





MARCH 26 2025

## Ensemble approach to deep learning seabed classification using multichannel ship noise<sup>a)</sup>

Ginger E. Lau ; Michael C. Mortenson ; Tracianne B. Neilsen ; David F. Van Komen ; William S. Hodgkiss; David P. Knobles



*J. Acoust. Soc. Am.* 157, 2127–2149 (2025)  
<https://doi.org/10.1121/10.0036221>



### Articles You May Be Interested In

Seabed classification from merchant ship-radiated noise using a physics-based ensemble of deep learning algorithms

*J. Acoust. Soc. Am.* (August 2021)

Estimation of seabed properties and range from vector acoustic observations of underwater ship noise

*J. Acoust. Soc. Am.* (April 2020)

Trans-dimensional inversion for seafloor properties for three mud depocenters on the New England shelf under dynamical oceanographic conditions

*J. Acoust. Soc. Am.* (March 2024)




**ASA**

Advance your science and career as a member of the **Acoustical Society of America**

[LEARN MORE](#)

## Ensemble approach to deep learning seabed classification using multichannel ship noise<sup>a)</sup>

Ginger E. Lau,<sup>1</sup>  Michael C. Mortenson,<sup>2</sup>  Tracianne B. Neilsen,<sup>2,b)</sup>  David F. Van Komen,<sup>3</sup>   
 William S. Hodgkiss,<sup>4</sup> and David P. Knobles<sup>5</sup>

<sup>1</sup>*Department of Physics, Emory University, Atlanta, Georgia 30322, USA*

<sup>2</sup>*Department of Physics and Astronomy, Brigham Young University, Provo, Utah 84602, USA*

<sup>3</sup>*Kahlert School of Computing, University of Utah, Salt Lake City, Utah, 84112, USA*

<sup>4</sup>*Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, California 92093, USA*

<sup>5</sup>*The Platt Institute, Austin, Texas 78731, USA*

### ABSTRACT:

In shallow-water downward-refracting ocean environments, hydrophone measurements of shipping noise encode information about the seabed. In this study, neural networks are trained on synthetic data to predict seabed classes from multichannel hydrophone spectrograms of shipping noise. Specifically, ResNet-18 networks are trained on different combinations of synthetic inputs from one, two, four, and eight channels. The trained networks are then applied to measured ship spectrograms from the Seabed Characterization Experiment 2017 (SBCEX 2017) to obtain an effective seabed class for the area. Data preprocessing techniques and ensemble modeling are leveraged to improve performance over previous studies. The results showcase the predictive capability of the trained networks; the seabed predictions from the measured ship spectrograms tend towards two seabed classes that share similarities in the upper few meters of sediment and are consistent with geoacoustic inversion results from SBCEX 2017. This work also demonstrates how ensemble modeling yields a measure of precision and confidence in the predicted results. Furthermore, the impact of using data from multiple hydrophone channels is quantified. While the water sound speed in this experiment was only slightly upward refracting, we anticipate increased advantages of using multiple channels to train neural networks for more varied sound speed profiles.

© 2025 Acoustical Society of America. <https://doi.org/10.1121/10.0036221>

(Received 24 August 2024; revised 26 February 2025; accepted 1 March 2025; published online 26 March 2025)

[Editor: James F. Lynch]

Pages: 2127–2149

### I. INTRODUCTION

Different seabed environments have acoustically distinct effects on underwater sound propagation. These effects are encoded in the sound signal as it travels from source to receiver. For example, noise from a transiting cargo ship or tanker encodes information about the ocean environment that can be used to obtain an estimate of effective seabed along the propagation path, capturing the cumulative impact of seabed interactions as sound travels from ships to receivers. Noise from these ships of opportunity (SOO) has been used to infer seabed properties with traditional optimization algorithms, and more recently, with deep learning. In this work, seabed inference is cast as a classification problem to bypass the parameter identifiability complications associated with parameter regression, i.e., the non-uniqueness caused by parameter correlations<sup>1</sup> and sloppiness.<sup>2</sup> By selecting catalogs of representative seabeds that are not too acoustically similar,<sup>3,4</sup> a residual convolutional

neural network (CNN) can be trained to predict seabed classes from ship noise spectrograms.

Ships of opportunity noise has been used for seabed characterization in many optimization algorithms. Several have noted that SOO noise contains enough information to obtain effective seabed properties. Studies have used one hydrophone,<sup>5,6</sup> a vertical line array,<sup>7–12</sup> a moored horizontal line array (HLA),<sup>13–16</sup> an L-shaped array,<sup>17</sup> a towed HLA,<sup>18,19</sup> a vector sensor,<sup>20,21</sup> and one hydrophone on an autonomous underwater vehicle.<sup>22</sup> Geoacoustic inversions with SOO noise have been performed with least squares based search algorithms,<sup>13,15,18,19,23</sup> blind ray deconvolution,<sup>7,11</sup> joint time-frequency inversion,<sup>10</sup> and Bayesian optimizations.<sup>14,16,21,22,24</sup> Studies have used lower frequencies (20–140 Hz)<sup>21</sup> as well as higher bands (1700–3300 Hz) for parameter estimation.<sup>11</sup>

As SOO noise has been used for geoacoustic inversion, researchers have noted several challenges and potential solutions. Nicholas *et al.*<sup>17</sup> noted that the vertical aperture of an L array provided environmental information, ship range, and depth information, while the horizontal aperture provided ship bearing information. Heaney<sup>5</sup> noted the challenges of predicting the full ocean response due to environmental mismatch between the model and data, compounded by variability with range; he instead moved to the

<sup>a)</sup>This paper is part of a special issue on Assessing Sediment Heterogeneity on Continental Shelves and Slopes.

<sup>b)</sup>Email: tbn@byu.edu

prediction of average field levels and time spreads of the acoustic field in the ocean. Tollefsen and Dosso<sup>14</sup> quantified the information content of five low frequency components received on a moored HLA and concluded that problems with low signal-to-noise ratio (SNR) could be overcome by combining multiple data segments. Stotts *et al.*<sup>23</sup> used 10–250 Hz spectrograms to examine the impact of interfering noise signals on the parameter estimates. Gervaise *et al.*<sup>6</sup> concluded that attenuation cannot be estimated accurately with their inversion scheme. Due to these limitations, we choose a seabed classification approach in which the closest effective seabed is selected instead of inferring individual parameters.

Machine and deep learning techniques also can be used to infer seabed properties.<sup>25,26</sup> Some approaches include neural and statistical classifiers,<sup>27</sup> artificial neural networks,<sup>28,29</sup> generalized additive models,<sup>30</sup> Bayesian learning,<sup>9,31</sup> and CNNs.<sup>32–35</sup> In these studies, the learned labels might be modal properties,<sup>31,36</sup> the channel impulse response,<sup>9</sup> individual seabed parameter values,<sup>29,30,37</sup> or a seabed type or class.<sup>32–35,38</sup> Different types of sound sources have been used for this purpose, including continuous wave signals,<sup>28,33,38</sup> linear frequency modulated signals,<sup>30</sup> and impulses.<sup>31,32,37</sup>

Most closely related to the current work are studies that use SOO noise and supervised learning. Van Komen *et al.*<sup>35</sup> used a CNN to obtain ranges and ship speed on synthetic spectrograms while providing seabed type predictions. Additionally, the difference between the types of input data: complex spectra, magnitude of the spectral density, and spectral density levels were investigated. They found that the complex spectral density input was better for range estimation, and spectral density levels were best for seabed classification. While the study only considered four representative seabed classes, a subsequent study by Escobar–Amado *et al.*<sup>34</sup> applied six different CNN architectures to classify between 34 seabeds and used the trained networks on SOO spectrograms measured during the Seabed Characterization Experiment 2017 (SBCEX 2017). Their work found that ResNet-18 gave the most consistent results because of the residual skip connections. The catalog of 34 seabeds was determined using the Pearson correlation of broadband transmission loss (TL) over range, as described in Forman *et al.*<sup>3</sup> While this 34 seabed catalog has performed well for the SOO spectrograms over the 360–1100 Hz frequency band of interest, the same approach could be used to identify an appropriate seabed catalog for different frequency bands or other types of input data.

The studies noted previously used only one hydrophone located in the middle of the water column. The main contributions of the current work are (1) to evaluate the impact of using multiple receivers, (2) to illustrate the benefits of ensemble learning, (3) to determine the most likely seabed class in the New England Mud Patch region, and (4) to introduce a new type of data mapping as a preprocessing step to facilitate more efficient deep learning. Care is taken to avoid data leakage (described in Sec. II E), which was

present in previous studies, in order to improve the ability of the trained networks to generalize. The ResNet-18 classification networks are trained using SOO spectrograms on combinations of one, two, four, or eight channels on a vertical line array (VLA). The trained networks are then applied to SOO spectrograms measured during SBCEX 2017. For SOO spectrograms with clear striation patterns, the trained networks predict one of two seabed classes. Our discussion of these results highlights important factors in developing generalizable deep learning algorithms for ocean acoustics applications that provide a measure of confidence through ensemble learning.

## II. METHODS

### A. Measured data

The multidisciplinary SBCEX 2017 studied a region of the continental shelf in the North Atlantic Ocean. While the experimental site, roughly 95 km south of Martha’s Vineyard, Massachusetts, is referred to as the “New England Mud Patch,” the seabed there is composed of a variety of sediments including mud, clay, silt, and sand. The bathymetry of the site is relatively flat, approximately 75 m deep, and is known to have a large region with relatively thick and uniform upper layer of fine-grained sediment. SBCEX 2017 occurred in March and April to take advantage of a stable, nearly isothermal the water column.<sup>39</sup>

The seabed in the experimental area has been studied extensively. Results of a full-waveform chirp sonar survey show a layer of mud on top of sand in Fig. 3 of Wilson *et al.*<sup>39</sup> The two-way acoustic travel time map in Fig. 2 of Wilson *et al.* indicates that the thickness of the mud varies around the measurement area up to a thickness of 12 m. For the sound speed in this first sediment layer, each of the 19 published estimates included in Fig. 7 of Wilson *et al.* found values near unity for the frequency-dependent sound speed ratio across the sediment-water interface and evidence of complexity in the exact layering of the sediment.<sup>39</sup>

Two identical 16-channel VLAs were deployed in different locations by the Marine Physical Laboratory of the Scripps Institution of Oceanography. VLA 1 and VLA 2 each contained 16 hydrophones spaced 3.75 m apart. The lowest hydrophone (#16) was located 3 m above the seafloor. A diagram illustrating the spacing of hydrophones is shown in Fig. 1. In previous deep learning seabed classification,<sup>34,35</sup> only data from the center hydrophone on these VLAs (#8, located 33 m from the seafloor) were considered. This work explores how using data from multiple sensors impacts the performance of a CNN and provides guidelines for the depths and spacing of the sensors to yield the most consistent results. Specifically, the input data for the CNNs are different combinations of one, two, four, and eight channels from the VLA.

In addition to the structured experiments conducted during SBCEX 2017, the VLAs recorded merchant vessels traveling in nearby shipping lanes. A schematic of the locations of the VLAs and the shipping lanes is shown in Fig. 2.

These passing merchant ships, which served as SOO, were identified using vessel traffic data collected by the U.S. Coast Guard through the Automatic Identification System (AIS) and retrieved from the Marine Cadastre.<sup>40</sup> An estimate for the ship speed in knots and closest-point-of-approach (CPA) distance to the VLA in km were calculated. A subset of seven of these measured ship spectrograms are used in this paper and listed in Table I.

For each channel, the time waveform over a 20 min window centered around CPA is used to generate spectrograms of the SOO passing the array. The sampling frequency of these waveforms is 25 kHz. A fast Fourier transform (FFT) is performed such that a frequency spacing of  $\Delta f = 6.016$  Hz is obtained, which matches the frequencies at which the synthetic spectrograms are computed (as described in Sec. II B). To achieve this  $\Delta f = 6.016$  Hz, the FFT is performed on blocks with  $n_s = 8310$  samples and 50% overlap. Also, a Hanning window was applied (from the numpy Python package). From each spectrum, frequencies from 360 to 1100 Hz are selected. The frequency range was chosen to be 360–1100 Hz to avoid corruption from the 350 Hz tones produced by research vessels during the experiment (the *RV Endeavor* emitted a loud source of noise in the 300–350 Hz band).

The previously noted FFT parameters are chosen for our specific application of deep learning. With the memory constraints often encountered while training deep learning networks, decisions must be made regarding the relative importance of the resolution of a single data sample compared to the total number of samples included in a training dataset. To evaluate the impact of the spectrogram resolution on the predicted seabed, Amos *et al.*<sup>41</sup> compared results

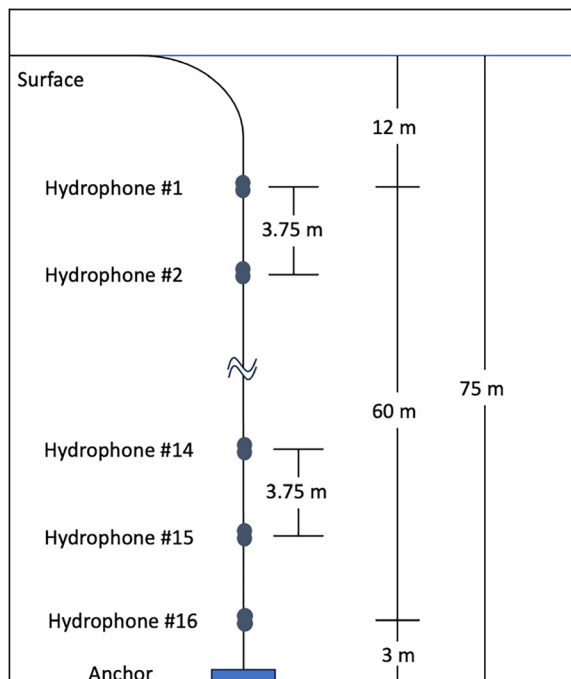


FIG. 1. Schematic of the vertical line arrays deployed in SBCEX 2017.

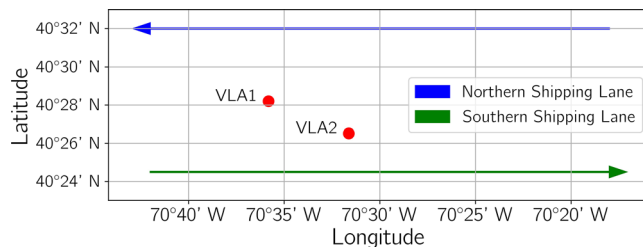


FIG. 2. SBCEX 2017 site in the New England Mud Patch. The deployment locations of VLA 1 and VLA 2 are marked for reference, as well as the shipping lanes. The SOO used in this work traveled in these shipping lanes, with the exception of the *Viking Bravery*, which traversed a path that was approximately parallel with the direction of a vector between VLA 1 and VLA 2.

for three frequency bands and time intervals. This study showed that using only 123 frequencies over the 360–1100 Hz band obtains the same results as using more frequencies. The 20 min time interval with 10 s between the spectra was found to be an ideal balance between using fewer times and obtaining consistent results. The FFT processing over the 20 min yields 7219 total time steps, but only the 121 time steps (spaced 10 s apart) centered on the CPA are selected for the input data samples. While this is a lower resolution than often used in ocean acoustics, data management considerations and memory limitations motivate the use of fewer frequencies and time steps.

No time averaging is done over the 10 s interval. This approach effectively produces a spectrogram consisting of snapshots of the spectra that are 10 s apart. Comparisons of a spectrogram before and after the time resampling are shown in Figs. 3(a) and 3(b). Although the resulting spectrograms are less smooth in appearance, this approach better matches how the synthetic spectrograms are calculated at discrete ranges, as explained in Sec. II B.

