

# Wave-equation reflection tomography: annihilators and sensitivity kernels

Maarten V. de Hoop,<sup>1</sup> Robert D. van der Hilst<sup>2</sup> and Peng Shen<sup>3</sup>

<sup>1</sup>Center for Computational and Applied Mathematics, Purdue University, 150 N. University Street, West Lafayette IN 47907, USA

<sup>2</sup>Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Rm 54-517A, Cambridge, MA 02139, USA.

E-mail: hilst@mit.edu

<sup>3</sup>Total E&P USA, 800 Gessner, Suite 700, Houston TX 7024, USA

Accepted 2006 July 6. Received 2006 July 2; in original form 2005 December 13

## SUMMARY

In seismic tomography, the finite frequency content of broad-band data leads to interference effects in the process of medium reconstruction, which are ignored in traditional ray theoretical implementations. Various ways of looking at these effects in the framework of *transmission* tomography can be found in the literature. Here, we consider inverse scattering of body waves to develop a method of wave-equation *reflection* tomography with broad-band waveform data—which in exploration seismics is identified as a method of wave-equation migration velocity analysis. In the transition from transmission to reflection tomography the usual cross correlation between modelled and observed waveforms of a particular phase arrival is replaced by the action of operators (annihilators) to the observed broad-band wavefields. Using the generalized screen expansion for one-way wave propagation, we develop the Fréchet (or sensitivity) kernel, and show how it can be evaluated with an adjoint state method. We cast the reflection tomography into an optimization procedure; the kernel appears in the gradient of this procedure. We include a numerical example of evaluating the kernel in a modified Marmousi model, which illustrates the complex dependency of the kernel on frequency band and, hence, scale. In heterogeneous media the kernels reflect proper wave dynamics and do not reveal a self-similar dependence on frequency: low-frequency wave components sample preferentially the smoother parts of the model, whereas the high-frequency data are—as expected—more sensitive to the stronger heterogeneity. We develop the concept for acoustic waves but there are no inherent limitations for the extension to the fully elastic case.

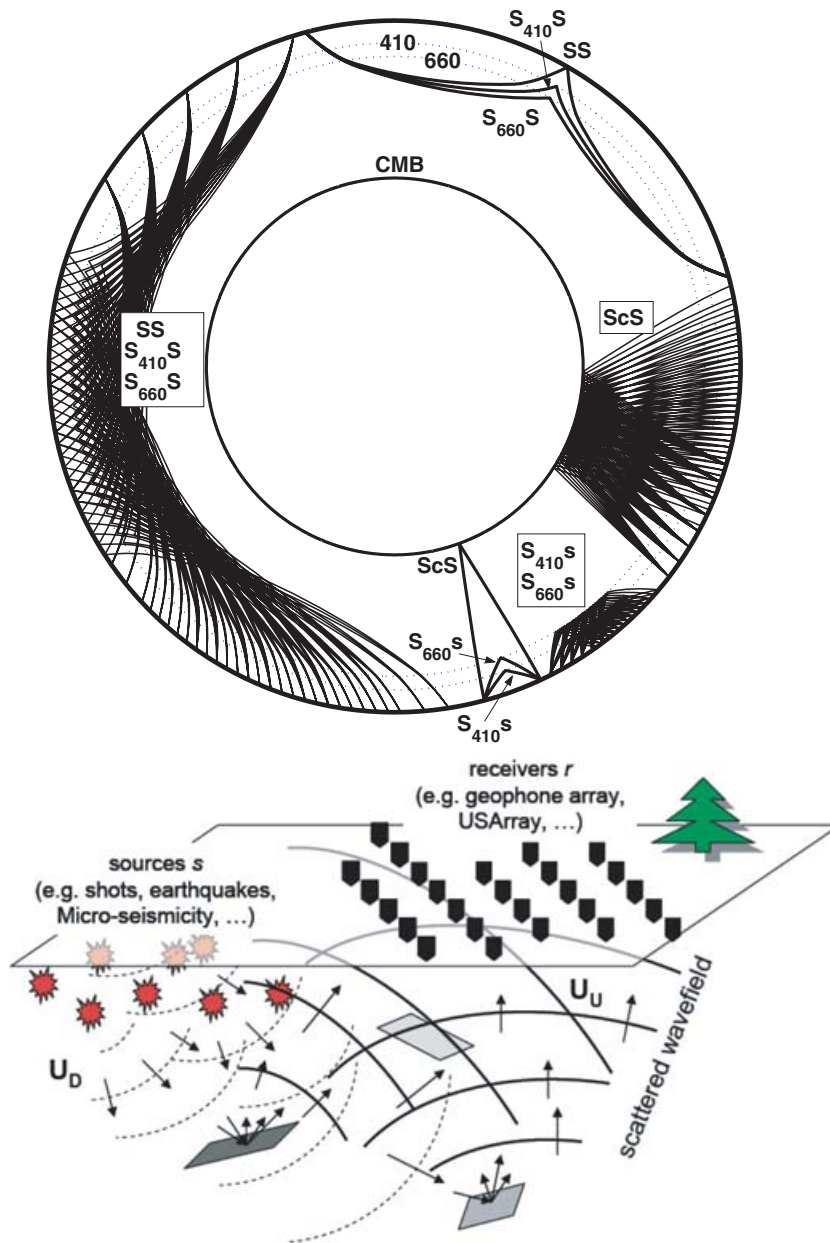
**Key words:** migration velocity analysis, reflection tomography, sensitivity kernels.

## 1 INTRODUCTION

The ever growing volumes of densely sampled broad-band data, both in industry applications and in global (array) seismology, are providing new opportunities and challenges for the development of techniques for subsurface imaging that exploit efficiently the rich information contained in seismic waveforms. Here, we analyse a method of wave-equation reflection tomography with finite-frequency data (derived from De Hoop & Van der Hilst 2003); in exploration seismics one refers to such methods in general as wave-equation migration velocity analysis (MVA, Biondi & Sava 1999). It makes use of the redundancy in the scattered wavefield data; the scattering is assumed to occur somewhere in the subsurface (Fig. 1, bottom), but the locations of the scatterers need not be known.

The general objective of linearized seismic tomography is to find by inversion a model, or a class of models, of elastic properties that through forward modelling predicts the data. The data fit is usually evaluated through a particular criterion, often expressed as a cost or penalty functional that is to be optimized. In the process of optimization, the key quantity to be found is the so called Fréchet or sensitivity kernel. The criterion—and, therefore, the character of the sensitivity kernel—depends on the type of observation being interpreted.

In traveltimes tomography the observed data are the measured arrival times of known seismic phases, and the inversion typically involves a backprojection algorithm for estimating medium wave speeds relative to some reference model. In the idealized case of infinitely high frequency wave propagation the traveltimes and kernel (ray paths) are well defined. With finite frequency data, however, the direct detection of the traveltimes is often obscured by wave interference and diffraction so that alternative ways of comparing modelled and observed data must be sought. In the framework of *transmitted* body waves various ways of accounting for finite frequency effects in the process of medium estimation have been proposed and analysed (Luo & Schuster 1991; Woodward 1992; Dahlen *et al.* 2000; De Hoop & Van der Hilst 2005). In such approaches to finite-frequency tomography one can distinguish between methods based on explicit backprojection over Fresnel-like



**Figure 1.** Top: Examples of reflected wavefields that can be considered in the wave equation reflection tomography developed in this paper. The medium reconstruction (that is, wave speed estimation) takes place where up and downgoing wavefields interact: the higher the density of scatterers (e.g. reflectors), the more constrained the solution. Bottom: Cartoon of source and receiver array concept exploited by the reflection tomography approach.

volumes (instead of backprojection along infinitesimally narrow rays), see, for instance, Dahlen *et al.* (2000) and De Hoop & Van der Hilst (2005), and those combining the wave equation with the Born approximation using an adjoint state method (Vasco *et al.* 1995; Tromp *et al.* 2005). In wave-equation tomography the wavefields in the kernel can be computed directly from the time-domain wave equation, using Green’s functions calculated, for instance, by spectral elements, normal mode summation (Zhao & Jordan 1998), or frequency-domain one-way wave propagation.

Here we focus on a method of wave equation reflection tomography, not to detect and investigate reflectors but to estimate subsurface wave speed variations. We use a frequency domain formulation and assume single scattering. In contrast to tomographic approaches that interpret specific phase arrivals—measured, for instance, by phase picking or waveform cross correlation—we consider the entire, broad-band wavefield (although some time windowing or other processing may need to be applied to mute the parts of the data influenced by multiple scattering). Instead of measuring single time-shifts, it involves the application of *annihilators* to the observed data. These annihilators are operators whose action on the wavefield vanishes if the wave speed model predicts the observations used: the level of annihilation is thus a measure for the accuracy of estimated wave speed variations. The annihilation criterion has a significant advantage over conventional travelttime mismatch criteria in that it explicitly accounts for the effects of scattering in Earth’s interior, while it exploits the redundancy in

the (scattered) wavefield. Such redundancy can arise from observations at multiple offsets (in exploration seismology) or at multiple angular epicentral distances and azimuths (in global seismology). For example, in the studies of sedimentary basins one can use waves scattered from the many subsurface reflectors and faults to elucidate the intervening seismic velocities (e.g. Zelt *et al.* 2003), and on a global scale one could use the wavefields formed by reflection off the core mantle boundary or the upper mantle discontinuities (see Fig. 1, top) to study mantle heterogeneity (instead of mantle layering, which has been done before Sipkin & Jordan 1976; Revenaugh & Jordan 1991).

The method presented here involves several steps, which also define the structure of this paper. First, we decompose the broad-band wavefield into up and downgoing waves (Section 2). This directional decomposition (e.g. Fishman & McCoy 1984a,b) leads to a coupled system of one-way wave equations. Second, we use this system to obtain a so-called double-square-root (DSR) equation for the downward continuation of the data to selected depths in the subsurface (Clayton 1978; Claerbout 1985), simultaneously for the source and the receiver side (Section 3). Third, after downward continuation, the decomposed wavefields are subjected to a (wave-equation) angle transform, which enables the exploitation of redundancy associated with different scattering angles and azimuths at these depths (Section 4.1). This transform, which is related to beamforming (Scherbaum *et al.* 1997), is used to generate common-image-point gathers (De Bruin *et al.* 1990); in Sava *et al.* (1999) and Sava & Fomel (2003) a different but related transform implemented as a post-processing operator is used. These gathers can be re-arranged into multiple images of the same part of Earth's reflecting structure from waves scattered over a range of angles (Stolk & De Hoop 2006). Fourth, in Section 4.2 we discuss how the annihilation of the downward continued data (again, at specific depths in the medium) forms the criterion for selecting acceptable wave speed models. In Section 5 we derive the sensitivity kernels underlying the optimization of this wavefield annihilation.

In Section 6 we illustrate the multiscale aspects of the finite-frequency sensitivity kernel, the character of the cost functional used for optimization, and the effectiveness of the angle transform. We show that in heterogeneous media the dependence of the kernels on frequency is complex and not self-similar. Low-frequency data sample heterogeneous models in ways that are very different from high-frequency data. It is this finite frequency behaviour that allows one to relate scales in the data (frequencies) to scales in the model (spatial scales) and in the physical processes that cause them.

The image gathers produced by the angle transform are without artefacts (that is, false events) if the medium wave speeds are correct, even in the presence of caustics. (This was not the case in the approach followed by Brandsberg-Dahl *et al.* 2003, which is the precursor of the development presented here). For incorrect wave speed models, however, the image gathers will reveal dependency on scattering angle and azimuth; that is, there will be a residual moveout. This moveout is the underlying diagnostic exploited here and is evaluated using data annihilators (zero moveout and uniform amplitudes will produce perfect annihilation). The annihilation operators depend on the wave speed model and determine how well observations are matched by model predictions (Stolk & De Hoop 2002); a wave speed model is acceptable if the data are in the range of, that is, when they can be predicted by our modelling operator.<sup>1</sup> The tomographic inversion can then be formulated as the problem of minimizing the action of these annihilators on the data. An important and unique property of annihilators is that they *smoothly* depend on the wave speed (Stolk & Symes 2003).

We develop the theory up to the explicit expression for the Fréchet kernel underlying the annihilator approach (Section 5). The evaluation of this kernel is formulated as an adjoint state method. Such methods have been exploited in seismic imaging and inverse scattering (see, for example, Tarantola 1987; Stolk *et al.* 2006) and have begun to be used in wave equation transmission tomography (Tromp *et al.* 2005). Once the kernel is available, standard optimization methods (such as conjugate gradients) can be invoked to select a particular wave speed model. We illustrate the method and its effectiveness with a synthetic data example. For simplicity of formulation we consider here acoustic (compressional) waves, possibly in transversely isotropic media with vertical symmetry axis, described by a scalar wave equation (see, e.g. Alkhalifah 1998; Schoenberg & De Hoop 2000), and we use the flattened Earth assumption. There are no obstructions, however, for extending the formalism to the full elastic case.

The annihilator-based approach to wave-equation reflection tomography (or to MVA) is akin to the notion of differential semblance in exploration seismics. Systematic methods for updating wave speed models by optimization using the differential semblance criterion have been introduced by Symes & Carazzone (1991), and further developed and applied by Chauris & Noble (2001), Mulder & Ten Kroode (2002), and others. Wave-equation MVA based on focussing in subsurface offset (related to annihilators of image gathers, as explained in Stolk & De Hoop 2001, 2006) and short record migration rather than downward continuation migration was more recently developed by one of the authors (Peng 2004). The work presented here builds on various previous developments: The generalized Bremmer coupling series to model seismic reflection data (De Hoop 1996), inverse scattering based upon the first-order term of this series, the wave-equation angle transform and annihilators (Stolk & De Hoop 2001, 2006), the relation between wavefield reciprocity and optimization (De Hoop & De Hoop 2000), and the generalized screen expansion for one-way wave propagation (De Hoop *et al.* 2000; Le Rousseau & De Hoop 2001).

## 2 DIRECTIONAL WAVEFIELD DECOMPOSITION: SINGLE SQUARE-ROOT OPERATOR

In our method for wave-equation reflection tomography the measure for selecting acceptable models is the successful annihilation of the observed wavefield. The annihilating action is, here, carried out—at selected depths in the model—on image gathers that are constructed from

<sup>1</sup> The principle of characterizing the range of modelling operators can be found in Guillemin & Uhlmann (1981) and has been developed for seismic body-wave scattering by De Hoop & Uhlmann (2005).

downward continued data (Section 4). The downward continuation of the observed wavefield involves decomposition into up and downgoing constituents. For this purpose we introduce here the one-way wave equations and their associated wavefield propagators. Initially, we carry out the directional decomposition for wave propagation in a smooth background model (for the general formulation, see Fishman & McCoy 1984a,b); in later sections we account for scattering of waves off reflectors superimposed on this background.

