# Evaluating the Efficacy of Acoustic Metrics for Understanding Baleen Whale Presence in the Western North Atlantic Ocean

Nicole Pegg[1,2], Irene T. Roca[3,4], Danielle Cholewiak[1], Genevieve E. Davis[1] and Sofie M. Van Parijs[1]*

[1] Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Woods Hole, MA, United States, [2] Harbor Branch Oceanographic Institute, Florida Atlantic University, Fort Pierce, FL, United States, [3] Helmholtz Institute for Functional Marine Biodiversity, University of Oldenburg, Oldenburg, Germany, [4] Ocean Acoustics Lab, Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Bremerhaven, Germany

Soundscape analyses provide an integrative approach to studying the presence and complexity of sounds within long-term acoustic data sets. Acoustic metrics (AMs) have been used extensively to describe terrestrial habitats but have had mixed success in the marine environment. Novel approaches are needed to be able to deal with the added noise and complexity of these underwater systems. Here we further develop a promising approach that applies AM with supervised machine learning to understanding the presence and species richness (SR) of baleen whales at two sites, on the shelf and the slope edge, in the western North Atlantic Ocean. SR at both sites was low with only rare instances of more than two species (out of six species acoustically detected at the shelf and five at the slope) vocally detected at any given time. Random forest classification models were trained on 1-min clips across both data sets. Model outputs had high accuracy ($>0.85$) for detecting all species' absence in both sites and determining species presence for fin and humpback whales on the shelf site ($>0.80$) and fin and right whales on the slope site ($>0.85$). The metrics that contributed the most to species classification were those that summarized acoustic activity (intensity) and complexity in different frequency bands. Lastly, the trained model was run on a full 12 months of acoustic data from on the shelf site and compared with our standard acoustic detection software and manual verification outputs. Although the model performed poorly at the 1-min clip resolution for some species, it performed well compared to our standard detection software approaches when presence was evaluated at the daily level, suggesting that it does well at a coarser level (daily and monthly). The model provided a promising complement to current methodologies by demonstrating a good prediction of species absence in multiple habitats, species presence for certain species/habitat combinations, and provides higher resolution presence information for most species/habitat combinations compared to that of our standard detection software.

**Keywords:** acoustic metrics, soundscapes, baleen whales, random forest classification model, species richness

# INTRODUCTION

Soundscapes comprise the complex variety of biological, anthropogenic, and environmental sounds of a given habitat (Krause, 2008; Pijanowski et al., 2011). They provide a unique perspective into a given ecosystem, whether terrestrial or marine, due to the integrative approach of studying all sounds of an environment together (e.g., Farina, 2013). However, recent capacity increases in long-term acoustic data collection are challenging scientists to develop new and creative ways to analyze these complex data (e.g., Gibb et al., 2019). Extracting variability in patterns and measuring diversity in these data frequently requires taking a number of approaches. McKenna et al. (submitted to this research topic) define soundscape analyses as viewing a question through different lenses by dividing them into biodiversity assessments, human impacts, and the acoustic scene analysis, with each view requiring different approaches and metrics. Which analyses or metrics are most appropriate depends on the "lens" with which one is interested in interpreting the data. Various approaches for analyzing soundscape data range from long-term ambient sound measurements, observing variation in sound pressure levels, and using individual or a combination of acoustic metrics (AMs) (e.g., Sueur et al., 2014; Kaplan et al., 2015; Haver et al., 2018; Bradfer-Lawrence et al., 2019).

Acoustic metrics refer to a number of standardized and automated metrics that can comprise acoustic diversity indices such as Acoustic Complexity Index, entropy, evenness, and roughness (e.g., Sueur et al., 2014; Bradfer-Lawrence et al., 2019; Farina et al., 2021), as well as specific metrics on the spectral and temporal patterns of the soundscapes such as the spectral integral, the envelope median, and background noise proportion (Boelman et al., 2007; Depraetere et al., 2012; Towsey, 2017). AM can be used singly or in tandem to aggregate complex information from acoustic recordings into a single value. Terrestrial soundscape analyses have applied the use of some AM successfully as proxies for estimating biodiversity (e.g., Sueur et al., 2008), community composition (e.g., Farina et al., 2011), habitat type and vegetation (e.g., Do Nascimento et al., 2020), and ecological condition (e.g., Tucker et al., 2014). Given the success of AM for understanding terrestrial soundscapes, these same AM were applied to the marine environment (e.g., Pieretti et al., 2017; Bohnenstiehl et al., 2018). However, direct application of these approaches to the marine environment have generally not been as successful due to the frequent overlap in frequency bands of biological sounds and background noise that are present, often leading to a lack of a clear relationship between biodiversity and acoustic indices (Mooney et al., 2020). One popular metric, the Acoustic Complexity Index, showed mixed results in describing patterns in the sound assemblages on coral reefs (Mooney et al., 2020). Acoustic indices also were not able to predict bioacoustic activity due to overlap with snapping shrimp and other anthropogenic sounds (Buxton et al., 2018). However, these studies focused on highly complex reef environments. When focusing only on low-frequency data (<125 Hz), Parks et al. (2014) successfully used a single acoustic entropy index to characterize baleen whale calls. Based on these mixed applications of AM there is a need for further developing approaches that can cope better with the heightened prevalence of overlapping sound sources in the marine environment. Solutions could include developing new indices specifically for the marine environment (Parks et al., 2014), using other methods to support assessments like clustering (Mooney et al., 2020), or combining AM in novel ways. Recently, an approach using a number of AM in combination with supervised learning algorithms was successfully used to classify species richness (SR) levels and species identities of marine mammal acoustic communities in the Southern Ocean (Roca and Van Opzeeland, 2019). Here, we evaluate whether this AM analysis approach might be helpful to understanding other marine mammal communities.

Seven baleen whale species are found in the western North Atlantic Ocean, where their distributions often overlap in space and time. These include the North Atlantic right whale (NARW; *Eubalaena glacialis*), Brydes' (*Balaenoptera edeni*), blue (*Balaenoptera musculus*), fin (*Balaenoptera physalus*), minke (*Balaenoptera acutorostrata*), sei (*Balaenoptera borealis*), and humpback whales (*Megaptera novaeangliae*). Recently, a decade-long dataset was analyzed using a low-frequency detection and classification system (LFDCS) with species-specific call libraries (Baumgartner and Mussoline, 2011), highlighting North Atlantic right whale year-round use of western North Atlantic Ocean habitat (Davis et al., 2017). In addition, blue, fin, humpback, and sei whales also exhibited wide habitat use in this area, especially in winter when all species were detected throughout the entire data collection range, from the Caribbean to the Greenland Sea (Davis et al., 2020). Minke whale acoustic daily presence was previously described using a different automated acoustic detector, showing seasonal variability and migratory movement throughout the western North Atlantic Ocean (Risch et al., 2014). Bryde's whales have been acoustically detected in the southern Caribbean (Oleson et al., 2003), but were not included in this study as their range does not typically include the region of the western North Atlantic where most of our acoustic monitoring efforts occur.

While the use and development of species-specific automated detectors are important for bioacoustic analyses, manually reviewing detector output can still be time consuming. Areas of species overlap within baleen whale feeding and migratory routes provide potential sites for methods, like AM, to be useful to gain more knowledge of the acoustic environments and marine mammal community composition in a faster and more streamlined manner. AM can also be applied across different datasets (Roca and Van Opzeeland, 2019), which is necessary when using a wide range of data aimed at understanding baleen whale presence across the western North Atlantic Ocean.

Baleen whales vary in their utilization of shelf, slope, and pelagic habitats throughout their range. In the western North Atlantic Ocean, some species, such as blue whales, have a predominantly pelagic distribution, but can be found seasonally in regions on the continental shelf (e.g., Lesage et al., 2017; Davis et al., 2020). Fin whales occur year-round in on-shelf areas, like Massachusetts Bay (Morano et al., 2012), but also in offshore waters near the Mid-Atlantic ridge (Nieukirk et al., 2012). In contrast, NARWs are generally found in coastal, on-shelf waters (Davis et al., 2017), and can be distributed very close to shore.

Due to this species variation in habitat usage, it is important to see how AM perform across different environments.

Previously, success was found in using a suite of AM to characterize the marine mammal community composition in the Southern Ocean with a combination of Balaenopterid, Physeterid, Delphinid, and Phocid species comprising the acoustic assemblages (Roca and Van Opzeeland, 2019). Here, we evaluate if a similar combination of AM and supervised machine learning could help determine the acoustic presence of individual species and SR levels specifically of baleen whales in the western North Atlantic Ocean at the continental shelf and slope sites.

