Geophysical Journal International

*Geophys. J. Int.* (2011) **184,** 439–451 doi: 10.1111/j.1365-246X.2010.04848.x

# Seismic attenuation imaging with causality

Bobby Hak[1] and Wim A. Mulder[1,2]

[1]*Department of Geotechnology, Faculty of Civil Engineering and Geosciences, Delft University of Technology, PO Box* 5048, 2600 *GA, Delft, The Netherlands. E-mail: bhak@fugro-jason.com*

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

Accepted 2010 October 13. Received 2010 October 12; in original form 2010 January 25

## SUMMARY

Seismic data enable imaging of the Earth, not only of velocity and density but also of attenuation contrasts. Unfortunately, the Born approximation of the constant-density visco-acoustic wave equation, which can serve as a forward modelling operator related to seismic migration, exhibits an ambiguity when attenuation is included. Different scattering models involving velocity and attenuation perturbations may provide nearly identical data. This result was obtained earlier for scatterers that did not contain a correction term for causality. Such a term leads to dispersion when considering a range of frequencies. We demonstrate that with this term, linearized inversion or iterative migration will almost, but not fully, remove the ambiguity. We also investigate if attenuation imaging suffers from the same ambiguity when using non-linear or full waveform inversion. A numerical experiment shows that non-linear inversion with causality convergences to the true model, whereas without causality, a substantial difference with the true model remains even after a very large number of iterations. For both linearized and non-linear inversion, the initial update in a gradient-based optimization scheme that minimizes the difference between modelled and observed data is still affected by the ambiguity and does not provide a good result. This first update corresponds to a classic migration operation. In our numerical experiments, the reconstructed model started to approximate the true model only after a large number of iterations.

**Key words:** Numerical solution; Inverse theory; Seismic attenuation.



## 1 INTRODUCTION

Seismic imaging provides qualitative, structural information about the subsurface geology. Inversion—a term that we use in the mathematical sense of finding the Earth's parameters that best explain the observed data for a given type of forward modelling—leads to a quantitative description of material properties. Even with a simplified wave propagation model such as constant-density visco-acoustics, the reconstruction of not only the velocity but also the attenuation, as a function of subsurface position, can help in distinguishing between a fluid- or gas-filled rock formation.

Inversion for attenuation has been attempted by several authors with different levels of success. Ribodetti & Virieux (1998) considered linearized inversion for density, velocity and attenuation perturbations in a given background model using ray tracing to model the wave propagation. The correction for causality was not included. They claimed successful reconstruction of the model perturbations, although their results had large errors near sharp interfaces. Hicks & Pratt (2001) considered the non-linear inversion problem on real data and obtained convincing results with alternating updates of the velocity and the quality factor. Kamei & Pratt (2008) applied the same approach to cross-well data. Simultaneous inversion caused problems, but inverting for the velocity model first and then for the quality factor, reproduced the correct model with synthetic data. They applied their approach to real data as well. Smithyman *et al.* (2009) did the same on shallow reflection data. Ribodetti *et al.* (2007) performed non-linear inversion in a vertical seismic profiling (VSP)-type configuration using synthetic data. The reconstruction is not so good for simultaneous inversion of both the velocity and quality factor, but fine if the initial quality factor model is the true one. The results in these papers provide a hint at an underlying problem in the simultaneous inversion of seismic data for velocity and attenuation.

In an earlier paper (Mulder & Hak 2009), we showed that an ambiguity occurs when performing linearized inversion with the constant-density visco-acoustic wave equation. In linearized inversion, the Born approximation for a given background model is used. The unknowns are the amplitudes of the scatterers. For the constant-density visco-acoustic wave equation, scattering is due to perturbations in the velocity as well as in the quality factor that measures the attenuation. We found that two different scattering models produced nearly identical seismic surface data. As a consequence, it will be impossible to recover both velocity and attenuation parameters for scatterers in this approach.

In another study (Hak & Mulder 2010), we investigated if we could overcome the ambiguity by changing the acquisition geometry for collecting the seismic data. In 2-D, placement of sources and receivers on a circle around a point scatterer removed the ambiguity completely, in agreement with results found by Ribodetti *et al.* (2000). For more common geometries, such as cross-well and surface acquisitions, the ambiguity remained. Also for a background model that had velocity increasing with depth, the ambiguity persisted, despite the fact that the target could be illuminated from above and below if it was not too deep.

In those papers (Mulder & Hak 2009; Hak & Mulder 2010), we simplified our model such that our parameters were independent of frequency. Consequently, the correction term required to make a signal causal was dropped in the scattering term, although it was kept for the background model. The question remains if the inclusion of this term in the scatterer will affect the occurrence of the ambiguity. The causality term introduces a frequency dependency that translates into dispersion. Inversion over a sufficiently large bandwidth might remove the ambiguity, as suggested by the work of Ribodetti & Hanyga (2004) and Innanen & Weglein (2007). Another question that was not addressed is if the ambiguity would persist if the Born approximation is abandoned, that is, in non-linear or full waveform inversion. At first sight, the problem should persist in non-linear inversion. If a gradient-based descent algorithm is used to minimize the least-squares norm of the error between modelled and measured seismic data, the first model update in a non-linear minimization scheme will be identical to the one obtained for linearized inversion with the Born approximation. It can be expected that a small number of iterations will not remove the ambiguity. At some point, however, the secondary scattering effects will create a difference between the non-linear updates and those for the linearized inversion. Although the differences will be quite small, they may, or may not, be sufficient to provide a unique answer in the non-linear case.

Here, we address these two questions: will causal scattering in the Born approximation and will non-linear waveform inversion, without or with causality, remove the ambiguity between velocity and attenuation perturbations? Section 2 lists the governing equations and describes the finite-difference code used for non-linear inversion. In Section 3, we present results for linearized inversion for a point scatterer in a constant background model. Section 4 addresses non-linear inversion for a homogeneous background model with added sombrero-shaped functions, without and with noise, and a slightly more realistic marine model. In Section 5, we summarize the conclusions.

## 2 METHODOLOGY

### 2.1 Forward modelling

We consider the constant-density visco-acoustic wave equation

$$-\omega^2 m(\omega,\boldsymbol{x})p(\omega,\boldsymbol{x}) - \Delta p(\omega,\boldsymbol{x}) = s(\omega,\boldsymbol{x}), \tag{1}$$

with $p$ the pressure wavefield, $s$ a source term, $\omega = 2\pi f$ the angular frequency, $f$ the frequency and $m$ the inverse square root of the complex-valued wave velocity obeying

$$m(\omega,\boldsymbol{x}) = \sigma^2(\boldsymbol{x})\left[1 + \frac{\mathrm{i} - a(\omega)}{Q(\boldsymbol{x})}\right], \quad a(\omega) = \frac{2}{\pi}\log\left(\frac{\omega}{\omega_r}\right). \tag{2}$$

Here $Q \gg 1$ is the quality factor that describes attenuation and $\sigma$ is the slowness, both real-valued. The logarithmic term $a$ is required for causality (Aki & Richards 2002) and is defined relative to a reference frequency $\omega_r = 2\pi f_r$, usually chosen as $f_r = 1$ Hz.

For the Born approximation, the model is split into a smooth background component $m_b$, that does not produce significant reflections in the seismic frequency band and an oscillatory perturbation $\delta m$, related by $m = m_b + \delta m$. The scattered wavefield $\delta p$ follows from the pair of equations

$$\begin{aligned}&\left[-\omega^2 m_b(\omega,\boldsymbol{x}) - \Delta\right]p_b(\omega,\boldsymbol{x}) = s(\omega,\boldsymbol{x}),\\ &\left[-\omega^2 m_b(\omega,\boldsymbol{x}) - \Delta\right]\delta p(\omega,\boldsymbol{x}) = \omega^2\delta m(\omega,\boldsymbol{x})p_b(\omega,\boldsymbol{x}).\end{aligned} \tag{3}$$

For a source term $s(\omega,\boldsymbol{x}) = w(\omega)\delta(\boldsymbol{x}-\boldsymbol{x}_\mathrm{s})$, with a wavelet $w$ and a delta function source located at $\boldsymbol{x}_\mathrm{s}$, the solution of eq. (3) at a receiver location $\boldsymbol{x}_\mathrm{r}$ can be expressed as

$$\delta p_{\mathrm{r(s)}}(\omega) = w(\omega)\omega^2 \int \mathrm{d}\boldsymbol{x}\; G(\omega,\boldsymbol{x}_\mathrm{s},\boldsymbol{x})\,\delta m(\omega,\boldsymbol{x})\,G(\omega,\boldsymbol{x},\boldsymbol{x}_\mathrm{r}), \tag{4}$$

where the background model $m_b$ determines the Green function $G(\omega,\boldsymbol{x}_1,\boldsymbol{x}_2)$, the wavefield at $\boldsymbol{x}_2$ due to a delta function source at $\boldsymbol{x}_1$ or vice versa. The subscript r(s) denotes the dependency on the receiver location $\boldsymbol{x}_\mathrm{r}$, which may itself depend on the source location $\boldsymbol{x}_\mathrm{s}$.

In earlier papers (Mulder & Hak 2009; Hak & Mulder 2010), we assumed frequency independency of the scatterer by setting $a = 0$ in the scatterer, but not in the background model. In that case, eq. (4) can be summarized as $\mathbf{p} = \widetilde{\mathbf{F}}\delta\widetilde{\mathbf{m}}$, with a linear map $\widetilde{\mathbf{F}}$ and using scalar complex values for the scattering model $\delta\widetilde{m}(\boldsymbol{x})$ and the data $\delta p_{\mathrm{r(s)}}(\omega)$. The Hessian of the problem $\widetilde{\mathbf{H}} = \sum_\omega \widetilde{\mathbf{F}}^\mathrm{H}\widetilde{\mathbf{F}}$ was Hermitian, where the superscript $(\cdot)^\mathrm{H}$ denotes the conjugate transpose. Applying the Hessian to a delta-function perturbation at position $\boldsymbol{x}_\mathrm{p}$ gave the resolution function.

If we do not ignore the causal term in the scatterer, we have to abandon the representation by complex-valued variables and treat the real and imaginary parts of the perturbation as separate model parameters. Let

$$\begin{aligned}\delta m^{(1)}(\boldsymbol{x}) &= \sigma^2(\boldsymbol{x}) - \sigma_b^2(\boldsymbol{x}),\\ \delta m^{(2)}(\boldsymbol{x}) &= \sigma^2(\boldsymbol{x})\,Q^{-1}(\boldsymbol{x}) - \sigma_b^2(\boldsymbol{x})\,Q_b^{-1}(\boldsymbol{x}).\end{aligned} \tag{5}$$

Then $\delta m = \delta m^{(1)} + (\mathrm{i} - a)\delta m^{(2)}$. The inclusion of the dispersive term introduces a frequency dependency that breaks the symmetry between the real and imaginary part, which caused difficulties in our earlier papers. This might alleviate the scattering ambiguity.

