A comparison of seismic velocity inversion methods for layered acoustics

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

[1]Delft University of Technology, Department of Geotechnology, P.O. Box 5048, 2600 GA Delft, The Netherlands.
[2]Shell International Exploration and Production, P.O. Box 60, 2280 AB Rijswijk, The Netherlands.

E-mail: `Tristan.vanLeeuwen@tudelft.nl`

Abstract. In seismic imaging, one tries to infer the medium properties of the subsurface from seismic reflection data. These data are the result of an active source experiment, where an explosive source and an array of receivers are placed at the surface. Due to the absence of low frequencies in the data, the corresponding inverse problem is strongly non-linear in the slowly varying component of the velocity. The least-squares misfit functional typically exhibits local minima and has a small basin of attraction. The usual approach of fitting the data in a least-squares sense by employing a gradient-based optimisation method will therefore most likely result in a wrong velocity model. In the geophysical community, this problem has long been recognised and alternative formulations of the inverse problem have been developed. We review several of these formulations and analyse the sensitivity to the error in the smooth velocity component. This analysis is carried out for laterally homogeneous velocities using an asymptotic solution of the wave equation. The analysis suggests that formulations which are geared towards fitting the phases of the data, rather than the amplitudes, have smooth corresponding misfit functionals with a large basin of attraction.

1. Introduction

In seismic imaging, one tries to infer the medium properties of the subsurface from seismic reflection data. These data are the result of an active source experiment, where an explosive source and an array of receivers are placed at or, in the case of a marine experiment, just below the surface. In the constant-density acoustic approximation, the wave propagation is governed by a scalar wave equation and the medium parameter to be recovered is the sound speed in the subsurface. The corresponding inverse problem can be formulated as a PDE-constrained least-squares (LS) problem. The sheer size of a typical problem in seismic exploration dictates that the resulting optimisation problem is solved using a gradient-based optimisation algorithm. This approach to seismic imaging

was pioneered by Tarantola (1984). However, it soon became apparent that the band-limited nature of the data causes local minima to exist in the least-squares objective functional. Hence, a very good initial guess of the subsurface velocity model is needed to converge to the global minimum.

Due to the absence of low frequencies in the source, the reflections in the data are caused by discontinuities in the velocity. This allows us to separate scales, and split the velocity into a smooth 'background' component and an oscillatory component, 'the reflectivity'. Ignoring multiple scattering, the data are now linear in the reflectivity. Under these assumptions, the inverse problem has become *separable*: the background velocity appears as a non-linear parameter, while the reflectivity appears as a linear parameter. A straightforward approach to solve such a separable LS problem, is to use a variable projection algorithm (Golub & Pereyra 1973). The basic idea is to resolve the reflectivity via the normal equations. In the literature this is often referred to as migration (Bednar 2005). Effectively, the data are projected onto the range of the modelling operator, and the LS difference between the data and their projection is minimised. This approach was applied to seismic imaging by Chavent et al. (1994). While this partly mitigates the nonlinearity of the inverse problem, the resulting misfit functional still has a small basin of attraction. This is because the range of the modelling operator changes very rapidly as the background velocity is perturbed. This causes the data and their projection for a perturbed velocity to be almost orthogonal.

To circumvent the problem of local minima and a small basin of attraction, a number of approaches exist that recast the optimisation problem as a tomographic problem. This is referred to as 'velocity analysis' in the literature (Claerbout 1984, Yilmaz 1987). We discern two distinct classes of such methods. The first exploits the redundancy in the data to introduce a redundant coordinate in the reflectivity estimate. For the correct velocity model, the reflectivity estimate should not depend on this redundant coordinate, and this serves as a criterion to obtain the correct background velocity. These methods have been well-established (Symes 1991, Chauris & Noble 2001, Mulder & ten Kroode 2002, Shen & Symes 2008) and a general theoretical framework exists that encompasses these methods. The main result here is that an exact *annihilator* can be constructed (Stolk & de Hoop 2006, de Hoop et al. 2006). This means that an exact criterion can be derived to measure the velocity error. It can be shown that the corresponding annihilators will indeed lead to misfit functionals that behave smoothly as a function of the velocity error (Stolk & Symes 2003). For an extensive overview we refer to (Symes 2008*b*) and references cited therein.

The second class of velocity analysis methods exploits the redundancy in the data directly. These methods are akin to the LS data-fitting approach, described above, except that they use a misfit criterion different from the LS error (Luo & Schuster 1989, van Leeuwen & Mulder 2008*a*). In this paper we review several velocity analysis methods in the context of laterally invariant velocities. First, we derive an asymptotic solution to the wave-equation for laterally invariant velocities and define the modelling operator. This type of modelling is known as 'convolutional' modelling and is well-established.

Nevertheless, we include an outline of the derivation for the sake of completeness. We present an example of seismic data modelled for a simple velocity model and compare it with the solution of the wave-equation modelled with finite differences. In section 4, we briefly discuss the linearised inverse problem. Several velocity analysis methods are discussed in sections 5 and 6. Although a quite general framework exists that encompasses some of these methods, we present only the very basics, providing an intuitive understanding of the general principles. Using stationary phase analysis, we examine the sensitivity to velocity errors. We include scans of the objective functionals, which show their typical behaviour. In section 7 we discuss the apparent relation between two particular methods. Finally, we present an example of several methods on synthetic data in section 8 and conclude the paper in section 9.

2. Notation and preliminaries

- $\mathbf{x}$ denotes a vector in $\mathbb{R}^3$, individual components are denoted by $x_{(i)}$. For variables in $\mathbb{R}$, italics will be used.

- The complex conjugate of a scalar as well as the complex conjugate transpose, or adjoint, of an operator will be denoted by $*$.

- $\langle \cdot, \cdot \rangle$ denotes the inner product. For vectors, $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_i x_{(i)}^* y_{(i)}$, for functions, $\langle f, g \rangle = \int \mathrm{d}x\, f^*(x)g(x)$.

- $|| \cdot ||$ denotes the L_2 norm, defined for vectors and functions in the usual way via the inner product $|| \cdot || = \sqrt{\langle \cdot, \cdot \rangle}$.

- [.] denotes functional dependence. For instance, $J[f] = ||f||$ and $C[f, g](x) = \int \mathrm{d}x'\, f(x')g(x - x')$.

- $\sim$ denotes an asymptotic relation. This indicates that we neglect smooth errors, as opposed to $\approx$ which indicates neglect of small errors.

An integral of the form

$$I = \int \mathrm{d}x\, f(x) \exp[\imath\omega\phi(x)],\tag{1}$$

can be approximated for large ω by

$$I \sim (1 + \imath \operatorname{sign}(\phi''(x_0)))\sqrt{\frac{\pi}{\omega|\phi''(x_0)|}} f(x_0) \exp[\imath\omega\phi(x_0)],\tag{2}$$

where x_0 is called the stationary point where $\phi'(x_0) = 0$. A derivation can be found in may textbooks on asymptotic analysis (Bleistein & Handelsman 1975, for example). We will deploy this stationary-phase approximation throughout the paper. Since we are mainly interested in the phase behaviour, we will neglect the amplitude factor $(1 \pm \imath)\sqrt{\pi/|\phi''|}$ most of the time.

3. Modelling

We use the so-called convolutional model with hyperbolic traveltimes to compute the seismic data. This type of modelling is valid for layered media at small source-receiver distances (offsets). Although this type of modelling is well-established, we present a derivation for the sake of completeness.

The seismic data sample the wavefield at some receiver position $\mathbf{x}_r$: $u(\omega, \mathbf{x}, \mathbf{x}_s)|_{\mathbf{x}=\mathbf{x}_r}$. Here, u is assumed to obey the scalar Helmholtz equation

$$L[c]u(\omega, \mathbf{x}, \mathbf{x}_s) = f(\omega)\delta(\mathbf{x} - \mathbf{x}_s), \tag{3}$$

where $L[c] = (\omega/c)^2 + \nabla^2$, $\mathbf{x}_s$ and $\mathbf{x}_r$ denote the source and receiver positions, $\mathbf{x}$ denotes the subsurface coordinate, $c(\mathbf{x})$ gives the soundspeed in the earth, and f is the source signature. It is common practise in seismic imaging to use a Born approximation and split the squared slowness into a smooth ('background velocity') and an oscillatory ('reflectivity') part: $c(\mathbf{x})^{-2} = c_0(\mathbf{x})^{-2} + r(\mathbf{x})$. The equations for the incident and scattered wavefield read:

$$L[c_0]u_0(\omega, \mathbf{x}, \mathbf{x}_s) = f(\omega)\delta(\mathbf{x} - \mathbf{x}_s), \tag{4}$$

$$L[c_0]u_1(\omega, \mathbf{x}, \mathbf{x}_s) = -\omega^2 u_0(\omega, \mathbf{x}, \mathbf{x}_s)r(\mathbf{x}). \tag{5}$$

In terms of the Green's function for the Helmholtz equation, the incident and scattered wavefields are given by

$$u_0(\omega, \mathbf{x}, \mathbf{x}_s) = f(\omega)G(\omega, \mathbf{x}, \mathbf{x}_s), \tag{6}$$

$$u_1(\omega, \mathbf{x}, \mathbf{x}_s) = -\omega^2 \int d\mathbf{x}' \, r(\mathbf{x}')u_0(\omega, \mathbf{x}', \mathbf{x}_s)G(\omega, \mathbf{x}, \mathbf{x}'). \tag{7}$$

For the Green's function we use a high-frequency (WKBJ) approximation

$$G(\omega, \mathbf{x}, \mathbf{x}') \sim A(\mathbf{x}, \mathbf{x}') \exp\left[\imath\omega T(\mathbf{x}, \mathbf{x}')\right], \tag{8}$$

where A is the amplitude factor which obeys a transport equation and T is the traveltime which obeys an eikonal equation:

$$\nabla T \cdot \nabla T = c(\mathbf{x})^{-2}, \tag{9}$$

$$\nabla \cdot \left(A^2 \nabla T\right) = 0. \tag{10}$$

Assuming that the data sample only the scattered field u_1, we define a linear operator that maps the reflectivity r to data p:

$$p(\omega, \mathbf{x}_s, \mathbf{x}_r) = F[c_0]r = u_1(\omega, \mathbf{x}_s, \mathbf{x}_r). \tag{11}$$

3.1. Laterally invariant velocity

To simplify the subsequent analysis we make the following assumptions

- the velocity is laterally invariant;
- there is one source at the origin and the receivers are positioned at the surface, i.e., $x_{r,(3)} = 0$;

- we are interested in relatively small source-receiver distances (offsets);

- we can neglect smoothly varying amplitude factors because we are interested in the singular behaviour of the data.

