Automatic velocity analysis with Reverse Time Migration

Wiktor Waldemar Weibull[1] *and Børge Arntsen*[1]

ABSTRACT

We derive a reverse-time migration based Differential Semblance objective function and its gradient with respect to seismic velocity. The Differential semblance objective is a function of common image point gathers constructed with reverse-time migration and parametrized by subsurface half-offset. The gradient is necessary for non-linear local optimization. The result of the optimization should be offset domain common image point gathers focused at zero lag. Which, under certain assumptions, are an indication of kinematically correct background velocites. We test the method in two 2D synthetic models, which prove the ability of the method to converge to accurate background velocities even from relatively vague initial estimates.

INTRODUCTION

An accurate estimate of the distribution of the subsurface seismic velocities is an indispensable component for obtaining an accurate image of Earth's reflectivity by prestack depth migration. Methods for velocity estimation in routine use today are often based on ray tomography and, although usually semi-automatic, requires time consuming picking of gathers and quality control. Furthermore, they generally only produce approximate solutions, and can fail to be adequately accurate when the velocity field is complex with strong and sharp spatial variations. Due to the shortcomings of ray-theoretical depth migration approaches in areas with complex geology (Arntsen et al., 2009), one-way and more recently two-way wave equation imaging methods have become popular. These approaches are often combined with velocity estimation based on ray-theory. For consistency and improved resolution, velocity building should preferably be based on wave equation methods.

Wave equation migration velocity analysis (WEMVA) is based on focusing of migrated data and uses an automatic optimization procedure to estimate the velocity field, avoiding manual picking. The approach is based on formulating an objective function measuring to what extent offset- or angle-gathers are, respectively, focused or flattened, and then minimizing the object function with respect to the velocity field.

Chavent and Jacewitz (1995) implemented Wave-Equation Migration Velocity Analysis (WEMVA) by using a similarity-index and reverse-time migration (RTM) to compute the velocity field. The procedure utilizes the complete wavefield and requires no picking. Biondi and Sava (1999) used one-way migration operators and image perturbations for computing corrections to the initial wavefield and Sava and Biondi (2004) extended this approach to a fully non-linear iterative scheme. The numerical implementation is described in detail by Sava and Vlad (2008).

Shen et al. (2003) used the Double Square Root approach to depth migration and an objective function based on Differential Semblance (Symes and Carazzone, 1991) to esti-

mate the velocity field. The approach was also extended to shot-profile migration based on one-way migration operators (Shen and Symes, 2008). Mulder (2008) used depth migration based on the two-way wave equation in the frequency-domain and an objective function related to the differential-semblance approach to implement a non-linear scheme for computing the velocity field. Gao and Symes (2009) proposed to use a differential semblance cost function and RTM to solve the velocity estimation problem, and also gave an initial theoretical framework.

We give a new objective function for wave-equation migration velocity analysis using differential semblance and a spatial differentiation operator. This choice of object function minimizes problems related to amplitude-sensitivity caused by two-way wave propagation. We also give complete expressions for computing the gradient of the cost function with respect to the velocity field, and show how the gradient can be used in a full non-linear optimization scheme illustrated with both simple and complex synthetic data examples. We show that the method can produce kinematically accurate models even when the initial estimate of the velocity model is far from the true model. The examples also show that the method can handle large spatial velocity variations.

The first section contains the definition of the objective function, and the equations necessary for implementing the non-linear optimization. We then illustrate the method with examples of the application of our velocity estimation for different synthetic datasets. Finally, we discuss about the applicability of the method to real field data and suggest some measures to improve the performance of the method.

REVERSE-TIME MIGRATION AND DIFFERENTIAL SEMBLANCE OPTIMIZATION (DSO)

In RTM, a Common Image Point gather(CIG), $\mathcal{R}$, can be produced by crosscorrelating a forward modeled source wavefield (D) with a reverse-time modeled scattered wavefield (U)

$$\mathcal{R}(\mathbf{x}, \mathbf{h}) = \sum_s \int_0^T dt \, U(\mathbf{x} + \mathbf{h}, t, s) D(\mathbf{x} - \mathbf{h}, t, s). \tag{1}$$

where $\mathbf{x} = (x_1, x_2, x_3)$ is the subsurface mid-point coordinate and $\mathbf{h} = (h_1, h_2, 0)$ is the subsurface horizontal half-offset.

This imaging condition is an extension of Claerbout's principle (Claerbout, 1971; Rickett and Sava, 2002). According to this principle, given an accurate estimate of the material velocities, the crosscorrelation of the reconstructed source and receiver wavefields will have a maximum at zero lag in time and space. In DSO, we explore this fact to set up a non-linear least squares inversion problem (Symes and Carazzone, 1991). By parametrizing the image with an additional lag parameter we can capture the deviation of the maximum in crosscorrelation from zero lag, and use this to quantify the error in the estimates of the velocities. In this paper we only consider horizontal spatial lags in the crosscorrelation, however vertical spatial lags, and even temporal lags can be incorporated in the image (Biondi and Symes, 2004; Sava and Fomel, 2006). The U and D wavefields are obtained through

$$D(\mathbf{x}, t, s) \quad = \quad \int d\mathbf{x}' \int_0^T dt \, G(\mathbf{x}, t; \mathbf{x}', 0) \sum_{sou=1}^{Nsou} \delta(\mathbf{x}' - \mathbf{x}_{sou}) S(\mathbf{x}', t, s)$$

$$U(\mathbf{x}, t, s) \quad = \quad \int d\mathbf{x}' \int_0^T dt \; G(\mathbf{x}, 0; \mathbf{x}', t) \sum_{rec=1}^{Nrec} \delta(\mathbf{x}' - \mathbf{x}_{rec}) P^{rec}(\mathbf{x}', t, s),$$

where G is the Green's function for the constant-density two-way acoustic wave equation. Also s represents the shot index (one for each seismic experiment), P^{rec} is the recorded pressure data, $\mathbf{x}_{sou}$ are source coordinates for one shot, $\mathbf{x}_{rec}$ are the receiver coordinates for one shot.

An important assumption in WEMVA is that when the velocity model is optimum, the CIGs are maximally focused at zero offset. Thus any deviation from perfect focus can be taken as an indication that the background velocity model must be improved. The differential semblance misfit represents a simple and direct way of quantifying the deviation from focus of CIGs, and can be used to formulate an object function measuring the error in the velocity field

$$\mathcal{DS} = \frac{1}{2} \left\| \mathbf{h} \frac{\partial \mathcal{R}}{\partial x_3}(\mathbf{x}, \mathbf{h}) \right\|^2 = \frac{1}{2} \int d\mathbf{x} \int d\mathbf{h} \; \mathbf{h}^2 \left[\frac{\partial \mathcal{R}}{\partial x_3}(\mathbf{x}, \mathbf{h}) \right]^2. \tag{2}$$

This differential semblance formulation is similar to that in Shen and Symes (2008). It is based on optimization of CIGs parameterized by subsurface half-offset, and is fundamentally different from the pioneering work by Symes and Carazzone (1991), where only surface coordinates are considered. Depth oriented CIGs carry less artifacts than their surface oriented counterparts, and are therefore better suited for velocity analysis (Symes, 2008).

Ideally, minimization of the error measured by the differential semblance object function will lead to an optimally focused image and accurate estimates of the background velocity field. However, because this objective function is a measure of both traveltime and amplitude information, and because both these quantities are sensitive to the velocities, non-optimal solutions are likely to exist. These solutions represent an equilibrium between minimization of image amplitudes and kinematic improvement. In order to avoid the influence of the amplitudes of the image on the solution of the velocity analysis, the sensitivity of the image amplitudes to the velocities needs to be kept to a minimum. This requires an understanding of how the velocities affect the image amplitudes.

Modeling the reconstructed source wavefield with the two-way wave equation generally results in reflection and refraction of the source wavefield at interfaces. These "upgoing waves" will crosscorrelate at low angles with the reconstructed scattered wavefield during imaging and produce a characteristic low wavenumber artifact in the final RTM image. It turns out that this effect can have high amplitudes and is very sensitive to the velocity model, thus dominating the optimization completely. Due to the low wavenumber character, and the predominantly vertical orientation of this effect, a simple vertical derivative filter acting over the image is sufficient to remove it (Guitton et al., 2007). A similar procedure, albeit involving multiplication in the wavenumber domain, has been applied by Mulder (2008) in his implementation of DSO using the two-way wave equation.

Excessive roughening of the velocity model is another possible detour for the velocity analysis, since it causes a decrease of the amplitudes of the image through the cumulative effect of reflection and transmission during wavefield extrapolation. One way to ensure a smooth solution to the estimate of the velocity model is to use a derivative regularization (Tikhonov and Arsenin, 1977).

Finally, velocity oscillations at the shallow parts of the model are very effective in narrowing down the angle coverage of the modeled wavefields without significantly changing the kinematics of the image. This will in turn cause a decrease of the misfit function that is not necessarily related to an improvement of the focusing of the CIGs. It is therefore sometimes necessary to bound the first layers of the model, which can be achieved by a bound constrained optimization algorithm (Byrd et al., 1995), or through regularization (Tikhonov and Arsenin, 1977).

These three measures (filtering of the image, derivative regularization and bounding the first layers of the model) seem to be generally enough to stabilize the optimization.

Formally, these measures are implemented by adding the following term to the objective function:

$$\mathcal{DS}_{reg} = \frac{\alpha(\mathbf{x})}{2} \left\| \frac{\partial v}{\partial \mathbf{x}}(\mathbf{x}) \right\|^2 + \frac{\beta(\mathbf{x})}{2} \left\| v(\mathbf{x}) - v_{prior}(\mathbf{x}) \right\|^2 , \tag{3}$$

where α and β are weight vectors, and v_{prior} is the vector containing a priori known values of velocity.

The final objective function is then given by:

$$\mathcal{J} = \mathcal{DS} + \mathcal{DS}_{reg}. \tag{4}$$

An optimum velocity field can then be obtained by minimizing equation 4 with respect to the P-wave velocities $v(\mathbf{x})$.

The optimization of equation 4 is performed using a gradient-based non-linear method (Byrd et al., 1995; Nocedal and Wright, 2000). The gradient of the misfit-function with respect to velocity is then required.

