Open-Source machine learning BANTER acoustic classification of beaked whale echolocation pulses

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

7/26/23

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, requires 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 (northwestern North Atlantic, southwestern North Atlantic, Hawaii, and Eastern Pacific) were analyzed using PAMGuard (acoustic detection), PAMpal (acoustic analysis) and BANTER (hierarchical random forest classifier). Overall classification accuracy ranged from 88% for the southwestern North Atlantic data to 97% for the northwestern 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 other 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 reliably be 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 the amount and number of recordings being made make it impossible for experienced acousticians to manually annotate it all. This has lead to an increased need for the development of automated classification routines that can provide accurate species determinations from acoustic recordings.

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 signals in the form of frequency-modulated (FM) pulses that have temporal and spatial characteristics that can be used to differentiate species (Baumann-Pickering et al. 2013). 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 pulses requires analysis by trained analysts using a number of visual aids to examine the call characteristics. Development of automated classification routines, if accurate, serve to improve the efficiency, reduce the subjectivity, and decrease the cost of analyzing large datasets.

BANTER (Bio-Acoustic eveNT classifiER)is a supervised machine learning method originally developed for species identification of dolphin schools from acoustic recordings (Rankin et al. 2017). This hierarchical Random Forest (Breiman (2001)) event classifier combines information from independent classification models for each call type (in the dolphin case study, this included one each for whistles, echolocation pulses, and burst pulses) into a model classifying a full acoustic event, which is defined as a discrete collection of calls during an encounter. The event classifier uses the distribution of species assignment probabilities for each call type, along with any event-level metrics (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 detector for a specific call type could yield different detector results (Rankin et al. 2017). In this case, settings for a whistle and 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 results were different. This approach was successfully applied to classified echolocation pulses for narwhals and belugas (Zahn et al. 2021). For species who primarily (or exclusively) produce echolocation pulses, 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, 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 species identification 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 the National Oceanographic and Atmospheric Administrations' research operations off the US East Coast, the Hawaiian Islands (Hawaii) and the US West Coast.

Recordings from the U.S. east coast (western North Atlantic) were collected using towed hydrophone arrays during the 2016 AMAPPS II (Atlantic Marine Assessment Program for Protected Species) survey Northeast Fisheries Science Center (U.S.). Southeast Fisheries Science Center (U.S.) (n.d.). 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, Appendix A, Northeast Fisheries Science Center (U.S.). Southeast Fisheries Science Center (U.S.) (n.d.)) and recordings used in this analysis consisted of two hydrophones (HTI-96-min, High Tech Inc., Long Beach, MS) which recorded at a 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, Appendix B, Northeast Fisheries Science Center (U.S.). Southeast Fisheries Science Center (U.S.) (n.d.)) and recordings used in this analysis consisted of two hydrophones (Reson [MODEL], Teledyne Marine, Slangerup, Denmark) which recorded at a 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, McCullough et al. (2021), Yano et al. (2018)). Drifting acoustic recorders were deployed at randomly selected locations within the main Hawaiian Islands. Data were collected using SoundTrap ST4300 recorders (Ocean Acoustics, Auckland, New Zealand) sampled at 288 kHz with a duty cycle of 2 min on for every 10 minutes. Individual buoys were deployed for 11 to 23 days (see McCullough et al. (2021), 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 (using 'save classifier set' in Pamguard click classification window). The general click classifiers were treated as spectral band click detectors (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)).

Echolocation signals that were identified as frequency-modulated pulses from beaked whales were then linked as PAMGuard "events" by experienced acoustic researchers. Events included all the pulses that were close to each other in time, had similar spectral and waveform characteristics, and were received at consistent bearing angles. For NAtlantic and SAtlantic data, acoustic detections of beaked whales using towed hydrophone arrays were further subdivided into putative individuals by identification of consecutive pulses along the same bearing angle in the bearing time plot (see DeAngelis et al. (2018)**Annamaria suggested citing 2017**).

