GEOPHYSICS®

Tackling cycle-skipping in full-waveform inversion with intermediate data

Journal:	*Geophysics*
Manuscript ID	GEO-2018-0096.R3
Manuscript Type:	Technical Paper
Keywords:	full-waveform inversion, tomography, nonlinear
Area of Expertise:	Seismic Inversion

Tackling cycle-skipping in full-waveform inversion with intermediate data

Gang Yao[1,2], Nuno V. da Silva[3], Michael Warner[3], Di Wu[1], Chenhao Yang[2]

Right Running Head: intermediate data for cycle-skipping

[1] *China University of Petroleum (Beijing), The Unconventional Natural Gas Institute, State Key Laboratory of Petroleum Resources and Prospecting, Beijing, China. E-mail: yaogang@cup.edu.cn; wudi@cup.edu.cn.*

[2] *Rice University, Department of Earth Science, Houston, Texas. E-mail: chenhao.yang@rice.edu.*

[3] *Imperial College London, Department of Earth Science and Engineering, London, UK. E-mail: n.vieira-da-silva@imperial.ac.uk; m.warner@imperial.ac.uk.*

ABSTRACT

Full-waveform inversion (FWI) is a promising technique for recovering the Earth models for both exploration geophysics and global seismology. FWI is generally formulated as the minimization of an objective function, defined as the L2-norm of the data residuals. The non-convex nature of this objective function is one of the main obstacles for the successful application of FWI. A key manifestation of this non-convexity is cycle-skipping, which happens if the predicted data is more than half a cycle away from the recorded data. We introduce the concept of intermediate data for tackling cycle-skipping. This intermediate data set is created to sit between predicted and recorded data, and it is less than half a cycle away from the predicted data. Inverting the intermediate data rather than the cycle-skipped recorded data can then circumvent cycle-skipping. We applied this concept to invert cycle-skipped first arrivals. Firstly, we picked up the first breaks of the predicted data and the recorded data. Secondly, we linearly scaled down the time difference between the two first breaks of each shot into a series of time shifts, the maximum of which is less than half a cycle, for each trace in this shot. Thirdly, we moved the predicted data with the corresponding time shifts to create the intermediate data. Finally, we inverted the intermediate data rather than the recorded data. Since the intermediate data is not cycle-skipped and contains the traveltime information of the recorded data, FWI with intermediate data updates the background velocity model in the correct direction. Thus, it produces a background velocity model accurate enough for carrying out conventional FWI to rebuild the intermediate- and short-wavelength components of the velocity model. Our numerical examples using synthetic data validate the intermediate-data concept for tackling cycle-skipping and demonstrate its effectiveness for the application to first arrivals.

INTRODUCTION

Full-waveform inversion (FWI) aims to recover the seismic properties of the Earth, including P-wave velocity, anisotropic parameters, and density. FWI is formulated as the optimization of an objective function defined by the misfit between recorded and predicted data, constrained by a wave equation (Tarantola, 1984). Compared to seismic imaging techniques that utilize the adjoint operator instead of the inverse operator, e.g., seismic migration (Claerbout, 1992), FWI, in which the optimization process acts as applying inverse operators to the seismic data, has the capability to recover quantitatively accurate models. Furthermore, the use of full wave equations allows for more accurate simulations of wavefield propagation, giving FWI the advantage over other techniques that employ simplified wave equations, e.g., migration with one-way wave equations and Kirchhoff migration with ray-tracing. For these two reasons, FWI is becoming one of the most promising techniques for exploration geophysics (Virieux and Operto, 2009; Warner et al., 2013; Debens et al., 2015; da Silva et al., 2016; da Silva et al., 2018; Morgan et al., 2016; Yao et al., 2018b) as well as global seismology (Zhu et al., 2012; Chen et al., 2015; Tao et al., 2017).

However, several aspects of FWI hinder its advance. Firstly, the high computational cost is a key obstacle to FWI for 3D field data applications because it requires solving wave equations repeatedly. This difficulty can be alleviated using more efficient inversion strategies, e.g., random shot selection (Herrmann et al., 2013; Warner et al., 2013), source encoding (Krebs et al., 2009), preconditioning (Baumstein et al., 2009; Burgess and Warner, 2015; Hu, 2016; Biondi et al., 2017; Yao et al., 2017), faster numerical-modeling codes (Zhang and Yao, 2012; Wang et al., 2014; Yao et al., 2016; Wang et al., 2017; Yao et al., 2018a), and faster hardware (Brown, 2007; Nemeth et al., 2008; Weiss and Shragge, 2013).

Secondly, because of the ill-posed nature of geophysical data inversion, FWI may produce a model that is very different from the true model but fits the recorded data due to incomplete acquisition coverage and the dimensionality of the model space including estimating multi-parameters (e.g., Baumstein, 2014; da Silva et al., 2016). The conditioning of the inverse problem can be ameliorated by developing better acquisition equipment to record more complete data (e.g., Shen et al., 2017), and incorporating suitable mathematical constraints (e.g., Tikhonov and Arsenin, 1977; Fehmers and Höcker, 2003; Esser et al., 2016; Trinh et al., 2017).

Thirdly, the objective function used in conventional FWI (square of the L2-norm of the residuals) is non-convex. Hence, convergence towards the global minimum is not guaranteed when using local gradient-based methods. In such cases, the iterative inversion often converges towards a local minimum if the initial model sits in a basin of attraction of the objective function that is away from the one containing the global minimum. As a result, the recovered model can be very different from the true model, and it is likely to be even worse than the initial model. Generally, seismic data is more nonlinear for long-wavelength background velocity than for short-wavelength impedance contrast (Jannane et al., 1989). Consequently, it is much easier to invert impedance contrast (reflectivity) with least-squares reverse-time migration (LSRTM) (e.g. Dai et al., 2012; Yao and Wu, 2015; Yao and Jakubowicz, 2016) than to recover the velocity model with FWI.

One key manifestation of this non-convexity is cycle-skipping. It happens if the events in the predicted data are more than half a cycle away from the corresponding events in the recorded data. As a result, cycle-skipping generally leads FWI to converge to a local minimum. This results in an incorrect estimation of the model parameters.

This problem has been addressed in several ways. One is to build an initial model that is accurate enough to produce the predicted data less than half a cycle away from the recorded data. Although this approach can guarantee, at least to some extent, that the successive estimates of the model parameters are carried out within the basin of attraction of the global minimum, it is usually difficult to generate such an accurate initial model.

The multi-scale strategy (Bunks et al., 1995) is perhaps the most widely and successfully used approach for mitigating cycle skipping. The lower the frequency, the wider the half cycle. Consequently, if the inversion starts from the lowest frequency in the recorded data and then the frequency is increased sequentially, the possibility of cycle-skipping to occur is substantially reduced. This is the main reason why techniques aiming to create low-frequency signals by using mathematical extrapolation (Shin and Ho Cha, 2008; Shin and Ho Cha, 2009; Li and Demanet, 2016), or acquiring low-frequency signals by modifying conventional acquisition (Kalinicheva et al., 2017), or inventing new acquisition systems (Baeten et al., 2013; Dellinger et al., 2016), have been investigated.

Alternative objective functions have been proposed to resolve the cycle-skipping issue. The fundamental mechanism of this type of method is to create a much broader convex region around the global minimum than that of conventional FWI, the objective function of which is the square of the L2-norm of the data residual (Tarantola, 1984). The envelope of an oscillating seismic trace has a much larger period than the original seismic signal. Thus, an objective function formulated by the difference of envelopes between the predicted data and the recorded data has a very broad convex region around the global minimum (Wu et al., 2014; Liu and Zhang, 2017). Consequently, the inversion is more likely to converge towards the global minimum even with a poorer starting model. Similarly, the superposition of the two single-frequency signals, the frequencies of which are close, produces a low-frequency envelope as well; therefore, the objective function with this envelope has a wide convex region around the global minimum. This method is referred to as the beat-tone method (Hu, 2014).

The alternative objective functions can also be formulated in the extended domain to solve the cycle-skipping issue. For example, by weighting the crosscorrelation function or designing a Wiener filter between the predicted traces and the corresponding recorded traces, the objective function can generate a large convex region around the global minimum, hence improving the convergence properties of the inversion algorithm (van Leeuwen and Mulder, 2010; Luo and Sava, 2011; Warner and Guasch, 2016; Zhu and Fomel, 2016). A similar mechanism has been applied to the image-domain inversion referred to as wave-equation migration velocity analysis (WEMVA). In this method, the objective function is formed by weighting the sub-surface offset image gather (Symes, 2008; Zhang and Biondi, 2013; Fu and Symes, 2017). There are also some other objective functions by domain extension, for instance, wavefield reconstruction inversion (WRI) (van Leeuwen and Herrmann, 2013; da Silva and Yao, 2018), matched source extension (Huang et al., 2017), tomographic full-waveform inversion (Biondi and Almomin, 2014), and differential semblance (Symes and Carazzone 1991; Plessix et al., 2000; Mulder and ten Kroode, 2002). The methods mentioned above have the common aspect of being based upon an unphysical extension that is penalized along with the inversion.

Alongside the extended domain methods, there are other objective functions that mitigate the cycle-skipping problem, including full-traveltime inversion (FTI) (Luo and Schuster, 1991; Luo et al., 2016), adjustive full-waveform inversion (AFWI) (Jiao et al., 2015), FWI with optimal transport distance (Métivier et al., 2016; Yang and Engquist, 2018; Yang et al., 2018), and the scaled-Sobolev objective function (Zuberi and Pratt, 2018), for example.

Wang et al. (2016) applied dynamic warping to shift the recorded data set to less than half a cycle away from the predicted data set to generate a series of data sets, which are used to update velocity models without cycle-skipping. In this paper, we further develop the methodology of Wang et al. (2016) to tackle the cycle-skipping issue using an intermediate data set, which is a generalized concept. The latter is a data set that retains some of the characteristics of the recorded data set. In addition, it is also sufficiently close to the predicted data set such that it is not cycle-skipped with respect to a current model. The new method is described as follows. Firstly, we create an intermediate data set by shifting the predicted data set towards the recorded data set but within half a cycle; secondly, we invert the intermediate data instead of the recorded data; then the first two steps are repeated until the time difference between the predicted data and the recorded data is less than half a cycle; finally, when this condition is met, we complete the inversion with conventional FWI (Tarantola, 1984). Since the intermediate data includes the missing information of the predicted data relative to the recorded data but also are not cycle-skipped to the predicted data, inverting the intermediate data can produce a correct model without cycle-skipping. Furthermore, since the intermediate data originates from the predicted data, the modeling kernel of the inversion is more compatible with the intermediate data than the recorded data, and consequently, the inversion with the intermediate data set is robust. We demonstrate the application of this method in the inversion of first arrivals with numerical examples.

METHOD

Conventional FWI (Tarantola, 1984) is generally formulated as the minimization of the L2-norm of the data residual, expressed as

$$\phi = \frac{1}{2}\left\| Ru - d \right\|_2^2 \ , \qquad (1)$$

subject to

$$\mathrm{A}u = s \ , \qquad (2)$$

where R represents the restriction operator to extract the wavefield at the position of receivers, u is the predicted wavefield, d is the recorded data, A is the operator of the wave equation, and s is the source wavelet. In equation 2, both A and u are a function of the model parameters, m. Generally, in practical applications, the number of the elements of m can reach the order of several millions or even higher. In addition, solving wave equations numerically is very computationally intensive. Consequently, local gradient-based methods, e.g., steepest-descent and conjugate-gradient, are commonly used in FWI to minimize the objective function (equation 1). However, the objective function shown in equation 1 has many local minima – mainly related to the oscillatory nature of seismic data. If the starting model is not close enough to the true model, then the events in the predicted data can be more than half a cycle away from their corresponding events in the recorded data. This phenomenon is known as 'cycle-skipping', and it can cause the iterative optimization to stall at a local minimum.

To illustrate this phenomenon, a trace generated with a 5-Hz Ricker wavelet is used as the record, which is shown as the solid curves in Figures 1b and 1c. The same trace with a different time shift is employed as the predicted data. The objective function (equation 1) as a function of time-lag can then be computed and is illustrated in Figure 1a. The nearest peaks to the global minimum are located at a time lag of -86 ms and 86 ms, which are half a cycle. The dashed curves in Figures 1b and 1c represent the predicted data with a time shift of 50 ms and 125 ms, respectively, which are corresponding to the starting point 1 and 2 in Figure 1a. As can be seen, the two starting points are positioned in different basins of attraction, and the gradients point towards opposite directions. Consequently, local gradient-based methods will converge towards the global minimum of the objective function when starting at point 1 but to a local minimum when starting at point 2. As a result, FWI with starting point 2 produces a wrong estimate of the model parameters. In addition, this wrongly estimated model is likely to be worse than the starting model. Therefore, to achieve a successful inversion with FWI, cycle-skipping should be avoided.

Herein, we introduce a method for tackling cycle-skipping by generating an intermediate data set. The events of this intermediate data sit between those in the predicted data and those in the recorded data. In addition, the events in the intermediate data are less than half a cycle away from the events in the predicted data set. The inversion then inverts intermediate data instead of the recorded data. The mechanism for this method is illustrated by the sketch in Figure 2. The red curve represents the recorded trace while the black curves are for the predicted traces. The initial traveltime difference between the recorded trace and the predicted trace 1 is Δt_0, which is much larger than half a cycle. To avoid cycle-skipping, we can shift the predicted data, which is represented by the blue curves, by Δt_s, which is smaller than half a cycle. This shifted predicted data is the intermediate data, which is closer to the recorded data than the originally predicted data. Consequently, if we invert the intermediate data instead of the recorded data, FWI can produce correct updates. As the correct updates are added into the initial model, the new predicted data, e.g., predicted data i, is closer to the record, but still more than half a cycle away from the record. We can then produce new intermediate data, e.g., intermediate data i, and invert it. By repeating this process, we can gradually improve the initial model to a point such that cycle-skipping is avoided. At this point, conventional FWI can be carried out without encountering issues related to cycle skipping.

In this paper, we only investigate generating an intermediate data set from first arrivals, for the sake of demonstrating the potential and validity of the concept. We also postulate the possibility of using more events in the data in order to generate such mapping. However, we do not explore that possibility further herein.

For a surface acquisition, the first arrivals are direct arrivals for short offsets but refractions for far offsets. The direct arrivals only carry the information for the surface update of the model while the refractions include crucial information for the background update of the deeper part of the model. For a cross-well setting, the first arrivals are transmitted waves, which carry the background update information of the whole model. As a result, the inversion with first arrivals in this paper can successfully update the background model. The algorithm can be outlined as follows:

1. *Calculating the time duration of half a cycle,* $0.5T$*, for the inverted frequencies;*
2. *Picking the first breaks of the record,* d*, which are indicated by the red curve in Figure 3;*
3. *Picking the first breaks of the predicted data,* d_0*, which are represented by the black curve in Figure 3;*
4. *Scaling down the difference between the two first breaks for each shot into a series of time shifts, the maximum of which is less than half a cycle;*
5. *Shifting the predicted data towards the record by the magnitude of the time shifts computed in step 4; the shift produces the intermediate data,* d_i*, the first breaks of which are indicated by the blue curve in Figure 3;*
6. *Producing a window function,* W*, the value of which is one around the first arrivals but decreases to zero when it is away from the first arrivals;*
7. *Minimizing the new objective function,*

$$\phi = \frac{1}{2}\left\| W\left(Ru - d_i\right)\right\|_2^2 ; \qquad (3)$$

8. *Repeating step 3 to step 7 until the difference of the first breaks of the record and the predicted data is smaller than half a cycle.*

In the first step, the quantity of half a cycle is obtained easily by firstly filtering the wavelet to keep the frequencies used for inversion, secondly computing the functional value between the filtered wavelet and its shifted version, which will be like the one in Figure 1a, and finally measuring the time lag of the nearest peak to the zero lag. The approach we outline herein is independent of the first-break picking method. Thus any method of picking first breaks used in exploration geophysics and global seismology can be used here to identify the arrival time of first arrivals. We used the method of Wong et al. (2009) for all the examples in this paper. In the fourth step, the scaling is defined by linearly mapping the difference of the two first breaks in the range from $-0.5\,T$ to $0.5\,T$. The window function, W, can be a Gaussian window or a cosine-squared window, which is used to select the data, so that the inversion is restricted to first arrivals only.

