NOVEMBER 01 2024

Validating seabed classification of spectrograms from multiple channels on a vertical line array FREE

Ginger E. Lau; Tracianne B. Neilsen

Proc. Mtgs. Acoust. 51, 055003 (2023)
https://doi.org/10.1121/2.0001971

ASA POMA Proceedings of Meetings on Acoustics

Volume 51 http://acousticalsociety.org/

184th Meeting of the Acoustical Society of America

Chicago, Illinois

7-12 May 2023

Signal Processing in Acoustics: Paper 2pSP7

Validating seabed classification of spectrograms from multiple channels on a vertical line array

Ginger E. Lau
Department of Physics, Emory University, Atlanta, GA; ginger.e.lau@gmail.com

Tracianne B. Neilsen
Department of Physics and Astronomy, Brigham Young University, Provo, Utah, 84602; tbn@byu.edu

The information contained in spectrograms of ocean shipping noise is explored with a deep learning method for seabed classification. Due to the lack of measured and labeled ship noise data, the residual convolutional neural network ResNet-18 is trained with synthetic data samples generated by physics-based models. Previously, spectrograms from a single hydrophone channel, each scaled by their standard deviation, were utilized to infer an effective seabed class. This work, in contrast, utilizes multiple channels and a new scaling of the data more suited to deep learning. Models are trained with data from one, two, or four channels corresponding to hydrophones of different depths. The validation results show that for two-channel models, models with data from hydrophones at a lower average depth perform slightly better. Model performance is shown to depend neither on whether a specific channel is included nor the distance between channels in the model. These conclusions allow future multichannel studies to narrow down the number of models needed to obtain comprehensive results.

Published by the Acoustical Society of America

 https://doi.org/10.1121/2.0001971

1. INTRODUCTION

The propagation of sound waves in the ocean is influenced by the absorptive and reflective properties of the seabed. These properties are encoded in underwater sound signals. For example, the noise from a transiting cargo ship or tanker, also known as a ship-of-opportunity (SOO), can be used to determine seabed parameters since it contains information about the ocean environment. Some early examples of optimizations based on SOO noise include Battle *et al.*,[1] Heaney,[2] Koch and Knobles,[3] and Park *et al.*[4] Spectrograms of SOO signals can be used to obtain an estimate of effective seabed properties along the propagation path.[5] While traditional optimization algorithms have been commonly used, deep learning has recently been applied to obtain seabed type based on SOO data.

In the recent deep learning studies, convolutional neural networks (CNNs) were trained on synthetic data generated with a range-independent normal model sound propagation model, ORCA,[6] and the Wales-Heitmeyer source spectrum for transiting cargo ships.[7] While Van Komen *et al.*[8] used training data generated with only four representative seabed types, Forman *et al.*[9] demonstrated how a correlation metric of frequency-dependent transmission loss over range can be used to identify seabed classes to include in a larger seabed catalog. This approach led to the creation of a 34 seabed catalog used to generate synthetic training data in Escobar *et al.*[10] Data simulated with these 34 seabeds were used to train six different deep learning architectures based on convolutional neural networks. The consistency of the seabed predictions from ResNet-18 led to the selection of that architecture for this study.

Previous work with SOO spectrograms for seabed classification used only a single hydrophone in the middle of the water column.[8–10] The single hydrophone was part of a 16 element vertical line array (VLA). This study considers the impact of including spectrograms from multiple hydrophones as different input channels to the ResNet-18 residual neural network.[11] This proceedings paper report is part of a larger study that considers the advantages and limitations of using multiple elements on the vertical line array. After a brief description of the training data and neural network architecture, the validation results are shown. Comparisons of the accuracy when data from one, two, and four channels are included. The two and four-channel results are also analyzed with respect to average depth and separation distance of the channels selected. While not exhaustive, these results provide some insight into how the selection of hydrophone depths and spacing influence prediction accuracy. Guidelines are provided for balancing the advantages of including more hydrophones with the increased computational costs.

