Received Date : 23-Jun-2016
Revised Date : 17-Oct-2016
Accepted Date : 23-Oct-2016
Article type : Article

Acoustic classification of dolphins in the California Current using whistles, echolocation clicks, and burst pulses

SHANNON RANKIN,[1] **FREDERICK ARCHER**, **JENNIFER L. KEATING**, Southwest Fisheries Science Center, NMFS, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.; **JULIE N. OSWALD**, **MICHAEL OSWALD**, Bio-waves, Inc., 364 2nd Street, Suite #3, Encinitas, California 92024, U.S.A.; **ALEX CURTIS**, Ocean Associates, Inc., under contract to Southwest Fisheries Science Center, NMFS, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.; **JAY BARLOW**, Southwest Fisheries Science Center, NMFS, NOAA, 8901 La Jolla Shores Drive, La Jolla, California 92037, U.S.A.

ABSTRACT

Passive acoustic monitoring of dolphins is limited by our ability to classify calls to species. Significant overlap in call characteristics among many species, combined with a wide range of call types and acoustic behavior, makes classification of calls to species challenging. Here, we introduce BANTER, a compound acoustic classification method for dolphins that utilizes information from all call types produced by dolphins rather than a single call type, as has been typical for acoustic classifiers. Output from the passive acoustic monitoring software, PAMGuard, was used to create independent classifiers for whistles, echolocation clicks, and burst pulses, which were then merged into a final, compound classifier for each species. Classifiers for five species found in the California Current ecosystem were trained and tested using 153 single-species acoustic events recorded during a 4.5 mo combined visual and acoustic shipboard cetacean survey off the west coast of the United States. Correct classification scores for individual species ranged from 71% to 92%, with an overall correct classification score of 84% for all five species. The conceptual framework of this approach easily lends itself to other species and study areas as well as to noncetacean taxa.

Key words: acoustic, classification, dolphins, delphinids, whistles, burst pulses, echolocation clicks.

Research, management, and mitigation of anthropogenic impacts on marine mammals increasingly rely on remote sensing of these animals using passive acoustic monitoring (PAM) technology. These methods are especially useful in remote oceanic regions, or where poor weather conditions inhibit traditional visual observation methods. Practical application of PAM requires some understanding of the vocal repertoire of the species under consideration. PAM has been particularly effective for species that produce distinctive sounds, such as porpoises and sperm whales (Barlow and Taylor 2005, Gillespie *et al*. 2005, Gallus *et al*. 2012), or for species that produce highly stereotyped calls such as those associated with some baleen whale song (McDonald *et al.* 2006). Acoustic classification can be challenging for species with highly variable calls, such as many dolphin species. Dolphins produce both tonal and pulsed vocalizations (referred to as “whistles,” “clicks,” and “burst pulses”) that have the widest range of spectral and temporal characteristics of any cetacean. Despite being highly vocal (Rankin *et al*. 2008*b, c*), acoustic classification has been difficult for these species due to their complex vocal behavior.

Dolphin whistles are narrow-band frequency-modulated tonal sounds that serve a social or communicative function (Janik and Slater 1998, Herzing 2000). Whistle characteristics are highly variable at an individual level, with some species producing individually identifiable “signature whistles” (Sayigh *et al*. 2007). In addition to high within-individual variability, geographic variation in whistles has been identified within a species (*e.g*. Azzolin *et al*. 2013, Papale *et al*. 2013). Moreover, there is great overlap in whistle characteristics across many species (Oswald *et al*. 2007, Azzolin 2014, Frasier 2016). Despite the high variation at an individual level and the extensive overlap in whistle characteristics among species, acoustic classifiers based on whistle features have shown reasonable success in acoustic species classification (Oswald *et al*. 2007, Oswald *et al*. 2013, Azzolin *et al.* 2014, Lin and Chou 2015). Nonetheless, there remains significant ambiguity in species classifications based solely on whistle characteristics. Also, not all dolphin species and schools produce whistles (Rankin *et al.* 2007, Oswald *et al.* 2008, Au *et al*. 2010), thus acoustic identification of some dolphins must rely on alternative call types.

Echolocation clicks are short-duration, directional, broadband clicks used for foraging or sensory tasks (Au 1993). Some species have been found to produce echolocation clicks with distinct spectral features that allow for accurate species classification (Soldevilla *et al*. 2008, Baumann-Pickering *et al.* 2010). However, there is extensive overlap in the characteristics of echolocation clicks for most dolphin species (Au 1993). The use of clicks for species identification is further complicated by the directional nature of these signals and the variation in the click characteristics based on the angle at which they are received. Echolocation clicks that are received off the longitudinal axis of the dolphins head are highly attenuated and variable in their spectral and temporal characteristics (Au 1993). Nonetheless, recent studies have shown varying levels of success in species classification based on received echolocation clicks and these sounds may have strong classification value for some species (Roch *et al.* 2011, Baumann-Pickering *et al.* 2015).

Burst pulses are broadband clicks produced in series with very short inter-click intervals (ICI), and are generally attributed to social communication (Herman and Tavolga 1980, Herzing 1996, Blomqvist and Amundin 2004). Differentiation of burst pulses from echolocation clicks is frequently based on subjective human perception of ICI, with burst pulses categorized as trains of clicks whose ICIs are too short for individual click resolution by the human ear, instead generating a more complex “buzz” or “creak” sound. In a similar vein, burst pulses are frequently identified as having tonal qualities and are occasionally misidentified as whistles; adjustment of the fast Fourier transform (FFT) values to improve the temporal resolution allows for identification of the individual clicks that comprise a burst pulse. With few exceptions, burst pulses have been neglected in the literature, and to our knowledge there are no publications that consider burst pulses for automated detection and classification. Nonetheless, burst pulses may be critical for successful classification of some species (Rankin *et al*. 2007).

Recent improvements to computer hardware and digital signal processing software allow for automated detection and measurement of various call types, and powerful multivariate statistical analysis tools provide an opportunity to analyze these data for classification purposes. Here, we present a method for acoustic classification of dolphin species based on automated detection and measurement of whistles, clicks, and burst pulses. Although this algorithm, called BANTER (Bio-Acoustic eveNT classifiER), was developed using data from five species of dolphins in the California Current off the west coast of the United States, it should be readily applicable to other taxa as well as other regions.

Methods

Data Collection

All data were collected during the 4.5 mo 2014 California Current Cetacean Ecosystem Assessment Survey (CalCurCEAS). This survey was a combined visual and acoustic shipboard line-transect survey of cetacean populations within the U.S. Exclusive Economic Zone off the west coasts of Washington, Oregon, and California conducted by the Southwest Fisheries Science Center on the R/V *Ocean Starr.* Visual methods consisted of a team of three experienced visual observers searching with “big-eye” 25 × 150 power binoculars, 7× binoculars, and unaided eye (Kinzey *et al.* 2000). All animals detected visually were approached for accurate species identification and group size estimation.

A custom hydrophone array was towed at a depth of approximately 12 m, 300 m behind the ship while traveling at a speed of 10 knots during daylight hours. The hydrophone array included either two or four hydrophones (HTI-96-min) with dual-stage preamplification contained within an oil-filled section of the array. The hydrophones had a flat frequency response from 2 kHz to 100 kHz (−158 dB at ±5 dB re 1v/µPa after internal amplification). Signals from the array were sent through a Magrec signal conditioner (high pass filter at 1 kHz) and were recorded to a computer hard drive (500 kHz sampling rate) using an analog-to-digital conversion card (National Instruments 6356) and PAMGuard software (http://pamguard.org).