EXAMPLES

In this section, we show numerical examples using a model, which contains two Gaussian anomalies over a homogeneous background, and the Marmousi model to demonstrate the effectiveness of the new method using intermediate data for tackling cycle-skipping.

A synthetic model with two Gaussian anomalies

The synthetic model with two Gaussian anomalies is shown in Figure 4a. Its background velocity is 3000 m/s while the two Gaussian anomalies have a velocity difference of ±1000 m/s to the background. In this example, a cross-well acquisition geometry is chosen. In total, 122 shots are fired at a depth of 100 m, from a distance of 160 m to 9840 m with a shot spacing of 80 m. A 10-Hz Ricker wavelet is used as the source wavelet. The receiver array with 1001 traces for each shot is fixed at a depth of 2900 m.

The initial model shown in Figure 4b has a constant velocity of 2800 m/s, comprising significant differences to the true model. These large differences cause the predicted data to be more than half a cycle away from the recorded data. This can be observed from the comparison of the shot gathers, depicted in Figure 5a. Conventional FWI (Tarantola, 1984) with this starting model will suffer from cycle-skipping. To demonstrate this, we carried out conventional FWI from 5 Hz to 9 Hz with a frequency increment of 1 Hz. Note that one frequency in the inversion means a narrow frequency band because the frequency was extracted using band-pass filtering. Each frequency is inverted for 5 iterations. Afterwards, 15 iterations are carried out with the full bandwidth. In total, the inversion encompasses 40 iterations. The gradient of the first iteration is shown in Figure 6a. One can observe that there are large areas dominated by a negative value of the gradient (in blue color), indicating that the velocity should decrease in these regions. However, the velocity in the starting model should be increased except in the area of the low-velocity blob. Thus, conventional FWI converges towards an inaccurate velocity model as a consequence of cycle-skipping as depicted in Figure 7a.

To deal with cycle-skipping, we generated an intermediate data set and inverted it in each iteration. Firstly, we measured the size of half a cycle, which is about 43 ms for the 10-Hz Ricker wavelet. Secondly, we picked the first breaks for both predicted data and recorded data. Thirdly, we linearly mapped the time difference of each trace between the two picks into a time shift that ranges from -30 ms to 30 ms, so that the maximum shift is less than half a cycle. Fourthly, we applied the shift to the predicted data to generate the intermediate data. A comparison between the predicted data and the intermediate data of one shot is shown in Figure 5b. Finally, we inverted the intermediate data instead of the record by minimizing the objective function shown in Equation 3. The residual is weighted by a cosine-squared function shown in Figure 5c. The inversion is carried out with full-bandwidth data in this example with the intermediate data. The gradient of the first iteration is shown in Figure 6b. One can observe the significant differences between this gradient (Figure 6b) and that for the conventional FWI (Figure 6a). The gradient with intermediate data is dominated by positive values (red color), meaning an increase in velocity. Hence, the inversion with the intermediate data leads to correct updates of the velocity model. The intermediate data is regenerated at the beginning of each iteration.

In this paper, we applied a steepest-descent method to minimize the two objective functions expressed in equations 1 and 3. The gradient is preconditioned by the pseudo-Hessian matrix following the approach of Shin et al. (2001), which is computed by stacking the scaled source-wavefield energy of all shots. The step-length is computed by assuming a linear relationship between the model and data perturbations. Its mathematical derivation is outlined in Appendix A.

The inversion result after 40 iterations is shown in Figure 7b. By comparison, one can see that the inversion with intermediate data properly fixed the background velocity and recovered the two Gaussian anomalies. This high-quality inversion result can also be verified by the good match between the predicted data with the inverted model and the originally generated record, which is shown in Figure 5d. After completing the intermediate data inversion, we then perform inversion with conventional FWI with the same inversion setting as the one shown in Figure 7a. Figure 7c shows the final result after 40 iterations of the conventional FWI. By comparison of Figures 7b and 7c, it can be seen that the conventional FWI further enhanced the Gaussian anomalies and converged to the global minimum.

The Marmousi model

The objective of this example is demonstrating the robustness of the outlined intermediate-data approach even when there is a discrepancy between the laws of physics of the inversion algorithm and that of the real world. This aspect is relevant as most inversion algorithms assume that the Earth is a fluid. However, elastic effects generally affect the acquired data.

In this test, we used both a surface geometry and a surface-to-horizontal-well geometry, which is a 90-degree rotated cross-well setting. In order to mimic the real world, we generated the record with the elastic wave equation (Virieux, 1986) in isotropic media with the true velocity models of the P-wave (Figure 8a) and the S-wave (Figure 8b) and the true density model. The pressure record is generated by summing the normal stress components, τ_{xx} and τ_{zz}. In each geometry, we generated the data firing 128 shots at a depth of 25 m from a distance of 0.187 km to 12.888 km with a spacing of 100 m. The source wavelet is a 10-Hz Ricker wavelet. The receiver array has 1051 receivers and their position is fixed at a depth of 25 m for the surface geometry and 2900 m for the surface-to-horizontal-well geometry. An absorbing boundary is applied around the entire domain for the surface-to-horizontal-well geometry but a free-surface boundary condition is used on the top boundary of the model for the surface geometry, introducing source and receiver ghosts as well as surface-related multiples.

In our inversion scheme, we invert for P-wave velocity only. In this test, the inversion includes two stages. In the first stage, we aim to use the proposed intermediate-data method to correct the background P-wave velocity. An acoustic wave equation with a constant density of 1000 kg/m^3 was used as the modeling kernel for FWI with intermediate data, which is achieved by minimizing equation 3. In the second stage, we applied conventional FWI, which minimizes equation 1 (Tarantola, 1984), to refine the P-wave velocity model recovered from the first stage. In order to honor the amplitude of the recorded data, an elastic wave equation was applied in this stage, in which the density is constrained with Gardner's relation (Gardner et al., 1974) and the S-wave velocity is formed using a constant vp/vs ratio of 1.9 throughout each iteration.

In the first stage of this test, we ran the inversion with the intermediate data from 5 Hz to 10 Hz with an increment of 1 Hz. Each frequency means a narrow band and was inverted for 10 iterations. This means that the inversion completed a total of 60 iterations. We used a maximum offset of 5 km and 10 km for the surface-to-horizontal-well geometry and the surface geometry, respectively. The resulting P-wave velocity model was further used as a starting model for conventional FWI in the second stage. This inversion was carried out starting at 5 Hz up to 24 Hz, incrementing with 1 Hz after completing a set of 5 iterations. Overall this inversion represents a total of 100 iterations of FWI. All offsets present in the data were used in this stage.

Surface-to-horizontal-well geometry

In the surface-to-horizontal-well test, we designed a 1D initial velocity model shown in Figure 8c, which has substantial errors. One shot of the predicted data from the initial model (Figure 8c) is shown in Figure 9b. Compared to the corresponding recorded shot, one can observe that the first arrivals of predicted data are fully cycle-skipped in relation to those in recorded data, which is indicated by the black dashed curves.

To overcome cycle-skipping, we applied the same method as in the first example, i.e., we shifted the predicted data towards the recorded data. The shift was calculated by linearly scaling down the difference between the first breaks into the range from -30 ms to 30 ms. Consequently, the maximum shift is less than half a cycle of the 10-Hz Ricker wavelet, which is about 43 ms. By shifting the predicted data towards the recorded data, we got the intermediate data, which carries the mismatch information between the predicted data and the recorded data but is not cycle-skipped. In addition, the shift only extracts the traveltime information from the recorded data but does not extract any elastic amplitude information, which cannot be properly handled by an acoustic modeling kernel. One shot gather of the intermediate data is shown in Figure 9c.

The gradient of the conventional FWI and the gradient of FWI with the intermediate data for the first iteration are shown in Figure 10. One can observe that the gradient of the conventional FWI has a negative background gradient on the left side but positive on the right side. However, since the initial velocity is much lower than that of the true velocity, a correct gradient should be positive (in red color). In addition, this gradient includes strong high-wavenumber events generated by the reflection events in the record. These high-wavenumber components do not carry information about the background model; hence they do not contribute to its update. By contrast, the gradient of FWI with the intermediate data is smooth and is dominated by the red color, which means the inversion will decrease the slowness (increase the velocities).

Figure 12 shows the recovered model. Compared with the true P-wave velocity model (Figure 8a), one can observe that the conventional FWI inversion converged towards an inaccurate estimate of the P-wave velocity model (Figure 12a) because of the cycle-skipping while the FWI with intermediate data inverted a smooth velocity model (Figure 12b), which has a correct background trend of the true P-wave velocity. The effectiveness of the intermediate-data method can be seen from the first-break picks of predicted data moving towards that of the recorded data progressively throughout iterations (Figure 11). This can also be seen from the model and gradient evolution shown in Figures B-3 and B-4 of Appendix B. However, if the intermediate data is generated by shifting the recorded data towards the predicted data, we obtained a much less accurate P-wave velocity, which is shown in Figure 12c.

The reason for this is that there is waveform discrepancy of the first arrivals between the predicted data and the recorded data due to different laws of physics applied for the data generation. One illustration is shown in Figures B-1 and B-2 of Appendix B. This less accurate result can also be seen from the vertical profile at the distance of 6.25 km and model errors shown in Figure 13. Since FWI with intermediate data recovered a good background P-wave velocity model (Figure 12b), subsequent conventional FWI with the surface record recovered a high-resolution correct velocity model, which is shown in Figure 12d.

Surface Geometry

For the surface acquisition geometry, we chose a simple initial velocity model shown in Figure 8d. As the turning waves and refractions have a good coverage on the top central part of this model, a Gaussian-shape low-velocity anomaly is embedded there for emphasizing the cycle-skipping.

Figure 14a depicts a shot gather of the recorded data. The first-break picks are represented by the dashed curves. The corresponding shot gather of the predicted data from the initial velocity model is shown in Figure 14b. By comparison, one can see that the first arrivals for the direct arrivals are matched very well but the refractions are not, which means that the background velocity of the initial velocity model is inaccurate, as it does not produce the correct traveltime for diving and refracted waves. The mismatch for the far offsets is even more than half a cycle, which means that the conventional FWI converges towards a local minimum.

To overcome cycle-skipping, we shifted the predicted data in the same way as we did for the surface-to-horizontal-well test to generate the intermediate data. Figure 14c shows the shot gather of the intermediate data set, corresponding to the shot gather in Figure 14a. As the intermediate data has less than half a cycle time shift to the predicted data, carrying out FWI with it generates a gradient that is not affected by cycle-skipping. Figure 15c shows the gradient at the first iteration. As can be seen, the gradient is dominated by positive values, which means the inversion decreases the slowness (increases the velocity). Note that the gradient shown in Figure 15c includes only low wavenumbers because the intermediate data includes only first arrivals, mainly refracted waves.

By contrast, Figure 15a shows the corresponding gradient from the conventional FWI. The high-wavenumber components of the gradient are generated by the reflection events while the low-wavenumber components, which are essential for the background update, are formed mainly from refracted waves. Note that during the inversion, the water layer, the bottom boundary of which is indicated by the green dotted line, is set as the true water velocity, 1500 m/s. In order to highlight the low-wavenumber components, we smoothed the gradient shown in Figure 15a with a Gaussian filter, $\frac{1}{2\pi\cdot\sigma^2}\cdot e^{-\frac{x^2+z^2}{2\sigma^2}}$, where the standard deviation parameter, σ, is 10, and x and z are in number of cells (Abdullah and Schuster, 2015). The smoothed version of the gradient is shown in Figure 15b. As can be seen, the gradient around the area at a distance of 6 km and a depth of 1 km is negative, which indicates that the inversion performs an incorrect update in that region due to cycle-skipping. Since the low-wavenumber components of the gradient have an incorrect sign in the middle of the top area due to cycle-skipping, the conventional FWI does not fix the background velocity but effectively migrates the reflection events in this area and below in a least-square sense. However, these migrated events are at the wrong depth and defocused as the background velocity model is incorrect, and it is not updated throughout the successive iterations. This can be observed in Figure 17a: the faults are obscured; an artificial low-velocity anomaly, which is a typical consequence of cycle-skipping during the application of FWI on the Marmousi model, appears close to the top of the faults.

By contrast, FWI with intermediate data produced a proper background update, which can be observed in Figure 17b. As a result, the predicted data matches the picked first break of the recorded data (Figure 14d). The effectiveness of the inversion can be further seen from the first-break picks of predicted data shifting towards those of recorded data gradually throughout iterations (Figure 16). Figures B-5 and B-6 in Appendix B provide further evidence of the effectiveness. However, the inversion becomes much less efficient with intermediate data generated by shifting the recorded data towards predicted data. Its result is shown in Figure 17c. This can be further verified by the vertical profile at a distance of 6.25 km and the model errors shown in Figure 18. We can also observe one interesting phenomenon from Figure 14: a smooth background model produces a simple record, where the first arrival is the primary and the other arrivals are predominantly multiples. Thus, the first arrivals may contain sufficient information for updating the background P-wave velocity. We then carried out the conventional FWI utilizing the P-wave velocity model estimated with the intermediate data (Figure 17b). Figure 17d shows the resulting high-resolution P-wave velocity model.

DISCUSSION

In this paper, we have demonstrated using intermediate data to tackle cycle-skipping of first arrivals. The first arrivals in cross-well geometries and at the far-offsets in the surface acquisitions contain information for background velocity updates. As a result, we have successfully recovered the background velocity, which is accurate enough for starting the conventional FWI to build the intermediate- and high-wavenumber components of the velocity model. In this application of the intermediate-data concept to first arrivals, we need to pick the first arrivals to the accuracy of half a period of the starting inversion frequency. Less than this accuracy may result in an inadequate result for the following conventional FWI. The routine first-arrival picking methods, e.g., the short-term-average over long-term-average ratio (STA/LTA) (Allen, 1982) and energy ratio (Wong et al., 2009), may be sufficient to process the predicted data. However, the first arrivals of the real data are usually contaminated by noise. In this case, we may need to seek other, more intelligent picking methods, for example, the methods based on artificial intelligence (e.g., Chen, 2018; Yuan et al. 2018), or even with the assistance of manual picking.

The intermediate data approach introduced herein includes only the traveltime errors of the first arrivals. Compared with other conventional velocity-building methods relying on the first arrivals, e.g., first-break traveltime tomography (e.g., Zelt and Smith 1992), one noticeable difference is that FWI with intermediate data estimates the velocity model with a wave equation instead of ray tracing. The key advantages are then twofold. First, the wave equation yields more accurate wave paths as it can deal more easily with multi-arrivals and is not affected by the existence of shadow areas and caustics, which exist in ray tracing. Second, it is a natural extension of any conventional FWI algorithm, only requiring the implementation of a method for traveltime picking. In fact, we can further extend the concept of intermediate data to much broader applications of FWI than using the traveltime information of the first arrivals only. Algorithms, such as dynamic warping (Hale, 2013; Ma and Hale, 2013), are potentially good alternatives when mapping multiple cycle-skipped events, including reflection events, in the predicted data to the corresponding events in the recorded data, for producing an intermediate data set. Inverting the intermediate data set will avoid cycle-skipping (e.g., Wang et al., 2016).

We can also investigate the concept of intermediate data to deal with other complications existing in FWI. It is well known that it is much easier to invert 'inverse-crime' data sets, which are generated by using the modeling kernels for inversion, than field data sets or data sets generated with different kernels. To mitigate this difficulty, we can generate an intermediate data set that has the information of the recorded data set, e.g., traveltime, missed in the predicted data set but also is compatible with the modeling kernel of FWI, e.g. the Marmousi example in this paper. Inverting the intermediate data set is then equivalent to inverting the desired information in the recorded data set only. This approach has the advantage of eliminating undesired events, such as records of S-wave, which cannot be handled by the modeling kernel if the Earth is considered to be an acoustic body. The existence of events that cannot be accounted for by the modeling kernel can introduce artifacts in the estimated models. Hence, this approach can, in principle, mitigate these artifacts.