2. METHODS

A. SYNTHETIC TRAINING DATA

The training data is modeled after ship noise spectrograms measured during the Seabed Characterization Experiment 2017 (SBCEX2017). The experiment took place in the New England Mud Patch, in an area with a water depth of approximately 75 m. Vertical line arrays (VLAs) were deployed by Scripps Institute of Oceanography - Marine Physical Laboratories (SIO:MPL) with 16 hydrophones spaced 3.75 m apart. The hydrophones were numbered from #1 (at a depth of 15.75 m) to #16 (at a depth of 72 m). The depths of each of the 16 hydrophones are shown in Table 1. Prior studies used data from the center hydrophone, #8 (at a depth of 42 m). For the present study, spectrograms were generated for all even-numbered hydrophones using a wide variety of ship speeds and ranges. For each ship-receiver combination, the received levels for the synthetic dataset are simulated by combining the empirical source spectral levels for shipping noise developed by Wales and Heitmeyer[7] with transmission loss computed with ORCA, a range-dependent normal mode model[6] for 34 different seabeds. The seabed classes are detailed in Table III of Escobar *et al.*[10] For each seabed, the training data for the one- and two-channel models were generated using one measured sound speed profile, one randomly drawn source depth, 15 closest-point-of-approach values ran-

Table 1: Hydrophone channel number and corresponding receiver depths (from the surface, in 75 m of water) for the SBCEX2017 VLAs deployed by SIO:MPL.

Hydrophone	Receiver depth (m)
1	15.75
2	19.50
3	23.25
4	27.00
5	30.75
6	34.50
7	38.25
8	42.00
9	45.75
10	49.50
11	53.25
12	57.00
13	60.75
14	64.50
15	68.25
16	72.00

domly drawn between 0.5 km to 15 km, and nine ship speeds randomly drawn between 8 kt to 20 kt. The four-channel models only differed by using three sound speed profiles instead. Therefore, to train one- and two-channel models, a total of 4590 synthetic ship spectrograms were simulated, and for the four-channel models, 13,770 synthetic spectrograms were simulated.

B. SEABED CLASSIFIER

The deep learning model used in this study is the ResNet-18, a CNN architecture with 18 layers, skip connections, and approximately 11.2 million learnable parameters.[11] Since this is a classification task, the last layer is linear and connects to the classifier output via a softmax activation. Additionally, k-fold cross validation[12] is used with $k = 5$. The following hyperparameters are used during training. The network is optimized with AdamW, an optimizer based on Adam but with weight decay corrections.[13] The learning rate begins at 0.001 and then changes each epoch according to a cosine scheduler. An early stopping mechanism is included which stops training when the validation loss does not improve by at least 0.001 for a patience length of three epochs. During the five-fold cross validation, one or more of the folds might not train properly and terminate training prematurely via the early stopping mechanisms. In this situation, 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, with the second trained model accepted if both have low accuracy.

Due to the use of ship noise spectrograms, each data sample is mapped individually to remove the impact of different source levels. For each data sample (one ship with multiple channels), the spectral density levels (in dB re 1 μPa) are adjusted to have zero mean and a standard deviation of one across all channels. Values

Table 2: Channel combinations for the four-channel models used in this study. Channel numbers correspond to hydrophone depths from Table 1.

Channel Combinations			
2	4	6	8
4	6	8	10
2	6	10	14
4	6	10	14
6	8	10	12
2	8	12	16
6	8	12	14
6	8	12	16
6	10	12	16
6	10	14	16
4	12	14	16
8	10	14	16
8	12	14	16

more than three standard deviations from the mean are reassigned to a value of mean $\pm$ three standard deviations. The spectral density levels are then linearly mapped to the domain [0,1].

The data are augmented with zero-mean Gaussian white noise,[14] which is different each epoch to ensure that models trained on synthetic data are learning the important underlying signals rather than memorizing specific instances of noise. This common machine learning procedure helps avoid overfitting and improves generalizability. 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. 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 the validation process to better mimic testing on measured data, as the noise on a measured spectrogram would not change from epoch to epoch.

While previous work used only single channel input,[8,10] machine learning models can utilize input data with multiple channels. When representing the input data as a tensor, the depth of the tensor is the number of channels. The channels correspond to the different hydrophones, and each two-dimensional layer of the tensor is the spectrogram data from one hydrophone. The additional information from the multichannel approach is hypothesized to increase accuracy and/or efficiency of training the deep learning models for seabed classification. For this newly expanded data, new machine learning training hyperparameters must be chosen. In particular, the batch size used to train one- and two-channel models was 512, but the optimal batch size for four-channel models was determined to be 1024.

For this study, the even-numbered channels on the 16-channel VLA are used to create different testing scenarios. Only even-numbered channels are used to yield more varied information, while minimizing the size of each data sample and thus lowering the computation costs. A one-channel model is trained for each of the even-numbered channels. For two-channel models, every possible combination of two channels from the eight available channels (${}_8C_2 = 28$) are trained. For four-channel models, thirteen total models are trained. The specific channel combinations are shown in Table 2. The accuracies obtained for different channel combinations are presented in Sec. 3.