## 2.1 The wave equation, evolution in depth

We make depth  $z$ —instead of time  $t$ —the evolution parameter for wave propagation and data continuation. The horizontal coordinates are collected in  $x$ . The partial derivatives are denoted by  $D_{x_1, \dots, x_{n-1}} = -i\partial_{x_1, \dots, x_{n-1}}$ ,  $D_z = -i\partial_z$ ,  $D_t = -i\partial_t$  so that their Fourier domain counterparts become multiplications by  $\xi_{1, \dots, n-1}$  (the negative of horizontal wave vector),  $\zeta$  (the negative of vertical wave vector) and  $\omega$  (frequency). Here,  $n = 2$  or  $3$  for 2-D or 3-D seismics, respectively. An alternative notation for the wave vector used in the seismic literature is  $(k_x, k_y, k_z)$  for  $(-\xi_1, -\xi_2, -\zeta)$  ( $n = 3$ ).

To bring out the use of  $z$  as the evolution parameter, the scalar wave equation for compressional waves,  $\Delta u - c_0(x, z)^{-2}\ddot{u} \equiv (\sum_{i=1}^{n-1} \partial_{x_i}^2 + \partial_z^2 - c_0(x, z)^{-2}\partial_t^2)u = f$ , is written as the first-order system

$$\frac{\partial}{\partial z} \begin{pmatrix} u \\ \frac{\partial u}{\partial z} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -L(x, z, D_x, D_t) & 0 \end{pmatrix} \begin{pmatrix} u \\ \frac{\partial u}{\partial z} \end{pmatrix} + \begin{pmatrix} 0 \\ f \end{pmatrix}, \quad (1)$$

where  $u$  is a scalar wavefield quantity such as pressure, the partial differential operator  $L(x, z, D_x, D_t) = c_0(x, z)^{-2}D_t^2 - \sum_{i=1}^{n-1} D_{x_i}^2$  (see, e.g. Aki & Richards 2002, p. 25),  $c_0$  is the wave speed in the (background) medium, and  $f$  is a source term. (N.B. here we take for  $f$  a volume injection source, but a general body force can be treated in a similar manner.) With  $\exp[i(\xi x + \omega t)]$  representing a Fourier factor in a plane wave,  $L(x, z, \xi, \omega)$  is defined as

$$L(x, z, \xi, \omega) = \exp[-i(\xi x + \omega t)]L(x, z, D_x, D_t)\exp[i(\xi x + \omega t)] = c_0(x, z)^{-2}\omega^2 - \|\xi\|^2. \quad (2)$$

For space dimension  $n > 2$ ,  $\xi x$  is a vector dot product. We note that, for given  $z$ ,  $(x, t, \xi, \omega)$  are coordinates on phase space, and eq. (2) defines a function on this space. For wave propagation in heterogeneous media, carrying out the analysis in phase space has important advantages over analysis in more conventional  $(n + 1)$ -D space–time. Heuristically, it allows a description of waves in terms of locally plane waves. The explicit definition and construction of the one-way wave equations below, their corresponding fundamental solutions (or propagators) in the form of path integrals, and the marching computational algorithms are all rooted in phase space (Fishman *et al.* 1987; Fishman 2004).

## 2.2 The system of one-way wave equations

In phase space, the roots of  $L(x, z, \xi, \omega)$  correspond, locally, to transitions from propagating ( $L > 0$ ) to evanescent ( $L < 0$ ) wave constituents. Away from these roots, in particular in the propagating regime, system (1) can be transformed into diagonal form (e.g. Stolk & De Hoop 2005). With D and U denoting the down- and upward propagating constituents of the wavefield, respectively, operator matrices  $Q(z) = Q(x, z, D_x, D_t)$  can be constructed such that

$$\begin{pmatrix} u_D \\ u_U \end{pmatrix} = Q(z) \begin{pmatrix} u \\ \frac{\partial u}{\partial z} \end{pmatrix}, \quad \begin{pmatrix} f_D \\ f_U \end{pmatrix} = Q(z) \begin{pmatrix} 0 \\ f \end{pmatrix},$$

effectively satisfy the one-way wave equations

$$\left( \frac{\partial}{\partial z} + iB_D(x, z, D_x, D_t) \right) u_D = f_D, \quad \left( \frac{\partial}{\partial z} - iB_U(x, z, D_x, D_t) \right) u_U = f_U. \quad (3)$$

Any coupling between D and U constituents is of the form  $Q(z)\frac{\partial}{\partial z}Q(z)^{-1}$ .

In the propagating wave regime, the operator  $B$  ( $= B_U$  or  $B_D$ ) admits an integral representation of the type

$$(Bu)(x, t, z) = (2\pi)^{-n} \int \int \int b(x', z, \xi, \omega) \exp[i\xi(x - x')] \exp[i\omega(t - t')] u(x', t', z) dx' dt' d\xi d\omega. \quad (4)$$

Here  $b(x', z, \xi, \omega)$  is a smooth function on phase space<sup>2</sup> and is, hence, better suited for manipulations (such as taking the wave speed derivatives needed below) than the operator  $B$  itself. For system (1), for high frequencies,<sup>3</sup>  $b = b(x, z, \xi, \omega) = \omega \sqrt{\frac{1}{c_0(x, z)^2} - \omega^{-2}\|\xi\|^2}$ , and for  $\omega > 0$ ,  $b$  coincides with  $\sqrt{L(x, z, \xi, \omega)}$  while for  $\omega < 0$ ,  $b$  coincides with  $-\sqrt{L(x, z, \xi, \omega)}$ . For this reason,  $B$  is often referred to as the ‘single-square-root operator’.

<sup>2</sup> In the mathematics literature  $b(x', z, \xi, \omega)$  would be referred to as the (antistandard) symbol of operator  $B$ . Likewise,  $L(x, z, \xi, \omega)$  in eq. (2) is the symbol of  $L(x, z, D_x, D_t)$ . We deviate from the calculus of standard symbols to facilitate the development of our preferred representation for the sensitivity kernel.

<sup>3</sup> For high frequencies the symbol of  $B$  reduces to what is called the principal symbol. For simplicity, we will denote the principal symbol of  $B$  by  $b$  from now on. We emphasize, however, that the principal symbol expression for  $b$  is not exact. The frequency dependence of operator  $B$  was discussed by Fishman & McCoy (1984a).

With  $b$  representing the vertical wavenumber,  $\omega^{-1}b$  has the appearance of a vertical wave slowness at the point  $(x, z)$  whereas  $\omega^{-1}\xi$  is a horizontal slowness. For  $(x, z, \xi, \omega)$  such that  $b$  is real, the one-way wave equations are of hyperbolic type, describing propagating waves.<sup>4</sup> (We remark that even though the separation between locally propagating and evanescent waves is, here, based upon the analysis of principal symbols, the basic picture of locally propagating and evanescent phase space regimes remains intact for a more complete description of the symbols.)

In the frequency domain, we introduce the operator notation  $\hat{B}$  in accordance with  $Bu = F_{\omega \rightarrow t}^{-1} \hat{B} F_{t \rightarrow \omega} u$  where  $F$  denotes the Fourier transform.<sup>5</sup> We note that  $\hat{B} = \hat{B}(x, z, D_x, \omega)$ .

We choose a normalization for operator  $Q(z)$  such that eq. (3) is self-adjoint; thus

$$Q = \frac{1}{2} \begin{pmatrix} (Q_D^*)^{-1} & -\mathcal{H}Q_D \\ (Q_U^*)^{-1} & \mathcal{H}Q_U \end{pmatrix}, \tag{5}$$

where  $*$  denotes the adjoint,  $\mathcal{H}$  denotes the Hilbert transform in time, and  $Q_{D,U} = Q_{D,U}(z) = Q_{D,U}(x, z, D_x, D_t)$  are operators with principal symbols  $(\frac{\omega^2}{c_0(x,z)^2} - \|\xi\|^2)^{-1/4}$ . The physical meaning of this choice of  $Q_{D,U}$  is that the down and upgoing fields are normalized in vertical-acoustic-power flux (De Hoop 1996). With this normalization, the coupling  $Q(z) \frac{\partial}{\partial z} Q(z)^{-1}$  between D and U constituents becomes asymptotically of lower order and, therefore, is neglected here; we get

$$u = Q_D^* u_D + Q_U^* u_U, \quad \begin{aligned} f_D &= -\frac{1}{2} \mathcal{H} Q_D f \\ f_U &= \frac{1}{2} \mathcal{H} Q_U f \end{aligned} \tag{6}$$

with  $Q_U^* u_U$  representing the up- and  $Q_D^* u_D$  representing the downgoing constituent of wavefield  $u$ .

### 2.3 One-way wave propagators

If singularities propagate nowhere horizontally between  $z$  and  $z_0$ , with  $z < z_0$ , there is a well-defined solution operator  $G_U(z, z_0)$  of the initial value problem for  $u_U$  given by eq. (3) with  $f_U = 0$  (Stolk 2004); this operator describes propagation in the (upward) direction from  $z_0$  to  $z$  (Stolk & De Hoop 2005, Section 2). (An alternative description of one-way propagators, in which is the issue of nowhere horizontal rays is suppressed, can be found in Fishman *et al.* 1997; De Hoop & Gautesen 2000). In the *Intermezzo* in Section (4.1) we briefly indicate how the full-wave solution can be used as well. A solution for the inhomogeneous eq. (3) is then given by

$$u_U(x, t, z) = \int_z^\infty \int \int (G_U(z, z_0))(x, t - t_0, x_0) f_U(x_0, t_0, z_0) dx_0 dt_0 dz_0, \tag{7}$$

or, in operator form by

$$u_U(\cdot, z) = \int_z^\infty G_U(z, z_0) f_U(\cdot, z_0) dz_0. \tag{8}$$

Here, the Green's function  $(G_U(z, z_0))$ , relating the wavefield at  $(x, t)$  and depth  $z$  to the force at  $(x_0, t_0)$  at depth  $z_0$ , is the kernel of one-way propagator  $G_U(z, z_0)$ .<sup>6</sup> The *adjoint*  $G_U(z, z_0)^*$  describes downward propagation from  $z$  to  $z_0$  of eq. (3) or upward from  $z_0$  to  $z$  in reversed time.<sup>7</sup>

### 2.4 Model representation and perturbation

For the optimization underlying our reflection tomography we need the (Fréchet) derivatives of the above mentioned square-root operators with respect to the background wave speed  $c_0(x, z)$ . Such derivatives, which are derived in Appendix A, are typically expressed in terms of a perturbation  $\delta c_0$  and can be interpreted as follows. We consider a finite-dimensional subspace of (smooth) wave speed models  $c_0$ , spanned by a finite set of basis functions  $\{\phi_k\}$ :

$$E_c : (c_k) \rightarrow c_0(x, z) = \sum_{k=1}^M c_k \phi_k(x, z) = E_c(c_k)(x, z).$$

This subspace could be generated, for example, by expanding the wave speed into cubic  $B$ -splines. The derivatives in eq. (A4) can then be expressed in terms of derivatives with respect to the coordinates  $\{c_k\}$  through  $\delta c_0(x, z) = \sum_{k=1}^M \delta c_k \phi_k(x, z)$ . The adjoint  $E_c^*$  of  $E_c$ , which satisfies

$$\langle E_c(c_k), c_0 \rangle_{(x,z)} = \langle (c_k), (E_c^* c_0) \rangle_{IR^M},$$

<sup>4</sup> In the regime where  $b$  is real, the principal symbols of the one-way wave operators— $\zeta \pm b(x, z, \xi, \omega)$  up to a factor  $i$ —can be identified as Hamiltonians for down and upgoing waves, and can be used to trace rays with  $z$  as evolution parameter. The next order symbol needs to be accounted for to obtain the transport equation leading to the correct asymptotic solutions.

<sup>5</sup>  $F_{t' \rightarrow \omega} u = \int \exp(-i\omega t') u(t') dt'$ ,  $F_{\omega \rightarrow t}^{-1} \hat{u} = (2\pi)^{-1} \int \exp(i\omega t) \hat{u}(\omega) d\omega$ , and similarly for  $F_{x' \rightarrow \xi}^{-1}$  and  $F_{\xi \rightarrow x}^{-1}$  in dimension  $n - 1$ .

<sup>6</sup> The upgoing constituent 'U' of the full-wave Green's function is generated by  $\frac{1}{2} \mathcal{H} Q_U^*(z) G_U(z, z_0) Q_U(z_0)$  (cf. 6).

<sup>7</sup> We note that, even though the principal symbol approximation is a high-frequency approximation, it can still be used to obtain an approximation to the propagator revealing a wave behaviour.

is the projection (interpolation)

$$c_0 \rightarrow \left( \iint c_0(x, z) \phi_k(x, z) dx dz \right) = (E_c^* c_0)_k.$$

This adjoint appears in the application of the adjoint state method to evaluating the sensitivity kernel.

### 3 DATA CONTINUATION IN DEPTH: DOUBLE-SQUARE-ROOT OPERATOR

The background medium  $c_0$  (with smooth perturbations  $\delta c_0$ ) considered in the previous section can produce caustics but no scattering (such as reflections). We now describe how, with the Born approximation, single-scattered phases are modelled in the framework of the one-way wave theory developed above. Making use of reciprocity, we model the data purely by upcoming propagators. Whereas up and downgoing waves are modelled with a single square-root operator  $B$  (see Section 2.2), the interaction of these contributions (through a time convolution) yields a DSR operator.