CONCLUSION

In this paper, we have proposed a generalized concept of intermediate data to tackle cycle-skipping in full-waveform inversion (FWI). The principle of this method is to create an intermediate data set, the events of which sit between the events of the predicted data and the recorded data and are less than half a cycle away from the predicted data, and then invert the intermediate data rather than the recorded data. We have successfully applied this concept to invert the first arrivals with both surface-to-horizontal-well and surface acquisition geometries. In these applications, we picked first breaks of both the recorded data set and the predicted data set, and then linearly scaled down the time difference between the two first breaks into a time shift that is less than half a cycle. We then create the intermediate data set by moving the predicted data set with the time shift, and finally invert the intermediate data set. The theoretical analyses and numerical examples have validated this method. Our tests also show that shifting the predicted data yields a more accurate inverted velocity model than shifting the recorded data, especially in the case of existing discrepancies between the laws of physics of the inversion algorithm and that of the real world. Moreover, this concept of intermediate data might be applied to much broader areas in FWI than just using the first arrivals to drive the inversion. With other multi-event mapping algorithms instead of first-break picking, the concept of intermediate data might be extended to overcome cycle-skipping for multi-events. There is also potential for the concept of intermediate data to deal with other complications in FWI, such as inverting data sets including S-waves with acoustic modeling kernels.

ACKNOWLEDGEMENTS

This research was supported by the National Key Research and Development Program of China (No. 2017YFC1500303), NSFC (Grant No. 41630209), NSF (Grant No. 1547228), Science Foundation of China University of Petroleum, Beijing (No. 2462018BJC001, No. 2462015YJRC012), and the FULLWAVE consortium. The authors would like to thank Editor, Anatoly Baumstein and Deyan Draganov, and Reviewer, Ettore Biondi and two anonymous reviewers, for their comments and suggestions, which helped to improve and clarify the manuscript significantly. Finally, the authors would also like to show gratitude to our colleague, George Stronge, for the proofreading.

APPENDIX A

CALCULATION OF STEP LENGTH

To find the optimal step length, α, that minimizes the functional, we start from the current model, $\mathbf{m}_0$, which gives the residual, $\delta\mathbf{d}_0$. We then perturb the current model by a small amount, $\delta\mathbf{m}$ to form a trial model,

$$\mathbf{m}_1 = \mathbf{m}_0 + \delta\mathbf{m}, \quad \text{(A-1)}$$

where $\delta\mathbf{m}$ can be generated by scaling the preconditioned gradient to a few percentages of $\mathbf{m}_0$ in magnitude. Our target is to find a new model,

$$\mathbf{m}_\alpha = \mathbf{m}_0 + \alpha\delta\mathbf{m} = \mathbf{m}_0 + \alpha\left(\mathbf{m}_1 - \mathbf{m}_0\right) \quad \text{(A-2)}$$

that minimizes the functional

$$\frac{1}{2}\left\|\delta\mathbf{d}_\alpha\right\|_2^2, \quad \text{(A-3)}$$

where $\delta\mathbf{d}_\alpha$ denotes the data residual corresponding to $\mathbf{m}_\alpha$. By assuming the data perturbation is linearly dependent on the model perturbation, we then have

$$\delta\mathbf{d}_\alpha = \delta\mathbf{d}_0 + \alpha\left(\delta\mathbf{d}_1 - \delta\mathbf{d}_0\right). \quad \text{(A-4)}$$

If we insert equation A-4 into equation A-3, and set the derivative of the functional with respect to α to zero, then we obtain the optimal step length,

$$\alpha = \frac{\delta\mathbf{d}_0^{\mathrm{T}}\left(\delta\mathbf{d}_1 - \delta\mathbf{d}_0\right)}{\left(\delta\mathbf{d}_1 - \delta\mathbf{d}_0\right)^{\mathrm{T}}\left(\delta\mathbf{d}_1 - \delta\mathbf{d}_0\right)}. \quad \text{(A-5)}$$

APPENDIX B

SUPPLEMENTARY FIGURES

In this section, we provide additional tests and figures to the main body of this paper. The first test is to compare the seismic shot gathers generated with the true P-wave, S-wave velocities, and densities of the Marmousi model. Figure B-1 shows one shot gather generated with the surface-to-horizontal-well geometry, which is used for the Marmousi example. The source is located at a distance of 3.1875 km. The source signature is a 10-Hz Ricker wavelet. Figure B-2 shows the counterpart of Figure B-1 with the surface geometry used for the Marmousi example. By comparison of these figures, we can see clearly that the records with different laws of physics share similar overall appearance and first-break times, but have a significant difference in wiggles and amplitudes.

Figures B-3 and B-5 show the gradients at different iterations in FWI with intermediate data while Figures B-4 and B-6 depict the evolution of the velocity model during the inversion.

REFERENCES

Abdullah, A., and G. T. Schuster, 2015, Inverting reflections using full-waveform inversion with inaccurate starting models: 85th Annual International Meeting, SEG, Expanded Abstracts, 1148-1153.

Allen, R., 1982, Automatic phase pickers: Their present use and future prospects: Bulletin of the Seismological Society of America, **72**, no. 6B, S225-S242.

Baeten, G., J. W. de Maag, R.-E. Plessix, R. Klaassen, T. Qureshi, M. Kleemeyer, F. t. Kroode, and Z. Rujie, 2013, The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study: Geophysical Prospecting, **61**, 701–711, doi: 10.1111/1365-2478.12010.

Baumstein, A., J. Anderson, D. Hinkley, and J. Krebs, 2009, Scaling of the objective function gradient for full wavefield inversion: 79th Annual International Meeting, SEG, Expanded Abstracts, 2243-2247.

Baumstein, A., 2014, Extended subspace method for attenuation of crosstalk in multi-parameter full wavefield inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1121-1125.

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

Biondi, E., G. Barnier, and B. Biondi, 2017, Preconditioned elastic full-waveform inversion with approximated Hessian: 79th Annual International Meeting, SEG, Expanded Abstracts, 1654-1658.

Brown, S., 2007, Performance comparison of finite-difference modeling on Cell, FPGA, and multicore computers: 77th SEG Annual meeting, SEG, Expanded Abstracts, 2110-2114.

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

Burgess, T., and M. Warner, 2015, Preconditioning FWI with approximate receiver Green's functions: 85th SEG Annual meeting, SEG, Expanded Abstracts, 1116-1121.

Chen, M., F. Niu, Q. Liu, J. Tromp, and X. Zheng, 2015, Multiparameter adjoint tomography of the crust and upper mantle beneath East Asia: 1. Model construction and comparisons: Journal of Geophysical Research: Solid Earth, **120**, 1762–1786, doi: 10.1002/2014JB011638.

Chen, Y, 2018, Automatic microseismic event picking via unsupervised machine learning: Geophysical Journal International, **212**, no. 1, 88-102, doi: 10.1093/gji/ggx420.

Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

da Silva, N. V., A. Ratcliffe, V. Vinje, and G. Conroy, 2016, A new parameter set for anisotropic multiparameter full-waveform inversion and application to a North Sea data set: Geophysics, **81**, no. 4, U25–U38, doi: 10.1190/geo2015-0349.1.

da Silva, N. V., and G. Yao, 2018, Wavefield reconstruction inversion with a multiplicative cost function: Inverse Problems, **34**, 015004, doi: 10.1088/1361-6420/aa9830.

da Silva, N., G. Yao, and M. Warner, 2018, Semi-global visco-acoustic full waveform inversion: Geophysics,1-78. doi: 10.1190/geo2017-0773.1.

Dai, W., P. Fowler, and G. T. Schuster, 2012, Multi-source least-squares reverse time migration: Geophysical Prospecting, **60**, 681–695.

Debens, H., M. Warner, A. Umpleby, and N. Silva, 2015, Global anisotropic 3D FWI: 85th Annual International Meeting, SEG, Expanded Abstracts, 1193–1197.

Dellinger, J., A. Ross, D. Meaux, A. Brenders, G. Gesoff, J. Etgen, J. Naranjo, G. Openshaw, and M. Harper, 2016, Wolfspar®, an "FWI-friendly" ultralow-frequency marine seismic source: 86th Annual International Meeting, SEG, Expanded Abstracts, 4891-4895.

Esser, E., L. Guasch, F. Herrmann, and M. Warner, 2016, Constrained waveform inversion for automatic salt flooding: The Leading Edge, 35, no. 3, 235-239, doi: 10.1190/tle35030235.1.

Fehmers, G. C., and Höcker, C. F. W., 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, no. 4, 1286-1293, doi: 10.1190/1.1598121.

Fu, L., and W. Symes, 2017, An adaptive multiscale algorithm for efficient extended waveform inversion: Geophysics, **82**, no. 3, R183–R197, doi: 10.1190/geo2016–0426.1.

Gardner, G., L. Gardner and A. Gregory, 1974, Formation velocity and density — the diagnostic basics for stratigraphic traps: Geophysics, **39**, no. 6, 770-780, doi: 10.1190/1.1440465.

Hale, D., 2013, Dynamic warping of seismic images: Geophysics, **78**, no. 2, S105–S115, doi: 10.1190/geo2012-0327.1.

Herrmann, F. J., I. Hanlon, R. Kumar, T. van Leeuwen, X. Li, B. Smithyman, H. Wason, A. J. Calvert, M. Javanmehri, and E. T. Takougang, 2013, Frugal full-waveform inversion: From theory to a practical algorithm: The Leading Edge, **32**, no. 9, 1082-1092, doi: 10.1190/tle32091082.1

Hu, W., 2014, FWI without low frequency data - beat tone inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1116–1120.

Hu, W., 2016, Optimal Green's function estimation for FWI acceleration: 86th Annual International Meeting, SEG, Expanded Abstracts, 1074-1078.

Huang, G., R. Nammour, and W. Symes, 2017, Full-waveform inversion via source-receiver extension: Geophysics, **82**, no. 3, R153–R171, doi: 10.1190/geo2016-0301.1.

Jannane, M., W. Beydoun, E. Crase, D. Cao, Z. Koren, E. Landa, M. Mendes, A. Pica, M. Noble, G. Roeth, S. Singh, R. Snieder, A. Tarantola, D. Trezeguet, and M. Xie, 1989, Wavelengths of earth structures that can be resolved from seismic reflection data: Geophysics, **54**, 906–910, doi: 10.1190/1.1442719.

Jiao, K., D. Sun, X. Cheng, and D. Vigh, 2015, Adjustive full waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1091–1095.

Kalinicheva, T., M. Warner, J. Ashley, and F. Mancini, 2017, Two- vs three-dimensional full-waveform inversion in a 3D world: 87th Annual International Meeting, SEG, Expanded Abstracts, 1383–1387.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, no. 6, WCC177-WCC188, doi: 10.1190/1.3230502.

Li, Y., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data: Geophysics, **81**, no. 6, R339-R348, doi: 10.1190/geo2016–0038.1.

Liu, Z., and J. Zhang, 2017, Joint traveltime, waveform, and waveform envelope inversion for near-surface imaging: Geophysics, **82**, no. 4, R235–R244, doi: 10.1190/geo2016-0356.1.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792.

Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, no. 5, 645–653, doi: 10.1190/1.1443081.

Luo, Y., Y. Ma, Y. Wu, H. Liu, and L. Cao, 2016, Full-traveltime inversion: Geophysics, **81**, no. 5, R261–R274, doi: 10.1190/geo2015-0353.1.

Ma, Y., and D. Hale, 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: Geophysics, **78**, no. 6, R223–R233, doi: 10.1190/geo2013-0004.1.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion: Geophysical Journal International, **205**, 345–377, doi: 10.1093/gji/ggw014.

Morgan, J. V., S. P. S. Gulick, T. Bralower, E. Chenot, G. Christeson, P. Claeys, C. Cockell, G. S. Collins, M. J. L. Coolen, L. Ferrière, C. Gebhardt, K. Goto, H. Jones, D. A. Kring, E. Le Ber, J. Lofi,

X. Long, C. Lowery, C. Mellett, R. Ocampo-Torres, G. R. Osinski, L. Perez-Cruz, A. Pickersgill, M. Poelchau, A. Rae, C. Rasmussen, M. Rebolledo-Vieyra, U. Riller, H. Sato, D. R. Schmitt, J. Smit, S. Tikoo, N. Tomioka, J. Urrutia-Fucugauchi, M. Whalen, A. Wittmann, K. E. Yamaguchi, and W. Zylberman, 2016, The formation of peak rings in large impact craters: Science, **354**, 878–882, doi: 10.1126/science.aah6561.

Mulder, W., and A. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, no. 4, 1184-1191, doi: 10.1190/1.1500380.

Nemeth, T., J. Stefani, W. Liu, R. Dimond, O. Pell, and R. Ergas, 2008, An implementation of the acoustic wave equation on FPGAs: 78th Annual International Meeting, SEG, Expanded Abstracts, 2874-2878.

Plessix, R. E., W. A. Mulder, and A. P. E. ten Kroode, 2000, Automatic cross-well tomography by semblance and differential semblance optimization: theory and gradient computation: Geophysical Prospecting, **48**, no. 5,913-935, doi: 10.1046/j.1365-2478.2000.00217.x.

Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2017, Salt model building at Atlantis with full-waveform inversion: 87th Annual International Meeting, SEG, Expanded Abstracts, 1507-1511.

Shin, C., and Y. Ho Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931, doi:10.1111/j.1365-246X.2008.03768.x.

Shin, C., and Y. Ho Cha, 2009, Waveform inversion in the Laplace-Fourier domain: Geophysical Journal International, **177**, 1067–1079, doi:10.1111/j.1365-246X.2009.04102.x.

Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606, doi: 10.1046/j.1365-2478.2001.00279.x.

Symes, W., and J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, no. 5,654-663. doi: 10.1190/1.1443082.

Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: 10.1111/j.1365-2478.2008.00698.x.

Tao, K., S. P. Grand, and F. Niu, 2017, Full-waveform inversion of triplicated data using a normalized-correlation-coefficient-based misfit function: Geophysical Journal International, **210**, no. 3, 1517-1524, doi: 10.1093/gji/ggx249.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, no. 8, 1259-1266, doi:10.1190/1.1441754.

Tikhonov, A. N. and V. Y. Arsenin, 1977, Solution of Ill-posed Problems: John Wiley & Sons.

Trinh, P.T., R. Brossier, L. Métivier, J. Virieux, and P. Wellington, 2017, Bessel smoothing filter for spectral-element mesh: Geophysical Journal International, **209**, no. 3, 1489–1512, doi: 10.1093/gji/ggx103.

van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, 182, **no**. 3, 1383-1394, doi: 10.1111/j.1365-246X.2010.04681.x.

van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, 195, **no**. 1, 661-667, doi: 10.1093/gji/ggt258.

Virieux, J, 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, no. 4, 889-901, doi: 10.1190/1.1442147.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.

Wang, M., Y. Xie, W. Xu, F. Loh, K. Xin, B. Chuah, T. Manning, and S. Wolfarth, 2016, Dynamic-warping full-waveform inversion to overcome cycle skipping: 86th Annual International Meeting, SEG, Expanded Abstracts, 1273-1277.

Wang, Y., W. Liang, Z. Nashed, X. Li, G. Liang, and C. Yang, 2014, Seismic modeling by optimizing regularized staggered-grid finite-difference operators using a time-space-domain dispersion-relationship-preserving method: Geophysics, **79**, no. 5, T277–T285, doi: 10.1190/geo2014-0078.1.

Wang, Z., J. Li, B. Wang, Y. Xu, and X. Chen, 2017, Time-domain explicit finite-difference method based on the mixed-domain function approximation for acoustic wave equation: Geophysics, **82**, no. 5, T237–T248, doi: 10.1190/geo2017-0012.1.

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Štekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, no. 2, R59–R80, doi: 10.1190/geo2012-0338.1.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81**, no. 6, R429–R445, doi: 10.1190/geo2015-0387.1.