Sounds were monitored by an acoustic technician both aurally, using headphones, and visually, using real-time scrolling spectrographic software (ISHMAEL, Mellinger 2001). Bearing angles to sound sources were estimated using the phone-pair bearing algorithm in ISHMAEL. Acoustic localization was performed using target motion analysis and was based on the convergence of bearing angles plotted to Whaltrak, a custom-written plotting program (Rankin *et al.* 2008*a*). Acoustic localization of dolphin schools allowed the acoustic technician to determine whether an acoustic detection could be matched to a visually detected group of dolphins, or if it was independent of any visually detected group. Postcruise examination of visual and acoustic detections confirmed identification of recording times that could be unambiguously assigned to a specific visual/acoustic detection; these were defined as acoustic “events.” The species composition of each acoustic event was based on the identification provided by the visual observers. Events not sighted by the visual observation team or those that could not be positively identified to species were labeled as “unknown.” To minimize confusion, from this point forward we will use 0the term “event” as defined above, and the term “detection” to identify individual sounds (either whistles, clicks, or burst pulses).

Call Detection

Preliminary detection of whistles, clicks, or burst pulses was obtained by analyzing recordings (raw WAV files) with a suite of individual detectors in PAMGuard software (Gillespie *et al.* 2008). Due to the large amount of data included in the analysis, a maximum of 30 min was examined per event. Echolocation clicks were found in recordings using a series of click detectors developed within PAMGuard. Recordings were decimated to 250 kHz prior to processing with PAMGuard’s Whistle and Moan (WM) detector using the spectrogram format (FFT length 4,096, hop size 2,048 samples). The WM detector identifies tonal sounds within the recording and has traditionally functioned as a whistle detector. The detector may trigger on the spectrographic representation of tonal bands for burst pulses, and the extent of these detections have been found to vary based on the spectrograph settings (JK, unpublished data).

The general methods for detecting whistles and burst pulses were the same, with changes made in the WM detector settings to improve detection of the desired call type (Table 1). The ceiling for the maximum frequency was set to avoid detection of the 38 kHz echosounder (maximum frequency 37 kHz for whistles and 36 kHz for burst pulses). All whistles and burst pulses that fell within the time boundaries for an event were considered; there was no attempt to exclude false positives or add missed detections. For each acoustic event, the WM detector passed detections to the ROCCA module in PAMGuard, which measured a suite of characteristics from each sound-type (Table S1). The same characteristics were measured from burst pulses using different settings for the WM detector. These characteristics were used to develop the acoustic classifier.

Detection of echolocation clicks was accomplished with a suite of click classifiers (JK, unpublished data). These detectors were used to detect clicks; however, the preliminary species classification provided by the click-classifiers was not considered during the analyses that followed. Clicks were marked manually within the bearing/time display in PAMGuard’s Viewer Mode to group into events; click measurements were made using the ROCCA module within PAMGuard (Table S1). The ROCCA module used in this study was an early beta version and the measure of intensity (dB) was not saved for clicks. No attempts were made to exclude false positives or to include missed detections.

Exported measurements from whistles, clicks, and burst pulses were merged with metadata associated with each event using custom scripts written in R (R Core Development Team 2014). A set of filters was applied to the data to reduce the number of false detections. Clicks with durations >0.002 s were eliminated to reduce the number of echosounder detections. The burst pulse and whistle detectors have a large degree of overlap; to improve discrimination based on these related detectors, burst pulse measures with beginning frequencies less than 20 kHz were eliminated. The appearance of “banding” in the spectrograph of burst pulses is related to the interpulse interval (Wenz 1964). In an attempt to provide an automated measure of the ICI, two additional “delta” measurements were calculated: the difference between the beginning frequency of one burst pulse and the next successive burst pulse (delta.FREQBEG) and the difference between the center frequency of one burst pulse and the next successive burst pulse (delta.FREQCENTER). This provided a total of 50 measures for whistle detections, 52 measures for burst pulse detections, and 11 measures for click detections.

BANTER (Bio-Acoustic eveNT classifiER)

The BANTER model is based on a two-stage process. In the first stage, a classification model was independently developed for each call type and individual calls were classified to species. In the second stage, the results from these models were combined to create a model that would classify the entire event to species based on the set of calls detected during that event. To differentiate these two stages, we refer to the first as the “call classifier,” and the second as the “event classifier.”

The strategy employed by BANTER for event classification considers the distribution of classification probabilities for each call type provided by Random Forest (Breiman 2001). Although classification accuracy for a given call type might be relatively low for a particular species, the pattern and frequency of misclassifications for this call type was found to be informative. For example, species A might have 30% of its misclassified whistles classified as species B and the other 70% as species C. If no other species had this pattern of misclassifications and we observe an event with a similar 30:70 ratio of B and C misclassifications, then there is a strong likelihood that this event contained species A.

Call Classification

Because our data set contained several hundred thousand calls for some species, data were distilled to a smaller number of representative whistles, clicks, and burst pulses for each species in order to make the classification models computationally tractable as well as to avoid overrepresentation of calls from very vocal species or clusters of similar calls. For each call type, data were distilled by first selecting a random subset of up to 10,000 calls from each species. From this subset, the 1,000 most representative calls were selected using the Density Clustering algorithm (Rodriguez and Laio 2014) as implemented in custom modifications to the *densityClust* v2.0.3 package in R (code available on request).

With these representative calls, Random Forest models were created to classify the whistles, clicks, and burst pulses to species using the *randomForest* v4.6-12 package (Liaw and Wiener 2002) in R. Each tree in a forest was built using a random selection of either 50 or half of the available calls for each species, whichever was smaller. This process served to balance the number of calls used in the model and avoided biasing classification accuracy towards extremely vocal or often-encountered species. Calls not selected to build a tree were referred to as “out-of-bag” (OOB) and were used to validate classification accuracy of the forest. Calls were selected without replacement and each model contained 10,000 trees which was found to produce stable OOB error estimates. The number of call measurements examined at every node was left at the default: the square root of the total number of measurements for each call type. All trees were built until every node contained a single call.

These models were used to estimate the species classification probability (or the probability of assignment to each species) for all calls across all events that had been collected. For calls that were not part of the representative subset used to build the model, probabilities were based on the fraction of 10,000 forest trees that “voted” for a given species, while for those that were used to build the model, only trees where the calls were OOB were used.

Event Classification

In order to create the Random Forest event classifier, a set of predictors was generated from summaries of the call classifiers. For every event, we calculated the mean of the species classification probabilities for each call type from their respective call classifiers. This produced seven predictors for whistles and clicks and six predictors for burst pulses (Table S1). For example, one predictor was the mean probability that all whistles in an event would be classified to striped dolphins (*Stenella coeruleoalba*), while another was the mean probability that all whistles would be classified as short beaked common dolphins (*Delphinus delphis*). A similar set of predictors (mean species classification probabilities) were generated for clicks and burst pulses. For an event, the full set of mean classification probabilities to all species for a call type summed to one. In addition to these mean species classification probabilities, the proportion of each type of call present in the event and the overall number of whistles per minute in each event were included as predictors.

Only species with at least two independent events (one for testing, one for training) were included in creating the event classifier. The Random Forest event model was composed of 15,000 trees, with each tree in the forest created from equal sample sizes as in the call type classifiers. For each tree, a random three events from each species (half of the smallest sample size) were used, with the remaining events designated as OOB, and events were sampled without replacement. Four predictors were examined at each node in a tree and all trees were built until each node contained one event.

For all four Random Forest models (the three call type classifiers and the event classifier), we recorded variable importance for each predictor as the mean decrease in prediction accuracy when the predictors are randomly permuted. In order to assess the performance of each model, we also calculated the classification rate expected under a null model of no relationship between the predictor variables and the species. This was simply taken as the fraction of calls or events represented by that species in the overall data set.

