Neural Nets and Nautical Noises

University of Washington Master's of Science in Data Science
Capstone Final Report

Kevin Sweet
Emily Linebarger
Castle Leonard
Khirod Sahoo

Introduction and Executive Summary

This report is the culmination of a University of Washington Masters in Data Science capstone, sponsored by Dr. Shima Abadi of the Ocean Data Lab. The Ocean Data Lab has tasked our team with leveraging a rich hydrophone recording dataset for ship sound classification. Although machine learning for underwater sound classification is still in its early stages, the literature supports the use of deep architectures for this task. Our research focused on finding the most promising model architectures for ship noise classification, as well as providing recommendations on training dataset processing.

We found that both power-spectral density vectors and greyscale spectrograms were effective in training classifiers. Both data formats are information-rich, as shown by model performance on our random sample. We found that including color as a dimension in spectrograms did not improve performance, and due to the file size increase full-color spectrograms are less desirable for training production models. All three of our model architectures had similar performance results, and showed that maintaining an independence assumption is critical when designing a sample. Our model performance on a time-stratified sample is more believable than performance on our random sample, especially given our concerns with our class labels. Finally, we did an in-depth analysis of the original "vessel type" label from the Automatic Identification System, and found that it has flaws which necessitate a relabeling scheme.

Problem Statement

Sound waves are the only waves that carry through water at any significant distance, whereas light and electromagnetic dissipate rapidly and so can only provide information at short distances. For this reason, sound waves are the signal of choice to observe ships, weather, and life underwater. This capstone will focus on ship noise classification which has applications in monitoring vessel traffic as well as assessing ecological impact.

Underwater recording technology has become increasingly adept at providing high-quality signals, but due in part to the lack of well-labeled datasets for training models, the data processing software remains underdeveloped. We are positioned to leverage the Ocean Observatories Initiative hydrophone data in conjunction with vessel traffic data from the Automatic Identification System to produce a more comprehensive benchmarking dataset for ship sound classification.

Our goal is to apply a variety of modeling approaches for classifying ships by their signals picked up through hydrophone sensors and benchmark them against each other. We anticipate that our modeling findings may lead to discoveries that can improve the dataset and vice versa. Our final deliverable will be a comprehensive report on the modeling approaches and their performances and the materials needed to recreate our findings.

Literature review

The literature on audio classification with hydrophone data emphasizes the difficulty of this modeling task. Simpler models do not perform well in distinguishing patterns across the time and frequency domains of audio signals. Therefore, all of the papers we reviewed in preparation for this project emphasize deep model architectures. Some of the model architectures proposed were convolutional neural networks (Peng Li et al. 2022; Yongchun Miao and Yuri V. Zakharov 2021; Sheng Shen et al. 2020), recurrent neural networks (Nandi 2021), and hybrid model architectures (Wang et al. 2021). Several papers emphasized time-series modeling techniques (Sheng Shen et al. 2020; Yongchun Miao and Yuri V. Zakharov 2021). Other components of models included chirplet transforms (Yongchun Miao and Yuri V. Zakharov 2021), additional processing/masking of spectrogram images (Peng Li et al. 2022), using pre-trained models (Peng Li et al. 2022), long short-term memory models (Nandi 2021), and Hough transforms (Lee 2022).

Data Collection

For the ocean noise classification problem, we will be primarily using 2 sources of data.

Automatic Identification System (AIS)

The Automatic Identification System (AIS) is vessel traffic data collected by the U.S. Coast Guard. It is collected through an onboard navigation safety device that transmits and monitors the location and characteristics of vessels in U.S. and international waters in real time. This data for the region of interest is already available to us with necessary features required for the purpose of this project. The data covers the region around 3 low frequency hydrophones (Axial base, Central Caldera and Eastern Caldera) and the period of data is from 2015 to 2020.

The AIS data covering regions around the two other hydrophones (Oregon slope and Southern hydrate) can be downloaded from the marinecadaster.gov website by applying filters in their interactive UI. This data is free to use with limited information like ship id, location, time, vessel type, speed, length, beam, and draft, and the period of data available is from 2017 to 2022. The AIS data that is currently available to us has two million rows with 11 columns.

The part of the data which is already downloaded and prepared for use is available in multiple files of .csv format. The additional AIS data we need to process is available in a remote server managed by The Bureau of Ocean Energy Management (BOEM), the National Oceanic and Atmospheric Administration (NOAA), and the U.S. Coast Guard Navigation Center. The AccessAIS toolbox in marinecadastre.gov/ais can be used to download the AIS data in .csv format.

Hydrophone Data

The audio data is collected by hydrophones placed in the northwest pacific ocean region around the Oregon coast. This data is periodically stored in the Ocean Observatory Initiative (OOI) raw data server and can be accessed through OOIPY python library. This data consists of five 200Hz hydrophones that have streamed audio data from December 2015 to October 2022.

AIS Label Generation

The Ocean Data Lab used AIS data to find continuous 1-minute time instances where only one ship is within the inner radius (5km) and all other ships are outside the outer radius (20km) of a hydrophone. This is defined as an "isolated ship event". A known issue is that there may be small ships present that do not report to AIS. For additional information on the data preprocessing by the Ocean Data Lab, see appendix 2.

Data Sampling

From the dataset of isolated ship recordings provided by the Ocean Data Lab, we dropped all vessel type classes that had fewer than ten 1-minute records. We also dropped out "Other", "Unspecified", and "Research Vessel" labels, as well as any records that did not match with a label. From these recordings, we took two samples:

Random Sample

For the random sample, we sampled 100 random 1-minute records from all classes that had 100 records or greater. We took 60% of observations for training (1199 records) and 40% for test (798 records).

Time-Stratified Sample

For each class, we found a point in time that split up isolated events roughly 60/40 across train and test sets (1154 records in train, 795 records in test). This ensured that the same original ship recording did not end up in both train and test.

Modeling Results

Transformer Model

The basis for using a transformer model originates from the promising results demonstrated in *STM: Spectrogram Transformer Model for Underwater Acoustic Target Recognition* by Li. (Peng Li et al. 2022) Li compares multiple computer vision model architectures and feature extraction pairs to classify hydrophone spectrograms by vessel type on a similar dataset, ShipsEar.

This study seeks to replicate the high performance results on two model architectures tested by Li: the Spectrogram Transformer Model (STM) pre-trained on the ImageNet dataset and the STM without pre-training. The architecture used for each of these models is outlined in Table 1.

Parameter	STM + ImageNet	STM
Image size	384 x 384	240 x 240
Patch size	16	16
Number of classes	20	20
Dimension of linear layer after self-attention block	768	1024
Number of transformer blocks	12	12
Number of heads in self-attention block	12	12
Feed forward layer dimension	3072	2048
Image channels	3	1
Dropout rate	0.1	0.1
Embedding dropout rate	0	0.1
Pooling	Token	Token

Table 1: Transformer model architectures

Each model was trained using STANDARD_DS13_V2 Instances in Azure ML with 8 core CPUS and 56GB of RAM. The models were both trained for 10 epochs with a batch size of 24 images using both sampling methods. The criterion used was cross entropy loss with Adam's optimizer for stochastic gradient descent at a learning rate of 10^{-4}.

The number of channels used in STM + ImageNet differs from STM because the ImageNet dataset was trained on 3 channel images, therefore it requires all 3 channels of the greyscale images. Each of these three channels are duplicative for grayscale images, therefore only 1 channel is required to capture the signal contained in the image, so that was all that was needed in STM. The other differences between the models were attempts to assess how STM performance changed with differing complexity.

Model Results

Pre-training on ImageNet resulted in performance improvements on both the random sample and time stratified dataset. Overall accuracy for STM + ImageNet was 84% on the random sample while STM alone had an accuracy of 79%. The time stratified sample in both cases performed very poorly: 29% for

STM + ImageNet and 31% for STM. Classification reports and confusion matrices for both models and both sampling methods can be found in appendices 8 - 11.

An interesting observation of the time-stratified classification report is the relatively high performance observed for two classes: Chemical Tanker and Container Ship. These two classes each reported an F-1 score of greater than 0.80 while most of the other classes had an F-1 score of less than 0.40. Upon further inspection, it was found that these two classes each contained only one unique ship across many isolation events. This is evidence that a new model trained using the unique ship identifier (MMSI) as the class labels could perform well at predicting a specific ship by its sound. Further analysis on the errors made by these models can be found in the Error Analysis section below.

Multilayer Perceptron on power-spectral density data

At our sponsor's request, we wanted to try training on the time-collapsed version of the spectrogram - the power spectral density (PSD). To best understand the PSD's significance, let's compare it with other audio signal visualizations. For all our models the data preparation process started with 1-minute increment audio files and converted them into power-spectral density vectors and spectrogram data, which encode and visualize the audio signals. Figure 1, to the right, shows how a waveform compares to a grayscale spectrogram. Both show the magnitude of sound over time, but the spectrogram adds the important dimension of frequency. Meanwhile the PSD line does away with the time dimension, effectively shrinking the data, while still capturing the audio "signature" seen in the fluctuating magnitude of the sound across frequencies. Note the horizontal white streaks in the spectrogram line up with peaks in the PSD and the vertical line matches with a spike in amplitude in the waveform.

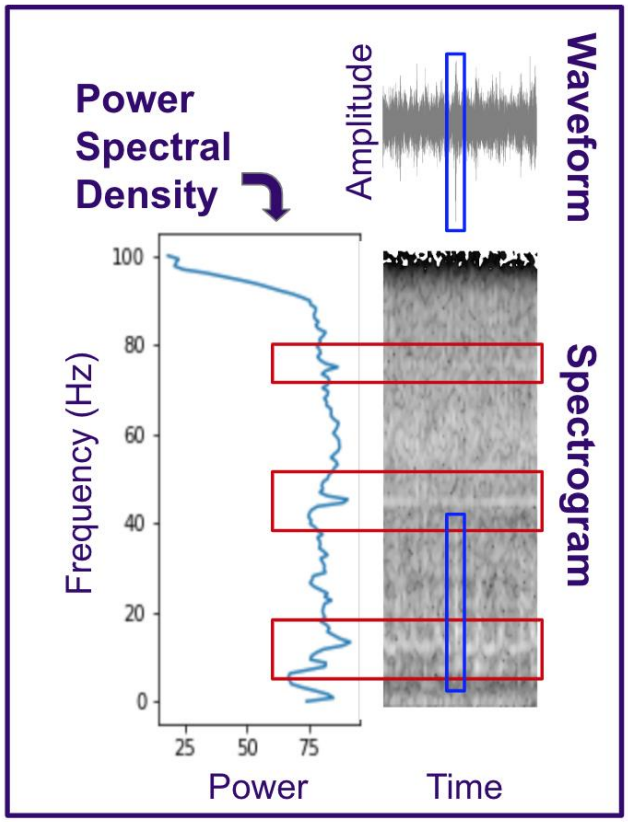


Figure 1: Comparing signal visualizations

At the onset we wanted to learn whether this consolidated version of the data contained sufficient information for class prediction without the cyclical nature of some sound signals captured or if the relative powers across frequency would carry enough distinctive information. There were advantages to using this consolidated format: firstly, the input data was a quite manageable size - just one vector of 129 powers at evenly spaced frequencies for each record. Additionally, without an explicit time dimension the model would theoretically be able to train and predict on recordings of variable length.

The model architecture was a modest sized multilayer perceptron (MLP) with three relu activation layers interleaved with linear layers for a total of approximately 30K trainable parameters. The layer thicknesses are detailed in the pytorch summary of the neural net, below. Note the input dimension is 128, not 129 as the last power value was dropped to benefit from greater training efficiently with power of two sized input data. The output dimension is 20, the number of classes.

```
NeuralNetwork(
  (flatten): Flatten(start_dim=1, end_dim=-1)
  (linear_relu_stack): Sequential(
    (0): Linear(in_features=128, out_features=128, bias=True)
    (1): ReLU()
    (2): Linear(in_features=128, out_features=64, bias=True)
    (3): ReLU()
    (4): Linear(in_features=64, out_features=64, bias=True)
    (5): ReLU()
    (6): Linear(in_features=64, out_features=20, bias=True)
  )
)
```

The above neural net architecture was trained on both the random sample dataset and the time-stratified dataset, each normalized using the mean and standard deviation per frequency over the training set, respectively producing two models. The loss function was a Cross Entropy loss, which is optimized to maximize the certainty of the correct classification.

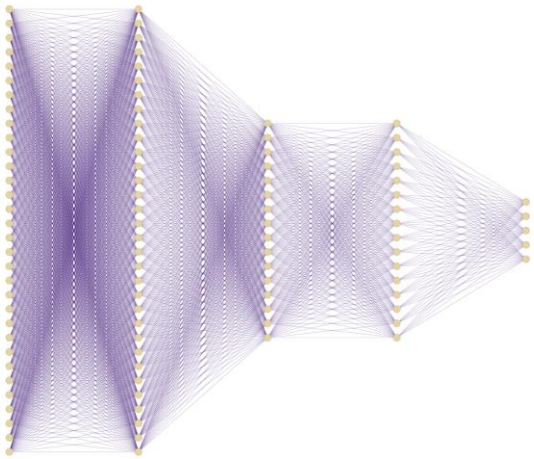


Figure 2: 1/4 scale visualization of MLP

Upon training the MLP, we found that despite being a smaller model trained on one dimensional data it performed on par with the larger transformer and CNN which both utilized richer two dimensional input data. Accuracy on the random sample set was 86% and accuracy for the time-stratified was 32%. One notable quirk was that for each class it performed best or near best compared to the other models except for the "No Ship" class, where it performed significantly worse (see Appendix 4: Model accuracy by class, location). It is known that not all ships report their locations to AIS and that sounds from ships outside the 10 kilometer boundary may still be faintly detected by the hydrophones, so we manually reviewed some of the incorrectly classified "No Ship" spectrograms. Some had no distinguishable ship signature (made more difficult for human eyes by the greyscale), however others displayed the clear pattern of a ship, such as this example from the Eastern Caldera hydrophone (Figure 3: instance_id EC_noship_20190531233000_20190531233100) that was classified as an "Offshore Supply Ship". The faint horizontal lines just above the vertical midpoint and near the bottom indicate a ship was indeed nearby. Upon further investigation, 8 of 20 "No Ship" labeled PSDs were classified as "Fishing" and 8

of 14 misclassified "Fishing" records were misclassified as "No Ship" in the random sample dataset. For the time-stratified sample, another pattern was discovered: 8 of the 21 misclassifications of the "No Ship" category were predicted to be an "Offshore Supply Ship", a class which had 58% accuracy by the MLP. These conflations may be the result of a combination of two factors: which ships are more likely to turn off their engine and remain stationary (as a fishing vessel naturally would) and which ships are less likely to report their location to AIS. This phenomenon warrants further investigation and experimentation.

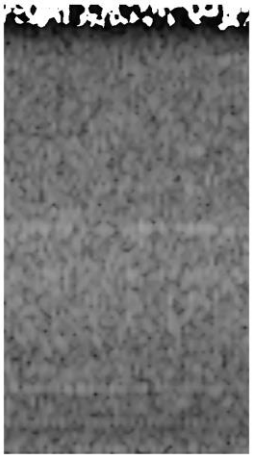


Figure 3: Misclassified spectrogram

Convolutional Neural Network

Several research papers have employed Convolutional Neural Networks (CNNs) for the classification of spectrogram images. Since spectrograms are essentially 2D images, CNNs prove to be a highly effective tool for analyzing spectrogram data. These networks have been successfully applied in various fields, including speech recognition, music classification, and environmental sound classification (Zhang et. al 2020, N.D. Roy et. al 2019, Mathur et. al 2018).

Inspired by this, we decided to test CNNs on raw spectrogram data. Our motivation for this stemmed from the fact that spectrogram images are processed, which may cause some information loss during the conversion. Additionally, there are numerous libraries available for processing spectrogram data to generate images, each of which processes the data differently. Therefore, we opted to use the raw spectrogram data, which is a 2D matrix of dimensions 129 x 60 (where 129 represents the frequency dimension, and 60 represents the time dimension). We believe that using raw data would help retain maximum information, as it has not undergone processing for visualization like spectrograms.

Additionally, using raw data also led to a significant decrease in data size from approximately 500,000 values to around 8,000. The large dataset size of spectrogram images made training on the data impractical. Nevertheless, the smaller dataset was closer to the pure signal data than the spectrogram visualizations generated using matplotlib.

As mentioned above, In our experiment, the input consisted of raw spectrogram data, represented as a 2D vector comprising 60 one-second data points across 129 frequencies. We normalized the data to ensure consistency across the different samples.

For the model architecture, we utilized a PyTorch network that incorporated a 4-layer convolutional neural network (CNN), along with batch normalization and dropout layers. The CNN layers were used to extract important features from the input spectrogram data, while the batch normalization and dropout layers helped regularize the model and prevent overfitting. The final layer of the network was a fully connected output layer that generated predictions based on the extracted features. Overall, the network had approximately 1.2 million parameters, which allowed it to learn complex patterns in the input data and make accurate predictions.

We ran several models on two different datasets and chose the architecture that gave the best accuracy scores on both the datasets.

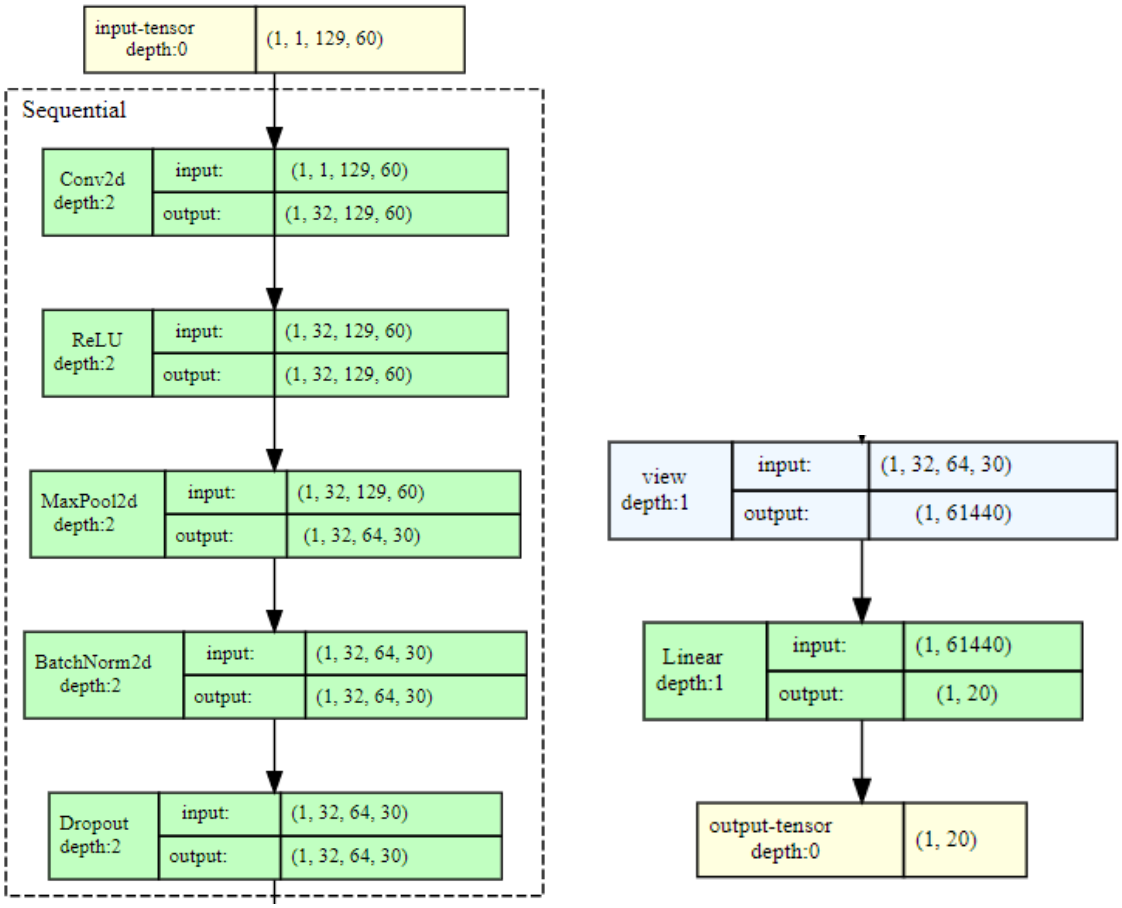


Figure 4: a scaled down visualization of the CNN architecture used (1 of 4 sequential blocks is shown)

The hyperparameters of the model are:

Number of epochs: num_epochs = 15
Optimization algorithm: opt_func = torch.optim.Adam
Learning rate: lr = 0.001
Dropout probability: 0.25 for all dropout layers
Number of input channels: 1 (2-D matrix)
Number of output classes: 20

Details of sequential blocks and layers used in the CNN architecture:

- Input layer: a single input channel (2-D matrix), with an input dimension of 129x 60
- Convolutional block 1: a convolutional layer with 32 filters, kernel size of 5 x 5, padding of 2, and ReLU activation function. This is followed by a 2 x 2 max pooling layer, batch normalization, and dropout with a probability of 0.25.

- Convolutional block 2: a convolutional layer with 64 filters, kernel size of 5 x 5, padding of 2, and ReLU activation function. This is followed by a 2 x 2 max pooling layer, batch normalization, and dropout with a probability of 0.25.
- Convolutional block 3: a convolutional layer with 128 filters, kernel size of 5 x 5, padding of 2, and ReLU activation function. This is followed by a 2 x 2 max pooling layer, batch normalization, and dropout with a probability of 0.25.
- Convolutional block 4: a convolutional layer with 256 filters, kernel size of 5 x 5, padding of 2, and ReLU activation function. This is followed by a 2 x 2 max pooling layer, batch normalization, and dropout with a probability of 0.25.
- Output layer: a fully connected linear layer with output dimension of 20

Error Analysis

The objective of the error analysis was to explore potential contributing factors towards the more difficult to predict classes in the random sample and the drop in performance from the random sample to the time-stratified sample.

In answering the first question - there were a few classes with particularly poor performance across all models - “Towing”, “Fishing”, and “Cargo, all ships of this type” all had accuracies below 70% for each of the three models and “No Ship” had remarkably low accuracy from the MLP and CNN, 46% and 59%, respectively, compared to the 70% of the transformer. Through visual exploration of the misclassified records’ spectrograms we identified clear ship signal among the “No Ship” class (which we expected and had chosen not to prematurely, manually exclude), but were not expecting that “Fishing” spectrograms would often lack a ship signal, presumably due to stopping for some purpose related to fishing. For “Towing” we theorize that a ship’s sound may be affected by whether it is towing another vessel and variations in the vessel. A maritime expert would be an excellent resource to better understand the unique behaviors of various ships, given their function.

To better understand the drop in accuracy from the random sample to the time-stratified one we conducted an analysis of the distribution of the training set versus the performance on the test set for both samples. The primary focus of the exploration was whether either the number of training examples per class, per hydrophone or the proportion of the distinct ships in the test set had been seen in the training set correlated with the accuracy. Our two theories were that perhaps the alternate sampling method led to test cases with too little support and that an individual ship would be more recognizable and would therefore be easier to predict. A fit line of accuracy per class, per hydrophone indicated a small correlation between the number of distinct ships in the test set seen in training. See Appendix 4: Model accuracy by class, location tables for full results.

Class Label Review

Our initial model results uncovered new questions about our data. One of them was "how well does the 'vessel type' label from AIS explain differences in ship noise"? If vessel type does not adequately distinguish ship noise, then we wanted to make a recommendation for a different labeling scheme.

The criteria we used for this new label were:

1. The label should minimize variance within classes
2. The label should maximize variance between classes
3. The label should make real-world sense (vessels of the same length, width, etc. classified together)

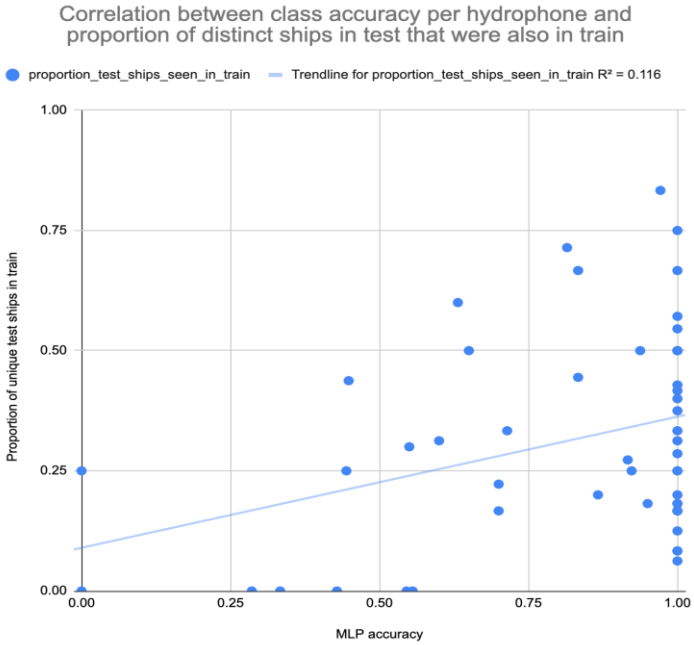


Figure 5: Faint correlation with accuracy

We started by reviewing the existing "vessel type" label from the Automatic Identification System data. First, we loaded all isolated ship incidents (the processed data that we took our samples from) and reviewed vessel type distributions. We found that the vessel types in the Axial Base and Oregon Slope clusters very rarely overlap, with the only common label being "Tug". This suggests an arbitrariness in this column, and the need for a relabeling strategy to build a model that would classify ships for any hydrophone. For each vessel type class, we reviewed characteristics like number of distinct ships, range of ship lengths, and range of average ship speeds. A table of these characteristics can be found in Appendix 5. We found that many classes only had a single vessel, and some categories like "Bulk Carrier" or "Cargo, all ships of this type" had over 100. We would likely get better model performance if vessels with similar characteristics were grouped into categories, and categories with only one ship were coalesced.

Figures 6 and 7 show the range of vessel lengths and speeds respectively for each "vessel type" label. Many categories only show a single line for the mean, which in this case means that there is only one ship in that category. These class labels might be more useful if ships with similar characteristics were clustered together.

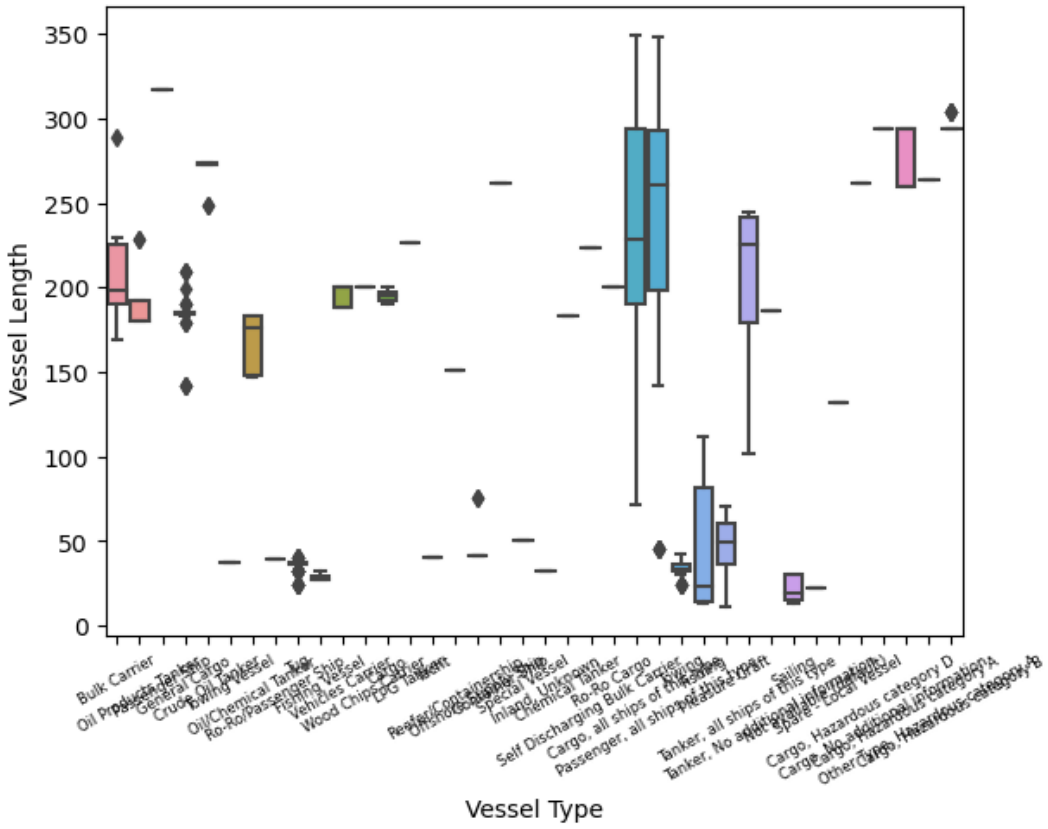


Figure 6: Vessel length boxplots by vessel type

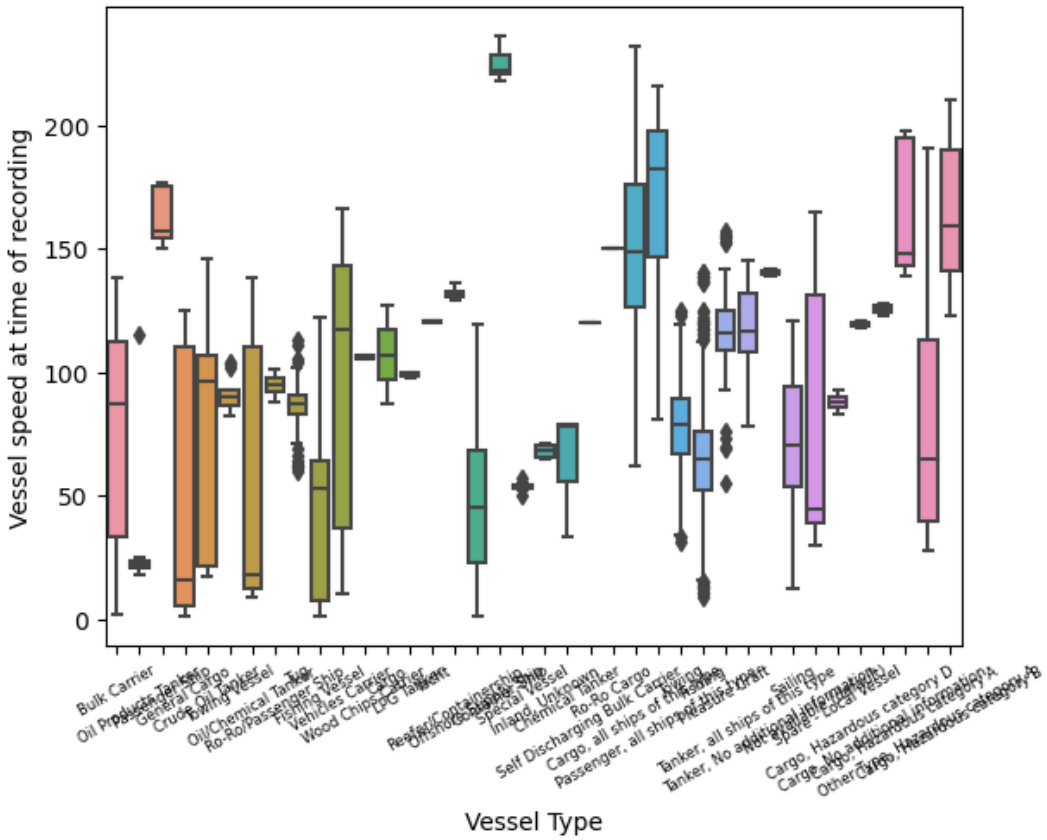


Figure 7: Vessel speed by vessel type

However, we didn't know how much ship length and speed would correlate with underwater audio signals. It's possible that these physical characteristics of the vessels would not explain differences in audio signals. We decided to run a K-means clustering algorithm on the power-spectral density vectors to get an understanding of the natural clusters in the audio signals.

