Final Report

Increasing Sample Size for Examining Changes in Vocal Behavior in Relation to Navy Sonar Activity

Submitted to:

Naval Facilities Engineering Command Atlantic under Contract No. N62470-10-D-3011, Task Order 21, issued to HDR, Inc.

Prepared by:

Bio-Waves, Inc.
364 2nd Street, Suite #3
Encinitas, CA 92024
(760) 452-2575

and

Centre for Research into Ecological and Environmental Modelling
University of St. Andrews
St. Andrews, Fife
Scotland

Submitted by:

HDR
Virginia Beach, VA

November 2016

Suggested Citation:

Oswald, J.N., C.S. Oedekoven, T.M. Yack., C.G.M. Paxton, V. Popov, L. Thomas, R. Walker, E. Ferguson, and T. Norris. 2016. Increasing Sample Size for Examining Changes in Vocal Behavior in Relation to Navy Sonar Activity. Final Report. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia, under Contract No. N62470-10-3011, Task Order 21, issued to HDR, Inc. November 2016.

Photo:

Atlantic spotted dolphins *(Stenella frontalis)* taken by Heather Foley, Duke University. Photo taken under NOAA Permit No. 16185.

Work conducted under following contract between Bio-Waves, Inc. and HDR, Inc.
MSA #: CON-005-4394-009
Subproject #164744, Agreement #105067, Task CTO21

This project is funded by US Fleet Forces Command and managed by Naval Facilities Engineering Command Atlantic as part of the US Navy's marine species monitoring program.

Executive Summary

The aim of this study was to develop and refine robust statistical methods to evaluate the effects of Navy sonar on the acoustic behaviors of cetaceans. Data collected using High-frequency Acoustic Recording Packages (HARPs) deployed off Jacksonville, Florida (n=1), Onslow Bay, North Carolina (n=2), and Cape Hatteras, North Carolina (n=1) were combined with data previously analyzed from Marine Autonomous Recording Units (MARUs) off Jacksonville and Onslow Bay (Oswald et al. 2015). These data were reviewed to log the occurrence of Navy sonar and sounds produced by marine mammals for the 24-hour period before sonar began, during the entire sonar event, and for the 24-hour period immediately following the end of sonar. Trained analysts scrolled through Long Term Spectral Averages (LTSAs) using the software package Triton (Wiggins 2007) with a 30-minute resolution setting to identify sonar events and acoustic encounters for minke whales, sperm whales, beaked whales, and delphinid species. Detections were grouped into "acoustic encounters," where an acoustic encounter was defined as a series of sounds produced by one species with less than 30 minutes between sounds. For delphinid acoustic encounters the number of sounds (whistles, clicks, buzzes) per minute within each encounter was tallied using PAMGuard's automated whistle & moan detector and click detector modules, as well as a custom written MATLAB script (Bin-it Counter.m). For beaked whales and sperm whales, classifiers within PAMGuard were used to classify clicks, which were subsequently marked by encounters. Clicks were analyzed and tallied for each encounter. Minke whale encounters were "sub-logged" (e.g., examined at a finer scale) to identify the start time, end time for each pulse train, and type of pulse trains within each encounter.

Spectrogram correlation templates for four types of active sonar (e.g. in frequency bands: 1 to 3 kilohertz [kHz], 3 to 5 kHz, 5 to 7 kHz, and 7 to 10 kHz) were created to detect individual sonar pings in the dataset. Sonar events were reviewed manually to ensure there were no missed pings and that any false detections were removed. Variables describing frequency content, duration, and received sound pressure level were measured from all true positive sonar detections using the MATLAB program SonarFinder (Bio-Waves, Inc. 2013).

Once data had been logged and characterized, five distinct statistical analyses were conducted across four cetacean taxa: delphinids, beaked whales, minke whales, and sperm whales. First, a regression analysis was conducted of cetacean acoustic absence or presence to address whether sonar activity influenced acoustic presence. Second, a hidden Markov model analysis of cetacean acoustic absence or presence was performed to address the same question using a different statistical method. The third analysis was a comparison of minke whale pulse train duration with and without sonar present. The fourth analysis was a regression analysis of dolphin signal type (i.e., whistles, clicks, and buzzes) presence-absence in a given acoustic encounter to address whether sonar occurrence influenced the occurrence of a particular signal type. Finally, a regression analysis of a composite index of delphinid whistles was conducted to examine whether whistle characteristics changed in to the presence of Navy sonar.

We used various covariates related or unrelated to sonar as potential explanatory covariates. Those related to sonar included the time period relative to sonar activity as well as various features of the sonar such as signal type and sound pressure level. When appropriate, data

from this study were added to similar datasets from earlier studies collected by Bio-Waves, Inc. and Cornell University using MARUs in order to increase the sample size and hence power.

For many of the species, acoustic presence or the characteristics of the detected sounds were associated with an effect of datatype (i.e., MARU or HARP), site/location, and/or day of year; however, because of the pattern of data collection, the potential effects of these variables could not be distinguished.

There was evidence that sonar activities influenced the acoustic detections at the study sites. A direct influence of sonar activity (measured as a change in the response variable before, during, between and after sonar activity) on delphinids was not found. However, the presence of a particular sonar signal (Type 1 long signal) slightly increased the probability of delphinid acoustic detection. Beaked whale detections declined after sonar activity had commenced, and this was possibly a response to the Type 1 long signal component. Detection of delphinid whistles, clicks, and buzzes increased during sonar activity and also increased in the presence of Type 1 long or Type 2 long sonar signals. Detection of whistles was negatively associated with the presence of clicks and positively associated with the presence of buzzes. When the effect of a specific component of sonar could be detected, the component was always a long signal.

Acoustic detection of beaked whales showed a noticeable decline during periods of sonar activity, with no evidence of recovery over the course of monitoring. Similarly, minke whale detections showed a decline once sonar activity started, but with some indication of subsequent recovery in the 24 hours after sonar ended. The effect on duration of minke whale pulse trains was equivocal. No effect of sonar on sperm whales was detected, but there was a marked difference between the numbers of day and night acoustic detections.

The results of this study have provided some insights to species-specific responses to Navy sonar and have resulted in the development of innovative statistical methods for assessing the impacts of a naval mid-frequency active sonar on some of the parameters describing the acoustic detections of cetaceans. The statistical methods used here can be applied to additional datasets in other geographic regions to provide further insights into the effects of naval mid-frequency active sonar on marine mammals.

Table of Contents

Figures

Tables

Appendices

Acronyms and Abbreviations

ACF	Autocorrelation Function
AIC	Akaike Information Criterion
BZ	buzz
D_{Mi}	Mahalanobis distances
EC	echolocation clicks
GEE	Generalized Estimating Equation
GLM	Generalized Linear Model
HARP	High-frequency Acoustic Recording Package
HMM	Hidden Markov Model
Hz	Hertz
IPI	inter-pulse interval
kHz	kilohertz
LTSA	Long-term Spectral Average
m	meter(s)
MARU	Marine Autonomous Recording Unit
MFA	mid-frequency active
NA	not any
PSD	Power Spectral Density
PSTGAE	presence of a signal type given a cetacean acoustic encounter
ROCCA	Real-time Odontocete Call Classification Algorithm
SPL	Sound Pressure Level
SNR	signal-to-noise ratio
TPM	transition probability matrix
W	whistles

This page intentionally left blank.

1. Introduction and Background

Passive acoustic monitoring using autonomous recorders deployed on the seafloor is an effective method for long-term monitoring of marine mammals (Mellinger et al. 2007, Van Parijs et al. 2009). Autonomous recorders have been used to investigate the distribution, occurrence, and acoustic behaviors of a variety of marine mammals in diverse habitats and geographic locations (Clark et al. 2002, Clark and Gagnon 2002, Clark and Clapham 2004, Baumgartner et al. 2008, Johnston et al. 2008, Sousa-Lima et al. 2013). Recently, researchers have analyzed autonomous recorder data to investigate the effects of noise, such as seismic airguns and active sonar, on the calling behavior of baleen whales (Nieukirk et al. 2004, Di Iorio and Clark 2010, Castellote et al. 2012, Melcón et al. 2012, Risch et al. 2012). However, only a few studies have focused on analyzing seafloor hydrophone data to examine the effects of naval sonar on the acoustic behaviors of odontocetes (McCarthy et al. 2011, Tyack et al. 2011, Oswald et al. 2015).

Very little is known about the acoustic behavioral responses of odontocetes to mid-frequency active (MFA) sonar. Rendell and Gordon (1999) reported that long-finned pilot whales (*Globicephala melas*) increased whistling rates during and after exposure to military sonar signals. DeRuiter et al. (2013a) analyzed acoustic data collected from DTAGs during controlled-exposure experiments using playbacks of MFA sonar and found that false killer whales (*Pseudorca crassidens*) and melon-headed whales (*Peponocephala electra*) increased whistling rates and appeared to mimic MFA signals immediately after exposures. Based on these results, DeRuiter et al. (2013a) suggested that changes in vocal activity may be one of the primary types of response to acoustic stimuli for delphinids. This may, in part, be because many odontocetes are highly social, occur in groups, and often rely on group awareness to detect and alert others to perceived threats by communicating in the form of whistles and other acoustic signals such as pulsed signals. Mysticete whales have also been documented to respond to sonar, for example humpback whales have been observed to change song length in response to low-frequency active sonar (Miller et al. 2000, Fristrup et al. 2003).

In a previous collaborative effort with the University of St. Andrews and Cornell University, statistical methods were compared to determine the best approach for examining the effects of naval MFA sonar on delphinid vocal behavior (Oswald et al. 2015). Passive acoustic data analyzed in that study were collected during two deployments of Marine Autonomous Recording Units (MARUs) off the coast of Jacksonville, Florida, and one deployment of MARUs in Onslow Bay, North Carolina. To identify potential changes in cetacean acoustic behaviors in association with MFA sonar, the occurrence of sonar pings and cetacean sounds during sonar exercises and 24 hours after the end of sonar exercises were compared to observations from a 24-hour control period prior to the commencement of sonar exercises. Results of that analysis indicated there was low statistical power by which to detect response effects primarily due to the limited number of independent sonar events available in the Jacksonville and Onslow Bay datasets.

The current work was conducted to: a) add more data with sonar and cetacean encounters, b) add additional species in order to examine species-specific differences in responses, and c) further develop and refine robust statistical methods previously developed in a related earlier study to evaluate the effects of MFA sonar on the acoustic occurrence of cetaceans.

This page intentionally left blank.

2. Statement of Navy Relevance

Changes in vocal behavior in response to Navy active sonar signals have been observed for several species of marine mammals (e.g., Rendell and Gordon 1999, Miller et al. 2000, Clark and Altman 2006, Di Iorio and Clark 2010, McCarthy et al. 2011, Tyack et al. 2011, Melcon et al. 2012, DeRuiter et al. 2013a, DeRuiter et al. 2013b, Stimpert et al. 2014, Martin et al. 2015, Isojunno et al. 2016). Oswald et al. (2015) examined marine mammal acoustic behavior in relation to the occurrence of MFA sonar in an effort to develop statistical methods to evaluate the relationships between the two. That study produced promising results; however, it was limited by the small sample sizes of sonar and marine mammal acoustic encounters. For the current effort, our goal was to increase the sample size to provide greater statistical power. The results of this work will provide a better understanding of whether and how Navy MFA sonar affects the detection of sounds produced by several marine mammal species. This, in turn, could provide important information for the conservation and management of these federally protected species.

This page intentionally left blank.

3. Methods

3.1 Data Collection

For this study, Duke University provided recordings from four High Frequency Acoustic Recording Packages (HARPs). HARPS were deployed to record before, during and after U.S. Navy sonar events between 2010 and 2012 (**Table 1**). One HARP was deployed off the coast of Cape Hatteras, North Carolina (HAT01A), one off the coast of Jacksonville, Florida (JAX05A), and two in Onslow Bay, North Carolina (USWTR05A and USWTR06E). The HARPs were deployed at varying depths ranging from 171 meters (m) to 952 m and had a sample rate of 200 kilohertz (kHz). The Cape Hatteras HARP was not duty-cycled, but the Jacksonville HARP had a 33 percent duty cycle (5 minutes on and 10 minutes off), and both Onslow Bay HARPs had 50 percent duty cycles (5 minutes on and 5 minutes off).

Table 1. HARPs and sonar events used for analysis. Total duration of sonar period represents the total period analyzed from the start of the first sonar event to the end of the last sonar event.

Site Number	Site ID	Deployment Start	Deployment End	Sample Rate (kHz)	Depth (m)	Latitude (°N)	Longitude (°W)	Sonar Start	Sonar End	Total Duration of Sonar Period (h)
1	Cape Hatteras (HAT01A)	03/15/2012 12:48	03/23/2012 00:12	200	950	35.34054	74.85761	03/17/2012 13:33	03/20/2012 14:30	72:57
2	Jacksonville (JAX05A)	10/06/2010 00:47	10/26/2010 1:35	200	91	30.26819	80.20894	10/07/2010 18:30	10/24/2010 13:49	403:19
3	Onslow Bay 1 (USWTR05A)	02/12/2011 23:20	02/21/2011 1:10	200	171	33.79316	76.51620	02/16/2011 02:22	02/18/2011 03:54	49:32
4	Onslow Bay 2 (USWTR06E)	08/18/2011 00:01	09/19/2011 1:11	200	952	33.77794	75.92641	08/21/2011 21:20	08/23/2011 04:04	30:44

3.2 Data Analysis

In order to prepare the acoustic data for analysis, all HARP data files (.wav format) were processed into Long Term Spectral Average (LTSAs) images using the MATLAB-based program Triton (Wiggins 2007). LTSAs are created by calculating frequency spectra for all the data and each of the frequency bands. For the analysis of minke whale sounds, data were decimated by a factor of 20, for all other cetacean and sonar logging the original 200 kHz files were used for initial logging. The LTSAs were created using a five-second time average with 100 Hz frequency resolution for high frequency sounds and 10 Hz resolution for minke whales. These LTSAs in combination with continuous spectrograms were used for logging and sub-logging of sonar events and cetacean acoustic encounters.

3.2.1 Sonar

3.2.1.1 LOGGING METHODS

Using the LTSAs for each HARP, trained data analysts examined LSTA images and spectrograms using Triton software (Wiggins 2007) to identify and annotate (i.e., log) sonar events in the recordings. A sonar "event" was defined as a continuous period of time containing sonar signals with no more than a 30-minute interval between sonar signals. When more than 30 minutes occurred between sonar signals, a new event was started. The start and end time of each sonar event was logged for each site, and these times were used to review and characterize marine mammal and MFA sonar sounds.

3.2.1.2 SONAR SIGNAL ANALYSIS

To increase the efficiency of the analysis, after sonar start and end times were initially logged from the 200 kHz files, acoustic data containing sonar events was down-sampled to 24 kHz. Spectrograms of downsampled .wav files were then examined to determine the different types of sonar present in the recordings. Based on this examination, spectrogram correlation templates were created for four types of sonar signals (e.g. in frequency bands: 1 to 3 kilohertz [kHz], 3 to 5 kHz, 5 to 7 kHz, and 7 to 10 kHz). A spectrogram correlation algorithm was then used to automatically detect sonar pings in the dataset. These sonar events were reviewed manually to ensure there were no missed detections and that any false detections were removed. All true positive sonar detections were then clipped and processed using the MATLAB program SonarFinder (Bio-Waves, Inc. 2013), to make sonar signal measurements. SonarFinder automatically extracted the following features for each sonar event: mean Power Spectral Density (PSD), minimum, maximum and peak frequencies, and maximum received sound pressure level (SPL_{max}). SonarFinder also measured start time and signal to noise ratio (SNR) for each individual ping. Detected sonar pings were categorized by frequency (Type 1: $\leq$4 kHz, Type 2: 4–7 kHz, Type 3: $\geq$7 kHz) and duration (Short: $\leq$1.5 seconds; Medium: 1.5–4.0 seconds; Long: $\geq$4.0 seconds).

3.2.2 Cetacean Encounter Logging Methods

3.2.2.1 TRITON LOGGING

Trained analysts scrolled through the LTSA for each HARP in Triton using a 30-minute page size to identify acoustic encounters for minke whales, sperm whales and delphinids. When signals were identified in the LTSA, analysts examined the corresponding spectrogram for the .wav file in a separate window and were able to zoom in or out in both the frequency and duration ranges to examine signals in more detail. A cetacean encounter was defined as a continuous period of time containing species- or group-specific sounds with no more than a 30-minute interval between the occurrence of sounds produced by the same species. When more than 30 minutes occurred between sounds, a new encounter was logged. Sounds for each species or species group were logged separately. Analysis focused on the 24-hour period before the first sonar event, during the sonar period irrespective of duration, and the 24-hour period after the end of the last sonar event. For each species, analysts also noted the sound types present in each event (**Table 2**).

Table 2. Sound types logged for each species or species group.

Species/Species Group	Call Type
Minke whales	Pulse Train
Sperm whales	Click Slow click Creak Coda
Delphinids	Whistle Click Buzz

For sperm whales and delphinids, the original 200-kHz LTSAs were used for logging analysis. However, in order to log low-frequency minke whale sounds, the data were down-sampled to 10 kHz using Adobe Audition CS5.5. Down-sampled files were then used to create low-frequency LTSAs and spectrograms.

Once cetacean encounters had been logged, analysts compared each encounter against the sonar event log and noted whether the encounter occurred with the 24-h period before a sonar event, during a sonar event, or within the 24-h period after a sonar event. Encounters were labeled as between when they occurred between sonar pings within an exercise.

3.2.2.2 BEAKED WHALE ENCOUNTER LOGGING

Logging all beaked whale encounters in Triton would be time-consuming and inaccurate, as it can be difficult to reliably distinguish between beaked whale and delphinid clicks using an LTSA or spectrogram. As such, additional click characteristics including peak frequency and Wigner-Ville plots available in PAMGuard ViewerMode (Gillespie et al. 2008) software were used to identify and log potential beaked whale sounds in the dataset (see **Appendix A** for details). All .wav files from all of the HARPs were initially processed using PAMGuard with customized click classifiers that were parameterized for each of four species groups (dolphins, blackfish, sperm whale, beaked whale). This post-processing resulted in “binary files” (a proprietary PAMGuard file type) and a populated Microsoft Access database that could be used to further analyze data in PAMGuard’s ViewerMode. PAMGuard ViewerMode was then used to identify and log beaked whale encounters. As with the other species, encounters were defined as periods of time during which beaked whale clicks were continuously present without an interval of more than 30 minutes between clicks. If there was more than a 30-minute interval between clicks, they were logged as a new and separate encounter. Using ViewerMode, trained analysts identified the clicks to species, whenever possible. Once all beaked whale click encounters were logged, analysts compared each encounter against the sonar event log and noted whether it occurred in the 24-hour period before sonar, during sonar, between sonar, or during the 24-hour period after a sonar event.

3.2.3 Sub-logging Methods

3.2.3.1 MINKE WHALES

In order to provide additional details for the analysis of minke whale acoustic encounters, each encounter was examined at a finer scale ("sub-logged") to identify the start time, end time, and type for individual pulse train within the encounter (Mellinger et al. 2000, Risch et al 2013). Pulse train type was defined as either speed up (inter-pulse interval [IPI] decreases throughout the duration of the pulse train), slow down (IPI increases throughout the duration of the pulse train), regular (IPI stays constant throughout the duration of the pulse train), or unidentified. The unidentified category was used when the signal-to-noise ratio (SNR) was too low to reliably determine the pulse train type, or alternatively when the pulse train was cut off due to duty-cycling (Onslow Bay dataset only). In addition, the relative SNR of the pulse train was defined subjectively by the analyst as either "high," "medium," or "low", and the frequency range of each pulse train was noted. Once all pulse trains had been sub-logged for all HARP data, the times were compared against the sonar event logs to determine whether each pulse train occurred within the 24-hour period before sonar, during sonar, between sonar, or within the 24-hour period after a sonar event.

3.2.3.2 DELPHINIDS

In order to provide greater detail for the analysis, each delphinid acoustic encounter was sub-logged into 1-minute sub-encounters to identify which sound types were present during each minute. To do this, analysts examined spectrograms in Triton using a window length between 10 and 30 seconds and noted which sound types occurred in each minute (Whistles [W], Clicks [EC], Buzzes [BZ]; **Table 2**). The frequency range of any whistles present was also noted (>10 kHz, <10 kHz, or both).

3.2.4 Click and Whistle Counts

3.2.4.1 BEAKED WHALES

PAMGuard ViewerMode was used to obtain click counts for beaked whale encounters (see **Appendix A**). In this mode, the user manually selected click trains and marked them by drawing a box around the click trains to signify an encounter. All of the marked clicks in the encounter were subsequently sent to the Real-time Odontocete Call Classification Algorithm (ROCCA; Oswald et al. 2013) to be measured and a suite of measurements for each click was saved in a database (**Appendix A**). This allowed analysts to estimate click counts and automatically make click measurements for later evaluation and confirmation of species identification.

After exporting data from PAMGuard, the click information was run through a custom MATLAB-based script known as "Bin-It Counter" in order to obtain click counts per minute for each HARP.

3.2.4.2 DELPHINIDS AND SPERM WHALES

In order to obtain estimated click counts for sperm whales and delphinids and whistle counts for delphinids, encounters for each HARP were processed using PAMGuard's click detection and classification modules, the whistle and moan detector module, and the Auto ROCCA module, which automatically identifies and exports measurements and counts for clicks and whistles (see **Appendix A** for more detail). Once exported, the click and whistle information was processed using Bin-It Counter to obtain click and whistle counts per minute for each HARP.

3.2.5 Delphinid Whistle Measurements

A subsample of whistles was randomly selected from each delphinid encounter and whistle features were measured using the ROCCA module in PAMGuard. Only encounters that contained at least 10 whistles with moderate to good SNR (i.e., at least 3 dB) were included in the analyses. For each encounter, analysts used the acoustic software Raven (Charif et al. 2004) to randomly select up to 30 whistles for encounters less than 2 hours in duration and up to 10 whistles per hour for encounters longer than 2 hours. To extract time-frequency contours from selected whistles, the analyst traced contours on ROCCA's spectrographic display using a computer touch-pad. ROCCA automatically measured 50 variables from each extracted contour, including duration, frequencies (e.g., minimum, maximum, beginning, ending, and at various points along the whistle), slopes, and variables describing shape of the whistles (e.g., number of inflection points and steps; see **Appendix B** for a complete list and description of variables measured). A subset of these features (minimum, maximum, mean frequency, standard deviation of the frequency, duration, mean slope, mean positive slope, mean negative slope, percent of whistle with positive, negative or zero slope) was sent to the University of St. Andrews to evaluate potential changes in whistle structure in the presence of MFA sonar. The extracted feature vectors were also analyzed using a random-forest classifier within ROCCA for species identification.

The random-forest model used in this analysis was a two-stage model trained using whistles recorded from single-species schools in the northwestern Atlantic Ocean. A two-stage model was used because it resulted in higher correct classification scores than a one-stage model that classified whistles directly to species (Oswald 2013). Five species were included in the model: bottlenose dolphins (*Tursiops truncatus)*, short-beaked common dolphins (*Delphinus delphis*), striped dolphins (*Stenella coeruleoalba*), Atlantic spotted dolphins (*S. frontalis*), and short-finned pilot whales (*Globicephala macrorhynchus*). The two-stage model first classified whistles to one of three categories: small dolphins (including common and striped dolphins), large dolphins (including bottlenose and spotted dolphins), or pilot whales. Whistles within each category were then classified to species in stage two (**Figure 1**). When the model was evaluated using a test dataset of visually validated recordings, 78 percent of whistles (n=1,034) and 86 percent of encounters (n=131) were correctly classified (Oswald 2013).

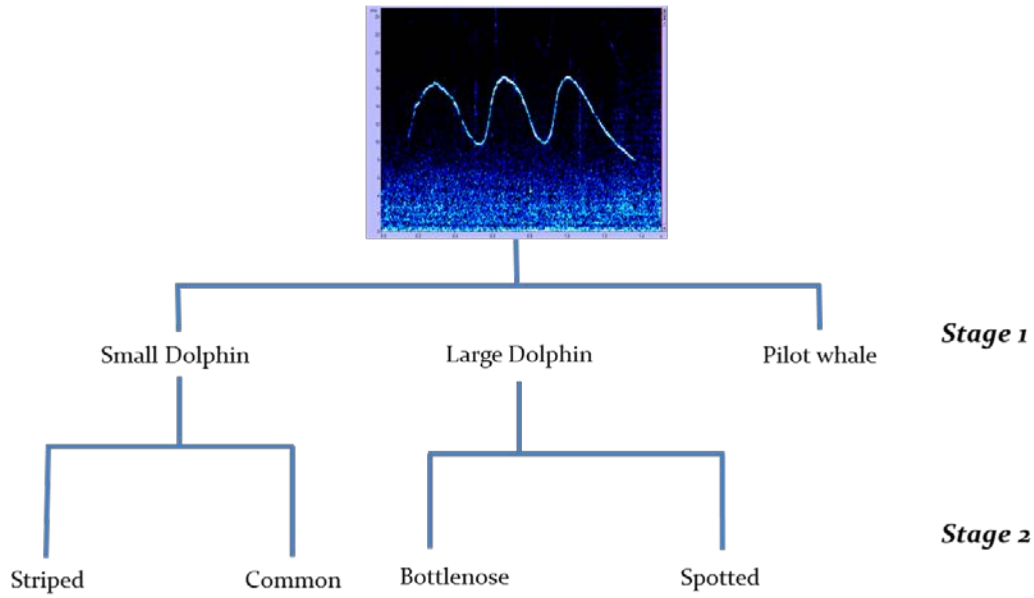

Figure 1. Schematic diagram of two-stage random forest model. In stage one whistles are classified as small dolphins (e.g., common and striped dolphins), large dolphins (e.g., bottlenose dolphins and spotted dolphins), or pilot whales. Whistles are then classified to species in stage two.

3.3 Statistical Modeling

3.3.1 Introduction

The primary dataset of cetacean acoustic detections processed from HARP recordings by Bio-Waves, Inc. consisted of start and end times of encounters and sub-encounters. The encounter and sub-encounter data were used to create an index of presence–absence of sound detections using 1-minute segments. This dataset was assembled for each taxon under consideration (i.e., minke, sperm, and beaked whales and dolphins). In the case of dolphins, additional data were available in the form of signal type (e.g., whistle, click, or buzz) and the number of clicks and whistles in an encounter. In addition, a sample of whistles was further characterized using the suite of characteristics described above, which might change in response to the presence of sonar. For minke whales, pulse train duration measurements were also available.

Time periods were categorized relative to the presence of sonar activity. Following Oswald et al. (2015), we defined a sonar exercise as the time period including all consecutively detected sonar pings at a given site with no gap in pings longer than 48 hours. When a gap was longer than 48 hours, a new exercise was started and the subsequent sonar pings were attributed to this new exercise. We used the 24-hour period before the commencement of each sonar exercise as a control period. Every cetacean acoustic sub-encounter or 1-minute segment occurring during these 24 hours was labelled as “before.” Using 24 hours as a control period represented a compromise between trying to capture a potentially ongoing effect after a previous sonar exercise and to avoid introducing additional variability by extending too far beyond the time of the sonar exercise while keeping a balance between the periods included before and after. Cetacean acoustic sub-encounters or 1-minute segments were labelled as “during” when sonar pings were detected during the sub-encounter or segment. Acoustic sub-encounters were labeled as “between” when they occurred between sonar pings within an exercise. Acoustic sub-encounters that occurred within 24 hours after a sonar exercise were labelled as “after.” Some of the HARP delphinid whistle components spanned two periods in relation to sonar, so they were classified as “between/during” or “during/after.” Subsequently both of these classes were converted to “during” for analysis.

Five separate statistical analyses of the data were undertaken to address different research questions. Response variables were defined differently for each analysis, but potential explanatory variables were roughly the same. The questions and analyses are as follows:

1) Does the probability of detecting acoustic signals from cetaceans change in the presence of sonar? Here a regression analysis of acoustic occurrence per 1-minute segment (separate models for sperm whales, beaked whales, minke whales, and delphinids where the presence/absence of acoustic occurrence was the response variable) was undertaken.

2) Do the transitions of states (producing sounds to quiet or vice versa) change in relation to sonar? Here a hidden Markov model analysis of these same acoustic occurrence data (as for 1) was undertaken. We note that state quiet does not distinguish between animals being present and not producing sounds and animals being absent.

3) Does the detected duration of minke whale pulse trains change in relation to sonar? Here a regression analysis of minke whale pulse train durations was undertaken.

4) Does the probability of detecting a signal type within a delphinid acoustic encounter change in relation to sonar? Here a regression analysis of presence/absence of a signal type (whistle, click or buzz) within an acoustic encounter of delphinids was undertaken.

5) Do the characteristics of detected delphinid whistles change in relation to sonar? Here a regression analysis of whistle characteristics (combined into a single metric using Mahalanobis distances, DeRuiter et al. 2013a) was undertaken.

All modelling was undertaken using R (version 3.2.4, R Developmental Core Team 2016). For delphinids, sperm whales, and minke whales, equivalent data logs existed from MARUs (see above and Oswald et al. 2015 and Charif et al. 2015). Combining the old MARU and new HARP data allowed increases in the sample size of transitions from a “before” to a “during/between,” and thence to an “after” period (**Table 3**), allowing greater power to detect an effect.

All of the above analyses used similar predictors, but not all datasets had the full set of predictors as in the MARU-only analyses (Oswald et al. 2015, Charif et al. 2015). **Table 4** provides a breakdown of the core predictors of interest.

Different sites had sensors deployed at different times of year (**Figures 2** and **3**). As a result, variables *Site* of a detector, day-of-year of detection and type of detector (*SurveyType*) are to an extent confounded (i.e., variance inflation factors $\approx$ 11 in the case of regressions containing *Site* and day-of-year, for example), so the individual effects of these three variables cannot be distinguished. For this reason, only the *Site* and *SurveyType* variables were considered out of these three variables and never in the same model at the same time.

For some of the data, additional predictor variables were available that described MFA sonar features (**Table 5**).

Table 3. Summary of data sets used in this study.

Sensor	Location	Site	Year of Operation	Number of sonar Exercises
MARU	Off Jacksonville	JAX2	2009	2
MARU	Off Jacksonville	JAX4	2009	2
MARU	Off Jacksonville	JAX5	2009	2
MARU	Off Jacksonville	JAX6	2009	2
MARU	Off Jacksonville	JAX7	2009	1
MARU	Off Jacksonville	JAX9	2009	2
MARU	Onslow Bay	OB152	2008	2
MARU	Onslow Bay	OB154	2008	4
MARU	Onslow Bay	OB159	2008	3
MARU	Onslow Bay	OB161	2008	4
HARP	Off Cape Hatteras	HAT01A	2012	1
HARP	Off Jacksonville	JAX05A	2010	1
HARP	Onslow Bay	USWTR05A	2011	1
HARP	Onslow Bay	USWTR06E	2011	1

Note: Sonar exercise defined as in **Section 3.3**.

Table 4. Potential core predictors used in the statistical analysis.

