Geophysical Journal International

Geophys. J. Int. (2010) **182,** 1383–1394 doi: 10.1111/j.1365-246X.2010.04681.x

A correlation-based misfit criterion for wave-equation traveltime tomography

T. van Leeuwen[1] and W. A. Mulder[1,2]

[1] *Delft University of Technology, Department of Geotechnology,* PO Box 5048, 2600 *GA Delft, the Netherlands. E-mail: tristan.vanleeuwen@tudelft.nl*
[2] *Shell International Exploration and Production,* PO Box 60,2280 *AB, Rijswijk, the Netherlands*

Accepted 2010 May 28. Received 2010 May 28; in original form 2010 March 8

SUMMARY

Wave-equation traveltime tomography tries to obtain a subsurface velocity model from seismic data, either passive or active, that explains their traveltimes. A key step is the extraction of traveltime differences, or relative phase shifts, between observed and modelled finite-frequency waveforms. A standard approach involves a correlation of the observed and measured waveforms. When the amplitude spectra of the waveforms are identical, the maximum of the correlation is indicative of the relative phase shift. When the amplitude spectra are not identical, however, this argument is no longer valid. We propose an alternative criterion to measure the relative phase shift. This misfit criterion is a weighted norm of the correlation and is less sensitive to differences in the amplitude spectra. For practical application it is important to use a sensitivity kernel that is consistent with the way the misfit is measured. We derive this sensitivity kernel and show how it differs from the standard banana–doughnut sensitivity kernel. We illustrate the approach on a cross-well data set.

Key words: Inverse theory; Body waves; Seismic tomography.

1 INTRODUCTION

In ray-based tomography, the aim is to construct a subsurface velocity model that explains the picked traveltimes of the measured data. Such a model can be obtained in an iterative manner by back projecting the traveltime differences along rays in the current velocity model. This procedure will lead to satisfactory results when the wave propagation is sufficiently well approximated by ray theory. To extract more information from the data than just the traveltimes of a few selected arrivals, seismologists are moving towards full-waveform processing and inversion of all available data. This trend is driven by the availability of high-quality broad-band data (earthquake data from USArray, for example) and a need to incorporate finite-frequency effects to process data from geologically complex areas (sub-salt exploration for the detection of hydrocarbons, for example). Also, the computing resources needed to routinely model 3-D wave-propagation in complex media are becoming a commodity. There have been two major developments regarding finite-frequency or full-waveform inversion in the last decades. The first is least-squares inversion—or waveform tomography—of seismic data pioneered by Tarantola & Valette (1982) that was aimed originally at inversion of active seismic reflection data. The second is the development of a finite-frequency analogue of ray-based traveltime tomography by Luo & Schuster (1991) (see also Tromp *et al.* 2005). Both aim at exploiting the full measured waveforms by posing the inversion as an optimization problem:

Given observed data, $\bar{d}$, find a subsurface model for which the modelled data, d, minimizes the misfit functional $\phi[\bar{d}, d]$.

The typical size of such optimization problems dictates that we employ an iterative, gradient-based optimization algorithm, such as non-linear Conjugate Gradients or a (Quasi-) Newton method (Vogel 2002), to solve the optimization problem. These methods require that the misfit does not exhibit local minima and has a large basin of attraction around the global minimum.

Whereas the least-squares misfit satisfies these criteria in some cases, resolving velocity structures that vary on a scale larger than the largest wavelength can be problematic. The basic rationale is as follows. These large-scale velocity variations cause significant phase shifts between the modelled and observed waveforms. This, in turn, may cause local minima in the least-squares error through loop-skipping. Because the traveltimes are more linearly related to such velocity perturbations (Cara & Leveque 1987), one usually resorts to (wave-equation) traveltime tomography to resolve these structures. One way of measuring the misfit relies on picking the traveltime difference from the correlation of the modelled and measured waveforms (Luo & Schuster 1991; de Hoop & van der Hilst 2005).

In this paper, we propose a new correlation-based misfit functional that can be used for wave-equation traveltime tomography. We show that the proposed misfit functional implicitly measures the phase shift between the modelled and observed waveforms. Hence, the procedure is in principle equivalent to the correlation-based approach proposed by Luo & Schuster (1991) where the phase shift

is measured explicitly by picking the maximum of the correlation. However, picking the maximum of the correlation will only reveal the phase shift correctly if the source spectra of the modelled and observed data are identical (Hörmann & de Hoop 2002; de Hoop & van der Hilst 2005). Our alternative procedure is less sensitive to errors in the estimated source spectrum.

As has been recognized by de Hoop & van der Hilst (2005), such measures of misfit will only lead to meaningful updates of the model if one uses a sensitivity kernel that is consistent with the misfit. We derive the sensitivity kernel for the newly proposed measure of misfit. The exposition and analysis closely follows de Hoop & van der Hilst (2005), and we repeat some of the main points and ideas of this work throughout the paper for completeness' sake.

The outline of the paper is as follows. First, we briefly introduce some notation and describe the mathematical modelling used for the subsequent exposition and analysis. In Section 3, we introduce the new correlation-based misfit functional, which is based on the weighted norm of the correlation, and compare it with a commonly used approach. We analyse them in some detail and show that the weighted norm of the correlation also measures the traveltime difference and is less sensitive to errors in the estimated source spectrum. The corresponding sensitivity kernels are derived and analysed in Section 4. We calculate the kernels explicitly for a velocity that increases linearly with depth and use these to illustrate their properties in Section 5. An inversion of real cross-well data using the proposed correlation-based approach is presented in Section 6. Finally, we discuss the results and present the conclusions.

2 MODELLING

The displacement, $\mathbf{u}(t, \mathbf{x})$, is modelled by a linear elastic wave equation of the form

$$\left[\delta_{il}\partial_t^2 - \partial_{x_j} c_{ijkl}\partial_{x_k}\right] u_l = f_i, \tag{1}$$

where c_{ijkl} is the stiffness tensor and f_i is the source function. The data consists of measurements of one or more components of the displacement for several distinct source–receiver pairs. Generally, waveform/traveltime tomography relies on solving the mentioned wave equation by some numerical method (e.g. finite differences). However, for the subsequent analysis we make a high-frequency assumption. We are interested in high-frequency body waves that, away from caustics, permit the following far-field representation of the Green's function

$$G_i^{(j)}(t, \mathbf{x}_s, \mathbf{x}_r) \simeq (2\pi)^{-1} a_i^{(j)}(\mathbf{x}_s, \mathbf{x}_r) \times \int \mathrm{d}\omega \, \exp[\imath\omega(t - T^{(j)}(\mathbf{x}_s, \mathbf{x}_r)) - \tfrac{n\pi}{2}], \tag{2}$$

where a is the amplitude, T is the traveltime, n is the KMAH index (indicating the number and type of caustics that the wave has encountered along the way) and $^{(j)}$ enumerates the modes and traveltime branches. For the analysis, we focus the exposition to one particular component, mode and branch. We will assume point sources and denote the corresponding source spectrum by f, which we assume to be hermitian (i.e. $f(\omega) = f^*(-\omega)$) and square-integrable (i.e. has finite energy). Also, we ignore the factor $\exp[\imath n\pi/2]$ for brevity of notation. Hence, we are considering data of the form

$$d(t, \mathbf{x}_s, \mathbf{x}_r) \simeq (2\pi)^{-1} a(\mathbf{x}_s, \mathbf{x}_r) \times \int \mathrm{d}\omega \, f(\omega) \exp[\imath\omega(t - T(\mathbf{x}_s, \mathbf{x}_r)]. \tag{3}$$

We will drop the dependence on the source–receiver coordinates, denoting $T = T(\mathbf{x}_s, \mathbf{x}_r)$, but $T_s = (\mathbf{x}_s, \mathbf{x})$, etc. Also, we will not distinguish between perturbations of different components of the stiffness tensor, simply denoting the relevant (scalar) medium parameter by c. The measured or true quantities are denoted with a bar. Hence, $\bar{d}$ denotes the measured data, $\bar{c}$ denotes the true velocity model, etc.

Note that the earlier definitions imply the following conventions for the Fourier transform

$$\hat{g}(\omega) = \int \mathrm{d}t \, g(t) \exp[-\imath\omega t], \tag{4}$$

$$g(t) = \frac{1}{2\pi} \int \mathrm{d}\omega \, \hat{g}(\omega) \exp[\imath\omega t], \tag{5}$$

which is the conjugate of what is usually used in geophysics.