For this analysis, we divided up all of the isolated ship recordings into 1-minute clips, and generated power-spectral density vectors for them (these were the same power-spectral density vectors used to take our random and time-stratified samples). There were 16,200 recordings over all five hydrophone locations. We read in the power spectral density vectors for each 1-minute recording, and normalized the values. Then, we used the elbow method to determine the right number of clusters[1]. From figure 8, we decided to group the recordings into six clusters.

[1] https://www.geeksforgeeks.org/elbow-method-for-optimal-value-of-k-in-kmeans/

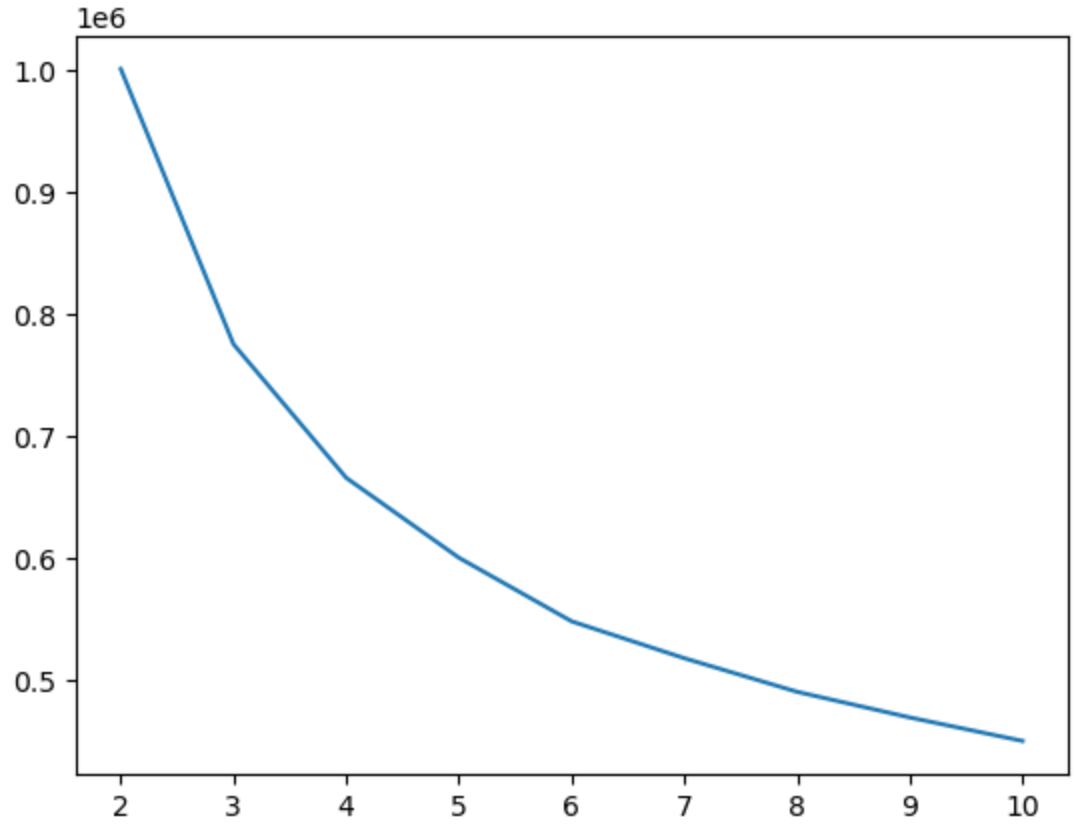


Figure 8: K-means SSE vs. Number of clusters

The K-means algorithm assigned a new class from 0-5 to each of the 1-minute PSD vectors. However, with 130-dimensional vectors, it was difficult to explain why the algorithm made the choices it did. Because explainability was important to us, we wanted to see if these audio clusters correlated with any real-world vessel attributes.

In figures 9 - 11, you can see boxplots of vessel length, current speed, and distance to hydrophone by the predicted audio cluster from the K-means algorithm.

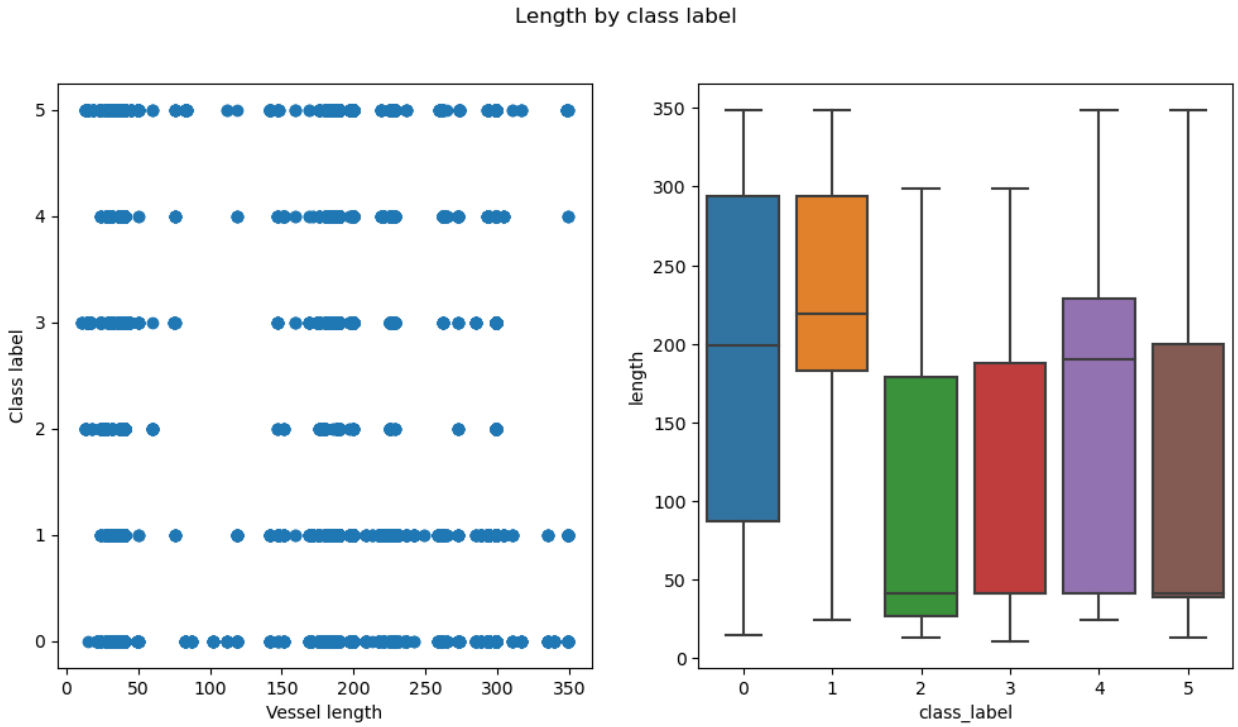


Figure 9: Vessel length by K-means class prediction

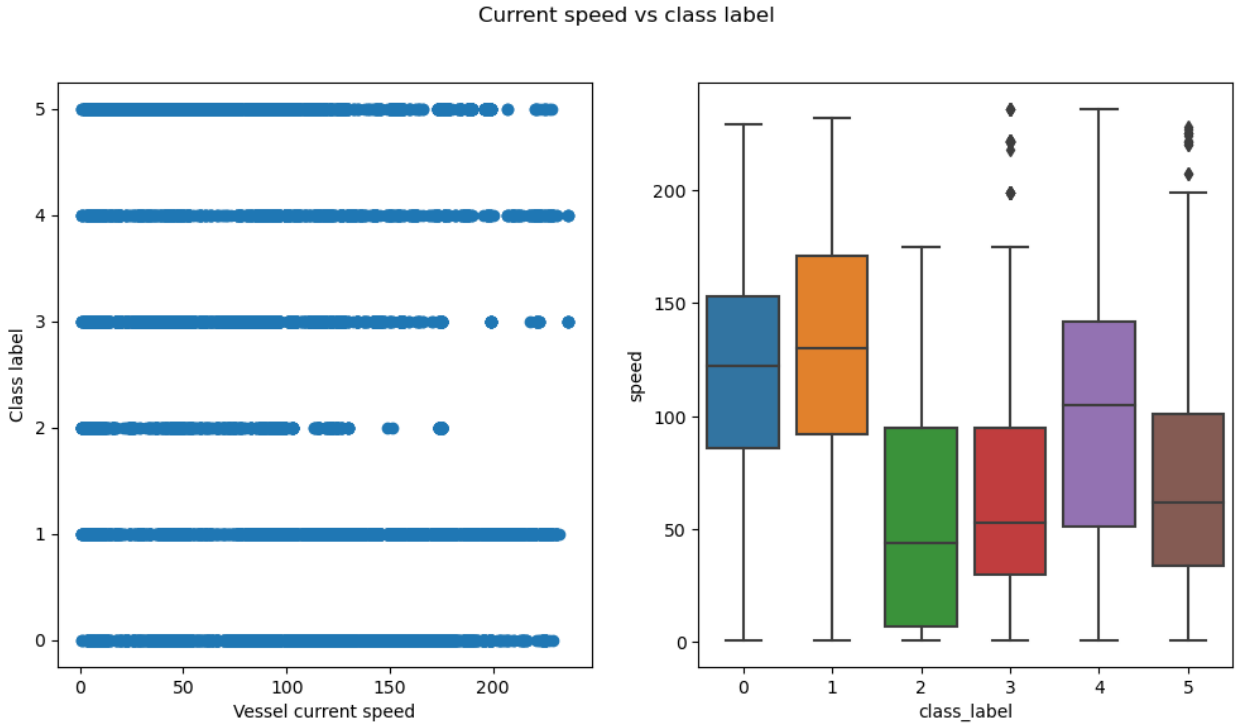


Figure 10: Current vessel speed by K-means class prediction

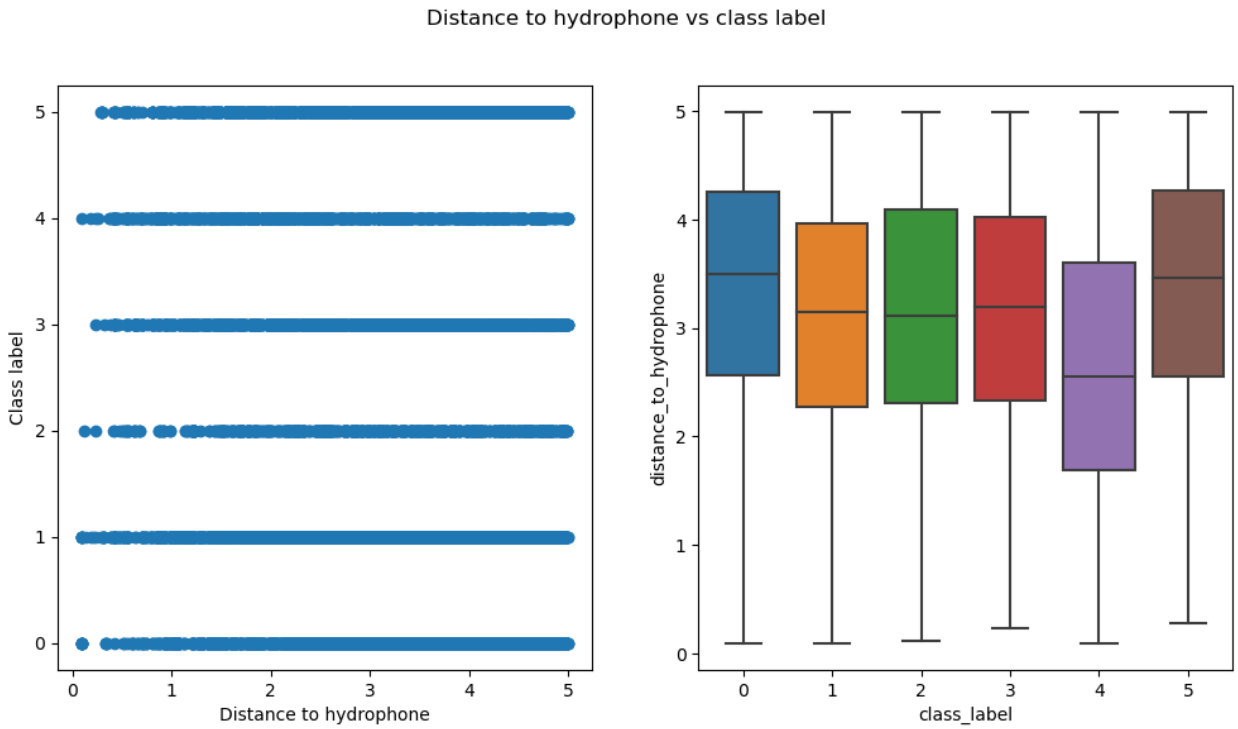


Figure 11: Distance to hydrophone by K-means class prediction

These graphs tell us that vessel length and speed at time of recording give some indication of audio signal, as seen by the marked difference in the mean for classes 0-1 versus 2-5. Classes 0 and 1 are faster, larger boats, while clusters 2 through 5 are more heterogeneous. Boats in these later clusters seem to be shorter in length, but have a range of speeds. Our theory here is that the clustering algorithm has lumped together small, fast boats like yachts and small, slow boats like fishing vessels.

An item for future work is to review the recordings in classes 2-5 and see if any additional real world attributes can separate these clusters. Our optimal end goal was to find the physical properties of the recording that explained the difference in ship noise, and that could form the basis for a relabeling algorithm. As it is, however, vessel speed and length alone would not make new satisfactory groupings.

Figure 12 shows a scatter plot of vessel speed and length, with color showing the predicted cluster from audio signal. This plot shows that speed and length do not neatly separate the clusters by color.

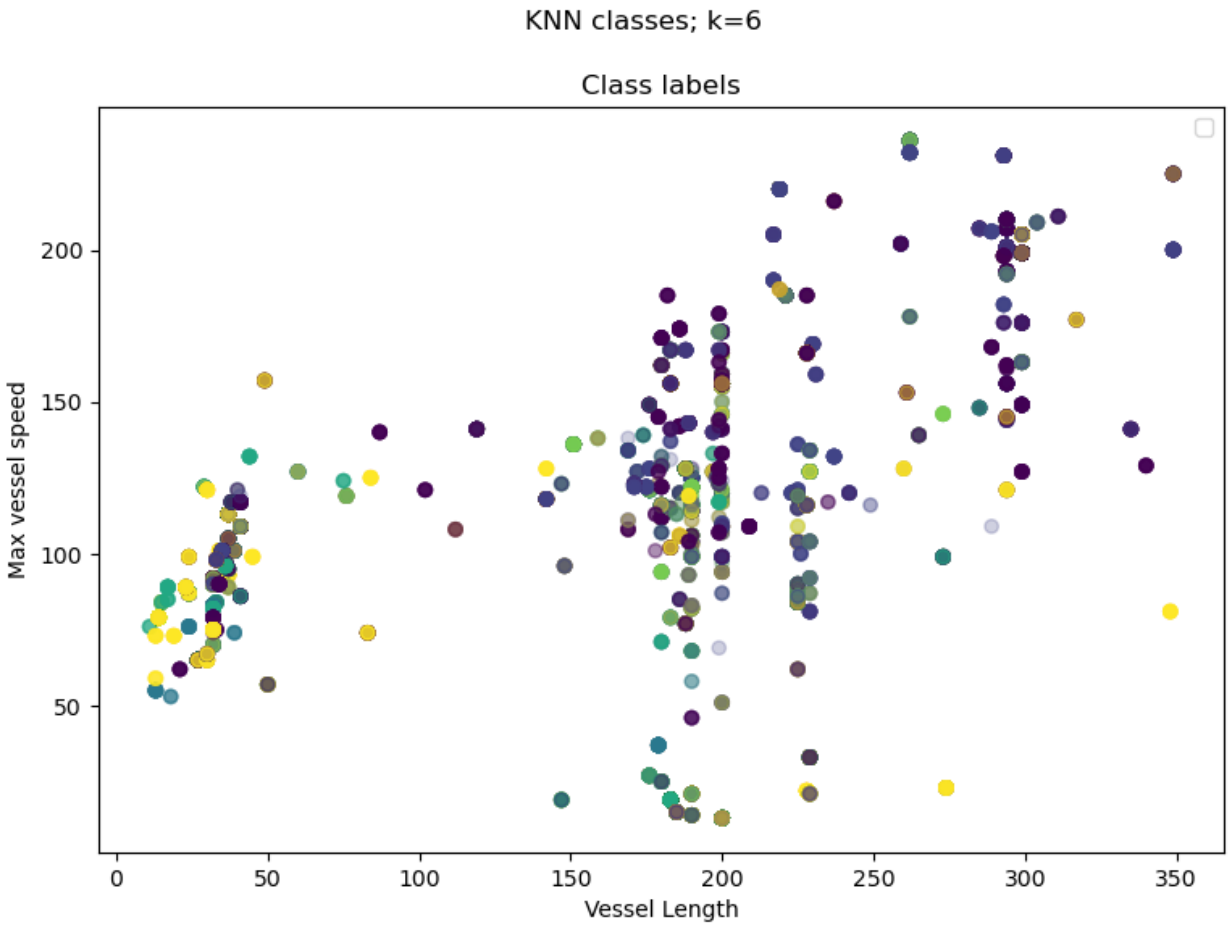


Figure 12: K-means clustering by vessel length and speed

We also made a few graphs showing the variance of PSD vectors within each vessel type class, as well as within each K-means predicted class. Figure 13 shows an example for the "Bulk Carrier" vessel type

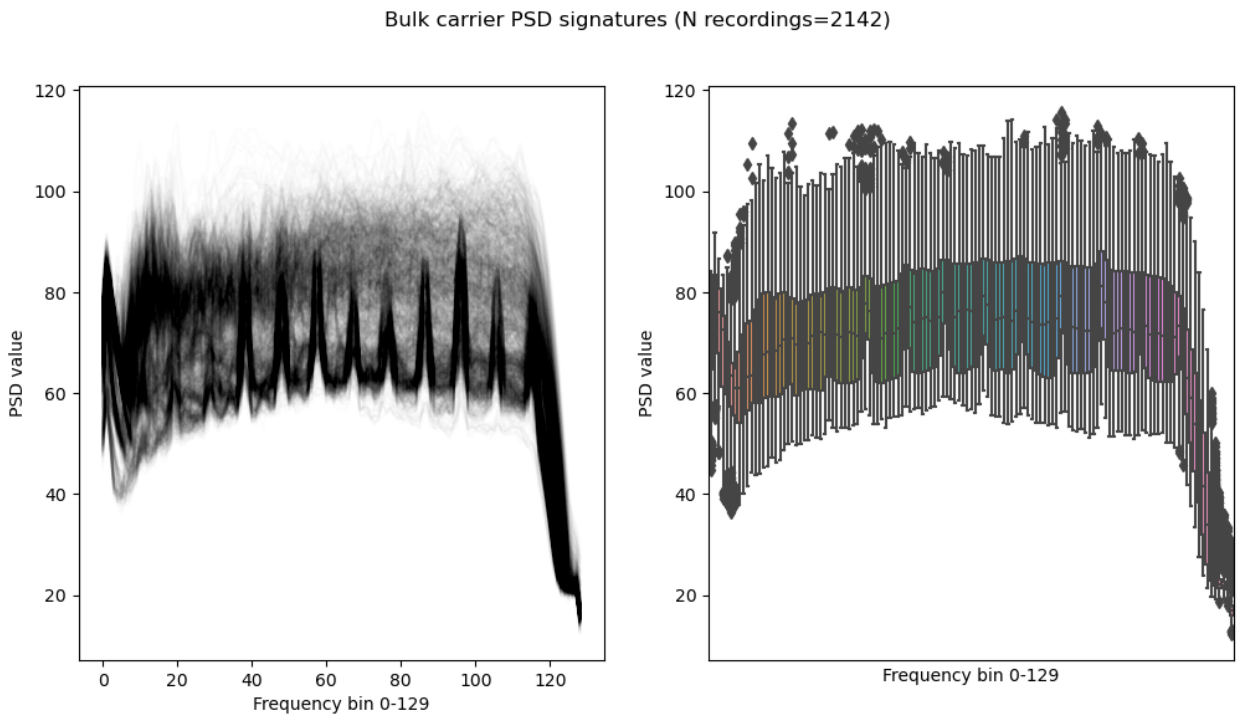


Figure 13: Bulk carrier PSD variance

class, and additional graphs of this type can be found in Appendix 6. Figures 14 - 19 show the variance between PSD vectors for each of the predicted K-means audio clusters.

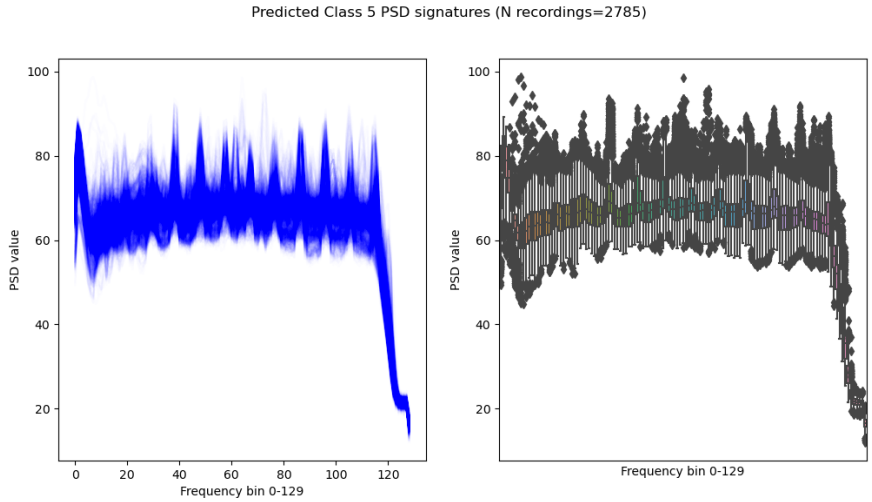


Figure 14: Class 5 PSD variance

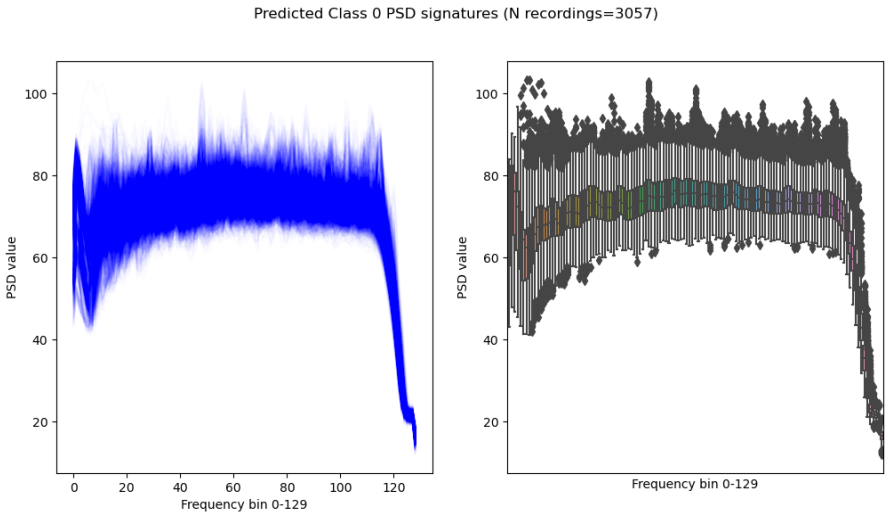


Figure 15: Class 0 PSD variance

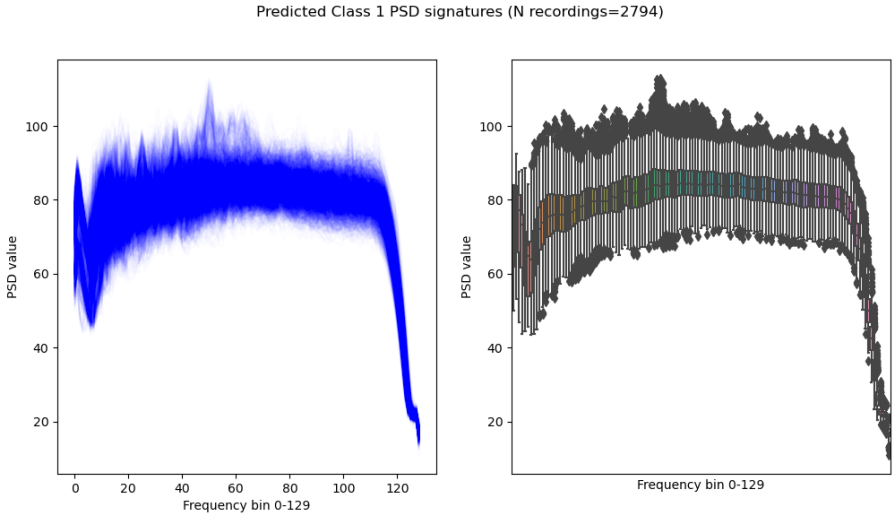


Figure 16: Class 1 PSD variance

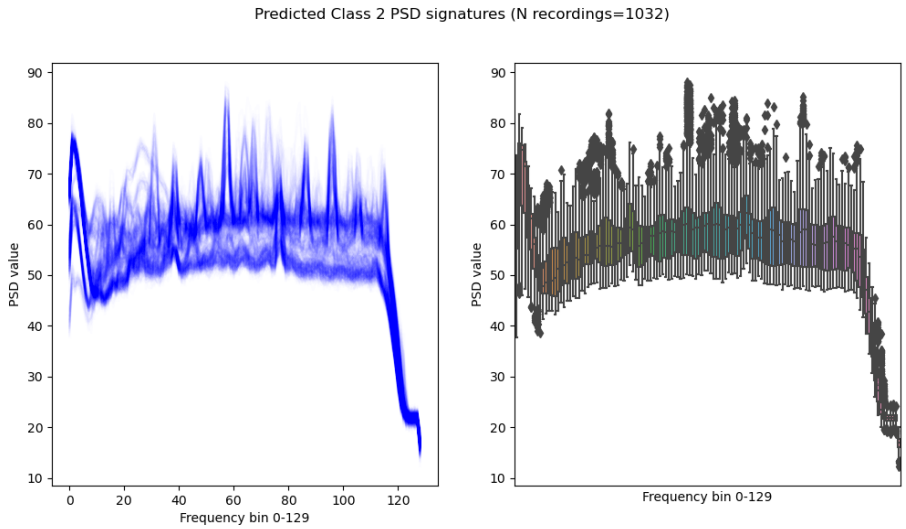


Figure 17: Class 2 PSD variance

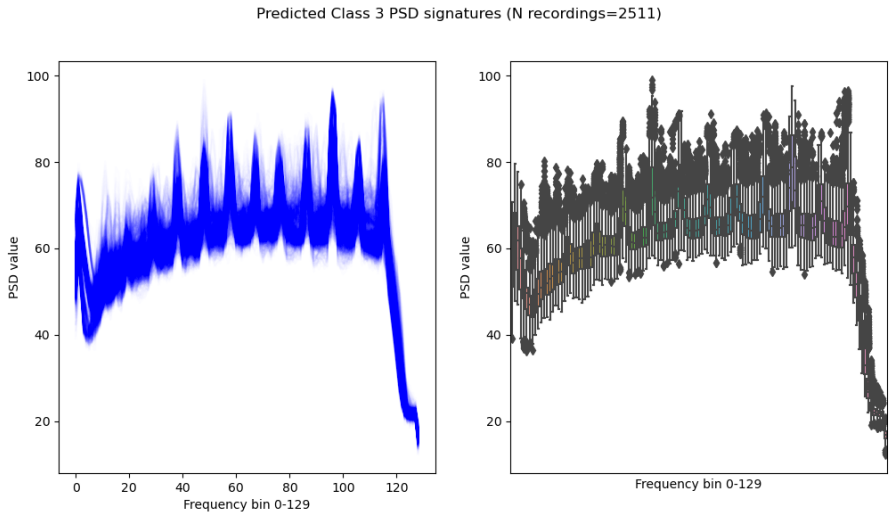


Figure 18: Class 3 PSD variance

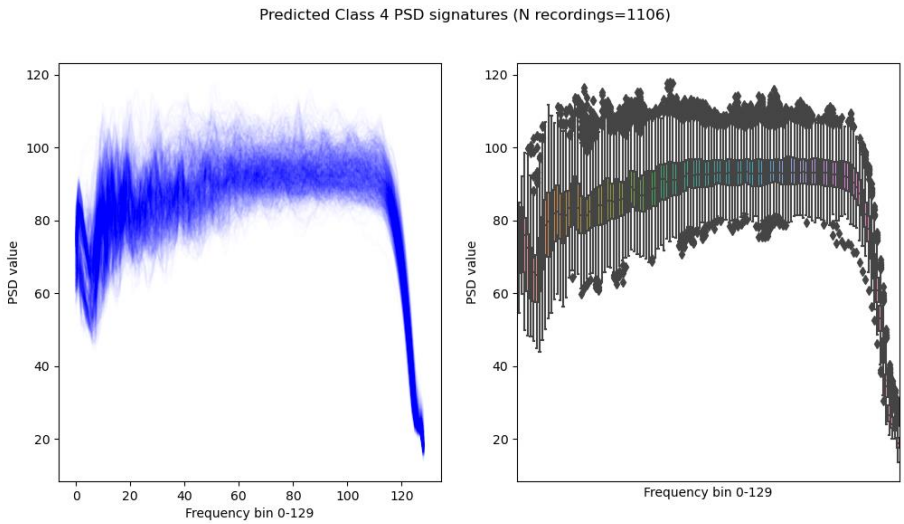


Figure 19: Class 4 PSD variance

Additional analyses on the K-means clustering results and ship characteristics can be found in Appendix 6.

