

Open-Source machine learning BANTER acoustic classification of a cryptic echolocating species

Shannon Rankin Taiki Sakai Frederick Archer Jay Barlow
Danielle Cholewiak Annamaria DeAngelis
Jennifer L. K. McCullough Erin Oleson Anne Simonis
Melissa Soldevilla Jenny Trickey

6/7/23

Passive acoustic monitoring is increasingly used for assessing populations of marine mammals; however, analysis of large datasets is limited by our ability to easily classify sounds detected. Classification of beaked whale acoustic events, in particular, require evaluation of multiple lines of evidence by expert analysts. Here we present a highly automated approach to acoustic detection and classification using supervised machine learning and open source software methods. Data from four large scale surveys of beaked whales (North Atlantic, South Atlantic, Hawaii, and Eastern Pacific) were analyzed PAMGuard (acoustic detection), PAMPal (acoustic analysis) and BANTER (hierarchical random forest classifier). Overall correct classification results ranged from 87% for the South Atlantic data to 98% for the North Atlantic. Results for many species could likely be improved with increased sample sizes, consideration of alternative automated detectors, and addition of relevant environmental features. These methods provide a highly automated approach to acoustic detection and classification using open source methods that can be readily adopted for species and geographic regions.

Introduction

Passive acoustic monitoring (PAM) has proven to be a valuable tool for studying populations of marine mammals ([Parijs et al. 2009](#)); however, the value of these studies depends on our ability to identify the sources of the sounds we are monitoring. Historically, experienced acousticians manually identify stereotyped sounds that could be reliably attributed to a given

species, based on their spectral or temporal characteristics ([Baumann-Pickering et al. 2013](#)); ([Bittle and Duncan 2013](#)); ([Rankin and Barlow 2005](#)); ([Soldevilla et al. 2008](#)). The dramatic increase in recordings make it impossible for experienced acousticians to manually annotate all data, and there has been an increasing need for the development of automated classification routines that can provide accurate determinations of the species responsible for the sounds.

Beaked whales are deep diving marine mammals found in offshore waters; their long dive intervals and cryptic surfacing behavior make them difficult to study using typical shipboard visual observation methods ([MacLeod 2018](#)). However, beaked whales make stereotyped echolocation clicks that have temporal and spatial characteristics that can be used to differentiate species ([Baumann-Pickering et al. 2014](#)). Detection of beaked whale signals in large datasets is greatly expanding our understanding of their population structure and the potential impact of human activities on these species ([Barlow, Moore, et al. 2022](#)); ([Baumann-Pickering et al. 2014](#)); ([Simonis 2020](#)). Unfortunately, manual classification of beaked whale echolocation clicks requires analysis by trained observers using a number of visual aids to examine the call characteristics. Development of automated classification routines, if accurate, serve to improve the efficiency and decrease the cost of analyzing large datasets.

BANTER is a supervised machine learning acoustic event classifier originally developed for classifying dolphin schools from acoustic recordings ([Rankin et al. 2017](#)). This hierarchical random forest event classifier relies on an initial call classifier for each call type (in the dolphin case study, this included one each for whistles, echolocation clicks, and burst pulses) followed by an event classifier. The event classifier considers the probability of assignment to species for each call in the call classifier, along with any event-level characteristics (such as call rate). BANTER is very flexible and can accommodate any number of measures from any number of detectors. While BANTER relies on multiple call types, relatively minor changes to a call detector could yield different detector results ([Rankin et al. 2017](#)). In this case, settings for a whistle & moan detector were modified to improve performance at detecting burst pulses. This approach of applying multiple call detectors (of the same type but with different settings) suggests that perhaps BANTER could be used on species that only produce one call type (or where only one call type was analyzed), as long as the settings of the automated call detector were modified such that the overall number of detections was different. This approach was successfully applied to classified echolocation clicks for narwhal and belugas ([Zahn et al. 2021](#)). For species who primarily (or exclusively) produce echolocation clicks, BANTER may serve as an option for automated machine learning classification.

Here we applied BANTER acoustic classification to beaked whale detections from four large acoustic datasets including the US East Coast (north and south), Hawaii, and the US West Coast. Acoustic data were analyzed by experienced acousticians to determine species identity, and these ‘ground truth’ classifications served as training data for the supervised BANTER trials. Our goals are to identify an efficient and accurate automated approach to acoustic classification of beaked whales, and to provide a framework for analysis that may serve for other PAM studies.

Materials and Methods

Field Data Collection

Passive acoustic data was collected during three surveys conducted by National Oceanographic and Atmospheric Administrations' research operations off the US East Coast, the Hawaiian Islands (Hawaii) and the US West Coast.

Recordings from the western North Atlantic Ocean (U.S. east coast) were collected using towed hydrophone arrays during the 2016 AMAPPS II (Atlantic Marine Assessment Program for Protected Species) survey (cite). AMAPPS II was subdivided into the northern (NAtlantic) and southern (SAtlantic) survey area. The NAtlantic study area ranged from Massachusetts south to New Jersey (HB1603, [CITE]) and recordings used in this analysis consisted of a single hydrophone (HTI-96-min, High Tech Inc., Long Beach, MS) recorded at 192 kHz sample rate with a 1 kHz high pass filter (National Instruments USB-6356 A/D card, see ([DeAngelis et al. 2018](#)) for more information). The SAtlantic area ranged from Delaware south to central Florida (GU1605, [CITE]) and recordings used in this analysis consisted of a single hydrophone (Reson [MODEL], Teledyne Marine, Slangerup, Denmark) recorded 500 kHz sample rate (soundcard) and decimated to 192 kHz with a 1 kHz high pass filter (see [CITE] for more information). Variation in hydrophone sensitivities precluded combining the NAtlantic and SAtlantic datasets.

