GEOPHYSICS, VOL. 90, NO. 6 (NOVEMBER-DECEMBER 2025); P. R389–R412, 42 FIGS., 3 TABLES.
10.1190/GEO2025-0153.1

Time-dependent receiver extension for full-waveform inversion: An alternative extension method for cycle-skipping mitigation

Marc Benziane[1], Romain Brossier[1], Ludovic Métivier[2], and Serge Sambolian[3]

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 by relaxing the nonconvexity of the problem and thereby alleviating the cycle-skipping issue. The receiver-based extension strategy introduces the receiver position as the additional degree of freedom to FWI 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-loop 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 receiver positions. This inner-loop problem is heavily nonlinear and nonconvex. Finding the global minimum is therefore a challenging task. To do so, we use 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 very easy to tune.

INTRODUCTION

Full-waveform inversion (FWI) is a high-resolution seismic imaging technique that aims to reconstruct subsurface parameters using 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 data sets. 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 with the observed data in the least-squares sense. That is, the L^2 norm of the difference between the observed and synthetic data sets is referred to as the misfit. The model is then updated in a manner that reduces this misfit. However, the L^2 misfit function is nonconvex, meaning 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 sufficiently accurate, 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 data sets 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); 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 also can be done from narrower to wider offsets (Shipp and Singh, 2002; Brossier et al., 2009) because the shorter the offset, the less is the time shift. This can be done in combination with time windowing, considering first the earlier events and

Manuscript received by the Editor 3 March 2025; revised manuscript received 21 May 2025; published ahead of production 9 July 2025; published online 12 September 2025.
[1]ISTerre, Université Grenoble Alpes, Grenoble, France. E-mail: benzianm@univ-grenoble-alpes.fr (corresponding author); romain.brossier@univ-grenoble-alpes.fr.
[2]ISTerre, Université Grenoble Alpes, Grenoble, France and LJK, CNRS, Université Grenoble Alpes, Grenoble, France. E-mail: ludovic.metivier@univ-grenoble-alpes.fr.
[3]Université de Strasbourg, CNRS, Institut Terre et Environnement de Strasbourg, Strasbourg, France. E-mail: sambolian@unistra.fr.

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 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 nonconvexity. The strategies that fall into 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 (DTW) (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 nonconvexity 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 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 workflow represents the so-called migration velocity analysis (MVA) or Inversion Velocity Analysis (IVA) algorithms, where it is potentially possible to fit the observed data in an incorrect background model (Biondi and Sava, 1999; Symes, 2008; Mulder, 2014; Biondi and Almomin, 2014; Chauris and Cocher, 2017). These extension approaches are commonly solved using a nested optimization approach: 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. Recently, FWI with Model Extension (FWIME) has been introduced, it replaces the outer loop with nonlinear FWI iterations for the velocity model update (Barnier et al., 2023a, 2023b). In all these model extension methods, a penalty term (an annihilator) in the objective function relaxes the additional degrees of freedom by enforcing the focus of the final image. 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 subsurface offsets (the extended parameter) leads to a 3D problem in the 2D case (x, y, h), with h being the horizontal subsurface offset. In three dimensions, 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) introduce wavefield reconstruction inversion (WRI), where the search space is extended by optimizing over 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 performed in the frequency domain, where the wave equation operator can be factorized. This factorization allows the classical wave equation to be replaced by a single linear system. 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 are 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, 2019a), which they call iteratively refined WRI (IR-WRI).

Source extension methods are obtained by reparameterizing 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 us to fit the observed data to the synthetic data. Huang et al. (2018a) call this method matched source waveform inversion, and it is equivalent to WRI. Similar to WRI, source extension methods are performed in the frequency domain (Huang and Symes, 2015; Huang et al., 2018a, 2018b, 2019).

A time-domain implementation of source extension methods is initially introduced by Wang et al. (2017), which they achieve with a source extension approximation (their equations 5–7). Aghamiry et al. (2020) introduce 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 by the accelerated convergence, and the improved model parameters estimate accuracy. Gholami et al. (2022) propose a time-domain implementation of IR-WRI, but 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 is 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 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 also have noticed a slow convergence of the method. In this study, we build on the work of Métivier and Brossier (2022) by 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 also is 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 and the state-of-the-art extended-receiver FWI is 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

FWI 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

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

subject to

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

where $d_{\text{obs},s}$ is the observed data; $d_{\text{cal},s}$ is 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 parameter vector and b_s being the source term. Here, $R_{s,r}$ 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 (equations 1 and 2) 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 require their storage in memory, which is infeasible for large-scale problems. In practice, the following reduced space approach is used:

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

Equation 4 is an unconstrained optimization problem, where the wavefield u_s has been eliminated from the optimization variables, which exacerbates the nonlinearity with respect to the model parameters. Equation 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 $\langle \bullet, \bullet \rangle$ indicates an inner-product operation. In time-domain FWI, it corresponds to the zero-lag crosscorrelation between the weighted incident wavefield and the adjoint wavefield. The latter is computed by adjoint propagation of the adjoint source, which is the data residuals for the L^2 misfit function case (difference between the observed and calculated data). This computation is achieved as follows:

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

where λ_s is the adjoint field, computed backward in time using the wave equation operator $A^T(m)$. The source term is the data residuals μ_s injected at the receiver positions by the operator $R_{s,r}^T$. After the gradient is obtained, the model can be updated. In all the examples shown in this paper, we use the preconditioned limited memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) (Nocedal, 1980) algorithm. Therefore, the new model m_{k+1} is obtained by updating the model at iteration k as follows:

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

where k is the iteration number, α_k is a step length obtained with a line search 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 us 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 incorrect medium velocity. We write the receiver extension misfit function as

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

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 on the right-hand side

is the data fit term, where $\tilde{d}_{\mathrm{cal},s}$ is 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. Thus, the calculated data are obtained as follows:

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

where $\tilde{R}_{s,r}$ represents 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 $\mathbf{x}_r + \Delta x_r$, with $\mathbf{x}_r$ 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 to prevent the relocalization from being too large. The expression of $\mathcal{P}_{1,s}$ is provided in Appendix A.

We illustrate the leading idea behind the receiver extension with a simple numerical experiment. In a homogeneous medium ($v_{\mathrm{true}} = 2000$ ms^{-1}), we consider one source-receiver couple. The observed trace is shown by a dashed black line in Figure 1b. The synthetic trace is computed in a medium with a higher velocity (2300 ms^{-1}), and it is shown by a solid red 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 different receiver relocalizations. This allows us to visualize the objective function of equation 8 for each velocity and each receiver position. This is shown in Figure 2a, in which 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.

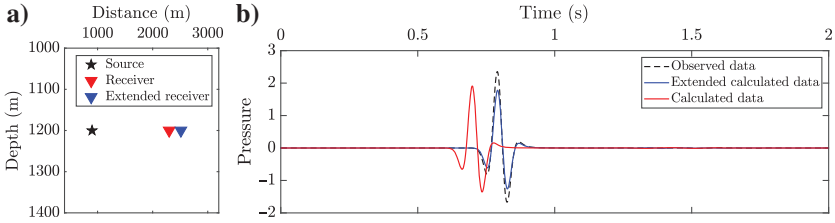

Figure 1. Receiver extension illustration. (a) Acquisition setup. The source is shown as a black star, the receiver position is shown as a red triangle, and the extended receiver is shown as a blue triangle. (b) Observed trace (dashed black line), calculated trace (solid red line), and calculated trace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

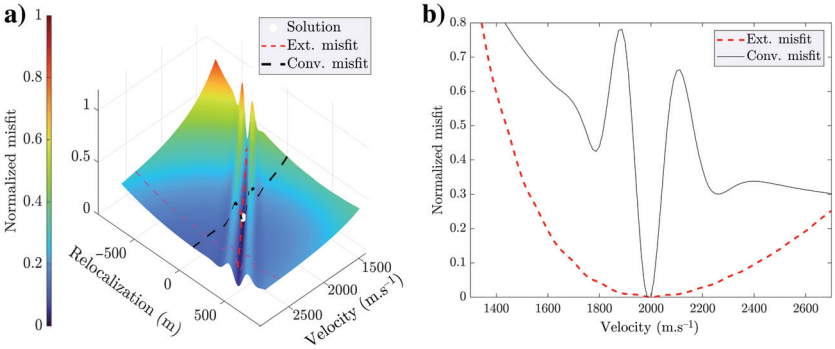

Figure 2. Illustration of the receiver extension misfit function. (a) Misfit function 2D view as a function of velocity and receiver relocalization. The minimum is shown in white, and the misfit values corresponding to the receiver extension are shown as a dashed red line. The dashed magenta line indicates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

Computing a numerical solution

Bivariate misfit functions, such as equation 8, are usually minimized using nested-loop strategy. The outer loop is the conventional FWI optimization over the model parameter m. The inner loop solves the subproblem 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 parameter 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), whereas 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 as follows:

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

According to 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 as follows:

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

where the adjoint field λ_s is obtained using

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

This means that the calculated data are now extracted at the extended receiver position according to equation 9, and the adjoint source position also corresponds to the extended receiver. 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, and 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, and the extended data are 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.

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 is performed in the framework of a single arrival. They also observe that the method suffers from a slow convergence rate. We aim in the present study to address these two main issues. Our first investigation step is to introduce more degrees of freedom to allow the receiver to move as a function of the acquisition time because 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, with transmitted and a reflected arrival. The observed data are simulated in a two-layer 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-layer starting model, where the first layer velocity is faster at 2500 m.s^{-1}, but the second layer velocity is the true one ($V_P = 3500$ m.s^{-1}). In Figure 4, the observed trace is shown by a dashed black line, the calculated trace is shown by a solid blue line, and the receiver relocalization as a function of time is shown in red. This curve indicates the receiver relocalization value at each time step, and we refer to this curve as the relocalization profile. Using receiver extension, only one arrival can be fitted. This occurs because the second (reflected) arrival would require another relocalization value 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. Figure 5 shows the FWI kernels for conventional FWI, static, and time-dependent receiver extension. The kernel for the

Algorithm 1. Nested-loop optimization.

while $\hat{f}(m_k) > \epsilon_{\text{outer}}$ **do**
  $u_s \leftarrow \text{do}_{\text{forward}}\text{modeling}(m)$
  while $f(m_k, \Delta x) > \epsilon_{\text{inner}}$ **do**
    $f(m_k, \Delta x) \leftarrow \text{compute}_{\text{inner}}\text{cost}(u_s)$
  end while
  $\overline{\Delta x} \leftarrow \Delta x$
  $\hat{f}(m_k) \leftarrow f(m_k, \overline{\Delta x})$
  $m_k \leftarrow \text{update}_{\text{model}}(m_{k-1}, \overline{\Delta x}, u_s)$
  $k \leftarrow k + 1$
end while

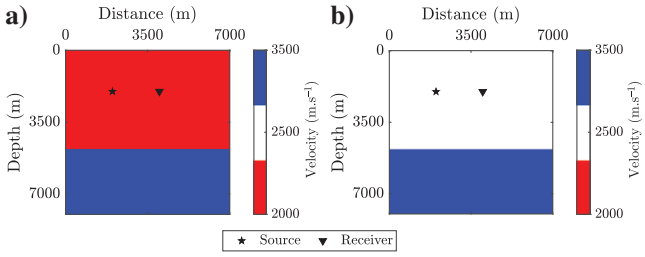

Figure 3. Experimental setup: (a) true velocity model and (b) the starting model used for gradient illustration hereafter.

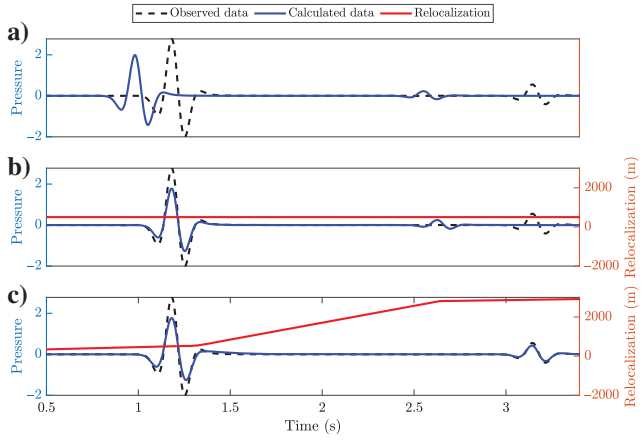

Figure 4. Observed and calculated data as well as relocalization time profiles for (a) conventional FWI, (b) static relocalization, and (c) time-dependent relocalization.

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 incorrect 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 due to the fit of the second arrival. In this particular case, where the starting velocity is higher than the true velocity, receiver extension virtually increases the illumination angle. This causes an update beyond the first Fresnel zone corresponding to the true receiver position (the observed raypath). In contrast, if the starting velocity is lower than the true velocity, receiver extension causes a narrower update (by reducing the offset). This is caused by the additional freedom that we introduce, which is relaxed along FWI iterations. As the model improves, the extended receiver positions become closer to the true receiver positions. The method provides the sought tomographic update by means of this intermediate step (the extension) when the model estimate is poor, where the gradient is evidently different from the conventional FWI gradient. This is why the method is able to mitigate cycle skipping. This time-dependent receiver extension strategy is detailed in the next section.

