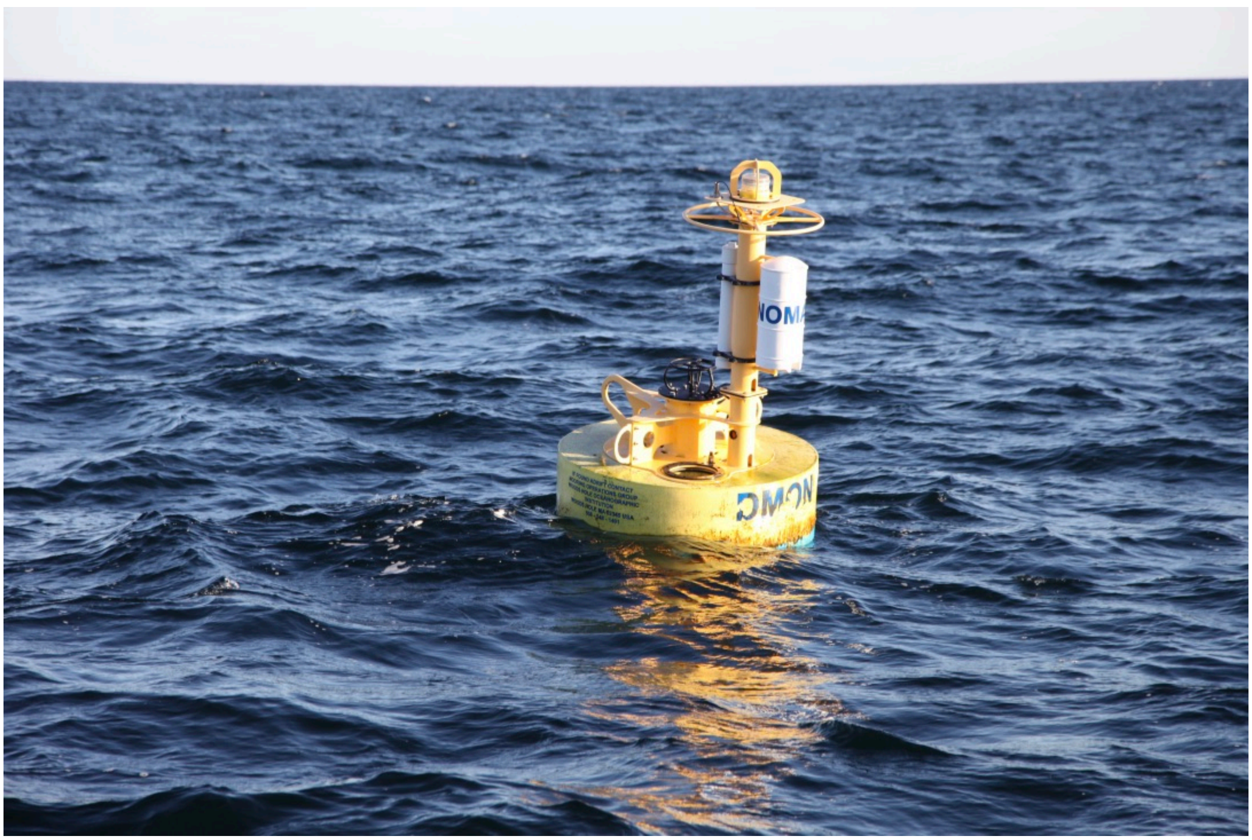
OCS Study
BOEM 2019-061

Evaluating the Accuracy and Detection Range of a Moored Whale Detection Buoy near the Massachusetts Wind Energy Area



US Department of the Interior
Bureau of Ocean Energy Management
Office of Renewable Energy Programs

MASSACHUSETTS CLEAN ENERGY CENTER®

BOEM
Bureau of Ocean Energy Management

OCS Study
BOEM 2019-061

Evaluating the Accuracy and Detection Range of a Moored Whale Detection Buoy near the Massachusetts Wind Energy Area

August 2019

Authors:
Mark Baumgartner, Woods Hole Oceanographic Institution
YT Lin, Woods Hole Oceanographic Institution

Prepared under Massachusetts Clean Energy Center and Bureau of Ocean Energy Management
Cooperative Agreement M17AC00012
By
Woods Hole Oceanographic Institution
266 Woods Hole Road
Woods Hole, MA 02543

US Department of the Interior
Bureau of Ocean Energy Management
Office of Renewable Energy Programs

MASSACHUSETTS CLEAN ENERGY CENTER®

BOEM
Bureau of Ocean Energy Management

DISCLAIMER

This report was prepared for the Massachusetts Clean Energy Center and Bureau of Ocean and Energy Management under BOEM Cooperative Agreement number M17AC00012 and a Grant Agreement with MassCEC. This report has been technically reviewed by BOEM and it has been approved for publication. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the opinions or policies of the US Government, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

REPORT AVAILABILITY

To download a PDF file of this report, go to the US Department of the Interior, Bureau of Ocean Energy Management Data and Information Systems webpage (http://www.boem.gov/Environmental-Studies-EnvData/), click on the link for the Environmental Studies Program Information System (ESPIS), and search on 2019-061. The report is also available at the National Technical Reports Library at https://ntrl.ntis.gov/NTRL/.

CITATION

Baumgartner, M.F and Lin, Y-.T. 2019. Evaluating the Accuracy and Detection Range of a Moored Whale Detection Buoy near the Massachusetts Wind Energy Area. Woods Hole (MA): Massachusetts Clean Energy Center and U.S. Department of the Interior, Bureau of Ocean Energy Management. OCS Study BOEM 2019-061. 72 p.

ABOUT THE COVER

The DMON buoy moored off Nomans Land Island just prior to recovery on March 30, 2016.

ACKNOWLEDGMENTS

Chapter 2 acknowledgements: We are grateful for the critical contributions of Julianne Bonnell (NOAA NEFSC), Sofie Van Parijs (NOAA NEFSC), Peter Corkeron (NOAA NEFSC), Cara Hotchkin (NAVFAC Atlantic), Keenan Ball (WHOI), Léo-Paul Pelletier (WHOI), Jim Partan (WHOI), Don Peters (WHOI), John Kemp (WHOI), Jeff Pietro (WHOI), Kris Newhall (WHOI), Andrew Stokes (U.S. Coast Guard), Tim Cole (NOAA NEFSC), Ester Quintana (New England Aquarium), and Scott Kraus (New England Aquarium). We also thank Annamaria Izzi, Danielle Cholewiak, and Genevieve Davis of the NOAA NEFSC for assistance in developing the analyst protocol. We are grateful to the NOAA NEFSC aerial survey observers (Leah Crowe, Pete Duley, Jen Gatzke, Allison Henry, Christin Khan, and Karen Vale) and the NEAq aerial survey observers (Angela Bostwick, Marianna Hagbloom, and Paul Nagelkirk).

Chapter 3 acknowledgements: We are grateful for the critical contributions of Hansen Johnson (Dalhousie University and WHOI), Arthur Newhall (WHOI) and Christopher Taggart (Dalhousie University). We are also indebted to Ken Houtler and Ian Hanley (WHOI) for their skilled operation of the R/V *Tioga*, to John Kemp, Meg Donohue, Jim Dunn, and Nico Llanos of the WHOI Mooring Operations Group for development, deployment, and recovery of the mooring systems, to Ed O'Brien and the WHOI Dive Group for assisting in mooring recovery, to Peter Koski, Julien Bonnel and Dan Zitterbart of WHOI Applied Ocean Physics & Engineering Department for guidance and advice, and to Kim Davies, Delphine Durette-Morin, Meg Carr, and Marcia Pearson at Dalhousie University for providing helpful feedback. Additional funding to support the glider deployment and Hansen Johnson's time as a PhD student for this project were provided by the Offshore Energy Research Association (OERA), Marine Environmental Prediction and Response Network (MEOPAR) Whales Habitat and Listening Experiment (WHaLE), Killam Foundation, Vanier Canada Graduate Scholarship program, Dalhousie University, and the Nova Scotia Graduate Scholarship program.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

BOEM	Bureau of Ocean Energy Management
DMON	Digital acoustic MONitoring instrument
DSP	Digital Signal Processor
GB	Gigabytes
GPS	Global Positioning System
HLA	Horizontal hydrophone Line Array
LFDCS	Low Frequency Detection and Classification System
MassCEC	Massachusetts Clean Energy Center
MWEA	Massachusetts Wind Energy Area
NEFSC	Northeast Fisheries Science Center
NEAq	New England Aquarium
NOAA	National Oceanic and Atmospheric Administration
PAM	Passive Acoustic Monitoring
SHRU	Several Hydrophone Receiving Unit
VLA	Vertical hydrophone Line Array
WHOI	Woods Hole Oceanographic Institution

1 Executive Summary

There is significant concern about the impact of the survey and construction phases of wind energy development on endangered large whales, particularly the critically endangered North Atlantic right whale (*Eubalaena glacialis*). Mitigation of these impacts will be required as part of a regulatory environmental compliance framework, and the use of near real-time passive acoustics to alert developers to the presence of whales will likely be part of an effective mitigation strategy. The Woods Hole Oceanographic Institution (WHOI) has developed the digital acoustic monitoring (DMON) instrument and the low-frequency detection and classification system (LFDCS) to detect, classify, and report the sounds of large whales in near real time from a variety of autonomous platforms, including moored buoys and electric ocean gliders. The moored buoy has been in operational use since its first deployment during 2015 on the northern edge of the Massachusetts Wind Energy Area. It was originally deployed there to monitor for right whales near three Coast Guard gunnery ranges, and the near real-time detections were used by the Coast Guard to schedule training exercises in the gunnery ranges. One limitation of this system is the absence of information about species-specific acoustic detection range, which remains an important knowledge gap for most passive acoustic monitoring systems. Our MassCEC-BOEM project sought to (1) evaluate the accuracy of the buoy's near real-time detections of right, humpback, sei and fin whales using contemporaneous acoustic recordings and visual sightings, and (2) characterize the detection range of the system for right whales using collocated hydrophone arrays capable of localizing calling whales. We believe this whale monitoring technology will help to reduce the impact of wind energy development activities on large whales; however, it is vital to characterize the system's performance and detection range before fully integrating it into a mitigation strategy.

1.1 Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: system description and evaluation

Managing interactions between human activities and marine mammals often relies on an understanding of the real-time distribution or occurrence of animals. Visual surveys typically cannot provide persistent monitoring because of expense and weather limitations, and while passive acoustic recorders can monitor continuously, the data they collect are often not accessible until the recorder is recovered. We have developed a moored passive acoustic monitoring system that provides near real-time occurrence estimates for humpback, sei, fin, and North Atlantic right whales from a single site for a year, and makes those occurrence estimates available via a publicly accessible website, email and text messages, a smartphone/tablet app, and the U.S. Coast Guard's maritime domain awareness software. We evaluated this system using a buoy deployed off the coast of Massachusetts during 2015-2016 and redeployed again during 2016-2017. Near real-time estimates of whale occurrence were compared to simultaneously collected archived audio as well as whale sightings collected near the buoy by aerial surveys. False detection rates for right, humpback, and sei whales were 0% and nearly 0% for fin whales, while missed detection rates at daily time scales were modest (12-42%). Missed detections were significantly associated with low calling rates for all species. We observed strong associations between right whale visual sightings and near real-time acoustic detections over a monitoring range of 30-40 km and temporal scales of 24-48 hours, suggesting that silent animals were not especially problematic for estimating occurrence of right whales in the study area. There was no association between acoustic detections and visual sightings of humpback whales. The moored buoy has been used to reduce the risk of ship strikes for right whales in a U.S. Coast Guard gunnery range, and can be applied to other mitigation applications.

1.2 Acoustic detection range of right whale upcalls detected in near real time from a moored buoy and a Slocum glider

Mitigation of anthropogenic impacts on North Atlantic right whales and other at-risk species is critical but challenging given limited survey resources and the cryptic nature of whale behavior. Using passive acoustic monitoring (PAM) to alert ocean users to whale presence in near real-time can support mitigation efforts. The Woods Hole Oceanographic Institution (WHOI) has developed the digital acoustic monitoring (DMON) instrument and low-frequency detection and classification system (LFDCS) to detect and classify baleen whales in near real-time from autonomous platforms (e.g., buoys and gliders). The species-specific acoustic detection range of many PAM systems, including the DMON/LFDCS, remains an important knowledge gap when applying such systems to mitigation actions. The goal of this study was to determine the range-dependent accuracy of the DMON/LFDCS for both a mobile and a fixed platform. Over a 4-week period (28 Feb to 30 Mar) during the spring of 2017, we deployed a DMON/LFDCS-equipped Slocum glider, vertical hydrophone array, and a horizontal hydrophone array alongside an extant DMON/LFDCS moored buoy at a shallow (30m) site approximately 15 km southwest of Martha's Vineyard, Massachusetts, USA. We used beamforming and a normal mode back-propagation technique with the array data to localize right whale upcalls, then conducted a call-by-call comparison between calls detected on the array and those detected by the glider or buoy to determine the probability of detecting localized calls for each platform. Both the buoy and the glider performed similarly at both close and far ranges, detecting 68 and 56% of localized calls within 5 km, respectively, and 17 and 25% of localized calls at 15-20 km, respectively; however, the glider detected a higher proportion of calls than the buoy between 5 and 15 km. Logistic regression analysis suggested that the probability of detecting localized calls was 0.333 at 9.4 and 17.2 km for the buoy and glider, respectively. The results help us to better characterize the performance of our monitoring system, which in turn allows us to disseminate more accurate information about whale distribution and occurrence to research, government, and industry stakeholders.

1.3 Summary and Recommendations

Having characterized the accuracy and detection range of the DMON/LFDCS moored buoy, we consider here its applicability to mitigating interactions between wind energy development activities and North Atlantic right whales. We focus on right whales because of their depleted status and the urgency with which solutions to human-caused mortality in their declining population must be implemented. The primary threats to right whales from wind energy development are (1) exposure to noise from construction activities and survey/construction/maintenance vessel traffic, and (2) ship strikes from survey/construction/maintenance vessel traffic. These threats can be managed by restricting the time during which industrial activities will occur to periods when right whales are historically scarce (e.g., summer and fall months in the MWEA). However, our observations (and those of other scientists) indicate that right whales can occur in the MWEA at any time of the year. Because the death of a single right whale has a significant impact on the population trajectory (e.g., Caswell et al. 1999), we believe that protections must be in place even for periods when right whale occurrence is historically low.

Near real-time passive acoustic detections can be part of a suite of monitoring approaches that may include visual observers on construction platforms and vessels, aerial surveys, and automated infrared imaging. Passive acoustic monitoring is particularly well suited for monitoring large areas (kilometers to tens of kilometers in radius; i.e., larger than the detection area of visual/infrared methods) persistently over long time periods. The DMON/LFDCS moored buoy is capable of operating at sea for a year at a time, and for right whales, it has a reliable detection range of 9.4 km (5 nautical miles; assuming a detection probability of ≥0.333; see Chapter 3) and a monitoring range of 30-40 km over time scales of 24-48 hours (see Chapter 2). The DMON/LFDCS currently relies on an omnidirectional hydrophone, and therefore cannot localize whale calls. Instead, it provides information on the occurrence of one or more whales that are within its maximum detection range (at least 20 km; see Chapter 3). As such, it can provide a general warning that whales are in the area, which can be used to directly manage industrial activities or to trigger more fine-scale (and perhaps more expensive) monitoring methods, such as aerial surveys. It can play a vital role as an early warning system for the presence of right whales, triggering a higher alert status, increased vigilance, and (or) mitigation activities.

Monitoring around construction sites may require more specificity about whale location than the current DMON/LFDCS moored buoy can provide. Limitations on power and communications bandwidth (i.e., the amount of data that can be transmitted to shore) severely constrain the ability to localize calls from an autonomous platform. For example, the processing described in Chapter 3 to localize calls using audio from a 4-channel vertical line array and an 8-channel horizontal line array is impossible to do on an autonomous platform because (1) the computer required to do the processing consumes too much energy to run on a power-limited autonomous platform for long periods of time, (2) even if power was not limiting, the localization process cannot be fully automated; a human still is needed to supervise the process, and (3) satellite communications to transmit 12 channels of audio data for processing on shore are extremely expensive and consume considerable energy (again, not suitable for a power-limited autonomous platform). Localization is not trivial, but there is one approach that could be used with the DMON/LFDCS that we recommend for future development.

A two-dimensional particle velocity sensor has both an omnidirectional hydrophone and two compass-corrected dipole (accelerometer) sensors that can be used to determine the bearing to low-frequency sounds. These sensors can be integrated with the DMON, as the instrument has 3 channels of audio input: one channel for the omnidirectional hydrophone and one channel each for the two dipole sensors (note that this integration has not yet occurred; we suggest it here for future development). The LFDCS can detect calls on the omnidirectional hydrophone, while the two dipole sensors can be used to calculate a bearing to detected calls in real time. The measurement of bearing for detected calls would allow two buoys with overlapping detection ranges to use cross-bearings to estimate a location of a whale (Figure 1.1). The method of cross bearings is an elegant approach that has been used extensively for monitoring baleen whales by DIFAR sonobuoys (e.g. Green et al. 2004) and toothed whales by towed arrays (e.g., Rankin et al. 2008). In concept, the changes to the system depicted in Figure 2.2 would be trivial, but the addition of bearing would allow occurrence estimates, localization, and a crude minimum count of vocalizing whales (provided calls were close in time, but whales were sufficiently separated in space to yield different bearings attributable to different whales). The DMON/LFDCS moored buoy is an ideal platform for such a system, since the bottom structure (multi-function node) of the mooring is extremely stable (making it ideal for sensing particle velocity), and all of the call detection, classification, transmission, and on-shore review components of the system would remain exactly the same.

