

Exploring some issues in acoustic full waveform inversion

W.A. Mulder* and R.-E. Plessix

Shell International Exploration and Production, PO Box 60, 2280 AB Rijswijk, The Netherlands

Received January 2007, revision accepted February 2008

ABSTRACT

The least-squares error measures the difference between observed and modelled seismic data. Because it suffers from local minima, a good initial velocity model is required to avoid convergence to the wrong model when using a gradient-based minimization method. If a data set mainly contains reflection events, it is difficult to update the velocity model with the least-squares error because the minimization method easily ends up in the nearest local minimum without ever reaching the global minimum.

Several authors observed that the model could be updated by diving waves, requiring a wide-angle or large-offset data set. This full waveform tomography is limited to a maximum depth. Here, we use a linear velocity model to obtain estimates for the maximum depth. In addition, we investigate how frequencies should be selected if the seismic data are modelled in the frequency domain. In the presence of noise, the condition to avoid local minima requires more frequencies than needed for sufficient spectral coverage.

We also considered acoustic inversion of a synthetic marine data set created by an elastic time-domain finite-difference code. This allowed us to validate the estimates made for the linear velocity model. The acoustic approximation leads to a number of problems when using long-offset data. Nevertheless, we obtained reasonable results. The use of a variable density in the acoustic inversion helped to match the data at the expense of accuracy in the inversion result for the density.

INTRODUCTION

In the 1980s, computers became sufficiently powerful to perform full waveform inversion on small seismic data sets with the aim of finding a subsurface model that produced the best fit to seismic data. Attempts were only partially successful, because the least-squares misfit error function often suffers from local minima. A good initial model is required to prevent gradient-based optimization methods from ending up in the wrong minimum. Examples can be found in papers by, among others, Crase *et al.* (1990, 1992), Pratt (1990, 1999), Sun and McMechan (1992), Djikpéssé and Tarantola (1999), Pratt, Plessix and Mulder (2001), Causse (2002) and Shipp and Singh (2002). Symes and Chavent and their co-workers studied reformulations of the least-squares error function to avoid local minima when using reflection data (Symes and

Carazzone 1991; Clement, Chavent and Gomez 2001; Shen *et al.* 2005). Also, it was realized that long offsets could help to avoid the problem of local minima when processing surface data (Mora 1989; Pratt *et al.* 1996; Ravaut *et al.* 2004; Sirgue and Pratt 2004; Brenders and Pratt 2007a,b). These data contain diving waves that allow for tomographic velocity inversion in parts of the subsurface traversed by these waves. Because diving waves reach a maximum depth that is often smaller than that used for migrating the data, full waveform inversion acts in two modes simultaneously, as discussed by Pratt *et al.* (1996). One mode is velocity updating, related to long wavenumbers in the model, in the shallower parts of the model reached by diving waves. This mode is called full waveform tomography. The other is iterative migration, related to the shorter wavenumbers in the model and covering both shallow and deeper parts. We call this type of iterative migration nonlinear migration to distinguish it from iterative migration applied to the linearized wave equation or Born

*E-mail: wim.mulder@shell.com

approximation (Østmo, Mulder and Plessix 2002; Plessix and Mulder 2004; Mulder and Plessix 2004a,b). For the latter, the velocity model is kept fixed and because this problem is linear, convergence is much more rapid than for iterative nonlinear migration.

Here, we will address several issues in full waveform tomography and nonlinear migration that may arise when a frequency-domain acoustic finite-difference modelling code is used. Diving waves determine the maximum depth for tomographic inversion. The choice of frequencies will be different for tomography and migration. The presence of noise may impose a smaller frequency sampling to avoid local minima. By noise we mean signals that cannot be explained by the forward modelling due to both measurement errors and approximations in the physics of wave propagation. In particular, the acoustic approximation may cause problems.

In the next section, we consider a linear velocity model which has exact solutions for the wavefield both in 2D (Kuvshinov and Mulder 2006) and in 3D (Pekeris 1946; Kuvshinov and Mulder 2006). This provides an estimate of the maximum depth that diving waves can reach. It also allows us to study the character of the least-squares error function that measures the differences between synthetic and observed data. The error function has many local minima already for this simple velocity model.

If the seismic data contain mainly reflection events, both linear and nonlinear migration require a careful choice of frequencies when using a frequency-domain finite-difference code. Mulder and Plessix (2004a) obtained a criterion that avoids aliasing in depth. This leads to equidistant frequency sampling. Sirgue and Pratt (2004) provided guidelines for choosing frequencies when inverting surface data for a velocity model with a frequency-domain finite-difference code. They require continuous vertical wavenumber coverage of the subsurface. This leads to a non-equidistant frequency sampling that is sparser than required for migration. Here, we further investigate the choice of frequencies along the lines of the multi-scale approach by Bunks *et al.* (1995). At each frequency, we examine the width of the basin around the global minimum of the least-squares error with the constant velocity-depth gradient as a single parameter. We also consider the effect of noise. Next, we study the effect on velocity updating when going from single-parameter inversion to multi-parameter inversion with an arbitrary dependence of the velocity on depth.

The linear velocity model allowed us to make quantitative estimates. For realistic complex subsurface models, these estimates cannot be used but still provide some insight into the

process. We then investigated acoustic waveform inversion of a marine data set computed by a time-domain finite-difference code in the elastic Marmousi2 model (Martin, Wiley and Marfurt 2006). The unknowns for the acoustic inversion were P-velocity, density and source wavelet (Mulder and Plessix 2004b). We could determine the maximum depth reached by diving waves by simple inspection of the first gradient used for model updating. Here, gradient refers to the derivative of the least-squares error with respect to the model parameters. Because the model extends beyond this depth, we had to use the choice of frequencies that avoids aliasing. For reference, we included the result of acoustic waveform inversion of unrealistic, purely acoustic data generated in the same model.

ESTIMATES FOR A SIMPLE MODEL

Velocity linear with depth

The velocity updates obtained by full waveform tomography are more or less valid up to the maximum depth reached by the diving waves. To obtain estimates of the depth reached by diving waves and the choice of frequencies for waveform tomography, we consider an acoustic model with constant density and a velocity $v(z)$ that increases linearly with depth z : $v(z) = v_0 + \alpha z$ where v_0 is the velocity at zero depth and α the velocity-depth gradient. Although this model is too simplistic for real applications, it will improve upon constant-velocity estimates and provide guidelines for waveform tomography.

In 3D, the pressure for a delta-function source at the origin is (Pekeris 1946; Kuvshinov and Mulder 2006)

$$p = \frac{1}{2\pi} A e^{i\mu \operatorname{arccosh} u}, \quad \mu = \sqrt{(\omega/\alpha)^2 - 1/4}, \quad (1)$$

where

$$A^{-1} = \sqrt{4v_0\alpha^{-2}(u^2 - 1)} = 2r \sqrt{1 + \frac{(\alpha r)^2}{4v_0(v_0 + \alpha z)}},$$

and

$$u = 1 + \frac{(\alpha r)^2}{2v_0(v_0 + \alpha z)}, \quad r = \sqrt{x^2 + y^2 + z^2}. \quad (2)$$

Here $\omega = 2\pi f$ for a frequency f . We have taken the complex conjugate of the formulas in Kuvshinov and Mulder (2006), because here we use the opposite Fourier convention. We place the source at the origin and receivers at $y = 0$ and $z = 0$. Note that equation (1) describes propagating waves only for $\omega > \alpha/2$, which, in practice, is usually the case. These waves are all diving waves when recorded at non-zero offsets.

Wavefronts in this velocity model are surfaces of constant u , given by equation (2). In the plane $y = 0$, they are perpendicular to circles which coincide with the turning rays in

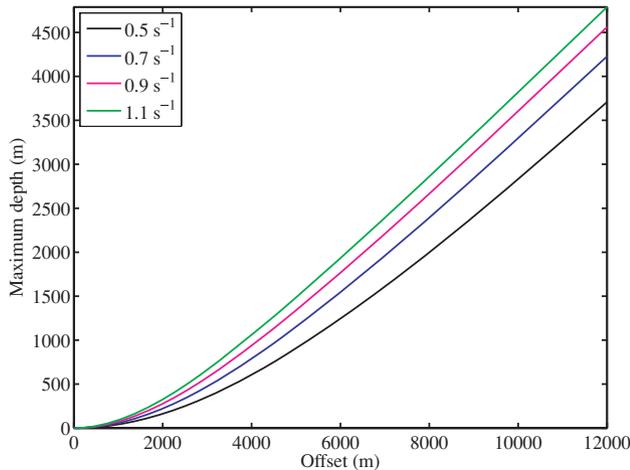


Figure 1 Maximum depth reached by turning rays as a function of offset, for various choices of the velocity-depth gradient α .

this model. The maximum depth reached for a given offset h is

$$z_{\max} = \frac{1}{2}h_c \left[-1 + \sqrt{1 + (h/h_c)^2} \right], \quad h_c = 2(v_0/\alpha). \quad (3)$$

The high-frequency approximation (ray tracing) will lead to the same result. Figure 1 shows the maximum depth as a function of offset for a number of velocity-depth gradients and $v_0 = 1.5$ km/s. Note that there is a significant increase in the maximum depth when going from 3 to 6 km. For very large offsets, z_{\max} tends to $h/2$.

In the past, marine surveys often had a maximum offset of 3 km. Diving waves would not penetrate beyond half a kilometre in that case. Nowadays, surveys often have a larger offset of 6 or even 12 km, allowing for waveform tomography of at least the shallower parts of the subsurface.

