GEOPHYSICS®

Time-dependent receiver extension for FWI: an alternative extension method for cycle skipping mitigation

Journal:	*Geophysics*
Manuscript ID	Draft
Manuscript Type:	Technical Paper
Keywords:	full-waveform inversion, algorithm, optimization, time-domain, inversion
Manuscript Focus Area:	Seismic Inversion, Seismic Velocities and Traveltimes

Time-dependent receiver extension for FWI: an alternative extension method for cycle skipping mitigation

(March 3, 2025)

Running head: **Time-dependent receiver extension**

ABSTRACT

Extension strategies for full waveform inversion (FWI) rely on introducing additional degrees of freedom to the FWI problem, which expands the search space. This search space extension helps relaxing the non-convexity of the problem, and thereby alleviating the cycle-skipping issue. Receiver-based extension strategy introduces the receiver position as the additional degree of freedom to full waveform inversion, in order to improve the fit between the observed and calculated data at early iterations. This helps circumvent the cycle-skipping phenomenon. In this study, we make this receiver position time-dependent, meaning that the receiver positions vary as a function of the acquisition time. The resulting mathematical problem is a two nested-loops minimization, where the outer loop is the conventional FWI loop to update the subsurface mechanical parameters, and the inner loop aims at finding the optimal time-dependent virtual receivers positions. This inner loop problem is heavily nonlinear and non-convex. Finding the global minimum is therefore a challenging task. We use for that a computational intelligence technique, Particle Swarm Optimization (PSO). PSO makes it possible to thoroughly explore the search space with few iterations. Numerical experiments using a North Sea exploration 2D synthetic model,

starting from crude initial models, illustrate that the method is robust and is very easy to tune.

INTRODUCTION

Full waveform inversion (FWI) is a high resolution seismic imaging technique, which aims at reconstructing the subsurface parameters using the full seismic waveforms. From a mathematical standpoint, FWI is formulated as a partial differential equation (PDE) constrained optimization problem, where the optimality criterion is the fit between the observed and the calculated datasets. This constrained optimization problem is solved iteratively, using gradient based approaches. The synthetic data which are computed in a given initial model (using a wave equation operator), are conventionally compared to the observed data in the least-squares sense. That is, the L_2 norm of the difference between the observed and synthetic datasets, which we call the misfit. The model is then updated in a manner that reduces this misfit. However, the L_2 misfit function is non-convex, meaning that it contains numerous local minima. Therefore, gradient based (local) optimization strategies might fail to converge to a meaningful solution. This occurs when the initial model is not accurate enough, causing the calculated data to be shifted in time by more than half the dominant period, with respect to the observed data (Virieux and Operto, 2009). This time shift between the two datasets is driven by the long wavelength (smooth) part of the velocity model (Jannane et al., 1989). This long wavelength part of the model controls the kinematics of the seismic data. Starting from an initial velocity model that does not contain the correct long wavelength velocity structure, which gives rise to a time shift larger than half the dominant period, causes FWI to converge to a local minimum. This is the well-known cycle-skipping issue, and it stems from the oscillatory nature of the seismic data.

Numerous strategies have been developed to circumvent this issue, such as multiscale approaches. In this framework, the inversion can be performed from lower to higher frequencies (Bunks et al., 1995; Sirgue and Pratt, 2004), as starting from lower frequencies results in broader phases, the support of which is likely to overlap, decreasing the apparent shift between the observed and calculated data. The inversion can also be done from narrower to wider offsets (Shipp and Singh, 2002; Brossier et al., 2009), as the shorter the offset, the less is the time-shift. This can be done in combination with time-windowing, considering at first the earlier events, and increasing this time window when the frequency and offsets are increased, as the model estimate improves. These strategies, rely on the availability of low frequencies and large offset in the data, which is not always possible. Furthermore, they do require heavy human intervention, making the FWI process less automatic.

More recent advances suggest different strategies to overcome the cycle-skipping issue, by reformulating the FWI problem. We divide these strategies into two categories. The first one relies on using alternative misfit functions, instead of the conventional L_2 norm. The second introduces additional degrees of freedom to the FWI problem, to relax the non-convexity. The strategies that fall in the first category use different metrics to measure the distance between the observed and calculated data. These alternative misfit functions can exhibit improved convexity, compared with the L_2 misfit function. They can be based on crosscorrelation (Luo and Schuster, 1991; van Leeuwen and Mulder, 2010), deconvolution (Luo and Sava, 2011; Guasch et al., 2019; Yong et al., 2022), instantaneous envelope (Bozdağ et al., 2011; Wu et al., 2014), dynamic time-warping (Ma and Hale, 2013) or optimal transport distances (Métivier et al., 2016, 2019; Yang et al., 2018b), to cite a few.

The second category encompasses strategies that rely on introducing additional degrees of freedom to the FWI problem. This is done to help relaxing the non-convexity by fitting the calculated data to the observed data, when the model estimate is poor. The strategy presented in this study belongs to this class of methods. The additional degrees of freedom can be introduced in the model space, giving the so-called model extension strategies. These methods rely on the assumption of scale separation between the background model and the reflectivity model, in a similar fashion as reflection waveform inversion (RWI) methods (Brossier et al., 2015; Yao et al., 2020). This is achieved during the image volume construction, introducing horizontal subsurface offsets or time lags as the additional degrees of freedom in the imaging condition. This depth oriented image construction workflow represents the so-called Migration Velocity Analysis (MVA) algorithms (Biondi and Sava, 1999; Symes, 2008; Mulder, 2014), where it is potentially possible to fit the observed data in a wrong background model. Similar to MVA, inversion velocity analysis introduces subsurface offsets or time lags as the additional degrees of freedom, using a nested optimization approach (Biondi and Almomin, 2014; Chauris and Cocher, 2017; Barnier et al., 2023a,b). An inner loop updates the reflectivity in a given background model in a migration process, while the outer loop updates the background model according to some focusing criterion. The additional degrees of freedom are relaxed during the process of FWI, thanks to the use of a penalty term in the misfit function (an annihilator). The use of an annihilator is common to all extension methods. When the extension is carried out in the model space, it gives rise to high dimensional problems. For instance, model extension with sub-surface offsets (the extended parameter) leads to 3D problem in the 2D case (x, y, h), h being the horizontal subsurface offset. In 3D it leads to a 5D problem (x, y, z, h_x, h_y) (Chauris and Cocher, 2017). The computation and storage related to these hypercubes can be prohibitive for large scale 3D problems.

van Leeuwen and Herrmann (2013, 2016) introduced wavefield reconstruction inversion (WRI), where the search space is extended by optimizing over both the wavefield and the model parameters. This is achieved by considering the wave equation as a soft constraint, using a penalty method. The minimization of the misfit associated with this problem gives rise to a linear system, which van Leeuwen and Herrmann (2013) call the augmented wave equation. This augmented wave equation gathers the data extraction from the wavefield constraint, and the wave equation. This is done in the frequency domain, because one linear linear system replaces the classical wave equation, thanks to the possibility of the wave equation operator factorization in the frequency domain. It is difficult to use this formulation with explicit time-marching in the time-domain, where such factorization is not possible. WRI is controlled by a penalty parameter, a scalar weight given to the penalty term of the wave equation. The data is matched well for small values of the penalty parameter, even with poor velocity models (less importance is given to the wave equation). We note that the penalty parameter needs to be increased during FWI iterations, which makes the tuning difficult. This can be circumvented using an augmented Lagrangian formulation instead of the penalty method (Aghamiry et al., 2018, 2019b,a), which they call Iteratively Refined WRI (IR-WRI).

Source extension methods are obtained by reparametrizing the WRI by means of a change of variables (Wang et al., 2017; Huang et al., 2018a). Through the change of variables, the reconstructed wavefield is replaced with the extended source (also referred to as the wave equation error, or scattering source in the literature). The extended source may contain energy away from the source position, when the velocity model is farther from the target one, which allows to fit the observed data to the synthetic. Huang et al. (2018a) call this method Matched Source Waveform Inversion (MSWI), and it is equivalent to WRI. Similar to WRI, source extension methods are performed in frequency domain (Huang and Symes, 2015; Huang et al., 2018a,b, 2019).

A time-domain implementation of source extension methods was initially introduced by Wang et al. (2017), which they achieve with a source extension approximation (their equations 5 to 7). Aghamiry et al. (2020) introduced another method for source extension using an explicit time-marching scheme, however it comes with a computational overhead, which stems from the backward-forward time-stepping recursion they use. They argue that this may be compensated for by the accelerated convergence, and the improved model parameters estimate accuracy. Gholami et al. (2022) propose a time-domain implementation of IR-WRI, however it does not account for the data-domain Hessian (it is approximated by a scaled identity matrix), which could be detrimental for complex geology, when starting from a crude initial model. More recently, a new time-domain extended source FWI (ES-FWI) implementation was introduced by Guo et al. (2024). ES-FWI does account for the data-domain Hessian, by means of a matching filter approximation (Liu and Peter, 2018). A comprehensive review of source extension is given by Huang et al. (2019), and more recently by (Operto et al., 2023).

Extended-receiver FWI is an alternative extension strategy, first introduced by Métivier and Brossier (2022). It relies on adding the additional degree of freedom at the receiver position, using a static relocalization strategy. Receiver extension is directly applicable to time-domain FWI, and has shown promising results in realistic numerical settings. However, the method is able to fit only the most energetic arrival, because the relocalization is static. The authors have also noticed a slow convergence of the method. In this study we build on the work of Métivier and Brossier (2022), introducing more freedom to the receiver position, using a time-dependent relocalization strategy. This is done to help obtain better fit for more complex data, and accelerate the convergence. Our new method is also directly applicable to time-domain FWI in a straightforward manner, and it is easy to tune. Time-dependent receiver extension relies on solving many small optimization problems (one problem per receiver), whose misfit functions are not convex, requiring the use of global optimization. The outline of the paper is as follows: a brief overview of conventional FWI as well as the state of the art extended-receiver FWI are given. Then we explain the time-dependent receiver extension, after which we discuss the underlying optimization problem. We then present a set of numerical experiments using a North Sea exploration scale synthetic model, providing an in-depth analysis of the sensitivity of our method to its tuning parameters. We conclude with a discussion about various aspects.

BACKGROUND

Full waveform inversion

Full waveform inversion is a PDE constrained optimization problem, where the fit between the observed and calculated data is improved by iteratively updating the model parameters. We write the FWI problem

$$\underset{m}{min}\ f(m) \ = \underset{m}{min} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T |d_{cal,s}[m](\mathbf{x}_r, t) - d_{obs,s}(\mathbf{x}_r, t)|^2 dt, \tag{1}$$

subject to

$$\begin{cases} A(m)u_s(\mathbf{x}, t) = b_s(\mathbf{x}, t) \\ d_{cal,s}[m](\mathbf{x}_r, t) = R_{s,r} u_s[m](\mathbf{x}, t), \end{cases} \tag{2}$$

where $d_{obs,s}$ are the observed data, and $d_{cal,s}$ are the calculated data. s and r are the source and receiver indices, respectively, N_s is the total number of sources, N_r is the total number of receivers, and $\mathbf{x}_r$ is the position of the receiver r. The calculated data are extracted from the wavefield u_s, which is computed using the wave equation operator $A(m)$ (equation 2) with m being the model parameters vector, and b_s being the source term. $R_{s,r}$ is the restriction operator, which extracts the wavefield values at the receiver positions using a convolution with a Dirac delta function

$$R_{s,r} u_s[m](\mathbf{x}, t) = \int_\Omega \delta(\mathbf{x} - \mathbf{x}_r) u_s[m](\mathbf{x}, t) d\mathbf{x}, \tag{3}$$

where $\delta(\mathbf{x})$ is the Dirac delta function and Ω is the computation domain. The constrained optimization problem (1) can be solved by finding the saddle-point of the associated Lagrangian (Haber et al., 2000). However, the computational cost of solving this problem using local optimization for large scale problems is prohibitive: the simultaneous update of the wavefields, the model parameters and the Lagrange multipliers requires their storage in memory, which is infeasible for large scale problems. In practice, the following reduced space approach is used

$$\underset{m}{min}\ f(m) = \underset{m}{min} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T |R_{s,r} A(m)^{-1} b_s(\mathbf{x}, t) - d_{obs,s}(\mathbf{x}_r, t)|^2 dt. \tag{4}$$

Problem (4) is an unconstrained optimization problem, where the wavefield u_s has been eliminated from the optimization variables, which exacerbates the non-linearity with respect to the model parameters. Problem (4) is solved -iteratively- using local optimization strategies (Nocedal and Wright, 2006). To do so, the gradient of the objective function needs to be computed. The adjoint state strategy is used to carry out this computation (Plessix, 2006). Following the adjoint state technique, the gradient is obtained using

$$\nabla_m f(m) = \sum_{s=1}^{N_s} \left\langle \frac{\partial A}{\partial m} u_s(\mathbf{x}, t), \lambda_s(\mathbf{x}, t) \right\rangle, \tag{5}$$

where the notation $\left\langle \bullet, \bullet \right\rangle$ indicates an inner-product operation. In time-domain FWI, it corresponds to the zero-lag cross-correlation between the weighted incident wavefield and the adjoint wavefield. The latter is computed by back-propagating the data residuals (difference between the observed and calculated data) injected at the receivers positions, following

$$\begin{cases} A(m)^T \lambda_s(\mathbf{x}, t) = \sum_{r=1}^{N_r} R_{s,r}^T \mu_s[m](\mathbf{x}_r, t) \\ \mu_s[m](\mathbf{x}_r, t) = d_{cal,s}[m](\mathbf{x}_r, t) - d_{obs,s}(\mathbf{x}_r, t), \end{cases} \quad (6)$$

where λ_s is the adjoint field, computed backwards in time using the wave equation operator $A^T(m)$. The source term is the data residuals μ_s injected at the receivers positions by the operator $R_{s,r}^T$. Once the gradient is obtained, the model can be updated. In all the examples we show in this paper, we use the preconditioned limited memory BFGS (l-$BFGS$) (Nocedal, 1980) algorithm. The new model m_{k+1} is therefore obtained by updating the model at iteration k following

$$m_{k+1} = m_k - \alpha_k Q_k \nabla_m f(m_k), \quad (7)$$

where k is the iteration number, α_k is a step-length obtained with a linesearch strategy (Nocedal and Wright, 2006), and Q_k is the inverse Hessian approximation obtained with the l-$BFGS$ algorithm, using l previously stored gradients.

Static receiver extension for FWI

Principle

Receiver extension introduces the receiver position as an additional degree of freedom (Métivier and Brossier, 2022). This additional degree of freedom allows to compensate for the kinematic mismatch between the observed and calculated data. In other words, by allowing the receiver to move in space, a fit can be obtained in the wrong medium velocity.

We write the receiver extension misfit function

$$\underset{m,\Delta x}{min}\,\tilde{f}(m,\Delta x) \;= \underset{m,\Delta x}{min}\,\frac{1}{2}\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\int_0^T |\tilde{d}_{cal,s}[m](\mathbf{x}_r+\Delta x_r,t)-d_{obs,s}(\mathbf{x}_r,t)|^2 dt+\alpha\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mathcal{P}_{1,s}[\Delta x_r]. \tag{8}$$

Equation (8) is a bivariate misfit function, depending on m and Δx_r, the latter being the additional degree of freedom, namely, the receiver relocalization. The first term in the right hand side is the data fit term, where $\tilde{d}_{cal,s}$ are the calculated (extended) data, which are extracted at the new receiver position. This new receiver position is shifted in space by the quantity Δx_r. The calculated data are thus obtained following

$$\tilde{d}_{cal,s}[m](\mathbf{x}_r+\Delta x_r,t) = \int_\Omega \delta(\mathbf{x}-(\mathbf{x}_r+\Delta x_r))u_s[m](\mathbf{x},t)d\mathbf{x} \stackrel{\text{def}}{=} \tilde{R}_{s,r}[\Delta x_r]u_s[m](\mathbf{x},t), \tag{9}$$

with $\tilde{R}_{s,r}$ being the extended restriction operator. The operator $\tilde{R}_{s,r}$ is similar to the conventional FWI restriction operator $R_{s,r}$, however, the extraction is now performed at the new receiver position $\mathbf{x}_r + \Delta x_r$, $\mathbf{x}_r$ being the true receiver position. The second term in the right hand side of equation (8) is a penalty term, which controls the receiver extension, and α is a tuning parameter. This term is used in order to prevent the relocalization from being too large. The expression of $\mathcal{P}_{1,s}$ is given in appendix A.

We illustrate the leading idea behind the receiver extension with a simple numerical experiment. In a homogeneous medium ($v_{true} = 2000$ m.s^{-1}) we consider one source/receiver couple. The observed trace is shown in a black dashed line in Figure 1b. The synthetic trace is computed in a medium with a higher velocity (2300 m.s^{-1}), and it is shown in a red solid line. By allowing the receiver to move (blue triangle in Figure 1a), the kinematic mismatch between the observed and calculated data is eliminated, and the traces are aligned in time. The extended (relocated) trace is shown in blue. We perform the same experiment for different velocities, and for different receiver relocalizations. This allows us to visualize the objective function of equation (8), for each velocity and for each receiver position. This is presented in Figure 2a, where the dashed black line shows the misfit variation at the original receiver position (conventional FWI). This misfit is not convex. However, if we take the minimum along the relocalization axis (the extended dimension) for each velocity following the red line, a convex misfit function is obtained.

[Figure 1 about here.]

[Figure 2 about here.]

Computing a numerical solution

Bivariate misfit functions such as equation (8), are usually minimized using nested-loops strategy. The outer loop is the conventional FWI optimization over the model parameters m. The inner-loop solves the sub-problem of finding the optimal receiver relocalization Δx for a given model m (we drop the subscripts s and r in this analysis for compactness). Consider the bivariate objective function $f(m, \Delta x)$ and the minimization problem

$$\min_{m,\Delta x} f(m, \Delta x), \tag{10}$$

where m is the model parameters vector and Δx is the receiver relocalization (a spatial shift). This problem is equivalent to

$$\min_{m} \hat{f}(m) \tag{11}$$

where we eliminate the variable Δx using

$$\hat{f}(m) = f(m, \overline{\Delta x(m)}), \tag{12}$$

with

$$\overline{\Delta x(m)} = \operatorname*{argmin}_{\Delta x} f(m, \Delta x). \tag{13}$$

The objective function $\hat{f}(m)$ is minimized in the outer loop (equation 11), while the inner loop carries out the minimization shown in equation (13) which defines $\Delta x(m)$. The inner loop aims at finding the optimal relocalization Δx for a given model iterate. In the framework of FWI, the gradient of the outer loop misfit function $\hat{f}(m)$ is required. This gradient is obtained following

$$\nabla_m \hat{f}(m) = \frac{\partial f(m, \overline{\Delta x(m)})}{\partial m} + \frac{\partial f(m, \overline{\Delta x(m)})}{\partial \Delta x} \frac{\partial \overline{\Delta x}}{\partial m}. \tag{14}$$

Per equation (12), $\overline{\Delta x}$ is a minimizer of $f(m, \Delta x)$ with respect to Δx, therefore, the first order optimality conditions tell us that

$$\frac{\partial f(m, \overline{\Delta x(m)})}{\partial \Delta x} = 0, \tag{15}$$

which yields

$$\nabla_m \hat{f}(m) = \frac{\partial f(m, \overline{\Delta x(m)})}{\partial m}. \tag{16}$$

Equation (16) shows that the gradient of the outer misfit function around m, is equal to the gradient with respect to m of the bivariate misfit function $f(m, \Delta x)$ calculated at $\overline{\Delta x}$. The gradient is obtained following

$$\nabla_m \tilde{f}(m) = \sum_{s=1}^{N_s} \left\langle \frac{\partial A}{\partial m} u_s[m](\mathbf{x}, t), \lambda_s[m, \overline{\Delta x}](\mathbf{x}, t) \right\rangle, \tag{17}$$

where the adjoint field λ_s is obtained using

$$\begin{cases} A(m)^T \lambda_s[m, \overline{\Delta x}](\mathbf{x}, t) = \displaystyle\sum_{r=1}^{N_r} \tilde{R}_{s,r}^T[\overline{\Delta x_r}] \tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r}, t) \\ \tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r}, t) = \tilde{d}_{cal,s}[m](\mathbf{x}_r + \overline{\Delta x_r}, t) - d_{obs,s}(\mathbf{x}_r, t). \end{cases} \tag{18}$$

This means that the calculated data are now extracted at the extended receiver position per equation (9), and the adjoint source position corresponds to the extended receiver as well. The latter is achieved using $\tilde{R}_{s,r}^T$, where the extended data residuals $\tilde{\mu}_s$ are injected at extended receiver positions. Throughout this paper, we use $\tilde{\bullet}$ to indicate extended quantities.

The extended receiver strategy with the nested-loops is summarized in algorithm 1. A forward modeling is performed in a given model, then the inner-loop computation provides the optimal receiver relocalizations. The extended calculated data are then extracted at the extended receiver positions. The model is then updated using gradient based optimization. It is crucial to note that the inner-loop computation does not require performing forward simulations, the extended data are simply extracted from the already computed incident wavefield. Métivier and Brossier (2022) use a grid-search to solve the inner loop problem. Indeed, finding one optimal receiver relocalization per receiver can be easily (and quickly) obtained with this global optimization scheme.