In the causal case, a forward modelling operator $\mathbf{F}$ maps parameter pairs $\{\delta m^{(1)}, \delta m^{(2)}\}$ to the data

$$\begin{pmatrix}\mathrm{Re}\,\mathbf{p}\\ \mathrm{Im}\,\mathbf{p}\end{pmatrix} = \begin{pmatrix}\mathbf{F}^{11} & \mathbf{F}^{12}\\ \mathbf{F}^{21} & \mathbf{F}^{22}\end{pmatrix}\begin{pmatrix}\delta\mathbf{m}^{(1)}\\ \delta\mathbf{m}^{(2)}\end{pmatrix}. \tag{6}$$

This can be related to the earlier, simplified choice $\widetilde{\mathbf{F}}$ with $a = 0$ by

$$\begin{aligned}&\mathbf{F}^{11} = \mathrm{Re}\,\widetilde{\mathbf{F}}, \quad \mathbf{F}^{12} = -\mathrm{Im}\,\widetilde{\mathbf{F}} - a\,\mathrm{Re}\,\widetilde{\mathbf{F}},\\ &\mathbf{F}^{21} = \mathrm{Im}\,\widetilde{\mathbf{F}}, \quad \mathbf{F}^{22} = \phantom{-}\mathrm{Re}\,\widetilde{\mathbf{F}} - a\,\mathrm{Im}\,\widetilde{\mathbf{F}}.\end{aligned} \tag{7}$$

Note that now all values are real. The corresponding Hessian becomes

$$\mathbf{H} = \sum_\omega \mathbf{F}^\mathrm{T}\mathbf{F} = \begin{pmatrix}\mathbf{H}^{11} & \mathbf{H}^{12}\\ \mathbf{H}^{21} & \mathbf{H}^{22}\end{pmatrix}. \tag{8}$$

The superscript $(\cdot)^\mathrm{T}$ denotes the transpose.

By applying this Hessian to a model with a delta-function perturbation at position $\boldsymbol{x}_\mathrm{p}$, we obtain four resolution functions now, one for each operator $\mathbf{H}^{ij}$. For $j = 1$, we obtain the pair for a delta-function perturbation of $\delta m^{(1)}$, producing $g^{11}$ and $g^{21}$, the images in the first and second component, respectively. Similarly, we find the pair for a delta function perturbation of the second model component, $\delta m^{(2)}$, for $j = 2$, producing images $g^{12}$ and $g^{22}$. The first pair corresponds mainly to a velocity perturbation and the second pair corresponds to an attenuation perturbation. Ideally, $g^{11}$ and $g^{22}$ should resemble spikes at the position of the scatterer whereas $g^{21}$ and $g^{12}$ should be small.

## 2.2 Linearized inversion

Least-squares inversion for the scatterer in a given background model that is kept fixed amounts to solving the linear problem

$$\mathsf{H}\,\delta\mathbf{m} = \mathsf{H}\,\delta\widehat{\mathbf{m}}, \tag{9}$$

where $\delta\widehat{\mathbf{m}}$ is the true scatterer and $\delta\mathbf{m}$ its reconstruction. As the Hessian $\mathsf{H}$ is singular, the linear system can be solved by means of the pseudo-inverse $\mathsf{H}^\dagger$. This provides the minimum-norm solution. The system may also be solved by an iterative method that can handle its null-space, for instance, the conjugate-gradient method. Without pre-conditioning, this method should also provide the minimum-norm solution. With pre-conditioning, the usual norm is replaced by a weighted norm, determined by the pre-conditioner, and the solution of the singular problem will differ from the one obtained without pre-conditioning.

## 2.3 Non-linear inversion

We used a frequency-domain finite-difference code (Mulder & Plessix 2004) to compute synthetic data and carry out the least-squares inversion. As pointed out by Marfurt & Shin (1989), the solution of eq. (1) for 2-D problems is more efficient in the frequency domain than in the time domain. The discretization of eq. (1) leads to a large but sparse matrix. Its lower–upper (LU)-decomposition can be computed by nested dissection (George & Liu 1981). This step is costly, but is required only once for each frequency. The result can then be used for all shots and also for backpropagation during migration and inversion. The spatial discretization is a compact fourth-order finite-difference scheme with a nine-point stencil. We used the generic fourth-order choice (Harari & Turkel 1995) rather than the parameters suggested by Jo *et al.* (1996). In their notation, our choice amounts to $a = c = 2/3$, $d = 1/12$ and $e = 1 - c - 4d = 0$.

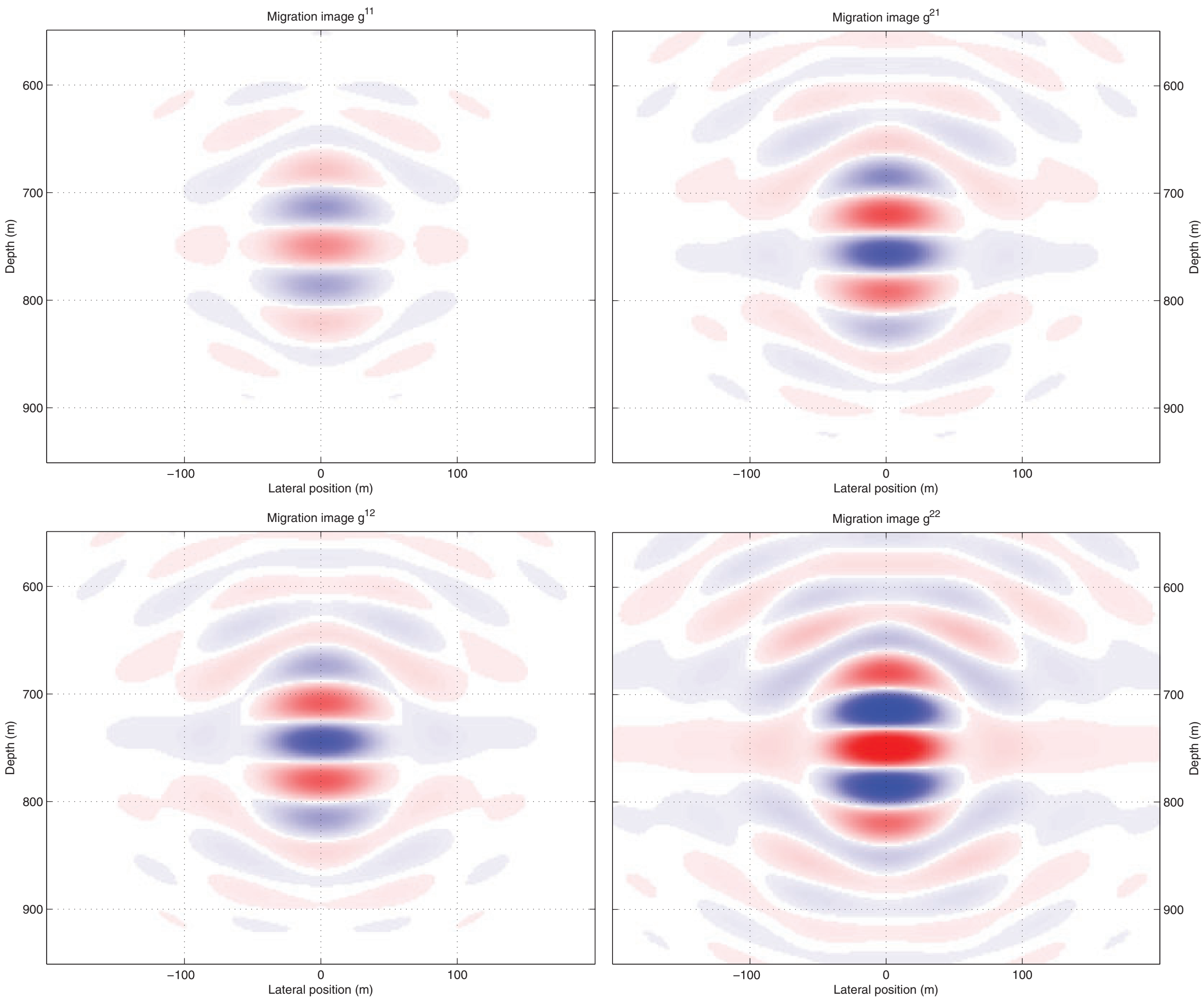


**Figure 1.** Resolution functions for a delta-function perturbation of the first model parameter (top row) and the second model parameter (bottom row), all plotted with the same colour scale. Ideally, the pictures on the diagonal should resemble a band-limited version of a spike at the point scatterer, as they do, and the off-diagonal pictures should have much smaller amplitudes, which they do not. Note that $g^{21}$ differs from $g^{12}$.

Non-linear inversion of the least-squares functional that measures the misfit between modelled and observed data was carried out by the L-BFGS method (Byrd *et al.* 1995). The unknowns where parametrized by $\sigma^2$ and $\sigma^2 Q^{-1}$, depending on subsurface position. The required gradient of the functional with respect to the model parameters was derived by the adjoint-state method, applied to the discretized equations. Details are given in Appendix A. For regularization purposes, we added a penalty term to the least-squares error functional that is a measure of the differences between the current estimated model and the given initial background model. The relative weight of this term was made very small.

