# Efficient seismic deghosting using convolutional neural networks

Khalid Almuteri & Paul Sava

*Center for Wave Phenomena and Dept. of Geophysics, Colorado School of Mines, Golden CO 80401*
*email kalmuteri@mymail.mines.edu*

**ABSTRACT**

Ghost reflections deteriorate the quality of seismic data in towed-streamer acquisitions. Various acquisition and processing solutions have been proposed to remove ghost reflections from seismic data. A common issue with the proposed solutions is their limited ability to remove source-side ghosts because of the sparse source sampling in towed-streamer seismic acquisitions. Additionally, satisfactory receiver-side deghosting solutions require complementary measurements, e.g., particle motion data, for wavefield separation. We propose a solution based on convolutional neural networks (CNNs) to remove source-side and receiver-side ghosts in the shot domain. The solution does not require complementary measurements, i.e. it can remove ghost reflections in conventional pressure data measured at a single reference surface. Our solution requires knowledge of the acquisition configuration to create training data that replicates the field acquisition geometry and requires the ocean floor bathymetry to be known. A CNN learns to map ghost-contaminated gathers to corresponding ghost-free gathers through an iterative training process. We show using a realistic synthetic example that the CNN-based deghosting operator can remove ghost reflections from previously unseen data. We demonstrate that the solution generalizes well to data generated using a velocity model intrinsically different from velocity models used to generate training data. Further, we show that our solution is robust to random noise, even though the training data are noise-free.

**Key words:** deghosting, convolutional neural networks, marine acquisition

## 1 INTRODUCTION

Ghost reflections are a pressing problem in towed-streamer seismic acquisition. They introduce frequency notches into the amplitude spectrum, limiting the usable data bandwidth (Amundsen and Zhou, 2013; Dondurur, 2018). This reduced bandwidth lowers the seismic resolution which is directly proportional to the bandwidth of the signal (Knapp, 1990). Further, ghost reflections distort the source signature by altering its phase, thus leading to erroneous impedance inversion (Jovanovich et al., 1983). Also, seismic data processing can benefit from ghost-free data, e.g., to improve normal-moveout velocity picking (Provenzano et al., 2020).

Ghost-free data are a prerequisite for many geophysical applications. Ghost reflections attenuate low frequencies essential for full waveform inversion and impedance inversion. The lack of sufficiently low frequencies makes full waveform inversion more susceptible to cycle-skipping (Bunks et al., 1995; Virieux and Operto, 2009). Low frequencies also help constrain and improve impedance inversion (Whitcombe and Hodgson, 2007; ten Kroode et al., 2013; Wang et al., 2017). Moreover, low frequencies are less prone to scattering (Bunks et al., 1995) and attenuation, making them ideal for subsalt imaging (Kapoor et al., 2005; ten Kroode et al., 2013). Conventionally, streamers are towed at a shallow depth to broaden the bandwidth of the seismic signal (Wang et al., 2017). However, shallow-tow attenuates low frequencies and yields data with a low signal-to-noise ratio (Kragh et al., 2010; He et al., 2013; Wang et al., 2017). A deep-tow has reciprocal effects; it allows for the acquisition of more low-frequency content and increases the signal-to-noise ratio, especially at the lower end of the spectrum, at the expense of a narrowband signal. Therefore, recovering the low frequencies attenuated by ghost reflections in conventional shallow-towed streamers is necessary for full waveform inversion, impedance inversion, and subsalt imaging.

There are three types of ghost reflections that co-exist in marine data: source-side ghosts, receiver-side ghosts, and source-receiver ghosts. A source-side ghost starts its propagation from a source as an up-going wave, whereas a receiver-side ghost ends its propagation at a receiver as a down-going wave. A source-receiver ghost is a combination of a source-side and receiver-side ghost. There is extensive literature with processing and acquisition solutions that address the ghost reflection problem, with each solution having its strengths and weaknesses. Deghosting solutions reported in the literature fall into two main categories: receiver-side solutions and source-side solutions. Both categories can be further classified into: acquisition-based and processing-based solutions, as discussed next.

Acquisition solutions for receiver-side deghosting include slanted streamer (Bearnth and Moore, 1989; Soubaras and Dowle, 2010), over/under streamer (Sønneland et al., 1986; Posthumus, 1993; Moldoveanu et al., 2007), shallow-over/deep-under streamer (Kragh et al., 2010), and dual-sensor streamer (Tenghamn et al., 2007; Carlson et al., 2007). In a slanted streamer acquisition, hydrophones vary in depth along the cable, which results in ghost notch diversity (Bearnth and Moore, 1989; Soubaras and Dowle, 2010). This diversity improves the bandwidth of poststack seismic data (Soubaras and Dowle, 2010; Provenzano et al., 2020), making the solution unsuitable for applications that require ghost-free prestack data or applications that require receivers to have the same frequency content. An over/under streamer, at which pressure data are acquired at two different reference surfaces, enables removing receiver-side ghosts through an upgoing/downgoing wavefield separation process (Sønneland et al., 1986; Posthumus, 1993; Moldoveanu et al., 2007). However, it requires twice as many streamers with accurate positioning of vertically-aligned cables (Reilly, 2016). Such requirements deem the solution complex to implement in the field. Kragh et al. (2010) propose a shallow-over with a sparse deep-under acquisition configuration. Shallow streamers contribute the mid- and high-frequency content to the seismic data, whereas deep streamers contribute the low-frequency content (Kragh et al., 2010). A shallow-over/deep-under acquisition requires fewer streamers than an over/under acquisition. This is because deep streamers are sparse since they contribute low frequencies only. However, a shallow-over/deep-under acquisition still requires accurate positioning of the cables and a proper interpolation of the low-frequency data to compensate for the sparse acquisition.

Dual-sensor streamer with pressure and particle motion sensors (Tenghamn et al., 2007; Carlson et al., 2007) exploit the fact that upgoing/downgoing wavefield separation is possible if pressure and particle velocity data are measured simultaneously (Claerbout, 1976; Amundsen, 1993; Day et al., 2013; Ikelle and Amundsen, 2018). However, the particle motion sensor is sensitive to mechanical noise below 20 Hz (Tenghamn et al., 2007; Carlson et al., 2007; Day et al., 2013). To overcome this issue, one needs to estimate particle motion data below 20 Hz from pressure data. Estimating such data requires computing the direct arrival and its ghost if the source is shallower than the streamer (Amundsen, 1993; Amundsen et al., 1995; Day et al., 2013; Ikelle and Amundsen, 2018). Computing direct arrival and its ghost require knowledge of the source function, which can be difficult to find. Moreover, data acquired with different sensor types requires calibration with an unknown factor (Grobbe et al., 2016). Such calibration can be carried using a deterministic approach (Day et al., 2013) or a statistical approach (Cambois et al., 2009; Alexandrov et al., 2014). Following a deterministic approach requires knowing the density and propagation velocity at the receivers (Day et al., 2013), and the receivers impulse responses to compute the scaling factor (Cambois et al., 2009). Using a statistical approach formulates the calibration process as a minimization problem (Cambois et al., 2009; Alexandrov et al., 2014). Despite its great potential in solving the ghost challenge, dual-sensor acquisition comes with an involved processing workflow, which is difficult to implement.

One can extend the acquisition solutions for receiver-side deghosting to address source-side ghosts. Haavik and Landrø (2015) propose varying the source depth during acquisition to diversify frequency notches in the amplitude spectra of different shot gathers. This improves the bandwidth of the seismic signal poststack in a way similar to slanted-streamer acquisition. An over/under source configuration, where data are acquired twice at the same inline location for sources at different depths (Moldoveanu, 2000), employs reciprocity to remove source-side ghosts in common-receiver gathers (CRGs) (Moldoveanu et al., 2007; Egan et al., 2007). However, source-side deghosting in the CRG domain requires dense shot sampling that towed-streamer acquisition generally lacks.

Various researchers have proposed processing solutions for receiver-side deghosting. Soubaras (2010) proposes an imaging solution to deghost and improve seismic data poststack. The method works by jointly deconvolving two migration images: a conventional migration image and a mirror migration image. Amundsen and Zhou (2013) derive a low-frequency deghosting filter that can improve the data at least up to half the frequency of the second ghost notch. The method is based on a truncated approximation of the ghost model. The solution can be applied in the time-space domain on a trace-by-trace basis (Amundsen and Zhou, 2013; Wang et al., 2017). However, it is only applicable to data acquired with horizontal streamers or streamers with mild variations in depth (Wang et al., 2017). Beasley et al. (2013a,b) propose a physics-based solution to separate single-component measurements, e.g., pressure data acquired at a single reference surface, into upgoing and downgoing wavefields using the wave equation. The wavefield separation process is iterative and employs the causality principle. The solution, however, is sensitive to the presence of noise in the data. Processing solutions for receiver-side deghosting are applicable to source-side ghosts in the CRG domain. However,

source-side deghosting is more challenging than receiver-side deghosting and remains mainly an unsolved problem (Amundsen et al., 2017; Ikelle and Amundsen, 2018), because shots are sparsely acquired.