Recordings from Hawaii were collected using drifting acoustic recorders during the 2017 HICEAS (Hawaiian Island Cetacean Ecosystem Assessment Survey, ([Yano et al. 2018](#))). Drifting acoustic recorders were deployed at randomly selected locations within the main Hawaiian Islands. Two types of drifting recorders were included: (1) Soundtrap ST4300 recorders (Ocean Acoustics, Auckland, New Zealand) and (2) Wildlife Acoustics SM3M recorders (Wildlife Acoustics, Maynard, MA). Soundtrap recorders were sampled at 288 kHz with a duty cycle of 2 min on for every 10 minutes. SM3M data were sampled at 256 kHz with continuous recordings. Individual buoys were deployed for 10 to 50 days (see ([Yano et al. 2018](#)) for more detailed information).

Recordings from the Eastern North Pacific Ocean off the U.S. west coast (EPacific) were collected using drifting acoustic recorders during the 2018 CCES (California Current Ecosystem Survey) survey ([Simonis 2020](#)). Buoys were deployed at 23 stations distributed throughout the California Current region; data from 15 of these included high quality acoustic data used in this analysis. Two types of drifting recorders were included: (1) Soundtrap ST4300 recorders (Ocean Acoustics, Auckland, New Zealand) and (2) Wildlife Acoustics SM3M recorders (Wildlife Acoustics, Maynard, MA). Recordings were duty cycled and recorded 2 minutes out of a variable 'off' time; recorders sampled at a minimum 256 sample rate. Individual buoys were deployed for 5-79 days (see ([Simonis 2020](#)) for more detailed information).

Acoustic Data Analysis

All acoustic recordings were analyzed using PAMGuard software (version 2.00.15c, 2.00.16e) with a suite of generic click classifiers within the Click Detector module (see ([Keating and Barlow 2013](#))). Classifier sets were saved such that any click may be classified as more than one click type. The general click classifiers were effectively click detectors, as each implemented a simple spectral band settings (e.g., 2 – 15 kHz, 15 – 30 kHz, 30 – 50 kHz, 50 – 80 kHz, and > 80 kHz). An additional detector within the 30 – 50 kHz peak frequency range considered the presence of a frequency sweep that is characteristic of beaked whale pulses (see ([Keating and Barlow 2013](#))).

Acoustic detection of beaked whale events were then classified to species by experienced acoustic researchers using multiple lines of evidence. If beaked whales were sighted in close proximity to the recording device, the visual confirmation of species identify was used to determine species identity. Otherwise, experienced acousticians followed strict protocol to define species identify based on acoustic characteristics. When the available characteristics were inconclusive, the species was considered an unidentified beaked whale; only acoustic events classified to species were considered for this study. Acoustic researchers with experience in classifying beaked whale echolocation clicks examined the bearing-time display for the click detector module in PAMGuard to assess the species identity of the acoustic event. For NEFSC data, acoustic detections of beaked whales using towed hydrophone arrays were further subdivided into click trains based on identification of consecutive clicks along the same bearing angle in the bearing time plot (see ([DeAngelis et al. 2018](#))). Data were saved within the PAMGuard database and binaries for downstream processing using PAMPal package ([Sakai 2021](#)) in R programming language ([Team 2022](#)).

Standard click calculations were measured from echolocation clicks using default values in PAMPal. Inter-click interval (ICI) was calculated using the ‘calculateICI’ function in PAMPal; the mode of the ICI was identified for each beaked whale acoustic event. In addition, for EPacific data, we used the ‘matchEnvData’ function in PAMPal to include ERDAPP environmental data ([Simons and John \(2022\)](#)), including seafloor depth and seafloor gradient (both from erdSrtm30plusSeafloorGradient), as well as sea surface temperature (from jplMURSST41).

Multiple trials were conducted to better understand the value (and potential limitations) of the different suite of measurements. Data subsets for each trial were exported for BANTER classification using the export_banter function in PAMPal. Trials included standard click calculations for all clicks in a beaked whale acoustic event (EC); the addition of ICI for each event (EC_ICI), and for the EPacific dataset we also included environmental variables (EC_ICI_ENV).

BANTER Acoustic Classification

BANTER models were created for each trial using the BANTER package in R programming language ([Archer 2022a](#)). Standard click calculations were used for the BANTER call classifier; event level measures of ICI and environmental variables were applied to the BANTER event classifier. A minimum of two events per species and the complete suite of variables (absence of NAs) were required for training and testing of the classification model. The BANTER random forest models contain two parameters, ntree and sampsize, which must be selected to ensure stability of the model. The sample size (sampsize) is the number of events considered for each trial of the model, and the ntree is the number of trees (iterations of the model) run in a given model. These values were modified to obtain the best results, with an emphasis on lower sample size and higher number of trees for computational performance ([Rankin and Archer 2021](#)).

The model output include the confusion matrix, and a number of visualizations provided by the rfpermutation package in R programming language ([Archer 2022b](#)).

