

Subsurface offset behaviour in velocity analysis with extended reflectivity images

W.A. Mulder^{1,2}

¹Shell Global Solutions International B.V., PO Box 60, 2280 AB Rijswijk, The Netherlands and ²Delft University of Technology, Department of Geoscience & Engineering, Faculty of Civil Engineering and Geosciences, PO Box 5048, 2600 GA Delft, The Netherlands

Received May 2012, revision accepted March 2013

ABSTRACT

Migration velocity analysis with the constant-density acoustic wave equation can be accomplished by the focusing of extended migration images, obtained by introducing a subsurface shift in the imaging condition. A reflector in a wrong velocity model will show up as a curve in the extended image. In the correct model, it should collapse to a point. The usual approach to obtain a focused image involves a cost functional that penalizes energy in the extended image at non-zero shift. Its minimization by a gradient-based method should then produce the correct velocity model. Here, asymptotic analysis and numerical examples show that this method may be too sensitive to amplitude peaks at large shifts at the wrong depth and to artefacts. A more robust alternative is proposed that can be interpreted as a generalization of stack power and maximizes the energy at zero-subsurface shift. A real-data example is included.

Key words: Velocity analysis, Seismic imaging, FD migration.

1 INTRODUCTION

The exponential growth of seismic data volumes follows that of computing power but at a slower rate. As a result, seismic processing and imaging algorithms allow for increasingly more accurate and realistic modelling of wave propagation. We are seeing a move from ray-based algorithms via one-way wave equations to fully acoustic modelling of wavefields, with elastic propagators as the next step. At the same time, staff levels remain fairly constant over time. This poses two challenges: (1) how can we apply the more realistic wave propagators for seismic processing, velocity model building and imaging and (2) how can these procedures be automated? This paper deals with one aspect, automatic velocity building with a constant-density acoustic ‘two-way’ wave equation.

One approach to migration velocity analysis with a wave equation is to build extended migration images with a subsurface spatial (MacKay and Abma 1992; Rickett and Sava 2002; Biondi and Sava 2004a, b) or temporal shift (Faye and Jeannot 1986; Sava and Fomel 2006; Brown, Higginbotham and Clapp 2008; van Leeuwen and Mulder 2010) and to focus

them at zero shift (Shen, Symes and Stolk 2003; Shen *et al.* 2005; Mulder 2008; Symes 2008; Shen and Symes 2008). The spatial or temporal shift compensates velocity errors. Because reflectors may be observable in the extended image as curved events, even with the wrong velocity, it is easier to improve their focusing than to maximize their stack power. With the latter, if the velocity model is too far away from the correct one, there is nothing that stacks, so nothing to use for velocity updating.

In practice, the focusing of extended images produces results of mixed quality. Sometimes, the method works very well, sometimes it has difficulties. The data should be free of multiples, otherwise, the method may converge to the wrong model (Mulder and van Leeuwen 2008). Point scatterers, diffractions and discontinuities in the background velocity model may cause problems (Vyas, Geco and Tang 2010). Artefacts in extended images may lead gradient-based optimization methods astray.

Asymptotic analysis and numerical tests show that the usual penalization of energy at non-zero shift may give velocity updates that are biased by amplitude peaks at large shifts

and incorrect depths and by artefacts in the gathers. Here, a modification of the focusing cost functional is proposed that improves the robustness of the method. Instead of penalizing energy at non-zero shift, amplitudes at zero shift are maximized.

In the next section, the cost functional for the extended image will be reviewed. Section 3 presents the result of a high-frequency asymptotic analysis applied to the simple case of a horizontal reflector in a constant velocity model. Section 4 compares the result of the asymptotic analysis with numerical examples. To improve the robustness of the method, a modification of the focusing cost functional is proposed. Section 5 contains a real-data example. Additional aspects of the method are discussed in Section 6.

2 COST FUNCTIONAL

Before defining the cost functional for migration velocity analysis based on extended migration images, the equations governing migration will be reviewed. Migration appears naturally in the setting of least-squares data fitting governed by the Born approximation of the wave equation. This approximation turns the non-linear full waveform inversion problem into a linear inverse problem. Migration is then the gradient of the least-squares cost functional with respect to perturbations in a given fixed background velocity model.

Classic migration

In the frequency domain, the constant-density acoustic wave equation is

$$L(m)p_s = f_s, \quad L = -\omega^2 m - \Delta, \quad (1)$$

with pressure $p_s(\mathbf{x}, \omega)$ as a function of position \mathbf{x} and angular frequency ω . The operator L depends on the model parameters $m(\mathbf{x}) = 1/c^2(\mathbf{x})$ with $c(\mathbf{x})$ the velocity. The source term f_s is usually taken as $f_s(\mathbf{x}, \omega) = w(\omega)\delta(\mathbf{x} - \mathbf{x}_s)$ with wavelet $w(\omega)$ and a delta function centred at the shot position \mathbf{x}_s for shot number s . Least-squares fitting of modelled to observed data involves minimization of the cost functional

$$J^{\text{LS}} = \frac{1}{2} \sum_{\omega} \sum_s \sum_{r(s)} \left| S_{r(s)} p_s - p_{r(s)}^{\text{obs}} \right|^2. \quad (2)$$

Here, $p_{r(s)}^{\text{obs}}$ are the observed data for receiver number $r(s)$ and shot s , $S_{r(s)}$ is a sampling or detection operator that reads off the wavefield $p_s(\mathbf{x}, \omega)$ at the receiver position $\mathbf{x}_{r(s)}$ and the summation involves all shots and receivers and relevant frequencies. The gradient of the functional J^{LS} with respect

to the model, $m(\mathbf{x})$, can be expressed as (Virieux and Operto 2009; Hak and Mulder 2011)

$$\nabla_m J = \text{Re} \sum_{\omega} \sum_s \left(\omega^2 p_s \right)^* q_s, \quad (3)$$

where q_s follows from the adjoint or ‘reverse-time’ problem

$$L^H q_s = \sum_{r(s)} S_{r(s)}^T \left[S_{r(s)} p_s - p_{r(s)}^{\text{obs}} \right]. \quad (4)$$

The asterisk denotes the complex conjugate, the superscript $(\cdot)^T$ the transpose and the superscript $(\cdot)^H$ the conjugate transpose.

Migration can be viewed as a minimization problem similar to full-waveform inversion but governed by the Born approximation of the wave equation (Østmo, Mulder and Plessix 2002; Mulder and Plessix 2004). The latter is obtained by splitting the model $m(\mathbf{x})$ into a smooth background, $m_0(\mathbf{x}) = c_0^{-2}(\mathbf{x})$ and an oscillatory part $m_1(\mathbf{x}) = c_1^{-2}(\mathbf{x})$ that represents the reflectivity. The background velocity model should not produce reflections in the seismic frequency band. The wavefield can be split as $p = p_0 + p_1$ just as the model is split into $m = m_0 + m_1$. Dropping the cross-term with $m_1 p_1$ leads to the Born approximation

$$L_0 p_0 = f_s, \quad L_0 p_1 = \omega^2 m_1 p_0, \quad (5)$$

with operator $L_0 = -\omega^2 m_0 - \Delta$ defined by the background model. The data error in the cost functional is replaced by $S_{r(s)} p_{1,s} - p_{1,r(s)}^{\text{obs}}$, where the observed data now should only contain the primary reflections. The gradient with respect to m_1 becomes

$$\nabla_{m_1} J = \text{Re} \sum_{\omega} \sum_s \left(\omega^2 p_{0,s} \right)^* q_{1,s}, \quad (6)$$

with

$$L_0^H q_{1,s} = \sum_{r(s)} S_{r(s)}^T \left[S_{r(s)} p_{1,s} - p_{1,r(s)}^{\text{obs}} \right]. \quad (7)$$

The first iteration with the Born approximation, assuming an initial guess with zero reflectivity, $m_1 = 0$, leads to a right-hand side with $p_{1,s} = 0$.

Migration velocity analysis with extended images

Least-squares fitting or full-waveform inversion suffers from local minima in the cost functional, due to the absence or unreliability of low frequencies in the data. To prevent a gradient-based minimization algorithm from ending up in the wrong minimum, the starting model needs to be accurate within half a wavelength. Migration velocity analysis attempts to construct such a model by exploiting the redundancy in the data,

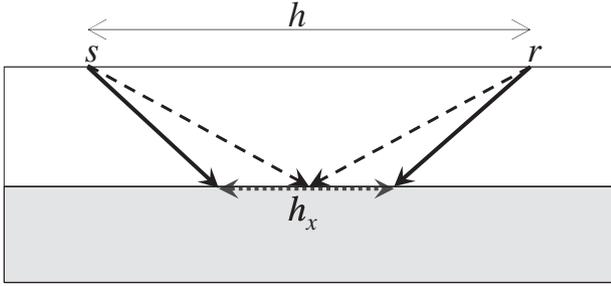


Figure 1 Migration of data with a surface offset h between shot s and receiver r involves the correlation of the forward wavefield from the shot and the time-reversed wavefield from the receiver. The dashed arrows indicate the point where this correlation would result in a peak at the reflector if the velocity model was correct and after summation of many shots and receivers. If the velocity in the upper layer is too small, constructive interference can still occur by following the black drawn lines and making up for lost time with an immediate jump along the dotted line, over a subsurface distance h_x .

requiring that subsurface images do not change for different subsets of the data that illuminate the same structure. With ray-tracing algorithms, the surface offset between sources and receivers, for instance, may be used to form different images for different offsets, which should be the same in the correct velocity model. In a wave-equation method, this would require solving for $q_{r(s)}$ in equation (4) or (7) one receiver at a time, which is far more costly than treating all receivers for one shot at once. However, in this way, we cannot easily use parameters like offset or scattering angle to describe the redundancy in the data.

One alternative is the introduction of a shift at depth (MacKay and Abma 1992; Rickett and Sava 2002; Shen *et al.* 2003; Biondi and Sava 2004a, 2004b), leading to the extended reflectivity

$$R(\mathbf{x}; \mathbf{h}) = \text{Re} \sum_{s, \omega} \omega^2 p_s^* \left(\mathbf{x} - \frac{1}{2} \mathbf{h}, \omega \right) q_s \left(\mathbf{x} + \frac{1}{2} \mathbf{h}, \omega \right). \quad (8)$$

Compared to equations (6) and (7), the subscripts $_0$ and $_1$ have been dropped. The wavefield $p_s(\mathbf{x}, \omega)$ is generated for a shot with index s , whereas $q_s(\mathbf{x}, \omega)$ is the reverse-time wavefield due to the observed data at the receivers that correspond to this source, as in equation (4) or (7) with modelled data $p_s = 0$. To avoid a data explosion, we can restrict the choice of \mathbf{h} to one of the coordinate directions. The subsurface shift can be interpreted as an action at distance, as sketched in Fig. 1. If the velocity model is incorrect, the reflectors will be mapped to the wrong subsurface location and will not focus. The spatial shift compensates velocity errors and may produce a reflector

at non-zero shift. Defining a cost functional that moves these events to zero shift should enable automatic velocity analysis.

