

# Adaptive waveform inversion: Theory

Michael Warner<sup>1</sup> and Lluís Guasch<sup>2</sup>

## ABSTRACT

Conventional full-waveform seismic inversion attempts to find a model of the subsurface that is able to predict observed seismic waveforms exactly; it proceeds by minimizing the difference between the observed and predicted data directly, iterating in a series of linearized steps from an assumed starting model. If this starting model is too far removed from the true model, then this approach leads to a spurious model in which the predicted data are cycle skipped with respect to the observed data. Adaptive waveform inversion (AWI) provides a new form of full-waveform inversion (FWI) that appears to be immune to the problems otherwise generated by cycle skipping. In this method, least-squares convolutional filters are designed that transform the predicted data into the observed data. The inversion problem is formulated

such that the subsurface model is iteratively updated to force these Wiener filters toward zero-lag delta functions. As that is achieved, the predicted data evolve toward the observed data and the assumed model evolves toward the true model. This new method is able to invert synthetic data successfully, beginning from starting models and under conditions for which conventional FWI fails entirely. AWI has a similar computational cost to conventional FWI per iteration, and it appears to converge at a similar rate. The principal advantages of this new method are that it allows waveform inversion to begin from less-accurate starting models, does not require the presence of low frequencies in the field data, and appears to provide a better balance between the influence of refracted and reflected arrivals upon the final-velocity model. The AWI is also able to invert successfully when the assumed source wavelet is severely in error.

## INTRODUCTION

Full-waveform inversion (FWI) of 3D seismic data is a technique for generating high-resolution high-fidelity models of physical properties in the subsurface that has become technically and commercially feasible for field data over the past few years. The method is able to produce dramatic improvements in depth-migrated images of conventional reflection data (Sirgue et al., 2010; Warner et al., 2013) and to produce interpretable images of physical properties directly at the reservoir (Lazaratos et al., 2011). Despite these and other successes (Vigh et al., 2010; Lu et al., 2013; Mancini et al., 2015), and its widening uptake across the industry (Kapoor et al., 2013), current implementations of FWI suffer from significant shortcomings; most prominent among these is the problem of local

minima in the objective function produced by cycle skipping (Virieux and Operto, 2009).

Adaptive waveform inversion (AWI), the topic of this paper, reformulates FWI using adaptive matching filters, and in so doing, it appears to provide a robust, effective, and efficient means to overcome cycle skipping during waveform inversion (Guasch and Warner, 2014; Warner and Guasch, 2014a); AWI also provides some additional advantages over conventional FWI (Warner and Guasch, 2014b, 2015). Here, we present the AWI methodology and the rationale behind it. We demonstrate the benefits and properties of AWI by applying it to a variety of synthetic data sets, and we discuss its antecedents and its relationship to other inversion methodologies. In a companion paper, we apply the AWI approach to a full anisotropic

Parts of this paper were presented at the 2014 SEG Annual Meeting as Warner M & L Guasch, "Adaptive waveform inversion" and at the 2015 SEG Annual Meeting as Warner M & L Guasch, "Robust adaptive waveform inversion." Manuscript received by the Editor 17 July 2015; revised manuscript received 16 June 2016; published online 21 September 2016.

<sup>1</sup>Imperial College London, Department of Earth Science and Engineering, London, UK and Sub Salt Solutions Limited, London, UK. E-mail: m.warner@imperial.ac.uk

<sup>2</sup>Sub Salt Solutions Limited, London, UK and Imperial College London, Department of Earth Science and Engineering, London, UK. E-mail: lluis.guasch@sub-salt.com

© The Authors. Published by the Society of Exploration Geophysicists. All article content, except where otherwise noted (including republished material), is licensed under a Creative Commons Attribution 4.0 Unported License (CC BY). See <http://creativecommons.org/licenses/by/4.0/>. Distribution or reproduction of this work in whole or in part commercially or noncommercially requires full attribution of the original publication, including its digital object identifier (DOI).

3D field data set and demonstrate that it significantly outperforms conventional FWI.

Cycle skipping occurs because seismic data are oscillatory. Conventional FWI seeks to find an earth model that minimizes an objective function formed from the sum of the squares of the sample-by-sample differences between the observed data and an equivalent synthetic data set predicted by applying the wave equation to that earth model. The FWI algorithm is sometimes configured instead to seek for a model that maximizes the zero lag of the crosscorrelation of the observed and predicted data sets (Routh et al., 2011). In either case, cycle skipping occurs when a model is found that provides a good match between the observed and predicted data, but for which all or part of the data are misaligned in time by approximately an integer number of wave cycles. Such a cycle-skipped model represents a local minimum in a conventional FWI objective function, so that perturbing this model in any direction will worsen the fit to the observed data even though it may improve the fit to the true model.

Cycle-skipped local minima provide a major challenge when designing robust FWI workflows. In most practical implementations, the problem is addressed by beginning the inversion at low frequencies that are typically approximately 2–4 Hz for surface seismic data, by beginning the inversion from an accurate starting model that will often have been generated using repeated applications of anisotropic reflection traveltimes tomography, by limiting the inversion to shallow depths and short traveltimes, and by experienced practitioners applying rigorous quality control during FWI together with careful parameter and data selection (Warner et al., 2013). Under these circumstances, conventional FWI performs well, and its benefits become manifest.

However, the disadvantages of this approach are clear. Many legacy data sets do not contain sufficiently low-frequency data, and enhanced low-frequency acquisition can add to the acquisition cost and may compromise data quality at the highest frequencies. Accurate model building can be slow and expensive, and it normally requires expert guidance when the data and/or models become complicated. In overly complicated geology, or with poorly illuminated or noisy field data, sufficiently accurate model building may not prove to be possible at all. The need for careful QC and expert oversight also provides a cost or a bar to entry for FWI. A version of waveform inversion, able to provide the benefits of FWI without its limitations related to cycle skipping, would therefore reduce the cost and the total elapsed time required to complete FWI, and would increase the number of data sets and range of problems to which this high-resolution technique could be usefully applied.

In contrast to FWI, AWI does not seek to minimize the difference between the observed and predicted data sets directly. Instead, it proceeds by finding a suite of data-adaptive matching filters that transform one of these data sets into the other. These data-adaptive filters necessarily depend on the observed and the predicted data, which in turn depend on the true and the assumed earth models. If the true and assumed earth models are the same, then these data-adaptive matching filters will necessarily degenerate to become trivial identity filters; an identity filter does not alter its input data, instead it simply passes the input data forward unaltered to its output.

The waveform inversion problem can therefore be setup so that it minimizes not the data misfit directly but instead minimizes the misfit between the filters required to match the two data sets and the ideal identity filter. This is the basis of AWI. In the simple version

that we describe here, we use 1D, least-squares, convolutional, Wiener filters, designed and applied trace-by-trace, to achieve the data-adaptive matching, configuring the inversion, so that it drives these Wiener filters toward zero-lag delta functions. When AWI is appropriately configured and parameterized, it appears to be immune to the effects of cycle skipping without any significant detrimental side effects, and with minimal computational overhead.

Several other waveform-inversion methods have been previously proposed to overcome cycle skipping. These methods currently appear to fall into one of two groups: either the methods are efficient and affordable on commercial data sets in three dimensions, but fail to converge when the data and/or the model are realistically complicated, or the methods work well on realistic data, but have not so far been proven to be affordably achievable on 3D field data. Methods in the first group include those of van Leeuwen and Mulder (2008, 2010), Luo and Sava (2011), and perhaps Ma and Hale (2013), and methods in the second group include those of van Leeuwen and Herrmann (2013) and Biondi and Almomin (2012, 2015). Mathematically, our approach is most closely related to the first two methods, but it takes two important concepts from the other methods, and it is these concepts that seem to be central to the success of AWI on realistic data sets.

The first of these two concepts is that the method is formulated in such a way that the predicted and observed data form a close match even in the presence of significant differences between the starting and true models. We expect this to be a common feature of any successful, broadly applicable, waveform-inversion method. Conventional FWI achieves this by using low frequencies and a good starting model; AWI and Ma and Hale's (2013) approach achieve this by using matching filters; and AWI, van Leeuwen and Herrmann (2013), and Biondi and Almomin (2015) all achieve it by reformulating the underlying problem so that the predicted data need no longer obey a physical wave equation.

The second concept, explored in depth by Symes (2008), is that the dimensionality of the unknown model space is hugely expanded — by the Wiener-filter coefficients in AWI, by adding an additional dimension to the subsurface model in Biondi and Almomin's (2015) approach, and by treating the entire wavefield as part of the unknown model space in the approach of van Leeuwen and Herrmann (2013). It is not clear if this is an essential feature for a successful method, but we note that it is common to all the methods that currently appear able to overcome cycle skipping during waveform inversion in complicated models.

We developed AWI by trying to incorporate both these concepts into conventional FWI, while at the same time trying to adapt the approach of Biondi and Almomin (2012, 2015), so that it became realistically affordable. Although the genesis of our approach originally lay elsewhere, AWI is mathematically more closely related to the approaches of van Leeuwen and Mulder (2010) and Luo and Sava (2011); we discuss the relationship of AWI with these and other methods later in the paper.

## THEORY

There is not as yet a uniformly agreed mathematical formulation within which to present FWI theory. Here, we adopt the simplest formulation that is consistent with our methodology. This formulation is not designed for its mathematical rigor, but it does have the benefits of clarity and of closely mimicking the operations that are performed within practical computational implementations. In

Appendix A, we show how the same result can be obtained using the more rigorous adjoint-state method.

We present the method entirely in the time domain. FWI made significant early progress by working in the frequency domain using mono- or sparse-frequency data (Pratt, 1999), but this is not an approach that is likely to be fruitful for AWI, which is fundamentally a finite-bandwidth concept. The simplest implementation of AWI operates at the level of a single trace representing one source-receiver pair. All our equations therefore refer to just one such time-domain data trace. In any practical implementation, there will always be a summation over many source-receiver pairs to calculate, for example, the gradient, the Hessian, or a model update; to simplify the notation, this implied summation over all source-receiver pairs is not made explicit in our equations.

### Formulation

We describe a single observed time-domain data trace using the column vector  $\mathbf{d}$ , and we describe the equivalent predicted data trace using the column vector  $\mathbf{p}$ . The predicted data trace will have been obtained using an earth model, which we write as the column vector  $\mathbf{m}$ . In simple applications,  $\mathbf{m}$  is likely to contain the values of slowness at points on a regular 3D grid, but in principle,  $\mathbf{m}$  can contain any useful description of the model. Conventional FWI then seeks to minimize an objective function  $f$  with respect to the model  $\mathbf{m}$ , where  $f$  is given by

$$f = \frac{1}{2} \|\mathbf{p} - \mathbf{d}\|^2, \quad (1)$$

and  $\|\mathbf{a}\|^2$  represents the sum of the squares of the elements of a column vector  $\mathbf{a}$ , or equivalently the inner product  $\mathbf{a}^T \mathbf{a}$ . If the starting model generates predicted data that differ from the observed data by more than half a cycle, then this objective function will tend to lead towards a model that is cycle skipped.

In contrast, AWI proceeds by defining a convolutional matching filter  $\mathbf{w}$  which, when convolved with the predicted data trace  $\mathbf{p}$ , provides the best least-squares match to the observed data trace  $\mathbf{d}$ . The coefficients that define the filter  $\mathbf{w}$  are, therefore, those that minimize the function  $g$  given by

$$g = \frac{1}{2} \|\mathbf{P}\mathbf{w} - \mathbf{d}\|^2, \quad (2)$$

where  $\mathbf{P}$  is a Toeplitz matrix containing the predicted data trace  $\mathbf{p}$  in each column, such that it represents the temporal convolution of  $\mathbf{p}$  with  $\mathbf{w}$  (e.g., Hansen, 2002). In equation 2,  $\mathbf{w}$  is a simple 1D acausal Wiener filter that is applied to a single predicted trace to generate a good match to the equivalent observed trace. The filter coefficients in  $\mathbf{w}$  are necessarily functions of  $\mathbf{p}$  and  $\mathbf{d}$ , which in turn are functions of the true and predicted earth models. When used for AWI, the matching filter represented by  $\mathbf{w}$  will typically be at least as long as the data trace  $\mathbf{d}$ . This filter is not designed to match each event in  $\mathbf{p}$  separately to an equivalent event in  $\mathbf{d}$ ; rather it matches an entire predicted trace  $\mathbf{p}$  to an entire observed trace  $\mathbf{d}$ . The AWI method is based directly on matching observed wavefields; it contains no concept of matching individual isolated events or matching explicit arrival times.