Localization of whale calls will certainly help with mitigating the impacts of wind energy development on right whales. However, it is important to keep in mind that right whales are neither trees nor metronomes. Right whales move (unlike trees), so while localizations are very accurate in real time, their accuracy decreases exponentially over time (i.e., the area in which the whale might occur after detection grows as $\pi[\text{swim speed} \times \text{time}]^2$). Right whales can also call irregularly in time (unlike metronomes), and if the interval between calls is long, there is substantial uncertainty in the whale's position between localizations. Moreover, infrequent calling often precludes inference about important behaviors. For example, determining the swim direction of a right whale from a track derived from regular localizations (e.g., to know if a whale is swimming toward a construction platform) is often impossible because of infrequent or irregular calling.

Despite these caveats about localization, there is no doubt that the development of real-time localization by cross-bearings with the DMON/LFDCS will provide more information than the current DMON/LFDCS, which uses an omnidirectional hydrophone. The current DMON/LFDCS represents the state-of-the-art in near real-time passive acoustic monitoring, and it will be a very useful monitoring tool to support mitigation efforts for right whales by wind energy developers and regulators. The addition of two-dimensional particle velocity sensing will extend the state-of-the-art in a system that is now well characterized in the Massachusetts Wind Energy Area.

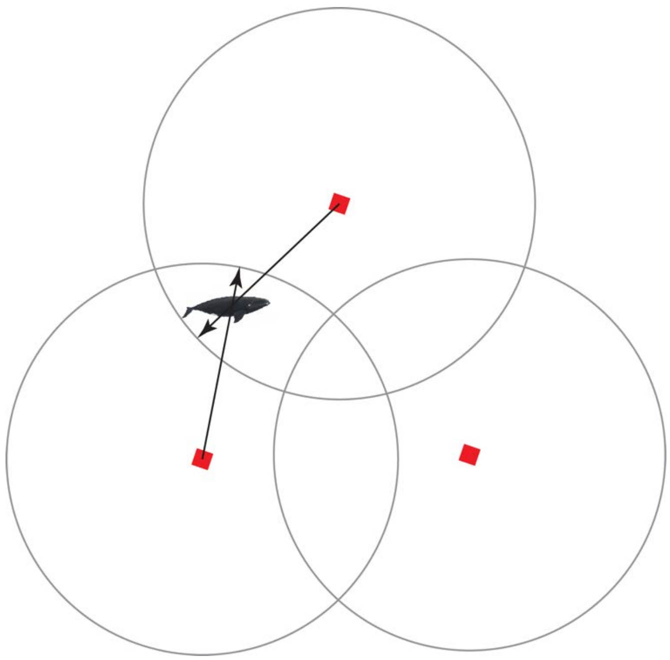


Figure 1.1. Cross-bearing localization.
Cross bearing localization from two-dimensional particle velocity instruments (red squares). Gray circles represent the acoustic detection range. Arrows indicate bearings to a detected right whale upcall from the north and west buoys; the location where the bearings cross is the estimate for the whale's location. Note that the whale is outside of the east buoy's acoustic detection range, so no detection or bearing is available. Localization is possible within the intersection of the detection range circles.

2 Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: system description and evaluation

2.1 Introduction

Marine mammals are an integral part of the ocean ecosystem and many are impacted by human activities, but like most marine organisms, their occurrence, distribution, and abundance are a challenge to monitor from unmanned ocean observing systems. Human observers have traditionally detected marine mammals during visual surveys, relying on the animals to return to the sea surface periodically to breathe where they can be visually detected. This approach is often expensive, as it requires a large team of observers and a ship or aircraft. Moreover, visual surveys are limited by weather and sighting conditions, such as fog, rain, heavy seas, and darkness. For their expense, visual surveys are often inefficient for persistent real-time monitoring of marine mammal occurrence, albeit for other tasks, such as photo identification, health assessment and abundance estimation, visual surveys remain an essential observing methodology for many species.

In recent decades, passive acoustic recorders have become extremely popular for detecting vocally active marine mammals, as they can operate continuously for periods of months to years (Mellinger et al., 2007; Van Parijs et al., 2009). Widespread use of passive acoustics for persistent marine mammal monitoring faces two challenges: (1) most passive acoustic recordings are only available for analysis after instruments are recovered, and (2) analysis of passive acoustic recordings is typically slow and tedious, involving trained human analysts that pore over large volumes of acoustic data or verify automated detections to assess occurrence. In many cases (particularly research applications), the delays in access and analysis are perfectly acceptable, but for mitigation applications or those involving real-time response, most passive acoustic recorders are unhelpful.

There is an urgent need for real-time information on the occurrence of marine mammals for both science and mitigation applications. Such a real-time capability can improve the efficiency of traditional visual-based research efforts by identifying areas where animals are likely to be located, and can provide critical occurrence information in sensitive areas where human activities must be managed to avoid harmful interactions with marine mammals. Van Parijs et al. (2009) reviewed several real-time or near real-time passive acoustic systems, including the Cornell University North Atlantic right whale detection buoy, which has been used to reduce ship strike risks from liquefied natural gas tankers transiting the shipping lanes approaching Boston, Massachusetts for over a decade. The work described here took inspiration from Cornell’s innovative and pioneering efforts.

We developed a system to monitor the occurrence of baleen whales in near real time from long-endurance Slocum ocean gliders (Baumgartner et al., 2013), and have in recent years adapted this system to operate from a purpose-built moored buoy. With the development of an analyst protocol, we have also formalized the review of detection data in near real time to substantially improve the accuracy of the system. This paper describes the moored buoy system and analyst protocol, and evaluates the accuracy of near real-time whale occurrence estimates derived from a buoy located near the Massachusetts coast. This evaluation compares occurrence estimates derived in near real time to those derived from (1) a review of simultaneously collected archived audio and (2) visual sightings collected by aerial surveys for humpback (*Megaptera novaeangliae*), sei (*Balaenoptera borealis*), fin (*Balaenoptera physalus*), and North Atlantic right whales (*Eubalaena glacialis*).

2.2 Materials and Methods

2.2.1 System Overview

A moored buoy was designed to deliver detection data in near real time from a passive acoustic instrument on the sea floor to a shore-side computer where an analyst could review the data to determine baleen whale occurrence. The passive acoustic instrument used in this system was the digital acoustic monitoring (DMON) instrument that is capable of running the low frequency detection and classification system (LFDCS) firmware developed to identify baleen whale calls. The mooring hardware allowed the delivery of power and data between the sea floor and a surface buoy via stretch hoses that isolated the motion of the surface buoy from an aluminum frame on the sea floor to which the DMON was attached. The surface buoy contained a platform computer to store DMON/LFDCS data and to transmit these data to shore every 2 hours via the Iridium satellite system. Upon reception, the DMON/LFDCS detection data were immediately displayed on a publicly accessible website and were reviewed once a day by an analyst. The results of the analyst review were posted on the website and disseminated automatically to researchers, managers, the United States Coast Guard, and other stakeholders via email and text messages.

2.2.2 Digital acoustic monitoring instrument

The DMON instrument is an acoustic hardware device that can (1) sample from up to three integrated hydrophones, (2) process and record the resulting audio with a programmable Texas Instruments TMS320C55 digital signal processor (DSP) and 32 GB of flash memory, and (3) communicate detection (and other) information to an external computer using serial input/output lines (Johnson & Hurst 2007; Baumgartner et al. 2013. The instrument is extremely low power, making it ideal for use on power-limited autonomous platforms. For the application described here, the electronics board, integrated lithium battery, and hydrophones were packaged in an oil-filled, acoustically transparent urethane housing.

The three hydrophones available for use with the DMON cover low- (8-7500 Hz), mid- (0.1-50 kHz), and high-frequency (1-160 kHz) bands. For our study, only the low-frequency hydrophone was used (WHOI custom-built, end-capped cylinders with Navy Type II ceramics) with a low-power and low-noise preamplifier (20 dB gain), an additional user programmable gain (4.6 and 13.2 dB available, selected 13.2 dB for this study), and a 6-pole Sallen-Key anti-alias filter. The hydrophone had a flat frequency response in the 8-7500 Hz band, 36 dB re μPa/√Hz noise floor at 2 kHz, and -169 dB re V/μPa sensitivity at 2 kHz (Baumgartner et al., 2013). Audio from the low-frequency hydrophone was digitized with a 16-bit analog-to-digital converter at 60 kHz, low-pass filtered and decimated to 2000 Hz, and spectrograms were created in real time using a short-time Fourier transform with a 512-sample frame, Hann window, and a frame-to-frame overlap of 384 samples (75%), yielding a spectrogram time step of 64 milliseconds and frequency resolution of 3.9 Hz.

2.2.3 Low-frequency detection and classification system

The LFDCS was originally developed to detect and classify the tonal sounds of baleen whales in archived audio (Baumgartner & Mussoline, 2011), but was later ported to run on the DMON instrument in real time (Baumgartner et al., 2013). Detailed descriptions of the LFDCS can be found in Baumgartner and Mussoline (2011) and Baumgartner et al. (2013), but briefly, the system builds a spectrogram using the short-time Fourier transform, creates pitch tracks of tonal calls in the spectrogram (a pitch track is a time series of frequency-amplitude pairs that describes a sound in a manner analogous to a series of notes on a page of sheet music), and classifies each call by comparing attributes of the pitch track to those of known call types in a call library using quadratic discriminant function analysis. For the application described here, audio was sampled at 2000 Hz, compressed using a lossless algorithm (Johnson, Partan, & Hurst, 2013) and archived to flash memory on a 50% duty cycle (30 minutes every hour). These recordings were accessible upon recovery of the mooring and used to evaluate the accuracy of near real-time detections.

During operation aboard an autonomous platform, the DMON/LFDCS regularly relays summary detection data, detailed detection data, status information (e.g., system voltage, available memory), and background noise estimates to the platform computer. Summary detection data consist of tallies of classified calls for every call type in the call library, which are relayed to the platform computer every 15 minutes (review and evaluation of detection data, described below, are organized in these 15-minute tally periods). Detailed detection data are sent to the platform computer in real time and include pitch tracks and associated classification information, but only up to a maximum of 8 kilobytes of detection data per hour. This data transmission limitation is designed solely to reduce operating costs by limiting (1) the amount of data sent through the Iridium satellite service and (2) time spent by the analyst reviewing pitch track data; however, the data transmission rate is configurable and can be eliminated altogether if the associated transmission and analysis costs can be accommodated.

2.2.4 Quiet mooring

We utilized a mature mooring design that allowed both quiet operation as well as delivery of digital data from the sea floor to shore (Figure 2.1). The DMON was housed in open cell foam and a urethane fairing and affixed to a bottom-mounted aluminum frame called the multi-function node (MFN), which in turn was attached to the surface buoy by stretch hoses. These hoses can stretch to nearly twice their relaxed length (Paul & Bocconcelli, 1994), thereby absorbing the motion of the buoy in rough wave conditions and keeping the MFN acoustically quiet. The hoses also contain helically wound conductors that allow power and data to be delivered between the buoy and the DMON. The surface buoy contains a platform computer, Iridium and global positioning satellite (GPS) antennas, and a 450-Ahr battery pack to power all system components. The platform computer receives and stores DMON/LFDCS data sent in real-time via the stretch hoses, and once every 2 hours, transmits these stored data to shore via an Iridium satellite modem (Figure 2.2). The buoy was designed to operate at sea for at least one year.

2.2.5 Near real-time analysis

All data are received by a dedicated shore-side server, immediately processed, and displayed on a publically accessible website (dcs.whoi.edu) for review by an experienced analyst (Figure 2.2). For each 15-minute tally period for which detailed detection data were transmitted, pitch track data and associated classification information are displayed on a single webpage in stacked 1-minute panels (e.g., Figures 2.3a, 2.4a, 2.5a). The analyst reviews these data and fills out a form on the webpage for each monitored 15-minute period to indicate whether each of the monitored species was “detected”, “possibly detected”, or “not detected” during the tally period; the form allows the entry of notes as well.

The analyst uses a standardized and documented protocol (available at dcs.whoi.edu/#protocol) developed jointly by the National Oceanic and Atmospheric Administration's Northeast Fisheries Science Center (NOAA NEFSC) Passive Acoustics Group and the Woods Hole Oceanographic Institution to determine how a tally period should be scored. In general, a tally period is scored as "detected" when there is convincing evidence of a species' acoustic presence, "possibly detected" when there is some evidence of acoustic presence, but the evidence is not completely convincing, or "not detected" when there is no reasonable evidence of a species' acoustic presence. We chose to emphasize minimizing false detections when developing the protocol, so the analyst is encouraged to be conservative (i.e., cautious) by only scoring tally periods as "detected" when there is strong evidence of acoustic presence.

Pitch tracks are an abstraction of the audio and spectrograms from which they are derived and are used to determine species presence because the audio and spectrograms are not available to the analyst in near real time. Despite being an abstraction, the analyst uses similar approaches to evaluating pitch tracks to those used when evaluating sounds in audio and spectrograms. Specifically, the characteristics of the call itself as well as the context of the call (i.e., sounds temporally adjacent to the call) are evaluated to decide if a call is part of a noise process or is genuinely produced by a particular species. When evaluating pitch tracks, the analyst protocol (dcs.whoi.edu/#protocol) instructs the analyst to evaluate putative whale calls using four characteristics: (1) shape, (2) amplitude, (3) isolation, and (4) classification.

1. *Shape:* The shape of a pitch track refers to how quickly and smoothly its frequency changes with time; poorly shaped pitch tracks have a jagged appearance with rapid changes in frequency with time. While some call types have inherent discontinuities or rapid changes in frequency with time (e.g., Antarctic blue whale z call, some humpback whale rapid up- or downsweeps), none of the call types used in our study for fin, sei, or North Atlantic right whales had these characteristics.

2. *Amplitude:* While legitimate calls can have a range of amplitudes, many noise processes produce pitch tracks with low amplitudes. For example, the noise from a distant passing ship may not be fully excluded with the tonal noise reduction step in the LFDCS algorithm, thereby leaving some low signal-to-noise ratio sounds available for pitch tracking. The resulting pitch tracks would have low amplitude, and by chance, some may take the shape of a genuine whale call. Thus, low-amplitude pitch tracks sometimes require extra scrutiny.

3. *Isolation:* Because noise processes often produce many spurious pitch tracks, the assessment of a pitch track of a putative whale call should include evaluation of all other pitch tracks in a temporal window adjacent to the call. In practice, if there are many pitch tracks that (1) surround the putative call and (2) are clearly not biological (because of their shape and lack of patterning), then the analyst should be skeptical of the putative call.

4. *Classification:* The DMON/LFDCS provides classification information for putative whale calls that is available to the analyst for evaluation. Classification of a call to a species' call type indicates that the pitch track has the correct shape for that call type. For fin, sei, and North Atlantic right whales, whose calls are stereotypical, the classification information is an objective assessment of how well the call conforms to a call type, and therefore helps minimize subjectivity.

Typically, if a putative call is lacking in three or four of these characteristics, it would not be considered genuine, whereas calls lacking in one or two of these characteristics might be considered more likely to be genuine depending on how far those characteristics deviate from the norm. Examples of right whale upcall pitch tracks are shown in Figure 2.6 to illustrate these characteristics. Figure 2.6a shows upcalls that have good shape (i.e., are well-formed with no rapid changes in frequency) and high amplitude, are isolated from pitch tracks that are associated with noise, and are classified as upcalls. Displaying all four characteristics, there is significant evidence of these calls being real (i.e., not produced by noise). Figure 2.6b, in contrast, shows a putative upcall that has been classified, but has poor shape, low amplitude, and is surrounded by pitch tracks that are very likely associated with noise. There is little evidence that this putative call is real, while it is more likely it was produced by whatever noise process produced the other pitch tracks.

Our evaluation protocol does not rely on the presence of a single call to determine occurrence, but instead requires several calls to be present during a tally period. Criteria to score tally periods as "detected", "possibly detected", and "not detected" are described for each of the whale species in the analyst protocol (dcs.whoi.edu/#protocol), but we will summarize the criteria for right whales here as an illustration of how the assessment of individual pitch tracks is used to determine a score for a tally period. A tally period should be scored as "detected" for right whales if three or more genuine upcalls are identified and one or more of these upcalls is classified as a right whale upcall. If pitch tracks associated with humpback whale singing are present, the analyst should check to see if the right whale calls are off the "rhythm" of the humpback song (i.e., right whale calls are not part of a recurring pattern) and/or there are significant differences in amplitude between the humpback song and the right whale upcalls (e.g., Figure 2.3). A tally period should be scored as "possibly detected" if one to two classified calls or three or more unclassified calls are identified. If two or fewer unclassified calls (including no calls) are present, then the analyst should score the tally period as "not detected." These criteria are not prescriptive, but are guidelines for the analyst; minor deviations from the criteria are allowed by experienced analysts exercising good judgement when appropriate, well justified, and documented (in the notes section of the scoring webpage).

The analyst reviews individual pitch tracks, associated classification information, and the context in which individual pitch tracks occur to assess species occurrence. Three of the four species monitored for this study make calls in distinct patterns that can be easily discerned in the pitch track displays. These include humpback whale song (Figure 2.3), sei whale low-frequency doublets or triplets (Figure 2.4), and fin whale 20-Hz pulse sequences (Figure 2.5). Assessing context (i.e., pitch tracks in temporal proximity to a pitch track of interest) is particularly helpful when identifying right whale upcalls, which can be confused with a similar upsweep sometimes present in humpback whale song (authors' personal observations).

