

JUNE 20 2022

Determining optimal time interval and frequency band of ship noise spectrograms for seabed classification

Stephen Michael Amos; Daniel B. Mortenson; Tracianne B. Neilsen; ... et. al



Proc. Mtgs. Acoust 47, 070016 (2022)

<https://doi.org/10.1121/2.0001625>



View
Online



Export
Citation

CrossMark

Related Content

Influence of seabed on very low frequency sound recorded during passage of merchant ships on the New England shelf

J Acoust Soc Am (May 2021)

Correlation of deep seabed structure with waveguide invariant β in $\pm\infty$ regime

J Acoust Soc Am (April 2021)

Mid-frequency sound propagation over a mud seabed with a compressional sound speed depth gradient

J Acoust Soc Am (March 2023)



Advance your science and career
as a member of the

ACOUSTICAL SOCIETY OF AMERICA

LEARN MORE





ICUA2022
International Conference on
Underwater Acoustics

International Conference on Underwater Acoustics

Southampton, UK

20-23 June 2022



Underwater Acoustics: Seabed and sediment classification

Determining optimal time interval and frequency band of ship noise spectrograms for seabed classification

Stephen Michael Amos, Daniel B. Mortenson and Tracianne B. Neilsen
Department of Physics and Astronomy, Brigham Young University, Provo, UT, 84602, USA;
amos.stephen.m@gmail.com; dbmort96@gmail.com; tbn@byu.edu

David P. Knobles
KSA LLC, Austin, TX, USA; dpknobles@kphysics.org

William S. Hodgkiss
Marine Physical Laboratories, University of California San Diego Scripps Institution of Oceanography, San Diego, CA, USA; whodgkiss@ucsd.edu

Because the seabed affects shallow ocean sound propagation, an interesting potential application of machine learning is to estimate acoustical properties of the seabed. In this study, surface ship noise spectrograms are employed in a residual neural network architecture to estimate an effective seabed class that produces similar sound propagation effects. As a follow up to the initial studies, this work seeks to determine the spectrogram time duration and frequency band that yield optimal results. The total time interval is varied between 5 and 20 minutes, and frequency bands range between 360-800Hz and 300-1500Hz to create nine different datasets. For each time-frequency band combination, a training dataset is created using a broadband source model and a range independent, normal-mode propagation model with 34 acoustically distinct seabed classes and a wide range of ship parameters. Networks are trained with five-fold cross validation then applied to ship noise spectrograms collected during a 2017 research excursion in the New England mud patch area, called Seabed Characterization Experiment. Results using a single hydrophone in the middle of the water column are presented along with conclusions about the total time and frequency band of spectrograms that yield the best generalization ability of trained networks.



1. INTRODUCTION

As sound propagates through the shallow ocean, it encodes geoacoustic information about the seabed based upon the reflection and transmission that occurs. A sound signal received by a hydrophone placed in the shallow ocean can thus be used to infer information about the seabed, a technique known as geoacoustic inversion. Different machine learning methods are now being applied to learn seabed properties, with some studies using regression for individual sediment parameters¹⁻⁵ or modal properties⁶⁻¹⁰ and others performing classification of seabed type.¹¹⁻¹³

The various studies have used different sound sources. Our work uses the noise from cargo ships, or Ships of Opportunity (SOO), which are an regularly occurring sound source in the shallow ocean. The sound received from a ship by a hydrophone can be represented as a spectrogram. Treating spectrograms as images, this study uses them as input data for a deep learning network to predict seabed type from a library of 34 seabeds. This study draws techniques from the broader field of image classification in machine learning to try to obtain accurate predictions for seabed type. This work builds on the initial studies of using SOO spectrograms for seabed classification^{12,13} to specifically address the question of how the selected time interval and frequency band influence the seabed classification.