The gradient of equation 4 is given by:

$$\nabla_{\mathbf{v}} \mathcal{J}(\mathbf{x}) = \nabla_{\mathbf{v}} \mathcal{DS}(\mathbf{x}) + \nabla_{\mathbf{v}} \mathcal{DS}_{reg}(\mathbf{x}). \tag{5}$$

The adjoint state method (Plessix, 2006; Chavent, 2009) offers an exact and efficient way to compute the gradient of the Differential Semblance misfit function. In Appendix A, I show how to derive the gradient of equation 2 with respect to the acoustic velocity by the adjoint state method . In this case the gradient is given by:

$$\nabla_{\mathbf{v}} \mathcal{DS}(\mathbf{x}) = -\sum_s \int_0^T dt \frac{2}{v^3(\mathbf{x})} \frac{\partial^2 D}{\partial t^2}(\mathbf{x}, t, s) D'(\mathbf{x}, t, s)$$

$$-\sum_s \int_0^T dt \frac{2}{v^3(\mathbf{x})} \frac{\partial^2 U}{\partial t^2}(\mathbf{x}, t, s) U'(\mathbf{x}, t, s) \tag{6}$$

The wavefields D' and U' are adjoint states associated with the constraints that the direct states (U and D), satisfy the constant density acoustic wave equation. These wavefields can be computed by the following adjoint modelings:

$$D'(\mathbf{x}, t, s) = \int d\mathbf{x}' \ G(\mathbf{x}, 0; \mathbf{x}', t) \int_0^T dt \ \int d\mathbf{h} \ \mathbf{h}^2 \frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x}' + \mathbf{h}, \mathbf{h}) U(\mathbf{x}' + 2\mathbf{h}, t, s) \tag{7}$$

$$U'(\mathbf{x}, t, s) = \int d\mathbf{x}' \ G(\mathbf{x}, t; \mathbf{x}', 0) \int_0^T dt \ \int d\mathbf{h} \ \mathbf{h}^2 \frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x}' - \mathbf{h}, \mathbf{h}) D(\mathbf{x}' - 2\mathbf{h}, t, s) \tag{8}$$

The cost of computing the gradient this way is then approximately the same as that of evaluating the misfit function, however, it requires the state variables (U and D) to be stored for each shot, which can be expensive. In the discussion we suggest some measures to reduce this cost.

From the gradient equations we can see that, similar to RTM, the gradient computation consists of a crosscorrelation of wavefields modeled by solving the two-way wave equation. The difference is that in RTM wavefields being correlated correspond to waves propagating in opposite directions, which naturally results in an image containing predominantly high wavenumbers, congruous with a reflectivity profile. While in the gradient computation, the wavefields are propagating in the same direction, giving rise to a "smoother" image. This characteristic, shared by all methods whose misfit functions are defined on the image space, ensures that the optimization starts by updating the low wavenumber components of the velocity model.

To complete the gradient of $\mathcal{J}$ we need to compute the gradient with respect to the regularization term:

$$\nabla_{\mathbf{v}} \mathcal{DS}_{reg}(\mathbf{x}) = \beta(\mathbf{x})(v(\mathbf{x}) - v_{prior}(\mathbf{x})) - \alpha(\mathbf{x}) \frac{\partial}{\partial \mathbf{x}} \left[\frac{\partial v}{\partial \mathbf{x}}(\mathbf{x}) \right], \qquad (9)$$

where $\frac{\partial v}{\partial \mathbf{x}}$ is taken to be zero at the boundaries.

NUMERICAL EXAMPLES

We present the results of optimizing two 2D datasets. Both datasets are generated using 2D finite difference modeling, with synthetic density and acoustic velocity models. The geometry simulates a typical marine acquisition. The sources simulate monopole point sources. The source time function is a Ricker wavelet with center frequencies of 25 Hz, and 15Hz for the first and second examples, respectively. The modeling is carried out with an algorithm that is 8th order accurate in space and 2nd order accurate in time (Virieux, 1986). The aperture is taken as twice the cable length. To avoid reflections from the boundaries of the models, PML absorbing boundary conditions were implemented at all sides (Qin et al., 2009). This way, the resulting datasets are devoid of free surface multiples. Interbeded multiples are, however, still present in the data. Preprocessing of the datasets consisted in muting the direct wave and the refracted waves at the receiver level.

Shallow gas accumulation and leakage model

The first example consists of a 4 kilometer long and 1km deep model shown in figures 1 and 2. This model simulates a dipping layered sediment succession with a small "reef" like structure in the center. The densities are varying between 1800 and 2400 kg/m^3. The velocity model has the same structure as the density model, but with a localized low velocity lens (Gaussian with a peak of -500 m/s) under the reef structure.

This model simulate a scenario of shallow gas accumulation and leakage, a typical situation that if not accurately predicted by the velocity model can produce significant distortion on depth images.

The acquisition geometry simulates a 2D marine acquisition with a 311 sources separated by 20 m. The cable length is 3.2 kilometers with 10 meters channel interval. Recording time length is 2 seconds.
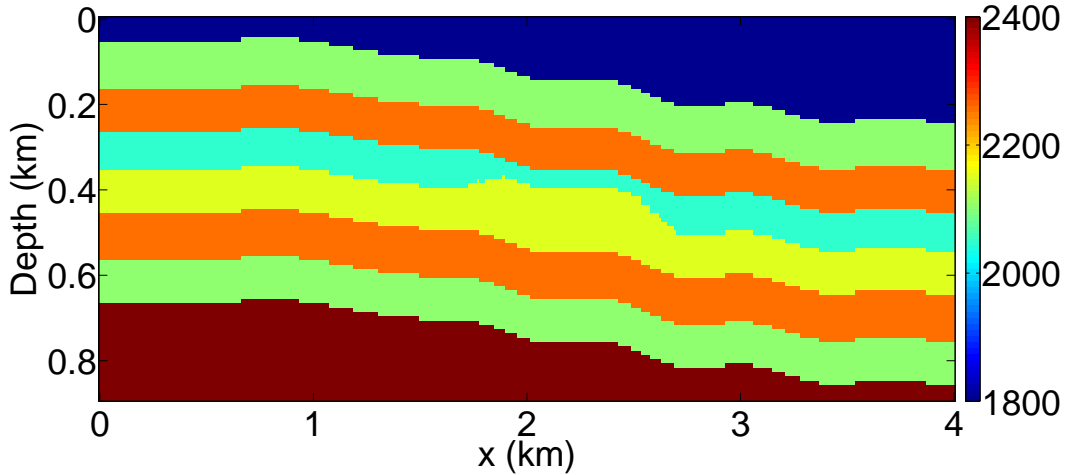

Figure 1: Density model (kg/m^3).

The initial model used for migration consists of a linear 1D velocity profile varying from 1.8 km/s to 2.4 km/s as shown in figure 3.

The CIGs are produced according to equation 1 with the half-offset axis varying between ±400 meters (81 offset samples).

The quality of our initial estimate of the velocity can be quantified by looking at the initial zero offset image shown in figure 4) and the CIGs output by RTM displayed in figure 5. The zero offset (ZO) image shows significant distortion, especially below the low velocity lens. At the same time, the CIGs are significantly spread across the offset axis.

For this example, we use regularization to fix the velocity of the first layer. Optimization is then carried out with a L-BFGS method (Byrd et al., 1995). Convergence is acchieved after 35 iterations, based on insufficient decrease criteria.

The optimized velocity model is shown in figure 6, and it is clear that the low velocity lens and the layered structure of the velocity model are successfully detected.

Figure 7 shows the results of the optimization on the zero-offset image, while figure 8 show the optimized CIGs. The ZO image migrated with the updated velocity is much better focused and the reflectors are now well positioned. At the same time, the CIGs are well focused at zero offset, indicating that the velocity is kinematically similar to the true velocity model.

Figure 9 show a comparison of the initial, updated and true traces of velocity at different spatial positions. While figure 10 does the same for the zero-offset reflectivity. The
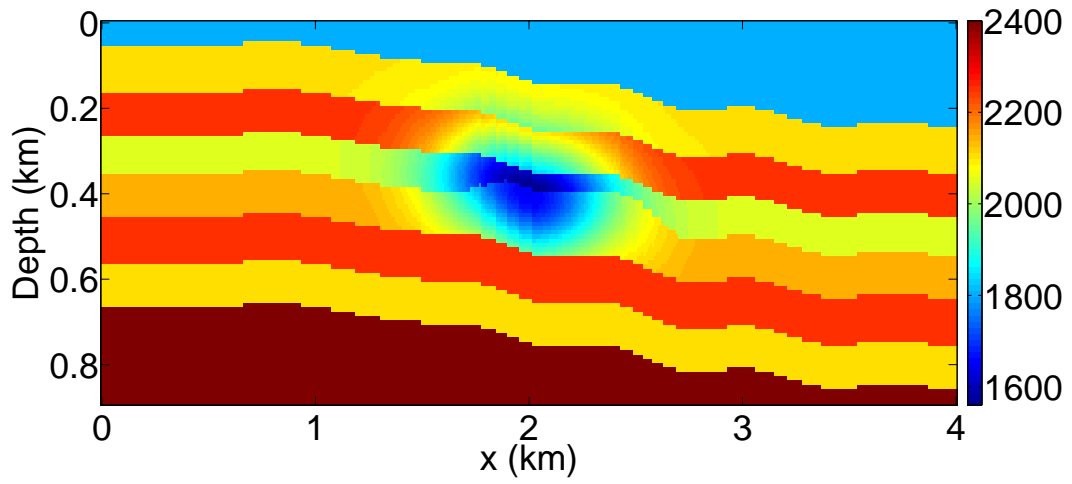
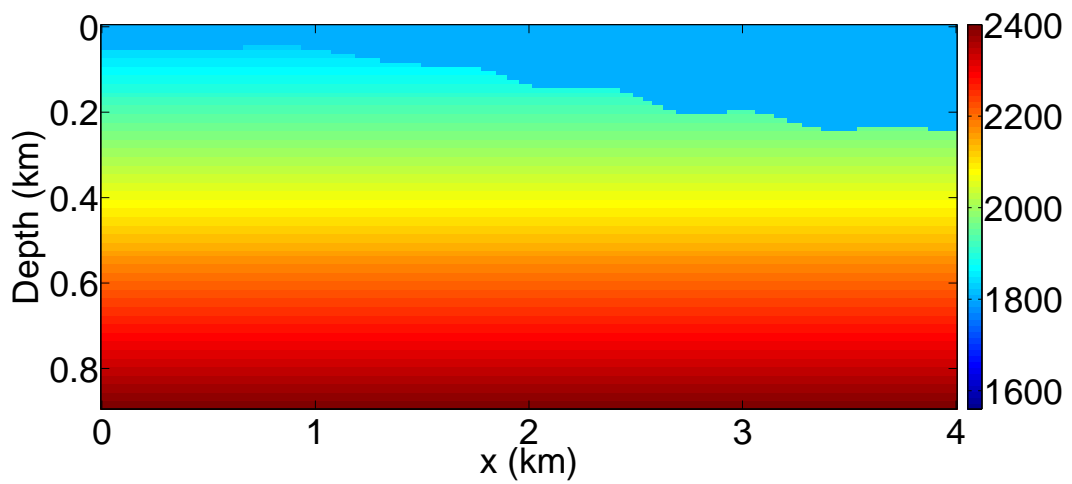Figure 2: Acoustic velocity model (m/s).

