Receiver-extension strategy for time-domain full waveform inversion using a relocalization approach

Ludovic Métivier, Romain Brossier

Receiver-extension strategy for time-domain full waveform inversion using a relocalization approach

Ludovic Métivier[1,2], Romain Brossier[2]

[1]CNRS, Univ. Grenoble Alpes, LJK, F-38000 Grenoble, France

[2]Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France

(August 26, 2021)

Running head: **L. Métivier & R. Brossier**

ABSTRACT

A receiver-extension strategy is presented as an alternative to recently promoted source-extension strategies, in the framework of high resolution seismic imaging by full waveform inversion. This receiver-extension strategy is directly applicable in time-domain full waveform inversion, and unlike source-extension methods it incurs negligible extra computational cost. After connections between difference source-extension strategies are reviewed, the receiver-extension method is introduced and analyzed for single-arrival data. The method results in a misfit function convex with respect to the velocity model in this context. The method is then applied to three exploration scale synthetic case studies representative of different geological environment, based on: the Marmousi model, the BP 2004 salt model, and the Valhall model. In all three cases the receiver-extension strategy makes it possible to start full waveform inversion with crude initial models, and reconstruct meaningful subsurface velocity models. The good performance of the method even considering inaccurate amplitude prediction due to noise, imperfect modeling, and source wavelet estimation, bodes well for

field data applications.

INTRODUCTION

Full waveform inversion (FWI) is a high resolution seismic imaging strategy. At the core of the method is a partial-differential-equations (PDE) constrained optimization problem, which is solved by iteratively reducing a misfit between calculated and observed data, as initially introduced by Lailly (1983) and Tarantola (1984). Continuous progress in the understanding of this geophysical imaging problem, as well as the design of wide azimuth/wide offset seismic acquisition systems and the development of high performance computing platforms have led to the current success of FWI. It is now routinely applied in the industry for exploration scale targets (Sirgue et al., 2010; Plessix and Perkins, 2010; Warner et al., 2013; Vigh et al., 2014; Operto et al., 2015; Raknes et al., 2015; Solano and Plessix, 2019), and in academia for crustal, regional and global scale imaging, yielding unprecedented high resolution 3D reconstruction of subsurface mechanical parameters (Fichtner and Villaseñor, 2015; Bozdağ et al., 2016; Górszczyk et al., 2017; Beller et al., 2018; Lei et al., 2020; Lu et al., 2020). A recent overview of FWI and its applications can be found in Virieux et al. (2017).

Despite this success, challenges remain for a wide and more automated application of FWI, especially at the exploration scale. The main reason is the absence of sufficient low frequency content in exploration data, yielding the well known cycle skipping problem (Virieux and Operto, 2009). From a mathematical perspective, cycle skipping is due to the non-convexity of the misfit function which is iteratively minimized. As FWI relies on local optimization techniques, the presence of local minima in the misfit function is harmful: if the starting model is not in the basin of attraction of the global minimum, the method converges to a possibly non-informative local minimum.

In practice, this issue is overcome through the careful design of data-based hierarchical

schemes. The main ingredient is a multi-scale approach, leading to the interpretation of the data from low to high frequency (Bunks et al., 1995). Interpreting the low frequency content first reduces the number of phases in the data and thus enlarges the basin of attraction of the global minimum. This strategy is usually complemented with time-windowing and offset selection strategies, to foster the interpretation of specific arrivals, such as diving waves, to constrain a specific part of the medium and again reduce the risk of cycle skipping (Shipp and Singh, 2002; Wang and Rao, 2009; Brossier et al., 2009). This complex design requires human expertise, can be time-consuming, and can also question the robustness of the results while increasing the uncertainty attached to them. What is the sensitivity of the inversion to the different choices made to design the workflow?

For this reason, research efforts are still dedicated to the design of more robust and efficient full waveform inversion schemes. To give an overview of this research field, it is convenient to split the proposed methods into two categories. In the first group, the focus is on the misfit measurement. Alternative misfit functions are proposed, with a desired improved convexity with respect to time-shifts, seen as a good proxy for convexity with respect to velocities (Jannane et al., 1989). Cross-correlation, (Luo and Schuster, 1991; van Leeuwen and Herrmann, 2013), deconvolution (Luo and Sava, 2011; Warner and Guasch, 2016), normalized integration (Donno et al., 2013), instantaneous envelope and phase (Fichtner et al., 2008; Bozdağ et al., 2011; Wu et al., 2014), optimal transport (Engquist and Froese, 2014; Métivier et al., 2016; Yang et al., 2018b; Métivier et al., 2019) are instances of the many methods which have been investigated in this frame. Some of these methods have been applied only on synthetic cases, while other have shown interesting properties in the frame of 3D field data applications. A common feature of all these methods is the presence of tuning parameters which might be sometimes difficult to control.

It is not our purpose here to elaborate on this first group. We focus instead on the second group, which could be labeled as "extension strategies". The fundamental idea is slightly different. The non-convexity of least-squares FWI is linked to the increased nonlinearity of the inverse problem with respect to the model parameters induced by the reduced space approach used to solve the PDE-constrained optimization problem. To overcome this difficulty, artificial degrees of freedom are injected in the problem, which shall be gradually eliminated along the convergence path to recover a physical solution. These degrees of freedom help to fit the data in the early iterations when the model estimate is poor.

When these extension methods are model based, they are generally known as migration velocity analysis (MVA) methods. A quite complete overview of these techniques is proposed in Symes (2008). Based on the scale separation assumption (subsurface parameters split in a smooth background and a sharp reflectivity model), the artificial degrees of freedom are introduced at the reflectivity level. The FWI problem is reformulated as the focusing of the extended reflectivity model at zero time-lag or zero subsurface offset or alternatively as the flattening of the extended reflectivity in the offset or angle direction. Mathematical analysis shows that in a transmission regime, under specific mathematical conditions which can be related to the absence of triplication the resulting problem asymptotically converges to a travel-time tomography problem, known to be convex (Symes, 2014).

More recently, a class of source extension strategies has emerged, named matched source waveform inversion (MSWI) (Huang et al., 2018a,b, 2019). Preliminary concepts on source extension had already been proposed in Almomin (2016). In this approach, the artificial degrees of freedom are introduced at the source instead of being introduced at the model level. This overcomes a series of limitations encountered by MVA approaches. In practice, the high computational cost for building extended reflectivity hypercubes makes it difficult

to apply MVA to 3D field data. A more fundamental difficulty is related to complex data with multi-arrival and multiple reflections (Cocher et al., 2017). As will be detailed further in this study, MSWI is equivalent to wavefield reconstruction inversion techniques (WRI), another class of methods previously introduced to relax cycle skipping in FWI van Leeuwen and Herrmann (2013, 2016); Aghamiry et al. (2019b). Note that the deconvolution approach introduced as adaptive waveform inversion (AWI) by Warner and Guasch (2016) can also be recast as a MSWI technique. This shows that the distinction between misfit function modification methods and extension strategies is not as clear as one could think, however it is convenient to draw a landscape of the numerous investigations performed in this field.

MSWI techniques have shown promising results on 2D synthetic applications in the frame of frequency-domain FWI. Theoretical results for a 1D transmission canonical case also show that, depending on the chosen formulation and particularly the choice of annihilator operator, MSWI can yield a convex misfit function. The use of the variable projection method to solve the extended inversion problem, detailed in the next subsection, seems also key to the success of such strategies (Symes et al., 2020). However, their implementation in the frame of time-domain FWI is still under development. Such an implementation is required to handle 3D field data applications. Indeed, frequency-domain FWI is for now limited to moderate size targets. This is due to the lack of scalability of the direct solvers on which they rely to solve harmonic equations (see Li et al., 2020, for a recent status on the capabilities of direct solvers to solve large scale harmonic wave equation problems). The reason why time-domain MSWI techniques are difficult to design is detailed in this study. In essence, MSWI requires the solution of a square wave propagation problem, which is possible in the harmonic case when a factorization of the wave propagation operator is available, but which is much more difficult to solve in the time-domain case through explicit time-stepping

algorithms.

This intrinsic difficulty for MSWI methods to be applied in the time-domain is the motivation of this study. We propose here an alternative extension strategy, based on the receivers rather than the source. We propose to introduce the receiver location as the artificial degree of freedom in the inversion. As will be shown, this avoids the introduction of a square wave propagation operator and thus makes this method applicable directly in the time-domain at a reasonable computational cost. In addition, introducing the receiver position as a new unknown makes possible to mitigate cycle skipping. The kinematic mismatch is compensated by the repositioning of the receivers which is slowly relaxed to the true receiver position. After presenting the method on a schematic cross-hole example, we illustrate how our algorithm works on 2D synthetic (visco-)acoustic examples based on the Marmousi, BP 2004 and 2D Valhall synthetic models. In all three cases, our receiver relocalization strategy makes possible to start FWI with crude initial models, outperforming standard least-squares based inversion.

The structure of the study is as follows. First, we give an overview of the theory behind MSWI methods. Then, we introduce our receiver extension strategy. We illustrate the fundamental properties of the algorithm on a schematic transmission case. We then present the application of our algorithm to three synthetic benchmark models. We propose finally a discussion, after what we conclude and we give some opening perspectives.

BACKGROUND AND STATE OF THE ART ON MSWI AND WRI METHODS

FWI as a PDE-constrained optimization problem

FWI can be cast as the following PDE-constrained optimization problem

$$\min_m \frac{1}{2}\sum_{s=1}^{N_s} \|Ru_s - d_{obs,s}\|^2_{\mathcal{D}}, \quad s.t. \;\; A(m)u_s = b_s, \quad s = 1, \dots, N_s, \tag{1}$$

where m denotes the subsurface model parameters which are to be reconstructed, $N_s \in \mathbb{N}$ is the number of source positions used to generate the data, $d_{obs,s}$ is the s-th shot gather, $A(m)$ is a general wave equation operator (from acoustic to visco-elastic), $u_s[m]$ is the synthetic wavefield solution of the wave equation for the s-th source position, $b_s(t)$ is the source term of the s-th wavefield, and R is a restriction operator mapping the wavefield u_s to the receivers location. Here and in the following, $\|.\|_{\mathcal{D}}$ will refer to the following L^2 norm in the data space: for a shot gather d, we will have

$$\|d\|^2_{\mathcal{D}} = \sum_{r=1}^{N_r} \int_0^T |d(x_r, t)|^2 dt, \tag{2}$$

where N_r corresponds to the number of receivers and x_r denotes the receiver positions.

The Lagrangian operator associated with this PDE-constrained optimization problem is

$$L(m, u, \lambda) = \frac{1}{2}\sum_{s=1}^{N_s} \|Ru_s - d_{obs,s}\|^2_{\mathcal{D}} + \sum_{s=1}^{N_s} \langle \lambda_s, A(m)u_s - b_s \rangle_{\mathcal{W}} \tag{3}$$

where $u = (u_1, \dots, u_{N_s})$ gathers the N_s synthetic wavefields, $\lambda = (\lambda_1, \dots, \lambda_{N_s})$ gathers the N_s adjoint wavefields, and $\langle ., . \rangle_{\mathcal{W}}$ is the Euclidean scalar product in the wavefield space. For

two wavefields u, v we have

$$\langle u, v \rangle_{\mathcal{W}} = \int_0^T \int_\Omega u(x,t) v(x,t) dx dt, \tag{4}$$

where Ω represents the subsurface.

Finding a solution to the PDE-constrained optimization problem 1 is equivalent to find a saddle point of the Lagrangian operator by solving the min max problem

$$\min_{u,m} \max_{\lambda} L(m, u, \lambda). \tag{5}$$

However, the computational cost for solving the problem 5 through local optimization is prohibitive: aside the convergence rate, it would imply all incident and adjoint wavefields in space and time, which is not affordable for realistic size FWI application. The reduced space approach is thus conventionally used. The problem 5 is transformed into the unconstrained optimization problem

$$\min_m \frac{1}{2} \sum_{s=1}^{N_s} \| RA(m)^{-1} b_s - d_{obs,s} \|_{\mathcal{D}}^2. \tag{6}$$

This conventional form for FWI is known to exhibit local minima into which local optimization solvers can converge. Compared with the problem 5, the nonlinearity with respect to the model parameter becomes more apparent in the term $RA(m)^{-1}b_s$, which corresponds to the solution of the wave equation for a given model parameter m.

WRI and MSWI formalism

As noted by van Leeuwen and Herrmann (2013), the problem 5 is only "mildly" nonlinear. Indeed, the Lagrangian $L(m, u, \lambda)$ depends linearly on λ. In addition, because of the

bilinearity of the wave equation operator it also depends linearly on m and quadratically on u. We express this bilinearity by introducing the operator $F(m,u)$

$$F(m,u) = A(m)u, \tag{7}$$

and the identity

$$F(m,u) = A(m)u = B(u)m. \tag{8}$$

This identity shows that the wave propagation problem can be rewritten equally as a linear operator $A(m)$ acting on u or a linear operator $B(u)$ acting on m. This identity is useful in the following developments. This property is true for general elastic and visco-elastic wave propagation, up to the choice of the parameterization for m, as is discussed in Aghamiry et al. (2019a).

This apparent "well behaved" property motivates the design of WRI (van Leeuwen and Herrmann, 2013). With the idea to make the nonlinearity with respect to m less stringent, they propose to reformulate the FWI problem using a quadratic penalty method instead of using the reduced space approach (Nocedal and Wright, 2006). This method, coined as wavefield reconstruction inversion (WRI), is expressed as

$$\min_{m,u} \frac{1}{2} \sum_{s=1}^{N_s} \|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta \sum_{s=1}^{N_s} \|F(m,u_s) - b_s\|_{\mathcal{W}}^2 . \tag{9}$$

where $\|.\|_{\mathcal{W}}$ is the Euclidean norm associated with the scalar product $\langle .,. \rangle_{\mathcal{W}}$. The wave equation is not imposed as a strict constraint, instead it should be fitted in the least-squares sense. This reformulation implies a change of paradigm: from a parameter estimation problem posed on m only (reduced space approach), FWI becomes a compatibility problem

where both the wavefield u and the model parameter m are reconstructed from partial observations d_{obs} and a priori knowledge of the physics of wave propagation (the operator $A(m)$). In this frame, solving exactly for the wave equation to compute u at each iteration while the model m is known to be only poorly approximated does not appear as a good choice, hence the freedom added on the reconstruction of u. The level of accuracy for the wavefield to satisfy the wave equation is controlled with the penalty parameter η.

Later on, Aghamiry et al. (2019b) have proposed an improvement of the WRI strategy where the FWI problem is reformulated following an augmented Lagrangian approach, which presents several advantages over the quadratic penalty method regarding convergence rate issues and selection of the parameter η (Nocedal and Wright, 2006). The Iteratively-Refined Wavefield Reconstruction Inversion (IR-WRI) is formulated as

$$\min_{m,u}\max_{\lambda} = \frac{1}{2}\sum_{s=1}^{N_s}\|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2 + \sum_{s=1}^{N_s}\langle\lambda_s, F(m,u_s)-b_s\rangle_{\mathcal{W}} + \eta\sum_{s=1}^{N_s}\|F(m,u_s)-b_s\|_{\mathcal{W}}^2 . \quad (10)$$

that is the standard Lagrangian augmented with the quadratic penalty term.

Please note however that there is no formal guarantee of the existence of a unique solution to the problem 5. Such a proof would require the operator $F(m,u) = A(m)u$ to be convex which is not the case (bilinearity does not imply convexity). A recent mathematical analysis of WRI also shows that WRI asymptotically tends to standard FWI in the context of pure 1D acoustic transmission and suffers from the same non-convexity problems in this case (Symes, 2020).

In parallel, Huang et al. (2018a,b) have proposed a matched source waveform inversion (MSWI) method. MSWI relies on an extended modeling operator making use of an extended source. In Huang et al. (2018a) this extension is proposed in space and time while in Huang

et al. (2018b) the extension is performed only in space, with the time signature of the source supposed to be known *a priori* and treated by deconvolution. In the general case of space and time extension, the extended source can be denoted by $\widetilde{b}(x,t) = \left(\widetilde{b}_1, \ldots, \widetilde{b}_{N_s}\right)$. MSWI is then formulated as

$$\min_{m,\widetilde{b}} = \frac{1}{2}\sum_{s=1}^{N_s} \|S(m)\widetilde{b}_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta \sum_{s=1}^{N_s} \|\widetilde{b}_s - b_s\|_{\mathcal{W}}^2, \tag{11}$$

where $S(m) = RA(m)^{-1}$ is the forward problem operator.

The philosophy of MSWI relies on the frame of extended inversion. Unphysical degrees of freedom are added to the modeling operator to help fit the data. In the case of MSWI the source is not punctual in space, and possibly the time signature becomes also an unknown. An annihilator is added to the misfit function to constrain the additional degrees of freedom towards physical values at convergence. In the case of MSWI, the extended source shall be localized on the correct source location with the correct time signature at convergence. For simplicity we restrict this annihilator here as the least-squares misfit but more general annihilator can be used (Huang et al., 2018a,b).

Interestingly, as noted by Wang et al. (2016) and Huang et al. (2018a), the change of variables $\widetilde{b}_s = F(m, u_s)$ yields

$$S(m)\widetilde{b}_s = RA(m)^{-1}F(m,u_s) = RA(m)^{-1}A(m)u_s = Ru_s. \tag{12}$$

Using this identity, we see that MSWI with a least-squares annihilator is equivalent to WRI. The difference between MSWI and WRI relies on the choice of unknown: $\widetilde{b}$ for MSWI, the source wavefield u for WRI.

Numerical solution and limitation for time-domain applications

We now explain the origin of the limitations of WRI, IR-WRI and MSWI when considering time-domain inversion. All three approaches rely on the minimization of a misfit function which depends on two parameters: the model parameter m and an additional parameter (wavefield u or extended source $\widetilde{b}$). The minimization is achieved by defining an outer minimization loop over the model parameter m and an inner loop on the additional parameter. This method is often referred to as variable projection approach (Golub and Pereyra, 2003). We recall it formally as it will be used throughout the paper.

Nested loop optimization

Consider the joint problem

$$\min_{x_1,x_2} f(x_1,x_2). \tag{13}$$

Assuming f is twice differentiable with respect to x_1 and x_2, the problem 13 is equivalent to

$$\min_{x_1} g(x_1), \tag{14}$$

where

$$g(x_1) = f\left(x_1, \overline{x_2}(x_1)\right), \quad \overline{x_2}(x_1) = \arg\min_{x_2} f(x_1,x_2). \tag{15}$$

The outer loop is the minimization of $g(x_1)$ and the computation of $\overline{x_2}(x_1)$ is the inner loop. This method is interesting in practice when the computation cost of the inner minimization over x_2 is cheap *i.e.* a quadratic problem with a closed form formula is solved. Gradient-based or quasi-Newton methods are then conventionally used to minimize $g(x_1)$ in the outer

loop. Interestingly, the gradient of $g(x_1)$ is given by

$$\nabla g(x_1) = \frac{\partial f}{\partial x_1}\left(x_1, \overline{x_2}(x_1)\right) + \frac{\partial f}{\partial x_2}\frac{\partial \overline{x_2}}{\partial x_1}, \tag{16}$$

however because of the definition of $\overline{x_2}(x_1)$ as a minimizer of $f(x_1, x_2)$ with respect to x_2 the second term in the right hand side vanishes and we have

$$\nabla g(x_1) = \frac{\partial f}{\partial x_1}\left(x_1, \overline{x_2}(x_1)\right). \tag{17}$$

This last equation shows that to compute the gradient of $g(x_1)$, one has only to solve the inner problem for x_2 and inject the solution in the gradient formula for $g(x_1)$.