TIME-DEPENDENT RECEIVER EXTENSION FOR FWI

Formalism

We write the new extended-receiver FWI misfit function as

$$\min_{m,\Delta x(t)} \tilde{f}(m,\Delta x(t)) = \min_{m,\Delta x(t)} \frac{1}{2}\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\int_0^T |\tilde{d}_{\mathrm{cal},s}[m](\mathbf{x}_r+\Delta x_r(t),t)-d_{\mathrm{obs},s}(\mathbf{x}_r,t)|^2 dt + \alpha\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mathcal{P}_{1,s}[\Delta x_r(t)]+\beta\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mathcal{P}_{2,s}[\Delta \dot{x}_r(t)]. \quad (19)$$

We aim to minimize the bivariate misfit function in equation 19, where m denotes 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 as

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

where $\Delta x_r(t)$ is the time-dependent receiver relocalization. The calculated data are extracted at the virtual receiver 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 side of equation 19 penalizes the receiver relocalization 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 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. Interested readers are referred to Appendix B. The full expressions of $\mathcal{P}_{1,s}$ and $\mathcal{P}_{2,s}$ are shown in Appendix A. The adjoint field is obtained using the adjoint system:

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

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

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

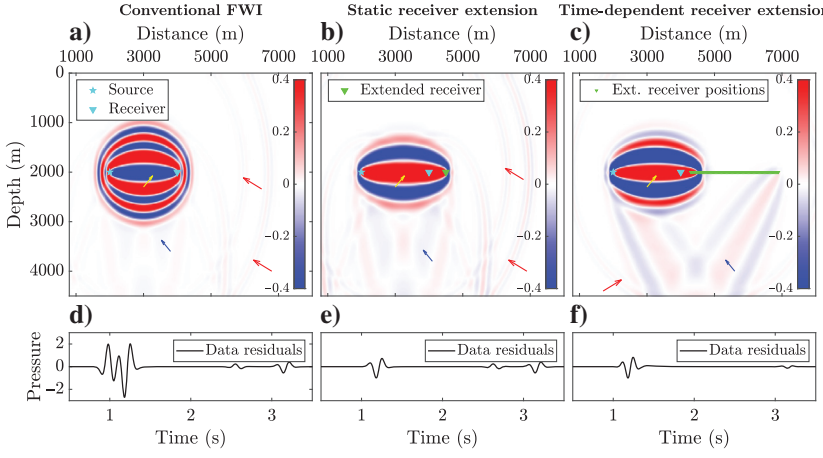

Figure 5. The FWI kernels and adjoint sources, respectively, for (a and d) conventional FWI, (b and e) static relocalization, and (c and f) time-dependent relocalization. The yellow arrow indicates the first Fresnel zone, the blue indicates the rabbit ears, and the red arrows show the migration isochrones.

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

Parameterization

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

Temporal parameterization

One possible choice of parameterization 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 will be used to solve the inner-loop problem. To keep a minimal parameterization, 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 $\ell_k^{N_\ell}(t)$ is Lagrange basis functions of order N_ℓ, n_s is the number of segments, and a_i $(i = k + N_\ell \times (j-1))$ is 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 receiver positions. In Figure 6, we illustrate this using three segments with first-order Lagrange polynomials. The time-dependent relocalization is shown by a 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-loop approach described in Algorithm 1.

Spatial parameterization

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 only 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 that allow fitting transmitted and reflected arrivals. We consider a one layer over a half-space model (Figure 7a). The traveltime 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 an incorrect 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 traveltime 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 traveltime for the transmitted arrival in an incorrect velocity model, the receiver is relocated as

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

Equating equations 28 and 29 gives

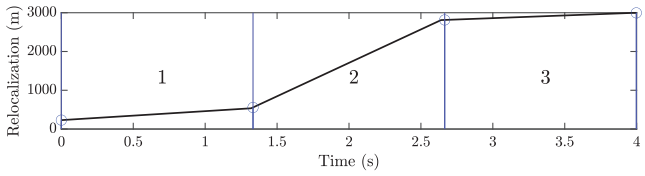

Figure 6. Time-dependent parameterization example using three segments with first-order Lagrange polynomials. The circles indicate the control point a_i.

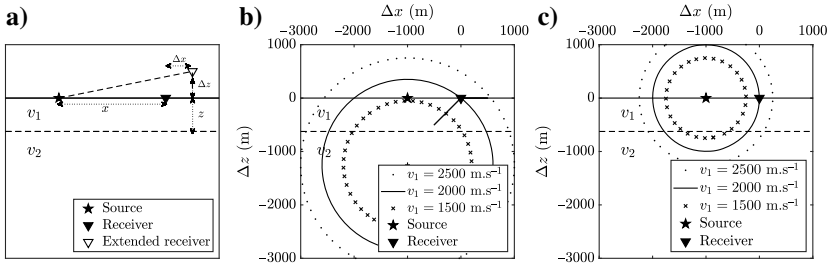

Figure 7. Simple reflection case analysis with receiver extension. The geometry is shown in (a), where the original and extended receiver positions are shown as black and white inverted triangles, respectively. Here, Δx and Δz are the horizontal and vertical receiver relocalizations, respectively. The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

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

Equations 27 and 30 are conic section equations. For the reflection case, equation 27 describes a circle with a center $(-x, -2z)$, which is the source image with respect to the reflector. The radius of this circle is $\sqrt{(4z^2 + x^2)(v_1^2/v_0^2)}$. 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 traveltime 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 "×" 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 toward the source following a $(\pi/4)$ angle with respect to the horizontal axis, a fit of the reflected and transmitted arrivals can be obtained. This is achieved by equating Δx to Δz when the receiver moves toward the source. Therefore, Δz can be obtained from Δx using

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

Please note that the aforementioned angle is a user-defined parameter because in some cases, it is not advisable to allow the receivers to move vertically in the medium, where physical parameters vary rapidly and the waveform amplitudes are very different in the medium (e.g., below the seafloor). In this case, we prefer to constrain the receiver relocalization to the water column. However, from an implementation standpoint, the receivers can in fact be repositioned anywhere in the computational domain to extract wavefield values. It is a choice of parameterization that we make; it is not a limitation of the algorithm. This is the choice of parameterization that we make to keep the parameterization minimal. Then, allowing the receiver to move freely in all the spatial dimensions would give rise to an expensive inner-loop computation. This is avoided because of our spatial parameterization.

We illustrate this parameterization with a numerical experiment using the same setup as shown 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 toward 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 by a dashed black line and the extended calculated data are shown by a solid blue line. The receiver needs to move following the z-axis 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.

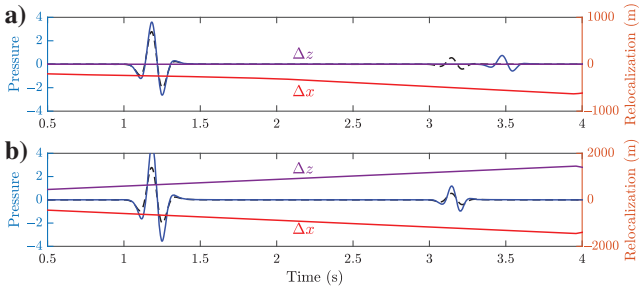

Figure 8. Data fit and relocalization profiles for a two-arrivals case using a lower top layer starting velocity. (a) Time-dependent horizontal relocalization and (b) time-dependent relocalization with vertical and horizontal components. The observed trace is shown as a dashed black line, the calculated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

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 does our misfit function contain numerous minima, but some local minima also 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 because the size of this inverse problem will grow large with the time dependence. Even with the parameterization used for the present example (two control points), it is still intractable because 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 converge to a good solution, and the latter suffers from premature convergence toward bad solutions. In contrast, PSO allows for a thorough exploration of the

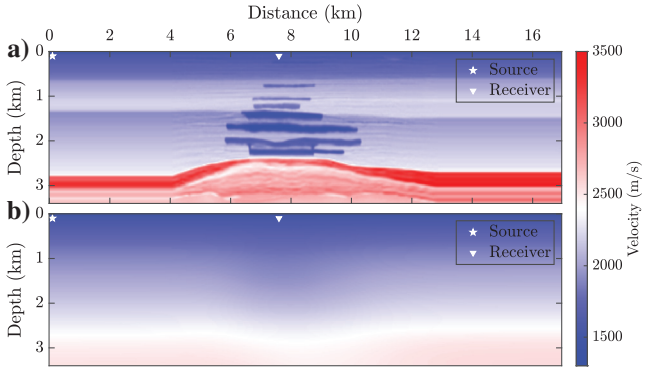

Figure 9. A North Sea exploration-scale synthetic model: (a) the true model and (b) the starting model obtained by Gaussian smoothing of the true model.

search space with reasonable cost. Interested readers are referred to the discussion part of the paper.

PSO

PSO 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 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 following equations). In contrast, the personal best ($\mathbf{x}_{p,j}$ in the equations hereafter) is the best solution obtained for each individual particle. Let $\mathbf{x}_j^i$ denote 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

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

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, which 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 the random variable vectors of the same dimension as the particle position $\mathbf{x}_j^i$, which are harvested from a uniform distribution. Here, 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

To showcase PSO, we carry out a simple numerical test using the same example we showed previously (Figure 9). We run PSO using 16 particles. The snapshots of the swarm configuration at iterations 1, 40, 80, and 160 are shown 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 by 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 particle positions is near the global minimum; nonetheless, PSO manages to converge relatively quickly. Figure 10 clearly shows that our misfit function contains 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 at the global minimum, other particles still search in the vicinity of other secondary minima because of the inertia term (equation 34). PSO can achieve convergence relatively quickly; however, setting the maximum number of 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 because we assume that 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 $n_{\text{stall}} = 200$. Here, 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 275th iteration. However, the optimization does not stop in this case because the global best needs to stall for 200 iterations. Next, we increase the number of particles to $N_p = 32$ to observe the impact of the swarm size on the convergence (Figure 12). Indeed, increasing the swarm size leads to a sooner stalling and therefore a better convergence. With 32 particles (shown in blue), stalling occurs at 127 iterations, and PSO has stopped at 327 iterations. When using 64 particles (shown in red), the global best does not change after 34 iterations, and PSO is 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 approximately 30 particles is common. A choice of too few

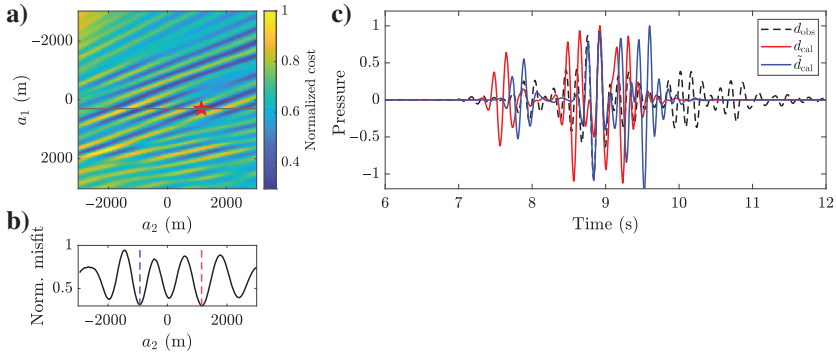

Figure 10. Inner-loop illustration: (a) misfit function map using one segment and two control points, where the global minimum is shown as a red star. (b) A cross section following the black line in (a). Note how close are the values of the global minimum (dashed red vertical line) and the secondary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

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, swarm size is adapted based on the inner-loop parameterization. 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 $\omega = -0.6485$.

APPLICATION TO A NORTH SEA EXPLORATION SCALE SYNTHETIC 2D MODEL

For all our testing, we use the SEISCOPE (acoustic and viscoacoustic) 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 (CPMLs) (Komatitsch and Martin, 2007) as absorbing boundary conditions to simulate an infinite medium. However, we note 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 only 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 use checkpointing-assisted reverse-forward simulation (CARFS) (Yang et al., 2016b). We note that the synthetic data extraction and adjoint sources injection also are performed using Kaiser-windowed sinc interpolation (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 and a sensitivity study. Second, we design a more realistic experimental setup using the same model. Please note that in all of our testing, we perform single component inversions, namely, the hydrophone (pressure) component.