3 CORRELATION-BASED CRITERIA

An important ingredient of a wave-equation traveltime tomography scheme is the procedure to detect the relative phase shift between two waveforms. A straight-forward approach relies on picking the relevant arrivals in the waveforms. More sophisticated techniques rely on correlating the waveforms before picking (e.g. Cara & Leveque 1987). Because we are casting the tomography as an optimization problem, we seek to define a functional that attains a minimum—or maximum—when the relative phase shift between the observed and modelled data is zero. Next, we discuss two such functionals that are based on the correlation of the measured and observed data. Part of this exposition can also be found in Luo & Schuster (1991) and de Hoop & van der Hilst (2005) and is repeated here for sake of completeness.

The correlation of the observed and modelled data is given by

$$C[\bar{d}, d](t) = \int \mathrm{d}t' \, \bar{d}(t' + t) d(t'). \tag{6}$$

Assuming the data are of the form presented in eq. (3) we have

$$C[\bar{d}, d](t) \simeq (2\pi)^{-1} \bar{a} a \int \mathrm{d}\omega \, \hat{F}(\omega) \exp[\imath\omega(t - \Delta T)] = F(t - \Delta T), \tag{7}$$

where $\Delta T = \bar{T} - T$ denotes the phase shift and $\hat{F} = \bar{f} f^*$ is the correlation of the source spectra. We will drop the amplitude factor $a\bar{a}$ in the rest of the paper.

3.1 Picking approach

Luo & Schuster (1991) proposed to extract the relative phase shift between the data by picking the maximum of the correlation:

$$\Delta t = \mathrm{argmax}_t \, C[\bar{d}, d](t). \tag{8}$$

The corresponding misfit functional is given by

$$\phi_{\max} = (\Delta t)^2. \tag{9}$$

However, as has been noted by Hörmann & de Hoop (2002) and de Hoop & van der Hilst (2005), this approach is only strictly valid if the source spectra of the observed and modelled data are identical. Indeed, eq. (7) suggests that the correlation has a maximum at $t = \Delta T$ when $F(t)$ has its maximum at $t = 0$. If, for example, the source spectra differ by a phase rotation over $\pi/2$, F is an odd function with a zero crossing at $t = 0$. Different scenarios are illustrated in Figs 1–3(a). To indicate the difference between the picked and true traveltime difference explicitly, we introduce

$$\epsilon_T = \Delta T - \Delta t. \tag{10}$$

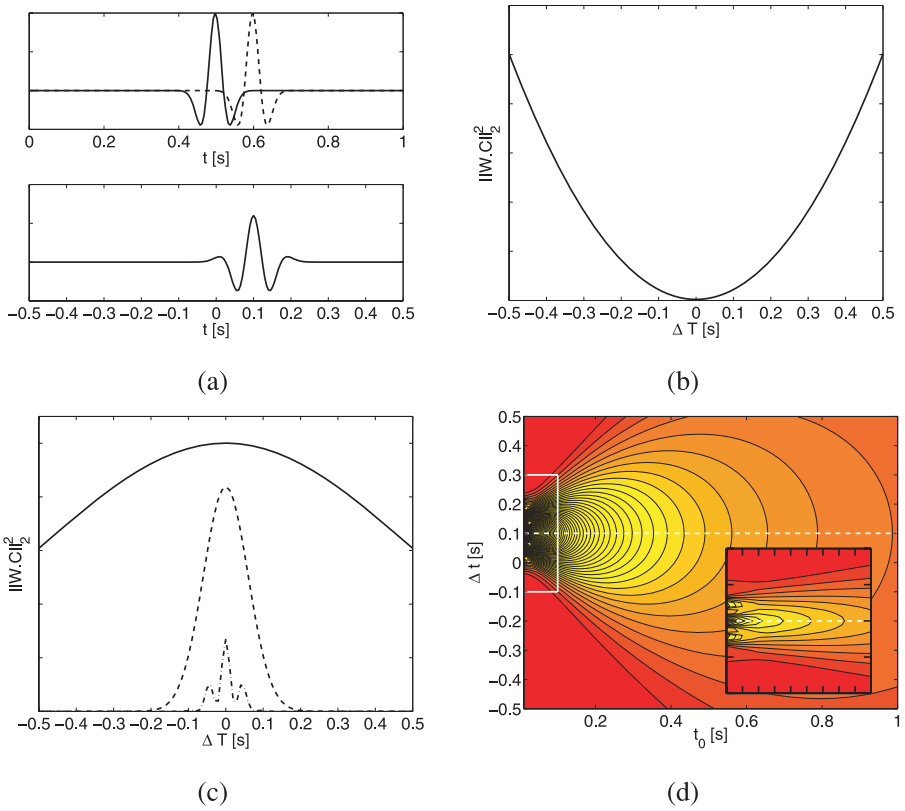

Figure 1. Example of different correlation-based misfit criteria. (a, top panel) Shows two waveforms (Ricker wavelet with a 10 Hz peak frequency) that are phase shifted by $\Delta T = 0.1$ s. The correlation of the waveforms is depicted in (a, bottom panel). The maximum of the correlation coincides with the phase shift. (b) Depicts the weighted norm of the correlation using the linear weight (solid), as a function of the phase shift. The weighed norm using the Gaussian weight is depicted in (c) for $t_0 = 1$ s (solid), $t_0 = 0.1$ s (dash) and $t_0 = 0.01$ s (dash-dot). The latter exhibits local maxima, indicating that t_0 is too small. The FBI transform of the correlation (a), $E_{\Delta t}$, as a function of t_0 and Δt and a fixed frequency of 10 Hz (*cf.* eq. 17) is depicted in (d). At $\Delta t = 0.1$ s, $E_{\Delta t}$ grows fastest as $t_0 \to 0$, indicating the phase shift. For t_0 chosen too small, local maxima occur. This indicates the transition between the picking approach and the weighted norm with the Gaussian weight.

Hence, the misfit functional, as a function of the phase shift ΔT, behaves as

$$\phi_{\max} \simeq (\Delta T + \epsilon_T)^2. \tag{11}$$

Minimizing this will lead to an erroneous velocity when $\epsilon_T \neq 0$.

3.2 Weighted norm approach

We propose an alternative misfit criterion, based on the weighted norm of the correlation

$$\phi_W = ||W \cdot C||_2^2. \tag{12}$$

This approach was introduced by the authors in the context of reflection tomography, or velocity analysis, of active seismic data (van Leeuwen & Mulder 2008, 2010). In Appendix A, we show that the weighted norm can be seen as a correlation of the squared weight with a regularizing kernel that depends on the source spectra:

$$\phi_W(\Delta T) \simeq \int \mathrm{d}t\, W(t)^2 \chi(t - \Delta T), \tag{13}$$

where $\chi(t) = F(t)^2$. So, the location of the minimum, or maximum, of the functional does not depend directly on the behaviour of F anymore, as it did for the picking approach. This resembles techniques were the envelope of the correlation is used for picking, instead of the correlation itself (Gee & Jordan 1992).

We investigate the behaviour of the misfit functional as a function of the phase shift, ΔT, in more detail for two specific choices of the weighting function

$$W_{t_0}^{(1)}(t) = \Omega_{t_0}(t_0)t, \tag{14}$$

$$W_{t_0}^{(2)}(t) = \exp[-(t/t_0)^2], \tag{15}$$

where the rectangle function $\Omega_{t_0}(t) = 1$ when $|t/t_0| \leq 1$ and zero otherwise. The parameter t_0 controls the width of the weighting function and hence the maximal distance over which events are allowed to interact in the correlation.

3.2.1 *Linear weight*

The linear weighting function, $W_{t_0}^{(1)}$, may be interpreted as an annihilator. It is readily verified that when the correlation can be represented as a shifted delta pulse, $C(t) \simeq \delta(t - \Delta T)$, the weighted norm is zero at $\Delta T = 0$. The idea is to penalize energy at non-zero shift. By minimizing, we are selecting a velocity that will focus the energy in the correlation around zero shift. This leads to an approach that is conceptually the same as the annihilator-based approach to velocity analysis that has found widespread use among exploration geophysicists. We refer to Symes (2008) for an excellent overview of these concepts.

