Learning location and seabed type from a moving mid-frequency source

T. B. Neilsen, C. D. Escobar-Amado, M. C. Acree, et al.

Citation: The Journal of the Acoustical Society of America **149**, 692 (2021); doi: 10.1121/10.0003361
View online: https://doi.org/10.1121/10.0003361
View Table of Contents: https://asa.scitation.org/toc/jas/149/1
Published by the Acoustical Society of America

JASA ARTICLE

Learning location and seabed type from a moving mid-frequency source[a)]

T. B. Neilsen,[1,b)] C. D. Escobar-Amado,[2] M. C. Acree,[1] W. S. Hodgkiss,[3] D. F. Van Komen,[1,c)] D. P. Knobles,[4] M. Badiey,[2] and J. Castro-Correa[2]

[1]*Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA*

[2]*Department of Electrical Engineering, University of Delaware, Newark, Delaware 19716, USA*

[3]*Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA*

[4]*Knobles Scientific and Analysis, LLC, Austin, Texas 78731, USA*

ABSTRACT:

While source localization and seabed classification are often approached separately, the convolutional neural networks (CNNs) in this paper simultaneously predict seabed type, source depth and speed, and the closest point of approach. Different CNN architectures are applied to mid-frequency tonal levels from a moving source recorded on a 16-channel vertical line array (VLA). After training each CNN on synthetic data, a statistical representation of predictions on test cases is presented. The performance of a single regression-based CNN is compared to a multitask CNN in which regression is used for the source parameters and classification for the seabed type. The impact of water sound speed profile and seabed variations on the predictions is evaluated using simulated test cases. Environmental mismatch between the training and testing data has a negative impact on source depth estimates, while the remaining labels are estimated tolerably well but with a bias towards shorter ranges. Similar results are found for data measured on two VLAs during Seabed Characterization Experiment 2017. This work shows the superiority of multitask learning and the potential for using a CNN to localize an acoustic source and detect the surficial seabed properties from mid-frequency sounds. © *2021 Acoustical Society of America.*

https://doi.org/10.1121/10.0003361

(Received 14 July 2020; revised 29 December 2020; accepted 29 December 2020; published online 27 January 2021)

[Editor: Peter Gerstoft] Pages: 692–705

I. INTRODUCTION

Environment variability in the world's oceans complicates the application of machine and deep learning to problems in ocean acoustics. The ocean environment is characterized by the sound speed in the water column, the water depth, and the seabed properties. Because the variable ocean environment impacts source localization and tracking and seabed characterization efforts,[1–3] all potential applications of machine and deep learning in ocean acoustics must be tested with this variability in mind. Van Komen *et al.*[4] contains evidence that deep learning, via a convolutional neural network (CNN), can find both a source range and a seabed type from a one-second pressure time series on a single hydrophone from an explosive source. This paper extends that work to mid-frequency tones emitted from a moving source received on a vertical line array (VLA) of hydrophones. The impact of environmental variability on CNN estimates of the moving source's location and seabed type are evaluated first using simulated data and then with data from two VLAs measured during the 2017 Seabed Characterization Experiment (SBCEX 2017).[5]

Acoustic source localization and tracking in the ocean have been performed using different passive acoustics techniques. Common optimization algorithms include matched-field processing (MFP),[6–9] multipath arrival estimation,[10,11] particle filtering,[12] and Bayesian approaches.[13,14] However, uncertainty in the ocean environment complicates source localization efforts.

Matched-field inversions[15] have been used in many cases to estimate environmental parameters with some using genetic algorithms,[16] simulated annealing,[17,18] or Bayesian-based sampling approaches.[3] Most studies focus on low-frequency sound that contains information about geoacoustic properties deeper in the seafloor. Recently, however, several studies in the mid-frequency range have been completed. Choi and Dahl[19] used reflection coefficients in the 2–20 kHz range to estimate sediment parameters using the Bootstrap method and concluded these frequencies could be used to obtain properties in the top 2 m for a sandy environment. Holland and Dosso[20] applied Bayesian methods to long-range reverberation to estimate depth-integrated intensity. Yang *et al.*[21] used bottom loss data in the 2–5 kHz range to estimate the properties of a half-space bottom model and found that averaging over source tracks helped reduce the impact of scattering over topographical changes. These studies point towards the potential of estimating surficial seabed properties from mid-frequency sounds.

[a)]This paper is part of a special issue on Machine Learning in Acoustics.

[b)]Electronic mail: atbn@byu.edu, ORCID: 0000-0002-9729-373X.

[c)]ORCID: 0000-0003-0610-0806.

https://doi.org/10.1121/10.0003361

The intertwined effects of source location and environmental characteristics on sound propagation have encouraged efforts to simultaneously obtain estimates for both. Approaches that have tackled the combined problem include Tabu optimizations,[22] simulated annealing with a rotated coordinate system,[23] and Bayesian inversions.[3] As the number of parameters in the inversions and their interdependence increases, so do the challenges associated with high-dimensional search spaces, nonlinear relationships between the unknowns, and large uncertainties due to varying information content about the inferred parameter values.

Recently, machine and deep learning approaches to the problems of source localization have gained interest. A review of learning applications in various fields of acoustics can be found in Bianco *et al.*,[24] including deep learning techniques based on neural networks. Steinberg *et al.*[25] used a neural network to localize a point source in a homogeneous medium in 1991. In ocean acoustics, neural networks have been employed for real-time range estimation[26] and source localization in the ocean.[27–32] Some studies, such as those done by Lefort *et al.*[29] and Niu *et al.*,[27,28] have found that deep learning classifiers outperform MFP, further suggesting the relevance of using deep learning in underwater environments. This work was extended in Ref. 33 to use a series of nine ResNet architectures to classify the range and depth of an air gun.

Machine and deep learning have also been used to make predictions about ocean seabed parameters, such as attenuation and sound speed. Early efforts included the use of neural and statistical classifiers by Michalopoulou *et al.*,[34] parallel feed-forward networks by Benson *et al.*,[35] and global and hierarchical approaches by Stephan *et al.*[36] Recently, Piccolo *et al.*[37] used generalized additive models for predicting compressional sound speed and attenuation from time-domain features from a single sensor. Vertical line array input was used by Niu *et al.*[38] to extract modal properties using a block sparse Bayesian learning and by Frederick *et al.*[39] in machine learning classifiers for a two-layer sediment model on synthetic data.

Simultaneous prediction of source range and seabed type using a CNN trained on synthetic data and applied to measured data were presented in Van Komen *et al.*[4] using one-second, broadband pressure time series from a single, bottom-mounted hydrophone. This work is extended in the current study to show that a CNN can learn both a moving source's location and a seabed type from mid-frequency tonal levels recorded on a VLA. Because of a lack of labeled field data, the CNN is trained on data simulated for measured sound speed profiles, four distinct seabed types, and a variety of source parameters. We present a comparison between a CNN based on regression and an implementation of the multitask deep learning (MTL) technique described in Kendall *et al.*[40] to perform classification for the seabed type and regression for the localization parameters simultaneously.

The MTL-CNN is compared to CNNs using regression for source and seabed labels, along with CNNs that learn only a single label. For all these networks, the CNN learning was ensured by high accuracy for randomly drawn validation data. The trained CNNs are applied to data simulated with sound speed profiles (SSPs) and seabed types not used in training to investigate the robustness of the predictions in cases of SSP and seabed mismatch. The trained CNNs are also applied to data measured on two VLAs during SBCEX 2017 to test the ability of the CNNs to generalize to ocean data. These tests show the potential for CNNs to estimate seabed type and source motion from mid-frequency tones.

II. EXPERIMENT

A. VLA data

The data were collected in the New England Mud Patch area during SBCEX 2017. The operational area for the experiment is depicted in Fig. 1. The black symbols represent acoustic sensors deployed by other research groups. A complete list of these sensors is found in Ref. 5.

The data used in this paper were collected at two VLAs deployed by the Marine Physical Laboratory of the Scripps Institution of Oceanography (MPL). The locations of the two MPL VLAs are depicted in Fig. 1 with green triangles: VLA 1 is centrally located, whereas VLA 2 is approximately 6 km SE. The coordinates of the VLAs are [40.470N, 70.597W] and [40.442N, 70.527W], respectively. The color bar in the plot represents the depth of the ocean in the region based on bathymetry data. The average ocean depth at VLA 1 is 71 m and VLA 2 is at 74 m depth.

During a portion of the experiment, an ITC2015 source was towed along roughly straight tracks in the vicinity of the VLAs (deviations due to fishing gear in the water). This source had a 45 m nominal tow depth and transmitted tones at frequencies of 2.0, 2.5, 3.0, 3.5, and 4.0 kHz. The source signal was received and recorded by both VLAs with a sampling rate of 25 kHz. Each VLA has 16 hydrophones with a 56.25 m aperture. The lowest one is 3 m from the ocean floor, and the hydrophone spacing is 3.75 m. The location of the ship during the time of interest, provided by GPS data, is shown as a blue line in Fig. 1. The (lower) NW travelling track (Track 1) began on March 24, 2017 (Julian Day 083)

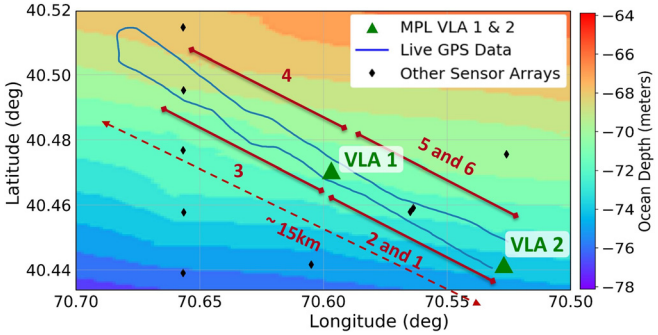

FIG. 1. (Color online) Bathymetric map of the area. Thin (blue) line shows the ship track from live GPS, and triangles (green) show the VLA locations. Thicker (red) lines and numbers 1–6 indicate the portion of the track assigned to different half-track/VLA data samples. The diamonds (black) show the locations of other sensor arrays deployed during SBCEX 2017.

at 0121 UTC, and the (upper) SE travelling track (Track 2) began at 0435 UTC on the same day. The full length of each track is nearly 15 km, however, the extracted portions of the tracks used in this study are approximately 6.5 km. These extracted portions are annotated with the red lines in Fig. 1. The corresponding red numbers indicate the sample number assigned to each half-track/VLA combination, and are used to identify the measured data samples in Figs. 2 and 10.