Results

Data Collection

The towed hydrophone array was deployed for 100 d between 8 August and 9 December 2014. A total of 14,949 km of trackline was surveyed in the study area, (Fig. 1) providing 935 h of recordings. Single species detections used in this study include: *Stenella coeruleoalba*, long-beaked common dolphin (*Delphinus capensis*), *Delphinus delphis*, Risso's dolphin (*Grampus griseus*), Pacific white-sided dolphin (*Lagenorhynchus obliquidens*), pilot whale (*Globicephala macrorhynchus*), and killer whale (*Orcinus orca*, Table 2). There were sightings of several additional species that could not be considered for this study due to poor quality or lack of recordings, including: bottlenose dolphins (*Tursiops truncatus*), pygmy killer whale

(*Feresa attenuata*), false killer whales (*Pseudorca crassidens*), and northern right-whale dolphins (*Lissodelphis borealis*). Additionally, Baird's beaked whales (*Berardius bairdii*) produce several call types that overlap with dolphin calls; however, recording complications prevented incorporation of this species for this analysis.

Call Classification

Over two million whistles, clicks, and burst pulses were detected in the 153 single-species events using the automated detectors in PAMGuard (Table 2); a subset of these data (1,000 calls per event) were used to develop the call classifiers. Click detections represent the entire click; however, whistles and burst pulses are detected in fragments, such that any one call may consist of one or more detections (Gillespie *et al*. 2013). The overall OOB correct classification rate for each call type was significantly greater than the 14% expected by chance alone for seven species (43% for whistles, 49% for echolocation clicks, and 43% for burst pulses, Table 3). Correct classification scores varied by species and by call type, as did the distribution of the misclassifications (Table 3). For example, for *S. coeruleoalba* 68% of the whistles, 40% of the echolocation clicks, and 54% of the burst pulses were correctly classified (Table 3).

The variable importance ranks in the Random Forest models varied by species (Fig. 2). Many variables that were high ranking for some species were of low importance for other species. There was less variability in the variable importance ranks for burst pulses. For this call type, the top ranking variables for all species were the delta variables, suggesting that inter-pulse interval may be important for species recognition (Fig. 2a).

Event Classification

The event classifier was built on the mean species assignment probability for all two million whistles, clicks, and burst pulses. Because *G. macrorhynchus* (*Gm*) and *O. orca* (*Oo*) were represented by only a single event, they were eliminated from the event classifier (although their calls were used in the call classifiers). The overall correct classification rate of 84% for the event classifier was greater than that expected by chance (39%, Table 4).

Short-beaked common dolphins (*D. delphis*) were misclassified to the greatest number of other species, with most misclassifications being *D. capensis*. Likewise, *D. capensis* had a single misclassification as *D. delphis* as well as one as *S. coeruleoalba*. However, the misclassification rate between *D. capensis* and *D. delphis* (1:5 *vs.* 11:96, respectively) was not significantly different (95% confidence interval of the difference in proportions = −0.36–0.55), as the sample size for the former is quite small.

Figure 3 illustrates the overlap of events for each species in the Random Forest model. It can be seen that although there were a large number of *D. delphis* events, most tend to be similar to one another and clustered together. However, there is still enough variability in this species to slightly overlap with *D. capensis* and *S. coeruleoalba*, leading to the misclassifications observed. Considerable overlap between *G. griseus* and *L. obliquidens* lead to one misclassification for each species (Table 4), but do not have significant overlap with the other three species.

The relative importance of the event predictor variables varied by species (Fig. 4). Two of the more important variables for predicting all species but *L. obliquidens* were the percent of whistles and burst pulses classified as *D. delphis*. The most important predictor for *D. capensis* was the percent of burst pulses classified to this species (bp.D capensis), however, this variable was only very important for this species. For *L. obliquidens*, the most important predictor was the fraction of echolocation clicks classified to this species.

Discussion

Acoustic classification of dolphins is complicated by the large diversity of species, overlap in call characteristics among species, individual variability in whistle characteristics, and the volume of data to be analyzed. Application of passive acoustic methods requires a large degree of automation, consideration of all species within a geographic region, and a relatively high level of classification certainty. Significant advances have been made on all three of these requisite needs for both whistles and echolocation clicks; however, a coherent, user-friendly classification method is needed. To our knowledge, this is the first study to attempt to combine detection and classification of all call types produced by dolphins to provide classification results at the level of the dolphin school (event).

Call Classification

The overall performance of the whistle classifier was similar to other studies (Oswald *et al*. 2007, 2013; Keating *et al.* 2015); however, species-level differences in correct classification scores suggest there is room for improvement. In particular, correct classification scores for *G. macrorhynchus*, *O. orca*, and the *Delphinus* species were lower than have been found in other studies. Given the large number of variables measured by ROCCA, it is unlikely that additional measures for individual whistles would provide added discriminatory power. However, filtering out low-intensity calls and false positive detections, or consideration of “delta” variables, such as those used in the burst pulse detector, may provide improvements in whistle classification.

The echolocation click classifier also provided mediocre performance on its own. However, it also has the greatest room for improvement as we only considered a small set of 11 click measures in this study and there are additional features that could be considered. For example, measures could be derived from the spectrum that compare the intensity at different frequencies to identify the possible “peaks” and “valleys” that have been useful for identifying *L. obliquidens* and *G. griseus* (Soldevilla *et al*. 2008). Also, preliminary analysis of similar data has shown that filtering out extremely low and high intensity clicks can improve classification results for echolocation clicks (JNO, unpublished data). Recent studies have also found that species discrimination of dolphin clicks may be improved by consideration of interclick interval (Frasier 2016). In addition, future studies should include cepstral feature analysis of echolocation clicks to provide improved resolution of these broadband impulse sounds across different platforms and noise environments.

This study includes a burst pulse classifier for dolphins based on a newly developed burst pulse detector implemented in the PAMGuard whistle and moan detector (JK, unpublished data). The burst pulse detector is essentially a variation of the whistle detector with slightly different settings, with changes to the method of crossing/joining as well as the connection type, minimum length and minimum total size (Table 1). In fact, burst pulses are detected by the whistle detector, and whistles are detected by the burst pulse detector. Nonetheless, despite relatively small differences in the detector settings between the whistle and burst pulse detectors, they resulted in very different numbers of detections for the same data (Table 2). Likewise, there were differences in the patterns of misclassification for the whistle and burst pulse classifiers (Table 3). Burst pulses are comprised of clicks with very short ICI, but they typically appear as stacked tonal calls in spectrographs (similar in appearance to harmonics, but with consistent frequency bands). The frequency separation between each of the stacked tonal sounds is related to the ICI (Wenz 1964). The two delta variables were intended to provide a measure of ICI for burst pulses by identifying the frequency difference between successive tonal calls. These two variables were consistently important for species classification (Fig. 2). The overall results of the burst pulse classifier were not impressive; however there is likely room for improvement by consideration of alternative delta measures or changing the filters applied to the data. For example, all calls with a beginning frequency of less than 20 kHz were omitted from this classifier; further analysis may identify preferred filter settings to maximize classification rate. Likewise, consideration of an alternative burst pulse detector tuned to burst pulses below 20 kHz may improve discrimination of some species.

Event Classification

Each independent call classifier provided an estimate of species identity for each call detected within an event. Studies have shown that correct classification of species identity at the level of the dolphin school (event) is much improved over the classification of individual calls (Oswald *et al.* 2013, Keating *et al*. 2015), and consideration of multiple call types can further improve classification results (Roch *et al.* 2011). Despite unexceptional results for each of the three call classifiers, when combined, they allow for a dramatic increase in correct classification, with an overall correct classification rate of 84%.