Algorithm 1 Nested loop optimization

while $\hat{f}(m_k) > \epsilon_{outer}$ **do**
  $u_s \leftarrow do\_forward\_modelling(m)$
  while $f(m_k, \Delta x) > \epsilon_{inner}$ **do**
    $f(m_k, \Delta x) \leftarrow compute\_inner\_cost(u_s)$
  end while
  $\overline{\Delta x} \leftarrow \Delta x$
  $\hat{f}(m_k) \leftarrow f(m_k, \overline{\Delta x})$
  $m_k \leftarrow update\_model(m_{k-1}, \overline{\Delta x}, u_s)$
  $k \leftarrow k + 1$
end while

Limitations of static receiver-extension

The extended-receiver FWI of Métivier and Brossier (2022) uses static relocalization, that is, the receiver position does not depend on time. Despite the promising results, the analysis was performed in the framework of a single arrival. They also observed that the method suffers from a slow convergence rate. We aim in the present study at addressing these two main issues. Our first investigation step is to introduce more degrees of freedom, in order to allow the receiver to move as a function of the acquisition time, as such a fit for more complex data can be obtained. By obtaining a better fit at earlier iterations, the convergence could be accelerated. This is carefully analyzed and illustrated in the present paper.

Before going into the formalism and the implementation details, we illustrate this strategy with a simple numerical experiment. We consider a two arrivals case, a transmitted and a reflected arrival. The observed data are simulated in a two layers medium, the first layer with $v_p = 2000$ m.s^{-1}, and a second layer with $v_p = 3500$ m.s^{-1} (Figure 3). For the sake of illustration, we choose a two layers starting model, where the first layer velocity is faster, at 2500 m.s^{-1}, while the second layer velocity is the true one ($v_p = 3500$ m.s^{-1}). The observed trace is shown in a dashed black line , and the calculated trace in a solid blue line. The receiver relocalization as a function of time is shown in red (Figure 4). This curve indicates the receiver relocalization value at each time step, we refer to this curve as the relocalization profile. Using receiver-extension, only one arrival can be fitted. This occurs since the second (reflected) arrival would require another relocalization value in order to align it with the observed reflected arrival. Therefore, only the most energetic arrival is fitted. Using time-dependent receiver extension (Figure 4c), a fit of both arrivals is obtained. We show the FWI kernels in Figure 5, for conventional FWI, static and time-dependent receiver extension. The kernel for the conventional FWI case suggests a positive velocity update (negative gradient), to a velocity that is already higher than the true value, in the first Fresnel zone (yellow arrows). With the static relocalization a correct velocity update is obtained in the first Fresnel zone, however two migration isochrones can be seen (red arrows), and the rabbit ears (blue arrows) suggest a positive velocity update. This is caused by the wrong fit of the second arrival (the adjoint source in Figure 5e shows two events related to the reflection). Using time-dependent receiver-extension makes it possible to fit both arrivals, obtaining the correct velocity update also in the rabbit-ears Fresnel zones. We also see a single migration isochrone, thanks to the fit of the second arrival. This time-dependent receiver-extension strategy is detailed in the next section.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

TIME-DEPENDENT RECEIVER EXTENSION FOR FWI

Formalism

We write the new extended-receiver FWI misfit function as

$$\min_{m,\Delta x(t)} \tilde{f}(m, \Delta x(t)) = \min_{m,\Delta x(t)} \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T |\tilde{d}_{cal,s}[m](\mathbf{x}_r + \Delta x_r(t), t) - d_{obs,s}(\mathbf{x}_r, t)|^2 dt$$

$$+\alpha \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \mathcal{P}_{1,s}[\Delta x_r(t)] + \beta \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \mathcal{P}_{2,s}[\Delta \dot{x}_r(t)]. \tag{19}$$

We aim at minimizing the bivariate misfit function in equation (19), where m are the model mechanical parameters, and $\Delta x_r(t)$ is now the time-dependent receiver relocalization. The first term in equation (19) is the data fit term, which is the L_2 norm of the data residuals. The calculated data are obtained using equation (9) however, the receiver position is now time-dependent. We write the new extended restriction operator

$$\tilde{R}_{s,r}[\Delta x_r(t)]u_s[m](\mathbf{x},t) = \int_{\Omega} \delta(\mathbf{x} - (\mathbf{x}_r + \Delta \mathbf{x}_r(t)))u_s(\mathbf{x},t)dx, \tag{20}$$

where $\Delta x_r(t)$ is the time-dependent receiver relocalization. The calculated data are extracted at the virtual receivers positions $(\mathbf{x}_r + \Delta x_r(t))$ from the wavefield $u_s(\mathbf{x},t)$. The term $\mathcal{P}_{1,s}$ in the right hand side of equation (19) penalizes the receiver relocalization, in order to prevent it from being too large, and to force it to become small along iterations. The parameter α is a tuning parameter, which weighs this penalty term. Similarly, the term $\mathcal{P}_{2,s}$ in the right hand side of equation (19) penalizes the receiver speed (the first order derivative with respect to time of the receiver relocalization), and β is a tuning parameter. This term is needed in order to mitigate potential Doppler effects, which stem from a moving receiver during the calculated data extraction, and also from a moving adjoint source during the adjoint simulation. The interested reader is referred to the discussion part of this paper. The full expressions of $\mathcal{P}_{1,s}$ and $\mathcal{P}_{2,s}$ can be found in appendix A. The adjoint field is obtained using the adjoint system

$$\begin{cases} A(m)^T \lambda_s[m, \overline{\Delta x(t)}](\mathbf{x}, t) = \sum_{r=1}^{N_r} \tilde{R}_{s,r}^T[\overline{\Delta x_r(t)}]\tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r(t)}, t), \\ \tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r(t)}, t) = \tilde{d}_{cal,s}[m](\mathbf{x}_r + \overline{\Delta x_r(t)}, t) - d_{obs,s}(\mathbf{x}_r, t). \end{cases} \tag{21}$$

We observe that now the adjoint source moves as a function of the acquisition time. The gradient is obtained following the adjoint state strategy

$$\nabla_m \tilde{f}(m) = \sum_{s=1}^{N_s} \Big\langle \frac{\partial A}{\partial m} u_s[m](\mathbf{x}, t), \lambda_s[m, \overline{\Delta x(t)}](\mathbf{x}, t) \Big\rangle. \tag{22}$$

As in the static case, the time-dependent strategy uses the nested-loops optimization approach. The inner-loop finds the optimal receiver relocalizations, while the outer-loop updates the model physical parameters. We now focus on the inner loop solution, starting with the question: how to parametrize the time-dependent receiver relocalization $\Delta x_r(t)$?

Parametrization

We seek to answer two questions: [1] how should we parametrize a receiver relocalization that depends on the acquisition time? [2] how should we parametrize the receiver motion in the physical space?

Temporal parametrization

One possible choice of parametrization is assigning a receiver relocalization to each time step. However, this would give rise to a problem with a large number of degrees of freedom (the number of time steps), which is prohibitive for the global optimization strategies, that is going to be used to solve the inner loop problem. To keep a minimal parametrization, we propose a piecewise polynomial interpolation. The time vector is divided into segments, each of which contains one Lagrange polynomial. We write the receiver relocalization as a function of the acquisition time

$$\Delta x(t) = \sum_{j=1}^{n_s} \sum_{k=1}^{N_\ell} a_{k+N_\ell \times (j-1)} \ell_k^{N_\ell}(t), \tag{23}$$

where $\ell_k^{N_\ell}(t)$ are Lagrange basis functions of order N_ℓ, n_s is the number of segments and a_i ($i = k + N_\ell \times (j-1)$) are the values at the control points. In the inner loop we aim at finding the optimal a_i values, that define the time-dependent virtual receivers positions. We illustrate in Figure 6, using three segments with first order Lagrange polynomials. The time-dependent relocalization is shown in black line plot, and the blue circles indicate the control points. The unknowns for the receiver relocalization subproblem are therefore $a = (a_1, a_2, .., a_{N_\ell \times n_s + 1})^T$, we rewrite the minimization problem of equation (19) as

$$\min_{m,a} \tilde{f}(m, a), \tag{24}$$

which we solve using the same nested loops approach described in algorithm 1.

[Figure 6 about here.]

Spatial parametrization

Métivier and Brossier (2022) use only horizontal relocalization, that is the receivers are allowed to move only following the horizontal axis, moving closer or farther from the source. In our case, relying solely on horizontal relocalization might not be advisable. This can be shown by a simple ray theory analysis (Benziane et al., 2023), to visualize the receiver positions which allow to fit transmitted and reflected arrivals. We consider a one layer over a half-space model (Fig. 7a). The travel-time expression of the reflected arrival in the true velocity model v_0 is

$$T_0^2 = \frac{x^2}{v_0^2} + \frac{4z^2}{v_0^2}, \tag{25}$$

where x is the offset and z is the receiver vertical distance to the reflector. To obtain the same arrival time in a wrong velocity model v_1, the receiver is relocated. We introduce the quantities Δx and Δz which provide the same arrival time

$$T_0^2 = \frac{(\Delta x + x)^2}{v_1^2} + \frac{(\Delta z + 2z)^2}{v_1^2}. \tag{26}$$

Equating equations (25) and (26) gives

$$(\Delta x + x)^2 + (\Delta z + 2z)^2 = (4z^2 + x^2)\frac{v_1^2}{v_0^2}. \tag{27}$$

Similarly, we perform the analysis for a transmitted arrival. The travel-time expression of the transmitted arrival in the true velocity model v_0 is

$$T_0^2 = \frac{x^2}{v_0^2}. \tag{28}$$

To obtain the same travel-time for the transmitted arrival in a wrong velocity model, the receiver is relocated

$$T_0^2 = \frac{\Delta z^2}{v_1^2} + \frac{(\Delta x + x)^2}{v_1^2}. \tag{29}$$

Equating equations (28) and (29) gives

$$(\Delta x + x)^2 + \Delta z^2 = x^2 \frac{v_1^2}{v_0^2}. \tag{30}$$

Equations (27) and (30) are conic section equations. For the reflection case, equation (27) describe a circle with a center $(-x, -2z)$, which is the source image with respect to the reflector. The radius of this circle is $\sqrt{(4z^2 + x^2)\frac{v_1^2}{v_0^2}}$. We plot the solution of equation (27) using different velocities, $v_1 = 1500\ m.s^{-1}$, $v_1 = v_0 = 1500$ m.s^{-1} and $v_1 = 2500$ m.s^{-1}. The possible receiver positions that give the same travel-time fall on a circle, the radius of which increases as the velocity increases. By allowing the receiver to move only horizontally, a fit cannot be obtained for the lower velocity case (crosses "$\times$" in Figure 7b). Similar to the reflected arrival case, equation (30) describes a circle. However, it is centered at the source position $(-x, 0)$. We plot the solution to our equation using three different velocities (Figure 7c). We observe that moving the receiver horizontally, makes it possible to fit the transmitted arrival, for all velocities.

By forcing the receiver to move towards the source following a $\frac{\pi}{4}$ angle with respect to the horizontal axis, a fit of the reflected and transmitted arrivals can be obtained. This is achieved by simply equating Δx to Δz when the receiver moves towards the source. Therefore, Δz can be obtained from Δx using

$$\Delta z = \begin{cases} |\Delta x| & \text{if the receiver is moving towards the source} \\ 0 & \text{otherwise.} \end{cases} \tag{31}$$

This is the choice of parametrization that we make, in order to keep the parametrization minimal. Again, allowing the receiver to move freely in all the spatial dimensions, would give rise to an expensive inner-loop computation. This is avoided thanks to our spatial parametrization.

[Figure 7 about here.]

We illustrate this parametrization with a numerical experiment, using the same setup as in Figure 3. However, we use now a top layer velocity with a lower velocity. Because the starting velocity is lower than the true one, the receiver needs to move towards the source. If the receiver moves only horizontally, a fit for the reflected arrival cannot be obtained. This is shown in Figure 8a, where the observed data are shown in a black dashed line, and the extended calculated data are shown in a blue solid line. The receiver needs to move following the z-axis in order to obtain a fit for the second arrival. This is in agreement with our ray theory analysis. By allowing the receiver to move vertically (equation 31), a fit for the reflected arrival is obtained.

[Figure 8 about here.]

INNER LOOP OPTIMIZATION

Overview

The solution of the inner loop problem using our time-dependent receiver extension raises a challenging optimization problem. We illustrate this using a North Sea exploration scale synthetic model (Figure 9). We consider a single source/receiver couple, and a single segment with first order Lagrange polynomial, leading to two control points a_1 and a_2. We compute the inner-loop misfit function map, using a fine discretization of a_1 and a_2, which we present in Figure 10. Not only our misfit function contains numerous minima: some local minima may have very close values (Figure 10b). This makes finding the global minimum a challenging task. A solution can be obtained using global optimization strategies. However, it appears not tractable to solve the inner problem using grid-search optimization, as the size of this inverse problem will grow large with the time-dependence. Even with the parametrization used for the present example (two control points), as the optimization needs to be performed for all receivers. Our investigations on global optimization strategies have led us to the choice of Particle Swarm Optimization (PSO) (Kennedy and Eberhart, 1995), over Markov-chain Monte Carlo (McMC) (Aster et al., 2013) and Very Fast Simulated Annealing (VFSA) (Ingber, 1993). The former requires a large number of misfit function evaluations to convergence to a good solution, while the latter suffers from premature convergence towards bad solutions. PSO on the other hand, allows for a thorough exploration of the search space, with reasonable cost. The interested reader is referred to the discussion part of the paper.

[Figure 9 about here.]

[Figure 10 about here.]

Particle Swarm Optimization (PSO)

Particle Swarm Optimization is a computational intelligence technique, proposed by Kennedy and Eberhart (1995). It is a heuristic optimization method where the search space is explored by so-called particles, in order to minimize some misfit function $g(\mathbf{x})$

$$\min_{\mathbf{x}} g(\mathbf{x}). \tag{32}$$

A swarm contains N_p particles, and each particle j explores the search space by its position $\mathbf{x}_j$. The best model from the swarm (from all the particles), associated with the lowest $g(\mathbf{x})$ value, is referred to as the global best ($\mathbf{x}_g$ in the equations below). The personal best ($\mathbf{x}_{p,j}$ in the equations hereafter) on the other hand, is the best solution obtained for each individual particle. Let $\mathbf{x}_j^i$ denotes a particle j position in a search space $\mathbb{R}^N$ at iteration i. The particle position is then updated to iteration $i+1$ as such

$$\mathbf{x}_j^{i+1} = \mathbf{x}_j^i + \mathbf{u}_j^{i+1}, \text{ with } \mathbf{x}_j^0 = U(\mathbf{x}_{min}, \mathbf{x}_{max}), \tag{33}$$

where $\mathbf{u}_j^i$, is the particle position update at iteration i. The starting particle positions are harvested from a uniform distribution (U in equation (33), where $\mathbf{x}_{min}$ and $\mathbf{x}_{max}$ are the search space bounds). In the literature, $\mathbf{u}^i$, is referred to as the particle speed, and is computed as

$$\mathbf{u}_j^{i+1} = \omega \mathbf{u}_j^i + c_1 \mathbf{r}_1^i : [\mathbf{x}_{p,j}^i - \mathbf{x}_j^i] + c_2 \mathbf{r}_2^i : [\mathbf{x}_g^i - \mathbf{x}_j^i]. \tag{34}$$

The first term in the right hand side is called the inertia term, it controls the contribution from the past iteration, with ω being the inertia weight (typically $\omega \in [0.9, 1.2]$). The second term in the right hand side is the contribution of the best position for each particle, where $\mathbf{x}_{p,j}^i$ is the best personal position for a particle j along its past trajectory. The third term in the right hand side is the contribution of the global best position of the swarm, where $\mathbf{x}_g^i$ is the global best position. Terms $\mathbf{r}_1^i$ and $\mathbf{r}_2^i$ are random variables vectors of the same dimension as the particle position $\mathbf{x}_j^i$, which are harvested from a uniform distribution. c_1 and c_2 are constants, usually set to equal values, and $\mathbf{a} : \mathbf{b}$ denotes the term to term product of vectors $\mathbf{a}$ and $\mathbf{b}$. If we wish to give more weight to either component, the constants c_1 and c_2 may be adjusted accordingly. This is the basic form of global best PSO (Engelbrecht, 2007).

Numerical example

In order to showcase PSO, we carry out a simple numerical test using the same example we showed earlier (Figure 9). We run PSO using 16 particles. We show snapshots of the swarm configuration at iterations 1, 40, 80 and 160 in Figure 11. The particles are shown in black, the personal best in red, the global best in magenta and the global solution which we obtain using a grid search is shown as a red star. The personal best positions are initialized with the particle positions and are updated at each iteration. The global best is selected from the personal bests (the particle whose personal best has the lowest cost). This process is repeated until convergence or the max number of iterations is reached. Note that none of the starting particles positions is near the global minimum, nonetheless, PSO manages to converge fairly quickly. As is clear from Figure 10, our misfit does contain many secondary minima, which is why a population based optimization is a good choice. It allows for a thorough exploration of the search space. Even though the global-best is already on the global minimum, other particles still search in the vicinity of other secondary minima, thanks to the inertia term (equation 34). PSO can achieve convergence fairly quickly, however, setting the maximum number iterations to a small value for all receivers is not a good idea. Indeed, convergence to a good solution is not guaranteed for the same number of iterations, for all receivers. A flexible way of handling this, is the use of what we call stalling detection. If the global best does not move in n_{stall} iterations the PSO is stopped, as we assume the global minimum has been reached. This helps to save computational time. For the first example (Figure 11), the max number of iterations is set to 400 and $n_{stall} = 200$. n_{stall} is particularly high in this example because the swarm is small ($N_p = 16$). For a larger swarm, stalling is observed earlier. For $N_p = 16$ the global best has indeed ceased to change from the 275^{th} iteration. However the optimization did not stop in this case, because the global best needs to stall for 200 iterations. Next, we increase the number of particles to $N_p = 32$, in order to observe the impact of the swarm size on the convergence (Figure 12). Indeed, increasing the swarm size, leads to a sooner stalling, therefore a better convergence. With 32 particles (shown in blue), stalling occurs at 127 iterations, PSO has stopped at 327 iterations. When using 64 particles (shown in red), the global best did not change after 34 iterations, PSO was stopped at 234 iterations in this case. Better convergence is obtained with a larger swarm size, because increasing the number of particles, allows for a better exploration of the search space. In other words, at each iteration, a larger swarm has a better "view" of the search space. It is clear that the swarm size is an important parameter. In the literature a choice of ≈30 particles is common. A choice of too few particles reduces the exploration abilities of the swarm, and choosing too many particles –although better in terms of convergence– requires more cost function evaluations (Luu et al., 2018). In our work, a swarm size is adapted based on the inner-loop parametrization. Higher dimensional problems would benefit from a larger swarm. All the tests performed in this paper (unless otherwise stated) use 45 particles. As for the choice of the PSO parameters, we use the results from Pedersen (2010). We set $c_1 = -0.6485$, $c_2 = 2.6475$ and $\omega = -0.6485$.

[Figure 11 about here.]

[Figure 12 about here.]

APPLICATION TO A NORTH SEA EXPLORATION SCALE SYNTHETIC 2D MODEL

For all our testing, we use the SEISCOPE (acoustic and visco-acoustic) modeling and FWI engine TOYxDAC_TIME (Yang et al., 2018a). The modeling is performed using fourth order finite-differences, with staggered grids (Virieux, 1986; Levander, 1988). A second order leap-frog scheme is used for the time-marching. We use Convolutional Perfectly Matched Layers (CPML) (Komatitsch and Martin, 2007) as absorbing boundary conditions, in order to simulate an infinite medium. We note however that we do not enable CPML when attenuation is used, instead, we use sponge layers (Cerjan et al., 1985). The gradient computation is performed using the time-decimated incident wavefield, which is saved solely at the boundary of the computation domain. This wavefield is interpolated using Kaiser-windowed sinc interpolator (Yang et al., 2016c), and is propagated from the boundaries during the adjoint simulation. When an attenuating medium is used, we employ Checkpointing Assisted Reverse-Forward Simulation or CARFS (Yang et al., 2016b). We note that the synthetic data extraction and adjoint sources injection are performed using Kaiser-windowed sinc interpolation as well (Hicks, 2002).

We use a North Sea exploration scale synthetic model in our testing. First we consider a constant density noise-free experimental setup (inverse crime), for which we carry out an in-depth analysis as well as a sensitivity study. Second, we design a more realistic experimental setup using the same model.

Inverse crime settings

We first generate the observed data using the true model shown in Figure 13a. The finite differences grid points spacing is set to 25 m, the time step is 0.003 s and the total number of time steps is 4000. The source wavelet is a Ricker wavelet with 4 Hz central frequency, it is filtered with a high-pass filter with cutoff frequency of 2 Hz (Figure 14). The acquisition used for the tests in this section is a fixed-spread acquisition, with 128 sources, spaced with 132 m, and 170 receivers spaced with 100 m. The starting model is obtained by a Gaussian smoothing of the true model (Figure 13b).

[Figure 13 about here.]

[Figure 14 about here.]

