Geophysical Prospecting, 2008, **56**, 791–803 doi:10.1111/j.1365-2478.2008.00704.x

Velocity analysis based on data correlation

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, and* [2]*Shell International Exploration and Production, PO Box 60, 2280 AB Rijswijk, The Netherlands*

Received January 2007, revision accepted January 2008

ABSTRACT

Several methods exist to automatically obtain a velocity model from seismic data via optimization. Migration velocity analysis relies on an imaging condition and seeks the velocity model that optimally focuses the migrated image. This approach has been proven to be very successful. However, most migration methods use simplified physics to make them computationally feasible and herein lies the restriction of migration velocity analysis. Waveform inversion methods use the full wave equation to model the observed data and more complicated physics can be incorporated. Unfortunately, due to the band-limited nature of the data, the resulting inverse problem is highly nonlinear. Simply fitting the data in a least-squares sense by using a gradient-based optimization method is sometimes problematic. In this paper, we propose a novel method that measures the amount of focusing in the data domain rather than the image domain. As a first test of the method, we include some examples for 1D velocity models and the convolutional model.

INTRODUCTION

Most successful methods to obtain a velocity model from seismic data automatically try to focus the migrated image in some sense. This approach to velocity analysis is called migration velocity analysis (MVA) and has a long history. The underlying principle of MVA has not changed but the precise way in which the amount of focusing of the image is measured and the migration algorithms themselves have evolved. A few well-known methods found in this class include differential semblance optimization, proposed by Symes and Carazzone (1991), which uses the flatness of image gathers as a measure, and methods based on a generalized imaging condition, proposed by Doherty and Claerbout (1974) and elaborated on by many others (Yilmaz and Chambers 1984; Stork 1992; MacKay and Abma 1992; Rickett and Sava 2002; Sava and Biondi 2004; Sava and Fomel 2006). These methods have been proven successful on data in which the multiples are not too strong (Mulder and ten Kroode 2002; Shen, Symes and Stolk 2003; Dussaud and Symes 2005; Li and Symes 2007). A common factor among all of these is that the migration algorithms are usually based on simplified physics (high-frequency asymptotics, one-way wave equations and so on), restricting the approach to dealing only with primary reflections. This motivates us to take a look at the data domain, where, in principle, we can include multiples by using the full wave equation to model the data. It should be mentioned that some attempts have been made to use multiples in migration (see, for example, Youn and Zhou 2001; Brown and Guitton 2005; Jiang *et al.* 2007).

Data-domain methods come in various shapes and sizes, however, we will focus on waveform inversion. Another class of methods that works in the data domain is known as traveltime tomography and requires the picking of events. Full waveform inversion relies on fitting the measured data. The full wave equation is used to model the data and the velocity model that minimizes the least-squares error between measured and simulated data is sought. A well-known problem here is that the gradient-based methods fail to converge if the initial velocity model is too far away from the true one. The least-squares misfit between observed and simulated data does not give a correct model update if seismic events are more than half a wavelength apart. We say that the basin of

E-mail: tristan.vanleeuwen@tudelft.nl

attraction of the least-squares functional is small. Numerous studies have been published on the use of the least-squares approach on the nonlinear problem (Tarantola and Valette 1982; Tarantola 1984; Pratt and Hicks 1998 and many others). Because of the absence of low frequencies in the seismic data, the inverse problem is linear in reflectivity (the oscillatory part of the velocity model) and strongly nonlinear in the background model (the smooth part of the velocity model). Using a gradient-based method will effectively perform a migration over the initial background (Lailly 1983) but will have great difficulty in updating the background model beyond the depth reached by diving waves (Mora 1988; Pratt *et al.* 1996). A method that tries to alleviate this problem is migration-based traveltime tomography, proposed by Chavent and Clément (1993) (see also Plessix, Chavent and De Roeck 1999; Clément, Chavent and Gómez 2001). Here, the iterations consist of two parts: first, a reflectivity for the current velocity model is constructed by one least-squares iteration; secondly, the data for this reflectivity are modeled and the least-squares criterion is used to update the background velocity model. This has been proven to work, however, the initial model can still not be too far away from the true model. In a way, migration-based traveltime tomography can be seen as a data-domain formulation of the migration velocity analysis (MVA) methods. The MVA methods measure the amount of focusing of the migrated image directly, whereas migration-based traveltime tomography generates data for the migrated image and compares this to the measured data. The main difference, apart from the domain the methods work in, is the way the misfit is measured. MVA measures the amount of focusing of a correlation of wavefields and migration based traveltime tomography measures the least-squares error, which can be seen as an analogue of measuring the stacking-power in the image domain. In this paper we propose a novel method that measures the amount of focusing of the data correlation. Although the method is intended to be used in a full waveform setting, our goals for this paper are less ambitious. To assess the feasibility of this approach, we restrict ourselves to 1D velocity models and the convolutional model to generate multiple-free data. The reflectivity and NMO (normal moveout) velocity are parameterized in vertical traveltime. To emulate migration based traveltime tomography we let the reflectivity as a function of vertical traveltime be known. This way, the zero-offset traveltimes are fixed and we can concentrate on updating the background model with the proposed functionals.

The paper is organized as follows. We first explore the properties of the correlation of seismograms and propose a family of functionals to measure the misfit. We give a short overview of the image domain methods and show there are some connections between the correlation in the depth-domain, as used in generalized imaging conditions and the correlation in the data domain. Finally, we present results on synthetic and real data.

DATA CORRELATION

Given two seismograms $p(t, h)$ and $q(t, h)$, where t and h denote time and offset respectively, we denote the temporal correlation of these seismograms by

$$C_t[p, q](\Delta t, h) = \int p(t, h)q(t + \Delta t, h)\,\mathrm{d}t, \tag{1}$$

the spatial correlation as

$$C_h[p, q](t, \Delta h) = \int p(t, h)q(t, h + \Delta h)\,\mathrm{d}h, \tag{2}$$

and the space-time correlation as

$$C_{t,h}[p, q](\Delta t, \Delta h) = \iint p(t, h)q(t + \Delta t, h + \Delta h)\,\mathrm{d}t\,\mathrm{d}h. \tag{3}$$

We consider the correlation for finite shifts: $-\Delta t_{\max} \leq \Delta t \leq \Delta t_{\max}$ and $-\Delta h_{\max} \leq \Delta h \leq \Delta h_{\max}$. In Fig. 1, we show an example of the auto correlation of a seismogram and the correlation of two seismograms with different moveout. It is readily observed that if the seismograms are identical (i.e., the velocity model is correct), there is a large peak at $\Delta t = \Delta h = 0$. The events at non-zero shifts are caused by 'cross-talk' between events in the seismograms. This cross-talk is considered noise in the current approach. Note that the temporal correlation suffers most from this cross-talk.

If the seismograms have different moveout (i.e., the velocity model is not correct) we observe that the energy is not focused in the centre. We use this focusing property in the data domain for velocity analysis. Below, we propose a family of functionals that are based on this property. We show by a few examples that these functionals respond smoothly to velocity perturbations and have a larger basin of attraction than the traditional least-squares functional. Consequently, these functionals should be fit for use with a gradient-based optimization scheme.