Results

Four study areas (NAtlantic, SAtlantic, Hawaii, EPacific) were analyzed for this study; species included: Baird's beaked whales (*Berardius bairdii*), Blainville's beaked whales (*Mesoplodon densirostris*), Cross Seamount beaked whale (McDonald et al. ([2009](#))), Cuvier's beaked whales (*Ziphius cavirostris*), Gervais' beaked whales (*Mesoplodon europaeus*), Hubb's beaked whale (*Mesoplodon carlhubbsi*), Longman's beaked whale (*Indopacetus pacificus*), Sowerby's beaked whales (*Mesoplodon bidens*), Stejneger's beaked whale (*Mesoplodon stejnegeri*), True's beaked whales (*Mesoplodon mirus*), and unidentified beaked whale 'BW43' ([Baumann-Pickering et al. 2013](#)). Analyses were conducted separately for the different study areas due to differences in the data collection methods that precluded combining of datasets. Specific species combinations and sample sizes varied by study area.

NAtlantic

Species encountered in the NAtlantic dataset include Cuvier's beaked whales (n = 120), Gervais' beaked whales (n = 4), Sowerby's beaked whales (n = 6), and True's beaked whales (n = 76). BANTER classification trials included (1) echolocation clicks (EC), and (2) echolocation clicks and ICI (EC_ICI).

For the NAtlantic dataset, the model with ICI had a modest increase in accuracy over the EC (only) model (EC_ICI model, Detector Model: sampsize = 3, ntrees = 10,000; Event Model: sampsize = 3, trees = 100,000). This model provided an overall correct classification rate of 98% for all four species (pct.correct in Confusion Matrix, Figure 1 a). Classification scores ranged from 75% for Gervais' beaked whales to 100% for Cuvier's and Sowerby's beaked

whales. All classification results were greater than the expected error rate (Priors in Figure 1 a). Results from the EC_ICI model are presented here; results from the EC model were similar and can be found in supplementary materials (Supplement Fig. 1).

The proximity plot (Figure 1 b) provides a view of the distribution of events within the feature space of the two ‘most important’ dimensions, or predictors. For each event in the plot, the color of the central dot represents the true species identity, and the circle represents the BANTER classification result. The importance heat map (Figure 1 c) shows the most important features for predicting each species. The strength of the classifications by species can be seen in the plot votes (Figure 1 d), which show the distribution of the votes for each of the 100,000 trials for each event. For each plot, the events are represented as vertical slices (with plot subdivided along the x-axes according to the event sample size), and the percentage of votes for each species is represented by their color along that vertical slice. In a perfect scenario, all 100,000 trees would have voted for the correct species and therefore the classification would be correct and the strength of these classifications would be maximized.

The proximity plot (Figure 1 b) shows overlap between Gervais’, Sowerby’s and True’s beaked whales in this two dimension feature space and the species classifications are resolved using additional features, as shown in the importance heat map (Figure 1 c). The plot vote figure (Figure 1 d) shows strong classification strength for Cuvier’s and Sowerby’s beaked whales, with increased variability in voting distributions for Gervais’ and True’s beaked whales.

SAtlantic

Species encountered in the SAtlantic study dataset include Blainsville’s beaked whales ($n = 46$), Cuvier’s beaked whales ($n = 111$), Gervais’ beaked whales ($n = 77$), Sowerby’s beaked whales ($n = 2$), and True’s beaked whales ($n = 13$). BANTER classification trials included (1) echolocation clicks (EC), and (2) echolocation clicks and ICI (EC_ICI).

For the SAtlantic dataset, the most accurate model considered the echolocation clicks and the ICI for an event (EC_ICI Model, Detector Model: sampsize = 4, ntrees = 10,000; Event Model: sampsize = 1, ntrees = 100,000) and provided an overall correct classification rate of 87.1% for all five species. Classification scores ranged from a low of 50% (Sowerby’s beaked whale) to a high of 100% (True’s beaked whale, Figure 2); all classification results were greater than the expected error rate (Priors in Figure 1 a). The sample sizes for the BANTER model was kept low (sampsize = 1) to retain Sowerby’s beaked whale, which had a sample size of 2. Analysis with a larger sample size resulted in improved classification results for Cuvier’s and Gervais’ beaked whale (see Supplement Fig. 3 for EC_ICI_ALT), but resulted in the loss of Sowerby’s in the final model. Results from the EC_ICI model are presented here; results from the EC and EC_ICI_ALT model can be found in supplementary materials (Supplement Fig. 2).

For SAtlantic, the two most important features are insufficient to differentiate species, as seen by the overlap in species distribution with these feature spaces in the proximity plot

(Figure 2 b). Importance variables vary by species (importance heat map Figure 2 c), with ICI measurements having relatively low importance for Blainsville's and Sowerby's beaked whales (Figure 2 c). The plot votes graph (Figure 2 d) shows relatively weak classification results; most events consisted of <50% of the trees voting for the correct species.

Hawaii

The Hawaii dataset consisted of 13 drifting buoys deployed within the main Hawaiian Islands (Yano et al. (2018)); species included Blainsville's beaked whale ($n = 521$), Cross Seamount beaked whale ($n = 76$), Cuvier's beaked whale ($n = 201$), and Longman's beaked whale ($n = 122$). BANTER classification trials included (1) echolocation clicks (EC), and (2) echolocation clicks and ICI (EC_ICI).

The most accurate model considered the echolocation clicks only (EC Model, Detector Model: sampsize = 10, ntrees = 5,000; Event model: sampsize = 4, ntree = 10,000) and provided an overall correct classification rate of 92.3% for all four species. Classification scores ranged from a low of 86% (Cuvier's Beaked Whale) to 97.5% (Longman's beaked whale, Figure 3). All classification results were greater than the expected error rate (Prior in Figure 3 a). Results from the EC model are presented here; results from EC_ICI model can be found in supplementary materials (Supplement Fig. 4).

