Deep learning for low-frequency extrapolation from multioffset seismic data

Deep learning for low-frequency extrapolation from multi-offset seismic data

Oleg Ovcharenko, Vladimir Kazei, Mahesh Kalita, Daniel Peter and Tariq Alkhalifah

King Abdullah University of Science and Technology (KAUST)

(April 15, 2019)

Running head: **Deep learning for low-frequency extrapolation**

ABSTRACT

Low-frequency seismic data are crucial for convergence of full-waveform inversion to reliable subsurface properties. However, it is challenging to acquire field with an appropriate signal-to-noise ratio in the low-frequency part of spectrum. Here, we extrapolate low-frequency data from its respective higher-frequency components of seismic wavefield by using deep learning. Through wavenumber analysis, we show that extrapolation per shot gather has broader applicability than per trace extrapolation. We numerically simulate marine seismic surveys for random subsurface models and train a deep convolutional neural network to derive a mapping between high and low frequencies. The trained network is then tested on sections from the BP and SEAM Phase I benchmark models. Our results show that we are able to recover 0.25 Hz data from 2-4.5 Hz frequencies. We also demonstrate that the extrapolated data are accurate enough for FWI application.

INTRODUCTION

Delineation of subsurface structures from a complicated seismic dataset, in particular of salt and subsalt regions, is a very challenging task (Etgen et al., 2009; Jones and Davison, 2014), with challenges at every stage, starting from the velocity model building process to migration, followed by its interpretation (Dellinger et al., 2017a,b). To retrieve subsurface parameters from seismic data, full-waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984; Pratt, 1990) proves to be a powerful tomography technique to invert for high-resolution images. It updates an initial model, that often lacks salt-body information, using the gradient of a data misfit functional, through an iterative optimization strategy. These nonlinear gradient-based optimization schemes allow to substantially deviate from the initial model assumptions and introduce strong subsalt features. Nevertheless, due to cycle-skipping problems, they often fail to retrieve reliable models as they converge to local minima rather than finding the global one, specifically when the initial model is too far from the true subsurface structure. Among the many challenging factors such as strong multiples, incomplete acquisition geometries and poor illumination due to complex overburden velocities, the absence of low frequencies in recorded seismic data contributes the most to the failure of FWI (Bunks et al., 1995). In this study, we aim to assist FWI by reconstructing the missing low frequencies, using an artificial neural network (ANN), which is a promising data-driven approach within the machine learning framework.

Seismic acquisition procedures and logistics have been significantly improved over the last two decades. However, they often fail to record the temporal low-frequency seismic data with good signal-to-noise ratio (SNR) (Maxwell and Lansley, 2011). Seismic vibrators and their required mechanical and hydraulic systems are still limited in their ability to

transmit sufficient seismic low frequency energy into the subsurface. In addition to ambient noise, the recording unit has its own system noise (thermal and quantization) that aggravates SNR of low-frequency components in data. This absence of usable low frequencies in seismic datasets leads FWI to reconstruct inaccurate long-scale features, eventually yielding to a local, rather than global, minimum model. To avoid local minima solutions, conventional FWI algorithms aim either to change the misfit function (Bozdağ et al., 2011; Yang and Engquist, 2017), modify the gradient (Alkhalifah, 2015; Kazei et al., 2016; Kalita and Alkhalifah, 2018), incorporate model domain regularization (Ovcharenko et al., 2018b; Kalita et al., 2019), or extrapolate the missing low-frequency part of the spectrum.

In the context of low-frequency extrapolation, several recent advances have been proposed: Hu (2014) introduced a beat-tone inversion to extract low frequencies from the interference of waveforms at neighboring frequencies. Wu et al. (2014) considered a seismogram as a modulated signal and used a de-modulation operator to extract the low-frequency envelope. Li and Demanet (2016) analytically extrapolated the low frequencies by decomposing selected seismic records into elementary events and taking into account inter-trace relations. Wang and Herrmann (2016) addressed frequency extrapolation as a convex optimization problem with a total-variation regularization that accounts for spatial correlation between traces. Discussions on bandwidth extrapolation in data-space and image-space approaches are provided in Li and Demanet (2017), who extrapolated frequency bandwidth of the data by wavenumber extrapolation of the extended images with subsequent extended Born modeling. Still, there is a challenge to find an approach that would fully exploit intrinsic connections between high- and low-frequency signals, together with the spatial behavior, between adjacent traces and shots for complex subsurface structures. With the recent advent of deep learning and artificial intelligence applications, machine learning ap-

proaches have emerged as powerful tools in data analytics. Our proposed method aims at extrapolating missing low frequencies by employing convolutional neural networks (CNN), using a deep learning approach.

Machine learning (ML) has gained a lot of attention over recent years due to its potential to replace human manual routine tasks. In exploration, the seismology community has also adopted ML techniques for various applications (Kong et al., 2018) such as facies classification (Qian et al., 2018), first-break picking (Akram et al., 2017), source mechanism inversions (Ovcharenko et al., 2018a), fault mapping (Guitton, 2018) and various kinds of seismic inversions (Araya-Polo et al., 2018; Mosser et al., 2018; Richardson, 2018; Zhang and Alkhalifah, 2018).

For frequency-bandwidth extrapolations, Ovcharenko et al. (2017) proposed a feed-forward ANN to extrapolate low-frequency data from multi-offset (shot gather) data and applied it to FWI (Ovcharenko et al., 2018c). Meanwhile, Sun and Demanet (2018) built a 1D CNN for trace-by-trace extrapolation and applied it to time-domain data. Another trace-wise approach has been proposed by Jin et al. (2018), who demonstrated the extrapolation by jointly processing beat-tone and raw waveform data using a deep Inception-based convolutional network. Here, we build on previous work by extending the neural network approach to deep CNNs and by further investigating its theoretical limits. Since it is well established in the FWI community, that long offsets can, to some extent, replace low frequencies (Sirgue and Pratt, 2004; Kazei et al., 2013b), we discuss the advantages of shot-to-shot data extrapolation from a theoretical point of view, and evaluate a 2D CNN algorithm for this purpose.

Training a deep CNN with general subsalt models, we highlight the subtleties in such

deep learning techniques, using on physics-based data inferences and the sampling limitations we may face in representing the associated model space. We then extrapolate missing low frequencies of seismic shot gathers in the frequency domain for different acquisitions. With those artificially enhanced datasets with low frequencies, we investigate the potential within a numerical benchmark model to invert long-scalelength features of a specific subsalt model. We start the theory section by discussing the relation between model-wavenumber and data-frequency spectra, followed by a detailed description of the deep learning framework in our study. To show the versatility of our proposed method, we consider datasets of BP-2004 and SEAM models in which the minimum available frequency is 2 Hz, and where the data inference is conducted with the same deep CNN to extrapolate lower frequency sampling points.

THEORY

For low-frequency extrapolation, any data inference technique is not only limited by acquisition geometry and instrumentation, but also by the physics of seismic wave propagation. Thus, let us first derive a simple theoretical model for frequency extrapolation, based on the wavenumber illumination theory. To simplify the theoretical framework, we only consider acoustic waves hereafter. In the frequency domain, the pressure wavefield $p(\mathbf{x},\omega)$ of a point source located at $\mathbf{x}_s$ satisfies the Helmholtz equation

$$(\Delta_{\mathbf{x}} + \frac{\omega^2}{v^2(\mathbf{x})})p(\mathbf{x},\omega) = \delta(\mathbf{x}-\mathbf{x}_s)s(\omega), \tag{1}$$

with $\Delta_{\mathbf{x}} \equiv \sum_i \frac{\partial^2}{\partial x_i}$, and where ω denotes the angular frequency. Given a source signature $s(\omega)$ and a velocity model $v(\mathbf{x})$, equation 1 determines the solution of p. Source signatures can, in principle, be estimated from data (Pratt et al., 1996). Therefore, we assume $s(\omega)$

to be known and normalized at frequencies used for the extrapolation. Without further restriction, we set $s(\omega) = 1$ for each frequency ω. In general, $s(\omega)$ can be set to different values at lower frequencies, e.g., to mimic a Ricker wavelet. Exact locations of source and receivers are also known in most exploration setups. Therefore, if the true velocity model $v(\mathbf{x})$ and exact physics of wave propagation in the associated rock materials are known, low frequencies can simply be modelled by numerical methods.

In the following, we consider the Born approximation for perturbed wavefields in a homogeneous model, where data and model spectra can be directly related (Devaney, 1984; Mora, 1989; Kazei et al., 2013a, 2015; Alkhalifah, 2016). We find the perturbed trace at receiver $\mathbf{g}$ by

$$\delta p(\mathbf{s}, \mathbf{g}, \omega) \propto \delta v(\mathbf{K}) \tag{2}$$

with the wavenumber $\mathbf{K} = \frac{\omega}{\mathbf{v}}(\mathbf{s} + \mathbf{g})$, and where $\mathbf{s}$ and $\mathbf{g}$ are unit vectors pointing towards the source and the receiver, respectively.

In seismic exploration, most subsurface models are dominated by horizontal structures. Thus, vertical wavenumbers in the model spectrum define the variations in these horizontal structures along the vertical direction. Focusing on vertical wavenumbers, equation 2 leads to

$$\delta p(\theta, \omega) \propto \delta v(K_z), \tag{3}$$

where the vertical wavenumber $K_z(\theta, \omega) = 2\frac{\omega}{v}\cos\frac{\theta}{2}$ and θ is the opening angle between $\mathbf{s}$ and $\mathbf{g}$, which encodes the offset. Equation 3 effectively connects data collected at different frequencies with the velocity model. It is evident, from equation 3, that if the angle θ between the source and the receiver is fixed, every wavenumber in the model is then illuminated by a unique frequency, which makes a trace-by-trace extrapolation impossible.

Furthermore, if the wavefield perturbation $\delta p(\theta_1, \omega_1)$ is a known entity, we can construct the model at the wavenumber $K_z(\theta_1, \omega_1)$ and subsequently model any wavefield $\delta p(\theta_2, \omega_2)$, so that $K_z(\theta_2, \omega_2) = K_z(\theta_1, \omega_1)$, according to equation 3. The extrapolation process can thus be summarized as

$$\delta p(\theta_1, \omega_1) \rightarrow \delta v(K_z) \rightarrow p(\theta_2, \omega_2). \tag{4}$$

Equation 4 could, in principle, be directly used for extrapolation, although this would limit us to a validity regime given by the Born approximation. Nevertheless, we use it to outline the limitations of what can be achieved by the artificial neural network approach proposed here.

To summarize, as long as the same model wavenumber K_z is illuminated by two different frequencies, extrapolation is possible. This sets a limit to the lowest frequency that can, in principle, be extrapolated from the model wavenumber perspective

$$K_{lowest} = 2\frac{\omega_{min}}{v}\cos(\theta_{max}/2) \tag{5}$$

where ω_{min} denotes the minimum frequency available in the dataset. Assuming that back-scattering is available, the same wavenumber should be available from the lowest extrapolated frequency

$$K_{lowest} = 2\frac{\omega_{lowest}}{v}\,, \tag{6}$$

which gives us an estimate of the frequency that illuminates the same wavenumber

$$\omega_{lowest} = \omega_{min}\cos(\theta_{max}/2). \tag{7}$$

The last equation defines how far the frequency can be extrapolated without explicit assumptions on the model, other than the Born approximation. For example, we can

estimate the lowest frequency for extrapolation at different locations in the subsurface model, based on equation 7, for a seismic acquisition limited by an available minimum frequency of 2 Hz (Figure 1).