Name	Description	Notes
Site	Factor variable pertaining to the individual recording devices	Should not be used with day of year as some sites were surveyed for a limited period only. Cannot be used with *Survey Type*.
Timeofday	Continuous variable describing time of day from midnight.	Should not be used with *Daynight.* Considered as a spline with 3 or 4 degrees of freedom.
Sonar	Factor variable indicating whether time period was 24 hours before, between, during, or 24 hours after a sonar event.	
Daynight	Factor variable indicating whether it is day or night. Relationship of time period to local sunrise/sunset.	Should not be used with *Timeofday*
SurveyType	A factor variable indicating whether MARUs or HARPS were used in collecting data.	Should not be used with *Site*

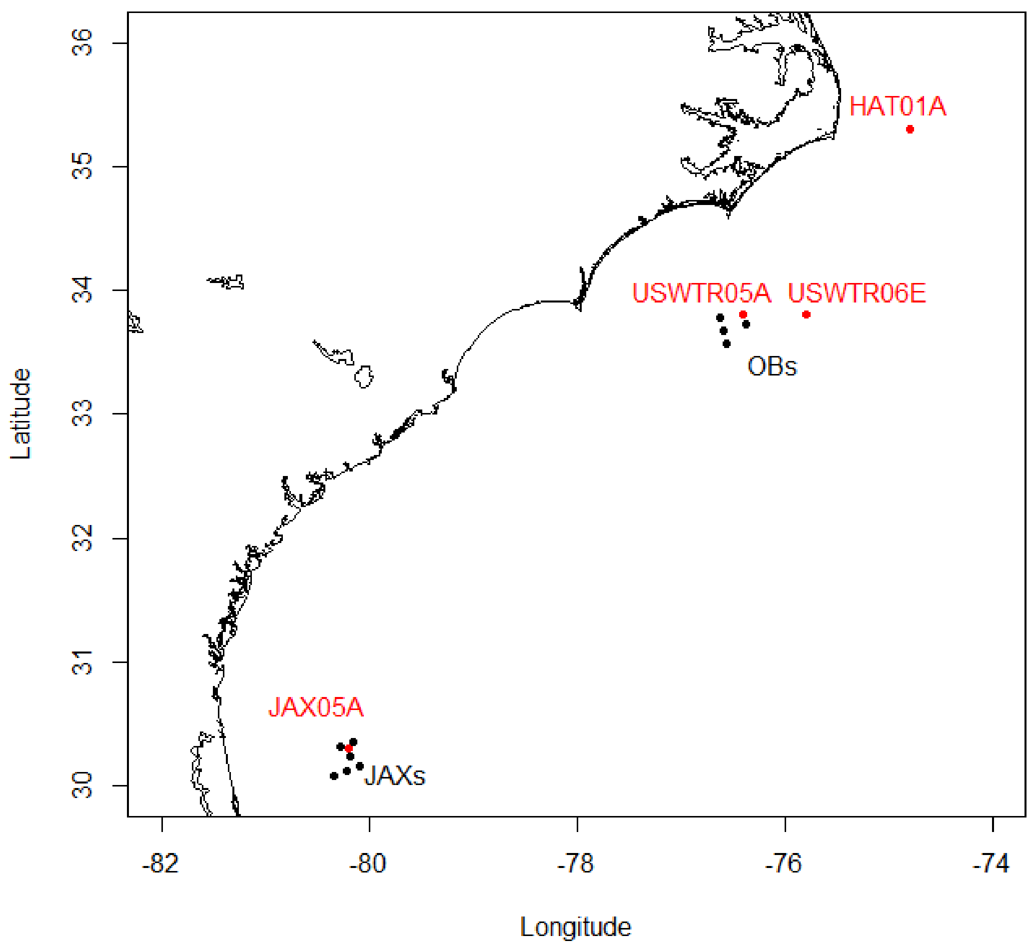

Figure 2. Locations of MARUs (black, see Oswald et al. 2015, Charif et al. 2015) and HARPs (red).

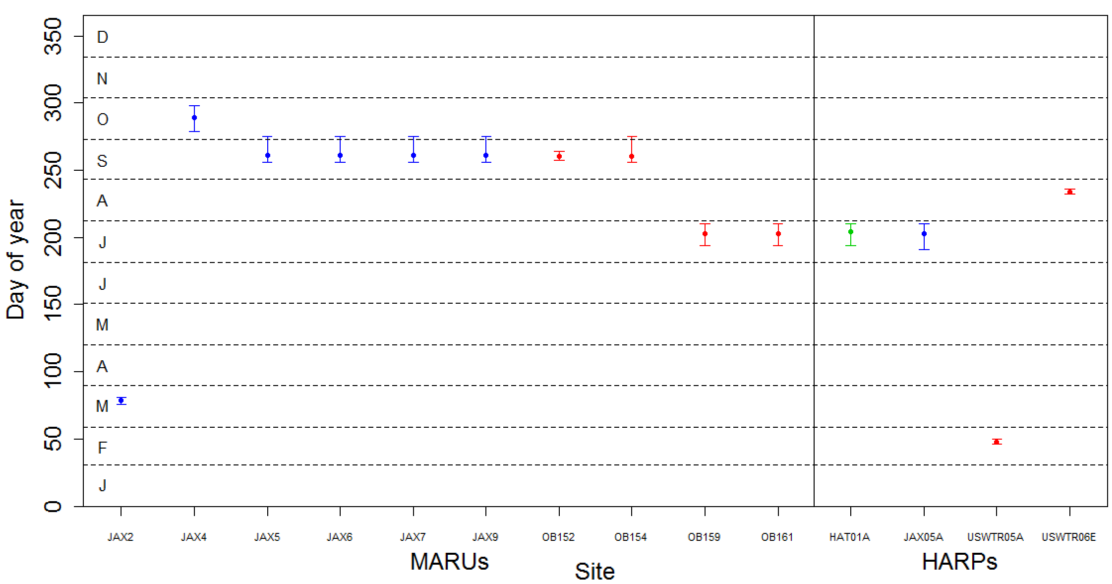

Figure 3. Sampling dates and locations of MARU and HARP recorders. Cape Hatteras HARP site green, the Onslow Bay sites red, and the Jacksonville sites blue. Dashed lines show the end of each month (given by capitals).

Table 5. Potential sonar descriptor predictors used in the regression analyses.

Name	Description	Notes
Sonaron	Factor variable that describes whether sonar activity is off or on.	Only considered in interaction terms with sonar descriptors. Never used as a main effect. *Sonar* used instead.
PFHz	Peak frequency	Considered only as an interaction with *Sonaron* either as a linear term or spline with 3 degrees of freedom.
SPL	Sound pressure level	Considered only as an interaction with *Sonaron* either as a linear term or spline with 3 degrees of freedom.
Type 1 long, *Type 1 med* *Type 1 short* *Type 2 long* *Type 2 med* *Type 2 short* *Type 3 long* *Type 3 med* *Type 3 short*	Presence/absence of these features of the sonar.	*Type 2 long* in the HARP data were negatively correlated with *Type 3 long* and *Type 3 med* signals, so these combinations were not considered together in the PSTGAE models (see below).

3.3.2 Regression modelling of presence or absence of the acoustic signal per 1-minute segment

For the presence–absence models, we assembled data records for each HARP/MARU site, which consisted of 1-minute segments encompassing the 24-hour period before a sonar exercise, the time during a sonar exercise, and the 24 hours after a sonar exercise (see Section 3.3.1 for how we define a sonar exercise in this report). For each 1-minute segment, the detection or non-detection of a cetacean signal (appropriate to the taxon under consideration) was recorded as a binary variable. In addition, each segment was labelled with the sonar condition (see above) according to when it occurred in relation to the sonar exercise at the same site. Effort segments that corresponded with periods of the sonar exercise were either labeled as during if a sonar ping had been recorded during the 1-minute segment or between if not. For this analysis, we assumed a binomial error structure and used a logit-link function.

Because the data were collected sequentially in time, there was the potential for unexplained residual autocorrelation in the fitted models. The presence of such autocorrelation means that assuming independence of the individual data points would underestimate the overall uncertainty leading to an increased risk of committing a type 1 error (rejecting the null hypothesis of no effect, when there actually is no effect). Therefore, it is practical to use a method that takes potential serial correlation of errors into account such as generalized estimating equations (GEEs; Hardin and Hilbe 2003, R library *geepack*). GEEs are an extension of Generalized Linear Models (GLMs) and allow the fitting of link functions to model data with residuals that are not normally distributed. Furthermore, splines can be used with GEEs allowing non-linear response covariates to be considered. Non-independence over a particular range can be addressed by grouping the data into blocks within which data are assumed to be correlated and where this correlation can be estimated allowing a correction for non-independence.

Model selection proceeded as follows. Initially, a logistic GLM was fitted to the non-duty cycled data with all the available non-mutually exclusive variables. The Pearson residuals were then subject to an autocorrelation analysis. The first lag at which no significant autocorrelation occurred was used to identify blocks in the GEE analysis. Forward model selection was then undertaken with a logistic GEE with variables included based on the lowest p-value for that round, with an inclusion criterion of p<0.01. We used p-values based on adjusted (Type II) sums of squares so they were unaffected by the order of the variables in the model. The terms in the final model were checked for variance inflation (using R library *car* on the equivalent GLM using a less-than-4 inclusion criterion) and only finally included if graphical inspection of the partial plots (using the R library *MRSea,* Scott-Haywood et al. 2014) suggested there really was a significant effect (i.e., the 99 percent confidence interval of the smooths could not contain a horizontal line that did not go over the upper or lower boundaries).

The predictor variables considered in the modelling in addition to the sonar period given above are provided in **Tables 4 and 5.** The covariate *Sonar* was confounded with other covariates pertaining to sonar, which prevented us from investigating the effect of either *Sonar* with the sonar relevant covariates (e.g. Type 1 long etc.). As we have a strong interest in both the effect of *Sonar* and the other covariates, we decided to add an extra step in model selection. In the initial model, we included *Sonar* and covariates not pertaining to sonar. After selecting the best model with these covariates, *Sonar* was removed from the model (if present) and model selection recommenced with the currently chosen variables and, in addition, considering all covariates in **Table 5** to determine any specific component of the sonar that was affecting the detection of the cetacean sound. Thus, there were two final models for each analysis, one where *Sonar* was considered and one where the individual components of the signal were considered. For some taxa, the additional sonar descriptors were not available so this second stage of modelling was not undertaken.

The data were regularly spaced in time except for the gaps caused by duty cycling. This had implications for defining the blocks for an appropriate covariance structure. We accommodated this by incorporating the temporal order of the data into the correlation structure using a *wave* variable (see Højsgaard et al. 2016, Højsgaard et al. 2005).

In **Appendix C**, we tested if model assumptions were met, using reference plots of residuals (the difference between the observed data and the fitted value from the model) in order and plots of the fitted values against the averages of binned (n=20) observed values. Binning of observed values (1's and 0's for presence and absence, respectively) was done as we used a binomial logistic model, which returns model predictions as probabilities as opposed to a binary variable. We evaluated the predictive success of the model by comparing observed presence and absence versus predicted presence and absence. Predicted presence (absence) values were obtained using the criterion that the predicted value had to be larger (smaller) than the mean of the model fitted values, as this is an indicator of expected presence in models of binary data with low predicted probability of presence.

The models were illustrated by plotting various predictions made from different relevant variables. Confidence estimation was by a non-parametric bootstrap (Davison & Hinkley 1997).

3.3.3 Hidden Markov modelling (HMM) of presence or absence of the acoustic signal per 1-minute segment of recording effort

3.3.3.1 INTRODUCTION

In all of the acoustic occurrence time series data (**Figures 4 through 7**), time periods of different types were observed, in particular we can distinguish two types:

i) periods of high acoustic activity followed by periods of low acoustic activity and vice versa, or

ii) periods with no acoustic activity interspersed with shorter periods with low levels of acoustic activity.

These patterns can be modelled using stochastic mixtures, with one mixture component corresponding to a certain type of behavior (e.g., high acoustic activity) and the other mixture component corresponding to a different type of behavior (e.g., low acoustic activity), with a random mechanism selecting which of the two components is active at a given time point. Independent mixture models can capture this pattern, but do not account for the serial correlation observed in the time series (as shown, for example, by the sample autocorrelation function for the beaked whales in HAT01A plotted in **Figure 8**). This justifies the use of HMMs, which are also mixture models, but ones in which the different component distributions, associated with the different states, are selected by a Markov chain, thereby inducing persistence in the sequence of observed behaviors (see for example Zucchini and MacDonald 2009).

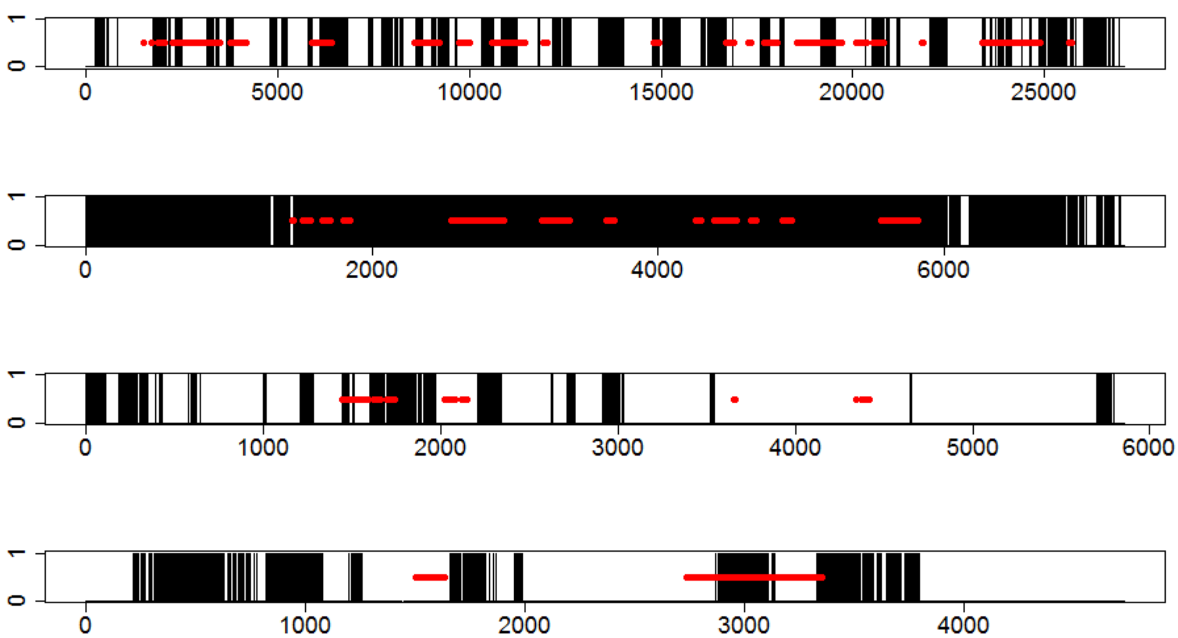

Figure 4. HARP time series of 1-minute segments where black bars indicate presence of delphinid acoustic signals and white indicates absence of acoustic signals. Data include “not available” (NAs, i.e. time points where data are missing). The Y-axis gives the value the binary variable takes for a given observation (0 or 1); the X-axis indexes the observations in the time series. From top to bottom – JAX05A, HAT01A, USWTR05A and USWTR06E. Red dots signify occurrence of sonar events.

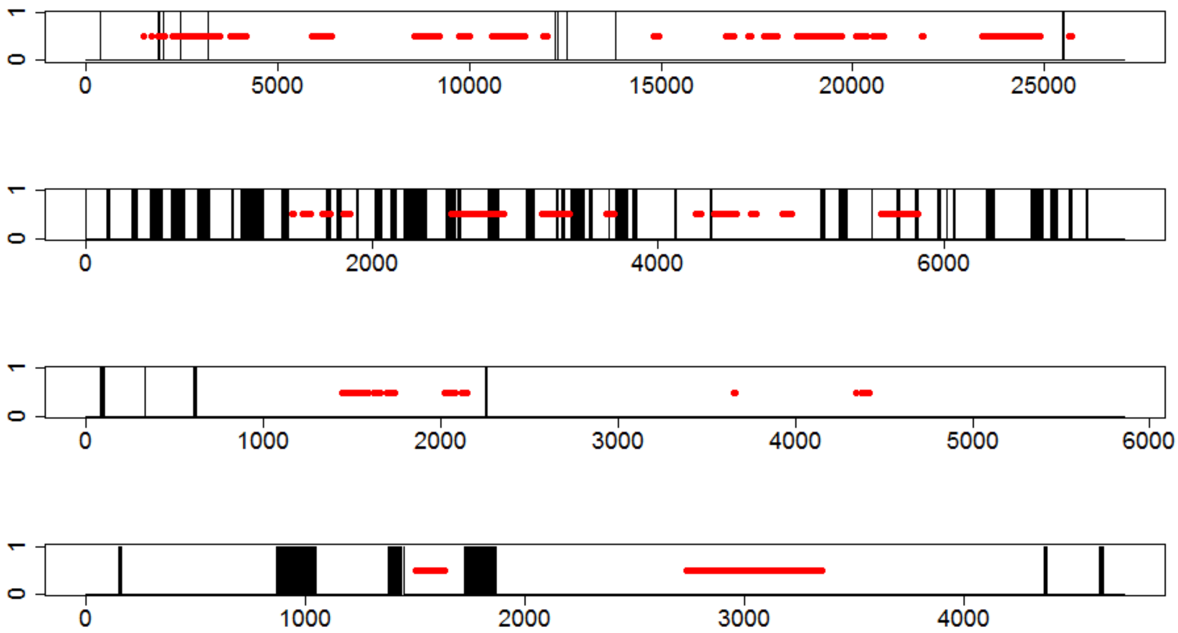

Figure 5. HARP time series of 1-minute segments where black bars indicate occurrence of beaked whale acoustic signals and white indicates absence of acoustic signals. Data include NAs (i.e. time points where data are missing). The Y-axis gives the value the binary variable takes for a given observation (0 or 1); the X-axis indexes the observations in the time series. From top to bottom – JAX05A, HAT01A, USWTR05A and USWTR06E. Red dots signify occurrence of sonar events.

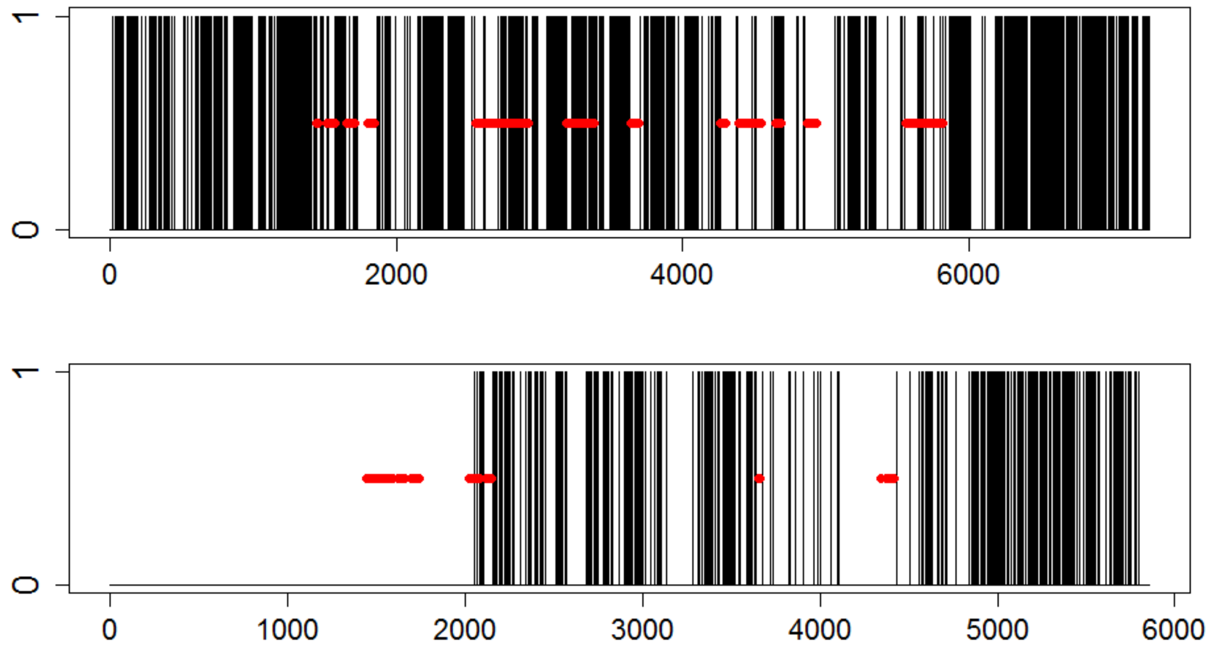

Figure 6. HARP time series of 1-minute segments where black bars show presence of minke whale acoustic signals and white indicates absence of acoustic signals. Data include NAs (i.e. time points where data are missing). The Y-axis gives the value the binary variable takes for a given observation (0 or 1); the X-axis indexes the observations in the time series. From top to bottom – HAT and USWTR05A. Red dots signify occurrence of sonar events.

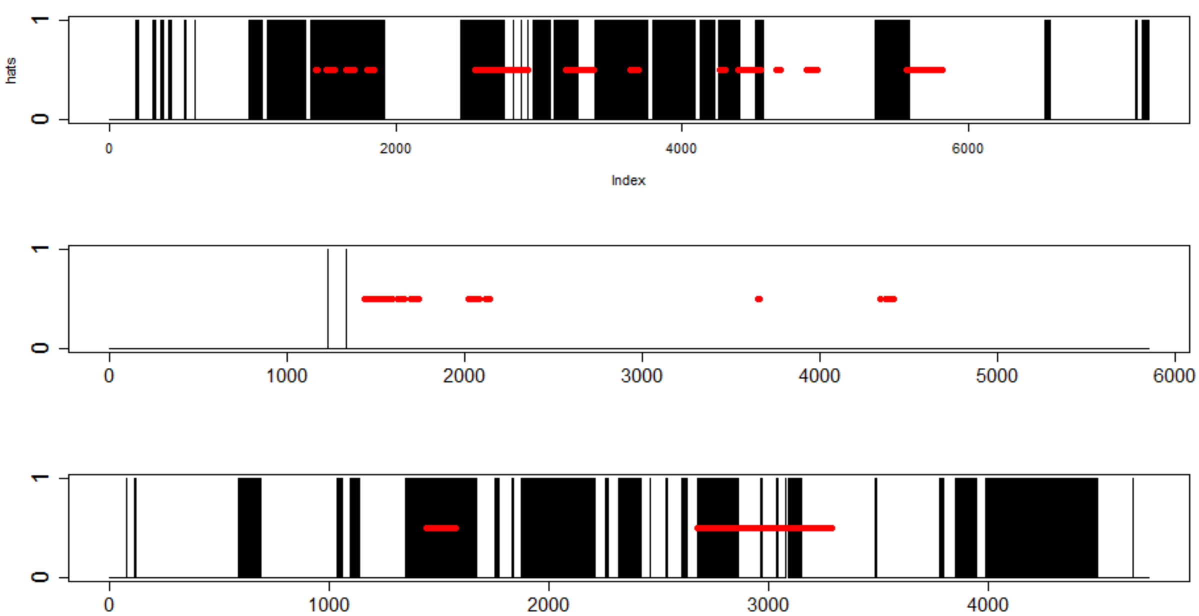

Figure 7: HARP time series of of 1-minute segments where black bars indicate sperm whale acoustic detections and white indicates absence of acoustic signal. Data include NAs (i.e. time points where data are missing). The Y-axis gives the value the binary variable takes for a given observation (0 or 1); the X-axis indexes the observations in the time series. From top to bottom – HAT01A, USWTR05A and USWTR06E. Red dots signify occurrence of sonar events.

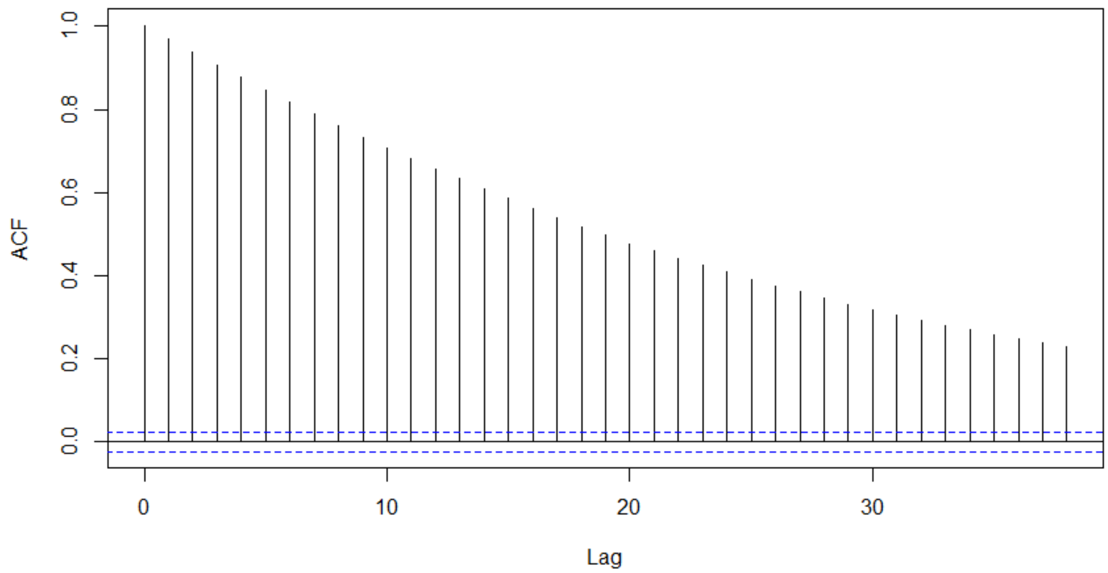

Figure 8. Sample autocorrelation function for beaked whales at the HAT01A site (HARP data).

3.3.3.1. FOCUS ON THE DATA

Here, we emphasize certain aspects of the HARP datasets that have particular importance for the analysis. Data analysis results were assembled at 1-minute time resolution. There were four locations (JAX05A, HAT01A, USWTR05A, and USWTR06E) and four species groups (dolphins; beaked, minke, and sperm whales) yielding a total of 16 potential time series. Four of the 16 were discarded because they had no or very few acoustic detections. There were no minke whale acoustic detections at JAX05A and USWTR06E and no sperm whale acoustic detections at JAX05A, and there were only two sperm whale acoustic detections in the 2929 1-minute samples at USWTR05A. We note that at USWTR05A, there was a long period without any minke whale acoustic detections (**Figure 6**). A reasonable explanation is that there were no minke whales in the area for an extended period until a certain time point. Since we did not explicitly model lack of acoustic detections in the absence of animals, we used a somewhat arbitrary, but reasonable approach of excluding the time periods prior to the first acoustic detection when fitting HMMs on this time series. Duty cycling lead to “not available” (NA) entries in the time series, which had to be dealt with. While this problem is easily overcome for baseline HMMs (see next section), this is not the case for HMMs with covariates. To cope with the NAs, we executed a simple imputation procedure: for a period with NA, assign the last observed value of the covariate to the missing entry. We note that, in the majority of the cases, the value of the covariate before the missing values and after the missing values is the same (e.g., 2,703 out of 2,864 cases for the dolphin data sets, i.e., 94.4 percent). Consequently, this imputation procedure does not occur at a major cost.

An HMM is a stochastic time-series model involving two components—an observable time series and an underlying latent state sequence. The latter is an N-state Markov Chain (i.e., a stochastic process that takes values in $\{1, \ldots, N\}$ and satisfies the “memorylessness” property that, given the present value of the process, future values are independent of the past). The observed time series, typically referred to as a “state-dependent process,” is such that its values are assumed to be generated by one of N component distributions, with the underlying Markov chain selecting which one is active at a given time. Usually the N distributions are chosen to belong to the same distribution family with the difference coming from different values of the respective parameters. In this study, the state-dependent process is a binary variable taking the value 1 if an acoustic detection occurred and 0 otherwise. In the following, we denote the observable state-dependent process by X_t, and the underlying latent N-state Markov Chain by S_t, and focus on the two-state case (i.e., N=2). We assume a basic dependence structure where, given the current state S_t, the variable X_t is conditionally independent from previous and future independent states, and where the Markov chain is of first order.

3.3.3.2. THE BASELINE MODEL

For the baseline model, the state transition probabilities, $\gamma_{ij} := \Pr(S_t = j | S_{t-1} = i)$, were assumed to be constant over time (i.e., the Markov chain was homogeneous). We summarized these probabilities in a 2×2 transition probability matrix (TPM) that is a matrix indicating the probability of changing (transitioning) from state to state, given as:

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}$$

The initial state probabilities were summarized in a row vector δ, where $\delta_i = \Pr(S_1 = i)$. When the Markov chain is homogeneous, we assumed that the process is in equilibrium when we start observing it, such that the initial distribution is the stationary distribution.

For the state-dependent distribution of X_t we assumed the Bernoulli distribution, with probability of detecting sound varying across the states. In other words, we assumed that:

$$X_t | S_t = i \sim \text{Bernoulli}(\pi_i), \quad i = 1,2$$

so that

$$\Pr(\text{sound recorded} | S_t = i) = 1 - \Pr(\text{no sound recorded} | S_t = i) = \pi_i,$$

with $\pi_1 \neq \pi_2$ in general.

The likelihood can be expressed using the following matrix product:

$$\mathcal{L}(\theta | x_1, \ldots, x_n) = \delta P(x_1) \Gamma P(x_2) \times \ldots \times \Gamma P(x_n) 1,$$

where θ denotes the vector with the parameters to estimate; $P(x_t)$ is an 2×2 diagonal matrix with the conditional probability mass functions

$$P(X_t = x_t | S_t = i) = \begin{cases} \pi_i & \text{when acoustic detection is recorded} \\ 1 - \pi_i & \text{when no acoustic detection is recorded} \end{cases}$$

of X_t, given $S_t = i$, $i = 1,2$, on the main diagonal; 1 is a column vector of ones. When data are missing at time t, $P(x_t)$ is simply replaced by the identity matrix in the equation above. The model fitting of the baseline models was conducted via numerical maximization of the log-likelihood as described in Zucchini and MacDonald (2009).

The baseline model was fitted separately to the 12 time series (4 for dolphins, 4 for beaked whales, 2 for minke whales, and 2 for sperm whales). A drawback of this approach is that there are too few transitions between states. However, pooling the time-series data does not seem reasonable. **Figures 4–7** show that it is unrealistic to assume the model parameters are identical either across species groups or across sites. The other alternative, the consideration of random-effect models, seems infeasible given the small number of component series and the computational complexity of HMMs incorporating random effects (see Schiele-Diecks et al. 2012).

3.3.3.3. HMMS WITH SONAR-RELATED COVARIATES

When considering covariates in an HMM, it is reasonable to assume that the external variables influence the transition of states rather than the state-dependent distributions, which remain fixed for a given state. This allowed us to draw conclusions about the effect of, for example, covariate *Sonaron* on the transitions between the different states. For any given series of covariates, $z_1, z_2, z_3, \ldots$, the model given above was modified by letting the transition probabilities depend on the covariate values as follows:

$$\gamma_{12}^{t} = \Pr(S_t = 2 | S_{t-1} = 1) = \text{logit}^{-1}(\beta_{1,0} + \beta_{1,1} z_t)$$
$$\gamma_{21}^{t} = \Pr(S_t = 1 | S_{t-1} = 2) = \text{logit}^{-1}(\beta_{2,0} + \beta_{2,1} z_t)$$

Note that the superscript t indicates that the transition probabilities were no longer constant but vary over time. The elements on the main diagonal of the TPM (γ_{11}^{t} and γ_{22}^{t}) are obtained by subtracting the other entry on the same row from one. The formulation of the state-dependent distribution remained unchanged as it was the construction of the likelihood.

We considered two covariates *Sonar* and *Sonaron* in single-covariate models. The model fitting was conducted using similar procedures as described above.

3.3.4 Regression modelling of presence or absence of dolphin signal types given an acoustic encounter.

For the "presence of signal type given a cetacean acoustic encounter" (PSTGAE) models, we considered delphinid data from the MARUs (defined as sub-events in Oswald et al. 2015) and the HARPs (defined as sub-encounters in **Section 3.2.3.2** above). MARU sub-events could be of different lengths whereas the HARP sub-encounters were of 1-minute length. Because of this, we opted to model the HARP data alone. The presence or absence of each of the three delphinid acoustic signal types (whistles, clicks, and buzzes) was modelled separately as a binomial logistic GEE regression, as before, with the variables in **Table 3–4** as before, but with the addition of variables indicating the presence or absence of other acoustic signal types (i.e., buzzes and clicks in the case of whistles, etc.). Because of the proximity of the signals in time, modelling was undertaken using GEEs. Blocks were identified as before, although they were now irregularly spaced in time.

Model selection and assumption checking was as in **Section 3.3.2**.

