

Source estimation for wavefield-reconstruction inversion

Zhilong Fang¹, Rongrong Wang^{2,3}, and Felix J. Herrmann^{1,4}

¹Department of Earth, Ocean and Atmospheric Sciences, University of British
Columbia

²Department of Mathematics, Michigan State University

³Department of Computational Mathematics Science and Engineering,
Michigan State University

⁴Now at Georgia Institute of Technology

(January 23, 2018)

Running head: **Source estimation for WRI**

ABSTRACT

Source estimation is essential to all the wave-equation-based seismic inversions, including full-waveform inversion and the recently proposed wavefield-reconstruction inversion. When the source estimation is inaccurate, errors will propagate into the predicted data and introduce additional data misfit. As a consequence, inversion results that minimize this data misfit may become erroneous. To mitigate the errors introduced by the incorrect and pre-estimated sources, an embedded procedure that updates sources along with medium parameters is necessary for the inversion. So far, such a procedure is still missing in the context of wavefield-reconstruction inversion, a method that is, in many situations, less prone to local minima related to the so-called cycle skipping, compared to full-waveform inversion through exact data-fitting. While wavefield-reconstruction inversion indeed helps to mitigate issues

related to cycle skipping by extending the search space with wavefields as auxiliary variables, it relies on having access to the correct source functions. To remove the requirement of having the accurate source functions, we propose a source estimation technique specifically designed for wavefield-reconstruction inversion. To achieve this task, we consider the source functions as unknown variables and arrive at an objective function that depends on the medium parameters, wavefields, and source functions. During each iteration, we apply the so-called variable projection method to simultaneously project out the source functions and wavefields. After the projection, we obtain a reduced objective function that only depends on the medium parameters and invert for the unknown medium parameters by minimizing this reduced objective. Numerical experiments illustrate that this approach can produce accurate estimates of the unknown medium parameters without any prior information of the source functions.

INTRODUCTION

The objective of full-waveform inversion (FWI) is to compute the best earth model that is consistent with observed data (Tarantola and Valette, 1982; Virieux and Operto, 2009). Most methods that fall into this category reconstruct the medium parameters by minimizing the misfit between synthetic and observed wavefields for different sources and at possibly different sets of receiver locations. As a consequence, the accuracy of the inversion result relies heavily on the correctness of the modeling kernel that generates synthetic wavefields. Being part of the wave-equation, the source signature is one key factor that affects the synthetic data and requires to be estimated during the inversion in practice.

Many source estimation approaches have been developed for the conventional FWI, also known as the adjoint-state method (Plessix, 2006; Virieux and Operto, 2009). Liu et al. (2008) proposed to treat source functions as unknown variables and update them with medium parameters simultaneously. However, the amplitudes of source functions and medium parameters are different in scale, and so are the amplitudes of their corresponding gradients. As a result, the gradient-based optimization algorithms would tend to primarily update the one at the larger scale. Pratt (1999) proposed another approach to estimate source functions during each iteration of the adjoint-state method. In this approach, the author estimates the source functions after each update of the medium parameters by solving a least-squares problem for each frequency. The solution of the least-squares problem is a single complex scalar that minimizes the difference between the observed and predicted data computed with the updated medium parameters. Because the updates of the source functions and medium parameters are in two separate steps, this method is not impacted by the different amplitude scales of the gradients. Indeed, Aravkin and van Leeuwen (2012)

and van Leeuwen et al. (2014) pointed out that the problem of FWI with source estimation falls into a general category called separable nonlinear least-squares (SNLLS) problems (Golub and Pereyra, 2003), which can be solved by the so-called variable projection method. The method proposed by Pratt (1999) falls into that category. By virtue of these variable projections, the gradient with respect to the source functions becomes zero, because the estimated source is the optimal solution that minimizes the misfit between the observed and predicted data given the current medium parameters (Aravkin and van Leeuwen, 2012). For this reason, the gradient with respect to the medium parameters no longer contains the non-zero cross derivative with respect to the source functions, and the inverse problem becomes source-independent. In fact, because the problem of FWI with source estimation is a special case of the SNLLS problem, the minimization of this source-independent objective converges in fewer iterations than that of the original source-dependent objective theoretically (Golub and Pereyra, 2003). Furthermore, Li et al. (2013) presented empirical examples to illustrate that the variable projection method is more robust to noise compared to the method that performs alternating gradient descent steps on two variables separately (Zhou and Greenhalgh, 2003), as well as to the one that minimizes the two variables simultaneously (Liu et al., 2008).

A well-known problem of the conventional adjoint-state method is that the resulting optimization problem may have local minima in common situations where the starting model for the velocity is not known accurately enough. As argued by van Leeuwen and Herrmann (2013b), this sensitivity to the starting model is due to the fact that the adjoint-state method eliminates the wave-equation constraint by explicitly solving it, which results in an objective function that is highly nonlinear in the unknown medium parameters. To mitigate the problem of these local minima, van Leeuwen and Herrmann (2013b) and van Leeuwen

and Herrmann (2015) proposed a penalty formulation instead, where the wave-equation constraint is replaced by a weak constraint in the form of an ℓ_2 -norm penalty. This method is known as wavefield-reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013b, 2015). Compared to the adjoint-state method, WRI relaxes the wave-equation constraint by introducing the wavefields as auxiliary variables and adding a penalty term to the data misfit. This idea of extending the search space was also recently utilized by Huang et al. (2016), where an additional source term is introduced as the auxiliary variable. Both of these approaches are more or less equivalent and are successful in mitigating the effects of local minima, because the extended search space allows for fitting the observed data even in cases where the velocity model is incorrect. As a result, we can carry out the inversion with poorer initial models and start at higher frequencies (Peters et al., 2014). While these developments are encouraging, WRI does, as illustrated in Figure 1, require accurate information on the source functions. For instance, a small difference in the origin time of the source can lead to a significant deterioration of the inversion result (cf. Figure 1).

[Figure 1 about here.]

Recognizing this sensitivity to source functions, we propose a modified WRI framework with source estimation integrated, making the method applicable to field data. In this new formulation, we continue to solve an optimization problem but now one that has the source functions, medium parameters, and wavefields all as unknowns. For fixed medium parameters, we observe that the objective function of this optimization problem is quadratic with respect to both wavefields and source functions. As a result, if we collect the wavefield and source function for each source experiment into one single unknown vector variable, the optimization problem becomes an SNLLS problem with respect to the medium parameters

and this new unknown variable. As before, we propose the variable projection method to solve the optimization problem by projecting out this new unknown variable during each iteration. After these least-squares problems, we compute single model updates either with the gradient descent or limited-memory Broyden–Fletcher–Goldfarb–Shanno (l-BFGS) method (Nocedal and Wright, 2006).

The outline of the paper is as follows. In the methodology section, we first start with a review of the source estimation technique for the conventional FWI using the variable projection method. Next, we introduce the standard WRI and propose our method for WRI source estimation. We conclude by presenting a computationally efficient implementation, which we evaluate on several synthetic examples to highlight the benefits of our approach.

METHODOLOGY

FWI with source estimation

During FWI, we solve the following least-squares problem for the discretized n_{grid} -dimensional unknown medium parameters, squared slowness $\mathbf{m} \in \mathbb{R}^{n_{\text{grid}}}$, which appear as coefficients in the acoustic time-harmonic wave-equation, i.e.,

$$\underset{\mathbf{u}, \mathbf{m}}{\text{minimize}} \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \|\mathbf{P}_i \mathbf{u}_{i,j} - \mathbf{d}_{i,j}\|_2^2 \quad (1)$$

$$\text{subject to } \mathbf{A}_j(\mathbf{m}) \mathbf{u}_{i,j} = \mathbf{q}_{i,j}.$$

Here, the i, j 's represent the source and frequency indices, and the vector $\mathbf{d} \in \mathcal{C}^{n_{\text{freq}} \times n_{\text{src}} \times n_{\text{rcv}}}$ represents monochromatic Fourier transformed data collected at n_{rcv} receiver locations from n_{src} seismic sources and sampled at n_{freq} frequencies. The matrix $\mathbf{A}_j(\mathbf{m}) = \omega_j^2 \text{diag}(\mathbf{m}) + \Delta$ is the discretized Helmholtz matrix at angular frequency ω_j . The symbol Δ represents the discretized Laplacian operator and the matrix \mathbf{P}_i corresponds to a restriction operator for

the receiver locations. Finally, the vectors $\mathbf{q}_{i,j}$ and $\mathbf{u}_{i,j}$ denote the unknown monochromatic source term at the i^{th} source location and j^{th} frequency and the corresponding wavefield, respectively. For simplicity, we omit the dependence of $\mathbf{A}_j(\mathbf{m})$ on the discretized squared slowness vector \mathbf{m} from now onwards.

Since the \mathbf{A}_j 's are square matrices, we can eliminate the variable \mathbf{u} by solving the discretized partial-differential equation (PDE) explicitly, leading to the so-called adjoint-state method, which has the following reduced form (Virieux and Operto, 2009):

$$\underset{\mathbf{m}}{\text{minimize}} f_{\text{red}}(\mathbf{m}) = \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \|\mathbf{P}_i \mathbf{A}_j^{-1} \mathbf{q}_{i,j} - \mathbf{d}_{i,j}\|_2^2. \quad (2)$$

In this form, the source function \mathbf{q} is assumed to be available. However, in practice, while spatial locations of the sources are often known quite accurately, the source function is generally unknown. If we ignore spatial directivity of the sources, we can handle this situation in the context of FWI by incorporating the source function through setting $\mathbf{q}_{i,j} = \alpha_{i,j} \mathbf{e}_{i,j}$ in the frequency-domain, where $\alpha_{i,j}$ is a complex number representing the frequency-dependent source weight for the i^{th} source experiment at the j^{th} frequency. In this expression, the symbol $\mathbf{e}_{i,j}$ represents a unit vector that equals one at the source location and zero elsewhere. When we substitute $\mathbf{q}_{i,j} = \alpha_{i,j} \mathbf{e}_{i,j}$ into the objective function in equation 2, we arrive at an optimization problem with two unknown variables, i.e., we have

$$\underset{\mathbf{m}, \alpha}{\text{minimize}} \bar{f}_{\text{red}}(\mathbf{m}, \alpha) = \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \|\mathbf{P}_i \mathbf{A}_j^{-1} \alpha_{i,j} \mathbf{e}_{i,j} - \mathbf{d}_{i,j}\|_2^2, \quad (3)$$

where the vector $\alpha = \{\alpha_{i,j}\}_{1 \leq i \leq n_{\text{src}}, 1 \leq j \leq n_{\text{freq}}}$ collects all the unknown source weights.

To solve the optimization problem in equation 3, one straightforward approach is to compute the gradient of the objective function \bar{f}_{red} with respect to \mathbf{m} and α , i.e.,

$$\bar{\mathbf{g}}_{\text{red}}(\mathbf{m}, \alpha) = \begin{bmatrix} \nabla_{\mathbf{m}} \bar{f}_{\text{red}}(\mathbf{m}, \alpha) \\ \nabla_{\alpha} \bar{f}_{\text{red}}(\mathbf{m}, \alpha) \end{bmatrix}, \quad (4)$$

and update \mathbf{m} and α alternately (Zhou and Greenhalgh, 2003) or simultaneously (Liu et al., 2008). While relatively straightforward to compute, this formulation may result in converging to a local minimum, which requires a very good initial guess of α to avoid. Moreover, while the gradients $\nabla_{\mathbf{m}} \bar{f}_{\text{red}}(\mathbf{m}, \alpha)$ and $\nabla_{\alpha} \bar{f}_{\text{red}}(\mathbf{m}, \alpha)$ seemingly provide all information that we would need to drive the inversion, their amplitudes may also differ significantly in scale. An undesired consequence of this mismatch between the two gradients is that the optimization may tend to primarily update the variable with the larger gradient contribution resulting in small updates for the other variable. This behavior may slow down the convergence, and the small updates may lead to errors that end up in the other variable resulting in a poor solution. Although we might mitigate this problem by manually scaling the two variables and the corresponding gradients to the approximately same levels, the crosstalk between \mathbf{m} and α may lead to more local minima in the objective function, resulting in the requirement of a good initial guess of α to ensure a successful inversion (Li et al., 2013).

Source estimation with the variable projection method

An alternative approach to solving the optimization problem in equation 3 is to use the so-called variable projection method (Golub and Pereyra, 2003) to project out the source weights α during each iteration (Aravkin and van Leeuwen, 2012; Li et al., 2013). The variable projection method is designed to solve SNLLS problems, which permit the following standard expression:

$$\underset{\theta, \mathbf{x}}{\text{minimize}} f(\theta, \mathbf{x}) = \frac{1}{2} \|\Phi(\theta)\mathbf{x} - \mathbf{y}\|^2, \quad (5)$$

where the matrix $\Phi(\theta)$ varies (nonlinearly) with respect to the variable vector θ . The vectors \mathbf{x} and \mathbf{y} represent a linear variable and the observations, respectively. More specifically, we can observe that the FWI problem with source estimation in equation 3 is a special case of

SNLLS problems by setting

$$\theta = \mathbf{m}, \quad \mathbf{x}_{i,j} = \alpha_{i,j}, \quad \mathbf{y}_{i,j} = \mathbf{d}_{i,j}, \quad \text{and} \quad \Phi_{i,j}(\theta) = \mathbf{P}_i \mathbf{A}_j^{-1} \mathbf{e}_{i,j}, \quad (6)$$

for each pair of (i, j) . In this case, the vectors \mathbf{x} and \mathbf{y} are block vectors containing all $\{\mathbf{x}_{i,j}\}_{1 \leq i \leq n_{\text{src}}, 1 \leq j \leq n_{\text{freq}}}$ and $\{\mathbf{y}_{i,j}\}_{1 \leq i \leq n_{\text{src}}, 1 \leq j \leq n_{\text{freq}}}$. Likewise, $\Phi(\theta)$ is a block-diagonal matrix.