Weiss, R. M., and J. Shragge, 2013, Solving 3D anisotropic elastic wave equations on parallel GPU devices: Geophysics, **78**, no.2, F7-F15, doi: 10.1190/geo2012-0063.1.

Wong, J., L. Han, J. C. Bancroft, and R. R. Stewart, 2009, Automatic time-picking of first arrivals on noisy microseismic data, CSEG conference Abstract 2009, https://www.crewes.org/ForOurSponsors/ConferenceAbstracts/2009/CSEG/Wong_CSEG_2009.pdf, accessed 25 May 2018

Wu, R., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, no. 3, WA13–WA24, doi: 10.1190/geo2013-0294.1.

Yang, Y., and B. Engquist, 2018, Analysis of optimal transport and related misfit functions in FWI: Geophysics, **83**, no. 1, A7-A12, doi: 10.1190/geo2017-0264.1.

Yang, Y., B. Engquist, J. Sun, and B. Froese, 2018, Application of Optimal Transport and the Quadratic Wasserstein Metric to Full-Waveform Inversion: Geophysics, **83**, no. 1, R43-R62, doi: 10.1190/geo2016-0663.1.

Yao, G., and D. Wu, 2015, Least-squares reverse-time migration for reflectivity imaging: Science China Earth Sciences, **58**, 1982–1992, doi: 10.1007/s11430-015-5143-1.

Yao, G., and H. Jakubowicz, 2016, Least-squares reverse-time migration in a matrix-based formulation: Geophysical Prospecting, **64**, no. 3, 611-621, doi: 10.1111/1365-2478.12305.

Yao, G., D. Wu, and H. A. Debens, 2016, Adaptive finite difference for seismic wavefield modelling in acoustic media: Scientific Reports, **6**, 30302, doi: 10.1038/srep30302.

Yao, G., N. V. da Silva, M. Warner, A. Umpleby, and D. Wu, 2017, Improved FWI convergence using efficient receiver-side spatial preconditioning employing ray theory: 79th EAGE Conference and Exhibition, EAGE, Extended Abstracts, We P1 14.

Yao, G., N. V. da Silva, H. A. Debens, and D. Wu, 2018a, Accurate seabed modeling using finite difference methods: Computational Geosciences, **22**, no. 2, 469-484, doi: 10.1007/s10596-017-9705-5.

Yao, G., N. V. Silva, M. Warner, and T. Kalinicheva, 2018b, Separation of migration and tomography modes of full-waveform inversion in the plane wave domain: Journal of Geophysical Research: Solid Earth, **123**, no. 2, 1486-1501, doi: doi:10.1002/2017JB015207.

Yuan, S., J. Liu, S. Wang, T. Wang, and P. Shi, 2018, Seismic Waveform Classification and First-Break Picking Using Convolution Neural Networks: IEEE Geoscience and Remote Sensing Letters, **15**, no. 2, 272-276, doi: 10.1109/LGRS.2017.2785834.

Zelt, C. A., and R. B. Smith, 1992, Seismic traveltime inversion for 2-D crustal velocity structure: Geophysical Journal International, **108**, no. 1, 16-34, doi: 10.1111/j.1365-246X.1992.tb00836.x.

Zhang, J., and Z. Yao, 2012, Optimized finite-difference operator for broadband seismic wave modeling: Geophysics, **78**, no. 1, A13–A18, doi: 10.1190/geo2012-0277.1.

Zhang, Y., and B. Biondi, 2013, Moveout-based wave-equation migration velocity analysis: Geophysics, **78**, no. 2, U31–U39, doi: 10.1190/geo2012-0082.1.

Zhu, H., E. Bozdağ, D. Peter, and J. Tromp, 2012, Structure of the European upper mantle revealed by adjoint tomography: Nature Geoscience, **5**, 493, doi: 10.1038/ngeo1501

Zhu, H., and S. Fomel, 2016, Building good starting models for full-waveform inversion using adaptive matching filtering misfit: Geophysics, **81**, no. 5, U61-U72, doi: 10.1190/geo2015-0596.1.

Zuberi, M. A., and R. G. Pratt, 2018, Mitigating nonlinearity in full waveform inversion using scaled-Sobolev pre-conditioning: Geophysical Journal International, **213**, no. 1, 706-725, doi: 10.1093/gji/ggx549.

List of figure caption:

Figure 1. Schematic illustration of cycle-skipping. (a) The value of the objective function, which is the square of the L2-norm of the difference between a 5-Hz Ricker wavelet and its shifts. The two black dots represent the global minimum and a local minimum. The nearest peaks to the global minimum has a time lag of 86 ms, which corresponds to half a cycle, $0.5T$. (b) The 5-Hz Ricker wavelet and its shift by 50 ms, which is less than half a cycle. (c) The same as (b) but the shift is 125 ms, which is greater than half a cycle. The functional value for the case of (b) and (c) is indicated by the two crosses in (a).

Figure 2. Schematic illustration for the method of tackling cycle-skipping with intermediate data. The predicted data 1 is generated with the initial model. The recorded data indicated by the red trace has a time difference of Δt_0 relative to predicted data 1. Δt_0 is greater than half a cycle. Shifting the predicted data by Δt_s , which is smaller than half a cycle, generates intermediate data 1. Replacing the record with intermediate data 1, FWI will update the model to make the predicted data fit intermediate data 1. Then shifting the new predicted data creates the new intermediate data. By repeating the process, the inversion will recover a model, which can produce the predicted data having a time difference to the record less than half a cycle.

Figure 3. Schematic illustration of creating intermediate data by using the first arrivals.

Figure 4. A model with two Gaussian anomalies. (a) The true velocity model, which includes a constant background of 3000 m/s, a high-velocity anomaly and a low-velocity anomaly. The anomalies are formed with a Gaussian function of $\pm 1000\, e^{-\frac{d^2}{5\text{x}10^5}}$, where d (unit: m/s) is the distance to the center

of the anomaly. The red dotted line indicates the location of sources while the blue dotted line represents the receiver arrays. (b) The initial velocity model, which is a constant velocity of 2800 m/s.

Figure 5. Overcoming cycle-skipping with intermediate data. (a) Alternating display of a shot gather of the predicted data from the initial velocity model and the record. (b) Alternating display of a shot gather of the predicted data from the initial velocity model and the intermediate data. (c) A cosine-squared weighting function used to restrict the inversion to the first arrivals only. (d) Alternating display of a shot gather of the predicted data from the recovered model and the recorded data. In all the alternating displays, the predicted data are shown first with the 'red-white-black' colormap, while the record and the intermediate data are depicted with the 'brown-white-black' colormap.

Figure 6. The gradient of (a) conventional FWI and (b) FWI with intermediate data in the first iteration. During the inversion, the model is updated through slowness, so that the negative gradient (blue color) indicates FWI will decrease velocity while the positive gradient (red color) means FWI will increase velocity.

Figure 7. The recovered velocity models from (a) conventional FWI starting from the initial model, (b) FWI with intermediate data starting from the initial model, and (c) conventional FWI starting from (b).

Figure 8. The Marmousi models and the initial models for inversion. (a) The true P-wave velocity model. (b) The true S-wave velocity model. (c) The simple 1D initial P-wave velocity model for the surface-to-horizontal-well test. (d) The simple initial P-wave velocity model, in which a Gaussian low-velocity anomaly is embedded around the fault area, for the test with a surface acquisition. In (a), the red dotted

line indicates the locations of the sources and the receiver array of the surface acquisition while the blue dotted line represents the locations of the receiver array of the surface-to-horizontal-well setting.

Figure 9. One shot gather in FWI with intermediate data for the surface-to-horizontal-well test. (a) Recorded data. (b) Predicted data with the initial velocity model shown in Figure 8c. (c) The intermediate data generated by shifting the predicted data shown in (b) within half a cycle. (d) The predicted data with the recovered model shown in Figure 12b. The dashed black curves indicate the first-break picks of the record shown in (a).

Figure 10. The gradient of (a) conventional FWI and (b) FWI with intermediate data in the first iteration. The green dotted lines indicate the sea bottom in the Marmousi model.

Figure 11. Evolution of first breaks throughout the iterations. The red curve indicates the first break of the recorded data while the blue curve represents the first break of the predicted data after 60 iterations. The black curves denote the first breaks for early iterations. The iteration numbers are drawn on top of these curves. The light gray strip around the red curve indicates the half period of the 5-Hz Ricker wavelet.

Figure 12. Recovered P-wave velocity models. (a) The recovered P-wave velocity model by using conventional FWI starting from the initial model shown in Figure 8c. (b) The recovered P-wave velocity model from the initial velocity model shown in Figure 8c by using FWI with intermediate data shifting of the predicted data towards the observed data. (c) The same as (b) but the intermediate data are generated by shifting the observed data towards the predicted data. (d) The recovered P-wave velocity with the conventional FWI starting from the velocity shown in (b).

Figure 13. (a) Velocity profile at a distance of 6.25 km. (b) Normalized L2-norm model errors. The black solid curves indicate the true velocity model while the dashed black curves are for the initial velocity model. The red and blue curves are for the recovered velocity model shown in Figures 12b and 12c, respectively.

Figure 14. One shot gather in FWI with intermediate data for the test of the surface acquisition geometry. (a) Recorded data. (b) Predicted data with the initial velocity model shown in Figure 8d. (c) The intermediate data generated by shifting the predicted data shown in (b) within half a cycle. (d) The predicted data with the recovered model shown in Figure 17b. The dashed black curves indicate the first-break picks of the record shown in (a).

Figure 15. The gradient of (a) conventional FWI and (c) FWI with the intermediate data in the first iteration. (b) A smoothed version of (a) using a Gaussian filter, i.e. $\frac{1}{2\pi\cdot\sigma^2}\cdot e^{-\frac{x^2+z^2}{2\sigma^2}}$, where $\sigma=10$, and x and z are in number of cells. The green dotted lines indicate the sea bottom in the Marmousi model.

Figure 16. Evolution of the first breaks throughout the iterations. The red curve indicates the first break of the recorded data while the blue curve represents the first break of the predicted data after 60 iterations. The black curves denote the first breaks for early iterations. The iteration numbers are drawn on top of these curves. The light gray strip around the red curve indicates the half period of the 5-Hz Ricker wavelet.

Figure 17. Recovered P-wave velocity models. (a) The recovered P-wave velocity model by using conventional FWI starting from the initial model shown in Figure 8d. (b) The recovered P-wave velocity

model from the initial velocity model shown in Figure 8d by using FWI with intermediate data shifting of the predicted data towards the observed data. (c) The same as (b) but the intermediate data are generated by shifting the observed data towards the predicted data. (d) The recovered P-wave velocity model with conventional FWI starting from the velocity model shown in (b).

Figure 18. (a) Velocity profile at a distance of 6.25 km. (b) Normalized L2-norm model errors. The black solid curves indicate the true velocity model while the dashed black curves are for the initial velocity model. The red and blue curves are for the recovered velocity model shown in Figures 17b and c, respectively.

Figure B-1. Comparison of one shot gather generated from the surface-to-horizontal-well geometry with different wave equations from the Marmousi models. (a) Using the elastic wave equation with the P-wave and S-wave velocities and density models. (b) Using the acoustic wave equation with the P-wave velocity and density models. (c) Using the acoustic wave equation with the P-wave velocity model. The trace comparison at trace (d) 600, (e) 700 and (f) (800). Blue, red and black curves represent the traces from (a), (b) and (c), respectively.

Figure B-2. Comparison of one shot gather generated from the surface geometry with different wave equations from the Marmousi models. (a) Using the elastic wave equation with the P-wave and S-wave velocities and density models. (b) Using the acoustic wave equation with the P-wave velocity and density models. (c) Using the acoustic wave equation with the P-wave velocity model. The trace comparison at trace (d) 600, (e) 700 and (f) (800). Blue, red and black curves represent the traces from (a), (b) and (c), respectively.

Figure B-3. Gradients of FWI with the intermediate data for the surface-to-horizontal-well geometry at different iterations. The green dotted lines indicate the sea bottom in the Marmousi model.

Figure B-4. Recovered P-wave velocity model using FWI with the intermediate data for the surface-to-horizontal-well geometry at different iterations.

Figure B-5. Gradients of FWI with the intermediate data for the surface geometry at different iterations. The green dotted lines indicate the sea bottom in the Marmousi model.

Figure B-6. Recovered P-wave velocity model using FWI with the intermediate data for the surface geometry at different iterations.

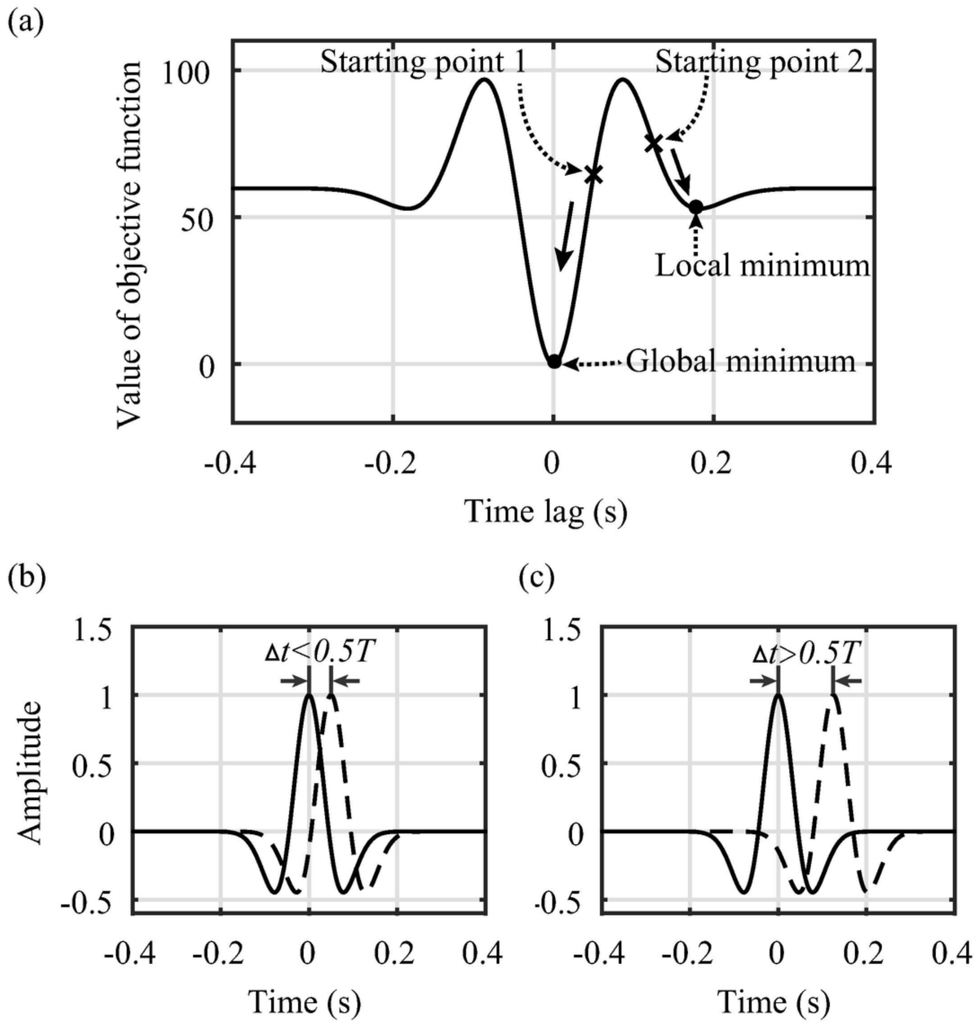

Figure 1. Schematic illustration of cycle-skipping. (a) The value of the objective function, which is the square of the L2-norm of the difference between a 5-Hz Ricker wavelet and its shifts. The two black dots represent the global minimum and a local minimum. The nearest peaks to the global minimum has a time lag of 86 ms, which corresponds to half a cycle, . (b) The 5-Hz Ricker wavelet and its shift by 50 ms, which is less than half a cycle *0.5T*. (c) The same as (b) but the shift is 125 ms, which is greater than half a cycle. The functional value for the case of (b) and (c) is indicated by the two crosses in (a).