WRI and IR-WRI

In van Leeuwen and Herrmann (2013), the nested loop optimization is employed with

$$x_1 = m, \quad x_2 = u. \tag{18}$$

The inner loop corresponds to the reconstruction of the wavefield u, by solving the problems

$$\min_{u_s} \frac{1}{2}\|Ru_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta \left\|F(m, u_s) - b_s\right\|_{\mathcal{W}}^2, \quad s = 1, \ldots, N_s. \tag{19}$$

Thanks to the bilinearity of the wave propagation operator, this problem is quadratic and a closed-form formula for u_s exists:

$$\left(\eta A(m)^T A(m) + R^T R\right) u_s = R^T d_{obs,s} + \eta A(m)^T b_s \tag{20}$$

Interestingly, the bilinearity of $F(m,u)$ makes also the outer minimization problem quadratic with respect to m, making possible to use a Newton method to solve the outer loop in a single step.

$$\left(\sum_{s=1}^{N_s} B(u_s)^T B(u_s)\right) m = \sum_{s=1}^{N_s} B(u_s)^T b_s \tag{21}$$

Another level of iteration further consists in reducing the weight η step by step. This iterative reduction of the weight can be difficult to adjust for practical applications.

IR-WRI circumvents this difficulty. It relies on a more sophisticated optimization scheme (ADMM method, see Boyd and Vandenberghe (2004); Combettes and Pesquet (2011) for instance), where such reduction of the weight "by hand" is not required. However, the core of the iteration is based on the same alternate reconstruction of the wavefield and the model. The same equations are solved, only with different right-hand-sides. For more details, the reader is referred to Aghamiry et al. (2019b).

MSWI

MSWI also relies on an alternate reconstruction between the extended source and the model parameters, with this time

$$x_1 = m, \quad x_2 = \widetilde{b} \tag{22}$$

As for WRI, the inner loop on $\widetilde{b}$ is equivalent to the following quadratic problems

$$\min_{\widetilde{b}_s} \frac{1}{2}\|S(m)\widetilde{b}_s - d_{obs,s}\|_{\mathcal{D}}^2 + \eta\|\widetilde{b}_s - b_s\|_{\mathcal{W}}^2, \quad s = 1,\ldots,N_s. \tag{23}$$

Therefore, closed-form formula exist for $\widetilde{b}_s$ such that

$$\widetilde{b}_s = \left[S(m)^T S(m) + \eta I\right]^{-1} \left(S(m)^T d_{obs,s} + b_s\right), \quad s = 1, \ldots, N_s. \tag{24}$$

Unlike WRI, the outer minimization problem is not quadratic with respect to m, therefore it should rely on a gradient-based algorithm (*i.e.* quasi-Newton methods). The gradient of the outer loop is computed following the adjoint state strategy (Plessix, 2006), as for conventional FWI. It is built as the correlation between incident and adjoint fields, where the adjoint is the backpropagation of the residuals at the receiver location. The difference is that the incident field and the residuals are computed using the extended source $\widetilde{b}$.

Extension to time-domain FWI

It can be shown that the operator $B(u_s)^T B(u_s)$ in equation 21 is diagonal for the acoustic wave equation (van Leeuwen and Herrmann, 2013; Aghamiry et al., 2019b), and block diagonal for general elastodynamics equations. In time-domain, each element of the diagonal blocks is accumulated by summation in time. The system in equation 21 therefore does not present particular difficulties for time-domain formulation.

However, this is not the case for the system in equation 20. The latter implies the operator $\eta A(m)^T A(m) + R^T R$. In the frequency-domain, $A(m)$ is a matrix after spatial discretization. It can be decomposed as a LU product and the system in equation 20 can be easily solved. In the time-domain, such technique is not available and solving the corresponding system is a real challenge. The difficulty actually comes from the component $R^T R$ in the operator which makes impossible the use of explicit time-domains schemes required for time-domain FWI. Neglecting $R^T R$ indeed yields the operator $A(m)^T A(m)$ which can be solved in two

steps through explicit time-domain schemes. Consider for instance, for a given right hand side z,

$$A^T A u = z \tag{25}$$

This can be solved using

$$A^T y = z, \;\; y = Au \tag{26}$$

The computation of y would require the solution of the adjoint wave equation with the right-hand-side z, and the computation of u the solution of the wave equation with the right-hand-side y. However, neglecting $R^T R$ amounts to an infinite weight η which goes back to solving the wave equation with infinite accuracy, *i.e.* the reduced space approach.

The same problem arises for MSWI. The reconstruction of the extended source implies the operator $S(m)^T S(m) + \eta I$. For the same reason mentioned above, this operator cannot be solved straightforwardly using explicit time-domain schemes because of the term ηI. Without it, the operator $S(m)^T S(m)$ can be solved through explicit time schemes in two steps, as in the WRI case. Circumventing this difficulty could imply giving an infinite weight to η: in this case MSWI also comes back to the reduced space approach as the extended source needs to conform with infinite accuracy to the true source b_s. Another option would be to make η tends to 0. However, this implies no regularization term in the MSWI problem 11, which is known to be an ill-posed problem because of the ambiguity between extended sources $\widetilde{b}$ and the model parameter m (Huang et al., 2018a).

Recent work proposed by Aghamiry et al. (2020) in the frame of WRI shows that an accurate reconstruction of the time-domain wavefields is however possible following a sophisticated backward-forward recursion where each iteration requires the solution of a wave propagation problem. The number of required iterations should be larger at the beginning of

the inversion. However preliminary result shows a computational extra cost approximately 8 times the cost of a gradient in the early stages of the inversion, which questions the feasibility of this strategy for field data application. The study by Aghamiry et al. (2020) also shows that the time-domain extension proposed in Wang et al. (2016) relies on a very crude approximation of the wavefield reconstruction step. Hence, the interest of the WRI approach tends to be lost following this time-domain approximation.

The difficulty of applying WRI, IR-WRI or MSWI in the frame of time-domain FWI has prompted us to investigate the alternative approach based on a receiver extension strategy we present in the next Section.

RECEIVER EXTENSION STRATEGY

Theory

In the same spirit as extended method strategies, we add an artificial degree of freedom to help fit the data when the subsurface parameter m is too far from the exact model. The difference is that this artificial degree of freedom is introduced at the receiver level, instead of being introduced at the source level.

The degree of freedom we introduce is the receiver position. As illustrated in the sequel, moving the receiver away from its true position can compensate for kinematic mismatch due to wrong subsurface model m. Formally, denote by $x_r, \;\; r = 1, \ldots, N_r$ the N_r receiver positions. Denote by $\Delta x_s \in \mathbb{R}^{N_r}$ a vector of N_r receiver corrections for receiver associated with source s, and $\Delta x = [\Delta x_1, \ldots, \Delta x_{N_s}] \in \mathbb{R}^{N_r \times N_s}$ the vector gathering the receiver position correction for each source/receiver pair. The receiver extension strategy consists in solving the problem

$$\min_{m, \Delta x} f(m, \Delta x) = \frac{1}{2} \sum_{s=1}^{N_s} \| R(\Delta x_s) A(m)^{-1} b_s - d_{obs,s} \|_{\mathcal{D}}^2 + \frac{1}{2} \| \Delta x \|_{\eta}^2, \tag{27}$$

where $\|.\|_\eta$ is a weighted least-squares norm

$$\| \Delta x \|_{\eta}^2 = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} \eta_{s,r} \Delta x_{s,r}^2, \tag{28}$$

with $\eta \in \mathbb{R}^{N_r \times N_s}$ a vector of weights $\eta_{s,r}$ (one per source/receiver couple), and $R(\Delta x_s)$ an extraction operator returning the values of the wavefield at the corrected receiver position

$x_r + \Delta x_{s,r}$, following the convolution

$$R(\Delta x_s)u = \int_\Omega u(x,t)\delta(x - (x_r + \Delta x_{s,r}))dx. \tag{29}$$

The second term in the right-hand-side of equation 27 is a least-squares annihilator, specifying that the receiver position correction should not be too large and converge to 0 for the correct model m. Note how close problem 27 is from reduced space problem 6. The only difference is in the receiver position correction introduction as a variable and the annihilator terms associated with this correction.

Numerical solution and implementation

Inner loop

To solve the problem 27, we use the nested optimization approach (equations 13 to 17) shared by WRI, IR-WRI, and MSWI techniques, with

$$x_1 = m, \;\; x_2 = \Delta x. \tag{30}$$

The inner loop problem thus consists in determining the receiver position correction for a given model m. We denote it by $\overline{\Delta x}(m)$. The key point for an efficient implementation is a fast solution of this inner problem. When using WRI, IR-WRI, or MSWI techniques, the inner problem is quadratic: it has a unique solution given by a closed-form formula. Using the receiver-extension strategy, the inner problem is highly non-linear. Thus, there is no closed-form formula for $\overline{\Delta x}(m)$. In addition, the associated misfit function presents local minima, condemning the use of local optimization methods. However, for the nested loop

optimization to be efficient, we need a fast and accurate solver for the solution of the inner problem.

Here, it is important to realize that, thanks to the use of L^2 norm both for data misfit and annihilator terms, the inner problem is separable for all source/receiver couples. It means the objective function in 27 can be decomposed as a sum of misfit function depending only on one source/receiver couple. Mathematically, we have

$$f(m, \Delta x) = \sum_{s=1}^{N_s} \sum_{r=1}^{N_r} f_{s,r}\left(m, \Delta x_{s,r}\right) \tag{31}$$

where

$$f_{s,r}\left(m, \Delta x_{s,r}\right) = \frac{1}{2} \int_0^T \left| u_s[m](x_r + \Delta x_{s,r}, t) - d_{obs,s}(x_r, t) \right|^2 dt + \frac{\eta_{s,r}}{2} |\Delta x_{s,r}|^2, \tag{32}$$

where $u_s[m] = A(m)^{-1} b_s$.

Hence, the solution of the inner loop can be obtained by solving independently for each receiver correction $\Delta x_{s,r}$ the subproblem

$$\min_{\Delta x_{s,r}} f_{s,r}\left(m, \Delta x_{s,r}\right) \tag{33}$$

The number of unknowns for each subproblem 33 is small: maximum 2 unknowns in 2D to and 3 unknowns in 3D to specify a receiver position correction. Global optimization methods can thus be employed to determine the optimal receiver position corrections $\overline{\Delta x}_{s,r}(m)$.

In practice, it is even possible and/or advisable to reduce this number of unknowns to a single parameter. For instance, in the 2D case, considering a seismic trace containing a single

event, there is an intrinsic ambiguity in the receiver correction making possible to fit the data. This ambiguity is related to the isochrones, which are 2D curves in the 2D approximation. This means that there would be an infinity of 2D receiver corrections (vertical and horizontal repositioning) yielding an equivalent data fit. To avoid this non-uniqueness, we consider in this study only horizontal repositioning. No vertical receiver position corrections are allowed. The additional benefit of this strategy is that the global optimization problems to be solved in the inner loop are single parameter problems.

In terms of implementation, we rely on a brute-force grid search approach. The misfit function in equation 33 is evaluated for different values of $\Delta x_{s,r}$ within bounds defined depending on the application. The time-history of the wavefield is stored on a line in 2D (or a plane in 3D) on which the receivers are confined. As explained above we restrict the receivers to move only laterally to avoid intrinsic ambiguity related to isochrones. From this stored time-history of the wavefield, the calculated data can be extracted at various receiver position without having to solve again the wave equation. For each receiver position, the misfit function is evaluated. We select the receiver position correction which provides the minimum misfit value. As we illustrate in the following, this provides an efficient method to solve the inner problem. In our 2D examples, the additional computational cost compared with conventional FWI is negligible.

Outer loop

We solve the outer problem by a conventional quasi-Newton strategy. We use a preconditioned l-BFGS method in this study (Nocedal, 1980). The gradient of the outer function can be computed, as in MSWI, following the adjoint source strategy (Plessix, 2006). We denote it

by $\nabla g(m)$, and in condensed form it can be expressed as

$$\nabla g(m) = \sum_{s=1}^{N_s} \left\langle \frac{\partial A}{\partial m} u_s, \lambda_s \right\rangle, \tag{34}$$

where $\langle ., . \rangle$ denotes the scalar product in time domain and

$$\begin{cases} A(m) u_s = b_s, \quad s = 1, \ldots, N_s \\ A(m)^T \lambda_s = R\left(\overline{\Delta x}_s(m)\right)^T \left(R\left(\overline{\Delta x}_s(m)\right) u_s - d_{obs,s}\right), \quad s = 1, \ldots, N_s. \end{cases} \tag{35}$$

The difference with the conventional reduced space approach is that the calculated data and the adjoint wavefields are computed using corrected receiver positions, both for the extraction of the wavefield values to build the calculated data with the operator $R\left(\overline{\Delta x}_s(m)\right)$ and the injection of the adjoint source with the operator $R\left(\overline{\Delta x}_s(m)\right)^T$. Using $\overline{\Delta x}_s(m) = 0$ in the previous equations yields the conventional least-squares gradient for FWI based on the reduced space approach.

Weight parameters $\eta_{s,r}$

The weights $\eta_{s,r}$ are computed following

$$\eta_{s,r} = \alpha \frac{\|d_{obs,s,r}\|_\infty}{L} \tag{36}$$

where

$$\|d_{obs,s,r}\|_\infty = \max_{t \in [0,T]} |d_{obs,s,r}(t)|, \tag{37}$$

while L is the maximum value we allow for $|\Delta x_{s,r}|$. The parameter α is a tuning parameter to control the constraint on the receiver position correction. The choice $\alpha = 1$ corresponds

to a simple dimensioning of the two terms in the misfit function (data fitting term and annihilator term). In the next numerical experiments, the sensitivity of the method to the choice of α is investigated. In synthetic experiments with inverse crime settings, low values of α (to the order of 10^{-2}) seem to yield satisfactory results (transmission case, Marmousi and BP2004 studies). When the amplitude cannot be predicted with perfect accuracy, higher values of α might be better adapted (Valhall case study).

A TRANSMISSION CASE ANALYSIS

We consider here a canonical transmission problem. We use a 2D cross-hole configuration, with two wells located 50 m apart (Fig. 1).

[Figure 1 about here.]

The source is a Ricker pulse with 250 Hz central frequency. We consider a single source/receiver couple at 50 m depth in each well. The source is in the left well, the receiver in the right well. We compute a reference seismic trace in a homogeneous medium at 2000 m.s^{-1}. We use for that a 2D constant density acoustic wave propagation model. Using this reference trace, we construct the misfit function $f(m, \Delta x)$ considering homogeneous velocity models m varying from 1000 m.s^{-1} to 3000 m.s^{-1}, and receiver position correction Δx varying only horizontally (following the x axis) from -37.5 m to 37.5 m. We select the weight α to be equal to 1. The resulting misfit function is presented in Figure 2.

[Figure 2 about here.]

We see that the misfit function $f(m, \Delta x)$ is not convex. Its minimum is hidden in a narrow valley, at position $m = 2000$ m.s^{-1} and $\Delta x = 0$, and surrounded by large barriers. The shape of the valley is driven by the shape of the Ricker function used to build the data: the lower frequency used, the wider the valley of attraction is.

Nevertheless, if we select, for each velocity value, the minimum reached in the receiver extension direction Δx, we can represent the function $g(m)$ that we aim at minimizing in the outer loop. This function is presented in Figure 3 for different values of the weight α.

[Figure 3 about here.]

We see in Figure 3 that for a proper selection of the weight α, a convex function depending on the velocity m can be obtained. The choice of α influences the size of the valley of attraction toward the global minimum.

To better understand why the receiver extension approach can yield a convex misfit function in this simple transmission case, we present in Figure 4 the synthetic traces computed for different values of velocity before and after the relocalization, and we compare it to the reference trace.

[Figure 4 about here.]

As expected, the relocalization of the receiver corrects for the kinematic mismatch. The relocalized synthetic traces are all in phases with the reference trace. However, the relocalization cannot compensate for the amplitude mismatch. This amplitude mismatch is related to energy conservation rules of wave propagation: the amplitude of the recorded signal depends on the rigidity of the medium in which it propagates, hence on the velocity in the simple constant density acoustic approximation we use here.

The result of this is that in the context of this single arrival canonical case, the misfit measured by the function $g(m)$ is related to this amplitude mismatch only. This mismatch increases with the velocity mismatch between the reference medium and the synthetic medium. Hence, the misfit function $g(m)$ is convex with respect to the velocity in this case. Note that the use of an amplitude sensitive misfit function, such as the least-square norm, to define $g(m)$, is crucial. A receiver extension approach based on a misfit function not sensitive to amplitude mismatch would not yield a convex function in this canonical case.

To end with this simple transmission case, we analyze the shape of the gradient using the same single source/receiver couple. We compare the conventional least-squares gradient

and the receiver extension gradient in two different homogeneous media: one at 1500 m.s^{-1}, the second at 2500 m.s^{-1}. The results are presented in Figures 5 and 6. In the least-squares case, the first Fresnel zone of the two kernels is negative, while we expect a change of sign: in one case the medium is slower than the reference one, in the other case it is faster. This is a clear indication of cycle skipping: starting from the faster medium, the least-squares gradient would produce a positive update (opposite of the gradient) of the velocity within the first Fresnel zone. Converging to the correct solution would require to slow down the velocity.

The receiver extension approach does not suffer from such inconsistencies. The sign of the first Fresnel zone is correct in both slower and faster media. This change of sign is directly related to the corresponding adjoint source. Let us remind that in the receiver extension approach, it is computed as the difference between observed and synthetic data *after* relocalization. As can be seen in Figure 6, the difference between observed and synthetic data changes of sign, depending on the velocity is faster or slower than the reference one, for the same reason as mentioned previously (energy conservation law). This explains the differences in the two kernels. Note also that the relocalization affects the shape of the kernels. In the slow medium, the receiver is relocalized closer from the source. As the adjoint source is injected at the corrected receiver position, the size of the kernel is smaller. In the faster medium, the receiver is relocalized farther from the source, and the size of the kernel is larger.

[Figure 5 about here.]

[Figure 6 about here.]

This simple experiment illustrates how the receiver extension approach can handle kinematic mismatch in the frame of FWI.