Functionals

We capture the amount of focusing of the correlation functions in a functional. This functional should attain an extremum for the correct velocity model. Luo and Schuster (1989) used the correlation of first arrivals for traveltime tomography. The

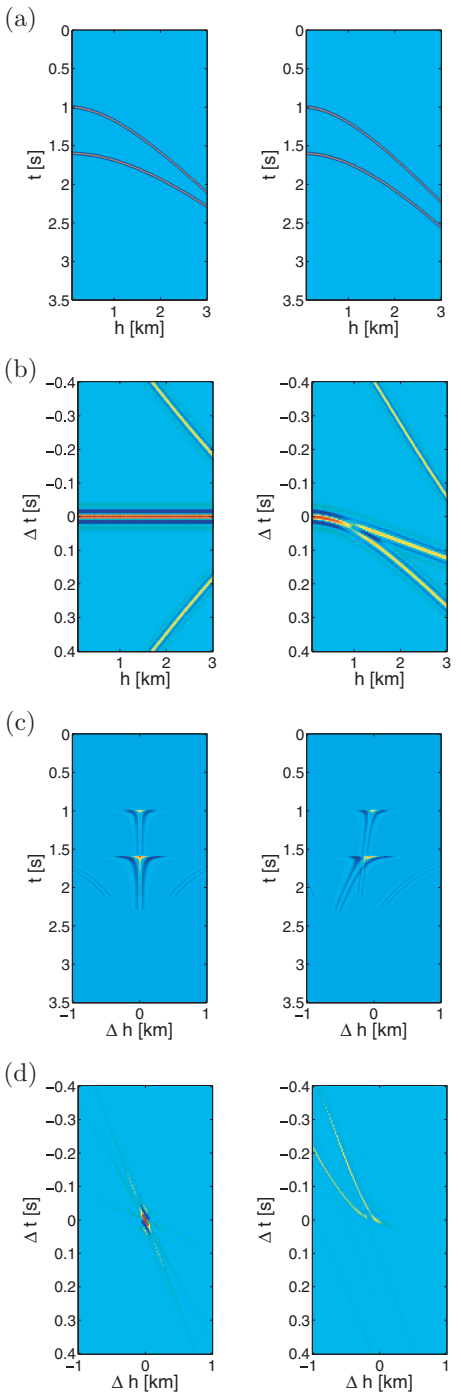

Figure 1 (a) Data, (b) temporal correlation, (c) spatial correlation and (d) space-time correlation for the correct (left) and a wrong (right) velocity model. There is a clear distinction between the focusing of the correlation for the correct and a wrong velocity model. The events that do not focus at zero shift for the correct velocity model are caused by correlations between neighbouring events (cross-talk). Note that in (b) and (d) all the events are superimposed around zero shift for all offsets and that in (c) the events all have different zero-offset traveltimes.

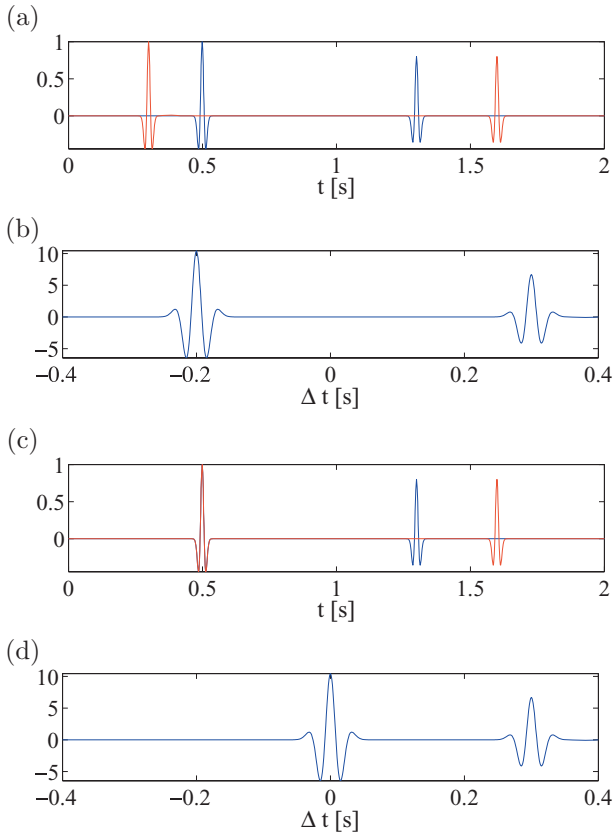

Figure 2 This example shows that the shift of the largest peak in the correlation is not an appropriate measure of similarity. In (a) we see two traces, the 'correct' one in blue and the 'wrong' one in red. The correlation of the two is depicted in (b). Indeed, the energy is not focused in the centre. After correction, the strongest event is in the right place (c) and the correlation is similar to (d). Here, we see that the shift of the largest peak is 0 *s* even though the traces are not identical. To properly measure the similarity of the traces, all the peaks in the correlation should be taken into account.

functional they proposed to measure the amount of focusing is based on the shifts of the largest peaks in the correlation. This is appropriate if there is only one event in the correlation but is not appropriate in the case of more than one event: Fig. 2 shows why.

To measure the amount of focusing of the data correlation, we propose to use a weighted norm. The weight should be such that the norm either measures the amount of energy around zero shift (this should be maximized) or the amount of energy away from zero shift (this should be minimized). This approach is also used in some migration-based velocity analysis methods, where the amount of focusing of a migrated image is measured in a similar way (Shen *et al.* 2003). The general form of the class of functionals for the temporal correlation

is:

$$J_t = \frac{1}{S_t} \iint_{-\Delta t_{\max}}^{\Delta t_{\max}} W(\Delta t) \Big(C_t[p,q](\Delta t, h) \Big)^2 \, \mathrm{d}\Delta t \, \mathrm{d}h, \tag{4}$$

$$S_t = \iint_{-\Delta t_{\max}}^{\Delta t_{\max}} \Big(C_t[p,q](\Delta t, h) \Big)^2 \, \mathrm{d}\Delta t \, \mathrm{d}h. \tag{5}$$

where $W(\Delta t)$ is a weighting function, to be specified below. The normalization S_t ensures that the functional is not minimized by simply shifting all the events outside the correlation window. For the spatial and space-time correlation, we define the functionals J_h and $J_{t,h}$ in a similar way.

The maximal shift plays a role in reducing the influence of the cross-talk. If the maximal shift is chosen too small, the weighted norm may fail to capture the deviation in the focusing of the correlation. If the maximal shift is chosen too large, the cross-talk may dominate over the focusing in the weighted norm.

Weighting functions

We will either measure the amount of energy away from zero shift or around zero shift. The weighting function we propose to measure the energy away from the centre is:

$$W_1(\Delta t) = \left(\frac{\Delta t}{\Delta t_{\max}} \right)^2, \tag{6}$$

and similarly for the spatial and space-time correlation. This weight will penalize peaks at non-zero shift and thus measure the amount of energy away from the centre. The weighted norm can only be zero if the wavelet is a delta pulse and all cross-talk is excluded by choosing an appropriate maximal shift. In the realistic case with a finite-bandwidth wavelet, where we cannot exclude all the cross-talk, the weighted norm will be very small for the correct velocity model.

A weight that measures energy around zero shift is a Gaussian weighting function:

$$W_2(\Delta t) = -e^{-\alpha \Delta t^2}. \tag{7}$$

The minus sign ensures that minimizing the functional will maximize the amount of energy around the centre and the parameter α controls the width of the weighting function. For large values of α the exponent decays quickly and the functional only measures the energy in a small region around the centre. Peaks that are far away are not taken into account. For small values of α, the weighting function is broader and events further away are taken into account. The weighting function should at least be wide enough to capture the event at zero shift. This means that the width should at least be in the order of the length of the wavelet.

To give an idea of the basin of attraction of the proposed functionals, we show how the functionals respond to a perturbation of the true velocity model in Fig. 3. The velocity model (blue) and the perturbed model (green) are depicted in Fig. 3(a). Figures 3(b), (c) and (d) display the temporal, spatial and space-time correlation based functionals using the quadratic weight defined in equation 6 (left) and the Gaussian weight, defined in equation 7 (right). Also depicted is the least-squares functional in red. For the correlation based functionals the plots are made for different values of the width parameters $\Delta t_{\max}$, $\Delta h_{\max}$, and α. The spatial and space-time correlation based functionals in particular respond smoothly to the perturbation and have a much larger basin of attraction than the least-squares functional.