94x96mm (300 x 300 DPI)

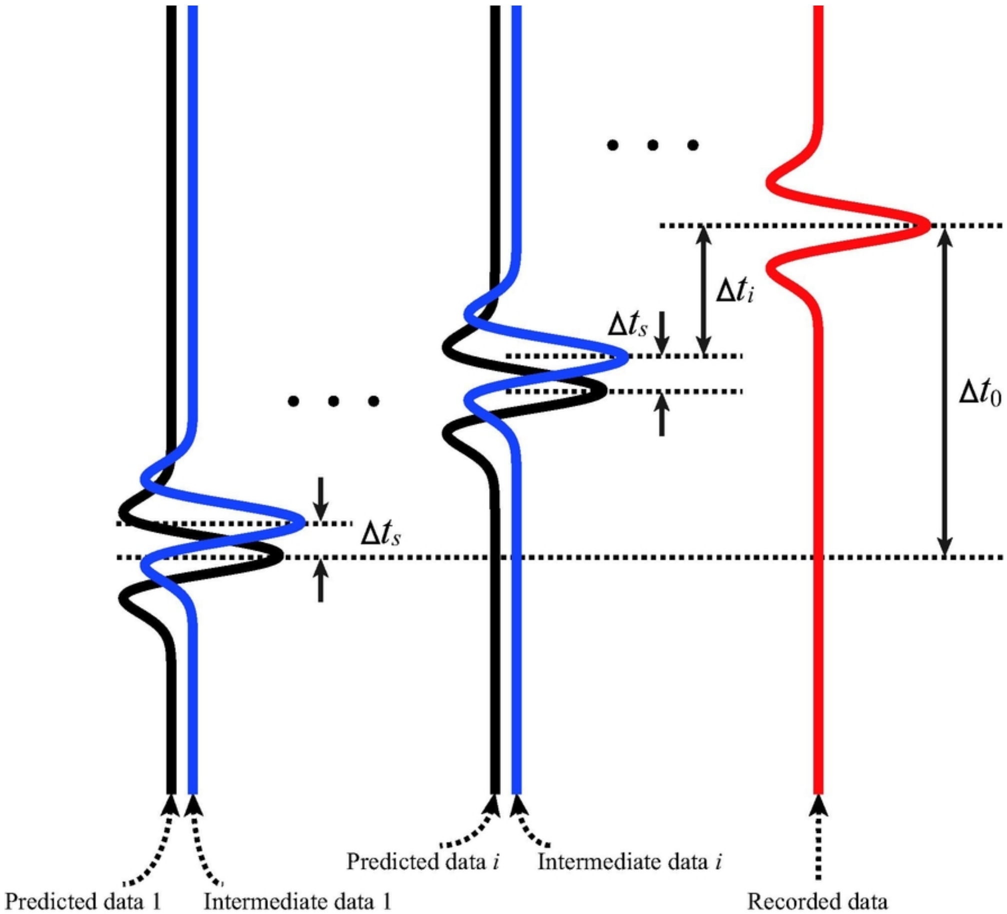

Figure 2. Schematic illustration for the method of tackling cycle-skipping with intermediate data. The predicted data 1 is generated with the initial model. The recorded data indicated by the red trace has a time difference of Δt_0 relative to predicted data 1. Δt_0 is greater than half a cycle. Shifting the predicted data by Δt_s, which is smaller than half a cycle, generates intermediate data 1. Replacing the record with intermediate data 1, FWI will update the model to make the predicted data fit intermediate data 1. Then shifting the new predicted data creates the new intermediate data. By repeating the process, the inversion will recover a model, which can produce the predicted data having a time difference to the record less than half a cycle.

75x68mm (300 x 300 DPI)

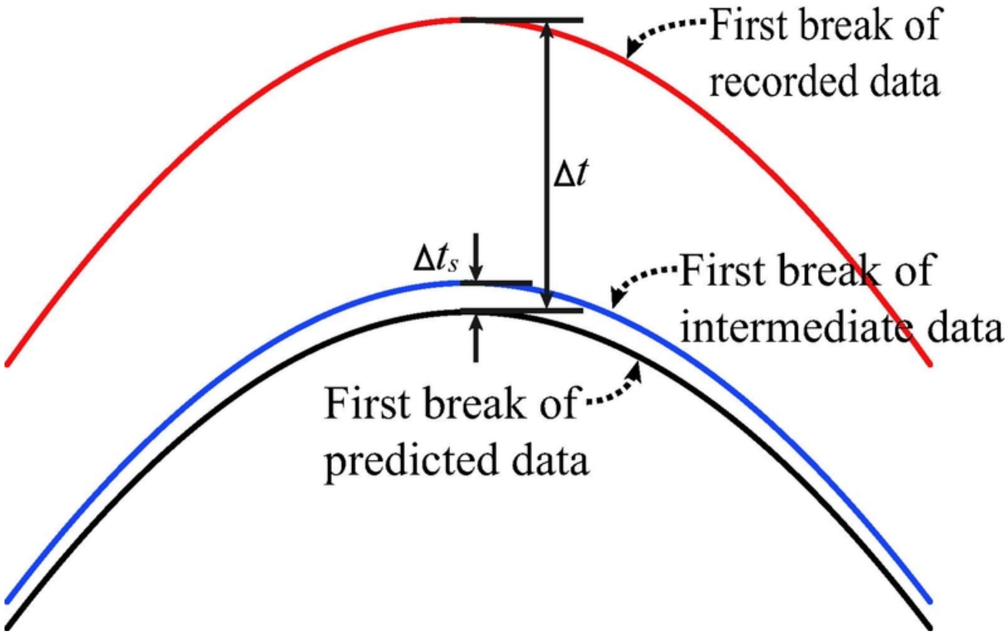

Figure 3. Schematic illustration of creating intermediate data by using the first arrivals.

67x42mm (300 x 300 DPI)

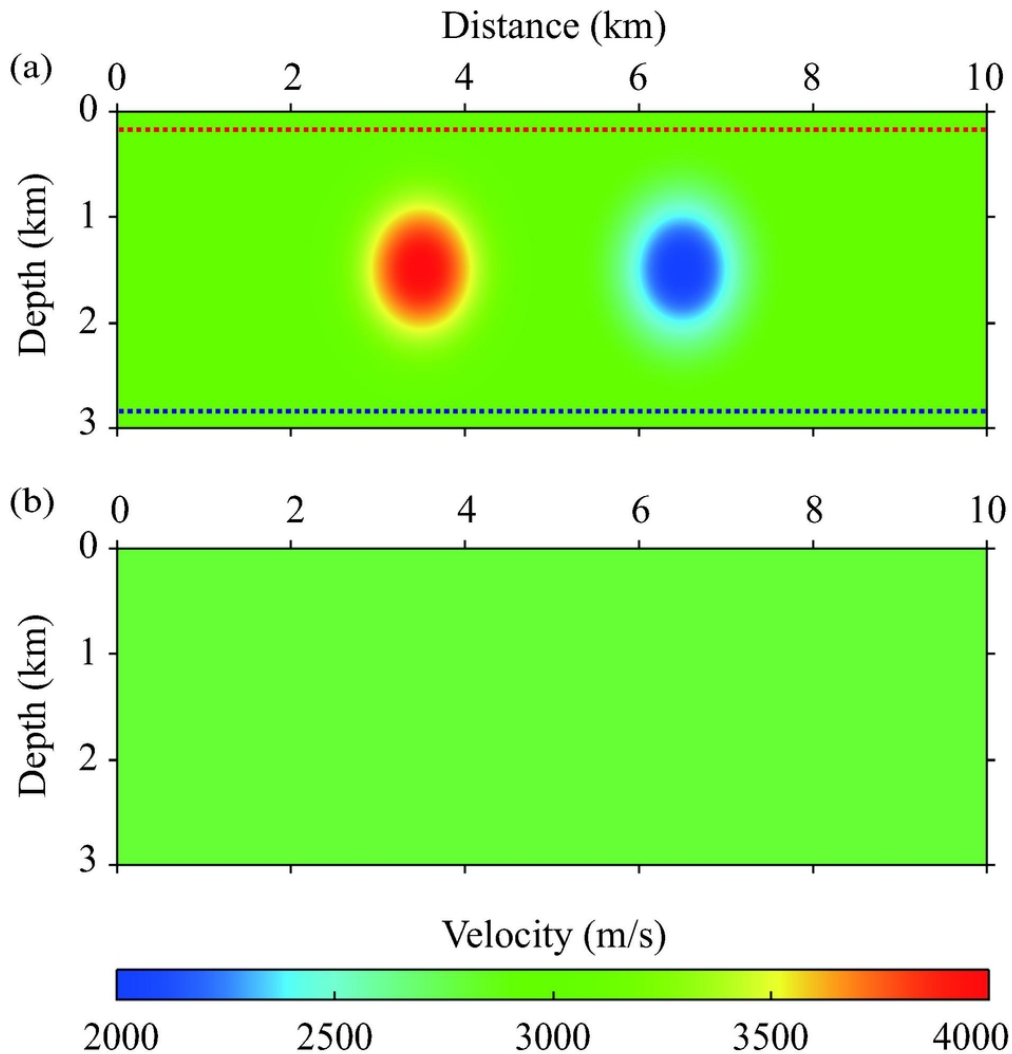

Figure 4. A model with two Gaussian anomalies. (a) The true velocity model, which includes a constant background of 3000 m/s, a high-velocity anomaly and a low-velocity anomaly. The anomalies are formed with a Gaussian function of $\pm 1000e^{-d^2/(5\times10^5)}$, where d (unit: m/s) is the distance to the center of the anomaly. The red dotted line indicates the location of sources while the blue dotted line represents the receiver arrays. (b) The initial velocity model, which is a constant velocity of 2800 m/s.

80x84mm (300 x 300 DPI)

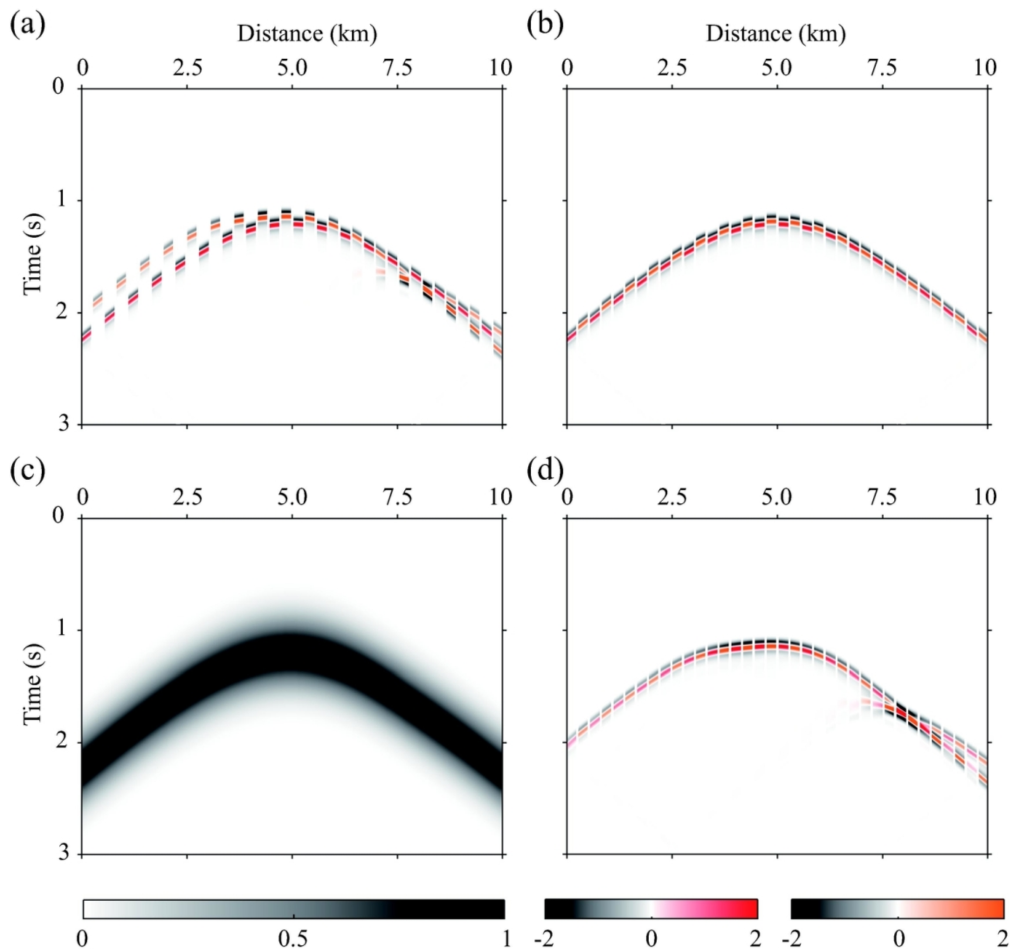

Figure 5. Overcoming cycle-skipping with intermediate data. (a) Alternating display of a shot gather of the predicted data from the initial velocity model and the record. (b) Alternating display of a shot gather of the predicted data from the initial velocity model and the intermediate data. (c) A cosine-squared weighting function used to restrict the inversion to the first arrivals only. (d) Alternating display of a shot gather of the predicted data from the recovered model and the recorded data. In all the alternating displays, the predicted data are shown first with the 'red-white-black' colormap, while the record and the intermediate data are depicted with the 'brown-white-black' colormap.

128x122mm (300 x 300 DPI)

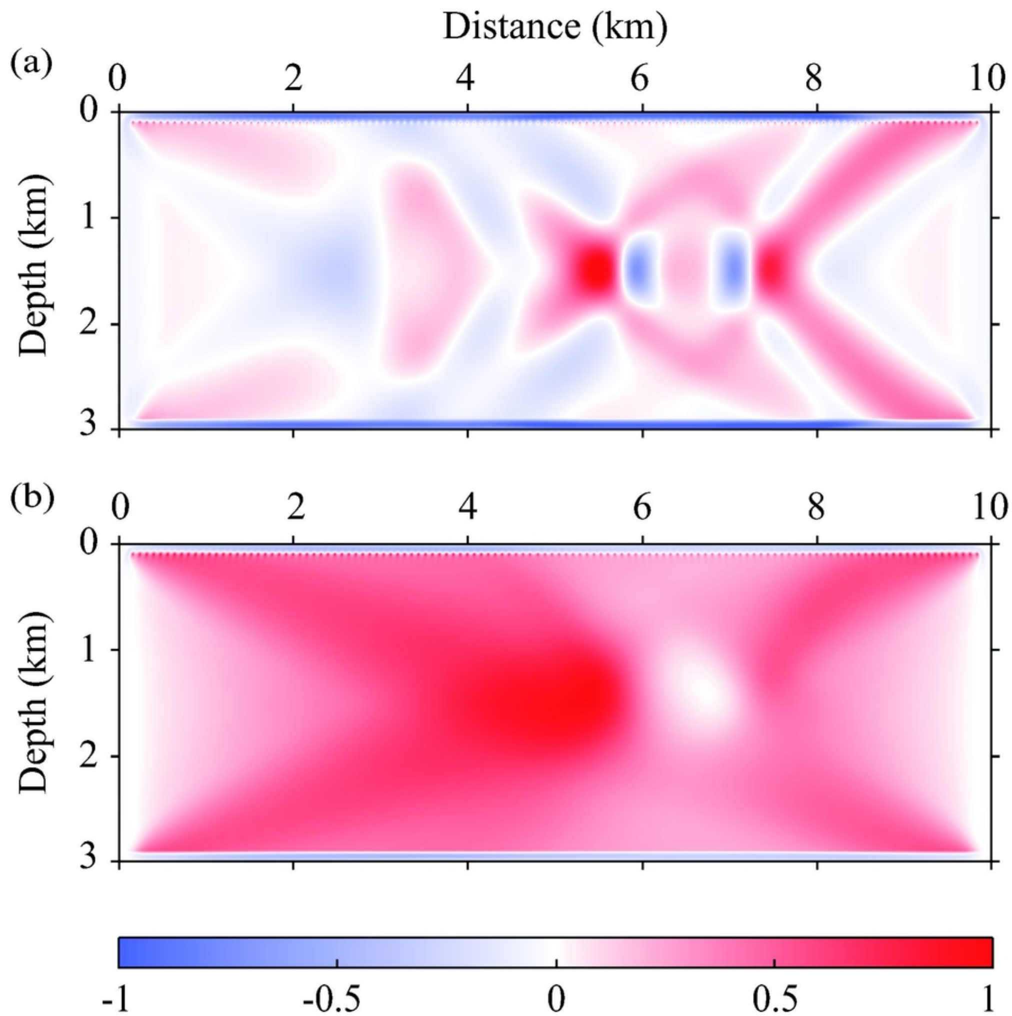

Figure 6. The gradient of (a) conventional FWI and (b) FWI with intermediate data in the first iteration. During the inversion, the model is updated through slowness, so that the negative gradient (blue color) indicates FWI will decrease velocity while the positive gradient (red color) means FWI will increase velocity.