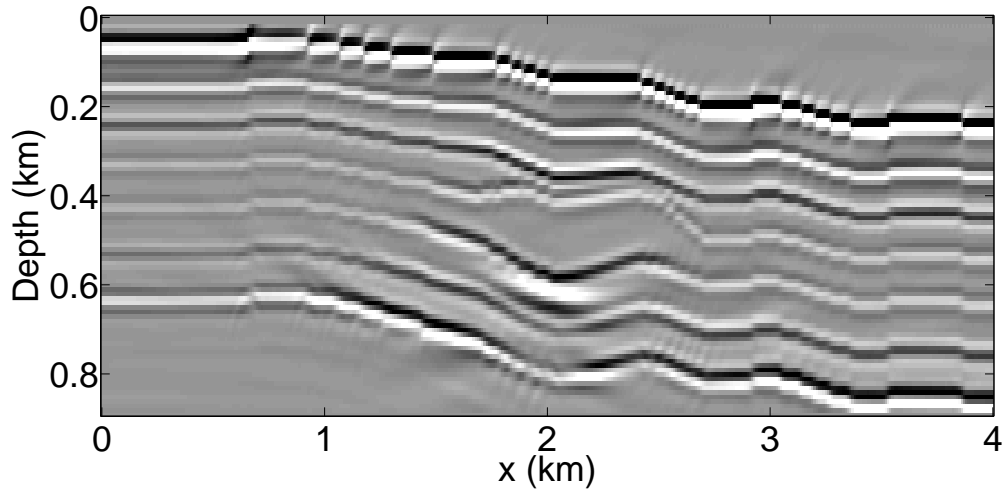Figure 3: Initial velocity model (m/s).

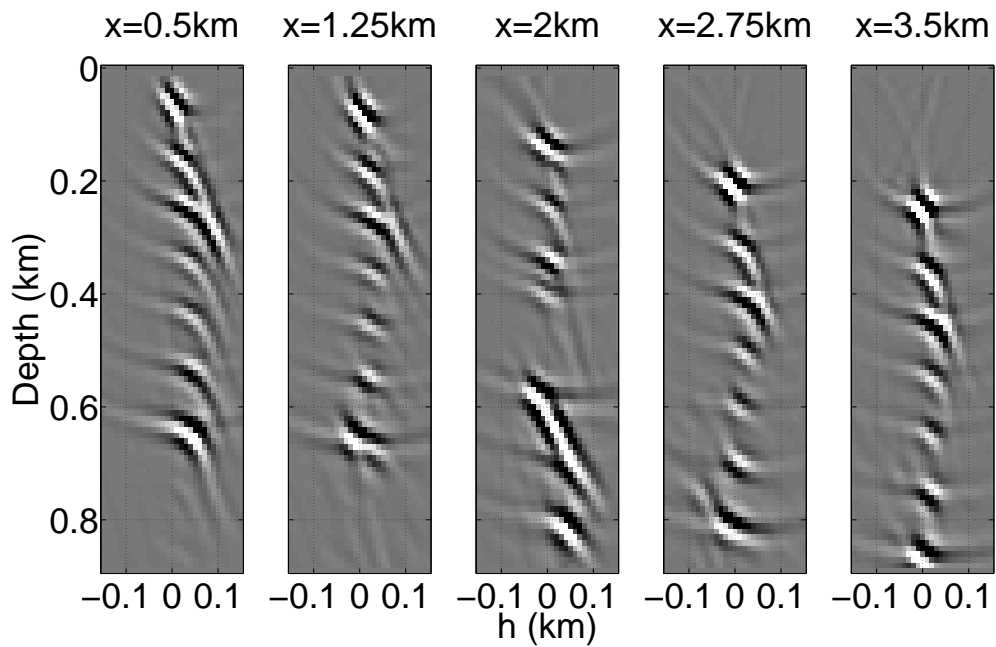Figure 4: Initial zero offset image.

Figure 5: Initial offset CIGs.

velocity traces show that, locally, the updated velocity can deviate significantly from the true velocity, though without compromising the reflectivity fit. This is a nice attestation to the non-uniqueness inherent to the solution of this type of problem. From the reflectivity, we can see that the method is able to correct traveltime errors that are, at times, larger than half-wavelength (see for example fig. 10 at $x = 2km$).
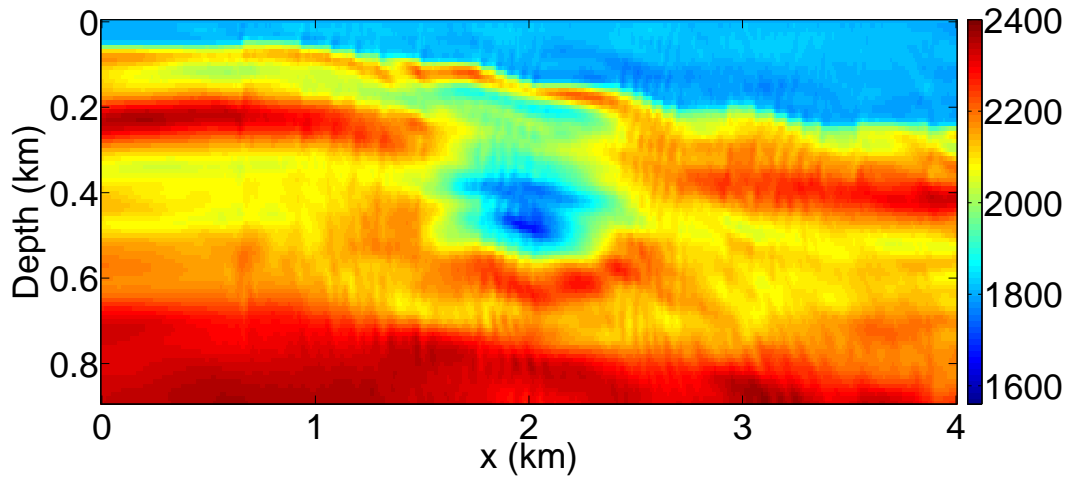
Figure 6: Updated velocity model (m/s).

To test the effect of noise in the results of the optimization, we repeat the same example, but with a noise contaminated dataset. The noise added consisted of bandlimited zero mean Gaussian noise. The resultant dataset had a signal to noise ratio of 20 (figure 11).

The updated velocities, resultant of optimization using the "noisy dataset", are shown in figure 12. In this figure, we can clearly see that the noise leaves a "stripy" inprint in the final velocity model. However, this effect does not affect the optimized reflectivities. As can be seen in figures 13 and 14. These results show that the presence of incoherent noise is not a major problem for the application of the method.

Gullfaks model

The second dataset is generated from a 2D synthetic model of the Gullfaks oil field, located in the Norwegian margin of the North Sea. The model is 3 km deep and 4 km across. At the top of the model there is 200 meter deep water layer. Beneath the water layer and down to about 1.8 km, the model consists largely of a layered sediment succession. The layered overburden unconformably overlies a set of rotated fault blocks. The densities vary between 1200 and 2400 kg/m^3 (fig. 15), while the velocities are varying between 1.48 and 3.4 km/s (fig. 16).