Before diving into a deeper analysis, we make a comparison of conventional FWI, extended-receiver FWI with static relocalization ($\alpha = 0.375$) and our time-dependent approach ($\alpha = 0.01$, $\beta = 0.25$, with one segment and two control points). We perform 300 iterations for all cases, the results are shown in Figure 15. As expected conventional FWI struggles to reconstruct the true velocity, from this crude starting model. Static receiver extension provides a better reconstruction, however with visible defects. The time-dependent approach on the other hand, is able to better reconstruct the velocity model from this crude starting model. We show the model error as a function of FWI iterations in Figure 16. The model error is computed using

$$E = \frac{100}{M} \sum_{i}^{M} = \frac{|m_{est,i} - m_{true,i}|}{|m_{true,i}|}, \tag{35}$$

where M is the total number of model points, $m_{est,i}$ is the estimated model at point i and $m_{true,i}$ is the true model at point i. The model error for the conventional FWI increases at first, then decreases to 9%, a value higher than the error at the starting model. Static receiver extension performs better: the model error is mainly decreasing. The time-dependent approach provides the best reconstruction, the model error decreases quickly to a lower model error.

To better understand why it works, we take a look at the data fit for both static and time-dependent extension approaches (Figures 17 and 18). At the first iteration a better fit is obtained using the time-dependent approach (blue and black on the Figures indicate a good fit). As the model estimate improves, the relocalization tends to zero, this is apparent at the last iteration. This can be better assessed by looking at the adjoint source (data residuals) in Figure 19. The adjoint source at the last iteration of the time-dependent approach has the lowest values, compared to the static approach, and the conventional FWI. This is indicative of a better model reconstruction. Time-dependent receiver extension is able to reach a lower model error, and a lower relocalization at the final iteration. This is not the case for the static approach, where at the last iteration, the relocalization is still important. Better explaining the data when the model estimate is poor, leads to a better model reconstruction.

[Figure 15 about here.]

[Figure 16 about here.]

[Figure 17 about here.]

[Figure 18 about here.]

[Figure 19 about here.]

To complement our analysis, we show the misfit function evolution as a function of FWI iterations, for the static and time-dependent cases in Figures 20a and 20b, respectively. The data fit term (L_2 norm of the data residuals without using relocalization) is shown as well (red line plot). The L_2 misfit increases at first while the extended-receiver FWI is decreasing, which indicates that L_2 FWI would have been stuck in a local minimum. We note also the rapid reduction of the time-dependent approach cost function, as opposed to the static counterpart. The L_2 norm of the data residuals for the static relocalization case increases at first (similar to the time-dependent approach), but then it stagnates after a brief decrease, which indicates a slow convergence. Next, we take a look at the receiver relocalization evolution as a function of outer-loop iterations. It is shown solely for the leftmost, center and rightmost shot gathers. We obtain it using

$$R_1 = \frac{1}{N_r} \sum_{r=1}^{N_r} \sqrt{\int_0^T |\Delta x_r(t)|^2 dt}. \tag{36}$$

The receiver relocalization can viewed as a proxy for model error, larger relocalization indicates a poor model estimate. The relocalization decreases almost monotonically for both approaches (Figure 21), however, the time-dependent approach reaches lower relocalization, faster. For the time-dependent case we can also visualize the receiver speed as a function of outer loop iterations, which is obtained using

$$R_2 = \frac{1}{N_r} \sum_{r=1}^{N_r} \sqrt{\int_0^T |\Delta \dot{x}_r(t)|^2 dt}. \tag{37}$$

The receiver speed decreases monotonically as the model estimates improves (Figure 22). The receiver speed is not too large (maximum of ≈ 500 m.s^{-1} for the leftmost and rightmost gathers), thanks to the second penalty term. This is an expected behavior, as the model estimate improves, there is less need for the receiver to move too fast.

For the tests we showed here, we set $\alpha = 0.375$ for the static approach, and for the time-dependent approach we set $\alpha = 0.01$ and $\beta = 0.25$. How would extended-receiver FWI behave when these tuning parameters are perturbed? We answer this question in the next paragraphs.

[Figure 20 about here.]

[Figure 21 about here.]

[Figure 22 about here.]

Sensitivity to the tuning parameters

In this section we investigate how the tuning parameters in our cost function (α and β) impact the model reconstruction, the data fit and the convergence. We carry out three sets of tests: [1] α variations impact on static receiver extension, [2] α variations impact on time-dependent receiver extension for a constant value of β, and [3] α and β variation impact on the time-dependent receiver extension. We keep the same experimental setup (Figure 13). For each experiment 300 outer loop iterations are performed. For the inner loop, we use one segment with first order Lagrange polynomial giving a two degrees of freedom problem, which we solve with a grid-search. We do this in order to avoid the PSO tuning parameters impacting our sensitivity testing.

Static receiver extension: sensitivity to α

For this first set of experiments we test 24 equally spaced values of $\alpha \in]0, 1]$. Extended-receiver FWI is then carried out using static relocalization. We show the cost function at the last iteration as well as the model error as function of α in Figure 23a and 23b, respectively. The cost function at the last iteration increases –in the most part– with increasing α with a few outliers. The first outlier corresponds to $\alpha = 0.125$, it is caused by a line-search failure (that is, the line-search process in the outer loop was unable to find an adequate step length). The second outlier corresponds to the lowest cost value that was achieved with $\alpha = 0.375$. Two other points, $\alpha = 0.333$ and $\alpha = 0.416$ do not follow the trend as well. These four points and two others, namely, $\alpha = 0.583$ and $\alpha = 0.791$ are shown as black circles in Figures 23a and 23b. The corresponding models are shown in Figure 24, this is discussed in the next paragraph. Increased cost at the last iteration with increasing α is -in theory- expected, as the relocalization is constrained more with increasing α. However, we note that the static relocalization is sensitive to small variation in the tuning parameter. As for the model error, it varies with α, but the variation is less apparent. We also note that the model error is high for all α values, which indicates that the model reconstructions are not satisfactory. The relocalization at the first and last iterations (Figures 23c and 23d, respectively) are as expected, the values decrease with increasing α values. It is interesting to note that the relocalization value at the last iteration, for small α values is high. This means that the model reconstruction is not satisfactory, otherwise, the relocalization would tend to zero.

We show the reconstructed models that correspond to the selected α values in Figure 24. The selected α values are shown as circles in Figures 23a 23b. This approach appears to be sensitive to the choice of α. Small variations of the tuning parameter, lead to a significant change in the reconstructed model. This means that the static relocalization approach is difficult to tune.

[Figure 23 about here.]

[Figure 24 about here.]

Time-dependent relocalization: sensitivity to α

We perform a similar numerical experiment using our time-dependent strategy, we set $\beta = 0.25$ and we test the 24 $\alpha \in]0,1]$. The results are shown in Figure 25. The choice of α appears to have less impact on the cost function at the last iteration, and lower values are reached, this is not the case for the static approach. Note that we use the same normalization to obtain the normalized cost, for the static and the time-dependent approaches, this is done in order to keep the cost function plots comparable. The tuning parameter does not have a significant impact on the model error, we note also that a lower model error is reached for all α values, compared to the static counterpart. The impact of this tuning parameter on the relocalization at the first iteration appears to be –roughly– linear, the relocalization decreases with increasing α, which is an expected behavior. As for the last iteration, α appears to have no observable effect, and the relocalization tends to zero. This is an indication of good model reconstruction. Our approach appears to be less sensitive to the choice of the tuning parameter α. We show the reconstructed models (Figure 26) corresponding to the same selected α values for the static case. These values are indicated by black labeled circles in Figure 25a and 25b. The effect of α on the reconstruction appears to be minimal. This is an encouraging observation, as it means that the method is easy to tune. For these tests, the β value has been kept constant, how does our method behave when both α and β are perturbed? To answer this question, we devise a parametric study which we discuss in the next paragraphs.

[Figure 25 about here.]

[Figure 26 about here.]

Time-dependent relocalization: sensitivity to α and β

In order to understand how the two tuning parameters α and β impact our strategy, we conduct a parametric study scanning for a different α and β values. This is a computationally expensive test, therefore, we use a non-regular grid. The tuning parameters values are shown in Table 1, where the first row shows the α and β values we consider, and the second row shows the additional values that are considered only for the β parameter.

[Table 1 about here.]

Following the same setup as before, we run extended-receiver FWI for 300 iterations for each combination of α and β. We show the result in Figure 27 using a logarithmic scale for both axes. Similar to the previous test, the time-dependent strategy appear to be less sensitive to the choice of α. This can be seen in the cost obtained at the last iteration, and also in the model error, shown in Figures 27a and 27a, respectively. The minimum cost is obtained for $\alpha = 0.025$ and $\beta = 0.25$, and the minimum model error for $\alpha = 0.0075$ and $\beta = 0.75$. We show the corresponding final models in Figures 28a and 28b. The reconstructed models are mostly good, regardless of the choice of the tuning parameters. However, for large β values, the model reconstruction is impacted, which can be seen on the relocalization (leftmost and rightmost shot points). This occurs when β is large: the method then behaves as static relocalization, the receiver speed being heavily constrained. The best results are obtained for reasonably low β values (not greater than 0.75 based on our findings from this experiment). This test concludes the investigations done in inverse crime settings. Next, we design a more realistic experimental setup.

[Figure 27 about here.]

[Figure 28 about here.]

Realistic settings

We now use the same North Sea exploration scale 2D synthetic model in a more realistic setting. First, the observed data are computed under the visco-acoustic approximation using the V_p, ρ and Q_p models, shown in Figure 29. The source wavelet is a Ricker centered at 4 Hz. It is filtered with a high-pass filter with cutoff frequency of 2 Hz. As for the acquisition, it is a fixed-spread setup, with 128 sources spaced with 117.5 m, and 150 receivers spaced with 100 m. The discretization step used for the finite differences is set to 12.5 meters, the time step is set to 0.0015 seconds, and the number of time steps is 6000. The data are decimated to a time step of 0.003 seconds, and a band-limited Gaussian noise is added (Figure 30). The discretization step used for the forward computations during the inversion is set to 25 meters. For the receiver extension, we set $\alpha = 0.01$ for both the static and time-dependent cases. Only for the time-dependent approach, we set $\beta = 0.0025$, and the time-dependent relocalization is parametrized with one segment and two control points.

[Figure 29 about here.]

[Figure 30 about here.]

The starting velocity model is a 1D model (obtianed with $V_P^{1D}(z) = 0.38z + V_P^{water}$), the starting density model is obtained using Gardner's law on the starting velocity model, which is given by $\rho(x) = 1740(10^{-3}V(x))^{\frac{1}{4}}$. The quality factor is set to 100 everywhere, except in the water layer where it is set to 1000. The starting models are shown in Figure 31. For these numerical tests we only update the velocity during the inversion. First we perform a source time function estimation in the starting velocity model (Pratt, 1999), the result of which is shown in Figure 32. The inversion is performed for 300 iterations using conventional FWI, static receiver extension, and time-dependent receiver extension, with the previously computed wavelet.

[Figure 31 about here.]

[Figure 32 about here.]

We show the reconstructed models after the 15, 110, 205 and 300 iterations in Figure 33. As expected, conventional FWI is unable to reconstruct the velocity from the 1D starting model. Static relocalization performed a bit better, however, some high velocity artifacts are present in the low velocity anomaly in the center of the model. The time-dependent approach provides the best model reconstruction. The low velocity anomaly is fully reconstructed, as well as most of the higher velocity basement. We show the data fit as well as the relocalization gathers, for the first and last iterations for both approaches in Figures 34 and 35. A good fit is obtained at the first iteration for both strategies. However, at the last iteration the data fit prior to relocalization is better for the time-dependent case. Moreover, the receiver relocalization is lower at the last iteration for the time-dependent case, which indicates a better model reconstruction. These results in a 2D realistic setting are very encouraging.

[Figure 33 about here.]

[Figure 34 about here.]

[Figure 35 about here.]

DISCUSSION

In this section we discuss the following points: [1] the use of the second penalty term ($\mathcal{P}_2$ in equation 19) and the Doppler effect, [2] the choice of PSO as our global optimization scheme, and [3] finally the computational overhead of our strategy with respect to more conventional FWI approaches.

The term $\mathcal{P}_2$ and the Doppler effect

The second penalty term is added to constrain the receiver speed. This is important as moving receivers, or moving sources alter the frequency content of the data. In our case, we encounter both situations. The calculated data extraction from the wavefield at a moving receiver positions, with a stationary source, causes a change in the frequency content. A moving source is encountered during the adjoint simulation, where the adjoint source (our receiver), is moving as a function of time, which in turn, can cause the frequency content change. These frequency content changes are attributed to the Doppler effect. For the moving receiver case, these effects can be better understood by looking at the Doppler effect formula

$$f = \frac{v_p \pm v_r}{v_p} f_o, \tag{38}$$

where f is the observed frequency, and f_o is the emitted frequency. v_p is the medium velocity and v_r is the receiver speed. The latter is added if the receiver is moving towards the source, and subtracted in the other case. Equation (38) is for the case where the source is stationary and the receiver is moving. If the source is moving and the receiver is stationary (the adjoint simulation), the observed frequency is given by

$$f = \frac{v_p}{v_p \pm v_s} f_o, \tag{39}$$

where v_s is the source speed, it is added if the source is moving away from the source, and it is subtracted in the other case. We illustrate the moving receiver case in Figure 36, using the same 2 layers setup shown in Figure 4. The spectrum of the calculated data, which is extracted at the time-dependent receiver position without using the second penalty term ($\beta = 0$) is shown in red. The blue curve is obtained with $\beta = 0.001$, and the green curve is for the static relocalization case (the receiver position does not depend on time). In the case where no constraint is imposed on the receiver speed (red plot) we can see that the spectrum is different. In particular, energy at low frequencies is added, this is due to Doppler effect for a receiver moving away from the source. When the receiver speed is constrained ($\beta = 0.001$), the spectrum is much closer to the one obtained with $\beta = \infty$, therefore mitigating the Doppler effect.

[Figure 36 about here.]

Another interesting observation can be made for the experiment where the source is moving (an adjoint source, our receiver). We show the adjoint field at time 3.18 s in Figure

37a. The source is moving to the right at a speed which is higher than the medium velocity, the structure that appears on the left of the moving source is called a Mach cone. The gradient resulting from this adjoint field is shown in Figure 37b, the positions occupied by the moving receiver are superimposed to the gradient, the color map indicates the receiver speed. Numerical artifacts are present in our gradient, which are caused by the frequency increase, which in turn, stems from the Doppler effect.

[Figure 37 about here.]

Choice of a global optimizer

In a previous work (Benziane et al., 2023), we have investigated various methods for the solution of the inner loop problem. Namely, Markov-chain Monte Carlo methods (Aster et al., 2013), simulated annealing (Sen and Stoffa, 2013), and a variation of it, Very Fast Simulated Annealing or VFSA (Ingber, 1992, 1993). McMC is in fact a very good method for exploring the parameter space, and inferring the posterior distribution. The sought distribution (the posterior) is inferred by randomly sampling a proposal distribution. Each sample is either accepted or rejected using the Metropolis-Hastings criterion (Metropolis et al., 1953; Hastings, 1970). However, it is very costly, because many misfit evaluations are required for the Markov-chain to converge to a stationary distribution. Simulated annealing is similar to McMC, as it also uses a Markov-chain, however, the probability of accepting a candidate solution is decreased along the iterations, making use of the so-called ”cooling schedules”. The cooling schedule forces the acceptance probability to be large for most candidate solutions at earlier iterations, which translates to a high acceptance rate. This probability is reduced using the cooling schedule, which could drastically reduce the acceptance rate. Simulated annealing is not very well adapted to our problem, as it suffers from premature convergence. This premature convergence to a local minimum occurs when the best candidate reached by the Markov-chain is a local minimizer, while the probability of acceptance decreases, making this local minimizer overwhelmingly probable. Furthermore, it was notoriously difficult to tune, namely, choosing a cooling schedule and its parameters. Very Fast Simulated Annealing suffers from convergence issues for our problem, although, it is easier to tune than conventional simulated annealing, as it relies on a single cooling schedule and a single generating distribution, which reduces the number of tuning parameters. Grid-search is of course costly, even for the simplest parametrization we can consider (one segment with first order Lagrange polynomial, giving a two dimensional problem).

In order to illustrate this, we use a North Sea Exploration Scale synthetic model. We keep the same setup, which is shown in Figure 9. We consider two segments with first order Lagrange polynomials, giving a three degrees of freedom problem. We compare different global optimization schemes, namely, McMC, VFSA and PSO. First, we sample the inner loop misfit function using McMC (Figure 38). This plot is obtained by cross-plotting each dimension against another, couple by couple. In other words, for a given point sampled by McMC in this 3-D space, with the coordinates (a_1, a_2, a_3), we plot its position in 2-D as such, a_1 is plotted against a_2 and then against a_3. We assign the value of the cost function to the sample providing the color map shown. The diagonal shows the histogram in each dimension. Thanks to the McMC sampling, we can see the inner loop misfit function, which contains many local minima. The misfit function corresponding to the couples a_3/a_1, and a_2/a_3, varies slowly compared to the other couples, and the histograms for a_3 is flat. This is expected, as the point points a_1 and a_3 correspond to the start and the end of the seismic trace, respectively. The data values at early times, as well as very late time (time steps closer to the end of the trace) are close to zero. This reduces the sensitivity to a_1 and a_3, this is particularly true for a_3.

We carry out the inner-loop optimization using McMC, VFSA and PSO. We show the evolution of the cost as a function of iterations in Figure 39. Note that the McMC plot shows only samples that reduce the cost function, otherwise, the plot would be cluttered. For this McMC example, a total of 2.5×10^6 misfit evaluations have been performed. But the plots show only 25000 samples, as we did a burn-in period of 5000 iterations and a skipping step of 500 iterations. This is done to avoid the correlation effects between samples, which stem from the pseudo-random number generation process (Aster et al., 2013). McMC did in fact provide a good solution, but it required a large number of cost function evaluations. VFSA did not converge to a meaningful solution, which is clear in the cost function plot. PSO gave the most interesting result, converging quickly to the global minimum. This population based strategy is well adapted to our problem, this can be seen in Figure 39b. The personal best cost for few particles are shown, different particles explore in the vicinity of various minima, leading to a convergent behavior of the swarm. That is why we use PSO for the inner loop optimization. Furthermore, PSO tuning was less challenging thanks to the few published meta-optimization studies (Pedersen, 2010; Mason et al., 2018), where we obtained the PSO parameters.

[Figure 38 about here.]

[Figure 39 about here.]

Computational cost

The CPU times we show in Table 2 correspond to the realistic setting example we showed earlier (Figure 29). Please note that we use Checkpointing Assisted Reverse Forward Simulation (CARFS) for the gradient computation (Yang et al., 2016a) for all our numerical experiments. We have a slight increase in the forward computation CPU time for both static and time-dependent receiver extension, this is caused by the decimation and storage of a portion of the wavefield that we use for the extension. The main computation overhead comes from the inner loop optimization. The computational burden for the static receiver extension is minor. However, for the time-dependent approach the inner loop computation is more important. An increase in the gradient computation CPU time is caused by the adjoint-source injection. We recall the adjoint equation

$$\begin{cases} A(m)^T \lambda_s[m, \overline{\Delta x(t)}](\mathbf{x}, t) = \sum_{r=1}^{N_r} \tilde{R}_{s,r}^T[\overline{\Delta x_r(t)}] \tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r(t)}, t), \\ \tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r(t)}, t) = \tilde{d}_{cal,s}[m](\mathbf{x}_r + \overline{\Delta x_r(t)}, t) - d_{obs,s}(\mathbf{x}_r, t). \end{cases} \tag{40}$$

The cost of adjoint simulation is more important for the time-dependent approach, as we use Kaiser windowed sinc interpolation (Hicks, 2002) to inject the adjoint the sources at moving receivers positions. This is performed for all receivers at every time step. Even though the cost increase is not negligible, it is certainly manageable. Also, the convergence to a good model is obtained faster.

To put the computation overhead into perspective, we take a look at the time complexities. The time complexity for the forward and adjoint simulations is $\mathcal{O}(n^2)$ in 2D and $\mathcal{O}(n^3)$ in 3D. As for the inner-loop, the complexity in term of the number of receivers is $\mathcal{O}(n)$ in 2D, and $\mathcal{O}(n^2)$ in 3D. The large number of receivers in 3D can constitute a bottleneck. We address this point in the conclusion of this paper.

[Table 2 about here.]

CONCLUSION AND PERSPECTIVES

Conclusion

Extended-receiver FWI with time-dependent relocalization have shown promising results. The ease with which this method is directly applied to time-domain FWI, as well as the simplicity of tuning its misfit function are very encouraging. Another advantage is that there is no need to adjust the tuning parameters during the inversion. The speed of convergence from very crude starting models is another attractive feature. However, the method comes with a non-negligible but certainly manageable computation cost.

Perspectives and ongoing work

The next step of our work is extending the method to 3D as well as an application to a field dataset. The 3D extension of this method relies on the idea of allowing the receivers to move either towards or away from the source, following a fixed angle defined by the source and receiver positions. The preliminary results in 3D synthetic settings are encouraging.

The 3D implementation comes with a caveat: the increasing the number of receivers. This has encouraged us to develop an alternative parametrization and optimization for the inner-loop, in order to make the cost manageable for 3D application. This strategy relies on the observation that the receiver relocalization problem broken down into a series of nested smaller problems, in a similar fashion to dynamic time warping (Hale, 2013), which can be solved deterministically. This approach is considerably less costly than the stochastic approach described herein. Testing in 3D settings using our dynamic programming approach is ongoing.