Distinctive clusters based on the two most important dimensions can be found for Longman's and Blainsville's beaked whales, and there is overlap in this feature space of the clusters for Cuvier's and Cross Seamount beaked whales as shown in the Proximity Plot (Figure 3 b). This overlap is resolved with different importance features for Cuvier's and Cross Seamount beaked whales as seen in the Importance Heat Map (Figure 3 c), resulting in the low misclassification rates for these species (Figure 3 a confusion matrix). The plot votes graph (Figure 3 d) show strong classification results for most Blainsville's, Cross Seamount, and Longman's beaked whales, with lower classification strength for many Cuvier's beaked whale events.

Eastern Pacific

The EPacific dataset consisted of 15 drifting recording buoys deployed off the west coast of the United States (Simonis 2020); species included Baird's beaked whale ($n = 29$), Unidentified Beaked Whale 'BW43' ($n = 125$), Cross Seamount beaked whale ($n = 6$), Cuvier's beaked whale ($n = 926$), Hubbs beaked whale ($n = 66$) and Stejneger's beaked whale ($n = 42$). BANTER classification trials included (1) echolocation clicks (EC), (2) echolocation clicks and ICI (EC_ICI), and (3) echolocation clicks, ICI, and environmental features (EC_ICI_ENV).

The best model considered the echolocation clicks, ICI, and environmental features (EC_ICI_ENV Model, Detector Model: sampsize = 3, ntrees = 10,000; Event model: sampsize = 4, ntree = 10,000) and provided an overall correct classification rate of 92.2% for all six species. Classification scores ranged from a low of 91.2% (BW43) to 100% for Baird's

beaked whales (Figure 4 a). All classification results were greater than the expected error rate (Prior in Figure 4 a). Results from the EC_ICI_ENV model are presented here; results from the EC and EC_ICI model can be found in supplementary materials (Supplement Figs. 5).

There is considerable overlap in the primary feature space for all but BW43, as shown in the Proximity Plot (Figure 4 b). ICI was among the most important classification variable for all species (Figure 4 c heat map). The plot votes (Figure 4 d) show relatively strong classification results Baird's, Hubbs, and Stejneger's beaked whales.

Discussion

The results for the four study areas varied from a low of 87.1% correct classification in the SAtlantic dataset to a high of 98% correct classification in the NAtlantic dataset. All species in all datasets had classifications rates well above those expected by chance; however Cuvier's (SAtlantic, Hawaii), Sowerby's (SAtlantic), and Gervais beaked whales (NAtlantic) had classification rates below 90%.

Several of our datasets included very low sample sizes for some species; however, low sample sizes did not always result in low classification scores. In the NAtlantic dataset, Sowerby's beaked whale ($n=6$) was correctly classified in 100% of the samples, and classification results were very strong. The Sowerby's beaked whale events in the SAtlantic study and the Gervais beaked whale events in the NAtlantic study consisted of weak classifications. This is likely due to both a small number of events ($n = 2$ for Sowerby's in SAtlantic and $n = 4$ for Gervais in NAtlantic) combined with the small sampsize used in the BANTER model. In general, larger sample sizes should lead to improved overall classification results, and improved strength of these classification results (as identified by plot vote graph).

While some species have distinctive differences in their click characteristics that lead to strong classification despite small sample sizes (e.g. Sowerby's beaked whales in NAtlantic); other species, such as Cuvier's beaked whales, have significant variation in their click measurements. For species with high variability in the predictor variables and an overlap in the range of these variables with other species, a large sample size is required to describe the true variability of these call measurements. Even with reasonably large sample sizes, the classification may suffer (e.g, SAtlantic Cuvier's beaked whale = 79.2% correct classification). We found by increasing the sampsize in BANTER, we could increase the classification rate for this species in this area to (91.8%, see Supplement Fig.3); however, this resulted in the loss of Sowerby's beaked whale in the final model. So, an increased sample size in Sowerby's would likely lead to increased classification results for other species in the model.

Previous application of BANTER to dolphin species in the California Current found that large sample sizes could result in strong classification of species where experienced acousticians are unable to differentiate species (long-beaked and short-beaked common dolphins, (Rankin et al.

2017)). This suggests that large increases in sample sizes may improve classification results for Cuvier's beaked whales.

Researchers have noted several 'types' of click characteristics for Cuvier's beaked whales (ADA, pers. comm) which may be related to variability in the propagation environment, different animal behaviors, gender or age classes, or acoustic populations. Future work should examine if these Cuvier's beaked whale events can be clustered in an objective and stable manner; if so, then these can be input into BANTER models as separate types of Cuvier's beaked whales. We recommend caution in this approach- this should only be considered if there is reasonable evidence that these different groups represent authentic acoustic groups. We do not recommend segregation of misclassified subgroups simply to improve classification scores.

Gervais' and True's beaked whales have similar click characteristics that make them difficult to differentiate and they require considerable expertise to classify manually (DeAngelis et al. 2018). Our results suggest that BANTER may serve as an efficient and effective means of classifying and differentiating these two species. Despite modest sample sizes, Gervais' and True's beaked whales showed high classification scores in the SAtlantic (93.5% and 100%, respectively), and True's performed well in the NAtlantic. Unfortunately, we were unable to combine the N and S Atlantic datasets (which would increase sample sizes of True's beaked whales) due to differences in hydrophone characteristics that result in differences in call measures. Calibration of signals (to make them comparable) may allow for combining datasets to improve sample sizes.

