Localized adaptive waveform inversion: theory and numerical verification

Peng Yong, Romain Brossier, Ludovic Métivier, Jean Virieux

Geophysical Journal International

Geophys. J. Int. (2023) **233,** 1055–1080
Advance Access publication 2022 December 15
GJI Seismology

https://doi.org/10.1093/gji/ggac496

Localized adaptive waveform inversion: theory and numerical verification

Peng Yong [1], Romain Brossier,[1] Ludovic Métivier[1,2] and Jean Virieux[1]

[1] *Université Grenoble Alpes, ISTerre, F-38000 Grenoble, France. E-mail: peng.yong@univ-grenoble-alpes.fr*
[2] *Université Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France*

Accepted 2022 December 9. Received 2022 December 5; in original form 2022 June 6

SUMMARY

Correctly interpreting phase events thanks to data processing techniques based on correlation or deconvolution has been the focus of numerous studies in the field of high-resolution seismic imaging using full-waveform inversion. To mitigate the non-convexity of the misfit function and the risk to converge towards non-informative local minima, correlation and deconvolution techniques make it possible to focus on phase information instead of amplitude information and to design more convex misfit function, alleviating the dependency of the full-waveform inversion process on the accuracy of initial models. Such techniques however rely on the assumption that phase events can be compared one by one, or that all the phase events are shifted in time in a similar way. This assumption is not satisfied in practice, which limits the effectiveness of these correlation/deconvolution-based methods. To overcome this issue, we propose to account for the non-stationary relation between observed and predicted data through a local in-time deconvolution technique, based on time–frequency analysis of the signal using a Gabor transform. This makes it possible to estimate instantaneous time-shift between locally coherent phase events. This strategy generalizes the conventional normalized deconvolution technique, which has been popularized under the name of adaptive waveform inversion. To support the introduction of our novel method, we compare it with four misfit functions based respectively on classical cross-correlation, penalized cross-correlation, penalized deconvolution, and adaptive waveform inversion. We analyse the behaviour of these methods on specific scenarios, and then propose a comparison on 2-D synthetic benchmarks. We show how our 'localized' adaptive waveform inversion applies in these realistic tests and overcomes some of the limitations of the aforementioned techniques.

Key words: Fourier analysis; Waveform inversion; Inverse theory; Computational seismology.

1 INTRODUCTION

Full-waveform inversion (FWI) has gradually been adopted in the practical workflows of seismic data process due to its capability to build high-resolution velocity models (Pratt 1999; Sirgue *et al.* 2010; Zhu *et al.* 2012; Operto *et al.* 2015; Lei *et al.* 2020; Huang *et al.* 2021). Seismic waveform inversion is recast as an iterative PDE-constrained optimization problem in FWI formulation, and the optimal model is obtained by using gradient-based optimization methods to minimize the misfit between observed and predicted waveforms from an initial model (Lailly 1983; Tarantola 1984). From an optimization point of view, FWI is a non-convex and non-linear problem, which means that it may be trapped into local minima when the starting model is not accurate enough using gradient-based methods. The non-convexity increases with frequency and, in practice, low-frequency content, if available in seismic data, together with the multiscale inversion strategy (Bunks *et al.* 1995), may prevent the risk of classical FWI approach converging to a spurious model (Virieux & Operto 2009; Virieux *et al.* 2017).

For field data applications, the initial models for FWI formulation are usually built by first-arrival traveltime tomography (Zhu *et al.* 1992; Zhang & Toksöz 1998; Taillandier *et al.* 2009), migration-based reflection traveltime inversion (Stork 1992; Chavent *et al.* 1994; Burdick *et al.* 2014), or stereotomography (Billette & Lambaré 1998; Prieux *et al.* 2013; Tavakoli F. *et al.* 2017; Sambolian *et al.* 2019). The initial low-wavenumber velocity models provided by these methods based on their kinematic information are agnostic to the frequency content of seismic waves (Červený 2001; Virieux *et al.* 2007): they cannot ensure that the classical FWI converges to a global minimum (Virieux *et al.* 2017). Besides, these kinematic methods rely on attributes picking and association (e.g. first

break, traveltime, slope), which already narrow the search of suitable models. Moreover, such careful extraction of observations can be a heavy task for large-amount of exploration data and prevents any correction on mistaken phase associations during the process of inversion.

The non-convexity issue of the misfit function can be alleviated by lowering the frequency generated by the source (Plessix *et al.* 2010; Ten Kroode *et al.* 2013; Dellinger *et al.* 2016): however it relies on long-offset and full-azimuth acquisitions (Plessix & Krupovnickas 2021; Vigh *et al.* 2021), as first-arrival phases are essentially driving the optimization workflow. Late low-frequency reflection phases, quite important for high-resolution contribution, are likely to interfere, increasing the non-linearity of the inverse problem if the current model is not precise enough (Ten Kroode *et al.* 2013). However, sources injecting low-frequency content inside the subsurface are economically expensive. Therefore, there is an interest for industry and academia in methodological developments of robust FWI for high-resolution subsurface imaging without these low-frequency content.

Designing a convex misfit function is one of the key research topics in methodology development of FWI (Pladys *et al.* 2021). Jannane *et al.* (1989) pointed out that medium-scale to large-scale velocity perturbation generates time-shifts in the data. Therefore, designing a convex misfit function with respect to time-shifts is a good proxy for making a convex data misfit function with respect to medium-scale to large-scale velocity perturbation (Luo & Schuster 1991; Woodward 1992). However, effectively and accurately capturing time-shifts is challenging for complex and noisy seismic data. Many techniques, such as the cross-correlation (Luo & Schuster 1991; van Leeuwen & Mulder 2010), the deconvolution (Luo & Sava 2011; Warner & Guasch 2016; Sun & Alkhalifah 2018, 2019a, b, 2020), the dynamic-time warping (Ma & Hale 2013; Chen *et al.* 2021), and the optimal transport (Engquist *et al.* 2016; Métivier *et al.* 2016, 2018; Yang *et al.* 2018; Yong *et al.* 2019; Górszczyk *et al.* 2021) have been used to capture time-shifts. Numerical and realistic studies have illustrated that these methods can make seismic waveform inversion less prone to cycle skipping issues and more tolerant to initial model design. In addition, low-frequency enhancement methods, when possible, can play an important role in mitigating cycle-skipping issues. Synthetic studies show that the signal envelope can produce ultra low-frequency content below the lowest physical frequency in the source spectrum, which can be exploited to improve the background velocity model (Bozdağ *et al.* 2011; Wu *et al.* 2014; Hu *et al.* 2019; Chen *et al.* 2020). On the other hand, low-frequency extrapolation with multiple signal classification (Li & Demanet 2016) and deep learning training (Sun & Demanet 2020; Hu *et al.* 2021) recently presents some encouraging synthetic and field data results for initial velocity design. Nevertheless, it is still an open question of how to maintain the effectiveness when applying a trained neural network to other data sets, especially to field data (Hu *et al.* 2021; Yu & Ma 2021).

In this work, we consider the deconvolution technique used in the adaptive waveform inversion (AWI). In this method, a stationary convolutional matching filter is designed to globally map one observed data trace into one predicted data trace (Warner & Guasch 2016; Guasch *et al.* 2019). For complex seismic data, a non-stationary convolutional approach (Margrave *et al.* 2011) for computing the matching filter seems to be more suitable. In a stationary convolutional operator, the filter is assumed invariant for all phase events. However, time-shifts of the different events are generally different, and therefore a local matching filter is expected to be more physically meaningful, since it accounts for the non-stationarity nature of seismic data. In fact, the interference on the global matching filter between different events could make the AWI misfit function non-convex with time-shift (Pladys *et al.* 2021), thereby it is necessary and beneficial to restrict the comparison of the data to selected arrivals by a local matching filter. In this paper, we detail how to efficiently obtain the local matching filter with a Gabor transform (Gabor 1946), and how to implicitly estimate instantaneous time-shifts with the local matching filter. Let us mention that the importance of taking non-stationarity into account when dealing with the cycle-skipping issue is also discussed in many studies (e.g. Baek *et al.* 2014; Díaz & Sava 2015; Zhu & Fomel 2016; Zhu 2018). Most of these methods are usually used to increase the accuracy of starting models, reducing the risk of the standard FWI converging into local minima. However, these methods seem lacking the capability to generate high-resolution models even with a broad-band data, meaning that they cannot fully use the information in seismic data. That is, they cannot make most of the precious low-frequency content in seismic data to make the starting velocity model reach the possible highest accuracy. In contrast, our approach can automatically generate high-resolution velocity models, thereby it can more robustly fill the wavenumber gap (Virieux *et al.* 2017, Fig. 2) in the road to velocity models of broadband spectra.

We shall review and explore relationships of AWI and the proposed localized AWI (LAWI) approaches with three relevant methods, namely, classical cross-correlation (Luo & Schuster 1991), penalization-based cross-correlation (van Leeuwen & Mulder 2010) and deconvolution (Luo & Sava 2011). The physical interpretation of AWI can be highlighted by the concept of centroid frequency used in quality-factor Q estimation (Quan & Harris 1997): from our point of view, the key point of AWI is the estimation of the centroid time of the matching filter, which implicitly matches observed and predicted traces. This physical interpretation motivates us in the development of a local matching filter for detecting instantaneous time-shifts to locally coherent phase events. Besides, we also illustrate how and why the normalization in AWI (and in LAWI) makes such approaches essentially different from penalization-based cross-correlation and deconvolution methods.

After this introduction, we shortly summarize the standard FWI pointing out the adjoint-source definition in the second section. Then, in the third section, we outline the key points of these five mentioned inversion methods based on the same wave-equation forward problem. Through simple numerical examples, we attempt to clarify the relationships among these matching-filter-based misfit functions by a sensitivity-kernel comparison and objective function analysis, also illustrating the benefit of local analysis when the relation between compared time-signals is non-stationary. Next, we investigate the capability of AWI and LAWI to handle local-minimum issues with a 2-D Valhall synthetic data through adjusting initial models. For a further comparison between AWI and LAWI, we apply them to the Chevron 2014 benchmark data set. We analyse the inversion results with well log, migrated images and corresponding common-image gathers in the offset domain. Finally, discussions and conclusions follow.

2 BRIEF SUMMARY OF FWI

Classical FWI can be written as a PDE-constrained optimization problem over all available sources denoted by $\mathbf{x}_s$ and related receivers $\mathbf{x}_r^s$. The misfit function can be written as

$$\min_{\mathbf{m}} \mathbb{J}_{L2}(\mathbf{m}) = \frac{1}{2}\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_r^s}\int_0^T (p[\mathbf{m}](\mathbf{x}_r^s,\mathbf{x}_s,t) - d(\mathbf{x}_r^s,\mathbf{x}_s,t))^2 \mathrm{d}t$$
$$\text{subject to} \quad A(\mathbf{m})u(\mathbf{x},\mathbf{x}_s,t) = s(t,\mathbf{x}_s), \tag{1}$$

where the physical model parameter set is denoted by $\mathbf{m}$ (we consider only wave speed in this work), the observed and predicted data at the receiver location $\mathbf{x}_r^s$ are denoted by $d(\mathbf{x}_r^s,\mathbf{x}_s,t)$ and $p[\mathbf{m}](\mathbf{x}_r^s,\mathbf{x}_s,t)$.

In practice, gradient-based local optimization methods are applied to solve this large-scale inverse problem. The optimal solution can be found through iterations. Knowing the model $\mathbf{m}_k$, the new model $\mathbf{m}_{k+1}$ is updated by

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \Delta\mathbf{m}_k, \tag{2}$$

where the step-length α_k at the kth iteration is determined, for example, by a linesearch method with classical Wolfe conditions (Métivier & Brossier 2016). The current model update direction $\Delta\mathbf{m}_k$ is given by

$$\Delta\mathbf{m}_k = -P_k \nabla\mathbb{J}_{L2}(\mathbf{m}_k), \tag{3}$$

where a pre-conditioner P_k is expected to accelerate convergence rate, after the computation of the gradient $\nabla\mathbb{J}_{L2}(\mathbf{m}_k)$. With the adjoint-state method, the gradient can be efficiently computed with an implicit summation over sources and receivers by

$$\nabla\mathbb{J}_{L2}(\mathbf{m}) = \int_0^T \lambda(t)\frac{\partial A(\mathbf{m})}{\partial \mathbf{m}}u(t)\,\mathrm{d}t. \tag{4}$$

The analytical expression $\partial A(\mathbf{m})/\partial\mathbf{m}$ defines the diffraction-radiation operator, which depends on the selected model parameter set. The adjoint field λ is calculated by back-propagating the receiver data residuals (Tarantola 1984; Plessix 2006) with the adjoint-state equation

$$A^T(\mathbf{m})\lambda = -\mathbf{r}_{L2}, \tag{5}$$

where the adjoint operator $A^T(\mathbf{m})$ could have a similar structure as the wave-modelling operator $A(\mathbf{m})$. The adjoint source $\mathbf{r}_{L2}$ is the data residual in the conventional FWI given for each (source, receiver) pair by

$$\mathbf{r}_{L2} = p(\mathbf{x}_r^s,\mathbf{x}_s,t) - d(\mathbf{x}_r^s,\mathbf{x}_s,t), \tag{6}$$

corresponding to the derivative of the misfit function $\mathbb{J}_{L2}$ with respect to the predicted data term.

In the next section, we introduce some misfit functions using matching filters for a more robust behaviour with respect to the initial model design.

3 MATCHING-FILTER BASED STRATEGIES FOR MITIGATING CYCLE-SKIPPING ISSUES

Comparison of observed and predicted waveforms may emphasize phase information in different ways. The wave-equation traveltime inversion (Luo & Schuster 1991) exploits only the phase information through a connective function, and its aim is the building of a more accurate low-resolution initial model for FWI. In order to avoid the pure time-shift assumption in Luo & Schuster (1991), a penalization-based cross-correlation misfit is developed (van Leeuwen & Mulder 2010), which increases the applicability of the wave-equation traveltime inversion. The alternative penalization-based deconvolution (Luo & Sava 2011) makes the matching filter more confined when the velocity model moves to the correct one. Although these two penalization-based methods are both designed to be driven by time lags, while numerical experiments show that they are sensitive to amplitude variations. Thanks to the normalization, AWI (Warner & Guasch 2016) can reduce the amplitude effect by emphasizing time-shift detection (Li 2018; Sun & Alkhalifah 2019a). Let us underline that the proposed LAWI is based on a local matching filter different from the global matching filter used in the AWI approach: therefore, it is more suitable to process complex seismic signals.

3.1 Wave-equation traveltime tomography

Wave-equation traveltime inversion attempts to determine a velocity model by minimizing the misfit function

$$\mathbb{J}_{WTI} = \frac{1}{2}(\delta\tau)^2, \tag{7}$$

where $\delta\tau$ denotes a time-shift between observed and predicted data. Such a delay is obtained through a connective function expressed by

$$\delta\tau = \arg\max_{\tau} c(\tau), \tag{8}$$

where the function $c(\tau)$ is defined by a cross-correlation between observed data $d(t)$ and predicted data $p(t)$. The time-shift is explicitly evaluated from the peak value of the connective function. For the sake of simplicity, we consider the data of one receiver here and below. Thanks to Fast Fourier transform, cross-correlation can be efficiently implemented in the frequency domain as the product

$$\tilde{c}(\omega) = \tilde{d}(\omega)^\dagger \tilde{p}(\omega), \tag{9}$$

where the complex conjugate is denoted by $\dagger$, and complex-valued quantities $\tilde{d}(\omega)$ and $\tilde{p}(\omega)$ are observed and predicted data in the frequency domain. The time-domain and frequency-domain representations of one trace can be connected through Fourier transform pair:

$$\tilde{f}(\omega) = \mathcal{F}[f](\omega) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(t)e^{-i\omega t}\,\mathrm{d}t, \tag{10}$$

$$f(t) = \mathcal{F}^{-1}[\tilde{f}](t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \tilde{f}(\omega)e^{i\omega t}\,\mathrm{d}\omega. \tag{11}$$

We shall introduce, in addition to the time, a time variable τ which will have the same sampling as the one used for the time variable t. Therefore, they are sharing the same frequency sampling. We distinguish the Fourier transform and its inverse by subscripts $\mathcal{F}_t$ for time variable and $\mathcal{F}_\tau$ for τ variable.

3.2 Penalization-based cross-correlation

When the amplitude spectra of observed and predicted data are not identical, the time-shift determined by the maximal value of the connective function is no longer valid (van Leeuwen & Mulder 2010). In addition, the adjoint source computation in the classical wave-equation traveltime inversion is under the assumption that observed and predicted data are purely time-shifted. In order to make wave-equation traveltime inversion more robust to field data, and

also inspired by differential semblance optimization used in migration velocity analysis (Symes & Carazzone 1991; Mulder & ten kroode 2002; Shen & Symes 2008), van Leeuwen & Mulder (2010) propose a penalization-based cross-correlation objective function in the data domain, defined as

$$\mathbb{J}_C = \frac{1}{2}\int_{\mathbb{R}} \mathcal{P}(\tau)c^2(\tau)\mathrm{d}\tau, \tag{12}$$

where the expression $\mathcal{P}(\tau)$ penalizes the signal energy at non-zero time lags: this makes the wave-equation traveltime inversion more robust to spectral variations. One simple penalty function is the absolute value function $\mathcal{P}(\tau) = |\tau|$ which is adopted here: it will help our introduction of the centroid frequency useful for the interpretation of AWI. In practice, other convex or concave functions named annihilators can also been used (van Leeuwen & Mulder 2010; Pladys *et al.* 2021): they do not modify the methodological aspect of such an approach. The adjoint source of this cross-correlation-based objective function can be expressed by

$$\mathbf{r}_C = \mathcal{F}_t^{-1}\left[\mathcal{F}_\tau\left[|\tau|c(\tau)\right]\tilde{d}(\omega)\right]. \tag{13}$$

3.3 Penalization-based deconvolution

For a bandlimited or non-impulsive source, Luo & Sava (2011) underline that the cross-correlation peak is centred at the zero lag for the correct velocity model, but could not be confined to zero lag. They mention that a deconvolution-based matching filter, behaving as an impulsive function for an unlimited frequency range, is more confined for realistic oscillatory signals. Besides, the resolution of the inversion result is expected to be improved (Luo & Sava 2011; Warner & Guasch 2016). The deconvolution-based misfit function can be defined as

$$\mathbb{J}_D = \frac{1}{2}\int_{\mathbb{R}} |\tau|w^2(\tau)\mathrm{d}\tau, \tag{14}$$

where a matching filter $w(\tau)$ is based on a stationary convolutional operator that maps the observed data $d(t)$ onto the predicted data $p(t)$ through the convolution expression

$$d(t) \otimes w(\tau) = p(t). \tag{15}$$

The matching filter can be obtained in the frequency domain as

$$\tilde{w}(\omega) = \frac{\tilde{d}(\omega)^\dagger \tilde{p}(\omega)}{\tilde{d}(\omega)^\dagger \tilde{d}(\omega) + \epsilon}, \tag{16}$$

where the so-called water-level parameter ϵ is a small positive number to stabilize the division (Clayton & Wiggins 1976). It can be empirically chosen as the average of the power spectrum of the observed data scaled by a value in the interval $[10^{-1}, 10^{-3}]$. Such a selection depends on signal-to-noise ratio (SNR) of the trace. Low SNR data may require high ϵ values. Note that when ϵ is an enormous value, the deconvolution will reduce to cross-correlation. ϵ controls the transition between deconvolution and cross-correlation. The corresponding adjoint source can be given by

$$\mathbf{r}_D = \mathcal{F}_t^{-1}\left[\mathcal{F}_\tau\left[|\tau|w(\tau)\right]\frac{\tilde{d}(\omega)}{\tilde{d}(\omega)\tilde{d}^\dagger(\omega) + \epsilon}\right]. \tag{17}$$

3.4 Adaptive waveform inversion

Using the same matching filter as in the deconvolution-based misfit, Warner & Guasch (2016) define the following objective function

$$\mathbb{J}_{AWI} = \frac{1}{2}\frac{\int_{\mathbb{R}} |\tau|w^2(\tau)\mathrm{d}\tau}{\int_{\mathbb{R}} w^2(\tau)\mathrm{d}\tau}. \tag{18}$$

The essential difference between the AWI misfit and the deconvolution-based misfit comes from a specific normalization: AWI normalizes the misfit with the L^2 norm of the matching filter. As detailed in the Appendix, the adjoint source of AWI can be written as

$$\mathbf{r}_{AWI} = \mathcal{F}_t^{-1}\left[\mathcal{F}_\tau\left[\frac{(|\tau| - 2\mathbb{J}_{AWI})w(\tau)}{\int_{\mathbb{R}} w^2(\tau)\mathrm{d}\tau}\right]\frac{\tilde{d}(\omega)}{\tilde{d}(\omega)\tilde{d}^\dagger(\omega) + \epsilon}\right]. \tag{19}$$

Such a normalization introduced by Warner & Guasch (2016) is a key element in mitigating local-minimum issues as shown by various publications (e.g. Li 2018; Sun & Alkhalifah 2019a, b). The definition (18) is quite similar to the expression for computing the centroid frequency of a signal, widely used for getting the quality factor parameter Q in geophysical applications (Quan & Harris 1997; Wang 2009). Given a signal $u(t)$, the centroid frequency f_u of the signal can be computed by the expression

$$f_u = \frac{\int_{\mathbb{R}} |\omega| A_u^2 \mathrm{d}\omega}{\int_{\mathbb{R}} A_u^2 \mathrm{d}\omega}, \quad A_u = |\tilde{u}(\omega)|. \tag{20}$$

Thus, the misfit in AWI can be understood as an estimation of the centroid time of the matching filter for the implicit estimation of the time-shift between observed and predicted data. With such an interpretation, the deconvolution-based misfit can be regarded as the time-shift estimation through a filter-weighted operation. Thanks to the normalization in eq. (18), the amplitude influence of the matching filter is reduced and AWI enhances the time-shift influence while still taking into account mildly the data amplitude contribution.