2D SYNTHETIC EXAMPLES

Choice of three models

We investigate the performance of the receiver relocalization approach on three benchmark models: Marmousi II (Martin et al., 2006), BP 2004 (Billette and Brandsberg-Dahl, 2004) and a 2D synthetic model built from 3D inversion results of the Valhall OBC data (North Sea) (Sirgue et al., 2010; Operto et al., 2015; Amestoy et al., 2016). We choose these three models to test the method in different geological contexts. Marmousi II is a useful framework to investigate in details the ability to mitigate cycle skipping issues. BP 2004 is representative of the gulf of Mexico geology and contains salt structures known to be challenging to reconstruct for seismic imaging methods, because of the high velocity contrasts between these structures and the surrounding water. The Valhall model contains an important gas cloud in its middle part, which significantly attenuates seismic wave energy and makes it difficult to image the reservoir located below.

Common framework

The three experiments we present in this Section are performed in the 2D (visco-) acoustic approximation. They rely on our 2D/3D (visco-)acoustic time-domain finite-difference based full waveform inversion code TOYxDAC_TIME, which implements the method described in Yang et al. (2018a). All are based on the reconstruction of the P-wave velocity $v_P(x)$.

The source which is used is a Ricker wavelet centered on 5 Hz and high pass filtered to remove all energy below 2.5 Hz. In the BP 2004 case study, we use an additional low-pass filter to remove energy above 8 Hz. The corresponding wavelets and their power spectrum are presented in Figure 7.

[Figure 7 about here.]

In all cases we use the bound constraint preconditioned l-BFGS solver from the SEISCOPE toolbox (Métivier and Brossier, 2016). The preconditioner chosen is either a simple linear scaling in depth for the Marmousi II case study, or a wavefield based pseudo-Hessian preconditioner (Choi and Shin, 2008) for BP 2004 and Valhall case studies. We also use a Gaussian smoothing of the gradient, with correlation lengths associated with the estimated local wavelength

$$\lambda(x) = \frac{v_P(x)}{f_{ref}} \tag{38}$$

where f_{ref} corresponds to the central frequency of the Ricker wavelet.

In all three experiments, a free surface condition is imposed on top of the model. Perfectly matched layers (PML) (Bérenger, 1994) (for Marmousi II and BP 2004 models) or sponge layers (Cerjan et al., 1985) (for the Valhall model) are applied on the other boundaries to mimic a medium of infinite extension.

While Marmousi II and BP 2004 experiments are performed in an "inverse crime" settings, using a constant density acoustic modeling, the Valhall case study intends to mimic a more realistic framework. In this case the observed data is computed using a variable density and variable quality factor under the visco-acoustic approximation. A Gaussian noise, filtered in the frequency band of the data, is added, with a signal to noise ratio (SNR) equal to 10. The mesh used to compute the observed data is finer than the inversion mesh.

For the relocalization strategy, the selection of the parameter α is discussed for each experiments. Regarding the choice of the parameter L (maximum absolute value for the receiver shifts $\Delta x_{s,r}$, see equation 36), we select it equal to the surface length for the

Marmousi II case (the receivers are allowed to be relocalized on the whole surface), while we take it equal to half of the surface length for BP 2004 and Valhall experiments. These rather unrestrictive choices yield meaningful results, and it seems not necessary at this stage to adapt L along the iterations.

Marmousi II

We use the Marmousi II P-wave velocity model introduced in (Martin et al., 2006), which is 3.5 km deep and 17 km long. We generate observed data using a fixed spread surface acquisition with 128 sources and 170 receivers. The source and receiver spacing is 125 m and 100 m respectively. We use a 25 m discretization mesh.

We investigate how the receiver relocalization approach can help mitigate the sensitivity to the initial model design. To this purpose we define four initial models, increasingly far from the exact model. Initial model 1,2 and 3 are obtained by applying a 2D Gaussian smoothing to the exact model, with correlation lengths equal to 1 km, 2 km, and 4 km in both horizontal and vertical directions respectively. Initial model 4 is a simple 1D linearly increasing model from the water bottom at 1500 m.s^{-1} to the bottom of the model at 4000 m.s^{-1}. For all initial models the correct water layer (same as exact model) is appended on top of the model. The exact and initial models are presented in Figure 8.

[Figure 8 about here.]

We compare inversion results obtained using a conventional L^2 FWI and the receiver relocalization approach starting from these 4 models in Figure 9 and 10. In this first experiment, α is set to 5×10^{-2}. Starting from model 1, both methods reconstruct satisfactory

estimates of the true model. Note that the receiver relocalization approach corrects the up-bending of the bottom left part observed in the L^2 reconstruction (around 2.5 km depth between $x = 0$ and $x = 4$ km). Starting from model 2, L^2 reconstruction starts introducing artifacts in the left part of the model, plus a low velocity anomaly at $x = 6$ km, $z = 2.5$ km. The receiver relocalization approach is more stable: there is no such artifacts, and the low velocity anomaly appears further from the center of the model (x=3 km, z=2.5 km). Starting from model 3 and 4, the L^2 reconstructions are not meaningful anymore. The receiver relocalization approach is more stable, preserving a correct estimate of the true model in the zone of main illumination (down to 3 km depth and between $x = 2$ and $x = 15$ km approximately).

[Figure 9 about here.]

This is confirmed by the analysis of the data fit presented in Figure 10. We overlay the exact left shot gather in red/blue color with the final shot gather in black and white in the different estimated models. While we observe a degradation of the data fit using the conventional L^2 approach, we see that the receiver relocalization approach is able to maintain a similar level of data-fit starting from the 4 different initial models.

[Figure 10 about here.]

One interest for working with synthetic models is the ability to quantify the model error. We use here a relative L^1 model misfit measure. For a given v_P model, discretized on a M points mesh, it is computed as

$$E_{v_P} = \frac{100}{M} \sum_{i=1}^{M} \frac{\left| v_{P,i} - v_{P,i}^{true} \right|}{\left| v_{P,i}^{true} \right|} \tag{39}$$

where v_P^{true} is the true P-wave velocity model. In Figure 11a we compare the decrease of the misfit function along the inversion iterations for both L^2 and receiver relocalization approaches, starting from the four initial models. The same plot for the model error is presented in Figure 11b. Finally, we present the model error decrease with respect with the misfit function decrease in Figure 11c. Interestingly, we see that the receiver relocalization approach provides a systematic lower model misfit error, even in the case where there is no cycle skipping and L^2 FWI works well. Starting from initial models 3 and 4, the receiver relocalization approach is able to decrease the model error, which is not the case for conventional L^2 FWI. Except for initial model 4, receiver relocalization always provides a monotonic decrease of the model error with respect to the misfit function (which is the expected behavior for a stable inversion). In case of initial model 4, there is an initial phase where the model error increases before decreasing, which corresponds to the first iterations of the process. Remember that initial model 4 is a vertically increasing model, therefore significantly far from the exact model.

[Figure 11 about here.]

To foster the analysis of the receiver relocalization strategy in itself, we present in Figure 12 the evolution through iterations of the relocalization error for the leftmost, central, and rightmost shot gathers, depending on the choice of initial model. This error, for a given shot gather s, corresponds to the quantity

$$E_{\Delta x} = \frac{1}{N_r}\sqrt{\sum_{r=1}^{N_r} |\Delta x_{s,r}|^2}. \tag{40}$$

This is an average over all the receivers of the relocalization error Δx for the shot s. We see

that this error tends to 0 along the iteration process. The speed of convergence depends on the initial model and on the shot gathers. For the central shot gathers, the convergence is much faster than for the leftmost and rightmost ones. For initial model 1 (easiest one) the convergence is also attained faster. For initial models 2 and 3, the speed of convergence is comparable. The values of the average relocalization are higher for initial model 3. The model being further from the exact one, stronger kinematic effects need to be accounted for through the relocalization process. This is even more visible for initial model 4. For this model, the convergence is the slowest, as well as the value of the mean relocalization error. As expected, stronger kinematic mismatch thus results in a higher compensation through relocalization of receivers.

[Figure 12 about here.]

A more qualitative visualization of the relocalization process is proposed in Figures 13, 14 and 15. In these figures, we present the leftmost shot gather data-fit before and after the relocalization, in P-wave velocity models obtained at iteration 0 (Fig. 13), iteration 100 (Fig. 14) and in the final model (Fig. 15). We have selected the experiment starting from the initial model 4 (1D linearly increasing model) for these Figures. The models with the receiver position represented as yellow ellipses are appended to the data. The effect of the relocalization step on the data-fit is strong: the receiver repositioning makes possible to compensate for the cycle-skipped diving wave visible on the left panel of Figure 13. Some events are not correctly matched: in particular we can see that the part of the diving waves in the synthetic data arriving at offset between 3 and 7 km are matched with strong reflections in the observed data. However, at further offset, the match seems better, and it certainly helps the method to mitigate this strong cycle skipping effect. We can also link this incorrect

initial matching to the rather slow convergence of the process in the initial iterations when starting from initial model 4. However, at iteration 100 (Fig. 14), the data fit is already much better, and we can see the same effect of the relocalization which compensates for the too fast diving wave. In the final model, the data fit is already very good, with much of the events in phase and correctly predicted. Therefore, as expected, the relocalization has very little effect on the data fit in the final stage of the iterations.

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

Finally, we analyze in Figure 16 the sensitivity of the relocalization error with respect to the choice of the regularization parameter α (equation 36). This parameter controls the weight on the annihilator term, which restrains the receivers from moving to far away from their true position. We present the model error evolution along the iteration of the inversion process. We vary α between 0.01 and 0.1 with 0.01 increment. Interestingly, we see that the model error follows the same trend for any of these parameters, with relatively few variations. This is encouraging toward a robust behavior of the receiver relocalization method regarding the tuning parameter α.

[Figure 16 about here.]

BP 2004

We use a rescaled version of the original BP 2004 model (rescaling by a factor 2), and focus on the left part of the model where the high velocity salt structures are the more complex.

The exact model we consider is almost 6 km deep and 16.2 km long (Fig.17a). We use a fixed spread acquisition with 128 sources and 161 receivers at 50 m depth in the water layer, from $x = 0.1$ km to $x = 16.1$ km. The source and receiver spacing is 125 m and 100 m respectively. To design the initial model, we first remove the salt from the exact model. We then smooth the resulting background model. The resulting initial model is presented in Figure 17b.

[Figure 17 about here.]

The leftmost shot gather is presented in Figure 18. The salt structure, especially the canyon structure at $x = 2$ km, generates energetic first order (red arrow) and higher order reflections (orange arrows), also with interactions with the free surface at $z = 0$ km. The blue arrows depict the refraction of the direct by the salt body. The event depicted by the green arrows corresponds to the transmission of the direct wave within the salt structure. Black arrows depict arrivals coming from below the salt after interacting with the canyon. Correctly matching the events depicted by the red, blue and green arrows is crucial to recover correctly the salt structure, especially starting from the model in Figure 17b.

[Figure 18 about here.]

To mitigate the complexity of the data, we use a time-windowing approach similar to the one we designed in Métivier et al. (2016). The inversion is decomposed in 7 time windows of increasing lengths: 6.9 s, 9.2 s, 10.35 s, 11.5 s, 12.65 s, 13.8 s and 14.95 s. We use such a long recording time to investigate the ability to reconstruct the subsalt velocity. Exploiting late events, which have traveled below the salt might help achieving this reconstruction.

Subsalt imaging is a knowledgeable challenge. As for the previous experiment, we select a low value for α, such that $\alpha = 5 \times 10^{-2}$.

We compare the results obtained using the receiver relocalization approach and conventional L^2 FWI. The comparison is shown for the 1st, 2nd, and last time-window. The reconstructed models are presented in Figure 19. Interestingly, the receiver relocalization method provides satisfactory reconstruction of the main salt body, including the canyon zone around $x = 2$ km, already from the inversion of the two first time-windows. The final results, obtained after the inversion of all the time windows, show that the subsalt velocity in the zone between $x = 6$ km and $x = 10$ km is correctly reconstructed, down to 5 km depth. We note also that the whole right part of the model, with no salt structure on top, between $x = 10$ km and $x = 16$ km, is accurately reconstructed, down to 6 km depth. The subsalt zone between $x = 0$ km and $x = 6$ km remains difficult to image.

Comparatively, results achieved using a conventional L^2 FWI are much less satisfactory. Inverting for the first time-window only yields the top-salt structure. The whole salt structure is reconstructed only after the last stage of inversion, with still a visibly incorrect recovery of the canyon structure on the left. The whole subsalt target is not correctly imaged either.

[Figure 19 about here.]

To interpret these results, we present the data fit in the final models in Figure 20. The true data in blue/red color is superposed with the synthetic data in black and white color. In the correct data fit, no black and white events should appear. The L^2 data fit is correct for the refracted and transmitted events depicted by blue and green arrows. However, the short offset reflections depicted by the red arrow, and multiples of these reflections (orange arrows) are not correctly matched. This is consistent with the incorrect geometry of the canyon

structure within the salt body which is recovered using the L^2 approach. Later arrivals coming from under the salt (black arrows) are also not correctly matched. On the contrary, the data fit achieved following the receiver relocalization strategy is more satisfactory. All the events depicted by the colored arrows are correctly matched. Even relatively late events ($t > 10$ s) are matched, which is consistent with the correct reconstruction of the subsalt part of the model. This experiment thus shows that the receiver relocalization strategy could be useful in the specific context of salt and subsalt imaging. The degree of freedom introduced on the receiver position level helps matching out of phase events, associated with complex paths within and below the salt structure, which cause strong artifacts in a conventional L^2 reconstruction. By progressively relaxing the receiver position towards their physical position, the receiver relocalization strategy makes it possible to improve the velocity model to match all these events and recover the correct geometry of the salt body, as well as information on the subsalt region.

[Figure 20 about here.]

Valhall

We end up this series of experiment with the synthetic Valhall case study. Here the model is representative of the North Sea geology, with shallow water, horizontally stratified structure, and gas bearing sediments. A layered gas cloud is located above a strong reflector with an anticlinal structure. The oil reservoir is located below. The presence of gas induces a rather strong attenuation effect (amplitude decrease and dispersion), which makes the reservoir imaging challenging. The exact P-wave velocity, density and quality factor model used to generate the data are presented in Figure 21. In the modeling, the quality factor

is considered independent of the frequency within the frequency band considered, which is approximately 2.5 - 15 Hz. This is enforced through the use of 3 standard linear solid (SLS) mechanisms (Yang et al., 2018a). The Valhall field is one of the first exploration scale target on which FWI has been applied successfully, yielding unprecedented high resolution images of the subsurface (Sirgue et al., 2010). Since then the Valhall data has served for testing different FWI methodologies including frequency-domain multiparameter FWI and time-domain visco-acoustic FWI (Operto et al., 2015; Operto and Miniussi, 2018; Kamath et al., 2021).

[Figure 21 about here.]

The initial models we consider are presented in Figure 22. The initial P-wave velocity model is obtained through a strong Gaussian smoothing of the exact model, with correlation lengths equal to 4 km. The initial density model is derived from a Gardner's law from this initial P-wave velocity model

$$\rho(x) = 1741 \left(10^{-3} v_P(x)\right)^{0.25}, \tag{41}$$

with $\rho = 1000$ kg.m^{-3} in the water layer. The initial quality factor model is built by setting its value to 1000 in the water layer and 100 below. During the inversion, these initial density and quality factor models are kept unchanged (passive parameters).

[Figure 22 about here.]

As a first step, we estimate the source wavelet in this initial model, following the frequency-domain deconvolution of Pratt (1999). We assume here the same wavelet for all shots. The

resulting estimated source wavelet is presented in Figure 23, where it is compared with the true source wavelet. Both time signature and amplitude spectrum are presented. We see that despite the noise and the inaccurate starting velocity and density models, the estimated wavelet remains relatively close to the true one. Differences are however visible in the normalized amplitude spectrum.

[Figure 23 about here.]

The P-wave velocity models obtained using L^2 FWI and the receiver relocalization approach are presented in Figure 24. As can be seen, the L^2 inversion fails to produce a meaningful estimate of the P-wave velocity model, except in the shallow part above 1 km depth. This part, sampled by diving and reflected waves, is relatively well reconstructed, except for the presence of high wavenumber artifacts around $x = 9$ km and $z = 0.8$ km. Below 1 km depth, a strong horizontally extended low velocity artifact is injected in the model reconstruction. The layered shape gas cloud below is not properly reconstructed. The continuity of the strong reflector at 2.5 km depth is broken, and its anticlinal shape is not reconstructed. All this indicates the convergence towards a non informative local minimum due to cycle skipping.

On the contrary, the P-wave velocity model obtained following the receiver relocalization approach is much closer to the exact model. The successive gas layers are properly reconstructed, as well as the main reflector at 2.5 km depth, which appears continuous, and with an anticlinal shape. Below, the medium is not sufficiently sampled by waves to make it possible to reconstruct it from the initial model which is used here. We can also note the presence of artifacts on the lateral edges of the model, which are also due to a lack of illumination in these part of the model. Low velocity V-shape artifacts also appear on both

sides of the gas cloud, which indicate still the presence of cycle skipped events. However, the overall estimation is correct down to 3 km depth.

[Figure 24 about here.]

These results indicate that the receiver relocalization approach is robust to relatively realistic settings where the amplitude of the data cannot be predicted to machine precision. This is comforting for perspectives of application to field data. To better understand the difference between the L^2 and receiver relocalization reconstruction, we compare the final data match using both approaches for the shot-gather associated with source position $x_S = 8$ km (Fig. 25). The superposition of exact (blue and red) and synthetic data (black and white) in the final model is intriguing: the L^2 data match seems relatively good, especially for diving waves. The receiver relocalization approach provides also a good data match, however less accurate regarding the larger offset arrivals. In Figure 26, we compare the normalized residuals computed between the observed data without noise, and the synthetic data in the final models provided by the two approaches. This comparison provides the explanation of the difference between the two reconstructed models. The L^2 approach is unable to correctly explain the short and medium offset reflections, associated with the gas cloud layers. Conversely, the receiver relocalization approach provides a model which explains significantly better these reflections, while increasing slightly the misfit with respect to largest offset diving waves.

The reason why the misfit related to these events remains large is that in the final model, the receiver position has still not converged towards the true position of the receivers. The average relocalization error for the shot considered here (x_S=8 km) indicates a systematic drift of 50 m even in the final model. This ambiguity shows that the weight associated to the

annihilator might not be not sufficiently high. We have tested different values of the weight α, however choosing a too large value prevents for adding sufficient freedom to the inversion in the early stage of the inversion to obtain a satisfactory reconstruction. The best results where achieved with $\alpha = 1000$, which is already a significantly higher value than what is used for Marmousi and BP 2004 case studies. This is in accordance with the presence of noise and the consequently higher value of the data matching term in the misfit function, which requires to strengthen the weight of the annihilator term in the relocalization approach.

[Figure 25 about here.]

[Figure 26 about here.]