## MATERIALS AND METHODS

### Study Sites and Acoustic Recordings

Data were collected from two recording sites in the western North Atlantic Ocean: one located on the continental shelf (hereafter referred to as "shelf"), and the other along the continental slope ("slope"). At the slope site, a High Frequency Acoustic Recording Package (HARP; Wiggins and Hildebrand, 2007) was deployed at a depth of 845 m at 41.06 N, 66.35 W, on the outer United States continental slope near Heezen Canyon, with the hydrophone approximately 20 m above the seafloor (**Figure 1** and **Table 1**). The HARP was programmed to record continuously at a bandwidth rate of 200 kHz from June 11, 2018–May 10, 2019. The HARP was comprised of two transducers; one high-frequency stage using an ITC-1042 hydrophone (International Transducer Corporation), and one low-frequency stage using Benthos AQ-1 transducers[1] (see **Table 1** for more technical information). For the purpose of this study, only the low frequency data (<2,000 Hz) were analyzed to best visualize the calls of baleen whale species found in the study area. At the shelf site, a Marine Autonomous Recording Unit (MARU; Clark et al., 2010) was deployed at a depth of 78.6 m at 40.393 N, 70.217 W near the Nantucket Shoals area on the United States continental shelf (**Figure 1** and **Table 1**). The MARU was programmed to record continuously at a 2 kHz sampling rate from December 21, 2016–February 14, 2018 with a gap in data due to battery life constraints from July 15 to 17, 2017 (see **Table 1** for additional technical information). Recording sites were selected based on previous analyses of baleen whale acoustic presence in slope and shelf environments around George's Bank and New England waters (e.g., Risch et al., 2014; Davis et al., 2017, 2020; Weiss et al., 2021).

### Acoustic Data Processing and Acoustic Metrics

We performed a stratified random sampling over the complete dataset from the slope and shelf sites to select the acoustic files to constitute our training set ($n$ = 695 and $n$ = 389, respectively). We searched for high quality recordings [i.e., high signal-to-noise ratio (SNR) for species spectral patterns] through as many months and days as possible in order to capture as much temporal, spatial, and species variation in the training set. All sound files for both locations were clipped to 1 min

[1] www.benthos.com

in length for the best call resolution based on baleen whale vocalizations in the study areas. Clipping was performed using Matlab R2017a (The Mathworks, Inc.).We manually assessed species presence/absence in the training set through visual and aural inspection of spectrograms using Raven Pro 2.0 (Cornell Lab of Ornithology, Ithaca, NY, United States). Spectrogram settings were chosen and changed to best help identify different species' call presence/absence (512pt. FFT for NARW, sei, minke, and humpback; 4096 FFT for fin and blue, Hamming window, 50% overlap).

We defined SR as the number of species acoustically present in a 1-min file, ranging from 0 to 6, and quantified SR for each file based on visual and aural review of the spectrograms. A species was marked absent based on a lack of acoustic presence. We computed 21 different AM (for details on the computed AM see **Supplementary Table 1**) for every acoustic file in each training set over the full bandwidth 0–1,000 Hz. The AMs that were chosen and the AM parameterization were adapted according to the target species' call patterns to optimize results. The AM selected were based on their mathematical principles, which had the potential to be the most relevant, and which had already shown to perform well in other marine, specifically marine mammal, acoustic contexts. To account for the complexity and variability in the spectral patterns of the species' calls, the AEI, ADI, ACI, BI, and NP metrics (see **Supplementary Table 1**) were computed over four other relevant bandwidths: 10–40; 40–100; 100–200; and 200–900 Hz, for a total of 44 AMs computed per acoustic file. All AMs were computed using R. We used functions from the R package seewave (Sueur et al., 2008) to calculate H, th, sh, ACI, NP, and M, and the R package soundecology (Villanueva-Rivera et al., 2018) to calculate AEI, ADI, BI, and NDSI.

### Random Forest Classification Models

We used random forest classification models (Breiman, 2001) to discriminate between the acoustic presence/absence of the different species comprising the acoustic community at each site and to evaluate the discrimination potential of the AMs. Random forest models are widely used tools that show high predictive accuracy and can cope with high-dimensional problems, complex interactions, and even highly correlated predictor variables.

We used the Boruta algorithm (Kursa and Rudnicki, 2010) to select relevant AMs to include as predictor variables in the random forest classification models. The Boruta algorithm iteratively removes the variables that are statistically less relevant than their randomly permuted copies (the random copies' importance can be non-zero only due to random fluctuations). We used the *Boruta* function from the Boruta package (Kursa and Rudnicki, 2010) in R.

We used the *randomForest* function in the R randomForest package (Liaw and Wiener, 2002) to develop the random forest classification models. We ran a hyperparameter grid search for each species model using the R package ranger (Wright and Ziegler, 2015) on values for the total number of trees necessary to stabilize prediction error rates, the number of predictor variables to randomly sample from at each node, the minimum number of samples within the terminal nodes and the maximum number of nodes (both define the degree of model complexity) and finally,
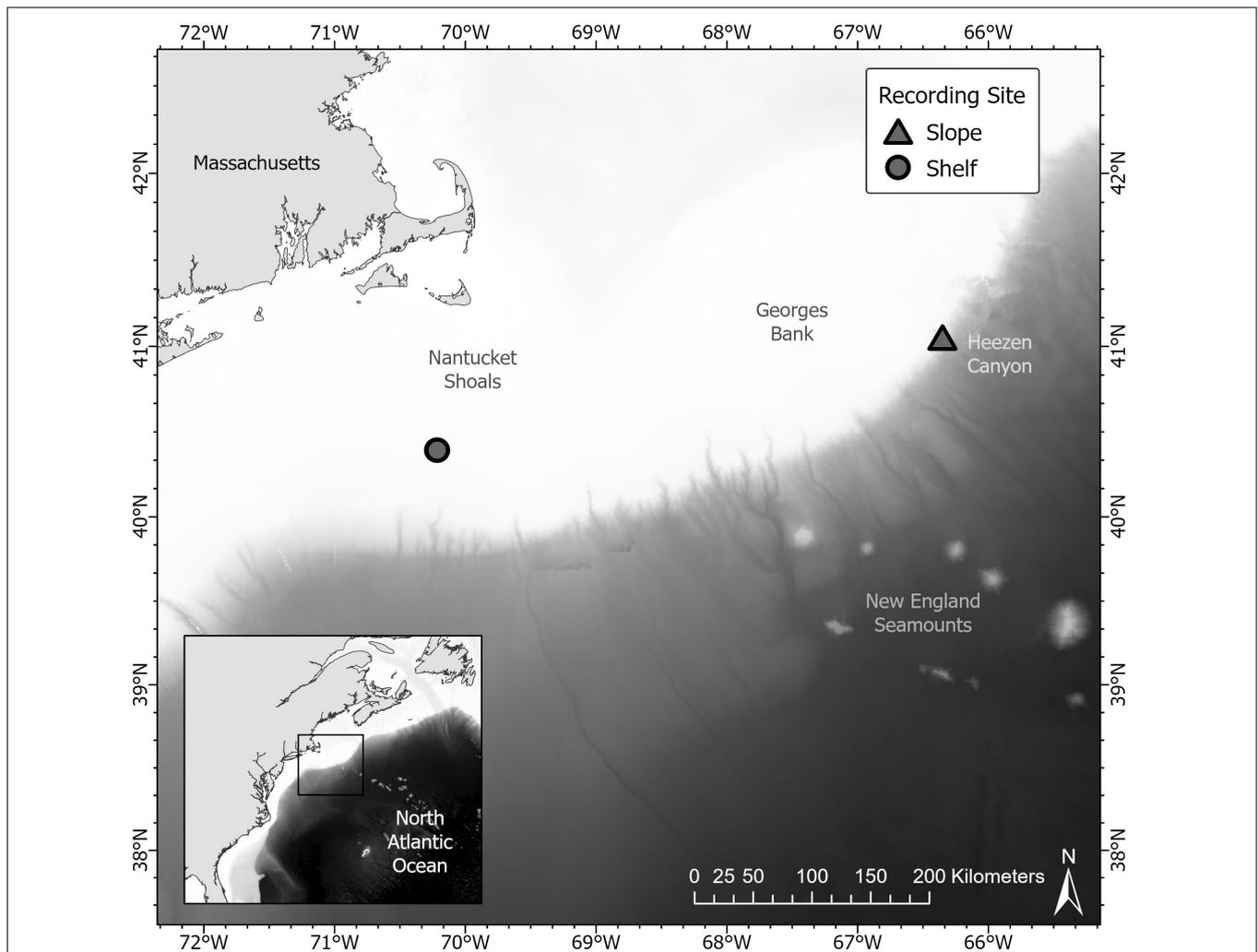
**FIGURE 1 |** This map shows the location of the two acoustic recording units in the western North Atlantic Ocean used in this study. At the slope site, a High Frequency Acoustic Recording Package (HARP) was located along the continental slope edge near Heezen canyon off of Georges Bank (triangle). At the shelf site, a Marine Autonomous Recording Unit was located near Nantucket Shoals (circle).