80x81mm (300 x 300 DPI)

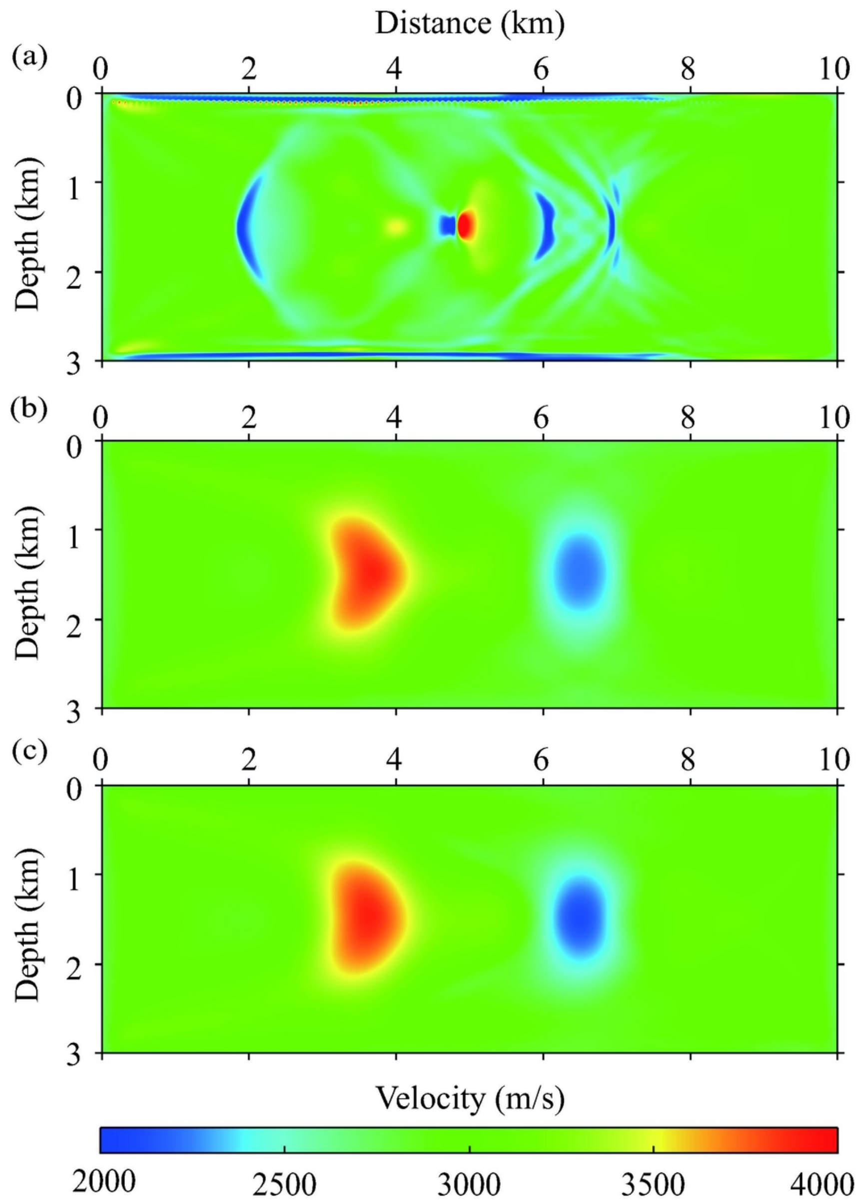

Figure 7. The recovered velocity models from (a) conventional FWI starting from the initial model, (b) FWI with intermediate data starting from the initial model, and (c) conventional FWI starting from (b).

80x113mm (300 x 300 DPI)

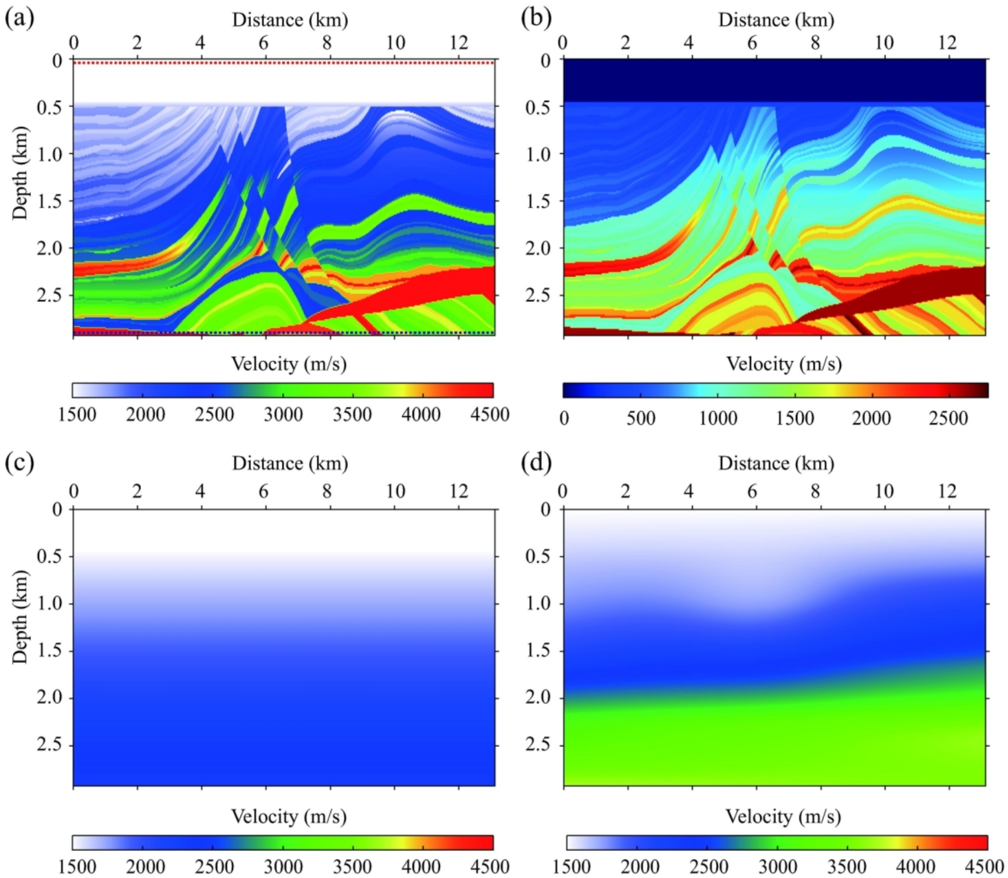

Figure 8. The Marmousi models and the initial models for inversion. (a) The true P-wave velocity model. (b) The true S-wave velocity model. (c) The simple 1D initial P-wave velocity model for the surface-to-horizontal-well test. (d) The simple initial P-wave velocity model, in which a Gaussian low-velocity anomaly is embedded around the fault area, for the test with a surface acquisition. In (a), the red dotted line indicates the locations of the sources and the receiver array of the surface acquisition while the blue dotted line represents the locations of the receiver array of the surface-to-horizontal-well setting.

155x136mm (300 x 300 DPI)

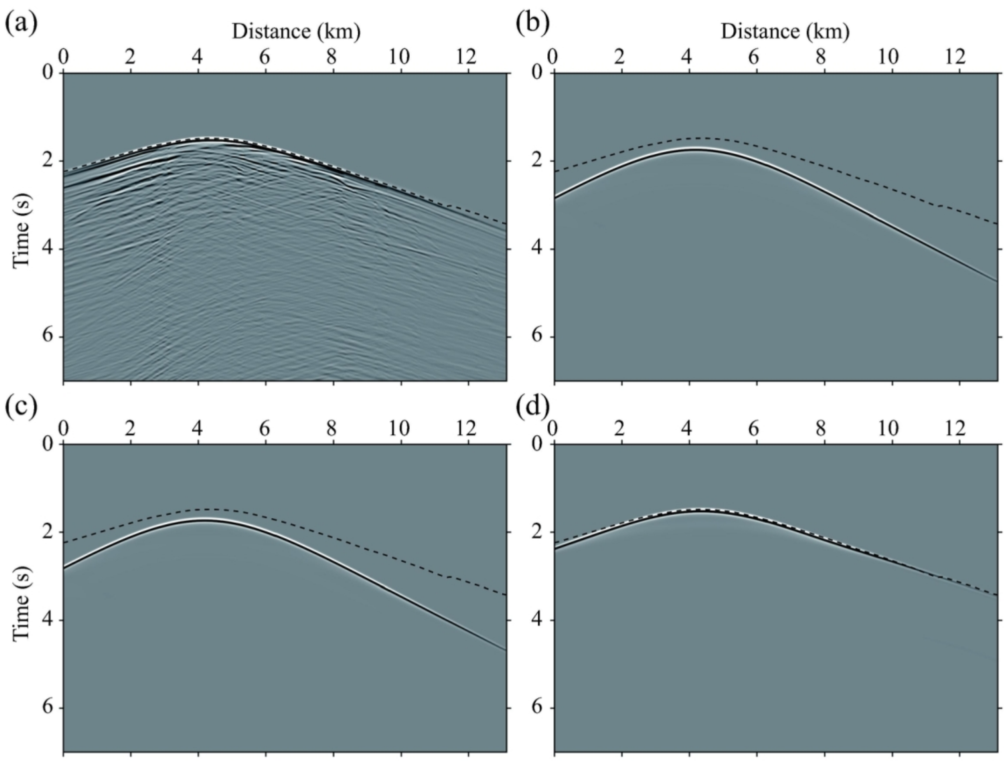

Figure 9. One shot gather in FWI with intermediate data for the surface-to-horizontal-well test. (a) Recorded data. (b) Predicted data with the initial velocity model shown in Figure 8c. (c) The intermediate data generated by shifting the predicted data shown in (b) within half a cycle. (d) The predicted data with the recovered model shown in Figure 12b. The dashed black curves indicate the first-break picks of the record shown in (a).

139x105mm (300 x 300 DPI)

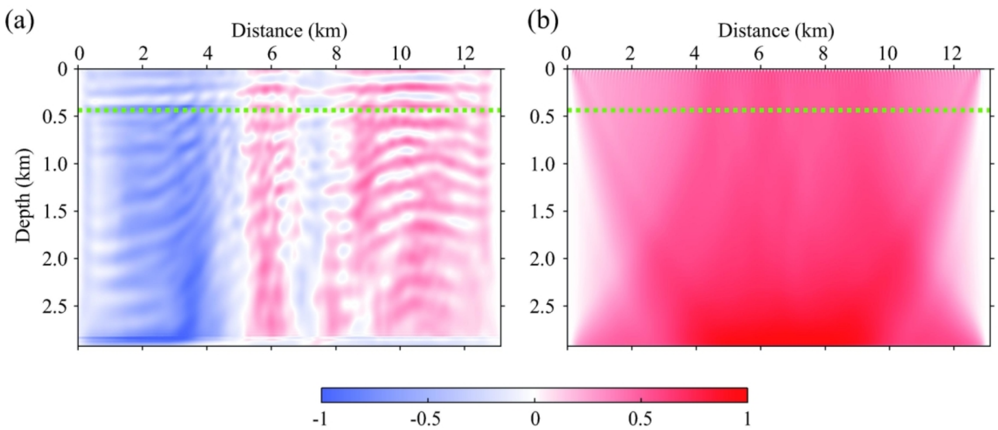

Figure 10. The gradient of (a) conventional FWI and (b) FWI with intermediate data in the first iteration. The green dotted lines indicate the sea bottom in the Marmousi model.

155x66mm (300 x 300 DPI)

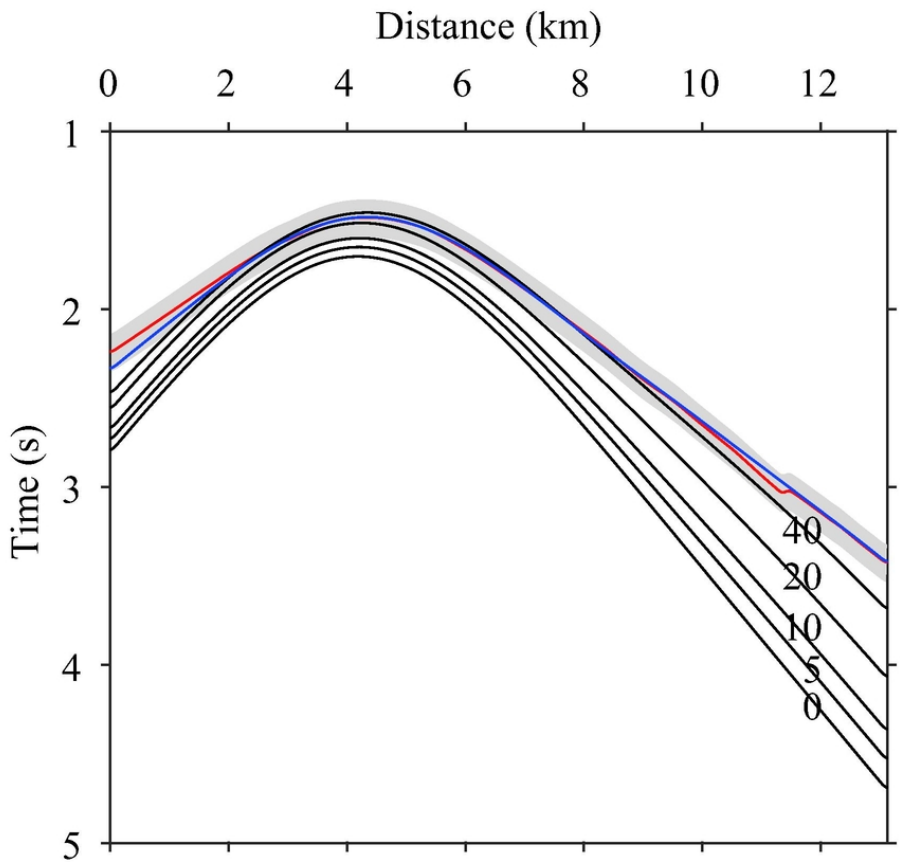

Figure 11. Evolution of first breaks throughout the iterations. The red curve indicates the first break of the recorded data while the blue curve represents the first break of the predicted data after 60 iterations. The black curves denote the first breaks for early iterations. The iteration numbers are drawn on top of these curves. The light gray strip around the red curve indicates the half period of the 5-Hz Ricker wavelet.