Computational cost

We end this Section with a comparison of computational cost of the receiver relocalization approach for each case study. The results are presented in Table 1. For each case study, we provide the computational time for a gradient computation, and provide the extra computational time associated with the receiver relocation strategy. The reference time for the extra computational cost is the one which would be obtained with a L^2 approach. We see that in the three cases, the extra cost remains below 5% which makes the receiver relocalization strategy relatively inexpensive. Note also that the overall computational cost associated with the Valhall model (approximately the same size as the Marmousi model), is significantly higher: this is related to the visco-acoustic modeling.

[Table 1 about here.]

DISCUSSION

The three case studies investigated in the previous Section illustrate the interesting properties of the receiver relocalization approach. In all cases, the method makes possible to start from crude initial models while still providing meaningful velocity estimations. The method is applicable directly in the frame of time-domain FWI. In the 2D (visco-)acoustic approximation considered here, the extra computational cost compared to a conventional least-squares approach is negligible. Based on the separability of the least-squares misfit function, the inner loop complexity depends linearly on the number of source/receiver pairs. We have considered here fixed spread acquisition systems. The computational cost increase would be even lower for corresponding streamer acquisition with constant offset, which would induce less source/receiver pairs.

We discuss here practical aspect and potential extensions of the method. First, it is important to control the design of the initial model with respect to the initial step of receiver relocalization. A too fast model could require to relocate the receivers outside the computational box. It might thus be advisable to start with initial velocity models underestimating the true velocity, or to adapt the computational box to the initial receiver relocation step.

Second, we have used here a grid search algorithm for the solution of the inner loop problem. Other possibilities could be considered if it becomes necessary to reduce the computational cost. Markov-Chain Monte-Carlo method could be used instead, in particular its recent Hamiltonian accelerated variant (Neal et al., 2011).

Third, we have observed that, for a given source s, the receiver relocalization $\Delta x_{r,s}$ can change rather abruptly for neighboring traces r_k, r_{k+1}. It is possible that these rapid

changes slow down the convergence of the whole method, which is observed for instance on the Marmousi II experiment where several hundreds iterations are required to converge. For this reason, it might be advisable to add a regularization term and/or constraints in the misfit function to promote smoother variations. This could be done by penalizing the discrete difference of the corrections between two traces for a given source, or by smoothing directly in the receiver direction the receiver correction vector $\Delta x_{r,s}$ solution of the inner loop.

Fourth, the convexity analysis and the resulting implementation performed in this study is done in the frame of single arrival traces. While the three case studies of the preceding Section illustrate that the method works in the frame of complex multi-arrival data, it might still be interesting to extend the analysis and implementation of the method to the case of multi-arrival traces. In this frame, the receiver position correction which we consider could depend on time. For a workable method, time windows should be defined prior to the application of the method, and a receiver position correction could be computed for each time-window. This could be interesting for instance to avoid mismatch of events (diving interpreted as strong reflections) in the initial iterations of the process.

Finally, we discuss the application of the receiver relocalization method in a 3D context. In 3D, isochrones are surfaces. Therefore, even if we restrict the receiver relocalization to the surface (forbidding vertical relocalization), an ambiguity would subsist in the case of single event traces. Again, this ambiguity can be prevented by restricting the receiver relocalization correction to a single parameter, which could be in this case a surface repositioning r in the direction of the source/receiver axis. The correction would thus not be aligned with horizontal axis x and y. The additional benefit would be again to obtain inner loop problems depending on a single degree of freedom, making the solution through global optimization

almost negligible. Thus it seems 3D extension of the method might be feasible.

CONCLUSION

We propose in this study a receiver relocalization method as a novel extension strategy, which is directly applicable to time-domain FWI. The receiver position is introduced as a degree of freedom in the FWI problem, which makes it possible to reduce the kinematic mismatch which would lead conventional least-squares FWI to converge a local minimum. Doing so, the data is fit progressively by the subsurface model as receivers converge towards their true positions. The method is implemented similarly as source extension strategies, using a variable projection approach, with an inner loop dedicated to the computation of the optimal receiver position and an outer loop dedicated to the subsurface model update. Our implementation solves the inner loop problem using a brute force grid search approach. The outer loop problem is solved using a conventional quasi-Newton l-BFGS approach.

We illustrate the properties of this receiver relocalization strategy first on a schematic cross-hole experiment, exhibiting the robustness of the approach with respect to strong kinematic mismatch and its resilience with respect to cycle skipping. Then we investigate three synthetic case studies, representative of different geological context. In all three cases, the receiver relocalization strategy is shown to successfully converge toward a correct estimation of the subsurface model starting from crude initial models, with a relatively inexpensive additional computational cost (no more than 5% more expensive).

The good results obtained, in particular in the Valhall case, where noise, inexact source wavelet, inexact density and attenuation models, make it not possible to predict the data amplitude with arbitrary precision, are encouraging towards application to field data. Finally, compared to misfit modification approaches based on optimal transport distances, which we have recently studied, our experiments indicate that the receiver relocalization approach

appears as a competitive alternative. The computational cost increase is of the same order or even lower for these 2D experiments, and the robustness to cycle skipping seems also comparable. Future studies will include comparisons between these different approaches and applications to 3D field data.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to William Symes for insightful discussions on source-extension strategies. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXON-MOBIL, JGI, SHELL, SINOPEC, SISPROBE and TOTAL. This study was granted access to the HPC resources of the Froggy platform of the CIMENT infrastructure (*https://ciment.ujf-grenoble.fr*), which is supported by the Rhône-Alpes region (GRANT CPER07_13 CIRA), the OSUG@2020 labex (reference ANR10 LABX56) and the Equip@Meso project (reference ANR-10-EQPX-29-01) of the programme Investissements d'Avenir supervised by the Agence Nationale pour la Recherche, and the HPC resources of CINES/IDRIS/TGCC under the allocation 046091 made by GENCI."

REFERENCES

Aghamiry, H., A. Gholami, and S. Operto, 2019a, Admm-based multi-parameter wavefield reconstruction inversion in VTI acoustic media with TV regularization: Geophysical Journal International, **219**, 1316–1333.

———, 2019b, Improving full-waveform inversion by wavefield reconstruction with alternating direction method of multipliers: Geophysics, **84(1)**, R139–R162.

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

Almomin, A., 2016, Tomographic full waveform inversion: PhD thesis, Stanford University.

Amestoy, P., R. Brossier, A. Buttari, J.-Y. L'Excellent, T. Mary, L. Métivier, A. Miniussi, and S. Operto, 2016, Fast 3D frequency-domain full waveform inversion with a parallel Block Low-Rank multifrontal direct solver: application to OBC data from the North Sea: Geophysics, **81**, R363 – R383.

Beller, S., V. Monteiller, L. Combe, S. Operto, and G. Nolet, 2018, On the sensitivity of teleseismic full waveform inversion to earth parametrisation, initial model and acquisition design: Geophysical Journal International, **212(2)**, 1344–1368.

Bérenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200.

Billette, F. J., and S. Brandsberg-Dahl, 2004, The 2004 BP velocity benchmark: Extended Abstracts, 67^{th} Annual EAGE Conference & Exhibition, Madrid, Spain, B035.

Boyd, S. P., and L. Vandenberghe, 2004, Convex optimization: Cambridge university press.

Bozdağ, E., D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire, 2016, Global adjoint tomography: first-generation model: Geophysical Journal International, **207**, 1739–1766.

870 Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inver-
871 sion based on instantaneous phase and envelope measurements: Geophysical Journal
872 International, **185**, 845–870.
873 Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore struc-
874 tures by 2D elastic frequency-domain full-waveform inversion: Geophysics, **74**, WCC105–
875 WCC118.
876 Bunks, C., F. M. Salek, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform
877 inversion: Geophysics, **60**, 1457–1473.
878 Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition
879 for discrete acoustic and elastic wave equations: Geophysics, **50**, 2117–2131.
880 Choi, Y., and C. Shin, 2008, Frequency-Domain Elastic Full Waveform Inversion Using the
881 New Pseudo-Hessian Matrix: Experience Of Elastic Marmousi 2 Synthetic Data: Bulletin
882 of the Seismological Society of America, **98**, 2402–2415.
883 Cocher, E., H. Chauris, and R. Plessix, 2017, Seismic Iterative Migration Velocity Analysis:
884 two strategies to update the velocity model: Computational Geosciences, **21**, 759–780.
885 Combettes, P. L., and J.-C. Pesquet, 2011, Proximal splitting methods in signal processing,
886 *in* Fixed-Point Algorithms for Inverse Problems in Science and Engineering: Springer New
887 York, volume **49** *of* Springer Optimization and Its Applications, 185–212.
888 Donno, D., H. Chauris, and H. Calandra, 2013, Estimating the background velocity model
889 with the normalized integration method: EAGE Technical Program Expanded Abstracts
890 2013, Tu0704.
891 Engquist, B., and B. D. Froese, 2014, Application of the Wasserstein metric to seismic
892 signals: Communications in Mathematical Science, **12**, 979–988.
893 Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background

for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, **175**, 665–685.

Fichtner, A., and A. Villaseñor, 2015, Crust and upper mantle of the western mediterranean – constraints from full-waveform inversion: Earth and Planetary Science Letters, **428**, 52 – 62.

Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse problems, **19**, R1.

Górszczyk, A., S. Operto, and M. Malinowski, 2017, Toward a robust workflow for deep crustal imaging by FWI of OBS data: The eastern nankai trough revisited: Journal of Geophysical Research: Solid Earth, **122**, 4601–4630.

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

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

Huang, G., R. Nammour, W. W. Symes, and M. Dolliazal, 2019, Waveform inversion via source extension, *in* SEG Technical Program Expanded Abstracts 2019: 4761–4766.

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

Kamath, N., R. Brossier, L. Métivier, A. Pladys, and P. Yang, 2021, Multiparameter full-waveform inversion of 3D ocean-bottom cable data from the Valhall field: Geophysics, **86**, B15–B35.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Con-

ference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia, Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia, 206–220.

Lei, W., Y. Ruan, E. Bozdağ, D. Peter, M. Lefebvre, D. Komatitsch, J. Tromp, J. Hill, N. Podhorszki, and D. Pugmire, 2020, Global adjoint tomography—model glad-m25: Geophysical Journal International, **223**, 1–21.

Li, Y., R. Brossier, and L. Métivier, 2020, On the comparison of MUMPS and STRUMPACK for 3D frequency-domain elastic wave modeling: Presented at the Expanded Abstracts, 82^{nd} Annual EAGE Meeting (Amsterdam).

Lu, Y., L. Stehly, R. Brossier, A. Paul, and A. W. ng Group, 2020, Imaging Alpine crust using ambient noise wave-equation tomography: Geophysical Journal International, **222**, 69–85.

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

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

Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: The Leading Edge, **25**, 156–166.

Métivier, L., and R. Brossier, 2016, The SEISCOPE optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: Geophysics, **81**, F11–F25.

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

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit

between seismograms using an optimal transport distance: Application to full waveform inversion: Geophysical Journal International, **205**, 345–377.

Neal, R. M., et al., 2011, Mcmc using hamiltonian dynamics: HandBook of markov chain monte carlo, **2**, 2.

Nocedal, J., 1980, Updating Quasi-Newton Matrices With Limited Storage: Mathematics of Computation, **35**, 773–782.

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

Operto, S., and A. Miniussi, 2018, On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea: Geophysical Journal International, **213**, 2037–2059.

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, **202**, 1362–1391.

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.

Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seismometer dataset: First Break, **28**, 71–78.

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.

Raknes, E. B., B. Arntsen, and W. Weibull, 2015, Three-dimensional elastic full waveform inversion using seismic data from the sleipner area: Geophysical Jounal International, **202**, 1877–1894.

Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture

marine seismic streamer data: Geophysical Journal International, **151**, 325–344.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: the next leap forward in imaging at Valhall: First Break, **28**, 65–70.

Solano, C. P., and R.-É. Plessix, 2019, Velocity-model building with enhanced shallow resolution using elastic waveform inversion — an example from onshore oman: GEOPHYSICS, **84**, R977–R988.

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

———, 2014, Seismic inverse poblems: recentd evelopments in theory and practice, *in* Proceedings of the Inverse Problems from Theory to Applications Conference (IPTA2014).

———, 2020, Wavefield reconstruction inversion: an example: Inverse Problems, **36**, 105010.

Symes, W. W., H. Chen, and S. E.Minkoff, 2020, Full-waveform inversion by source extension: Why it works, *in* SEG Technical Program Expanded Abstracts 2020: 765–769.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

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

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

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, R63–R77.

Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An

introduction to Full Waveform Inversion, *in* Encyclopedia of Exploration Geophysics: Society of Exploration Geophysics, R1–1–R1–40.

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

Wang, C., D. Yingst, P. Farmer, and J. Leveille, 2016, Full-waveform inversion with the reconstructed wavefield method, *in* SEG Technical Program Expanded Abstracts 2016: 1237–1241.

Wang, Y., and Y. Rao, 2009, Reflection seismic waveform tomography: Journal of Geophysical Research, **114**, 1978–2012.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, R429–R445.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, R59–R80.

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

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

LIST OF FIGURES

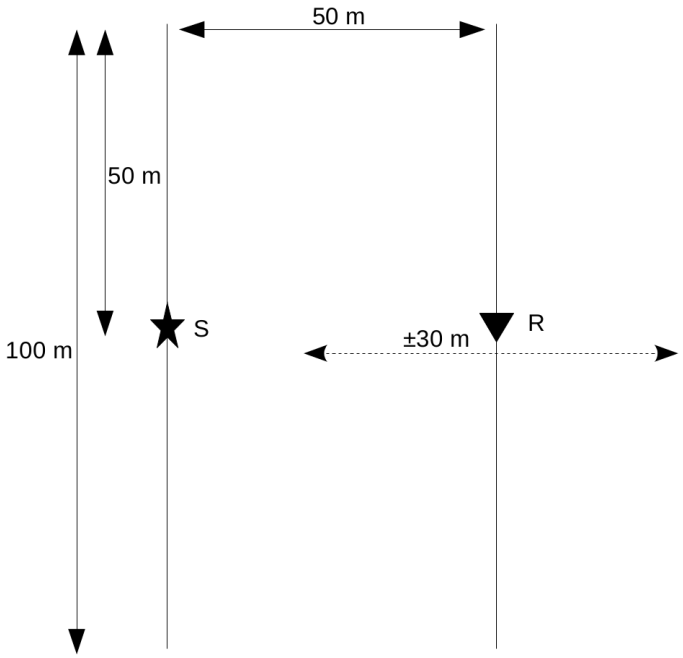

Figure 1: Acquisition configuration for the transmission case.

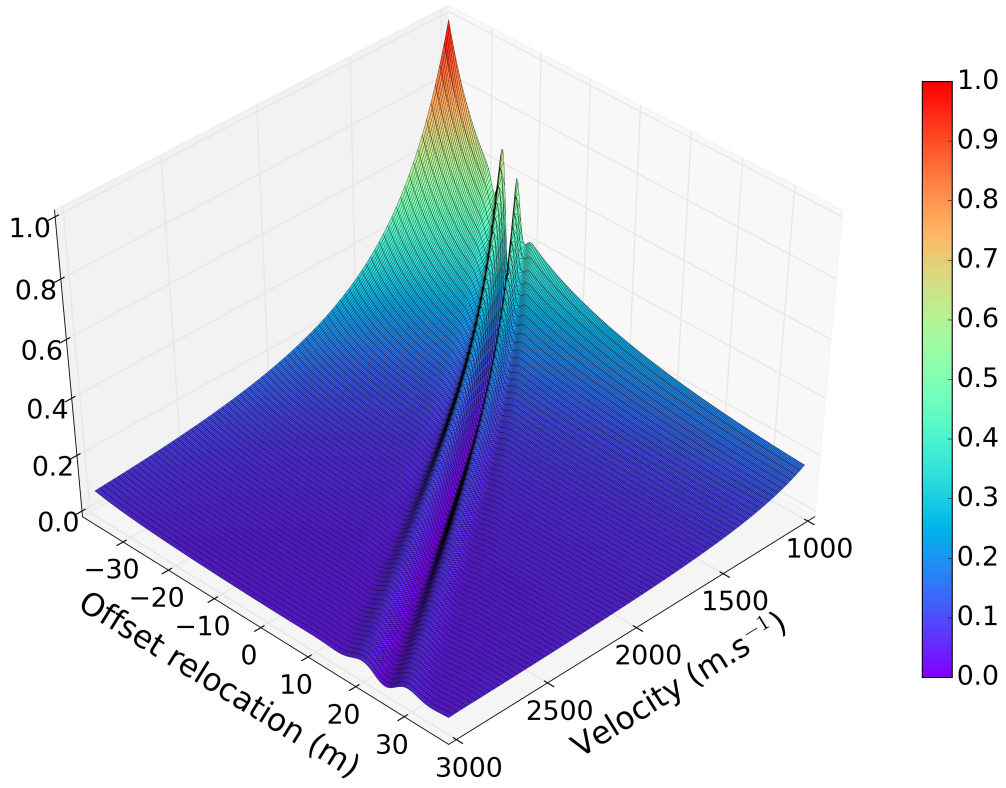

Figure 2: Map of the misfit function $f(m, \Delta x)$ in the cross-hole transmission case, using a single source/receiver couple at 50 m depth. The misfit function is normalized such that its maximum reaches 1. Its minimum is located in a narrow valley of attraction at position $\Delta x = 0$ and $m = 2000$ m.s^{-1}.

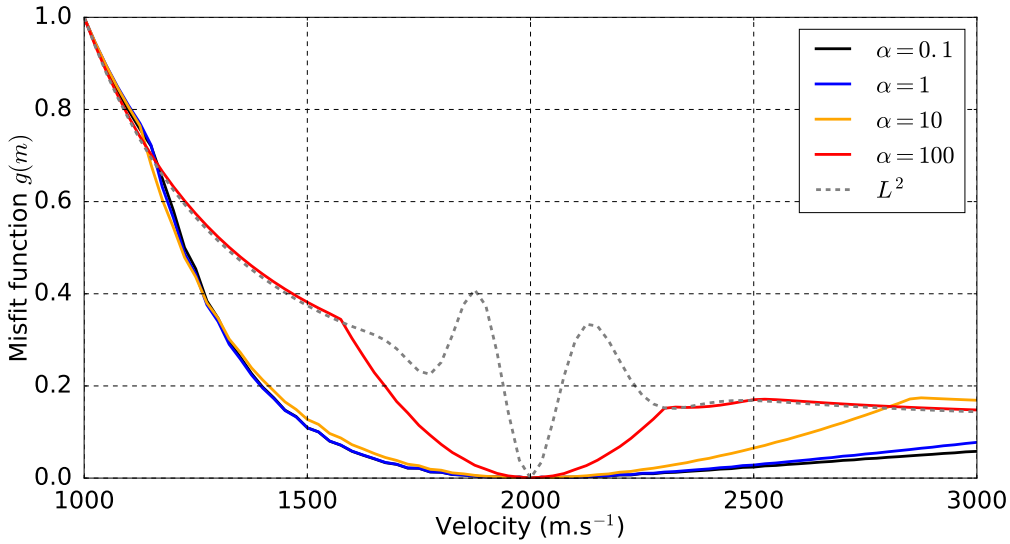

Figure 3: Profile of the misfit function $g(m)$ in the single couple source/receiver transmission case, depending on the weight α. Small weights $(0.1, 1)$ yield a convex function with respect to the velocity m, while larger weights $(10, 100)$ tend to reduce the width of the valley of attraction. In all cases the minimum is reached at the correct velocity $m = 2000$ m.s^{-1}.