2. METHODS

A. MEASURED DATA

The measured data for this experiment were collected during the Seabed Characterization Experiment in 2017 (SBCEX 2017) in the New England Mudpatch area, near two shipping lanes.¹⁴ The position of two vertical line arrays (VLAs), deployed by Marine Physical Laboratories: Scripps Institute of Oceanography, are shown in Fig. 1 in relation to the two shipping lanes. For this study, only the data from a single channel at approximately the center of each VLA were used. With an average water depth of 75 m, the middle hydrophone on each VLA was at an approximate depth of 42 m. While the VLAs were recording, 31 SOO passed by while these two VLAs were deployed; the names and positions of the ship were later identified using AIS data from the Marine Cadastre and are listed with their corresponding closest point of approach (CPA) range and speed in columns R-1 and R-2 of Table II in Ref. 13. Travelling in a straight line at a constant speed, the noise recorded from most of these SOO yielded spectrograms. The frequency band of interest was chosen to be above 300 Hz for two reasons. First, lower frequencies contained more background noise. Second, lower frequencies are more sensitive to the structure of deeper sediment layers; by looking at frequencies above 300 Hz, the seabed classification efforts focus on the upper properties of the seabed. Examples of these spectrograms are shown in Fig. 2.

B. SYNTHETIC DATA

Deep learning requires large quantities of training data, and collecting measured data samples for this study is expensive and time-consuming. SBCEX 2017 produced an insufficient number of data samples to use for training, and not enough is known about the seabeds where the samples were collected to provide labels for seabed type, beyond being muddy environments. To provide a sufficient quantity of training data, synthetic data samples were produced using the

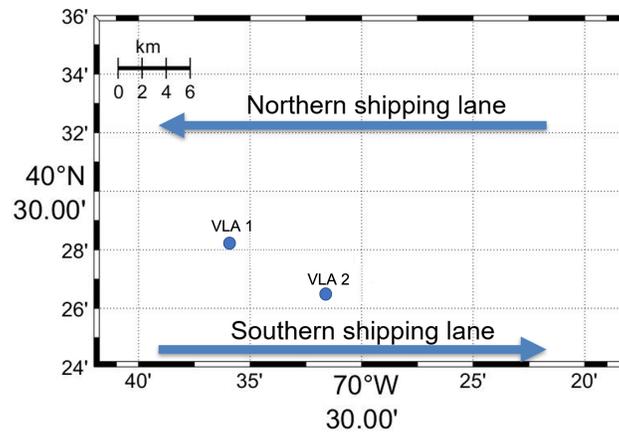


Figure 1: Two VLAs from Marine Physical Laboratories: Scripps Institute of Oceanography were placed between two shipping lanes in the New England Mudpatch area.

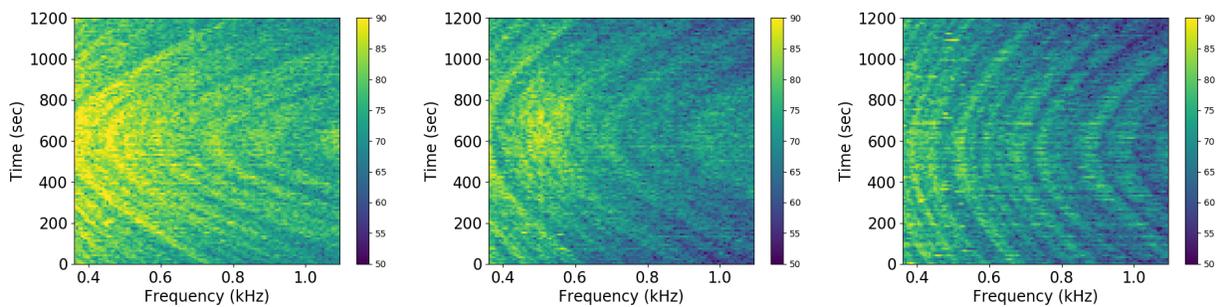


Figure 2: SOO Spectrograms. The ships shown from left to right are Kalamata, received on VLA2; Viking Bravery, VLA1; and Maersk Matsuma, VLA2.

range-independent, normal-mode propagation model, ORCA,¹⁵ with the Wales-Heitmeyer source spectrum¹⁶ for ship noise as the simulated source, as in Refs. 12 and 13.