GENERALIZED IMAGING CONDITIONS

The classical imaging condition states that the reflectivity can be recovered by correlating the source and receiver wavefields u_s and u_r at depth and by extracting the part at zero shift. The generalized imaging condition (Doherty and Claerbout 1974; Yilmaz and Chambers 1984; MacKay and Abma 1992; Rickett and Sava 2002; Sava and Fomel 2006) allows a finite spatial and temporal shift between the wavefields. The correct velocity model will focus energy in the correlation at zero shift. This approach to velocity analysis is known as migration velocity analysis (MVA).

The generalized image is given by

$$I(t_0, \Delta t, \Delta h) = \iint u_s(t_0, t + \Delta t, h + \Delta h) u_r(t_0, t, h) \, \mathrm{d}t \, \mathrm{d}h, \tag{8}$$

where t_0 denotes depth measured in vertical two-way traveltime. For $\Delta t = \Delta h = 0$ we obtain the classical imaging condition. With the convolutional model in 1D, the wavefields u_s and u_r are simply time-shifted versions of the wavelet and seismogram:

$$u_s(t_0, t, h) = w\Big(t - \tfrac{1}{2}\tau(t_0, h)\Big), \tag{9}$$

$$u_r(t_0, t, h) = p\Big(t + \tfrac{1}{2}\tau(t_0, h), h\Big). \tag{10}$$

Here, $\tau(t_0, h)$ denotes the two-way traveltime as a function of zero-offset two-way traveltime t_0 and offset h. Thus we can

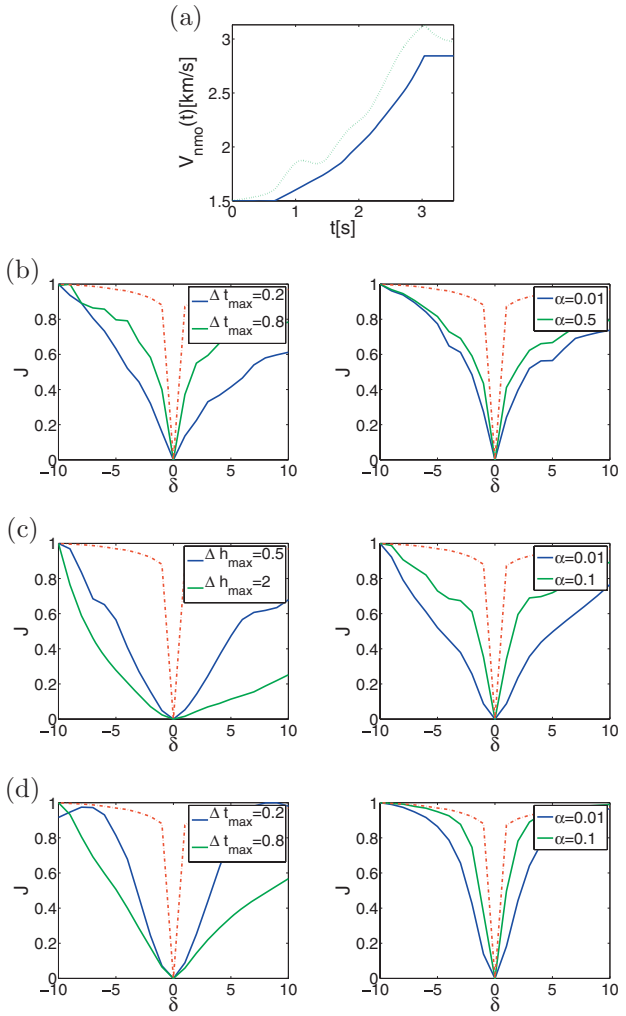

Figure 3 Dependence of the functionals on a perturbation of the correct velocity model. The functionals are shifted and scaled to allow for visual comparison. The amplitude of the perturbation, δ, is represented on the horizontal axis. The correct velocity model (blue) is depicted in (a), as well as the perturbed model for $\delta = 1$ (green). (b) shows the temporal correlation-based functional, (c) the spatial correlation and (d) the space-time correlation. The functionals using the quadratic weight are shown on the left and the ones using the Gaussian weight on the right. The red dashed lines show the least-squares functional. The correlation-based functionals are shown for two different values of the width parameter. It seems that for the spatial and space-time correlation there is a direct relation between the width of the weighting function and the width of the basin of attraction; a larger maximal shift or smaller α results in a wider weighting function and a broader basin of attraction.

write the generalized image as

$$I(t_0, \Delta t, \Delta h) = \iint p(t + \Delta t, h + \Delta h) w\Big(t - \tau(t_0, h)\Big)\, \mathrm{d}t\, \mathrm{d}h \tag{11}$$

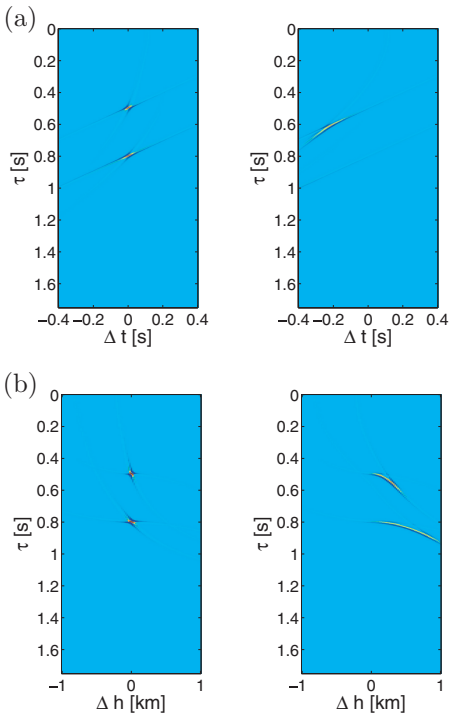

Figure 4 Examples of the generalized image for the data displayed in Fig. 1 (a) for the correct (left) and a wrong (right) velocity model. (a) Shows the image at $\Delta h = 0$ and (b) at $\Delta t = 0$. As with the data correlation the energy focuses around zero shift for the correct velocity model. Note the absence of crosstalk.

In Fig. 4, we show an example of the generalized image for the correct and for a wrong velocity model. As with the data correlation, the energy is focused around zero shift for the correct velocity model. Note that there is no cross-talk present in the generalized image.

Functionals

Shen *et al.* (2003) propose a functional based on the generalized image and use this in combination with an optimization method to perform automatic MVA (see also Sava and Biondi 2004). This approach is akin to differential semblance optimization, proposed by Symes and Carazzone (1991). The functional is a weighted norm of the generalized image:

$$J_h = \iint \Delta h^2 \Big(I(t_0, \Delta t = 0, \Delta h)\Big)^2\, \mathrm{d}t_0\, \mathrm{d}\Delta h. \tag{12}$$

Of course, we can also use the functionals proposed before to measure the focusing of the generalized image. Crossplots of the functionals are depicted in Fig. 5. These can be compared to the crossplots of the data-domain functionals. The

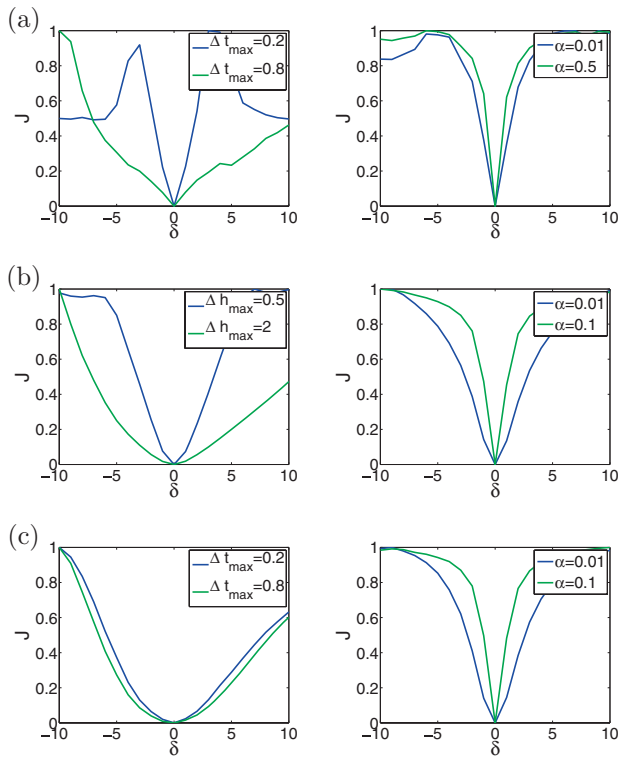

Figure 5 Crossplots for the image-domain functionals. The velocity model and perturbation from Fig. 3 are used. The functionals based on the temporal, spatial and space-time correlations are shown in (a), (b), and (c), respectively. The ones on the left use the quadratic weight and on the right use the Gaussian weight. All the functionals are shown for two different values of the width parameter. The functionals are scaled between 0 and 1 to allow for visual comparison. The functionals show a similar behaviour as their data-domain analogues.