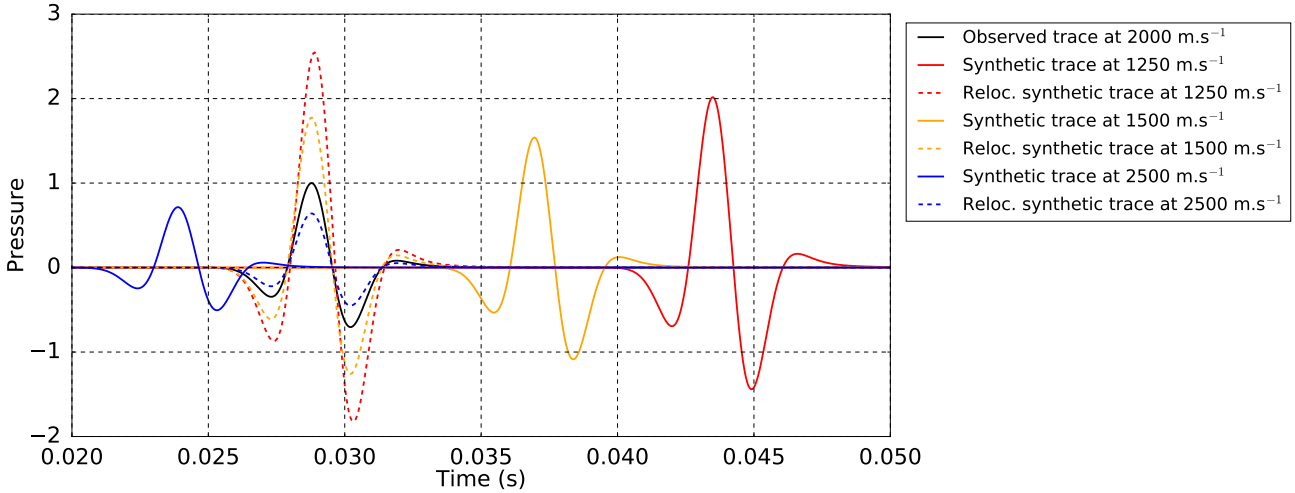

Figure 4: Comparison of the reference trace in black, obtained in the medium at 2000 m.s^{-1}, and synthetic traces before and after relocalization, obtained in media at 1250 m.s^{-1}, 1500 m.s^{-1}, and 2500 m.s^{-1} respectively. The traces have been normalized according to the maximum amplitude of the reference trace. The receiver relocalization is able to correct for the wrong kinematic and put the traces in phase with the reference trace. The amplitude difference remains however incorrect. The amplitude mismatch is convex with respect to the velocity: this explains why the receiver extension misfit function is convex with respect to the velocity in this case.

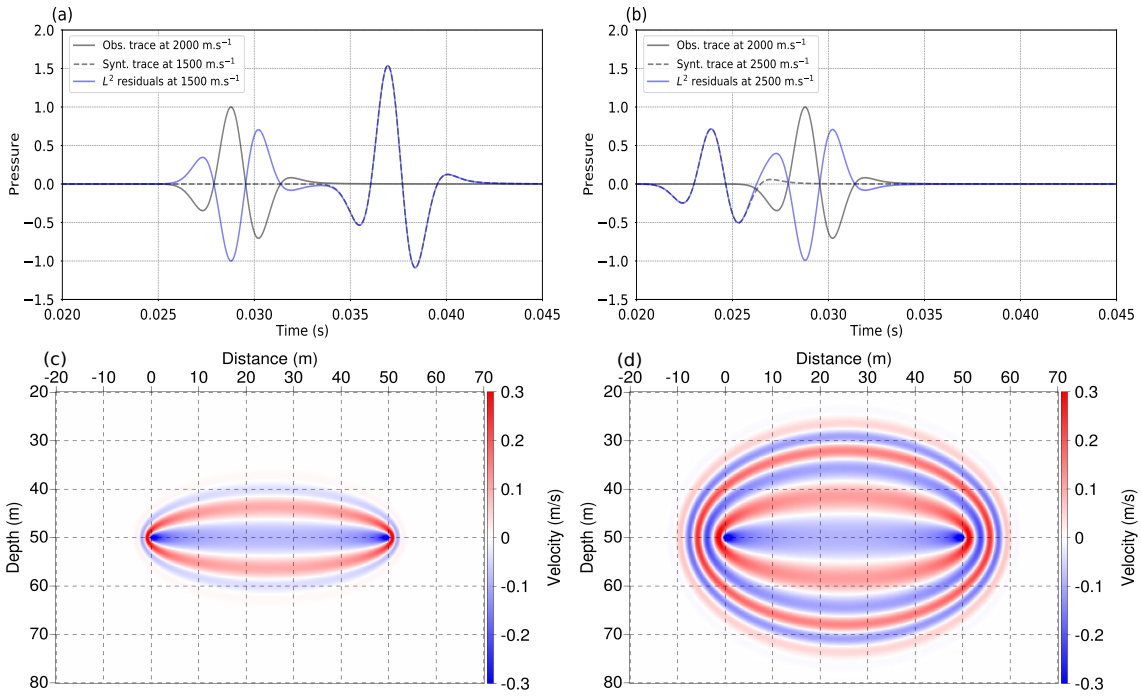

Figure 5: Transmission case, cross-hole configuration, single source/receiver pair. Reference, synthetic and adjoint traces in the least squares case for the slower medium at 1500 m.s^{-1} (a), faster medium at 2500 m.s^{-1} (b). Least-squares kernel in the slower medium at 1500 m.s^{-1} (c), faster medium at 2500 m.s^{-1} (d). In both cases the first Fresnel zone exhibits the same negative value, while a change of sign would be expected from a misfit function convex with respect to the velocity.

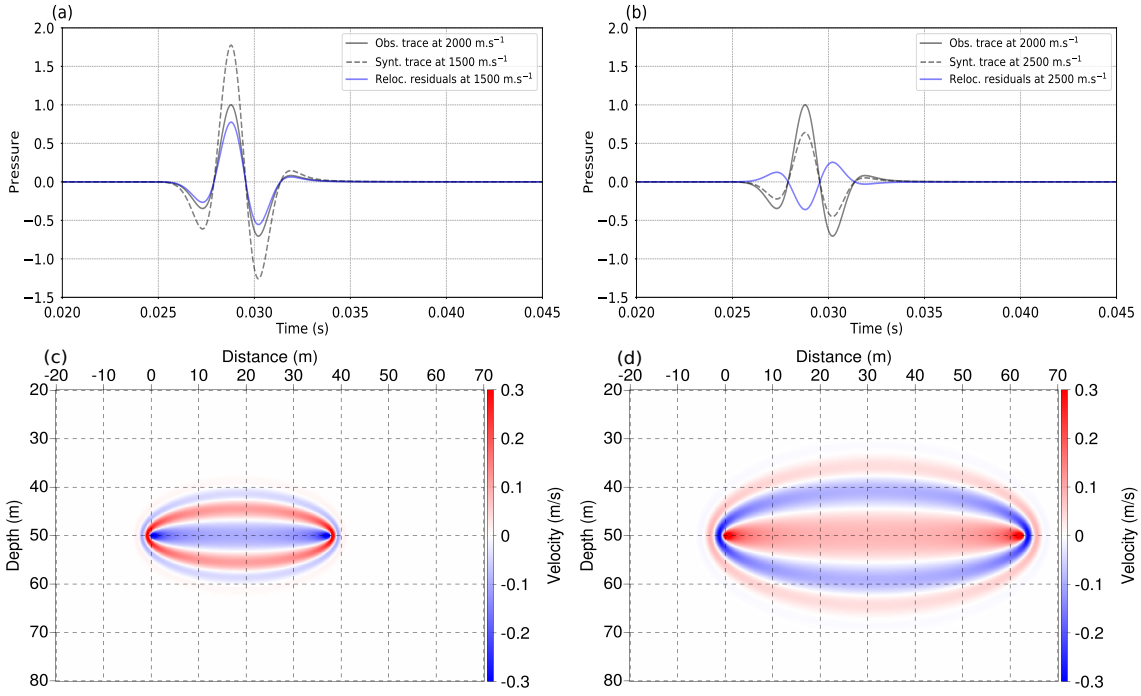

Figure 6: Transmission case, cross-hole configuration, single source/receiver pair. Reference, synthetic and adjoint traces in the receiver extension case for the slower medium at 1500 m.s^{-1} (a), faster medium at 2500 m.s^{-1} (b). Receiver extension kernel in the slower medium at 1500 m.s^{-1} (c), faster medium at 2500 m.s^{-1} (d). The synthetic and reference traces are in phase, with higher amplitude for the slower medium (a) and lower amplitude for the faster medium (b). The adjoint source therefore changes its sign from one medium to the other. The first Fresnel zone of the corresponding kernels thus exhibits the same change of sign depending whether it is calculated in the slower (c) or faster medium (d).

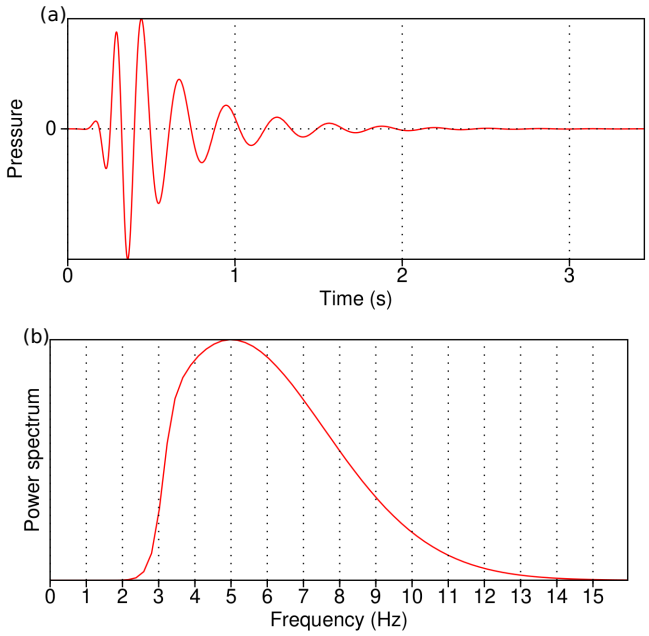

Figure 7: Time signature of the source wavelet used to generate the data for Marmousi and Valhall case (a), together with its power spectrum (b). No energy is present below 2.5 Hz.

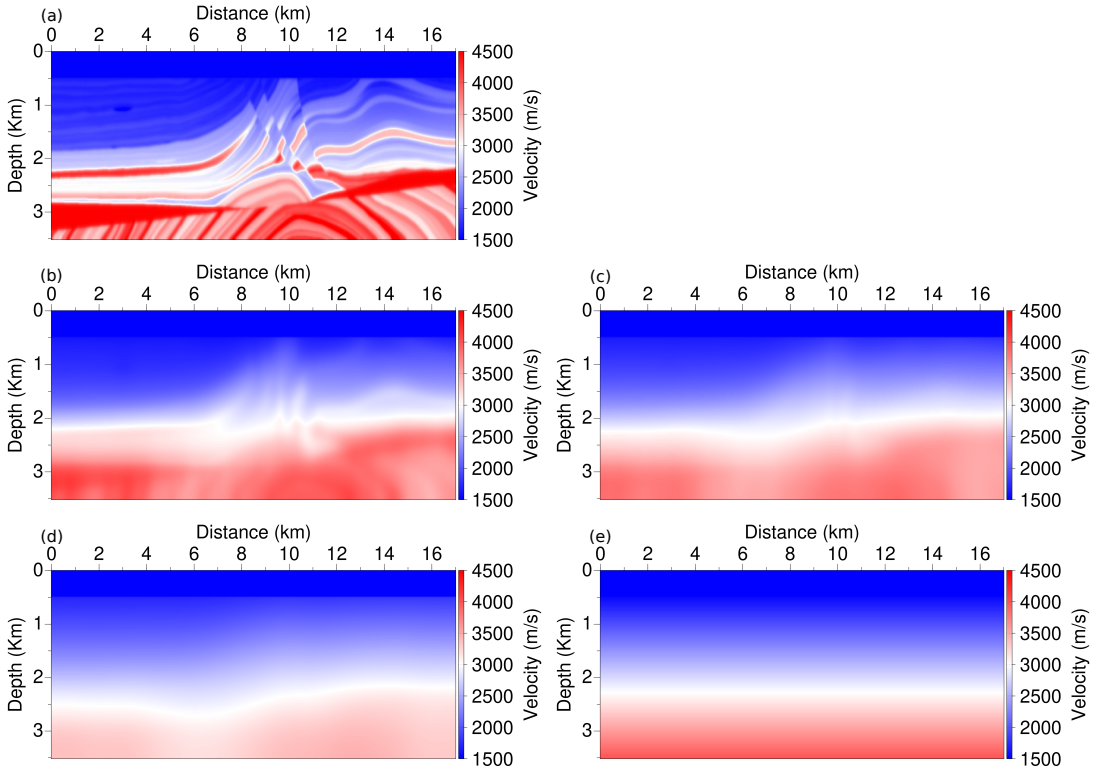

Figure 8: Marmousi II case study. Exact (a) and initial models 1 (b), 2 (c), 3 (d), and 4 (e).

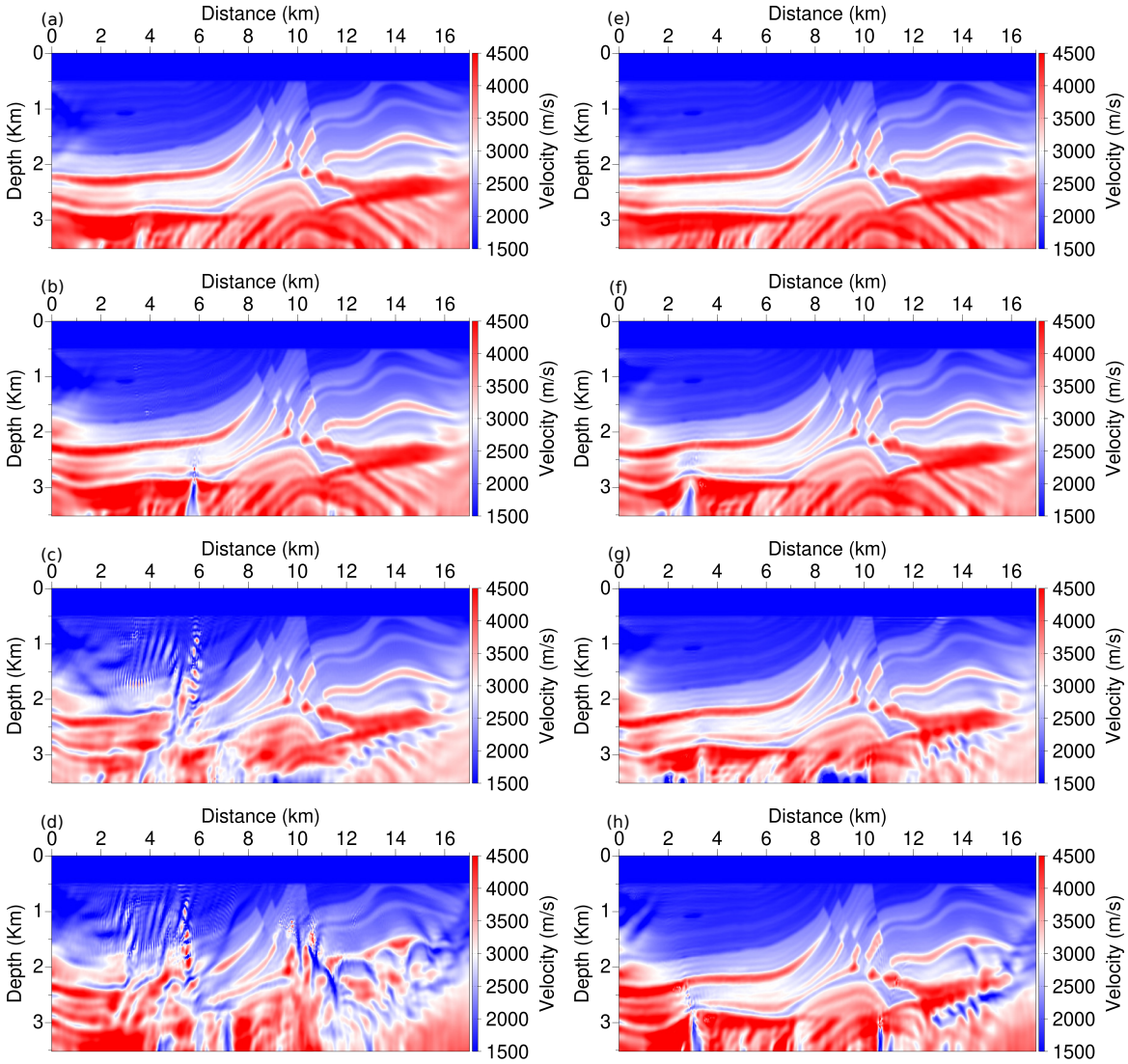

Figure 9: Marmousi II case study. Reconstructed models starting from the initial models in Figure 8 using a conventional L^2 approach (left column) and the receiver relocalization approach (right column). L^2 results starting from models 1 (a), 2 (b), 3 (c), 4 (d). Receiver relocalization results starting from models 1 (e), 2 (f), 3 (g), 4 (h).