The least-squares error

The least-squares error measures the differences between modelled data, p , and observed seismic data, p^{obs} . For a linear velocity model, the error function depends on the two parameters v_0 and α . Its data-normalized version is

$$J(v_0, \alpha) = \frac{1}{2N_b} \sum_b |(p - p^{\text{obs}})/p^{\text{obs}}|^2.$$

Here N_b is the number of offsets involved in the summation over h . The pressure at the receiver is denoted by p and depends on offset and on the model parameters v_0 and α . The ‘observed’ data p^{obs} are those obtained in the true model, for which we will choose $v_0 = 1.5$ km/s and $\alpha = 0.7$ s $^{-1}$ in the examples.

In the linear velocity model, the pressure amplitude in 3D for realistic frequencies and velocity-depth gradients ($\omega > \alpha/2$)

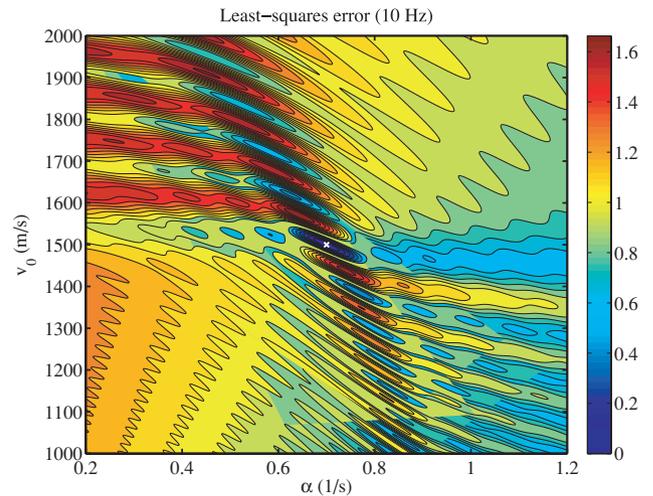


Figure 2 The weighted least-squares error for 3D data at 10 Hz obtained in a linear velocity model of the form $v(z) = v_0 + \alpha z$. The minimum is located at $\alpha = 0.7$ s $^{-1}$ and $v_0 = 1500$ m/s.

recorded by a receiver at offset h and position $x = r = h$, $y = 0$ and $z = 0$, is:

$$|p(h, 0, 0)| = \left(4\pi h \sqrt{1 + (h/h_c)^2} \right)^{-1},$$

with h_c as defined in equation (3). For short offsets, $|p|$ decays with h^{-1} , just as in the homogeneous case but for larger offsets $|p| \sim h^{-2}$. The transition from h^{-1} to h^{-2} decay occurs around the offset h_c , which is 4.3 km for $v_0 = 1.5$ km/s and $\alpha = 0.7$ s $^{-1}$.

Figure 2 shows the least-squares error at $f = 10$ Hz as a function of v_0 and α . The offsets h ranged from 100 to 6000 m at an interval of 25 m. The reference solution p^{obs} was obtained for $v_0 = 1.5$ km/s and $\alpha = 0.7$ s $^{-1}$. Apart from this global minimum, there are many local minima even for this simple velocity model. To reach the global minimum with a gradient-based descent method, the initial model should lie in the basin of the global minimum.

Choice of frequencies

Mulder and Plessix (2004a) determined the choice of frequencies for migration. Migration of the seismic data is usually carried out to depths that exceed those reached by diving waves. The requirement to avoid spatial aliasing leads to an estimated bound on the frequency spacing:

$$\Delta f \leq \frac{v/2}{-(h_{\max}/2) + \sqrt{z_{\max}^2 + (h_{\max}/2)^2}}, \quad (4)$$

where h_{\max} is the maximum offset in the data. The velocity v is assumed to be constant. In practice, we can use an estimate of

the effective normal move-out velocity. The maximum depth z_{\max} is determined by the maximum two-way traveltime in the data and the velocity v .

Sirgue and Pratt (2004) derived a criterion for choosing frequencies when performing waveform inversion in an acoustic frequency-domain finite-difference code. By requiring continuous vertical wavenumber coverage, they found that the frequencies should obey:

$$f_{n+1} = \beta f_n, \quad \beta = \sqrt{1 + [h_{\max}/(2z)]^2} > 1, \quad (5)$$

starting from the smallest frequency f_0 . Here z is the depth of the target. This usually leads to fewer frequencies than required by equation (4). Formula (5) is based on earlier papers by Miller, Oristaglio and Beylkin (1987), Wu and Toksöz (1987) and Mora (1989). These authors show that the wavenumber for a given frequency f ranges from $\beta^{-1} (2f/v)$ to $(2f/v)$. On the basis of equation (5), Pratt *et al.* (1996), Sirgue and Pratt (2004) and Operto *et al.* (2004) proposed to invert the data for one frequency at the time going from low to high frequencies.

We further investigate this idea of finding the velocity model by using one frequency at a time in a multi-scale approach, similar to Bunks *et al.* (1995). Instead of a constant velocity, we assume a velocity that increases linearly with depth. Our main goal is to avoid the local minima. The minimum frequency should be chosen such that the initial guess lies in the domain of attraction of the global minimum for that frequency. In practice, we have to start with the lowest frequency in the data. The next frequency can be based on formula (5) to fill in more details of the model. However, we should also stay within the basin of attraction of the global minimum. We can quantify these ideas for the linear velocity model. We assume that the velocity at the recording surface is known, so we fix v_0 . Figure 3 displays the least-squares error at frequencies between 6 and 30 Hz with an increment of 1 Hz for a fixed value of $v_0 = 1500$ m/s. The maximum offset was 6 km. The valley around the global minimum at $\alpha = 0.7$ s⁻¹ becomes narrower for increasing frequency. Note that in realistic velocity models, the location of the minimum may actually change with frequency. The width of the valley for a fixed, true v_0 is related to the diagonal of the Hessian $H(\alpha) = \partial^2 J / \partial \alpha^2$ evaluated at $\alpha = \alpha_{\min}$ where J has its global minimum. In appendix A, we show that $H(\alpha_{\min}) \simeq (\alpha_{\min} w_\alpha)^{-2}$ where $w_\alpha = \alpha_{\min} / (C_b f)$ and C_b is a constant depending on the maximum offset in the data. Therefore, the least-squares error behaves as $J \simeq \frac{1}{2} [(\alpha - \alpha_{\min}) / (\alpha_{\min} w_\alpha)]^2$ for α close to the minimum α_{\min} . This shows that w_α is a scaled measure of the width of the basin around the minimum. The quantity C_b depends on the

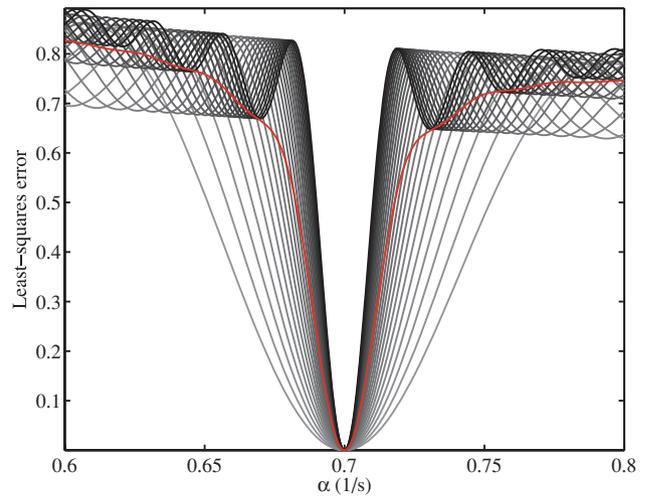


Figure 3 The weighted least-squares error for frequencies between 6 and 30 Hz with an increment of 1 Hz. The higher the frequencies, the narrower the valley around the minimum. The red line marks the average over the frequency range.

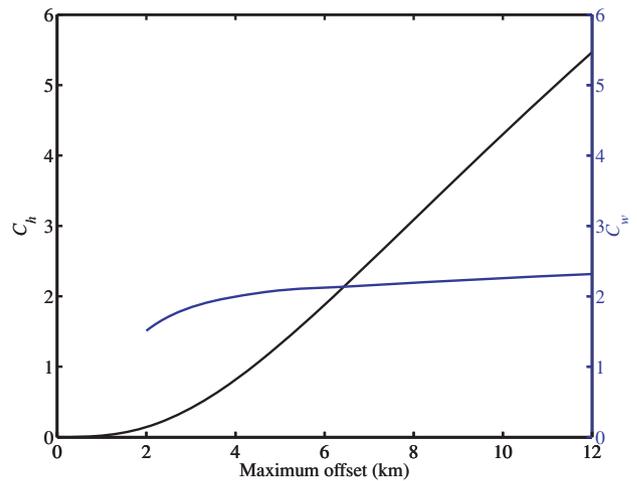


Figure 4 C_b (black) and C_w (blue) as a function of the maximum offset.

maximum offset h_{\max} and is displayed in Fig. 4. Here the summation in equation (A1) involved offsets between 100 m and h_{\max} at a 25 m interval. Because C_b increases with maximum offset, the width of the basin decreases as the maximum offset increases. Sirgue (2006) reached a similar conclusion on the basis of different arguments.

