Geophysical Journal International

Geophys. J. Int. (2023) **235,** 1761–1777
Advance Access publication 2023 August 21
GJI Seismology

https://doi.org/10.1093/gji/ggad330

Learning with real data without real labels: a strategy for extrapolated full-waveform inversion with field data

Hongyu Sun [1,*] Yen Sun,[2] Rami Nammour,[2] Christian Rivera,[2] Paul Williamson[2] and Laurent Demanet[1,3]

[1] *Earth Resources Laboratory, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA* 02139, *USA. E-mail: hongyu-sun@outlook.com*
[2] *TotalEnergies E&P Research and Technology, Houston, TX* 77002, *USA*
[3] *Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA* 02139, *USA*

Accepted 2023 August 10. Received 2023 June 7; in original form 2022 May 5

SUMMARY

Full-waveform inversion (FWI) relies on low-frequency data to succeed if a good initial model is unavailable. However, field seismic data excited by active sources are typically band-limited above 3 Hz. By extrapolated FWI, we can start inversion from computational low frequencies extrapolated from band-limited data. However, low-frequency extrapolation with deep learning is challenging for field data since a neural network trained on synthetic data usually generalizes poorly on real seismic data. Here we use a semi-supervised learning method to extrapolate low frequencies for field data by training with real data without real labels. Specifically, by training CycleGAN with unpaired images of field 4–10 Hz band-limited and synthetic 0–4 Hz low-frequency shot gathers, we can extrapolate the 0–4 Hz low frequencies for the field data band-limited above 4 Hz. The source wavelet for the simulation of synthetic low-frequency data is used as the source in FWI using the extrapolated data. The inverted velocity model using only the extrapolated low frequencies is comparable to the tomography model. Our method strengthens the ability of FWI for mapping fine Earth structures by mitigating the cycle-skipping problem effectively.

Key words: Computational seismology; Controlled source seismology; Waveform inversion; Low-frequency extrapolation; Deep learning.

1 INTRODUCTION

Full-waveform inversion (FWI) is well-known for its potential to reveal the Earth's structures with high resolution by computationally interpreting the seismic data. Starting from an initial guess about the Earth (initial model), FWI can provide physical models by gradually matching the calculated data with the observed data, usually with an L_2-norm objective function (Tarantola 1984; Pratt *et al.* 1998). Since the L_2-norm objective function using low-frequency data is more convex and has fewer local minima than that using high-frequency data, we usually start FWI from the lowest frequency component in the seismic data (Bunks *et al.* 1995). However, seismic waves excited by a conventional active source lack the low-frequency parts. FWI using the band-limited data can easily fall into a local minimum (the cycle-skipping problem). To mitigate cycle-skipping, FWI has been reformulated through different objective functions, such as Laplace-domain inversion (Shin & Cha 2008) and optimal transport (Engquist *et al.* 2016; Métivier *et al.* 2016). Successful field data applications of FWI rely on tomography to build a good starting model, but this tomography requires arduous seismic phase identification and traveltime picking.

The advance of artificial intelligence enables researchers to solve inverse problems with the assistance of deep-learning methods (Richardson 2018; Adler *et al.* 2021). Several groups have experimented with directly mapping data to models using convolutional neural networks (Araya-Polo *et al.* 2018; Wu & Lin 2019; Yang & Ma 2019; Zhang & Lin 2020; Feng *et al.* 2021; Kazei *et al.* 2021; Du *et al.* 2022; Geng *et al.* 2022) or generative adversarial networks (Mosser *et al.* 2018; Zhong *et al.* 2020; Sun *et al.* 2021; Cai *et al.* 2022; Wang *et al.* 2022a). Within the Bayesian seismic inversion framework, deep generative models have been trained to provide priors (Herrmann *et al.* 2019; Fang *et al.* 2020b; Mosser *et al.* 2020). Li *et al.* (2021, 2022) propose a wide-band butterfly network for solving wave-based inverse problems in the super-resolution regime. He & Wang (2021), Zhu *et al.* (2022) and Zhang *et al.* (2023) reparametrize FWI using deep learning frameworks. Chen & Saygin (2021) calculate the FWI misfit in the latent space of an

*Now at: Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA.

autoencoder and show that it is less affected by the cycle-skipping problem compared to the waveform mismatch in a high-dimensional space. Yang *et al.* (2021, 2023) accelerate FWI with Neural Operators. In addition, some groups propose to generate initial models for FWI with deep learning (Dhara & Sen 2022; Vantassel *et al.* 2022; Yao & Wang 2022)

Instead of directly training deep neural networks to solve inverse problems, many researchers extrapolate low-frequency data from band-limited data with deep learning so that conventional FWI can benefit from the extrapolated low-frequency data (Sun & Demanet 2018, 2020; Ovcharenko *et al.* 2019; Hu *et al.* 2021; Jin *et al.* 2022b; Wang *et al.* 2022b). Subsequently, Sun & Demanet (2022a) and Ovcharenko *et al.* (2022) have extended their frameworks to extrapolate the low frequencies for elastic data. Fabien-Ouellet (2020) performs low-frequency extrapolation using a recurrent neural network. Meanwhile, a progressive transfer learning workflow is designed to modify the training data set using inverted models so it comes closer to the test data set (Zhao *et al.* 2020; Jin *et al.* 2022a). Moreover, Plotnitskii *et al.* (2020) choose to extrapolate the low-wavenumber components of the FWI gradient using deep neural networks. Nakayama & Blacquière (2021) include low-frequency extrapolation into a multitask learning workflow for seismic processing. Sun & Demanet (2022b) attempt to extrapolate low frequencies for surface waves to increase the investigation depth of surface wave tomography. As for bandwidth extension to a high-frequency regime, Choi *et al.* (2021) use a U-Net deep-learning model to broaden the frequency spectrum of high-cut-filtered seismic data. Inspired by these works on seismic data, Lin *et al.* (2021) apply low-frequency extrapolation to electromagnetic scattered field data. Robins *et al.* (2021) use the method in ultrasound computed tomography to image the breast with more diagnostically relevant quantitative tissue properties.

Only a few attempts of low-frequency extrapolation have been tried on field data. Aharchaou & Baumstein (2020) extrapolate the low frequencies for towed-streamer data band-limited above 4 Hz using deep neural networks trained on 1–4 Hz field data recorded by an ultrasparse set of ocean-bottom nodes. Zhang *et al.* (2022b) recover low frequencies for field data from post-stack data and use the extrapolated low-frequency data to invert the *P*-wave impedance. However, the training and test data sets in both field-data studies are collected from the same region. Moveover, Wang *et al.* (2020) propose a self-supervised learning method for low-frequency extrapolation and show its performance on field seismic shot gathers without FWI results. Fang *et al.* (2020a) compare both FWI and reverse-time migration results on field data with and without low-frequency data predicted by the neural network trained on the Marmousi model. Ovcharenko *et al.* (2022) simultaneously predict low-frequency data and background velocity models with multitask learning and show the performance on a field marine data set.

If we process real seismic data with a supervised learning framework, we may train the neural network on either real or synthetic data. Training on synthetic data may introduce large generalization errors when applying the trained neural network to real seismic data. Thus, a target-orientated synthetic training data set has to be carefully designed to improve the generalizability of the neural network to the target real data. For example, Alkhalifah *et al.* (2022) propose a series of data processing operations to introduce similarity to data distributions between synthetic training and real application data sets (Zhang *et al.* 2022a). Alternatively, training on real data may enable the neural network to perform better on field data, due to a smaller generalization gap. However, real labels, especially low-frequency field data, are generally unavailable.

In this work, we propose a workflow of low-frequency extrapolation for field band-limited data. To solve the problem of missing real low-frequency data for training, we train CycleGAN (Zhu *et al.* 2017), a semi-supervised learning method, using unpaired images of field band-limited shots, $\{h_i\}_{i=1}^{N}$, $h_i \in H$ and synthetic low-frequency shots, $\{l_i\}_{i=1}^{M}$, $l_i \in L$, in the time domain. In particular, 'unpaired' means that the correspondence between h_i and l_i does not exist and there is no information provided as to which h_i matches which l_i. Our field data example shows that the CycleGAN trained with only synthetic low-frequency shot gathers is able to extrapolate the low-frequency data for field data. Since a well-trained generative model creates data under the similar distribution as the training data in the output domain (the low-frequency shot gathers), we use the low-frequency wavelet in the simulation of the synthetic low-frequency data as the source signature for FWI with the extrapolated low frequencies. The inverted low-wavenumber velocity model is comparable with the benchmark model obtained by tomography and geological correction in the same wavenumber range.

2 METHOD

2.1 Low-frequency extrapolation with semi-supervised learning

We formulate low-frequency extrapolation from band-limited data as an image-to-image translation problem. The source domain H contains 2-D images of the band-limited shot gathers $\{h_i\}_{i=1}^{N}$ and the target domain L is a collection of 2-D images of low-frequency shot gathers $\{l_i\}_{i=1}^{M}$. Both input h and output l have equivalent dimensions of $nt \times ntr$ where nt and ntr are the number of time samples in each trace and the total number of receivers, respectively. Our goal is to learn a mapping function from an input high-frequency image to its output low-frequency image.

Learning the task of image-to-image translation with a supervised method requires paired training data, where the correspondence between h_i and l_i exists. Compared to training with synthetic data, training the neural network with real data can reduce the generalization error and provide much better performance when we apply the neural network to field data. However, the lack of availability of the low frequencies make it difficult to train the neural network with field data in a supervised manner. To solve the problem of missing real low-frequency data as labels when training directly with field data, we use Cycle-Consistent Adversarial Networks (CycleGAN (Zhu *et al.* 2017)), a semi-supervised method to learn a mapping from input to output images from unpaired training data. Specifically, 'unpaired' means that the high-frequency data h_i and low-frequency data l_i do not belong to the same shot gather.

Fig. 1 summarizes the flowchart of the CycleGAN. The CycleGAN contains two generators, $G_{HL}: H \rightarrow L$ and $G_{LH}: L \rightarrow H$, and associated adversarial discriminators, D_L and D_H. Since each sample in H is unlabelled, no explicit loss exists between $G_{HL}(H)$ and L. Instead, G_{HL} is trained via D_L to generate images from domain H indistinguishable from domain L by an adversarial loss of G_{HL} and D_L. Since training with unpaired images is highly underconstrained, we couple $G_{HL}: H \rightarrow L$ with an inverse mapping $G_{LH}: L \rightarrow H$ and introduce a cycle consistency loss to enforce $G_{LH}(G_{HL}(H)) \approx H$ (and vice versa). Likely, D_H encourages G_{LH} to generate images from domain L indistinguishable from domain H by training

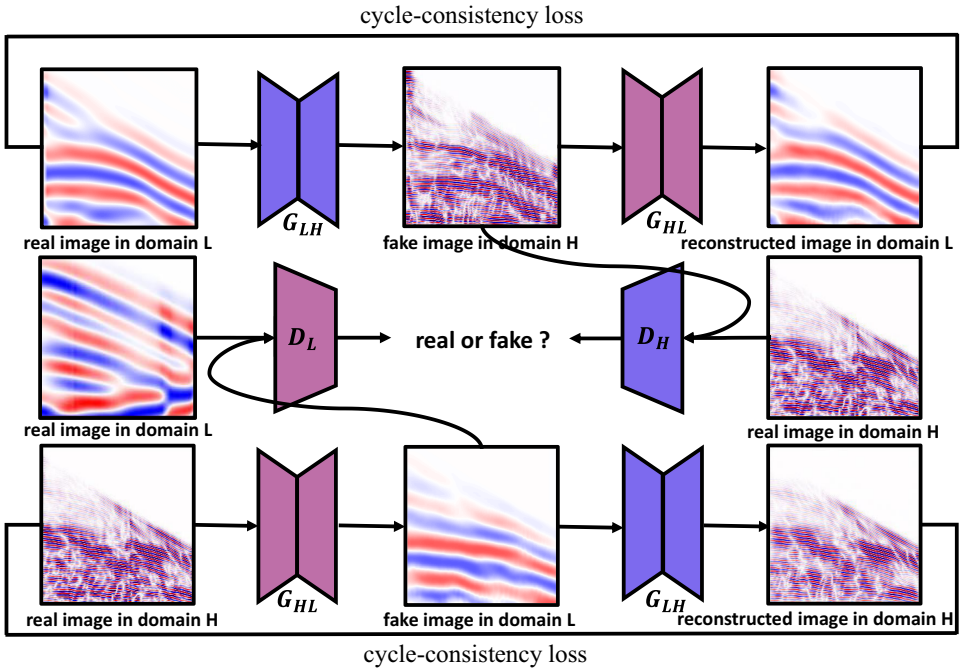

Figure 1. Flowchart of seismic bandwidth extension using CycleGAN. The deep-learning model contains two generators and two discriminators. Generator G_{HL} extrapolates low frequencies from their input high-frequency shot gathers in domain H. Discriminator D_L distinguishes between real and fake low-frequency shot gathers in domain L. Likewise, generator G_{LH} translates one low-frequency shot gather to its high-frequency counterpart. Discriminator D_H distinguishes between real and fake high-frequency shot gathers.