**TABLE 1 |** This table provides details of the two passive acoustic recorders used for this study, one of which was located along the continental slope and another was located on the shelf.

| Recorder | Location (latitude, longitude) | Depth (m) | Recording date range | Sampling rate (kHz) | Bit depth (bit) | Frequency range (Hz) | Sensitivity (dB re: 1 $\mu$V/$\mu$ Pa) |
|---|---|---|---|---|---|---|---|
| HARP (slope site) | 41.062, −66.352 | 845 | June 11, 2018–May 10, 2019 | 200 | 16 | 10–100,000 | −187 |
| MARU (shelf site) | 40.393, −70.217 | 78.6 | December 21, 2016–July 15, 2017 and July 17, 2017–February 14, 2018 | 2 | 12 | 10–1,000 | −151.2 |

*The acoustic recorder on the former was a High Frequency Acoustic Recording Package (HARP) and the latter was a Marine Autonomous Recording Unit (MARU).*

the sizes of training and test data subsets to find the best model parameterization according to the above mentioned criteria. We grew 2,001 trees with a node size of 1 and tested between 6 and 16 predictor variables at each split according to the species model. For each tree constructed in the random forest, ∼ 2/3 (0.66–0.75% of data according to species) of the training set were sub-sampled with replacement to train the classification model

and ∼ 1/3 was left out as a test subset (i.e., out-of-bag or OOB cases). The general misclassification rate of the model (general OOB estimate) is computed as the average across all OOB cases and trees. We used a conditional permutation scheme (Strobl et al., 2008) to assess variable importance, in order to account for correlations that occurred between some of the AMs. We used the *permimp* function from the permimp package in R

(Debeer and Strobl, 2020) with a 0.80 threshold value. We used the area under the curve (AUC) to compute the permutation importance (Janitza et al., 2013).

## Model Predictions and Second-Step Evaluation

To evaluate the performance of the models across a long time series (12 months from January 1 to December 31, 2017), we made a case-study of the models developed for the species recorded at the shelf site. This site was chosen instead of the slope site due to six species being acoustically present instead of six, more diverse combinations of species vocalizing at higher SR levels, and a higher SNR for some species, especially humpback whales, in the clips. We used the generic *predict* function in R to generate predictions of species presence probabilities on the complete dataset using the trained random forest classification models. We randomly selected 400 1-min acoustic files from the complete dataset and annotated species' presence/absence using Raven Pro 2.0 as mentioned in section "Acoustic Data Processing and Acoustic Metric" as an independent test set to use as second-step cross validation and evaluation of the species classification models' performance. To assess the optimal threshold to generate presence/absence scores from the model predicted probabilities in order to evaluate model performance for each species, we used the *optimal.thresholds* function from the PresenceAbsence package in R (Freeman and Moisen, 2008). We used the *confusionMatrix* function from the caret package in R (Kuhn, 2020) to conduct the second-step evaluation of models' performance. We computed a relative daily proportion of acoustic presence for every species as the sum of the 1-min species presence scores per day (obtained by applying the optimal threshold to the estimated probabilities) divided by the total number of minutes recorded per day (i.e., 1,440).

We then compared the modeled outputs summarized as relative daily proportion of acoustic presence with the manually assessed daily acoustic presence of the different baleen whale species for the shelf data set, derived using two acoustic detection software methodologies: LFDCS (Baumgartner and Mussoline, 2011) for NARW (upcall), sei (doublet or triplet down sweeps), blue (A, B, and AB phrases), fin (20 Hz pulse), and humpback (song across multiple years and social sounds) whales (according to Davis et al., 2017, 2020), and an automated pulse train detector (according to Risch et al., 2014) for minke whales (**Figure 2**). The call types listed above for each individual species were the only call types used when selecting and annotating the 1-min clips for analysis.

Low-frequency detection and classification system detections were manually reviewed by trained acoustic analysts to determine daily presence of each of the baleen whale species. Given the variability of each species' vocalizations, the specific methodology to determine daily acoustic presence varied by species (see Davis et al., 2017, 2020 for a more complete description). A true detection was defined as a pitch track that correctly classified a call or song unit to the species that produced it (Bonnell et al., 2016). All detections were reviewed by an analyst until a true detection was found for NARW, sei, humpback, and

blue whales. In contrast, only hours with 29 or more detections for fin whales were reviewed manually for acoustic presence due to a logistic regression application revealing 29 to be the minimum number of detections in an hour to ensure a fin whale was acoustically present with 90% confidence (Davis et al., 2020). This was implemented in order to reduce the time of manually reviewing detections. In addition, the automated pulse train detector which used a Ripple-Down Rule learner trained to identify the three main types of minke whale low-frequency pulse trains implemented by Risch et al. (2014), was applied to examine selected recordings for the acoustic presence of North Atlantic minke whales. It is important to note that the output from the classification model represents the relative daily proportion of species presence across 1-min clips while in our manual detection protocol, a day is marked as "present" for the full 24 h as soon as a single target species call type is observed. The manual detection protocol represents a minimum acoustic presence for each species in order to maximize confidence of acoustic presence. Therefore first, the model resolution presents information at a much finer scale than the latter approach and second, the output comparison between both methodologies concerns the period of the species-specific acoustic activity peak and its duration but not the magnitude of the daily activity or its monthly average.

## RESULTS

### Species Richness of Training Data Set

All six baleen whale species were observed in the shelf training dataset for the random forest classification models, while only five target species were detected in the slope training dataset. Minke whale acoustic presence was only included in the analyses of the shelf dataset, and not the slope dataset due to the lack of minke whale vocalizations during the time periods analyzed when other baleen whale species were vocalizing in the slope dataset. Fin whales were the most represented species in the slope dataset, comprising just over 50% of the SR 1 clips. They were also the species most present in clips with the higher SR levels. NARW were present in the least number of clips in the slope dataset (27/695 clips). The shelf dataset had a more even representation of species in the clips than that of the slope dataset. The 1-min clips with SR of 0–2 co-occurring species were the most commonly observed in both training datasets (slope: 639/695 clips; shelf: 338/389 clips). The highest SR level found in the slope dataset was 3 (55/695 clips), and the highest in the shelf dataset was 4 (4/389 clips), indicating that most often 2 or fewer species were acoustically active at any given time and only rarely were there more than 2 species vocalizing.