The neural network architecture and training algorithms are implemented in PyTorch version 1.5.1 and

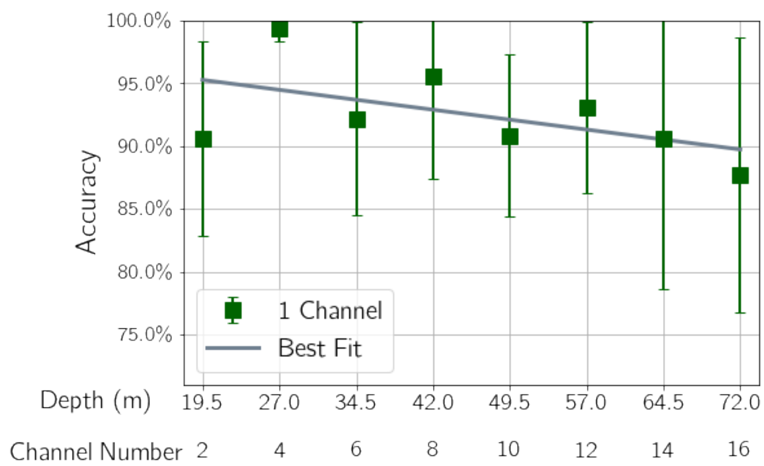

Figure 1: Seabed classification accuracy on validation samples for the one-channel models. The horizontal axis units are listed as both channel number and receiver depth in meters from the surface. Dots indicate the means, and bars indicate the standard deviation over five models (trained with k-fold cross validation). Error bars that extend past 100% are truncated to represent the maximum accuracy physically possible.

run in Python 3.6.9 on an NVIDIA Tesla T4 GPU. Under these conditions, the entire process from loading training dataset to training completion takes approximately 15 minutes for single-channel models, 18 minutes for two-channel models, and 40 minutes for four-channel models for the 34-seabed classification task. During this training process, validation is conducted using the "testing" portion of the dataset as determined by the k-fold splitting of the dataset. After training, the networks are then applied to the data samples from SBCEX2017 to test generalizability. Generalization results are being prepared for a peer-reviewed manuscript.[15]

3. RESULTS

Validation tests are performed to determine how accurately the model performs on synthetic data not used in training but drawn from the same statistical distribution as the training data. Five-fold cross validation is used, as described in Sec. 2.2. The mean accuracy and standard deviation over the resulting five trained models are calculated for each channel combination. The models are subsequently grouped by a given characteristic to investigate factors affecting model accuracy.

The seabed classifier output for each testing data sample (i.e., synthetic ship spectrogram) is a vector 34 elements long containing numbers that indicate the probability of a specific seabed being the correct answer. For the validation cases, the classifier output (for each data sample) is averaged over the five folds (for each model), and the seabed with the highest probability is designated as the selected seabed. In this validation testing, the predicted seabed can then be compared with the correct seabed to determine the accuracy of the model. Results are shown for models trained using both single and multiple channels with the 34 seabed catalog.

Validation metrics help test the hypothesis that utilizing additional hydrophone channels increases accuracy and efficiency of training deep learning models for seabed classification. Regarding training efficiency, the introduction of additional channels increases the amount of time required to train a model. Therefore, it should only be used if an increase in accuracy or generalization is offered. The validation results are described in Sec. 3.1.