#### 3.1 Upward continuation and modelling of reflection data

We assume wavefield scattering off a contrast  $\delta c$  that contains singular variations in medium wave speed (for instance, reflectors). The total medium wave speed follows the decomposition (with  $\delta c^{-2} \simeq -2c_0^{-3} \delta c$ )

$$c^{-2}(x, z) = \gamma_0^{-2}(z) + [c_0^{-2}(x, z) - \gamma_0^{-2}(z)] - 2c_0^{-3}(x, z) \delta c(x, z).$$

To obtain a downward/upward continuation representation of the single scattered waves, we introduce the extended medium contrast (Stolk & De Hoop 2005),

$$R(s', r', t', z') = \frac{1}{2} \delta(t') \delta(r' - s') (c_0^{-3} \delta c) \left( \frac{r' + s'}{2}, z' \right). \tag{9}$$

One can view  $R$  as a data-like representation of the contrast. Using the footnote below eq. (8) concerning the relation between one-way Green's functions and the Green's function of the original wave equation and using that  $\mathcal{H}^2 = -I$ , the singular part of the data (as recorded at Earth's surface, that is, at  $z = 0$ ) can be modelled as

$$d(s, r, t) = Q_{U,s}^*(0) Q_{U,r}^*(0) \int_0^\infty (H(0, z') Q_{U,s}(z') Q_{U,r}(z') D_t^2 R(\cdot, z'))(s, r, t) dz', \tag{10}$$

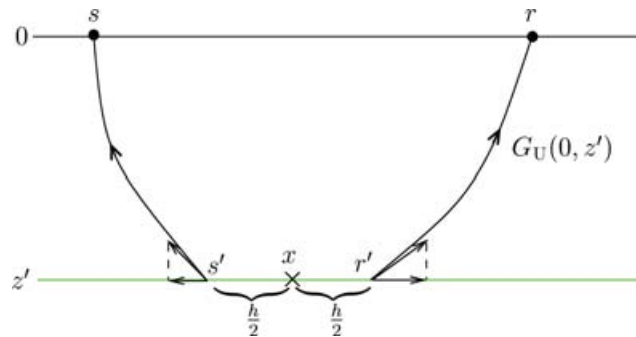
where  $Q_{U,s}(z)$  is short for  $Q_U(s, z, D_s, D_t)$  and  $Q_{U,r}(z)$  is short for  $Q_U(r, z, D_r, D_t)$ , and the kernel of  $H(z, z')$  is given by the time convolution

$$(H(z, z'))(s, r, t, s', r', t') = \int_{IR} (G_U(z, z'))(s, t - t' - \bar{t}, s', 0) (G_U(z, z'))(r, \bar{t}, r', 0) d\bar{t}, \quad \text{for } z > z'. \tag{11}$$

Through  $H$  as in eq. (11), we recognize in eq. (10) the time convolution (integration over  $\bar{t}$ ) of an incoming Green's function that connects the source at  $s$  (at zero depth) with the scattering point at  $r' = s'$  (at depth  $z'$ ) with the outgoing Green's function that connects this scattering point with the receiver at  $r$  (at zero depth). This process gives rise to a DSR propagator; the spatial notation is illustrated in Fig. 2 (see also Stolk & De Hoop 2005).

The data, in eq. (10), can be regarded as the solution to a (Cauchy) initial value problem: at the surface ( $z = 0$ ) the single scattered wavefield  $u$  represents the data  $d$ , that is,  $d(s, r, t) = Q_{U,s}^*(0) Q_{U,r}^*(0) u(s, r, t, z = 0)$ . For  $z > 0$ ,  $u = u(s, r, t, z)$  represents the (decomposed) wavefield—treated as ‘virtual’ data—as a function of time, generated at depth  $z$  below the surface by fictitious sources  $s$  and observed by fictitious receivers  $r$ :

$$u(s, r, t, z) = \int_z^\infty (H(z, z') Q_{U,s}(z') Q_{U,r}(z') D_t^2 R(\cdot, z'))(s, r, t) dz'.$$



**Figure 2.** DSR propagator (eq. 11). Data at depth  $z'$  are propagated towards the surface at  $z = 0$ ;  $s'$  and  $r'$  indicate fictitious source and receiver locations whereas  $s$  and  $r$  indicate actual source and receiver locations. The subsurface offset  $h$  and midpoint  $x$  occurring in the angle transform are indicated as well.

Then,  $u$  solves an inhomogeneous DSR equation

$$\left(\frac{\partial}{\partial z} - iB_U(s, z, D_s, D_t) - iB_U(r, z, D_r, D_t)\right)u = g, \quad g = Q_{U,s}(z)Q_{U,r}(z)D_t^2 R(\cdot, z). \quad (12)$$

This equation is solved in the (upward) direction of decreasing  $z$  (with vanishing initial condition for some large  $z$ , say, at the bottom of the model).

The DSR equation plays a role later in the evaluation of the sensitivity kernel for reflection tomography. In fact,  $H$  is the propagator associated with eq. (12), and is hence referred to as the DSR propagator. For later use, we introduce the DSR operator

$$C(s, r, z, D_s, D_r, D_t) = B_U(s, z, D_s, D_t) + B_U(r, z, D_r, D_t), \quad (13)$$

so that eq. (12) attains the form  $\left(\frac{\partial}{\partial z} - iC(s, r, z, D_s, D_r, D_t)\right)u = g$ . The DSR operator  $C$  depends on the background model, and a perturbation  $\delta c_0$  in  $c_0$  yields a perturbation  $\delta C$  in  $C$ , which is derived from the results in Appendix A, Subsection A2.<sup>8</sup>

### 3.2 Downward continuation and imaging of reflection data

The *adjoint* operator  $H(0, z)^*$  is used to propagate the data backward to depth  $z$ . We consider

$$\psi d \mapsto D = H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d, \quad (14)$$

where  $\psi$  is a (source, receiver, and time) taper that suppresses the parts of the data that one does not want to consider in the imaging process (such as data associated with turning rays). In eq. (14),  $D$  is a function of  $(s, r, t, z)$ ; for  $t > 0$ ,  $D(s, r, t, z)$  can be identified with  $u(s, r, t, z)$  in eq. (12), and is sometimes referred to as a ‘sunken survey’. Indeed,  $D$  can be obtained by solving the evolution equation

$$\left(\frac{\partial}{\partial z} - iC(s, r, z, D_s, D_r, D_t)^*\right)u = 0, \quad (15)$$

in the direction of increasing  $z$  (downward), subject to the initial condition  $u(s, r, t, z = 0) = -Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d$ . In the propagating regime,  $C(s, r, z, D_s, D_r, D_t)^* = C(s, r, z, D_s, D_r, D_t)$ .

Following Claerbout (1985), an image of  $\delta c$ —that is the singular wave speed variations (including reflectors)—can be obtained from  $D$  by applying imaging conditions:

$$I(x, z) = D\left(s = x - \frac{h}{2}, r = x + \frac{h}{2}, t, z\right) \Big|_{h=0, t=0}, \quad (16)$$

where we have introduced subsurface horizontal midpoint-offset coordinates,  $(x, h)$ . Our tomography aims to estimate wave speed variations in the background, but we use image gathers to evaluate associated model updates. These image gathers are obtained by replacing the imaging condition by an angle transform (see the next section).

## 4 WAVE-EQUATION ANGLE TRANSFORM AND DATA ANNIHILATORS

Building on the work by Stolk & De Hoop (2001) and Brandsberg-Dahl *et al.* (2003) we discuss here how redundancy in the data can be exploited for the purpose of tomography. We assume that we have data from multiple sources and multiple receivers (see Fig. 1, bottom). In the imaging process, the reflection point becomes the image point, and the redundancy in the data becomes manifest in the multiple images that are generated (at that image point) from the different scattering angles and azimuths. This is accomplished by the angle transform. In this transform, we encounter a variable  $p$  that is essentially half the difference between the horizontal slowness vectors associated with source and receiver rays at the depth where the transform is applied. The relation between  $p$  and scattering angle and azimuth is described by De Hoop *et al.* (2003, eqs 88–90); by virtue of this relation, in imaging, the angle transform can be thought of as a replacement of the generalized Radon transform.

For an acceptable wave speed model the images are independent of scattering angle and azimuth. Equivalently, the wave speed model is acceptable if the data are predictable by (in other words, they are in the range of) the modelling operator in eq. (10). In this section, we derive the annihilators used to evaluate this criterion. (We emphasize that the notion that a model is acceptable does *not* imply that it is uniquely determined; in fact, there typically is a class of acceptable models. It can be shown, however, that the higher the scatterer density, the smaller the collection of acceptable models).

<sup>8</sup> For completeness sake, in Appendix B we show that generalized screen expansions for  $Q_{U,s}(z)$  and  $Q_{U,r}(z)$  lead to wave speed derivatives in a fashion similar to the ones for  $B_U$ . These derivatives are suppressed in the treatment presented here, however.

#### 4.1 Angle transform and common image-point gathers

With, as before,  $u = u(s, r, t, z)$ , and  $h$  the horizontal offset between  $r$  and  $s$  (Fig. 2), we introduce

$$\begin{aligned} (Ru)(x, z, p) &= \int_{\mathbb{R}^{n-1}} u\left(x - \frac{h}{2}, x + \frac{h}{2}, ph, z\right) \chi(x, z, h) dh \\ &= \frac{1}{2\pi} \int \int_{\mathbb{R}^{n-1}} \hat{u}\left(x - \frac{h}{2}, x + \frac{h}{2}, \omega, z\right) \exp(-i\omega ph) \chi(x, z, h) dh d\omega, \end{aligned} \quad (17)$$

(note the zero intercept time:  $t = 0 + ph$ , with  $ph$  a multiplication for  $n = 2$  and a dot product for  $n \geq 3$ ; this transform was originally introduced by De Bruin *et al.* 1990). Here,  $\chi(x, z, h)$  is a taper in  $h$ , such that  $\chi(x, z, 0) = 1$ . We apply  $R$  to the downward continued data,  $D(s, r, t, z)$  (cf. eq. 14), to obtain the wave-equation angle transform,  $\mathcal{A}$ ,

$$\mathcal{A}d = RH(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d. \quad (18)$$

The result of this transform applied to the data is

$$\bar{l}(x, z, p) = (\mathcal{A}d)(x, z, p) = (RD)(x, z, p). \quad (19)$$

For each  $x$ ,  $\bar{l}(x, z, p)$  is a common image-point gather in  $(z, p)$ .<sup>9</sup> With an appropriate choice of  $\chi$  the common image-point gathers are artefact free (that is, free of false events), even in the presence of caustics (Stolk & De Hoop 2001, 2006). As mentioned above, the dependence on  $p$  reflects the redundancy in the data (a dimension count confirms this:  $(s, r, t) \in \mathbb{R}^{2n-1}$  while  $(x, z) \in \mathbb{R}^n$  and  $2n - 1 - n = n - 1$  is the dimension of  $p$ ). For an acceptable model  $c_0$ ,  $\mathcal{A}$  maps data  $d$  to a  $p$ -family of reconstructions of  $\delta c$ , and for each  $p$  the same image ( $\bar{l}(x, z, p)$ ) of  $\delta c(x, z)$  is obtained.

#### 4.2 Intermezzo: Angle transform for global Earth applications

In global Earth applications, in particular in view of turning rays, the one-way wave propagators introduced in Section 2 may be replaced by full-wave propagators. According to eq. (12), the (fictitious) sources  $s$  and receivers  $r$  are at the same depth ( $z$ ). For application to earthquake data, however, we consider (clusters) of earthquakes with hypocentres at  $z = z_s$  and stations at  $z = 0$  (see Fig. 1, bottom). Then, we can replace the kernel  $(H(0, z'))(s, r, t, s', r', t')$  of  $H(0, z')$  by

$$(H(0, z'))(s, r, t, s', r', t') := \int_{\mathbb{R}} G(s, z_s, t - t' - \bar{t}, s', z') G(r, 0, \bar{t}, r', z') d\bar{t}. \quad (20)$$

In view of reciprocity the kernel of the adjoint is the same, so that the angle transform becomes

$$D\left(x - \frac{h}{2}, x + \frac{h}{2}, t', z\right) = \int \left( \int_{\mathbb{R}} G\left(s, z_s, -(t' - t + \bar{t}), x - \frac{h}{2}, z\right) G\left(r, 0, \bar{t}, x + \frac{h}{2}, z\right) d\bar{t} \right) (\psi d)(s, r, t) ds dr dt, \quad (21)$$

followed by eq. (17); here,  $z > z_s$  for all  $s$ . We can use any numerical method to compute the Green's functions, for example normal mode summation, leading then to a 'normal mode angle transform'. Eq. (21) can be rewritten, upon a change of variable of integration  $\bar{t} = -(\bar{t} - t)$ , in the form

$$D\left(x - \frac{h}{2}, x + \frac{h}{2}, t', z\right) = \int \left\{ G\left(s, z_s, \bar{t} - t', x - \frac{h}{2}, z\right) \left\{ \int G\left(r, 0, -(\bar{t} - t), x + \frac{h}{2}, z\right) (\psi d)(s, r, t) dr dt \right\} d\bar{t} \right\} ds; \quad (22)$$

at  $t' = 0$  and  $h = 0$ , this form reveals the structure of shot record migration, while one recognizes the notion of double focussing (Berkhout 1997; Thorbecke 1997).

Essentially, eq. (17) is a special case of beamforming with the downward continued data (cf. 18) in sources and receivers (see also Scherbaum *et al.* 1997).

#### 4.3 Annihilators of the downward continued data

Since the outcome  $(\mathcal{A}d)(x, z, p)$  should be independent of  $p$  we can define annihilators  $W$ , whose action on the data  $d$  is to yield zero, as follows: for  $i = 1, \dots, n-1$ ,  $W_i = (\frac{\partial}{\partial t})^{-1} \langle \mathcal{A}^{-1} \rangle \frac{\partial}{\partial p_i} \mathcal{A}$  (where  $\langle \mathcal{A}^{-1} \rangle$  indicates a regularized inverse of  $\mathcal{A}$ ), which indeed yields  $W_i d = 0$ . We consider the annihilators not in the data but in the image domain. Therefore, removing the mapping  $\langle \mathcal{A}^{-1} \rangle$  from image gathers to data, we consider the companion operators  $\frac{\partial}{\partial p_i} \mathcal{A} (\frac{\partial}{\partial t})^{-1}$  (note that  $(\frac{\partial}{\partial t})^{-1}$  has to act in the data domain), which, when applied to eq. (17), brings out a

<sup>9</sup> Sava *et al.* (1999) use a related but different transform, viz.,  $\bar{l}(x, Z, p) = \int_{\mathbb{R}^{n-1}} D(x - \frac{h}{2}, x + \frac{h}{2}, t, z) \chi(x, z, h)|_{t=0, z=Z+ph} dh$ . Unlike  $\mathcal{A}$ , this transform cannot be cast, with appropriately chosen weights, into an estimation of the reflection coefficient induced by  $c_0, \delta c$ .



multiplication by factor  $h_i$ :

$$\begin{aligned} (R'_i u)(x, z, p) &= \int_{\mathbb{R}^{n-1}} u\left(x - \frac{h}{2}, x + \frac{h}{2}, ph, z\right) h_i \chi(x, z, h) dh \\ &= \frac{1}{2\pi} \int \int_{\mathbb{R}^{n-1}} \hat{u}\left(x - \frac{h}{2}, x + \frac{h}{2}, \omega, z\right) \exp(-i\omega ph) h_i \chi(x, z, h) dh d\omega. \end{aligned} \quad (23)$$

The annihilation of the data is thus replaced by an annihilation of the set of subsurface image gathers,  $R'_i D$ , parametrized by  $p$ .