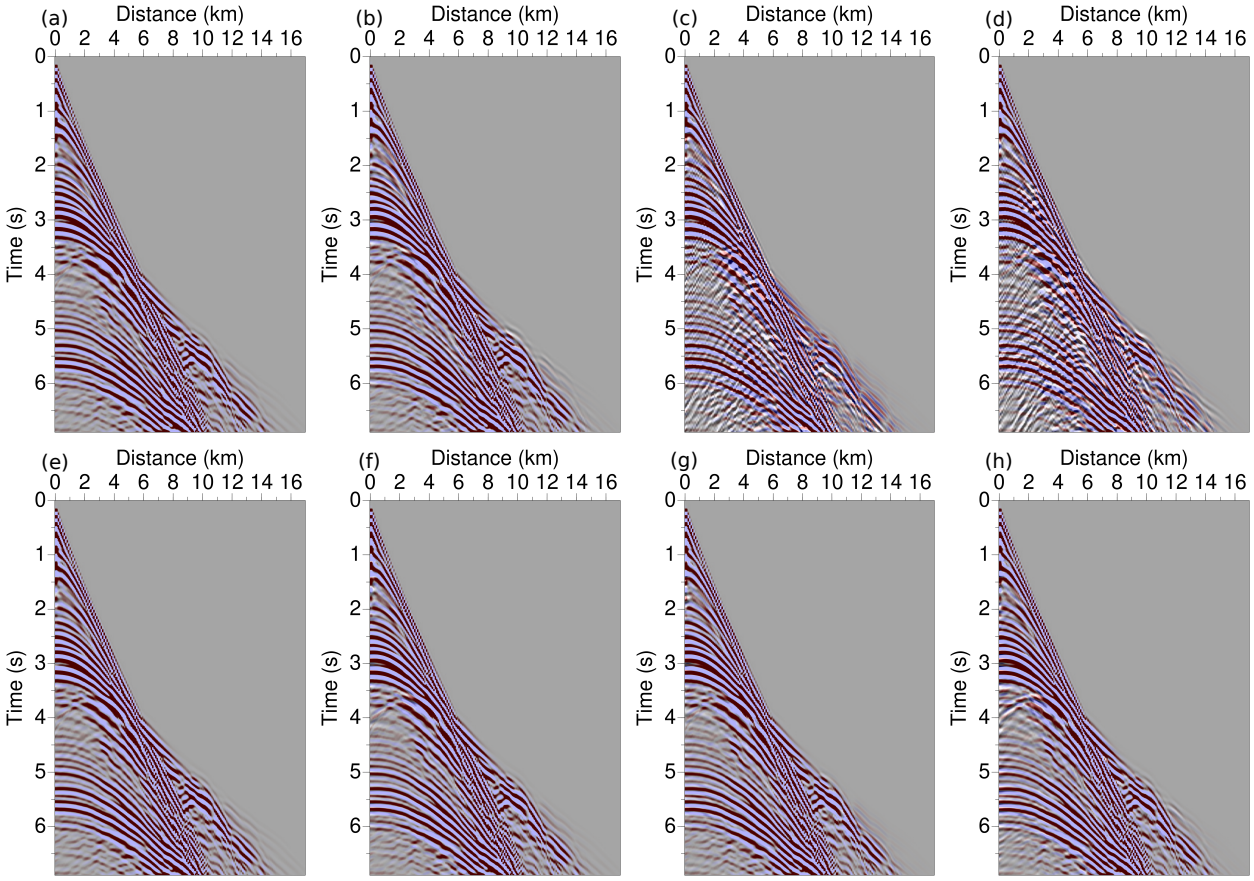

Figure 10: Marmousi II case study. Final data fit using a conventional L^2 approach (top row) and the receiver relocalization approach (bottom row) starting from the initial models presented in Figure 8. The exact data appears in blue and red colors in transparency. The data in the final models is in black and white. L^2 data fit starting from model 1 (a), 2 (b), 3 (c), 4 (d). Receiver relocalization data fit starting from model 1 (e), 2 (f), 3 (g), 4 (h).

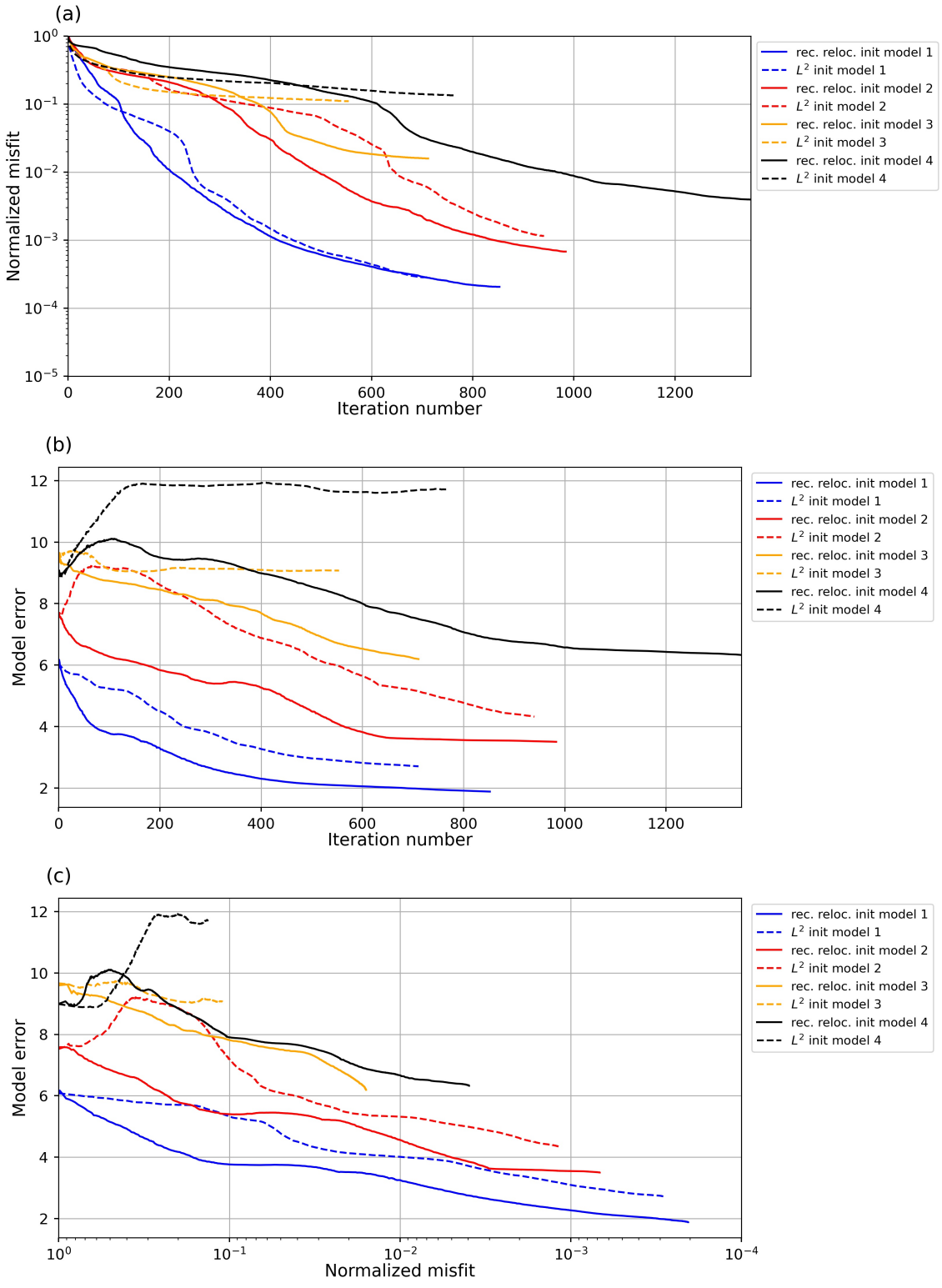

Figure 11: Marmousi II case study. Quantitative analysis of error evolution. Relative decrease of the misfit function along iterations (a). Relative decrease of the model error along iterations (b). Relative decrease of the model error depending on the decrease of the misfit function (c).

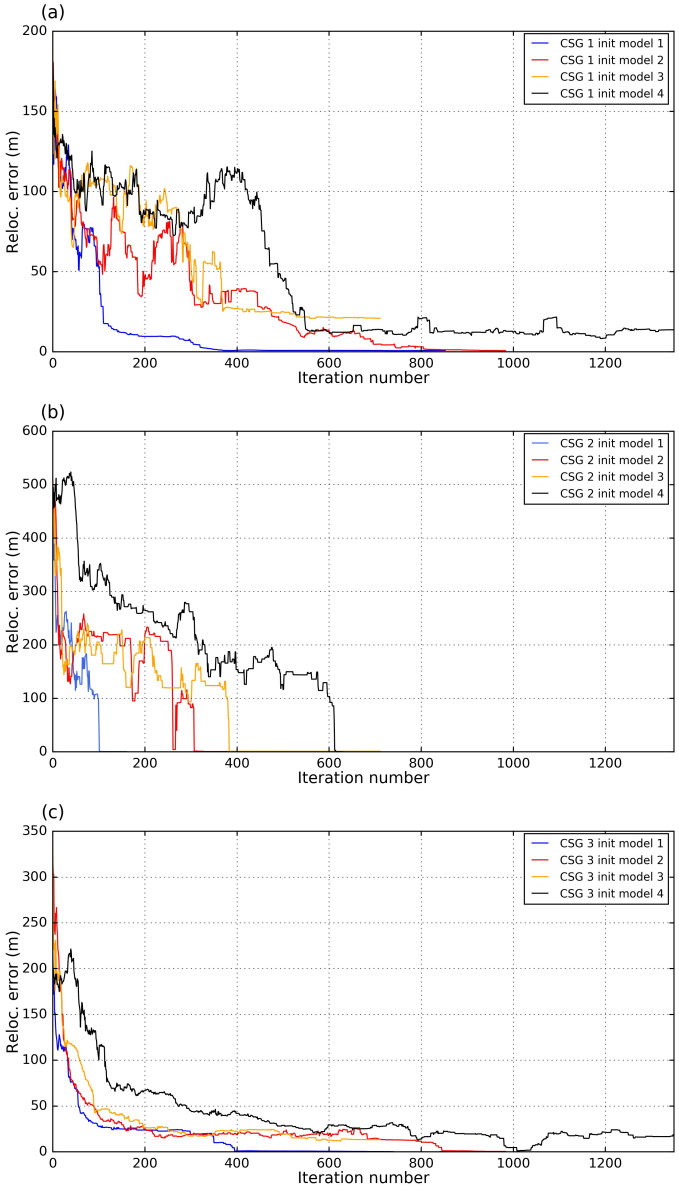

Figure 12: Marmousi II case study. Evolution of the relocalization error along iterations, depending on the choice of initial model for leftmost shot gather (a), central shot gather (b), right shot gather (c).

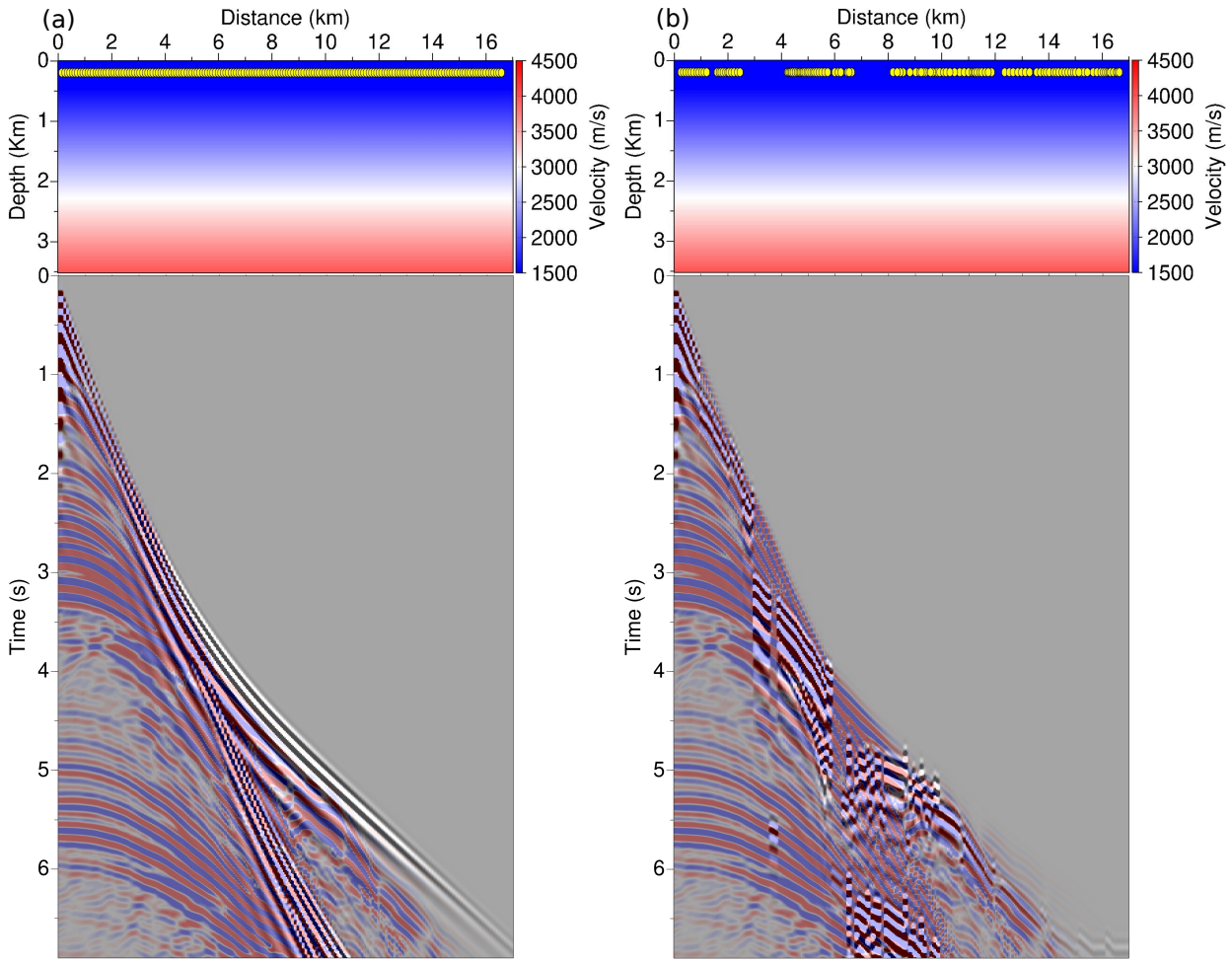

Figure 13: Marmousi II case study. Data fit in the initial model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.

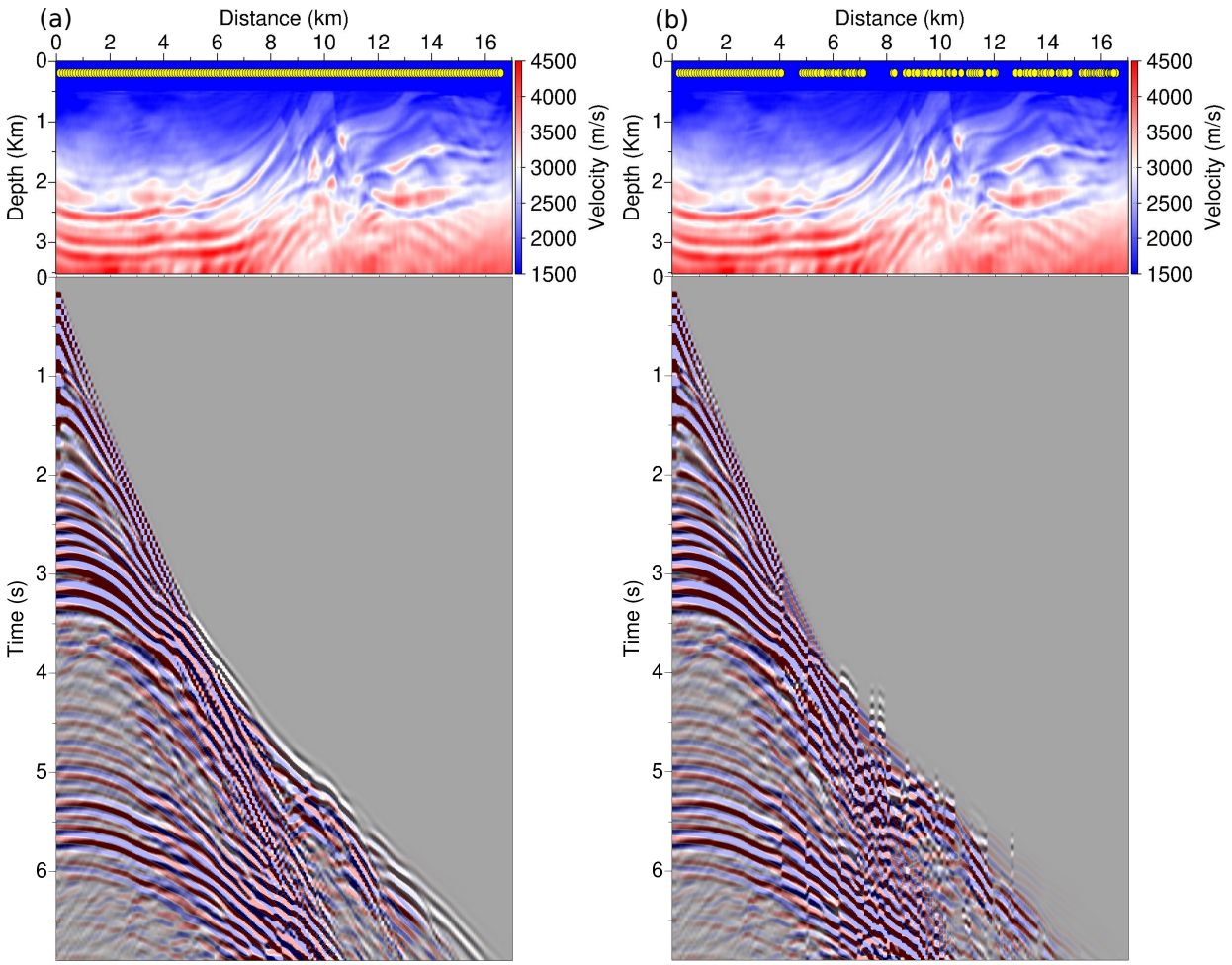

Figure 14: Marmousi II case study. Data fit in the iteration 100 model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.

