the unperturbed ray connecting r with s: In the direction locally orthogonal to the unperturbed ray the regularization of the kernel is oscillatory, whereas it is smoothly varying in level sets of $\phi_{s,r}$.

In general one does not know, however, where the value of the regularization of the kernel vanishes. Indeed, the results discussed by Dahlen *et al.* (2000) represent a specific case, which can be obtained from the above analysis: If $n = 3$ (or any odd dimension), $\sigma = 0$ (no caustics, hence a quasi-homogeneous reference medium), and if we take for χ_ε the autocorrelation of $W_m = W_d$ (that is, we assume that we know—and can use—the exact source signature), then the *regularization* of the kernel shows a zero ('doughnut hole') on the unperturbed ray. In this subsection, we elaborate on the intimate connection between the mathematical notion of regularization and the seismic notion of 'finite-frequency bandwidth'.

Band limitation. A regularization of a distribution can be written in the form of a convolution with a mollifier. The effect of band limitation on the kernel in (16) is precisely of this form. As before, we parametrize the bandwidth by ε. Since $\chi_\varepsilon(t)$ represents a regularization of $\delta(t)$ as $\varepsilon \to 0$, $\chi_\varepsilon^{(n)}(t)$ represents a regularization of the nth derivative, $\delta^{(n)}(t)$.

A mollifier, ρ say, is a smooth, compactly supported function such that $\int \rho(t)\,\mathrm{d}t = 1$. We model the regularization χ_ε as the net

$$\chi_\varepsilon(t) = \frac{1}{\varepsilon}\rho\left(\frac{t}{\varepsilon}\right). \tag{17}$$

As $\varepsilon \to 0$, its support shrinks to $\{0\}$ and $\chi_\varepsilon(t)$ tends to $\delta(t)$ (having a white spectrum), hence the name delta net. We reserve the symbol ρ for the case $W_m = W_d = W_\varepsilon$; that is, when ρ is even (or symmetric). It is immediate that in this case $\chi_\varepsilon^{(n)}(0) = 0$ for n odd, for any finite value of ε, which explains the 'hole' observed at the unperturbed ray in the regularization of the sensitivity kernel in the absence of caustics. It also explains why for n even, non-zero values occur at the unperturbed ray [for instance, for $n = 2$—surface waves—the value on the (unperturbed) wave path would be finite, and negative].

[7]The support of the kernel is determined as follows. We can restrict the kernel to a distribution in $B \subset \mathbb{R}^n$ by restricting the arbitrary test functions (possibly even or odd)—on which the kernel acts—to the ones compactly supported in B. The support of the kernel is then the set of points having no open neighborhood (B) to which the restriction of the kernel is zero. Applying this procedure to our kernel (for all scaling parameters ε)—realizing that B cannot be a point set—reveals that the unperturbed ray is always contained in the support of the sensitivity kernel.

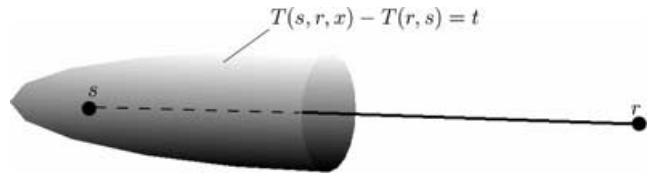

Figure 2. A level set of $\phi_{s,r}$, or isochron (in a homogeneous, isotropic medium), used in eq. (16).

However, instead of eq. (17), with symmetric mollifier ρ, one can also analyse sensitivity kernels via model delta nets with mollifiers that are not symmetric. This is not merely a mathematical exercise in mollifiers. Indeed, asymmetry of χ_ε would arise naturally if W_m (the modelled source signature) differs (slightly) from W_d (the real, but generally unknown, signature): That is, if we allow for uncertainty in such source parameters as the moment rate function. For example, let $W_m = T^*_{-\varepsilon\alpha} W_d$ where α may be a random variable while $\varepsilon\alpha \to 0$ as $\varepsilon \to 0$ so that W_m tends to W_d as $\varepsilon \to 0$. (Indeed, the shift $\varepsilon\alpha$ does not represent a fixed shift in time because such a shift should be independent of ε.) In this context, an equally valid mollifier, ρ, is obtained from ρ by the shift α,

$$\check{\rho}(t) = \rho(t - \alpha)$$

when

$$\chi_\varepsilon(t) = \frac{1}{\varepsilon}\check{\rho}\left(\frac{t}{\varepsilon}\right) = \frac{1}{\varepsilon}\rho\left(\frac{t - \varepsilon\alpha}{\varepsilon}\right). \tag{18}$$

The mollifier ρ is no longer even or symmetric about the origin, hence $\chi_\varepsilon^{(n)}(0)$ is not equal to 0 for n odd.

The effect of caustics. In the presence of caustics, χ_ε and its underlying mollifier may be Hilbert transformed (depending on whether σ is even or odd). A Hilbert transform applied to the mollifier ρ,

$$\mathcal{H}\rho(t) = \frac{1}{\pi} \fint_{-\infty}^{\infty} \frac{1}{t - t'}\rho(t')\,\mathrm{d}t',$$

naturally changes its form. For example, the Hilbert transform of an even mollifier, such as ρ, is odd. Indeed, $\mathcal{H}\chi_\varepsilon^{(n)}(0)$ with χ_ε as in (17) and n odd is not equal to 0. The recognition that the hole in the kernel can disappear in the presence of caustics is not new; we emphasize, however, that caustics form generically in heterogeneous media and they should not be considered as being exceptional.

4.3 Illustration of the above concepts

In this section we illustrate the above analysis, with emphasis on the oscillatory behaviour of the regularizations of the kernel eq. (16) for $n = 3$. For clarity, we use separate illustrations for the argument $\phi_{s,r}$, the regularization $\chi_\varepsilon * \delta^{(n)}$ and its Hilbert transforms, and the action on test functions γ_k.

Isochrons. The level set of points x of the function $\phi_{s,r}$ is an isochron corresponding to one value of time t, the argument in $\delta^{(n)}$ or its regularization (Fig. 2). A level set has the shape of the boundary of a Fresnel volume (Kravtsov & Orlov 1990), which is determined by $\phi_{s,r} = \frac{1}{2}f_0^{-1}$ if f_0 is the dominant frequency—indicated for our examples in Fig. 3. In simple media (such as media with a weak gradient in elastic parameters), the Fresnel volumes look like 'bananas'.