For the purpose of tomography we also need the *adjoint*  $(R'_i)^*$  of  $R'_i$ . Let  $l$  denote a trial image as a function of  $(x, z, p)$ , then

$$\begin{aligned} \langle R'_i u, l \rangle_{(x,p)} &= \int_{\mathbb{R}^{2n-2}} \left( \frac{1}{2\pi} \int \int_{\mathbb{R}^{n-1}} \hat{u}\left(x - \frac{h}{2}, x + \frac{h}{2}, \omega, z\right) \exp(-i\omega ph) h_i \chi(x, z, h) dh d\omega \right) l(x, z, p) dx dp \\ &= \frac{1}{2\pi} \int \int_{\mathbb{R}^{2n-2}} \hat{u}(s, r, \omega, z) \\ &\quad \left( \int_{\mathbb{R}^{n-1}} (r-s)_i \exp[i\omega p(r-s)] l\left(\frac{1}{2}(s+r), z, p\right) dp \right) \chi\left(\frac{1}{2}(s+r), z, r-s\right) ds dr d\omega \\ &= \langle \hat{u}, (\hat{R}'_i)^* l \rangle_{(s,r,\omega)} \text{ for given } z, \end{aligned} \quad (24)$$

so that

$$(\hat{R}'_i)^* l(s, r, \omega, z) = \int_{\mathbb{R}^{n-1}} (r-s)_i \exp[i\omega p(r-s)] l\left(\frac{1}{2}(s+r), z, p\right) dp. \quad (25)$$

Here,  $\overline{\phantom{x}}$  indicates complex conjugation. We note that  $(\hat{R}'_i)^*$ , in eq. (24), yields an extension of a (differentiated) image gather to fictitious data, in the nature of  $R$  defined in eq. (9). In eq. (24) we changed variables of integration from  $(x, h)$  to  $(s, r)$ . By removing the factor  $h_i$  in the integrand, we immediately obtain an expression for  $\hat{R}^*$ . By Parseval's theorem, we also find  $R^*$ ,

$$(R^* l)(s, r, t, z) = \int_{\mathbb{R}^{n-1}} \delta(t - p(r-s)) l\left(\frac{1}{2}(s+r), z, p\right) dp$$

with the property that  $\langle Ru, l \rangle_{(x,p)} = \langle u, R^* l \rangle_{(s,r,t)}$ . Notice that neither  $R$  nor  $R'_i$  depends on  $c_0$  and, thus, that they are insensitive to smooth perturbations  $\delta c_0$  in  $c_0$ .

## 5 AN OPTIMIZATION PROCEDURE FOR REFLECTION TOMOGRAPHY

With the reflection tomography developed here we aim to estimate the background medium ( $c_0 + \delta c_0$ ), with as a measure of success the annihilation of image gathers. Indeed, if annihilation with operator  $R'_i$  applied to the downward continued data in eq. (23) occurs, then the background medium is considered acceptable.

### 5.1 Cost functional

With this notion, the model estimation is cast into the minimization of the functional

$$\mathcal{J}[c_0] = \frac{1}{2} \sum_i \int \int |(R'_i H(0, z)^* Q_{\hat{u},s}^*(0)^{-1} Q_{\hat{u},r}^*(0)^{-1} \psi d)(x, z, p)|^2 dx dz dp, \quad (26)$$

which corresponds with the effective data annihilation integrated over all image (scattering) points  $(x, z)$  and 'angles'  $p$ . Here,  $H(0, z)^*$  (and  $\psi$ ) depend on  $c_0$  and are sensitive to perturbations  $\delta c_0$ ; however, since we assumed that  $\delta c_0$  vanishes near  $z = 0$ , we need not consider wave speed derivatives of  $Q_{\hat{u},s}^*(0)^{-1} Q_{\hat{u},r}^*(0)^{-1}$ . Compared with conventional tomography, the measure of traveltimes mismatch (or waveform cross correlation) has thus been replaced by the overall effect of the annihilation operator  $R'_i$ .

### 5.2 Gradient of cost functional; sensitivity kernel

A standard method for optimization can be invoked to carry out the minimization of  $\mathcal{J}$ . Here, we discuss a method for evaluating the Fréchet kernel or gradient of  $\mathcal{J}$ . The method for evaluating the gradient of a functional derived from the solution of a partial (or pseudo-) differential equation is known as the adjoint state method (Wunsch 1996). It has been introduced in seismology by Tarantola (1987) and used by many others. The gradient is required for the optimization, whereas the Fréchet or sensitivity kernel provides information about the resolution of our approach to wave-equation reflection tomography.

The perturbation of the functional  $\mathcal{J}$ , under a smooth perturbation  $\delta c_0$  of  $c_0$ , is derived from

$$\delta\mathcal{J}[c_0] = \int \left\{ \int \left( \sum_i (\hat{R}'_i)^* R'_i H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d \right) (s, r, \omega, z) \overline{(\delta H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d)} (s, r, \omega, z) ds dr d\omega \right\} dz, \quad (27)$$

and evaluated in Appendix C. The evaluation procedure, which is common to many different imaging schemes, consists of solving two evolution equations followed by an imaging operation. We use the observation made below eq. (14) and relate  $D$  to solutions,  $u$ , of the DSR equation. The first evolution equation is eq. (15), which is solved in the direction of increasing  $z$  (downward); in the frequency domain this equation reads:

$$\left( \frac{\partial}{\partial z} - i\hat{C}(s, r, z, D_s, D_r, \omega) \right) \hat{u} = 0, \quad \hat{u}(s, r, \omega, z = 0) = - (Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d)^\wedge (s, r, \omega). \quad (28)$$

The second is the adjoint field equation,

$$\left( \frac{\partial}{\partial z} - i\hat{C}(s, r, z, D_s, D_r, \omega) \right) \hat{v} = \sum_i (\hat{R}'_i)^* R'_i u, \quad (29)$$

which is solved in the (upward) direction of decreasing  $z$  (with vanishing initial condition for some large  $z$ , say, at the bottom of the model). The right-hand side quantifies the failure of annihilation and represents a mismatch source distribution.

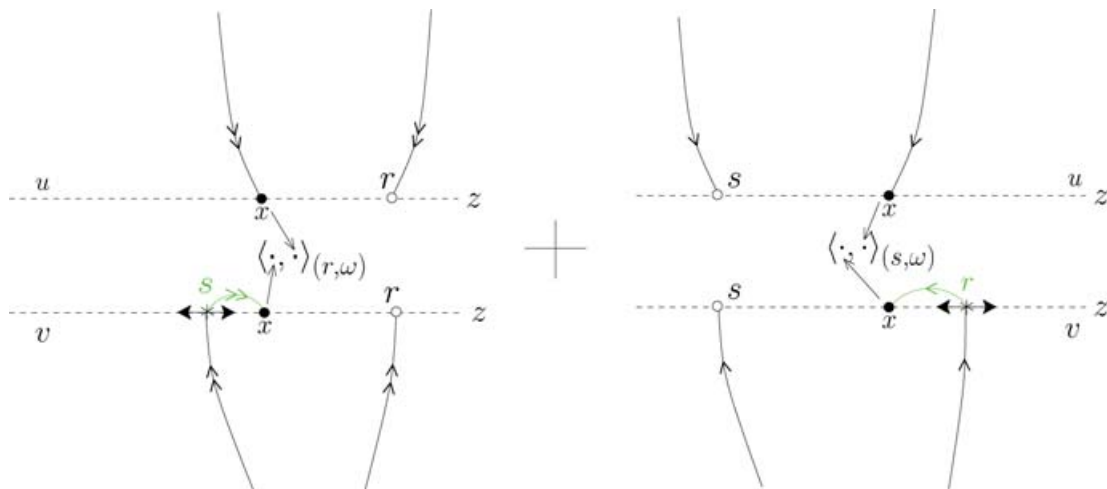
The solutions  $\hat{u}$  and  $\hat{v}$  combined form the kernel (or ‘image’)  $\mathcal{K}$  in  $\delta\mathcal{J} = \int \mathcal{K}(x, z) \delta c_0(x, z) d(x, z)$ , and is given by (cf. C12 and A7)

$$\mathcal{K}(x, z) \sim \frac{1}{\pi} \text{Re} \int_0^\infty d\omega (-i) \sum_{j=0}^N \left[ \int \overline{\hat{u}(x, r, \omega, z)} \underbrace{S'_j(x, z) F_{\sigma \rightarrow x}^{-1} A_j(\sigma, \omega, z) F_{s \rightarrow \sigma}}_{\text{generalized screen}} \hat{v}(., r, \omega, z) dr + \int \overline{\hat{u}(s, x, \omega, z)} \underbrace{S'_j(x, z) F_{\rho \rightarrow x}^{-1} A_j(\rho, \omega, z) F_{r \rightarrow \rho}}_{\text{generalized screen}} \hat{v}(s, ., \omega, z) ds \right]. \quad (30)$$

Here,  $S_j$  and  $A_j$  appear in the generalized-screen expansion of the single-square-root operator, which is explained in detail in Appendix A (cf. A1 and A4). We used the symmetry in frequency to restrict the evaluation to positive values. Invoking the model representation in Section 2.4 yields  $\delta\mathcal{J} = \langle \mathcal{K}, \delta c_0 \rangle_{(x,z)} = \langle (\delta c_k), (E_c^* \mathcal{K}) \rangle_{IRM}$ .

Eq. (30) is of the form of a time cross correlation of the downward continued data and the adjoint field excited by a mismatch force nested in a generalized screen operation. It differs from a standard imaging condition in particular through the integrations over  $s$  and  $r$ . The procedure to evaluate eq. (30) is illustrated in Fig. 3. It consists of the following steps:

- (i) starting at the top, downward continue step-by-step the data (in the frequency domain) all the way to the bottom (cf. 14) while storing the results at all intermediate depths;
- (ii) starting at the bottom, evaluate the success of annihilation, that is, the mismatch source, with the downward continued data,



**Figure 3.** The evaluation of the kernel. The single and double arrows refer to the two terms in between brackets in eq. (30); the grey arrows indicate the generalized screen operations. The field  $v$  is upward continued with sources determined by the failure to annihilate the data in the image domain;  $u$  represents the downward continued data. At any depth,  $z$ , a screen action is applied to  $v$  (indicated by grey arrows) in parallel in the source variable (left) and in the receiver variable (right). Then, in parallel, the inner products (including time correlation) are taken in receivers (left) and sources (right). Finally, the two contributions are added together.

- (iii) upward continue (in the frequency domain) step-by-step the adjoint field (cf. 29) and
- (iv) at each depth, evaluate the frequency contribution to the kernel (cf. 30).

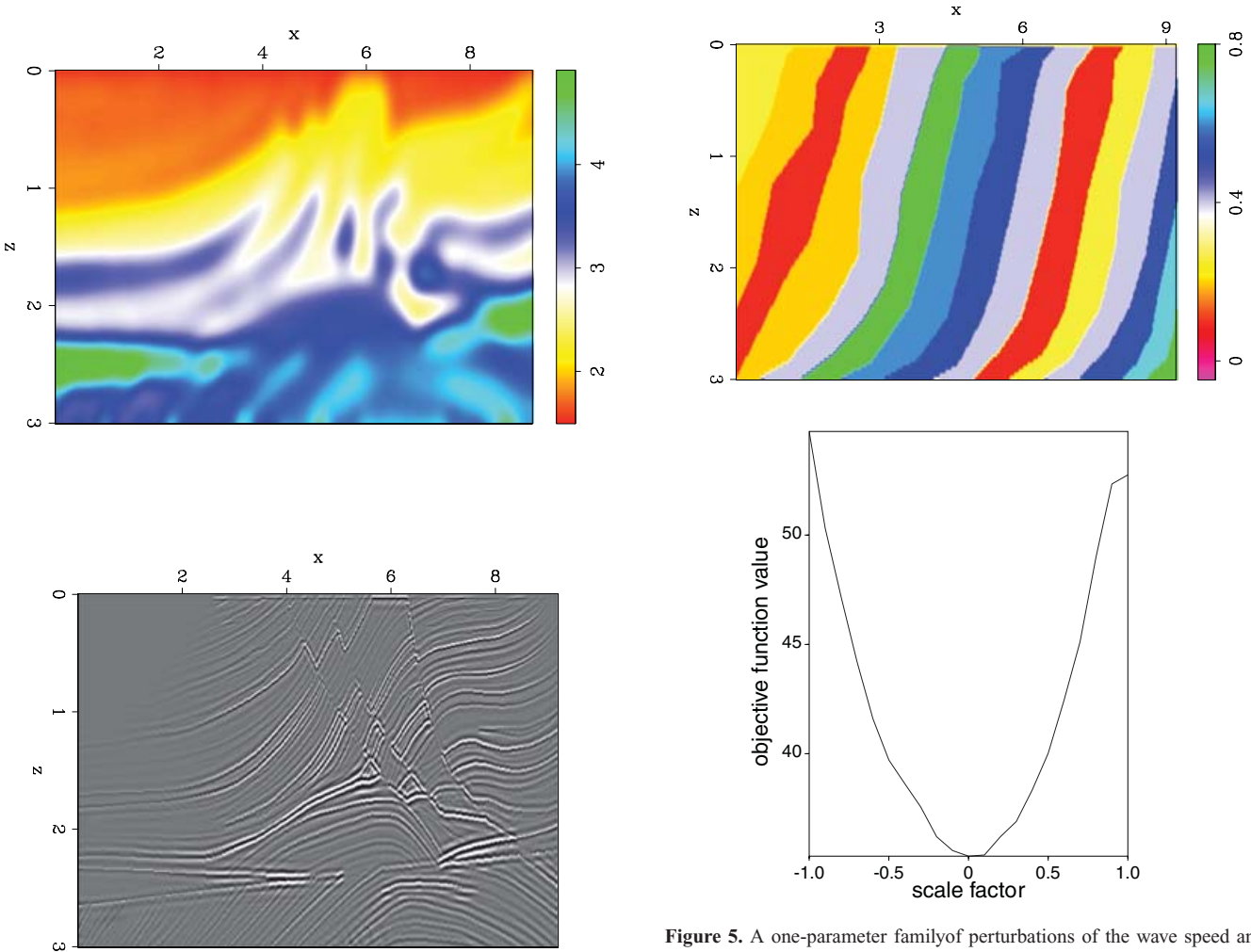
We stress that even though the kernel has been derived from generalized screen expansions of the relevant operators, the downward and upward continuation can be carried out with more precise and sophisticated methods.

In eq. (30), the frequency integral is kept as the outside integral, because the kernel computation is to be carried out in the  $\omega$ -domain. It becomes clear that subjecting the data to a timescale decomposition leads to a decomposition of the kernel. Such a decomposition can be accomplished by applying frequency window functions  $\hat{\psi}_k$ , corresponding with a wavelet transform at dyadic scales ( $2^k$ ), to the data in the right-hand side of eq. (28). The effect of this propagates through the right-hand side of eq. (29); we then make use of the observation that  $(R_i^t \psi_k)^*$  yields  $\hat{\psi}_k(R_i^t)^*$ . Thus, upon inserting in the  $\omega$  integration in eq. (30) the factor  $|\hat{\psi}_k(\omega)|^2$  we obtain the contribution  $\mathcal{K}_k$  to  $\mathcal{K}$ .

