ANALYSIS OF ACOUSTIC ECOLOGY OF NORTH ATLANTIC SHELF BREAK CETACEANS AND EFFECTS OF ANTHROPOGENIC NOISE IMPACTS

FY 2021 PROGRESS REPORT

PI's : Sofie Van Parijs & Danielle Cholewiak (Northeast Fisheries Science Center)

Collaborators: Alba Solsona Berga, Kaitlin E. Frasier, Jennifer Trickey, Annebelle C. M. Kok, Taylor Ackerknecht, Chelsea Field, Rebecca Cohen, John A. Hildebrand, Simone Baumann-Pickering (Scripps Institution of Oceanography), Liam Mueller-Brennan, Annamaria DeAngelis (Northeast Fisheries Science Center)

Introduction

Over 25 species of cetaceans utilize the shelf break regions of the US eastern seaboard, including several endangered species. Understanding patterns in species distribution, and the anthropogenic and environmental drivers that may impact their distribution, are critical for appropriate management of marine habitats. To better understand patterns in species distribution and vocal activity, NOAA's Northeast Fisheries Science Center and Scripps Institution of Oceanography (SIO) collaboratively deployed long-term high-frequency acoustic recording packages (HARPs) at eight sites along the western North Atlantic shelf break. This work was conducted from 2015-2019, in coordination with the Bureau of Ocean Energy Management (BOEM). Likewise, the US Navy has been monitoring the shelf break region at 3 to 4 sites since 2007. Together these combined efforts bring the total to 11 recording sites spanning the U.S. eastern seaboard, from New England to Georgia.

Data from earlier HARP recorders have been analyzed in multiple previous studies (e.g. Davis et al. 2017; Stanistreet et al. 2017, 2018). This project focuses on analyses of the new datasets collected from 2015-2019. The focus of our efforts in 2021 have been to refine species occurrence analyses, including extensive work to improve the classification algorithms for odontocetes; applying frameworks to assess impacts of anthropogenic noise on the acoustic ecology and acoustic behavior of protected species; and finalizing and publishing work on new acoustic metrics to describe species occurrence and diversity.

Objectives

The work this year was aimed at advancing the analytical components for these key objectives:

I. Continuing to improve tools for automated classification for beaked whales
II. Assessing effects of anthropogenic noise on beaked whale vocal activity
III. Assessing the prevalence of seismic survey noise along the eastern seaboard
IV. Novel broad-scale approach to assessing acoustic niche and anthropogenic contributors, and assessing the utility of new acoustic metrics

Acoustic Data Collection

Continuous passive acoustic recordings were collected along the Atlantic continental shelf break of the United States at eleven sites beginning in 2015 by both NEFSC and the U.S. Navy. The sites deployed in 2015 include Heezen Canyon, Oceanographer Canyon, and Nantucket Canyon (3 northernmost sites), and Norfolk Canyon, Hatteras, and JAX (U.S. Navy deployments). These were expanded in 2016 to include Wilmington Canyon & Babylon Canyon north of Cape Hatteras, and Gulf Stream, Blake Plateau and Blake Spur south of Cape Hatteras. (**Figure 1, Table 1**). HARPs were deployed at depths of 750-1100 m, with the hydrophones suspended approximately 20 m above the seafloor. Each HARP was programmed to record continuously at a sampling rate of 200 kHz with 16-bit quantization, providing an effective recording bandwidth from 0.01-100 kHz. HARPs include a hydrophone comprised of two types of transducers: a low-frequency (< 2 kHz) stage utilizing Benthos AQ-1 transducers (frequency response -187 dB re: 1V/µPa, ± 1.5 dB, www.benthos.com), and a high-frequency stage (> 2 kHz) utilizing an ITC-1042 hydrophone (International Transducer Corporation, frequency response -200 dB re: 1V/µPa, ±2dB), connected to a custom built preamplifier board and bandpass filter. Further details of HARP design are described in Wiggins and Hildebrand,2007.

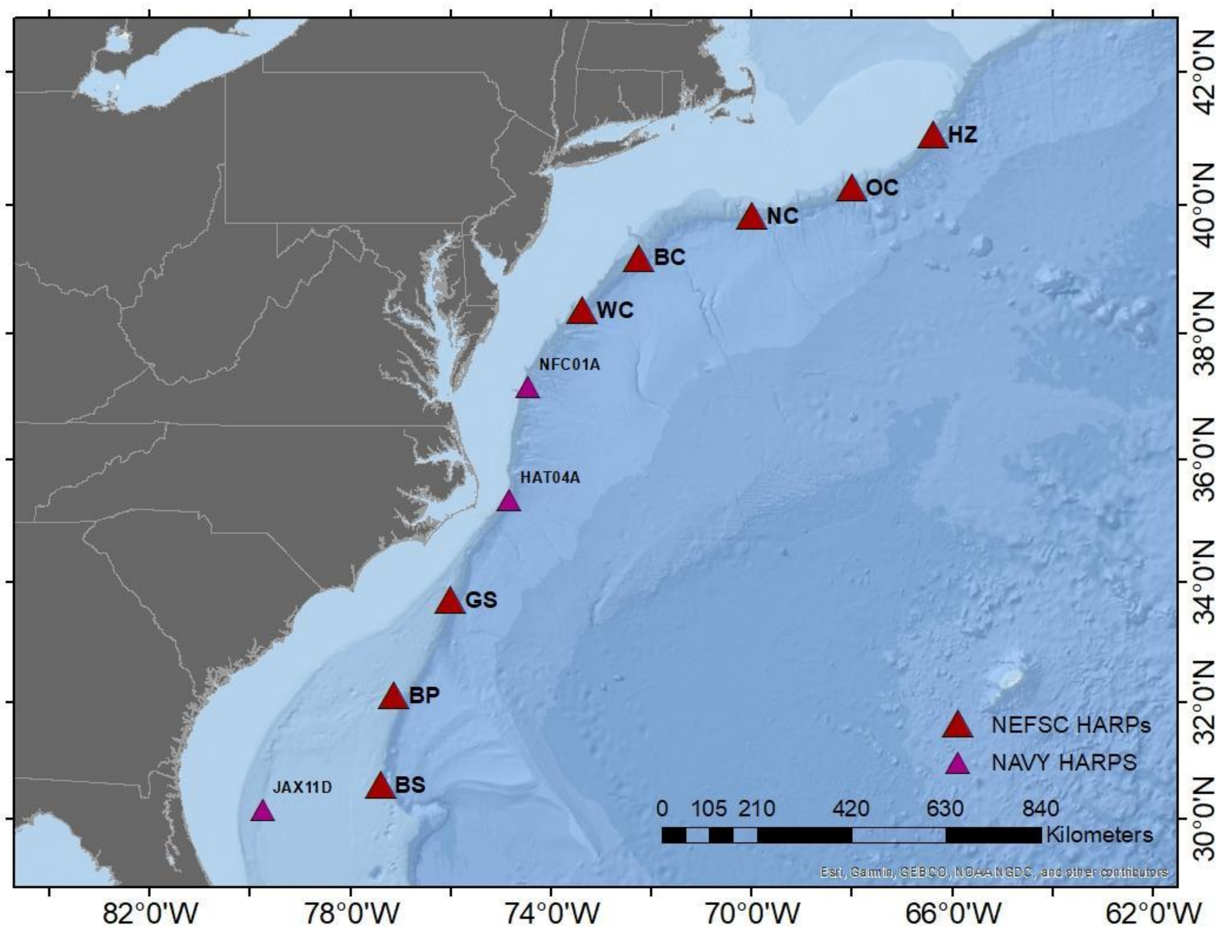

***Figure 1**. HARP deployment sites for data collected from 2015 through 2019.*

Table 1. *HARP deployment sites, recording dates and recording durations for 2015-2019. All HARPs recorded continuously at a sampling rate of 200 kHz. The first and last day of each deployment represent partial recording days.*

Site Name, Location	Recording Date Range	Latitude	Longitude	Recorder Depth (m)
WAT_HZ; Heezen Canyon	Jun 2015 - Mar 2016 Apr 2016 - Jun 2017 Jul 2017 - Jan 2018 Jun 2018 - May 2019	41.0619	-66.3515	845
WAT_OC; Oceanographer Canyon	Apr 2015 - Feb 2016 Apr 2016 - May 2017 Jul 2017 - May 2019	40.2633	-67.9862	1000
WAT_NC; Nantucket Canyon	Apr 2015 - Sep 2015 Apr 2016 - May 2017 Jul 2017 - Apr 2018 Jun 2018 - Jun 2019	39.8325	-69.9821	977
WAT_BC; Babylon Canyon	Apr 2016 - May 2019	39.1911	-72.2287	1000
WAT_WC; Wilmington Canyon	Apr 2016 - May 2019	38.3742	-73.3707	1000
NCF; Norfolk Canyon	Apr 2016 – May 2019	37.166	-74.466	1000
HAT; Hatteras	Apr 2016 – May 2019	35.584	-74.749	1100
WAT_GS; Gulf Stream	Apr 2016 - Jun 2019	33.6656	-76.0014	954
WAT_BP; Blake Plateau	Apr 2016 - May 2019	32.1060	-77.0943	945
WAT_BS; Blake Spur	Apr 2016 - Jun 2019	30.5838	-77.3907	1005
JAX; Jacksonville	Apr 2016 – Jun 2019	30.152	-79.771	750

Methods

I. Improving automated classification for beaked whales

The volume of data generated from the 11 recording sites during 2015-2019 presented a challenge for classification of beaked whales to the species level as it requires expertise and time to manually label echolocation clicks. The purpose of this effort was to design a system to streamline and automate the process of detecting and classifying beaked whale echolocation clicks using deep-learning neural networks. The classification pipeline consisted of multiple steps targeted to efficiently detect beaked whales, often challenging to detect when other species dominate the soundscape. The steps included (1) a generic detector to detect clicks above a received level threshold, (2) a discrimination phase to remove dominant non-beaked whale detections, (3) an unsupervised learning to derive clusters of distinct clicks types based on similarities in the spectral shape, and (4) a trained deep neural network to classify clusters of echolocation clicks based on spectral shape, inter-click interval, and click duration.

a. Generic impulse detection

In the first step of the workflow, an energy detector was applied using the friendly user-interface from the *SPICE-Detector Remora* (github.com/MarineBioAcousticsRC/Triton/wiki/SPICE-Detector) within the open-source data processing software package *Triton* (Wiggins et al., 2010). The energy detector was configured to band-pass the data with a five-pole Burtterworth filter from 5 to 100 kHz, and return signals with a received level ≥ 118 dB peak-to-peak re 1μPa² and durations between 30 and 1200 ms. All information on the detector settings can be found in **Figure 2**.

b. Pruning non-beaked whale detections

The generic detector returned thousands or hundreds of detections per day, depending on the site and time. Previously, a clustering method that involved unsupervised learning and neural networks (Frasier et al., 2017, Frasier, 2021) was applied directly to the output of the generic detector. This method generated a significant amount of false-positive detections and overall false labels for beaked whales. The effort during this phase was to implement a new step to remove non-beaked whale detection that dominated the recordings before the clustering phase. Then, a targeted clustering method was developed to effectively classify beaked whale detections to the species level.