Previous studies found classification of *Delphinus* species to be extremely difficult (Oswald *et al*. 2007, 2013; Roch *et al*. 2011; Lin and Chou 2015) and thus, we had low expectations for our ability to differentiate *Delphinus* species from each other (or from *S. coeruleoalba*) at the event level. Although most species had high correct classification scores in the event classifier, only 71% of *Delphinus capensis* (Dc) events were correctly classified, with one event misclassified as *Delphinus delphis* (Dd) and another as *Stenella coeruleoalba* (Sc). *D. delphis* was represented by a very large sample size for each of the three call types as well as at the event level, providing an overwhelming amount of data to characterize the vocal characteristics of this species, and to differentiate them from other species for most events.

Typically, when Random Forest models are presented with unequal sample sizes, they will tend to yield high correct classification scores for the dominant class, at the expense of the other classes. To mitigate this effect, our models used equally-sized subsets of calls, although a slight advantage may remain for *D. delphis* due to better representation of the variability within the subset. Sample sizes for *D. capensis* were extremely low in comparison to *D. delphis*. A larger sample size should provide improved resolution of the distribution of the vocal characteristics for this species, and better discrimination from the closely related *D. delphis*. If the model is constructed with sufficient sample sizes to define call and event characteristics for all species, then misclassifications can be attributed to "outliers." In this case, outliers may provide additional information about the group composition or behavior.

Improvements and Future Applications

The hierarchical approach of BANTER allows for independent incremental improvement of both the individual call classifiers as well as the overall event classifier. Specifically, improvements to the stability in classification (and misclassification) at the call classifier level will improve the overall event classification. Classification stability for individual call classifiers may be improved by considering additional features or reducing errors in call detection or measurement. Additional information used to "describe" the call characteristics can be considered, such as the delta variables included in the burst pulse classifier. The click classifier has the greatest potential for improvement by considering additional measures, especially ones more immune to environmental variability (such as cepstral measures). We also suggest that future studies should use a generic "dolphin" click detector rather than the suite of detectors used in this study. While this will most likely have little or no effect on the classification rates at the click classification or event level, it will allow for a more simple and standardized application of these methods. The event classifier could be further improved by the consideration of additional event-level acoustic measures (*e.g.*, mean ICI) or environmental features (*e.g.*, water depth or temperature).

At the event classification level, the results of each call classifier provide an alternative view of the overall event. For the whistle and burst pulse detectors, these alternative views are of the *same calls,* with an overlap in call detection for the two detectors. While these detectors may not exclusively identify a particular call type, sampling these calls with slightly different WM detector settings changed number and fragmentation of the detections, which lead to differences in the distribution of the call classifications. Theoretically, a third WM detector, using alternative settings, may provide yet another set of fragmented detections with a different distribution of call classification. Again, additional WM detectors may not identify specific call types, but rather they provide additional views of the data that may improve discrimination of some species at the event level.

Our initial goal was to include all dolphin species whose normal distribution fell within the California Current. During this survey, anomalously warm waters resulted in the unusual detection of several tropical species in the study area, including: *F. attenuata*, *P. crassidens*, and *G. macrorhynchus*. Because we required a minimum of two independent events to include a species in the classifier, these rare species were excluded from the model. Data recording errors further eliminated other species of interest, resulting in a classifier that is not complete for the study area. Given the likelihood of changes in cetacean distribution due to climate change, future acoustic classifiers should attempt to consider all species that *may* fall within the area of interest.

In theory, once the single-species classification results are consistently strong, BANTER can be used to identify mixed-species events. One of the strengths of the Random Forest algorithm is that it provides not only a species classification, but also a measure of the strength of that classification in the form of the assignment probabilities. If a single-species event classifier has large assignment probabilities to the correct species (*i.e.*, a large percentage of all events are being correctly assigned and they are being assigned with high certainty), then mixed-species events should be represented by a mix of assignment probabilities that reflects the species composition of the school.

A preliminary examination of the membership probability for mixed-species events including *D. delphis* and *S. coeruleoalba* identified potential behavioral complications to this effort. In this study, *S. coeruleoalba* tended to avoid the vessel while *D. delphis* tended to approach the vessel. When combined with the high vocal rate of *D. delphis*, this difference in approach behavior may tend to lead to situations where recordings are dominated by the vocalizations of *D. delphis* within mixed-species schools. This complication does not preclude the possibility of acoustic classification of mixed-species schools; however, it must be considered. One means of improving these results would be to apply classifications to localized calls to identify species composition of subgroups.

The data for this analysis were collected using a single hydrophone within a hydrophone array towed behind a single vessel. We expect that our results would be applicable to data collected in the same manner using the same equipment. Future tests should examine how differences in hardware (*e.g.*, hydrophone sensitivity) or vessel and environmental noise affect variables used in the classifiers, and identify variables that are relatively immune to these differences (see Roch *et al*. 2015). Exclusion of variables that are not stable across platforms will ultimately improve the robustness and overall efficacy of the classifier. Also, a robust classifier that is insensitive to differences in background noise or hydrophone sensitivity is more useful in a wider range of applications.

Hydrophone depth may also affect the variables used in this classifier. Studies have shown variation in the measurement of echolocation clicks characteristics based on hydrophone depth (Baumann-Pickering *et al.* 2013), and studies are currently being conducted to test how hydrophone depth affects whistle and click measurements and classification (JNO, unpublished data). In addition to the mean species assignment probabilities of calls, the event classifier also considers the proportion of calls by call type. The detection of calls can vary based on the depth, distance, and orientation of the sound source (dolphin) and receiver (hydrophone); therefore, the proportion of calls will be affected by the location of the source and receiver (especially in relation to the thermocline). This classifier was developed and tested using recordings collected from a relatively shallow hydrophone array towed at 12 m depth; we would expect lower correct classification scores for data collected at depth.

Data used in this study were collected from a towed hydrophone survey conducted in “closing mode,” where all visually obtained detections were approached and most recordings were made in close physical proximity to the dolphin schools. The classification results provided in this study pertain to this type of survey, and the classifiers may not have the same level of performance for calls detected at greater distances. All call types are subjected to degradation during propagation, and distant (faint) calls will have different call characteristics than calls produced in closer proximity to the hydrophone. Future studies should examine the effect of distance on call classification, and ideally call classifiers may be adjusted based on distance between the vocalizing dolphin school and the recorder.

Conclusions

The results of this study show that BANTER should prove to be a simple and effective tool for acoustic research, management, and mitigation. The high correct classification scores we observed can be attributed to the hierarchical approach. This approach considers data collected with automated systems and minimal human intervention, providing more consistent results with fewer biases and errors. The flexible framework also allows for incremental testing and improvement of the classifier at the call and event level. Relatively minor improvements to single-species classification and localization should allow for classification of mixed species aggregations. We expect that improvements to computer processing power will allow real-time use of this classifier in the near future. BANTER may be implemented as-is to classify data collected using similar survey design in the California Current. BANTER is currently being implemented into a new real-time ROCCA classifier within PAMGuard (expected release date early 2017). The flexible methodology we have developed can also be easily modified for application in other geographic regions with different species composition, or perhaps even for other *noncetacean* taxa.

Acknowledgments

This study could not have been performed without the help of the officers and crew of the R/V *Ocean Starr*. Special thanks to the scientists for their cooperation, especially Emily Griffiths for her work as a lead acoustician. Tina Yack and Kerry Dunleavy of Bio-Waves provided a beta version of the ROCCA click detector. Funding for the field survey was provided by Southwest Fisheries Science Center, U.S. Navy, and BOEM. Funding for the analysis was provided by NOAA's Advanced Science and Technology Working Group, and from Southwest Fisheries Science Center. This manuscript was improved by helpful comments from Jim Carretta and two anonymous reviewers.