### 6 COST FUNCTIONAL AND SENSITIVITY KERNEL; A MULTISCALE PERSPECTIVE

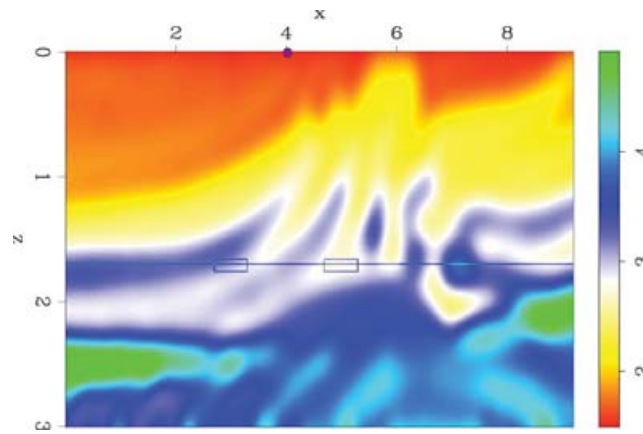
With synthetic data we illustrate here the effectiveness of the angle transform (eq. 19), the character of the cost functional used for optimization (eq. 26), and the multifrequency aspects of the sensitivity kernel (that is, the gradient, 30) associated with the annihilator-based method for wave-equation reflection tomography. Furthermore, through a wavelet-like decomposition ( $\psi_k$ ) in time frequency of the sources generating the data, we demonstrate the spatial multiscale aspects of the kernels underlying wave-equation reflection tomography.

As the background wave speed model we consider a smooth version of Marmousi (Fig. 4, top). For reference, in Fig. 4 (bottom) we show the image of the singular medium variations, obtained with the imaging condition in eq. (16). These singularities produce the scattered wavefield exploited in the annihilation procedure. This wavefield has been generated with a finite difference approach and forms the synthetic data. Furthermore, we consider a one-parameter family of model perturbations (that is, in Subsection 2.4 we take  $M = 1$ ), generated by the

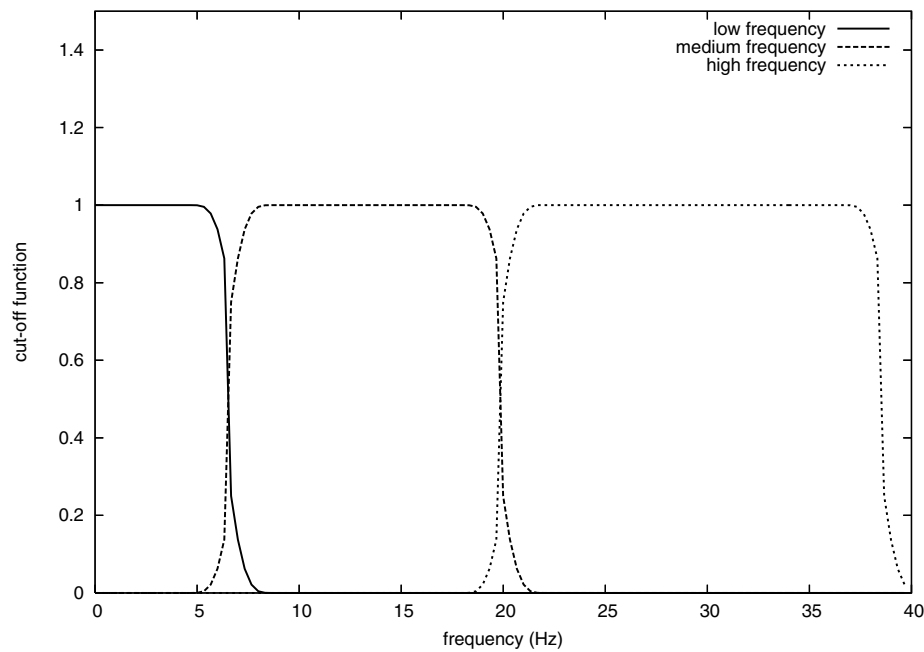


**Figure 4.** One-way wave-equation migration in a modified Marmousi model. Top: wave speed model ( $c_{\text{true}}$ ) used to generate the data by DSRE equation simulation (cf. 12). Bottom: Image obtained by wave-equation migration (cf. 16) using the true wave speed model in the Top.

**Figure 5.** A one-parameter family of perturbations of the wave speed and the associated objective function (the functional in 26) curve. Top: An inhomogeneous wave speed perturbation,  $\Delta c_0$ . Bottom: Functional evaluated at  $c_0 = \alpha \Delta c_0 + c_{\text{true}}$ , where  $\alpha$  is a scalar ranging from  $-1$  to  $1$ , and  $c_{\text{true}}$  is the true wave speed model shown in Fig. 4 (top), respectively.



**Figure 6.** The wave speed model of Fig. 4 (top); indicated are a source location (red dot), and a fictitious reflector (horizontal line, fixed  $z$ ) that contains two scattering points (indicated by boxes). The data from these source and scattering points are used to obtain the kernel contribution in Fig. 8.

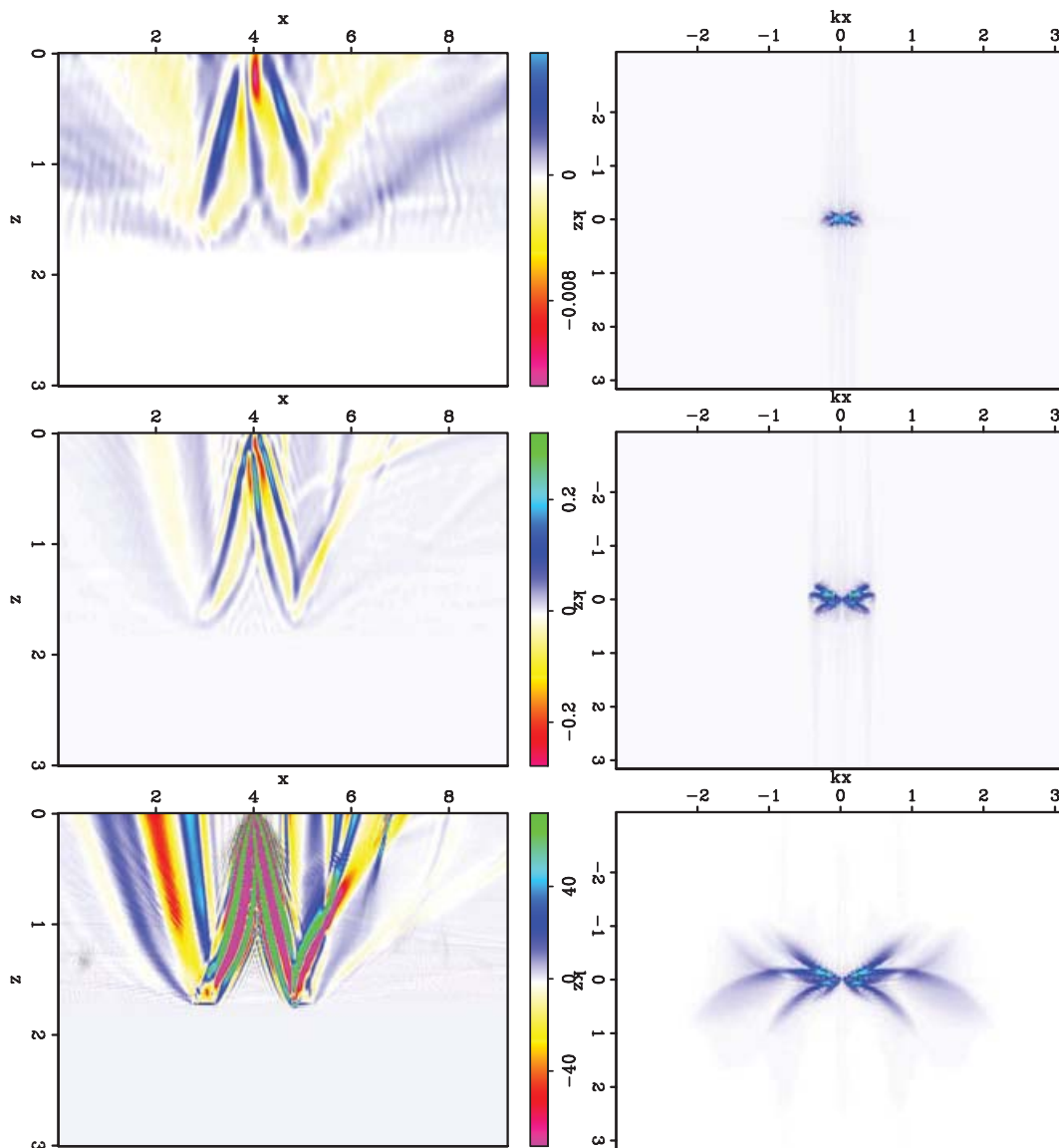


**Figure 7.** Amplitude spectra of  $\psi_k$ , representing, here, the multi frequency windowing applied to the data. We distinguish only three bands: low, middle and high frequencies. The windows can also be derived from a wavelet transform.

differential model ( $\phi_1(x, z)$ ), illustrated in Fig. 5 (top). In our examples, we use the generalized screen expansion only to find  $A_j$  and  $S'_j$  in eq. (30).

First, we evaluate kernel contributions (functions of  $(x, z)$ ) from data restricted to a single seismic source ( $s$ ) indicated by a solid red dot on Fig. 6, and reflections from two scattering points within the yellow boxes lying on a fictitious reflector indicated by a horizontal line in the same figure. The key result of this analysis is presented in Fig. 8. This figure reveals the complexity of the kernel resulting from finite-frequency (interference) effects and multipathing. An imprint of the rays connecting the two mismatch source points to the source and receivers determined by the reflector dip is apparent.

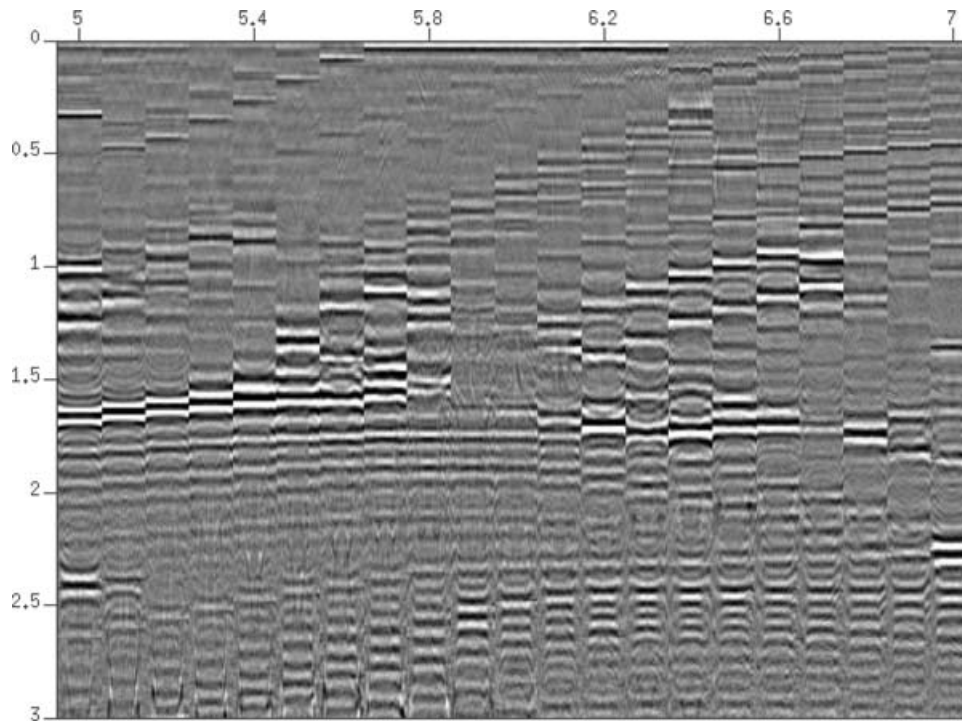
In order to appreciate the nature of finite frequency kernels and, thus, the multiscale aspects of wave-equation tomography, we subjected the (synthetic) data to low-, middle- and high-frequency windowing, according to Fig. 7. In Fig. 8(top, left), we illustrate the contribution to the kernel for low frequencies. The wavenumber spectrum of this contribution is displayed on the right. Fig. 8 (middle) and (bottom) illustrate the middle- and high-frequency contributions to the kernel. We observe how differently, through the wave dynamics, the contributions sample the subsurface. Indeed, the kernels are not self-similar, in that they are not simply scaled versions of one another, as the frequency window changes. In fact, the low-frequency data are more sensitivity to the smooth parts of the model in Fig. 4 (top), whereas the high-frequency data have (relatively) higher sensitivity to the fine scale structures in the model. (This is confirmed by viewing the wavenumber spectra of the different contributions to the kernel, in the panels on the right.) Different frequencies thus provide information on different



**Figure 8.** The contributions to the gradient (*cf.* 30) due to a pair of scattering points (indicated by boxes in Fig. 6) on a single reflector and data from a single source located at the red disk in Fig. 6. Top, left: low frequencies, coarse scale; top, right: wavenumber spectrum of this contribution. Middle, left: middle range frequencies, intermediate scale; middle, right: wavenumber spectrum of this contribution. Bottom, left: high frequencies, fine scale; bottom, right: wavenumber spectrum of this contribution. The frequency bands are given in Fig. 7. We note how differently the contributions sense the subsurface. We also note the multidirectional illumination of the wave speed model.

parts and different scales in the model. In regions where the medium varies only smoothly, the kernel reveals a (stereo) banana–doughnut type behaviour, in accord with analyses by, e.g. (Woodward 1992; Dahlen *et al.* 2000). In regions of stronger heterogeneity, however, the kernels are rich in complexity, and include caustics and multipathing, as predicted by, e.g. (De Hoop & Van der Hilst 2005; Zhao *et al.* 2005).

In Fig. 5 (bottom) we illustrate the local convexity and smoothness (expected on the basis of the work of Stolk & Symes (2003)) of the cost functional by considering the above mentioned (one-parameter family of) perturbations of the model depicted in Fig. 4 (top). In Figs 9–11 we show that the annihilator does indeed detect the error in background wave speed. Fig. 9 shows the result of the application of the angle transform to the downward continued data using the true model; the uniformity of and zero residual moveout in each of the image gathers indicates successful annihilation (that is, the gathers do not depend on  $p$  so that differentiation with respect to  $p$  would yield zero, see Section 4.2). To illustrate what happens if the wave speed is not correct we also applied the transform to the downward continued data using wave speeds lower (Fig. 10) or higher than the true ones (Fig. 11). The image gathers are artefact free—in the sense that they do not contain false events (which would show residual moveout even with an acceptable wave speed model)—but they show residual moveouts that will be detected in full waveform sense by the annihilators (because differentiation with respect to  $p$  would not yield zero).