An algorithm was developed to discriminate beaked whale detections based on temporal and spectral features (Baumann-Pickering et al., 2013). Clicks with peak and center frequencies of at least 32 and 25 kHz, respectively, durations of at least 355 ms, and frequency-modulated upsweeps with a sweep rate of at least 23 kHz/ms were considered potential beaked whale signals. After this initial discrimination step, an additional set of criteria was applied, requiring the waveform envelope of each click to increase over the first 0.1 ms and to remain above a 50 % energy threshold for a duration of at least 0.1 ms. Clicks meeting these criteria were further evaluated at 75-s time bins. If bins had more than seven clicks or 13 % of valid clicks, those clicks were retained for the classification task.

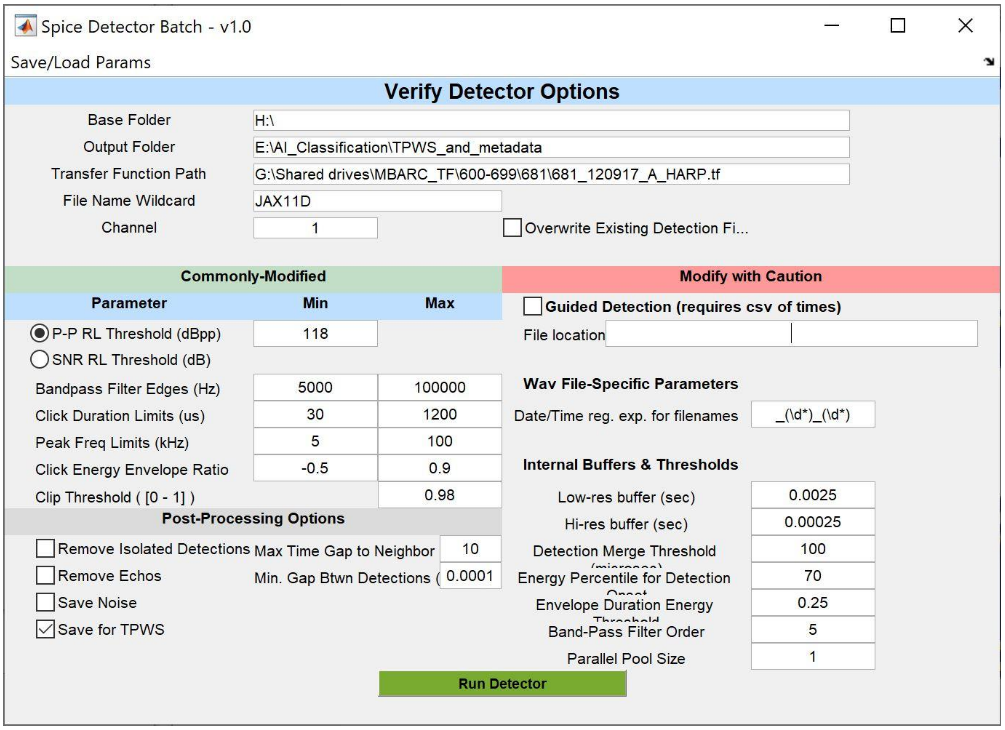

***Figure 2**. Generic impulse detector settings example from SPICE-Detector Remora interface.*

c. Unsupervised clustering of signal types in short time intervals

An unsupervised clustering approach developed by Frasier et al. (2017) was applied in successive time bins of 5 minutes to identify and group similar signals types based on similarities in the spectral shape. Spectra of all detections were truncated at 10 and 90 kHz and normalized [0, 1]. A similarity metric was computed (Frasier et al, 2017) between pairs of spectra resulting in a matrix of [0 ,1] edge weights. A network was constructed in which nodes represented individual clicks, and edge weights connections. Weak edges were pruned to reduce the size of the distance matrix input into the clustering algorithm. An edge pruning threshold at 80 % was used. This approach improved cluster formation but could result in exclusion of highly dissimilar events from any identified clusters. Therefore, different pruning thresholds were tested. Clusters of similar nodes were defined through the Chinese Whispers (CW) clustering algorithm (Biemann, 2006), using a maximum of 25 assignments iterations and a maximum network size of 40,000 clicks for each 5-min bin.

Mean spectra, waveform envelopes and the mode of inter-click interval distributions were calculated for each signal type found in each 5-min bin. ICIs were calculated for sequential clicks with each cluster and sorted in 10 ms bins up to 800 ms. A time bin could contain multiple clusters, which could represent different signal types. Clusters were formed with a minimum of ten clicks. The binned average features of all clusters were used as the input for the next classification phase. The unsupervised clustering was applied using the friendly user-interface from the *Cluster Tool Remora* (github.com/MarineBioAcousticsRC/Triton/wiki/Cluster-Tool, **Figure 3**) within the software package *Triton* (Wiggins et al., 2010).

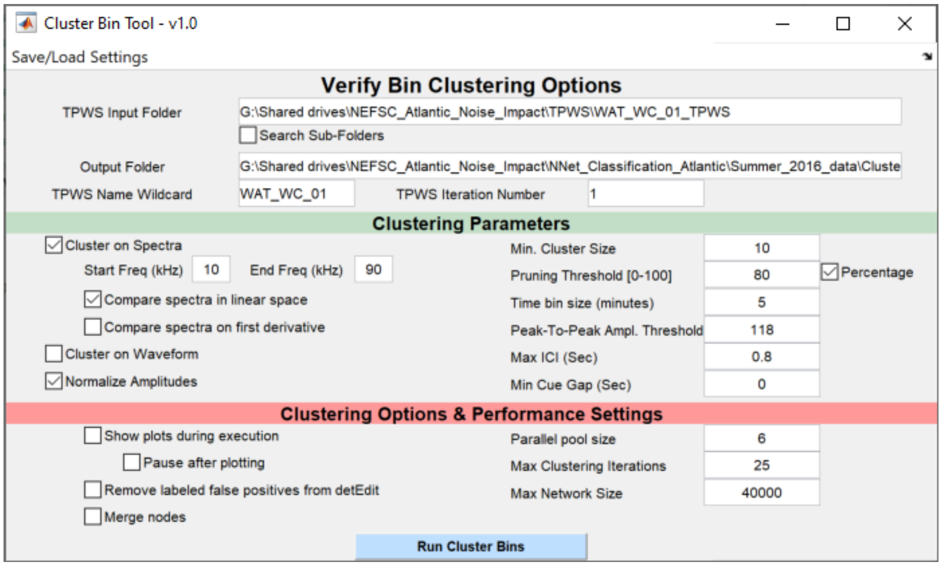

***Figure 3**. Unsupervised clustering settings phase I example from Cluster Tool Remora interface.*

d. Development of a deep neural network classifier

The purpose of this step was to develop a labeled dataset from which to train a deep neural network classifier. The steps included: (1) compiling a representative dataset of beaked whale signals and other possible signals using an unsupervised clustering method, (2) train a neural network, (3) evaluate the network performance, and (4) classify data with the trained neural network.

i. Compilation of a representative dataset

The generic energy detector was used to detect signals within periods of time with known occurrences of different beaked whale species. Detections were gathered from multiple sites and different years. To identify distinct click types, a 2-step unsupervised clustering approach (Fraiser et al., 2017) was applied. The method was used to cluster signals within each species' known occurrence times and sites independently.

In the first step, detections were divided into 5-minute time bins and grouped into similar signal types based on similarities in the spectral shape. Spectra of all detections were truncated at 10 and 90 kHz and normalized [0, 1]. A network was constructed in which nodes represented individual clicks, and edge weights connections. Clusters of similar nodes were defined through the Chinese Whispers (CW) clustering algorithm (Biemann, 2006), using a maximum of 25 assignment iterations and a maximum network size of 40,000 clicks for each 5-min bin. Multiple clusters were permitted to form per bin with a minimum of ten clicks per cluster. An edge pruning threshold at 95 % was used to remove any weakly-connected nodes. ICIs were calculated for sequential clicks with each cluster and sorted in 10 ms bins up to 800 ms. Mean spectra, waveform envelopes and the mode of inter-click interval (ICI) distributions were calculated for signal type found in each 5-min. The binned average features were stored as summary nodes for the input of the second step.

In the second step, the same algorithm was applied to cluster the set of summary nodes within each species' known occurrence times and sites independently. Similarities were again computed by comparing mean spectral shape as well as mean waveform envelope. Euclidean distances between modal ICIs were calculated to determine ICI distances values and converted into a similarity metric (Frasier et al., 2017). These two similarity scores were then combined and subsequently used in the CW clustering algorithm, allowing 25 interactions and a maximum network size of 40,000 clicks with a pruning threshold of 95 % and at least 5 nodes remaining in each resulting cluster (**Figure 4**). Due to high variability of the click duration of this species' echolocation clicks, similarities were computed only by comparing the mean spectral shape and modal ICI to assemble a representative set of signal types for this species.

After the second step of the clustering process, the binned average features grouped into distinct signal types were visually evaluated by multiple trained analysts (ASB, LMB, and AD) and assigned to possible classes of known species or sound sources. Analysts evaluated the distinct click types by inspecting plots of mean summary spectra per cluster, ICI distributions, concatenations of contributing bin-level spectra, and concatenations of contributing bin-level mean waveforms envelopes. Multiple clusters from a given site were allowed to contribute to a signal class, with the assumption that click types show substantial natural variability and the stringency of the clustering process could have led to higher cluster separation.