When introducing the variable projection method, it is often useful to recognize that for fixed θ , the objective function $f(\theta, \mathbf{x})$ becomes quadratic with respect to \mathbf{x} . As a consequence, the variable \mathbf{x} has a closed-form optimal least-squares solution $\bar{\mathbf{x}}(\theta)$ that minimizes the objective $f(\theta, \mathbf{x})$:

$$\bar{\mathbf{x}}(\theta) = (\Phi(\theta)^\top \Phi(\theta))^{-1} \Phi(\theta)^\top \mathbf{y}, \quad (7)$$

where the symbol $^\top$ denotes the (complex-conjugate) transpose. For each pair of (i, j) , we now have

$$\bar{\mathbf{x}}_{i,j}(\theta) = (\Phi_{i,j}(\theta)^\top \Phi_{i,j}(\theta))^{-1} \Phi_{i,j}(\theta)^\top \mathbf{y}_{i,j}. \quad (8)$$

By substituting equation 7 into equation 5, we project out the variable \mathbf{x} by the optimal solution $\bar{\mathbf{x}}(\theta)$ and arrive at a reduced optimization problem over θ alone, i.e., we have

$$\underset{\theta}{\text{minimize}} \bar{f}(\theta) = f(\theta, \bar{\mathbf{x}}(\theta)) = \left\| \left(\mathbf{I} - \Phi(\theta) (\Phi(\theta)^\top \Phi(\theta))^{-1} \Phi(\theta)^\top \right) \mathbf{y} \right\|^2. \quad (9)$$

The gradient of the reduced objective function $\bar{f}(\theta)$ can be computed via the chain rule, yielding

$$\bar{\mathbf{g}}(\theta) = \nabla_\theta \bar{f}(\theta) = \nabla_\theta f(\theta, \bar{\mathbf{x}}(\theta)) = \nabla_\theta f(\theta, \mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}(\theta)} + \nabla_{\mathbf{x}} f(\theta, \mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}(\theta)} \nabla_\theta \bar{\mathbf{x}}. \quad (10)$$

Note that from the definition of $\bar{\mathbf{x}}(\theta)$, we have

$$\nabla_{\mathbf{x}} f(\theta, \mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}(\theta)} = 0. \quad (11)$$

With this equality, the gradient in equation 10 can be written as:

$$\bar{\mathbf{g}}(\theta) = \nabla_{\theta} f(\theta, \mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}(\theta)}. \quad (12)$$

Equation 12 implies that although the optimal solution $\bar{\mathbf{x}}(\theta)$ is a function of θ , the construction of the gradient $\bar{\mathbf{g}}(\theta)$ does not need to include the cross-derivative term $\nabla_{\mathbf{x}} f(\theta, \mathbf{x})|_{\mathbf{x}=\bar{\mathbf{x}}(\theta)} \nabla_{\theta} \mathbf{x}$ by the virtue of equation 11.

As shown by Aravkin and van Leeuwen (2012) and van Leeuwen et al. (2014), we can through substituting equation 6 into equations 7, 9, and 12 successfully project out the source weights by computing the optimal source weight $\alpha_{i,j}(\mathbf{m})$ for each source and frequency:

$$\bar{\alpha}_{i,j}(\mathbf{m}) = \frac{\bar{\mathbf{d}}_{i,j}^{\top} \mathbf{d}_{i,j}}{\mathbf{d}_{i,j}^{\top} \bar{\mathbf{d}}_{i,j}}, \quad \text{with} \quad \bar{\mathbf{d}}_{i,j} = \mathbf{P}_i \mathbf{A}_j^{-1} \mathbf{e}_{i,j}, \quad (13)$$

and arrive at minimizing the following reduced objective:

$$\underset{\mathbf{m}}{\text{minimize}} \tilde{f}_{\text{red}}(\mathbf{m}) = \bar{f}_{\text{red}}(\mathbf{m}, \bar{\alpha}(\mathbf{m})), \quad (14)$$

with a gradient given by

$$\tilde{\mathbf{g}}_{\text{red}}(\mathbf{m}) = \nabla_{\mathbf{m}} \bar{f}_{\text{red}}(\mathbf{m}, \alpha)|_{\alpha=\bar{\alpha}(\mathbf{m})}. \quad (15)$$

Compared to the objective function $\bar{f}_{\text{red}}(\mathbf{m}, \alpha)$ in equation 3, the objective function $\tilde{f}_{\text{red}}(\mathbf{m})$ only depends on \mathbf{m} . As a consequence, we avoid having to optimize over two different variables that differ in scale. Moreover, because the objective $\tilde{f}_{\text{red}}(\mathbf{m})$ is source-independent, the optimization with it does not require any initial guesses of the source weights, while the optimization with the objective $\bar{f}(\mathbf{m}, \alpha)$ does need a sufficiently accurate initial guess to ensure reaching the correct solution. Empirically, Li et al. (2013) showed the outperformances of minimizing $\tilde{f}_{\text{red}}(\mathbf{m})$ over minimizing $\bar{f}(\mathbf{m}, \alpha)$, especially when the data contain noise.

WRI with source estimation

Introduction of WRI

In the conventional FWI or adjoint-state method in equation 2, the dependence of the objective function $f_{\text{red}}(\mathbf{m})$ on \mathbf{m} runs through the nonlinear operator $\mathbf{A}_j^{-1}\mathbf{q}_{i,j}$, instead of via the original linear operator $\mathbf{A}_j\mathbf{u}_{i,j}$ in equation 1. As a result, the price we pay is that the objective function $f_{\text{red}}(\mathbf{m})$ becomes highly nonlinear in \mathbf{m} . The gain, of course, is that we no longer have to optimize over the wavefields. However, this gain is perhaps a bit short-lived, because the reduced objective function $f_{\text{red}}(\mathbf{m})$ may contain local minima (Warner et al., 2013), which introduces more difficulties in the search for the best medium parameters using the local derivative information only.

To make FWI less susceptible to these parasitic minima, van Leeuwen and Herrmann (2013b) and van Leeuwen and Herrmann (2015) proposed WRI—a penalty formulation of FWI that reads

$$\underset{\mathbf{m}, \mathbf{u}}{\text{minimize}} f_{\text{pen}}(\mathbf{m}, \mathbf{u}) = \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} (\|\mathbf{P}_i\mathbf{u}_{i,j} - \mathbf{d}_{i,j}\|_2^2 + \lambda^2 \|\mathbf{A}_j\mathbf{u}_{i,j} - \mathbf{q}_{i,j}\|_2^2). \quad (16)$$

In this optimization problem, we keep the wavefields \mathbf{u} as unknown variables instead of eliminating them as in FWI, i.e., we replace the PDE constraint by an ℓ_2 -norm penalty term. The scalar parameter λ controls the balance between the data and the PDE misfits. As λ increases, the wavefield is more tightly constrained by the wave-equation, and the objective function $f_{\text{pen}}(\mathbf{m}, \mathbf{u})$ in equation 16 converges to the objective function of the reduced problem in equation 2 (van Leeuwen and Herrmann, 2015). Different from the nonlinear objective function of the reduced problem in equation 2, the WRI objective $f_{\text{pen}}(\mathbf{m}, \mathbf{u})$ is a bi-quadratic function with respect to \mathbf{m} and \mathbf{u} , i.e., for fixed \mathbf{m} , the objective $f_{\text{pen}}(\mathbf{m}, \mathbf{u})$ is quadratic

with respect to \mathbf{u} and vice versa. In addition, van Leeuwen and Herrmann (2013b) and van Leeuwen and Herrmann (2015) indicated that WRI explores a larger search space by not satisfying the PDE constraints exactly, which results in that WRI is less “nonlinear” and may contain less local minima compared to the conventional adjoint-state method.

In the optimization problem of equation 16, both the wavefields \mathbf{u} and medium parameters \mathbf{m} are unknown. Searching for both \mathbf{u} and \mathbf{m} by methods like gradient-descent and l-BFGS method requires storing the two unknown variables. However, the memory cost of storing \mathbf{u} can be extremely large, because $\mathbf{u} \in \mathcal{C}^{n_{\text{grid}} \times n_{\text{freq}} \times n_{\text{src}}}$. To mitigate the storage cost, van Leeuwen and Herrmann (2013b) and van Leeuwen and Herrmann (2015) noticed that the optimization problem in equation 16 is also a special case of SNLLS problems and applied the variable projection method to solve it. Indeed, we have

$$\theta = \mathbf{m}, \quad \mathbf{x}_{i,j} = \mathbf{u}_{i,j}, \quad \mathbf{y}_{i,j} = \begin{bmatrix} \lambda \mathbf{q}_{i,j} \\ \mathbf{d}_{i,j} \end{bmatrix}, \quad \text{and} \quad \Phi_{i,j}(\theta) = \begin{bmatrix} \lambda \mathbf{A}_j \\ \mathbf{P}_i \end{bmatrix}. \quad (17)$$

Therefore, after substituting equation 17 into equations 7, 9, and 12, we can successfully project out the wavefields \mathbf{u} by computing the optimal wavefield $\bar{\mathbf{u}}_{i,j}(\mathbf{m})$ for each source and frequency:

$$\bar{\mathbf{u}}_{i,j}(\mathbf{m}) = \left(\lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}_i^\top \mathbf{P}_i \right)^{-1} \left(\lambda^2 \mathbf{A}_j^\top \mathbf{q}_{i,j} + \mathbf{P}_i^\top \mathbf{d}_{i,j} \right), \quad (18)$$

and arrive at minimizing the following reduced objective:

$$\begin{aligned} \underset{\mathbf{m}}{\text{minimize}} \bar{f}_{\text{pen}}(\mathbf{m}) &= f_{\text{pen}}(\mathbf{m}, \bar{\mathbf{u}}(\mathbf{m})) \\ &= \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \left(\|\mathbf{P}_i \bar{\mathbf{u}}_{i,j}(\mathbf{m}) - \mathbf{d}_{i,j}\|_2^2 + \lambda^2 \|\mathbf{A}_j \bar{\mathbf{u}}_{i,j}(\mathbf{m}) - \mathbf{q}_{i,j}\|_2^2 \right), \end{aligned} \quad (19)$$

with a gradient given by

$$\bar{\mathbf{g}}_{\text{pen}}(\mathbf{m}) = \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \lambda^2 \omega_i^2 \text{diag} \left(\text{conj}(\bar{\mathbf{u}}_{i,j}(\mathbf{m})) \right) \left(\mathbf{A}_j \bar{\mathbf{u}}_{i,j}(\mathbf{m}) - \mathbf{q}_{i,j} \right). \quad (20)$$

Here the term $\text{diag}\left(\text{conj}(\bar{\mathbf{u}}_{i,j}(\mathbf{m}))\right)$ represents a diagonal matrix with the elements of the complex conjugate of the vector $\bar{\mathbf{u}}_{i,j}(\mathbf{m})$ on the diagonal. With the projection of the wavefields \mathbf{u} in equation 18, the new reduced objective $\bar{f}_{\text{pen}}(\mathbf{m})$ varies only with respect to the variable \mathbf{m} . As a result, we do not need to store the \mathbf{u} 's reducing the storage costs from $\mathcal{O}(n_{\text{grid}} \times (n_{\text{freq}} \times n_{\text{src}} + 1))$ to $\mathcal{O}(n_{\text{grid}})$.

Source estimation for WRI

The standard formulation of WRI in equation 16 does not take situations with unknown source functions into account. When the source functions are unknown, we face to solve an optimization problem with three unknown variables, i.e., we have

$$\underset{\mathbf{m}, \mathbf{u}, \alpha}{\text{minimize}} f_{\text{pen}}(\mathbf{m}, \mathbf{u}, \alpha) = \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \left(\|\mathbf{P}_i \mathbf{u}_{i,j} - \mathbf{d}_{i,j}\|_2^2 + \lambda^2 \|\mathbf{A}_j \mathbf{u}_{i,j} - \alpha_{i,j} \mathbf{e}_{i,j}\|_2^2 \right), \quad (21)$$

where the vector $\alpha = \{\alpha_{i,j}\}_{1 \leq i \leq n_{\text{src}}, 1 \leq j \leq n_{\text{freq}}}$ contains all the source weights.

Because this new objective $f_{\text{pen}}(\mathbf{m}, \mathbf{u}, \alpha)$ contains more than two unknowns, it no longer corresponds to a standard SNLLS problem, so we cannot directly apply the variable projection method to solve it. One simple approach that comes to mind to convert the problem in equation 21 to a standard SNLLS problem would be to reduce the number of unknown variables by collecting two of the three variables into a new variable. Since for fixed \mathbf{m} and α , $f_{\text{pen}}(\mathbf{m}, \mathbf{u}, \alpha)$ is still quadratic with respect to \mathbf{u} , the most straightforward approach would be to lump the medium parameters \mathbf{m} and the source weights α together. During each iteration, we can apply the variable projection method to project out \mathbf{u} , yielding an objective function $\tilde{f}(\mathbf{m}, \alpha)$ with respect to both \mathbf{m} and α . However, as discussed in the context of FWI with source estimation, minimizing $\tilde{f}(\mathbf{m}, \alpha)$ by either simultaneously or alternately updating \mathbf{m} and α would face the convergence-related problems caused by the

crosstalk between the two unknowns and the mismatch of their amplitudes. Additionally, a straightforward application of the variable projection method to project out α from $\tilde{f}(\mathbf{m}, \alpha)$ is not viable, because it destroys the underlying assumption of the projection of \mathbf{u} , i.e., $\nabla_{\mathbf{u}} f_{\text{pen}}(\mathbf{m}, \mathbf{u}, \alpha)|_{\mathbf{u}=\bar{\mathbf{u}}(\mathbf{m}, \alpha)} \neq 0$ with the projected α .