In more realistic scenarios, subsurface velocities increase with depth, which provides larger illumination angles at depth from refractions and diving waves, and makes it possible to estimate lower frequencies. Additionally, we assume that the geological model has layered, block-like structures, and we build this *a priori* geological knowledge into the training set of subsurface models. This allows to push further down the limits of the lowest available frequency for extrapolation.

DEEP LEARNING FRAMEWORK

Neural network architectures date back as early as McCulloch and Pitts (1943). Since then, major advances, supported by powerful computational hardware progresses, have been made, including the recent progress in architectural representations, training capacities and data inferences in various scientific fields and engineering applications (Jordan and Mitchell, 2015). Artificial neural networks represent a mathematically simplified model of neurons connected in a biological brain. The simple mathematical unit of a neuron is called a perceptron. It mimics a neuron that accumulates charge and passes it to a non-linear thresholding output function. Multiple perceptrons assembled into a layered structure shape a simple, fully-connected artificial neural network where all neurons within a layer are connected to each other. Each incoming connection to a neuron has a weight that defines the contribution of this connection to the total output. Neural networks have become one of the most powerful and appealing data analysis tools for non-linear adaptive data regression, using multiple input and multiple output data (Schmidhuber, 2015).

In supervised learning, the neural network is given pairs of input and target data. The training stage consists of a two-step optimization problem of tuning the variable weights of all neurons. In a first step, input data are passed through all layers of the multi-layer network before a misfit is computed between the results from its final output layer and the desired target data. The formulation of misfit is optional, although in generic cases, it is set to a L_n - norm. In a second step, the misfit is back-propagated through each connection in the network, changing each weight according to the respective gradient (for a more complete overview we refer to LeCun et al. (2012).

Convolutional Neural Network (CNN)

Based on our previous experience with feed-forward neural networks (Ovcharenko et al., 2017, 2018c), we see that the major challenge for a fully-connected neural network architecture is the steep growth of the number of trainable parameters, when increasing the size of input and target data. This impedes scalability and leads to an impractical amount of data samples needed for the training stage. A convolutional neural network efficiently addresses this challenge by implementing a concept of local spatial connectivity, meaning that only data within a receptive field are fully-connected. The size of a receptive field corresponds to the spatial extent of a kernel, which always shares the same depth with the input volume, but is usually smaller along width and height axes, as shown schematically in Figure 2.

The convolution of the input data with a kernel results in a feature map that maximizes locations where a pattern in the data matches the kernel.

Feature maps from convolutions with different kernels are stacked together to build an input volume for the next layer. This implies that the number of kernels used in each convo-

lutional layer is equal to the depth of the output volume from this layer. Kernels are usually initiated randomly and are trained by back-propagation of misfits between expected and predicted outputs from the network. After training, each kernel matches a certain feature in the data, with more complex features learned at deeper layers of the network (Eldan and Shamir, 2016). For the purpose of low-frequency extrapolation, we discretize each shot gather in the frequency domain, and treat it as a digital image for feature detection. Adding several convolutional layers together will then allow to generalize these features across all shot gathers and ultimately increase the robustness of the low-frequency extrapolation with CNNs.

Input and target data

Marine seismic acquisition delivers a suitable framework for supervised ML applications. Towed streamers carrying a fixed number of hydrophones naturally matches the limitation imposed on the dataset shape for a neural network. The restriction is that all pairs of training and testing data should share the same dimensions throughout the dataset. Input and target data for the network designed in this study are high- and low-frequency parts of the spectrum of a shot gather, respectively. For each shot, we extract the observed high-frequency part from the frequency spectrum of seismic data recorded by a linear array of hydrophones and treat it as the input, whereas a single low-frequency representation of the shot gather is the target.

In general, neural networks in supervised learning are trained on pairs of input and target data. So far, most neural network setups are determined empirically, meaning that a trial-and-error approach is then necessary to find an optimal network architecture that

will lead to the best inference result. In that sense, both training and validation data are of primary importance, whereas a method to treat these data is optional. This means that the same dataset might be processed by a number of different methods to find the best option. Thus, it is crucial to have a representative initial dataset that will constitute a solid basis to build a ML framework around it. Let us therefore explain in more details how we generate synthetic data, based on physical approximations, to determine a neural network able to generalize across different seismic acquisitions.

To investigate our deep learning approach to extrapolate low-frequencies, we generate a synthetic dataset that mimics a marine seismic acquisition. We solve the Helmholtz wave equation to obtain a pressure wavefield for a single frequency. For each frequency and source-receiver pair, we obtain a complex value describing the pressure field excited by the source and recorded at the receiver location. Data from a set of receivers sharing the same source are shaping a single-frequency shot-gather, which is a complex-valued vector with as many elements as there are receivers (Figure 3).

The entire set of observed mono-frequency data $p(x_s, x_r)$ shapes a data matrix (Figure 4a). We assume an evenly-spaced acquisition, so the matrix is symmetric and squared. When the seismic acquisition covers the entire target area, the data matrix is densely populated. On the other hand, in marine seismic acquisition, the maximum offset is limited by the length of the streamer, which results in missing off-diagonal elements in the data matrix (Figure 4b).

A stack of mono-frequency data matrices for a set of consecutive single frequencies shapes a data cube (Figure 5). Top and bottom surfaces of the cube are the data matrices for the highest and the lowest frequency, respectively. A slice of the cube, normal to the

source axis, represents a shot gather with its spectrum content along the vertical. To preserve continuity in the wavenumber domain, we sample the frequency spectrum as a geometrical set, meaning that every following frequency is obtained from the previous one as the product with a constant coefficient. A similar selection of frequencies is used in the multi-scale approach (Bunks et al., 1995), which is a common practice in frequency-domain FWI.

Exploring the data cube, one could notice continuous stripes extending within a single shot gather, as shown in Figure 6. These features correspond to transmission and reflection modes of the wavefield. Event tracking (Li and Demanet, 2016) essentially stretches data at a short offset and a high frequency to the data at a long offset and a low frequency, along the depicted lines. For a simple homogeneous velocity model, their shape could be derived analytically from a wavenumber analysis, whereas, when the velocity model is inhomogeneous, these features become distorted in a non-linear way. This makes their extrapolation challenging.

METHOD

The standard workflow for a machine learning algorithm consists of only three stages:

- The data. The whole dataset is separated in two parts, one to serve for training of the algorithm, and the other part to be used for validation and performance evaluation. Data preprocessing is also an important step within the preparation stage, as it helps to equalize the contribution of features in the training.

- Selection and training of the algorithm. The dataset is a constant which can be manipulated in a variety of ways, so our goal is to identify an approach that delivers

the best performance result on the validation data.

- Inference. The algorithm infers about the unseen target data based on its training experience.

For a streamer acquisition with R hydrophones, the input data has the shape of $F \times R \times 2$. The last factor 2 is due to the real and imaginary parts of the wavefield in the frequency domain, and F indicates the number of frequencies in the high-frequency part in which the shot gather spectrum has been split. Similarly, the shape of the mono-frequency target data is $1 \times R \times 2$. We discretize the known frequency range into $F = R$ frequencies to get a squared volume with two channels. Equal spatial dimensions in the image assure equal contributions from both vertical and horizontal features within the training stage of the CNN. Figure 7 shows a schematic representation of our input data matrix and output data vector. Note that the desired target output is specific to a single low frequency predicted at R receivers, i.e., each single low-frequency extrapolation is associated with an individual neural network.

Random model generation

For all artificial neural network applications, the training stage is the most crucial step. In supervised learning, multiple pairs of data are needed to train the network, where massive datasets are often required for more complex data analysis applications. For our purposes, many real-world seismic data acquisitions would be required to train a CNN capable of generalizing subsurface features, which would become unfeasible. Instead, we generate all synthetic data for training based on approximate physics, all synthetic data for training and then use the trained CNN for inference on samples of true data generated for bench-

mark models. All synthetic seismic datasets are modeled for marine seismic acquisitions in randomly generated velocity models, and are used to train and validate the neural network.

The generation of realistic velocity models itself is a challenging task and deserves separate consideration (Christakos, 2012). Here, we only list empirical practices that outperformed others in our particular framework. We find that approaches for random model generation such as random Gaussian fields, layered models, and linear gradient models with salt-body intrusions, do not succeed in building a representative set of velocity models that would lead to successful training. A principal component analysis of those random velocity models shows that, often, only a few principal components dominate all others, which may be the result of improper parameter tuning or an implementation bias. However, when building random models from interpolation in-between random 1D velocity profiles (Figure 8), the total model variance is spread among a larger amount of principal components, and thus results in a better network inference on validation data.

The number of points to sample a random 1D profile, as well as the number of profiles needed to produce a velocity model with dominant lateral structures, is defined by the minimum and maximum frequencies to be extrapolated. We compute the range of expected wavelengths within the physical dimension of the velocity model, and then use these wavelengths to limit the thicknesses of layers within a profile. Random profiles are then stacked side-by-side and reshaped to match the size of the original velocity model where the acquisition takes place.

We also introduced a linear background trend into all generated random subsurface models such that it serves as a baseline to limit the lowest velocities in the generated models. To that end, we conducted a rough grid search for trends that minimize the misfit between

observed and modeled data. Each newly generated random model is then compared, in terms of Euclidean distance, with the set of all previously built models to ensure a more even sampling of the model space.

Normalization of features within the dataset is crucial for successful training of any deep learning model. Proper scaling of the data accelerates training and equalizes contributions from each feature in the training stage. We split complex valued wavefield into its real and imaginary parts and scale each of them individually to fit the range $[-1, 1]$. Scaling coefficients are saved and then applied when needed to restore the data values to the original scale that is suitable for FWI applications.

CNN architecture

We design and train each CNN to serve as a non-linear extrapolator from high- to low-frequency representation of a shot gather. Figure 9 shows our CNN architecture, which consists mainly of four convolutional blocks with two fully-connected layers at the end. Input and target data to the network are represented by a complex-valued matrix and a vector, respectively (as shown in Figure 7). For the CNN, each convolutional block includes two consecutive convolutional layers sharing the same set of hyperparameters. Each block is followed by a batch-normalization and a max-pooling layer aimed at reducing the dimensionality and equalizing contributions from each block to the learning process. A stack of convolutional layers with small-sized kernels is more efficient in capturing larger features in the dataset rather than an equivalent single layer with a larger kernel size (Karpathy, 2018). This means that a sequence of two consecutive convolutional layers with a stride of 1 and receptive field of 3×3 is equivalent to a single layer with a receptive field of 5×5,

whereas fewer trainable parameters which are beneficial for preventing overfitting. A set of convolutional blocks compresses input data into a latent space which is then fully-connected by two dense layers to shape the target output.

In more details, there are 16, 32, 64 and 128 kernels used in each of convolutional blocks. We pad each convolution with zeros and use exponential linear units (Clevert et al., 2015) as an activation function for all convolutional layers. The output from the last convolutional layer is flattened into a vector of 512 units, which is then passed to the fully-connected part of the network composed of layers with 136 and 68 neurons, respectively. At the latest stage, we reshaped the output from the last layer of the network to be 34×2, which stands for one complex value for each of the 34 receivers in the streamer line. To meet the range of normalized training data ($[-1, 1]$), we equip the first of two dense layers with a hyperbolic tangent activation function that spans the same interval. The last dense layer, which has a linear activation, serves as a summator for its inputs. Such architecture leads to a total of 373,212 trainable parameters.