## 3 LINEARIZED INVERSION IN CONSTANT BACKGROUND

We start with the same example as in our earlier paper (Hak & Mulder 2010): a point scatterer in a 2-D constant background model with velocity $c_b = 2000\,\mathrm{m\,s^{-1}}$ and quality factor $Q_b = 100$. The scatterer is located at $x = 0$ and $z = 750\,\mathrm{m}$. Sources are placed between $-1887.5$ and $+1887.5\,\mathrm{m}$ at a 25-m interval and receivers between $-1900$ and $+1900\,\mathrm{m}$ at the same interval, all at zero depth. We computed the Hessian on a grid with $101 \times 101$ points and a spacing of 4 m, having the scatterer at the centre. This resulted in a

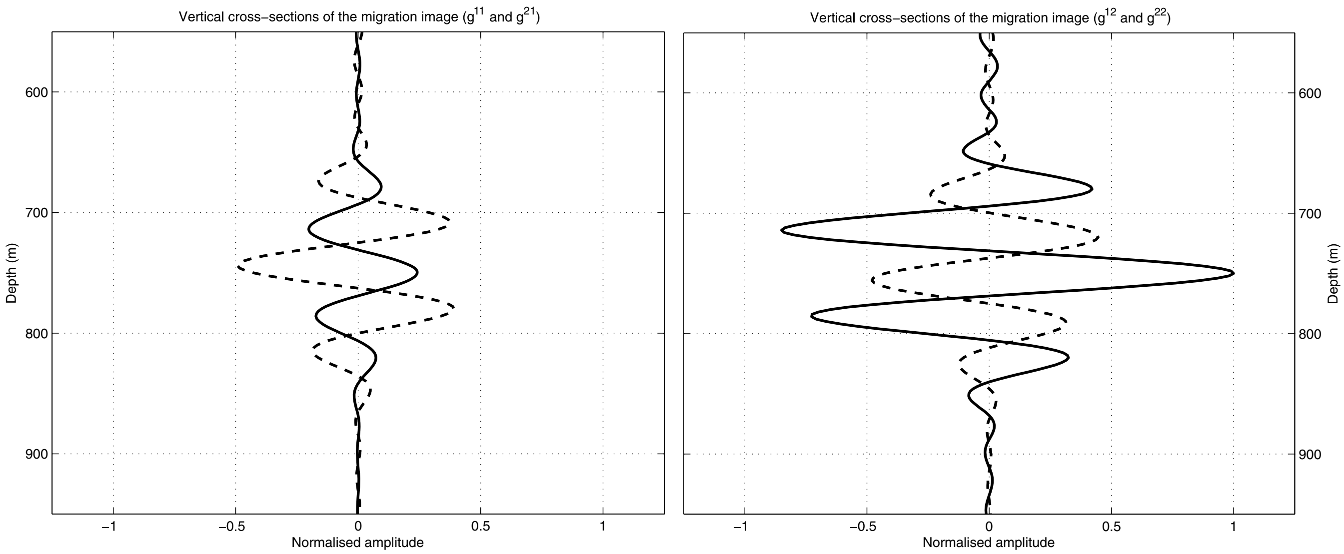


**Figure 2.** Vertical cross-sections of the resolution function through the scattering point. The left-hand panel corresponds to a delta-function perturbation of the first component, the right one to a perturbation of the second component. The solid line represents the perturbed component, the dashed line is its contribution in the unperturbed component ($g^{21}$ and $g^{12}$ for the left- and right-hand panel, respectively). There is a significant contribution (crosstalk) in the unperturbed component, similar to the case of a scatterer without a causal term. The shape of the crosstalk, however, is rather different from the one obtained without the causal term, where there is a 90° phase difference. Here, the crosstalk has almost the opposite phase.

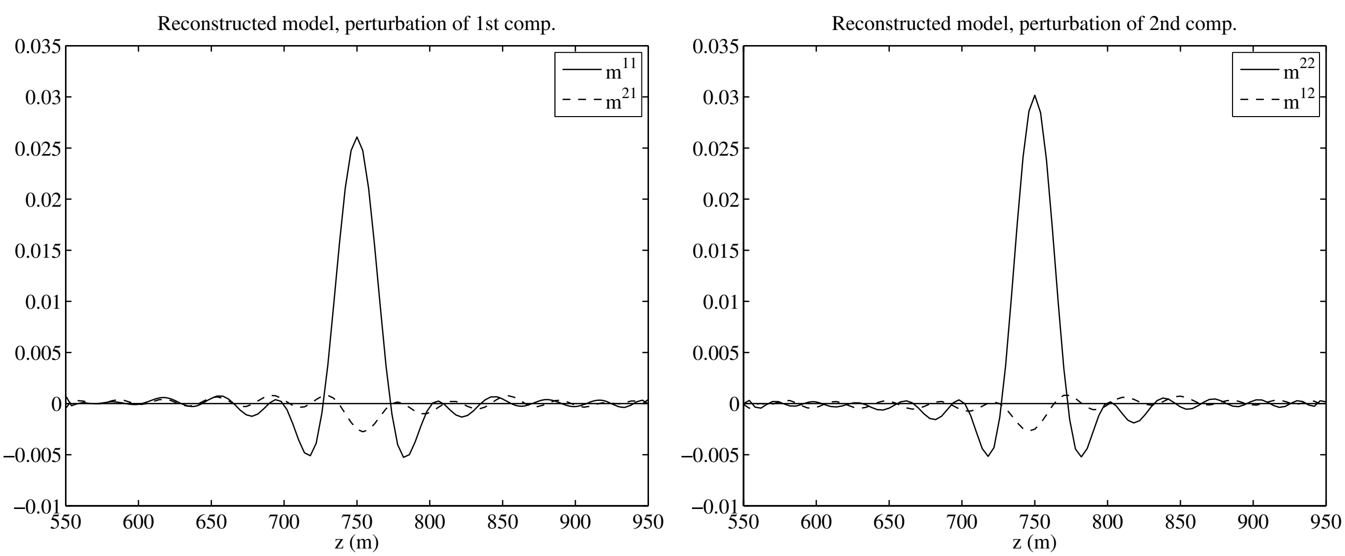


**Figure 3.** Vertical cross-sections through the scattering point of the reconstructed perturbation. The left-hand panel corresponds to a delta-function perturbation of the first component, the right one to a perturbation of the second component. The solid line represents the reconstruction of the perturbed component, the dashed line is its contribution (crosstalk) in the unperturbed component ($m^{21}$ and $m^{12}$ for the left- and right-hand panel, respectively). In both cases, a band-limited reconstruction is provided. A remnant in the unperturbed component is still present, but is considerably smaller than in the case of a scatterer without a causal term.

Hessian matrix of size 20 402 × 20 402. We used a Ricker wavelet with a peak frequency of 15 Hz and considered frequencies in the range between 4 and 20 Hz at a 0.5-Hz interval.

Fig. 1 shows the resolution functions, all on the same colour scale, with red for positive and blue for negative values. A delta-function perturbation of the first component produces a band-limited version, $g^{11}$, after migrating the data, but also produces a substantial contribution in the other component, $g^{21}$, with an even larger amplitude. Similar crosstalk occurs for a delta-function perturbation of the second component. Its migration image, $g^{22}$, looks reasonable, but there is substantial contribution to $g^{12}$, although with a smaller amplitude than $g^{21}$. Fig. 2 displays vertical cross-sections through the scatterer point of the resolution functions for a delta-function perturbation of the first (left-hand panel) and second component (right-hand panel). Compared to the case of a scatterer with causal correction term, presented elsewhere (Hak & Mulder 2010), we note that the crosstalk into the wrong component is still substantial, but now has a phase difference of almost 180° instead of 90°.

Because the crosstalk is still large, simultaneous imaging of velocity and attenuation perturbations will lead to erroneous results if just a single step is carried out. However, if we try to solve eq. (9) by iterative migration, the result may improve. In the current example, we directly computed the pseudo-inverse of the Hessian and obtained the reconstructions shown in Fig. 3. The left-hand panel displays a vertical cross-section obtained for a delta-function perturbation of the first component, the right-hand panel for a perturbation of the second. The inversion has considerably reduced the amount of crosstalk and provided an acceptable band-limited reconstruction of the scatterer.

The results show that inclusion of the frequency-dependent causal correction term in the scatterer provides sufficient additional information to enable a better reconstruction. This agrees with the work of Ribodetti & Hanyga (2004) and Innanen & Weglein (2007). However, in a iterative migration approach, the substantial crosstalk in the first iteration suggests that a large number of iterations will be required before the crosstalk is reduced to an acceptable level. Indeed,

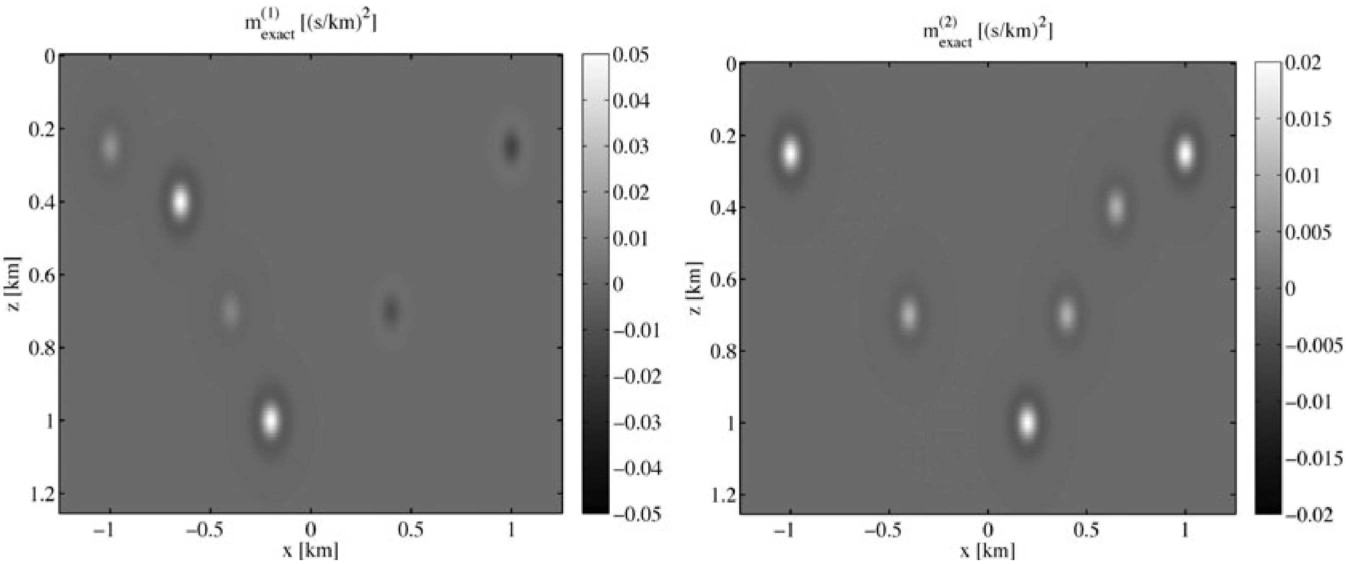


**Figure 4.** First (left-hand panel) and second (right-hand panel) component of the given model after subtraction of the background model.