### Acoustic Metrics and Classification Models

The random forest classification models trained with a combination of relevant AM (**Supplementary Table 1**) showed high average accuracy values for all species around the slope and the shelf sites (**Tables 2**, **3**). Overall model accuracy ranged between 80 and 92% for species around the slope site and
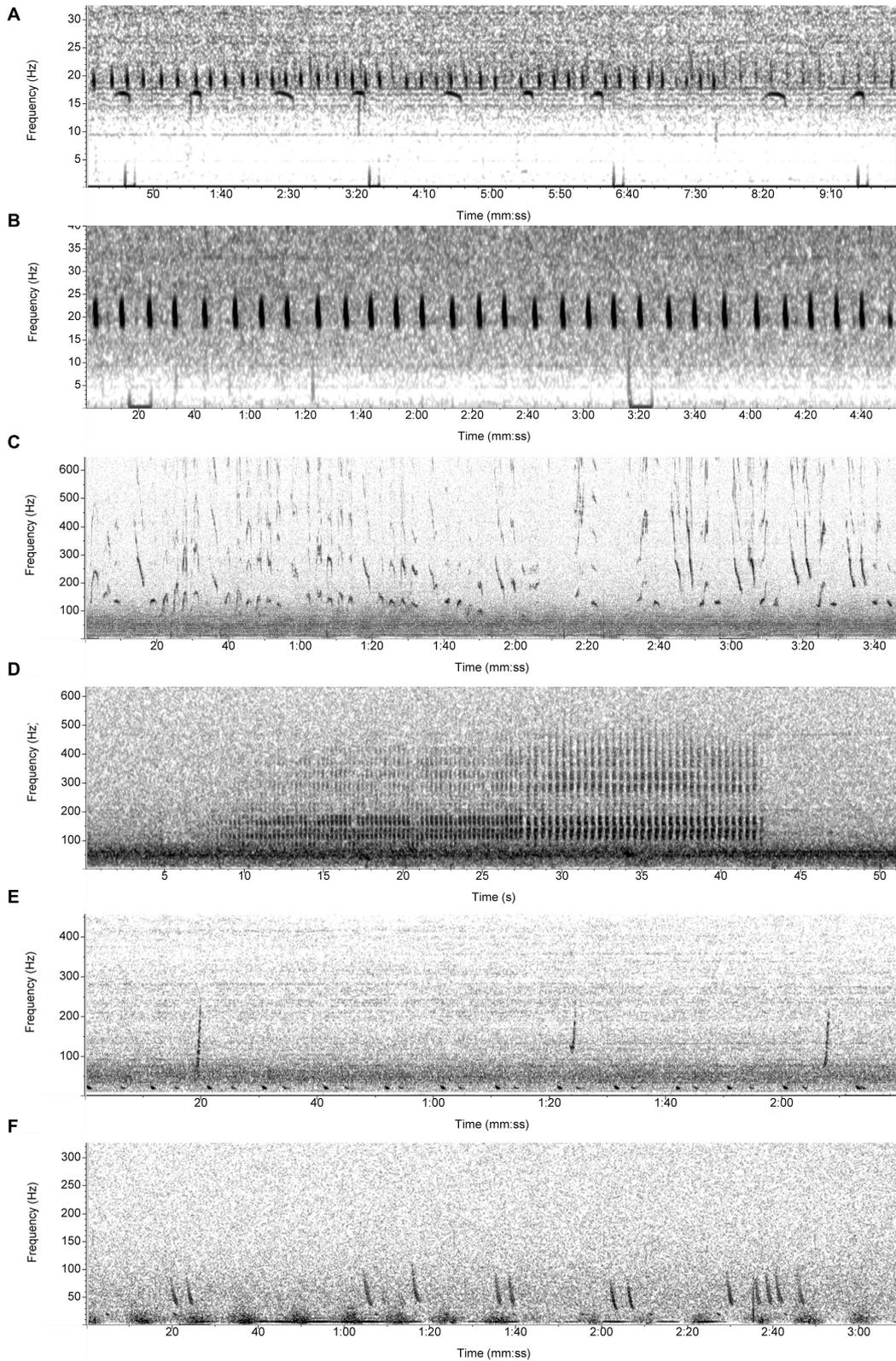
**FIGURE 2 |** Spectrograms of the vocalization types used to detect species presence for the **(A)** blue whale (A/B song), **(B)** fin whale (20 Hz pulse), **(C)** humpback whale (song), **(D)** minke whale (pulse train), **(E)** North Atlantic right whale (upcall), and **(F)** sei whale (downsweep).

between 82 and 95% for species around the shelf site. The false omission rate (class 0 error) was low for all species around both sites (0–12%), indicating that the models had high precision for predicting true species' absence (**Tables 2**, **3**). The classification models trained to discriminate the calls from the species around the slope site performed best for fin and right whales with false discovery rates (class 1 error) <0.15 (**Table 2**). However, model performance for blue, sei, and humpback whales was low (0.42, 0.54, 0.79, class 1 error respectively; **Table 2**). The precision of the classification models was higher in general for the species recorded around the shelf site, with false discovery rates (class 1 error) ranging from 0.43 (sei whales) to 0.12 (fin whales; **Table 3**). While the model's ability to discriminate between SR levels was better for the shelf than the slope site (0.07 and 0.29 OOB errors respectively for SR 0), the model did not perform well at the higher SR levels for both sites in discriminating between the SR levels (**Supplementary Table 2**).

**Figure 3** shows the relative importance of the different AM used in the random forest classification models to discriminate between the different species presence. The number of AM that were included in the classification models for species identification can be found in **Tables 2**, **3**. Results on the conditional importance of the AMs used to train the classification models showed a good agreement in general between the AMs that were found to be most important for the classification of the different species acoustically present around the shelf and slope sites (**Figure 3**). Overall, the most important AMs were AMP, ACI, and BI (see **Supplementary Table 1** for details). However, the importance of these AMs was highest when computed on the frequency bands corresponding with the bands in which the respective species' call patterns show most of their acoustic energy (**Figure 3**). Further, there were also clear differences in the number of AM that were relatively most important between slope and shelf sites, with species in the latter being captured by a wider range of AM.

## Second-Step Model Evaluation of on Shelf Acoustic Recordings

The resulting model was applied to 12 months of near continuous acoustic recordings on the shelf site to evaluate its performance. Results on the second-step model evaluation showed high average accuracy scores for all species models regardless of the criteria used to determine the optimal threshold to produce presence/absence scores from the estimated probabilities. However, the threshold that in general yielded the best balance between sensitivity and precision for all species was the threshold that maximized Kappa (**Table 4**). The Kappa statistic provides a measure of agreement between the predicted and observed classes above that expected by chance (Kuhn and Johnson, 2013). Model performance was very high for fin, minke, and humpback whales with sensitivity and precision scores >0.70 (minke whale model precision = 0.67). Blue, NARW, and sei whales' models showed high average accuracy and specificity but, in general, low sensitivity and precision (0.30–0.40; **Table 4**).

We then compared the long-trend output of the classification model to the results from the manual detection software approach that we standardly use to evaluate baleen whale presence in all of our western Atlantic data sets (**Figure 4**). The classification model estimation represents the species' average daily relative presence proportion compared to a single verification of a species' acoustic presence per day to determine daily presence in our manual detection methodology. Accordingly, the former provides much higher resolution detection information compared with the latter. The comparison made therefore is not between the raw proportion of presence in a given month (since the measures are not comparable) but in the timing and duration of the relative monthly acoustic presence peak for each species. The patterns in species presence showed good correspondence between the two methodologies for most species and time periods including peaks for blue whales in winter months, humpback, minke, and sei whales in spring months, and fin whales in late winter/early spring months (**Figure 4**). There were some discrepancies between the results of the model and manual verification. For example, fin whale proportion of acoustic occurrence showed higher proportions in the summer and fall months for the model approach, while fin whale presence showed zero proportion of acoustic occurrence during those same months in the detector and manual verification approach (**Figure 4**). Although the model did not perform well for blue and sei whale species when estimating acoustic presence at a 1-min clip level, when pooled across a day, the model provided a good approximation to the standard manual detection approach regarding the species relative presence and high activity periods.

## DISCUSSION

This study showed that the shelf and slope sites include habitat that is utilized by the majority of baleen whale species that occur in the Northwestern Atlantic Ocean, with six and five different species detected at each site, respectively. However, in our training dataset, it was difficult to find more than two species vocalizing in the same 1-min time periods, resulting in low overall SR values. Due to the short length of the 1-min clips and the characteristics of the baleen whale calls around the study sites, it was not surprising that the SR levels were low in the clips analyzed. In some cases, different combinations of species' vocalizations were detected at different times of day, or on nearby days, indicating that there were more species in the area than what may be represented by the SR values alone. Nevertheless, model discrimination for SR levels was poor (**Supplementary Table 2**), probably due to the higher intra- than inter-level acoustic variability, so it was not feasible to directly predict the number of co-occurring species in the same 1-min time period across the full datasets.

In this study, our random forest classification models trained with a combination of relevant AM were successful at discriminating the presence/absence of the call repertoire of different baleen whale species at the shelf and slope sites. Overall model accuracy was high for all species at both sites and in general, it was primarily driven by the model's strong ability to predict the absence of the baleen whale species, meaning that they were highly successful at detecting time periods when specific

**TABLE 2 |** Model results for the slope site.

| Species | N | n AMs | OOB error (%) | Class 0 error | Class 1 error |
|---|---|---|---|---|---|
| Blue whale | 197 | 36 | 17 | 0.06 | 0.42 |
| Fin whale | 333 | 39 | 12.7 | 0.12 | 0.12 |
| Humpback whale | 57 | 29 | 8 | 0.01 | 0.79 |
| Right whale | 27 | 33 | 0.7 | 0.00 | 0.15 |
| Sei whale | 121 | 35 | 10 | 0.01 | 0.54 |

*The random forest classification models (one per species) were trained to discriminate between the acoustic activity of five baleen whale species using the training set consisting of 1-min clips (n = 695). n AM shows the number of predictors (AMs) that were used in the classification model according to the Boruta test results. OOB error stands for out-of-bag error {OOB error = [false positive (FP) + false negative (FN)/(true positive (TP) + true negative (TN) + FP + FN]; 1-overall accuracy}. Class 0 error represents the false omission rate [(FN)/(TN) + FN] and Class 1 error represents the false discovery rate [(FP)/(TP) + FP; 1-precision].*

**TABLE 3 |** Model results for the shelf site.

| Species | N | n AMs | OOB error (%) | Class 0 error | Class 1 error |
|---|---|---|---|---|---|
| Blue whale | 46 | 24 | 4 | 0.00 | 0.32 |
| Fin whale | 135 | 39 | 6 | 0.03 | 0.12 |
| Humpback whale | 112 | 36 | 9.7 | 0.05 | 0.21 |
| Minke whale | 52 | 37 | 5 | 0.01 | 0.30 |
| Right whale | 71 | 38 | 11.5 | 0.05 | 0.42 |
| Sei whale | 93 | 35 | 13 | 0.04 | 0.43 |

*The random forest classification models (one per species) were trained to discriminate between the six baleen whale species call types using the training set consisting of 1-min clips (n = 389). n AMs shows the number of predictors (AMs) that were used in the classification model according to the Boruta test results. OOB error stands for out-of-bag error [OOB error = (FP + FN)/(TP + TN + FP + FN); 1-overall accuracy]. Class 0 error represents the false omission rate [false negative (FN)/true negative (TN) + FN] and Class 1 error represents the false discovery rate [false positive (FP)/true positive (TP) + FP; 1-precision].*

species calls were not present. The false omission rate ranged from 0 to 5% on the shelf site, and 0 to 12% at the slope site, depending on species. Understanding when species are likely not to be acoustically present can be incredibly useful for marine soundscape planning, such as when trying to plan anthropogenic activities at times that could minimize acoustic disturbance to protected species (e.g., Van Opzeeland and Boebel, 2018). Being able to predict a species acoustic absence using rapid techniques can facilitate more efficient processing of large datasets.