Finally, we train the network using Adam's optimizer (Kingma and Ba, 2014), with a batch size of 32 and a learning rate dropping down to 10^{-5}. We initialize random weights following Glorot and Bengio (2010), and use early stopping to prevent overfitting when training does not advance for more than six iterations. For the numerical implementation of the CNN, we used Keras (Chollet et al., 2015) with a TensorFlow (Abadi et al., 2015) backend.

EXAMPLES

In this section, we illustrate our low-frequency extrapolation approach for the central and left parts of the BP 2004 (Billette and Brandsberg-Dahl, 2005) and SEAM Phase I (Fehler and Keliher, 2011) benchmark models. To demonstrate its validity, we conduct a multi-scale acoustic FWI with the extrapolated low frequencies and show results for the central part of BP 2004 velocity model.

The central segment of the BP 2004 model is 24 km long and 7 km deep. Surface acquisition involves 68 collocated sources and receivers evenly placed with 320 m spacing. Input and target data for the network are assembled from samples of equal size extracted from the acquisition data matrix, according to the scheme shown in Figure 4b where we limit the maximum offset in the data is limited to half of the model width (which is 12 km). This mimics a standard streamer acquisition and allows us to process all the shots in an efficient manner, using reciprocity to process all data that are not inferred directly. All benchmark sections share the same acquisition geometry and model dimensions, the only difference being in the distribution of acoustic parameters.

Dataset

The synthetic dataset includes shot gathers from marine surveys generated for 400 random subsurface velocity models. With 68 unique receiver locations, whereas the network extrapolates mono-frequency data for only 34 of them at a time. Thus, to shape a squared input volume for the network, we need to model 34 mono-frequencies for each shot gather, within the known range from 2 Hz to 4.5 Hz and a single target low-frequency for output.

Figures 10a and 10b compare the extrapolation accuracies of the network for multiple

models, different training set sizes and different misfit definitions, respectively. From a total of 27,200 data pairs, 20% (5,440 samples) are used as a validation dataset, whereas the rest of 21,760 samples serve as a training dataset. The extrapolation accuracy is then measured as the total sum of all misfits between the target and inferred outputs for the validation set. When compared with a training dataset of twice this size, we find that the extrapolation accuracy of the network improves, although only slightly. This may be due to a relatively homogeneous dataset among the generated random velocity models, where more models only lead to little extra information.

Furthermore, we do not observe significant differences when solving the network optimization for an L_1 or L_2 misfit function. In general, we find that very low frequencies are better extrapolated. This is somewhat expected, as for longer wavelength signals, the effects of complex subsurface features become less pronounced, thus easier to detect for a general regression system.

Low-frequency extrapolation results

The ability to generalize data inference is one of the most beneficial features of artificial neural networks (Giles and Maxwell, 1987). Complex non-linear relations learned by the neural network enable it to produce also reasonable results when applied to data it has never seen. We check this generalization ability of our networks, trained each for a single low-frequency extrapolation, by feeding them high-frequency data generated for several benchmark velocity models. All benchmark models are mimicking a realistic geology of salt-induced media, thus the same neural network should be able to infer the low-frequency signal for each of them.

In the following, we compare real parts of true and extrapolated data matrices and their differences for 0.25, 0.5 and 1 Hz. We also compare data for central and left parts of the BP 2004 benchmark model, as well as for a section from the SEAM Phase I model. Extrapolated data for each shot gather in the acquisition, at different frequencies, are generated by individual CNNs. As previously shown in Figure 7, amplitudes of the extrapolated low-frequency wavefield for a shot gather are produced from high-frequency data within the known frequency range. The known frequency range is sampled by as many individual frequencies as there are receivers.

BP 2004 central. The massive salt body in the middle of the BP 2004 benchmark model is one of its distinctive features (Figure 11a). We extract this salt-containing section to show the benefits from using the extrapolated low-frequency data in such a complex geology. The main complexity comes from the presence of both flat and oblique flanks of the salt body, which cause strong reflections and multi-scattering. In particular, the strong reflections from the salt-sediment surface lower the signal energy propagating through the salt body, which then causes tomographic imaging techniques to fail. Additionally, the steep flanks cause multi-scattering, which becomes a challenge for methods relying on the Born approximation. Finally, there is a low-velocity anomaly below the salt, which may be a hydrocarbon collector. Targeting the inversion to this collector, we first extrapolate low-frequency data and will show its use in a multi-scale FWI.

As noted above, the maximum offset in the data is limited to half of the model width (12 km). Due to this limited acquisition offset, the illumination in deep parts of the model becomes poorer (Figure 11b). We can therefore expect that the neural network training will mostly corroborate connections in the data, due to variations in the upper part of the

velocity model, where data coverage is the highest.

Extrapolated low-frequency data (Figure 11c) follows the major trends in the true data. The mismatch between data matrices increases at higher frequencies, possibly caused by more complex contributions of subsurface features into the total misfit. During the training stage of the network, we search for the minimum of the L_2 loss function that is mostly sensitive to the largest deviations in the data. However, we find that optimizing for a more sensitive L_1 misfit does not improve the overall extrapolation results (Figure 10b). From other side, parametrization of the input data defines features of the wavefield that contribute to the norm. In this study, we treat the complex wavefield by explicitly feeding its real and imaginary parts to the network. A different parametrization of the wavefield by its amplitude and phase didn't lead to significantly different extrapolation results, however there is still a room further exploration. For this reason, in future studies, we plan to investigate both misfit functions that are more sensitive to phase variations and other parametrizations of the data.

BP 2004 left. In the section from the left-side of the BP 2004 model, there is an elongated salt body covering almost half of the section's bottom part (Figure 12a). Without low-frequency data, conventional FWI stagnates, retrieving only the top of the salt and corrupting the rest of the image, due to a lack of illumination. This is another case where low-frequencies would highly contribute to the success of full-waveform inversions.

Distinctive features in the data matrix in Figure 12b are parallel stripes, shaped by off-diagonal elements. These are built by strong reflections coming from the top of the salt in which its surface is lying below water and a thin layer of sediments. Extrapolation

results show a mismatch in fine details at higher frequencies, whereas smooth data is well-extrapolated at lower frequencies.

SEAM Phase I. Despite some similarities in the placement of the salt body, at the center of the velocity model, data matrices from the SEAM Phase I (Figure 13a) and the central section from the BP 2004 model (Figure 11a) look very different. The shape of the salt surface, as well as the water depth, strongly affect the structure of data matrices, due to reflections of different amplitudes arriving at receivers with different phase shifts.

Extrapolated data for the section from the SEAM Phase I model (Figure 13b) show a fit at low frequencies that is worse than the fit for other test examples listed above. The largest errors occur in parts of the data matrices where seismic sources are placed right above the salt surface (centers of top and bottom edges of the data matrix). A reason for this is the resizing (for testing purposes) of the original model, in order to fit the same dimensions as those for the randomly generated data, which results in a very shallow water layer above the salt. For low frequencies, such a small gap is less than a quarter wavelength in the water, and thus becomes negligible, leading both the modeling of the synthetic training data, as well as the extrapolation by the network, to poor reconstruction results for those particular areas.

FWI application

To test the usefulness of the extrapolated low-frequency data, we follow a multi-scale approach running an acoustic FWI (Bunks et al., 1995) for the central part of the BP 2004 benchmark velocity model (Figure 11a). We initiate the inversion at the lowest extrapolated frequency, and use the inversion result at that frequency as an initial model for FWI at the

next higher frequency. The set of frequencies used in this strategy is that of a geometric sequence. Assuming that observed data are available within the range from 2 Hz to 4.5 Hz, we extrapolate the missing data needed for this multi-scale approach, starting from 0.25 Hz with a geometric multiplier of 1.3. The corresponding total sequence consists of twelve frequencies, with eight extrapolated ones $\{0.25, 0.33, 0.42, 0.55, 0.72, 0.93, 1.21, 1.57\}$ Hz, and four within the known range $\{2.05, 2.66, 3.46, 4.50\}$ Hz.

The CNN learns to map a single low frequency from a set of high frequencies. Therefore, to extrapolate eight target low frequencies used for the multi-scale approach, we train eight neural networks sharing the same architecture and input training data set. The only part that differs is the target data, different for each target low frequency. The inversion is then sensitive both to phase and amplitude of the extrapolated data. However, since these values may not be entirely accurate, we compensate for inaccuracies of poorly extrapolated data by adding an additional regularization term to the misfit function. As shown in the Figure 10b, the extrapolation errors increase for higher frequencies, meaning that using poorly predicted data at later inversion stages will hamper inverted velocity models. Thus, to successfully run the inversion we use the first five extrapolated low frequencies only, and fill the remaining gap by adding regularization terms to the misfit function (Kazei et al., 2017).

The largest velocity anomalies in the resulting subsurface model are placed in their correct locations when inverting the lowest extrapolated frequency data of 0.25 Hz. According to Figure 10b, the best match of true and extrapolated data is reached for the lowest frequencies, which is confirmed in the image domain, when running an inversion of the extrapolated data. The resulting image from FWI, at the highest frequency of 4.5 Hz, recovers most features in the central part of the model. We also see the target low-velocity arch,

below the massive salt body. Examples of constrained FWI applied to similar benchmark models are shown in Esser et al. (2016); Ovcharenko et al. (2018b); Kalita et al. (2019).

To summarize, FWI applied to extrapolated low-frequency data converges to a reasonable initial model, which then leads the inversion at higher frequencies to a more robust final velocity model than that obtained without extrapolated frequencies. As a result, we see the current neural network approach to low-frequency extrapolation as being potentially interesting for FWI applications in complex subsalt models.

DISCUSSION

Deep neural networks are powerful data-driven mathematical models able to derive relations directly from data. This allows accounting for realistic physics such as visco-elastic anisotropic effects, as no explicit modeling is involved at the inference stage. We use an acoustic wave equation for modeling, which is most likely not realistic. A common drawback is that there is no unambiguous approach to determine the architecture and set of hyperparameters of the neural network. The architecture plays a crucial role, but there are only empirical practices on how to design it. The structure of the CNNs we implemented in this study was motivated by the need to compress input data into a latent space, and then map it onto a space with the dimensions of the target data. It is therefore an exploratory setup, where the current experience can only guide to future improvements in such a machine learning framework.

Generation of synthetic training data can be prohibitively expensive when attempting to train a network to extrapolate a complete low-frequency dataset from entire survey data recorded for numerous seismic sources. In such scenario of survey-to-survey mapping,

one random velocity initialization only delivers one training sample. However, targeting to interpolate for a shot gather in a single pass, rather than for entire survey, drastically reduces the time needed to generate a training dataset, while preserving a promising extrapolation capability. For a single random subsurface model, we generate as many training data pairs as there are sources. We benefit from this approach even more when we run the wave propagation modeling in the frequency domain where computational costs do not scale as fast as that in the time domain, when modeling for multiple sources. Still, it remains unclear which particular features from the synthetic dataset generated with approximate physics – purely acoustic in this study – can help in training neural networks for them to be able to generalize data inference across different subsalt acquisitions.

Time-domain data is natural for real-world seismic surveys. However, diversity in duration and sampling rates of recorded traces lead to variable input dimensions of data and make it challenging to directly use such time-domain data in ML applications, unless these are designed to process time series data. Proper preprocessing and compression of time-domain data seems to be the key point to address prior to developing an deep learning model. Otherwise, conversion of time-domain data to frequency domain also leads us to the sufficient framework for frequency bandwidth extrapolation.