Literature Cited

Au, W. W. L. 1993. The sonar of dolphins. Springer-Verlag, Berlin, Germany.

Au, W. W. L., M. O. Lammers and S. Yin. 2010. Acoustics of dusky dolphins. Pages 75–97 *in* B. Würsig and M. Würrsig, eds. The dusky dolphin: Master acrobat off different shores, Academic Press, San Diego, CA.

Azzolin, M., E. Papale, M. O. Lammers, A. Gannier and C. Giacoma. 2013. Geographic variation of whistles of the striped dolphin (*Stenella coeruleoalba*) within the Mediterranean Sea. Journal of the Acoustical Society of America 134:694–705.

Azzolin, M., A. Gannier, M. O. Lammers, J. N. Oswald, E. Papale, G. Buscaino and C. Giacoma. 2014. Combining whistle acoustic parameters to discriminate Mediterranean odontocetes during passive acoustic monitoring. Journal of the Acoustical Society of America 135:502–512.

Barlow, J., and B. L. Taylor. 2005. Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. Marine Mammal Science 21:429–445.

Baumann-Pickering, S., S. M. Wiggins, J. A. Hildebrand, M. A. Roch and H. U. Schnitzler. 2010. Discrimination of echolocation clicks of melon-headed whales (*Peponocephala electra*), bottlenose dolphins (*Tursiops truncatus*), and Gray’s spinner dolphins (*Stenella longirostris longirostris*). Journal of the Acoustical Society of America 128:2212–2224.

Baumann-Pickering, S., T. M. Yack, J. Barlow, S. M. Wiggins and J. A. Hildebrand. 2013. Baird’s beaked whale echolocation signals. Journal of the Acoustical Society of America 133:4321-4331.

Baumann-Pickering, S., A. E. Simonis, E. M. Oleson, R. W. Baird, M. A. Roch and S. M. Wiggins. 2015. False killer whale and short-finned pilot whale acoustic identification. Endangered Species Research 28(2):97–108.

Blomqvist, C., and M. Amundin. 2004. High-frequency burst-pulse sounds in agonistic/aggressive interactions in bottlenose dolphins, *Tursiops truncatus*. Pages 425–431 *in* J. A. Thomas, C. F. Moss and M. Vater, eds. Echolocation in bats and dolphins. University of Chicago Press, Chicago, IL.

Breiman, L. 2001. Random forests. Machine Learning 45:5–32.

Frasier, K. E., E. Henderson, H. R. Bassett and M. A. Roch. 2016. Automatic identification and clustering of subunits within delphinid vocalizations. Marine Mammal Science

32:911–930.

Gallus, A., M. Dähne, U. K. Verfuß, S. Bräger, S. Adler, U. Siebert and H. Benke. 2012. Use of static passive acoustic monitoring to assess the status of the 'Critically Endangered' Baltic harbour porpoise in German waters. Endangered Species Research 18:265–278.

Gillespie, D., P. Berggren, S. Brown, *et al*. 2005. Relative abundance of harbor porpoises (*Phocoena phocoena*) from acoustic and visual surveys in the Baltic Sea and adjacent waters during 2001 and 2002. Journal of Cetacean Research and Management 7:51–57.

Gillespie, D., D. K. Mellinger, J. Gordon, *et al*. 2008. PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. Proceedings of the Institute of Acoustics 30:54–62.

Gillespie, D., M. Caillat, J. Gordon and P. White. 2013. Automatic detection and classification of odontocete whistles. Journal of the Acoustical Society of America 134:2427–2437.

Herman, L. M., and W. N. Tavolga. 1980. Communication systems of cetaceans. Pages 149–197 *in* L. M. Herman, ed. Cetacean behavior: Mechanisms and function. Wiley Interscience, New York, NY.

Herzing, D. L. 1996. Vocalizations and associated underwater behavior of free-ranging Atlantic spotted dolphins, *Stenella frontalis*, and bottlenose dolphins, *Tursiops truncatus*. Aquatic Mammals 22:61–79.

Herzing, D. L. 2000. Acoustics and social behavior of wild dolphins: implications for a sound society. Pages 225–272 *in* W. W. L. Au, A. N. Popper and R. R. Fay, eds. Hearing by

whales and dolphins. Springer-Verlag, New York, NY.
Janik, V. M., and P. J. B. Slater. 1998. Context-specific use suggests that bottlenose dolphin signature whistles are cohesion calls. Animal Behavior 56:829–838.
Keating, J. L., J. N. Oswald, S. Rankin and J. Barlow. 2015. Whistle classification in the California Current: A complete whistle classifier for a large geographic region with high species diversity. U.S. Department of Commerce, National Marine Fisheries Service Technical Memorandum NOAA-TM-NMFS-SWFSC-552 (unpublished). 13 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037.
Kinzey, D., P. Olson and T. Gerrodette. 2000. Marine mammal data collection procedures on research ship line-transect surveys by the Southwest Fisheries Science Center. Southwest Fisheries Science Center Administrative Report LJ-00-08 (unpublished). 32 pp. Available from SWFSC, 8901 La Jolla Shores Dr., La Jolla, CA 92037.
Liaw, A., and M. Wiener. 2002. Classification and regression by randomForest. R News 2(3):18–22.
Lin, T. H., and L. S. Chou. 2015. Automatic classification of delphinids based on the representative frequencies of whistles. Journal of the Acoustical Society of America 138:1003–1011.
McDonald, M. A., S. L. Mesnick and J. A. Hildebrand. 2006. Biogeographic characterization of blue whale song worldwide: Using song to identify populations. Journal of Cetacean Research and Management 8:55–65.
Mellinger, D. K. 2001. ISHMAEL 1.0 user’s guide: Integrated System for Holistic Multichannel Acoustic Exploration and Localization. U.S. Department of Commerce, NOAA Technical

Memorandum OAR-PMEL-120 (unpublished). 30 pp. Available from PMEL, 7600 Sand Point Way NE, Seattle, WA 98115.

Oswald, J. N., S. Rankin, J. Barlow and M. O. Lammers. 2007. A tool for real-time acoustic species identification of delphinid whistles. Journal of the Acoustical Society of America 112:587–595.

Oswald, J. N., S. Rankin and J. Barlow. 2008. To whistle or not to whistle? Geographic variation in the whistling behavior of small odontocetes. Aquatic Mammals 34:288–302.

Oswald, J. N., S. Rankin, J. Barlow and M. Oswald. 2013. Real-time odontocete call classification algorithm: Software for species identification of delphinid whistles. Pages 245–266 *in* O. Adam and F. Samaran, eds. Detection, classification, and localization of marine mammals using passive acoustics, 2003–2013: 10 years of international research. DIRAC NGO, Paris, France.

Papale, E., M. Azzolin, I. Cascao, *et al*. 2013. Macro and micro geographic variation of short-beaked common dolphin’s whistles between the Mediterranean Sea and Atlantic Ocean. Ethology Ecology and Evolution 26:392-404.

R Core Development Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rankin, S., J. Oswald, J. Barlow and M. Lammers. 2007. Patterned burst-pulse vocalizations of the northern right whale dolphin, *Lissodelphis borealis*. Journal of the Acoustical Society of America 121:1213–1218.

Rankin, S., J. Barlow and J. Oswald. 2008*a*. An assessment of the accuracy and precision of localization of a stationary sound source using a two-element towed hydrophone array.

U.S. Department of Commerce, National Marine Fisheries Service Technical Memorandum NOAA-TM-NMFS-SWFSC-416 (unpublished). 29 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037.