The geometry of the data in this example is a half-spread marine setup with 6 km long streamer equipped with receivers every 10 meters. The recording time length is 4 seconds.
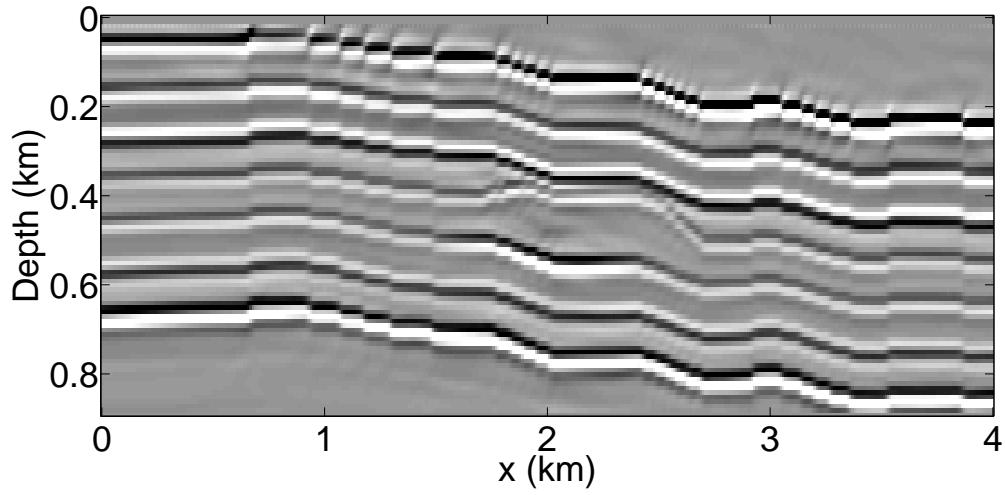
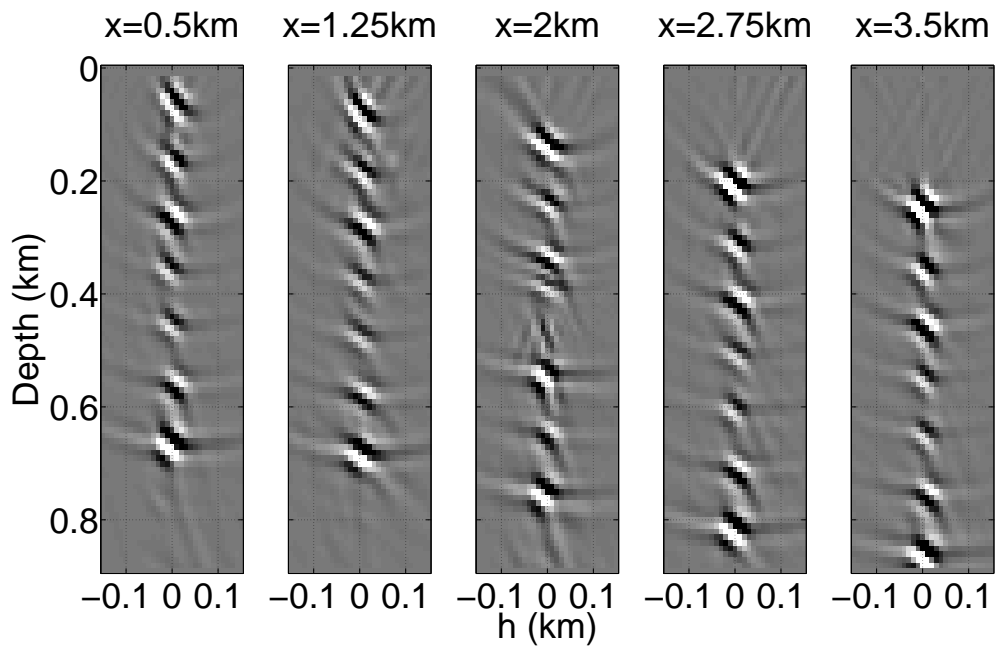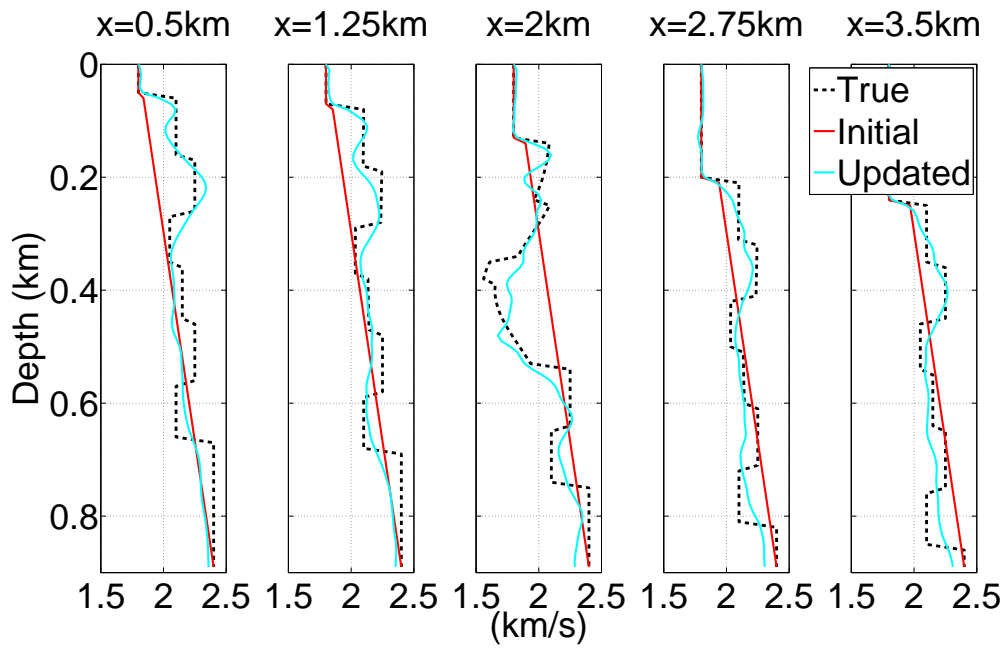Figure 7: Updated zero offset image.

Figure 8: Updated offset CIGs.
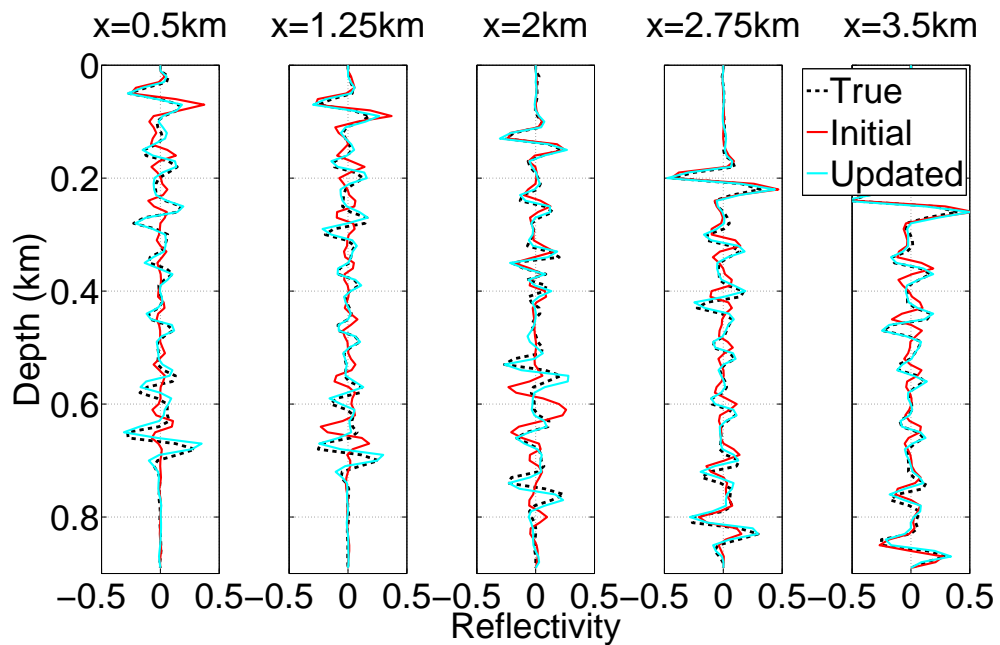
Figure 9: Comparison of velocity traces.

Figure 10: Comparison of reflectivity traces.

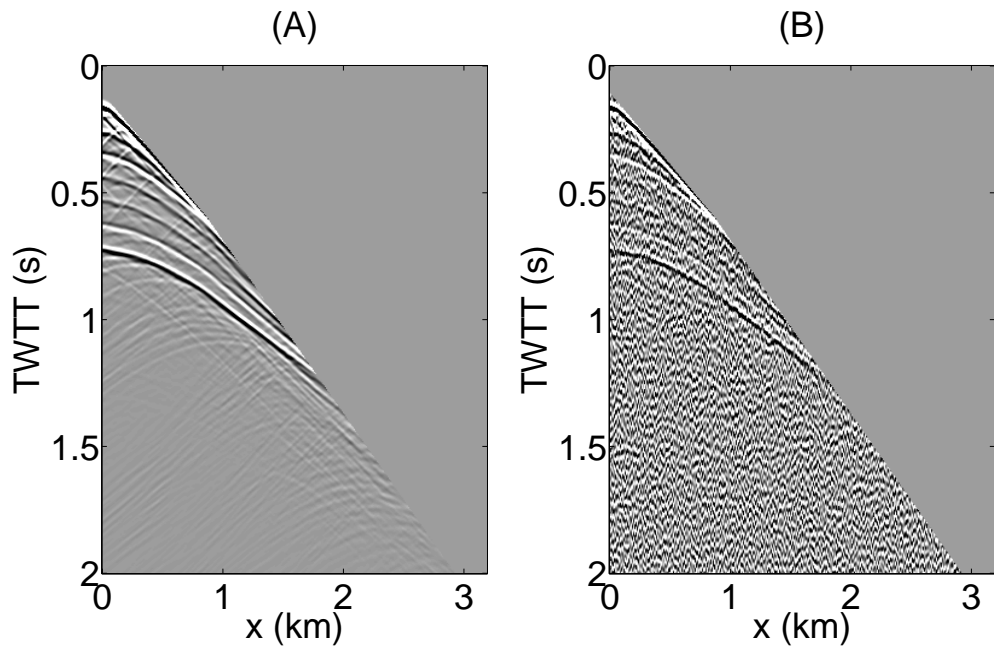Figure 11: A- Shot gather from source position x=1.5 km; B- Same shot gather contaminated with zero mean Gaussian noise ($S/N = 20$).
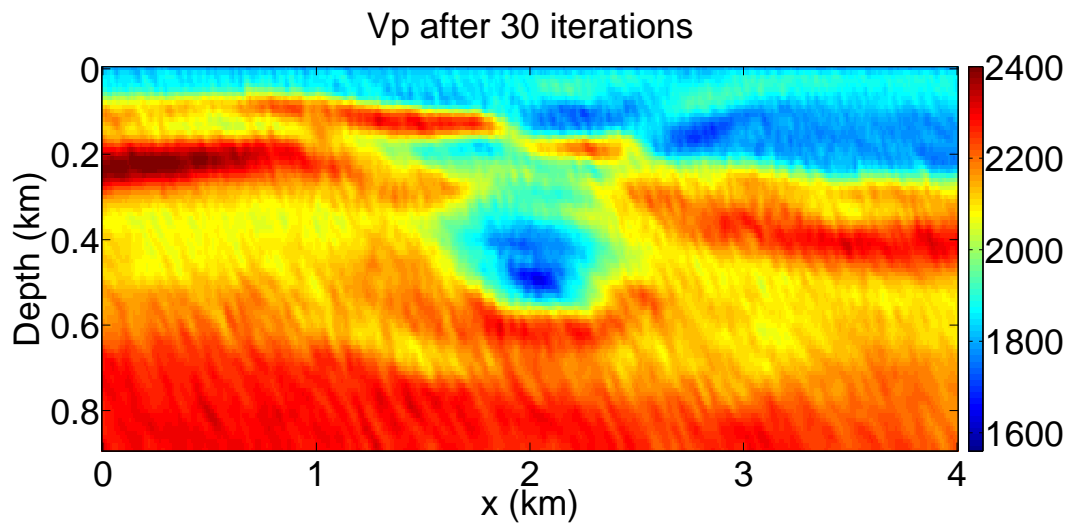

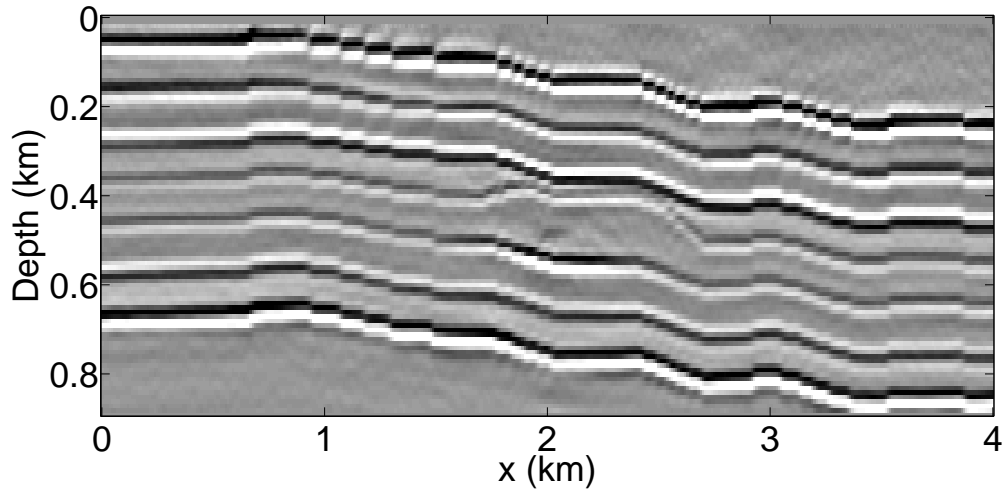Figure 12: Updated velocity model from noise contaminated data (m/s).