A cost function for migration velocity analysis with subsurface shifts in two-space dimensions can be taken as the weighted sum $J_x + \gamma J_z$ of

$$J_x = \frac{1}{2} \sum_{x, z, b_x} W(x, z) \eta(b_x) |\Xi_x R(x, z; b_x, 0)|^2, \quad (9)$$

and

$$J_z = \frac{1}{2} \sum_{x, z, b_z} W(x, z) \eta(b_z) |\Xi_z R(x, z; 0, b_z)|^2. \quad (10)$$

The functional J_z may be necessary in the presence of steep dips (Biondi and Shan 2002). The filters Ξ_x and Ξ_z should remove long-wavelength components and events with large or small dip, respectively, as proposed earlier (Mulder 2008). The long-wavelength migration artefacts can be removed by a spatial low-cut filter (Mulder and Plessix 2004). A dip filter can easily be constructed in 2D with a spatial FFT in x and z on the extended images $R(x, z; b_x, 0)$ and $R(x, z; 0, b_z)$. If the wavenumber domain has wavenumbers k_x and k_z , then the Fourier symbol of Ξ_x can be taken as $\hat{\Xi}_x(k_x, k_z) = \sqrt{k_z^2 / (k_x^2 + k_z^2)}$ and for Ξ_z as $\hat{\Xi}_z(k_x, k_z) = \sqrt{k_x^2 / (k_x^2 + k_z^2)}$. Weibull (2011) introduced a simpler but possibly more noise-sensitive filter, involving a first derivative in the orthogonal direction, so $\Xi_x = \partial / \partial z$ and $\Xi_z = \partial / \partial x$.

To update a velocity model in such a way that the migration image is focused for zero shift, a common choice is to penalize larger shifts by letting $\eta(b_x) = b_x^2$ and $\eta(b_z) = b_z^2$ (Shen *et al.* 2003, 2005; Symes 2008; Mulder 2008; Shen and Symes 2008). The cost functional is then minimized by a gradient-based optimization algorithm. In this paper, an alternative weighting function will be proposed.

A spatial weighting function $W(x, z)$ can be used to balance the amplitudes of shallower and deeper reflectors. In the real-data example, it will be taken as a simple depth weighting of the form $W(x, z) = \max(0, z - z_{\min})^p$, with a power $p > 0$, rather than more complicated true-amplitude migration weights based on an approximation of the Hessian (Shin, Jang and Min 2001; Plessix and Mulder 2004).

The gradient of the cost functional with respect to the model is given in Appendix B.

3 ASYMPTOTIC ANALYSIS

To study the behaviour of the extended reflectivity, a single horizontal reflector in a constant velocity model will be considered for a marine acquisition with positive surface

offsets in the high-frequency limit using the stationary phase approximation. A derivation is given in Appendix A and the main results are summarized next.

For a delta-function reflector at depth z_0 , the observed data are $p^{\text{obs}}(\bar{h}, \omega) = A_0 \exp(2ik_0 r_0)$, where $r_0 = (\bar{h}^2 + z_0^2)^{1/2}$, with amplitude $A_0^{2\text{D}} = \frac{(i-1)}{8z_0} \sqrt{\frac{\omega c_0^3 r_0}{2\pi}}$ in 2D or $A_0^{3\text{D}} = \frac{i\omega c_0}{16\pi z_0}$ in 3D as a function of surface half-offset $\bar{h} = \frac{1}{2}h$ for a source and receiver at zero depth and for a constant velocity c_0 and wavenumber $k_0 = \omega/c_0$. Note that in 2D, the amplitude increases with offset.

If we consider only a horizontal subsurface shift $h_x = 2\bar{h}_x$, the stationary phase approximation shows that the extended reflectivity in a model with constant velocity $c \neq c_0$ follows the curve

$$z = z_{c,0} = \gamma \sqrt{z_0^2 - \bar{h}_x^2/\beta}, \quad \bar{h}_x = -\beta \bar{h}, \quad (11)$$

where $\gamma = c/c_0$ and $\beta = \gamma^2 - 1 > -1$. If $\beta > 0$, the curve truncates at $z_0 = \bar{h}\sqrt{\beta}$ or $\bar{h}_x = -z_0\sqrt{\beta}$, assuming $\bar{h} > 0$. Note that the surface and subsurface offsets are related via β , which depends on the velocity error.

The amplitude along curve (11) is

$$A^{2\text{D}} = \frac{\gamma^5}{128z_0 z_{c,0}^2} \sqrt{\frac{2\pi c_0^2 r_{c,0}^5}{|\beta|}}, \quad (12)$$

when using the 2D Green function, whereas in the 3D case, we have

$$A^{3\text{D}} = \frac{(\gamma c_0)^3 r_{c,0}}{512\pi z_0 z_{c,0}^2} \sqrt{\frac{2}{|\beta|}}. \quad (13)$$

Here, $r_{c,0} = \sqrt{z_0^2 + (\bar{h}_x/\beta)^2}$.

This stationary-phase estimate assumes an infinite acquisition. Truncation at the shortest or longest offset yields a curve

$$z = \sqrt{\gamma^2(\bar{h}_m^2 + z_0^2) - (\bar{h}_x - \bar{h}_m)^2}, \quad (14)$$

where \bar{h}_m denotes either the shortest or longest surface half-offset.

4 EXAMPLES

Extended image for a horizontal reflector

Figure 2 shows the curves (a) and amplitude (b) along the main curve for a velocity of 1450 m/s using data obtained for a velocity of 1500 m/s and a horizontal reflector at 750 m depth in the 2D case. Surface offsets ranged from 10 to 4000 m with a 10-m interval in a marine-type acquisition (positive offsets to the right of the source) and shots ran from -3000 to 3000 m at a 25-m interval. The dash-dotted and dashed curves are

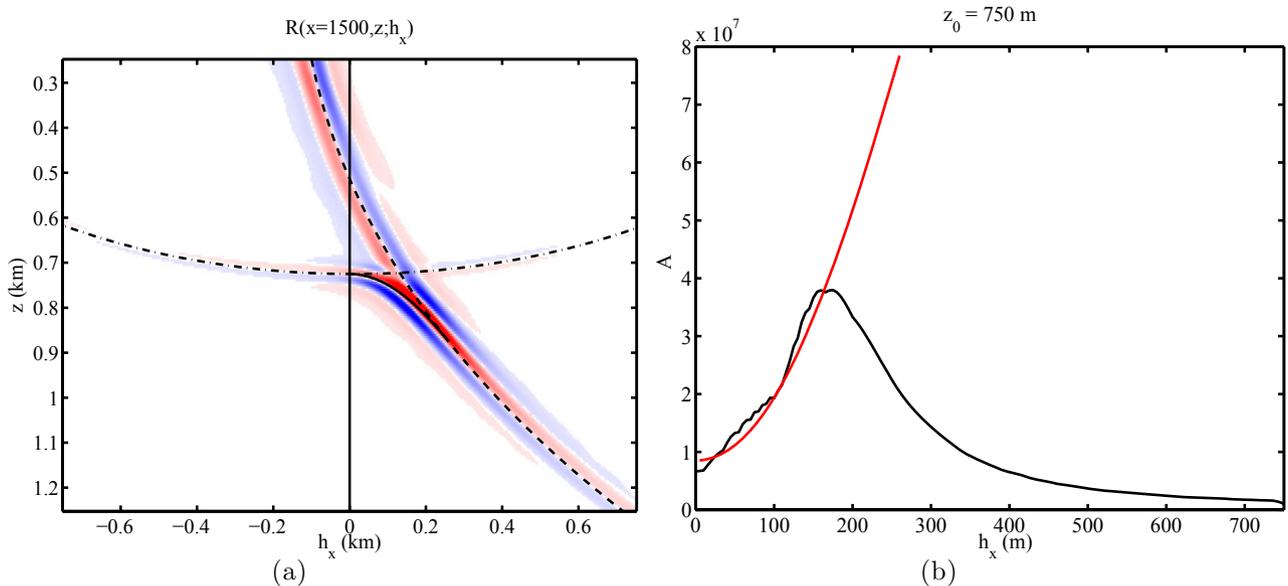


Figure 2 (a) Extended image at a fixed lateral position x as a function of horizontal subsurface offset $h_x = 2\bar{h}_x$ and depth z for too low a velocity of 1450 m/s. The dash-dotted line corresponds to the shortest offset, the dashed line to the longest and the drawn line to the stationary phase curve. The colours show the result of a numerical simulation, with positive values in red and negative values in blue. (b) Observed amplitude behaviour (black), measured in a window around the stationary curve, together with the asymptotic result (red), using arbitrary scaling.

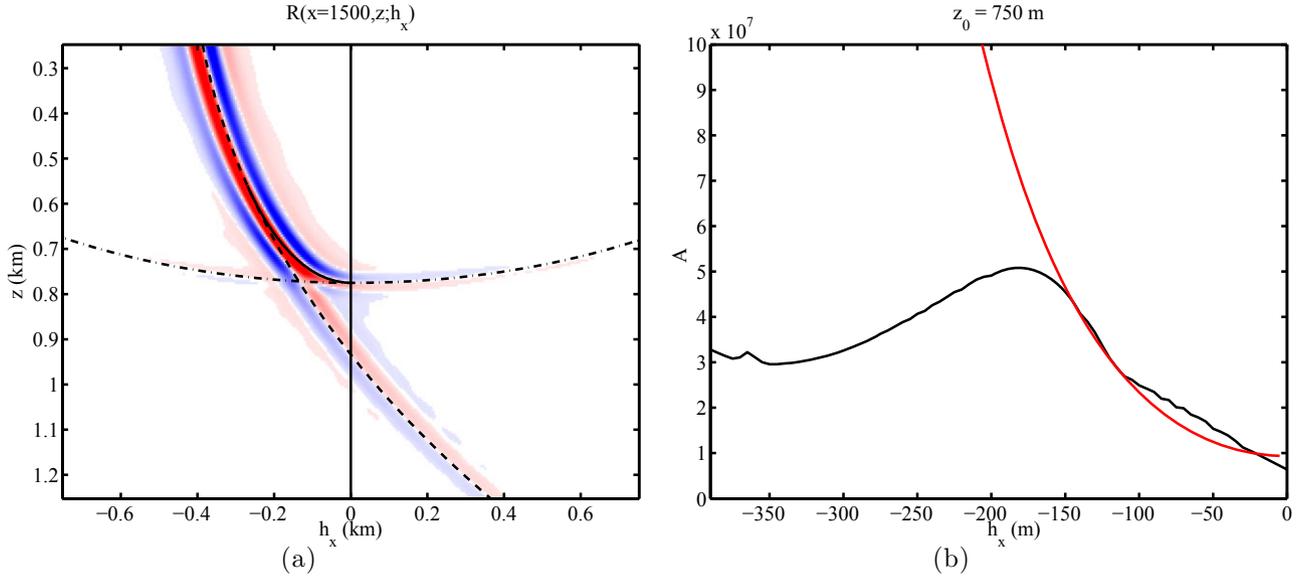


Figure 3 As Fig. 2 but for too high a velocity of 1550 m/s.

caused by truncation of the acquisition at the shortest and longest offsets, respectively, as described by equation (14).