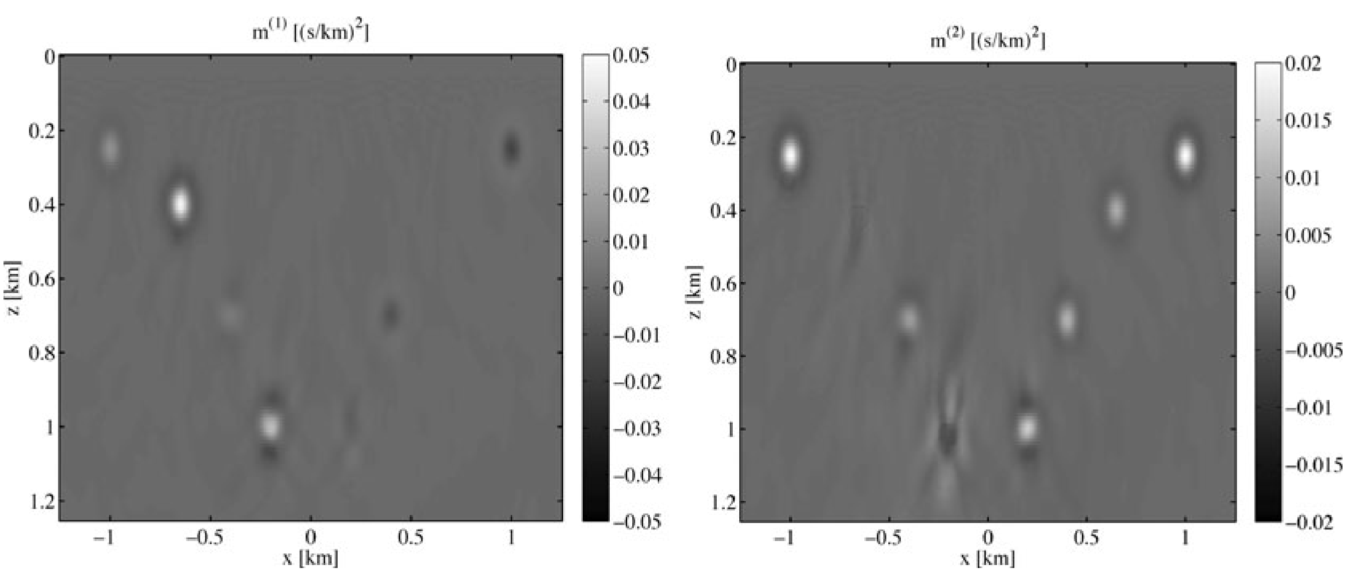


**Figure 5.** First (left-hand panel) and second (right-hand panel) component of the reconstructed perturbation model after about 30 000 iterations.

we found that the solution of eq. (9) by a conjugate-gradient method without pre-conditioning took of the order of $10^4$ iterations to obtain the same result as in Fig. 3. More precisely, 11 445 iterations were required for a decrease of the residual by a factor $10^{-7}$. With pre-conditioning by a diagonal matrix with 1 for the first unknown, $m^{(1)}$, and $Q_b$ for the second unknown, $m^{(2)}$, at each gridpoint the number was reduced to 8016. This is still substantial and perhaps too large to be of practical use in applications of iterative migration (Østmo *et al.* 2002; Mulder & Plessix 2004).

## 4 NON-LINEAR INVERSION

### 4.1 Constant velocity model

Given the severe limitations of the Born approximation, non-linear inversion is more natural. We take a homogeneous background model with a velocity $c_b = 1500\,\mathrm{m\,s^{-1}}$ and quality factor $Q_b = 100$ and add a number of sombrero-shaped functions to it. We take this sombrero function as minus the Laplace operator applied to a Gaussian, or

$$(1-\eta)\exp(-\eta), \text{ with } \eta = \frac{(x-x_0)^2 + (z-z_0)^2}{2\sigma_g^2}, \tag{10}$$

which in 1-D would have the shape of a Ricker wavelet. Fig. 4 shows $\delta m^{(1)}$ and $\delta m^{(2)}$ as defined in eq. (5), the difference between the model and the background represented by the squared slowness and the squared slowness divided by the quality factor, respectively. For the sombrero function, we chose a standard deviation $\sigma_g$ of 40 m and used various amplitudes. The peak value for the first component, shown in the left-hand panel of Fig. 4, was 5 $10^{-8}$ $(\mathrm{s\,m^{-1}})^2$. The second component, shown on the right-hand panel, had a peak value of 2 $10^{-8}$ $(\mathrm{s\,m^{-1}})^2$. The amplitudes of the sombrero functions for each component can be deduced from the vertical cross-sections in the figures to follow.

We generated synthetic data for the constant-density visco-acoustic wave equation with the frequency-domain finite-difference code mentioned earlier, without and with the correction term for causality. Absorbing boundary conditions were imposed on all four sides of the computational domain. We used the same acquisition

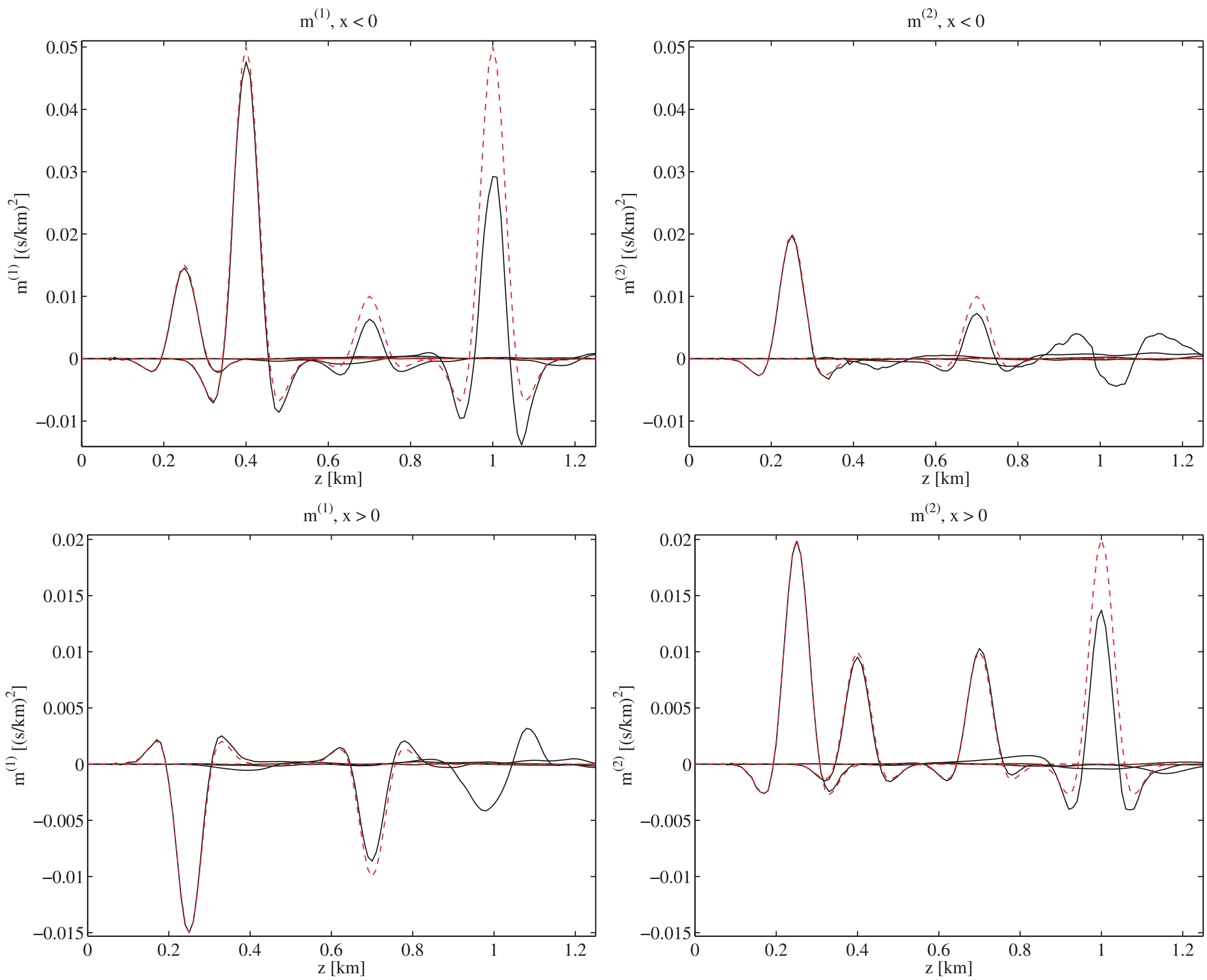


**Figure 6.** Vertical cross-sections of the reconstructed model after about 30 000 iterations. The left-hand panels show the first component, those on the right the second component. The top row contains sections for $x = -1000$, $-650$, $-400$ and $-200$ m and the bottom for $x = 200$, 400, 650 and 1000 m. The solid lines represent the reconstructed model perturbations, the true model perturbations are drawn as red dashed lines.

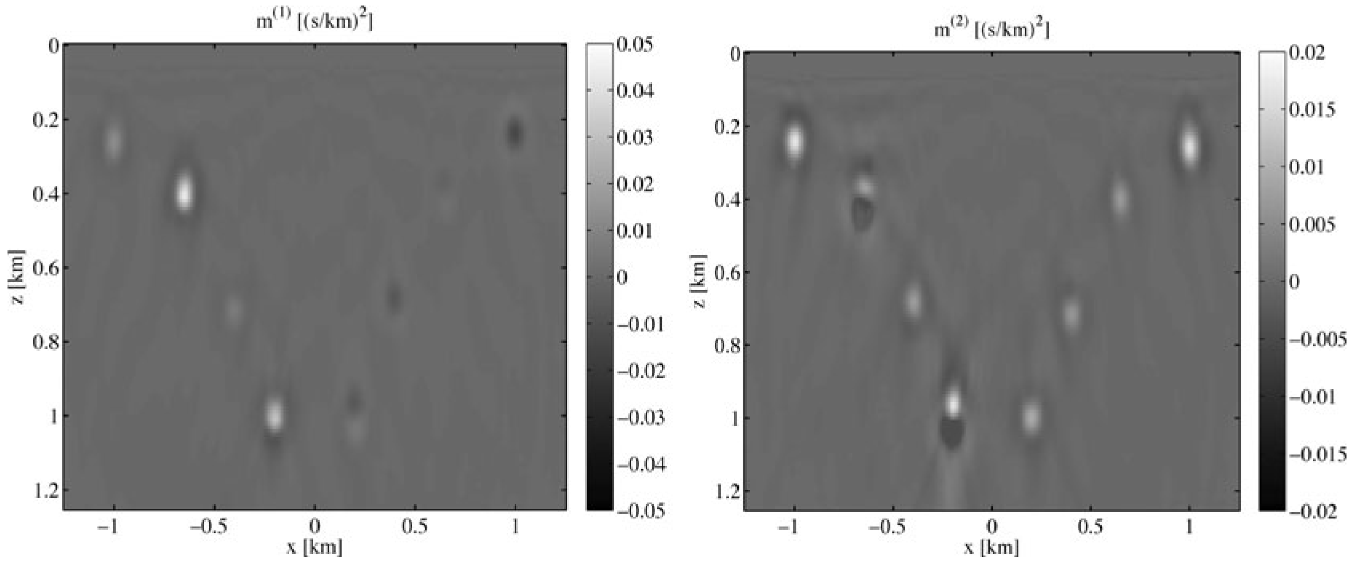


**Figure 7.** First (left-hand panel) and second (right-hand panel) component of the reconstructed perturbation model after about 60 000 iterations. The correction for causality was not applied in the data generation or the inversion.