To address the challenge arising from the different amplitude scales of \mathbf{m} and α and their gradients, we propose another strategy to reduce the original problem in equation 21. Instead of collecting \mathbf{m} and α together, we lump \mathbf{u} and α into a single vector \mathbf{x} . To understand the potential advantage of this alternative strategy, let us rewrite the optimization problem in equation 21 as

$$\begin{aligned}
\underset{\mathbf{m}, \mathbf{u}, \alpha}{\text{minimize}} f_{\text{pen}}(\mathbf{m}, \mathbf{u}, \alpha) &= \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \left\| \begin{bmatrix} \lambda \mathbf{A}_j & -\lambda \mathbf{e}_{i,j} \\ \mathbf{P}_i & 0 \end{bmatrix} \begin{bmatrix} \mathbf{u}_{i,j} \\ \alpha_{i,j} \end{bmatrix} - \begin{bmatrix} 0 \\ \mathbf{d}_{i,j} \end{bmatrix} \right\|^2 \\
&= \frac{1}{2} \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \left\| \begin{bmatrix} \lambda \mathbf{A}_j & -\lambda \mathbf{e}_{i,j} \\ \mathbf{P}_i & 0 \end{bmatrix} \mathbf{x}_{i,j} - \begin{bmatrix} 0 \\ \mathbf{d}_{i,j} \end{bmatrix} \right\|^2 \\
&:= \hat{f}_{\text{pen}}(\mathbf{m}, \mathbf{x}).
\end{aligned} \tag{22}$$

For fixed \mathbf{m} , the objective function $\hat{f}_{\text{pen}}(\mathbf{m}, \mathbf{x})$ continues to be quadratic in the new variable \mathbf{x} , where \mathbf{u} and α are grouped together. As a result, we can rewrite the optimization problem in equation 22 into the standard SNLLS form (see equation 5) when we use the following notations:

$$\theta = \mathbf{m}, \quad \mathbf{x}_{i,j} = \begin{bmatrix} \mathbf{u}_{i,j} \\ \alpha_{i,j} \end{bmatrix}, \quad \mathbf{y}_{i,j} = \begin{bmatrix} \mathbf{0} \\ \mathbf{d}_{i,j} \end{bmatrix}, \quad \text{and} \quad \Phi_{i,j}(\theta) = \begin{bmatrix} \lambda \mathbf{A}_j & -\lambda \mathbf{e}_{i,j} \\ \mathbf{P}_i & \mathbf{0} \end{bmatrix}. \tag{23}$$

Now by substituting the notations in equation 23 into equation 8, we can apply the variable projection method again to project out \mathbf{x} from the objective function $\hat{f}_{\text{pen}}(\mathbf{m}, \mathbf{x})$ for fixed \mathbf{m}

by computing the optimal solution

$$\bar{\mathbf{x}}_{i,j}(\mathbf{m}) = \begin{bmatrix} \bar{\mathbf{u}}_{i,j}(\mathbf{m}) \\ \bar{\alpha}_{i,j}(\mathbf{m}) \end{bmatrix} = \begin{bmatrix} \lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}_i^\top \mathbf{P}_i & -\lambda^2 \mathbf{A}_j^\top \mathbf{e}_{i,j} \\ -\lambda^2 \mathbf{e}_{i,j}^\top \mathbf{A}_j & \lambda^2 \mathbf{e}_{i,j}^\top \mathbf{e}_{i,j} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{P}_i \mathbf{d}_{i,j} \\ \mathbf{0} \end{bmatrix}. \quad (24)$$

The linear system in equation 24 is solvable if and only if the rank of the matrix $\Phi_{i,j}(\theta)$ in equation 23 is $n_{\text{grid}} + 1$, which requires

$$\mathbf{P}_i \mathbf{A}_j^{-1} \mathbf{e}_{i,j} \neq \mathbf{0}. \quad (25)$$

In fact, this means that the wavefields corresponding to a point source do not vanish at all the receiver locations simultaneously, which holds in almost all seismic situations. Indeed, even in the extremely rare situation where a source is made nearly “silent” or non-radiating due to its surrounding geological structures, we can simply discard this source in the inversion to avoid inverting a rank deficient matrix. After the projection in equation 24, we obtain an objective function that only depends on \mathbf{m} :

$$\hat{f}(\mathbf{m}) = f_{\text{pen}}(\mathbf{m}, \bar{\mathbf{u}}(\mathbf{m}), \bar{\alpha}(\mathbf{m})), \quad (26)$$

whose gradient $\hat{\mathbf{g}}(\mathbf{m})$ can be derived as

$$\begin{aligned} \hat{\mathbf{g}}(\mathbf{m}) &= \nabla_{\mathbf{m}} \hat{f}(\mathbf{m}) = \nabla_{\mathbf{m}} f_{\text{pen}}(\mathbf{m}, \mathbf{u}, \alpha) \Big|_{\mathbf{u}=\bar{\mathbf{u}}(\mathbf{m}), \alpha=\bar{\alpha}(\mathbf{m})} \\ &= \sum_{i=1}^{n_{\text{src}}} \sum_{j=1}^{n_{\text{freq}}} \lambda^2 \omega_i^2 \text{diag}(\text{conj}(\bar{\mathbf{u}}_{i,j}(\mathbf{m}))) (\mathbf{A}_j \bar{\mathbf{u}}_{i,j}(\mathbf{m}) - \bar{\alpha}_{i,j}(\mathbf{m}) \mathbf{e}_{i,j}). \end{aligned} \quad (27)$$

As a result, we avoid optimizing over two different variables that differ in scale. Moreover, the optimization with $\hat{f}(\mathbf{m})$ does not require any initial guesses of the source weights.

Optimization scheme

With the gradient $\hat{\mathbf{g}}(\mathbf{m})$ we derived in the previous section we can compute updates for the medium parameters via the steepest-descent method (Nocedal and Wright, 2006):

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \beta_k \hat{\mathbf{g}}(\mathbf{m}_k). \quad (28)$$

Here, the value β_k is an appropriately chosen step length at the k^{th} iteration, which requires a line search to determine. Even with an optimized step length, as a first-order method, the steepest-descent method can be slow. To speed up the convergence, second-order methods including Newton method, Gauss-Newton method, and quasi-Newton method may be more desirable (Pratt, 1999; Akcelik, 2002; Askan et al., 2007). During each iteration, Newton method and Gauss-Newton method use the full or Gauss-Newton Hessian to compute the update direction. However, the construction and inversion of the full or Gauss-Newton Hessian for both FWI (Pratt, 1999) and WRI (van Leeuwen and Herrmann, 2015) involve a large amount of additional PDE solves, which makes these two methods less attractive in this context. On the other hand, the quasi-Newton method, especially the l-BFGS method (Nocedal and Wright, 2006; Brossier et al., 2009; van Leeuwen and Herrmann, 2013a), utilizes the gradient at the current iteration k and a few previous gradients (typically, 3-20 iterations) to construct an approximation of the inverse of the Hessian. Hence, except for the increased memory, the l-BFGS method constructs the approximation of the inverse of the Hessian for free. For this reason, we use the l-BFGS method as our optimization scheme and denote the l-BFGS update as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \beta_k \mathbf{H}(\mathbf{m}_k)^{-1} \hat{\mathbf{g}}(\mathbf{m}_k), \quad (29)$$

where the matrix $\mathbf{H}(\mathbf{m}_k)$ is the l-BFGS Hessian and the step length β_k is determined by a weak Wolfe line search. We give a detailed description of the l-BFGS method for the proposed WRI framework with source estimation in Algorithm 1.

Algorithm 1 l-BFGS algorithm for WRI with source estimation

1. Initialization with an initial model \mathbf{m}_0
 2. **for** $k = 1 \rightarrow n_{\text{iter}}$
 3. **for** $j = 1 \rightarrow n_{\text{freq}}$
 4. **for** $i = 1 \rightarrow n_{\text{src}}$
 5. Compute $\bar{\mathbf{u}}_{i,j}(\mathbf{m}_k)$ and $\bar{\alpha}_{i,j}(\mathbf{m}_k)$ by equation 24
 6. **end**
 7. **end**
 8. Compute \hat{f}_k and $\hat{\mathbf{g}}_k$ by equations 26 and 27
 9. Apply l-BFGS Hessian with history size n_{his}
 10. $\mathbf{p}_k = \text{lbgfs}(-\hat{\mathbf{g}}_k, \{\mathbf{t}_l\}_{l=k-n_{\text{his}}}^k, \{\mathbf{s}_l\}_{l=k-n_{\text{his}}}^k)$
 11. $\{\mathbf{m}_{k+1}, \hat{f}_{k+1}, \hat{\mathbf{g}}_{k+1}\} = \text{line search}(\hat{f}_k, \hat{\mathbf{g}}_k, \mathbf{p}_k)$
 12. $\mathbf{t}_{k+1} = \mathbf{m}_{k+1} - \mathbf{m}_k$
 13. $\mathbf{s}_{k+1} = \hat{\mathbf{g}}_{k+1} - \hat{\mathbf{g}}_k$
 14. **end**
-

Fast solver for WRI with source estimation

Most of the computational burden for the objective function and its gradient in equations 26 and 27 lies within inverting the matrix

$$\mathbf{C}_{i,j}(\mathbf{m}) = \Phi_{i,j}(\mathbf{m})^\top \Phi_{i,j}(\mathbf{m}) = \begin{bmatrix} \lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}_i^\top \mathbf{P}_i & -\lambda^2 \mathbf{A}_j^\top \mathbf{e}_{i,j} \\ -\lambda^2 \mathbf{e}_{i,j}^\top \mathbf{A}_j & \lambda^2 \mathbf{e}_{i,j}^\top \mathbf{e}_{i,j} \end{bmatrix}. \quad (30)$$

For 2D problems, we can invert the matrix $\mathbf{C}_{i,j}(\mathbf{m})$ by direct solvers such as the Cholesky method, while for 3D problems, we may need an iterative solver equipped with a proper preconditioner. In either case, following computational challenges arise from the fact that the matrix $\mathbf{C}_{i,j}(\mathbf{m})$ differs from source to source and frequency to frequency. First, to

apply the Cholesky method, we need to calculate the Cholesky factorization for each i and j , i.e., for each source and frequency. As a result, the computational cost arrives at $\mathcal{O}(n_{\text{src}} \times n_{\text{freq}} \times (n_{\text{grid}}^3 + n_{\text{grid}}^2))$, which prohibits the application to problems with a large number of sources. Second, for iterative solvers, because the matrix $\mathbf{C}_{i,j}(\mathbf{m})$ varies with respect to i and j , the corresponding preconditioner may as well as depend on i and j . Therefore, we may have to design $n_{\text{src}} \times n_{\text{freq}}$ different preconditioners, which can be computationally difficult and intractable. Moreover, the additional terms $-\lambda^2 \mathbf{A}_j^\top \mathbf{e}_{i,j}$, $-\lambda^2 \mathbf{e}_{i,j}^\top \mathbf{A}_j$, and $\lambda^2 \mathbf{e}_{i,j}^\top \mathbf{e}_{i,j}$ may lead to the condition number of the matrix $\mathbf{C}_{i,j}$ worse than that of the matrix $\lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}_i^\top \mathbf{P}_i$. As a result, without a good preconditioner, the projection procedure in the framework of WRI with source estimation may be much slower than that without source estimation.

To lighten the computational cost of inverting the matrix $\mathbf{C}_{i,j}(\mathbf{m})$, we describe a new inversion scheme to implement the algorithm when the projection operator \mathbf{P} remains the same for all sources. This situation is common in ocean bottom node acquisition and the land acquisition where all sources “see” the same receivers. When we replace \mathbf{P}_i by a single \mathbf{P} in equation 30, the matrix $\mathbf{C}_{i,j}(\mathbf{m})$ can be simplified as

$$\mathbf{C}_{i,j}(\mathbf{m}) = \begin{bmatrix} \lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}^\top \mathbf{P} & -\lambda^2 \mathbf{A}_j^\top \mathbf{e}_{i,j} \\ -\lambda^2 \mathbf{e}_{i,j}^\top \mathbf{A}_j & \lambda^2 \mathbf{e}_{i,j}^\top \mathbf{e}_{i,j} \end{bmatrix}. \quad (31)$$

Unfortunately, this matrix $\mathbf{C}_{i,j}(\mathbf{m})$ still depends on the source and frequency indices, and a straightforward inversion still faces the aforementioned computational challenges. However, this form, equation 31, allows us to use an alternative block matrix formula to invert $\mathbf{C}_{i,j}(\mathbf{m})$.

To arrive at this result, let us first write $\mathbf{C}_{i,j}(\mathbf{m})$ in a simpler form

$$\mathbf{C}_{i,j}(\mathbf{m}) = \begin{bmatrix} \mathbf{M}_1 & \mathbf{M}_2 \\ \mathbf{M}_3 & \mathbf{M}_4 \end{bmatrix}, \quad (32)$$

where the matrix $\mathbf{M}_1 = \lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}^\top \mathbf{P}$, the vector $\mathbf{M}_2 = \mathbf{M}_3^\top = -\lambda^2 \mathbf{A}_j^\top \mathbf{e}_{i,j}$, and the scalar $\mathbf{M}_4 = \lambda^2 \mathbf{e}_{i,j}^\top \mathbf{e}_{i,j}$. Now if we apply the block matrix inversion formula (Bernstein, 2005) to compute $\mathbf{C}_{i,j}^{-1}(\mathbf{m})$, we arrive at the following closed form analytical expression

$$\mathbf{C}_{i,j}^{-1}(\mathbf{m}) = \begin{bmatrix} (\mathbf{I} + \mathbf{M}_1^{-1} \mathbf{M}_2 (\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1} \mathbf{M}_3) \mathbf{M}_1^{-1} & -\mathbf{M}_1^{-1} \mathbf{M}_2 (\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1} \\ -(\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1} \mathbf{M}_3 \mathbf{M}_1^{-1} & (\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1} \end{bmatrix}. \quad (33)$$

While complicated the right-hand side of this equation only involves the inversion of the source-independent matrix \mathbf{M}_1 , all other terms are all scalar inversions that can be evaluated at negligible costs. Substituting equation 33 into equation 24, we obtain an analytical solution for the optimal variable

$$\begin{aligned} \bar{\mathbf{x}}_{i,j}(\mathbf{m}) &= \mathbf{C}_{i,j}^{-1}(\mathbf{m}) \begin{bmatrix} \mathbf{P}^\top \mathbf{d}_{i,j} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{I} + \mathbf{M}_1^{-1} \mathbf{M}_2 (\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1} \mathbf{M}_3) \mathbf{M}_1^{-1} \mathbf{P}^\top \mathbf{d}_{i,j} \\ -(\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1} \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{P}^\top \mathbf{d}_{i,j} \end{bmatrix}. \end{aligned} \quad (34)$$