The resulting measured SOO data samples used by the ResNet-18 networks for prediction contain spectral density levels in dB re  $1 \mu\text{Pa}^2/\text{Hz}$ . As a visual example, spectral density levels from two of the ships are shown. Spectrograms from *MSC Kalamata* and *Hafnia Green* measured on VLA 2 at four different channel depths are shown on the left columns of Figs. 4 and 5. The “bathtub” striation pattern typical of SOO spectrograms (and other moving sources) is due to the shifting of

TABLE I. Details about seven transiting SOO (cargo ships and tankers) that have strong striations in their spectrograms. CPA range and ship speed are estimated from AIS data (Ref. 40). Ship transits occurred from Julian Days 83–90 in 2017.

Ship name	Type	VLA	CPA (km)	Speed (kt)	Day
<i>Corrido</i>	Tanker	2	3.95	14.6	84
<i>MSC Kalamata</i>	Cargo	2	3.11	16.7	83
<i>MSC Kalamata</i>	Cargo	1	5.95	16.7	83
<i>Tombarra</i>	Cargo	2	3.24	16.3	90
<i>Viking Bravery</i>	Cargo	2	3.09	14.7	90
<i>Viking Bravery</i>	Cargo	1	3.30	14.7	90
<i>Hafnia Green</i>	Tanker	2	2.77	10.9	89



interference patterns as the ship approaches and then leaves CPA.<sup>5</sup> The impact of receiver depth on the spectrogram images also can be observed. These inter-channel differences are caused by interference effects and encode information about the seafloor due to the phase of the Green’s function during shallow water propagation. This present study seeks to determine how the additional information affects the performance of the trained deep learning networks.

**B. Synthetic SOO spectrograms**

Due to a lack of labeled measured data, we train the neural networks using synthetic data. To train a neural

network on a synthetic dataset for generalization to measured data, the synthetic data must capture the character and variation of the measured data samples, i.e., sufficient overlap is needed in the distributions of the real and synthetic datasets. To this end, the received levels (RL) for the synthetic dataset are simulated by combining the empirical source spectral levels (SL) for shipping noise developed by Wales and Heitmeyer<sup>42</sup> with modeled TL,

$$RL(f, \theta, z_s, z_r, \mathbf{r}) = SL(f) - TL(f, \theta, z_s, z_r, \mathbf{r}), \quad (1)$$

where  $f$  is frequency,  $\theta$  is a vector of environmental parameters,  $z_s$  is the source depth,  $z_r$  is the receiver depth, and  $\mathbf{r}$  is a vector of source-receiver ranges. The environmental parameters in  $\theta$  include the water depth, the water sound speed profile (SSP), and the geoacoustic profile of the seabed. The geoacoustic profile includes parameters such as sound speed, compressional attenuation, density, shear speed and attenuation, and thickness of the layers. Receiver depths correspond with measured hydrophone depths as shown in Fig. 1.

The source spectrum of radiated ship noise is approximated by the empirical source model described by the unlabeled equation in Sec. 3 of Wales and Heitmeyer.<sup>42</sup> The source spectral density level in dB/Hz as a function of frequency,  $f$ , is calculated as

$$SL(f) = S_0 - 10\log(f^{3.594}) + 10\log((1 + (f/340)^2)^{0.917}), \quad (2)$$

where  $S_0$  might be described as the “y-intercept” of the source spectrum. Wales and Heitmeyer recommend a value of  $S_0 = 230$  dB, but a randomly selected value between 220 and 240 dB can simulate ship noise corresponding to different overall levels.

The TL is modeled using ORCA, a range-independent normal-mode model for acousto-elastic sound propagation.<sup>43</sup> ORCA treats the ocean environment as a waveguide and computes the frequency-dependent modal eigenvalues  $k_n(f, \theta)$  and normalized depth-dependent mode functions  $\bar{\phi}_n(f, z, \theta)$ . The mode functions are vertical standing waves caused by the interference of downward and upward traveling waves at specific angles.

These modal eigenvalues and depth-dependent mode functions are used to calculate the complex Green’s function for the wave equation in the ocean environment  $\theta$  for each source-receiver combination at each frequency  $f$ ,

$$p(z_s, z_r, r) \approx \sqrt{2\pi e^{i\pi/4}} \frac{1}{\rho_s} \sum_n \bar{\phi}_n(z_s) \bar{\phi}_n(z_r) \frac{e^{ik_n r}}{\sqrt{k_n r}}, \quad (3)$$

where  $\phi_n(z_s)$  and  $\phi_n(z_r)$  are the  $n$ th depth-dependent normal mode functions evaluated at the source depth  $z_s$  and receiver depth  $z_r$ , respectively;  $k_n(f)$  is the  $n$ th modal eigenvalue;  $\rho_s$  is the density at the source depth; and  $r$  is the horizontal range between source and receiver.

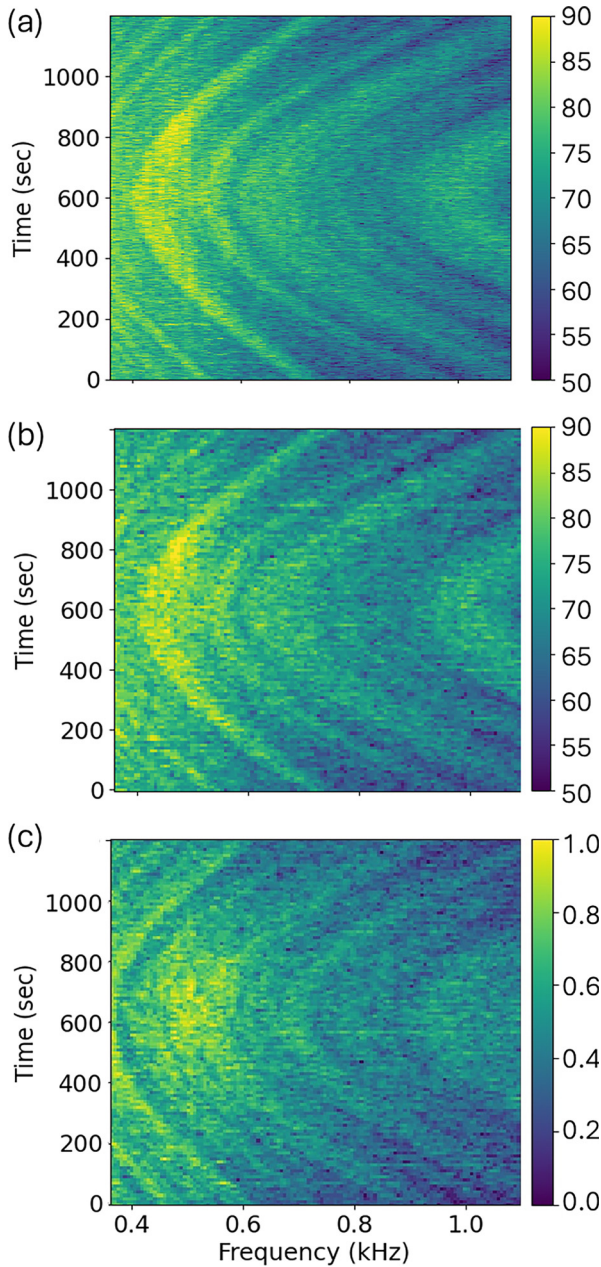


FIG. 3. Data preprocessing steps applied to measured SOO spectrograms. (a) The bandpassed spectrogram with 123 frequencies. (b) Spectrogram resampled to the coarser time sampling of 121 time steps over 20 min. (c) Spectrogram mapped to a domain of [0,1].

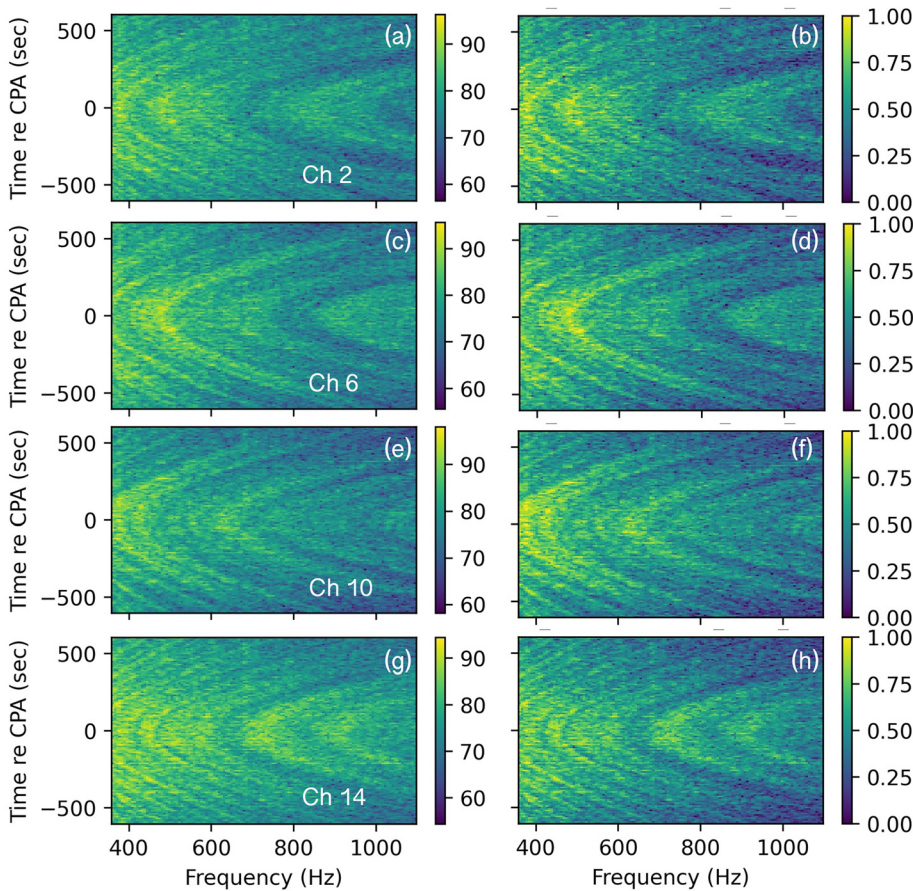


FIG. 4. SOO spectrograms of the *MSC Kalamata* ship during SBCEX 2017 from Channels #2 (a) and (b); #6 (c) and (d); #10 (e) and (f); and #14 (g) and (h) on VLA 2. The left column shows the extracted spectrograms and the right column shows the same spectrograms after they have been mapped, as described in Sec. II E 1. The AIS data for *MSC Kalamata* gives an approximate speed of 16.7 kt and a CPA range of 3.11 km from VLA 2.

From the modulus of the Green’s function, the frequency-dependent TL is calculated for each source-receiver configuration for each environment  $\theta$ ,

$$TL(z_s, z_r, \mathbf{r}) = -20 \log_{10}(|p(z_s, z_r, \mathbf{r})/p_0|), \quad (4)$$

with TL in dB relative to  $p_0$ , the pressure 1 m from the source. Examples of synthetic spectrograms are displayed in Fig. 6 for different  $\theta$ . These spectrograms have been mapped to facilitate more effective deep learning, as discussed in Sec. II E 1.

### C. Environmental parameters

The sound propagation model requires the ocean environment  $\theta$  be parameterized using sound speed, density, and attenuation as a function of depth both in the water column and in the seabed (shear parameters can also be included in the seabed). Since the measured data were obtained during SBCEX 2017, the measured water column sound speeds and bathymetry of the area where the VLAs were located, as described in Wilson *et al.*,<sup>39</sup> are used in generating the synthetic spectrograms. The ocean is modeled as 75 m of water. Three different SSPs are used, each increasing approximately linearly. The sound speeds are modeled after SSPs measured during SBCEX 2017, as shown in Fig. 7. Water density is held constant at 1.04 g/cm<sup>3</sup>. Previous studies used ten different measured SSPs, but synthetic data generated with such similar SSPs were found to cause significant data

leakage, as discussed in Sec. II E. We selected only three of the previous ten measured SSPs: the mean, minimum, and maximum, which span the space of possible SSPs.

The purpose of this work is to evaluate the potential for using multichannel SOO spectrograms in seabed classification. To accomplish this, a catalog of suitable seabed classes has been created that capture the variety of seabeds in shallow ocean regions and yet are acoustically distinct. Initially, in Van Komen *et al.*,<sup>35</sup> four seabed classes were used with parameterizations  $\theta$  representing deep mud, mud over sand, sandy silt, and sandy seabeds, taken from previous geoaoustic inversions at different locations. However, more environments were needed to obtain better resolution and improve the applicability of a deep learning seabed classifier.

To expand the catalog of seabed classes, a measure of acoustic similarity was identified to help select seabed classes that are acoustically distinct enough to be learned by the classifier. As explained in Forman *et al.*,<sup>3</sup> the determination of “acoustically distinct” seabeds was made by calculating the commonly-used Pearson correlation for vectors of TL as a function of frequency and range. For each of approximately 60 seabeds from the literature, TL vectors at one-third octave band center frequencies 250–1250 Hz and ranges 1–15 km were generated, and the Pearson correlation of every combination was calculated. One-third octave bands were chosen to represent the overall effect of the frequency-dependence of the TL as a function of range.



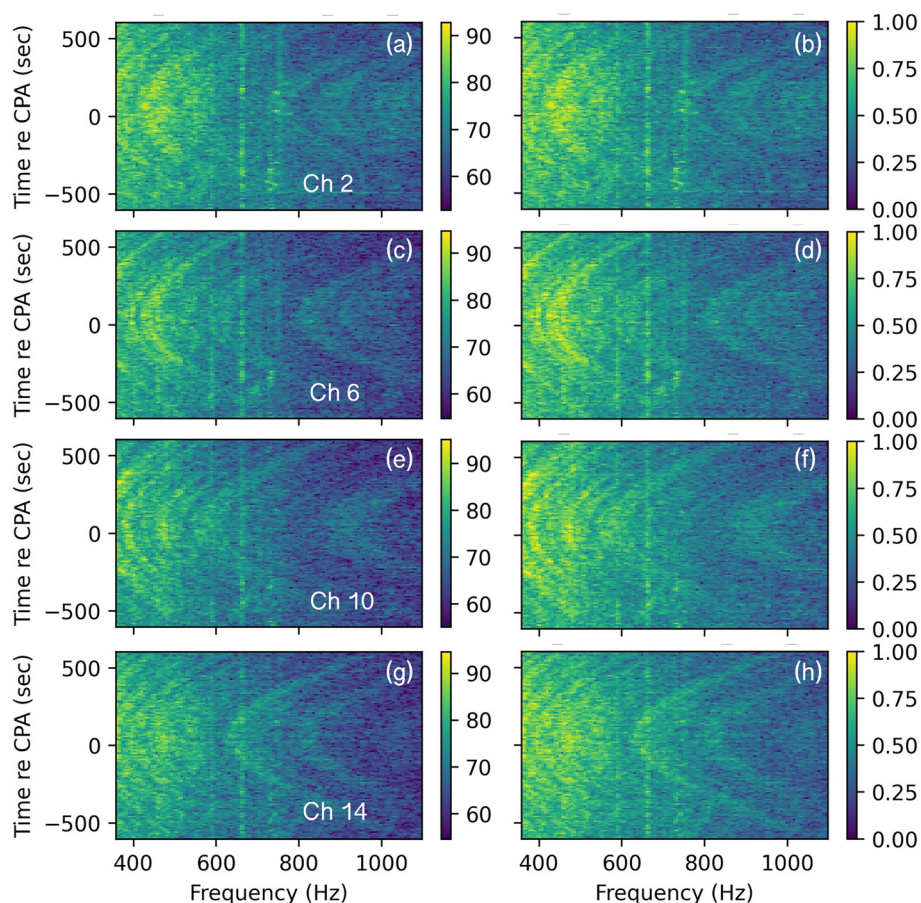


FIG. 5. SOO spectrograms of *Hafnia Green* during SBCEX 2017 from VLA 2, similar to Fig. 4. The AIS data for *Hafnia Green* gives an approximate speed of 10.9kt and a CPA range of 2.27 km from VLA 2.

Different frequency spacings were tested and yielded similar results.