spatial and space-time correlation based functionals in particular respond smoothly to the perturbation. The functionals show a similar behaviour as their data-domain analogues

SIMILARITIES BETWEEN DATA- AND IMAGE-DOMAIN FOCUSING

In this section, we briefly explore the relations between the data- and image-domain functionals in the high-frequency limit, under the assumptions of the convolutional model.

Given a reflectivity $r(t_0)$, traveltime $\tau(t_0, h)$ and a wavelet $w(t)$ we model the data as:

$$p(t,h) = \int r\Big(\tau_0(t',h)\Big) w(t-t')\,\mathrm{d}t', \tag{13}$$

where $\tau_0(t, h)$ is the inverse of the traveltime such that $\tau(\tau_0(t, h), h) = t$. In the high-frequency limit (the wavelet becomes a delta-pulse) we can write this as:

$$p(t,h) = r\Big(\tau_0(t,h)\Big). \tag{14}$$

In the following we will denote the measured or true quantities with a hat ($\hat{}$) and the simulated quantities without one. In the high-frequency limit we can write the data correlation as:

$$C_t[p,\hat{p}](\Delta t,h) = \int r\Big(\tau_0(t,h)\Big)\hat{r}\Big(\hat{\tau}_0(t+\Delta t,h)\Big)\,\mathrm{d}t, \tag{15}$$

$$C_h[p,\hat{p}](t,\Delta h) = \int r\Big(\tau_0(t,h)\Big)\hat{r}\Big(\hat{\tau}_0(t,h+\Delta h)\Big)\,\mathrm{d}h, \tag{16}$$

$$C_{t,h}[p,\hat{p}](\Delta t,\Delta h) = \iint r\Big(\tau_0(t,h)\Big)\hat{r}\Big(\hat{\tau}_0(t+\Delta t,h+\Delta h)\Big)\,\mathrm{d}t\,\mathrm{d}h. \tag{17}$$

The generalized image can be written as:

$$I(t_0,\Delta t,\Delta h) = \int \hat{r}\Big(\hat{\tau}_0\Big(\tau(t_0,h)+\Delta t\Big),h+\Delta h\Big)\,\mathrm{d}h. \tag{18}$$

Note that for the correct velocity model, $I(t_0, 0, 0) \propto \hat{r}(t_0)$.

With a few extra integrals and coordinate transforms we find the following relations:

$$\int C_t[p,\hat{p}](\Delta t,h)\,\mathrm{d}h = \iint r(t_0)\hat{r}\Big(\hat{\tau}_0\Big(\tau(t_0,h)+\Delta t,h\Big)\Big)s(t_0,h)\,\mathrm{d}t_0\,\mathrm{d}h, \tag{19}$$

$$\int C_h[p,\hat{p}](t,\Delta h)\,\mathrm{d}t = \iint r(t_0)\hat{r}\Big(\hat{\tau}_0\Big(\tau(t_0,h),h+\Delta h\Big)\Big)s(t_0,h)\,\mathrm{d}h\,\mathrm{d}t_0, \tag{20}$$

$$C_{t,h}[p,\hat{p}](\Delta t,\Delta h) = \iint r(t_0)\hat{r}\Big(\hat{\tau}_0\Big(\tau(t_0,h)+\Delta t,h+\Delta h\Big)\Big)s(t_0,h)\,\mathrm{d}t_0\,\mathrm{d}h. \tag{21}$$

where $s(t_0) = \frac{\partial\tau}{\partial t_0}(t_0,h)$ is a stretch factor due to the coordinate change from time to depth. If $s(t_0, h) \approx 1$ – this is also the range for which the convolutional model is valid (Symes 1999), typically offsets larger than $1000t$ are blanked out – we find the following relations between the data correlation and the generalized image:

$$\int C_t[p,\hat{p}](\Delta t,h)\,\mathrm{d}h \sim \int r(t_0)I(t_0,\Delta t,\Delta h=0)\,\mathrm{d}t_0, \tag{22}$$

$$\int C_h[p,\hat{p}](t,\Delta h)\,\mathrm{d}t \sim \int r(t_0)I(t_0,\Delta t=0,\Delta h)\,\mathrm{d}t_0, \tag{23}$$

$$C_{t,h}[p,\hat{p}](\Delta t,\Delta h) \sim \int r(t_0)I(t_0,\Delta t,\Delta h)\,\mathrm{d}t_0. \tag{24}$$

This shows that the space-time correlation is a weighted integral of the generalized image over depth.

Functionals

With the relations derived above we can express the data-domain functional as (leaving out the normalization factor S for the moment):

$$J^{\text{data}}_{t,h} = \iint W(\Delta t, \Delta h)(C_{t,h}[p, \hat{p}](\Delta t, \Delta h))^2 \, \mathrm{d}\Delta t \, \mathrm{d}\Delta h$$
$$= \iint W(\Delta t, \Delta h)\left(\int r(t_0) I(t_0, \Delta t, \Delta h) \, \mathrm{d}t_0\right)^2 \mathrm{d}\Delta t \, \mathrm{d}\Delta h. \quad (25)$$

The corresponding image-domain functional is given by

$$J^{\text{image}}_{t,h} = \iiint W(\Delta t, \Delta h)(I(t_0, \Delta t, \Delta h))^2 \, \mathrm{d}\Delta t \, \mathrm{d}\Delta h \, \mathrm{d}t_0. \quad (26)$$

The main difference between the data and image-domain functionals here is the order in which the weighting and integration over depth is carried out. Also, the image is weighted with the reflectivity in the data-domain functional. Despite the apparent difference between the image- and data-domain approaches, it seems that the functionals exhibit similar behaviour when we consider the crossplots in Figs 3 and 5.

TESTS WITH SYNTHETIC DATA

Here, we present results for velocity analysis on synthetic data for two 1D velocity models. The data are generated by the convolutional model. The experiments are intended as a first test, to see if the method proposed here is promising. We will give a brief overview of the experimental setup (generation of synthetic data and optimization method), followed by the results of these experiments.

Synthetic data

The synthetic data were modelled by the convolutional model, see equation 13. The traveltimes are given by

$$\tau(t_0, h) = \sqrt{t_0^2 + (h/v_{\text{nmo}}(t_0))^2}, \quad (27)$$

where t_0 denotes depth in vertical traveltime and $v_{\text{nmo}}(t_0)$ is the NMO velocity. The reflectivity is a spike train (see Fig. 6(a), for example) and as wavelet we used a Ricker wavelet with a peak frequency of 30 Hz.

For the experiments, we have used two velocity models, depicted in Figs 6 and 7. We started the optimization from a constant initial velocity model. The synthetic data were generated with the 'true' reflectivity, which basically means that the zero-offset traveltimes were fixed and we only had to correct for a wrong moveout.