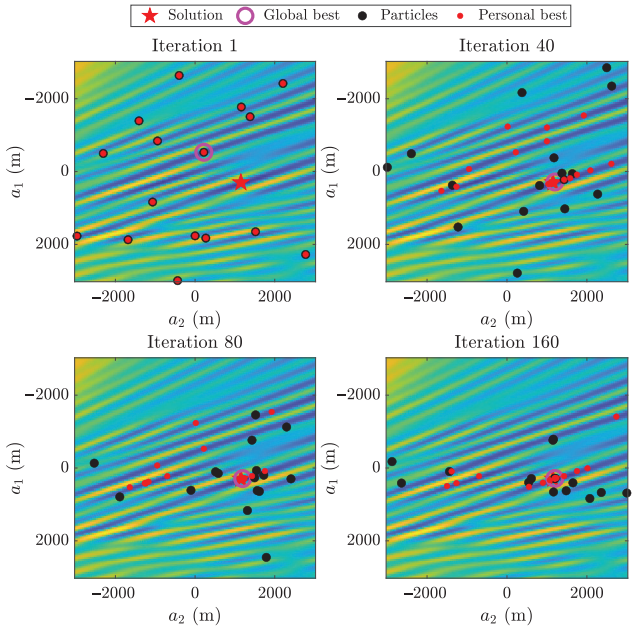

Figure 11. The PSO example showing the swarm configuration at different iterations superimposed on the misfit map. The personal best positions are shown in cyan, the global best is shown in magenta, and the global solution (obtained with grid search) is shown as a red star.

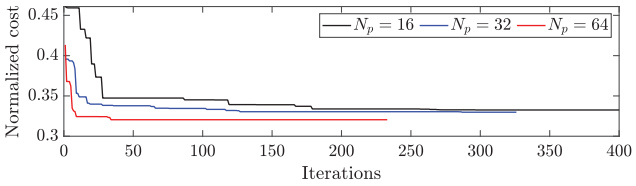

Figure 12. Global best normalized cost as a function of the number of iterations using 16, 32, and 64 particles.

Inverse crime settings

We first generate the observed data using the true model shown in Figure 13a. The finite-difference grid point 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 a central frequency of 4 Hz, and it is filtered with a high-pass filter with a cutoff frequency of 2 Hz (Figure 14). The acquisition used for the tests in this section is a fixed-spread acquisition, with 128 sources with 132 m spacing and 170 receivers with 100 m spacing. The starting model is obtained by a Gaussian smoothing of the true model (Figure 13b).

Before diving into a deeper analysis, we make a comparison of conventional FWI, extended-receiver FWI with static relocalization with $\alpha = 0.375$, and our time-dependent approach with $\alpha = 0.01$ and $\beta = 0.25$. The latter is parameterized with one segment and two control points. Because we use a single segment with first-order Lagrange polynomials, there is an implicit monotonicity; the relocalization is either increasing or decreasing. We have done experiments with more segments and higher-order polynomials, but the best results were obtained when we use this minimal parameterization. Note that we use horizontal relocalization only for this particular model because the water column is shallow (75 m). This is done to avoid receivers moving deeper into the model, where the properties vary rapidly, leading to some artifacts in the reconstruction. We

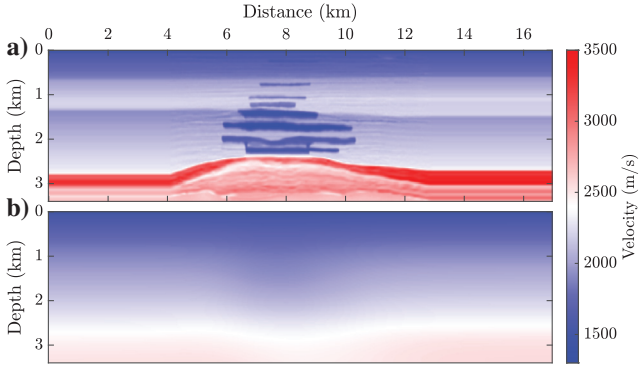

Figure 13. A North Sea exploration-scale synthetic model: (a) the true model used to generate the observed data and (b) the starting model.

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 but with visible defects. In contrast, the time-dependent approach is able to better reconstruct the velocity model from this crude starting model. The model error as a function of FWI iterations is shown in Figure 16. The model error is computed using

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

where M is the total number of model points, $m_{\text{est},i}$ is the estimated model at point i, and $m_{\text{true},i}$ is the true model at point i. The model error in 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, and the model error decreases quickly to a lower model error.

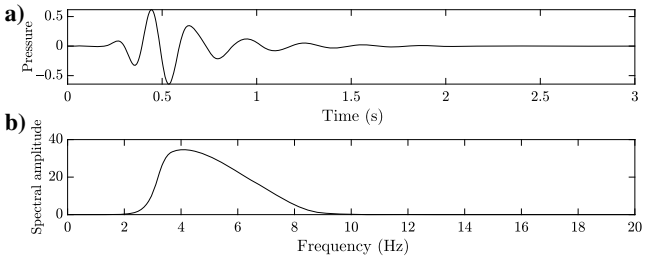

Figure 14. The source wavelet used for the numerical experiment: (a) time domain and (b) frequency domain.

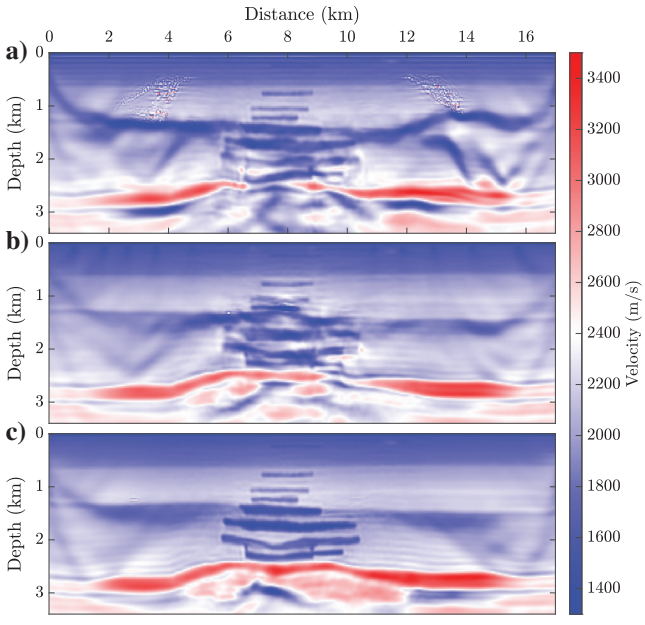

Figure 15. Reconstructed velocity models after 300 iterations: (a) conventional FWI, (b) static extended-receiver FWI, and (c) time-dependent extended-receiver FWI.

To better understand why it works, we take a look at the data fit for 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 in Figures 17 and 18 indicate a good fit). As the model estimate improves, the relocalization tends to zero, and 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 with the static approach

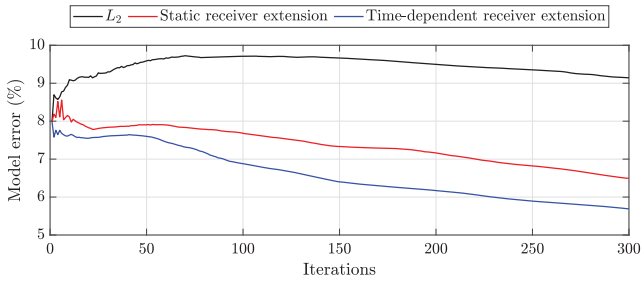

Figure 16. Receiver relocalization as a function of iterations: static extended-receiver FWI (red line) and time-dependent FWI (blue line).

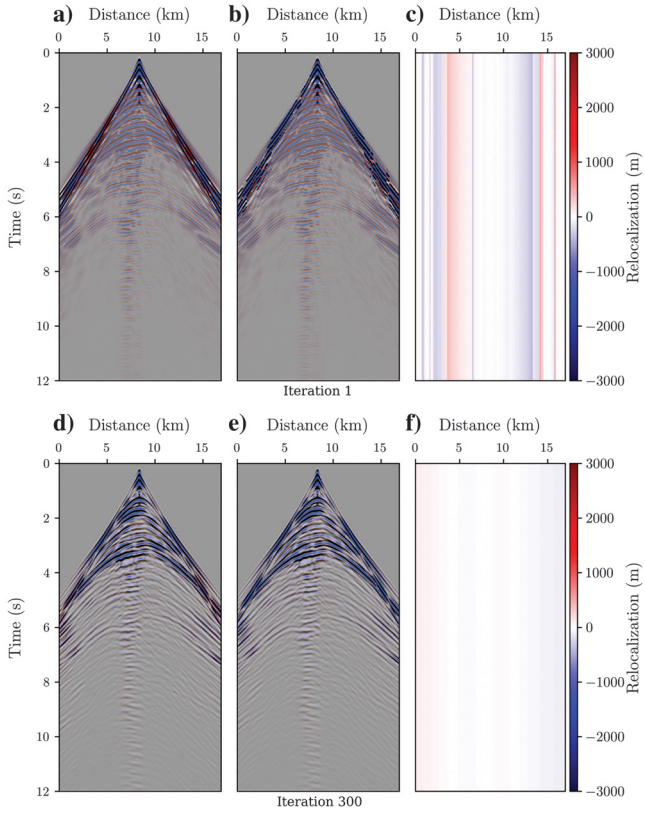

Figure 17. Synthetic and observed shot gathers as well as the relocalization gathers for the static case. The calculated data are displayed in a red and blue color scale and the observed data in grayscale. If only blue and black are apparent on the wiggles, a good fit is obtained; if red is shown, then the fit is not satisfactory. (a and d) Data fit before relocalization; (b and e) data fit after relocalization; and (c and f) relocalization gather, which has the same dimension as the shot gather. The first iteration is shown in (a)–(c), and the last iteration is shown in (d)–(f).

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 the relocalization is still important at the last iteration. Better explaining the data when the model estimate is poor leads to a better model reconstruction.

To complement our analysis, we show the misfit function evolution as a function of FWI iterations for the static and time-dependent cases in Figure 20a and 20b, respectively. The data fit term (L^2 norm of the data residuals without using relocalization) also is shown (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 also note 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 only 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 be viewed as a proxy for model error, and 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 also can visualize the receiver speed as a function of outer loop iterations, which is obtained using

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

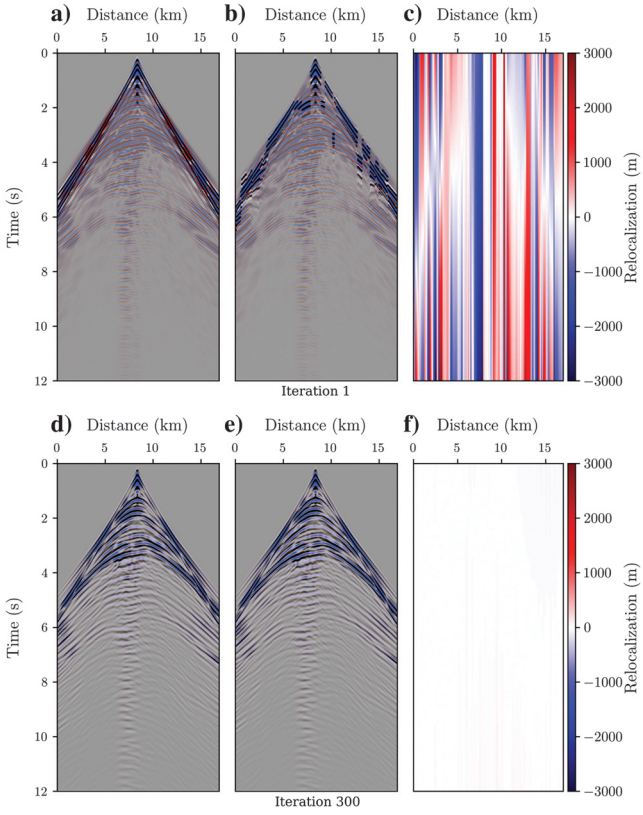

Figure 18. Synthetic and observed shot gathers as well as the relocalization gathers for the time-dependent case. We display the calculated data in a red and blue color scale and the observed data in grayscale. If only blue and black are apparent on the wiggles, a good fit is obtained; if red is shown, then the fit is not satisfactory. (a and d) Data fit before relocalization; (b and e) data fit after relocalization; and (c and f) relocalization gather, which has the same dimension as the shot gather. The first iteration is shown in (a)–(c), and the last iteration is shown in (d)–(f).

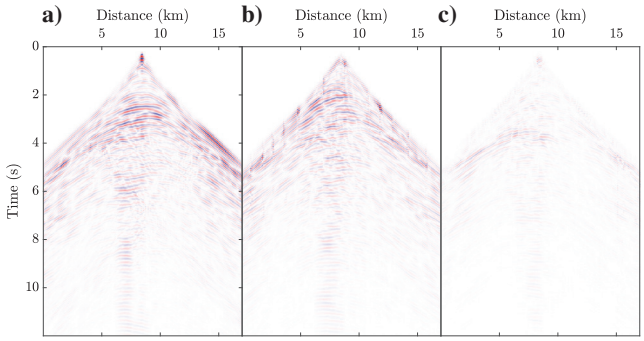

Figure 19. Adjoint source of the center shotpoint at the 300th iteration for (a) conventional FWI, (b) static receiver extension, and (c) time-dependent receiver extension. The same color scale is used for all figures.

