Deep Learning for Low Frequency Extrapolation of Multicomponent Data in Elastic FWI

Hongyu Sun, *Student Member, IEEE*, and Laurent Demanet

***Abstract*—Full waveform inversion (FWI) strongly depends on an accurate starting model to succeed. This is particularly true in the elastic regime: The cycle-skipping phenomenon is more severe in elastic FWI compared to acoustic FWI, due to the short S-wave wavelength. In this paper, we extend our work on extrapolated FWI (EFWI) by proposing to synthesize the low frequencies of multi-component elastic seismic records, and use those "artificial" low frequencies to seed the frequency sweep of elastic FWI. Our solution involves deep learning: we can either train the same convolutional neural network (CNN) on two training datasets, one with vertical components and one with horizontal components of particle velocities, or train with two components together, to extrapolate the low frequencies of elastic data for 2D elastic FWI. The architecture of this CNN is designed with a large receptive field by dilated convolution. Numerical examples on the Marmousi2 model show that the 2-4 Hz low frequency data extrapolated from band-limited data above 4 Hz provide good starting models for elastic FWI of P-wave and S-wave velocities. Additionally, we study the generalization ability of the proposed neural network from acoustic to elastic data. For elastic test data, collecting the training dataset by elastic simulation shows better extrapolation accuracy than acoustic simulation, i.e., a smaller generalization gap.**

***Index Terms*—Neural networks, Waveform inversion, Computational seismology, Controlled source seismology, Numerical solutions.**

I. Introduction

FULL waveform inversion is well-known for its great potential to provide quantitative Earth properties of complex subsurface structures. Acoustic FWI is widely used and has been successfully applied to real seismic data. However, most seismic data have strong elastic effects [1]. The acoustic approximation is insufficient to estimate correct reflections and introduces additional artifacts to FWI results [2], [3]. Therefore, it is desirable to develop a robust elastic FWI method for high-resolution Earth model building.

Foundational work has shown the ability of elastic FWI to retrieve realistic properties of the subsurface [4], [5]. However, it has difficulty handling real data sets. Elastic FWI is very sensitive to: accuracy of the starting model; correct estimation of density; proper definition of multi-parameter classes; and noise level [6]. The complex wave phenomena in elastic wavefields bring new challenges to FWI.

Among the many factors that affect the success of elastic FWI, the lowest starting frequency is an essential one, given that an accurate starting model is generally unavailable. Compared to acoustic FWI, the nonlinearity of elastic FWI is more severe due to the short S-wave propagating wavelength. Therefore, elastic FWI always requires a lower starting frequency compared to acoustic FWI. Additionally, the parameter cross-talk problem exists in elastic FWI and becomes more pronounced at higher frequencies, so ultra-low frequencies are required for a successful inversion of S-wave velocity and density.

In synthetic studies of elastic FWI, [7] invert the overthrust model [8] from 1.7 Hz. [6] invert the Valhall model [9] from 2 Hz. Both inversion workflows start from Gaussian smoothing of true models. Moreover, [10] invert the Marmousi2 model [11] using a velocity-gradient starting model but a very low frequency (0.16 Hz). For a successful inversion of the Marmousi2 density model, [12] use 0-2 Hz in the first stage of multi-scale FWI. [13] invert the same model from 0.2 Hz.

Few applications of elastic FWI to real data sets are reported in the literature [1], [14], [15]. [16] use 3.5 Hz as the starting frequency of elastic FWI given that the initial models are accurate enough. [17] apply 3D elastic FWI to update P-wave velocity and obtain S-wave velocity and density using empirical relationships. [18] perform elastic FWI involving surface waves in the band of 5-15 Hz for a land data set.

New developments in acquisition enhance the recent success of FWI by measuring data with lower frequencies and longer offsets [19], [20]. However, only acoustic FWI was applied to the land data set with low frequencies down to 1.5 Hz [21]. In addition to the expensive acquisition cost for the low-frequency signals, direct use of the field low-frequency data requires dedicated pre-processing steps, including travel-time tomography, for an accurate enough model to initialize FWI. The final inversion results strongly rely on the starting tomography model. Hence, attempting to retrieve reliable low-frequency data offers a sensible pathway to relieve the dependency of elastic FWI on starting models.

Deep learning is an emerging technology in many aspects of exploration geophysics. In seismic inversion, several groups have experimented with directly mapping data to model using deep learning [22]–[26]. Within Bayesian seismic inversion framework, deep learning has been applied for formulating priors [27]–[29]. Other groups use deep learning as a signal processing step to acquire reasonable data for inversion. For instance, [30] use deep learning to remove elastic artifacts for acoustic FWI. [31] remove the numerical dispersion of wavefields by transfer learning.

Computationally extrapolating the missing low frequencies from band-limited data is the cheapest way for FWI to mitigate

The authors thank TotalEnergies for support. Hongyu Sun acknowledges SEG scholarships and MIT MathWorks Science Fellowship for funding. *(Corresponding author: Hongyu Sun.)*

The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: hongyus@mit.edu; laurent@math.mit.edu).

the cycle-skipping problem. [32], [33] seperate the shot gather to atomic events and then change the wavelet to extrapolate the low frequencies. [34] extend the frequency spectrum based on the redundancy of extended forward modeling. [35]–[37] have utilized CNN to extrapolate the missing low frequencies from band-limited data. They have proposed different architectures of CNN to learn the mapping between high and low frequency data from different features in the training datasets. However, only acoustic data are considered in these studies. Recently, [38] extrapolated the low frequencies for prestack land data. Meanwhile, our preliminary work [39] shows the feasibility to extrapolate reliable low frequencies for elastic FWI using the deep learning model in [40].

Although the mechanism of deep learning is hard to explain, the feasibility of low frequency extrapolation has been discussed in terms of sparsity inversion [41] and wavenumber illumination [42]. With multiple-trace extrapolation, the low wavenumbers of far-offset data have been proposed as the features in the frequency domain detected by CNN to extrapolate the missing low frequencies [42]. In contrast, for trace-by-trace extrapolation [40], the features to learn are the structured time series themselves. The feasibility of trace-by-trace frequency extrapolation has been mathematically proved in simple settings in [43], [44], as a by-product of super-resolution.

In this paper, we extend our workflow of extrapolated FWI with deep learning [40] into the elastic regime. We train the neural network with elastic training datasets to predict the low-frequency data of the horizontal components (v_x) and the vertical components (v_z). The extrapolated low frequency data are used to initialize elastic FWI from a crude starting model. For the architecture design of CNN, a large receptive field is efficiently achieved by dilation convolution. Moreover, to investigate the generalization ability of neural networks from acoustic to elastic data, we compare the extrapolation results of the neural networks trained on elastic data and acoustic data to predict the elastic low-frequency data. The elastic test data are simulated on the modified Marmousi2 model (referred to as Marmousi2) in a deep-water marine environment, so surface waves are ignored in this work.

II. Method

We first give a brief review of elastic FWI as implemented in this paper. Then we illustrate the feasibility of low frequency extrapolation, and design two deep learning models for this purpose. Afterwards, the training and test datasets are provided to train and verify the performance of the proposed neural networks.

A. Review of elastic FWI

Elastic FWI is implemented in the time domain to invert the P-wave velocities ($\mathbf{v}_p$), S-wave velocities ($\mathbf{v}_s$) and density (ρ) simultaneously. The object function E is formulated as

$$E = \frac{1}{2}\delta\mathbf{d}^T\delta\mathbf{d} = \frac{1}{2}\sum_s\sum_r\int[\mathbf{u}_{cal}-\mathbf{u}_{obs}]^2dt, \qquad (1)$$

where $\mathbf{d}$ are the residuals between observed wavefields $\mathbf{u}_{obs}$ and calculated wavefields $\mathbf{u}_{cal}$. In 2D, both $\mathbf{u}_{obs}$ and $\mathbf{u}_{cal}$ contain the v_x and v_y components of elastic wavefields. The gradient $\frac{\delta E}{\delta \mathbf{m}}$ relative to the model parameters $\mathbf{m}$ is calculated in terms of $\mathbf{v}_p$, $\mathbf{v}_s$ and ρ using the velocity-stress formulation of the elastic wave equation [12]. The starting models $\mathbf{m_0}$ are updated using the L-BFGS method [45].