In this paper, we propose seismic deghosting using convolutional neural networks (CNNs), which is a special type of artificial neural networks (ANNs). Our solution does not require any particular acquisition configuration, but requires that its geometry and the bathymetry of the ocean floor be known. It can remove source-side and receiver-side ghosts simultaneously in the shot gather domain. This feature offers an opportunity to completely deghost conventional pressure data acquired at a single reference surface. To achieve our objective, we develop a CNN that functions as a deghosting operator that takes ghost-contaminated data as its input and outputs a ghost-free prediction. In the following sections, we provide the theoretical foundation of CNNs, and ANNs in general. We explain how they work and discuss their benefits and pitfalls. Then we demonstrate how CNNs can remove ghost reflections using a synthetic example inspired by a realistic acquisition scenario.

## 2 CNN-BASED DEGHOSTING

### 2.1 Background

ANNs are powerful machine learning (ML) models inspired by biological neurons found in the brain (Russell and Norvig, 2016; Goodfellow et al., 2016; Géron, 2019). Neural networks are capable of tackling complex problems from various fields such as natural language processing (Sutskever et al., 2014), object recognition (Krizhevsky et al., 2017), and speech recognition (Hinton et al., 2012) with a great degree of success (LeCun et al., 2015). The interest of the geophysics community in ANNs is not new (Van der Baan and Jutten, 2000), but recent advances in ML renewed the interest in the field. Geophysical applications include velocity model building (Araya-Polo et al., 2018; Wu and McMechan, 2018; Yang and Ma, 2019), normal-moveout velocity picking (Ma et al., 2018; Biswas et al., 2018), petrophysical properties and impedance inversion (Biswas et al., 2019; Das et al., 2019; Das and Mukerji, 2020), and seismic deblending (Sun et al., 2020).

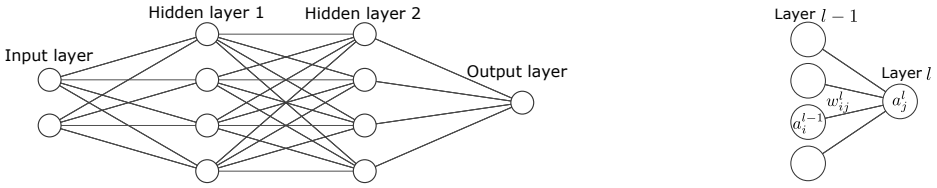
Various types of ANNs exist, with fully-connected NN, which consists of layers of neurons stacked together as shown in Figure 1a, being the simplest. The links that connect different neurons represent learnable parameters known as weights. One can find such weights through an iterative data-fitting procedure called training. The weights map inputs to outputs according to

$$a_j^l = g\left(\sum_{i=0}^{n} w_{ij}^l a_i^{l-1}\right), \tag{1}$$

where $a_j^l$ is the value at the $j$th neuron at the $l$th layer, $w_{ij}^l$ is the weight that connects the $i$th neuron at the $(l-1)$th layer with the $j$th neuron at the $l$th layer, $n$ is the number of neurons in the $(l-1)$th layer, and $g$ is a nonlinear activation function. Figure 1b gives a visual representation of equation 1. The activation function in the NN is critical; a linear activation function allows only linear mapping between inputs and outputs, while a nonlinear activation function enables the NN to learn nonlinear mappings.

A fully-connected NN suffers from multiple issues that limit its usefulness in some applications. One such issue is fixed input and output size (LeCun et al., 1998, 2015; Goodfellow et al., 2016). Processing seismic data using NNs requires them to have a structure that can intrinsically handle data with variable recording lengths and variable number of traces. Also, the memory requirement of fully-connected NNs with a large input size can be prohibitively expensive (LeCun et al., 1998; Goodfellow et al., 2016; Géron, 2019). Additionally, as the input size to a NN increases, the number of learnable parameters increases as well. In such cases, one needs to use more training examples to avoid overfitting (LeCun et al., 1998, 2015; Géron, 2019). Fully-connected NNs are not robust to any translation of features (e.g., faults in a stacked section or first-break in a shot gather) in the inputs. This implies that if a fully-connected NN is trained, for example, to detect faults in a stacked section, it will only detect faults in locations where they existed in training examples. LeCun et al. (1998) and Goodfellow et al. (2016) discuss in depth the limitations of fully-connected layers and point out the advantages of using CNNs as an alternative. In CNNs, convolutional filters replace the weights characterizing a fully-connected layer. CNNs generalize better than fully-connected NNs, i.e. they outperform fully-connected NNs on previously unseen data (LeCun et al., 2015). Also, CNNs have fewer parameters than fully-connected NNs, making them ideal for complex and large-scale problems.

In the framework of ANNs, a supervised learning approach is the most common form of learning (LeCun et al., 2015). The goal of supervised learning is to find the optimal weights that minimize an objective function (e.g., least-squares) given a set of $N$ training examples, where the training examples come in input-output pairs: $(\mathbf{x}_i, \mathbf{y}_i)$, $i = 1, \ldots, N$. The relationship between the input and output is given by an unknown function $f$ such that $\mathbf{y}_i = f(\mathbf{x}_i)$. The goal of training is to find a function $h_{\mathbf{w}}$, which is represented by an ANN and parameterized by the learnable weights ($\mathbf{w}$), that approximates the function $f$. In our work, the input ($\mathbf{x}_i$) is a ghost-contaminated shot gather, the output ($\mathbf{y}_i$) is a true ghost-free gather, and $h_{\mathbf{w}}$ is the deghosting operator. We arrive at




(a) (b)

**Figure 1.** (a) An example of a fully-connected NN with two input units, two hidden layers with four units in each layer, and one output unit. (b) A simple model of a neuron. The neuron's output is $a_j^l = g\left(\sum_{i=0}^{n} w_{ij}^l a_i^{l-1}\right)$, where $a_j^l$ is the value at the $j$th neuron at the $l$th layer, $w_{ij}^l$ is the weight that connects the $i$th neuron at the $(l-1)$th layer with the $j$th neuron at the $l$th layer, $n$ is the number of neurons in the $l-1$th layer, and $g$ is a nonlinear activation function.

function $h_{\mathbf{w}}$ by minimizing the mean-squared-error objective function, given by

$$E(\mathbf{w}) = \frac{1}{N}\sum_{i=1}^{N} \left\|\mathbf{y}_i - h_{\mathbf{w}}(\mathbf{x}_i)\right\|^2, \tag{2}$$

with respect to the weights ($\mathbf{w}$) of the NN. A common minimization method used in ANNs is the gradient descent, which iteratively updates the weights using the gradient of the objective function

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \alpha \frac{\partial E(\mathbf{w}_k)}{\partial \mathbf{w}_k}, \tag{3}$$

where $\alpha$ is the learning rate. One should select the optimal learning rate experimentally; a small learning rate results in slow convergence, while a large learning potentially rate results in divergence (Géron, 2019).

ANNs are useful and powerful machine learning models. According to the universal approximation theorem (Cybenko, 1989; Hornik et al., 1989; Hornik, 1991), ANNs are universal approximators, that is, ANNs with a single hidden layer can approximate any continuous function up to an arbitrary order of accuracy. ANNs with two hidden layers, on the other hand, can approximate any function with an arbitrarily small error (Lapedes and Farber, 1987). An exact mapping of input-output pairs is achievable under some constraints (Huang and Huang, 1990; Sartori and Antsaklis, 1991; Tamura and Tateishi, 1997). NNs with a single hidden layer can still approximate piecewise continuous functions given knowledge about the discontinuities (Llanas et al., 2008). However, NNs with a single hidden layer are impractical for two reasons. First, multi-layer NNs can approximate functions with fewer learnable parameters by using significantly fewer neurons in each layer (Tamura and Tateishi, 1997; Eldan and Shamir, 2016; Goodfellow et al., 2016). Second, multi-layer NNs generalize better than single layer NNs (Goodfellow et al., 2016). Another important property of ANNs is their ability to simultaneously approximate functions and their derivatives (Pinkus, 1999), which is critical for physics-informed NNs (e.g., Raissi et al., 2019 and Moseley et al., 2020), because training a physics-informed NN to approximate a solution to a partial differential equation (PDE) involves taking derivatives of the solution with respect to the inputs of the NN inside the objective function.

The ability to approximate any continuous function arbitrarily well is not unique to ANNs. One can approximate any continuous function up to an arbitrary order of accuracy with polynomials (including trigonometric polynomials) as well (Jeffreys and Jeffreys, 1950; Carothers, 1998; Stein and Shakarchi, 2011). The advantage of using ANNs as opposed to polynomials for approximation is merely computational. The number of coefficients one needs to solve for in polynomial regression makes it prohibitive to approximate high-dimensional functions. The exact number of polynomial coefficients is given by

$$\frac{(n+d)!}{(n!)(d!)}, \tag{4}$$

where $n$ is the number of dimensions and $d$ is the degree of the polynomial (Géron, 2019). For trigonometric polynomials it is the computational complexity rather than the number of coefficients that makes such an approach infeasible. The prohibitive cost of approximating high-dimensional functions, which includes the number of training examples one needs to avoid overfitting, is commonly known as the "curse of dimensionality" (Hanka and Harte, 1997; Griebel, 2005; Géron, 2019; Grohs et al., 2019; E, 2020). The ability of ANNs to approximate any function, including high-dimensional nonlinear functions, make them useful and gives rise to a wide range of applications that would not be possible otherwise.