Equation 2 represents a linear least-squares problem; it has a single global minimum and no secondary local minima, so that cycle skipping is not an issue. Its solution is given by

$$\mathbf{w} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{d}. \quad (3)$$

This well-known equation represents the crosscorrelation of the observed and predicted data, deconvolved by the autocorrelation of the predicted data. Any practical implementation would normally also apply some form of stabilization, for example, by adding a small positive number to the diagonal of the autocorrelation matrix  $\mathbf{P}^T \mathbf{P}$ .

For a convolutional Wiener filter, the identity filter that leaves the input data unchanged is simply a unit-amplitude delta function at zero lag. We therefore setup the seismic inversion problem using an objective function  $f$  that acts to force  $\mathbf{w}$  toward a delta function at zero lag. There are several potential ways to achieve this. The method that we describe here does this by minimizing (or maximizing) an objective function  $f$  with respect to the model  $\mathbf{m}$ , where  $f$  is given by

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{w}\|^2}{\|\mathbf{w}\|^2}, \quad (4)$$

where  $\mathbf{T}$  is a diagonal matrix that acts to weight the coefficients of  $\mathbf{w}$  by some continuous monotonic function of the magnitude of their temporal lag. The simplest usable form for  $\mathbf{T}$  weights the coefficients directly by the magnitude of their lag up to some maximum value, but more sophisticated forms for  $\mathbf{T}$  can provide more rapid and more stable convergence. If  $\mathbf{T}$  is a function that is zero at zero lag, and that increases with increasing absolute lag, then  $f$  should be minimized. If instead,  $\mathbf{T}$  has a maximum value at zero lag and decreases toward zero at larger lags, then  $f$  should be maximized. Symes (2008) and van Leeuwen and Mulder (2010) discuss strategies for choosing a particular form for  $\mathbf{T}$  in applications of this kind; we have not, however, found AWI to be especially sensitive to the exact form of  $\mathbf{T}$ .

Equation 4 works by penalizing Wiener coefficients at large lags, and the penalty increases as the lag increases; the latter is an important feature of the AWI approach. As the method is iterated, this objective function forces the Wiener filter  $\mathbf{w}$  toward a delta function at zero lag. As equation 4 is written, no account is taken of any absolute mismatch in amplitude between the observed and predicted traces, but it is straightforward to devise forms that are analogous to equation 4 that do consider absolute amplitudes in circumstances in which that is advantageous. One way to achieve this would be to multiply the objective function by a factor of  $1 + (\Delta_{\text{rms}}/\Sigma_{\text{rms}})$ , where  $\Delta_{\text{rms}}$  is the absolute difference of the root-mean-square (rms) amplitudes, and  $\Sigma_{\text{rms}}$  is the sum of the rms amplitudes, of the two traces being compared, but many other possibilities exist.

The normalization of equation 4, as represented here by the term in the denominator, is important. Without some form of normalization,  $f$  could be minimized simply by decreasing all the coefficients of  $\mathbf{w}$ . This would be equivalent to increasing all the amplitudes within the predicted data trace  $\mathbf{p}$ , for example, by increasing all reflection coefficients in the model. This is not a behavior that will move toward the true earth model, and so normalization of the objective function is always necessary in some form to obtain a practical AWI algorithm.

The form of AWI represented by equations 3 and 4 is conceptually the most straightforward, but it does not lead to the simplest mathematical outcome. A simpler outcome is obtained instead by finding a filter  $\mathbf{v}$  that transforms the observed data  $\mathbf{d}$  into the predicted data  $\mathbf{p}$ , that is by solving

$$\mathbf{v} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{p}, \quad (5)$$

and minimizing

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{v}\|^2}{\|\mathbf{v}\|^2}, \quad (6)$$

where  $\mathbf{D}$  represents convolution by the observed data. Note that the two filters  $\mathbf{w}$  and  $\mathbf{v}$  are least-squares convolutional inverses; that is, if  $\mathbf{w}$  and  $\mathbf{v}$  are convolved together, then the result will be a band-limited, zero-lag, unit-amplitude delta function. In the discussion below, we refer to the version of AWI represented by equations 3 and 4 as forward AWI, and that represented by equations 5 and 6 as reverse AWI; both versions, however, seek to solve analogous seismic inverse problems.

### Gradient and adjoint source

To solve the problem at hand, which is to find the best-fitting earth model, we must minimize (or maximize)  $f$  with respect to the earth model in equation 4 or 6. First, we write the discretized wave equation as

$$\mathbf{A}\mathbf{u} = \mathbf{s}, \quad (7)$$

where  $\mathbf{A}$  is a matrix representation of the numerical operator used to implement the wave equation. Most typically this will be a finite-difference time-stepping explicit implementation of the 3D, anisotropic, variable density, acoustic-wave equation, but it can be any useful numerical implementation of any appropriate wave equation. The form of  $\mathbf{A}$  depends on the numerical implementation, and also upon the earth model  $\mathbf{m}$ . In equation 7,  $\mathbf{s}$  is the seismic source defined at all points and all times in the model, and  $\mathbf{u}$  is the wavefield generated at all points and all times by the source  $\mathbf{s}$  in the model  $\mathbf{m}$ . We note that there is a linear relationship between the wavefield  $\mathbf{u}$  and the source  $\mathbf{s}$ , and a nonlinear relationship between the wavefield and model  $\mathbf{m}$ . We also note that the predicted data  $\mathbf{p}$  form a subset of the wavefield  $\mathbf{u}$ , so that  $\mathbf{p} = \mathbf{R}\mathbf{u}$ , where  $\mathbf{R}$  is a restriction operator that acts to select the wavefield only at the receiver positions.

Because the source  $\mathbf{s}$  does not depend on the model  $\mathbf{m}$ , equation 7 implies that

$$\frac{\partial \mathbf{u}}{\partial \mathbf{m}} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u}, \quad (8)$$

which, when restricted to the receivers becomes

$$\frac{\partial \mathbf{p}}{\partial \mathbf{m}} = -\mathbf{R}\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u}. \quad (9)$$

Following the approach of Pratt et al. (1998), the gradient of an objective function  $f$  with respect to the model parameters  $\mathbf{m}$  for any objective function of the form  $f = (1/2)\mathbf{x}^T\mathbf{x}$ , where  $\mathbf{x}$  is a function of the observed and predicted data, is given by

$$\begin{aligned} \frac{\partial f}{\partial \mathbf{m}} &= \frac{\partial \mathbf{x}^T}{\partial \mathbf{m}} \mathbf{x} = \left( \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \frac{\partial \mathbf{p}}{\partial \mathbf{m}} \right)^T \mathbf{x} = - \left( \frac{\partial \mathbf{x}}{\partial \mathbf{p}} \mathbf{R}\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u} \right)^T \mathbf{x} \\ &= -\mathbf{u}^T \left( \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \mathbf{A}^{-T} \mathbf{R}^T \frac{\partial \mathbf{x}^T}{\partial \mathbf{p}} \mathbf{x}. \end{aligned} \quad (10)$$

This quantity is invariably referred to as the gradient. This equation is more conveniently written as

$$\frac{\partial f}{\partial \mathbf{m}} = -\mathbf{u}^T \left( \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \mathbf{A}^{-T} \mathbf{R}^T \delta \mathbf{s}, \quad (11)$$

where  $\delta \mathbf{s}$  is the adjoint source, which we derive below for AWI. Reading from right to left, equation 11 says, that to calculate the gradient, we must find the adjoint source  $\delta \mathbf{s}$ , inject this into the model, back propagate to obtain the adjoint wavefield everywhere in the model, modify this wavefield in a way that depends on the details of the numerical implementation and the model, and crosscorrelate the result of this with the forward wavefield  $\mathbf{u}$ , which was generated using the original source  $\mathbf{s}$ .

In equation 11, the adjoint source is given by the gradient of the objective function with respect to the predicted data

$$\delta \mathbf{s} = \frac{\partial \mathbf{x}^T}{\partial \mathbf{p}} \mathbf{x} = \frac{\partial f}{\partial \mathbf{p}}. \quad (12)$$

In conventional FWI, as defined in equation 1, the adjoint source is simply the data residual  $(\mathbf{p} - \mathbf{d})$ . However, in AWI, the adjoint source is more complicated. If we use equation 2 to define the matching filter, then the AWI adjoint source is given by

$$\delta \mathbf{s} = \mathbf{W}^T \mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \left( \frac{2f\mathbf{I} - \mathbf{T}^2}{\mathbf{w}^T \mathbf{w}} \right) \mathbf{w}, \quad (13)$$

where  $\mathbf{W}^T$  represents crosscorrelation with the filter  $\mathbf{w}$ , and  $\mathbf{I}$  is the identity matrix. The derivation of equation 13 is given in Appendix B.

Reading this equation from right to left, it says that to calculate the adjoint source, we must find the Wiener filter  $\mathbf{w}$  that matches the predicted to the observed data in a least-squares sense, we must normalize this by its inner product  $\mathbf{w}^T \mathbf{w}$ , we must weight the coefficients using the function represented by the matrix  $(2f\mathbf{I} - \mathbf{T}^2)$  that depends on the lag of each coefficient and the current value of the objective function, we must deconvolve this by the autocorrelation  $\mathbf{P}^T \mathbf{P}$  of the predicted data, convolve the result with the predicted data  $\mathbf{P}$ , and finally crosscorrelate this result with the original Wiener filter  $\mathbf{W}$ . These are all computationally inexpensive 1D operations applied to a single trace corresponding to a particular source-receiver pair. The result of this is to generate the adjoint source that corresponds to this source-receiver pair.

If instead, we use equations 5 and 6 to define the reverse AWI functional, then the adjoint source is given by

$$\delta \mathbf{s} = \mathbf{D} (\mathbf{D}^T \mathbf{D})^{-1} \left( \frac{\mathbf{T}^2 - 2f\mathbf{I}}{\mathbf{v}^T \mathbf{v}} \right) \mathbf{v}, \quad (14)$$

where  $f$  is now as defined in equation 6. We derive this equation using the adjoint-state method in Appendix A.

The AWI methodology, in its simplest manifestation, therefore, consists of first computing a forward wavefield for each physical source, from which the predicted data can be extracted. A suite of 1D acausal Wiener filters can then be generated, each of which transforms a predicted data trace into its equivalent observed data trace (or vice versa if using reverse AWI). There will be a different filter for each source-receiver pair. These filters are then manipulated, according to equation 13 or 14, to obtain a suite of adjoint



sources. In practice, there is no requirement to find the Wiener filters explicitly; all that is required is that we manipulate the observed and predicted data to generate the appropriate adjoint source  $\delta s$ . These adjoint sources are then back propagated in place of the conventional data residuals, and the results manipulated and cross-correlated with the forward wavefield, as in conventional FWI, to obtain the gradient. This gradient can then be used in a steepest descent or more sophisticated local-inversion scheme to generate a model update, and the process iterated. The AWI algorithm is, therefore, straightforward and inexpensive to incorporate into any existing time-domain FWI computer code, and it can use the same workflows and all of the preprocessing, preconditioning, regularization, stabilization, and other heuristic methodologies that are typically found within practical FWI implementations.

## RESULTS

In this section, we demonstrate the performance and advantages of AWI by applying it to three synthetic data sets, each of which exemplifies different beneficial characteristics of the method. In the first example, reflections and refractions contribute to model updates approximately equivalently; we discuss this example at some length. In the other two examples, we examine AWI performance for a data set that is dominated by back-scattered reflected arrivals, and for a data set inverted using an incorrect estimate of the source wavelet. Elsewhere (Guasch and Warner, 2014), we apply AWI directly to 3D field data and discuss its practical parameterization and quality control. For each of the synthetic examples below, the detailed modeling and inversion parameters used to generate the data and to perform the inversions are listed in Appendix C.

### Example 1: Combined reflection and refraction inversion

The Marmousi model is well-known. It is a small, shallow model that has a significant vertical velocity gradient; it is an easy model to invert given low-frequency data, sufficient long-offset illumination, and an accurate long-wavelength starting model. In many ways, the Marmousi model is ideally preconditioned for low-frequency combined refraction-reflection wide-angle FWI. Figure 1a shows the version of Marmousi that we use here to demonstrate AWI.