Under these assumptions, the data, $p(\omega, h) = u_1(\omega, \mathbf{x}_s, \mathbf{x}_r)$, depend only on half-offset $h = \frac{1}{2}||\mathbf{x}_r - \mathbf{x}_s||$ and frequency. Using a stationary-phase approximation to get rid of the integrals over the lateral coordinates and taking into account that $x_{r,(3)} = x_{s,(3)} = 0$, we can approximate eq. (11) by

$$p(\omega, h) = F[c_0]r \sim -\omega f(\omega) \int \mathrm{d}z \, r(z) \exp\left[\imath\omega\tau[c_0](z, h)\right], \tag{12}$$

where

$$\tau[c_0](z, h) = 2 \int_0^z \mathrm{d}\zeta \, c_0(\zeta)^{-1} + h^2 \left(\int_0^z \mathrm{d}\zeta \, c_0(\zeta)^{-1}\right)^{-1} m(z) \tag{13}$$

denotes the two-way traveltime and m is the effective squared slowness:

$$m(z) = \frac{\int_0^z \mathrm{d}\zeta \, c_0(\zeta)^{-1}}{\int_0^z \mathrm{d}\zeta \, c_0(\zeta)}. \tag{14}$$

This approximation to the two-way traveltime relates to the solution of the eikonal equation as $T(0, \mathbf{x}) = \frac{1}{2}\tau(x_3, \frac{1}{2}\sqrt{x_1^2 + x_2^2}) + \mathcal{O}(h^4)$. For a formal derivation we refer to (Dix 1955, Symes 1999). The quantity $1/\sqrt{m}$ is sometimes referred to as the root-mean-square (RMS) velocity. The traveltime $\tau(z, h)$ is called the normal moveout (NMO) traveltime.

Next, we parameterise depth by the vertical two-way traveltime $t_0(z) = 2 \int_0^z \mathrm{d}\zeta \, c_0(\zeta)^{-1}$ and define the convolutional modelling operator as

$$\tilde{F}[\tilde{m}]\tilde{r} = -\omega f(\omega) \int \mathrm{d}t_0 \, \tilde{r}(t_0) \exp\left[\imath\omega\tilde{\tau}[\tilde{m}](t_0, h)\right], \tag{15}$$

where $\tilde{r}(t_0) = r(z(t_0))$, etc., and $\tilde{\tau}[\tilde{m}](t_0, h) = t_0 + \frac{1}{2}h^2 t_0^{-1}\tilde{m}(t_0)$. Note that we neglect the Jacobian of the coordinate change $2/\tilde{c}(t_0)$.

We also introduce the inverse of the traveltime function $\tilde{\tau}_0(t, h)$ such that $\tilde{\tau}(\tilde{\tau}_0(t, h), h) = t$ and $\tilde{\tau}_0(\tilde{\tau}(t_0, h), h) = t_0$. For the inverse to exist globally, we need the traveltime to be strictly increasing as a function of t_0:

$$\tilde{s}(t_0, h) = \frac{\partial\tilde{\tau}}{\partial t_0}(t_0, h) > 0. \tag{16}$$

The factor $\tilde{s}$ also determines the amount of coordinate stretch. Since this stretch is unphysical, we will want to restrict the range of h such that $\tilde{s} \approx 1$. We have

$$\tilde{s}(t_0, h) = 1 - \frac{h^2}{2t_0^2}\left(t_0\frac{\mathrm{d}\tilde{m}}{\mathrm{d}t_0}(t_0) - \tilde{m}(t_0)\right). \tag{17}$$

Typically, we use offsets $h \lesssim 10^3 t_0$. In the remainder of the paper, we implicitly assume that an appropriate cut-off is applied whenever necessary, and neglect the factor $\tilde{s}$.

4. Linearised inversion

Given the correct background velocity, the reflectivity can be retrieved from the data by solving the normal equations:

$$Nr = F^*p, \tag{18}$$

where $N = F^*F$. The adjoint of the modelling operator can be directly derived from eq. (15), but it is insightful to derive it directly from eq. (11), using the adjoint-state technique. This yields:

$$F^*[c_0]p = -\int d\mathbf{x}_r \int d\mathbf{x}_s \int d\omega\, \omega^2 u_0^*(\omega, \mathbf{x}_s, \mathbf{x}_r) v_0(\omega, \mathbf{x}_s, \mathbf{x}_r), \tag{19}$$

where the *source wavefield*, u_0, satisfies eq. 4 and the *receiver wavefield*, v_0, satisfies

$$L^*[c_0]v_0 = \int d\mathbf{x}\, p(\omega, \mathbf{x}_r, \mathbf{x}_s)\delta(\mathbf{x} - \mathbf{x}_r). \tag{20}$$

Under the assumptions mentioned above, the source and receiver wavefields are given by

$$\tilde{u}_0[\tilde{m}](\omega, t_0, h) \sim f(\omega) \exp\!\left[\imath\omega \tfrac{1}{2}\tilde{\tau}[\tilde{m}](t_0, h)\right], \tag{21}$$

$$\tilde{v}_0[\tilde{m}](\omega, t_0, h) \sim p(\omega, h) \exp\!\left[-\imath\omega \tfrac{1}{2}\tilde{\tau}[\tilde{m}](t_0, h)\right]. \tag{22}$$

We arrive at

$$\tilde{F}^*p = -\int dh \int d\omega\, \omega f^*(\omega) p(\omega, h) \exp\!\left[-\imath\omega\tilde{\tau}[\tilde{m}](t_0, h)\right]. \tag{23}$$

This operation is called 'NMO-correction', since it removes the 'moveout' effect in the data.

We study the normal operator, given by

$$\tilde{N}[\tilde{m}]\tilde{r} = \int d\omega \int dt_0' \int dh\, \omega^2 |f(\omega)|^2 \tilde{r}(t_0') \exp\!\left[\imath\omega\left(\tilde{\tau}[\tilde{m}](t_0', h) - \tilde{\tau}[\tilde{m}](t_0, h)\right)\right]$$

$$\sim \int dt_0' \int d\omega\, \omega^2 |f(\omega)|^2 \tilde{r}(t_0') \exp\!\left[\imath\omega(t_0' - t_0)\right]. \tag{24}$$

In case the wavelet is a delta function, we have $\tilde{N}\tilde{r} \sim \frac{d^2}{dt_0^2}\tilde{r}$. This result is consistent with a more general analysis of the normal operator (Beylkin 1985), which states that the normal operator is asymptotically diagonal in phase space. Throughout the rest of the paper we assume that we have applied appropriate filtering to account for the ω^2 factor.

Example

Fig. 1 shows a simple velocity model together with the corresponding effective velocity and reflectivity. Data modelled with a time-domain finite-difference code (Mulder & Plessix 2002) are depicted, as are the data resulting from the convolutional model operator (cf. eq. (15)). For comparison, two traces at $h = 0$ m and $h = 2000$ m are depicted, showing that the traveltimes of the events in the convolutional model data are accurate. The amplitudes are not correctly modelled, especially for large offsets.

5. Velocity analysis in the image domain

How do we get a handle on the velocity error? The linearised inverse problem mentioned above can be solved for any given background velocity, but the result will only be useful if the background velocity is correct. First, we observe that the source and receiver wavefields that constitute the reflectivity estimate will constructively interfere at the location of a reflector. Thus, at a reflector position, t_0, we have

$$u_0(\omega, t_0, h) \sim v_0(\omega, t_0, h), \tag{25}$$

(cf. eq. (21,22)). If the background velocity is incorrect, however, the wavefields will *not* match at a reflector. In image-domain velocity analysis this is exploited in various ways. We will now discuss several of them.

We will denote the observed data, and the corresponding 'true' reflectivity and background velocity with a bar, $\bar{p} = F[\bar{m}]\bar{r}$. Also, $\bar{\tau} = \tau[\bar{m}]$ and $\tau = \tau[m]$. The functional argument will be left out whenever there is no confusion possible.

5.1. Stackpower

The stackpower criterion relies on maximising the inner product of the wavefields, which will attain a maximum if the wavefields are identical. The corresponding functional is formulated as (Toldi 1989)

$$J = ||r||^2, \quad r(t_0) = \int dh \int d\omega \, u_0^*(\omega, t_0, h) v_0(\omega, t_0, h). \tag{26}$$

Using the expressions derived above we can write

$$r(t_0) =$$
$$\int dh \int d\omega \int dt_0' \, \bar{r}(t_0') |f(\omega)|^2 \exp\left[\iota\omega\left((t_0' - t_0) + \tfrac{1}{2}h^2\left(\frac{\bar{m}(t_0')}{t_0'} - \frac{m(t_0)}{t_0}\right)\right)\right]. \tag{27}$$

The phase has a stationary point at $h = 0$, implying that singularities in $\bar{r}$ are mapped onto themselves, or $r(t_0) \sim \bar{r}(t_0)$. The velocity error is to be found in the corresponding amplitude factor.

As an example, we present a scan of the functional. We use the data from the example presented before and evaluate the functional for different background velocities, which we represent on a 2-node cubic spline grid $\{t_0, t_1\} = \{1, 1.5\}$. Thus, we parametrise the background velocity with two parameters: $\{m_1, m_2\}$. We then evaluate the functional at $\{\bar{m}_1 + \delta m_1, \bar{m}_2 + \delta m_2\}$. As a source function we assume a point source with isotropic radiation pattern and a Ricker wavelet for the signature. In the frequency domain, the wavelet is given by (Kanasewich 1985):

$$f_\epsilon(\omega) = \omega^2 \exp[-\epsilon\omega^2]. \tag{28}$$

The peak-frequency of this wavelet is $1/(2\pi\sqrt{\epsilon})$. This wavelet is commonly used in the geophysical community.