Beaked whale acoustic detection events were then classified to species by experienced acoustic researchers using multiple lines of evidence. Experienced acousticians followed strict protocols to define species identity based on acoustic characteristics. Acoustic researchers with experience in classifying beaked whale echolocation pulses examined the bearing-time display for the click detector module in PAMGuard to assess the species identity of the acoustic event. When the available characteristics were inconclusive, the species was considered an unidentified beaked whale. Data were saved within the PAMGuard database and binaries for downstream processing using the *PAMpal package* (Sakai 2021) in R v2022.07.02 (Team 2022).

Standard click metrics were measured from echolocation pulses using default values in *PAMpal* (**CITE Taiki draft manuscript**). Inter-click interval (IPI) was calculated using the 'calculateICI' function in *PAMpal*; the mode of the IPI 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). A full consideration of environmental data for all regions was beyond the scope of this study.

Multiple models were created to better understand the value (and potential limitations) of the different suite of measurements. Models included standard click metrics (only) for all pulses in a beaked whale acoustic event (EC); the addition of IPI for each event (EC_IPI), and for the EPacific dataset we also included environmental variables (EC_IPI_ENV).

BANTER Acoustic Classification

BANTER models were created for each trial using the *banter* package (Archer 2022a) in R v2022.07.02 (Team 2022). The BANTER approach includes two stages of classification: first call classification and then event classification, which is based on the call classification values and (optionally) event-level parameters such as ICI and environmental variables. Standard click metrics were used for the BANTER call classifier. A minimum of two events per species and the complete suite of variables (no missing data) were required for training and testing of the classification model. The Random Forest models that BANTER is based on are largely driven by two parameters, the number of trees in the forest (*ntree*) and the number of samples randomly drawn to build each tree (*sampsize*). Small values of *sampsize* were selected to improve computational performance, but large values of *ntree* were used to obtain models with stable classification results (Rankin and Archer 2021). Model results were summarized using the *rfpermute* package (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*), Hubbs' 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. Species 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 pulses (EC), and (2) echolocation pulses and IPI (EC_IPI).

For the NAtlantic dataset, the most accurate model included echolocation pulse metrics with IPI; this model had a modest increase in accuracy over the EC (only) model (EC_IPI model, Detector Model: *sampsize* = 3, *ntree* = 10,000; Event Model: *sampsize* = 3, *ntree* = 100,000). This model provided an overall correct classification rate of 97.5% 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 beaked whales. All classification results were greater than expected (Priors in Figure 1 a). Results from the EC_IPI 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 classification model space. 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. This plot shows the degree of overlap between Gervais', Sowerby's and True's beaked whales. Species classifications are resolved using additional features, as seen in the ten msot important features for predicting each species (importance heat map, Figure 1 c). The strength of classifications for each event can be seen in the distribution of votes for each species across the forest (Figure 1 d). In each frame of this figure, the events are represented as vertical slices along the x-axis, and the percentage of votes for each species is represented by their color along that vertical slice in the y-axis. This distribution shows strong assignment probabilities for most Cuvier's and Sowerby's beaked whale events, with greater variability in assignment for Gervais' and True's beaked whale events.

SAtlantic

Species encountered in the SAtlantic study dataset include Blainville'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 pulses (EC), and (2) echolocation pulses and IPI (EC_IPI).

For the SAtlantic dataset, the most accurate model considered the echolocation pulse metrics and the IPI for an event (EC_IPI Model, Detector Model: *sampsize* = 4, *ntree* = 10,000; Event Model: *sampsize* = 1, *ntree* = 100,000) and provided an overall correct classification rate of 88.7% 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 expected (Priors in Figure 1 a). The sample sizes for the BANTER model were kept low (event 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_IPI_ALT), but resulted in the loss of Sowerby's in the final model. Results from the EC_IPI model are presented here; results from the EC and EC_IPI_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 IPI measurements having relatively low importance (Figure 2 c). The distribution of votes (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 McCullough et al. (2021); species included Blainville'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 pulses (EC), and (2) echolocation pulses and IPI (EC_IPI).

The most accurate model considered the echolocation pulses only (EC Model, Detector Model: *sampsize* = 10, *ntree* = 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 expected (Prior in Figure 3 a). Results from the EC model are presented here; results from EC_IPI model can be found in supplementary materials (Supplement Fig. 4).