APPENDIX A

FULL MISFIT FUNCTION EXPRESSION FOR EXTENDED-RECEIVER FWI

836 We write the full expression of the receiver-extension misfit function

$$\min_{m,\Delta x} \tilde{f}(m, \Delta x) = \min_{m,\Delta x} \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T |\tilde{d}_{cal,s}[m](\mathbf{x}_r + \Delta x_r, t) - d_{obs,s}(\mathbf{x}_r, t)|^2 dt$$
$$+ \frac{\alpha}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} ||d_{obs,s,r}||_\infty^2 \frac{||\Delta x_{s,r}||_2^2}{L^2}. \quad \text{(A-1)}$$

837 the second term in the right hand side is the penalty term ($\mathcal{P}_1$ in equation 8), it prevents
838 the relocalization from being too large, and forces the receiver to its original position as
839 the model estimate improves. In this penalty term, L is the maximum allowed receiver
840 relocalization, and α is a tuning parameter.

841 Similarly, we write the misfit function of the extended-receiver FWI with time-dependent
842 relocalization

$$\min_{m,\Delta x} \tilde{f}(m, \Delta x) = \min_{m,\Delta x(t)} \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int_0^T |\tilde{d}_{cal,s}[m](\mathbf{x}_r + \Delta x_r(t), t) - d_{obs,s}(\mathbf{x}_r, t)|^2 dt$$
$$+ \frac{\alpha}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} ||d_{obs,s,r}||_2^2 \frac{||\Delta x_{s,r}||_2^2}{L^2}$$
$$+ \frac{\beta}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} ||d_{obs,s,r}||_2^2 \frac{||\Delta \dot{x}_{s,r}||_2^2}{V_{max}^2} \quad \text{(A-2)}$$

843 the second term in the right hand side of equation A-2 ($\mathcal{P}_1$ in equation 19) penalizes the receiver relocalization, in order to prevent it from being too large, and to force it to become small along iterations. L is the maximum allowed receiver relocalization, and α is a tuning parameter, for weighting this penalty term. Similarly, the third term in the right hand side of equation A-2 ($\mathcal{P}_2$ in equation 19) penalizes the receiver speed (the first order derivative with respect to time of the receiver relocalization). V_{max} is the maximum allowed receiver speed, and β is a tuning parameter.

REFERENCES

Aghamiry, H., A. Gholami, and S. Operto, 2018, Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers: Geophysics, **84**.

———, 2019a, Admm-based multi-parameter wavefield reconstruction inversion in VTI acoustic media with TV regularization: Geophysical Journal International, **219**, 1316–1333.

———, 2019b, Implementing bound constraints and total-variation regularization in extended full waveform inversion with the alternating direction method of multiplier: application to large contrast media: Geophysical Journal International, **218**, 855–872.

———, 2020, Accurate and efficient wavefield reconstruction in the time domain: Geophysics, **85(2)**, A7–A12.

Aster, R. C., B. Borchers, and C. H. Thurber, 2013, Chapter eleven - bayesian methods, *in* Aster, R. C., B. Borchers, and C. H. Thurber, eds., Parameter Estimation and Inverse Problems (Second Edition), 253–280. Academic Press, second edition edition.

Barnier, G., E. Biondi, R. G. Clapp, and B. Biondi, 2023a, Full-waveform inversion by model extension: Practical applications: GEOPHYSICS, **88**, R609–R643.

———, 2023b, Full-waveform inversion by model extension: Theory, design, and optimization: GEOPHYSICS, **88**, R579–R607.

Benziane, M., R. Brossier, and L. Métivier, 2023, Receiver extension strategy using a time-dependent relocalization approach for time-domain fwi: 84th EAGE Annual Conference & Exhibition, 1–5, European Association of Geoscientists & Engineers.

Biondi, B. and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79(3)**, WA129–WA140.

Biondi, B. and P. Sava, 1999, Wave-equation migration velocity analysis: Annual Meeting Abstracts, 1723–1726, Society Of Exploration Geophysicists.

Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

Brossier, R., S. Operto, and J. Virieux, 2009, 2D elastic frequency-domain full-waveform inversion for imaging complex onshore structures: 71^{th} Annual EAGE Conference & Exhibition, Amsterdam, Expanded Abstracts, U019, EAGE.

———, 2015, Velocity model building from seismic reflection data by full waveform inversion: Geophysical Prospecting, **63**, no. 2, 354–367.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition for discrete acoustic and elastic wave equations: Geophysics, **50**, 2117–2131.

Chauris, H. and E. Cocher, 2017, From migration to inversion velocity analysis: Geophysics, **82**, S207–S223.

Engelbrecht, A. P., 2007, Computational intelligence: an introduction, chapter 16, 289–358. John Wiley & Sons, Ltd, 2nd edition. OCLC: ocn133465571.

Gholami, A., H. S. Aghamiry, and S. Operto, 2022, Extended-space full-waveform inversion in the time domain with the augmented lagrangian method: GEOPHYSICS, **87**, R63–R77.

894 Guasch, L., M. Warner, and C. Ravaut, 2019, Adaptive waveform inversion: Practice: Geophysics, **84(3)**, R447–R461.

896 Guo, G., S. Operto, A. Gholami, and H. S. Aghamiry, 2024, Time-domain extended-source full-waveform inversion: Algorithm and practical workflow: GEOPHYSICS, **89**, R73–R94.

899 Haber, E., U. Ascher, D. Aruliah, and D. Oldenburg, 2000, Fast simulation of 3d electromagnetic problems using potentials: Journal of Computational Physics, **163**, 150–171.

901 Hale, D., 2013, Dynamic warping of seismic images: Geophysics, **78**, S105–S115.

902 Hastings, W. K., 1970, Monte carlo sampling methods using markov chains and their applications.

904 Hicks, G. J., 2002, Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions: Geophysics, **67**, 156–166.

906 Huang, G., R. Nammour, and W. W. Symes, 2018a, Source-independent extended waveform inversion based on space-time source extension: Frequency-domain implementation: Geophysics, **83**, R449–R461.

909 ———, 2018b, Volume source-based extended waveform inversion: Geophysics, **83**, R369–387.

911 Huang, G., R. Nammour, W. W. Symes, and M. Dolliazal, 2019, Waveform inversion via source extension, 4761–4766.

913 Huang, G. and W. W. Symes, 2015, Full waveform inversion via matched source extension: SEG Technical Program Expanded Abstracts, 1320–1325.

915 Ingber, L., 1992, Very fast simulated re-annealing: Mathematical and Computer Modelling, **12**, 967–973.

———, 1993, Simulated annealing: Practice versus theory: Mathematical and Computer Modelling, **18**, 29–57.

Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, and D. Trezeguet, 1989, Wavelengths of Earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910.

Kennedy, J. and R. Eberhart, 1995, Particle swarm optimization: Proceedings of ICNN'95 - International Conference on Neural Networks, 1942–1948 vol.4.

Komatitsch, D. and R. Martin, 2007, An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation: Geophysics, **72**, SM155–SM167.

Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436.

Liu, Q. and D. Peter, 2018, One-step data-domain least-squares reverse-time migration: Geophysics, **83**, 1–46.

Luo, S. and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: SEG Technical Program Expanded Abstracts, **30**, 2788–2792.

Luo, Y. and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Luu, K., M. Noble, A. Gesret, N. Belayouni, and P. F. Roux, 2018, A parallel competitive particle swarm optimization for non-linear first arrival traveltime tomography and uncertainty quantification: Computers and Geosciences, **113**, 81–93.

Ma, Y. and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion: Geophysics, **78**, R223–R233.

Mason, K., J. Duggan, and E. Howley, 2018, A meta optimisation analysis of particle swarm optimisation velocity update equations for watershed management learning: Applied Soft Computing, **62**, 148–161.

Métivier, L. and R. Brossier, 2022, Receiver-extension strategy for time-domain full-waveform inversion using a relocalization approach: Geophysics, **87**, R13–R33.

Métivier, L., R. Brossier, Q. Mérigot, and E. Oudet, 2019, A graph space optimal transport distance as a generalization of L^p distances: application to a seismic imaging inverse problem: Inverse Problems, **35**, 085001.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: Application to full waveform inversion: Geophysical Journal International, **205**, 345–377.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, Equation of state calculations by fast computing machines: The journal of chemical physics, **21**, 1087–1092.

Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–773.

Nocedal, J. and S. J. Wright, 2006, Numerical optimization: Springer, 2nd edition.

Operto, S., A. Gholami, H. Aghamiry, G. Guo, S. Beller, K. Aghazade, F. Mamfoumbi, L. Combe, and A. Ribodetti, 2023, Extending the search space of full-waveform inversion beyond the single-scattering born approximation: A tutorial review: GEOPHYSICS, **88**, R671–R702.

Pedersen, M. E. H., 2010, Good parameters for particle swarm optimization: Presented at the .

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

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model: Geophysics, **64**, 888–901.

Sen, M. K. and P. L. Stoffa, 2013, Global optimization methods in geophysical inversion (second edition): Elsevier Science Publishing Co.

Shipp, R. M. and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: Geophysical Journal International, **151**, 325–344.

Sirgue, L. and R. G. Pratt, 2004, Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

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

van Leeuwen, T. and F. Herrmann, 2016, A penalty method for PDE-constrained optimization in inverse problems: Inverse Problems, **32(1)**, 1–26.

van Leeuwen, T. and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195(1)**, 661–667.

van Leeuwen, T. and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394.

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

Virieux, J. and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, C., D. Yingst, P. Farmer, I. Jones, G. Martin, and J. Leveille, 2017, Reconstructed full-waveform inversion with the extended source, 1449–1453.

Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, WA13–WA24.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016a, Checkpointing-assisted reverse forward simulation: an optimal recomputation method for fwi and rtm: SEG Technical Program Expanded Abstracts 2016, 1089–1093.

———, 2016b, Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method: Geophysics, **81**, R349–R362.

———, 2018a, 3-D time-domain VTI viscoacoustic full waveform inversion: Application to Valhall field data: Presented at the Expanded Abstracts, 80^{th} Annual EAGE Meeting (Copenhagen).

Yang, P., R. Brossier, and J. Virieux, 2016c, Wavefield reconstruction from significantly decimated boundaries: Geophysics, **80**, T197–T209.

Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018b, Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion: Geophysics, **83**, R43–R62.

Yao, G., D. Wu, and S.-X. Wang, 2020, A review on reflection-waveform inversion: Petroleum Science, **17**, 334–351.

Yong, P., R. Brossier, L. Métivier, and J. Virieux, 2022, Localized adaptive waveform

1009 inversion: theory and numerical verification: Geophysical Journal International, **233**,
1010 1055–1080.

LIST OF FIGURES

1 Receiver extension illustration. (a) Acquisition setup, the source is shown as a black star, the receiver position as a red triangle, and the extended receiver in blue, (b) observed trace (black dashed line), calculated trace (red solid line), and calculated trace extracted at the extended receiver position (blue solid line). 64

2 Illustration of the receiver-extension misfit function. (a) Misfit function 2D view as a function of velocity and receiver relocalization. The minimum is shown in white, the misfit values corresponding to the receiver-extension are shown in a red dashed line. The dashed magenta line indicates the velocity value at which the calculated data in Figure 1 are computed. (b) Misfit function profile along the zero-relocalization (conventional L_2 misfit) in a solid black line, misfit function profile along the minimum value over the receiver relocalization $\Delta x_{s,r}$ in a dashed red line. 65

3 Experimental setup (a): true velocity model, (b): the starting model used for gradient illustration hereafter. 66

4 Observed and calculated data, as well as relocalization time profiles for: (a) conventional FWI, (b): static relocalization and (c): time-dependent relocalization. 67

5 FWI kernels and adjoint sources for (a,d): conventional FWI, (b,e): static relocalization and (c,f): time-dependent relocalization. The yellow arrow indicates the first Fresnel zone, the blue indicates the rabbit-ears and the red arrows show the migration isochrones. 68

6 Time-dependent parametrization example, using three segments with first order Lagrange polynomials. The circles indicate the control points a_i. . . . 69

7 Simple reflection case analysis with receiver extension. The geometry is shown in (a), the original and extended receiver positions shown as a black, and white inverted triangles, respectively. Δx and Δz are the horizontal and vertical receiver relocalizations. The possible receiver relocalizations at different velocities for the reflection and transmission cases are shown in (b) and (c), respectively. 70

8 Data fit and relocalization profiles for a two arrivals case using a a lower top layer starting velocity. (a): Time-dependent horizontal relocalization, and (b): time-dependent relocalization with both vertical and horizontal components. The observed trace is shown in a black dashed line, the calculated is shown in a blue solid line, the horizontal relocalization is shown in a red solid line, and the vertical in a purple solid line. 71

9 A North Sea exploration scale synthetic model, (a): true model, (b): starting model obtained by Gaussian smoothing of the true model. 72

10 Inner-loop illustration, (a): misfit function map, using one segment and two control points, the global minimum is shown as a red star. (b): A cross section following the black line in (a). Note how close are the values of the global minimum (red vertical dashed line), and the secondary minimum (blue vertical dashed line). (c): the observed shown in a black dashed line, calculated in a red solid line and the extended data corresponding to the global minimum shown in a blue solid line. 73

11 PSO example showing the swarm configuration at different iterations superimposed on the misfit map, the personal best positions are shown in cyan and the global best in magenta, the global solution (obtained with grid-search) is shown as a red star. 74

12 Global best normalized cost as a function of the number of iterations using 16, 32 and 64 particles. 75

13 A North Sea exploration scale synthetic model. (a): true model used to generate the observed data, (b): the starting model. 76

14 The source wavelet used for the numerical experiment. (a): time-domain (b): frequency domain. 77

15 Reconstructed velocity models after 300 iterations. (a): conventional FWI, (b): static extended-receiver FWI and (c): time-dependent extended-receiver FWI. 78

16 Receiver relocalization as a function of iterations,(a): static extended-receiver FWI, (b): time-dependent FWI 79

17 Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f). 80

18 Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f). 81

19 Adjoint source of the center shot point at the 300^{th} iteration for (a): conventional FWI, (b): static receiver extension and (c): time-dependent receiver extension. The same color scale is used for all figures. 82

20 Cost function evolution as a function of iterations, extended-receiver FWI cost is plotted in black and the L_2 norm of the data residuals at each iteration is shown in red. (a): static extended-receiver FWI, (b): time-dependent extended-receiver FWI. 83

21 Receiver relocalization as a function of iterations,(a): static extended-receiver FWI, (b): time-dependent FWI. 84

22 Receiver speed as a function of iterations for the time-dependent extended-receiver FWI. 85

23 Tuning parameter α impact on extended-receiver FWI with static relocalization. (a): normalized cost at the last iteration, (b): reconstructed model error, (c): relocalization at the first iteration and (d): relocalization at the last iteration. The lower two panels show plots for the leftmost, center and rightmost shot points. The black labeled circles in (a) and (b) correspond to selected α values, whose corresponding final models are shown in Figure 24, where each model label correspond to an α value with the same label. . 86

24 Tuning parameter α impact on the velocity reconstruction using extended-receiver FWI with static relocalization. (a): $\alpha = 0.125$, (b): $\alpha = 0.333$, (c): $\alpha = 0.375$, (d): $\alpha = 0.4167$, (e): $\alpha = 0.5833$ and (f) $\alpha = 0.7917$. Each sub-figure label corresponds to an α value indicated by black labeled circles in Figure 23a and 23b. 87

25 Tuning parameter α impact on extended-receiver FWI with time-dependent relocalization. (a): normalized cost at the last iteration, (b): reconstructed model error, (c): relocalization at the first iteration and (d): relocalization at the last iteration. The lower two panels show plots for the leftmost, center and rightmost shot points. The black labeled circles in (a) and (b) correspond to selected α values, whose corresponding final models are shown in Figure 26, where each model label correspond an α value with the same label. 88

26 Tuning parameter α impact on the velocity reconstruction using extended-receiver FWI with time-dependent relocalization. (a): $\alpha = 0.125$, (b): $\alpha = 0.333$, (c): $\alpha = 0.375$, (d): $\alpha = 0.4167$, (e): $\alpha = 0.5833$ and (f) $\alpha = 0.7917$. Each sub-figure label corresponds to an α value indicated black labeled circles in Figure 23a and 23b. 89

27 Tuning parameters impact on the (a): cost function at the last iteration, (b): model error and (c,d) receiver relocalization, shown here for (c) the leftmost and (d) the rightmost shot points. 90

28 Reconstructed models corresponding to the (a): lowest cost and the (b): lowest model error (Figure 27). 91

29 True models used to generate the observed data. (a): velocity, (b): density and (c): quality factor. The density and the quality factor in the water layer are 1000 kg.m^{-3} and 1000, respectively. The color-bars in (b) and (c) are clipped for clarity, therefore, the color in the water layer is not representative of the true values. 92

30 The observed data computed in a North Sea exploration scale synthetic model, (a) leftmost shot-gather, (b): center shot-gather and (c): rightmost shot-gather. 93

31 Starting models. (a): 1D velocity model, (b): density obtained using Gardner's law and (c): quality factor, set to 100 everywhere and 1000 in the water column. he color-bars in (b) and (c) are clipped for clarity, as a result, the color in the water layer is not representative of the true values. . . . 94

32 Source wavelet estimation in the starting velocity model. (a): source time function, (b): source frequency amplitude spectrum. 95

33 Reconstructed velocity models for (a-d) conventional FWI, (e-h) static receiver extension and (i-l): time-dependent receiver extension. This is shown for (a,e,i): 15 iterations, (b,f,j): 110 iterations (c,g,k): 205 iterations and (d,h,l) 300 iterations. 96

34 Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f). 97

35 Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f). 98

36 Calculated data amplitude spectrum for three cases: time-dependent relocalization without the second penalty term ($\beta = 0$, shown in red) and with the second penalty term ($\beta = 0.001$), as well as the static case ($\beta = \infty$). The notches in the spectra are caused by the free-surface reflection. 99

37 Gradient computation with $\beta = 0$. (a): adjoint field snapshot at $3.18\ s$, the extended receiver (adjoint source) is moving to the right causing the Mach cone shape in the wavefield, (b) extended receiver FWI kernel, the receiver positions are shown, the color indicates the receiver speed. 100

38 McMC sampling of the inner misfit function, samples of each degree of freedom are cross-plotted against another, the color map indicates the value of the cost function. The plots on the diagonal show the histograms of samples of each degree of freedom. 101

39 Cost function values as a function of inner-loop iterations for one receiver. (a): McMC candidates that lower the cost functions are shown in black, PSO global-best cost is shown in green, and Very Fast Simulated Annealing best solutions are shown in magenta. (b): the personal-best cost function of few particles as a function of iterations. 102

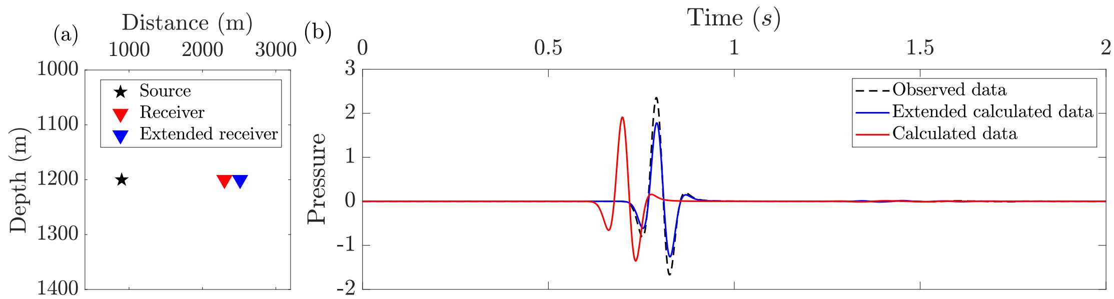

Figure 1: Receiver extension illustration. (a) Acquisition setup, the source is shown as a black star, the receiver position as a red triangle, and the extended receiver in blue, (b) observed trace (black dashed line), calculated trace (red solid line), and calculated trace extracted at the extended receiver position (blue solid line).

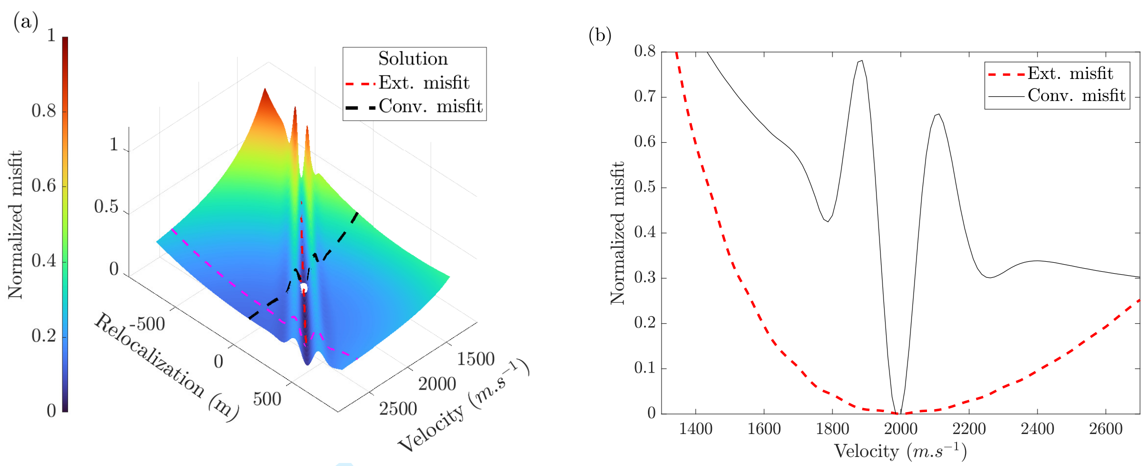

Figure 2: Illustration of the receiver-extension misfit function. (a) Misfit function 2D view as a function of velocity and receiver relocalization. The minimum is shown in white, the misfit values corresponding to the receiver-extension are shown in a red dashed line. The dashed magenta line indicates the velocity value at which the calculated data in Figure 1 are computed. (b) Misfit function profile along the zero-relocalization (conventional L_2 misfit) in a solid black line, misfit function profile along the minimum value over the receiver relocalization $\Delta x_{s,r}$ in a dashed red line.

