

Full-waveform inversion with extrapolated low-frequency data

Yunyue Elita Li¹ and Laurent Demanet¹

ABSTRACT

The availability of low-frequency data is an important factor in the success of full-waveform inversion (FWI) in the acoustic regime. The low frequencies help determine the kinematically relevant, low-wavenumber components of the velocity model, which are in turn needed to avoid convergence of FWI to spurious local minima. However, acquiring data less than 2 or 3 Hz from the field is a challenging and expensive task. We have explored the possibility of synthesizing the low frequencies computationally from high-frequency data and used the resulting prediction of the missing data to seed the frequency sweep of FWI. As a signal-processing problem, bandwidth extension is a very nonlinear and delicate operation. In all but the simplest of scenarios, it can only be expected to lead to plausible recovery of the low frequencies, rather than their accurate reconstruction. Even so, it still requires a high-level interpretation of band-limited seismic records into individual events,

each of which can be extrapolated to a lower (or higher) frequency band from the nondispersive nature of the wave-propagation model. We have used the phase-tracking method for the event separation task. The fidelity of the resulting extrapolation method is typically higher in phase than in amplitude. To demonstrate the reliability of bandwidth extension in the context of FWI, we first used the low frequencies in the extrapolated band as data substitute, to create the low-wavenumber background velocity model, and then we switched to recorded data in the available band for the rest of the iterations. The resulting method, extrapolated FWI, demonstrated surprising robustness to the inaccuracies in the extrapolated low-frequency data. With two synthetic examples calibrated so that regular FWI needs to be initialized at 1 Hz to avoid local minima, we have determined that FWI based on an extrapolated [1, 5] Hz band, itself generated from data available in the [5, 15] Hz band, can produce reasonable estimations of the low-wavenumber velocity models.

INTRODUCTION

Since proposed by [Tarantola \(1984\)](#), full-waveform inversion (FWI) has established itself as the default wave-equation-based inversion method for subsurface model building. In contrast to ray-based traveltime tomography ([Woodward et al., 2008](#)), FWI includes phase and amplitude information in the seismograms for elastic-parameter estimation. When working with reflection data, the model update of FWI is similar to a migration image of the data residual, given the background propagation velocity ([Claerbout, 1985](#); [Biondi and Symes, 2004](#)). The accuracy of the resolved model is controlled by the frequency band in the data and the accuracy of the initial (background and macro) model. When this initial macro model is not sufficiently accurate, the iterative process of FWI gets trapped in undesirable local minima or valleys ([Virieux and Operto, 2009](#)). The lack of convexity is intrinsic and is due to the relatively high frequencies of the seismic waveforms.

Global optimization techniques could overcome the nonconvexity in theory; however, their computational cost for million-dimensional model space is often prohibitively expensive for FWI. Therefore, most FWI algorithms are based on local gradient-based iterative techniques. In spite of an extensive literature on the subject, a convincing solution has yet to emerge for mitigating these convergence issues. So far, the community's efforts can be grouped into three categories.

In the first category, misfit functions different from least squares have been proposed to emulate traveltime shifts between the modeled and recorded waveforms. [Luo and Schuster \(1991\)](#) propose to combine traveltime inversion with FWI, in which the misfit is measured over the maximum correlation time lag. [Ma and Hale \(2013\)](#) and [Baek et al. \(2014\)](#) propose variants of FWI augmented with warping functions directly estimated from the data. However, these methods move away from the attractive simple form of the least-squares formulation and require additional data processing steps

Manuscript received by the Editor 20 January 2016; revised manuscript received 9 May 2016; published online 7 September 2016.

¹Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. E-mail: yunyueli@mit.edu; laurent@math.mit.edu.

© 2016 Society of Exploration. All rights reserved.

that are themselves not guaranteed to succeed in complex propagation geometries.

In the second category, additional degrees of freedom are introduced to (attempt to) convexify the waveform inversion in higher dimensions. The extension of velocity with a nonphysical dimension was first introduced by [Symes and Carazzone \(1991\)](#). Recently, [Shen \(2004\)](#), [Symes \(2008\)](#), [Sun and Symes \(2013\)](#), and [Biondi and Almomin \(2014\)](#) show that a velocity extension along the subsurface-offset, plane-wave ray-parameter, or time-lag axis is able to describe the large time shift in wave propagation. [Warner and Guasch \(2014\)](#) extend the model space to include trace-based Wiener filters comparing the modeled and recorded data. [van Leeuwen and Herrmann \(2013\)](#) extend the model space to include the whole wavefield so that the reconstructed wavefield fits the data by design. All these methods rely on an iterative formulation to gradually restrict the extended nonphysical model space to the physical model space, which is often a delicate process without guarantees. In addition, they introduce significant computational cost and memory usage in addition to the already very expensive FWI.

In the third category, the tomographic and migration components in the FWI gradient are separated and enhanced at different stages of the iterations. Based on a nonlinear iterative formulation, [Mora \(1989\)](#) decomposes the wavefields according to their propagation directions. Then, the low- and high-wavenumber components of the velocity model are extracted using correlations of these directional wavefields. Following a similar approach, [Tang et al. \(2013\)](#) propose to enhance the tomographic components at the early iterations and gradually reduce its weights toward convergence. [Alkhalifah \(2015\)](#) further separates the gradient components based on the scattering angle at the imaging point. Although these methods enhance the low-wavenumber components of the FWI gradient, the essential difficulty of ensuring correctness of the tomographic component is still mostly untouched.

The most straightforward way to increase the basin of attraction of the least-squares objective function is to seed it with low-frequency data only, and slowly enlarge the data bandwidth as the descent iterations progress. However, until several years ago, the low-frequency energy less than 5 Hz was often missing due to instrument limitations. More recently, as the importance of the low frequencies became widely recognized by industry, broadband seismic data with a high signal-to-noise ratio between 1.5 and 5 Hz started being acquired at a significantly higher cost than previously. To synthesize some form of low-frequency information, [Wu et al. \(2013\)](#) propose using the envelope function. Similarly, [Hu et al. \(2014\)](#) propose to use the low-frequency information hiding in the difference between data at two adjacent frequencies. Both studies demonstrated that fitting “manufactured” low-frequency information produces somewhat improved low-wavenumber models that enabled closer initialization for the subsequent FWI with the band-limited recordings. However, neither method attempts to approximate the actual low-frequency recordings, and neither choice of nonlinear low-frequency combination is justified by a rationale that they succeed at convexifying the FWI objective function. The situation is worse: it is easy to find examples in which the wave-equation Hessian of an augmented FWI functional with, say, the envelope function and a low-pass filter, still have undesirable large eigenvalues growing as the frequency increases.

The premise of this paper is that the phase-tracking method, proposed in [Li and Demanet \(2015\)](#), is a reasonably effective algorithm

for extrapolating the low-frequency data based on the phases and amplitudes in the observed frequency band. A tracking algorithm is able to separate each seismic record into atomic events, the amplitude and phase functions of which are smooth in space and frequency. With this explicit parameterization, the user can now fit smooth nonoscillatory functions to represent and extrapolate the wave physics to the unrecorded frequency band. Although the resulting extrapolated data can only be expected to accurately reproduce the low-frequency recordings in very controlled situations, they are nevertheless adequate substitutes that appear physically plausible in a broad range of scenarios. Although existing literatures have tried to expand the frequency band of seismic data implicitly via model space extension and/or sparse regularization ([Sun and Symes, 2013](#); [Pan et al., 2016](#)), we are not aware that there is any other attempt at synthesizing low frequencies directly in the data space. The resulting synthesized seismic data are comparable with the low-frequency data that would have been recorded in the field (the mathematical problem of providing tight guarantees concerning extrapolation of smooth functions from the knowledge of their noisy samples has, however, been solved in our companion paper; [Demanet and Townsend, 2016](#)).