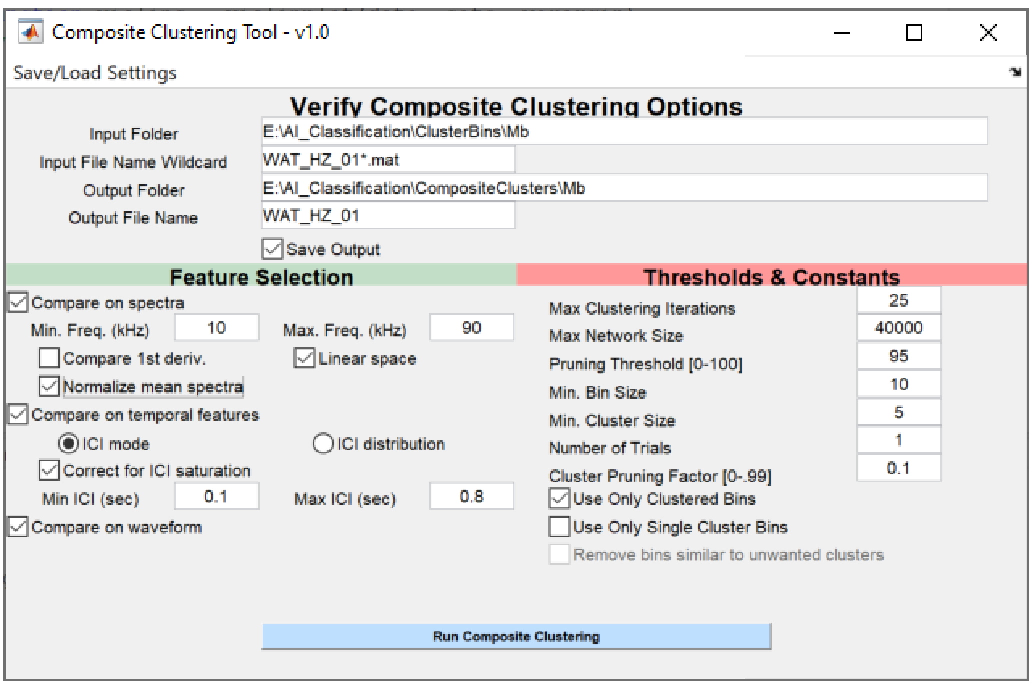

Figure 4*. Unsupervised clustering settings phase II example from Cluster Tool Remora interface.*

ii. Training a deep-learning neural network

Once detections from the representative dataset were categorized into signal classes, the dataset was used to train a supervised deep learning neural network to recognize the differences between the different classes by comparing the mean spectral shape, mean waveform envelope and modal ICI (Frasier, 2021). The neural network was trained using the friendly user-interface from the *Neural Net Tool Remora* (github.com/MarineBioAcousticsRC/Triton/wiki/NNet-Tool) within the software package *Triton* (Wiggins et al., 2010). The interface allowed the dataset to be divided into training, validation and test sets. The training set was used to train the network, the validation set to evaluate the model fit while training, and the test set to evaluate the performance of the trained network. For each signal class, detections were split first into encounters (separated by a minimum of 15 minutes without detections), and these encounters were randomly assigned to the different sets (70 % for training, and 30 % for evaluation). Several classes of the dataset had far more examples than others. To obtain a balanced dataset, 1,000 bins were randomly selected for each signal class across the training encounters and were subdivided for training and validation using an 80/20 split: 80 % for training and 20 % for validating performance. To form a test set, 500 bins were randomly selected across the evaluation encounters for each signal class.

The deep network was constructed with a similar design applied by Frasier et al. (2021). The network architecture consisted of an input layer, four 512-node fully connected layers with rectified linear unit (ReLU) (Maas, 2013) activation, 50 % dropout between layers, and a softmax output layer. Deep networks were trained with a batch size of 100 with a patience of three training epochs, after which if performance on the validation set was not improving , training ceased. A maximum of 15 epochs were allowed (**Figure 5**).

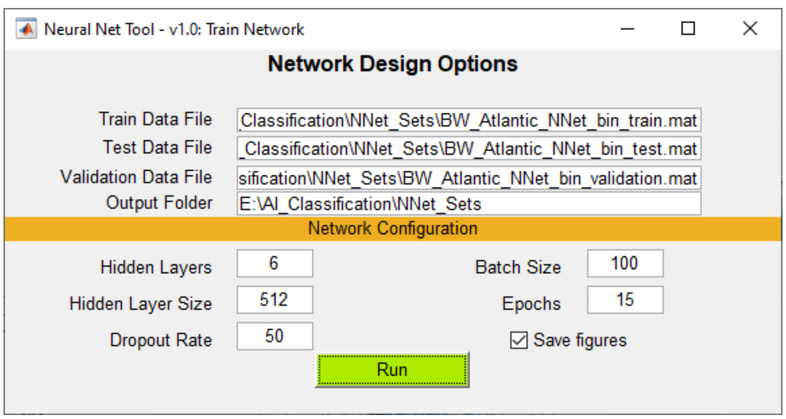

***Figure 5**. Deep network architecture settings example from Neural Net Tool Remora interface.*

e. Analysis of the summer 2016 data via manual review

As part of a separate project, the beaked whale data that was run through the first two steps of the improved classification method (impulse detector and pruning of non-beaked whale detections) collected from July 1 - August 31 2016 at all HARP sites were manually reviewed using the open-source software DetEdit (Solsona-Berga et al., 2020). All clicks that exhibited an FM upsweep and contained spectral and temporal characteristics that matched previously described click types in the literature for North Atlantic species were marked. These marked clicks were assigned a species class, and minutes containing 5+ clicks were marked as containing presence for that beaked whale species. The results from this analysis were used to test the neural net's performance and are also being combined with towed array data that were collected at the same time as part of a separate project (Atlantic Marine Assessment Program for Protected Species [AMAPPS]).

II. Assessing effects of anthropogenic noise on beaked whale vocal activity

The goal for this component of the project is to refine a statistical approach to investigate the potential impacts of mid-frequency active (MFA) sonar on beaked whale acoustic activity in the Western North Atlantic. The analyses include data for several species of beaked whales for acoustic behavioral response to sonar operations in areas with varying naval activity. The relationship between MFA sonar and the acoustic behavior of beaked whales is complex and requires the inclusion of natural temporal and spatial variability in click densities, e.g., caused by species or population-level seasonality, habitat preference, the behavioral context of echolocating, and individual variability. For this part of the project, analyses focus on the Navy HARP sites, as presence of MFA sonar is higher there than on the WAT sites.

We previously documented the progress made on data preparation, defining methods for automated identification of beaked whales to click-level and parameters to be used in statistical analysis (Van Parijs et al, 2021). The proposed statistical analysis to investigate impact entails presence/absence-level decisions in 1-min segments, which requires beaked whale data to be classified to a finer resolution of at least 1-minute granularity. The previous classification methodology included a clustering method that had a significant proportion of false detections and false classifications, which needed to be addressed. Therefore, this FY, the majority of the effort was focused on the refinement of the species-specific classifier (see above).

A short summary of the progress on this component of the project as reported previously is as follows: Automatic detection of MFA sonar was implemented using a modified version of the *silbido* detection system (Roch et al., 2011) designed for characterizing toothed whale whistles. Parameters in *silbido* were adjusted to detect tonal contours ≥ 2 kHz (in data decimated to a 10 kHz sample rate) with a signal-to-noise ratio ≥ 5 dB and contour durations > 200 ms with a frequency resolution of 100 Hz. Detections were compiled into MFA sonar events, defined as MFA sonar detections separated by more than 5 min. For each event, start and stop times were saved, as well as peak-to-peak received level (RL_{pp}, in dB) and sound exposure level (SEL). We selected generalized estimating equations (GEES) as the modeling framework for statistical data analysis. We explored the power that various explanatory variables have to the response variable, including the time of day and season, sonar presence, and sonar signal characteristics. As a first approach, we focused on two of the Navy sites (NFC, JAX) and limited the response variables to the different beaked whale species' presence in 1-min segments. To investigate the probability of beaked whale signals changing in the presence of sonar, we used a binary response variable which was equal to 1 (presence) for those 1-minute segments during which at least one signal was detected and 0 (absence) for those during which no signal was detected. This was done for the four beaked whale species click types. The explanatory covariates were defined to capture the potential effects of sonar on the response variable in various ways, e.g., the amount of sonar pings, the intensity of sonar received level at the monitoring site, the recovery time since sonar stopped. Non-sonar-related variables such as time of day, date, or year were included to account for natural variability in the response.

The current effort is to apply the classification pipeline to automatically label beaked whale acoustic events to species level in 1-min level to all Atlantic sites between 2015-2019. Progress per site and deployment is shown in the preliminary results section. The results from the Navy sites (NFC, JAX, and HAT) will be used to continue the modeling effort to examine anthropogenic effects of sonar activities.

III. Assessing the prevalence of seismic survey noise along the eastern seaboard

The goal for this component of the project is to describe and quantify the extent to which seismic airgun activity is detected along U.S. Atlantic shelf-break waters, and consider these results within the context of potential impacts on baleen whale acoustic ecology. Work on this component of the project in FY21 was two-fold. First, analyses of airgun prevalence were completed for 11 HARP sites, including both WAT and Navy sites, recording from 2016-2017. The resulting data were then used to localize all events in which corresponding airgun signals were detected across four or more hydrophones. The initial presence of airguns was automatically detected using a matched filter detector, where the time series was filtered with a 10th order Butterworth bandpass filter between 25 and 200 Hz. A cross-correlation was computed on the filtered time series; when a correlation coefficient reached a threshold of $2*10^{-6}$ above the median, a trained analyst manually verified the detections (Rafter et al. 2020). A second trained analyst reviewed the entire dataset, to identify periods with gaps in airgun activity (for example, **Figure 6**) that could be used to match signals across multiple hydrophones. Custom-written Matlab code was used to align gaps in airgun activity and estimate the bearing to the signals via time-of-arrival differences between hydrophones. Putative locations with corresponding localization errors were plotted to assess ocean basin-wide sources for airgun signals detected along the US eastern seaboard.

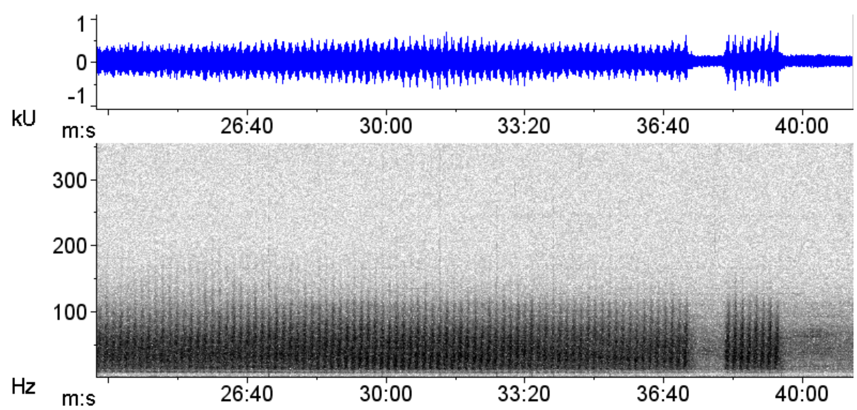

Figure 6. *Example of a "gap" and "stop" in airgun activity from Wilmington Canyon on 05/22/2016.*

IV. Novel broad-scale approach to assessing acoustic niche and anthropogenic contributors, and assessing the utility of new acoustic metrics

The goal for this component of the project is to develop and apply new techniques for visualization and rapid extraction of soundscape information from large acoustic datasets. For the former objective, in FY21 we published the following manuscript in Marine Policy: *Weiss SG, Cholewiak D, Frasier KE, Trickey JS, Baumann-Pickering SM, Hildebrand JA, Van Parijs SM. 2021. [Monitoring the acoustic ecology of the shelf break of Georges Bank, Northwestern Atlantic Ocean: New approaches to visualizing complex acoustic data](). Mar Pol. 130:104570.* This manuscript includes the summary results and data visualization from the deployment of three HARPs in 2015-2016, which were presented in our FY19 annual report. The manuscript was submitted during 2020; delays due to COVID-19 resulted in a lengthy review process, but the paper was published in August 2021.