Regularizations. For W_d we use the time derivative of a model time-rise function of a large earthquake ($M_w = 6.5$) that occurred on 2000 May 4 at 516 km depth beneath Fiji–Tonga (Persh &

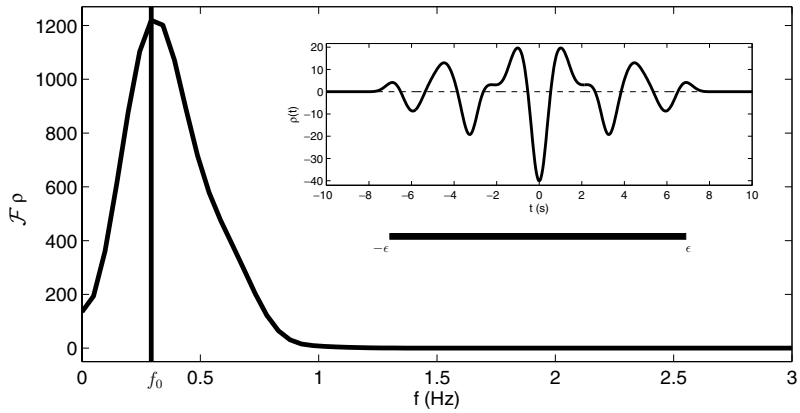

Figure 3. A mollifier $\rho(t)$ (inset) *cf.* eq. (17) and its amplitude spectrum derived from a moment rate function of a deep earthquake beneath Fiji–Tonga (time: 2004 May 4 −20:36:33.40; location: −17.801S, −178.412W, 516 km deep; magnitude: $M_w = 6.5$ Persh & Houston 2004, fig. 5 (event 000504)). The mollifier is normalized such that its integral is 1.

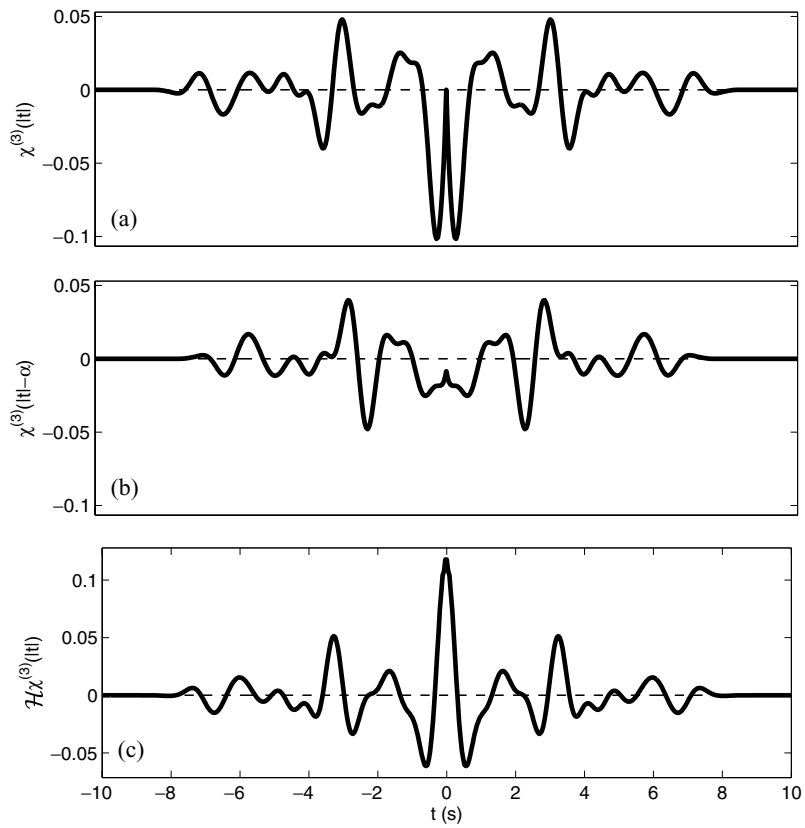

Figure 4. Illustration of the effect of symmetry and caustics on the regularization of a 3-D kernel. We plot $\chi_\varepsilon^{(n)}(|t|)$ ($n = 3$), since, for simple media, $|t|$ will be proportional to distance from the unperturbed ray. (a) A symmetric regularization (*cf.* 17) that is generated from the mollifier, $\rho(t)$, depicted in Fig. 3. Notice the zero value at $t = 0$. (b) An asymmetric regularization (*cf.* 18) generated with the mollifier $\rho(t)$ derived from Fig. 3 ($\alpha = 0.8$s). Notice that the zero value at $t = 0$ has disappeared. (c) Illustration of the effect of caustics: We show $(\mathcal{H}\chi)^{(n)}(|t|)$ (*cf.* 17) generated from the mollifier, $\rho(t)$, of Fig. 3. Notice that the zero value at $t = 0$ has disappeared.

Houston 2004, fig. 5 event 000504), and we generate the mollifier ρ in eq. (17) by autocorrelation (*cf.* eq. 15) and normalization. We illustrate $\rho(t)$ and its amplitude spectrum in Fig. 3. Note that, indeed, ρ is even. In Fig. 4(a) we show for this mollifier the regularization $\chi_\varepsilon^{(n)}$ of $\delta^{(n)}$. In this case, the regularization has a zero ('doughnut hole') precisely on the unperturbed ray (where the argument $\phi_{s,r} = 0$. This regularization is representative of the 3-D case in which the source signature is precisely known and if caustics are absent (that is, the medium must be simple) in dimension 3.