Another option for reclassifying ships other than by their length, distance and speed is to consider grouping the various vessel types into fewer, broader categories. The potential benefit of doing this is that this allows the models to train on much larger samples of fewer ship categories. In this dataset there exists "ud_group" which is a logical combining of these vessel types, formed by the Ocean Data Lab team. During exploratory analysis of the differences between ud_group and vessel_type, it was discovered that the average spectrogram of ud_group was vastly different from the average vessel type spectrograms that composed that group. The result of this is that much of the variance between classes is lost when using ud_group. As shown below in key findings, it may be beneficial to reevaluate using a different ud group once better class labels are found.

Examples of the comparison of the average spectrogram image for ud group and vessel type can be found in Appendix 7.

Key Findings

The hydrophone data from the Ocean Data Lab is information rich, as shown by model performance on the random sample.
All three of our models were able to classify ship noise from spectrogram and PSD inputs. If the class label issues from the original AIS data were addressed, we are confident that any of these model architectures would be able to classify ship noise with a reasonable level of accuracy. We did not find major differences in performance between the three model architectures we tried, so model selection might be based on other means, like model size or time needed to train in production.

Grayscale spectrogram images and power-spectral density vectors were more effective for training a model than full-color spectrograms.
We found that while color is helpful for a human when reviewing spectrograms, breaking down a continuous variable like decibel level into three color channels does not improve model performance. The greyscale spectrograms performed as well or better than color spectrograms. Also, we found very little difference between the models that used spectrogram inputs versus power spectral density vectors, which tells us that both of these formats convey information well. Power spectral density vectors might be preferred, because they are significantly smaller than images and therefore easier to train on.

Using a time-stratified sampling strategy is one way to ensure independence between train and test sets.
An independence assumption is key to evaluating model performance. In our random sample, we realized that we had made an error of allowing instances from the same ship recording to be seen in both train and test. We corrected this by using a time-stratified sampling strategy, and model performance on this sample was closer to what we expected (30% accuracy), given our reservations about the original class labels from AIS.

Class definitions from the Automatic Identification System are imperfect.
Through our data analysis, we found that there are some critical issues with the “vessel type” label in the original Automatic Identification System data. To start, the labels are arbitrary, and similar types of vessels are given different labels in different geographic locations. Only one label, “Tug”, was seen in both the Axial Base cluster and Oregon Slope cluster. The Ocean Data Lab relabeled the recordings into new clusters, with a variable called "ud_group". From our analysis on the average spectrograms, we found that ud_group resulted in a lower variance between classes and therefore a worse model performance than "vessel_type". In addition, by running a clustering algorithm on power-spectral density vectors, we found that vessel function is not a good predictor of audio signal. However, we did not find an easy answer on how to reclassify these vessels. Additional research is needed to determine what physical characteristics of vessels correlate with ship noise.

Future Work

This work could be expanded in many ways. These are some that we feel would be most impactful.

Use an independent sampling strategy that keeps all records from a single isolation event within either train or test, but also accounts for the effects of seasonality.
The goal here is to control for external variables that might influence the overall noise in the ocean, namely water temperature, storms, and noise from ocean life. We did not look at whether this was possible with the five years of hydrophone data we had access to. There may be other ways to control for these external variables, such as merging on variables like ocean temperature and weather measurements.

Explore alternatives to the ship function class label of vessel type.

Some potential options are:

- MMSI (unique ship identifier in AIS)
- Ship build / model from Spire Vessel API called “Individual type”. See “Understanding how ship types are assigned” section at https://documentation.spire.com/vessels-api/
- Vessel and voyage attributes, such as length, engine type, cargo weight

Continue to search for physical features that explain the variations within and between PSD clusters.

Some things to explore would be distance to the hydrophone, weight, or environmental factors like storms. One metric to quantify the goodness of clusters is the silhouette score, which combines mean intra-cluster distance and extra- cluster distance over all samples to measure the separation of the clusters. Clusters with clear “silhouettes” distinguishing them from other clusters may indicate true, distinct groups within the population and aid in identifying better labels.

About the Authors

Castle Leonard is a Data Scientist at Microsoft as well as a current student in the University of Washington's Masters in Data Science program. Within the MSDS Capstone, Castle is excited to contribute in the areas of stakeholder communication, solution design, and consideration toward social responsibility.

Emily Nelson is a Software Engineer at ExtraHop networks and a current student in the University of Washington's Master's of Science in Data Science program. Emily will serve as the point of contact with our capstone sponsor and will contribute to data engineering, report writing, and modeling.

Kevin Sweet is the Co-Founder of the Fin-tech startup CommandFi and part-time UW MSDS student. He has a technical background in product management, software engineering, and recommender systems development. His role on this project will be to assess the performance and tradeoffs of multiple classifiers and report on his findings.

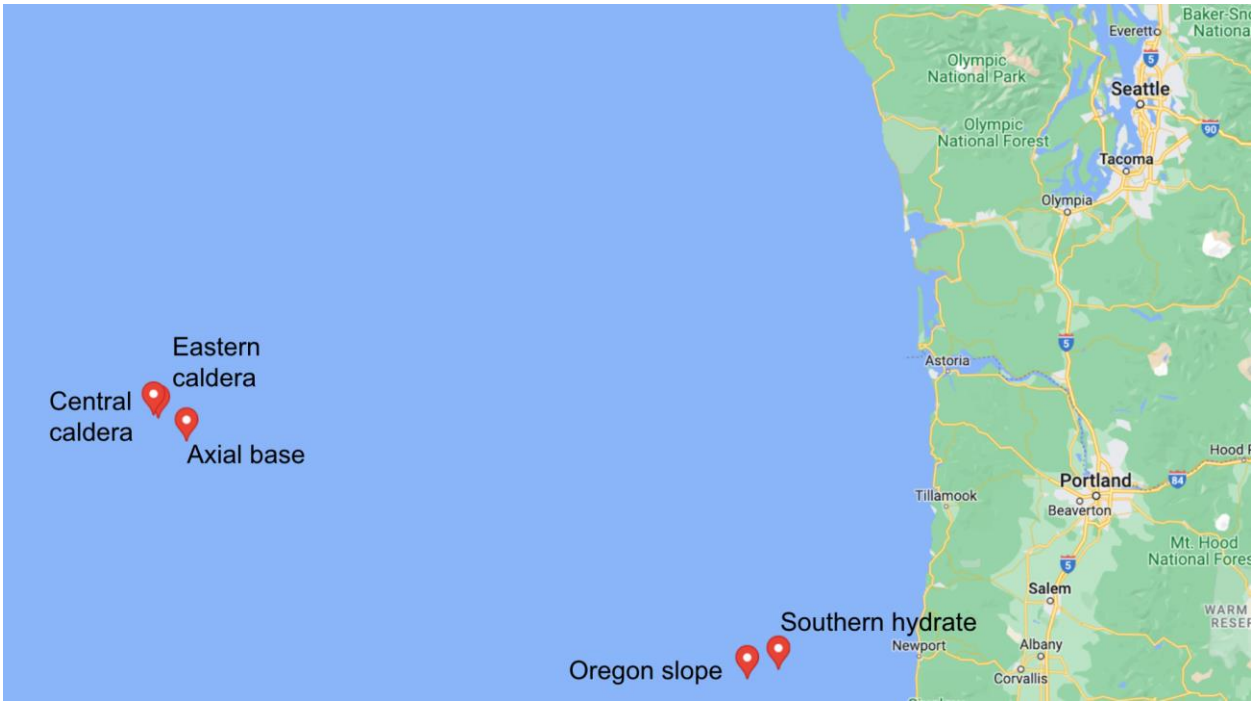
Khirod is a full-time student in the MSDS program. He has 3+ years of professional experience in data analysis and model development. He has proficiency in algorithm development, Machine Learning, and predictive modeling. He has a good domain understanding of ocean data and will assist in the data preparation and pre-processing part of the capstone project along with ML model evaluation.

References

1. "AIS Frequently Asked Questions | Navigation Center." n.d. Accessed December 13, 2022. https://www.navcen.uscg.gov/ais-frequently-asked-questions.
2. "Decision Learning Algorithm for Acoustic Vessel Classification." 2012. *Homeland Security Affairs* (blog). March 26, 2012. https://www.hsaj.org/articles/211.
3. Hanly, Steve. "Why the Power Spectral Density (PSD) Is the Gold Standard of Vibration Analysis." Accessed March 14,2023 .https://blog.endaq.com/why-the-power-spectral-density-psd-is-the-gold-standard-of-vibration-analysis.
4. Lee, Socret. 2022. "Lines Detection with Hough Transform." Medium. January 30, 2022. https://towardsdatascience.com/lines-detection-with-hough-transform-84020b3b1549.
5. Nandi, Papia. 2021. "Recurrent Neural Nets for Audio Classification." Medium. March 1, 2021. https://towardsdatascience.com/recurrent-neural-nets-for-audio-classification-81cb62327990.
6. Peng Li, Ji Wu, Yongxian Wang, Qiang Lan, and Wenbin Xiao. 2022. "STM: Spectrogram Transformer Model for Underwater Acoustic Target Recognition." *Journal of Marine Science and Engineering* 10 (1428).
7. Pollara, Alexander, Alexander Sutin, and Hady Salloum. 2016. "Improvement of the Detection of Envelope Modulation on Noise (DEMON) and Its Application to Small Boats." In *OCEANS 2016 MTS/IEEE Monterey*, 1–10. https://doi.org/10.1109/OCEANS.2016.7761197.
8. Sheng Shen, Honghui Yang, Xiohui Yao, and Junhao Li. 2020. "Ship Type Classification by Convolutional Neural Networks with Auditory-Like Mechanisms." *Sensors* 20 (253).
9. Wang, Biao, Chengxi Wu, Yunan Zhu, Mingliang Zhang, Hanqiong Li, and Wei Zhang. 2021. "Ship Radiated Noise Recognition Technology Based on ML-DS Decision Fusion." *Computational Intelligence and Neuroscience* 2021. https://doi.org/10.1155/2021/8901565.
10. Yongchun Miao and Yuri V. Zakharov. 2021. "Underwater Acoustic Signal Classification Based on Sparse Time-Frequency Representation and Deep Learning." *IEEE Journal of Oceanic Engineering* 46 (3).
11. Li, P.; Wu, J.; Wang, Y.; Lan, Q.; Xiao, W. STM: Spectrogram Transformer Model for Underwater Acoustic Target Recognition. *J. Mar. Sci. Eng.* 2022, *10*, 1428. https://doi.org/10.3390/jmse10101428
12. M. A. Naderi and H. Mahdavi-Nasab, "Analysis and classification of EEG signals using spectral analysis and recurrent neural networks," *2010 17th Iranian Conference of Biomedical Engineering (ICBME)*, 2010, pp. 1-4, doi: 10.1109/ICBME.2010.5704931.
13. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems* (p./pp. 5998--6008), .
14. "A Deep Learning Approach for Ship Noise Classification" by Dongyong Zhang, Xinyu Li, Jianzhuang Liu, and Xiaochuan Pan (2020). This paper presents a method for ship noise classification using a CNN with multiple convolutional and pooling layers..
15. "Automated Classification of Shipping Noise Spectrograms Using Convolutional Neural Networks" by N. D. Roy, D. D. Joshi, and R. Ramachandran (2019). This paper proposes a CNN-based method for automated classification of shipping noise spectrograms.
16. "A CNN Based Ship Noise Detection and Classification System" by Gaurav Mathur, Piyush Kuchhal, and Manoj Kumar Sharma (2018). This paper presents a CNN-based system for ship noise detection and classification using spectrograms.
17. "NN SVG." Accessed March 14, 2023. http://alexlenail.me/NN-SVG/index.html.
18. scikit-learn. "Sklearn.Metrics.Silhouette_score." Accessed March 15, 2023. https://scikit-

learn/stable/modules/generated/sklearn.metrics.silhouette_score.html.

Appendix 1: Map of hydrophone locations



Appendix 2: Pre-processing of hydrophone data into spectrogram images

This appendix describes some of this data pre-processing for our research, which was done by the Ocean Data Lab.

The hydrophone data consists of five 200Hz hydrophones that have streamed audio data from December 2015 to October 2022. For the hydrophone data, the Ocean Data Lab used the OOIPY library to extract acoustic data in a specified time period for a ship. Figure 1 shows a snippet of the code to extract the spectrogram images using a start and an end time, and Figure 2 shows the resulting spectrogram.

```
In [10]: start_time=datetime.datetime(2020,10,1,12,0,0)
         end_time=datetime.datetime(2020,10,1,12,3,0)

In [4]: node='HYSB1'
        data_trace = ooipy.get_acoustic_data_LF(start_time, end_time, node, verbose=True, zero_mean=True)
        Downloading mseed file...

In [6]: spec = data_trace.compute_spectrogram(L = 256,avg_time=10, overlap=0.9)

In [7]: spec.values.shape
Out[7]: (18, 129)

In [8]: # spec.time

In [9]: ooipy.plot(spec,fmin=0,fmax=90,vmax=110) #xlabel changed from 30 to 10
        plt.show()
```

Figure 1: Code to isolate audio signal by time

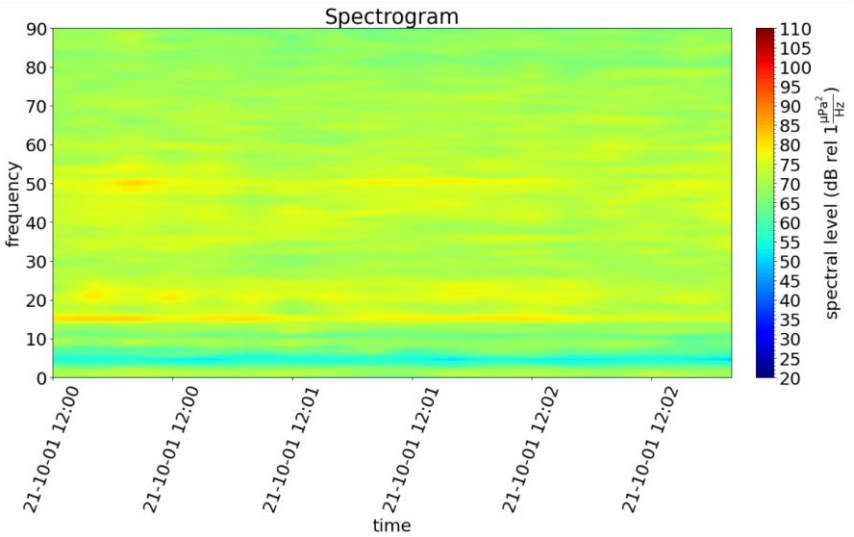


Figure 2: Spectrogram produced by code in figure 1

The AIS data has around 75 classes in the raw data, which is not ideal for machine learning modeling. Therefore, the Ocean Data Lab clustered these vessel types together to simplify the modeling task. To decide on the new classes, two major analyses were done:

1. Checking the distributions of ships attributes like length, speed, and distance from hydrophones
2. Creating Spectral Probability Density Function (SPDF)[2] charts to understand the signatures of the ship groups. This helped determine how similar and different ship types were from each other.

Figure 3 shows the distribution of ships with "passenger" and "fishing type" groups around the Axial Base hydrophone:

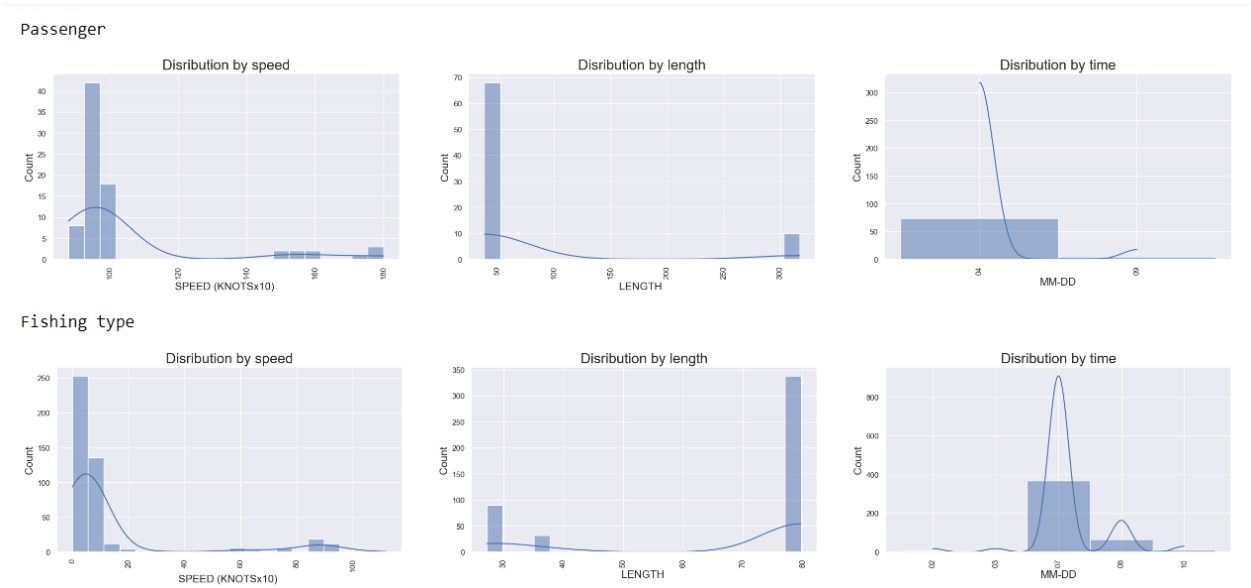


Figure 3: Distribution of ships around Axial Base

The vessel types "fishing vessel", "towing vessel", and "tug l" are grouped together into the new category "fishing type". We can see that most of the ships in the fishing type class have length around 80m and speed between 0-2 knots. We may need to revisit this decision as modeling progresses, because a fishing vessel or towing vessel may vary in length depending upon its size. The speed of a ship is also variable. The speed of a fishing vessel may increase and match the speed of a passenger ship if the ship is returning from the ocean.

Therefore, spectral density probability function charts will give us a better understanding of the groups which take into account the spectral level of the ship's noises at different frequencies.

[2] *SPDF is a probability function of spectral densities.*

Figure 12 shows the SPDF for passenger-type vessels. The passenger group has only 2 vessel types - Passenger ship and Ro/Ro ship. Hence we don't see many horizontal lines separating the audio signals radiated by the two vessel types. The two vessel types are very similar in terms of noise radiated.

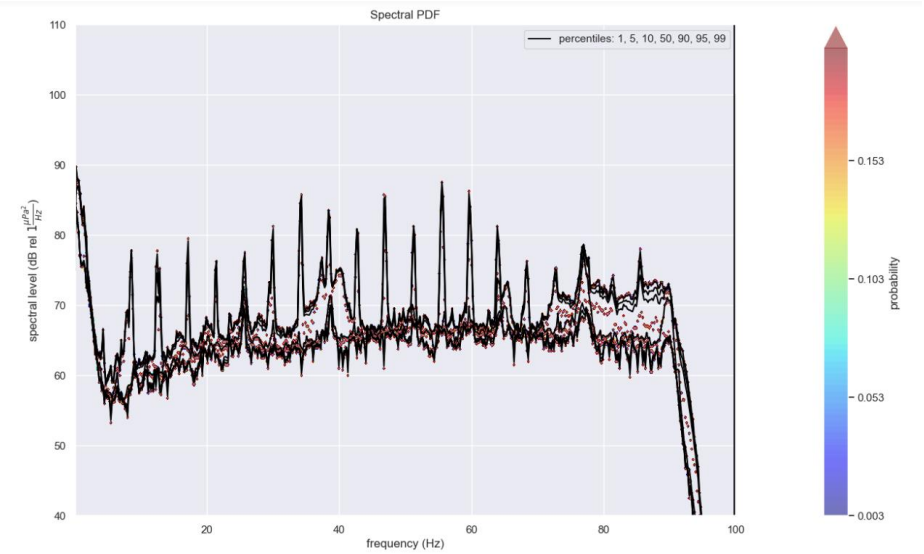


Figure 4: SPDF for passenger vessels

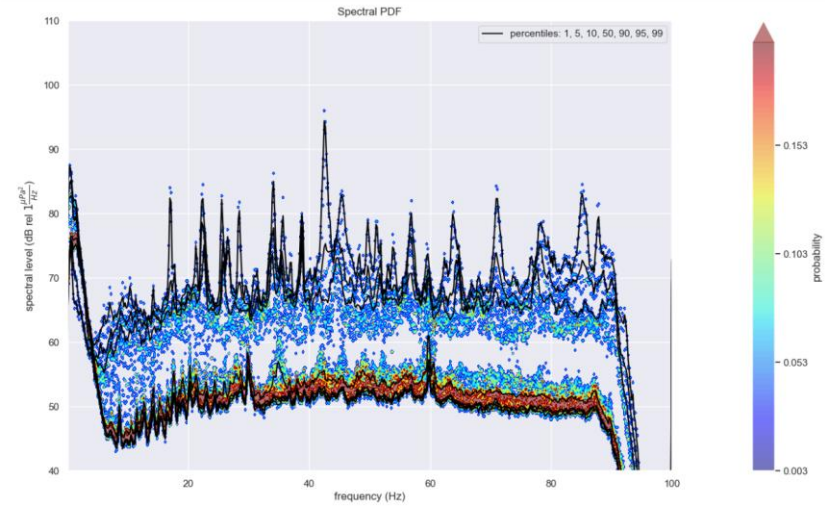


Figure 5: Fishing type vessels

"Fishing type" has 3 different vessel types. as shown in Figure 13. The horizontal lines separating the audio signals could be an indicator of the presence of the different vessel types, but it is difficult to conclude this from the SPDF chart alone. The separation could also be due to a ship traveling at different speeds and radiating different levels of noise, or the distance of the ships from the hydrophones. Therefore, the distribution of ship attributes along with the SPDF charts is needed to distinguish vessels further.

However, since the two charts shown in figures 12 and 13 are quite different from each other, two different classes for “passenger” and “fishing type” make sense. A similar analysis is performed for all the vessel types, and they are clustered into 5-6 categories. These classes and the sub-groups within each class may change in the future if needed.

Appendix 3: Data Exploration from Research Proposal

Spectrogram Images from Hydrophone Signals

Spectrograms are an image representation of audio frequency and amplitude over time. Due to the sound traveling underwater, the amplitude of the wave is expressed in decibels relative to the underwater reference pressure of $1\frac{\mu Pa^2}{Hz}$. This relative amplitude is referred to as spectral level. For the purpose of this analysis, the terms "amplitude" and "spectral level" are used interchangeably.

Figure 1A is an example of a hydrophone recording in which no ships are present near the hydrophone. This is evident by the lack of observations of a "straight line" of abnormally high spectral levels in the 30-60Hz range. Figure 1B is an example of when a ship is present near the hydrophone, evidenced by distinct horizontal lines of relatively high spectral level in the range of 30-60Hz.

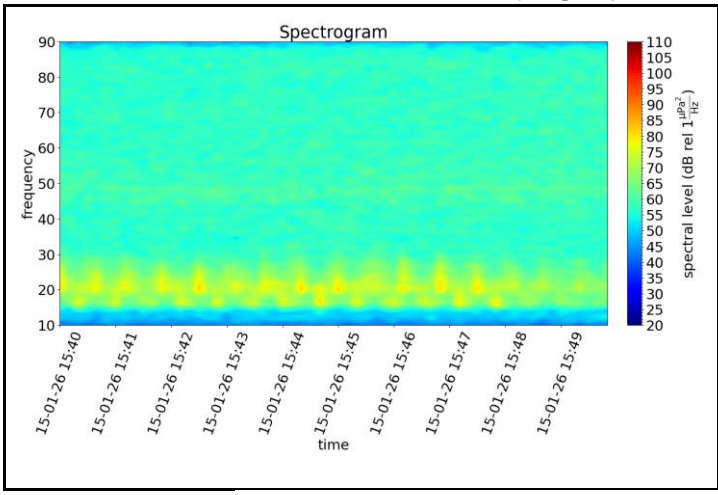


Figure 1A - No Ships Present

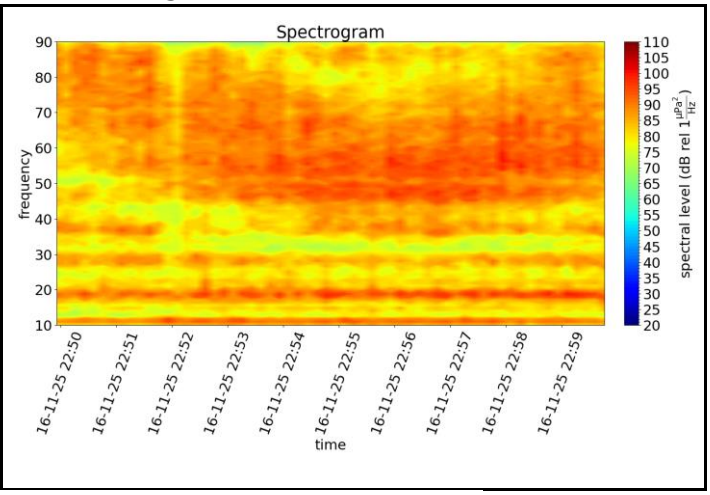


Figure 1B - Bulk Carrier Ship

The data used to generate these spectrograms originates from the Ocean Observatories Initiative (OOI) API. This data source has five years of continuous hydrophone data, using hardware located off the coast of Oregon and Washington. For a map of hydrophone locations, refer to Appendix 1. This OOI audio data was queried into 10-minute spectrograms where exactly one ship was around the hydrophone. As discussed in Appendix 2, these spectrograms were identified and labeled using the AIS data and an algorithm that identifies the duration in which a single ship is around a particular hydrophone. There are between 3000 - 5000 instances of these "single ship" spectrograms in each of the 5 hydrophones that we will use for our modeling project.

In addition to the spectrograms containing a ship, there are many more 10-minute spectrograms that characterize the ambient noise without any ships. The "no-ship" spectrograms were polled uniformly over the available data and manually verified using subject matter expertise. More information on the manual labeling of "no-ship" spectrograms can be found in Appendix 2.

To profile the differences between spectrograms that have ships and those with no ships, all of the images of each category were overlaid and averaged together to form a new spectrogram. Figure 2A shows only one distinct line on average at around 45Hz, while Figure 2B shows a prominent line at 60Hz and multiple fainter lines at lower frequencies.

