A robust waveform inversion using a global comparison of modeled and observed data

Bingbing Sun and Tariq Alkhalifah

bingbing.sun@kaust.edu.sa,

tariq.alkhalifah@kaust.edu.sa

Physical Sciences and Engineering

King Abdullah University of Science and Technology

Thuwal, 23955, Saudi Arabia

ABSTRACT

A high-resolution model of the subsurface is the product of a successful Full-Waveform Inversion (FWI) application. However, this optimization problem is highly nonlinear, and thus, we iteratively update the subsurface model by minimizing a misfit function that measures the difference between observed and modeled data. The often celebrated L2 norm misfit function provides a simple, sample-by sample comparison between the observed and modeled data. However, it is susceptible to local minima in the objective function and it can converge to it if the low wavenumber components of the initial model are not accurate enough. In this article, we review an alternative formulation of full-waveform inversion based on a more global comparison. A combination of Radon transforms and utilizing a matching filter allows for comparisons beyond the sample to sample or even the trace to trace. We combine two of our recent developments to suggest the following algorithm for optimal inversion: 1) We compute the matching filter between the observed and modeled data in the Radon domain, which helps reduce the crosstalk introduced in the deconvolution step of computing the matching filter. 2) We use Wasserstein distance to measure the distance between the resulting matching filter in the radon domain and a representation of the Dirac delta function, which provides us with the optimal transport between two distribution functions. We use a modified Marmousi model to show how this Radon-domain optimal-transport based matching-filter approach can mitigate "cycle-skipping". Starting from a rather simplified v(z) media as the initial model, the proposed method is able to invert for the Marmousi model with considerable accuracy while standard L2-norm formulation is trapped in a meaningless, local minimum. The application of the proposed method to an offshore dataset further demonstrate its robustness and effectiveness.

INTRODUCTION

Full waveform inversion (FWI) (Virieux and Operto, 2009) aims to use the whole recorded waveform signal, and all seismic events (diving waves, pre-and post-critical reflections, and converted waves) to invert for the physical properties that produced and characterized these events, such as P- and S-wave velocities, attenuation, density and anisotropy parameters. FWI normally yields high-resolution estimates of the medium compared to tomographic methods, up to the theoretical limit of half of the shortest wavelength of the recorded signal. Although the basic concept and theory of FWI in the time domain was introduced and developed by Lailly (1983) and Tarantola (1984), there are variations of the FWI formulation such as frequency domain FWI (Pratt, 1999; Sirgue and Pratt, 2004), Laplace domain and Laplace-Fourier domain FWI (Shin and Cha, 2008, 2009). Due to the relatively high computational cost of FWI, only recently, and courtesy of the significant developments in high-performance computing, we managed to apply it to real data in a more consistent fashion. There are numerous reported successful applications to real data (Operto et al., 2006; Plessix and Perkins, 2010; Sirgue et al., 2010; Warner et al., 2013; Vigh et al., 2014; Shen et al., 2018; Alkhalifah et al.,2018a).

However, there is a dark side to FWI. The model we get or converge to might not be an accurate one. This can be attributed to limitations of the physics considered in our modeling as we ignore anisotropy, attenuation, and other phenomena. However, most of the time, the inaccuracy in the inverted model is due to falling a local minima for parts of our model domain. Designing a misfit function is an important component of FWI: a well-behaved misfit function would relax the requirement for a good initial velocity model or usable low-frequency signals in the data and

resolve the so-called “local minima” problem. In typical implementations, we either use a kinematically accurate initial models or start the inversion from a relatively low frequency. Current velocity model building technologies rely highly on tomography (Woodward et al., 2008). In some circumstances, where the subsurface is complex or has large velocity contrasts, like salt bodies (Wang et al., 2008), the low-resolution model output from tomography is insufficient and can hardly meet the requirement for FWI to avoid "local minima". Due to the low signal-to-noise ratio (SNR) and acquisition limitations, obtaining ultra-low frequency seismic signals tend to be difficult. Currently, a practical, cost-effective solution to "cycle skipping" in FWI is designing a misfit function that is reasonably convex and allows the model to gradually converge to the exact solution from a far away, "cycle-skipped", initial model. Over the past decade, the least square L2 norm was widely used as a misfit function for its simplicity and its potential for high-resolution models, but it suffers from "cycle skipping." One potential solution to "cycle skipping" issue in FWI is to use global optimization schemes such as Monte Carlo (Jin and Madariaga, 1994; Sambridge and Mosegaard, 2002), genetic algorithms (Sen and Stoffa,1992; Jin and Madariaga, 1993), simulating annealing (Kirkpatrick et al., 1983; Datta and Sen, 2016). However, Global optimization schemes typically require an evaluation of the misfit function for tens of thousands of models, considering the large model space we tend to work within seismic applications. For real applications, especially in 3D, where we have hundreds of millions of unknown parameters, these strategies become impractical considering the current computational capabilities. Thus, current FWI methods rely on local optimization schemes and update the model along the descent directions. In addition to the data domain misfit functions, image domain misfit functions, based on the concept of "Differential Semblance Optimization" (Symes and Carazzone, 1991), have provided potential solutions for the cycle skipping dilemma. In these methods, we assume that the

velocity can be separated into low wavenumber background and high wavenumber reflectivity components. Based on the extended imaging condition, an extended image (reflectivity) can be computed and used to evaluate the accuracy of the background velocity model. The misfit function can be defined to measure the focusing of energy in subsurface offsets (Shen and Symes, 2008; Sun and Alkhalifah, 2018b) to zero offset, time lags (Sava and Fomel, 2006) to zero lag, or measuring the coherency in angle domain gathers (Biondi and Symes, 2004). Although this category of methods shows more robustness in dealing with "cycle skipping", the computation of extended images is costly. In addition, these methods assume that only primary reflections are used to generate the images, and thus, considerable preprocessing is needed to meet this requirement.

Recently, new and more advanced misfit functions were proposed, such as a matching-filter misfit function (Van Leeuwen and Mulder, 2008, 2010; Luo and Sava, 2011; Warner and Guasch, 2016) and the optimal transport misfit function (Engquist and Froese,2014; Engquist et al., 2016; Metivier et al., 2016b; Yang et al., 2018; Yang and Engquist, 2018; Qiu et al., 2017). Those newly proposed methods transform the local, sample-by-sample comparison, to a global one, trace by trace. As expected, the resulting misfit function shows more convex behavior and can reasonably mitigate the "local-minima" issue. The idea behind the matching filter approach is that instead of comparing the seismic trace sample by sample through for example subtraction in the L2 norm, we first compute a matching filter by deconvolving the measured data from the modeled ones. If the velocity used in modeling is correct, the resulting matching filter would focus energy mainly at zero lag, providing a band limited Dirac delta function. Otherwise, by penalizing the coefficients at non zero-time lag, we can formulate an optimization problem (Luo and Sava, 2011; Huang et al., 2017; Warner and Guasch, 2016). In principle, all those approaches try to measure the departure of the matching filter from a Dirac delta function. In their approaches, the misfit function

uses deconvolution instead of cross-correlation (Van Leeuwen and Mulder, 2008, 2010), and thus, the objective function is less sensitive to a band limited or non-impulsive source function, which may result in a nonzero gradient of the objective function even when the inverted velocity model matches the true model. Warner and Guasch (2016) introduced an extra normalization term, and in their study, the normalization term plays an important role in providing better convexity for the misfit function and accelerating its convergence. As explained by Warner and Guasch (2016), without some form of normalization, the misfit function could be minimized simply by using zero predicted data. This is not a behavior that will help us move toward the true earth model, and thus, the normalization of the objective function is always necessary in some form to obtain a practical AWI algorithm. This normalization term has its roots in the instantaneous travel time objective function (Alkhalifah and Choi, 2014) as demonstrated by Sun and Alkhalifah (2018c). As shown in Sun and Alkhalifah (2018a), the normalization term will occur naturally, if we try to measure the Wasserstein distance between the resulting matching filter and the Dirac delta function. The normalization term is a strict requirement for the comparison of two distributions, specially to satisfy the mass conversation rule. As long as we include the normalization by preconditioning the resulted matching filter, we form an optimization that focuses the resulting matching filter to a Dirac delta function. A simple penalty term although appears to try to focus the matching filter in the same way, but it can also actually reduce the amplitude of the matching filter rather than push it to be Dirac delta function.