In Fig. 4(b) we show another regularization of $\delta^{(n)}$, now with the mollifier ρ derived from ρ in Fig. 3 using eq. (18) and $\alpha = 0.8$ s. The zero on the unperturbed ray has disappeared. This regularization is representative for the case in which the modelled and real source

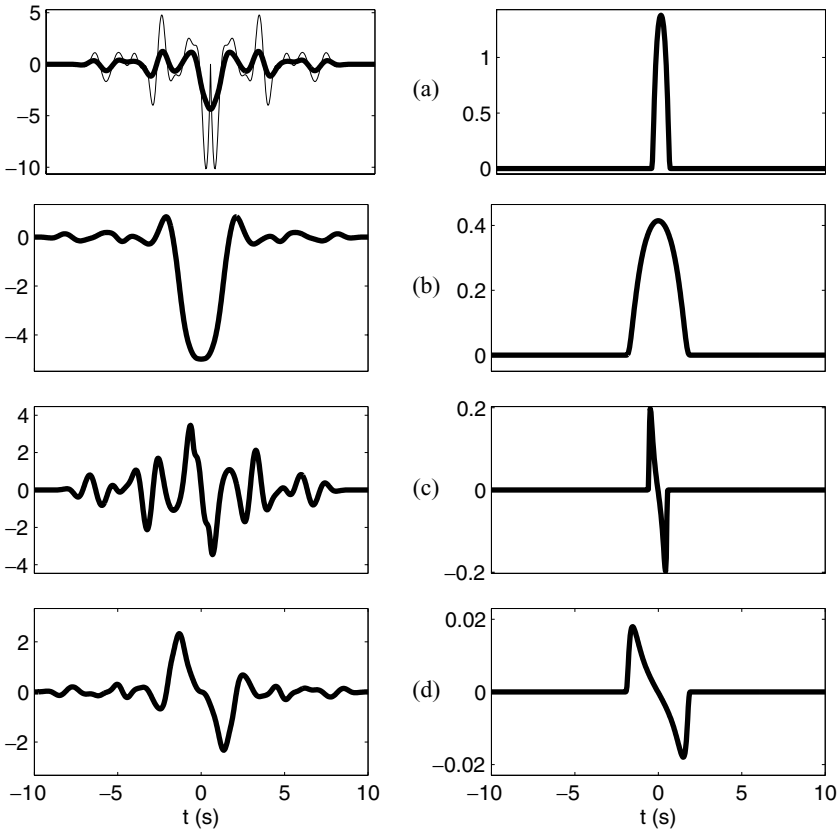

Figure 5. Illustration of the action of kernels—calculated from the symmetric $\chi^{(n)}(|t|)$, $n = 3$ in Fig. 4—on different test functions (*cf.* 16); the action is illustrated on the left, while the test functions are depicted on the right: (a) even, high resolution test function, (b) even, low resolution test function, (c) odd, high resolution test function, (d) odd, low resolution test function. (The even test functions are normalized to unit integral; the odd test functions are obtained from the even ones by differentiation.) For reference, the original regularization is shown as a thin line in the top, left panel.

signatures differ (slightly) from one another. We note that whereas the regularizations in Figs 4(a) and (b) differ, in the broad-band limit they tend to the same, (ray) geometrical, sensitivity kernel.

The Hilbert transform, $(\mathcal{H}\chi)^{(n)}(|t|)$, associated with the regularization $\chi^{(n)}(|t|)$ in Fig. 4(a) is shown in Fig. 4(c); clearly, the zero on the unperturbed ray has disappeared. Since $\mathcal{H}^2 = -1$ this is the only case that needs to be considered for illustrating the effect of caustics on the kernel regularization.

Test functions. The regularizations shown in Fig. 4 do not reveal the full behaviour of the sensitivity kernel. As mentioned above, we must consider the action of the kernel on different test functions. In this respect, the choice of (shape of) test function should be unimportant. Indeed, let the support of the test functions be shaped according to an elongated pebble or fat needle centred at x_0, say. We orient the needle close to orthogonal to the unperturbed ray; we consider both even and odd functions relative to x_0 in the direction of the needle's axis. In the directions orthogonal to the needle's axis (one of which is parallel to the unperturbed ray) we multiply with a symmetric cut-off function. Fig. 5 illustrates the action of the wave-equation sensitivity kernel on such test functions as a function of distance—between x_0 and the unperturbed ray and mapped to time accordancing to the level sets of $\phi_{s,r}$—with x_0 moving in a direction orthogonal to the ray. The test functions are plotted along the needle's axis—also as a function of time—on the right.

We consider four types of test function: High resolution, low resolution, even or odd. The imprint of the test functions on the result is obvious in the low resolution cases. In the case of an even test function centred at the unperturbed ray, the application of the wave-equation kernel—as well as the application of the ray-geometrical kernel—yield a non-vanishing outcome, see the left panels of Figs 5(a) and (b). For odd test functions centred at the unperturbed ray, the application of the wave-equation kernel—as well as the application of the ray-geometrical kernel—yield a vanishing outcome, see the left panels of Figs 5(c) and (d). As regards the action of kernels there is thus no discrepancy between wave equation and ray geometrical kernels.

5 MULTIRESOLUTION ASPECTS

With finite frequency sensitivity kernels for wave equation transmission tomography one can, in principle, account for different scales in the data (frequencies) and in the model (wavenumbers). Here, we briefly introduce a novel approach to multiresolution analysis of sensitivity kernels; a detailed analysis is beyond the scope of this paper and will be presented elsewhere (Douma *et al.* 2004, in preparation).

In the spirit of time-frequency or space-wavenumber localization, which are the subjects of wavelet or multi-taper analysis [e.g. (Simons *et al.* 2003)], we subject the kernel as represented by eq. (16) to a decomposition in terms of a frame of curvelets (Candes & Guo 2002). Curvelets can be thought of as some extension of wavelets to higher dimensions that incorporate geometry; they are optimally localized in space and wave vector, and provide a tiling of phase space, see Fig. 6. The microlocal analysis of wave propagation