To intuitively interpret AWI by connecting with the concept of centroid frequency, we consider one specific form of AWI that the penalty function is defined as the absolute value of the time-shift ($|\tau|$). Note that other penalty functions can be used, which could lead to other interpretations of AWI and behaviours in applications (Warner & Guasch 2016; Li 2018; Sun & Alkhalifah 2019b, 2020; Pladys *et al.* 2021). For instance, Sun & Alkhalifah (2018) use the time-shift τ as the penalty function and interpret AWI as an instantaneous traveltime approach (Sun & Alkhalifah 2019b). Besides, different penalty functions can affect the convergence rate of AWI, and comprehensive numerical studies can be found in Sun & Alkhalifah (2019b, 2020).

3.5 Localized adaptive waveform inversion

The global matching filter builds the connection between observed and predicted data in AWI: the same filter is applied for all phase events detected in a given seismogram. For complex traces, a local matching filter based on a non-stationary convolutional expression seems advisable. This makes it possible to capture local time-shifts related to locally coherent events. Let us describe first the LAWI method by using the Gabor transform. Then, we discuss the time window width to be selected for the Gabor transform in connection with the local aspect of phase events.

3.5.1 Methodology

The local matching filter used in LAWI relies on the Gabor transform, which can be defined as

$$\hat{f}(t,\omega) = \mathcal{G}[f](t,\omega) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} f(\xi)h_\sigma^\dagger(\xi - t)e^{-i\omega\xi}\mathrm{d}\xi, \tag{21}$$

$$f(t) = \mathcal{G}^{-1}[\hat{f}](t) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}^2} \hat{f}(\xi,\omega)h_\sigma(t - \xi)e^{i\omega t}\mathrm{d}\xi\mathrm{d}\omega. \tag{22}$$

h_σ denotes the time window function, and it is chosen as the following Gaussian function (Strang & Nguyen 1996; Fichtner *et al.* 2008)

$$h_\sigma(t) = (\pi\sigma^2)^{-\frac{1}{4}} e^{\frac{-t^2}{2\sigma^2}}, \tag{23}$$

where the parameter σ controls the radius of the window function. Such a window definition ensures the energy conservation $\int_{\mathbb{R}} h_\sigma^2(t)\mathrm{d}t = 1$. A time-varying matching filter in the frequency domain can be efficiently computed by Gabor deconvolution with a water-level parameter ϵ for stability:

$$\hat{w}(t,\omega) = \frac{\hat{d}^\dagger(t,\omega)\hat{p}(t,\omega)}{\hat{d}^\dagger(t,\omega)\hat{d}(t,\omega) + \epsilon}. \tag{24}$$

Here, $\hat{d}(t,\omega)$ and $\hat{p}(t,\omega)$ are the time–frequency spectra of predicted and observed data. The corresponding time-varying matching filter in the time domain can be obtained through

$$w(t,\tau) = \mathcal{F}_\tau^{-1}\left[\hat{w}(t,\omega)\right], \tag{25}$$

where $\mathcal{F}_\tau^{-1}$ denotes applying inverse Fourier transform for the variable τ (see the definition in subsection 3.1). The instantaneous centroid time-shift at the local time t can now be estimated by

$$T(t) = \frac{\int_{\mathbb{R}} |\tau| w^2(t,\tau)\mathrm{d}\tau}{\int_{\mathbb{R}} w^2(t,\tau)\mathrm{d}\tau + \eta}, \tag{26}$$

where the small positive water-level parameter η prevents instabilities for local null data. From numerical experiences, we find that the strategy used for the ϵ parameter selection also works for the parameter η.

Integrating all instantaneous time-shifts under the L^2 norm, we can define the misfit function of LAWI as

$$\mathbb{J}_{LAWI} = \frac{1}{2}\int_{\mathbb{R}} W(t)T^2(t)\mathrm{d}t, \tag{27}$$

where a positive weighting function $W(t)$ makes the misfit more robust for noisy data. In this work, we assume a simple weight $W(t) = 1$ for synthetic data, and, for low-SNR data applications, the following weight

$$W(t) = \log(1 + \int_{\mathbb{R}} |\hat{d}(t,\omega)|\mathrm{d}\omega). \tag{28}$$

This weight function is a variation of the one ($\log(1 + |\hat{d}(t,\omega)|)$) used in Fichtner *et al.* 2008 to stabilize time–frequency attributes calculation in noisy data. As the estimated instantaneous shifts is in the time domain, an integral of the time–frequency spectrum over frequency is made in the weight function. Besides, to mitigate large-amplitude dominance effect, the logarithmic function is used, instead of absolute function (Fichtner *et al.* 2008). Other weighting strategies are also possible (Fichtner *et al.* 2008, section 2.2.2).

As detailed in the Appendix the adjoint source of LAWI can be written as

$$\mathbf{r}_{LAWI} = 2\mathcal{G}^{-1}\left[\mathcal{F}_\tau\left[\frac{W(t)T(t)(|\tau| - T(t))w(t,\tau)}{\int_{\mathbb{R}} w^2(t,\tau)\mathrm{d}\tau + \eta}\right] \times \frac{\hat{d}(t,\omega)}{\hat{d}(t,\omega)\hat{d}^\dagger(t,\omega) + \epsilon}\right]. \tag{29}$$

3.5.2 *Selection of Gaussian window parameter*

When using time–frequency transformations for the analysis of time signals, one always has to face the problem of the parameter selection related to time-window width. In the Gabor transform and its inverse, defined respectively by eqs (21) and (22), the width of the sliding window h_σ does not explicitly depend on time or frequency. In general, the optimal choice of the width of the sliding window is case-dependent. In some applications, a flexible sliding window h may be preferable (Strang & Nguyen 1996; Mallat 2008).

In the LAWI approach, time–frequency analysis is applied to efficiently capture the local time-shift between observed and predicted data. Staying on the safe side when facing the cycle-skipping issue, we focus on the detection of local time-shift of a single phase event, instead of a precise time-shift of a specific seismic wiggle. In other words, the effective length of the sliding window T_w should be larger than the duration of the phase event T_e related to the seismic frequency content. Considering the time-shift T_s between observed and predicted data, a reasonable choice would be

$$T_w = T_e + T_s. \tag{30}$$

If the window width T_w becomes too large, the Gabor transform will lose time resolution. As a result, LAWI would behave like AWI, losing capability to locally analyze events. For the Gaussian function defined by eq. (23), the effective length T_w can be regarded as 4σ, because the range of $[-2\sigma, 2\sigma]$ encapsulates about 95 % of the surface of the normal distribution. Thus, the window-width parameter σ, required to be defined in the LAWI method, can be computed by

$$\sigma = \frac{T_w}{4} = \frac{T_e + T_s}{4}. \tag{31}$$

Now let us illustrate two key points in the strategy with the numerical example displayed in Fig. 1: (1) The effective length of the Gaussian window T_w can be considered as 4σ. (2) It is reasonable to make the sliding window covers two compared events.

A zero-phase Ricker wavelet with a peak frequency of 5 Hz, shown in Fig. 1 (a), is defined by

$$R_w(t) = \left(1 - 2\pi^2 f_d^2 t^2\right) e^{-\pi^2 f_d^2 t^2}. \tag{32}$$

The time support of a Ricker wavelet can be roughly estimated as $T_e = 2\sqrt{6}/(\pi f_d)$, which corresponds to the double length of the time interval between two local minima. The Gaussian function with $\sigma = T_e/4$ is plotted using the blue line in Fig. 1 (a). We can see that the Ricker wavelet can be appropriately covered by the Gaussian function.

Then, we consider two phase events with time-shift $T_s = 0.5$ s. With the analysis above, the effective length of Gaussian function should be $T_e + 0.5$ s, and the recommended width parameter in this case is $\sigma_0 = (T_e + T_s)/4$. From Fig. 1 (b), we can see that the Gaussian function with $\sigma = \sigma_0$ can cover the combined surface of the two Ricker wavelets. Three choices of window width are plotted in Fig. 1 (c). When the window width is too small, the estimation of the time-shift is below the expected value 0.5 s. Increasing the window width will improve the estimation, but a too large increase would affect the time resolution. Besides, a large window width will increase the risk of combining multiple phase events, reducing the resolution which is, however, less crucial at the beginning of the inversion process. A proper selection of σ values could be driven by numerical experiments.

For LAWI applications, the information (T_e and T_s) used for computing the window-width parameter σ can be easily obtained ahead of the inversion: The duration of a single-phase event T_e can be of the order of the characteristic source wavelet duration, and better preserved in a 3D geometry than in a 2D geometry. The time-shift T_s between observed and predicted data can be empirically estimated,

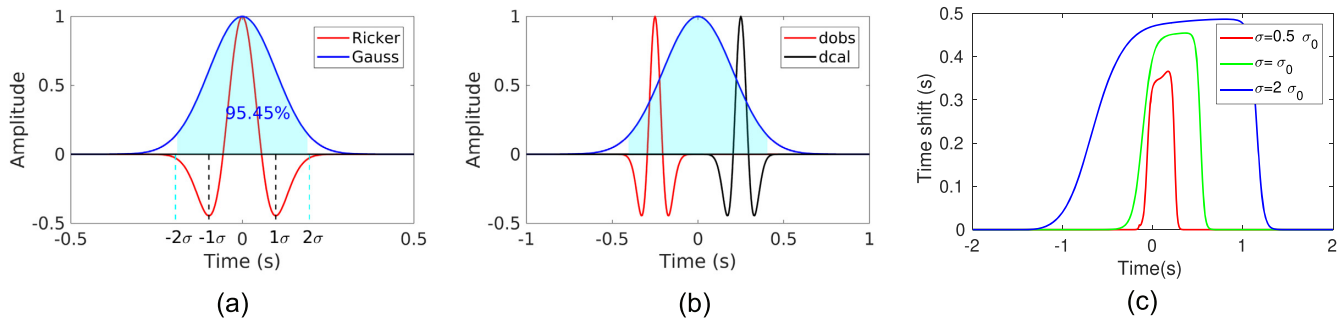

Figure 1. An illustration for the choice of Gaussian window: (a) a 5 Hz Ricker wavelet covered by the Gaussian function with a standard deviation $\sigma = T_e/4$, where T_e denotes the double length of the time interval between two local minima of the Ricker wavelet, (b) two Ricker wavelets with time-shift $T_s = 0.5$ s still covered by the Gaussian function with a recommended standard deviation $\sigma_0 = (T_e + T_s)/4$, (c) estimated instantaneous time-shifts with a too large width in blue, the one suggested in green which includes the time-shift of 0.5 s and a too small value in red.

normally less than 0.5 s in practical applications. Overall, the time-window definition is a common problem in seismology (Maggi *et al.* 2009) and our simple strategy does work well in the following numerical tests.

4 NUMERICAL ILLUSTRATION OF THESE FIVE DIFFERENT APPROACHES

Through simple numerical examples in cycle-skipped circumstances, we notice that approaches AWI and LAWI are able to distinguish whether the starting velocity is slower or faster than the true one, whereas the L^2 norm (Tarantola 1984), penalization-based cross-correlation (van Leeuwen & Mulder 2010) and deconvolution (Luo & Sava 2011) methods do not succeed. Through sensitivity kernel analysis with related adjoint sources and matching filters, parametric objective function evolution, we illustrate these two drastically different behaviours. Finally, we demonstrate the benefit of a local analysis when the relation between compared signals is non-stationary and show how the LAWI strategy handles the situation that the number of phase events is not the same in observed and calculated signals.

4.1 Analysis of sensitivity kernel and one-event time signal

Let us consider two 2-D models of 4 km by 4 km with constant velocities of 3.0 and 2.0 km s^{-1} to be reconstructed from an initial model of velocity of 2.5 km s^{-1}. A transmission configuration is given by a source and a receiver respectively at ($z = 2.0$ km, $x = 0.5$ km) and ($z = 2.0$ km, $x = 3.5$ km). The three seismograms for the corresponding models are shown in Fig. 2(a) for a 5 Hz Ricker source wavelet with a significant time-shift between blue (fast) or red (slow) single observed signals and the black synthetic signal inducing the cycle-skipping issue. For the L^2 norm, penalization-based cross-correlation and penalization-based deconvolution, the two in-phase blue and red adjoint sources shown in Figs 2 (b)–(d) have similar contribution nearby the initial trace, while the two opposite-phase blue and red adjoint sources of AWI and LAWI formulations, shown in Figs 2(e) and (f) will provide a different direction for updating the velocity. The amplitude variations of these adjoint sources induce patterns of the initial sensitivity kernels shown in Fig. 3. Let us underline that only the sensitivity kernels generated by AWI and LAWI strategies have opposite signs in the first Fresnel zone: leading to the correct direction for updating the velocity for fast (blue) and slow (red) target ones. Other methods fail to distinguish whether the starting velocity is faster or slower than the true one.

Fig. 4 displays the matching filters involved in this example. The time-shift between calculated and observed data can be inferred from the locations of the matching filters. The annihilator $\mathcal{P}(\tau)$, shown in Fig. 4(e), penalizes the energy located at non-zero lag and the penalization increases with time-shifts, expecting to reduce the objective function by moving $c(t)$ or $w(t)$ towards zero lag. However, the objective function of penalization-based cross-correlation method can be minimized by either pushing $c(t)$ towards zero lag or reducing the amplitude of $c(t)$. The objective function reduction of the penalization-based deconvolution method can also be achieved through these two ways, which is also described in Sun & Alkhalifah (2019a). Differently, with the normalization in eq. (18), the AWI formulation mainly relies on time-shift reduction to minimize the objective function. With the analogy provided by the physical concept of centroid frequency, the objective function of AWI workflow can be interpreted as computing the centroid time of the matching filter to implicitly measure the time-shift between observed and calculated data. In addition, the time–frequency analysis used in LAWI strategy makes it possible to capture instantaneous time-shift.

For an illustration of the reduction mechanisms followed by the five objective functions, let us plot the variations of these objective functions with respect to time-shift τ and amplitude scaling A on the calculated signal:

$$d_{\text{obs}}(t) = A \times d_{\text{cal}}(t + \tau). \tag{33}$$

All objective functions are centred at $\tau = 0$ (Fig. 5) and symmetric about the face of $\tau = 0$. The L^2-norm based objective function is non-convex with respect to time-shift. Although the objective functions of cross-correlation and deconvolution are convex, they are sensitive to amplitude changes, while AWI and LAWI objective functions are almost flat with amplitude variation, indicating that they are nearly insensitive to amplitude scaling and mainly sensitive to time-shifts. Namely, the model update of AWI or LAWI approaches is mainly driven by minimizing time-shift between observed and calculated data. Similar analysis about the AWI method can also be found in Li (2018, section 2.3) and Sun & Alkhalifah (2018, 2019a, b, 2020).

4.2 Analysis of multi-event time signals

Let us now consider various configurations of seismic events with different time-shifts and/or with a different number of events.

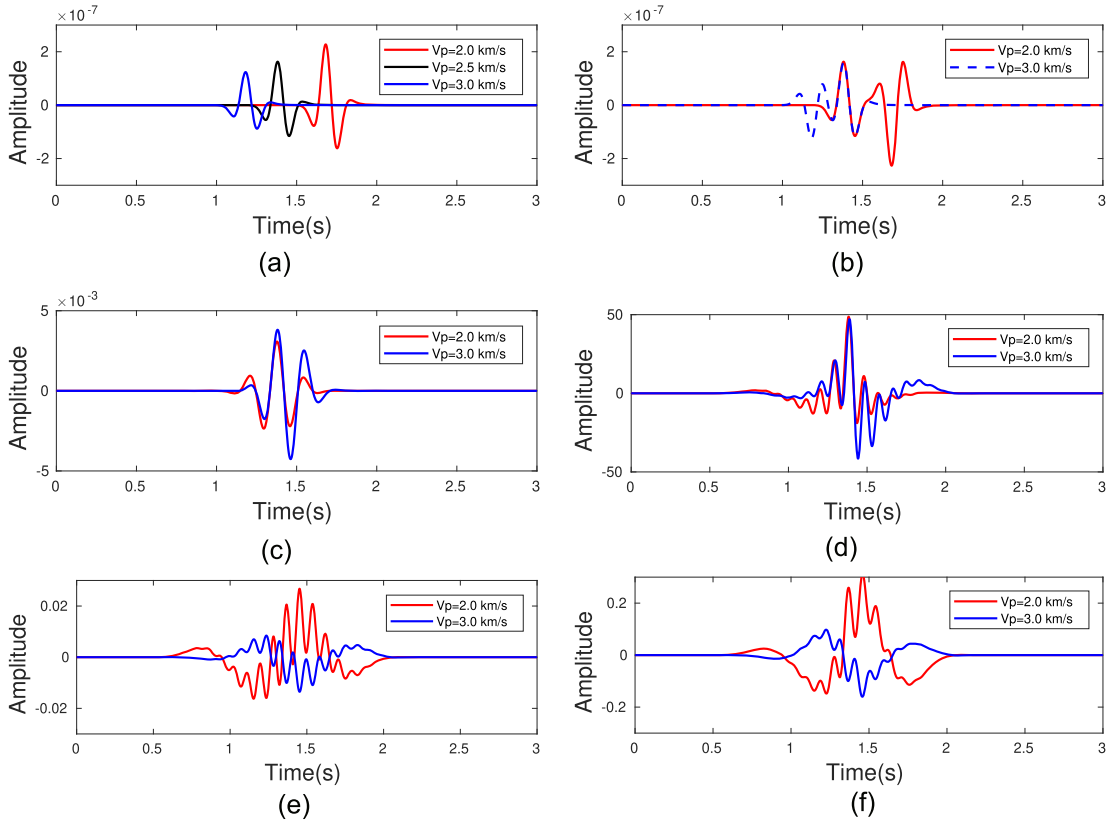

Figure 2. Three seismograms in three homogeneous models with velocity of 2.0, 2.5 and 3.0 km s^{-1}. The red and blue traces are two independent observed traces while the black one is the predicted trace in the initial model (a) to be moved onto the blue one in one experiment or onto the red in another experiment. Related blue and red adjoint sources of L^2 norm (b), cross-correlation (c), deconvolution (d) show, respectively, in-phase shape, while the blue and red adjoint sources of AWI (e), and LAWI (f) have, respectively, an opposite-phase pattern. Strong oscillations in the last two cases destroy somehow contributions of interference fringes.