In addition to poor classification of Gervais in the NAtlantic study area, the plot votes show that for one of the four events, only a few of the 100,000 trees 'voted' for Gervais, and the majority of the votes were for True's beaked whales. Analysis notes indicate uncertainty in manual species classification of one of the events in this dataset. While small sample size can at times provide good classification scores, as we found in Sowerby's beaked whale in this SAtlantic dataset, they can be devastating if there are inaccuracies in the training data. These two examples of small sample sizes highlights a conflict in preferred protocol. Ideally, (1) all species would be included in a classification model, so as to better represent the local species diversity, (2) all events would ideally be included in the classification model, so as to better represent the variability found in the area, and (3) only confident 'ground truth' classifications would be considered in the training model. Unfortunately, in species where identity must be determined based on call characteristics (rather than visual confirmation of species identity in the field), it can sometimes be difficult to confidently determine species identity when calls are highly variable and do not include at least one call that provides high confidence of species identity. In the case of beaked whales, we recommend that training data include a reasonable level of confidence for inclusion. This is especially critical for species with small sample sizes. An alternative is to require agreement from multiple analysts, as provided by the EPacific datasets. This concern about accurate labeling of training data is further complicated by the potential existence of species that have not yet been identified. For example, (Barlow, Cárdenas-Hinojosa, et al. 2022) recently discovered what appears to be a new species of

beaked whale off Baja California, Mexico; this putative new species may have been detected but misclassified in other datasets (including the EPacific data presented here).

Interclick Interval (ICI) was found to be the most important variable for NAtlantic and EPacific, and ICI were the third and fourth ranked variable for SAtlantic (see Importance Heat Maps for each survey area). While the addition of ICI did not improve results Hawaii, results were similar and ICI ranked #9 in importance for the Hawaii study area. There was a difference in how ICI was calculated for these surveys, and it is possible these differences may affect the value of the ICI variable in species classification for some species. Beaked whale events were defined by encounter time for both the Hawaii and EPacific data, which were collected using drifting acoustic recorders where animals could not be localized. The Atlantic data were collected using a towed hydrophone array with localization capabilities. Each event was defined as an individual click train, so ICI for these events more likely represented true ICI. ([Baumann-Pickering et al. 2013](#)) showed an overlap in the ICI between Cuvier's and Blainville's beaked whales, but a strong difference in ICI between these species and Cross Seamount beaked whales. It is possible that classification of these species could be improved by refining an event to consider ICI from individual click trains. Subdividing click trains in PAMGuard is time consuming (and complicated in the absence of localization), but alternative options include consideration of the PAMGuard click train detector module in BANTER, or by developing a similar function in R to apply to events.

For this study, only a few simple environmental features were considered for one of the study areas (seafloor depth, seafloor gradient, and sea surface temperature). In the EPacific study area, inclusion of these variables increased the overall classification rate from 89.2% to 92.2%. These variables had the most impact on classification results for the Cross Seamount beaked whale ($n = 6$), raising the classification scores to from 83.3% to 100%. Alternative environmental features may improve results for other species.

While BANTER provides an efficient and consistent approach to classification, there are significant limitations that must be considered. BANTER is a supervised machine learning tool and requires reliable training data for success. Training data should consist of labels with strong confidence in species identify (ideally determined by agreement from more than one analyst) and sample sizes should be large enough to explain the natural variability in the data.

The workflow presented here provides a highly automated approach to detection of acoustic events (PAMGuard), integration of environmental data (PAMPal) and acoustic event classification (BANTER). These methods significantly reduce manual analysis, provide more consistent classification results with fewer biases, and provide an estimate of classification error. The greatest improvement to classification results for beaked whales would likely result from improved sample sizes, and examination of individual click trains in measuring ICI. Consideration of additional detectors (e.g. matched filter detector, click train detector) or additional environmental variables may further improve classification results. Improved alignment of detection methods across studies (i.e., definition of events, hydrophone calibration) may allow for combination of data within geographic regions to improve assessments within regions for all species, and across regions for species with global distributions, such as Cuvier's beaked whales. These

highly automated methods may allow for analysis of data across large spatial and temporal scales to address large ecological and population level questions.

Acknowledgements

The authors would like to acknowledge the large number of scientists and shipboard crew who were responsible for data collection during multiple large scale studies. Sea surface temperature data were provided by JPL under support by NASA MEASUREs program. Funding for data collection and analysis were provided by U.S. Navy, Bureau of Ocean Energy Management, and the National Oceanographic and Atmospheric Administration. The manuscript was improved thanks to reviews from XX and XX.