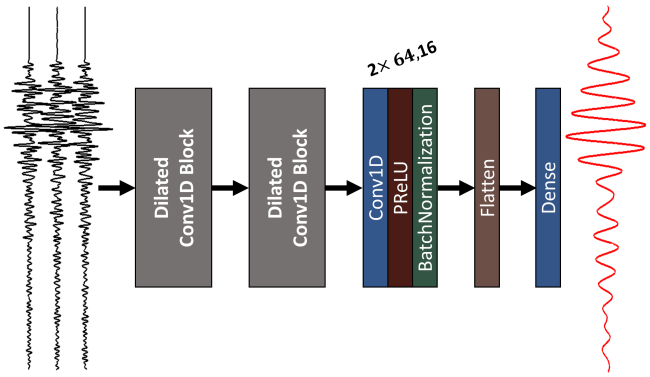
B. Deep learning models for low-frequency extrapolation

We choose CNN to perform the task of low-frequency extrapolation. The output of the neural network is a one-dimensional time series in the low frequency band. The input of the neural network involves bandlimited traces recorded by several adjacent receivers where the middle trace corresponds to the high frequency part of the output. Low-frequency extrapolation with the multi-trace setup is expected to benefit from the coherence among different traces [40]. However, the moderate improvement of the performance with multiple traces also increases the training effort, so we choose to use three traces as the input in our experiments. In 2D, the elastic data contain horizontal and vertical components. As a result, we can choose to separately train two neural networks with the same architecture on two different training datasets: one contains v_x and the other contains v_z. Alternatively, we can feed into one neural network with both components and train the neural network to simultaneously process two components.

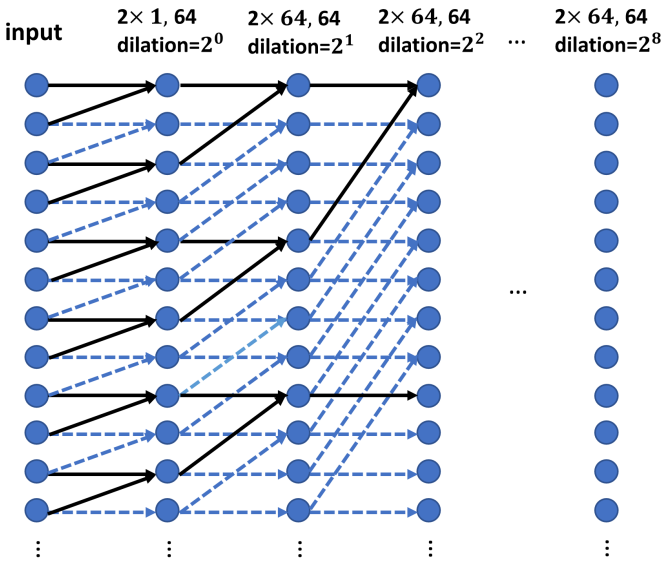
We emprically find that the proposed CNN model requires a large receptive field for low-frequency extrapolation. A very large receptive field enables each feature in the final output to include a large range of input pixels. Since any single frequency component is related to the entire waveform in the time domain, extrapolation from one frequency band to the other requires a large receptive field to cover the entire input signal. Typically, the receptive field is increased by stacking layers. For example, stacking two convolutional layers (without pooling) with 3×3 filter results in a layer of 5×5 filter. However, it requires many layers to result in a large enough receptive field and is computationally inefficient. In our previous work, a large receptive field is achieved by directly using a large filter on each convolutional layer. Instead, here we propose to use a small filter with dilated convolution.

The designed CNN architecture uses dilated convolution to increase the receptive field by orders of magnitude. A dilated convolution [46] (convolution with holes) is a convolution where the filter is applied over an area larger than its length by skipping input values with a certain step (dilation). It effectively allows the network to operate on a coarser scale than with a normal convolution. This is similar to pooling or stride, but here the output has the same size as the input. As a special case, dilated convolution with dilation of one yields the standard convolution. Stacked dilated convolutions enable networks to have very large receptive fields with just a few layers. In addition to save computational cost, this method helps to preserve the input resolution throughout the network [46]. Moreover, we use causal convolution to process time series [47], although this choice does not appear to be essential in our case.

The architecture has two dilated convolutional blocks (Fig. 1). Each block consists of nine 1D convolutional layers.



(a)



(b)

Fig. 1. The deep learning model designed with dilated convolution for low-frequency extrapolation. Three traces are input together to extrapolate the low frequencies of the middle trace. (a) The CNN model. The filter length is 2 on each convolutional layer but two dilated convolutional blocks are stacked to increase the receptive field exponentially with depth. (b) The dilated causal convolutional block. Each block has 9 convolutional layers. The filter size, number of channel and dilation of each convolutional layer are labeled on the top. The number of convolutional layers and dilation may be increased according to the size of the input time series. Here the receptive field of the architecture is 1024, which can cover the entire input of 1000 after downsampling the signal with a factor of 3.

Each convolutional layer is followed by a PReLU layer and a batch normalization layer. On each convolutional layer, there are 64 causal convolutional filters with a length of two (Fig. 1b). The dilations of the nine convolutional layers are 2^0, 2^1, ..., 2^8, respectively. The exponential increase in dilation results in exponential growth, with depth, of the receptive field [48].

Without pooling layer, the size of the receptive field RF_{l+1} on the $l+1$ layer is

$$RF_{l+1} = RF_l + (k_{l+1} - 1) \times s_{l+1} \times d_{l+1}, \quad l = 0, ..., n \quad , \tag{2}$$

$$RF_0 = 1, \tag{3}$$

where k_{l+1} is the kernel size of the $l+1$ convolutional layer. s_{l+1} is the stride size. d_{l+1} is the dilation on the $l+1$ layer if the layer contains a dilated convolution. Otherwise, d_{l+1} equals one for regular convolutional layers.

TABLE I
COMPARISON BETWEEN TWO DEEP LEARNING ACHITECTURES FOR LOW-FREQUENCY EXTRAPOLATION

deep learning architecture	receptive field	trainable parameters
ARCH1 (this work)	1,024	17,314,200
ARCH2 ([40])[a]	496	35,119,048

[a] We use a kernel length of 100 for all the filters in this model.

Table I compares the receptive field and trainable parameters of the deep learning model used in this work (called ARCH1) and that in [40] (ARCH2, with a kernel length of 100) for low-frequency extrapolation. Although both neural networks are able to perform low frequency extrapolation, the new achitecture has a much larger receptive field with fewer trainable parameters, thanks to a smaller kernel size. Thus, convolution with dilation is more efficient than directly using a large convolutional kernel in terms of low frequency extrapolation.

C. Training and test datasets

The training and test datasets are simulated on the elastic training and test models. The Marmousi2 elastic model (Fig. 2) is referred to as the test model in deep learning. This is also the true model in the subsequent elastic FWI. We randomly extract seven batches from the 3D Overthrust benchmark model [49] and use six batches as the training models and one as the validation model. The size of each model is 500×200 with a grid spacing of 20 m, including a water layer on the top of each model with a depth of 420 m. Fig. 3 shows the P-wave velocities of these models. In the elastic regime, each model consists of three parameters: $\mathbf{v}_p$, $\mathbf{v}_s$ and ρ. The S-wave velocities and densities of the models are calculated based on P-wave velocities using empirical relations between elastic wavespeeds and densities in the Earth's crust [50], [51]. The very different structures between the training and test models enable us to check the generalization ability for the neural network to extrapolate low frequencies on the Marmousi2 model.

Both training and test datasets are simulated using a 2D time domain stress-velocity PSV finite-difference (FD) code [52], [53] with an eighth-order spatial FD operator. A Ricker wavelet with a dominant frequency of 10 Hz is used as the source signal. The sampling rate and the recording time is 0.002 s and 6 s, respectively. To collect the test dataset, 50 shots are excited evenly from 800 m to 8640 m in the water layer at the same depth of 40 m. 400 receivers are placed from 800 m to 8780 m under the water layer with a depth of 460 m to record v_x and v_z of the elastic wavefields. It is not necessary to collect the training and test datasets using the same acquisition geometry. For the training model, there are 100 shots evenly spaced from 500 m to 8420 m with a depth of 40 m on each training model. 400 receivers are placed from 480 m to 8460 m. Additionally, a free surface

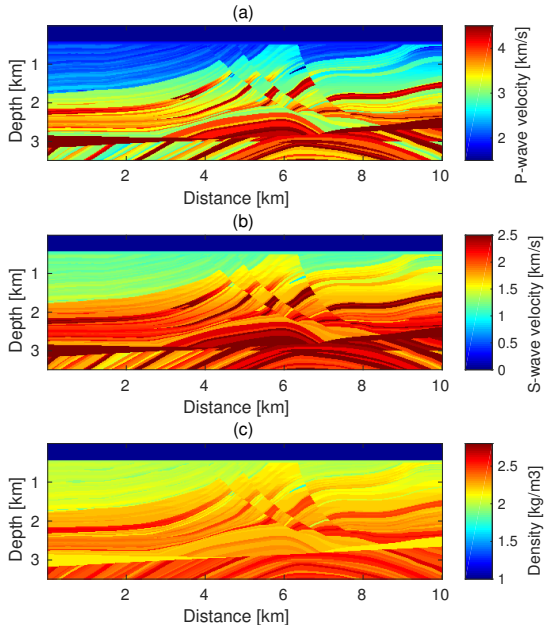


Fig. 2. The Marmousi2 model: (a)$\mathbf{v}_p$, (b)$\mathbf{v}_s$ and (c)ρ. The Marmousi2 model is the true model for elastic FWI and the test model for synthesizing the test dataset for deep learning.