A more precise measure of the width of the basin of attraction is twice the distance between the minimum at $\alpha = \alpha_{\min}$ and the $\alpha = \alpha_1$ for which the maximum of J nearest to the minimum is attained. We define the corresponding scaled half-width as $\tilde{w}_\alpha = |\alpha_1 / \alpha_{\min} - 1|$. From the above, we may expect that w_α

and \tilde{w}_α should have approximately the same size. We therefore computed $C_w(h_{\max}) = \tilde{w}_\alpha(f, h_{\max})/w_\alpha(f, h_{\max})$ by determining the distance from the global minimum to the nearest local maximum of $J(\alpha)$ for various frequencies and offsets h_{\max} and for fixed $\alpha_{\min} = 0.7$. Note that we assume that $C_w(h_{\max})$ does not depend on the frequency. Figure 4 shows C_w as a function of h_{\max} between 2 and 12 km. Because the width of the basin is proportional to $1/f$, a multi-scale approach in the style of Bunks *et al.* (1995) is natural. Starting from the lowest frequency in the data, higher frequencies can be added to improve the velocity model according to equation (5). Note that this equation combined with equation (3) provides:

$$\beta = \sqrt{\frac{2}{1 - [1 + (h_{\max}/h_c)^2]^{-1/2}}}, \quad (6)$$

which is always larger than $\sqrt{2}$ and decreases with the maximum offset h_{\max} .

In the presence of noise, we may have to select a smaller frequency interval than allowed by equation (5) if we want to stay inside the basin of the global minimum. As an example, we again consider $\alpha_{\min} = 0.7$, fixed $v_0 = 1500$ m/s, offsets from 100 to 6000 m at a 25 m interval and a starting frequency of 6 Hz. The basin of the global minimum lies between $\alpha = 0.6073$ and 0.7941 . Suppose the noise energy amounts to $J = 0.5$. If we start with an initial value of around 0.61 and descend towards the minimum but stop when we reach the noise floor at $J = 0.5$, the resulting value of α is 0.651. The error function at 11.4 Hz has a basin around the global minimum between $\alpha = 0.651$ and 0.749 , so the next frequency should not exceed that value. If we choose 11.4 Hz as the next frequency and repeat the minimization until we reach $J = 0.5$, the resulting α is 0.674, corresponding to the left side of the basin at 21.4 Hz. In the same way, we find a subsequent frequency of 39.7 Hz. Had we applied equation (6), we would have obtained $\beta = 2.19$ and frequencies of 6.0, 13.1, 28.7 and 62.6 Hz, causing us to end up in local minimum. A more conservative value of $\beta = 1.8$ avoids this. In practice, there is always noise in the form of measurement errors and, more importantly, because of simplifications made in the modelling. Often, the effect of ‘noise’ will be worse for the higher frequencies.

Instead of selecting one frequency at the time, we may use a set of frequencies at once, although this is computationally expensive. The red line in Fig. 3 represents the error function obtained by averaging over the frequencies, in this case 6 to 30 Hz at a 1 Hz increment. The stacking narrows the basin around the minimum but also removes the ripples further away, effectively enlarging the range of α ’s for which the

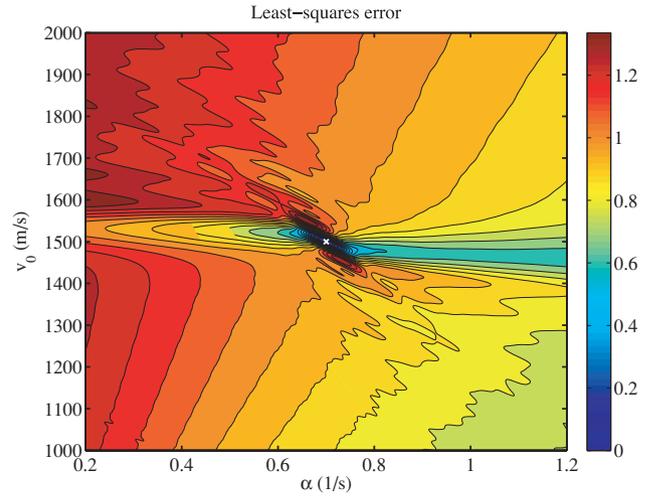


Figure 5 The weighted least-squares error for 3D data using frequencies between 6 and 20 Hz at an interval of 0.5 Hz.

global minimum can be reached by a descent method. The narrowing means that minimization by subsequently selecting a single frequency from low to high is a reasonable approach. The reduction of ripples means that it can still be beneficial to use a range of frequencies simultaneously, in spite of the added cost. When moving to the next frequency, the frequencies treated earlier are then included in the error function and model updates. Given the fact that in a frequency-domain finite-difference code, lower frequencies can be computed on a coarser grid, the added cost is not extremely large. Figure 5 shows the effect of using a range of frequencies on the error function in both parameters α and v_0 . Overall, the smoothness improves although some ripples remain.

Depth dependence

So far, we have considered a single-parameter model. One may rightfully argue that this is too specific for generalization to practical problems. In what follows, we consider model updates as a function of depth, thereby stepping outside the model space.

Given the least-squares error function $J(v(z)) = \frac{1}{2} \sum_b |p/p^{\text{obs}} - 1|^2$, its gradient with respect to $v(z) = 1/v^2(z)$ is (Plessix and Mulder 2004):

$$\frac{\partial J}{\partial v} = \omega^2 \sum_b \int_{x_{\min}}^{x_{\max}} dx m(b; x, z),$$

with

$$m(b; x, z) = \text{Re}[G(0, 0, 0; x, 0, z) G(b, 0, 0; x, 0, z)(p - p^{\text{obs}})^* / |p^{\text{obs}}|^2]. \quad (7)$$

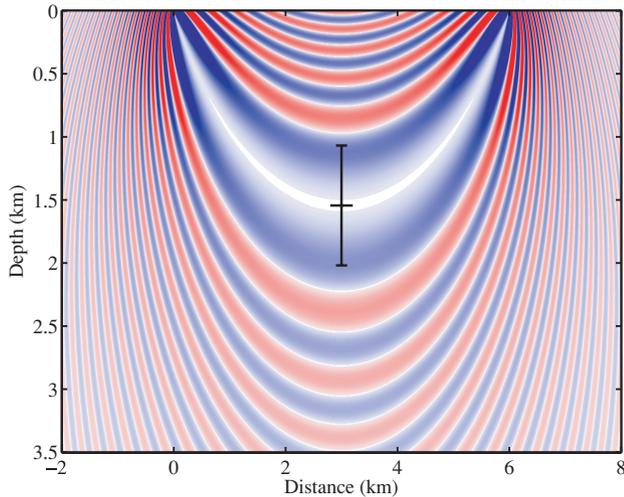


Figure 6 The contribution $m(b;x, z)$ to the gradient for a single offset $b = 6$ km, using 3D Green functions.

$G(x_1, y_1, z_1; x_2, y_2, z_2)$ is the Green function between two points (x_1, y_1, z_1) and (x_2, y_2, z_2) . The asterisk denotes the complex conjugate and $\text{Re}(\cdot)$ the real part. The pressures p and p^{obs} are recorded at $x = b, y = 0$ and $z = 0$ for a source at the origin. The gradient with respect to $v(z)$ is:

$$g(z) = \frac{\partial J}{\partial v} = -\frac{2}{v(z)^3} \frac{\partial J}{\partial v}.$$

A ‘preserved-amplitude’ migration image requires the diagonal of the Hessian:

$$H(z, z) = \frac{\partial^2 J}{\partial v^2}(z) = \frac{4}{v(z)^6} \omega^4 \sum_b \int_{x_{\min}}^{x_{\max}} dx |G(0, 0, 0; x, 0, z)G(b, 0, 0; x, 0, z)/p^{\text{obs}}(b)|^2,$$

leading to a depth-dependent model update $\delta v(z) = -H(z, z)^{-1}g(z)$.

Figure 6 displays $m(b; x, z)$, the contribution to the gradient of the least-squares error with respect to $1/v^2$ at 10 Hz before integration over x and using only a single offset $b = 6$ km. A delta-function source is located at the origin. The true model had $v_0 = 1500$ m/s and $\alpha = 0.7$. The ‘observed’ data p^{obs} were computed in that model, whereas the modelled data were based on $\alpha = 0.7 + 10^{-4}$ and the same v_0 . Offsets ranged from 100 to 6000 m with spacing of 25 m. The contribution $m(b; x, z)$ to the gradient is proportional to the Born wavepath given by $2(\omega/v)^2 G(0, 0, x, z)G(b, 0, x, z)$ (Woodward 1992). The broad blue band corresponds to the first Fresnel zone, cf. Pulliam and Snieder (1998) or Spetzler and Snieder (2004). In appendix B, we estimate that the half-width of the first Fresnel zone is: $w_F \simeq \sqrt{Lv(z_{\max})/\omega}$, with $v(z_{\max})$ the velocity

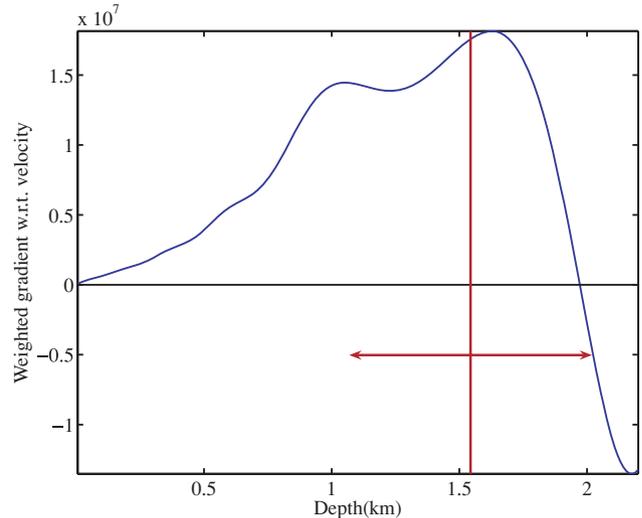


Figure 7 Weighted gradient w.r.t. $v(z)$.

at the deepest point reached by the ray and L the length along the ray path from source to receiver, given by:

$$L = (2z_{\max} + b_c) \arctan(b/h_c). \tag{8}$$

The vertical bar in Fig. 6 marks the interval from $z_{\max} - w_F$ to $z_{\max} + w_F$. Integration over x will lead to destructive interference except in the first Fresnel zone for x values around half the offset of 6 km used to generate the figure. After summation over offsets b between 100 and 6000 m at an interval of 25 m, we obtain the result $-\delta v(z) = H(z, z)^{-1}g(z)$ shown in Fig. 7. We used $x_{\min} = \max(x_L, -b)$ and $x_{\max} = \min(x_R, 2b)$, where $x_L = -2$ km and $x_R = 8$ km. A proper update of the velocity would be proportional to a linear function in depth. The vertical line in Fig. 7 marks z_{\max} , the maximum depth reached by the ray between source and receiver. The horizontal arrows indicate the width of the Fresnel zone between $z_{\max} - w_F$ and $z_{\max} + w_F$. We apparently obtain an update that resembles a linear function down to $z_{\max} - w_F$ but then starts to decay. Beyond z_{\max} , the update gradually becomes useless.