Figure 1b shows the 1D starting model that was used to invert synthetic data generated by the true model. In this study, it is significant that the starting model was not obtained using the true model, by smoothing or some similar operation. Such a procedure would have ensured that the long-wavelength macrovelocity in the starting model was unduly accurate, and such an unjustified assumption is one of the reasons that synthetic inversions of the Marmousi model are often unrealistically successful. Here, the starting velocity model is correct above the seabed, below which it increases smoothly and monotonically toward higher velocities at the bottom of the model. It is trivial to obtain such a model from the observed data after a few minutes of analysis without any knowledge of the true model.

Figure 2a shows a shot record obtained from the true model, and Figure 2b shows the equivalent record obtained from the starting model. Both data sets use a free surface, and so include the effects of surface multiples and source and receiver ghosts. The power spectrum of both data sets, shown in the inset in Figure 1b, peaks at approximately 10 Hz, and the power rolls off rapidly at lower and higher frequencies, principally as result of these ghosts. We have added a low

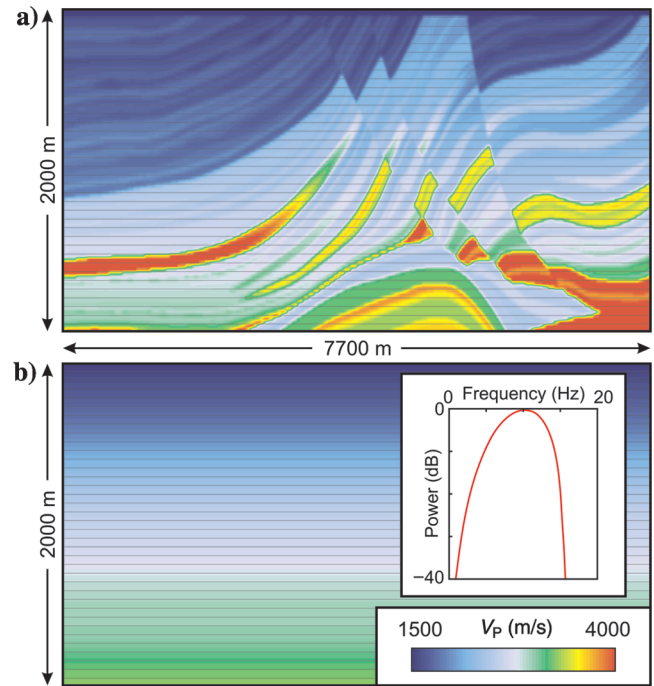


Figure 1. Marmousi acoustic velocity model. (a) True model. (b) Starting model. Insert shows mean power spectrum of the true synthetic data. The model includes a free surface; 91 sources and 187 receivers were used throughout. The velocity scale is identical for all subsequent Marmousi figures.

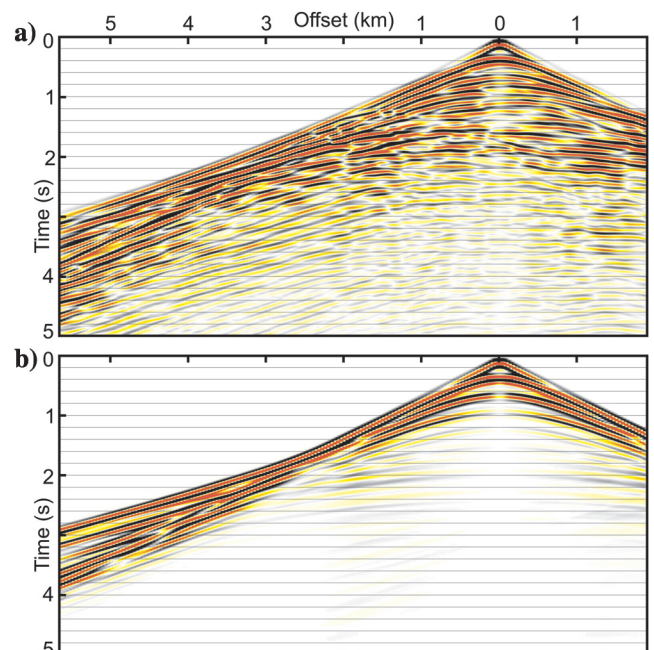


Figure 2. (a) Shot record generated using the true velocity model from Figure 1a. (b) Shot record generated using the starting model from Figure 1b; the first-arrival data are cycle skipped with respect to Figure 2a.

level of band-limited noise to the true data to ensure that low frequencies cannot be spuriously recovered during the inversion.

Figure 3 demonstrates that parts of the predicted data, generated using the starting model, are cycle skipped at 10 Hz with respect to the true data. Consequently, to invert these data successfully with conventional FWI, the starting model would need to be improved, the inversion would need to begin at less than 10 Hz, or a sophisticated strategy of depth, time, and offset windowing would need to be implemented. For this shallow, simple, Marmousi model, such strategies are straightforward to implement, and can be made to work well; but for deeper, more complicated, noisy, band limited, 3D, field data sets, such strategies typically range from difficult and expensive to near impossible to achieve with certainty.

Figure 4 shows the result of using conventional FWI to invert the synthetic data generated by Figure 1a, beginning from the starting model in Figure 1b, using the full data bandwidth, and the full time and offset range, at all iterations. The detrimental effects of cycle skipping are immediately apparent, and the resultant model is far from the true model. Several generic features in Figure 4 are noteworthy. The original long-wavelength macrovelocity model has not

changed significantly from the starting model. Fine structure has appeared in the model, but it is generally poorly focused, misplaced in depth, and of insufficient contrast with the macromodel. The effects of cycle skipping are most readily apparent within the shallow fault block in the right center of the model. This block should be of higher velocity than its surroundings, but cycle skipping has pushed the velocity within it toward lower values as the seismic data have become misaligned on the wrong cycle. The detailed structure of this cycle-skipped model is a strong function of the exact parameters, starting model, and software that are used to generate it; cycle-skipped FWI is unstable, and so small changes in any one of these will tend to lead the model to a different local minimum. The significant result is that these local minima are all spurious and all share the generic features outlined above.

It is also worth noting that the recovered model in Figure 4 has evolved features that could appear to be geologically reasonable, and that these match qualitatively to analogous features in the true model. Although familiarity with Marmousi makes it clear that Figure 4 is badly in error, if this were a result obtained from real field data, then the apparent match to geologic expectation might provide undue confidence in its veracity. Although it is difficult to demonstrate definitively, we suspect that the detrimental effects of cycle skipping during commercial FWI of field data may be more prevalent than is generally believed, though these effects are not typically of the intensity displayed by Figure 4.

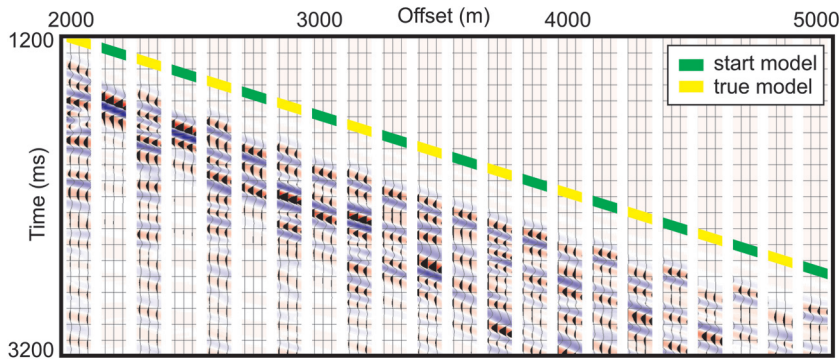


Figure 3. A portion of the shot record from Figure 2 showing the observed and predicted data interleaved. The observed data, from the true model, are indicated by yellow; the predicted data, from the start model, are indicated by green. The starting model generates first-arrival data that are late by about a cycle in the offset range 2000–3000 m, and that are early by about a cycle at longer offsets in the range 4000–5000 m. The starting model, therefore, generates data that are badly cycle skipped with respect to the true data.

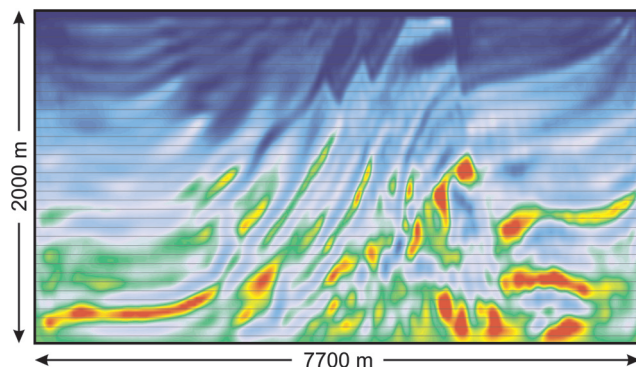


Figure 4. Velocity model recovered by conventional FWI obtained when the starting data are badly cycle skipped. The deep structure is poorly focused; the intense low-velocity region within the shallow high-velocity fault block at center right is a symptom of cycle skipping — the FWI updates here have the wrong sign because they act to move the predicted data toward the wrong cycle.

Figure 5 shows data generated by the recovered FWI model; it demonstrates how FWI attempts to deal with cycle skipping in the starting data. It brings the two data sets into good phase alignment, and there are many features that suggest a good data match; however, this match uses the wrong cycle over a significant portion of the early arrivals. Conventional FWI also modifies the model, so that the unmatched cycles in the predicted data have their amplitudes suppressed. Watching the model evolution, and using these perfectly modeled low-noise data, the effects of cycle skipping are obvious. But in noisier, less-perfectly matched field data, cycle-skipped models can produce synthetic

data that give even careful and experienced practitioners significant difficulty in detecting the presence of cycle skipping in the final result. We note that the FWI objective function here drops systematically at every iteration, and that the zero lag of the crosscorrelation of the field and predicted data increases almost everywhere during the inversion, so that neither of these parameters can be used easily to detect cycle skipping.

Figure 6 shows the result of applying AWI to these synthetic data. This result is now unaffected by cycle skipping, and the dramatically improved match to the true model is immediately apparent. The AWI result is not perfect. It is, as is conventional FWI, still limited by the bandwidth, receiver aperture, shot location, and subsurface illumination provided by the observed data set. We show below, in example 2, that AWI can also help to push successfully a little harder against these acquisition limitations than conventional FWI. The only significant difference between the two computer codes that generated Figures 4 and 6 was the adjoint source. For FWI, the adjoint source is just the data residual, whereas for AWI, it is



given by either equation 13 or 14 as appropriate. Everything else in the two codes, including the approximate Hessian, data preprocessing, gradient preconditioning, data bandwidth, step-length calculation, and inversion parameterization were identical between the two runs, and the parameters that were used correspond to those that we would typically adopt for conventional FWI (Warner et al., 2013).

Figure 7, which is directly analogous to Figure 5, shows the predicted data generated by the AWI-recovered model. The match to the true data, qualitative and quantitative, is now much improved; the correct phases are aligned, and there is no evidence that the predicted data have been affected by cycle skipping during inversion. The inversion here proceeds following a distinctive pattern that appears to characterize the AWI process (Figure 8). AWI initially improves the model where the data coverage is best, and brings deeper structure rapidly into focus even though the shallow model is not yet correct. As the inversion proceeds, and the shallow model improves, the deep structure evolves smoothly toward the correct depth, and the improvement in model quality and in data match move successively outward and downward into regions of the model and regions of the data where the coverage is less complete. FWI and AWI are ultimately trying to solve similar problems, but their different objective functions result in two significantly different solution spaces. Consequently, the path that each method follows, and the models that they visit during their respective attempts to reach a global solution, can be completely different.

### Example 2: Reflection dominated inversion

Although we developed AWI as a means of overcoming cycle skipping, it rapidly became clear that the method has additional advantages over conventional FWI even for data sets for which cycle skipping is not an issue. One such advantage appears to be that AWI is better able to deal with reflections than is conventional FWI. In particular, AWI appears more able to push deeper into a model, below the deepest refracted arrivals, to recover macrovelocities at intermediate scale lengths.