In equation 34, we have to invert two terms, i.e., \mathbf{M}_1 and $\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2$. Because \mathbf{M}_4 is a scalar, the term $(\mathbf{M}_4 - \mathbf{M}_3 \mathbf{M}_1^{-1} \mathbf{M}_2)^{-1}$ is a scalar inversion, whose computational cost is negligible. Therefore, to construct $\bar{\mathbf{x}}_{i,j}(\mathbf{m})$, we only need to compute the following two vectors:

$$\begin{aligned} \mathbf{w}_1 &= \mathbf{M}_1^{-1} \mathbf{M}_2 \\ &= -\lambda^2 (\lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}^\top \mathbf{P})^{-1} \mathbf{A}_j^\top \mathbf{e}_{i,j}, \end{aligned} \quad (35)$$

and

$$\begin{aligned} \mathbf{w}_2 &= \mathbf{M}_1^{-1} \mathbf{P}^\top \mathbf{d}_{i,j} \\ &= (\lambda^2 \mathbf{A}_j^\top \mathbf{A}_j + \mathbf{P}^\top \mathbf{P})^{-1} \mathbf{P}^\top \mathbf{d}_{i,j}. \end{aligned} \quad (36)$$

From these two expressions, we can observe that the computation of the vectors \mathbf{w}_1 and \mathbf{w}_2 for each source is independent from that of other sources, therefore, we can sequentially

compute them from source to source, yielding a negligible requirement of storing only 2 additional wavefields during the inversion. For each frequency, because the matrix \mathbf{M}_1 is source-independent, we only need one Cholesky factorization, whose computational cost is $\mathcal{O}(n_{\text{grid}}^3)$. With the pre-computed Cholesky factors, for each source, solving \mathbf{w}_1 and \mathbf{w}_2 by equations 35 and 36 requires inverting the matrix \mathbf{M}_1 twice with a computational cost of $\mathcal{O}(n_{\text{grid}}^2)$. Consequently, we reduce the total cost from $\mathcal{O}(n_{\text{src}} \times n_{\text{freq}} \times (n_{\text{grid}}^3 + n_{\text{grid}}^2))$ to $\mathcal{O}(n_{\text{freq}} \times (n_{\text{grid}}^3 + 2 \times n_{\text{src}} \times n_{\text{grid}}^2))$. As a reference, the computational complexity of WRI without source estimation is of a close order $\mathcal{O}(n_{\text{freq}} \times (n_{\text{grid}}^3 + n_{\text{src}} \times n_{\text{grid}}^2))$ (van Leeuwen and Herrmann, 2015). A comparison of the computational complexity of WRI with and without source estimation using the Cholesky method is shown in Table 1. In this table, we refer to the scheme that directly inverts the matrix $\mathbf{C}_{i,j}(\mathbf{m})$ as the old form, and refer to the proposed scheme in equation 34 as the new form. Compared to the old form, instead of requiring n_{src} Cholesky factorizations, the proposed new form only requires 1 Cholesky factorization for each frequency, which significantly reduces the computational cost. Furthermore, besides reducing the cost for direct solvers, the proposed inversion scheme can also benefit iterative solvers. For each frequency, since all sources only need to invert the same matrix \mathbf{M}_1 , the proposed new form avoids inverting the potentially ill-conditioned matrix $\mathbf{C}_{i,j}$ directly and only requires one preconditioner instead of n_{src} , which significantly simplifies the projection procedure. In the following section of numerical experiments, we will only show the computational gain for direct solvers.

[Table 1 about here.]

NUMERICAL EXAMPLES

BG model with frequency-domain data

To compare the performance of the two source estimation methods for WRI described in the previous section, we conduct numerical experiments on a part of the BG compass model \mathbf{m}_t shown in Figure 2a, which is a geologically realistic model created by BG Group and has been widely used to evaluate performances of different FWI methods (Li et al., 2012; van Leeuwen and Herrmann, 2013a). We will refer to the method that combines \mathbf{m} and α as WRI-SE-MS and the one that combines \mathbf{u} and α as WRI-SE-WS. For our discretization, we use an optimal 9-point finite-difference frequency-domain forward modeling code (Chen et al., 2013). Sources and receivers are positioned at the depth of $z = 20$ m with a source distance of 50 m and a receiver distance of 10 m, resulting a maximum offset of 4.5 km. Data is generated with a source function given by a Ricker wavelet centered at 20 Hz with a time shift of 0.5 s. We do not model the free-surface, so the data contain no free-surface related multiples. As is commonly practiced (see e.g. Pratt (1999)), we perform frequency continuation using frequency bands ranging from $\{2, 2.5, 3, 3.5, 4, 4.5\}$ Hz to $\{27, 27.5, 28, 28.5, 29, 29.5\}$ Hz with an overlap of one frequency between every two consecutive bands. During the inversion, the result of the current frequency band serves as a warm starter for the inversion of the next frequency band. During each iteration, we also apply a point-wise bound constraint (Peters and Herrmann, 2016) to the update to ensure that each gridpoint of the model update remains within a geologically reasonable interval. In this experiment, because the lowest and highest velocities of the true model are 1480 m/s and 4500 m/s, we set the interval of the bound constraint to be $[1300, 5000]$ m/s. The initial model \mathbf{m}_0 (Figure 2b) is generated by smoothing the original model, followed by an average along the horizontal direction. The

difference between the initial and true models is shown in Figure 2c. We use this initial model and apply the source estimation method proposed by Pratt (1999) to obtain an initial guess of the source weights for WRI-SE-MS. Because this guess minimizes the difference between the observed and predicted data with the initial model, it can be considered as the best choice, when there is no more additional information. For WRI-SE-WS, as described in the previous sections, it does not need an initial guess of the source weights. To control the computational load, we fix the maximum number of l-BFGS iterations to be 20 for each frequency band. We select the penalty parameter $\lambda = 1e4$ according to the selection criteria in van Leeuwen and Herrmann (2015) that optimizes the performance of WRI.

[Figure 2 about here.]

As a benchmark, we depict the inversion result obtained with the true source weights in Figure 3a. Figures 3b and 3c show the inversion results obtained by WRI-SE-MS and WRI-SE-WS, respectively. Figure 3d shows the difference between the results in Figures 3a and 3b, and Figure 3e displays the difference between the results in Figures 3a and 3c. We observe that while both the results obtained by WRI-SE-MS and WRI-SE-WS are close to the result obtained with the true source weights, the difference shown in Figure 3d is almost 10 times larger than that shown in Figure 3e. In addition, we compare the true source weights and the estimated source weights obtained with WRI-SE-MS and WRI-SE-WS in Figures 4a (phase) and 4b (amplitude). From these two figures, we observe that both methods can provide reasonable estimates of the source weights while the estimates with WRI-SE-WS contain smaller errors. These errors found in the source weights may result in the large model differences shown in Figure 3d. In Figure 5, we display the relative model errors $\frac{\|\mathbf{m}_t - \mathbf{m}\|}{\|\mathbf{m}_t - \mathbf{m}_0\|}$ versus the number of iterations during the three inversions. The dashed, solid, and

dotted curves correspond to inversions with the true source weights and those estimates using WRI-SE-MS and WRI-SE-WS. The relative model errors of WRI-SE-WS are almost the same as those using the true source weights, while the errors of WRI-SE-MS are clearly larger than the other ones. Figure 6 depicts a comparison of the data residuals corresponding to the inversion results of WRI-SE-MS and WRI-SE-WS. Clearly, the data residual of WRI-SE-WS is much smaller than that of WRI-SE-MS. In Table 2, we quantitatively compare the inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals, which also illustrates the outperformances of WRI-SE-WS over WRI-SE-MS. These observations imply that compared to iteratively updating α , projecting out α and \mathbf{u} together during each iteration can provide more accurate estimates of source weights, which further benefits the inversions of medium parameters.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Table 2 about here.]

To evaluate the computational performance of the fast inversion scheme of WRI-SE-WS, we compare the time spent in evaluating one objective for three cases, namely (i) WRI without source estimation; (ii) WRI-SE-WS with the old form, equation 24; and (iii) WRI-SE-WS with the new form, equation 34. The computational time for each of the three cases is 94s,

2962 s, and 190 s, respectively. As expected, case (iii) spends twice the amount of time as case (i), because of the additional PDE solves for each frequency and each source. As only one Cholesky factorization is required for case (iii), the computational time is only 1/15 of that for case (ii). This result illustrates that the proposed approach can estimate source weights in the context of WRI with a small computational overhead.

BG model with time-domain data

To test the robustness of the two source estimation techniques in less ideal situations, we perform the inversion tests on non-inverse-crime data. For this purpose, we generate time-domain data with a recording time of 4 seconds using iWAVE (Symes et al., 2011) and transform the data from the time domain to the frequency domain for the inversions. The data are generated on uniform grids with a grid size of 6 m, while the inversions are carried out on uniform grids with a coarser grid of 10 m. As a result, modeling errors arise from the differences between the modeling kernels and the grid sizes. All other experimental settings in this example are the same as example 1. Figures 7a and 7b show the inversion results with WRI-SE-MS and WRI-SE-WS, respectively. From Figure 7a, we can observe that method WRI-SE-MS fails to converge to a reasonable solution, as there is a large low-velocity block at the depth of $z = 1500$ m, which does not exist in the true model. On the other hand, the result of WRI-SE-WS (Figure 7b) is more consistent with the true model. The average model errors presented in the inversion results of WRI-SE-MS and WRI-SE-WS are 0.18 km/s and 0.11 km/s, respectively. Moreover, compared to the true source weights, the amplitudes of the estimated weights obtained with WRI-SE-MS contain very large errors, while the results obtained with WRI-SE-WS are almost identical to the true source weights (see Figure 8). Additionally, the residual between the observed and predicted data computed by the inversion

results of WRI-SE-WS is much smaller than that of WRI-SE-MS (see Figure 9). We can also obtain similar observations from the quantitative comparison presented in Table 3. These results imply that compared to WRI-SE-MS, WRI-SE-WS is more robust with respect to the modeling errors.

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

[Table 3 about here.]

BG model with noisy data

To test the stability of the proposed two techniques with respect to measurement noise in data, we add 40% Gaussian noise into the data used in example 2. Again, other experimental settings remain the same as in example 2. As expected, due to the noise in the data, both inverted models (Figures 10a and 10b) contain more noise than the inverted models in example 2. Similar to example 2, the result of WRI-SE-MS still contains a large incorrect low-velocity block at the depth of $z = 1500$ m, which we do not find in the result of WRI-SE-WS. The final average model errors of WRI-SE-MS and WRI-SE-WS are 0.22 km/s and 0.16, km/s, respectively. A comparison of the true source weights (+), estimated source weights obtained with WRI-SE-MS (\times) and WRI-SE-WS (\circ) is depicted in Figure 11. The estimated source weights with WRI-SE-WS agree with the true source weights much better than those obtained with WRI-SE-MS. We also compare the data residuals corresponding to

the inversion results of WRI-SE-WS and WRI-SE-MS in Figure 12. Clearly, WRI-SE-MS produces larger data residuals than WRI-SE-WS. Moreover, the quantitative comparison in Table 4 illustrates that the inversion results of WRI-SE-WS exhibit smaller model errors, source errors, and data residuals when compared to that of WRI-SE-MS. These observations imply that compared to WRI-SE-MS, WRI-SE-WS is more robust and stable with respect to measurement noise.

[Figure 10 about here.]

[Figure 11 about here.]

[Figure 12 about here.]

[Table 4 about here.]

Comparison with FWI

Finally, we intend to compare the performances of FWI with source estimation and WRI-SE-WS under bad initial models. We use the same experimental settings as example 1 except with frequency bands ranging from $\{7, 7.5, 8, 8.5, 9, 9.5\}$ Hz to $\{27, 27.5, 28, 28.5, 29, 29.5\}$ Hz and the selection of the penalty parameter $\lambda = 1e0$. During the inversions, we use the initial model displayed in Figure 13a. This model is difficult for both FWI and WRI with the given frequency range, because in the shallow part of the model, i.e., from the depth of 0 m to 120 m, the velocity of the initial model is 0.2 km/s higher than the true one (shown in Figure 13b), which can produce cycle-skipped predicted data shown in Figure 14. Moreover, as the maximum offset is 4.5 km, the transmitted waves that the conventional FWI uses to

build up long-wavelength structures can only reach the depth of 1.5 km (Virieux and Operto, 2009). When the transmitted data are cycle-skipped, the resulting long-wavelength velocity structures would be erroneous, which would further adversely affect the reconstruction of the short-wavelength velocity structures, especially those below 1.5 km.

[Figure 13 about here.]

[Figure 14 about here.]

Figures 15a and 15b show the inversion results obtained by WRI and FWI, respectively. As expected, due to the cycle-skipped data and absence of low-frequency data, the conventional FWI fails to correctly invert the velocity in the shallow area where the transmitted waves arrive, i.e., $z \leq 1.5$ km, which subsequently yields larger errors within the inverted velocity in the deep part of the model, i.e., $z > 1.5$ km. On the other hand, WRI mitigates the negative effects of the cycle-skipped data to the inversion and correctly reconstructs the velocity in both areas that can and cannot be reached by the transmitted turning waves. The final average model error of WRI is 0.1 km/s, which is much smaller than that of FWI—0.22 km/s. Moreover, as shown in Table 5, the ℓ_2 -norm model error of FWI is twice as large as that of WRI. This implies that besides the transmitted waves WRI also uses the reflection waves to invert the velocity model. Clearly, the main features and interfaces in the model are reconstructed correctly by WRI. Figure 16 displays three vertical cross-sections through the model. Compared to FWI, WRI provides a result that matches the true model much better. Figure 17 shows the comparison of the estimated source weights obtained by FWI and WRI. From the comparisons in Figures 17a and 17b, we observe that compared to FWI, WRI provides estimates of source weights with mildly better phases and significantly better

amplitudes. The ℓ_2 -norm source error of FWI is much larger than that of WRI as shown in Table 5. This is not surprising as the quality of the source estimation and the inverted velocity model are closely tied to each other. Figure 18 shows the comparisons between the observed data and predicted data computed with the final results of WRI and FWI. The predicted data computed from the WRI inversion (Figures 18a and 18b) almost matches the observed data, while the mismatch of that from the FWI inversion is much larger (Figures 18c and 18d), which can also be found in the quantitative comparison in Table 5. All these results imply that besides providing a better velocity reconstruction, WRI also produces a better estimate of the source weights.

[Table 5 about here.]

[Figure 15 about here.]

[Figure 16 about here.]

[Figure 17 about here.]

[Figure 18 about here.]