Figure 3 shows similar results for a velocity of 1550 m/s. Again, the curve for the predicted amplitude is truncated at the subsurface shift that corresponds to the maximum surface offset. The amplitude in Fig. 3(b) increases with subsurface offset until it truncates at the point where the maximum surface offset is reached, at $h_x = -217$ m and $z = 558$ m, obtained from equation (14) with $z = z_{c,0}$. The curve obtained from the numerical simulation continues beyond with a smaller amplitude, due to the truncation of the surface offset range and eventually will vanish. Note that the truncation point given by $\bar{h}_x = -z_0\sqrt{\beta}$ would occur further away at $h_x = -391$ m. The differences between the numerical and asymptotic results near the minimum and maximum offsets are most likely caused by the finite ranges for frequency, shot and offset. The truncation of the stationary phase integral provides additional contributions at the endpoints that were not included in the asymptotic estimates.

At the correct velocity of 1500 m/s, the curve $z = z_c$ collapses to a single point at a zero shift, $h_x = 0$ and at the correct depth of the reflector. Only the endpoint contributions of the truncated acquisition remain, as can be seen in Fig. 4(a). Their effect can be mitigated by a suitable offset taper that smoothly decreases the data amplitudes towards the endpoints of the seismic line and the shot range, as illustrated in Fig. 4(b). In this case, the recorded data were multiplied by the window function $w_r = [4\xi(1 - \xi)]^{p_w}$, with $\xi = j/(n_r + 1)$ for $j = 1, \dots, n_r$ corresponding to offsets $h_j = h_{\min} + (j - 1)\Delta h$,

with $h_{\min} = \Delta h = 10$ m. For the power, we used $p_w = 1$. Also, we used a shot taper 10% of the full shot range at the beginning and at the end.

These figures demonstrate that the peak amplitude occurs away from the zero-subsurface offset in the wrong velocity model and at a depth that can be far away from the true depth z_0 . Penalizing this peak by a weighting function η equal to

$$\eta_1(h_x) = h_x^2, \quad (15)$$

will give a contribution to the gradient that should update the velocity model at a depth that may be significantly different from where an update is desired. Figure 5 provides an illustration for a velocity of 1350 m/s, which is too low. The extended image in Fig. 5(a) at a lateral position of $x = 0$ extends to depths beyond the true horizontal reflector depth of 750 m. In this example, the recorded amplitudes at the receivers were tapered towards the beginning and end of the receiver line as well as the ends of the shot range. Also, the dip filter, Ξ_x , was applied. Figure 5(b) displays the contribution to the slowness gradient, with red for positive and blue for negative values, when the extended image is only computed on the vertical line at $x = 0$. The downward curving event in Fig. 5(a) shows up symmetrically as the starting point of the gradient contribution towards zero depth. It extends well beyond the true reflector depth. In the general case with many reflectors, it could potentially harm the deeper layers, particularly if the reflector amplitude would be much higher than that of the deeper ones. Adding the contributions from

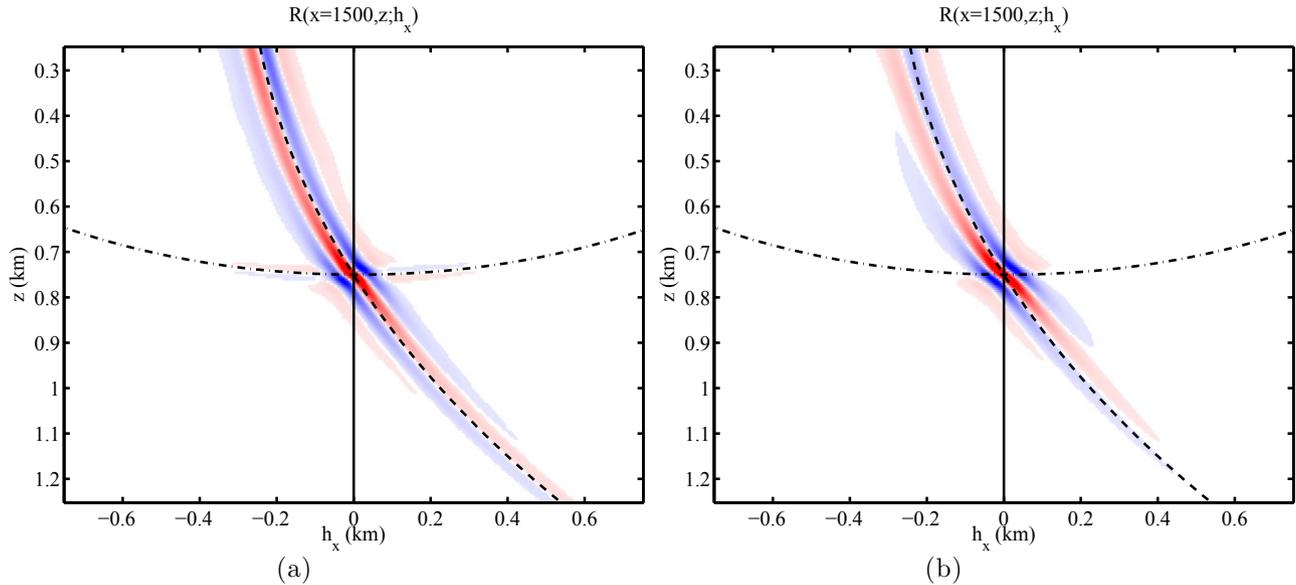


Figure 4 As Fig. 2 but for the correct velocity of 1500 m/s. Only the extended image is shown without (a) and with (b) tapers.

all lateral positions provides the full gradient, which should be translation invariant in the lateral direction for an infinite shot range. Fig. 5(c) depicts a vertical line of this gradient, mapped to a spline grid with a 50-m vertical spacing. The part that is shallower than the reflector will update the velocity in the correct direction but there is a significant contribution at larger depths as well. This effect would clearly be less severe if the amplitudes in the extended image would not increase with the surface offset. Therefore, an alternative weighting function is proposed that is less sensitive to large amplitudes far away. Instead of minimization, the cost function is maximized by

$$\eta_2(b_x) = 1/\left[1 + (b_x/\ell_x)^2\right]^p, \quad (16)$$

with, for instance, $p = 1$ or $p = 2$ and a length scale ℓ_x . The power p should be large enough to reduce the sensitivity to amplitudes and artefacts at large b_x but not so large that the convexity of the functional is destroyed. With this new functional, Fig. 5(d), though a bit noisier, has smaller amplitudes at larger depths. As a consequence, the full gradient in Fig. 5(e) shows a better localization around the true reflector depth.

In the limit for small ℓ_x , the weighting function, $\eta_2(b_x)$, decreases rapidly with b_x and the cost function only measures energy close to zero-subsurface shift. In this case, it starts to resemble a stack or imaging power maximization (Toldi 1989; Soubaras and Gratacos 2007; van Leeuwen and Mulder 2010). The need for including stack power as an additional term or factor was already noted earlier (Mulder 2008; Shen

and Symes 2008). Note that stack power by itself is computationally far less expensive than working with extended images.

Velocity scan

To test the quality of the cost functional, a scan over a set of constant-velocity models was performed, using the same parameters as before and including source and receiver tapers. The reference velocity is $c_0 = 1500$ m/s.

Figure 6(a) depicts the peak amplitude of $R(x, z; b_x, 0)$ as a function of the velocity error $c/c_0 - 1$. Note the decrease of the amplitude followed by an increase for increasing positive velocity errors. Figure 6(b) displays scans of the cost functional J_1 , being J_x in equation (9) to be minimized with $\eta(b_x) = \eta_1(b_x) = b_x^2$. The new functional requires maximization of J_x with the proposed penalization function $\eta = \eta_2$ and will be denoted by J_2 .

Note that J_1 is more convex but does not have its minimum at the correct model. The same is true for J_2 but the discrepancy is much smaller. Given the asymmetry in the extended image for a one-sided acquisition, this was to be expected. Less trivial geological structures will also cause asymmetry.

The convexity of J_2 can be controlled by ℓ_x . Figure 7 shows examples for different choices of the parameters ℓ_x and p in the penalization function in equation (16). The effective penalizing weighting function is controlled both by the length

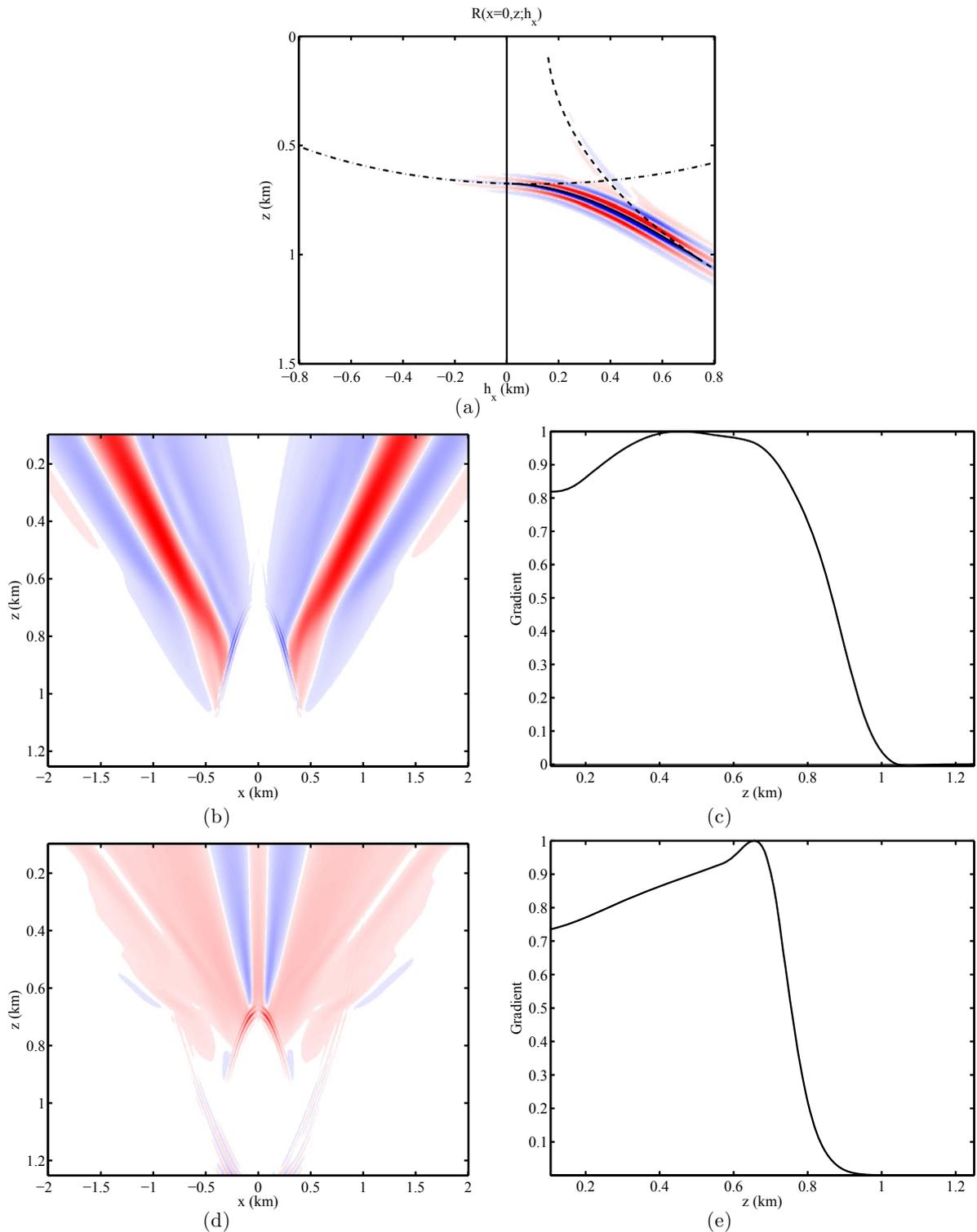


Figure 5 Gradients for too low a velocity of 1350 m/s and a single reflector at 750-m depth. (a) The extended image at the central lateral position, $x = 0$. (b) Contribution to the slowness gradient when only the extended image at $x = 0$ is considered, for the functional J_1 based on $\eta_1(b_x) = b_x^2$. The full gradient is a sum of laterally translated copies of this image. (c) Vertical cross-section of the full slowness gradient at $x = 0$. This gradient will not only update the velocity between the depth of the sources and receivers and the true reflector location but also deeper parts. (d) As for (b) but for the functional J_2 . (e) For $\eta_2(b_x)$, the gradient has less impact on the velocity below the true reflector depth.