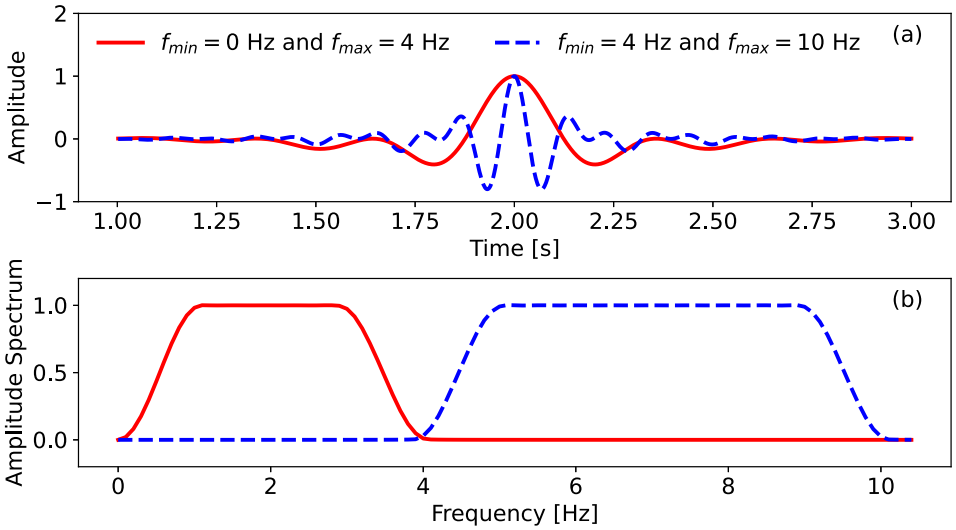

Figure 2. Source wavelets (a) in the time domain and (b) in the frequency domain. The band-limited data in the synthetic Marmousi experiment are simulated using the band-limited wavelet with $f_{\min} = 4$ Hz and $f_{\max} = 10$ Hz. The wavelet with $f_{\min} = 0$ Hz and $f_{\max} = 4$ Hz is the source signal $S(t)$ for the simulation of synthetic low-frequency shot gathers for both the Marmousi synthetic example and the field-data example.

an adversarial loss of G_{LH} and D_H. The specific formulas of the adversarial loss and cycle consistency loss of G_{HL}, G_{LH}, D_L and D_H are in the Appendix.

After training, we feed the band-limited data into G_{HL} and extrapolate the low frequencies of the field data shot by shot. The generated images are computational low frequencies of the field data. They contain information about the propagation media and can be used to infer the Earth structures. Since the adversarial loss is optimized to match the distribution of generated images to the data distribution in the target domain, the generated images share the same source signature and frequency components as the synthetic shot gathers in the target domain.

2.2 Training data sets

We design two examples to illustrate the workflow of our method: One synthetic example simulated on the Marmousi model and one

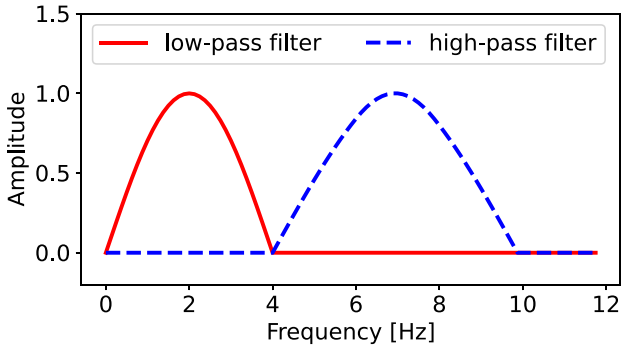

Figure 3. A low-pass filter with $f_{min} = 0$ Hz and $f_{max} = 4$ Hz and a high-pass filter with $f_{min} = 4$ Hz and $f_{max} = 10$ Hz. The high-pass filter is applied to the field data for preparing the 4–10 Hz band-limited data, that is the input high-frequency shot gathers in domain H.

field-data example from a marine field data set. The marine field data set contains 428 shots. Each shot is recorded by 648 receivers with an interval of 12.5 m. The total recording time is 6.6 s with a sampling rate of 6 ms. For simplicity, all simulations are performed using the same acquisition geometry and sampling rate as the field data.

The input high-frequency shot gathers in domain H come from the band-limited data for FWI. For the Marmousi example, we assume that data below 4 Hz are totally missing from the observed data and directly simulate 4–10 Hz band-limited data using a band-limited wavelet (blue dash line in Fig. 2). For the field data set, we apply a smooth high-pass filter (blue dash line in Fig. 3) to the field data and obtain 4–10 Hz band-limited shot gathers for the collection of H. We preprocess the raw field data shot-by-shot following these steps:

(i) Perform dip filtering to remove noise in the $f - k$ domain.
(ii) Filter the data into 4–10 Hz in the frequency domain.
(iii) Mute and remove everything before the approximated first arrivals in the time domain.

Each shot gather is considered as one sample in the input image domain H of the training data set.

To prepare images of synthetic shot gathers in L, we perform forward modelling by solving wave equations to simulate the low-frequency shot gathers. The source wavelet $S(t)$ contains only components in the target low-frequency range. To enrich the shot gathers in the training data set, we should design the training models with reasonable geological structures. The prior knowledge of the model that is being inverted, for instance, the water bottom depth, can be used when designing the training models.

We select two training models for the simulation of synthetic low-frequency shot gathers in L: one slice of the SEG/EAGE Overthrust model (Fig. 4a) for the Marmousi example and one von-Karman type random model (Fig. 4b) for the field-data example. The random model is a superposition of von-Karman type random media with a 1-D linear velocity model where velocities range from 1500 to 3000 m s^{-1} and are linearly increased with depth. With the same acquisition geometry as the field data, we simulate 428 low-frequency shot gathers on each training model by solving an isotropic acoustic wave equation. The wavelet (Fig. 2) with $f_{min} = 0$ Hz and $f_{max} = 4$ Hz is the source signal for the simulation of synthetic low-frequency shot gathers.

To train the neural network efficiently, we preprocess both real band-limited and synthetic low-frequency shot gathers with the following steps. First, all shot gathers are downsampled by a factor of three in the time axis to reduce the computation cost. After downsampling, the new sampling rate of each trace is 18 ms. Each image has a dimension of $nt = 364$ and $ntr = 648$. Secondly, all shot gathers are weighted by $gain(t) = te^{0.5t}$ to enhance the amplitudes of events arriving at a later time. Thirdly, all shots are normalized to one by dividing by the maximum amplitude trace-by-trace, which helps to compensate for weaker amplitudes at further offsets. Note that it is not necessary to recover the original amplitudes if FWI uses an amplitude-insensitive cost function.

2.3 Network architecture and training procedure

We adopt the neural network architecture from Zhu *et al.* (2017), which has shown great performance on many image-to-image translation tasks with unpaired labels. The generator sequentially contains one 7 × 7 convolutional layer (64 channels), two downsampling blocks (128 and 256 channels), nine residual blocks (256 channels in each block), two upsampling blocks (256 and 128 channels), and another 7 × 7 convolutional layer with one channel that maps feature to an output image. Each downsampling block contains one 3 × 3 convolutional layer with a stride of two. Each upsampling block carries one 3 × 3 transpose convolutional layer with a stride of two. We use reflection padding before convolution and instance normalization after convolution. For the discriminator, we use four 4 × 4 Convolution-InstanceNorm-LeakyReLU layers with a stride of two except the last layer where we use a stride of one. There are 64, 128, 256 and 512 filters, respectively, on the four convolutional layers. Instance normalization is used after convolutional layers except for the first layer. The slope of the leaky ReLUs is 0.2. We use a convolutional layer with one channel and a stride of one as the output layer to predict labels of one on 'real' images and zero on 'fake' images (see the Appendix).

During training, we use adaptive moment estimation (Adam, Kingma & Ba 2014) to simultaneously update G_{HL}, G_{LH}, D_L and D_H from scratch with a mini-batch of one. In practice, we slow down the update of discriminators compared with generators by dividing their loss functions by five while optimizing the discriminators. The learning rate is set as 2×10^{-5}. Weights are initialized from a Gaussian distribution with zero mean and a standard deviation of 0.02.

2.4 Extrapolated FWI

With the extrapolated low-frequency data, we are able to mitigate the cycle-skipping problem of FWI on an uninformative initial model using extrapolated FWI (Li & Demanet 2016; Sun & Demanet 2020). Extrapolated FWI contains two inversion stages. In the first stage, we perform FWI using the generated low-frequency data and the designed source wavelet $S(t)$. Low-frequency data are much less affected by the cycle-skipping problem of FWI and can thus provide a good initial model for the subsequent inversion using band-limited data. In the second stage, with the resulting low-wavenumber velocity model from the first inversion stage, we perform FWI with band-limited data to resolve finer velocity structures of the subsurface. We can extract another source wavelet from the field data for the inversion of band-limited data. However, in our example, a synthetic band-limited wavelet seems to be sufficient for the inversion

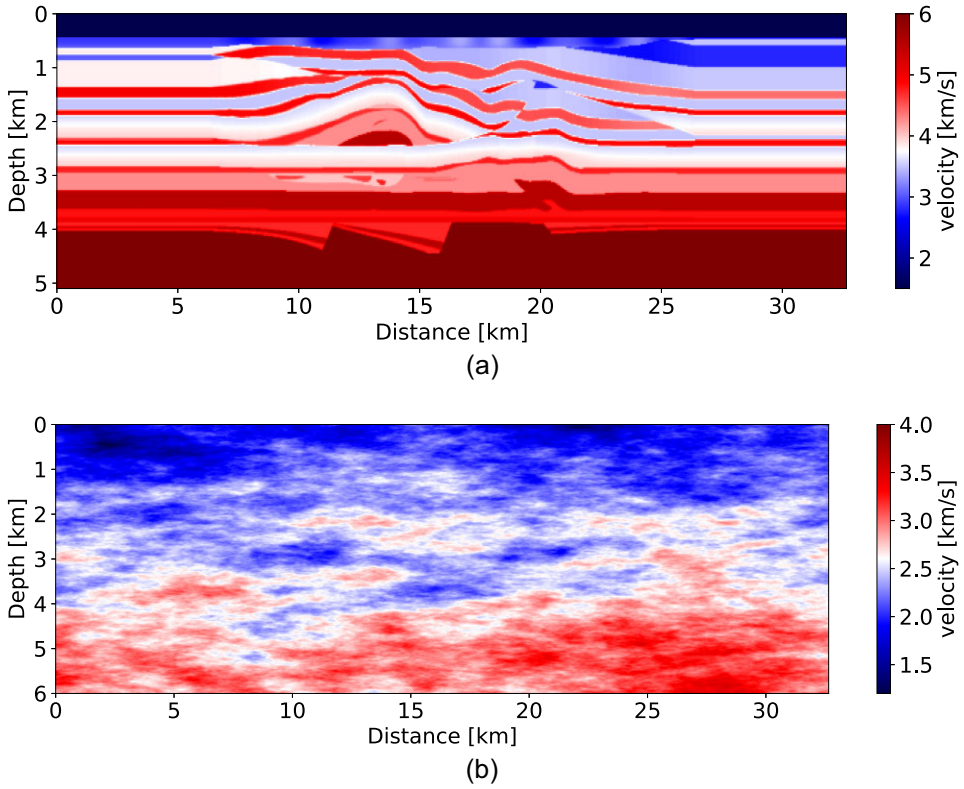

Figure 4. Training models for the simulation of synthetic low-frequency shot gathers. (a) One slice of the SEG/EAGE overthrust model for the Marmousi example. The model has been extended from both sides. (b) Random media characterized by a von Karman correlation function for the field-data example.