From the available literature, 34 acoustically distinct seabeds – each with a Pearson correlation for acoustic similarity of 0.8 or less¹⁷ – were chosen to represent the range of possible seabeds. These are listed in order of increasing sound speed in Table III of Ref. 13. Values for ship CPA, ship speed and source depth are varied randomly between 0.5 and 15 km, 8 and 20 knots, and 6 and 12 m, yielding 405 data samples for each of the 34 environments, a total of 13770 samples. Further details are found in Table IV of Ref. 13. Each spectrogram has the CPA in the center of the time interval. Previous studies used ten different but similar sound speed profiles when generating synthetic data, whereas this study only uses one to reduce data leakage as discussed later.

The main question of this study is which time intervals and frequency bands for input spectrograms are optimal. In Ref. 12, 15 minute spectrograms with a 300-1500 Hz frequency band were used. Each data sample had 301 time steps and 198 frequencies. These data were used in three- and five-layer convolutional neural networks (CNNs) and half of AlexNet¹⁸ using multitask learning to classify between four seabed classes and estimate ship speed and CPA range via regression. In Ref. 13, 20 minute spectrograms with a 360-1100 Hz frequency band were used. A low frequency of 360 Hz was used as the measured data samples contained a strong tone at approximately 350 Hz from the research vessel. Each data sample had 243 time steps and 123 frequencies. These data

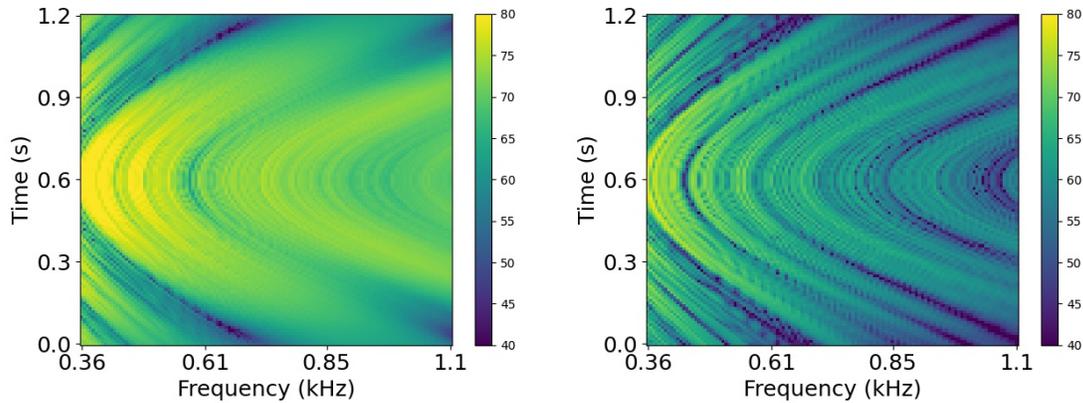


Figure 3: Synthetic spectrograms produced using ORCA for seabeds 13 (left) and 11 (right) with the same source-receiver configuration.

were used in six different CNNs to classify between the 34 seabed classes previously mentioned. From among these six networks, ResNet-18¹⁹ was found to have the best generalisation ability.

The current work uses ResNet-18 with smaller spectrograms (121 time steps, 123 frequencies) to study what time interval and frequency band are needed for a deep learning model to adequately classify seabeds from the 34 representative classes. This study uses time intervals of 20, 10 and 5 minutes and frequency bands of 300-1500, 360-1100 and 360-800 Hz for a total of 9 training datasets to consider whether smaller time intervals and frequency bands yield comparable results to wider intervals and bands, which would reduce the amount of data needed for seabed classification.