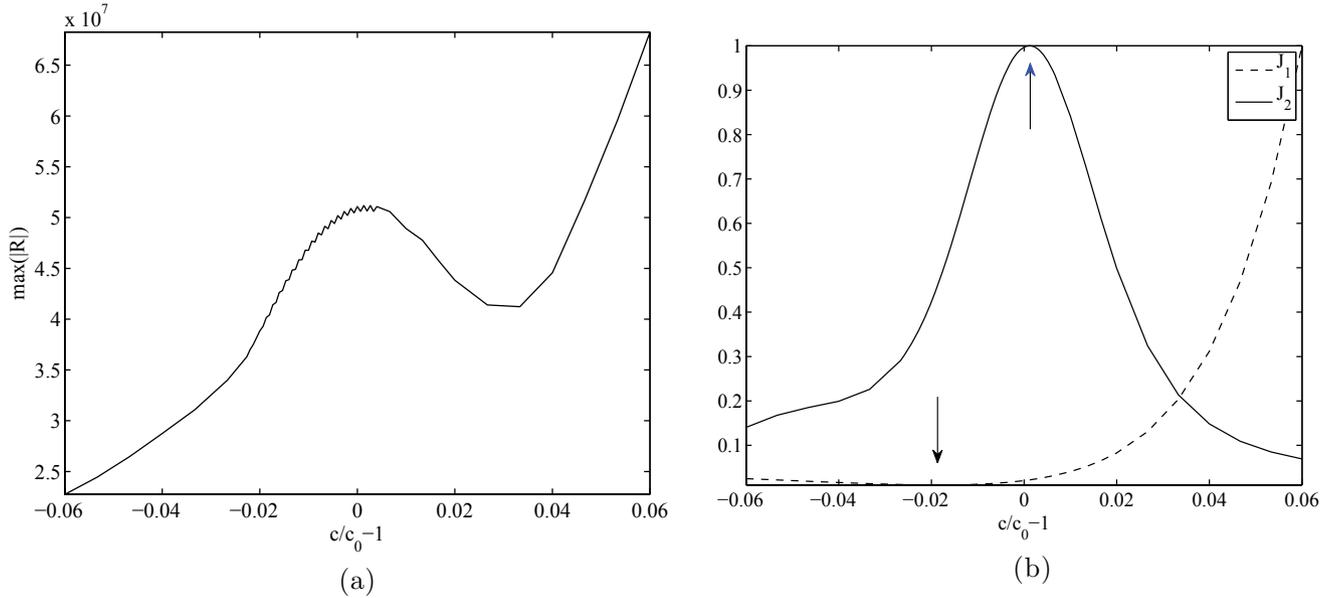


Figure 6 Scan over a set of constant-velocity models. (a) Observed maximum of $|R(x, z; b_x, 0)|$ at $x = 1.5$ km for single-reflector data as a function of the relative velocity error $(c/c_0) - 1$. (b) Cost functionals J_1 , to be minimized with $\eta_1(b_x)$ and J_2 , to be maximized as a function of the velocity error. Note the wrong location of the minimum for J_1 and the maximum of J_2 , indicated by the arrows.

scale ℓ_x and the decay power p . If the function is too narrow, for small ℓ_x or large p , convexity is lost but the maximum is better defined. If the function is too wide, convexity is lost as well and the maximum moves to the wrong position. In general, the closer the initial model is to the true model, the sharper the weighting function can be. At the same time, the maximum value for $|b_x|$ can be decreased.

5 REAL DATA EXAMPLE

The method was applied to a marine seismic line, courtesy of Saga Petroleum A.S., now part of Statoil. The data were acquired above the Haltenbanken terrace, offshore Norway, where the sea bottom is fairly hard. The direct arrival through the seawater as well as multiple reflections were removed with the method described by Verschuur and Berkhout (1997). The coordinates in that paper were changed here, with x replaced by $(17200 - x)$ m, to have positive surface offsets as in the earlier sections.

To obtain an initial guess of the model, a one-dimensional optimization based on values of the functional J_x for the new penalizing function $\eta_2(b_x)$ was carried out over velocity models that are linear in depth, of the form $c(z) = c_0 + \alpha(z - z_0)$, with c_0 set equal to the known water velocity and $z_0 = -100$ m. Also, a single iteration of full-waveform inversion in a model with a constant seawater velocity model was per-

formed. The depth of the sea bottom was picked from the computed gradient and used to insert the seawater velocity and sea bottom into the earlier velocity model that was linear in depth, leaving it unaltered below the sea bottom. Starting from the modified one-parameter result, the velocity model in Fig. 8(a) was obtained after 20 iterations of gradient-based optimization with a limited-memory BFGS method (Byrd *et al.* 1995). Fig. 8(b) displays the corresponding migration image and Fig. 8(c) the amplitudes of the extended image at one lateral position. This computation required the gradient of the functional with respect to the model, listed in Appendix B for completeness. Wavefields were obtained with a 2D frequency-domain code (Mulder and Plessix 2004), using a fourth-order finite-difference discretization (Harari and Turkel 1995) of the constant-density acoustic wave equation. The model is represented by cubic splines on a grid that is coarser than the one used for modelling and imposes some degree of smoothness.

The method produced the subsurface image $R(x, z; 0, 0)$ in Fig. 8(b). The first reflector towards the top of the figure is the sea bottom. The extended image in Fig. 8(c) shows reasonable though not perfect focusing towards zero shift, $b_x = 0$. Around a depth of 700 m, there is significant energy at a subsurface shift b_x around 300 m. This means that, according to equation (11), $\beta < 0$, so the reconstructed velocity is too low. At the same time, at about 900 m, there is another event that

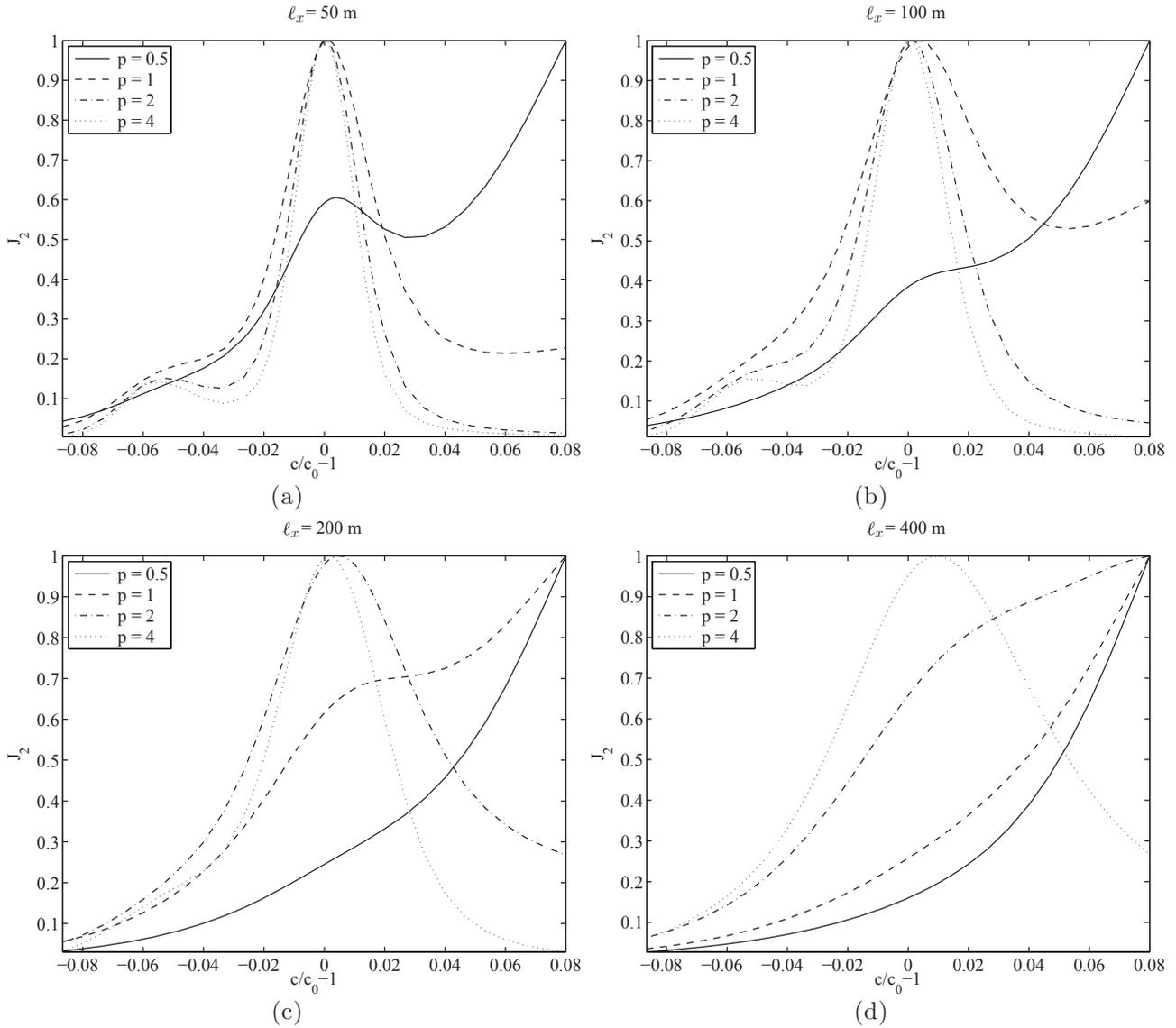


Figure 7 Cost functional J_2 for various parameters when scanning over a set of constant-velocity models. The length scale ℓ_x equals 50, 100, 200, or 400 m, from (a)—(d). The power p is 0.5, 1, 2, or 4 in each figure. Ideally, the curve should peak at zero perturbation and have a convex shape. If the penalizing weighting function is too narrow, for small ℓ_x or large p , convexity is lost. If it is too wide, convexity is lost as well and the peak moves to the wrong location.

is focused. This suggests that a non-negligible amount of multiple energy is still present in the data, despite the applied multiple elimination. Multiples are a well-known problem for migration-based velocity analysis (Mulder and ten Kroode 2002; Li and Symes 2007). Mulder and van Leeuwen (2008) introduced an asymmetry in the penalizing function to bias the optimization towards higher velocities under the assumption that surface multiples tend to lead to a lower velocity. There are, of course, cases where this assumption does not hold, for instance, in the presence of strong inversion layers.

The asymmetric version has a penalizing weighting function of the form $\eta(\min(0, h_x) + b \max(0, h_x))$. In the present setting with positive surface offsets, $b > 1$ will favour the higher velocities.