Figure 9a shows a synthetic model that generates data that are dominated by back-scattered, subcritical, reflected arrivals. It contains a water layer, a weak positive vertical velocity gradient, and strongly reflecting heterogeneities with horizontal and vertical boundaries. The model contains a free surface, and the resultant surface seismic data display free-surface and interbed multiples. The

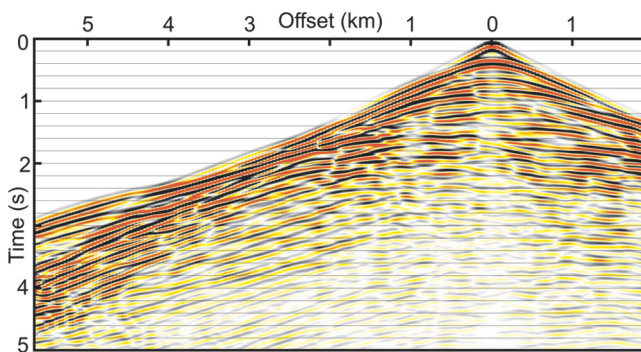


Figure 5. Shot record generated using the FWI-recovered model from Figure 4. Careful comparison with Figure 2a shows that the data are still cycle skipped.

starting model (Figure 9b) is 1D; the starting model is not cycle skipped at 5 Hz, but it also does not correspond to the correct long-wavelength velocity model. The acquisition geometry used in this experiment represents a single towed marine streamer with a maximum offset of 5 km, which is equal to the maximum target depth in the model. With this acquisition geometry, recorded refracted arrivals do not penetrate the model any deeper than the second layer of heterogeneities.

Figure 10a shows the results of applying conventional FWI to these data. Here, we have stopped iterating after 50 iterations on each source, which is equivalent to considerably more computational effort than would typically be used to invert 3D field data. Because these synthetic data are noise free, and the forward and inverse modeling are identical, continued iteration would slowly continue to improve the recovered model almost without limit by matching ever-finer details within the observed data; such over-iterated synthetic results though have little relevance to behavior on real data. In Figure 10a, FWI has recovered the velocity structure accurately within the top two layers, principally by matching the shallow refracted arrivals. At greater depth, FWI is able to partially recover the fine structure using back-scattered reflected arrivals, but the image is rather weak, and there has been only minimal change to the underlying long and intermediate-wavelength velocity model. It is

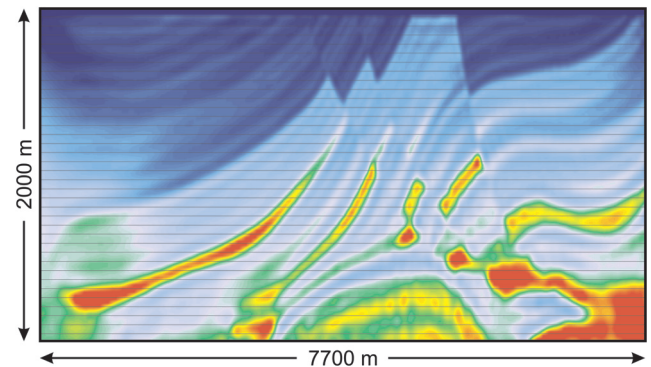


Figure 6. Velocity model recovered by AWI when the starting-model data are badly cycle skipped. The AWI result is clearly superior to the conventional FWI result from Figure 4. The result shows no evidence of cycle skipping, and its limitations are caused principally by the limited bandwidth and limited coverage of the observed data.

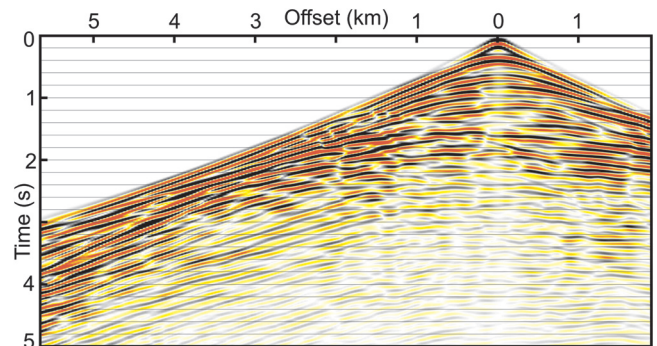


Figure 7. Shot record generated using the AWI-recovered model from Figure 6. The data are no longer cycle skipped.

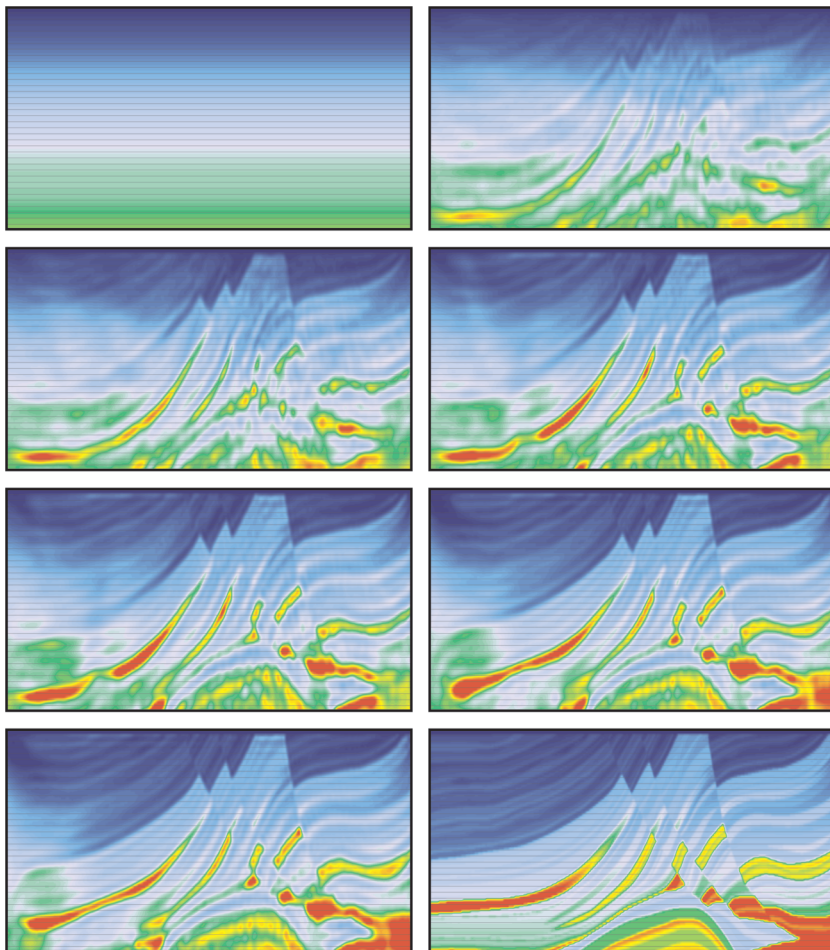


Figure 8. Model evolution during AWI. The starting model is top left, the final model is bottom left, and the true model is bottom right. Intermediate models are spaced at approximately equal steps in the logarithm of the total number of iterations. The central shallow portion of the model is recovered early, and improvements in the model tend to spread outward and downward from that region. Note that the high-velocity layer at bottom left is formed initially at the wrong depth, but it evolves toward the correct depth as the shallow velocities improve.

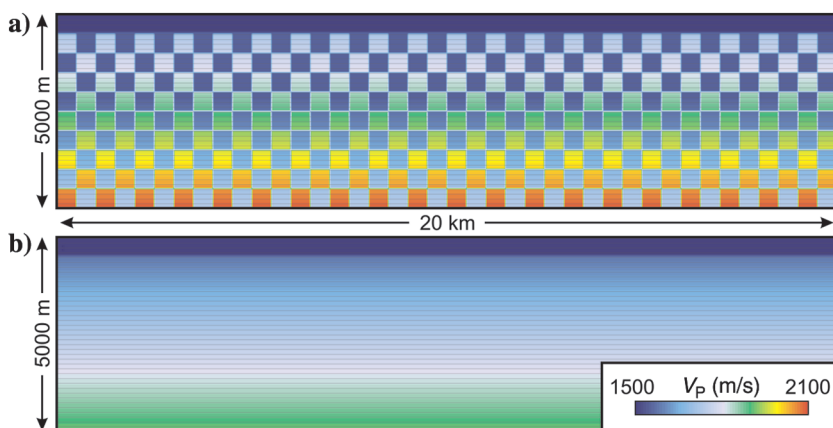


Figure 9. Model used to demonstrate the performance of FWI and AWI on reflection-dominated data sets. (a) True model. (b) Starting model.

possible to add several enhancements to FWI (and AWI) to improve model recovery from pure-reflection data (for example, Yao and Warner, 2015); no such enhancements have been included in this example.

Figure 10b shows exactly analogous results obtained using AWI. As before, the AWI adjoint source was the only difference between the algorithms used to generate the results shown in Figure 10a and 10b. The AWI result is clearly superior to the FWI result though the differences are not as significant as in the cycle-skipped example 1. The AWI method is able to make more use of reflected phases, and to push more deeply into the model below the deepest turning rays, to recover a more-accurate velocity model, especially at intermediate wavelengths. AWI has also moved the long-wavelength velocity model closer to the true model in that region of the model that is imaged only by reflections. The AWI algorithm is also able, in synthetic and field data, to push a little harder into regions at the edge of the model where the data coverage is less complete.

Although this enhanced sensitivity was unexpected, in retrospect it is perhaps not so surprising. In AWI, missing reflectors within the current model will typically contribute to Wiener coefficients at large lags, and these will have a strong effect upon the objective function; FWI does not up-weight reflections in this way. In addition, FWI responds directly only to point-by-point amplitude differences between observed and predicted data; it responds to arrival-time differences only indirectly through their effect upon these amplitude differences. AWI, however, is able to respond directly to arrival-time mismatches because these are directly related to the lags that appear in the Wiener filters, and larger lags are more strongly penalized by the AWI objective function. Consequently, AWI has enhanced sensitivity to build missing reflectors and enhanced sensitivity to mismatches in the moveout of the reflections that subsequently appear from these reflectors; AWI has correspondingly reduced sensitivity to amplitude mismatches, which also tends to favor the recovery of longer wavelength velocity structure from reflections.

The characteristics of AWI displayed in Figure 10, and the additional sensitivity to reflection data beyond that achieved by FWI, appear to be general and repeatable features of AWI. We stress, however, that this additional uplift is relatively modest, and AWI alone does not provide a complete solution to the difficult problem of recovering an unknown long-wavelength velocity model from pure-reflection data via waveform inversion, but AWI can make a greater contribution to the solution of this problem than can conventional FWI. We see exactly this behavior



replicated in tests with blind synthetics (Warner and Guasch, 2015) and 3D field data (Guasch and Warner, 2014).

### Example 3: Inversion assuming an incorrect source wavelet

Because AWI has a strong relationship to source-wavelet inversion — see the “Discussion” section — the method might be supposed to have undue sensitivity to the quality of the source wavelet assumed during inversion. In the examples above, the source wavelet used during inversion was exact; in this example, we explore the consequences for FWI and AWI when the assumed source wavelet is not correct. For this test, we use the same true model and acquisition parameters as for example 1, but we begin from a more-accurate starting model because we wish to compare FWI and AWI in the absence of cycle skipping.

Figure 11 shows the two wavelets that were used; the true source is the wavelet required to generate the real data. We first established a benchmark for AWI and FWI by using the correct wavelet and using a good starting model for the inversion. Figure 12 shows the results of this for AWI and FWI; this figure also shows a useful comparison between the two approaches for this ideal case. The two recovered models are similar, they differ in their fine details, and they are both limited in their accuracy by the limitations of the observed data. The AWI result is marginally sharper and better focused, but it is also a little more noisy and more influenced by the finite-source spacing. The practical differences here though are subtle.

Figure 13 shows FWI and AWI inversion results obtained when using a poor estimate of the true source; in this case, the source in Figure 11a was used to generate the data, but the source in Figure 11b was used to invert them. The bandwidth of both sources is similar, but their waveforms are quite different; there is approximately a 90° phase shift between the major peaks in the two wavelets. Figure 13a shows the FWI result. Not surprisingly, this result is poor; the assumed wavelet is badly in error, and FWI cannot recover from that position without, for example, including an explicit inversion for the source wavelet into the algorithm. The result in Figure 13a shares some characteristics with the cycle-skipped result in Figure 4. Indeed, this FWI result is now cycle skipped because the spurious source wavelet acts to cause some arrivals to become incorrectly aligned between observed and predicted data even though the starting model is accurate.

Figure 13b shows the analogous result for AWI. Surprisingly, this result is much better than the FWI result, and it is clear that AWI is more robust against errors in the assumed source wavelet than is FWI. Comparison of Figures 12b and 13b demonstrates that use of the correct wavelet is beneficial, but that even a wavelet as significantly in error as that shown in Figure 11b can produce a reasonably accurate outcome. The principal differences introduced here by use of the wrong wavelet are an increase in noise, spurious shallow layering that presumably acts to extend the duration of the effective wavelet to fit the observed data better, and an over-intense recovery of the higher velocity regions within the deeper portion of the model. None of these errors is catastrophic, and the resultant model is clearly superior to its FWI equivalent.