3.3.5 Whistle Characteristic Models

The available variables describing the characteristics of the whistles obtained from the ROCCA analysis were: minimum frequency, maximum frequency, standard deviation of the frequency, duration, mean frequency, slope of the whistle, mean of the positive slope, variables describing the whistle contour, mean of the negative slope values in the whistle contour, percent positive slope, and percent zero slope. Other whistle characteristics including mean frequency and percent negative slope were excluded due to collinearity with other characteristics. The observed values were normalized by subtracting the mean value and dividing by the standard deviation for each vector of observed whistle characteristics (using the scale function of the *base* package in R). For each whistle the information from these characteristics was combined into one response variable using Mahalanobis distances (D_{Mi}, cf. DeRuiter et al. 2013a). These were calculated for the *i*th whistle using

$$D_{Mi}(x_i) = \sqrt{(\boldsymbol{x_i} - \boldsymbol{\mu_c})^T \boldsymbol{S_c^{-1}} (\boldsymbol{x_i} - \boldsymbol{\mu_c})}$$

where x_i represents a vector of the observed characteristics of the *i*th whistle, μ_c represents a vector of means for each whistle characteristic obtained from a set of control whistles (see below for the definition), S_c is the covariance matrix of the control whistles, and *T* indicates the transposition (in this case of a vector).

All of the above variables could be affected by the presence of duty cycling because the record of the whistle may not be complete. In addition, the HARP recordings were limited in length.

For defining the control whistles, two strategies were possible. In the first case, all whistles that occurred in the 24 hours before a sonar exercise were included in the control group. Each whistle, including those from before, during, between and after sonar, was compared to the control group. An alternative strategy was to define the control whistles for each individual whistle as the set of *n* whistles that preceded it. The benefit of the second strategy is that it may reveal sharper contrasts in the case of a short-term change. However, the difficulty for this strategy is choosing *n* such that it is biologically meaningful. Here we applied the first strategy, and used Mahalanobis distances as proxies for potential response intensity following the example of DeRuiter et al. (2013a); the distances were then related to the explanatory covariates. The control period was derived from each site separately.

Model selection was conducted as in **Section 3.3.1** however now a Gamma error was assumed using an inverse link function. However, here we used a plot of the Pearson residuals verses the fitted values for diagnostic purposes, as well as the plot of the fitted values against the actual observations. Block size was re-estimated based on the final model selection.

4. Results

4.1 Sonar Summary

A total of 57 sonar events were identified in the entire dataset, with an overall cumulative sonar event duration of approximately 157 hours (**Table 6**). The Jacksonville dataset contained 36 events with a total duration of approximately 116 hours. The Onslow Bay dataset contained eight events with a total duration of 19 hours. The Hatteras dataset contained 13 events with a total duration of approximately 22 hours (**Table 6**).

Table 6. Summary of sonar events per site. The total sonar duration per site represents the cumulative duration of the actual sonar events, excluding gaps between events.

Site ID	Total Number of Sonar Events Per Site	Total Sonar Duration Per Site (hh:mm:ss)	Total Number of Pings Per Site	Mean SPL Per Site (dB re: 1 μPa)	Median SPL Per Site (dB re: 1 μPa)
Cape Hatteras (HAT01A)	13	21:39:12	1391	71.7	70.4
Jacksonville (JAX05A)	36	115:58:06	4740	81.3	78.8
Onslow Bay 1 (USWTR05A)	6	6:59:23	127	86.2	83.2
Onslow Bay 2 (USWTR06E)	2	12:23:45	759	78.7	75.7

4.2 Cetacean Encounter Summary

4.2.1 Minke whales

A total of 50 minke whale encounters was logged in the dataset with an overall duration of approximately 124 hours (**Table 7**). Minke whale pulse trains were not detected in the Jacksonville and Onslow Bay 2 datasets.

Table 7: Summary of minke whale encounters per site.

Site ID	Total Number of Minke Whale Encounters Per Site	Total Duration of Minke Whale Encounters Per Site (hh:mm:ss)
Cape Hatteras (HAT01A)	32	84:07:51
Onslow Bay 1 (USWTR05A)	18	40:08:31

A total of 1,213 minke whale pulse trains were identified in the sub-logging analysis (941 at Cape Hatteras and 272 at Onslow Bay 1). Of the pulse trains that could be identified to type, the majority were the slow-down type (**Figures 9** and **10**). Speed-up type pulse trains were only identified at Cape Hatteras.

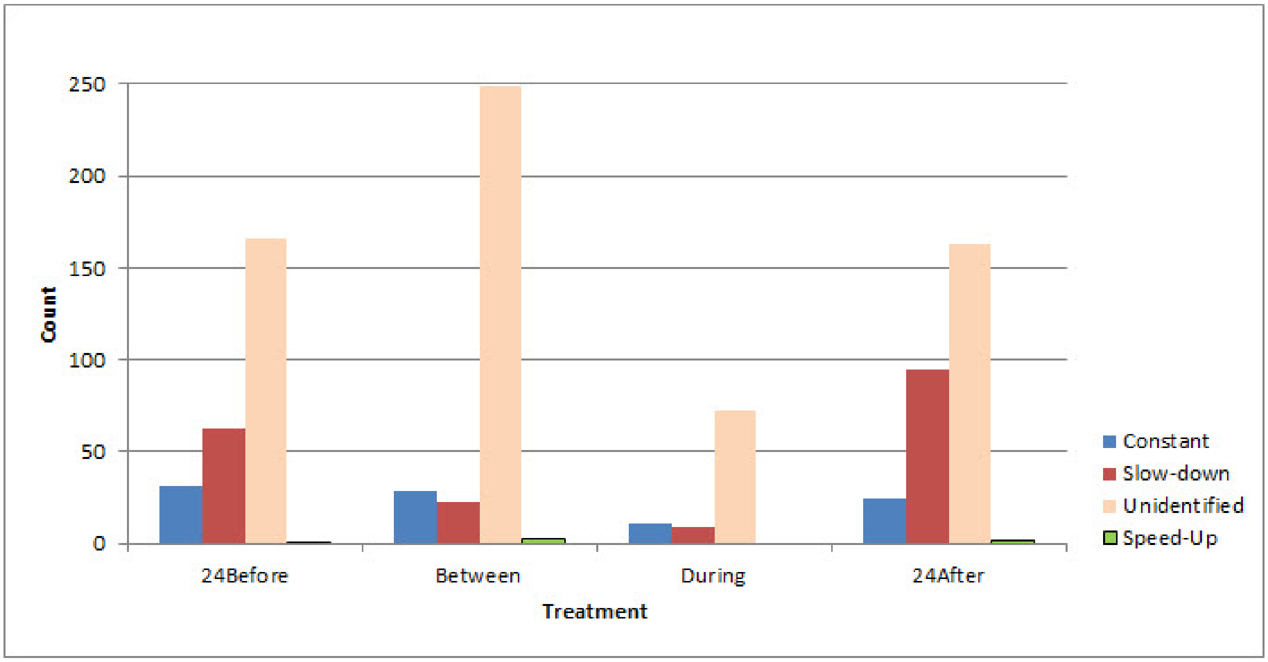

Figure 9. Summary of minke whale pulse train type detections at the Cape Hatteras site (HAT01A) for each of the four sonar occurrence categories.

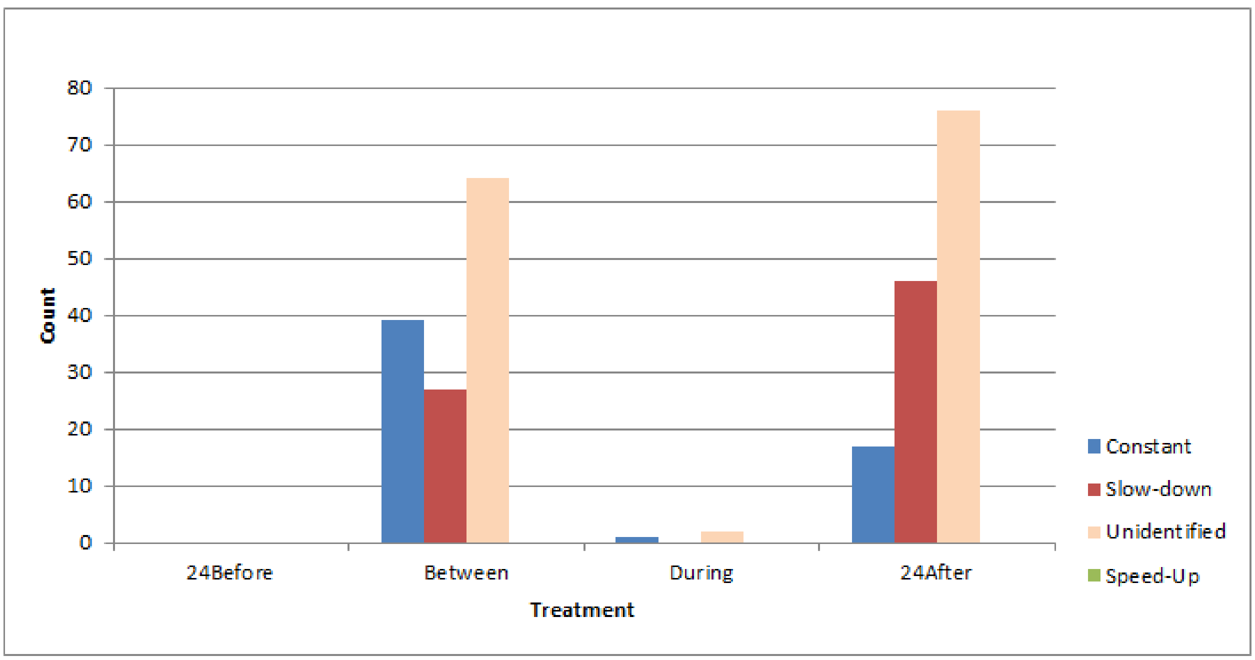

Figure 10. Summary of minke whale pulse train types at the Onslow Bay 1 site (USWTR05A) for each of the four sonar occurrence categories.

4.2.2 Delphinids

A total of 88 delphinid encounters with an overall duration of approximately 334 hours was logged in the dataset (**Table 8**).

Table 8. Summary of delphinid encounters per site.

Site ID	Total No. of Delphinid Encounters per Site	Total No. of Delphinid Clicks per Site	Total No. of Delphinid Whistles per Site	Total Duration of Delphinid Encounters Per Site (hh:mm:ss)
Cape Hatteras (HAT01A)	4	1,249,318	881,897	117:38:14
Jacksonville (JAX05A)	62	852,551	26,524	162:57:43
Onslow Bay 1 (USWTR05A)	15	30,485	2,840	22:33:45
Onslow Bay 2 (USWTR06E)	7	36,493	46,838	31:20:10
Total	**88**	**2,168,847**	**958,099**	**334:29:52**

4.2.3 Beaked Whales

A total of 61 beaked whale encounters with an overall duration of approximately 37 hours was logged in the dataset (**Table 9**).

Table 9. Summary of beaked whale encounters by site. The total number of clicks is shown in parentheses.

Site ID	Total No. of Unidentified *Mesoplodon* Encounters per Site	Total Duration of Unidentified *Mesoplodon* Encounters per Site (hh:mm:ss)	Total No. of Cuvier's Beaked Whale Encounters per Site	Total Duration of Cuvier's Beaked Whale Encounters per Site (hh:mm:ss)	Total No. of Sowerby's Beaked Whale Encounters per Site	Total Duration of Sowerby's Beaked Whale Encounters per Site (hh:mm:ss)
Cape Hatteras (HAT01A)	3 (330)	0:34:02	38 (23,654)	27:10:01	-	-
Jacksonville (JAX05A)	4 (62)	0:52:51	6 (129)	0:21:07	-	-
Onslow Bay 1 (USWTR05A)	1 (35)	0:13:42	3 (110)	0:35:50	-	-
Onslow Bay 2 (USWTR06E)	4 (2,687)	2:00:51	-	-	2 (1,913)	5:23:16
Total	**12 (3,114)**	**3:41:26**	**47 (23,893)**	**28:06:58**	**2 (1,913)**	**5:23:16**

4.2.4 Sperm Whales

A total of 49 sperm whale encounters with an overall duration of approximately 79 hours was logged in the dataset (**Table 10**).

Table 10. Summary of sperm whale encounters per site. The total number of clicks is shown in parentheses.

Site ID	Total No. of Sperm Whale Encounters per Site	Total Duration of Sperm Whale Encounters Per Site (hh:mm:ss)
Cape Hatteras (HAT01A)	24 (28,625)	47:41:36
Onslow Bay 1 (USWTR05A)	2 (43)	0:01:03
Onslow Bay 2 (USWTR06E)	23 (11,997)	31:18:55
Total	**49 (40,665)**	**79:01:34**

4.2.5 Cetaceans and Sonar

For each site and species, encounters were plotted with sonar events overlaid with periods of darkness depicted with shading (**Figures 11 through 14**).

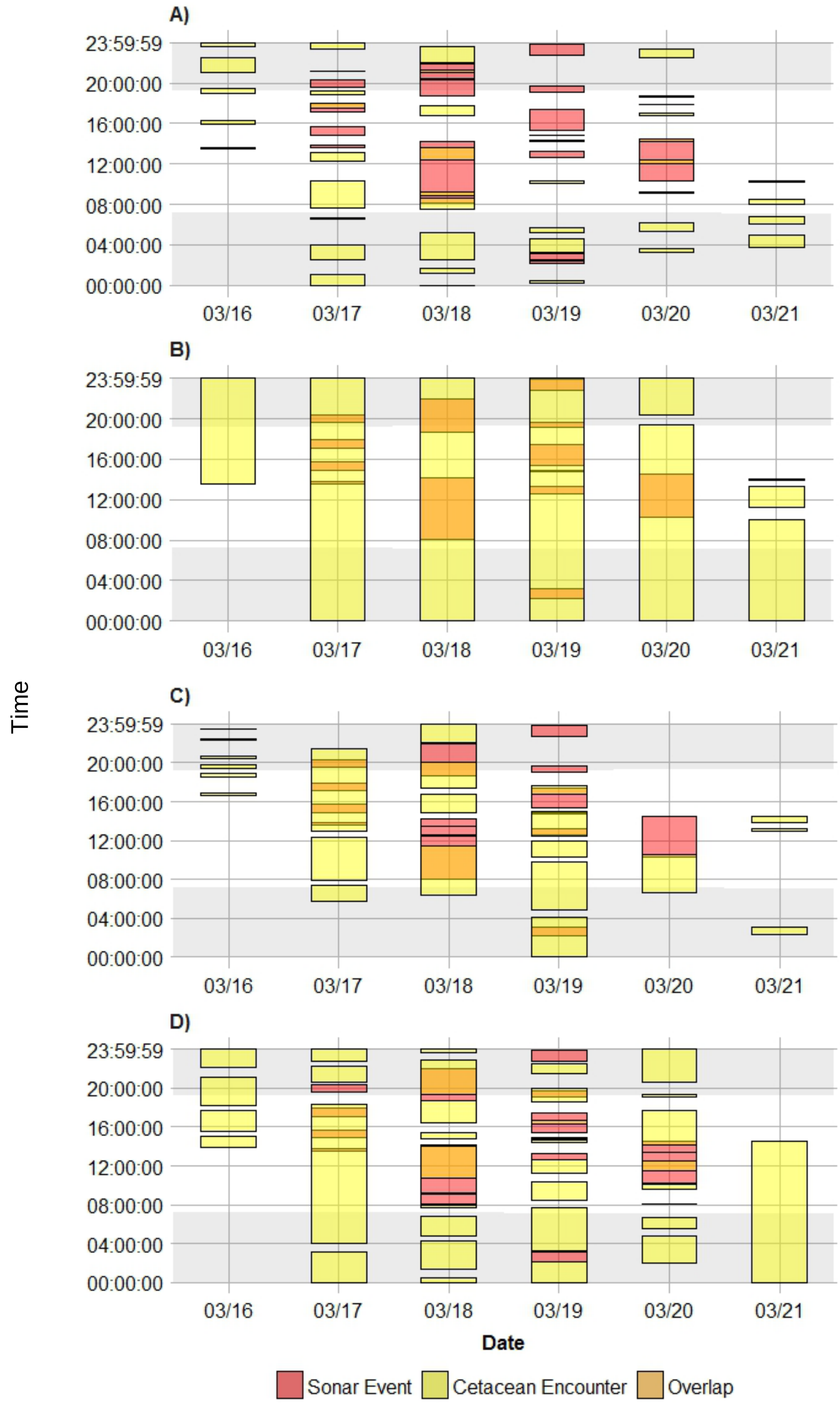

Figure 11. Plot of sonar events (pink) and acoustic encounters (yellow) by species for the HARP deployed off Cape Hatteras, NC (HAT01A). A) Beaked whales, B) Delphinids, C) sperm whales, and D) minke whales. Time of day is plotted on the y-axis, date is plotted on the x-axis, and shading represents periods of light and darkness. "Overlap" refers to periods with the co-occurrence of sonar events and acoustic encounters.

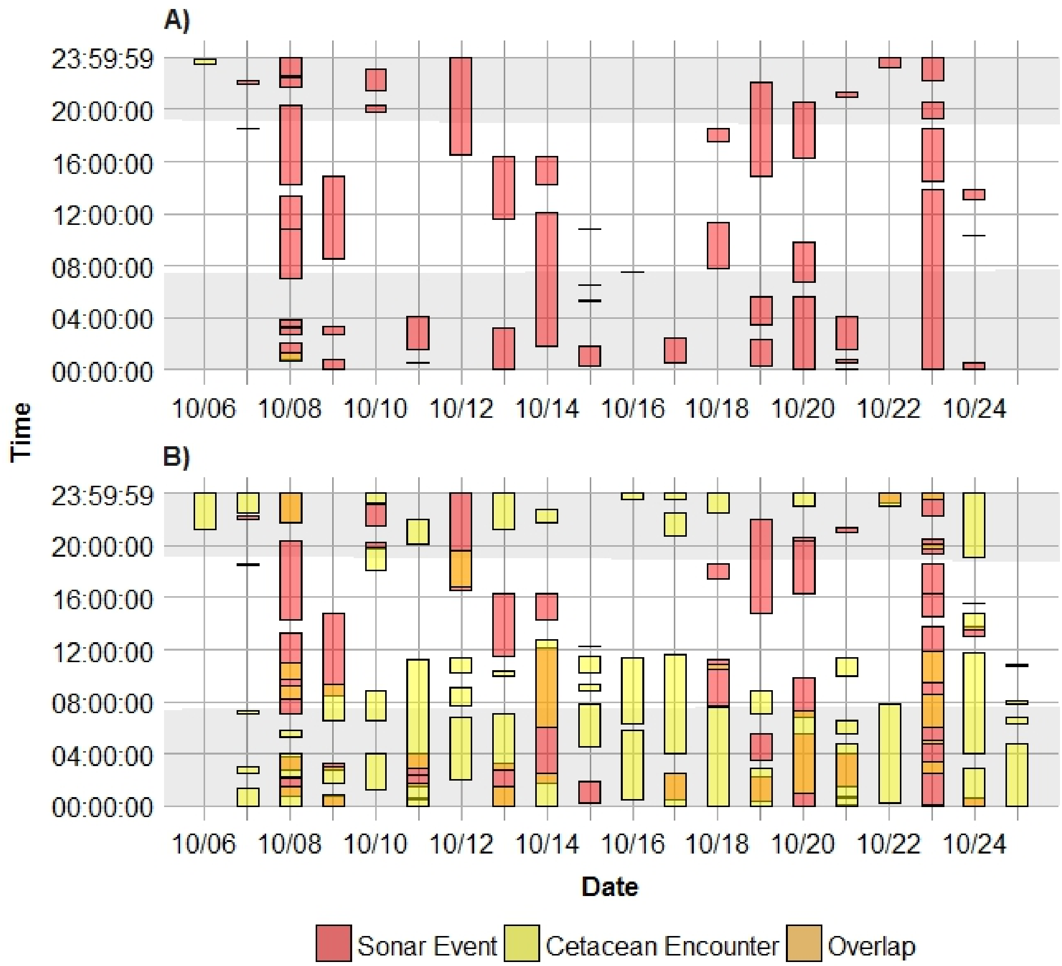

Figure 12. Plot of sonar events (pink) and acoustic encounters (yellow) by species for the HARP deployed off Jacksonville, Florida (JAX05A). A) Beaked whales and B) delphinids. Time of day is plotted on the y-axis, date is plotted on the x-axis, and shading represents periods of light and darkness. “Overlap” refers to periods with the co-occurrence of sonar events and acoustic encounters.

Figure 13. Plot of sonar events (pink) and acoustic encounters (yellow) by species for the Onslow Bay 2 HARP deployed in Onslow Bay, North Carolina (USWTR06E). A) Beaked whales, B) delphinids, and C) sperm whales. Time of day is plotted on the y-axis, date is plotted on the x-axis, and shading represents periods of light and darkness. "Overlap" refers to periods with the co-occurrence of sonar events and acoustic encounters.

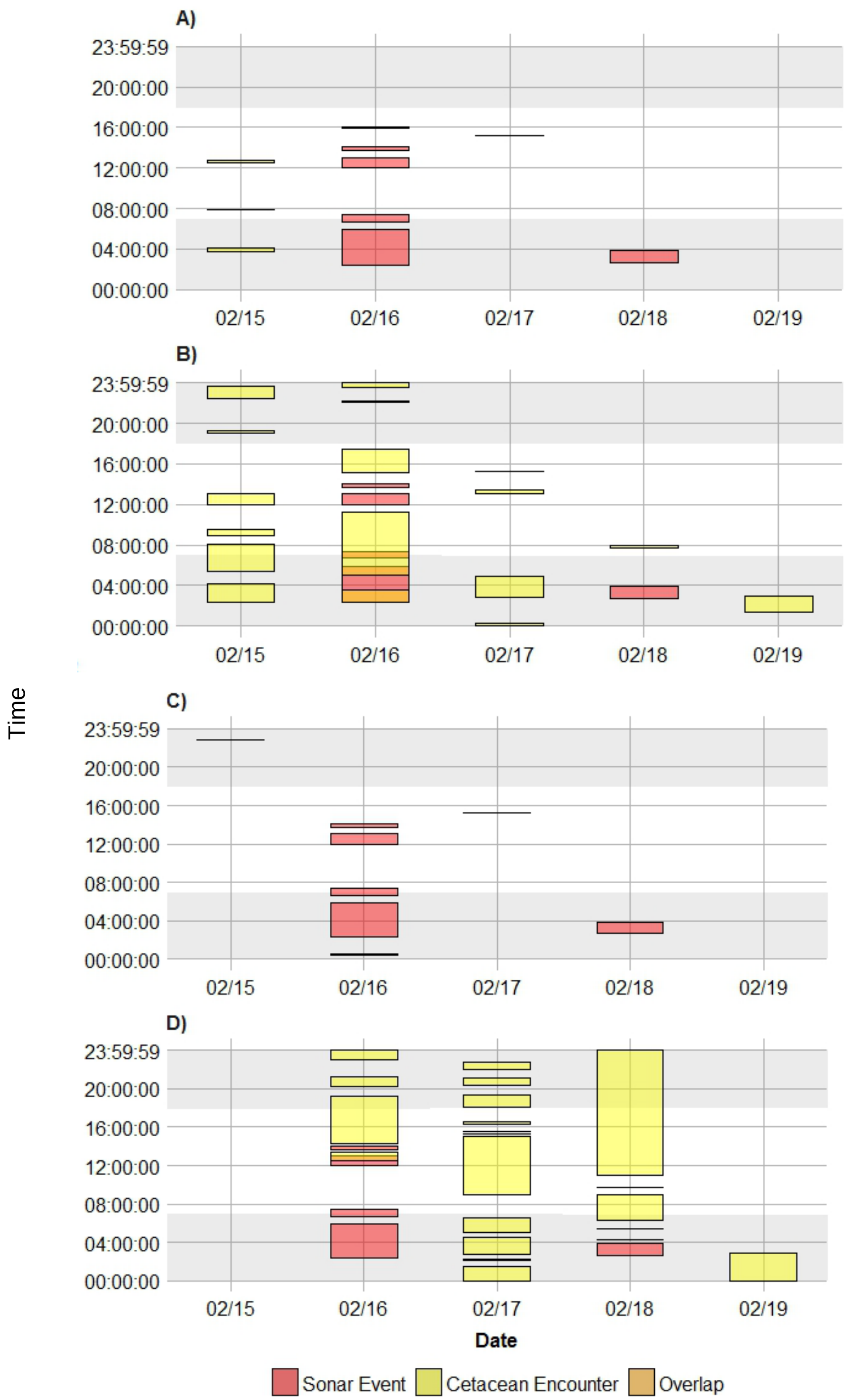

Figure 14. Plot of MFA sonar events (pink) and acoustic encounters (yellow) by species for the Onslow Bay 1 HARP deployed in Onslow Bay, North Carolina (USWTR05A). A) Beaked whales, B) delphinids, C) sperm whales, and D) minke whales.Time of day is plotted on the y-axis, date is plotted on the x-axis, and shading represents periods of light and darkness. “Overlap” refers to periods with the co-occurrence of MFA sonar events and acoustic encounters.

4.3 Classification Analysis

A total of 44 delphinid encounters contained enough whistles to be included in the ROCCA analysis. Most of these encounters (n=28) were recorded on the HARP deployed off Jacksonville. The majority of these encounters (61 percent) were classified as Atlantic spotted dolphins or bottlenose dolphins, and small percentages of encounters were classified as pilot whales or striped dolphins (**Figure 15)**. There were six encounters analyzed from the Onslow Bay 1 data and six encounters analyzed from the Onslow Bay 2 data. Most of the encounters were classified as bottlenose dolphins and a small percentage were classified as pilot whales or striped dolphins at both sites (**Figure 15**). One Onslow Bay 2 encounter was classified as spotted dolphins. All of the five encounters analyzed from the Cape Hatteras data were classified as bottlenose dolphins.

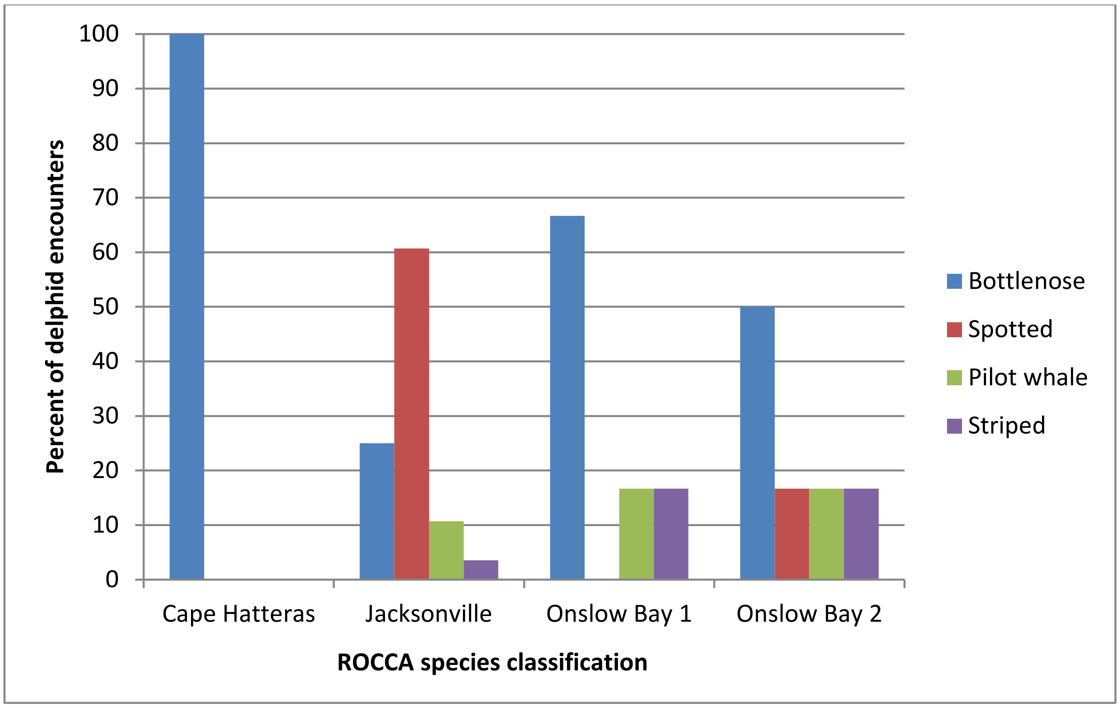

Figure 15. Percentage of delphinid acoustic encounters classified by species for Cape Hatteras (n=5), Jacksonville (n=28), and Onslow Bay 1 (n=6), Onslow Bay 2 (n=6).

4.4 Statistical Modeling

4.4.1 Delphinids

4.4.1.1 REGRESSION ANALYSIS OF PRESENCE-ABSENCE OF ACOUSTIC SIGNAL IN 1 MINUTE SEGMENTS

To remain consistent with previous analyses, all delphinids were considered together for this analysis. Given that delphinids were present at all sites, the total of 21,581 minutes of HARP data were combined with 148,359 minutes of MARU data (both pilot whales and other delphinids). Delphinids were present in 50 percent of the HARP 1-minute segments and 7 percent of the MARU 1-minute segments. The initial analysis of the residuals suggested significant autocorrelation up to 861 minutes; hence this was established as the block size for the GEE models for delphinids. The final selected model (PA model 1) consisted of *Site*, (*p*<0.001) (**Figure 16**). No effect of *Sonar* was found in this combined model of HARP and MARU data (but see below). The model predicted 88 percent of observed values correctly using the fitted value > mean fitted value criterion above to identify presences etc. (**Appendix C,** Table 1). Other diagnostics are given in **Appendix C,** Figure 1.

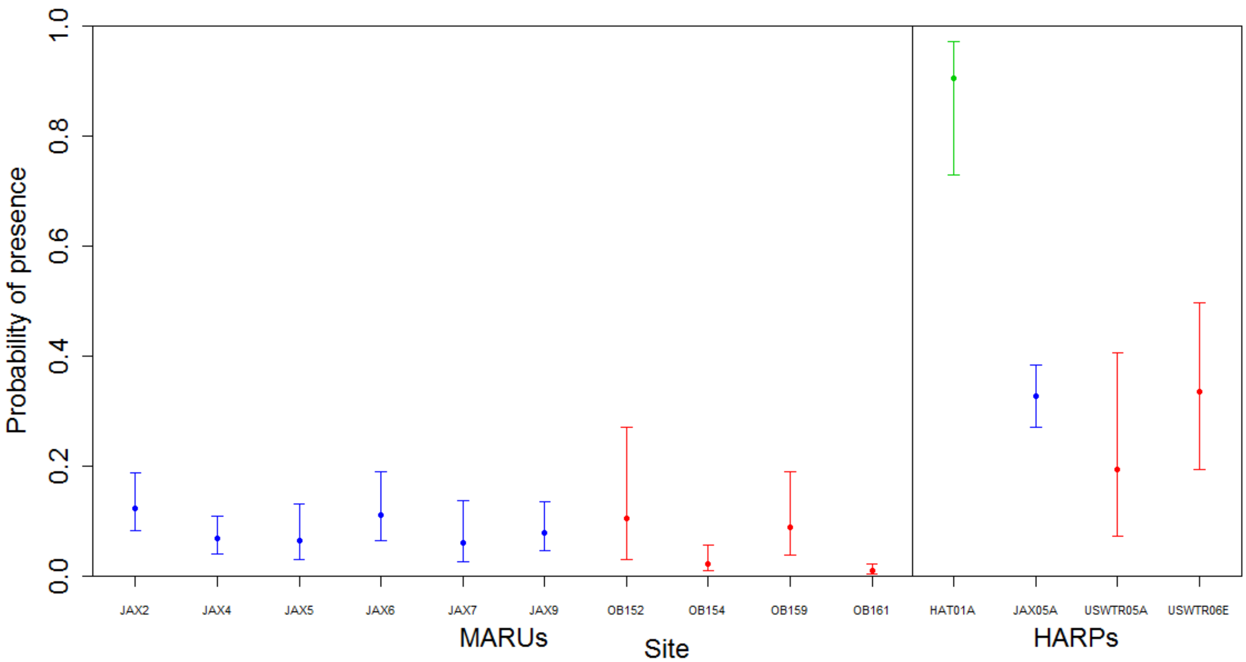

Figure 16. Predicted probability of detected delphinid acoustic presence by *Site* (from PA model 1)*.* Jacksonville sensors in blue, Onslow Bay sensors in red, and the Cape Hatteras (HAT01A) sensor is in green. Bars indicate 99% confidence intervals.