Figure 9 shows results obtained for a rather large bias, $b = 8$. The velocity model in the shallow part down to a depth of 1.5 km is substantially different and events in the migration image are more clearly defined in that part. Fig. 9(c) shows an asymmetric distribution of energy, with larger amplitudes at negative values of the subsurface shift, h_x .

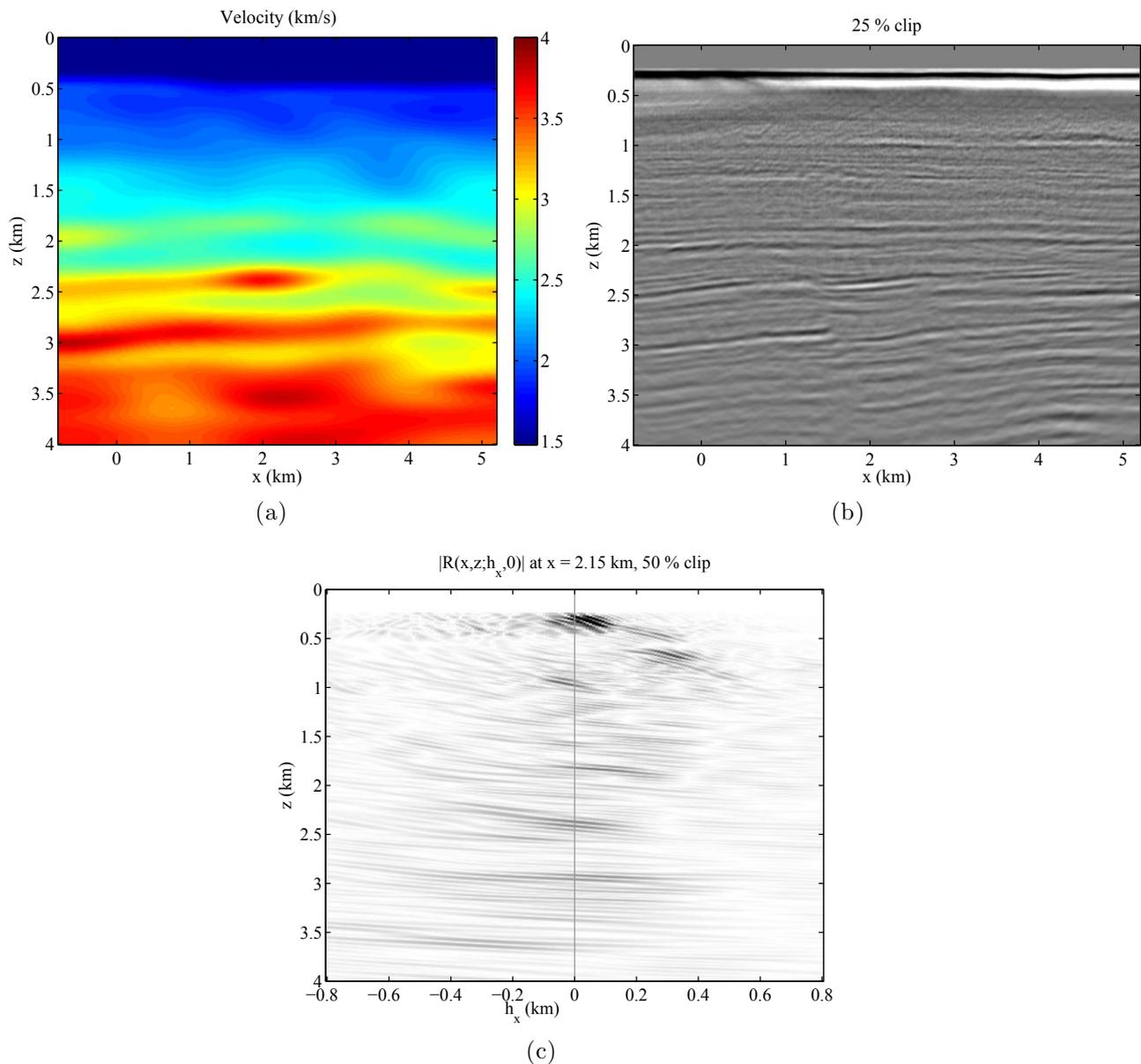


Figure 8 (a) Velocity model after 20 iterations, (b) corresponding migration image and (c) amplitudes of the extended image at $x = 2.15$ km.

Figure 10 provides a comparison between the velocity models without and with use of the bias and includes a velocity model based on a well log. The latter was obtained by scanning Fig. 8.12(A) from the thesis of van Wijngaarden (1998). The velocity model is based on the well log for well B at a lateral position of 17675 m, corresponding to $x = -475$ m after the coordinate transform. The drawn line that represents the reconstructed velocity at the well location with the bias in the inversion algorithm follows the true velocity more closely in the shallower part. The velocity inversion

layer around 2-km depth is missed when the bias towards higher velocities is used but this also happens in the unbiased case. Note that the cubic-spline representation leads to a model that is much smoother than the model based on the well log.

6 REMARKS

This section contains a discussion of various details and aspects of the method.

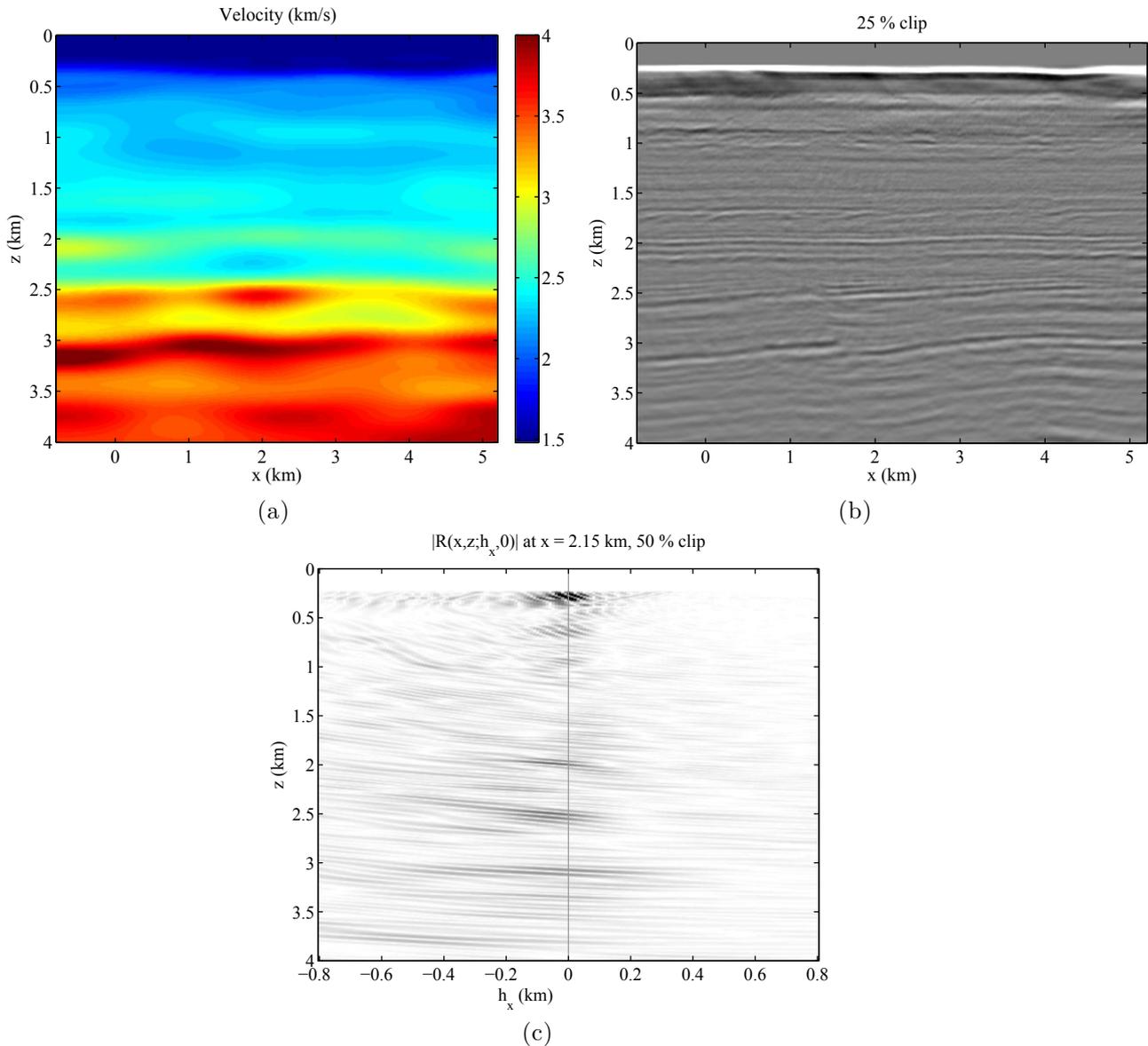


Figure 9 (a) Velocity model after 20 iterations, using a bias towards higher velocities in the cost functional, (b) corresponding migration image and (c) amplitudes of the extended image at $x = 2.15$ km.

The computational cost of forming the extended images by far exceeds that of wavefield computations. To solve the 2D wave equation in the frequency domain, a direct sparse-matrix solver based on nested dissection (George and Liu 1981) was used. Its LU-factorization has a relatively high-computational cost but can be reused for all shots and gradient computations at each frequency. Because the computation of the extended images on the same finite-difference takes much longer, it was accelerated by using a less dense grid. For instance, in the example in Fig. 8 the modelling was carried out on a grid

with a 10-m spacing, the extended image had a spacing of 100 m in x and 10 m in z and the model was represented by cubic splines on a grid with a spacing of 400 m in x and 80 m in z .

The definition of the extended image in equation (8) is based on the real part of the correlated forward and ‘reverse-time’ wavefield. One may ask what would happen if the imaginary part would be included, making $R(\mathbf{x}; \mathbf{h})$ complex-valued. For classic migration, without the extension to non-zero \mathbf{h} , the imaginary part is related to perturbations in the

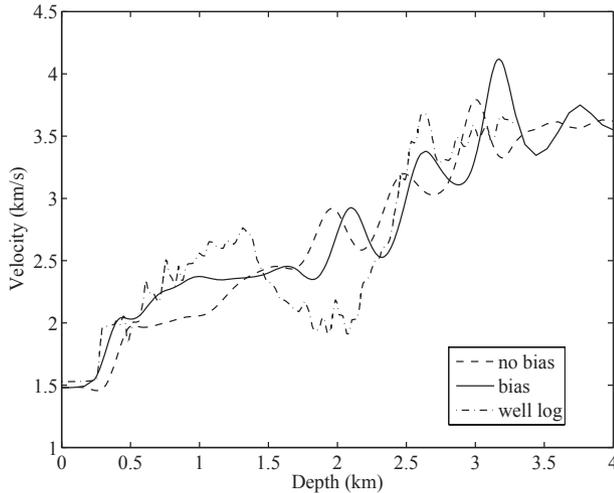


Figure 10 Velocity at a lateral position of -475 m obtained without (dashed line) and with a bias (drawn line) in the penalizing weighting function, compared to a velocity model based on a well log (dot-dashed line).

background attenuation (Mulder and Hak 2009; Hak and Mulder 2010, 2011). Unfortunately, migration tends to produce images of the attenuation that are completely unphysical. Even in the absence of attenuation perturbations of the background model, the imaginary part of the migration image has amplitudes that are comparable to its real part. For a horizontally layered medium, the imaginary part approximately equals a weighted Hilbert transform in depth of the real part (Mulder and Hak 2009), although this property is lost when steep dips are present. As a consequence, a smooth migration image is obtained that resembles the envelope of the true reflectivity model if the absolute value of the migration result is taken for a problem without steep dips. In the present context of migration velocity analysis, including the imaginary part for $R(x, z; h_x, 0)$ will result in smoother pictures for $|R(x, z; h_x, 0)|$. This will affect the reconstructed velocity model on a small-length scale. For instance, instead of having a saw-tooth like pattern of the reflector with a zero-crossing at the position of the impedance contrast, we will obtain a maximum there. With the smooth model representation by cubic splines, however, the effect should not be large. As an illustration, Fig. 11 shows the amplitudes of the extended image at a lateral position of 2.15 km when using the complex-valued extended image during the optimization. Otherwise, the parameters were the same as for Fig. 9. Figure 11(c) is now much smoother than Fig. 9(c). Nevertheless, the reconstructed velocity model and the corresponding migration image, based on only the real part, are similar.