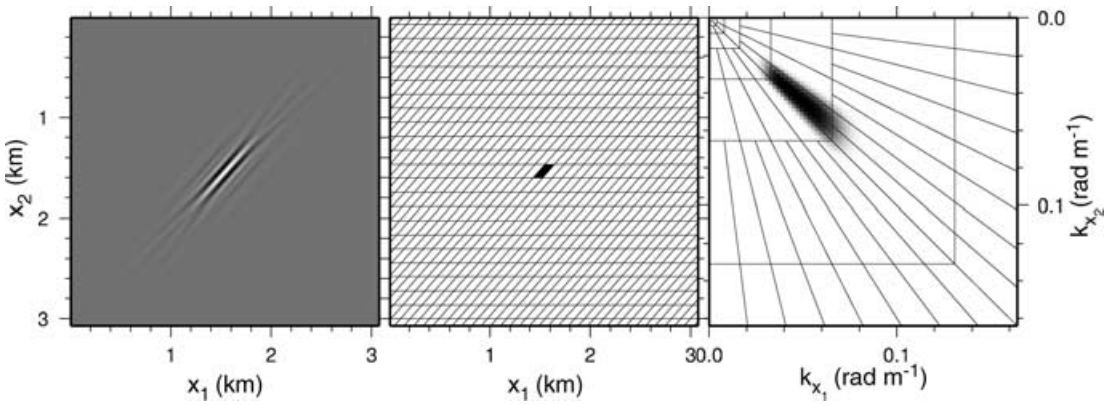

Figure 6. Illustration of a curvelet, in space (left), defining a grid (middle), and in wavenumber (right). Curvelets (Candes & Guo 2002), which are related to coherent wave packets, look like needles whose envelopes are 'ridges' of effective length $2^{-j/2}$ and width 2^{-j} with oscillatory character across the main 'ridge', where $j = 0, 1, 2, \ldots$ is a scale parameter.

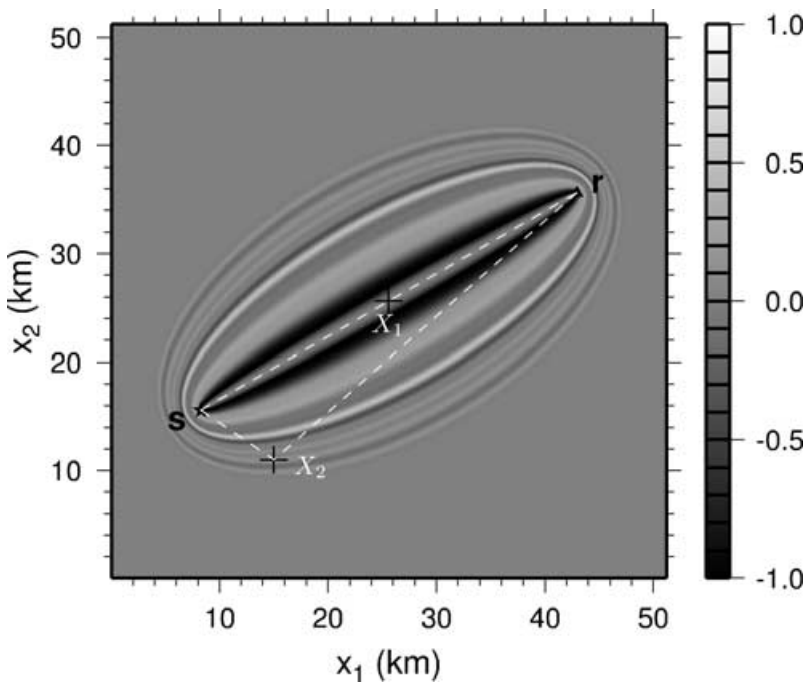

Figure 7. A regularization $\{[\chi_\varepsilon * (-1)^n \delta^{(n)}][\phi_{s,r}(x)]\}$ of the sensitivity kernel ($n = 3$) in a 2-D plane that contains the unperturbed ray. We assume a homogeneous, isotropic medium and use the even regularization (and mollifier) of Fig. 4 (that is, knowledge of the source signature). Note that everywhere on the unperutbed ray the value of the regularization of the kernel is zero. Figs 8 and 9 depict the sensitivity in point X_1 midway on the unperturbed source-receiver path and in X_2 away from this path, respectively.

concerns the (space,time) location and direction of singularities (*cf.* eq. 4), and curvelets and their 3-D realization provide a natural frame for this, as well as for wave equation tomography; in eq. (16) we chose them as γ_k, with k representing a multi-index describing translation, dilation, and rotation.

We illustrate the multiresolution concept with a simple (isotropic, quasi-homogeneous) medium in 3-D that does not produce caustics ($\sigma = 0$), and we assume that we know the precise source signature ($W_d = W_m = W_\varepsilon$ so that χ_ε is even). In this case the regularization of the 3-D kernel resembles a 'banana-doughnut' (Dahlen *et al.* 2000). Fig. 7 depicts a 2-D cross section that contains the unperturbed ray, connection receiver r with source s. We use the same source time function (and mollifier) as before (see Figs 3 and 4).

Following eq. (16), we first project the kernel in Fig. 7 on γ_k's. The goal is to quantify the contribution from curvelets—or, the sensitivity to spatial scales of structure—at a point on the unperturbed ray. For this purpose we restrict the sum over k so that the translation index corresponds to a point, X_1, half-way between the source and receiver on the unperturbed ray (indicated by a cross in Fig. 7); scale and orientation are then the remaining degrees of freedom. Fig. 8 illustrates the sensitivity at X_1 to different scales and directions. The column on the left depicts the contribution in space to the (normalized) sensitivity. The panels on the right are obtained from the ones on the left by a spatial Fourier (space wavenumber) transform; they illustrate the directions and scales of structure sensed at point X_1. Figs 8(a)–(c) show contributions from the smallest to the largest scales constrained by the data, and (d) depicts the contribution to the kernel for all scales combined.