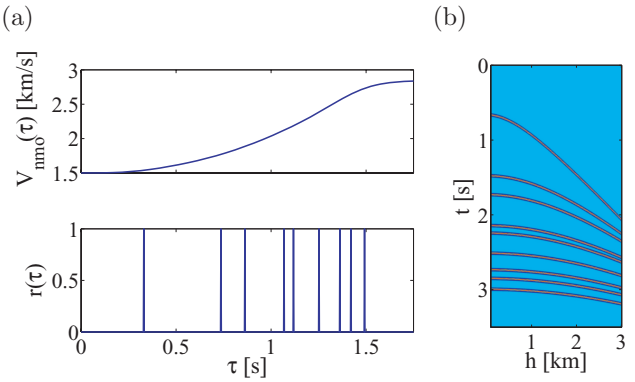

Figure 6 1D velocity model used in the experiments. (a) NMO velocity and reflectivity. (b) Synthetic data.

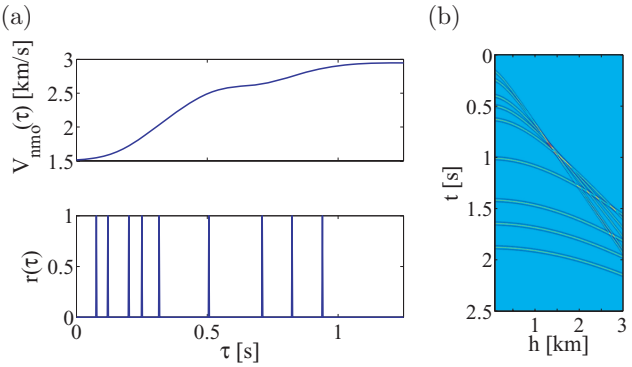

Figure 7 1D velocity model used in the experiments. (a) NMO velocity and reflectivity. (b) Synthetic data.

Optimization method

We represent the NMO velocity on a cubic-spline grid and use the values at the nodes as our model parameters. For the optimization we use a BFGS (Broyden-Fletcher-Goldfarb-Shanno) method with the exact gradient. This gradient is derived in the Appendix. It proved desirable to use a smoothness constraint to avoid unnecessary peaks in the solutions. For an overview of the BFGS method and Tikhonov regularization we refer to Vogel (2002).

Results

Figures 8–11 show the results of the experiments. For each experiment, we depict the initial, final, and true velocity model. Obviously, the inversion was successful if the true velocity model was recovered. In some instances, we found perfectly matched simulated data for somewhat different velocity mod-

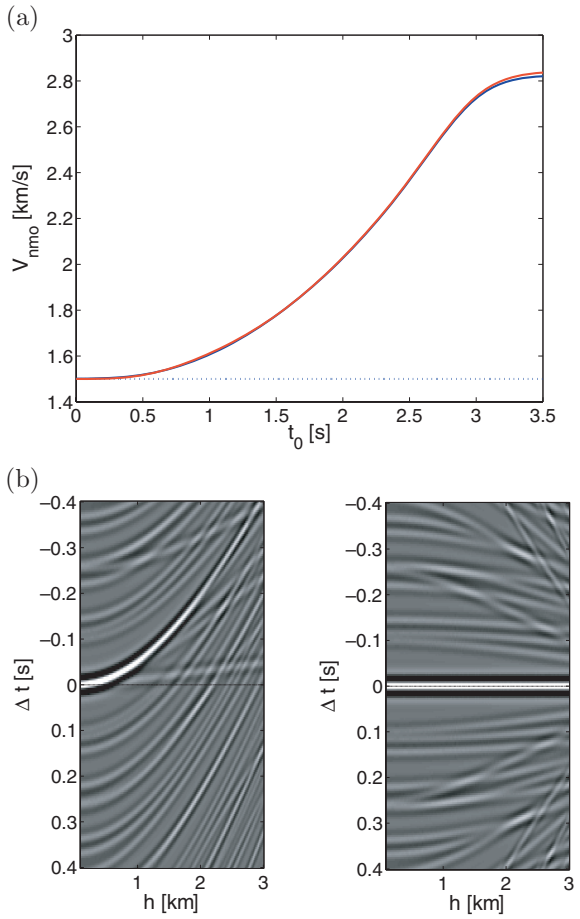

Figure 8 Results obtained for the correlation-based functional with the Gaussian weighting function. (a) Correct (blue), initial (dots) and final (red) velocity models. (b) Initial (left) and final (right) correlations. The iterations converged to the true velocity model.

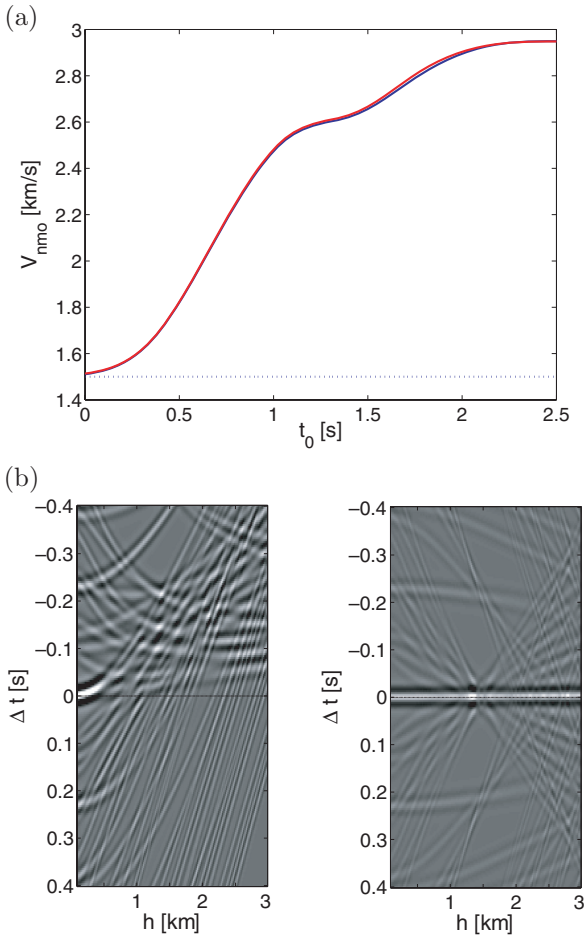

Figure 9 Same as figure 8 but for a different velocity model.

els. This is caused by the null-space of the problem and can be dealt with by regularization. We chose to use a Tikhonov regularization that penalizes large variations in the velocity model. The data correlation for the initial and final velocity models are displayed to show the difference between the focusing of the correlation.

Figures 8 and 9 display the results for the functional based on the temporal correlation with the Gaussian weighting function. Although the final velocity models are not identical to the true ones, they do reproduce the data perfectly. The correlations for the initial velocity models show a deviation in the focusing (Figs 8b and 9b, left). For the final velocity model, the correlation is focused (same figures, right). For the more complicated velocity model (Fig. 9a), the focusing is less distinct due to the cross-talk. Nevertheless, the functional was able to capture the focusing.

In Figs 10 and 11, the results for the functional based on the spatial correlation with the Gaussian weighting function are shown. As in the previous results, the initial and true velocity models are not identical but the data are. Again, this is caused by a lack of reflectors in the regions where the velocity models disagree. In the correlation (Figs 10b and 11b, left) we observe that the events are separated in traveltime, hence the focus of every event can be measured independently. It also shows that the deviation in the focusing increases with time. This agrees with the error in the initial velocity model, which also increases with time.

We present only a few of the several successful experiments. The functionals based on the space-time correlation ($J_{t,b}$) also proved successful on the test problems. Overall, the Gaussian weighting function worked best on our two test problems. This is probably because this weighting function is better able to cope with the cross-talk.