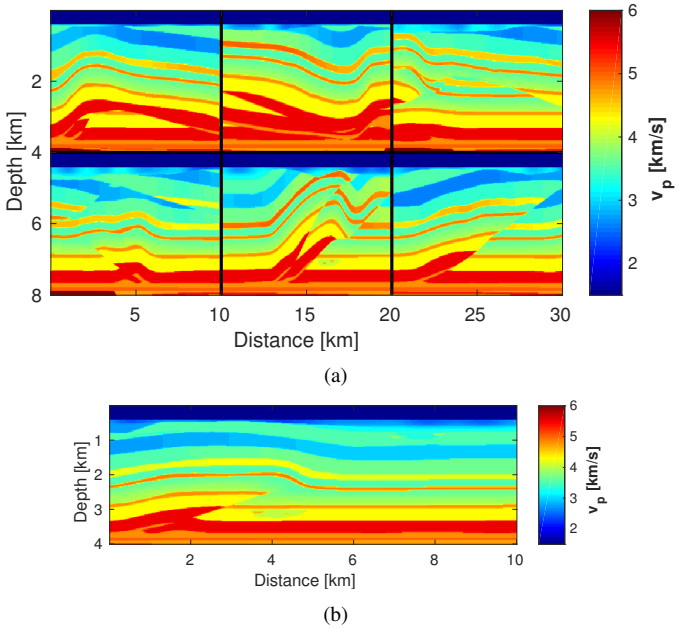


Fig. 3. The P-wave velocities of the 2D training and validation models extracted from 3D Overthrust benchmark model. (a) The six training models. (b) The validation model. The size of each model is equally 500×200 with a grid spacing of 20 m including a 420 m depth water layer. Each model contains three parameters: $\mathbf{v}_p$, $\mathbf{v}_s$ and ρ.

boundary condition is applied to the top of the model to simulate the realistic marine exploration environment. The free surface condition damages the low frequency data and thus the energy in the low frequency band is very low in the simulated full-band data. This brings a new challenge to the low frequency extrapolation.

After the forward modeling, two training datasets are collected, one with a dataset of horizontal components and one with a dataset of vertical components. The 2D elastic data on the test model is also separated into two test datasets to process each component individually. With a three-to-one extrapolation setup, there are $6\times100\times398 = 238,800$ training samples in each training dataset and $1 \times 50 \times 398 = 19,900$ test samples in each test dataset.

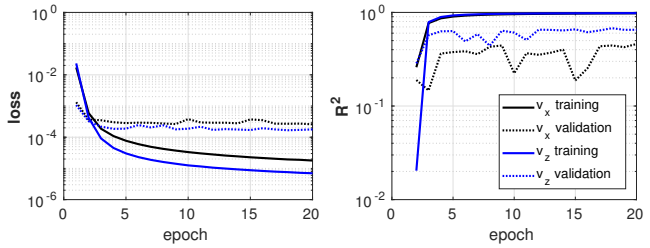
A simple preprocessing step can be used to improve the deep learning performance. Each sample in the training and test datasets is normalized by dividing the raw signal by its maximum. Then all the data are scaled with a constant (for instance, 100) to stabilize the training process. The values used to normalize and scale the raw data are recorded to recover the original observed data for elastic FWI. After this process, each sample in the training and test dataset is separated into a low-frequency signal and a high-frequency signal using a smooth window in the frequency domain. Then, each time series in the high-frequency band is fed into the neural network to predict the low-frequency time series. We also downsample the time series with a factor of three to save the computational cost after spliting the full-band data into low and high frequency parts. Thus, the actual input size of the neural network is 1000×3. After extrapolation, the original low frequency signal is recovered using bandlimited interpolation.

III. Numerical Examples

The numerical examples section is divided into four parts. In the first part, we train the neural network (Fig. 1) to extrapolate the low-frequency data of bandlimited multi-component recordings simulated on the Marmousi2 model (Fig. 2). Then, we use the extrapolated low-frequencies of multi-component band-limited data to seed the frequency sweep of elastic FWI on the Marmousi2 model. In the third part, we study the generalization ability of the proposed neural network from acoustic to elastic data. In the last part, we compare the performance of the architecture in this paper with that in [40] for low-frequency extrapolation.

A. Low frequency extrapolation of multicomponent data

We first extrapolate the low frequency data below 4 Hz on the Marmousi2 model (Fig. 2) using $4-30$ Hz band-limited data. Each sample in the training and test datasets is separated into a $0.125-4$ Hz low-frequency signal and a $4-30$ Hz high-frequency signal using a smooth window in the frequency domain. The time series in the high-frequency band is directly fed into the neural network to predict the $0.125-4$ Hz low-frequency time series. To deal with the multicomponent data, the neural network is trained twice: once on the training dataset of v_x and once on the training dataset of v_z. Both training processes use the ADAM method with a



(a) (b)

Fig. 4. (a) The learning curves and (b) evaluation metrics (R^2) of the neural network trained to extrapolate the $0.125-4$ Hz low frequencies of v_x and v_z from the $4-30$ Hz band-limited elastic recordings on the training and validation sets.

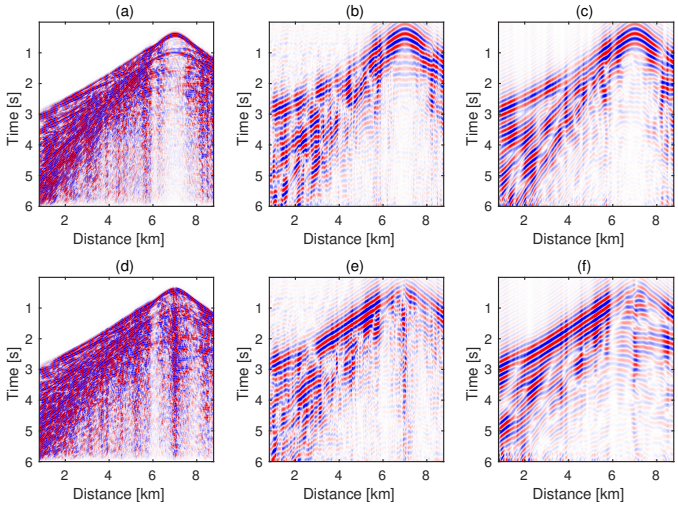


Fig. 5. Extrapolation results on the Marmousi2 model (test model) for the neural network trained on the Overthrust model: comparison among the (a) band-limited recordings ($4-30$ Hz), (b) predicted and (c) true low-frequency recordings ($0.125-4$ Hz) of v_z and (d) band-limited recordings ($4-30$ Hz), (e) predicted and (f) true low-frequency recordings ($0.125-4$ Hz) of v_x.

mini-batch of 32 samples. We refer readers to [40] for more details about training. Fig. 4(a) and 4(b) show the learning curves and evaluation metrics (R^2) over 20 epochs to predict the low frequencies of v_x and v_z on the training and validation sets. The curves of training loss decay over epochs on both the training and validation datasets, which indicate that the neural network does not overfit.

Fig. 5 shows the extrapolation results of both v_x and v_z where the source is located at 7.04 km. Most predicted low frequency events are comparable with the true low frequencies. Especially, the diving waves are well extrapolated. Fig. 6 compares the amplitude specturm of data with extrapolated low frequencies, true low frequencies, and their difference, for the shot in Fig. 5. Despite minor prediction errors, the neural network can successfully recover the low frequencies of v_x and v_z recordings with satisfactory accuracy.