The universal approximation theorem (Cybenko, 1989; Hornik et al., 1989; Hornik, 1991) only addresses the ability of ANNs to represent continuous functions and does not consider the learnability aspects of the problem. The existence of an ANN that can approximate an unknown function does not imply one can find it. Further, one cannot characterize which functions can and cannot be approximated by a particular ANN structure (Russell and Norvig, 2016). An ANN architecture is given by a set of parameters

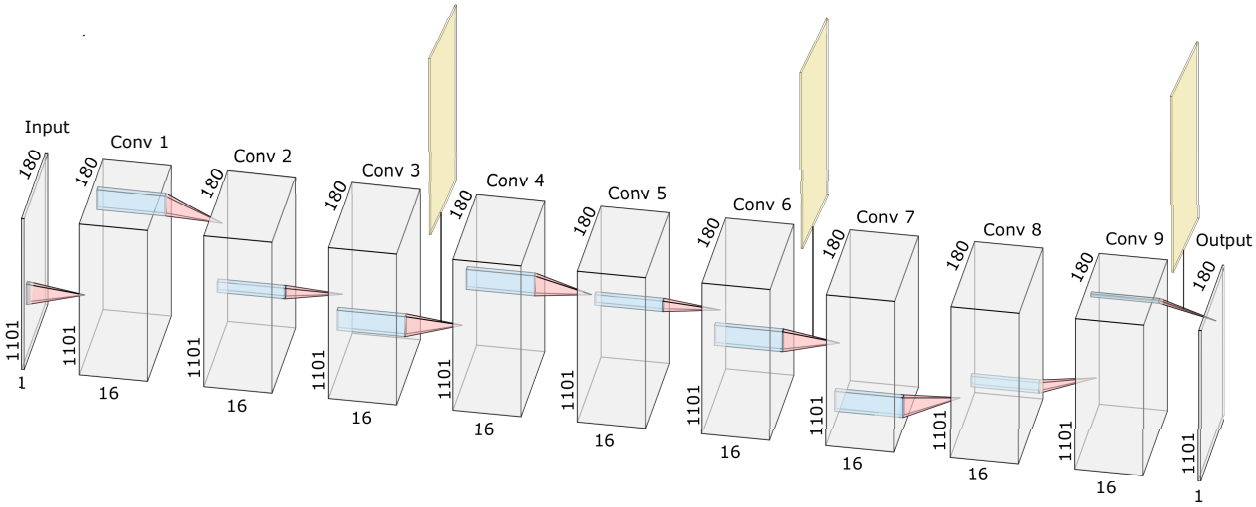



**Figure 2.** An illustration of the architecture of the network we use for seismic deghosting on 2D data. Each gray cuboid labeled with Conv represents a multi-channel feature map generated using N convolutional filters ($N = 16$). The height of each cuboid corresponds to the number of time samples ($n_t = 1101$) and the width corresponds to the number of traces in the gather ($n_r = 180$). Each yellow cuboid represents a temporal masking layer in the network.

(e.g., number of layers, learning rate, and the type of activation functions) commonly known as hyperparameters. One has to fine-tune the hyperparameters, i.e. to experimentally search for the best set of parameters to solve a given problem with adequate accuracy (Goodfellow et al., 2016; Géron, 2019). In the context of seismic deghosting, we make use of the ability of ANNs to exploit any intrinsic relationship between inputs and outputs to construct a deghosting operator. Such an ANN-based operator takes time to train, but after training it can predict ghost-free data in a negligible amount of time. However, to construct the deghosting operator, we must first search for the best set of parameters. The hyperparameters that define the neural network used in our method are: the number of layers, the number of filters, the size of each filter, the learning rate, and the activation function. Our approach to selecting these parameters is given in the following section with motivations to our specific choices.

### 2.2 Neural network architecture

Our neural network consists of three convolutional layers followed by a masking layer, repeated three times for a total of 12 layers (Figure 2). Each convolutional layer has 16 filters. The first two convolutional layers in the sequence use filters with a temporal length of 101 coefficients and a spatial width of three traces in the inline and crossline directions. The third convolutional layer in the sequence has filters with 11 temporal coefficients and identical spatial width as the preceding two layers. We use tanh as activation function throughout our NN because it is centered around zero and bounded between $-1$ and 1.

In CNNs, it is common to use filters with a small size (e.g., $3\times3$) because they usually outperform large filters and require fewer computations (Géron, 2019). Given that our objective is to remove ghost reflections in time domain, we use temporally long filters, of 101 samples, to reconstruct the low-frequency information of the data. In our numerical example, a filter with 101 samples cover six times the dominant period of the data. Smaller filters can achieve the same goal if one uses many more layers. We use temporally short filters for every third convolutional layer, of 11 samples, to reconstruct the high-frequency information. By choosing both temporally long and short filters, we give the CNN freedom to work at different scales. Our choices of filter sizes are experimental because a theoretically rigorous approach to select such a parameter remains an open problem in machine learning. As reported in the literature, optimizing such hyperparameters is an experimental process (Goodfellow et al., 2016; Géron, 2019).

Deghosting can be performed on individual traces using 1D filters or can utilize information from adjacent traces using 3D filters. 3D filters should outperform 1D filters because ghost reflections have offset dependency, which can be observed using the 3D ghost function in the Fourier domain (Amundsen, 1993) given by

$$G(k_z, z) = 1 + re^{2ik_z z}, \tag{5}$$

where $k_z = \sqrt{k^2 - k_x^2 - k_y^2}$ is the vertical wavenumber, $k = \frac{\omega}{c}$, is the wavenumber, $\omega = 2\pi f$ is the angular frequency, $c$ is the water velocity, $k_x$ and $k_y$ are the horizontal wavenumbers in the inline and crossline directions, $z$ is the source or receiver depth, and $r$ is the reflection coefficient at the sea surface.

We update the weights $(\mathbf{w})$ using the Nesterov's Accelerated Gradient (NAG) method (Sutskever et al., 2013; Géron, 2019), which is a modified version of the gradient decent (Equation 3)

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mathbf{m}_{k+1}, \tag{6}$$

where

$$\mathbf{m}_{k+1} = \beta \mathbf{m}_k - \alpha \frac{\partial E(\mathbf{w}_k + \beta \mathbf{m}_k)}{\partial \mathbf{w}_k} \tag{7}$$

is the momentum $(\mathbf{m}_0 = 0)$, $\beta \in [0, 1]$ is the momentum coefficient, and $\alpha$ is the learning rate. The momentum coefficient $(\beta)$ defines how much previous gradients contribute towards new updates, and if the momentum coefficient is zero, the method reduces to gradient descent. During training we use $\beta = 0.9$, i.e. we put much weight on previous directions to speed up the convergence rate. To further speed up convergence, the NAG method computes the gradient of the objective function not at the current location of weights $(\mathbf{w})$ but at the location of weights with a slide in the momentum direction $(\mathbf{w} + \beta \mathbf{m})$. We experimentally determine that the optimum learning rate to train the NN is $\alpha = 0.1$.

The masking layers in our CNN are essential for two closely related reasons: (1) to eliminate any noise the NN produces before any first-arrival; (2) to help the NN converge to a solution during training. Without masking layers, training fails because the NN tries to fit the data by removing noise the CNN generates and by removing ghost reflections. To create a masking layer, we use the Modified Coppens's method (Sabbione and Velis, 2010) to pick the first seismic signal of any kind (whether it is a water bottom reflection or refraction from a subsequent layer) along the time axis. The masking array defines the regions where a seismic signal could exist, and it accordingly mutes irrelevant signal, i.e. not generated by the source before the first-arrival.

## 3 NUMERICAL EXAMPLE

In this section, we demonstrate CNN ghost removal in the shot domain on 2D data. First, we describe the process of generating the training and testing data. Then, we illustrate the performance of the CNN-based deghosting operator using a synthetic example.

### 3.1 Neural network training

To train our CNN, we use 128 input-output pairs of ghost-free and ghost-contaminated shot gathers. For this purpose, we randomly extract from the Marmousi model (Martin et al., 2002) 117 sub-models (Figure 3a) and semi-randomly extract from the Sigsbee model (Paffenholz et al., 2002) 11 additional sub-models (Figure 3b). We use this semi-random approach in order to avoid salt-dominated models which would have only a few sparse seismic interfaces. To test our method, we extract a sub-model (Figures 4) from the Amoco statics test model (O'Brien, 1994). We add a water layer with 0.35 km depth to the top of all training and testing models. Each sub-model is 1.5 km wide and 1.41 km deep to lower the computational cost of generating data.

For each velocity model, we generate two shot gathers, with and without ghosts. In all cases, we use a primary source at $z = 0.025$ km and $x = 0.303$ km, and a mirror source for source-side ghost modeling at $z = -0.025$ km with the same lateral position of the primary source. We use 180 receivers at $z = 0.030$ km with a spacing of 0.005 km. Also, we use 180 mirror receivers for receiver-side ghost modeling placed at $z = -0.030$ km. We use the same source wavelet with a flat spectrum between $1 - 50$ Hz to simulate the training and testing data. We model seismic data for 2.2 s and a sampling interval of 0.2 ms, later resampled to 2 ms. Before training the CNN, we normalize the data to the range of the chosen activation function.

### 3.2 Seismic deghosting

To test the CNN-based deghosting operator, we use a sub-model with velocity values not within the range of the velocity values of the training sub-models. We choose this testing model to show that with simpler training models, i.e. training models with a narrow velocity range, one can remove ghost reflections from more complex models and lower the computational burden of generating training data simultaneously. By complex models, we mean models with stronger velocity contrasts and velocity values outside the range of that of the training models.