Figures and tables

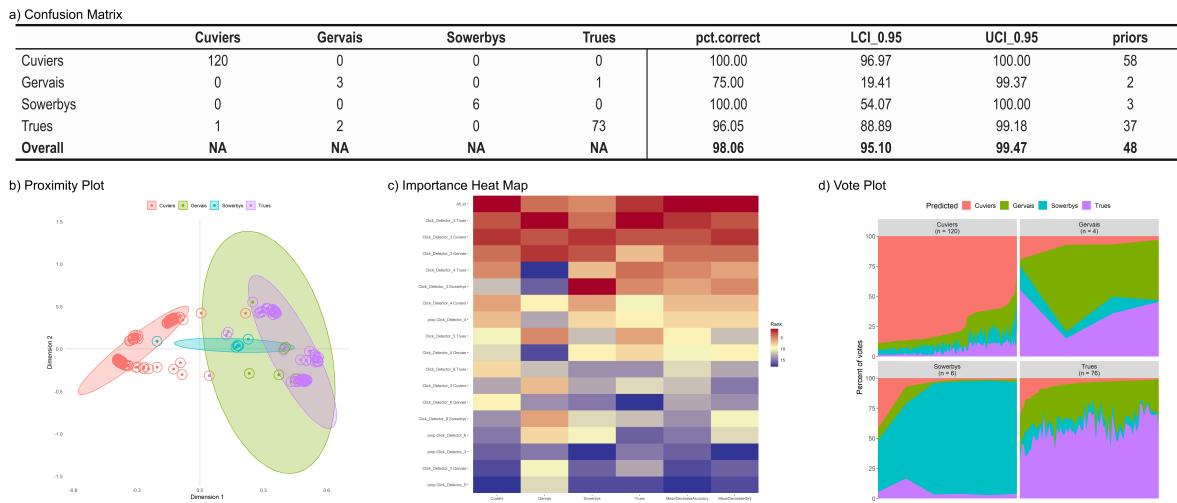


Figure 1: BANTER classification results from the NAtlantic dataset including echolocation clicks and interclick interval (EC_ICI). Confusion matrix (a) provides the percent correct classification for each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error rate). Proximity plot (b) for species events from BANTER model (central dot color represents true species identity; color of circle surrounding dot represents BANTER species classification). Heat map (c) for ranks of variable importance for each species; colors scale from most important predictors (dark red) to least important predictors (dark blue). Vote Plot (d) shows the vote distribution for each event (vertical slice) for each species; distribution of votes by species is shown by their representative color.

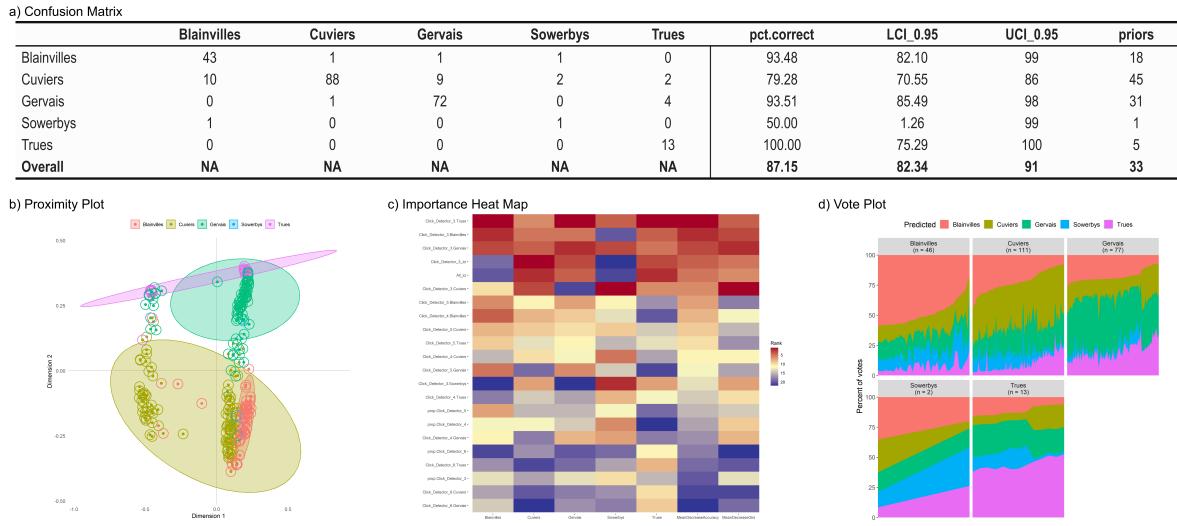


Figure 2: BANTER classification results from the SAtlantic dataset (EC_ICI). Confusion matrix (a) provides the percent correct classification for each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error rate). Proximity plot (b) for species events from BANTER model (central dot color represent true species identity; color of circle surrounding dot represents BANTER species classification). Heat map (c) for ranks of variable importance for each species; colors scale from most important predictors (dark red) to least important predictors (dark blue). Vote Plot (d) shows the vote distribution for each event (vertical slice) for each species; distribution of votes by species is shown by their representative color.