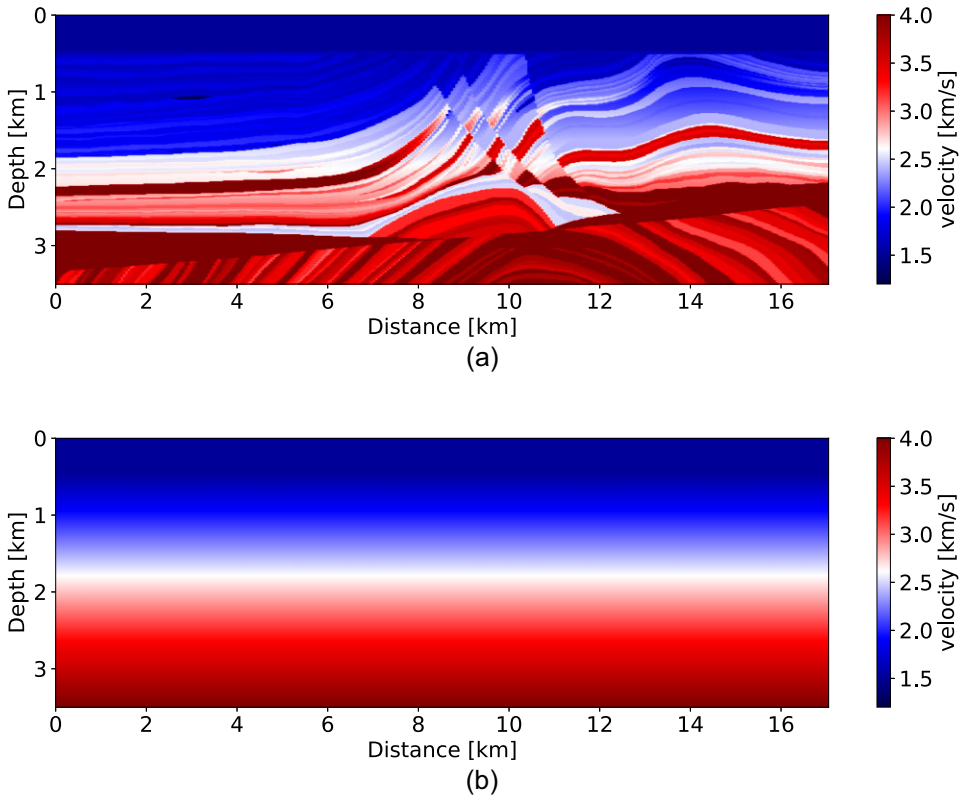

Figure 5. Synthetic data example: (a) Marmousi model and (b) linear initial velocity model.

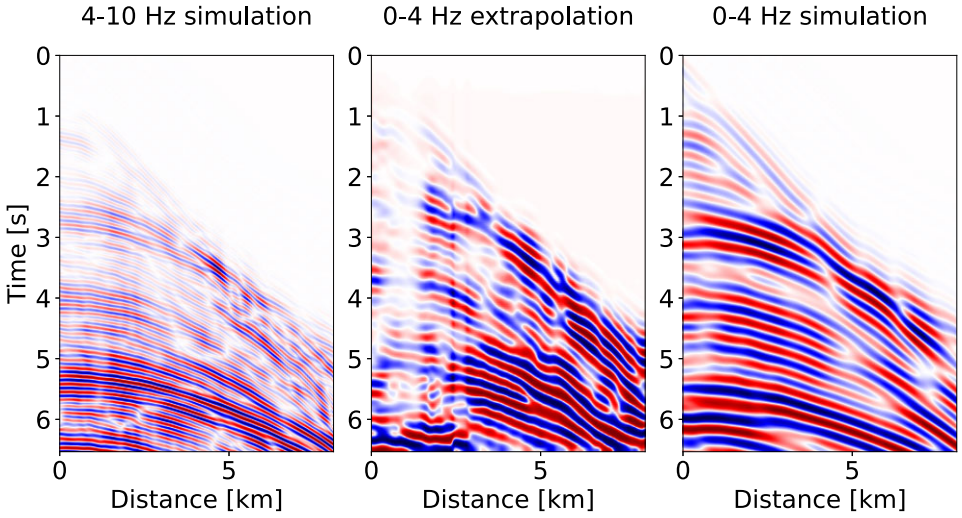

Figure 6. Comparison of 4–10 Hz band-limited data, 0–4 Hz extrapolated low-frequency data, and 0–4 Hz simulated low-frequency data.

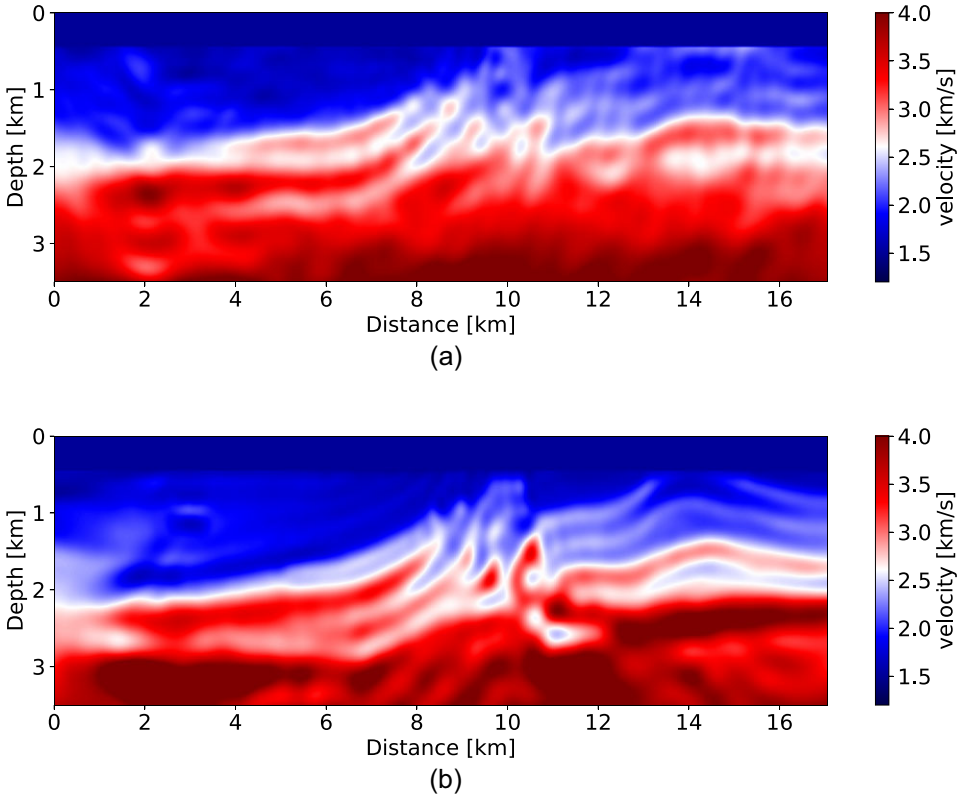

Figure 7. FWI results started from the linear initial velocity model with only (a) 0–4 Hz extrapolated low-frequency data and (b) 0–4 Hz simulated low-frequency data for the Marmousi model. A multiscale approach is used by first performing the inversion with 0–2 Hz and then 0–4 Hz data.

of the band-limited field data. We choose the L_2 norm of the normalized (trace by trace) difference in time domain as the objective function of FWI in both inversion stages (Shen 2010, 2014). With this objective function, the observed data and calculated data have approximately the same relative amplitudes, making the inversion focus more on phase comparison.

3 NUMERICAL RESULTS

In this section, we demonstrate the performance of low-frequency extrapolation with the semi-supervised learning method using both synthetic and field-data experiments. For the synthetic experiment, we evaluate extrapolation results by directly comparing with

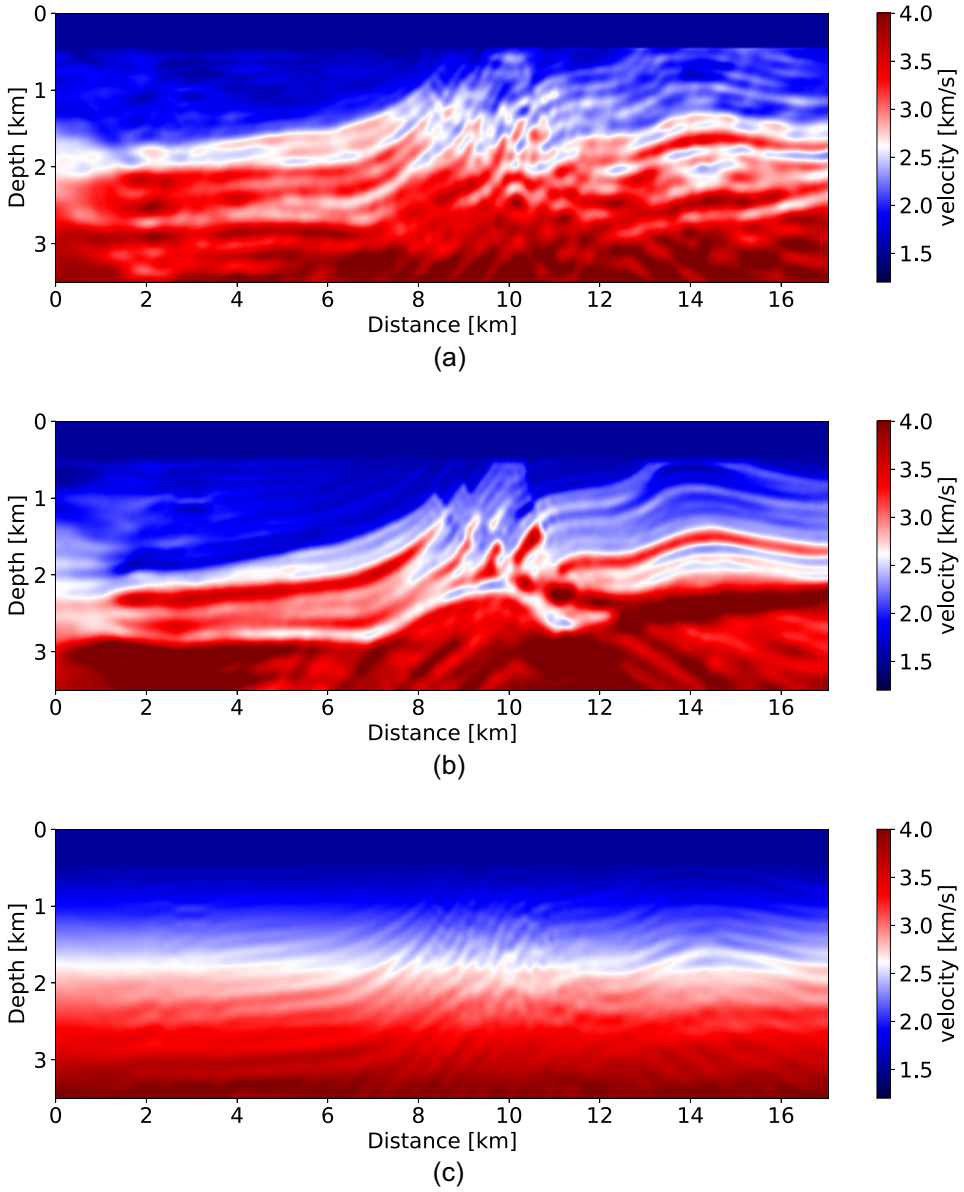

Figure 8. FWI results using 4–7 Hz band-limited data. (a) EFWI result started from the low-wavenumber velocity model from only 0 to 4 Hz extrapolated low-frequency data in Fig. 7(a). (b) FWI result started from the low-wavenumber velocity model from 0 to 4 Hz simulated low-frequency data in Fig. 7(b). (c) FWI result directly started from the linear initial velocity model in Fig. 5(b).

simulated low frequencies. As for the field data, since real low frequencies are unavailable, we evaluate results by comparing FWI resulting models with a benchmark model obtained by tomography followed with human correction and interpretation.