Although low-frequency extrapolation results are not entirely accurate, we were still able to demonstrate that, if used within a multi-scale inversion strategy combined with additional regularizations, they can help the FWI to correct large scale-length velocity features in the initial model. The neural network approach demonstrated here has the advantage of being a purely data-driven regression tool, able to generalize across different seismic acquisitions. It therefore minimizes assumptions on data connectivity or physics-related extrapolations. However, implicitly, we still include *a priori* geological information in form of the dataset

generated for the training stage. Future research needs to evaluate how this sampling of the model space affects, and may improve the network inference accuracy. That said, the current neural network extrapolation results help to further enhance the robustness of FWI in complex subsurface regions.

With recent advances in the domain of multiparameter scattering in the wavenumber domain the theory can easily be extended to elastic and anisotropic cases. Wave propagation on HPC architectures, in frequency and time domains, paves the way for the extension of the numerical part of the studies. In future studies, we also plan to apply the method to real data.

CONCLUSIONS

We explain the feasibility limits of the frequency bandwidth extrapolation from a wavenumber illumination perspective. Through wavenumber analysis, we show that high-frequency data recorded at long offsets are linked, through the subsurface velocity distribution, to low-frequency data at short offsets. Therefore, bandwidth extensions of individual traces are not viable without strong constraints on the model. However, extensions of full data sets are possible, assuming that the inverse problem is, in principle, resolvable with the available data.

We propose the framework for multi-offset low-frequency data extrapolation. Treating the entire acquisition as a collection of independent shot gathers is beneficial for several reasons. First, shot-wise low-frequency data extrapolation makes the proposed neural network-based technique applicable without retraining to a range of exploration setups with fixed-offset geometry such as in marine streamer acquisition. Second, the generation of the

synthetic training dataset in the frequency domain drastically reduces computational costs as we are able, from a single velocity model, to produce as many training data pairs as there are sources.

We designed a deep convolutional neural network to generate a low-frequency representation of a shot gather, given its high-frequency part, and trained it on data generated from random velocity models. Inputs and outputs from the network are multiple high-frequency and single low-frequency representations of a shot gather, respectively. We subsequently apply the trained network to extrapolate the spectrum, for several benchmark data sets. Extrapolation excels at capturing general trends in the data, which is enough for extrapolating very low frequencies. On the other hand, fine features at higher frequencies are missing. Finally, we run an acoustic FWI for a benchmark model, using the extrapolated low-frequency data. The synthetic inversion tests show that the artificial data of low frequencies are accurate enough to correct most of the large scale-length error in the initial model, and help FWI to converge.

APPENDIX A

APPLICATION TO NOISY DATA

Seismic data are always corrupted by noise in real-world scenarios, thus we want to examine the capability of the network to treat noisy data.

We train the network on synthetic data for two scenarios. First, we add an artificial Gaussian noise, SNR=14, to the input part of the training dataset and subsequently test the network on a dataset with SNR=7, for the central part of BP 2004 benchmark model (Figure 11a). A reduced level of added noise, during training acts as a regularization

(Bishop, 1995), which helps to prevent overfitting and to handle noise at the inference stage. Second scenario demonstrates the case when the network uses a noise-free dataset for training, whereas noisy data is given as input for inference.

Low-frequency data reconstructed from noisy data exhibit features similar to those in true data for both scenarios (Figure A-1). Noise in the high-frequency data affects mostly the amplitude of the extrapolated low-frequency data, whereas the phase remains close to the inference results from noise-free data. However, the network trained on clean data did not succeed in extrapolating for the higher frequencies from noisy input data.

ACKNOWLEDGMENTS

We are grateful to Gerhard Pratt and to Wim Mulder for their comments and to Xiangliang Zhang who advised us on machine learning. We also give a credit to Tristan Van Leuween whose open source FWI code was used as a building block in our inversion scheme (https://github.com/tleeuwen/SimpleFWI). We thank members of the Seismic Modeling and Inversion group (SMI) and the Seismic Wave Analysis Group (SWAG) at KAUST for constructive discussions. The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia.

REFERENCES

Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, 2015, TensorFlow: Large-scale machine learning on heterogeneous systems. (Software available from tensorflow.org).

Akram, J., O. Ovcharenko, and D. Peter, 2017, A robust neural network-based approach for microseismic event detection, *in* SEG Technical Program Expanded Abstracts 2017: SEG, 2929–2933.

Alkhalifah, T., 2015, Scattering-angle based filtering of the waveform inversion gradients: Geophysical Journal International, **200**, 363–373.

———, 2016, Full-model wavenumber inversion: An emphasis on the appropriate wavenumber continuation: Geophysics, **81**, no. 3, R89–R98.

Araya-Polo, M., J. Jennings, A. Adler, and T. Dahlke, 2018, Deep-learning tomography: The Leading Edge, **37**, 58–66.

Billette, F., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: Presented at the 67th EAGE Conference & Exhibition.

Bishop, C. M., 1995, Training with noise is equivalent to tikhonov regularization: Neural computation, **7**, 108–116.

Bozdağ, E., J. Trampert, and J. Tromp, 2011, Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements: Geophysical Journal International, **185**, 845–870.

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

Chollet, F., et al., 2015, Keras: `https://keras.io`.

Christakos, G., 2012, Random field models in earth sciences: Courier Corporation.

Clevert, D.-A., T. Unterthiner, and S. Hochreiter, 2015, Fast and accurate deep network learning by exponential linear units (elus): arXiv preprint arXiv:1511.07289.

Dellinger, J., A. Brenders, X. Shen, I. Ahmed, J. Sandschaper, and J. Etgen, 2017a, What mistakes are we making while interpreting salt? Could FWI help?: Presented at the 79th EAGE Conference and Exhibition 2017-Workshops.

Dellinger, J., A. J. Brenders, J. Sandschaper, C. Regone, J. Etgen, I. Ahmed, and K. Lee, 2017b, The garden banks model experience: The Leading Edge, **36**, 151–158.

Devaney, A. J., 1984, Geophysical diffraction tomography: IEEE Transactions on Geoscience and Remote Sensing, **GE-22**, 3–13.

Eldan, R., and O. Shamir, 2016, The power of depth for feedforward neural networks: Conference on Learning Theory, 907–940.

Esser, E., L. Guasch, F. J. Herrmann, and M. Warner, 2016, Constrained waveform inversion for automatic salt flooding: The Leading Edge, **35**, 235–239.

Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, WCA5–WCA17.

Fehler, M., and P. J. Keliher, 2011, SEAM Phase 1: Challenges of subsalt imaging in tertiary basins, with emphasis on deepwater Gulf of Mexico: Society of Exploration Geophysicists.

Giles, C. L., and T. Maxwell, 1987, Learning, invariance, and generalization in high-order neural networks: Applied optics, **26**, 4972–4978.

Glorot, X., and Y. Bengio, 2010, Understanding the difficulty of training deep feedfor-

ward neural networks: Proceedings of the thirteenth international conference on artificial intelligence and statistics, 249–256.

Guitton, A., 2018, 3d convolutional neural networks for fault interpretation: Presented at the 80th EAGE Conference and Exhibition 2018.

Hu, W., 2014, FWI without low frequency data-beat tone inversion, *in* SEG Technical Program Expanded Abstracts 2014: Society of Exploration Geophysicists, 1116–1120.

Jin, Y., W. Hu, X. Wu, and J. Chen, 2018, Learn low wavenumber information in fwi via deep inception based convolutional networks, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 2091–2095.

Jones, I. F., and I. Davison, 2014, Seismic imaging in and around salt bodies: Interpretation, **2**, SL1–SL20.

Jordan, M. I., and T. M. Mitchell, 2015, Machine learning: Trends, perspectives, and prospects: Science, **349**, 255–260.

Kalita, M., and T. Alkhalifah, 2018, Multiscale full-waveform inversion using flux corrected transport, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 1153–1157.

Kalita, M., V. Kazei, Y. Choi, and T. Alkhalifah, 2019, Regularized full-waveform inversion with automated salt-flooding: Geophysics, **84**, 1–74.

Karpathy, A., 2018, CS231n Convolutional Neural Networks for Visual Recognition.

Kazei, V., M. Kalita, and T. Alkhalifah, 2017, Salt-body inversion with minimum gradient support and sobolev space norm regularizations: Presented at the 79th EAGE Conference and Exhibition 2017.

Kazei, V., B. Kashtan, V. Troyan, and W. Mulder, 2015, FWI spectral sensitivity analysis in the presence of a free surface, *in* SEG Technical Program Expanded Abstracts 2015:

Society of Exploration Geophysicists, 1415–1419.

Kazei, V., E. Tessmer, and T. Alkhalifah, 2016, Scattering angle-based filtering via extension in velocity, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 1157–1162.

Kazei, V., V. Troyan, B. Kashtan, and W. Mulder, 2013a, On the role of reflections, refractions and diving waves in full-waveform inversion: Geophysical Prospecting, **61**, 1252–1263.

Kazei, V. V., B. M. Kashtan, V. N. Troyan, and W. A. Mulder, 2013b, Spectral sensitivity analysis of FWI in a constant-gradient background velocity model: Presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013.

Kingma, D. P., and J. Ba, 2014, Adam: A method for stochastic optimization: arXiv preprint arXiv:1412.6980.

Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft, 2018, Machine learning in seismology: Turning data into insights: Seismological Research Letters.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: theory and application, Society for Industrial and Applied Mathematics, Philadelphia, PA, 206–220.

LeCun, Y. A., L. Bottou, G. B. Orr, and K.-R. Müller, 2012, Efficient backprop, *in* Neural networks: Tricks of the trade: Springer, 9–48.

Li, Y., and L. Demanet, 2017, Extrapolated full-waveform inversion: An image-space approach, *in* SEG Technical Program Expanded Abstracts 2017: Society of Exploration Geophysicists, 1682–1686.

Li, Y. E., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency

data: Geophysics, **81**, R339–R348.

Maxwell, P., and M. Lansley, 2011, What receivers will we use for low frequencies?, *in* SEG Technical Program Expanded Abstracts 2011: Society of Exploration Geophysicists, 72–76.

McCulloch, W., and W. Pitts, 1943, A logical calculus of the ideas immanent in nervous activity: Bulletin of Mathematical Biophysics, **7**, 115–133.

Mora, P., 1989, Inversion = migration + tomography: Geophysics, **54**, 1575–1586.

Mosser, L., W. Kimman, J. Dramsch, S. Purves, A. De la Fuente Briceño, and G. Ganssle, 2018, Rapid seismic domain transfer: Seismic velocity inversion and modeling using deep generative neural networks: Presented at the 80th EAGE Conference and Exhibition 2018.

Ovcharenko, O., J. Akram, and D. Peter, 2018a, Feasibility of moment tensor inversion from a single borehole data using artificial neural networks: Search and Discovery.

Ovcharenko, O., V. Kazei, D. Peter, and T. Alkhalifah, 2017, Neural network based low-frequency data extrapolation, *in* 3rd SEG FWI workshop: What are we getting?: Society of Exploration Geophysicists.

———, 2018b, Variance-based model interpolation for improved full-waveform inversion in the presence of salt bodies: Geophysics, **83**, 1–60.

Ovcharenko, O., V. Kazei, D. Peter, X. Zhang, and T. Alkhalifah, 2018c, Low-frequency data extrapolation using a feed-forward ann: Presented at the 80th EAGE Conference and Exhibition 2018.