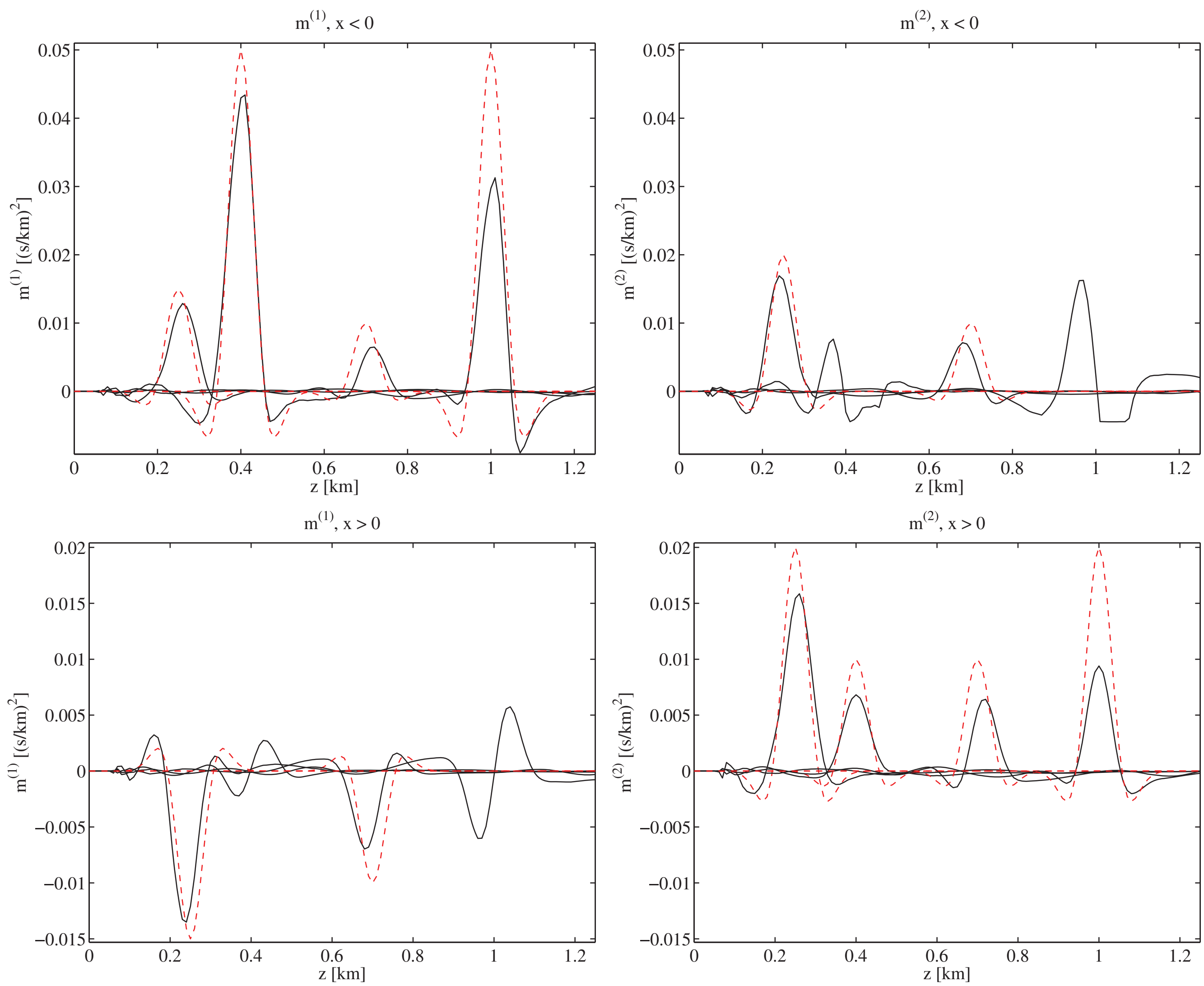


**Figure 8.** As Fig. 6, but without applying the corrections for causality in the data generation or the inversion and after about 60 000 iterations.

geometry and wavelet as in the earlier example. A different computational grid was used with a 10-m grid spacing on a domain with $x$ between $-2500$ and 2500 m and $z$ between $-500$ and 1500 m. The model parameters were kept fixed for $z < 50$ m.

Fig. 5 shows the reconstructed model after 29 924 iterations, where the causal correction term was applied in the generation of the original data and in the inversion. The least-squares functional $\mathcal{J}$, given in eq. (A1), decreased from its initial value by a factor $10^{-7}$. Vertical cross-sections are displayed in Fig. 6. The figure shows multiple curves, each one corresponding to a vertical line through the centre of one of the sombrero functions. The dashed lines correspond to the true model perturbations. There are separate graphs for the the first and second component of the model perturbations as well as the cross-sections at negative and positive $x$. From the relative positions of the peaks, the reader should be able to find the corresponding scatterer. The shallower part of model is reconstructed accurately, whereas the deeper parts have not converged to the true model.

We repeated this exercise without applying the causal correction in the generation of the original data or in the inversion. The least-squares function $\mathcal{J}$ decreased from its initial value by a factor $10^{-7}$ after 59 904 iterations. Fig. 7 shows the reconstructed model. Vertical cross-sections are displayed in Fig. 8. Compared to the causal case, the deviations from the true model remain substantial even after so many iterations. This suggests that the non-linearity by itself does not fully remove the ambiguity between the first and second component of the model parameter, at least not when the number of iterations is limited, as will happen in practical applications.

### 4.2 Noise

We performed the same numerical experiment for the causal case with noise added to the data. The earlier frequency domain data were transformed to the time domain for a Ricker wavelet with a peak frequency of 15 Hz. We took the maximum amplitude of the first shot, multiplied it by $10^{-4}$ and added random numbers between $-1$ and $+1$ times this amplitude to all the data. The result was then transformed back to the frequency domain.

Fig. 9 shows vertical cross-sections of the reconstructed model after about 5000 iterations. The result is qualitatively correct but the amplitude discrepancies still are substantial. The result after about 30 000 iterations, shown in Fig. 10, is clearly better, but the computational cost is huge. Also, the noise level is higher. This might be improved by adding smoothing penalty terms to the