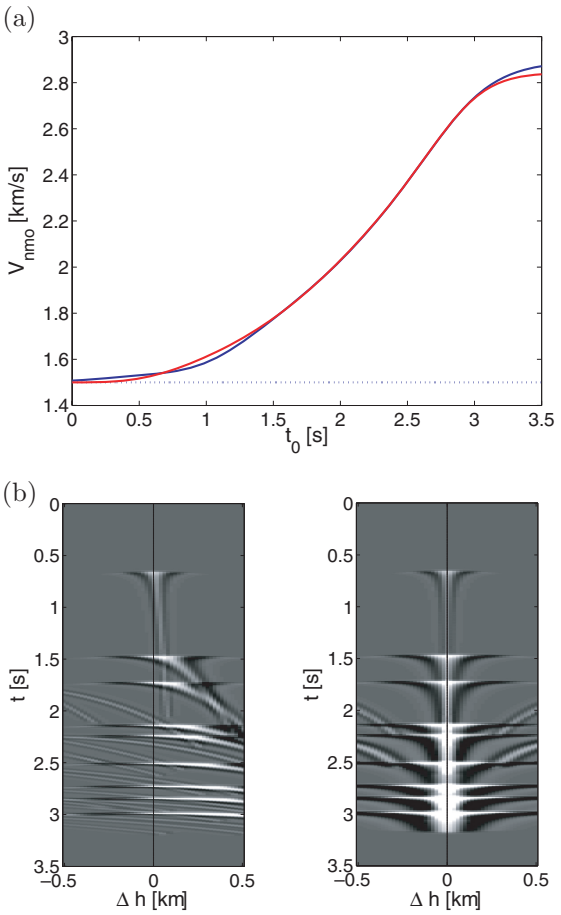

Figure 10 Same as figure 8 but for the space correlation.

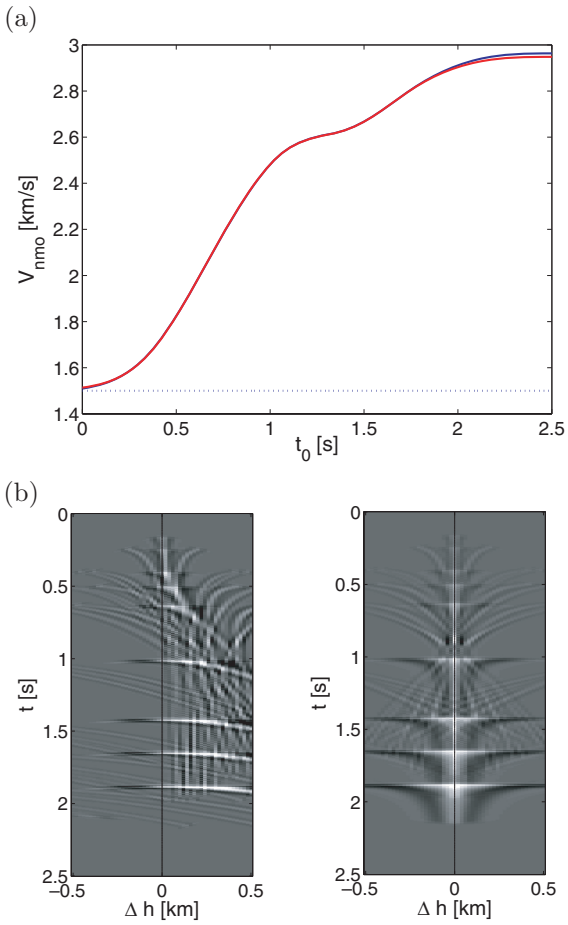

Figure 11 Same as figure 10 but for a different velocity model.

Combining the spatial and temporal correlation in one functional by adding J_t and J_h is another possibility we did not consider. We have not compared the methods in terms of convergence speed, since computational cost is not an issue with the convolutional model.

TEST WITH REAL DATA

We also tested the proposed functionals on real data. The synthetic data are again modelled by the convolutional model. The reflectivity for the synthetic data is taken to be the first trace of the observed data. The wavelet used is a Ricker wavelet with a peak frequency of 30 Hz. The real data were processed slightly in an attempt to remove the surface related multiples.

In Figs 12–14 the results of an inversion using the spatial correlation and the Gaussian weighting function are depicted. The initial and final velocity models are shown in Fig. 12(b). Also, the NMO-corrected data for the initial and final velocity models are shown in Fig. 12(a). The flatness of the events indicates that the final velocity model is indeed the 'correct' NMO velocity. The final velocity model focuses both the data (Fig. 13) and image (Fig. 14) correlation.

CONCLUSIONS AND DISCUSSION

We explored the use of correlation of seismograms for velocity analysis for 1D velocity models. The synthetic seismograms were generated by the convolutional model. The examples showed that the correlation focuses at zero shift for the correct velocity model and responds to errors in the velocity model by de-focusing. To measure the amount of focusing, we proposed a family of functionals. These functionals are weighted norms of the correlation function. In the correlation, peaks that do not focus at zero shift are also present. The functionals have

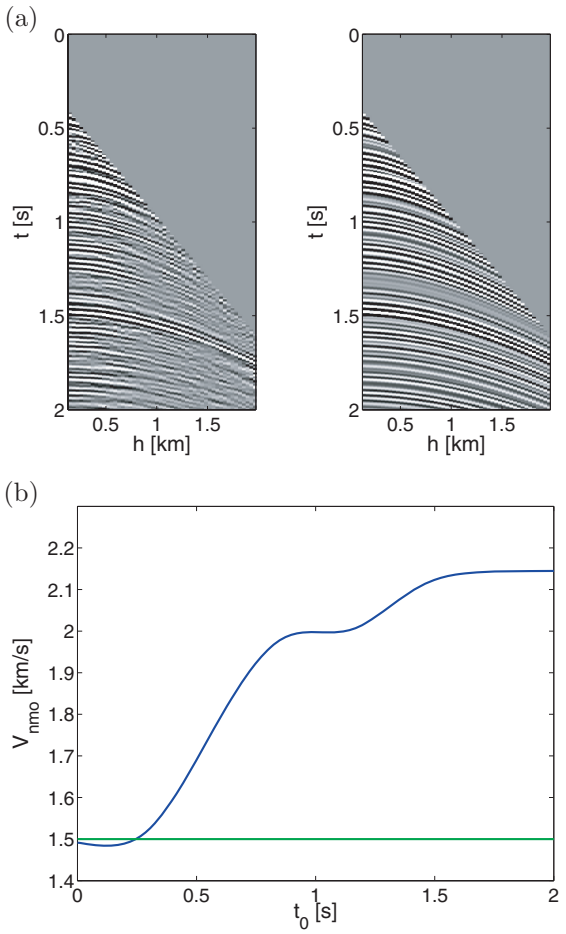

Figure 12 Real data example. The simulated data are generated by the convolutional model using the first trace of the data as reflectivity and a Ricker wavelet with a peak frequency of 30 Hz. (a) Shows the real data (left) and the synthetic data (right) for the final velocity model, which resulted from using the functional based on the spatial correlation with the Gaussian weighting function. (b) Shows the initial (green) and final (blue) velocity models.