Rankin, S., J. Barlow, J. Oswald and L. Ballance. 2008*b*. Acoustic studies of marine mammals during seven years of combined visual and acoustic line-transect surveys for cetaceans in the eastern and central Pacific Ocean. U.S. Department of Commerce, National Marine Fisheries Service Technical Memorandum NOAA-TM-NMFS-SWFSC-429 (unpublished). 58 pp. Available from SWFSC, 8901 La Jolla Shores Drive, La Jolla, CA 92037.

Rankin, S., J. N. Oswald and J. Barlow. 2008*c*. Acoustic behavior of dolphins in the Pacific Ocean: Implications for using passive acoustic methods for population studies. Canadian Acoustics 36:88–92.

Roch, M. A., H. Klinck, S. Baumann-Pickering, *et al*. 2011. Classification of echolocation clicks from odontocetes in the Southern California Bight. Journal of the Acoustical Society of America 129:467–4475.

Roch, M. A., J. Stinner-Sloan, S. Baumann-Pickering and S. M. Wiggins. 2015. Compensating for the effects of site and equipment variation on delphinid species identification from their echolocation clicks. The Journal of the Acoustical Society of America 137:22–29.

Rodriguez, A., and A. Laio. 2014. Clustering by fast search and find of density peaks. Science 344:1492–1496.

Sayigh, L. S., H. C. Esch, R. S. Wells and V. M. Janik. 2007. Facts about signature whistles of bottlenose dolphins, *Tursiops truncatus*. Animal Behaviour 74:1631–1642.

Soldevilla, M. S., E. E. Henderson, G. S. Campbell, S. M. Wiggins, J. H. Hildebrand and M. A. Roch. 2008. Classification of Risso's and Pacific white-sided dolphins using spectral properties of echolocation clicks. Journal of the Acoustical Society of America 124:609-624.

Wenz, G. M. 1964. Curious noises and the sonic environment in the ocean. Pages 101–119 *in* W. N. Tavolga, ed. Marine bio-acoustics. Pergamon, New York, NY.

Received: 23 June 2016

Accepted: 23 October 2016

Figure 1. Map showing shipboard acoustic line-transect trackline coverage using a towed hydrophone array during the 2014 CalCurCEAS (California Current Cetacean Ecosystem Assessment Survey) off the west coast of the United States.

Figure 2. Ranks of variable importance from each call type Random Forest classification model for each species. Colors scale from most important (dark red) to least important (dark blue). Variables have been filtered to include the most important predictor for each species, identified by a "1."

Figure 3. Proximity plot for species events from Random Forest model. Central color of dots represent the true species identity, while color of circle surrounding the dot is the predicted species based on the event classifier. The solid lines connect the outermost events for each species.

Figure 4. Ranks of variable importance from event Random Forest classification model for each species. Colors scale from most important (dark red) to least important (dark blue).

[1] Corresponding author (e-mail: shannon.rankin@noaa.gov).

Table 1. Settings for PAMGuard Whistle and Moan detector for automated detection of whistles and burst pulses.

Maximum frequency	Connection type	Minimum length	Minimum total size	Crossing/joining	Maximum cross length	Median filter length	Subtraction constant	Smoothing	Threshold
37 kHz	8 sides/diagonals	10 slices	50 pixels	Relinkin	5 slices	61	0.02	ON	8 dB
36 kHz	4 sides only	7 slices	10 pixels	Discard branched regions	OFF	61	0.02	ON	7 dB

Table 2. Sample sizes for each species. Species names (scientific and common name) and species code are given for all species included in this study. The total number of unique events per species is given, as well as the number of calls per species for whistles, echolocation clicks, burst pulses. For species with fewer than two events, marked with an asterisk (*), calls were used for development of the call classifier but not for final event classification.

Species name	Common name	Species code	# Events	Whistles	Echolocation clicks	Burst pulses	Total
Stenella coeruleoalba	striped dolphin	Sc	13	8,271	29,146	142	37,559
Delphinus capensis	long-beaked common dolphin	Dc	7	61,902	119,358	13,529	194,789
Delphinus delphis	short-beaked common dolphin	Dd	116	501,288	982,187	138,170	1,621,645
Grampus griseus	Risso's dolphin	Gg	5	570	50,286	4,047	54,903
Lagenorhynchus obliquidens	Pacific white-sided dolphin	Lo	10	4,794	91,787	5,076	101,657
*Globicephala macrorhynchus**	pilot whale	Gm	1	734	5,401	157	6,292
*Orcinus orca**	killer whale	Oo	1	2	1,316	0	1,318
		Total	153	577,561	1,279,481	161,121	2,018,163

Table 3. Confusion matrices (number of calls classified as each species), out-of-bag correct classification scores, and correct classification scores for call classifier for (a) whistles, (b) echolocation clicks, (c) burst pulses. The species codes represent the following species: *Stenella coeruleoalba* (Sc), *Delphinus capensis* (Dc), *Delphinus delphis* (Dd), *Grampus griseus* (Gg), *Lagenorhynchus obliquidens* (Lo), *Globicephala macrorhynchus* (Gm), and *Orcinus orca* (Oo). No burst pulses were detected for *Orcinus orca*.

(a) Whistles: Out-of-bag estimate of correct classification rate: 43%.

	Sc	Dc	Dd	Gg	Lo	Gm	Oo	Correct classification
Sc	2,483	399	297	12	316	150	0	68%
Dc	1,270	757	853	198	541	114	0	20%
Dd	999	578	1,254	259	459	64	0	35%
Gg	28	3	11	471	34	23	0	83%
Lo	542	183	99	611	1,240	388	0	40%
Gm	196	34	48	31	77	348	0	47%
Oo	2	0	0	0	0	0	0	0%

(b) Echolocation clicks: Out-of-bag estimate of correct classification rate: 49%.

	Sc	Dc	Dd	Gg	Lo	Gm	Oo	Correct classification
Sc	2,911	599	128	660	715	874	1468	40%
Dc	448	2,734	819	956	271	222	241	48%
Dd	664	1,809	1,622	947	564	387	425	25%
Gg	325	406	152	3,312	202	240	40	71%
Lo	214	48	56	129	3,753	264	434	77%
Gm	605	192	109	427	577	1,780	533	42%

Oo	260	29	10	34	98	111	608	53%

(c) Burst pulses: Out-of-bag estimate of correct classification rate: 43%.

	Sc	Dc	Dd	Gg	Lo	Gm	Oo	Correct classification
Sc	76	15	10	8	3	30	—	54%
Dc	1,107	2,034	872	562	543	442	—	37%
Dd	649	1,217	2,030	741	616	708	—	34%
Gg	101	68	87	2,609	433	247	—	74%
Lo	331	463	374	1,119	1,581	427	—	37%
Gm	30	9	24	16	20	58	—	37%
Oo	—	—	—	—	—	—	—	—

Table 4. Confusion matrix (showing number of events classified as each species, by species), out-of-bag correct classification rate, and correct classification scores (with expected correct classification results in parentheses) for the event classifier. The species codes represent the following species: *Stenella coeruleoalba* (Sc), *Delphinus capensis* (Dc), *Delphinus delphis* (Dd), *Grampus griseus* (Gg), *Lagenorhynchus obliquidens* (Lo). Expected classification scores are based on the proportion of all events composed of by each species.