All these correlations and deconvolution-based filters extend the data comparison along the whole trace. However, though needed to avoid cycle skipping, the global comparison can introduce crosstalk between events in the measured and modeled data that are not related. Thus, such crosstalk in the global comparison can cause drawbacks in the application. For example, Huang et

al. (2017) reported that the matching-filter approach deals mainly with diving waves and surfers in the presence of strong multiples. Thus, we propose to introduce to the general workflow for FWI a Radon transform of the data prior to computing the matching filter. Radon transform is a mathematical technique that has been widely used in seismic data processing (Stoffa et al., 1981; Thorson and Claerbout, 1985; Sacchi and Ulrych, 1995; Luo et al., 2009; Gholami and Zand, 2018). A number of researchers have worked on the plane wave and Radon domain related waveform inversion (Diebold and Stoffa, 1981; Pan et al., 1988; Kormendi and Dietrich, 1991; Stoffa and Sen, 1991; Minkoff and Symes, 1997; Vigh and Starr, 2008; Tao and Sen, 2013; Pan et al., 2015; Kwon et al., 2016). Those approaches try to utilize the Radon transform to improve the SNR of the seismic data and the efficient encoding of the plane wave record for accelerating the computation. Here, we expand the application of Radon transform to FWI for designing a robust misfit function by integrating with the matching filter approach. For the Radon-domain matching-filter FWI, instead of computing the matching -filter for each trace directly, we first transform a common-shot gather into the tau-p domain using Radon transform and then compute the matching filter for each trace indexed by the same p-value. Thus, instead of designing the matching filter by applying the matching on a single trace, we apply the matching on traces that have been stacked over slopes and represented in the tau-p domain, and consequently, we intuitively take the coherency of the events into consideration. The criterion for a successful convergence is still the same in the p domain, i.e., the energy of the matching filter focusses at zero lag.

Recently, optimal-transport based misfit functions have demonstrated their potential. The idea of using the optimal transport measure of the misfit for waveform inversion was first proposed by Engquist and Froese (2014). The Wasserstein metric is a concept based on optimal transportation (Villani, 2003). To use seismic data in optimal transport, the datasets of the modeled

and measured seismic signals need to be modified and preconditioned to form density functions of probability distributions, which can be imaged as the distribution of two piles of sand with equal mass. Different plans of transportation of one collection into the other results in a different amount of energy spent in the transport. The plan with the lowest amount of energy needed in the transportation is the optimal map, and this lowest cost is defined as the Wasserstein metric. If we consider time signals, the resulting Wasserstein metric yields a convex function concerning time-shifted signals (Engquist and Froese, 2014); This property makes the Wasserstein metric a viable remedy to the limitations of the L2 norm misfit function in FWI applications.

The main issue in the implementation of optimal transport in FWI is the requirement that we need measure the distance between distributions. The density function for a probability distribution has two features: 1) its elements are non-negative and 2) and it satisfy the mass conservation rule (its integration equals one). The non-negative rule requirement is more critical. As long as the function is non-negative, we can easily divide the function by its summation to fulfill the mass conservation rule. As seismic signals are oscillatory with zero mean, they hardly meet these two requirements for distribution. Specially, for the non-negative rule, transforming the seismic signal to be positive faces the danger of destroying the polarity sensitivity (Pladys et al., 2017).

There are currently two categories of optimal-transport based misfit functions. The first category is related to the quadratic Wasserstein metric W2 (Engquist et al., 2016; Yang et al., 2018; Yang and Engquist, 2018). The W2 metric has more convex properties, and for a 1D problem, the explicit solution for the optimal map exists and helps in solving the optimal transport optimization problem (Villani, 2003). In this category, the seismic signal is modified to satisfy the requirements of a distribution. Yang and Engquist (2018) discussed different data modification strategies. Qiu et al. (2017) implemented an exponentially encoded optimal-transport norm. However, as the

seismic signal is not a distribution naturally, any precondition or normalization applied to the seismic signal directly would either alter its phase or amplitude. Metivier et al. (2018) gives a nice review of current modification strategies. They, specially, showed that the current application of such modifications to the seismic signal to satisfy the distribution requirements have their drawbacks: for example, one strategy of positive and negative value separated comparison (Engquist and Froese,2014) although ensures convexity, it is in violation of the mass conversation rule, and it is highly sensitive to the source phase rotation. Another strategy based on an affine scaling of the data (adding a relatively large value to make the signal positive) (Yang et al., 2018) will destroy the convexity with regard to time shifts.

Another category of optimal transport is related to using the Kantorovich-Rubinstein (KR) norm (Metivier et al., 2016a,b,c). KR norm is a relaxation of the Wasserstein distance, which is another optimal transport metric with the absolute value as a cost function. The advantage of the KR norm is that it does not require the data to satisfy non-negativity or mass conservation conditions. However, in this approach, unlike Wasserstein W2, there is no explicit solution for the optimal map and an optimal transportation optimization problem has to be solved. Besides, this approach also does not maintain the convexity of the quadratic Wasserstein metric with respect to time shifts (Metivier et al., 2018).

Thus, we propose a framework for misfit function designing by combining the matching filter approach with optimal transport theory resulting in a more elegant way of measuring the distance between the matching filter and the Dirac delta function. Instead of applying the optimal transport between seismic signals, we apply it between the filter that matches the observed and modeled data and the form we want this filter to be, which is a representation of the Dirac delta function. This matching filter naturally adheres to the requirements of the optimal transport theory as we

seek to transport it to a distribution given by the Dirac delta function. Since we use the Wasserstein W2 measure, unlike the dual optimal transport strategy (Metivier et al., 2016a, b, c), we inherently obtain the explicit solution for the optimal map, which makes the computation more efficient and accurate. A simple precondition would transform the classic deconvolution measured matching filter into a distribution, and since we are not modifying the modeled or measured data directly, the phase and amplitude of the original seismic signals are preserved and not altered. As we will show later, the comparison of the matching filter with the Dirac delta function maintains the convexity related to time shift.

In the following, we first review the matching-filter approach for FWI and then discuss the Radon-domain application to the matching-filter approach. After we introduce the optimal transport theory and propose a new misfit function by comparing the resulting matching filter with the Band-limited Dirac Delta function based on Wasserstein distance. In the example section, we first use a shifted wavelet to analysis the properties of the proposed misfit function as well as compare it to other misfit functions and then we use a modified Marmousi model and an offshore field data example to demonstrate the good performance of the proposed method for resolving cycle skipping.

FWI BY MATCHING FILTER

The Radon transform, with its ability to highlight coherency in the data, can potentially benefit nonlinear misfit functions including the normalized global cross-correlation, envelope, optimal transport and matching-filter approach. However, in this study we will focus on its features

in combination with the matching-filter based FWI. Thus, we start with a brief review of the matching-filter approach.