In an iterative process, seabeds were included in the final catalog if their TL vectors had a Pearson correlation of less than 0.8 with all other seabeds already in the catalog. The threshold of 0.8 was selected because correlations greater than 0.8 indicate a strong positive correlation. The resulting catalog of 34 seabeds was used in Escobar-Amado *et al.*<sup>34</sup> These seabeds are listed in Table II along with the references for where the geoacoustic profiles were obtained. The seabeds are ordered according to the sound speed at the top of the sediment. The sediments with “sbc” at the end of the label and a “\*” next to the number came from geoacoustic inversions using data measured during SBCEX 2017. Details about how these seabeds were parameterized may be found in Howarth *et al.*<sup>4</sup> Schematics of Seabeds 0–21, 28, and 33 are shown in Figs. 8 and 9.

The seabed catalog was used to create a synthetic dataset of simulated spectral density levels (in dB re  $1 \mu\text{Pa}^2/\text{Hz}$ ) for training and validation testing. For each environment, sets of CPA ranges, ship speeds, and effective source depth were randomly chosen. These quantities were chosen to provide a reasonable variety of ship properties for each combination of seabed and SSPs. The bounds for ship CPA, ship speed, and effective source depth encompass a common range of values for transiting ships. Specifically, the bounds on CPA are between 0.5 and 15.5 km, ship speeds are

between 8 and 20 knots, and effective source depth is between 6 and 9 m.

To ensure that the random selections of ship speed and CPA range are not clumped, the span for each parameter is divided into bins, and three random samples are drawn from each bin. Specifically, three ship speeds are randomly chosen between 8 and 12 knots, 12 and 16 knots, and 16 and 20 knots. Three CPA ranges are randomly chosen between 0.5 and 3.5 km, 3.5 and 6.5 km, 6.5 and 9.5 km, 9.5 and 12.5 km, and 12.5 and 15.5 km. In previous work, the center of each bin was included for every environment, but these are not included in the current study. Thus, for each environment (combination of seabed and SSP), 15 CPA ranges, nine ship speeds, and one source depth were randomly chosen. For each data sample, the receiver depths are selected from those shown in Fig. 1. The resulting synthetic dataset has a total of 41 310 spectrograms with unique source properties. The training dataset is intentionally small to avoid data leakage during training and validation, which improves the generalizability of the trained networks.<sup>44</sup> More input data will be required when greater variations in SSPs are included in the training data, as described in Van Komen *et al.*<sup>35</sup>

#### D. Deep learning

A ResNet-18 neural network architecture was chosen for seabed classification following the results of Escobar–Amado

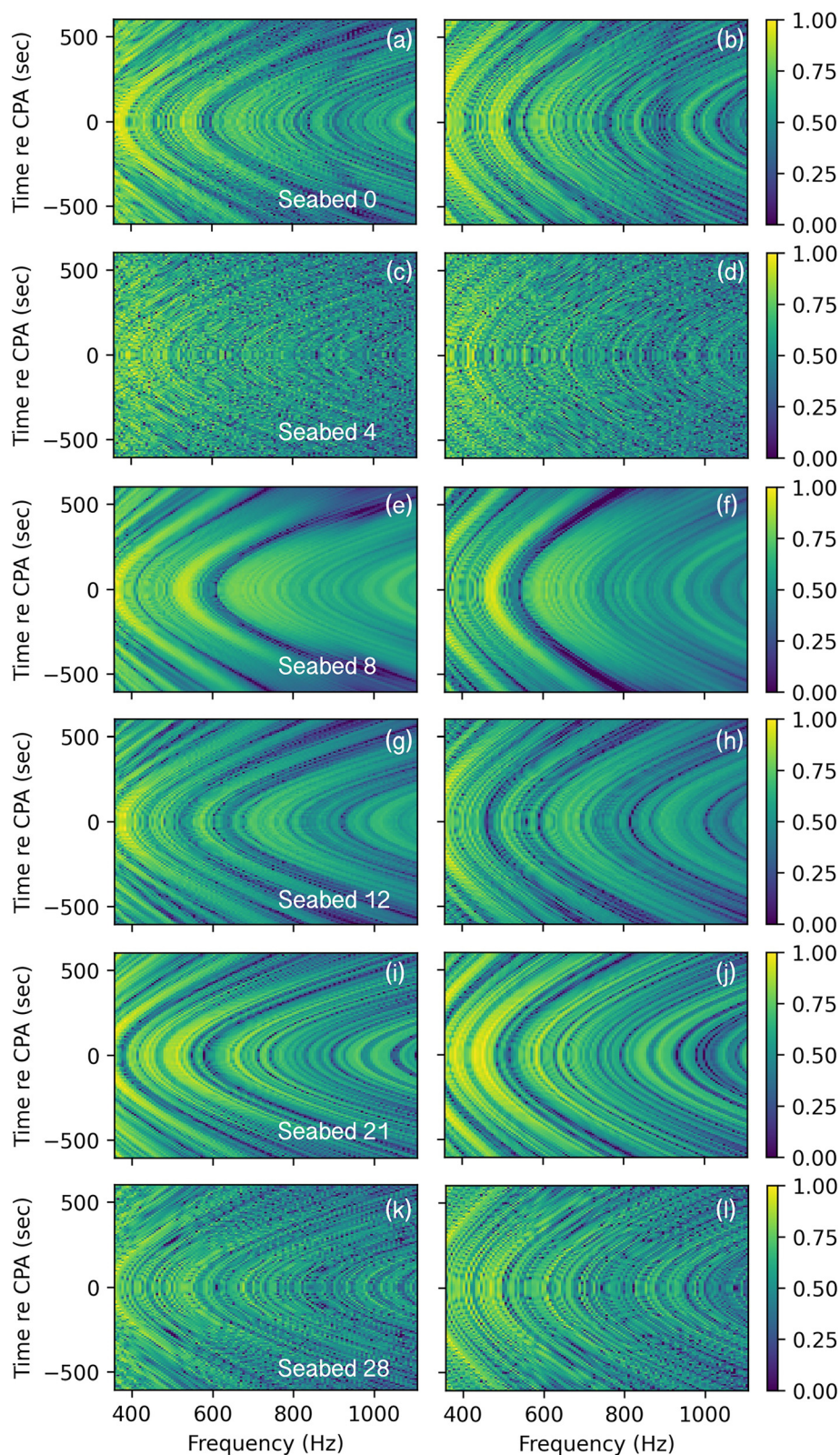


FIG. 6. Synthetic SOO spectrograms for Seabeds 0 (a) and (b); 4 (c) and (d); 8(e) and 9(f); 12 (g) and (h); 21 (i) and (j); and 28 (k) and (l) using the ship speed and CPA range of *MSC Kalamata* (left) and *Hafnia Green* (right) from VLA 2 Channel #10. These spectrograms have been mapped as discussed in Sec. II E 1.

31 March 2025 19:12:44

*et al.*<sup>34</sup> which compared six different CNN architectures. The networks in that experiment were two three-layer CNNs, three five-layer CNNs, and one 18-layer residual CNN called ResNet-18. The ResNet-18 networks had the most consistent performance in cross-validation testing and generalization on 69 measured SOO spectrograms.

Residual neural networks were first proposed by He *et al.*<sup>66</sup> and introduced identity mapping skip connections between layers. The skip connections combat vanishing gradients during the backpropagation step of training a CNN, which worsens as more layers are added to a network, and thus skip connections make deeper networks more feasible.



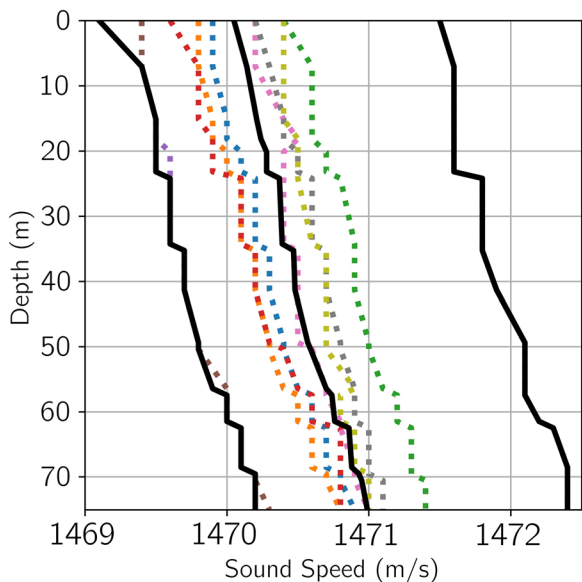


FIG. 7. Ten water SSPs measured during SBCEX 2017. The three SSPs shown as solid black lines are the profiles used in generating synthetic SOO spectrograms for the present study. The dotted, colored lines are SSPs not used in this study.

ResNet-18 is a CNN architecture containing 18 layers and skip connections, with  $11.2 \times 10^6$  learnable parameters. A final linear layer connects to the classifier output via a soft-max activation.<sup>67</sup>

While previous work used only one-channel input data,<sup>34,35,41</sup> deep learning networks can utilize input data with multiple channels. Signals recorded on multiple hydrophone depths can be stacked together into one input data tensor. When representing the input spectrogram data as a tensor, the depth of the tensor is the number of channels in the data samples (one, two, four, or eight channels in this study). Each two-dimensional layer of the tensor is the

spectrogram data from one hydrophone. The additional information from this multichannel approach is hypothesized to increase the performance of the deep learning networks for seabed classification. For the majority of this study, the even-numbered channels on the 16-channel VLA are used to create different testing scenarios. Using the even-numbered channels captures the range of information across the array, while keeping the size of each data sample low, thus, decreasing the overall computational load.

A one-channel network is trained for each of the even-numbered channels. For two-channel networks, the 16 different combinations listed in the left column of Table III are trained. For four-channel networks, the 13 combinations from the right column of Table III are trained. For eight-channel networks, one network is trained using all available even-numbered channels, another network is trained using the odd-numbered channels, and a third network is trained using only the deeper eight channels (Channels #9–#16), for a total of three 8-channel networks. The specific channel combinations were chosen following the results of a numerical study by Lau and Neilsen,<sup>68</sup> which concluded that the spacing between hydrophone channels used in a multichannel network does not significantly impact validation accuracy for the case of slightly downward refracting SSP. However, a slightly higher validation accuracy was obtained for networks with a deeper average depth of the hydrophones included.

E. Hyperparameters

The ResNet-18 network is trained on the synthetic dataset discussed in Sec. II B. The input dataset is randomly divided for *k*-fold cross-validation<sup>69</sup> with *k* = 5. This method splits the training dataset into five equally-sized groups. Training occurs in four of the five groups, and

TABLE II. List of the 34 seabeds used for this study. Column “#” corresponds to the seabed number used for identification in the results.  $c_1^{top}$  is the value of sound speed at the top of the first layer reported in the paper specified in the “Ref” column. The seabeds with “\*” next to the number come from geoaoustic inversions using data from SBCEX 2017, as indicated in the references.

#	Seabed label	$c_1^{top}$	Reference	#	Seabed label	$c_1^{top}$	Reference
0*	lin_sbc	1387	45	17*	dahl2020_sbc	1479	21
1*	knobles_sbc	1436	46	18*	brown_sbc	1479	47
2*	dahl_sbc	1445	20	19*	michalopoulou_sbc	1491	48
3*	belcourt_2_sbc	1445	49	20	clay_35m	1500	50
4*	potty_sbc	1452	51	21*	tollefsen_sbc	1508	16
5*	wan_2_sbc	1452	52	22	malta_a	1510	53
6*	belcourt_1_sbc	1453	49	23	west_florida	1540	54
7	deep_mud	1454	55	24	korean_c	1553	56
8*	barclay_2_sbc	1455	57	25	korean_a	1558	55
9*	barclay_3_sbc	1459	57	26	korean_k	1567	58
10*	bonnel_1_sbc	1464	59	27	silt_35m	1575	50
11*	wan_1_sbc	1467	60	28	sand	1592	61
12*	mud_over_sand_sbc	1469	46	29	sandy_silt	1592	51
13*	barclay_1_sbc	1470	57	30	knobles14j	1650	62
14*	mud_35m	1470	50	31	newJersey_a	1704	63
15*	bonnel_2_sbc	1471	64	32	korean_h	1757	65
16*	bonnel_3_sbc	1474	64	33	gravel_35m	1800	50

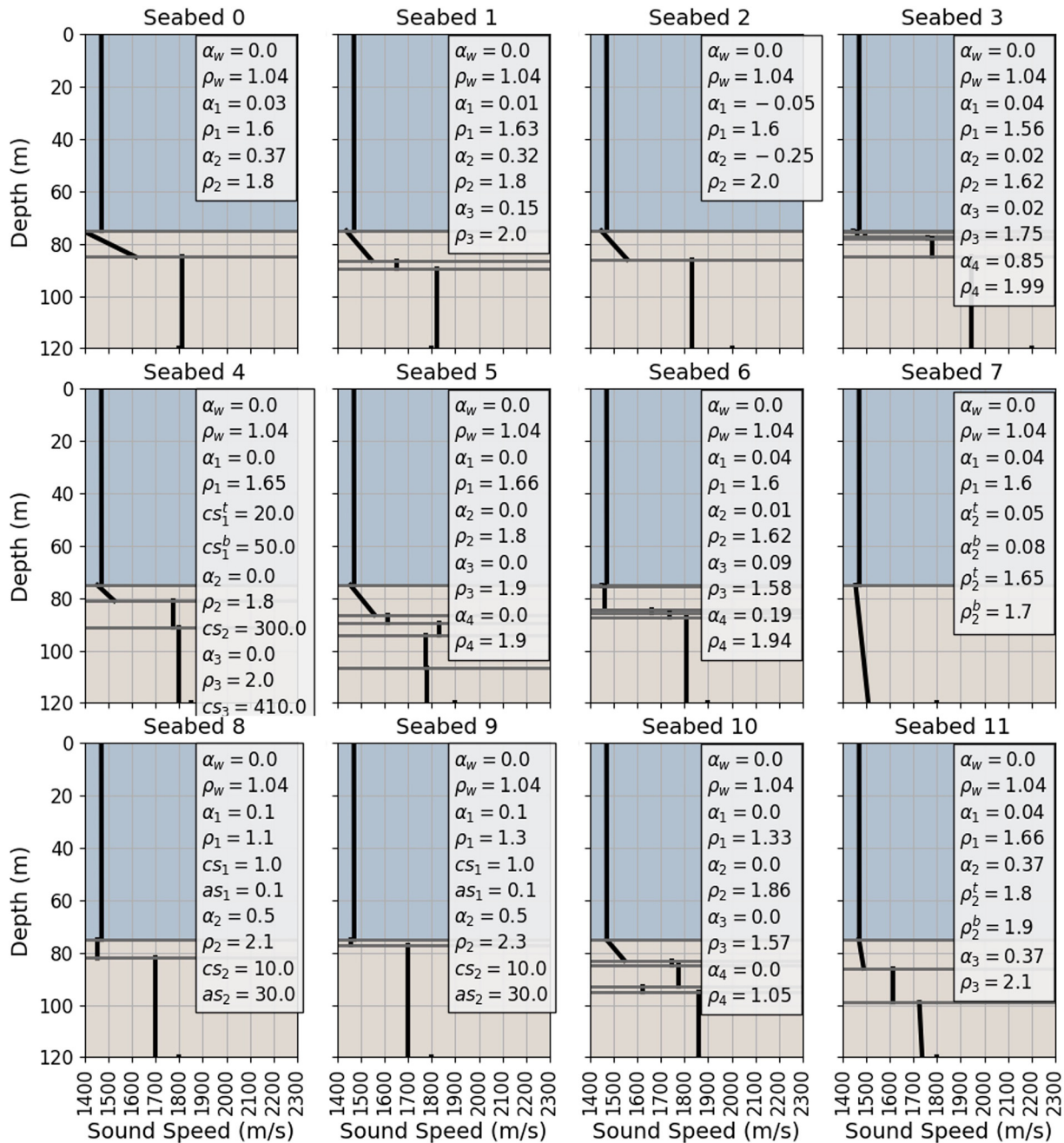


FIG. 8. Schematics of Seabeds 0 to 11 from Table II. The depth-dependent sound speed  $c(z)$  is shown as the thick black line through 75 m of water (blue background) and the upper 45 m of seafloor (brown background). The density  $\rho$  ( $\text{g/cm}^3$ ) and compressional attenuation  $\alpha_p$  (dB/m/kHz) are listed in the side bar. For seabeds that contain shear properties, the shear speed  $c_s$  (m/s) and shear attenuation  $a_s$  (dB/m/kHz) are also listed. The subscript  $w$  indicates the water; superscripts  $t$  and  $b$  differentiate between the top and bottom of the sediment layer; the subscript number corresponds to the sediment layers.

validation testing occurs in the remaining group. The purpose of  $k$ -fold cross-validation is to reduce bias in trained networks since five networks are effectively trained with the same amount of input data. Each trained network yields a classifier output vector corresponding to the probability that the different classes correspond to an input data sample.