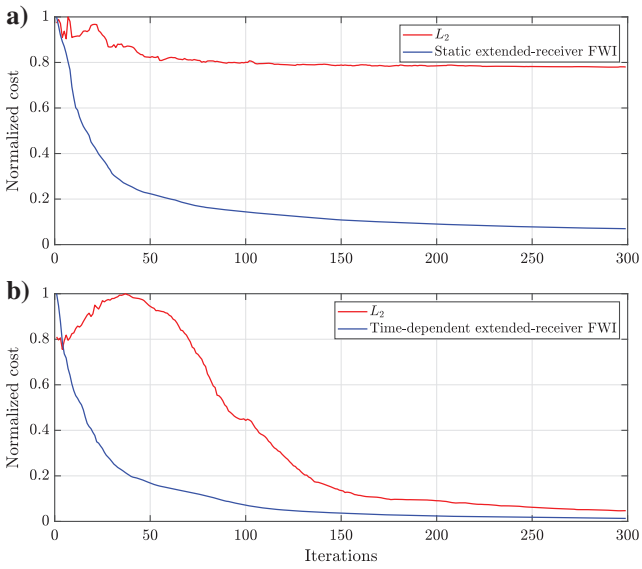

Figure 20. Cost function evolution as a function of iterations. The extended-receiver FWI cost is plotted in blue, and the L^2 norm of the data residuals at each iteration is shown in red. (a) Static extended-receiver FWI and (b) time-dependent extended-receiver FWI.

The receiver speed decreases monotonically as the model estimates improve (Figure 22). The receiver speed is not too large (maximum of approximately 500 ms^{-1} for the leftmost and rightmost gathers) because of the second penalty term. This is an expected behavior because 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.

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 to avoid the PSO tuning parameters impacting our sensitivity testing.

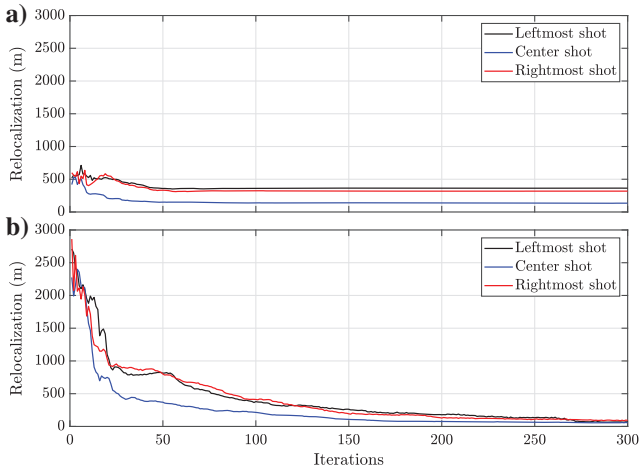

Figure 21. Relocalization as a function of iterations. The values are averaged from all receivers in each shown gather: (a) static extended-receiver FWI and (b) time-dependent FWI.

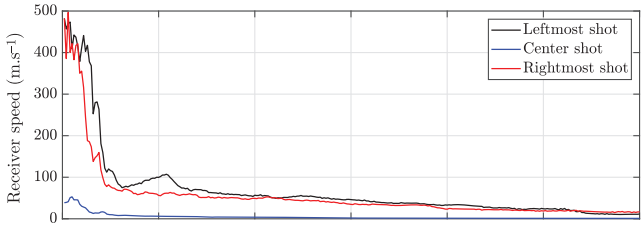

Figure 22. Receiver speed as a function of iterations for the time-dependent extended-receiver FWI. The values are averaged from all receivers in each shown gather.

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. The cost function at the last iteration and the model error as a function of α are shown 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 $\alpha = 0.125$, and it is caused by a line-search failure (i.e., 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 Figure 23a and 23b. The corresponding models are shown in Figure 24, and this is discussed in the next paragraph. Increased cost at the last iteration with increasing α is expected in theory because 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 (Figure 23c and 23d, respectively) is 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.

The reconstructed models that correspond to the selected α values are shown in Figure 24. The selected α values are shown as circles in Figure 23a and 23b. This approach appears to be sensitive to the choice of α. Small variations in the tuning parameter lead to

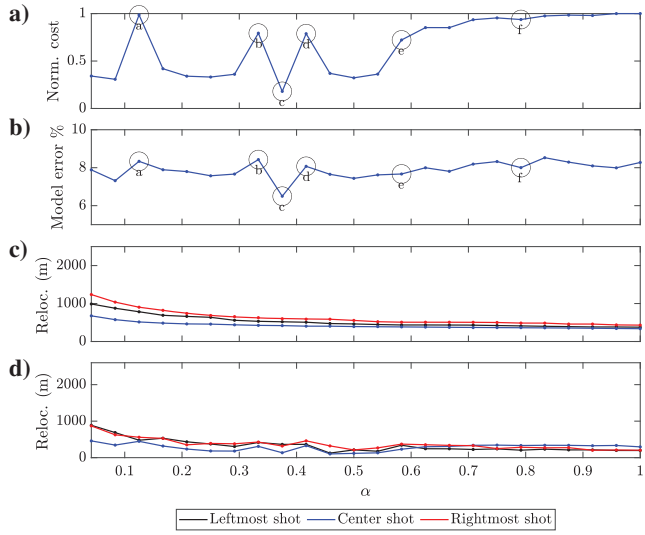

Figure 23. Tuning parameter α impact on extended-receiver FWI with static relocalization. (a) Normalized cost at the last iteration, (b) reconstructed model error, (c) relocalization at the first iteration, and (d) relocalization at the last iteration. The lower two panels show plots for the leftmost, center, and rightmost shotpoints. 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 corresponds to an α value with the same label.

a significant change in the reconstructed model. This means that the static relocalization approach is difficult to tune.

Time-dependent relocalization: Sensitivity to α

We perform a similar numerical experiment using our time-dependent strategy, in which we set $\beta = 0.25$ and test $24\alpha \in [0, 1]$. 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, which is done to keep the cost function plots comparable. The tuning parameter does not have a significant impact on the model error, and we also note that a lower model error is reached for all α values compared with the static counterpart. The impact of this tuning parameter on the relocalization at the first iteration appears to be roughly linear because 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 the black labeled circles in Figure 25a and 25b. The effect of α on the reconstruction appears to be minimal. This is an encouraging observation because it means that the method is easy to tune. For these tests, the β value has been kept constant, so how does our method behave when α and β are perturbed? To answer this question, we devise a parametric study, which we discuss in the next paragraphs.

Time-dependent relocalization: Sensitivity to α and β

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

Following the same setup as previously, 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 appears to be less sensitive to the choice of α. This can be seen in the cost obtained at the last iteration and in the model error, as shown in

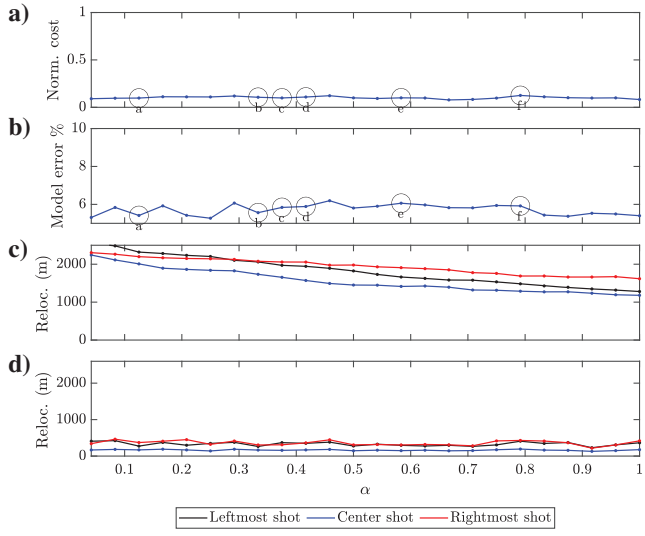

Figure 25. Tuning parameter α impact on extended-receiver FWI with time-dependent relocalization. (a) Normalized cost at the last iteration, (b) reconstructed model error, (c) relocalization at the first iteration, and (d) relocalization at the last iteration. The lower two panels show plots for the leftmost, center, and rightmost shotpoints. 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 corresponds to an α value with the same label.

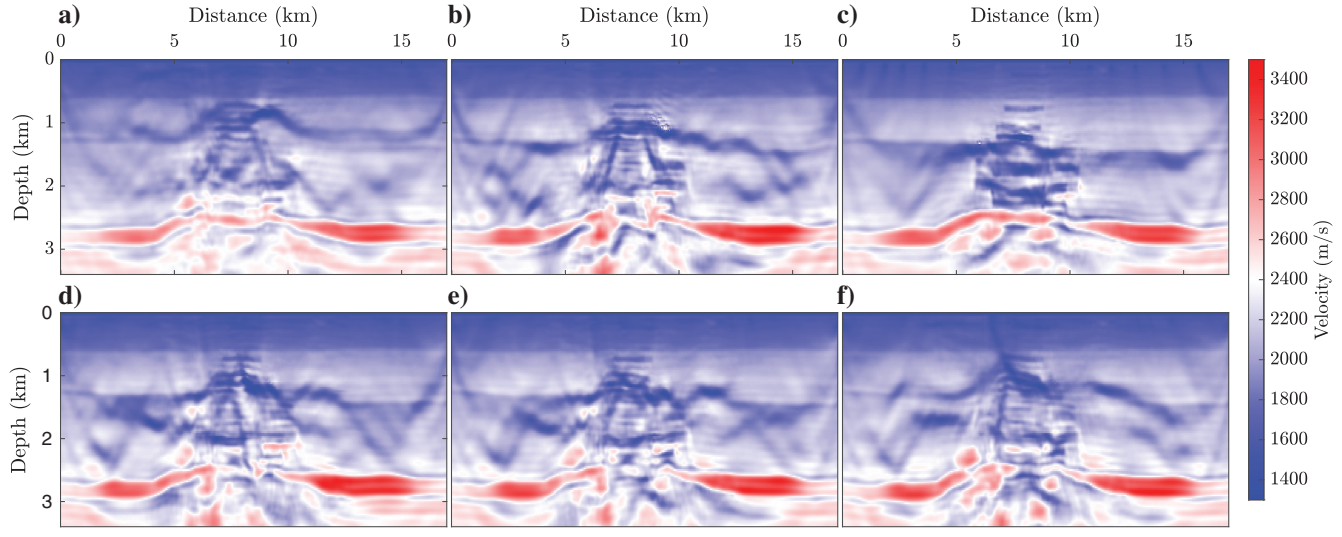

Figure 24. Tuning parameter α impact on velocity reconstruction using extended-receiver FWI with static relocalization: (a) $\alpha = 0.125$, (b) $\alpha = 0.333$, (c) $\alpha = 0.375$, (d) $\alpha = 0.4167$, (e) $\alpha = 0.5833$, and (f) $\alpha = 0.7917$. Each subfigure label corresponds to an α value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

Figure 27a and 27b, respectively. The minimum cost is obtained for $\alpha = 0.025$ and $\beta = 0.25$, and the minimum model error is obtained for $\alpha = 0.0075$ and $\beta = 0.75$. We show the corresponding final models in Figure 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 shotpoints). This occurs when β is large; the method then behaves as static relocalization, with the receiver speed being heavily constrained. To verify this, we select two points, which correspond to a higher cost and model error; they are shown as diamonds in Figure 29. The yellow diamond corresponds to $\alpha = 0.0005$ and $\beta = 0.1$, and the black diamond corresponds to $\alpha = 0.05$ and $\beta = 1$. The corresponding models are shown in Figure 29. The yellow point model reconstruction is satisfactory even though it corresponds to a relatively high cost and model error. As for the black point (upper right corner), the reconstruction is not satisfactory; because the receiver speed is constrained very much, it is essentially static extension. In fact, the reconstructed model is somewhat similar to the one obtained with static extension (Figure 15). 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.

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 viscoacoustic approximation using the V_P, ρ, and Q_P models shown in Figure 30. The source wavelet is a Ricker wavelet centered at 4 Hz. It is filtered with a high-pass filter with a cutoff

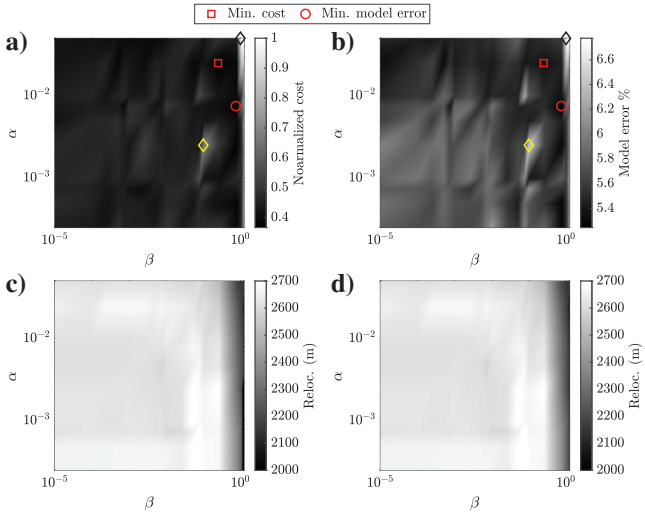

Figure 27. Tuning parameters impact on the (a) cost function at the last iteration, (b) model error, and (c and d) receiver relocalization, shown here for (c) the leftmost and (d) the rightmost shotpoints. The diamonds indicate two experiments: the yellow diamond corresponds to $\alpha = 0.0005$ and $\beta = 0.1$, and the black diamond in the upper right corner corresponds to $\alpha = 0.05$ and $\beta = 1$. We show the corresponding models in subsequent figures.