Two additional considerations were made in creating training data. In Ref. 12 they considered whether it is more beneficial to train a network using spectral density represented as complex pressure or levels. They found that seabed classification was better using levels, likely because the impact of the seabed on the sound propagation is strongly reflected in the transmission loss. In this and the two companion studies mentioned, each spectrogram is individually scaled by its standard deviation to remove the impact of an unknown source level. In Ref. 16, they assume the base value of the mean spectrum (S_0 in Eq. (1) of Ref.12) to be 230 dB. In practice, however, the actual value best suited to the noise for each ship varies and is unknown. To enable the network to generalise independent of S_0 , each spectrogram is individually scaled by its standard deviation prior to using the data for network training or predictions. The normalization in this and the companion studies have been performed on the squared spectral density values that are then converted to levels (dB re $1\mu\text{Pa}$). Additional options for scaling need to be investigated in future work.

C. RESNET-18

The deep learning model used for this study was ResNet-18, an 18-layer residual network containing seventeen convolutional layers followed by one fully connected layer and a softmax classifier output that gives a value between 0 and 1 for each of the 34 seabed classes. With all the advantages of a CNN, it also contains identity mapping through shortcut connections between convolutional layers, which help to preserve information throughout the network. This allows for a deeper network with more layers, while avoiding the associated increase in training error that can occur in CNNs.¹⁹

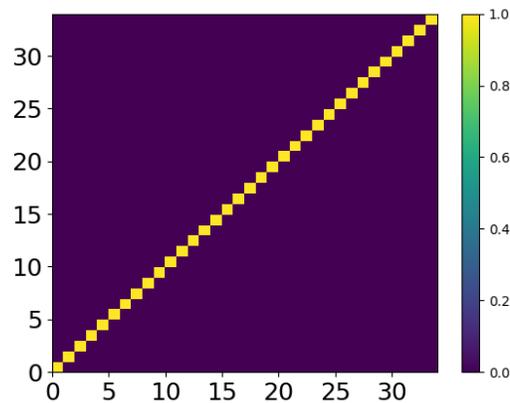


Figure 4: Confusion Matrix showing average validation accuracy for a model trained on the 360-1100 Hz, 20 minutes dataset.

The following hyperparameters were used in training the network. A batch size of 512 was used. The AdamW stochastic optimizer²⁰ was used along with a cosine annealing learning rate scheduler. The network was allowed to train for a maximum of 100 epochs with early stopping enabled, which ended the training process if a validation accuracy improvement of 0.001 did not occur within five epochs. This early stopping was intended to reduce overfitting.

Five-fold cross-validation was used to obtain a statistical representation of model performance.²¹ The training dataset was split into five sections or folds, four of which were used to train the model while the fifth was reserved as a testing dataset. The model was trained five times, each time using a different fold as the test dataset. This approach produced five separate models with the same parameters for input spectrograms but potentially different prediction performance.

The model's performance when predicting seabeds from the simulated test data is the *validation accuracy* of the model because the test data are drawn from the same statistical distribution as the training data. The *generalisation accuracy* of the model relates to the predictions on the measured data.

3. RESULTS

In this work, it was considered that using ten similar sound speeds in synthetic data generation, as was done in Ref. 13, leads to data samples similar enough to each other that the model is predicting on test spectrograms almost identical to those seen in training, thus giving an inflated validation accuracy metric without really learning to generalise. This study chose to train on only one sound speed profile to reduce this effect.