In practice, the analyst reviewed detection data for this study once a day, usually between 0700 and 1000 local time, and the resulting near real-time occurrence estimates were displayed on the website within minutes of the analyst's review. The near real-time occurrence estimates were also (1) distributed directly to interested users via email and text messages, eliminating the need for users to check the website constantly, (2) made available in Whale Alert (www.whalealert.org), a smartphone/tablet app for iOS and Android platforms, and (3) viewable in the U.S. Coast Guard's One View software to easily allow Coast Guard personnel to monitor whale presence.

2.2.6 Evaluation of real-time occurrence estimates with archived audio

Contemporaneous estimates of whale occurrence derived from the DMON/LFDCS recorded audio were used to assess the accuracy of whale occurrence estimates derived in real time. For the 2015-2016 buoy deployment near Nomans Land Island, Massachusetts described below, all 15-minute tally periods with both audio available and at least 3.75 minutes of detailed real-time detection data available between 24 March 2015 and 31 August 2015 were retrospectively analyzed for species occurrence in the recorded audio. This 5-month period (March-August 2015) was chosen to span the time when all four of the monitored species were present and to make the manual audio analysis manageable. Spectrograms and audio were reviewed visually and aurally, respectively, to determine species occurrence during the entirety of each 15-minute tally period (regardless of the duration that the same tally period was actually monitored in near real time). Like in the near real-time analysis, each 15-minute tally period in this audio analysis was scored as “detected”, “possibly detected” or “not detected” based on how convincing the acoustic evidence was. We assessed the accuracy of the near real-time analysis by treating the retrospective audio analysis as the truth and comparing the results of the two analyses using confusion matrices. Only periods scored as either “detected” or “not detected” in both the near real-time and audio analysis were assessed (periods scored as “possibly detected” in either the near real-time or audio analysis were assessed separately). A variety of performance metrics were used to quantify the accuracy of the near real-time analysis (Figure 2.7). Cases in which there was disagreement between the near real-time and retrospective audio analyses were examined to determine the reason for the disagreement. Finally, logistic regression was used to determine if the probability of missed occurrences in near real time was related to the amount of daily calling activity.

2.2.7 Evaluation of real-time occurrence estimates with visual sightings

The accuracy of near real-time whale occurrence estimates was also evaluated with whale sightings collected by aerial surveys conducted near the DMON buoy. Comparison of occurrence estimates derived from passive acoustics and visual observations is challenging because of the significant differences in the detectability of whales between the two methods. Neither passive acoustics nor visual surveys are perfect detection systems; nevertheless, when one system correctly detects a whale, there is a reasonable expectation that the other system should detect it as well. We compared occurrence estimates derived from aerial surveys flown near the Nomans Land buoy site to the near real-time passive acoustic occurrence estimates derived from the buoy using log odds ratio tests. Aerial surveys were conducted by the New England Aquarium (NEAq) and the NOAA NEFSC using standard large whale survey protocols (two observers on either side of the plane, 229-305 m altitude, 185 km hr^{-1} speed).

Visual occurrence was evaluated on a daily basis within particular radii of the buoy for the aerial survey observations (within 20-60 km in 10 km increments), and acoustic occurrence was evaluated within particular time intervals before the start of the aerial survey for the near real-time passive acoustic observations from the buoy (within 12-72 hours in 12 hour increments; note that only the period before a survey was examined so that acoustic occurrence prior to the survey could be used prospectively to predict visual occurrence during the survey – see end of this paragraph). The log odds ratio test evaluates the ratio of the odds of acoustic detection when a species is visually present to the odds of acoustic detection when a species is visually absent. The log odds ratio was evaluated using a logistic regression between the near real-time passive acoustic observations (dependent variable) and the visual observations (independent variable). To account for multiple comparisons over several radii and time intervals, we used a Bonferroni adjusted alpha threshold of 0.00167 ($\alpha_{Bonferroni} = \alpha \div 5$ radii $\div$ 6 time intervals, where $\alpha = 0.05$) to determine the significance of log odds ratios. In addition to comparing daily occurrence estimates, we also used logistic regression to assess whether the probability of detecting a species during an aerial survey was related to the percentage of near real-time tally periods scored as “detected” within 12-72 hours prior to the start of the survey.

2.2.8 Statistical treatment

Whenever percentages were used in correlation or regression analyses, they were transformed using the arcsine square-root transform: $\hat{X} = \sin^{-1}\left[\sqrt{\frac{X}{100}}\right]$ (Sokal & Rohlf, 1995). Axes of transformed values were back-transformed into percentages for clarity. Regression analyses were deemed appropriate based on evaluating linearity (logistic/linear), normality (linear only), and homoscedasticity (linear only) using scatterplots, binned logit scatterplots, and histograms.

2.3 Results

A moored DMON/LFDCS buoy was deployed 9 km southwest of Nomans Land Island near Martha's Vineyard, Massachusetts, USA (41.1418, -70.9292; Figure 2.1) from the M/V *Scarlett Isabella* in 34 m water depth on 24 March 2015. The mooring was recovered by the R/V *Tioga* on 30 March 2016. A second, identical moored buoy was deployed in the same location on 28 September 2016 by the R/V *Armstrong* and recovered on 19 October 2017 by the M/V *Scarlett Isabella*. The 2015-2016 moored buoy was used for both the audio and visual evaluations, while the 2016-2017 moored buoy was used for the visual evaluation only.

From 24 March 2015 to 30 March 2016, the DMON/LFDCS generated 7,379,987 pitch tracks with associated classification information, of which 1,464,471 (20%) were delivered to the buoy's platform computer for transmission to shore (owing to the 8-kilobyte limit on transmitting detailed detection data). A total of 11,239 tally periods were reviewed in near real time for the 2015-2016 deployment, and the retrospective analysis of archived audio was conducted for 4,606 of these tally periods (selected as all 15-minute tally periods that occurred between 24 March and 31 August 2015 that had audio available and at least 3.75 minutes of pitch track data transmitted in near real time). The NEAq and NOAA NEFSC conducted 22 and 14 flights, respectively, near Nomans Land Island while the buoy was operational in 2015 and 2017. Each flight flew 222-1298 km of trackline within 60 km of the buoy in conditions of Beaufort 5 or less and visibility greater than 3 km.

2.3.1 Evaluation of real-time occurrence estimates with archived audio

Comparisons between occurrence estimates determined in near real time and those determined during the audio analysis indicated remarkably low false detection rates (Tables 2.1 and 2.2). Of all the species, only fin whales had a false detection, and this occurred in only a single 15-minute period. Missed detections rates for all species ranged from 27 to 67% during 15-minute tally periods and 12-42% over daily time scales (Table 2.2). Fifteen-minute tally periods were scored in near real-time as "not detected" when there was evidence of acoustic presence in the archived audio for several reasons (Table 2.3). For right whales, the most common reason (67% of missed detections) was because upcalls occurred after the 8-kilobyte per hour limit was reached and before the end of the 15-minute tally period (i.e., upcalls were available for the audio analyst to detect, but not available for the near real-time analyst to detect). More often for other species, tally periods were scored as "not detected" in near real time because calls were poorly pitch tracked (or not pitch tracked at all) owing to low amplitude or interfering sounds (e.g., other whales, vessel noise). For fin whales, which require the detection of several 20-Hz pulses with a constant inter-pulse interval, tally periods were often scored as "not detected" because not enough pulses were identified in near real time to be confident of the species' presence. Over daily time scales, the probability of missed detection was significantly related to the amount of calling activity, measured as the percentage of tally periods that were scored as "detected" in the audio analysis during a single day (Figure 2.8). Fitted logistic regression models suggested that if 12, 33, 19, and 22% or more tally periods were scored as "detected" during a day in the audio analysis (i.e., if observed calling rates were modest or high), then the probability of daily missed detections in near real time dropped to 10% or less for right, humpback, sei, and fin whales, respectively (i.e., then the chance of missing occurrence in near real time was low) (Figure 2.8).

The vast majority of tally periods scored as "possibly detected" in near real time for right, humpback, and fin whales were scored as "detected" during the audio analysis (Table 2.4). Together with the very low false detection rates, this indicates that the analyst was quite cautious in scoring periods as "detected" (as encouraged by the protocol). For sei whales, roughly half of the tally periods scored as "possibly detected" were determined to have evidence of sei whale presence in the audio analysis.

The time series of "detected" and "possibly detected" scores determined from the near real-time analysis closely mirrored the time series determined from the audio analysis (Figure 2.9a-d), suggesting that variability in the near real-time assessment of occurrence is nearly as accurate as one can derive from a manual audio analysis. The percentage of near real-time detections per day was significantly correlated with the percentage of detections per day from the audio analysis for all species ($p < 0.0001$; Figure 2.9e-h). These correlations were particularly high for right and fin whales ($r^2 = 0.904$ for both). Slopes of the corresponding regressions were less than 1 for all species, indicating that acoustic detection rates were underestimated in near real time. This is not surprising considering the missed detection rates described previously.

2.3.2 Evaluation of real-time occurrence estimates with visual sightings

Of all the species examined, the best agreement between visual and near real-time acoustic detections was observed for right whales (Figure 2.10a; Table 2.5). The log odds ratio test was significant ($p < \alpha_{Bonferroni}$) for most radii and time intervals, but the best agreement between visual and acoustic occurrence was within 30-40 km of the buoy and 24-48 hours prior to an aerial survey (Table 2.5). For example, within 48 hours of an aerial survey, right whales were acoustically detected on 13 of the 14 days (92.9%) when right whales were detected within 40 km of the buoy by the aerial surveys, and right whales were acoustically detected on only 3 of the 22 days (13.6%) when right whales were not detected by the aerial surveys, yielding a log odds ratio of 4.41 (95% CI: 1.95-6.87; $p < 0.0001$; Table 2.5). In contrast to right whales, there were no associations observed between visual and near real-time acoustic detections of humpback whales at any radii or time interval (Figure 2.10b; Table 2.5).

Sei whale occurrence estimates from aerial surveys and near real-time passive acoustic monitoring were significantly associated only at 30 and 40 km radii around the buoy, and only within 24 hours of an aerial survey (Figure 2.10c; Table 2.5). Acoustic detection rates were modest when sei whales were encountered by the aerial surveys (6 of 9 days; 66.7%), but acoustic detection rates were appropriately low when sei whales were not encountered by the aerial surveys (3 of 27 days; 11.1%). Fin whale occurrence estimates from aerial surveys and near real-time passive acoustic monitoring were significantly associated within 40 km of the buoy and 24, 36, and 72 hours prior to an aerial survey (Figure 2.10d; Table 2.5). Fin whales were most often (10-11 of 11 days; $\geq 91\%$) acoustically detected on days when they were sighted by the aerial surveys within 40 km of the buoy, but fin whales were also acoustically detected when the aerial surveys did not encounter fin whales (7-14 of 25 days; 28-56%).

The probability of encountering right, sei, and fin whales during an aerial survey was significantly associated with near real-time acoustic detection rates of those species prior to the aerial survey (Figures 2.10e,g,h; Table 2.6), but there was no such association for humpback whales (Table 2.6). Fitted logistic regression models suggested that detecting right whales in just 1-4% of all reviewed tally periods within 24-72 hours prior to an aerial survey was associated with a 50% probability of encountering a right whale within 30-50 km of the buoy during the aerial survey (Figure 2.10e; Table 2.6). Similarly, detecting right whales in just 6-15% of tally periods over 24-72 hours prior to an aerial survey was associated with a 90% probability of encountering a right whale within 30-50 km of the buoy during the aerial survey (Table 2.6). Logistic regression models suggested that detecting sei whales in 4-6% and 13-20% of tally periods over 48-72 hours prior to an aerial survey was associated with a 50% and 90% probability of encountering a sei whale within 30-60 km of the buoy during the aerial survey, respectively (Figure 2.10g; Table 2.6). Detecting fin whales in 16-17% and 61-64% of tally periods within 24-48 hours prior to an aerial survey was associated with a 50% and 90% probability of encountering a fin whale within 40 km of the buoy during the aerial survey, respectively (Figure 2.10h; Table 2.6).

2.4 Discussion

The mooring design was successful, allowing for quiet continuous operation for two yearlong deployments in an area that is exposed to intense New England storms and oceanic swell owing to unlimited fetch from the south. Near real-time false detection rates were virtually zero, indicating that when a tally period is scored as “detected”, the analyst is nearly 100% correct. Such high accuracy in near real-time is attributable to (1) having an analyst as part of the detection process, and (2) having a protocol that stresses conservatism in scoring. The greatest advantage of having an analyst review the detection data is the assessment of context. The human analyst can consider context in a way that is not yet available in automated detection and classification systems for marine mammal sounds. Many automated detectors attempt to determine species presence based on a single call with little or no regard for the noise environment, other sounds in temporal proximity to a call of interest, or patterning in calls. An analyst can take such contextual information into account (when reviewing either archived audio or time series of pitch tracks, e.g., Figure 2.3), which increases accuracy significantly. The need for low false detections provided by the analyst must always be weighed against the cost of the analyst; we found in our study that the analyst spent about 30-45 minutes per day per platform reviewing pitch tracks and scoring tally periods.

We have developed a protocol that encourages the analyst to score “detected” only with convincing evidence of a species’ acoustic presence. It is important to recognize that the protocol was designed a-priori to minimize false detections by encouraging the analyst to be conservative (i.e., cautious) in their analysis. We chose to do this because marine mammal mitigation applications often have significant costs associated with false detections. For example, stopping construction activities or at-sea training exercises, slowing ships down, or moving fishing operations in response to marine mammal presence have substantial costs, so we sought to minimize these costs by minimizing false detections. An equally compelling argument can be made that mitigation should be precautionary, emphasizing protection of whales over cost to industry or government; thus, missed detections should be minimized at the expense of false detections. These are policy decisions that can affect how the protocol is developed. For applications that require a much lower missed detection rate than what we observed here (with accompanying higher false detection rates), a relaxation of the protocol may be warranted. For example, the protocol could be changed such that an analyst would score a tally period as “detected” if there was any evidence of a species' acoustic presence (rather than convincing evidence, which was the criterion used in this study). Such a change in the protocol for our study would have resulted in a modest increase in false detection rates for right, humpback, and fin whales, but a substantial increase in false detection rates for sei whales (Table 2.4).

The daily missed detection rate for right whales was 27% (Table 2.2), but observed daily calling rates (i.e., rates of received calls) were low on all of the 7 days when presence was missed in near real time (i.e., on days when less than 10% of tally periods were scored as "detected" during the audio analysis; Figure 2.8a). For such days with low calling rates, right whales are only acoustically available to be detected in near real time in just a few tally periods, and with missed detection rates of 42% for individual 15-minute tally periods (Table 2.2), positive detection is not always possible. Unlike humpback or fin whales that are prodigious callers once a calling bout is initiated, right whales often produce sporadic upcalls in low numbers without pattern, so the opportunities for detection are fewer. However, if the 15-minute missed detection rate is constant and missing detections in a tally period is independent of missing detections in the next tally period, then the probability of missing right whale acoustic presence in two tally periods in a day is $42\% \times 42\% = 17.6\%$, and the probability of missing right whale presence in three tally periods in a day is $(42\%)^3 = 7.4\%$. Hence, as calling rates increase, we expect the probability of daily missed detection to decrease, even for a sporadic caller (as observed in Figure 2.8a). There may be some benefit to sending more pitch track data than allowed by the 8-kilobyte limit used in this study, since many missed detections at the 15-minute time scale were caused by the cessation of pitch track transmission (Table 2.3); however, the additional costs of data transmission and analyst time must be weighed against the potential reduction in missed detection rates at the 15-minute time scale, which presumably will help to lower the missed detection rate at the daily time scale. The strong association between acoustic detections and aerial survey sightings (Figure 2.10a) suggest that silent right whales were not especially problematic for estimating occurrence of the species in the study area; when right whales were seen, they were typically also heard, particularly over 30-40 km spatial scales and 24-48 hour temporal scales.

Humpback whales had higher missed detection rates than right whales at 37% on daily time scales (Table 2.2), but missed detections were also strongly associated with low calling rates (Figure 2.8b). Our analysis of missed detections at the 15-minute temporal scale suggested that faint calling was often the cause of missed humpback whale detections (Table 2.3). Because humpbacks are most easily identified by the numerous patterned calls that make up their songs, faint singing is more detectable in spectrograms by an analyst than if they made few sporadic calls like right whales (i.e., the faint pattern can be recognized in the spectrograms better than a faint single call). While an analyst can identify this faint singing, such faint song units are difficult to pitch track. Therefore, it is likely that the higher missed detection rate for humpbacks is attributable to the difference in detection capabilities of a human and the pitch-tracking algorithm. Although there was a significant correlation between occurrence estimates derived from the near real-time and audio analyses (Figure 2.9f), there was no association between acoustic detections and visual sightings (Figure 2.10b). We suspect that this lack of association is related to our use of song to identify humpback whales acoustically. Song is produced by males (Payne & McVay, 1971), and one can imagine a situation where a single male is near the buoy singing; this single animal is easily detected acoustically, but difficult to detect visually during an aerial survey. Conversely, one can imagine a group of several females that are easy to detect visually, but very difficult to detect acoustically since none of the females are singing. Hence when using song for humpback whale detection, there may not be a strong relationship between what one hears and what one sees.