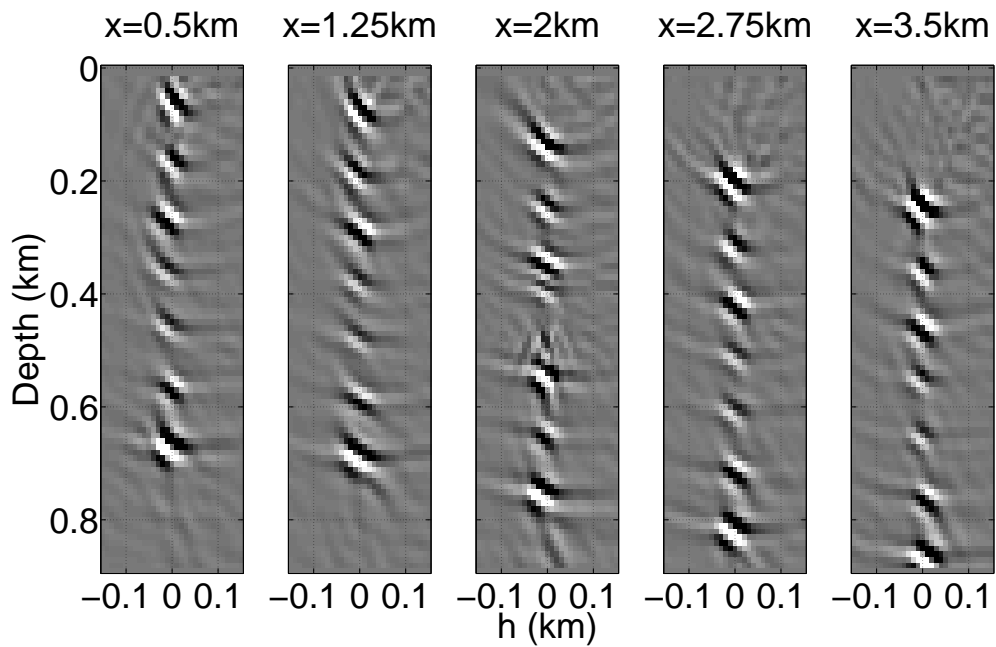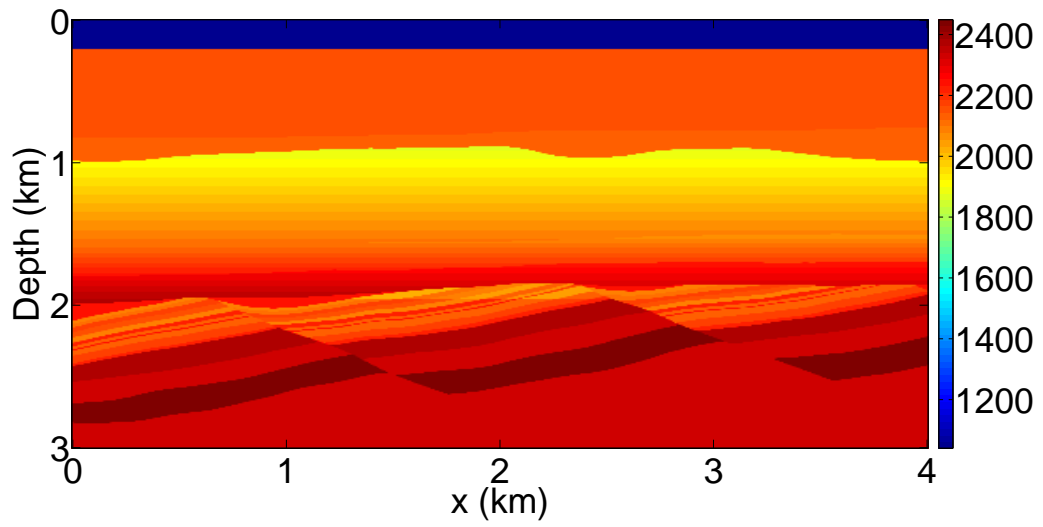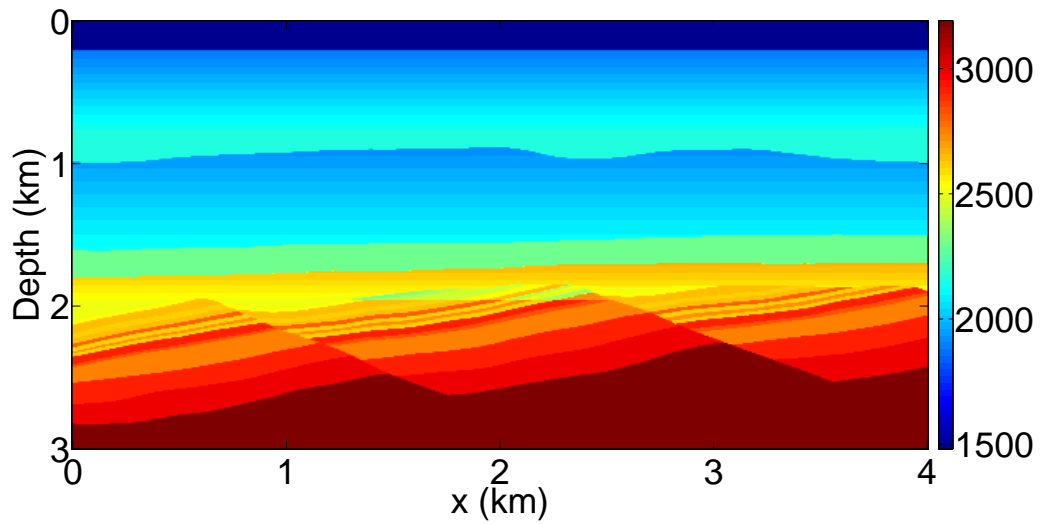Figure 13: Updated zero offset image from noise contaminated data.

Figure 14: Updated offset CIGs from noise contaminated data.

Figure 15: Density model (kg/m^3).

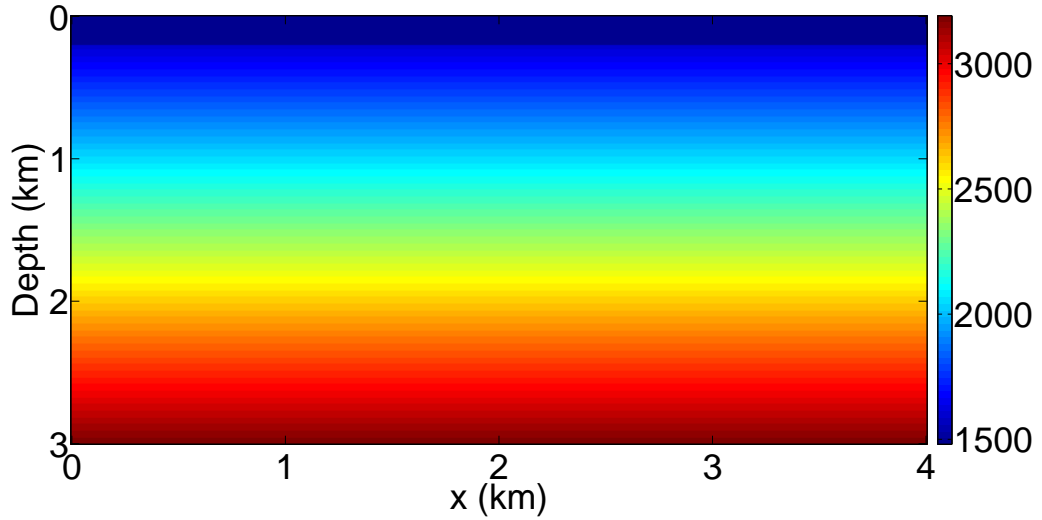Figure 16: Acoustic velocity model (m/s).

Figure 17: Initial velocity model (m/s).

The initial model used for migration consists of a 1D velocity profile linearly varying from 1.48 to 3.4 km/s (fig. 17). This model carries large deviations from the true model, and these stretch over several hundred meters. This results in significant traveltime errors (more than one wavelength at the considered frequencies).

Migration and optimization are carried out in a similar way as in the first example. Also here we assume knowledge of the first layer (water layer). However we use a bound constrain to fix its velocity at exactly 1.48 km/s. This is because, in this case, the first migrated reflector has strong artifacts in the CIGs due to shot aliasing, and using a soft regularization was significantly slowing down the convergence. In addition to that, we use regularization to avoid roughening of the velocity model.

The results of migration with the initial model are shown in figures 18 and 19.

Despite the Gullfaks synthetic model being a simple model, consisting mostly of flat layers with only mild variations in velocity, the convergence rate for this example is relativelly slow. We interpret this as a consequence of the fact that there are not many reflectors in the upper 1.8 km of the model, which makes the inversion poorly constrained. Nevertheless, the optimization is stopped after 50 iterations, at which point the CIGs were deemed sufficiently focused.