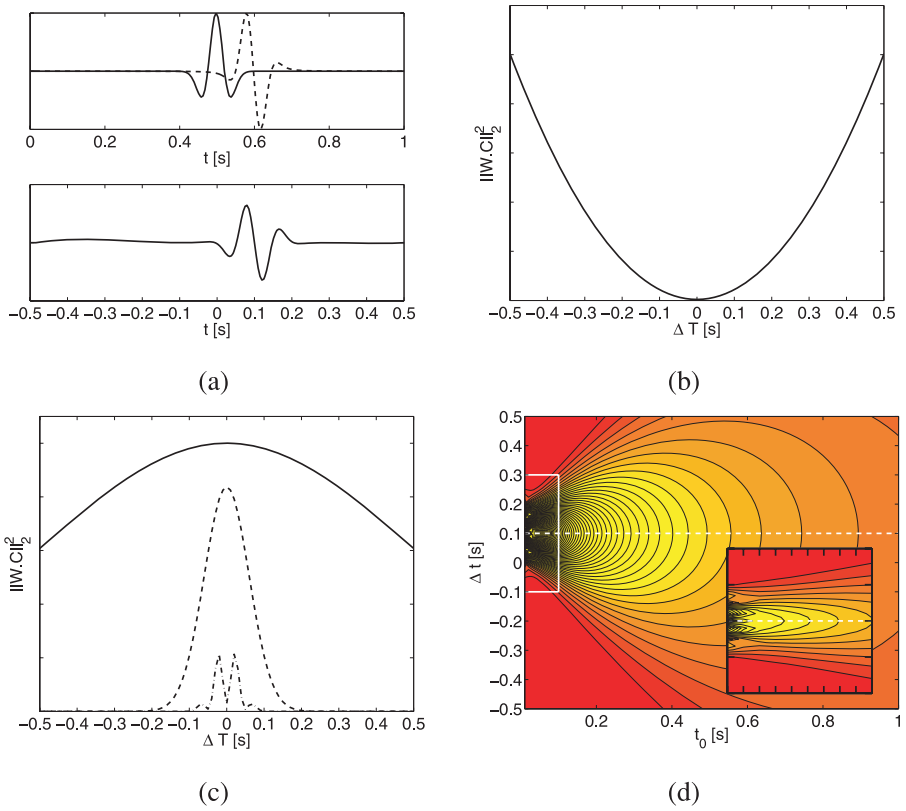

Figure 2. Example of different correlation-based misfit criteria. (a, top panel) Shows two waveforms (Ricker wavelet with a 10 Hz peak frequency) that are phase shifted by $\Delta T = 0.1$ s and one of them has its phase rotated by $\pi/2$. The correlation of the waveforms is depicted in (a, bottom). The maximum of the correlation does not coincide with the phase shift. Instead, the correlation has a zero-crossing at $t = 0.1$ s. (b) Depicts the weighted norm of the correlation (*cf.* eq. 17) using the linear weight (solid), as a function of the phase shift. The weighed norm using the Gaussian weight is depicted in (c) for $t_0 = 1$ s (solid), $t_0 = 0.1$ s (dash) and $t_0 = 0.01$ s (dash-dot). Again, the correlation and exhibits local maxima, indicating that t_0 is too small. The FBI transform of the correlation (a), $E_{\Delta t}$, as a function of t_0 and Δt and a fixed frequency of 10 Hz (*cf.* eq. 17) is depicted in (d). At $\Delta T = 0.1$ s, $E_{\Delta t}$ grows fastest as $t_0 \to 0$, indicating the phase shift. For t_0 chosen too small local maxima occur. This indicates the transition between the picking approach and the weighted norm with the Gaussian weight.

Figs 1–3(b) illustrate the properties of the misfit functional as a function of the phase shift, ΔT, for three different scenarios: the source spectra are identical, phase rotated over $\pi/2$ or phase rotated over $\pi/3$. In all cases, the misfit attains a minimum at $\Delta T = 0$ and is quadratic in ΔT.

Under the assumption that χ is even (this includes phase rotation over $\pi/2$ or a factor $\imath\omega$ between the source spectra), we show in Appendix A that

$$\phi_{W_{t_0}^{(1)}}(\Delta T) \simeq ||W_{t_0}^{(1)} \cdot F||_2^2 + (\Delta T)^2 ||F||_2^2, \tag{16}$$

which confirms the quadratic dependency of the misfit on ΔT. This indicates that there may be more to this approach than is suggested in the annihilator framework, which deals with the propagation of singularities rather than finite-frequency waveforms (*cf.* Stolk & de Hoop 2006). Having made this remark, we leave the matter for further research.

3.2.2 Gaussian weight

The Gaussian weighting function does not act as an annihilator. Instead, the misfit functional attains a maximum at $\Delta T = 0$ (Appendix A). Figs 1–3(c) illustrate this. The picking approach, as explained in Section 3.1, may be derived from this approach by letting $t_0 \downarrow 0$. The Gaussian weight then collapses to a delta function and maximizing the weighted norm is equivalent to maximizing the correlation at zero shift. This can be seen from Figs 1–3(c,d), where the Gaussian-weighted norm develops local maxima for very small t_0.

A tentative link may be made with an approach discussed by Hörmann & de Hoop (2002) (see also de Hoop & van der Hilst 2005). They propose to detect the phase shift by a time–frequency analysis of the correlation as follows. Consider the windowed Fourier transform of the correlation:

$$E_{\Delta t}(t_0, \omega) = \int \mathrm{d}t\, W_{t_0}^{(2)}(t - \Delta t) C(t) \exp[-\imath\omega t]. \tag{17}$$

This is also known as the FBI (Fourier-Bros-Iagolnitzer) or Gabor transform (*cf.* Folland 1989; Strichartz 2003). The singular support of the correlation (i.e. the location of the singularity) may be detected by inspecting the growth properties of $E_{\Delta t}(t_0, \omega)$ as $t_0 \downarrow 0$. More precisely, when Δt is *not* in the singular support of the correlation we have, for any $N \in \mathbb{N}$,

$$|E_{\Delta t}(t_0, \omega)| \leq c_N (t_0)^N. \tag{18}$$

(*cf.* Hörmann & de Hoop 2002). So when Δt does *not* coincide with the singular support of the correlation, $E_{\Delta t}(t_0, \omega)$ should be rapidly decreasing as $t_0 \downarrow 0$. By inspecting the graph of $E_{\Delta t}(t_0, \omega)$ as a function of t_0 for each Δt and a fixed ω we should select the Δt for which $E_{\Delta t}(t_0, \omega)$ grows fastest as $t_0 \downarrow 0$. This is clearly illustrated in Figs 1–3(d). Again, for very small t_0 this approach reduces to the picking approach.