Within the framework of FWI, we use an initial or available velocity model to simulate predicted data p to compare the measured data d. The fitting criteria forms the optimization problem, in which we minimize, for example, the L2-norm objective function:

$$J = \frac{1}{2}\int [p(t) - d(t)]^2 \, \mathrm{d}t. \tag{1}$$

It is obvious that the L2-norm misfit function measures the mismatch between two traces through a sample-by-sample comparison. As a result, it is prone to cycle skipping if the samples we are comparing do not share the same half cycle of the signal. In practice, the application of the L2-norm misfit objective function either requires a kinematically accurate initial model or very low frequencies in the data so that the half-cycle can tolerate poor initial models and thus, converge to the objective, which is referred to as the global minimum.

A strong remedy to such FWI limitations is to compute a matching filter that describes the difference between the measured and predicted data over the whole trace. Thus, a matching filter w satisfies:

$$w(t) * d(t) = p(t), \tag{2}$$

where * denotes the convolution operation. Equation 2 is a linear equation and it can be solved effectively either in the time domain by deconvolution or in frequency domain by division. When the velocity model is correct, the energy in the matching filter focusses to zero lag, more like a Delta function. Thus, we can formulate an optimization problem that penalizes energy in the matching filter beyond zero lag. As a result, the objective function has the following form:

$$J = \frac{1}{2}[P(t)w(t)]^2 \mathrm{d}t, \tag{3}$$

where $P(t)$ is the penalty function, e.g., $P(t) = |t|$, will enhance energy beyond zero lag, and the objective is to minimize such enhancement.

MATCHING FILTER IN THE RADON DOMAIN

The matching filter computed using equations 2 takes the whole trace of the measured and predicted data as input and would suffer from crosstalk between different events. We use Radon transform to separate the events based on their space-time coherency, i.e., the slope information. This would lead to a more robust computation of the matching filter. For a common-shot gather, we apply Radon transform to both the modeled and observed data, and then compute a matching filter in tau-p domain:

$$\tilde{p} = \Re[p],\ \widetilde{d} = \Re[d],\ \widetilde{d} * \widetilde{w} = \tilde{p}. \tag{4}$$

Here $\Re$ represents the forward Radon transform. In applications, we use local Radon transform (Wang et al., 2010) to capture the local coherence of seismic events. After transformation, the events should be separated based on their local slope. For example, in Figure 1, the red and blue correspond to different events in the measured and predicted data. It is obvious that if we compute the matching filter per trace by deconvolution of the measured and predicted data, the red and blue events would cross-talk with each other. Alternatively, the Radon-transform panels in Figure c) and d), show the blue and red events separated due to their distinct slope values. Deconvolution of the trace indexed by the same slope value would thus avoid the cross-talks in the original time-space domain and as a result lead to a more robust inversion process.

The other question is related to how we measure the focusing of the resulting matching filter. Equation 3 uses a penalty function. As we will show later, this penalty approach has a potential

drawback: the misfit function is sensitive to the amplitude of the resulting matching filter. We can reduce the misfit value by decreasing the amplitude of the matching filter. However, the expected evolution of the matching filter should reduce its amplitude at large time lags while increasing its amplitude at near zero time lags, or, in other words, transport the energy of the filter to zero lag. As our target for the resulting matching filter is a band-limited Dirac Delta function, we can use optimal transport theory in designing such a misfit function naturally by measuring the Wasserstein distance between the resulting matching filter and the Dirac delta function.

WASSERSTEIN DISTANCE BETWEEN MATCHING FILTER ANDDIRAC DELTA FUNCTION

Conventional optimal transport approaches measure the Wasserstein distance between the predicted data *p(t)* and measured data *d(t)* directly. The difficulty in transforming the seismic data to a probability distribution limited its application (Metivier et al., 2018). We use the Wasserstein W_2 norm to measure the distance between the matching filter *w(t)* and the Dirac delta function. To fulfill the requirements of the optimal transport theory, we need transform *w(t)* to a distribution. We elect, here, to square it and normalize it as follows:

$$w'(t) = \frac{w^2(t)}{\int w^2(t)dt} = \frac{w^2}{||w||_2^2}. \quad (5)$$

Other kinds of preconditioning, like the envelope, can be used. The modification of the matching filter to probability distribution is important: Firstly, it is a strict requirement for the application of the optimal transport theory. Secondly, it makes the misfit function insensitive to the amplitude of

the matching-filter, and thus, avoids the problem faced by the conventional matching-filter misfit function of equation 3.

When the model parameters are accurate, the resulting matching filter reduces to a Dirac delta function. This implies that the Dirac delta function $\delta(t)$ is the target, thus the misfit function can be formulated as

$$J = W_2^2(w', \delta)\,. \tag{6}$$

For the Wasserstein W_2 distance, the readers can refer to Engquist et al. (2016); Yang et al. (2018); Yang and Engquist (2018). While for complete mathematical development of this proposed approach, such as the adjoint source computation, the readers can look into Sun and Alkhalifah (2018a).

MISFIT FUNCTION CONVEXITY ANALYSIS

In this section, we analyze the convexity properties of the discussed misfit functions. We will compare the least square L2 norm approach of equation 1, the conventional matching filter approach (MF) of equation 3, conventional optimal transport (OT) with precondition using the affine scaling method (Yang and Engquist, 2018), and the proposed method optimal-transport matching-filter (OTMF) approach of equation 6. For the proposed method, we use a Gaussian standard deviation of 0.008s to represent the Dirac delta function.

In first setup, as shown in Figure 3a, we set the measured data *d(t)* to a Ricker wavelet with a peak frequency of 10 Hz, while the predicted data *p(t)* is given by a time shift version of the measured data, i.e., $p(t) = d(t + \tau)$. Here τ is the time shift ranging from -0.8s to 0.8s. We show the comparison results for the different misfit functions in Figure 3b. We can observe that the L2

norm misfit function (red curve) and the optimal transport approach with an affine scaling (green curve) have local minima. As discussed by Metivier et al. (2018), a precondition of the seismic signal by the affine scaling method artificially creates mass at each point. The optimal mass transport becomes more local instead of having global exchange properties between the initial and target signals. This makes the corresponding misfit function less sensitive to the time shift, and therefore, destroy its convexity property. For the conventional matching filter approach (the blue curve) and the proposed approach (the black curve), we obtain good convexity for this setup.

In an alternative setup, as shown in Figure 4a, we will not only shift the signal but also scale the amplitude of the signal as well. Now, the predicted data are given by formula: $p(t) = e^{-2\tau} d(t+\tau)$. As the time shift τ ranges from -0.8s to 0.8s, the scaling of the signal would vary from 0.2019 ($e^{2\ 0:8}$) to 4.95 ($e^{2\ 0:8}$). Scaling plus shifting is reasonable, as the waves propagation in the subsurface experience geometric spreading among other phenomena. Thus, waves arriving early would have relatively larger amplitudes than later arrivals. The performance of the misfit functions is shown in Figure 4b. Now we can see that the conventional matching filter approach (the blue curve) loses its convexity due to its sensitivity to the amplitude, while the proposed method (the black curve) still maintains its convexity property. In Figure 4c, we show a zoom of Figure 4b and focus on the part where the residual is small. The non-convexity of the conventional matching filter approach (the blue curve) becomes clearer. Due to the sensitivity to amplitude, even when the time shift increases, e.g., from 0.6s to 0.8s, the conventional matching filter results in decreased misfit value and this, as a result, hampers the convexity property.

Through this simple, but often used, test to measure the convexity of the objective function, we observe that compared to the conventional optimal transport method with precondition by the affine scaling method, our proposed method shows better convexity. Compared with the

conventional matching filter approach, the proposed misfit function can ensure convexity, even if we scale and shift the signal, while the conventional matching filter approach loses the convexity due to its sensitivity to the amplitude.

EXAMPLES

We will test the proposed approach first on the modified Marmousi model. The second example is a marine real dataset from offshore Australia (Sun and Alkhalifah, 2018b); we will focus on the potential of the proposed method to converge to a high-resolution model.