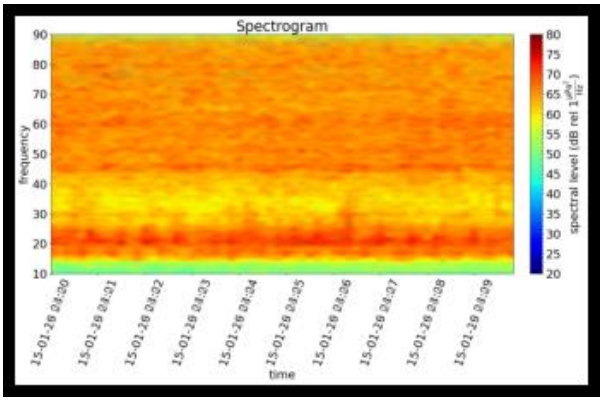


Figure 2A - Average No Ships Present

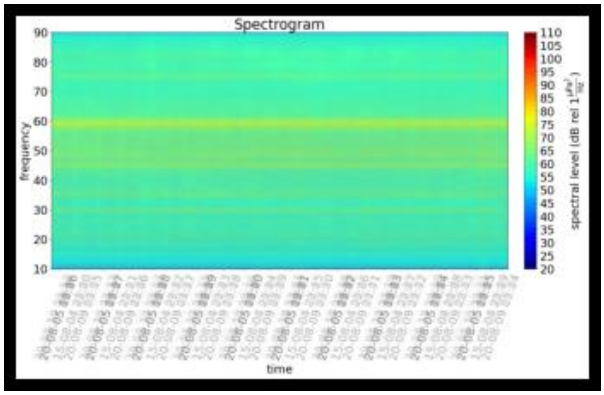


Figure 2B - Average Ships Present

Automatic Identification System (AIS) Ship Labels

There are two distinct automatic identification system downloads that we are using for our analysis. The first dataset covers the area around the Axial base, Central Caldera, and Eastern Caldera hydrophones, and contains ~2 million rows with 11 features. Figure 3 shows a snapshot of the dataset.

	Unnamed: 0.1	Unnamed: 0	MMSI	SHIPNAME	VESSEL TYPE	STATUS	SPEED (KNOTSx10)	LAT	LON	COURSE	HEADING	TIMESTAMP UTC	LENGTH	Year
0	0	0	209605000	AKILI	Bulk Carrier	0	108	46.09859	-129.6550	83	84	2015-01-01 00:28:00	189.99	2014
1	1	1	256832000	NaN	NaN	0	115	45.07486	-128.9430	178	181	2015-01-01 00:28:00	NaN	2014
2	2	2	352358000	ANNA G	Bulk Carrier	0	132	45.86138	-130.5627	86	85	2015-01-01 00:28:00	229.00	2014
3	3	3	356566000	GLOBAL SAIKAI	General Cargo	0	104	46.50261	-129.0129	272	275	2015-01-01 00:31:00	188.50	2014
4	4	4	477293500	JIN XIU FENG	Bulk Carrier	0	112	46.74633	-129.5912	299	298	2015-01-01 00:48:00	229.00	2014

Figure 3: Snapshot of AIS download 1

This dataset is clean and well-labeled with only a few missing values. The right column in Figure 4 is the percentage of missing values in the whole dataset.

Unnamed: 0.1	0.000000
Unnamed: 0	0.000000
MMSI	0.000000
SHIPNAME	0.002452
VESSEL TYPE	0.002766
STATUS	0.000000
SPEED (KNOTSx10)	0.000000
LAT	0.000000
LON	0.000000
COURSE	0.000000
HEADING	0.000000
TIMESTAMP UTC	0.000000
LENGTH	0.002452
Year	0.000000

Figure 4: Missing values in AIS download 1

The second AIS data download covers the Oregon Slope and Southern Hydrate hydrophones and contains ~8.5 million rows with 12 columns. Figure 5 shows the snapshot of the dataset.

	Unnamed: 0.1	Unnamed: 0	MMSI_x	BaseDateTime	LAT	LON	SOG	COG	Heading	VesselName	IMO	CallSign	Status	LENGTH
0	0	0	636092819	2021-01-01T00:17:40	44.5682	-125.62248	11.0	180.6	181.0	NaN	NaN	NaN	0.0	NaN
1	1	0	636092819	2021-01-01T00:17:40	44.5682	-125.62248	11.0	180.6	181.0	NaN	NaN	NaN	0.0	NaN
2	2	0	636092819	2021-01-01T00:17:40	44.5682	-125.62248	11.0	180.6	181.0	NaN	NaN	NaN	0.0	NaN
3	3	0	636092819	2021-01-01T00:17:40	44.5682	-125.62248	11.0	180.6	181.0	NaN	NaN	NaN	0.0	NaN
4	4	0	636092819	2021-01-01T00:17:40	44.5682	-125.62248	11.0	180.6	181.0	NaN	NaN	NaN	0.0	NaN

Figure 5: Snapshot of AIS download 2

In this dataset, around 20-30% of the ship's information is missing, and these missing labels will be excluded from our analysis.

Unnamed: 0.1	0.000000
Unnamed: 0	0.000000
MMSI_x	0.000000
BaseDateTime	0.000000
LAT	0.000000
LON	0.000000
SOG	0.000000
COG	0.000000
Heading	0.000000
VesselName	0.219643
IMO	0.270544
CallSign	0.228178
Status	0.042849
LENGTH	0.275054
Width	0.327129
Draft	0.452229
Cargo	0.543584
TransceiverClass	0.703205
TIMESTAMP UTC	0.000000

Figure 6: Missing values in AIS download 2

Data Quality Concerns

We are aware of several sources of bias that might impact our modeling results. First and foremost, our labels are subjective. They simplify complex, real-world data, and even though they are informed by subject matter expertise, they are subject to human error. There may be multiple ships in a single hydrophone recording, or there may be ship noise from a ship that does not report its coordinates to AIS. Smaller fishing vessels are not required to report. ("AIS Frequently Asked Questions | Navigation Center" n.d.) Figure 7 is an example of an image that was labeled "no ship" yet has some audio signal in frequencies that signal ship presence.

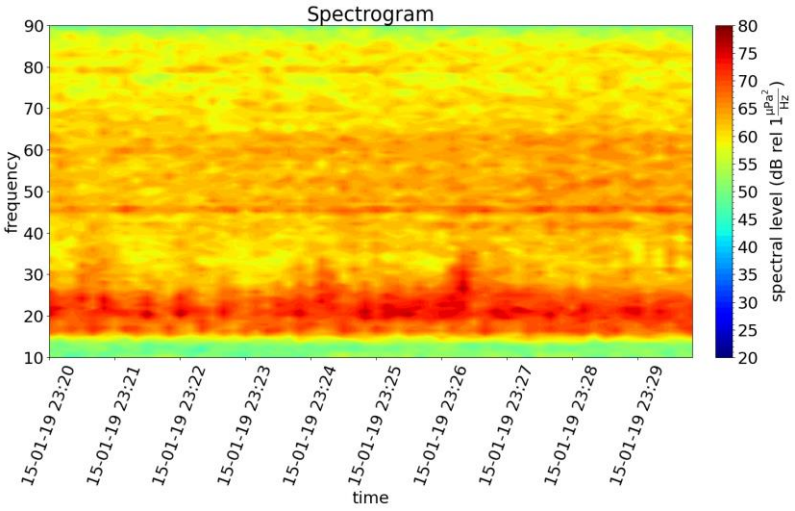
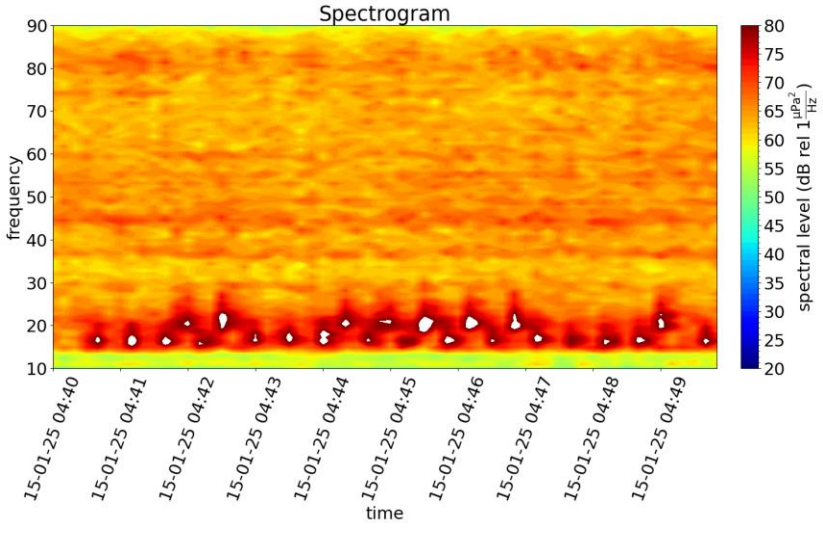


Figure 7: "No ship" image with data quality concerns

Other sources of error might be mechanical issues with the hydrophones or errors in converting the audio signals to images. For example, in the below spectrogram, the concerning white patches are due to the out-of-range values in the colormap. If we increase the range of the colormap, the white patches would turn to dark red but the faint horizontal red lines representing the signature of a ship's presence may become fainter making it harder to visually inspect the spectrogram.

Figure 8: Spectrogram image with white patches



As with any real-world process, we expect these imperfections in data quality. Our model will need to be robust to these issues in order to be useful. We anticipate that modeling and data review will be an interactive process with Dr. Abadi's lab, and additional data can be added where the model underperforms. One way we might keep these quality issues in mind is by considering model uncertainty.

Appendix 4: Model accuracy by class, location

Accuracy by vessel type for random sample dataset (aka dataset A)

true_vessel_type	num train records	num test records	train nships	test nships	intersect nships	% test ships in train	trans-former accuracy	mlp accuracy	cnn accuracy
Bulk Carrier	66	34	34	16	9	0.5625	0.617647	0.735294	0.852941
Cargo, all ships of this type	65	35	35	20	10	0.5	0.514286	0.685714	0.171429
Chemical Tanker	61	39	13	12	7	0.583333	1	1	1
Container Ship	52	48	6	6	6	1	1	1	0.979167
Crude Oil Tanker	61	39	16	12	7	0.583333	1	1	1
Fishing	58	39	15	10	10	1	0.692308	0.641026	0.307692
Fishing\t	56	44	14	10	8	0.8	0.977273	1	0.931818
Fishing Vessel	53	47	6	4	4	1	0.93617	0.957447	0.93617
General Cargo	67	33	21	12	9	0.75	0.909091	1	0.878788
No Ship	63	37	63	37	0	0	0.702703	0.459459	0.594595
Offshore Supply Ship	55	45	3	4	2	0.5	0.933333	0.955556	0.866667
Oil Products Tanker	68	32	12	11	6	0.545455	0.90625	0.96875	0.875
Oil/Chemical Tanker	64	36	14	11	8	0.727273	0.916667	0.944444	0.944444
Passenger, all ships of this type	61	39	13	12	11	0.916667	0.846154	0.871795	0.74359
Pleasure Craft	63	37	7	7	7	1	0.864865	1	0.972973
Sailing	60	42	6	6	6	1	0.880952	0.928571	0.857143
Tanker, all ships of this type	63	37	10	7	7	1	0.972973	0.864865	0.891892
Towing	54	47	16	16	11	0.6875	0.574468	0.446809	0.595745
Tug	56	44	16	18	13	0.722222	0.681818	0.772727	0.886364
Vehicles Carrier	56	44	19	16	12	0.75	0.840909	1	0.977273

Accuracy by vessel type, hydrophone for random sample dataset (aka dataset A)

true vessel type	hydro-phone	num train records	num test records	train nships	test nships	intersect nships	% test ships in train	trans-former accuracy	mlp accuracy	cnn accuracy
Bulk Carrier	AB	12	1	9	1	1	0.0625	1	1	1
	CC	34	13	18	7	4	0.25	0.692308	0.923077	0.923077
	EC	20	20	11	10	5	0.3125	0.55	0.6	0.8
Cargo, all ships of this type	OS	30	15	15	8	4	0.2	0.6	0.866667	0.333333
	SH	35	20	20	12	6	0.3	0.45	0.55	0.05
Chemical Tanker	CC	26	16	9	8	5	0.416667	1	1	1
	EC	35	23	8	5	2	0.166667	1	1	1
Container Ship	CC	11	15	4	5	4	0.666667	1	1	0.933333
	EC	8	5	3	1	1	0.166667	1	1	1
	OS	33	28	2	2	2	0.333333	1	1	1
Crude Oil Tanker	AB	7	4	1	1	1	0.083333	1	1	1
	EC	54	35	15	11	6	0.5	1	1	1
Fishing	OS	23	19	7	6	6	0.6	0.789474	0.631579	0.157895
	SH	35	20	9	5	5	0.5	0.6	0.65	0.45
Fishing\t	CC	19	15	9	6	5	0.5	0.933333	1	0.933333
	EC	17	15	6	5	4	0.4	1	1	0.866667
	OS	20	14	2	2	2	0.2	1	1	1
Fishing Vessel	AB	49	42	3	3	3	0.75	1	1	1
	EC	4	1	3	1	1	0.25	1	0	1
General Cargo	AB	39	20	9	6	6	0.5	0.95	1	0.8
	CC	22	9	10	5	2	0.166667	1	1	1
	EC	6	4	3	2	1	0.083333	0.5	1	1
No Ship	AB	16	3	16	3	0	0	0.666667	0.333333	0.666667

	CC	19	11	19	11	0	0	0.818182	0.545455	0.818182
	EC	12	9	12	9	0	0	0.777778	0.555556	0.777778
	OS	6	7	6	7	0	0	0.428571	0.428571	0.428571
	SH	10	7	10	7	0	0	0.714286	0.285714	0.142857
Offshore Supply Ship	CC	39	32	3	4	2	0.5	0.90625	0.9375	0.84375
	EC	16	13	2	2	1	0.25	1	1	0.923077
Oil Products Tanker	AB	24	12	8	7	3	0.272727	0.75	0.916667	0.75
	CC	10	7	3	3	2	0.181818	1	1	0.857143
	EC	34	13	3	2	2	0.181818	1	1	1
Oil/Chemical Tanker	AB	1	1	1	1	0	0	0	0	0
	CC	26	20	3	5	2	0.181818	0.9	0.95	0.95
	EC	37	15	12	6	6	0.545455	1	1	1
Passenger, all ships of this type	OS	11	9	4	3	3	0.25	0.777778	1	0.888889
	SH	50	30	10	9	8	0.666667	0.866667	0.833333	0.7
Pleasure Craft	OS	33	18	4	4	4	0.571429	0.888889	1	0.944444
	SH	30	19	3	3	3	0.428571	0.842105	1	1
Sailing	OS	53	35	5	5	5	0.833333	0.971429	0.971429	0.857143
	SH	7	7	2	2	2	0.333333	0.428571	0.714286	0.857143
Tanker, all ships of this type	OS	46	27	6	5	5	0.714286	0.962963	0.814815	0.888889
	SH	17	10	4	2	2	0.285714	1	1	0.9
Towing	OS	17	18	5	7	4	0.25	0.444444	0.444444	0.444444
	SH	37	29	11	9	7	0.4375	0.655172	0.448276	0.689655
Tug	AB	6	10	4	5	4	0.222222	0.5	0.7	1
	CC	16	10	8	7	3	0.166667	0.5	0.7	0.9
	EC	34	24	10	11	8	0.444444	0.833333	0.833333	0.833333
Vehicles Carrier	AB	30	21	7	7	6	0.375	0.761905	1	1
	CC	13	13	8	6	5	0.3125	0.846154	1	0.923077

	EC	13	10	7	4	2	0.125	1	1	1

Accuracy by vessel type for time-stratified sample dataset (aka dataset C)

true_vessel_type	num train records	num test records	train nships	test nships	intersect nships	% test ships in train	trans-former accuracy	mlp accuracy	cnn accuracy
Bulk Carrier	60	40	23	19	9	0.473684	0.525	0.625	0.575
Cargo, all ships of this type	60	40	31	26	8	0.307692	0.425	0.475	0.475
Chemical Tanker	60	40	14	9	5	0.555556	0.725	0.725	1
Container Ship	60	40	4	5	3	0.6	1	0.95	0.975
Crude Oil Tanker	14	40	1	14	0	0	0	0	0
Fishing	60	40	13	12	5	0.416667	0.175	0.125	0.025
Fishing\t	60	36	15	3	2	0.666667	0.027778	0.027778	0.027778
Fishing Vessel	60	40	8	3	0	0	0	0.075	0.05
General Cargo	60	40	20	5	0	0	0	0.25	0.175
No Ship	60	40	60	40	0	0	0.825	0.475	0.7
Offshore Supply Ship	60	40	3	5	1	0.2	0.025	0.575	0.45
Oil Products Tanker	60	40	13	2	1	0.5	0.025	0.025	0.025
Oil/Chemical Tanker	60	40	4	15	0	0	0	0	0
Passenger, all ships of this type	60	40	9	9	4	0.444444	0.2	0.15	0.075
Pleasure Craft	60	40	4	4	1	0.25	0.025	0.025	0.025
Sailing	61	40	5	3	2	0.666667	0.175	0	0
Tanker, all ships of this type	60	40	9	4	2	0.5	0.375	0.55	0.575
Towing	60	40	16	10	5	0.5	0.225	0.4	0.35
Tug	60	40	16	10	5	0.5	0.75	0.35	0.2
Vehicles Carrier	60	39	18	14	9	0.642857	0.179487	0.25641	0.307692

Accuracy by vessel type, hydrophone for time-stratified sample dataset (aka dataset C)

true vessel type	hydro-phone	num train records	num test records	train nships	test nships	intersect nships	% test ships in train	trans-former accuracy	mlp accuracy	cnn accuracy
Bulk Carrier	AB	4	8	2	4	0	0	0.625	0.875	0.875
	CC	19	20	11	12	6	0.315789	0.45	0.55	0.45
	EC	37	12	14	8	4	0.210526	0.583333	0.583333	0.583333
Cargo, all ships of this type	OS	25	19	12	12	5	0.192308	0.473684	0.631579	0.631579
	SH	35	21	19	14	3	0.115385	0.380952	0.333333	0.333333
Chemical Tanker	EC	16	40	4	9	3	0.333333	0.725	0.725	1
Container Ship	CC	7	18	3	4	2	0.4	1	0.888889	1
	EC	6	10	2	2	2	0.4	1	1	0.9
	OS	47	12	2	2	2	0.4	1	1	1
Fishing	OS	21	29	4	7	1	0.083333	0.241379	0.103448	0
	SH	39	11	9	5	2	0.166667	0	0.181818	0.090909
Fishing\t	EC	26	1	7	1	1	0.333333	1	1	1
General Cargo	AB	41	10	9	2	0	0	0	0.9	0
	CC	12	23	7	3	0	0	0	0.043478	0.26087
	EC	7	7	4	1	0	0	0	0	0.142857
No Ship	AB	10	5	0	0	0	0	1	1	0.8
	CC	14	8	0	0	0	0	1	0.625	0.75
	EC	19	8	0	0	0	0	1	0.5	0.875
	OS	7	8	0	0	0	0	0.5	0.375	0.875
	SH	10	11	0	0	0	0	0.727273	0.181818	0.363636
Offshore Supply Ship	CC	38	40	2	5	1	0.2	0.025	0.575	0.45
Oil Products Tanker	AB	14	39	10	1	0	0	0	0	0
	CC	15	1	2	1	1	0.5	1	1	1

Oil/Chemical Tanker	**AB**	5	4	2	3	0	0	0	0	0
	CC	31	5	2	3	0	0	0	0	0
	EC	24	31	1	11	0	0	0	0	0
Passenger, all ships of this type	**OS**	18	9	3	2	1	0.111111	0	0	0
	SH	42	31	6	7	2	0.222222	0.258065	0.193548	0.096774
Pleasure Craft	**OS**	14	40	1	4	1	0.25	0.025	0.025	0.025
Sailing	**OS**	49	40	3	3	1	0.333333	0.175	0	0
Tanker, all ships of this type	**OS**	34	40	4	4	2	0.5	0.375	0.55	0.575
Towing	**OS**	12	22	5	5	2	0.2	0	0.181818	0.227273
	SH	48	18	11	5	3	0.3	0.5	0.666667	0.5
Tug	**AB**	2	8	2	3	0	0	0.75	0	0
	CC	24	11	12	5	2	0.2	0.363636	0.090909	0
	EC	34	21	7	5	2	0.2	0.952381	0.619048	0.380952
Vehicles Carrier	**AB**	23	29	6	9	5	0.357143	0.241379	0.344828	0.413793
	EC	10	10	7	5	2	0.142857	0	0	0

Appendix 5: Characteristics of ships by "Vessel Type" Label

Vessel Type	Distinct Ships	Min Length	Mean Length	Max Length	Min Speed	Mean Speed	Max Speed
Cargo, all ships of this type	109	72	228.45	349	66.40	145.78	223.50
Bulk Carrier	84	169	200.23	289	7.70	94.33	138.00
Fishing	23	13	43.26	112	39.35	75.81	138.50
Towing	21	24	34.90	42	53.16	83.25	124.00
Passenger, all ships of this type	11	45	238.91	348	81.00	159.81	208.50
Pleasure Craft	10	11	42.60	71	68.60	114.35	154.50
Vehicles Carrier	9	188	198.67	200	58.14	128.94	163.00
Tug	9	24	34.67	40	62.00	84.53	90.73
Tanker, all ships of this type	8	102	198.00	245	96.80	124.03	143.50
General Cargo	8	142	184.00	209	7.15	88.21	123.00
Oil/Chemical Tanker	7	147	162.29	183	12.33	89.26	136.50
Crude Oil Tanker	4	249	267.25	274	20.00	93.60	143.67
Sailing	4	13	19.25	30	39.25	67.85	100.81
Fishing Vessel	3	27	29.33	32	26.00	67.00	119.00
Oil Products Tanker	3	180	196.00	228	20.00	52.33	115.00
Towing Vessel	2	37	37.00	37	90.60	90.80	91.00
Offshore Supply Ship	2	41	58.50	76	50.32	84.41	118.50
Cargo	2	190	195.00	200	87.00	107.00	127.00
Cargo, Hazardous category B	2	294	299.00	304	125.00	149.25	173.50

Cargo, Hazardous category A	2	260	277.00	294	144.44	170.47	196.50
Wood Chips Carrier	1	200	200.00	200	106.00	106.00	106.00
Tanker, No additional information	1	186	186.00	186	140.50	140.50	140.50
Special Vessel	1	50	50.00	50	53.50	53.50	53.50
Spare - Local Vessel	1	132	132.00	132	88.00	88.00	88.00
Self Discharging Bulk Carrier	1	223	223.00	223	120.00	120.00	120.00
Other Type, Hazardous category A	1	264	264.00	264	84.74	84.74	84.74
Ro-Ro/Passenger Ship	1	39	39.00	39	94.77	94.77	94.77
Ro-Ro Cargo	1	200	200.00	200	150.00	150.00	150.00
Reefer/Container Ship	1	151	151.00	151	131.83	131.83	131.83
Passenger Ship	1	317	317.00	317	163.43	163.43	163.43
Not available (default)	1	22	22.00	22	41.19	41.19	41.19
LPG Tanker	1	226	226.00	226	99.00	99.00	99.00
Inland, Unknown	1	32	32.00	32	68.00	68.00	68.00
Container Ship	1	262	262.00	262	225.13	225.13	225.13
Chemical Tanker	1	183	183.00	183	63.33	63.33	63.33
Cargo, No additional information	1	294	294.00	294	125.50	125.50	125.50
Cargo, Hazardous category D	1	262	262.00	262	119.50	119.50	119.50
Yacht	1	40	40.00	40	120.50	120.50	120.50

Appendix 6: Additional recording classification questions

Did the heading provide any information about why the clustering algorithm predicted a certain class?

We did not have time to do a deep dive into this analysis. We were given headings from 0-360 degrees, and ideally you would do a geographic analysis to track the ships direction relative to the hydrophone it was reporting to at a given time. We did not do this, and instead just separated heading into 4 quadrants (0-90 degrees, 91-180 degrees, etc.) This had no correlation with the class prediction from the PSD clustering.

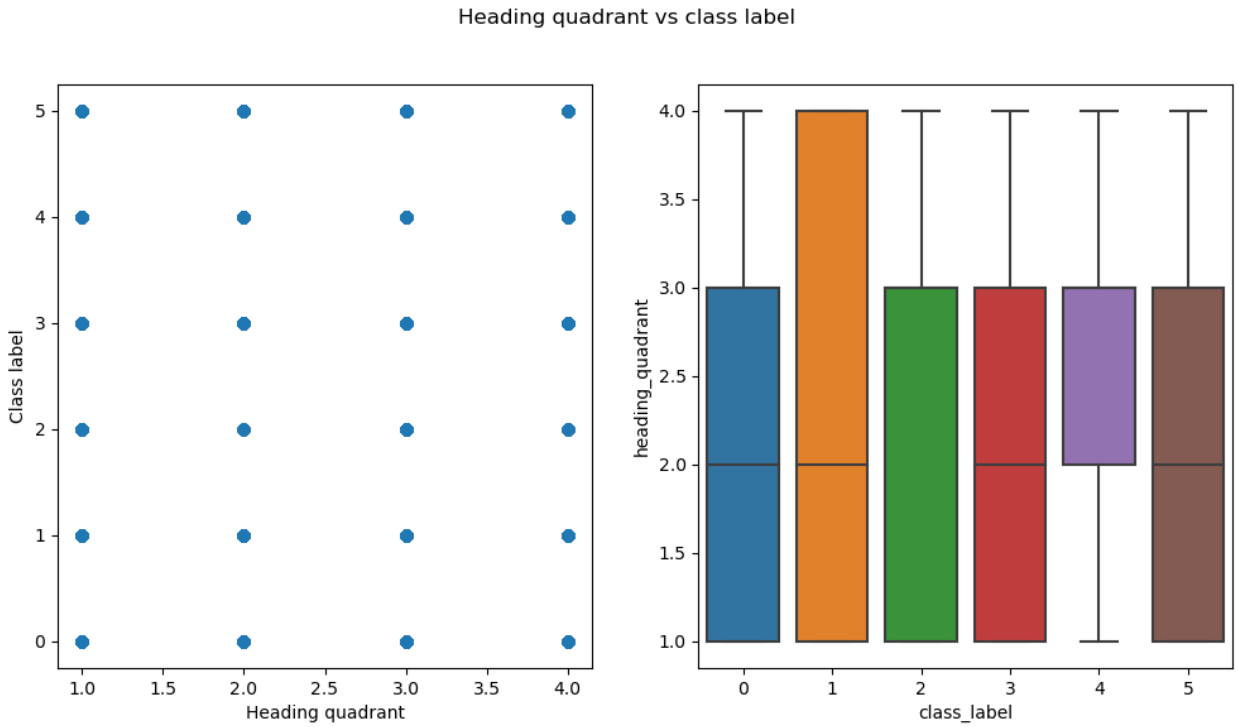


Figure 1: Heading quadrant versus predicted K-means class

Did redoing the clustering algorithm for just Axial Base hydrophone provide any additional information?

The results of the clustering algorithm on just Axial Base ships were different than on all locations, but it is still difficult to draw conclusions. In particular, cluster 3 in Axial Base had specific features that made it distinct from all other classes. Vessels with this label tended to be shorter in length and slow at the time of the recording.

Figures 1 and 2 show the vessel speed and length plots for Axial Base ships only.