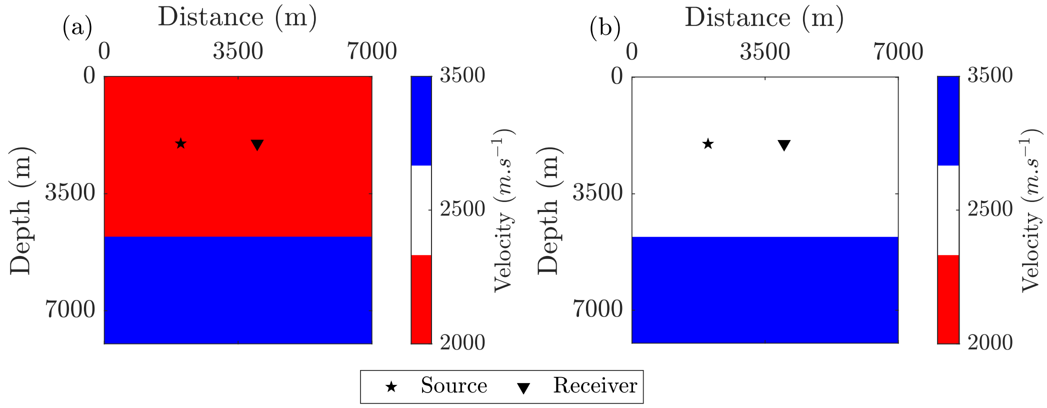

Figure 3: Experimental setup (a): true velocity model, (b): the starting model used for gradient illustration hereafter.

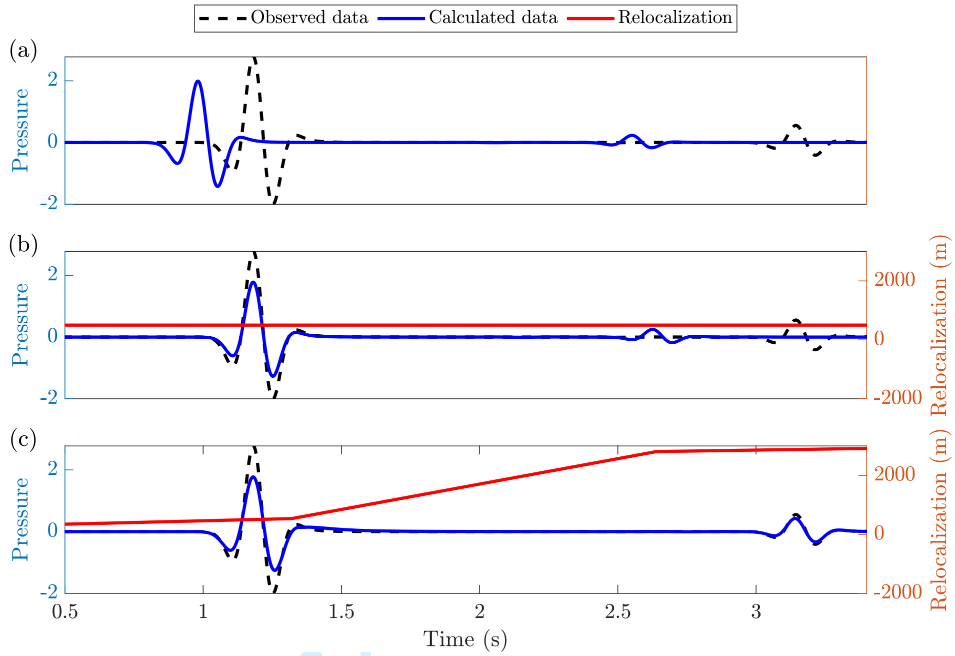

Figure 4: Observed and calculated data, as well as relocalization time profiles for: (a) conventional FWI, (b): static relocalization and (c): time-dependent relocalization.

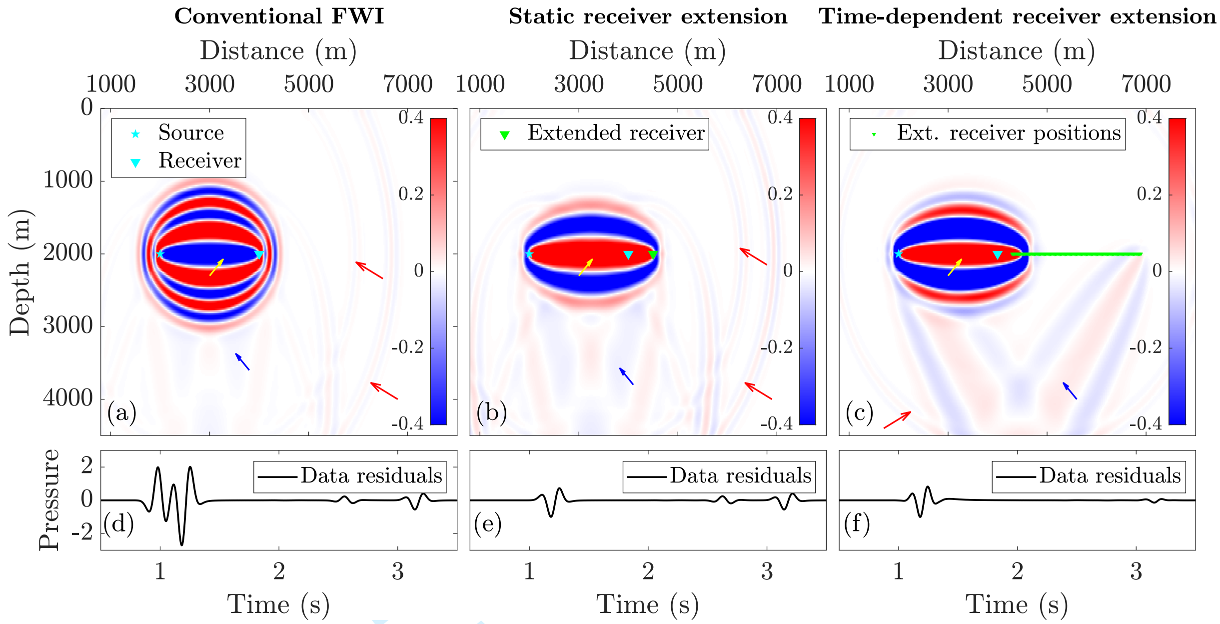

Figure 5: FWI kernels and adjoint sources for (a,d): conventional FWI, (b,e): static relocalization and (c,f): time-dependent relocalization. The yellow arrow indicates the first Fresnel zone, the blue indicates the rabbit-ears and the red arrows show the migration isochrones.

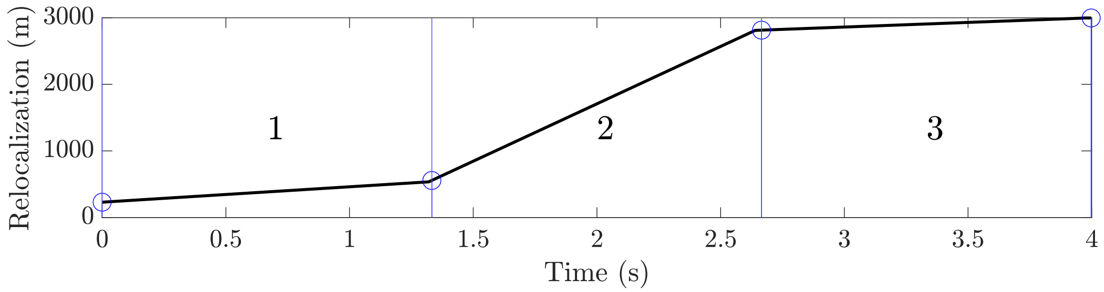

Figure 6: Time-dependent parametrization example, using three segments with first order Lagrange polynomials. The circles indicate the control points a_i.

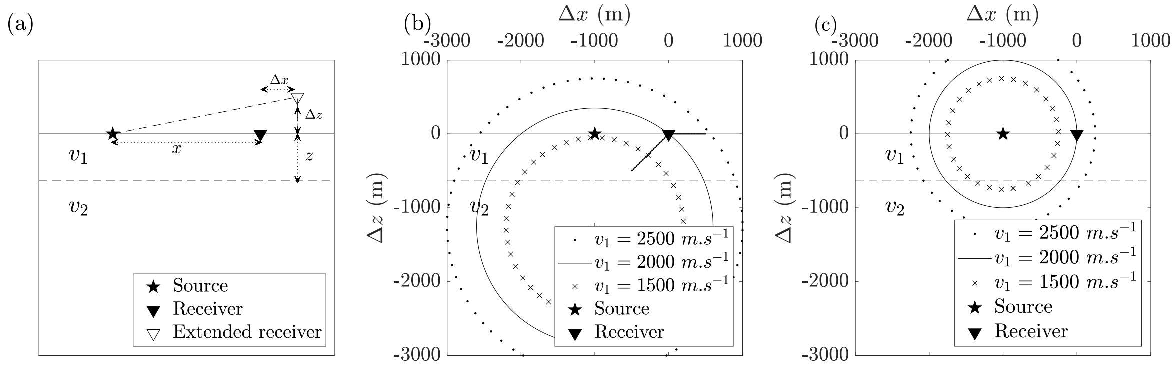

Figure 7: Simple reflection case analysis with receiver extension. The geometry is shown in (a), the original and extended receiver positions shown as a black, and white inverted triangles, respectively. Δx and Δz are the horizontal and vertical receiver relocalizations. The possible receiver relocalizations at different velocities for the reflection and transmission cases are shown in (b) and (c), respectively.

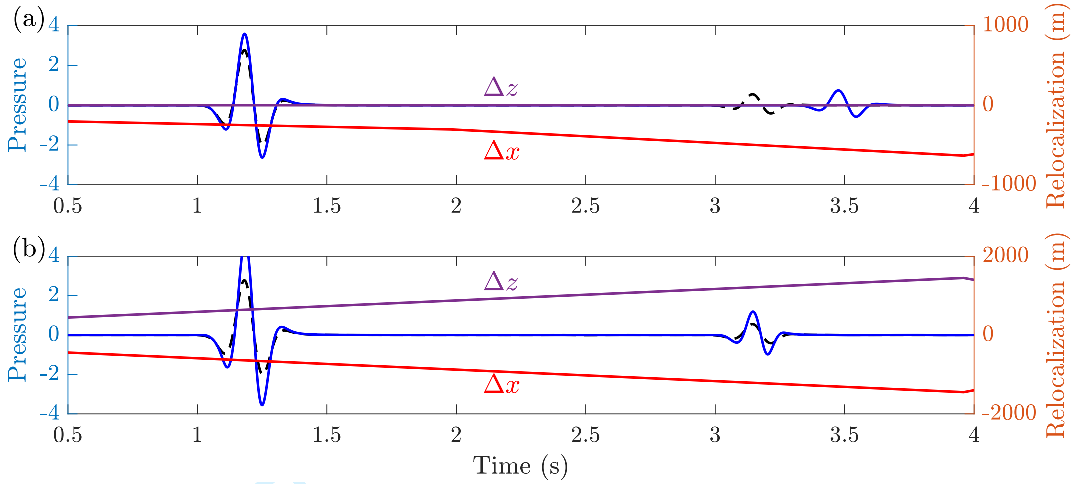

Figure 8: Data fit and relocalization profiles for a two arrivals case using a a lower top layer starting velocity. (a): Time-dependent horizontal relocalization, and (b): time-dependent relocalization with both vertical and horizontal components. The observed trace is shown in a black dashed line, the calculated is shown in a blue solid line, the horizontal relocalization is shown in a red solid line, and the vertical in a purple solid line.

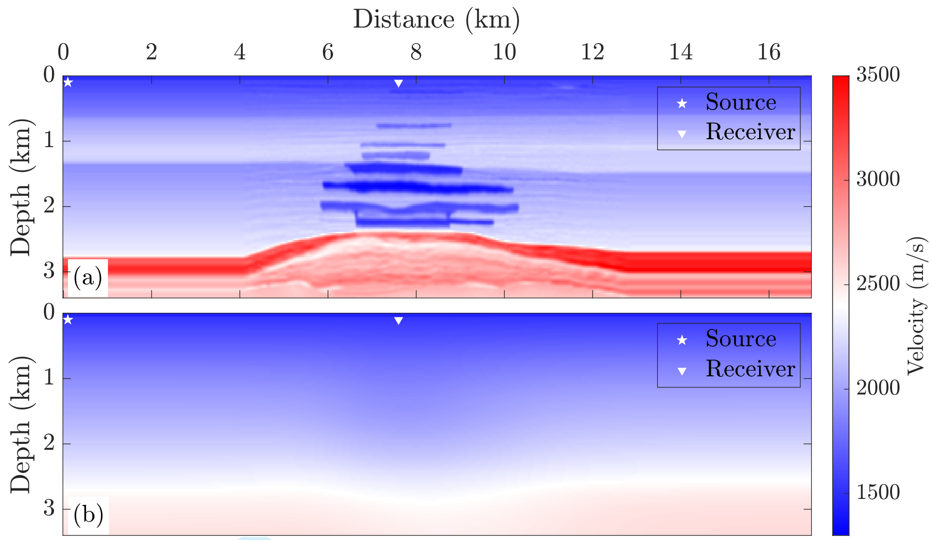

Figure 9: A North Sea exploration scale synthetic model, (a): true model, (b): starting model obtained by Gaussian smoothing of the true model.

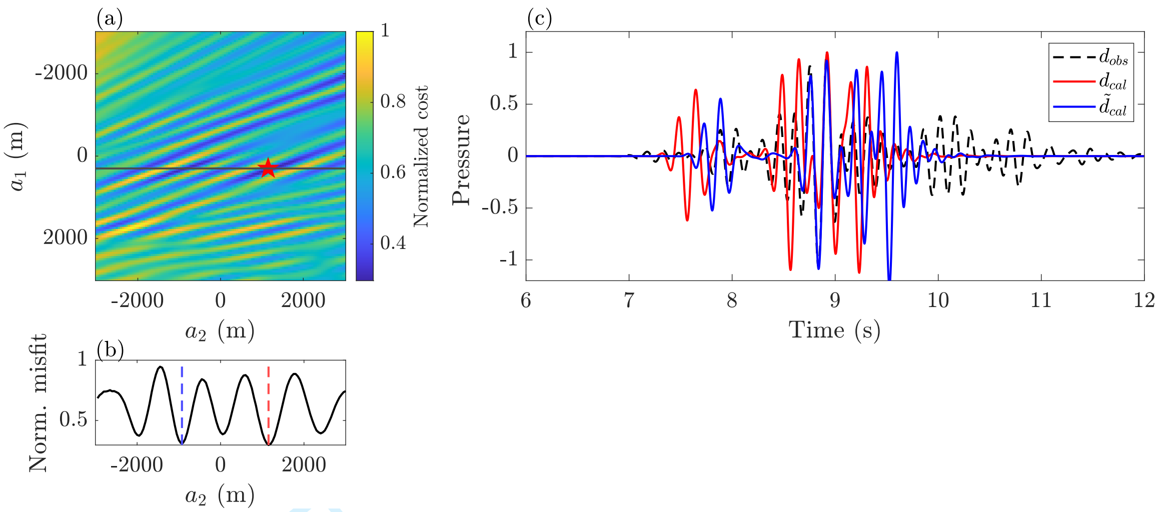

Figure 10: Inner-loop illustration, (a): misfit function map, using one segment and two control points, the global minimum is shown as a red star. (b): A cross section following the black line in (a). Note how close are the values of the global minimum (red vertical dashed line), and the secondary minimum (blue vertical dashed line). (c): the observed shown in a black dashed line, calculated in a red solid line and the extended data corresponding to the global minimum shown in a blue solid line.

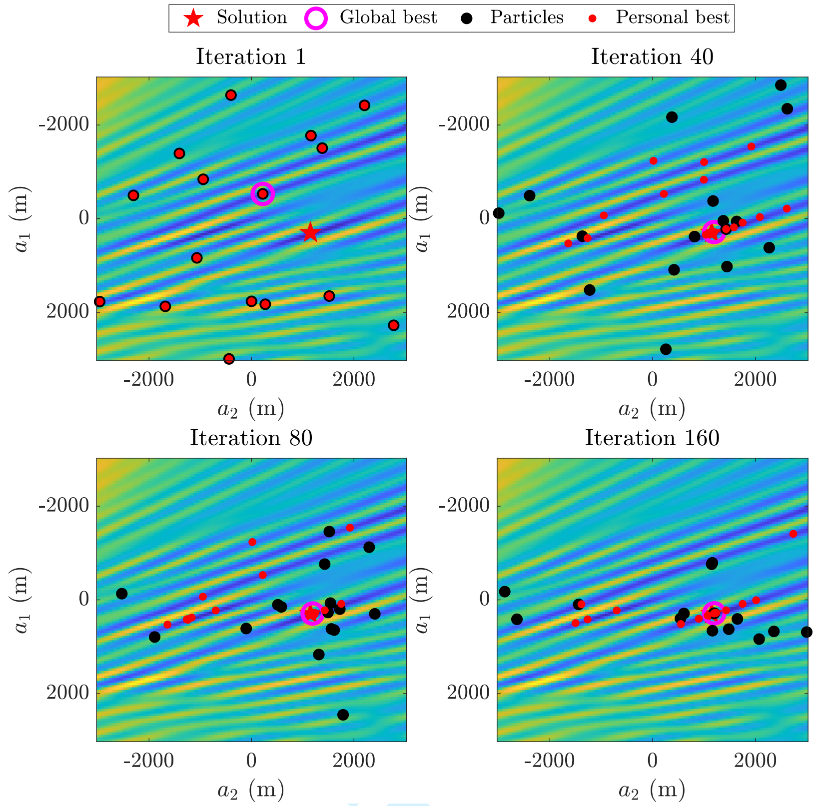

Figure 11: PSO example showing the swarm configuration at different iterations superimposed on the misfit map, the personal best positions are shown in cyan and the global best in magenta, the global solution (obtained with grid-search) is shown as a red star.

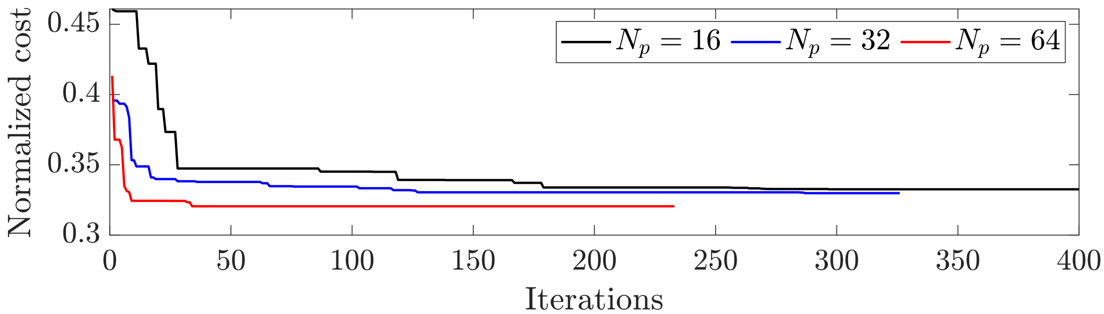

Figure 12: Global best normalized cost as a function of the number of iterations using 16, 32 and 64 particles.

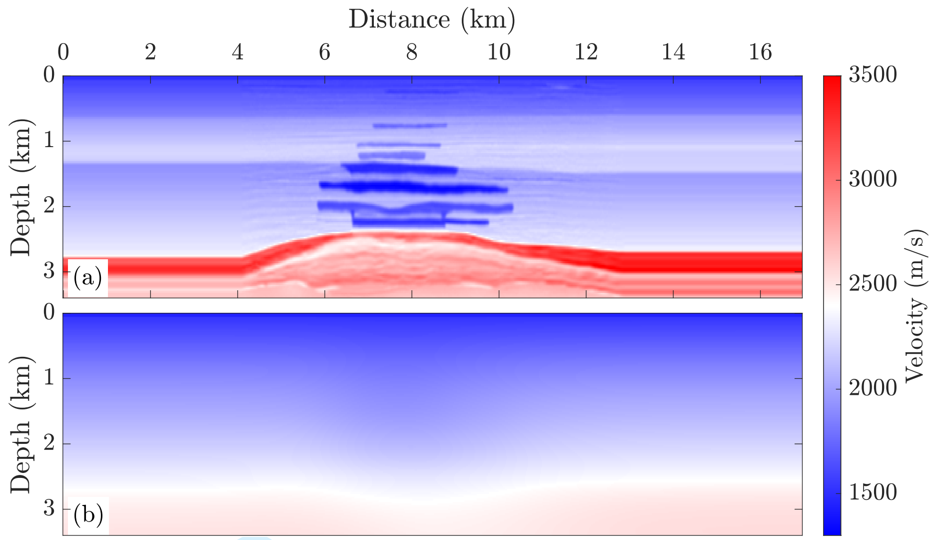

Figure 13: A North Sea exploration scale synthetic model. (a): true model used to generate the observed data, (b): the starting model.