B. Extrapolated elastic full waveform inversion

We perform extrapolated elastic FWI using $4-20$ Hz band-limited data on the Marmousi2 model. The lower band of the band-limited data is 4 Hz. To quantitatively evaluate inverted models, a metric, relative model error (RME), is calculated

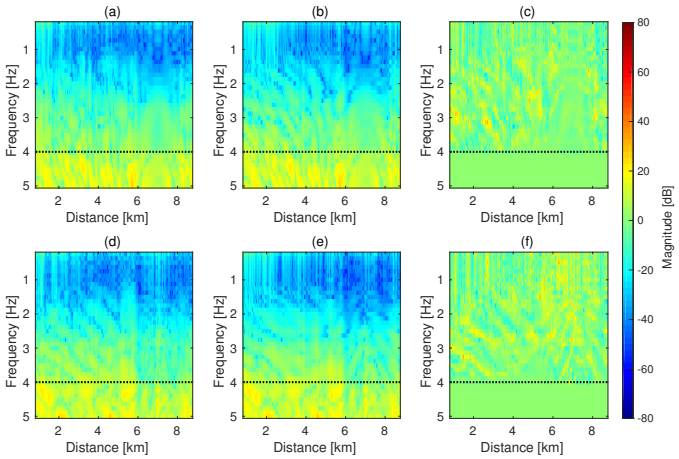


Fig. 6. Extrapolation results on the Marmousi2 model (test model) for the neural network trained on the Overthrust model: comparison of the amplitude spectrum of v_z (top row) and v_x (bottom row). (a) and (d) Data with $0.125-4$ Hz predicted low frequencies. (b) and (e) Data with true low frequencies. (c) and (f) The difference between data with predicted and true low frequencies. On each panel, the data above the dash line are low frequencies. The data below the dash line are part of the bandlimited data. The lowest frequency reported on the vertical axis of each figure is 0.125 Hz, but we estimate that frequencies below 0.5 Hz are significantly affected by windowing artifacts coming from the numerical Fourier transform, and should not be assigned a physical meaning.

using $\|(\mathbf{m}_{inv}-\mathbf{m}_{true})/\mathbf{m}_{true}\|_2$, where $\mathbf{m}_{inv}$ and $\mathbf{m}_{true}$ denote inverted and true models. The water layer is excluded in this calculation. Fig. 7(a), (d) and (g) show the initial models of $\mathbf{v}_p$, $\mathbf{v}_s$ and ρ, respectively. Starting from the crude initial models, Fig. 7(b) and (c) show the resulting P-wave velocity models after 30 iterations using extrapolated and true $2-4$ Hz low-frequency data, respectively. The inverted S-wave velocity models using extrapolated and true low frequencies are shown in Fig. 7(e) and (f). Also, Fig. 7(h) and (i) compare the inverted density models using the extrapolated and true $2-4$ Hz low frequencies. The RME values of these models are labeled on the top of each model. Note that the starting RME of density is smaller than velocity, due to a smaller range of values on the density model compared with velocity. The inverted low-wavenumber models of $\mathbf{v}_p$, $\mathbf{v}_s$ and ρ using extrapolated data are roughly the same as those using true data. However, the inversion of density model is not successful since $2-4$ Hz data are relatively high frequencies for the inversion of density model. Similar to the observation in [51], we have an increased RME of the density model after elastic FWI, due to the overestimation of the density.

Then the inversion is continued with the $4-20$ Hz band-limited data. We utilize a multiscale method [54] and sequentially explore the $4-6$ Hz, $4-10$ Hz and $4-20$ Hz band-limited data in the elastic FWI. In each frequency band, the number of iterations is 30, 30 and 20, respectively. Fig. 8 (a)-(f) show the resulting $\mathbf{v}_p$ and $\mathbf{v}_s$ models started from different low wavenumber models. The inversion results of $\mathbf{v}_p$ and $\mathbf{v}_s$ started from $2-4$ Hz extrapolated data are very close to the results started from $2-4$ Hz true data. Conversely, elastic FWI directly starting from the crude initial models using the band-limited data shows large errors.

Fig. 8(g)-(i) shows the resulting density models using the $4-20$ Hz band-limited data but started from different models in Fig. 7(g)-(i). Since the starting frequency band ($2-4$ Hz) is relatively high for the inversion of density on the crude initial model (Fig. 7(g)), the inverted models using $2-4$ Hz data (Fig. 7(h) and (i)) only show high-wavenumber structure of the density model. With band-limited data involved in the inversion, the inverted density models (see Fig. 8(h) and (i)) resemble migration results but show the density perturbation [5]. For a successful inversion of the density model, a much lower starting frequency band is required to recover the low-wavenumber structures.

C. Generalization ability of the neural network

In this work, we study the generalization ability of the proposed neural network for low-frequency extrapolation in three situations: different subsurface structures, acoustic to elastic data and different elastic simulation codes. First, the previous example has shown that the neural network can generalize from the Overthrust model to the Marmousi2 model. The extrapolated low frequencies predicted by the neural network trained on the Overthrust model can be used to initialize elastic FWI on the Marmousi2 model.

Then, we study the generalizability of the proposed neural network from acoustic to elastic data. Here the acoustic data are simulated using the same elastic solver (DENISE [55]) by setting the S-wave velocities as zeros. Instead of training the neural network twice with two components in the previous example, here we train the neural network once and simultaneously predict the low frequencies of both v_x and v_z in the same elastic test data set. Fig. 9 compares the $0.125-4$ Hz low frequencies of the shot located at 7.04 km extrapolated by the neural network trained using acoustic v_z, only elastic v_z, only elastic v_x and both elastic v_x and v_z. Training with either acoustic data or single component of the elastic data degrades the accuracy of the extrapolated low frequencies, compared with training using both elastic v_x and v_z components. Additionally, we use the structural similarity image metric (SSIM, [56]) to quantitatively measure the similarity between the extrapolated and true low frequency data. Table II and table III respectively compare the SSIM index for the shot in Fig. 9 and all shots in the test data set. Training with both elastic v_x and v_z achieved the best similarty for the multicomponent data, thanks to a smaller generalization gap between the elastic data compared with the gap from acoustic to elastic data.

Then, we study the generalization ability of the proposed neural network among different forward modeling codes. We use 'fdelmodc' [57] to simulate the acoustic or elastic training data sets, whereas the elastic test data set is simulated using 'DENISE' [12]. The acoustic training data set is simulated using the 2D variable density acoustic wave equation on the P-wave velocity models (see Fig. 3) and associated density models used in elastic simulation. As can be seen in Fig. 10, training with the acoustic v_z or elastic v_z cannot retrieve the P-S polarity reversal effect on elastic v_x. Training with elastic v_x cannot recover the continuity of events at near zero-offset.

TABLE II
COMPARISON OF THE SSIM INDEX BETWEEN EXTRAPOLATED AND TRUE LOW FREQUENCY DATA FOR THE SHOT IN FIG 9

component	acoustic v_z	elastic v_z	elastic v_x	elastic v_x and v_z
elastic v_x	0.6379	0.6080	0.6074	0.6565
elastic v_z	0.6328	0.6248	0.5886	0.6635

TABLE III
COMPARISON OF THE SSIM INDEX BETWEEN EXTRAPOLATED AND TRUE LOW FREQUENCY DATA FOR ALL SHOTS IN THE TEST DATA SET

component	acoustic v_z	elastic v_z	elastic v_x	elastic v_x and v_z
elastic v_x	0.6390	0.6145	0.6275	0.6595
elastic v_z	0.6388	0.6089	0.5762	0.6493

Instead, training with both elastic v_z and v_x enables the neural network to retrieve the P-S polarity reversal effect on elastic v_x and continuity of events at near zero-offset on elastic v_z. The numerical results demonstrate that the generalization gap from acoustic to elasic data becomes larger when the training and test data sets are simulated using different solvers. The extrapolation accuracy is much poorer if the elastic training data set is simulated using a different code, which is even harder than the generalization over different orders of the same solver [40]. We speculate that the poorer performance may be attributed to different shape of the spectra simulated using 'fdelmodc' and 'DENISE'.

Finally, we compare the performance of ARCH1 and ARCH2 using the same training dataset and hyperparameters. Each network is trained twice with the training dataset of v_x and the training dataset of v_z. According to the learning curves, the training of ARCH1 is much more stable than that of ARCH2. ARCH1 also requires less training time, due to the less trainable parameters compared with ARCH2. Fig. 11 shows the extrapolated elastic FWI results started from the 2-4 Hz extrapolated low frequency data using ARCH2. According to the comparison of the inverted models started from the extrapolated data using ARCH1 and ARCH2, both neural networks are able to provide sufficient accuracy for the inversion of $\mathbf{v}_p$ and $\mathbf{v}_s$.