Data leakage<sup>44</sup> occurs when statistical independence between the validation and training data is violated. Networks trained on leaky datasets “cheat” by learning the validation labels instead of generalizable attributes of the training data. Data leakage was hypothesized to have

occurred in previous studies, including Escobar-Amado *et al.*,<sup>34</sup> where the training data were generated with ten measured SSPs capable of producing nearly-indistinguishable samples. Additionally, previous networks were trained on synthetic data in which ship speeds and CPA ranges contained both specified values (bin centers, as previously described) and randomly drawn values. Consequently, the training dataset contained multiple spectrograms with the exact same ship speed and CPA range, similar SSPs, and similar source depths for each environment. The similarity in the resulting spectrograms led to

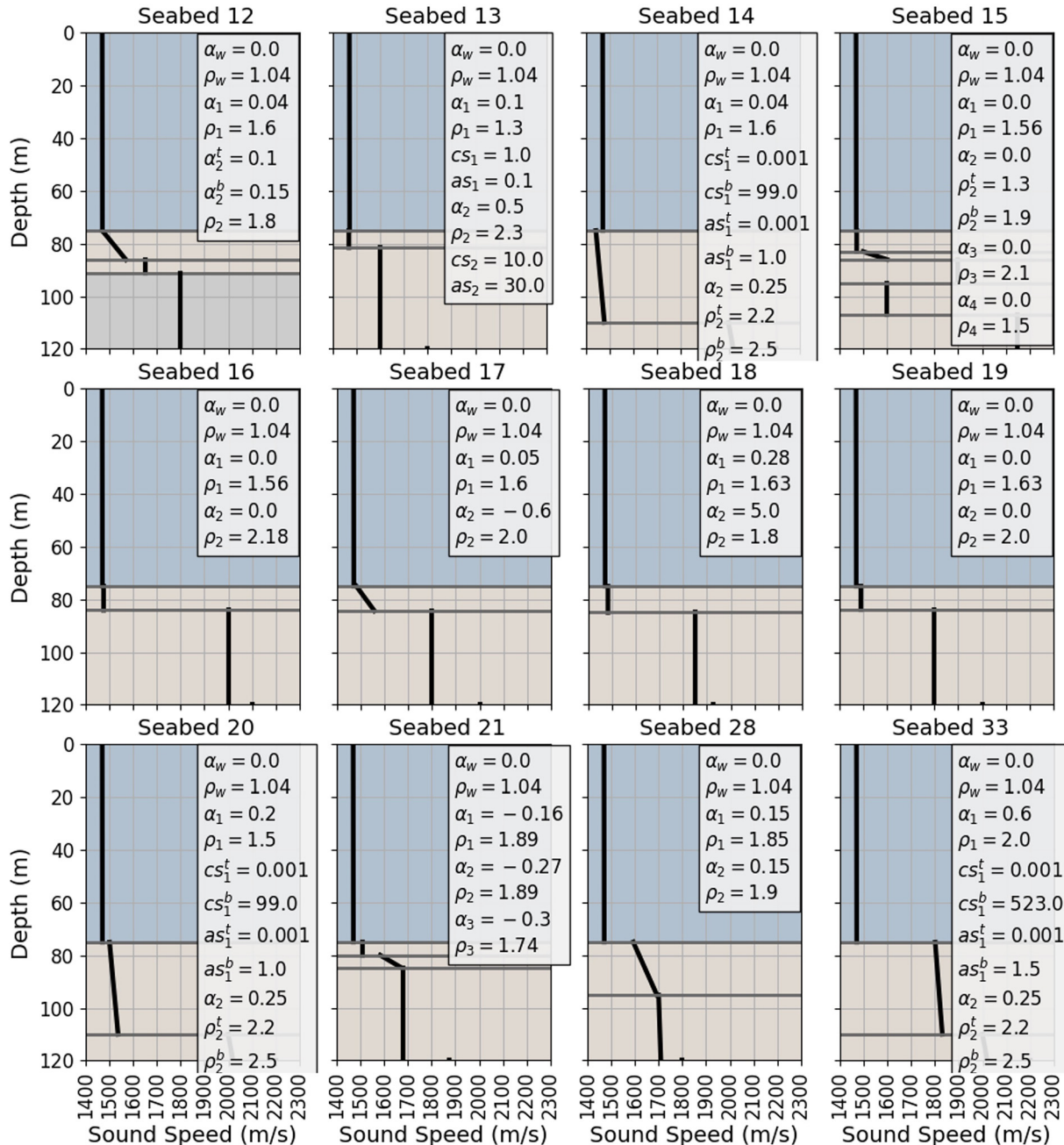


FIG. 9. Schematics of Seabeds 12 to 21, 28, and 33 from Table II, similar to Fig. 8. Negative attenuation values indicate units of dB/λ. Files listing all the parameter values are available from the corresponding author, upon request.

data in both the training and validation sets containing near-identical samples when the  $k$ -fold split was made. For the present study, data leakage is mitigated by using only three measured SSPs, one randomly drawn source depth, and CPA ranges and ship speeds randomly chosen for each data sample. In future work, significant variation in SSPs should also be included when generating the training dataset, as discussed in Van Komen *et al.*<sup>70</sup> Additionally, the source depth variation may have an impact, as described in Knobles *et al.*<sup>71</sup> and should be more fully investigated.

Other hyperparameters used during training are as follows. The network is optimized with the AdamW optimizer,

an optimizer based on Adam but with weight decay corrections.<sup>72</sup> The learning rate begins at 0.001 and then changes each epoch according to a cosine annealing scheduler. An early stopping mechanism is included that stops training when the validation loss does not improve by at least 0.001 for three epochs, otherwise referred to as an early stopping patience of three. However, for eight-channel networks, a patience of five epochs, rather than three epochs, was used.

In deep learning, at each epoch, the training data is randomly divided into batches. Preliminary testing of various batch sizes (number of samples in one batch) indicated that a larger batch size improved training with ResNet-18. A



TABLE III. Channel combinations for the two-channel and four-channel networks used in this study. Channel numbers correspond to hydrophone depths from Fig. 1.

Two channels	Four channels
4-8	2-4-6-8
4-10	4-6-8-10
4-12	2-6-10-14
6-8	4-6-10-14
6-10	6-8-10-12
6-12	2-8-12-16
8-10	6-8-12-14
8-12	6-8-12-16
8-14	6-10-12-16
8-16	6-10-14-16
10-12	4-12-14-16
10-14	8-10-14-16
10-16	8-12-14-16
12-14	
12-16	
14-16	

batch size of 512 samples is used for one-channel and two-channel networks. However, four-channel and eight-channel networks appear to require a batch size of at least 1024 samples in order to train properly. Empirical tests to determine optimal batch sizes and patience indicated that as the number of channels increases, increasing batch size has a far greater effect on validation accuracy than changes to the early stopping patience.

During the fivefold cross-validation, one or more of the folds might terminate training prematurely via the early stopping mechanisms and result in an improperly trained network. In this circumstance, if the validation accuracy for any of the folds is below 70%, a re-training mechanism is implemented where the weights in the network receive a different random initialization, and the fold is trained again. Each fold has a maximum of two tries to reach the 70% validation accuracy threshold, and the second trained network is retained if both have low accuracy.

The neural network architecture and training algorithms are implemented in PyTorch version 1.5.1 and run in Python 3.6.9 on an NVIDIA Tesla T4 GPU. Under these conditions, the entire process from loading the training dataset to completing training takes approximately 15 min for one-channel networks, 18 min for two-channel networks, 40 min for four-channel networks, and 5.5 h for eight-channel networks. During this training process, validation is conducted using the holdout portion of the training dataset as determined by the  $k$ -fold split. After training, the networks are then applied to the data samples from SBCEX 2017 to test generalizability.

### 1. Mapped spectrogram input

In order to enhance learning, data samples are often scaled or mapped before training. Due to the use of SOO spectrograms, this study maps each data sample individually

to remove the impact of different source levels. In previous work,<sup>34,35</sup> the levels of the one-channel data samples were converted to spectral density values in  $\mu\text{Pa}^2/\text{Hz}$ , scaled by the standard deviation of values across the data sample, and then converted back to relative levels. In this current work, a different technique is used to map the spectral density levels to a domain of [0,1]. For each data sample (one ship recorded on multiple channels), the spectral density levels (in dB re 1  $\mu\text{Pa}^2/\text{Hz}$ ) are adjusted to have zero mean and a standard deviation of one across all channels. This data mapping has several purposes. From a deep learning point of view, the network trains more efficiently when the input features are in the same range as the initialized weights: [0,1]. From an acoustical perspective, this mapping removes information about the source level of the ship and the overall impact of the range and seabed attenuation. While this mapping obfuscates a lot of information typically used in ocean acoustics, this approach allows the deep learning network to focus on learning the interference pattern in the striations instead of the overall level.

The second component of the mapping procedure performs imputation on outlier values more than three standard deviations from the mean. These values are reassigned to the mean  $\pm$  exactly three standard deviations. As a result, outliers are filtered out, reducing the depth of the nulls in synthetic spectrograms and the impact of loud intermittent noise on measured spectrograms. The values are then linearly mapped to the domain [0,1]. Visual comparisons of spectrograms before and after the mapping procedure are shown in Figs. 3(b) and 3(c) for measured data and Figs. 10(a) and 10(b) for synthetic data.

This imputation has two benefits. First, the nulls in the synthetic spectrograms are not as deep after the mapping procedure. Thus, the imputation effectively introduces a noise floor, which is more realistic to the measured data samples as they contain ambient noise. An example spectrogram of this ambient noise is provided in Fig. 11(a). Second, imputation on values greater than  $3\sigma$  from the mean reduces the potential impact of loud interference in the measured SOO spectrograms.

### 2. Additive noise

To ensure that networks trained on synthetic data are learning important patterns rather than memorizing the specific simulated spectrograms, the data are augmented with additive noise.<sup>73,74</sup> This common deep learning procedure helps avoid overfitting and improves generalizability. Specifically, a zero-mean Gaussian white noise<sup>75</sup> is utilized. For each frequency at each time step of a spectrogram, a value from the Gaussian distribution is randomly drawn, and its magnitude is added to the point on the spectrogram. A visual comparison of a spectrogram before and after the additive noise is shown in Figs. 10(b) and 10(c).

In previous work, simulated wind noise was used, but in this present study (with the new mapping), Gaussian noise is used instead. The physical justification for using a Gaussian



distribution for additive noise is twofold: (1) The spectrogram in Fig. 11(a) shows that the noise exhibits minimal frequency dependence in the 360–1100 Hz band. (2) Statistical analyses of ambient noise measured during the experiment, shown in Fig. 11(b), reveal a low skewness value for noise samples.

An appropriate standard deviation for the distribution,  $\sigma$ , was determined through a hyperparameter search experiment. In this hyperparameter search, first the ideal early stopping patience length and training batch size were determined through training networks for combinations of

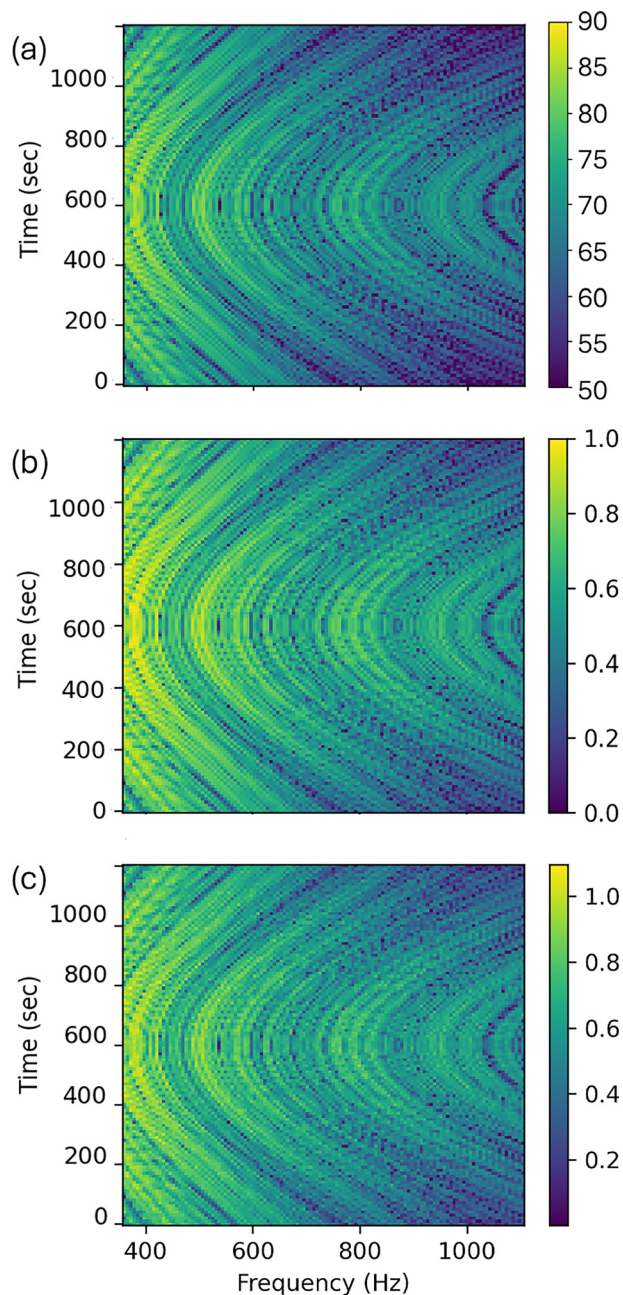


FIG. 10. Data preprocessing steps applied to synthetic spectrograms. (a) The synthetic spectrogram generated at a resolution of 121 time steps by 123 frequencies. (b) Spectrogram mapped to a domain of [0,1]. (c) Spectrogram with additive Gaussian noise generated from a zero-mean distribution with  $\sigma = 0.04$ .

candidate values spanning a wide range. The evaluation criterion was validation accuracy. Then, the noise  $\sigma$  was determined by training networks with differing  $\sigma$  and evaluating validation accuracy. The largest  $\sigma$  value with the highest accuracy was  $\sigma = 0.04$ . In future work, additional forms of data augmentation could be applied, as described in Castro-Correa *et al.*<sup>76</sup>

The implementation of noise differs between the training dataset and the validation set. For the training set, new noise is randomly drawn and added to each spectrogram of each batch, before each epoch of the training process. The addition of different noise to the training data each epoch allows the network to learn to ignore this type of noise. For the current study, the amplitude of the Gaussian was constant across all noise signals but could be chosen randomly in future work.