Towards the goal of assessing the utility of acoustic metrics for the rapid soundscape assessment in long-term datasets, we pursued an approach to apply a suite of acoustic metrics, using supervised machine learning, to assess the presence and species richness (SR) of baleen whales at two sites in the western North Atlantic: the Heezen Canyon HARP dataset (2018-2019), and a MARU recorder deployed at Nantucket Shoals (2016-2018). These are referred to as the "slope" and "shelf" sites, respectively. Sound files were clipped into 1-minute segments, and we performed stratified random sampling over the entire dataset to select files to constitute the training dataset for the model. We quantified species presence or absence, as well as the number of species acoustically present in a 1-min file (SR), using aural and visual review of spectrograms. We computed 21 different acoustic metrics for every acoustic file in each training set over the full bandwidth 0–1,000 Hz. We used random forest classification models to discriminate between the acoustic presence and absence of the different species comprising the acoustic community at each site, and to evaluate the discrimination potential of the acoustic metrics. Finally, the trained model was run on a full 12 months of acoustic data from the Nantucket Shoals site, and compared with our LFDCS detection software and manual verification outputs.

Preliminary Results

I. Improving automated classification for beaked whales

a. Development of a deep neural network classifier

A representative dataset for the North Atlantic shelf break to train a deep-neural network was collected from multiple sites and different years. The total amount of binned average features of clustered clicks varied among classes. Classes had signals from more than three sites to increase variability of the representative features within each signal class (**Table 5**). Twelve signal classes were identified which were attributed to six different species of beaked whales: *Mb* - Blainville's beaked whale (*M. densirostris*), *Zc* - Cuvier's beaked whale (*Z. cavirostris*), *Me* - Gervais' beaked whale (*M. europaeus*), *Mm* - True's beaked whale (*M. mirus*), *Mb* - Sowerby's beaked whale (*M. bidens*), and *BWG* - the unknown *Mesoplodon sp.* designated as Beaked Whale Gulf. Three signal classes were attributed to non-beaked whale species representing *Kogia sp.*, *Gg* - Risso's dolphin (*G. griseus*), and a large class categorizing most of the delphinid click types. One class was attributed to different pings of echosounders. Three signal classes were categorized to generic noise sources - one representing noise from frequencies above 50 kHz (labeled as *High-noise*), another with frequency between 20-50 kHz (labeled as *Mid-noise*), and a class with signals with frequencies below 20 kHz (labeled as *Ship-Pm*) containing possibly shipping noise and sperm whale echolocation clicks.

Unique training binned average features for beaked whale species classes ranged from approximately 1,000 bins for *Mb*, *Md* and *Mm*, as low as 29 bins for *BWG,* to more than 8,000 bins for *Me* and *Zc*. For the other non-beaked whale signal classes, the total amount of bins varied greatly, with the highest amount of examples for the delphinid class and the *Ship-Pm* class (**Table 5**).

Table 5*. Number of binned averages of clustered clicks per 5-minute bins within each signal class per deployment.*

	BWG	Mb	Md	Me	Mm	Zc	Gg	Delphinid	Kogia	Ship-Pm	Echosounder	Mid-Noise	High-Noise
Heezen													
WAT_HZ_01		1121			49	649	146	3295		2262		170	
Nantucket Canyon													
WAT_NC_01		23			152	3		312		91			
Norfolk Canyon													
NFC_A_02		335	2		302	123		1604		393		55	19
NFC_A_03		369	4		385	472		1189		487		24	28
Hatteras													
HAT_B_03_01			5	120		2758		15367		1809	92	85	
Bear Seamount													
WAT_BR_01					227	232	45	490		547	170		
Blake Spur													
WAT_BS_01	7		558	765		187		145	18	76		21	
Jacksonville													
JAX11D			3	28				121					
JAX_D_15		4											
Bermuda													
BM01A			24			2		4		49			
BM02A			78	2		9				16			
BM_A_03			297			14		128		183			
USWTR													
USWTR06E				1653		15		532	19	796			
USWTR07E			2	1420		5	66	88		400			
Gulf of Mexico													
GofMX_DT01				571		604				228		13	
GofMX_DT02				520		1117	53	50		420		86	
GOM_DT_10	16		3	1468		2726	26	1012		252		63	
GofMX_GC01	6		6	166				92		33			
GofMX_GC02				76				16		145			
GofMX_GC03			4	142		37		78		215			
GofMX_GC04				155		74		24		253			
GofMX_GC05			40			68							
GofMX_HH01				996		1738	1703		13	586		107	
GofMX_MC05				160		40		211		318			
GofMX_MC06				445		44		558		852		12	
Binned averages	**29**	**1852**	**1035**	**8687**	**1115**	**10917**	**2039**	**25316**	**50**	**10411**	**262**	**636**	**47**

A balanced dataset for training the neural network was selected with a 1,000 bin examples per signal class. Classes with more than 1,000 bin examples were subsampled such as the delphinid class, or the *Zc* and *Me* class. For the minority classes such as *BWG* and *Kogia* examples were resampled to reach the required number of examples to train the network (**Figure 7**).

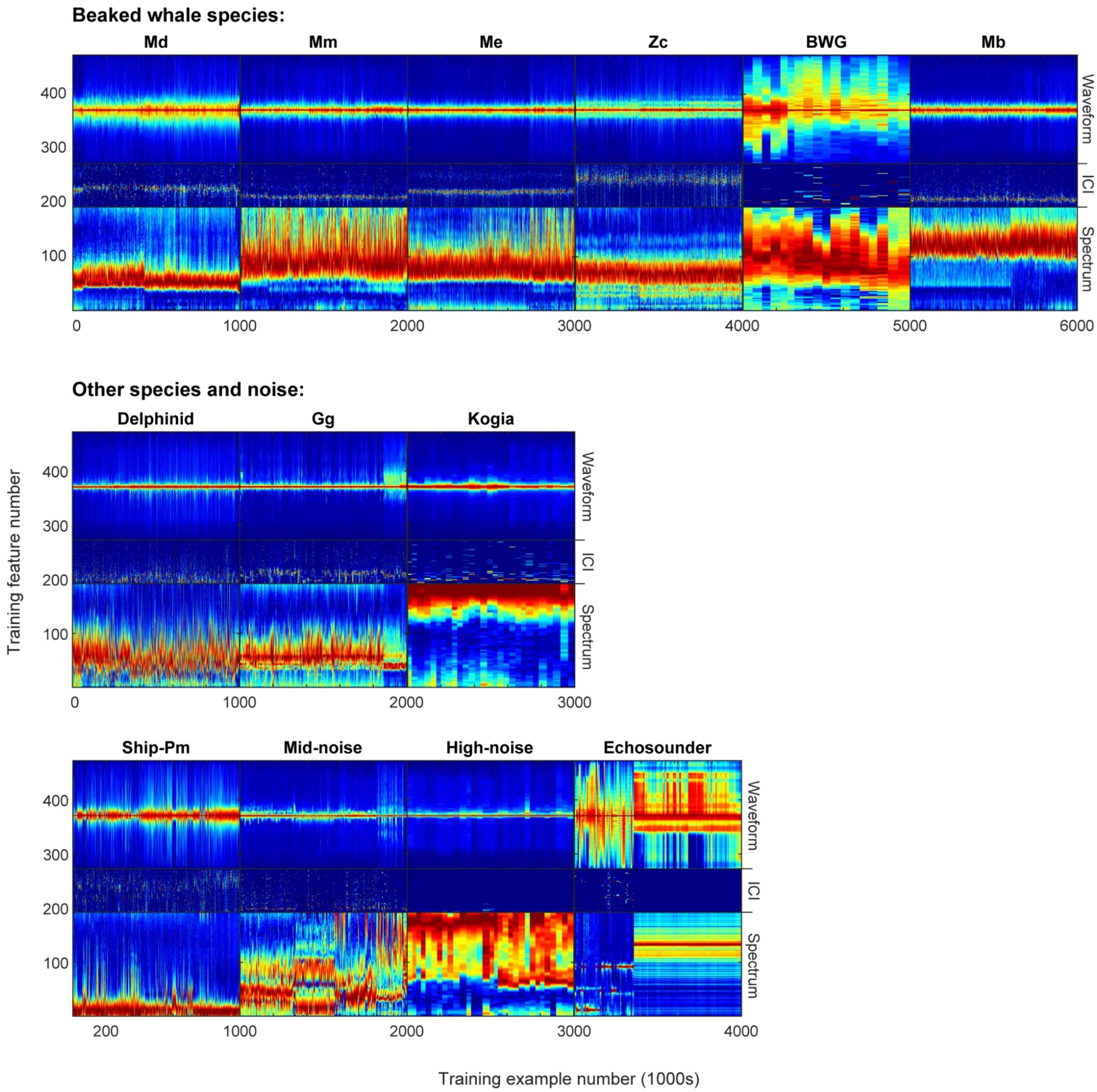

Figure 7. Binned average features of the balanced dataset for neural network training, consisting of concatenated mean spectra, ICI distributions, and mean waveforms of 1,000 bins per signal class. Features are normalized [0,1] prior to being fed into the neural network.

The learning algorithm required 15 epochs to train the neural network, passing through the entire training dataset at every epoch. The network achieved 98 % classification overall accuracy on the balanced test set with 500 bin examples per signal class. Confusion was very low for all beaked whale classes, with accuracy rates above 99 % (**Figure 8**). Performance on the non-beaked whale classes was lower, as expected since effort was to gather enough non-beaked whale classes to reduce spurious assignments of unrepresented non-target signals to beaked whale categories. BWG, Sowerby's, and Blainville's beaked whales were 100 % correctly classified. Little confusion was observed between Gervais' and True's beaked whale, with 2 binned examples of Gervais' classified as True's. Other 2 binned examples of Gervais's were classified as delphinid, and 2 binned examples of delphinids were misclassified as True's. Approximately

1 % of binned examples of Cuvier's beaked whales were misclassified as delphinids. Some confusion was observed with *Mid-noise* and delphinid examples that were classified as Cuvier's beaked whales, which some of those examples could have been incorrectly labeled by the unsupervised clustering process. Therefore, the *Mid-noise* class could contain some true Cuvier's beaked whale examples. Performance was the lowest with the delphinid class. This class contained multiple species of delphinids and after being resampled to create a balanced dataset, this could have resulted in a brittle class characterization. Nonetheless, the unsupervised clustering approach does not ensure the creation of perfect training and testing sets, and some of the binned examples could have been correctly classified by the neural network although were clustered pertaining to a distinct signal type.