Distinctive clusters can be seen for Longman's and Blainville's beaked whales, although there is overlap on these first two dimensions of the clusters for Cuvier's and Cross Seamount beaked whales (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 distribution of votes (Figure 3 d) show strong classification results for most Blainville'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 pulses (EC), (2) echolocation pulses and IPI (EC_IPI), and (3) echolocation pulses, IPI, and environmental features (EC_IPI_ENV).

The best model considered the echolocation pulses, IPI, and environmental features (EC_IPI_ENV Model, Detector Model: *sampsize* = 1, *ntree* = 10,000; Event model: *sampsize* = 5, *ntree* = 10,000) and provided an overall correct classification rate of 91.9% for all six species. Classification scores ranged from a low of 91.2% (BW43) to 100% for Cross

Seamount beaked whale (Figure 4 a). All classification results were greater than expected (Prior in Figure 4 a). Results from the EC_IPI_ENV model are presented here; results from the EC and EC_IPI 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). IPI was among the most important classification variable for all species (Figure 4 c importance heat map). The distribution of 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 88.7% correct classification in the SAtlantic dataset to a high of 97.5% correct classification in the NAtlantic dataset. All species in all datasets had classification 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 events, with strong assignment probabilities. 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 indicated in the vote distributions).

While some species have distinctive differences in their pulse characteristics that can 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 pulse 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 = 82.8% correct classification). We found that 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 inability to include 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.

Gervais' and True's beaked whales have similar pulse 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 metrics. 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 heavily influential if there are inaccuracies in the training data.

These two examples of small sample sizes highlight 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 high level of confidence for inclusion. This is especially critical for species with small sample sizes. An alternative is to require agreement from multiple analysts. 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).

Interpulse interval (IPI) was found to be the most important variable for NAtlantic and EPacific, and IPI were the fourth and fifth ranked variables for SAtlantic (see importance heat maps for each survey area). While the addition of IPI did not improve results for data from Hawaii, results were similar and IPI ranked #9 in importance for the Hawaii study area. Baumann-Pickering et al. (2013) showed an overlap in the IPI between Cuvier's and Blainville's beaked whales, but a strong difference in IPI between these species and Cross Seamount beaked whales. It is possible that classification of these species could be improved by refining an event to annotation at the individual level, as was done with the NAtlantic and SAtlantic datasets. Subdividing pulse trains into individuals in PAMGuard is time consuming, but alternative options include consideration of the PAMGuard click train detector module as an additional first stage detector 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 91.1% to 91.9%. 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 identity (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 individuals to accurately measure IPI. 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 combining 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. Thanks to Liam Mueller-Brennan for their help with data analysis. Sea surface temperature data were provided by JPL under support by NASA MEaSUREs program. Funding for data collection and analysis were provided by the 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

a) Confusion Matrix

	Cuviers	Gervais	Sowerbys	Trues	pct.correct	LCI_0.95	UCI_0.95	priors
Cuviers	120	0	0	0	100.00	96.97	100.00	58
Gervais	0	3	0	1	75.00	19.41	99.37	2
Sowerbys	0	0	6	0	100.00	54.07	100.00	3
Trues	0	4	0	72	94.74	87.07	98.55	37
Overall	**NA**	**NA**	**NA**	**NA**	**97.57**	**94.43**	**99.21**	**48**