The time series data from linear frequency-modulated pulses on VLA 2 were used by Michalopoulou and Gerstoft[41] to obtain source, array element, and environmental properties. The received time signals across the 16 channels were used to obtain arrival times of different propagation paths via particle filtering and a cross correlation cost function. The resulting arrival times were used to obtain source range and depth, array element depth, water column depth, and the sediment sound speed and thickness. Their results agreed with expected values, and their work illustrates the richness of the MPL VLA data from SBCEX 2017.

B. Data extraction

During the times shown in Fig. 1 (Julian day 83, 0056 to 0732 Zulu time), the ITC 2015 towed source emitted tones in the mid-frequency range with a 50% duty cycle. Because of high signal-to-noise ratios, the five tones selected for this work were 2, 2.5, 3, 3.5, and 4 kHz. The 50% duty cycle was required due to marine mammal activity in the area; tones were on for 10 s and then off for 10 s. Because of the duty cycle, some preprocessing was required to obtain signals representative of a continuous source. A spectrogram was obtained using a 1 s window and 50% overlap. Because the energy of each tone was spread over several frequency bins around the center frequency, a summation of the squared pressure values over these frequency bins was performed to get the total energy associated with each tone.

Due to the 50% duty cycle, signal processing was done to approximate a continuous signal that preserved the information content but could be more easily modeled than a moving source that was on for 10 s then off for 10 s. The time-varying tonal levels for each of the five frequencies were smoothed via an envelope function on the peaks over a sliding 30 step (time sample) window using a spline fit. The ends of the spline fit were removed to prevent spuriously large values near the end points. The resulting smoothly varying tonal levels at 0.5 s time intervals are shown in Fig. 2.

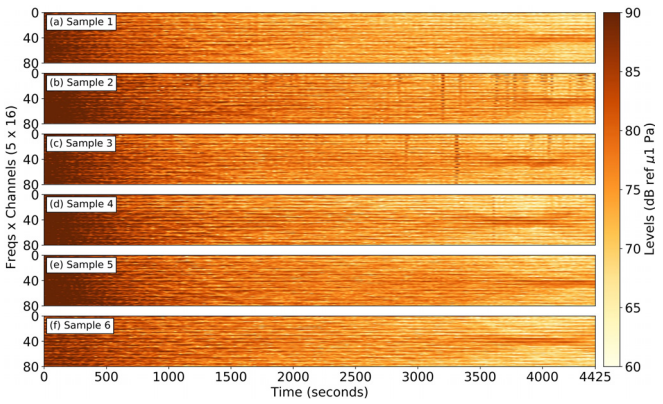

FIG. 2. (Color online) Extracted tonal levels from the VLAs during tracks 1 and 2 (Julian day 83, 0056 to 0732 Zulu time). The sample number corresponds to the numbers in Fig. 1 with samples 1 and 6 corresponding to VLA2 recordings and the remaining samples coming from VLA1.

Each resulting data sample contains the power spectral levels (in decibels) of the five tones on each of the 16 VLA channels for 8850 time steps covering 4424 s (approximately 74 min). Each data sample contains multichannel tonal levels stored as an 80×8850 matrix. The data samples (Fig. 2) correspond to the six half-track/VLA combinations identified in Fig. 1. The horizontal axis corresponds to the time. The high amplitude portions starting around sample 3500 is due to modal interference and contains information about the modal properties of the ocean environment.

Some of the extracted signals have the closest point of approach (CPA) to the VLA at the beginning of the track, while CPA is at the end of the track for others. To provide consistency in the input data structure, the time axis on these latter signals was flipped such that CPA occurs at the beginning of the matrix. Similar tonal spectrograms are simulated for training the neural networks, as described in Sec. IV D.

Selection of the time corresponding to CPA for the towed source was challenging. Due to the complex interference patterns, peak levels and change in Doppler shift appeared different for the different frequencies. Thus, the start time for the extracted signals corresponds roughly to CPA. Because the simulated training data begin with the source at CPA, this uncertainty in start time for the measured samples translates into an uncertainty in expected CPA range of approximately 20 m, with CPA range defined as the horizontal distance between the source and receiver array when the source is at the closest point of approach, as illustrated in Fig. 3.

III. METHODS

CNNs and other supervised deep learning algorithms are typically trained with a large number of labeled samples. The lack of labeled field data is one of the primary challenges for applying deep learning in ocean acoustics. Our approach is to simulate data using source-receiver configuration similar to the measured data in representative ocean environments. These simulated training data are used for training and validating the CNNs. Additional simulated data from environments not represented in the training data are then used to test the robustness of the trained CNNs predictions to environmental mismatch. Finally, the trained CNNs are applied to measured data (Fig. 2), as described in Sec. II B. Details for each of these steps are provided in this section.

A. Simulated data

The data simulation process models the interaction of the source spectrum with the ocean waveguide for different

https://doi.org/10.1121/10.0003361

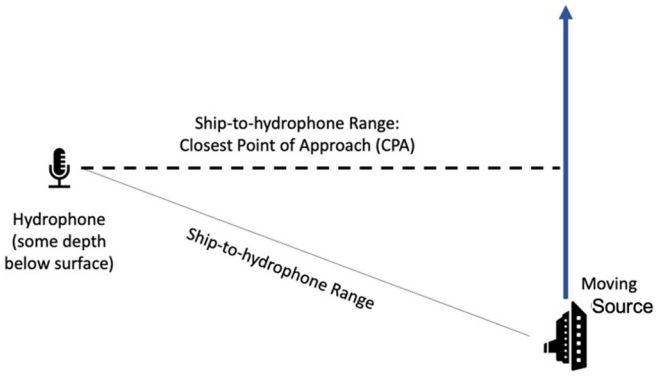

FIG. 3. (Color online) Geometry assumed for the path of the moving source relative to the sensor array.

ocean environments. The ITC2015 source line levels are used as the source spectral levels. For these five frequencies, the ocean response is modeled with ORCA, a range-independent, elastic, normal-mode model.[42] The two assumptions about the source motion are that the first time in each data sample corresponds to CPA, and second, that the source moves along a track perpendicular to the line between the receiver and CPA, as illustrated in Fig. 3. Source line levels relative to 1 μPa at 1 m are 142 dB at 2 kHz, 144 dB at 2.5 kHz, 140 dB at 3 kHz, 140 dB at 3.5 kHz, 139 dB at 4 kHz. These levels were obtained experimentally and are included because the relative levels between tones are presented to the CNN during training. Data simulation occurs for all 16 channels with the assumption that the VLA is vertical.

Data are simulated for several combinations of SSP and seabed type. Three SSPs measured during the experiment are used to generate the training data. These SSPs are shown as black lines in Fig. 4, with a water depth of 74.4 m. While there are many ways to parameterize the seabed, this first mid-frequency study in characterizing the seabed with deep

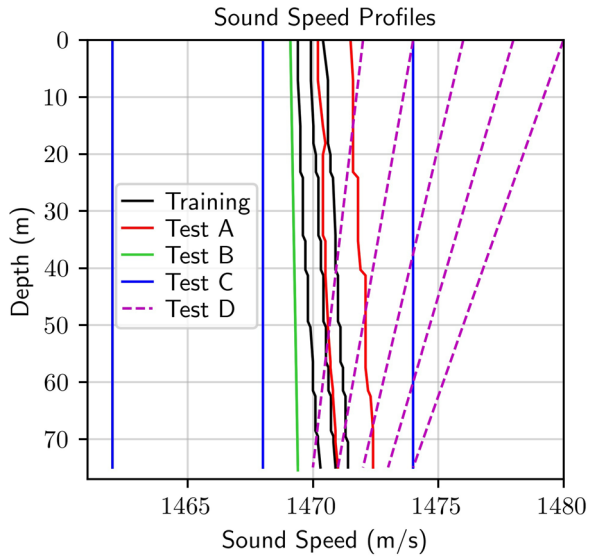

FIG. 4. (Color online) Sound speed profiles (SSPs) used for simulating training and testing datasets.

learning uses four representative seabed types that span a wide range of seabed characteristics from a soft, deep mud to a hard sand.