Sei whales had the highest missed detection rates of any of the other species on both 15-minute and daily time scales (Table 2.2). No one factor stood out strongly as the reason for these higher missed detection rates at the 15-minute temporal scale (Table 2.3), but like all of the other species, missed detections in near real time were strongly associated with low calling rates at the daily time scale (Figure 2.8c). Sei whales produce sporadic calls in low numbers like right whales, but sometimes with very short patterns (doublets or triplets; Baumgartner et al. 2008; Figure 2.4). Near real-time occurrence estimates were significantly correlated with occurrence estimates derived from the audio analysis (Figure 2.9g), and were significantly associated with sightings at 30-40 km spatial scales and 24-hour time scales (Figure 2.10c). Acoustic detections were appropriately low when sei whales were not sighted during aerial surveys, but acoustic detections were only modest when sei whales were encountered during aerial surveys. This could certainly be a consequence of missed detections, but also silent animals.

Fin whales had the lowest missed detection rates of any of the other species on both 15-minute and daily time scales (Table 2.2). Fin whales call in trains of 20-Hz pulses that are separated by a nearly constant inter-pulse interval (Watkins et al., 1987; Morano et al., 2012). The pattern of these pulses is easily recognized both in an audio analysis and in pitch tracks when correctly classified (Figure 2.5). We rely strongly on the automated classification of 20-Hz calls since the frequency resolution of the spectrogram used by the DMON/LFDCS in the 20-Hz call band is very coarse. When calls are not classified because of interfering sound (including calls from other fin whales) or a clear pattern with a constant inter-pulse interval is not apparent, our protocol encourages the analyst to be skeptical. As with the other species at daily time scales, missed detections in near real-time were strongly associated with low calling rates (Figure 2.8d), but there was very good agreement between daily calling activity derived from the audio and near real-time analyses (Figure 2.9h). There was a significant association between acoustic and visual occurrence estimates at 40 km spatial scales and 24-72 hour temporal scales (Figure 2.10d). At these scales, fin whales were nearly always acoustically detected when sighted by the aerial surveys, but they were also acoustically detected when not seen by the aerial surveys. This pattern could be caused by false detections, but we observed that the near real-time false detection rate is nearly 0% for fin whales (Table 2.2). It is more likely that fin whales sometimes go undetected by the aerial surveys, perhaps because they do not often aggregate in large groups (Hain et al., 1992) or their acoustic detection range exceeds the spatial scales examined here (> 60 km; we are unaware of published acoustic detection range estimates for fin whales in shallow neritic waters, so this hypothesis is currently difficult to address). Interestingly, fin whale 20-Hz pulse trains are thought to be a reproductive display by males (Croll et al. 2002) like humpback singing, but the association between acoustic and visual occurrence estimates for fin whales was much stronger than that for humpback whales.

The acoustic detection range for the monitored species is much lower than the spatial scales at which we observed significant associations between aerial survey and near real-time acoustic occurrence estimates. Right whales are estimated to have detection ranges of up to 9 km in shallow continental shelf waters (Clark, Brown, & Corkeron, 2010), and humpback whales, producing calls at similar frequencies and source levels as right whales (Clark et al., 2008-2010; Thompson, Cummings, & Ha, 1986; Au et al., 2006), likely have a similar detection range. Sei whales produce lower frequency calls at louder source levels (Baumgartner et al., 2008; Newhall et al., 2012), and Baumgartner et al. (2008) estimated an acoustic detection range of 10-15 km. Fin whales produce the loudest and lowest frequency calls of all of the species studied here (Charif et al., 2002), and may have detection ranges of several tens of kilometers in shallow neritic waters. With detection ranges of 9-15 km for right, humpback, and sei whales, why would the best associations between acoustic and visual occurrence estimates be observed at 30-40 km? While the instantaneous detection range of the buoy may be 10-20 km for these three species, whales move over the time scales of the analysis presented here (e.g., 24-48 hours), so the time and location when they are acoustically detected is rarely the same time and location when they are visually detected. A whale that is calling near the buoy on one day may be 30 km away on the next day when it is detected by the aerial survey. The implications of this are important. If acoustic detections are to be used for mitigation over time scales of a few days, then the movement of whales must be taken into account. The spatial scale over which there are significant associations between aerial and acoustic occurrence estimates can be thought of as the monitoring range of the acoustic system, which is different from its detection range. We define the monitoring range as the area over which whales that are acoustically detected will move over a specified time scale (see supporting information). It is dependent on short-term (tens of hours to days) movement behavior, of which we know little for whales, but we have estimated the monitoring range empirically here using associations between acoustic detections and visual sightings.

We chose to describe the area over which whales move around a stationary acoustic instrument in a specified time window as the "monitoring range" of the instrument, which is different from the acoustic detection range of the instrument. We use the word "monitoring" in this context in the same way the word monitoring would be used to describe the survey efforts of an airplane. In both cases, the term "monitoring" describes a process of detecting whales that occurs over both time and space. An aerial survey has an instantaneous detection range of hundreds of meters to a few kilometers to either side of the trackline (depending on species' detectability), but it moves over that trackline and can therefore detect whales over thousands of square kilometers in several hours. While a whale on the trackline may not be detectable by the aerial survey at time A because the whale is not in the detection range of the airplane at that time, that same whale can be detected at time B when the plane has moved such that the whale is now within its detection range. The same principle applies for a stationary acoustic instrument: a whale that is outside the acoustic detection range at time A will later be detectable when it moves into the acoustic detection range of the instrument at time B. Just as we would refer to the aerial survey as monitoring for whales during its flight (i.e., when the movement of the plane is considered), a stationary acoustic instrument can similarly monitor for whales when the movement of the whales is taken into account. For this reason, we chose the term "monitoring range" to refer to the spatial scale of this monitoring over a given time scale.

The near real-time estimates of occurrence from the DMON/LFDCS buoy were accurate, producing false detection rates of 0% for right, humpback, and sei whales, and nearly 0% for fin whales. The analysis protocol was purposely designed to be conservative to produce low false detection rates for marine mammal mitigation applications at the expense of higher missed detection rates. There are several U.S. Coast Guard gunnery training ranges near Nomans Land Island, and the DMON/LFDCS buoy was used to deliver near real-time detections of right whales directly to the Coast Guard operations center in Woods Hole, Massachusetts to aid in minimizing interactions between Coast Guard vessels and right whales during training exercises. In addition to reducing ship strike risks to right whales by postponing training exercises when whales were present, the system saved the Coast Guard time and mobilization costs by reducing the chances that right whales will be encountered during an exercise, which would force the immediate cancellation of the exercise and the return of the training ships to port. We hope to expand the use of the system in the near future for other applications, including mitigating ship strikes in areas heavily trafficked by commercial ships and noise exposure during wind farm construction.

Table 2.1. Confusion matrices comparing near real-time analysis to audio analysis for right, humpback, sei, and fin whales over 15-minute and daily time scales.

		Audio analysis				
		15-minute			**Daily**	
Species	***Near real-time analysis***	**Detected**	**Not detected**		**Detected**	**Not detected**
Right	Detected	87	0		19	0
	Not detected	62	4304		7	110
Humpback	Detected	192	0		39	0
	Not detected	303	3852		23	68
Sei	Detected	63	0		31	0
	Not detected	129	4244		22	77
Fin	Detected	1036	1		99	0
	Not detected	390	2951		13	41

Table 2.2. Performance metrics for the near real-time analysis (when treating the audio analysis as the truth) over 15-minute and daily time scales.

Performance metric	Right	Humpback	Sei	Fin
15-minute				
False detection rate (%)	0.0	0.0	0.0	0.1
False positive rate (%)	0.0	0.0	0.0	0.0
False omission rate (%)	1.4	7.3	2.9	11.7
Missed detection rate (%)	41.6	61.2	67.2	27.3
Precision (%)	100.0	100.0	100.0	99.9
Recall (%)	58.4	38.8	32.8	72.7
Accuracy (%)	98.6	93.0	97.1	91.1
n (15-min periods)	4453	4347	4436	4378
Daily				
False detection rate (%)	0.0	0.0	0.0	0.0
False positive rate (%)	0.0	0.0	0.0	0.0
False omission rate (%)	6.0	25.3	22.2	24.1
Missed detection rate (%)	26.9	37.1	41.5	11.6
Precision (%)	100.0	100.0	100.0	100.0
Recall (%)	73.1	62.9	58.5	88.4
Accuracy (%)	94.9	82.3	83.1	91.5
n (days)	136	130	130	153

Table 2.3. Reasons for missed calls. Values are percentages of missed occurrences during 15-minute tally periods. Some tally periods had more than one reason for a missed occurrence, so columns do not add to 100%.

Reason for missed detection	Right	Humpback	Sei	Fin
Calls occurred after the 8 kB per hour data limit was reached	66.7	19.0	38.8	34.5
Calls were not pitch tracked at all because of low amplitude	12.1	81.9	53.2	18.4
Calls were not pitch tracked accurately/completely because of low amplitude	34.8	58.0	15.8	0.0
Calls were not pitch tracked accurately because of interfering sound	31.8	67.8	55.4	64.2
Not enough calls to trigger a “detected” or “possibly detected” score in near real time	7.6	14.7	3.6	64.2
Uncertainty due to interfering species calls	4.5	2.8	1.4	0.0
Human error (analyst chose wrong score erroneously)	1.5	0.0	0.0	3.5
Consecutive fin whale pulses were not classified for reasons other than faintness or background noise (e.g., two pulses were joined to create a longer pitch track)	—	—	—	13.4

Table 2.4. Audio analysis scores for 15-minute tally periods scored as “possibly detected” in near real time.

Audio analysis score	Right	Humpback	Sei	Fin
Not detected (%)	6.8	2.6	19.4	6.6
Possibly detected (%)	5.5	0.0	33.3	1.2
Detected (%)	87.7	97.4	47.2	92.2
n (15-minute periods)	73	76	36	167

Table 2.5. Results of log odds ratio test for aerial survey sightings at various radii around the buoy and near real-time passive acoustic detections at various time intervals prior to an aerial survey. An asterisk after a p-value indicates significance at the Bonferroni-adjusted alpha level ($p < \alpha_{Bonferroni}$ where $\alpha_{Bonferroni} = 0.00167$). Percentages of acoustically detected days are shown in Figure 2.10a-d. The term "whales" in the column titles indicate whales of the particular species in the table (e.g., "whales" refers only to right whales for the right whale rows in the table).

Radii (km)	Time interval (hours)	Number of surveys with whales seen	Of surveys when whales seen, % acoustically detected	Number of surveys with whales not seen	Of surveys when whales not seen, % acoustically detected	Log odds ratio	p-value
Right whale							
20	12	6	50.0	28	10.7	2.12	0.03806
20	24	6	83.3	28	17.9	3.14	0.00204
20	36	6	100.0	28	25.0	19.66	0.00021*
20	48	6	100.0	28	32.1	19.31	0.00070*
20	60	6	100.0	28	32.1	19.31	0.00070*
20	72	6	100.0	28	32.1	19.31	0.00070*
30	12	11	54.5	25	4.0	3.36	0.00056*
30	24	11	81.8	25	8.0	3.95	0.00001*
30	36	11	100.0	25	12.0	21.56	0.00000*
30	48	11	100.0	25	20.0	20.95	0.00000*
30	60	11	100.0	25	24.0	20.72	0.00000*
30	72	11	100.0	25	24.0	20.72	0.00000*
40	12	14	42.9	22	4.5	2.76	0.00417
40	24	14	64.3	22	9.1	2.89	0.00037*
40	36	14	78.6	22	13.6	3.15	0.00006*
40	48	14	92.9	22	13.6	4.41	0.00000*
40	60	14	92.9	22	18.2	4.07	0.00000*
40	72	14	92.9	22	18.2	4.07	0.00000*
50	12	16	37.5	20	5.0	2.43	0.01169
50	24	16	56.3	20	10.0	2.45	0.00219
50	36	16	68.8	20	15.0	2.52	0.00076*
50	48	16	81.3	20	15.0	3.20	0.00004*
50	60	16	81.3	20	20.0	2.85	0.00015*
50	72	16	81.3	20	20.0	2.85	0.00015*
60	12	18	33.3	18	5.6	2.14	0.02799
60	24	18	50.0	18	11.1	2.08	0.00909
60	36	18	61.1	18	16.7	2.06	0.00512
60	48	18	72.2	18	16.7	2.56	0.00054*

Table 2.5. (continued)

Radii (km)	Time interval (hours)	Number of surveys with whales seen	Of surveys when whales seen, % acoustically detected	Number of surveys with whales not seen	Of surveys when whales not seen, % acoustically detected	Log odds ratio	p-value
Right whale (cont.)							
60	60	18	72.2	18	22.2	2.21	0.00210
60	72	18	72.2	18	22.2	2.21	0.00210
Humpback whale							
20	12	7	57.1	27	44.4	0.51	0.54870
20	24	7	57.1	27	51.9	0.21	0.80233
20	36	7	57.1	27	55.6	0.06	0.93986
20	48	7	57.1	27	55.6	0.06	0.93986
20	60	7	57.1	27	55.6	0.06	0.93986
20	72	7	57.1	27	55.6	0.06	0.93986
30	12	9	55.6	27	44.4	0.45	0.56320
30	24	9	55.6	27	51.9	0.15	0.84702
30	36	9	55.6	27	55.6	0.00	1.00000
30	48	9	55.6	27	55.6	0.00	1.00000
30	60	9	55.6	27	55.6	0.00	1.00000
30	72	9	55.6	27	55.6	0.00	1.00000
40	12	10	60.0	26	42.3	0.72	0.34028
40	24	10	60.0	26	50.0	0.41	0.58924
40	36	10	60.0	26	53.8	0.25	0.73862
40	48	10	60.0	26	53.8	0.25	0.73862
40	60	10	60.0	26	53.8	0.25	0.73862
40	72	10	60.0	26	53.8	0.25	0.73862
50	12	13	53.8	23	43.5	0.42	0.54948
50	24	13	53.8	23	52.2	0.07	0.92309
50	36	13	61.5	23	52.2	0.38	0.58595
50	48	13	61.5	23	52.2	0.38	0.58595
50	60	13	61.5	23	52.2	0.38	0.58595
50	72	13	61.5	23	52.2	0.38	0.58595
60	12	14	50.0	22	45.5	0.18	0.79003
60	24	14	50.0	22	54.5	-0.18	0.79003
60	36	14	57.1	22	54.5	0.11	0.87842
60	48	14	57.1	22	54.5	0.11	0.87842
60	60	14	57.1	22	54.5	0.11	0.87842
60	72	14	57.1	22	54.5	0.11	0.87842

Table 2.5. (continued)

Radii (km)	Time interval (hours)	Number of surveys with whales seen	Of surveys when whales seen, % acoustically detected	Number of surveys with whales not seen	Of surveys when whales not seen, % acoustically detected	Log odds ratio	p-value
Sei whale							
20	12	4	50.0	30	6.7	2.64	0.03616
20	24	4	75.0	30	20.0	2.48	0.02886
20	36	4	75.0	30	26.7	2.11	0.06091
20	48	4	100.0	30	33.3	18.26	0.00500
20	60	4	100.0	30	43.3	17.83	0.01367
20	72	4	100.0	30	53.3	17.43	0.03171
30	12	9	33.3	27	3.7	2.57	0.02386
30	24	9	66.7	27	11.1	2.77	0.00141*
30	36	9	66.7	27	18.5	2.17	0.00823
30	48	9	77.8	27	25.9	2.30	0.00560
30	60	9	88.9	27	37.0	2.61	0.00459
30	72	9	88.9	27	48.1	2.15	0.02220
40	12	9	33.3	27	3.7	2.57	0.02386
40	24	9	66.7	27	11.1	2.77	0.00141*
40	36	9	66.7	27	18.5	2.17	0.00823
40	48	9	77.8	27	25.9	2.30	0.00560
40	60	9	88.9	27	37.0	2.61	0.00459
40	72	9	88.9	27	48.1	2.15	0.02220
50	12	11	27.3	25	4.0	2.20	0.05041
50	24	11	54.5	25	12.0	2.17	0.00823
50	36	11	54.5	25	20.0	1.57	0.04194
50	48	11	63.6	25	28.0	1.50	0.04429
50	60	11	72.7	25	40.0	1.39	0.06659
50	72	11	81.8	25	48.0	1.58	0.04964
60	12	11	27.3	25	4.0	2.20	0.05041
60	24	11	54.5	25	12.0	2.17	0.00823
60	36	11	54.5	25	20.0	1.57	0.04194
60	48	11	63.6	25	28.0	1.50	0.04429
60	60	11	72.7	25	40.0	1.39	0.06659
60	72	11	81.8	25	48.0	1.58	0.04964
Fin whale							
20	12	5	60.0	29	31.0	1.20	0.22145
20	24	5	80.0	29	41.4	1.73	0.10186

Table 2.5. (continued)

Radii (km)	Time interval (hours)	Number of surveys with whales seen	Of surveys when whales seen, % acoustically detected	Number of surveys with whales not seen	Of surveys when whales not seen, % acoustically detected	Log odds ratio	p-value
Fin whale (cont.)							
20	36	5	80.0	29	48.3	1.46	0.17446
20	48	5	80.0	29	48.3	1.46	0.17446
20	60	5	80.0	29	51.7	1.32	0.22218
20	72	5	100.0	29	62.1	18.07	0.03789
30	12	9	55.6	27	29.6	1.09	0.16683
30	24	9	88.9	27	33.3	2.77	0.00249
30	36	9	88.9	27	40.7	2.45	0.00807
30	48	9	88.9	27	44.4	2.30	0.01362
30	60	9	88.9	27	48.1	2.15	0.02220
30	72	9	100.0	27	59.3	18.19	0.00518
40	12	11	63.6	25	24.0	1.71	0.02369
40	24	11	90.9	25	28.0	3.25	0.00025*
40	36	11	90.9	25	36.0	2.88	0.00124*
40	48	11	90.9	25	40.0	2.71	0.00254
40	60	11	90.9	25	44.0	2.54	0.00493
40	72	11	100.0	25	56.0	18.32	0.00155*
50	12	14	50.0	22	27.3	0.98	0.16787
50	24	14	71.4	22	31.8	1.68	0.01878
50	36	14	71.4	22	40.9	1.28	0.07027
50	48	14	71.4	22	45.5	1.10	0.12186
50	60	14	71.4	22	50.0	0.92	0.19870
50	72	14	85.7	22	59.1	1.42	0.07997
60	12	14	50.0	22	27.3	0.98	0.16787
60	24	14	71.4	22	31.8	1.68	0.01878
60	36	14	71.4	22	40.9	1.28	0.07027
60	48	14	71.4	22	45.5	1.10	0.12186
60	60	14	71.4	22	50.0	0.92	0.19870
60	72	14	85.7	22	59.1	1.42	0.07997
60	72	14	85.7	22	59.1	1.42	0.07997

Table 2.6. Results of logistic regression analysis relating the probability of detecting whales within a particular radius of the buoy during an aerial survey to the percentage of near real-time acoustic tally periods scored as “detected” within a particular time interval prior to the start of the aerial survey. An asterisk after the p-value indicates significance at the Bonferroni-adjusted alpha level ($p < \alpha_{Bonferroni}$ where $\alpha_{Bonferroni} = 0.00167$). Significant logistic regression model fits are shown in Figure 2.10e-h.