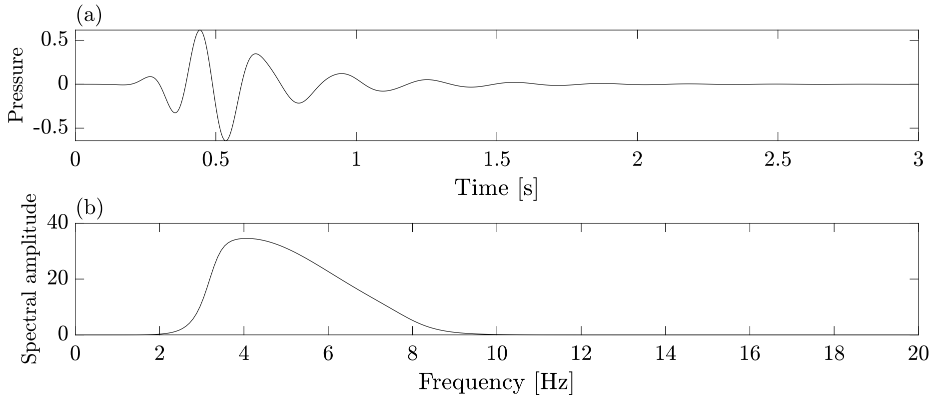

Figure 14: The source wavelet used for the numerical experiment. (a): time-domain (b): frequency domain.

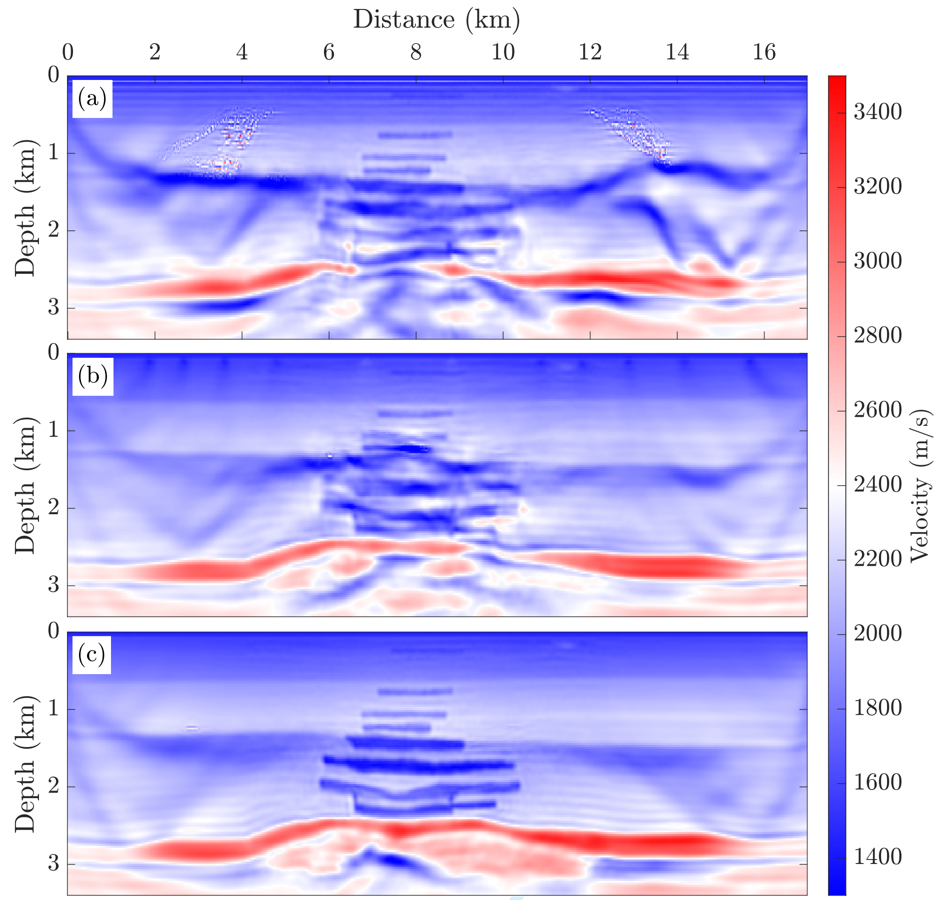

Figure 15: Reconstructed velocity models after 300 iterations. (a): conventional FWI, (b): static extended-receiver FWI and (c): time-dependent extended-receiver FWI.

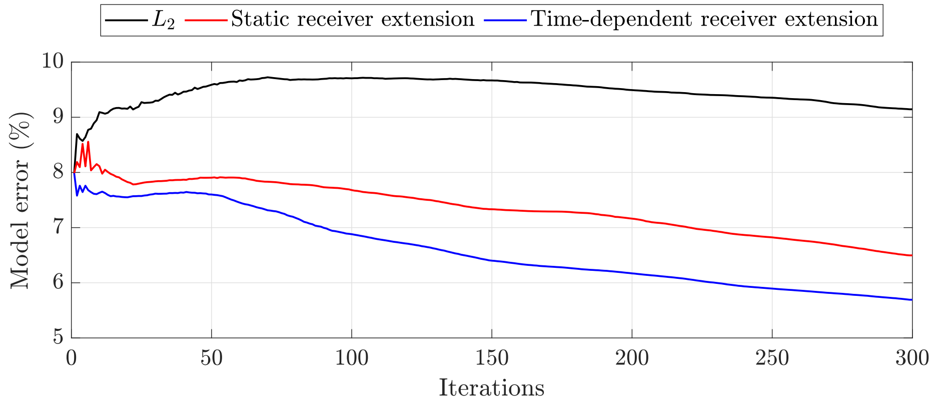

Figure 16: Receiver relocalization as a function of iterations,(a): static extended-receiver FWI, (b): time-dependent FWI

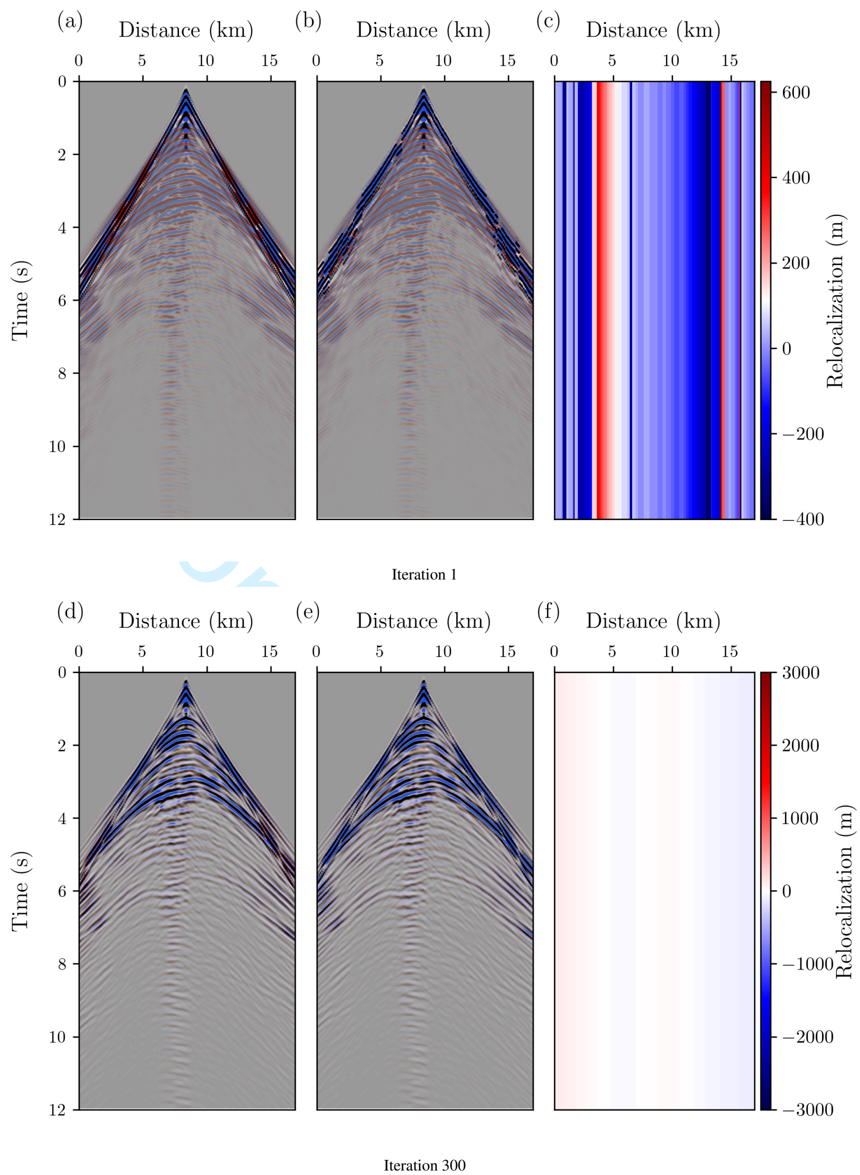

Figure 17: Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f).

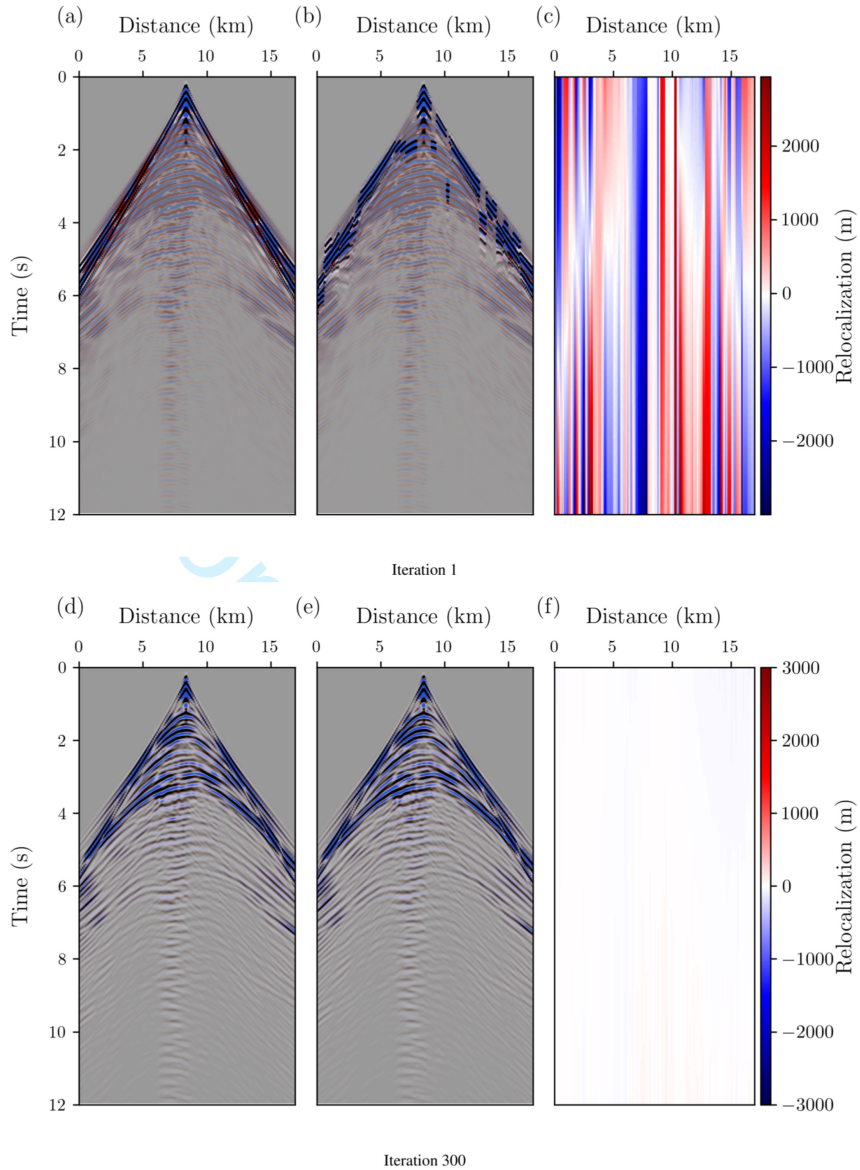

Figure 18: Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f).

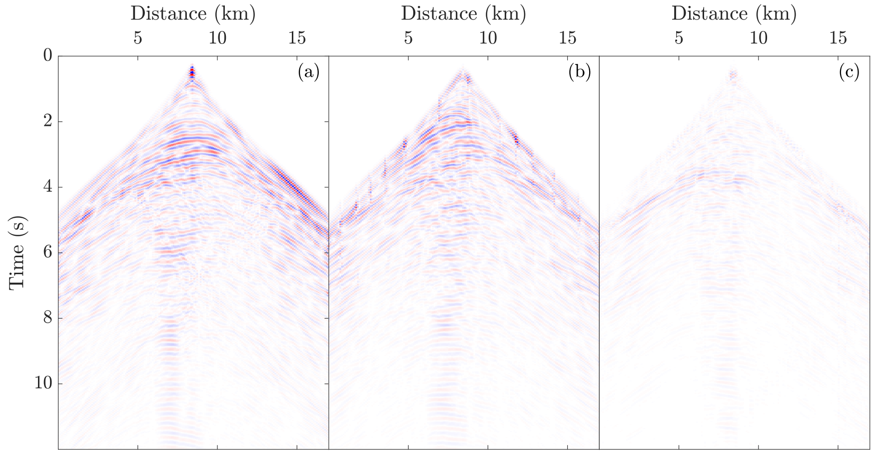

Figure 19: Adjoint source of the center shot point at the 300^{th} iteration for (a): conventional FWI, (b): static receiver extension and (c): time-dependent receiver extension. The same color scale is used for all figures.

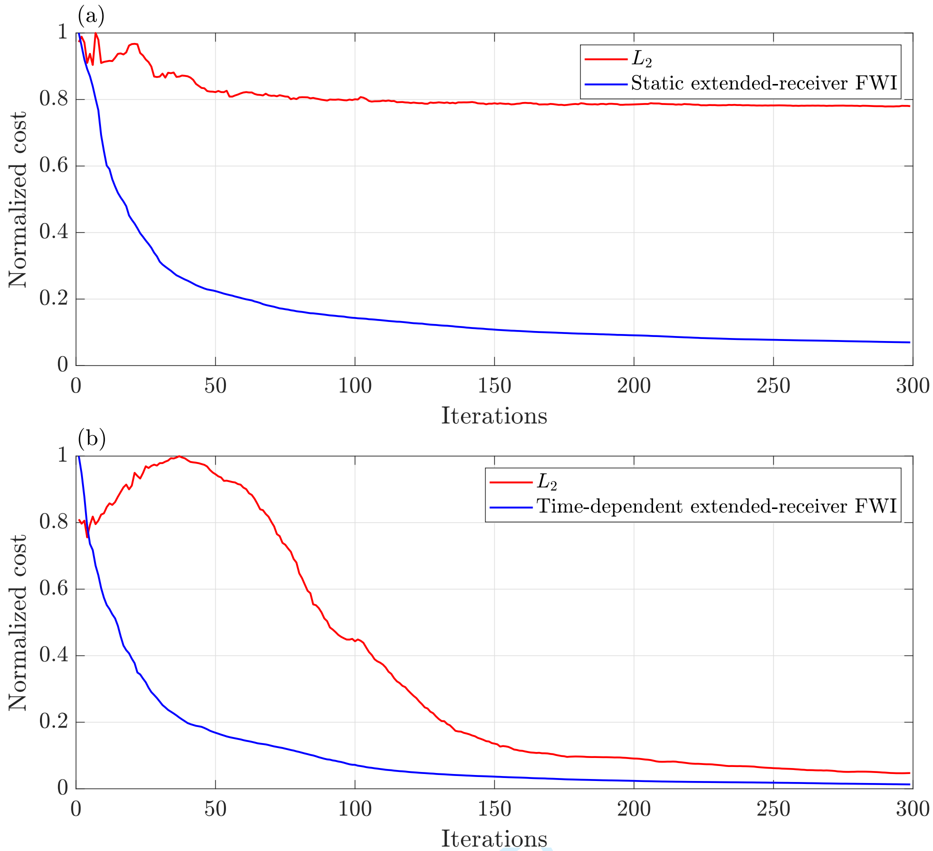

Figure 20: Cost function evolution as a function of iterations, extended-receiver FWI cost is plotted in black and the L_2 norm of the data residuals at each iteration is shown in red. (a): static extended-receiver FWI, (b): time-dependent extended-receiver FWI.

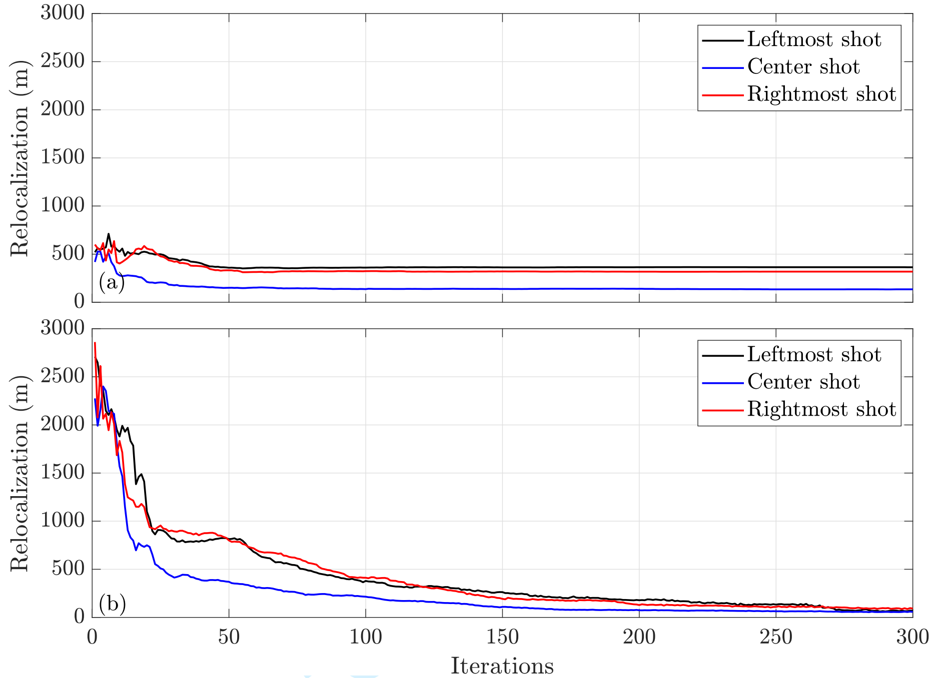

Figure 21: Receiver relocalization as a function of iterations,(a): static extended-receiver FWI, (b): time-dependent FWI.

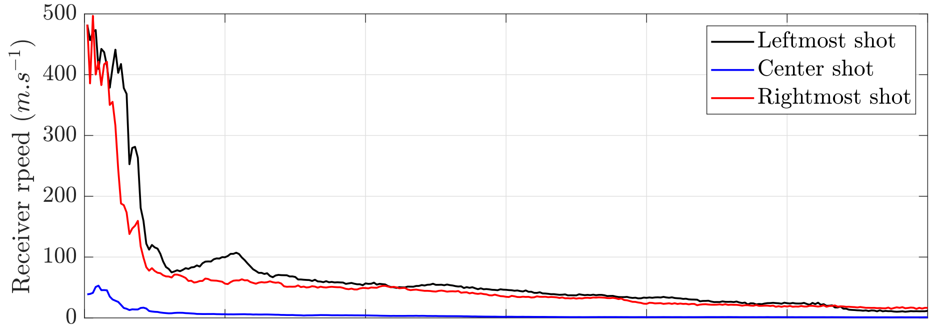

Figure 22: Receiver speed as a function of iterations for the time-dependent extended-receiver FWI.

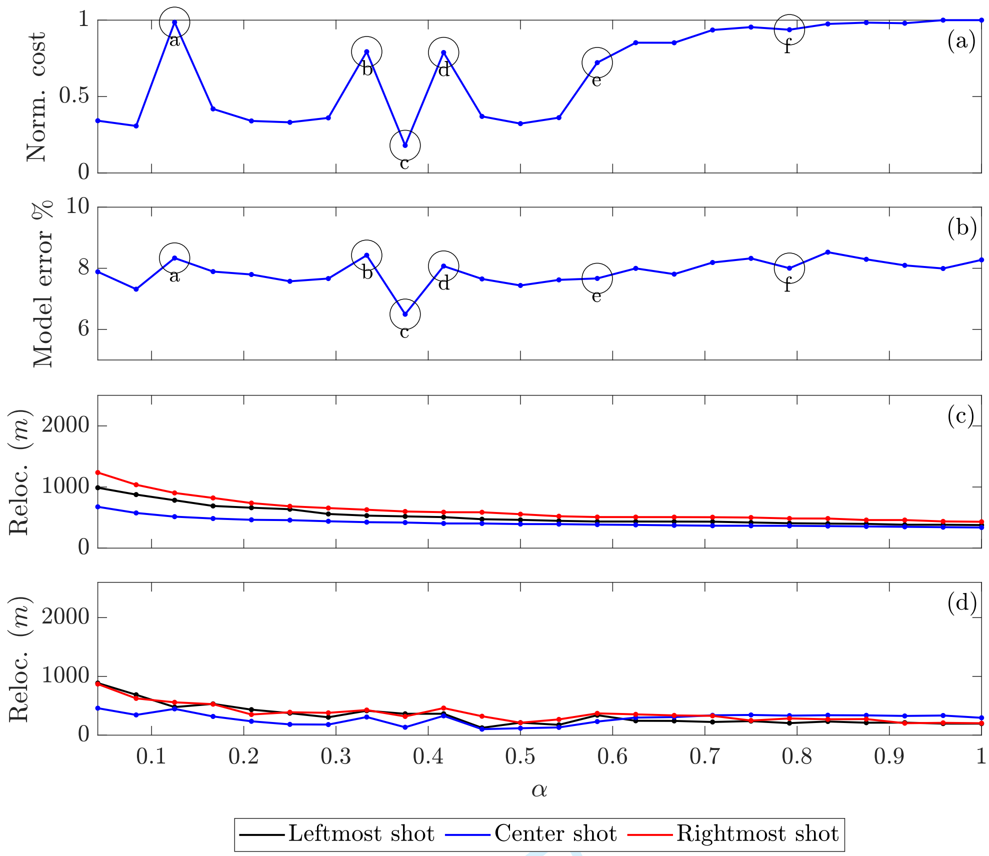

Figure 23: Tuning parameter α impact on extended-receiver FWI with static relocalization. (a): normalized cost at the last iteration, (b): reconstructed model error, (c): relocalization at the first iteration and (d): relocalization at the last iteration. The lower two panels show plots for the leftmost, center and rightmost shot points. The black labeled circles in (a) and (b) correspond to selected α values, whose corresponding final models are shown in Figure 24, where each model label correspond to an α value with the same label.