At the shelf and slope sites, the models were also successful at predicting presence for a subset of our target species. The models performed the best for fin whales, predicting their presence with an 88% precision at both sites. This is similar performance to that reported for fin whale detectors in other studies (e.g., Buchan et al., 2019; Fregosi et al., 2020). The model also performed well for NARW at the slope site (15% false discovery rate) and humpback whales at the shelf site (21% false discovery rate). However, the ability to accurately predict presence decreased for the other species/site combinations, with false discovery rates ranging from 30 to 79%. In general, the random forest classification models performed somewhat better on the shallower shelf site as compared to the deep-water slope site. This may be due to a combination of differences in ambient noise backgrounds, propagation influence at the slope site based on hydrophone depth, as well as overall species occurrence. While the false discovery rates were lower for the shelf site than the slope site for blue, humpback, and sei whales, the model had a higher false discovery rate at the shelf site for NARW than the slope site. This could be attributed to the quality of signal in the NARW clips in the slope site even though they were only

acoustically present in around 4% of the slope clips. The low SR and minimum amount of overlap between target signals may well have improved our model performance in some instances. Our North Atlantic sites may be considered less "acoustically cluttered" compared with tropical waters; we did not have to contend with issues such as fish chorusing and snapping shrimp as previous studies did (e.g., Buxton et al., 2018). However, instead our region is subject to high anthropogenic noise levels (Rice et al., 2014), which impacted our model performance for species like humpback whales whose variety of vocalizations clearly overlap the same frequency ranges.

In our dataset, the metrics that contributed the most to species classification were those that summarized acoustic activity (intensity) and complexity in different frequency bands, such as AMP and ACI. In particular, important contributors to model performance were the AM computed across the lower frequency bands corresponding with the bands in which the respective species' call patterns show most of their acoustic energy (**Table 3**). Furthermore, for specific species/sites combinations, BI, Ht, and BL (see **Supplementary Table 1** for details) also showed high relative importance suggesting that the acoustic spectral and temporal heterogeneity, together with the background noise level were important drivers in the discrimination process between the species acoustic presence/absence. While it is interesting to note the relative importance of these AMs within the species-specific models, in the current study, the AMs are merely an automated way to parameterize and train the model when combined all together.

The role of acoustic indices, and AM in general, is essential to quantify the complexity of the soundscape in
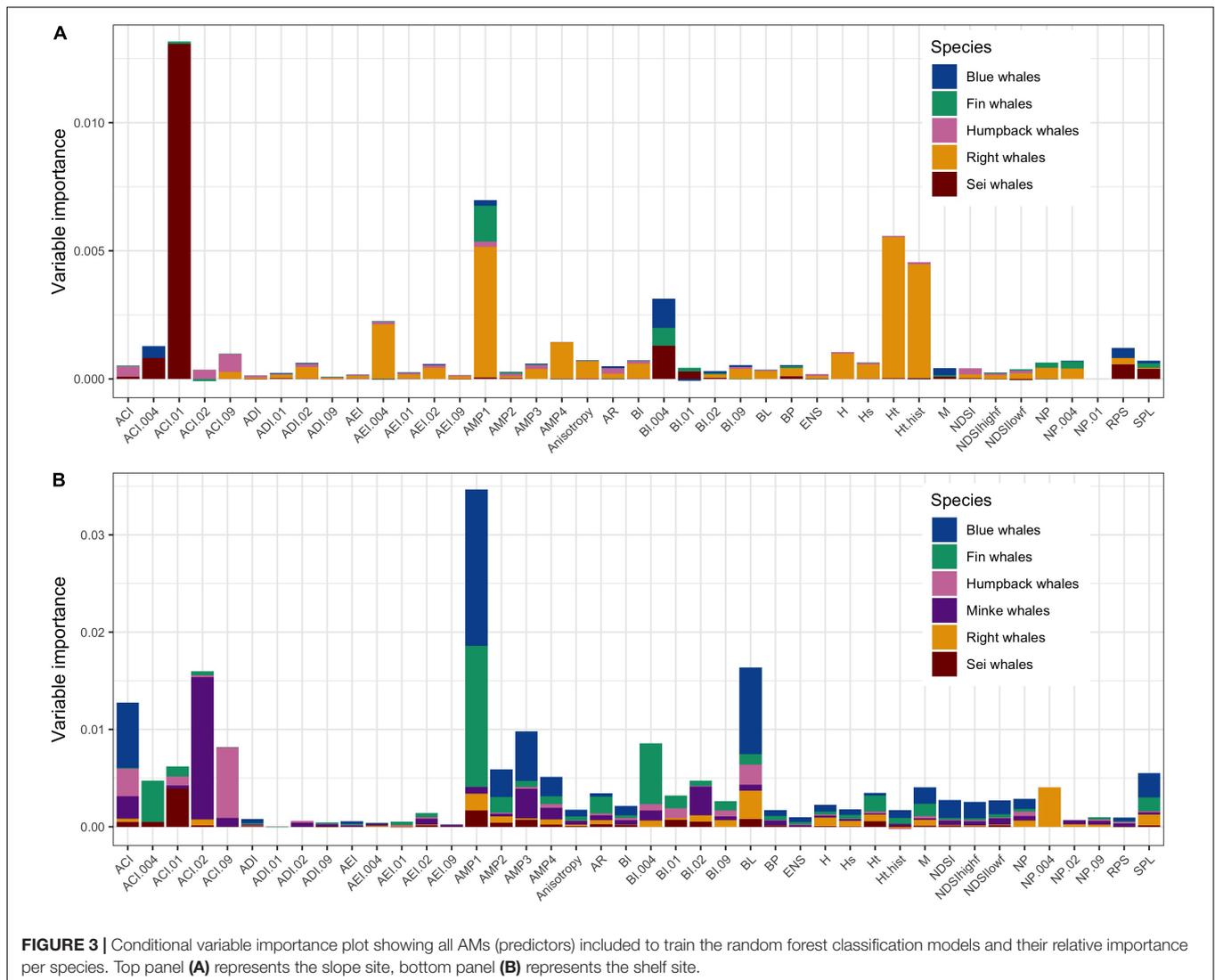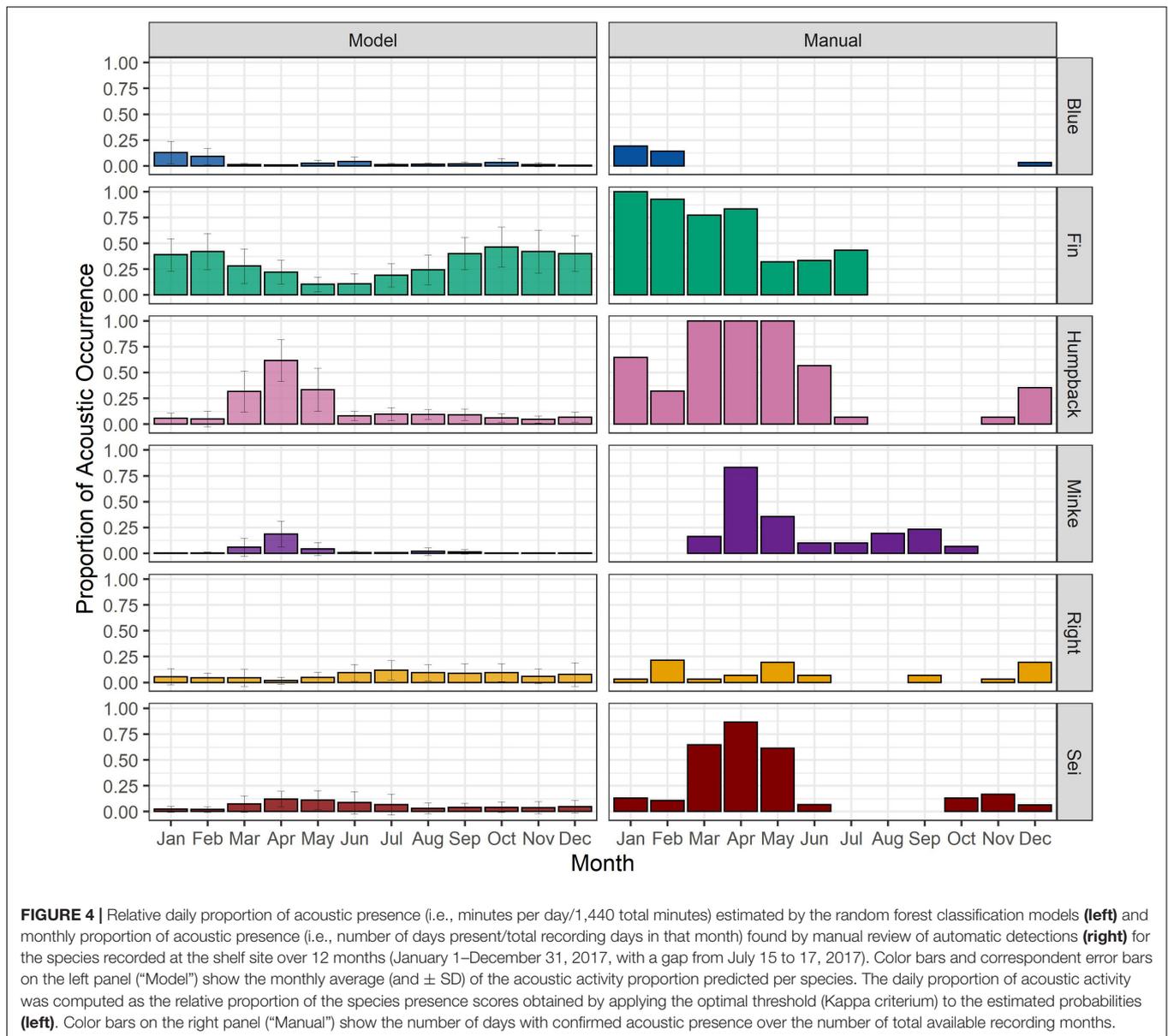
**FIGURE 3 |** Conditional variable importance plot showing all AMs (predictors) included to train the random forest classification models and their relative importance per species. Top panel **(A)** represents the slope site, bottom panel **(B)** represents the shelf site.