Numerical test

We verified the conclusion that a single frequency suffices to update the velocity trend by running a numerical inversion with the finite-difference frequency-domain code described elsewhere (Mulder and Plessix 2004b). Two-dimensional data were computed for the profile $v(z) = 1500 + 0.7z$ with the same acquisition geometry as before. The initial model had a velocity profile $v(z) = 1500 + 0.7 \min(0, z) + 0.6 \max(0, z)$, with z in m and v in m/s. This matches the true model for

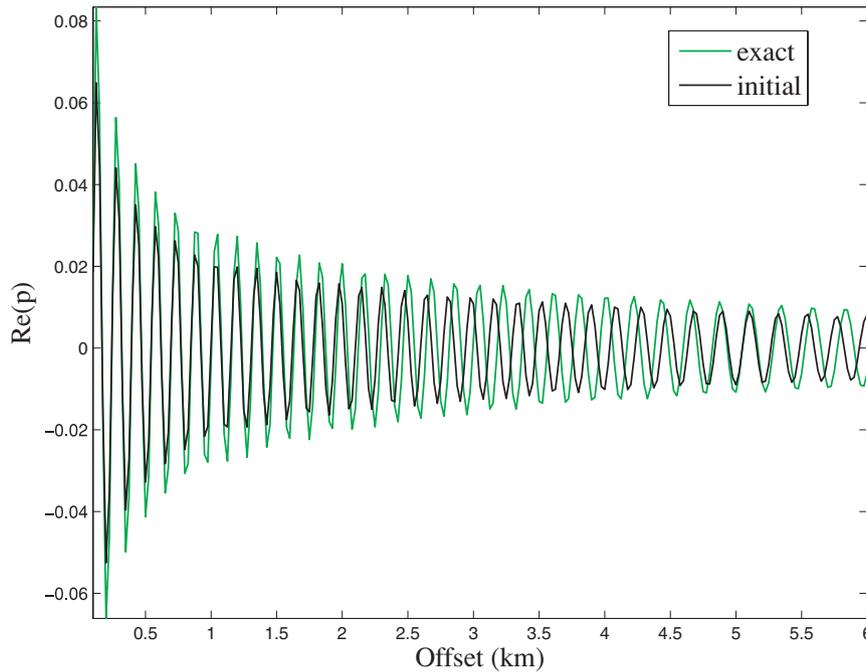


Figure 8 Real part of the 6 Hz data for the initial model.

negative depths but has a significant perturbation of the slope for positive depths. We parameterized the velocity model by a cubic-spline representation for the slowness with 100 m distance between spline nodes and depths ranging from -800 to 3600 m. This ensures some smoothness of the velocity model. In addition, we employed a smoothness penalty term to minimize differences between neighbouring values of the spline nodes.

Figure 8 shows the real part of the pressure data at zero depth for the initial and true model at 6 Hz. For the minimization, we used the least-squares error function without the normalization by p^{obs} : $J = \frac{1}{2} \sum_b |\beta p - p^{\text{obs}}|^2$. Here β accounts for the wavelet in the data. It was determined by $\beta = \sum_b p^{\text{obs}*} p / \sum_b |p|^2$ (Mulder and Plessix 2004b). We imposed a zero-phase wavelet by setting the imaginary part of β to zero.

Note that the absence or presence of data normalization does not change the kinematics. The problem of loop skipping between observed and computed data remains the same. However, data weighting or offset weighting allows us to emphasize that part of the data for which the difference in arrival times between observed and computed waveforms is smaller than half a wavelength. Once the waveforms match in these parts, the minimization can proceed to fit other parts of the data that have a smaller weight. We might hope that the loop

skipping problem has disappeared or has become less severe after the model has been optimized for the part of the data that had a larger weight.

Figure 9 displays the result of minimization of the least-squares error, together with the initial and true model. The model was kept fixed for depths $z \leq 0$ m. We did not apply an approximation of the inverse of the Hessian. The data for the resulting velocity model cannot be visually distinguished from the true data. Note that the model is correct down to $z_{\text{max}} = 1.54$ km and starts to flatten beyond this, due to the smoothness penalty term.

A numerical experiment starting from the same initial model but with the least-squares error normalized by p^{obs} failed to reproduce the correct model. The initial model lies outside the basin of attraction shown in Fig. 3. Without the scaling by p^{obs} , we nevertheless managed to obtain the correct result. We can explain the difference in behaviour by considering Fig. 8. At short offsets, the loops of the exact and modelled data coincide. Between 2.5 and 4.5 km offset, loop skipping occurs and around 5 km offset, the wrong loops match. With the scaling by p^{obs} , the long-offset data have the same weight as the near-offset data and the minimization process stagnates by ending up in a local minimum. Without the scaling, the near-offset data dominate the error. As a result, first the shallow part of the model is updated and later, the deeper parts.

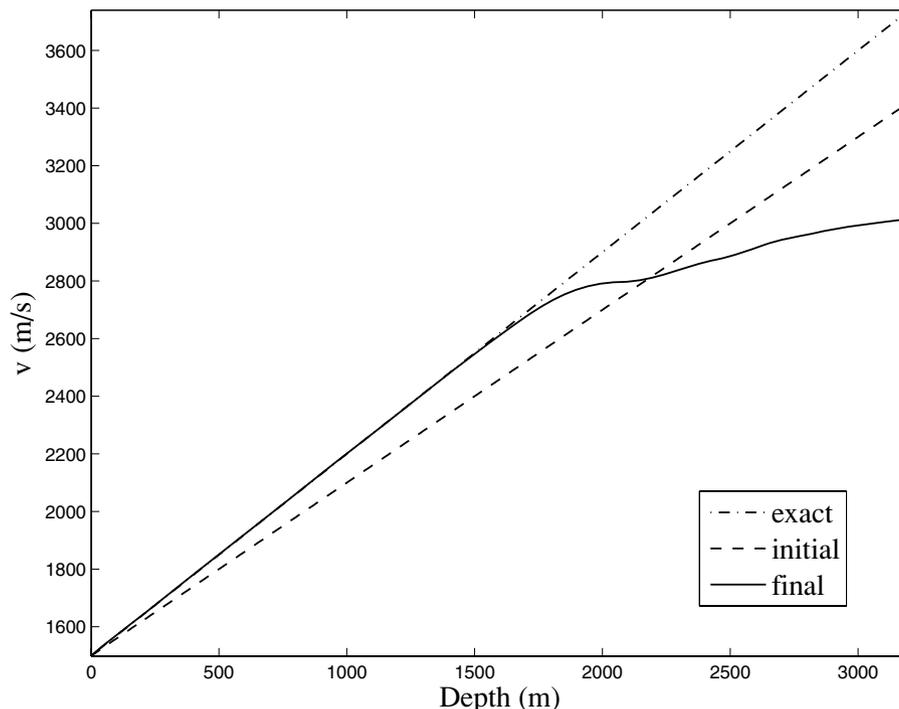


Figure 9 Initial, final, and exact velocity model.

The short offsets are matched first, and gradually this also improves the fit for the longer offsets. In this way, the local minimum could be avoided.

In practice, one generally favours the long-offset data that contain diving waves with significant energy to update the velocity model with full waveform tomography, starting from an initial model built by refraction tomography (Pratt *et al.* 1996; Shipp and Singh 2002; Operto *et al.* 2004; Brenders and Pratt 2007a). Our result suggests that we could also start with a velocity model constructed from reflection data, using any of the standard methods for velocity analysis based on data move-out. A drawback of the long-offset approach is the isotropic approximation. In the presence of anisotropy, full waveform tomography and refraction tomography are more sensitive to the horizontal velocities, whereas the reflection tomography is more sensitive to the vertical velocities.

From this exercise, we conclude that a valid update of the solution based on the weighted least-squares error can be obtained with just a single frequency in the absence of noise and for an initial model that is not too far from the true one, at least for this simple model. This agrees with the previous section. Higher frequencies will add more detail and suppress noise, as discussed before.