Figure 2 displays the stackpower for two different peak-frequencies, 5 Hz and 30 Hz, as a function of the slowness perturbations. The stackpower is very sensitive to the frequency

content of the data. As the frequency increases, the basin of attraction becomes smaller and local maxima start to develop. As seismic data are in general dominated by high frequencies, this approach is less suitable for velocity inversion.

5.2. Differential semblance

An alternative way to measure the difference between the wavefields is the so-called differential semblance criterion. The idea is that we can obtain a reflectivity estimate from just a single trace and we can use the offset as the redundant coordinate. The reflectivity estimate effectively becomes a function of both depth and offset, denoted as $r(t_0, h)$. Note that the conventional reflectivity estimate is obtained by integrating $r(t_0, h)$ over offset. For the correct velocity model, the reflectivity estimate should be independent of offset. The velocity error can now be measured by differentiation w.r.t. offset, hence the name *differential semblance*. Here, the offset-differentiation is called an *annihilator*, since it annihilates the reflectivity estimate for the correct velocity model, i.e., $\partial_h r[\bar{m}] = 0$. The corresponding error functional is given by

$$J = ||\partial_h r||^2. \tag{29}$$

This approach was introduced by Symes (1991) and has been elaborated upon by others (Chauris & Noble 2001, Mulder & ten Kroode 2002, Li & Symes 2007).
The reflectivity estimate is given by

$$\begin{aligned} r(t_0, h) &= \int \mathrm{d}\omega\, u_0^*(\omega, t_0, h) v_0(\omega, t_0, h) \\ &= \int \mathrm{d}\omega \int \mathrm{d}t_0'\, \bar{r}(t_0') |f(\omega)|^2 \exp\left[\imath\omega\left((t_0' - t_0) + \tfrac{1}{2}h^2\left(\tfrac{\bar{m}(t_0')}{t_0'} - \tfrac{m(t_0)}{t_0}\right)\right)\right] \\ &\sim \bar{r}\left(t_0 + \frac{h^2}{2t_0}(\bar{m} - m)\right). \end{aligned} \tag{30}$$

This shows that the reflectivity estimate is independent of h for the correct velocity. Fig. 3 shows an example of $r(t_0, h)$ for different velocities.
Fig. 4 depicts the behaviour of the DSO functional as a function of the velocity error for different peak frequencies. It shows that the DSO functional has a large basin of attraction and does not depend on the frequency content of the data. In fact, Symes (1999) proved that, under similar assumptions as introduced here, the DSO functional is asymptotically stable. Numerical evidence exists that the same holds under more general conditions (Chauris & Noble 2001, Mulder & ten Kroode 2002, Li & Symes 2007). Finally, an interesting observation can be made by considering the spatial Fourier transform of $r(t_0, h)$, denoted by $\hat{r}(t_0, k)$. Whereas the events will flatten in r, they will focus at $k = 0$ in $\hat{r}$. It is readily verified that $\int \mathrm{d}h\, r(t_0, h) = \hat{r}(t_0, 0)$. This indicates that, in the stackpower approach, we are actually throwing away valuable information by integrating over offset.

5.3. Correlation

A third way to measure the difference between the wavefields comes from the observation that a correlation of two identical signals will have a large contribution at zero lag. If there is a pertinent shift between the signals, the correlation will reveal this by having a large contribution at some non-zero lag. Introducing a temporal and spatial shift, we write the reflectivity estimate as

$$r(t_0, \Delta t, \Delta h) = \int dh \int d\omega \, u_0^*(\omega, t_0, h) v_0(\omega, t_0, h + \Delta h) \exp[-\imath \omega \Delta t]. \qquad (31)$$

We can introduce the scattering angle as redundant coordinate by taking the angle-transform of $r(r_0, \Delta t, \Delta h)$. The angle transform is defined as (Stolk & de Hoop 2006):

$$r(t_0, \theta) = \int d\Delta h \, r(t_0, \theta \Delta h, \Delta h). \qquad (32)$$

For small Δh this can be re-written as

$$r(t_0, \theta) \sim \bar{r}\left(t_0 + \frac{\theta^2}{2m^2 t_0}(\bar{m} - m)\right). \qquad (33)$$

This may be seen as a generalisation of the differential semblance method, discussed before. de Hoop et al. (2006) have proven rigorously that this approach is valid for laterally varying media as well.

Next we discuss the use of Δh and Δt separately as redundant coordinate.

5.3.1. Spatial shift The use of the spatial shift for velocity analysis has been studied by many authors (MacKay & Abma 1992, Rickett & Sava 2002, Sava & Fomel 2006, Shen & Symes 2008). Intuitively, the focusing may be understood by observing that the shift that is introduced actually represents an instantaneous displacement of the wavefield, or 'action at a distance'. This is unphysical and the correct velocity model does not allow for this, making the wavefields meet each other at zero shift. The amount of focus in the reflectivity can be now be measured by a multiplication by Δh, which serves as the annihilator. The misfit functional is defined accordingly as (Shen & Symes 2008):

$$J = ||W \cdot r||^2, \quad W(\Delta h) = \Delta h. \qquad (34)$$

The reflectivity estimate is given by

$$r(t_0, \Delta h) = $$
$$\int dh \int d\omega \int dt_0' \, \bar{r}(t_0') |f(\omega)|^2 \exp\left[\imath \omega \left(\bar{\tau}(t_0', h) - \tau(t_0, h + \Delta h)\right)\right]. \qquad (35)$$

We examine the contribution to r of a singularity in $\bar{r}$ by stationary phase analysis. A stationary point h_0 satisfies

$$\frac{\partial \bar{\tau}}{\partial h}(t_0', h_0) - \frac{\partial \tau}{\partial h}(t_0, h_0 + \Delta h) = 0, \qquad (36)$$

which yields

$$h_0 = \left(\frac{\bar{m}(t_0')t_0}{m(t_0)t_0'} - 1\right)^{-1} \Delta h. \qquad (37)$$

The phase at the stationary point is given by

$$\bar{\tau}(t_0', h_0) - \tau(t_0, h_0 + \Delta h) = (t_0' - t_0) - \frac{\Delta h^2}{2}\left(\frac{\bar{m}(t_0')m(t_0)}{\bar{m}(t_0')t_0 - m(t_0)t_0'}\right). \quad (38)$$

Note that the singularity that occurs at $m = \bar{m}$ is not a problem, since $\Delta h \to 0$ as $m \to \bar{m}$ as can be seen from eq. (37). With this we can express the reflectivity estimate as

$$r(t_0, \Delta h) \sim \bar{r}\left(t_0 + \frac{\Delta h^2}{2t_0}\left(\frac{\bar{m}m}{\bar{m} - m}\right)\right). \quad (39)$$

This shows how a singularity in $\bar{r}$ is spread out along a certain trajectory in the $(t_0, \Delta h)$ space. For the correct velocity model, this trajectory collapses to a point at $\Delta h = 0$ (cf. eq. (37)) and then $r(t_0, \Delta h) \sim \bar{r}(t_0)\delta(\Delta h)$. Fig. 5 displays an example. Fig. 6 shows the behaviour of the functional as function of velocity error for different peak frequencies. Like the DSO functional, this functional has a large basin of attraction and is insensitive to the frequency content of the data.

Note that, analogously to the earlier observation, the conventional reflectivity estimate is obtained by evaluating $r(t_0, \Delta h)$ at $\Delta h = 0$.

5.3.2. Temporal shift The correlation with a temporal shift also exhibits a focusing behaviour, though it is quite different from the spatial shift discussed previously. The reflectivity estimate is given by

$$r(t_0, \Delta t) =$$
$$\int \mathrm{d}h \int \mathrm{d}\omega \int \mathrm{d}t_0' \, \bar{r}(t_0')|f(\omega)|^2 \exp\left[\imath\omega\left(\bar{\tau}(t_0', h) - \tau(t_0, h) + \Delta t\right)\right]. \quad (40)$$

Again, we calculate the stationary point h_0 from

$$\frac{\partial\bar{\tau}}{\partial h}(t_0', h_0) - \frac{\partial\tau}{\partial h}(t_0, h_0) = 0, \quad (41)$$

yielding $h_0 = 0$. The phase of the stationary point is now given by

$$\bar{\tau}(t_0', 0) - \tau(t_0, 0) + \Delta t = t_0 - t_0' + \Delta t. \quad (42)$$

So, singularities in $\bar{r}$ are spread out along straight lines, independent of the velocity error. The velocity dependence is to be found in the corresponding amplitude factor which depends on the second derivative of the phase function:

$$\frac{\partial^2\bar{\tau}}{\partial h^2}(t_0', 0) - \frac{\partial^2\tau}{\partial h^2}(t_0, 0) = \frac{\bar{m}(t_0')}{t_0'} - \frac{m(t_0)}{t_0}. \quad (43)$$

With this, we may express the reflectivity estimate as

$$r(t_0, \Delta t) \sim \sqrt{\left|\frac{(t_0 + \Delta t)t_0}{t_0\bar{m}(t_0 + \Delta t) - (t_0 + \Delta t)m(t_0)}\right|}\,\bar{r}(t_0 + \Delta t). \quad (44)$$

An example is shown in Fig. 7.

The location of the peak in amplitude at $\Delta t = (\bar{m}/m - 1)t_0$ can be used to update the

velocity model by determining the location of the peaks in the image automatically. A similar approach is described by **?**. Although in this case a multiplication by Δt does *not* annihilate the reflectivity estimate, we define the functional analogously to eq. (34). The behaviour of this functional is depicted in Fig. 8. This functional has the minimum in the wrong place. Although the amplitude peaks in the image are located at zero shift for the correct velocity model, the functional does not capture this.

6. Data domain velocity analysis

A more intuitive approach to obtain the velocity model is to treat it as a non-linear least-squares problem. That is, we try to find $\{m, r\}$ that minimise

$$J[m, r] = ||F[m]r - \bar{p}||^2. \tag{45}$$

This approach was pioneered by Tarantola & Valette (1982) and elaborated on by many others. Exploiting the fact that the problem is *separable*, we can eliminate r via the normal equations:

$$r = N^{-1}F^*\bar{p}. \tag{46}$$

Substituting this into eq. (45) we obtain a modified functional:

$$\tilde{J}[m] = ||(\Pi[m] - I)\bar{p}||^2, \tag{47}$$

where

$$\Pi[m] = F[m]N[m]^{-1}F^*[m], \tag{48}$$

is an orthogonal projector onto the range of F. This approach is know as the variable projection approach (Golub & Pereyra 1973). Note that $(\Pi - I)$ is an annihilator acting on the data, i.e., $(\Pi[\bar{m}] - I)\bar{p} = 0$.
We also note that the zero-offset trace of the data is invariant under the projection, that is

$$\bar{p}(t, 0) \sim (\Pi[m]\bar{p})(t, 0). \tag{49}$$

This follows immediately from the earlier observation that $F^*\bar{p} \sim \bar{r}(t_0)$.
As with the image-domain approach, we now have two wavefields, $\bar{p}$ and $\Pi\bar{p}$, that will be identical for the correct velocity model. Next, we discuss several methods of comparing the two wavefields.