Pratt, R. G., 1990, Inverse theory applied to multi-source cross-hole tomography.: Geophysical Prospecting, **38**, 311–329.

Pratt, R. G., Z.-M. Song, P. R. Williamson, and M. R. Warner, 1996, Two-dimensional

velocity models from wide-angle seismic data by waveform inversion: Geophysical Journal International, **124**, 323–340.

Qian, F., M. Yin, X.-Y. Liu, Y.-J. Wang, C. Lu, and G.-M. Hu, 2018, Unsupervised seismic facies analysis via deep convolutional autoencoders: Geophysics, **83**, A39–A43.

Richardson, A., 2018, Generative adversarial networks for model order reduction in seismic full-waveform inversion: arXiv preprint arXiv:1806.00828.

Schmidhuber, J., 2015, Deep learning in neural networks: An overview: Neural networks, **61**, 85–117.

Sirgue, L., and R. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.

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

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

Wang, R., and F. Herrmann, 2016, Frequency down extrapolation with tv norm minimization, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 1380–1384.

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

Yang, Y., and B. Engquist, 2017, Analysis of optimal transport and related misfit functions in FWI: Geophysics, **83**, 1–25.

Zhang, Z., and T. Alkhalifah, 2018, Regularized elastic full waveform inversion using deep learning: Geophysics, submitted.

LIST OF FIGURES

10 Comparison of misfits between target and inferred data for the networks trained on datasets built from 200 and 400 random velocity models (a), and optimization for L_1 and L_2 loss functions (b).

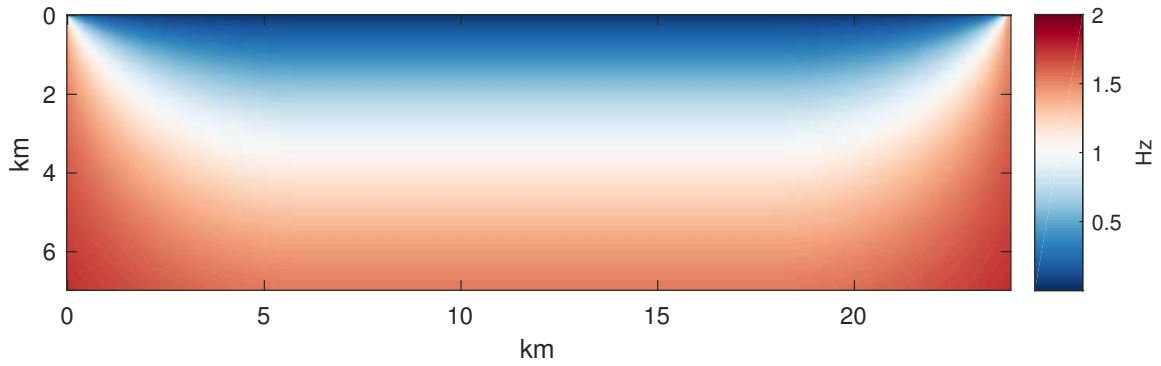
11 Central section from the BP 2004 benchmark velocity model (a). Normalized illumination in the model accounting for truncated half-offset acquisition. White lines indicate ray paths of diving wave, green points indicate locations of collocated sources and receivers (b). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (c).

12 Left section from BP 2004 benchmark velocity model (a). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (b).

13 A section from SEAM Phase I benchmark velocity model (a). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (c).

14 Multiscale full-waveform inversion of low-frequency data extrapolated by CNN (a-b) and data from the known interval (c-d).

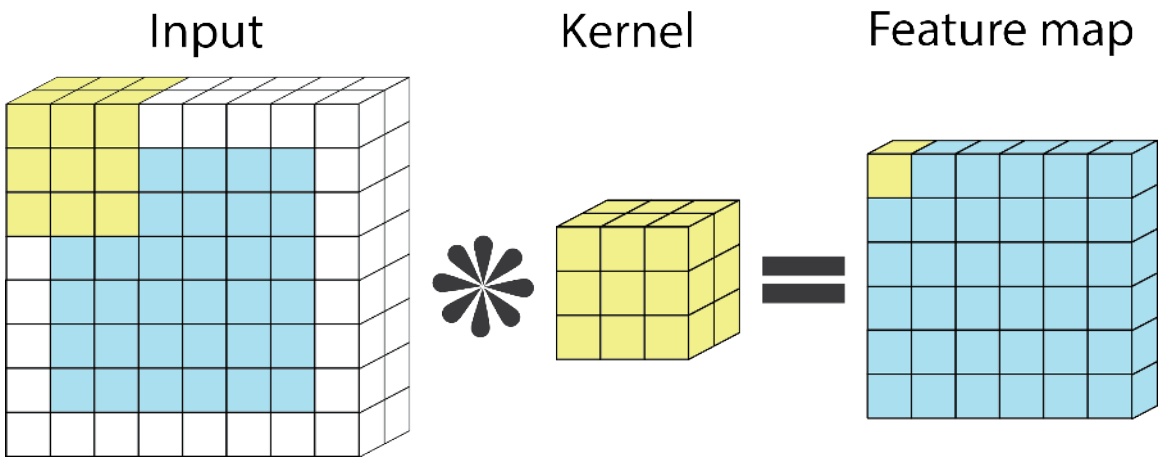
A-1 Extrapolation of low-frequency data for 0.25, 0.5 and 1 Hz from noisy high-frequency data within 2-4.5 Hz range, and for a SNR value of 7. The network was trained for two scenarios, being given input training data without noise and with noise. The later application delivers better inference result.



a

Figure 1: Frequency estimation ω_{lowest} for extrapolation based on wavenumber illumination theory, for a given minimum signal of 2 Hz available in the acquisition dataset and a maximum source-receiver offset of 12 km. This model size setup will be used in the subsequent benchmark study.

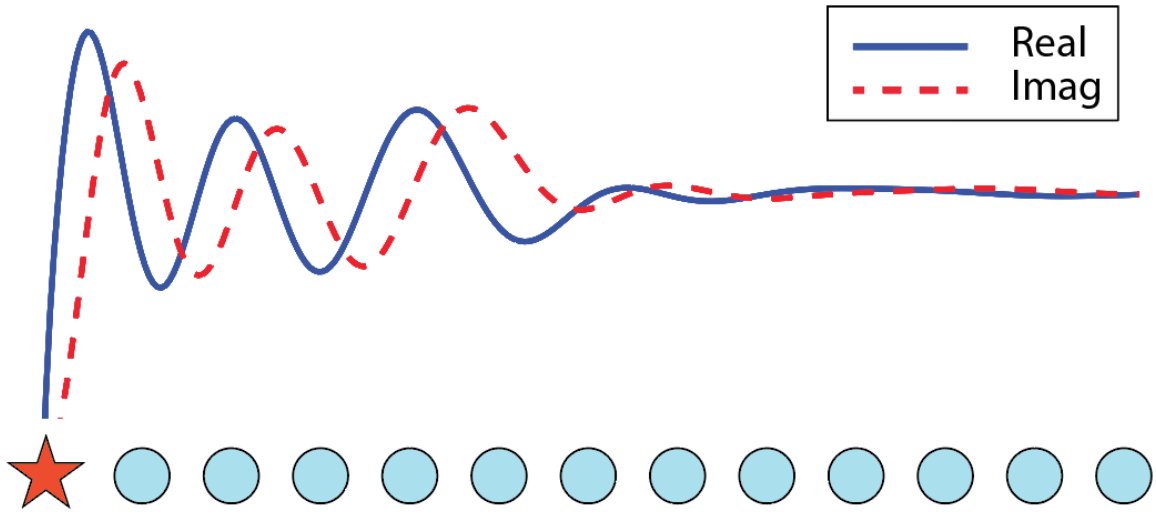
–



a

Figure 2: Convolution of input volume with a kernel resulting in a feature map. Depth of the kernel is always equal to the depth of the input volume, whereas its spatial extent may differ.

–



a

Figure 3: Example of mono-frequency representation of a shot gather. This is a complex-valued vector with as many elements as there are receivers in the acquisition. The red star indicates the source position and blue dots represent receivers.

_

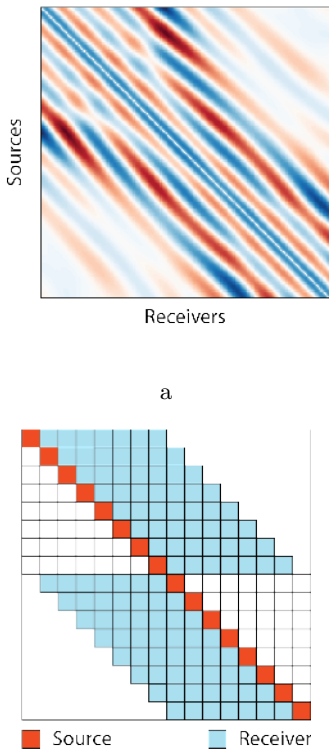
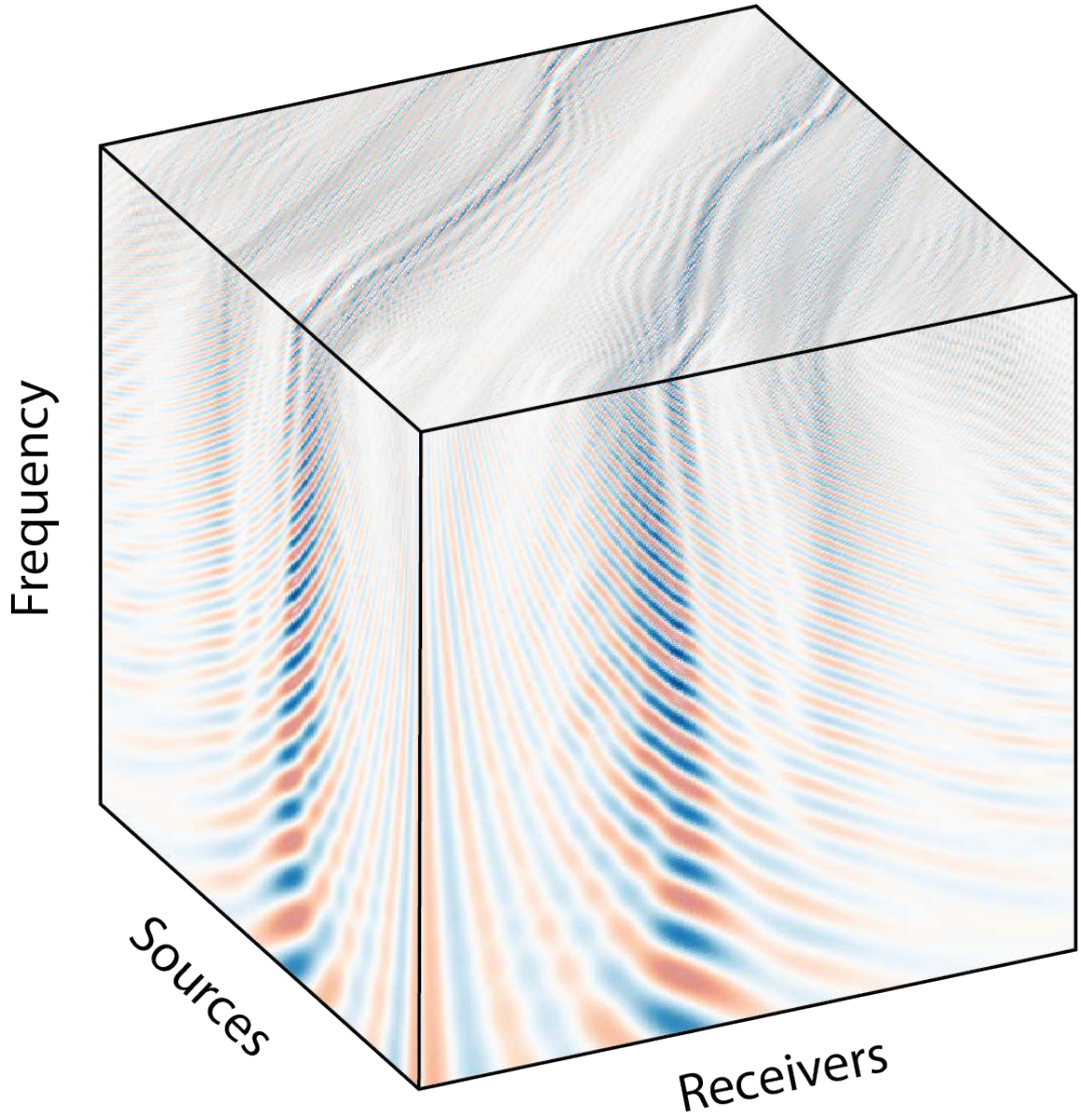


Figure 4: Real part of a mono-frequency data matrix (a) and its mapping onto fixed-offset streamer acquisition layout (b).