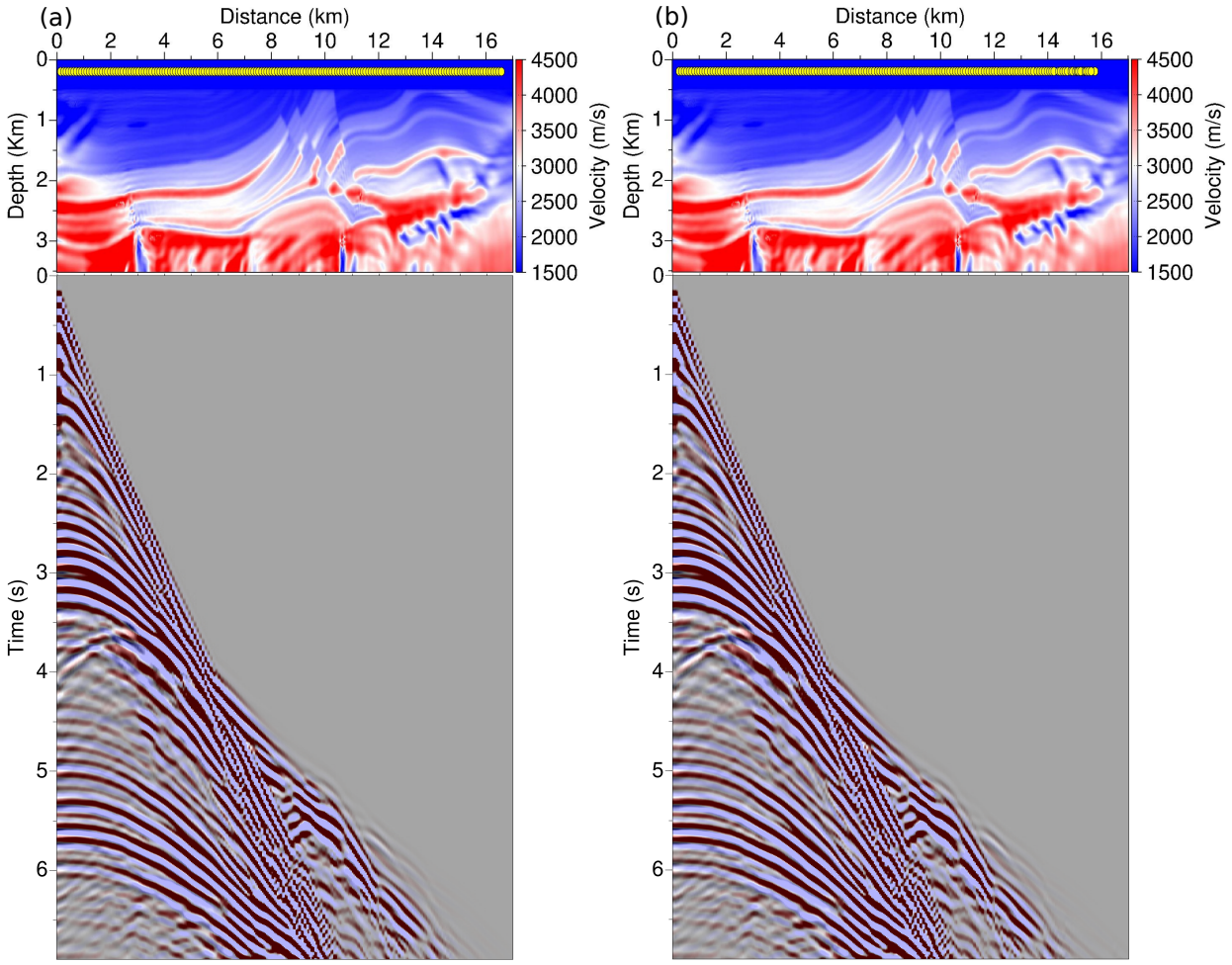

Figure 15: Marmousi II case study. Data fit in the final model before (a) and after (b) the receiver relocalization step. The receiver positions are represented in yellow ellipses.

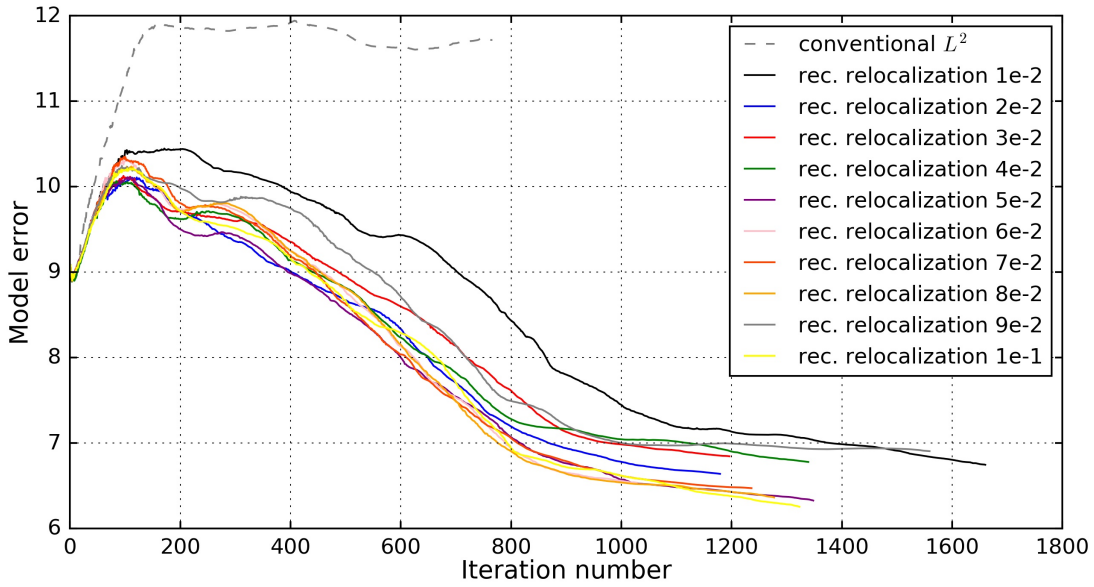

Figure 16: Marmousi II case study. Sensitivity to the weighting parameter α. Model error along the iteration depending on the choice of α, ranging from 10^{-2} to 10^{-1}.

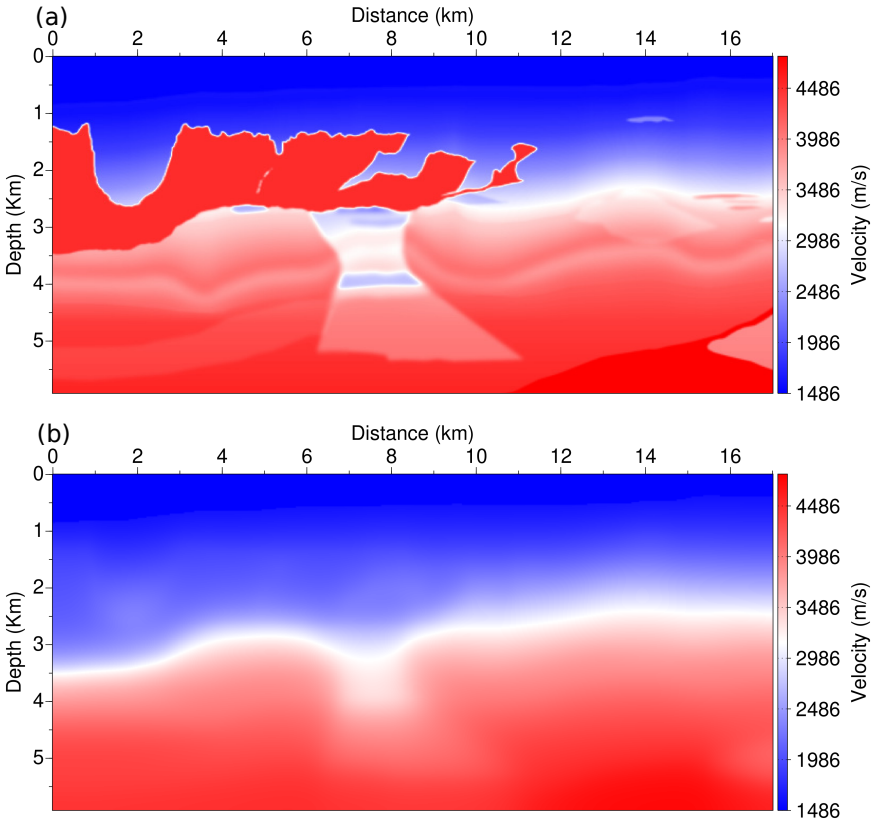

Figure 17: Exact (a) and initial (b) models used for the BP 2004 case study.

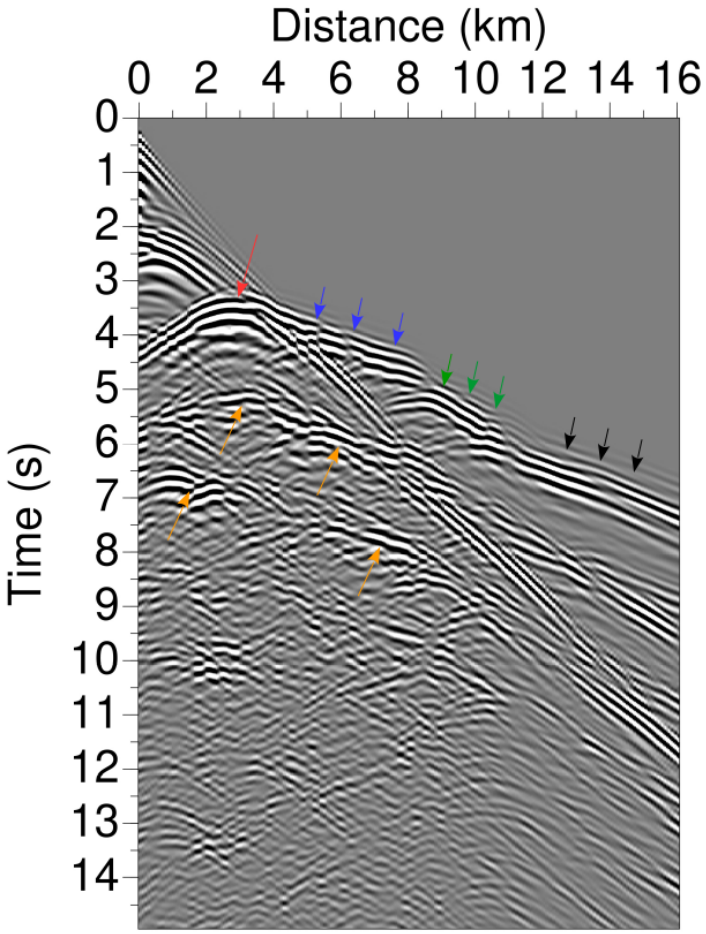

Figure 18: BP 2004 case study. Left most common shot gather. The canyon structure at $x = 2$ km, generates energetic first order (red arrow) and higher order reflections (orange arrows). The blue arrows depict the refraction of the direct by the salt body. The event depicted by the green arrows corresponds to the transmission of the direct wave within the salt structure. Black arrows depict arrivals coming from below the salt after interacting with the canyon

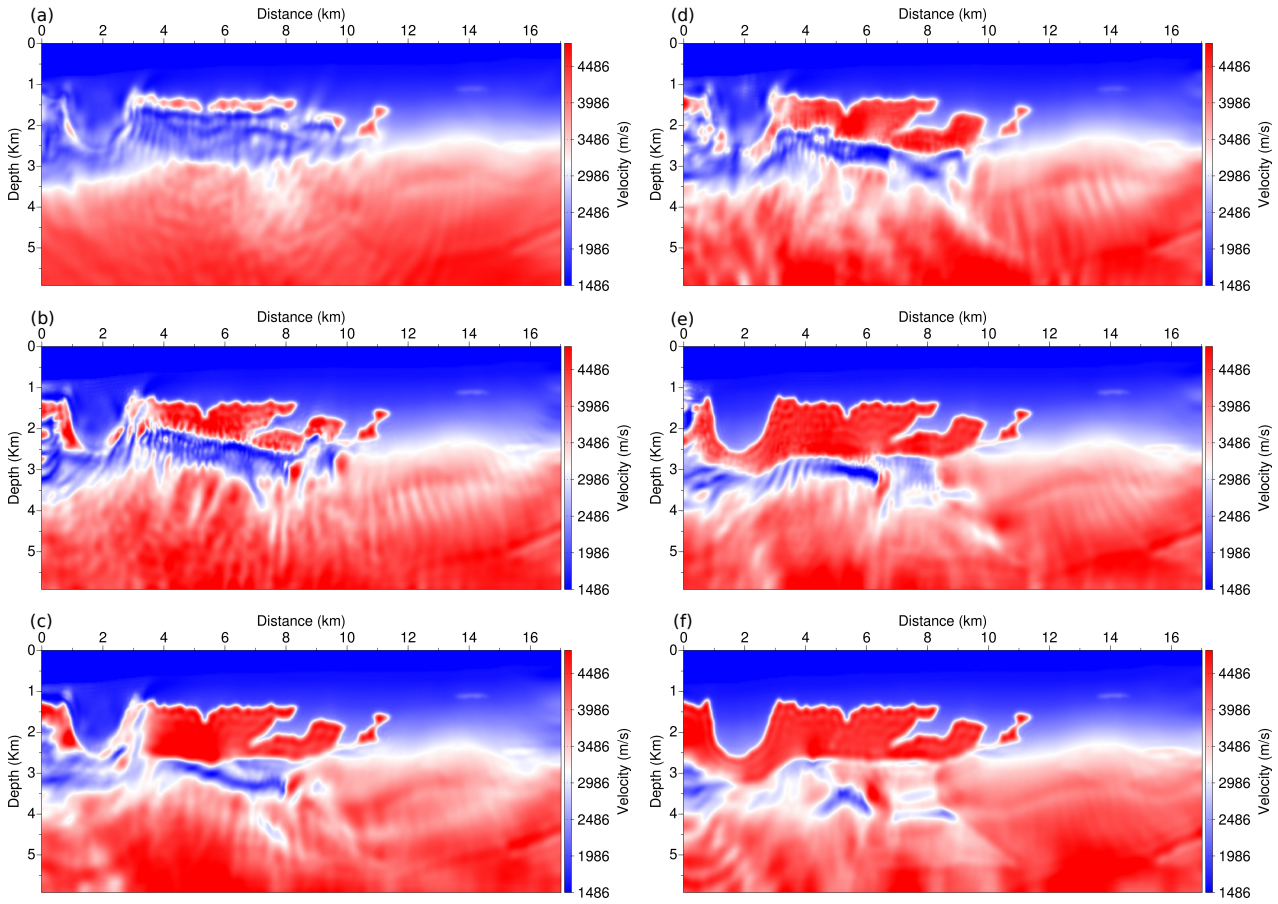

Figure 19: BP 2004 case study. P-wave velocity reconstruction using conventional FWI after 1st time-window (a), 2nd time-window (b), 3rd time-window (c). P-wave velocity reconstruction using receiver relocalization approach after 1st time-window (d), 2nd time-window (e), 3rd time-window (f).

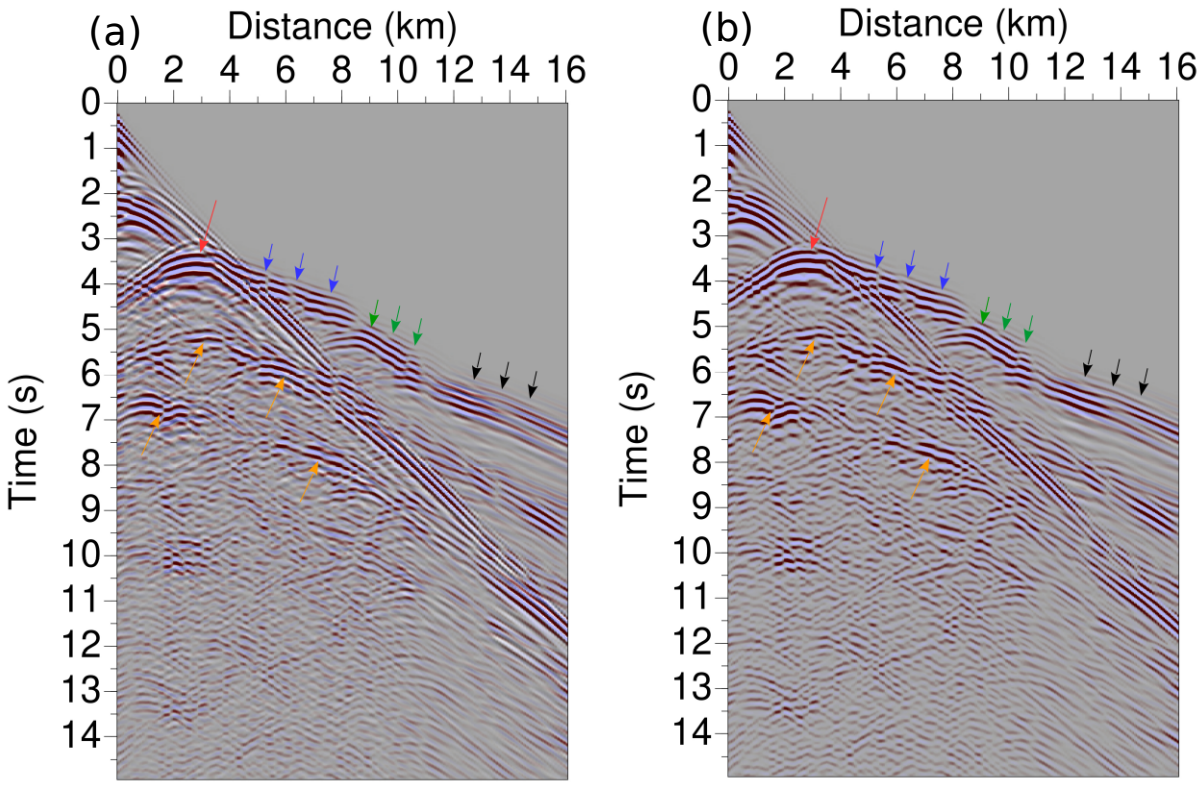

Figure 20: BP 2004 case study. Data fit in the P-wave velocity models reconstructed using a conventional FWI approach (a), using the receiver relocalization approach (b). The true data in blue/red color is superposed with the synthetic data in black and white color. For a correct data fit no black and white color should appear. Blue and green arrows depict refracted and transmitted events. Red arrow depicts short offset reflection on the left (canyon shape) part of the model. Multiples of these reflections are depicted by orange arrows. Black arrow depict later arrivals coming from below the salt. While using conventional FWI, mostly transmitted events are matched, using the relocalization approach all of these events are correctly matched.