The Marmousi model

The first example is the modified Marmousi model. As shown in Figure 5a, the true velocity model extends 2 km in depth and 8 km laterally. The initial velocity is a linearly increasing velocity with depth as shown in Figure 5b. The dataset is modeled using 80 shots with a source interval of 100m and 400 receivers with an interval of 20 m. The source is a Ricker wavelet with a 10 Hz peak frequency. We apply an absorbing boundary condition at the surface. In the inversion, we mute data below 3 Hz to verify that our proposed method is capable of overcoming the "cycle skipping" problem without low frequencies. We apply time domain inversions and utilize a low pass filter allowing frequencies up to 10Hz in the inversion. We do not sequentially increase the frequency as often implemented. Instead, we use the frequency range from 3Hz to 10Hz simultaneously in the inversion. This is a challenge for the conventional L2 misfit function as for this frequency band we face cycle skipping. We use a Gaussian function with a constant standard deviation of 0.004 s to approximate the Dirac delta function in optimal-transport matching-filter

misfit function of equation 6. We run the inversion over 200 iterations using a nonlinear conjugate gradient method.

The inverted result for the L2-norm misfit function is shown in Figure 6a. Due to cycle skipping, the result has very strong artifacts in the left part of the model and it is generally far from the true model. Figure 6c is the result by proposed Radon-domain optimal-transport matching-filter approach. Compared to the true model, the inverted model recovers well both of the low-wavenumber and the fine-scale parts of the true model although there are some artifacts and reduced accuracy at the boundary due to the limited illumination. Compared to the conventional optimal-transport matching-filter result of Figure 6b, the Radon-domain approach obviously shows fewer artifacts and higher accuracy especially in the deeper part of the model. Considering the initial velocity is far away from the true one, the proposed method can mitigate cycle-skipping and effectively converge towards the target model.

The offshore field dataset

The second example is a marine real dataset from offshore Australia. The offset range is from 160 m to 8200 m. The initial velocity model converted form RMS velocity is given in Figure 7a. We perform the inversion using the proposed misfit function with a low pass filter applied to the data equal to 3 Hz, 5 Hz, 10 Hz, 20 Hz, 40 Hz, sequentially. During the inversion, TV regularization (Esser et al., 2018; Alkhalifah et al., 2018b) is used to reduce the noise. The inverted model is shown in Figure 7b. The updated model shows consistent structures and high resolution due to the high frequency inversion and TV regularization. In the left panels of Figures 8a and 8b, we show one selected common shot gather from the initial and inverted models. We compare it with the recorded shot gather at the same location in the right panel. Clearly, the observed model reproduces the data that batter match the observed data, especially at the larger offsets where cycle

skipping usually happens, and it is evident for the initial model. Considering the initial velocity model is obtained from a crude RMS velocity, we attribute the reasonable good result to the proposed misfit function ability to handle cycle skipped data.

CONCLUSION

The measure of distance between the deconvolution-based matching filter and a Dirac delta function using optimal transport theory provides a new framework for designing misfit functions for FWI. In combination, with Radon domain representation of observed and modeled data to have a more global comparison, the resulting form of the objective function is based on a solid mathematical foundation supported by optimal transport theory. It also admits good results in the examples we shared. The current comparison between measured and modeled data using optimal transport theory, though it may provide good results in some examples, as demonstrated, it is unnatural as data are not distributions, and do not lend itself easily to the requirements of optimal transport theory. On the other hand, the matching filter, provided by a deconvolution of the measured data with the modeled one, is an ideal candidate for a distribution. The most valuable information embedded in the matching filter is the way it is distributed over time away from zero lag. The polarity information is less significant for our objective of transforming the filter to a Dirac delta function. Besides, the distribution of its energy over the time axis is consistent with the definition of a probability distribution. As a result, we can easily transform the matching filter to a distribution, and this transformation maintains the convexity related to time shifts promised by the optimal transport theory. In other words, most of the useful information in the resulting

matching filter is given by how its non-zero lag coefficients are distributed over the time axis. The preconditioning required by the optimal transport theory will not change this fact.

REFERENCES

Alkhalifah, T., and Y. Choi, 2014, From tomography to full-waveform inversion with a single objective function: Geophysics, 79, no. 2, R55-R61.

Alkhalifah, T., B. Sun, Y. Choi, F. Alonaizi, and M. Almalkl, 2018a, Sparse frequencies data inversion: an application to a near surface experiment: 80th EAGE Conference and Exhibition.

Alkhalifah, T., B. Sun, and Z. Wu, 2018b, Full model wavenumber inversion (FWMI): identifying sources of information for the elusive middle model wavenumbers: Geophysics,83, no. 6, R587-R610.

Biondi, B., and W. W. Symes, 2004, Angle domain common image gathers for migration velocity analysis by wavefield continuation imaging: Geophysics, 69, no. 5, 1283-1298.

Datta, D., and M. K. Sen, 2016, Estimating a starting model for full-waveform inversion using a global optimization method: Geophysics, 81, no. 4, R211-R223.

Diebold, J. B., and P. L. Stoffa, 1981, The traveltime equation, taup mapping, and inversion of common midpoint data: Geophysics, 46, 238-254.

Engquist, B., and B. Froese, 2014, Application of the Wasserstein metric to seismic signals: Communications in Mathematical Sciences, 9, no. 1, 79-88.

Engquist, B., B. D. Froese, and Y. Y., 2016, Optimal transport for seismic full waveform inversion: Communications in Mathematical Sciences, 14, no. 2, 79-88.

Esser, E., L. Guasch, T. van Leeuwen, A. Aravkin, and F. Herrmann, 2018, Total variation

regularization strategies in full-waveform inversion: SIAM Journal on Imaging Sciences,11, no. 1, 376-406.

Gholami, A., and T. Zand, 2018, Three-parameter radon transform based on shifted hyperbolas: Geophysics, 83, no. 1, V39-V48.

Huang, G., R. Nammour, and W. Symes, 2017, Full-waveform inversion via source-receiver extension: Geophysics, 82, no. 3, R153-R171.

Jin, S., and R. Madariaga, 1993, Background velocity inversion with a genetic algorithm: Geophysical Research Letters, 20, 93-96.

-----, 1994, Nonlinear velocity inversion by a two step monte carlo method.: Geophysics, 59, no. 4, 577-590.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi, 1983, Optimization by simulated annealing: Science, 220, no. 4598, 671-680.

Kormendi, F., and M. Dietrich, 1991, Nonlinear waveform inversion of planewave seismograms in stratified elastic media: Geophysics, 56, no. 5, 664-674.

Kwon, T., S. J. Seol, and J. Byun, 2016, Reduction of noise effect on elastic full-waveform inversion by using plane-wave data: 1131-1135.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and application, Society for Industrial and Applied Mathematics, Philadelphia, Expanded Abstracts, 206-220.

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

Luo, Y., J. Xia, R. D. Miller, Y. Xu, J. Liu, and Q. Liu, 2009, Rayleigh-wave mode separation by high-resolution linear radon transform: Geophysical Journal International, 179,no. 1, 254-264.

Metivier, L., A. Allain, R. Brossier, Q. Mrigot, E. Oudet, and J. Virieux, 2018, Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach: Geophysics, 83, no. 5, R515-R540.

Metivier, L., R. Brossier, Q. Mrigot, E. Oudet, and J. Virieux, 2016a, Increasing the robustness and applicability of full-waveform inversion: An optimal transport distance strategy: The Leading Edge, 35, no. 12, 1060-1067.

----, 2016b, Measuring the mis t between seismograms using an optimal transport distance: application to full waveform inversion: Geophysical Journal International, 205,no. 1, 345-377.