ACOUSTIC INVERSION OF ELASTIC DATA

We inverted a marine data set based on an elastic model to further explore the aspects of the acoustic approximation and to evaluate some of the formulas derived in the previous section. Real data are more accurately described by an elastic than an acoustic model. However it is still common practice to apply acoustic full waveform tomography on real data. If we insist on using an acoustic 2D code for cost and complexity reasons, we will run into a number of problems. Firstly, let us assume that we are dealing with marine data and that the sea water velocity model and sea bottom position are known. The long-offset data related to the sea bottom reflections will then display a phase difference when comparing acoustic to elastic reflections and refractions. An obvious remedy is the use of a smooth velocity model to first process the diving waves and then fill in the sea bottom and layering by introducing the shorter offsets. Secondly, the actual velocity of a refracted wave in an elastic medium may differ from the velocity of the same wave in an acoustic medium with the same densities and P-velocities. If we only have sediments with small contrasts, events will consist of diving waves and pre-critical reflections. Acoustics should then suffice. Strong contrasts, such as salt bodies or basalt layers, will generate post-critical events that

may lead to the wrong velocity below the top of the salt or basalt. Thirdly, thin-bed layering of sediments will result in an effective anisotropy. If that is ignored, the velocities produced by tomography of diving waves will represent the effective horizontal rather than vertical velocity. Fourthly, converted waves and 3D effects may cause unwanted effects.

To study some of these effects, we generated 5 seconds of data with a time-domain elastic finite-difference code for the elastic Marmousi2 model (Martin *et al.* 2006) with explosive shots between 5 and 12 km at a 50 m interval and 7.5 m depth in the water layer. The receivers had the same depth as the shots and offsets between 200 and 6000 m at a 25 m interval on the right side of each shot. Only the pressure was recorded in this marine example. A free-surface boundary condition was included. We also generated acoustic data for the same acquisition geometry. Subsets of the P-velocity and of the density model are shown in Figure 10. The S-velocity model is not displayed. We selected frequencies to avoid depth aliasing in nonlinear migration (Mulder and Plessix 2004a). They ranged from 6 to 15 Hz at a 0.2 Hz interval. Because the original wavelet was not zero-phase, it was estimated by fitting the direct wave for one shot. The result was used to change the phase of the data as if they had been generated by a zero-phase wavelet. The modified wavelet peaked at 12 Hz. Waveform inversion requires an initial model that is sufficiently close to the true one. We should have built such a model with any of the classic methods for velocity analysis. Instead, we obtained a kinematically correct model from the true one by adjoint interpolation onto a spline grid with a 400 m spacing, which was then interpolated back to a 10 m grid. Next, a shallow model was constructed for the sea bottom and the sediments just below. This was spliced into the very smooth background velocity model. The initial density was based on Gardner's rule (Gardner, Gardner and Gregory 1974). Figure 11 displays the initial model.

The acoustic inversion has the P-velocity, density and source wavelet as unknowns. For details, we refer the reader to an earlier paper (Mulder and Plessix 2004b). There, we describe the estimation of the wavelet at each iteration and the total-variation smoothness penalty terms.

The maximum depth reached by diving waves can be found by inspecting the gradient of the least-squares error function. As an example, Fig. 12 shows the gradient with respect to slowness in the initial model, on the top for all the frequencies, on the bottom for a single frequency of 12 Hz. For both, all shots and receivers were used. In spite of the presence of reflectors and guided waves, one might discern diving waves. With some effort, we can distinguish a bluish, banana-shaped

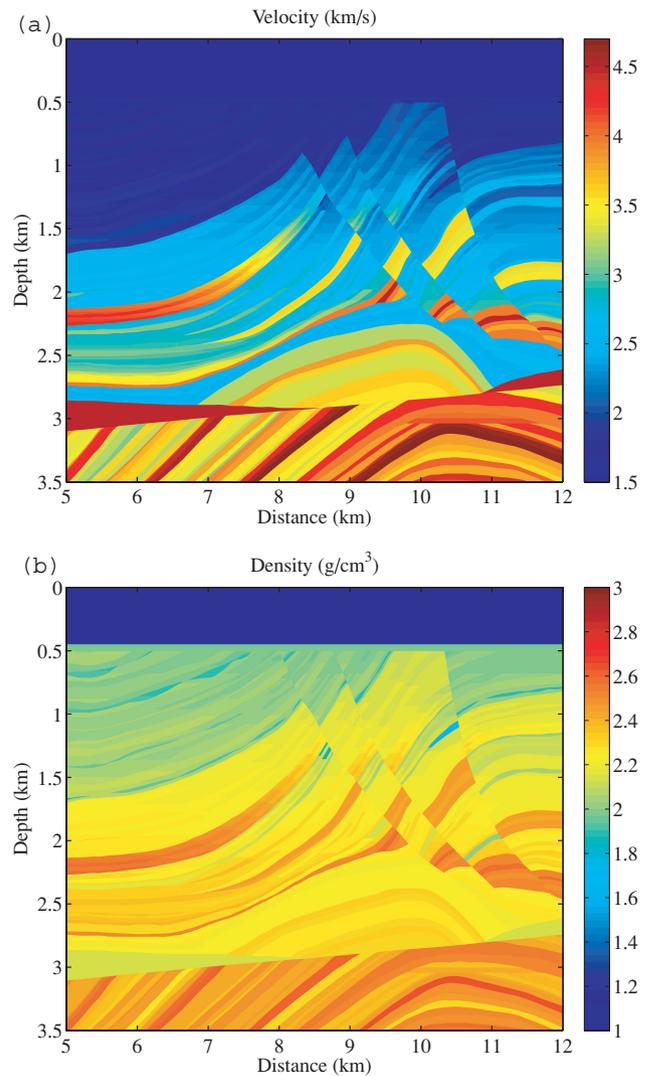


Figure 10 The original velocity (top) and density (bottom) model.

event with low spatial frequency starting at $x = 5$ km and 0.5 km depth and reaching a depth of 1.7 km around $x = 7$ km. Between $x = 9$ and 11 km, a similar bluish low-frequency event can be seen at a depth around 2.2 km.

If we try to use equation 3 to estimate the maximum depth reached by diving waves, we run into the problem that the velocity is constant at 1500 m/s down to the sea bottom at $z_b = 450$ m. We therefore modified our estimate as outlined in the following. We consider a vertical velocity profile that is constant down to a depth z_b and then increases linearly with depth. Rays in this model will be straight between $z = 0$ and $z = z_b$ and part of a circle for $z > z_b$. Appendix C lists the modified result for z_{\max} , the maximum depth reached by a ray and w_F , an estimate of half the width of the Fresnel zone. A

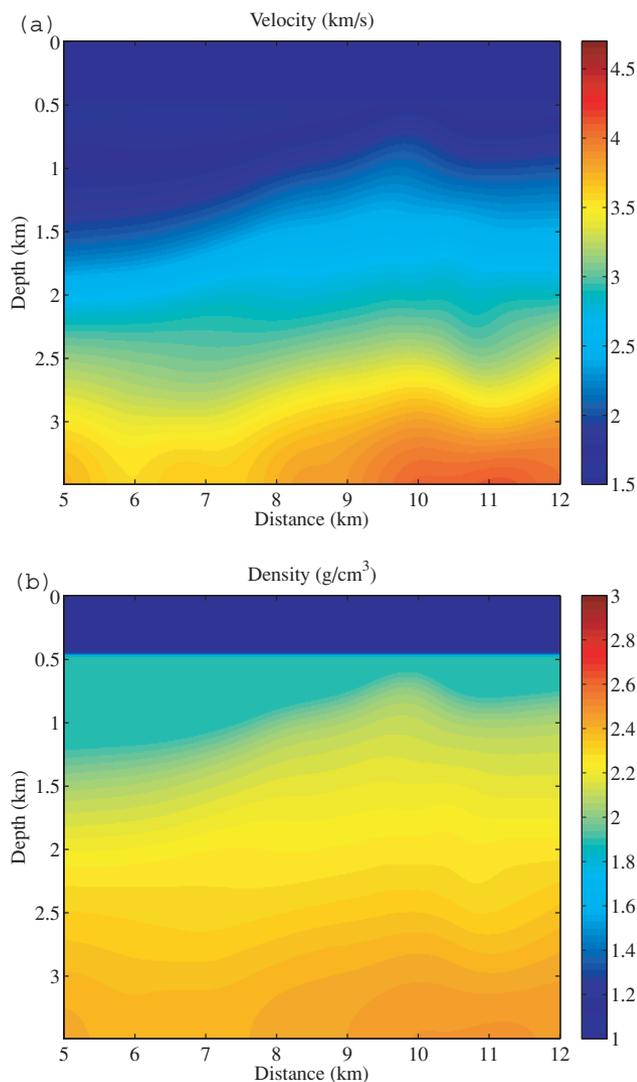


Figure 11 The starting model. The top panel displays the velocity, the bottom one the density determined by Gardner's rule.

fit of the velocity profile at $x = 6$ km down to 2 km depth yields $\alpha \simeq 0.5$, whereas at $x = 10$ km, $\alpha \simeq 0.9$. For the first, we obtain $z_{\max} \simeq 1.3$ km and $w_F \simeq 0.39$ km, for the second $z_{\max} \simeq 2.0$ km and $w_F \simeq 0.53$ km, using a frequency of 12 Hz. The observed depths mentioned earlier lie between the estimated z_{\max} and $z_{\max} + w_F$.

Both waveform tomography and nonlinear migration will play a role down to that depth. Beyond this, the model updates will be dominated by reflections and we are operating in 'migration mode'. This implies that we should work with a set of frequencies that avoids aliasing. The result after hundreds of iterations is shown in Figs 13, 14 and 17, using all frequencies simultaneously.