Out-of-bag estimate of correct classification: 84% (Expected = 39%)

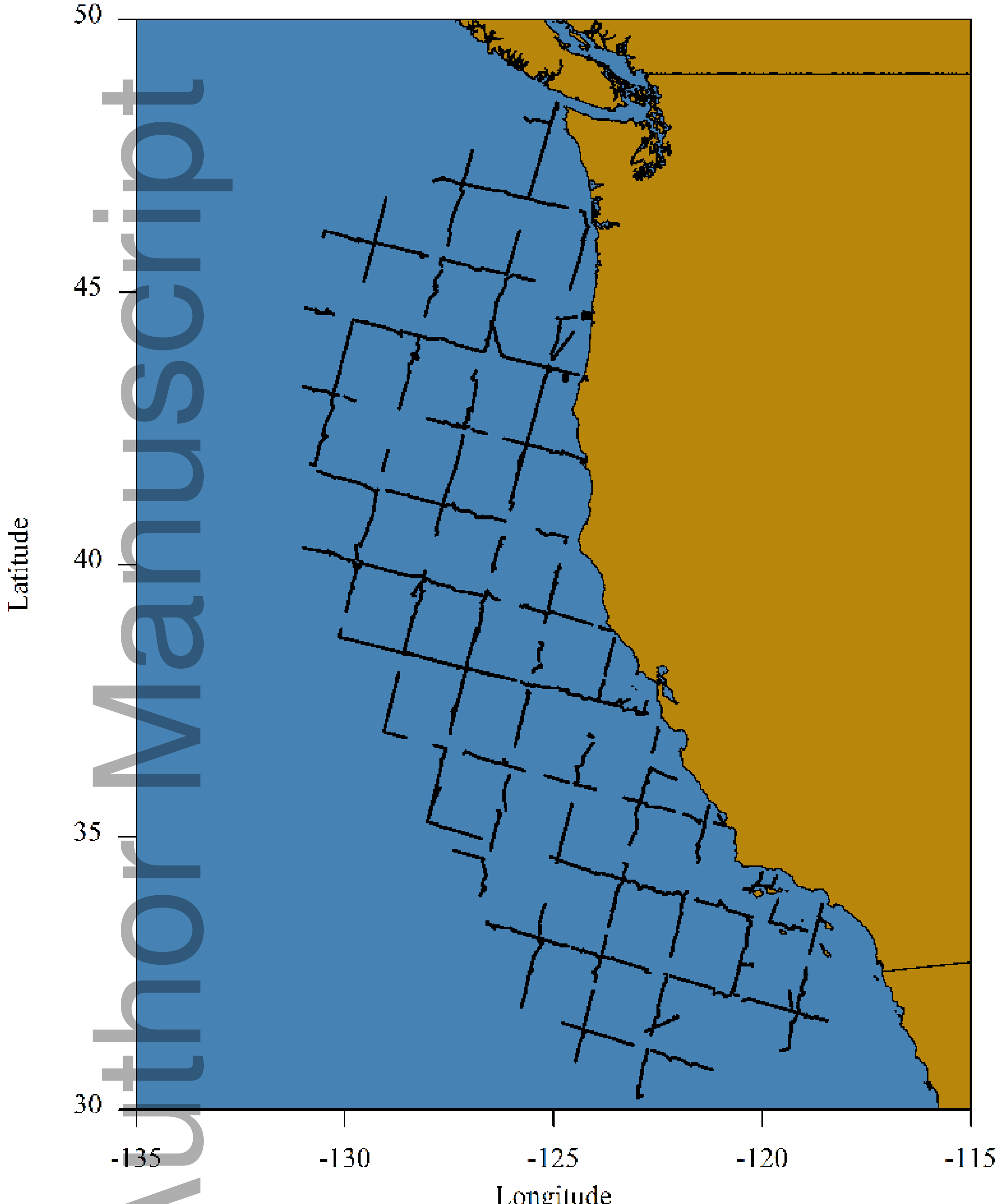
	Sc	Dc	Dd	Gg	Lo	Correct classification
Sc	12	1	0	0	0	92% (7%)
Dc	1	5	1	0	0	71% (5%)
Dd	7	11	96	1	0	83% (77%)
Gg	0	0	0	4	1	80% (4%)

Lo	0	0	0	1	9	90% (7%)

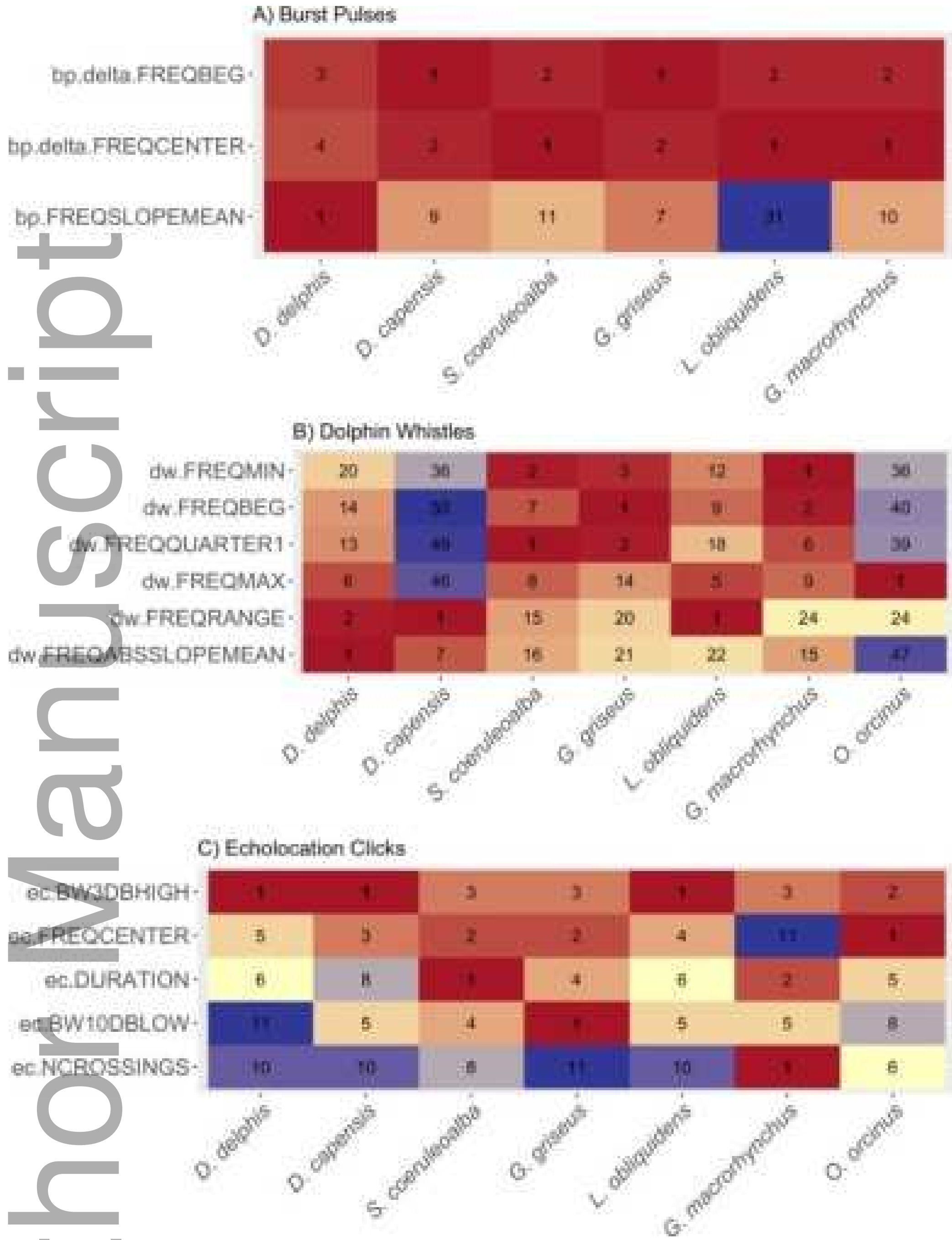
Supporting Information

The following supporting information is available for this article online:

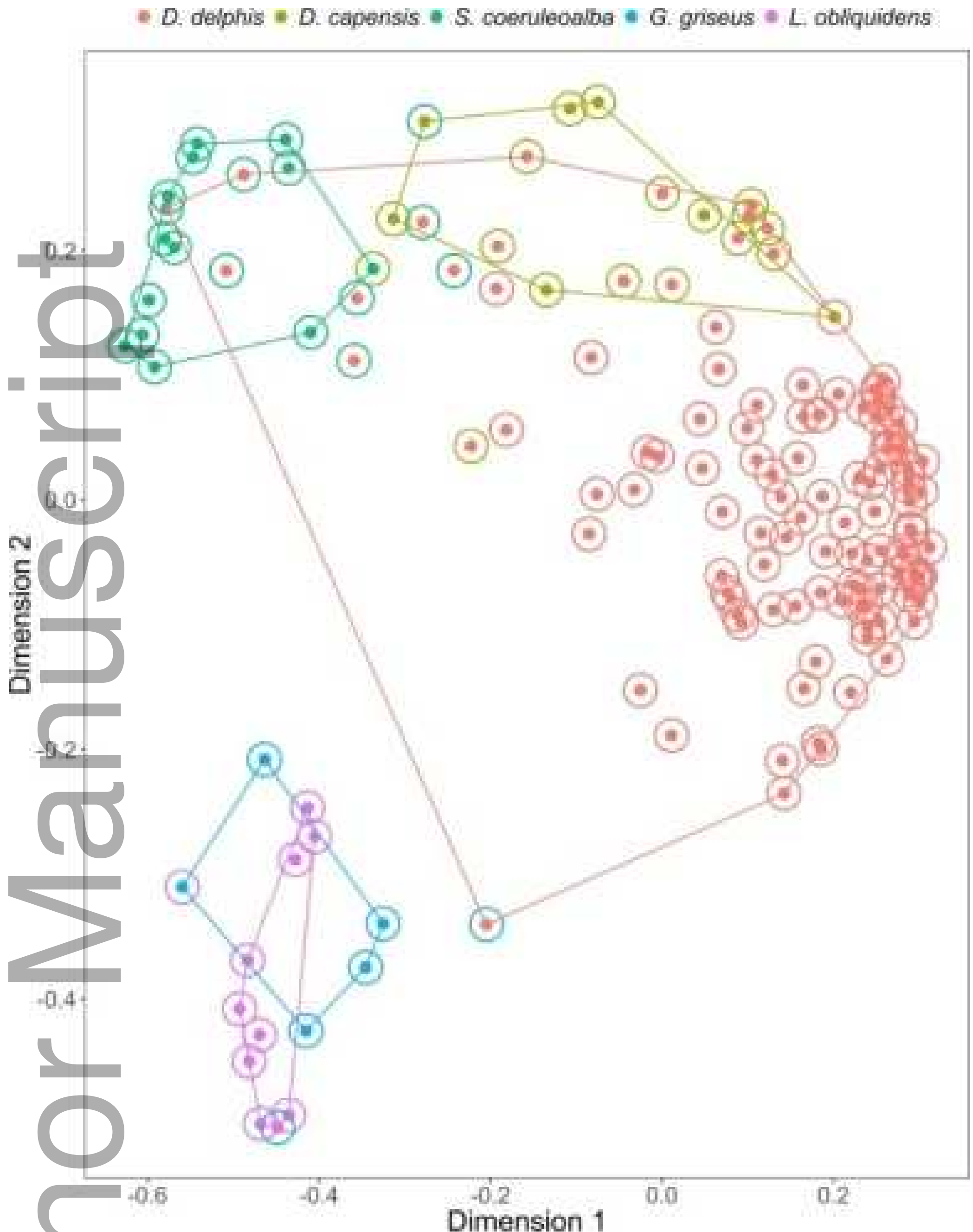
Table S1. Description of variables used in call type and event Random Forest models. Classifier models include whistle detector (DW), burst pulse detector (BP), echolocation click detector (EC), and event classifier.



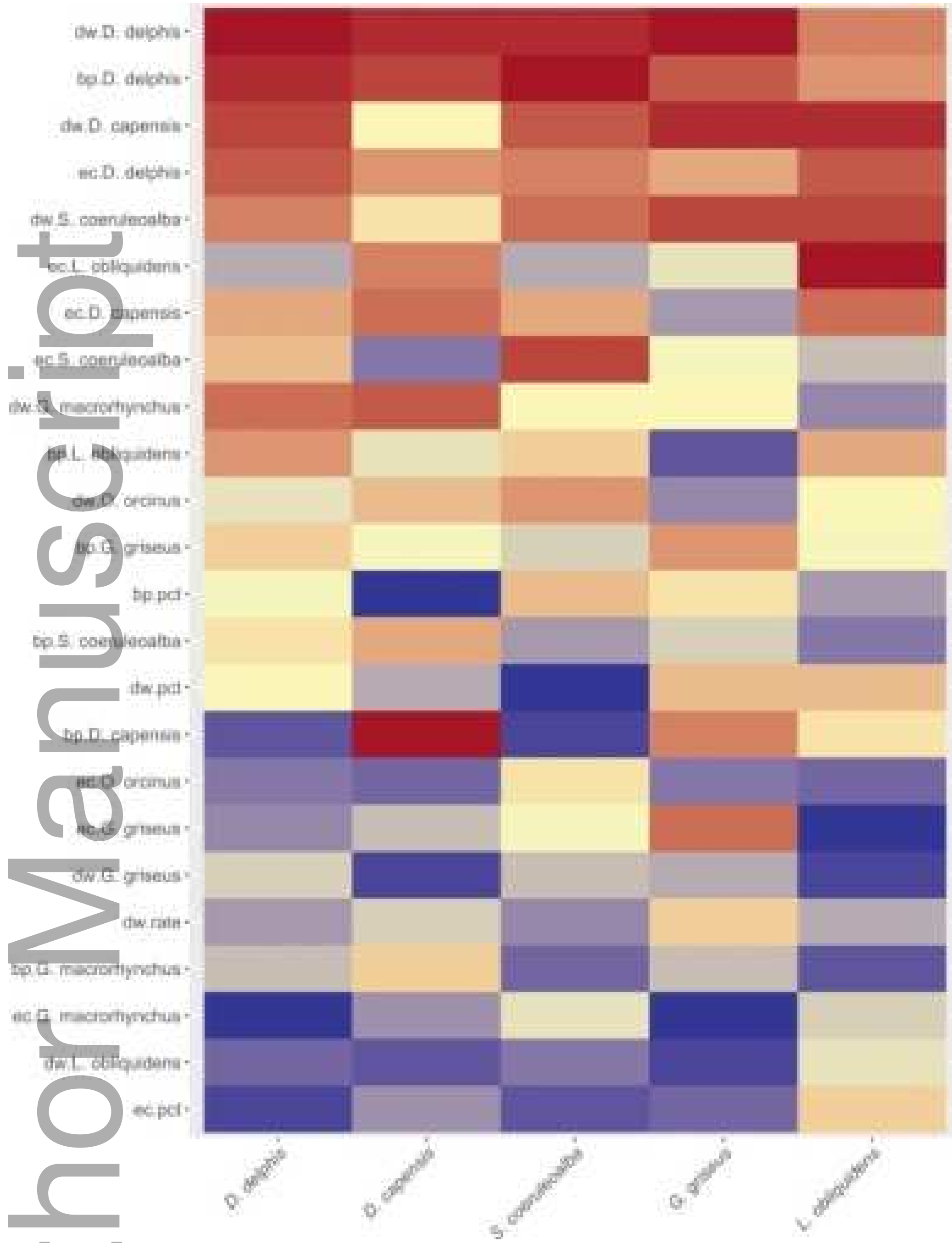
mms_12381_f1.tif



mms_12381_f2.tiff



mms_12381_f3.tiff



mms_12381_f4.tiff