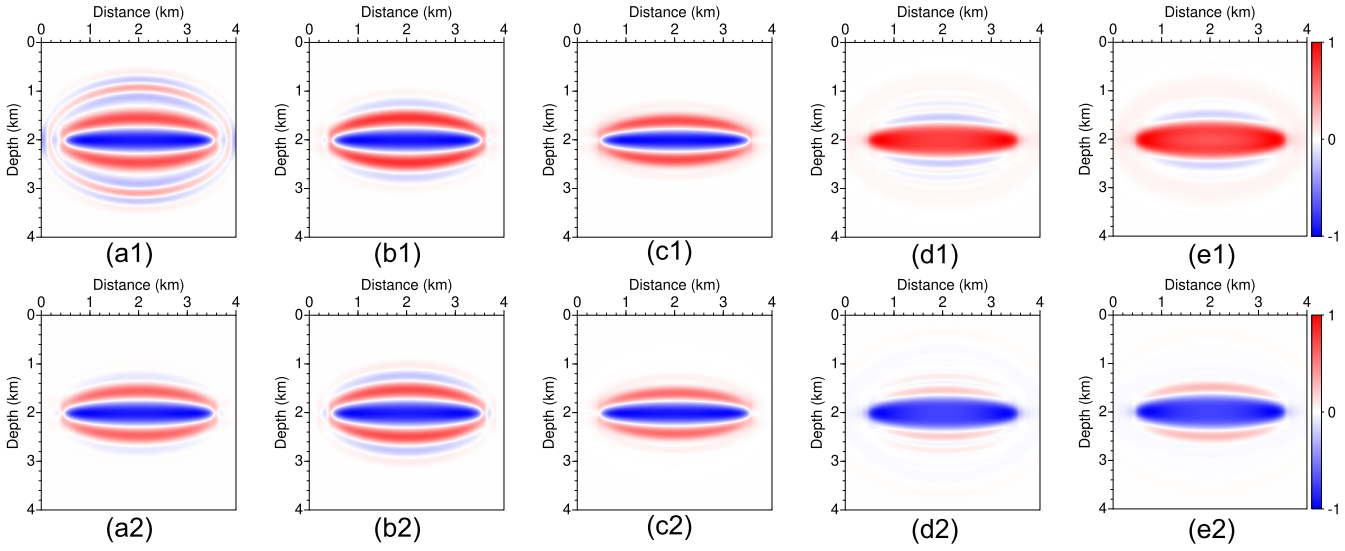

Figure 3. Initial sensitivity kernels (normalized) in a transmission configuration with single source/receiver pair with an initial velocity of 2.5 km s^{-1}. The top row is for a target slower velocity of 2.0 km s^{-1} while the bottom row is for a target faster velocity of 3.0 km s^{-1}. Amplitude variations of the adjoint sources induce the Fresnel interference pattern on different panels: respectively L^2 norm (a), cross-correlation (b), deconvolution (c), AWI (d) and LAWI (e). Only AWI and LAWI show the correct sign in the first-Fresnel zone which should be positive for the slow target (negative gradient) and negative for the fast target, which is an illustration of its strong capability to capture time-shifts.

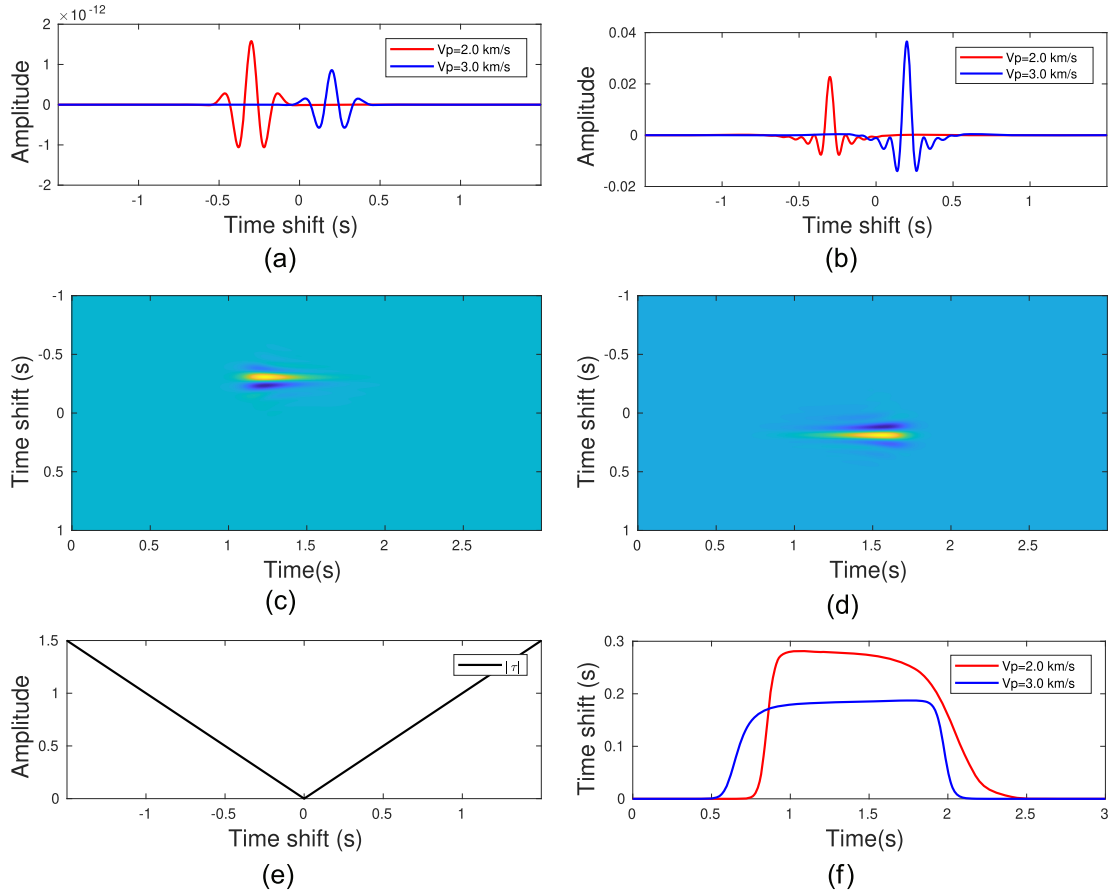

Figure 4. The connective function $c(\tau)$ (a), the global matching filter $w(\tau)$ (b), the local matching filters $w(t, \tau)$ relating to observed data with velocity of 2.0 km s^{-1} (c) and 3.0 km s^{-1} (d), the annihilator (e) and the estimated instantaneous time-shift (f). Except for the time-shift information, the local matching filter also can provide the time location information of the compared events. This makes it possible to estimate the instantaneous time-shift.

4.2.1 Different time-shifts of events

From the analysis, we know that the success of AWI and LAWI formulations rely on the time-shift estimation. However, the global matching filter, involved in the AWI workflow, is a time-invariant operator for all phase events occurring in the time signal. Let us consider two phase events, with different time-shifts, in observed and calculated signals. As shown in Fig. 6 (a), we assume that the first event has no time-shift, leading to a correct prediction while the second one has a 0.4 s time-shift. Amplitudes are normalized to one for all events. Although we can also observe two signals located at 0.0 and 0.4 s in the global matching filter (Fig. 6b), there are several other non-zero contributions mapping the first observed phase event onto the second predicted phase event: these contributions might be removed through iterations during the AWI procedure if successful. To make the matching filter more robustly evolve to zero-time lag, Sun & Alkhalifah (2020) design a misfit by minimizing the mean and information entropy of the matching filter distribution from a statistics point of view. Differently, we use a non-stationary matching filter to build the relation between observed and calculated data. Apparently, the interpretation of the local maching filter (Fig. 6c) is much simpler with a zero time-shift for the first event and a 0.4 s time-shift for the second event. Finally, Fig. 6(d) shows both AWI and LAWI objective functions: the AWI objective function becomes non-convex with respect to local time-shifts, while LAWI objective function remains convex. One can see the advantage of a local analysis in processing complex seismic data.

4.2.2 Unequal numbers of events

In synthetics tests of waveform inversion for exploration applications, the observed data usually contains direct, diving and reflected waves. When inversion starts from a smooth velocity model, the calculated data may only contain direct and diving waves. Thus the number of events in calculated data and observed data can not be matched. This case can be mimicked with the signals presented in Fig. 7(a). In the AWI procedure, the global matching filter can take all events in observed data into account (Fig. 7b), whereas LAWI approach only accounts for the event in observed data locally coherent to the event occurring in calculated data. One can observe that some non-physical events occur in the AWI adjoint source, which indicates that the AWI formulation does not properly make use of the information of second event. From the test in the Section 4.2.1, we know that it is useful to estimate instantaneous time-shift in waveform inversion. While directly applying the LAWI procedure in this case, a non-convergence issue is met, because the estimated time-shifts are initially related mainly to diving and direct waves. However, with model update, reflections will arise in the predicted data, which leads to time-shifts of the new reflections (not existing at the beginning) and may make the objective function increase.

To avoid such a non-convergence issue in LAWI applications, one can use a mask function to fix the compared events for several iterations. Specifically, diving and direct waves are first taken into account. With model updates, the mask is then gradually enlarged to gently include more reflected waves. This strategy is known as the data-domain layer stripping (Wang & Rao 2009; Pladys *et al.*

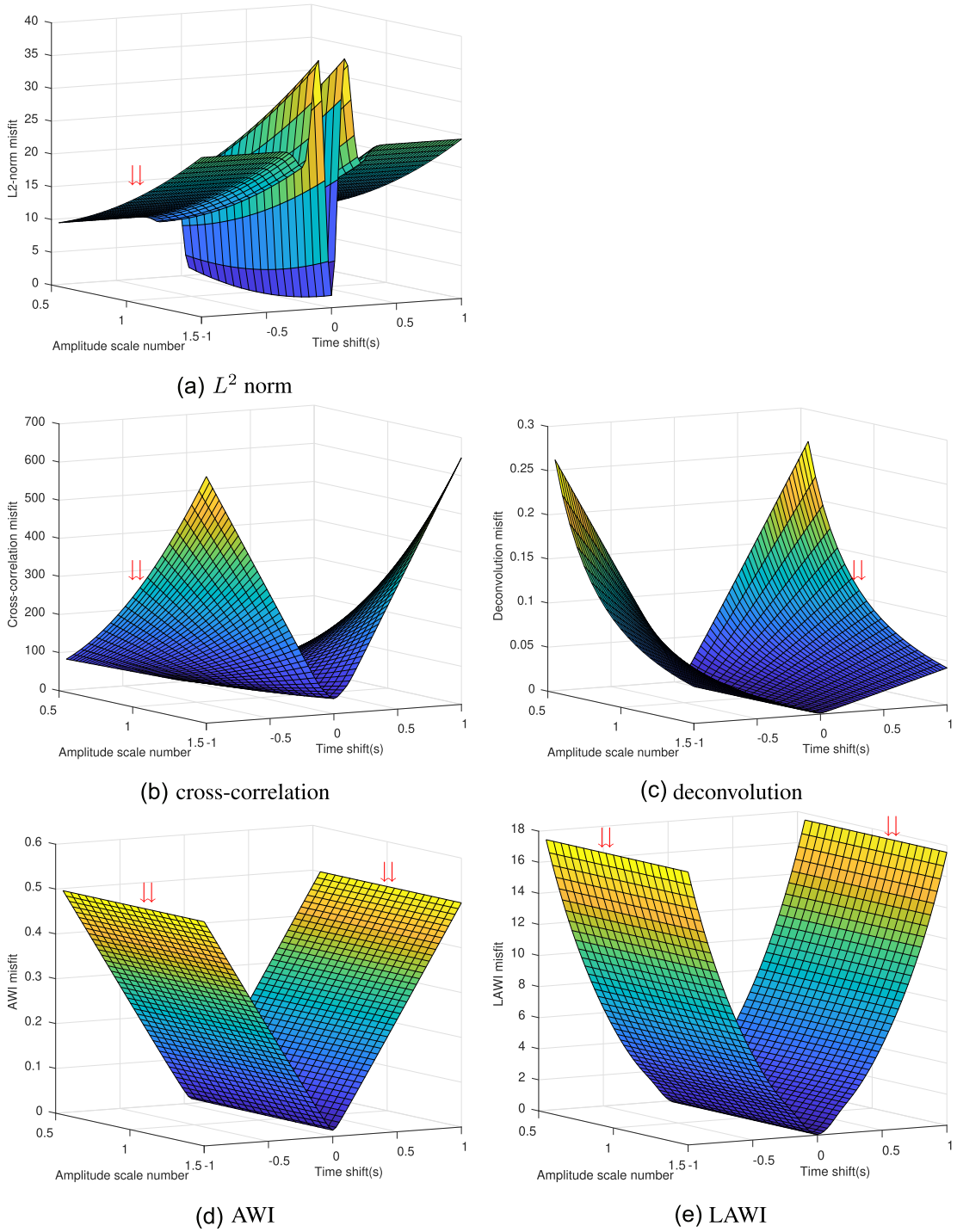

Figure 5. Misfit functions with respect to time-shift τ and amplitude scaling A on the predicted data. The objective function of the conventional L^2 norm is non-convex (a). The objective functions of cross-correlation (b), and deconvolution (c) are convex but sensitive to amplitude change, whereas the objective functions of AWI (d) and LAWI (e) capture essentially time-shifts. Red double arrows highlight amplitude influence of penalization-based cross-correlation and deconvolution.

2022). Let us point out that, for field-data applications, the multi-scale strategy (Bunks *et al.* 1995) is generally adopted to deal with the cycle-skipping issue: inversion starts with low-frequency data and gradually includes more high-frequency content. In realistic low-frequency data, the usable signals are mainly diving and direct waves, and reflections gradually come into the scene with frequency increases, as shown later in the Chevron benchmark data (Fig. 21). Namely, the multiscale strategy can automatically generate a suitable mask for the LAWI application. Therefore, for realistic data applications, manual mask design becomes not mandatory when the multi-scale strategy is used in LAWI approach. More practical details will be given by the Chevron benchmark data later. In addition,

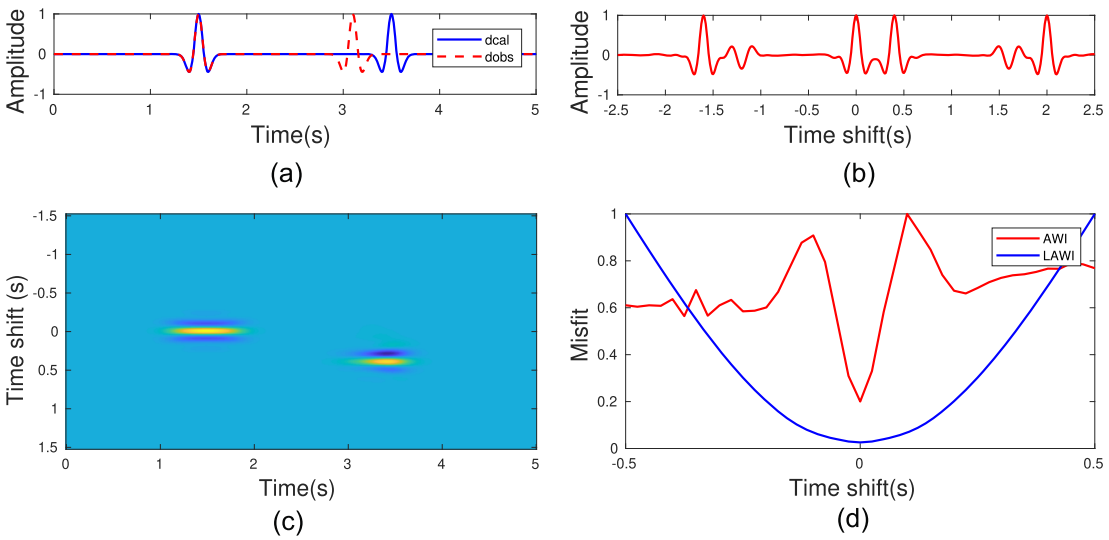

Figure 6. Comparison of AWI and LAWI using two phase events in the time signal: observed in red and predicted in blue data with 0 and 0.4 s time-shifts of the 5 Hz Ricker wavelet (a), the corresponding global matching filter (b) and the local matching filter $w(t, \tau)$ (c), normalized misfit function with different time-shifts (d), in which the first phase event is kept fixed while the second event is shifted with time delays from 3 to 4 s.

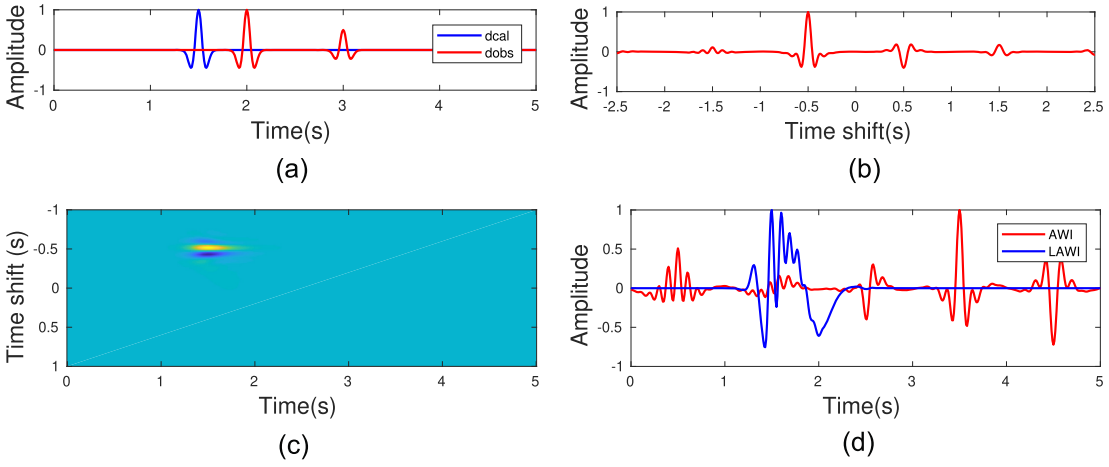

Figure 7. The predicted data in blue has less phase events than observed data in red, and the first event suffers from a 0.5 s time-shift (a), the corresponding global matching filter (b) and the local matching filter $w(t, \tau)$ (c), adjoint sources (d). AWI tries to map also the second observed event with the single predicted event, while LAWI only maps the first observed event into the single predicted event, leading to a rather different adjoint source.