Although the seabed in the New England Mud Patch where SBCEX 2017 took place has been studied extensively,[5] this work is evaluating the ability of a CNN to distinguish between different types of seabeds when such *a priori* information is not available. Thus, the four selected seabeds are inspired by geoacoustic inversions from different areas of the world. The properties of these four seabeds are displayed in Fig. 5. The deep mud seabed type (#1) was estimated for the Gulf of Mexico, as reported by Knobles *et al.*[43] The mud over sand seabed type (#2) was obtained from maximum entropy statistical inference on data collected during the SUS circle experiments in SBCEX 2017.[44] The sandy silt seabed type (#3) resulted from geoacoustic inversions for data collected in the New England Bight by Potty *et al.*[45] The sand seabed type (#4) was the conclusion of a study on several sandy sea bottoms by Zhou *et al.*[46] For each seabed type, the sediment sound speeds are shown as thick black lines, while the legend contains the remaining ORCA input parameters: attenuation, α, and density, ρ, for the water (subscript w), and the top (superscript t) and bottom (superscript b) of each layer (subscript $i = 1 - 3$) and the half-space (subscript hs). For each seabed, the ratio across the water-sediment interface [$r_c = c_1^t/c(h_w)$, where $c(h_w)$ is the sound speed at the bottom of the water column] is held constant when different SSPs are used in the simulations. A linear gradient g_i is assumed for each sediment layer, i.

These four seabeds are used to generate a synthetic training dataset as a proof-of-concept that a CNN can learn some of the basic properties of the ocean seabed, such as degree of reflectivity. In this work, the seabeds have been ordered from highest to lowest bottom loss. Seabed number 1 (deep mud) has a water-sediment sound speed ratio, r_c, less than 1, which causes the modal wavenumbers to have a large imaginary part. Seabed number 2 (mud over sand) has a similar structure at the top of the sediment but an additional reflection occurs because of a deeper, more reflective layer. Seabed number 3 (sandy silt) has $r_c > 1$, mitigating penetration into the seabed and, thus, enhancing the specular component of the seabed reflection. Seabed number 4 (sand) has the largest r_c, providing an even larger specular response. Seabed types 1 and 2 are characterized by an angle of intromission, whereas seabed types 3 and 4 are characterized by a critical angle. Significant work is needed to discover the best way to extend the number and variety of seabed types required for *in situ* applications.

1. Training data

The simulated training data set were designed to share characteristics with the extracted data. Levels for five tonal frequencies were simulated using all combinations of three measured sound speed profiles (black in Fig. 4), the four seabed types (Fig. 5), and a variety of source parameters.

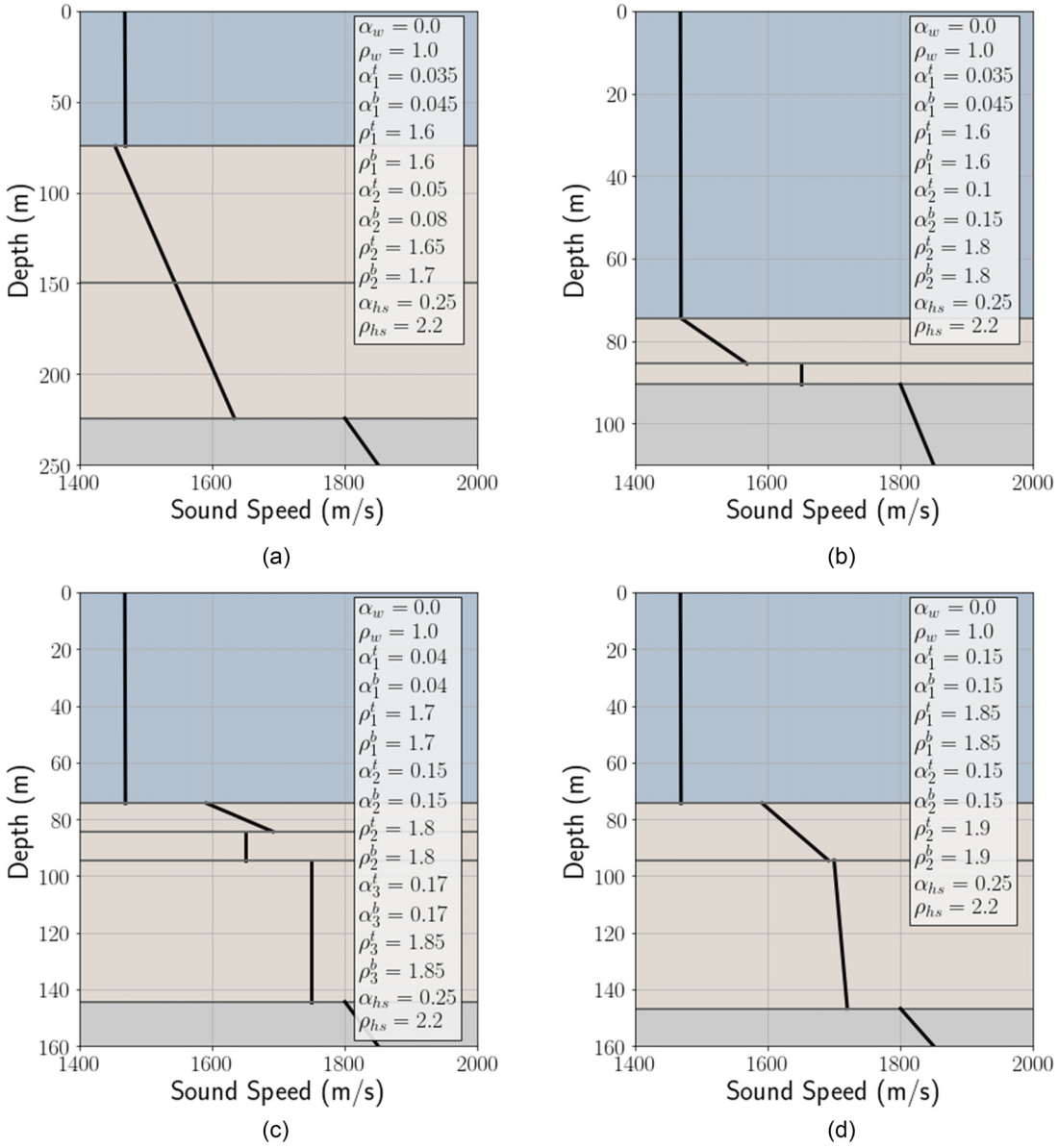

FIG. 5. (Color online) Four seabed models used for simulating the synthetic training data with upper sediment layers of (a) 1: deep mud (Ref. 43), (b) 2: mud over sand (Ref. 44), (c) 3: sandy silt (Ref. 45), and (d) 4: sand (Ref. 46). Sound speed profiles are shown as thick black lines, while the legend contains attenuation, α, and density, ρ, for the water (subscript w), and the top (superscript t) and bottom (superscript b) of each layer (subscript $i = 1 - 3$) and the half-space (subscript hs).

The minimum and maximum source parameters are shown in Table I. To make sure the entire source parameter space is sampled, the data were simulated at specified values (evenly spaced between the minimum and maximum values) that are constant for each environment; the combination of specified source parameters for each of the four seabed types and three SSPs yielded 3600 samples. In addition, data were simulated using randomly selected sets of source parameters for each environment resulting in an additional 3000 samples. Thus, the entire training dataset contained 6600 samples. The total number of specified and random source parameters are also listed in the Table I. For comparison, the last column contains approximate source parameters expected for the measured samples from the ship GPS and assumed distance to the towed source.

TABLE I. Span and quantity of source parameters used in generating the simulated training data along with approximate expected values for the measured samples. The source parameters are CPA range r, ship speed v, and source depth z_s.

	Min Value	Max Value	Number Specified	Number Random	Expected Values
r, m	100	1100	10	10	200–900
v, kts	1	5	5	5	3
z_s, m	5	65	6	5	45

The synthetic data samples were arranged in the same manner as the extracted data. Levels at five towed tonal frequencies (2 to 4 kHz) were simulated for the 16 channels of the VLA and 8850 time steps. Each data sample was

https://doi.org/10.1121/10.0003361

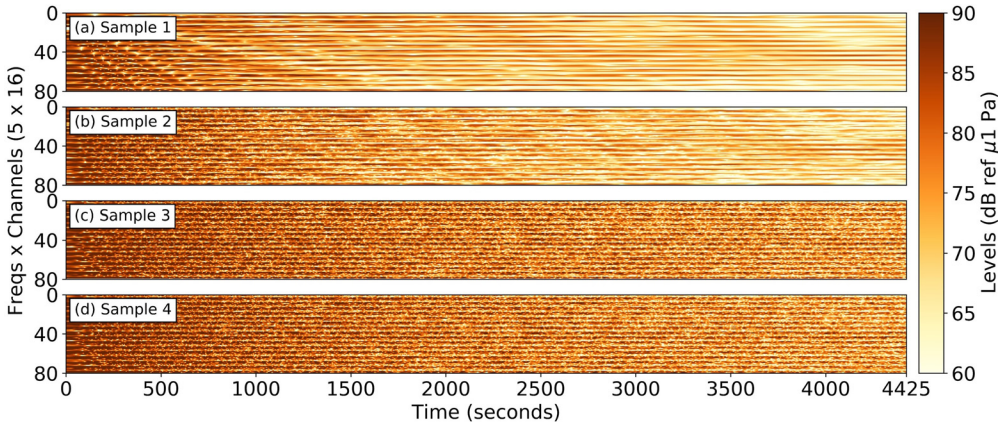

FIG. 6. (Color online) Examples of the synthetic data for CPA range of 400 m, source depth = 45 m, and ship speed = 3 knots for each of the four seabeds in Fig. 5.

smoothed along the horizontal axis, using a moving median with window size of eight time samples then downsampled by a factor of 8 resulting in an input size of 80×1106. Examples of the resulting 80×1106 matrices are shown in Fig. 6. This preprocessing reduced the computational memory requirement and better matches the envelope processing described in Sec. II B.