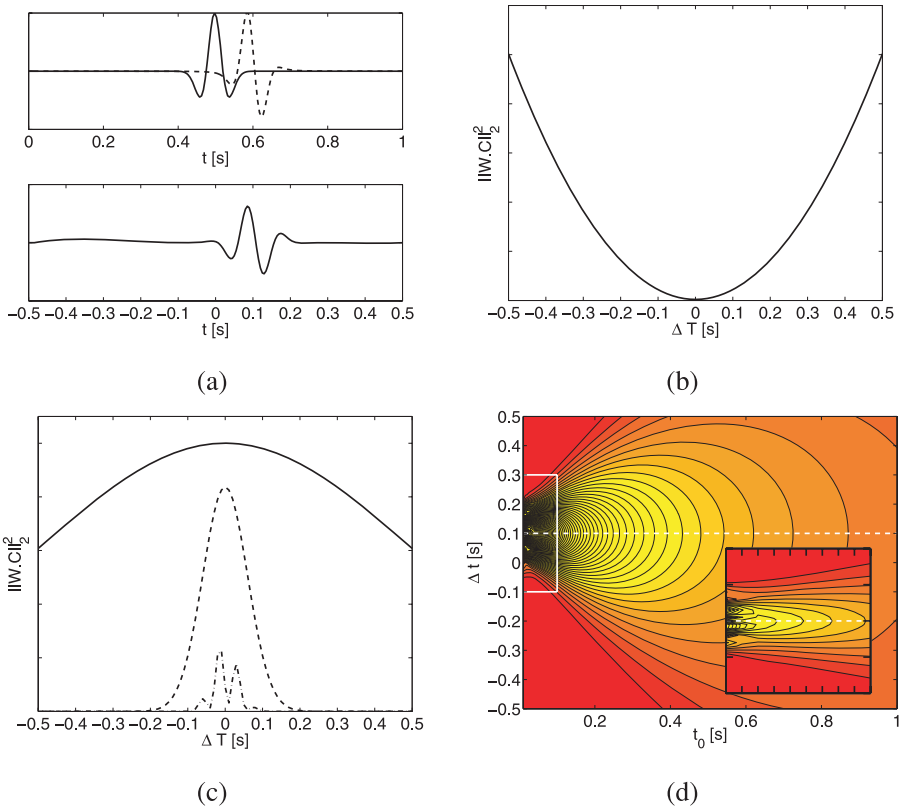

Figure 3. Example of different correlation-based misfit criteria. (a, top panel) Shows two waveforms (Ricker wavelet with a 10 Hz peak frequency) that are phase shifted by $\Delta T = 0.1$ s and one of them has its phase rotated by $\pi/3$. The correlation of the waveforms is depicted in (a, bottom panel). The maximum of the correlation does not coincide with the phase shift. (b) The weighted norm of the correlation (*cf.* eq. 17) using the linear weight (solid), as a function of the phase shift. The weighed norm using the Gaussian weight is depicted in (c) for $t_0 = 1$ s (solid), $t_0 = 0.1$ s (dash) and $t_0 = 0.01$ s (dash-dot). Again, the latter exhibits local maxima, indicating that t_0 is too small. The FBI transform of the correlation, $E_{\Delta t}$, as a function of t_0 and Δt and a fixed frequency of 10 Hz (*cf.* eq. 17) is depicted in (d). At $\Delta T = 0.1$ s, $E_{\Delta t}$ grows fastest as $t_0 \to 0$, indicating the phase shift. For t_0 chosen too small local maxima occur. This indicates the transition between the picking approach and the weighted norm with the Gaussian weight.

The weighted norm approach arises when we consider $E_{\Delta t}$ for a fixed t_0 and all frequencies. In fact, it is readily verified that

$$||E_0(t_0, \cdot)||_2^2 = ||W_{t_0}^{(2)} C||_2^2, \tag{19}$$

which in light of the above seems a reasonable way to measure the shift of the singular support.

4 SENSITIVITY KERNELS

The sensitivity kernel relates perturbations in the medium parameters to changes in the misfit. It should be noted that the sensitivity kernel only gives a meaningful update of the model if it is consistent with the misfit (de Hoop & van der Hilst 2005). Although the derivation of such kernels is rather straightforward we include it for completeness' sake.

Via a Taylor-series argument we have

$$\phi[c_0 + \delta c] - \phi[c_0] = \int \mathrm{d}\mathbf{x}\, \partial_c \phi(\mathbf{x}) \delta c(\mathbf{x}) + \mathcal{O}(||\delta c||_2^2), \tag{20}$$

where $\partial_c \phi$ $(\mathbf{x})$ is the sensitivity kernel, or the Fréchet derivative of the misfit functional w.r.t. to the medium parameters. In the context of transmitted body waves, one usually employs the Born approximation to derive this kernel, which can then be efficiently computed by correlating a forward and adjoint wavefield (see, e.g. Tarantola 1984). We briefly illustrate this procedure with an example. Concentrating on a single component, one source–receiver pair and a scalar medium perturbation, we assume that the wavefield obeys a scalar wave equation

$$\underbrace{\left[c^{-2}\partial_t^2 - \nabla^2\right]}_{\mathcal{L}[c]} u = f\delta(\cdot - \mathbf{x}_s). \tag{21}$$

We apply a Born approximation, writing $c = c_0 + \delta c$, $u = u_0 + \delta u$

$$\mathcal{L}[c_0] u_0(t, \mathbf{x}) = f(t)\delta(\mathbf{x} - \mathbf{x}_s), \tag{22}$$

$$\mathcal{L}[c_0]\delta u(t, \mathbf{x}) = 2\frac{\delta c(\mathbf{x})}{c_0(\mathbf{x})^3}\partial_t^2 u_0(t, \mathbf{x}). \tag{23}$$

We also introduce the wavefield v_0 satisfying

$$\mathcal{L}^*[c_0] v_0 = \partial_d \phi(t)\delta(\mathbf{x} - \mathbf{x}_r), \tag{24}$$

where $\partial_d \phi$ is the Fréchet derivative of the misfit functional w.r.t. to the data, or the *adjoint source*. The sensitivity kernel is then given by

$$\partial_c \phi(\mathbf{x}) = \frac{2}{c_0(\mathbf{x})^3}\int \mathrm{d}t\, v_0(t, \mathbf{x})\partial_t^2 u_0(t, \mathbf{x}). \tag{25}$$

For more details concerning the adjoint-state technique, we refer to Plessix (2006).

Asymptotically, the kernel is then given by (dropping the factor $2c_0^{-3}$):

$$\partial_c \phi(\mathbf{x}) \simeq a_s a_r \int \mathrm{d}\omega\, \widehat{\partial_d \phi}(\omega)\omega^2 f^*(\omega) \exp[\imath\omega(T_s + T_r)]. \tag{26}$$

We derive the adjoint sources for the different correlation-based approaches in Appendix B. This yields the following expressions for the sensitivity kernels:

$$\partial_c \phi(\mathbf{x}) = a_s a_r \int \mathrm{d}\omega\, \omega^2 \hat{F}(\omega)\hat{H}(\omega) \exp[\imath\omega(T_s + T_r - T)]. \quad (27)$$

The factor $\hat{H}$ depends on the misfit criterion, which for source signatures that are band-limited delta functions is given by

$$\hat{H}_{\max}(\omega) \simeq (\text{amplitude terms}) \times \imath\omega\Delta T \exp[\imath\omega\epsilon_T], \quad (28)$$

$$\hat{H}_W(\omega) \simeq (\text{amplitude terms}) \times W(\Delta T)^2. \quad (29)$$

When $\hat{F}$ is real and the velocity varies only mildly, eq. (27) leads to the famous banana–doughnut shaped kernel with zero sensitivity along the central ray (Marquering *et al.* 1999). When $\hat{F}$ is not real, which happens in general when the source spectra are not identical, the shape of the kernel changes and loses the zero-sensitivity along the ray. This is a subtle, but important, point first made by de Hoop & van der Hilst (2005).

The kernel for the weighted norm approach does not have zero-sensitivity along the ray. For general source spectra, we find that the factor $\hat{H}_W$ contains a factor $\hat{F}^*$. Under the assumption that χ is an even function, the factor $\hat{F}\hat{H}_W$ (*cf.* eq. 27) is real and even and defines an even function around the central ray. Therefore, the sensitivity kernel for the weighted norm approach has non-zero sensitivity along the central ray, as long as χ is even. We might call these kernels bananas instead of banana–doughnuts. This confirms our earlier assertion that the weighted-norm approach is less sensitive to differences in the source spectra. A similar, more detailed multiresolution analysis of the sensitivity kernel for the picking approach is given by de Hoop & van der Hilst (2005).

To illustrate the above-mentioned properties, we calculate the adjoint sources for the different approaches for two different scenarios: (1) the source spectra are identical and (2) the source spectra are phase rotated over $\pi/2$ w.r.t. each other. The different adjoint sources are depicted in Fig. 4. As predicted, the phase rotation leads to a phase shift in the adjoint source for the picking approach. The adjoint sources for the weighted-norm approach respond to the phase rotation of the source spectrum by phase rotation, as argued earlier.

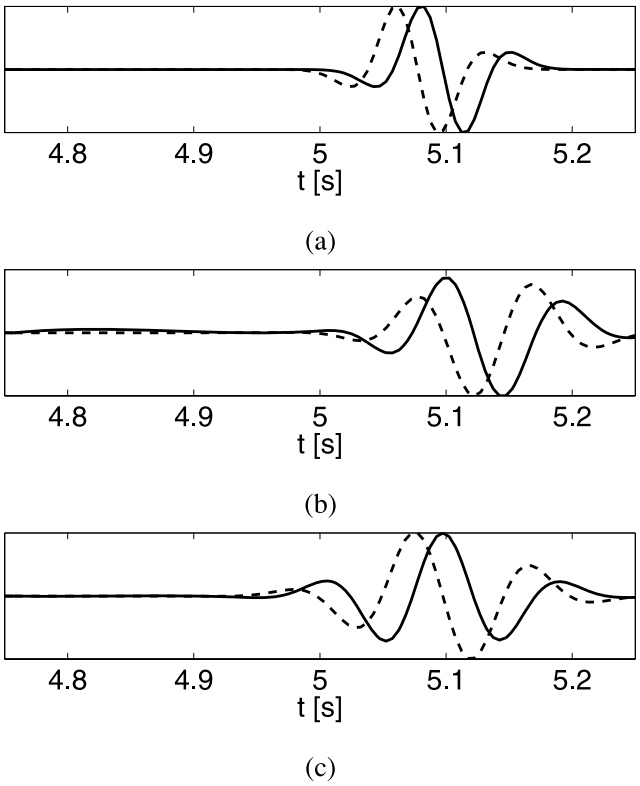

Figure 4. Example of adjoint sources for the different correlation-based approaches: (a) picking, (b) linear weight and (c) Gaussian weight for $t_0 = 1$. The waveforms are phase-shifted by 0.1 s and their spectra are either the same (solid) or phase-rotated over $\pi/2$ (dash) w.r.t. each other. The shift of the adjoint sources for the picking approach depends on the source spectra. For the weighted norm approach, the phase rotation of the source spectrum causes a phase rotation of the adjoint source.