DISCUSSION

Source functions are essential for wave-equation-based inversion methods to conduct a successful reconstruction of the subsurface structures. Because the correct source functions are typically unavailable in practical applications, an on-the-fly source estimation procedure is necessary for the inversion. In this study, we proposed an on-the-fly source estimation technique for the recently developed WRI by adapting its variable projection procedure for

the wavefields to include the source functions. Through simultaneously projecting out the wavefields and source functions, we obtained a reduced objective function that only depends on the medium parameters. As a result, we are able to accomplish the inversion without prior information of the source functions.

The main computational cost of the proposed source estimation method lies within the procedure of projecting out the wavefields and source weights, in which we have to invert a potentially ill-conditioned source-dependent matrix. For ocean bottom node and land acquisitions where all sources “see” the same receivers, we applied the block matrix inversion formula and arrived at a new inversion scheme that only involves inverting a source-independent matrix. This new scheme brings benefits to both direct and iterative solvers when compared to the direct inversion of the original source-dependent matrix. For 2D problems, we illustrated that without losing accuracy, the inversion of the proposed scheme only needs one Cholesky factorization for each frequency, while the direct inversion requires n_{src} . Indeed, this benefit does not only apply to the Cholesky method, but also to other faster direct solvers including the multifrontal massively parallel sparse direct solver (MUMPS, Amestoy et al. (2016)), which can further speed up the projection procedure. When solving 3D problems with iterative solvers, the proposed inversion scheme reduces the number of preconditioners for each frequency from n_{src} to one; therefore, it significantly reduces the computational complexity. Furthermore, the source-independent matrix required to invert is the same one as in the conventional WRI without source estimation. Thus, we avoid iteratively inverting the potentially ill-conditioned source-dependent matrix and can straightforwardly apply any efficient and scalable methods designed for the conventional WRI (Peters et al., 2015) to the proposed approach, which allows us to extend the application of this approach to 3D frequency-domain waveform inversions.

Aside from benefiting inversions with ocean bottom node and land acquisitions, the proposed scheme may also benefit inversions with conventional 3D marine towed-streamer acquisitions where all sources “see” different receivers. In this case, we may lose the benefit of reducing the number of preconditioners due to the acquisition geometry, however, the benefit that avoids the direct inversion of the potentially ill-conditioned source-dependent matrix remains. As a result, we still can straightforwardly apply any preconditioners designed for the conventional WRI without source estimation to solve the proposed scheme. With carefully designed preconditioners, the computational cost of the proposed WRI with source estimation can be roughly equal to that of standard 3D frequency-domain FWI using iterative solvers, because the total numbers of the iterative PDE solves required by the two approaches are approximately the same. As a result, it should be viable to apply the proposed WRI with source estimation to inversions with the conventional 3D marine towed-streamer acquisitions.

While the presented examples illustrated the feasibility of the proposed approach—WRI with source estimation—for frequency-domain inversions, its extension to the time-domain inversions is not trivial. The main computational challenge lies in the fact that in the time domain, the projection procedure requires us to solve a very large linear system, whose size is n_{time} (number of time slices) times larger than that in the frequency domain. As a result, the computational and storage cost in the time domain is larger than that in the frequency domain. Moreover, the time-domain source function is a time series rather than a complex number in the frequency domain. In order to estimate the time series successfully, we may need additional constraints, such as the duration of the source function, to regularize the unknown source functions. Due to the two facts, the extension of the proposed approach to the time-domain inversions requires further investigations.

CONCLUSION

We showed that the recently proposed wavefield-reconstruction inversion method can be modified to handle unknown source situations. In the proposed modification of wavefield-reconstruction inversion, we considered the source weights as unknown variables and updated them jointly with the medium and wavefields parameters. To update the three unknowns, we proposed an optimization strategy based on the variable projection method. During each iteration of the inversion, for fixed medium parameters, the objective function is quadratic with respect to both wavefields and source weights. This fact motivates us to use the variable projection method to first project out the wavefields and source weights simultaneously, and then to update the medium parameters. As a result, we obtained an objective that does not depend on the wavefields and the source weights. Numerical experiments illustrated that equipped with the proposed source estimation method wavefield-reconstruction inversion can produce accurate inversion results without prior knowledge of the true source weights. Moreover, this method can also produce reliable results when the observations contain noise.

We also compared the proposed on-the-fly source estimation technique for wavefield-reconstruction inversion to the conventional source estimation technique for full-waveform inversion. The numerical result illustrated that by extending the search space and the inexact PDE-fitting, the proposed source estimation technique for wavefield-reconstruction inversion is less sensitive to inaccurate initial models and can start the inversion with data without low-frequency components.

ACKNOWLEDGMENTS

We would like to acknowledge Charles Jones (BG Group) for providing the BG Compass model. This research was carried out as part of the SINBAD project with the support of the member organizations of the SINBAD Consortium.

REFERENCES

- Akcelik, V., 2002, Multiscale Newton-Krylov methods for inverse acoustic wave propagation: PhD thesis, Carnegie Mellon 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.
- Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: *Inverse Problems*, **28**, 115016.
- Askan, A., V. Akcelik, J. Bielak, and O. Ghattas, 2007, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures: *Bulletin of the Seismological Society of America*, **97**, 1990–2008.
- Bernstein, D. S., 2005, *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*: Princeton University Press Princeton.
- Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion: *Geophysics*, **74**, WCC105–WCC118.
- Chen, Z., D. Cheng, W. Feng, and T. Wu, 2013, An optimal 9-point finite difference scheme for the Helmholtz equation with PML: *International Journal of Numerical Analysis and Modeling*, **10**, 389–410.
- Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: *Inverse Problems*, **19**, R1.
- Huang, G., W. Symes, and R. Nammour, 2016, Matched source waveform inversion: Volume extension: 86th Annual International Meeting, SEG, Expanded Abstracts, 1364–1368.
- Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme for

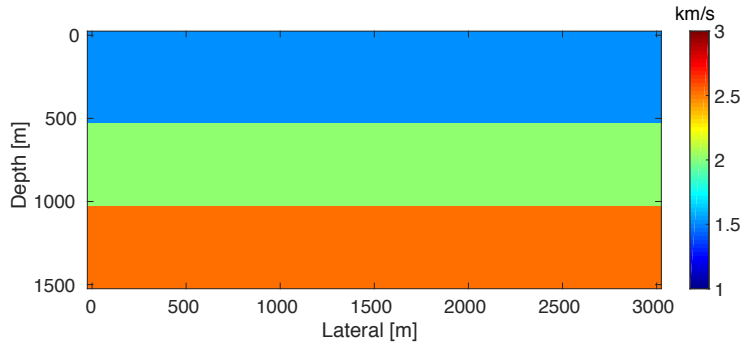
- frequency-domain full-waveform inversion: *Geophysics*, **78**, R249–R257.
- Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: *Geophysics*, **77(3)**, A13–A17.
- Liu, J., A. Abubakar, T. Habashy, D. Alumbaugh, E. Nichols, and G. Gao, 2008, Nonlinear inversion approaches for cross-well electromagnetic data collected in cased-wells: 78th Annual International Meeting, SEG, Expanded Abstracts, 304–308.
- Nocedal, J., and S. J. Wright, 2006, *Numerical optimization*: Springer-Verlag New York.
- Peters, B., C. Greif, and F. J. Herrmann, 2015, An algorithm for solving least-squares problems with a Helmholtz block and multiple right-hand-sides: Presented at the International Conference On Preconditioning Techniques For Scientific And Industrial Applications.
- Peters, B., and F. J. Herrmann, 2016, Constraints versus penalties for edge-preserving full-waveform inversion: *The Leading Edge*, **36**, 94–100.
- Peters, B., F. J. Herrmann, and T. van Leeuwen, 2014, Wave-equation based inversion with the penalty method: adjoint-state versus wavefield-reconstruction inversion: Presented at the 76th EAGE Conference and Exhibition 2014.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: *Geophysics*, **64**, 888–901.
- Symes, W. W., D. Sun, and M. Enriquez, 2011, From modelling to inversion: designing a well-adapted simulator: *Geophysical Prospecting*, **59**, 814–833.
- Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: *Reviews of Geophysics*, **20**, 219–232.
- van Leeuwen, T., A. Y. Aravkin, and F. J. Herrmann, 2014, Comment on: “Application

- of the variable projection scheme for frequency-domain full-waveform inversion” (M. Li, J. Rickett, and A. Abubakar, *Geophysics*, 78, no. 6, R249–R257): *Geophysics*, **79(3)**, X11–X17.
- van Leeuwen, T., and F. J. Herrmann, 2013a, Fast Waveform Inversion without Source-Encoding: *Geophysical Prospecting*, **61**, 10–19.
- , 2013b, Mitigating local minima in full-waveform inversion by expanding the search space: *Geophysical Journal International*, **195**, 661–667.
- , 2015, A penalty method for PDE-constrained optimization in inverse problems: *Inverse Problems*, **32**, 015007.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74(6)**, WCC1–WCC26.
- Warner, M., T. Nangoo, N. Shah, A. Umpleby, J. Morgan, et al., 2013, Full-waveform inversion of cycle-skipped seismic data by frequency down-shifting: 83th Annual International Meeting, SEG, Expanded Abstracts, 903–907.
- Zhou, B., and S. A. Greenhalgh, 2003, Crosshole seismic inversion with normalized full-waveform amplitude data: *Geophysics*, **68**, 1320–1330.

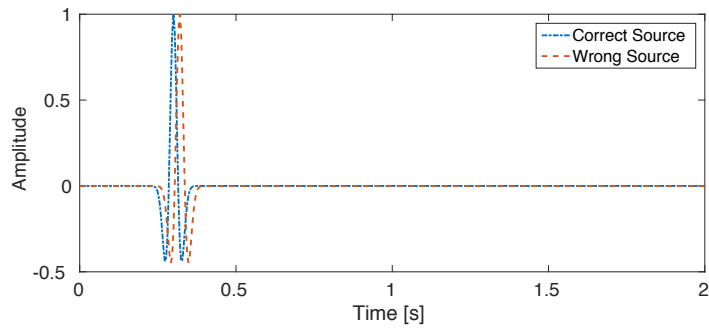
LIST OF FIGURES

1	(a) The true model. (b) Comparison of the correct (blue) and wrong (red) source wavelets. (c) Inversion result with the correct source wavelet. (d) Inversion result with the wrong source wavelet.	38
2	(a) True model, (b) initial model, and (c) the difference between them.	39
3	(a) Inversion result with the true source weights. (b) Inversion result with WRI-SE-MS. (c) Inversion result with WRI-SE-WS. (d) Model difference between (a) and (b). (e) Model difference between (a) and (c).	40
4	Comparison of the true source weights (+) and the estimated source weights with WRI-SE-MS (\times) and WRI-SE-WS (\circ). (a) Phase comparison. (b) Amplitude comparison.	41
5	Comparison of the relative model errors.	42
6	Comparison of the data residuals corresponding to the inversion results of WRI-SE-MS (dashed line) and WRI-SE-WS (dotted line).	43
7	Inversion results with (a) WRI-SE-MS and (b) WRI-SE-WS.	44
8	Comparison of the true source weights (+) and the estimated source weights with WRI-SE-MS (\times) and WRI-SE-WS (\circ). (a) Phase comparison. (b) Amplitude comparison.	45
9	Comparison of the data residuals corresponding to the inversion results of WRI-SE-MS (dashed line) and WRI-SE-WS (dotted line).	46
10	Inversion results with (a) WRI-SE-MS and (b) WRI-SE-WS.	47
11	Comparison of the true source weights (+) and the estimated source weights with WRI-SE-MS (\times) and WRI-SE-WS (\circ). (a) Phase comparison. (b) Amplitude comparison.	48
12	Comparison of the data residuals corresponding to the inversion results of WRI-SE-MS (dashed line) and WRI-SE-WS (dotted line).	49
13	(a)Initial model and (b) difference between the initial model and true models.	50
14	Comparisons between the observed (dashed line) and predicted data (dotted line) computed with the initial model for a source located at the center of the model. (a) Real part comparison. (b) Imaginary part comparison.	51
15	Inversion results with (a) WRI and (b) FWI.	52
16	Vertical profiles of the true model (solid line), initial model (dashdot line), and the inversion results with WRI (dashed line) and FWI (dotted line) at (a) $x = 1000$ m, (b) $x = 2500$ m, and (c) $x = 4000$ m.	53
17	Comparison of the true source weights (+) and the estimated source weights with FWI (\times) and WRI (\circ). (a) Phase comparison. (b) Amplitude comparison.	54

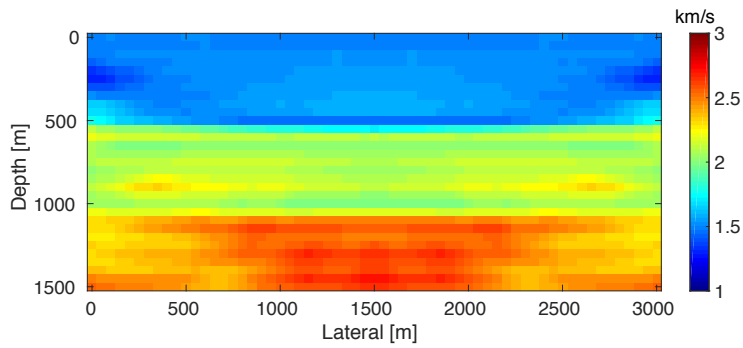
18 Comparisons between the observed (dashed line) and predicted data (dotted line) computed with the final results of WRI and FWI. (a) Real part and (b) imaginary part of WRI result. (c) Real part and (d) imaginary part of FWI result. 55