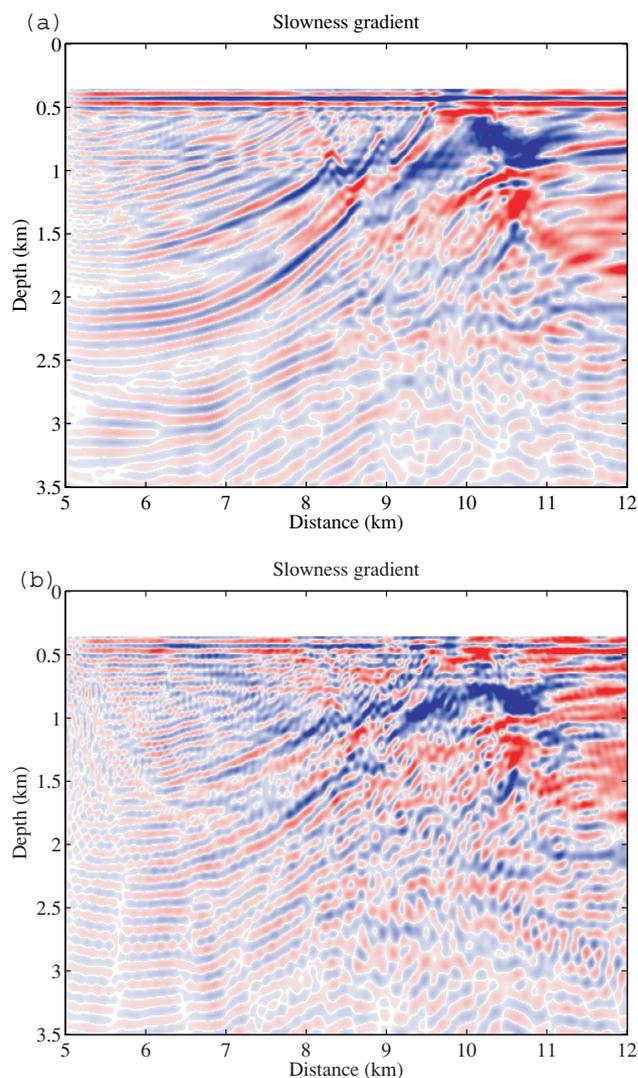


Figure 12 The gradient of the least-squares error with respect to the slowness gradient for the first iteration, using all 141 shots. The figure at the top shows the result for all frequencies, whereas the one at the bottom was obtained for a single frequency.

For reference, we repeated the acoustic inversion on unrealistic, purely acoustic data generated for the same subsurface model. This provided the velocity and density displayed in Fig. 15. The velocity reconstruction is fairly accurate in the shallower part. A vertical cross section is shown in Fig. 16. The density has a rather large error because we used Gardner's rule, which does not apply very well to the Marmousi2 model and because the inversion is more accurate in terms of the impedance (Tarantola 1984). As a result, the density compensates the lack of detail in the reconstructed velocity model.

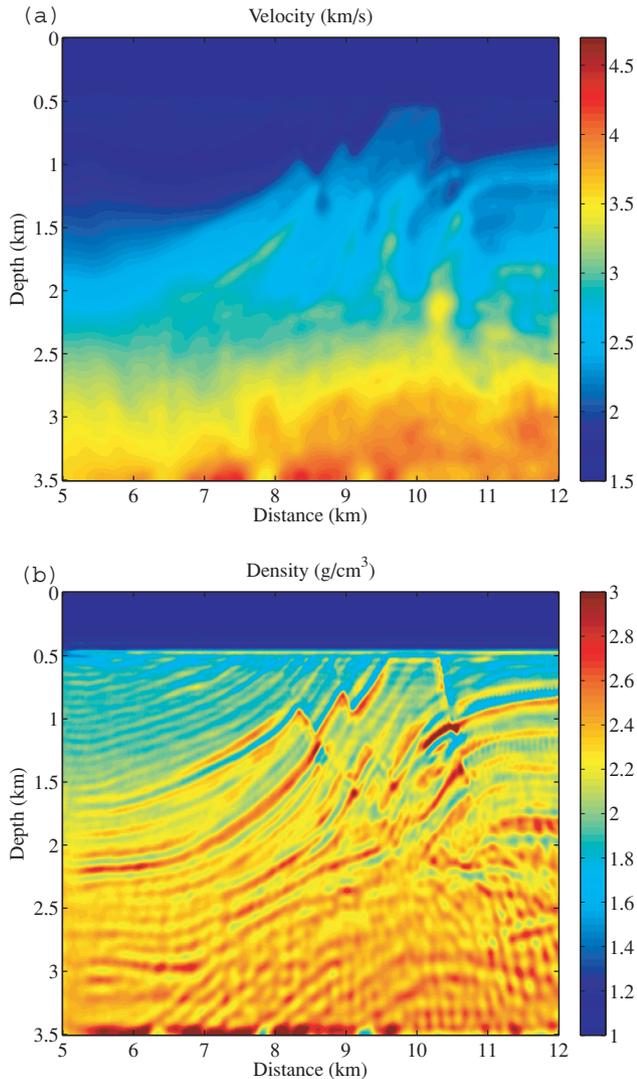


Figure 13 Velocity (top) and density (bottom) model obtained by inverting the elastic data.

Figure 13 shows the velocity and density obtained with elastic data. The density differs strongly from the true density as it tries to mimic elastic amplitudes and phases in an acoustic model. Density contrasts can partly handle amplitude differences between acoustics and elastics but not the phase differences that are particularly large for post-critical events. Nevertheless, we are able to capture the overall structure of the model.

In this example, the least-squares error reached a value of 0.15 times the data energy, which may seem quite large but is considerably smaller than usually found for real marine data sets. If we repeat the inversion with the exact P-velocity and density as a starting model, we find a least-squares error of

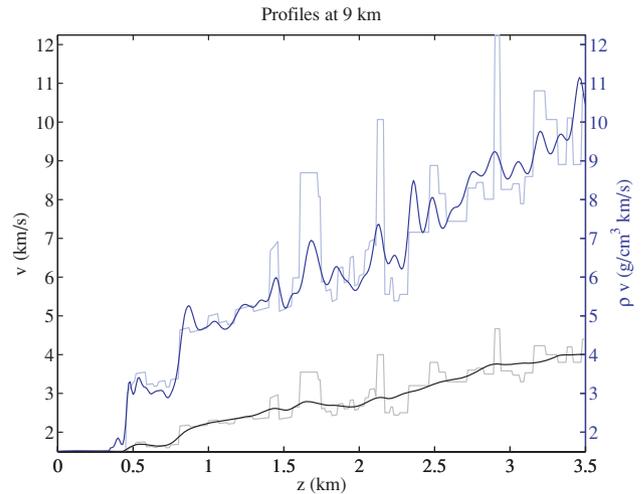


Figure 14 Vertical cross section of the P-velocity (black) and impedance (blue) at a distance of 9 km obtained for the elastic data. The true model is shown in lighter shades.

0.12 relative to the data energy. This shows that the other run has ended in a local minimum. It is obvious that the results of acoustic inversion of acoustic data should be better than with elastic data. In spite of this, the global minimum of the least-squares error was not found either in this case.

The rough part of the true impedance obtained by applying a spatial high-pass filter is displayed in Fig. 18. It resembles a typical migration image. This exact result is shown for comparison. Acoustic inversion of the acoustic data provided an image (top of Fig. 17) that closely resembles the true one. In this case, the least-squares error has decreased to 0.08 relative to the data energy. Acoustic inversion of the elastic data produces an image (bottom of Fig. 17) that shows a reasonable agreement with the true one in the less complex parts of the model, namely in the sediment part where the acoustic approximation is the most valid as explained before, but more pronounced differences in the complex part of the model between $x = 9$ and 11 km, already at smaller depths. In the deeper parts below the maximum depth reached by the diving waves, about 2.2 km, it becomes difficult to determine the structure of the reflectors.

Our investigation suggests that variable-density acoustic inversion of marine data can have some value if the subsurface is not too complex and the target is not too deep.

CONCLUSIONS

We have discussed a number of issues and problems in acoustic full waveform tomography and nonlinear migration when

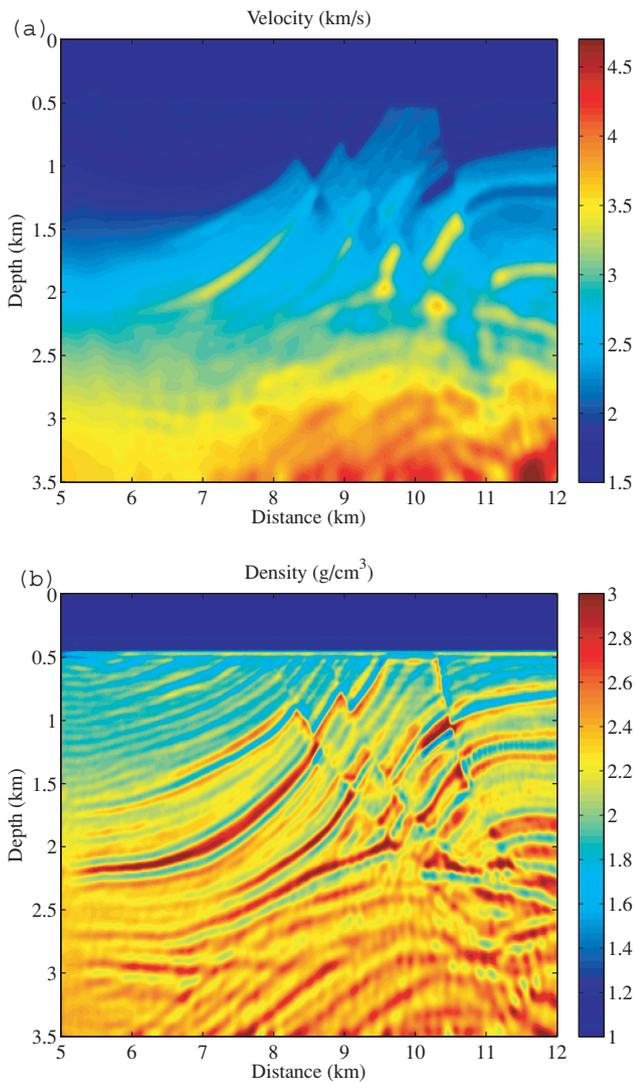


Figure 15 Velocity (top) and density (bottom) model obtained by inverting the acoustic data.

using a frequency-domain code. Tomographic inversion based on diving waves can be used down to a depth that can be estimated by assuming a simple linear velocity model with depth or by inspection of the gradient for a single frequency in a more realistic model. The computational domain should be limited to that depth if tomography with one or a few frequencies is the goal. Beyond, we are in migration mode and sufficiently many frequencies should be used to avoid aliasing in depth.