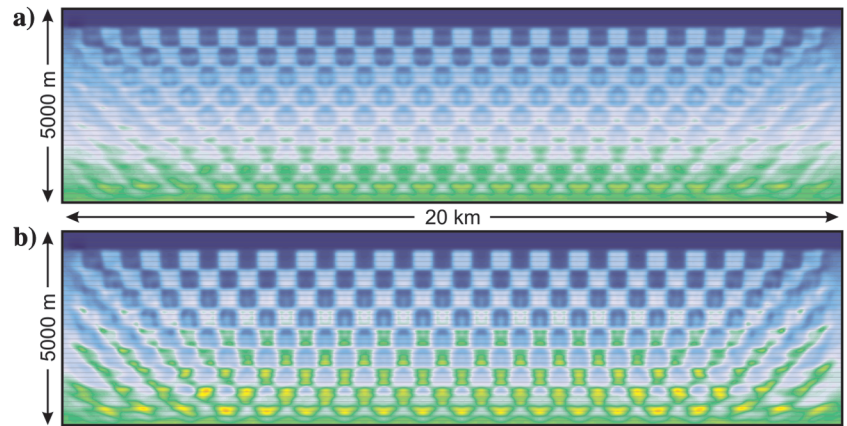


Figure 10. Results of inverting data generated by the model in Figure 9a. (a) Using FWI. (b) Using AWI.

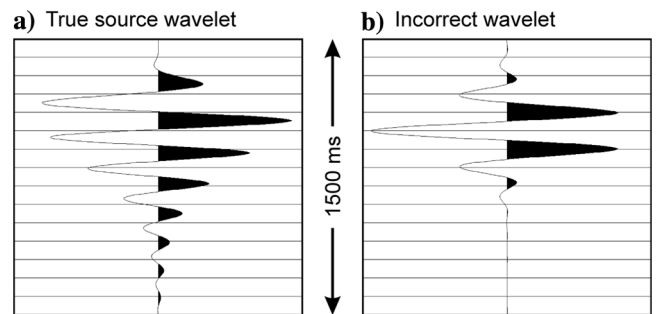


Figure 11. Source wavelets used to test FWI and AWI. (a) Wavelet used to generate synthetic data, and to produce the inversions shown in Figure 12. (b) Incorrect wavelet used to produce the inversions shown in Figure 13.

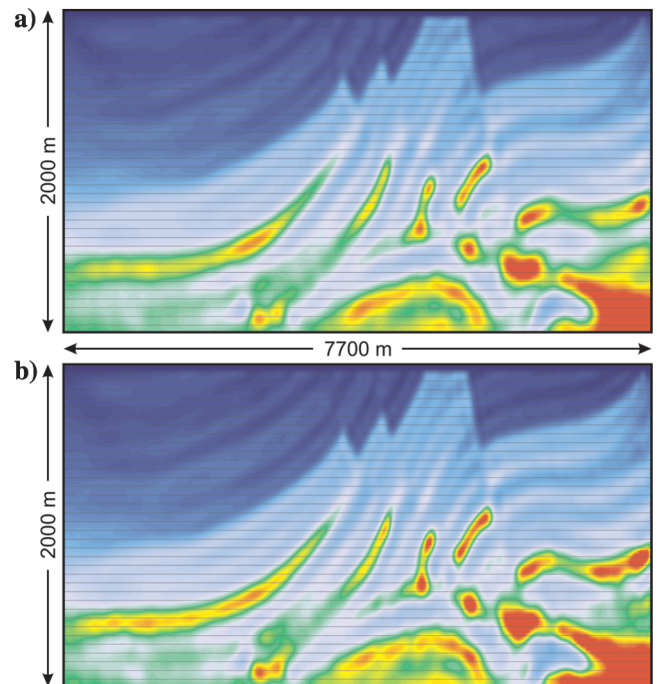


Figure 12. Inversion results for the Marmousi model generated using the correct source wavelet. (a) FWI result. (b) AWI result.

Interestingly, the synthetic data generated by model 13b, when combined with the wrong wavelet, do not match in phase to the true data, and the  $90^\circ$  phase shift in the assumed source is largely retained in the final synthetics. This characteristic of AWI means that source errors do not readily corrupt the velocity model. In contrast, the FWI result does generate synthetic data that are predominantly locked in phase to the true data, but this match in phase is now spurious. This is a more general characteristic of FWI and AWI. The former always acts to try to bring predicted and observed data into close phase alignment, whereas there is no direct requirement on AWI to phase lock the data in this way. When there is already a good match between predicted and observed data, phase locking is desirable, and it will typically lead to the highest resolution and highest fidelity in the final model. In contrast, when there are differences still to be explained in the observed data, phase locking has the potential to lead to spurious outcomes, and its avoidance by AWI is then desirable. This difference in behavior and outcome will typically lead to workflows in which inversion should begin with AWI, subsequently switching to FWI as the inversion proceeds.

## ANALYSIS AND DISCUSSION

AWI is a new method. Consequently, we would like to develop insight into how and why it works beyond the mathematics of its formulation and its performance on test data sets. Here, we explore its relationship to other methods for seismic inversion, and outline some of the ways in which AWI can be advanced beyond the simple scheme that we have described above.

### Crosscorrelation

Conventional FWI can be configured to use an objective function that is based on the crosscorrelation in time of the observed and

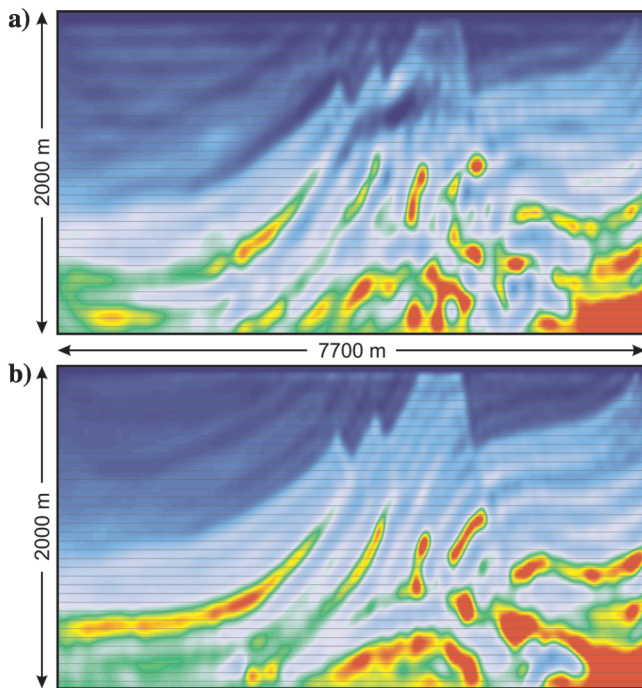


Figure 13. Inversion results for the Marmousi model generated using the incorrect source wavelet. (a) FWI result. (b) AWI result.

predicted data at the receivers. AWI is one of this class of methods, which at least conceptually, represent evolution from early work by Luo and Schuster (1991). In Appendix D, we list the objective functions and corresponding adjoint sources, using a consistent notation, for the correlation-based methods that have the most mathematical similarity to AWI.

The simplest version of a crosscorrelation method behaves similarly to conventional FWI. In that version, an objective function is formed that seeks to maximize the zero lag of the temporal crosscorrelation of the observed and predicted data. This approach can be less sensitive to missing data when it is used in conjunction with methods that form composite sources (Routh et al., 2011), but it confers no particular immunity to cycle skipping.

A crosscorrelation method that is immune to cycle skipping is introduced by van Leeuwen and Mulder (2008, 2010) for wave-equation traveltome tomography, and this method is also applicable to FWI. In this approach, temporal crosscorrelation between observed and predicted data is performed at the receivers, but instead of seeking to maximize only the amplitude of the zero lag of this, the method seeks also to reduce the amplitudes of all nonzero lags, and to increase the penalty as the magnitude of the lag increases. Their approach, therefore, is similar to AWI except that they use simple crosscorrelation rather than Wiener filtering. Two versions of this approach have been proposed: one normalized and one not; we show adjoint sources for both versions in Appendix D.

This method works well in simple models that have a single main arrival. However, the method is not effective in more complicated models that have several arrivals. In that case, there is crosstalk in the crosscorrelation between the different arrivals, and the method does not then converge to a useful result. The reason for this is clear. When the recovered model is near perfect, the observed and predicted data will be near identical. The crosscorrelation will then have high amplitudes close to zero lag. However, there will also be significant amplitude at large nonzero lags that represents the crosscorrelation of events that are separated in their arrival times. The objective function will, therefore, continue to penalize these amplitudes at large lags, and the resultant adjoint source will tend to act to drive the near-perfect model toward a new less-perfect model in an attempt to minimize the energy present at these large lags. The resultant model will not then provide a good fit to the true model.

A Wiener filter also involves the crosscorrelation of the two data sets that it seeks to match, and so it might be thought that AWI would suffer from the same problem. However, a Wiener filter also involves a deconvolution by the autocorrelation of one of the data sets. It is this feature of AWI that removes the problem of crosstalk because, when the true and recovered models are near identical, the crosscorrelation of the two data sets is practically the same as the autocorrelation of one of them, so that deconvolution by the autocorrelation collapses the crosstalk into near-zero lags. AWI, therefore, retains the cycle-skipping immunity of van Leeuwen and Mulder's (2010) approach, but it exactly overcomes the problem of crosstalk in the crosscorrelation, which otherwise limits the effectiveness of that approach for field data.

Luo and Sava (2011) introduce a method that was designed to improve the resolution of van Leeuwen and Mulder's (2010) approach. They introduce a deconvolution-based objective function, deconvolving the crosscorrelation by the autocorrelation of the observed data, with the primary motivation that this deconvolution would increase the resolution of the objective function and hence

the spatial resolution of the gradient. This approach is algorithmically similar to AWI, but does not include normalization in the objective function; that is, there is no denominator in the equivalent of equation 4. Our experience is that some form of normalization is essential in any AWI-like algorithm to produce a useful outcome. In the forward version of AWI, its omission will lead the algorithm to tend to increase the amplitudes of the predicted data without limit, so that all the Wiener coefficients become small. And in the reverse version of AWI that is directly analogous to Luo and Sava's (2011) formulation, omission of some form of normalization will tend to drive the predicted data toward zero. The best-fitting model, for the un-normalized reverse method for surface data, will then tend to evolve toward one that contains no reflectors and that includes a strong negative vertical velocity gradient, such that no energy reaches the receivers; appropriately normalized AWI does not suffer from this problem.

Examination of the adjoint sources in Appendix D for forward and reverse AWI, normalized and un-normalized weighted cross-correlation, and deconvolution approaches, shows the mathematical relationships between these methods. The adjoint sources in equations D-1 and D-2 for the two methods that are un-normalized differ significantly from those that are normalized in that they contain only one term in the adjoint source. The form of these adjoint sources also demonstrates a systematic progression in complexity as the method moves from correlation to Wiener filter, from un-normalized to normalized, and from reverse to forward AWI.

### Source inversion

Consider a single observed seismic trace  $\mathbf{d}$  generated by a single source and a single receiver. Consider also the corresponding predicted trace  $\mathbf{p}$  generated using an initial earth model  $\mathbf{m}$  and an estimated source wavelet  $\mathbf{s}$ . There are two ways that could be used in principle to modify this prediction to improve the match to the observed data. We could modify the earth model  $\mathbf{m}$ ; this is the primary objective of FWI and AWI. There is a complicated nonlinear relationship between the model  $\mathbf{m}$  and the predicted data  $\mathbf{p}$ , and so waveform-inversion methods are themselves complicated and are invariably iterative. We could, however, leave the model unchanged, and instead modify the source wavelet  $\mathbf{s}$ . Because the relationship between the predicted data and the source wavelet is linear, this second method is much more straightforward.

For a single source and receiver, and for any nonpathological model, there will always be a source wavelet that can explain the observed data, to any required degree of accuracy. The required wavelet is likely to be acausal, and it may be infinite in time in both directions, but nonetheless it can provide an accurate explanation of the observed data without modifying the earth model.