We did not find a *Sonar* effect, nevertheless after consideration of the individual components of sonar, one additional predictor was kept in the final model: *Type 1 long* (*p*<0.001, **Figure 17**). This model (PA model 2) predicted 86 percent of its observations correctly **Appendix C,** Table 2). Other diagnostics are given in **Appendix C,** Figure 2**.**

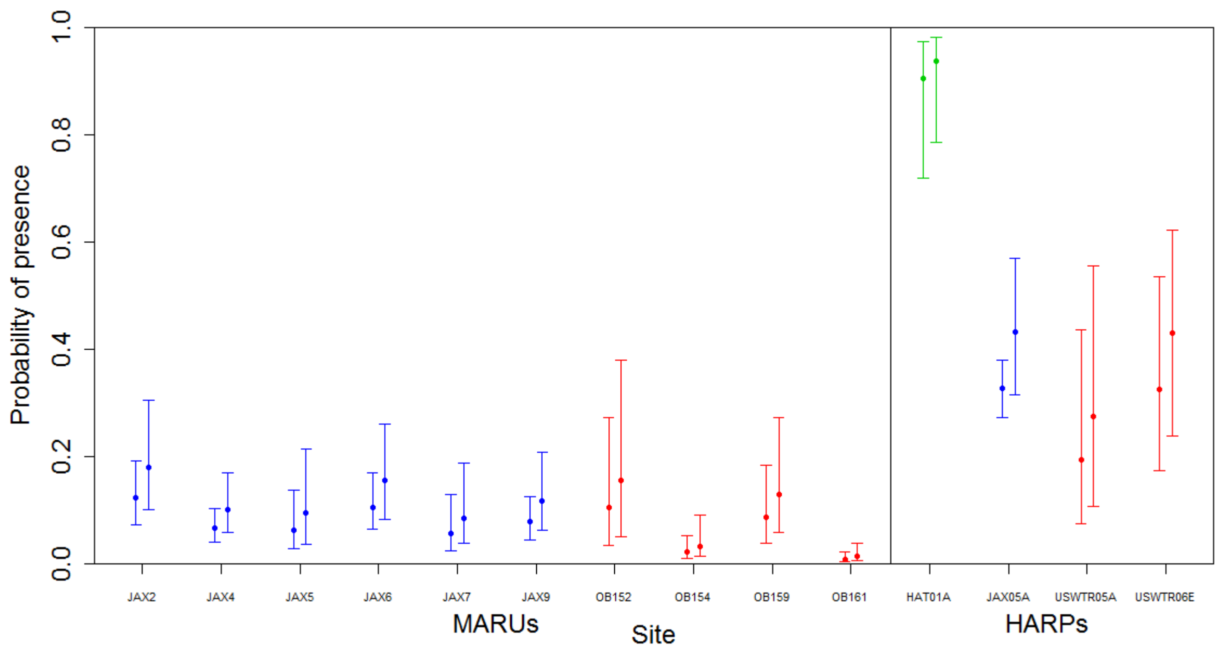

Figure 17. Predicted probability of detected delphinid acoustic presence from a model with *Site* and *Type 1 long* predictors (from PA model 2). Jacksonville sensors in blue, Onslow Bay sensors in red and the Cape Hatteras (HAT01A) sensor is in green. Bars indicate 99% confidence intervals. The estimates for each site are in clusters representing absence and presence respectively of *Type 1 long* signals, represented by the two adjacent bars.

4.4.1.2 HIDDEN MARKOV MODEL.

In this section, we first provide an overview of the baseline HMMs and the HMMs with covariates for each species and each site based on the model selection criterion Akaike Information Criterion (AIC). We note that, compared to the GEEs, model selection for the HMM was only between the baseline model and single-covariate models (either *Sonar Sonaron*). A closer look at the estimated parameters of the baseline models provides interesting insights. In the last part of this section we focus on one particular model fit with the aim to investigate the influence of the covariate *Sonar* on the TPM) that is a matrix indicating the probability of changing (transitioning) from state to state, and the stationary distribution.

Altogether 36 models were fit, three for each time series (baseline with no covariates, covariate *Sonar*, covariate *Sonaron*). The AICs of all models are given in **Table 11**.

Table 11. AIC values for the 36 models. Models with the lowest AIC are given in bold. Blue color indicates data sets without NAs (missing data).

Taxa	Site ID	Baseline	*Sonar*	*Sonaron*
Delphinids	Jacksonville (JAX05A)	**3243.50**	3244.28	3261.04
	Cape Hatteras (HAT01A)	1985.08	**1853.40**	1900.79
	Onslow Bay 1 (USWTR05A)	931.35	**931.00**	931.02
	Onslow Bay 2 (USWTR06E)	**882.59**	883.69	885.59
Beaked whales	Jacksonville (JAX05A)	**201.97**	210.14	205.31
	Cape Hatteras (HAT01A)	**871.99**	878.40	880.91
	Onslow Bay 1 (USWTR05A)	**89.01**	95.26	92.63
	Onslow Bay 2 (USWTR06E)	**141.31**	146.21	142.84
Minke whales	Cape Hatteras (HAT01A)	6091.10	**6019.22**	6080.88
	Onslow Bay 1 (USWTR05A)	1815.81	**1811.43**	1813.69
Sperm whales	Cape Hatteras (HAT01A)	584.39	**578.47**	604.29
	Onslow Bay 2 (USWTR06E)	**473.60**	479.20	476.46

No single model outperformed its competitors in terms of AIC for all time-series data. However, certain patterns became obvious:

- The models with *Sonaron* were never preferred over the other alternatives. In other words, if a model with a covariate was chosen, *Sonar* provided more relevant information about the transition probabilities than *Sonaron*.
- The baseline model was chosen for all beaked whales time-series data. There was, therefore, an indication that covariate *Sonar* did not affect the probability of acoustic detection of the beaked whale species group.
- For all other species groups (provided they were present), there was a significant effect of the covariate *Sonar* for the sites HAT05A and USWTR05A.

Since the baseline model performed relatively well in the model comparison, we investigated the baseline model results in more detail. **Table 12** provides the estimates of the six parameters in the 12 baseline models.

Judging by the values of the estimated parameters of the state-dependent distributions, state 1 can be labelled as the “silent” state and state 2 as the “acoustically active” state. **Table 12** confirms the impression that we obtained from the acoustic occurrence plots—that the models differ across species groups and sites with the possible exceptions of beaked whales and sperm whales. However, a closer look at the estimates of π_1 and π_2 for beaked whales and sperm whales suggests that the state-dependent process is deterministic and thus a Markov Chain (with two observable states—acoustically active/not acoustically active) rather than an HMM is the more appropriate model. As seen in **Figure 5** this is not much of a surprise - the periods with presences of acoustic occurrence are hardly interspersed with absences (ignoring the NAs), nor do periods of absences contain presences.

Table 12. Estimates of the parameters in the baseline models. Blue color indicates data sets without Ns (see Section 3.3.3).

Taxa	Site ID	π_1	π_2	γ_{11}	γ_{21}	γ_{12}	γ_{22}
Delphinids	Jacksonville (JAX05A)	0.006	0.953	0.994	0.011	0.006	0.989
	Cape Hatteras (HAT01A)	0.168	1.000	0.896	0.013	0.104	0.987
	Onslow Bay 1 (USWTR05A)	0.000	0.859	0.995	0.018	0.005	0.982
	Onslow Bay 2 (USWTR06E)	0.001	0.922	0.992	0.014	0.008	0.986
Beaked whales	Jacksonville (JAX05A)	0.000	1.000	0.999	0.105	0.001	0.895
	Cape Hatteras (HAT01A)	0.000	1.000	0.993	0.024	0.007	0.976
	Onslow Bay 1 (USWTR05A)	0.000	1.000	0.999	0.069	0.001	0.931
	Onslow Bay 2 (USWTR06E)	0.000	1.000	0.999	0.013	0.001	0.987
Minke whales	Cape Hatteras (HAT01A)	0.000	0.794	0.897	0.283	0.103	0.717
	Onslow Bay 1 (USWTR05A)	0.006	0.403	0.974	0.020	0.026	0.980
Sperm whales	Cape Hatteras (HAT01A)	0.000	1.000	0.995	0.008	0.005	0.992
	Onslow Bay USWTR06E	0.000	1.000	0.992	0.014	0.008	0.986

Finally, we investigated the model fit for the minke whales at the HAT01A site in more detail. The motivation for this stems from the data as well as from the model fit itself.HAT01A offered better quality data (no missing values). Moreover, the data in **Table 12** suggested that, for minke whales, including the covariate *Sonar* brought the largest improvement in terms of AIC over the baseline model. In other words, the effect of the covariate was more pronounced for these species.

For the minke whale data from the HAT05A site, the TPMs associated with the four different levels of the categorical variable *Sonar* were:

$$\hat{\Gamma}(\text{Before}) = \begin{pmatrix} 0.864 & 0.136 \\ 0.231 & 0.769 \end{pmatrix}; \qquad \hat{\Gamma}(\text{During}) = \begin{pmatrix} 0.929 & 0.071 \\ 0.427 & 0.573 \end{pmatrix};$$

$$\hat{\Gamma}(\text{Between}) = \begin{pmatrix} 0.913 & 0.087 \\ 0.352 & 0.648 \end{pmatrix}; \qquad \hat{\Gamma}(\text{After}) = \begin{pmatrix} 0.838 & 0.162 \\ 0.267 & 0.733 \end{pmatrix}$$

The associated stationary (or equilibrium) distributions, which indicate the expected proportion of time spent in the two states (for a given covariate level), were:

$$\hat{\delta}(\text{Before}) = (0.629{,}0.371)$$

$$\hat{\delta}(\text{During}) = (0.858{,}0.142)$$

$$\hat{\delta}(\text{Between}) = (0.802{,}0.198)$$

$$\hat{\delta}(\text{After}) = (0.623{,}0.377)$$

In general, the persistence in the "silent" state was higher than the persistence in the "acoustically active" state. The largest persistence in the "silent" state was during the 1-minute segments when *Sonar* equalled "during." Furthermore, the probability of transition from the "acoustically active" to the "silent" state was the highest in this case. The patterns exhibited in the TPMs (and the related stationary distributions) were similar for "before" and "after" the sonar exposure conditions as well as for the "during" and "between" exposure conditions. It is important to note that in terms of the stationary distribution, minke whales were more likely to be acoustically undetected in the "during" and "between" exposure conditions than in the "before" or "after" exposure condition.

4.4.1.3 *PRESENCE OF SIGNAL TYPE GIVEN CETACEAN ACOUSTIC ENCOUNTER* (PSTGAE) MODELS

There were 10,881 delphinid acoustic sub-encounters in the HARP data. Despite there being no overall effect of MFA sonar on acoustic occurrence (see **Section 4.4.1.1**), there was an effect of MFA sonar on the probability of acoustic occurrence for some particular delphinid signal types (not to be confused with the sonar signal types used as predictors). The final fitted model (PSTGAE model 1) for whistles consisted of *Site* (p<0.001), *Sonar* (p<0.001), presence of buzzes *,Buzzes,* (p<0.001), and presence of clicks *Clicks,* (p<0.001) (**Figures 18 through 20**). The block size was 281 and 86 percent of the observed values were predicted correctly (**Appendix C,** Table 3). Diagnostics are given in **Appendix C,** Figure 3.

Consideration of individual MFA sonar components resulted in a model with *Site* (p<0.001), *Buzzes* (p<0.001), *Clicks* (p<0.001) and *Type 1 long* (p<0.001). Block size for this model (PSTGAE model 2) was 733. The *Type 1 long* signal effect is illustrated in **Figure 21,** and 86 of the observed values were predicted correctly (**Appendix C,** Table 4). Diagnostics are given in **Appendix C,** Figure 4**.**

The fitted model (PSTGAE model 3) for clicks as the dependent variable consisted of *Site* (p<0.001), presence or absence of whistles (*Whistles* (p<0.001), *Buzzes* (p<0.001), and *Sonar* (p<0.001). The final block size was 94. The effect of *Sonar* is shown in **Figure 22**. The effect of *Buzzes* is shown in **Figure 23.** Diagnostic plots are given in **Appendix C,** Figure 5. The model correctly predicted 80 percent of the observed values (**Appendix C,** Table 5).

Further model selection (PSTGAE model 4) of the click data considering the components of the MFA sonar led to a model with *Site* (p<0.001), *Buzzes* (p<0.001), *Whistles* (p<0.001), and *Type 2 long* (p<0.001); (**Figure 24**) diagnostics in **Appendix C.** Figure 6. The model correctly predicted 70 percent of the observed values (**Appendix C**, Table 6). The final block size was 150.

The fitted model (PSTGAE model 5) for buzzes consisted of *Site* (p<0.001), *Whistles* (p<0.001)*,* *Buzzes* (p<0.001), and *Sonar* (p<0.001). The final block size was 223. The effect of *Sonar* is shown in **Figure 25**. Diagnostic plots are given in **Appendix C,** Figure 7. The model successfully predicted 70 percent of all observations (**Appendix C,** Figure 7).

Model selection considering the components of the MFA sonar led to a buzz model (PSTGAE model 6) with *Site* (p<0.001), *Clicks* (p<0.001), *Whistles* (p<0.001), and *Type 3 med* (p<0.001). The effect of the latter two variables is shown in **Figure 26**. Diagnostics can be found in **Appendix C,** Figure 8. The block size was 253. The model successfully predicted 66 percent of the observed values (**Appendix C,** Table 8).

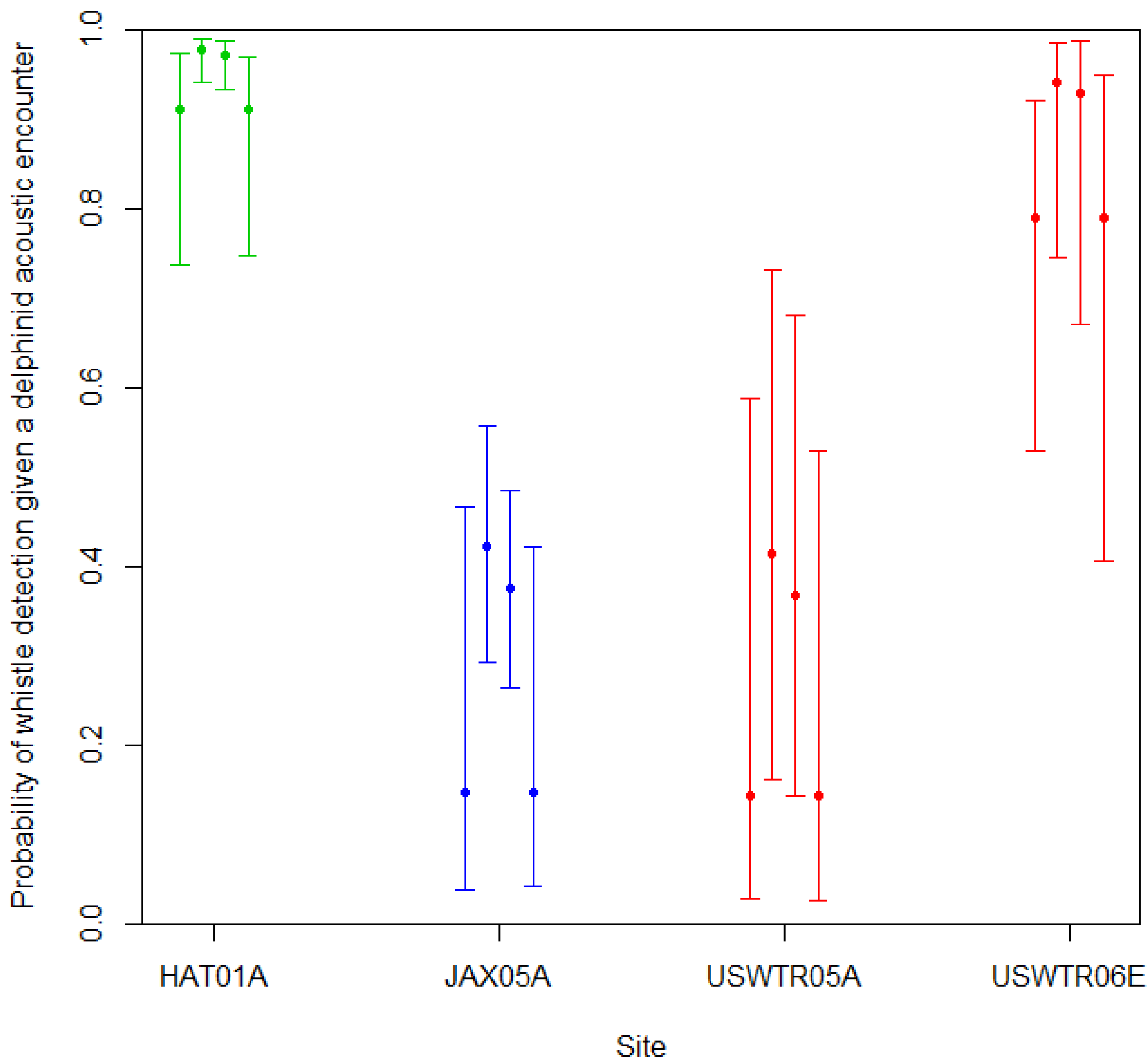

Figure 18. Predicted probability of whistle detection given delphinid acoustic activity from a model with *Site*, *Sonar*, *Buzzes* and *Clicks* as predictors (from PSTGAE model 1). Predictions assume *Clicks* = 1 (presence) *Buzzes* = 1 (presence). Vertical bars indicate 99% confidence intervals. For each site, predictions are shown here (from left to right) for "before," "during," "between," and "after" sonar activity, respectively.

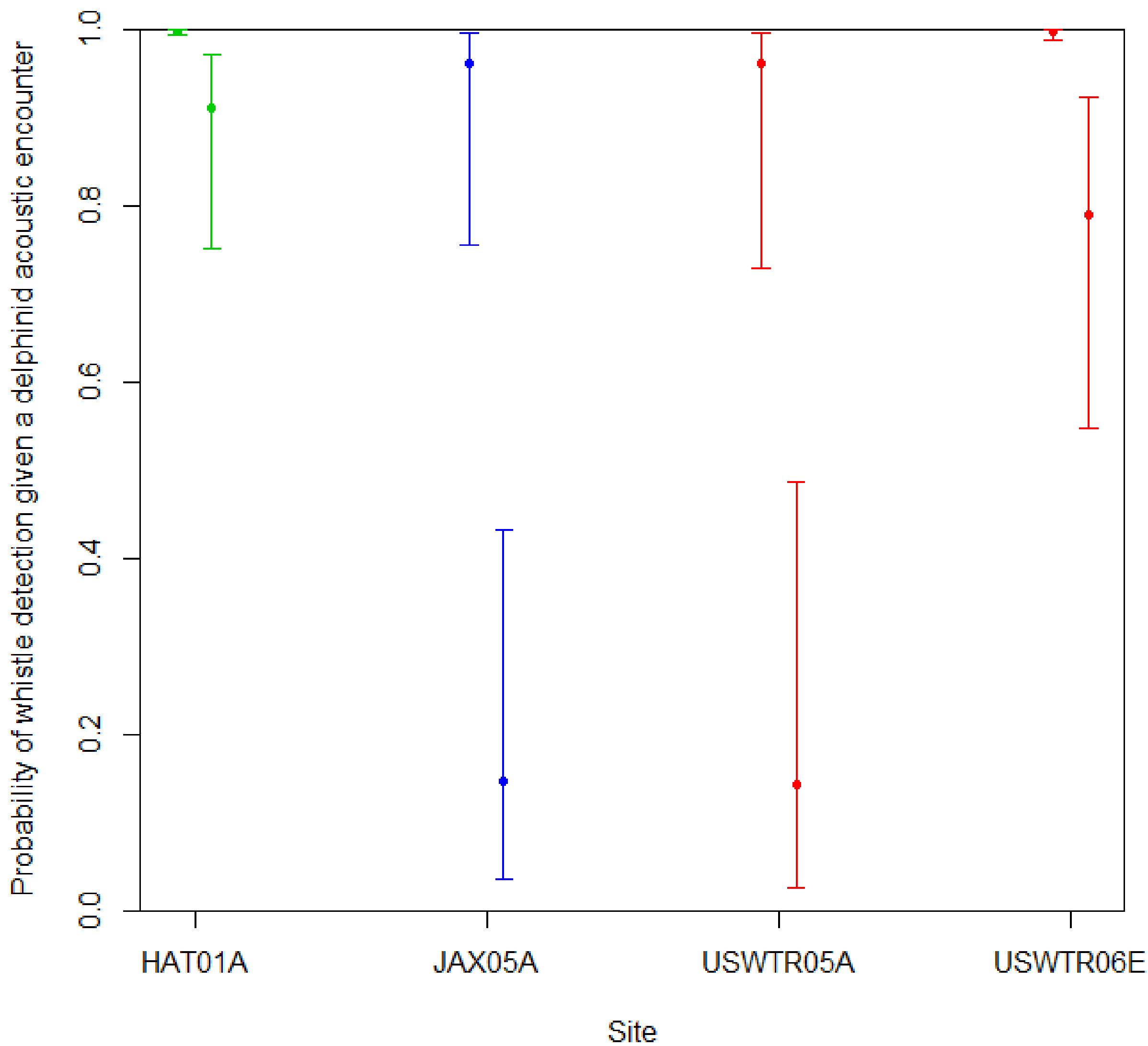

Figure 19. Predicted probability of whistle detection given delphinid acoustic activity with *Site*, *Sonar, Buzzes* and *Clicks* as predictors (PSTGAE model 1). Vertical bars indicate 99% confidence intervals. For each site, predictions are shown here for the absence (left) and presence (right) of clicks, respectively, assuming "before" MFA sonar with *Buzzes* = 1 (presence).

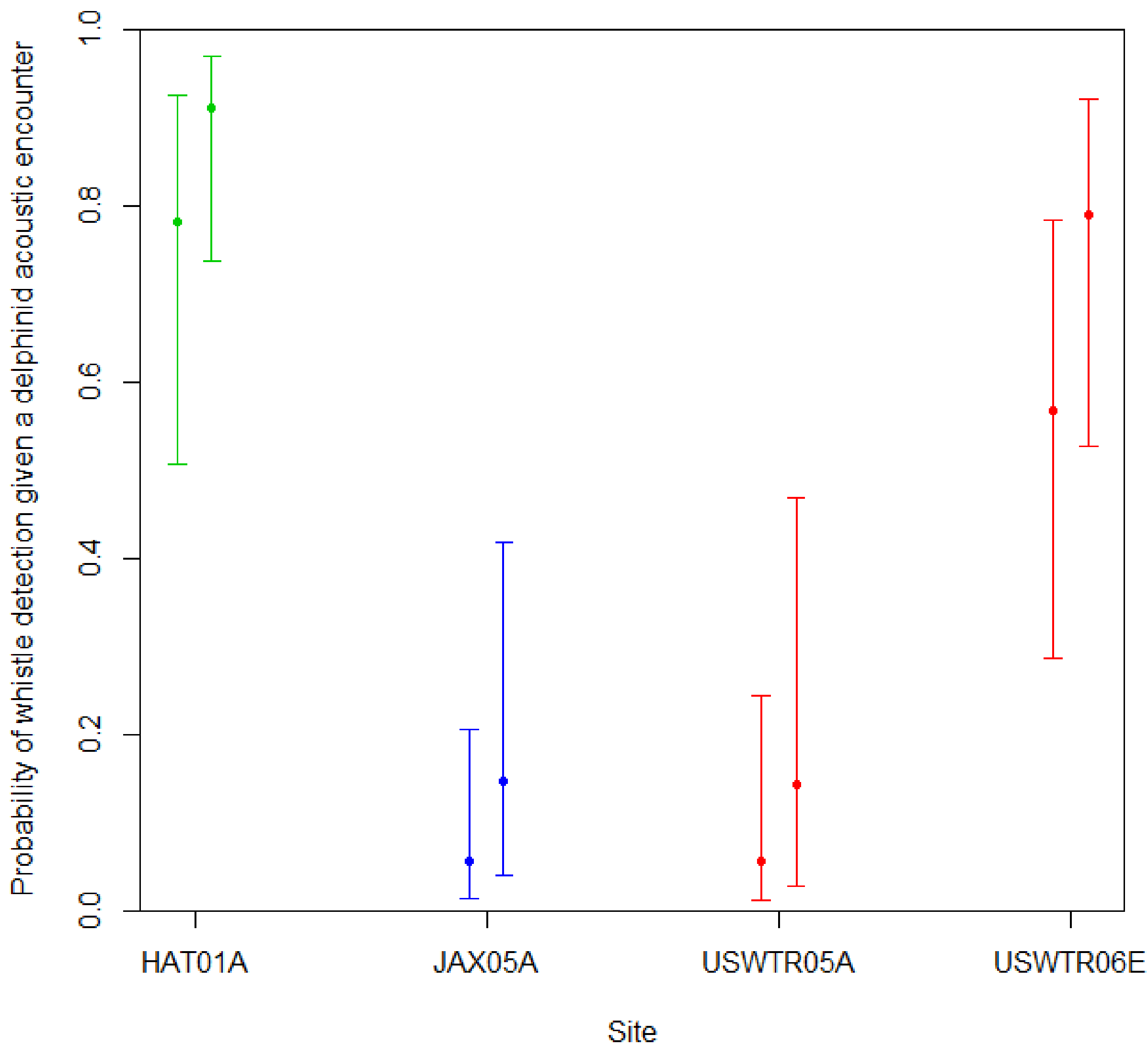

Figure 20. Predicted probability of whistle detection given delphinid acoustic activity with *Site*, *Sonar, Buzzes* and *Clicks* as predictors (PSTGAE model 1). Vertical bars indicate 99% confidence intervals. For each site, predictions are shown here for the absence (left) and presence (right) of buzzes respectively assuming "before" MFA sonar with *Clicks* = 1 (presence).

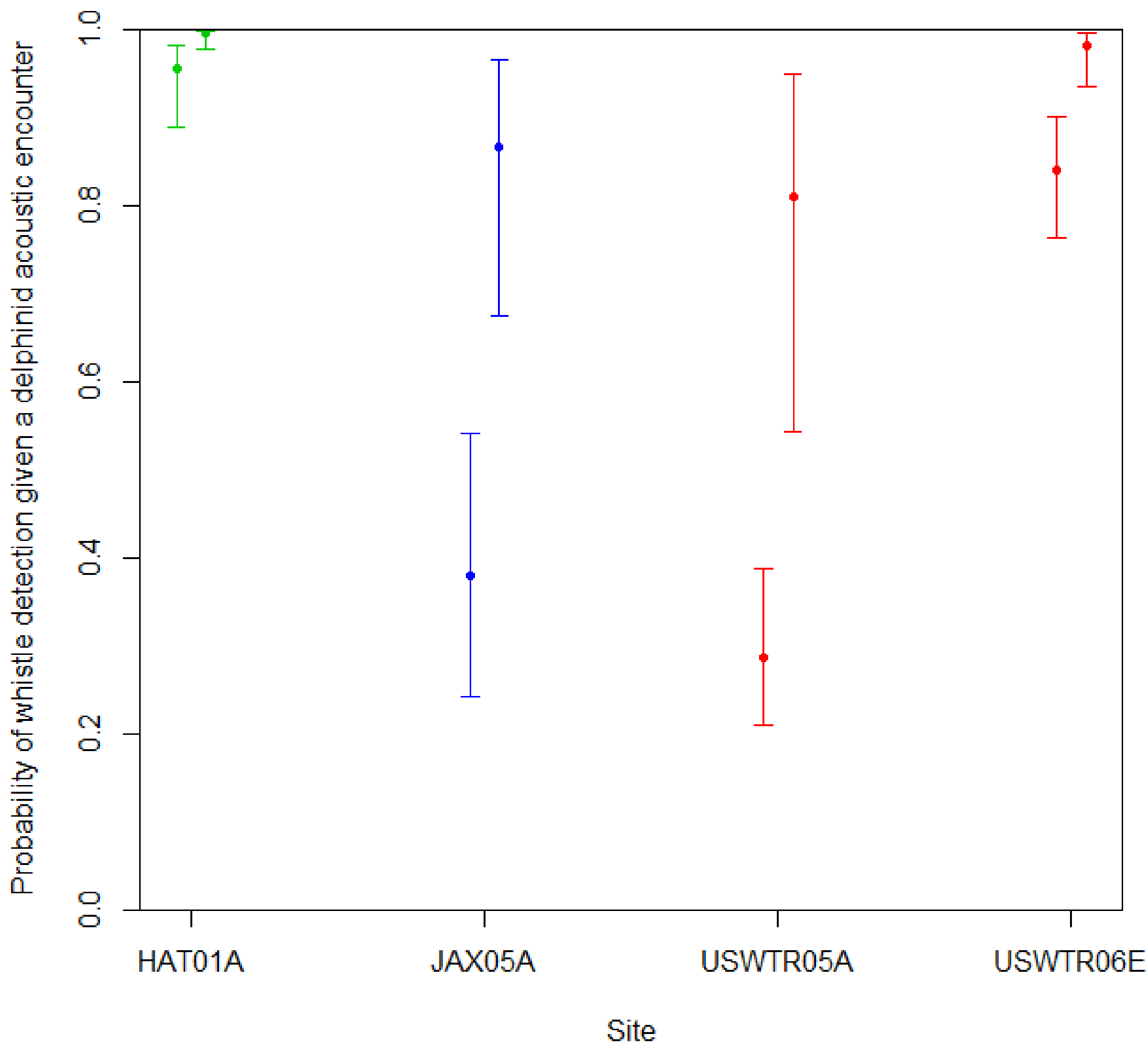

Figure 21. Predicted probability of whistle detection given delphinid acoustic activity (PSTGAE model 2) with *Site, Type 1 Long, Buzzes* and *Clicks* as the predictors. Vertical bars indicate 99% confidence intervals. For each site, predictions are shown here for absence (left) or presence (right) of *Type 1 long* signals respectively with *Clicks* and *Buzzes = 1* (assumed present).

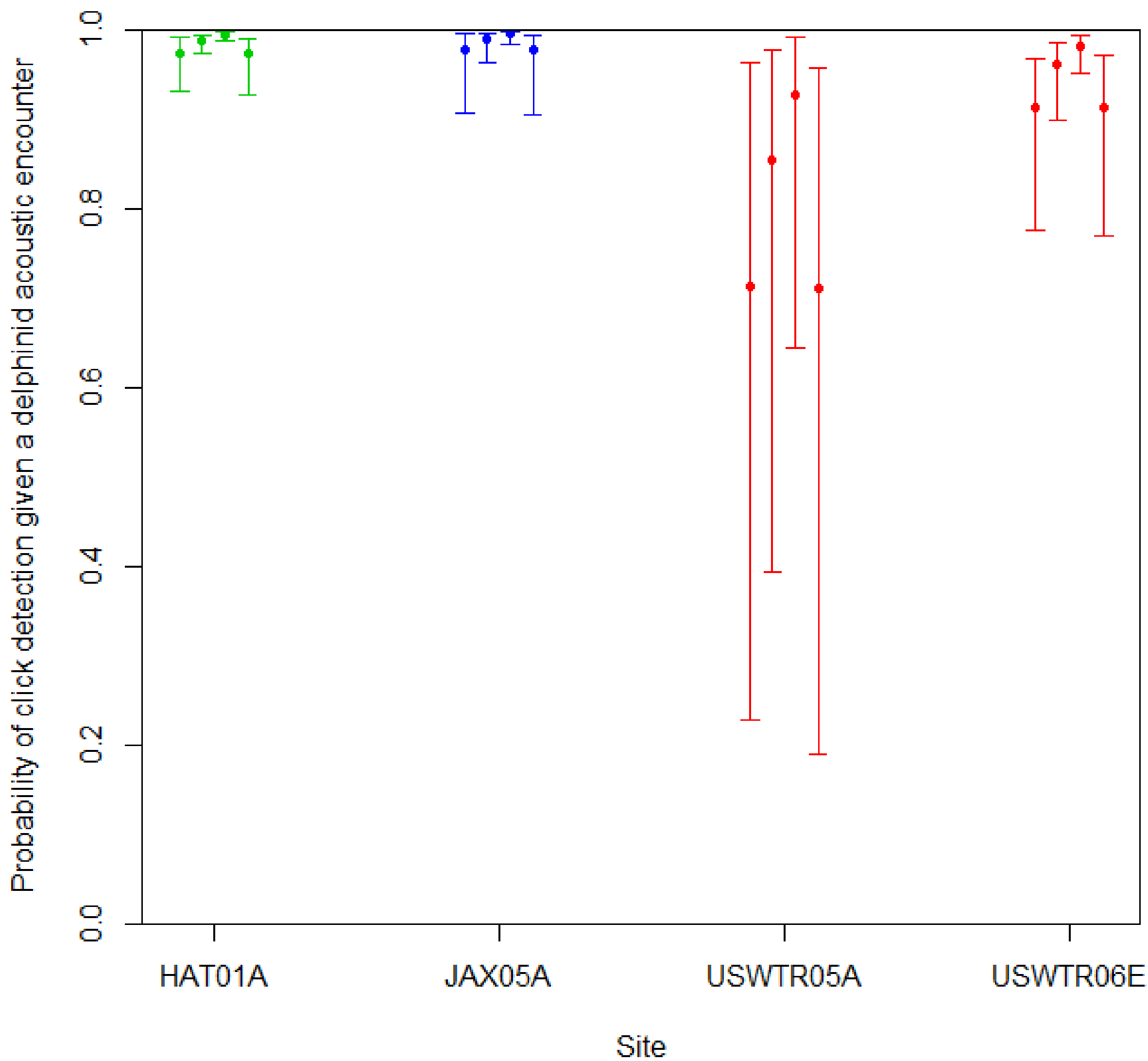

Figure 22. Predicted probability of click detection given delphinid acoustic activity with *Site*, *Whistles*, *Buzzes* and *Sonar* (PSTGAE model 3). Predictions are assuming *Whistles* = 1 and *Buzzes* = 1 (presence). Vertical bars indicate 99 percent confidence intervals. For each site, predictions are shown for "before," "during," "between" and "after" MFA sonar activity, respectively.