3.1 Synthetic data example: Marmousi model

We use the *P*-wave velocity of the Marmousi 2 model as the true model for the synthetic experiment (Fig. 5a).The starting model is a 1-D linear velocity model where velocities are linearly increased with depth from the water bottom to the maximum depth (Fig. 5b). After training with the overthrust training data set, we feed the 4–10 Hz band-limited data to the trained generator G_{HL} for the extrapolation of 0–4 Hz low-frequency data shot-by-shot.

Fig. 6 compares the 4–10 Hz input band-limited data, 0–4 Hz extrapolated low-frequency data and 0–4 Hz simulated low-frequency data for one shot on the Marmousi model. We observe that the extrapolated result shares comparable low-frequency ranges as the simulated ones. The arrival time of several events with large amplitude aligns well with their high-frequency counterparts on the band-limited shot gather. However, some mismatches exist between extrapolation and simulation, especially for events with weak amplitudes. Although extrapolation accuracy depends on network architecture, loss formulation, and other hyperparameters of learning, we speculate that low-frequency extrapolation with semi-supervised learning seems to be less constrained compared to supervised-learning methods, and thus we should expect an approximation of the low-frequency signals rather than the exact low-frequency events as those simulated directly with a low-frequency wavelet.

Fig. 7 compares FWI results started from the 1-D linear initial velocity model with only 0–4 Hz extrapolated and simulated

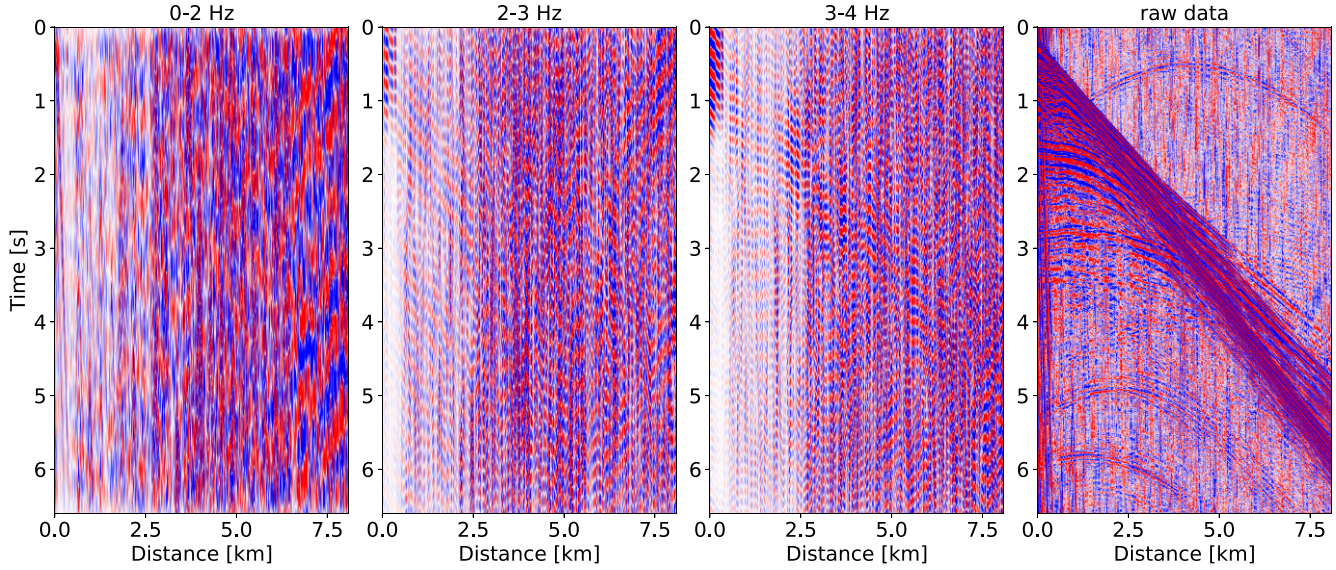

Figure 9. Field-data properties. Comparison of one shot gather from the field data in the 0–2Hz, 2–3 Hz, 3–4 Hz and full bands (raw data). The data below 3 Hz are fully contaminated by noise. The effective signal starts to be visible in the 3–4 Hz band and becomes clear in the band above 4 Hz.

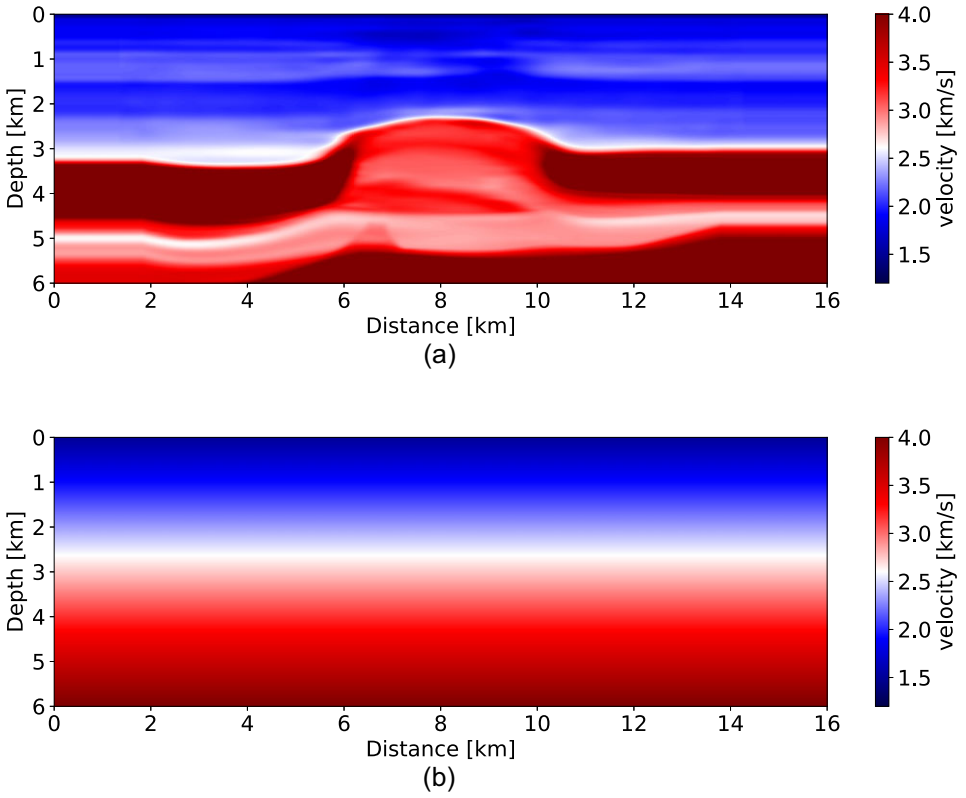

Figure 10. (a) The benchmark model of the field data obtained by reflection travel-time tomography following additional human interpretation and correction. (b) An uninformative initial model which is far from the benchmark model.

low-frequency data. A multiscale approach is used by first performing the inversion with 0–2 Hz and then 0–4 Hz data. Despite the mismatch between extrapolated and simulated low frequencies, the FWI result with only extrapolated 0–4 Hz data shares similar velocity structures as the result with simulated data. However, without regularization, inversion with only extrapolated data shows high-frequency artefacts, owing to the imperfect extrapolation.

With the low-wavenumber results as initial models, we continue inversion with 4–7 Hz band-limited data. Compared with the inversion result started from the 0–4 Hz simulated data (Fig. 8b), FWI started from the extrapolated data (Fig. 8a) reveals several major velocity structures similar to those from the true data. However, inevitable extrapolation errors with the less constraint semi-supervised learning approach bring inversion artefacts to the resulting model from only the extrapolated data. Thus, it is challenging

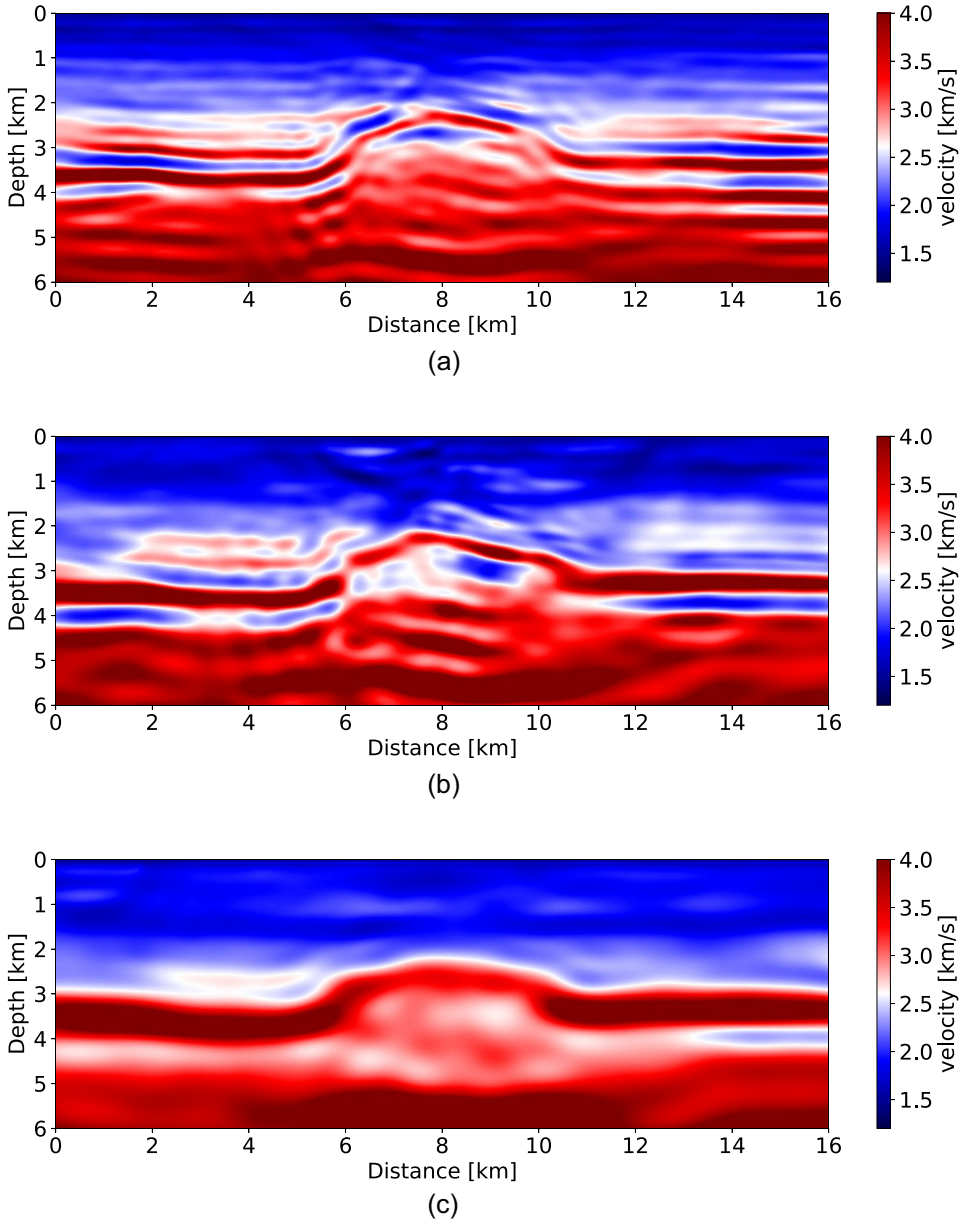

Figure 11. Determination of the starting frequency band to avoid cycle-skipping of FWI for the field data using the uninformative initial model in Fig. 10b. FWI results using synthetic data simulated on the benchmark model in Fig. 10a. The maximum frequencies of the synthetic data are (a) 4 Hz, (b) 3 Hz and (c) 2 Hz, respectively. In each result, we use the same source wavelets in both forward modelling and inversion. The maximum starting frequency should be less than 2 Hz for FWI with the uninformative initial model.

for FWI with band-limited data to deal with the artefacts in the initial model. Regularization for the inversion with only extrapolated data may be helpful to mitigate this problem. Nevertheless, compared with the inversion result directly started from the linear initial model (Fig. 8c), the extrapolated data mitigate the cycle-skipping problem for the inversion with the linear initial model to a large extent. The synthetic experiment confirms the effectiveness of our method to relieve the dependence of FWI on initial models.