2. Simulating environmental mismatch

To analyze the performance of the networks in cases of environmental mismatch, five additional datasets were simulated with environments not used in training. The data for tests A, B, C, and D were created to analyze the impact of a mismatch in the SSP in the water column and in the water depth (test B, +/ –0.5 m) while using the same seabeds as the training data (Fig. 5). The data for test E were created using the same SSPs as the training data but six seabeds based on different parameterizations of the viscous grain shearing (VGS) model that represent various types of mud and sand.[47,48]

The cases of SSP mismatch simulated different propagation conditions than the slightly upward refracting SSPs measured during the experiment, as shown in Fig. 4. Test A had SSPs with the same slope but slightly offset from the measured SSPs used for the training data. Test B had a single isovelocity SSP but random water depths of $74.4\,\text{m} \pm 0.5\,\text{m}$. Test C used three isovelocity SSPs (with a single water depth) with sound speeds above and below those used in the training data. Test D contains downward refracting sound speed profiles, with different slopes and represent a significantly different sound propagation environment. For these test cases, the data samples were generated with randomly selected source parameters to form 2000 samples for tests A and B, 3000 samples for test C, and 5000 samples for test D. These test cases are used to evaluate the performance of the CNN in cases of SSP mismatch; results are shown in Sec. IV B.

Test E investigates the issue of seabed mismatch. The four seabed types used to generate the training data (in Fig. 5) do not represent every possible type of seabed in the ocean, which leads to the important question: How will the CNNs perform when none of the seabeds used in training match the real one? To investigate this question, six additional seabeds were created using the VGS model.[47,48] The parameters for four of these seabeds are chosen to represent different muddy seabed types, potential candidates for the SBCEX 2017 experimental area (Sec. II A); the other two represent sandy seabeds. A description of each VGS seabed is provided in Table II. To evaluate the impact of seabed mismatch, a simulated test case was generated with these six VGS seabeds, the same SSPs and a subset of the source parameters used in the training data yielding a total of 4536 testing samples. Results are shown in Sec. IV C.

B. Convolutional neural networks

A convolutional neural network (CNN) is a standard deep learning tool designed to operate on gridded data similar to images[49] (e.g., evenly spaced in frequency, depth, time, etc.). CNNs are particularly useful for data in which patterns are important and when translations (e.g., time delays) can be ignored when comparing different data samples. More information on CNNs can be found in Goodfellow *et al.*[50]

The base architecture used in this study is a five-layer CNN. Five two-dimensional convolution kernels are employed, with max-pooling in the first two layers. The algorithm is written in PyTorch[51] and employs the Adam

TABLE II. Six seabeds defined by VGS parameters.

	Description
1	Lossy mud over sand
2	Deep mud over sand based on low-frequency geoacoustic inversions from the Gulf of Mexico in Ref. 43
3	Medium loss mud over sand
4	Low loss mud over sand
5	Sandy silt based on geoacoustic inversions from Ref. 45
6	Coarse sand based on geoacoustic inversions of New Jersey sand ridge experiment in Ref. 46

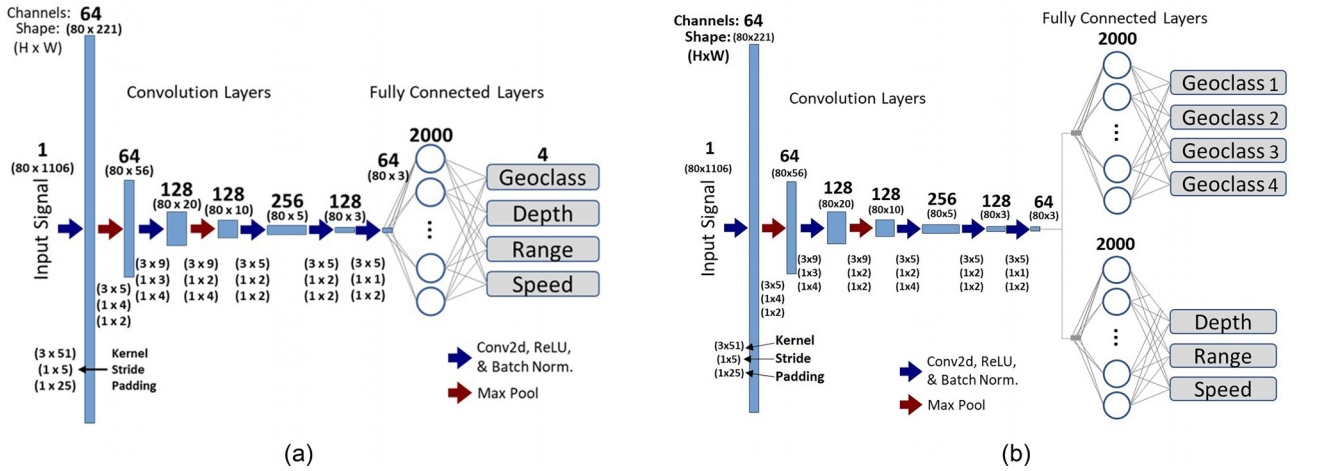

FIG. 7. (Color online) Convolutional neural network architectures using (a) regression for four labels (networks #3 and #4) and (b) multitask learning (networks #1 and #2).

optimizer,[52] ReLU activation functions,[53] and batch normalization. As with the procedure detailed in Ref. 4, a cosine annealing learning rate is used.

A set of preliminary tests to tune the hyperparameters of the network architectures provided insights into a few things that improved learning on these long (in time) data samples. Maximum pooling in the first two layers helped the network learn but not in the lower layers, as did large (in time) kernel sizes. In addition, the labels are first divided their maximum value and then multiplied by 100 (close to the peak value of the power spectral levels in the data), as this scaling expedites the learning process by putting the data and the labels on the same scale.[54] This base architecture was designed to fit the dimensions of the problem and fit the memory requirements. Future work can explore more complex architectures with many more parameters, such as ResNet.[55]

In this paper, a comparison is made between networks with different output layers. The architecture details are shown in Fig. 7. The loss function used to train the network determines if a classification or regression approach is taken. For networks where regression is done for all desired labels, a single fully connected layer is used prior to the output and the mean-squared error over all labels defines the loss function, as shown in Fig. 7(a). However, to accomplish multitask learning (MTL), two independent fully connected layers are used to add flexibility, as shown in Fig. 7(b).

The MTL approach is implemented to simultaneously classify the seabed type and do regression for the localization parameters. Homoscedastic uncertainty, which does not depend on the input data but in the network architecture, is taken into account by weighting the losses for each task.[40] Each label is learned separately and then the total loss function, L, combines the losses from the individual tasks

$$L = \frac{1}{2\sigma_1^2}L_1(w) + \frac{1}{2\sigma_2^2}L_2(w) + \frac{1}{2\sigma_3^2}L_3(w) + \frac{1}{\sigma_4^2}L_4(w) + log(\sigma_1\sigma_2\sigma_3\sigma_4), \quad (1)$$

where $L_1(w)$, $L_2(w)$, and $L_3(w)$ correspond to the mean-squared error for the CPA range, the ship speed, and the source depth regression tasks, respectively; $L_4(w)$ is the cross-entropy loss for the seabed classification; and σ_i is the noise parameter for each task. When σ_i increases, the objective is penalized and, therefore, the weight for that specific task decreases. The values of σ_i are learned during the training process to balance the different learning tasks and relate to the relative "noise" in the different learning tasks. In practice, $\log(\sigma_i^2)$ is used to avoid any division by zero.[40]

An example of the values of σ_i^2 learned during are displayed in Fig. 8. The plot shows the distribution of weights for all 20 instances; the dark line corresponds to the mean value and the colored area corresponds to ± 1 standard deviation to show consistency cross-the different networks. CPA range exhibits a higher noise that is penalized with a lower weight. The σ_i^2 learned for the seabed type is closer to zero which indicates that the classification task requires a weight several orders of magnitude higher than the regression tasks due to their differences in scale. While the difference in weights for the three regression parameters may be

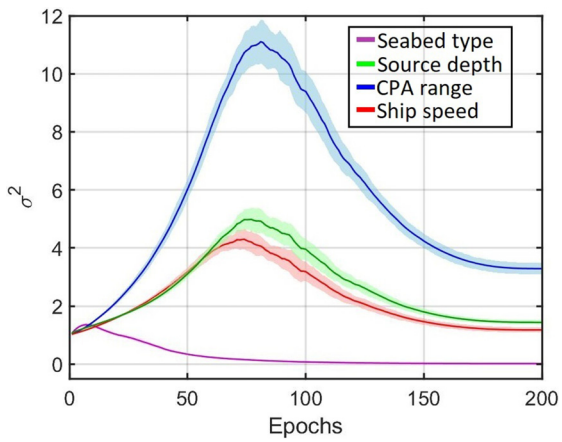

FIG. 8. (Color online) Values of σ_i^2 in Eq. (1) learned during training. The dark line corresponds to the mean value across 20 training instances, and the colored area corresponds to ± 1 standard deviation.

https://doi.org/10.1121/10.0003361

interpreted to correspond to their relative impact on the error, the difference between the seabed weight and the others is more related to the fundamental difference between the regression and classification tasks for this dataset.