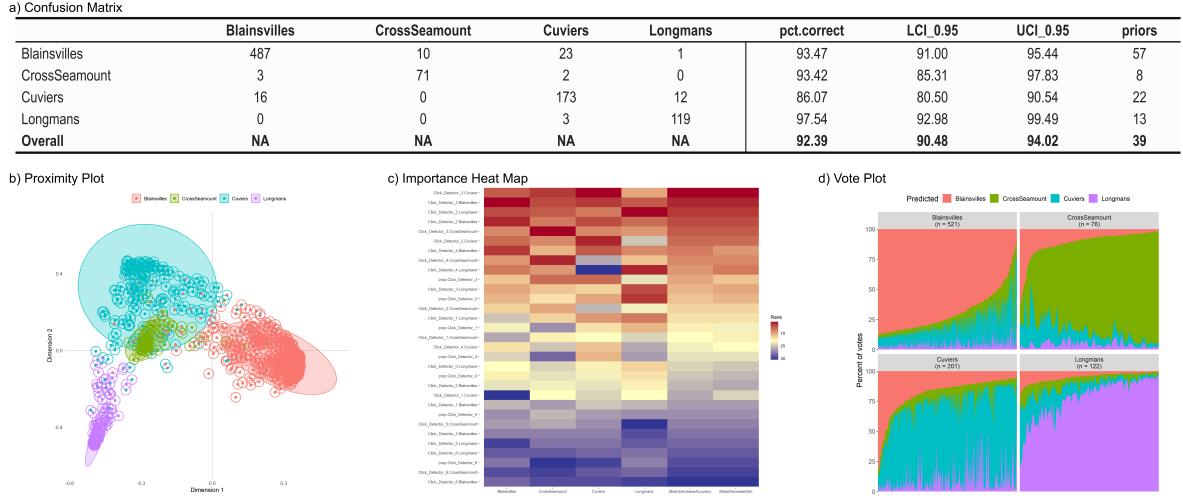


Figure 3: BANTER classification results from the Hawaii dataset (EC). Confusion matrix (a) provides the percent correct classification or each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error rate). Proximity plot (b) for species events from BANTER model (central dot color represent true species identity; color of circle surrounding dot represents BANTER species classification). Heat map (c) for ranks of variable importance for each species; colors scale from most important predictors (dark red) to least important predictors (dark blue). Vote Plot (d) shows the vote distribution for each event (vertical slice) for each species; distribution of votes by species is shown by their representative color.

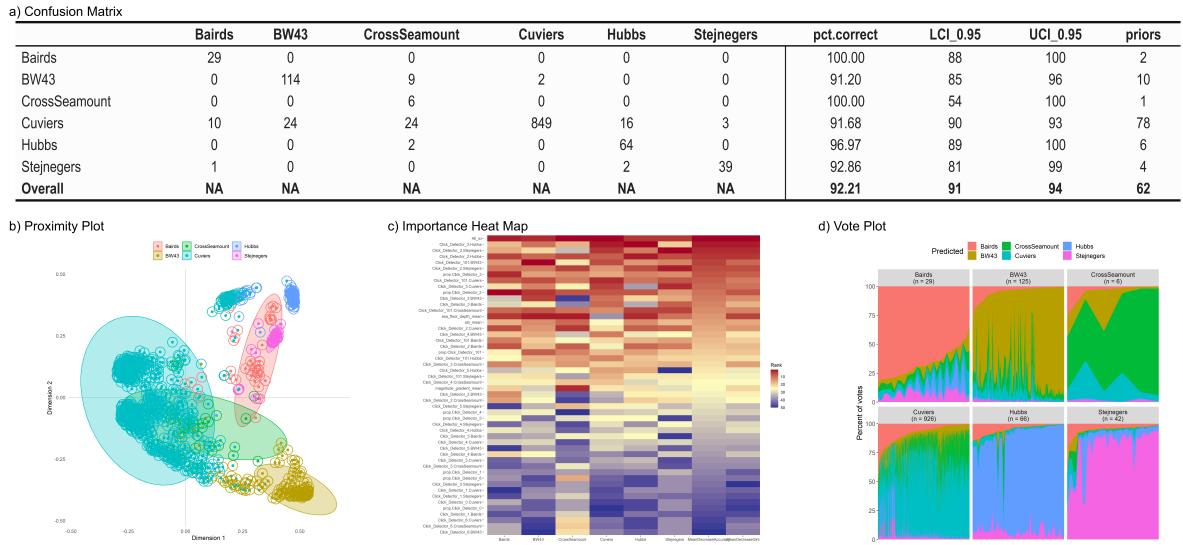


Figure 4: BANTER classification results from the EPacific dataset with environmental data (EC_ICI_ENV). Confusion matrix (a) provides the percent correct classification or each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error rate). Proximity plot (b) for species events from BANTER model (central dot color represent true species identity; color of circle surrounding dot represents BANTER species classification). Heat map (c) for ranks of variable importance for each species; colors scale from most important predictors (dark red) to least important predictors (dark blue). Vote Plot (d) shows the vote distribution for each event (vertical slice) for each species; distribution of votes by species is shown by their representative color.