Even with using just one sound speed profile, validation accuracy was typically above 98%, often achieving 100% accuracy across three or more folds. For the case represented by the confusion matrix in Fig. 4, four out of five folds achieved 100% validation accuracy, with a fifth fold achieving 99.383%. Validation accuracy this high is highly uncommon, indicating that data leakage is almost certainly occurring. Data leakage occurs when there is significant overlap between the training and testing data, meaning that the validation accuracy shows the model's ability to memo-

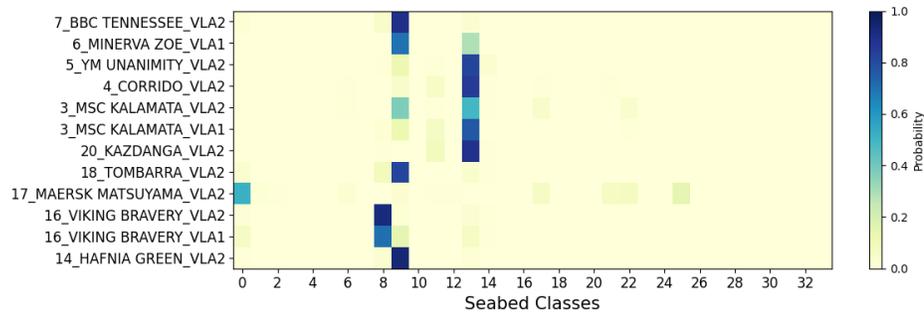


Figure 5: Classifier output for predictions on SOO ship data. Averaged output of 5 folds for the model trained on the 360-1100 Hz, 20 minutes dataset.

size rather than generalise.²² With no data leakage occurring, one would expect the generalisation accuracy to be comparable to the validation accuracy.

While validation accuracy is used to see if the model is learning and to inform the tuning of hyperparameters, the real test of model accuracy comes when it is asked to generalise, or make predictions on measured data. Measured spectrograms of twelve ships were given as input to the five trained models from a network trained on the dataset containing spectrograms with a frequency range of 360-1100 Hz and a time interval of twenty minutes. Fig. 5 shows the averaged output. Seabeds zero through sixteen indicate muddy environments with a sound speed ratio at the sediment-water interface that is less than one. While the unlabeled measured data makes it hard to validate generalizations, the measured data samples were taken in a muddy area, so predicted seabeds should be within this zero through sixteen range. Two seabeds are clearly predicted with higher confidence than any others, both in this range.

Measured spectrograms of twelve ships were given as input to the five trained models produced by each of the nine training datasets. Fig. 6 shows a comparison between the nine cases used in this study for different frequency bands (columns) and time intervals (rows).

One question asked in this study was the minimum time interval needed to make accurate predictions. Shown in the bottom row of the figure, five minute spectrograms appear to give a good general indication of environment, with most of the predictions occurring in the zero through sixteen range, along with some outliers. Surprisingly, increasing the length of the time interval to ten minutes, as shown in the middle row, doesn't show significant improvement and increases the number of outliers in predictions made with the 360-1100 Hz frequency band. Increasing the time interval to twenty minutes, however, decreases the spread of results and increases the probability with which the model predicts specific seabeds.

The second question concerned the optimal frequency band. The first column shows the 300-1500 Hz results. While this frequency band was used in Ref. 12, the lower limit was raised to 360 Hz in order to exclude the tones emitted by the research vessel. Increasing this lower limit appears to increase prediction certainty as seen in the second and third columns. Comparison between the columns indicates that frequencies between 800 Hz to 1100 Hz do contain information which gives the most certainty with fewest outliers. Overall, the 20 minute spectrograms with a frequency band of 360-1100 Hz have the fewest outliers.