To analyze the impact on performance due to how the labels and cost functions are defined, six different networks have been investigated: #1: MTL-CNN using regression tasks for the ship speed, CPA range, and source depth and a classification task for the seabed type, with seabeds ordered according to bottom loss as in Fig. 5. #2: MTL-CNN similar to #1 but with seabed type labels in a different order (i.e., with no physical meaning to the order). #3: CNN using regression tasks for all four labels with seabeds ordered according to bottom loss. #4: CNN similar to #3 but with seabed type labels in a different order. #5: Four independent CNNs that each learned only a single label using regression. #6: CNN using classification for seabed type only, i.e., without predicting any source localization parameters. Networks #5 and #6 have seabeds ordered according to bottom loss, similar to networks #1 and #3.

Each network has been trained 20 times to account for the uncertainty due to random initialization. For networks #2 and #4, the labels for the seabed type have been shuffled in such a way their ordering no longer has physical meaning. This mixing is done to determine the importance of using MTL instead of regression for the seabed type and to assess how network performance is affected when the order of the seabed types is not tied to a physical quantity.

C. Validation vs generalization

Tests of machine and deep learning algorithms come in two flavors: validation and generalization. Unfortunately, many papers on deep learning do not distinguish between the two. We feel this distinction is important, especially when developing algorithms for ocean acoustics applications.

The validation error of a network comes from testing the CNN on data drawn from the same statistical distribution as the training dataset. Typically, the data for validation is a randomly selected portion of the training data withheld for testing. For example, in an 80/20 train/test split, the CNN trains on 80% and then tests on the remaining 20%.

The generalization error quantifies the network's ability to take what it has learned during training and generalize to data drawn from a different statistical distribution. In the case of ocean acoustics, such differences in distribution could come from naturally occurring random variations in the SSP, bathymetry, seabed structure, ambient noise, etc. Additionally, modeling limitations can yield different statistical distributions. In this paper, the generalizability of the networks is evaluated for both simulated environmental mismatch and measured VLA data.

Both the validation and generalization errors of networks can be influenced by the random initialization of the weights in the network and the random draw of training data samples. If the training data set is large enough and the network performance is optimized, results should be independent of the initialization. Because this work is a proof-of-concept exercise for obtaining both source parameters and seabed type in face of ocean variability, time was not spent optimizing every aspect of the network. Instead, the robustness of each network is evaluated by comparing the results from multiple initialization (instances) of the CNN. The reported errors are averaged over 20 instances of each CNN.

IV. RESULTS

Metrics must be defined to concisely evaluate the performance of the neural networks. The metrics should be physically meaningful with regard to the different labels (source depth, speed, CPA range, and seabed type) and provide an indication of the uncertainty in results due to the random initialization. The metric used for the source labels is the mean absolute percentage error (MAPE). For each CNN instance n,

$$\mathrm{MAPE_n} = \frac{1}{M}\sum_m 100 \times \frac{|l_{n,m} - \hat{l}_{n,m}|}{l_{n,m}},$$

where m indicates the testing data sample, M is the total number of testing samples, l is the correct label, and $\hat{l}$ is the label predicted by the CNN. Statistics of the MAPE over the 20 training instances are presented for each network.

For seabed type, accuracy is presented. For the MTL and classification only networks (networks #1, #2, and #6, respectively), the accuracy was computed by counting the number of times the network predicted the correct seabed class. For the remaining three networks, the CNN-predicted (regressed) value for seabed type is rounded to the nearest integer and then compared to the correct seabed label; the seabed type accuracy is the percentage of the testing samples for which the rounded seabed type is correct.

A. Validation

The six networks were each trained for 200 epochs on 95% of the 6600 training samples and validated on the remaining 5%. All six networks had a MAPE for the validation set of less than 2% on the training samples and less than 5% on the validation samples for the source localization parameters and over 99% accuracy for the seabed type classification. The trained networks are applied to the simulated test cases and the measured data.

B. SSP mismatch

Test cases A–D are used to evaluate the impact of SSP mismatch on the CNN predictions of the source parameters and seabed type (when using the same seabeds as the training data). The statistical distribute of the MAPE in predicted labels, for the six networks over 20 training instances of each, are shown as box-and-whisker-plots in Fig. 9.

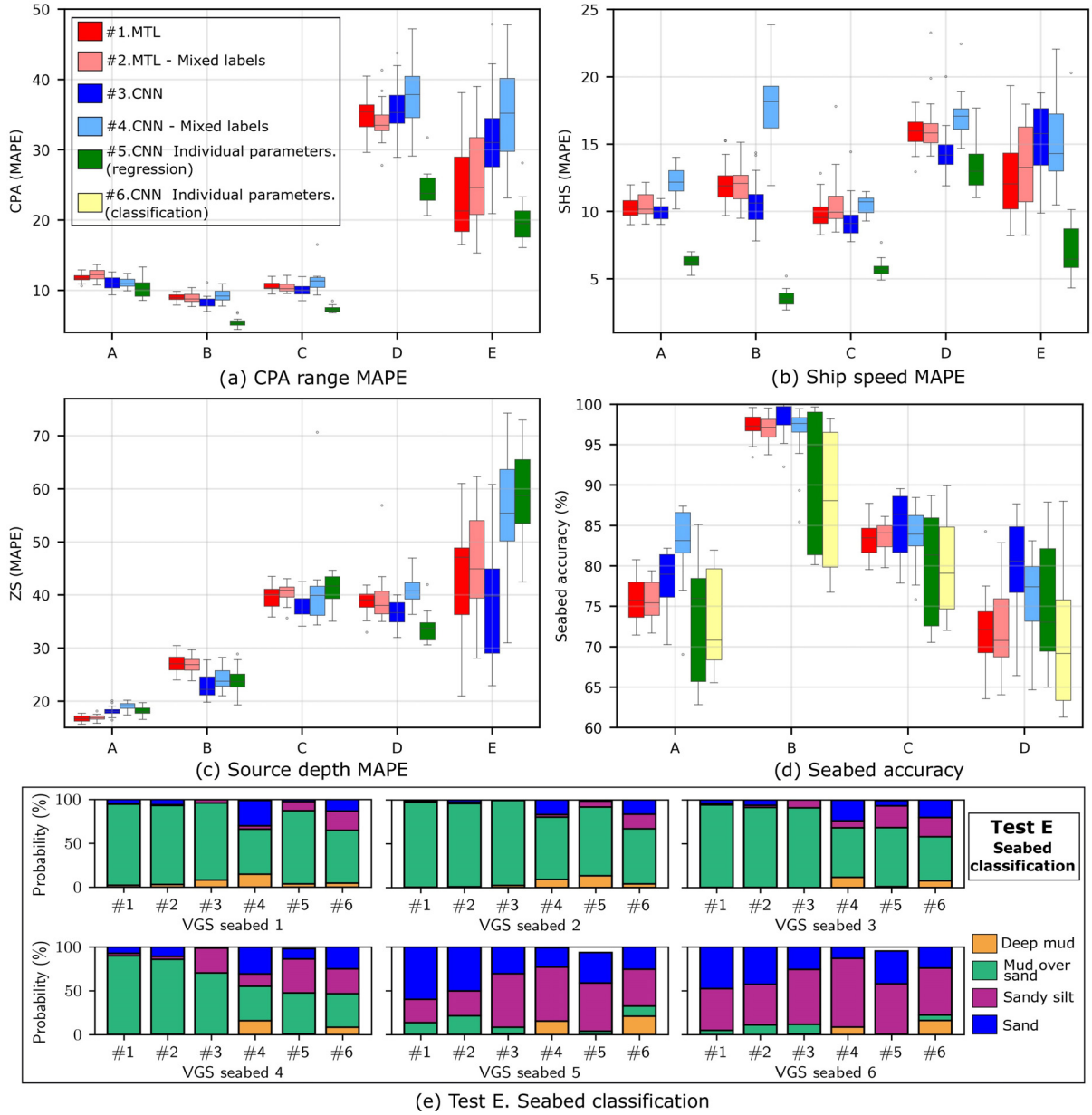

FIG. 9. (Color online) Simulated test cases results for the 20 instances of each network. Box and whisker error plots of the MAPE for (a) CPA range, (b) ship speed, (c) source depth, and (d) accuracy for seabed type. (e) Predicted seabed labels (stacked from 1 on the bottom to 4 on top) for each of the six VGS seabeds in test E.

The MAPE for the source labels exhibit different trends with SSP mismatch for networks #1–#4. In Fig. 9(a), the MAPE for CPA range is approximately 10% for tests A–C but increases to more than 30% for test D. Predictions for ship speed, in Fig. 9(b), show the most consistency in the face of SSP mismatch. The MAPE is approximately 10% for tests A–C and 15% for test D. SSP mismatch had the greatest effect on estimates of source depth. Even the slight variation in SSP in test A causes 20% MAPE in the seabed type; the water depth variation in test B causes 25% MAPE; and the larger SSP variations in tests C and D cause 40% MAPE. The trend is for the source depth to be underestimated in all these cases.

Even though the statistics of the MAPE describe the overall performance of the CNNs, the effect of the SSP mismatch on predictions of source labels is strongly correlated with the seabed type. Data samples generated with the deep mud seabed are most likely to have large errors. The sensitivity of source depth estimations to environmental mismatch is tied to the variation in the depth-dependent mode functions that occur as either the SSP or seabed changes.

For tests A–D, network #5 showed similar results for source depth as the other networks but had the best results for CPA range and ship speed. Both of these labels are predicted more accurately when the network is only trying to learn the one label even though the same variability exists in the other parameters in the training data.