Radii (km)	Time interval (hours)	Logistic regression slope	p-value	Periods acoustically detected to predict 50% probability of encounter during aerial survey (%)	Periods acoustically detected to predict 90% probability of encounter during aerial survey (%)
Right whale					
20	12	3.29	0.11584		
20	24	3.01	0.14787		
20	36	3.06	0.13342		
20	48	3.29	0.10098		
20	60	3.16	0.11299		
20	72	3.03	0.13381		
30	12	8.16	0.00317		
30	24	13.02	0.00005*	2.3	10.0
30	36	18.62	0.00001*	1.9	6.5
30	48	13.31	0.00002*	3.2	11.5
30	60	13.31	0.00001*	3.4	11.8
30	72	11.29	0.00005*	4.3	15.5
40	12	6.63	0.01370		
40	24	9.60	0.00089*	1.6	12.0
40	36	12.20	0.00033*	1.4	8.6
40	48	10.92	0.00016*	2.0	11.3
40	60	12.92	0.00004*	1.8	9.0
40	72	10.99	0.00010*	2.4	12.1
50	12	5.81	0.02966		
50	24	8.06	0.00379		
50	36	9.74	0.00212		
50	48	8.73	0.00125*	1.4	13.2
50	60	10.24	0.00039*	1.3	10.5
50	72	8.69	0.00099*	1.7	14.1
60	12	5.06	0.05753		
60	24	6.79	0.01250		
60	36	7.84	0.00923		
60	48	7.07	0.00620		
60	60	8.30	0.00246		
60	72	6.96	0.00590		

Table 2.6. (continued)

Radii (km)	Time interval (hours)	Logistic regression slope	p-value	Periods acoustically detected to predict 50% probability of encounter during aerial survey (%)	Periods acoustically detected to predict 90% probability of encounter during aerial survey (%)
Humpback whale					
20	12	0.79	0.48078		
20	24	1.04	0.46028		
20	36	1.14	0.46493		
20	48	1.02	0.53036		
20	60	0.89	0.59369		
20	72	0.96	0.58470		
30	12	0.21	0.84403		
30	24	0.33	0.80919		
30	36	0.51	0.73507		
30	48	0.44	0.78082		
30	60	0.29	0.86078		
30	72	0.21	0.90302		
40	12	0.77	0.45528		
40	24	1.32	0.30977		
40	36	1.64	0.25799		
40	48	1.79	0.23754		
40	60	1.82	0.24091		
40	72	1.71	0.28842		
50	12	0.68	0.48525		
50	24	0.79	0.51686		
50	36	1.11	0.41470		
50	48	1.17	0.41441		
50	60	1.15	0.43171		
50	72	0.98	0.51828		
60	12	0.38	0.69544		
60	24	0.39	0.74849		
60	36	0.64	0.63141		
60	48	0.69	0.62488		
60	60	0.68	0.64060		
60	72	0.49	0.74464		
Sei whale					
20	12	1.95	0.37012		
20	24	4.45	0.03477		

Table 2.6. (continued)

Radii (km)	Time interval (hours)	Logistic regression slope	p-value	Periods acoustically detected to predict 50% probability of encounter during aerial survey (%)	Periods acoustically detected to predict 90% probability of encounter during aerial survey (%)
Sei whale (cont.)					
20	36	4.59	0.04185		
20	48	5.37	0.02178		
20	60	5.69	0.02564		
20	72	5.67	0.03766		
30	12	30.16	0.01502		
30	24	6.37	0.00496		
30	36	7.58	0.00370		
30	48	9.97	0.00059*	4.6	17.9
30	60	14.43	0.00004*	4.4	12.6
30	72	13.35	0.00012*	5.6	15.4
40	12	30.16	0.01502		
40	24	6.37	0.00496		
40	36	7.58	0.00370		
40	48	9.97	0.00059*	4.6	17.9
40	60	14.43	0.00004*	4.4	12.6
40	72	13.35	0.00012*	5.6	15.4
50	12	29.63	0.02510		
50	24	5.23	0.01541		
50	36	5.89	0.01434		
50	48	7.49	0.00391		
50	60	9.47	0.00096*	4.2	18.0
50	72	9.22	0.00139*	5.3	20.5
60	12	29.63	0.02510		
60	24	5.23	0.01541		
60	36	5.89	0.01434		
60	48	7.49	0.00391		
60	60	9.47	0.00096*	4.2	18.0
60	72	9.22	0.00139*	5.3	20.5
Fin whale					
20	12	1.89	0.12332		
20	24	2.03	0.12168		
20	36	1.88	0.16090		
20	48	2.00	0.14050		

Table 2.6. (continued)

Radii (km)	Time interval (hours)	Logistic regression slope	p-value	Periods acoustically detected to predict 50% probability of encounter during aerial survey (%)	Periods acoustically detected to predict 90% probability of encounter during aerial survey (%)
Fin whale (cont.)					
20	60	2.02	0.13555		
20	72	2.11	0.12371		
30	12	1.52	0.14590		
30	24	2.30	0.04180		
30	36	2.14	0.06487		
30	48	2.31	0.05264		
30	60	2.22	0.06216		
30	72	2.23	0.06442		
40	12	2.81	0.01135		
40	24	4.38	0.00122*	17.2	64.2
40	36	4.29	0.00201		
40	48	4.60	0.00151*	16.4	60.8
40	60	4.46	0.00203		
40	72	4.70	0.00169		
50	12	1.87	0.07344		
50	24	2.82	0.01827		
50	36	2.66	0.02897		
50	48	2.82	0.02439		
50	60	2.69	0.03096		
50	72	2.73	0.03113		
60	12	1.87	0.07344		
60	24	2.82	0.01827		
60	36	2.66	0.02897		
60	48	2.82	0.02439		
60	60	2.69	0.03096		
60	72	2.73	0.03113		

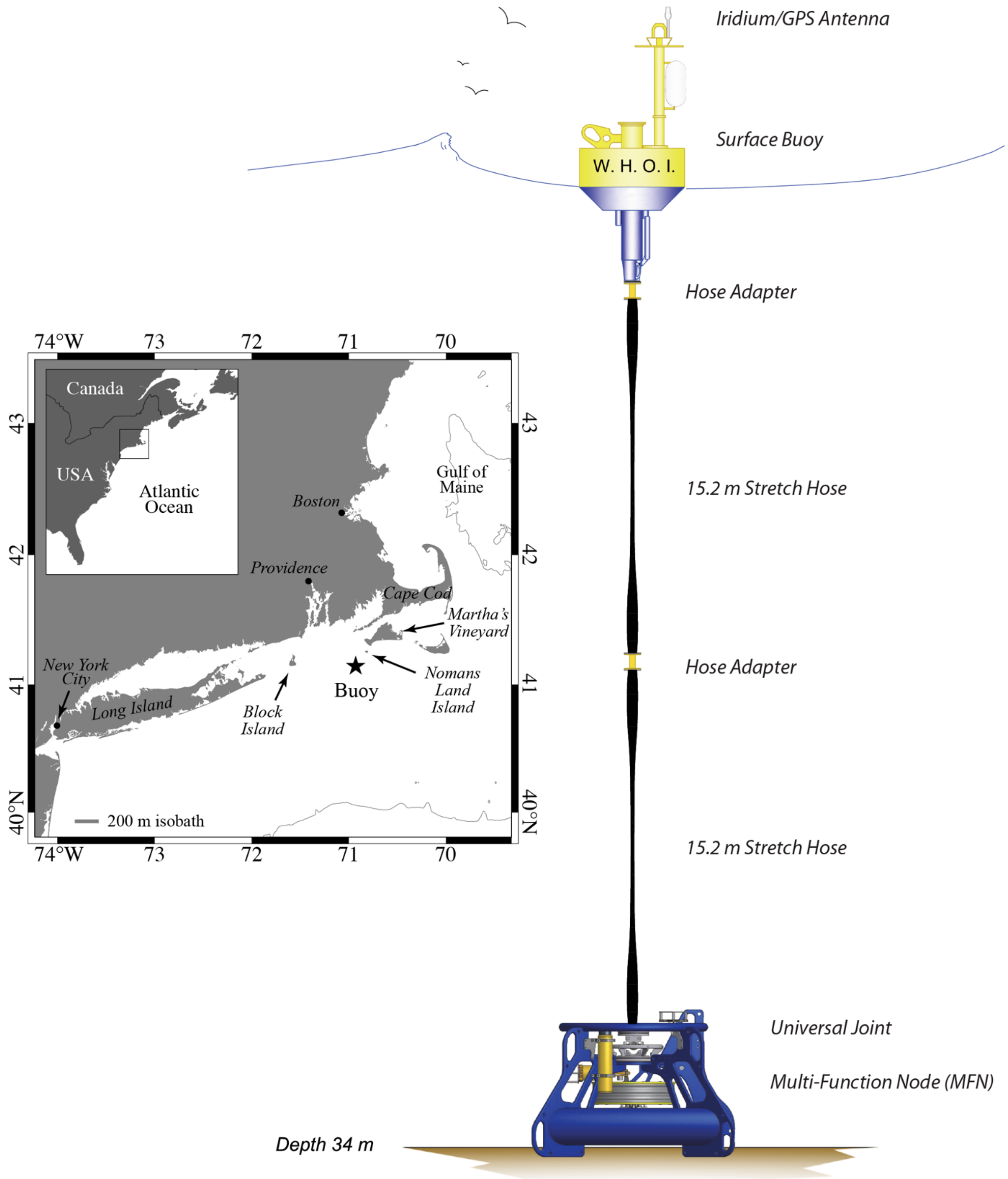


Figure 2.1. Buoy design and location
Design of the DMON/LFDCS mooring, including surface buoy, stretch hoses, and multi-function node (MFN) to which the DMON was affixed. The location of the DMON/LFDCS buoy southwest of Martha's Vineyard, Massachusetts is also shown.

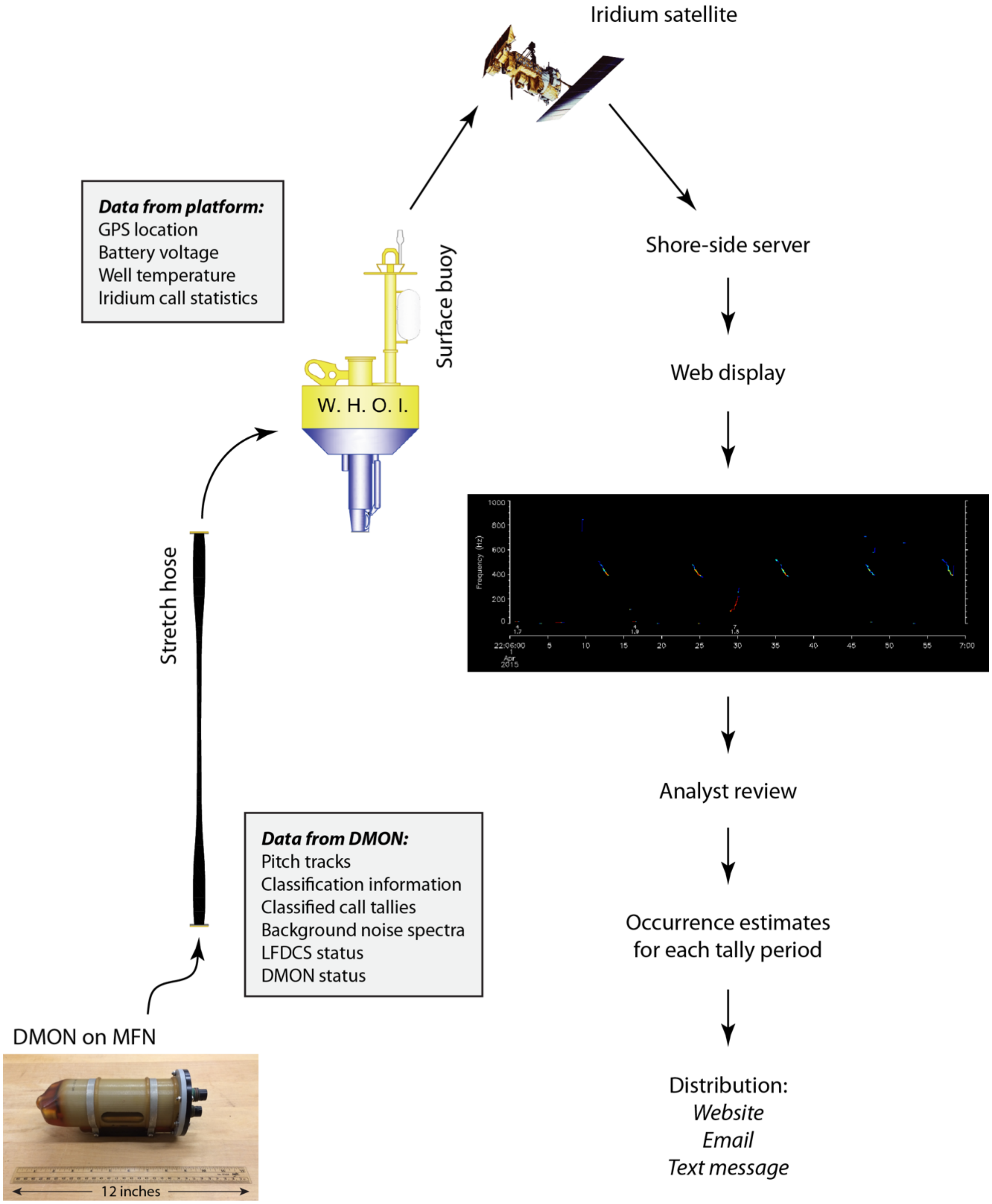


Figure 2.2. Detection data flow
Diagram of data flow from the DMON mounted on the multi-function node (MFN) to a shore-side server via the stretch hoses, surface buoy, and Iridium satellite service. These data are displayed on a website and reviewed by an analyst to produce species-specific occurrence estimates for each monitored tally period. Occurrence estimates are then distributed to users via a publically accessible website as well as email and text messages.

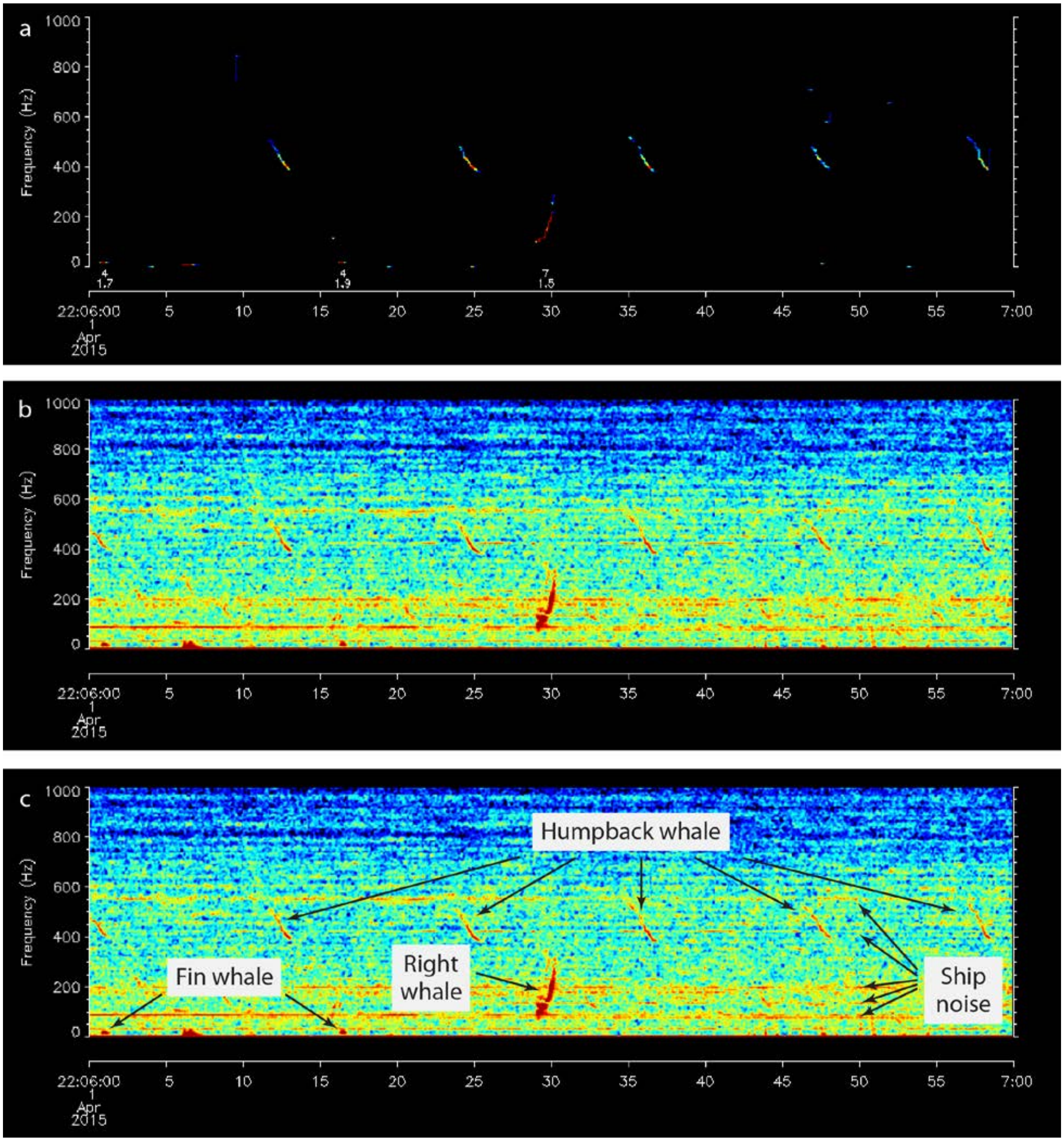


Figure 2.3. Examples of pitch tracks and associated spectrograms
(a) Display of detection information transmitted in near real-time as it appears on the website, which includes pitch tracks (colored lines; quiet sounds in cool colors, loud sounds in warm colors) and associated classification information for classified calls (numbers below some pitch tracks). (b) Corresponding spectrogram for time period shown in (a) (2000-Hz sampling rate, 512-sample frame, 75% overlap, Hann window). While processed in real time to generate the detection information shown in (a), the archived audio used for the spectrogram is only accessible upon recovery of the mooring. (c) Same spectrogram in (b) with annotations of sounds.