Elastic effects and effective anisotropy may limit the validity of long-offset tomographic inversion by means of an acoustic modelling code. Our example of acoustic inversion of pressure data obtained for a marine elastic model shows that the overall structure of the resulting model is reasonable, despite its

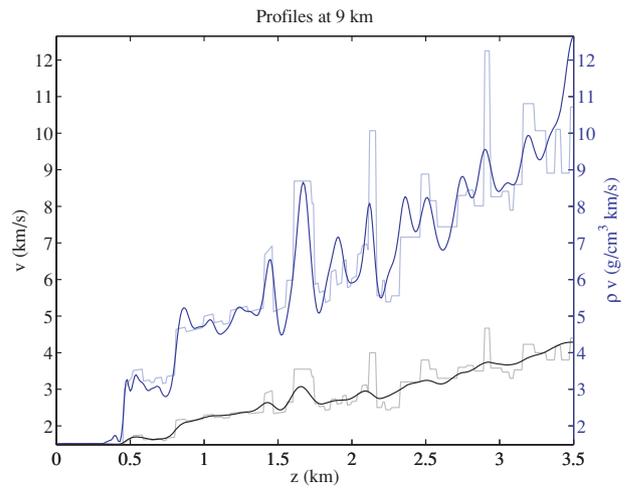


Figure 16 Vertical cross section of the P-velocity (black) and impedance (blue) at a distance of 9 km obtained for the acoustic data.

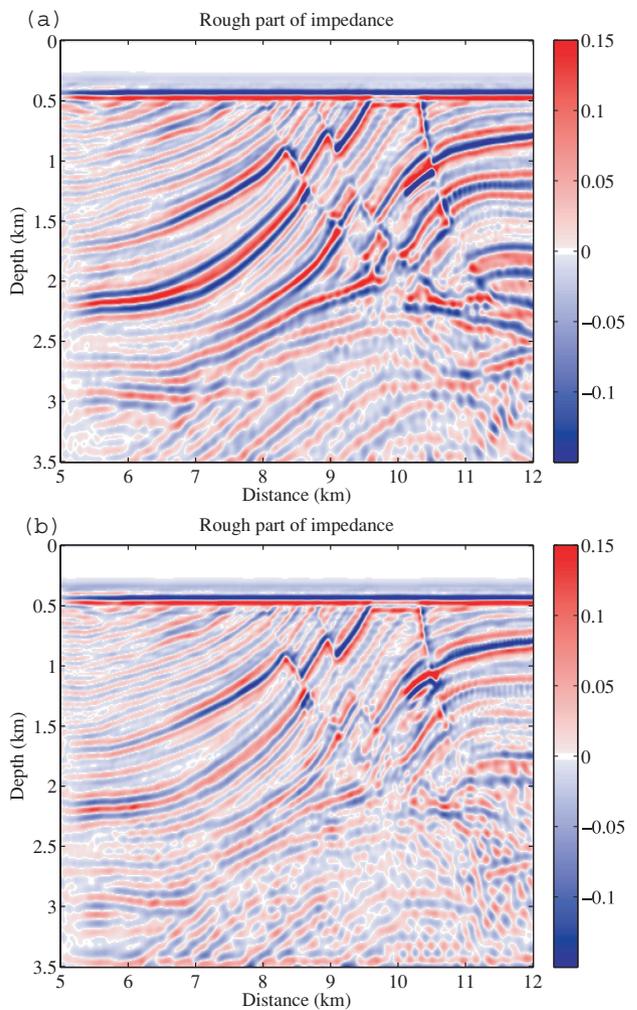


Figure 17 Rough part of the impedance after inversion of the acoustic (top) and elastic (bottom) data.

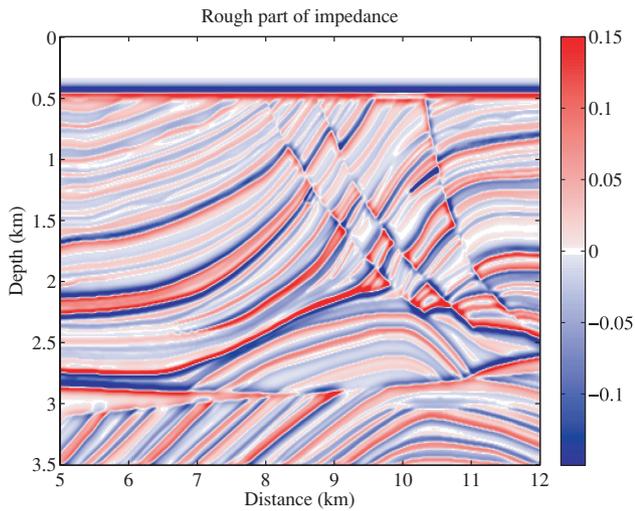


Figure 18 Rough part of the impedance for the exact model.

inherent non-uniqueness. The use of a variable density helps to compensate amplitude and phase differences between acoustics and elastics at the expense of the accuracy of the recovered density model.

REFERENCES

- Brenders A.J. and Pratt R.G. 2007a. Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International* **168**, 133–151.
- Brenders A.J. and Pratt R.G. 2007b. Efficient waveform tomography for lithospheric imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data. *Geophysical Journal International* **168**, 152–170.
- Bunks C., Saleck F.M., Zaleski S. and Chavent G. 1995. Multiscale seismic waveform inversion. *Geophysics* **60**, 1457–1473.
- Causse E. 2002. Avoiding pitfalls of least-squares inversion by using the energy-flux error criterion. *Geophysical Prospecting* **50**, 317–327.
- Clement F., Chavent G. and Gomez S. 2001. Migration-based traveltimes waveform inversion of 2-D simple structures: A synthetic example. *Geophysics* **66**, 845–860.
- Crase E., Pica A., Noble M., McDonald J. and Tarantola A. 1990. Robust elastic nonlinear waveform inversion: Application to real data. *Geophysics* **55**, 527–538.
- Crase E., Wideman C., Noble M. and Tarantola A. 1992. Nonlinear elastic waveform inversion of land seismic reflection data. *Journal of Geophysical Research* **97**, 4685–4704.
- Djirkpéssé H.A. and Tarantola A. 1999. Multiparameter L1 norm waveform fitting: interpretation of Gulf of Mexico reflection seismograms. *Geophysics* **64**, 1023–1035.
- Gardner G.H.F., Gardner L.W. and Gregory A.R. 1974. Formation velocity and density: The diagnostic basis for stratigraphic traps. *Geophysics* **39**, 770–780.
- Kuvshinov B.N. and Mulder W.A. 2006. The exact solution of the time-harmonic wave equation for a linear velocity profile. *Geophysical Journal International* **167**, 659–662.
- Lailly, P. 1983. The seismic inverse problem as a sequence of before stack migration. In: *Proceedings of the Conference on Inverse Scattering, Theory and Applications*. SIAM, Philadelphia.
- Martin G.S., Wiley R., and Marfurt K.J. 2006. Marmousi2: An elastic upgrade for Marmousi. *The Leading Edge* **25**, 156–166.
- Miller D., Oristaglio M.L. and Beylkin G. 1987. A new slant on seismic imaging – Migration and integral geometry. *Geophysics* **52**, 943–964.
- Mora P. 1987. Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics* **52**, 1211–1228.
- Mora P. 1988. Elastic wave-field inversion of reflection and transmission data. *Geophysics* **53**, 750–759.
- Mora P. 1989. Inversion=migration+tomography. *Geophysics* **54**, 1575–1586.
- Mulder W.A. and Plessix R.-E. 2004a. How to choose a subset of frequencies in frequency-domain finite-difference migration. *Geophysical Journal International* **158**, 801–812.
- Mulder W.A. and Plessix R.-E. 2004b. A comparison between one-way and two-way wave-equation migration. *Geophysics* **69**, 1491–1504.
- Operto S., Ravaut C., Improta L., Virieux J., Herrero A. and Dell'Aversana P. 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multi-scale traveltimes and waveform inversions: a case study. *Geophysical Prospecting* **52**, 625–651.
- Østmo S., Mulder W.A. and Plessix R.-E. 2002. Finite-difference iterative migration by linearized waveform inversion in the frequency domain. 72nd SEG Meeting, Salt Lake City, Utah, USA, Expanded Abstracts, 1384–1387.
- Pekeris C.L. 1946. Theory of propagation of sound in a half-space of variable sound velocity under conditions of the formation of a shadow zone. *Journal of the Acoustical Society of America* **18**, 295–315.
- Plessix R.-E. and Mulder W.A. 2004. Frequency-domain finite-difference amplitude-preserving migration. *Geophysical Journal International* **157**, 975–987.
- Pratt R.G. 1990. Frequency-domain elastic wave modeling by finite differences: a tool for crosshole seismic imaging. *Geophysics* **55**, 626–632.
- Pratt R.G., Song Z.-M., Williamson P.R. and Warner M.R. 1996. Two-dimensional velocity models from wide angle seismic data by waveform inversion. *Geophysical Journal International* **124**, 323–340.
- Pratt R.G. 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model. *Geophysics* **64**, 888–901.
- Pratt R.G., Plessix, R.E. and Mulder W.A. 2001. Seismic waveform tomography: crosshole results from the Nimr field in Oman. 71st SEG Meeting, San Antonio, Texas, USA, Expanded Abstracts, 714–717.
- Pulliam J. and Snieder R. 1998. Ray perturbation theory, dynamic ray tracing, and the determination of Fresnel zones. *Geophysical Journal International* **135**, 463–469.