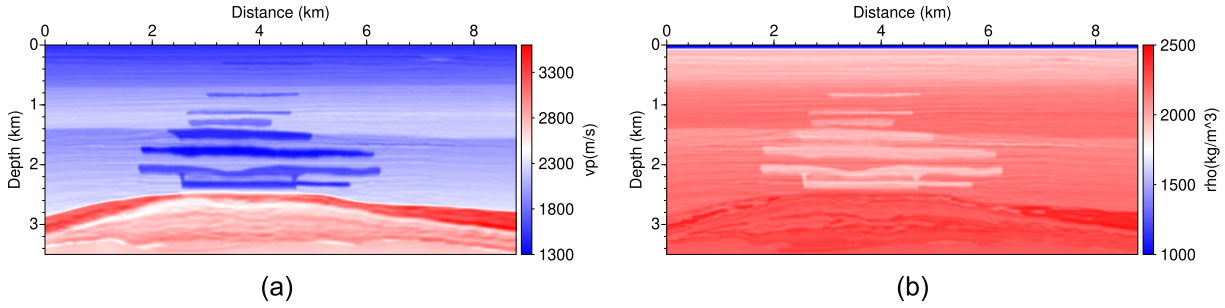

Figure 8. 2-D Valhall models: (a) P-wave velocity model, (b) density model.

we can combine LAWI procedure with reflection FWI formulation (Xu *et al.* 2012; Brossier *et al.* 2015; Zhou *et al.* 2015) for updating the deep zone of the model, in which special care needed to treat this issue becomes unnecessary, as (more or less) the same number of main reflections can be produced by migration results.

5 FULL-WAVEFORM INVERSION APPLICATIONS

In this section, we consider more complex situations for highlighting similarities and differences between AWI and LAWI workflows with two different synthetic models: the 2-D Valhall model for a standard

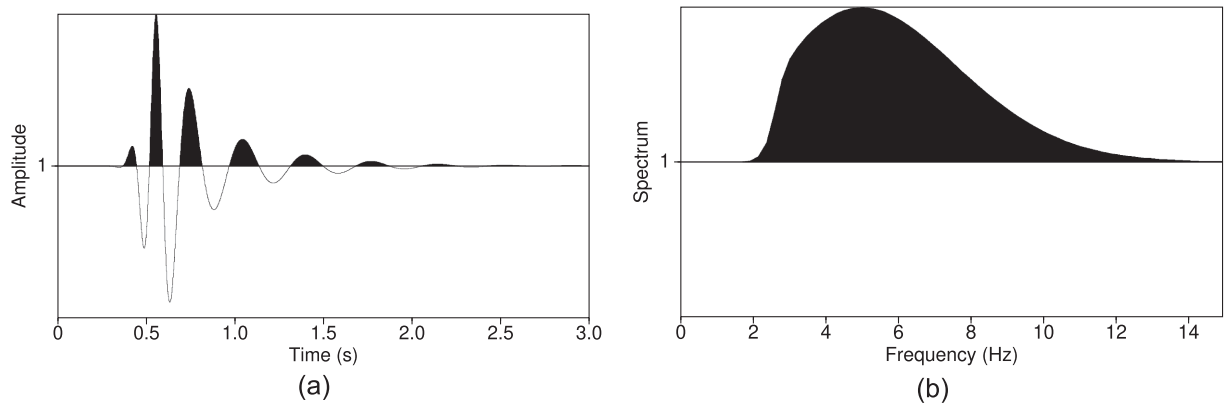

Figure 9. The source wavelet (a) used to generate the synthetic data set on the Valhall model: a 5 Hz Ricker wavelet with the application of a zero-phase Butterworth filter below 2.0 Hz. (b) its related amplitude spectrum.

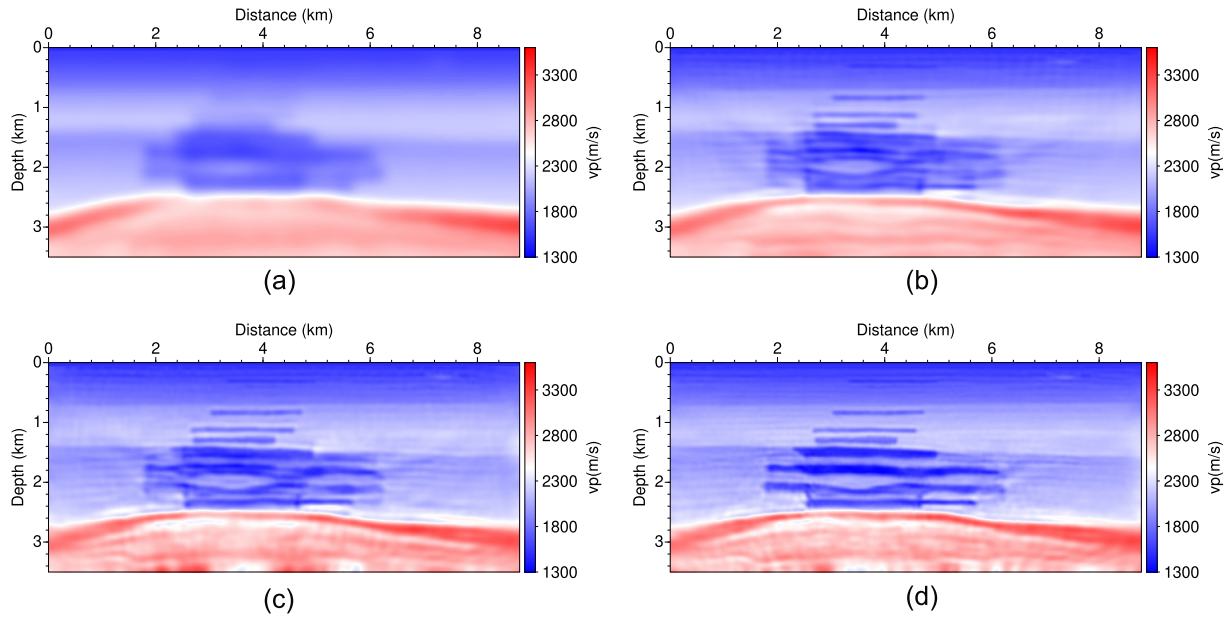

Figure 10. Case 1: relatively accurate initial model (a), and inversion results after 50 iterations using standard FWI (b), AWI (c) and LAWI (d).

marine environment and the Chevron benchmark data with strong localized near-surface variations, often met for on-shore data.

5.1 2-D Valhall model

The velocity and density models are presented in the Fig. 8, which represent the shallow water environment of the real Valhall oil field (Sirgue *et al.* 2010). They are defined on a Cartesian grid of 281 × 704 with a spatial interval of 12.5 m in both directions. Since the ratio between the maximum offset and depth is only 2.5, direct and diving waves could not illuminate depth below 1.5–2.0 km. We use a fixed spread surface acquisition with 32 sources and 352 receivers at a depth of 25 m. The synthetic observed data are generated with a 5 Hz Ricker wavelet and the energy below 2 Hz is removed. The low-cut wavelet and its spectrum are displayed in Figs 9(a) and (b). The effective time duration of the source wavelet can be regarded as 1.2 s, the time-shift between observed and predicted data should be less than 0.2 s in such a case. The recommended width parameter in the Gabor transform is $\sigma = 0.35$ s. Besides, the two scaling values for water-level parameters ϵ and η are chosen as 10^{-3} and 10^{-2} for this example. Optimization is carried out with the preconditioned ℓ-BFGS algorithm ($\ell = 5$), and the pre-conditioner is based on energy compensation (Yong *et al.* 2022). For synthetic tests, there is an interest to track the model error evolution with iterations. A relative L_1-norm model error is used here. For a given discrete velocity model of M degrees of freedom, such a model error is computed as

$$E_V = \frac{100}{M} \sum_{i=1}^{M} \frac{|V_i - V_i^{\text{true}}|}{|V_i^{\text{true}}|}. \tag{34}$$

For testing the robustness of these three methods (standard FWI, AWI and LAWI approaches) with respect to the local minima issue, we implement different inversions starting from three velocity models. Fig. 10(a) displays the first starting model that are obtained by applying a triangle smoother (Claerbout & Fomel 2014) with a radius of 250 m to the true model. The other two starting models are generated by repeating the same Gaussian smoother 10 times (Fig. 11a) and 15 times (Fig. 12a). For an illustration of how LAWI works during inversion, we keep the true density model that allows the same number of events in predicted and observed data. We will use a smooth density model as the starting model at the last part

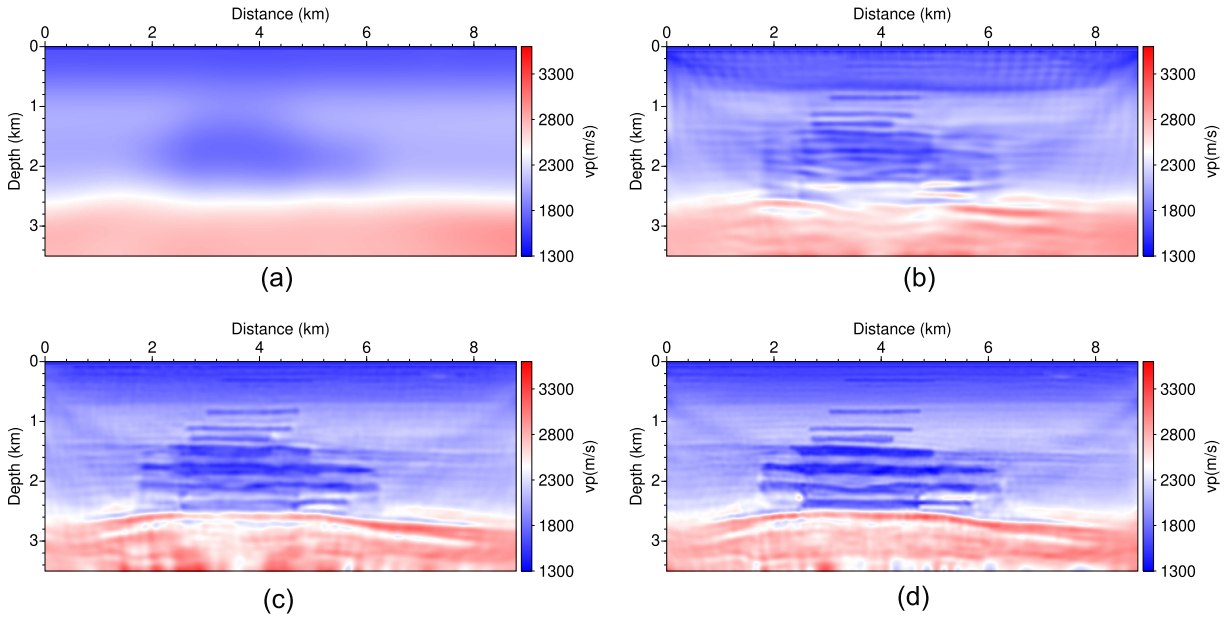

Figure 11. Case 2: less accurate initial model (a), and inversion results after 50 iterations using standard FWI (b), AWI (c), and LAWI (d).

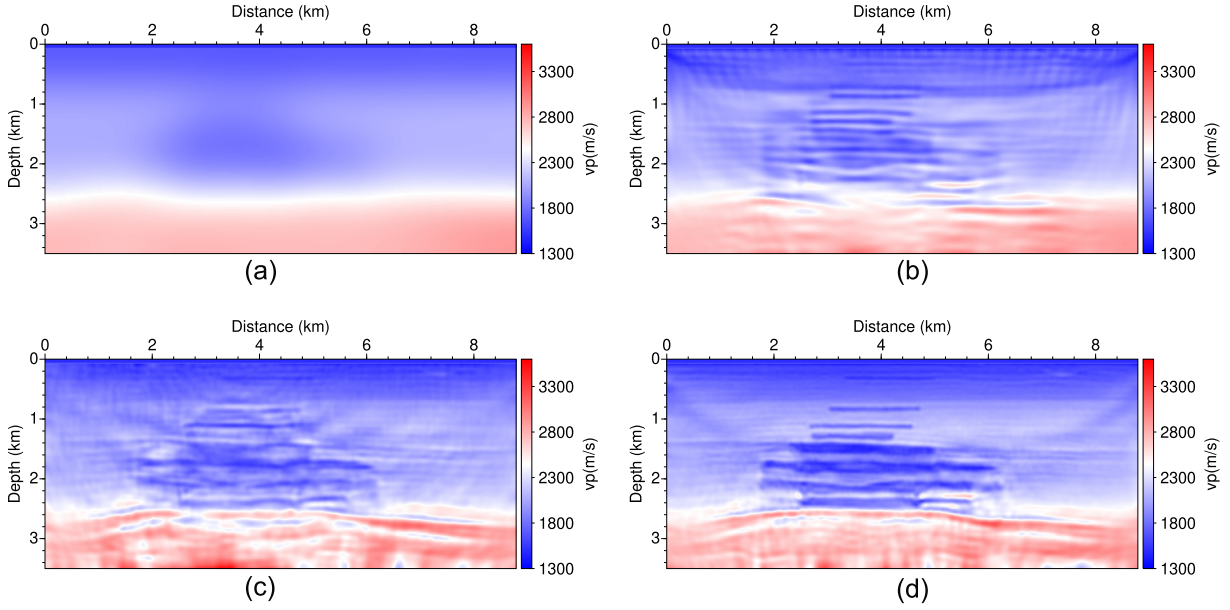

Figure 12. Case 3: poor initial model (a), and inversion results after 50 iterations using standard FWI (b), AWI (c) and LAWI (d).

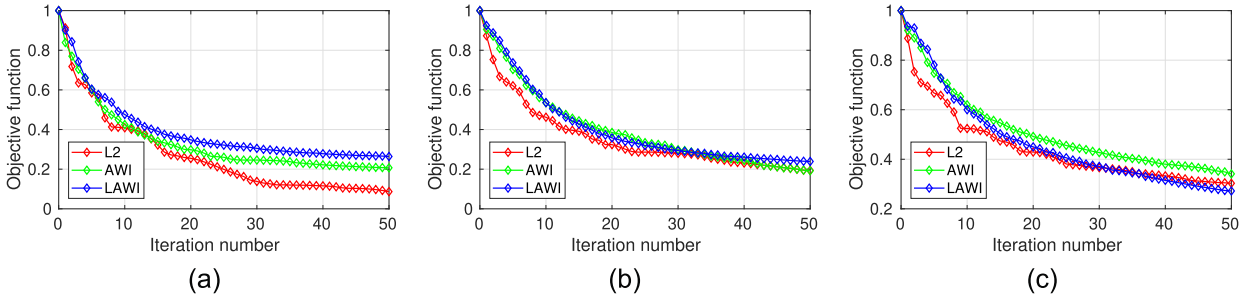

Figure 13. Comparison of normalized data misfits decrease with iterations: case 1 (a), case 2 (b) and case 3 (c).

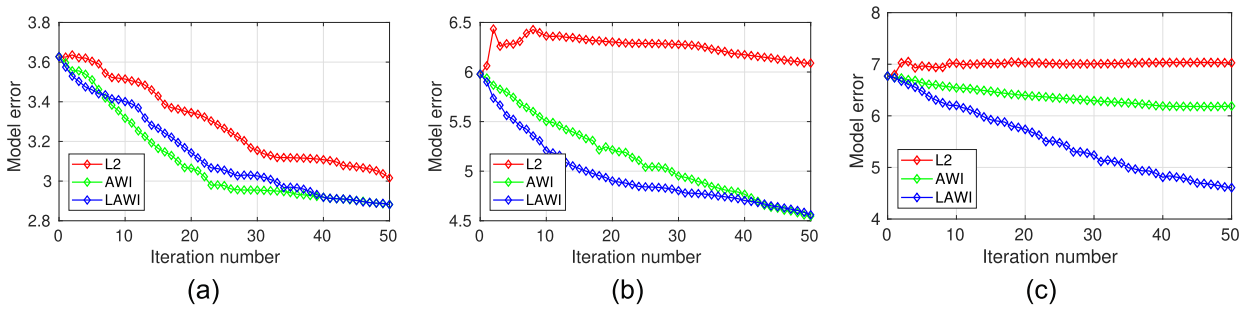

Figure 14. Comparison of model error decrease with iterations: case 1 (a), case 2 (b) and case 3 (c).

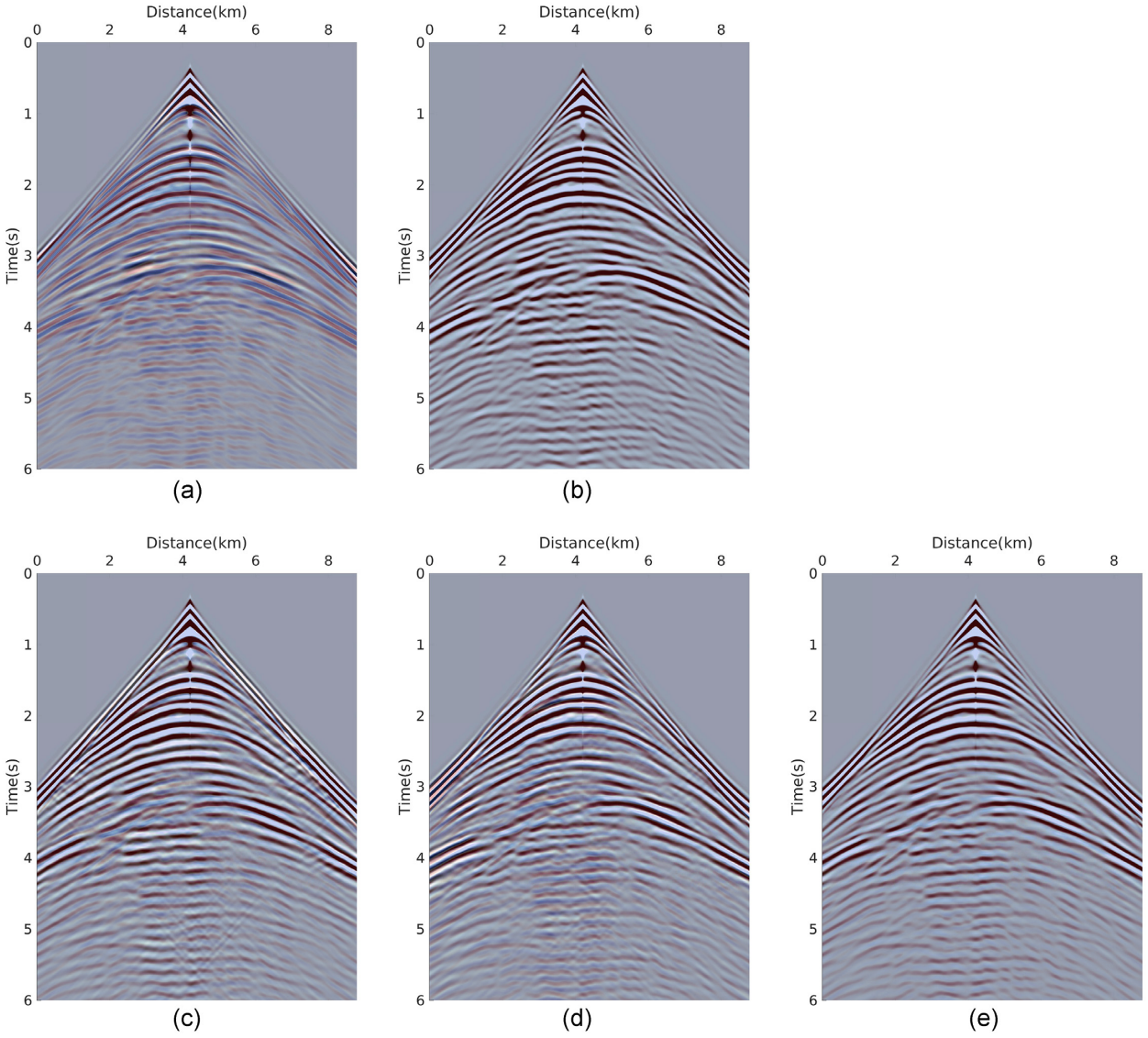

Figure 15. Data quality check in the case of the poor initial model: the observed data is plotted in blue-white-red colour as background, and compared data are overlapped in greyscale with transparency. The compared data are generated with initial model (a), true mode (b), and reconstructed models by FWI (c), AWI (d), LAWI (e). When the compared data match the observed data, colourful pixels will be covered by grey pixels.

of this subsection, and introduce a weight-windowing strategy to handle the fore-mentioned non-convergence issue.