IV. DISCUSSION AND LIMITATIONS

A. Physical interpretation

Is the problem at all solvable? We claim that neural networks might be able to extend the bandwidth of a survey, but it is important to understand why the problem might even be solvable in the first place. We believe the setting for bandwidth extension is favorable when 1) the waves are impulsive, and 2) the medium in which the waves propagate is mostly nondispersive. In that case, changing the frequency content of a wave is a matter of identifying isolated waves, and changing their time signature while keeping the arrival time unchanged. This setting is quite similar to changing the pitch of speech – phonemes should be spoken in a different voice, but at the same time as the original recording.

Can we expect the waves to be nondispersive? If the characteristic length scale of the medium in which the waves travel is either much larger than the wavelength (geometrical

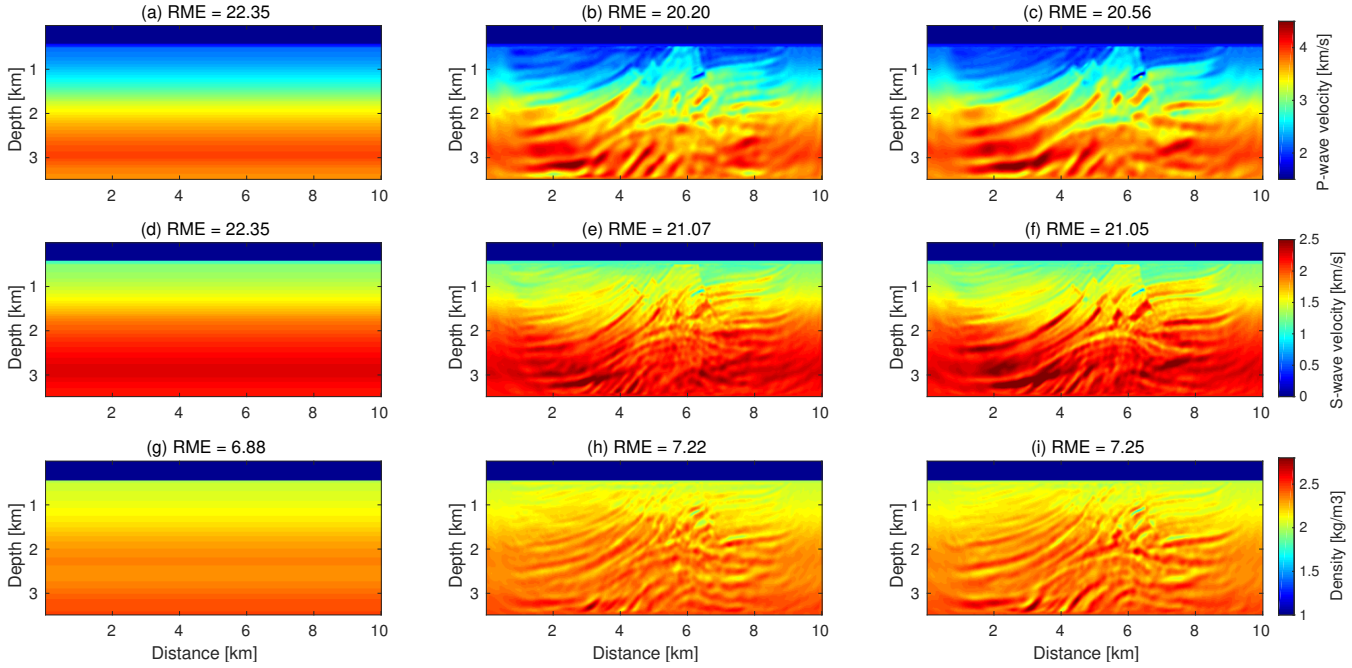


Fig. 7. Comparison among (a) the initial $\mathbf{v}_p$ model, the inverted low-wavenumber P-wave velocity models using (b) $2.0-4.0$ Hz extrapolated data and (c) $2.0-4.0$ Hz true data, (d) the initial $\mathbf{v}_s$ model, the inverted low-wavenumber S-wave velocity models using (e) $2.0-4.0$ Hz extrapolated data and (f) $2.0-4.0$ Hz true data, (g) the initial ρ model, the inverted low-wavenumber density models using (h) $2.0-4.0$ Hz extrapolated data and (i) $2.0-4.0$ Hz true data.

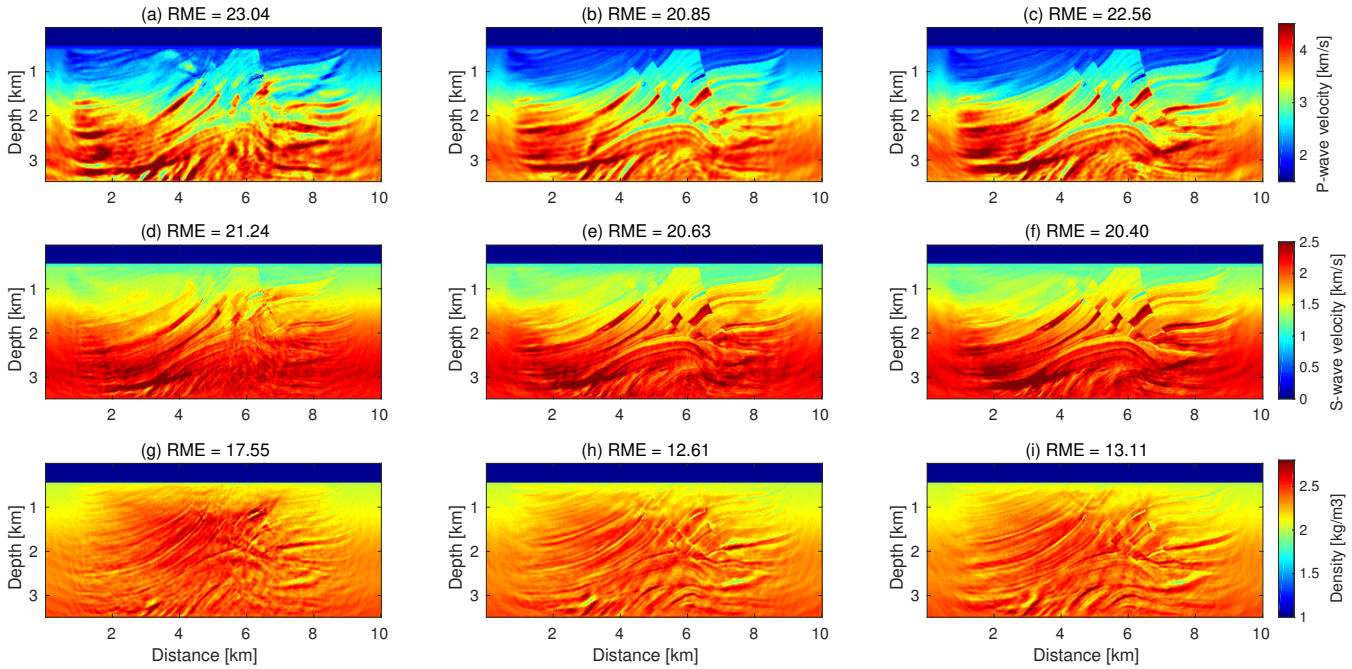


Fig. 8. Comparison of the inverted (a)-(c) $\mathbf{v}_p$, (d)-(f) $\mathbf{v}_s$, and (g)-(i) ρ models from elastic FWI using $4-20$ Hz band-limited data. (a), (d), and (g) The resulting models start from the original initial models. (b), (e), and (h) The resulting models start from the inverted low-wavenumber models using $2.0-4.0$ Hz extrapolated data. (c), (f), and (i) The resulting models start from the inverted low-wavenumber models using $2.0-4.0$ Hz true data.