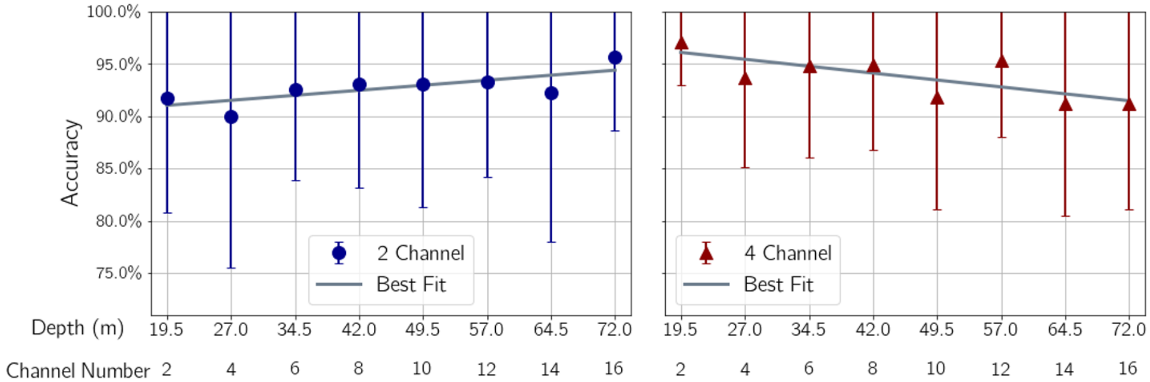

Figure 2: Seabed classification accuracy on the validation data for two-channel models (left) and four-channel models (right). On the left, all two-channel combinations of the even-numbered channels are represented for a total of 140 models (28 channel pairs $\times$ 5 folds). On the right, selected combinations of four channel models are represented by the channel numbers they contain. The horizontal axis units are listed as both channel number and receiver depth in meters from the surface. Dots indicate the means, and bars indicate the standard deviation over the models belonging to each group defined by the inclusion of a specific channel. Error bars that extend past 100% are truncated to represent the maximum accuracy physically possible.

A. VALIDATION

The validation accuracy of models trained with the 34 seabed catalog are analyzed to evaluate how best to use multiple channels from a VLA in deep learning. For each channel or channel combination tested, the seabed classification accuracy on the testing (validation) data is averaged over the five folds; the mean and standard deviation are plotted for comparison along with a line of ”Best Fit” for the mean values. For one-channel models, shown in Fig. 1, no significant difference is found in validation accuracy between models trained on different channels.

The two-channel models are evaluated in three ways: the inclusion of a specific channel, distance between the two channels, and average depth of the two channels used. All combinations of the even-numbered channels are represented for a total of 28 channel pairs, which yields 140 trained models (28 channel pairs $\times$ 5 folds each). Grouping trained models by the inclusion of a specific channel means that, for example, all seven pairs using Channel #2 (#2 and #4; #2 and #6; #2 and #8; #2 and #10; #2 and #12; #2 and #14; #2 and #16) belong to a group, all models using Channel #4 belong to a group, and so on. This analysis reveals no significant difference between performance of models when comparing which channel they contain, as shown in Fig. 2 (left). The mean and standard deviation are calculated for the 35 models (7 pairs $\times$ 5 folds) associated with each channel. The accuracy is fairly constant across the hydrophone depths.

The four-channel models are evaluated in a similar way, as shown in Fig. 2 (right). The mean and standard deviations of validation accuracy for the thirteen models (each with 5 folds) are plotted based on which channels they include. Similarly to the two-channel models, there appears to be no significant difference when analyzing multichannel model performance by the inclusion of data from a specific hydrophone. For the given dataset, the two-channel and four-channel models have similar validation accuracy. Therefore, the additional time required to train four-channel models is likely not beneficial to validation accuracy. However, generalization accuracy, or the ability for the model to predict on measured data, must be evaluated differently.

Another way to analyze validation accuracy is to consider the average depth of the channels included in the two- and four-channel models. The mean validation accuracy of the models are plotted in Fig. 3 as a function of the average depth of the channels used in each model. For the two-channel models, performance

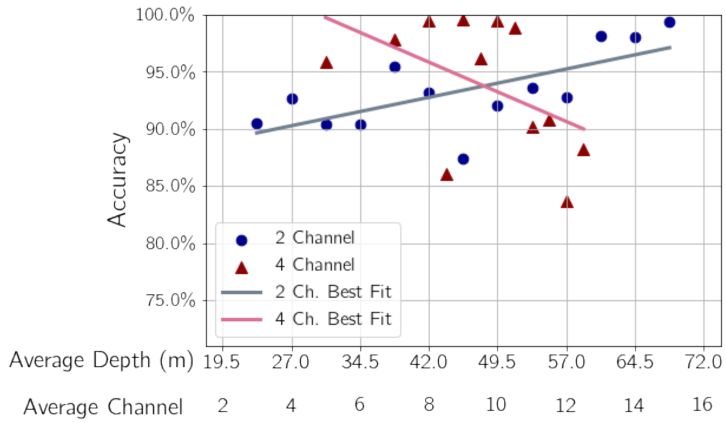

Figure 3: Mean seabed classification accuracy on the validation data for two- and four-channel models as a function of the average depth of the channels included. The horizontal axis units are listed as both average channel number and average receiver depth in meters from the surface.