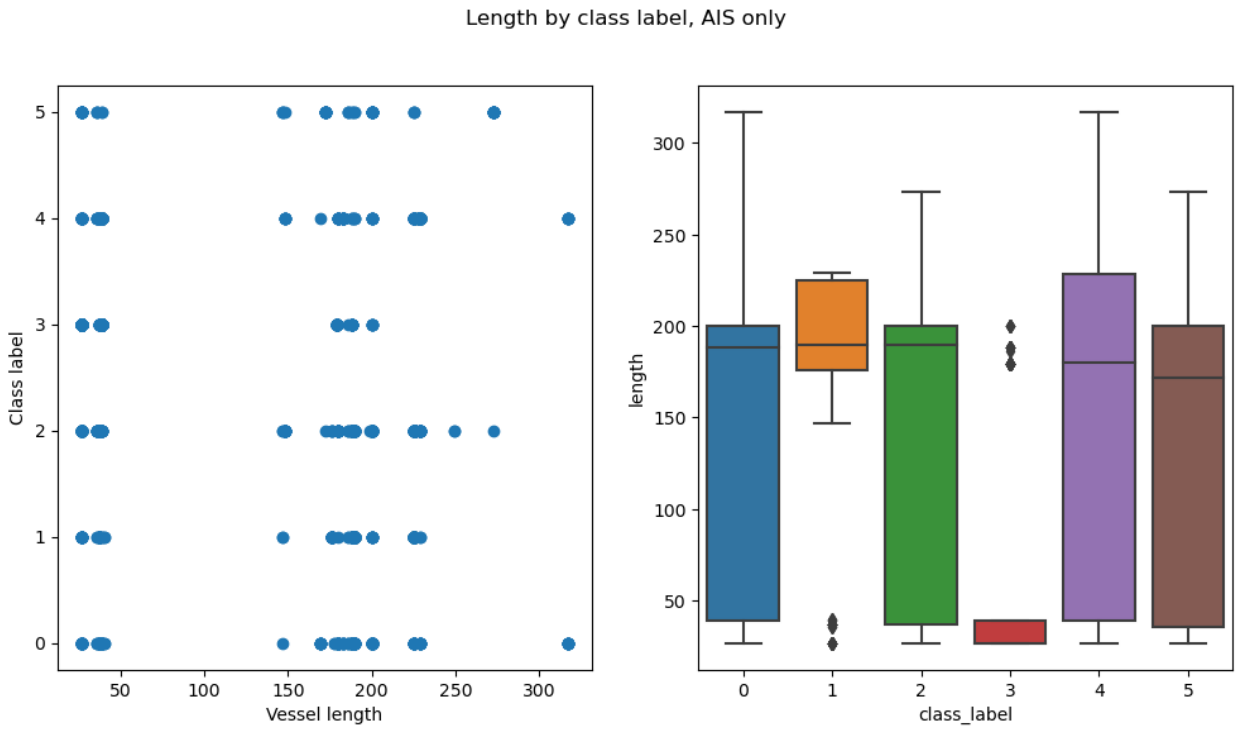


Figure 1: Vessel length by predicted K-means class, Axial Base only

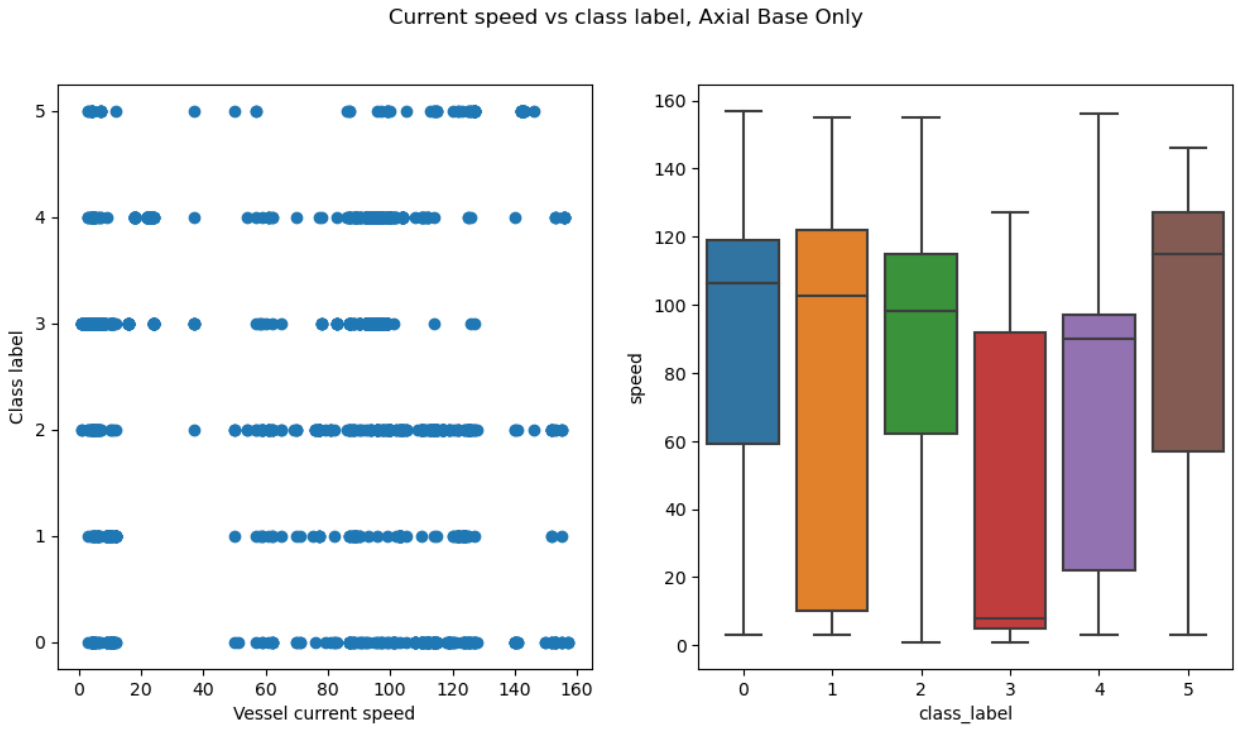
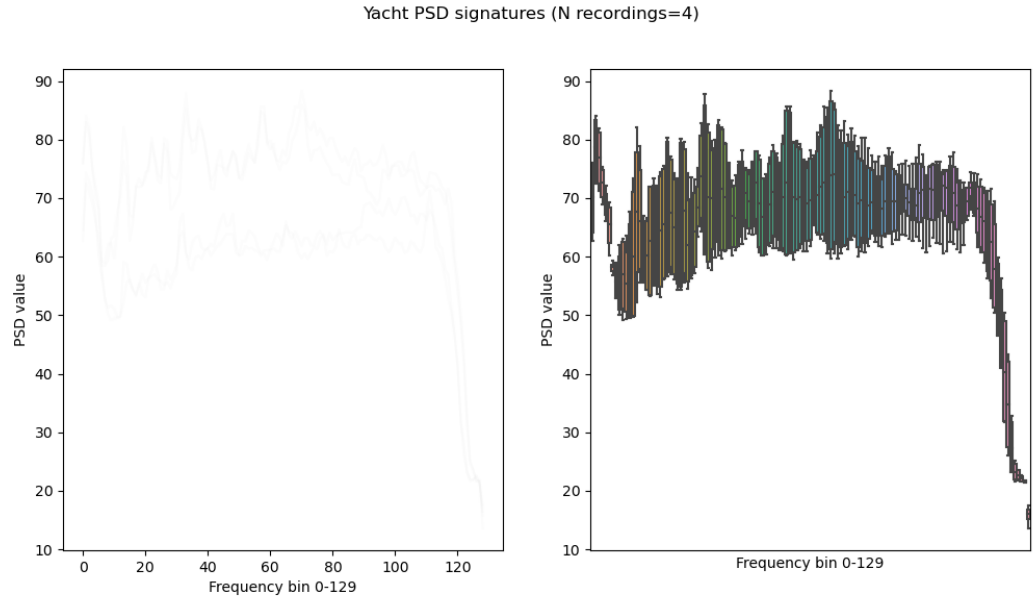
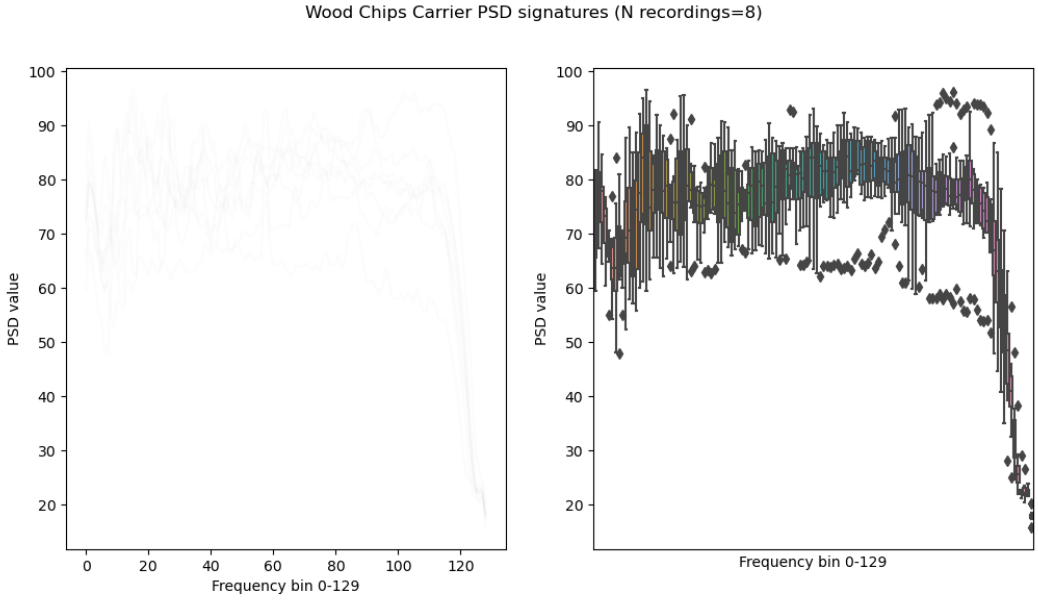
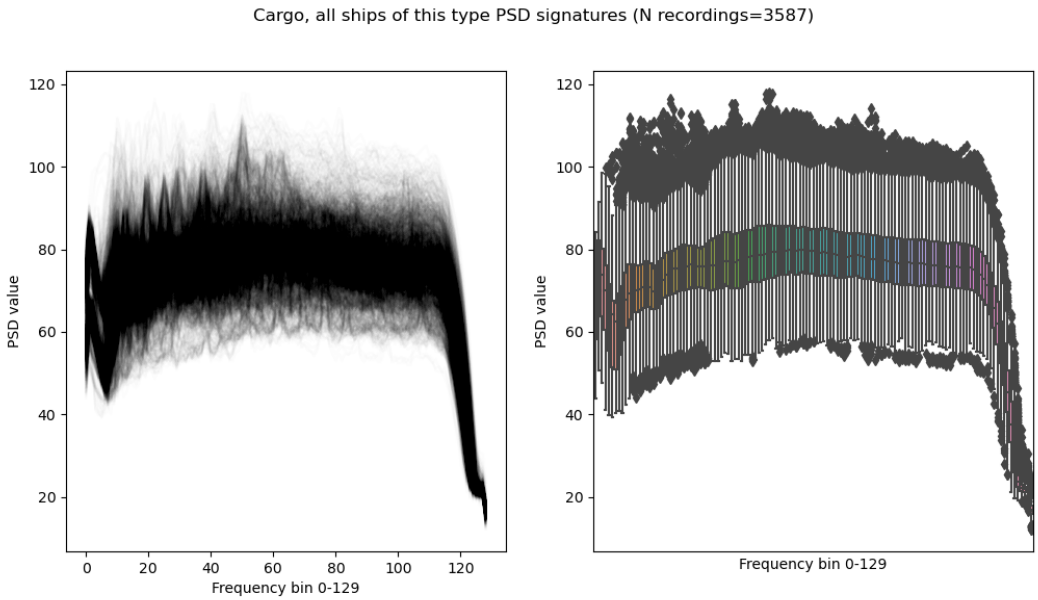


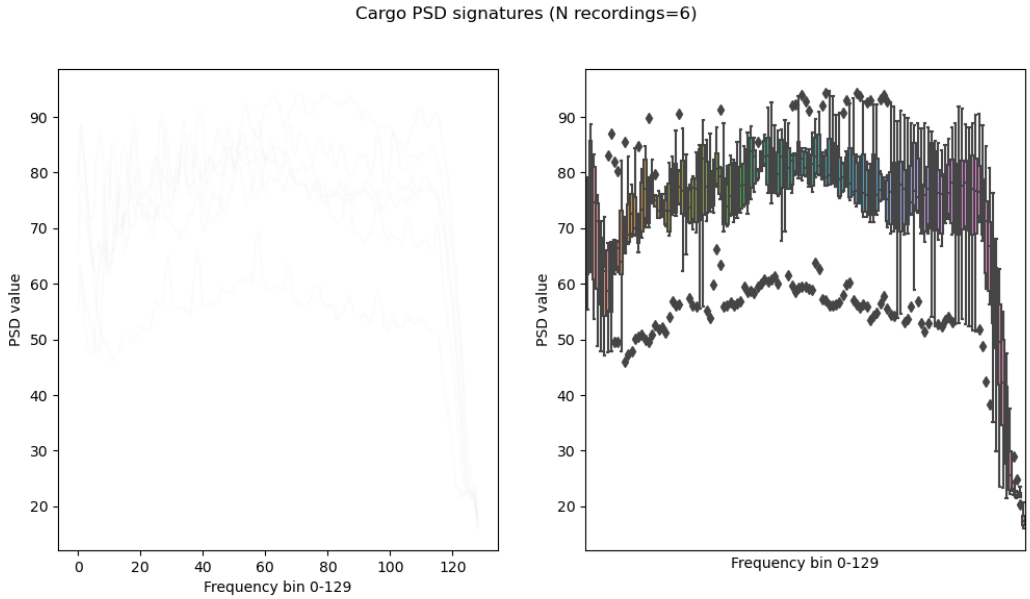
Figure 2: Current vessel speed by predicted K-means class, Axial Base only

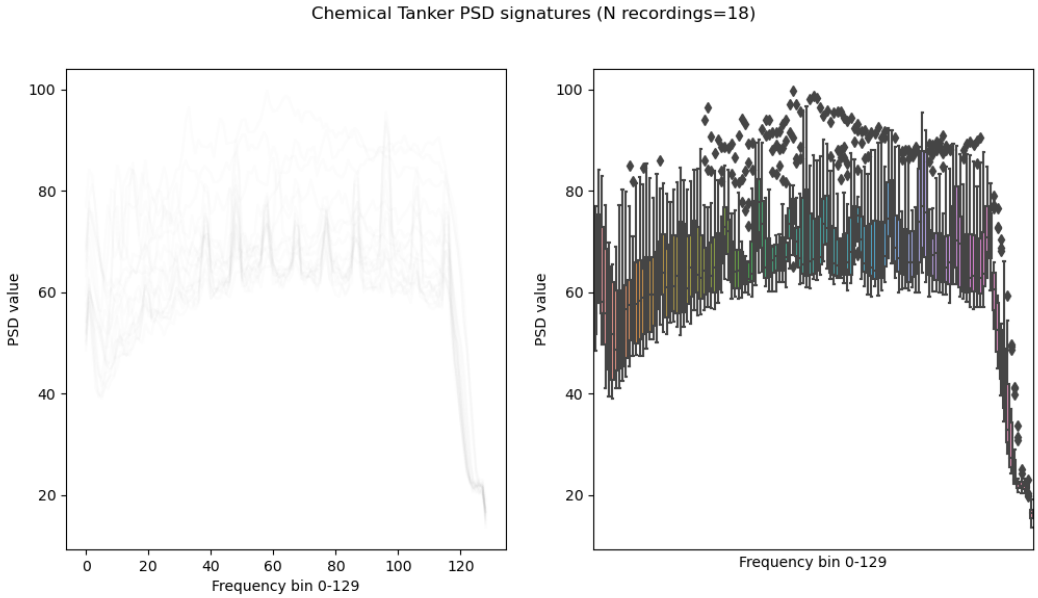
The following graphs show the variance in power-spectral density vectors for each vessel type class. This is a complement to the analysis presented in the “Data Exploration” section.

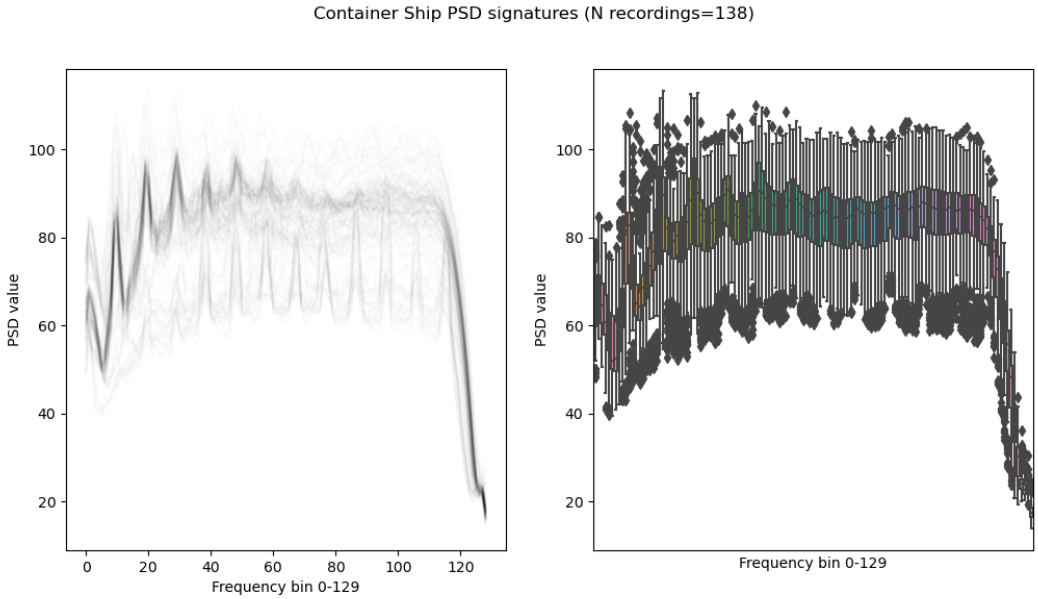

Yacht PSD signatures (N recordings=4)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

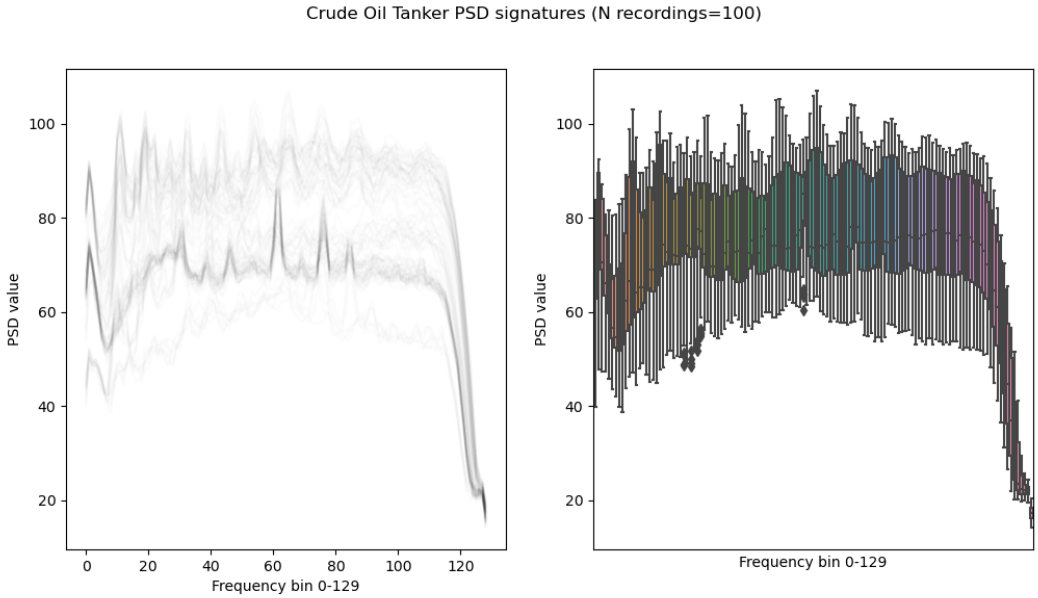

Wood Chips Carrier PSD signatures (N recordings=8)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

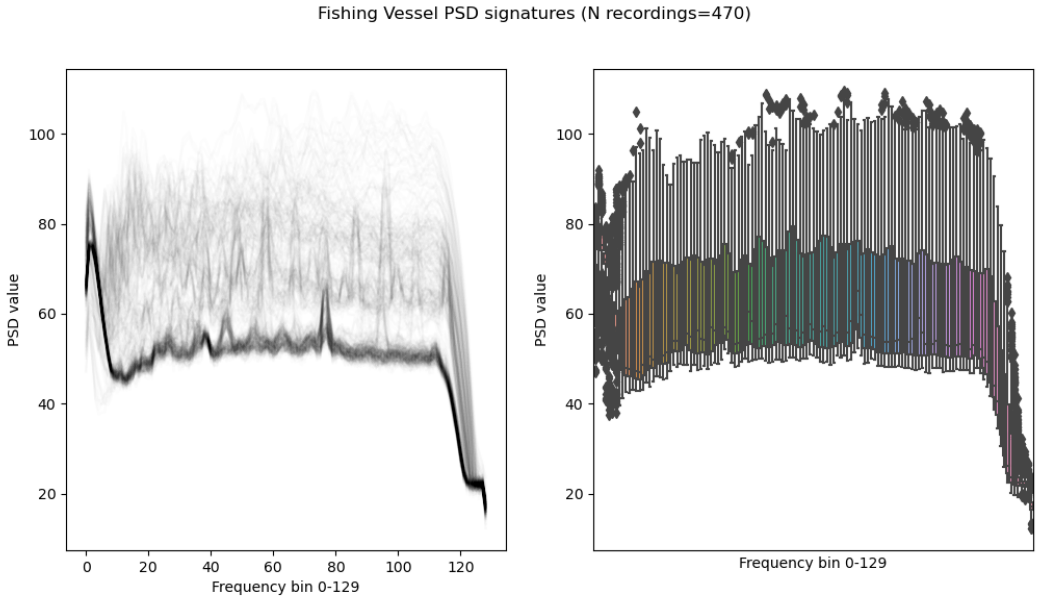

Cargo, all ships of this type PSD signatures (N recordings=3587)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

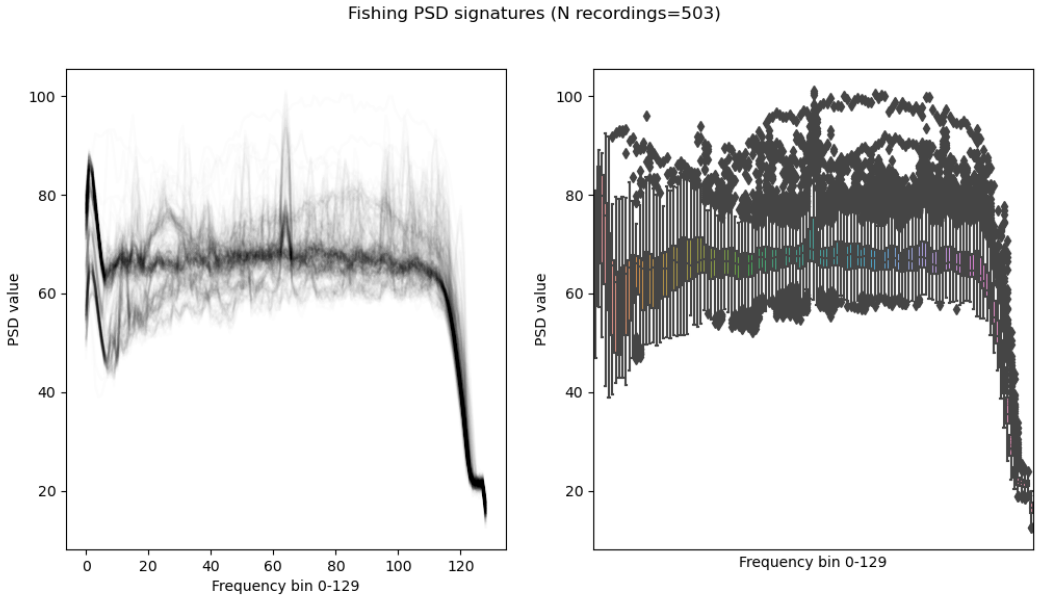

Cargo PSD signatures (N recordings=6)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

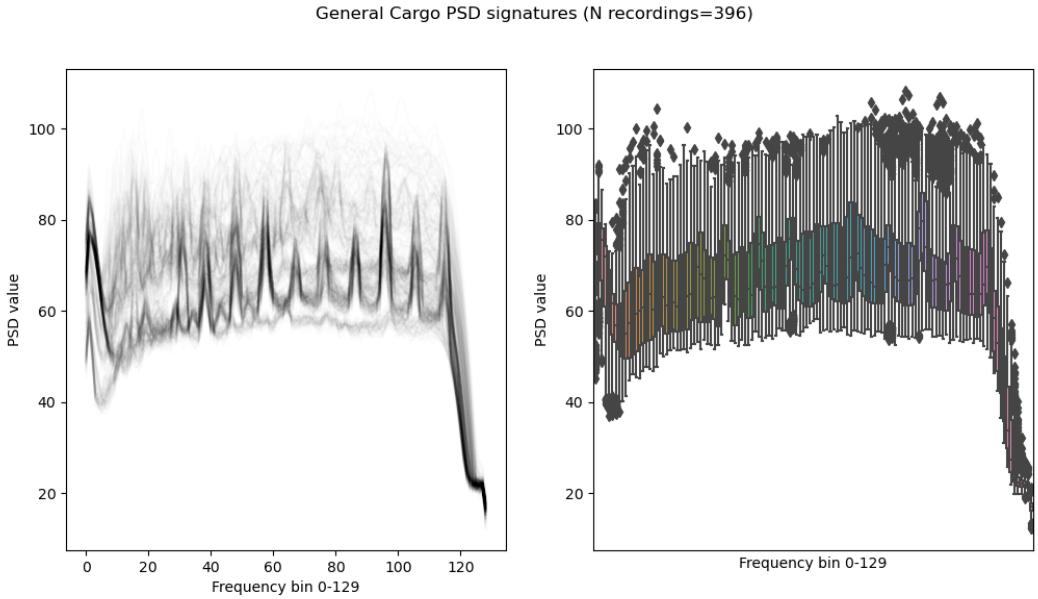

Chemical Tanker PSD signatures (N recordings=18)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

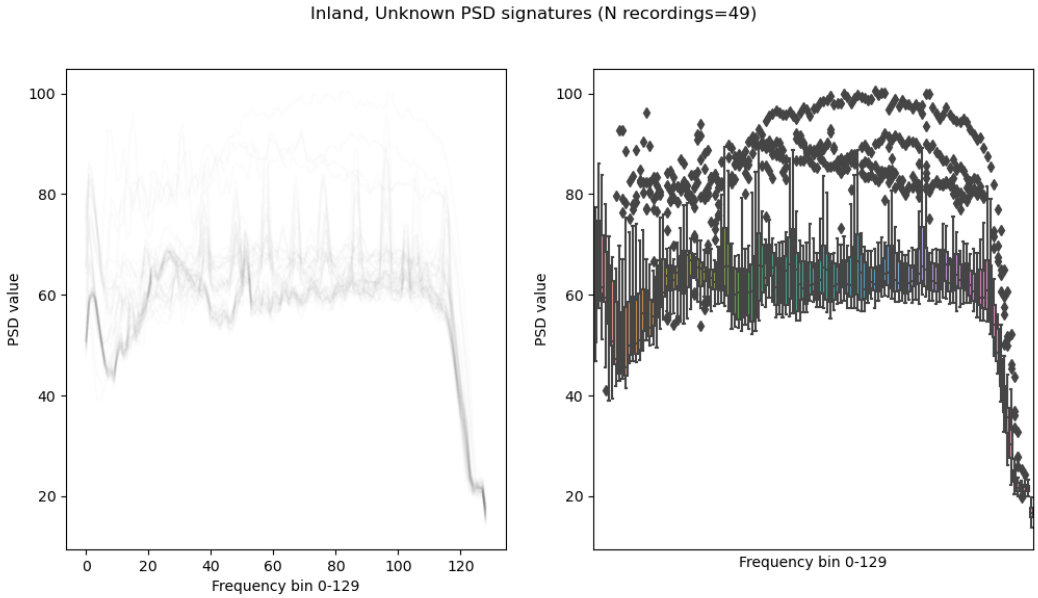

Container Ship PSD signatures (N recordings=138)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

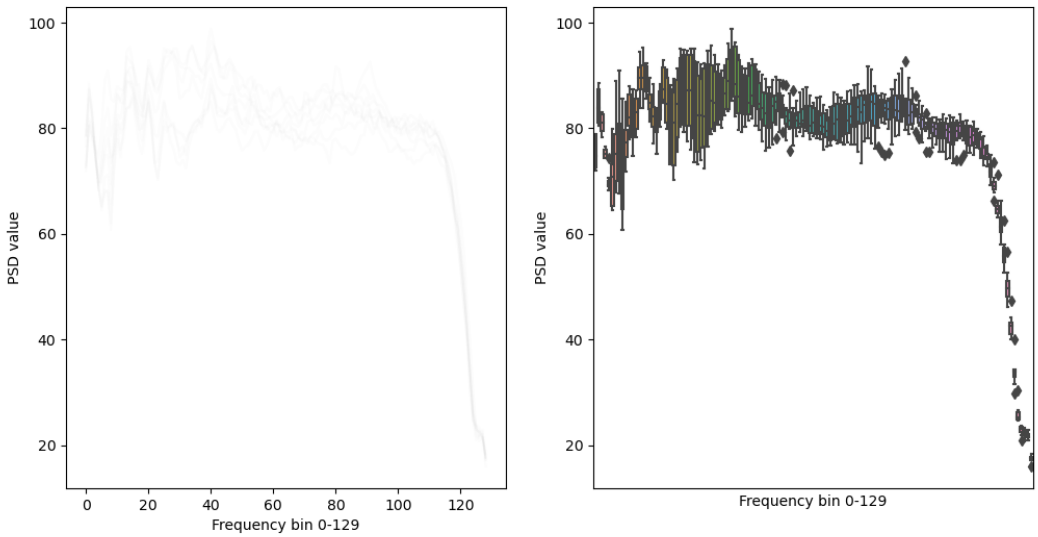

Crude Oil Tanker PSD signatures (N recordings=100)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

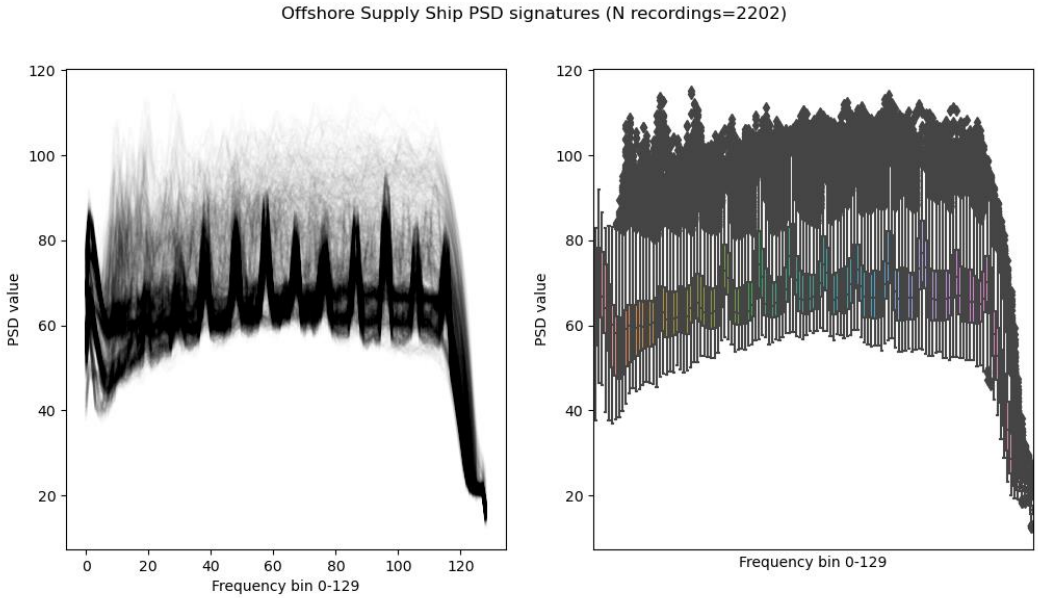

Fishing Vessel PSD signatures (N recordings=470)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