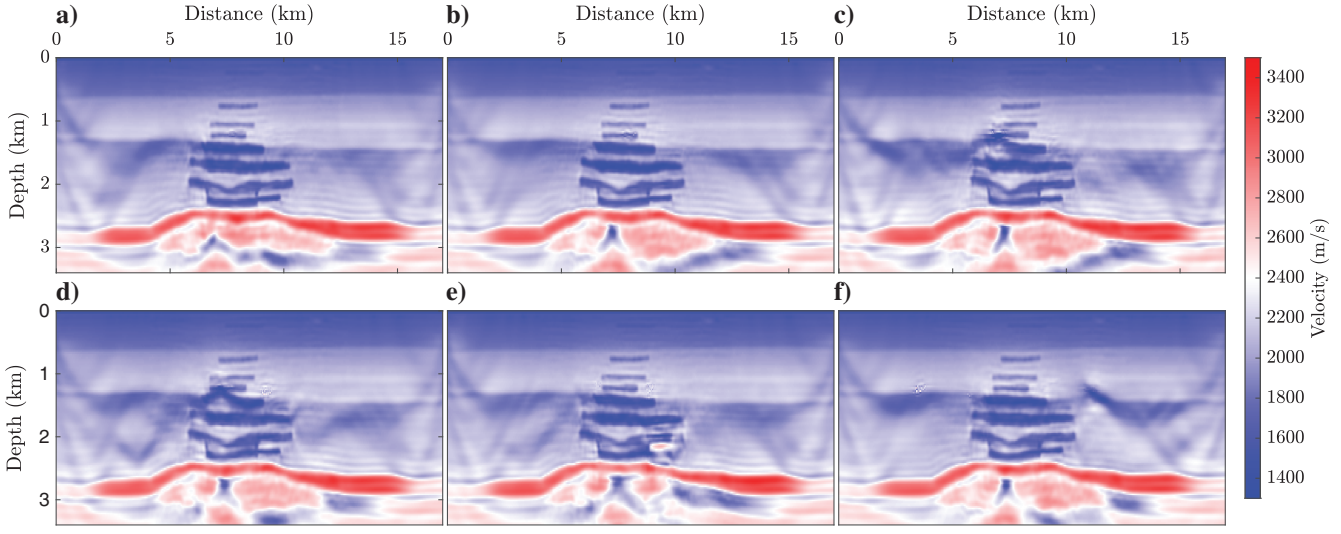

Figure 26. Tuning parameter α impact on velocity reconstruction using extended-receiver FWI with time-dependent relocalization: (a) $\alpha = 0.125$, (b) $\alpha = 0.333$, (c) $\alpha = 0.375$, (d) $\alpha = 0.4167$, (e) $\alpha = 0.5833$, and (f) $\alpha = 0.7917$. Each subfigure label corresponds to an α value indicated by black labeled circles in Figure 25a and 25b.

Table 1. The α and β values used in the parametric study.

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	1.0	1.25

The first row shows the values considered for α and β, and the second row shows the additional values we test only for β.

frequency of 2 Hz. As for the acquisition, it is a fixed-spread setup, with 128 sources with 117.5 m spacing and 150 receivers with 100 m spacing. The discretization step used for the finite differences is set to 12.5 m, the time step is set to 0.0015 s, and the number of time steps is 6000. The data are decimated to a time step of 0.003 s, and a band-limited Gaussian noise is added (Figure 31). The discretization step used for the forward computations during the inversion is set to 25 m. For the receiver extension, we set $\alpha = 0.01$ for the static and time-dependent cases. Only for the time-dependent approach, we set $\beta = 0.0025$, and the time-dependent relocalization is parameterized with one segment and two control points.

The starting velocity model is a 1D model (obtained with $V_P^{1D}(z) = 0.38z + V_P^{water}$), and the starting density model is obtained using Gardner's law on the starting velocity model, which is given by $\rho(x) = 1740(10^{-3}V(x))^{(1/4)}$. 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 32. 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 33. The inversion is performed for 300 iterations using conventional FWI, static receiver extension, and time-dependent receiver extension, with the previously computed wavelet.

We show the reconstructed models after the 15, 110, 205, and 300 iterations in Figure 34. As expected, conventional FWI is unable to reconstruct the velocity from the 1D starting model. Static relocalization performs slightly 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 and the relocalization gathers for the first and last iterations for both approaches in Figures 35 and 36. 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 encouraging.

DISCUSSION

In this section, we discuss the following three points: (1) the choice of PSO as our global optimization scheme, (2) the computational overhead of our strategy with respect to more conventional FWI approaches, and (3) some analogies of our time-dependent receiver extension approach with source extension methods and DTW.

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, MCMC methods (Aster et al., 2013), simulated annealing (Sen and

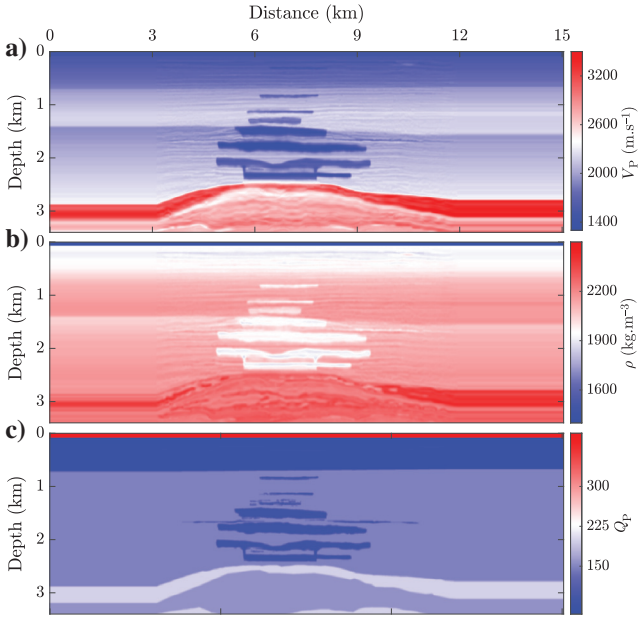

Figure 30. True models used to generate the observed data. (a) Velocity, (b) density, and (c) quality factor. The density and the quality factor in the water layer are 1000 kgm^{-3} and 1000, respectively. The color bars in (b) and (c) are clipped for clarity; therefore, the color in the water layer is not representative of the true values.

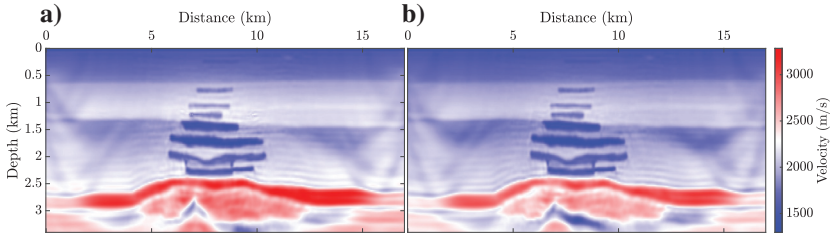

Figure 28. Reconstructed models corresponding to the (a) lowest cost and (b) lowest model error (Figure 27).

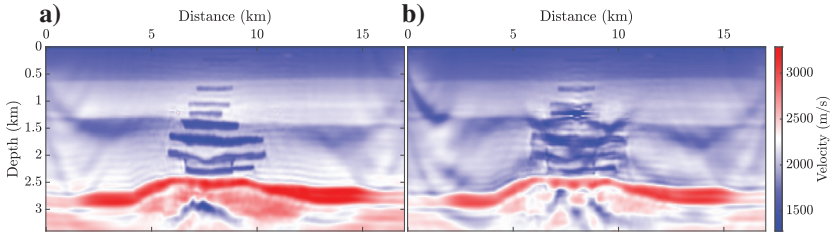

Figure 29. Reconstructed models corresponding to the two experiments shown as diamonds in Figure 27. (a) Yellow diamond with $\alpha = 0.0005$ and $\beta = 0.1$ and (b) black diamond with $\alpha = 0.05$ and $\beta = 1$.

Stoffa, 2013), and a variation of it, VFSA (Ingber, 1992, 1993). MCMC is a 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 because 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 because 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 is notoriously difficult to tune, namely, choosing a cooling schedule and its parameters. VFSA suffers from convergence issues for our problem, though it is easier to tune than conventional simulated annealing because it relies on a single cooling schedule and a single generating distribution, which reduces the number of tuning parameters. Grid search is costly even for the simplest parameterization we can consider (one segment with a first-order Lagrange polynomial, giving a 2D problem).

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 37). This plot is obtained by crossplotting each dimension against another, couple by couple. In other words, for a given point sampled by MCMC in this 3D space with the coordinates (a_1, a_2, a_3), we plot its position in two dimensions as follows: a_1 is plotted against a_2 and then against a_3. We assign the value of the cost function to the sample by providing the color map shown. The diagonal shows the histogram in each dimension. Because of 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 with the other couples, and the histograms for a_3 are flat. This is expected because points a_1 and a_3 correspond to the start and end of the seismic trace, respectively. The data values at early times and at 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, and this is particularly true for a_3.

We carry out the inner-loop optimization using MCMC, VFSA, and PSO. The evolution of the cost as a function of iterations is shown in Figure 38. 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

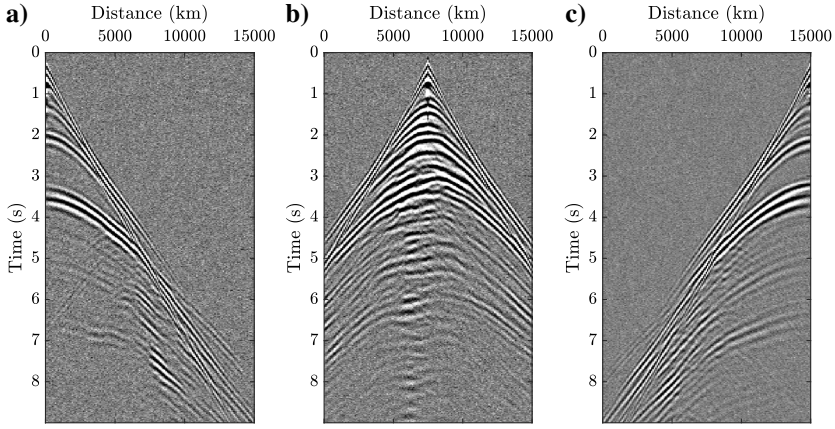

Figure 31. The observed data computed in a North Sea exploration-scale synthetic model: (a) leftmost shot gather, (b) center shot gather, and (c) rightmost shot gather.

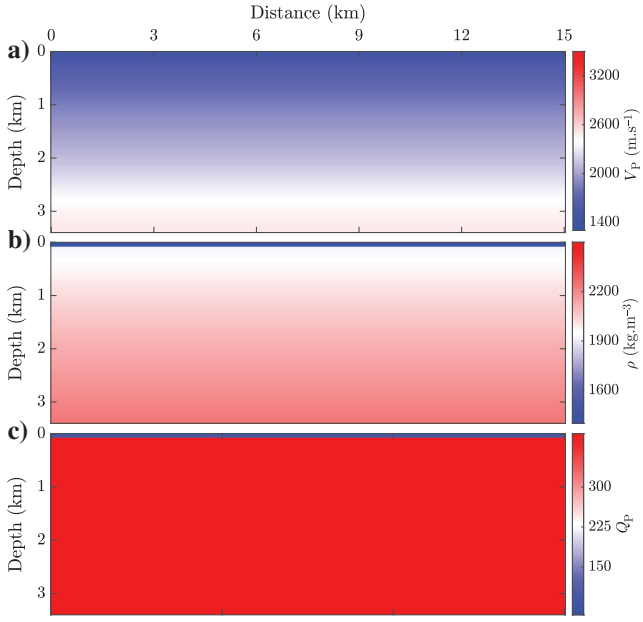

Figure 32. Starting models: (a) 1D velocity model, (b) density obtained using Gardner's law, and (c) quality factor set to 100 everywhere and 1000 in the water column. The color bars in (b) and (c) are clipped for clarity. As a result, the color in the water layer is not representative of the true values.

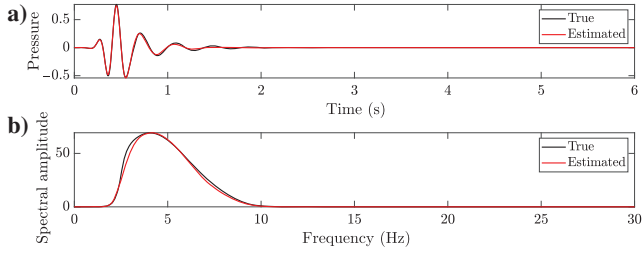

Figure 33. Source wavelet estimation in the starting velocity model: (a) source time function and (b) source frequency amplitude spectrum.