**TABLE 4 |** Results on second-step evaluation of classification model performance on 14 months of acoustic data from the shelf site.

| Species | n | Acc (95% CI) | Sensitivity | Specificity | Precision | NPV | Prevalence |
|---|---|---|---|---|---|---|---|
| Blue whale | 17 | 0.95 (0.92, 0.97) | 0.35 | 0.97 | 0.38 | 0.97 | 0.04 |
| Fin whale | 165 | 0.80 (0.76, 0.84) | 0.73 | 0.86 | 0.78 | 0.82 | 0.41 |
| Humpback whale | 57 | 0.92 (0.89, 0.95) | 0.72 | 0.96 | 0.75 | 0.95 | 0.14 |
| Minke whale | 11 | 0.98 (0.96, 0.99) | 0.73 | 0.99 | 0.67 | 0.99 | 0.03 |
| Right whale | 26 | 0.90 (0.87, 0.93) | 0.38 | 0.94 | 0.30 | 0.96 | 0.06 |
| Sei whale | 33 | 0.90 (0.87, 0.93) | 0.36 | 0.95 | 0.40 | 0.94 | 0.08 |

*Optimal thresholds to transform species predicted probabilities into presence/absence scores were selected to maximize Kappa. Overall accuracy values (Acc) are shown with the correspondent 95% confidence interval. Sensitivity (Recall) represents the ratio of correctly predicted presence observations to the total (true) presence observations in the dataset. Specificity represents the equivalent of the later but for absences. Precision (1-class 1 error) is the ratio of correctly predicted presence observations to the total predicted presence observations (denotes the accuracy of the predicted presences). NPV stands for negative predictive value (1-class 0 error). Species predicted prevalence represented their respective prevalence in the test set.*

a single or handful of values. However, their success has been variable, and it seems that a combination of AM (rather than a single metric) is more effective at predicting bioacoustic activity (Sueur and Farina, 2015; Mooney et al., 2020). Drawing from our results, the utility of these metrics

may well reflect and capture the complexity and stereotypy of individual species call types with NARWs and sei whales generally producing sporadic bouts of calls, while fin, blue, and minke whales produce song sequences which are more frequent and longer in duration (**Figure 2**; e.g., Stafford et al., 2007;

**FIGURE 4 |** Relative daily proportion of acoustic presence (i.e., minutes per day/1,440 total minutes) estimated by the random forest classification models **(left)** and monthly proportion of acoustic presence (i.e., number of days present/total recording days in that month) found by manual review of automatic detections **(right)** for the species recorded at the shelf site over 12 months (January 1–December 31, 2017, with a gap from July 15 to 17, 2017). Color bars and correspondent error bars on the left panel ("Model") show the monthly average (and ± SD) of the acoustic activity proportion predicted per species. The daily proportion of acoustic activity was computed as the relative proportion of the species presence scores obtained by applying the optimal threshold (Kappa criterium) to the estimated probabilities **(left)**. Color bars on the right panel ("Manual") show the number of days with confirmed acoustic presence over the number of total available recording months.

Risch et al., 2014; Cholewiak et al., 2018; Davis et al., 2017, 2020). In noisy, shallower waters where snapping shrimp choruses also dominate, the soundscape can be homogenized because of the wide frequency range, pervasiveness, and intensity of this characteristic sound. This acoustic context may challenge the use of AMs to train classification models to discriminate other concurring calls (e.g., fish). In contrast, acoustic contexts dominated by whale calls, or even song, as the ones studied at these sites, generally have a narrower band and less intense background and therefore, they may provide a promising route for using AM or an adapted version thereof as was shown in Parks et al. (2014).

Second-step evaluation of model performance in new data showed relatively high performance for three species: fin, humpback, and minke whales (**Table 4**). For these species, the

model performed well in predicting species presence as well as absence, similarly to the results of the first stage analysis. Overall predicted prevalence in the dataset for these three species ranged from 41% (fin whales) to 3% (minke whales; **Table 4**). However, the model performed worse for blue, NARW, and sei whales when applied to the entire dataset, with low rates of precision and sensitivity.

The comparison between model and detector outputs was aimed at understanding how the trained model would perform across a long time series (12 months of data), and whether it would provide outcomes that were comparable to the results obtained from the more time-consuming acoustic detection software approaches. When we averaged the model predictions and compared to the average daily presence across each month, the model predicted clear seasonal patterns of occurrence for all

species well (**Figure 4**). In comparison to our standard acoustic detection software process, the model provided much finer resolution and increased presence information. It is important to remember that the model provided information on species presence at the resolution of 1-min clips, while our standard acoustic detection approach provided species presence at the resolution of 24 h (daily). Therefore, although the model performed poorly when predicting certain species presence at the 1-min clip level, when aggregated across a day, it predicted similar patterns of occurrence as our standard acoustic detection software approach. However, there were time periods where the model accurately showed a larger proportion of daily acoustic presence for some species than the detector, like for fin whales during the summer and fall in 2017. Some of the discrepancies between the two methodologies could be explained by the missed detection rates of the detectors and the constrains of the manual verification process used (see section "Model Predictions and Second-Step Evaluation"). As mentioned in section "Model Predictions and Second-Step Evaluation", the detector methodologies were aimed at determining species' minimum acoustic presence. False negative rates for humpback, blue, fin, and sei whales were found to be 5, 10, 10, and 14%, respectively (Davis et al., 2020). Davis et al. (2017) found 31% of days where NARW were acoustically present were missed by the LFDCS detector. Lastly, the minke whale pulse train detector was found to have a 27% false negative rate (Risch et al., 2013). Because of these slight discrepancies, a detector method could be used in instances where the results aim to be more conservative, and species presence should be manually verified. In contrast, the AMs classification model can provide accurate species absence and presence information at daily-scale resolution with minimal human verification of the models' predictions. This demonstrates that this approach is valuable at this coarser reporting level.