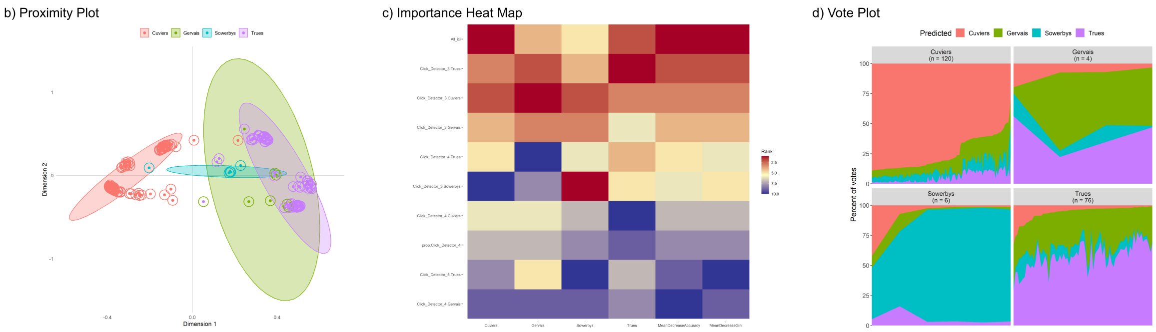

Figure 1: BANTER classification results from the NAtlantic dataset including echolocation pulses and interclick interval (EC_IPI). Confusion matrix (a) provides the percent correct classification for each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), 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 ten most important variables; 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

a) Confusion Matrix

	Blainvilles	Cuviers	Gervais	Sowerbys	Trues	pct.correct	LCI_0.95	UCI_0.95	priors
Blainvilles	43	1	1	1	0	93.48	82.10	99	18
Cuviers	9	92	7	2	1	82.88	74.57	89	45
Gervais	0	1	72	0	4	93.51	85.49	98	31
Sowerbys	1	0	0	1	0	50.00	1.26	99	1
Trues	0	0	0	0	13	100.00	75.29	100	5
Overall	**NA**	**NA**	**NA**	**NA**	**NA**	**88.76**	**84.16**	**92**	**33**

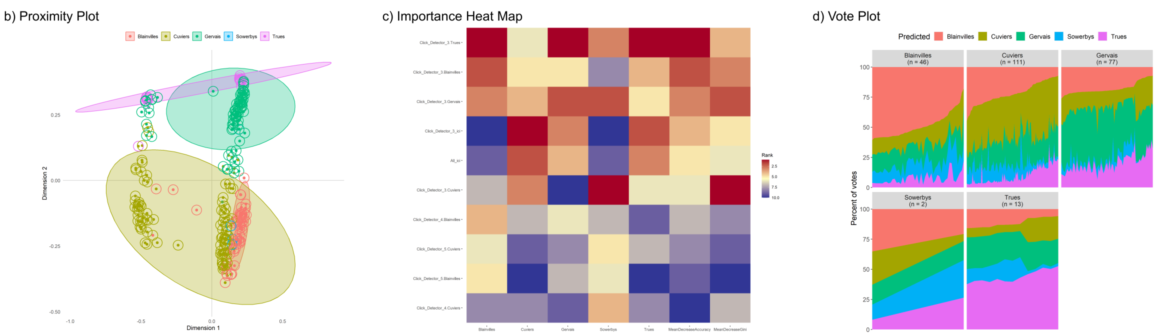

Figure 2: BANTER classification results from the SAtlantic dataset (EC_IPI). Confusion matrix (a) provides the percent correct classification for each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error rate). Proximity plot (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 ten most important variables; 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.

a) Confusion Matrix

	Blainsvilles	CrossSeamount	Cuviers	Longmans	pct.correct	LCI_0.95	UCI_0.95	priors
Blainsvilles	487	10	23	1	93.47	91.00	95.44	57
CrossSeamount	3	71	2	0	93.42	85.31	97.83	8
Cuviers	16	0	173	12	86.07	80.50	90.54	22
Longmans	0	0	3	119	97.54	92.98	99.49	13
Overall	**NA**	**NA**	**NA**	**NA**	**92.39**	**90.48**	**94.02**	**39**