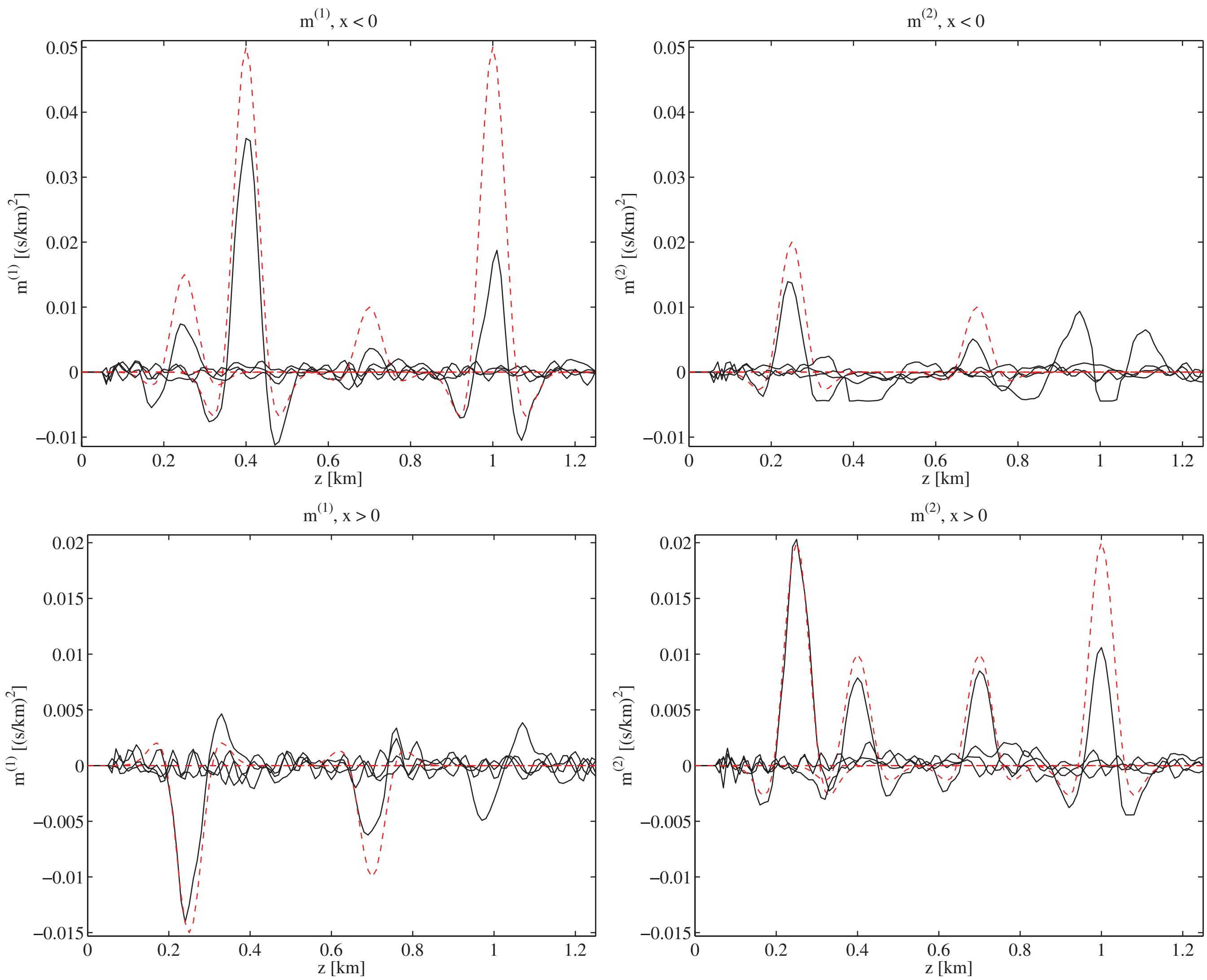


**Figure 9.** As Fig. 6, but for data with a relative noise level of $10^{-4}$ and after about 5000 iterations.

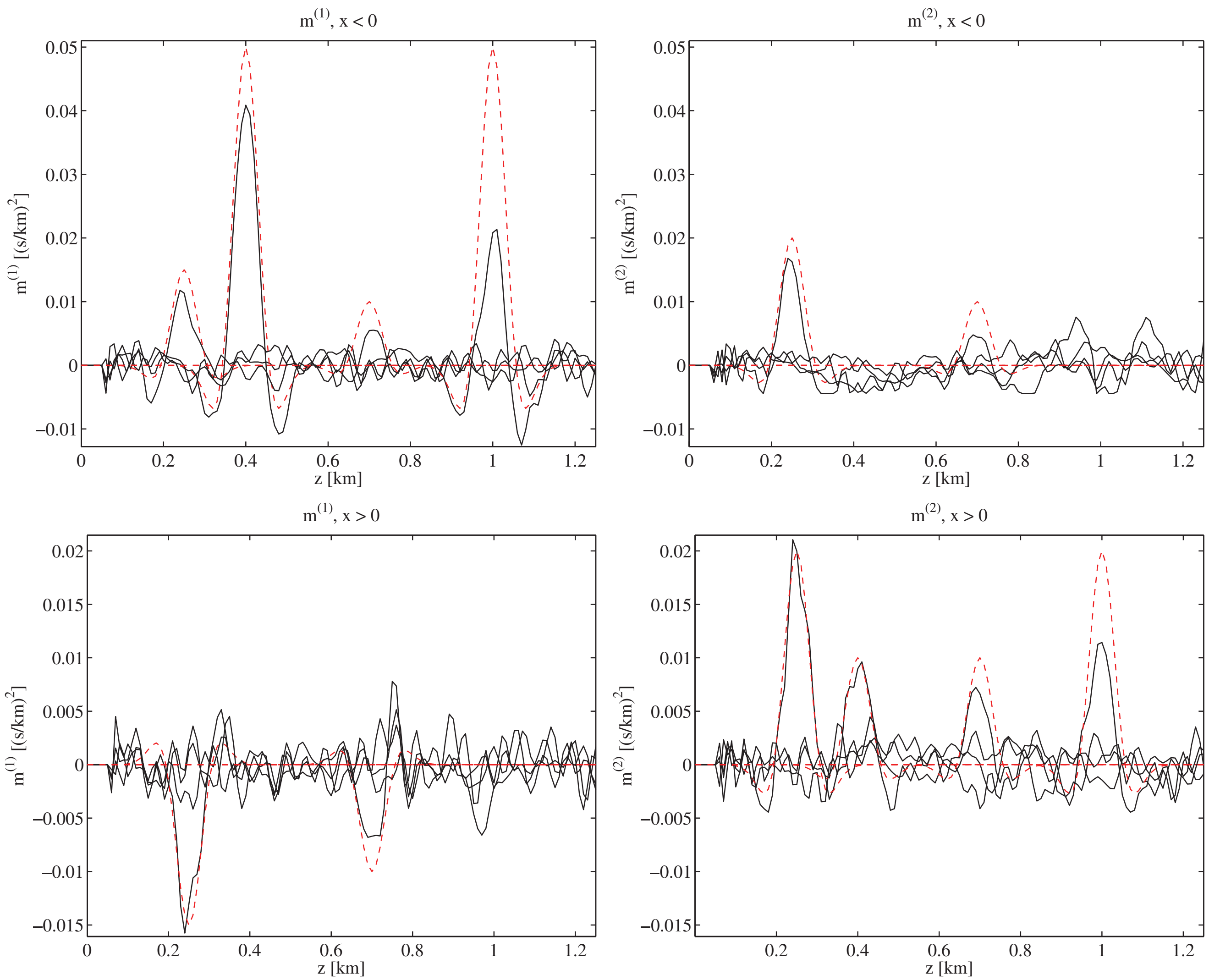


**Figure 10.** As Fig. 9, but after about 30 000 iterations. The reconstruction of the scatterers has improved, but the noise level in the image has increased as well.

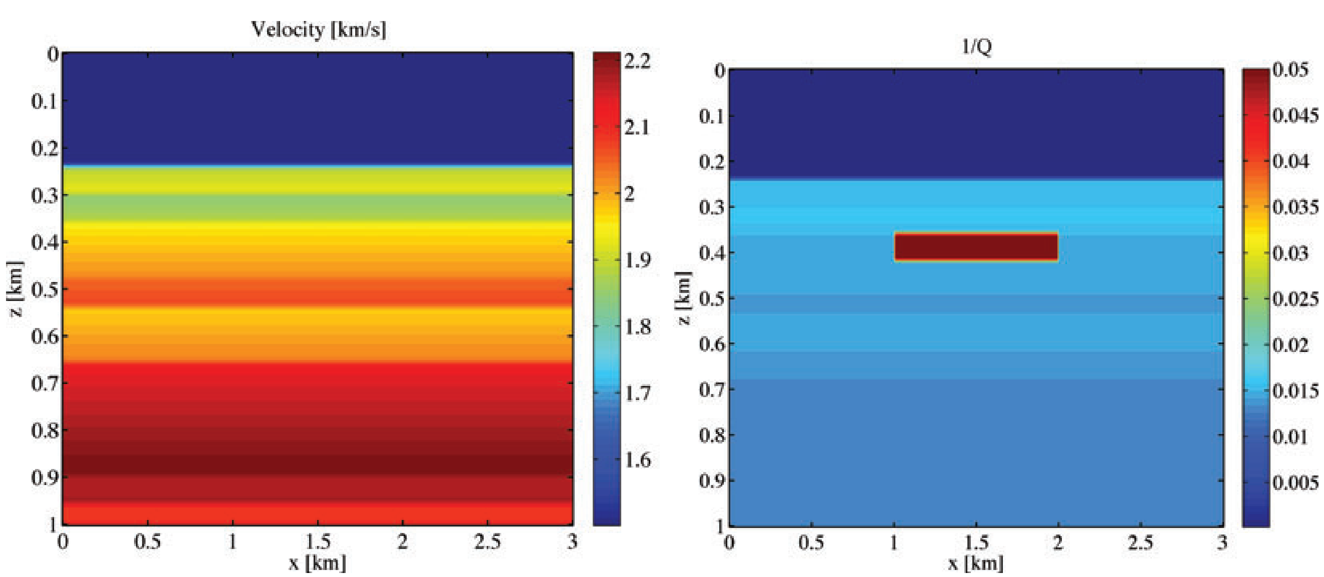


**Figure 11.** Velocity (left-hand panel) and inverse quality factor (right-hand panel) of the given model.

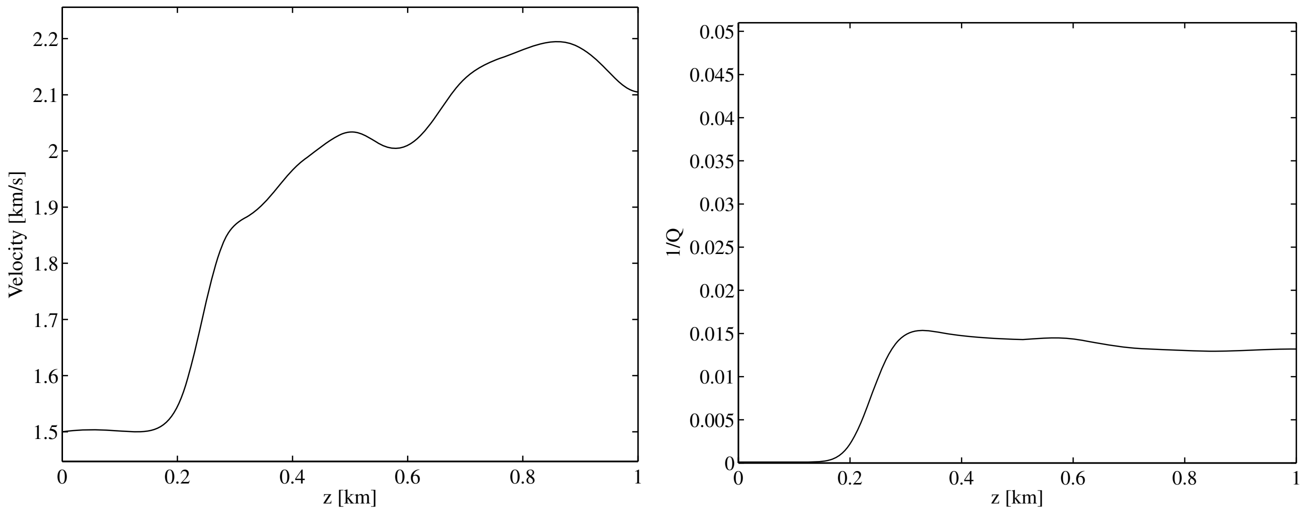


**Figure 12.** Vertical cross-section of the velocity (left-hand panel) and inverse quality factor (right-hand panel) of the initial model.

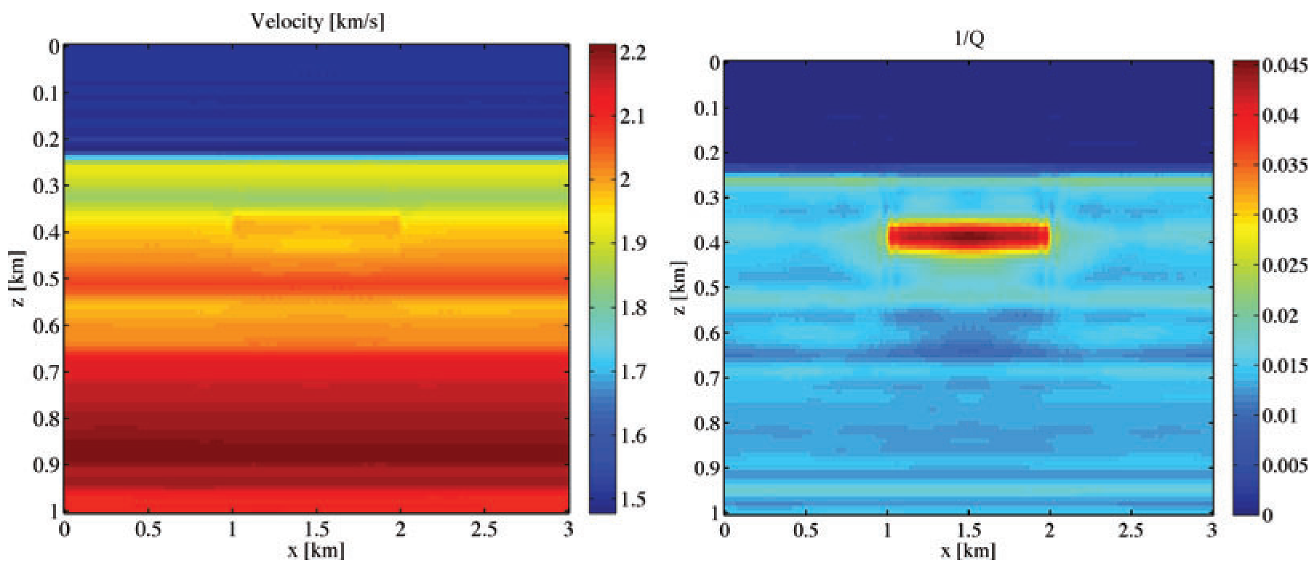


**Figure 13.** Velocity (left-hand panel) and inverse quality factor (right-hand panel) of the reconstructed model after about 8000 iterations.