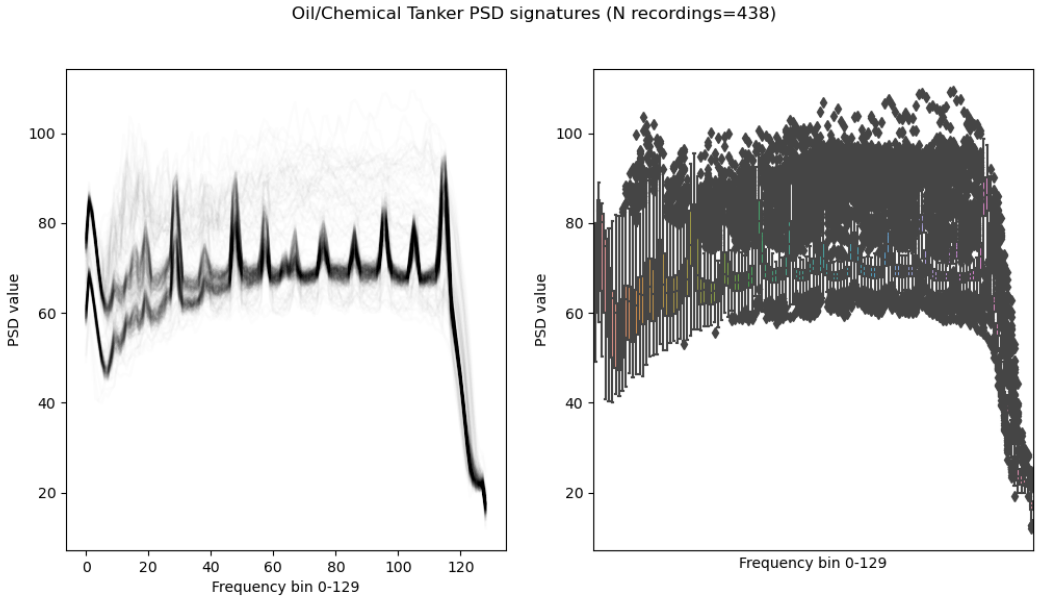

Fishing PSD signatures (N recordings=503)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

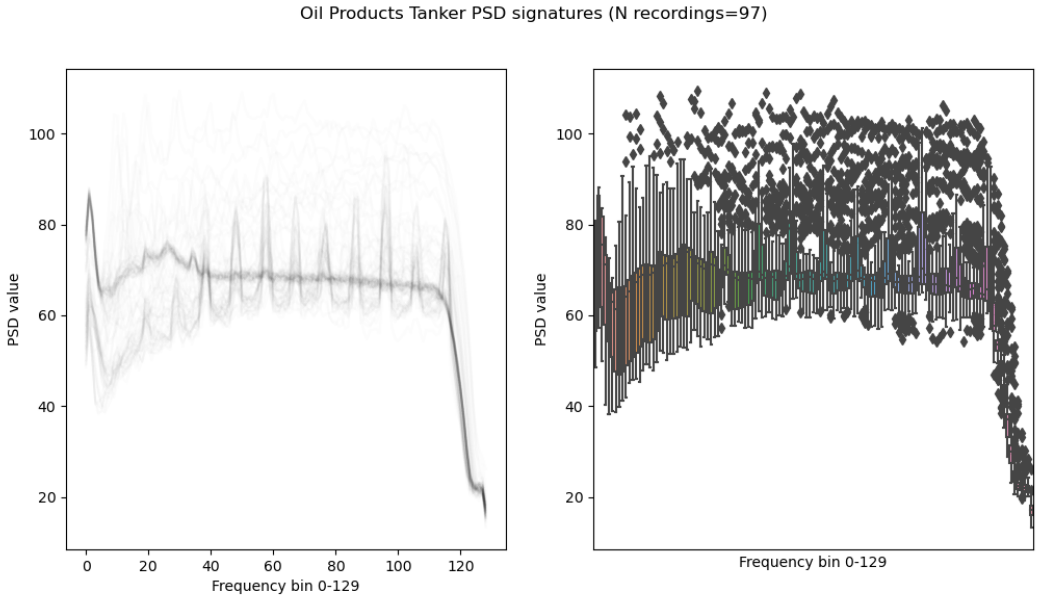

General Cargo PSD signatures (N recordings=396)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

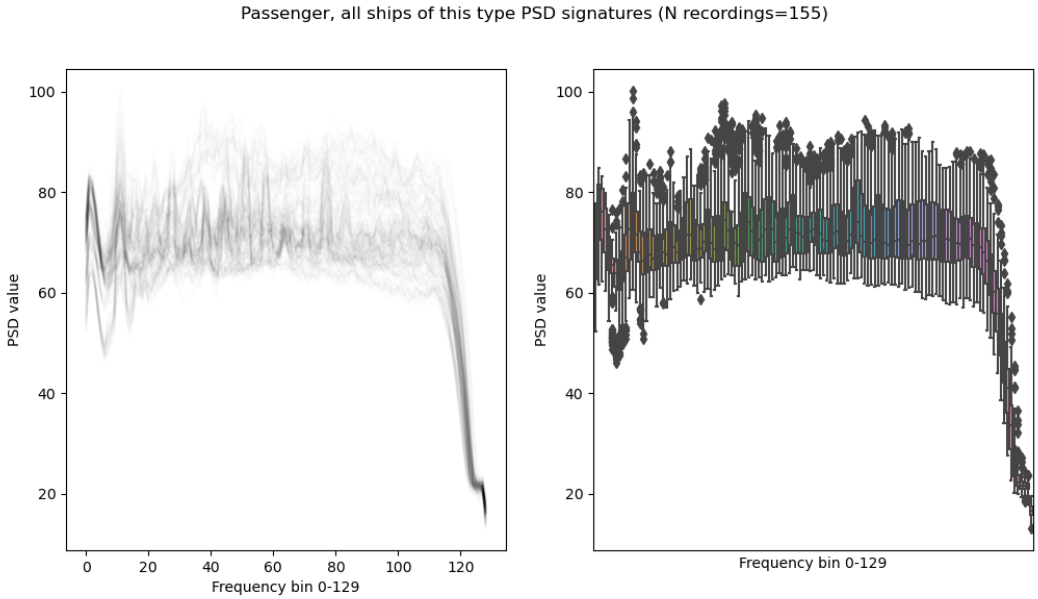

Inland, Unknown PSD signatures (N recordings=49)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

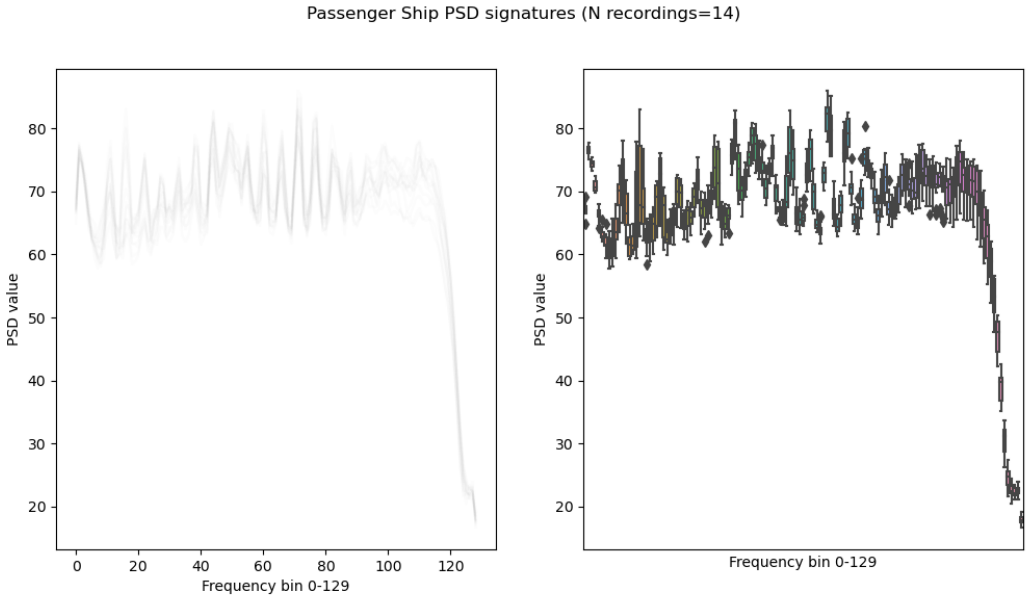

LPG Tanker PSD signatures (N recordings=8)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

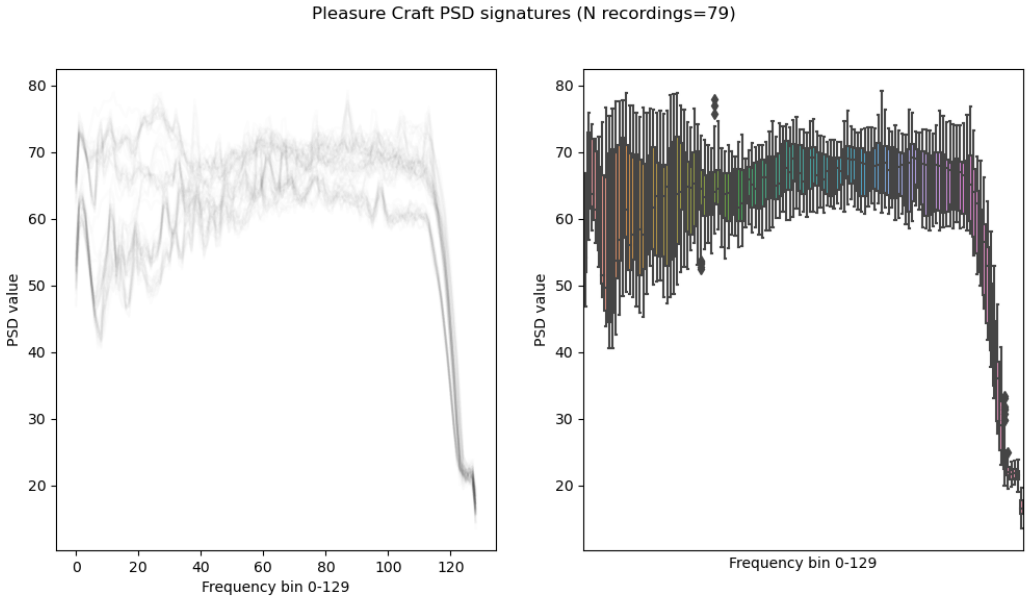

Offshore Supply Ship PSD signatures (N recordings=2202)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

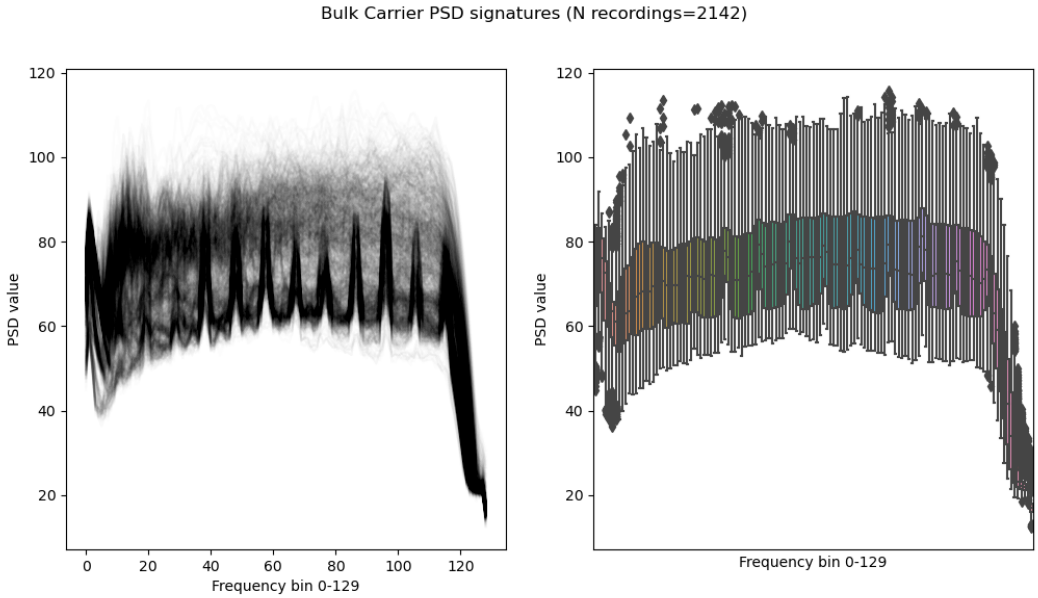

Oil/Chemical Tanker PSD signatures (N recordings=438)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

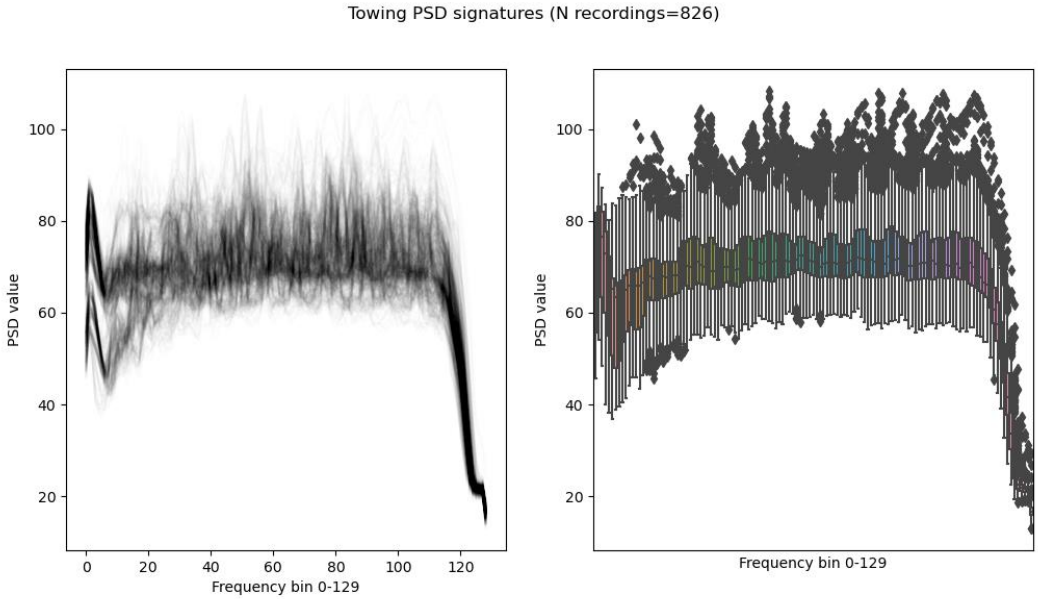

Oil Products Tanker PSD signatures (N recordings=97)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

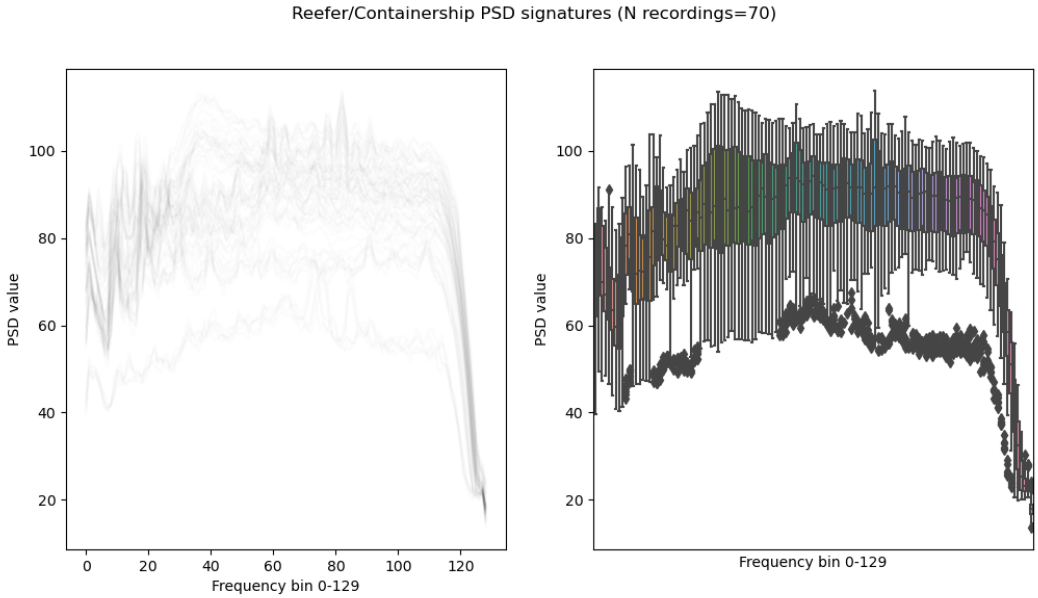

Passenger, all ships of this type PSD signatures (N recordings=155)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

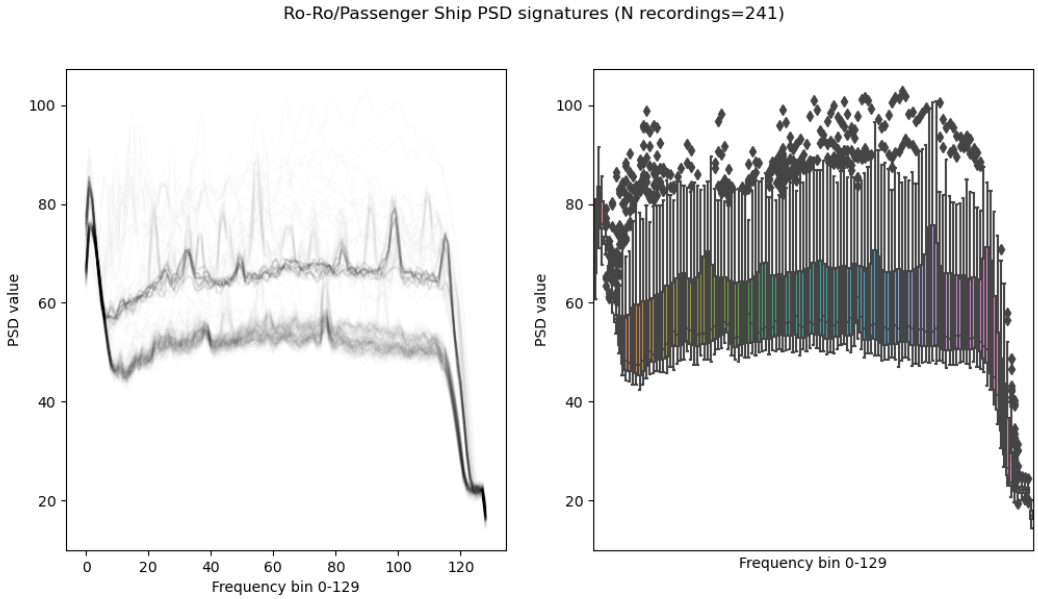

Passenger Ship PSD signatures (N recordings=14)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

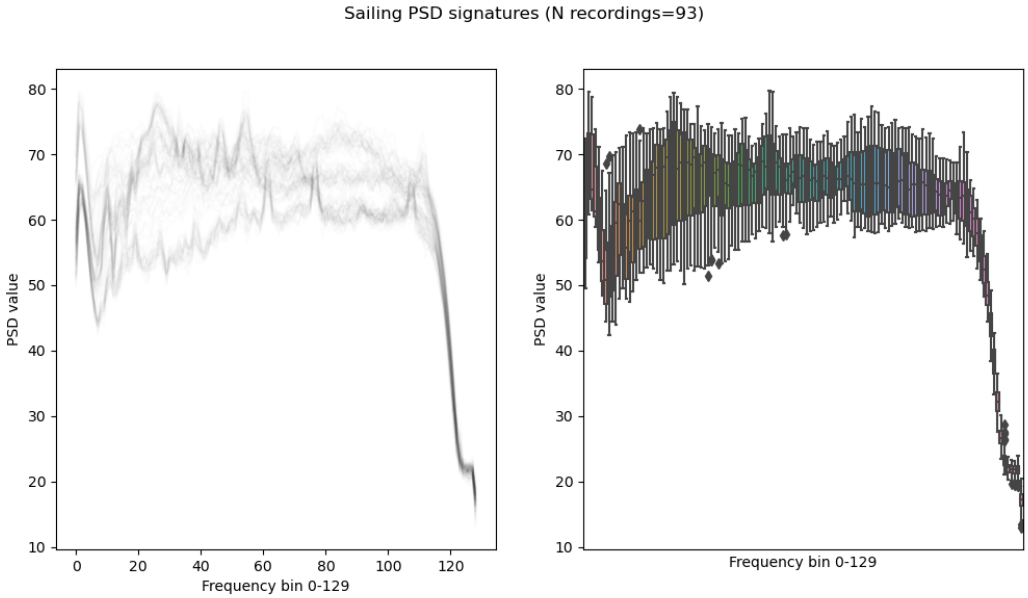

Pleasure Craft PSD signatures (N recordings=79)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

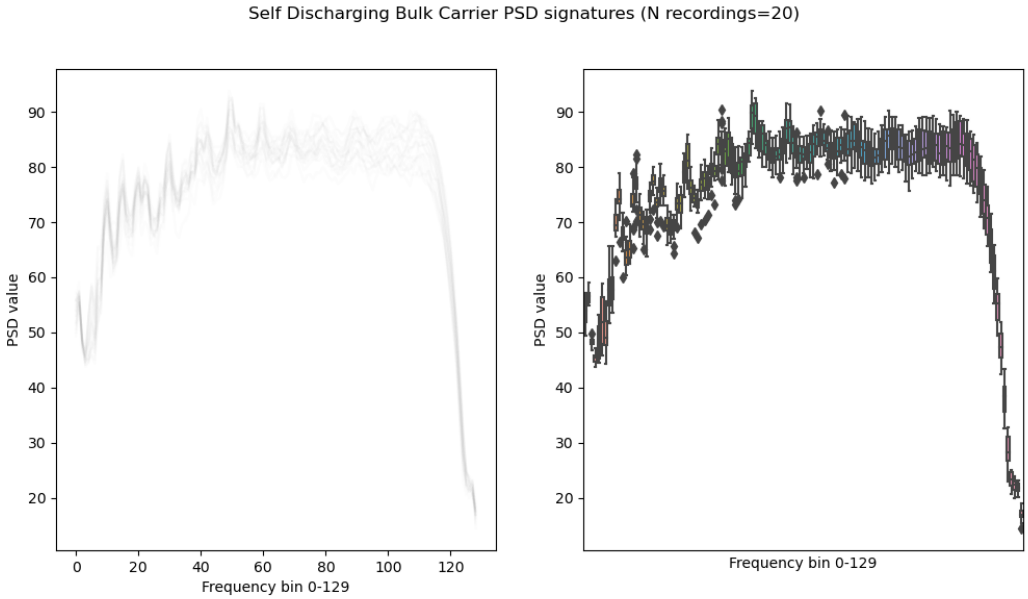

Bulk Carrier PSD signatures (N recordings=2142)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

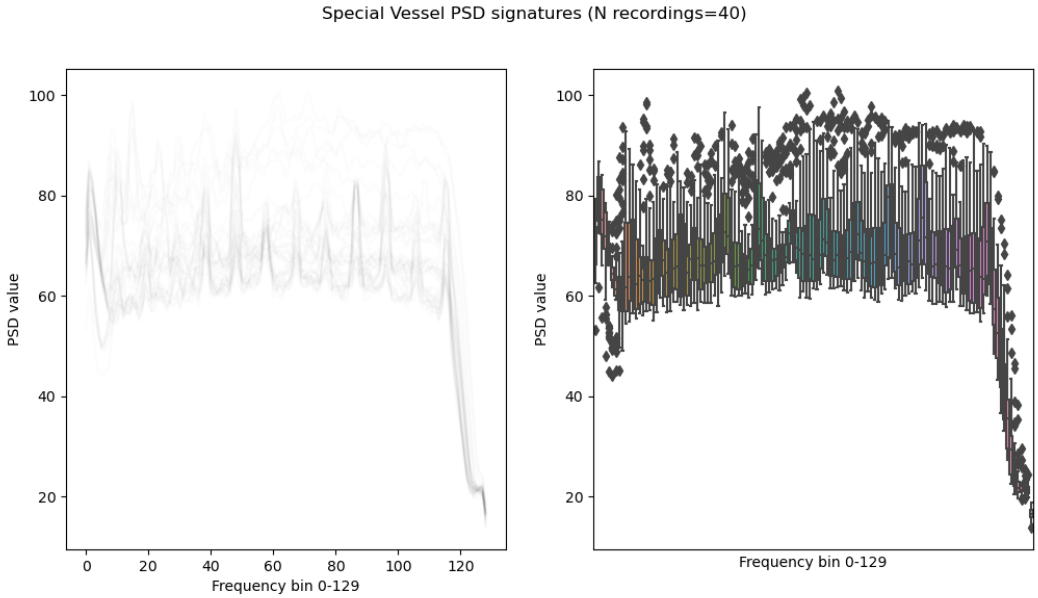

Towing PSD signatures (N recordings=826)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

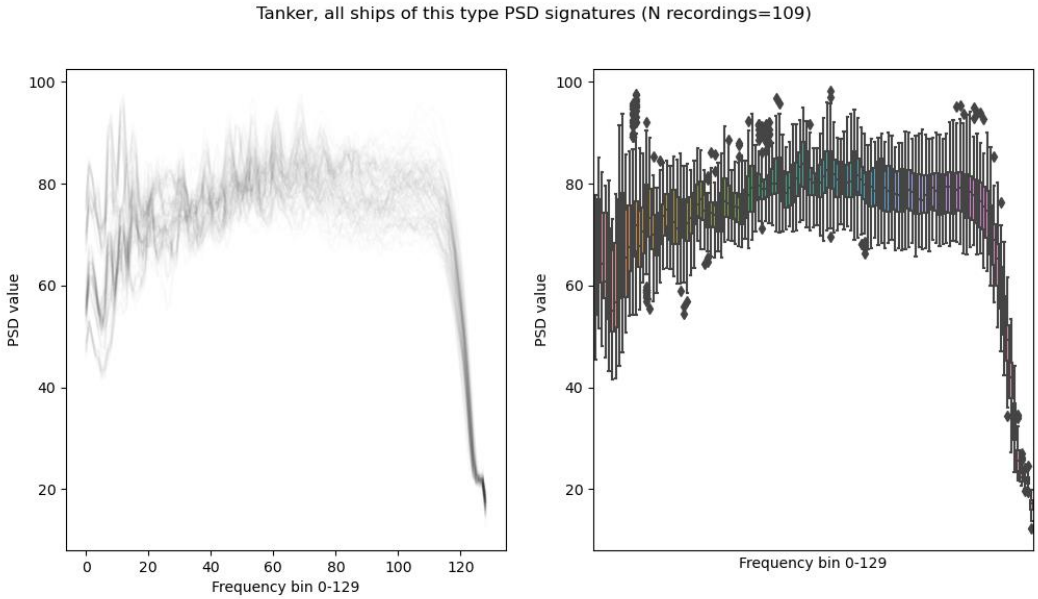

Reefer/Containership PSD signatures (N recordings=70)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

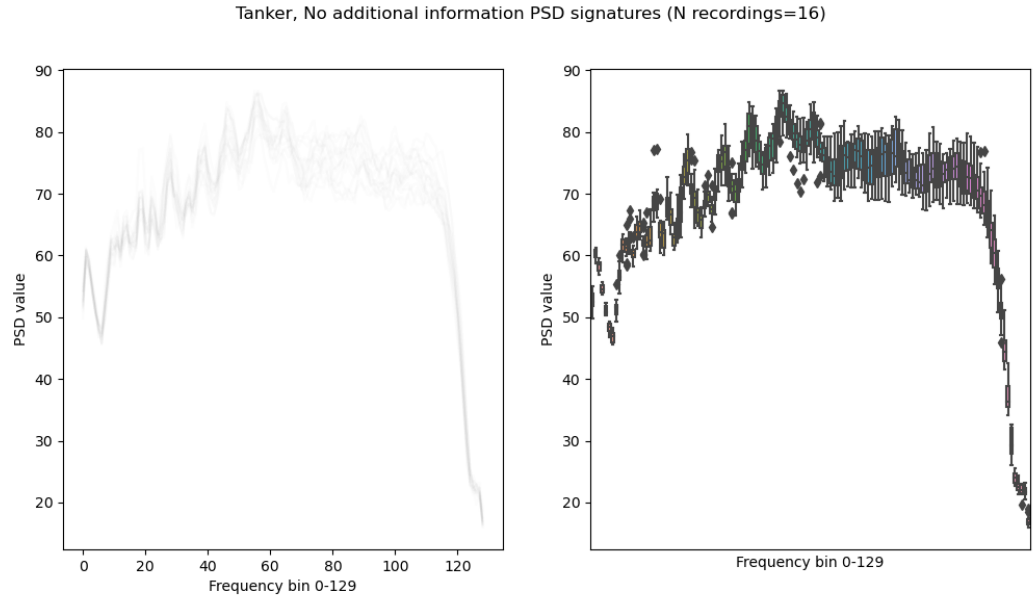

Ro-Ro/Passenger Ship PSD signatures (N recordings=241)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