5 EXAMPLE: DIVING WAVE TOMOGRAPHY

We illustrate the properties of the sensitivity kernels derived in the previous section for constant-density acoustic velocity models that increase linearly with depth

$$c(x_3) = c_0 + \alpha x_3. \quad (30)$$

The 3-D Green's function for such models can be expressed analytically (Pekeris 1946; Kuvshinov & Mulder 2006) and is given by

$$G(\omega, \mathbf{x}, \mathbf{x}') = (4\pi\sqrt{cc_0(\rho^2-1)/\alpha})^{-1} \times \exp[\imath\sqrt{(\omega/\alpha)^2 - 1/4}\,\mathrm{arccosh}(\rho)], \quad (31)$$

where

$$\rho = 1 + \frac{1}{2}\frac{||\mathbf{x}-\mathbf{x}'||_2}{(x_3 + c_0/\alpha)(x_3' + c_0/\alpha)}. \quad (32)$$

The kernel is explicitly given by

$$\partial_c \phi(\mathbf{x}, \mathbf{x}_s, \mathbf{x}_r) = \int \mathrm{d}\omega\, \omega^2 f^*(\omega)\widehat{\partial_d \phi}(\omega, \mathbf{x}_s, \mathbf{x}_r) G(\omega, \mathbf{x}_s, \mathbf{x})^* G(\omega, \mathbf{x}, \mathbf{x}_r)^*. \quad (33)$$

Figs 5–7 illustrate the different kernels for $c_0 = \bar{c}_0 = 1500$ and $\alpha = 0.7\,\mathrm{s}^{-1}, \bar{\alpha} = 0.5\,\mathrm{s}^{-1}$, again for identical and phase-rotated source spectra. The central ray is also indicated. The kernel for the picking approach is affected by the phase rotation of the source spectra, while the weighted norm kernels are visibly indiscernible, confirming our earlier assertion.

6 APPLICATION: CROSS-WELL TOMOGRAPHY

We use the weighted norm of the correlation with the Gaussian weight as a misfit criterion for cross-well tomography. That is, we solve the optimization problem

$$c_{\text{optimal}} = \mathrm{argmax}_c\, \phi_{W_{t_0}^{(2)}}[c]. \quad (34)$$

The data are modelled with a frequency-domain finite-difference code (Mulder & Plessix 2002). We employ a limited-memory BFGS (Broyden–Fletcher–Shanno–Goldfarb) method to solve the optimization problem. As regularization, we represent the model on linear splines on a coarse grid. The gradient is calculated with the adjoint-state technique. More details on how to implement the

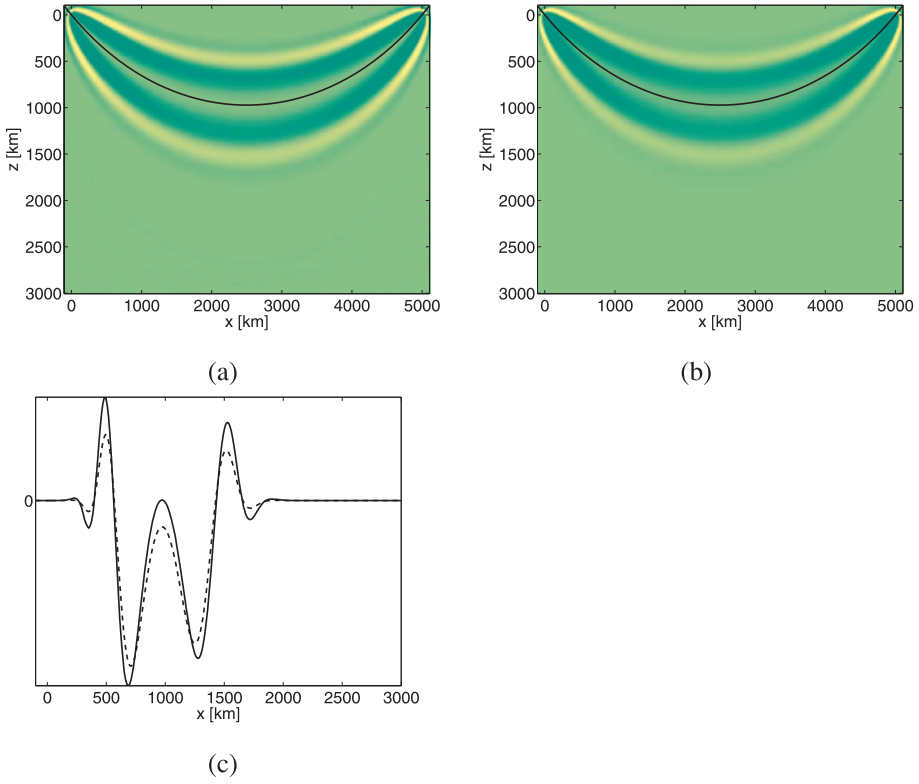

Figure 5. Example of sensitivity kernels for the picking approach for a velocity that increases linearly with depth. [a, c(solid)] Depicts the kernel where the source spectra are identical, whereas (b) and (c, dash) depicts a situation where the source spectra are phase rotated over $\pi/2$. The slice through the kernels at $x = 2500$ m shows that kernel has zero sensitivity on the central ray when the source spectra are identical and non-zero sensitivity when the source spectra are not identical.

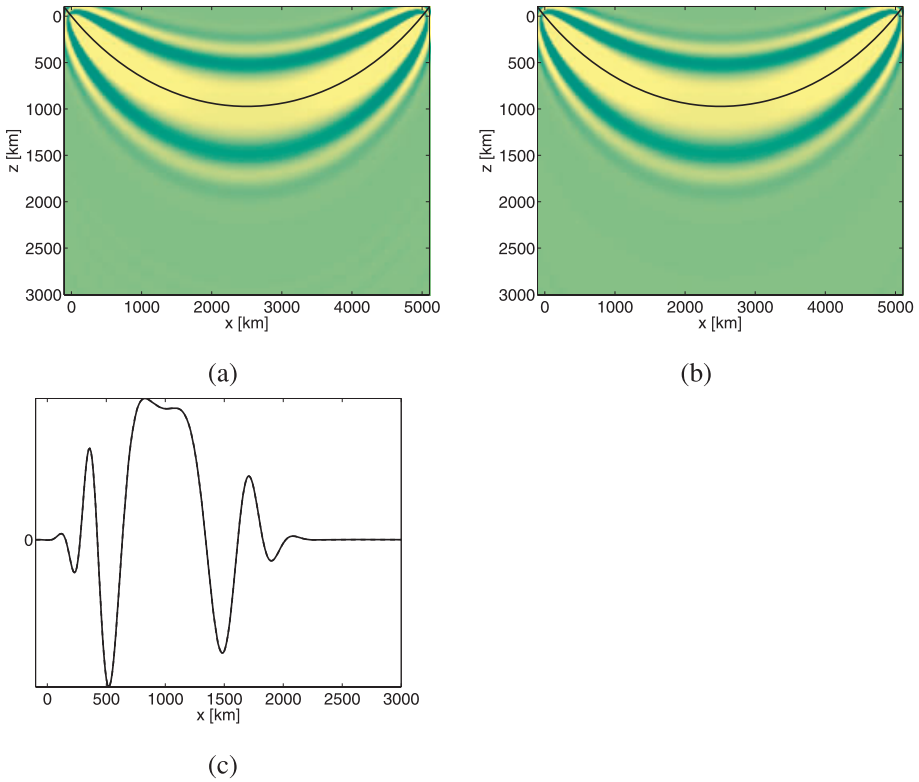

Figure 6. Example of sensitivity kernels for the weighted norm approach with the linear weight. [a, (c, solid)] Depicts the kernel where the source spectra are identical, whereas (b) and (c, dash) depicts a situation where the source spectra are phase rotated over $\pi/2$. The kernel is not affected by the phase-rotation.