----, 2016c, An optimal transport approach for seismic tomography: application to 3d full waveform inversion: Inverse Problems, 32, no. 11, 115008.

Minkoff, S. E., and W. W. Symes, 1997, Full waveform inversion of marine reflection data in the planewave domain: Geophysics, 62, no. 2, 540-553.

Operto, S., J. Virieux, J.-X. Dessa, and G. Pascal, 2006, Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: Application to the eastern nankai trough: Journal of Geophysical Research: Solid Earth,111, B09306.

Pan, G. S., R. A. Phinney, and R. I. Odom, 1988, Fullwaveform inversion of planewave seismograms in stratified acoustic media: Theory and feasibility: Geophysics, 53, no. 1, 21-31.

Pan, W., K. A. Innanen, G. F. Margrave, and D. Cao, 2015, Efficient pseudo-gauss-newton full-waveform inversion in the -p domain: Geophysics, 80, no. 5,R225-R14.

Pladys, A., R. Brossier, and L. Metivier, 2017, Fwi alternative misfit functions: What properties should they satisfy: 79th EAGE Conference and Exhibition, Tu P1 01.

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

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, 64, no. 3, 888-901.

Qiu, L., J. Ramos-Martnez, A. Valenciano, Y. Yang, and B. Engquist, 2017, Full-waveform inversion with an exponentially encoded optimal-transport norm: SEG Technical Program Expanded Abstracts, 1286-1290.

Sacchi, M. D., and T. J. Ulrych, 1995, High resolution velocity gathers and offset space reconstruction: Geophysics, 60, no. 4, 1169-1177.

Sambridge, M., and K. Mosegaard, 2002, Monte carlo methods in geophysical inverse problems.: Reviews of Geophysics, 40, no. 3, 1009.

Sava, P., and S. Fomel, 2006, Time-shift imaging condition in seismic migration: Geophysics,71, no. 6, S209-S217.

Sen, M. K., and P. L. Sto a, 1992, Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion: Geophysical Journal International, 108, no. 1, 281-292.

Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot pro le migration: Geophysics, 73, no. 5, VE49-VE59.

Shen, X., I. Ahmed, A. Brenders, J. Dellinger, J. Etgen, and S. Michell, 2018, Full-waveform inversion: The next leap forward in subsalt imaging: The Leading Edge, 37, no. 1, 67b1-67b6.

Shin, C., and Y. Cha, 2008, Waveform inversion in the laplace domain: Geophysical Journal International, 173, no. 3, 922-931.

----, 2009, Waveform inversion in the laplacefourier domain: Geophysical Journal International, 177, no. 3, 1067----1079.

Sirgue, L., O. I. Barkved, J. Dellinger, U. A. J. Etgen, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at valhall: First Break, 28, 65-70.

Sirgue, L., and R. G. Pratt, 2004, E cient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, 69, no. 1, 231-248.

Stoffa, P. L., P. Buhl, J. B. Diebold, and F. Wenzel, 1981, Direct mapping of seismic data to the domain of intercept time and ray parameter planewave decomposition: Geophysics, 46, no. 3, 255-267.

Stoffa, P. L., and M. K. Sen, 1991, Nonlinear multiparameter optimization using genetic algorithms: Inversion of planewave seismograms: Geophysics, 56, no. 11, 1794-1810.

Sun, B., and T. Alkhalifah, 2018a, The application of an optimal transport to a preconditioned data matching function for robust waveform inversion: SEG Technical Program Expanded Abstracts, 5168-5172.

----, 2018b, Automatic wave-equation migration velocity analysis by focusing subsurface virtual sources: Geophysics, 83, no. 2, U1-U8.

----, 2018c, Mitigate cycle skipping in fwi: A generalized instantaneous travel-time approach: 80th EAGE Conference and Exhibition.

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

Tao, Y., and M. K. Sen, 2013, Frequency-domain full waveform inversion with plane-wave data: Geophysics, 78, no. 1, R13-R23.

Tarantola, A., 1984, Inversion of seismic re ection data in the acoustic approximation: Geophysics, 49, no. 8, 1259-1266.

Thorson, J. R., and J. F. Claerbout, 1985, Velocity stack and slantstack stochastic inversion: Geophysics, 50, no. 12, 2727-2741.

Van Leeuwen, T., and W. Mulder, 2008, Velocity analysis based on data correlation: Geophysical Prospecting, 56, no. 6, 791-803.

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.

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, 79, no. 2,R63-R77.

Vigh, D., and E. W. Starr, 2008, 3d prestack plane-wave, full-waveform inversion: Geophysics, 73, no. 5, VE135-VE144.

Villani, C., 2003, Topics in optimal transportation, graduate studies in mathematics: American Mathematical Society, 58, 1-145.

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

Wang, B., Y. Kim, C. Mason, and X. Zeng, 2008, Advances in velocity model-building technology for subsalt imaging: Geophysics, 73, no. 5, VE173-VE181.

Wang, J., M. Ng, and M. Perz, 2010, Seismic data interpolation by greedy local radon transform: Geophysics, 75, no. 6,WB225-WB234.

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, 81, no. 6, R429-R445.

Warner, M., A. Ratcli e, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. tekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3d full-waveform inversion: Geophysics, 78, no. 2, R59-R80.

Woodward, M. J., D. Nichols, O. Zdraveva, P. Whit eld, and T. Johns, 2008, A decade of tomography: Geophysics, 73, no. 5, VE5-VE11.

Yang, Y., and B. Engquist, 2018, Analysis of optimal transport and related mis t functions in full-waveform inversion: Geophysics, 83, no. 1, A7-A12.

Yang, Y., B. Engquist, J. Sun, and B. F. Hamfeldt, 2018, Application of optimal transport and the quadratic wasserstein metric to full-waveform inversion: Geophysics, 83, no. 1, R43-R62.