Instead of spatial shifts, a temporal shift h_t can be applied, as already mentioned in the introduction, leading to an extended image

$$R(\mathbf{x}; h_t) = \sum_{s, \omega} \omega^2 p_s^*(\mathbf{x}, \omega) q_s(\mathbf{x}, \omega) e^{-i\omega h_t}, \quad (17)$$

with or without taking the real part. The corresponding functional is

$$J_t = \frac{1}{2} \sum_{x, z, h_t} W(x, z) \eta(h_t) |\Xi_t R(x, z; h_t)|^2, \quad (18)$$

and can be maximized with a similar weighting function $\eta(h_t) = \eta_2(h_t)$ as before. The stationary-phase analysis for the time-shifted case was not carried out and may suggest better weighting functions but the general idea of having a function that decays with $|h_t|$ remains. The filter Ξ_t should remove long-wavelength migration artefacts. Forming one time-shifted extended image requires less computational time than two or even one space-shifted image.

Methods for migration velocity analysis are sensitive to multiples (Mulder and ten Kroode 2002; Verm and Symes 2006; Mulder and van Leeuwen 2008). Multiple suppression is therefore important and should be done thoroughly in some cases. A data-domain approach with shifts in time or offset may be a viable alternative (van Leeuwen and Mulder 2008), although this approach still has not been matured to the level of the full-acoustic wave equation.

From a mathematical point of view, the cost functional has some less desirable properties. In the limit of an infinitely wide frequency band, the stationary-phase curve collapses to a point when the extended image is focused, so there is a singular limit. A similar behaviour occurs with time shifts, where two curves collapse to one, not considering the additional curves due to finite-acquisition effects. The time-shifted images are less singular than the space-shift extended images. The remaining curve in a time-shifted image, which is a straight line in a constant-velocity model, has an amplitude peak at zero shift for the correct model. All this leads to quite singular behaviour, which also shows up in the gradients that update the velocity model. With finite frequencies and proper scale separation (Claerbout 1985, Fig. 1.4-3) and a sufficiently smooth velocity model representation, this singular behaviour can be dealt with. However, given the problems reported by Fei and Williamson (2010) and Vyas, Geco and Tang (2010) in the presence of discontinuities in the velocity model, there seems to be room for improvement.

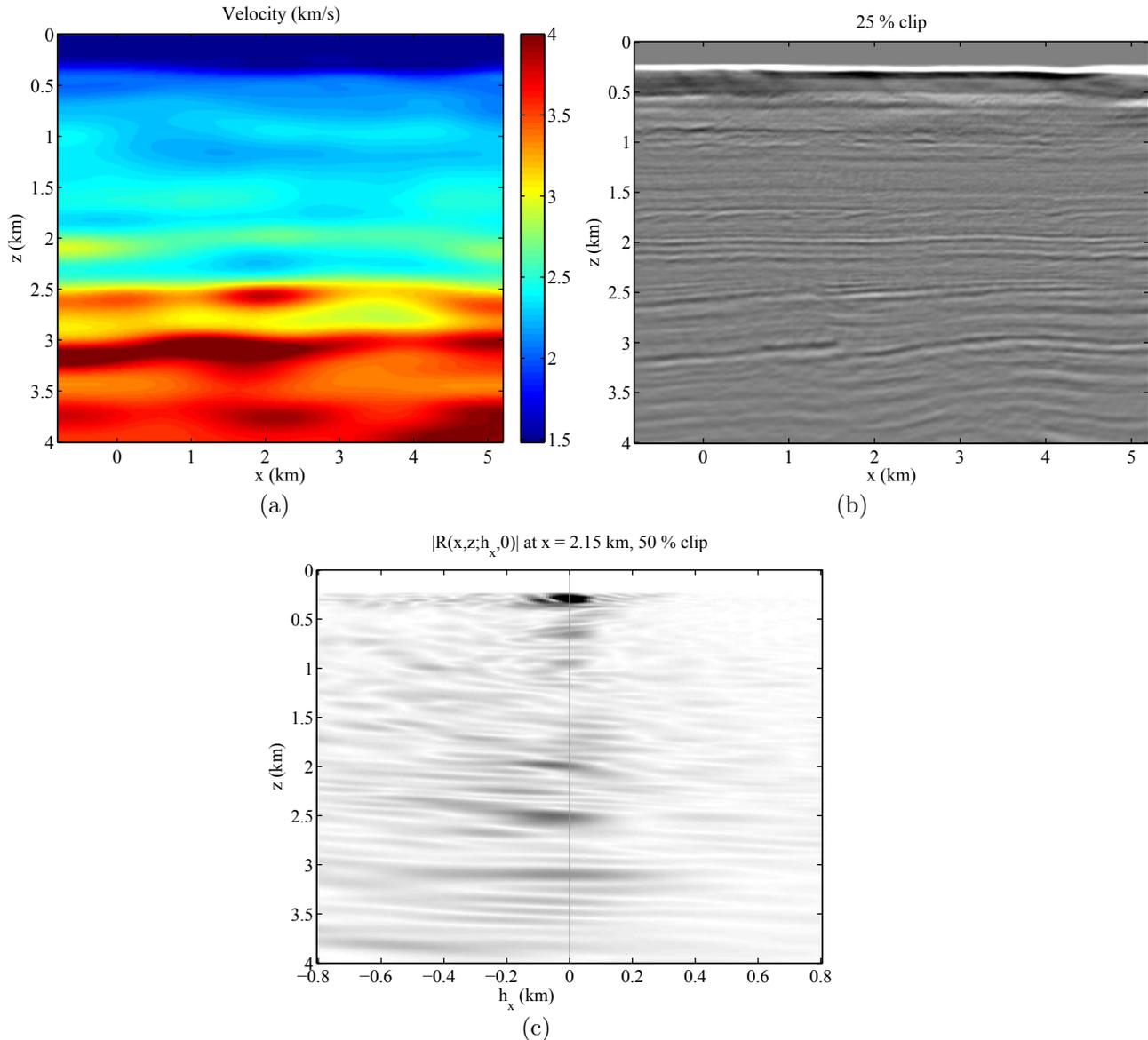


Figure 11 As Fig. 9 but including the imaginary part of the extended image.

7 CONCLUSION

Automatic migration velocity analysis with extended images based on a subsurface shift can be made more robust by maximizing the energy around zero shift rather than minimizing energy at non-zero shift. The approach generalizes stack power maximization.

The analysis of amplitudes in extended migration images reveals an increase of amplitudes with subsurface offset along a curve of varying depth if the velocity model is not correct. Penalizing these peaks at non-zero subsurface shift may result

in a velocity correction at the wrong depth, possibly harming reflectors that are already focused. Maximization of amplitudes around zero shift is less sensitive to this effect, because energy at larger shifts receives less weight.

A 2D marine real-data example confirmed the usefulness of the approach. Although the data were preprocessed to suppress multiples, the velocity model after inversion still was affected by the remaining energy of multiples. A bias towards higher velocities, imposed by an asymmetric penalizing weighting function, improved the result.

ACKNOWLEDGEMENTS

The author is indebted to Eric Verschuur for providing the 2D marine data after processing and multiple suppression.

REFERENCES

- Biondi B. and Sava P. 2004a. Wave-equation migration velocity analysis. I. Theory. *Geophysical Prospecting* 52(6), 593–606.
- Biondi B. and Sava P. 2004b. Wave-equation migration velocity analysis. II. Subsalt imaging examples. *Geophysical Prospecting* 52(6), 607–623.
- Biondi B. and Shan G. 2002. Prestack imaging of overturned reflections by reverse time migration. *SEG Technical Program Expanded Abstracts* 21(1), 1284–1287.
- Bleistein N. and Handelsman R.A. 1975. *Asymptotic Expansions of Integrals*. Holt, Rinehart and Winston, New York.
- Brown M.P., Higginbotham J.H. and Clapp R.G. 2008. Velocity model building with wave equation migration velocity focusing analysis. *SEG Technical Program Expanded Abstracts* 27(1), 3078–3082.
- Byrd R.H., Lu P., Nocedal J. and Zhu C. 1995. A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16(5), 1190–1208.
- Claerbout J.F. 1985. *Imaging the Earth's Interior*. Blackwell Scientific Publications.
- Faye J. and Jeannot J. 1986. Prestack migration velocities from focusing depth analysis. *SEG Expanded Abstracts* 5, 438–440.
- Fei W. and Williamson P. 2010. On the gradient artifacts in migration velocity analysis based on differential semblance optimization. *SEG Technical Program Expanded Abstracts* 29(1), 4071–4076.
- George A. and Liu J. 1981. *Computer Solution of Large Sparse Positive Definite Systems*. Prentice-Hall, Inc.
- Giles M.B., Duta M.C., Müller J. and Pierce N.A. 2003. Algorithm developments for discrete adjoint methods. *AIAA Journal* 41(2), 198–205.
- Hak B. and Mulder W.A. 2010. Migration for velocity and attenuation perturbations. *Geophysical Prospecting* 58(6), 939–952.
- Hak B. and Mulder W.A. 2011. Seismic attenuation imaging with causality. *Geophysical Journal International* 184(1), 439–451.
- Harari I. and Turkel E. 1995. Accurate finite difference methods for time-harmonic wave propagation. *Journal of Computational Physics* 119(2), 252–270.
- van Leeuwen T. and Mulder W.A. 2008. Velocity analysis with multiples – NMO modeling for layered velocity structures. *SEG Technical Program Expanded Abstracts* 27, 1925–1929.
- van Leeuwen T. and Mulder W.A. 2010. A comparison of seismic velocity inversion methods for layered acoustics. *Inverse Problems* 26(1), 015008.
- Li J. and Symes W.W. 2007. Interval velocity estimation via NMO-based differential semblance. *Geophysics* 72(6), U75–U88.
- Lions J. and Magenes E. 1972. *Nonhomogeneous Boundary Value Problems and Applications*. Springer Verlag, Berlin.
- MacKay S. and Abma R. 1992. Imaging and velocity analysis with depth-focusing analysis. *Geophysics* 57(12), 1608–1622.
- Mulder W.A. 2008. Automatic velocity analysis with the two-way wave equation. *70th EAGE Conference & Exhibition. Extended Abstracts*, P165, European Association of Geoscientists & Engineers.
- Mulder W.A. and Hak B. 2009. An ambiguity in attenuation scattering imaging. *Geophysical Journal International* 178(3), 1614–1624.
- Mulder W.A. and ten Kroode A.P.E. 2002. Automatic velocity analysis by differential semblance optimization. *Geophysics* 67(4), 1184–1191.
- Mulder W.A. and van Leeuwen T. 2008. Automatic migration velocity analysis and multiples. *SEG Technical Program Expanded Abstracts* 27(1), 3128–3132.
- Mulder W.A. and Plessix R.É. 2004. A comparison between one-way and two-way wave-equation migration. *Geophysics* 69(6), 1491–1504.
- Østmo S., Mulder W.A. and Plessix R. 2002. Finite-difference iterative migration by linearized waveform inversion in the frequency domain. *SEG Technical Program Expanded Abstracts* 21, 1384–1387.
- 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(2), 495–503.
- Plessix R.E. and Mulder W.A. 2004. Frequency-domain finite-difference amplitude-preserving migration. *Geophysical Journal International* 157(3), 975–987.
- Rickett J.E. and Sava P.C. 2002. Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics* 67(3), 883–889.
- Sava P. and Fomel S. 2006. Time-shift imaging condition in seismic migration. *Geophysics* 71(6), S200–S217.
- Shen P. and Symes W.W. 2008. Automatic velocity analysis via shot profile migration. *Geophysics* 73(5), VE49–59.
- Shen P., Symes W.W., Morton S., Hess A. and Calandra H. 2005. Differential semblance velocity analysis via shot profile migration. *SEG Technical Program Expanded Abstracts* 24(1), 2249–2252.
- Shen P., Symes W.W. and Stolk C.C. 2003. Differential semblance velocity analysis by wave-equation migration. *SEG Technical Program Expanded Abstracts* 22(1), 2132–2135.
- Shin C., Jang S. and Min D.J. 2001. Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting* 49(5), 592–606.
- Soubaras R. and Gratacos B. 2007. Velocity model building by semblance maximization of modulated-shot gathers. *Geophysics* 72(5), U67–U73.
- Symes W.W. 2008. Migration velocity analysis and waveform inversion. *Geophysical Prospecting* 56(6), 765–790.
- Toldi J.L. 1989. Velocity analysis without picking. *Geophysics* 54(2), 191–199.
- Verm R.W. and Symes W.W. 2006. Practice and pitfalls in NMO-based differential semblance velocity analysis. *SEG Technical Program Expanded Abstracts* 25(1), 2112–2116.
- Verschuur D.J. and Berkhouit A.J. 1997. Estimation of multiple scattering by iterative inversion, Part II: Practical aspects and examples. *Geophysics* 62(5), 1596–1611.