Fig. 8 demonstrates that even for the 'banana-doughnut'-like regularization shown in Fig. 7, the data will be sensitive to structure

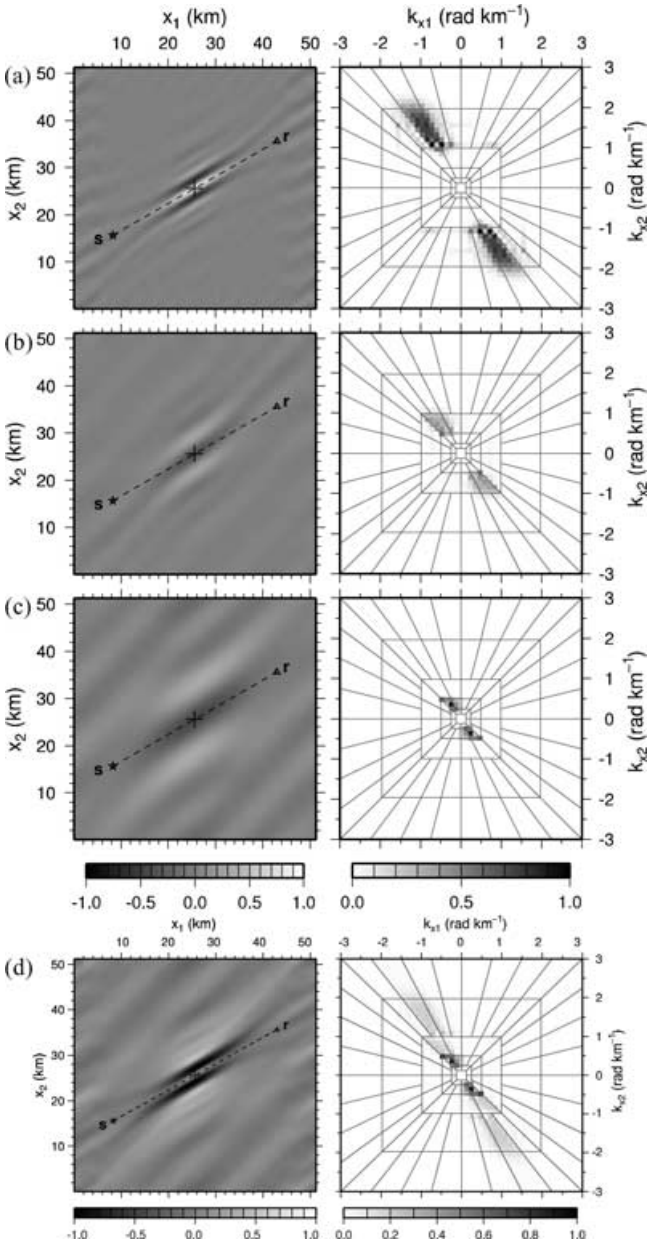

Figure 8. Multiresolution analysis of the sensitivity kernel shown in Fig. 7 using curvelets (Fig. 6). We use curvelets to decompose the kernel. In this figure we illustrate the contributions to the kernel from curvelets located on the unperturbed ray at X_1, half-way between the source and the receiver. The remaining degrees of freedom are scale and orientation. Left: space domain representations; right: Fourier (wavenumber) domain representations (the latter subjected to an amplification by a factor of 2 between scales). The wedges in the figures on the right indicate the resolved directions whereas the annulus in which a wedge lies indicates the resolved scale, with small wavenumbers (long wavelengths) plotting near the centre whereas large wavenumbers smaller wavelengths) plotting closer to the edge of the diagrams. Panels (a)–(c) show the contributions from the highest to the lowest scale, and (d) shows the total contribution by all scales and directions.

on the unperturbed ray, but only for orientations and (spatial) scales contained in the data. As expected, the panels on the right (wavenumber domain) in Fig. 8 demonstrate that the kernel would—for all scales available from the data—resolve variations in structure primarily in directions contained in wedges perpendicular to the unperturbed ray. The annulus in which a wedge lies indicates the resolved scale. The panels on the left (space domain) illustrate the lack of resolution in particular along the direction of the unperturbed ray.

In a similar fashion we can quantify the sensitivity to structure at an arbitrary position X_2 away from the unperturbed ray. For this purpose we restrict the sum over k so that the translation index corresponds to X_2 on the broken ray in Fig. 7. Fig. 9 (right) reveals a rotation of the wedges of resolved orientations with scale: The lowest scale senses contributions near the unperturbed ray (in directions primarily perpendicular to the unperturbed ray), whereas the highest scale senses contributions near the point where the ray is broken (in directions primarily along the bisectrix of the ray paths from s and from r).

In principle, for each point in space we can quantify the relationship between the time scales in the cross correlation and the resolved structural length scales (and directions) and thus conduct a full multiresolution analysis of wave-equation tomography.

6 DISCUSSION

It is still often impossible to draw conclusions from the behaviour of sensitivity kernels derived from any optimization criterion in the context of inverse problems. Any such conclusions would assume a precise understanding of the degree of uniqueness (Uhlmann 2001), that is, in what sense the optimization criterion applied to the data determines the real variations in the elastic properties of the Earth. Even for perfect traveltime data in an anisotropic, elastic Earth without boundaries or interfaces the question of uniqueness is as yet unresolved. In fact, in the context of local optimization, as long as significant damping (regularization) is required because of data errors and uneven data coverage, the precise behaviour of the kernels may be less important than model parametrization and regularization.

Apart from the above argument, one may wonder whether the banana-doughnut structure with a 'hole' occurs in a sensitivity kernel. In the previous sections we answered this question. Taking (9) as a point of departure, and assuming the specific case of (i) $n = 3$, (ii) a simple medium (that is, the absence of caustics), (iii) high-frequency observations of the delta wave part of the wavefield, (iv) the use of the symmetric autocorrelation of the time derivative of the time-rise function as a regularization (that is, knowing the exact source signature), a banana-doughnut structure appears in the regularization of sensitivity kernel as the result of a *pointwise* evaluation of its values. However, an evaluation of this kernel in the sense of generalized functions leads to a structure without the 'hole'. This is confirmed by a multiresolution analysis using curvelets of its regularization.

Furthermore, as recognized by Dahlen *et al.* (2000), the banana-doughnut kernels strictly relate to a particular class of cross-correlation traveltime delays. The precise analysis of the above framework reveals, however, that the traveltime-delay mismatch criterion does not honor the detection of traveltime (delays) from broad-band data. Thus the criterion is not a 'frequency dependent' traveltime delay criterion. Other measurements, including ones that do detect the phase arrivals proper, imply different kernels.