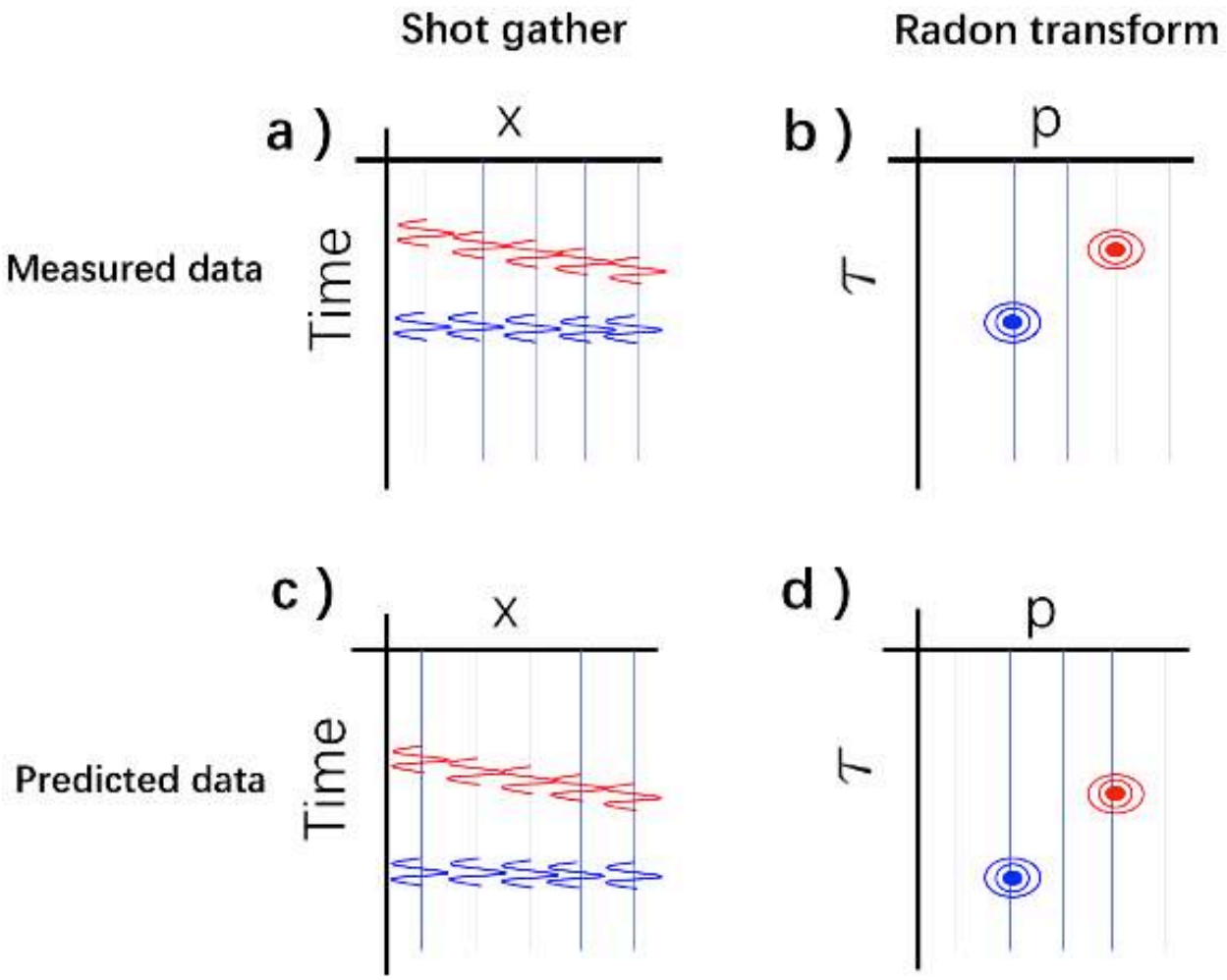


Figure 1: Cartoon illustration of the Radon transform: The common shot gathers of the measured data in a) the time-space domain; b) the Radon domain. The common shot gathers for the predicted data in c) the time-space domain; d) the Radon domain. Note after Radon transform, the events would be separated based on their slopes and this would reduce the cross-talks between different events (the red and the blue here) when deconvolution of the measured data with the predicted data.

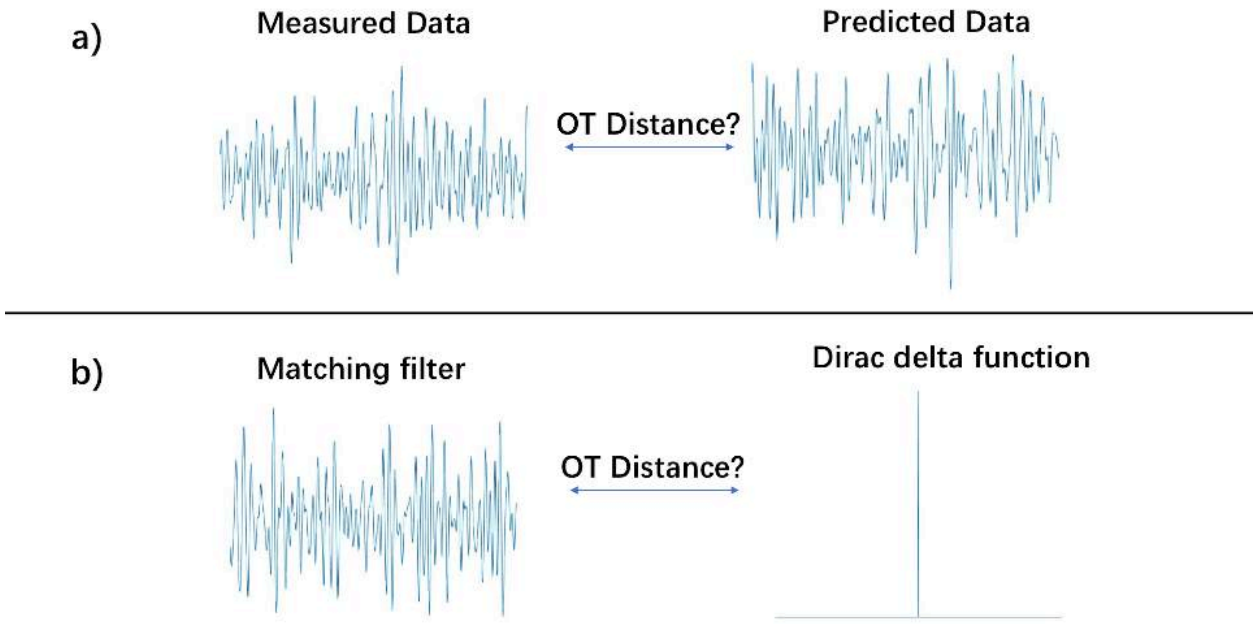
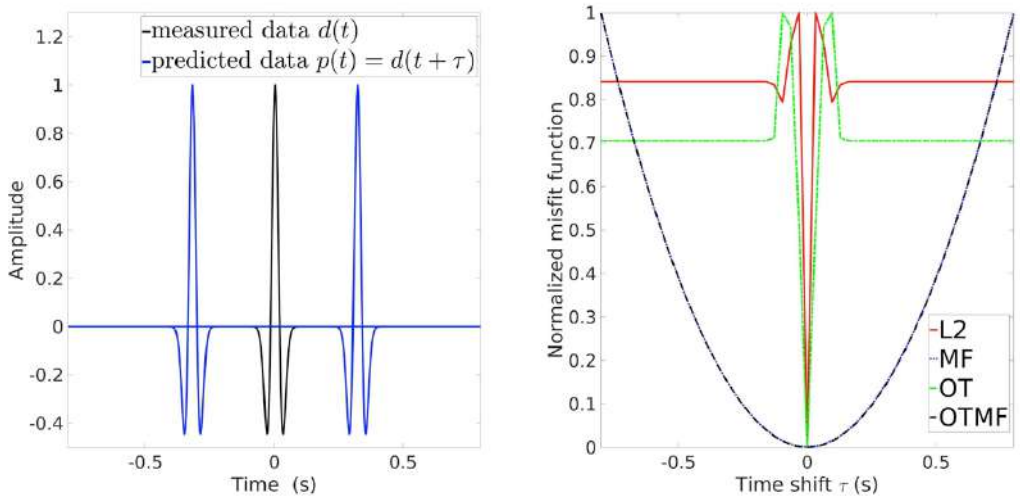
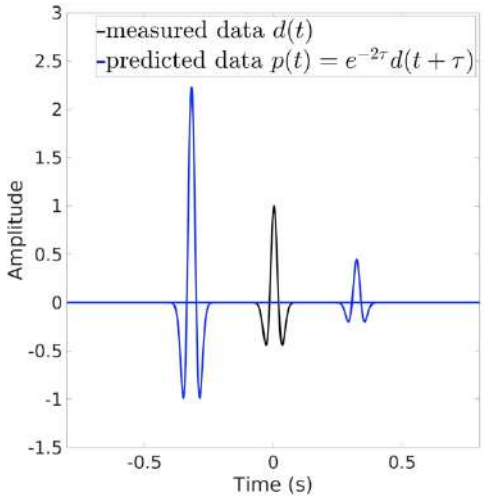


Figure 2: Cartoon illustration of the different strategy of applying optimal transport (OT) in full-waveform inversion : a) measure the OT distance between the measured data and the predicted data; b) measure the OT distance between the resulting matching filter and the Dirac delta function. In OT, the data for comparison should be probability distribution, i.e., they should be positive and mass equals to 1. The strategy a) faces the difficulty of transforming the data into a distribution, as any modification to the data directly would changes its amplitude or phase. And worse, it may destroy the convexity of the misfit function related to time shift. While strategy b) can safely precondition the resulting matching filter to be a distribution without touch the original data and this approach also issues the convexity regrading to time shift.