Fig. 13 shows that all objective functions decrease with iterations in the three cases. However, such a decrease does not mean that reconstructed velocity models converge to the true one. When the starting model is close to the true model, the main geological structures have been reconstructed by three methods after 50 iterations (Fig. 10). Besides, from Fig. 14(a), we can observe that, compared to AWI and LAWI approaches, the conventional FWI workflow seems to have a slow convergence rate in reconstructing

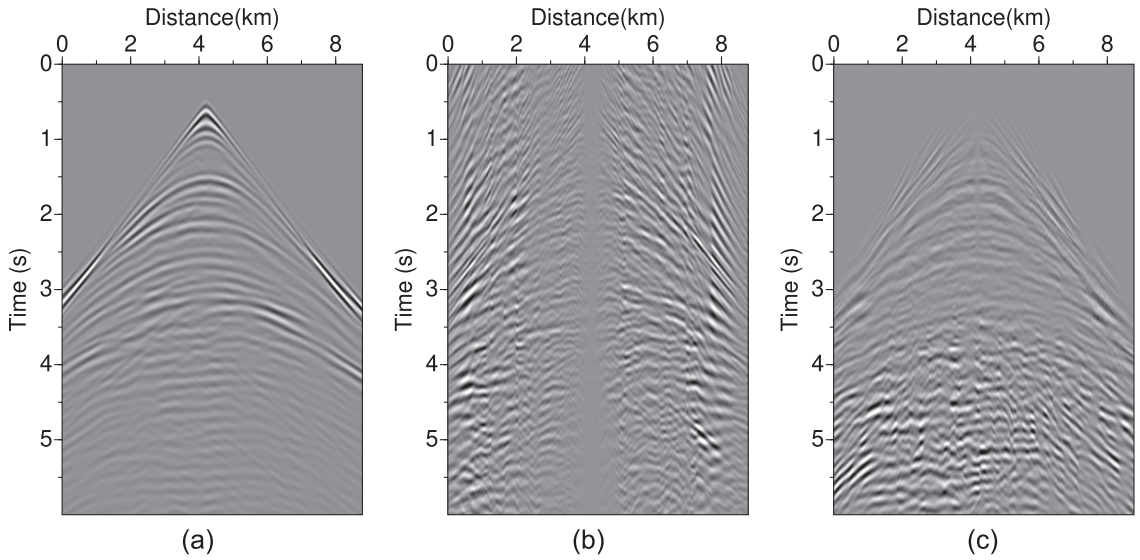

Figure 16. Adjoint sources for the first iteration in the case of the poor initial model: (a) standard FWI, (b) AWI and (c) LAWI.

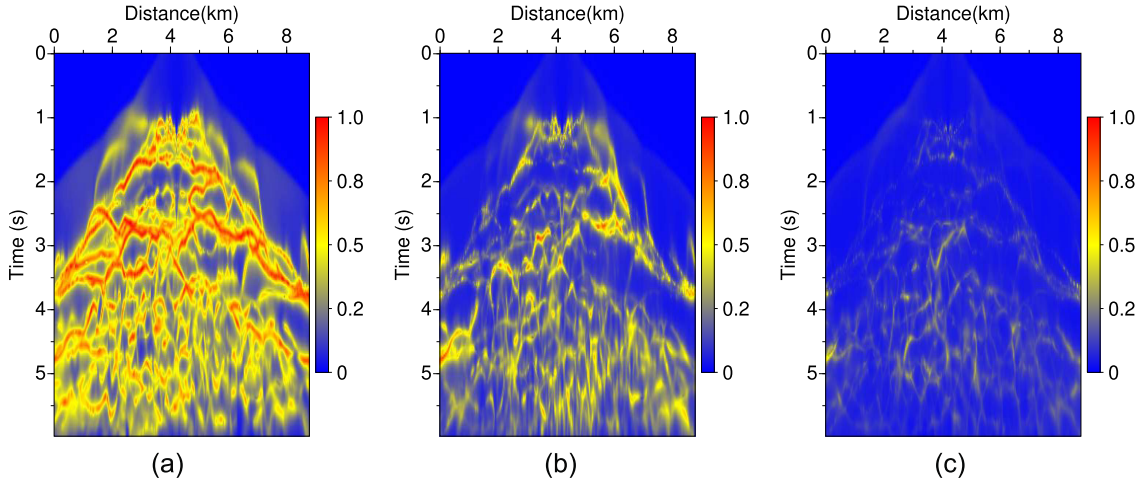

Figure 17. The evolution of the centroid time of the local matching filter in the case of the poor initial model: the 1st iteration (a), the 15th iteration (b) and the 50th iteration (c).

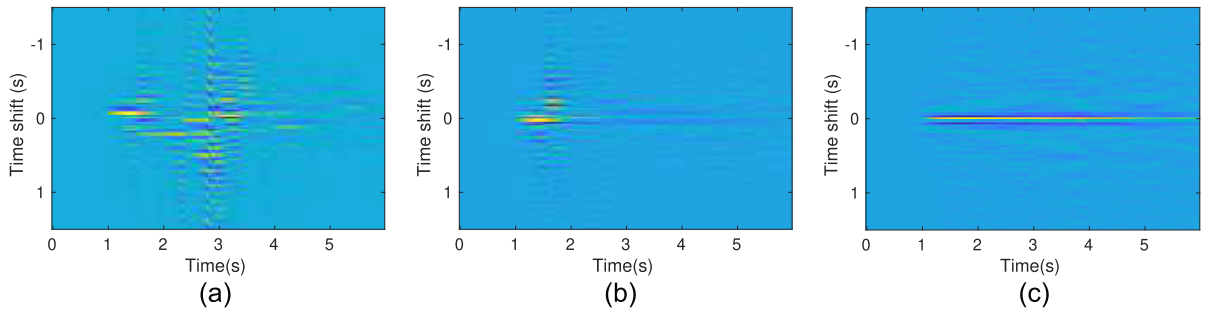

Figure 18. The evolution of the local matching filter $w(t, \tau)$ ($x = 3$ km): the 1st iteration (a), the 15th iteration (b) and the 50th iteration (c).

the low-velocity gas layer shown in the middle of Fig. 10(b) due to the lack of low frequency content in the data. When the starting model becomes less accurate (Fig. 11a), conventional FWI is trapped into local minima and fails to recover the low-velocity gas layer. An incorrect low-wavenumber anomaly occurs in the shallow part (Fig. 11b). The results (Figs 11c and d) obtained by AWI and LAWI approaches are similar with those obtained by the first test. Figs 12(b) and (c) displays the inversion results from the most inaccurate starting model of FWI and AWI procedures. Due to the high-velocity contrast between the gas layer and sediment, the relation between the waveform and the velocity perturbation is strongly non-linear. Although AWI can reduce the risk of getting stuck in local minima, it cannot fully recover the gas layer. From Fig. 14(c), we know that increasing iterations can hardly improve the final result

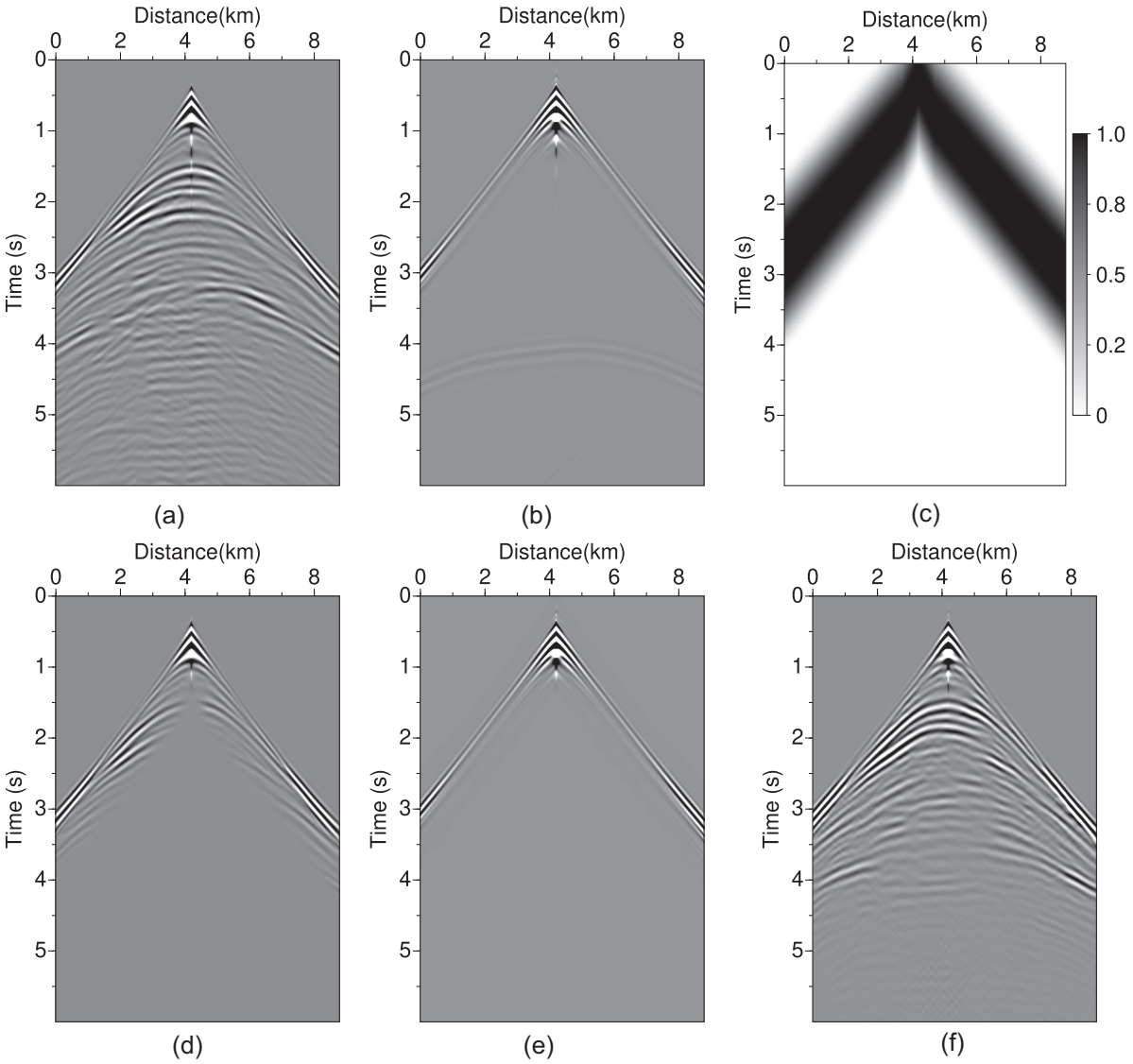

Figure 19. The weighting-window strategy: (a) the observed data, (b) the predicted data generated by using smooth velocity and density models, (c) the weights for data windowing, (d) the weighted observed data, (e) the weighted predicted data and (f) the predicted data generated by the reconstructed velocity presented in Fig. 20(e) and smooth density models.

presented in Fig. 12(c). This indicates that the AWI approach gets trapped into local minima. The *P*-wave velocity model obtained by the LAWI formulation (Fig. 12d) is the closest to the exact model, thanks to the non-stationary nature between the compared predicted and observed data taken into account. In addition, Fig. 15 displays the data quality check in the third case: one can see that the calculated data in the final model match quite well the observed data.

For the last case with the most inaccurate initial model, the adjoint sources are displayed in Fig. 16 for the first iteration. Non-zero acausal contributions can be observed in the adjoint source time function of AWI due to the global matching filter (Fig. 16b). Local matching filter attempts to estimate the time-shifts using locally coherent events, which reduce the effects of delayed events, preventing acausal non-zero contributions (Fig. 16c). Fig. 17 displays the decrease of the instantaneous centroid times with iterations. At the same time, the local matching filter converges towards zero lag during iterations as does the global matching filter when successful (Fig. 18). It is necessary to point out that the instantaneous centroid times cannot reflect exact time-shifts between calculated and observed data (Fig. 17). The reason is, for this complex data the multi-phase events interactively present in the Gaussian window, and the Gabor convolutional model is not able to map these with a simple and focused filter. In fact, the local matching filter becomes complex and its energy is diffusely distributed along the time-shift axis (vertical axis in Fig. 18). The time-shifts, implicitly estimated by the instantaneous centroid times of the local matching filter allow convergence of the inversion. One may consider shortening the window width for a more local comparison of signals. However, it would move to the wiggle matching, which is the basic feature of FWI procedure.

Now let us consider a more realistic situation where the starting velocity and density models are both very smooth. Here, we apply the same smoothing filter, used for generating the third smooth velocity model (Fig. 12a), to the true density model. The generated smooth density and the third smooth velocity are used as starting

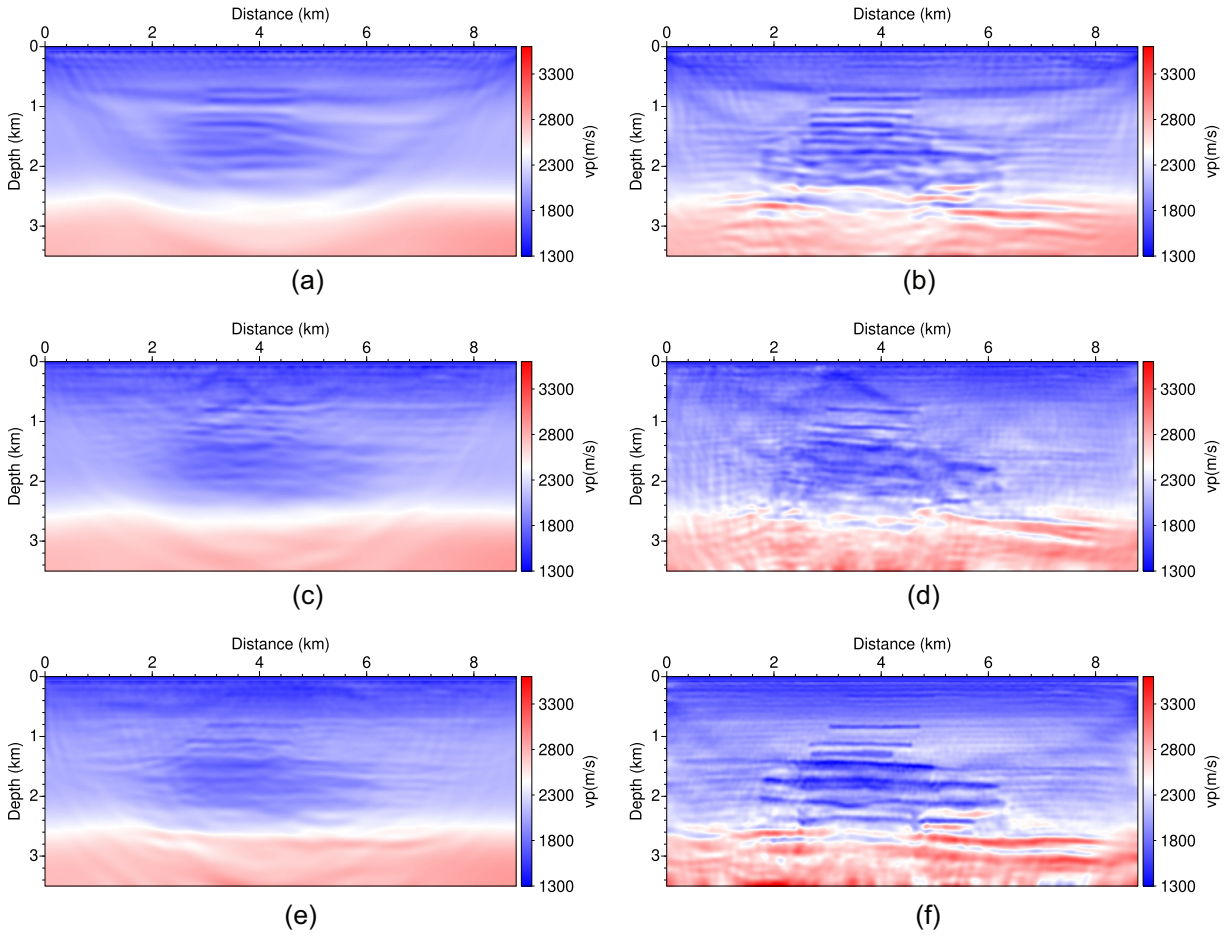

Figure 20. Inversion results starting with the third velocity model and a smooth density model. Pictures on the left-hand side show the recovered velocity models computed with weighted data after 20 iterations: (a) FWI, (c) AWI and (e) LAWI. Pictures on the right-hand side show the recovered velocity models subsequently computed with full data after 50 iterations: (b) FWI, (d) AWI and (f) LAWI.

models. During the inversion, we only update velocity model. The predicted data is shown in Fig. 19(b). One can notice that there is a lot of reflections in the observed data, while they are absent in the predicted data as we start from the smooth density and velocity models. From the Section 4.2.2, we know that LAWI procedure is driven by matching the predicted data, therefore it can not take all reflections in the observed data into consideration at the beginning of inversion. If we directly apply LAWI formulation, with model update, reflections will occur in the predicted data. The newly arising reflections could lead to an increase of misfit function and cause a non-convergence issue, because more events are taken into account. In order to avoid this issue, we can use a simple weighting-window strategy to determine signals to be compared. Specifically, we design weights to extract direct waves and diving waves from predicted and observed data, as shown in Fig. 19(c). At the first stage, we focus on inverting the data in the windowed area. With model update, reflections would gradually be presented in the predicted data, then we can use full data for LAWI procedure. This strategy is known as data-domain layer stripping, and it has been widely used to mitigate the cycle-skipping issue (Wang & Rao 2009; Pladys *et al.* 2022).