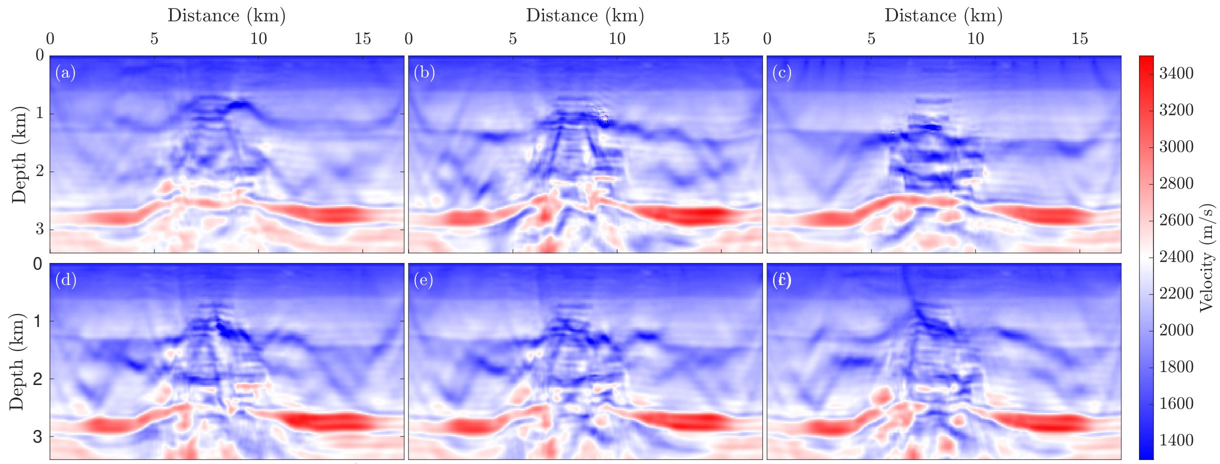

Figure 24: Tuning parameter α impact on the velocity reconstruction using extended-receiver FWI with static relocalization. (a): $\alpha = 0.125$, (b): $\alpha = 0.333$, (c): $\alpha = 0.375$, (d): $\alpha = 0.4167$, (e): $\alpha = 0.5833$ and (f) $\alpha = 0.7917$. Each sub-figure label corresponds to an α value indicated by black labeled circles in Figure 23a and 23b.

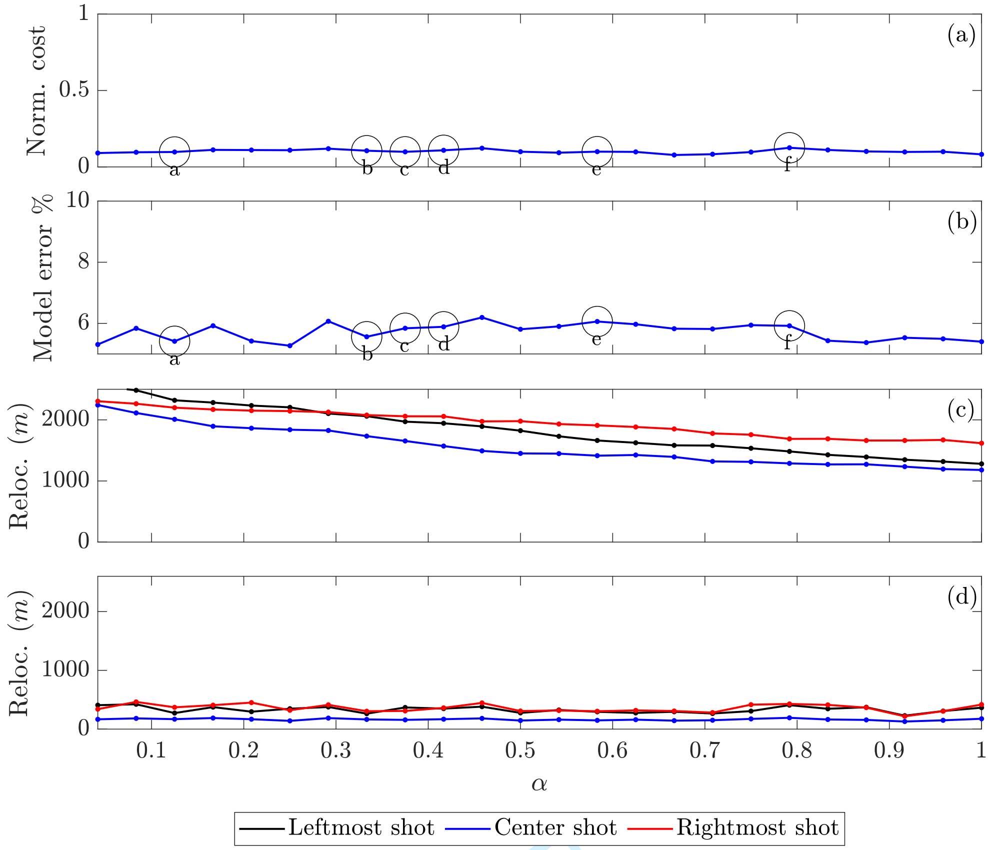

Figure 25: Tuning parameter α impact on extended-receiver FWI with time-dependent relocalization. (a): normalized cost at the last iteration, (b): reconstructed model error, (c): relocalization at the first iteration and (d): relocalization at the last iteration. The lower two panels show plots for the leftmost, center and rightmost shot points. The black labeled circles in (a) and (b) correspond to selected α values, whose corresponding final models are shown in Figure 26, where each model label correspond an α value with the same label.

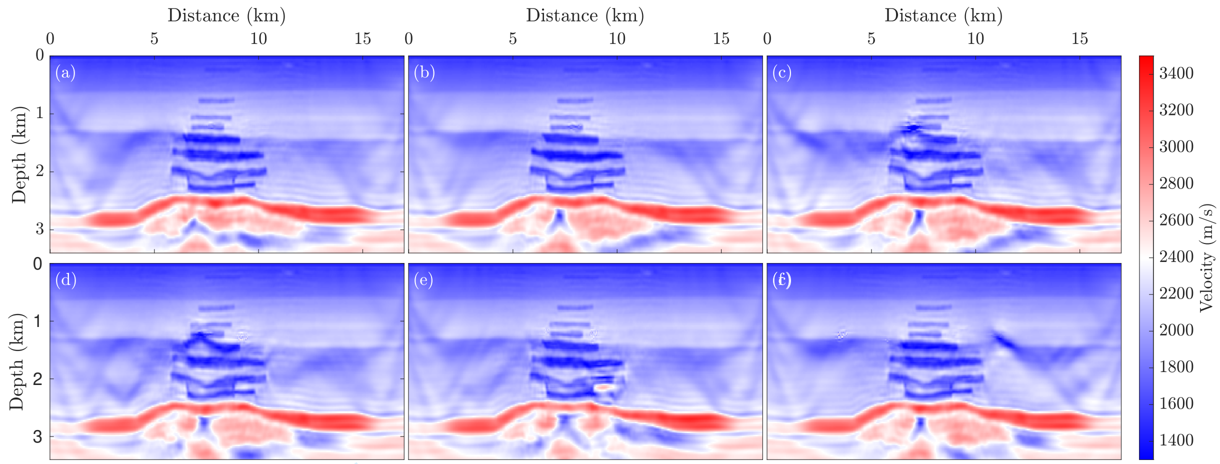

Figure 26: Tuning parameter α impact on the velocity reconstruction using extended-receiver FWI with time-dependent relocalization. (a): $\alpha = 0.125$, (b): $\alpha = 0.333$, (c): $\alpha = 0.375$, (d): $\alpha = 0.4167$, (e): $\alpha = 0.5833$ and (f) $\alpha = 0.7917$. Each sub-figure label corresponds to an α value indicated black labeled circles in Figure 23a and 23b.

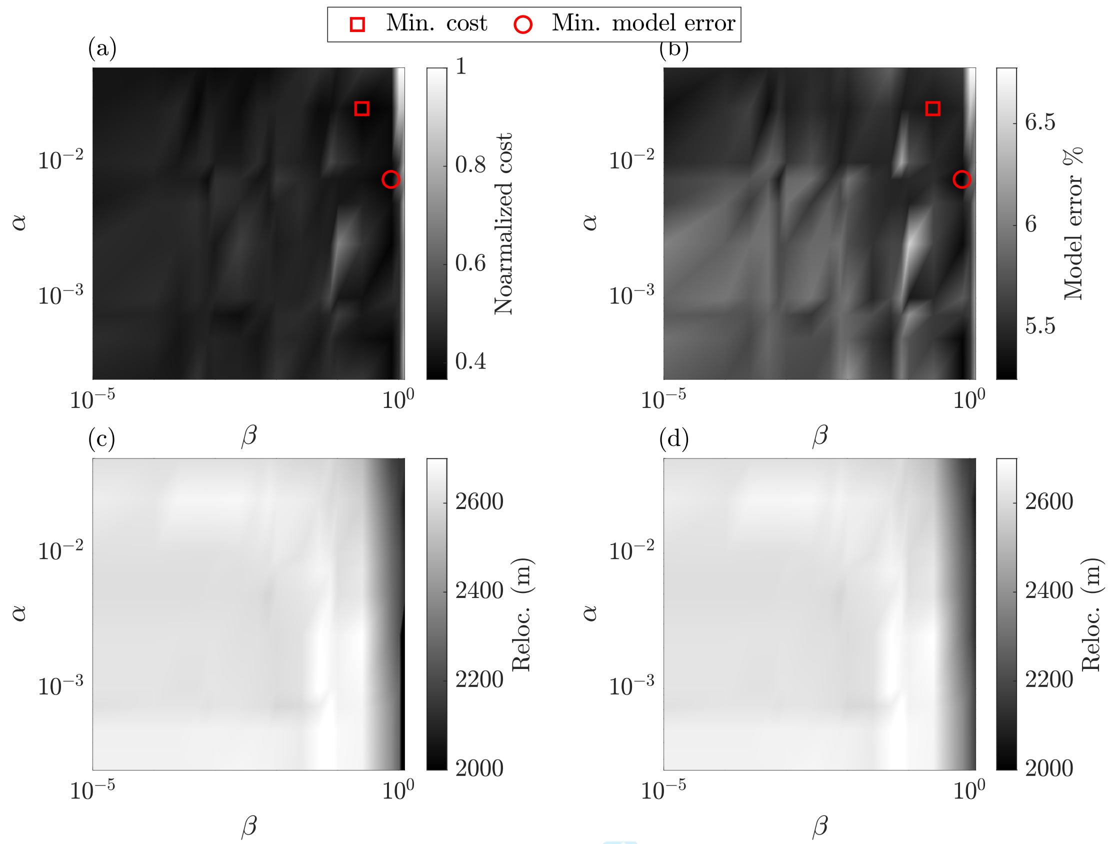

Figure 27: Tuning parameters impact on the (a): cost function at the last iteration, (b): model error and (c,d) receiver relocalization, shown here for (c) the leftmost and (d) the rightmost shot points.

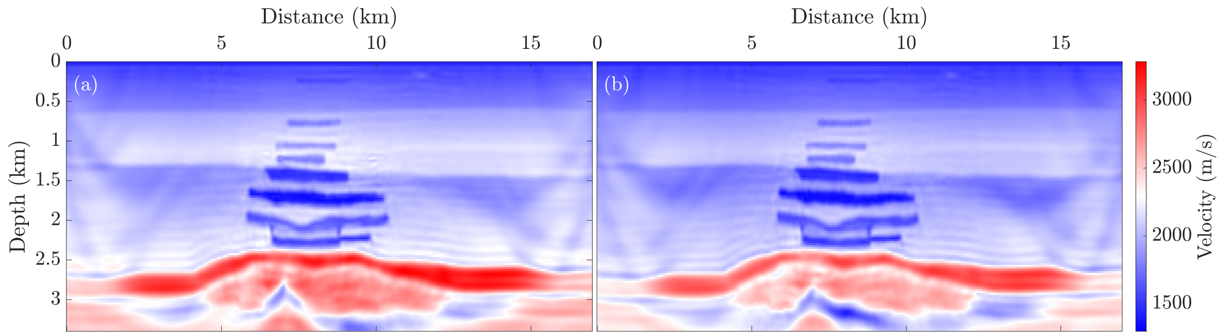

Figure 28: Reconstructed models corresponding to the (a): lowest cost and the (b): lowest model error (Figure 27).

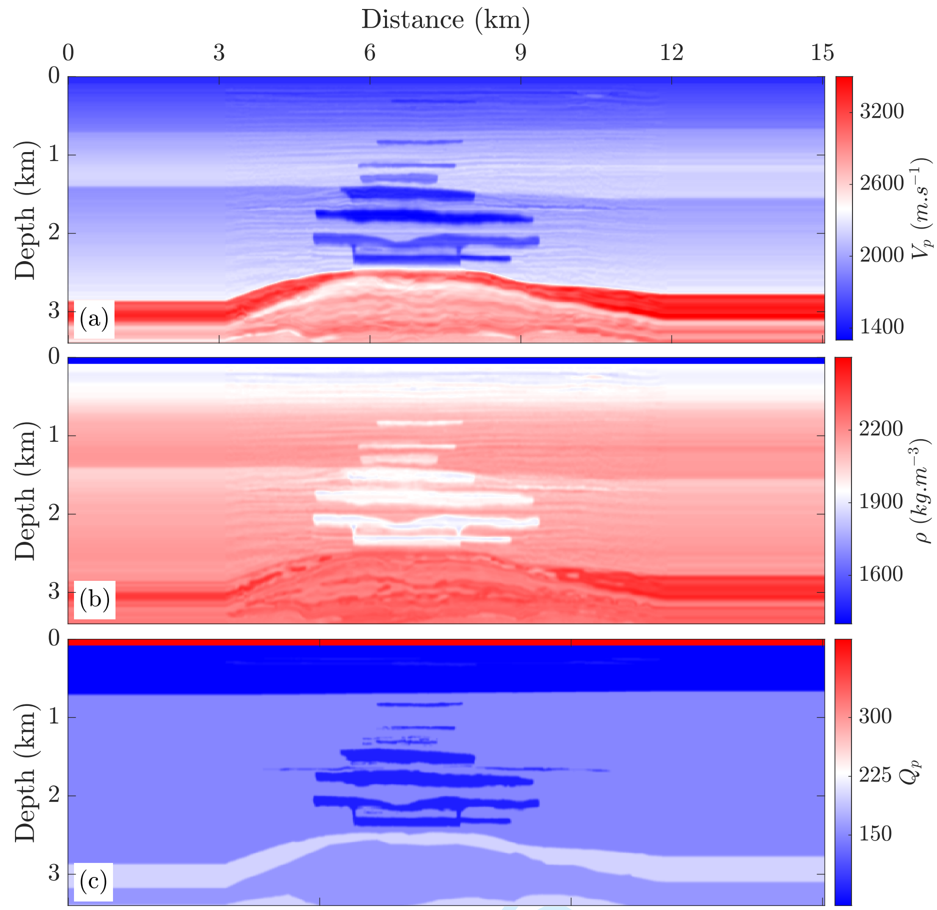

Figure 29: True models used to generate the observed data. (a): velocity, (b): density and (c): quality factor. The density and the quality factor in the water layer are 1000 kg.m^{-3} and 1000, respectively. The color-bars in (b) and (c) are clipped for clarity, therefore, the color in the water layer is not representative of the true values.

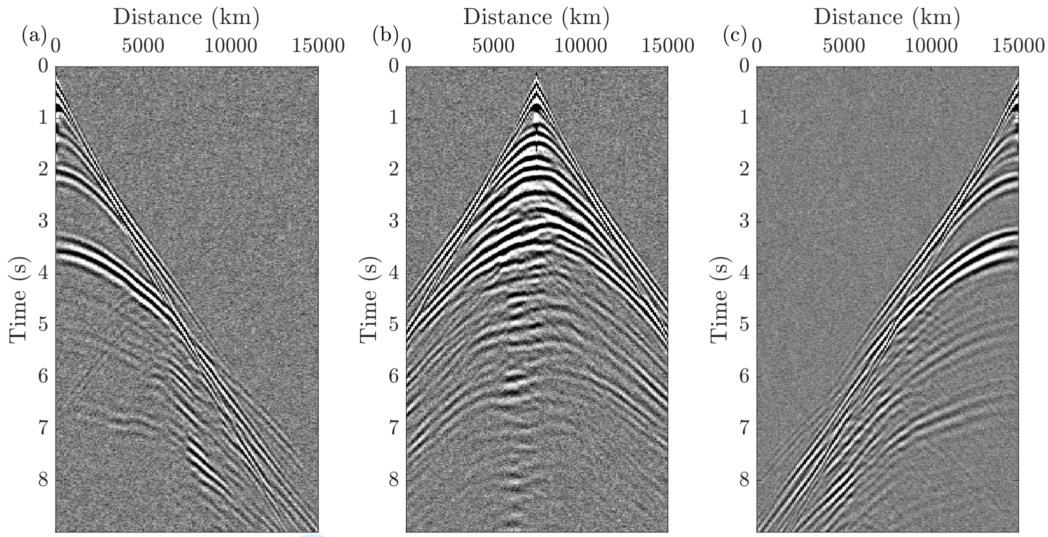

Figure 30: The observed data computed in a North Sea exploration scale synthetic model, (a) leftmost shot-gather, (b): center shot-gather and (c): rightmost shot-gather.

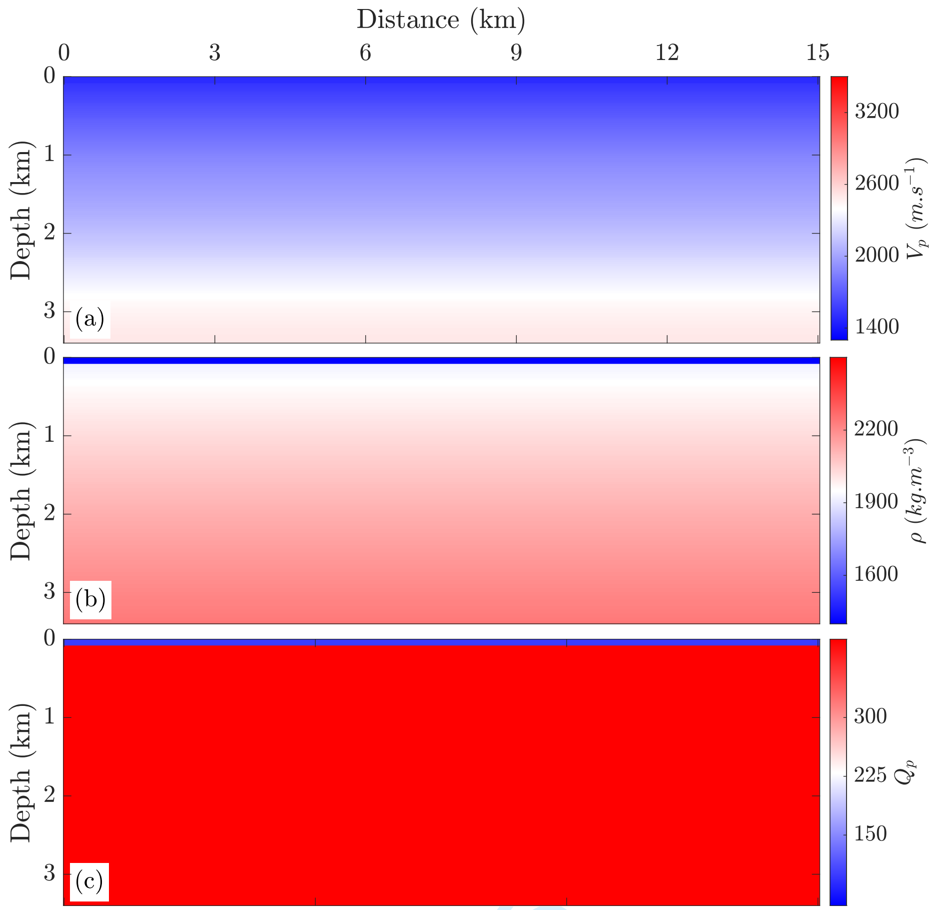

Figure 31: Starting models. (a): 1D velocity model, (b): density obtained using Gardner's law and (c): quality factor, set to 100 everywhere and 1000 in the water column. he colorbars in (b) and (c) are clipped for clarity, as a result, the color in the water layer is not representative of the true values.