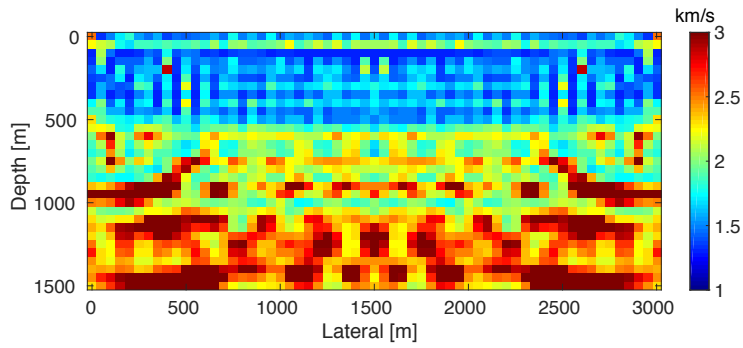
(a) True model



(b) Correct source v.s. wrong source

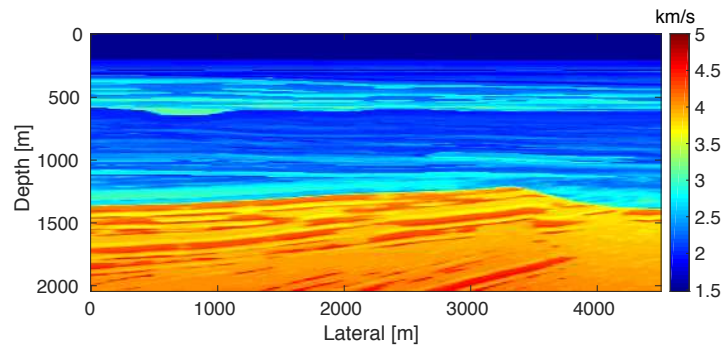


(c) Result with the true source

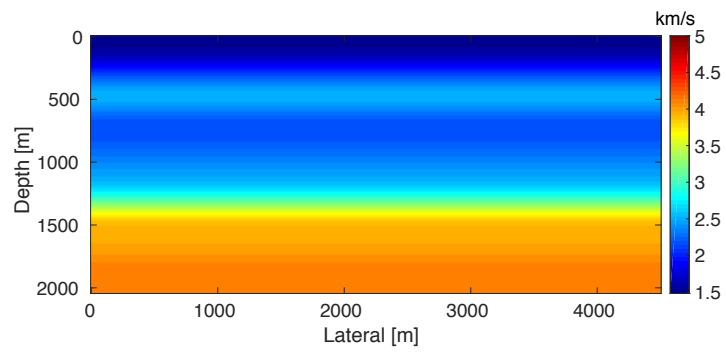


(d) Result with the wrong source

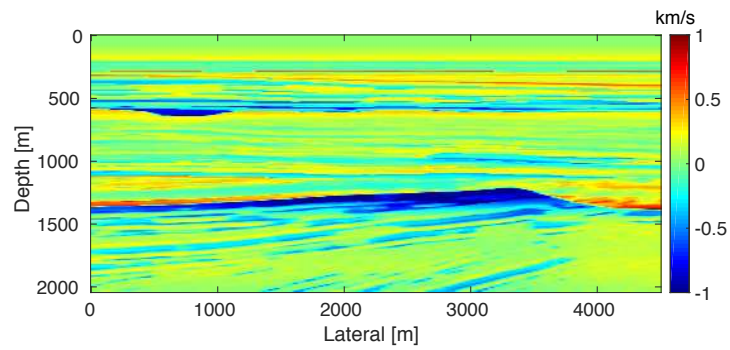
Figure 1: (a) The true model. (b) Comparison of the correct (blue) and wrong (red) source wavelets. (c) Inversion result with the correct source wavelet. (d) Inversion result with the wrong source wavelet.



(a) True model

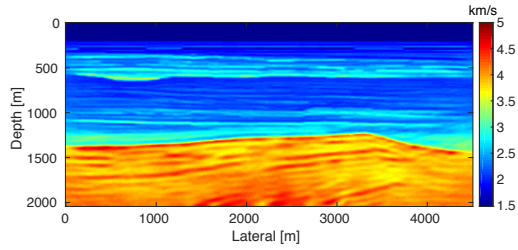


(b) Initial model

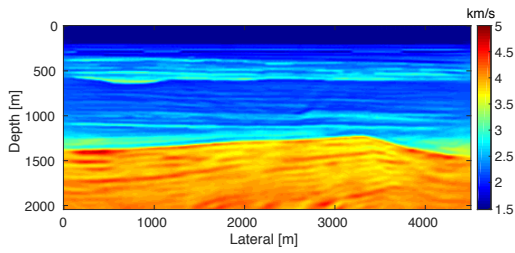


(c) Difference between the initial and true models

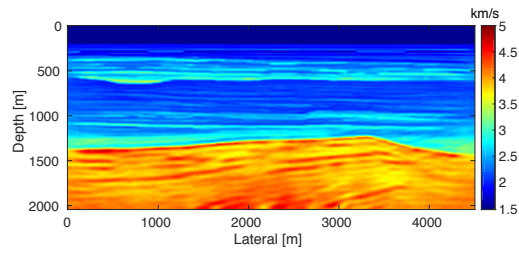
Figure 2: (a) True model, (b) initial model, and (c) the difference between them.



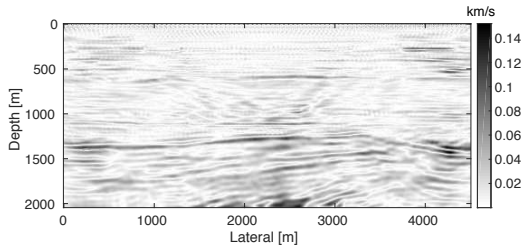
(a) Inversion result with the true source weights



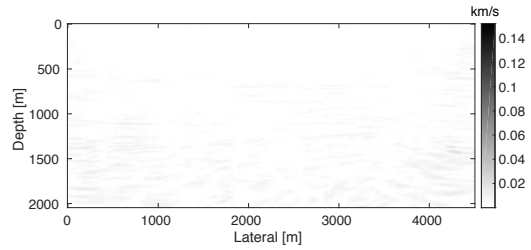
(b) Inversion result with WRI-SE-MS



(c) Inversion result with WRI-SE-WS

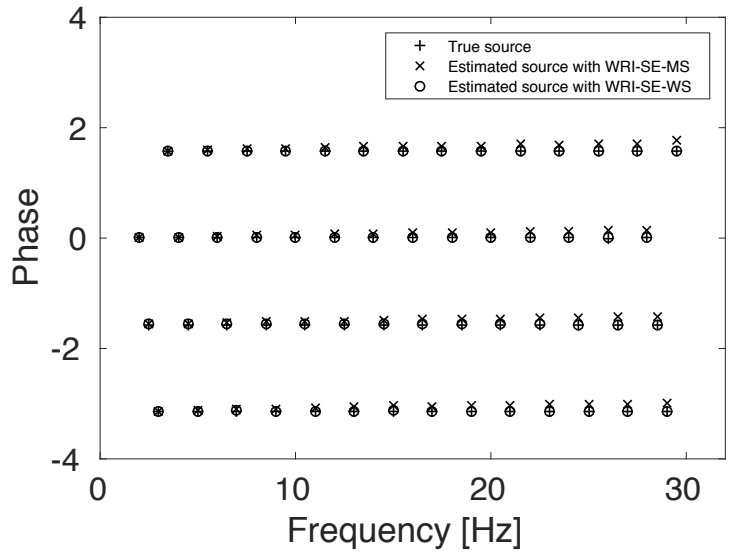


(d) Model difference for WRI-SE-MS

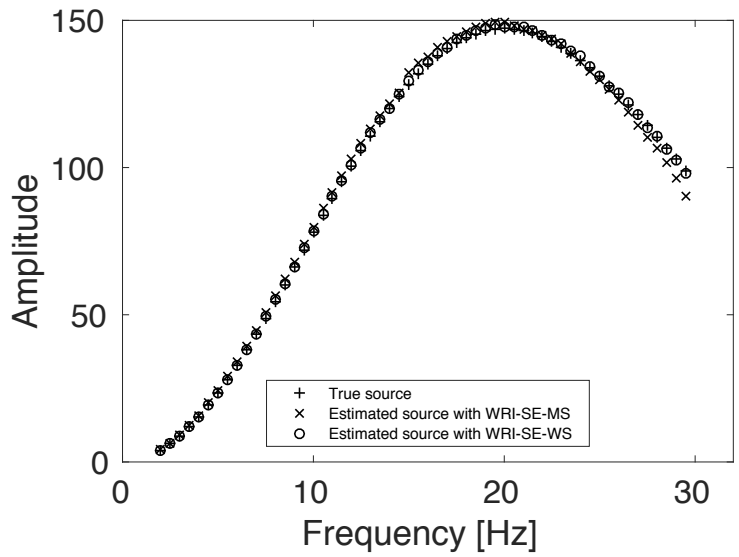


(e) Model difference for WRI-SE-WS

Figure 3: (a) Inversion result with the true source weights. (b) Inversion result with WRI-SE-MS. (c) Inversion result with WRI-SE-WS. (d) Model difference between (a) and (b). (e) Model difference between (a) and (c).



(a) Phase comparison



(b) Amplitude comparison

Figure 4: Comparison of the true source weights (+) and the estimated source weights with WRI-SE-MS (\times) and WRI-SE-WS (\circ). (a) Phase comparison. (b) Amplitude comparison.

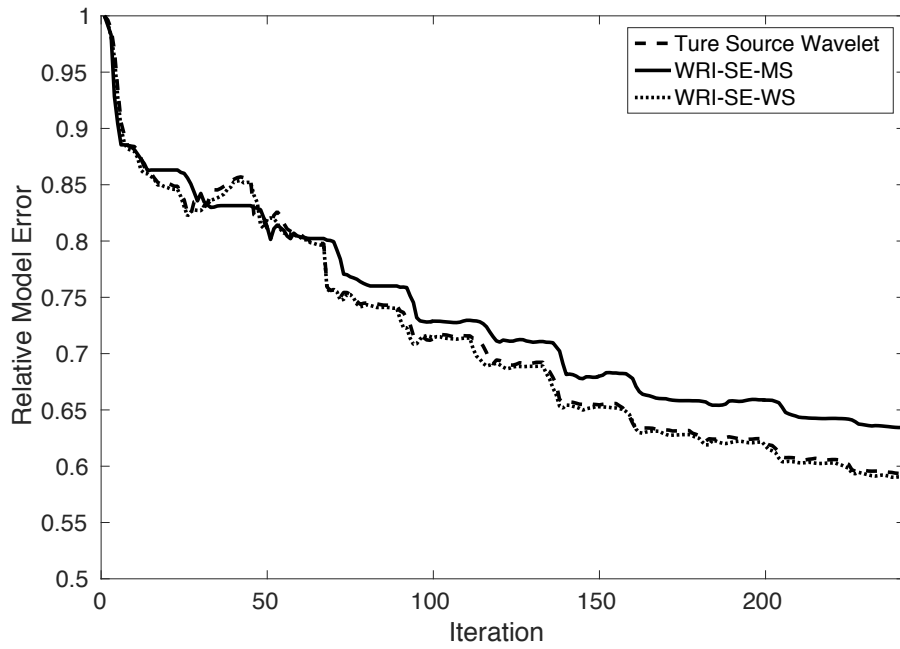


Figure 5: Comparison of the relative model errors.

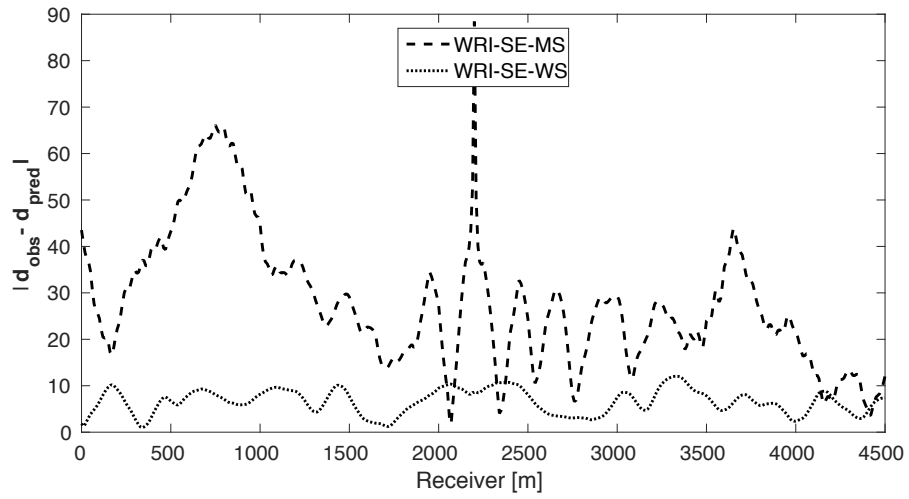
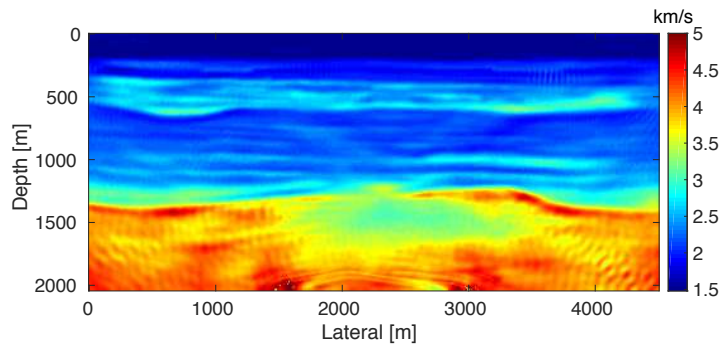
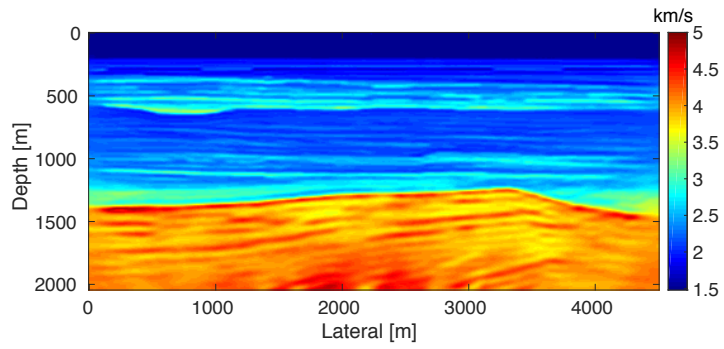


Figure 6: Comparison of the data residuals corresponding to the inversion results of WRI-SE-MS (dashed line) and WRI-SE-WS (dotted line).