3.2 Field-data example

We use a field seismic data set recorded by a streamer survey in a marine environment to illustrate the performance of our method. Fig. 9 shows one shot gather from the field data set. We first check the minimum available frequencies of the field data. Data below 3 Hz are fully contaminated by noise. The effective signal starts to be visible in the 3–4 Hz band and becomes clear in the band above 4 Hz (Fig. 9). Hence, we decide to remove the data below 4 Hz from the raw shot gather and use 4–10 Hz data to extrapolate the low frequencies below 4 Hz.

A conventional velocity model building workflow contains both tomography and FWI when tomography provides an initial model for the subsequent FWI. Fig. 10(a) shows the tomography result of the field data. The tomography model is obtained by reflection traveltime tomography following additional human effort to correct the boundary of the high-velocity layer in the deep area. Using the tomography model as a benchmark, the goal of our work is to build

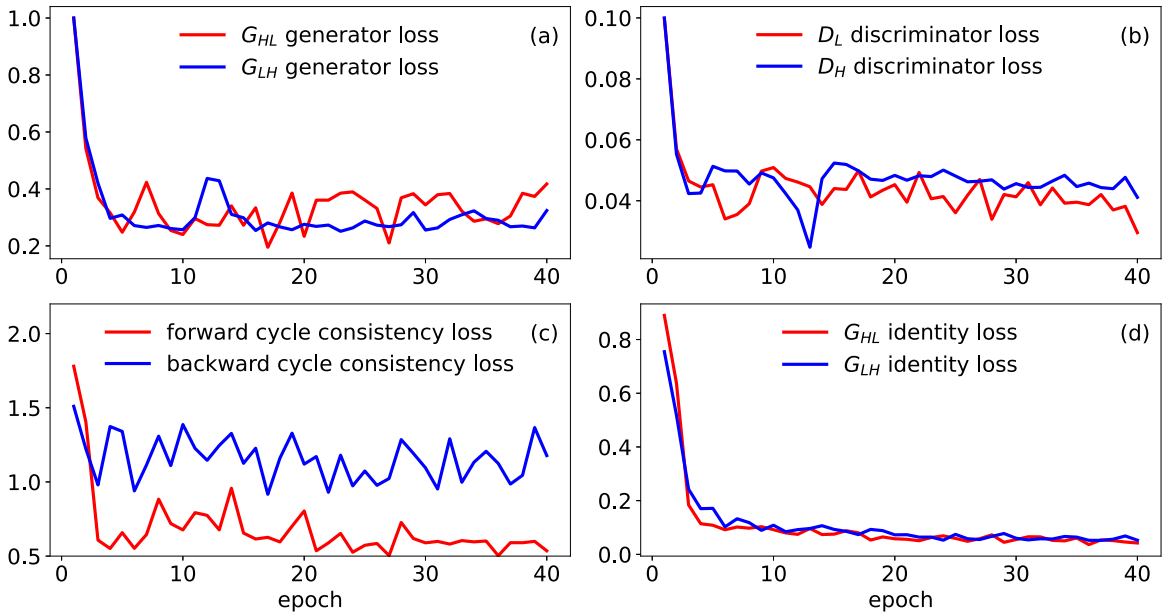

Figure 12. Learning curves after training with 40 epochs for the field-data example. (a) Generator loss of $\mathcal{L}_{\text{LSGAN}}(G_{HL}, H)$ and $\mathcal{L}_{\text{LSGAN}}(G_{LH}, L)$. (b) Discriminator loss of $\mathcal{L}_{\text{LSGAN}}(D_L, H, L)$ and $\mathcal{L}_{\text{LSGAN}}(D_H, H, L)$. (c) Forward cycle-consistency loss $\mathcal{L}_{\text{cycle}}(G_{HL}, G_{LH}, H)$ and backward cycle-consistency loss $\mathcal{L}_{\text{cycle}}(G_{LH}, G_{HL}, L)$. (d) Identity loss $\mathcal{L}_{\text{identity}}(G_{HL}, L)$ and $\mathcal{L}_{\text{identity}}(G_{LH}, H)$.

an initial model for FWI using only the extrapolated low-frequency data. The extrapolated low-frequency data are meaningful if the FWI result starts with the uninformative initial model (Fig. 10b) and using only the extrapolated low-frequency data is comparable with the tomography model in the same wavenumber range.

3.2.1 Determination of starting frequency band to avoid cycle-skipping

The low-frequency data below 4 Hz are missing from the field data. FWI may suffer from the cycle-skipping problem if we directly use 0–4 Hz extrapolated data in inversion given the poor initial model (Fig. 10b). Hence, we perform a benchmark test using synthetic low-frequency data simulated on the benchmark model to determine the starting frequency band for FWI. First, we simulate 0–2 Hz, 0–3 Hz and 0–4 Hz 'observed' data on the benchmark model using low-frequency wavelets with maximum frequencies of 2, 3 and 4 Hz, respectively. Then, we perform FWI using the synthetic 'observed' data to recover the benchmark model starting with the given initial model. With only *P*-wave velocities, both forward modelling and inversion are performed using an isotropic acoustic wave equation.

The benchmark result shows that strong artefacts exist on the inverted models (Figs 11a and 11b) using the synthetic 0–4 Hz or 0–3 Hz data. Instead, in spite of small artefacts, FWI using 0–2 Hz synthetic low-frequency data can reconstruct the low wavenumber velocity structure of the benchmark model (Fig. 11c). Therefore, we need to seed FWI using 0–2 Hz low-frequency data given the uninformative initial model for the field data.

3.2.2 Frequency extrapolation and FWI results

Fig. 12 shows the learning curves after training with 40 epochs for the field data. Because of the adversary between generators and discriminators, the discriminator tends to predict an average probability of 0.5 on both real and fake images if the neural networks are well-trained. With a scale factor of 0.1, the discriminator loss should finally stabilize around $0.1 \times [(1-0.5)^2 + (0-0.5)^2] = 0.05$ (eqs A.2 and A.8) where 1 and 0 are the labels for real and fake images, respectively (the Appendix). Likewise, the generator loss should stabilize around $(1-0.5)^2 = 0.25$ (eqs A.1 and A.10). Both generator losses (G_{HL} and G_{LH}) and discriminator losses (D_L and D_H) stabilize around the expected values, which indicates that D_L and D_H approximately output an average probability of 0.5 on a real image, as well as an average probability of 0.5 on a fake image. A well-trained CycleGAN model would generate images that share the same characteristics as the provided synthetic low-frequency images and real high-frequency images.

After training the CycleGAN with synthetic shot gathers in the band of 0–4 Hz, we extrapolate the 0–4 Hz data from the field 4–10 Hz band-limited data. Fig. 13 shows results of the CycleGAN trained with 40 epochs for the extrapolation of $0-4$ Hz data. For the purpose of quality control, we check both forward and backward cycle consistencies using two shot gathers randomly selected from field band-limited and synthetic low-frequency data sets, respectively. $G_{LH}(G_{HL}(h)) \approx h$ and $G_{HL}(G_{LH}(l)) \approx l$ are approximately satisfied on the results. However, we observe that $G_{LH}(G_{HL}(h))$ did not recover all the information on h, implying that the extrapolation from low to high frequencies is harder than the opposite direction. Even so, $G_{LH}(l)$ seems to provide sufficient high-frequency information for G_{HL} to recover l. Overall, the positions of the low-frequency events are comparable with those on the band-limited shot gathers. Since the low-frequency data below 4 Hz are missing, it is impossible to check the accuracy of the extrapolated data by comparing them with the real data. Hence, we perform FWI using the extrapolated data and compare the results with models inverted using synthetic data simulated on the benchmark model to check the reliability of our extrapolation.

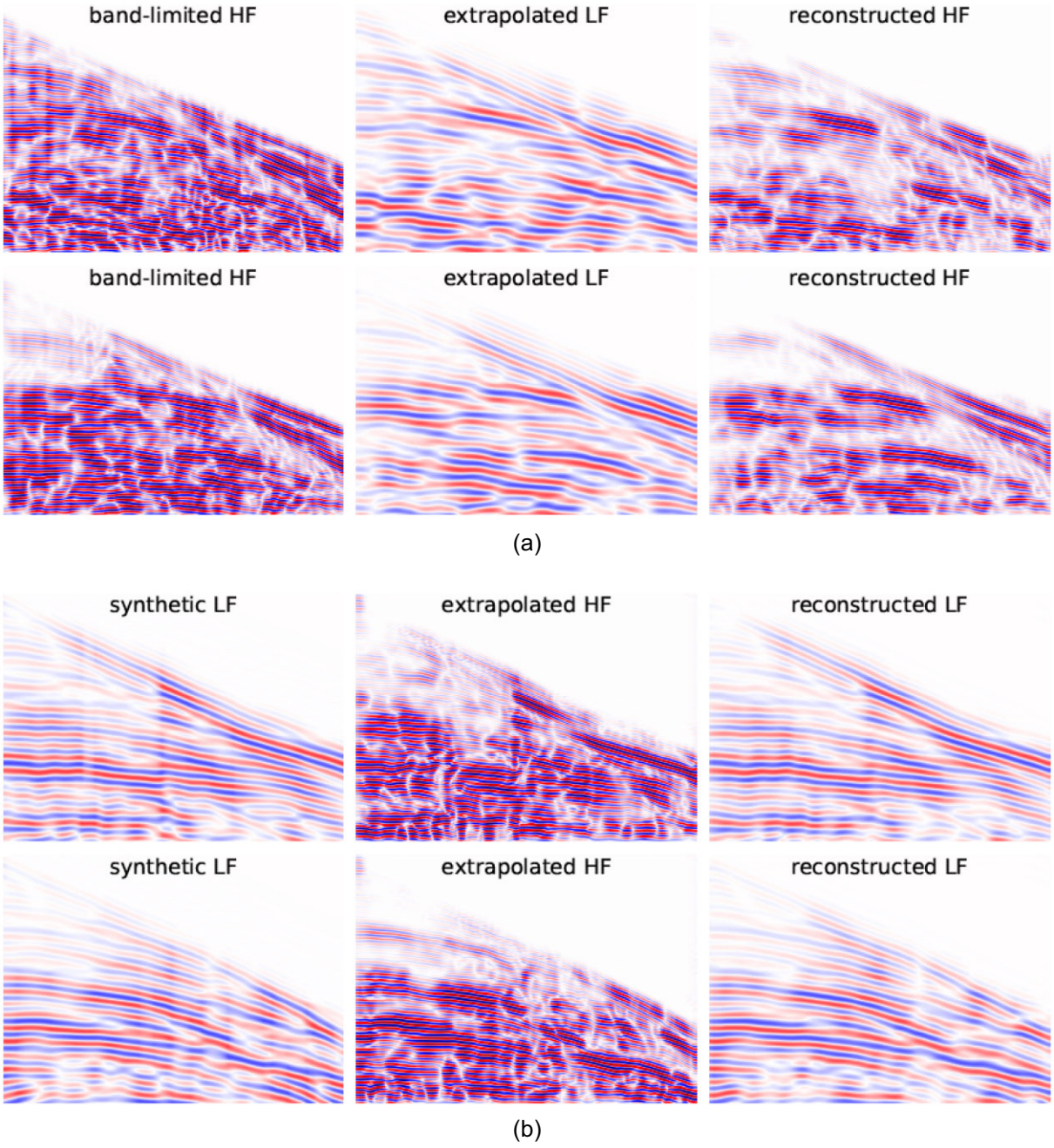

Figure 13. Frequency extrapolation results for the field data: (a) Forward cycle starting from 4–10Hz field band-limited data, $h \rightarrow G_{HL}(h) \rightarrow G_{LH}(G_{HL}(h))$. The second column shows the resulting 0–4Hz data extrapolated from the field band-limited data. (b) Backward cycle starting from 0–4Hz synthetic low-frequency data, $l \rightarrow G_{LH}(l) \rightarrow G_{HL}(G_{LH}(l))$. The cycle consistency loss enforces $G_{LH}(G_{HL}(h)) \approx h$ and $G_{HL}(G_{LH}(l)) \approx l$.