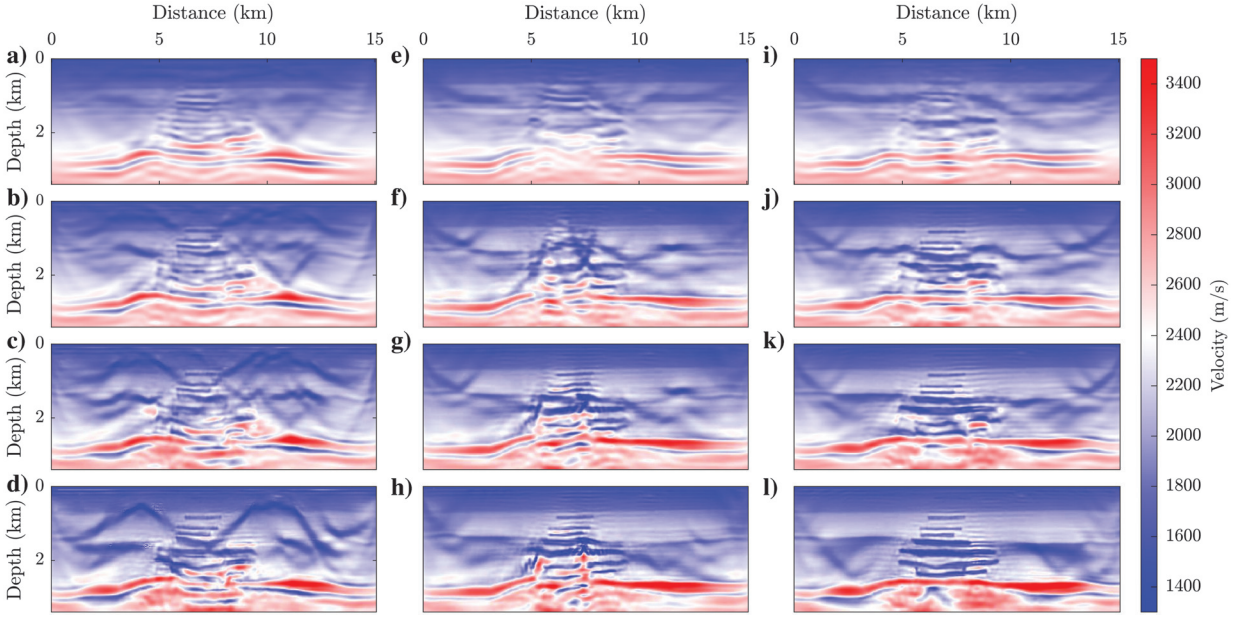

Figure 34. Reconstructed velocity models for (a–d) conventional FWI, (e–h) static receiver extension, and (i–l) time-dependent receiver extension: (a, e, and i) 15 iterations, (b, f, and j) 110 iterations, (c, g, and k) 205 iterations, and (d, h, and l) 300 iterations.

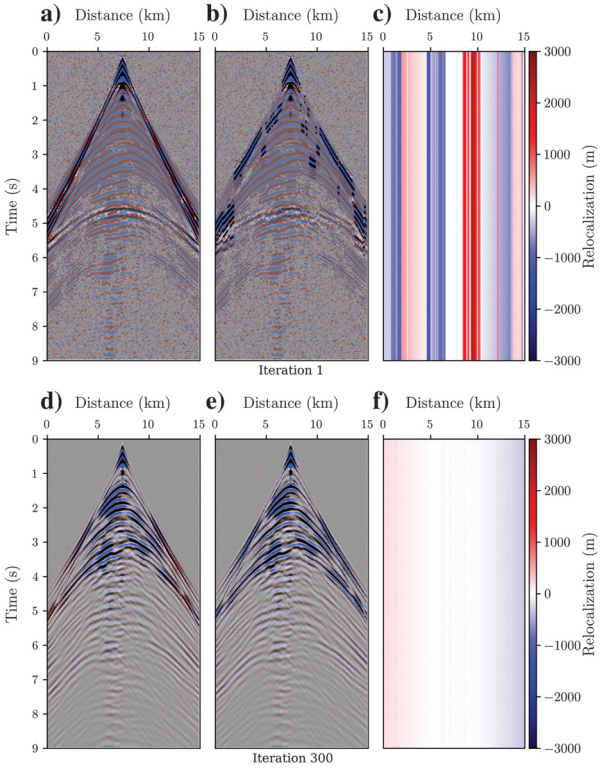

Figure 35. Synthetic and observed shot gathers as well as the relocalization gathers for the static case. The calculated data are displayed in a red and blue color scale, and the observed data are shown in grayscale. If only blue and black are apparent on the wiggles, a good fit is obtained; if red is shown, then the fit is not satisfactory. (a and d) Data fit before relocalization; (b and e) data fit after relocalization; and (c and f) relocalization gather, which has the same dimension as the shot gather. The first iteration is shown in (a)–(c), and the last iteration is shown in (d)–(f).

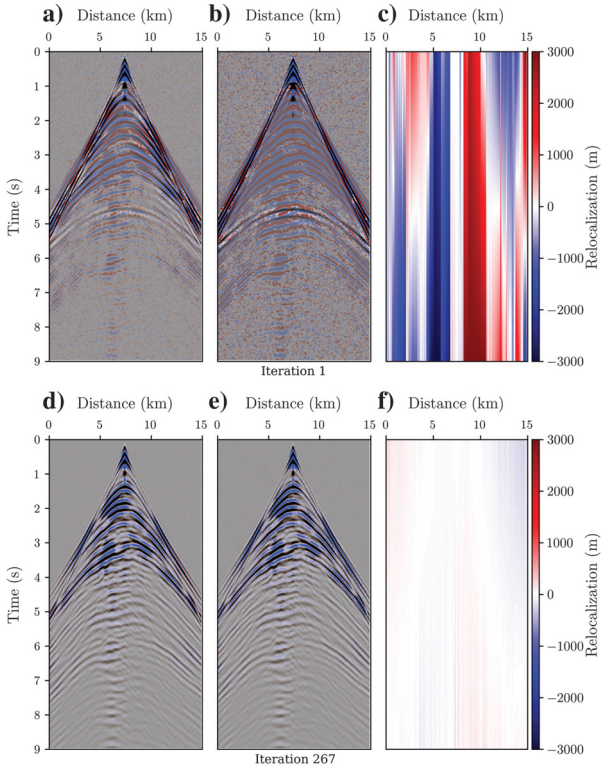

Figure 36. Synthetic and observed shot gathers as well as the relocalization gathers for the time-dependent case. The calculated data are displayed in a red and blue color scale, and the observed data are shown in grayscale. If only blue and black are apparent on the wiggles, a good fit is obtained; if red is shown, then the fit is not satisfactory. (a and d) Data fit before relocalization; (b and e) data fit after relocalization; and (c and f) relocalization gather, which it has the same dimension as the shot gather. The first iteration is shown in (a)–(c), and the last iteration is shown in (d)–(f).

have been performed. However, the plots show only 25,000 samples because 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 pseudorandom number generation process (Aster et al., 2013). MCMC provides a good solution, but it requires a large number of cost function evaluations. VFSA does 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, as shown in Figure 38b. The personal best cost for few particles is shown; different particles explore in the vicinity of various minima, leading to a convergent behavior of the swarm. Therefore, we use PSO for the inner-loop optimization. Furthermore, PSO tuning is less challenging given the few published meta-optimization studies (Pedersen, 2010; Mason et al., 2018) where we obtained the PSO parameters.

Computational cost

The central processing unit (CPU) times shown in Table 2 correspond to the realistic setting example we showed previously (Figure 30). Please note that we use 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 static and time-dependent receiver extension, and 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. Recall the adjoint equation

$$\begin{cases} A(m)^T \lambda_s[m,\overline{\Delta x(t)}](\mathbf{x},t) = \sum_{r=1}^{N_r} \tilde{R}^T_{s,r}[\overline{\Delta x_r(t)}]\tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r(t)},t) \\ \tilde{\mu}_s[m](\mathbf{x}_r + \overline{\Delta x_r(t)},t) = \tilde{d}_{\mathrm{cal},s}[m](\mathbf{x}_r + \overline{\Delta x_r(t)},t) - d_{\mathrm{obs},s}(\mathbf{x}_r,t) \end{cases}. \tag{38}$$

The cost of adjoint simulation is more important for the time-dependent approach because we use Kaiser-windowed sinc interpolation (Hicks, 2002) to inject the adjoint sources at moving receiver positions. This is performed for all receivers at every time step. Even though the cost increase is not negligible, it is certainly manageable. In addition, 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 complexities for the forward and adjoint simulations are $\mathcal{O}(n^2)$ in two dimensions and $\mathcal{O}(n^3)$ in three dimensions. As for the inner loop, the complexity in terms of the number of receivers is $\mathcal{O}(n)$ in two dimensions and $\mathcal{O}(n^2)$ in three dimensions. The large number of receivers in three dimensions can constitute a bottleneck. We address this point in the conclusion of this paper.

Analogies with source extension and DTW

Please note that receiver extension methods (described here) are not the reciprocal of source extension methods. Receiver extension is a receiver relocalization problem, where the added freedom is the

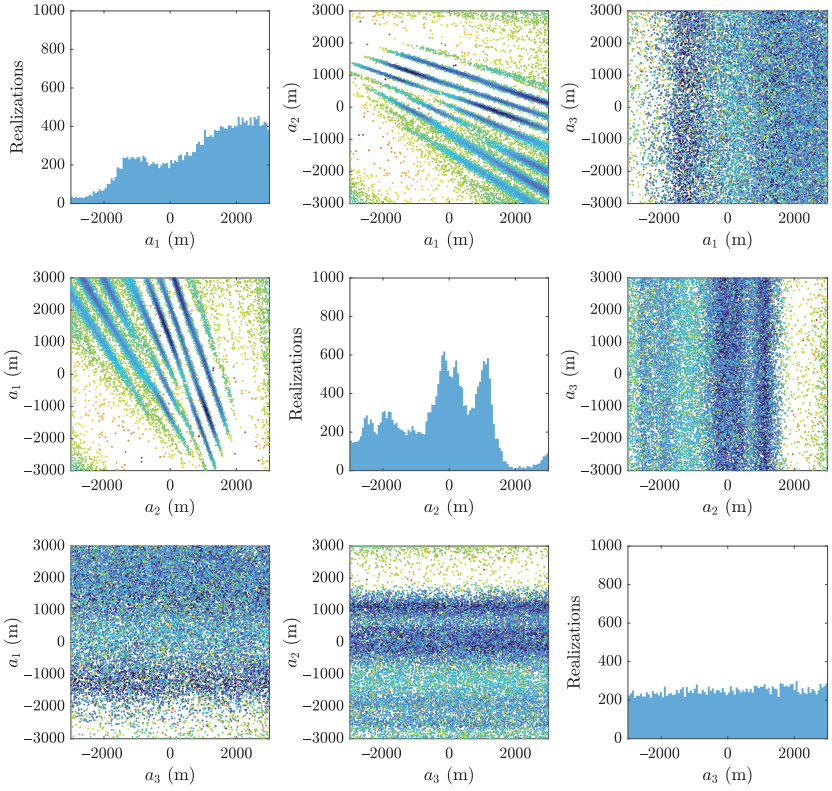

Figure 37. Corner plot of the misfit function sampled by MCMC for three model parameters (a_1 , a_2, a_3). The off-diagonal panels show the 2D joint distributions for each parameter pair, with each sample colored by its corresponding misfit value. The diagonal panels show the 1D marginal distribution for each parameter as a histogram.

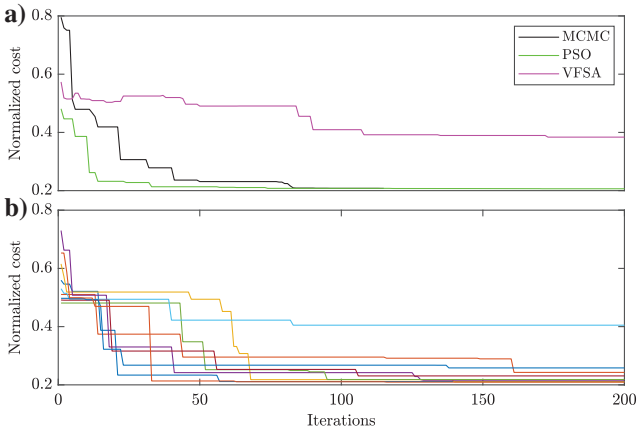

Figure 38. Cost function values as a function of inner-loop iterations for one receiver. (a) MCMC candidates that lower the cost functions are shown in black, PSO global-best cost is shown in green, and VFSA best solutions are shown in magenta. (b) The personal-best cost function of few particles as a function of iterations.

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

	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

receiver position. In contrast, source extension methods are a wavefield extension method. The wavefield (or equivalently, the source term) serves to relax the original FWI problem: instead of being computed as a solution of the wave equation up to machine precision, a least-squares trade-off between the satisfaction of the wave equation and the fit to the data is performed. Thus, source extension and receiver extension are fundamentally different.