The optimized velocity model is shown in figure 20. The results of optimization show that again, in this case the method is capable of improving the kinematics of the velocity model. This can be seen clearly if we compare the initial images (figs. 18 and 19) to the updated images (figs. 21 and 22). While figures 23 and 24 show that the updated velocities and reflectivities are close to their true value (from the synthetic model).
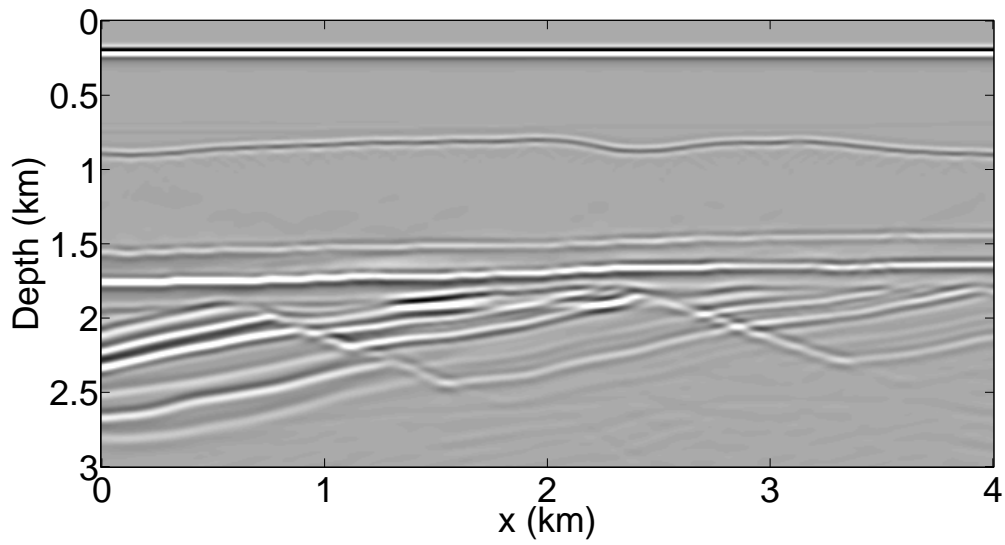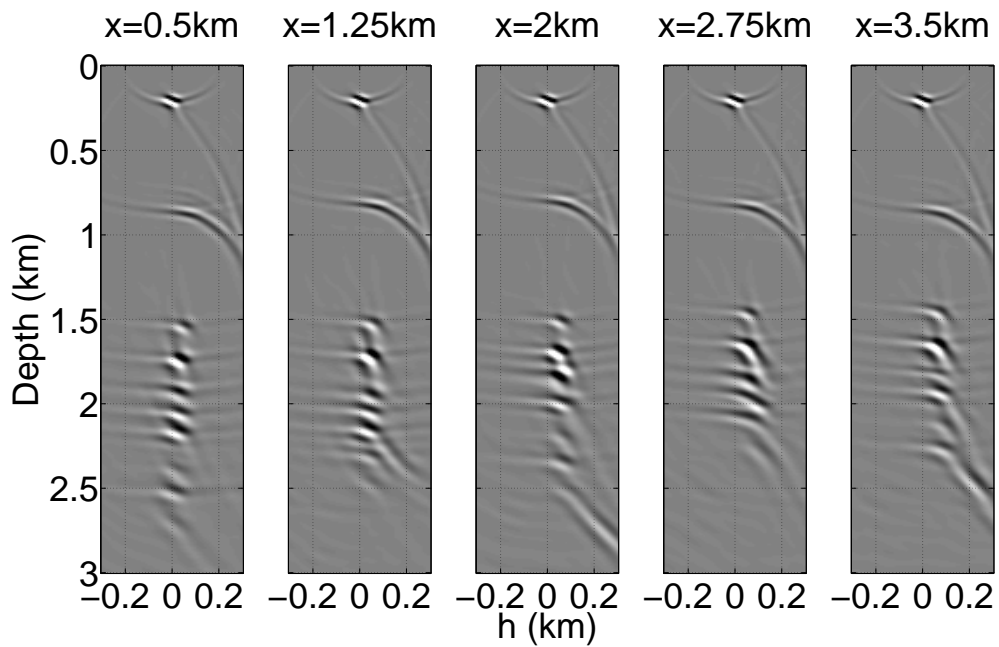
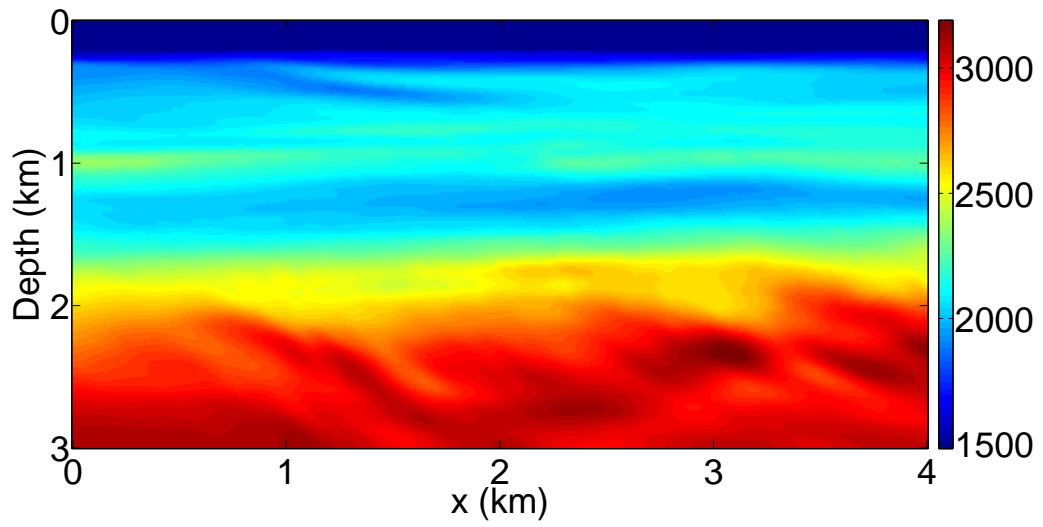Figure 18: Initial zero offset image.

Figure 19: Initial offset CIGs.

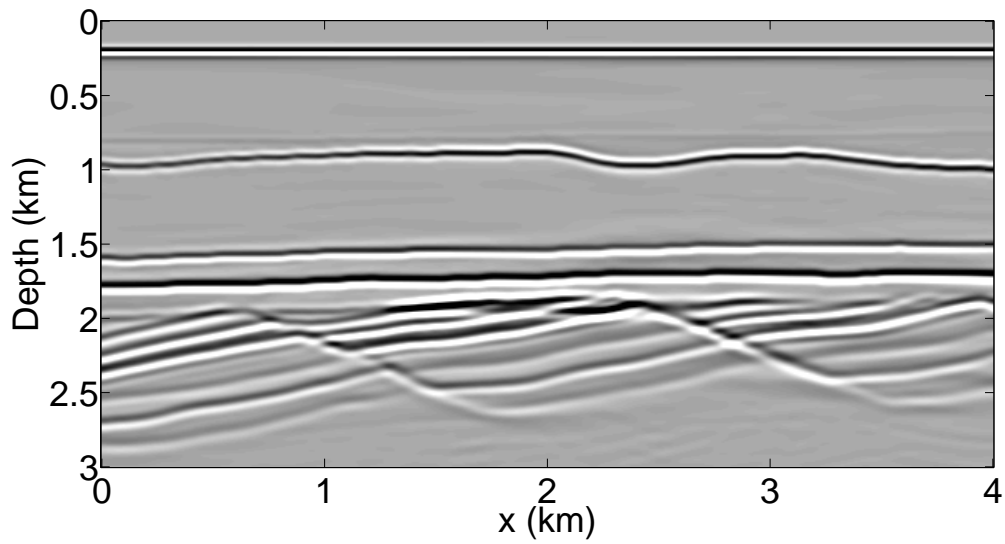Figure 20: Updated velocity model (m/s).
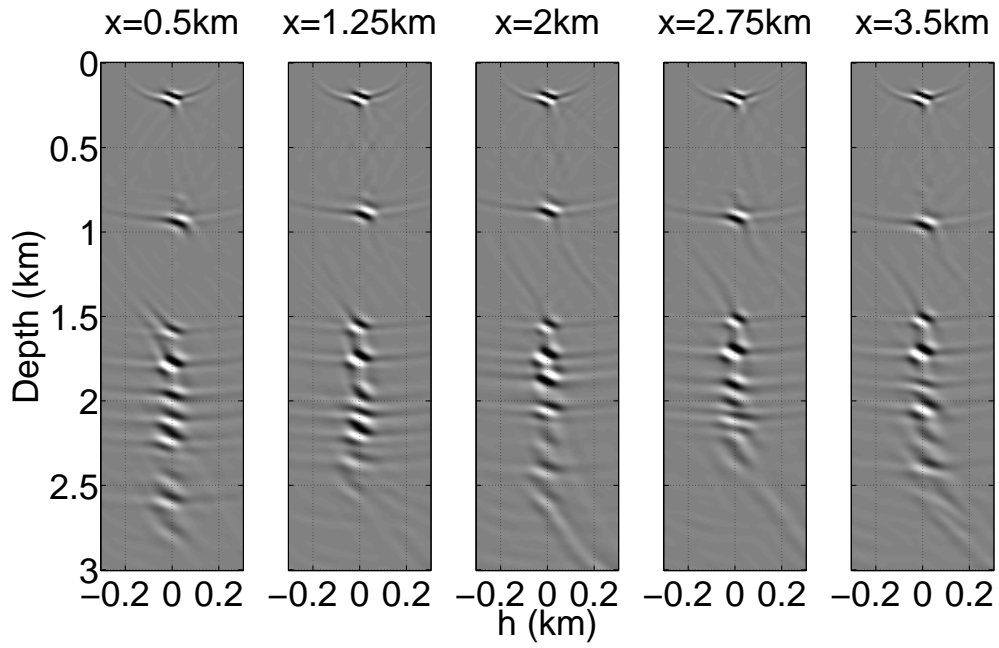

