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 Melissa Soldevilla

2/7/23

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 86% for the South Atlantic data to 99% for the North Atlantic. Results for many species could likely be improved with increased sample sizes and consideration of alternative automated detectors. 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.

```r
library(tidyverse)
```

```
-- Attaching packages --------------------------------------- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.8      v dplyr   1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 1.0.0
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()    masks stats::lag()
```

```r
library(knitr)
library(here)
```

```
here() starts at C:/Users/shannon.rankin/Documents/GitHub/BANTER_BeakedWhales
```

Introduction

Passive acoustic monitoring (PAM) from 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. 2013a); (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. 2013b). 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 et al. 2022); (Baumann-Pickering et al. 2014); (Simonis et al. 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. 2017a). This hierarchical random forest event classifier relies on an initial call classifier (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. 2017b). 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 off 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 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. 2018a) 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).

Recordings from Hawaii were collected using drifting acoustic recorders during the 2017 HICEAS (Hawaiian Island Cetacean Ecosystem Assessment Survey, (Yano et al. 2018)). Drifting recording buoys 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 2016 PASCAL (Passive Acoustic Survey of Cetacean Abundance Levels) survey (J. Keating et al. 2018a). Buoys were deployed at a number of stations in the California Current, including pre-determined locations (for abundance estimation studies) and in close proximity to seamounts (Ad Hoc Seamount Experiment). 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 from pre-determined locations were duty cycled and recorded 2 minutes out of a variable 'off' time; recordings from the Seamount Experiment were continuous. Soundtrap recorders were sampled at 288 kHz and the SM3M recorders were recorded at 256 sample rate (with some sampling at 96 kHz for ocean noise measurements). Individual buoys were deployed for 11-23 days (see [J. Keating et al. (2018b)] for more detailed information).

Acoustic Data Analysis

All acoustic recordings were analyzed using PAMGuard software (version 2.00.15c) with a suite of generic click classifiers within the Click Detector module (see (J. L. Keating and Barlow 2013a)). 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/classifier within the 30 – 50 kHz peak frequency range considered the presence of a frequency sweep that is characteristic of beaked whale pulses (see (J. L. Keating and Barlow 2013b)).

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. 2018b)). Data were saved within the PAMGuard database and binaries for downstream processing using PAMpal package (Sakai 2021) in R programming language (R Core Team 2022).

Standard click calculations were measured from echolocation clicks using default values in PAMpal. For EPacific data, any events > 2 minutes were subdivided into 2-minute events to ensure that all events were of equal value (because most of our data was duty cycled into 2-minute blocks). 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.

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. At a minimum, each trial included standard click calculations for all clicks in a beaked whale acoustic event (EC); and some trials included the mode of the ICI for each event (EC_ICI).

BANTER Acoustic Classification

BANTER models was 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 were applied to the BANTER event classifier. A minimum of two events per species and 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 rfpermute 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*), Blainsville's beaked whales (*Mesoplodon densirostris*), Cuvier's beaked whales (*Ziphius cavirostris*), Cross Seamount beaked whale (McDonald et al., 2009), Gervais' beaked whales (*Mesoplodon europaeus*), Longman's beaked whale (*Indopacetus pacificus*), Sowerby's beaked whales (*Mesoplodon bidens*), True's beaked whales (*Mesoplodon mirus*), Stejneger's beaked whale (*Mesoplodon stejnegeri*), and unidentified beaked whale 'BW43' (Baumann-Pickering et al. 2013a). 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=70), Sowerby's beaked whales (n=5), True's beaked whales (n=65), and Gervais' beaked whales (n=3). BANTER classification trials included (1) echolocation clicks (EC), and (2) echolocation clicks and ICI (EC_ICI).

For the NAtlantic dataset, the most accurate model considered the echolocation clicks and the ICI for an event (EC_ICI, Detector Model: sampsize = 4, ntrees = 30,000; BANTER Model: sampsize = 4, trees=300,000) and provided an overall correct classification rate of 100% for all four species (Fig. 1). There were insufficient samples of Gervais' beaked whales to include in the NAtlantic model (n=2). All classification results were greater than the expected error rate (Prior in Fig. 1a). Results from the EC model are presented here; results from the EC_ICI model can be found in supplementary materials (Supplement Fig. 1).

[Insert Figure 1][Figure 1. BANTER classification results from the NAtlantic 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.

The proximity plot (Fig.1b) 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 (Fig. 1c) shows that the most important features for predicting each species.

The strength of the classifications by species can be seen in the plot votes (Fig. 1d), which show the istribution of the votes for each of the 300,000 trials. 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 300,000 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 plot vote for the NAtlantic dataset show that all species

SAtlantic

Species encountered in the SAtlantic study dataset include Blainsville's beaked whales (n=35), Cuvier's beaked whales (n=66), Gervais' beaked whales (n=45), Sowerby's beaked whales (n=2), and True's beaked whales (n-12). 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 (Detector Model: sampsize = 2, ntrees = 30,000; BANTER Model: sampsize = 1, ntrees = 300,000) and provided an overall correct classification rate of 86.9% for all four species. 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 beaked whale (see Supplement Fig. 3 for EC_ICI_ALT), but resulted in the loss of Sowerby's in the final model. Classification scores ranged from a low of 74.2% (Cuvier's beaked whale) to 100% (Sowerby's and True's beaked whale, Fig. 2); all classification results were greater than the expected error rate (Prior in Fig. 2a). Results from the EC_ICI model are presented here; results from the EC model can be found in supplementary materials (Supplement Fig. 2).

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

For SAtlantic, the proximity plot (Fig. 2b) shows that True's beaked whales are differentiated from the other species based solely on the two most important features. The importance heat map (Fig. 2c) shows that Cuvier's, Gervais' and Sowerby's beaked whales rely on a number of predictors that are lower in importance and there is likely greater distinction between these species than is represented by the proximity plot (Fig. 2b). The plot votes (Fig. 2d) shows a moderate number of trees in the forest were misclassified and these misclassifications included all species in the model.

Hawaii

The Hawaii dataset consisted of 13 drifting buoys deployed within the main Hawaiian Islands (Yano et al. 2018); species included Longman's beaked whale (n=121), Blainsville's beaked whale (n=518), Cross Seamount beaked whale (n=84) and Cuvier's beaked whale (n=201). 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 (Detector Model: sampsize = 10, ntrees = 5,000; BANTER model: sampsize = 4, ntree = 10,000) and provided an overall correct classification rate of 91.8% for all four species. Classification scores ranged from a low