**Figure 9.** A collection of image gathers  $I(., z, p)$  in  $p$  (at selected horizontal positions  $x$  displayed at the centre on top of each gather over the most complex part of the model; compare with Fig. 4 (bottom))—obtained with the angle transform (cf. 19) at the correct wave speed model. The  $p$  value varies from  $-0.5 \text{ s km}^{-1}$  to  $0.5 \text{ s km}^{-1}$  in each gather. In these gathers there is no dependency on  $p$  (and no residual moveout); therefore, these gathers will be annihilated by  $\frac{\partial}{\partial p_i}$ .

## 7 DISCUSSION

We present a method of *wave-equation* reflection tomography—or, in exploration seismics, MVA—with finite-bandwidth data. The approach followed here makes use of scattered phases in the data and has previously been exploited in the framework of the *ray-geometrical* generalized Radon transform (Brandsberg-Dahl *et al.* 2003).

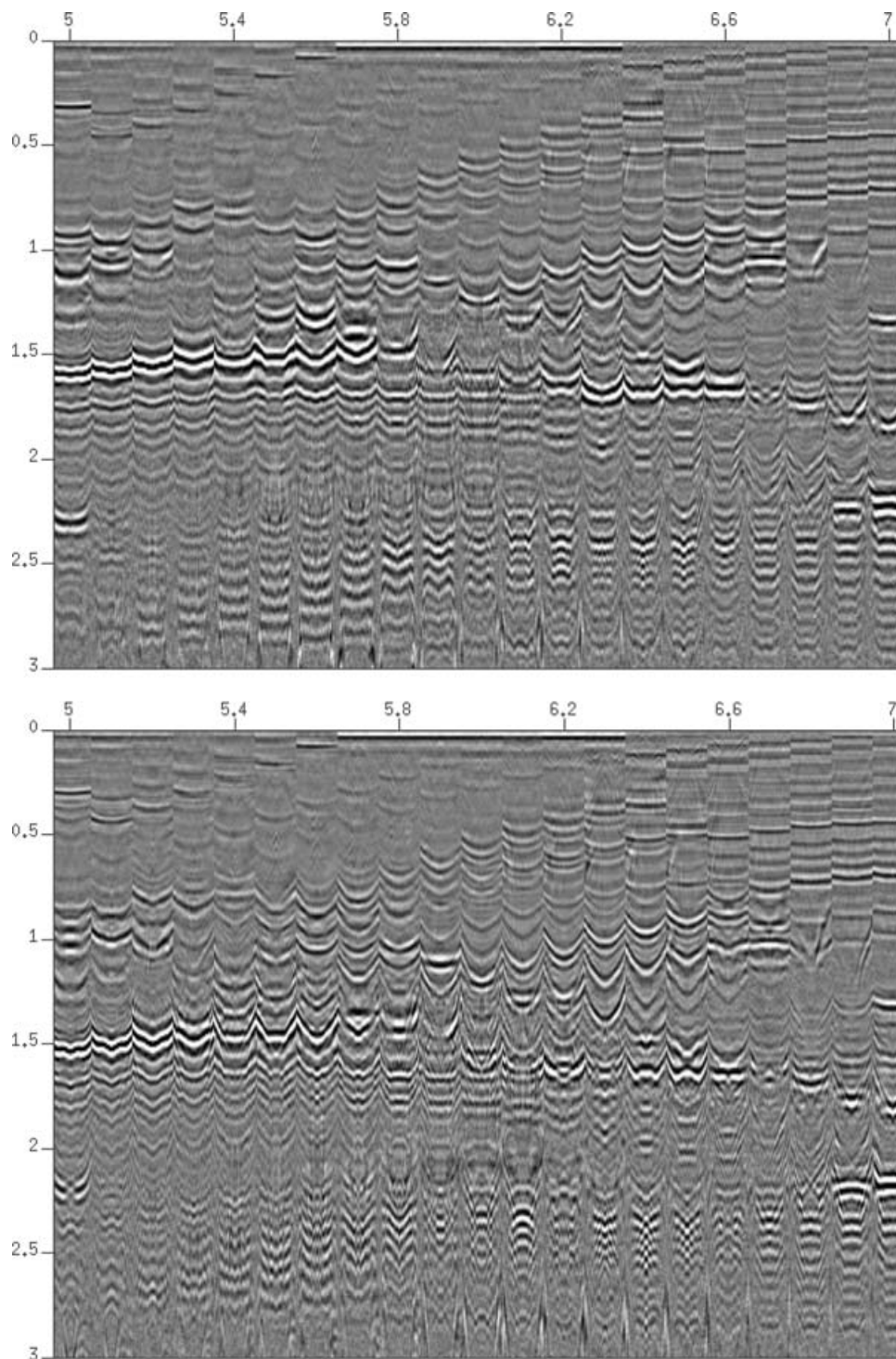
The method presented has aspects in common with the methodology developed by Pratt and co-workers (e.g. Pratt 1999; Sirgue & Pratt 2004). Notably, both approaches take advantage of a frequency domain formulation and implementation. However, there are also important differences. An attractive feature of the approach by Pratt is that it can account for some multiple scattering, unlike our current single-scattering implementation. Pratt's model selection, however, is based on a waveform misfit criterion (Pratt 1999), which may run into problems of multiple local minima when wave speed is at half wavelength away from the true model for the given frequency range. In contrast, our method yields (a class of) models through the use of image gathers and data annihilators. It has been established that the class of acceptable models reduces to a unique solution for an ideal spatial distribution of scatterers (De Hoop, Lassas and Uhlmann, unpublished). Numerical experiments indicate that the annihilator-based approach provides a single minimum of the functional in a wide range of wave speed variations, in particular, when not restricted to the low-frequency regime. Another practical advantage of the wave-equation reflection tomography approach proposed here is that the transition from 2-D to 3-D is straightforward and (computationally) affordable.

The key result presented here is an explicit expression for and characterization of the kernels used in wave-equation reflection tomography. This is all that is needed for, for example, a conjugate gradient implementation of the optimization. The generalized screen expansion for the downward continuation of the seismic data appears as a natural tool in this process, since it is based on a particular wave speed expansion. Our criterion is based upon the condition that a wave speed model is acceptable if the data are in the range of our modelling operator, here developed in the single scattering approximation. The evaluation of the kernel has a few unusual aspects:

- (i) the mismatch occurs and is evaluated in the subsurface rather than at the surface,
- (iii) the adjoint state method, derived from the reciprocity theorem of the time correlation type, leads to a procedure of propagating the adjoint field upwards and the data downwards and taking their cross correlation at the depths where the fields meet and
- (iii) a generalized screen operator is nested in the cross correlation in time.

The mismatch force can attain non-vanishing values only in regions that contain reflectors. Note that an error in, say, a shallow part of the wave speed model will result in unsuccessful annihilation even much deeper in the model.

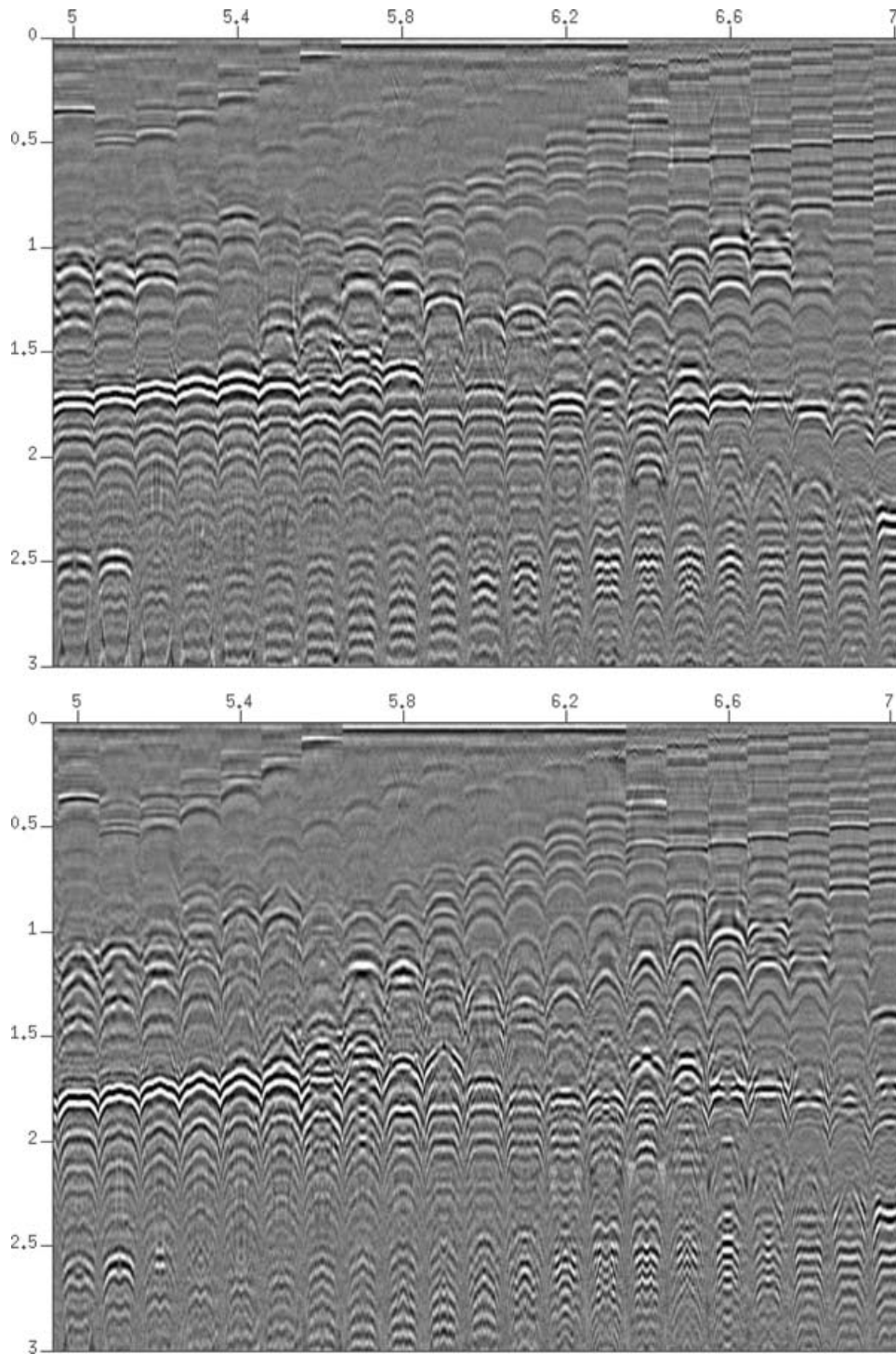
The kernel can be computed with one-way propagators that are more accurate and of a higher degree of sophistication than the generalized screen ones. Relaxing the desire of wide-angle propagation accuracy somewhat, in the presence of rapid and large lateral wave speed variations, the optimal rational operator approximation might be the approach of choice (Van Stralen *et al.* 1998). Exact propagators (Fishman 1992;



**Figure 10.** Image gathers in  $p$ —obtained with eq. (19)—at incorrect wave speed models. Panels of image gathers are obtained at  $c_0 = c_{\text{true}} - 0.1\Delta c_0$  (top) and  $c_0 = c_{\text{true}} - 0.2\Delta c_0$  (bottom), respectively. Notice the degree of residual moveout and how it increases with  $\alpha$  if  $c_0 = c_{\text{true}} + \alpha\Delta c_0$ ; also notice that the events in each gather (in particular at  $p = 0$ ) are mapped at incorrect depths (compare with Fig. 9).

Fishman *et al.* 2000) can be obtained for particular wave speed profiles, most usefully ones quadratic in  $x$ , and applied locally. Uniformly asymptotic propagators (Fishman *et al.* 1997; De Hoop & Gautesen 2000, 2003) have been developed to overcome the inaccuracies associated with the transition from propagating to evanescent wave constituents and apply in general, smoothly varying wave speed profiles.

The procedure presented here can be applied to broad-band (albeit band-limited) data. Any effects related to wave front healing are accounted for in the downward continuation restricted to the frequency band of the data irrespective of the formation of caustics. The full waveform is used, not just the phase of the data. The example demonstrates that the multifrequency band kernels are not merely scaled



**Figure 11.** Image gathers in  $p$ —obtained with eq. (19)—at incorrect wave speed models. Panels of image gathers are obtained at  $c_0 = c_{\text{true}} + 0.1\Delta c_0$  (top) and  $c_0 = c_{\text{true}} + 0.2\Delta c_0$  (bottom), respectively. Notice the degree of residual moveout and how it increases with  $\alpha$  if  $c_0 = c_{\text{true}} + \alpha\Delta c_0$ ; as in Fig. 10, the events in each gather (in particular at  $p = 0$ ) have moved in  $z$  as compared with Fig. 9.

versions of a generic kernel shape. Indeed, the low-frequency data sense different aspects (the smoother parts) of the inhomogeneities than the high-frequency data (which are more sensitive to small scale variations).

There are various practical issues to overcome when applying the procedure to field array data. The data should be deconvolved jointly to ensure a uniform source signature. Also, the sources and receivers may be irregularly spaced so that a regularization of the data would be required prior to downward continuation. If the data are caustic free, such a regularization could in principle be accomplished with a partial differential equation approach (Fomel 2003). In the presence of caustics, the single scattered wave constituents can, in principle, be continued as well with a procedure derived from the annihilators (De Hoop & Uhlmann 2006).



As mentioned above, in the process of downward continuing the data, the results may need to be stored at all intermediate depths – compression techniques may need to be used to reduce memory usage. (The computational complexity of one iteration in the optimization is of the order of two to three depth migrations.) It may be necessary, prior to applying an annihilator, to carry out statistical estimates over image gathers. In passive seismics applications, with an extensive array, interferometric techniques applied between all pairs of receivers may yield the data; for deep Earth applications the sources will need to be imaged from the direct upgoing wavefield observations prior to applying the method of reflection tomography.

The current method is developed for single scattering, that is, primary reflections, and single modes using a scalar wave equation. Replacing the acoustic wavefield decomposition and generalized screen expansion by their elastic counterparts, the method can be extended to the displacement vector wavefield (Le Rousseau & De Hoop 2003), in which case it can be applied to, for instance, P-to-S mode conversions (e.g. via receiver functions). For a recent development and application in global seismology of one-way theory in anisotropic elastic media, see Angus *et al.* (2004).