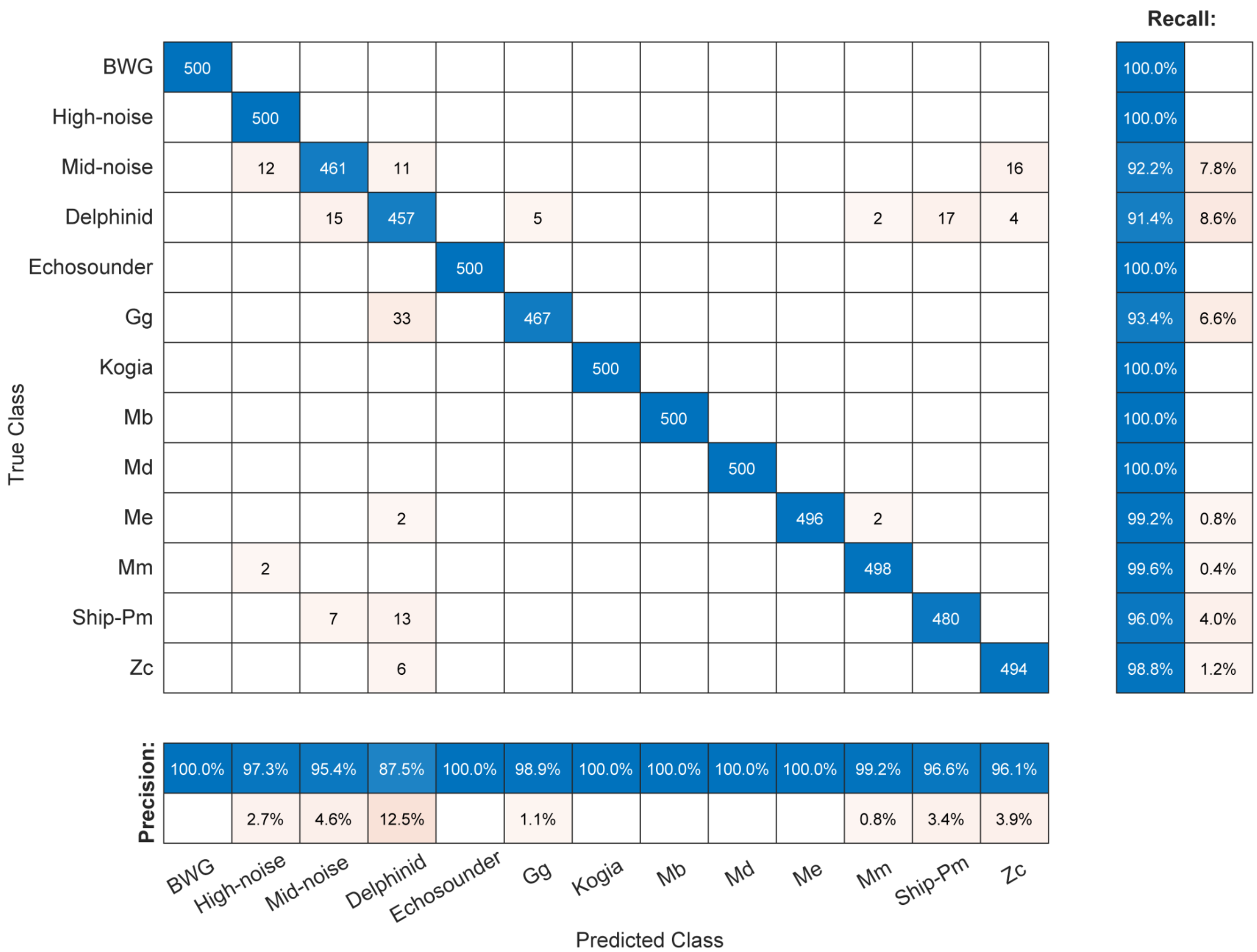

Figure 8. Confusion matrix for the deep neural network classification of the balanced test set, consisting of 500 examples per signal class. Values in the matrix indicate total number of 5-min bins classified.

b. Testing the classification pipeline with an independent dataset

The deep neural network was found to achieve high overall performance on the representative dataset compiled from periods of time with known beaked whale presence. The classification pipeline was also tested on a small dataset of several deployments from July to August of 2016, which was previously labeled for another purpose (described in section I.e). This created an opportunity to test the performance of the classification pipeline against a manually labeled dataset and the performance of the previous neural network developed in previous fundings for odontocetes in general.

For the unsupervised clustering process, different thresholds of pruning were tested to improve the formation of clusters of similar signals. The first configuration that was applied was the same configuration used to compile the representative dataset to train the neural network. This configuration had a high pruning threshold of 95 %. This pruning led to a large number of isolated clicks of beaked whales that were off-axis, with a less characteristic spectral shape (**Figure 9**). Lowering the pruning threshold to 80 % improved the inclusion of those clicks while still balancing the distinction of different signal types.

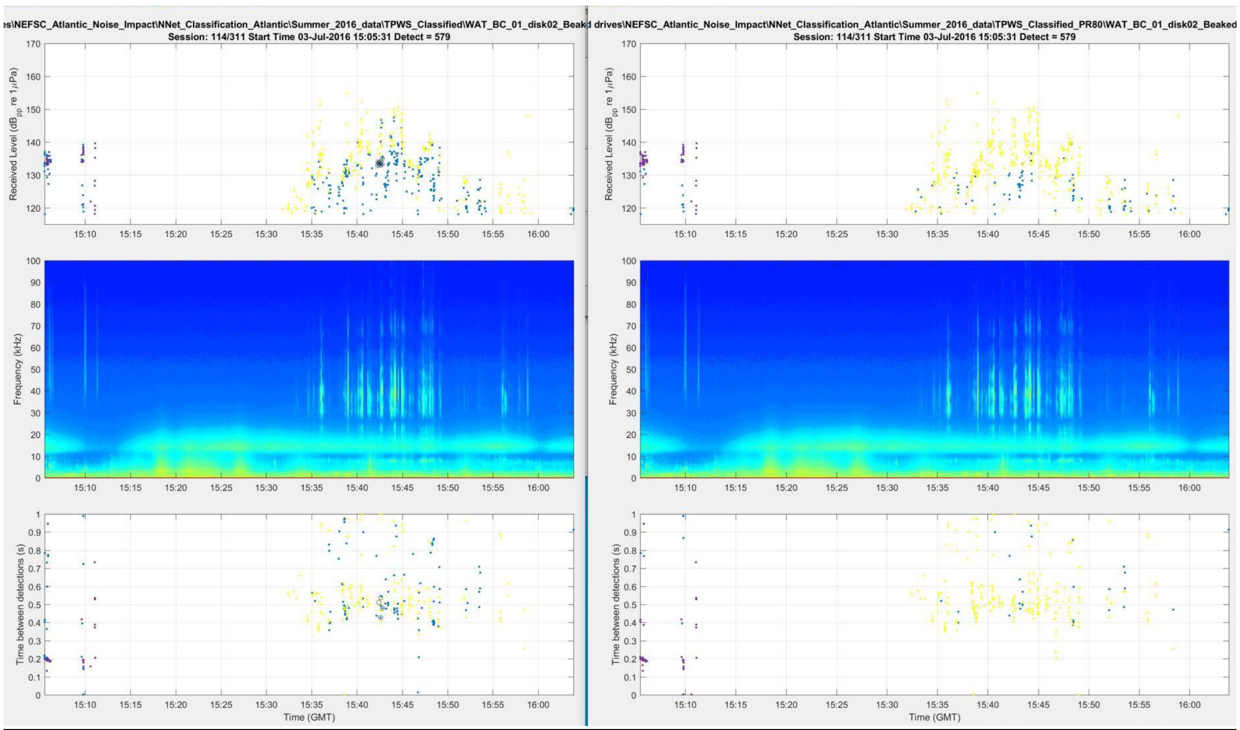

Figure 9. Example of Cuvier's beaked whale clicks clustered in yellow and not clustered in blue using different pruning thresholds: left panel using a 95 % pruning threshold and on the right panel using a 80 % pruning threshold.

The small dataset from summer 2016 was classified using the trained deep neural network. However, only results from the Babylon Canyon (BC) site are shown. The manual labels were considered for this purpose as the true species classes and compared with the classification given by the neural network. The results from a previous trained neural network targeted to classified odontocetes in general were compared with the manual labels, and differences between both neural networks were evaluated. To facilitate comparison, the developed neural network during this funding will be referenced from now on as the "BW neural network", as it was targeted to improve beaked whale classification, and the other neural network as "DE neural network". For cases in which no labels were given during the manual labeling, a label of "No label" was applied to account for the missing bins. Similarly, if no cluster formed, this was counted as a "No label" by the classifier for purposes of computing accuracy metrics.

The BW network had a high prediction rate for beaked whale events (**Figure 10**), with 70 % of the 5-minute bins correctly predicted for all beaked whales species present at site BC. The majority of bins not labeled had less than 10 clicks (**Figure 11**). These cases were not directly available to the classifier during the labeling stage due to the clustering choices in prior steps. The unsupervised clustering method was set to retain clusters of signals of at least 10 clicks within a 5-min time bin. Examination of the misclassified events of the beaked whale classes revealed that a large number of the incorrectly labeled events had a low classification probability (**Figure 11**). Examination of the labeled bins by the neural network that did not have a label from the manual dataset revealed a large number of bins correctly assign to the species signal type (**Figure 12**). Therefore, those bins were missed in the manual dataset. The high probability (> 0.8) of Sowerby's and True's beaked whale labels appeared to be of those species, as well as those labels with a high probability (> 0.99) of Cuvier's beaked whale (**Figure 12**). In contrast, most of those bins labeled as BWG, Gervais' and Blainville's beaked whales (not labeled in the manual dataset) had lower probabilities. Therefore, discarding low probability labels, the precision of the network could increase considerably without severely affecting recall.