We perform FWI using the 0–4 Hz extrapolated data. The wavelet with $f_{\min} = 0$ Hz and $f_{\max} = 4$ Hz (Fig. 2) is used as the source signature. The extrapolated data generated by the neural network are directly used in FWI without post-processing. Owing to inevitable errors in extrapolation, we remove artefacts less than 500 m by smoothing the gradients calculated using the extrapolated data during inversion. Since the starting frequency band should not be larger than 2 Hz on the given initial model (Fig. 10b), we perform inversion firstly with 0–2 Hz data and then continue with 0–4 Hz data. Starting with the uninformative initial model, Fig. 14a shows the FWI result using only 0–4 Hz extrapolated data. The inverted model is comparable with the FWI result (Fig. 14b) using the simulated 0–4 Hz data on the tomography model, which means that the accuracy of the extrapolated 0–4 Hz data is sufficient to build a low-wavenumber velocity model from the uninformative initial model.

Starting with the low-wavenumber velocity models in Fig. 14, we continue FWI using 3–10 Hz band-limited field data with a multiscale approach. In particular, FWI starting with the extrapolated low-frequency data successfully reveals the positions of gas clouds indicated by the low velocity zones (Fig. 15a). The inversion results

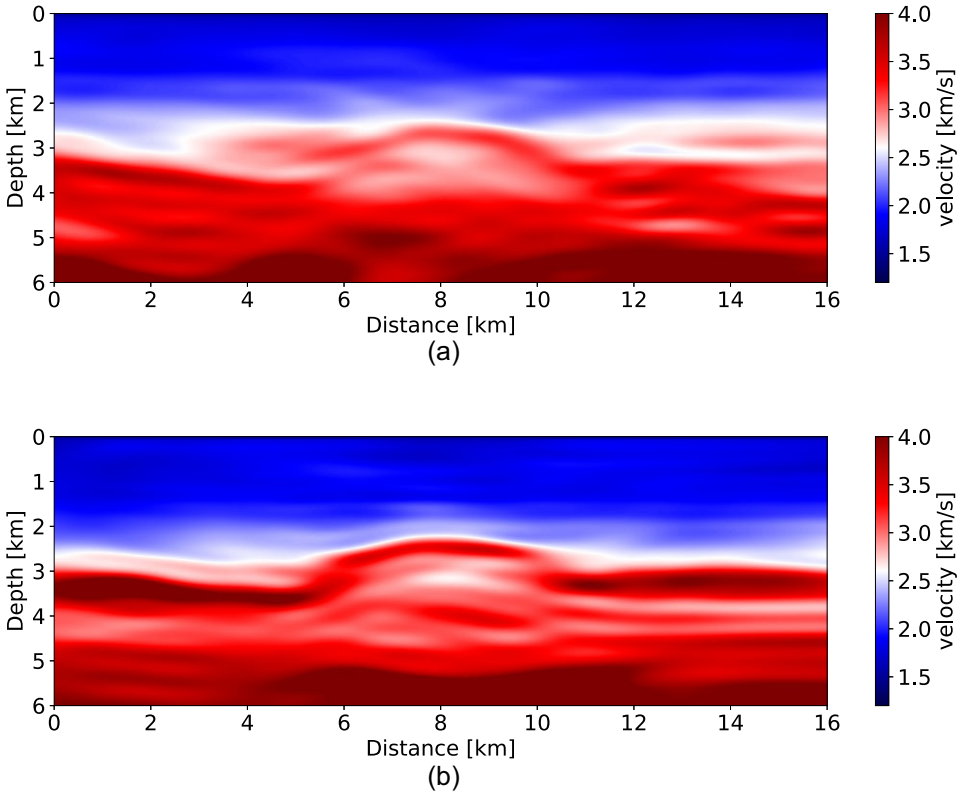

Figure 14. Comparison of FWI results started from the uninformative linear velocity model (Fig. 10b) using (a) extrapolated 0–4 Hz data and (b) synthetic 0–4 Hz data simulated on the benchmark model (Fig. 10a). In each result, we use a multiscale approach (Bunks *et al.* 1995) with 0–2 Hz and 0–4 Hz data.

starting with the 0–4 Hz extrapolated data and 0–4 Hz simulated data on the tomography model are very similar. However, FWI directly from the uninformative initial model using the band-limited field data falls into a local minimum and cannot reconstruct the correct geological structures for this area (Fig. 15c). Fig. 16 compares calculated and observed data before and after FWI using 3–6 Hz band-limited data. We clearly see cycle-skipping of events in FWI initiated without low frequencies. By contrast, less cycle-skipping exists in the inversion initiated with extrapolated low frequencies, which is consistent with the inversion initiated with simulated low frequency data. The inversion results with field data demonstrate that the extrapolated low-frequency data are reliable to mitigate the cycle-skipping problem of FWI starting with an uninformative initial model with band-limited data.

3.3 Discussion and limitations

3.3.1 Influence of training models on extrapolation accuracy

Our previous work (Sun & Demanet 2020) shows that the neural network trained for low-frequency extrapolation can generalize among very different geological structures (i.e. from a layer to a salt model). In this field data example, we also investigate the influence of training models on extrapolation accuracy. In addition to the training model which resembles salt structures, we also train the neural network on synthetic data simulated on the Marmousi model. Following the same procedure, we perform FWI using the low-frequency data extrapolated by the neural network trained on the Marmousi model. Although the background velocity is correctly decreased from the initial model, inversion artefacts occur on the sides of the model. The extrapolation accuracy of the shot gathers at these locations where artefacts exist is not as good as those trained on the random model. Comparing with the results in Sun & Demanet (2020, 2022a), we speculate that shot-by-shot extrapolation is more difficult to generalize than trace-by-trace extrapolation since the hidden geological information on the shot gathers may be remembered by the neural network. Therefore, the training models in the shot-by-shot experimental setup should be more diverse and include representative geological structures as in the model that is being inverted.

3.3.2 Stability of extrapolation: variations among epochs

Without regularization, FWI results from the conventional L_2-norm misfit are very sensitive to the accuracy of the extrapolated low-frequency data. If the neural network is not well trained, the low-frequency data generated by the neural network may vary among epochs. In addition to checking the generated shot gathers, it is necessary to compare the FWI results on generated data among different epochs and then decide on a good solution. If the neural network is well trained, inversion should be stable among epochs.

3.3.3 Limitations and future work

The method proposed in this study directly uses field band-limited data as one part of the training data set. As a result, we may need to retrain the neural networks from scratch to achieve superior

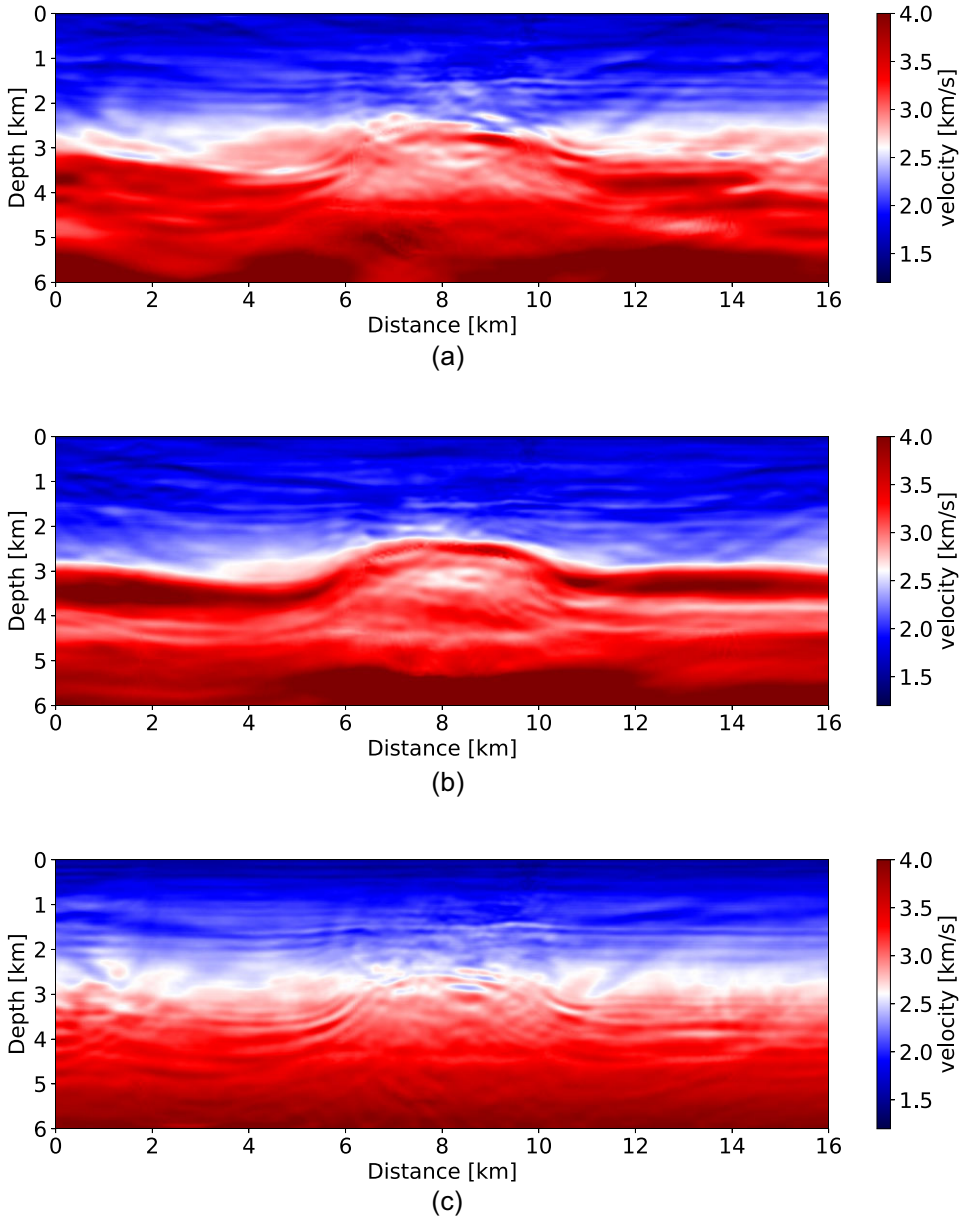

Figure 15. Comparison among FWI results using 3–10 Hz field band-limited data. In (a), the resulting model starts with the low-wavenumber velocity model inverted using 0–4 Hz extrapolated data (Fig. 14a). In (b), the resulting model starts with the low-wavenumber velocity model inverted using the 0–4 Hz synthetic data simulated on the benchmark model (Fig. 14b). In (c), the starting model is the uninformative initial model (Fig. 10b).

performance on each new field data set. Since training is much more time-consuming than testing, the proposed method may not be efficient when applied to a large 3-D seismic data volume. Randomly selecting partial shot gathers from the whole field data set as a training data set may be sufficient to train the neural networks. However, the extrapolation accuracy of the remaining unseen shot gathers may be degraded, depending on the generalization ability of the designed neural network.