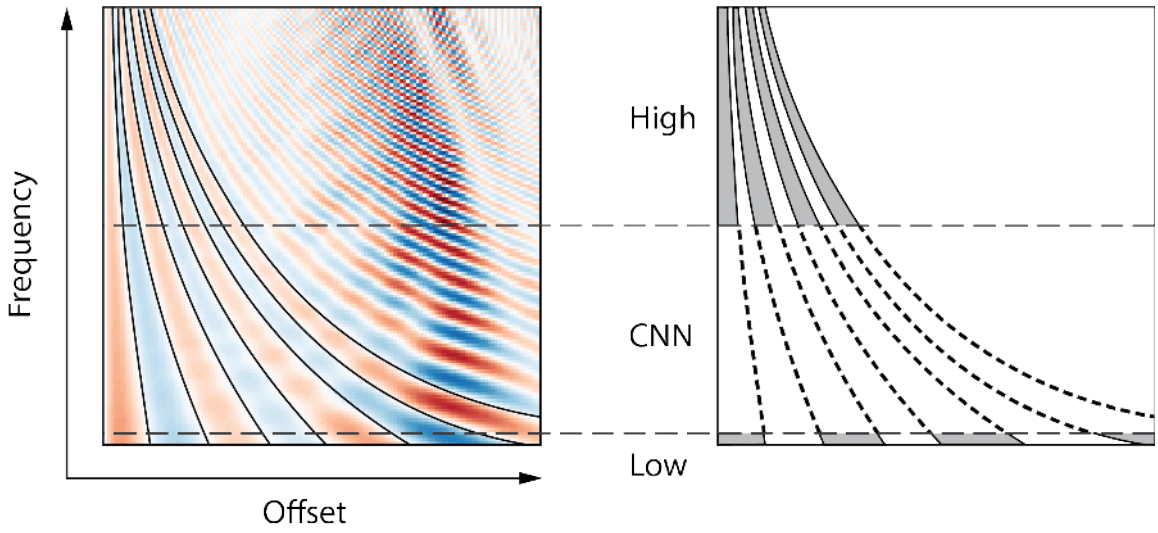
–



a

Figure 5: A data cube created from a stack of mono-frequency data matrices. Each side slice of the cube is a shot gather, decomposed into a number of frequencies.

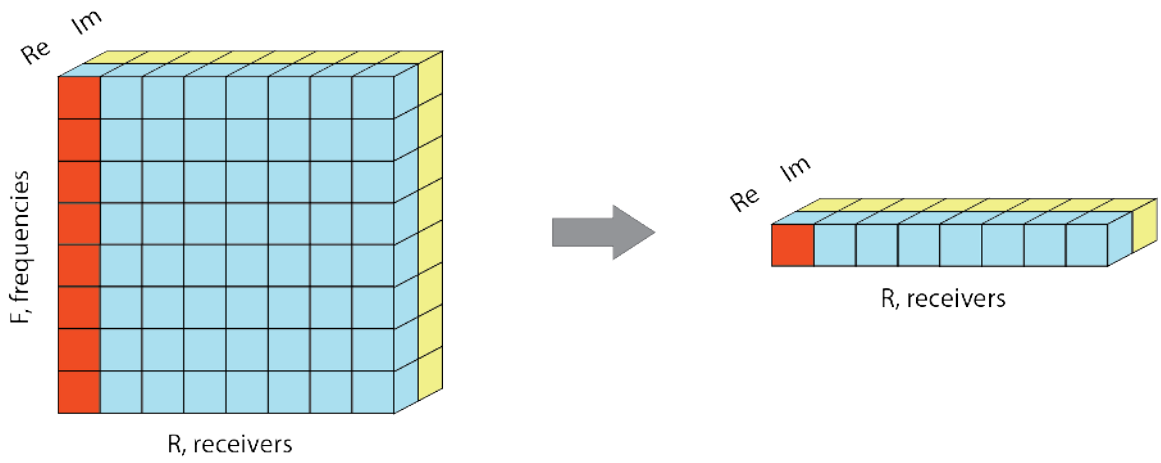
–



a

Figure 6: Single shot gather in the frequency domain. Lines indicate the stationary phase for transmitted waves. For reflections, the stationary phases are in the opposite direction. The data regime of the convolutional neural network (CNN) connects high to low frequencies.

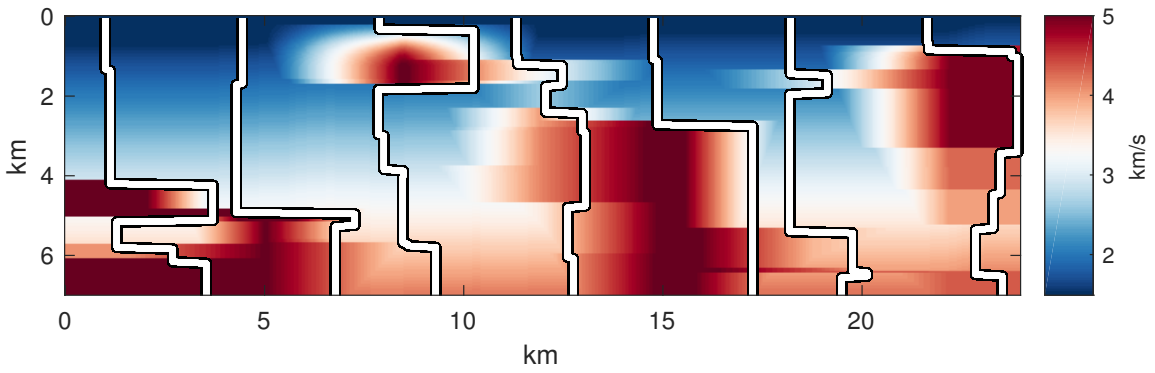
–



a

Figure 7: Input data is a sampled high-frequency part of a shot gather spectrum, whereas target data is a single low-frequency shot gather representation.

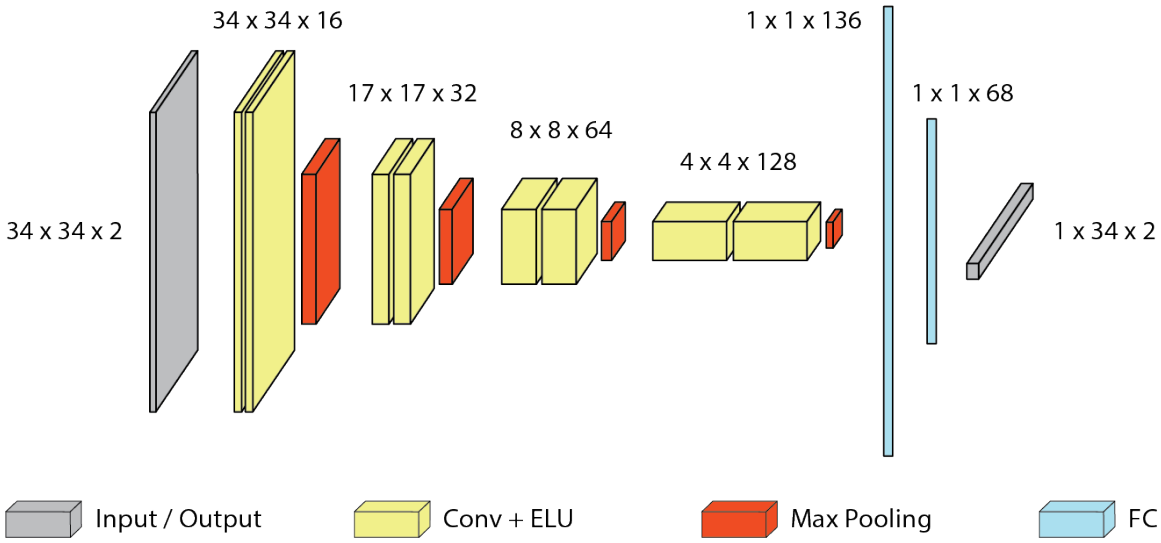
–



a

Figure 8: Random velocity model created by interpolation in-between random 1D velocity profiles.

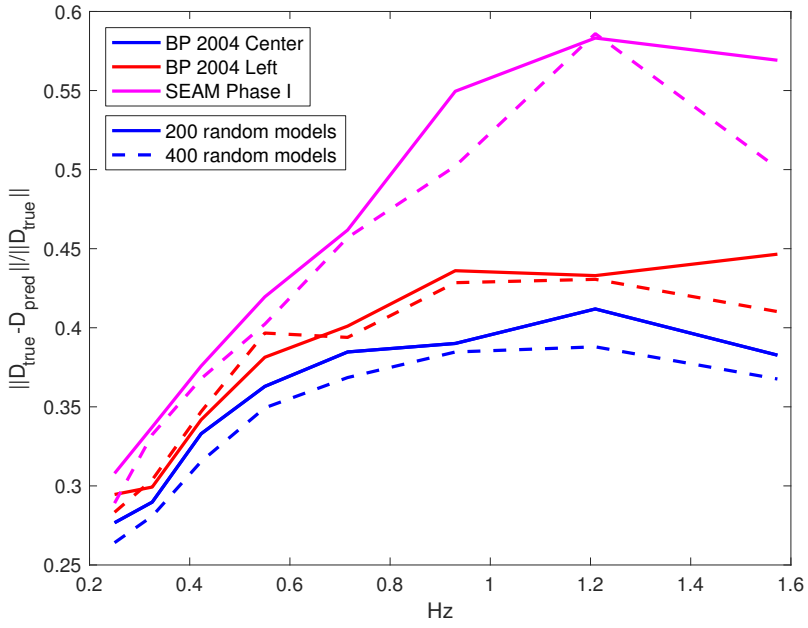
–



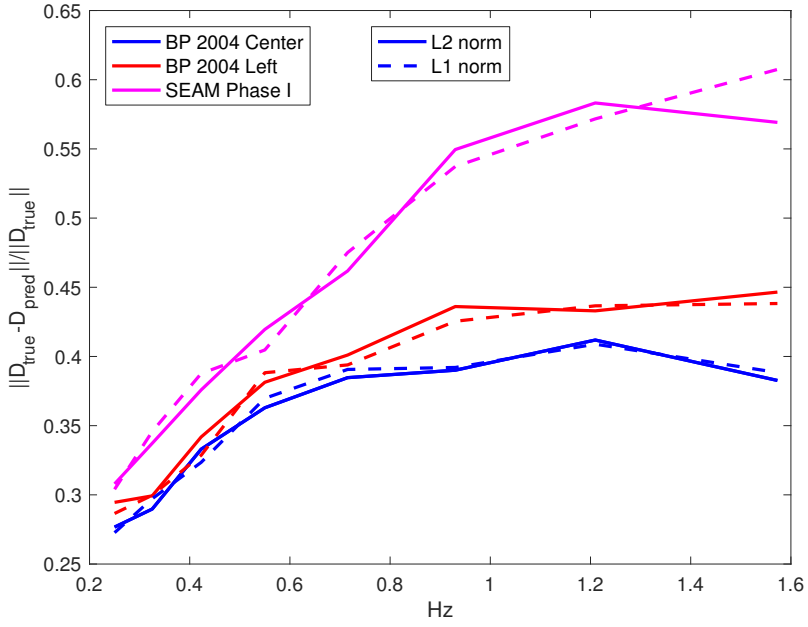
a

Figure 9: Architecture of a convolutional neural network designed for low-frequency extrapolation. There are four convolutional blocks followed by two fully-connected layers.