80x76mm (300 x 300 DPI)

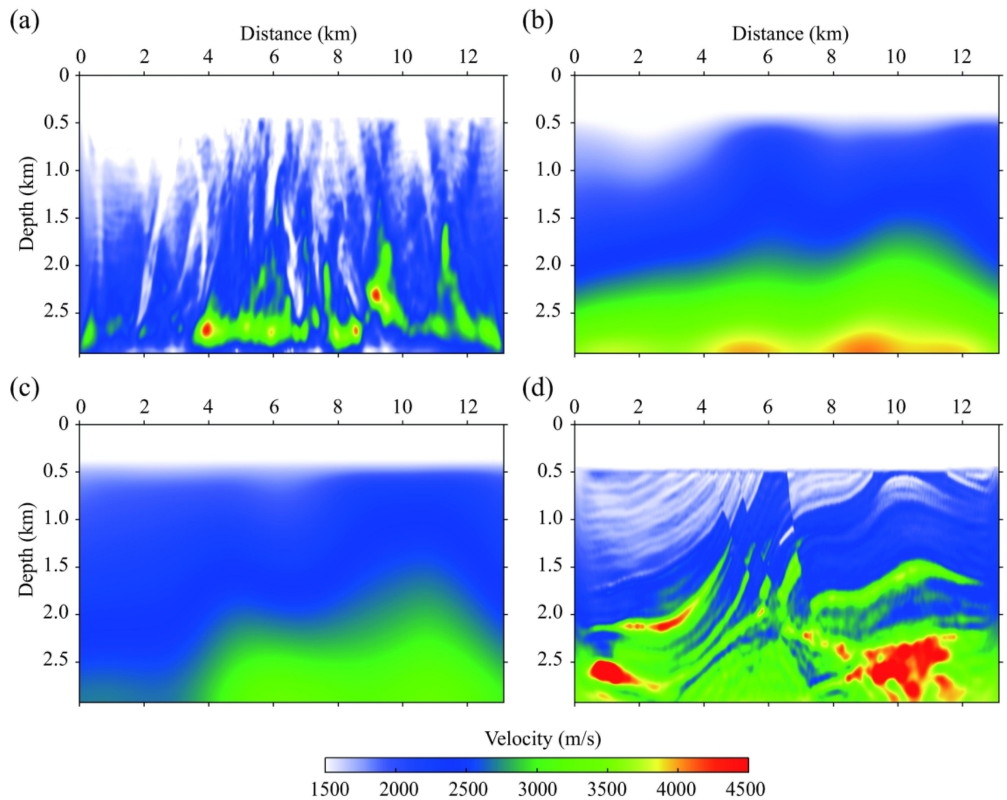

Figure 12. Recovered P-wave velocity models. (a) The recovered P-wave velocity model by using conventional FWI starting from the initial model shown in Figure 8c. (b) The recovered P-wave velocity model from the initial velocity model shown in Figure 8c by using FWI with intermediate data shifting of the predicted data towards the observed data. (c) The same as (b) but the intermediate data are generated by shifting the observed data towards the predicted data. (d) The recovered P-wave velocity with the conventional FWI starting from the velocity shown in (b).

154x123mm (300 x 300 DPI)

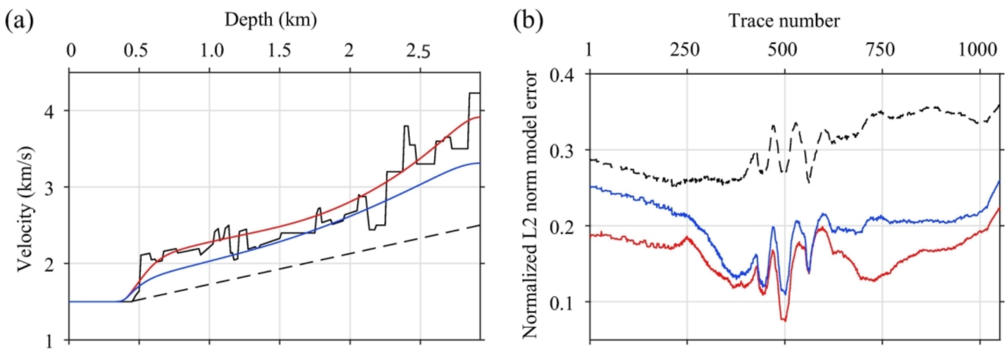

Figure 13. (a) Velocity profile at a distance of 6.25 km. (b) Normalized L2-norm model errors. The black solid curves indicate the true velocity model while the dashed black curves are for the initial velocity model. The red and blue curves are for the recovered velocity model shown in Figures 12b and 12c, respectively.

158x55mm (300 x 300 DPI)

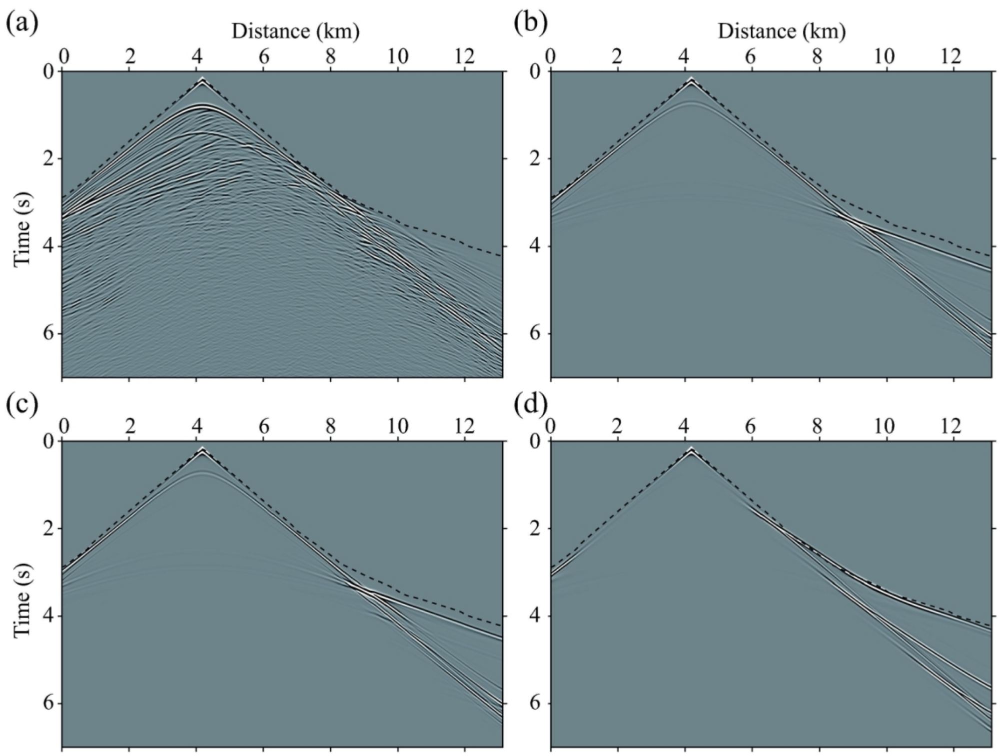

Figure 14. One shot gather in FWI with intermediate data for the test of the surface acquisition geometry. (a) Recorded data. (b) Predicted data with the initial velocity model shown in Figure 8d. (c) The intermediate data generated by shifting the predicted data shown in (b) within half a cycle. (d) The predicted data with the recovered model shown in Figure 17b. The dashed black curves indicate the first-break picks of the record shown in (a).

140x105mm (300 x 300 DPI)

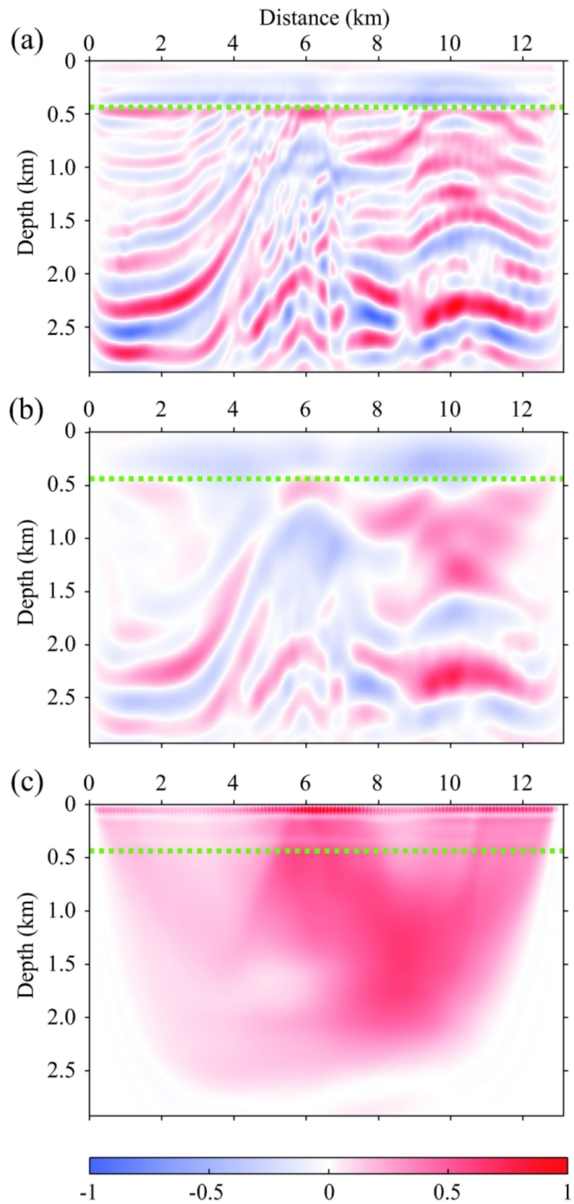

Figure 15. The gradient of (a) conventional FWI and (c) FWI with the intermediate data in the first iteration. (b) A smoothed version of (a) using a Gaussian filter, i.e. $1/(2\pi\cdot\sigma)e^{(x2+z2)/2\sigma2}$, where $\sigma=10$, and x and z are in number of cells. The green dotted lines indicate the sea bottom in the Marmousi model.

78x165mm (300 x 300 DPI)

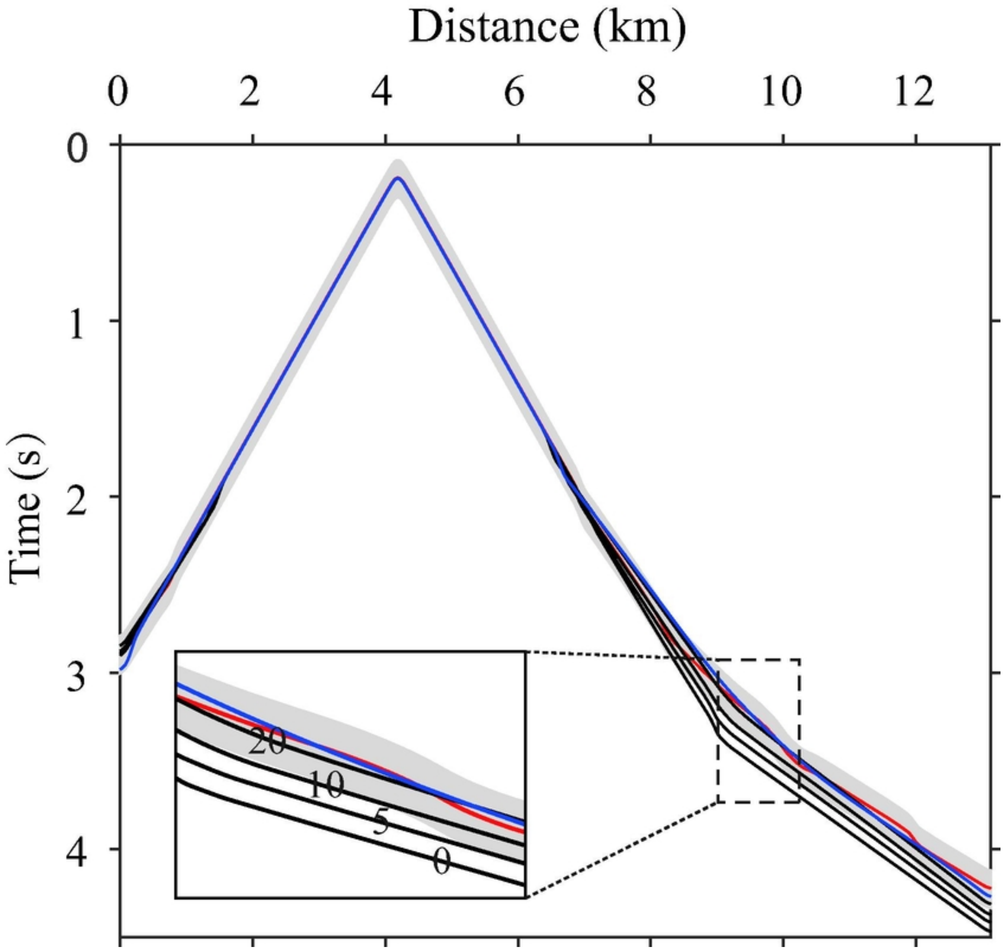

Figure 16. Evolution of the first breaks throughout the iterations. The red curve indicates the first break of the recorded data while the blue curve represents the first break of the predicted data after 60 iterations. The black curves denote the first breaks for early iterations. The iteration numbers are drawn on top of these curves. The light gray strip around the red curve indicates the half period of the 5-Hz Ricker wavelet.

71x67mm (300 x 300 DPI)

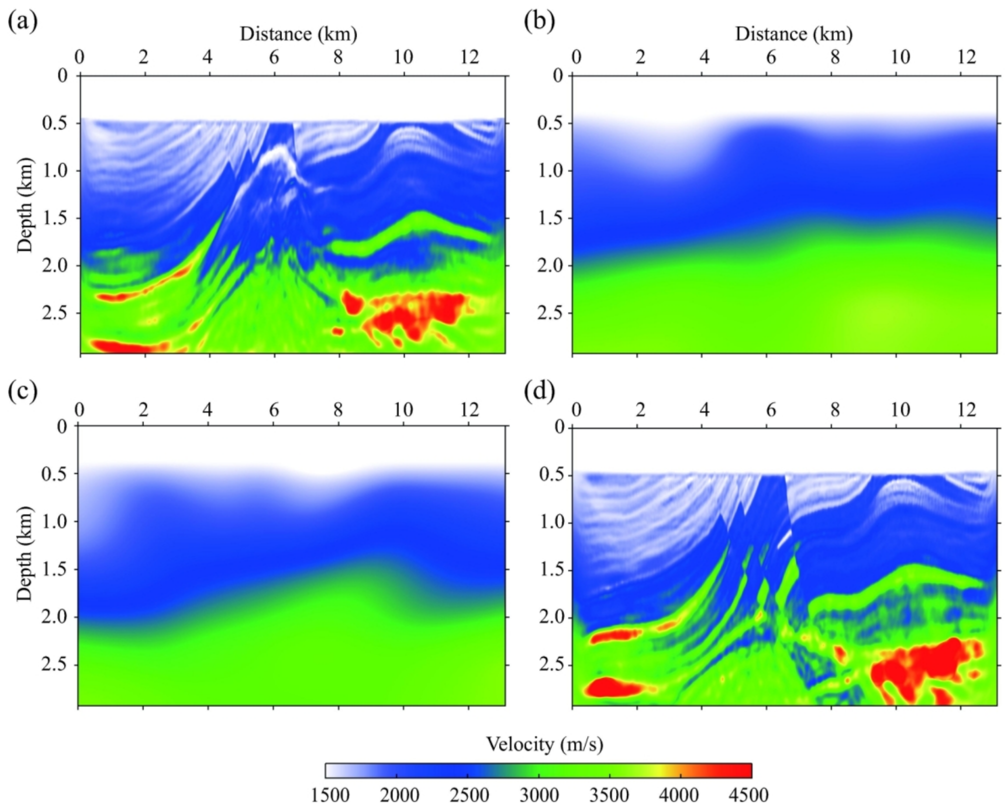

Figure 17. Recovered P-wave velocity models. (a) The recovered P-wave velocity model by using conventional FWI starting from the initial model shown in Figure 8d. (b) The recovered P-wave velocity model from the initial velocity model shown in Figure 8d by using FWI with intermediate data shifting of the predicted data towards the observed data. (c) The same as (b) but the intermediate data are generated by shifting the observed data towards the predicted data. (d) The recovered P-wave velocity model with conventional FWI starting from the velocity model shown in (b).