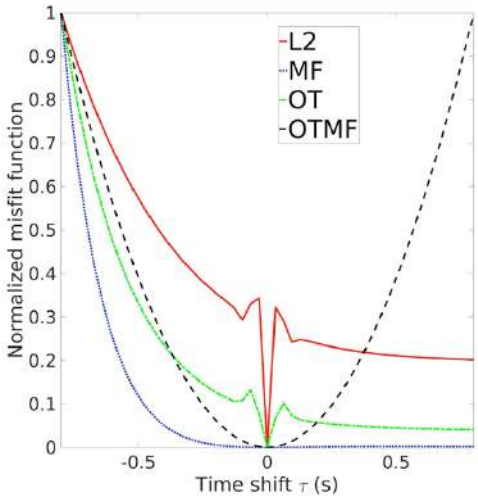


(a) (b)

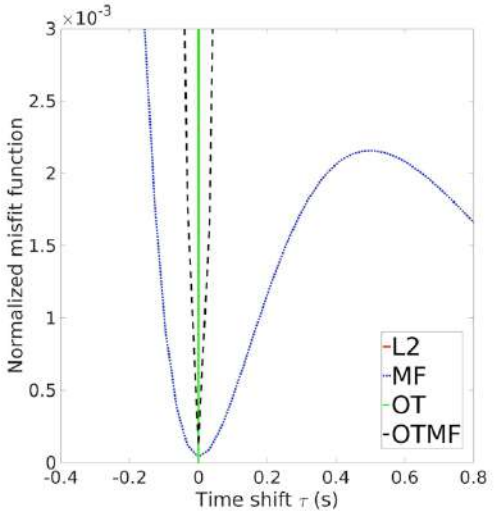
Figure 3: a) The measured (black curve) and the shifted predicted (blue curve) signals; and b) the misfit functions comparison. Here, L2 denotes L2 norm misfit function; MF is for conventional matching-filter approach based on penalty misfit function; OT presents the conventional optimal-transport approach which measures the OT distance between the predicted and measured data directly; OTMF is the proposed optimal-transport matching-filter approach which tries to measure the OT distance between the resulting matching filter and the Dirac delta function. In Figure b, it is obvious that the conventional optimal transport approach has a local minimum while the proposed method (the dash black line) shows better convexity related to time shift.



(a)

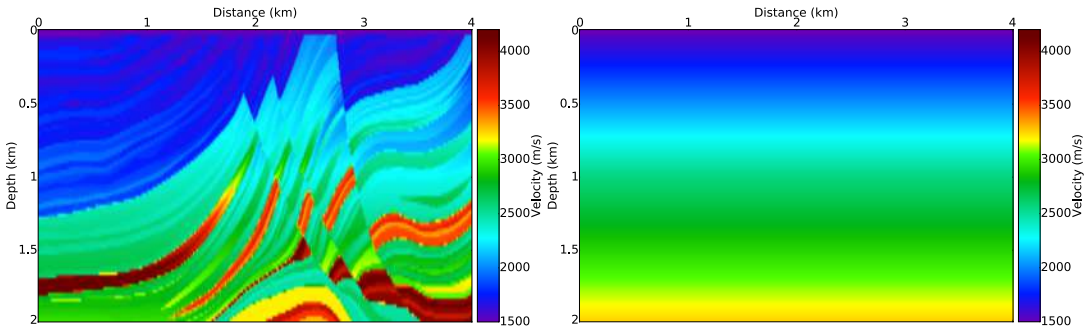


(b)



(c)

Figure 4: a) The measured (black curve) and the shifted and scaled predicted (blue curve) signals; b) and the mis t functions comparison; c) a zoomed view of Figure b). In Figure c, it is obvious that due to the sensitivity to the amplitude, the conventional matching filter approach (the dash blue curve) shows local minimum while the proposed method still ensures the convexity. Refer to the caption of Figure 3 for extra information.



(a) (b)

Figure 5: a) The true velocity; and b) the initial velocity for the modified Marmousi model.

In the example, 80 shots and 400 receivers are distributed at depth 20m with a source interval of 100 m and receiver interval of 20 m. An absorbing boundary condition is used at the surface.

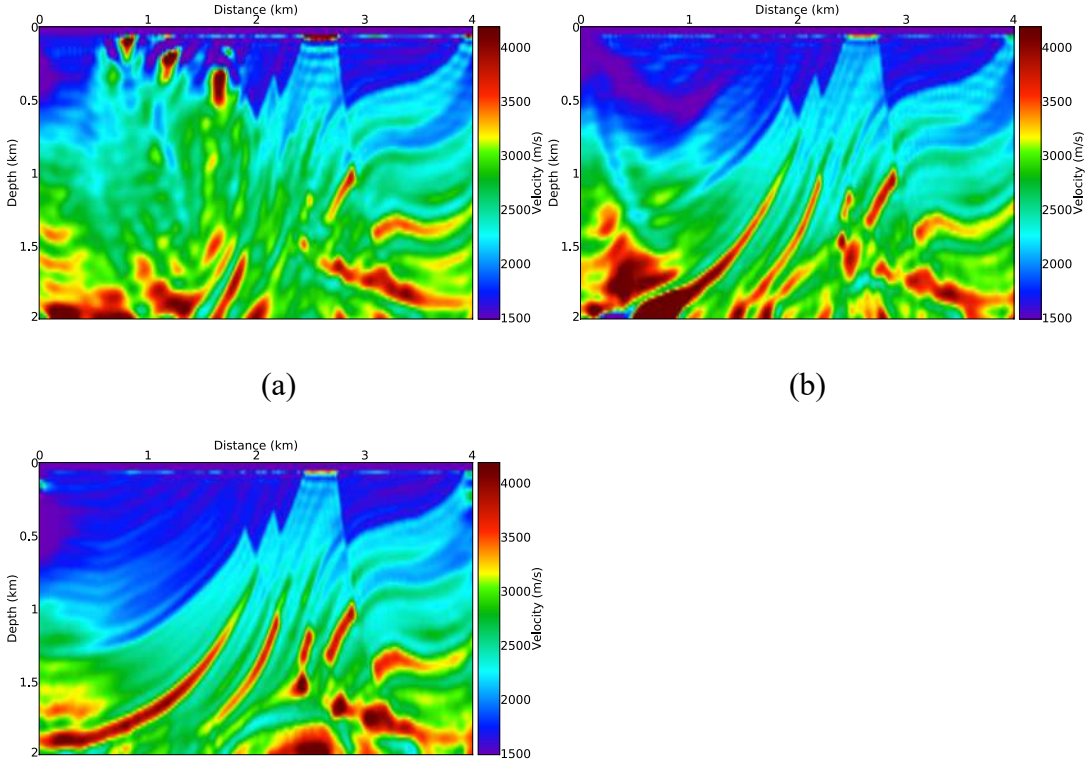
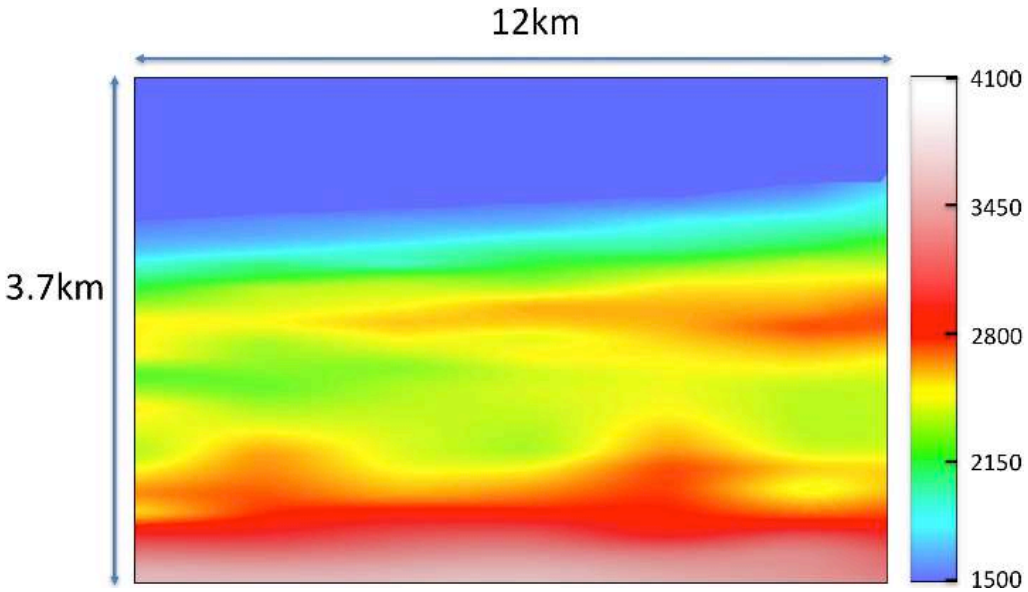
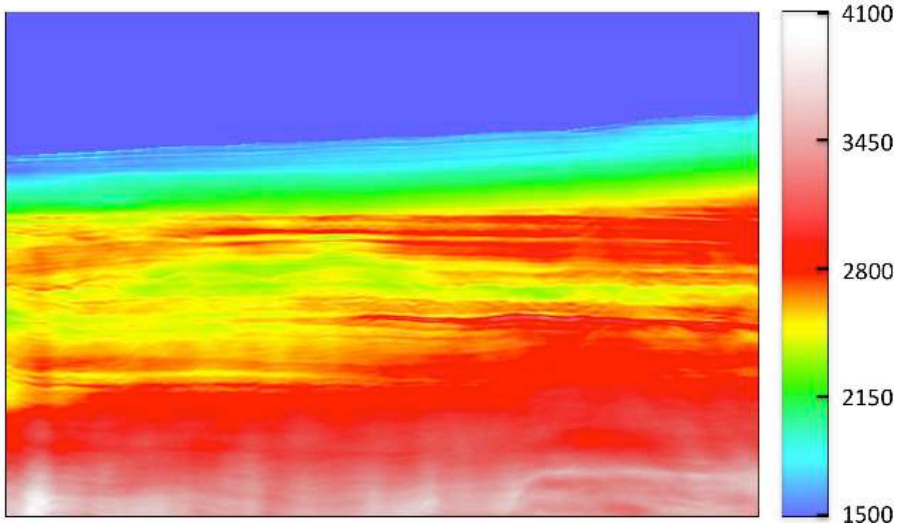