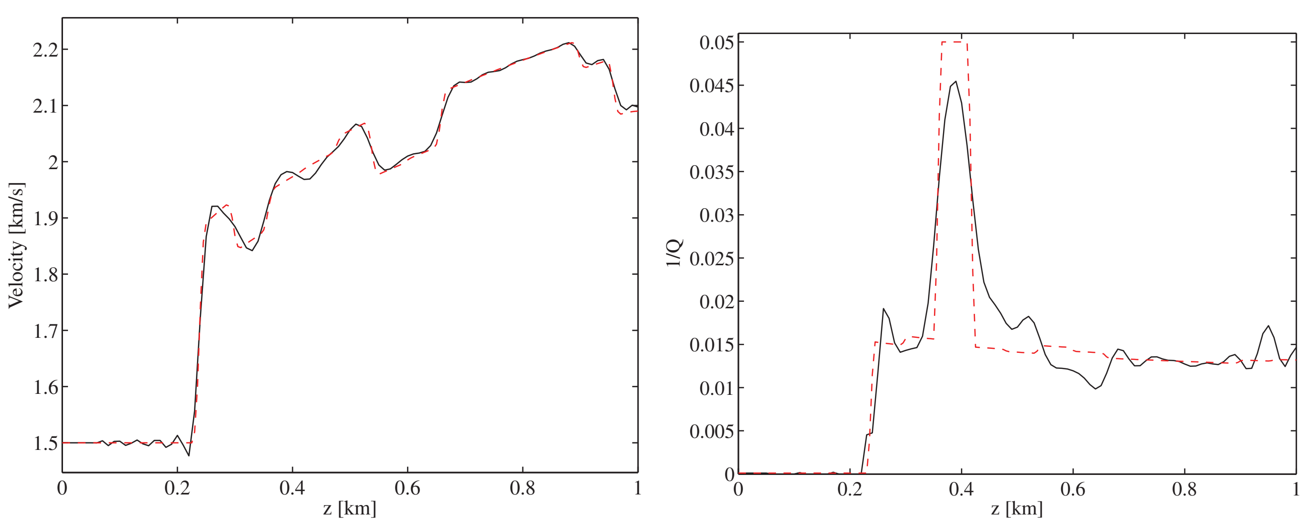


**Figure 14.** Vertical cross-section at $x = 1.5$ km of the reconstructed velocity (left-hand panel) and inverse quality factor (right-hand panel) after about 8000 iterations. The solid line represents the reconstructed model, the true model is drawn as red dashed line.

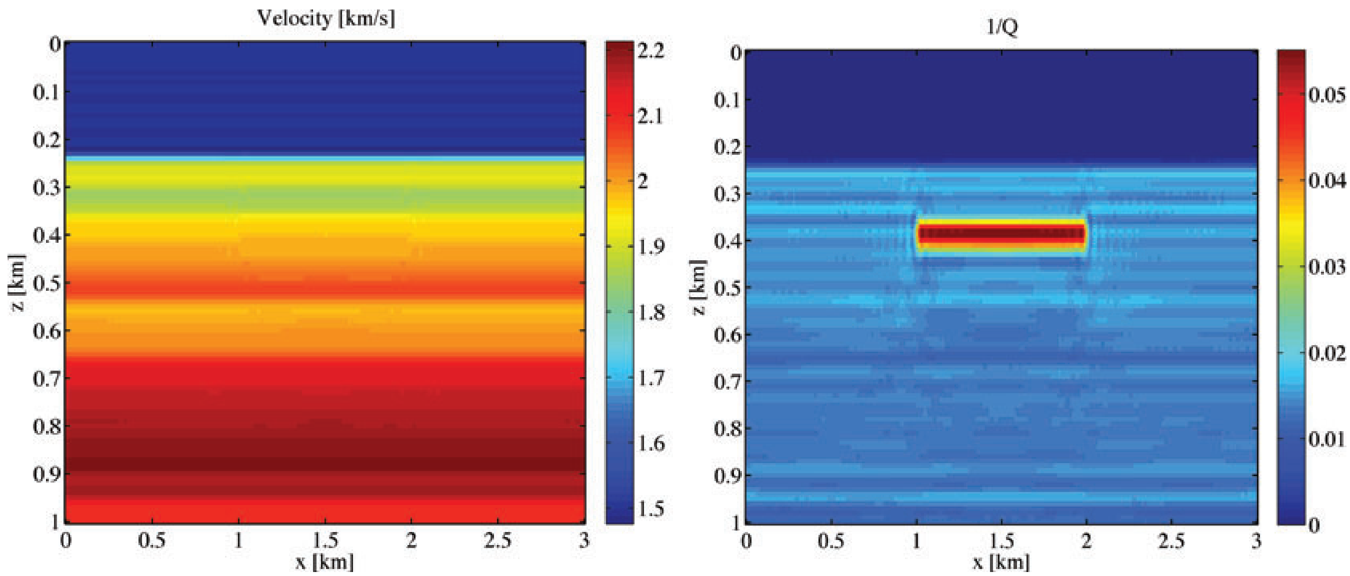


**Figure 15.** Velocity (left-hand panel) and inverse quality factor (right-hand panel) of the reconstructed model after about 50 000 iterations.

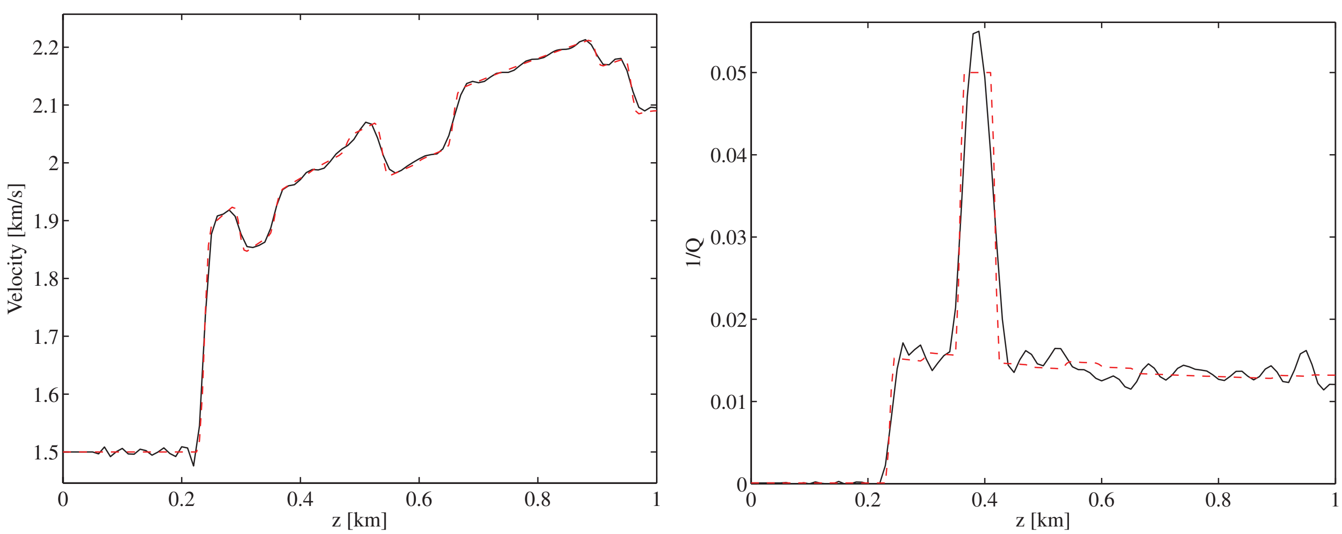


**Figure 16.** Vertical cross-section at $x = 1.5$ km of the reconstructed velocity (left-hand panel) and inverse quality factor (right-hand panel) after about 50 000 iterations. The solid line represents the reconstructed model, the true model is drawn as red dashed line.

least-squares functional. Compared to the acausal case, the reconstruction result appears to be better.

### 4.3 Marine model

Fig. 11 shows a geophysically slightly more realistic model than used so far. Below the sea bottom, the velocity is horizontally layered. A rectangular region with low quality factor, as may happen with shallow gas, is included. Fig. 12 displays a vertical cross-section of a smoothed version of the model without the strongly attenuating rectangular region, which served as an initial guess for the non-linear inversion. For our computations, we used a grid spacing of 10 m. Causality was included.

Fig. 13 shows the result of the reconstruction after 8164 iterations, when the least-squares cost function had decreased by a factor $1.1\ 10^{-5}$ from its initial value. The velocity still shows an imprint of the rectangular region. Fig. 14 displays vertical cross-sections through the centre.

We continued the iterations to a total of 49 363 and a corresponding decrease of the least-squares cost function by a factor $9.3\ 10^{-7}$. Figs 15 and 16 show an improvement over the earlier result. The imprint of the strongly attenuating region on the reconstructed velocity model has almost disappeared and the quality factor is closer to the true model. We conclude that the inversion produces the correct result, but at a discouragingly large cost.

## 5 CONCLUSIONS

We have investigated how two methodologies are affected by an ambiguity that occurs for attenuation scattering imaging based on the Born approximation. The first one is the inclusion of the correction for causality in the scatterer. This introduces a frequency-dependent dispersion term. When using a range of frequencies, this term almost, but not completely, removes the ambiguity that otherwise occurs between the real and imaginary parts of the scattering model.

Because the Born approximation offers no handle on the energy of the scattered data and is invalid for larger scattering angles when, for instance, flat scatterers are considered, a non-linear approach is more natural and we took this as a second methodology. In an example, we observed that non-linearity is not able to completely remove the ambiguity if the causal correction term is dropped, at least not for a large but limited number of iterations. With the causal term, the reconstruction succeeds.

In both cases, the initial update of a gradient-based optimization algorithm that tries to minimize the difference between modelled and observed data differs significantly from the correct update. Many iterations are therefore required to approach the true solution. We observed the same when adding noise and for a slightly more realistic example, both with causality included. This implies that, although causality enabled us to resolve the ambiguity in attenuation imaging in the examples considered here, it will be difficult to actually find the true model in realistic cases when noise is present and the number of iterations has to be limited because of computational cost.

## ACKNOWLEDGMENTS

This work is part of the Industrial Partnership Programme (IPP) the Physics of Fluids and Sound Propagation (I06) of the 'Stichting voor Fundamenteel Onderzoek der Materie (FOM)', which is financially supported by the 'Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO)'. The IPP is co-financed by 'Sichting Shell Research'.

## REFERENCES

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, 2nd edn, University Science Books, Sausalito, CA.

Byrd, R.H., Lu, P., Nocedal, J. & Zhu, C., 1995. A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.*, **16**(5), 1190–1208.

George, A. & Liu, J.W.H., 1981. *Computer Solution of Large Sparse Positive Definite Systems*, Prentice-Hall, Englewood Cliffs, NJ.

Hak, B. & Mulder, W.A., 2010. Migration for velocity and attenuation perturbations, *Geophys. Prospect.*, **58**(6), 939–952.

Harari, I. & Turkel, E., 1995. Accurate finite difference methods for time-harmonic wave propagation, *J. Comput. Phys.*, **119**(2), 252–270.

Hicks, G.J. & Pratt, R.G., 2001. Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit, *Geophysics*, **66**(2), 598–612.