For the validation set of spectrograms, noise is drawn once and added during the first training epoch. The new noisy validation spectrograms replace the original validation spectrograms for use in each subsequent epoch. This approach allows more consistent monitoring of validation accuracy during training. The benefits of this approach are illustrated when the network predicts on measured data. Although the *Hafnia Green* (Fig. 5) contains intermittent

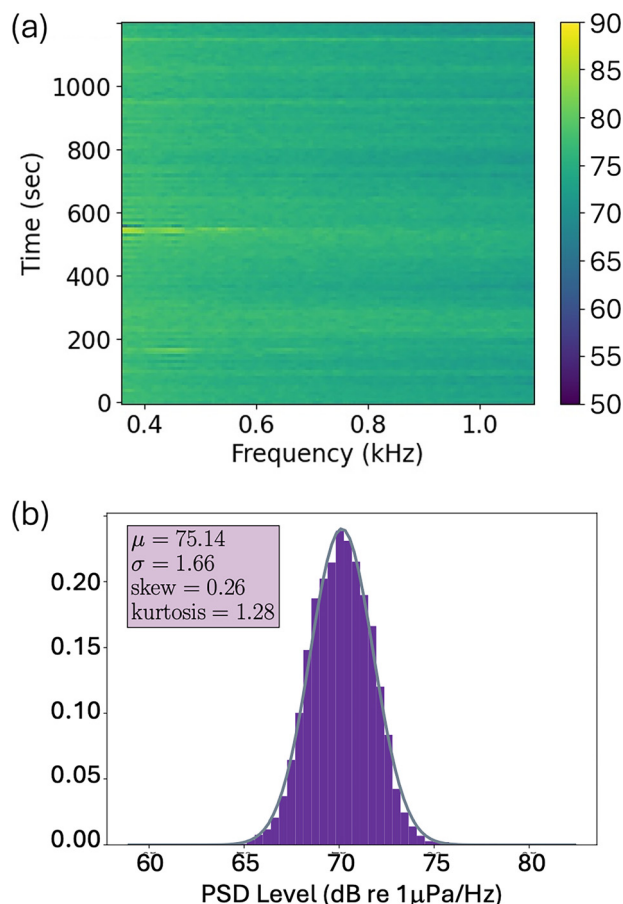


FIG. 11. Example of ambient noise measured on VLA 1 during SBCEX 2017. (a) Spectral density levels in dB re  $1 \mu\text{Pa}^2/\text{Hz}$  processed in the same manner as the SOO spectrograms. (b) Statistical analysis of this noise sample.

vertical banding, the network is still able to make reasonable predictions, as discussed in Sec. IV C.

## F. Simple ensemble approach

One challenge of real-world applications of machine learning is obtaining a measure of confidence in the predicted results. What constitutes a correct versus an incorrect prediction? In ocean acoustics applications, the correct answer is usually unknown, and therefore a measure of confidence in the predictions is necessary. To obtain a measure of confidence, we currently use a simple ensemble approach.<sup>77</sup>

In this simple ensemble approach, multiple ResNet-18 networks are trained, each with a different training dataset from a unique combination of hydrophones (depths shown in Fig. 1). For each training dataset, fivefold cross-validation is performed, which produces five trained networks. Then, all the trained networks (sometimes referred to as base learners<sup>77</sup>) are applied to measured SOO spectrograms. The results are analyzed in two ways.

First, for each measured SOO spectrogram, majority voting is performed across the output of the base learners. In majority voting, also referred to as hard voting, the seabed with the highest classifier output for each base learner is awarded one vote, and the seabed with the majority of the votes is deemed the selected seabed for that data sample. Second, soft voting is also used. In soft voting, the full classifier output vector from all base learners trained on datasets with the same number of channels is averaged (the lists of two- and four-channel combinations are shown in Table III). Soft voting can reveal if two or more seabed classes have similar predicted probabilities, something which would not be represented in the majority (hard) voting scheme. The resulting average classifier outputs for one-, two-, four-, and eight-channel networks are evaluated in Sec. III A.

These examples of simple ensemble learning yield a more informed prediction than the application of a single trained network. Furthermore, ensemble learning yields a measure of confidence. More complicated methods to apply ensemble learning are described in Polikar<sup>77</sup> and should be investigated in future work.

## III. RESULTS

Two methods are used to evaluate the performance of trained networks: validation and generalization. Validation tests are reported in Lau and Neilsen<sup>68</sup> to illustrate how accurately the network predicts synthetic data not used in training, but drawn from the same statistical distribution as the training data. This current study performs generalization testing to determine whether the network can make adequate predictions on real, measured data. Five-fold cross-validation was used to train networks, as described in Sec. II E. Therefore, in the following testing, the mean accuracy and standard deviation over the resulting five trained folds are calculated for each channel combination. For generalization testing, only networks with above 70% validation accuracy

were included to make predictions on measured data samples. This led to 95% of one-channel networks (38/40 networks), 90% of two-channel networks (72/80 networks), 100% of four-channel networks (65/65 networks), and 100% of eight-channel networks (15/15 networks) used.

## A. Generalization

To test the generalizability of the networks, data samples from the SBCEX 2017 are used. The data come from hydrophone recordings of passing surface ships at one of two VLA locations. Seven SOO spectrograms with clear striations are used to examine how different trained networks compare. These seven ships are listed in Table I. The seabeds predicted by networks trained on each of the different channel combinations are shown in Fig. 12. The seabed classifier output for each testing data sample (i.e., ship) is a 34 element vector containing values that indicates the probability of a specific seabed being the correct answer. A classifier output is obtained for each trained network. For each of the five networks obtained through  $k$ -fold cross-validation, the seabed with the highest classifier output is selected, and then majority voting yields the predicted seabed. Each majority-voted seabed prediction for a given channel combination is displayed in Fig. 12 with a unique marker, as indicated in the legend.

However, the majority voting does not mean that the other seabeds have zero probability of being the correct answer. To display the probabilities obtained for all seabeds, the classifier output was averaged over all networks trained with the same number of channels. Comparisons of the seabed probabilities are shown in Fig. 13 for the trained one-, two-, four-, and eight-channel networks. For each of the seven ships, shown along the  $y$ -axis, the averaged seabed probabilities are plotted across the  $x$ -axis with the color corresponding to the probability of the seabed being the correct seabed. The colorbar represents the averaged probabilities over all trained channel combinations for each category of one, two, four, or eight channels included, normalized so the total seabed probabilities for each ship sum to one.

To illustrate how this ensemble approach to seabed classification might work in practice, the average classifier output for all 32 SOO spectrograms from VLA 1 and VLA 2 recorded during SBCEX 2017 are provided in the Appendix.

## IV. DISCUSSION

### A. Multichannel network performance

One major goal of this study is to analyze the impact of using multiple receivers in deep learning seabed classification. Results for the validation analysis are in Lau and Neilsen<sup>68</sup> for the case of slightly downward refracting SSP. The efficacy of using multiple channels for seabed classification is evaluated in this current study via the generalization results of measured SOO spectrograms (Figs. 12 and 13). Evaluating generalization results is challenging because the correct seabed is unknown. In such cases, accuracy cannot be determined, but one can estimate the precision of the

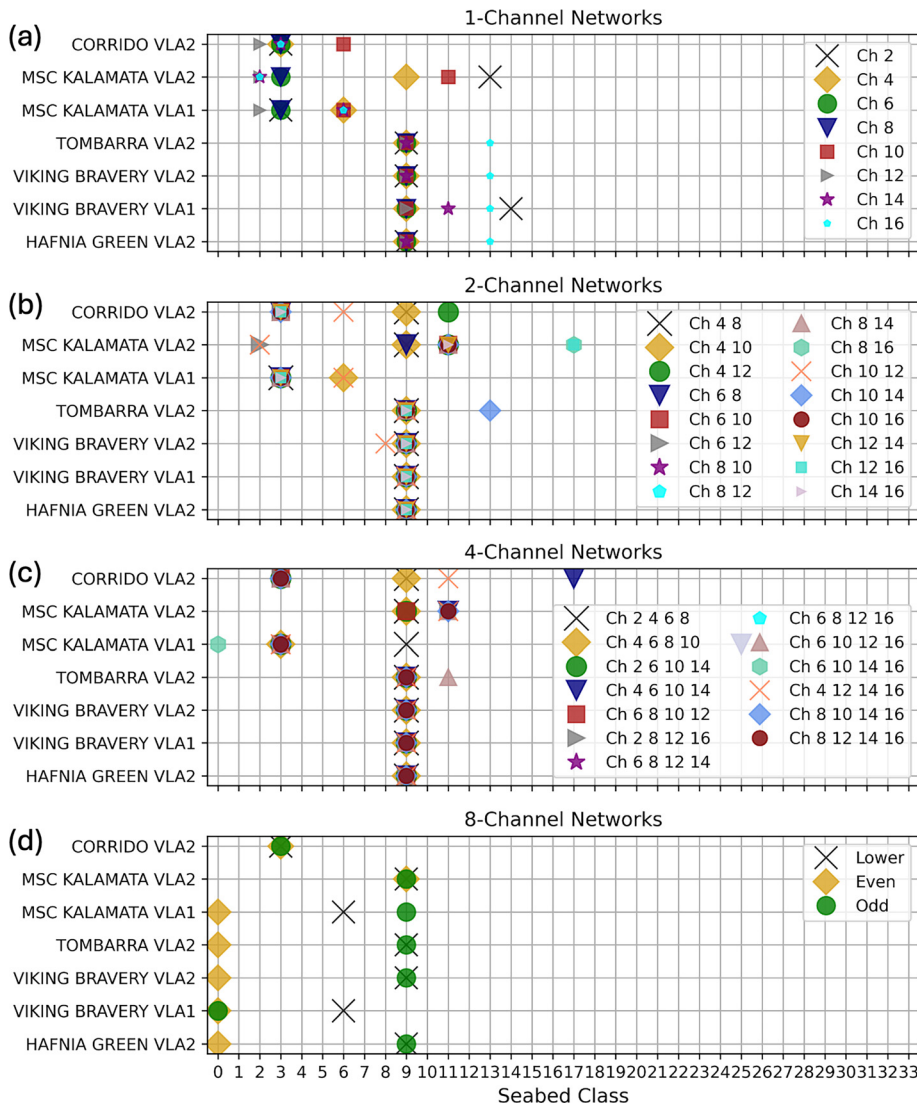


FIG. 12. Generalization results of the predicted seabeds for seven SOO spectrograms for the (a) one-channel, (b) two-channel, (c) four-channel, and (d) eight-channel combinations. The symbols indicate the predicted seabed resulting from majority voting. The names on the left correspond to the ship names and VLA listed in Table I.

results. A high level of precision or consistency in the predictions from an ensemble of trained networks can yield a measure of confidence in the predictions.

The predicted seabed distributions in Figs. 13 and 18 contain the average classifier output distributions over the ensembles of networks. To quantify the precision based on the number of channels, the entropy of the classifier outputs is computed for each distribution. The entropy of a distribution yields a measure of the preciseness of the distribution: a lower entropy indicates a more concentrated distribution with fewer probable outcomes. The entropy of the predicted seabed distributions for seven SOO spectrograms is shown in Fig. 14, and the average entropy over the ship predictions, grouped by the number of channels used in the networks, are indicated as horizontal lines. Entropies of the predicted seabed distributions for all ships are shown in Fig. 19.

The average entropy decreases as the channel number is increased from one to two to four channels. For the seven ship example in Fig. 14, the eight-channel average entropy is slightly higher than four channels, whereas the 32 ship example in Fig. 19 shows the opposite. Thus, the entropy

appears to converge around four channels. In practice, the potential increase in precision by including extra channels must be weighed against practical constraints on data size and resulting computational costs. For example, the seabed predictions in Figs. 12 and 13 show that the ensemble modeling using two, four, and eight channels does not yield significantly different predicted seabed distributions, so the added channels may not be worth the computational cost. For a more complicated depth-dependent SSP, however, the advantage of using more channels may become more apparent.

### B. Ensemble learning

The next major goal of this study is to evaluate the effects of ensemble learning. We aim to determine if ensemble learning can increase the precision of the network predictions, as well as provide a measure of confidence in the seabed predictions. First, the results shown in Fig. 12 are analyzed. These plots display the predicted seabed from majority voting for the networks trained on each



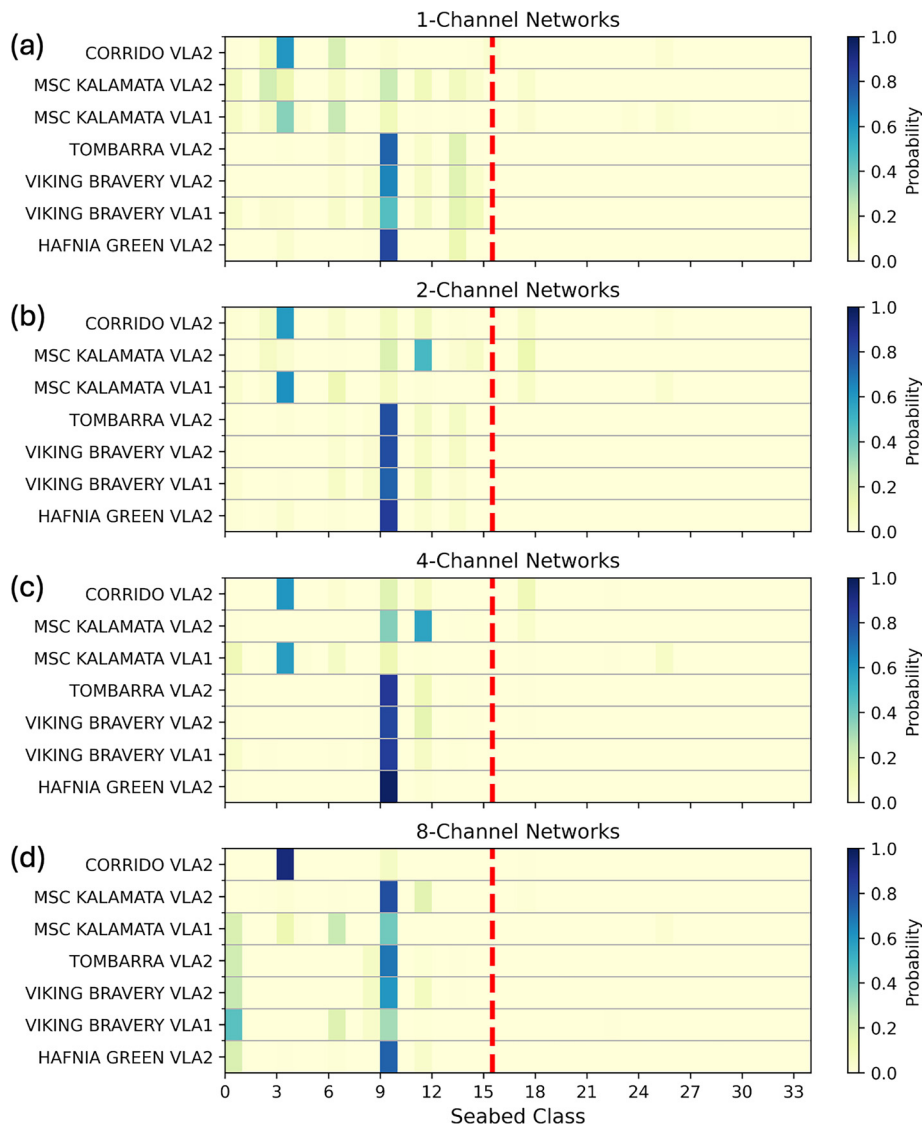


FIG. 13. Ensemble generalization results for seven SOO spectrograms: average classifier output over the (a) one-channel networks, (b) two-channel networks, (c) four-channel networks, and (d) eight-channel networks. The names on the left correspond to the measured data sample: ship name and which VLA. The  $x$ -axis ticks lie to the left of their corresponding shaded rectangular cells. The red vertical dotted line separates a sound speed ratio across the water-sediment interface of less than one to the left (i.e., where the sound speed at the top of the sediment is less than the sound speed at the bottom of the water column), and greater than one to the right of the line.

combination of channels. These results show that not all channel combinations yield the same predicted seabed, especially in the one-channel cases. Therefore, if only one network from a single channel or set of channels is used, no information is available regarding the likelihood that the selected seabed represents reality. However, analyzing seabed predictions from multiple networks trained on different channel combinations provides a measure of precision or confidence in the most frequently predicted seabed.

For example, consider the ship with the largest spread in seabed predictions, the *MSC Kalamata* on VLA 2. For the one-channel predictions in Fig. 12(a), the predicted seabeds are split between Seabeds 2, 3, 9, 11, 13, and 17. Fewer seabeds are predicted when four or eight channels are used [see Figs. 12(c) and 12(d)].