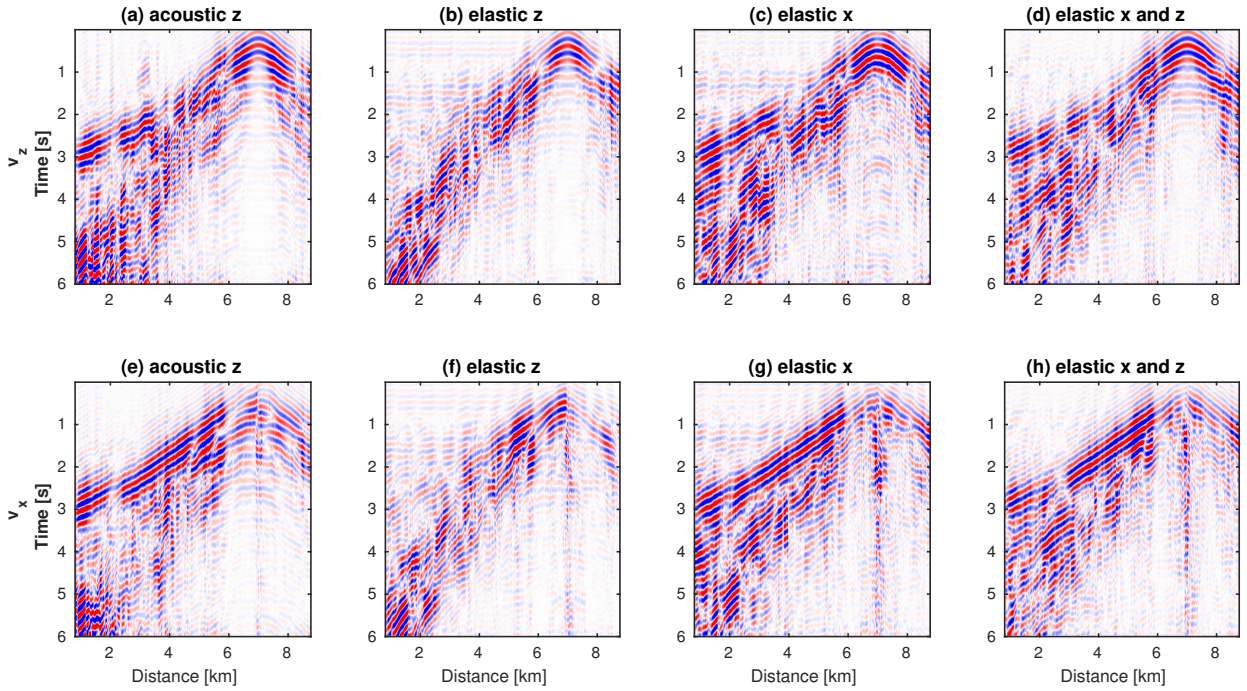


Fig. 9. Extrapolation results of the shot located at 7.04 km on the Marmousi2 model (test model) for the neural network trained using (a) and (e) acoustic v_z, (b) and (f) only elastic v_z, (c) and (g) only elastic v_x, (d) and (h) both elastic v_x and v_z. The top row shows the v_z component of the $0.125 - 4$ Hz extrapolated low frequencies. The bottom row shows the v_x component.

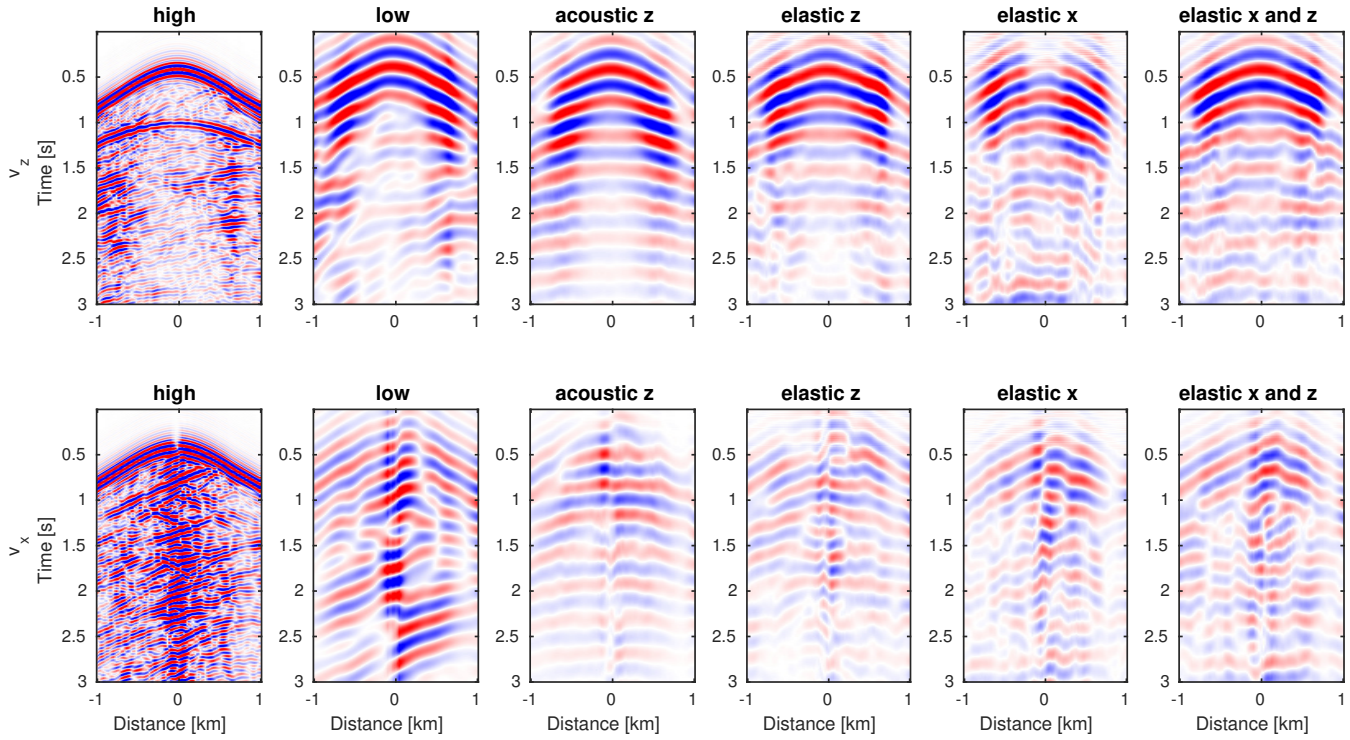


Fig. 10. Generalization over different simulation codes: comparison of the extrapolation results for the neural network trained on different training data sets. The top row shows the vertical component of the elastic data. The bottom row shows the horizontal component. The first (leftmost) and second columns show the true high and low frequency data. The third to the sixth columns show the prediction of the neural network trained with the acoustic data (v_z), only elastic v_z, only elastic v_x, and both elastic v_z and elastic v_x. Training with the acoustic v_z or elastic v_z cannot retrieve the P-S polarity reversal effect on elastic v_x. Training with elastic v_x cannot recover the continuity of events at near zero-offset. Instead, training with both elastic v_z and v_x enables the neural network to retrieve the P-S polarity reversal effect on elastic v_x and continuity of events at near zero-offset on elastic v_z.

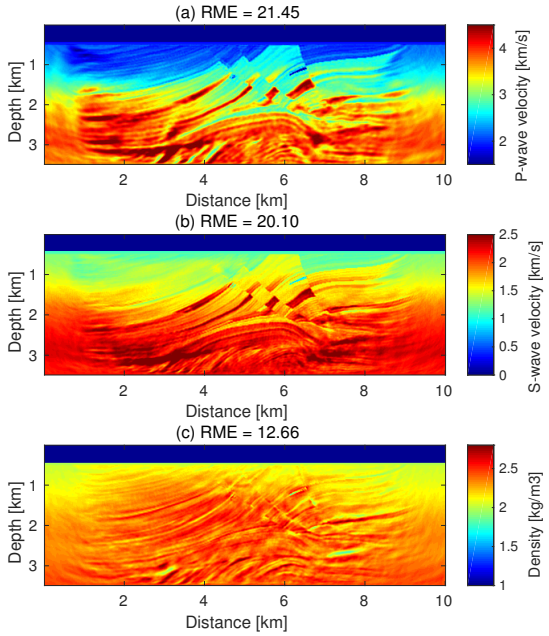


Fig. 11. Comparison of the inverted (a) $\mathbf{v}_p$, (b) $\mathbf{v}_s$ and (c) ρ models from extrapolated elastic FWI. The resulting models start from the inverted low-wavenumber models using $2.0-4.0$ Hz data extrapolated using ARCH2.

optics limit), or much smaller than the wavelength (homogenization), we can indeed expect that the wave speed is mostly independent of the frequency content of the wave.

Is the problem necessarily hard? Frequency extrapolation is in general ill-posed, but as we mention above, there are (ideal) settings where it is a well-posed problem. Changing the frequency content of an impulsive wave from high to low will in general create interference, constructive or destructive, because the footprint of the wave is now larger. This is fine. But changing the frequency content from low to high frequencies might in general be harder, since it would require undoing such interference. In this paper, we do not attempt to generate unobserved high frequencies.

Does a neural network know the right physics to compute low frequencies? Neural networks are built as universal approximators, and we observe in this paper that they do not need to be "enhanced" with physical principles to work – only the training set needs to be representative of the physics. It is unclear to us what operation the network performs to be successful, but it is an important research question going forward.

Might a neural network even have the capability to perform frequency extrapolation? Besides being a universal approximator, yes, there are simple examples to support the idea that a simple neural network can peform the required nonlinear signal processing for frequency extrapolation. A crude example is the following sequence of operations, which can easily be implemented by a shallow network with ReLU activators:

- Take a high-frequency waveform, remove its negative part. This operation creates low frequencies all the way to $\omega = 0$;
- Convolve the result by an adequate filter that substitutes the low-frequency wavelet for the positive part of the high-frequency wavelet. No division by zero should be required in the frequency domain to obtain such a filter.