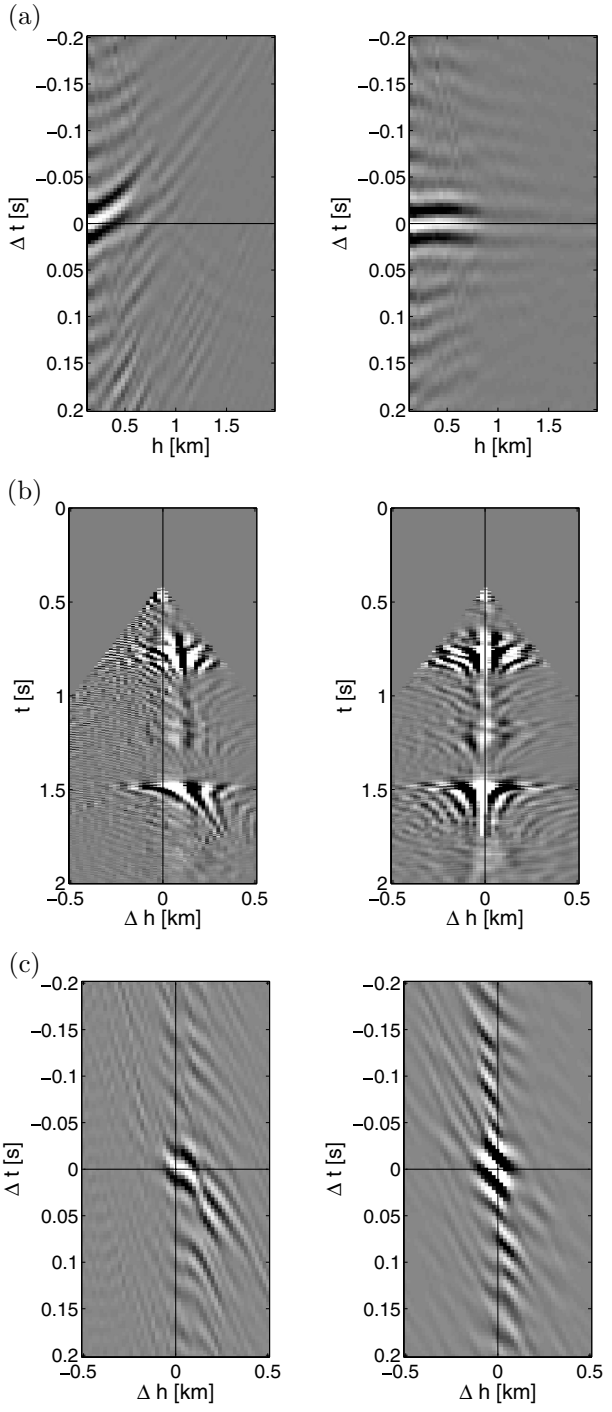

Figure 13 Data correlation for the real data example from Fig. 12. The temporal (a), spatial (b) and space-time correlations (c) are depicted for the initial (left) and final (right) velocity models. Note that although we used the functional based on the spatial correlation, the final velocity model focuses all three.

a width parameter to suppress this cross-talk. From the plots of the data-focusing functionals it seems that the time correlation suffers most from the cross-talk. The Gaussian weighting function deals with the cross-talk successfully. Compared to the least-squares functional, most of the weighted norms have a large basin of attraction and respond smoothly to a perturbation of the true velocity model.

We showed that under the assumptions of the convolutional model, there is a strong connection between the functionals based on the space-time correlation in the depth and the data-domain, the main difference being the order in which the weighting and integration over depth is carried out. In the data

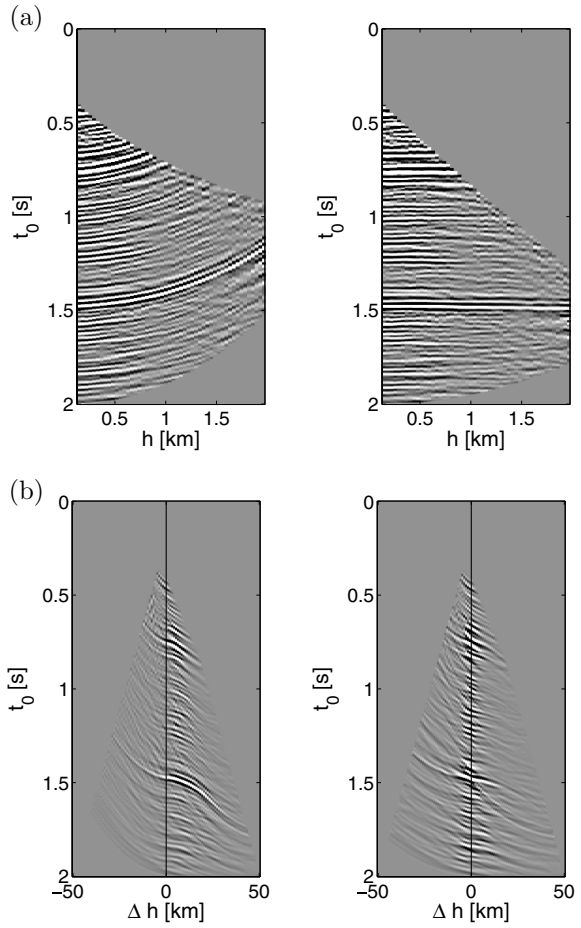

Figure 14 Focusing in the image-domain for the real data example from Fig. 12. (a) Shows the NMO corrected data for the initial (left) and final (right) velocity models. The events are flattened by the final velocity model, which indicates that this is indeed the correct NMO velocity. In (b) the generalized image is shown for $\Delta t = 0$, again for the initial (left) and the final (right) velocity models. The final velocity model focuses the generalized image. This suggests an equivalence between the focusing in the data- and image-domains.

focusing approach the weighting is applied to the squared sum of the images, while in the data focusing approach the weighting is applied to the sum of the squared images. Overall, the data and image-domain functionals respond similarly to a perturbation.

To test the functionals proposed here, we used a gradient-based optimization scheme to obtain the true velocity model, starting from a constant initial model. We have shown only a few experiments but all the functionals yielded good results. The Gaussian weighting function proved especially successful, whereas the quadratic weighting function showed more sensitivity to the width parameter.

The real data example showed that the velocity model resulting from optimizing the data-focusing also focused the generalized image and flattened the events after NMO correction.

It remains to be seen if the method can be successfully extended beyond the convolutional model for 1D models. The next step would be to test the proposed functionals in a full waveform setting. A point of special interest is whether this method would succeed when multiples are present because this is where the image-domain methods tend to break down.

ACKNOWLEDGEMENTS

This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM), which is financially supported by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). The authors thank NAM (Nederlands Aardolie Maatschappij) for providing the real data.

REFERENCES

Brown M.P. and Guitton A. 2005. Least-squares joint imaging of multiples and primaries. *Geophysics* **70**, S79–S89.

Chavent G. and Clément F. 1993. Waveform inversion through MBTT formulation. Technical Report 1893 INRIA.

Clément F., Chavent G. and Gómez S. 2001. Migration-based traveltime waveform inversion of 2-D simple structures: A synthetic example. *Geophysics* **66**, 845–860.

Doherty S.M. and Claerbout J.F. 1974. Velocity analysis based on the wave equation. Technical Report 1, Stanford Exploration Project. http://sepwww.stanford.edu/oldreports/sep01/01_12.pdf.

Dussaud E.A. and Symes W.W. 2005. Velocity analysis from interferometric data. 75th SEG meeting, Houston, Texax, USA, Expanded Abstracts, pp. 2237–2241.

Jiang Z., Sheng J., Yu J., Schuster G.T. and Hornby B.E. 2007. Migration methods for imaging different-order multiples. *Geophysical Prospecting* **51**, 1–19.

Lailly P. 1983. The seismic inverse problem as a sequence of before stack migrations. In: *Conference on Inverse Scattering: Theory and Application* (eds J.B. Bednar, R. Redner, E. Robinson and A. Weglein). Society for Industrial and Applied Mathematics, Philadelphia.

Li J. and Symes W.W. 2007. Interval velocity estimation via NMO-based differential semblance. *Geophysics* **72**, U75–U88.

Luo Y. and Schuster G.T. 1989. Wave-equation traveltime inversion. *Geophysics* **56**, 645–653.

MacKay S. and Abma R. 1992. Imaging and velocity analysis with depth-focussing analysis. *Geophysics* **57**, 1608–1622.

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

Mulder W.A. and ten Kroode A.P.E. 2002. Automatic velocity analysis by differential semblance optimization. *Geophysics* **67**, 1184–1191.

Plessix R.-E. 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International* **167**, 495–503.

Plessix R.-E., Chavent G. and Roeck Y.-H. De 1999. Waveform inversion of reflection seismic data for kinematic parameters by local inversion. *SIAM Journal of Scientific Computing* **20**, 1033–1052.

Pratt R.G., Song Z.M., Williamson P. and Warner M. 1996. Two-dimensional velocity models from wide-angle seismic data by wavefield inversion. *Geophysical Journal International* **124**, 232–340.

Pratt R.G. and Hicks G.J. 1998. Gauss-Newton and full-Newton methods in frequency space seismic waveform inversion. *Geophysical Journal International* **133**, 341–362.