The seabed type results for tests A–D lead to several observations about the ability to predict seabed type in these

https://doi.org/10.1121/10.0003361

cases. First, the SSP mismatch in tests A, C, and D caused seabed prediction accuracy to fall below 90%, compared to the higher accuracy of Test B when only the water depth was varied by ±0.5 m. While this decrease in accuracy is understandable, the relative behavior of the six networks is a bit surprising. While seabed classification in the MTL-CNN (#1 and #2) are expected to function independent of the seabed type ordering, a larger difference is expected between networks #3 and #4. However, the apparent similarity in the errors, even though one has seabeds ordered by bottom loss and the other does not, indicates that the CNN based on regression apparently learns a transformation that puts those seabeds 1–4 in a reasonable representation that works when only these four seabeds are used. Another observation is that networks #5 and #6 in which only the seabed label is learned—via regression and classification, respectively—have higher variability than networks that were trying to predict the source localization parameters simultaneously.

C. Seabed mismatch

Test E was designed to evaluate the impact of seabed mismatch. The testing data were created with six seabeds not used in training. While the source labels predictions are shown in Figs. 9(a)–9(c), the seabed type predictions for test E are shown in a separated panel in Fig. 9(e). The MAPE for source labels for test E are similar to test D, however, the variation is larger because of the seabed mismatch in test E. For the seabed predictions, the results have been divided based on the VGS seabed used in generating the data samples. For each VGS seabed, the horizontal axis corresponds to the six trained networks, and the bars show the percentage of times each of the representative seabed was selected by the CNN for all the data samples over the 20 instances of each network.

These seabed results [Fig. 9(e)] show that the networks can learn patterns that represent the overall impact of the sediment on the acoustic propagation. The MTL-CNNs and the CNN using ordered seabed labels (networks #1–#3) classify the first four VGS seabed types, corresponding to muddy environments, mainly as (1) deep mud or (2) mud over sand and the last two VGS environments, which correspond to sandy environments, as (3) sandy silt or (4) sand. Network #4, in which regression is used and the order of the seabeds is mixed, has more varied predictions indicating the inability of the network to interpolate between the four seabeds not ordered in a physically meaningful way. As in tests A–D, networks #5 and #6 also show more variability in the predicted seabed. The consistent seabed type predictions from the MTL-CNN provide encouragement that a representative seabed type can be learned even when the exact seabed parameterization is not included in the training dataset.

D. Experimental data

The trained models are tested on experimental data measured in two VLAs in the SBCEX 2017. [Details are given in Sec. II A and for convenience, the measurement configuration is also shown in Fig. 10(d).] The predictions for these six samples are shown in Fig. 10, along with the expected values (dashed lines). As expected from the simulated test cases, the ship speed predictions are most robust; the ship speeds are consistently predicted close to the ground truth of ∼3 kn. CPA range is underestimated, but the CNNs are able to distinguish between closer and longer ranges, with the latter predicted more accurately, although there is more uncertainty in the expected value for this data sample.

The source depth is underestimated as well but shows consistent predictions at about 30–40 m for networks #1–#4. However, the MTL-CNN predictions have a narrower distribution. For network #5, the variance increases. The large variance and bias in the source depth predictions can be related to mismatch in either SSP or the seabed, as shown with the simulated test cases. To further address the reason for underestimation in source depth, an additional study has been done to evaluate the data-model mismatch using a Bayesian approach.

For seabed classification, results from the six data samples for each of the networks are shown as bar plots in Fig. 10(e). Focusing on networks #1–#3, which performed best in the simulated test cases, the mud over sand and sandy silt seabeds are selected most of the time for data samples 1–5. For data sample 6, with the longest CPA range, the MTL-CNNs (#1 and #2) select mud over sand, and network #3 selects deep mud or mud over sand. The remaining networks #4–#6 show more variability as in the simulated test cases.

V. DISCUSSION

The results of these CNNs applied to mid-frequency towed tones offer insights that can help guide further developments of deep learning in ocean acoustics.

A. Impact of training data on generalization

While it is commonly thought that more training data yield better results, recent advances in machine learning have indicated that variety in the training data is more important. We also found the diversity of the training data is more important than the quantity of data. For example, a training data set with randomly selected source parameters outperformed one that was trained on a 20% larger dataset that had gridded (evenly spaced) source parameter values. This result indicates that inclusion of random parameters is more efficient than simply reducing the grid spacing and removes the need for the user to select the grid spacing. Network performance will be influenced by the diversity of source parameters and ocean environments included in the training data set.

B. Impact of loss function

The six networks compared in this paper show several differences even though the base architecture was the same. The MTL-CNNs (networks #1 and #2) perform identically

since the order of the seabed types is irrelevant when doing classification. When using only regression for predicting four labels simultaneously in CNNs #3 and #4, if the labels of the seabed type are sorted at random with no physical meaning, the performance of the network decreases, particularly in cases of seabed mismatch. On the other hand, if one parameter at a time is being learned as in CNN #5, the predictions for ship speed and CPA range show a higher accuracy, while accuracy in source depth is significantly decreased (especially in the case of seabed mismatch). The approach of learning a single label has also been used in Niu *et al*. for independently learning source range and depth *via* classification using a series of ResNets.[33]

In CNNs #5 and #6, when only the seabed type is learned, higher uncertainty and lower accuracy occur. This finding implies that for this data type in a CNN, learning all four labels allows the network to better generalize the seabed type predictions for either regression or classification tasks. This increased variability pertains to the mid-frequencies used in this work, as studies on pressure time series, band-passed filtered (30–2500 Hz), showed seabed only output worked well.[4]

The difference in predictions when using different loss functions highlights an important point: Based on the type of input data, the design of the network and the loss function need to optimized. While one might assume that when the network trains on all the parameters at the same time, it finds the optimal set of weights that minimizes all of the errors simultaneously, this is not always the optimal solution. For example, in this work, the CPA range and ship speed, shown in Figs. 9(a) and 9(b), have lower MAPE values when learned independently from the other parameters. Seabed classification, on the other hand, is more robust when learned simultaneously with the source parameters, as shown by the narrower distributions in Fig. 9(d). However, in other cases, seabed-only CNN predictions were shown to improve seabed predictions (based on one-second impulse pressure time series from a single sensor) compared to those in which range and seabed were found simultaneously through regression. Similarly, Niu *et al*.[33] report good results using a series of residual networks that invert for source range and depth separately (from 100 to 200 Hz broadband signals on a single sensor). These results emphasize that different options for the loss function should be investigated based on the type of input data of interest.

C. Impact of SSP variability

Tests on data simulated for different SSPs (particularly tests B and D) gives an indication of the impact of SSP variability on the CNN predictions. Ideally, the training data should include the SSP variability expected during *in situ* application of the CNN. However, tests A–D provide some insights into the behavior of the CNN on mid-frequency tones in the case of SSP mismatch between the training data and the *in situ* testing data. The estimations of source depth are most affected by variation in the SSP. In addition, SSP mismatch has a larger impact on estimates of ship speed and range when the seabed has higher bottom loss. The variability caused by SSP mismatch would be similar to those caused by a seabed slope.[56] These results emphasize the need to account for SSP variability when training neural networks for ocean acoustics.

D. Impact of seabed mismatch

Because of the infinite variety in the ocean, no training dataset can represent every possible seabed. Thus, this work has investigated the impact of seabed mismatch on the mid-frequency CNN results. Test E showed that testing on data samples simulated with seabed types not included in the training data showed that the variability in the source label predictions increased and tended to be underestimated. The predicted seabed type label, however, was reasonable using the two MTL-CNNs and the CNN with seabed types ordered according to bottom loss. The CNNs chose a seabed type in the training data with a similar bottom loss to the correct one. These results indicate that perhaps the CNNs learned something about the porosity of the seabed that was generalizable and highlights the potential ability of CNNs to find sensitive seabed parameters.

E. Application to measured data

The CNNs trained on simulated data were applied to the tonal levels from a towed source measured on 16-channel VLAs as a proof-of-concept that CNNs can learn from mid-frequency sounds. The biases and variances of the network predictions over 20 training instances, shown via the distributions in Fig. 10, are related to both source and environmental considerations:

- The largest complication with applying the trained CNNs to the measured data was the 50% duty cycle required during the experiment. To approximate a continuous signal, an enveloping function was used (Sec. II B). This data extraction process caused the measured samples to have considerably less detail than the simulated samples, which is why the training data were smoothed and downsampled.
- When extracting the data, determining CPA range (time) was difficult. The GPS data indicated when the ship was at CPA, but not the towed source. The start time for each extracted sample (in Fig. 2) was estimated by looking for the change in Doppler shift and level near CPA. Due to the complicated modal interference pattern, however, the different frequencies exhibited these changes at different times. The uncertainty in the location of the towed source at the beginning of the extracted samples may be linked to underestimations of range and speed due to the geometry assumed in the simulations. Data augmentation techniques that include time shifts could be included during training to overcome the CPA extraction issue.