In this paper, we test the reliability of the extrapolated low-frequency data on two numerical examples in the constant-density acoustic regime. In both cases, the low frequencies between 1 and 5 Hz are extrapolated from the recordings at 5 Hz and above. We demonstrate that although the extrapolated low frequencies are sometimes far from exact, the low-wavenumber models obtained from the extrapolated low frequencies are often suitable for initialization of FWI at higher frequencies.

METHODS

Review of full-waveform inversion with truncated Gauss-Newton iterations

Conventional FWI is formulated in data space via the minimization of the least-squares mismatch between the modeled seismic record u with the observed seismic record d

$$J(m) = \frac{1}{2} \sum_{r,s,t} (u_s(x_r, t; m) - d_{r,s,t})^2, \quad (1)$$

where m is the slowness of the pressure wave, s indexes the shots, and x_r are the receiver locations. We also write sampling at the receiver locations with the sampling operator S as $u_s(x_r, t; m) = Su_s(x, t; m)$. The modeled wavefield $u_s(x, t; m)$ is the solution of a wave equation (discretized via finite differences in space and time):

$$\left(m^2 \frac{\partial^2}{\partial t^2} - \nabla^2 \right) u_s = f_s, \quad (2)$$

with f_s a source wavelet at location x_s and $\nabla^2 = (\partial^2/\partial x^2) + (\partial^2/\partial y^2) + (\partial^2/\partial z^2)$ the Laplacian operator.

Starting with an initial model $m^{(0)}$, we use a gradient-based iterative scheme to update the model

$$m^{(i+1)} = m^{(i)} - \alpha (\mathbf{J}_r^T \mathbf{J}_r)^{-1} \mathbf{J}_r^T r(m^{(i)}), \quad (3)$$

where $m^{(i)}$ is the model at the i th iteration, $r = Su(m^{(i)}) - d$ is the data residual, \mathbf{J}_r is the Jacobian matrix, and α is the step length for the update.

In practice, the normal matrix $\mathbf{J}_r^T \mathbf{J}_r$ is too large to build explicitly and is often approximated by an identity matrix. In this paper, we choose to precondition this matrix by approximately solving the following system using a few iterations of the conjugate gradient method (Claerbout, 1985; Metivier et al., 2013):

$$\min_{\delta m} \|r(m^{(i)}) - \mathbf{J}_r(m^{(i)})\delta m\|_2^2, \quad (4)$$

where δm is the unknown increment in model. The linear iteration in equation 4 is also known as least-squares reverse time migration, which effectively removes the source signature and produces “true-amplitude” velocity perturbation at convergence. Then, a careful line search is performed to make sure the objective function (equation 1) decreases in the nonlinear iterations (another effective way of preconditioning the normal operator is to use randomized matrix probing, see Demanet et al., 2012).

To help the iterative inversion avoid local minima, we perform frequency-continuation FWI starting from the lowest available frequency with a growing window. Algorithm 1 shows the workflow for FWI with the truncated Gauss-Newton iterations.

Review of phase tracking and frequency extrapolation

In a previous paper (Li and Demanet, 2015), we demonstrated that there exist interesting physical scenarios, in which low-frequency data can be synthesized from the band-limited field recordings using nonlinear signal processing. This processing step is performed before FWI in the frequency domain. To extrapolate the data from the recorded frequency band to lower (and higher)

Algorithm 1. The FWI with truncated Gauss-Newton iterations.

```

Initialize  $m^0$ 
for  $i \in 0 \dots N$  do
     $r^{(i)} \leftarrow Su(m^{(i)}) - d$ 
    if  $\|r_i\| < \epsilon$  then
        Converged with model  $m^{(i)}$ 
    else
        Initialize Gauss-Newton iterations
         $p_0 = r^{(i)}$ ;  $\mathbf{A} = \mathbf{J}_r(m^{(i)})$ ;  $g_0 = \mathbf{A}^T p_0$ 
        First iteration as steepest descent
         $\delta p_0 = \mathbf{A}g_0$ ;  $\alpha_0 \leftarrow \frac{\delta p_0 \cdot \delta p_0}{\delta p_0 \cdot p_0}$ 
         $\delta m_1 = -\alpha_j p_0$ ;  $p_1 = p_0 - \alpha_0 \delta p_0$ 
        for  $j \in 1 \dots n$  do
             $g_j = \mathbf{A}^T p_j$ ;  $\delta p_j = \mathbf{A}g_j$ 
             $\begin{bmatrix} \alpha_j \\ \beta_j \end{bmatrix} \leftarrow \begin{bmatrix} \delta p_{j-1} \cdot \delta p_{j-1} & \delta p_{j-1} \cdot \delta p_j \\ \delta p_j \cdot \delta p_{j-1} & \delta p_j \cdot \delta p_j \end{bmatrix}^{-1} \begin{bmatrix} p_j \cdot \delta p_{j-1} \\ p_j \cdot \delta p_j \end{bmatrix}$ 
             $\delta p_j \leftarrow \alpha_j \delta p_{j-1} + \beta_j \delta p_j$ ;  $g_j \leftarrow \alpha_j g_{j-1} + \beta_j g_j$ 
             $p_{j+1} \leftarrow p_j - \delta p_j$ ;  $\delta m_{j+1} \leftarrow \delta m_j - g_j$ 
        end for
        Line search to find  $\gamma^{(i)}$ 
         $m^{(i+1)} \leftarrow m^{(i)} - \gamma^{(i)} \delta m_{n+1}$ 
    end if
end for

```

frequencies, the phase-tracking method consists of solving the following optimization problem to separate the measured data to its atomic event components: minimize

$$J_{\text{tracking}}(\{a_j, b_j\}) = \frac{1}{2} \|\hat{u}(\omega, x_r) - \hat{d}(\omega, x_r)\|_2^2 + \lambda \sum_j \|\nabla_\omega^2 b_j(\omega, x_r)\|_2^2 + \mu \sum_j \|\nabla_{x_r} b_j(\omega, x_r)\|_2^2 + \gamma \sum_j \|\nabla_{\omega, x_r} a_j(\omega, x_r)\|_2^2, \quad (5)$$

where \hat{d} is the measured data in the frequency domain at a receiver location x_r ; ∇_k and ∇_k^2 , with $k = \omega, x_r$, respectively, denote the first- and second-order partial derivatives; ∇_{ω, x_r} denotes the full gradient; and the predicted data record \hat{u} is modeled by the summation of r individual events:

$$\hat{u}(\omega, x_r) = \sum_{j=1}^r \hat{v}_j = \sum_{j=1}^r \hat{w}(\omega) a_j(\omega, x_r) e^{ib_j(\omega, x_r)}, \quad (6)$$

where wavelet $\hat{w}(\omega)$ is assumed known to a certain level of accuracy. The constants λ , μ , and γ are chosen empirically based on the available frequency, accuracy of the source signature, and the sampling of the offset. These parameters are tested on a couple of shot gathers, so that a majority of the events are consistently picked.

The optimization problem for event tracking here is reminiscent of FWI with high-frequency data, hence shares a similar level of nonconvexity. Yet, by posing it as a data processing problem, the nonconvexity can be empirically overcome with an explicit initialization scheme using multiple signal classification (MUSIC), coupled with a careful trust-region “expansion and refinement” scheme to track the smooth phase and amplitude. We refer the reader to the detailed algorithm in the previous paper (Li and Demanet, 2015).