Another point we would like to bring to the attention of readers is the potential analogy between the time-dependent receiver extension described in this study and DTW of seismic images (Hale, 2013). DTW aims to align two time series by applying a time-varying time shift. In other words, the sought trace is obtained by assigning samples from one trace onto the other. In contrast, our time-dependent receiver extension aligns two traces by applying a time-varying space shift of the receiver position. Therefore, although the former obtains the trace by a nonphysical distortion and dilation of the phases, the latter (our strategy) provides a solution, which is always a solution of the wave equation. This point is crucial and is at the core of the efficacy of the strategy, especially the convexity of the misfit function with respect to m in the outer loop. This point is illustrated in a schematic example in Appendix C.

CONCLUSION AND PERSPECTIVES

Extended-receiver FWI with time-dependent relocalization has shown promising results. The ease with which this method is directly applied to time-domain FWI and 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 nonnegligible but certainly manageable computation cost.

We would like to mention a possible limitation of the method for reconstructing attenuation parameters. Because receiver extension impacts the amplitude of the data (geometrical spreading) and the frequency content (Doppler effect), this could constitute an ambiguity to reconstruct attenuation parameters such as quality factors. A potential strategy to overcome this difficulty is to use our method as a robust velocity model-building tool and then invert attenuation using conventional FWI when the velocity estimate has improved. Nevertheless, further investigation is required.

The next step of our work is extending the method to three dimensions as well as an application to a field data set. The 3D extension of this method relies on the idea of allowing the receivers to move either toward 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 number of receivers. This has encouraged us to develop an alternative parameterization and optimization for the inner loop to make the cost manageable for 3D application. This strategy relies on the observation that the receiver relocalization problem is broken down into a series of nested smaller problems, in a similar fashion to DTW (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.

Another point of interest is multicomponent FWI (Cao et al., 2022) because the method has been developed and tested for 1C (hydrophone) data to date. It also might be interesting to investigate the feasibility of receiver extension using a polarization-based misfit function (Sambolian et al., 2022).

ACKNOWLEDGMENTS

This study was partially funded by the SEISCOPE consortium (http://seiscope2.osug.fr), sponsored by Aker BP, DUG, ExxonMobil, Geolinks, JGI, Petrobras, Shearwater, Shell, Sinopec, TotalEnergies, and Viridien. This study was granted access to the high performance computing (HPC) resources provided by the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, the HPC resources of Cray Marketing Partner Network (https://partners.cray.com), and those of CINES/IDRIS/TGCC under the allocation 046091 made by GENCI.

DATA AND MATERIALS AVAILABILITY

Data and code associated with this research belong to the SEISCOPE consortium. All requests related to the materials availability need to be addressed to Romain Brossier or Ludovic Métivier.

APPENDIX A

FULL MISFIT FUNCTION EXPRESSION FOR EXTENDED-RECEIVER FWI

We write the full expression of the receiver extension misfit function:

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

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

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

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

The second term in the right-hand side of equation A-2 ($\mathcal{P}_1$ in equation 19) penalizes the receiver relocalization to prevent it from being too large and to force it to become small along iterations. Here, 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). Here, $V_{\max}$ is the maximum allowed receiver speed and β is a tuning parameter.

APPENDIX B

THE TERM $\mathcal{P}_2$ AND THE DOPPLER EFFECT

The second penalty term is added to constrain the receiver speed. This is important because 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:

$$f = \frac{v_p \pm v_r}{v_p} f_o, \quad \text{(B-1)}$$

where f is the observed frequency, 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 toward the source and subtracted in the other case. Equation B-1 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

$$f = \frac{v_p}{v_p \pm v_s} f_o, \quad \text{(B-2)}$$

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 B-1 using the same two-layer 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.) We can see that the spectrum is different in the case where no constraint is imposed on the receiver speed (red plot). In particular, energy at low frequencies is added due to the 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.

Another interesting observation can be made for the experiment where the source is moving (an adjoint source, our receiver). The adjoint field at time 3.18 s is shown in Figure B-2a. The source is moving to the right at a speed that is higher than the medium velocity, and the structure that appears on the left of the moving source is called a Mach cone. The gradient resulting from this adjoint field is

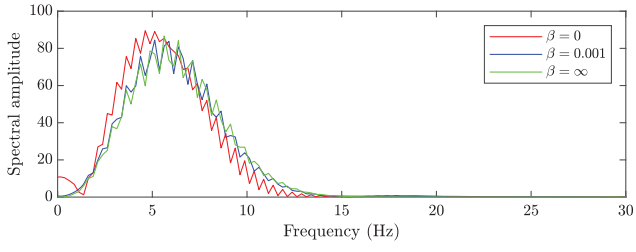

Figure B-1. Calculated data amplitude spectrum for three cases: time-dependent relocalization without the second penalty term ($\beta = 0$, shown in red) and with the second penalty term ($\beta = 0.001$) as well as the static case ($\beta = \infty$). The notches in the spectra are caused by the free-surface reflection.

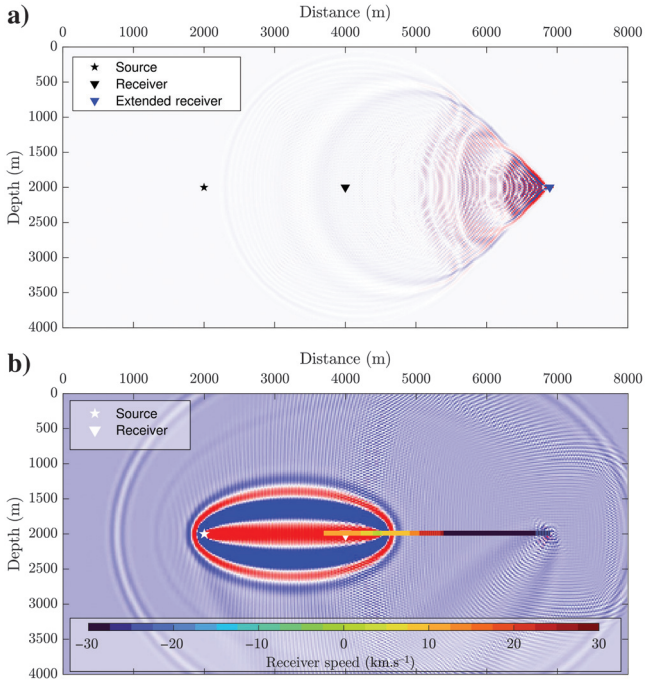

Figure B-2. Gradient computation with $\beta = 0$: (a) adjoint field snapshot at 3.18. s where the extended receiver (adjoint source) is moving to the right, causing the Mach cone shape in the wavefield. (b) Extended-receiver FWI kernel, with receiver positions shown, and the color indicates the receiver speed.

shown in Figure B-2b. The positions occupied by the moving receiver are superimposed to the gradient, and 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.

APPENDIX C

ON THE ANALOGY BETWEEN TIME-DEPENDENT RECEIVER EXTENSION AND DTW

DTW aligns two time series by applying sample-by-sample shifts; in other words, it applies a time-varying time shift. This technique has been used for decades in speech recognition (Sakoe and Chiba, 1978). Hale (2013) proposes a DTW algorithm for seismic traces and images. DTW can be used as a misfit function, as introduced by Ma and Hale (2013). The authors formulate a wave equation reflection traveltime tomography method, where the traveltimes are obtained using DTW. Wave equation reflection traveltime tomography writes

$$\min_{\tau} f_{\mathbf{T}}(m) = \min_{\tau} \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int \tau_{s,r}(t)^2 dt, \tag{C-1}$$

where τ is the optimal time-dependent time shift, which is obtained using

$$\tau_{s,r} = \underset{l}{\operatorname{argmin}}\, D_{s,r}(l), \quad \text{subject to} \left| \frac{\partial \tau_{s,r}}{\partial t} \right| \leq \sigma_t, \tag{C-2}$$

with

$$D_{s,r}(l) = \int |d_{\mathrm{cal},s}(\mathbf{x}_r, t + l_r(t)) - d_{\mathrm{obs},s}(\mathbf{x}_r, t)|^2 dt, \tag{C-3}$$

where $l_r(t)$ is a time-dependent time shift. The constrained minimization in equation C-2 is solved using dynamic programming (Hale, 2013). The constraint in dynamic programming controls the amount of shift allowed from one time sample to the next, tuned by the quantity σ_t. Then, the wave equation tomography misfit function in equation C-1 is minimized.

DTW also can be used as an extension method, aiming to align the data in inaccurate models. The following two-variable optimization problem is therefore considered (Huang et al., 2021):

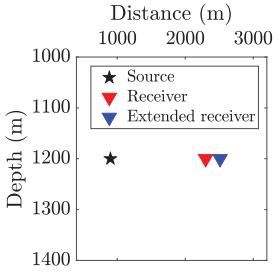

Figure C-1. Experimental setup. The source is shown as a star, the receiver as a red triangle, and the extended receiver as a blue triangle. Please note that the extended receiver position changes for each velocity.

$$\tilde{f}_{\mathbf{T}}(m, l(t)) = \frac{1}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int |d_{\mathrm{cal},s}(\mathbf{x}_r, t + l_r(t)) - d_{\mathrm{obs},s}(\mathbf{x}_r, t)|^2 dt + \frac{\alpha}{2} \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \int l_{s,r}(t)^2 dt. \tag{C-4}$$

The analogy between our time-dependent receiver extension method and DTW is thus clear when DTW is used as an extension method. As a reminder, our time-dependent receiver extension method writes

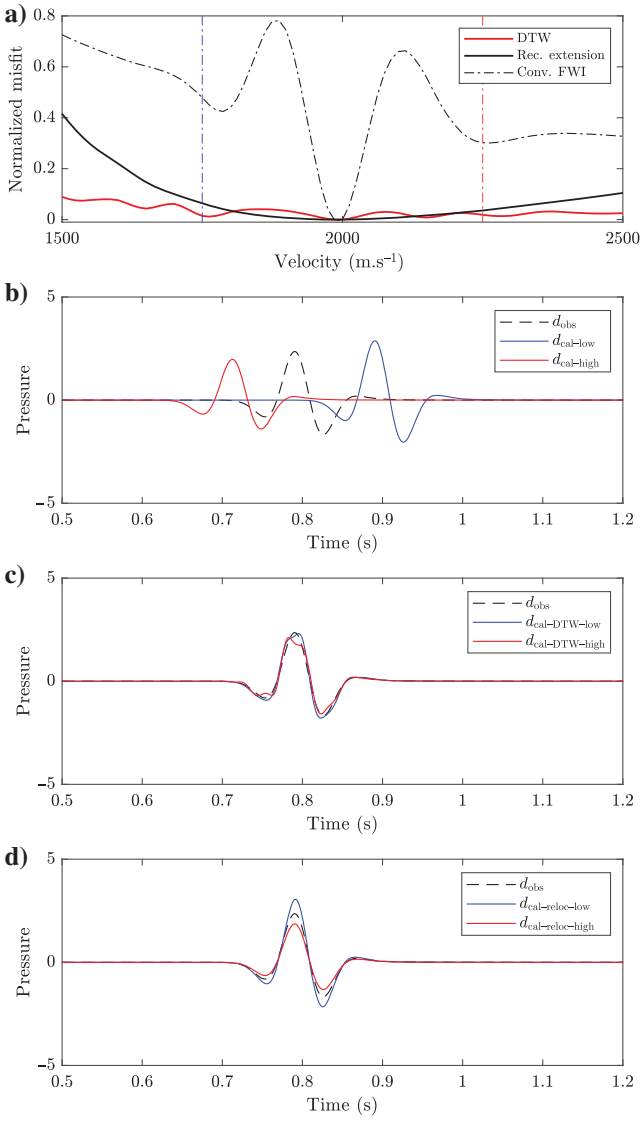

Figure C-2. Comparison of receiver extension and DTW: (a) misfit computed for different velocities; (b) conventional FWI case: observed data (dashed black line) and two synthetic traces corresponding to low and high velocities (blue and red squares in [a], respectively); (c) DTW case: observed trace synthetic traces corresponding to low and high velocities (blue and red squares, respectively, in [a]); and (d) receiver extension case: observed trace, synthetic traces corresponding to low and high velocities (blue and red squares in [a]).

$$
\begin{aligned}
&f(m,\Delta x(t))\\
&=\frac{1}{2}\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\int |d_{\mathrm{cal},s}(\mathbf{x}_r+\Delta x_{s,r}(t),t)-d_{\mathrm{obs},s}(\mathbf{x}_r,t)|^2 dt\\
&+\alpha\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mathcal{P}_1[\Delta x_{s,r}(t)]\\
&+\beta\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\mathcal{P}_2[\Delta \dot{x}_{s,r}(t)(t)],
\end{aligned}
\tag{C-5}
$$

where the synthetic data are extracted at time-dependent receiver positions $\mathbf{x}_r+\Delta x_{s,r}(t)$ from the synthetic wavefield. The strong correspondence between DTW as an extension method and our time-dependent receiver extension method is clear from equations C-4 and C-5.