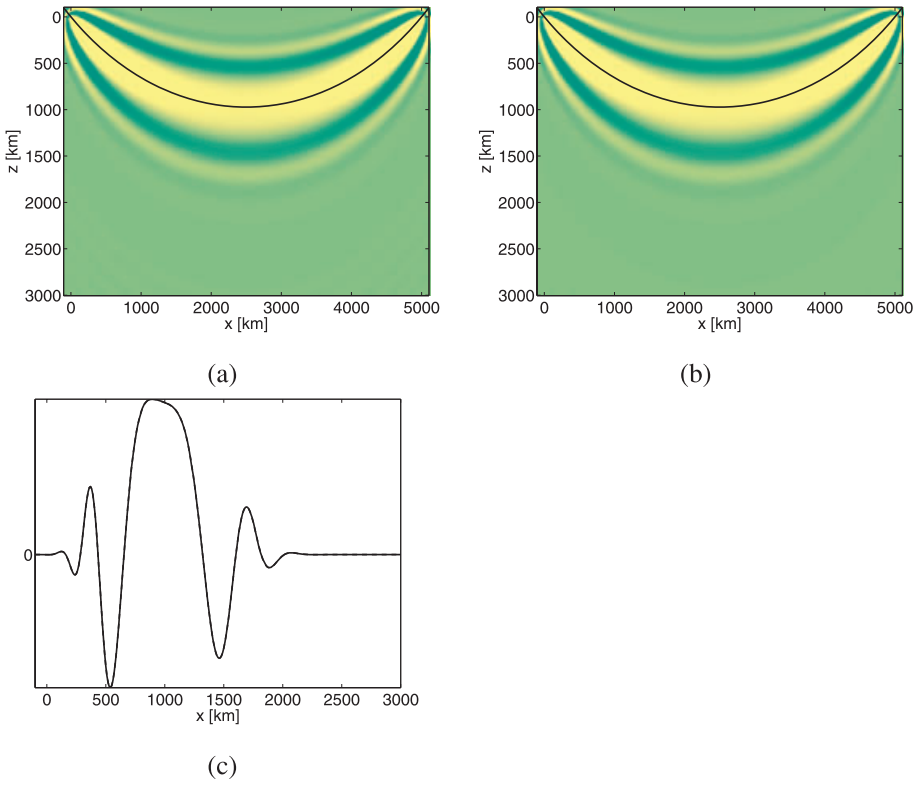

Figure 7. Example of sensitivity kernels for the weighted norm approach with the Gaussian weight. [a, (c, solid)] Depicts the kernel where the source spectra are identical, whereas (b) and (c, dash) depicts a situation where the source spectra are phase rotated over $\pi/2$. The kernel is not affected by the phase-rotation.

adjoint calculations can be found in (e.g. Tromp *et al.* 2005; Plessix 2006).

The cross-well data set, which was also use for a virtual source study (Mehta *et al.* 2008), contains 147 shot positions in the well at $x = 205.74$ m between 7.19 and 567.9 m depth at a 3.84 m interval and 150 receiver positions at $x = 26.82$ m between 6.63 and 578.86 m depth at the same depth interval. The source and 3C geophones were clamped to the casing.

We used the x-component of the data for an x-source to estimate the P velocity and the y-component of the data for a y-source to estimate the S velocity separately. We assumed that the separate components were not too different from acoustic pressure data from an explosive source. The data where damped to emphasize the first arrival by multiplication with $e^{-\gamma(t-t_r)}$. This translates into a complex angular frequency with γ as its imaginary part. We inverted frequencies from 70 to 160 Hz with an interval of 2 Hz. Attenuation is taken into account: $\tilde{c} = c/\sqrt{1+(\imath - a)/Q}$, $a = (2/\pi)\log(f/f_r)$. The free surface was also modelled by imposing a zero-pressure boundary condition at the top of the model. As a wavelet we used a band-limited delta function.

The inversion was done in several stages. First, an optimal (i.e. resulting in the lowest misfit) linear velocity $c = c_0 + \alpha x_3$ was sought at a fixed quality factor $Q = Q_r$. Then, we inverted for a layered model, represented by linear splines at a 15.24 m interval, again at a fixed $Q = Q_r$. Finally, we allowed Q to vary and used a finer spline grid to represent the velocity. The inversion parameters for the P and S velocities are given in Table 1.

The final results are shown in Figs 8 and 9. We compare the final models with the interpolated well-logs. This shows that the general trend of the velocity matches pretty well. Also, the traveltimes are reasonably well explained, as can be seen from the correlation panels. To allow for visual inspection of the correlation panels we normalized the correlation by dividing by the absolute value of the spectrum.

Table 1. Inversion parameters for P and S velocities.

	γ	t_r	Q_r	f_r	c_0
P	8	0.05	80	100	1551
S	4	0.1	80	100	430

A perfect match, in terms of velocity and traveltime, is not be expected because important effects such as anisotropy have not been taken into account. Also, the damping parameters used do probably not reflect the reality, which may seriously effect the result. We have generated synthetic data for the interpolated well-logs and these data did not give a significantly better traveltime match than our inverted model. This indicates that the resulting traveltime and velocity errors are most probably caused by systematic modelling errors. We did not try to adjust the damping parameters or introduce anisotropy to improve the results, as this example serves as a proof of concept of the method.

7 DISCUSSION

The weighted norm of the correlation can be seen as a regularized version of the correlation and resembles techniques that use the envelope or a time–frequency analysis of the correlation to pick the phase shift. The main difference with such approaches is that the weighted norm can be used directly as an optimization criterion, without picking. In particular, this makes it easier to derive the corresponding sensitivity kernels, as we do not need to use implicit relations between the maximum of the correlation and the phase shift.