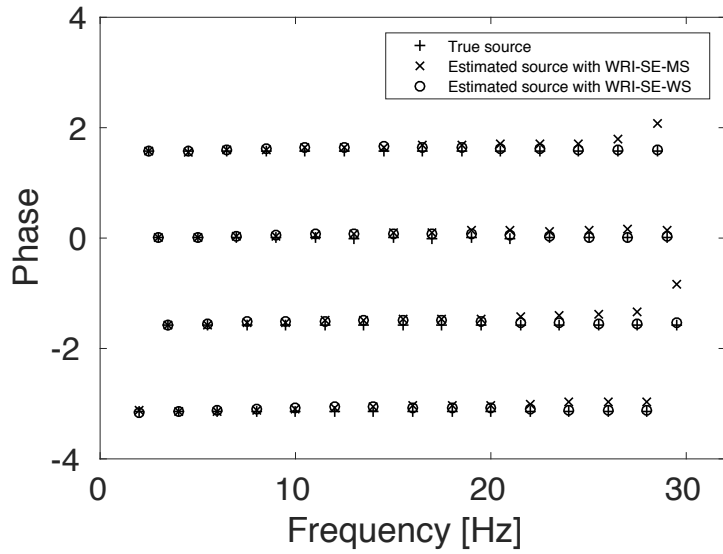


(a) Inversion result with WRI-SE-MS

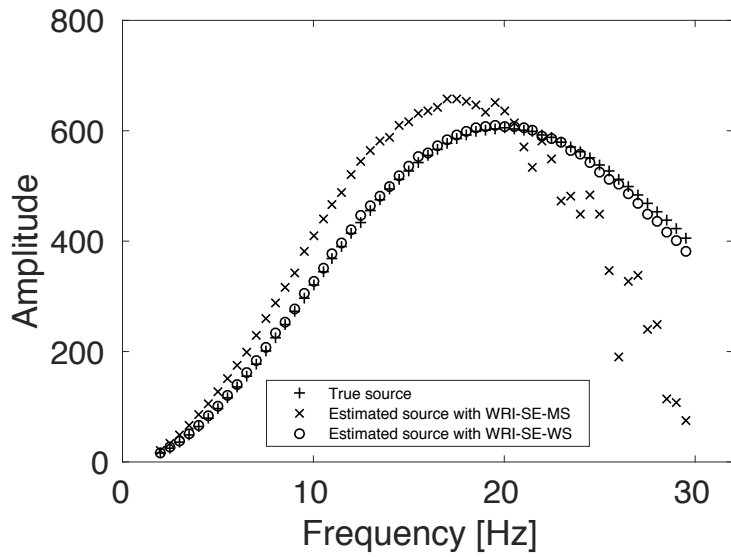


(b) Inversion result with WRI-SE-WS

Figure 7: Inversion results with (a) WRI-SE-MS and (b) WRI-SE-WS.



(a) Phase comparison



(b) Amplitude comparison

Figure 8: Comparison of the true source weights (+) and the estimated source weights with WRI-SE-MS (\times) and WRI-SE-WS (\circ). (a) Phase comparison. (b) Amplitude comparison.

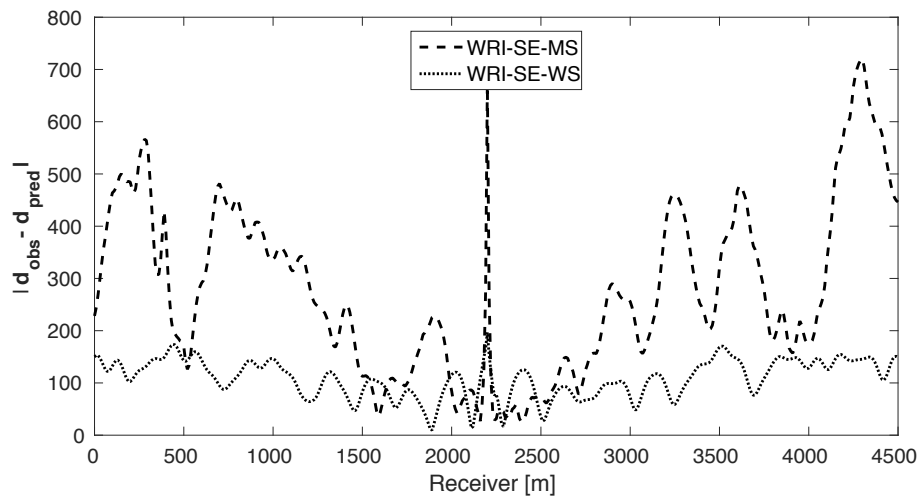
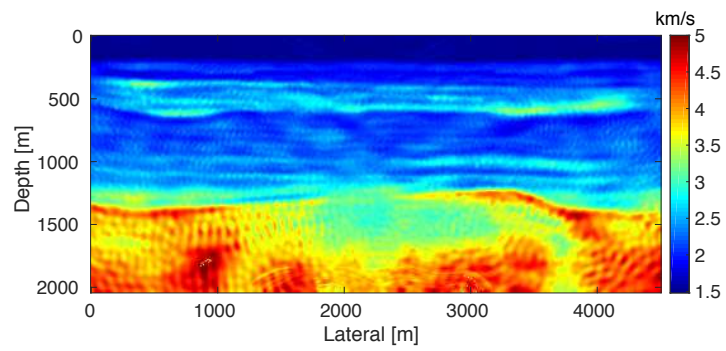
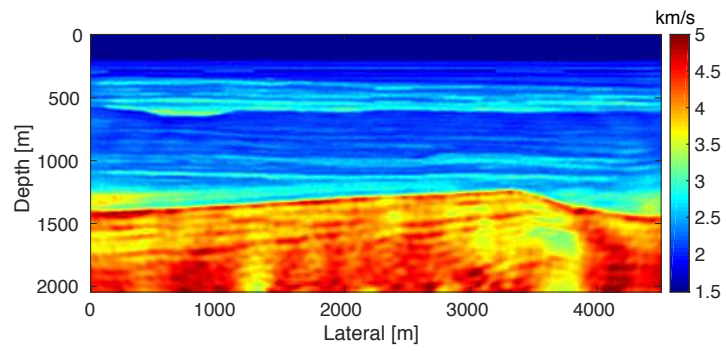


Figure 9: Comparison of the data residuals corresponding to the inversion results of WRI-SE-MS (dashed line) and WRI-SE-WS (dotted line).

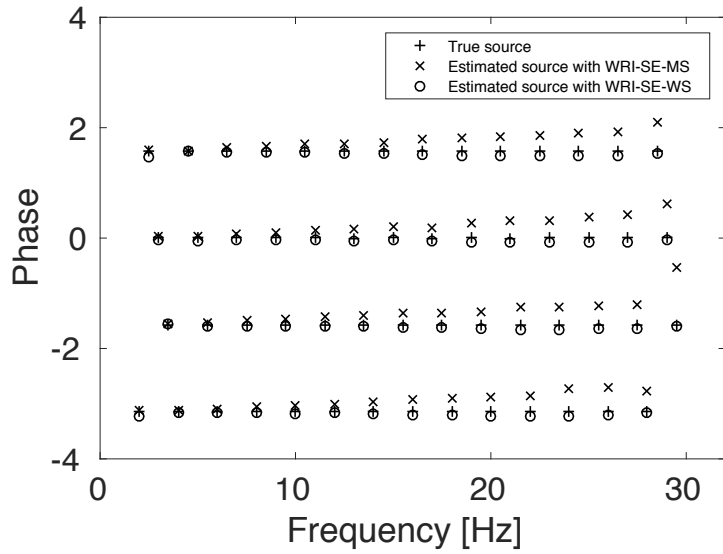


(a) Inversion result with WRI-SE-MS

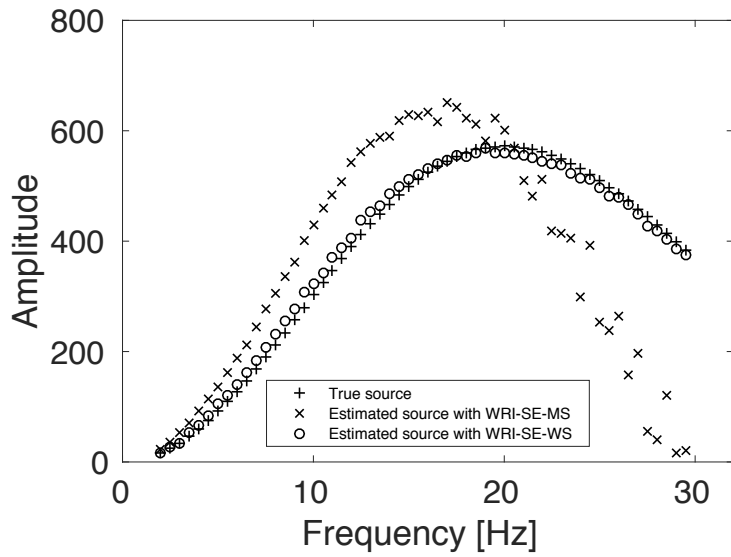


(b) Inversion result with WRI-SE-WS

Figure 10: Inversion results with (a) WRI-SE-MS and (b) WRI-SE-WS.



(a) Phase comparison



(b) Amplitude comparison

Figure 11: Comparison of the true source weights (+) and the estimated source weights with WRI-SE-MS (\times) and WRI-SE-WS (\circ). (a) Phase comparison. (b) Amplitude comparison.

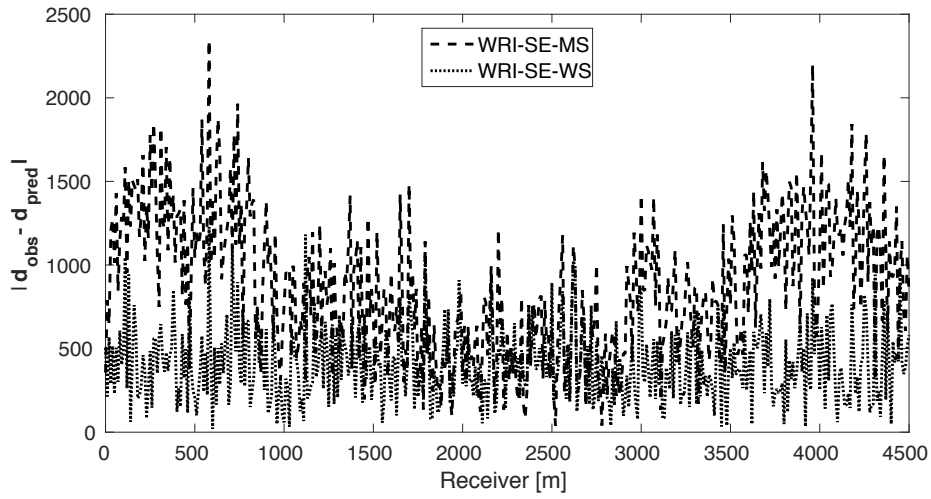


Figure 12: Comparison of the data residuals corresponding to the inversion results of WRI-SE-MS (dashed line) and WRI-SE-WS (dotted line).

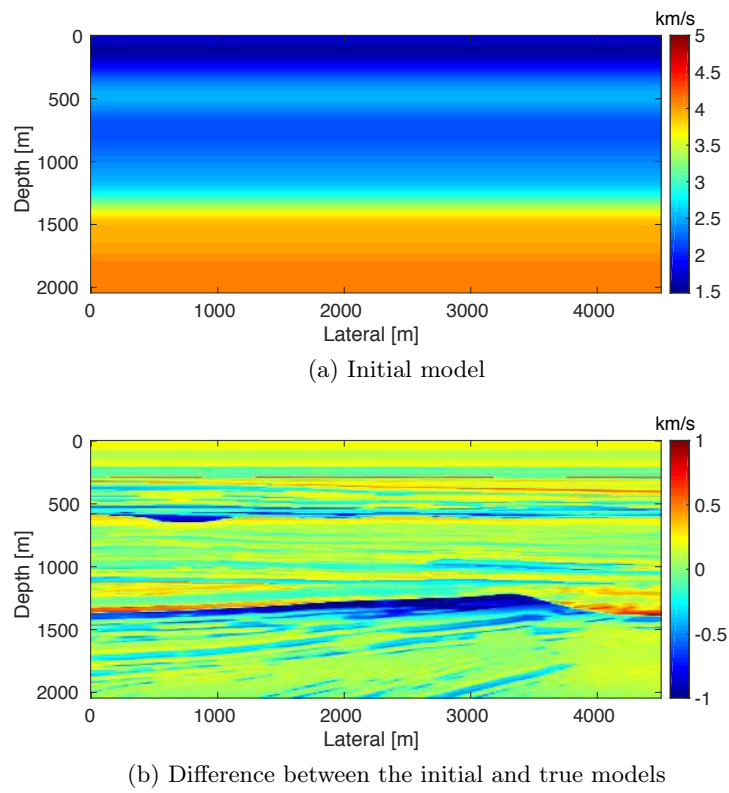
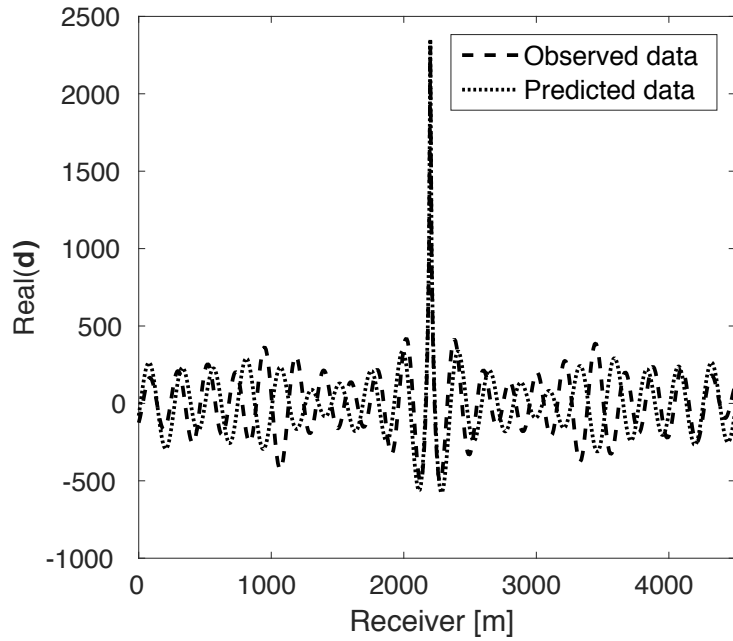
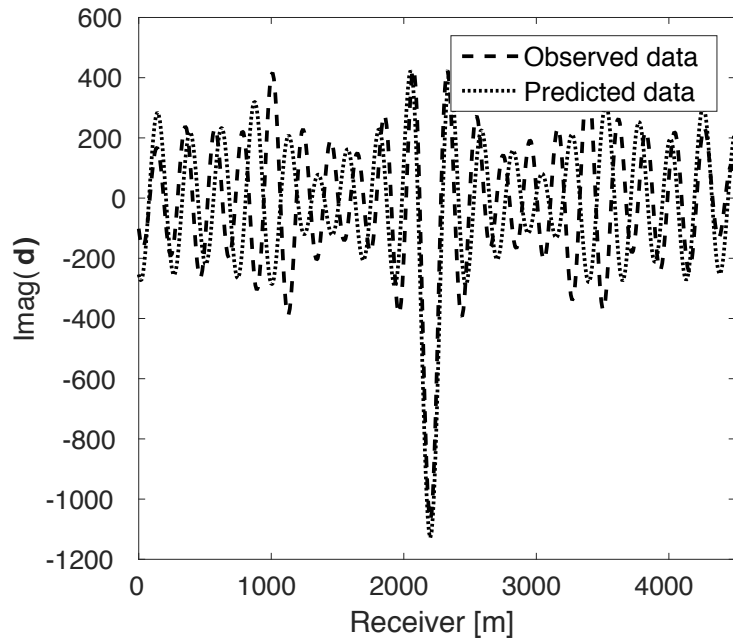


Figure 13: (a)Initial model and (b) difference between the initial model and true models.