6.1. Least-squares

The obvious way to match the wavefields is by minimising the modified least-squares functional. This approach is also known as *Migration Based Traveltime Tomography* (MBTT) and was first proposed by Clément et al. (2001). The functional may be written as

$$\tilde{J} = ||\bar{p}||^2 - \langle\Pi\bar{p}, \bar{p}\rangle = ||\bar{p}||^2 - \langle N^{-1}F^*\bar{p}, F^*\bar{p}\rangle \sim ||\bar{p}||^2 - ||F^*\bar{p}||^2. \tag{50}$$

This reveals that this approach is akin to the stackpower approach; the wavefields are matched by maximising the inner product. Fig. 9 shows the behaviour of the functional and confirms the assertion. This similarity also holds under more general conditions (Stolk & Symes 2003, Symes 2008a).

6.2. Correlation

Here, we use a correlation of the data and their projection to judge the similarity. As with the correlation of the wavefields u_0 and v_0 discussed before, the correlation of the data and their projection will have a large contribution at zero shift if the velocity is correct. In an earlier paper (van Leeuwen & Mulder 2008b), we successfully exploited this to obtain a velocity estimate. We consider a spatial and temporal shift.

6.2.1. Spatial shift Given data $\bar{p}$ and $p = \Pi\bar{p}$, the spatial correlation is given by

$$C_h(t, \Delta h) = \int \mathrm{d}h\, p(t, h)\bar{p}(t, h + \Delta h). \tag{51}$$

Since the zero-offset trace of the data is invariant under the projection we may write this as

$$C_h(t, \Delta h) \sim \int \mathrm{d}h\, \bar{r}(\tau_0(t, h))\bar{r}(\bar{\tau}_0(t, h + \Delta h)), \tag{52}$$

We expect a significant contribution to C_h if $\tau_0(t, h) = \bar{\tau}_0(t, h + \Delta h)$.
This relation allows us to investigate different events that turn up in the correlation.
For the two-reflector example $\bar{r}(t_0) = \delta(t_0 - T_1) + \delta(t_0 - T_2)$, the contributions to C_h are given by

$$\Delta h = \pm \left(T_i \sqrt{\frac{t - T_i}{m(T_i)}} \pm T_j \sqrt{\frac{t - T_j}{\bar{m}(T_j)}} \right), \quad i, j \in \{1, 2\} \tag{53}$$

The contributions for $i = j$ can be interpreted as the result of the interaction of a reflector with itself. It shows that this interaction cause two distinct events in the correlation; one that will focus as $\bar{m} \to m$ and one that will not. Fig. 11 displays an example. The contributions for $i \neq j$ result from interaction between the different reflectors. This crosstalk is not present in Fig. 11 because it manifests itself at larger shifts. For a more detailed discussion of the crosstalk, we refer to section 8 and our earlier paper (van Leeuwen & Mulder 2008a).
Unfortunately, because of the non-focusing events, we cannot formulate an annihilator. Instead, we adopt a heuristic approach, measuring the amount of energy concentrated around zero shift with the functional:

$$J[m] = \frac{||W_h \cdot C_h||^2}{||C_h||^2}, \quad W_\sigma(\Delta h) = \exp[-\sigma \Delta h^2] \tag{54}$$

Here, σ determines the width of the weighting function. Note that this functional is to be maximised. Fig. 12 depicts the behaviour of this functional. The functional is smooth as a functional of the velocity error and has a large basin of attraction. Also, the behaviour does not change significantly with the frequency.

6.2.2. Temporal shift We also consider the temporal correlation given by

$$C_t(\Delta t, h) = \int \mathrm{d}t \, p(t, h) \bar{p}(t + \Delta t, h), \tag{55}$$

which may be written as

$$C_t(\Delta t, h) = \int \mathrm{d}t \, \bar{r}(\tau_0(t, h)) \bar{r}(\bar{\tau}_0(t + \Delta t, h)). \tag{56}$$

Again, we expect a contribution when $\tau_0(t, h) = \bar{\tau}_0(t + \Delta t, h)$. For our two-reflector model, $\bar{r}(t_0) = \delta(t_0 - T_1) + \delta(t_0 - T_2)$, this yields

$$\Delta t = (T_i - T_j) + \frac{h^2}{2} \left(\frac{\bar{m}(T_i)}{T_i} - \frac{m(T_j)}{T_j} \right), \quad i, j \in \{1, 2\}. \tag{57}$$

So, for $i = j$ each reflector contributes one event to the correlation. This event will focus for the correct velocity model. As in the spatial correlation, there is crosstalk present. Fig. 10 depicts an example. Here, the crosstalk is also visible.

Again, we cannot formulate an annihilator for the temporal correlation. So, we adopt the heuristic approach, described above and define the functional analogous to eq. (54). Fig. 13 depicts its behaviour. This functional has a large basin of attraction and its behaviour does not change with frequency.

7. Focusing in the image and data domain

In this section, we will study the focusing of the spatial correlation in the image and data-domain more closely. We have seen that the focusing of the extended reflectivity $r(t_0, \Delta h)$ is described by stationary points of the phase function. There is a contribution to this image when the derivatives of the traveltimes are equal:

$$\frac{\partial \bar{\tau}}{\partial h}(t_0, h) = \frac{\partial \tau}{\partial h}(t_0, h + \Delta h), \tag{58}$$

which yields (cf. eq. (37))

$$\Delta h = - \left(1 + \sqrt{\frac{\bar{m}}{m}} \right) \left(1 - \sqrt{\frac{\bar{m}}{m}} \right) h. \tag{59}$$

There is a contribution to the correlation $C_h(t, \Delta h)$ when $\bar{\tau}_0(t, h) = \tau_0(t, h + \Delta h)$. For a fixed t_0 this is equivalent to

$$\bar{\tau}(t_0, h) = \tau(t_0, h + \Delta h), \tag{60}$$

which yields

$$\Delta h = - \left(1 \pm \sqrt{\frac{\bar{m}}{m}} \right) h. \tag{61}$$

Fig. 14 displays an example of the phase function $\bar{\tau}(t_0, h) - \tau(t_0, h + \Delta h)$ for a fixed t_0 and various relative velocity perturbations. This shows how the contributions to $r(t_0, \Delta h)$ and $C_h(t, \Delta h)$ gradually shift towards $\Delta h = 0$ as the velocity error decreases. Effectively, by weighting with Δh and taking the norm, the slope of the line is measured.

In the DSO approach, the slope of the line is measured by taking the h-derivative at $\Delta h = 0$. The non-focusing artifact that appears in the data correlation is also explained by these pictures; one of the solid lines hardly changes slope and is due to contributions of small offsets. This suggests that the artifact may be suppressed by removing the small offsets from the data.

8. Synthetic-data example

We applied the differential semblance method and the spatial data-correlation, described above to synthetic data. The data where produced by solving the linearised wave equation for a 1D-slice of the Marmousi velocity model, a well-known benchmark model (Versteeg 1993), with a time-domain finite-difference code. The RMS velocity was represented using cubic splines on an equidistant grid of 7 nodes. For the optimisation, we used a BFGS (Broyden-Fletcher-Shanno-Goldfarb) algorithm. The gradients of the functionals were calculated using the adjoint-state technique (Plessix 2006).

The functional for the data-domain is slightly different from the one defined earlier:

$$J = ||S^{-1} \cdot W_\sigma \cdot C_h||^2, \quad S(t) = \int \mathrm{d}\Delta h \, (C_h(t, \Delta h))^2. \tag{62}$$

The scaling factor S normalises the amplitudes of the events.

Since the zero-offset trace of the data is invariant under the projection, we may write the projection of the data as

$$\Pi[m]\bar{p} = F[m](\bar{p}(t_0, 0)). \tag{63}$$

Fig. 15 depicts the observed data as well as the true, initial, and final velocity model that resulted from optimising the functional defined in eq. (62). The velocity model is almost perfectly recovered. A perfect fit is not to be expected, since we are representing the velocity model with splines. Fig. 17 shows the spatial correlation of the data and their projection for the initial and final velocity, illustrating how the focusing is improved.

9. Conclusions

The seismic inverse problem, where we want to recover the soundspeed in the subsurface, can be cast as a separable least-squares problem. The resulting optimisation problem is linear in the oscillatory component of the velocity model and non-linear in the smooth background velocity. Typically, local minima exist and this requires a reformulation of the non-linear optimisation problem. We have reviewed several such alternative formulations, concentrating mainly on an intuitive understanding of the underlying principles of the different methods.

The methods discussed fall in two categories; methods that rely on *amplitude* fitting and methods that rely on *phase* fitting. The least-squares and stack-power approach, discussed in sections 5.1 and 6.1, are both relying on amplitude information to obtain the correct velocity model. These methods turn out to be very sensitive to the frequency

content of the data, as can be seen in figs. 2 and 9. For higher peak frequencies the basin of attraction of the functionals becomes smaller and local minima start to develop. Also, one can imagine that field data will not be completely model-consistent. That is, even for the correct background velocity and reflectivity, the modelled amplitudes may not completely match the observed amplitudes. In this case, fitting the amplitudes may lead to a wrong background velocity. The other methods discussed rely on fitting the phases of the data. The phases, or traveltimes, of the data are expected to carry more information about the background velocity than the amplitudes do. Also, this approach is far less sensitive to the frequency content of the data. This is revealed in figs 4,6,12, and 13. The basin of attraction is large, independent of the peak frequency.

Additional analysis of two methods that use a spatial correlation reveals how the correlation will *gradually* focus as the velocity error decreases. Of course, one expects the traveltimes of the data to change gradually with velocity, and this translates to the focusing of the correlation, which basically measures a traveltime difference.