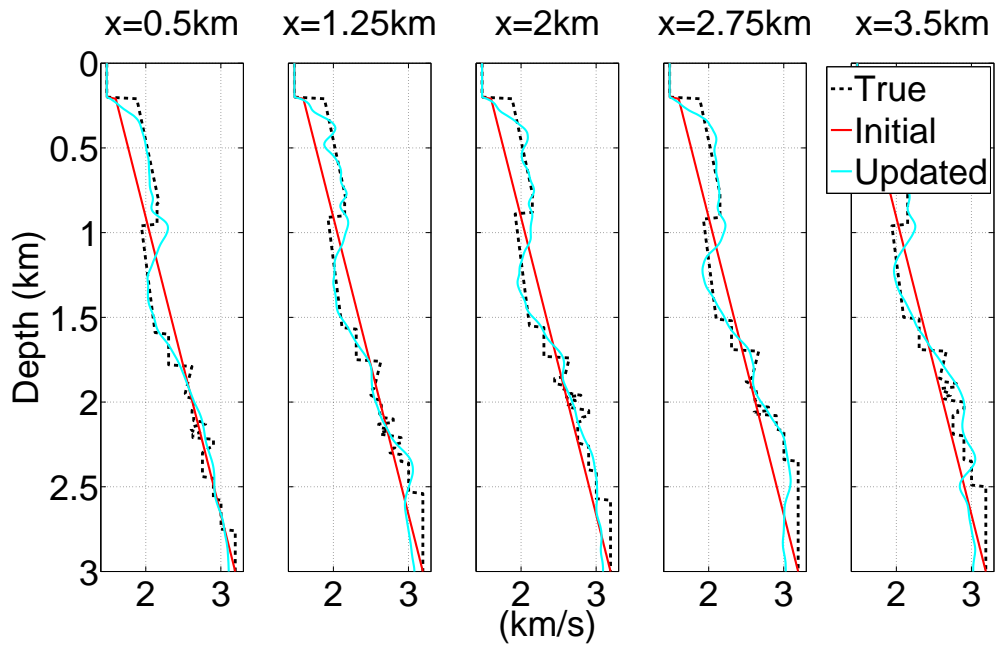Figure 21: Updated zero offset image.

Figure 22: Updated offset CIGs.

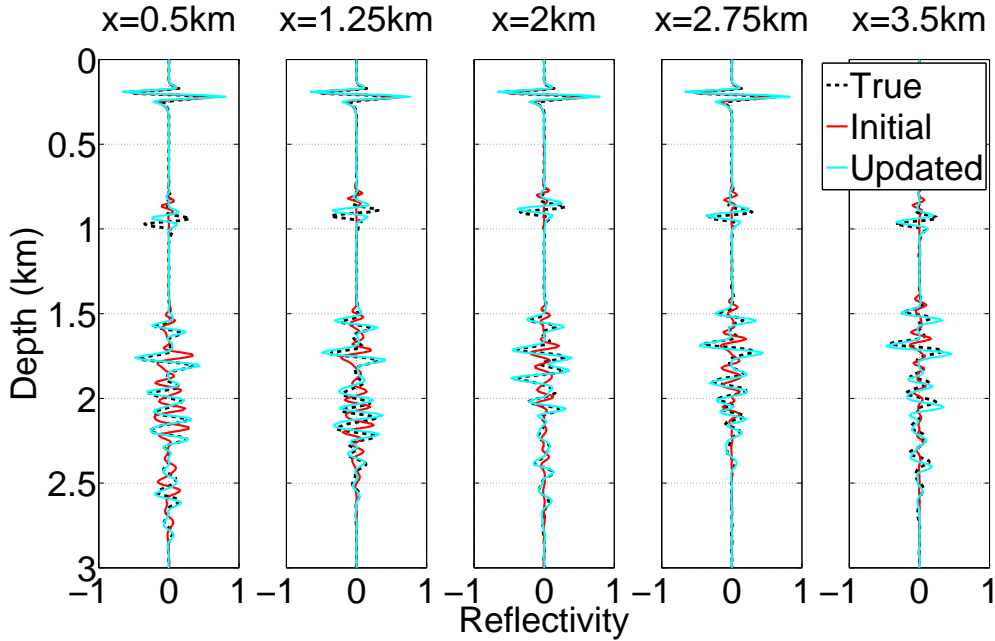Figure 23: Comparison of velocity traces.

Figure 24: Comparison of reflectivity traces.

DISCUSSION

The results presented here show that RTM based DSO works well for synthetic data. In all examples the method was able to obtain a velocity model that produces an accurate image in a few tens of iterations. The presence of coherent noise in form of interbeded multiples in the datasets used for the presented examples does not seem to have prevented the method from converging to a reasonable solution. Equally, the presence of incoherent noise in the form of zero mean Gaussian noise have little effect on the result of optimization. A question remains as to how would the method perform in the presence of other more realistic types of noise. This is of crucial importance if this method is to be used for velocity analysis in real field examples. Real field data results of DSO from other authors using different methods of migration suggest that incoherent noise does not seem to have a strong effect on the solution(Shen and Symes, 2008; Mulder and ten Kroode, 2002). Multiples, on the other hand, break with the assumptions inherent to the method and can be very detrimental (Mulder and ten Kroode, 2002).

One of the major difficulties in applying this method is the high cost, both in terms of computation and storage. Even in 2D, the cost can be large if the dataset consists of many shots and/or has a large recording array/time length. There are, however, several measures that can be used to reduce these costs. First in terms of computational cost, we suggest two strategies to improve the runtime of the method on large datasets. The first strategy is related to reducing the number of shots, and the second is related to reducing the cost of the finite difference modeling.

The cost of RTM can be reduced significantly if shots can be combined and migrated together. If only two shots are combined the cost of migration is already halved. The speedup does not come for free though, since this approach introduces crosstalk in the result

image. The crosstalk can be attenuated by using some sort of source encoding (Romero et al., 2000) . This approach has been widely experimented in RTM and FWI (Ben-Hadj-Ali et al., 2011). It is, however, not clear how the crosstalk artifacts could affect the results of the velocity analysis with RTM based DSO.

Parallelization of the finite difference modeling can be combined with the previous approach to further reduce the runtime of RTM based DSO. This can be easily achieved through domain decomposition (Smith et al., 1996).

In RTM a typical problem has been the need to model separetely the incident wave field and the scattered wavefield, which means that one of the two wavefields must be stored and reacessed later for the imaging step (checkpointing). In the case of differential semblance optimization, both fields need to stored and accessed during gradient computation. Therefore the cost of storage is the double of that of RTM. In this case the cost of storing these fields can be reduced by optimal checkpointing as suggested in Symes (2007).

CONCLUSION

Migration velocity fields can be estimated by minimizing an object function based on differential-semblance and a spatial differentiation operator. The chosen object function reduces the sensitivity to artifacts due to the two-way wave propagation used in the non-linear estimation scheme. It is shown that the new velocity estimation method produces kinematically correct models for both simple and more complex synthetic cases.

APPENDIX A

DERIVATION OF GRADIENT OF THE DIFFERENTIAL SEMBLANCE OBJECTIVE FUNCTION

We now present the main steps required to derive the gradient of equation 2 by the adjoint state method (Plessix, 2006; Chavent, 2009). To simplify the notation at most I introduce the subsurface source and receiver coordinates $\mathbf{x_s} = (x_1 + h_1, x_2 + h_2, x_3)$ and $\mathbf{x_r} = (x_1 - h_1, x_2 - h_2, x_3)$. Subscripts often refer to the Einstein summation convention with the exception to when I need to refer to the subsurface coordinates, then the indexes of Einstein summation convention are cast as superscripts, such as in $\frac{\partial u_i}{\partial x_s{}^i}$.

A lagrangian function associated with the problem of minimizing $\mathcal{DS}$ with respect to $\mathbf{v}$ can be written as:

$$
\begin{aligned}
\mathcal{L}(U, D, U', D', \mathbf{c}) = & \frac{1}{2} \int d\mathbf{x_r} \int d\mathbf{x_s} \; h^2(\mathbf{x_r}, \mathbf{x_s}) \frac{\partial \mathcal{R}}{\partial x_3}^2 (\mathbf{x_r}, \mathbf{x_s}) \\
& + \sum_s \int d\mathbf{x_r} \int_0^T dt \; \left[\left(\frac{1}{v^2(\mathbf{x_r})} \frac{\partial^2}{\partial t^2} + \nabla^2 \right) U(\mathbf{x_r}, -t, s) \right. \\
& \left. - \sum_{rec=1}^{Nrec} \delta(\mathbf{x_r} - \mathbf{x}_{rec}) P^{rec}(\mathbf{x}_{rec}, -t, s) \right] U'(\mathbf{x_r}, t, s) \\
& + \sum_s \int d\mathbf{x_s} \int_0^T dt \; \left[\left(\frac{1}{v^2(\mathbf{x_s})} \frac{\partial^2}{\partial t^2} + \nabla^2 \right) D(\mathbf{x_s}, t, s) \right.
\end{aligned}
$$

$$- \sum_{sou=1}^{Nsou} \delta(\mathbf{x_s} - \mathbf{x}_{sou})S(\mathbf{x}_{sou}, t, s) \Bigg] D'(x_s, t, s) \tag{A-1}$$

Where $U'(x_r, t, s)$ and $D'(x_s, t, s)$ are lagrange multipliers (adjoint states).

But $\mathcal{L}$ is related to DS by:

$$\mathcal{DS} = \mathcal{L}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v}), \tag{A-2}$$

where $U(\mathbf{v})$, $D(\mathbf{v})$ denote one realization of the direct states for a particular shot with a velocity vector $\mathbf{v}$.