References

- Archer, Frederick. 2022a. “Banter: BioAcoustic eveNT classifiER.” <https://cran.r-project.org/web/packages/banter/index.html>.
- . 2022b. “rfPermute: Estimate Permutation p-Values for Random Forest Importance Metrics.” <https://cran.r-project.org/web/packages/rfPermute/index.html>.
- Barlow, Jay, Gustavo Cárdenas-Hinojosa, E. Elizabeth Henderson, Dawn Breese, Diana López-Arzate, Eva Hidalgo Pla, and Barbara L. Taylor. 2022. “Unique Morphological and Acoustic Characteristics of Beaked Whales (*Mesoplodon* Sp.) Off the West Coast of Baja California, Mexico.” *Marine Mammal Science* 38 (1): 383–90. <https://doi.org/10.1111/mms.12853>.
- Barlow, Jay, Jeffrey E. Moore, Jennifer L. K. McCullough, and Emily T. Griffiths. 2022. “Acoustic-Based Estimates of Cuvier’s Beaked Whale (*Ziphius Cavirostris*) Density and Abundance Along the U.S. West Coast from Drifting Hydrophone Recorders.” *Marine Mammal Science* 38 (2): 517–38. <https://doi.org/10.1111/mms.12872>.
- Baumann-Pickering, Simone, Mark A. McDonald, Anne E. Simonis, Alba Solsona Berga, Karolina P. B. Merkens, Erin M. Oleson, Marie A. Roch, et al. 2013. “Species-Specific Beaked Whale Echolocation Signals.” *The Journal of the Acoustical Society of America* 134 (3): 2293–2301. <https://doi.org/10.1121/1.4817832>.
- Baumann-Pickering, Simone, Marie A. Roch, Robert L. Brownell Jr, Anne E. Simonis, Mark A. McDonald, Alba Solsona-Berga, Erin M. Oleson, Sean M. Wiggins, and John A. Hildebrand. 2014. “Spatio-Temporal Patterns of Beaked Whale Echolocation Signals in the North Pacific.” *PLOS ONE* 9 (1): e86072. <https://doi.org/10.1371/journal.pone.0086072>.
- Bittle, Michael, and Alec Duncan. 2013. “A Review of Current Marine Mammal Detection and Classification Algorithms for Use in Automated Passive Acoustic Monitoring.” *Proceedings of Acoustics* 2013: 8.
- DeAngelis, Annamaria Izzi, Joy E. Stanistreet, Simone Baumann-Pickering, and Danielle M. Cholewiak. 2018. “A Description of Echolocation Clicks Recorded in the Presence of True’s Beaked Whale (*Mesoplodon Mirus*).” *The Journal of the Acoustical Society of America* 144 (5): 2691–2700. <https://doi.org/10.1121/1.5067379>.
- Keating, Jennifer L., and Jay Barlow. 2013. “Summary of PAMGuard Beaked Whale Click Detectors and Classifiers Used During the 2012 Southern California Behavioral Response Study.” {NOAA} {Tech} {Memo} NOAA-TM-NMFS-SWFSC-517.
- MacLeod, Colin D. 2018. “Beaked Whales, Overview.” In *Encyclopedia of Marine Mammals (Third Edition)*, edited by Bernd Würsig, J. G. M. Thewissen, and Kit M. Kovacs, 80–83. Academic Press. <https://doi.org/10.1016/B978-0-12-804327-1.00062-5>.
- McDonald, Mark A., John A. Hildebrand, Sean M. Wiggins, David W. Johnston, and Jeffrey J. Polovina. 2009. “An Acoustic Survey of Beaked Whales at Cross Seamount Near Hawaii.” *The Journal of the Acoustical Society of America* 125 (2): 624–27. <https://doi.org/10.1121/1.3050317>.
- Parijs, Sofie M. Van, Chris W. Clark, Renata S. Sousa-Lima, Susan E. Parks, Shannon Rankin, Denise Risch, and Ilse C. Van Opzeeland. 2009. “Management and Research Applications of Real-Time and Archival Passive Acoustic Sensors over Varying Temporal and Spatial

- Scales.” *Marine Ecology Progress Series* 395 (December): 21–36. <https://doi.org/10.3354/meps08123>.
- Rankin, Shannon, and Frederick Archer. 2021. “BANTER: A User’s Guide to Acoustic Classification.” <https://taikisan21.github.io/PAMpal/banterGuide.html>.
- Rankin, Shannon, Frederick Archer, Jennifer L. Keating, Julie N. Oswald, Michael Oswald, Alex Curtis, and Jay Barlow. 2017. “Acoustic Classification of Dolphins in the California Current Using Whistles, Echolocation Clicks, and Burst Pulses.” *Marine Mammal Science* 2 (33): 520–40. <https://doi.org/10.1111/mms.12381>.
- Rankin, Shannon, and Jay Barlow. 2005. “Source of the North Pacific ‘Boing’ Sound Attributed to Minke Whales.” *The Journal of the Acoustical Society of America* 118 (5): 3346–51. <https://doi.org/10.1121/1.2046747>.
- Sakai, Taiki. 2021. “PAMpal: Load and Process Passive Acoustic Data.” <https://cran.r-project.org/web/packages/PAMpal/index.html>.
- Simonis, Anne E. 2020. “Passive Acoustic Survey Of Deep-Diving Odontocetes in the California Current Ecosystem 2018: Final Report.” *NOAA-TM-NMFS-SWFSC-630*. <https://doi.org/10.25923/W5XX-JZ73>.
- Simons, R. A., and Chris John. 2022. “ERDDAP.” Monterey, CA: NOAA/NMFS/SWFSC/ERD. <https://coastwatch.pfeg.noaa.gov/erddap>.
- Soldevilla, Melissa S., E. Elizabeth Henderson, Gregory S. Campbell, Sean M. Wiggins, John A. Hildebrand, and Marie A. Roch. 2008. “Classification of Risso’s and Pacific White-Sided Dolphins Using Spectral Properties of Echolocation Clicks.” *The Journal of the Acoustical Society of America* 124 (1): 609–24. <https://doi.org/10.1121/1.2932059>.
- Team, R Core. 2022. “R: A Language and Environment for Statistical Computing.” Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Yano, Kymberly M., Erin M. Oleson, Jennifer L. Keating, Lisa Ballance, Marie C. Hill, Amanda Bradford, Ann N. Allen, Trevor W. Joyce, Jeffrey E. Moore, and Annette E. Henry. 2018. “Cetacean and Seabird Data Collected During the Hawaiian Islands Cetacean and Ecosystem Assessment Survey (HICEAS), July–December 2017.” {OCS} {Study} {BOEM} OCS Study BOEM 2018-044. Bureau of Ocean Energy Management. <https://repository.library.noaa.gov/view/noaa/18660>.
- Zahn, Marie J., Shannon Rankin, Jennifer L. K. McCullough, Jens C. Koblitz, Frederick Archer, Marianne H. Rasmussen, and Kristin L. Laidre. 2021. “Acoustic Differentiation and Classification of Wild Belugas and Narwhals Using Echolocation Clicks.” *Scientific Reports* 11 (1): 22141. <https://doi.org/10.1038/s41598-021-01441-w>.