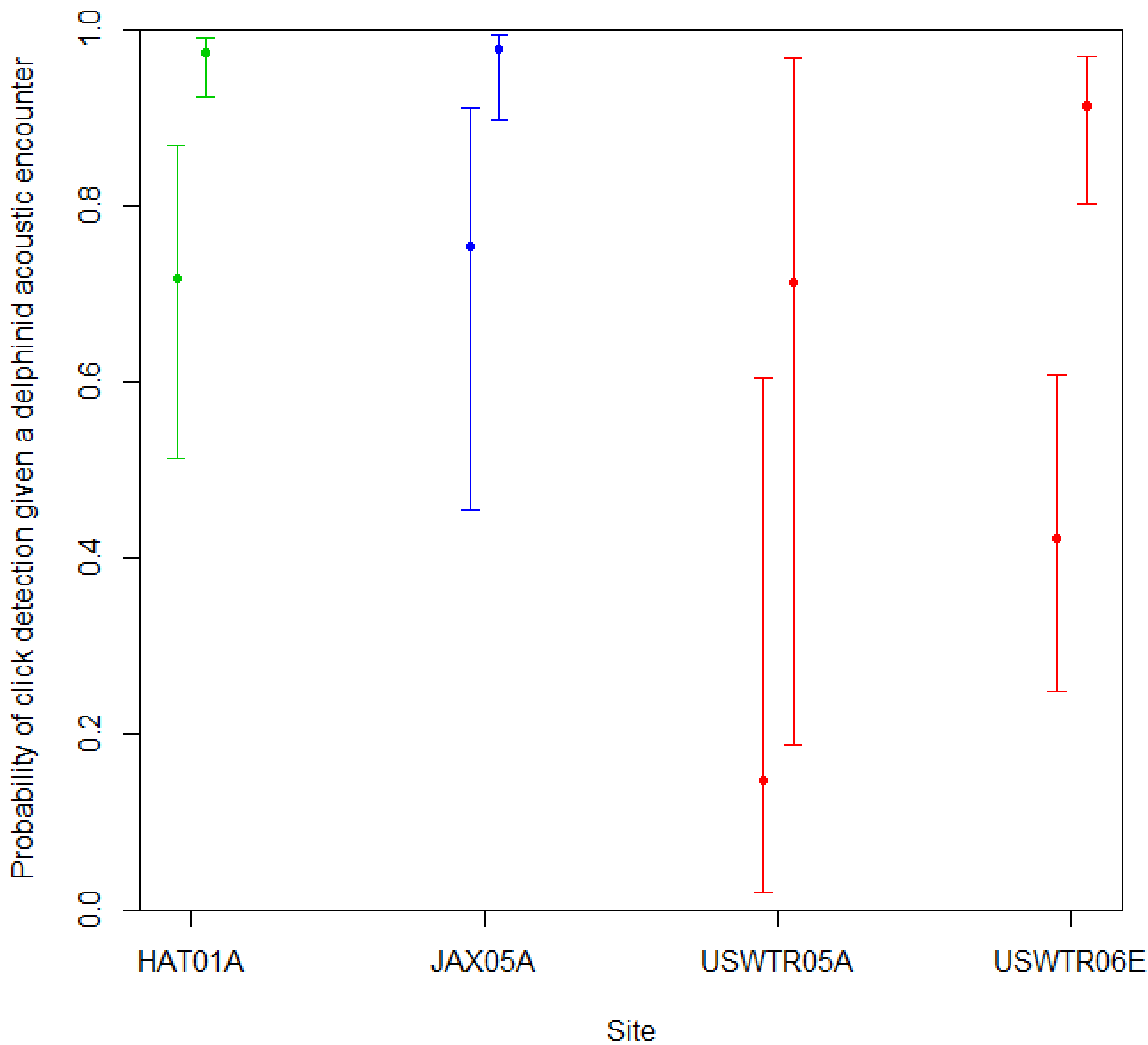

Figure 23. Predicted probability of click detection given delphinid acoustic activity with *Site*, *Whistle*s, *Buzzes* and *Sonar* (PSTGAE model 3). Vertical bars indicate 99% confidence intervals. For each site, predictions are shown for for the absence (left) and presence (right) of buzzes, respectively, with *Whistles* = 1 (assumed present).

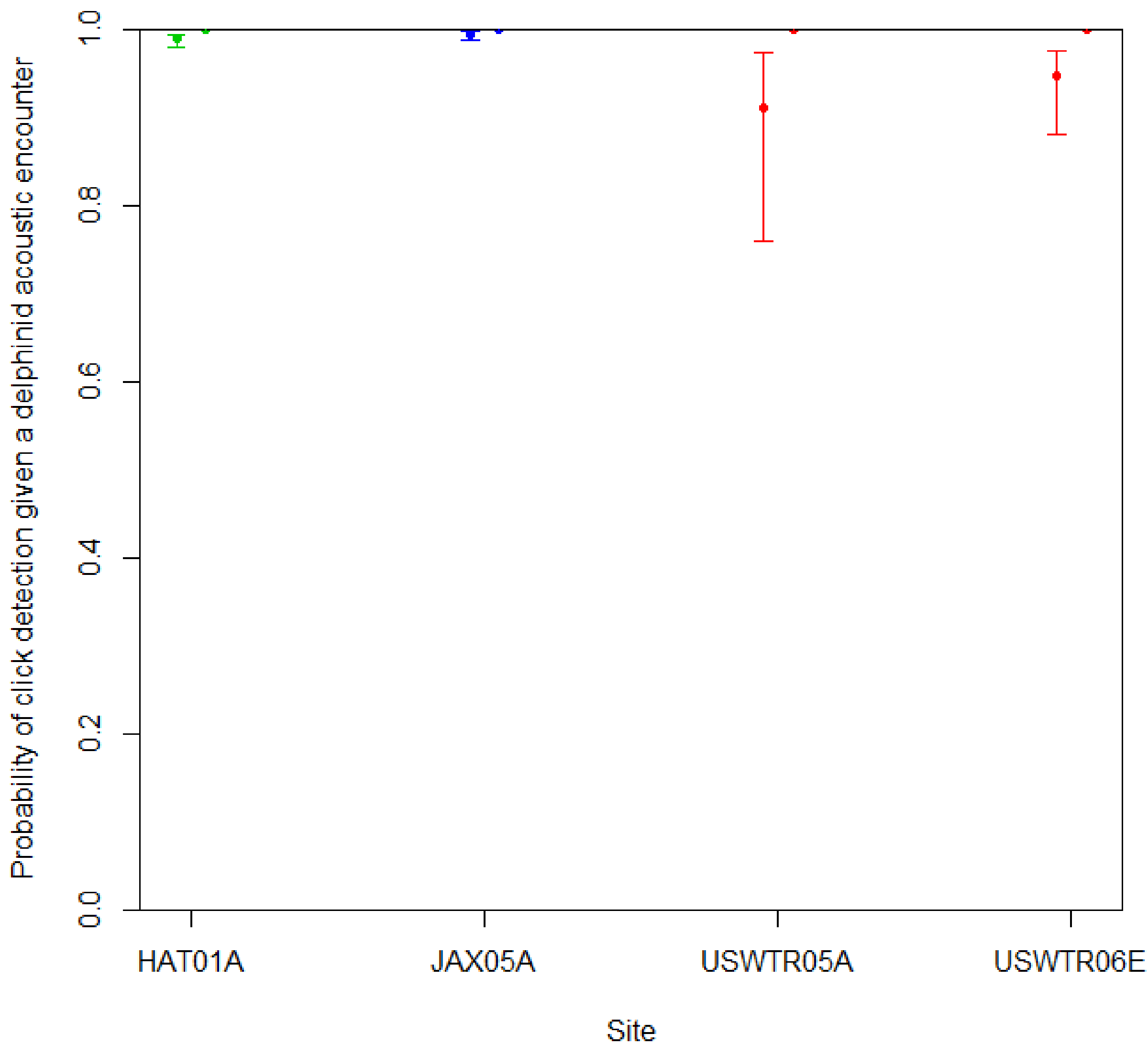

Figure 24. Predicted probability of click detection given delphinid acoustic activity with Site, Whistles, buzzes and Type 2 long signals (PSTGAE model 4). Vertical bars indicate 99% confidence intervals. For each site predictions are shown for absence (left) or presence (right) of *Type 2 long* signals, respectively, with *Whistles* and *Buzzes* = 1 (assumed present).

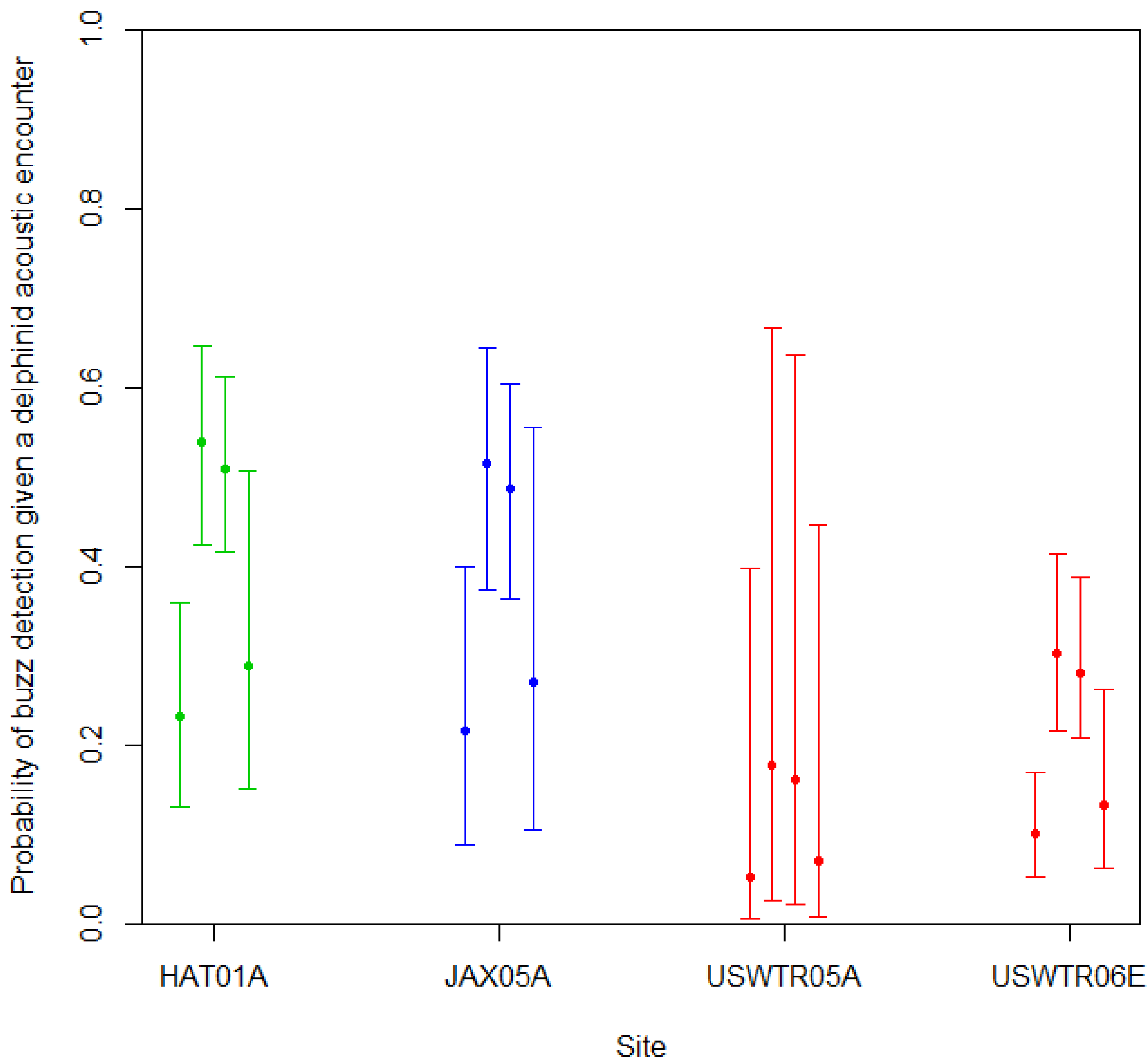

Figure 25. Predicted probability of buzz detection given delphinid acoustic activity with *Site, whistles, Clicks* and *Sonar* as predictors (PSTGAE model 5). Vertical bars indicate 99% confidence intervals. For each site, predictions are shown for for "before," "during," "between" and "after" MFA sonar events, respectively, with *Whistles* and *Clicks* = 1 (assumed present).

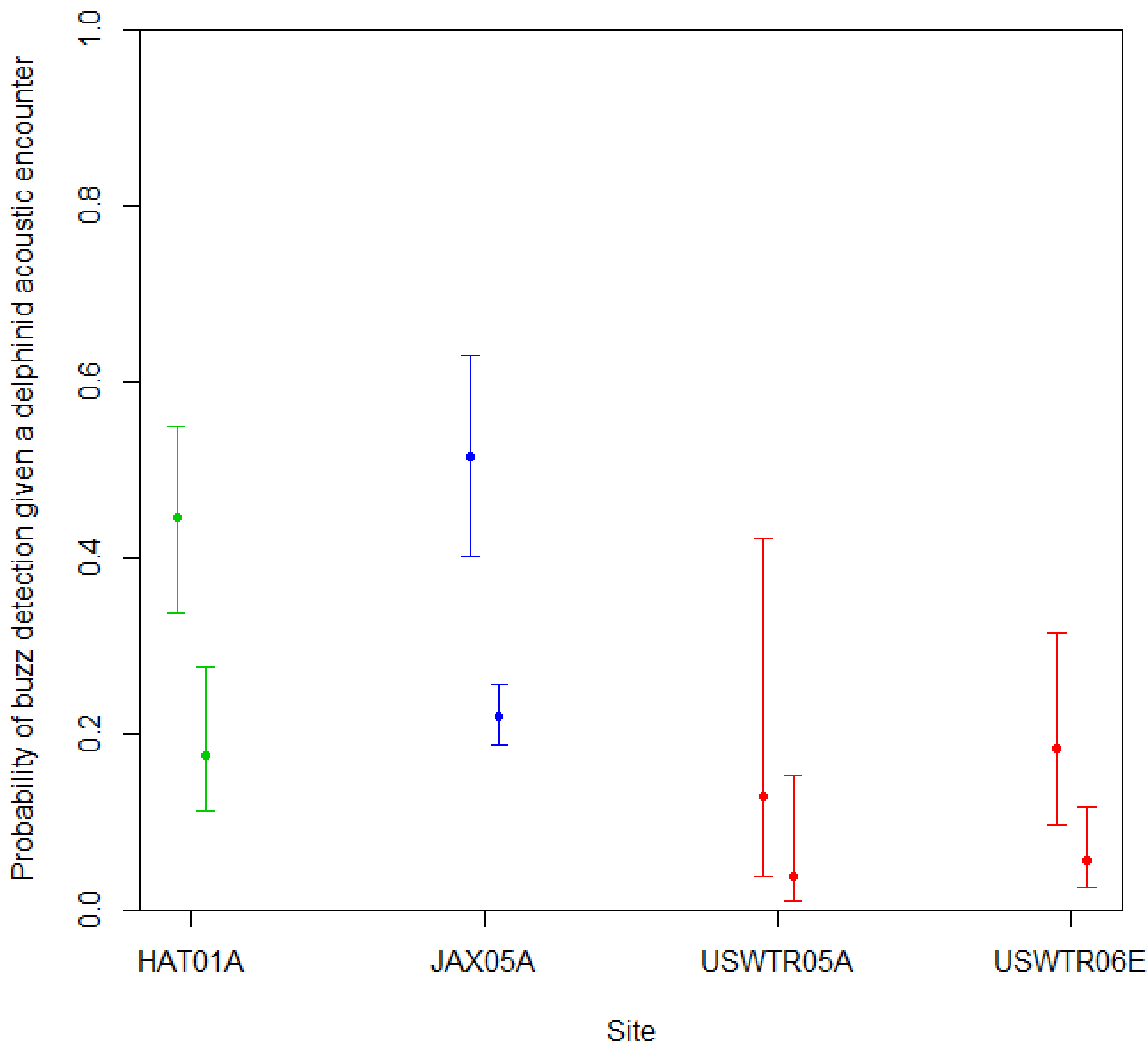

Figure 26. Predicted probability of buzz detection given delphinid acoustic activity in a model (PSTGAE model 6) with *Site*, *Clicks*, *Whistles*, *Type 3 med* sonar signals. Vertical bars indicate 99% confidence intervals. For each site, predictions are shown for *Whistles* and *Clicks* = 1 (assumed present).

4.4.1.4 WHISTLE CHARACTERISTIC MODELS

In the case of this analysis, delphinid acoustic encounters could be divided into pilot whales and other delphinids. However, as there were only 58 pilot whale acoustic encounters in the HARP data, there were not enough data to create a Mahalanobis distance data set. In addition, there were six categories of time periods in relation to MFA sonar events ("before," "during," "between," "during/between," and "during/after", "after"); the latter two categories were reclassified as "during". HARP (n=1734) and MARU (n=2234) data were combined for this analysis. No additional variables other than *Site* and *Sonar* were available for this analysis.

As mentioned above, the control ("before") period data set for referencing the Mahalanobis distances was taken from the "before" periods of the relevant data set (MARU or HARP)

Model fitting was by means of a GEE, assuming a Gamma error function with an inverse link function because of heteroscedasticity in the residuals from a model with an identity link. The final model (**Figure 27**) consisted of *Site* (*p*<0.001) and *Sonar* (*p*=0.002). Model diagnostics were not perfect for this model (see **Appendix C,** Figure 9) with slight heteroscedasticity in the residuals even with the use of the inverse link function.

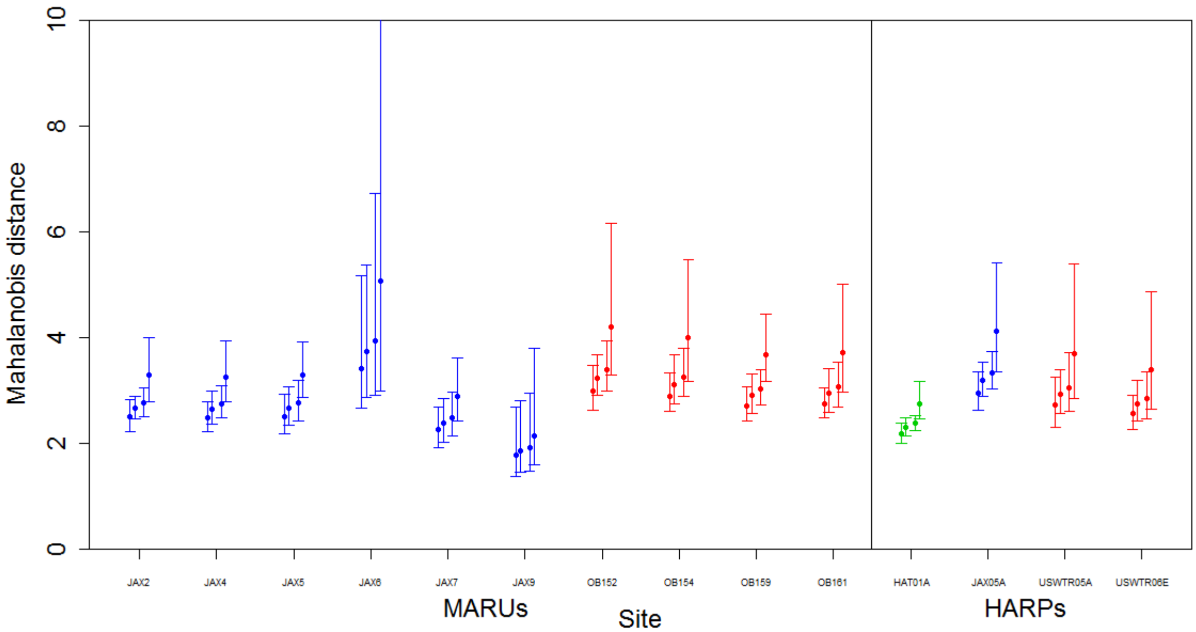

Figure 27. Predicted mean Mahalanobis distance given with *Site* and *Sonar* MFA sonar variables. Jacksonville data in blue, Onslow Bay in red and the Cape Hatteras (HAT01A) in green. For each site, predictions are for "before," "during," "between" and "after" (left to right) MFA sonar events assuming no *Type 3 long* signal and the presence of whistles and clicks. Bars indicate 99% confidence intervals.

4.4.2 Beaked whales

Beaked whales occurred only at the HARP sites, so there was no contribution from the MARU data. After removal of gaps in the data due to duty cycling, there were 21,581 minutes of recording effort (3,322 before, 3,560 during, 11,342 between and 3,557 after). A total of 9.3 percent of these minutes had beaked whale acoustic detections. Temporal independence was achieved after 53 minutes in the non-duty cycled data (see also **Figure 8**).

The final fitted model (PA model 3) consisted of *Site* (*p*-<0.001), *Sonar* (*p*=0.006), and *Timeofday* (fitted as a smooth with 3 degrees of freedom, *p*=0.005). Predicted mean values and 99 percent confidence intervals are given in **Figure 28**. Diagnostic plots are given in **Appendix C**, Figure 8. The model predicted the correct observations 71 percent of the time.

The only specific component of signal type that could be associated with detecting acoustic beaked whale presence was *Type 2 long* signals in a model with *Site* and *Timeofday* (**Figure 29**, PA model 4). Diagnostic plots are given in **Appendix C,** Figure 9. The model predicted 68 percent of the observations correctly. The presence of *Type 2 long signals* was associated with lower probability of presence of beaked whale clicks.

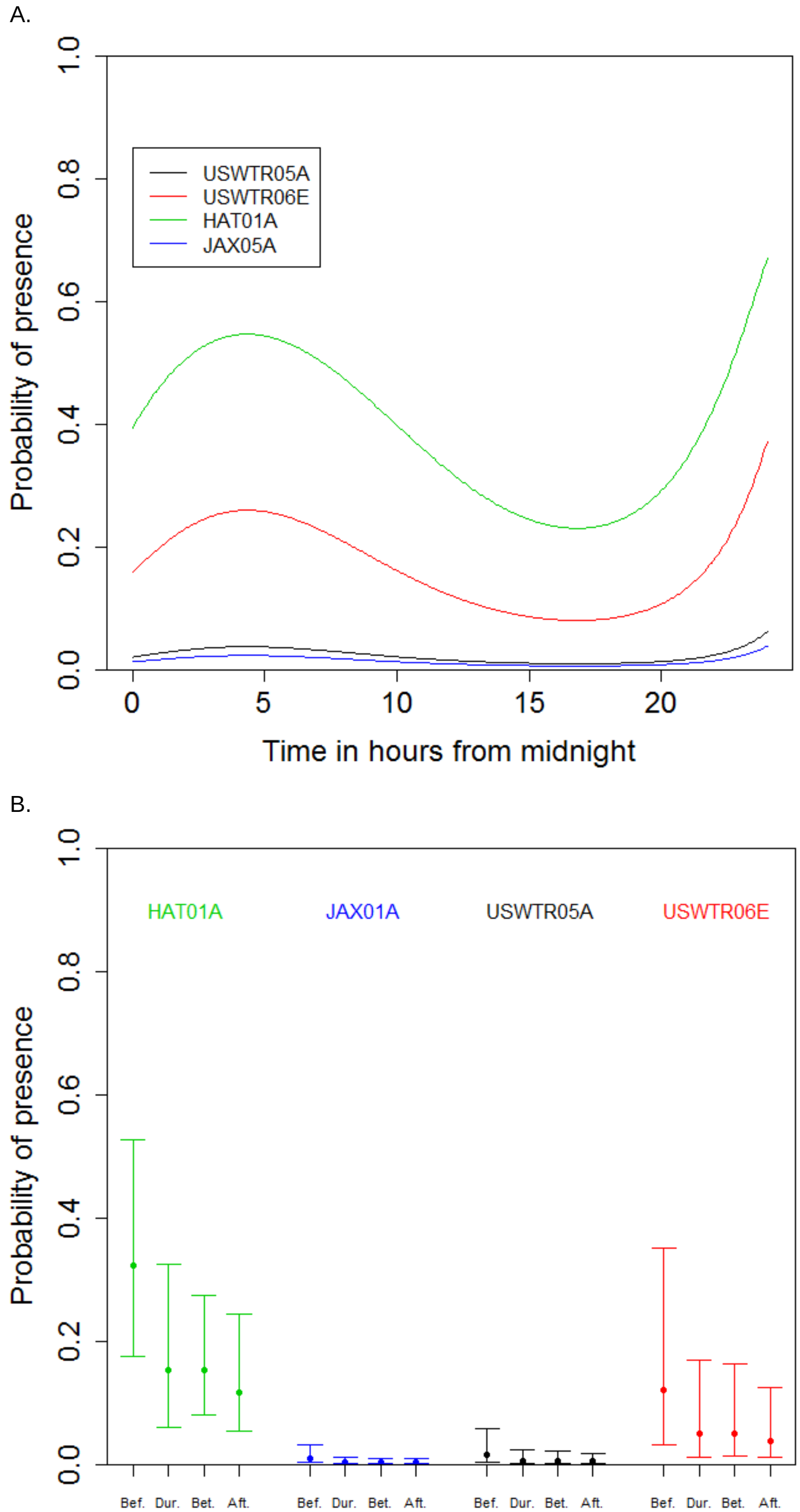

Figure 28. Predicted probability of detecting presence of beaked whale clicks from HARPs from four sites. Cape Hatteras (HAT01A) site (black), the Onslow Bay sites (USWTR05A & USWTR06E) and the Jacksonville (JAX01A) site for: A) time effect "before" MFA sonar events (for clarity, confidence intervals are not shown); and B) "before," "during," "between" and "after" MFA sonar events assuming the time of day is noon. Vertical lines indicate 99% confidence intervals on the predictions.

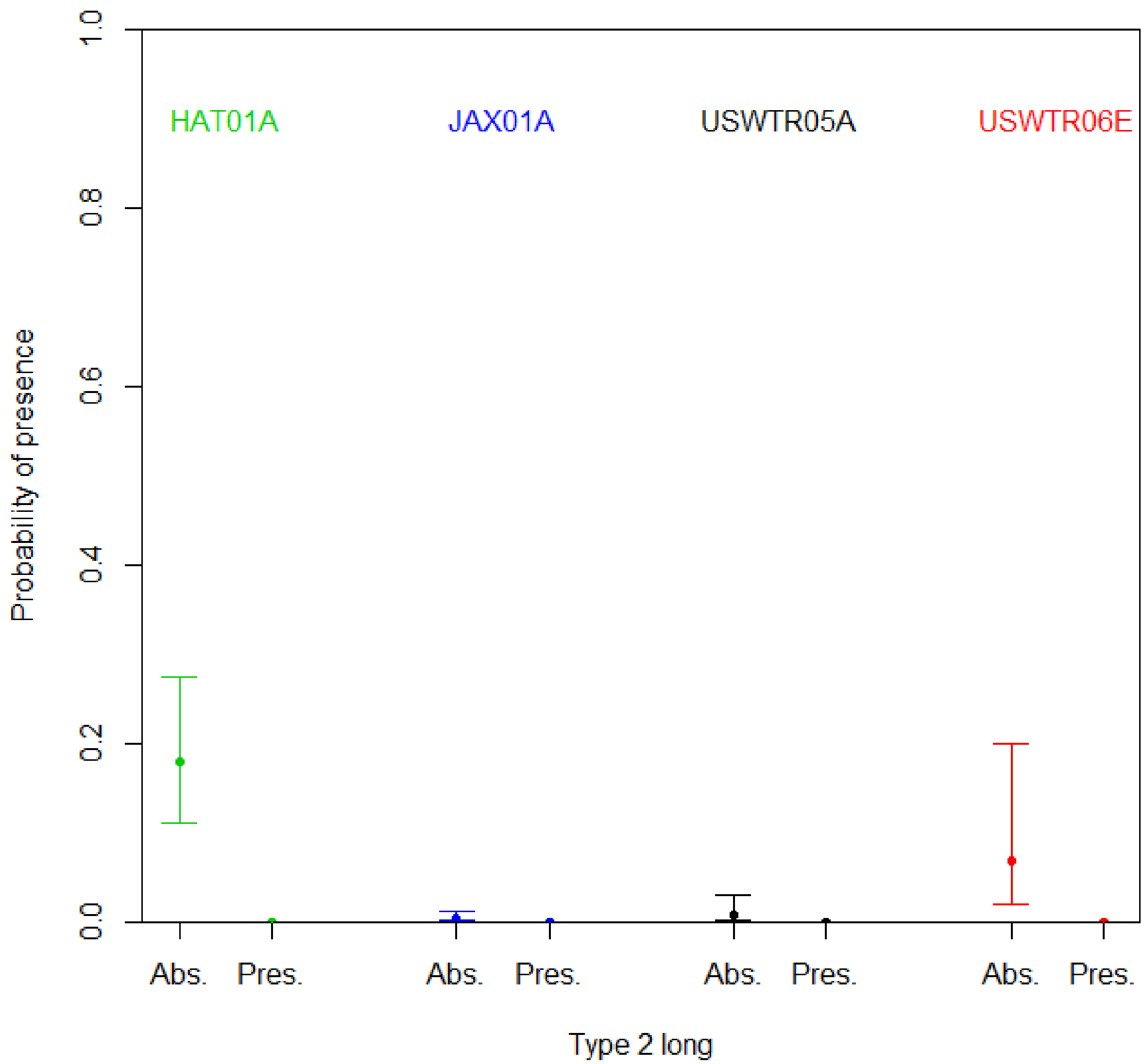

Figure 29. Predicted probability of detecting presence of beaked whale clicks from HARPs from four sites. Cape Hatteras (HAT01A) site (black), the Onslow Bay sites (USWTR05A & USWTR06E) and the Jacksonville (JAX01A) site showing the probability of detected presence in response to the presence of *Type 2 long* signals assuming the time of day is noon. Vertical lines indicate 99% confidence intervals on the predictions.

4.4.3 Minke whales

4.4.3.1 MODELLING PRESENCE-ABSENCE

For this species, data were available from both the MARU and HARP recordings. After removal of gaps in the data due to duty cycling, there were 19,006 minutes of recording effort (10,187 HARP; 8,819 MARU; 5,040 before; 5,040 during; 7,133 between; and 1,793 after). A total of 13 percent of these minutes of effort recorded minke whale presence. Temporal independence was achieved after 836 minutes in the non-duty cycled data. Only three sites were represented in these data JAX2 (MARU), HAT01A, and USWTR05A.

The final probability of presence model (PA model 5) consisted of *SurveyType* and *Sonar* (**Figure 30**,"during" coefficient=1.524, SE=0.309). No minke whales were detected at *USWTR05A* prior to sonar activity. Diagnostics are given in **Appendix C,** Figure 11. The model predicted 59 percent of outcomes correctly. The block size was 836.