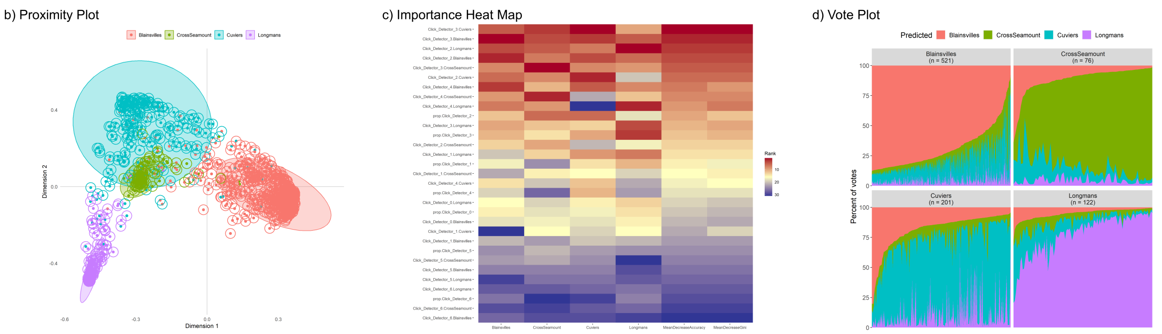

Figure 3: BANTER classification results from the Hawaii dataset (EC). Confusion matrix (a) provides the percent correct classification for each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error rate). Proximity plot (b) for 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 ten most important variables; 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.

a) Confusion Matrix

	Bairds	BW43	CrossSeamount	Cuviers	Hubbs	Stejnegers	pct.correct	LCI_0.95	UCI_0.95	priors
Bairds	28	0	0	0	1	0	96.55	82	100	2
BW43	0	114	9	2	0	0	91.20	85	96	10
CrossSeamount	0	0	6	0	0	0	100.00	54	100	1
Cuviers	12	31	23	846	11	3	91.36	89	93	78
Hubbs	0	0	1	0	65	0	98.48	92	100	6
Stejnegers	1	0	0	0	2	39	92.86	81	99	4
Overall	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**91.96**	**90**	**93**	**62**

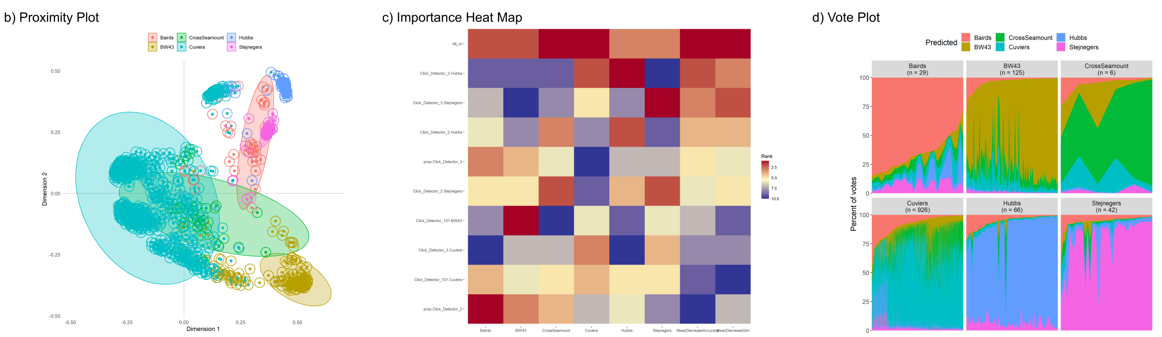

Figure 4: BANTER classification results from the EPacific dataset with environmental data (EC_IPI_ENV). Confusion matrix (a) provides the percent correct classification for each species (pct.correct), lower confidence intervals (LCI_0.95), upper confidence intervals (UCI_0.95), and priors (expected error 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 ten most important variables; 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.

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, Karlina 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.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. https://doi.org/10.1023/A:1010933404324.

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.

McCullough, Jennifer L. K., Erin M. Oleson, Jay Barlow, Ann N. Allen, and Karlina P. B. Merkens. 2021. “An Acoustic Survey in the Main Hawaiian Islands Using Drifting Recorders.” *PIFSC Administrative Report* H-21-04: 26. https://doi.org/10.25923/RZZZ-0V38.

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.

Northeast Fisheries Science Center (U.S.). Southeast Fisheries Science Center (U.S.). n.d. “2016 Annual Report of a Comprehensive Assessment of Marine Mammal, Marine Turtle, and Seabird Abundance and Spatial Distribution in US Waters of the Western North Atlantic Ocean – AMAPPS II.” Accessed June 30, 2023. https://doi.org/10.25923/GBAP-G480.

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.