–



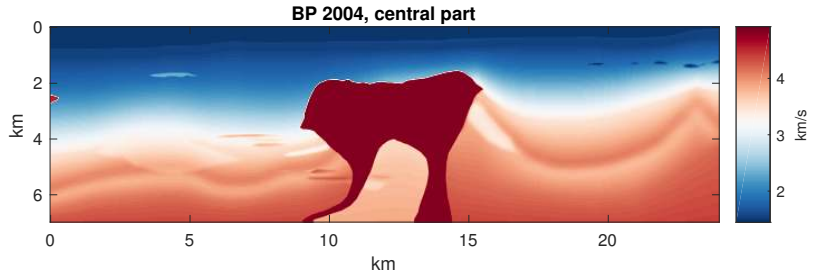
a



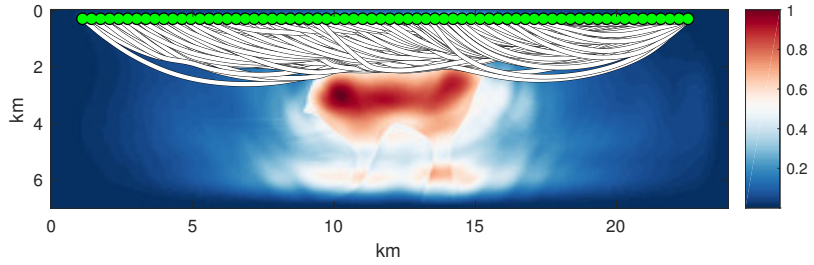
b

Figure 10: Comparison of misfits between target and inferred data for the networks trained on datasets built from 200 and 400 random velocity models (a), and optimization for L_1 and L_2 loss functions (b).

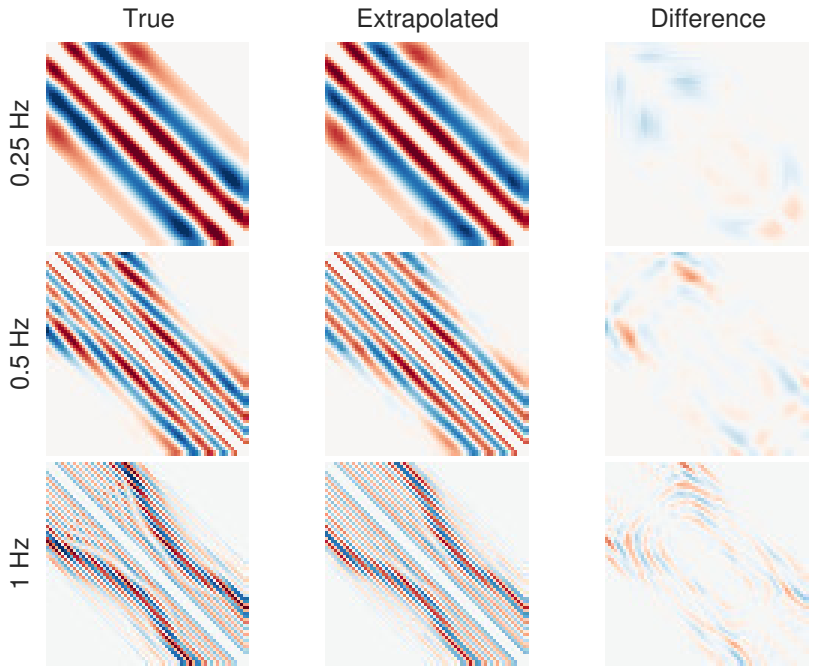
–



a

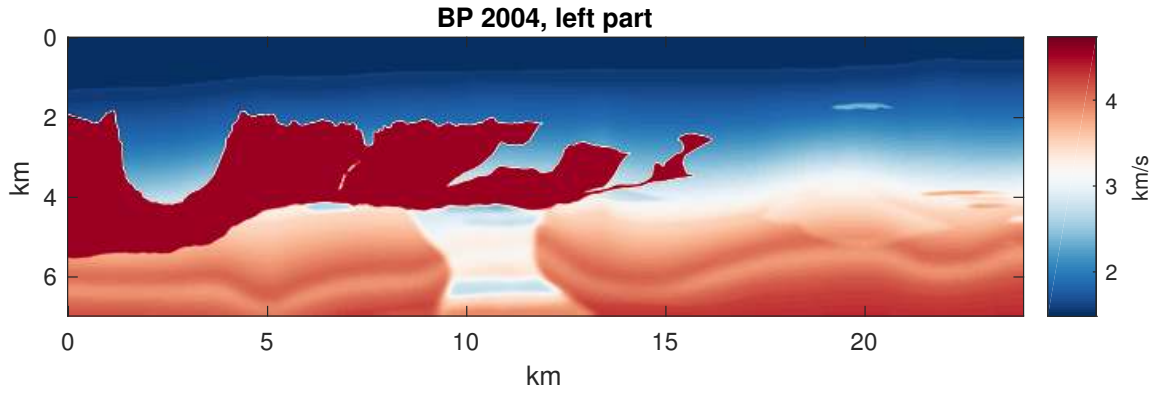


b

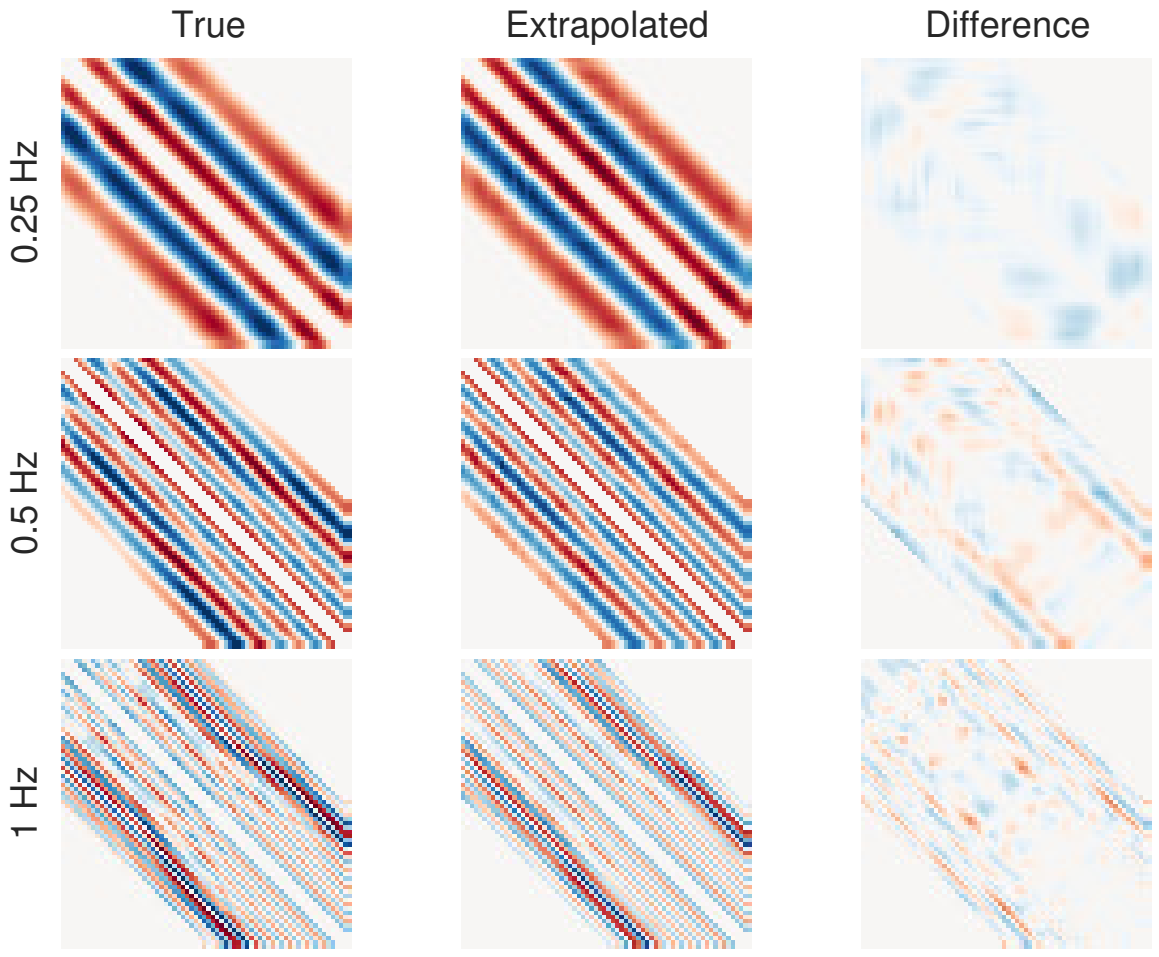


c

Figure 11: Central section from the BP 2004 benchmark velocity model (a). Normalized illumination in the model accounting for truncated half-offset acquisition. White lines indicate ray paths of diving wave, green points indicate locations of collocated sources and receivers (b). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (c).