However, despite this analogy, there is a fundamental difference that we want to stress. Although DTW aligns the data by applying a time-varying time shift and therefore compressing and dilating the traces, impacting the phase of the data, and the amplitudes, in a nonphysical way, our time-dependent receiver extension method generates "extended" traces, which are always solutions of the wave equation. This difference might appear anecdotic, but we illustrate in the following example why this is fundamental.

Consider the canonical setting (Figure C-1): in a homogeneous medium with the true velocity 2000 ms^{-1}, we compute the misfit function depending on velocities ranging from 1500 to 2500 ms^{-1}. This is straightforward for least-squares-based conventional FWI. For DTW and time-dependent receiver extension, we compute this misfit for the optimal time-shift function (in the case of DTW) and time-dependent receiver shift (in the case of time-dependent receiver extension). This is the misfit function that would correspond to the outer loop in both cases, the one we hope the extension makes convex.

As shown in Figure C-2a, the resulting misfit function in the case of DTW is nonconvex, but the one corresponding to the receiver extension is convex. To understand why, it is good to see how both extensions align the traces in time in incorrect velocity models. What we can see from Figure C-2b–C-2d is that the time-shift alignment is imperfect, leading to a nonzero data-misfit term in the misfit function (equation C-4). The alignment with receiver extension is imperfect as well. The main difference is that the "aligned" data, in the receiver extension case, are kinematically compatible with the observed data, with only an amplitude difference associated with the physics of wave propagation (geometrical spreading). This amplitude attribute depends quadratically on the velocity and hence the convexity of the receiver extension misfit function. In contrast, the misfit between the DTW aligned data and observed data does not behave monotonically with respect to the velocity. One way to overcome this difficulty is to switch to the misfit function modification version of DTW (as shown in equation C-1), which is a purely tomographic misfit function; however, this involves other difficulties slightly outside the scope of our study, including the requirement to use a connectivity function to compute the gradient of the misfit function (Luo and Schuster, 1991) and less robustness or effectiveness of this purely kinematic misfit function compared with the extension method where the least-squares fit to the data is kept. It also might be possible to investigate appropriate penalty weight to make the DTW extension behave better; however, this also is beyond the scope of this study. However, we would like to stress the fact that time-dependent receiver extension is easy to tune because it is not very sensitive to its tuning parameters, as illustrated in this study. This fundamental difference is at the core of our interest in the receiver extension strategy.

REFERENCES

Aghamiry, H., A. Gholami, and S. Operto, 2018, Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers: Geophysics, **84**, no. 1, R125–R148, doi: 10.1190/geo2018-0093.1.

Aghamiry, H., A. Gholami, and S. Operto, 2019a, ADMM-based multiparameter wavefield reconstruction inversion in VTI acoustic media with TV regularization: Geophysical Journal International, **219**, 1316–1333, doi: 10.1093/gji/ggz369.

Aghamiry, H., A. Gholami, and S. Operto, 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, doi: 10.1093/gji/ggz386.

Aghamiry, H., A. Gholami, and S. Operto, 2020, Accurate and efficient wavefield reconstruction in the time domain: Geophysics, **85**, no. 2, A7–A12, doi: 10.1190/geo2019-0535.1.

Aster, R. C., B. Borchers, and C. H. Thurber, 2013, Chapter eleven — Bayesian methods, *in* R. C. Aster, B. Borchers, and C. H. Thurber, eds., Parameter estimation and inverse problems, 2nd ed.: Academic Press, 253–280.

Barnier, G., E. Biondi, R. G. Clapp, and B. Biondi, 2023a, Full-waveform inversion by model extension: Practical applications: Geophysics, **88**, no. 5, R609–R643, doi: 10.1190/geo2022-0382.1.

Barnier, G., E. Biondi, R. G. Clapp, and B. Biondi, 2023b, Full-waveform inversion by model extension: Theory, design, and optimization: Geophysics, **88**, no. 5, R579–R607, doi: 10.1190/geo2022-0350.1.

Benziane, M., R. Brossier, and L. Métivier, 2023, Receiver extension strategy using a timedependent relocalization approach for time-domain FWI: 84th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.202310583.

Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, no. 3, WA129–WA140, doi: 10.1190/geo2013-0340.1.

Biondi, B., and P. Sava, 1999, Wave-equation migration velocity analysis: 69th Annual International Meeting, SEG, Expanded Abstracts, 1723–1726, doi: 10.1190/1.1820867.

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, doi: 10.1111/j.1365-246X.2011.04970.x.

Brossier, R., S. Operto, and J. Virieux, 2009, 2D elastic frequency-domain full-waveform inversion for imaging complex onshore structures: 71st Annual International Conference and Exhibition, EAGE, Extended Abstracts, U019, doi: 10.3997/2214-4609.201400376.

Brossier, R., S. Operto, and J. Virieux, 2015, Velocity model building from seismic reflection data by full waveform inversion: Geophysical Prospecting, **63**, 354–367, doi: 10.1111/1365-2478.12190.

Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473, doi: 10.1190/1.1443880.

Cao, J., R. Brossier, and L. Métivier, 2022, Elastic full-waveform inversion of 4C oceanbottom seismic data: Model parameterization analysis: Second International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts, 957–961, doi: 10.1190/image2022-3751342.1.

Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition for discrete acoustic and elastic wave equations: Geophysics, **50**, 705–708, doi: 10.1190/1.1441945.

Chauris, H., and E. Cocher, 2017, From migration to inversion velocity analysis: Geophysics, **82**, no. 3, S207–S223, doi: 10.1190/geo2016-0359.1.

Engelbrecht, A. P., 2007, Computational intelligence: An introduction, 2nd ed.: John Wiley & Sons, Ltd, 289–358.

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**, no. 1, R63–R77, doi: 10.1190/geo2021-0186.1.

Guasch, L., M. Warner, and C. Ravaut, 2019, Adaptive waveform inversion: Practice: Geophysics, **84**, no. 3, R447-R461, doi: 10.1190/geo2018-0377.1.

Guo, G., S. Operto, A. Gholami, and H. S. Aghamiry, 2024, Time-domain extended-source full-waveform inversion: Algorithm and practical workflow: Geophysics, **89**, no. 2, R73–R94, doi: 10.1190/geo2023-0055.1.

Haber, E., U. Ascher, D. Aruliah, and D. Oldenburg, 2000, Fast simulation of 3D electro-magnetic problems using potentials: Journal of Computational Physics, **163**, 150–171, doi: 10.1006/jcph.2000.6545.

Hale, D., 2013, Dynamic warping of seismic images: Geophysics, **78**, no. 2, S105–S115, doi: 10.1190/geo2012-0327.1.
Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97–109, doi: 10.1093/biomet/57.1.97.
Hicks, G. J., 2002, Arbitrary source and receiver positioning in finite-difference schemes using Kaiser windowed sinc functions: Geophysics, **67**, 156–165, doi: 10.1190/1.1451454.
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**, no. 5, R449–R461, doi: 10.1190/geo2017-0333.1.
Huang, G., R. Nammour, and W. W. Symes, 2018b, Volume source-based extended waveform inversion: Geophysics, **83**, no. 5, R369–R387, doi: 10.1190/geo2017-0330.1.
Huang, G., R. Nammour, W. W. Symes, and M. Dolliazal, 2019, Waveform inversion via source extension: 89th Annual International Meeting, SEG, Expanded Abstracts, 4761–4766, doi: 10.1190/segam2019-3216338.1.
Huang, G., J. Ramos-Martínez, Y. Yang, and N. Chemingui, 2021, FWI in extended domain using time-warping: First International Meeting for Applied Geoscience & Energy, SEG, Expanded Abstracts, 817–821, doi: 10.1190/segam2021-3594800.1.
Huang, G., and W. W. Symes, 2015, Full waveform inversion via matched source extension: 85th Annual International Meeting, SEG, Expanded Abstracts, 1320–1325, doi: 10.1190/segam2015-5872566.1.
Ingber, L., 1992, Very fast simulated re-annealing: Mathematical and Computer Modelling, **12**, 967–973, doi: 10.1016/0895-7177(89)90202-1.
Ingber, L., 1993, Simulated annealing: Practice versus theory: Mathematical and Computer Modelling, **18**, 29–57, doi: 10.1016/0895-7177(93)90204-C.
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, Wave-lengths of earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910, doi: 10.1190/1.1442719.
Kennedy, J., and R. Eberhart, 1995, Particle swarm optimization: Proceedings of International Conference on Neural Networks, 1942–1948.
Komatitsch, D., and R. Martin, 2007, An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation: Geophysics, **72**, no. 5, SM155–SM167, doi: 10.1190/1.2757586.
Levander, A. R., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436, doi: 10.1190/1.1442422.
Liu, Q., and D. Peter, 2018, One-step data-domain least-squares reverse-time migration: Geophysics, **83**, no. 4, R361–R368, doi: 10.1190/geo2017-0622.1.
Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792, doi: 10.1190/1.3627773.
Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653, doi: 10.1190/1.1443081.
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, doi: 10.1016/j.cageo.2018.01.016.
Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion: Geophysics, **78**, no. 6, R223–R233, doi: 10.1190/geo2013-0004.1.
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, doi: 10.1016/j.asoc.2017.10.018.
Métivier, L., and R. Brossier, 2022, Receiver-extension strategy for time-domain fullwaveform inversion using a relocalization approach: Geophysics, **87**, no. 1, R13–R33, doi: 10.1190/geo2020-0922.1.
Métivier, L., R. Brossier, Q. Mérigot, and E. Oudet, 2019, A graph space optimal transport distance as a generalization of Lp distances: Application to a seismic imaging inverse problem: Inverse Problems, **35**, 085001, doi: 10.1088/1361-6420/ab206f.
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, doi: 10.1093/gji/ggw014.
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, doi: 10.1063/1.1699114.
Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33, doi: 10.1111/1365-2478.12073.
Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782, doi: 10.1090/S0025-5718-1980-0572855-7.
Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.
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**, no. 6, R671–R702, doi: 10.1190/geo2022-0758.1.
Pedersen, M. E. H., 2010, Good parameters for particle swarm optimization.
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, doi: 10.1111/j.1365-246X.2006.02978.x.
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, doi: 10.1190/1.1444597.
Sakoe, H., and S. Chiba, 1978, Dynamic programming algorithm optimization for spoken word recognition: IEEE Transactions on Acoustics, Speech, and Signal Processing, **26**, 43–49, doi: 10.1109/TASSP.1978.1163055.
Sambolian, S., R. Brossier, and L. Métivier, 2022, Exploiting the richness of multicomponent data: A time-dependent polarization-based FWI approach: 83rd Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.202210482.
Sen, M. K., and P. L. Stoffa, 2013, Global optimization methods in geophysical inversion, 2nd ed.: 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, doi: 10.1046/j.1365-246X.2002.01645.x.
Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, doi: 10.1190/1.1649391.
Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: 10.1111/j.1365-2478.2008.00698.x.
van Leeuwen, T., and F. Herrmann, 2016, A penalty method for PDE-constrained optimization in inverse problems: Inverse Problems, **32**, 1–26, doi: 10.1088/0266-5611/32/1/015007.
van Leeuwen, T., and F. J. Herrmann, 2013, Mitigatingfin local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667, doi: 10.1093/gji/ggt258.
van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, 1383–1394, doi: 10.1111/j.1365-246X.2010.04681.x.
Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite difference method: Geophysics, **51**, 889–901, doi: 10.1190/1.1442147.
Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.
Wang, C., D. Yingst, P. Farmer, I. Jones, G. Martin, and J. Leveille, 2017, Reconstructed full-waveform inversion with the extended source: 87th Annual International Meeting, SEG, Expanded Abstracts, 1449–1453, doi: 10.1190/segam2017-17736054.1.
Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, no. 3, WA13–WA24, doi: 10.1190/geo2013-0294.1.
Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016a, Checkpointing-assisted reverse forward simulation: an optimal recomputation method for FWI and RTM: 86th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093, doi: 10.1190/segam2016-13685603.1.
Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016b, Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method: Geophysics, **81**, no. 6, R349–R362, doi: 10.1190/geo2016-0082.1.
Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018a, A time-domain preconditioned truncated newton approach to multiparameter visco-acoustic full waveform inversion: SIAM Journal on Scientific Computing, **40**, B1101–B1130.
Yang, P., R. Brossier, and J. Virieux, 2016c, Wavefield reconstruction from significantly decimated boundaries: Geophysics, **81**, no. 5, T197–T209, doi: 10.1190/geo2015-0711.1.
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**, no. 1, R43–R62, doi: 10.1190/geo2016-0663.1.
Yao, G., D. Wu, and S.-X. Wang, 2020, A review on reflection-waveform inversion: Petroleum Science, **17**, 334–351, doi: 10.1007/s12182-020-00431-3.
Yong, P., R. Brossier, L. Métivier, and J. Virieux, 2022, Localized adaptive waveform inversion: Theory and numerical verification: Geophysical Journal International, **233**, 1055–1080, doi: 10.1093/gji/ggac496.

Biographies and photographs of the authors are not available.