## ACKNOWLEDGMENTS

We thank Dr Xander Campman (ERL-MIT), Dr John Haines (University of Cambridge) and Dr Louis Fishman for critical comments, which helped us improve the manuscript. The research reported here was supported by the National Science Foundation under grant EAR-0409816 and by Total E&P.

## REFERENCES

- Aki, K. & Richards, P., 2002. *Quantitative Seismology*, Cambridge University Press, Cambridge.
- Alkhalifah, T., 1998. Acoustic approximations for processing in transversely isotropic media, *Geophysics*, **63**, 623–631.
- Angus, D., Thomson, C. & Pratt, R., 2004. A one-way wave equation for modelling seismic waveform variations due to elastic anisotropy, *Geophys. J. Int.*, **156**, 595–614.
- Berkhout, A., 1997. Pushing the limits of seismic imaging, part i: prestack migration in terms of double dynamic focusing, *Geophysics*, **62**, 937–953.
- Biondi, B. & Sava, P., 1999. Wave-equation migration velocity analysis, 69th Ann. Internat. Meeting, Soc. of Exploration Geophysicists, *Expanded Abstracts*, 1723–1726.
- Brandsberg-Dahl, S., Ursin, B. & De Hoop, M., 2003. Seismic velocity analysis in the scattering-angle/azimuth domain, *Geophys. Prospect.*, **51**, 295–314.
- Chauris, H. & Noble, M., 2001. Two-dimensional velocity macromodel estimation from seismic reflection data by local differential semblance optimization: applications to synthetic and real data sets, *Geophys. J. Int.*, **144**, 14–26.
- Claerbout, J., 1985. *Imaging the earth's interior*, Blackwell Scientific Publishing, Oxford.
- Clayton, R., 1978. Common midpoint migration, *Technical Report*, Stanford University, **SEP-14**.
- Dahlen, F., Hung, S.-H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.*, **141**, 157–174.
- De Bruin, C., Wapenaar, C. & Berkhout, A., 1990. Angle-dependent reflectivity by means of prestack migration, *Geophysics*, **55**, 1223–1234.
- De Hoop, M., 1996. Generalization of the Bremmer coupling series, *J. Math. Phys.*, **37**, 3246–3282.
- De Hoop, M. & De Hoop, A., 2000. Wavefield reciprocity and optimization in remote sensing, *Proc. R. Soc. Lond. A (Mathematical, Physical and Engineering Sciences)*, **456**, 641–682.
- De Hoop, M. & Gautesen, A., 2000. Uniform asymptotic expansion of the generalized Bremmer series, *SIAM J. Appl. Math.*, **60**, 1302–1329.
- De Hoop, M. & Gautesen, A., 2003. Uniform asymptotic expansion of the square-root helmholtz operator and the one-way wave propagator, *SIAM J. Appl. Math.*, **63**, 777–800.
- De Hoop, M. & Uhlmann, G., 2006. Characterization and ‘source-receiver’ continuation of reflection seismic data, *Comm. Math. Phys.*, **263**, 1–19.
- De Hoop, M. & Van der Hilst, R., 2003. Global wave-equation reflection tomography, *Center for Wave Phenomena Preprint*, **453**.
- De Hoop, M. & Van der Hilst, R., 2005. On sensitivity kernels for ‘wave-equation’ transmission tomography, *Geophys. J. Int.*, **160**, doi:10.1111/j.1365-246X.2004.02509.
- De Hoop, M., Le Rousseau, J. & Wu, R., 2000. Generalization of the phase-screen approximation for the scattering of acoustic waves, *Wave Motion*, **31**, 43–70.
- De Hoop, M., Le Rousseau, J. & Biondi, B., 2003. Symplectic structure of wave-equation imaging: a path-integral approach based on the double-square-root equation, *Geoph. J. Int.*, **153**, 52–74.
- Fishman, L., 1992. Exact and operator rational approximate solutions of the helmholtz, weyl composition equation in underwater acoustics. the quadratic profile, *J. Math. Phys.*, **33**, 1887–1914.
- Fishman, L. & McCoy, J., 1984a. Derivation and application of extended parabolic wave theories i. the factorized helmholtz equation, *J. Math. Phys.*, **25**, 285–296.
- Fishman, L. & McCoy, J., 1984b. Derivation and application of extended parabolic wave theories ii. path integral representations, *J. Math. Phys.*, **25**, 297–308.
- Fishman, L., McCoy, J. & Wales, S., 1987. Factorization and path integration of the helmholtz equation: numerical algorithms, *J. acoust. Soc. Am.*, **81**, 1355–1376.
- Fishman, L., Gautesen, A. & Sun, Z., 1997. Uniform high-frequency approximations of the square-root helmholtz operator symbol, *Wave Motion*, **26**, 127–161.
- Fishman, L., De Hoop, M. & Van Stralen, M., 2000. Exact constructions of square-root helmholtz operator symbols: the focusing quadratic profile, *J. Math. Phys.*, **41**, 4881–4938.
- Fomel, S., 2003. Theory of differential offset continuation, *Geophysics*, **68**, 718–732.
- Guillemin, V. & Uhlmann, G., 1981. Oscillatory integrals with singular symbols, *Duke Math. Journal*, **48**, 251–267.
- Le Rousseau, J. & De Hoop, M., 2001. Modeling and imaging with the scalar generalized-screen algorithms in isotropic media, *Geophysics*, **66**, 1551–1568.
- Le Rousseau, J. & De Hoop, M., 2003. Generalized-screen approximation and algorithm for the scattering of elastic waves, *Q. J. Mech. Appl. Math.*, **56**, 1–33.
- Luo, Y. & Schuster, G., 1991. Wave-equation travel time inversion, *Geoph.*, **56**, 645–653.
- Mulder, W. & Ten Kroode, A., 2002. Automatic velocity analysis by differential semblance optimization, *Geophysics*, **67**, 1184–1191.
- Nilsson, B. & Fishman, L., eds., 2004. *One-way wave equation modeling in two-way wave propagation problems*, Vol. 7 of Mathematical Modelling in Physics, Engineering and Cognitive Sciences, Växjö University Press, Växjö, Sweden, pp. 91–111.
- Peng, S., 2004. Automatic wave equation migration velocity analysis using differential semblance, *PhD thesis*, Rice University, Houston.
- Pratt, R., 1999. Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model, *Geophysics*, **64**, 888–901.

- Revenaugh, J. & Jordan, T., 1991. Mantle layering from ScS reverberations. 1. wave form inversion of zeroth-order reverberations, *J. geophys. Res.*, **96**, 19 749–19 762.
- Sava, P., Biondi, B. & Fomel, S., 1999. Amplitude-preserved common image gathers by wave-equation migration, in *Expanded Abstracts*, pp. 824–827, Society of Exploration Geophysicists.
- Sava, P. & Fomel, S., 2003. Angle-domain common-image gathers by wave-field continuation methods, *Geophysics*, **68**, 1065–1074.
- Scherbaum, F., Krüger, F. & Weber, M., 1997. Double beam forming: mapping lower mantle heterogeneities using combinations of source and receiver arrays, *J. geophys. Res.*, **102**, 507–522.
- Schoenberg, M. & De Hoop, M., 2000. Approximate dispersion relations for {qP-qSV} waves in transversely isotropic media, *Geophysics*, **65**, 919–933.
- Sipkin, S. & Jordan, T., 1976. Lateral heterogeneity of upper mantle determined from travel times of multiple ScS, *J. geophys. Res.*, **81**, 6307–6320.
- Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, **69**, 231–248.
- Stolk, C., 2004. A pseudodifferential equation with damping for one-way wave propagation in inhomogeneous acoustic media, *Wave Motion*, **40**, 562–575.
- Stolk, C. & De Hoop, M., 2001. Seismic inverse scattering in the ‘wave-equation’ approach, *MSRI preprint*, #2001–047.
- Stolk, C. & De Hoop, M., 2002. Microlocal analysis of seismic inverse scattering in anisotropic elastic media, *Comm. Pure Appl. Math.*, **55**, 261–301.
- Stolk, C. & De Hoop, M., 2005. Modeling of seismic data in the downward continuation approach, *SIAM J. Appl. Math.*, **65**, 1388–1406.
- Stolk, C. & De Hoop, M., 2006. Seismic inverse scattering in the downward continuation approach, *Wave Motion*, in press.
- Stolk, C. & Symes, W., 2003. Smooth objective functionals for seismic velocity inversion, *Inverse Problems*, **19**, 73–89.
- Stolk, C., De Hoop, M. & Symes, W., 2006. Kinematics of shot-geophone migration, *Geophysics*, submitted.
- Symes, W. & Carazzone, J., 1991. Velocity inversion by differential semblance optimization, *Geophysics*, **56**, 654–663.
- Tarantola, A., 1987. *Inverse problem theory*, Elsevier, Amsterdam, The Netherlands.
- Thorbecke, J., 1997. Common focus point technology, *PhD thesis*, Delft University of Technology, Delft, the Netherlands.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.*, **160**, 195–216.
- Van Stralen, M., De Hoop, M. & Blok, H., 1998. Generalized bremsler series with rational approximation for the scattering of waves in inhomogeneous media, *J. acoust. Soc. Am.*, **104**, 1943–1963.
- Vasco, D., Peterson, J. & Majer, E., 1995. Beyond ray tomography: wavepaths and fresnel volumes, *Geophysics*, **60**, 1790–1804.
- Woodward, M., 1992. Wave equation tomography, *Geophysics*, **57**, 15–26.
- Wunsch, C., 1996. *The ocean circulation inverse problem*, Cambridge University Press, Cambridge.
- Zelt, C., Sain, K., Naumenko, J. & Sawyer, D., 2003. Assessment of crustal velocity models using seismic refraction and reflection tomography, *Geophys. J. Int.*, **153**, 609–626.
- Zhao, L. & Jordan, T., 1998. Sensitivity of frequency-dependent travel times to laterally heterogeneous, anisotropic earth structure, *Geophys. J. Int.*, **133**, 683–704.
- Zhao, L., Jordan, T., Olsen, K. & Chen, P., 2005. Frechet kernels for imaging regional earth structure based on three-dimensional reference models, *Bull. seism. Soc. Am.*, **95**, 2066–2080.

## APPENDIX A: GENERALIZED SCREEN EXPANSION AND ITS APPLICATION

For the optimization underlying our reflection tomography we need to know the (Fréchet) derivatives of the single-square-root operator with respect to the background wave speed  $c_0(x, z)$ . We take this derivative by expanding its symbol,  $b$ , into a sum of symbols each of which allows a separation of the phase-space variables  $(x, \xi)$ . This applies where  $b$  is smooth as a function of  $(x, \xi)$ , which is the case in the propagating regime (or, more generally, away from horizontal propagation which occurs when  $\omega^{-1} \|\xi\| = c_0^{-1}$ ). Throughout, we will assume appropriate cut-offs have been applied to ensure that the smoothness condition is satisfied. For computational efficiency, we seek a frequency domain formulation. Such a formulation allows a multifrequency strategy for optimization.

The expansion mentioned above falls in the well-established category of generalized screen expansions, which also leads to the introduction of fast algorithms for one-way wave propagation. We provide, here, an overview of such expansions and their use in reflection tomography.

One way of arriving at a generalized screen expansion is by choosing a reference medium with a wave speed  $\gamma_0$  that depends on depth  $z$  only. In the process of model updating to be developed below, we assume that at any particular depth  $z$ ,  $\gamma_0(z)$  is the largest lower bound on the wave speed at that depth according to any of the allowable models. With this requirement,  $b$  is expanded in terms of positive or zero contrasts  $[c_0(x, z)^{-2} - \gamma_0(z)^{-2}]$  with  $c_0$  representing the actual wave speeds in the medium; thus, the  $c_0$  used in the tomography still depends on all  $n$  space variables.

### A1 The single-square-root operator

The generalized screen expansion of  $b$  up to order  $N$  is of the form

$$b(x, z, \xi, \omega) = \sum_{j=0}^N A_j(\xi, \omega, z) S_j[c_0](x, z), \quad (\text{A1})$$

with  $A_0(\xi, \omega, z) = \omega \sqrt{\frac{1}{\gamma_0(z)^2} - \omega^{-2} \|\xi\|^2}$  and  $S_0[c_0](x, z) = 1$  (cf. Le Rousseau & De Hoop 2001, 16).<sup>10</sup> The factors  $A_j$  mostly control the shape (bending) of the local slowness surface ( $b$  as a function of  $\xi$  for given  $(x, z)$  and  $\omega$ ) while the factors  $S_j$  (loosely referred to as screen functions) account for the change in the slowness surface due to the horizontal medium fluctuations relative to the background  $\gamma_0 = \gamma_0(z)$ . The factors  $A_j$  depend on  $\gamma_0$  but not on the horizontal medium fluctuations.

<sup>10</sup> The generalized screen expansion leads to  $A_j(\xi, \omega, z) = \omega a_j [\gamma_0(z)^{-2} - \omega^{-2} \|\xi\|^2]^{(-2j+1)/2}$  with  $a_j = (-1)^{j+1} \frac{1 \cdot 3 \cdots (2j-1)}{j! 2^j}$ , and  $S_j(x, z) = [c_0(x, z)^{-2} - \gamma_0(z)^{-2}]^j$ ,  $j = 1, 2, \dots$

Substituting  $b$  as in eq. (A1) into eq. (4) shows that the single-square-root operator  $B_U$  in eq. (3) acts on the upgoing wavefield  $u_U$  as

$$(B_U u_U)(x, t, z) \sim F_{\omega \rightarrow t}^{-1} \sum_{j=0}^N F_{\xi \rightarrow x}^{-1} A_j(\xi, \omega, z) F_{x' \rightarrow \xi} S_j[c_0](x', z) F_{t' \rightarrow \omega} u_U(x', t', z),$$

where  $F$  denotes the Fourier transform as in the main text. For computational efficiency the calculations leading to the sensitivity kernel for wave equation reflection tomography (see Appendix C) are carried out in the frequency ( $\omega$ ) domain. In the frequency domain, which is indicated by  $\hat{\cdot}$ , the above expression becomes

$$(\hat{B}_U \hat{u}_U)(x, \omega, z) \sim \sum_{j=0}^N F_{\xi \rightarrow x}^{-1} A_j(\xi, \omega, z) F_{x' \rightarrow \xi} S_j[c_0](x', z) \hat{u}_U(x', \omega, z), \quad (\text{A2})$$

(cf. Le Rousseau & De Hoop 2001, 31). The dependency of the operator on the horizontally varying component of the background medium is completely contained in the factors  $S_j$ .