Rickett J.E. and Sava P.C. 2002. Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics* **67**, 883–889.

Sava P.C. and Biondi B. 2004. Wave-equation migration velocity analysis. I. Theory. *Geophysical Prospecting* **52**, 593–606.

Sava P.C. and Fomel S. 2006. Time-shift imaging condition in seismic migration. *Geophysics* **71**, S209–S217.

Shen P., Symes W.W. and Stolk C.C. 2003. Differential semblance velocity analysis by wave-equation migration. 73th SEG meeting, Dallas, Texas, USA, Expanded Abstracts, pp. 2132–2135.

Stork C. 1992. Reflection tomography in the postmigrated domain. *Geophysics* **57**, 680–692.

Symes W.W. 1999. All stationary points of differential semblance are global minimizers: Layered acoustics. Technical Report 100, Stanford Exlorayion Project.

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

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

Tarantola A. and Valette A. 1982. Generalized nonlinear inverse problems solved using the least squares criterion. *Reviews of Geophysics and Space Physics* **20**, 129–232.

Vogel C. R. 2002. *Computational Methods for Inverse Problems*, vol. 23. SIAM, Philadelphia.

Yilmaz O. and Chambers R. 1984. Migration velocity analysis by wavefield extrapolation. *Geophysics* **32**, 1664–1674.

Youn O.K. and Zhou H. 2001. Depth imaging with multiples. *Geophysics* **66**, 246–255.

APPENDIX: GRADIENT

We use the exact gradients for the optimization. We explicitly derive the gradient for the functional based on the temporal correlation using the adjoint-state technique (see Plessix 2006, for example). The gradients for the spatial and space-time correlation can be derived in the same way. Some of the arguments of the functions have been omitted for notational clarity.

We represent the velocity model with parameters $\{v_0, v_1, \ldots, v_{N-1}\}$ on an equidistant grid $\{t_0, t_1, \ldots, t_{N-1}\}$ of spline nodes. Given these parameters the NMO velocity is represented by

$$v_{\text{nmo}}(t) = \sum_{i=0}^{N-1} v_i \psi(t - t_i), \tag{A1}$$

where $\psi(t)$ is a cubic spline. We use the convolutional model to generate the data, so the simulated seismogram can be written as

$$\begin{aligned} p(t,h) &= \int w(t-t')r\Big(\tau_0(t',h)\Big)\,\mathrm{d}t' \\ &= \int w(t-\tau(t_0,h))r(t_0)s(t_0,h)\,\mathrm{d}t_0, \end{aligned} \tag{A2}$$

where $w(t)$ is the wavelet, $r(t)$ is the reflectivity, in our case a spike train and $\tau_0(t, h)$ is the inverse of the traveltime:

$$\tau(t_0,h) = \sqrt{t_0 + \Big(h/v_{\text{nmo}}(t)\Big)^2}. \tag{A3}$$

Here h denotes offset and t_0 is depth measured in vertical two-way traveltime. We assume that the stretch factor $s(t_0, h) = \frac{\partial \tau}{\partial t_0}(t_0, h)$ can be neglected for the rest of the derivation. In fact, the convolutional model is only valid for $s(t_0, h) \approx 1$ (Symes 1999). Typically, offsets larger than $1000t$ are blanked out. The functional is given by:

$$\begin{aligned} J_t = \frac{1}{S_t} \iint W(\Delta t)\Big(C_t[p,q](\Delta t,h)\Big)^2\,\mathrm{d}\Delta t\,\mathrm{d}h \\ +\beta \sum_i (v_{i+1}-v_i)^2, \end{aligned} \tag{A4}$$

$$S_t = \iint \Big(C_t[p,q](\Delta t,h)\Big)^2\,\mathrm{d}\Delta\mathrm{d}\,\Delta h t \tag{A5}$$

The second term in the functional is the Tikhonov regularization, which penalizes large variations in the velocity model. The parameter β controls the amount of regularization applied. To avoid a large penalty at the edges of the domain we let $v_{-1} = v_0$ and $v_N = v_{N-1}$.

To find an expression for the gradient $\frac{\partial J_t}{\partial v_i}$, we consider the Lagrangian:

$$\begin{aligned} \mathcal{L} = \frac{1}{S_t}\iint W(\Delta t)\Big(C_t[p,q](\Delta t,h)\Big)^2\,\mathrm{d}\Delta t\,\mathrm{d}h + \beta\sum_i (v_{i+1}-v_i)^2 \\ + \iint \mu_1(\Delta t,h)\Big(C_t[p,q] - \int p(t,h)q(t+\Delta t,h)\mathrm{d}t\Big)\,\mathrm{d}\Delta t\,\mathrm{d}h \\ + \iint \mu_2(t,h)\Big(p(t,h) - \int w\Big(t-\tau(t_0,h)\Big)r(t_0)\,\mathrm{d}t_0\Big)\,\mathrm{d}t\,dh \\ + \iint \mu_3(t_0,h)\Big(\tau(t_0,h) - \sqrt{t_0^2 + \Big(h/v_{\text{nmo}}(t_0)\Big)^2}\Big)\,\mathrm{d}t_0\,\mathrm{d}h \\ + \int \mu_4(t_0)\left(v_{\text{nmo}}(t_0) - \sum_{i=0}^{N-1} v_i\psi(t_0-t_i)\right)\mathrm{d}t_0, \end{aligned} \tag{A6}$$

where $\mu_1 \ldots \mu_4$ are the Lagrange multipliers. The gradients of the Lagrangian with respect to the state variables C_t, p, τ and v_{nmo} are given by:

$$\frac{\partial \mathcal{L}}{\partial C_t} = \frac{2}{S_t}(W - J_t)C_t[p, q](\Delta t, h) + \mu_1(\Delta t, h), \tag{A7}$$

$$\frac{\partial \mathcal{L}}{\partial p} = -\int \mu_1(\Delta t, h)\, p(t + \Delta t, h)\, \mathrm{d}\Delta t + \mu_2(t, h), \tag{A8}$$

$$\frac{\partial \mathcal{L}}{\partial \tau} = \int \mu_2(t, h)\dot{w}\Big(t - \tau(t_0, h)\Big) r(t_0)\, \mathrm{d}t \tag{A9}$$

$$+ \mu_3(t_0, h), \tag{A10}$$

$$\frac{\partial \mathcal{L}}{\partial v_{\mathrm{nmo}}} = \int \mu_3(t_0, h) \frac{h^2 v_{\mathrm{nmo}}^{-3}}{\tau(t_0, h)}\, \mathrm{d}h + \mu_4(t_0). \tag{A11}$$

Here $\dot{w}(t)$ denotes the temporal derivative of $w(t)$.

At a stationary point, $\nabla \mathcal{L} = 0$. Using this, we can eliminate the multipliers by successive substitution. This yields an expression for μ_4 which can be used to determine the gradient of the functional

$$\begin{aligned}\frac{\partial J_t}{\partial v_i} &= \frac{\partial \mathcal{L}}{\partial v_i} - \int \mu_4(t_0)\psi(t_0 - t_i)\, \mathrm{d}t_0 - 2\beta(v_{i+1} - 2v_i + v_{i-1}) \\ &= \frac{2}{S_t}\int \mathrm{d}t_0\, r(t_0)\psi(t_0 - t_i) \\ &\times \left(\iint C_t\,[(W - J_t)C_t[p, q], q]\, \dot{w}\Big((t - \tau(t_0, h)\Big) \frac{h^2 v_{\mathrm{nmo}}^{-3}}{\tau(t_0, h)}\, \mathrm{d}t\, \mathrm{d}h\right)\end{aligned} \tag{A12}$$

To calculate the gradient we: *(i)* perform weighted correlation of the data correlation and the observed data,*(ii)* correlate this with a wavelet (the derivative of the original wavelet), *(iii)* stack the NMO corrected result and *(iv)* project this back onto the spline basis.