- Virieux J. and Operto S. 2009. An overview of full-waveform inversion in exploration geophysics. *Geophysics* 74(6), WCC1–WCC26.
- Vyas M., Geco W. and Tang Y. 2010. Gradients for wave-equation migration velocity analysis. *SEG Technical Program Expanded Abstracts* 29(1), 4077–4081.
- Weibull W.W. 2011. Automatic migration velocity analysis using reverse time migration. *73rd EAGE Conference & Exhibition. Extended Abstracts*, B012, European Association of Geoscientists & Engineers.
- van Wijngaarden A.J. 1998. Imaging and characterization of angle-dependent seismic reflection data. PhD thesis, Delft University of Technology.
- Wirtinger W. 1927. Zur formalen Theorie der Funktionen von mehr komplexen Veränderlichen. *Mathematische Annalen* 97(1), 357–375.

APPENDIX A: ASYMPTOTIC ANALYSIS

Asymptotic analysis in the high-frequency limit for a constant-velocity model offers some insight into how the reflectors appear in the extended image. Below, different velocities for the data and for the extended migration image will be considered. First, an expression for the data due to a single horizontal reflector in a constant velocity model is derived. Then, the shape of a reflector and its amplitude in the extended image with a different constant-velocity model is determined.

Reflection data for horizontal reflectors

The principle of stationary phase yields the approximation

$$\int_{-\infty}^{\infty} dz a(z) e^{i\omega\phi(z)} \simeq \sqrt{\frac{\pi}{|\omega\phi^{(2)}(z_0)|}} a(z_0) e^{i\omega\phi(z_0)} \times [1 + i \operatorname{sign}(\omega\phi^{(2)}(z_0))], \quad (\text{A1})$$

when ω is large (Bleistein and Handelsman 1975). The phase $\phi(z)$ is expressed as $\phi(z) \simeq \phi(z_0) + \frac{1}{2}(z - z_0)^2 \phi^{(2)}(z_0)$, where the first derivative $\phi^{(1)}(z_0) = 0$ at $z = z_0$. If $\phi^{(1)}$ vanishes at multiple points, $z_{0,j}$, then a summation over j should be performed.

In a constant model with velocity c_0 , the solution of the Born approximation (5) is

$$p_{1,r(s)} = \int_{-\infty}^{\infty} d\mathbf{x} G(\mathbf{x}_s, \mathbf{x}) m_1(\mathbf{x}) G(\mathbf{x}, \mathbf{x}_r), \quad (\text{A2})$$

where $G(\mathbf{x}_a, \mathbf{x}_b)$ denotes the Green function for the wave operator L_0 for a delta-function source at \mathbf{x}_a and a receiver at \mathbf{x}_b

as well as a delta-function wavelet $w(\omega) = 1$. In the 3D case, the Green function is

$$G^{3D}(r) = \frac{1}{4\pi r} e^{ikr}, \quad k = \frac{\omega}{c}, \quad r = \|\mathbf{x}_b - \mathbf{x}_a\|, \quad (\text{A3})$$

with r the distance between \mathbf{x}_a and \mathbf{x}_b . In 2D, we have

$$G^{2D}(r) = \frac{i}{4} H_0^{(1)}(kr) \sim \frac{(1+i)}{4\sqrt{\pi kr}} e^{ikr}, \quad (\text{A4})$$

where the approximation holds for large kr . It is assumed that the reflectors are all horizontal, as may happen in simple sedimentary geology. The source and receiver are placed on the surface, at $z = 0$ and the offset, the horizontal distance between the source and receiver, is $2\bar{h}$. The sources are then located at $\mathbf{x}_s = (x_m - \bar{h}, 0, 0)$ and the receivers at $\mathbf{x}_r = (x_m + \bar{h}, 0, 0)$. We integrate over y for the 3D case, with a stationary point at $y = 0$, and over the midpoints x_m in both the 2D or 3D case, with a stationary point at $x_m = 0$. The stationary-phase result for the scattered field, p_1 in equation (5), becomes either

$$p^{3D}(\bar{h}, \omega) \simeq \frac{i\omega w(\omega) c_0}{16\pi} \int_0^\infty dz z^{-1} m_1(z) e^{2ik_0\bar{r}}, \quad (\text{A5})$$

with

$$\bar{r} = \sqrt{\bar{h}^2 + z^2}, \quad k_0 = \omega/c_0, \quad (\text{A6})$$

or, in the 2D case,

$$p^{2D}(\bar{h}, \omega) \simeq \frac{(i-1)w(\omega)}{8} \sqrt{\frac{\omega c_0^3}{2\pi}} \int_0^\infty dz \frac{\sqrt{\bar{r}}}{z} m_1(z) e^{2ik_0\bar{r}}. \quad (\text{A7})$$

Note that $\omega \geq 0$ and that real-valued data in the time domain are obtained with $p(-\omega, \bar{h}) = p^*(\omega, \bar{h})$, $\omega > 0$.

In the following, only a delta-function reflector at $z = z_0$ will be considered, of the form $m_1(z) = \delta(z - z_0)$, which represents a single horizontal reflector. Recall that $m_1 = c^{-2} - c_0^{-2}$, with c_0 the background velocity. Then,

$$p(\bar{h}, \omega) \simeq w(\omega) A_0 e^{2ik_0 r_0}, \quad r_0 = \sqrt{\bar{h}^2 + z_0^2}, \quad (\text{A8})$$

with

$$A_0^{2D} = \frac{(i-1)}{8z_0} \sqrt{\frac{\omega c_0^3 r_0}{2\pi}}, \quad \text{or } A_0^{3D} = \frac{i\omega c_0}{16\pi z_0}. \quad (\text{A9})$$

Extended image

If the data are acquired for a seismic line in the plane $y = 0$ with an infinite and continuous distributions of shots, the

extended image of equation (8) in this plane becomes

$$R(x, z; 2\bar{h}_x, 0) = \text{Re} \int_{-\infty}^{\infty} d\omega \omega^2 \mathcal{F}(\omega) \int d\bar{h} \\ \times \int_{-\infty}^{\infty} dx_m p^*(x - \bar{h}_x, z, \omega) q_b(x + \bar{h}_x, z, \omega), \quad (\text{A10})$$

where the dependence on y in the 3D case has been dropped. The wavelet and additional data filters are included as $\mathcal{F}(\omega)$. In this expression, the wavefields p and q are determined by the Green functions for a different model with constant velocity c but the data fed into q correspond to a velocity c_0 . We have replaced $q(\cdot)$ by $\int d\bar{h} q_b(\cdot)$ corresponding to observed data at half-offset \bar{h} , that is, each receiver is treated separately in the analysis. The integration interval for \bar{h} depends on the acquisition geometry. With a seismic line on land, we typically have $\bar{h}_{\min} < 0 < \bar{h}_{\max}$. For a marine survey with a towed receiver cable, we can consider $0 < \bar{h}_{\min} < \bar{h}_{\max}$, which will be assumed to be the case below. The case $\bar{h}_{\max} < \bar{h}_{\min} < 0$ follows by symmetry.