Where the AM methodology really came into its own was both in the reduction of time needed to process large quantities of data by reducing the processing time from days to hours. In addition, the increased detection resolution that the AM model provides allows for a much more in-depth analyses of the data, allowing for tidal and diel trends to be evaluated in addition to having much finer resolution on species presence. The acoustic detector currently provides a manageable but time consuming and coarse resolution of information. Further improvements to this model are likely possible by adding higher resolution clips and more training data aimed at improving the positive detection accuracy and decreasing false detections. Also, in areas where the model and manual detector methods showed differing species presence proportions, clips could be browsed from those areas to confirm or deny species presence to further inform the model's test training datasets. However, results show that this methodology has promise to stand as an alternative or complementary analysis to current methods used for understanding large scale daily distribution of baleen whale species.

Overall, a relevant combination of AM in the context of a random forest classification modeling approach provides a promising methodology for less time consuming and laborious species detection for baleen whales based on the AM processes taking half as long as manually reviewing the detectors for these two sites. In addition, it could equally provide significantly higher resolution of information than what is currently available. Passive acoustic monitoring in the marine environment continues to rapidly expand as a methodology for understanding species occupancy and distributions to avoid human impacts and understand changing ocean environments for these species (e.g., Van Parijs et al., 2009; Mann, 2012; Rettig et al., 2013; Wall et al., 2013). With the end user in mind (Mooney et al., 2020) further development of this type of multi-step evaluation of AM could provide highly valuable advances in speed and improved resolution of data extraction.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## AUTHOR CONTRIBUTIONS

## FUNDING

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmars.2021.749802/full#supplementary-material

# REFERENCES

Baumgartner, M. F., and Mussoline, S. E. (2011). A generalized baleen whale call detection and classification system. *J. Acoust. Soc. Am.* 129, 2889–2902. doi: 10.1121/1.3562166

Boelman, N. T., Asner, G. P., Hart, P. J., and Martin, R. E. (2007). Multi-trophic invasion resistance in Hawaii: bioacoustics, field surveys, and airborne remote sensing. *Ecol. Appl.* 17, 2137–2144. doi: 10.1890/07-0004.1

Bohnenstiehl, D. R., Lyon, R. P., Caretti, O. N., Ricci, S. W., and Eggleston, D. B. (2018). Investigating the utility of ecoacoustic metrics in marine soundscapes. *J. Ecoacoustics* 2:56. doi: 10.22261/JEA.R1156L

Bonnell, J., Davis, G., Cholewiak, D., DeAngelis, A., Gerlach, D., Van Parijs, S., et al. (2016). *Guide to monitoring real-time marine mammal detections using autonomous platforms*. Falmouth, MA: Woods Hole Oceanographic Institution.

Bradfer-Lawrence, T., Gardner, N., Bunnefeld, L., Bunnefeld, N., Willis, S. G., and Dent, D. H. (2019). Guidelines for the use of acoustic indices in environmental research. *Methods Ecol. Evol.* 10, 1796–1807. doi: 10.1111/2041-210X.13254

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Buchan, S. J., Gutierrez, L., Balcazar-Cabrera, N., and Stafford, K. M. (2019). Seasonal occurrence of fin whale song off Juan Fernandez. *Chile. Endangered Species Res.* 39, 135–145. doi: 10.1007/s12526-020-01087-3

Buxton, R. T., McKenna, M. F., Clapp, M., Meyer, E., Stabenau, E., Angeloni, L. M., et al. (2018). Efficacy of extracting indices from large-scale acoustic recordings to monitor biodiversity. *Conserv. Biol.* 32, 1174–1184. doi: 10.1111/cobi.13119

Chase, J. M., and Knight, T. M. (2013). Scale-dependent effect sizes of ecological drivers on biodiversity: why standardised sampling is not enough. *Ecol. Lett.* 16, 17–26. doi: 10.1111/ele.12112

Cholewiak, D., Clark, C. W., Ponirakis, D., Frankel, A., Hatch, L. T., Risch, D., et al. (2018). Communicating amidst the noise: modeling the aggregate influence of ambient and vessel noise on baleen whale communication space in a national marine sanctuary. *Endangered Species Res.* 36, 59–75. doi: 10.3354/esr00875

Clark, C. W., Brown, M. W., and Corkeron, P. (2010). Visual and acoustic surveys for North Atlantic right whales, *Eubalaena glacialis*, in Cape Cod Bay, Massachusetts, 2001-2005: Management implications. *Mar. Mamm. Sci.* 26, 837–854. doi: 10.1111/j.1748-7692.2010.00376.x

Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C., Thornton, J. B., et al. (2017). Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. *Sci. Rep.* 7:13460. doi: 10.1038/s41598-017-13359-3

Davis, G. E., Baumgartner, M. F., Corkeron, P. J., Bell, J., Berchok, C., Bonnell, J. M., et al. (2020). Exploring movement patterns and changing distributions of baleen whales in the western North Atlantic using a decade of passive acoustic data. *Glob. Change Biol.* 26:15191. doi: 10.1111/gcb.15191

Debeer, D., and Strobl, C. (2020). Conditional permutation importance revisited. *BMC Bioinform.* 21:307. doi: 10.1186/s12859-020-03622-2

Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., and Sueur, J. (2012). Monitoring animal diversity using acoustic indices: implementation in a temperate woodland. *Ecol. Indic.* 13, 46–54. doi: 10.1016/j.ecolind.2011.05.006

Do Nascimento, L. A., Campos-Cerqueria, M., and Beard, K. H. (2020). Acoustic metrics predict habitat type and vegetation structure in the Amazon. *Ecol. Indic.* 117:106679. doi: 10.1016/j.ecolind.2020.106679

Farina, A. (2013). *Soundscape ecology: principles, patterns, methods and applications*. New York, NY: Springer Science & Business Media. doi: 10.1007/978-94-007-7374-5

Farina, A., Pieretti, N., and Piccoili, L. (2011). The soundscape methodology for long-term bird monitoring: a mediterranean europe case-study. *Ecol. Inform.* 6, 354–363. doi: 10.1016/j.ecoinf.2011.07.004

Farina, A., Righini, R., Fuller, S., Li, P., and Pavan, G. (2021). Acoustic complexity indices reveal the acoustic communities of the old-growth Mediterranean forest of Sasso Fratino Integral Natural Reserve (Central Italy). *Ecol. Indic.* 120:106927. doi: 10.1016/j.ecolind.2020.106927

Freeman, E. A., and Moisen, G. (2008). Presence absense: an r package for presence absense analysis. *J. Stat. Softw.* 23:11. doi: 10.18637/jss.v023.i11

Fregosi, S., Harris, D. V., Matsumoto, H., Mellinger, D. K., Negretti, C., Moretti, D. J., et al. (2020). Comparison of fin whale 20 Hz call detections by deep-water mobile autonomous and stationary recorders. *J. Acoustical Soc. Am.* 147, 961–977. doi: 10.1121/10.0000617

Gasc, A., Sueur, J., Pavoine, S., Pellens, R., and Grandcolas, P. (2013). Biodiversity sampling using a global acoustic approach: contrasting sites with microendemics in new caledonia. *PLoS ONE* 8:e65311. doi: 10.1371/journal.pone.0065311

Gibb, R., Browning, E., GloverKapfer, P., and Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methods Ecol. Evol.* 10, 169–185. doi: 10.1111/2041-210X.13101

Haver, S. M., Gedamke, J., Hatch, L. T., Dziak, R. P., Van Parijs, S., McKenna, M. F., et al. (2018). Monitoring long-term soundscape trends in US Waters: the NOAA/NPS Ocean noise reference station network. *Mar. Policy* 90, 6–13. doi: 10.1016/j.marpol.2018.01.023

Janitza, S., Strobl, C., and Boulesteix, A. L. (2013). An AUC-based permutation variable importance measure for random forests. *BMC Bioinform.* 14:119. doi: 10.1186/1471-2105-14-119

Kaplan, M. B., Mooney, T. A., Partan, J., and Solow, A. R. (2015). Coral reef species assemblages are associated with ambient soundscapes. *Mar. Ecol. Prog. Ser.* 533, 93–107. doi: 10.3354/meps11382

Kasten, E. P., Gage, S. H., Fox, J., and Joo, W. (2012). The remote environmental assessment laboratory's acoustic library: An archive for studying soundscape ecology. *Ecol. Inform.* 12, 50–67. doi: 10.1016/j.ecoinf.2012.08.001

Krause, B. L. (2008). Anatomy of the Soundscape: Evolving Perspectives. *J. Audio Eng. Soc.* 56, 73–80.

Kuhn, M. (2020). *caret: Classification and Regression Training. R package version 6.0-86*. https://CRAN.R-project.org/package=caret.