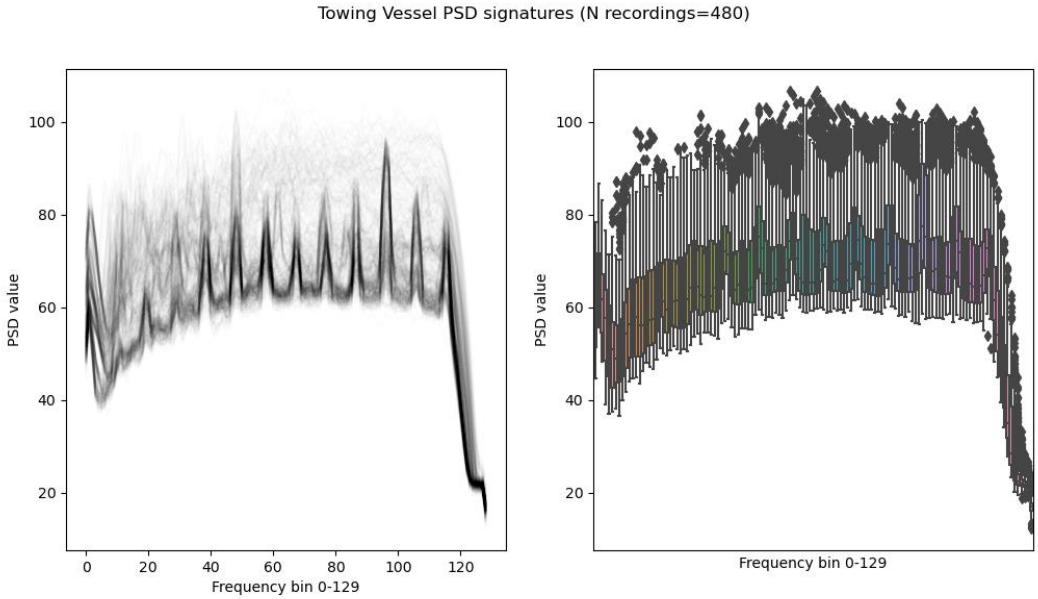

Sailing PSD signatures (N recordings=93)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

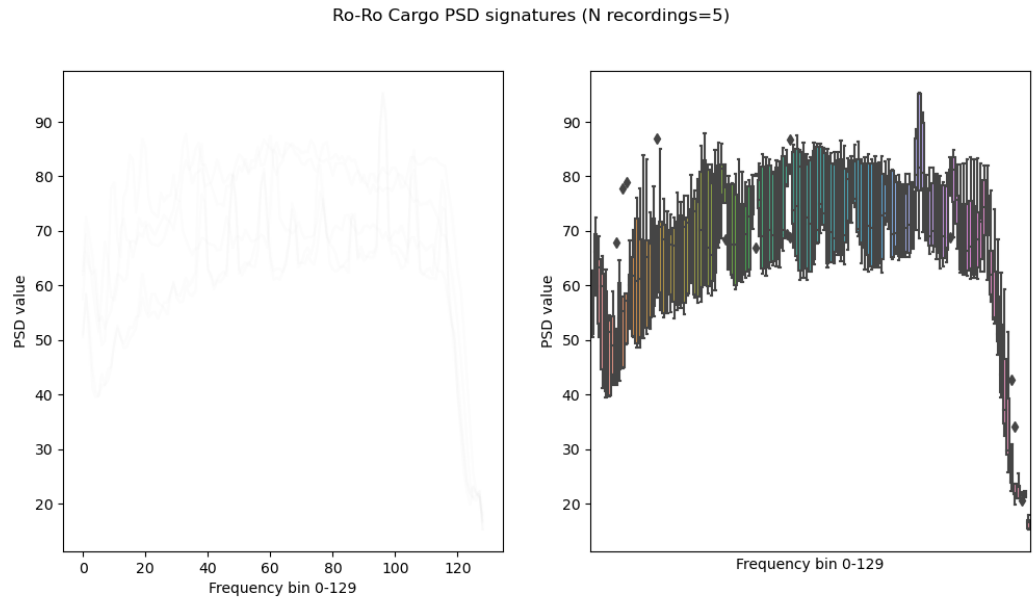

Self Discharging Bulk Carrier PSD signatures (N recordings=20)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

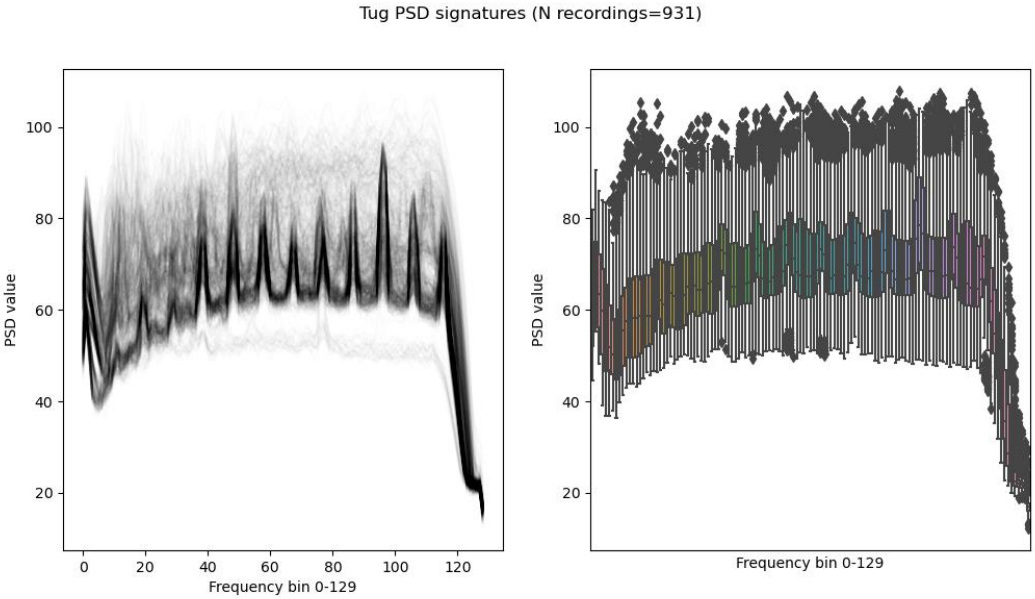

Special Vessel PSD signatures (N recordings=40)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

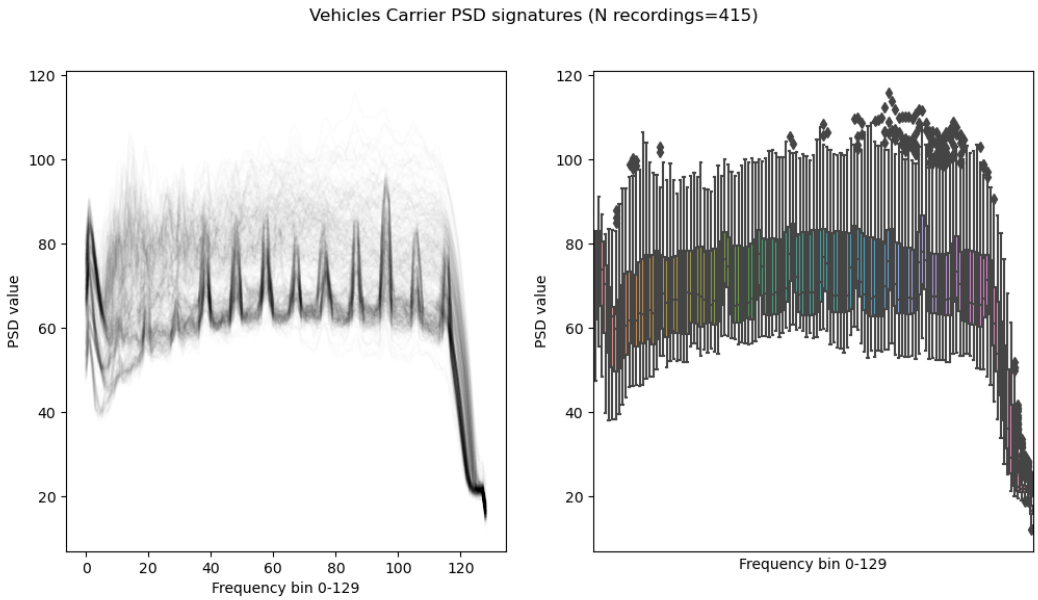

Tanker, all ships of this type PSD signatures (N recordings=109)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129


Tanker, No additional information PSD signatures (N recordings=16)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129


Towing Vessel PSD signatures (N recordings=480)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

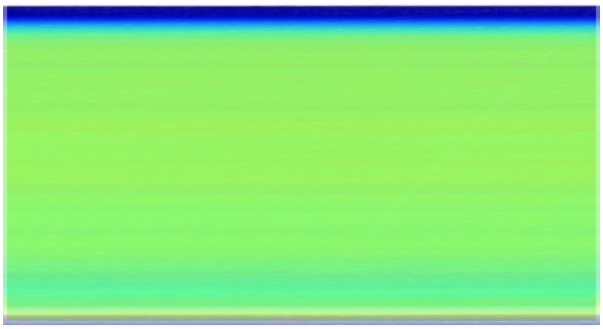
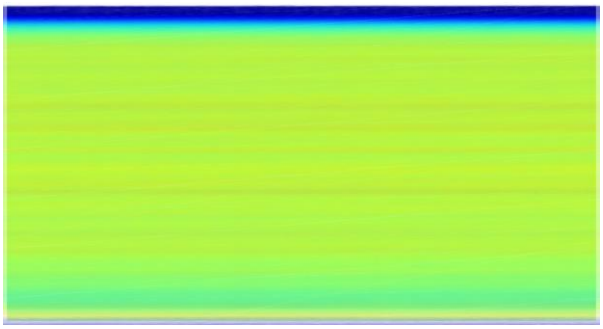
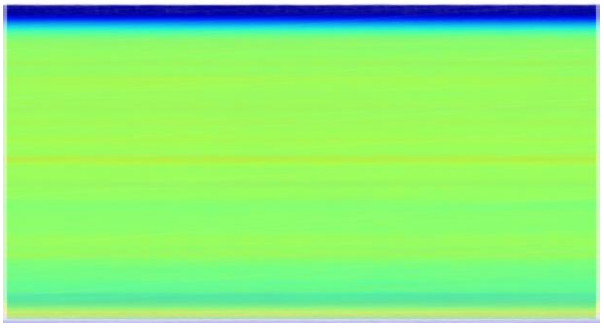
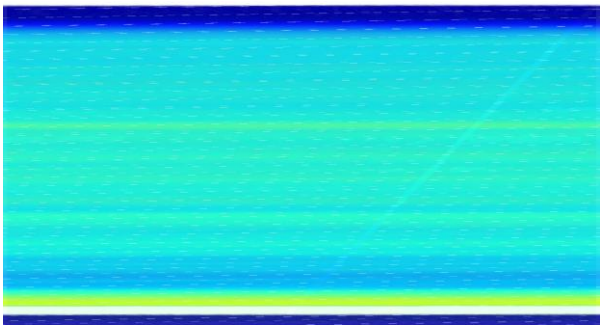

Ro-Ro Cargo PSD signatures (N recordings=5)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

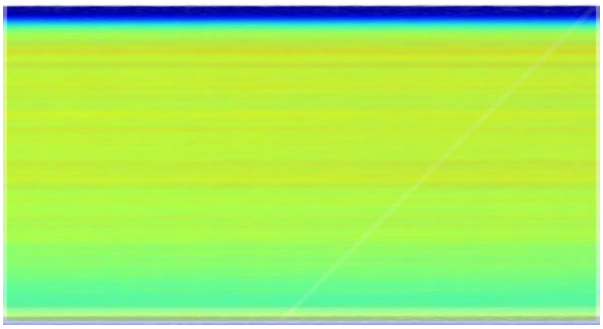
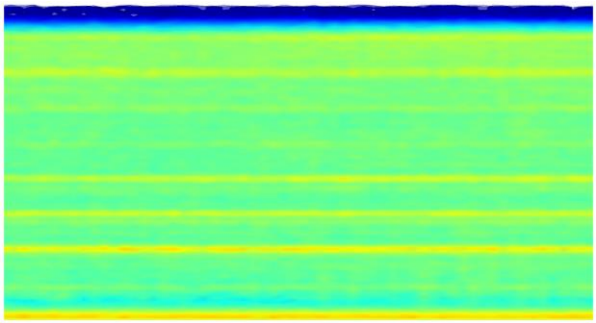

Tug PSD signatures (N recordings=931)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129


Vehicles Carrier PSD signatures (N recordings=415)
PSD value
Frequency bin 0-129
PSD value
Frequency bin 0-129

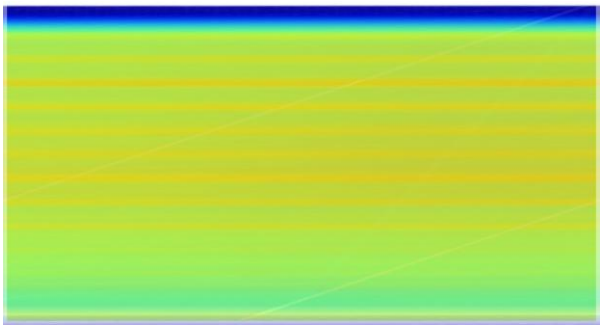
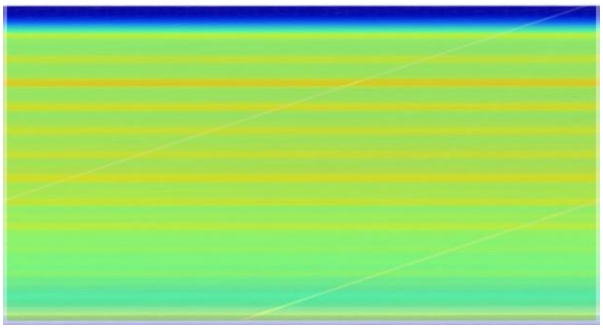
Appendix 7: Average Spectrogram: ud group vs vessel type

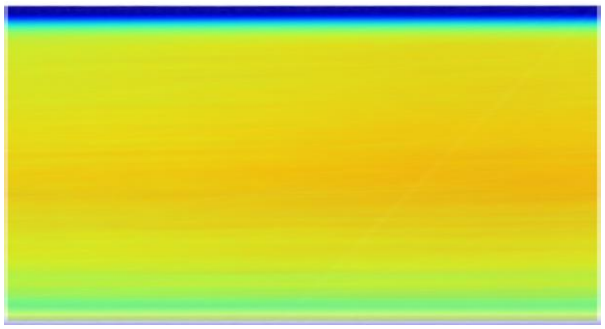
Average Image Comparison of “Vessel Type” for ud_group “Fishing type”:

ud_group	Vessel Type
 “Fishing type”	 ”Towing”
	 “Fishing”
	 “Fishing Vessel”

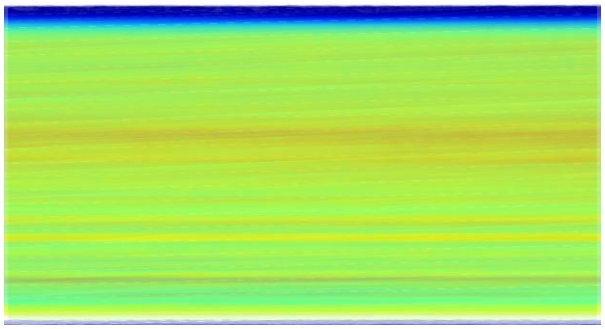
	 “Tug”
	 “Towing Vessel”

Average Image Comparison of “Vessel Type” for ud_group “Cargo”:

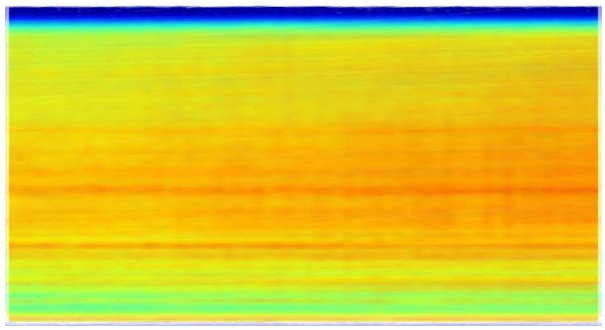
ud_group	Vessel Type
 “Cargo”	 ”Bulk Carrier”



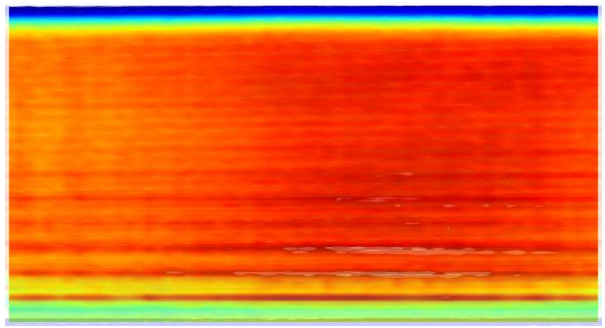
"Cargo, all ships of this type"



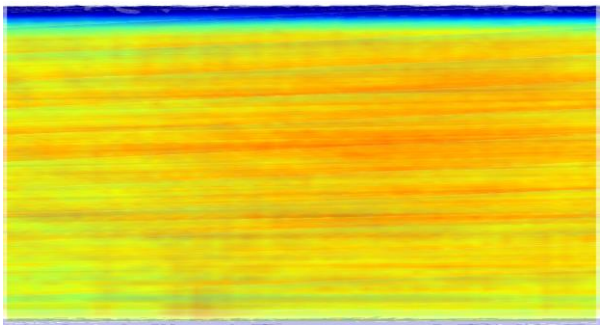
"General Cargo"



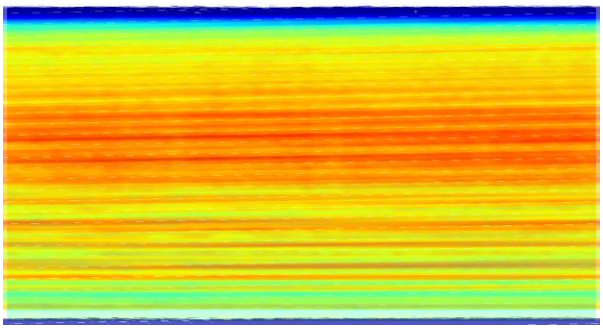
"Vehicles Carrier"



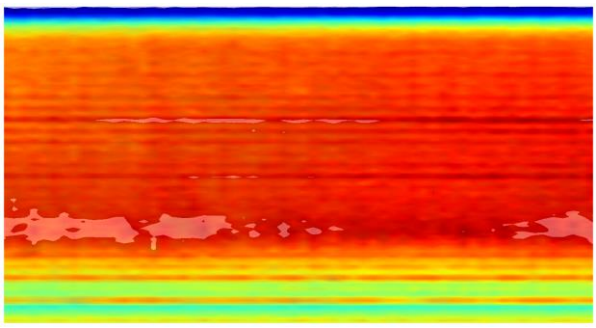
"Container Ship"



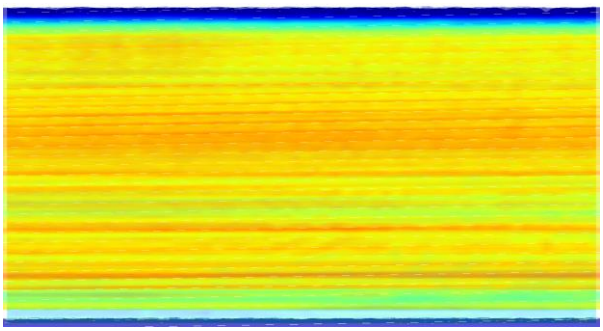
"Cargo, Hazardous category B"



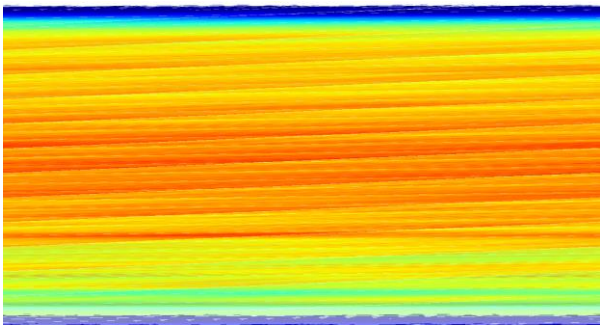
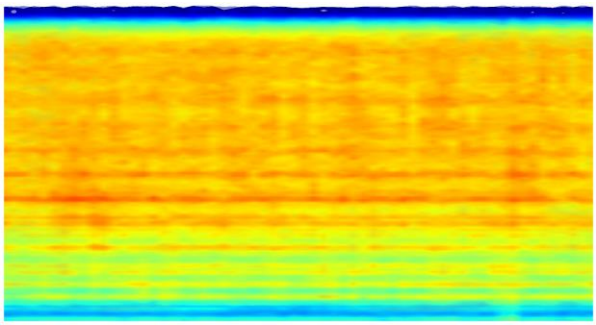
"Wood Chips Carrier"



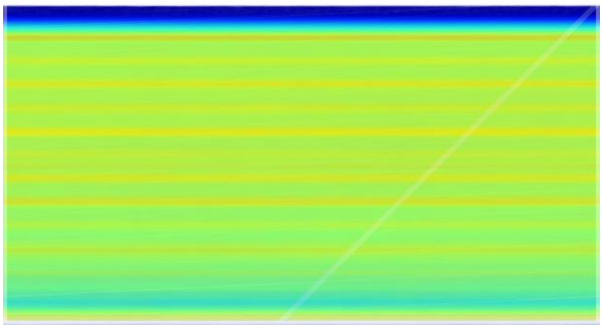
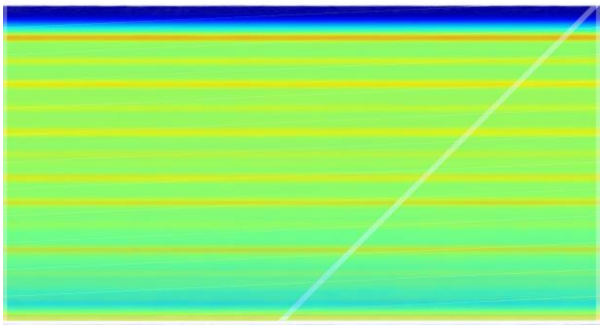
"Reefer"

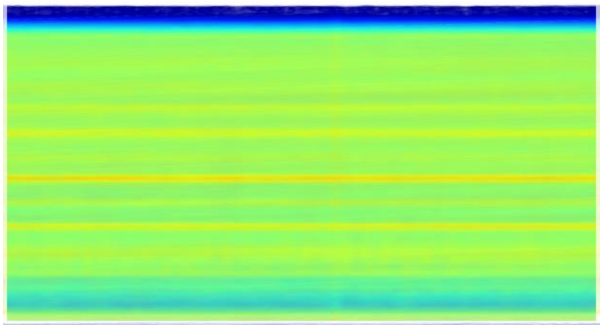


"Cargo"

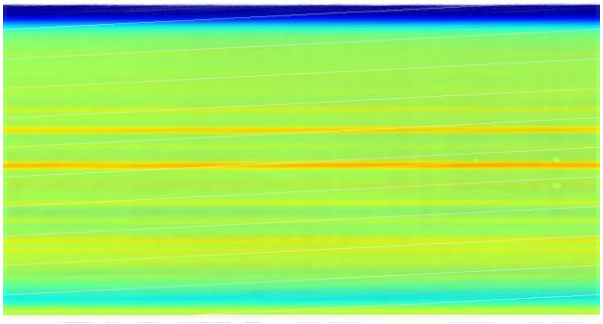
	 “Ro-Ro Cargo”
	 “Self Discharging Bulk Carrier”

Average Image Comparison of “Vessel Type” for ud_group “Tanker”:

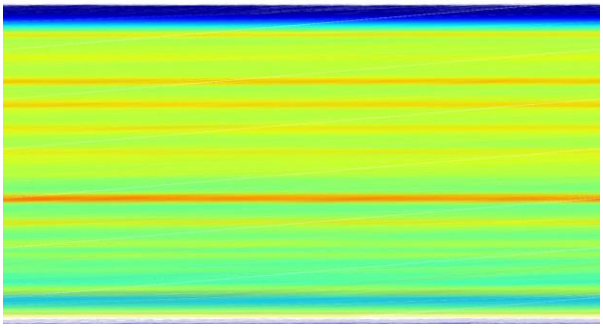
ud_group	Vessel Type
 “Tanker”	 ”Oil/Chemical Tanker”



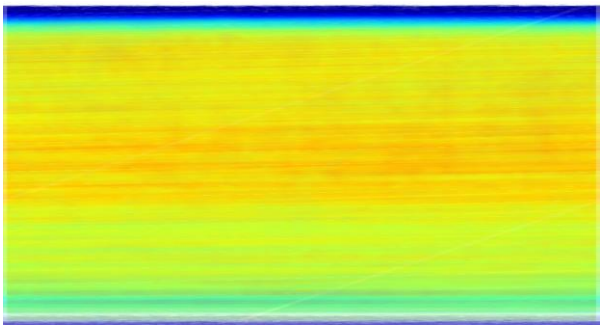
"Oil Products Tanker"



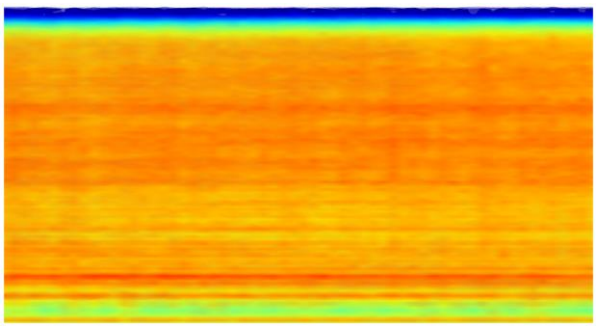
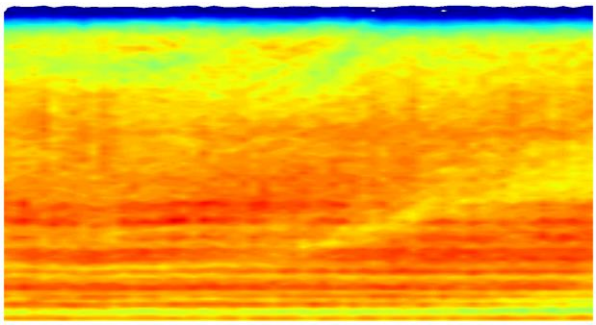
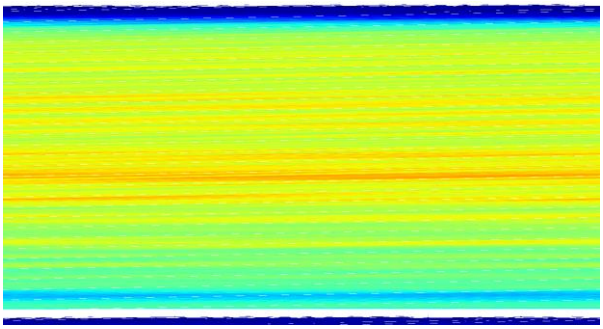
"Crude Oil Tanker"



"Chemical Tanker"



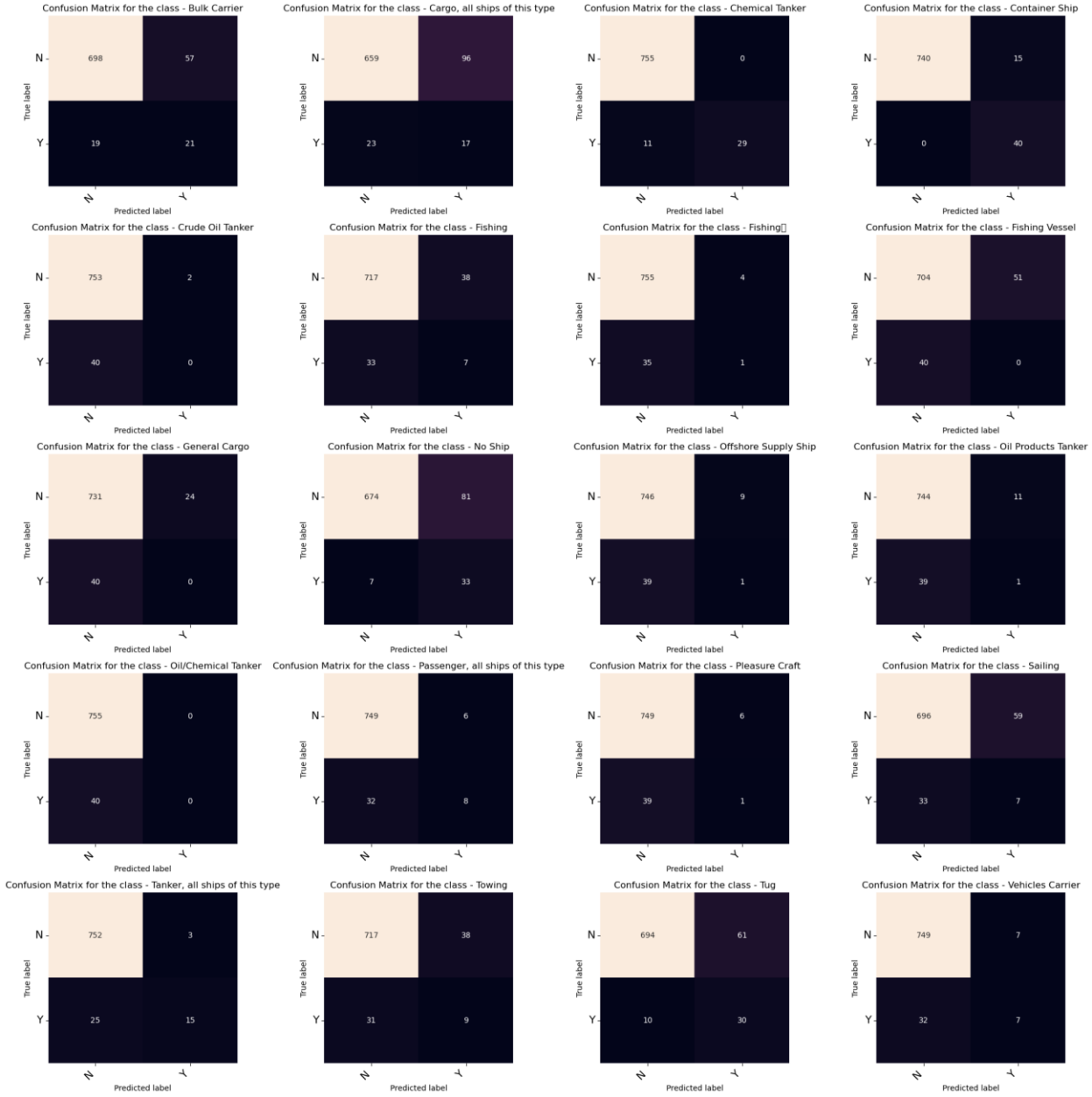
"Tanker, all ships of this type"