Figures 5a and 5b show the true ghost-free and ghost-contaminated data generated using the model shown in Figure 4, respectively. The characteristics of the waveform of the source signature differ significantly between the two gathers. The waveform in the ghost-contaminated data has a reverberatory character that lowers the resolution of the data. This character is not observed in the true ghost-free data; the waveform in the ghost-free data is more compact than the waveform in the ghost-contaminated data. The average amplitude spectra of the two gathers indicate that ghost reflections destructively interfere with the primary signal at 0 Hz

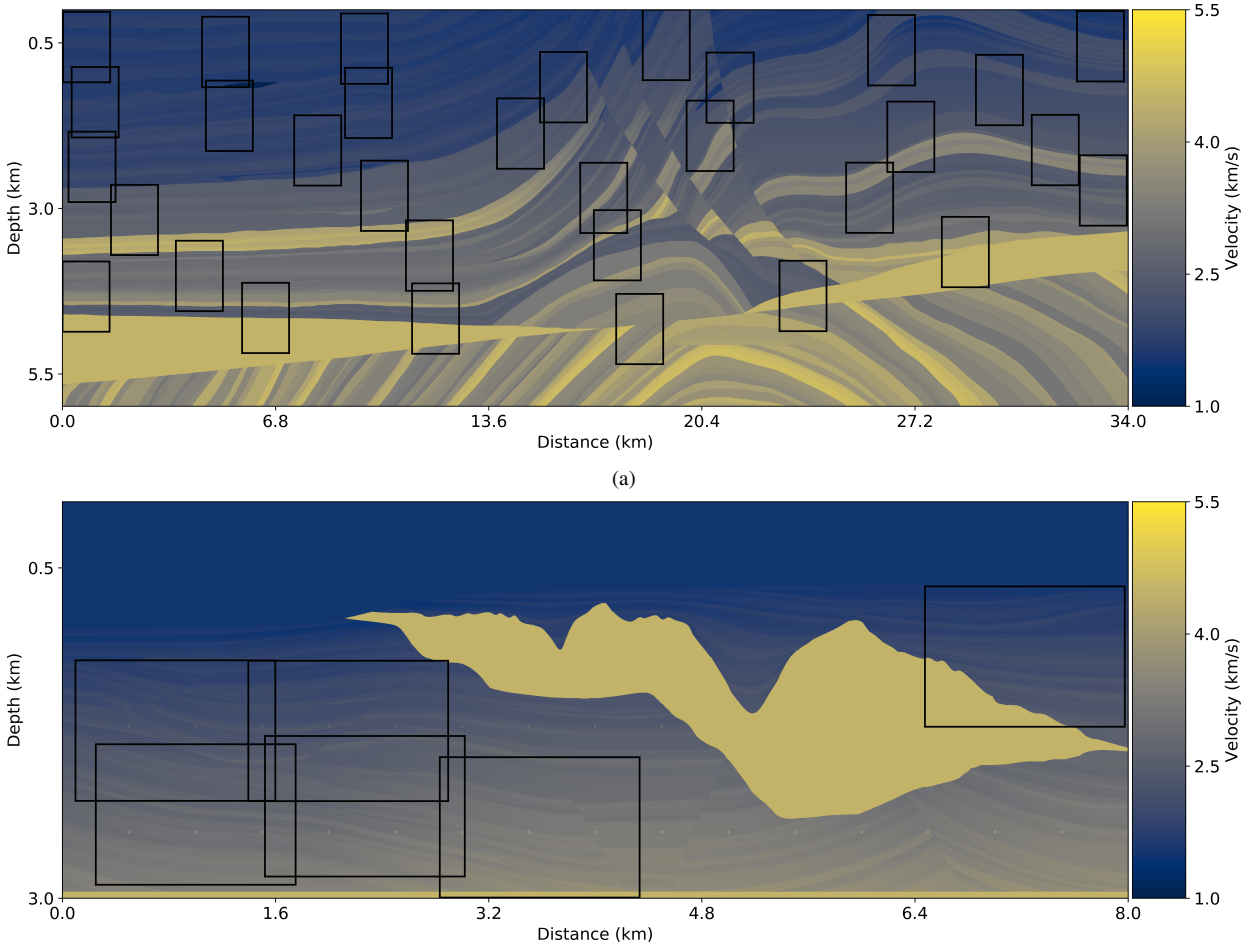



**Figure 3.** (a) Marmousi model showing only 32 sub-models used for training. (b) Sigsbee model showing only 6 sub-models used for training.

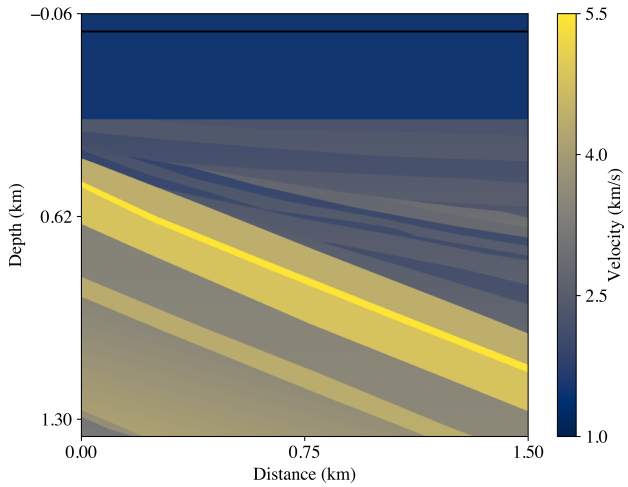



**Figure 4.** Sub-model extracted from the Amoco statics test model to test the CNN. The velocity values of this model are not within the range of the velocity values of the training models. We add a water layer with 0.35 km depth to the top. The black line corresponds to the air-water interface.

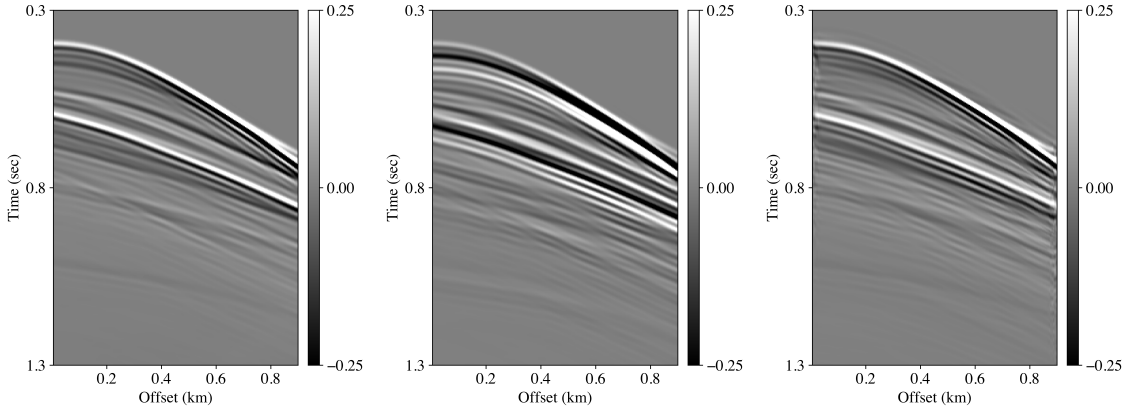



**Figure 5.** Comparison between (a) the true ghost-free data, (b) the ghost-contaminated data, and (c) the ghost-free prediction.

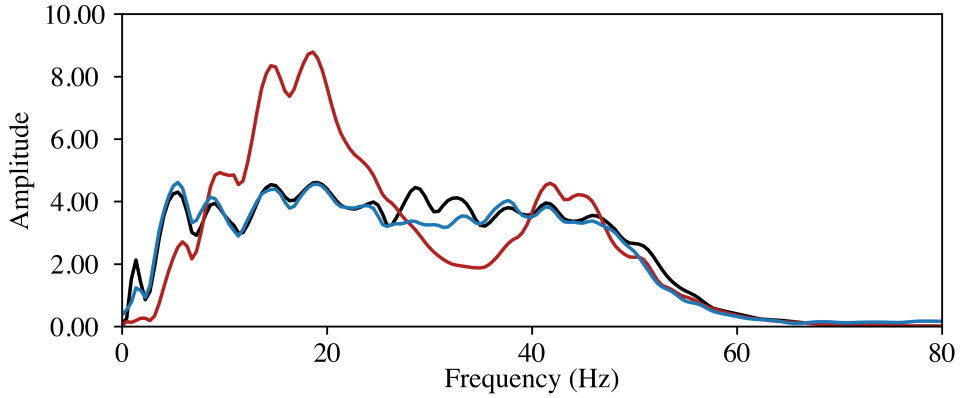



**Figure 6.** Comparison between the average amplitude spectra of the ghost-free data (black line) shown in Fig 5a, the ghost-contaminated data (red line) shown in Fig 5b, and the ghost-free prediction data (blue line) shown in Fig 5c.

and 34 Hz, and constructively interfere at 16 Hz (Figure 6). The destructive interference is more significant at low frequencies than at high frequencies. The amount of energy loss at 1 Hz is 21 dB, whereas the amount of energy loss at 34 Hz is 6 dB.

Using the ghost-contaminated gather (Figure 5b) as an input to the CNN-based deghosting operator gives the ghost-free prediction shown in Figure 5c. The true and predicted ghost-free gathers closely match in character. The average amplitude spectra (Figure 6) indicate a nearly perfect reconstruction. The CNN recovers the spectrum correctly within the constructive interference frequency bands but less accurately within the destructive interference frequency bands.