## A2 The perturbed single-square-root operator

The previous subsection shows that a wave speed perturbation of operator  $B_U$  follows from the wave speed perturbation of its symbol  $b$ , through the perturbation of the factors  $S_j$ . By the nature of the derivation of the generalized screen expansion, up to first order, we have  $b - A_0 S_0 = A_1 S_1$  (cf. A1); hence,  $A_1$  can be viewed as the Fréchet derivative of  $b$  with respect to  $c_0^{-2}$  evaluated at  $\gamma_0^{-2}$ . Now, we extend the evaluation of the derivative of  $b$ , and  $B_U$ , to the general case.

For the later application of tomography, we consider how  $B_U$  is perturbed under a smooth perturbation  $\delta c_0(x, z)$  of  $c_0$  subject to the constraint that  $\gamma_0 = \gamma_0(z)$  is kept fixed. (In principle, any perturbation in  $\gamma_0(z)$  can be absorbed in  $\delta c_0(x, z)$ , but for fixed  $N$  the accuracy of the propagator depends on  $\delta c_0(x, z)$  and may suffer from keeping  $\gamma_0(z)$  fixed, because  $\gamma_0(z)$  needs to remain the largest bound on the wave speed under perturbations at each  $z$ ). In view of eq. (A2) we have

$$(\delta \hat{B}_U) \hat{u}_U(x, \omega, z) \sim \sum_{j=1}^N F_{\xi \rightarrow x}^{-1} A_j(\xi, \omega, z) F_{x' \rightarrow \xi} \delta S_j(x', z) \hat{u}_U(x', \omega, z). \quad (\text{A3})$$

The perturbation  $\delta S_j$  in  $S_j$  is expressed in terms of Fréchet derivatives  $S_j'^1$  as the multiplication

$$\delta S_j(x, z) = S_j'[c_0](x, z) \delta c_0(x, z). \quad (\text{A4})$$

Substituting eq. (A4) into eq. (A3) gives the integral operator

$$\hat{B}'_U(\hat{u}_U) \cdot \sim \sum_{j=1}^N F_{\xi \rightarrow x}^{-1} A_j(\xi, \omega, z) F_{x' \rightarrow \xi} S_j'(x', z) \hat{u}_U(x', \omega, z) \cdot_{(x', z)}, \quad (\text{A5})$$

such that

$$(\delta \hat{B}_U) \hat{u}_U \sim \hat{B}'_U(\hat{u}_U) \delta c_0. \quad (\text{A6})$$

Eq. (A5) shows how the field  $\hat{u}_U$  is absorbed as a factor in the screen function,  $S_j' \hat{u}_U$ ; the  $\cdot_{(x', z)}$  notation reveals that the function (to be inserted at  $\cdot$ ), such as  $\delta c_0$  in eq. (A6), on which the derivative  $\hat{B}'_U$  acts, is evaluated at  $(x', z)$  inside the Fourier transform from  $x'$  to  $\xi$ .

## A3 The adjoint of the perturbed single-square-root operator

For the purpose of tomography (that is, the adjoint state calculation to obtain an image), we will also need the *adjoint*  $\hat{B}'_U(\hat{u}_U)^*$  of  $\hat{B}'_U(\hat{u}_U)$ . Since  $(F_{x \rightarrow \xi})^* = F_{\xi \rightarrow x}^{-1}$  (up to factors of  $2\pi$ ), we have

$$\hat{B}'_U(\hat{u}_U)^* \cdot \sim \sum_{j=1}^N \overline{\hat{u}_U(x', \omega, z)} S_j'(x', z) F_{\xi \rightarrow x'}^{-1} A_j(\xi, \omega, z) F_{x \rightarrow \xi} \cdot_{(x, \omega, z)}, \quad (\text{A7})$$

with the property that

$$\langle \hat{B}'_U(\hat{u}_U) \delta c_0, \hat{v} \rangle_{(x)} = \langle \delta c_0, \hat{B}'_U(\hat{u}_U)^* \hat{v} \rangle_{(x)}$$

where  $\langle \cdot, \cdot \rangle_{(x)}$  indicates the inner product in the space of complex, square-integrable functions (the subscript in parentheses indicates the integration variable);  $\hat{v}$  represents a trial function of  $x$  and  $\omega$  (and  $z$ ).

<sup>11</sup> Here,  $S_j'(x, z) = j [c_0(x, z)^{-2} - \gamma_0(z)^{-2}]^{j-1} (-2) c_0(x, z)^{-3}$ ,  $j = 1, 2, \dots$ . It is natural to absorb the factors  $j$  and  $-2$  in  $A_j$  and re-label  $j =: j - 1$ . If one parametrizes the wave speed model by  $c_0^{-2}$  rather than  $c_0$ , the factor  $(-2) c_0(x, z)^{-3}$  disappears and the expansion retains its original structure.

## APPENDIX B: DECOMPOSITION AND RECOMPOSITION

The decomposition and recombination operators are subjected to a generalized screen expansion also. It is noted that these operators also occur in the WKB approximation to the Green's functions in layered media.

The principal symbol of  $Q_U$  is given by

$$q(x, z, \xi, \omega) = (\omega^2 c_0(x, z)^{-2} - \|\xi\|^2)^{-1/4}.$$

Following the methodology in eq. (A1), we find an expansion of the type

$$q(x, z, \xi, \omega) \sim \zeta(\xi, \omega, z)^{-1/2} \left( 1 - \frac{1}{4} \frac{\omega^2 \mathcal{S}_1[c_0](x, z)}{\zeta(\xi, \omega, z)^2} + \frac{5}{32} \frac{\omega^4 \mathcal{S}_2[c_0](x, z)}{\zeta(\xi, \omega, z)^4} + \dots \right), \quad (\text{B1})$$

where

$$\zeta = \zeta(\xi, \omega, z) = (\omega^2 \gamma_0(z)^{-2} - \|\xi\|^2)^{1/2}.$$

In the split-step approximation, one would evaluate the expression in between parentheses for  $\xi = 0$ . A similar expansion can be found for  $Q_U^{-1}$ .

## APPENDIX C: PERTURBING THE FUNCTIONAL

Under a smooth perturbation  $\delta c_0$  of  $c_0$  subject to the constraint that  $\gamma_0$  is kept the same, the perturbation of the downward-continued data  $u$  is  $\delta u$ ; the perturbation of the functional then follows to be

$$\delta \mathcal{J} = \sum_i \int \int (R'_i u) (R'_i \delta u) dx dz dp, \quad (\text{C1})$$

because  $R'$  is independent of  $c_0$ . Recognizing an inner product in  $(x, p)$ , it follows that

$$\delta \mathcal{J} = \frac{1}{2\pi} \int \int \int \underbrace{\left( \sum_i \hat{R}'_i{}^* R'_i u \right)}_{\text{source}} \underbrace{\left( \widehat{\delta u} \right)}_{\text{field}} ds dr dz d\omega, \quad (\text{C2})$$

by definition of the adjoint in eq. (24). The integration over  $(s, r, \omega)$  is identified as an inner product  $\langle \cdot, \cdot \rangle_{(s,r,\omega)}$ . Substituting the expression for downward continuation yields eq. (27):

$$\delta \mathcal{J}[c_0] = \int \left\{ \int \left( \sum_i \hat{R}'_i{}^* R'_i H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d \right) (s, r, \omega, z) \overline{\left( \delta H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d \right)} (s, r, \omega, z) ds dr d\omega \right\} dz. \quad (\text{C3})$$

In view of the complex conjugation we recognize in eq. (C2) a time correlation between a 'source' and a field. A standard adjoint calculation results in the kernel associated with  $\delta \mathcal{J}$ ; the details are presented in the following subsections.

### C1 Evolution equations

First, we consider the perturbation of the downward continued data,  $\delta u(\cdot, z) = \delta H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d$  as it appears in eq. (C3). Making use of the self-adjointness of  $C$ , in the linearization, we find that

$$\delta H(0, z)^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d = \int_{z'=0}^z H(z', z)^* i \delta C(z') H(0, z')^* Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi d dz'. \quad (\text{C4})$$

The right-hand side solves the equation

$$\left( \frac{\partial}{\partial z} - i \hat{C}(s, r, z, D_s, D_r, \omega) \right) \delta \hat{u} = i \delta \hat{C}(s, r, z, D_s, D_r, \omega) \hat{u}, \quad (\text{C5})$$

subject to a vanishing initial condition at  $z = 0$ , where  $\delta \hat{C}$  is the perturbation of operator  $\hat{C}$  with  $\delta c_0$ , which follows from eqs (A3) and (12).

Eq. (C5) is solved in the direction of increasing  $z$  (downward) with homogeneous initial conditions at  $z = 0$ . These conditions follow from the assumption that  $c_0$  is known where the data are measured (that is, near  $z = 0$ ). Eq. (C5) follows the wave speed perturbation for the downward continuation of the data, and its solution can be generated by a composition of thin-slab propagators (De Hoop *et al.* 2003): within each thin slab,  $[z, z + \Delta]$ , with  $\Delta$  small, the wave speed is treated as if it were depth independent. For the unperturbed DSR eq. (15), with  $c$  the symbol of DSR operator  $C$  in eq. (13), this gives

$$\begin{aligned} u(s', r', t', z + \Delta) &= (H(z, z + \Delta)^* u(\cdot, z))(s', r', t') \\ &\sim (2\pi)^{-(2n-1)} \int \int \exp[i\sigma(s' - s)] \exp[i\rho(r' - r)] \exp[i\omega(t' - t)] \\ &\quad \exp[i\Delta c(s, r, z, \sigma, \rho, \omega)] d\sigma d\rho d\omega u(s, r, t, z) ds dr dt. \end{aligned} \quad (\text{C6})$$

Taking the wave speed perturbation of this expression, making explicit that  $u$  depends on wave speed, yields

$$\begin{aligned} \delta u(s', r', t', z + \Delta) &\sim (2\pi)^{-(2n-1)} \int \int \exp[i\sigma(s' - s)] \exp[i\rho(r' - r)] \exp[i\omega(t' - t)] \\ &\quad \exp[i\Delta c(s, r, z, \sigma, \rho, \omega)] d\sigma d\rho d\omega \delta u(s, r, t, z) ds dr dt \\ &+ (2\pi)^{-(2n-1)} \int \int \exp[i\sigma(s' - s)] \exp[i\rho(r' - r)] \exp[i\omega(t' - t)] \\ &\quad i\Delta \delta c(s, r, z, \sigma, \rho, \omega) \exp[i\Delta c(s, r, z, \sigma, \rho, \omega)] d\sigma d\rho d\omega u(s, r, t, z) ds dr dt, \end{aligned} \quad (C7)$$

which represents the thin-slab propagation associated with eq. (C5).

Second, in preparation of the application of the volume integral form of the reciprocity theorem of the time-correlation type in the framework of one-way wave theory, we now distinguish two states: one state with *contrast* source distribution  $i\delta C u$  and field  $\delta u$ , and one state with *mismatch* source distribution  $\sum_i (\hat{R}'_i)^* R'_i u$  and field, say  $\hat{v}$ . The mismatch source distribution is generated at any image point where full annihilation of a downward continued reflection ( $R'_i u$ ) has failed (is not zero) and vanishes if the local background wave speed  $c_0$  is correct (so that full annihilation is achieved).

More specifically, we have

$$v(., z') = \int_{z'}^{\infty} H(z', z) \sum_i (\hat{R}'_i)^* R'_i H(0, z) Q_{U,s}^*(0)^{-1} Q_{U,r}^*(0)^{-1} \psi dz, \quad (C8)$$

which solves the equation

$$\left( \frac{\partial}{\partial z} - i\hat{C}(s, r, z, D_s, D_r, \omega) \right) \hat{v} = \sum_i (\hat{R}'_i)^* R'_i u, \quad (C9)$$

which is also known as the adjoint field equation (Wunsch 1996). This equation is solved in the (upward) direction of decreasing  $z$ , with vanishing initial condition for some large  $z$  (say, at the bottom of the model).

## C2 The sensitivity kernel

The reciprocity theorem of the time-correlation type, applied to eq. (C5), (C2),<sup>12</sup> now implies that we can write eq. (C2) in the equivalent form

$$\delta \mathcal{J} = \frac{1}{2\pi} \int \int \int \widehat{\hat{v}} \left( \widehat{i\delta \hat{C} \hat{u}} \right) ds dr dz d\omega. \quad (C10)$$

The frequency integral is kept as the outside integral because the kernel computations is to be carried out in the  $\omega$ -domain. We note that  $v$  is real-valued ( $\widehat{\hat{v}}(\omega) = \hat{v}(-\omega)$ ) while the same holds for  $F_{\omega \rightarrow i}^{-1} i\delta \hat{C} \hat{u}$  in view of the frequency dependency of  $b$ ; see the remark below eq. (4).

We can now combine the concepts developed in the previous appendices, and formulate the sensitivity kernel. To do this, we write the perturbation  $\delta \hat{C}$  in eq. (C10) in the form of a derivative  $\delta \hat{C} \hat{u} = \hat{C}'(\hat{u}) \delta c_0$ , with

$$\begin{aligned} \hat{C}'(\hat{u}) \cdot &= \hat{B}'_{U,s}(\hat{u}) \cdot + \hat{B}'_{U,r}(\hat{u}) \cdot \sim \sum_{j=0}^N F_{\sigma \rightarrow s}^{-1} A_j(\sigma, \omega, z) F_{s' \rightarrow \sigma} S'_j(s', z) \hat{u}(s', r, \omega, z) \cdot_{(s',z)} \\ &+ \sum_{j=0}^N F_{\rho \rightarrow r}^{-1} A_j(\rho, \omega, z) F_{r' \rightarrow \rho} S'_j(r', z) \hat{u}(s, r', \omega, z) \cdot_{(r',z)}. \end{aligned}$$

Substituting this expression into eq. (C10) and taking the adjoint, we obtain

$$\delta \mathcal{J} = \frac{1}{2\pi} \int \int \int \hat{v} \left( \widehat{i\delta \hat{C} \hat{u}} \right) ds dr dz d\omega = \int [(iC'(u))^* v] \delta c_0 dx dz. \quad (C11)$$

To extract the kernel of the derivative of  $\mathcal{J}$  out of eq. (A11), we make use of relation (A7) with  $x'$  playing the role of  $s$  or  $r$ . We write  $\delta \mathcal{J} = \int \mathcal{K}(x, z) \delta c_0(x, z) dx dz$ , in which

$$\mathcal{K}(x, z) := (iC'(u))^* v(x, z) = \frac{1}{2\pi} \int d\omega (-i) \left[ \int \hat{B}'_{U,s}(\hat{u})^* \hat{v} dr + \int \hat{B}'_{U,r}(\hat{u})^* \hat{v} ds \right]. \quad (C12)$$

<sup>12</sup> This theorem, in the frequency domain, follows upon right multiplication of eq. (C9) by  $(\delta \hat{u})$  and left multiplication of the complex conjugate of eq. (C5) by  $\hat{v}$ , and subtracting the resulting equalities.