Figure 6: The inverted velocity using a) the L2-norm misfit function; b) the conventional optimal-transport matching-filter approach; d) the Radon-domain optimal-transport matching-filter approach. Due to "cycle skipping", L2-norm misfit function failed to get a meaningful result. Compared to the result of figure b) by optimal-transport matching-filter approach in time-space domain, the Radon-domain result of figure c) shows higher accuracy, especially for the deeper part of the model.

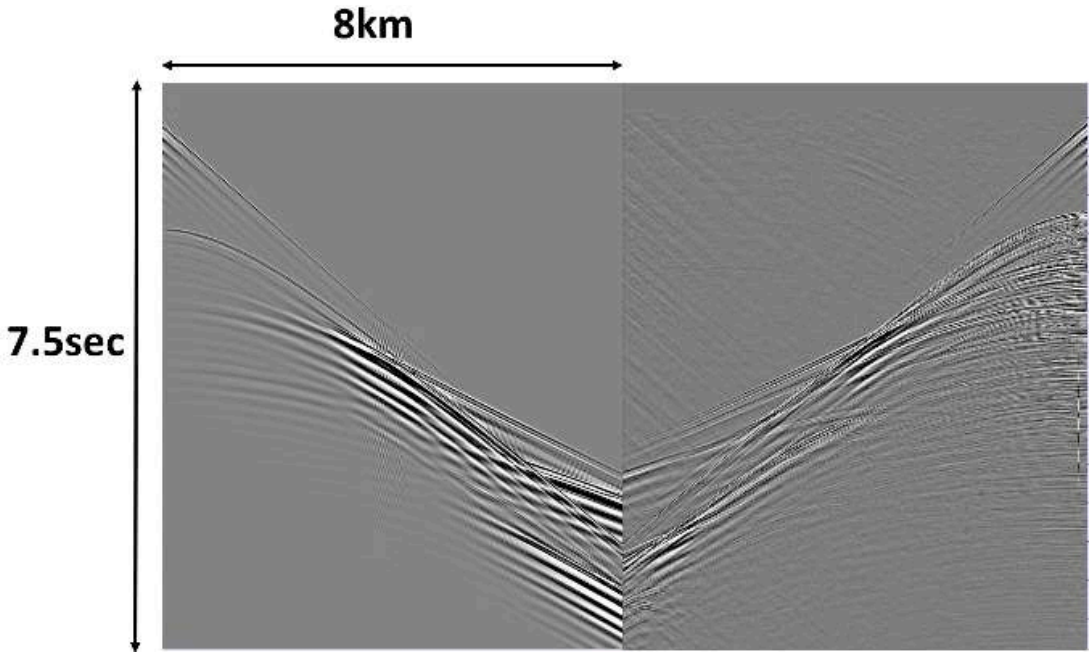


(a)

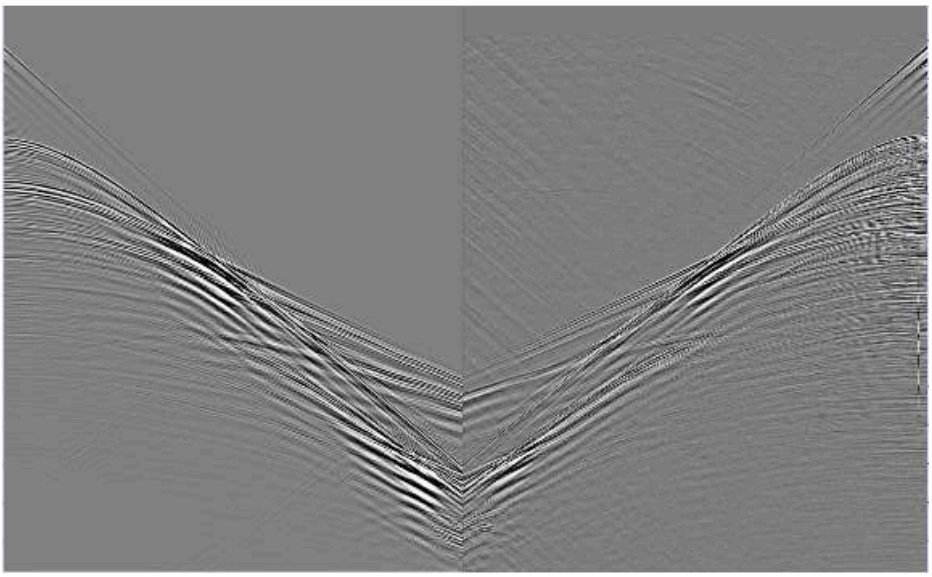


(b)

Figure 7: a) The initial model; and b) the inverted model for the offshore real dataset. Note the initial model is obtained from a raw RMS velocity.



(a)



(b)

Figure 8: Right panel is the shot record of the real seismic data while left panel is the modeled one from a) the initial model; b) the inverted model. The well-fitting of the record demonstrates the effectiveness of the proposed method for reducing cycle skipping.