In AWI, we do not use source wavelets that are of infinite duration; instead, we use Wiener filters to find the best least-squares finite-length approximation to these infinite-source wavelets. We can, therefore, regard AWI as performing a form of source inversion that allows for a different source wavelet for every source-receiver pair. These effective source wavelets are simply the convolution of the AWI Wiener filters with the source wavelet used to synthesize the predicted data. The AWI approach then modifies the earth model so that these different effective sources evolve toward the true physical source that was used to acquire the observed data. In this process, the source inversion is linear, whereas the earth-model inversion is nonlinear; like FWI, AWI must, therefore, be iterated,

and the linear source inversion must be recomputed at each iteration.

The effective source wavelets introduced by AWI are clearly non-physical because they require a different experiment to have been performed for each source-receiver pair. Thus, an alternative view of AWI would be to regard it as a method that involves a form of model extension that introduces nonphysical sources. The inversion then modifies the earth model, such that it squeezes this nonphysical extension out of the final model to produce a physical outcome that fits the observed data and real-world physics. From this viewpoint, AWI is immune to cycle skipping because the nonphysical predicted data always provide a near-perfect fit to the observed data, and it achieves this by modifying the effective source wavelet trace by trace.

AWI involves an objective function that contains the Wiener filter coefficients. These Wiener coefficients are oscillatory, and so it might be argued that this objective function could reintroduce cycle skipping. In this objective function though, no comparison is made between two oscillating signals that might move out of phase. Instead, the objective function has regard to the properties of just one signal, and it forces this signal smoothly and continuously toward zero lag. Because there is no comparison between different signals in this objective function, there can be no possibility of cycle skipping between them.

AWI is not the only method that has been proposed, which uses a form of source inversion to form an objective function that can be solved to recover the earth model. Pratt and Symes (2002), following ideas from Symes (1994), use differential semblance between different source estimates at a single frequency to form such an objective function. In a synthetic crosshole experiment in a simple model, they show that this function has a much broader zone of attraction than that of conventional FWI, so that this approach has immunity to cycle skipping. This early work was not developed further, in part because the approach does not appear to work well for models that generate more than one arrival (W. W. Symes, personal communication, 2014). AWI is able to deal successfully with such models, and this appears to be in part because the AWI objective function also includes a feature that encourages focusing toward zero lag; such focusing is of course a feature of many other inversion schemes.

Regarding AWI as a perverse form of source inversion also perhaps indicates why the method works well when using only simple, temporal, single-channel, time-invariant Wiener filters. The wave equation itself is linear in the source wavelet, and it involves a temporal, single-channel, time-invariant convolution of the source wavelet that is exactly analogous to the operation performed during AWI. We suspect that the close match between the AWI algorithm and this characteristic of the wave equation helps to make AWI particularly effective for seismic inversion.

### Matching filters

AWI uses Wiener filters to adapt one data set to another. This is an entirely general approach, and many other forms of matching filter might be used to similar effect. We note first that conventional FWI can also be considered to proceed by using a perverse form of matching. In FWI, the matching filter is trivial. It operates on the predicted data by adding the observed data and subtracting the predicted data. This provides a perfect match to the observed data. The identity filter is obtained when this matching filter leaves



the input data unchanged, which means that the observed minus the predicted data must be zero, and so we arrive at conventional FWI. With this view of FWI, the matching filter is immune to cycle skipping because the data match is perfect, but the objective function reintroduces the cycle skipping because it involves the comparison of two oscillating signals.

We could attempt to construct a similar trivial matching filter in which the predicted data are multiplied by the observed data and divided by the predicted data. This also provides a perfect matching provided that the predicted data are never zero. This is the approach taken by AWI with the additional feature that the required multiplication and division take place, suitably stabilized to avoid division by zero, in the temporal frequency domain. Multiplication and division in the frequency domain are of course equivalent to convolution and deconvolution in the original time domain.

We can then write the problems that FWI and AWI seek to solve, at least conceptually, as

$$\text{FWI: } \mathbf{p} - \mathbf{d} \rightarrow 0, \quad \text{AWI: } \mathbf{p}/\mathbf{d} \rightarrow 1, \quad (15)$$

so that FWI acts to force the difference of two data sets to zero, whereas AWI acts to force the ratio of two data sets to unity, and Wiener filters provide the stabilized least-squares version of this conceptual division of one data set by the other. If the observed and predicted data are perfect, that is if the former has no noise and the latter involves all the correct physics, and if both inversion schemes are driven to their global minima, then both methods will produce the same final outcome. However, if these methods are implemented using a linearized iterative local-descent method, then the routes that the two methods will follow through model space in their attempts to reach the global minimum will typically be quite different; one can lead to a cycle-skipped local minimum, whereas the other can lead successfully to the global minimum. We note also that, if the data or its simulation is imperfect, then the two methods need not lead to the same global minimum.

Ma and Hale (2013) introduce a more explicit form of matching into FWI, using dynamic time warping to match the two data sets. Unlike AWI, this method and others like it assume that the matching is bijective and diffeomorphic; that is, they assume that each point or event in one data set corresponds uniquely and smoothly to an equivalent point or event in the other. This restriction means that the predicted data must have the correct number of events present to match the observed data, and more significantly that geometric complications should have the same topology in both data sets. This requires events in the two data sets to appear in the same temporal order, and that caustics, shadows, multipathing, and analogous features should appear with similar geometries in both data sets. In complicated models, this represents a significant restriction that is difficult to ensure in field data; typically, it will require that the starting model must already provide a close match to the true model, or that the true model be very simple.

We reiterate that AWI does not assume any relationship between events in one data set and the other, it does not explicitly match one event to another, and it is not analogous to the use of short-length Wiener filters to match local wavelets. Instead, AWI uses a filter that is typically longer than the trace length to match one entire trace to another entire trace that may contain significantly different numbers and categories of events that may arrive in a different temporal sequence. AWI makes no assumption about starting from a close

match, and it will tend to act to move, to transpose, to create, or to destroy events in the predicted data as required to match the observed data. In contrast, in conventional seismic processing, most of the operations that are labeled as matching filters do not behave in this way; rather, they are designed to operate locally in a more limited fashion. Long-duration Wiener filters instead operate globally across a trace, matching every part of one trace into every part of the other.

### Wave-equation migration velocity analysis

Wave-equation migration velocity analysis (WEMVA) is an automated inversion scheme that seeks to recover the background macrovelocity model from pure-reflection data (Shen et al., 2003; Sava and Biondi, 2004). It operates in the image domain rather than in the data domain that is used in conventional FWI and in AWI. It does not suffer from cycle skipping, but it cannot easily deal with surface or interbed multiples, or with other nonprimary arrivals. In 3D, WEMVA is typically more expensive than FWI or AWI.

Most WEMVA-like schemes involve the concept of a nonphysical extended model (Symes, 2008), and they setup the inversion such that models evolve toward physical outcomes. In conventional WEMVA, the model is extended by introducing a subsurface offset. This represents nonphysical scattering in the subsurface, whereby an incident wavefield at one location generates a coeval scattered wavefield at another location. The inversion is then formulated to focus the energy to zero subsurface offset, thus producing a physical outcome in which the incident and scattered wavefields are coincident.

AWI has important features in common with WEMVA, and it extends the model in an analogous nonphysical way; it is this feature that makes AWI immune to cycle skipping. In AWI, the Wiener filters can be regarded as a means of redistributing energy, nonphysically, in time. In AWI, energy arriving at a receiver at a particular time produces a signal at that receiver that extends to earlier and later times. These nonphysical arrivals disappear when the filter becomes a zero-lag delta function; this corresponds to a physical outcome. Conventional WEMVA involves nonphysical action at a distance in the subsurface, whereas AWI involves nonphysical interaction across time at the receivers. We can therefore regard AWI as a data-domain analog of WEMVA that seeks to focus energy — all energy and not only primary reflections — to zero temporal lag at all the receivers just as WEMVA seeks to focus primary reflections to zero spatial lag at all points in the subsurface.

WEMVA, however, has some important disadvantages that are not shared by AWI. First, WEMVA involves a convolution in the subsurface in space at all points in the model, for all time steps. In 3D models, this WEMVA convolution can become prohibitively expensive, especially if subsurface offset is treated as a vector rather than a scalar quantity. Second, because WEMVA operates in the image domain, it can properly image only primary events; multiples of all kinds and other nonprimary events will not focus in the same velocity model as do the primaries. If nonprimary arrivals are included within its input data, then the model recovered by WEMVA will represent a compromise between the various different velocity models that fit these various different types of event. In practice, this can be a serious problem (Mulder, 2008) even when the input data are nominally free of surface-related multiples, interbed multiples, and mode conversions. In contrast because AWI operates in the data domain, it does not suffer from this problem, and it does not require



data to be preprocessed to remove particular categories of arrival. Finally, WEMVA is not a high-resolution method. It aims to match the kinematics of primary reflections, and it does not attempt to fit detailed waveforms; some version of the latter is normally a requirement of methods that attempt to resolve more finely than that of a single Fresnel zone.

### Advanced methods

Several new waveform-inversion methods have appeared recently that also have abilities to overcome cycle skipping. In common with AWI, these methods have yet to move substantially beyond the groups that initially developed them, and until that happens, it will be difficult for others to assess fully their utility, potential, and limitations.

Biondi and Almomin (2012, 2015) introduce a methodology that combines a form of WEMVA with a version of FWI to produce the method of tomographic FWI (T-FWI). This approach appears able to combine the high spatial resolution of FWI with the cycle-skipping immunity of WEMVA. In T-FWI, focusing in subsurface spatial offset is replaced by focusing in subsurface temporal offset; AWI uses a similar approach, but limits its attention only to receiver locations with a consequent reduction in computational cost. T-FWI also introduces temporal offset directly into the velocity model and the wave equation such that the observed data can always be well-matched by the predicted data. T-FWI appears to work well on synthetic data, and the early success of this method on the Marmousi model (Biondi and Almomin, 2012) was one of the drivers that encouraged us to pursue AWI to a successful conclusion. However, T-FWI also appears to have two disadvantages that are not shared by AWI: The method is much more computationally demanding than conventional FWI, especially in 3D, and the observed data require preprocessing so that only primary arrivals are used within the WEMVA portion of the inversion.

van Leeuwen and Herrmann (2013) and van Leeuwen et al. (2014) introduce wavefield reconstruction inversion (WRI) as an alternative to FWI that is able to circumvent local minima in the conventional FWI objective function. Like AWI, this method introduces a nonphysical model extension, and then attempts to recover an earth model by collapsing this extended model onto a physical model that is still able to explain the observed data. Unlike AWI, this method extends the model by relaxing the requirement that the subsurface wavefield should honor the wave equation. The wave equation is then used to focus this extended model onto a physical model. In contrast, AWI relaxes the requirement that the wavefield should honor the wave equation at the receivers. WRI appears to work well in 2D in the frequency domain; it remains an open question as to whether it can be made to work efficiently in 3D, and whether it can be extended to the anisotropic time-stepping codes that are commonly used in commercial FWI.

Several authors (Alkhalifah and Choi, 2012; Shah et al., 2012; Choi and Alkhalifah, 2014) have attempted to solve cycle skipping during FWI by unwrapping phase in various ways. The most developed of these approaches appears to be that of Jiao et al. (2015), who introduce matching pursuit FWI. Automated phase unwrapping of field data is difficult, and although this suite of methods appears to be promising, their potential has yet to be fully demonstrated. These approaches do not seem to be obviously related to AWI; they appear to work best, and perhaps only, for models that

generate clear isolated arrivals, or that can be preprocessed to generate such an outcome.

AWI, T-FWI, WRI, and phase-unwrapping FWI schemes are all attempts to overcome the most important limitation of conventional FWI. All these methods appear able to achieve that on appropriate synthetic data; the challenge for each of these methods, including AWI, is to demonstrate them robustly and efficiently on full 3D field data.

### Pathological models

It is possible to devise simple pathological models for which AWI seemingly behaves inappropriately; however, none of these examples will easily arise during the practical inversion of real data. We discuss some of these pathological cases here.

The objective function expressed by equation 4 is insensitive to the polarity of the predicted trace with respect to the observed trace. Thus, it might be supposed to be possible for AWI to recover a model, at the global minimum, for which the polarity of the predicted data was reversed, perhaps by spuriously reversing the polarity of every reflector for one or more source-receiver pairs. Note that, for this effect to manifest, it would be necessary to reverse the polarity of all arrivals on a particular trace; the effect cannot appear merely by reversing a subset of arrivals, which will always produce an increased AWI mismatch. In any real-world data set, there will always be at least one reference arrival that is impossible, or at least extremely difficult, to reverse using any reasonable physical model. These include the direct arrival, the primary sea-bottom reflection, and most primary refractions, head waves, and turning arrivals. Provided that at least one of these reference arrivals is present in the field data, then the global minimum will always correspond to a model that also has the correct polarity for all other events. We note, however, that, unlike most other waveform-inversion methodologies, AWI will deal correctly with a data set within which a subset of traces have had their polarity spuriously reversed, or for which the assumed source polarity is erroneous.