While these two operations are far from providing an accurate predictor, they are indicative of the kind of operation that one might require of the neural network. It is also worth pointing out that there are reasonable solutions to the question of frequency extrapolation that do not involve neural networks, such as the work of one of us on phase tracking [32], [33].

B. Limitations and variations

Recovering the density using FWI is very challenging, independently of the bandwidth extension question, for the following reasons. (1) Cross-talk happens using short-offset data since P-wave velocity and density have the same radiation patterns at short apertures. (2) The variations in density are smaller than those in velocities. (3) Inversion of density requires ultra-low frequencies. Although elastic FWI does not always allow to correctly estimate density, it stands a better chance of properly reconstructing velocities, with either extrapolated or true low frequencies.

In extrapolated FWI, the choice of starting frequency is a trade-off between the accuracy of extrapolated low frequency data and the lowest frequency to mitigate the cycle-skipping problem. We start elastic FWI with $2-4$ Hz extrapolated low frequency data due to the insufficient extrapolation accuracy in the near-zero frequency range. The accuracy of the $2-4$ Hz extrapolated low frequency data is sufficient for elastic FWI of P-wave and S-wave velocities when starting from 4 Hz band-limited data. However, the starting frequency range $2-4$ Hz is still relatively high for the inversion of density, so the inverted density model is lack of the low-wavenumber structure.

C. Future work on the application to field data

The main challenge in the application of this papers idea to field data is the lack of availability of the low frequencies, i.e., the data come unlabeled. Training the neural network on synthetic data and testing it on field data may be one solution. It is helpful to collect a realistic training dataset in the following ways to reduce the generalization error on the field test dataset:

- First, prior knowledge about the model that is being inverted can be used to build the training models. For instance, we can build training models with the same depth of water bottom if the depth has been estimated in advance.
- Second, considering the inevitable discrepancy of amplitude between field and simulated data, we should normalize the band-limited data in the same way as synthetic data. Note that it is not necessary to recover the original amplitude of the field data if FWI uses an amplitude-insensitive cost function.
- Furthermore, the source signal is assumed to be known for extrapolated elastic FWI in this paper. However, for field data, the source signal may vary shot by shot. We can retrieve the source wavelet of the field dataset first, and then artificially boost the low-frequency energy after

denoising. The new source signal can be used to synthesize the training dataset for low-frequency extrapolation. It can also be the source wavelet in the following FWI using the extrapolated low-frequency data. In this way, the uncertainty of the source can be controlled to some extent.

With a careful design of the training data set, the generalization error from synthetic to real data is expected to be small. Otherwise, we may need to propose a new deep learning model that can deal with the unlabeled band-limited data directly.

V. Conclusion

To relieve the dependency of elastic FWI on starting models, low-frequency extrapolation of multi-component seismic recordings is implemented to computationally recover the missing low frequencies from band-limited elastic data. The deep learning model is designed with a large receptive field using dilated convolution to increase the receptive field exponentially with depth. By training the neural network with both horizontal and vertical components, we can extrapolate the low frequencies of multi-component band-limited recordings. The extrapolated $0.125 - 4$ Hz low frequencies match well with the true low-frequency data on the Marmousi2 model. Elastic FWI using $2 - 4$ Hz extrapolated data shows similar results to the true low frequencies. The accuracy of the extrapolated low frequencies is enough to provide low-wavenumber starting models for elastic FWI of P-wave and S-wave velocities on data band-limited above 4Hz.

The generalization ability of the neural network from acoustic to elastic data is studied in this paper. The neural network trained on purely acoustic data shows larger prediction error on elastic test dataset compared to the neural network trained on elastic data. Therefore, collecting more realistic elastic training dataset will help to process the field data with strong elastic effects. Furthermore, it might be worthwhile for future studies to consider designing neural networks with multi-channel inputs and multi-channel outputs, which may enable bandwidth extension of elastic data to benefit from the implicit relationship among different components.

Acknowledgment

Tensorflow and Keras are used for deep learning. Elastic FWI code (DENISE), the modified Marmousi2 model, and the 1D initial models can be accessed at https://github.com/daniel-koehn/DENISE-Black-Edition. Acoustic training datasets are simulated using Pysit [58].

References

[1] M. Marjanović, R.-É. Plessix, A. Stopin, and S. C. Singh, "Elastic versus acoustic 3-d full waveform inversion at the east pacific rise 9 50 n," *Geophysical Journal International*, vol. 216, no. 3, pp. 1497–1506, 2018.

[2] R. Plessix, P. Milcik, H. Rynja, A. Stopin, K. Matson, and S. Abri, "Multiparameter full-waveform inversion: Marine and land examples," *The Leading Edge*, vol. 32, no. 9, pp. 1030–1038, 2013.

[3] A. Stopin, R.-É. Plessix, and S. Al Abri, "Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in oman," *Geophysics*, vol. 79, no. 3, pp. WA69–WA77, 2014.

[4] A. Tarantola, "A strategy for nonlinear elastic inversion of seismic reflection data," *Geophysics*, vol. 51, no. 10, pp. 1893–1903, 1986.

[5] P. Mora, "Nonlinear two-dimensional elastic inversion of multioffset seismic data," *Geophysics*, vol. 52, no. 9, pp. 1211–1228, 1987.

[6] R. Brossier, S. Operto, and J. Virieux, "Which data residual norm for robust elastic frequency-domain full waveform inversion?" *Geophysics*, vol. 75, no. 3, pp. R37–R46, 2010.

[7] ——, "Seismic imaging of complex onshore structures by 2d elastic frequency-domain full-waveform inversion," *Geophysics*, vol. 74, no. 6, pp. WCC105–WCC118, 2009.

[8] F. Aminzadeh, B. Jean, and T. Kunz, *3-D salt and overthrust models*. Society of Exploration Geophysicists, 1997.

[9] L. Sirgue, O. Barkved, J. Van Gestel, O. Askim, and J. Kommedal, "3d waveform inversion on valhall wide-azimuth obc," in *71st EAGE Conference and Exhibition incorporating SPE EUROPEC 2009*. European Association of Geoscientists & Engineers, 2009, pp. cp–127.

[10] Y. Choi, D.-J. Min, and C. Shin, "Frequency-domain elastic full waveform inversion using the new pseudo-hessian matrix: Experience of elastic marmousi-2 synthetic data," *Bulletin of the Seismological Society of America*, vol. 98, no. 5, pp. 2402–2415, 2008.

[11] G. S. Martin, R. Wiley, and K. J. Marfurt, "Marmousi2: An elastic upgrade for marmousi," *The leading edge*, vol. 25, no. 2, pp. 156–166, 2006.

[12] D. Köhn, D. De Nil, A. Kurzmann, A. Przebindowska, and T. Bohlen, "On the influence of model parametrization in elastic full waveform tomography," *Geophysical Journal International*, vol. 191, no. 1, pp. 325–345, 2012.

[13] W. Jeong, H.-Y. Lee, and D.-J. Min, "Full waveform inversion strategy for density in the frequency domain," *Geophysical Journal International*, vol. 188, no. 3, pp. 1221–1242, 2012.

[14] E. Crase, A. Pica, M. Noble, J. McDonald, and A. Tarantola, "Robust elastic nonlinear waveform inversion: Application to real data," *Geophysics*, vol. 55, no. 5, pp. 527–538, 1990.

[15] T. J. Sears, P. J. Barton, and S. C. Singh, "Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: Application to alba field, uk north sea," *Geophysics*, vol. 75, no. 6, pp. R109–R119, 2010.

[16] D. Vigh, K. Jiao, D. Watts, and D. Sun, "Elastic full-waveform inversion application using multicomponent measurements of seismic data collection," *Geophysics*, vol. 79, no. 2, pp. R63–R77, 2014.

[17] E. B. Raknes, B. Arntsen, and W. Weibull, "Three-dimensional elastic full waveform inversion using seismic data from the sleipner area," *Geophysical Journal International*, vol. 202, no. 3, pp. 1877–1894, 2015.

[18] D. Borisov, F. Gao, P. Williamson, and J. Tromp, "Application of 2d full-waveform inversion on exploration land data," *Geophysics*, vol. 85, no. 2, pp. R75–R86, 2020.