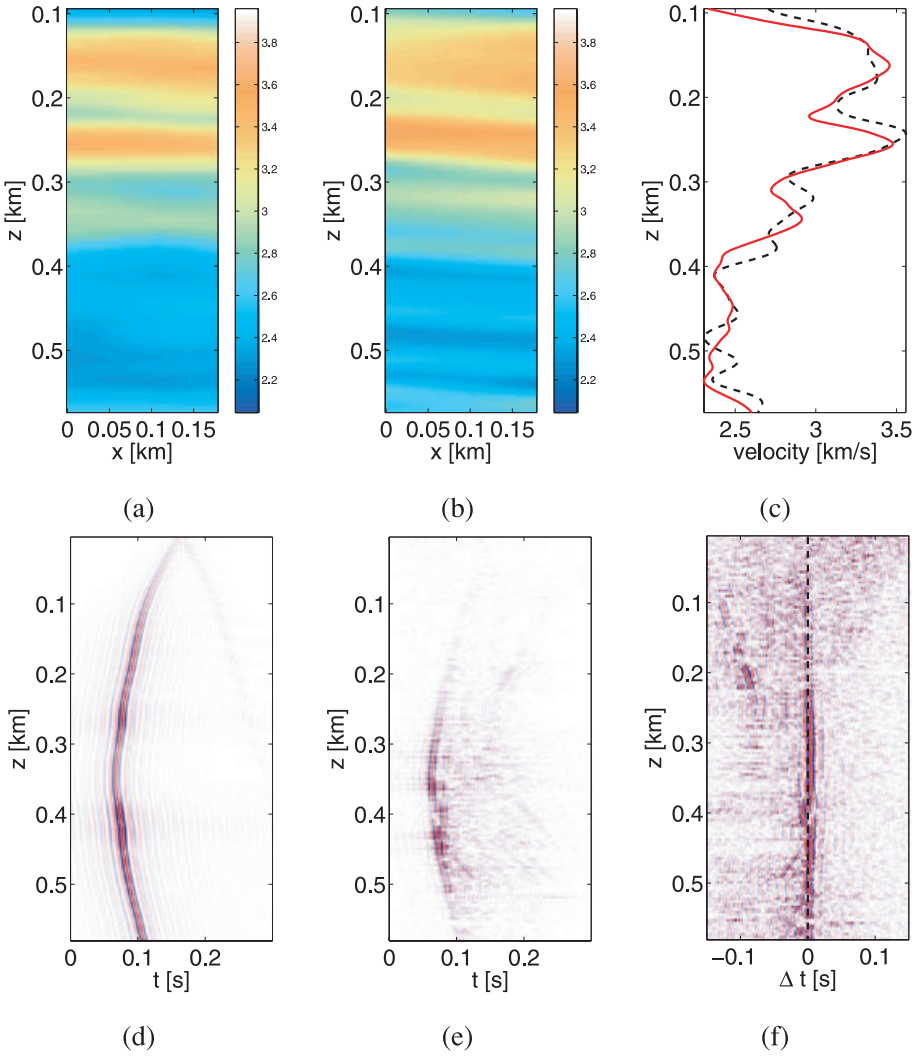

Figure 8. Inversion of the x-component of cross-well data for an x-source using the weighted norm of the correlation as optimization criterion. We used the Gaussian as weighting function. (a) Depicts the final velocity, (b) depicts the interpolated and smoothed well-logs. (c) Depicts a central slice through the final (solid) and well-log (dash) models. (d,e) Show a shotgather at 387.4 m of the modelled and observed data. The correlation of the latter two is depicted in (f). To allow for visual inspection, the correlation is normalized by dividing by the spectrum.

The behaviour of the sensitivity kernel for the picking approach is determined by the correlation of the source spectra, $\hat{F}(\omega) = \bar{f}(\omega) f^*(\omega)$. If, for example, the source spectra differ by a phase rotation, this will show up in $\hat{F}$ and hence directly in the kernel, causing the zero-sensitivity along the central ray to disappear. In the sensitivity kernels for the weighted norm approach an extra factor $\hat{F}^*$ appears, making these kernels less sensitive to errors in the source spectra. For example, phase rotation or differentiation of the source spectrum will not dramatically affect the weighted norm approach. The link to the time–frequency analysis, presented in Section 3, suggests that the weighting function plays a vital role in mitigating this sensitivity. If we choose the width of the Gaussian weight too small, the weighted norm degrades to the picking approach. If we choose it too large, we might not be sensitive enough to time shifts. The width of the Gaussian should somehow reflect the frequency content of the data. One might even envision a scheme were different frequency bands of the data are treated separately with an optimal width of the Gaussian.

Intuitively, the weighted norm approach should be able to handle multiple arrivals at the same time, as long as each arrival in the observed data has a corresponding arrival in the modelled data. An issue that arises here is the cross-talk between different arrivals. Moreover, the correlation could be extended to measure phase differences of whole wavefields. Then, not only the shift of the singular support but also the slowness difference between the wavefields can be measured. This would allow for a wave-equation stereo-tomography approach.

8 CONCLUSION

We have reviewed several correlation-based misfit functionals that are sensitive to the time shift between complex waveforms. Such misfit functionals are used to cast wave-equation traveltime tomography as a PDE-constrained optimization problem. Ideally, the misfit functional attains a minimum—or maximum—if the phase shift between the measured and observed data is zero. The commonly used procedure that relies on picking the maximum of the correlation has the drawback that it is very sensitive to errors in the estimated source signature. When the modelled and observed source

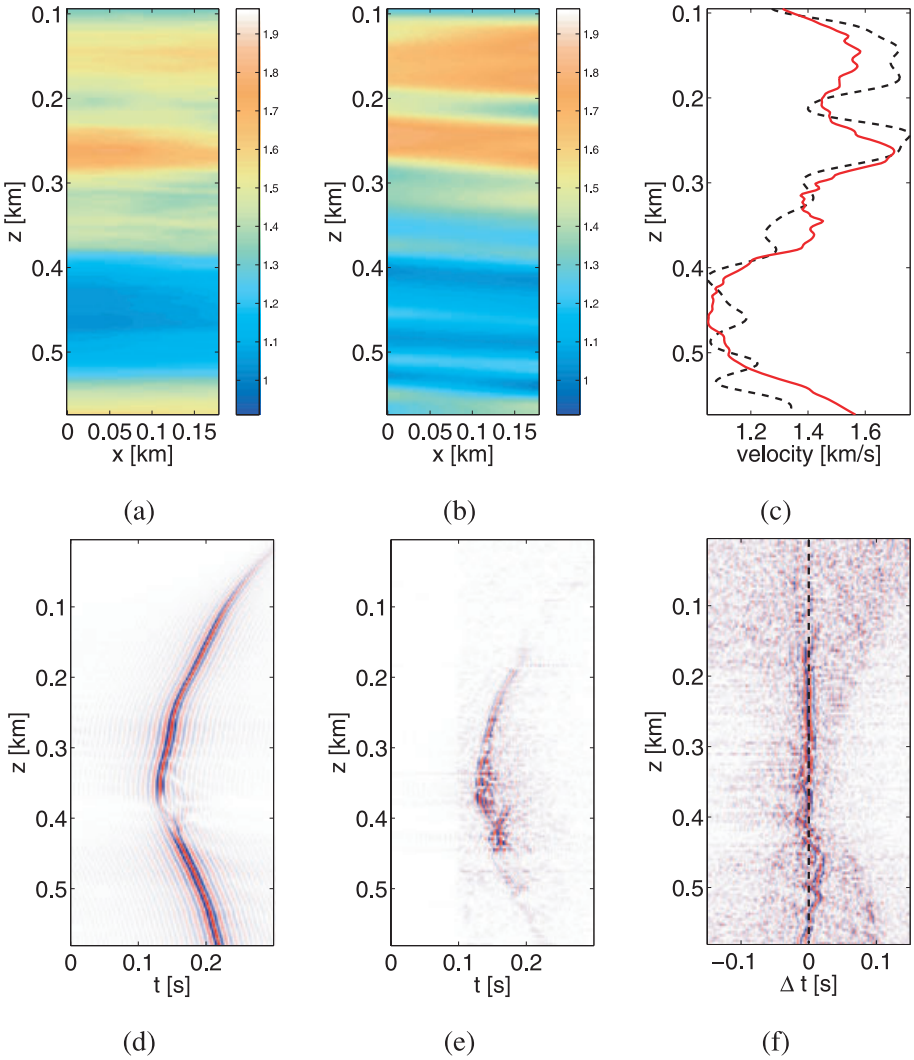

Figure 9. Inversion of the *y*-component of cross-well data for a *y*-source using the weighted norm of the correlation as optimization criterion. We used the Gaussian as weighting function. (a) Depicts the final velocity, (b) depicts the interpolated and smoothed well-logs. (c) Depicts a central slice through the final (solid) and well-log (dash) models. (d,e) Show a shotgather at 387.4 m of the modelled and observed data. The correlation of the latter two is depicted in (f). To allow for visual inspection, the correlation is normalized by dividing by the spectrum.

spectra are not identical, the corresponding misfit functional may attain its minimum at non-zero phase shift. As an alternative to the picking approach, we propose to use a weighted norm of the correlation as misfit functional. We discuss two particular instances of this approach, using a linear and a Gaussian weighting function. For the linear weighting function, we may view the procedure as an annihilator-based criterion which attains a minimum when the phase shift is zero. Using the Gaussian weight, the weighted norm attains a maximum. We show that this approach is intimately connect to techniques from time–frequency analysis.

The sensitivity kernels corresponding to the weighted norm approach do not exhibit the characteristic hole in the centre and are less sensitive to errors in the source spectra. One might say that these kernels look more like regular bananas instead of banana–doughnuts. We illustrate the properties of the sensitivity kernels for a velocity model that increase linearly with depth, using an analytic solution of the constant-density acoustic wave-equation.

The cross-well examples illustrate that the weighted norm approach is a viable alternative to the picking approach which is more robust w.r.t. errors in the estimated source spectrum.

ACKNOWLEDGMENTS

This work is part of the research programme of the Foundation for Fundamental Research on Matter (FOM), which is financially supported by the Netherlands Organisation for Scientific Research (NWO). The authors thank Shell Oil for permission to use the cross-well data.

REFERENCES

Cara, M. & Leveque, J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.,* **14**(10), 1046–1049.

de Hoop, M.V. & van der Hilst, R.D., 2005. On sensitivity kernels for 'wave-equation' transmission tomography, *Geophys. J. Int.,* **160**(3), 621–633.