https://doi.org/10.1121/10.0003361

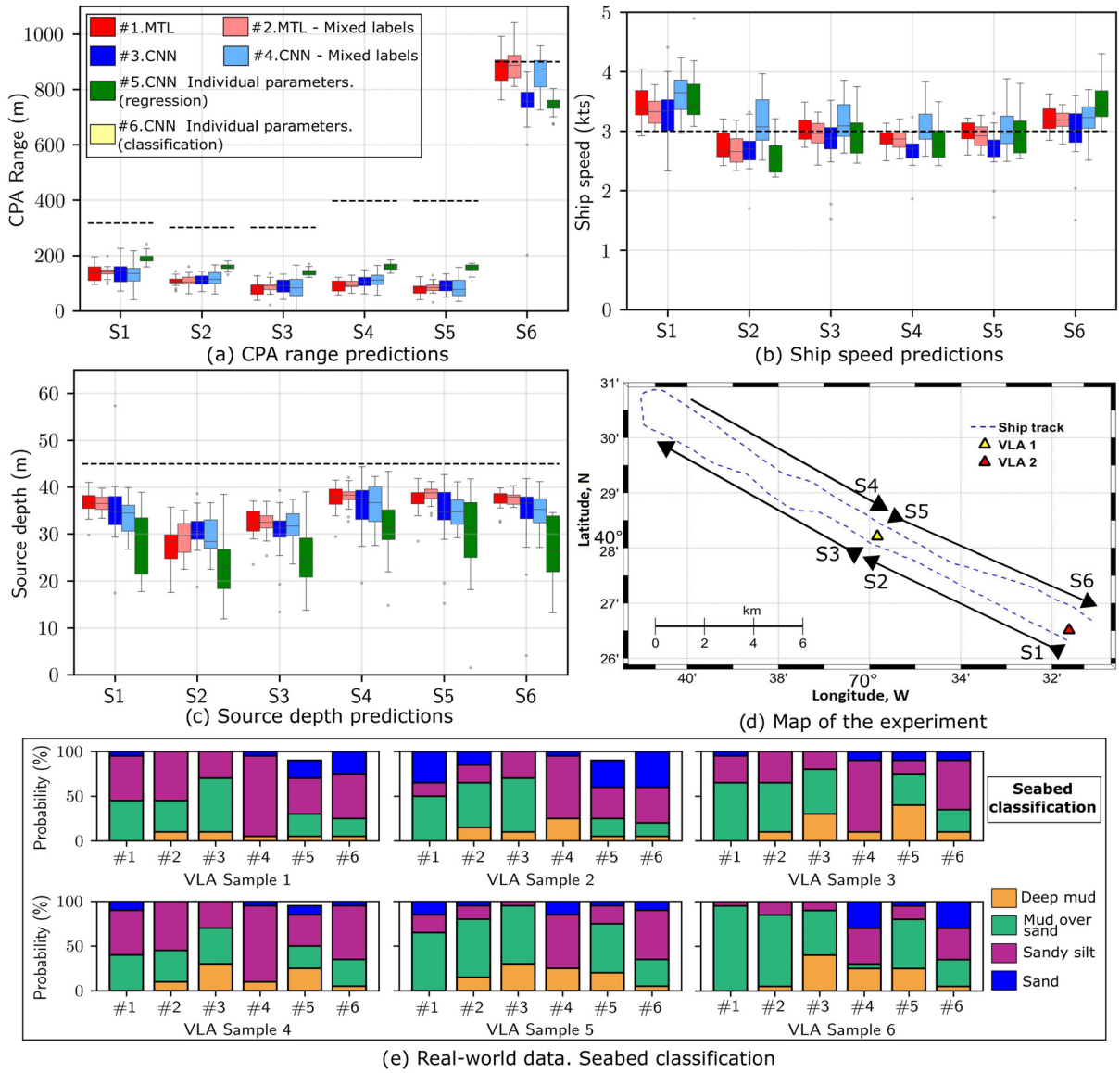

FIG. 10. (Color online) Experimental data results for the 20 instances of each network. Box and whisker plots of the predictions for (a) CPA range, (b) ship speed, and (c) source depth. (d) Source track for measured data samples S1–S6. (e) Predicted seabed labels, similar to those in Fig. 9, when tested on the six samples.

- Although the SNR was high for all six data samples, the data extraction process essentially increased the noise floor by using the peak windowing method that smoothed out the naturally occurring dips. This raised noise floor was approximated by the data preprocessing (smoothing) done on the synthetic data. In future applications, variable SNR can be accounted for by adding noise to the simulated training data.
- All 16 channels of each VLA were used but no modifications were introduced to account for array tilt. This omission was less significant than in other potential applications because this work used received levels and relatively large time blocks over which an enveloping function was applied when extracting the measured data. Receiver location variability could be included in future applications, especially if coherent processing or finer time resolution are included.
- The behavior of the predictions is similar to the simulated test cases, where the ship speed predictions are most accurate and the source depth is underestimated, as seen in cases of environmental mismatch

A limitation of this work was the small number of ocean environments used to simulate the training data. The water depth was constant for all samples. While this water depth was approximately that found at VLA2, it was approximately 3 m deeper than the measured water depth at VLA 1. The simulated training data came from three measured SSPs and four seabed types, representative of a deep mud, mud over sand, sandy silt, and sand sediment profiles obtained from previous geoacoustic inversions.[43–46] Even though four is a limited number of seabeds, the CNNs appear to distinguish between the mud-related seabeds with an angle of intromission and the sand-based seabeds. A significant amount of work is needed to determine the optimal

way to include environmental variability in the training data.

Environmental mismatch introduce biases and increase variances in the CNN results and, thus, need to be considered carefully in future applications of machine and deep learning.

VI. CONCLUSIONS

Convolutional neural networks (CNNs) were trained on simulated data and tested on mid-frequency tonal levels from a moving source in SBCEX 2017. The input data consisted of five tonal levels on a 16-channel VLA over approximately 74 min. Considerations for designing the CNN were explored using a single base architecture and different output layers and loss functions designed to obtain values for the source depth, speed, and CPA range, and a seabed type (1–4 as shown in Fig. 5). The advantages of multitask learning were shown on both simulated test cases and measured data.

The performance of the networks with different output layers and labels were compared in the face of environmental mismatch. Each network type was trained 20 times to account for uncertainties due to the random initializations. Data from five testing cases were then passed through the trained networks. The statistical distributions of the mean absolute percentage error in each source label were shown along with the accuracy of the seabed label.

Test cases with increasing degrees of SSP mismatch showed increased error in predictions of the source labels, however, seabed type accuracy was still above 70%. The impact of SSP mismatch was more significant for data samples simulated with the deep mud seabed type than the other (more reflective) seabeds. Thus, appropriate ways to include SSP variability in the training data must be found to account for the spatial and temporal changes in the real ocean.

The trained CNNs were also applied to data simulated for seabed mismatch. The CNNs predicted a seabed type 1–4 that matched the porosity of the different seabeds used in the test case. This seabed mismatch, however, produced a general bias towards smaller values of all the source parameters and increased the variance in the source label predictions. These results point to the potential for seabed type characterization using MTL-CNNs and emphasize the importance of including realistic and representative seabed variations in the training data.

The trained networks were applied to six data samples extracted from the SBCEX 2017 measurements. Downsampling and smoothing the simulated data approximated the data extraction process (required by the 50% duty cycle), although biases were still apparent in CPA range and source depth. The ship speed was the most reliably predicted label by all the networks. Source depths and ranges were underestimated, similar to what was seen in the tests with simulated seabed type mismatch. The seabed type predictions were generally around the expected seabed type using the MTL-CNN and the CNN based on regression for all four labels, as long as the seabeds were ordered according to bottom loss.

This work demonstrated the importance of accounting for environmental variability and network uncertainties when developing machine and deep learning algorithms for applications in a dynamic ocean. The variability of the training data can help avoid overfitting. The homoscedastic uncertainty associated with the random initializations of weights and random draws during training must be understood. Questions remain as to the best form of input data for source localization and tracking and seabed type predictions. In all cases, however, the variability of the environment must be accounted for in training machine and deep learning algorithms for ocean applications.

ACKNOWLEDGMENTS

This work was funded by Office of Naval Research Contract No. N00014-19-C-2001. The Seabed Characterization Experiment 2017 was funded by the U.S. Navy Office of Naval Research. The authors acknowledge and thank the captains and crews of RV Endeavor. We also thank David Ensberg for assistance with the measured data and the reviewers for their helpful comments.