Individual traces from the true ghost-free, the ghost-contaminated, and the ghost-free prediction gathers for a receiver 0.1 km away from the source (Figure 7a) confirm the ability of the CNN to remove ghost reflections. The correlation coefficient between the true ghost-free and ghost-contaminated traces is 0.45, indicating that ghost reflections severely alter the waveform of the source signature. On the other hand, the correlation coefficient between the true and predicted ghost-free traces is 0.92, indicating that the CNN-based deghosting operator can recover with a great degree of accuracy both the amplitude (Figure 7b) and the phase (Figure 7c) of the true ghost-free data. As expected, stronger amplitudes are better fit than weaker amplitudes because stronger amplitudes contribute the most to the objective function. Therefore, removing direct arrival and its ghost improves the ghost-free prediction of reflection data.

Figure 8 shows the true ghost-free, the ghost-contaminated, and the ghost-free prediction gathers after applying a 5 Hz low-pass filter to the gathers shown in Figure 5. The ghost-contaminated gather is characterized by weaker amplitudes, indicating that ghost reflections severely reduce the signal-to-noise ratio at low frequencies (Figure 9). The proposed method can recover the attenuated low frequencies with a great degree of accuracy, improving the signal-to-noise ratio at this frequency band. Low-frequency traces

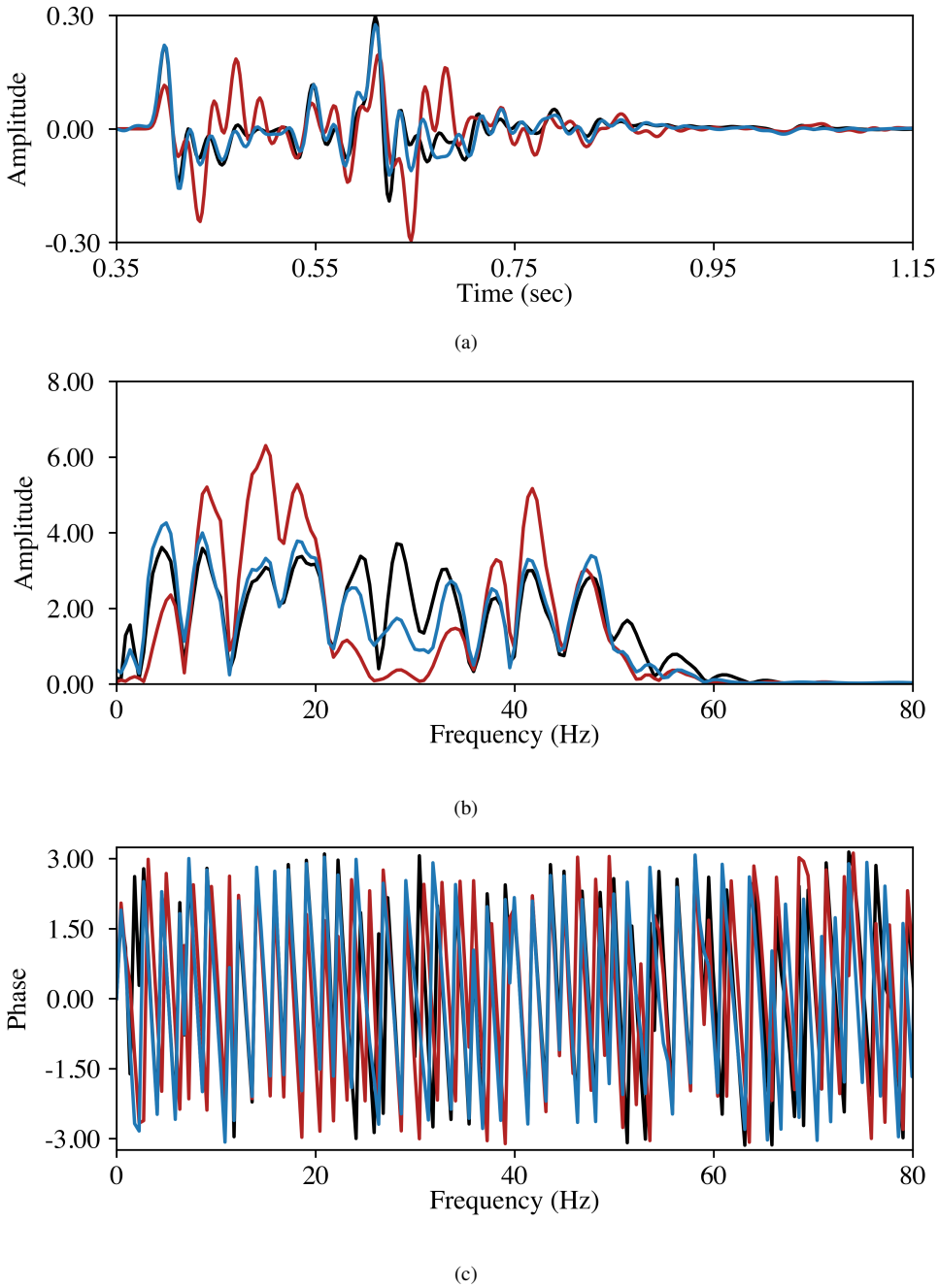



**Figure 7.** (a) Comparison between ghost-free (black line), ghost-contaminated (red line), and ghost-free prediction (blue line) traces recorded at a receiver that is 0.1 km away from the source. (b) Comparison between the amplitude spectra of the shown traces. (c) Comparison between the phase spectra of the shown traces.

from the true ghost-free, the ghost-contaminated, and the ghost-free prediction gathers for a receiver 0.1 km away from the source are shown in Figure 10a. Figures 10b and 10c show the corresponding amplitude and phase spectra, respectively. Although the reconstructed trace matches well the true ghost-free trace in both amplitude and phase, the CNN-deghosting operator recovers the phase of the signal more accurately than its amplitude. The low-frequency ghost-free prediction is nearly noise-free, indicating that noise introduced by the CNN is mainly at intermediate and high frequencies.

The presence of noise in field data is inevitable. To examine the robustness of the CNN-based deghosting operator to noise,

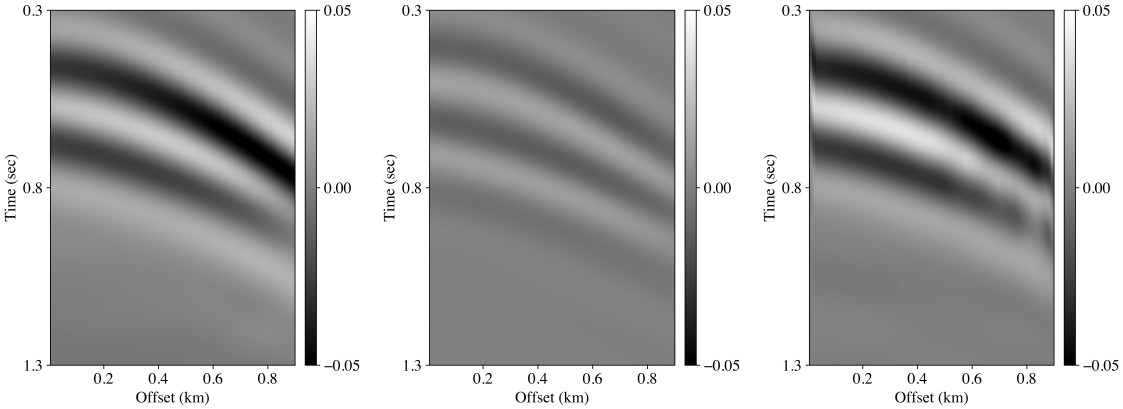



**Figure 8.** 5 Hz low-pass filter applied to (a) the ghost-free gather shown in Fig 5a, (b) the ghost-contaminated gather shown in Fig 5b, and (c) the ghost-free prediction gather shown in Fig 5c.

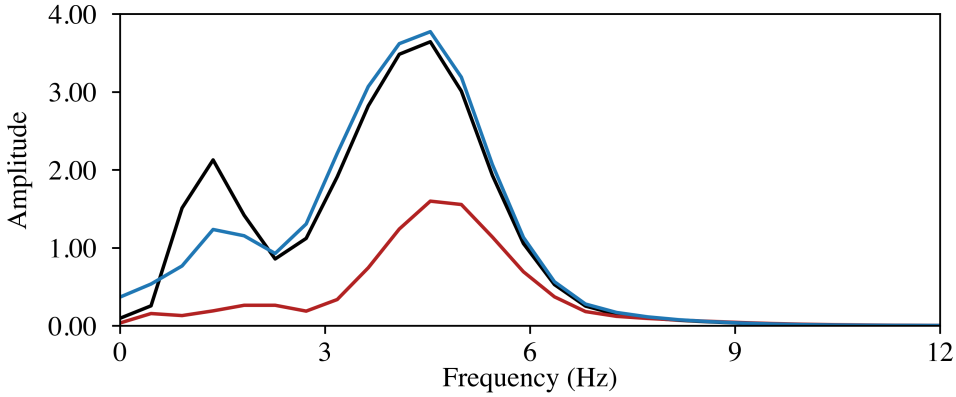


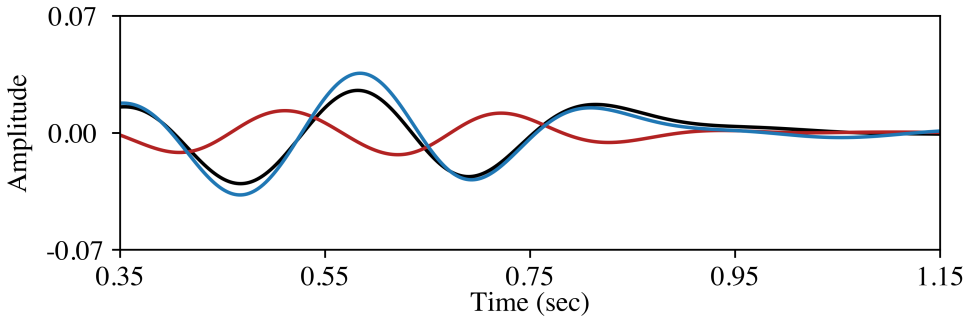

**Figure 9.** Comparison between the average amplitude spectra of the ghost-free data (black line) shown in Fig 8a, the ghost-contaminated data (red line) shown in Fig 8b, and the ghost-prediction data (blue line) shown in Fig 8c.