Kuhn, M., and Johnson, K. (2013). *Measuring Performance in Classification Models in Applied Predictive Modeling*. New York, NY: Springer, doi: 10.1007/978-1-4614-6849-3_11

Kursa, M. B., and Rudnicki, W. R. (2010). Feature Selection with the Boruta Package. *J. Stat. Softw.* 36, 1–13. doi: 10.18637/jss.v036.i11

Lesage, V., Gavrilchuk, K., Andrews, R. D., and Sears, R. (2017). Foraging areas, migratory movements and winter destinations of blue whales from the western North Atlantic. *Endangered Species Res.* 34, 27–43. doi: 10.3354/esr00838

Liaw, A., and Wiener, M. (2002). Classification and Regression by randomForest. *R News* 2, 18–22.

Mann, D. A. (2012). Remote sensing of fish using passive acoustic monitoring. *Acoustics Today* 8, 8–15. doi: 10.1121/1.4753916

Mooney, T. A., Di Iorio, L., Lammers, M., Lin, T. H., Nedelec, S. L., Parsons, M., et al. (2020). Listening forward: Approaching marine biodiversity assessments using acoustic methods. *Royal Soc. Open Sci.* 7:1287. doi: 10.1098/rsos.201287

Morano, J. L., Salisbury, D. P., Rice, A. N., Conklin, K. L., Falk, K. L., and Clark, C. W. (2012). Seasonal and geographical patterns of fin whale song in the western North Atlantic Ocean. *J. Acoustical Soc. Am.* 132, 1207–1212. doi: 10.1121/1.4730890

Nieukirk, S. L., Mellinger, D. K., Moore, S. E., Klinck, K., Dziak, R. P., and Goslin, J. (2012). Sounds from airguns and fin whales recorded in the mid-Atlantic Ocean, 1999-2009. *J. Acoustical Soc. Am.* 131, 1102–1112. doi: 10.1121/1.3672648

Oleson, E. M., Barlow, J., Gordon, J., Rankin, S., and Hildebrand, J. A. (2003). Low frequency calls of Bryde's whales. *Mar. Mamm. Sci.* 19, 407–419. doi: 10.1111/j.1748-7692.2003.tb01119.x

Parks, S. E., Miksis-Olds, J. L., and Denes, S. L. (2014). Assessing marine ecosystem acoustic diversity across ocean basins. *Ecol. Inform.* 21, 81–88. doi: 10.1016/j.ecoinf.2013.11.003

Pekin, B. K., Jung, J., Villanueva-Rivera, L. J., Pijanowski, B. C., and Ahumada, J. A. (2012). Modeling acoustic diversity using soundscape recordings and LIDAR-derived metrics of vertical forest structure in a neotropical rainforest. *Landscape Ecol.* 27, 1513–1522. doi: 10.1007/s10980-012-9806-4

Pieretti, N., Farina, A., and Morri, D. (2011). A new methodology to infer the singing activity of an avian community: the acoustic complexity index (ACI). *Ecol. Indic.* 11, 868–873. doi: 10.1016/j.ecolind.2010.11.005

Pieretti, N., Lo Martire, M., Farina, A., and Danovaro, R. (2017). Marine soundscape as an additional biodiversity monitoring tool case: A case study from the Adriatic Sea. *Ecol. Indic.* 83, 13–20. doi: 10.1016/j.ecolind.2017.07.011

Pijanowski, B. C., Farina, A., Gage, S. H., Dumyahn, S. L., and Krause, B. L. (2011). What is soundscape ecology? An introduction and overview of an emerging

new science. *Landscape Ecol.* 26, 1213–1232. doi: 10.1007/s10980-011-9600-8

Rettig, S., Boebel, O., Menze, S., Kindermann, L., Thomisch, K., and van Opzeeland, I. (2013). "Local to basin scale arrays for passive acoustic monitoring in the Atlantic sector of the Southern Ocean".in *Poster presented at: International Conference and Exhibition on Underwater Acoustics*. Corfu Island.

Rice, A. N., Tielens, J. T., Estabrook, B. J., Muirhead, C. A., Rahaman, A., Guerra, M., et al. (2014). Variation of ocean acoustic environments along the western north Atlantic coast: a case study in context of the right whale migration route. *Ecol. Inform.* 21, 89–99. doi: 10.1016/j.ecoinf.2014.01.005

Risch, D., Castellote, M., Clark, C. W., Davis, G. E., Dugan, P. J., Hodge, L. E. W., et al. (2014). Seasonal migrations of North Atlantic minke whales: novel insights from large-scale passive acoustic monitoring networks. *Mov. Ecol.* 2, 24–24. doi: 10.1186/s40462-014-0024-3

Risch, D., Clark, C. W., Dugan, P. J., Popescu, M., Siebert, U., and Van Parijs, S. M. (2013). Minke whale acoustic behavior and multi-year seasonal and diel vocalization patterns in Massachusetts Bay, USA. *Mar. Ecol. Prog. Ser.* 489, 279–295. doi: 10.3354/meps10426

Roca, I. T., and Van Opzeeland, I. (2019). Using acoustic metrics to characterize underwater acoustic biodiversity in the Southern Ocean. *Rem. Sens. Ecol. Conserv.* 6, 262–273. doi: 10.1002/rse2.129

Stafford, K. M., Mellinger, D. K., Moore, S. E., and Fox, C. G. (2007). Seasonal variability and detection range modeling of baleen whale calls in the Gulf of Alaska, 1999–2002. *J. Acoustical Soc. Am.* 122, 3378–3390. doi: 10.1121/1.2799905

Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., and Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinform.* 9:307. doi: 10.1186/1471-2105-9-307

Sueur, J., and Farina, A. (2015). Ecoacoustics: the ecological investigation and interpretation of environmental sound. *Biosemiotics* 8, 493–502. doi: 10.1007/s12304-015-9248-x

Sueur, J., Farina, A., Gasc, A., Pieretti, N., and Pavoine, S. (2014). Acoustic indices for biodiversity assessment and landscape investigation. *Acta Acustica United Acustica* 100, 772–781. doi: 10.3813/AAA.918757

Sueur, J., Pavoine, S., Hamerlynck, O., and Duvail, S. (2008). Rapid acoustic survey for biodiversity appraisal. *PloS one* 3:e4065. doi: 10.1371/journal.pone.0004065

Towsey, M. (2017). *The Calculation of Acoustic Indices Derived from Long-Duration Recordings of the Natural Environment*. Brisbane, QLD: QUT ePrints, 1–11.

Tucker, D., Gage, S. H., Williamson, I., and Fuller, S. (2014). Linking ecological condition and the soundscape in fragmented Australian forests. *Landscape Ecol.* 29, 745–758. doi: 10.1007/s10980-014-0015-1

Van Opzeeland, I., and Boebel, O. (2018). Marine soundscape planning: seeking acoustic niches for anthropogenic sound. *J. Ecoacoustics* 2:8. doi: 10.22261/JEA.5GSNT8

Van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., et al. (2009). Management and research applications of real-time and archival passive acoustic sensors over varying temporal and spatial scales. *Mar. Ecol. Prog. Ser.* 395, 21–36. doi: 10.3354/meps08123

Villanueva-Rivera, L. J., Pijanowski, B. C., Doucette, J., and Pekin, B. (2011). A primer of acoustic analysis for landscape ecologists. *Landscape Ecol.* 26:1233. doi: 10.1007/s10980-011-9636-9

Villanueva-Rivera, L. J., Pijanowski, B. C., and Villanueva-Rivera, M. L. J. (2018). *Package 'soundecology'. R package version*, 1. http://cran.ms.unimelb.edu.au/web/packages/soundecology/soundecology.pdf

Wall, C. C., Simard, P., Lembke, C., and Mann, D. A. (2013). Large-scale passive acoustic monitoring of fish sound production on the West Florida Shelf. *Mar. Ecol. Prog. Ser.* 484, 173–188. doi: 10.3354/meps10268

Weiss, S. G., Cholewiak, D., Frasier, K. E., Trickey, J. S., Baumann-Pickering, S., Hildebrand, J. A., et al. (2021). Monitoring the acoustic ecology of the shelf break of Georges Bank, Northwestern Atlantic Ocean: New approaches to visualizing complex acoustic data. *Mar. Policy* 130:4570. doi: 10.1016/j.marpol.2021.104570

Wiggins, S. M., and Hildebrand, J. A. (2007). "High-frequency Acoustic Recording Package (HARP) for broad-band, long-term marine mammal monitoring". *International Symposium on Underwater Technology 2007 and International Workshop on Scientific Use of Submarine Cables & Related Technologies 2007*. Tokyo: Institute of Electrical and Electronics Engineers.

Wright, M. N., and Ziegler, A. (2015). ranger: A fast implementation of random forests for high dimensional date in c++ and r. *J. Stat. Softw.* 77, 1–17. doi: 10.18637/jss.v077.i01