A different approach to ensemble learning is to average the classifier output from all networks trained with a specified number of channels. The averaged classifier output for the seven SOO spectrograms is displayed in Fig. 13. With this type of ensemble learning, even the one and two-channel predictions show less variation. For example, these

ensemble results from *MSC Kalamata* on VLA 2 show that Seabed 9 or 11 has the highest average classifier output for most of the multichannel cases. The results in Fig. 13 illustrate that averaging the classifier output from an ensemble of networks trained with different channel combinations yields a more precise result than majority voting or taking the prediction from a single network. The average classifier output distributions provide a measure of confidence in the answer, as shown in Fig. 14.

While effective, the ensemble learning techniques used in this work are basic, and future work should consider more advanced ensemble learning approaches.<sup>77</sup>

### C. Seabed predictions

A third goal of this study is to evaluate if the predicted seabeds are good predictions. The correct seabed is unknown, so this evaluation is based on the similarity between predicted seabeds, visual inspection of spectrograms, and a discussion of why different ships yield different seabed predictions.



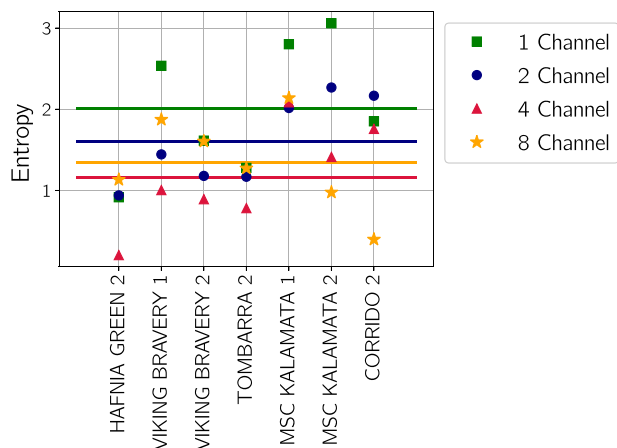


FIG. 14. Entropy of the predicted seabed distributions for seven SOO spectrograms in Fig. 13. Values were calculated over the average classifier outputs obtained for networks trained using the same number of channels. Average entropy values over the ships for each number of channels are shown as horizontal lines, with values of approximately 2.01 for the ensemble of one-channel networks, 1.60 for two-channel networks, 1.16 for four-channel networks, and 1.34 for eight-channel networks.

The seabeds with the highest average classifier output over the ensemble of networks are Seabeds 3 and 9 (Fig. 13). These seabeds came from previous geoaoustic inversions done in the same New England Mud Patch area using data collected during SBCEX 2017, as denoted by the “\*” in Table II. Seabeds 3 and 9 share two key features: (1) similar sound speed ratio across the water-sediment interface (i.e., the sound speed at the top of the sediment layer is less than the sound speed at the bottom of the water column), and (2) no sound speed gradient in the top few meters of the sediment. Deeper features of the sediments are likely less significant because 360–1100 Hz spectrograms are used in this work, which contains less information about the deeper sediment than lower frequency ranges.

To search for visual clues as to why certain seabeds were selected, the measured spectrograms are mapped according to Sec. II E 1 and then compared to spectrograms simulated with different seabed classes. Two examples are provided. Synthetic spectrograms are generated with the ship speed and CPA range of the *Hafnia Green* and the *MSC Kalamata* on hydrophone Channel #10 of VLA 2. They are displayed beside the measured spectrograms in Figs. 15 and 16. Synthetic spectrograms for an additional six seabeds are represented in Fig. 6.

An examination of the *Hafnia Green* spectrograms in Fig. 15 shows that the selected Seabed 9 provides a more reasonable visual match for the measured data than many of the other seabeds. Although the measured spectrogram contains intermittent vertical banding not present in the synthetic spectrograms, the predicted seabed still produces the characteristics of the measured spectrogram overall, indicating little effect of the banding on prediction accuracy. A different example is shown in Fig. 16 for the measured and synthetic spectrograms from Channel #10 of *MSC Kalamata* on VLA 2. A visual examination of the spectrograms for different seabeds shows that the ensemble-selected Seabeds 9

and 11 (in Fig. 13) not only provide reasonable matches to the measured spectrogram, but also look remarkably similar to each other. This similarity illustrates why both seabeds have high classifier output (Fig. 13).

This qualitative comparison between the spectrograms provides a preliminary check that the network is learning key features of the spectrogram. A quantitative comparison of the measured and synthetic spectrograms would be considerably more useful; however, the best metric for quantifying this comparison is unclear. In deep learning applications for photographs and videos, the structural similarity index metric (SSIM) is used to compare the perceptual qualities of two images.<sup>78</sup> These qualities include luminance, contrast, and structural information. We calculated the SSIM on the measured and synthetic SOO spectrograms for different seabeds, and the results did not follow logical trends based on the visual evidence. Thus, we conclude that the SSIM is not a good metric for comparing spectrograms and a useful metric needs to be developed using domain-specific knowledge, i.e., to consider specific features of spectrograms that are tied to the seabed. This question will be considered in future work.

What seabed class, then, is the best prediction from our networks for the experimental area of the SBCEX 2017? It appears that some ships have a classifier output with high probability for Seabed 3, and other ships for Seabed 9. This difference is neither related to the VLA nor the shipping lane used. The main correlation between the selection of Seabed 3 versus Seabed 9 appears to be the date of the SOO passage. From Table I (corresponding to Fig. 13), the upper three SOO were recorded on Julian days 83–84 and the lower four SOO were recorded on Julian days 89–90.

To investigate this trend, the statistics of the ambient ocean noise were analyzed. Samples of ambient noise on days 83–84 and 89–90 were identified that were free from the sound of ships and towed sources. From these ambient noise samples, recordings closest in time to the ships from Table I were selected. The statistics across 20-min spectrograms of the noise samples in the 360–1100 Hz band are provided in Fig. 17. The statistics of these noise spectrograms reveal that the ambient noise samples had a lower mean, greater standard deviation, and more variability in skewness and kurtosis on days 89–90 than noise samples from days 83–84. These differences may be tied to the seabeds selected by the trained deep learning networks.

This potential dependence on ambient noise statistics at the time of recording could explain why the prior single-channel study summarized in Fig. 7 of Escobar–Amado *et al.*<sup>34</sup> found Seabed 3 to be the most commonly predicted seabed (i.e., single seabed with the highest classifier output) over 2070 network predictions. Their study included 18 additional ships that were recorded on the University of Delaware (UD) VLA prior to Julian day 83, because the UD VLA started recording on Julian day 66 (March 7, 2017). The UD VLA was located approximately halfway between VLA 1 and VLA 2 as shown in Fig. 1 of Escobar–Amado *et al.*<sup>34</sup>

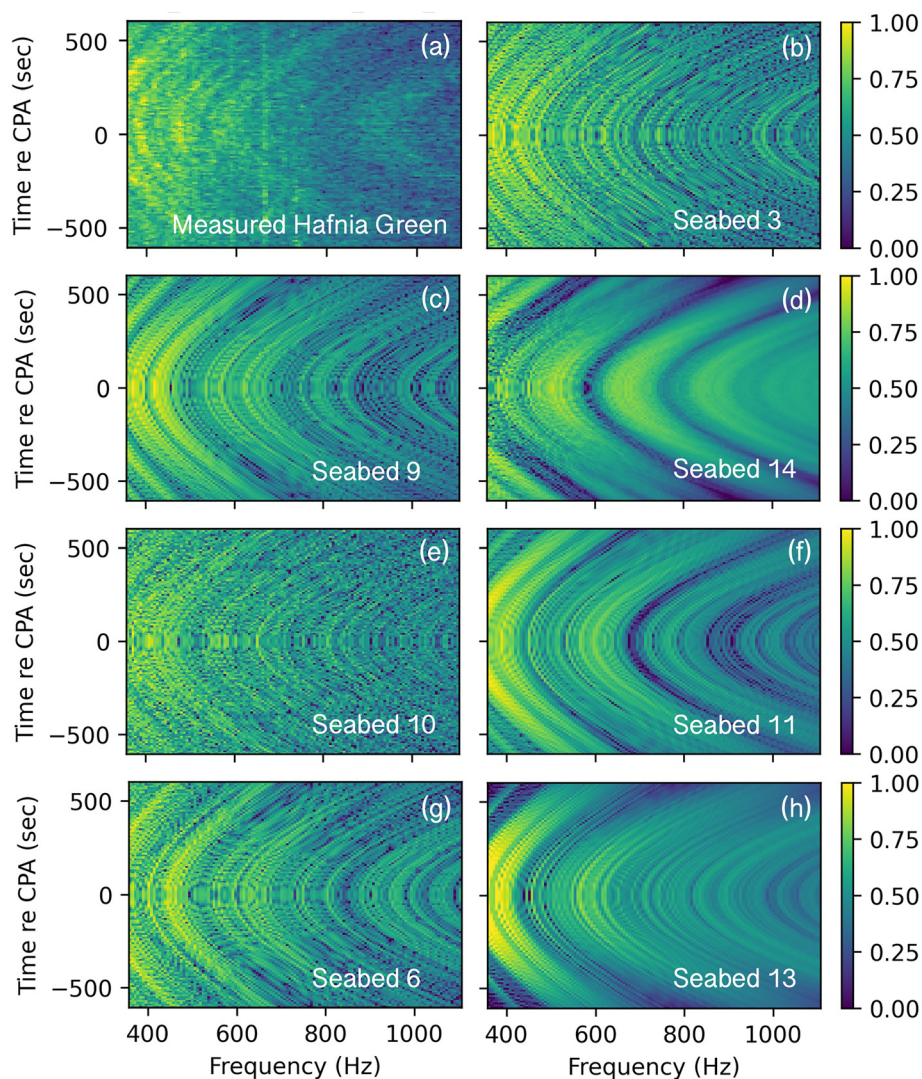


FIG. 15. (a) Mapped SOO spectrogram for *Hafnia Green* from Channel #10 on VLA 2, and (b)–(h) mapped synthetic SOO spectrograms generated with the listed seabeds using the ship speed and CPA range of *Hafnia Green*. The ensemble results of the multichannel networks predict Seabed 9. Additional synthetic spectrograms for seabeds not chosen are shown in the right column of Fig. 6.

More research is needed regarding the impact of ambient noise statistics on the seabed predictions. The encouraging fact is that, although two seabed classes were commonly predicted instead of just one, the two seabeds share similar geoacoustic profiles in the upper portion of the sediment. We anticipate that the role of noise statistics will gain further importance when more water sound speed variation is present during the data collection period, as was the case in SBCEX 2022 [see Fig. 1(b) of Vardi and Bonnel<sup>37</sup> for examples of measured SSPs].

**D. Robust deep learning**

Several key decisions were made about the deep learning process that offered improvements to previous seabed classification work.<sup>4,34,35</sup> With regard to the network architecture, a convolutional neural network works well for detecting spectral features such as striations from SOO spectrograms. The skip connections in ResNet-18 enable deeper networks that learn more features. A larger batch size helps the ResNet-18 train more effectively. The batch size needs to increase as the number of channels increases to obtain the

same validation accuracy during training. Early stopping helps guard against overfitting and helps the networks train faster. Similar seabed predictions are obtained with and without early stopping.

Our new approach to mapping has a positive impact on the performance. The spectral density levels allow the networks to be trained in fewer epochs. Previously, each data sample was scaled to have a standard deviation of one in Pascals and then converted back to decibels, as is conventional in acoustics. In this current work, the spectral density levels are mapped to lie between zero and one, as explained in Sec. II E 1, which is a common deep learning data scaling procedure. The validation performance is not impacted by this change, but this new mapping allows the networks to train faster. The generalization results, however, are impacted by the choice of mapping. The multichannel results with the previous scaling in Pascals do not appear to be learning the subtler features associated with different ships.

Another preprocessing step is the addition of Gaussian noise to help the networks learn to separate signal from noise. This additive noise is a minor form of data



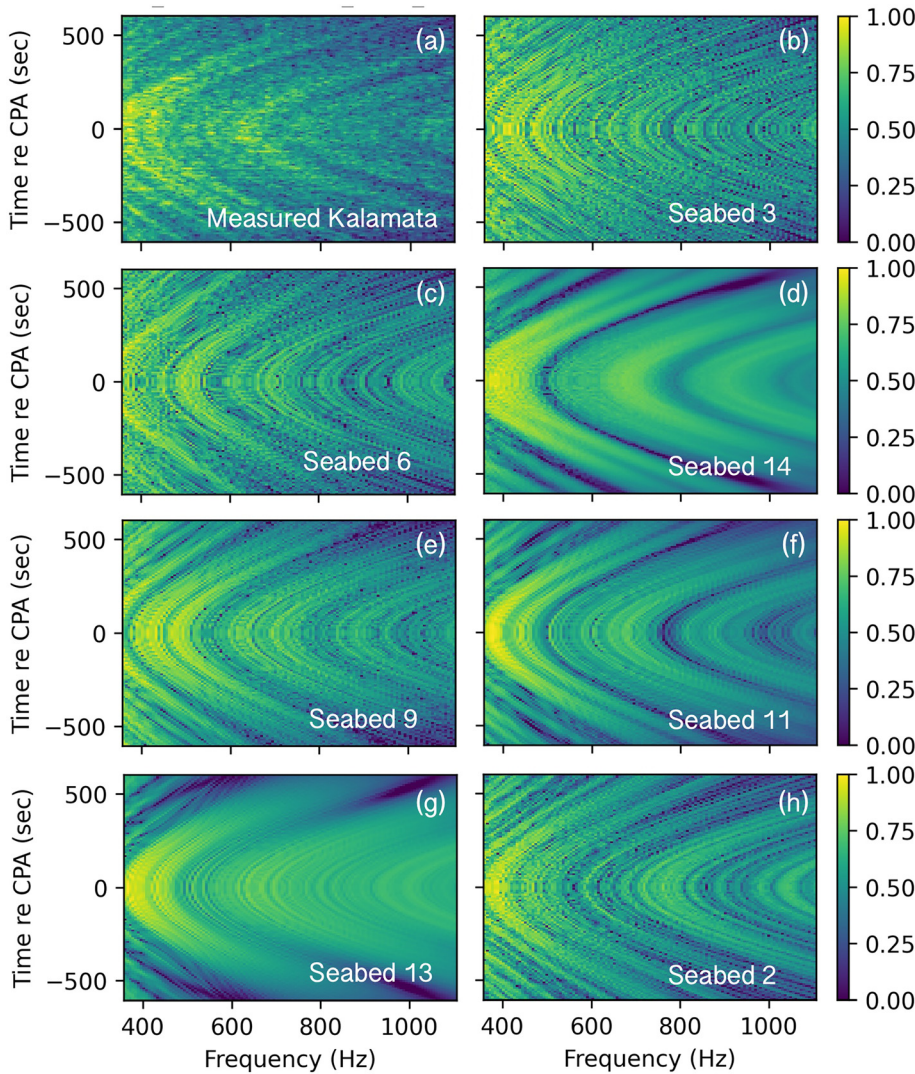


FIG. 16. (a) Mapped SOO spectrogram for *MSC Kalamata* from Channel #10 on VLA 2, and (b)–(h) mapped synthetic SOO spectrograms generated with the listed seabeds using the ship speed and CPA range of *MSC Kalamata*. The ensemble results of the multichannel networks predicted Seabeds 9 and 11. Additional synthetic spectrograms for seabeds not chosen are shown in the left column of Fig. 6.

augmentation, and a unique noise signal is randomly added to each data sample at each epoch. In subsequent tests, we observed that the average classifier output results were similar between networks trained with and without this additive noise, for the ships with clear striation (Fig. 13). The benefit is seen, however, when a wider range of ships is considered, as in Fig. 18. For ships with less clear striations, the average seabed classifier output shows less uncertainty when using networks trained with the additive noise. Future work should be done to identify cases where the Gaussian additive noise is essential. Additionally, in this work, the amplitude of the Gaussian distribution was not varied but could be randomly selected in future work.