Although the analysis here is limited to layered media, the general principles of amplitude and phase fitting are still valid in laterally varying media. Therefore, we expect that the general conclusions will hold. In fact, for some of the methods abundant numerical evidence exists that is consistent with this analysis (Chauris & Noble 2001, Mulder & ten Kroode 2002, Shen & Symes 2008). Based on the analysis it is impossible to conclude which would be the preferred method, although it is clear that a reformulation in terms of phase fitting is beneficial for these kind of band-limited inverse problems. The computational complexity would also very much depend on the implementation and the modelling kernel used to solve the wave equation.

References

Bednar J B 2005 *Geophysics* **70**(3), 3MJ–20MJ.

Beylkin G 1985 *Journal of Mathematical Physics* **26**(1), 99–108.

Bleistein N & Handelsman R 1975 *Asymptotic Expansions of Integrals* Dover New York.

Chauris H & Noble M S 2001 *Geophysical Journal International* **144**(1), 14–26.

Chavent G, Clément F & Gómez S 1994 *SEG Technical program expanded abstracts* **13**, 1179–1182.

Claerbout J F 1984 *Imaging the earth's interior* Blackwell Scientific Publishers.

Clément F, Chavent G & Gómez S 2001 *Geophysics* **66**(3), 845–860.

de Hoop M V, van der Hilst R D & Shen P 2006 *Geophysical Journal Int.* **167**, 1332–1352.

Dix C H 1955 *Geophysics* **20**(1), 68–86.

Faye J P & Jeannot J P 1986 *SEG Technical Program Expanded Abstracts* **5**, 438–440.

Golub G & Pereyra V 1973 *SIAM J. Numer. Anal.* **10**(2), 413–432.

Higginbotham J H, Brown M P & Clapp R G 2008 *SEG Technical Program Expanded Abstracts* **27**, 3083–3087.

Kanasewich E 1985 *Time Sequence Analysis in Geophysics* University of Alberta Press.

Li J & Symes W 2007 *Geophysics* **72**(6), U75–U88.

Luo Y & Schuster G T 1989 *Geophysics* **56**(5), 645–653.

MacKay S & Abma R 1992 *Geophysics* **57**(12), 1608–1622.

Mulder W A & ten Kroode A P E 2002 *Geophysics* **67**(4), 1184–1191.

Mulder W & Plessix R E 2002 *EAGE expanded abstracts* pp. E015–E015.

Plessix R E 2006 *Geophysical Journal International* **167**(2), 495–503.

Rickett J E & Sava P C 2002 *Geophysics* **67**(3), 883–889.

Sava P C & Fomel S 2006 *Geophysics* **71**(6), S209–S217.

Shen P & Symes W W 2008 *Geophysics* **73**(5), VE49–VE59.

Stolk C C & Symes W W 2003 *Inverse Problems* **19**, 73–89.

Stolk C & de Hoop M 2006 *Wave Motion* **43**, 579–598.

Symes W 2008*a Geophysical Prospecting* **56**(6), 765–790.

Symes W W 1991 *SIAM Journal on Mathematical Analysis* **22**(3), 680–716.

Symes W W 1999 All stationary points of differential are global minimizers: Layered acoustics Technical Report 100 Stanford Exploration Project.
URL: *http://sepwww.stanford.edu/public/docs/sep100/bill3.pdf*

Symes W W 2008*b Geophysics* **73**(2), R23–R35.

Tarantola A 1984 *Geophysics* **49**(8), 1259–1266.

Tarantola A & Valette A 1982 *Reviews of Geophysics and Space Physics* **20**(2), 129–232.

Toldi J L 1989 *Geophysics* **54**(2), 191–199.

van Leeuwen T & Mulder W 2008*a Geophysical Prospecting* **56**(6), 791–803.

van Leeuwen T & Mulder W 2008*b EAGE Extended Abstracts* .

Versteeg R 1993 *Geophysics* **58**(6), 873–882.

Yilmaz O 1987 *Seismic data processing* Atlantic Books.