[19] S. Mahrooqi, S. Rawahi, S. Yarubi, S. Abri, A. Yahyai, M. Jahdhami, K. Hunt, and J. Shorter, "Land seismic low frequencies: acquisition, processing and full wave inversion of 1.5-86 hz," in *SEG Technical Program Expanded Abstracts 2012*. Society of Exploration Geophysicists, 2012, pp. 1–5.

[20] A. Brenders, J. Dellinger, C. Kanu, Q. Li, and S. Michell, "The wolfspar® field trial: Results from a low-frequency seismic survey designed for fwi," in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 1083–1087.

[21] R.-É. Plessix, G. Baeten, J. W. de Maag, F. ten Kroode, and Z. Rujie, "Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set," *Geophysical Prospecting*, vol. 60, no. Simultaneous Source Methods for Seismic Data, pp. 733–747, 2012.

[22] M. Araya-Polo, J. Jennings, A. Adler, and T. Dahlke, "Deep-learning tomography," *The Leading Edge*, vol. 37, no. 1, pp. 58–66, 2018.

[23] F. Yang and J. Ma, "Deep-learning inversion: A next-generation seismic velocity model building method," *Geophysics*, vol. 84, no. 4, pp. R583–R599, 2019.

[24] Y. Wu and Y. Lin, "Inversionnet: An efficient and accurate data-driven full waveform inversion," *IEEE Transactions on Computational Imaging*, 2019.

[25] Z. Zhang and Y. Lin, "Data-driven seismic waveform inversion: A study on the robustness and generalization," *IEEE Transactions on Geoscience and Remote Sensing*, 2020.

[26] V. Kazei, O. Ovcharenko, P. Plotnitskii, D. Peter, T. Alkhalifah, I. Silvestrov, A. Bakulin, and P. Zwartjes, "Elastic near-surface model estimation from full waveforms by deep learning," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 3872–3876.

[27] F. J. Herrmann, A. Siahkoohi, and G. Rizzuti, "Learned imaging with constraints and uncertainty quantification," *arXiv preprint arXiv:1909.06473*, 2019.

[28] L. Mosser, O. Dubrule, and M. J. Blunt, "Stochastic seismic waveform inversion using generative adversarial networks as a geological prior," *Mathematical Geosciences*, vol. 52, no. 1, pp. 53–79, 2020.
[29] Z. Fang, H. Fang, and L. Demanet, "Deep generator priors for bayesian seismic inversion," *Advances in Geophysics*, vol. 61, pp. 179–216, 2020.
[30] D. Li, F. Gao, and P. Williamson, "A deep learning approach for acoustic fwi with elastic data," in *SEG Technical Program Expanded Abstracts 2019*. Society of Exploration Geophysicists, 2019, pp. 2303–2307.
[31] A. Siahkoohi, M. Louboutin, and F. J. Herrmann, "The importance of transfer learning in seismic modeling and imaging," *Geophysics*, vol. 84, no. 6, pp. A47–A52, 2019.
[32] Y. E. Li and L. Demanet, *Phase and amplitude tracking for seismic event separation*. Society of Exploration Geophysicists, 2015, vol. 80, no. 6.
[33] ——, *Full-waveform inversion with extrapolated low-frequency data*. Society of Exploration Geophysicists, 2016, vol. 81, no. 6.
[34] Y. Li and L. Demanet, *Extrapolated full-waveform inversion: An image-space approach*. Society of Exploration Geophysicists, 2017.
[35] H. Sun and L. Demanet, "Low frequency extrapolation with deep learning," in *SEG Technical Program Expanded Abstracts 2018*. Society of Exploration Geophysicists, 2018, pp. 2011–2015.
[36] O. Ovcharenko, V. Kazei, D. Peter, X. Zhang, and T. Alkhalifah, "Low-frequency data extrapolation using a feed-forward ann," in *80th EAGE Conference and Exhibition 2018*, vol. 2018, no. 1. European Association of Geoscientists & Engineers, 2018, pp. 1–5.
[37] Y. Jin, W. Hu, X. Wu, J. Chen *et al.*, "Learn low wavenumber information in fwi via deep inception based convolutional networks," in *2018 SEG International Exposition and Annual Meeting*. Society of Exploration Geophysicists, 2018.
[38] O. Ovcharenko, V. Kazei, P. Plotnitskiy, D. Peter, I. Silvestrov, A. Bakulin, and T. Alkhalifah, "Extrapolating low-frequency prestack land data with deep learning," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 1546–1550.
[39] H. Sun and L. Demanet, "Elastic full-waveform inversion with extrapolated low-frequency data," in *SEG Technical Program Expanded Abstracts 2020*. Society of Exploration Geophysicists, 2020, pp. 855–859.
[40] ——, "Extrapolated full-waveform inversion with deep learning," *Geophysics*, vol. 85, no. 3, pp. R275–R288, 2020.
[41] W. Hu, Y. Jin, X. Wu, and J. Chen, "Progressive transfer learning for low-frequency data prediction in full waveform inversion," *Geophysics*, vol. 86, no. 4, pp. 1–82, 2021.
[42] O. Ovcharenko, V. Kazei, M. Kalita, D. Peter, and T. Alkhalifah, "Deep learning for low-frequency extrapolation from multioffset seismic data," *Geophysics*, vol. 84, no. 6, pp. R989–R1001, 2019.
[43] L. Demanet and N. Nguyen, "The recoverability limit for superresolution via sparsity," *arXiv preprint arXiv:1502.01385*, 2015.
[44] L. Demanet and A. Townsend, "Stable extrapolation of analytic functions," *Foundations of Computational Mathematics*, vol. 19, no. 2, pp. 297–331, 2019.
[45] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.
[46] A. v. d. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, N. Kalchbrenner, A. Senior, and K. Kavukcuoglu, "Wavenet: A generative model for raw audio," *arXiv preprint arXiv:1609.03499*, 2016.
[47] B. Moseley, A. Markham, and T. Nissen-Meyer, "Fast approximate simulation of seismic waves with deep learning," *arXiv preprint arXiv:1807.06873*, 2018.
[48] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," *arXiv preprint arXiv:1511.07122*, 2015.
[49] F. Aminzadeh, N. Burkhard, J. Long, T. Kunz, and P. Duclos, "Three dimensional seg/eaeg modelsan update," *The Leading Edge*, vol. 15, no. 2, pp. 131–134, 1996.
[50] T. M. Brocher, "Empirical relations between elastic wavespeeds and density in the earth's crust," *Bulletin of the seismological Society of America*, vol. 95, no. 6, pp. 2081–2092, 2005.
[51] K. Xu and G. A. McMechan, "2d frequency-domain elastic full-waveform inversion using time-domain modeling and a multistep-length gradient approach," *Geophysics*, vol. 79, no. 2, pp. R41–R53, 2014.
[52] J. Virieux, "P-sv wave propagation in heterogeneous media; velocity-stress finite-difference method," *Geophysics*, vol. 51, no. 4, pp. 889–901, 1986.
[53] A. R. Levander, "Fourth-order finite-difference p-sv seismograms," *Geophysics*, vol. 53, no. 11, pp. 1425–1436, 1988.
[54] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, "Multiscale seismic waveform inversion," *Geophysics*, vol. 60, no. 5, pp. 1457–1473, 1995.
[55] D. Köhn, "Time domain 2d elastic full waveform tomography," Ph.D. dissertation, 2011.
[56] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
[57] J. W. Thorbecke and D. Draganov, "Finite-difference modeling experiments for seismic interferometry," *Geophysics*, vol. 76, no. 6, pp. H1–H18, 2011.
[58] R. Hewett and L. Demanet, "the pysit team, 2013," *PySIT: Python seismic imaging toolbox v0*, vol. 5, 2013.

Hongyu Sun (S'21) is pursuing the Ph.D. degree in geophysics with Department of Earth, Atmospheric and Planetary Sciences at Massachusetts Institute of Technology, Cambridge, MA, USA. She received her B.S. degree in geophysics in 2014 from Jilin University, Changchun, China. Her research interests range from computational to experimental studies of the Earth, including but not limited to machine learning, seismic imaging, seismic interferometry, inverse problems, signal processing, and geomechanics.

Laurent Demanet received his undergraduate degrees in mathematical engineering and theoretical physics from UCLouvain, Belgium, and his Ph.D. degree in applied and computational mathematics in 2006 under Emmanuel Candès, from Caltech, Pasadena, CA, USA. He is currently a Professor of applied mathematics with the Department of Mathematics, MIT, Cambridge, MA, USA. He is also the Director of MIT's Earth Resources Laboratory. His research interests include signal processing, inverse problems, and wave propagation. He is the recipient of a Sloan research fellowship, a CAREER award from NSF, and a Young Investigator award from AFOSR.