A second pathological possibility is apparently obtained when the predicted and observed data differ only by a constant time shift. In this case, the Wiener filter appears to consist of one nonzero coefficient at a nonzero lag. This circumstance leads to a value of zero for the expression  $(2f\mathbf{I} - \mathbf{T}^2)\mathbf{w}$  in equation 13, so that the adjoint source, and hence the calculated gradient, both become zero even though the earth model is clearly in error. In any real data set, however, this circumstance will not arise because it rests on an implicit assumption of infinite bandwidth. More correctly, in a discrete rather than a continuous formulation, this conclusion assumes that the bandwidth extends to the Nyquist frequency with amplitudes that lie well above the level of stabilization applied in deriving the Wiener filter. In any real observed or simulated data set, the bandwidth will be finite, and in any discrete data set, the bandwidth will not extend to the Nyquist frequency at high amplitude. Consequently, the Wiener filter that is obtained by matching two time-shifted and otherwise identical data sets does not contain only a single nonzero coefficient. Rather, it consists of a band-limited delta function of finite temporal duration. This filter does not then lead to a value of zero in equation 13; rather it produces the correct nonzero adjoint source, which leads to a finite gradient that continues to indicate the direction required to improve the model. Thus, in any real problem, this apparent pathology is not manifest. It is worth noting that, by using an assumption of infinite bandwidth, conventional

FWI can also be made to produce analogous pathological outcomes in which the gradient is zero for arbitrarily incorrect earth models.

The final possibility that we consider here is that obtained when either the observed or predicted data contain no arrivals. Such a circumstance will of course break AWI, and any practical implementation will therefore need to deal sensibly with dead traces. More significantly though is the case of reflection-dominated field data sets when the starting model is sufficiently smooth that it does not generate reflections. It might be thought therefore that AWI cannot deal with this circumstance; Figure 10 demonstrates that it can. All that is required is that there is some arrival of some kind present in each predicted data set. This arrival can be a direct arrival, it can be a water-bottom reflection, it can be a multiple, and indeed it can be anything that is source generated. The long-duration Wiener filters used by AWI can and do generate a full suite of reflected arrivals from one unrelated predicted arrival, and these then serve to bootstrap AWI by migrating those reflections into the starting model.

### Enhancements and developments

Wiener filters are well-understood. They can operate in space rather than time, they can be extended into two or more dimensions, they can be time varying, they can be regularized so that they vary smoothly in some dimensions, and they can be extended so that they represent a generalized multidimensional convolution that uses a different operator for every output sample. All these variations potentially lead to variations in the AWI methodology.

Our early numerical experiments appear to show that the temporal version of AWI that we have presented here is significantly superior to AWI schemes that use Wiener filters implemented in space within individual shot records. We speculate that this is because time-domain convolution arises naturally within the wave equation if the earth model does not vary with time over the duration of a shot record, which it does not normally do. In contrast, the earth model does typically vary over space within the extent of a shot record, so that spatial convolution does not have the same intimate relationship to the wave equation. Spatial convolution in the shot domain for field data will often also be complicated by irregular acquisition geometry, and by the necessity for multidimensional Wiener filters in 3D surveys. Spatial implementations, however, would perhaps represent one way that an AWI-like scheme could be configured in the sparse temporal-frequency domain.

Wiener filters that vary in time are widely used in seismic processing. They are an obvious choice around which to build an AWI-like scheme, although, they do not have the same relationship to the wave equation that the time-stationary Wiener filter has. If the time window used to construct the time-variant filter is rather short, then this version of AWI comes more to resemble FWI based on dynamic time warping. Although the time-variant version of AWI will not be strictly diffeomorphic, it will begin to map single events in one data set into single events in the other. A great advantage of conventional FWI, and of the simple time-invariant AWI method, is that there need be no initial one-to-one correspondence between events in the predicted and observed data sets. Indeed, there seldom is such a correspondence because the starting model is typically smooth, whereas the true earth model is not. A time-variant version of AWI, therefore, will have nothing to match in the predicted data at late times in typical starting models, and the method will not then be able to build the required structure because the time-variant filter will not properly

capture this from the observed data. Using long, time-stationary Wiener filters for AWI does not suffer from this problem.

The Wiener filters generated by AWI can vary laterally quite rapidly. Some form of regularization of the Wiener filters, or of the adjoint source, can serve to suppress this effect, and Wiener filters that vary smoothly in space are often used in conventional seismic processing. It is also possible to configure AWI so that it uses an objective function other than that expressed in equation 4; all that is required is that the function serves to drive the Wiener coefficients toward zero lag, that it is constructed so that it does not also drive them in some other undesirable direction, for example, toward very small or very large values, and that the functional itself does not reintroduce cycle skipping. There are likely to be many such possibilities.

The AWI method appears to have complete and unconditional immunity to cycle skipping. This is because it uses an objective function that does not form differences of oscillatory functions, and consequently does not have minima when the predicted data are shifted by an integer number of cycles with respect to the observed data. This immunity to cycle skipping does not of course confer immunity to other causes of nonlinearity and local minima that are unrelated to cycle skipping, and it is probable that at least some of these other causes will become increasingly important as the starting velocity model moves further from reality. Only extensive testing using synthetic and field data will ultimately reveal how far AWI can be pushed before it fails. Our limited tests to date suggest that AWI can be expected to function satisfactorily when starting-model errors are around three times as large as those for which FWI fails; that is, AWI seems to be able to deal successfully with starting models that predict data that are incorrect in time by approximately one and a half cycles, whereas FWI will succeed only when this initial uncertainty lies below half a cycle. In practice, for many field data sets, the former can be achieved readily by picking stacking velocities on prestack time-migrated reflection data, whereas the later will more typically often require iterated reflection traveltime tomography applied to prestack depth-migrated data.

### CONCLUSION

AWI appears to provide a form of FWI in the data domain that is immune to the effects of cycle skipping. It achieves this by extending the model space nonphysically, so that it becomes of a similar size to the data space. The extension can be thought of as either allowing for a different effective source wavelet to be used for each source-receiver pair, or equivalently by allowing for a nonphysical interaction across time at the individual receivers that is analogous to the nonphysical interaction in subsurface offset that is introduced in WEMVA-like methods. However, in contrast to most other useful model extensions, AWI is computationally inexpensive and is straightforward to implement in 3D time-domain codes. It is applied trace by trace, it requires no communication between different sources, it requires no additional operations to be undertaken in the subsurface, it involves only 1D operations, and the model extension is performed only at the individual receiver positions.

Our analysis does not rule out that there may be other types of local minima to which AWI might be especially susceptible that are unrelated to cycle skipping. The synthetic tests presented here, however, demonstrate that any such susceptibility is not readily manifest, and AWI appears to be effective over a wide range of models and data sets. Conventional FWI is itself susceptible to various

types of local minima that are unrelated to cycle skipping. AWI will undoubtedly be susceptible to at least some of these same issues, but it is not susceptible to the more serious problem caused by cycle skipping, and it does not appear to have any obvious new susceptibilities that are not also displayed by conventional FWI.

## ACKNOWLEDGMENTS

We thank Sub Salt Solutions Limited for permission to publish this work. AWI and adaptive waveform inversion are registered trademarks. AWI is subject to GB patent number GB1319095.4; international patent protection is pending.

## APPENDIX A

### DERIVATION OF THE REVERSE-AWI GRADIENT VIA THE ADJOINT-STATE METHOD

To find the gradient of the objective function  $f$  with respect to the model parameters  $\mathbf{m}$ , where

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \quad \text{subject to} \quad \mathbf{v} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{R}\mathbf{u} \quad \mathbf{A}\mathbf{u} = \mathbf{s} \quad (\text{A-1})$$

we define Lagrange multipliers  $\lambda_{\mathbf{v}}$  and  $\lambda_{\mathbf{u}}$ , and form the Lagrangian functional  $\mathcal{L}$

$$\mathcal{L}(\mathbf{m}, \mathbf{v}, \mathbf{u}) = f + \langle \lambda_{\mathbf{v}}, \mathbf{v} - (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{R}\mathbf{u} \rangle + \langle \lambda_{\mathbf{u}}, \mathbf{A}\mathbf{u} - \mathbf{s} \rangle. \quad (\text{A-2})$$

Now, we find the derivatives of the Lagrangian with respect to  $\mathbf{v}$  and  $\mathbf{u}$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{v}} &= \frac{1}{\mathbf{v}^T\mathbf{v}} (\mathbf{T}^2 - 2\mathbf{I}f)\mathbf{v} + \lambda_{\mathbf{v}} \\ \frac{\partial \mathcal{L}}{\partial \mathbf{u}} &= -[(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\mathbf{R}]^T\lambda_{\mathbf{v}} + \mathbf{A}^T\lambda_{\mathbf{u}}. \end{aligned} \quad (\text{A-3})$$

Setting these to zero, solving for the Lagrange multipliers, and substituting these results into the derivative of the Lagrangian with respect to the model parameters  $\mathbf{m}$  gives

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \mathbf{m}} &= \left( \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \mathbf{u} \right)^T \lambda_{\mathbf{u}} \\ &= -\mathbf{u}^T \left( \frac{\partial \mathbf{A}}{\partial \mathbf{m}} \right)^T \mathbf{A}^{-T} \mathbf{R}^T \mathbf{D} (\mathbf{D}^T \mathbf{D})^{-1} \left( \frac{\mathbf{T}^2 - 2f\mathbf{I}}{\mathbf{v}^T \mathbf{v}} \right) \mathbf{v}, \end{aligned} \quad (\text{A-4})$$

which is the same as the gradient obtained by combining equations 11 and 14.

## APPENDIX B

### DERIVATION OF THE ADJOINT SOURCE FOR FORWARD AWI BY DIRECT DIFFERENTIATION

To derive the expression for the adjoint source  $\delta\mathbf{s}$  shown in equation 13, first, find the differential of the Wiener filter  $\mathbf{w}$  with respect to the predicted data  $\mathbf{p}$

$$\begin{aligned} \frac{\partial \mathbf{w}}{\partial \mathbf{p}} &= (\mathbf{P}^T\mathbf{P})^{-1} \frac{\partial \mathbf{P}^T}{\partial \mathbf{p}} \mathbf{d} - (\mathbf{P}^T\mathbf{P})^{-1} \frac{\partial \mathbf{P}^T}{\partial \mathbf{p}} \mathbf{P} (\mathbf{P}^T\mathbf{P})^{-1} \mathbf{P}^T \mathbf{d} \\ &\quad - (\mathbf{P}^T\mathbf{P})^{-1} \mathbf{P}^T \frac{\partial \mathbf{P}}{\partial \mathbf{p}} (\mathbf{P}^T\mathbf{P})^{-1} \mathbf{P}^T \mathbf{d}. \end{aligned} \quad (\text{B-1})$$

Now because  $\mathbf{P}\mathbf{w} = \mathbf{d}$ , we have  $\mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{d} = \mathbf{d}$ . Thus, the first two terms in equation B-1 cancel, and the remaining term simplifies to give

$$\frac{\partial \mathbf{w}}{\partial \mathbf{p}} = -(\mathbf{P}^T\mathbf{P})^{-1} \mathbf{P}^T \frac{\partial \mathbf{P}}{\partial \mathbf{p}} \mathbf{w} = -(\mathbf{P}^T\mathbf{P})^{-1} \mathbf{P}^T \mathbf{W}, \quad (\text{B-2})$$

where the last step is obtained because convolution is commutative so that

$$\frac{\partial \mathbf{P}}{\partial \mathbf{p}} \mathbf{w} = \mathbf{W} \frac{\partial \mathbf{p}}{\partial \mathbf{p}} = \mathbf{W}, \quad (\text{B-3})$$

where  $\mathbf{W}$  is a Toeplitz matrix containing  $\mathbf{w}$  in each column.