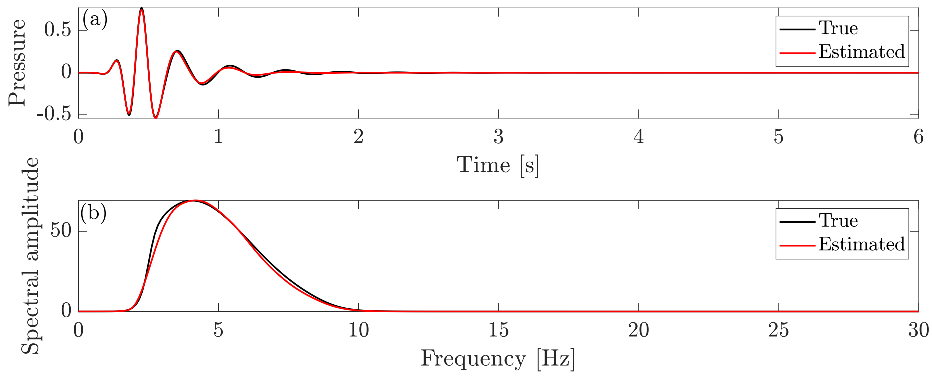

Figure 32: Source wavelet estimation in the starting velocity model. (a): source time function, (b): source frequency amplitude spectrum.

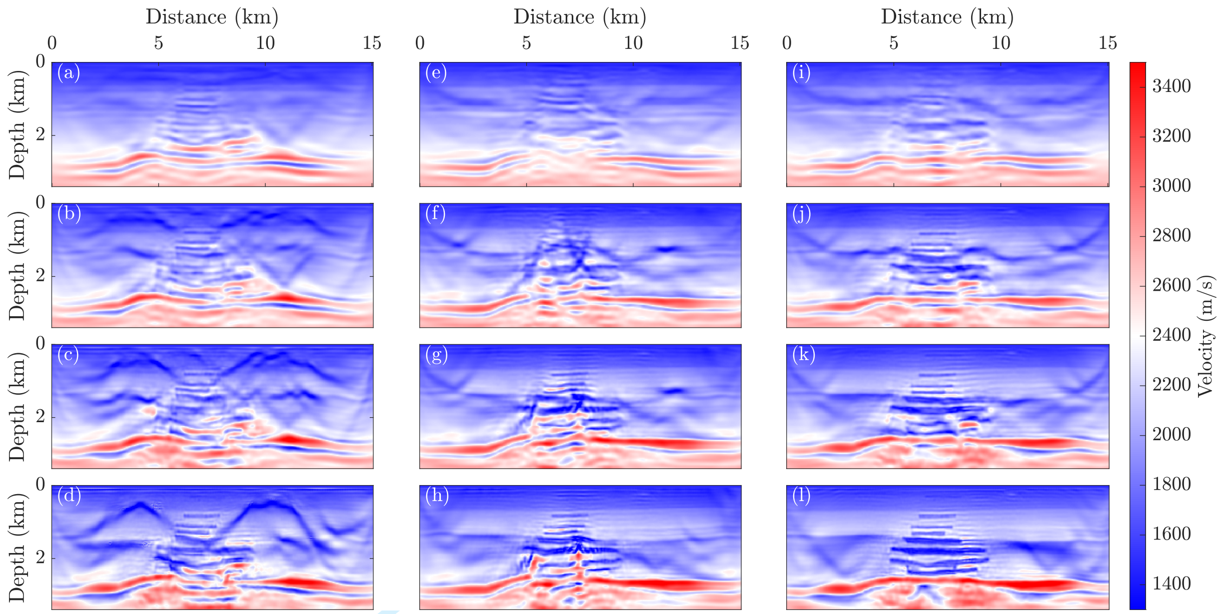

Figure 33: Reconstructed velocity models for (a-d) conventional FWI, (e-h) static receiver extension and (i-l): time-dependent receiver extension. This is shown for (a,e,i): 15 iterations, (b,f,j): 110 iterations (c,g,k): 205 iterations and (d,h,l) 300 iterations.

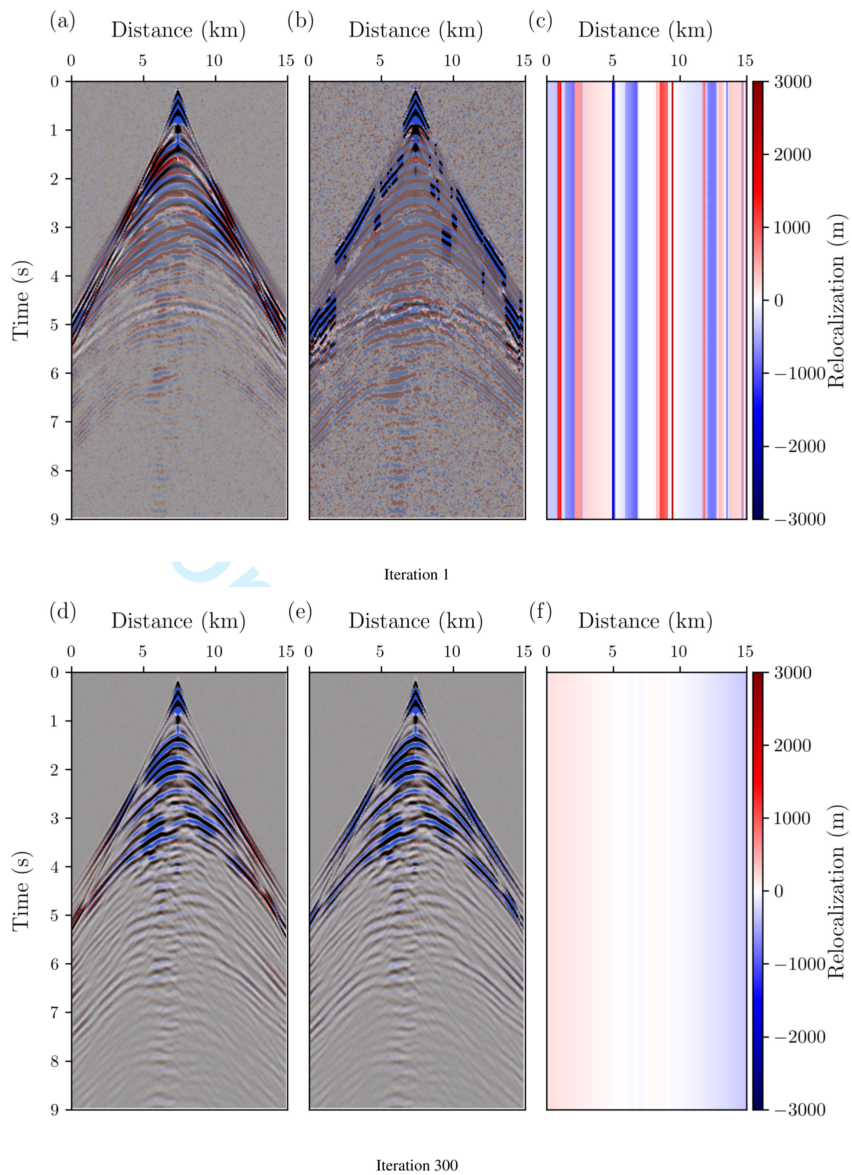

Figure 34: Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f).

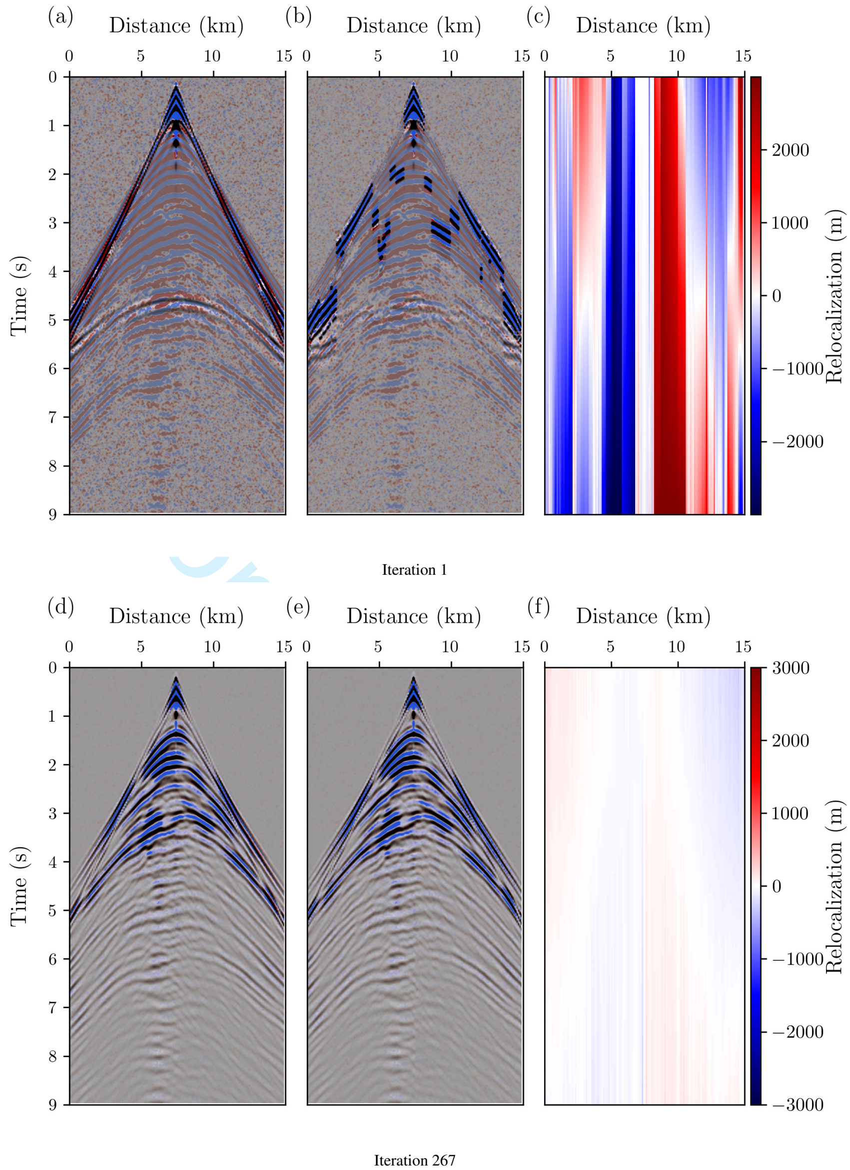

Figure 35: Synthetic and observed shot-gathers as well as the relocalization gathers for the static case. We display the calculated data in a red and blue color scale while the observed data is shown in gray-scale. If solely blue and black are apparent on the wiggles, a good fit is obtained, if the red shows than the fit is not satisfactory. (a,d): Data fit before relocalization, (b,e): data fit after relocalization and (c,f): relocalization gather, it has the same dimension as the shot gather. The first iteration is shown in the first row(a,b,c), and the last iteration is shown in the second (d,e,f).

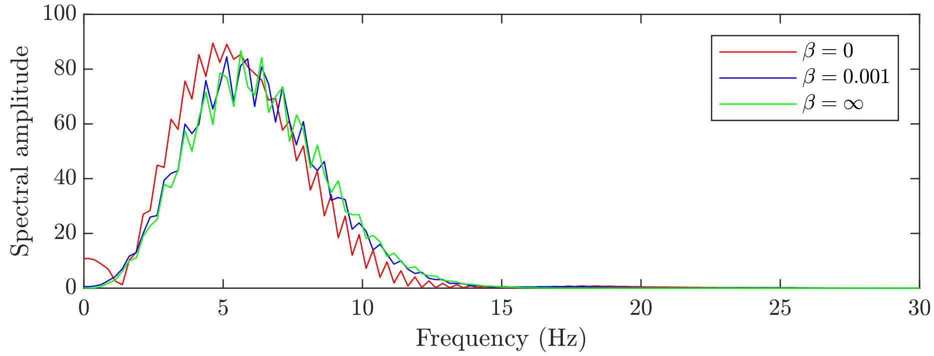

Figure 36: Calculated data amplitude spectrum for three cases: time-dependent relocalization without the second penalty term ($\beta = 0$, shown in red) and with the second penalty term ($\beta = 0.001$), as well as the static case ($\beta = \infty$). The notches in the spectra are caused by the free-surface reflection.

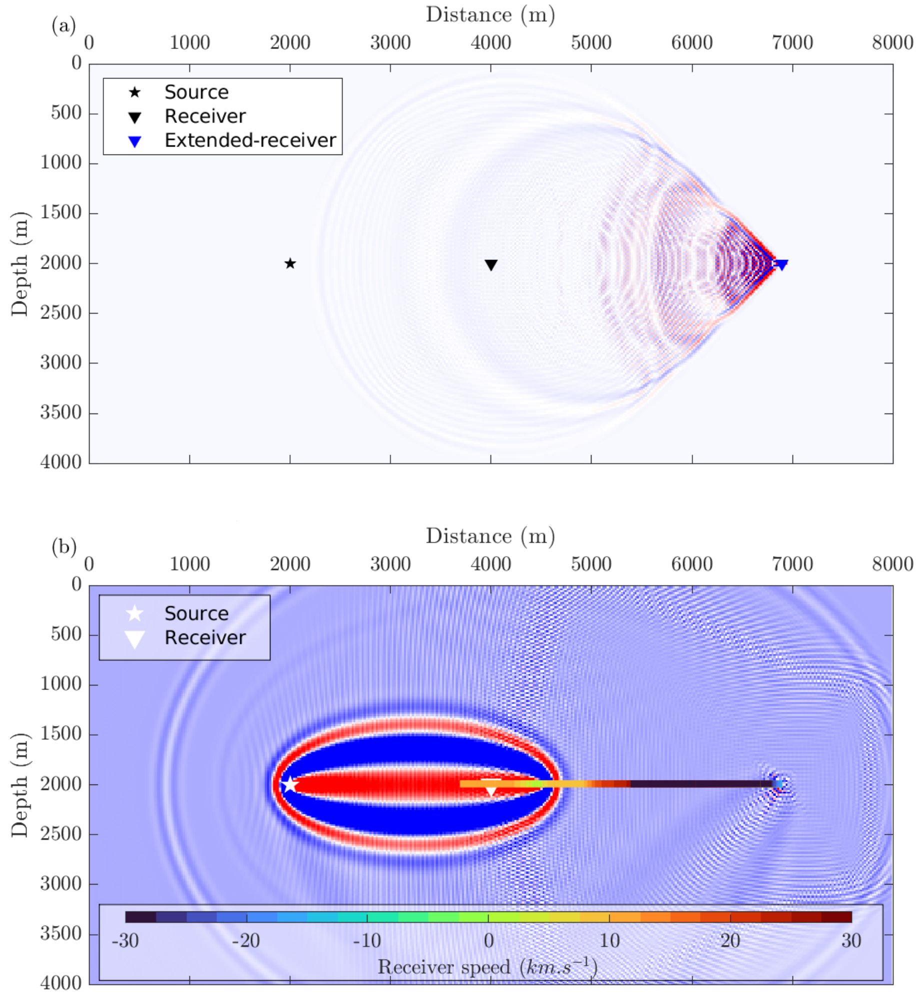

Figure 37: Gradient computation with $\beta = 0$. (a): adjoint field snapshot at $3.18\ s$, the extended receiver (adjoint source) is moving to the right causing the Mach cone shape in the wavefield, (b) extended receiver FWI kernel, the receiver positions are shown, the color indicates the receiver speed.

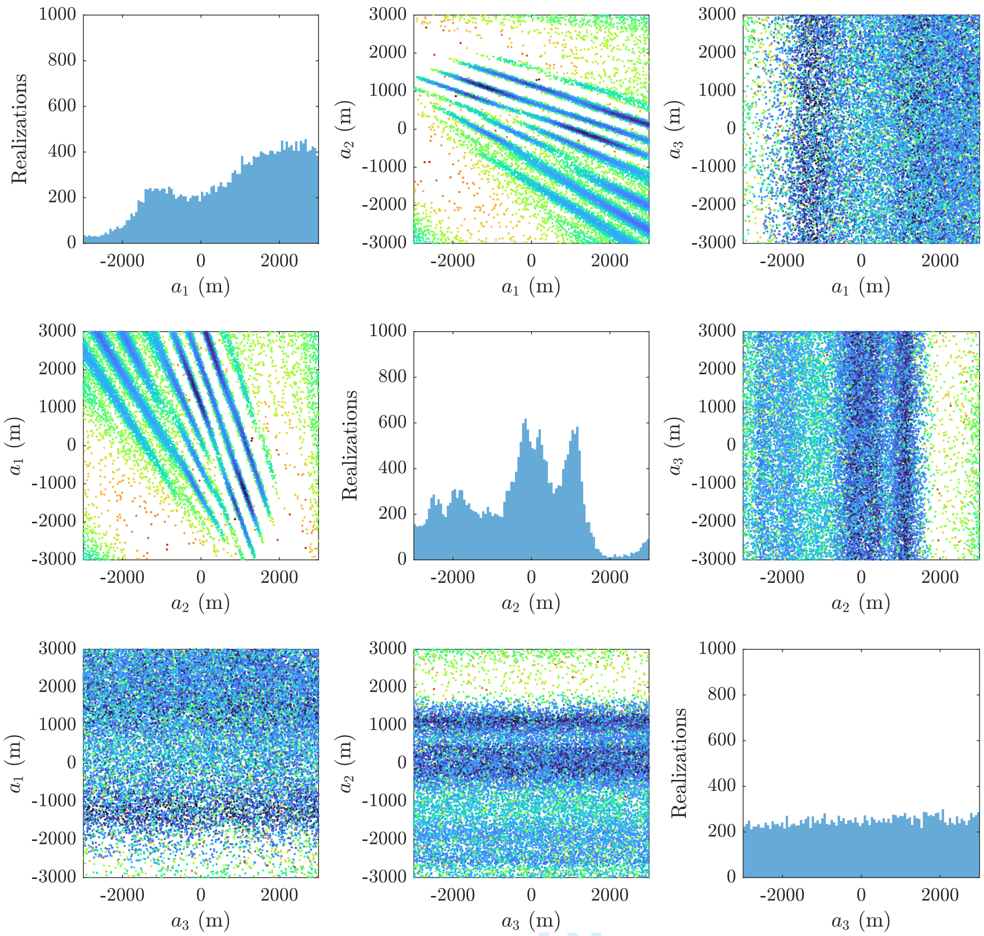

Figure 38: McMC sampling of the inner misfit function, samples of each degree of freedom are cross-plotted against another, the color map indicates the value of the cost function. The plots on the diagonal show the histograms of samples of each degree of freedom.

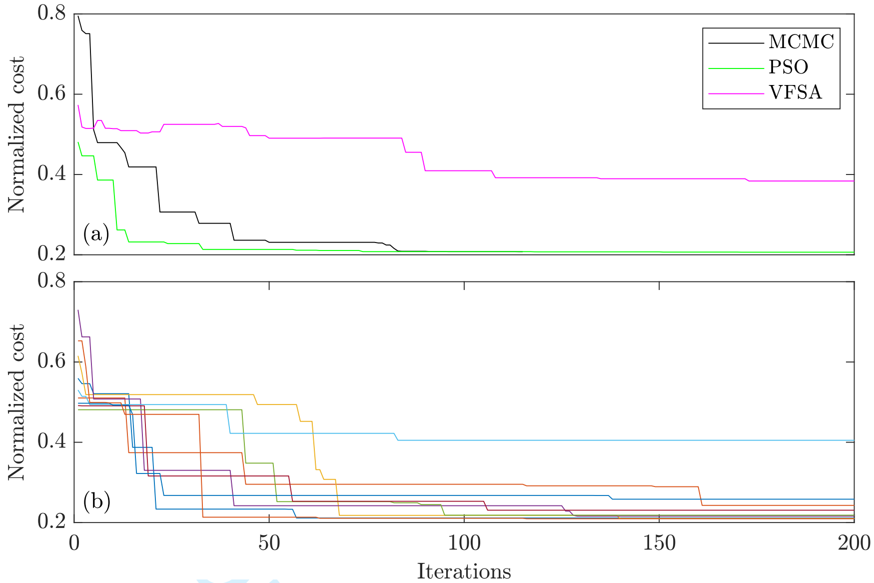

Figure 39: Cost function values as a function of inner-loop iterations for one receiver. (a): McMC candidates that lower the cost functions are shown in black, PSO global-best cost is shown in green, and Very Fast Simulated Annealing best solutions are shown in magenta. (b): the personal-best cost function of few particles as a function of iterations.

LIST OF TABLES

1 α and β values we use for the parametric study. The first row shows the values considered for both α and β, while the second row shows the additional values we test only for β. 104

2 CPU times in seconds for the different steps of the computation. 105

α,β	0.0003	0.0005	0.0008	0.001	0.002	0.005	0.0075	0.01	0.025	0.05		
β	0.0025	0.005	0.0075	0.01	0.025	0.0500	0.075	0.1	0.25	0.75	1.	1.25

Table 1: α and β values we use for the parametric study. The first row shows the values considered for both α and β, while the second row shows the additional values we test only for β.

	Forward	Inner loop	Inner loop per receiver	Adjoint simulation+ gradient summation
Conventional FWI	17.6	-	-	66.4
Static	21.9	2.1	1.4×10^{-2}	70.2
Time-dependent	22.1	53.7	0.3	74.6

Table 2: CPU times in seconds for the different steps of the computation.