	 "Tanker"
	 "LPG Tanker"
	 "Tanker, No additional information"

Appendix 8: STM + ImageNet Confusion Matrices

Random Sample

Time-Stratified Sample


Confusion Matrix for the class - Bulk Carrier
698 57
19 21
Confusion Matrix for the class - Cargo, all ships of this type
659 96
23 17
Confusion Matrix for the class - Chemical Tanker
755 0
11 29
Confusion Matrix for the class - Container Ship
740 15
0 40
Confusion Matrix for the class - Crude Oil Tanker
753 2
40 0
Confusion Matrix for the class - Fishing
717 38
33 7
Confusion Matrix for the class - Fishing
755 4
35 1
Confusion Matrix for the class - Fishing Vessel
704 51
40 0
Confusion Matrix for the class - General Cargo
731 24
40 0
Confusion Matrix for the class - No Ship
674 81
7 33
Confusion Matrix for the class - Offshore Supply Ship
746 9
39 1
Confusion Matrix for the class - Oil Products Tanker
744 11
39 1
Confusion Matrix for the class - Oil/Chemical Tanker
755 0
40 0
Confusion Matrix for the class - Passenger, all ships of this type
749 6
32 8
Confusion Matrix for the class - Pleasure Craft
749 6
39 1
Confusion Matrix for the class - Sailing
696 59
33 7
Confusion Matrix for the class - Tanker, all ships of this type
752 3
25 15
Confusion Matrix for the class - Towing
717 38
31 9
Confusion Matrix for the class - Tug
694 61
10 30
Confusion Matrix for the class - Vehicles Carrier
749 7
32 7
True label
Predicted label
N
Y

Appendix 9: STM + ImageNet Classification Report

Random Sample

	precision	recall	f1-score	support
Bulk Carrier	0.84	0.62	0.71	34
Cargo, all ships of this type	0.53	0.51	0.52	35
Chemical Tanker	0.97	1.00	0.99	39
Container Ship	1.00	1.00	1.00	48
Crude Oil Tanker	0.91	1.00	0.95	39
Fishing	0.72	0.71	0.72	41
Fishing\t	1.00	0.98	0.99	44
Fishing Vessel	0.98	0.94	0.96	47
General Cargo	0.71	0.91	0.80	33
No Ship	1.00	0.70	0.83	37
Offshore Supply Ship	0.93	0.93	0.93	45
Oil Products Tanker	0.91	0.91	0.91	32
Oil/Chemical Tanker	0.92	0.92	0.92	36
Passenger, all ships of this type	0.67	0.85	0.75	39
Pleasure Craft	0.94	0.86	0.90	37
Sailing	0.88	0.88	0.88	42
Tanker, all ships of this type	0.84	0.97	0.90	37
Towing	0.51	0.57	0.54	47
Tug	0.75	0.68	0.71	44
Vehicles Carrier	0.93	0.84	0.88	44
accuracy			0.84	800
macro avg	0.85	0.84	0.84	800
weighted avg	0.85	0.84	0.84	800

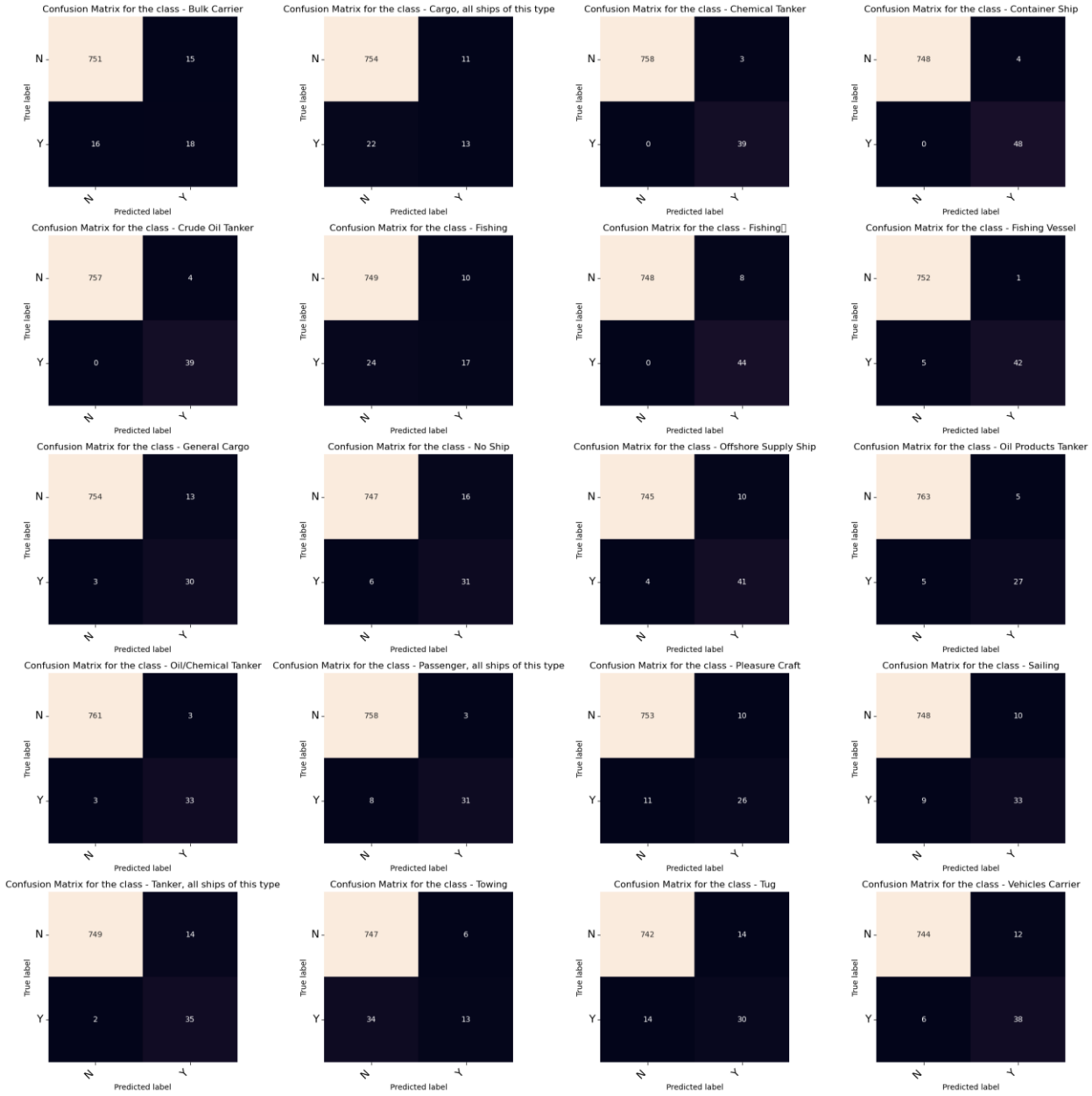
Time-Stratified Sample

	precision	recall	f1-score	support
Bulk Carrier	0.27	0.53	0.36	40
Cargo, all ships of this type	0.15	0.42	0.22	40
Chemical Tanker	1.00	0.72	0.84	40
Container Ship	0.73	1.00	0.84	40
Crude Oil Tanker	0.00	0.00	0.00	40
Fishing	0.16	0.17	0.16	40
Fishing\t	0.20	0.03	0.05	36
Fishing Vessel	0.00	0.00	0.00	40
General Cargo	0.00	0.00	0.00	40
No Ship	0.29	0.82	0.43	40
Offshore Supply Ship	0.10	0.03	0.04	40
Oil Products Tanker	0.08	0.03	0.04	40

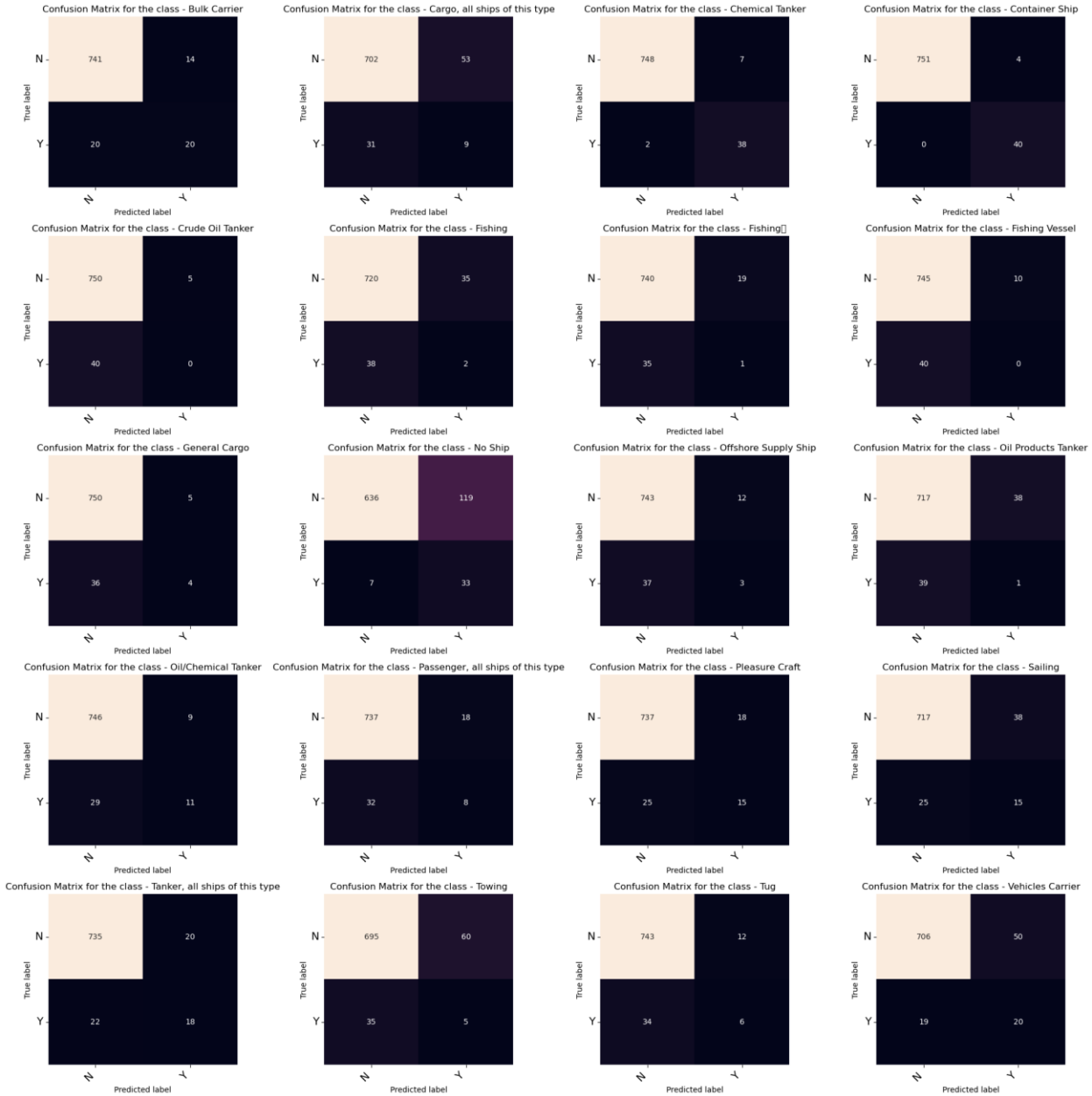
Oil/Chemical Tanker	0.00	0.00	0.00	40
Passenger, all ships of this type	0.57	0.20	0.30	40
Pleasure Craft	0.14	0.03	0.04	40
Sailing	0.11	0.17	0.13	40
Tanker, all ships of this type	0.83	0.38	0.52	40
Towing	0.19	0.23	0.21	40
Tug	0.33	0.75	0.46	40
Vehicles Carrier	0.50	0.18	0.26	39
accuracy			0.29	795
macro avg	0.28	0.28	0.24	795
weighted avg	0.28	0.29	0.25	795

Appendix 10: STM Confusion Matrix

Random Sample



Time-Stratified Sample



Appendix 11: STM Classification Report

Random Sample

	precision	recall	f1-score	support
Bulk Carrier	0.55	0.53	0.54	34
Cargo, all ships of this type	0.54	0.37	0.44	35
Chemical Tanker	0.93	1.00	0.96	39
Container Ship	0.92	1.00	0.96	48
Crude Oil Tanker	0.91	1.00	0.95	39
Fishing	0.63	0.41	0.50	41
Fishing\t	0.85	1.00	0.92	44
Fishing Vessel	0.98	0.89	0.93	47
General Cargo	0.70	0.91	0.79	33
No Ship	0.66	0.84	0.74	37
Offshore Supply Ship	0.80	0.91	0.85	45
Oil Products Tanker	0.84	0.84	0.84	32
Oil/Chemical Tanker	0.92	0.92	0.92	36
Passenger, all ships of this type	0.91	0.79	0.85	39
Pleasure Craft	0.72	0.70	0.71	37
Sailing	0.77	0.79	0.78	42
Tanker, all ships of this type	0.71	0.95	0.81	37
Towing	0.68	0.28	0.39	47
Tug	0.68	0.68	0.68	44
Vehicles Carrier	0.76	0.86	0.81	44
accuracy			0.79	800
macro avg	0.77	0.78	0.77	800
weighted avg	0.78	0.79	0.77	800

Time-Stratified Sample

	precision	recall	f1-score	support
Bulk Carrier	0.59	0.50	0.54	40
Cargo, all ships of this type	0.15	0.23	0.18	40
Chemical Tanker	0.84	0.95	0.89	40
Container Ship	0.91	1.00	0.95	40
Crude Oil Tanker	0.00	0.00	0.00	40
Fishing	0.05	0.05	0.05	40
Fishing\t	0.05	0.03	0.04	36
Fishing Vessel	0.00	0.00	0.00	40
General Cargo	0.44	0.10	0.16	40
No Ship	0.22	0.82	0.34	40
Offshore Supply Ship	0.20	0.07	0.11	40
Oil Products Tanker	0.03	0.03	0.03	40
Oil/Chemical Tanker	0.55	0.28	0.37	40

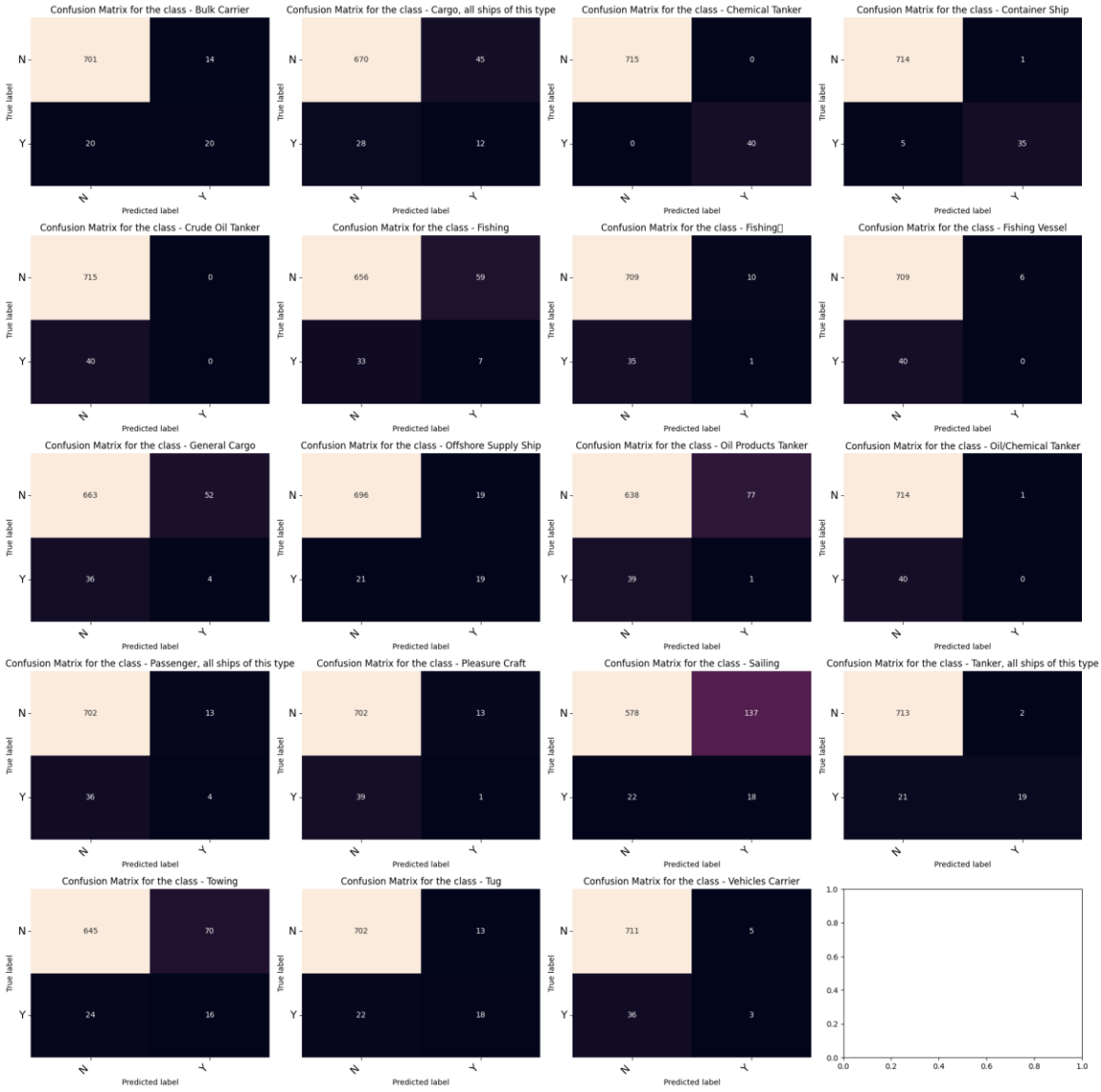
Passenger, all ships of this type	0.31	0.20	0.24	40
Pleasure Craft	0.45	0.38	0.41	40
Sailing	0.28	0.38	0.32	40
Tanker, all ships of this type	0.47	0.45	0.46	40
Towing	0.08	0.12	0.10	40
Tug	0.33	0.15	0.21	40
Vehicles Carrier	0.29	0.51	0.37	39
accuracy			0.31	795
macro avg	0.31	0.31	0.29	795
weighted avg	0.31	0.31	0.29	795

Appendix 12: CNN Confusion Matrix

Find below the details results of model on different datasets:

Random Sample

Time-Stratified Sample



Appendix 13: CNN Classification Report

Random Sample

	precision	recall	f1-score	support
Bulk Carrier	0.40	0.85	0.54	34
Cargo, all ships of this type	0.86	0.17	0.29	35
Chemical Tanker	0.97	1.00	0.99	39
Container Ship	1.00	0.98	0.99	48
Crude Oil Tanker	0.97	1.00	0.99	39
Fishing	0.50	0.34	0.40	50
Fishing	1.00	0.93	0.96	44
Fishing Vessel	1.00	0.94	0.97	47
General Cargo	0.83	0.88	0.85	33
No Ship	0.29	0.59	0.39	37
Offshore Supply Ship	1.00	0.87	0.93	45
Oil Products Tanker	0.97	0.88	0.92	32
Oil/Chemical Tanker	0.97	0.94	0.96	36
Passenger, all ships of this type	0.94	0.74	0.83	39
Pleasure Craft	0.82	0.97	0.89	37
Sailing	1.00	0.88	0.94	41
Tanker, all ships of this type	0.92	0.89	0.90	37
Towing	0.80	0.60	0.68	47
Tug	0.93	0.89	0.91	44
Vehicles Carrier	0.96	0.98	0.97	44
accuracy			0.81	808
macro avg	0.86	0.82	0.81	808
weighted avg	0.86	0.81	0.82	808

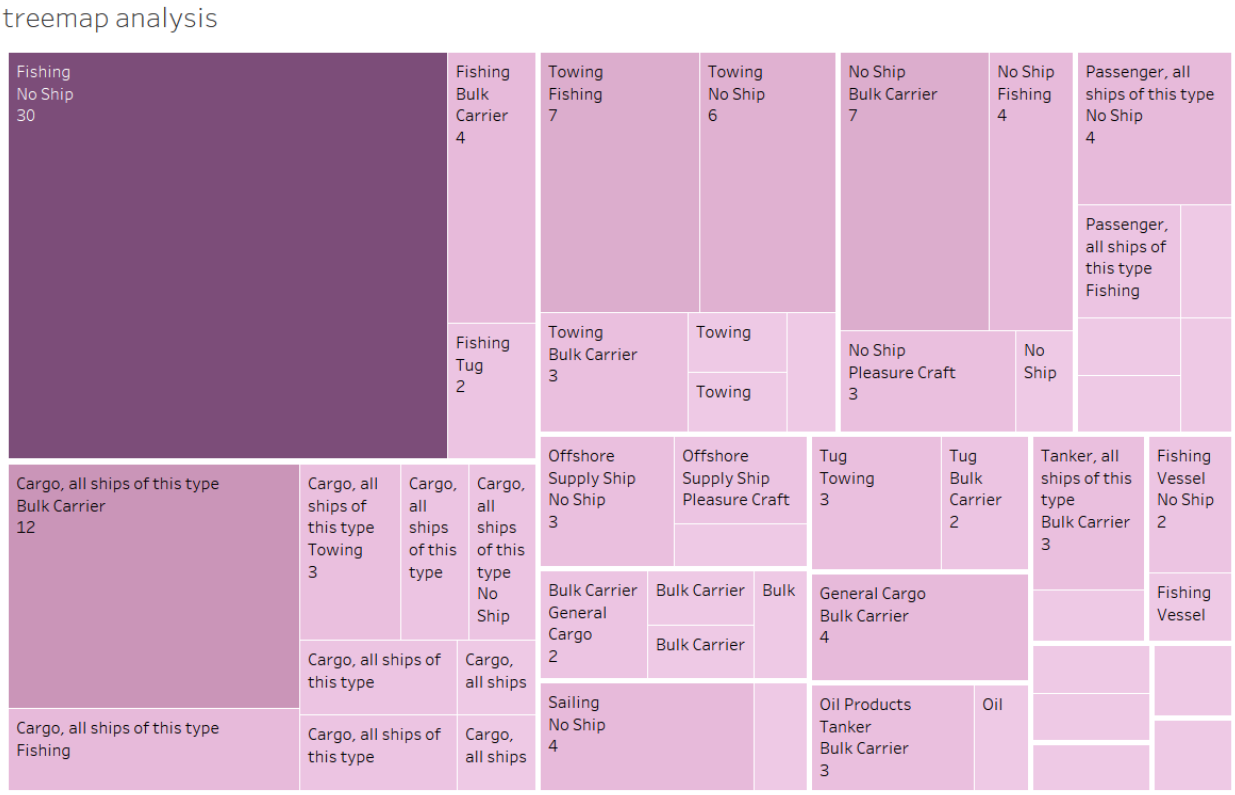
Time-stratified Sample

	precision	recall	f1-score	support
Bulk Carrier	0.59	0.50	0.54	40
Cargo, all ships of this type	0.21	0.30	0.25	40
Chemical Tanker	1.00	1.00	1.00	40
Container Ship	0.97	0.88	0.92	40
Crude Oil Tanker	0.00	0.00	0.00	40
Fishing	0.11	0.17	0.13	40
Fishing	0.09	0.03	0.04	36
Fishing Vessel	0.00	0.00	0.00	40
General Cargo	0.07	0.10	0.08	40
Offshore Supply Ship	0.50	0.47	0.49	40
Oil Products Tanker	0.01	0.03	0.02	40
Oil/Chemical Tanker	0.00	0.00	0.00	40
Passenger, all ships of this type	0.24	0.10	0.14	40
Pleasure Craft	0.07	0.03	0.04	40
Sailing	0.12	0.45	0.18	40
Tanker, all ships of this type	0.90	0.47	0.62	40
Towing	0.19	0.40	0.25	40
Tug	0.58	0.45	0.51	40
Vehicles Carrier	0.38	0.08	0.13	39
accuracy			0.29	755
macro avg	0.32	0.29	0.28	755
weighted avg	0.32	0.29	0.28	755

Appendix 14: CNN Model performance analysis

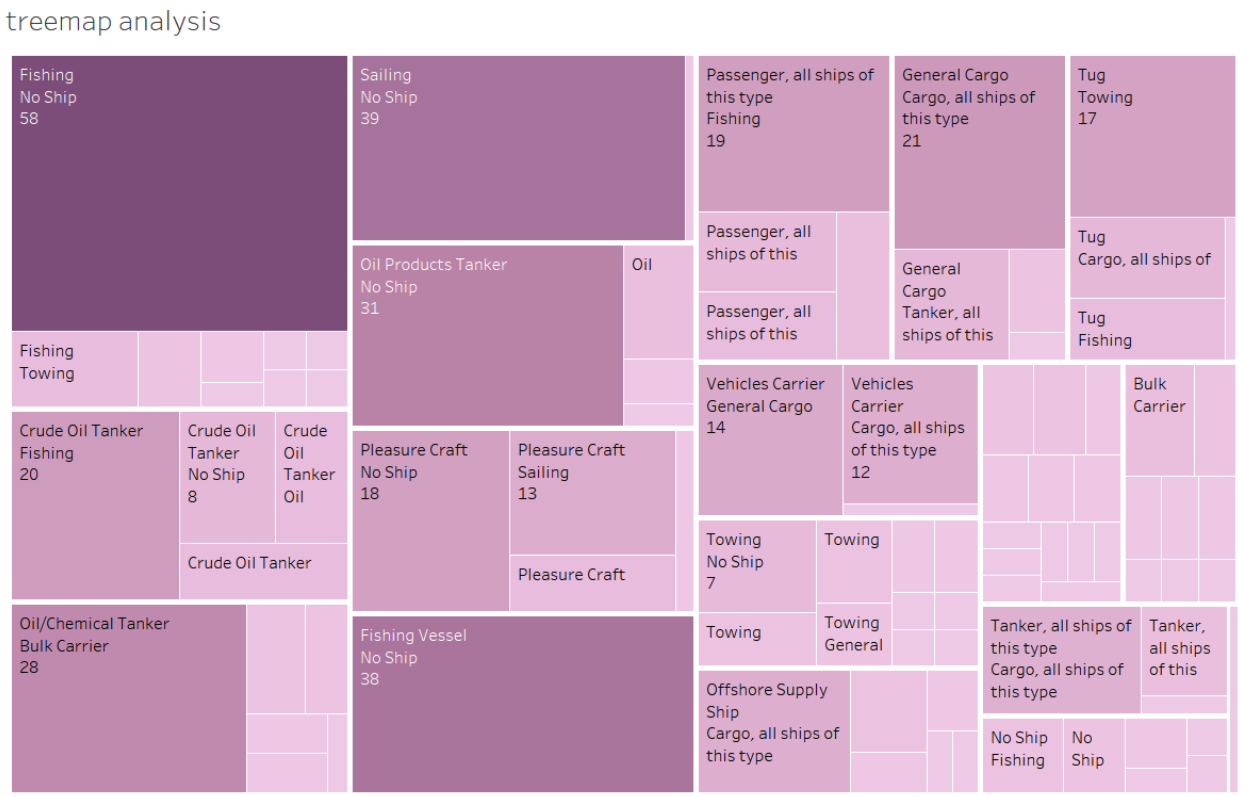
Random Sample

Displayed below is a tree map that illustrates the misclassified classes within the test sample and their corresponding incorrect classifications. The size of each rectangle corresponds to the number of misclassifications associated with each category. For instance, in the depicted graphic, the largest number of misclassifications observed was for 'Fishing' type vessels that were incorrectly classified as 'No ship' with a total number of misclassifications as 30.



Time-stratified sample:

A similar treemap analysis in time-stratified sample is shown below:



The complete analysis is available on this public tableau dashboard: https://public.tableau.com/app/profile/khirod.sahoo/viz/Noiseclassificationanalysis/treemapanalysis