- Ravaut C., Operto S., Impropa L., Virieux J., Herrero A. and Dell'Aversana P. 2004. Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International* **159**, 1032–1056.
- Spetzler J. and Snieder R. 2004. The Fresnel volume and transmitted waves. *Geophysics* **69**, 653–663.
- Shen P., Symes W. W., Morton S., Hess A. and Calandra H. 2005. Differential semblance velocity analysis via shot profile migration. 75th SEG Conference, New Orleans, Louisiana, USA, Extended Abstracts, 2249–2253.
- Shipp R.M. and Singh S.C. 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data. *Geophysical Journal International* **151**, 325–344.
- Sirgue L. and Pratt R.G. 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics* **69**, 231–248.
- Sirgue L. 2006. The importance of low frequency and large offset in waveform inversion. 68th EAGE Conference, Vienna, Austria, Extended Abstracts, A037.
- Sun R. and McMechan G. 1992. 2-D full-wavefield inversion for wide-aperture, elastic, seismic data. *Geophysical Journal International* **111**, 1–10.
- 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. 1986. A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics* **51**, 1893–1903.
- Woodward M.J. 1992. Wave equation tomography. *Geophysics* **57**, 15–26.
- Wu R.S. and Toksöz M.N. 1987. Diffraction tomography and multi-source holography applied to seismic imaging. *Geophysics* **52**, 11–25.

APPENDIX A

The diagonal of the Hessian

We consider the weighted least-squares error function $J(v_0, \alpha) = \frac{1}{2} N_b^{-1} \sum_b |p/p^{\text{obs}} - 1|^2$. The valley around the minimum at α_{\min} becomes narrower for higher frequencies. The width of the valley is measured by the diagonal of the Hessian $H(\alpha) = \partial^2 J / \partial \alpha^2$ for fixed v_0 , here set to the value at the minimum. For $\alpha = \alpha_{\min}$ also at the minimum, we have

$$H(\alpha_{\min}) = \frac{1}{\alpha_{\min}^2 N_b} \sum_b \frac{4\xi^2 [\hat{\omega}^2 (1 + \xi^2) - 1/4]}{(1 + \xi^2)^2} - \frac{4\hat{\omega}^2 \xi}{(1 + \xi^2)^{1/2}} \operatorname{arccosh}(1 + 2\xi^2) + \frac{\hat{\omega}^4}{\hat{\omega}^2 - 1/4} \operatorname{arccosh}^2(1 + 2\xi^2),$$

where $\xi = h/h_c$ and $\hat{\omega} = \omega/\alpha_{\min}$.

For $\omega \gg \alpha_{\min}$,

$$H(\alpha_{\min}) \simeq \left(C_b f / \alpha_{\min}^2 \right)^2, \quad C_b = \sqrt{N_b^{-1} \sum_b c(h)^2}, \quad (\text{A1})$$

with

$$c(h) = 2\pi \left(\operatorname{arccosh}(1 + 2\xi^2) - \frac{2\xi}{\sqrt{1 + \xi^2}} \right).$$

An approximation for C_b^2 can be found from:

$$C_b^2 N_b \Delta h \simeq C_b^2 h_{\max} \simeq \int_{b=0}^{h_{\max}} c(h)^2 dh$$

resulting in:

$$C_b^2 \simeq 4\pi^2 \left[20 - 8\sqrt{1 + 1/\xi_{\max}^2} \operatorname{arccosh}(1 + 2\xi_{\max}^2) + \operatorname{arccosh}^2(1 + 2\xi_{\max}^2) - 4 \arctan(\xi_{\max})/\xi_{\max} \right].$$

Note that, typically, ξ_{\max} has a value around 1, so all terms are of the same order of magnitude.

APPENDIX B

The width of the Fresnel zone

We will derive an estimate of the width of the Fresnel zone for the linear velocity model. If $\omega \gg \alpha$, the pressure in equation (1) becomes

$$p \simeq \frac{1}{2\pi} A e^{i\omega\tau}, \quad \tau = \alpha^{-1} \operatorname{arccosh} u.$$

Given an offset h for a single receiver, we want to find the depths where the migration image defined by equation (7) becomes zero along the vertical line through the midpoint at $x = h/2$:

$$m(h/2, 0, z) = \operatorname{Re} G(0, 0, 0; h/2, 0, z)^2 (p - p^{\text{obs}})^* / |p^{\text{obs}}|^2 = 0, \quad (\text{B1})$$

for depths z nearest to the depth z_{\max} . If the data p were obtained for a small perturbation $\delta\alpha$ of the original model, then at the receiver

$$p - p^{\text{obs}} \simeq i\omega \frac{\partial \tau}{\partial \alpha} \delta\alpha p^{\text{obs}}.$$

Here we have ignored the dependence of the amplitude on α , which would provide a much smaller contribution than the given term for $\omega \gg \alpha$. Equation (B1) becomes approximately

$$\sin\left(\omega[2\tau(h/2, 0, z) - \tau(h, 0, 0)]\right) = 0.$$

Because $\tau(h, 0, 0) = 2\tau(h/2, 0, z_{\max})$, we obtain

$$\tau(h/2, 0, z) - \tau(h/2, 0, z_{\max}) = 1/(4f).$$

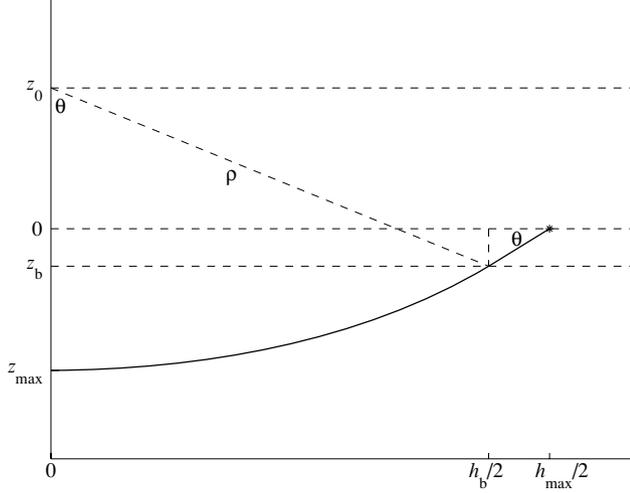


Figure A1 The black line shows half of a diving ray which is straight down to the sea bottom at depth z_b and follows part of the circle onwards.

A series expansion in z around z_{\max} yields

$$(z - z_{\max})^2 \simeq \frac{(v_0 + \alpha z_{\max})h}{4f}.$$

The half-width $w_F = |z - z_{\max}|$ obeys

$$w_F^2 = \frac{Lv(z_{\max})}{4f\psi(\eta)}, \quad \psi(\eta) = \frac{\arcsin(\eta)}{\eta},$$

$$\eta = \frac{(h/h_c)}{\sqrt{1 + (h/h_c)^2}}.$$

Here $L = h\psi(\eta) = (2z_{\max} + h_c) \arctan(h/h_c)$ is the length along the ray given in equation (8) and $v(z_{\max}) = v_0 + \alpha z_{\max}$.

Note that $1 \leq \psi(\eta) \leq \pi/2$. The upper bound provides the conservative estimate:

$$w_F = \sqrt{\frac{Lv(z_{\max})}{\omega}}.$$

APPENDIX C

Linear model with constant part

We estimate the depth that rays will reach when the velocity is constant to the sea bottom at depth z_b and linear beyond:

$$v(z) = v_0 + \alpha \max(0, z - z_b).$$

Rays in this model will be straight between $z = 0$ and $z = z_b$ and part of a circle for $z > z_b$. This part of the circle should be tangent to the straight line at $z = z_b$. Figure A1 sketches half the diving ray between the surface and its deepest point at z_{\max} . The centre of the circle is located at the depth z_0 obeying $v_0 + \alpha(z_0 - z_b) = 0$, so $z_0 = z_b - \frac{1}{2}h_c$ using $h_c = 2v_0/\alpha$. The radius of the circle is $\rho = z_{\max} - z_0$. We readily find

$$z_{\max} = z_b + \frac{h_c}{2} \left(-1 + \sqrt{1 + (h_b/h_c)^2} \right), \quad (\text{C1})$$

where

$$h_b = \frac{h_{\max}}{2} \left(1 + \sqrt{1 - (8h_c z_b)/h_{\max}^2} \right)$$

is the distance between the point where the ray enters the sea bottom and the point where it leaves. The width of the Fresnel zone $w_F \simeq \sqrt{Lv(z_{\max})/\omega}$ should now be evaluated for a length $L = 2(z_b \sin \theta + \rho \theta)$ with $\theta = \arctan(h_b/h_c)$.