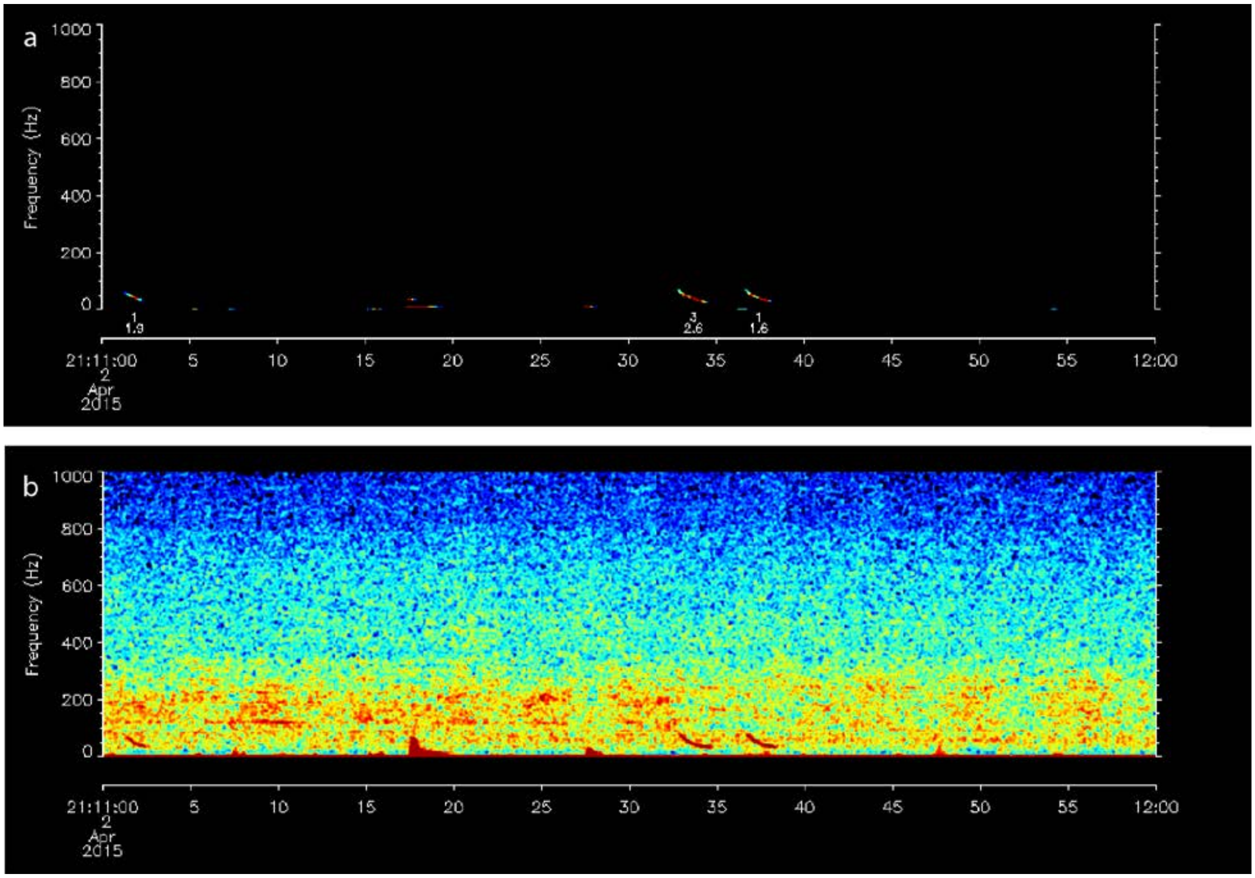


Figure 2.4. Sei whale pitch track example
(a) Display of detection information transmitted in near real-time as it appears on the website showing sei whale downsweeps, including a doublet. Classification information for classified calls (text below calls) show call type on top (call types 1 and 3 are variants of the sei whale downsweep) and Mahalanobis distance (described in Baumgartner and Musoline 2011) on the bottom. (b) Corresponding spectrogram for the time period shown in (a); note that the audio used to produce the spectrogram is only available after the mooring is recovered.

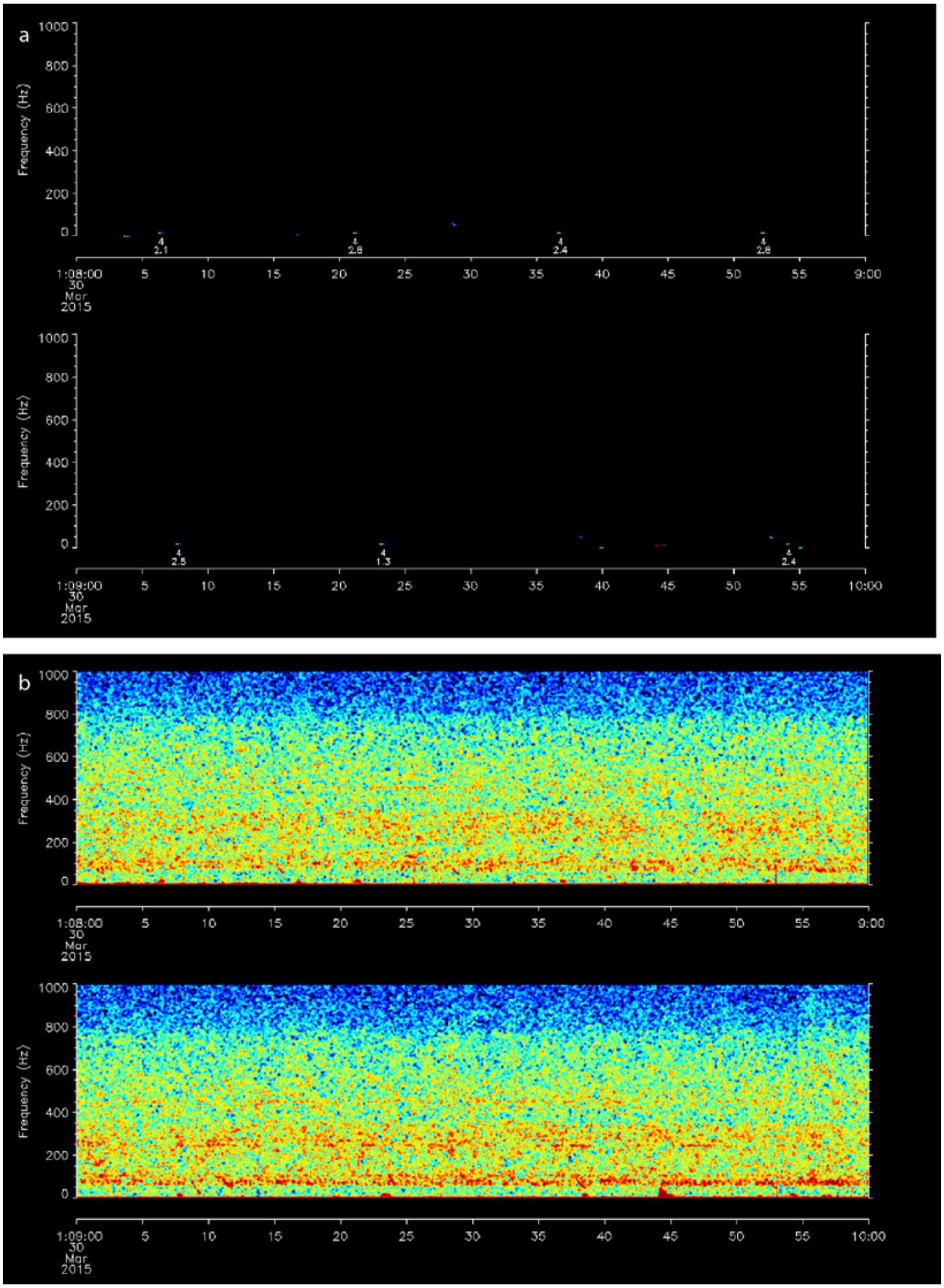


Figure 2.5. Fin whale pitch track example
(a) Display of detection information transmitted in near real-time as it appears on the website showing a sequence of fin whale 20-Hz pulses with a 15-16 s inter-pulse interval. Figure similar to Figure 2.4 (call type 4 is the fin whale 20-Hz pulse). (b) Corresponding spectrogram for the time period shown in (a); note that the audio used to produce the spectrogram is only available after the mooring is recovered.

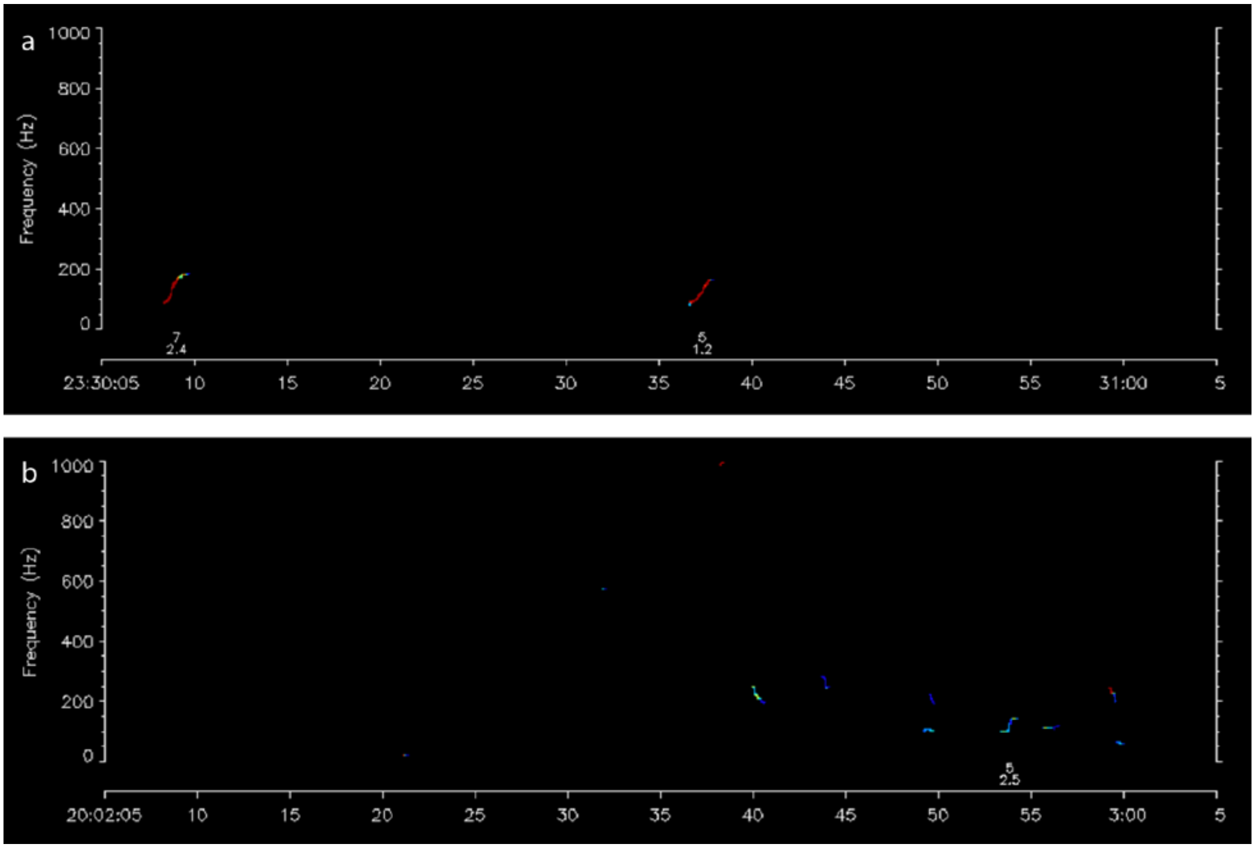


Figure 2.6. Call evaluation example
(a) Pitch tracks of North Atlantic right whale upcalls that (1) have good shape (i.e., no rapid frequency changes in time), (2) are isolated from pitch tracks likely caused by noise, (3) are classified as right whale upcalls (call types 5-8 are variants of right whale upcalls), and (4) have high amplitude. (b) Pitch track of a putative right whale upcall at 20:02:54. The putative call is classified as a right whale upcall (call type 5), but its amplitude is low, it has poor shape (rapid frequency change in the middle of the call), and it is surrounded by pitch tracks that are most likely produced by noise.

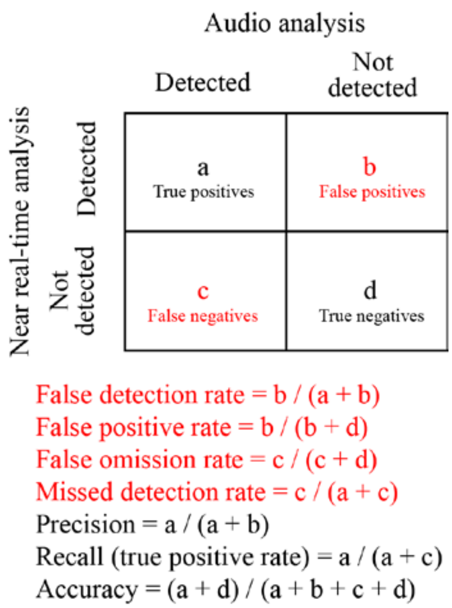


Figure 2.7. Performance metrics
Definitions of performance metrics for comparing occurrence estimates from the near real-time and audio analyses. A good detection process minimizes the quantities in red and maximizes the quantities in black.

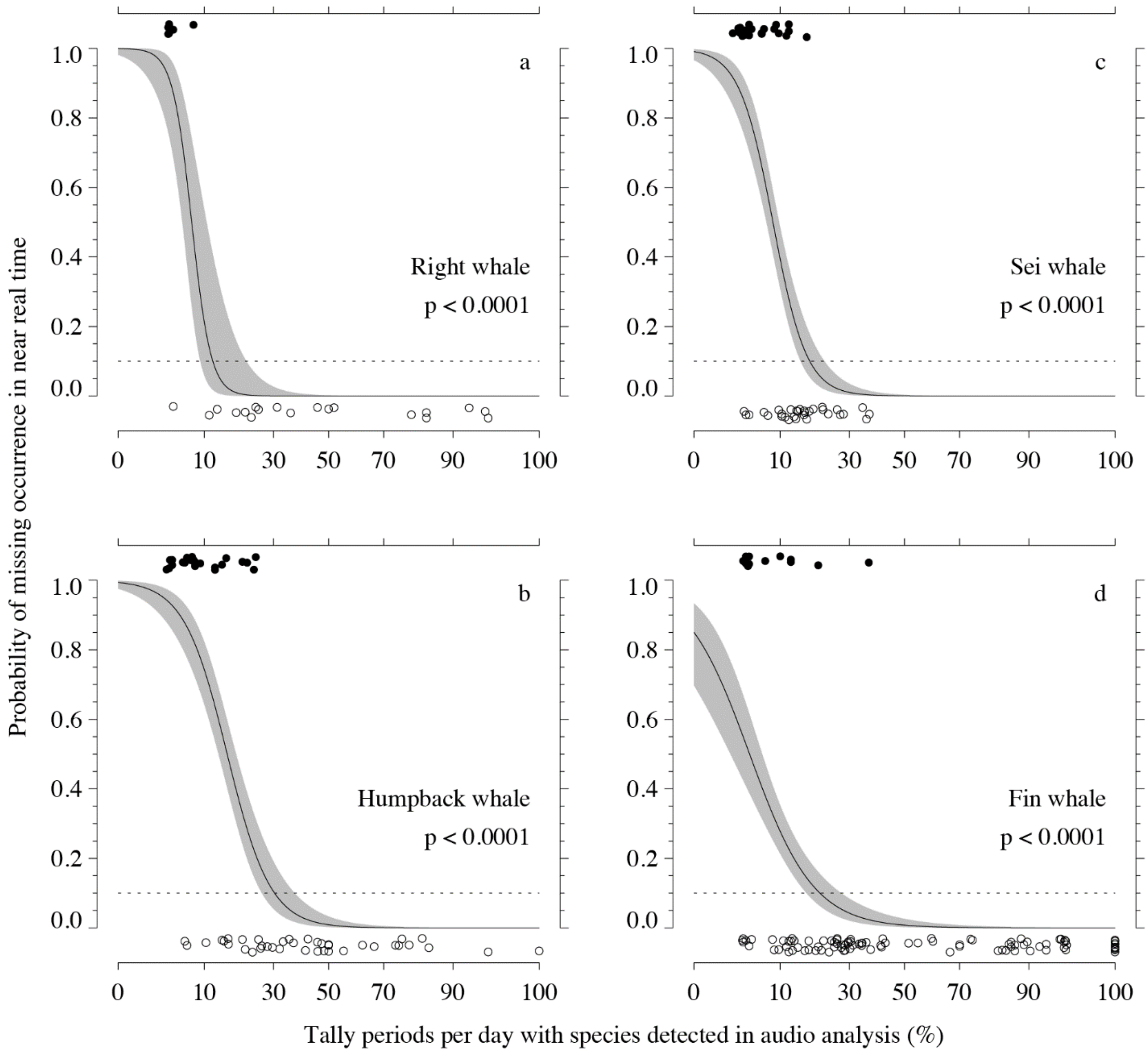


Figure 2.8. Probability of missing occurrence
Probability of missing occurrence in near real time over daily time scales as a function of daily calling rates derived from the audio analysis (i.e., the daily percentage of tally periods scored as "detected" in the audio analysis) for (a) right, (b) humpback, (c) sei, and (d) fin whales. Jittered open circles at the bottom of the plot indicate days when whales were detected both in the audio and near real-time analyses (i.e., not missed). Jittered filled circles at top of plot indicate days when whales were detected in the audio analysis, but missed in the near real-time analysis. The line indicates the logistic regression model fit, the grey area indicates the standard error of the fitted line, and the reported p-value is from the model's drop-in-deviance test.