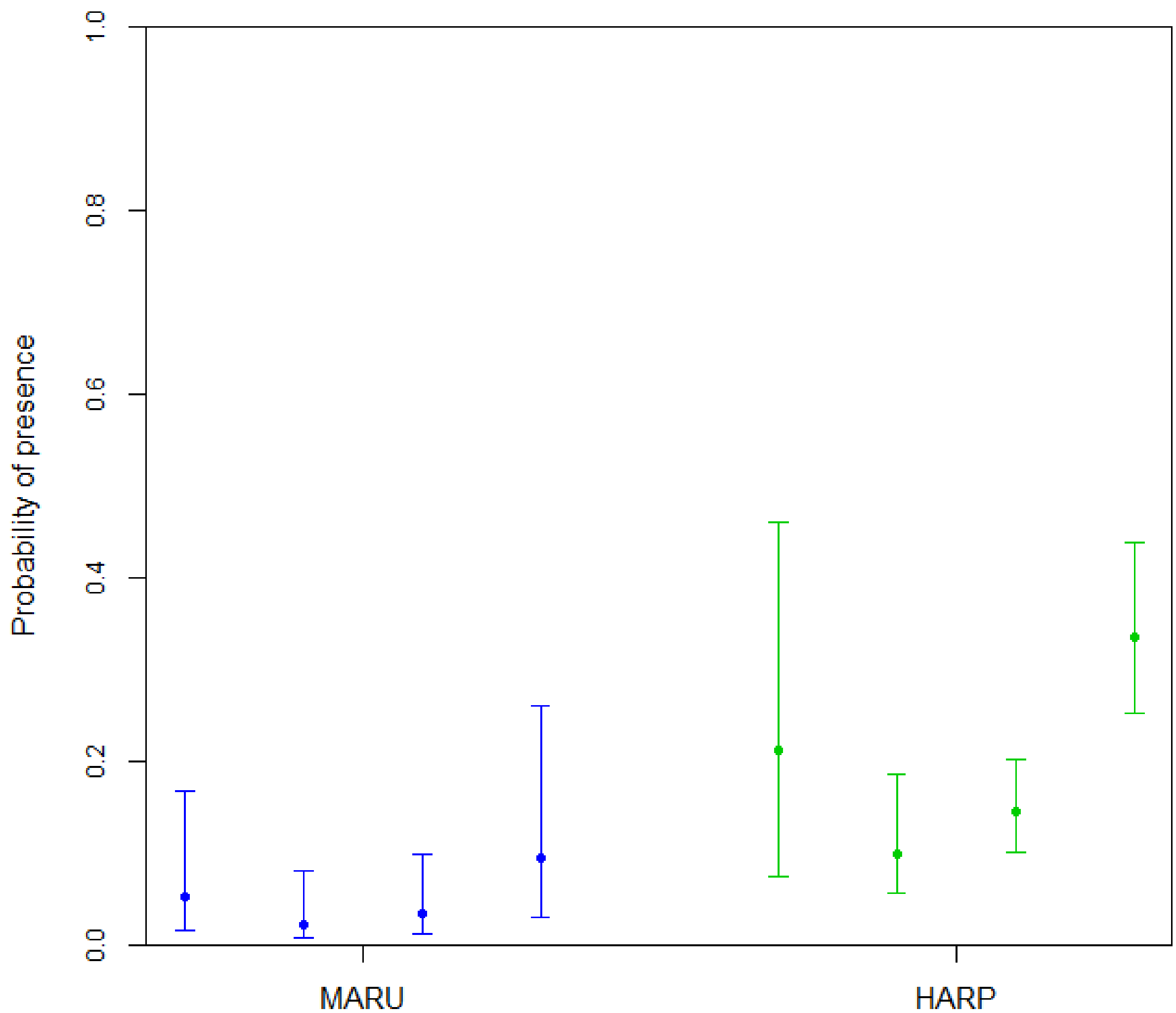

Figure 30. Predicted probability of detecting presence of minke whale pulse-trains at noon from the HARP and MARU data. Predictions are "before," "during," "between" and "after" (left to right) MFA sonar events for each data type. Vertical lines indicate 99% confidence intervals on the predictions.

4.4.3.2 MODELLING DURATION OF CALL

There were 415 MARU and 1,212 HARP minke whale pulse train duration data values. Minke pulse train durations ranged from 5 to 197 seconds. The data were fitted to a GEE, assuming a Gamma error structure and an inverse link function. The best-fit model consisted of *Site* (p<0.001) (**Figure 31**). The block size was 13. Diagnostics are given in **Appendix C,** Figure 12. There was some heteroscedasticity in the fits residual plot, although the widest apparent spread of the data was still at middling fitted values. The extra MFA sonar components of **Table 4** were not available for all of this dataset, so the effect of MFA sonar components could not be considered.

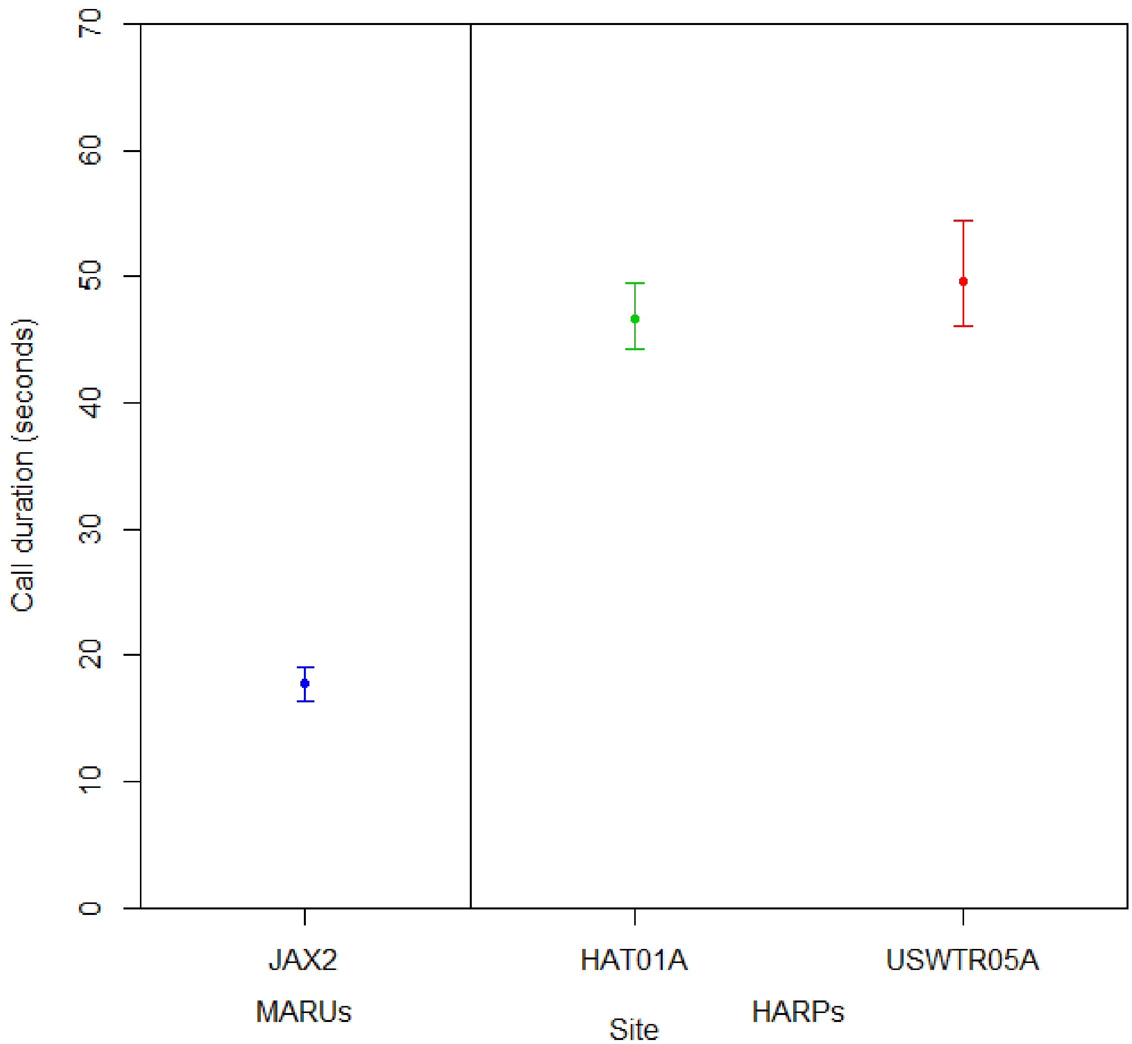

Figure 31. Predicted duration of minke whale pulse trains from a model with *Site*: Jacksonville MARU (blue), Cape Hatteras (HAT01A) HARP (green) and Onslow Bay HARP (red). Vertical lines indicate 99% confidence intervals on the predictions.

4.4.4 Sperm whales

For this species, data were available from both the MARU and HARP recordings, although common MFA sonar descriptor data were not available across both datasets because of the use of different detectors. After removal of gaps in the data due to duty cycling and removal of data from site USWTR05A with only two detected presences, there was a total of 41,983 minutes of recording effort (10,333 before, 5,850 during, 16,420 between, and 9,360 after). A total of 25.2 percent of these 1-minute effort segments had sperm whale acoustic presence. Temporal independence was achieved after 470 minutes in the non-duty cycled data.

The final fitted model (**PA model 6**) consisted of *Daynight.* No significant effect of *Sonar* was found. Predicted mean values and 99 percent confidence intervals are given in **Figure 32.** Diagnostics are given in **Appendix C,** Figure 13. With only 62 percent of correct model predictions, the fit of the model was not very much above 50 percent which is what is expected by chance.

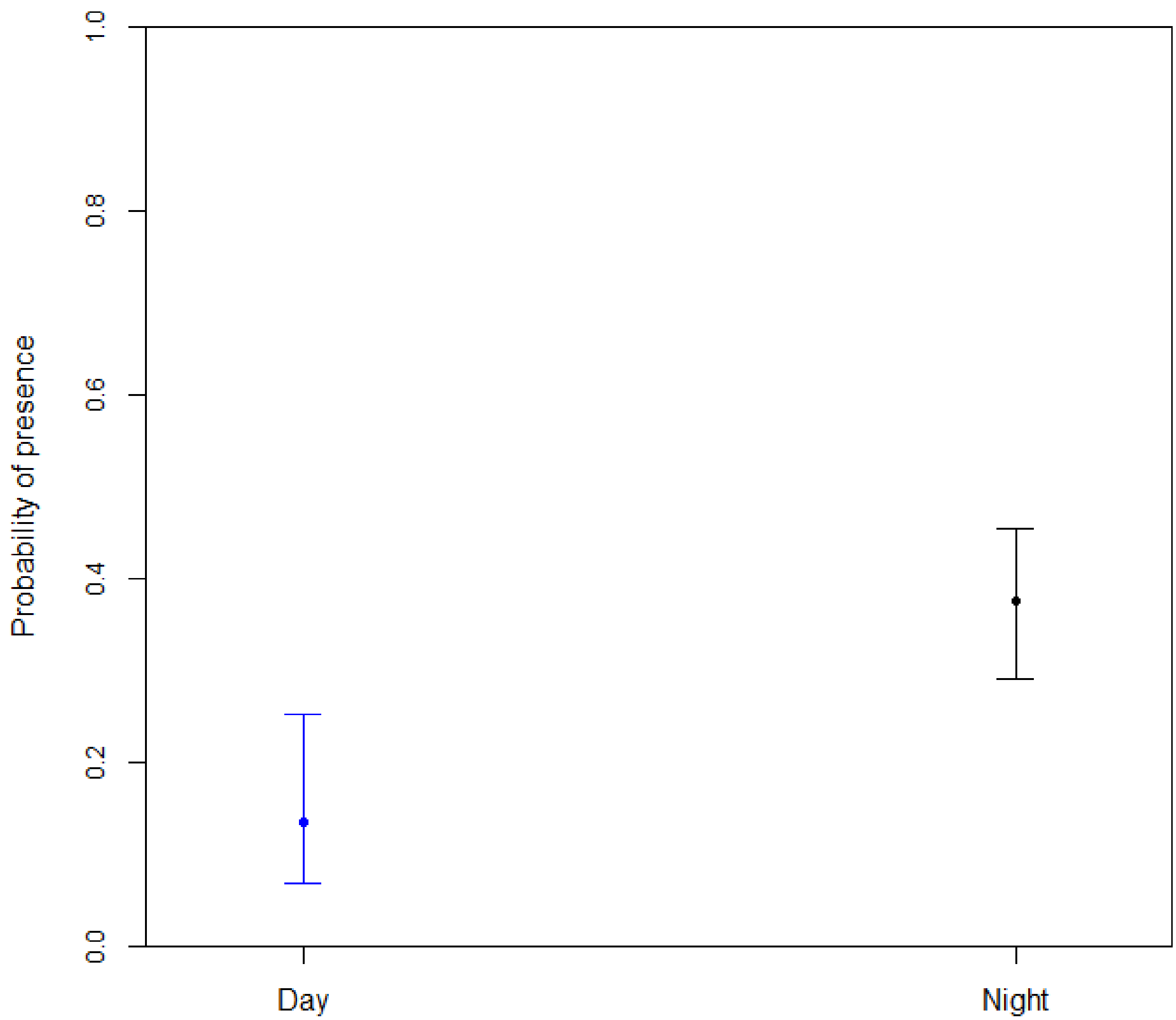

Figure 32. Predicted probability of detecting presence of sperm whale clicks from the HARP/MARU combined data from a model with covariate *Daynight*.

4.4.5 Presence-absence modelling versus HM modelling

The GEE models are not directly comparable to the HMM results, as the former combine both MARU and HARP records from all sites, whereas the HMMs only included HARP data and were run separately for each site. To provide a direct comparison, a new set of GEE models were fitted for each taxa for each HARP site and modelled with *Sonar* as a predictor with error families and link functions and models as described above. However now a *Sonar* only model and a null model (i.e., a model fitting a mean probability of presence only, approximately equivalent to the baseline HMMs) were compared using the quasi-likelihood under the independence model criterion of Pan (2001) as the nearest equivalent to AIC the model selection criterion used in the HMMs.

When a GEE could be fitted per site, a model containing *Sonar* was always selected (**Table 13**) suggesting GEEs may be more powerful than the current implementation of HMMs for detecting effects on cetacean acoustic presences, albeit less versatile.

Table 13. Comparison of HARP-only model selection on the presence-absence data between GEEs and HMMs.

Taxa	Site ID	GEE model with lowest QIC*	HMM with lowest AIC (see Table 12)
Delphinids	Jacksonville (JAX05A)	*Sonar*	*Baseline*
	Cape Hatteras (HAT01A)	*Sonar*	*Sonar*
	Onslow Bay 1 (USWTR05A)	*Sonar*	*Sonar*
	Onslow Bay 2 (USWTR06E)	*Sonar*	*Baseline*
Beaked whales	Jacksonville (JAX05A)	*Could not be fitted*	*Baseline*
	Cape Hatteras (HAT01A)	*Sonar*	*Baseline*
	Onslow Bay 2 (USWTR06E)	*Sonar*	*Baseline*
	Onslow Bay 2 (USWTR06E)	*Sonar*	*Baseline*
Minke whales	Cape Hatteras (HAT01A)	*Sonar*	*Sonar*
	Onslow Bay 1 (USWTR05A)	*Could not be fitted*	*Sonar*
Sperm whales	Cape Hatteras (HAT01A)	*Sonar*	*Sonar*
	Onslow Bay 2 (USWTR06E)	*Sonar*	*Baseline*

* After Pan (2001) using the QIC function in the R library MESS (Eckstom 2014).

5. Discussion

5.1 General comments on statistical methods

5.1.1 The dependent data

The dependent data in these analyses cannot be unambiguously related to the behavior, status, or acoustic presence of the animals. A decrease in the probability of detecting acoustic signals could result from decreased acoustic activity by animals that are present, the absence of animals from the region, a change in an animal's orientation relative to the sensor, or some other unknown behavioral or environmental factor. Similarly, a recorded duration may reflect, in part, changes in the orientation of the animal to the sensor. Likewise, features of acoustic signals for example may be compromised by environmental variables.

5.1.2 Regression analyses

These data were analyzed in a robust manner to reduce the risk of Type I errors (rejecting the null hypothesis when it ought to be accepted, i.e. seeing an effect where there is not one) caused by the lack of independence between adjacent segments in time. Unexplained serial correlation was dealt with by the use of GEEs, an alpha (the Type 1 error rate) of 0.01 rather than 0.05 was used as a model-selection inclusion criterion, and there was graphical inspection of the relationships of the data. Further work could consider the use of family-wide error adjustment on alpha to reflect the large number of tests being undertaken.

Variance inflation factors allow for confounding of variables to be recognized, which is important for making the correct inference about the models. Nevertheless, care should be taken in inferring the correct effect of *Sonar*, for example, given the presence of other variables that may be correlated with the dependent variable.

5.1.3 Hidden Markov Models

The conclusions of our HMM analysis could be strengthened if data were not duty cycled and longer sample sizes for a given time series become available. Further improvement could be achieved by accounting in a more systematic way for the difference between "no acoustic occurrence" due to presence of an animal that is not acoustically active (i.e., silent) and "no acoustic occurrence" due to absence of the species in the study area being monitored. However, in practice it may be difficult to distinguish between these two possibilities. Despite these limitations, the analysis of cetacean acoustic signals using HMMs has provided important insights into their behaviour and reaction to MFA sonar exposure. Specifically, we found evidence for a MFA sonar effect on the acoustic detections for three out of the four taxa considered. For these three taxa, there was always an effect at HAT01A, which was the site without duty cycling.

5.2 Dolphins

5.2.1 Regression analysis of presence-absence data

A direct influence of MFA sonar time period on delphinids was not found, but the presence of a *Type 1 long* MFA sonar signal paradoxically resulted in a slightly increased probability of delphinid acoustic detection. The presence of a *Site* effect in these models potentially covers a large number of sources of variation: geography, seasonality, and both individual and recorder type (i.e., MARU or HARP) effects, and is often by far the biggest contributor to the variation (e.g., **Figures 16 and 17**). No consistent patterns for *Site* across both sets of data were apparent except that HARP recordings contained more detections than MARUs. This difference could be caused by different sample rates, instrument self-noise, or frequency responses of the instrumentation. The HARP sampling rate was higher than the MARU sampling rate (200 kHz vs. 32 kHz, respectively). Many of the sounds produced by dolphins contain much of their energy above 16 kHz, which would make them difficult or impossible to detect in the MARU recordings. However, as the HARPs and MARUs were deployed at different times and in different locations, it is not possible to compare them directly to determine the extent to which recording bandwidth affected the observed differences in delphinid acoustic detection rates.

5.2.2 Hidden Markov Models

The model-selection criteria did not suggest that a consistent model was suitable for all sites for delphinids, although an MFA sonar effect was found at HAT01A and USWTR05A, a pattern replicated, in part, by results for minke whales and sperm whales. This raises the question of whether the MFA sonar at these two sites is in any way different from the other sites and requires further investigation.

5.2.3 *Presence of signal type given cetacean acoustic encounter* (PSTGAE) models

The occurrences of delphinid whistles and clicks were negatively correlated with each other in this analysis (**Figure 19**), whereas occurrences of whistles and buzzes were positively correlated (**Figure 20**), as were, paradoxically, occurrences of clicks and buzzes (**Figure 23**). As in the case of overall occurrence, there was a strong *Site* effect. *Sonar* had an effect of an increased probability of detecting whistles, buzzes, and clicks (given acoustic presence) “during” and “between” MFA sonar events.

Again, there was an association of cetacean signal type with some of the components of MFA sonar, with elevated presences of whistles, clicks and buzzes in response to the presence of long MFA sonar signals.

These results are consistent with previous examinations of delphinid responses to anthropogenic noise from other studies. For example, DeRuiter et al. (2013a) found that false killer whales increased whistle production rates after MFA sonar exposure. Similarly, Rendell and Gordon (1999) reported increased whistle rates from long-finned pilot whales during post-exposure periods. Additionally, bottlenose dolphin whistle rates have been reported to increase and whistle modulation to decrease in response to stress (Caldwell et al. 1970, Esch et al. 2009), during increased ambient noise (Morisaka et al. 2005), and during vessel approaches (Buckstaff 2004). Increasing vocal activity in response to an acoustic event may represent an alert/startle response and/or an attempt to overcome the effects of masking in active acoustic space.

5.2.4 Whistle-characteristics models

The results of the whistle-characteristic models were interesting in that while there was a *Site* effect, it was less marked than in other analyses. The *Sonar* effect was also interesting in that—compared to “before” MFA sonar—Mahalanobis distances increased in the “during,” “between,” and “after” time periods, indicating increased variability in the whistle parameters.

Mahalanobis distance provides a useful method for characterizing the overall whistle at different points in time and hence acts as a useful initial consideration of the data. However, it might be useful to follow up a significant influence on Mahalanobis distance with an investigation of the specific features of the whistles affected by *Sonar.*

5.3 Beaked whales

Acoustic detections of beaked whales declined after MFA sonar activity commenced (**Figure 29b**). This result is not surprising, given the known sensitivity of this taxon to MFA sonar and other types of acoustic events (Tyack et al. 2011; DeRuiter et al. 2013b, Stimpert et al. 2014). Again, it appears that the occurrence of the long MFA sonar signal resulted in decreased acoustic detections. Previous work during controlled exposure experiments has shown observed behavioral (vigorous swimming, directed travel, changes in dive behavior) and acoustic responses (cessation of echolocation activity) from multiple beaked whale species, including Cuvier’s (*Ziphius cavirostris)* (DeRuiter et al. 2013b), Blainville’s (*Mesoplodon densirostris*) (Tyack et al. 2011), and Baird’s (*Berardius bairdii*) (Stimpert et al. 2014) beaked whales. There was also a marked diurnal pattern in acoustic detections in this taxon (**Figure 29a**), showing a decrease in vocal activity between 10:00 and 20:00 hours, as well as a site effect (**Figure 29b**), with the deeper Hatteras site recording more detections, which is in agreement with the observational data (McLellan et al. in prep.) for the region.

No effect of MFA sonar on beaked whales was detected at any site using the HMM method despite the known sensitivity of beaked whales (see above).

5.4 Minke whales

Similar to the patterns found for beaked whales, minke whales also showed a decline in acoustic detections once MFA sonar activity started, but there was a recovery thereafter. There was an effect of *SurveyType* as well, but this could easily have been a *Site* effect because the variables are confounded.

The hidden Markov models detected an MFA sonar effect at both HARP sites where minke whales occurred. We also found that minke whales at HAT01A tend to spend more time in the “silent” state both during and between MFA sonar exposures. However, there was no MFA sonar effect on minke pulse train duration, merely a difference in durations between sites. These results are consistent with other observations of minke whale response to naval sonar. For example, progressive aversion of minke whales to sonar playback has been shown, whereby the tagged animal increased speed and changed dive patterns to move away from the sound source (Sivle et al. 2015). Additionally, the density of minke whales, as evidenced by pulse train detections, has been shown to decrease in response to naval sonar activity (Martin et al. 2015). Opportunistic observations of minke whale displacement and avoidance in response to MFA sonar have also been noted (Dolman et al. 2011, Parsons et al. 2000).

5.5 Sperm whales

A marked change in the probability of detecting sperm whale sounds between day and night is the only characteristic of this species revealed by this analysis. The significantly higher probability of detecting sperm whale clicks at night is in line with previous studies, which have shown that sperm whales in the northwestern Atlantic exhibit a diel foraging behavior pattern (Hodge et al. 2013, Yack et al. 2016). No effect of *Sonar* or even *Site* was found. The presence-absence results are consistent with previous studies that did not detect changes in sperm whale foraging behavior in response to MFA sonar (Isojunno et al. 2016, Sivle et al. 2012).

In slight disagreement with the presence-absence modelling, the hidden Markov models detected an MFA sonar effect in the HARP data at the Cape Hatteras site, but not at the USWTR06E site.

5.6 Future Work

Recordings from three of the four HARPs were duty-cycled. This created several issues for fitting the models. We used control periods of 24 hours before the first MFA sonar ping from each MFA sonar exercise to compare with cetacean acoustic signals that were considered unaffected by MFA sonar with those from periods during, between, and after exposure to MFA sonar. When analyzing duty-cycled data, it was impossible to determine whether the first MFA sonar ping detected for each exercise was actually the beginning of the exercise, or whether the exercise began when the recorder was off. If data are not duty-cycled, the only uncertainty is in whether the detector detected it. When data are duty-cycled, the uncertainty increases as the first (and consecutive) pings may have occurred when the recorder was not recording. In the worst-case scenario, all of the 1-minute segments that are labelled as “before” from a given exercise could be labelled in error if sonar pings were missed at the beginning of the 24-hour period. This has further implications for calculating the Mahalanobis distances as the whistles from the “before” periods are used as the control data against which all other whistles are compared. Similar issues arise when determining the last MFA sonar ping of an exercise and labelling the segments belonging to the 24 hours “after.” A simulation study could address the severity of this issue; however, as there are many data sets potentially available without duty cycling, a simulation study should not be a high priority.

Further issues arise from duty cycling. When fitting GEEs, we used a correlation structure that assumed consecutive observations were correlated. For the presence of acoustic activity models, for example, we assumed that the correlation decreased with increasing number of 1-minute segments that lie between two records and used a block structure for modelling this correlation. With missing observations due to duty cycling, the uncertainty in determining the size of these blocks increases, potentially affecting inference on the model parameters.

For fitting HMMs, uninterrupted time series data are required. A way to accommodate missing observations (here, the 1-minute segments) is by filling in the missing time points in the sequence and assigning NAs to the response variable. However, when including covariates in the model, no Ns are allowed for the observed values of these covariates. Hence, we needed to add the extra step of imputing these covariate values, adding further uncertainty to the possible inference of these models.

Due to these issues arising from duty cycling, we recommend that when using these methods, data should not be duty-cycled.

Analysis of the HARP data has confirmed that the response variables were strongly site dependent (i.e., covariate *Site* was almost always in the final model). One exception was the presence of acoustic signal models fitted to the 1-minute segment data for sperm whales and minke whales (although for minke whales, *Surveytype* was in the final model). Covariate *Site* was in all final models for the MARU data for pilot whales and for the remaining delphinids combined, but not in the final model for sperm whales. Due to this site dependency, we fitted separate HMMs to the time series at each site. The disadvantage of fitting separate HMMs for each site is that this drastically reduces the number of times that the observed state switches from one state to the other. However, fitting one HMM to data from all sites simultaneously would require developing new models where some parameters are allowed to vary between sites while others are assumed constant between sites. These site-specific parameters can either be fitted as fixed or random effects. However, the development of these methods would require more time than the current project allows. Computational issues would likely need to be overcome as the HMM analysis machinery can be very slow when applied to large datasets.

A major focus of future work should be in testing our methods via a simulation study. With such a study, we could assess the efficacy of our model selection strategy. Model selection for GEEs is an ongoing area of research where no clear strategy has emerged. The discrepancy in best-fitting models between the GEE and the HMM approaches in this report highlight the importance of this issue.

An alternative strategy to both our GEEs and HMMs is to use occupancy modelling, which attempts to distinguish between animal presence and detection given presence. Initial work in this area has been undertaken by Mevin Hooten and colleagues at Colorado State University (Hooten, pers. comm.). Further investigation of this approach may prove useful.

Future work should also focus on working towards a unified data-processing protocol, both for cetacean data and sonar data. Discrepancies in the cetacean data existed in that, for the HARP data, delphinid detections were provided as sub-encounters (1-minute counts of signals), while for the MARU data, delphinid detections were provided as sub-encounters with varying length (see Oswald et al. 2015 for the definition of sub-encounters). In short, sub-encounters included time spans of consecutive detections with no gap longer than 1 minute. Hence, the length of these varied between 1 and 7,279 seconds. On the other hand, the sub-encounters logged for the HARP data were all of 1-minute length. This difference had implications, in particular for the PSTGAE analyses. In comparison, we regard the sub-encounters as an improvement, as they have consistent time lengths and provide more information compared to the sub-encounter. This information included click counts and whistle counts, which provide another opportunity for further analyses.

In the analyses of the effects of sonar on delphinid acoustic behavior, all small delphinid species were examined together as a "delphinid" species group. It is possible, and even likely, that species-specific differences exist in responses to anthropogenic sounds. To examine this more closely, additional data should be included to provide sufficient sample sizes for individual species.

Lastly, future work should entail identifying the most important covariates that should be tested as potential predictors in the models.

5.7 Summary and Conclusion

This study has resulted in the development of innovative statistical methods for assessing the potential impacts of naval MFA sonar on marine mammal acoustic behavior. These methods can be applied to other types of anthropogenic acoustic datasets in other geographic regions. The work herein has highlighted some species-specific responses to MFA sonar in the northwestern Atlantic. Our results indicate that for the sonar activities, regions, and populations of marine mammals included in the analysis, delphinids may increase acoustic activity in response to MFA sonar (Type 1 long component, **Figure 17**). Beaked whales cease acoustic activity once MFA sonar begins and show no evidence of a recovery within 24 hours after a sonar exercise (**Figure 28**); minke whales cease acoustic activity during and after sonar but show some evidence of recovery within 24 hours (**Figure 30**); and sperm whales show no response to MFA sonar. The results increase our knowledge of the responses to MFA sonar for the taxa included in our study. Although responses were not consistent across species or in some cases were subtle or limited, there were clear effects for several species, which imply at a minimum, cessation of important activities (e.g. courtship and breeding for minke whales, and foraging for beaked whales). These results can be used, with the caveats we provided, to better inform species-specific management and mitigation strategies, especially in the surveyed area.

6. Acknowledgements

We would like to thank Shannon Coates, Kerry Dunleavy, and Cory Hom-Weaver for many hours spent analyzing whistle and sonar data and to Gabriela Alongi for her help with R programming. Thank you to Andrew Read and Lynne Hodge for providing HARP data and to Russ Charif for providing MARU data. We thank US Fleet Forces Command for providing funding for this analysis under the US Navy's Marine Species Monitoring Program. Project management and technical review was provided by NAVFAC Atlantic (NAVFAC LANT). We are especially grateful for the logistic support and advice from Joel Bell (NAVFAC LANT) and from Dan Engelhaupt and Michael Richlen (HDR). Finally, we would like to thank Chris Clark and Robert Kenney for their very thorough review of the report.

This page intentionally left blank.

7. References

Baumgartner, M.F., L. Freitag, J. Partan, K.R. Ball, and P.E. Kenneth. 2008. Tracking large marine predators in three dimensions: The Real-Time Acoustic Tracking System. *IEEE Journal of Oceanic Engineering* 33(2):146–157.

Bio-Waves, Inc. 2013. SonarFinder Users Manual. Submitted to HDR, Inc., MSA # CON-005-4394-009, Subproject # 164744-KB10. Bio-Waves, Inc., Encinitas, California. 30 pp.

Buckstaff, K.C. 2004. Effects of watercraft noise on the acoustic behavior of bottlenose dolphins, *Tursiops truncatus*, in Sarasota Bay, Florida. *Marine Mammal Science* 20:709–725.

Caldwell, M.C., D.K. Caldwell, and R.H. Turner. 1970. Statistical analysis of the signature whistle of an Atlantic bottlenosed dolphin with correlations between vocal changes and level of arousal. Technical Report no. 8. Los Angeles County Museum of Natural History Foundation, Los Angeles, California. 31 pp.

Castellote, M., C.W. Clark, and M.O. Lammers. 2012. Acoustic and behavioral changes by fin whales (*Balaenoptera physalus*) in response to shipping and airgun noise. *Biological Conservation* 147(1):115–122.

Charif, R.A., C.W. Clark, and K.M. Fristrup. 2004. Raven 1.2 User's Manual. Cornell Laboratory of Ornithology, Ithaca, NY.

Charif, R.A., C.S. Oedekovan, A. Rahaman, B.J. Estabrook, L. Thomas, and A.N. Rice. 2015. Development of Statistical Methods for Assessing Changes in Whale Vocal Behaviour in Response in Mid-frequency Active Sonar. Final Report. Prepared for U.S. Fleet Forces Command. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia. Under contract N. N62470-10-3011, Task Order 39, issued to HDR Inc., Virginia Beach, Virginia, 20 March 2015.