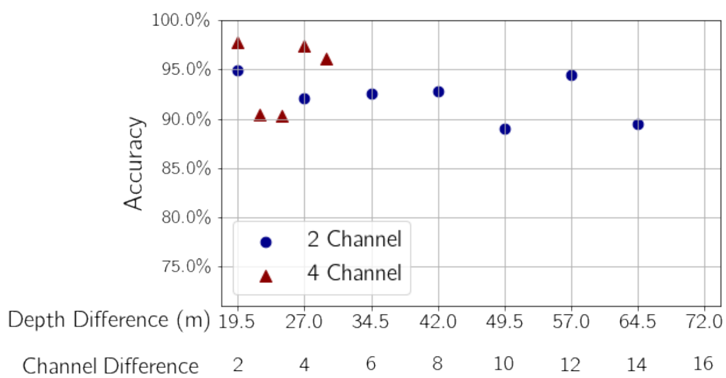

Figure 4: Mean seabed classification accuracy on the validation data for two- and four-channel models as a function of the average distance between the hydrophone depths included. The horizontal axis units are listed as both channel number difference and receiver depth difference in meters from the surface.

increases for models with data closer to the seafloor. For the four-channel models, however, maximum accuracy occurs for an average channel depth between 42.0 and 49.5 m from the surface, or between Channels #8 and #10. The final validation analysis investigates the impact of the distance between the selected channels on model performance. The distance between the channels does not have a significant effect on validation accuracy, as shown in Fig. 4.

4. CONCLUSION

In this study, a deep learning model for seabed classification was expanded to utilize data from hydrophones at multiple depths. This classification task used data from transiting SOOs to classify seabed type, given a set of 34 seabed classes. Validation on the holdout set in the k-fold cross-validation procedure offers suggestions for how many, and which, hydrophones should be used in a multichannel ResNet-18 model. The results indicate that for two-channel models, an average hydrophone depth closer to the seafloor performs slightly better. The results also indicate that neither the inclusion of a specific hydrophone nor the distance between hydrophones in a model has a significant impact on validation accuracy. However, to fully determine whether multichannel models perform better than single-channel models, the ability of the deep learning model to generalize to measured spectrograms must be analyzed. Furthermore, with the availability of data from multiple channels, an ensemble method of combining results from models trained on differing subsets can also be investigated. Both these tasks are done in Lau *et al.*[15] In conclusion, when data from multiple hydrophone depths is available to train a ResNet-18 for seabed classification, more consideration should be placed on where the average depth of the included channels lies, rather than the distance between channels in the model.

ACKNOWLEDGMENTS

This work has been supported by the Office of Naval Research grant number N00014-22-1-2402 ($298k). G. Lau also thanks the National Science Foundation for Research Experiences for Undergraduates grant number 2051129.

REFERENCES

[1] 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 Journal of Oceanic Engineering **28**(3), 454–467 (2003).

[2] K. D. Heaney, "Rapid geoacoustic characterization using a surface ship of opportunity," IEEE Journal of Oceanic Engineering **29**(1), 88–99 (2004).

[3] R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," The Journal of the Acoustical Society of America **117**(2), 626–637 (2005).

[4] C. Park, W. Seong, and P. Gerstoft, "Geoacoustic inversion in time domain using ship of opportunity noise recorded on a horizontal towed array," The Journal of the Acoustical Society of America **117**(4), 1933–1941 (2005).

[5] D. Tollefsen, S. E. Dosso, and D. P. Knobles, "Ship-of-opportunity noise inversions for geoacoustic profiles of a layered mud-sand seabed," IEEE Journal of Oceanic Engineering **45**(1), 189–200 (2020).

[6] 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).

[7] 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).

[8] 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," The Journal of the Acoustical Society of America **149**(2), 1198–1210 (2021).

[9] 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) (2021).

[10] 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," The Journal of the Acoustical Society of America **150**(2), 1434–1447 (2021).

[11] 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.

[12] M. Stone, "Cross-validatory choice and assessment of statistical predictions," Journal of the royal statistical society: Series B (Methodological) **36**(2), 111–133 (1974).

[13] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," arXiv (2017).

[14] 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.

[15] G. Lau, T. B. Neilsen, D. P. Knobles, and W. S. Hodgkiss, "Seabed classification with resnet-18 using multichannel ship noise spectrograms," IEEE Journal of Oceanic Engineering (Submitted May 2023).

[16] R. A. Koch and D. P. Knobles, "Geoacoustic inversion with ships as sources," The Journal of the Acoustical Society of America **117**(2), 626–637 (2005).