Having obtained the individual events, we make explicit assumptions about their phase and amplitude functions to extrapolate outside of the recorded frequency band. Namely, we assume that the earth is nondispersive, i.e., the phase is affine (constant + linear) in frequency, and that the amplitude is to a good approximation constant in frequency — though both are variable in x , of course. A least-squares fit is then performed to find the best-constant approximations $a_j(\omega, x_r) \simeq \alpha_j(x_r)$, and the best-affine approximations $b_j(\omega, x_r) \simeq \omega \beta_j(x_r) + \phi_j(x_r)$, from values of ω within the useful frequency band. These phase and amplitude approximations can be evaluated at values of ω outside this band, to yield synthetic flat-spectrum atomic events of the form

$$\hat{v}_j^e(\omega, x_r) = \alpha_j(x_r) e^{i(\omega \beta_j(x_r) + \phi_j(x_r))}. \quad (7)$$

These synthetic events are multiplied by a band-limited wavelet, and summed up, to create a synthetic data set.

The effectiveness of this method for event identification is limited by many factors, chiefly the resolution of the MUSIC algorithm and the signal-to-noise ratio of the data. The algorithm often tracks the strong events and treats the weak events as noise. Moreover, the amplitudes of the events are less predictable than the phases, due to propagation and interfering effects. Therefore, the extrapolated data record is inexact, typically with higher fidelity in phase than in amplitude. In the following section, we test the reliability of

the extrapolated low frequencies by initializing the frequency continuation of FWI. Our goal is to bring reliable low-wavenumber information in the model by fitting the phase of the extrapolated data and to help enlarge the basin of attraction for FWI when the low frequencies are missing from data.

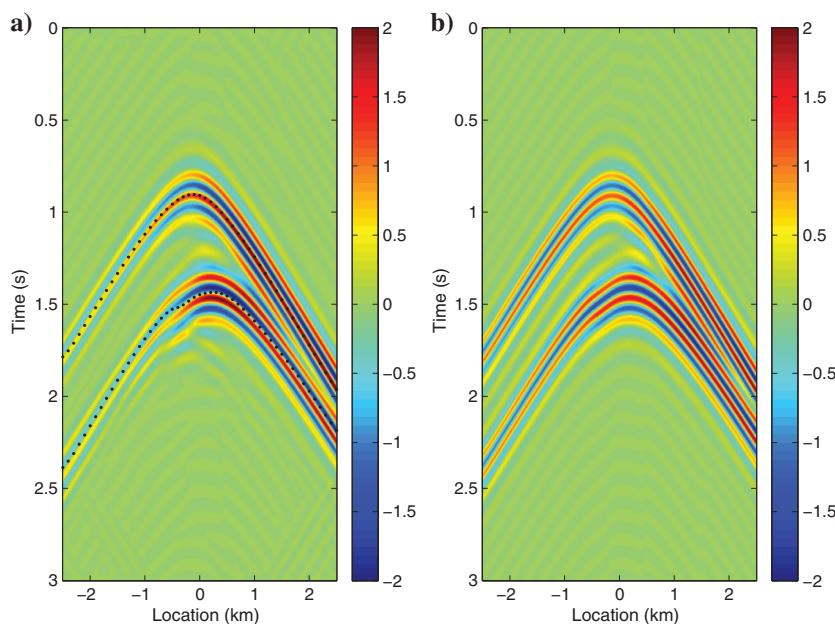


Figure 1. Comparison between (a) the recorded shot profile and (b) the inverted shot profile on the Camembert model within the recorded bandwidth (5–15 Hz). The inverted shot profile recovers the phase and the amplitude of the two strong reflection events, whose estimated arrival times are denoted by the dotted line in panel (a). The inversion does not recover the weak multiple reflection events behind the second reflection event. Energy from these missing events will show up as phase and amplitude errors at lower frequency when events begin interfering with each other.

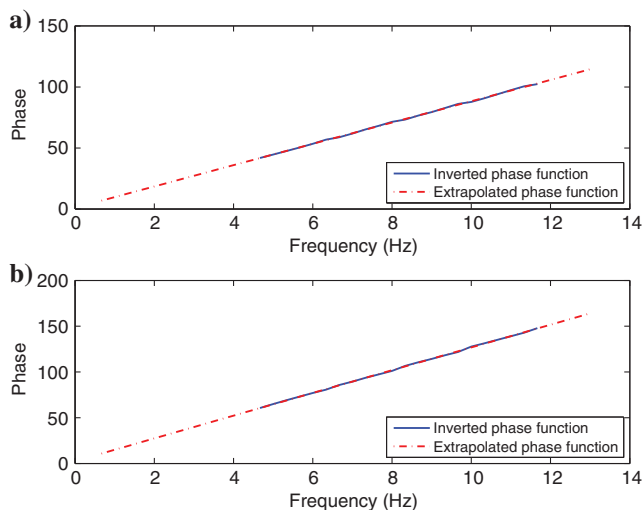


Figure 2. Comparison between the inverted phase function (solid line) and the extrapolated phase function (dashed line) for (a) top and (b) bottom reflection events. The extrapolated phase function represents a reasonable approximation to the inverted phase function.

NUMERICAL EXAMPLES

In this section, we demonstrate the reliability of the extrapolated low frequencies on two synthetic examples. The first model is a wide-aperture, one-sided version of the classic Camembert model (Tarantola, 1984; Gauthier et al., 1986), in which the accuracy of frequency extrapolation can be analyzed carefully on the data record. The second model is the Marmousi model, in which the frequency extrapolation is tested in a more complex and geologically relevant environment. In both examples, we compare three cases while keeping the initial models fixed:

- 1) In the control case, frequencies from 1 to 15 Hz are used in the frequency continuation of FWI.
- 2) In the extrapolated FWI (EFWI) case, we first extrapolate the data between 5 and 15 Hz to the frequency band between 1 and 5 Hz. The extrapolated data are used to build the low-wavenumber model to further initialize the frequency continuation starting at 5 Hz.
- 3) In the missing low-frequency case, frequency continuation of FWI starts from the lowest frequency at 5 Hz.

For the extrapolated case, data are processed prior to the FWI iterations. For both test cases, we set $\lambda = 1.0$, $\mu = 10.0$, and $\gamma = 0.002$. With our sequential implementation, phase tracking and frequency extrapolation costs approximately 5% of a single FWI iteration. We truncate the Gauss-Newton inversion in Algorithm 1 at three iterations for each nonlinear step, when it shows a good balance between convergence and computational cost.

Synthetic example: Camembert model

In this example, we model the synthetic seismic records on the classic Camembert model. We work with the reflection setting in which 41 sources and 401 receivers are evenly spaced on the surface. The maximum offset between the sources and receivers is 4 km. A circular low-velocity anomaly ($v = 1700$ m/s) is placed in a constant velocity background ($v = 2000$ m/s).

Figure 1a shows one record between 5 and 15 Hz. The tracking algorithm picked up the two strong reflection events from the top and the bottom of the low-velocity anomaly. The dotted black line overlaid in Figure 1 denotes the inverted arrival time for each event. The inverted two reflection events are shown in Figure 1b. The tracking algorithm does not perfectly reproduce the input recording, leaving behind the weaker multiple reflection events trailing the second reflection event. Energy from these missing events will appear as phase and amplitude errors at lower frequency when events interfere with each other.

Following the extrapolation strategy in equation 7, we extend each event from the recorded frequency band to higher and lower frequency bands. Figure 2 shows the comparison between inverted phase function (solid line) and the extrapolated phase function

(dashed line). Although the inverted phase function is not perfectly linear with respect to frequency for each event, the extrapolated phase function represents a reasonable approximation of the inverted phase function.