we add Gaussian noise to the ghost-contaminated data such that the signal-to-noise ratio (SNR) is 4. Figure 11 shows the true ghost-free, the noisy ghost-contaminated, and the ghost-free prediction, respectively. The CNN predicts the ghost-free data from a noisy input with a great degree of accuracy. The average amplitude spectra show that the amplitude reconstruction is reliable for a ghost- and noise-contaminated input. By comparing the average amplitude spectra for data with (Figure 12) and without (Figure 6) noise, we note that random noise has limited effects on the ability of the CNN to recover the correct amplitude of the data.

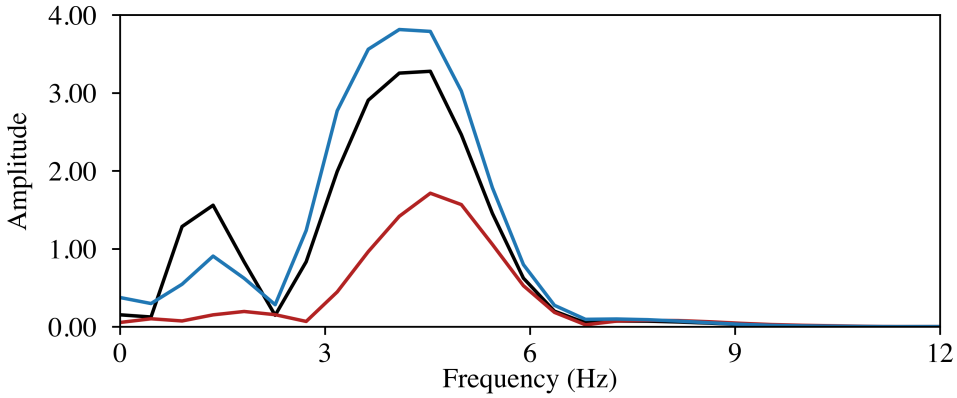
Figure 13a shows traces from the true ghost-free, the noisy ghost-contaminated, and the ghost-free prediction gathers for a receiver 0.1 km away from the source. The traces demonstrate the ability of the CNN to remove ghost reflections in the presence of random noise. Comparing the amplitude spectra of the traces with (Figure 13b) and without (Figure 7b) noise, and the corresponding phase spectra (Figure 7c and 13c) indicates that the CNN gives comparable results for input with and without noise, even though we train the CNN with noise-free data. The ability of the CNN to remove ghost reflections from noise-contaminated input, under a noise-free assumption, demonstrates the robustness of our deghosting solution to random noise.

## 4 DISCUSSION

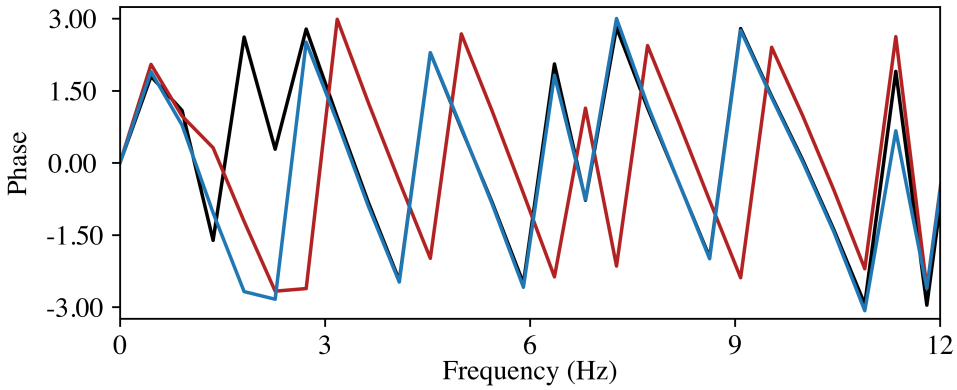
The numerical example demonstrates the ability of CNNs to remove source-side and receiver-side ghosts from conventional pressure data acquired at a single reference surface in the shot gather domain. This solution does not make assumptions about the acquisition configuration, but it requires knowledge of such configuration to create training data that replicates the actual field acquisition




(a)




(b)




(c)

**Figure 10.** (a) 5 Hz low-pass filter applied to the traces shown in Fig 7a. (b) Comparison between the amplitude spectra of the shown traces. (c) Comparison between the phase spectra of the shown traces.

geometry. We assume that data are acquired under a flat sea surface, which is not always a feature of marine acquisitions. Cecconello et al. (2018), Blacquière and Sertlek (2019), and Konuk and Shragge (2020) show that a dynamic sea surface significantly affects the phase and amplitude of seismic data. Therefore, evaluating the ability of CNNs to remove ghost reflections for data acquired under a time-varying sea surface condition is an important extension of our method and subject to future work.

Although ghost reflections are independent of the source function, they are dependent on the bandwidth of the seismic signal, as shown by Equation 5. However, given that the solution is based on a data-fitting procedure, the applicability of the proposed solution to field data depends on the ability of the solution to tolerate source wavelet uncertainties. In other words, the ability to

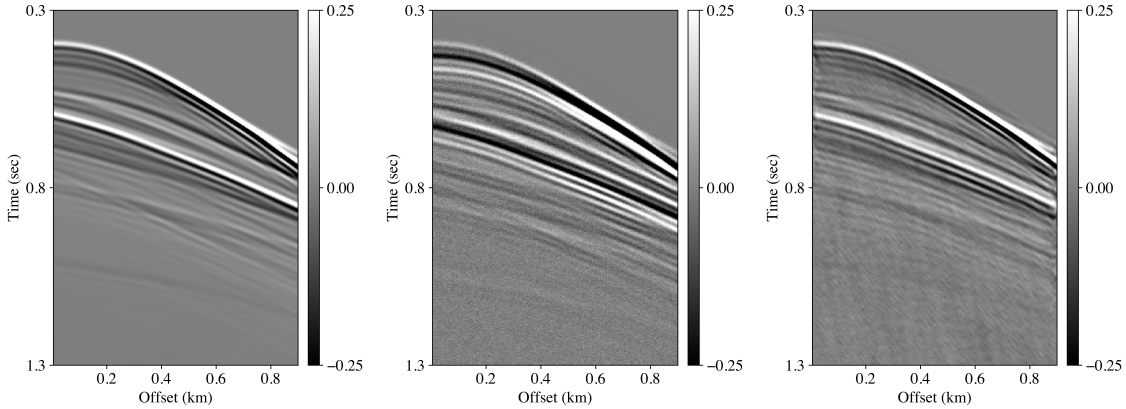



**Figure 11.** Comparison between (a) the true ghost-free data, (b) the ghost-contaminated data with random noise (SNR=4), and (c) the ghost-free prediction.

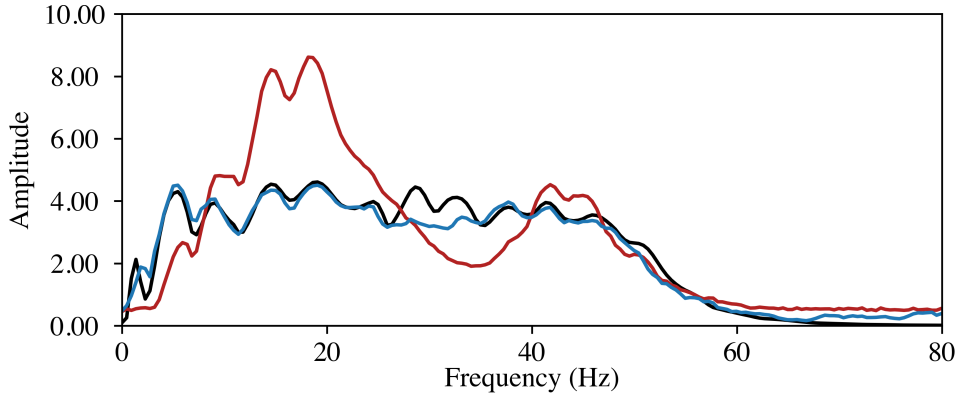


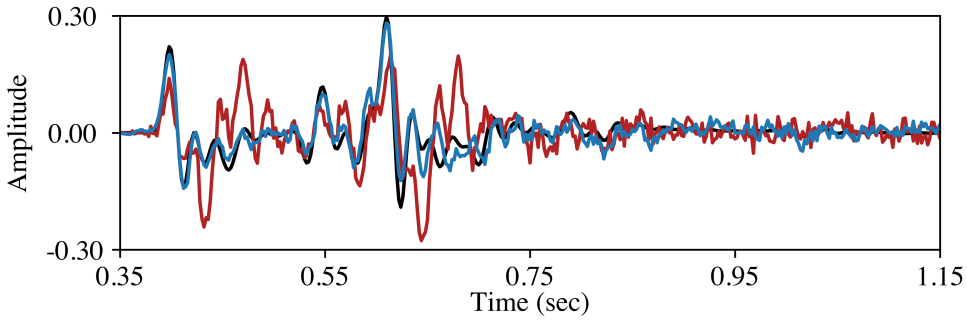

**Figure 12.** Comparison between the average amplitude spectra of the ghost-free data (black line) shown in Fig 11a, the ghost-contaminated data with random noise (red line) shown in Fig 11b, and the ghost-free prediction data (blue line) shown in Fig 11c.

apply the solution to field data depends on the ability of the CNN to learn to remove ghost reflections regardless of the source function. Therefore, evaluating the robustness of the CNN-based deghosting operator to different source wavelets remains an open problem for future work.