In this study, data leakage is avoided by strategically selecting a smaller dataset with more independent samples. Specifically, nearly duplicate data samples with the same source properties are not included in the training dataset. While this dramatically decreased the total number of data samples (from 137 000 in Escobar–Amado *et al.*<sup>34</sup> to 41 000), it helped reduce data leakage, which led to trained networks that generalize better to measured data samples. Additionally, a representation of the spectrograms that is

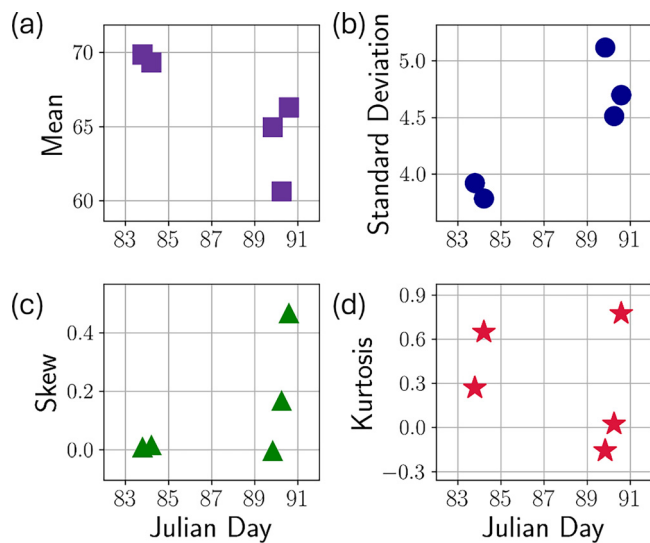


FIG. 17. Statistical analysis of five different noise samples recorded at times where no obvious ship noise or tones were present. The days and times are chosen to be close to the transits of the ships in Table I. The values calculated were (a) mean, (b) standard deviation, (c) skewness, and (d) kurtosis.

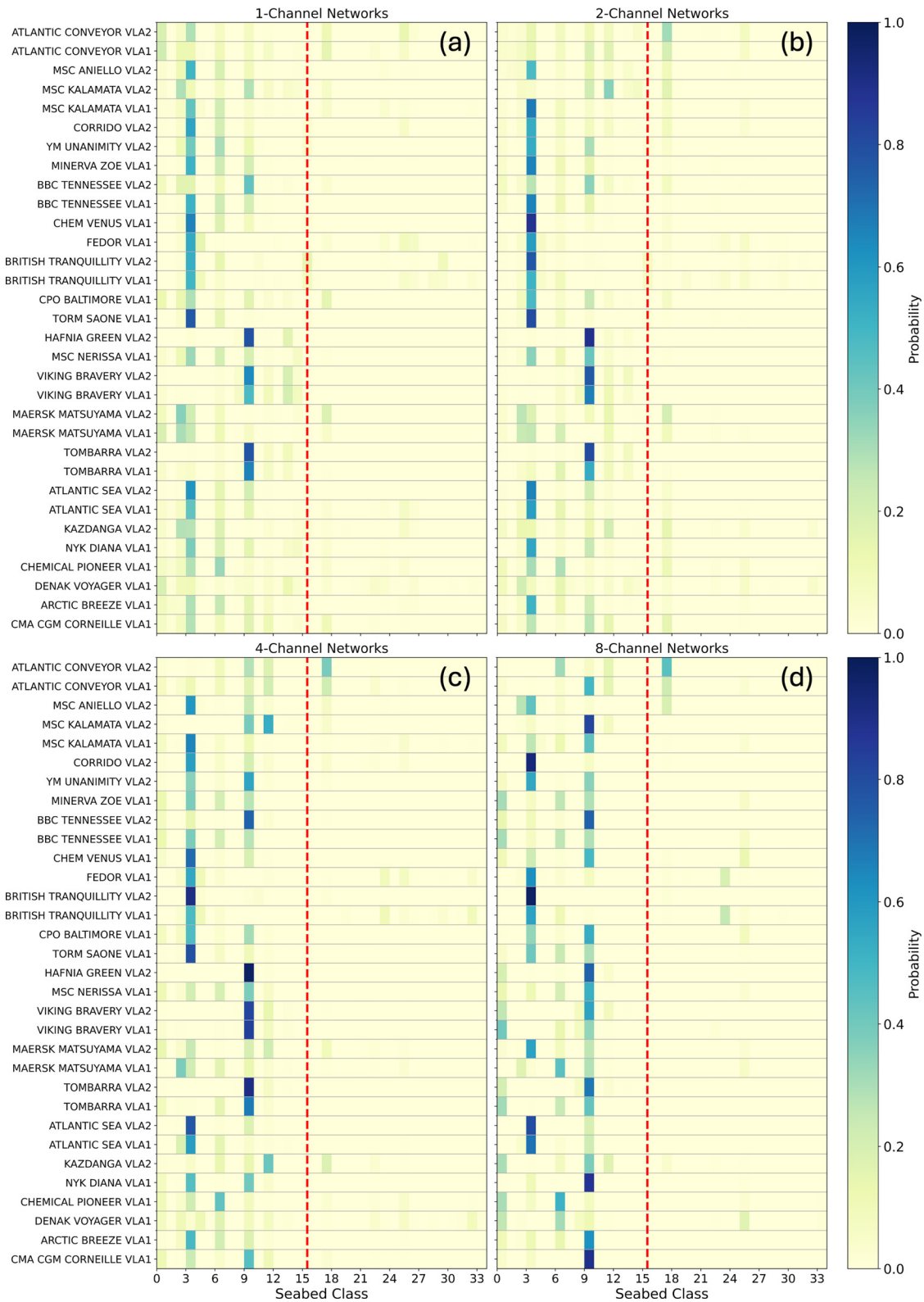


FIG. 18. Generalization results for 32 SOO spectrograms recorded during SBCEX 2017: average classifier output over all folds for all trained networks (similar to Fig. 13) for (a) one-channel networks, (b) two-channel networks, (c) four-channel networks, and (d) eight-channel networks. Ships are listed in chronological order from top to bottom. The red vertical dotted line is the same as in Fig. 13 (additional details about the ships are provided in Table II of Escobar *et al.* Ref. 34).

coarser in time proves to be sufficient and allows data from multiple hydrophone channels to be used. The spectrograms in this work are half as large as in previous work (only 121 time steps instead of 241).

This sparse time sampling still contains enough information for the networks to train well, as shown by Amos *et al.*<sup>41</sup> For the multichannel networks, we hypothesize that the impact of the reduced time resolution will be offset by



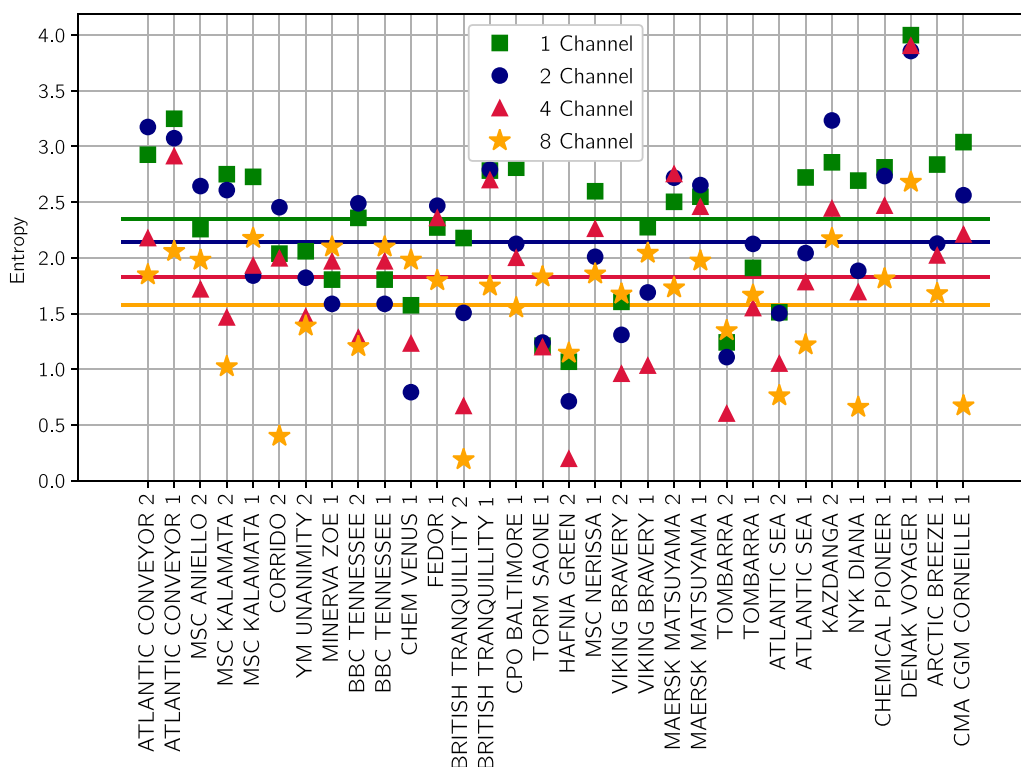


FIG. 19. Entropy of the predicted seabed distributions for each ship in Fig. 18. Values were calculated over the average classifier outputs obtained for networks trained using the same number of channels. Average entropy values over all ships for each number of channels are shown as horizontal lines, with values of approximately 2.34 for the ensemble of one-channel networks, 2.14 for two-channel networks, 1.83 for four-channel networks, and 1.58 for eight-channel networks.

using multiple channels. Specific applications will need to determine if the spatial variation due to multiple channels is more or less important than increased temporal resolution. Decisions regarding the size of the spectrograms and the number of channels must be balanced with the increased memory requirements and computational time required to train larger networks.

Our decision to use ensemble learning to compare the results from different channel combinations provides an estimate of the precision of the results and a measure of confidence when the correct seabed answer is not known. The advantages of an ensemble of networks trained on different combinations of channels are anticipated to increase as the variability in the water sound speed increases.

### V. CONCLUSIONS

This work has presented seabed classification results using multichannel SOO spectrograms via deep learning with ResNet-18. The networks, trained with relatively sparse synthetic spectrograms, have been applied to data samples of transiting ships measured in the New England Mud Patch during SBCEX 2017. The trained networks used data from one, two, four, or eight hydrophone channels on a VLA, and a catalog of 34 seabeds in the training data. In this work, we have illustrated the benefits of ensemble learning in providing a measure of precision and confidence, while evaluating the impact of using multiple receivers in the training data.

We additionally obtain an estimate of which seabed class provides a reasonably effective description of the seabed.

This study demonstrated how the ensemble approach increases the precision of seabed predictions. Our current preferred method is to average the classifier output over multiple networks, which have been trained using different combinations of channels on the VLA. The resulting classifier output provides an estimate of the precision of the results, which may be quantified by calculating the entropy of the distribution. The entropy of the ensemble-averaged distribution can then be used as a measure of confidence in the predictions. Future work should consider more advanced ensemble learning approaches.<sup>77</sup>

The results in this study show the advantages of using multiple channels in deep learning. These multiple channels are fed into the ResNet-18 network in parallel, similar to the RGB channels of a picture. In this work, a slight decrease in the entropy (increase in precision) is obtained when more channels are used. The resulting seabed predictions do not, however, change as the number of channels increases from two to four to eight. Thus, the increased computational time required to use more channels needs to be balanced with the resulting increase in precision.

The answer for the optimal number of channels to use in the deep learning networks likely depends on the variability in the water SSPs used in generating the synthetic training data. For the work presented in this paper, the SSPs were slightly upward refracting, similar to those measured

during SBCEX 2017. We anticipate that when more SSP variability is present in both the synthetic and the measured data samples, the advantages of using more channels will increase. One approach to finding optimal channel combinations has been introduced in Kurniawan *et al.*<sup>79</sup>

The seabed predictions from SOO spectrograms measured during SBCEX 2017 separate into two main groups, depending on what day the measurements were made. In both cases, however, the chosen seabeds were obtained from geoacoustic inversions of data from SBCEX 2017 and share similar properties. In particular, all chosen seabeds shared the similarity of having a sound speed ratio of less than one across the water-sediment interface and no large gradients of sound speed in the first sediment layer.

The lessons learned from this study can help guide development of more advanced networks for seabed classification or other applications in underwater acoustics. As more data samples become available from more recent experiments, the impact of changing SSPs and bathymetry should be investigated.

**ACKNOWLEDGMENTS**

The authors thank those who collected the SBCEX2017 data, the reviewers for helpful comments, and Mark Transtrum for ideas about how to interpret the results and quantify the entropy of the distributions. This work has been supported by the Office of Naval Research Grant No. N00014-22-1-2402. G.E.L. also thanks the National Science Foundation for Research Experiences for Undergraduates Grant No. 2051129.

**AUTHOR DECLARATIONS**

**Conflict of Interest**

The authors have no conflicts to disclose.

**DATA AVAILABILITY**

The data that support the findings of this study are available from the corresponding author upon reasonable request.

**APPENDIX: ADDITIONAL GENERALIZATION RESULTS**

For an indication of how the seabed classifier would perform in real-world applications, the average classifier outputs examined for 32 ships recorded on the two VLAs during SBCEX 2017 are shown in Fig. 18. The complete list of ships is found in Table II of Escobar-Amado *et al.*<sup>34</sup> Some of these spectrograms have unclear striations, changing speed, irregular striations, or additional artifacts. The performance of the seabed classifier on all the ship spectrograms illustrates how this method is likely to perform *in situ*. To facilitate the discussion of ambient noise statistics in Sec. IVC, the ships are listed chronologically from top to bottom.

Some of the SOO predictions in Fig. 18 clearly show Seabeds 3 and 9 with the highest classifier output. The SOO spectrograms that do not yield a strong preference for a single seabed were less clear than the others. For example, they may have less distinct SOO interference patterns (striations) or contain extraneous noise resulting in vertical or horizontal banding. Even for ships with less clear striations, the seabeds with the highest ensemble classifier output largely lie to the left of the red vertical dotted line. This line separates the seabeds with a sound speed ratio of less than one to the left, and greater than one to the right. With the exception of a few outliers, the predicted seabeds have a sound speed ratio of less than one.

The entropy of the distributions in Fig. 19 show a similar decrease in entropy as in Fig. 14. In this case, however, the eight-channel results have a slightly lower average entropy than the four-channel results. Because the generalization results are similar between the two cases, the use of four-channel models is recommended due to the reduced memory requirements.