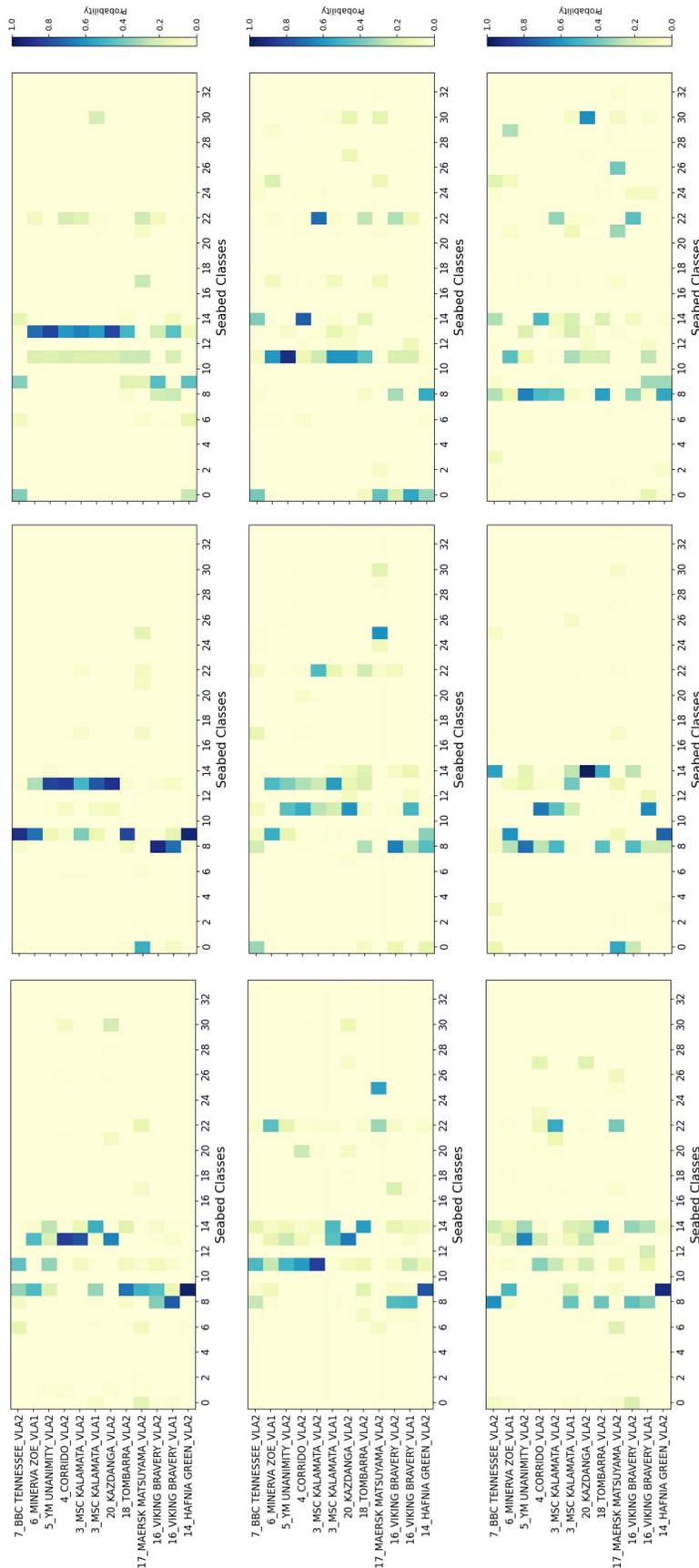


Figure 6: Classifier Output from models trained with the nine different datasets. The columns represent the three frequency bands: 300-1500 Hz (left), 360-800 Hz (middle) and 360-1100 Hz (right). The rows represent the three time intervals: 20 minutes (top), 10 minutes (middle) and 5 minutes (bottom). In all cases, the input spectrograms contain 123 frequencies and 121 time steps.