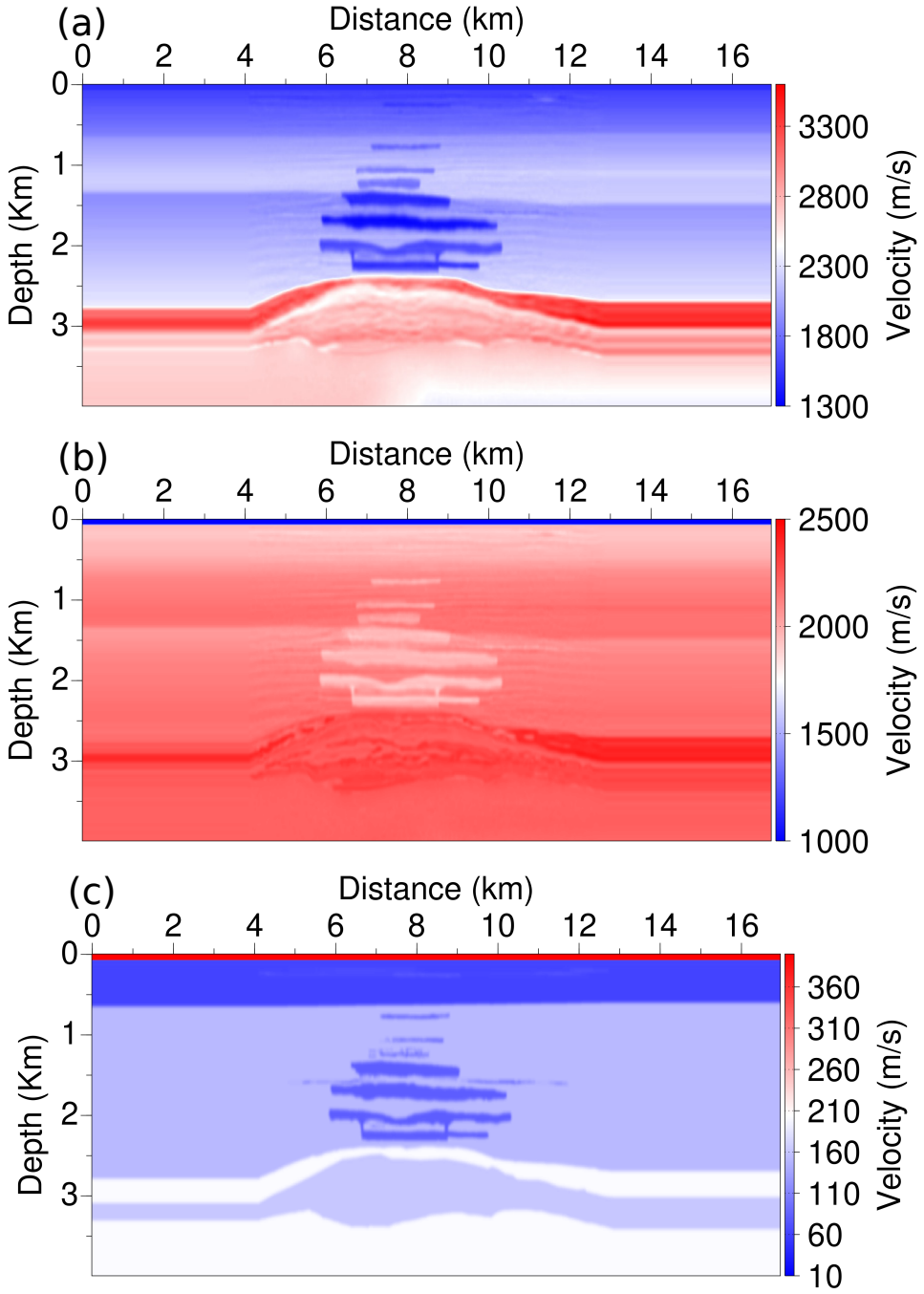

Figure 21: Valhall case study. Exact P-wave velocity model (a), exact density model (b), and exact quality factor (c) used to generate the data.

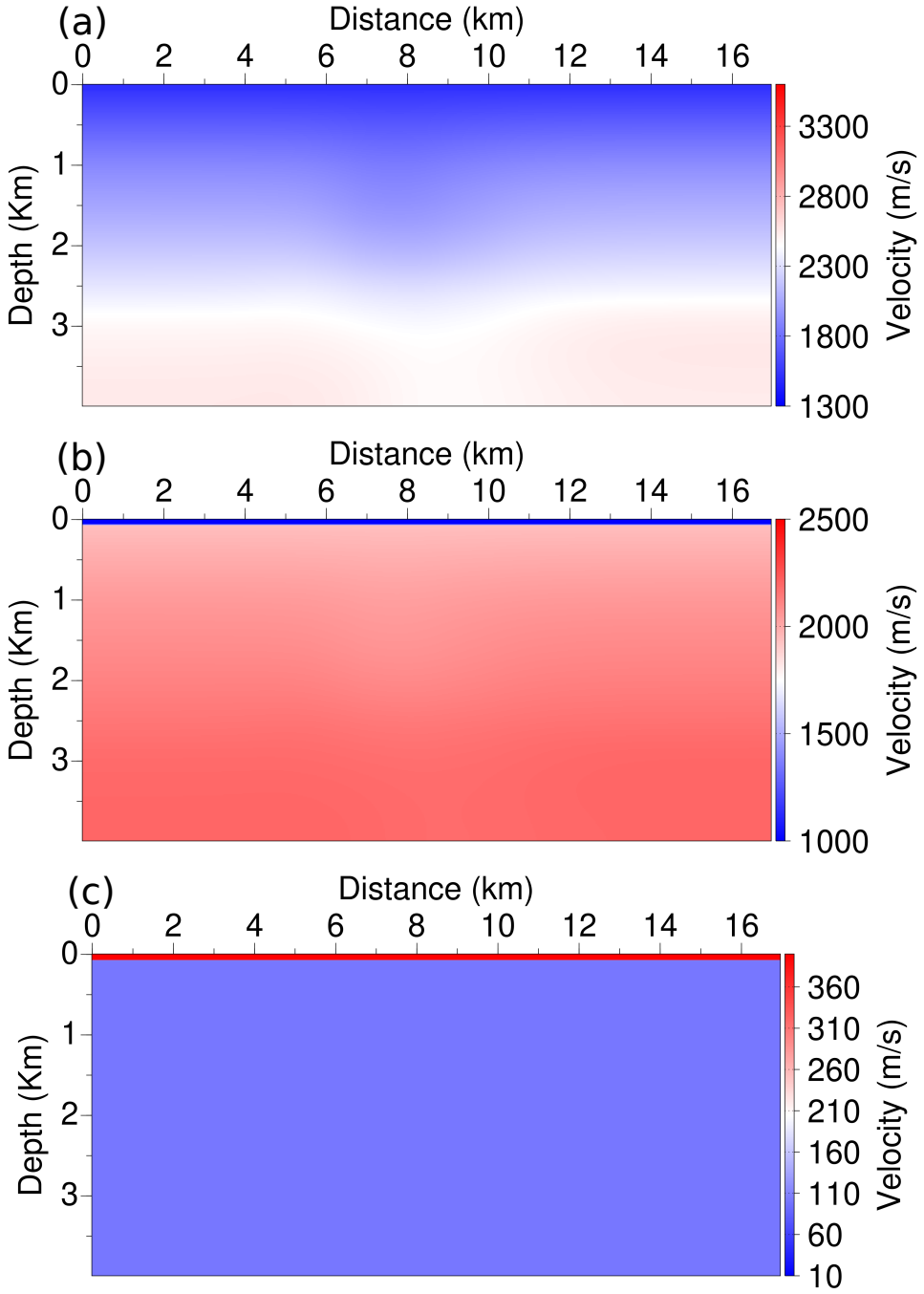

Figure 22: Valhall case study. Initial P-wave velocity model (a), initial density model (b), and initial quality factor (c).

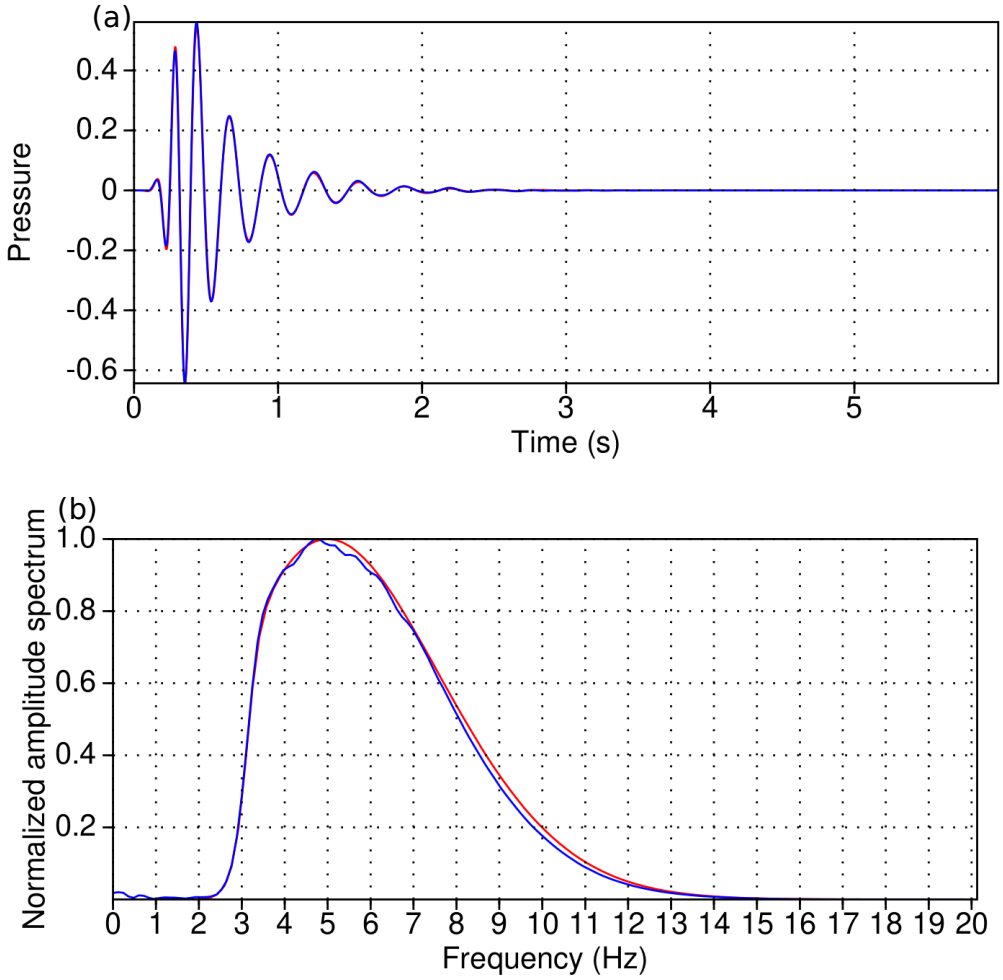

Figure 23: Valhall case study. Estimated source wavelet (blue) compared with the true source wavelet (red). Time-signature (a), normalized amplitude spectrum (b).

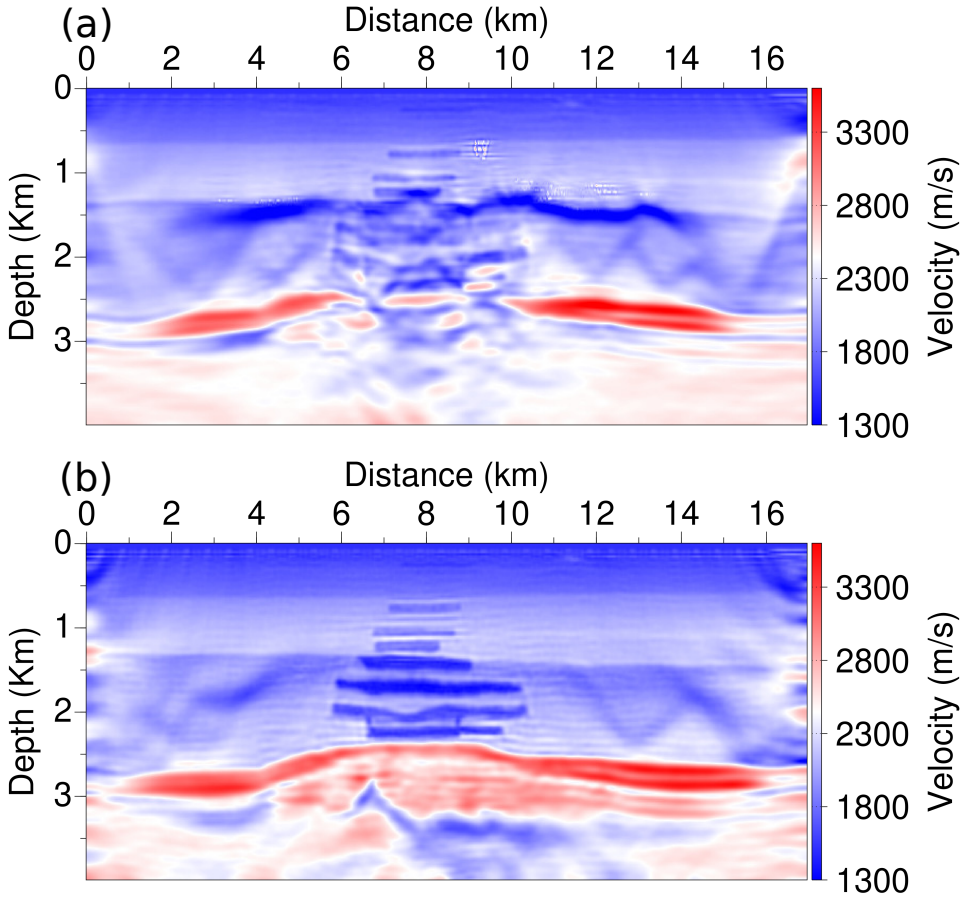

Figure 24: Valhall case study. Final models obtained using L^2 FWI (a), and the receiver relocalization strategy (b).

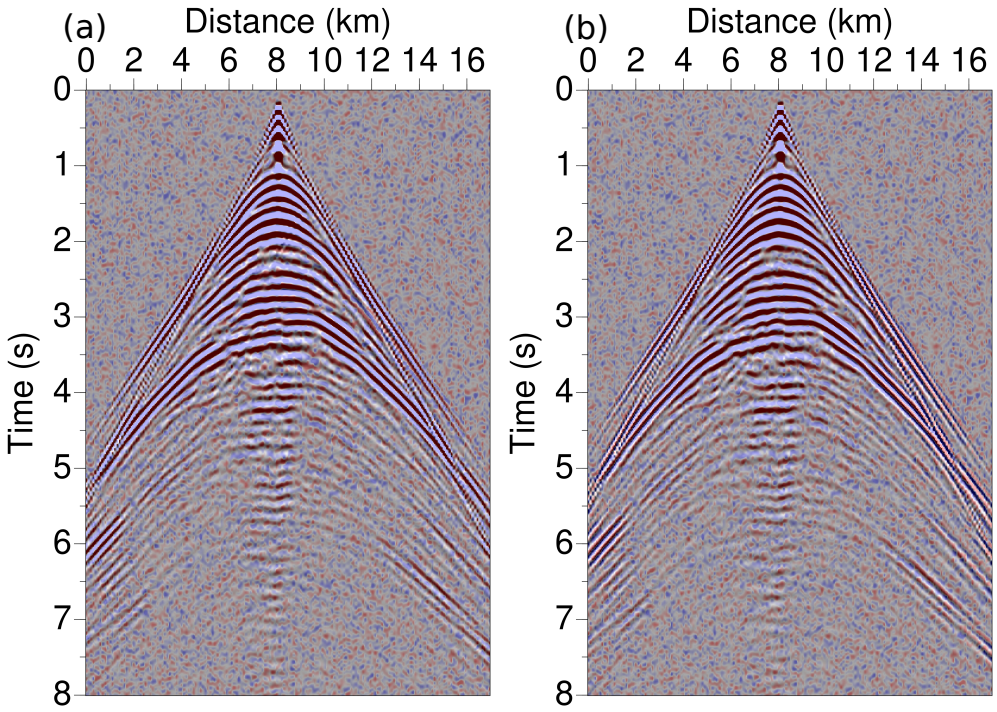

Figure 25: Valhall case study. Comparison of observed (blue and red) and synthetic (black and white) data for the shot gather associated with the source located at $x_S = 8$ km. L^2 result (a), receiver relocalization result (b). A good data match is indicated by the dominance of purple and black color. The presence of white and red indicates an incorrect data match. In this respect the L^2 data match is relatively good. The receiver relocalization data match is also correct, except for large offset diving waves events.

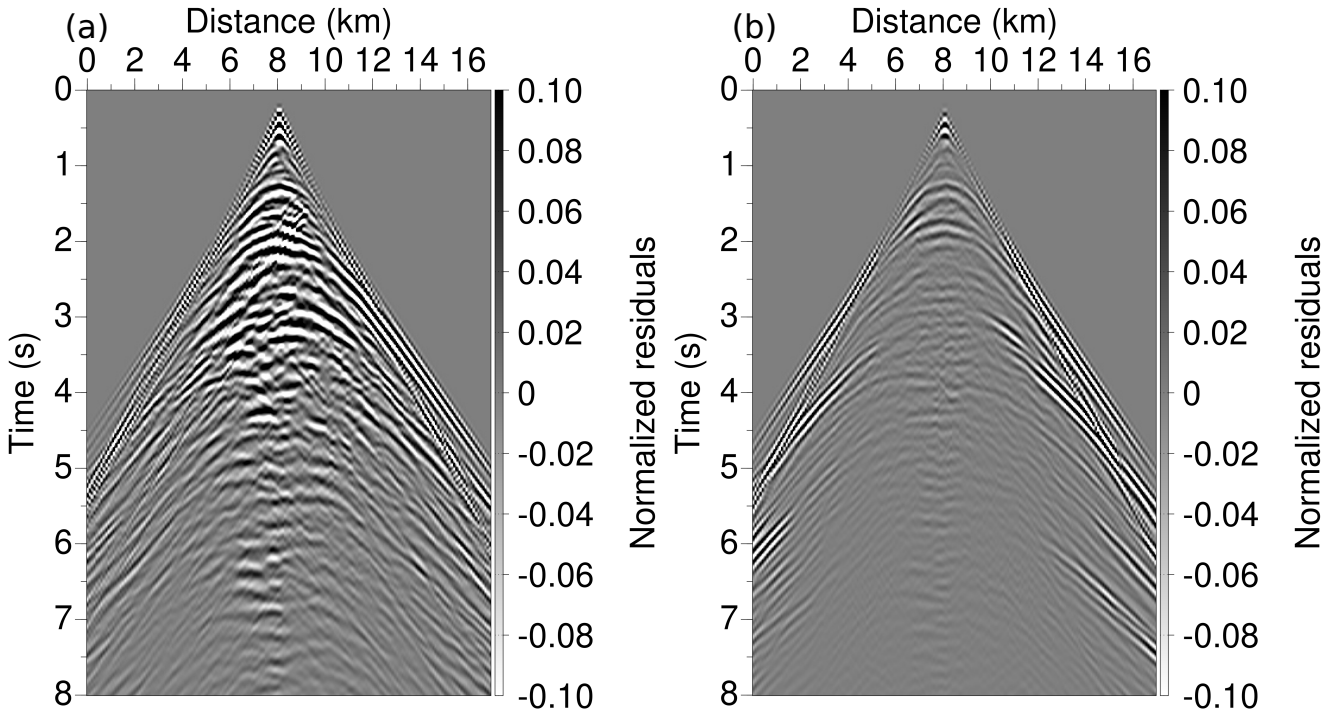

Figure 26: Valhall case study. Comparison of normalized residuals for the shot gather associated with the source located at $x_S = 8$ km. L^2 result (a), receiver relocalization result (b). The residuals are computed as the difference between the synthetic data in the final model using both approaches and the observed data without noise. The L^2 residuals exhibit a clear mismatch of the short and medium offsets event associated with reflection on the gas cloud layers.

LIST OF TABLES

Case study	$n_z \times n_x$	Inc. field	Adj. field	Rec. reloc. loop	Other	Total	Extra cost
Marmousi	141 × 681	6.7 s	19.1 s	**1.4 s**	2.4 s	29.6 s	**4.9**%
BP 2004	237 × 651	10.6 s	29.9 s	**1.1 s**	2.9 s	44.5	**2.53**%
Valhall	160 × 679	11.5 s	44.5 s	**1.1 s**	10.6 s	67.7	**1.65** %

Table 1: Computational time for different gradient building steps.