Due to inevitable errors of extrapolation, involving more fake data before real data may accumulate inversion errors during the frequency sweep of FWI. Monitoring the inversion results using fake data is therefore crucial to build a reasonable initial model for inversion with band-limited data. Otherwise, we may continue to modify the architecture of generators and discriminators within the CycleGAN framework in order to improve the extrapolation accuracy.

Another undetermined factor that may change the extrapolation accuracy is attenuation. Since the neural networks are trained to generate data under the distribution of output data, only features that exist in the synthetic output data can be learned during the training procedure. Since our synthetic low-frequency shot gathers are simulated with an isotropic wave equation, the neural network may have to learn to explain attenuation and anisotropy effects present in real data but under the isotropic acoustic approximation. The extrapolated data may be contaminated by this lack of physics if the implicit data conversion is not effective. However, it is unclear if accounting for real physics in simulation would provide a better performance since learning wavefield conversion in addition

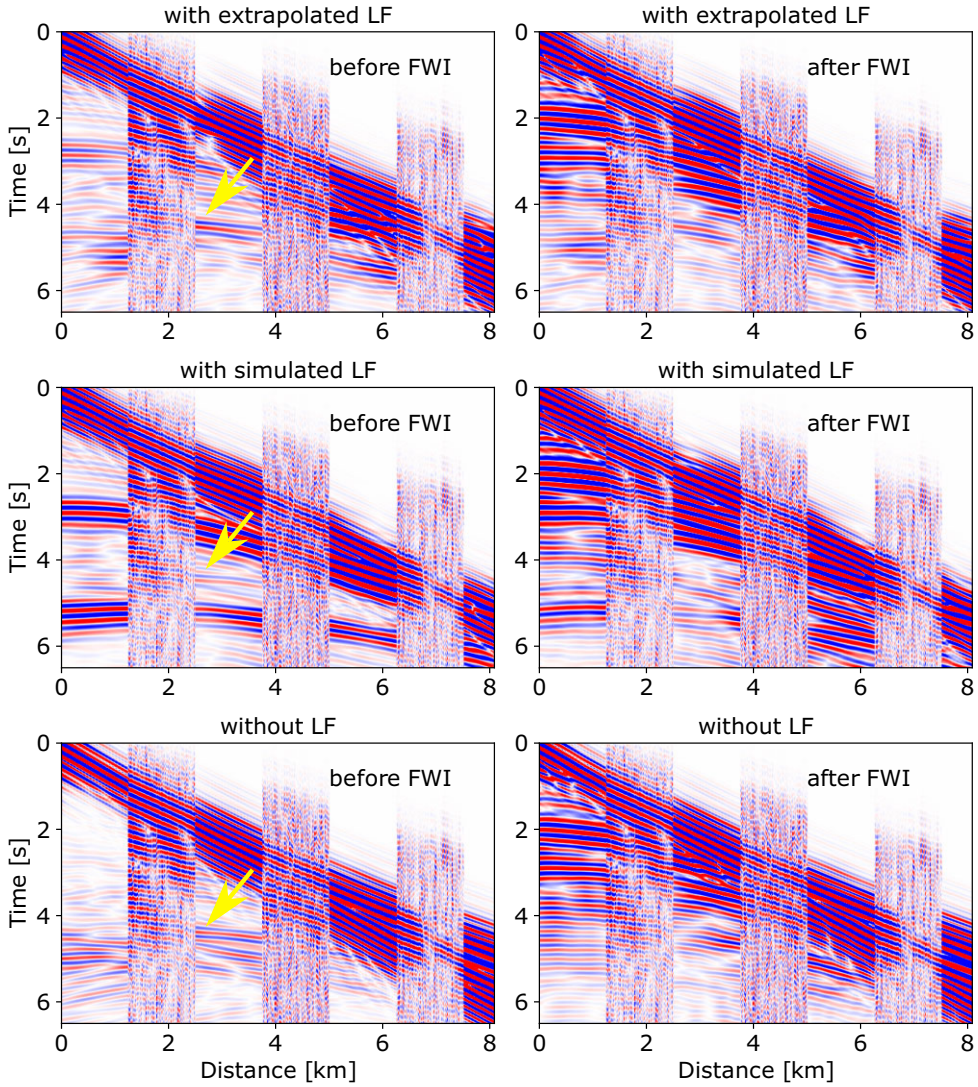

Figure 16. Interval comparison between 3–6Hz calculated and observed recordings. The calculated data are simulated on the velocity models recovered by FWI with 3–6 Hz band-limited data after initiated with extrapolated low frequencies (LF), with simulated LF and without LF.

to extrapolation would be more complex. Furthermore, attenuation is frequency-dependent. Hence, it is necessary to study how attenuation may affect low-frequency extrapolation, which we leave for future work.

4 CONCLUSIONS

We propose a semi-supervised learning method to extrapolate the missing low-frequency data from field band-limited data in an active-source seismic survey. With unpaired images of field band-limited and synthetic low-frequency shot gathers simulated on training models, we can train the CycleGAN for low-frequency extrapolation of the field data without real labels. Although we directly use the field data in training, real low-frequency data are not required. The FWI result using only the 0–4 Hz extrapolated data is comparable with that using 0–4 Hz synthetic data simulated on the benchmark model. The wavelet to simulate the synthetic low-frequency data on the training model should be used for FWI using the extrapolated low frequencies. An amplitude-insensitive objective function is preferred for FWI since the original amplitudes may be lost during pre-processing. Our field data results validate that the extrapolated data are reliable for building a low wavenumber velocity model and for mitigating the cycle-skipping problem of FWI starting with an uninformative initial model.

ACKNOWLEDGMENTS

We thank the editor Andrew Valentine and three anonymous reviewers for providing valuable suggestions. This research is funded by TotalEnergies. HS is also supported by MIT MathWorks Science Fellowship. We thank Fuchun Gao, Mohamed Dolliazal and Russell Jones for additional technical support of this work. We thank TotalEnergies for permission to publish this work.

DATA AVAILABILITY

The data underlying this research belong to TotalEnergies and cannot be shared publicly.

REFERENCES

Adler, A., Araya-Polo, M. & Poggio, T., 2021. Deep learning for seismic inverse problems: toward the acceleration of geophysical analysis workflows, *IEEE Sig. Process. Mag.,* **38**(2), 89–119.

Aharchaou, M. & Baumstein, A., 2020. Deep learning-based artificial bandwidth extension: training on ultrasparse OBN to enhance towed-streamer FWI, *Leading Edge,* **39**(10), 718–726.

Alkhalifah, T., Wang, H. & Ovcharenko, O., 2022. MLReal: bridging the gap between training on synthetic data and real data applications in machine learning, *Artif. Intell. Geosci.,* **3,** 101–114.

Araya-Polo, M., Jennings, J., Adler, A. & Dahlke, T., 2018. Deep-learning tomography, *Leading Edge,* **37**(1), 58–66.

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

Cai, A., Qiu, H. & Niu, F., 2022. Semi-supervised surface wave tomography with Wyzasserstein cycle-consistent GAN: method and application to Southern California Plate Boundary Region, *J. geophys. Res.,* **127**(3), e2021JB023598, doi:10.1029/2021JB023598.

Chen, Y. & Saygin, E., 2021. Seismic inversion by hybrid machine learning, *J. geophys. Res.: Solid Earth,* **126**(9), e2020JB021589, doi:10.1029/2020JB021589.

Choi, Y., Jo, Y., Seol, S.J., Byun, J. & Kim, Y., 2021. Deep learning spectral enhancement considering features of seismic field data, *Geophysics,* **86**(5), 1–60.

Dhara, A. & Sen, M.K., 2022. Physics-guided deep autoencoder to overcome the need for a starting model in full-waveform inversion, *Leading Edge,* **41**(6), 375–381.

Du, M., Cheng, S. & Mao, W., 2022. Deep-learning-based seismic variable-size velocity model building, *IEEE Geosci. Remote Sens. Lett.,* **19,** 1–5.

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

Fabien-Ouellet, G., 2020. Low-frequency generation and denoising with recursive convolutional neural networks, in *SEG Technical Program Expanded Abstracts 2020,* pp. 870–874, Society of Exploration Geophysicists.

Fang, J., Zhou, H., Elita Li, Y., Zhang, Q., Wang, L., Sun, P. & Zhang, J., 2020a. Data-driven low-frequency signal recovery using deep-learning predictions in full-waveform inversion, *Geophysics,* **85**(6), A37–A43.

Fang, Z., Fang, H. & Demanet, L., 2020b. Deep generator priors for Bayesian seismic inversion, *Adv. Geophys.,* **61,** 179–216.

Feng, S., Lin, Y. & Wohlberg, B., 2021. Multiscale data-driven seismic full-waveform inversion with field data study, *IEEE Trans. Geosci. Remote Sens.,* **60,** 1–14.

Geng, Z., Zhao, Z., Shi, Y., Wu, X., Fomel, S. & Sen, M., 2022. Deep learning for velocity model building with common-image gather volumes, *Geophys. J. Int.,* **228**(2), 1054–1070.

He, Q. & Wang, Y., 2021. Reparameterized full-waveform inversion using deep neural networks, *Geophysics,* **86**(1), V1–V13.

Herrmann, F.J., Siahkoohi, A. & Rizzuti, G., 2019. *Learned imaging with constraints and uncertainty quantification,* preprint (arXiv preprint arXiv:1909.06473).

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.

Jin, Y., Hu, W., Wang, S., Zi, Y., Wu, X. & Chen, J., 2022a. Efficient progressive transfer learning for full-waveform inversion with extrapolated low-frequency reflection seismic data, *IEEE Trans. Geosci. Remote Sens.,* **60,** doi:10.1109/TGRS.2021.3129810.

Jin, Y., Zi, Y., Hu, W., Hu, Y., Wu, X. & Chen, J., 2022b. A robust learning method for low-frequency extrapolation in gpr full waveform inversion, *IEEE Geosci. Remote Sens. Lett.,* **19,** 1–5.

Kazei, V., Ovcharenko, O., Plotnitskii, P., Peter, D., Zhang, X. & Alkhalifah, T., 2021. Mapping full seismic waveforms to vertical velocity profiles by deep learning, *Geophysics,* **86**(5), R711–R721.

Kingma, D.P. & Ba, J., 2014. *Adam: A method for stochastic optimization,* preprint (arXiv:1412.6980).

Li, M., Demanet, L. & Zepeda-Núñez, L., 2021. *Accurate and robust deep learning framework for solving wave-based inverse problems in the super-resolution regime,* preprint (arXiv:2106.01143).

Li, M., Demanet, L. & Zepeda-Núñez, L., 2022. Wide-band butterfly network: stable and efficient inversion via multi-frequency neural networks, *Multiscale Model. Simulat.,* **20**(4), 1191–1227.

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

Lin, Z., Guo, R., Li, M., Abubakar, A., Zhao, T., Yang, F. & Xu, S., 2021. Low-frequency data prediction with iterative learning for highly nonlinear inverse scattering problems, *IEEE Trans. Microwave Theory Techniq.,* **69**(10), 4366–4376.

Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z. & Paul Smolley, S., 2017. Least squares generative adversarial networks, in *Proceedings of the IEEE International Conference on Computer Vision,* pp. 2794–2802.

Mosser, L., Kimman, W., Dramsch, J., Purves, S., De la Fuente Briceño, A. & Ganssle, G., 2018. Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks, in *Proceedings of the 80th EAGE Conference and Exhibition 2018,* Vol. **2018,** pp. 1–5, European Association of Geoscientists & Engineers.

Mosser, L., Dubrule, O. & Blunt, M.J., 2020. Stochastic seismic waveform inversion using generative adversarial networks as a geological prior, *Math. Geosci.,* **52**(1), 53–79.