Although the weights are designed to interpret the direct and diving waves of predicted data (Fig. 19e), some reflections in the observed data also appears in the weighting window (Fig. 19d). As a result, we can see layered structures appearing in the reconstructed velocity model after 20 iterations by LAWI, shown in Fig. 20(e). The corresponding predicted data to this velocity model are depicted in Fig. 19(f). We can find that the reflections and multiples are fully distributed in the predicted seismograms, thus it is ready to move to the second stage: using full data in the LAWI approach. Fig. 20(f) presents the reconstructed velocity model with subsequent 50 iterations. For comparison, we also show the inversion results by FWI and AWI formulations with the same workflow. Due to the inaccuracy of the density model, the quality of reconstructed velocity models has a slight downgrade. Overall, the velocity model generated by the LAWI workflow is still the closest to the true one. From the weighted data test (the first stage), we can see that the LAWI formulation can account for all events in the given weighting-window. Although the number of events at the beginning is not the same in the predicted and observed data, the LAWI approach has the capacity to deal with this circumstance. To make better use of LAWI, we can combine it with the reflection FWI (Xu *et al.* 2012; Brossier *et al.* 2015; Zhou *et al.* 2015) for velocity update of deep areas that diving waves can hardly reach.

5.2 Chevron 2014 benchmark data

In order to further investigate the performance of the LAWI approach in dealing with the cycle-skipping issue, we will apply it to Chevron 2014 blind test data, which has a strong noise in the low-frequency band, especially in 0–3 Hz. Such a synthetic data set

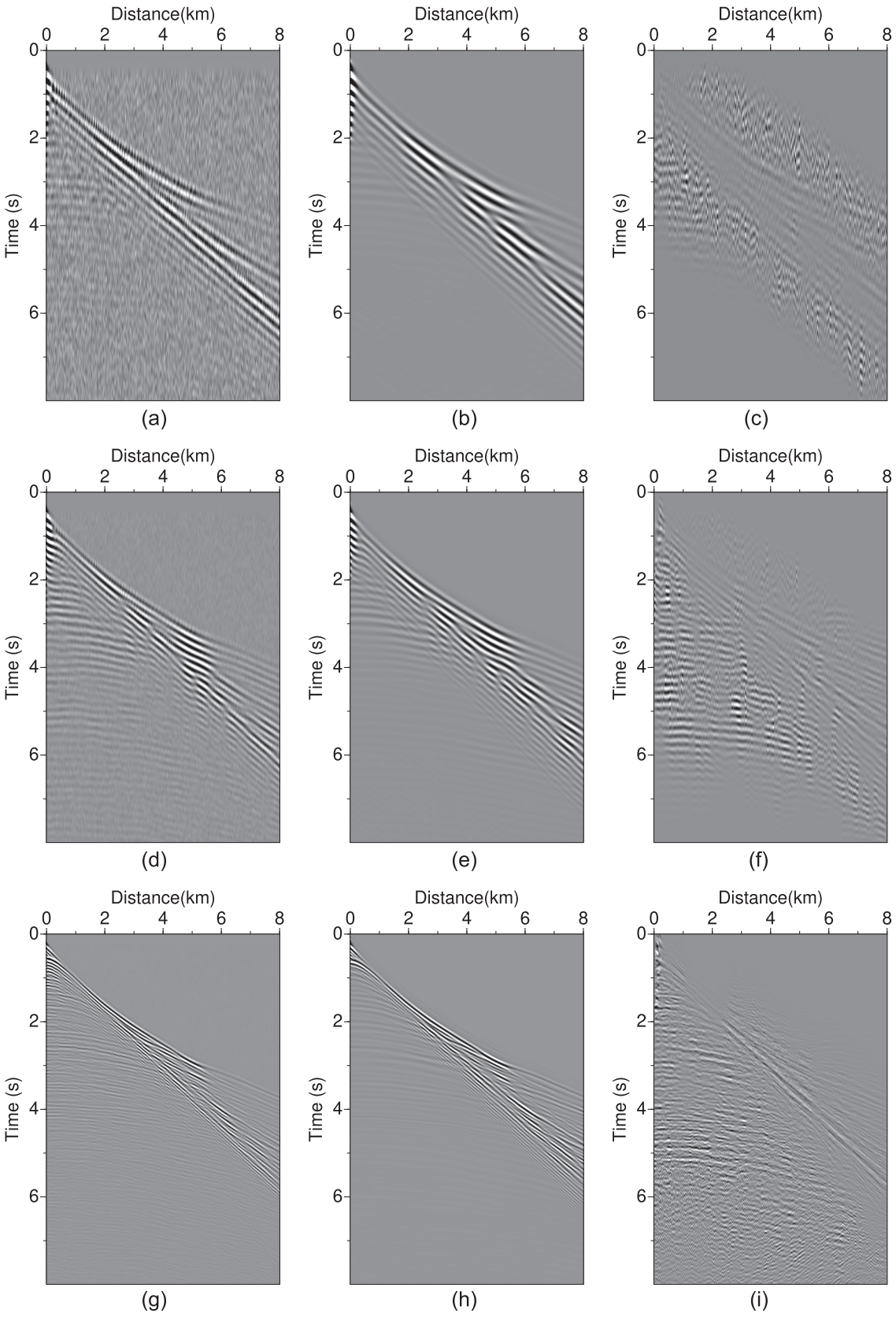

Figure 21. The observed data: 2–4 Hz (a), 2–6 Hz (d) and 2–15 Hz (g). The calculated data (b, e, h) with the reconstructed models using data on the left-hand side. The pictures (c, f, i) on the 3rd column display the corresponding adjoint sources of LAWI for the first iteration, which indicates that LAWI can gradually take reflections into account with multi-scale strategy.

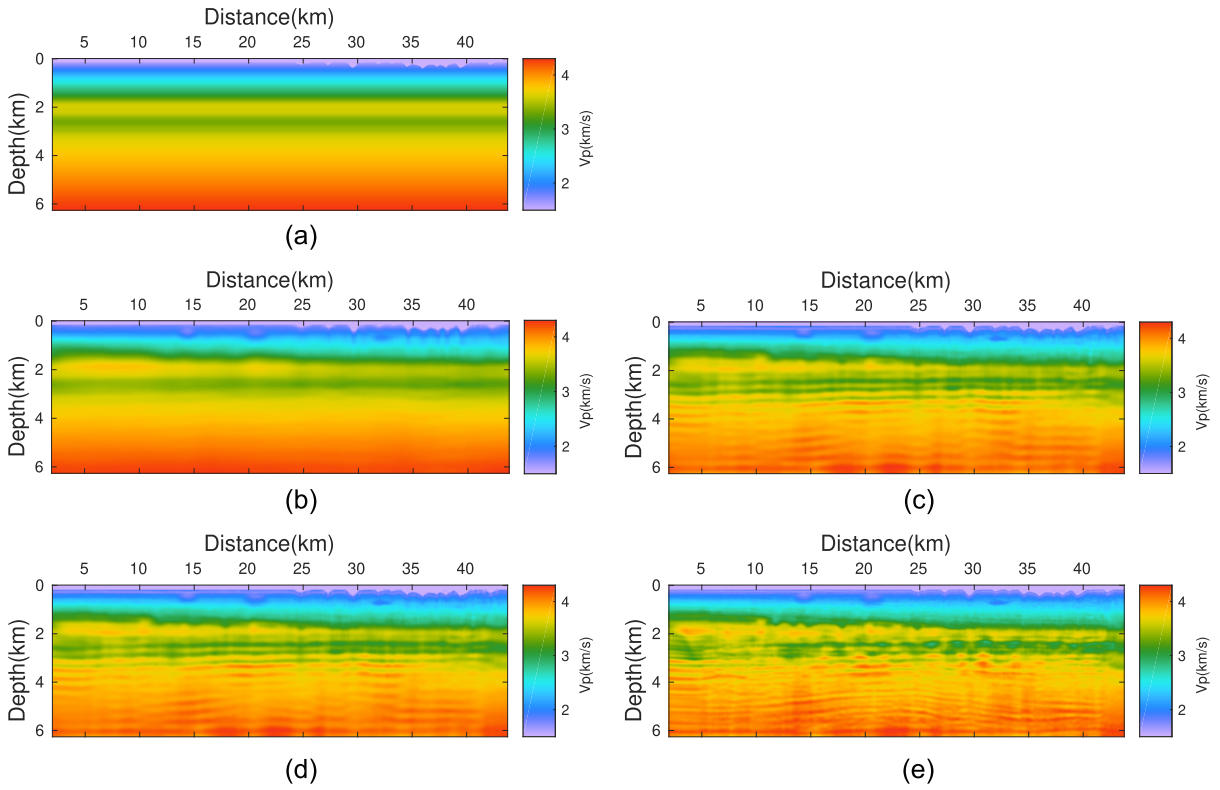

Figure 22. Velocity models: initial model (a), the reconstructed velocity with AWI method using the data of the frequency band 2–4 Hz (b), 2–6 Hz (c), 2–10 Hz (d) and 2–15 Hz (e).

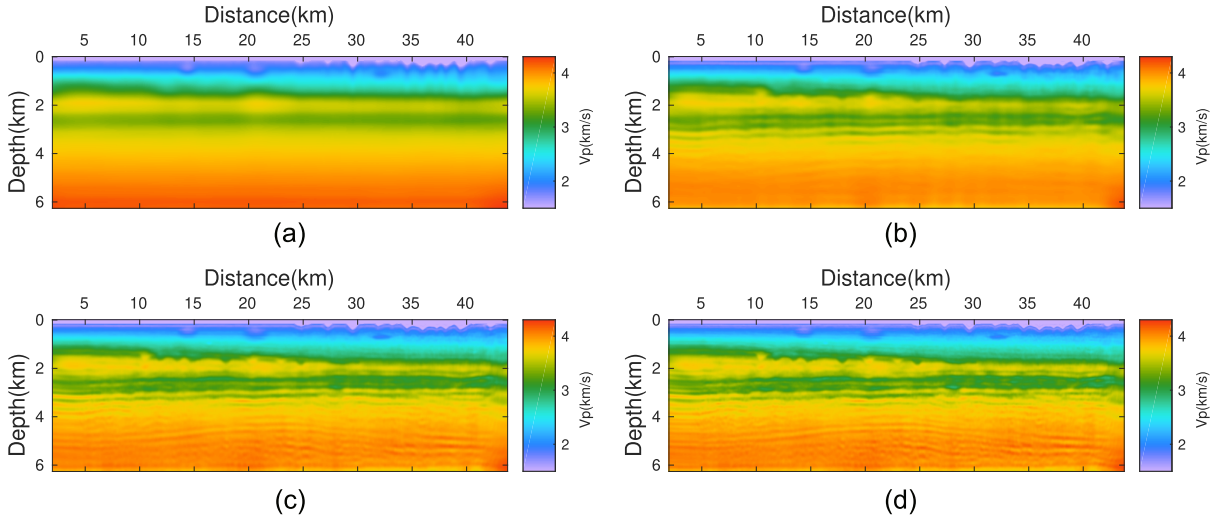

Figure 23. Reconstructed velocity models with LAWI method using the data of the frequency band 2–4 Hz (a), 2–6 Hz (b), 2–10 Hz (c) and 2–15 Hz (d).

mimics an exploration field data set with missing low-frequency content. Besides, the 'observed' data are generated with an elastic modelling engine. The first column in Fig. 21 displays the data in different frequency bands. The original data set contains 1600 shots with a single-side acquisition, and the maximum offset is 8 km. This makes it difficult to reconstruct velocity below 2 km, which is hardly probed by diving waves (Vigh *et al.* 2016). In our experiment, 160 shots, equally extracted from the original data, are used for inversion. The given initial model is presented in Fig. 22(a), which only provides layer information below sea water. To mitigate the cycle-skipping issue, the widely used frequency continuation strategy, proposed by Bunks *et al.* (1995), is adopted in this case study. The wavelets are estimated by solving a linear problem in the frequency domain (Pratt 1999). We compute one wavelet for one frequency-band data and update it after finishing iterations of each frequency band with the latest velocity model. All wavelets used in LAWI are shown in Fig. 24. In total, seven frequency bands are used in inversion, and they are 2–4 Hz, 2–5 Hz, 2–6 Hz, 2–8 Hz, 2–10 Hz, 2–12 Hz and 2–15 Hz. The same processing of frequency continuation and wavelet estimation are used in AWI. The window width parameter σ in Gabor transform is chosen as 0.5 s in this study. The two scaling numbers in water-level values ϵ and η are,

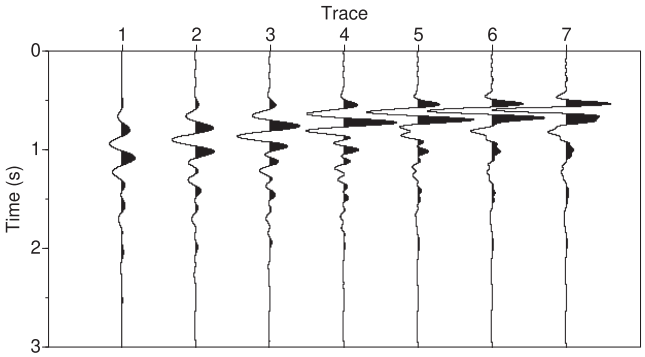

Figure 24. The estimated wavelets in different frequency bands: 2–4 Hz, 2–5 Hz, 2–6 Hz, 2–8 Hz, 2–10 Hz, 2–12 Hz and 2–15 Hz (from left to right).

respectively, 10^{-2} and 10^{-1} for the first two frequency-band tests, and they are modified to 10^{-3} and 10^{-2} for other frequency-band tests. Besides, ℓ-BFGS method is used for updating the velocity model, and for the first three stages, the iterations are 20 and are changed to 30 in the subsequent frequency bands. Note that L^2-norm based FWI fails to generate geologically meaningful velocity with this workflow (Métivier *et al.* 2016; Yong *et al.* 2019). Therefore, we only show the results generated by AWI and LAWI formulations in this paper. It is necessary to point out that we do not use the weighting-window strategy in this case test, because the usable signals in the low-frequency observed data are mainly direct and diving waves. Restricting the compared areas in seismic profiles becomes not mandatory, and the reflections are gradually taken into account with the increase of the frequency band.

Figs 22(b)–(e) displays the reconstructed velocity models by the AWI workflow at four stages. The inversion results obtained by LAWI at the same stage are shown in Figs 23(a)–(d). Both methods generate informative velocity models, and the geological structures above 2.5 km are similar. However, it seems that there is a low-velocity zone located in 2.5–3 km depth, which has been reported in previous studies (Warner & Guasch 2015; Wu & Alkhalifah 2015; Vigh *et al.* 2016; Sun & Alkhalifah 2019a). Compared with the AWI procedure, the LAWI formulation can better reconstruct this low-velocity layer, which can also be observed in the well log (Fig. 25). From Fig. 25, we can see that the reconstructed models match the true velocity well above 2.5 km. One may notice that the low-velocity layer at the right place has better update, because the update is related to the thickness of the high-velocity overburden, which decreases from left to right. Since AWI and LAWI approaches are less sensitive to amplitude than the FWI formulation and that only low-frequency data are used, the reconstructed models do not have a high resolution as that of the well-log data. In addition, the given data is generated in elastic media, our inversion uses an acoustic wave equation to simulate wave propagation, which can not take *S* waves and converted waves into account. When comparing the observed data with the calculated data displayed in Fig. 21, one can notice that the events in the calculated data match well with that in the observed data in phase, while the calculated data has less reflections and elastic amplitude-versus-offset phenomenon can not be matched due to the acoustic modelling. In the reconstructed velocity models, one may notice some vertical-strip artefacts (especially in the results of AWI), which may be caused by time inconsistency of different offset data (Valensi & Baina 2021). These artefacts could be mitigated by either structure-oriented smoother (Yao *et al.* 2019; Provenzano *et al.* 2021) or edge-preserving regularization (Yong *et al.* 2018; Aghamiry *et al.* 2019).

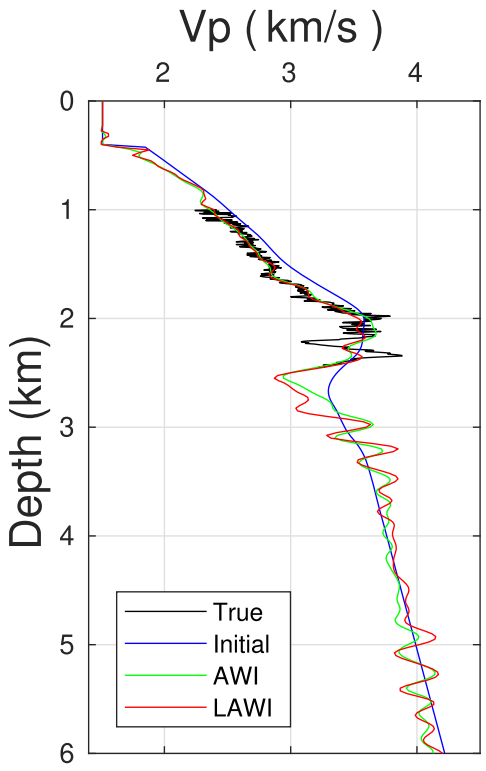

Figure 25. Comparison of the well logs taken at $x = 39{,}375$ m of the reconstructed models by LAWI (red line) and AWI (green line), the initial model (blue line) and given true one (black line). Due to the velocity update between 2.5 and 3 km, the well log generated by LAWI has an up shift below 3 km depth, compared with that extracted from the velocity estimated by AWI.

The true models have not been released so far. For a better assessment of the quality of reconstructed models, three reverse time migration using respectively initial model and two final AWI and LAWI reconstructed models are considered. The migrated images with the inverse scattering imaging condition (Op't Root *et al.* 2012; Whitmore & Crawley 2012), namely the impedance gradients of FWI with the parametrization of impedance and velocity (Zhou *et al.* 2015), are displayed in Fig. 26. Overall, the velocity obtained by the AWI approach can greatly improve the quality of the image: reflectors become more continuous due to the kinematic correction in the velocity. One may notice that LAWI formulation can further make the energy of the migrated event more focusing. Because of the low-velocity zone, the reflectors in the images also have an uplift below 3 km depth, which can be better observed in the common image gathers (CIG) shown in Fig. 27. Note that the arrows plotted in the migrated images (also in CIG gathers) have the same relative coordinates, which highlight not only the improvements but also can be regarded as references to observe the uplift of reflectors. One may notice that the migrated images seem to contain more reflectors compared to the structures presented in the velocity model, which could be possible artefacts, because no deghosting and demultiple have been applied to the data before migration. This could be one reason for pushing the non-linear process of FWI towards high frequencies (Huang *et al.* 2021). At last, from the overlay of velocity and migration results in Fig. 28, we can observe the increase of consistency between the velocity model and imaging structures with the velocity update.