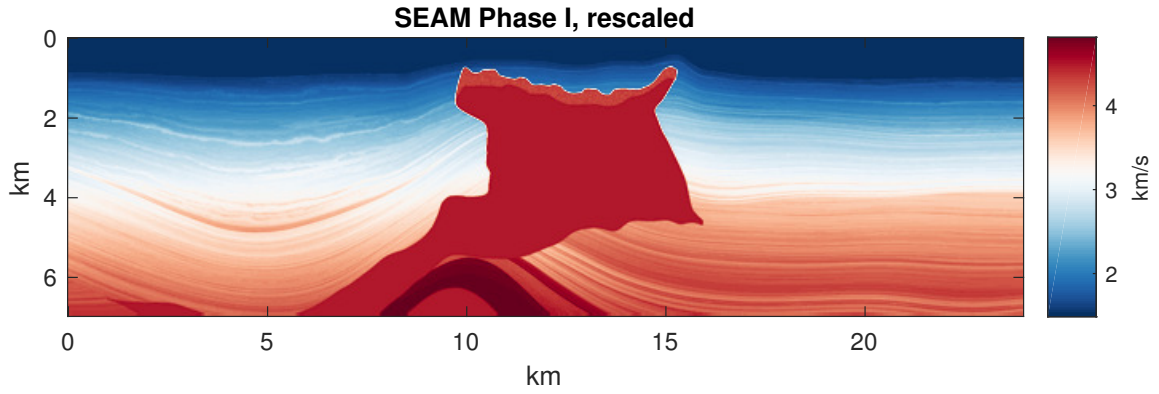


a

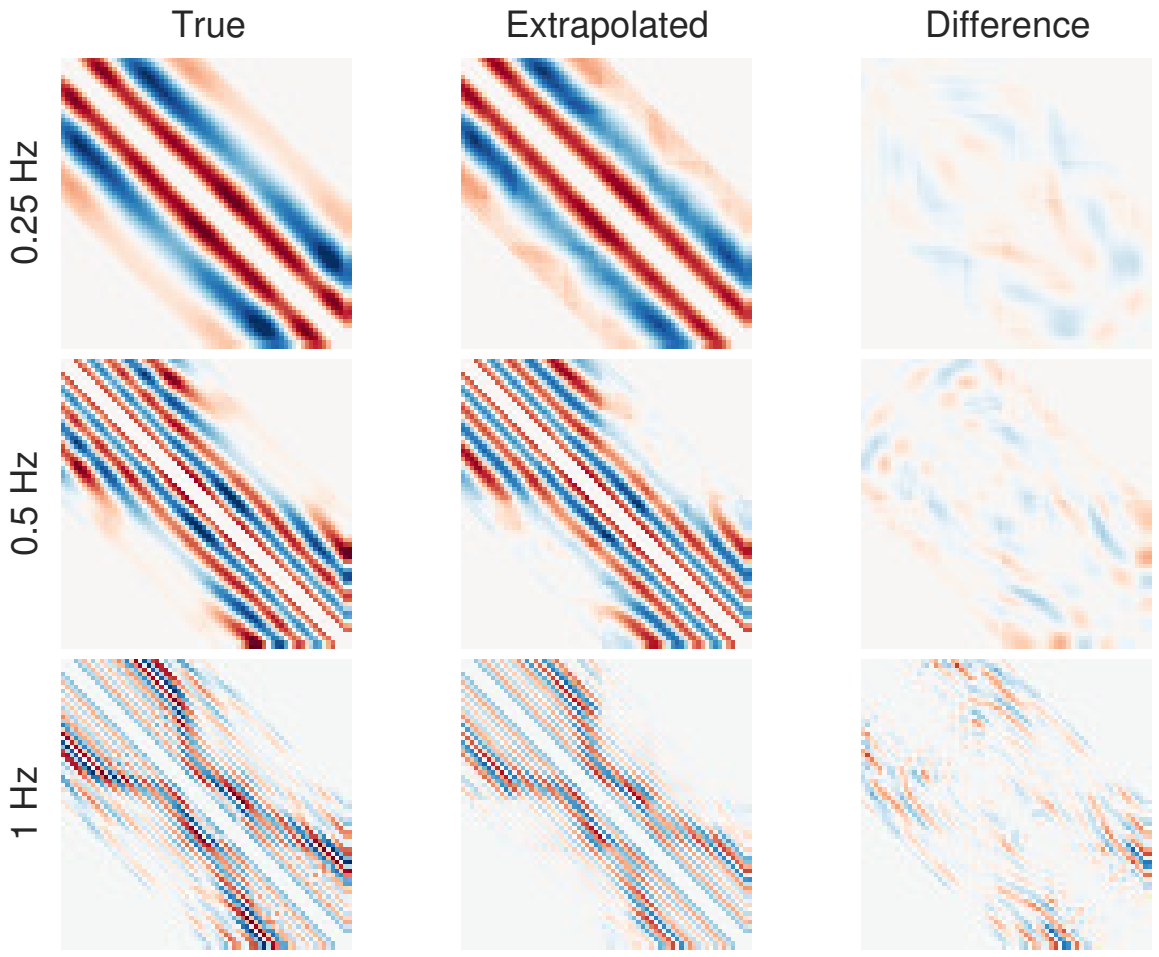


b

Figure 12: Left section from BP 2004 benchmark velocity model (a). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (b).



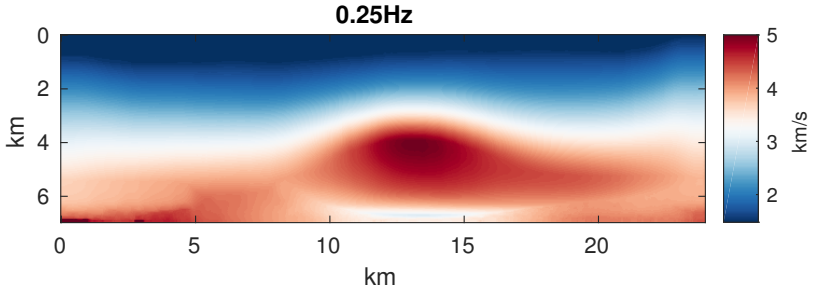
a



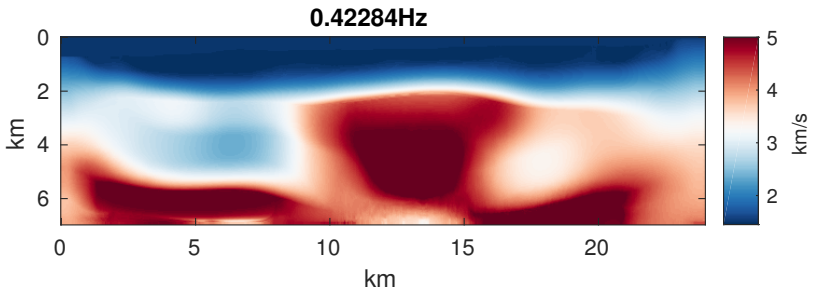
b

Figure 13: A section from SEAM Phase I benchmark velocity model (a). Real part of data matrices for extrapolated low-frequencies at 0.25, 0.5 and 1 Hz (c).

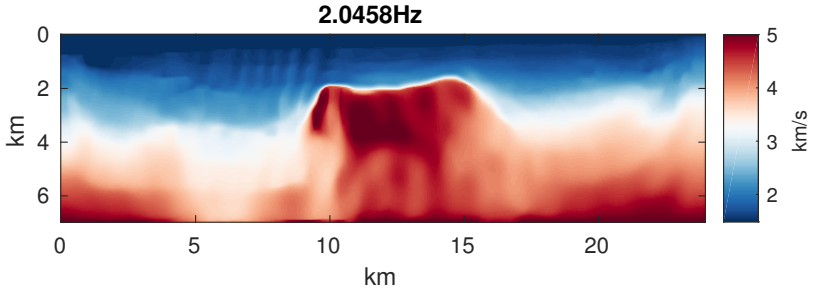
–



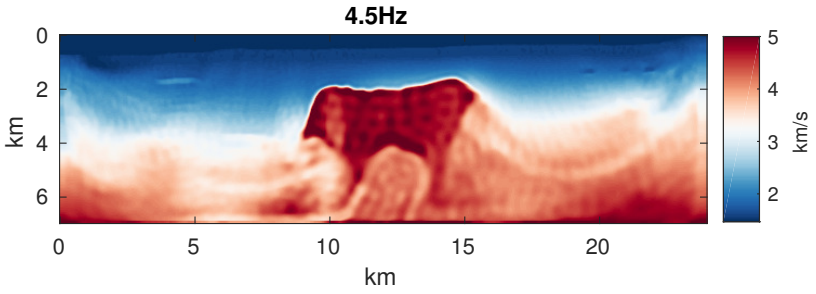
a



b



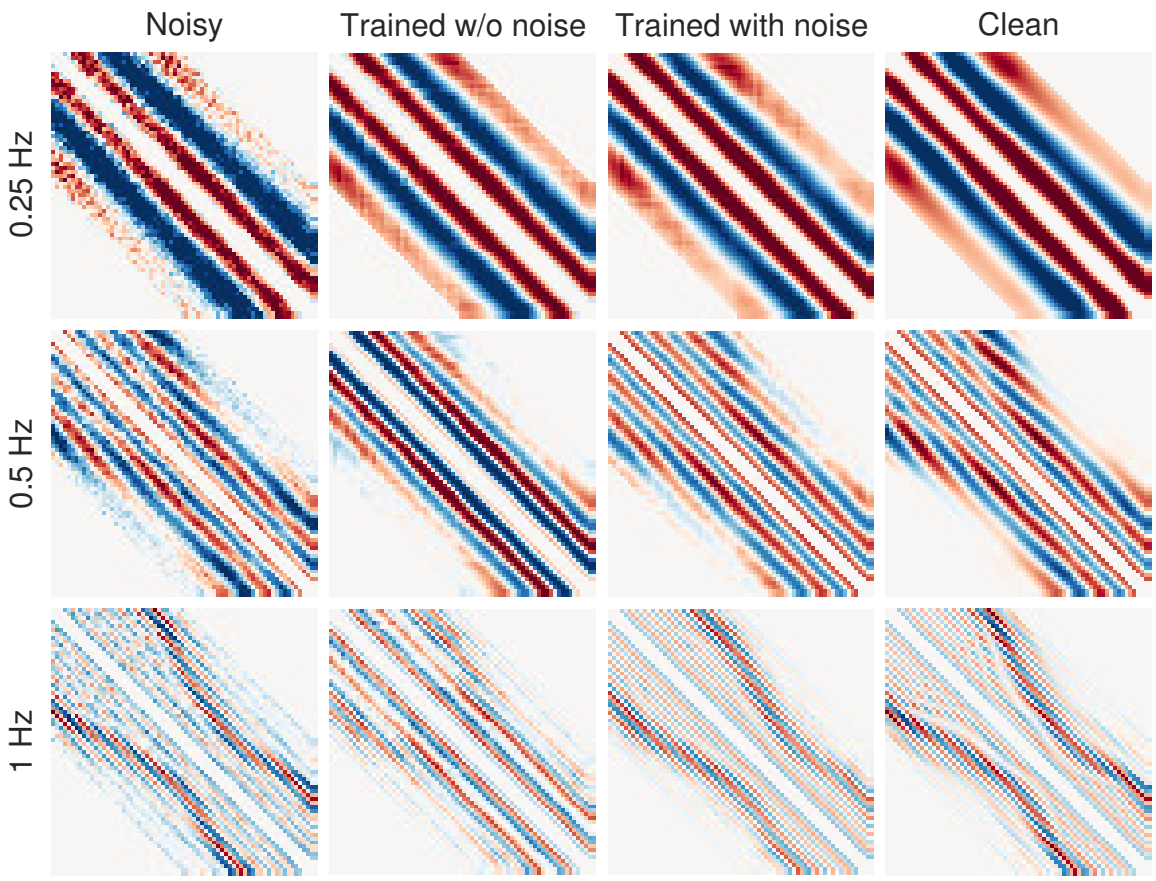
c



d

Figure 14: Multiscale full-waveform inversion of low-frequency data extrapolated by CNN (a-b) and data from the known interval (c-d).

–



a

Figure A-1: Extrapolation of low-frequency data for 0.25, 0.5 and 1 Hz from noisy high-frequency data within 2-4.5 Hz range, and for a SNR value of 7. The network was trained for two scenarios, being given input training data without noise and with noise. The later application delivers better inference result.