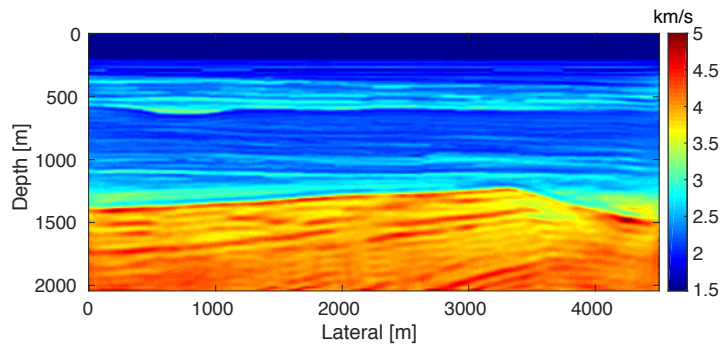


(a) Real part

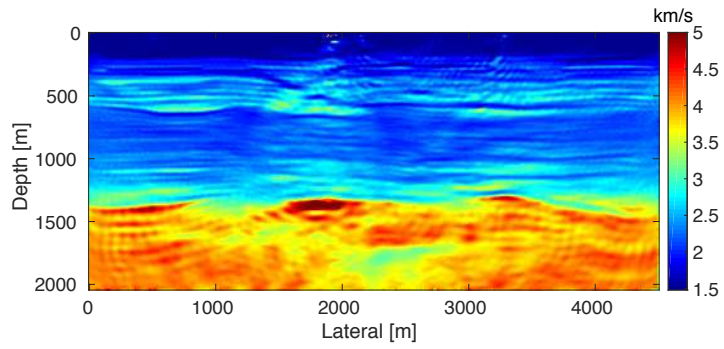


(b) Imaginary part

Figure 14: Comparisons between the observed (dashed line) and predicted data (dotted line) computed with the initial model for a source located at the center of the model. (a) Real part comparison. (b) Imaginary part comparison.



(a) Inversion result with WRI



(b) Inversion result with FWI

Figure 15: Inversion results with (a) WRI and (b) FWI.

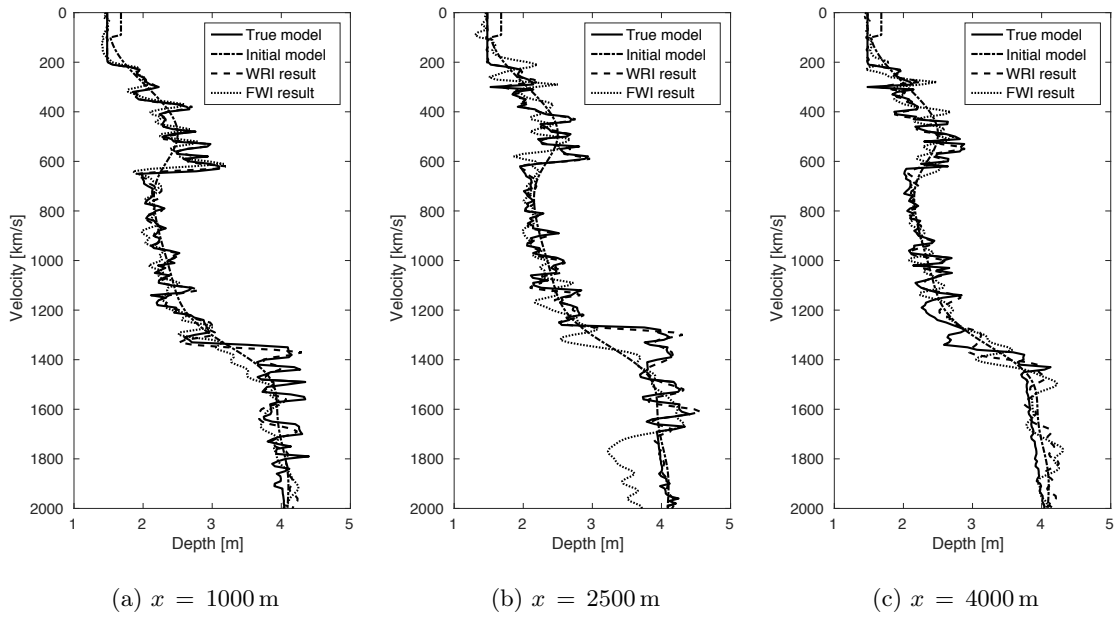
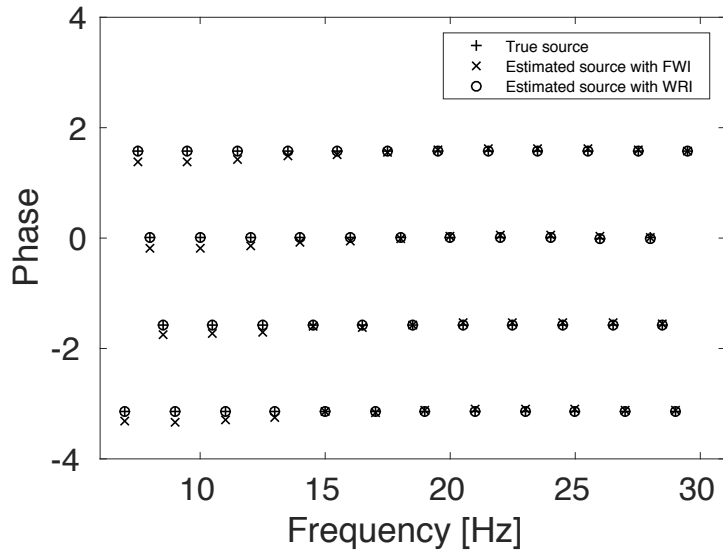
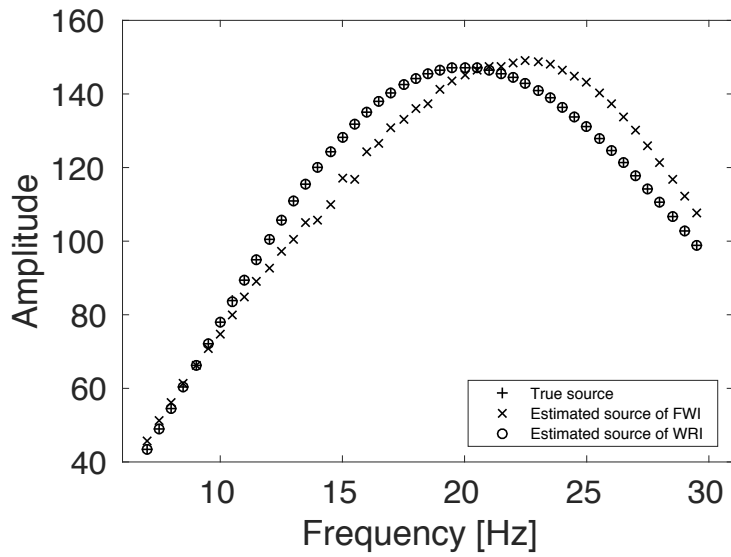


Figure 16: Vertical profiles of the true model (solid line), initial model (dashdot line), and the inversion results with WRI (dashed line) and FWI (dotted line) at (a) $x = 1000$ m, (b) $x = 2500$ m, and (c) $x = 4000$ m.

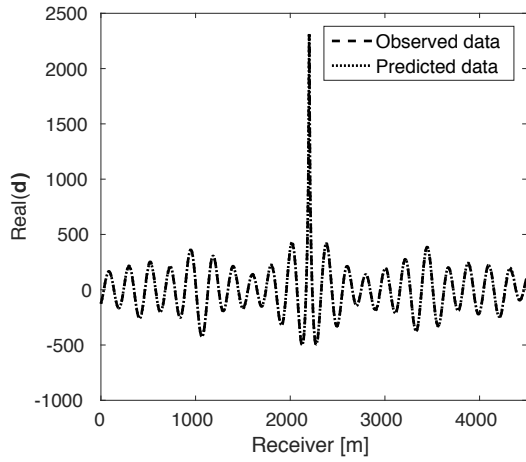


(a) Phase comparison

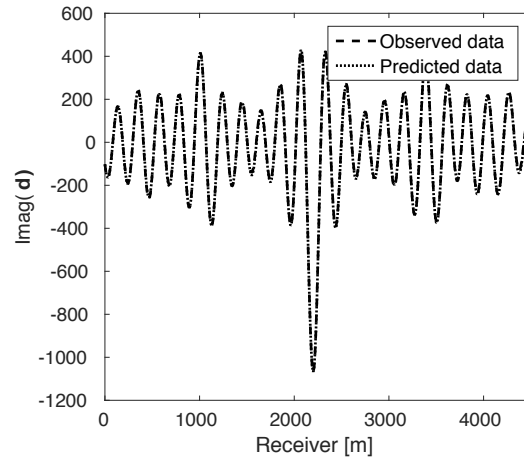


(b) Amplitude comparison

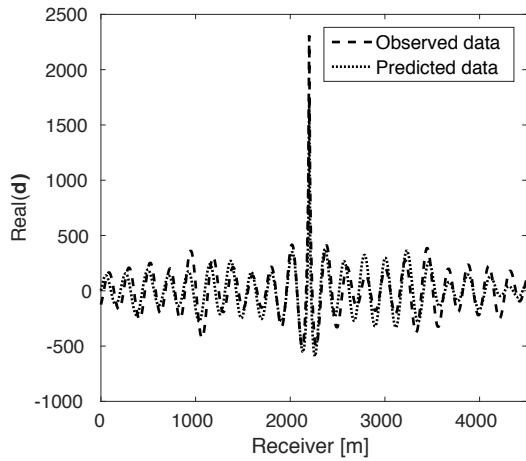
Figure 17: Comparison of the true source weights (+) and the estimated source weights with FWI (\times) and WRI (\circ). (a) Phase comparison. (b) Amplitude comparison.



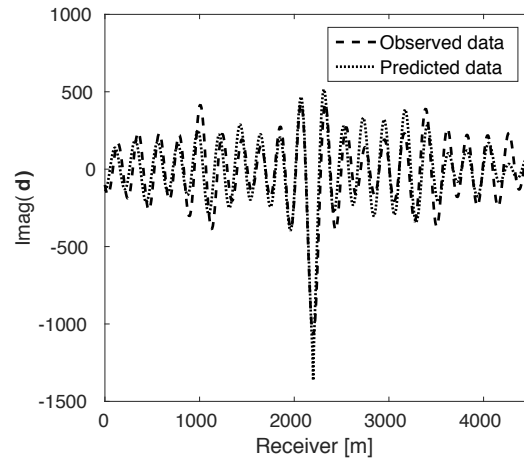
(a) Real part - WRI



(b) Imaginary part - WRI



(c) Real part - FWI



(d) Imaginary part - FWI

Figure 18: Comparisons between the observed (dashed line) and predicted data (dotted line) computed with the final results of WRI and FWI. (a) Real part and (b) imaginary part of WRI result. (c) Real part and (d) imaginary part of FWI result.

LIST OF TABLES

1	Comparison of the computational costs for different algorithms	57
2	Comparisons between inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals. .	58
3	Comparisons between inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals. .	59
4	Comparisons between inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals. .	60
5	Comparisons between inversion results obtained by FWI and WRI in terms of the final model errors, source errors, and data residuals.	61

	nr. of Cholesky factorization	nr. of inversion with Cholesky factors
w/o source estimation	1 per frequency	2 per source per frequency
w/ source estimation – old form	n_{src} per frequency	2 per source per frequency
w/ source estimation – new form	1 per frequency	4 per source per frequency

Table 1: Comparison of the computational costs for different algorithms

	WRI-SE-MS	WRI-SE-WS
$\ \mathbf{m}_t - \mathbf{m}_f\ _2$	48	45
$\ \alpha_t - \alpha_f\ _2$	9.3e2	8e1
$\ \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}}\ _2$	6.6e3	1.5e3

Table 2: Comparisons between inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals.

	WRI-SE-MS	WRI-SE-WS
$\ \mathbf{m}_t - \mathbf{m}_f\ _2$	80	53
$\ \alpha_t - \alpha_f\ _2$	1.1e3	2.1e2
$\ \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}}\ _2$	6.8e3	2.4e3

Table 3: Comparisons between inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals.

	WRI-SE-MS	WRI-SE-WS
$\ \mathbf{m}_t - \mathbf{m}_f\ _2$	96	76
$\ \alpha_t - \alpha_f\ _2$	1.5e3	2.3e2
$\ \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}}\ _2$	2.1e4	9.8e3

Table 4: Comparisons between inversion results obtained by WRI-SE-MS and WRI-SE-WS in terms of the final model errors, source errors, and data residuals.

	FWI	WRI
$\ \mathbf{m}_t - \mathbf{m}_f\ _2$	100	50
$\ \alpha_t - \alpha_f\ _2$	1.3e3	6e1
$\ \mathbf{d}_{\text{obs}} - \mathbf{d}_{\text{pred}}\ _2$	3.8e4	1.7e3

Table 5: Comparisons between inversion results obtained by FWI and WRI in terms of the final model errors, source errors, and data residuals.