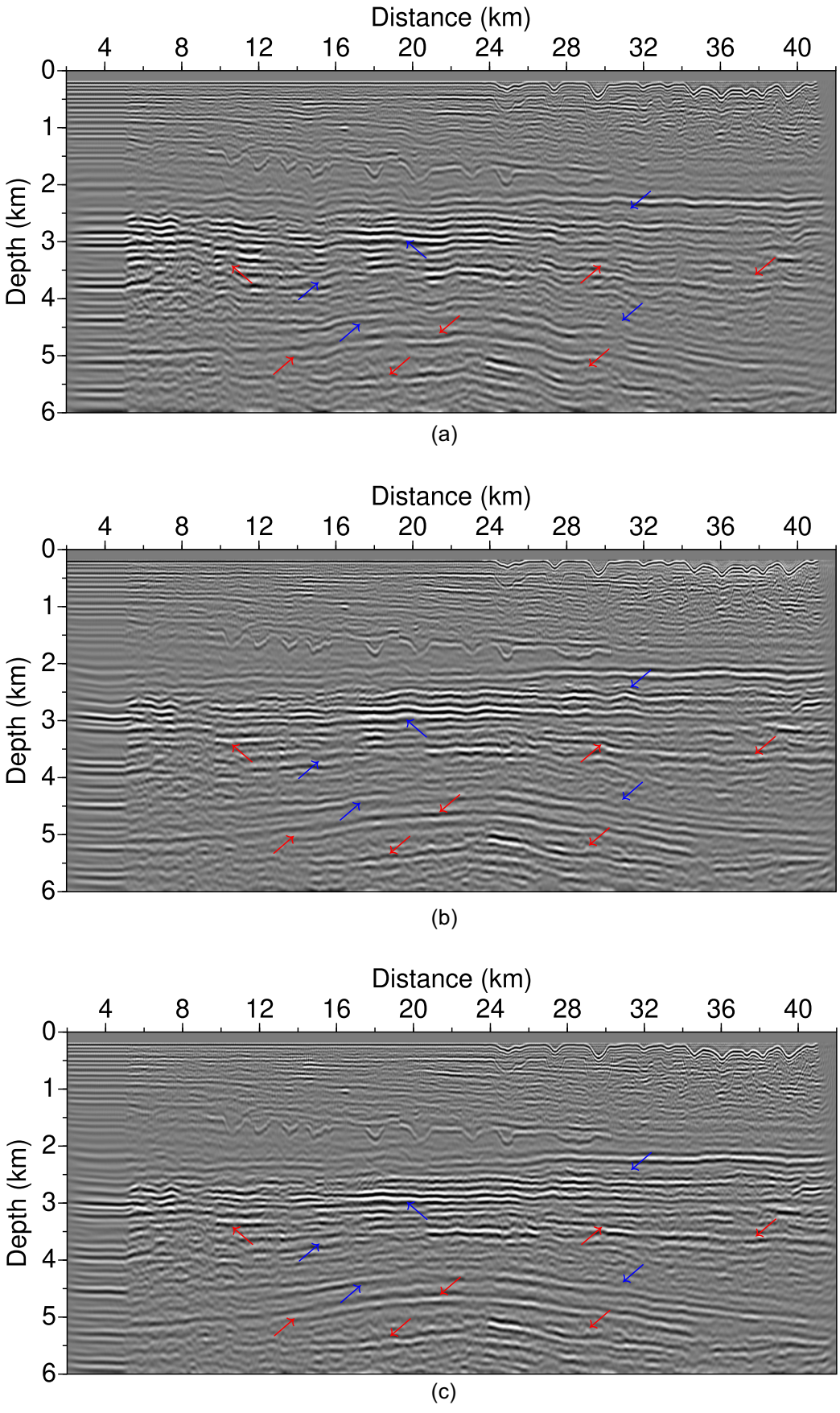

Figure 26. The migrated images using the initial model (a), the reconstructed velocities by AWI (b) and LAWI (c) with 2–20 Hz data. Arrows (red and blue) suggest the improvements with velocity update, and the blue arrows highlight the difference between results from AWI and LAWI.

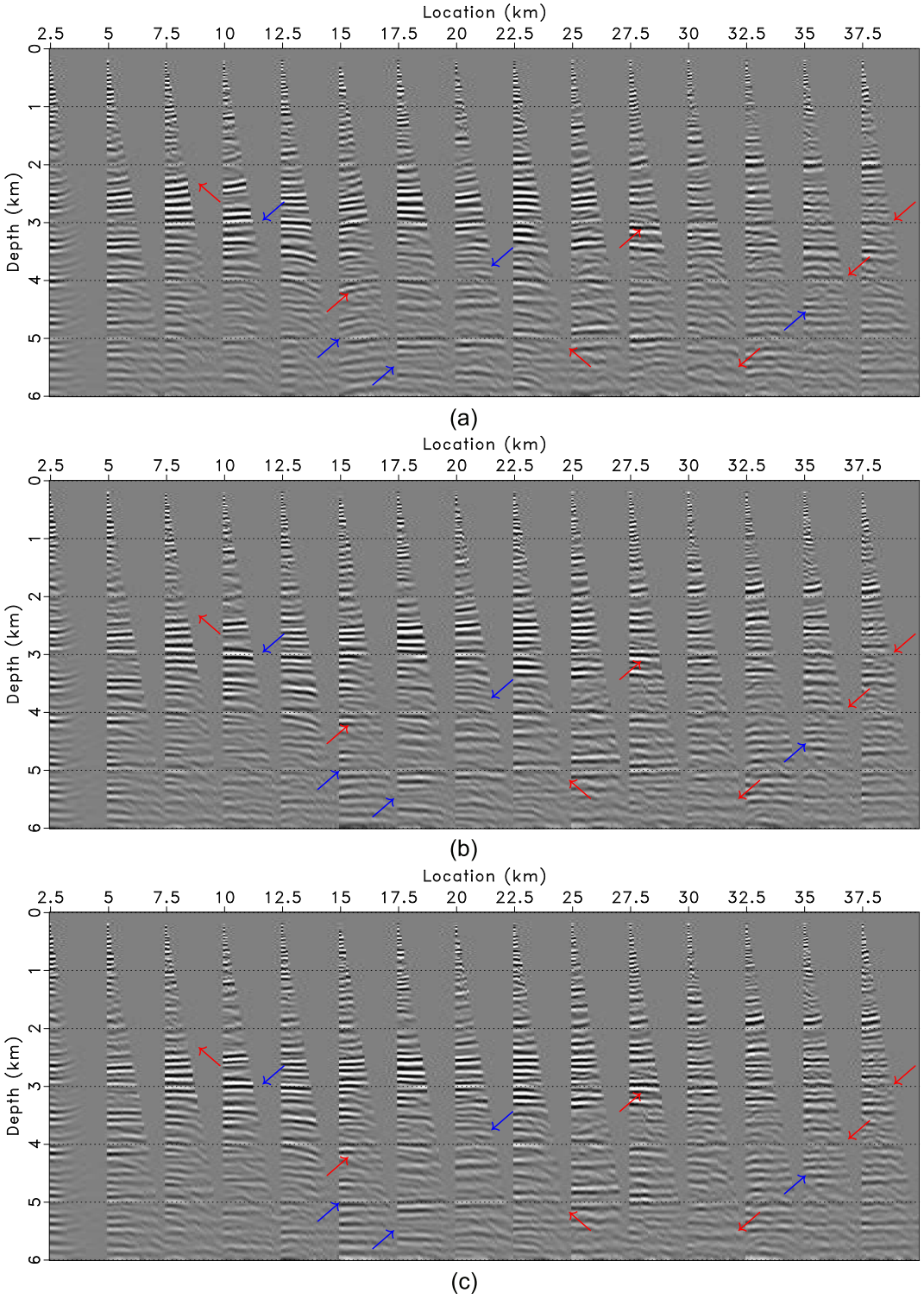

Figure 27. Common image gathers: (a) the initial model, estimated models by AWI (b) and LAWI (c). With auxiliary dash lines, one can observe that the gathers below the depth of 3 km have an uplift due to the update of the low-velocity layer in the 2.5–3 km deep zone. The improvements of the flatness of the gathers are indicated by the arrows. The gathers are extracted with 2.5 km interval, and the surface locations in the images are denoted by the value of horizontal coordinates. The maximum offset of each gather is 5 km.

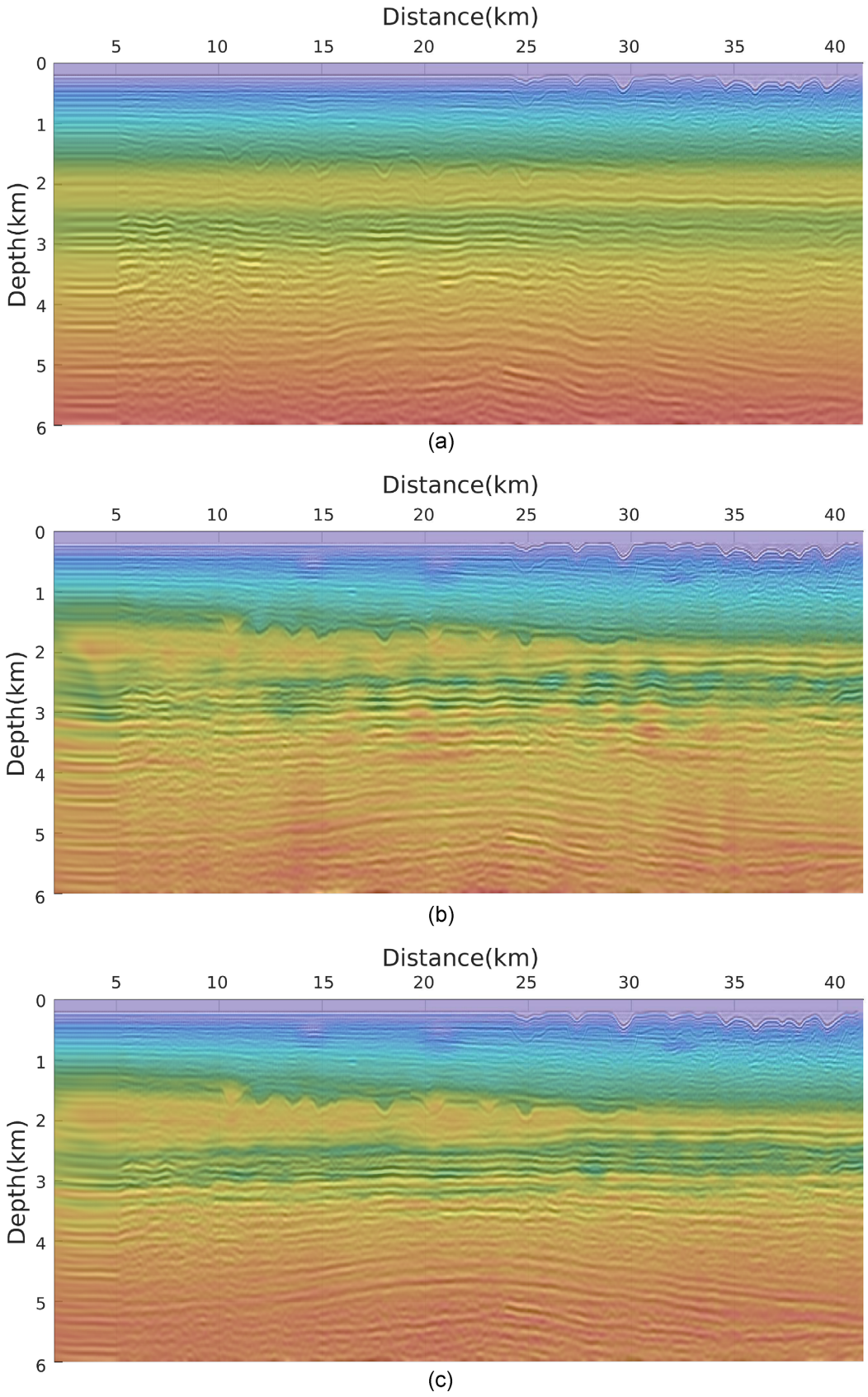

Figure 28. Overlay of velocity models and their corresponding migration images: (a) the initial velocity model, (b) the reconstructed velocity model by AWI and (c) the reconstructed velocity model by LAWI.

6 DISCUSSION AND CONCLUSION

For waveform fitting, different strategies have been proposed for a better interpretation of traveltime information. Time analysis are essential for boosting time delays compared to amplitude variations between observed and calculated waveforms. Understanding how standard time operations, such as the cross-correlation, the deconvolution and the matching filter with or without normalization, operate on simple time signals or on realistic synthetic examples is crucial before tackling field data sets.

We highlight that the normalization is a key element of the AWI workflow, which enables the positive or negative updating direction of the velocity model while direct cross-correlation or deconvolution methods fails to identify the polarity of the velocity perturbation. These methods can be understood as a way to estimate a filter-weighted time-shift, and such an estimation highly depends on the amplitude of the filter. By a normalization, the AWI approach moves the matching filter for each trace towards zero-time lag for the minimization of the objective function. The total duration of the trace is considered, leading to a global matching filter involving various more or less significant associations of different phase events.

We propose in this work to consider a local matching filter in order to capture instantaneous time-shifts based on a non-stationary convolutional model. It can be regarded as a generalization form of the AWI formulation, which can further release the accuracy requirement of initial models. In addition, thanks to Fast Fourier transform, the overhead cost induced by time–frequency analysis in the LAWI approach is affordable, and it is about 5 per cent of the computation complexity of gradient construction in the inversion tests. Let us mention that, for the sake of heuristic interpretation for AWI, by showing the similarity between the AWI misfit and the centroid frequency, the absolute-value function is used in this paper as the penalty function. Since both Warner & Guasch (2016) and Sun & Alkhalifah (2020) suggest that alternative forms of penalty function may provide a more rapid and more stable convergence for the AWI approach, it could be an interesting work to investigate the effect of the penalty function on the LAWI method in the future.

In the AWI approach, a least-squares convolutional filter is used to transform one-trace observed data onto one-trace calculated data. The LAWI formulation uses the Gaussian window to locally accomplish such mapping, which can take the non-stationary nature between predicted and observed data into account. Numerical analysis and inversion tests illustrate that time-localized manipulations can improve the capability to handle cycle-skipping issues. It is necessary to point out that, when several events interactively appear in a local time window, the centroid time of the local matching filter can only quantitatively reflects the instantaneous time-shift of these combined phase events, as the energy of the local matching filter, the optimal solution of the Gabor convolutional model under least-squares sense, is diffusely distributed. Identifying each event may resolve this problem, while facing difficulties encountered in dynamic time warping (Ma & Hale 2013; Chen *et al.* 2021). Nevertheless, it could be quite difficult to achieve this for complex exploration data. One possible way to make the centroid time closer to the true time-shift is through wavelet transform, which may generate a more focusing matching filter with a hierarchy of scales and with different mother-wavelet shapes.

In the numerical experiments, the regularization parameter ϵ in Gabor deconvolution is a constant for each time point. Such a parameter plays an important role especially for noisy data. One may consider using a time-varying ϵ to account for the amplitude variation of phase events. After different numerical experiments, we observe that the inversion could be unstable when such a parameter ϵ varies too much with the local energy estimation. One can recognize that, with a constant ϵ, the local matching filter will be still affected by amplitude to some extent. Certainly, time-varying ϵ can make the LAWI workflow more traveltime-driven. However, for noisy data applications, using a constant ϵ is recommended. This observation has been found for other applications of Gabor deconvolution (Margrave 1998; Margrave *et al.* 2011). In future work, we shall apply the LAWI approach to an anisotropic 3-D field data, and we shall discuss in more details how to stabilize Gabor deconvolution when considering low SNR data from a point view of regularization.

Apart from exploration applications, the LAWI method is also applicable to global-scale and regional-scale passive applications, because the LAWI approach is dedicated to enhance traveltime information for model update and to be nearly agnostic to amplitude errors. In global seismology, the earthquake source time function is difficult to estimate and the amplitude could vary for each station and for each phase, depending on the corresponding location point of the focal sphere (Tromp 2019). Thus, the developed LAWI method can be an alternative to the existing phase and envelope misfits (Fichtner *et al.* 2008; Bozdağ *et al.* 2011) for these global and regional FWI applications.

DATA AND MATERIALS AVAILABILITY

Data and code associated with this research belong to the SEISCOPE consortium. Any request related to the availability of the materials needs to be addressed to Romain Brossier (romain.brossier@univ-grenoble-alpes.fr).

ACKNOWLEDGMENTS

The authors would like to thank Yubing Li and Weiguang He for sharing their insights on adaptive waveform inversion. PY also appreciates many useful discussions with Giuseppe Provenzano, Jian Cao and Serge Sambolian. This study was partially funded by the SEISCOPE consortium (*http://seiscope2.osug.fr*), sponsored by AKERBP, CGG, CHEVRON, EXXON-MOBIL CORPORATION, GEOLINKS, JGI, PETROBRAS, SHELL, SINOPEC, SISPROBE and TOTALENERGIES. This study was granted access to the HPC resources provided by the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, the HPC resources of Cray Marketing Partner Network (http://partners.cray.com), and those of IDRIS/TGCC under the allocation 046091 made by GENCI. We thank the editor Frederik Simons and two reviewers Hejun Zhu and Tariq Alkhalifah for their careful reading and helpful comments.

REFERENCES

Aghamiry, H., Gholami, A. & Operto, S., 2019. Implementing bound constraints and total-variation regularization in extended full waveform inversion with the alternating direction method of multiplier: application to large contrast media, *Geophys. J. Int.,* **218**(2), 855–872.

Baek, H., Calandra, H. & Demanet, L., 2014. Velocity estimation via registration-guided least-squares inversion, *Geophysics,* **79**(2), R79–R89.

Billette, F. & Lambaré, G., 1998. Velocity macro-model estimation from seismic reflection data by stereotomography, *Geophys. J. Int.,* **135**(2), 671–680.

Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185**(2), 845–870.

Brossier, R., Operto, S. & Virieux, J., 2015. Velocity model building from seismic reflection data by full waveform inversion, *Geophys. Prospect.,* **63,** 354–367.

Bunks, C., Salek, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics,* **60**(5), 1457–1473.

Burdick, S., de Hoop, M.V., Wang, S. & van der Hilst, R.D., 2014. Reverse-time migration-based reflection tomography using teleseismic free surface multiples, *Geophys. J. Int.,* **16**(2), 996–1017.

Červený, V., 2001. *Seismic Ray Theory,* Cambridge Univ. Press.

Chavent, G., Clément, F. & Gòmez, S., 1994. Automatic determination of velocities via migration-based traveltime waveform inversion: a synthetic data example, *SEG Tech. Prog. Expanded Abstracts 1994,* 1179–1182.

Chen, F., Peter, D. & Ravasi, M., 2021. Misfit functions based on differentiable dynamic time warping for waveform inversion, in *SEG/AAPG/SEPM First International Meeting for Applied Geoscience & Energy,* OnePetro.

Chen, G., Yang, W., Chen, S., Liu, Y. & Gu, Z., 2020. Application of envelope in salt structure velocity building: from objective function construction to the full-band seismic data reconstruction, *IEEE Trans. Geosci. Remote Sens.,* **58**(9), 6594–6608.

Claerbout, J.F. & Fomel, S., 2014. *Geophysical Image Estimation by Example,* Lulu. com.

Clayton, R.W. & Wiggins, R.A., 1976. Source shape estimation and deconvolution of teleseismic bodywaves, *Geophys. J. Int.,* **47**(1), 151–177.

Dellinger, J. *et al.*, 2016. Wolfspar®, an "FWI-friendly" ultralow-frequency marine seismic source, in *SEG Technical Program Expanded Abstracts 2016,* pp. 4891–4895, Society of Exploration Geophysicists.

Díaz, E. & Sava, P., 2015. Data domain wavefield tomography using local correlation functions, in *2015 SEG Annual Meeting,* OnePetro.

Engquist, B., Froese, B.D. & Yang., Y., 2016. Optimal transport for seismic full waveform inversion, *Commun. Math. Sci.,* **14**(8), 2309–2330.

Fichtner, A., Kennett, B. L.N., Igel, H. & Bunge, H.P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *Geophys. J. Int.,* **175,** 665–685.

Gabor, D., 1946. Theory of communication. Part 1: the analysis of information, *J. Inst. Electr. Eng.-Part III: Radio Commun. Eng.,* **93**(26), 429–441.