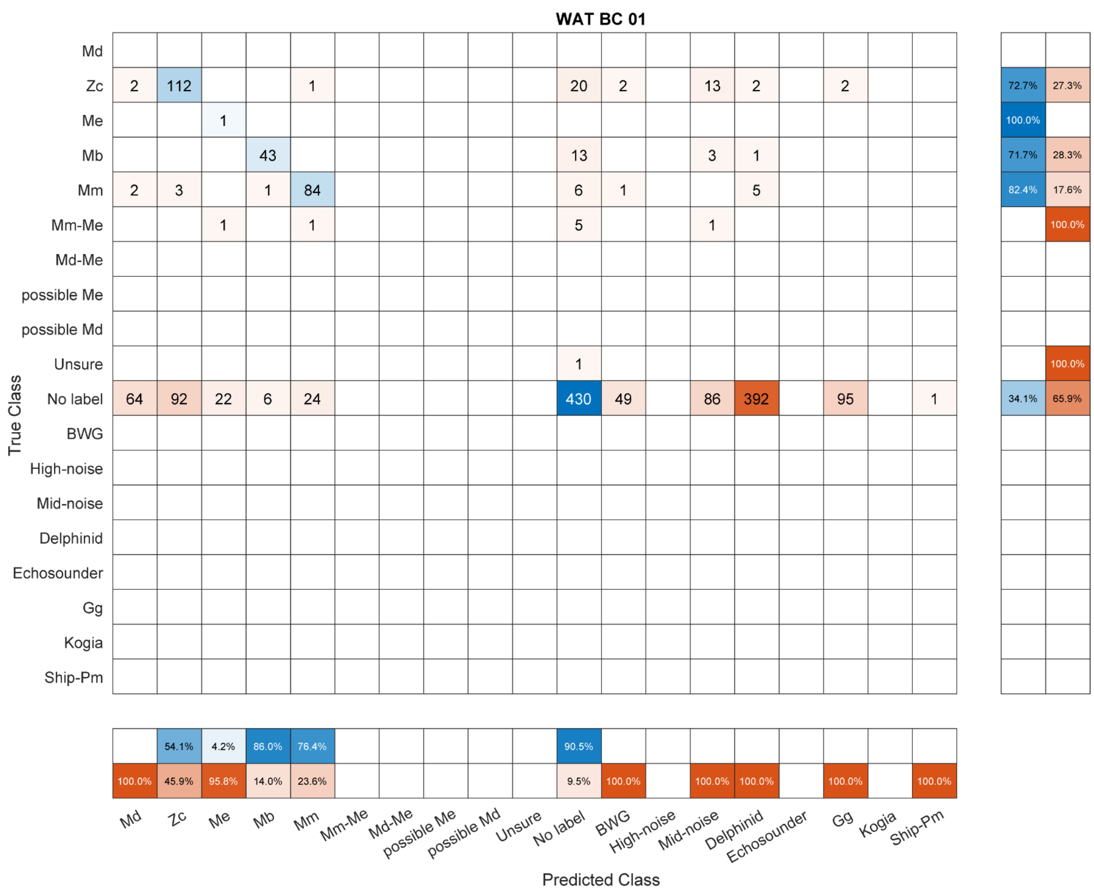

Figure 10. Confusion matrix for the BW deep neural network classification of site BC detections during summer months. Values in the matrix indicate total number of 5-minute time bins classified.

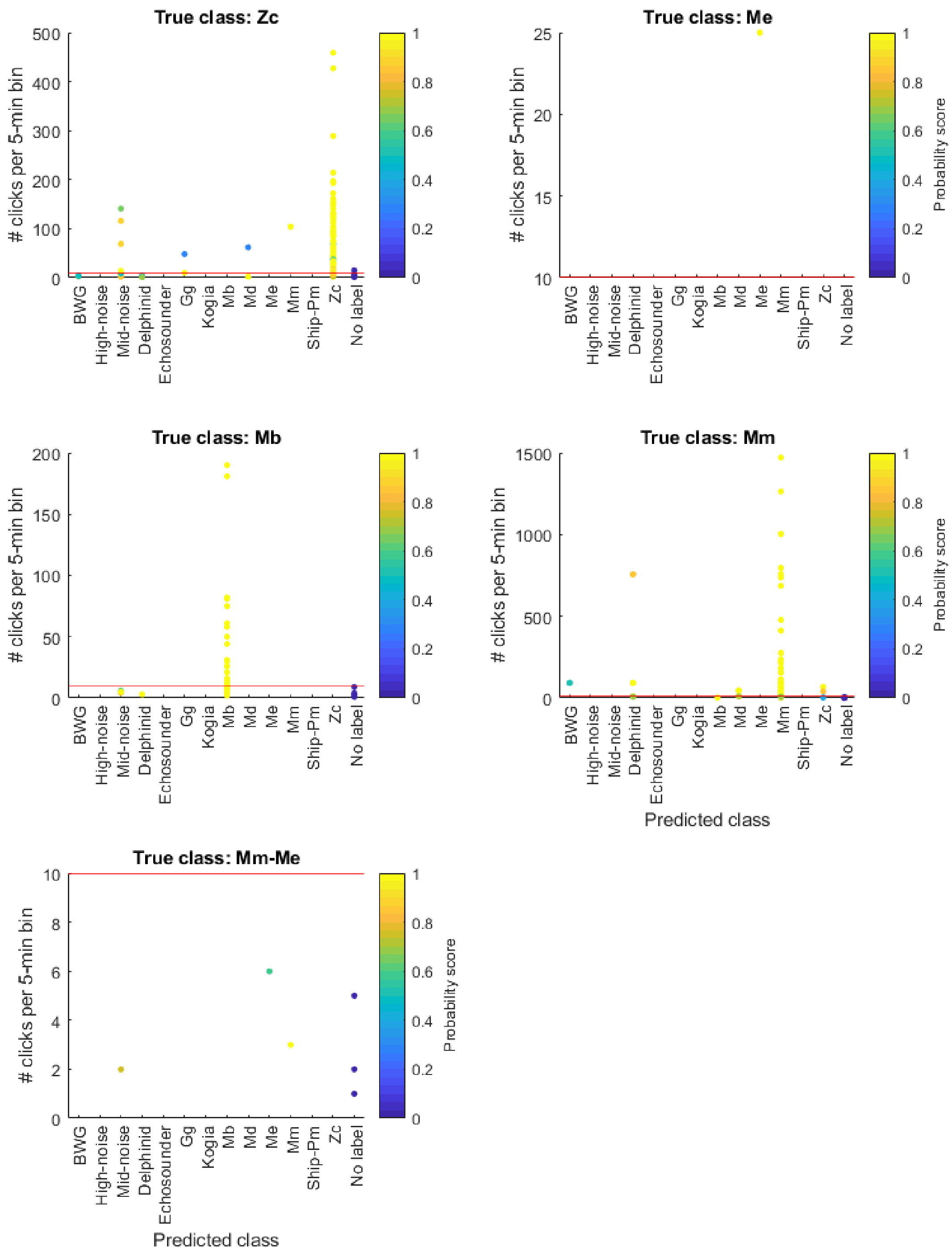

Figure 11. Total number of signals per cluster within 5-min time bins and label probability scores for each class at site BC. Each subplot represents a true class and those predicted classes assigned by the BW neural network.

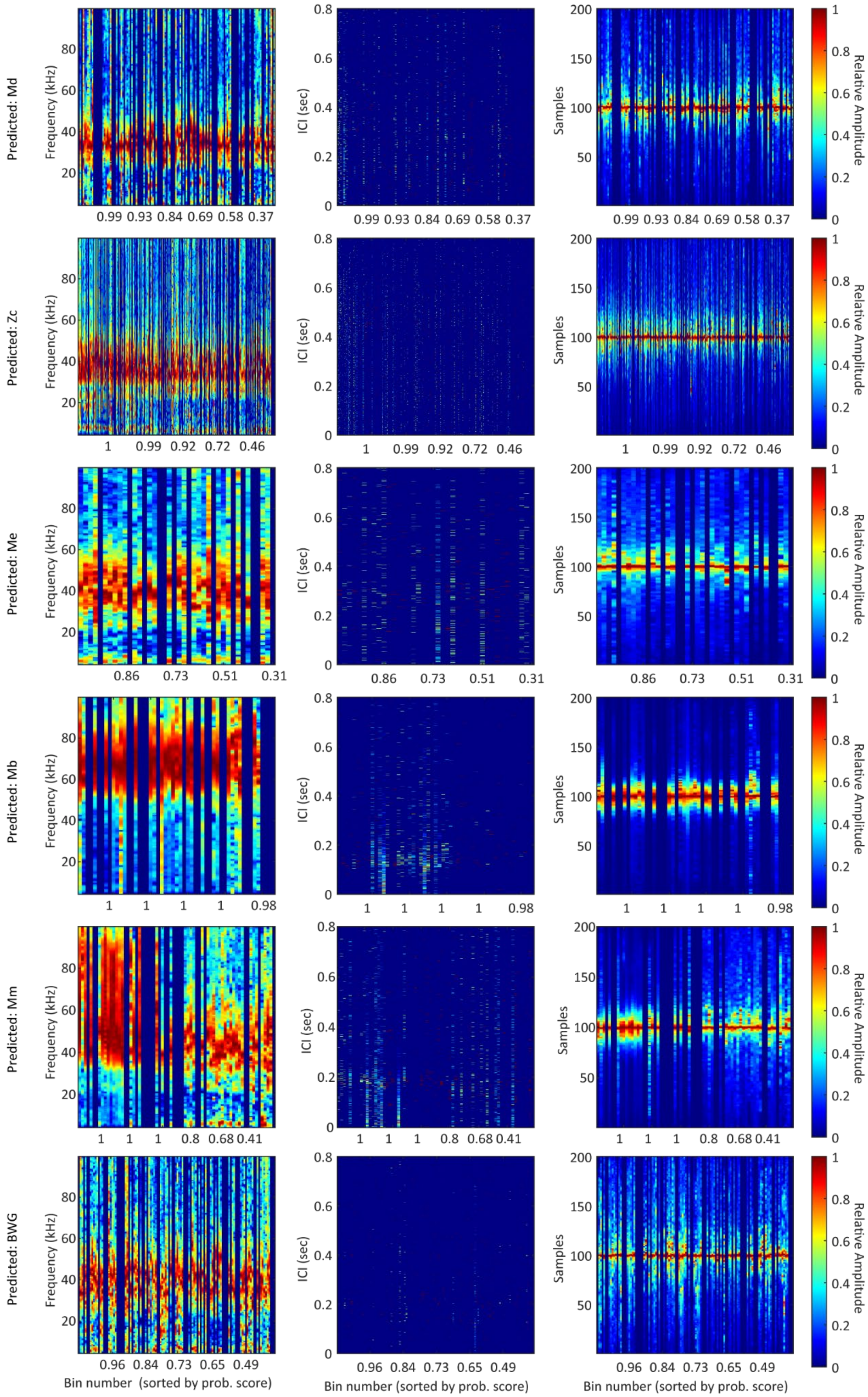

Figure 12. Averaged features of bins predicted by the neural network and not labeled in the manual dataset at site BC. Only predictions of beaked whale classes are shown. Bins are sorted by label probability.