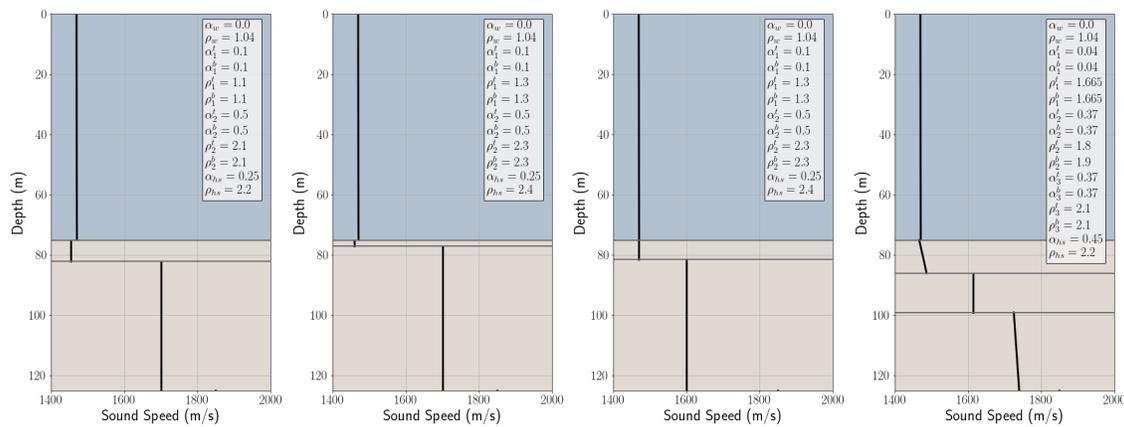


Figure 7: Properties of seabeds 8, 9, 13 and 11 from left to right.

While predictions varied between all nine cases, four seabeds were chosen most often: 8, 9, 11 and 13. Properties of the four seabeds are shown in Fig. 7. The first three have very similar geoacoustic profiles, indicating that more distinct seabed classes may be needed to improve classification. The fourth has a similar sound speed ratio at the sediment-water interface, which is likely the most important factor in these frequency ranges.

4. CONCLUSION

This work examined the necessary time interval and frequency bands to perform seabed classification on ship noise spectrograms. Five minute spectrograms are enough to get a general idea of seabed sediment type but predictions contain outliers. Twenty minute spectrograms give better precision. Omitting the 350 Hz tones that were emitted by the research vessel improves generalization. Another approach to account for these tones would be to include the tones in the training data so the model learns to ignore them. Including frequencies from 800 to 1100 Hz appears to reduce uncertainty. Future studies could explore including even higher frequencies such as up to 1500 Hz with the increased lower limit of 360 Hz. Across the networks trained with the nine datasets, only four seabeds had a high probability of being selected. As the geoacoustic profiles of these four are very similar, more work is needed to improve the library of seabeds for more robust classification.

ACKNOWLEDGMENTS

This work is sponsored by the Office of Naval Research, Grant No. N00014-22-1-2402. The undergraduate research assistantship for Stephen M. Amos was funded by the College of Physical and Mathematical Sciences, Brigham Young University, made possible by the generosity of donors.