Implicit differentiation of the above equation with respect to $\mathbf{v}$ gives:

$$\delta\mathcal{DS} = \frac{\partial\mathcal{L}}{\partial U}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta U + \frac{\partial\mathcal{L}}{\partial D}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta D$$
$$+ \frac{\partial\mathcal{L}}{\partial \mathbf{v}}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta\mathbf{v} \tag{A-3}$$

Now if we choose $\mathbf{U}'$ and $\mathbf{D}'$ as to satisfy:

$$\frac{\partial\mathcal{L}}{\partial U}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta U = 0 \tag{A-4}$$

$$\frac{\partial\mathcal{L}}{\partial D}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta D = 0 \tag{A-5}$$

for all $\delta\mathbf{U}$ and $\delta\mathbf{D}$, then equation A-3 reduces to:

$$\delta\mathcal{DS} = \frac{\partial\mathcal{L}}{\partial \mathbf{v}}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta\mathbf{v}. \tag{A-6}$$

The problem now is to solve equations A-4 and A-5 for the adjoint states for each shot. Starting with equation A-5:

$$\frac{\partial\mathcal{L}}{\partial D} = -\int_0^T dt \int d\mathbf{x_s} \int d\mathbf{x_r}\ h^2(\mathbf{x_r}, \mathbf{x_s})\frac{\partial^2\mathcal{R}}{\partial x_3^2}(\mathbf{x_r}, \mathbf{x_s})U(\mathbf{x_r}, t, s)\delta D(\mathbf{x_s}, t, s)$$
$$+ \int_0^T dt \int d\mathbf{x_s}\ \left[\left(\frac{1}{v^2(\mathbf{x_s})}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\delta D(\mathbf{x_s}, t, s)\right] D'(s) = 0 \tag{A-7}$$

where we applied integration by parts to move the derivative with respect to x_3 from δD to $\mathcal{R}$.

Applying first order perturbation theory to the constant density acoustic wave equation yields an equation for δD:

$$\left(\frac{1}{v^2(\mathbf{x})}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\delta D(\mathbf{x}, t) = \frac{2\delta v(\mathbf{x})}{v^3(\mathbf{x})}\frac{\partial^2 D}{\partial t^2}(\mathbf{x}, t) \tag{A-8}$$

The solution to this equation can be written in terms of Green's functions as:

$$\delta D(\mathbf{x}, t, s) = \int d\mathbf{x}'\ G(\mathbf{x}, t; \mathbf{x}', 0) * \left(\frac{2\delta v(\mathbf{x}')}{v^3(\mathbf{x}')}\frac{\partial^2 D}{\partial t^2}(\mathbf{x}', t)\right) \tag{A-9}$$

Where * means time convolution.

Substitution of δD and $\left[\left(\frac{1}{v^2(\mathbf{x_s})}\frac{\partial^2}{\partial t^2} + \nabla^2\right)\delta D\right]$ in equation A-7 by the RHS terms of equations A-8 and A-9 yields:

$$-\int_0^T dt \int d\mathbf{x_s} \int d\mathbf{x_r}\ h^2(\mathbf{x_r}, \mathbf{x_s})\frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x_r}, \mathbf{x_s})U(\mathbf{x_r}, t, s)$$

$$\times \int d\mathbf{x'_s}\ G(\mathbf{x_s}, t; \mathbf{x'_s}, 0) * \left(\frac{2\delta v(\mathbf{x'_s})}{v^3(\mathbf{x'_s})}\frac{\partial^2 D}{\partial t^2}(\mathbf{x'_s}, t, s)\right)$$

$$+\int_0^T dt \int d\mathbf{x'_s}\ \frac{2\delta v(\mathbf{x'_s})}{v^3(\mathbf{x'_s})}\frac{\partial^2 D}{\partial t^2}(\mathbf{x'_s}, t)D'(s) = 0 \tag{A-10}$$

Using the reciprocity of the Green's functions $[G(x'_r, t; x_r, 0) = G(x_r, t; x'_r, 0)]$ and the identity:

$$\int_0^T dt\ h(t)[G(t) * f(t)] = \int_0^T dt\ f(t)[G(-t) * h(t)] \tag{A-11}$$

Equation A-10 can now be written:

$$-\int_0^T dt \int d\mathbf{x'_s}\ \left(\frac{2\delta v(\mathbf{x'_s})}{v^3(\mathbf{x'_s})}\frac{\partial^2 D}{\partial t^2}(\mathbf{x'_s}, t, s)\right)$$

$$\times \int d\mathbf{x_s}\ G(\mathbf{x'_s}, 0; \mathbf{x_s}, t) * \left(\int d\mathbf{x_r}\ h^2(\mathbf{x_r}, \mathbf{x_s})\frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x_r}, \mathbf{x_s})U(\mathbf{x_r}, t, s)\right)$$

$$+\int_0^T dt \int d\mathbf{x'_s}\ \frac{2\delta v(\mathbf{x'_s})}{v^3(\mathbf{x'_s})}\frac{\partial^2 D}{\partial t^2}(\mathbf{x'_s}, t)D'(s) = 0 \tag{A-12}$$

The solution of equation A-12 then gives:

$$D'(\mathbf{x}', t, s) = \int d\mathbf{x_s}\ G(\mathbf{x}', 0; \mathbf{x_s}, t) * \left(\int d\mathbf{x_r}\ h^2(\mathbf{x_r}, \mathbf{x_s})\frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x_r}, \mathbf{x_s})U(\mathbf{x_r}, t, s)\right) \tag{A-13}$$

By similar procedure we can solve equation A-4 to get:

$$U'(\mathbf{x}', t, s) = \int d\mathbf{x_r}\ G(\mathbf{x}', t; \mathbf{x_r}, 0) * \left(\int d\mathbf{x_s}\ h^2(\mathbf{x_r}, \mathbf{x_s})\frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x_r}, \mathbf{x_s})D(\mathbf{x_s}, t, s)\right) \tag{A-14}$$

Change of coordinates from subsurface source and receiver coordinates $(\mathbf{x_s}, \mathbf{x_r})$ to subsurface midpoint and half-offset $(\mathbf{x}, \mathbf{h})$, gives:

$$D'(\mathbf{x}', t, s) = \int d\mathbf{x}\ G(\mathbf{x}', 0; \mathbf{x}, t) * \left(\int d\mathbf{h}\ h^2\frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x} + \mathbf{h}, \mathbf{h})U(\mathbf{x} + 2\mathbf{h}, t, s)\right)$$

$$U'(\mathbf{x}', t, s) = \int d\mathbf{x}\ G(\mathbf{x}', t; \mathbf{x}, 0) * \left(\int d\mathbf{h}\ h^2\frac{\partial^2 \mathcal{R}}{\partial x_3^2}(\mathbf{x} - \mathbf{h}, \mathbf{h})D(\mathbf{x} - 2\mathbf{h}, t, s)\right)$$

Finally we turn to the problem of finding the derivative of $\mathcal{DS}$ with respect to velocity $(\mathbf{v})$. After equations A-4 and A-5 the formula for $\delta\mathcal{DS}$ (equation A-3) simplifies to:

$$\delta\mathcal{DS} = \frac{\partial \mathcal{L}}{\partial \mathbf{v}}(U(\mathbf{v}), D(\mathbf{v}), U', D', \mathbf{v})\delta\mathbf{v} \tag{A-15}$$

Differentiation of equation A-1 with respect to $\mathbf{v}$ yields:

$$
\begin{aligned}
\delta\mathcal{DS} \;=\; & -\sum_s \int_0^T dt \int d\mathbf{x}\, \frac{2\delta v(\mathbf{x})}{v^3(\mathbf{x})} \frac{\partial^2 D}{\partial t^2}(\mathbf{x},t,s) D'(\mathbf{x},t,s) \\
& -\sum_s \int_0^T dt \int d\mathbf{x}\, \frac{2\delta v(\mathbf{x})}{v^3(\mathbf{x})} \frac{\partial^2 U}{\partial t^2}(\mathbf{x},t,s) U'(\mathbf{x},t,s)
\end{aligned}
$$

From which we obtain the gradient by picking coefficients for $\delta\mathbf{v}$:

$$
\begin{aligned}
\nabla_v \mathcal{DS}(\mathbf{x}) \;=\; & -\sum_s \int_0^T dt\, \frac{2}{v^3(\mathbf{x})} \frac{\partial^2 D}{\partial t^2}(\mathbf{x},t,s) D'(\mathbf{x},t,s) \\
& -\sum_s \int_0^T dt\, \frac{2}{v^3(\mathbf{x})} \frac{\partial^2 U}{\partial t^2}(\mathbf{x},t,s) U'(\mathbf{x},t,s)
\end{aligned}
$$

REFERENCES

Arntsen, B., C. Gerea, and T. Rosten, 2009, Imaging salt bodies using explicit migration operators offshore norway: Geophysics, **74**, S25–S32.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, **76**, R109–R124.

Biondi, B., and P. Sava, 1999, Wave-equation migration velocity analysis: SEG Houston 1999 International Exposition and Anual Meeting, Soc. of Expl. Geophys., 1723–1726.

Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Byrd, R. H., P. Lu, J. Nocedal, and C. Zhu, 1995, A limited memory algorithm for bound contrained optimization: SIAM Journal on Scientific Computing, **16**, 1190–1208.

Chavent, G., 2009, Non-linear least squares for inverse problems: Theoretical foundations and step-by-step guide for applications: Springer.

Chavent, G., and C. A. Jacewitz, 1995, Determination of background velocities by multiple migration fitting: Geophysics, **60**, 476–490.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

Gao, F., and W. Symes, 2009, Differential semblance velocity analysis by reverse time migration: Image gathers and theory: SEG Houston 2009 International Exposition and Anual Meeting, Society of Exploration Geophysicists, 2317–2321.

Guitton, A., B. Kaelin, and B. Biondi, 2007, Least-squares attenuation of reverse-time-migration artifacts: Geophysics, **72**, S19–S23.

Mulder, W. A., 2008, Automatic velocity analysis with the two-way wave equation: Presented at the 70th EAGE Conference & Exhibition, European Association of Geoscientists and Engineers.

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

Nocedal, J., and S. J. Wright, 2000, Numerical optimization: Springer.

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**, 495–503.

Qin, Z., M. Lu, X. Zheng, Y. Yao, C. Zhang, and J. Song, 2009, The implementation of an improved NPML absorbing boundary condition in elastic wave modeling: Applied Geophysics, **6**, 113–121.

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

Romero, L. A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

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

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

Sava, P., and L. Vlad, 2008, Numeric implementation of wave-equation migration velocity analysis operators: Geophysics, **73**, 145–159.

Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, 49–59.

Shen, P., W. W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by waveequation migration: SEG, Expanded Abstracts, **22**, 2132–2135.

Smith, B., P. E. Bjorstad, and W. D. Gropp, 1996, Domain decomposition: Parallel multi-level methods for elliptic partial differential equations: Cambridge university press.

Symes, W. W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, SM213–221.

———, 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **5**, 654–663.

Tikhonov, A. N., and V. Y. Arsenin, 1977, Solution of ill-posed problems: W. H. Winston and Sons.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media; velocity-stress finite-difference method: Geophysics, **51**, 889–901.