Figure 3 compares the modeled low-frequency shot profile (Figure 3a) with the extrapolated low-frequency shot profile (Figure 3b). The two shot profiles differ from each other in amplitude and phase. Discrepancies in amplitude are caused by the crude “constant amplitude” assumption, whereas discrepancies in phase are caused by the interference of the unmodeled weak events. However, the extrapolated low-frequency shot profile is a reasonable approximation of the modeled one for FWI.

To test the accuracy and reliability of the extrapolated record, we initialize the frequency continuation of FWI using the extrapolated data. Starting from a constant-velocity model ($v = 2000$ m/s), Figure 4b shows the inversion result using the extrapolated low frequency (1–5 Hz) data in comparison with the inversion result using the modeled low-frequency data (Figure 4a). Although the

positive velocity side lobes are stronger in the extrapolated result, both inversion tests resolve the low-wavenumber structure of the velocity model.

We continue the frequency continuation of FWI starting from the initial models in Figure 4 and obtain the final-inversion results in Figure 5. Figure 5a shows the inversion result in the control case when all the frequencies (1–15 Hz) are used in the frequency continuation. The inversion successfully delineates the boundary of the low-velocity anomaly. Figure 5b shows the inversion result when low-frequency (1–5 Hz) energy is missing from the inversion. Thanks to the low-wavenumber information extracted from the low-frequency extrapolated data, FWI without low frequencies has converged to a meaningful velocity model. In comparison, FWI starting from a constant background velocity model and 5 Hz data cannot resolve the low-wavenumber velocity structure, yielding mispositioned boundaries of the velocity anomaly, as shown in Figure 5c.

One important issue to notice is that the velocity value of the anomaly has not been fully recovered in any of the three cases. This is primarily due to the limited number of iterations that only enable

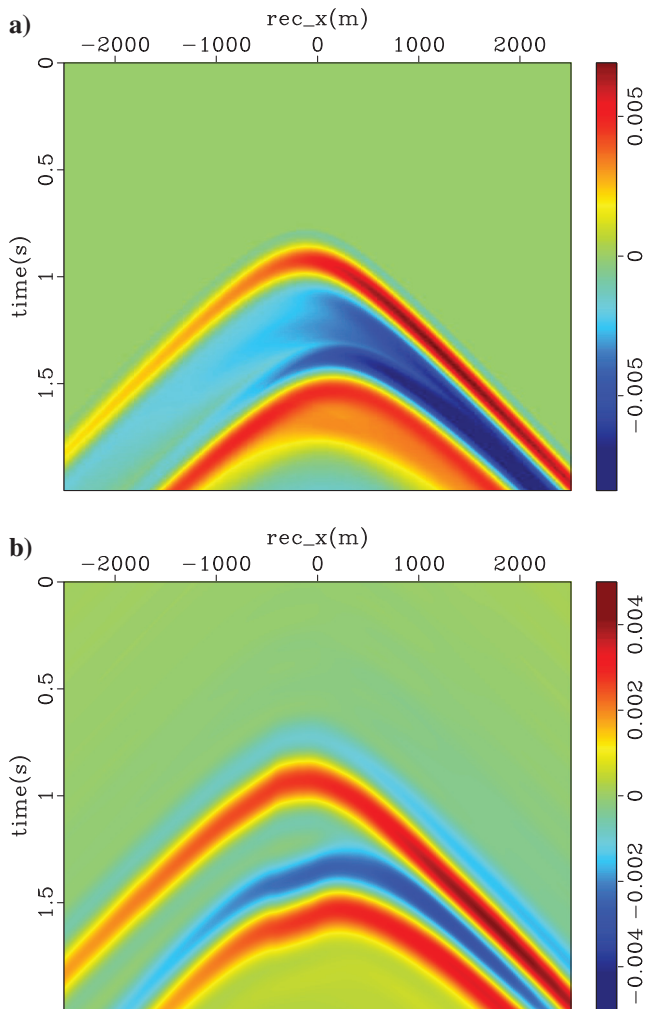


Figure 3. Comparison between (a) the modeled low-frequency (1–5 Hz) shot profile and (b) the extrapolated low-frequency shot profile on the Camembert model. Discrepancies in amplitude are caused by the crude “constant amplitude” assumption. Discrepancies in phase are caused by the interfering of the unmodeled weak events.

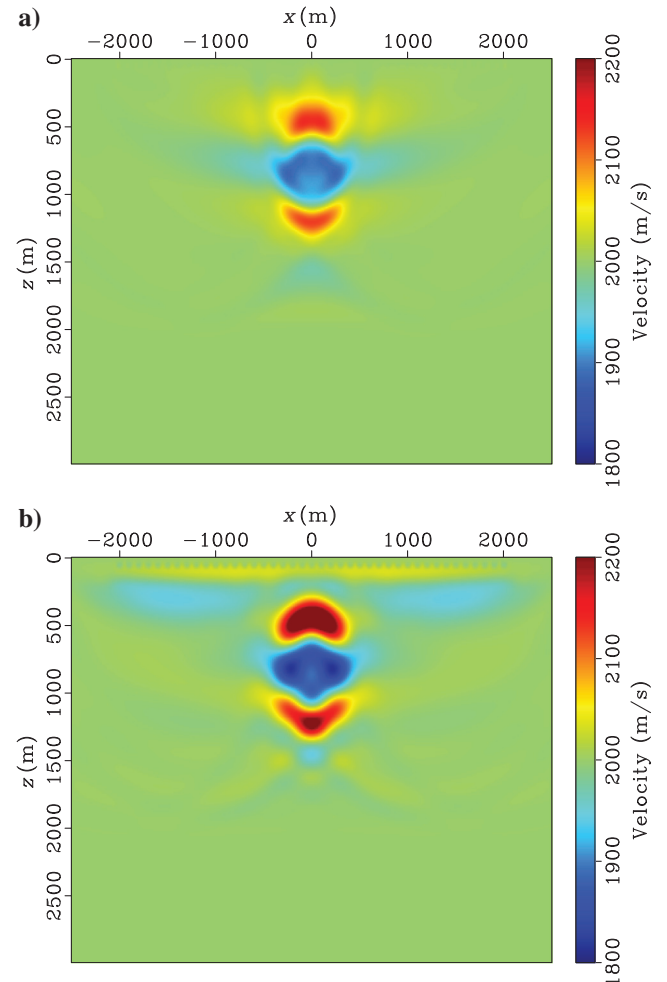


Figure 4. Inversion results from FWI with (a) the modeled low frequency (1–5 Hz) shot profile and (b) the extrapolated low-frequency shot profile. These inversion tests resolve similar low-wavenumber structure in the velocity model.