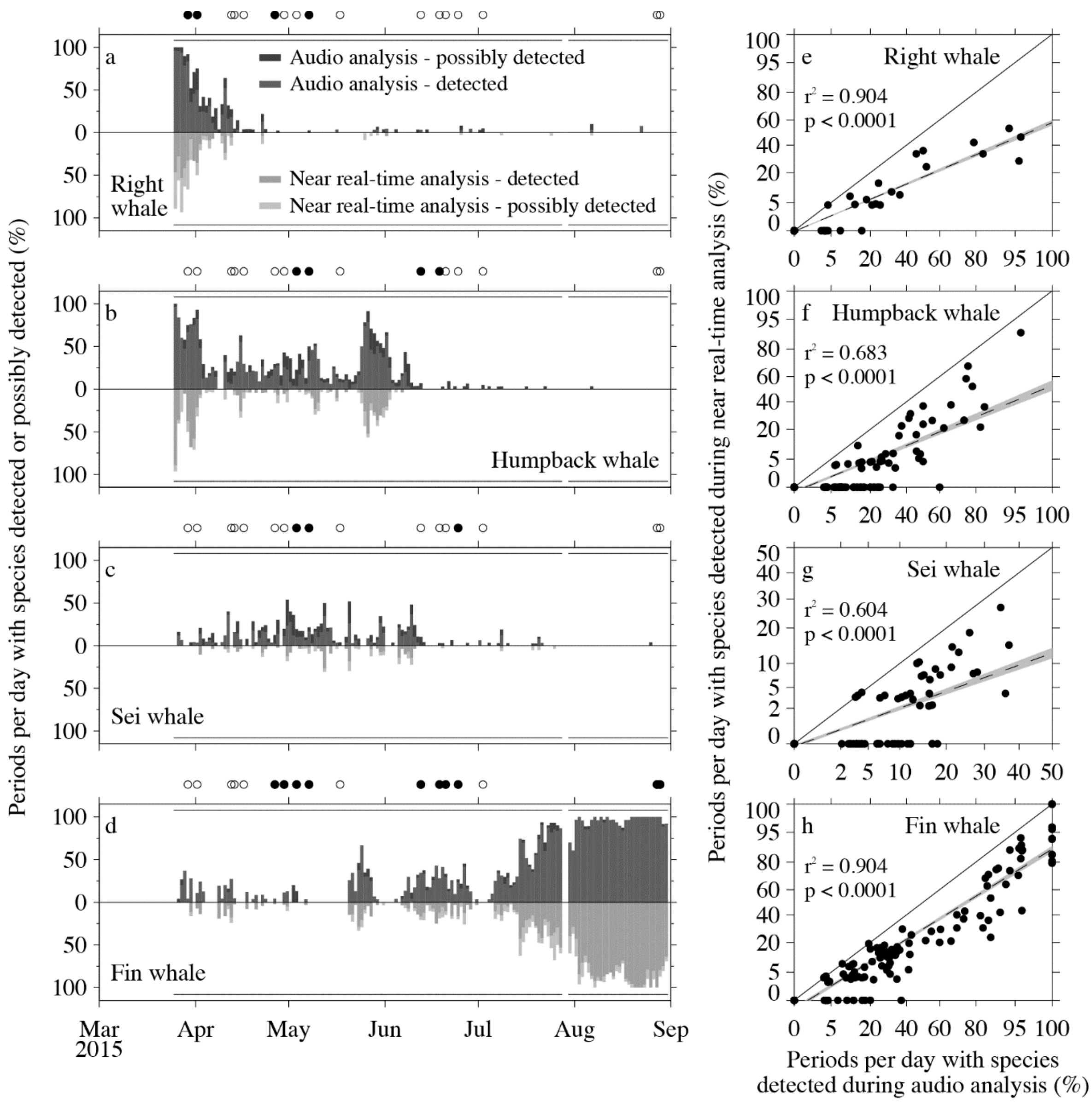


Figure 2.9. Daily detections
Time series of daily detections derived from the retrospective audio analysis (above zero line) and daily detections derived from the near real-time analysis (below zero line) for (a) right, (b) humpback, (c) sei, and (d) fin whales. Detections and possible detections are shown for both analyses. Horizontal lines near the x-axes show when periods were analyzed (the single gap in late August indicates a brief interruption of near real-time data). Filled and open circles above each plot indicate days when aerial surveys were conducted and whales were visually present or absent, respectively. Scatterplots of near real-time versus retrospective audio detections for (e) right, (f) humpback, (g) sei, and (h) fin whales. Coefficients of determination (r2) and associated p values are shown as well as a 1:1 line (solid), simple linear regression line (dashed), and standard error of the linear regression line (grey area).

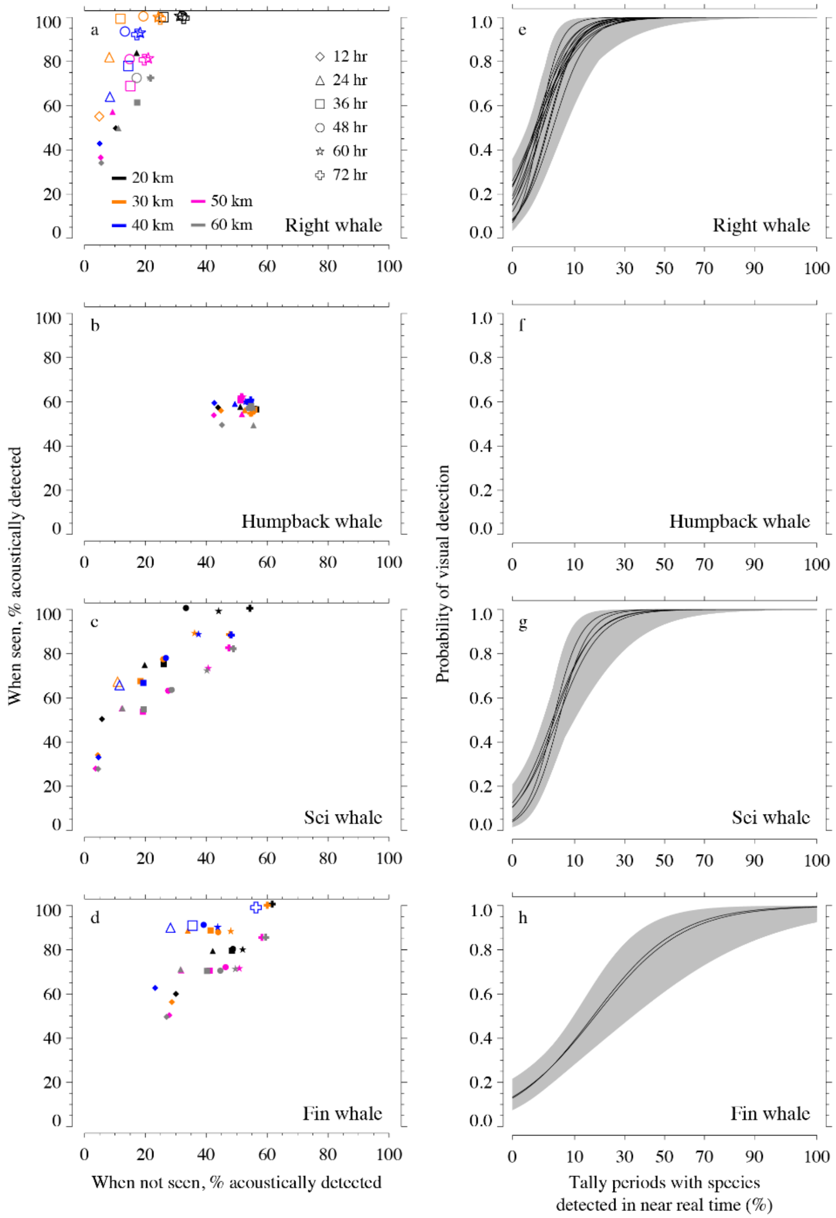


Figure 2.10. Visual-acoustic comparison
(a-d) Acoustic detection rates when (a) right, (b) humpback, (c) sei, and (d) fin whales are visually detected during aerial surveys (y-axis) and when not visually detected during aerial surveys (x-axis) within particular radii of the buoy and within particular time intervals prior to the start of an aerial survey (data in Table 2.5). Large open symbols are for radii and time intervals that have significant log odds ratio tests ($p < \alpha_{Bonferroni}$); small filled symbols have non-significant log odds ratio tests. Symbols are jittered by less than $\pm 1\%$ to improve clarity. Symbols located in the upper left-hand corner of the plot would indicate excellent agreement between the visual and acoustic observations. (e-h) Logistic regression model results showing the probability of encountering a (e) right, (f) humpback, (g) sei, or (h) fin whale within particular radii of the buoy during an aerial survey against the percentage of tally periods with those species scored as acoustically detected in near real time within particular time intervals prior to the start of an aerial survey. Fitted regression lines are shown for significant models only (drop-in-deviance test has $p < \alpha_{Bonferroni}$; data in Table 2.6), while the gray area indicates the standard error for all fitted lines plotted on top of one another.

3 Acoustic detection range of right whale upcalls detected in near real time from a moored buoy and a Slocum glider

3.1 Introduction

Mitigation of anthropogenic impacts on North Atlantic right whales and other at-risk species is critical but challenging given limited survey resources and the cryptic nature of whale behavior. Nearly all risk mitigation and management strategies rely on knowledge of whale distribution collected by monitoring surveys. Traditional visual survey methods provide critically important information used for population and health assessment, but they alone cannot cover the time and space scales required to resolve range-scale distribution patterns. Passive acoustic monitoring (PAM) can complement visual survey methods by offering the ability to autonomously monitor remote areas persistently for months to years at a time.

Numerous efforts have demonstrated the efficacy of PAM for right whale monitoring. Clark et al. (2010) conducted an extensive comparison between aerial and acoustic surveys for right whales in Cape Cod Bay. They found that visual surveys detected right whales on only two-thirds of the days in which they were detected acoustically. They concluded that PAM is a more reliable mechanism for determining right whale presence over daily timescales, and strongly recommended that it be used to inform management decisions. In a similar comparison in Roseway Basin, Durette-Morin et al. (2019) not only came to similar conclusions, but also highlighted the capacity of PAM to extend monitoring beyond visual surveys constrained by resources and poor sightings conditions. Davis et al. (2017) collated and analyzed a massive acoustic dataset spanning 35,600 days over 2004-2014 on recorders located from the Caribbean up the eastern seaboard of North America north to Iceland and the Davis Strait. They were able to use that dataset to document shifts in the range-scale distribution pattern of the species since 2010 as well as persistent wintertime presence in most regions, both of which would likely not have been possible using only the existing, sporadic visual survey effort.

Archival PAM data, while information rich, is typically not available on timescales required to inform dynamic management and risk-mitigation strategies. The Woods Hole Oceanographic Institution (WHOI) has developed a PAM system comprised of a low-power acoustic instrument (digital acoustic monitoring instrument; DMON) and on-board detection algorithm (low-frequency detection and classification system; LFDCS) that detects, classifies and reports the sounds of at-risk baleen whales (right, fin, sei, blue, and humpback) in near real-time from autonomous platforms. Briefly, the LFDCS algorithm produces smoothed spectrograms of the audio data, removes spurious broadband noise and continuous tonal noise, then uses a contour-following algorithm to create pitch tracks of tonal sounds from the spectrogram (Figure 3.1). It then sends a subset of these pitch tracks back to shore via Iridium satellite approximately every 2 hours where they are divided into ~15-minute analysis periods that can be manually reviewed for acoustic presence by a trained analyst (Figure 3.2; Baumgartner & Mussoline 2011, Baumgartner et al. 2013, in press).

The LFDCS detector and validation protocol have been extensively used and quantitatively evaluated for right whales. Davis et al. (2017) used the LFDCS on a desktop computer for their massive analysis of archival recordings, and Baumgartner et al. (in press) recently evaluated the accuracy of the LFDCS on the DMON for near real-time detections. Baumgartner et al. (in press) found that the false positive rate was 0%, meaning that right whales were never detected in near real-time when they were not acoustically present, and that the system missed right whale presence 27% of the time on daily time scales. The protocol was designed to be conservative (cautious) in recognition of the high operational costs of a false detection but can be adjusted depending on the application (Baumgartner et al. in press).

The DMON/LFDCS system is fully operational on both Slocum gliders (Baumgartner et al. 2013) and moored buoys (Baumgartner et al. in press). These platforms are particularly useful for management applications because they can monitor persistently for months to years at a time, regardless of weather conditions, at no risk to human operators, and at a relatively low cost compared to traditional visual surveys. Since 2013, the DMON/LFDCS system has been deployed on 31 Slocum glider and 7 moored buoy missions in the Northwest Atlantic, amassing nearly 3,300 days and 25,000 kilometers at sea, and recording over 1,200 right whale detections. All of these data are made available in a variety of ways, including email and text messages, websites (robots4whales.whoi.edu, whalemap.ocean.dal.ca) and a mobile app (Whale Alert App). The system has already demonstrated its effectiveness in several monitoring initiatives with the National Oceanographic and Atmospheric Association (NOAA) Northeast Fisheries Science Center (NEFSC), the U.S. Navy and Coast Guard, Department of Fisheries and Oceans Canada (DFO), and the Department of National Defense Canada (DND).

As with visual surveys, PAM performance depends on a variety of biological and environmental factors. For PAM, the source, propagation conditions, receiver, and detection processes all influence the probability of detection. A challenge to applying many PAM systems, including the DMON/LFDCS, to science, conservation, and mitigation applications is uncertainty in the species-specific acoustic detection range from a particular monitoring platform. The platforms in which the DMON/LFDCS has been integrated currently relay only the position of the platform when sounds are detected, not the position of a sound source. Determining whether positional uncertainty is tolerable for a particular application depends a great deal on the acoustic detection range for a species of concern; for small detection ranges (e.g., hundreds of meters), the position of the platform may be an acceptable proxy for the position of the animal, but for large detection ranges (e.g., tens of kilometers), lack of location specificity may severely limit mitigation options.

Because of their critically endangered status, right whales are a conservation priority, and as the use of autonomous platforms for near real-time PAM increases, so does the need to evaluate acoustic detection range for these PAM systems in a variety of habitats. The goal of the present study was to assess the range-dependent accuracy of the DMON/LFDCS for detecting right whale upcalls on a mobile (glider) and a fixed (buoy) platform in a relatively shallow environment on the continental shelf that will be developed for wind energy in the near future.

3.2 Methods

We deployed collocated horizontal and vertical line arrays of hydrophones and a Slocum glider from the R/V *Tioga* adjacent to an extant monitoring buoy 9 km southwest of Noman's Island, Massachusetts, USA from 28 Feb to 30 Mar 2017. The water depth was approximately 30 m at the buoy and remained relatively flat to a range of 15 km with the notable exception of a steep shoal near Noman's Island (Figure 3.3). The glider and buoy were equipped with DMON/LFDCS near real-time PAM systems configured to sample audio continuously at 2 kHz. These systems generated and classified pitch tracks of tonal signals in real time and transmitted them back to shore every 2 hours. More details on the specifications of the PAM system on the glider and buoy are available in Baumgartner et al. (2013) and Baumgartner et al. (in press), respectively.