Taking the cross correlation criterion as the point of departure, there are obstructions to the formation of the 'hole' in the kernel: Indeed, from a physical perspective, caustics form generically in a heterogeneous medium (White *et al.* 1988). Moreover, from a mathematical perspective it is clear that the singular behaviour of the kernel should not depend upon the (data driven) regularization. The latter problem can be addressed with multiresolution techniques using the actual time-rise function in the data. Furthermore, one must

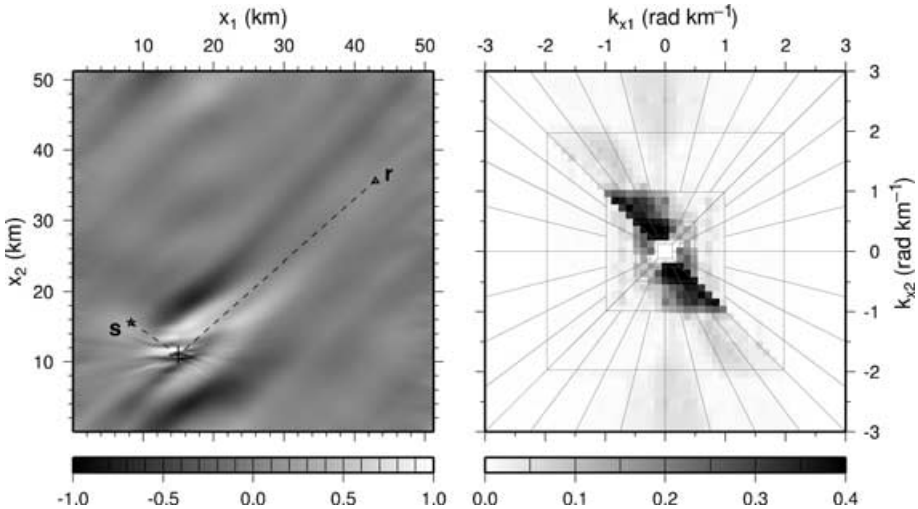

Figure 9. Contributions from curvelets located at X_2 away from the unperturbed ray as indicated on the broken ray in Fig. 7—that is, an arbitrary point away from the unperturbed source-receiver ray. We observe the rotation of wedge with scale: The lowest scale senses contributions near the unperturbed ray whereas the highest scale sense contributions near the point where the ray is broken.

be careful with the use of approximate finite frequency kernels to linearize tomographic inversions, because the kernels calculated in the starting model (usually simple, possibly with a 'banana-doughnut' feature) may well differ significantly from the sensitivity kernels implied by the heterogeneous model produced by inversion.

We would also like to discuss the analysis of the kernels based on the Born approximation in the context of 'wavefield resolution'. On the one hand, the physical phenomenon of wave front healing [for a numerical investigation, see, e.g. Igel & Gudmundsson (1997); the analysis can be carried out, e.g. for the diffraction off a sphere (Nussenzveig 1992)] occurs in the process of scattering waves in a finite frequency window from an obstacle with compact support. Comparing the scattered field with the incident field (in the absence of the obstacle) by cross correlation would reveal that an obstacle that is 'small' compared to the predominant wavelength would not show up. On the other hand, in odd dimensions with the choice of a symmetric mollifier, the sensitivity kernel for wave-equation tomography goes to zero on the unperturbed ray connecting the epicentre with the seismic station. Thus the behaviour of the kernel becomes compatible with the phenomenon of wave front healing if the obstacle (now scatterer) is visited by unperturbed rays connecting the epicentre with the seismic station and fits into the region where the kernel is close to zero (Hung *et al.* 2001). However, this compatibility disappears readily (i) in odd space dimensions if the scatterer is located away from the unperturbed ray but is still in the range of the kernel, because the kernel will then sense the object—and thus yield a contribution—while in reality in this case the object would not be detected owing to physical wave phenomena, (ii) in the presence of caustics forming in the unperturbed medium, and (iii) in even space dimensions if the scatterer is visited by unperturbed rays. An elegant approach to quantify these statements is the one of generalized screens, which connects ray theory to wave theory through the notion of a frequency dependent generalized slowness surface (Fishman *et al.* 2000), and contains various aspects of scattering from obstacles; the development of this is beyond the scope of this paper, however, and we refer the reader to (Le Rousseau & de Hoop 2003).

Wave-equation sensitivity kernels for transmission tomography will prove to be important in the problem of data fusion. We have elucidated this by means of a multiresolution analysis of their regularizations using curvelets.

ACKNOWLEDGMENTS

The authors thank H. Douma and A. E. Malcolm (Colorado School of Mines) for developing numerical procedures and for generating the figures, E. Candes (California Institute of Technology) for providing the curvelet software, G. Hörmann (University of Vienna) for stimulating and helpful discussions, H. Houston for providing the source time functions used in Fig. 3, and J. Tromp, B. Romanowicz and V. Farra for constructive reviews. This research was supported under NSF grant EAR-0409816 and a Dutch National Science Foundation grant (NWO:VICI 865.03.007) for Innovative Research.

REFERENCES

Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.,* **103,** 30 055–30 078.

Burridge, R., De Hoop, M.V., Miller, D. & Spencer, C., 1998. Multiparameter inversion in anisotropic elastic media, *Geophys. J. Int.,* **134,** 757–777.

Càndes, E. & Guo, F., 2002. New multiscale transforms, minimum total variation synthesis: Application to edge-preserving image reconstruction, *Sig. Process. (Special issue in Image and Video Coding Beyond Standards),* **82,** 1519–1543.

Červený, V. & Ravindra, R.J., 1971. *Theory of seismic head waves.* University of Toronto Press, Toronto.

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

De Hoop, M.V., Spencer, C. & Burridge, R., 1999. The resolving power of seismic amplitude data: An anisotropic inversion/migration approach, *Geophysics,* **64,** 852–873.

Farra, V. & Le Bégat, S., 1995. Sensitivity of *qP*-wave traveltimes and polarization vectors to heterogeneity, anisotropy and interfaces, *Geophys. J. Int.,* **121,** 371–384.

Fishman, L., De Hoop, M.V. & van Stralen, M.J.N., 2000. Exact constructions of square-root Helmholtz operator symbols: The focusing quadratic profile, *J. Math. Phys.,* **41,** 4881–4938.

Fukao, Y., Obayashi, M., Inoue, H. & Nenbai, M., 1992. Subducting slabs stagnant in the mantle transition zone, *J. geophys. Res.,* **97,** 4809–4822.

Grand, S.P., 1994. Mantle shear structure beneath the Americas and the surrounding oceans, *J. geophys. Res.,* **99,** 11 591–11 621.