Acknowledgements
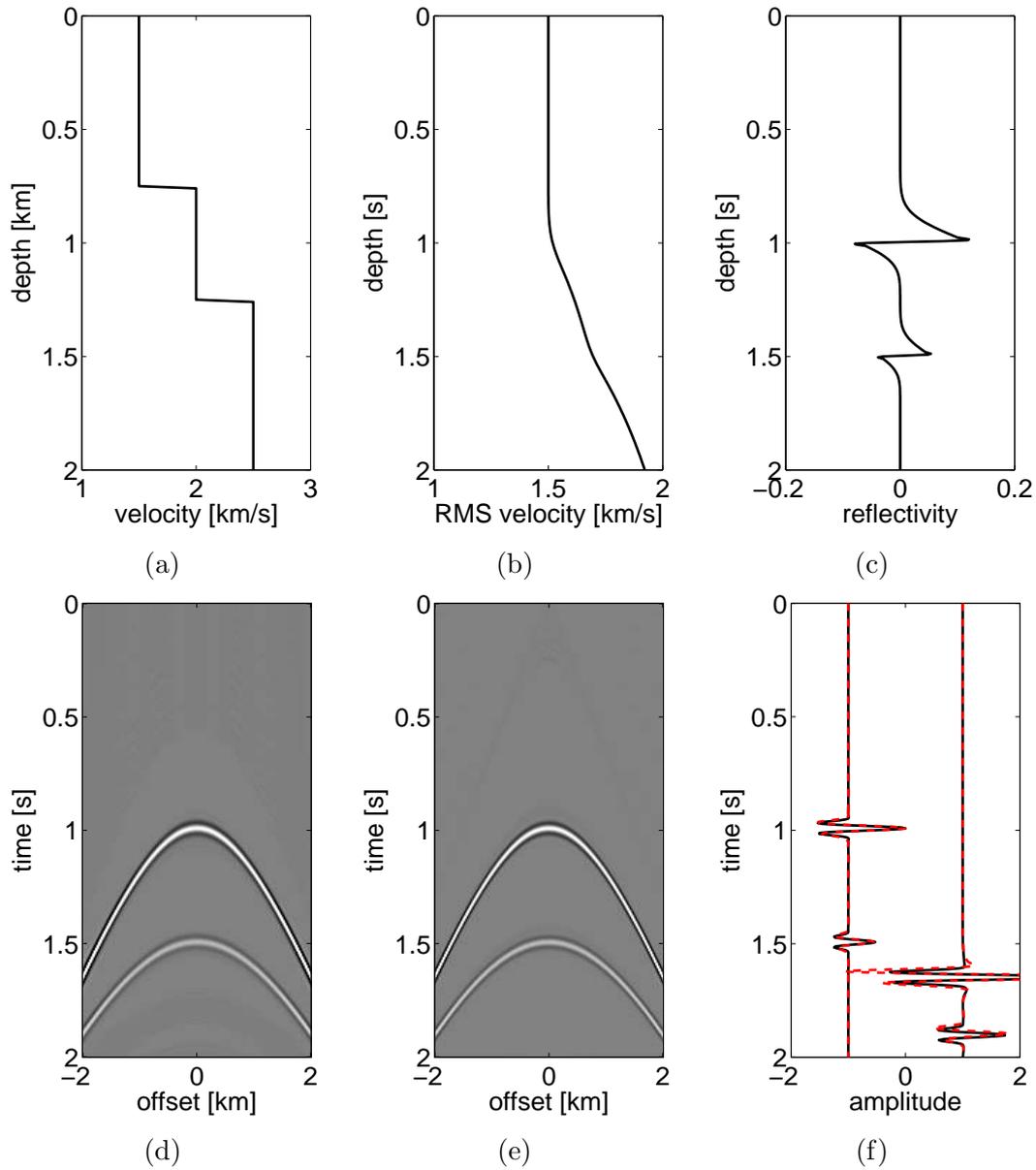
Figure 1. Example of data modelled with the convolutional model. (a) Velocity model $c(z)$, (b) RMS velocity and (c) reflectivity $\tilde{r}(t_0)$. (d) Born data for velocity model (a) modelled with finite differences, (e) data for the convolutional model and (f) traces at $h = 0$ km (left) and $h = 2$ km (right) of the finite-difference and convolutional-model data (dash and solid lines respectively).
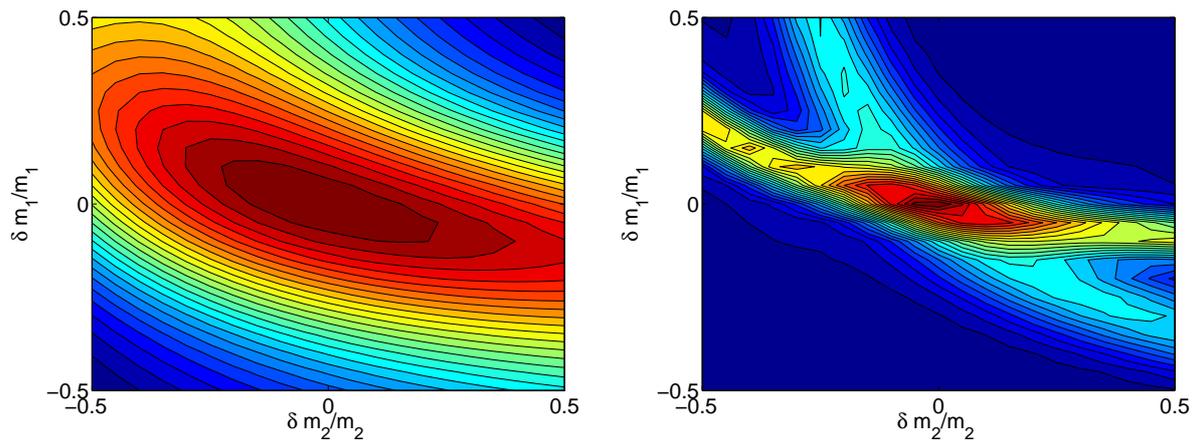
Figure 2. Stackpower functional (eq. (26)) as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. As the peak frequency of the source increases, the behaviour of the functional becomes more singular and local minima start to develop.
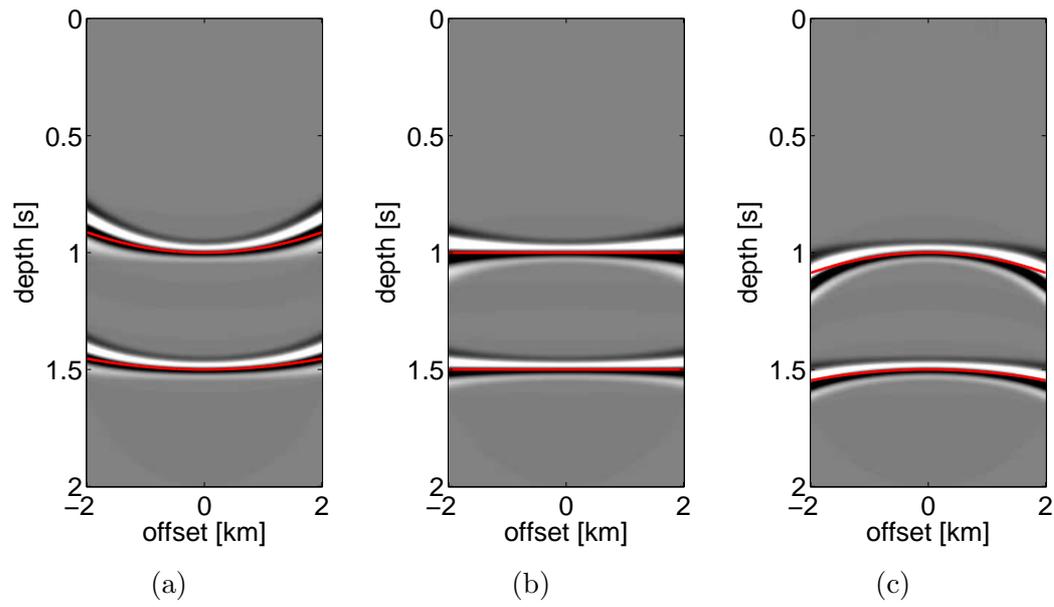
Figure 3. Reflectivity estimate $r(t_0, h)$, used for differential semblance, for a velocity that is 10% too low (a), correct (b) and 10% too high (c). The trajectories along which the singularities in the reflectivity are smeared out, cf. eq. (30), are indicated in red.
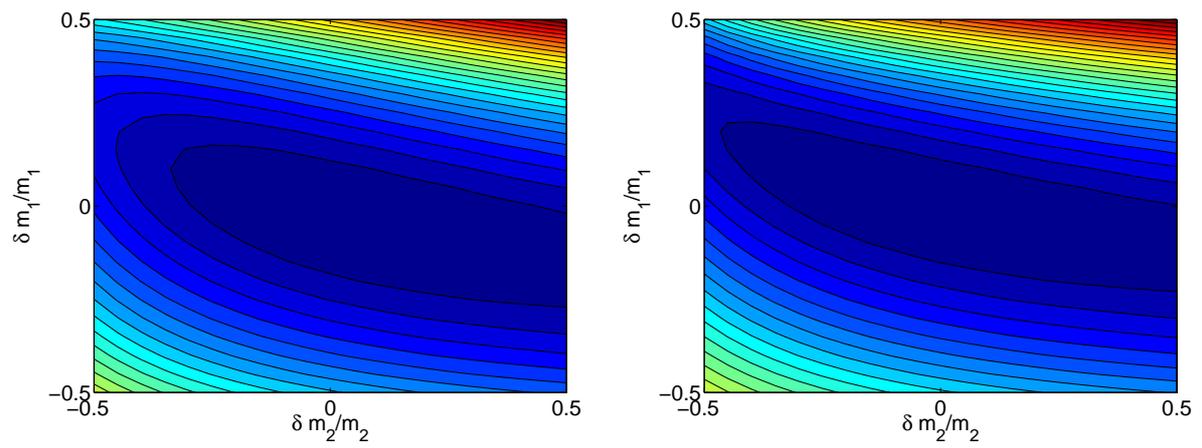
Figure 4. Differential semblance functional (eq. (29)) as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. This functional does not depend on the frequency content of the data.
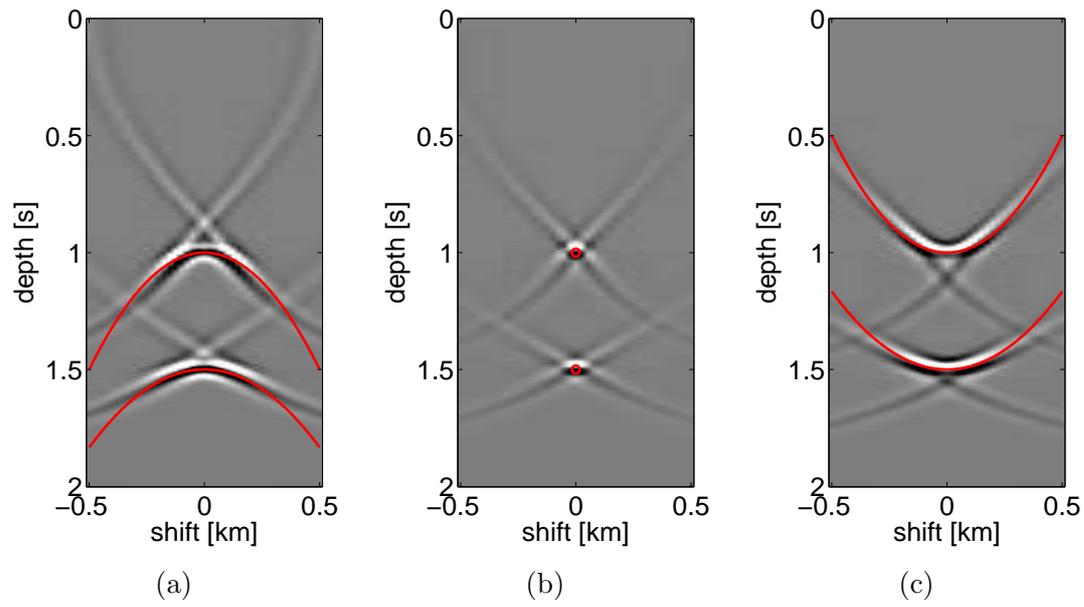
Figure 5. Reflectivity estimate $r(t_0, \Delta h)$, used in the spatial correlation, for a velocity that is 10% too low (a), correct (b) and 10% too high (c). The trajectories along which the singularities in the reflectivity are smeared out, cf. eq. (39), are indicated in red.
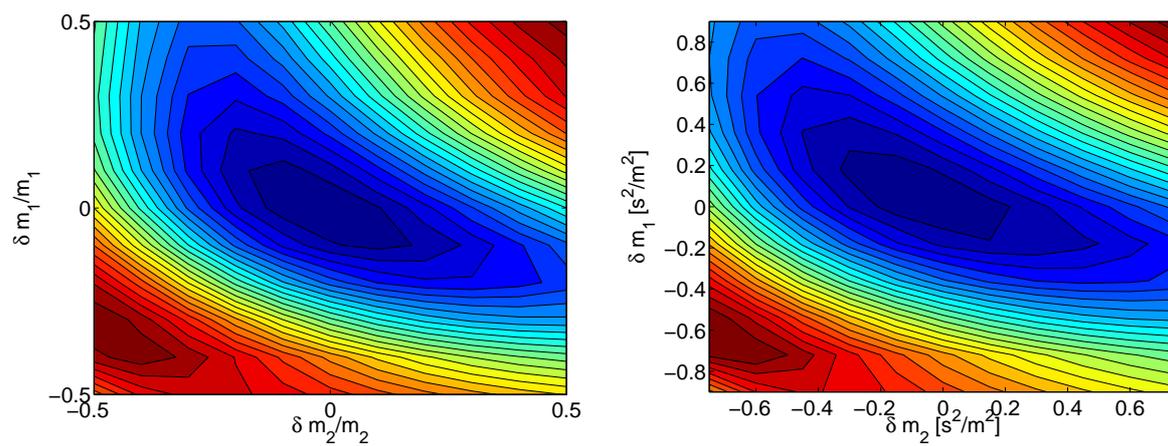
Figure 6. Spatial shift functional (eq. (34)) as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. This functional does not depend on the frequency content of the data.