<sup>1</sup>M. D. Collins and L. Fishman, “Efficient navigation of parameter landscapes,” *J. Acoust. Soc. Am.* **98**(3), 1637–1644 (1995).  
<sup>2</sup>M. C. Mortenson, T. B. Neilsen, M. K. Transtrum, and D. P. Knobles, “Accurate broadband gradient estimates enable local sensitivity analysis of ocean acoustic models,” *J. Theor. Comput. Acoust.* **31**(02), 2250015 (2023).  
<sup>3</sup>D. J. Forman, T. B. Neilsen, D. F. Van Komen, and D. P. Knobles, “Validating deep learning seabed classification via acoustic similarity,” *JASA Express Lett.* **1**(4), 040802 (2021).  
<sup>4</sup>K. Howarth, T. B. Neilsen, D. F. Van Komen, and D. P. Knobles, “Seabed classification using a convolutional neural network on explosive sounds,” *IEEE J. Oceanic Eng.* **47**(3), 670–679 (2022).  
<sup>5</sup>K. D. Heaney, “Rapid geoacoustic characterization using a surface ship of opportunity,” *IEEE J. Oceanic Eng.* **29**(1), 88–99 (2004).  
<sup>6</sup>C. Gervaise, B. G. Kinda, J. Bonnel, Y. Stéphan, and S. Vallez, “Passive geoacoustic inversion with a single hydrophone using broadband ship noise,” *J. Acoust. Soc. Am.* **131**(3), 1999–2010 (2012).  
<sup>7</sup>S.-H. Byun, C. Verlinden, and K. G. Sabra, “Blind deconvolution of shipping sources in an ocean waveguide,” *J. Acoust. Soc. Am.* **141**(2), 797–807 (2017).  
<sup>8</sup>L. Muzi, M. Siderius, and C. M. Verlinden, “Passive bottom reflection-loss estimation using ship noise and a vertical line array,” *J. Acoust. Soc. Am.* **141**(6), 4372–4379 (2017).  
<sup>9</sup>K. L. Gemba, J. Sarkar, B. Cornuelle, W. S. Hodgkiss, and W. Kuperman, “Estimating relative channel impulse responses from ships of opportunity in a shallow water environment,” *J. Acoust. Soc. Am.* **144**(3), 1231–1244 (2018).  
<sup>10</sup>L. Xu, K. Yang, and Q. Yang, “Joint time-frequency inversion for seabed properties of ship noise on a vertical line array in South China Sea,” *IEEE Access* **6**, 62856–62864 (2018).  
<sup>11</sup>X. Zhang, N. C. Durofchalk, H. Niu, L. Wu, R. Zhang, and K. G. Sabra, “Geoacoustic inversion using ray-based blind deconvolution of shipping sources,” *J. Acoust. Soc. Am.* **147**(1), 285–299 (2020).  
<sup>12</sup>A. A. Lunkov and B. G. Katsnelson, “Using discrete low-frequency components of shipping noise for gassy sediment characterization in shallow water,” *J. Acoust. Soc. Am.* **147**(5), EL428–EL433 (2020).  
<sup>13</sup>R. A. Koch and D. P. Knobles, “Geoacoustic inversion with ships as sources,” *J. Acoust. Soc. Am.* **117**(2), 626–637 (2005).  
<sup>14</sup>D. Tollefsen and S. E. Dosso, “Bayesian geoacoustic inversion of ship noise on a horizontal array,” *J. Acoust. Soc. Am.* **124**(2), 788–795 (2008).  
<sup>15</sup>T. W. Tan, O. A. Godin, B. G. Katsnelson, and M. Yarina, “Passive geoacoustic inversion in the Mid-Atlantic Bight in the presence of strong water column variability,” *J. Acoust. Soc. Am.* **147**(6), EL453–EL459 (2020).

- <sup>16</sup>D. Tollefsen, S. E. Dosso, and D. P. Knobles, "Ship-of-opportunity noise inversions for geoacoustic profiles of a layered mud-sand seabed," *IEEE J. Oceanic Eng.* **45**(1), 189–200 (2020).
- <sup>17</sup>M. Nicholas, J. S. Perkins, G. J. Orris, L. T. Fialkowski, and G. J. Heard, "Environmental inversion and matched-field tracking with a surface ship and an L-shaped receiver array," *J. Acoust. Soc. Am.* **116**(5), 2891–2901 (2004).
- <sup>18</sup>D. J. Battle, P. Gerstoft, W. A. Kuperman, W. S. Hodgkiss, and M. Siderius, "Geoacoustic inversion of tow-ship noise via near-field-matched-field processing," *IEEE J. Oceanic Eng.* **28**(3), 454–467 (2003).
- <sup>19</sup>C. Park, W. Seong, and P. Gerstoft, "Geoacoustic inversion in time domain using ship of opportunity noise recorded on a horizontal towed array," *J. Acoust. Soc. Am.* **117**(4), 1933–1941 (2005).
- <sup>20</sup>P. H. Dahl and D. R. Dall'Osto, "Vector acoustic analysis of time-separated modal arrivals from explosive sound sources during the 2017 seabed characterization experiment," *IEEE J. Oceanic Eng.* **45**(1), 131–143 (2020).
- <sup>21</sup>P. H. Dahl and D. R. Dall'Osto, "Estimation of seabed properties and range from vector acoustic observations of underwater ship noise," *J. Acoust. Soc. Am.* **147**(4), EL345–EL350 (2020).
- <sup>22</sup>S. E. Crocker, P. L. Nielsen, J. H. Miller, and M. Siderius, "Geoacoustic inversion of ship radiated noise in shallow water using data from a single hydrophone," *J. Acoust. Soc. Am.* **136**(5), EL362–EL368 (2014).
- <sup>23</sup>S. A. Stotts, R. A. Koch, S. M. Joshi, V. T. Nguyen, V. W. Ferreri, and D. P. Knobles, "Geoacoustic inversions of horizontal and vertical line array acoustic data from a surface ship source of opportunity," *IEEE J. Oceanic Eng.* **35**(1), 79–102 (2010).
- <sup>24</sup>D. P. Knobles, "Maximum entropy inference of seabed attenuation parameters using ship radiated broadband noise," *J. Acoust. Soc. Am.* **138**(6), 3563–3575 (2015).
- <sup>25</sup>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).
- <sup>26</sup>Y. Stephan, X. Demoulin, and O. Sarzeaud, "Neural direct approaches for geoacoustic inversion," *J. Comput. Acoust.* **06**(01n02), 151–166 (1998).
- <sup>27</sup>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).
- <sup>28</sup>J. Benson, N. R. Chapman, and A. Antoniou, "Geoacoustic model inversion using artificial neural networks," *Inverse Prob.* **16**(6), 1627 (2000).
- <sup>29</sup>L. Mao, X. Pan, and Y. Shen, "Geoacoustic inversion based on neural network," in *Proceedings of Oceans 2021*, San Diego–Porto (2021), pp. 1–5.
- <sup>30</sup>J. Piccolo, G. Haramuniz, and Z.-H. Michalopoulou, "Geoacoustic inversion with generalized additive models," *J. Acoust. Soc. Am.* **145**(6), EL463–EL468 (2019).
- <sup>31</sup>H. Niu, P. Gerstoft, R. Zhang, Z. Li, Z. Gong, and H. Wang, "Mode separation with one hydrophone in shallow water: A sparse Bayesian learning approach based on phase speed," *J. Acoust. Soc. Am.* **149**(6), 4366–4376 (2021).
- <sup>32</sup>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).
- <sup>33</sup>T. Neilsen, C. Escobar-Amado, M. Acree, W. Hodgkiss, D. Van Komen, D. Knobles, M. Badiey, and J. Castro-Correa, "Learning location and seabed type from a moving mid-frequency source," *J. Acoust. Soc. Am.* **149**(1), 692–705 (2021).
- <sup>34</sup>C. D. Escobar-Amado, T. B. Neilsen, J. A. Castro-Correa, D. F. Van Komen, M. Badiey, D. P. Knobles, and W. S. Hodgkiss, "Seabed classification from merchant ship-radiated noise using a physics-based ensemble of deep learning algorithms," *J. Acoust. Soc. Am.* **150**(2), 1434–1447 (2021).
- <sup>35</sup>D. F. Van Komen, T. B. Neilsen, D. B. Mortenson, M. C. Acree, D. P. Knobles, M. Badiey, and W. S. Hodgkiss, "Seabed type and source parameters predictions using ship spectrograms in convolutional neural networks," *J. Acoust. Soc. Am.* **149**(2), 1198–1210 (2021).
- <sup>36</sup>A. Varon, J. Mars, and J. Bonnel, "Approximation of modal wavenumbers and group speeds in an oceanic waveguide using a neural network," *JASA Express Lett.* **3**(6), 066003 (2023).
- <sup>37</sup>A. Vardi and J. Bonnel, "End-to-end geoacoustic inversion with neural networks in shallow water using a single hydrophone," *IEEE J. Oceanic Eng.* **49**, 380–388 (2024).
- <sup>38</sup>C. Frederick, S. Villar, and Z.-H. Michalopoulou, "Seabed classification using physics-based modeling and machine learning," *J. Acoust. Soc. Am.* **148**(2), 859–872 (2020).
- <sup>39</sup>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).
- <sup>40</sup>Bureau of Ocean Energy Management and National Oceanic and Atmospheric Administration, "Automatic identification system, vessel traffic data [AIS\_2017\_03\_Zone19 and AIS\_2017\_04\_Zone19]," available at [MarineCadastre.gov](https://www.marinecadastre.gov) (Last viewed October 10, 2024).
- <sup>41</sup>S. M. Amos, D. B. Mortenson, T. B. Neilsen, D. P. Knobles, and W. S. Hodgkiss, "Determining optimal time interval and frequency band of ship noise spectrograms for seabed classification," *Proc. Mtgs. Acoust.* **47**, 070016 (2022).
- <sup>42</sup>S. C. Wales and R. M. Heitmeyer, "An ensemble source spectra model for merchant ship-radiated noise," *J. Acoust. Soc. Am.* **111**(3), 1211–1231 (2002).
- <sup>43</sup>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).
- <sup>44</sup>S. Kaufman, S. Rosset, C. Perlich, and O. Stitelman, "Leakage in data mining: Formulation, detection, and avoidance," *ACM Trans. Knowl. Discov. Data* **6**(4), 1–21 (2012).
- <sup>45</sup>Y. T. Lin, J. Bonnel, D. P. Knobles, and P. S. Wilson, "Broadband waveform geoacoustic inversions with absolute travel time," *IEEE J. Oceanic Eng.* **45**(1), 174–188 (2020).
- <sup>46</sup>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**(1), 161–173 (2020).
- <sup>47</sup>E. M. Brown, Y. T. Lin, J. D. Chaytor, and W. L. Siegmann, "Geoacoustic inversion for a New England Mud Patch sediment using the silt-suspension theory of marine mud," *IEEE J. Oceanic Eng.* **45**(1), 144–160 (2020).
- <sup>48</sup>Z. H. Michalopoulou and P. Gerstoft, "Multipath broadband localization, bathymetry, and sediment inversion," *IEEE J. Oceanic Eng.* **45**(1), 92–102 (2020).
- <sup>49</sup>J. Belcourt, C. W. Holland, S. E. Dosso, J. Dettmer, and J. A. Goff, "Depth-dependent geoacoustic inferences with dispersion at the New England Mud Patch via reflection coefficient inversion," *IEEE J. Oceanic Eng.* **45**(1), 69–91 (2020).
- <sup>50</sup>F. B. Jensen, W. A. Kuperman, M. B. Porter, and H. Schmidt, *Computational Ocean Acoustics*, 2nd ed. (Springer Publishing Company, New York, 2011), p. 39.
- <sup>51</sup>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).
- <sup>52</sup>L. Wan, M. Badiey, D. P. Knobles, and P. S. Wilson, "The Airy phase of explosive sounds in shallow water," *J. Acoust. Soc. Am.* **143**(3), EL199–EL205 (2018).
- <sup>53</sup>S. E. Dosso, P. M. Giles, G. H. Brooke, D. F. McCammon, S. Pecknold, and P. C. Hines, "Linear and nonlinear measures of ocean acoustic environmental sensitivity," *J. Acoust. Soc. Am.* **121**(1), 42–45 (2007).
- <sup>54</sup>D. P. Knobles, R. A. Koch, and M. S. Haire, "Geoacoustic inversion studies," in *Oceans Conf. Rec.* (IEEE, New York, 2001), Vol. 2, pp. 737–744.
- <sup>55</sup>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).
- <sup>56</sup>W.-H. Ryang, J.-H. Jin, and J. Hahn, "Geoacoustic model at the YSDP-105 long-core site in the mid-eastern Yellow Sea," *J. Korean Earth Sci. Soc.* **40**(1), 24–36 (2019).
- <sup>57</sup>D. R. Barclay, D. A. Bevans, and M. J. Buckingham, "Estimation of the geoacoustic properties of the New England Mud Patch from the vertical coherence of the ambient noise in the water column," *IEEE J. Oceanic Eng.* **45**(1), 51–59 (2020).
- <sup>58</sup>H. Kwon, J. W. Choi, W.-H. Ryang, S.-U. Son, and S.-K. Jung, "Measurements of mid-frequency bottom-interacting signals and geoacoustic inversion in Jinhae Bay, southeast Korea," *J. Acoust. Soc. Am.* **145**(3), 1205–1211 (2019).
- <sup>59</sup>J. Bonnel, Y.-T. Lin, D. Eleftherakis, J. A. Goff, S. Dosso, R. Chapman, J. H. Miller, and G. R. Potty, "Geoacoustic inversion on the New England

- Mud Patch using warping and dispersion curves of high-order modes,” *J. Acoust. Soc. Am.* **143**(5), EL405–EL411 (2018).
- <sup>60</sup>L. Wan, M. Badiéy, D. P. Knobles, P. S. Wilson, and J. A. Goff, “Estimates of low-frequency sound speed and attenuation in a surface mud layer using low-order modes,” *IEEE J. Oceanic Eng.* **45**(1), 201–211 (2020).
- <sup>61</sup>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).
- <sup>62</sup>D. P. Knobles, J. A. Goff, R. A. Koch, P. S. Wilson, and J. A. Shooter, “Effect of inhomogeneous sub-bottom layering on broadband acoustic propagation,” *IEEE J. Oceanic Eng.* **35**(4), 732–743 (2010).
- <sup>63</sup>Y.-M. Jiang, N. R. Chapman, and M. Badiéy, “Quantifying the uncertainty of geoacoustic parameter estimates for the New Jersey shelf by inverting air gun data,” *J. Acoust. Soc. Am.* **121**(4), 1879–1894 (2007).
- <sup>64</sup>J. Bonnel, S. E. Dosso, D. Eleftherakis, and N. R. Chapman, “Trans-dimensional inversion of modal dispersion data on the New England Mud Patch,” *IEEE J. Oceanic Eng.* **45**(1), 116–130 (2020).
- <sup>65</sup>W.-H. Ryang, S.-P. Kim, D.-C. Kim, and J. Hahn, “Geoacoustic model of coastal bottom strata at Jeongdongjin in the Korean continental margin of the East Sea,” *J. Korean Earth Sci. Soc.* **37**(4), 200–210 (2016).
- <sup>66</sup>K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV (2016), pp. 770–778.
- <sup>67</sup>Y. Bengio, I. Goodfellow, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2017), Vol. 1, pp. 180–184.
- <sup>68</sup>G. E. Lau and T. B. Neilsen, “Validating seabed classification of spectrograms from multiple channels on a vertical line array,” *Proc. Mtgs. Acoust.* **51**, 055003 (2023).
- <sup>69</sup>M. Stone, “Cross-validators choice and assessment of statistical predictions,” *J. R. Statistical Soc.: Ser. B* **36**(2), 111–133 (1974).
- <sup>70</sup>D. F. Van Komen, K. Howarth, T. B. Neilsen, D. P. Knobles, and P. H. Dahl, “A CNN for range and seabed estimation on normalized and extracted time-series impulses,” *IEEE J. Oceanic Eng.* **47**(3), 833–846 (2022).
- <sup>71</sup>D. Knobles, T. B. Neilsen, W. S. Hodgkiss, and J. A. Goff, “Inference of source signatures of merchant ships in shallow ocean environments,” *J. Acoust. Soc. Am.* **155**(5), 3144–3155 (2024).
- <sup>72</sup>I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” *arXiv:1711.05101* (2017).
- <sup>73</sup>L. Holmstrom and P. Koistinen, “Using additive noise in back-propagation training,” *IEEE Trans. Neural Networks* **3**(1), 24–38 (1992).
- <sup>74</sup>G. An, “The effects of adding noise during backpropagation training on a generalization performance,” *Neural Comput.* **8**(3), 643–674 (1996).
- <sup>75</sup>S. Zheng, Y. Song, T. Leung, and I. Goodfellow, “Improving the robustness of deep neural networks via stability training,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 4480–4488.
- <sup>76</sup>J. Castro-Correa, M. Badiéy, T. Neilsen, D. Knobles, and W. Hodgkiss, “Impact of data augmentation on supervised learning for a moving mid-frequency source,” *J. Acoust. Soc. Am.* **150**(5), 3914–3928 (2021).
- <sup>77</sup>R. Polikar, *Ensemble Learning* (Springer, New York, 2012), pp. 1–34.
- <sup>78</sup>Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- <sup>79</sup>Y. Kurniawan, T. B. Neilsen, B. L. Francis, A. M. Stankovic, M. Wen, I. Nikiforov, E. B. Tadmor, V. V. Bulatov, V. Lordi, and M. K. Transtrum, “An information-matching approach to optimal experimental design and active learning,” *arXiv:2411.02740* (2024).