The wavefields in equation (A10) are given by

$$p^*(x - \bar{h}_x, z, \omega) = G_{[c]}^*(r_1), \quad (\text{A11})$$

with

$$r_1 = \sqrt{\{(x - \bar{h}_x) - (x_m - \bar{h})\}^2 + z^2}, \quad (\text{A12})$$

and

$$q_b(x + \bar{h}_x, z, \omega) = G_{[c]}^*(r_2) p_{[c_0]}(\bar{h}, \omega), \quad (\text{A13})$$

with

$$r_2 = \sqrt{\{(x + \bar{h}_x) - (x_m + \bar{h})\}^2 + z^2}, \quad (\text{A14})$$

where the dependence on c and c_0 has been made explicit with the square brackets. Then

$$R(x, z; 2\bar{h}_x, 0) = \text{Re} \int_{-\infty}^{\infty} d\omega \int d\bar{h} \int_{-\infty}^{\infty} dx_m A_1 e^{i\psi_0}, \quad (\text{A15})$$

with amplitude

$$A_1^{2D} = \mathcal{F}(\omega) \omega^{3/2} \frac{(1+i)c}{64z_0} \sqrt{\frac{c_0^3 r_0}{2\pi^3 r_1 r_2}}, \quad (\text{A16})$$

or

$$A_1^{3D} = \mathcal{F}(\omega) \omega^3 \frac{ic_0}{256\pi^3 z_0 r_1 r_2}, \quad (\text{A17})$$

and phase

$$\psi_0 = \omega \left(\frac{2r_0}{c_0} - \frac{r_1 + r_2}{c} \right). \quad (\text{A18})$$

Stationarity of the phase requires zero derivatives with respect to half-offset \bar{h} and midpoint x_m , which happens at

$$x_m = x, \quad \bar{h}_x = \bar{h} \left(1 - \gamma^2 \frac{z}{z_c} \right), \quad (\text{A19})$$

where $\gamma = c/c_0$, $\beta = \gamma^2 - 1$, and $z_c = \gamma \sqrt{z_0^2 - \beta \bar{h}^2}$. Note that \bar{h} is defined implicitly and that we assume $z_0 \geq \bar{h} \sqrt{\beta}$ if $\beta \geq 0$. At the stationary point, the phase is $-\omega \frac{2r_0(z-z_c)}{c_0 z_c}$. Its Hessian at this point is diagonal with

$$\frac{\partial^2 \psi_0}{\partial \bar{h}^2} = \frac{2\omega(\gamma^4 z z_0^2 - z_c^3)}{c_0 \gamma^4 r_0^3 z} = \frac{2\omega z_0^2 (z - z_1)}{c_0 r_0^3 z}, \\ \frac{\partial^2 \psi_0}{\partial x_m^2} = \frac{-2\omega z_c^3}{c_0 \gamma^4 r_0^3 z} < 0, \quad (\text{A20})$$

where $z_1 = z_c^3 / (\gamma^4 z_0^2)$. The amplitudes reduce to

$$A_1^{2D} = \mathcal{F}(\omega) \frac{(1+i)z_c}{64z_0} \sqrt{\frac{\omega^3 c_0^5}{2\pi^3 r_0}}, \quad (\text{A21})$$

or

$$A_1^{3D} = \mathcal{F}(\omega) \frac{i\omega^3 z_c^2 c_0}{256\gamma^2 \pi^3 z_0 z^2 r_0^2}. \quad (\text{A22})$$

After integration over x_m and \bar{h} , the phase is $\psi_1 = -\omega \frac{2r_0(z-z_c)}{c_0 z_c}$, as mentioned above and the amplitudes become

$$A_2^{2D} = \mathcal{F}(\omega) \frac{[1 + i \text{sign}(z - z_1)] z_c}{64z_0^3} \sqrt{\frac{\omega c_0^7 r_0^5}{2\pi z_1 |z - z_1|}}, \quad (\text{A23})$$

or

$$A_2^{3D} = \mathcal{F}(\omega) \frac{[1 + i \text{sign}(z - z_1)] (1+i)r_0 z_c^2 c_0^2 \omega^2}{512\gamma^2 \pi^2 z z_0^3 \sqrt{z_1 |z - z_1|}}. \quad (\text{A24})$$

For the last step, we have to evaluate

$$A = 2 \text{Re} \int_0^{\infty} d\omega A_2 e^{i\omega\phi_1}, \quad \phi_1 = \psi_1/\omega = -\frac{2r_0(z-z_c)}{c_0 z_c}. \quad (\text{A25})$$

To obtain simple closed-form expressions, the frequency dependence of the amplitudes is removed by letting $\mathcal{F}(\omega) = \omega^{-1/2}$ in the 2D case and $\mathcal{F}(\omega) = \sqrt{2}/(\omega^2(1+i)) = \omega^{-2} e^{-i\pi/4}$ in the 3D case. For the integration, we can use $A = \int_0^{\omega_{\max}} d\omega [A_2 e^{i\omega\phi_1} + A_2^* e^{-i\omega\phi_1}]$ and

$$\lim_{\omega_{\max} \rightarrow \infty} \frac{\sin(\omega_{\max} \phi_1)}{\phi_1} = \pi \delta(\phi_1) = \pi \frac{\delta(z - z_{c,0})}{\left| \frac{d\phi_1}{dz}(z_{c,0}) \right|}. \quad (\text{A26})$$

To evaluate $d\phi_1/dz$, we can determine $d\bar{h}/dz$ from $\bar{h}_x = \bar{h}(1 - \gamma^2 \frac{z}{z_c})$ and $z_c = \gamma \sqrt{z_0^2 - \beta \bar{h}^2}$. Then,

$$\frac{d\phi_1}{dz}(z_{c,0}) = -\frac{2z_{c,0}}{\gamma^2 c_0 r_{c,0}}. \quad (\text{A27})$$

On the stationary-phase curve, $\bar{h}_x = -\beta\bar{h}$, $z = z_{c,0} = \gamma(z_0^2 - \bar{h}_x^2/\beta)^{1/2}$ and $r_0 = r_{c,0} = (z_0^2 + \bar{h}_x^2/\beta^2)^{1/2}$. If $\beta > 0$, then $\bar{h}_x < 0$ and we should have $(-\bar{h}_x) \leq z_0\sqrt{\beta}$ and $\bar{h} \leq z_0/\sqrt{\beta}$. The resulting extended image is $R(x, z; \bar{h}_x, 0) = A(x, z; \bar{h}_x)\delta(z - z_{c,0})$ with amplitude

$$A^{2D} = \frac{\gamma^5}{128z_0z_{c,0}^2} \sqrt{\frac{2\pi c_0^3 r_{c,0}^5}{|\beta|}}, \quad A^{3D} = \frac{(\gamma c_0)^3 r_{c,0}}{512\pi z_0 z_{c,0}^2} \sqrt{\frac{2}{|\beta|}}, \quad (\text{A28})$$

in 2D and 3D, respectively.

So far, an infinite acquisition was assumed. If the offset range is truncated at \bar{h}_m , we can solve z from $\psi_0 = 0$ at $x_m = x$, resulting in equation (14).

APPENDIX B: GRADIENT OF THE COST FUNCTIONAL

The functionals defined in equations (9) and (10) can be treated simultaneously. The forward problem is $Lp_s = f_s$ and the ‘reverse-time’ equations are

$$L^H q_s = - \sum_{r(s)} S_{r(s)}^T p_{r(s)}^{\text{obs}}, \quad (\text{B1})$$

similar to equations (1) and (7). The contributions to the extended reflectivity per shot and frequency are

$$R_s(\mathbf{x}, \mathbf{h}, \omega) = \sum_{r(s)} \omega^2 p_s^* \left(\mathbf{x} - \frac{1}{2}\mathbf{h}, \omega \right) q_{r(s)} \left(\mathbf{x} + \frac{1}{2}\mathbf{h}, \omega \right), \quad (\text{B2})$$

with summed versions $R(\mathbf{x}; \mathbf{h}) = \sum_{s,\omega} R_s(\mathbf{x}, \mathbf{h}, \omega)$. For simplicity, only a single shot and single frequency will be considered in what follows, given that the required summations are straightforward. The functional

$$J = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{h}} W(\mathbf{x}) \eta(\mathbf{h}) |\Xi R(\mathbf{x}; \mathbf{h})|^2 = \frac{1}{2} \sum_{\mathbf{x}, \mathbf{h}} R^H(\mathbf{x}; \mathbf{h}) \mathcal{A} R(\mathbf{x}; \mathbf{h}), \quad (\text{B3})$$

with

$$\mathcal{A}(\mathbf{x}, \mathbf{h}) = \Xi^H(\mathbf{x}) \eta^{1/2}(\mathbf{h}) W(\mathbf{x}) \eta^{1/2}(\mathbf{h}) \Xi(\mathbf{x}), \quad (\text{B4})$$

is not analytic, so complex derivatives cannot be used. One can either resort to the Wirtinger calculus (Wirtinger 1927) or consider the real and imaginary parts of a complex number separately. The last approach is followed here. The real part is denoted by a superscript r , the imaginary part by i .

The adjoint-state approach (Lions and Magenes 1972; Giles *et al.* 2003; Plessix 2006) for computing the gradient of the cost functional with respect to the model parameters starts

with a Lagrangian for J , which in this case for a single shot and one frequency reads

$$\begin{aligned} \mathcal{L} = J - \langle \lambda^r, L^r p^r - L^i p^i - f^r \rangle - \langle \lambda^i, L^r p^i + L^i p^r - f^i \rangle \\ - \langle \mu^r, (L^r)^T q^r + (L^i)^T q^i + S^T p^{\text{obs},r} \rangle \\ - \langle \mu^i, (L^r)^T q^i - (L^i)^T q^r + S^T p^{\text{obs},i} \rangle. \end{aligned} \quad (\text{B5})$$

The multipliers are λ^r , λ^i , μ^r and μ^i and the scalar product $\langle \cdot, \cdot \rangle$ involves integration over the spatial variables (or summation in the discretized case on a finite-difference mesh). Stationarity, with respect to the wavefields p and q , leads to the forward and reverse-time or adjoint problems

$$L\mu = \mathcal{G}_q, \quad L^H \lambda = \mathcal{G}_p, \quad (\text{B6})$$

respectively. Here,

$$\begin{aligned} \frac{\partial J}{\partial p^r}(\mathbf{x}) = \text{Re } \mathcal{G}_p(\mathbf{x}), \quad \frac{\partial J}{\partial p^i}(\mathbf{x}) = \text{Im } \mathcal{G}_p(\mathbf{x}), \\ \frac{\partial J}{\partial q^r}(\mathbf{x}) = \text{Re } \mathcal{G}_q(\mathbf{x}), \quad \frac{\partial J}{\partial q^i}(\mathbf{x}) = \text{Im } \mathcal{G}_q(\mathbf{x}), \end{aligned} \quad (\text{B7})$$

with

$$\mathcal{G}_p(\mathbf{x}) = \omega^2 \sum_{\mathbf{h}} \left[\mathcal{A} \left(\mathbf{x} + \frac{1}{2}\mathbf{h}, \mathbf{h} \right) R \left(\mathbf{x} + \frac{1}{2}\mathbf{h}, \mathbf{h} \right) \right]^* q(\mathbf{x} + \mathbf{h}), \quad (\text{B8})$$

and

$$\mathcal{G}_q(\mathbf{x}) = \omega^2 \sum_{\mathbf{h}} p(\mathbf{x} - \mathbf{h}) \left[\mathcal{A} \left(\mathbf{x} - \frac{1}{2}\mathbf{h}, \mathbf{h} \right) R \left(\mathbf{x} - \frac{1}{2}\mathbf{h}, \mathbf{h} \right) \right]. \quad (\text{B9})$$

Note that the wave operator L is analytic in the model parameters $m(\mathbf{x})$, so we can use the Cauchy-Riemann relations $\partial L^r / \partial m^r = \partial L^i / \partial m^i$ and $\partial L^r / \partial m^i = \partial L^i / \partial m^r$. The gradient of the cost functional with respect to the model becomes

$$\frac{\partial J}{\partial m^r} + i \frac{\partial J}{\partial m^i} = -\langle \lambda, \frac{\partial L}{\partial m^r} p \rangle_{\mathbf{c}} - \langle q, \frac{\partial L}{\partial m^r} \mu \rangle_{\mathbf{c}} \quad (\text{B10})$$

with the complex scalar product $\langle a, b \rangle_{\mathbf{c}} = \langle a, b^* \rangle$. For velocity updating, only its real part is needed and the summation over frequencies and shots should be included. Note that in the above, $R(\mathbf{x}; \mathbf{h})$ was assumed to be complex rather than real-valued, as in equation (8). In this case, the gradient follows from the same expressions as above but with the imaginary part of R set to zero in equations (B8) and (B9).