(a) (b) (c)
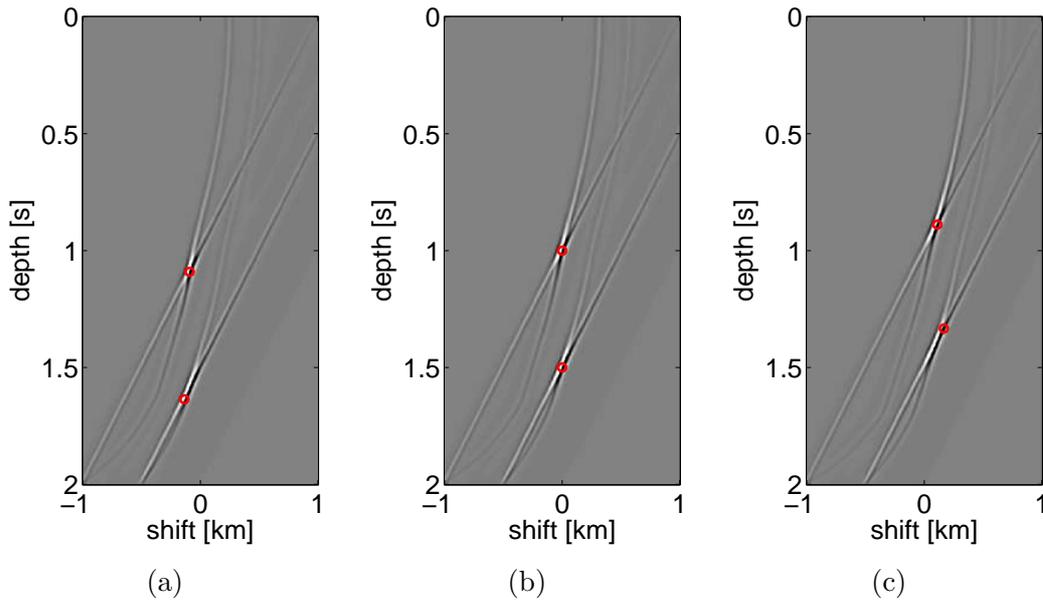
Figure 7. Reflectivity estimate $r(t_0, \Delta t)$, used for the temporal correlation, for a velocity that is 10% too low (a), correct (b) and 10% too high (c). The trajectories along which the singularities in the reflectivity are smeared out, cf. eq. (44), are indicated in red.
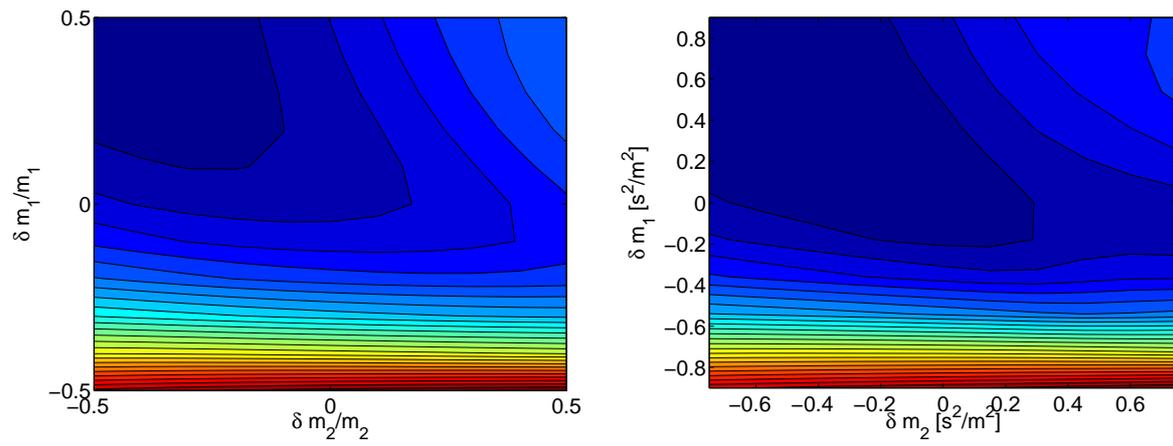
Figure 8. Temporal shift functional as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. This functional does not appear to be suited for velocity analysis since the minimum is in the wrong place.
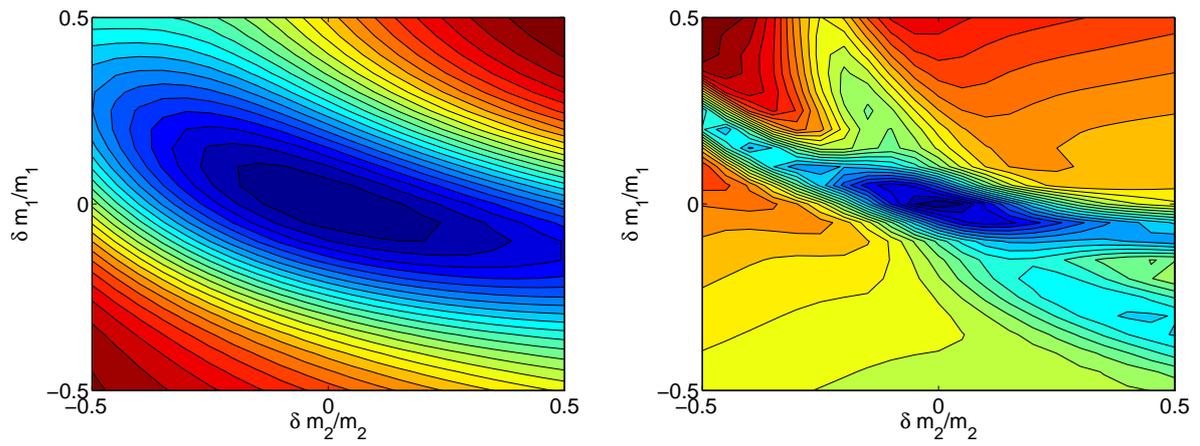
Figure 9. MBTT functional (eq. (47)) as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. As the peak frequency increases, the behaviour of the functional becomes less smooth.
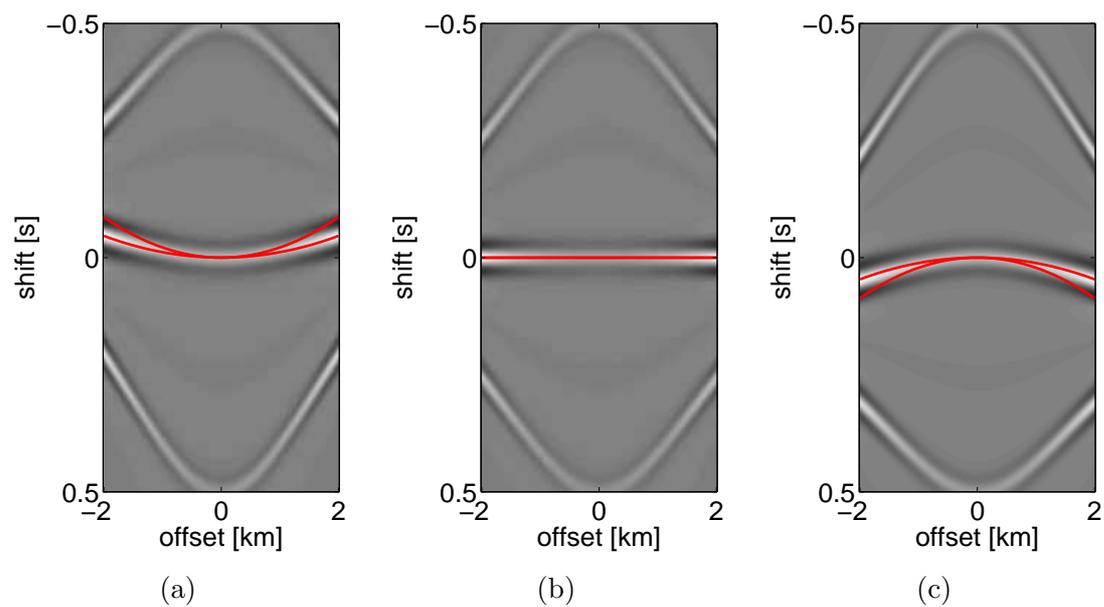
Figure 10. Data correlation C_t for a velocity that is 10% too low (a), correct (b) and 10% too high (c). The focusing events, cf. eq. (44), are indicated in red. Note the presence of the crosstalk.
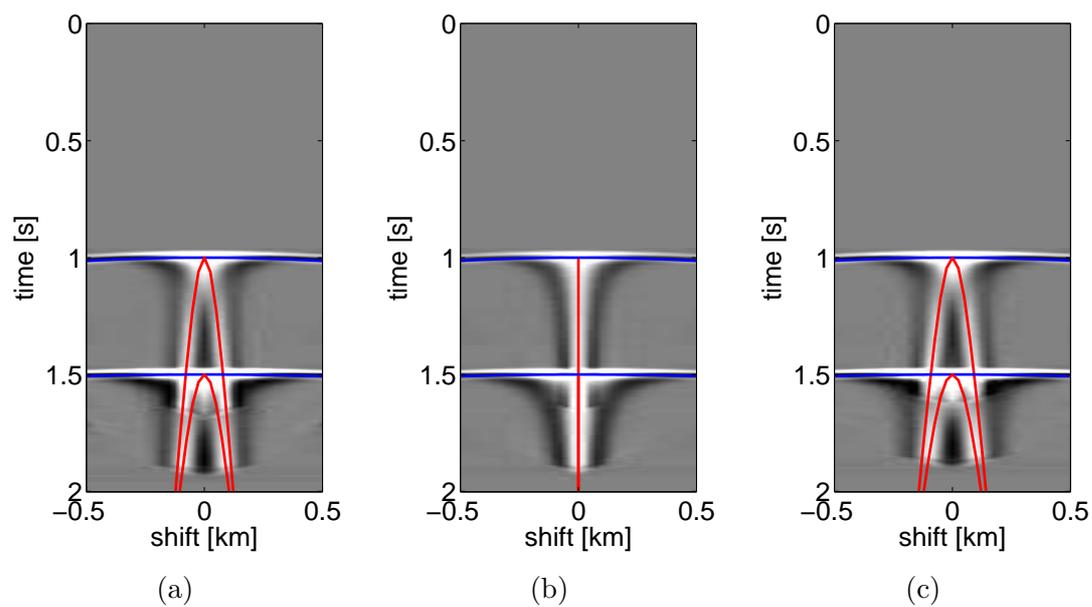
Figure 11. Data correlation C_h for a velocity that is 10% too low (a), correct (b) and 10% too high (c). The focusing events (cf. eq. (53)) are indicated in red while the *non*-focusing events are indicated in purple.
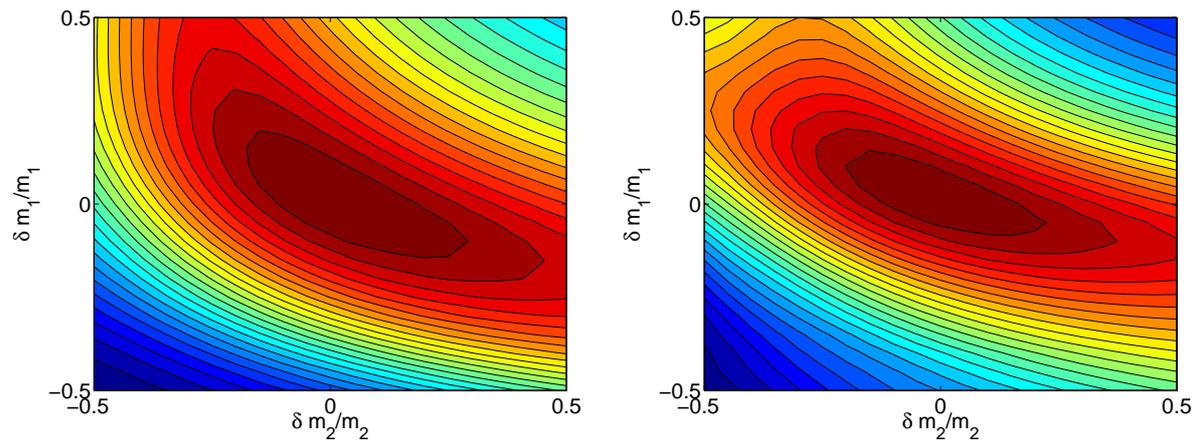
Figure 12. Spatial correlation functional (eq. (54)) as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. The functional does not change its behaviour dramatically as the frequency content of the data changes.
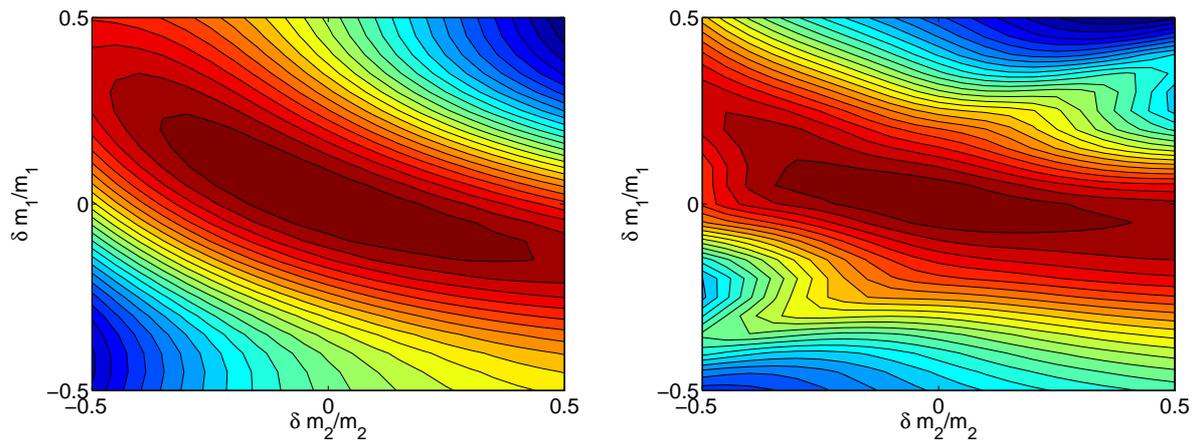
Figure 13. Temporal correlation functional (eq. (54)) as a function of the slowness perturbation $\{\delta m_1, \delta m_2\}$, for peak frequencies 5 Hz and 30 Hz, respectively. The functional does not change its behaviour dramatically as the frequency content of the data changes.
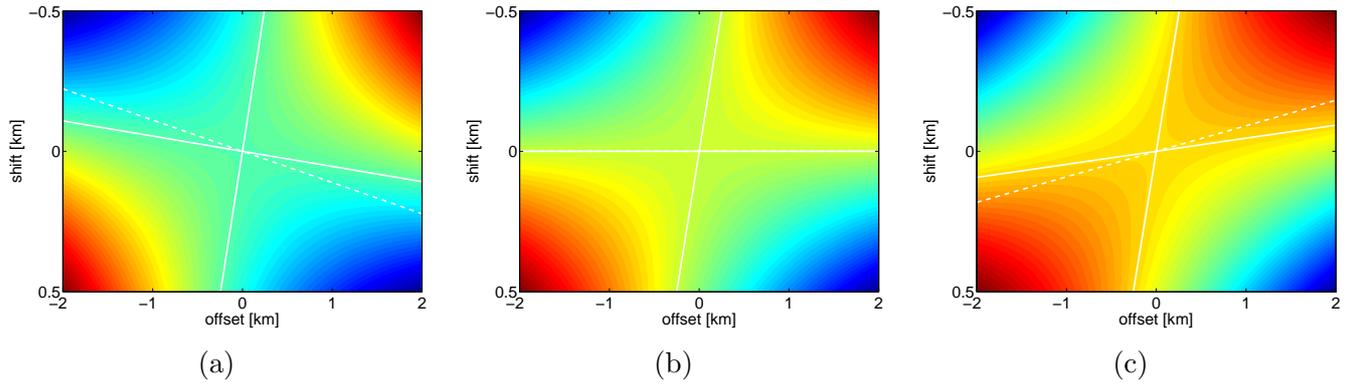
Figure 14. Phase function $\bar{\tau}(t_0, h) - \tau(t_0, h + \Delta h)$ at a fixed t_0 for a velocity that is (a) 10% too low, (b) correct and (c) 10% too high. The solid lines indicate where the phase function is zero. One of the trajectories moves towards $\Delta h = 0$ as $m \to \bar{m}$ and creates the focusing event in the data correlation. The other trajectory creates the non-focusing artifact in the data correlation. This analysis suggests that this artifact is caused by the small offsets. The dotted line indicates where the derivative of the phase function is zero. The trajectory moves towards $\Delta h = 0$ as $m \to \bar{m}$ and drives the focusing in the image-domain.