The DE network had a lower prediction rate for beaked whale events than the BW network (**Figure 13**), with less than 50 % of the bins correctly predicted for Cuvier's and True's beaked whale at site BC. The only bin with Gervai's beaked whale presence was correctly classified by both neural networks (**Figure 10, 13**). The DE network achieved a higher predictability for Sowerby's beaked whale than the BW network, with approximately 85 % of the bins correctly classified. In contrast, only 72% of the bins were correctly classified by the BW network with double the number of unlabeled bins than the other network. In this case, all the unlabeled bins had less than ten clicks per cluster (**Figure 11**).

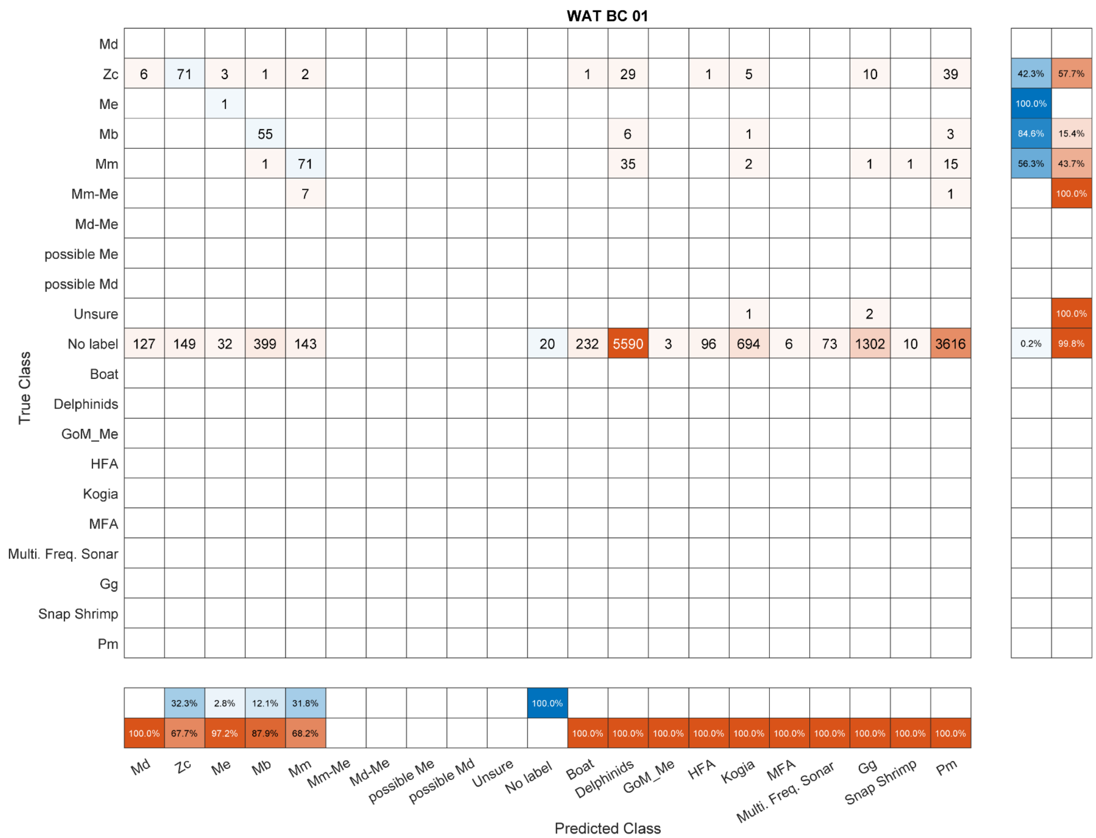

Figure 13. Confusion matrix for the DE deep neural network classification of site BC detections during summer months. Values in the matrix indicate the total number of 5-minute time bins classified.

The described classification pipeline facilitated the classification of beaked whale events, reducing the time spent running different detectors and analyst revision for different target signals. Overall higher performance was accomplished by implementing the classification pipeline with a targeted clustering method to beaked whale encounters, improving the classification of those rare events often obscured by large acoustic presence of delphinids, sperm whales or shipping noise. By operating on clusters formed from multiple similar events co-occurring within 5-minute time bins, the neural network had some contextual information. The inclusion of modal ICI distributions as an input feature for the network classification facilitated the discrimination between Gervais' and True's beaked whale acoustic encounters, which are otherwise very similar in spectral shape and waveform. After the network classification, the binned labels can be propagated down to all individual signals contained in the each cluster giving a label per signal. In this case, manual review and editing of the labels using a batch review tool such as *DetEdit* (Solsona-Berga et al., 2020) remains particularly important to achieve high confidence of beaked whale detections in 1-min segments.

II. <u>Fine-scale detection of beaked whale signals and Navy sonar events</u>

To achieve fine granularity of detections at one-minute level, all beaked whale acoustic encounters and Navy sonar pings were detected at all Atlantic sites between 2015-2019 using automated methods (described in sections I. and II.). Beaked whale encounters were automatically detected and are being classified to the species level using the trained deep neural network developed during this funding. Progress on classification of encounters per site and deployment can be found in **Table 6**.

Table 6. *Summary of data analysis of all Atlantic sites using automated methods for beaked whale and Navy sonar detection events at one-minute granularity.*

Completed - previous funding | Completed - during this funding | In progress

			BEAKED WHALES					NAVY SONAR	
			Generic impulse detector	Prune non-beaked whale detections	Unsupervised clustering	Deep network classification	Validate species labels	Detection	Params for covariates
	Heezen								
	2015-2016	WAT_HZ_01	Manual classification						
	2016-2017	WAT_HZ_02							
	2017-2018	WAT_HZ_03							
	2018-2019	WAT_HZ_04							
	Oceanographer's Canyon								
	2015-2016	WAT_OC_01	Manual classification						
	2016-2017	WAT_OC_02							
	2017-2018	WAT_OC_03							
	2018-2019	WAT_OC_04							
	Nantucket Canyon								
	2015-2016	WAT_NC_01	Manual classification						
	2016-2017	WAT_NC_02							
	2017-2018	WAT_NC_03							
	2018-2019	WAT_NC_04							
	Babylon Canyon								
	2016-2017	WAT_BC_01							
	2017-2018	WAT_BC_02							
	2018-2019	WAT_BC_03							
	Wilmington Canyon								
	2016-2017	WAT_WC_01							
	2017-2018	WAT_WC_02							
	2018-2019	WAT_WC_03							
IMPACT ANALYSIS	**Norfolk Canyon**								
	2015-2016	NFC01A							
	2016-2017	NFC_A_02	Manual classification						
	2017-2018	NFC_A_03	Manual classification						
	2018-2019	NFC_A_04	Manual classification						
	2019-2020	NFC_A_05							
IMPACT ANALYSIS	**Hatteras**								
	2015-2016	HAT_A_05							
	2016-2017	HAT_A_06							
	2017-2018	HAT_B_01_01 HAT_B_03_01 HAT_B_04_01							
	2018-2019	HAT_B_05							
	2019	HAT_B_06_01							
	2019-2020	HAT_B_07							
	Gulf Stream								
	2016-2017	WAT_GS_01							
	2017-2018	WAT_GS_02							
	2018-2019	WAT_GS_03							
	Blake Plateau								
	2016-2017	WAT_BP_01							
	2017-2018	WAT_BP_02							
	2018-2019	WAT_BP_03							
	Blake Spur								
	2016-2017	WAT_BS_01							
	2017-2018	WAT_BS_02							
	2018-2019	WAT_BS_03							
IMPACT ANALYSIS	**Jacksonville**								
	2016-2017	JAX_D_13	Manual classification						
	2018-2019	JAX_D_15	Manual classification						
	2019-2020	JAX_D_16							

III. Assessing the prevalence of seismic survey noise along the eastern seaboard

Acoustic data were analyzed from eleven HARP sites, recording from Apr 20th, 2016 - Jun 29th, 2017, totaling over 110,000 hours. Recordings ranged in length from 390-436 days, with an average of 417 days per site. Airguns were heard on 21-292 days of recordings, representing 5%-69% percent of the study period, spanning all months of the year at all sites (**Figure 14**). With the exception of the Norfolk Canyon site, airguns were detected on 50% or more of days for all sites from Cape Hatteras to the north at Heezen Canyon. Seasonally, airgun detections were highest from April-November.

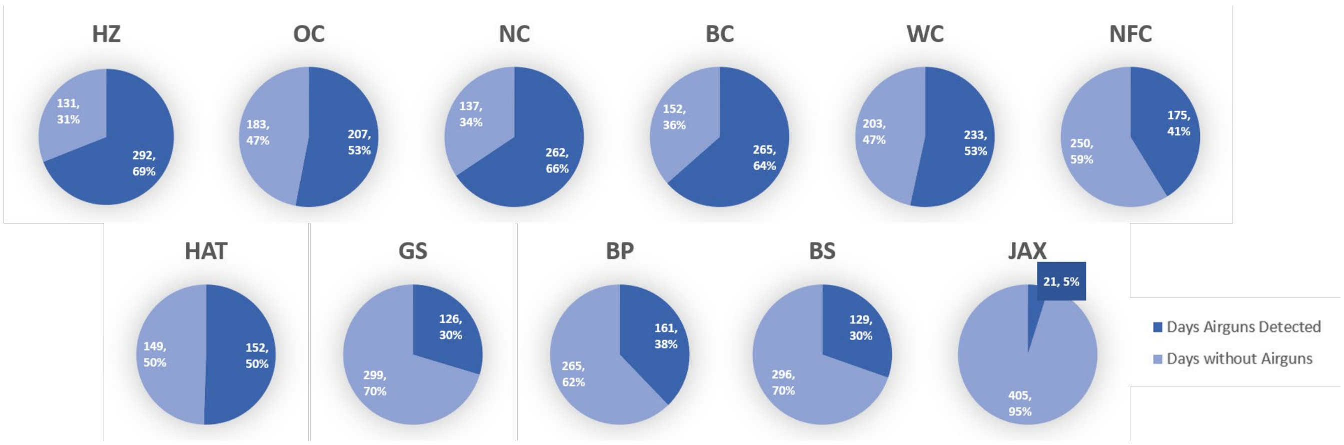

***Figure 14**. Number and percent of days with (dark blue) and without (light blue) airgun detections on each HARP site. Sites are organized from north (Heezen Canyon) to south (Jacksonville, FL), and site names correspond to FIgure 1.*

Corresponding airgun pulses were detected across at least 4 hydrophones on over 350 instances. Localization analyses suggest that there are four main sources of airgun activity: one along the North American eastern seaboard, two off the northeastern coast of South America, and one towards the mid-Atlantic **(Figure 15**). Most localized signals originated at great distance, presumably from oil fields along the South American coast. In addition, airguns were detectable across 7 hydrophones in over 30 cases, and across 10 hydrophones in at least 11 cases, indicating that in some periods, airgun signals were simultaneously detectable across all US shelf-break Atlantic waters. These findings signify the large spatial scale at which airgun pulses can impact local soundscapes, and suggest that marine seismic surveys have the potential to affect marine animals across an ocean basin.