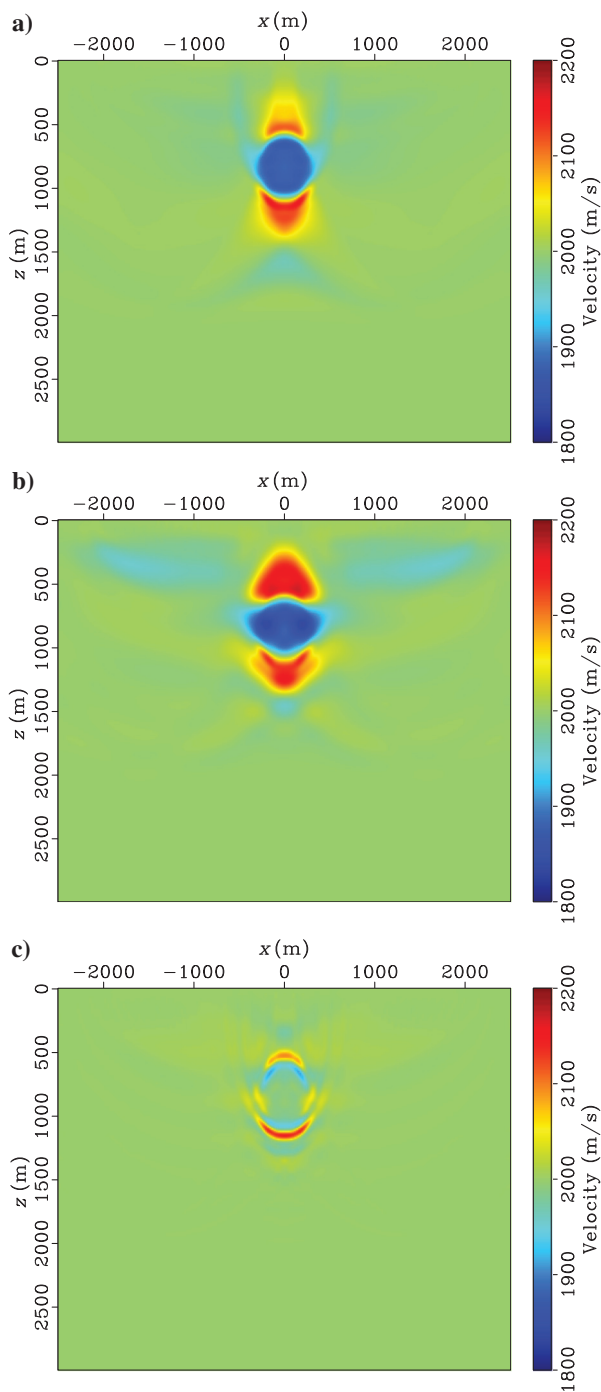


Figure 5. Comparison between the inverted model from FWI after a full-bandwidth continuation. (a) Resulting model from the control case (frequency continuation from 1 to 15 Hz). (b) Resulting model from the extrapolated case (initialization using extrapolated low frequencies [1–5 Hz] and frequency continuation with recorded data from 5 to 15 Hz). (c) Resulting model from the missing low-frequency case (frequency continuation from 5 to 15 Hz). When all the frequencies (1–15 Hz) are used in the frequency continuation, FWI successfully delineates the boundary of the low-velocity anomaly. When low frequencies (1–5 Hz) are missing from the data, FWI converges to a velocity model with meaningful low-wavenumber components with the aid from the extrapolated low-frequency data. Otherwise, FWI cannot resolve the low-wavenumber velocity structure, yielding mispositioned velocity boundaries.

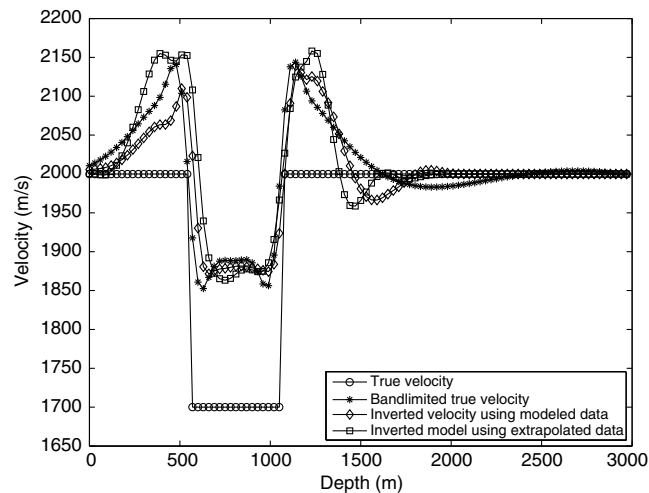


Figure 6. Comparison of the pseudovelocity logs at $x=0$ m between the inversion results and the true velocity model. The FWI cannot fully resolve the perturbation of the anomaly due to the limited bandwidth in the data. Both inversion results match well with the band-limited (between 1 and 15 Hz) true velocity model.

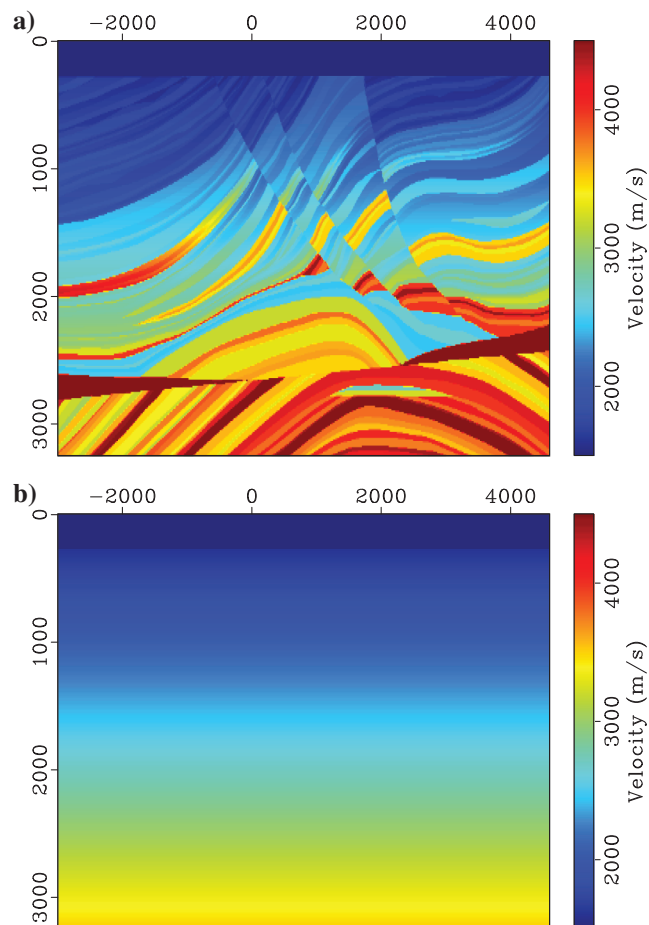


Figure 7. (a) Marmousi model and (b) the starting model for FWI. The starting model is a 1.5D linearly increasing velocity profile from 1500 m/s at the water bottom to 3500 m/s at 3.2 km. The initial model is far from the true especially in the deeper section.

the updates within the available frequency. In the control case, the inversion will resolve the exact model if we let FWI iterate to convergence — an unrealistic proposition when the data become noisy. Due to the inaccuracies in the extrapolated data, we stop the inversion after 50 iterations to avoid overfitting the unreliable fine scale features in the data. We compare the pseudologs at $x = 0$ between the true and the inverted velocities in Figure 6. The true velocity anomaly is -300 m/s with respect to the background, whereas the inversion results only recover half the perturbation with side lobes of similar opposite perturbations. We find that these inverted results match well with a band-limited version of the true velocity model, in which frequencies less than 1 Hz and more than 15 Hz are not present in the model.

Further tests reveal that the lowest frequency that is needed to recover the circular shape of the Camembert model is 1 Hz (acquisition and iterative scheme fixed). The FWI starting from 2 Hz will only resolve the boundary of the Camembert model, which might lead to erroneous interpretations. For this simple model, our extrapolation algorithm continues to yield stable results, as we increase the lowest available frequency in the recorded data. With the highest available frequency fixed at 15 Hz, the lowest frequency for a reliable extrapolation can be as high as 8 Hz.

Synthetic example: Marmousi model

In this example, we test the reliability of the extrapolated low frequency on the Marmousi model (Figure 7a). We restrict FWI to reflection events only by limiting the maximum offset to 500 m in the inversion. The starting model for FWI is a 1.5 D linearly increasing velocity profile (Figure 7b). It is an extremely challenging task for FWI to recover large low-wavenumber discrepancies between the initial and the true models, especially in the deeper section (more than 2 km).