Clark, C.W., and G.C. Gagnon. 2002. Low-frequency vocal behaviors of baleen whales in the North Atlantic: Insights from IUSS detections, locations and tracking from 1992 to 1996. *Journal of Underwater Acoustics (USN)* 52(3):609-640. Unclassified publication.

Clark, C.W., J.F. Borsani, and G. Notarbartolo-di-Sciara. 2002. Vocal activity of fin whales, *Balaenoptera physalus*, in the Ligurian Sea. *Marine Mammal Science* 18(1):286–295.

Clark, C.W., and P.J. Clapham. 2004. Acoustic monitoring on a humpback whale (*Megaptera novaeangliae*) feeding ground shows continual singing into late spring. *Proceedings of the Royal Society of London, Series B: Biological Sciences* 271:1,051–1,057.

Clark, C.W., and N.S. Altman. 2006. Acoustic detections of blue whale (*Balaenoptera musculus*) and fin whale (*B. physalus*) sounds during a SURTASS LFA exercise. *IEEE Journal of Oceanic Engineering* 31:120-128.

Davison, A.C. & Hinkley D.V. (1997) *Bootstrap Methods and their Application* (Cambridge Series in Statistical and Probabilistic Mathematics). Cambridge University Press. Cambridge. 592pp.

DeRuiter, S.L., I.L. Boyd, D.E. Claridge, C.W. Clark, B.L. Southall, and P.L. Tyack. 2013a. Delphinid whistle production and call matching during playback of simulated military sonar. *Marine Mammal Science* 29:46–59*.*

DeRuiter, S.L., B.L. Southall, J. Calambokidis, W.M. Zimmer, D. Sadykova, E.A. Falcone, A.S. Friedlaender, J.E. Joseph, D. Moretti, G.S. Schorr, L.Thomas, and P.L. Tyack. 2013b. First direct measurements of behavioural responses by Cuvier's beaked whales to mid-frequency active sonar. *Biology Letters* 9(4): 2013.0223.

Di Iorio, L., and C.W. Clark. 2010. Exposure to seismic survey alters blue whale acoustic communication. *Biology Letters* 6(1): 51–54.

Dolman, S.J., P.G. Evans, G. Notarbartolo-di-Sciara, and H. Frisch. 2011. Active sonar, beaked whales and European regional policy. *Marine Pollution Bulletin* 63: 27–34.

Eckstom, C. 2014. *MESS: Miscellaneous Esoteric Statistical Scripts* package (for R).

Esch, H.C., L.S. Sayigh, J.E. Blum, and R.S. Wells. 2009. Whistles as potential indicators of stress in bottlenose dolphins (*Tursiops truncatus*). *Journal of Mammalogy* 90:638–650.

Fristrup, K.M., L.T. Hatch, and C.W. Clark. 2003. Variation in humpback whale (*Megaptera novaeangliae*) song length in relation to low-frequency sound broadcasts. *Journal of the Acoustical Society of America* 113:3411-3424.

Gillespie, D., J. Gordon, R. McHugh, D. McLaren, D.K. Mellinger, P. Redmond, A. Thode, P. Trinder, and D. Xiao. 2008. PAMGUARD: Semiautomated, open-source software for real-time acoustic detection and localization of cetaceans. *Journal of the Acoustical Society of America* 30(5):54–62.

Hardin, J.W, and J.M. Hilbe. 2003. *Generalized Estimating Equations.* Chapman and Hall, London. 240 pp.

Hodge, L.E., J.T. Bell, A. Kumar, and A.J. Read. 2013 The influence of habitat and time of day on the occurrence of odontocete vocalizations in Onslow By, North Carolina. *Marine Mammal Science* 29(4):E411-E427.

Højsgaard, S., R. Halekoh, and J. Yan. 2016. Package *geepack* manual package (for R).

Højsgaard, R., J. Yan, and S. Halekoh. 2005. The R package GEEPACK for generalized estimating equations. *Journal of Statistical Software* 15(i02) · November 2005

Isojunno, S., C. Curé, P.H. Kvadsheim, F.P.A. Lam, P.L. Tyack, P.J. Wensveen, and P.J.O.M. Miller. 2016. Sperm whales reduce foraging effort during exposure to 1–2 kHz sonar and killer whale sounds. *Ecological Applications* 26(1): 77–93.

Johnston, D.W., M. McDonald, J. Polovina, R. Domokos, S. Wiggins, and J. Hildebrand. 2008. Temporal patterns in the acoustic signals of beaked whales at Cross Seamount. *Biology Letters* 4:208–211.

Martin, S.W., C.R. Martin, B.M. Matsuyama, and E.E. Henderson. 2015. Minke whales (*Balaenoptera acutorostrata*) respond to navy training. *Journal of the Acoustical Society of America* 137(5):2533-2541.

McCarthy, E., D. Moretti, L. Thomas, N. DiMarzio, R. Morrissey, S. Jarvis, J. Ward, A. Izzi, and A. Dilley. 2011. Changes in spatial and temporal distribution and vocal behavior of Blainville’s beaked whales (*Mesoplodon densirostris*) during multi-ship exercises with mid-frequency sonar. *Marine Mammal Science* 27:E206–E226.

McLellan, W.A., McAlarney, R.J., Cummings, E.W., Read, A.J., Paxton, C.G.M., Bell, J.T. and Pabst, D.A. (submitted) Year-round presence of multiple beaked whale species (Family Ziphiidae) off Cape Hatteras, North Carolina, USA. *Journal of Cetacean Research and Management.*

Melcón, M.L., A.J. Cummins, S.M. Kerosky, L.K. Roche, S.M. Wiggins, and J.A. Hildebrand. 2012. Blue whales respond to noise. *PLoS ONE* 7(2):e32681. doi:10.1371/journal.pone.0032681.

Mellinger, D., C. Carson, C. Clark. 2000. Characteristics of minke whale (*Balaenoptera acutorostrata*) pulse trains recorded near Puerto Rico. *Marine Mammal Science* 16: 739−756.

Mellinger, D.K., K.M. Stafford, S.E. Moore, R.P. Dziak, and H. Matsumoto. 2007. An overview of fixed passive acoustic observation *methods* for cetaceans. *Oceanography* 20(4):36–45.

Miller, P.J., N. Biassoni, A. Samuels, and P.L. Tyack. 2000. Whale songs lengthen in response to sonar. *Nature* 405(6789):903–903.

Morisaka, T., M. Shinohara, F. Nakahara, and T. Akamatsu. 2005. Effects of ambient noise on the whistles of Indo–Pacific bottlenose dolphin populations. *Journal of Mammalogy* 86:541–546.

Nieukirk, S.L., K.M. Stafford, D.K. Mellinger, R.P. Dziak, and C.G. Fox. 2004. Low-frequency whale and seismic airgun sounds recorded in the mid-Atlantic Ocean*. Journal of the Acoustical Society of America* 115(1):1,832–1,843.

Oswald, J.N. 2013. Development of a Classifier for the Acoustic Identification of Delphinid Species in the Northwest Atlantic Ocean. Final Report. Submitted to HDR Environmental, Operations and Construction, Inc. Norfolk, Virginia under Contract No. CON005-4394-009, Subproject 164744, Task Order 003, Agreement #105067. Prepared by Bio-Waves, Inc., Encinitas, California.

Oswald, J.N., S. Rankin, J. Barlow, M. Oswald and M.O. Lammers. 2013. *Real-time Call Classification Algorithm (ROCCA): 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.

Oswald, J.N, C.S.S. Oedekoven, T.M. Yack, R. Langrock, L. Thomas, E. Ferguson, and T. Norris. 2015. Development of Statistical Methods for Examining Relationships Between Odontocete Vocal Behaviour and Navy Sonar Signals. Final Report. Prepared for U.S. Fleet Forces Command. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia, under Contract No N62470-10-3011, Task Order 39, issued to HDR Inc. Virginia Beach, Virginia. 30 March 2015.

Pan, W. (2001). Akaike's Information Criterion in Generalized Estimating Equations. *Biometrics,* 57(1):120–125.

Parsons, E.C.M., I. Birks, P.G.H. Evans, J.C.D. Gordon, J.H. Shrimpton, and S. Pooley. 2000. The possible impacts of military activity on cetaceans in west Scotland. *European Research on Cetaceans* 14:185–190.

R Developmental Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rendell, L.E., and J.C.D. Gordon. 1999. Vocal response of long-finned pilot whales (*Globicephala melas*) to military sonar in the Ligurian Sea. *Marine Mammal Science* 15:198–204.

Risch, D., C.W. Clark, P.J. Dugan, M. Popescu, U. Siebert, and S.M. Van Parijs. 2013. Minke whale acoustic behavior and multi-year seasonal and diel vocalization patterns in Massachusetts Bay, USA. *Marine Ecology Progress Series* 489: 279-295.

Risch, D., P.J. Corkeron, W.T. Ellison, and S.M. Van Parijs. 2012. Changes in humpback whale song occurrence in response to an acoustic source 200 km away. *PLoS ONE* 7(1):e29741. doi:10.1371/journal.pone.0029741.

Schliehe-Diecks, S., P.M. Kappeler, and R. Langrock. 2012. On the application of mixed hidden Markov models to multiple behavioural time series. *Interface Focus* 2:180–189.

Scott-Haywood, L., C. Oedekoven, M. Mackenzie, C.Walker, and E. Rexstad. 2014. *MRSea* package (for R).v.0.2.2.

Sivle, L.D., P.H. Kvadsheim, C. Curé, S. Isojunno, P.J. Wensveen, F.P.A. Lam, F. Visser, L. Kleivane, P.L. Tyack, C.M. Harris, and P.J. Miller. 2015. Severity of expert-identified behavioural responses of humpback whale, minke whale, and northern bottlenose whale to naval sonar. *Aquatic Mammals* 41(4): 469–502.

Sivle, L.D., P.H. Kvadsheim, A. Fahlman, F.P.A. Lam, P.L. Tyack, and P.J.O. Miller. 2012. Changes in dive behavior during naval sonar exposure in killer whales, long-finned pilot whales, and sperm whales. *Frontiers in Physiology* 3:400.

Sousa-Lima, R.S., T.F. Norris, J.N. Oswald, and D.P. Fernandes. 2013. A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals. *Aquatic Mammals* 39:23–53.

Stimpert, A.K., S.L. DeRuiter, B.L. Southall, D.J. Moretti, E.A. Falcone, J.A. Goldbogen, A. Friedlaender, G.S. Schorr, and J. Calambokidis. 2014. Acoustic and foraging behavior of a Baird's beaked whale, *Berardius bairdii,* exposed to simulated sonar. *Scientific Reports* 4:7031.

Tyack, P.L., W.M.X. Zimmer, D. Moretti, B.L. Southall, D.E. Claridge, J.W. Durban, C.W. Clark, A. D'Amico, N. DiMarzio, S. Jarvis, E. McCarthy, R. Morrissey, J. Ward, and I.L. Boyd. 2011. Beaked whales respond to simulated and actual navy sonar. *PloS ONE* 6(3):e17009.

Van Parijs, S.M., C.W. Clark, R.S. Sousa-Lima, S. Parks, S. Rankin, D. Risch, D., and I.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* 395:21–36.

Wiggins, S. 2007. Triton (Version 1.80) [Acoustic Processing Software]. Scripps Institution of Oceanography, UC San Diego, La Jolla, California. Retrieved August 1, 2011. Available from www.cetus.ucsd.edu.

Yack, T.M., E.L. Ferguson, R.P. Walker, G.C. Alongi, C.A. Hom-Weaver, and T.F. Norris. 2016. Assessment of Vocal Behavior of Sperm Whales in the Northwestern Atlantic Ocean. Final Report. Prepared for U.S. Fleet Forces Command. Submitted to Naval Facilities Engineering Command Atlantic, Norfolk, Virginia, under Contract No. N62470-10-3011, Task Order 21, issued to HDR, Inc., Virginia Beach, Virginia. 30 March 2016

Zucchini, W. and I. MacDonald. 2009. *Hidden Markov Models for Time Series: An introduction using R.* 2nd edition. Chapman and Hall/CRC, London. 298pp.

This page intentionally left blank.

A

PAMGuard Tools for Click Detection and Measurement

Appendix A. PAMGuard Tools for Click Detection and Measurement

PAMGuard (Gillespie et al. 2008) is an acoustic data-processing software platform that has been widely adopted by the marine mammal bioacoustic research, mitigation and monitoring communities. PAMGuard is freely available (www.pamguard.org), and users who are familiar with the Java programming language can create custom modules to meet their needs. Researchers from Bio-Waves, Inc. have worked closely with the developers of PAMGuard (Sea Mammal Research Unit/University of St. Andrews) to integrate our tools and algorithms into their program.

PAMGuard contains an automated click detector module that can be parameterized (i.e., configured) to detect clicks from specific species or species-groups. Bio-Waves, Inc. has parameterized generalized automated classifiers for sperm whales, several species of beaked whale, dolphins, and blackfish species groups. These classifiers have been tested and validated in the field during various Navy and NOAA funded research projects and marine mammal surveys (e.g., GOALSII, PODS and AMAPPS). They have also been used for a variety of research and monitoring projects, which required efficient post-processing and analysis of large autonomous acoustic recorder datasets. These generalized classifiers have proven reliable for both autonomous acoustic recorder and towed hydrophone array data.

In order to train classifiers to classify clicks to species (within each species-group) Bio-Waves, Inc. has created a ”software bridge” between PAMGuard’s click detector module and the classification module, ROCCA (Real-time Odontocete Call Classification Module). This bridge allows the click detector to pass detected clicks to ROCCA in real-time via one of two user-selected methods: 1) *Automated*: all clicks from user-selected PAMGuard species groups are sent to ROCCA or 2) *Semi-automated*: only specific clicks selected by the user are sent to ROCCA (**Figure 1**).

Once clicks have been sent to ROCCA, JAVA code written by Bio-Waves, Inc. automatically measures features (e.g., peak Hz, inter-click interval, signal-to-noise ratio, etc.) from them. Click measurement capabilities are also available in PAMGuard Viewer Mode. “Viewer Mode”, allows efficient visual review of click detections from large datasets by allowing data analysts to rapidly review automated detections, select click train events, verify species ID’s, and localize individual animals. In this mode the user manually selects click train “events” (e.g., individual whale trains) and marks them by drawing a box around the clicks to signify an event. All of the marked clicks in the “event” are subsequently sent to ROCCA to be measured and measured values are saved in a database (**Figure 2**).

Figure 1. PAMGuard click detector display showing the bearing (y-axis) vs. time (x-axis) display with detected clicks represented as filled ovals with the color indicating automatic classification of species or species groups. Using the semi-automated method, selected clicks can be manually assigned by the user to a "whale train" and these click train clicks are then sent to ROCCA for measurement. In contrast, in the automated method all clicks colored as the species of interest (e.g., beaked whale [orange]) would be sent to ROCCA for measurement.

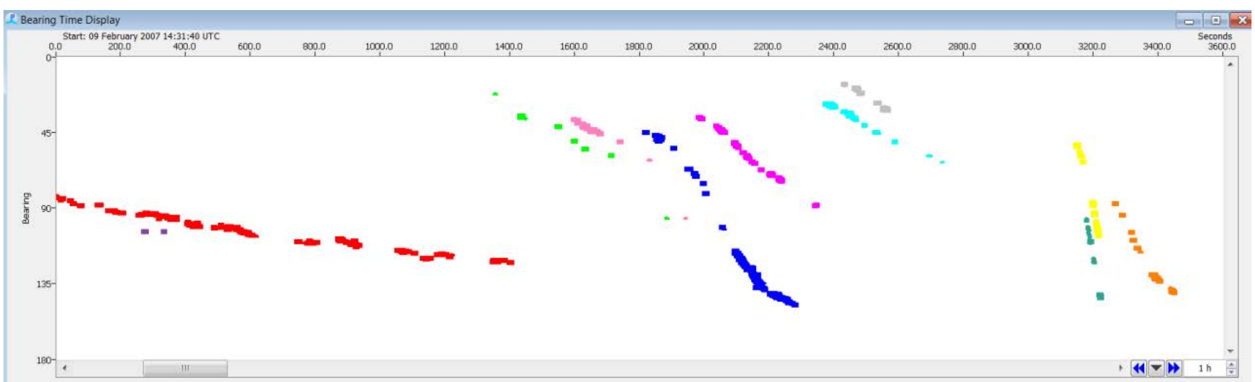

Figure 2. PAMGuard Viewer Mode click detector display illustrating the post-processing method of sending clicks to ROCCA. Events are marked as individual colors, and clicks from each event are sent to ROCCA for measurement.

B

Variables Measured by ROCCA

Appendix B: Variables Measured by ROCCA

Variable	Explanation
Begsweep	slope of the beginning sweep (1 = positive, -1 = negative, 0 = zero)
Begup	binary variable: 1 = beginning slope is positive, 0 = beginning slope is negative
Begdwn	binary variable: 1 = beginning slope is negative, 0 = beginning slope is positive
Endsweep	slope of the end sweep (1 = positive, -1 = negative, = 0 zero)
Endup	binary variable: 1 = ending slope is positive, 0 = ending slope is negative
Enddwn	binary variable: 1 = ending slope is negative, 0 = ending slope is positive
Beg	beginning frequency (Hertz [Hz])
End	ending frequency (Hz)
Min	minimum frequency (Hz)
Dur	duration (seconds)
Range	maximum frequency - minimum frequency (Hz)
Max	maximum frequency (Hz)
mean freq	mean frequency (Hz)
median freq	median frequency (Hz)
std freq	standard deviation of the frequency (Hz)
Spread	difference between the 75th and the 25th percentiles of the frequency
quart freq	frequency at one-quarter of the duration (Hz)
half freq	frequency at one-half of the duration (Hz)
Threequart	frequency at three-quarters of the duration (Hz)
Centerfreq	(minimum frequency + (maximum frequency-minimum frequency))/2
rel bw	relative bandwidth: (maximum frequency - minimum frequency)/center frequency
Maxmin	maximum frequency/minimum frequency
Begend	beginning frequency/end frequency
Cofm	coefficient of frequency modulation (COFM): take 20 frequency measurements equally spaced in time, then subtract each frequency value from the one before it. COFM is the sum of the absolute values of these differences, all divided by 10,000
tot step	number of steps (10 percent or greater increase or decrease in frequency over two contour points)
tot inflect	number of inflection points (changes from positive to negative or negative to positive slope)
max delta	maximum time between inflection points
min delta	minimum time between inflection points
maxmin delta	maximum delta/minimum delta
mean delta	mean time between inflection points
std delta	standard deviation of the time between inflection points
median delta	median of the time between inflection points
mean slope	overall mean slope
mean pos slope	mean positive slope
mean neg slope	mean negative slope
mean absslope	mean absolute value of the slope

Variable	Explanation
Posneg	mean positive slope/mean negative slope
perc up	percent of the whistle that has a positive slope
perc dwn	percent of the whistle that has a negative slope
perc flt	percent of the whistle that has zero slope
up dwn	number of inflection points that go from positive slope to negative slope
dwn up	number of inflection points that go from negative slope to positive slope
up flt	number of times the slope changes from positive to zero
dwn flt	number of times the slope changes from negative to zero
flt dwn	number of times the slope changes from zero to negative
flt up	number of times the slope changes from zero to positive
step up	number of steps that have increasing frequency
step dwn	number of steps that have decreasing frequency
step.dur	number of steps/duration
inflect.dur	number of inflection points/duration

C

Diagnostics for Statistical Models

This page intentionally left blank.

Appendix C: Diagnostics for Statistical Models

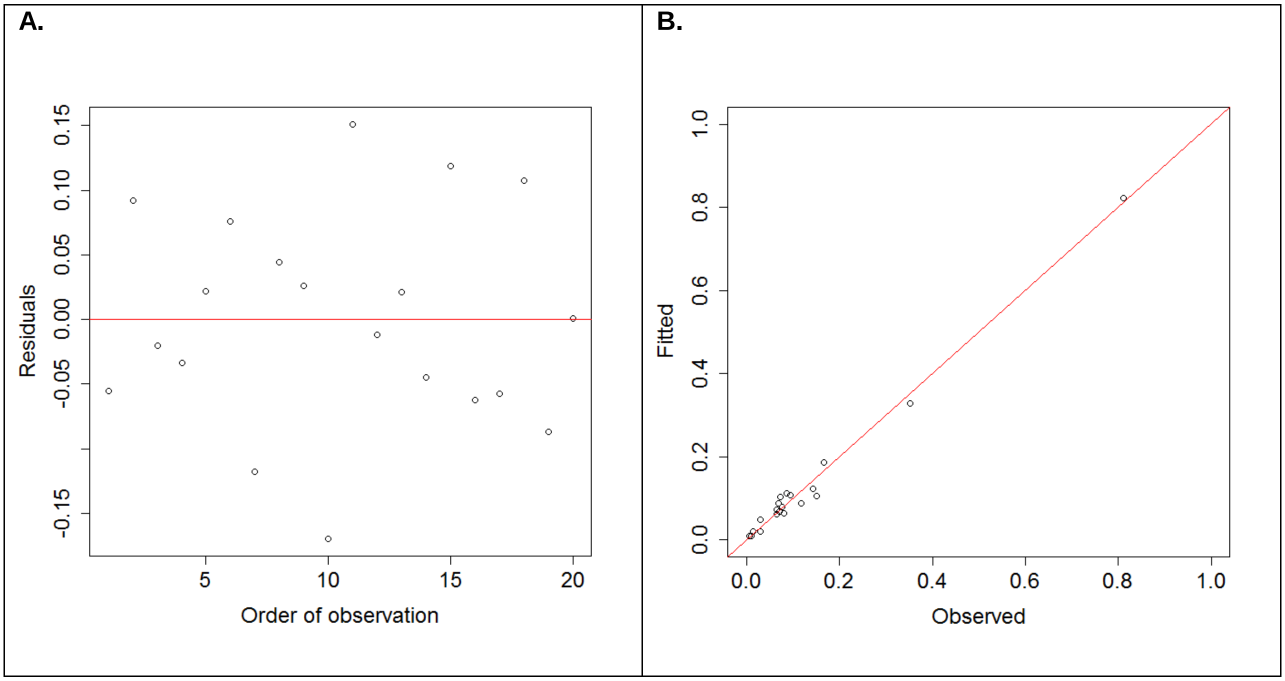

Figure 1. Diagnostics for the final model of presence/absence (PA model 1) of acoustic detections of delphinids with *Site* the predictor. (A.) the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. (B.) mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line, is desired in B.

Table 1. Observed versus predicted values from the dolphin presence-absence model with *Site.* Ideally few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	137862	10497
Presence	10699	10882

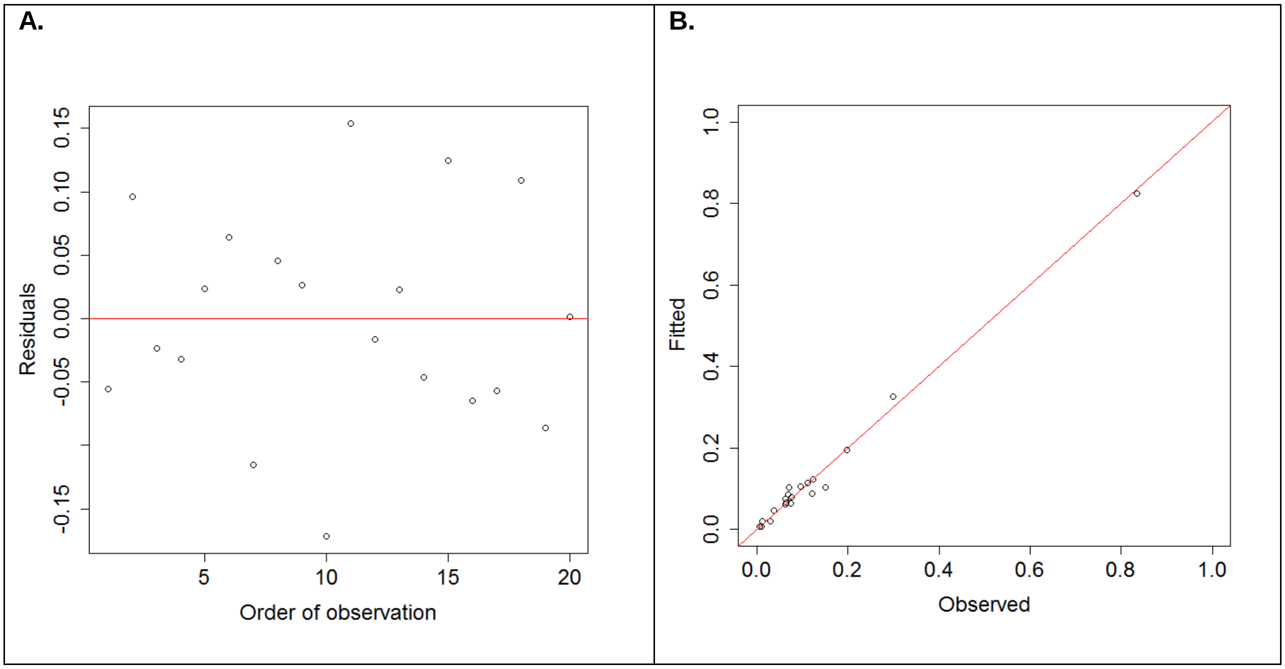

Figure 2. Diagnostics for the final model (PA model 2) presence/absence of acoustic detections of delphinids with *Site* and *Type 1 long* as the predictor (A.) the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. (B.) mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 2. Observed versus predicted values from the dolphin presence-absence model with *Site.* Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	135258	10106
Presence	13303	11273

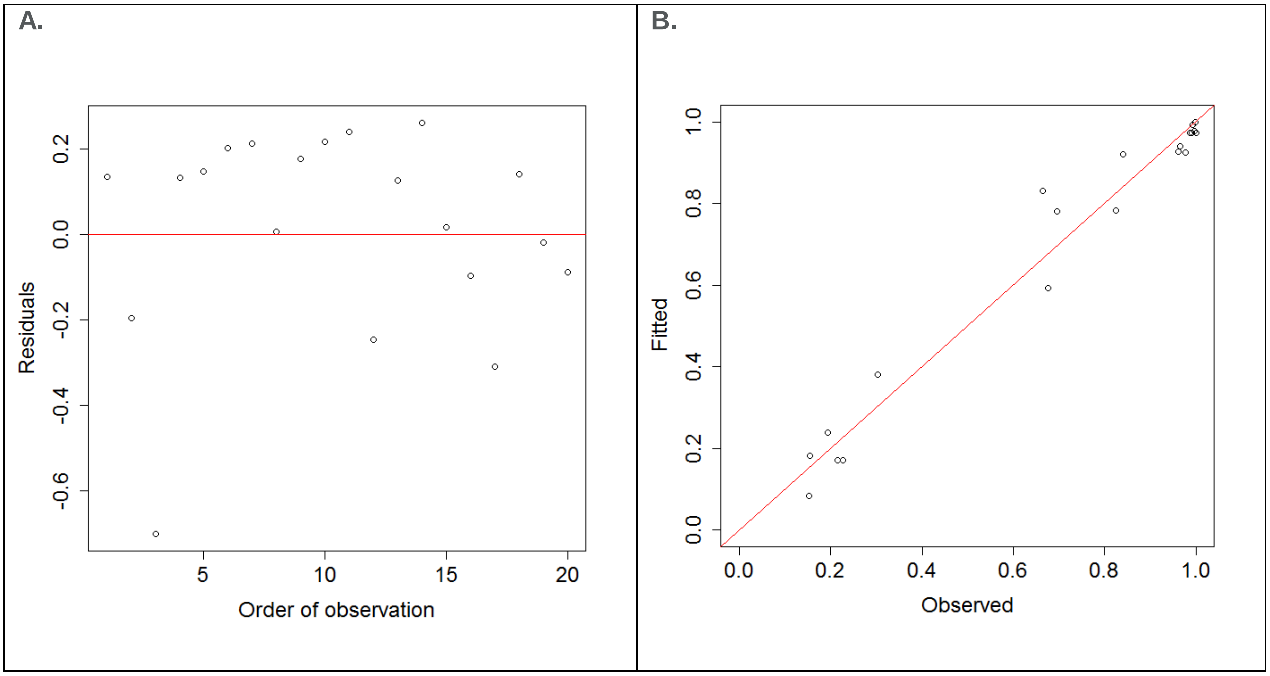

Figure 3. Diagnostics for dolphin PSTGAE model 1 – whistle mode with *Site, Sonar, Buzzes* and *clicks* as the predictors (A.) the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. (B.) mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 3. Observed versus predicted values PSTGAE whistle model 1 with *Site, Sonar, Buzzes* and *Clicks* as predictors. Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	2750	921
Presence	620	6590

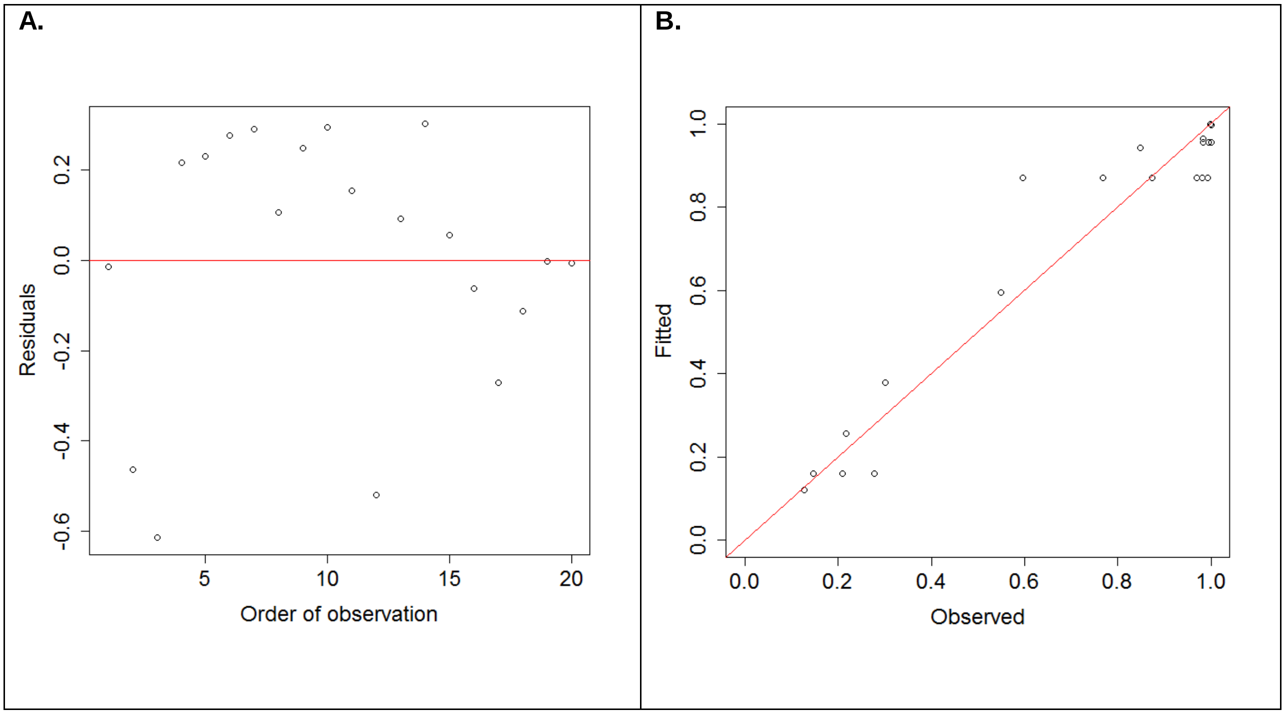

Figure 4. Diagnostics for dolphin PSTGAE model 2 – whistle model with *Site, Type 1 Long, Buzzes* and *Clicks* as the predictors (PSTGAE model 2). (A.) the means of binned fitted values the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. (B.) mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 4. Observed versus predicted values whistle model (PSTGAE model 2) with *Site, Type 1 long, Buzzes* and *Clicks* as predictors. Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	2812	953
Presence	558	6558