153x123mm (300 x 300 DPI)

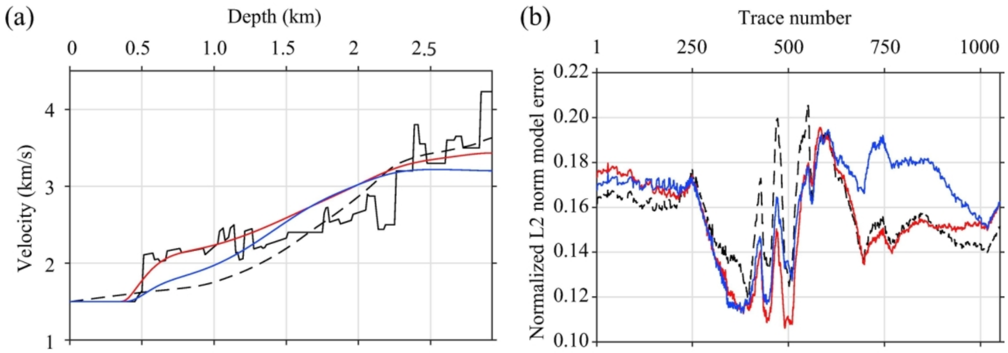

Figure 18. (a) Velocity profile at a distance of 6.25 km. (b) Normalized L2-norm model errors. The black solid curves indicate the true velocity model while the dashed black curves are for the initial velocity model. The red and blue curves are for the recovered velocity model shown in Figures 17b and c, respectively.

156x64mm (300 x 300 DPI)

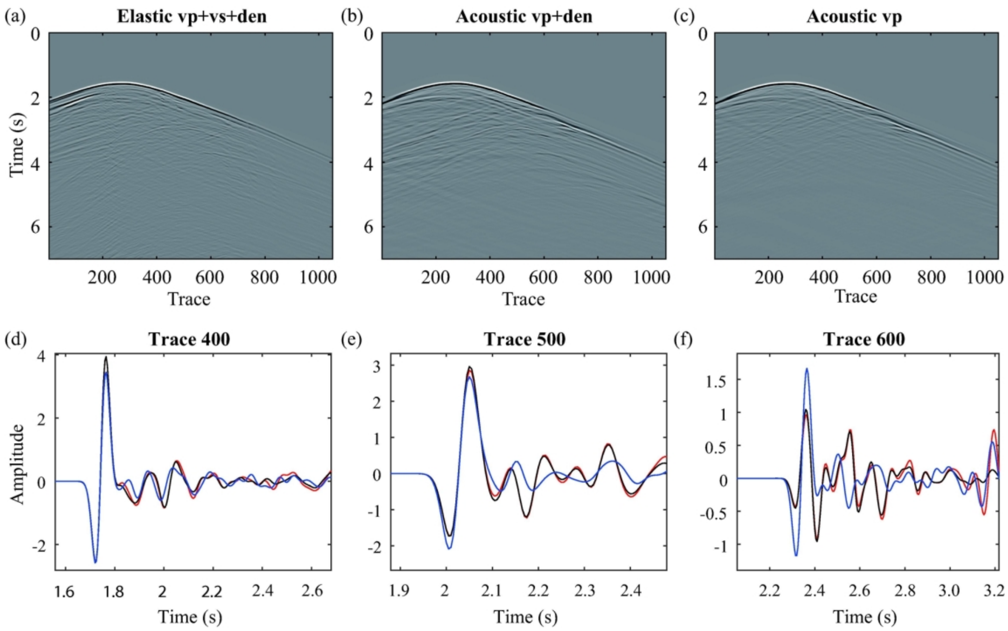

Figure B-1. Comparison of one shot gather generated from the surface-to-horizontal-well geometry with different wave equations from the Marmousi models. (a) Using the elastic wave equation with the P-wave and S-wave velocities and density models. (b) Using the acoustic wave equation with the P-wave velocity and density models. (c) Using the acoustic wave equation with the P-wave velocity model. The trace comparison at trace (d) 600, (e) 700 and (f) (800). Blue, red and black curves represent the traces from (a), (b) and (c), respectively.

163x108mm (300 x 300 DPI)

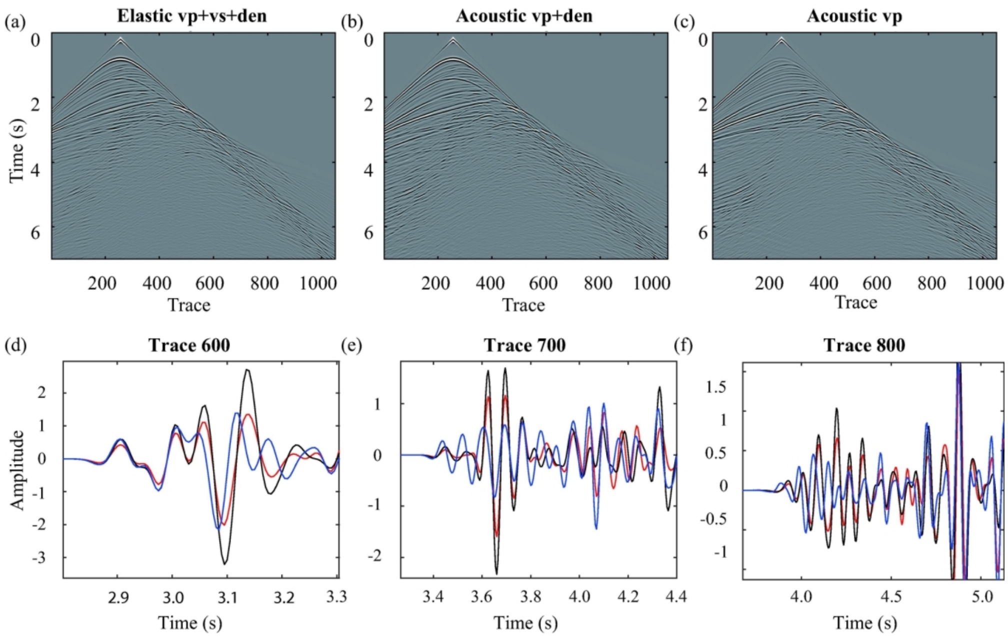

Figure B-2. Comparison of one shot gather generated from the surface geometry with different wave equations from the Marmousi models. (a) Using the elastic wave equation with the P-wave and S-wave velocities and density models. (b) Using the acoustic wave equation with the P-wave velocity and density models. (c) Using the acoustic wave equation with the P-wave velocity model. The trace comparison at trace (d) 600, (e) 700 and (f) (800). Blue, red and black curves represent the traces from (a), (b) and (c), respectively.

163x109mm (300 x 300 DPI)

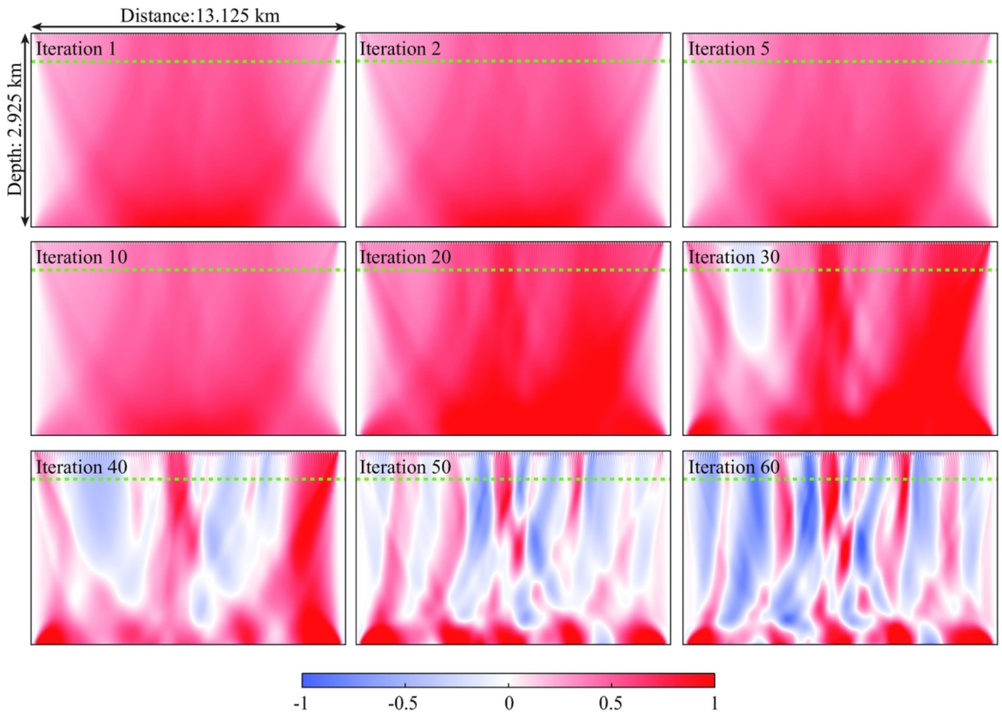

Figure B-3. Gradients of FWI with the intermediate data for the surface-to-horizontal-well geometry at different iterations. The green dotted lines indicate the sea bottom in the Marmousi model.

160x113mm (300 x 300 DPI)

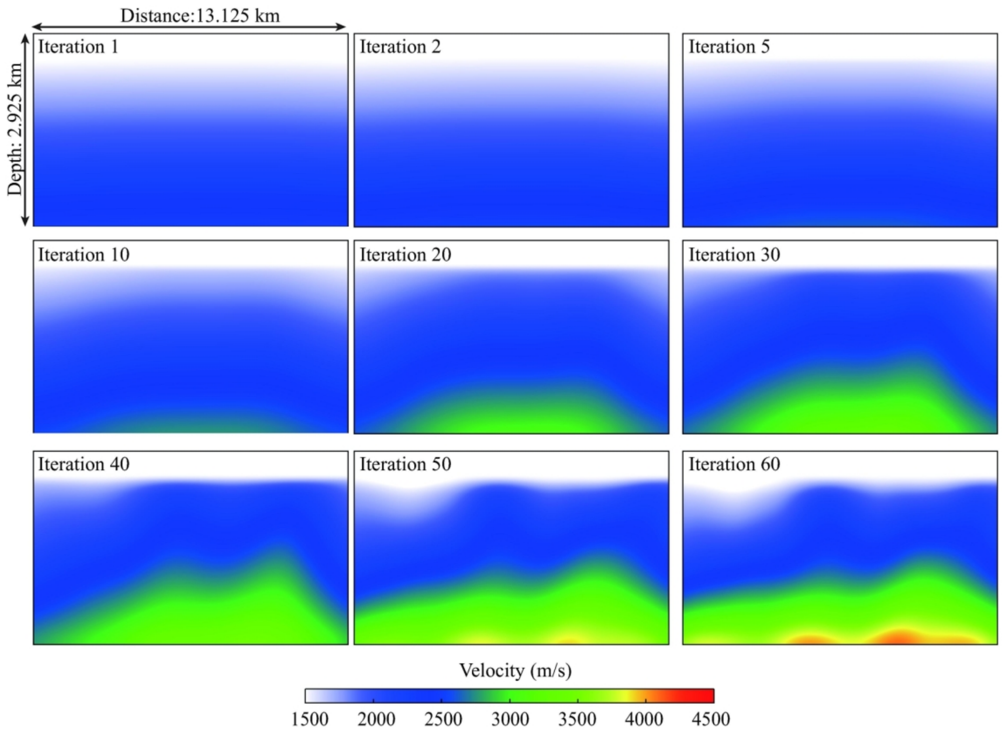

Figure B-4. Recovered P-wave velocity model using FWI with the intermediate data for the surface-to-horizontal-well geometry at different iterations.

160x116mm (300 x 300 DPI)

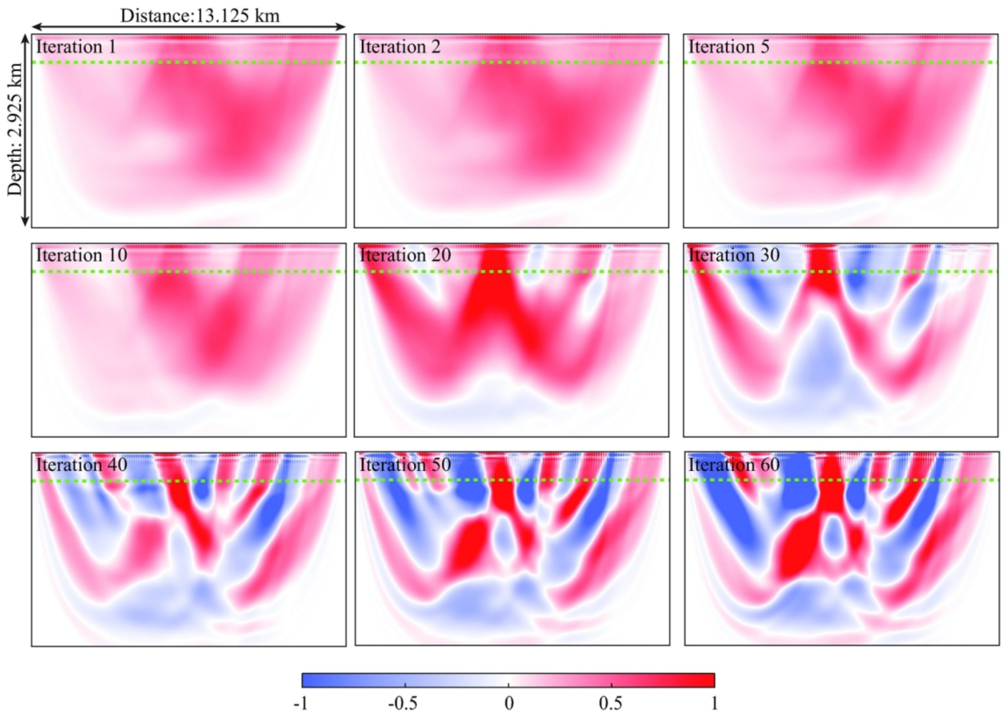

Figure B-5. Gradients of FWI with the intermediate data for the surface geometry at different iterations. The green dotted lines indicate the sea bottom in the Marmousi model.

160x113mm (300 x 300 DPI)

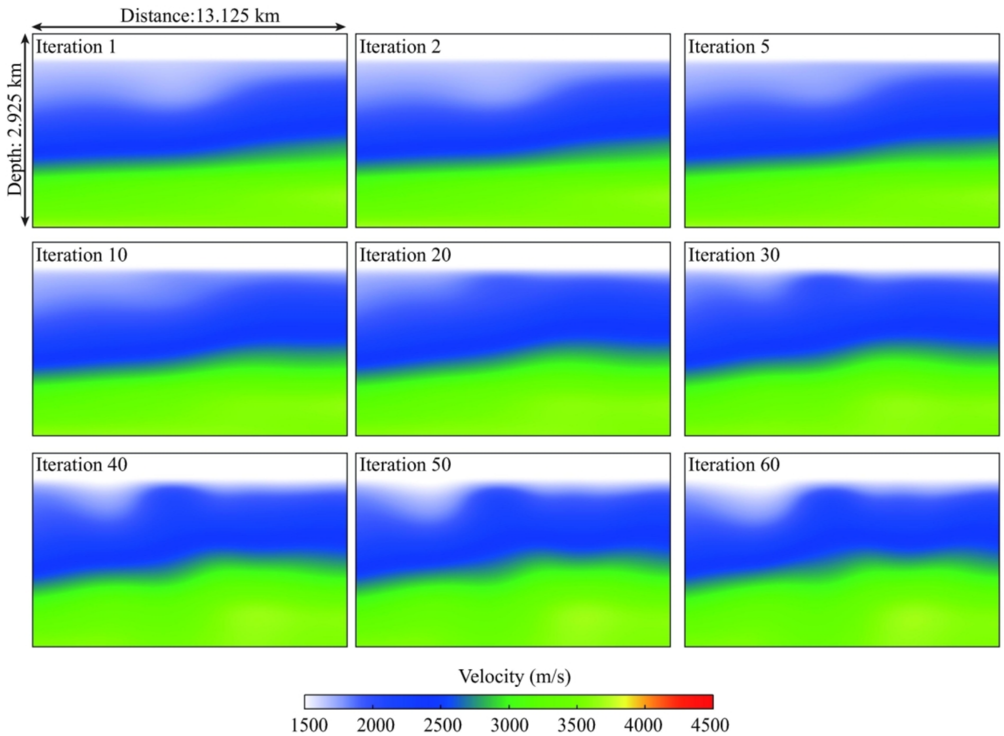

Figure B-6. Recovered P-wave velocity model using FWI with the intermediate data for the surface geometry at different iterations.

160x117mm (300 x 300 DPI)

DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.