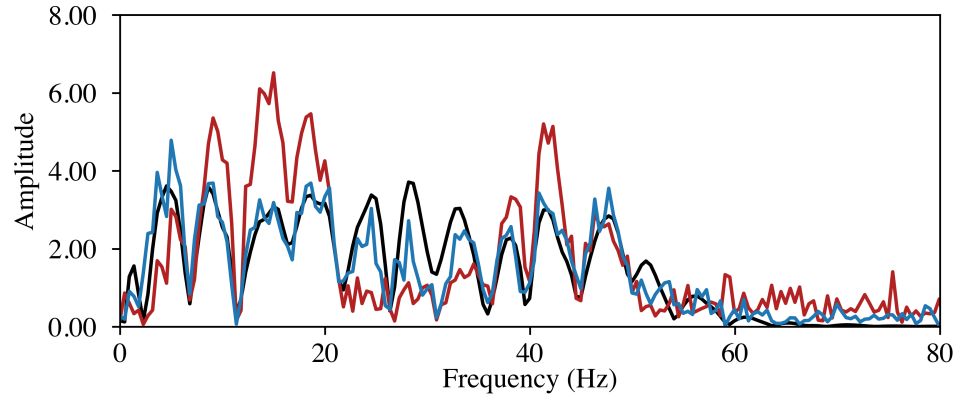
Understanding the effects of different physics between training and testing data on deghosting is important for applying the proposed solution to field data. Different physics assumptions generate data with different characteristics. We conjecture that it is important for field and training data to match in character, but this remains open for investigation.

## 5 CONCLUSION

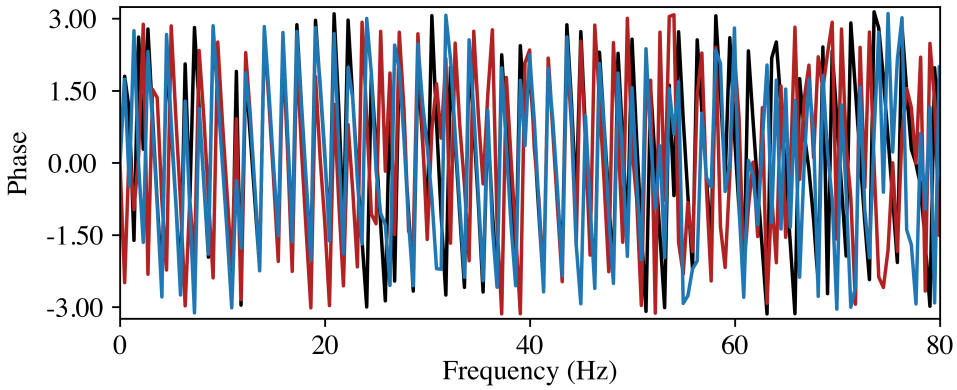
We propose a method to remove source-side and receiver-side ghosts in the shot domain. The solution works for pressure data acquired at a single reference surface and uses a convolutional neural network with a supervised learning approach to map ghost-contaminated data to ghost-free data. Such a solution enables ghost removal from legacy data acquired using conventional streamers. Further, our solution helps reduce costs and alleviates acquisitions and processing complexities associated with multi-measurement acquisition systems. Since our solution is based on learning through a data-fitting procedure, the acquisition geometry and the bathymetry of the ocean floor must be known to create training data that closely mimic field data acquisition conditions. We demonstrate with a synthetic example that CNNs can predict ghost-free data with a great degree of accuracy. We show that the solution is robust to data generated using a model with strong velocity contrasts and velocity values beyond that of the models used




(a)




(b)




(c)

**Figure 13.** (a) Comparison between ghost-free (black line), ghost-contaminated with noise (red line), and ghost-free prediction (blue line) traces recorded at a receiver that is 0.1 km away from the source. (b) Comparison between the amplitude spectra of the shown traces. (c) Comparison between the phase spectra of the shown traces.

to generate the training data. Thus, one can lower the computational burden of generating training data by using models with a narrow range of velocities. Additionally, we show that the solution is robust to inputs with random noise even though the training data are noise-free. The presence of random noise increases the uncertainty of the accuracy of ghost-free predictions. Therefore, the robustness of our solution to noise provides a framework to remove ghost reflections and mitigate any uncertainty due to random noise simultaneously.

## 6 ACKNOWLEDGMENTS

The first author would like to thank Saudi Aramco for graduate study sponsorship. We thank the sponsors of the Center for Wave Phenomena at Colorado School of Mines for the support of this research. We also thank Tugrul Konuk (CWP) for many useful discussions and for providing the forward-modeling code that made this research possible.

## REFERENCES

Alexandrov, D., B. Kashtan, A. Bakulin, and P. Leger, 2014, Dual-sensor summation with buried land sensors, *in* SEG Technical Program Expanded Abstracts 2014: Society of Exploration Geophysicists, 1929–1933.

Amundsen, L., 1993, Wavenumber-based filtering of marine point-source data: Geophysics, **58**, 1335–1348.

Amundsen, L., Ø. Pedersen, A. Osen, J. O. Robertsson, and M. Landrø, 2017, Broadband seismic over/under sources and their designature-deghosting: Geophysics, **82**, P61–P73.

Amundsen, L., B. G. Secrest, and B. Arntsen, 1995, Extraction of the normal component of the particle velocity from marine pressure data: Geophysics, **60**, 212–222.

Amundsen, L., and H. Zhou, 2013, Low-frequency seismic deghosting: Geophysics, **78**, WA15–WA20.

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

Bearnth, R. E., and N. A. Moore, 1989, Air gun-slant cable seismic results in the gulf of mexico, *in* SEG Technical Program Expanded Abstracts 1989: Society of Exploration Geophysicists, 649–652.

Beasley, C. J., R. Coates, Y. Ji, and J. Perdomo, 2013a, Wave equation receiver deghosting: A provocative example, *in* SEG Technical Program Expanded Abstracts 2013: Society of Exploration Geophysicists, 4226–4230.

Beasley, C. J., R. Coates, and C. Lapilli, 2013b, Wave equation receiver deghosting: 2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), IEEE, 280–283.

Biswas, R., M. K. Sen, V. Das, and T. Mukerji, 2019, Prestack and poststack inversion using a physics-guided convolutional neural network: Interpretation, **7**, SE161–SE174.

Biswas, R., A. Vassiliou, R. Stromberg, and M. K. Sen, 2018, Stacking velocity estimation using recurrent neural network, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 2241–2245.

Blacquière, G., and H. Ö. Sertlek, 2019, Modeling and assessing the effects of the sea surface, from being flat to being rough and dynamic: Geophysics, **84**, T13–T27.

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

Cambois, G., D. Carlson, C. Jones, M. Lesnes, W. Söllner, and H. Tabti, 2009, Dual-sensor streamer data: calibration, acquisition QC and attenuation of seismic interferences and other noises, *in* SEG Technical Program Expanded Abstracts 2009: Society of Exploration Geophysicists, 142–146.

Carlson, D. H., A. Long, W. Söllner, H. Tabti, R. Tenghamn, and N. Lunde, 2007, Increased resolution and penetration from a towed dual-sensor streamer: First Break, **25**, 71–77.

Carothers, N. L., 1998, A short course on approximation theory: Bowling green State University.

Cecconello, E., E. G. Asgedom, O. C. Orji, M. W. Pedersen, and W. Söllner, 2018, Modeling scattering effects from time-varying sea surface based on acoustic reciprocity: Geophysics, **83**, T49–T68.

Claerbout, J. F., 1976, Fundamentals of geophysical data processing: McGraw-Hill, New York.

Cybenko, G., 1989, Approximation by superpositions of a sigmoidal function: Mathematics of Control, Signals, and Systems, **2**, 303–314.

Das, V., and T. Mukerji, 2020, Petrophysical properties prediction from pre-stack seismic data using convolutional neural networks: Geophysics, **85**, N41–N55.

Das, V., A. Pollack, U. Wollner, and T. Mukerji, 2019, Convolutional neural network for seismic impedance inversion: Geophysics, **84**, R869–R880.

Day, A., T. Klüver, W. Söllner, H. Tabti, and D. Carlson, 2013, Wavefield-separation methods for dual-sensor towed-streamer data: Geophysics, **78**, WA55–WA70.