[1]M. Siderius, P. L. Nielsen, J. Sellschopp, M. Snellen, and D. Simons, "Experimental study of geo-acoustic inversion uncertainty due to ocean sound-speed fluctuations," J. Acoust. Soc. Am. **110**(2), 769–781 (2001).
[2]C.-F. Huang, P. Gerstoft, and W. S. Hodgkiss, "Effect of ocean sound speed uncertainty on matched-field geoacoustic inversion," J. Acoust. Soc. Am. **123**(6), EL162–EL168 (2008).
[3]S. E. Dosso and J. Dettmer, "Bayesian matched-field geoacoustic inversion," Inv. Probl. **27**(5), 055009 (2011).
[4]D. F. Van Komen, T. B. Neilsen, K. Howarth, D. P. Knobles, and P. H. Dahl, "Seabed and range estimation of impulsive time series using a convolutional neural network," J. Acoust. Soc. Am. **147**(5), EL403–EL408 (2020).
[5]P. S. Wilson, D. P. Knobles, and T. B. Neilsen, "Guest editorial an overview of the seabed characterization experiment," IEEE J. Oceanic Eng. **45**(1), 1–13 (2020).
[6]A. B. Baggeroer, W. A. Kuperman, and H. Schmidt, "Matched field processing: Source localization in correlated noise as an optimum parameter estimation problem," J. Acoust. Soc. Am. **83**(2), 571–587 (1988).
[7]E. K. Westwood, "Broadband matched–field source localization," J. Acoust. Soc. Am. **91**(5), 2777–2789 (1992).
[8]A. Tolstoy, *Matched Field Processing for Underwater Acoustics* (World Scientific, Singapore, 1993).
[9]M. B. Porter and A. Tolstoy, "The matched field processing benchmark problems," J. Comput. Acoust. **02**(03), 161–185 (1994).
[10]E. K. Westwood and D. P. Knobles, "Source track localization via multipath correlation matching," J. Acoust. Soc. Am. **102**(5), 2645–2654 (1997).
[11]M. S. Ballard, "Estimation of source range using horizontal multipath in continental shelf environments," J. Acoust. Soc. Am. **134**(4), EL340–EL344 (2013).
[12]I. Zorych and Z.-H. Michalopoulou, "Particle filtering for dispersion curve tracking in ocean acoustics," J. Acoust. Soc. Am. **124**(2), EL45–EL50 (2008).
[13]S. E. Dosso and M. J. Wilmut, "Bayesian focalization: Quantifying source localization with environmental uncertainty," J. Acoust. Soc. Am. **121**(5), 2567–2574 (2007).
[14]J. Dettmer, S. E. Dosso, and C. W. Holland, "Trans-dimensional geoacoustic inversion," J. Acoust. Soc. Am. **128**(6), 3393–3405 (2010).
[15]M. D. Collins, W. A. Kuperman, and H. Schmidt, "Nonlinear inversion for ocean–bottom properties," J. Acoust. Soc. Am. **92**(5), 2770–2783 (1992).

[16]P. Gerstoft, "Inversion of seismoacoustic data using genetic algorithms and *a posteriori* probability distributions," J. Acoust. Soc. Am. **95**(2), 770–782 (1994).
[17]M. R. Fallat, P. L. Nielsen, and S. E. Dosso, "Hybrid geoacoustic inversion of broadband Mediterranean Sea data," J. Acoust. Soc. Am. **107**(4), 1967–1977 (2000).
[18]T. B. Neilsen and E. K. Westwood, "Extraction of acoustic normal mode depth functions using vertical line array data," J. Acoust. Soc. Am. **111**(2), 748–756 (2002).
[19]J. W. Choi and P. H. Dahl, "Mid-to-high-frequency bottom loss in the East China Sea," IEEE J. Oceanic Eng. **29**(4), 980–987 (2004).
[20]C. W. Holland and S. E. Dosso, "Mid frequency shallow water fine-grained sediment attenuation measurements," J. Acoust. Soc. Am. **134**(1), 131–143 (2013).
[21]J. Yang, D. R. Jackson, and D. Tang, "Mid-frequency geoacoustic inversion using bottom loss data from the Shallow Water 2006 Experiment," J. Acoust. Soc. Am. **131**(2), 1711–1721 (2012).
[22]Z.-H. Michalopoulou and U. Ghosh-Dastidar, "Tabu for matched-field source localization and geoacoustic inversion," J. Acoust. Soc. Am. **115**(1), 135–145 (2004).
[23]T. B. Neilsen, "An iterative implementation of rotated coordinates for inverse problems," J. Acoust. Soc. Am. **113**(5), 2574–2586 (2003).
[24]M. J. Bianco, P. Gerstoft, J. Traer, E. Ozanich, M. A. Roch, S. Gannot, and C.-A. Deledalle, "Machine learning in acoustics: Theory and applications," J. Acoust. Soc. Am. **146**(5), 3590–3628 (2019).
[25]B. Z. Steinberg, M. J. Beran, S. H. Chin, and J. H. Howard, Jr., "A neural network approach to source localization," J. Acoust. Soc. Am. **90**(4), 2081–2090 (1991).
[26]L. Houégnigan, P. Safari, C. Nadeu, M. van der Schaar, and M. André, "A novel approach to real-time range estimation of underwater acoustic sources using supervised machine learning," in *OCEANS 2017-Aberdeen*, IEEE (2017), pp. 1–5.
[27]H. Niu, E. Ozanich, and P. Gerstoft, "Ship localization in Santa Barbara Channel using machine learning classifiers," J. Acoust. Soc. Am. **142**(5), EL455–EL460 (2017).
[28]H. Niu, E. Reeves, and P. Gerstoft, "Source localization in an ocean waveguide using supervised machine learning," J. Acoust. Soc. Am. **142**(3), 1176–1188 (2017).
[29]R. Lefort, G. Real, and A. Drémeau, "Direct regressions for underwater acoustic source localization in fluctuating oceans," J. Appl. Acoust. **116**, 303–310 (2017).
[30]Z. Huang, J. Xu, Z. Gong, H. Wang, and Y. Yan, "Source localization using deep neural networks in a shallow water environment," J. Acoust. Soc. Am. **143**(5), 2922–2932 (2018).
[31]Y. Wang and H. Peng, "Underwater acoustic source localization using generalized regression neural network," J. Acoust. Soc. Am. **143**(4), 2321–2331 (2018).
[32]E. Ozanich, P. Gerstoft, and H. Niu, "A feedforward neural network for direction-of-arrival estimation," J. Acoust. Soc. Am. **147**(3), 2035–2048 (2020).
[33]H. Niu, Z. Gong, E. Ozanich, P. Gerstoft, H. Wang, and Z. Li, "Deep-learning source localization using multi-frequency magnitude-only data," J. Acoust. Soc. Am. **146**(1), 211–222 (2019).
[34]Z.-H. Michalopoulou, D. Alexandrou, and C. de Moustier, "Application of neural and statistical classifiers to the problem of seafloor characterization," IEEE J. Oceanic Eng. **20**(3), 190–197 (1995).
[35]J. Benson, N. R. Chapman, and A. Antoniou, "Geoacoustic model inversion using artificial neural networks," Inv. Probl. **16**(6), 1627–1639 (2000).
[36]Y. Stephan, X. Demoulin, and O. Sarzeaud, "Neural direct approaches for geoacoustic inversion," J. Comput. Acoust. **06**, 151–166 (1998).
[37]J. Piccolo, G. Haramuniz, and Z.-H. Michalopoulou, "Geoacoustic inversion with generalized additive models," J. Acoustical Soc. America **145**(6), EL463–EL468 (2019).
[38]H. Niu, P. Gerstoft, E. Ozanich, Z. Li, R. Zhang, Z. Gong, and H. Wang, "Block sparse Bayesian learning for broadband mode extraction in shallow water from a vertical array," J. Acoust. Soc. Am. **147**, 3729–3739 (2020).
[39]C. Frederick, S. Villar, and Z.-H. Michalopoulou, "Seabed classification using physics-based modeling and machine learning," J. Acoust. Soc. Am. **148**(6), 859–872 (2020).
[40]A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 18–22 June 2018, pp. 7482–7491.
[41]Z.-H. Michalopoulou and P. Gerstoft, "Multipath broadband localization, bathymetry, and sediment inversion," IEEE J. Oceanic Eng. **45**(1), 92–102 (2020).
[42]E. K. Westwood, C. T. Tindle, and N. R. Chapman, "A normal mode model for acousto–elastic ocean environments," J. Acoust. Soc. Am. **100**(6), 3631–3645 (1996).
[43]D. P. Knobles, R. A. Koch, L. A. Thompson, K. C. Focke, and P. E. Eisman, "Broadband sound propagation in shallow water and geoacoustic inversion," J. Acoust. Soc. Am. **113**(1), 205–222 (2003).
[44]D. P. Knobles, P. S. Wilson, J. A. Goff, L. Wan, M. J. Buckingham, J. D. Chaytor, and M. Badiey, "Maximum entropy derived statistics of sound-speed structure in a fine-grained sediment inferred from sparse broadband acoustic measurements on the New England continental shelf," IEEE J. Oceanic Eng. **45**, 161–173 (2020).
[45]G. R. Potty, J. H. Miller, and J. F. Lynch, "Inversion for sediment geoacoustic properties at the New England bight," J. Acoust. Soc. Am. **114**(4), 1874–1887 (2003).
[46]J.-X. Zhou, X.-Z. Zhang, and D. P. Knobles, "Low-frequency geoacoustic model for the effective properties of sandy seabottoms," J. Acoust. Soc. Am. **125**(5), 2847–2866 (2009).
[47]M. J. Buckingham, "Compressional and shear wave properties of marine sediments: Comparisons between theory and data," J. Acoust. Soc. Am. **117**(1), 137–152 (2005).
[48]M. J. Buckingham, "On pore-fluid viscosity and the wave properties of saturated granular materials including marine sediments," J. Acoust. Soc. Am. **122**(3), 1486–1501 (2007).
[49]Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE **86**(11), 2278–2323 (1998).
[50]I. Goodfellow, Y. Bengio, and, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016).
[51]A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, and L. Antiga, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems* (2019), pp. 8024–8035.
[52]D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," arXiv:1412.6980 (2014).
[53]V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)* (Omnipress, Madison, WI, 2010). pp. 801–814.
[54]C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1995).
[55]Z. Allen-Zhu and Y. Li, "What can ResNet learn efficiently, going beyond kernels?," in *33rd Conference on Neural Information Processing Systems* (2019), pp. 1–12.
[56]P. Abbot, S. Celuzza, I. Dyer, B. Gomes, J. Fulford, J. Lynch, G. Gawarkiewicz, and D. Volak, "Acoustic propagation," IEEE J. Oceanic Eng. **26**(2), 266–284 (2001).