Hörmann, G. & de Hoop, M.V., 2002. Detection of wave front set perturbations via correlation: Foundation for wave-equation tomography, *Appl. Anal.,* **81,** 1443–1465.

Horváth, J., 1966. *Topological vector spaces and distributions,* Addison-Wesley, Reading, MA, USA.

Hörmander, L., 1990. *The Analysis of Linear Partial Differential Operators,* Vol. 1, 2nd edn, Springer-Verlag, Berlin.

Hung, S.-H., Dahlen, F.A. & Nolet, G., 2001. Wavefront healing: a banana-doughnut perspective, *Geophys. J. Int.,* **146,** 289–312.

Igel, H. & Gudmundsson, O., 1997. Frequency-dependent effects on travel times and waveforms of long-period S and SS waves, *Phys. Earth. planet. Inter.,* **104,** 229–246.

Inoue, H., Fukao, Y., Tanabe, K. & Ogata, Y., 1990. Whole mantle *P*-wave travel-time tomography, *Phys. Earth planet. Inter.,* **59,** 291–328.

Kárason, H. & van der Hilst, R.D., 2000. Constraints on mantle convection from seismic tomography, in, *The History and Dynamics of Global Plate Motion,* Vol. 121, pp. 277–288. eds Richards, M.R., Gordon, R. & Van der Hilst, R.D., Geophysical Monograph, American Geophysical Union, Washington, DC, USA.

Kárason, H. & van der Hilst, R.D., 2001. Tomographic imaging of the lowermost mantle with differential times of refracted and diffracted core phases (PKP, Pdiff), *J. Geoph. Res.,* **106,** 6569–6588.

Kárason, H., Van der Hilst, R.D. & Masters, G., 2001. Unpublished manuscript.

Katzman, R., Zhao, L. & Jordan, T.H., 1998. High-resolution, two-dimensional vertical tomography of the central Pacific mantle using ScS reverberations and frequency-dependent travel times, *J. geophys. Res.,* **103,** 17 933–17 971.

Kendall, J.M. & Thomson, C.J., 1993. Maslov ray summation, pseudo-caustics, Lagrangian equivalence and transient seismic waveforms, *Geophys. J. Int.,* **113,** 186–214.

Kennett, B.L.N., Widiyantoro, S. & van der Hilst, R.D., 1998. Joint seismic tomography for bulk-sound and shear wavespeed, *J. geophys. Res.,* **103,** 12 469–12 493.

Kravtsov, Yu.A. & Orlov, Yu.I., 1990. *Geometrical optics of inhomogeneous media,* Springer-Verlag, New York.

Le Rousseau, J.H. & de Hoop, M.V., 2003. Generalized-screen approximation and algorithm for the scattering of elastic waves, *Q. J. Mech. Appl. Math.,* **56,** 1–33.

Li, X.D. & Romanowicz, B., 1996. Global mantle shear velocity model developed using nonlinear asymptotic coupling, *J. geophys. Res.,* **101,** 22 245–22 272.

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

Marquering, H., Nolet, G. & Dahlen, F.A., 1998. Three-dimensional waveform sensitivity kernels, *Geophys. J. Int.,* **132,** 521–534.

Montelli, R., Nolet, G., Masters, G., Dahlen, F.A. & Hung, S.-H., 2004. Global *P* and *PP* travel time tomography: rays versus waves, *Geophys. J. Int.,* submitted.

Nussenzveig, M.H., 1992. *Diffraction effects in semiclassical scattering,* Cambridge University Press, Cambridge.

Persh, S.E. & Houston, H., 2004. Deep earthquake rupture histories determined by global stacking of broadband P waveforms, *J. geophys. Res.,* **109,** doi:10.1029/2003JB002762.

Simons, F.J., Van der Hilst. R.D. & Zuber, M.T., 2003. Spatio-spectral localization of isostatic coherence anisotropy in Australia and its relation to seismic anisotropy: Implications for lithopsheric deformation, *J. geophys. Res.,* **108,** 10.1029/2001JB000704.

Spetzler J., Trampert J. & Snieder, R., 2001. Are we exceeding the limits of the great circle approximation in global surface wave tomography?, *Geophys, Res. Lett.,* **28,** 2341–2344.

Stolk, C.C. & de Hoop, M.V., 2004. Seismic inverse scattering in the downward continuation approach, *SIAM J. Appl. Math.,* in press.

Su, W.-J. & Dziewonski, A.M., 1997. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth planet. Inter.,* **100,** 135–156.

Symes, W.W. & Carazzone, J., 1991. Velocity inversion by differential semblance optimization, *Geophysics,* **56,** 654–663.

Taylor, M.E., 1975. Reflection of singularities of solutions to systems of differential equations, *Comm. Pure Appl. Math.,* **28,** 457–478.

Uhlmann, G., 2001. Travel time tomography, Mathematics in the new millennium, *J. Korean Math. Soc.,* **38,** 711–722.

Van der Hilst, R.D., Widyantoro, S. & Engdahl, E.R., 1997. Evidence for deep mantle circulation from global tomography, *Nature,* **386,** 578–584.

VanDecar, J. & Crosson, R., 1991. Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares, *Bull. seism. Soc. Am.,* **80,** 150–159.

White, B.S., Nair, B. & Bayliss, A., 1988. Random rays and seismic amplitude anomalies, *Geophysics,* **53,** 903–907.

Woodward, M.J., 1992. Wave-equation tomography, *Geophysics,* **57,** 15–26.

Woodward, R.L. & Masters, G., 1991. Global upper mantle structure from long-period differential traveltimes, *J. geophys. Res.,* **96,** 6351–6377.

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

APPENDIX A: CONNECTION WITH RAY AND TRAVELTIME PERTURBATION

In this Appendix we will show (i) that with delta waves expression (12) leads to the travel time difference predicted by ray perturbation theory (e.g. Farra & Le Bégat (1995)), and (ii) that from the travel time perturbation we can extract a sensitivity kernel in the framework of distribution theory. Both, (i) and (ii) are manifestations of well-known results in the theory of Fourier integral operators.