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**(1), 345–377.

Nakayama, S. & Blacquière, G., 2021. Machine-learning-based data recovery and its contribution to seismic acquisition: Simultaneous application of deblending, trace reconstruction, and low-frequency extrapolation, *Geophysics,* **86**(2), P13–P24.

Ovcharenko, O., Kazei, V., Kalita, M., Peter, D. & Alkhalifah, T., 2019. Deep learning for low-frequency extrapolation from multioffset seismic data, *Geophysics,* **84**(6), R989–R1001.

Ovcharenko, O., Kazei, V., Alkhalifah, T.A. & Peter, D.B., 2022. Multi-task learning for low-frequency extrapolation and elastic model building from seismic data, *IEEE Trans. Geosci. Remote Sens.,* **60,** 1–17.

Plotnitskii, P., Kazei, V., Ovcharenko, O., Peter, D. & Alkhalifah, T., 2020. Extrapolation of low wavenumbers in FWI gradients by a deep convolutional neural network, in *Proceeding of the 82nd EAGE Annual Conference and Exhibition,* Vol. **2020,** pp. 1–5, European Association of Geoscientists & Engineers.

Pratt, R.G., Shin, C. & Hick, G., 1998. Gauss–newton and full newton methods in frequency–space seismic waveform inversion, *Geophys. J. Int.,* **133**(2), 341–362.

Richardson, A., 2018. *Seismic full-waveform inversion using deep learning tools and techniques,* preprint (arXiv:1801.07232).

Robins, T., Camacho, J., Agudo, O.C., Herraiz, J.L. & Guasch, L., 2021. Deep-learning-driven full-waveform inversion for ultrasound breast imaging, *Sensors,* **21**(13), 4570.

Shen, X., 2010. Near-surface velocity estimation by weighted early-arrival waveform inversion, in *SEG Technical Program Expanded Abstracts 2010,* pp. 1975–1979, Society of Exploration Geophysicists.

Shen, X., 2014. *Early-arrival waveform inversion for near-surface velocity estimation, PhD thesis*, Department of Geophysics, Stanford University, Stanford, CA.

Shin, C. & Cha, Y.H., 2008. Waveform inversion in the laplace domain, *Geophys. J. Int.,* **173**(3), 922–931.

Sun, H. & Demanet, L., 2018. Low frequency extrapolation with deep learning, in *SEG Technical Program Expanded Abstracts 2018,* pp. 2011–2015, Society of Exploration Geophysicists.

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

Sun, H. & Demanet, L., 2022a. Deep learning for low-frequency extrapolation of multicomponent data in elastic FWI, *IEEE Trans. Geosci. Remote Sens.,* **60,** doi:10.1109/TGRS.2021.3135790.

Sun, H. & Demanet, L., 2022b. Extrapolated surface-wave dispersion inversion, in *Second International Meeting for Applied Geoscience & Energy,* pp. 2060–2065, Society of Exploration Geophysicists.

Sun, Y., Araya-Polo, M. & Williamson, P., 2021. Data characterization and transfer learning for DL-driven velocity model building, in *First International Meeting for Applied Geoscience & Energy,* pp. 1475–1479, Society of Exploration Geophysicists.

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

Vantassel, J.P., Kumar, K. & Cox, B.R., 2022. Using convolutional neural networks to develop starting models for near-surface 2-d full waveform inversion, *Geophys. J. Int.,* **231**(1), 72–90.

Wang, M., Xu, S. & Zhou, H., 2020. Self-supervised learning for low frequency extension of seismic data, in *SEG Technical Program Expanded Abstracts 2020,* pp. 1501–1505, Society of Exploration Geophysicists.

Wang, Y.-Q., Wang, Q., Lu, W.-K., Ge, Q. & Yan, X.-F., 2022a. Seismic impedance inversion based on cycle-consistent generative adversarial network, *Petrol. Sci.,* **19**(1), 147–161.

Wang, Z., Liu, G., Du, J., Li, C. & Qi, J., 2022b. Low-frequency extrapolation of prestack viscoacoustic seismic data based on dense convolutional network, *IEEE Trans. Geosci. Remote Sens.,* **60,** 1–13.

Wu, Y. & Lin, Y., 2019. InversionNet: an efficient and accurate data-driven full waveform inversion, *IEEE Trans. Comput. Imag.,* **6,** 419–433.

Yang, F. & Ma, J., 2019. Deep-learning inversion: a next-generation seismic velocity model building method, *Geophysics,* **84**(4), R583–R599.

Yang, Y., Gao, A.F., Castellanos, J.C., Ross, Z.E., Azizzadenesheli, K. & Clayton, R.W., 2021. Seismic wave propagation and inversion with neural operators, *Seismic Record,* **1**(3), 126–134.

Yang, Y., Gao, A.F., Azizzadenesheli, K., Clayton, R.W. & Ross, Z.E., 2023. Rapid seismic waveform modeling and inversion with neural operators, *IEEE Trans. Geosci. Remote Sens.,* **61,** 1–12.

Yao, J. & Wang, Y., 2022. Building a full-waveform inversion starting model from wells with dynamic time warping and convolutional neural networks, *Geophysics,* **87**(2), R223–R230.

Zhang, H., Alkhalifah, T., Liu, Y., Birnie, C. & Di, X., 2022a. Improving the generalization of deep neural networks in seismic resolution enhancement, *IEEE Geosci. Remote Sens. Lett.,* **20,** doi:10.1109/LGRS.2022.3229167.

Zhang, H., Yang, P., Liu, Y., Luo, Y. & Xu, J., 2022b. Deep learning-based low-frequency extrapolation and impedance inversion of seismic data, *IEEE Geosci. Remote Sens. Lett.,* **19,** doi:10.1109/LGRS.2021.3123955.

Zhang, T., Sun, J., Trad, D. & Innanen, K., 2023. Multilayer perceptron and Bayesian neural network based elastic implicit full waveform inversion, *IEEE Trans. Geosci. Remote Sens.,* **61,** doi:10.1109/TGRS.2023.3265657.

Zhang, Z. & Lin, Y., 2020. Data-driven seismic waveform inversion: a study on the robustness and generalization, *IEEE Trans. Geosci. Remote Sens.,* **58**(10), 6900–6913.

Zhao, T., Abubakar, A., Cheng, X. & Fu, L., 2020. Augment time-domain FWI with iterative deep learning, in *SEG Technical Program Expanded Abstracts 2020,* pp. 850–854, Society of Exploration Geophysicists.

Zhong, Z., Sun, A.Y. & Wu, X., 2020. Inversion of time-lapse seismic reservoir monitoring data using CycleGAN: a deep learning-based approach for estimating dynamic reservoir property changes, *J. geophys. Res.,* **125**(3), e2019JB018408.

Zhu, J.-Y., Park, T., Isola, P. & Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks, in *Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV),* 22–29 October 2017, Venice, Italy, pp. 2223–2232.

Zhu, W., Xu, K., Darve, E., Biondi, B. & Beroza, G.C., 2022. Integrating deep neural networks with full-waveform inversion: reparametrization, regularization, and uncertainty quantification, *Geophysics,* **87**(1), R93–R109.

APPENDIX: LOSS FUNCTION FOR TRAINING

Within the CycleGAN (Zhu *et al.* 2017) framework, we use least squares loss function for the discriminators (LSGANs, Mao *et al.* 2017). Suppose 1 and 0 are the labels for real data and fake data, respectively. The objective functions for the forward mapping G_{HL}: $H \rightarrow L$ can be expressed as follows:

$$\mathcal{L}_{\text{LSGAN}}(G_{HL}, H) = \mathbb{E}_{h \backsim p(h)}[\|1 - D_L(G_{HL}(h))\|_2^2], \tag{A1}$$

$$\mathcal{L}_{\text{LSGAN}}(D_L, H, L) = \mathbb{E}_{h \backsim p(h)}[\|D_L(G_{HL}(h))\|_2^2] + \mathbb{E}_{l \backsim p(l)}[\|1 - D_L(l)\|_2^2], \tag{A2}$$

where $h \backsim p(h)$ and $l \backsim p(l)$ denote the data distribution. Then D_L can be obtained by:

$$D_L^* = \arg\min_{D_L} \mathcal{L}_{\text{LSGAN}}(D_L, H, L). \tag{A3}$$

We use a cycle-consistency loss $\mathcal{L}_{\text{cycle}}(G_{HL}, G_{LH}, H)$ and an identity loss $\mathcal{L}_{\text{identity}}(G_{HL}, L)$ to regularize G_{HL}. For generators, we use L_1 norm in the loss functions since L_1 norm can prevent blur better than L_2 norm:

$$\mathcal{L}_{\text{cycle}}(G_{HL}, G_{LH}, H) = \mathbb{E}_{h \backsim p(h)}[\|G_{LH}(G_{HL}(h)) - h\|_1], \tag{A4}$$

$$\mathcal{L}_{\text{identity}}(G_{HL}, L) = \mathbb{E}_{l \backsim p(l)}[\|G_{HL}(l) - l\|_1]. \tag{A5}$$

The full objective function for optimizing the generator G_{HL} is:

$$\mathcal{L}(G_{HL}, H, L) = \lambda_1 \mathcal{L}_{\text{LSGAN}}(G_{HL}, H) + \lambda_2 \mathcal{L}_{\text{cycle}}(G_{HL}, G_{LH}, H) + \lambda_3 \mathcal{L}_{\text{identity}}(G_{HL}, L), \tag{A6}$$

where λ_1, λ_2 and λ_3 are trade-off parameters controlling the relative importance of the three terms. Then G_{HL} can be obtained by:

$$G_{HL}^* = \arg\min_{G_{HL}} \mathcal{L}(G_{HL}, H, L). \tag{A7}$$

As for the backward mapping G_{LH}: $L \rightarrow H$, the associated discriminator D_H can be obtained by:

$$\mathcal{L}_{\text{LSGAN}}(D_H, H, L) = \mathbb{E}_{l \backsim p(l)}[\|D_H(G_{LH}(l))\|_2^2] + \mathbb{E}_{h \backsim p(h)}[\|1 - D_H(h)\|_2^2], \tag{A8}$$

$$D_H^* = \arg\min_{D_H} \mathcal{L}_{\text{LSGAN}}(D_H, H, L). \tag{A9}$$

Likewise, G_{LH} can be obtained by optimizing a full objective function $\mathcal{L}(G_{LH}, L, H)$ regularized by a cycle-consistency loss $\mathcal{L}_{\text{cycle}}(G_{LH}, G_{HL}, H)$ and an identity loss $\mathcal{L}_{\text{identity}}(G_{LH}, H)$:

$$\mathcal{L}_{\text{LSGAN}}(G_{LH}, L) = \mathbb{E}_{l \backsim p(l)}[\|1 - D_H(G_{LH}(l))\|_2^2], \tag{A10}$$

$$\mathcal{L}_{\text{cycle}}(G_{LH}, G_{HL}, L) = \mathbb{E}_{l \backsim p(l)}[\|G_{HL}(G_{LH}(l)) - l\|_1], \tag{A11}$$

$$\mathcal{L}_{\text{identity}}(G_{LH}, H) = \mathbb{E}_{h \backsim p(h)}[\|G_{LH}(h) - h\|_1], \tag{A12}$$

$$\mathcal{L}(G_{LH}, L, H) = \lambda_1 \mathcal{L}_{\text{LSGAN}}(G_{LH}, L) + \lambda_2 \mathcal{L}_{\text{cycle}}(G_{LH}, G_{HL}, L) + \lambda_3 \mathcal{L}_{\text{identity}}(G_{LH}, H), \tag{A13}$$

$$G_{LH}^* = \arg\min_{G_{LH}} \mathcal{L}(G_{LH}, L, H). \tag{A14}$$

During training, the four neural networks G_{HL}, D_L, G_{LH} and D_H are optimized simultaneously. We follow Zhu *et al.* (2017) and use $\lambda_1 = 1$, $\lambda_2 = 10$ and $\lambda_3 = 5$ in our experiments.