Górszczyk, A., Brossier, R. & Métivier, L., 2021. Graph-space optimal transport concept for time-domain full-waveform inversion of ocean-bottom seismometer data: Nankai trough velocity structure reconstructed from a 1D model, *J. geophys. Res.,* **126**(5), e2020JB021504, doi:10.1029/2020JB021504.

Guasch, L., Warner, M. & Ravaut, C., 2019. Adaptive waveform inversion: practice, *Geophysics,* **84(3),** R447–R461.

Hu, W., Jin, Y., Wu, X. & Chen, J., 2021. Progressive transfer learning for low-frequency data prediction in full waveform inversion, *Geophysics,* **86**(4), 1–82.

Hu, Y., Wu, R.-S., Han, L.-G. & Zhang, P., 2019. Joint multiscale direct envelope inversion of phase and amplitude in the time–frequency domain, *IEEE Trans. Geosci. Remote Sens.,* **57**(7), 5108–5120.

Huang, R., Zhang, Z., Wu, Z., Wei, Z., Mei, J. & Wang, P., 2021. Full-waveform inversion for full-wavefield imaging: decades in the making, *Leading Edge,* **40**(5), 324–334.

Jannane, M. *et al.*, 1989. Wavelengths of Earth structures that can be resolved from seismic reflection data, *Geophysics,* **54**(7), 906–910.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering, Theory and Application,* pp. 206–220, eds Bednar, J.B., Robinson, E. & Weglein, A., Society for Industrial and Applied Mathematics.

Lei, W. *et al.*, 2020. Global adjoint tomography—model glad-m25, *Geophys. J. Int.,* **223**(1), 1–21.

Li, Y., 2018. Subsurface seismic imaging based on inversion velocity analysis in both image and data domains, *PhD thesis*, PSL Research University-MINES ParisTech.

Li, Y.E. & Demanet, L., 2016. Full-waveform inversion with extrapolated low-frequency data, *Geophysics,* **81**(6), R339–R348.

Luo, S. & Sava, P., 2011. A deconvolution-based objective function for wave-equation inversion, *SEG Tech. Prog. Expanded Abstracts,* **30**(1), 2788–2792.

Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics,* **56**(5), 645–653.

Ma, Y. & Hale, D., 2013. Wave-equation reflection traveltime inversion with dynamic warping and full waveform inversion, *Geophysics,* **78**(6), R223–R233.

Maggi, A., Tape, C., Chen, M., Chao, D. & Tromp, J., 2009. An automated time-window selection algorithm for seismic tomography, *Geophys. J. Int.,* **178,** 257–281.

Mallat, S., 2008. *A Wavelet Tour of Signal Processing,* 3rd edn: The Sparse Way, Academic Press.

Margrave, G.F., 1998. Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering, *Geophysics,* **63**(1), 244–259.

Margrave, G.F., Lamoureux, M.P. & Henley, D.C., 2011. Gabor deconvolution: estimating reflectivity by nonstationary deconvolution of seismic data, *Geophysics,* **76**(3), W15–W30.

Métivier, L. & Brossier, R., 2016. The SEISCOPE optimization toolbox: a large-scale nonlinear optimization library based on reverse communication, *Geophysics,* **81**(2), F11–F25.

Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. & Virieux, J., 2016. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, *Geophys. J. Int.,* **205,** 345–377.

Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E. & Virieux, J., 2018. Optimal transport for mitigating cycle skipping in full waveform inversion: a graph space transform approach, *Geophysics,* **83**(5), R515–R540.

Mulder, W.A. & ten kroode, A. P.E., 2002. Automatic velocity analysis by differential semblance optimization, *Geophysics,* **67**(4), doi:10.1190/1.1500380.

Op 't Root, T., Stolk, C.C. & Maarten, V., 2012. Linearized inverse scattering based on seismic reverse time migration, *Journal de mathématiques pures et appliquées,* **98**(2), 211–238.

Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A. & Virieux, J., 2015. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *Geophys. J. Int.,* **202**(2), 1362–1391.

Pladys, A., Brossier, R., Li, Y. & Métivier, L., 2021. On cycle-skipping and misfit function modification for full-wave inversion: comparison of five recent approaches, *Geophysics,* **86**(4), R563–R587.

Pladys, A., Brossier, R., Kamath, N. & Métivier, L., 2022. Robust FWI with graph space optimal transport: application to 3D OBC Valhall data, *Geophysics,* **87**(3), 1–76.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Plessix, R.-É. & Krupovnickas, T., 2021. Low-frequency, long-offset elastic waveform inversion in the context of velocity model building, *Leading Edge,* **40**(5), 342–347.

Plessix, R.E., Baeten, G., de Maag, J.W., Klaassen, M., Rujie, Z. & Zhifei, T., 2010. Application of acoustic full waveform inversion to a low-frequency large-offset land data set, *SEG Tech. Prog. Expanded Abstracts,* **29**(1), 930–934.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, part I: theory and verification in a physical scale model, *Geophysics,* **64,** 888–901.

Prieux, V., Lambaré, G., Operto, S. & Virieux, J., 2013. Building starting model for full waveform inversion from wide-aperture data by stereotomography, *Geophys. Prospect.,* **61**(Issue supplement: 60 year anniversary issue), 109–137.

Provenzano, G., Brossier, R. & Métivier, L., 2021. Joint reflection and diving FWI using graph-space optimal transport and structure-guided smoothing on benchmark data, in *Proceedings of the 82th Annual EAGE Meeting,* Amsterdam, Vol. 2021, pp. 1–5, European Association of Geoscientists & Engineers.

Quan, Y. & Harris, J.M., 1997. Seismic attenuation tomography using the frequency shift method, *Geophysics,* **62**(3), 895–905.

Sambolian, S., Operto, S., Ribodetti, A., Tavakoli, B. & Virieux, J., 2019. Parsimonious slope tomography based on eikonal solvers and the adjoint-state method, *Geophys. J. Int.,* **218**(1), 456–478.

Shen, P. & Symes, W.W., 2008. Automatic velocity analysis via shot profile migration, *Geophysics,* **73**(5), VE49–VE59.

Sirgue, L., Barkved, O.I., Dellinger, J., Etgen, J., Albertin, U. & Kommedal, J.H., 2010. Full waveform inversion: the next leap forward in imaging at Valhall, *First Break,* **28,** 65–70.

Stork, C., 1992. Reflection tomography in the postmigrated domain, *Geophysics,* **57**(5), 680–692.

Strang, G. & Nguyen, T., 1996. *Wavelets and Filter Banks,* SIAM.

Sun, B. & Alkhalifah, T., 2018. Mitigate cycle skipping in FWI: a generalized instanatenous travel-time approach, in *Proceedings of the 80th EAGE Conference & Exhibition,* 11–14 June 2018, Copenhagen, Denmark.

Sun, B. & Alkhalifah, T., 2019a. The application of an optimal transport to a preconditioned data matching function for robust waveform inversion, *Geophysics,* **84**(6), R935–R957.

Sun, B. & Alkhalifah, T., 2019b. Adaptive traveltime inversion, *Geophysics,* **84**(4), U13–U29.

Sun, B. & Alkhalifah, T.A., 2020. Joint minimization of the mean and information entropy of the matching filter distribution for a robust misfit function in full-waveform inversion, *IEEE Trans. Geosci. Remote Sens.,* **58**(7), 4704–4720.

Sun, H. & Demanet, L., 2020. Extrapolated full-waveform inversion with deep learning, *Geophysics,* **85**(3), R275–R288.

Symes, W.W. & Carazzone, J.J., 1991. Velocity inversion by differential semblance optimization, *Geophysics,* **56,** 654–663.

Taillandier, C., Noble, M., Chauris, H. & Calandra, H., 2009. First-arrival travel time tomography based on the adjoint state method, *Geophysics,* **74**(6), WCB1–WCB10.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tavakoli, F. B., Operto, S., Ribodetti, A. & Virieux, J., 2017. Slope tomography based on eikonal solvers and the adjoint-state method, *Geophys. J. Int.,* **209**(3), 1629–1647.

Ten Kroode, F., Bergler, S., Corsten, C., de Maag, J.W., Strijbos, F. & Tijhof, H., 2013. Broadband seismic data—the importance of low frequencies, *Geophysics,* **78**(2), WA3–WA14.

Tromp, J., 2019. Seismic wavefield imaging of earth's interior across scales, *Nat. Rev. Earth Environ.,* **1,** 40–53.

Valensi, R. & Baina, R., 2021. A time consistent waveform inversion (twin) method, in *Proceedings of the 82nd EAGE Annual Conference & Exhibition,* Vol. 2021, pp. 1–5, European Association of Geoscientists & Engineers.

van Leeuwen, T. & Mulder, W.A., 2010. A correlation-based misfit criterion for wave-equation traveltime tomography, *Geophys. J. Int.,* **182**(3), 1383–1394.

Vigh, D., Jiao, K., Cheng, X., Sun, D. & Lewis, W., 2016. Earth-model building from shallow to deep with full-waveform inversion, *Leading Edge,* **35,** 1535–1554.

Vigh, D., Cheng, X., Jiao, K., Kang, W. & Brand, N., 2021. The impact of acquisition geometry on full-waveform inversion updates, *Leading Edge,* **40**(5), 335–341.

Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

Virieux, J., Lambaré, G., Romanowicz, B. & Diewonski, A., 2007. Theory and observations–body waves: Ray methods and finite frequency effects, in *Seismology and Structure of the Earth: Treatise on Geophysics,* pp. 127–155, eds Romanowicz, B.A. & Dziewonski, A.M., Elsevier.

Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A. & Zhou, W., 2017. An introduction to full waveform inversion, in *Encyclopedia of Exploration Geophysics,* pp. R1–1-R1-40, eds Grechka, V. & Wapenaar, K., Society of Exploration Geophysics.

Wang, Y., 2009. *Seismic Inverse Q Filtering,* John Wiley & Sons.

Wang, Y. & Rao, Y., 2009. Reflection seismic waveform tomography, *J. geophys. Res.,* **114**(B3), 1978–2012.

Warner, M. & Guasch, L., 2015. Robust adaptive waveform inversion, in *SEG Technical Program Expanded Abstracts 2015,* pp. 1059–1063, Society of Exploration Geophysicists.

Warner, M. & Guasch, L., 2016. Adaptive waveform inversion: Theory, *Geophysics,* **81**(6), R429–R445.

Whitmore, N. & Crawley, S., 2012. Applications of RTM inverse scattering imaging conditions, in *SEG Technical Program Expanded Abstracts 2012,* pp. 1–6, Society of Exploration Geophysicists.

Woodward, M.J., 1992. Wave-equation tomography, *Geophysics,* **57,** 15–26.

Wu, R.-S., Luo, J. & Wu, B., 2014. Seismic envelope inversion and modulation signal model, *Geophysics,* **79**(3), WA13–WA24.

Wu, Z. & Alkhalifah, T., 2015. Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion, *Geophysics,* **80**(6), R317–R329.

Xu, S., Wang, D., Chen, F., Lambaré, G. & Zhang, Y., 2012. Inversion on reflected seismic wave, in *SEG Technical Program Expanded Abstracts 2012,* pp. 1–7, Society of Exploration Geophysicists.

Yang, Y., Engquist, B., Sun, J. & Hamfeldt, B.F., 2018. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, *Geophysics,* **83**(1), R43–R62.

Yao, G., da Silva, N.V. & Wu, D., 2019. Reflection-waveform inversion regularized with structure-oriented smoothing shaping, *Pure appl. Geophys.,* **176,** 5315–5335.

Yong, P., Liao, W., Huang, J. & Li, Z., 2018. Total variation regularization for seismic waveform inversion using an adaptive primal dual hybrid gradient method, *Inverse Problems,* **34**(4), doi:10.1088/1361-6420/aaaf8e.

Yong, P., Liao, W., Huang, J., Li, Z. & Lin, Y., 2019. Misfit function for full waveform inversion based on the Wasserstein metric with dynamic formulation, *J. Comput. Phys.,* **399,** doi:10.1016/j.jcp.2019.108911.

Yong, P., Brossier, R. & Métivier, L., 2022. Parsimonious truncated newton method for time-domain full-waveform inversion based on the Fourier-domain full-scattered-field approximation, *Geophysics,* **87**(1), R123–R146.

Yu, S. & Ma, J., 2021. Deep learning for geophysics: current and future trends, *Rev. Geophys.,* **59**(3), e2021RG000742, doi:10.1029/2021RG000742.

Zhang, J. & Toksöz, M.N., 1998. Nonlinear refraction traveltime tomography, *Geophysics,* **63**(5), 1726–1737.

Zhou, W., Brossier, R., Operto, S. & Virieux, J., 2015. Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation, *Geophys. J. Int.,* **202**(3), 1535–1554.

Zhu, H., 2018. Seismogram registration via Markov Chain Monte Carlo optimization and its applications in full waveform inversion, *Geophys. J. Int.,* **212**(2), 976–987.

Zhu, H. & Fomel, S., 2016. Building good starting models for full-waveform inversion using adaptive matching filtering misfit, *Geophysics,* **81**(5), U61–U72.

Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.,* **5,** 493–498.

Zhu, X., Sixta, D.P. & Angstman, B.G., 1992. Tomostatics: turning-ray tomography + static corrections, *Leading Edge,* **11**(12), 15–23.

APPENDIX: DERIVATION OF ADJOINT SOURCES FOR AWI AND LAWI FORMULATIONS

We give the details about the derivation of adjoint sources in the hybrid time and frequency domains. Combining the objective function definition of the AWI approach (eq. 18) and the Fourier transform pair (eqs 10 and 11), also using the chain rule, we have

$$\begin{aligned}\frac{\partial \mathbb{J}_{AWI}}{\partial p(t)} &= \frac{\partial \mathbb{J}_{AWI}}{\partial w(\tau)} \times \frac{\partial w(\tau)}{\partial \tilde{w}(\omega)} \times \frac{\partial \tilde{w}(\omega)}{\partial \tilde{p}(\omega)} \times \frac{\partial \tilde{p}(\omega)}{\partial p(t)} \\ &= \int_{\mathbb{R}} \frac{(|\tau| - 2\mathbb{J}_{AWI})w(\tau)}{\int_{\mathbb{R}} w^2(\tau)\mathrm{d}\tau}\mathrm{d}\tau \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\omega\tau}\mathrm{d}\omega \times \frac{\tilde{d}^{\dagger}(\omega)}{\tilde{d}(\omega)\tilde{d}^{\dagger}(\omega) + \epsilon} \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-i\omega t}\mathrm{d}t \\ &= \left\{ \mathcal{F}_{\tau} \left[\frac{(|\tau| - 2\mathbb{J}_{AWI})w(\tau)}{\int_{\mathbb{R}} w^2(\tau)\mathrm{d}\tau} \right] \times \frac{\tilde{d}(\omega)}{\tilde{d}^{\dagger}(\omega)\tilde{d}(\omega) + \epsilon} \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i\omega t}\mathrm{d}t\mathrm{d}\omega \right\}^{\dagger} \\ &= \left\{ \int_{\mathbb{R}} \mathcal{F}_t^{-1} \left[\mathcal{F}_{\tau} \left[\frac{(|\tau| - 2\mathbb{J}_{AWI})w(\tau)}{\int_{\mathbb{R}} w^2(\tau)\mathrm{d}\tau} \right] \times \frac{\tilde{d}(\omega)}{\tilde{d}^{\dagger}(\omega)\tilde{d}(\omega) + \epsilon} \right] \mathrm{d}t \right\}^{\dagger}. \end{aligned} \tag{A1}$$

Since the AWI adjoint source is a real-valued signal, we can write it as

$$\mathbf{r}_{AWI} = \mathcal{F}_t^{-1} \left[\mathcal{F}_{\tau} \left[\frac{(|\tau| - 2\mathbb{J}_{AWI})w(\tau)}{\int_{\mathbb{R}} w^2(\tau)\mathrm{d}\tau} \right] \frac{\tilde{d}(\omega)}{\tilde{d}(\omega)\tilde{d}^{\dagger}(\omega) + \epsilon} \right]. \tag{A2}$$

Following a similar step, using the LAWI objective function (eq. 27) and the Gabor transform (eqs 21 and 22), the derivative of the misfit of the LAWI approach with respect to predicted data can be obtained as

$$\begin{aligned}\frac{\partial \mathbb{J}_{LAWI}}{\partial p(\xi)} &= \frac{\partial \mathbb{J}_{LAWI}}{\partial T(t)} \times \frac{\partial T(t)}{\partial w(t,\tau)} \times \frac{\partial w(t,\tau)}{\partial \hat{w}(t,\omega)} \times \frac{\partial \hat{w}(t,\omega)}{\partial \hat{p}(t,\omega)} \times \frac{\partial \hat{p}(t,\omega)}{\partial p(\xi)} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \underbrace{\frac{2W(t)T(t)(|\tau| - T(t))w(t,\tau)}{\int_{\mathbb{R}} w^2(t,\tau)\mathrm{d}\tau + \eta}}_{\mathbf{A}(t,\tau)} \mathrm{d}\tau\mathrm{d}t \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\omega\tau}\mathrm{d}\omega \times \frac{\partial \hat{w}(t,\omega)}{\partial \hat{p}(t,\omega)} \times \frac{\partial \hat{p}(t,\omega)}{\partial p(\xi)} \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{A}(t,\tau)\mathrm{d}\tau\mathrm{d}t \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{i\omega\tau}\mathrm{d}\omega \times \underbrace{\frac{\hat{d}^{\dagger}(t,\omega)}{\hat{d}(t,\omega)\hat{d}^{\dagger}(t,\omega) + \epsilon}}_{\mathbf{B}(t,\omega)} \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h_{\sigma}^{\dagger}(t-\xi)e^{-i\omega\xi}\mathrm{d}\xi \\ &= \left\{ \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbf{A}(t,\tau)\mathrm{d}\tau\mathrm{d}t \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-i\omega\tau}\mathrm{d}\omega \times \mathbf{B}^{\dagger}(t,\omega) \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h_{\sigma}(t-\xi)e^{i\omega\xi}\mathrm{d}\xi \right\}^{\dagger} \\ &= \left\{ \int_{\mathbb{R}} \mathrm{d}\xi \mathcal{F}_{\tau} \left[\mathbf{A}(t,\tau) \right] \times \mathbf{B}^{\dagger}(t,\omega) \times \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \int_{\mathbb{R}} h_{\sigma}(t-\xi)e^{i\omega\xi}\mathrm{d}t\mathrm{d}\omega \right\}^{\dagger} \\ &= \left\{ \int_{\mathbb{R}} \mathcal{G}^{-1} \left[\mathcal{F}_{\tau} \left[\mathbf{A}(t,\tau) \right] \times \mathbf{B}^{\dagger}(t,\omega) \right] \mathrm{d}\xi \right\}^{\dagger}. \end{aligned} \tag{A3}$$

Consequently, the LAWI adjoint source can be given by

$$\mathbf{r}_{LAWI} = 2\mathcal{G}^{-1} \left[\mathcal{F}_{\tau} \left[\frac{W(t)T(t)(|\tau| - T(t))w(t,\tau)}{\int_{\mathbb{R}} w^2(t,\tau)\mathrm{d}\tau + \eta} \right] \frac{\hat{d}(t,\omega)}{\hat{d}(t,\omega)\hat{d}^{\dagger}(t,\omega) + \epsilon} \right]. \tag{A4}$$