Figure 8 shows the comparison between the modeled band-limited (5–15 Hz) shot profile (Figure 8a) with the inverted shot profile by event tracking (Figure 8b). The tracking algorithm is set up to identify up to 10 strongest events in each shot record. On this particular shot profile, nine events are identified. The tracking algorithm recovers the input record very well. Figure 9 shows the comparison between the modeled and the extrapolated low-frequency (1–5 Hz) shot profile, amplitude-normalized. Despite the amplitude and slight phase discrepancy, the extrapolated shot profile approximates the modeled shot profile sufficiently well in the low-frequency band for the purpose of initializing FWI.

The reliability of the extrapolated low frequencies (1–5 Hz) is tested with FWI at these low

frequencies. Figure 10 shows the inverted velocity model using modeled data (Figure 10a) and using extrapolated data (Figure 10b). Although not as detailed as Figure 10a, the velocity model inverted using the extrapolated data correctly captures the very low-wavenumber component of the true model. These models are used to initialize FWI with data at higher frequencies.

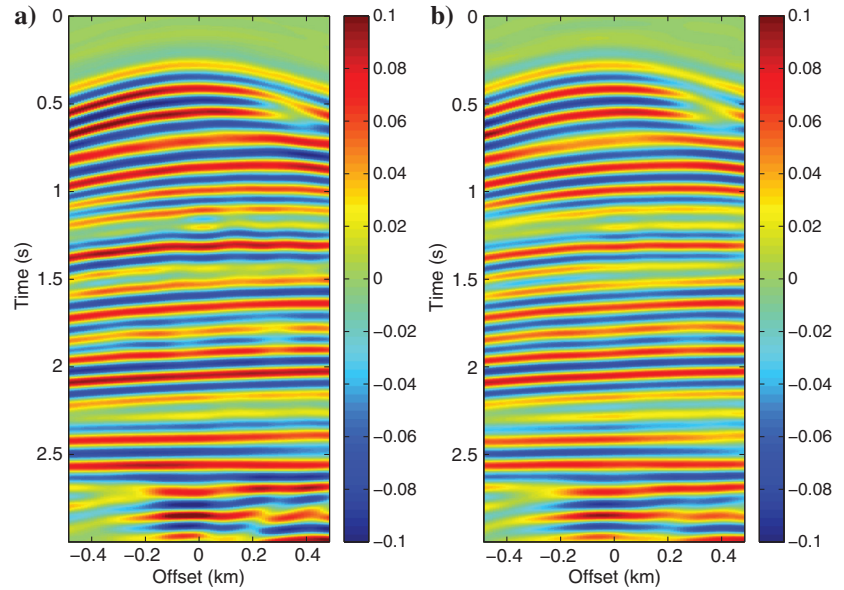


Figure 8. Comparison between (a) the modeled band-limited (5–15 Hz) shot record and (b) the inverted shot record. The tracking algorithm identifies nine individual events and recovers the modeled data very well.

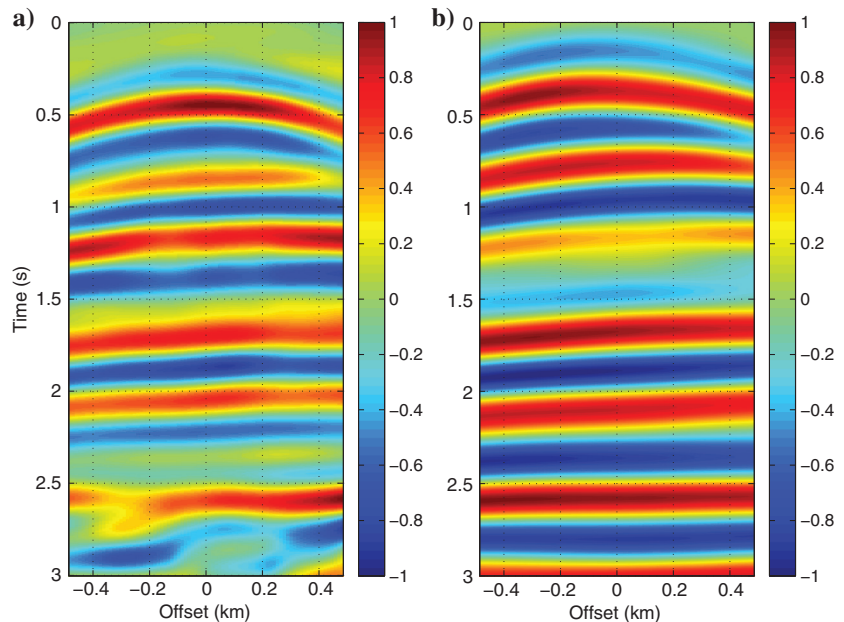


Figure 9. Comparison between (a) the modeled low-frequency (1–5 Hz) shot record and (b) the extrapolated low frequency (1–5 Hz) shot record. The color bar denotes the normalized amplitude. Despite the imperfect reconstruction of phase and amplitude, the extrapolated low frequencies are adequate for the purpose of initializing FWI.

Figure 11 compares the final-inverted results after a full-bandwidth FWI continuation. In the shallow region (less than 2 km), velocity models resolved in the control and the extrapolated cases are very similar with accurately imaged fine layers and normal faults. Both models have trouble resolving a high-resolution and accurate velocity model in the deep region because reflections from the dipping reflectors and the anticline structure have not been sufficiently recorded due to the limited offset. In comparison, FWI starting at 5 Hz yields little meaningful information about the subsurface. The inversion failed to update the low-wavenumber structure of the velocity model and placed reflectors at wrong positions. Compared with the true model, relative velocity errors in Figure 11a–11c are 14.45%, 15.33%, and 18.39%, respectively.

Figure 12 compares pseudovelocity logs at three surface locations. Velocity models in the control and extrapolated cases recover the true velocity model very well above more than 1 km. Quality of the inverted model degrades with depth. However, both velocity models capture the low-wavenumber components of the velocity model. The maximum updates in the deeper section are as high

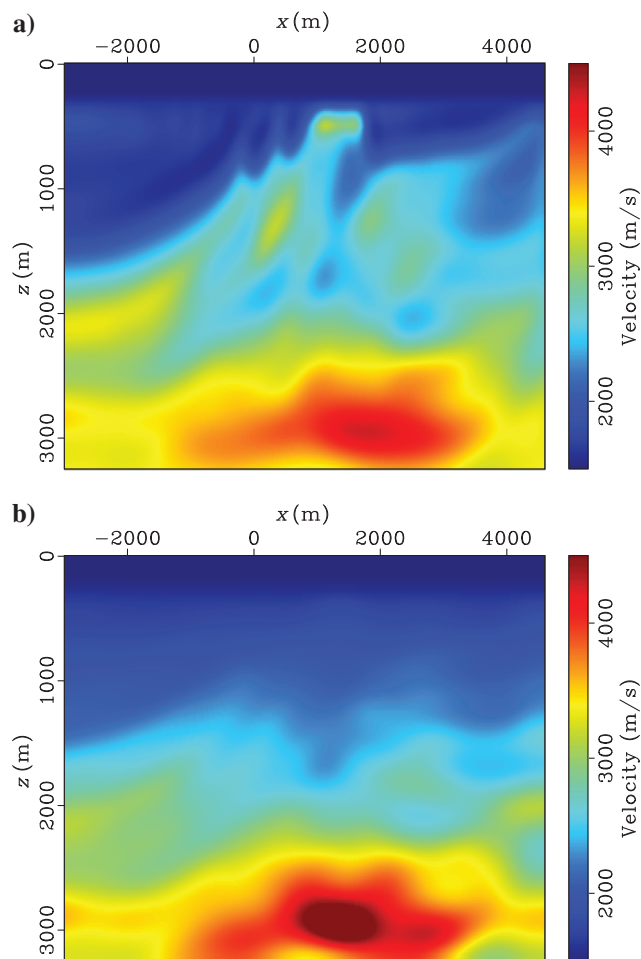


Figure 10. Comparison between (a) the inverted model after FWI using modeled low-frequency (1–5 Hz) data and (b) using extrapolated low-frequency data (1–5 Hz). Both models capture the low-wavenumber structure of the Marmousi model, although the inverted model using modeled data contains more details in the shallow part.