Innanen, K.A. & Weglein, A.B., 2007. On the construction of an absorptive-dispersive medium model via direct linear inversion of reflected seismic primaries, *Inverse Probl.*, **23**(6), 2289–2310.

Jo, C.-H., Shin, C. & Suh, J.H., 1996. An optimal 9-point, finite-difference, frequency-space, 2-D scalar wave extrapolator, *Geophysics*, **61**(2), 529–537.

Kamei, R. & Pratt, R.G., 2008. Waveform tomography strategies for imaging attenuation structure with cross-hole data, 70th EAGE Conference & Exhibition incorporating SPE EUROPEC, Rome, Italy, Extended Abstracts, F019.

Marfurt, K.J. & Shin, C.S., 1989. The future of iterative modeling in geophysical exploration, in *Supercomputers in Seismic Exploration*, ed. Eisner, E., Pergamon Press, Oxford.

Mulder, W.A. & Hak, B., 2009. An ambiguity in attenuation scattering imaging, *Geophys. J. Int.*, **178**(3), 1614–1624.

Mulder, W.A. & Plessix, R.-E., 2004. A comparison between one-way and two-way wave-equation migration, *Geophysics*, **69**(6), 1491–1504.

Østmo, S., Mulder, W.A. & Plessix, R.-E., 2002. Finite-difference iterative migration by linearized waveform inversion in the frequency domain, SEG 72nd Annual Meeting, Salt Lake City (UT), USA, Expanded Abstracts, 1384–1387.

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

Ribodetti, A. & Hanyga, A., 2004. Some effects of the memory kernel singularity on wave propagation and inversion in viscoelastic media – II. Inversion, *Geophys. J. Int.*, **158**(2), 426–442.

Ribodetti, A. & Virieux, J., 1998. Asymptotic theory for imaging the attenuation factor Q, *Geophysics*, **63**(5), 1767–1778.

Ribodetti, A., Operto, S., Virieux, J., Lambaré, G., Valéro, H.P. & Gibert, D., 2000. Asymptotic viscoacoustic diffraction tomography of ultrasonic laboratory data: a tool for rock properties analysis, *Geophys. J. Int.*, **140**(2), 324–340.

Ribodetti, A., Plessix, R.-E., Operto, S. & Virieux, J.M., 2007. Viscoacoustic frequency-domain full-waveform inversion: application to numerical VSP data, 69th EAGE Conference & Exhibition Incorporating SPE EUROPEC, London, Extended Abstracts, C029.

Smithyman, B.R., Pratt, R.G., Hayles, J.G. & Wittebolle, R.J., 2009. Detecting near-surface objects with seismic waveform tomography, *Geophysics*, **74**(6), WCC119–WCC127.

## APPENDIX A: GRADIENT OF THE LEAST-SQUARES FUNCTIONAL

The gradient of the least-squares functional that measures the difference between synthetic and observed data can be derived by the adjoint-state method. The derivation for complex parameters, though straightforward, is somewhat less obvious than that for real-valued parameters, listed for instance by Plessix (2006), so we present it below.

In the frequency domain, the least-squares error functional for the non-linear problem is

$$\mathcal{J} = \frac{1}{2} \sum_{\omega,\mathrm{s},\mathrm{r(s)}} |\varepsilon_{\mathrm{r(s)}}(\omega)|^2, \tag{A1}$$

$$\varepsilon_{\mathrm{r(s)}}(\omega) = w(\omega) \int \mathrm{d}\boldsymbol{x}\, \mathbf{R}_{\mathrm{r(s)}} \overline{p}(\omega, \boldsymbol{x}) - p^{\mathrm{obs}}_{\mathrm{r(s)}}(\omega), \tag{A2}$$

with $p^{\mathrm{obs}}_{\mathrm{r(s)}}$ the observed data and $\overline{p}$ the modelled wavefield obtained for a temporal delta function at zero time and a spatial delta source at $\boldsymbol{x}_\mathrm{s}$. Here s indexes shot positions and r(s) indexes receivers positions for shot s. We have included a wavelet $w$ and a real-valued sampling operator $\mathbf{R}$ that is a restriction of the modelled wavefield to the receiver location. If the wavelet is the same for all shots, minimization of $\mathcal{J}$ with respect to the wavelet leads to

$$w(\omega) = \frac{\sum_{\mathrm{s},\mathrm{r(s)}} \left[\int \mathrm{d}\boldsymbol{x}\, \mathbf{R}_{\mathrm{r(s)}}\, \overline{p}(\omega, \boldsymbol{x})\right]^* p^{\mathrm{obs}}_{\mathrm{r(s)}}(\omega)}{\sum_{\mathrm{s},\mathrm{r(s)}} \left|\int \mathrm{d}\boldsymbol{x}\, \mathbf{R}_{\mathrm{r(s)}}\, \overline{p}(\omega, \boldsymbol{x})\right|^2}. \tag{A3}$$

This step can be skipped if the wavelet is known and specified in advance.

To determine the gradient of $\mathcal{J}$ with respect to the model parameters, we can use the adjoint-state method. For brevity, we only consider a single shot at a single frequency such that we can drop the indexing subscripts and writing dependency on the frequency. The gradients for all frequencies and shots can be obtained by summation in the end.

Let the wave equation for a temporal delta function at zero time and a spatial delta source at $\boldsymbol{x}_s$ be given by $\mathbf{P} = \mathbf{L}\overline{p} - \delta(\boldsymbol{x} - \boldsymbol{x}_s) = 0$, where the operator $\mathbf{L} = -\omega^2 m - \Delta$. We choose a Lagrangian

$$\mathcal{L} = \mathcal{J} - \int \mathrm{d}\boldsymbol{x}\, q_r(\boldsymbol{x})\, \mathbf{P}_r(\boldsymbol{x}) - \int \mathrm{d}\boldsymbol{x}\, q_i(\boldsymbol{x})\, \mathbf{P}_i(\boldsymbol{x}), \tag{A4}$$

where the subscripts $r$ and $i$ denote the real and imaginary parts, respectively, and $q_r$ and $q_i$ are the multipliers. Let $m = m_r + \mathrm{i}m_i$, such that $\mathbf{L} = \mathbf{L}_r + \mathrm{i}\mathbf{L}_i$. Stationarity of $\mathcal{L}$ with respect to $\overline{p} = \overline{p}_r + \mathrm{i}\overline{p}_i$ leads to

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \overline{p}_r} &= \mathrm{Re}\left[\sum_{\mathrm{r(s)}} w^* \mathbf{R}^{\mathrm{T}} \varepsilon\right] - \left(\mathbf{L}_r^{\mathrm{T}} q_r + \mathbf{L}_i^{\mathrm{T}} q_i\right) = 0,\\ \frac{\partial \mathcal{J}}{\partial \overline{p}_i} &= \mathrm{Im}\left[\sum_{\mathrm{r(s)}} w^* \mathbf{R}^{\mathrm{T}} \varepsilon\right] - \left(\mathbf{L}_r^{\mathrm{T}} q_i - \mathbf{L}_i^{\mathrm{T}} q_r\right) = 0,\end{aligned} \tag{A5}$$

which, defining $q = q_r + \mathrm{i}q_i$, can be summarized as

$$\mathbf{L}^{\mathrm{H}} q = \sum_{\mathrm{r(s)}} w^* \mathbf{R}^{\mathrm{T}} \varepsilon. \tag{A6}$$

Let $g_r = \partial\mathcal{J}/\partial m_r$ and $g_i = \partial\mathcal{J}/\partial m_i$ be the gradient of the least-squares functional with respect to the real and imaginary part of $m$, respectively. Then

$$\begin{aligned} g_r &= (\mathbf{L}_{rr}\, p_r) q_r - (\mathbf{L}_{ir}\, p_i) q_r + (\mathbf{L}_{rr}\, p_i) q_i + (\mathbf{L}_{ir}\, p_r) q_i,\\ g_i &= -(\mathbf{L}_{ir}\, p_r) q_r - (\mathbf{L}_{rr}\, p_i) q_r - (\mathbf{L}_{ir}\, p_i) q_i + (\mathbf{L}_{rr}\, p_r) q_i,\end{aligned} \tag{A7}$$

where $\mathbf{L}_{rr} = \partial\mathbf{L}_r/\partial m_r$ and $\mathbf{L}_{ir} = \partial\mathbf{L}_i/\partial m_r$, and the Cauchy-Riemann relations imply that $\mathbf{L}_{rr} = \mathbf{L}_{ii} = \partial\mathbf{L}_i/\partial m_i$ and $\mathbf{L}_{ir} = -\mathbf{L}_{ri} = -\partial\mathbf{L}_r/\partial m_i$. Using $g = g_r + \mathrm{i}g_i$ the above can be summarized as

$$g = \left(\frac{\partial \mathbf{L}}{\partial m_r} p\right)^* q = (\omega^2 p)^*\, q. \tag{A8}$$

This includes cases with complex $\omega$.

When the correction term for causality is non-zero, we need another step. For $a = (2/\pi)\log(\omega/\omega_r)$, we have $m = \sigma^2[1 + (\mathrm{i} - a)/Q]$. If we choose $m^{(1)} = \sigma^2$ and $m^{(2)} = \sigma^2 Q^{-1}$ as unknowns, the corresponding gradient can be computed with the chain rule, resulting in

$$\begin{pmatrix} \dfrac{\partial \mathcal{J}}{\partial\, m^{(1)}} \\[2ex] \dfrac{\partial \mathcal{J}}{\partial\, m^{(2)}} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -a & 1 \end{pmatrix} \begin{pmatrix} g_r \\ g_i \end{pmatrix}. \tag{A9}$$

A similar approach can be taken for the Born approximation in eq. (3). The wave equation can be chosen as $\delta\mathbf{P} = \mathbf{L}_b \delta p - \omega^2 \delta m p_b = 0$ and the corresponding Lagrangian may be written as

$$\mathcal{L}(\boldsymbol{x}) = \mathcal{J}(\boldsymbol{x}) - \int \mathrm{d}\boldsymbol{x}\, q_r(\boldsymbol{x})\, \delta\mathbf{P}_r(\boldsymbol{x}) - \int \mathrm{d}\boldsymbol{x}\, q_i(\boldsymbol{x})\, \delta\mathbf{P}_i(\boldsymbol{x}). \tag{A10}$$

The adjoint equation is similar to eq. (A6), but now the right-hand side is based on primaries and the operator is the one for the background, $\mathbf{L}_b$, applying to the background model only. The gradient with respect to $\delta m$ is similar to the one before, $g = (\omega^2 p_b)^*\, q$, and the correction for the causal term of eq. (A9) can be applied if necessary.

The Hessian for the non-linear inversion problem becomes the same as the Hessian for the Born approximation if the Gauss–Newton approximation is used, that is, if we assume that we are sufficiently close to the true model that the data differences, $\varepsilon$, and related expressions, such as $q$ in eq. (A6), can be neglected.