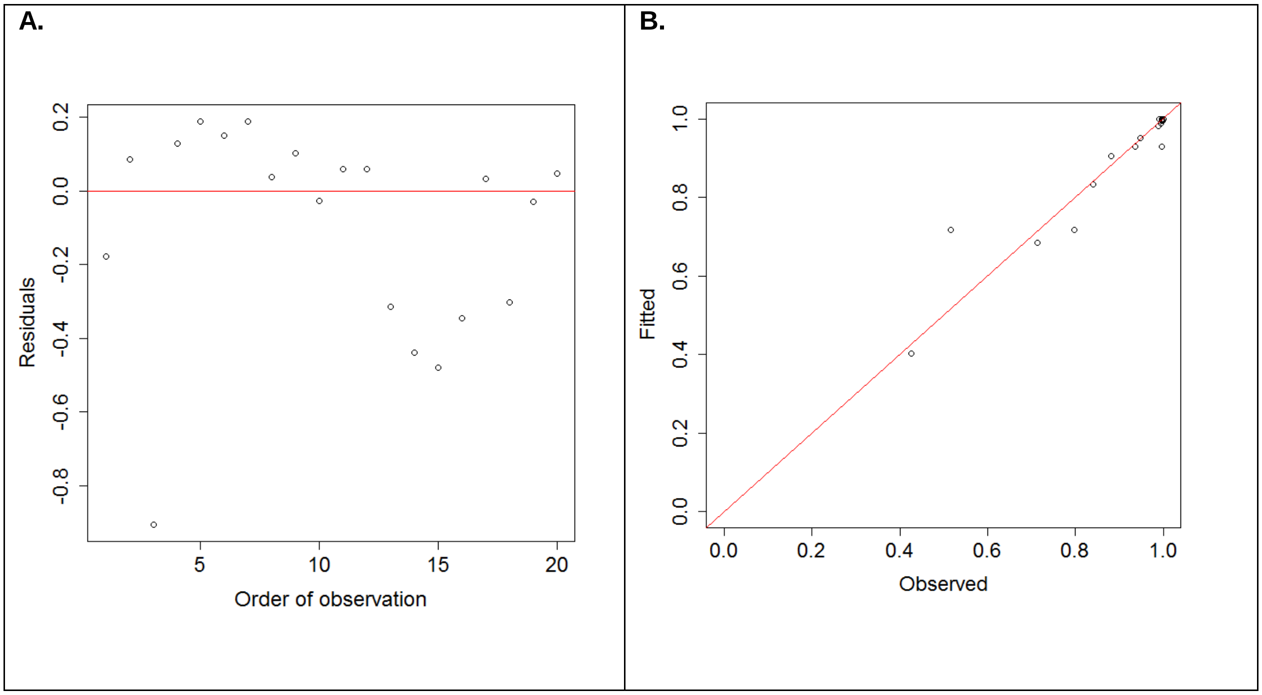

Figure 5. Diagnostics for dolphin PSTGAE model 3 – clicks with *Site, Sonar, Whistles and Buzzes* as predictors (PSTGAE model 3). A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 5. Observed versus predicted values PSTGAE click model 3 with *Site, Sonar, Whistles and Buzzes* as predictors. Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	945	1998
Presence	137	7801

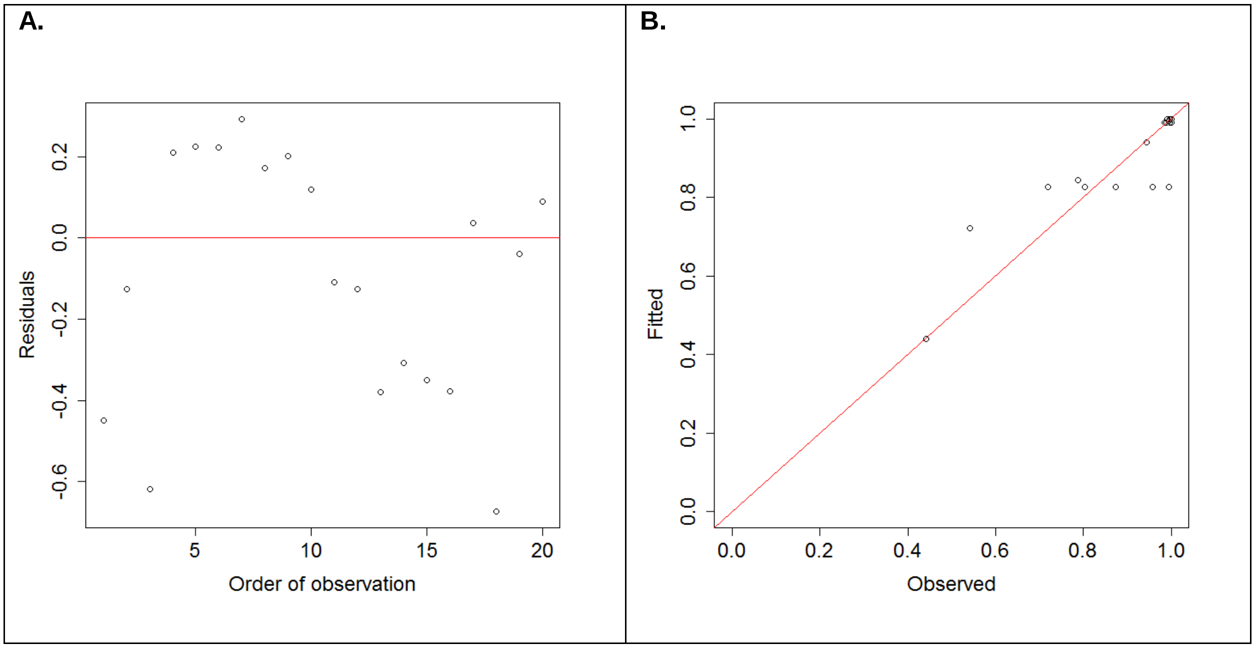

Figure 6. Diagnostics for dolphin PSTGAE model 4 – clicks with *Site, Type 2 long* signal, *Buzzes* and *Whistles as* predictors. A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 6. Observed versus predicted values PSTGAE click model 4 with *Site, Type 2 long*, *Buzzes* and *Whistles* as predictors. Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	1020	3220
Presence	62	6579

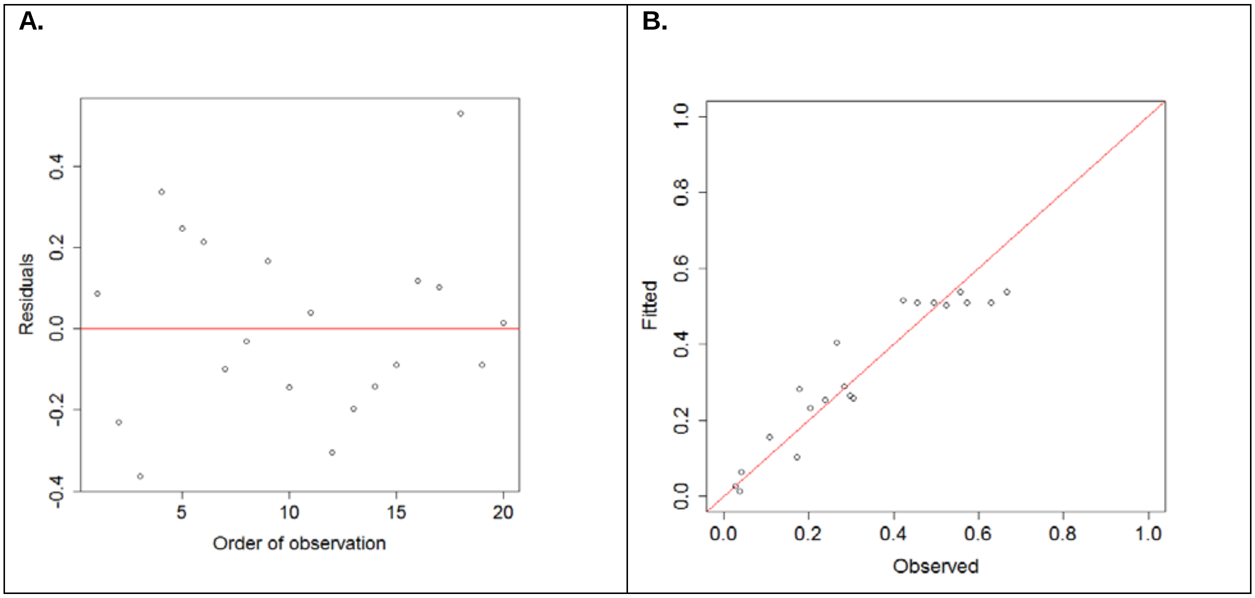

Figure 7. Diagnostics for dolphin PSTGAE model 5 – buzzes with *Site, Whistles, Clicks and Sonar* as the predictors. A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 7. Observed versus predicted values PSTGAE buzz model 5 with *Site, Whistles, Clicks and Sonar* as predictors. Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	5153	1076
Presence	2205	2447

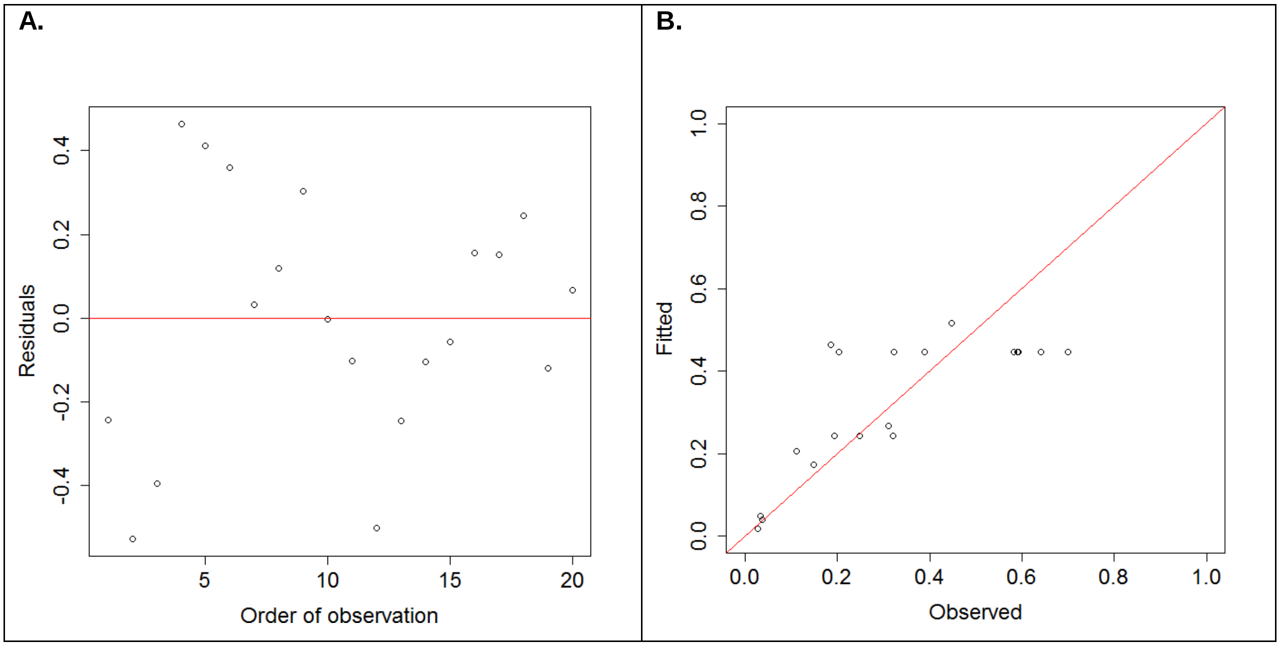

Figure 8. Diagnostics for dolphin PSTGAE model 6 – buzzes with *Site*, *Clicks*, *whistles* and *Type 3 med* sonar signals. A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 8. Observed versus predicted values PSTGAE model 6 buzzes with *Site*, *Clicks*, *whistles*, and *Type 3 med* sonar signals. Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	4064	774
Presence	3294	2749

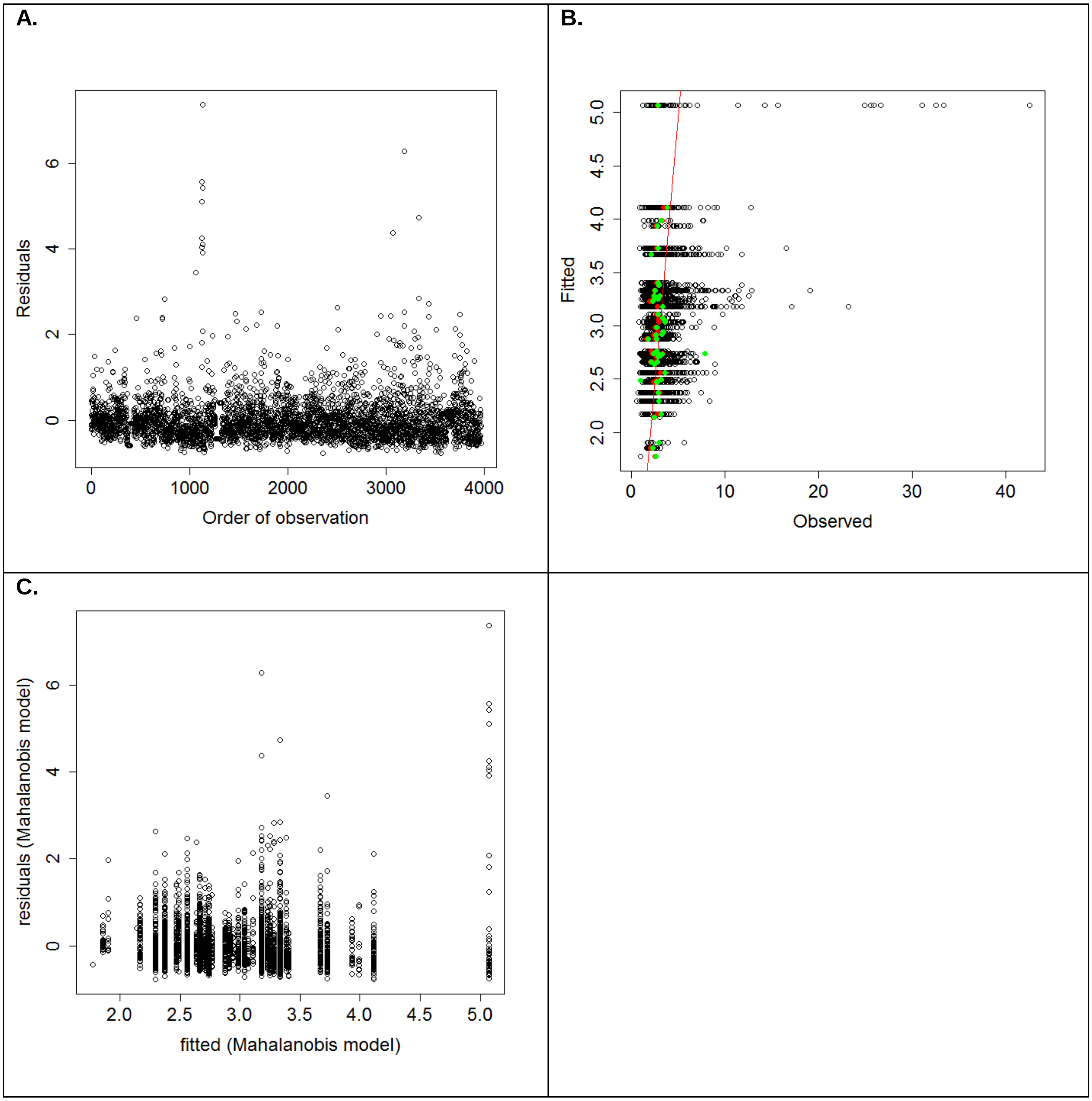

Figure 9. Diagnostic plots for the Mahalanobis distance analysis. A. residual plot, B. fitted values from the whistle characteristic model for delphinids against observed values. The red line indicates a perfect fit of the model to the observed data. Green and red points represent, respectively, the median and mean of the observed values corresponding to the unique fitted values. C. Plot of the fits and residuals. No trend in the pattern of the residuals is desired in A, most of the points lying close to the line is desired in B, and no pattern of the residuals is desired in C.

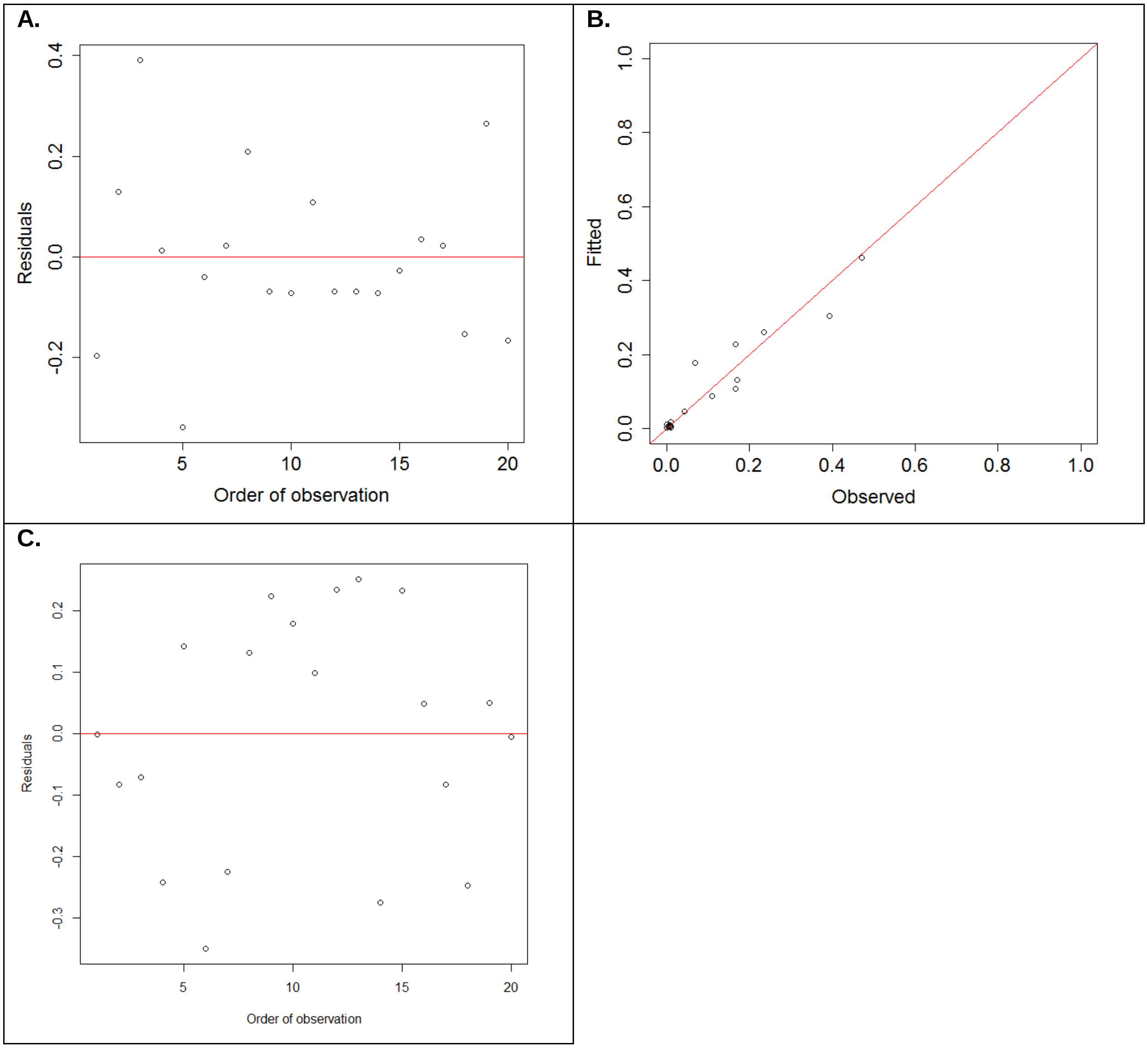

Figure 10. Diagnostics for the final predicted model (PA model 3) presence/absence of acoustic detections of beaked whales with predictors *Site, Sonar and Time of day.* A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 9. Observed versus predicted values for the beaked whale presence-absence model with *Site, Sonar* and *Time of day.* Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	13562	167
Presence	5996	1856

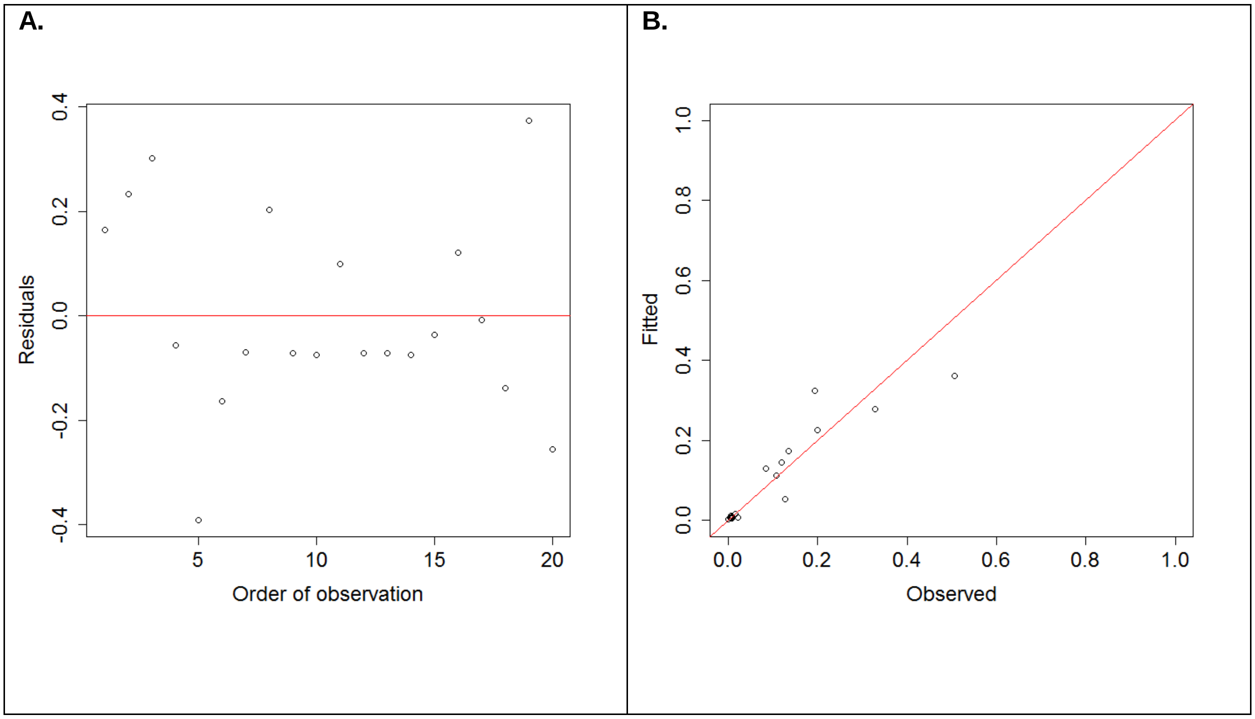

Figure 11. Diagnostics for the final predicted model presence/absence of acoustic detections of beaked whales (PA model 4) with predictors *Site, Type 2 long* signal *and Time of day.* A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 10. Observed versus predicted values for the beaked whale presence-absence model with *Site, Type2 long* and *Time of day.* Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	12905	225
Presence	6653	1798

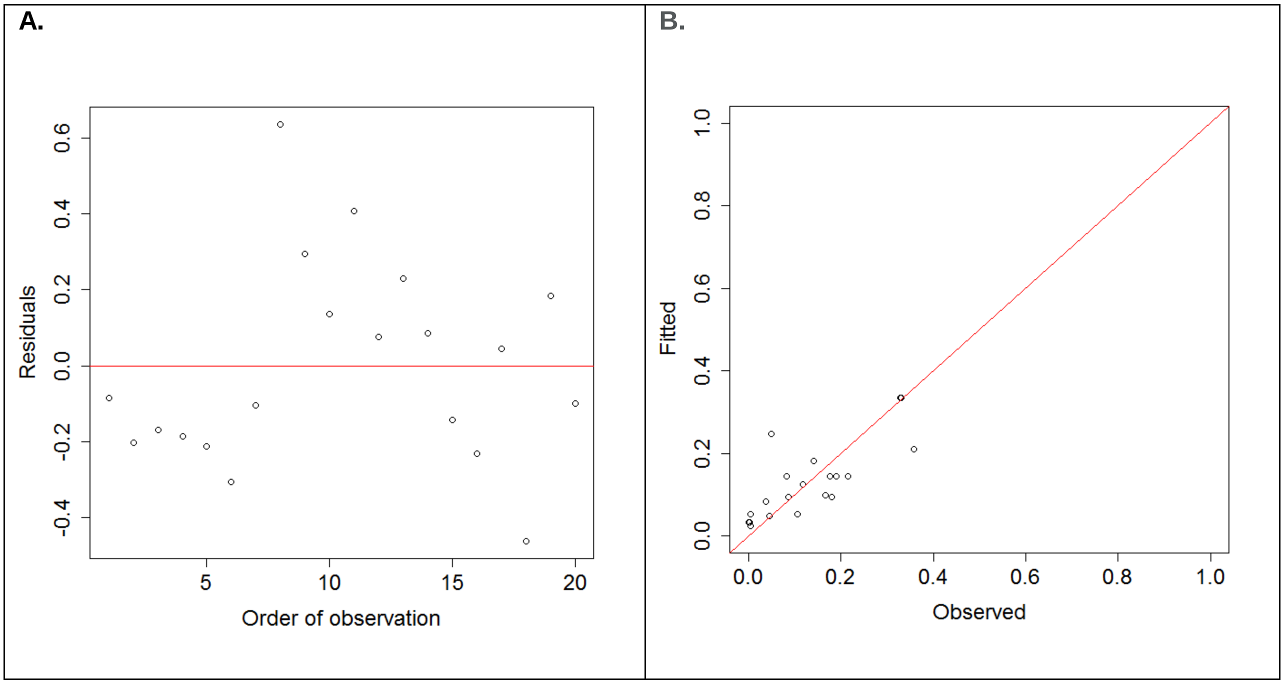

Figure 12. Diagnostics for (PA model 5) presence/absence of acoustic detections of minke whales with predictors *SurveyType* and *Sonar.* A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

Table 11. Observed versus predicted values for the minke whale presence-absence model (PA model 5) with *SurveyType,* and *Sonar.* Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	9275	639
Presence	7235	1857

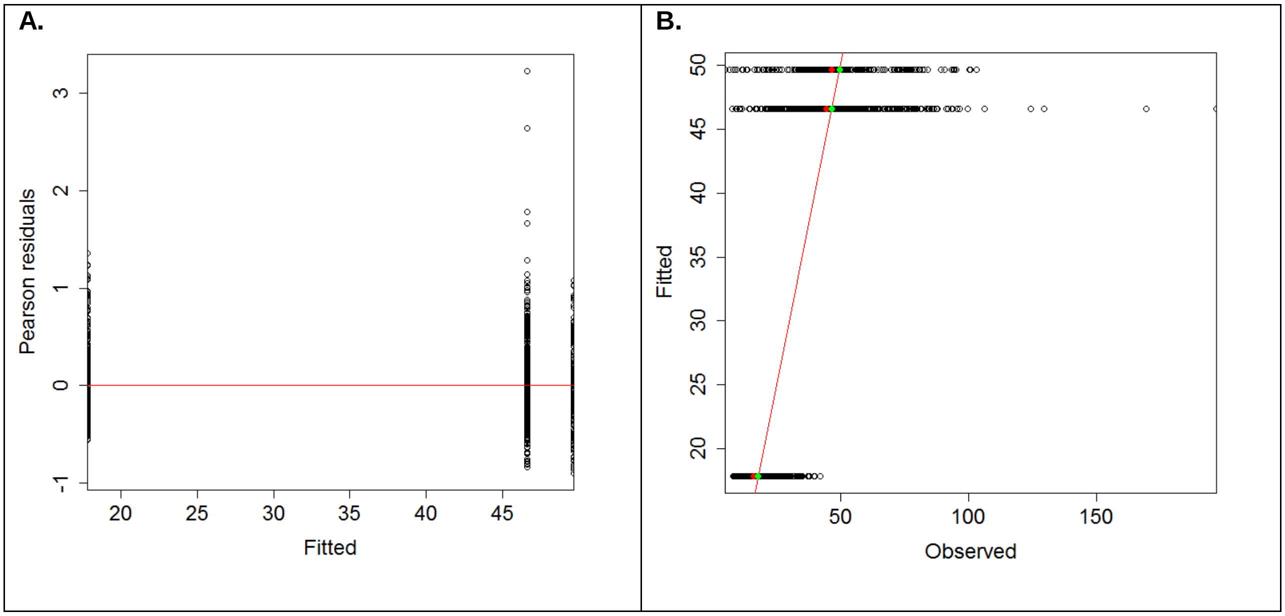

Figure 13. Diagnostics for the minke whale call duration model with *Site.* A. fits-residual plot. The residuals are Pearson residuals. B. The red line indicates a perfect fit of the model to the observed data. Green and red points represent respectively the median and mean of the observed values corresponding to the unique fitted values. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B.

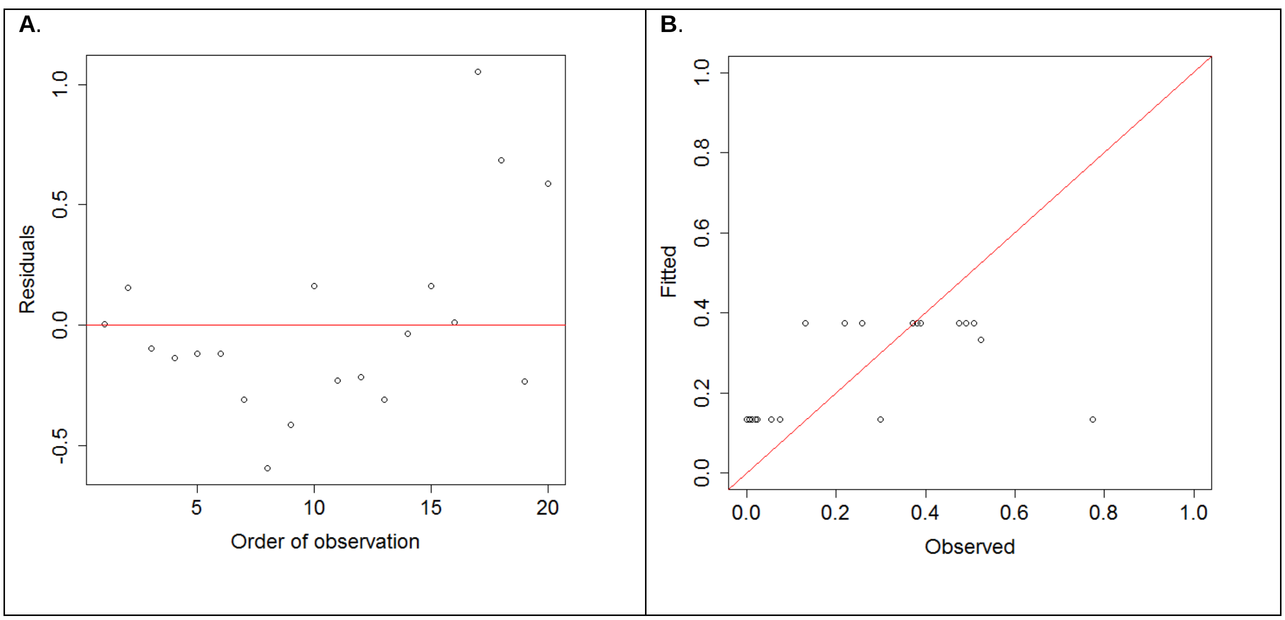

Figure 14. Diagnostics for the final model (PA model 6) presence/absence of acoustic detections of sperm whales with predictors *Daynight.* A. the means of binned fitted values versus the means of corresponding residuals. Binning occurred by splitting the fitted values into 20 equally sized bins in ascending order of observation. B. mean observed versus mean fitted values from presence/absence of cetacean acoustic detections. No trend in the pattern of the residuals is desired in A, and most of the points lying close to the line is desired in B, in this there is a strong discrete predictor leading to the pattern observed.

Table 12. Observed versus predicted values for the sperm whale presence-absence model (PA model 5) with *Daynight.* Ideally, few cases of disagreement should be found.

	Observed	
Site Number	Absence	Presence
Predicted		
Absence	18506	2857
Presence	12884	7716