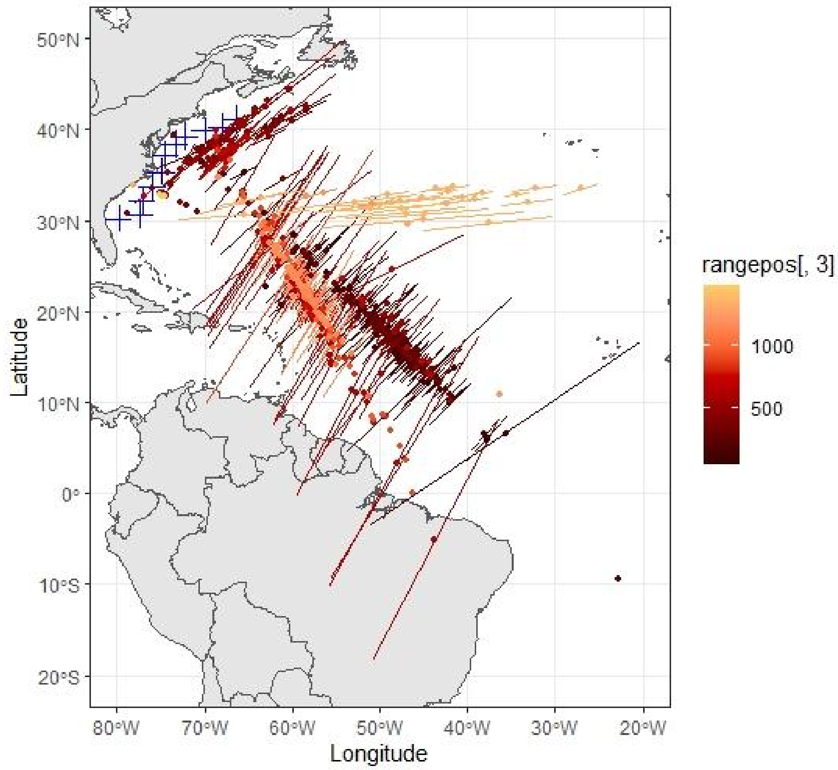

Figure 15. *Localization of seismic survey airgun pulses which were detected on four or more hydrophones, with corresponding error bars. Four main regions of seismic survey activity were identified in the acoustic data.*

IV. <u>Novel broad-scale approach to assessing acoustic niche and anthropogenic contributors, and assessing the utility of new acoustic metrics</u>

Our training dataset included 695 one-minute clips for the Heezen Canyon HARP dataset (slope site), and 389 clips for the Nantucket Shoals MARU dataset (shelf site). Five baleen whales species were present in the slope dataset (blue, fin, humpback, North Atlantic right, sei whales); the shelf site also included minke whales. Fin whales were the most prevalent species, with North Atlantic right whales being the least common in the HARP dataset (4% of acoustic clips). The highest species richness level (number of species acoustically present in the same one-minute clip) was 3 for the slope site.

The Random Forest classification models trained with the suite of acoustic metrics showed high overall model accuracy (80-92%) for species at the slope site and 82-95% at the shelf site. False negative rates were low for all species at both sites, indicating that the models predicted true species' absence with high precision. The false discovery rate varied by species, and was generally low for fin and right whales at the slope site, but higher for blue, humpback and sei whales (0.49-0.79 class 1 error). Overall the most important acoustic metrics for the models were AMP (amplitude of peak frequency computed across four bandwidths), ACI (acoustic complexity index) and BI (bioacoustic index) (see **Figure 16** for slope site). There were clear differences in the acoustic metrics that were most important between the slope and shelf sites.

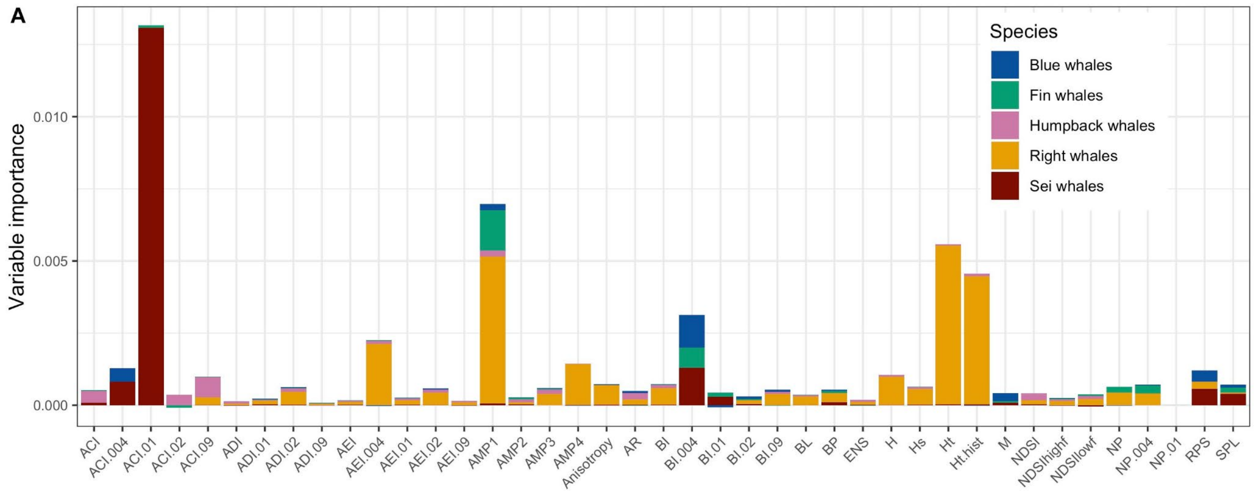

***Figure 16**. Adapted from Pegg et al. 2021 (Figure 3). Conditional variable importance plots showing all acoustic metrics included to train the random forest models, and their relative importance per species. Data shown for HARP site only.*

In summary, the random forest classification models, trained with a combination of acoustic metrics, were successful at predicting absence for multiple baleen whales species at both sites, which can be extremely useful information to inform marine soundscape planning. Models were also successful at predicting presence for a subset of the target species. The acoustic metrics that contributed most to model classification were those that summarized acoustic activity and complexity. For some site/species combinations, certain metrics contributed more significantly to species classification. The approach of using multiple acoustic metrics within a random forest modeling framework is a promising avenue for future soundscape monitoring. For more details, please refer to Pegg et al. 2021.

References

Baumann-Pickering S, McDonald MA, Simonis AE, Solsona Berga A, Merkens K, Oleson EM, Roch MA et al. 2013. Species-specific beaked whale echolocation signals. J. Acoust. Soc. Am 134: 2293-2301.

Baumgartner MF and Mussoline SE. 2011. A generalized baleen whale call detection and classification system. J. Acoust. Soc. Am 129: 2889-2902.

Biemann, C. 2006. Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems. Paper presented at the First workshop on graph-based methods for natural language processing. p. 73-80.

Davis GE, Baumgartner MF, Bonnell JM, Bell J, Berchok C, Bort Thornton J, Brault S et al. 2017. Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (Eubalaena glacialis) from 2004 to 2014. Sci. reports 7, 1: 1-12.

Frasier KE. 2021. A machine learning pipeline for classification of cetacean echolocation clicks in large underwater acoustic datasets. PLoS Computational Biology. 17(12):e1009613.

Frasier KE, Roch MA, Soldevilla MS, Wiggins S M, Garrison LP, and Hildebrand JA. 2017. Automated classification of dolphin echolocation click types from the Gulf of Mexico. PLOS Comput. Biol., 13: e1005823.

Maas AL, Hannun AY, Ng AY, editors. 2013. Rectifier nonlinearities improve neural network acoustic models. Proc icml: Citeseer.

Pegg N, Roca IT, Cholewiak D, Davis GE, Van Parijs SM. 2021. Evaluating the Efficacy of Acoustic Metrics for Understanding Baleen Whale Presence in the Western North Atlantic Ocean. Frontiers in Marine Science, 8:749802.

Rafter MA, Rice AC, Thayre BJ, O'Neill E, Wiggins SM, Frasier KE, Baumann-Pickering S, Hildebrand JA Cholewiak DM, Van Parijs SM. 2020. Passive Acoustic Monitoring for Marine Mammals in the Western Atlantic June 2017–June 2018. Final Report. Marine Physical Laboratory Technical Memorandum 640.

Roca IT and Van Opzeeland I. 2019. Using acoustic metrics to characterize underwater acoustic biodiversity in the Southern Ocean. Remote Sensing in Ecology and Conservation 6: 262-273.

Roch MA, Scott Brandes T, Patel B, Barkley Y, Baumann-Pickering S, and Soldevilla MS. 2011. Automated extraction of odontocete whistle contours, J. Acoust. Soc. Am., 130:2212.

Pegg N, Roca IT, Cholewiak D, Davis GE, Van Parijs SM. 2021. Evaluating the Efficacy of Acoustic Metrics for Understanding Baleen Whale Presence in the Western North Atlantic Ocean. Frontiers in Marine Science. https://doi.org/10.3389/fmars.2021.749802

Solsona-Berga A, Frasier KE, Baumann-Pickering S, Wiggins SM, and Hildebrand JA. 2020. DetEdit: A graphical user interface for annotating and editing events detected in long-term acoustic monitoring data. D. Schneidman-Duhovny, Ed. PLOS Comput. Biol., **16**: e1007598.

Stanistreet JE, Nowacek DP, Baumann-Pickering S, Bell JT, Cholewiak DM, Hildebrand JA, Hodge L, Moors-Murphy HB, Van Parijs SM, and Read AJ. 2017. Using passive acoustic monitoring to document the distribution of beaked whale species in the western North Atlantic Ocean. Canadian Journal of Fisheries and Aquatic Sciences 74: 2098-2109.

Stanistreet JE, Nowacek DP, Bell JT, Cholewiak DM, Hildebrand JA, Hodge LE, Van Parijs SM and Read AJ. 2018. Spatial and seasonal patterns in acoustic detections of sperm whales Physeter macrocephalus along the continental slope in the western North Atlantic Ocean. End. Sp. Res. 35:1-13.

Weiss SG, Cholewiak D, Frasier KE, Trickey JS, Baumann-Pickering S, Hildebrand JA, Van Parijs SM. 2021. Monitoring the acoustic ecology of the shelf break of Georges Bank, Northwestern Atlantic Ocean: New approaches to visualizing complex acoustic data. Marine Policy 130:104570.

Wiggins SM and Hildebrand JA. 2007. High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring. In: Symposium on Underwater Technology and Workshop on Scientific Use of Submarine Cables and Related Technologies, pp. 551-557, IEEE.

Wiggins SM, Roch MA, and Hildebrand JA. 2010. TRITON software package: Analyzing large passive acoustic monitoring data sets using MATLAB. J. Acoust. Soc. Am., 128:2299.