REFERENCES

- ¹ M. Liu, H. Niu, Z. Li, Y. Liu, and Q. Zhang, “Deep-learning geoacoustic inversion using multi-range vertical array data in shallow water,” *The Journal of the Acoustical Society of America* **151**(3), 2101–2116 (2022).
- ² L. Mao, X. Pan, and Y. Shen, “Geoacoustic inversion based on neural network,” in *OCEANS 2021: San Diego–Porto*, IEEE (2021), pp. 1–5.
- ³ J. Piccolo, G. Haramuniz, and Z.-H. Michalopoulou, “Geoacoustic inversion with generalized additive models,” *The Journal of the Acoustical Society of America* **145**(6), EL463–EL468 (2019).
- ⁴ Y. Shen, X. Pan, Z. Zheng, and P. Gerstoft, “Matched-field geoacoustic inversion based on radial basis function neural network,” *The Journal of the Acoustical Society of America* **148**(5), 3279–3290 (2020) 10.1121/10.0002656.
- ⁵ B. B. Thompson, R. J. Marks, M. A. El-Sharkawi, W. J. Fox, and R. T. Miyamoto, “Inversion of neural network underwater acoustic model for estimation of bottom parameters using modified particle swarm optimizers,” in *Proceedings of the International Joint Conference on Neural Networks, 2003.*, IEEE (2003), Vol. 2, pp. 1301–1306.
- ⁶ B. Gao, J. Pang, X. Li, W. Song, and W. Gao, “Recovering reverberation interference striations by a conditional generative adversarial network,” *JASA Express Letters* **1**(5), 056001 (2021) 10.1121/10.0004907.
- ⁷ X. Li, W. Song, D. Gao, W. Gao, and H. Wang, “Training a U-Net based on a random mode-coupling matrix model to recover acoustic interference striations,” *The Journal of the Acoustical Society of America* **147**(4), EL363–EL369 (2020) 10.1121/10.0001125.
- ⁸ H. Niu, P. Gerstoft, E. Ozanich, Z. Li, R. Zhang, Z. Gong, and H. Wang, “Block sparse Bayesian learning for broadband mode extraction in shallow water from a vertical array,” *The Journal of the Acoustical Society of America* **147**, 3729–3739 (2020) 10.1121/10.0001322.
- ⁹ 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,” *The Journal of the Acoustical Society of America* **149**(6), 4366–4376 (2021) 10.1121/10.0005312.
- ¹⁰ T. Paviet-Salomon, J. Bonnel, C. Dorffer, B. Nicolas, T. Chonavel, D. Tollefsen, D. P. Knobles, P. S. Wilson, and A. Drémeau, “Estimation of frequency-wavenumber diagrams using a physics-based grid-free compressed sensing method,” *IEEE Journal of Oceanic Engineering* **45**(1), 565–577 (2021) 10.1109/JOE.2021.3109432.
- ¹¹ C. Frederick, S. Villar, and Z. H. Michalopoulou, “Seabed Classification using Physics-based Modeling and Machine Learning,” *The Journal of the Acoustical Society of America* **148**, 859–872 (2020) 10.1121/10.0001728.

-
- ¹² D. F. Van Komen, T. B. Neilsen, D. B. Mortenson, M. C. Acree, D. P. Knobles, M. Badiy, and W. S. Hodgkiss, “Seabed type and source parameters predictions using ship spectrograms in convolutional neural networks,” *The Journal of the Acoustical Society of America* **149**(2), 1198–1210 (2021).
- ¹³ C. D. Escobar-Amado, T. B. Neilsen, J. A. Castro-Correa, D. F. Van Komen, M. Badiy, D. P. Knobles, and W. S. Hodgkiss, “Seabed classification from merchant ship-radiated noise using a physics-based ensemble of deep learning algorithms,” *The Journal of the Acoustical Society of America* **150**(2), 1434–1447 (2021).
- ¹⁴ P. S. Wilson, D. P. Knobles, and T. B. Neilsen, “Guest editorial an overview of the seabed characterization experiment,” *IEEE Journal of Oceanic Engineering* **45**(1), 1–13 (2020).
- ¹⁵ E. K. Westwood, C. T. Tindle, and N. R. Chapman, “A normal mode model for acousto-elastic ocean environments,” *The Journal of the Acoustical Society of America* **100**(6), 3631–3645 (1996) 10.1121/1.417226.
- ¹⁶ S. C. Wales and R. M. Heitmeyer, “An ensemble source spectra model for merchant ship-radiated noise,” *The Journal of the Acoustical Society of America* **111**(3), 1211–1231 (2002).
- ¹⁷ D. J. Forman, T. B. Neilsen, D. F. Van Komen, and D. P. Knobles, “Validating deep learning seabed classification via acoustic similarity,” *JASA Express Letters* **1**(4), 040802 (2021).
- ¹⁸ A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems* (2012), pp. 1097–1105.
- ¹⁹ 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* (2016), pp. 770–778.
- ²⁰ I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” arXiv preprint arXiv:1711.05101 (2017).
- ²¹ R. Kohavi, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Ijcai*, Montreal, Canada (1995), Vol. 14, pp. 1137–1145.
- ²² A. Elangovan, J. He, and K. Verspoor, “Memorization vs. generalization: quantifying data leakage in nlp performance evaluation,” arXiv preprint arXiv:2102.01818 (2021).