Folland, G., 1989. *Harmonic Analysis in Phase Space,* Annals of Mathematics Studies, Princeton University Press, Princeton, NJ.

Gee, L. & Jordan, T., 1992. Generalized seismological data functionals, *Geophys. J. Int.,* **111**(2), 363–390.

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.

Kuvshinov, B. & Mulder, W.A., 2006. The exact solution of the time-harmonic wave equation for a linear velocity profile, *Geophys. J. Int.,* **167**(2), 659–662.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), 645–653.

Marquering, H., Nolet, G. & Dahlen, F., 1999. Three-dimensional sensitivity kernels for finite frequency traveltimes: the banana-doughnut paradox, *Geophys. J. Int.,* **132**(3), 521–534.

Mehta, K., Bakulin, A., Kiyashchenko, D. & Lopez, J., 2008. Comparing virtual versus real crosswell surveys, *SEG Expanded Abstracts,* **27,** 1372–1376.

Mulder, W.A. & Plessix, R.-E., 2002. Time-versus frequency-domain modelling of seismic wave propagation, *EAGE Expanded Abstracts,* pp. E015–E015.

Pekeris, C., 1946. Theory of propagation of sound in a half-space of variable sound velocity under conditions of the formation of a shadow zone, *J. acoust. Soc. Am.,* **18,** 295–315.

Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Stolk, C.C. & de Hoop, M.V., 2006. Seismic inverse scattering in the downward continuation approach, *Wave Motion,* **43,** 579–598.

Strichartz, R., 2003. *A Guide to Distribution Theory and Fourier Transforms,* World Scientific, Singapore.

Symes, W.W., 2008. Migration velocity analysis and waveform inversion, *Geophys. Prospect.,* **56**(6), 765–790.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tarantola, A. & Valette, A., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.,* **20**(2), 129–232.

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

van Leeuwen, T. & Mulder, W.A., 2008. Velocity analysis based on data correlation, *Geophys. Prospect.,* **56**(6), 791–803.

van Leeuwen, T. & Mulder, W.A., 2010. A comparison of velocity inversion methods for layered acoustics, *Inverse problems,* **26**(1).

Vogel, C.R., 2002. *Computational Methods for Inverse Problems,* vol. 23 of Frontiers in Applied Mathematics, SIAM, Philadelphia.

APPENDIX A: WEIGHTED NORM OF THE CORRELATION

In the following we will drop the amplitude and source–receiver dependence from the notation, expressing the data as

$$d(t) = (2\pi)^{-1} \int \mathrm{d}\omega\, f(\omega) \exp[\imath\omega(t - T)]. \tag{A1}$$

The correlation is then given by

$$C(t) = (2\pi)^{-1} \int \mathrm{d}\omega\, \hat{F}(\omega) \exp[\imath\omega(t - \Delta T)] \tag{A2}$$

$$= F(t - \Delta T), \tag{A3}$$

where $\hat{F} = \bar{f} f^*$ is the correlation of the source spectra. The weighted norm of the correlation, as a function of the phase shift, is then given by

$$\begin{aligned}\phi_W(\Delta T) &= \int \mathrm{d}t\ (W(t)F(t - \Delta T))^2 \\ &= \int \mathrm{d}t\, B(t + \Delta T)\chi(t),\end{aligned} \tag{A4}$$

where $B = W^2$ and $\chi = F^2$. The behaviour of the functional, as a function of ΔT depends intricately on the interplay between the weighting function and the correlation of the source spectra. In the following, we assume that χ is an even function. This restricts the validity of the analysis to cases where $\hat{F}$ is either purely real or purely imaginary (this includes the source spectra differing by a phase rotation over $\pi/2$ degrees or a factor $\imath\omega$). The example in Fig. 3 suggests, however, that the validity of the approach is not restricted to this case. Also, the more general result on the detection of wavefront sets by using the Gaussian weight (Hörmann & de Hoop 2002) suggests a more general validity. The following analysis is meant to illustrate why the weighted norm approach might work.

We investigate the behaviour of the misfit functional as a function of ΔT. The first derivate is given by

$$\phi'_W(\Delta T) = \int \mathrm{d}t\, B'(t + \Delta T)\chi(t). \tag{A5}$$

Since χ is even and B' is odd, $\phi'_W(0) = 0$. For the specific cases $B(t) = t^2$ (linear weight) and $B(t) = e^{-2t^2}$ (Gaussian weight), it is readily verified that $\Delta T = 0$ is the *only* stationary point of ϕ'_W. The second derivative is given by

$$\phi''_W(\Delta T) = \int \mathrm{d}t\, B''(t + \Delta T)\chi(t). \tag{A6}$$

For the linear weight, $B''(0) = 2$, so $\phi''_W(0) > 0$, i.e. the misfit functional has a minimum at $\Delta T = 0$. Likewise, we find that the misfit functional has a maximum for the Gaussian weight.

For the linear weight, we can derive an even more explicit expression

$$\begin{aligned}\phi_{W^{(1)}}(\Delta T) &= \int \mathrm{d}t\, (t + \Delta T)^2 \chi(t) \\ &= (\Delta T)^2 ||F||_2^2 + 2\Delta T \int \mathrm{d}t\, t\, \chi(t) + ||W^{(1)} \cdot F||_2^2.\end{aligned} \tag{A7}$$

The second term vanishes because the integrand is odd.

APPENDIX B: SENSITIVITY KERNELS

The sensitivity kernel is given by

$$\partial_c \phi \simeq \int \mathrm{d}\omega\, \omega^2 f^*(\omega) \widehat{\partial_d \phi}(\omega) \exp[\imath\omega(T_s + T_r)]. \tag{B1}$$

The adjoint sources for the different functionals are given by

$$\partial_d \phi_{\max}(t) = -2\Delta t \frac{\partial_t \bar{d}(t + \Delta t)}{C[\partial_t^2 \bar{d}, d](\Delta t)} \tag{B2}$$

$$\partial_d \phi_W(t) = \int \mathrm{d}t'\, W(t')^2 C(t') \bar{d}(t' + t). \tag{B3}$$

For a derivation of eq. (B2), we refer to (Luo & Schuster 1991).

Upon substituting the asymptotic expression for the data in eq. (B2), we get

$$\widehat{\partial_d \phi}_{\max}(\omega) \simeq \frac{2(\Delta T + \epsilon_T)}{C[\partial_t^2 \bar{d}, d](\Delta t)} \imath\omega \bar{f}(\omega) \exp[\imath\omega(\epsilon_T - T)]. \tag{B4}$$

This leads to the kernel for the picking approach

$$\begin{aligned}\partial_c \phi_{\max} = &\frac{2(\Delta T + \epsilon_T)}{C[\partial_t^2 \bar{d}, d](\Delta t)} \\ &\times \int \mathrm{d}\omega\, \imath\omega^3 \hat{F}(\omega) \exp[\imath\omega(T_s + T_r - T + \epsilon_T)].\end{aligned} \tag{B5}$$

For the weighted-norm approach we find, by inserting the asymptotic expressions for the data into eq. (B3) and taking the Fourier

transform:

$$\widehat{\partial_d \phi}_W(\omega) \simeq \int \mathrm{d}t \int \mathrm{d}t' \int \mathrm{d}\omega' \int \mathrm{d}\omega'' \, \hat{F}(\omega') \bar{f}(\omega'') B(t') \times \exp[-\imath\omega t + \imath\omega'(t' - \Delta T) + \imath\omega''(t' + t - \bar{T})]. \quad \text{(B6)}$$

In case the source spectra are band-limited delta functions, the integrals can be collapsed directly to yield

$$\widehat{\partial_d \phi}_W(\omega) \simeq B(\Delta T) \exp[-\imath\omega T], \quad \text{(B7)}$$

which leads to

$$\partial_c \phi_W \simeq B(\Delta T) \int \mathrm{d}\omega \, \exp[\imath\omega(T_s + T_r - T)]. \quad \text{(B8)}$$

For general source spectra, we may insert the expression into eq. (B1). Rewriting in terms of the Fourier transform of B, we get

$$\partial_c \phi_W \simeq \int \mathrm{d}\omega \int \mathrm{d}\omega' \, \omega^2 \hat{F}(\omega) \hat{F}^*(\omega') \hat{B}(\omega' - \omega) \times \exp[\imath\omega(T_s + T_r - T) + \imath(\omega' - \omega)(\Delta T)]. \quad \text{(B9)}$$