The denominator of (12), with the aid of Parseval's theorem, allows the frequency integral representation

$$\int |a_0(r,s)|^2 |\mathcal{F}W_\varepsilon|^2 (-\mathrm{i}\omega)^{n-1} \mathrm{d}\omega.$$

Intuitively, the denominator should provide a normalization in (12) that combined with (14) leads to the travel time perturbation obtained by ray perturbation theory.

In (13) consider $\delta(T(r, s) - T(s, r, x))$, and introduce 'isochrones' as the level sets $T(s, r, x) = \tau$ for given (s, r). Isochrones define the boundaries of volumes containing the direct ray. For isochrones in the neighbourhood of the direct ray (s to r), i.e. $0 \leq \tau - T(r, s)$ sufficiently small, we can construct coordinates on the associated volume starting with $q = (q_1, \ldots, q_{n-1})$ representing ray centred (Fermi) coordinates, so that the paraxial ray approximation applies,

$$T(s, r, x) - T(r, s) = \tfrac{1}{2}\langle q, Mq\rangle. \tag{A1}$$

Here $\langle .,. \rangle$ denotes a $(n-1)$-dimensional inner product, and we identify a point on the direct ray such that the difference vector connecting it with x has coordinates q. M is the sum of wavefront curvature matrices, associated with the rays connecting s to x and r to x. In mentioned volumetric neighbourhood, defined by

$$0 \leq \tfrac{1}{2}\langle q, Mq\rangle \leq \tau - T(r, s), (s, r) \text{ given},$$

we complete the coordinates by including $\tilde{t}$ as the evolution parameter along the direct (in the main text unperturbed) ray, $\boldsymbol{x}(\tilde{t})$ say, connecting r (at $\tilde{t} = T(r, s)$) with s (at $\tilde{t} = 0$). The change of coordinates, $x \mapsto (q_1, q_2, \tilde{t})$, implies for the volume element

$$\mathrm{d}V(x) = (\partial_\tau \mathcal{A})\, \mathrm{d}\tau\, |\dot{\boldsymbol{x}}(\tilde{t})|\, \mathrm{d}\tilde{t}, \tag{A2}$$

where $\mathcal{A}$ is a cross section, for $n = 3$ the ellipse with area

$$\mathcal{A} = \frac{2\pi(\tau - T(r,s))}{\sqrt{M_1 M_2}} H(\tau - T(r,s)), \tag{A3}$$

in which M_1, M_2 are the eigenvalues of the 2×2 matrix M and are evaluated at x which position determines $\tilde{t}$. Using symplectic properties associated with the system of equations for dynamic ray tracing, it follows that

$$\frac{1}{\sqrt{M_1 M_2}} = \frac{4\pi a_0(r,s)}{(4\pi)^2 A_0(s,r,x)} \frac{1}{|\dot{\boldsymbol{x}}(\tilde{t})|}, \tag{A4}$$

where the amplitudes have been identified with geometrical spreadings, and radiation patterns have been taken out.

We now return to the numerator of expression (12) given in the equation below it. Within this numerator we encounter $\delta\ G_r(t,s)$. In its volume integral representation, we will apply the change of coordinates introduced above, viz.

$$\begin{aligned}
\delta G_r(t,s) &\sim -\partial_t^2 \int A_0(s,r,x)\Omega(s,r,x)\delta c(x)\delta(t - T(s,r,x))\,\mathrm{d}V(x) \\
&\simeq -\frac{1}{2}\int_{\tilde{t}=0}^{T(r,s)} \int_{\tau=T(r,s)}^{\infty} a_0(r,s)\Omega(s,r,\boldsymbol{x}(\tilde{t}))\delta c(\boldsymbol{x}(\tilde{t}))\delta(t-\tau)\,\mathrm{d}\tau\,\mathrm{d}\tilde{t} \\
&= \underbrace{a_0(r,s)\delta'(t - T(r,s))}_{\partial_t G_r(t,s)} \frac{1}{2}\int_{\tilde{t}=0}^{T(r,s)} (-)\Omega(s,r,\boldsymbol{x}(\tilde{t}))\delta c(\boldsymbol{x}(\tilde{t}))\,\mathrm{d}\tilde{t},
\end{aligned} \tag{A5}$$

where we have used that $\Omega(s,r,.)\,\delta c(.)$ varies smoothly in the vicinity of the direct ray, whereas s and r are fixed. In this expression, we identify

$$\delta T = -\frac{1}{2}\int_0^{T(r,s)} (-)\Omega(\boldsymbol{x}(t))\delta c(\boldsymbol{x}(t))\,\mathrm{d}t, \tag{A6}$$

the travel time shift predicted by ray perturbation theory. To complete the estimation of the numerator, we need to evaluate the autocorrelation,

$$-\int_{\mathbb{R}} \partial_t G_r(t,s)\partial_t G_r(t,s)\,\mathrm{d}t = \int_{\mathbb{R}} \partial_t^2 G_r(t,s)\partial G_r(t,s)\,\mathrm{d}t;$$

we resort to the application of the source signature and Parseval's theorem, yielding

$$\int |a_0(r,s)|^2 |\mathcal{F}W_\varepsilon|^2 (-i\omega)^{n-1}\,\mathrm{d}\omega, \quad n = 3,$$

as in the representation of the denominator initially discussed in the appendix. Indeed, the ratio in (12) reduces to the travel time shift predicted by ray perturbation theory.

We conclude the appendix by extracting from (A6) the ray perturbation sensitivity kernel as a distribution. To this end, let us introduce the translation operator $T_{\boldsymbol{x}(\tilde{t})}$, then we can define the mapping $d : \mathbb{R} \to \mathcal{D}'(\mathbb{R}^n)$ through the pull-back $d(\tilde{t}) = T^*_{\boldsymbol{x}(\tilde{t})}\delta$. Then $H \overset{(\tilde{t})}{*} d \in C^\infty(\mathbb{R}, \mathcal{D}'(\mathbb{R}^n))$ which allows us to write (A6) in the form

$$\delta T = -\frac{1}{2}\int \left\{ \int_0^{T(r,s)} (-)\Omega(\boldsymbol{x}(t))\delta(x - \boldsymbol{x}(t))\,\mathrm{d}t \right\} \delta c(x)\,\mathrm{d}V(x). \tag{A7}$$