Dondurur, D., 2018, Acquisition and processing of marine seismic data: Elsevier.

E, W., 2020, Machine learning and computational mathematics: Communications in Computational Physics, **28**, 1639–1670.

Egan, M., K. G. El-Kasseh, and N. Moldoveanu, 2007, Full deghosting of obc data with over/under source acquisition, *in* SEG Technical Program Expanded Abstracts 2007: Society of Exploration Geophysicists, 31–35.

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

Géron, A., 2019, Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems: O'Reilly Media.

Goodfellow, I., Y. Bengio, A. Courville, and Y. Bengio, 2016, Deep learning: MIT Press.

Griebel, M., 2005, Sparse grids and related approximation schemes for higher dimensional problems: SFB.

Grobbe, N., J. van der Neut, E. Slob, K. Wapenaar, C. A. Vidal, and G. Drijkoningen, 2016, Unified multi-depth-level field decomposition: Geophysical Prospecting, **64**, 361–391.

Grohs, P., A. Jentzen, and D. Salimova, 2019, Deep neural network approximations for monte carlo algorithms: arXiv preprint arXiv:1908.10828.

Haavik, K. E., and M. Landrø, 2015, Variable source depth acquisition for improved marine broadband seismic data: Geophysics, **80**, A69–A73.

Hanka, R., and T. P. Harte, 1997, Curse of dimensionality: classifying large multi-dimensional images with neural networks, *in* Computer Intensive Methods in Control and Signal Processing: Springer, 249–260.

He, J.-W., W.-K. Lu, and Z.-X. Li, 2013, An adaptive over/under data combination method: Applied Geophysics, **10**, 469–476.

Hinton, G., L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, et al., 2012, Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups: IEEE Signal Processing Magazine, **29**, 82–97.

Hornik, K., 1991, Approximation capabilities of multilayer feedforward networks: Neural Networks, **4**, 251–257.

Hornik, K., M. Stinchcombe, H. White, et al., 1989, Multilayer feedforward networks are universal approximators.: Neural Networks, **2**, 359–366.

Huang, S., and Y. Huang, 1990, Bounds on number of hidden neurons of multilayer perceptrons in classification and recognition: IEEE International Symposium on Circuits and Systems, IEEE, 2500–2503.

Ikelle, L. T., and L. Amundsen, 2018, Introduction to petroleum seismology: Society of Exploration Geophysicists.

Jeffreys, H., and B. Jeffreys, 1950, Methods of mathematical physics: Cambridge University Press.

Jovanovich, D. B., R. D. Sumner, and S. L. Akins-Easterlin, 1983, Ghosting and marine signature deconvolution: A prerequisite for detailed seismic interpretation: Geophysics, **48**, 1468–1485.

Kapoor, J., C. Stork, and M. Egan, 2005, Benefits of low frequencies for subsalt imaging: SEG Technical Program Expanded Abstracts 2005, Society of Exploration Geophysicists, 1993–1996.

Knapp, R. W., 1990, Vertical resolution of thick beds, thin beds, and thin-bed cyclothems: Geophysics, **55**, 1183–1190.

Konuk, T., and J. Shragge, 2020, Modeling full-wavefield time-varying sea-surface effects on seismic data: A mimetic finite-difference approach: Geophysics, **85**, T45–T55.

Kragh, E., E. Muyzert, T. Curtis, M. Svendsen, and D. Kapadia, 2010, Efficient broadband marine acquisition and processing for improved resolution and deep imaging: The Leading Edge, **29**, 464–469.

Krizhevsky, A., I. Sutskever, and G. E. Hinton, 2017, Imagenet classification with deep convolutional neural networks: Communications of the ACM, **60**, 84–90.

Lapedes, A., and R. Farber, 1987, How neural nets work: Conference on Neural Information Processing Systems, 442–456.

LeCun, Y., Y. Bengio, and G. Hinton, 2015, Deep learning: Nature, **521**, 436–444.

LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998, Gradient-based learning applied to document recognition: Proceedings of the IEEE, **86**, 2278–2324.

Llanas, B., S. Lantarón, and F. J. Sáinz, 2008, Constructive approximation of discontinuous functions by neural networks: Neural Processing Letters, **27**, 209–226.

Ma, Y., X. Ji, T. Fei, and Y. Luo, 2018, Automatic velocity picking with convolutional neural networks, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 2066–2070.

Martin, G. S., K. J. Marfurt, and S. Larsen, 2002, Marmousi-2: An updated model for the investigation of avo in structurally complex areas, *in* SEG Technical Program Expanded Abstracts 2002: Society of Exploration Geophysicists, 1979–1982.

Moldoveanu, N., 2000, Vertical source array in marine seismic exploration, *in* SEG Technical Program Expanded Abstracts 2000: Society of Exploration Geophysicists, 53–56.

Moldoveanu, N., L. Combee, M. Egan, G. Hampson, L. Sydora, and W. Abriel, 2007, Over/under towed-streamer acquisition: A method to extend seismic bandwidth to both higher and lower frequencies: The Leading Edge, **26**, 41–58.

Moseley, B., A. Markham, and T. Nissen-Meyer, 2020, Solving the wave equation with physics-informed deep learning: arXiv preprint arXiv:2006.11894.

O'Brien, M., 1994, Amoco statics test: Amoco.

Paffenholz, J., J. Stefani, B. McLain, and K. Bishop, 2002, Sigsbee_2a synthetic subsalt dataset - image quality as function of

migration algorithm and velocity model error: Presented at the 64th EAGE Conference & Exhibition, European Association of Geoscientists & Engineers.

Pinkus, A., 1999, Approximation theory of the mlp model in neural networks: Acta numerica, **8**, 143–195.

Posthumus, B. J., 1993, Deghosting of twin streamer configuration: Geophysical Prospecting, **41**, 267–286.

Provenzano, G., T. J. Henstock, J. M. Bull, and G. Bayrakci, 2020, Attenuation of receiver ghosts in variable-depth streamer high-resolution seismic reflection data: Marine Geophysical Research, **41**, 1–15.

Raissi, M., P. Perdikaris, and G. E. Karniadakis, 2019, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations: Journal of Computational Physics, **378**, 686–707.

Reilly, J. M., 2016, Marine broadband technology: History and remaining challenges from an end-user perspective: The Leading Edge, **35**, 316–321.

Russell, S., and P. Norvig, 2016, Artificial intelligence: a modern approach: Pearson Education Limited.

Sabbione, J. I., and D. Velis, 2010, Automatic first-breaks picking: New strategies and algorithms: Geophysics, **75**, V67–V76.

Sartori, M. A., and P. J. Antsaklis, 1991, A simple method to derive bounds on the size and to train multilayer neural networks: IEEE Transactions on Neural Networks, **2**, 467–471.

Sønneland, L., L. E. Berg, P. Eidsvig, A. Haugen, B. Fotland, and J. Vestby, 1986, 2-d deghosting using vertical receiver arrays, *in* SEG Technical Program Expanded Abstracts 1986: Society of Exploration Geophysicists, 516–519.

Soubaras, R., 2010, Deghosting by joint deconvolution of a migration and a mirror migration, *in* SEG Technical Program Expanded Abstracts 2010: Society of Exploration Geophysicists, 3406–3410.

Soubaras, R., and R. Dowle, 2010, Variable-depth streamer–a broadband marine solution: first break, **28**, 89–96.

Stein, E. M., and R. Shakarchi, 2011, Fourier analysis: An introduction: Princeton University Press.

Sun, J., S. Slang, T. Elboth, T. Larsen Greiner, S. McDonald, and L.-J. Gelius, 2020, A convolutional neural network approach to deblending seismic data: Geophysics, **85**, WA13–WA26.

Sutskever, I., J. Martens, G. Dahl, and G. Hinton, 2013, On the importance of initialization and momentum in deep learning: International conference on machine learning, PMLR, 1139–1147.

Sutskever, I., O. Vinyals, and Q. V. Le, 2014, Sequence to sequence learning with neural networks: Advances in Neural Information Processing Systems, 3104–3112.

Tamura, S., and M. Tateishi, 1997, Capabilities of a four-layered feedforward neural network: four layers versus three: IEEE Transactions on Neural Networks, **8**, 251–255.

ten Kroode, F., S. Bergler, C. Corsten, J. W. de Maag, F. Strijbos, and H. Tijhof, 2013, Broadband seismic data - the importance of low frequencies: Geophysics, **78**, WA3–WA14.

Tenghamn, R., S. Vaage, and C. Borresen, 2007, A dual-sensor towed marine streamer: Its viable implementation and initial results, *in* SEG Technical Program Expanded Abstracts 2007: Society of Exploration Geophysicists, 989–993.

Van der Baan, M., and C. Jutten, 2000, Neural networks in geophysical applications: Geophysics, **65**, 1032–1047.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

Wang, Y., A. C. Ramirez, and A. Osen, 2017, A low-frequency deghosting method: Analysis and numerical tests: Geophysics, **82**, V285–V296.

Whitcombe, D., and L. Hodgson, 2007, Stabilizing the low frequencies: The Leading Edge, **26**, 66–72.

Wu, Y., and G. A. McMechan, 2018, Feature-capturing full waveform inversion using a convolutional neural network, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 2061–2065.

Yang, F., and J. Ma, 2019, Deep-learning inversion: A next-generation seismic velocity model building method: Geophysics, **84**, R583–R599.