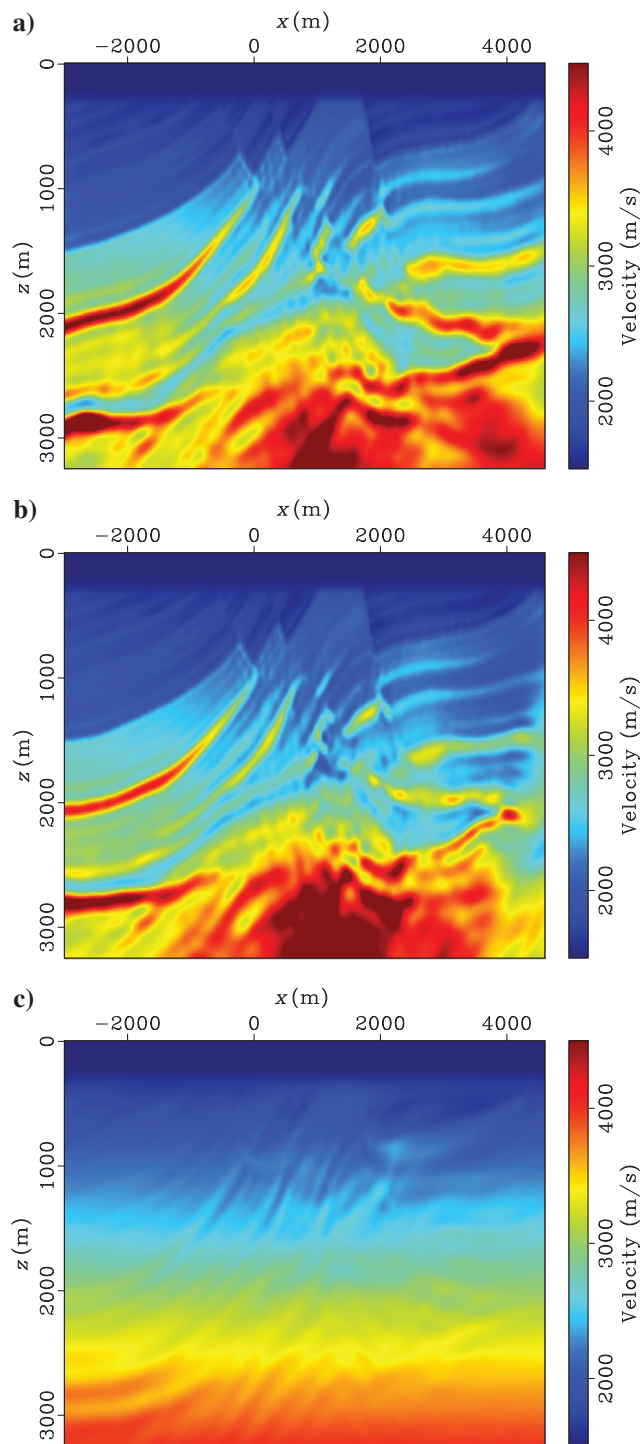


Figure 11. Comparison between the inverted model from FWI after a full-bandwidth continuation. (a) Resulting model from the control case (frequency continuation from 1 to 15 Hz). (b) Resulting model from the extrapolated case (initialization using extrapolated low frequencies [1–5 Hz] and frequency continuation with recorded data from 5 to 15 Hz). (c) Resulting model from the missing low-frequency case (frequency continuation from 5 to 15 Hz). A better inverted model can be obtained in the control case, if we iterate to convergence at the lowest frequencies. However, we limit the number of iterations in the control case to ensure a fair comparison with the extrapolated case.

as 1000 m/s. The huge velocity error prevents FWI starting at 5 Hz from converging to the true model.

DISCUSSION

In the Camembert example, we have carefully examined the accuracy of the phase-tracking method for low-frequency extrapolation. There are two main reasons that the amplitude and the phase of the extrapolated data are inexact. First, the tracking algorithm determines the number of individual events as an initialization step. This number may decrease (for event truncation), but it may not increase (for event bifurcation) as the tracking expands. These untracked events are the main contributors to the errors in the extrapolated phase function. An aggregation method, with event fragments tracked in subsets of traces and merged into actual composite events, may help to improve the accuracy of the event tracking and phase extrapolation. Second, compared with the phase extrapolation based on nonattenuative physics, the amplitude extrapolation is less constrained by physical principles. Still, it has been demonstrated that FWI is robust against inaccuracies in modeling the amplitudes of seismic data (Pratt and Shipp, 1999; Shin and Cha, 2008). Consequently, the extrapolated data provide adequate phase information for initializing FWI, although they might not be suitable for absolute impedance inversion and amplitude-based rock-property interpretation.

Due to the inaccuracy in their phase and amplitude, we do not allow FWI to fully fit the data at the extrapolated low frequencies. This limits the resolution of the inverted FWI model. With hundreds

of more iterations in the control case, FWI starting at 1 Hz can reduce the data residual to 1% at each frequency band. Hence, the velocity model can be fully resolved (with all wavenumber components) because of the availability of low frequency and long offsets. To the contrary, overfitting the extrapolated data at low frequencies would lead the inversion to undesired local minima and spurious models. The FWI with a limited number of iterations resolves a good estimate of the velocity model within the increased available frequency band, but it cannot perfectly resolve the model at all wavenumbers due to the slow convergence and the potential local minima introduced by the inaccuracies in the extrapolated data.

In our numerical examples, the extrapolated low frequencies are used only to initialize FWI to obtain a low-wavenumber model. As soon as the frequency continuation moves to the recorded frequency band, the extrapolated low frequencies are abandoned. This leaves the low-wavenumber components of the model space unconstrained in later FWI iterations. A proper combination of the extrapolated and the recorded data needs to be studied to ensure a fully constrained inversion for velocity in the whole wavenumber band.

CONCLUSION

To mitigate the nonconvexity of FWI, we propose to start the frequency continuation using the extrapolated low-frequency data. The extrapolation is only feasible after decomposing the seismic records into individual atomic events via phase tracking for each isolated arrival. Numerical examples demonstrate that FWI is surprisingly tolerant to inaccuracies in the amplitude and phase of the