(a)   (b)
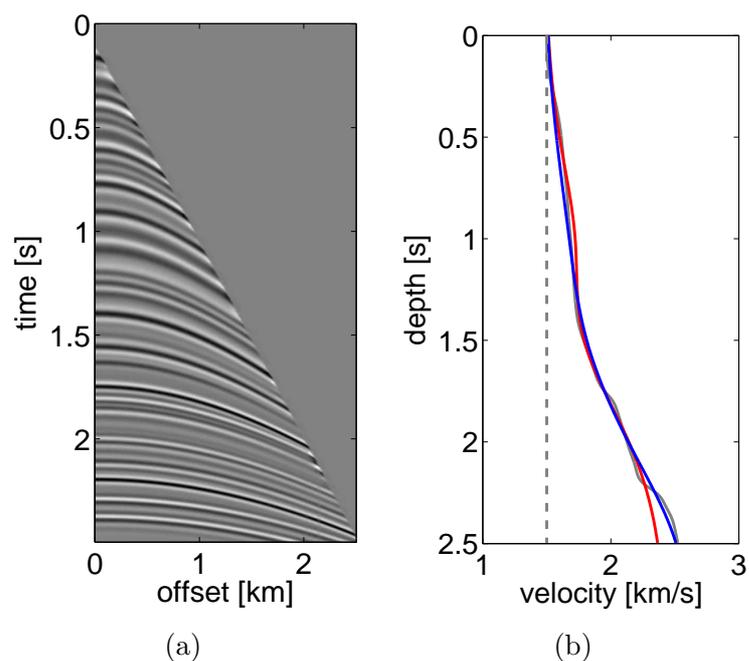
Figure 15. Synthetic-data example. The simulated data are generated with the convolutional model, using the first trace of the data as reflectivity and a Ricker wavelet with a peak frequency of 30 Hz. (a) Shows the synthetic data and (b) shows the initial (dashed) and final (solid) velocity model that resulted from the DSO (red) and data-correlation method(blue). The true RMS velocity is depicted in grey. The match is not prefect because we represent the NMO velocity with cubic splines.

(a)   (b)
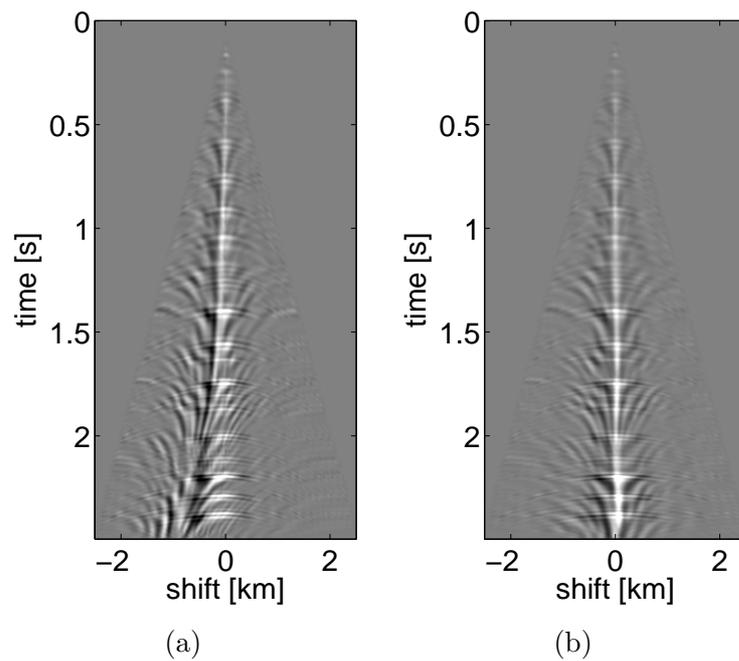
Figure 16. Data correlation for the synthetic-data example from Fig. 15 (a). The spatial correlation is depicted for the initial (left) and final (right) velocity model (15 (b), solid blue). The latter is clearly focused, whereas the former is not. It seems that the focusing is not quite as perfect around t= 2.25 s. This probably due to the rapid variations in the RMS velocity model that could not be captured by splines.
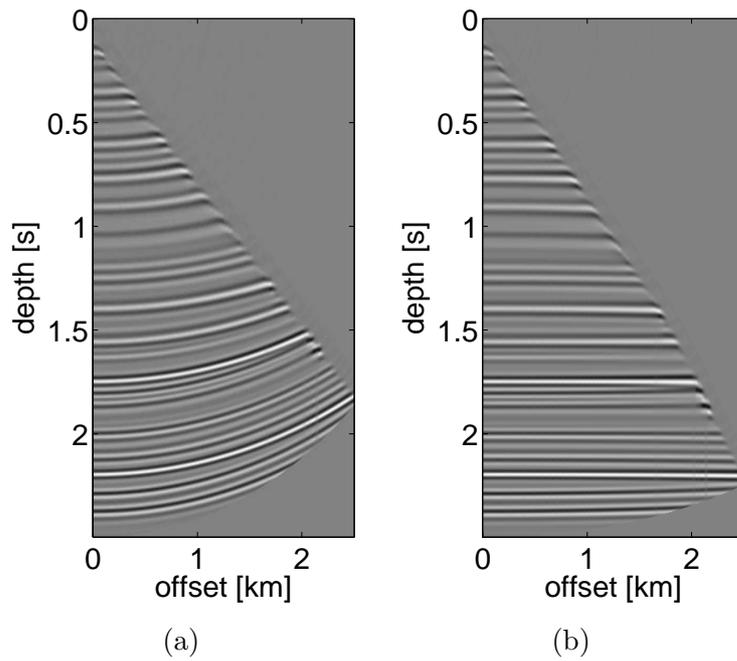
Figure 17. Reflectivity estimate $r(t_0, h)$ for the synthetic-data example from Fig. 15 (a). The reflectivity is depicted for the initial (left) and final (right) velocity model (15 (b), solid red). The latter is clearly 'flat', whereas the former is not.