Using the definition of  $f$  given in equation 4, and the definition of the adjoint source given in equation 12, the adjoint source can thus be written as

$$\delta\mathbf{s} = \frac{\partial f}{\partial \mathbf{p}} = \frac{1}{(\mathbf{w}^T\mathbf{w})^2} \left[ \mathbf{w}^T\mathbf{w} \left( \mathbf{T} \frac{\partial \mathbf{w}}{\partial \mathbf{p}} \right)^T \mathbf{T}\mathbf{w} - (\mathbf{T}\mathbf{w})^T \mathbf{T}\mathbf{w} \frac{\partial \mathbf{w}^T}{\partial \mathbf{p}} \mathbf{w} \right], \quad (\text{B-4})$$

and this can be simplified using equation B-2 and the definition of  $f$  to give the final result

$$\delta\mathbf{s} = \mathbf{W}^T \mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \left( \frac{2f\mathbf{I} - \mathbf{T}^2}{\mathbf{w}^T \mathbf{w}} \right) \mathbf{w}. \quad (\text{B-5})$$

## APPENDIX C

### MODELING AND INVERSION PARAMETERS FOR SYNTHETIC EXAMPLES

The parameters used for the modeling and inversion are stated below. These parameters may not be optimal, necessary, or sufficient for a particular outcome.

#### All examples

The time-domain forward modeling is isotropic acoustic and includes a variable density parameterized using Gardner's law. Free-surface ghosts and all multiples are included. The inversions use steepest descent, preconditioned using an approximate diagonal Hessian; the code includes numerous heuristics designed to speed convergence and mitigate against noise and spurious amplitude mismatches — the more important of these are outlined in Warner et al. (2013). No model-domain regularization was used. For the AWI inversions, we used a simple reverse AWI algorithm with 10% noise stabilization, a maximum lag equal to the data length, and a Gaussian weighting function configured so that the objective function is to be maximized. In all three examples, the weighting function was the same, and the standard deviation of the Gaussian function was 5% of the maximum lag used.

### Example 1

The model measures  $7680 \times 2000$  m with a mesh spacing of 40 m. We used 91 sources and 187 receivers; sources and receivers were located at a depth of 37.5 m below the free surface giving a peak amplitude at 10 Hz. The 1D starting model was built by picking typical moveouts on time-migrated gathers; it increases approximately linearly from 1550 m/s at the seabed to 2550 m/s at 1000 m depth, then less slowly to reach a velocity of 3250 m/s at the base of the model. The source was a Ricker wavelet at 10 Hz. The inversion was run in the time domain for 200 iterations, using all sources every iteration and a low-pass filter at 10 Hz.

### Example 2

The model measures  $20 \times 5$  km with a mesh spacing of 50 m. We used 200 sources and 400 receivers; sources were one cell below the free surface, and receivers were close to the sea bed. The checkers measure  $500 \times 500$  m, and represent a  $\pm 10\%$  perturbation of velocity except in the shallowest section, where the velocity is constrained not to fall below 1550 m/s so that the perturbation falls to  $\pm 7\%$ . The starting velocity model increases linearly from 1550 m/s at the sea bed to 2000 m/s at the base of the model, and it is approximately 7% lower than the mean of the true model. The source was a Ricker wavelet at 4 Hz. The inversion was run in the time domain for 50 inversions using all sources every iteration using a single 5 Hz low-pass filter.

### Example 3

Details are as in example 1 except that the source signatures used were as shown in Figure 11, the starting model was a smoothed version of the true model, and the inversion was run for 25 iterations at each of 4, 5, 6, 8, 10, and 12 Hz low-pass filters.

## APPENDIX D

### OBJECTIVE FUNCTIONS AND ADJOINT SOURCES FOR RELATED METHODS

#### Crosscorrelation without normalization (van Leeuwen and Mulder, 2010)

$$f = \frac{1}{2} \|\mathbf{T}\mathbf{c}\|^2, \quad \mathbf{c} = \mathbf{D}^T \mathbf{p}, \quad \delta \mathbf{s} = \mathbf{D} \mathbf{T}^2 \mathbf{c}. \quad (\text{D-1})$$

#### Deconvolution without normalization (Luo and Sava, 2011)

$$f = \frac{1}{2} \|\mathbf{T}\mathbf{v}\|^2, \quad \mathbf{v} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{p}, \quad \delta \mathbf{s} = \mathbf{D} (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{T}^2 \mathbf{v}. \quad (\text{D-2})$$

#### Crosscorrelation with normalization (van Leeuwen and Mulder, 2008)

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{c}\|^2}{\|\mathbf{c}\|^2}, \quad \mathbf{c} = \mathbf{D}^T \mathbf{p}, \quad \delta \mathbf{s} = \mathbf{D} \left( \frac{\mathbf{T}^2 - 2f\mathbf{I}}{\mathbf{c}^T \mathbf{c}} \right) \mathbf{c}. \quad (\text{D-3})$$

### Reverse AWI (this paper)

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{v}\|^2}{\|\mathbf{v}\|^2}, \quad \mathbf{v} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{p}, \quad \delta \mathbf{s} = \mathbf{D} (\mathbf{D}^T \mathbf{D})^{-1} \left( \frac{\mathbf{T}^2 - 2f\mathbf{I}}{\mathbf{v}^T \mathbf{v}} \right) \mathbf{v}. \quad (\text{D-4})$$

### Forward AWI (this paper)

$$f = \frac{1}{2} \frac{\|\mathbf{T}\mathbf{w}\|^2}{\|\mathbf{w}\|^2}, \quad \mathbf{w} = (\mathbf{P}^T \mathbf{P})^{-1} \mathbf{P}^T \mathbf{d}, \quad \delta \mathbf{s} = \mathbf{W}^T \mathbf{P} (\mathbf{P}^T \mathbf{P})^{-1} \left( \frac{2f\mathbf{I} - \mathbf{T}^2}{\mathbf{w}^T \mathbf{w}} \right) \mathbf{w}. \quad (\text{D-5})$$

## REFERENCES

- Alkhalifah, T., and Y. Choi, 2012, Taming waveform inversion non-linearity through phase unwrapping of the model and objective functions: *Geophysical Journal International*, **191**, 1171–1178.
- Biondi, B., and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: [10.1190/segam2012-0275.1](https://doi.org/10.1190/segam2012-0275.1).
- Biondi, B., and A. Almomin, 2015, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: *Geophysics*, **79**, no. 3, WA129–WA140, doi: [10.1190/geo2013-0340.1](https://doi.org/10.1190/geo2013-0340.1).
- Choi, Y., and T. Alkhalifah, 2014, Two-dimensional unwrapped phase inversion with damping and a Gaussian filter: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20140702](https://doi.org/10.3997/2214-4609.20140702).
- Guasch, L., and M. Warner, 2014, Adaptive waveform inversion: Applications: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20141093](https://doi.org/10.3997/2214-4609.20141093).
- Hansen, P., 2002, Deconvolution and regularization with Toeplitz matrices: *Numerical Algorithms*, **29**, 323–378, doi: [10.1023/A:1015222829062](https://doi.org/10.1023/A:1015222829062).
- Jiao, K., D. Sun, X. Cheng, and D. Vigh, 2015, Matching pursuit full-waveform inversion: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201412911](https://doi.org/10.3997/2214-4609.201412911).
- Kapoor, S., D. Vigh, E. Wiarda, and S. Alwon, 2013, Full waveform inversion around the world: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20130827](https://doi.org/10.3997/2214-4609.20130827).
- Lazaratos, S., I. Chikichev, and K. Wang, 2011, Improving the convergence rate of full wavefield inversion using spectral shaping: 81st Annual International Meeting, SEG, Expanded Abstracts, 2428–2432.
- Lu, R., S. Lazaratos, K. Wang, Y. H. Cha, I. Chikichev, and R. Prosser, 2013, High-resolution elastic FWI for reservoir characterization: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20130113](https://doi.org/10.3997/2214-4609.20130113).
- Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792.
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltimes inversion: *Geophysics*, **56**, 645–653, doi: [10.1190/1.1443081](https://doi.org/10.1190/1.1443081).
- Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltimes inversion with dynamic warping and full-waveform inversion: *Geophysics*, **78**, no. 6, R223–R233, doi: [10.1190/geo2013-0004.1](https://doi.org/10.1190/geo2013-0004.1).
- Mancini, F., J. Moss, K. Prindle, and T. Ridsdill-Smith, 2015, Where can full waveform inversion have the biggest impact in the exploration and production cycle?: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201413199](https://doi.org/10.3997/2214-4609.201413199).
- Mulder, W. A., 2008, Automatic velocity analysis with the two-way wave equation: 70th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20147941](https://doi.org/10.3997/2214-4609.20147941).
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain. Part I: Theory and verification in a physical scale model: *Geophysics*, **64**, 888–901, doi: [10.1190/1.1444597](https://doi.org/10.1190/1.1444597).
- Pratt, R. G., C. Shin, and G. H. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362, doi: [10.1046/j.1365-246X.1998.00498.x](https://doi.org/10.1046/j.1365-246X.1998.00498.x).



- Pratt, R. G., and W. Symes, 2002, Semblance and differential semblance optimization for waveform tomography: A frequency domain implementation: *Journal of Conference Abstracts*, **7**, 183–184.
- Routh, P. S., J. R. Krebs, S. Lazaratos, A. I. Baumstein, I. Chikichev, S. Lee, N. Downey, D. Hinkley, and J. E. Anderson, 2011, Full-wavefield inversion of marine streamer data with the encoded simultaneous source method: 73rd Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20149730](https://doi.org/10.3997/2214-4609.20149730).
- Sava, P., and B. Biondi, 2004, Wave-equation migration velocity analysis. I: Theory: *Geophysical Prospecting*, **52**, 593–606, doi: [10.1111/gpr.2004.52.issue-6](https://doi.org/10.1111/gpr.2004.52.issue-6).
- Shah, N. K., M. R. Warner, J. K. Washbourne, L. Guasch, and A. P. Umpleby, 2012, A phase-unwrapped solution for overcoming a poor starting model in full-wavefield inversion: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20148455](https://doi.org/10.3997/2214-4609.20148455).
- Shen, P., W. W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2132–2135.
- 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, doi: [10.3997/1365-2397.20100012](https://doi.org/10.3997/1365-2397.20100012).
- Symes, W. W., 1994, Inverting waveforms for velocities in the presence of caustics: *Proceeding of the Mathematical Methods in Geophysical Imaging II*, SPIE, **2301**, doi: [10.3997/10.1117/12.187492](https://doi.org/10.3997/10.1117/12.187492).
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790, doi: [10.1111/gpr.2008.56.issue-6](https://doi.org/10.1111/gpr.2008.56.issue-6).
- van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: *Geophysical Journal International*, **195**, 661–667, doi: [10.1093/gji/ggt258](https://doi.org/10.1093/gji/ggt258).
- van Leeuwen, T., F. J. Herrmann, and B. Peters, 2014, A new take on FWI: Wavefield reconstruction inversion: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20140703](https://doi.org/10.3997/2214-4609.20140703).
- van Leeuwen, T., and W. A. Mulder, 2008, Velocity analysis based on data correlation: *Geophysical Prospecting*, **56**, 791–803, doi: [10.1111/gpr.2008.56.issue-6](https://doi.org/10.1111/gpr.2008.56.issue-6).
- van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation travel time tomography: *Geophysical Journal International*, **182**, 1383–1394, doi: [10.1111/j.1365-246X.2010.04681.x](https://doi.org/10.1111/j.1365-246X.2010.04681.x).
- Vigh, D., B. Starr, J. Kapoor, and H. Li, 2010, 3D full waveform inversion on a Gulf of Mexico WAS data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 957–961.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, no. 6, WCC1–WCC26, doi: [10.1190/1.3238367](https://doi.org/10.1190/1.3238367).
- Warner, M., A. Ratcliffe, T. Nangoo, J. V. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic full-waveform inversion: *Geophysics*, **78**, no. 2, R59–R80, doi: [10.1190/geo2012-0338.1](https://doi.org/10.1190/geo2012-0338.1).
- Warner, M. R., and L. Guasch, 2014a, Adaptive waveform inversion: Theory: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.20141092](https://doi.org/10.3997/2214-4609.20141092).
- Warner, M. R., and L. Guasch, 2014b, Adaptive waveform inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093.
- Warner, M. R., and L. Guasch, 2015, Robust adaptive waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstract, 1059–1063.
- Yao, G., and M. Warner, 2015, Bootstrapped waveform inversion: Long-wavelength velocities from pure reflection data: 77th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: [10.3997/2214-4609.201413208](https://doi.org/10.3997/2214-4609.201413208).