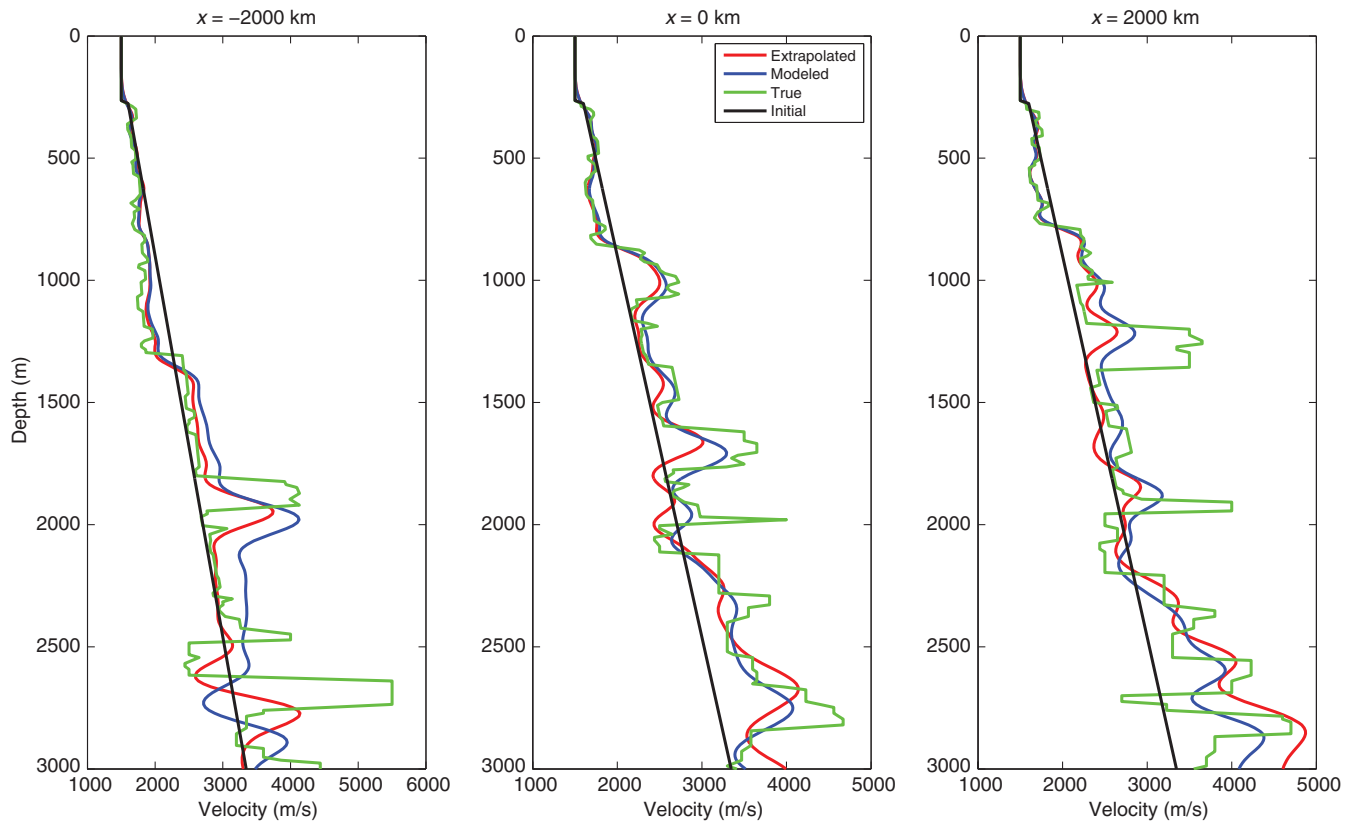


Figure 12. Pseudovelocity logs at three surface locations. The black line denotes the initial model. The green line denotes the true model. The black, green, red, and blue lines denote the pseudolog from the initial, true, control, and extrapolated model, respectively.

extrapolated events. Initializing with the extrapolated low frequencies mitigates the severe nonconvexity that FWI suffers from when only high-frequency data are available. By explicitly obtaining the phase and amplitude of each individual event, our method shares an important feature with traveltome tomography: its ability to extract kinematic information from high frequencies only. We call the method EFWI for short.

ACKNOWLEDGMENTS

This project was funded by Total S.A. L. Demanet is grateful to AFOSR, ONR, and NSF for providing funding. Y. E. Li acknowledges the sponsors of the Earth Resources Laboratory at MIT for providing funding.

REFERENCES

- Alkhalifah, T., 2015, Scattering-angle based filtering of the waveform inversion gradients: *Geophysical Journal International*, **200**, 363–373, doi: [10.1093/gji/ggu379](https://doi.org/10.1093/gji/ggu379).
- Baek, H., H. Calandra, and L. Demanet, 2014, Velocity estimation via registration-guided least-squares inversion: *Geophysics*, **79**, no. 2, R79–R89, doi: [10.1190/geo2013-0146.1](https://doi.org/10.1190/geo2013-0146.1).
- Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: *Geophysics*, **79**, no. 3, WA129–WA140, doi: [10.1190/geo2013-0340.1](https://doi.org/10.1190/geo2013-0340.1).
- Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: *Geophysics*, **69**, 1283–1298, doi: [10.1190/1.1801945](https://doi.org/10.1190/1.1801945).
- Claerbout, J. F., 1985, *Fundamentals of geophysical data processing*: Pennwell Books.
- Demanet, L., P.-D. Letourneau, N. Boumal, H. Calandra, J. Chiu, and S. Snelson, 2012, Matrix probing: A randomized preconditioner for the wave-equation Hessian: *Applied and Computational Harmonic Analysis*, **32**, 155–168, doi: [10.1016/j.acha.2011.03.006](https://doi.org/10.1016/j.acha.2011.03.006).
- Demanet, L., and A. Townsend, 2016, Stable extrapolation of analytic functions, [10.1016/j.acha.2011.03.006](https://doi.org/10.1016/j.acha.2011.03.006), accessed 1 June 2016.
- Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: *Geophysics*, **51**, 1387–1403, doi: [10.1190/1.1442188](https://doi.org/10.1190/1.1442188).
- Hu, W., 2014, FWI without low frequency data-beat tone inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1116–1120.
- Li, Y. E., and L. Demanet, 2015, Phase and amplitude tracking for seismic event separation: *Geophysics*, **80**, no. 6, D363–D383, doi: [10.1190/geo2014-0535.1](https://doi.org/10.1190/geo2014-0535.1).
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltome 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 traveltome 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).
- Metivier, L., R. Brossier, J. Virieux, and S. Operto, 2013, Full waveform inversion and the truncated Newton method: *SIAM Journal on Scientific Computing*, **35**, B401–B437, doi: [10.1137/120877854](https://doi.org/10.1137/120877854).
- Mora, P., 1989, *Inversion = migration + tomography*: Springer.
- Pan, W., K. Innanen, G. Margrave, and S. Keating, 2016, Mitigating cycle-skipping for FWI by bandlimited impedance inversion and POCS: Presented at the CSEG GeoConvention.
- Pratt, R. G., and R. M. Shipp, 1999, Seismic waveform inversion in the frequency domain. Part 2: Fault delineation in sediments using crosshole data: *Geophysics*, **64**, 902–914, doi: [10.1190/1.1444598](https://doi.org/10.1190/1.1444598).
- Shen, P., 2004, *Wave equation migration velocity analysis by differential semblance optimization*: Ph.D. thesis, Rice University.
- Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: *Geophysical Journal International*, **173**, 922–931, doi: [10.1111/j.1365-246X.2008.03768.x](https://doi.org/10.1111/j.1365-246X.2008.03768.x).
- Sun, D., and W. W. Symes, 2013, Waveform inversion via nonlinear differential semblance optimization: Presented at the 75th EAGE Conference and Exhibition-Workshops.
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical prospecting*, **56**, 765–790, doi: [10.1111/j.1365-2478.2008.00698.x](https://doi.org/10.1111/j.1365-2478.2008.00698.x).
- Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: *Geophysics*, **56**, 654–663, doi: [10.1190/1.1443082](https://doi.org/10.1190/1.1443082).
- Tang, Y., S. Lee, A. Baumstein, and D. Hinkley, 2013, Tomographically enhanced full wavefield inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 1037–1041.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266, doi: [10.1190/1.1441754](https://doi.org/10.1190/1.1441754).
- 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).
- 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., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093.
- Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: *Geophysics*, **73**, no. 5, VE5–VE11, doi: [10.1190/1.2969907](https://doi.org/10.1190/1.2969907).
- Wu, R.-S., J. Luo, and B. Wu, 2013, Ultra-low-frequency information in seismic data and envelope inversion: 83rd Annual International Meeting, SEG, Expanded Abstracts, 3078–3082.