The vertical line array (VLA) used in this study consisted of a Several Hydrophone Receiving Unit (SHRU), 4 hydrophones, multiple environmental sensors, and a number of additional mooring components. The SHRU was suspended several meters above the anchor and acoustic release system, and it sampled the hydrophones continuously at a rate of 9.7656 kHz for the full deployment period (see Newhall et al. (2010) for a description of the SHRU electronics, data formatting, data processing, and additional engineering specifications). The hydrophones and environmental sensors were secured to a 15-m wire rope that extended from the top of the SHRU to a steel sphere suspended approximately 8 m below the surface. The environmental sensors included two temperature loggers and a temperature-pressure logger positioned at intervals along the extent of the array to measure the water column structure, depth, and array tilt at 0.5 Hz throughout the entire deployment.

The horizontal line array (HLA) was comprised of 8 hydrophones positioned at 7.5 m intervals along a 60-m cable coated with hairy fairing. These were all sampled continuously at 4 kHz using a multichannel recorder built by Webb Research Corporation (see Newhall et al. (2010) for more detailed system specifications). The HLA also had a single temperature-pressure sensor to record bottom water properties for the full deployment.

The full 12-channel acoustic record from the HLA/VLA was displayed as spectrograms and visually/aurally reviewed for right whale upcalls. The pitch tracks and classification information generated in real time by the DMON/LFDCS on the buoy and glider were also independently analyzed for the presence of right whale upcalls. Upcalls were chosen because they are a stereotypical call used by the LFDCS to determine right whale presence and were amenable to localization (see below). Each call discovered in the HLA/VLA record was displayed as a spectrogram and manually cropped to isolate the call from spurious noise prior to localization.

The HLA and VLA were deployed concurrently to facilitate call localization using a normal mode back-propagation method (Lin et al. 2012, Newhall et al. 2012). The utility of this method is that it allows 3-D localization of low-frequency signals from a single station, as opposed to the distributed arrays required for conventional arrival time difference methods (e.g., Cato 1998). The general steps of the localization workflow were to (1) isolate the call in the array data (2) use a normal mode model (KRAKEN; Porter 1992) and pseudo-inverse mode filter to isolate the modal arrivals of a given call on the VLA, (3) use the estimated group velocities of each modal arrival to beamform with the HLA to determine the arrival angle (bearing) of the call, (4) use the same mode model to estimate mode structures along the arrival path, and (5) back-propagate the received signal along the arrival path until the filtered modes converge. The range with the best convergence was used as the estimated range to the call. With this estimated bearing and range, the position of the calling whale could be calculated. For more detail on the methods, see Lin et al. (2012) and for an application to sei whale localization, see Newhall et al. (2012).

The normal mode back-propagation method requires the excitation of two or more acoustic modes. The cut-off frequency for mode 2 at the study site was approximately 80 Hz, which prevented localization of any calls with substantial energy at lower frequencies. This meant that right whale upcalls and (some) humpback whale calls were amenable to localization, but fin whale 20-Hz pulses and sei whale downsweeps were not. The cut-off frequency for mode 3 was approximately 300 Hz, so mode 3 was not reliably present in all upcalls. We repeated the localization routine using both 2 and 3 modes to exploit the presence of mode 3 when possible.

We assumed that the HLA/VLA would allow better detection as well as detect calls over a greater range than either of the two single-hydrophone DMON/LFDCS systems. As such, the HLA/VLA record was used as the ground truth for comparison between platforms, and the performance of each DMON/LFDCS was assessed relative to the HLA/VLA. For each call detected and localized on the HLA/VLA, a score of zero was assigned if the call was missed and a score of one was assigned if the call was detected by the buoy's DMON/LFDCS. Likewise, scores of zero and one were assigned for HLA/VLA-detected and localized calls that were missed and detected, respectively, by the glider's DMON/LFDCS. The series of scored pitch tracks was used as the dependent variable and the range to each localized call was used as the independent variable in a logistic regression to characterize the relationship between the probability of detecting localized calls and range. Separate logistic regressions were fit for the buoy and the glider. For all missed calls, we also examined the buoy and glider pitch track records to determine why they were not detected by the DMON/LFDCS systems.

3.3 Results

During recovery of the HLA, it was immediately evident that energy from late-winter storms moved the array from its original position. There were two occasions (2-3 Mar and 14-15 Mar) with substantial storm-induced noise, indicating times when the HLA may have moved (Figure 3.4). Precise estimates of the location of each HLA element was critical to our localization methodology, as errors in HLA element location prevent accurate beamforming for call bearing estimation. Fortunately, the R/V *Tioga* revisited the deployment location on several occasions while the arrays were recording. The known position and source of the R/V *Tioga* was used to re-locate the HLA elements to correct for storm-induced movement. This analysis confirmed that the array moved in both storm events, but was stable (i.e., usable) during the period before, between and after the storms.

A total of 489 right whale upcalls were detected on the HLA/VLA over the two-week period between 28 Feb and 14 Mar. The DMON/LFDCS systems on the glider and buoy convincingly pitch tracked 340 and 196 right whale upcalls, respectively, during the same period. Calls occurred throughout the monitoring period, but the majority occurred on a single day (08 March; Figure 3.5). Of the calls detected on the array, 42.5% (208 of 489) could be accurately localized. Figure 3.6 shows an example of the localization workflow for a single successfully localized call. Most calls originated to the south of the DMON/LFDCS buoy and the arrays (Figure 3.7). The distances to localized calls from each platform ranged from 0.6 km to 19.6 km on the glider (median = 7.2 km), and from 0.2 km to 19.8 km on the buoy (median = 6.7 km; Figure 3.8).

The 208 localized calls were used to determine the range-dependent accuracy of the DMON/LFDCS for each platform. For the buoy, the proportion of localized calls that were detected decreased with range (Figure 3.8); 68.2% of localized calls (58 of 85) within 5 km were detected while 17.0% of localized calls (2 of 12) beyond 15 km were detected. Calls were missed for a variety of reasons: 43.5% of the 208 localized calls were missed because of absent or poor pitch tracks, 3.8% were missed because of interfering biological sounds (i.e., humpback whale song), 7.2% were missed because of interfering non-biological sounds (e.g., other platform noise, ship noise), and 1.0% were missed because of human error in scoring the pitch tracks (Table 3.1; Figure 3.9). The logistic regression provided strong evidence that the probability of detecting localized calls was related to range ($p < 0.0001$; drop-in-deviance test), and the fitted regression curve suggested that the average probability of detecting a localized call at 5.4 and 9.4 km was 0.50 (95% CI: 0.335-0.665) and 0.333 (95% CI: 0.170-0.549), respectively (Figure 3.10).

For the glider, many calls (n=52) occurred during periods when the glider caused acoustical or electrical noise during activation of the buoyancy pump (typically 30 seconds every 3.5 minutes in 30-35 m water depths) or during satellite communications (typically 10-15 minutes every 2-2.25 hours), respectively (Table 3.1). Because these calls were not available for detection and therefore could not inform our assessment of the effect of range on the accuracy of the DMON/LFDCS, they were excluded from the analysis. The proportion of the remaining 156 localized calls that were detected decreased with range, albeit not as dramatically as the buoy (Figure 3.8); 55.8% of localized calls (24 of 43) within 5 km were detected while 25.0% of localized calls (2 of 8) beyond 15 km were detected. Calls were missed for a variety of reasons: 39.1% of the localized 156 localized calls were missed because of absent or poor pitch tracks, 7.7% were missed because of interfering biological sounds (i.e., humpback whale song), 1.9% were missed because of interfering non-biological sounds (e.g., other platform noise, ship noise), and 0.6% were missed because of human error in scoring the pitch tracks (Table 3.1; Figure 3.9). The logistic regression provided modest evidence that the probability of detecting localized calls was related to range ($p = 0.0437$; drop-in-deviance test), and the fitted regression curve suggested that the average probability of detecting a localized call at 7.9 and 17.2 km was 0.50 (95% CI: 0.420-0.580) and 0.333 (95% CI: 0.185-0.523), respectively (Figure 3.10).

Some pitch tracks did not conform to right whale upcalls well enough to be completely convincing, so they were scored as “possible” and treated in the analyses described above as undetected. If a science or mitigation application seeks to minimize missed calls at the expense of occasional false detections, these possible detections could be considered equivalent to convincing detections. In our study, treating “possible” pitch tracks as detected calls caused the probability of detecting localized calls to increase for both the buoy and the glider across all ranges (Figure 3.11). For the buoy, the logistic regression still provided strong evidence that the probability of detecting localized calls decreased with range ($p < 0.0001$; drop-in-deviance test), and the fitted regression curve suggested that the average probability of detecting a localized call at 8.9 and 12.7 km was 0.50 (95% CI: 0.295-0.705) and 0.333 (95% CI: 0.144-0.596), respectively (Figure 3.11). For the glider, there was no evidence that the probability of detecting localized calls was related to range ($p = 0.2588$; drop-in-deviance test), while the fitted regression curve suggested that the average probability of detecting a localized call at 16.3 and 32.9 km was 0.50 (95% CI: 0.330-0.670) and 0.333 (95% CI: 0.071-0.766), respectively (Figure 3.11).

3.4 Discussion

Our primary motivation is to improve conservation outcomes for right whales by using an effective and reliable near real-time passive acoustic monitoring system. One such system, the DMON/LFDCS, has been operational for several years, but has not been used extensively to inform dynamic management measures owing partly to uncertainty in the acoustic detection range. We sought to address this uncertainty by conducting a dedicated study that would allow us to empirically quantify the range-dependent probability of detecting localized right whale upcalls from a mobile and a fixed autonomous platform equipped with the DMON/LFDCS. The study site near Noman’s Land Island, USA was chosen because of the extant DMON/LFDCS buoy located there, which was originally deployed to monitor right whale presence in near real time near several Coast Guard gunnery ranges. We deployed the glider and HLA/VLA arrays in the early spring to capitalize on known right whale presence in the region at that time of year (Davis et al. 2017).

Many efforts have succeeded in ranging and localizing baleen whale calls for purposes such as density estimation (e.g., Harris et al. 2013) or call attribution (e.g., Baumgartner et al. 2008), but few studies have attempted to empirically quantify the range-dependent probability of detecting localized calls. This is operationally difficult because it requires knowledge of the source locations of calls that were both detected and not detected by the monitoring platform. As a result, most efforts to characterize detection range rely on modeling efforts. These can be informative, but require many simplifying assumptions be made about the source, transmission, receiver and detector characteristics, which are typically poorly constrained and highly variable. Such an effort in the Bay of Fundy suggested a maximum upcall detection range of 16 km under ideal ambient noise conditions (Tennessen & Parks 2016). Earlier observations in the same region documented maximum detection ranges of approximately 30 km (Laurinolli et al. 2003). The reasons for this discrepancy are unclear, which highlights the challenges of simulating detection range.

We chose an empirical approach to avoid making assumptions about the source (e.g., source depth, level, frequency) or detector characteristics (e.g., ambient or platform noise level, detection threshold). The single-station ranging method we employed does require some assumptions be made about signal transmission, but this was a necessary compromise given the logistical constraints of the alternative approach of deploying a sparse, large aperture array for localization using arrival time differences. These assumptions are well justified, as several studies have demonstrated the efficacy of normal mode ranging of low-frequency signals in shallow water environments (e.g., Wiggins et al. 2004, Munger et al. 2011, Newhall et al. 2012). We made efforts to account for variation in bathymetry by using a range-dependent backpropagation method. Furthermore, the array- and glider-based environmental sensors revealed that the water column was entirely mixed throughout the study, so depth-varying sound speed was unlikely to contribute to ranging error.

The vast majority of calls detected by both the buoy and glider occurred within 15 km of the platforms, with some calls detected between 15 and 20 km (Figures 3.8, 3.9). From the empirically derived detection proportions (Figure 3.8), the buoy and glider performed similarly at close ranges, but the glider had higher detection probabilities than the buoy between 5-15 km. We believe this observation is explained by the different depths of the hydrophones for each platform. For the buoy, the DMON was affixed to a bottom-mounted structure approximately 1.25 m from the sea floor, whereas the glider-borne DMON was constantly moving between just a few meters below the surface and several meters off the sea floor. We assume that the glider’s time away from the surface and bottom boundaries of the water column improved the reception of distant calls. The fitted regression model for the buoy suggested that the probability of detecting localized calls was 0.5 at 5.4 km and 0.333 at 9.4 km, whereas the probability of detecting localized calls on the glider was 0.5 at 7.9 km and 0.333 at 17.2 km; these observations also support the notion that the glider had slightly better performance at farther range than the buoy. There were few localized calls beyond 15 km, but a fraction of those calls was detected by both the buoy (17.0%) and the glider (25.0%).

Clark et al. (2010) studied acoustic detections of right whale upcalls in Cape Cod Bay, a shallow habitat similar to our study area, and they stated that the “acoustic detection area was reliably found to be within a range of approximately 9 km (~5 nmi) from a recorder” (Clark et al. 2010, p. 842). The comparability of our observations for the DMON/LFDCS to this acoustic detection range estimate depends on the definition of “reliably”. If the definition of acoustic detection range is the range at which the probability of detecting a calling whale is 0.5 (i.e., reliable is defined as a 50:50 chance of detection), then our detection range estimates are smaller than those of Clark et al. (2010). However, if we define detection range as the range at which the probability of detecting a calling whale is 0.333 (i.e., reliable is defined as a 1 in 3 chance of detection), then our detection range estimate is very similar to that of Clark et al. (2010). If Clark et al. (2010) were reporting a maximum detection range, then our observed maximum detection range of 15-20 km exceeds the detection range that they reported. These comparisons highlight something that is likely obvious, but perhaps underappreciated: the use of a single number for detection range is an incomplete description of how far away a whale can be detected by a passive acoustic system. From our own study, we observed that whales calling at 15-20 km can be detected by the DMON/LFDCS carried aboard either a glider or a buoy, but the chances of those whales being detected are low.

The maximum detection ranges we observed are similar to model-based estimates (Tennessen & Parks 2016), but half of measured ranges (Laurinolli et al. 2003) in the Bay of Fundy. It is not surprising that our maximum detection ranges are lower than those measured by Laurinolli et al. (2003) given the greater water column depth in the Bay of Fundy (~200 m) than in our study area (~30 m). This discrepancy emphasizes that the results presented here are specific to the conditions in the area and at the time of our study. They provide an indication of how these PAM systems might perform in similar conditions, but caution is warranted when applying our results to other areas or times. Many similar studies to ours must be conducted to characterize variability in the source, transmission, receiver, and detector before detection probability estimates can be generalized.

Effective monitoring area increases exponentially with range from the receiver (e.g., Helble et al. 2013), so it is critically important to resolve the DMON/LFDCS performance at long ranges. Though the probability of detection decreased for both platforms as a function of range, it never reached zero. This is likely because the detection range of the DMON/LFDCS platforms matched or exceeded the localization range of the array (i.e., the distances over which we could successfully localize calls with the HLA/VLA). We did not exploit any array gain in detecting or localizing calls, and this is an area of future work to try to resolve the probability of detecting calling whales between 20 and 30 km. While our logistic regression curves could be used to extrapolate how detectability changes with these longer ranges, we prefer to develop and use observational methods with the HLA/VLA to definitively measure the distance at which the probability of detection falls to zero.

Our approach depends upon right whales vocalizing within the detection range of the PAM system during the study period, as well as our ability to accurately localize those calls. Our success rate for localizing calls was 43% (208 of 489), which was nearly identical to the success rate of Laurinolli et al. (2003) for loud tonal sounds in the Bay of Fundy using traditional cross-correlation methods. Our success rate was substantially higher than in other studies (e.g., Cummings & Holliday 1986).

The empirical detection functions do not reach a value of one at close range (Figure 3.10), suggesting that factors other than range are responsible for missing calls. The majority of the undetected calls within 5 km of the buoy were not detected because of poor pitch tracks, meaning that the pitch tracks were present but could not be confidently classified by the analyst because of poor shape or low amplitude. Poorly formed pitch tracks from close range calls could be caused by competing platform noise processes and/or variable source level (e.g., Parks & Tyack 2005). It is also important to stress that the validation process used here was conducted on a call-by-call basis to facilitate estimating the range-dependent accuracy of the system. The real-time validation procedure operates on nominal 15-minute "tally periods" (Baumgartner et al. in press), and may therefore be more robust to occasional missed detections.

Our analysis makes no attempt to quantify the likelihood that a right whale will produce a call; we merely assess the probability that a platform will detect a call that has already been produced, detected on the HLA/VLA, and localized. Call types, rates, depths, and spectral characteristics (e.g., frequency, amplitude) vary depending on the time of day, season, location, environment, behavior, and individual. Some of this variability has been characterized for right whales (e.g., Parks et al. 2011a,b), but small sample sizes have often precluded range-wide characterization. The current dataset and others like it can contribute to this research area by quantifying aspects of right whale acoustic ecology, including acoustically derived movement patterns, calling rates and source levels.

Our results have considerable implications for right whale management. They provide evidence that the ranges to calls detected in near real-time by the DMON/LFDCS are likely consistent with detection ranges of archival PAM systems, thereby demonstrating the viability of the DMON/LFDCS as a monitoring system. Furthermore, the results reported here have been used by Johnson et al (in prep) to show that, given these detection ranges and allowing for whale movement, acoustic and visual detections provide nearly identical estimates of whale location on dynamic management timescales (typically 1-3 days).

The urgency of right whale conservation compels us to take advantage of every monitoring tool at our disposal to inform population management. We do not have the luxury of time to develop and validate new methods, nor the resources to sustain the necessary monitoring with visual surveys alone. The results presented here allow us to confidently continue to recommend the DMON/LFDCS as a viable near real-time acoustic right whale monitoring and risk mitigation tool. The demonstrated, reliable, and quantified performance of these platforms emphasizes their readiness for use in areas where long-term, persistent, and low-cost monitoring is required.

Table 3.1. Results from manual scoring of glider and buoy pitch track records of calls localized by the HLA/VLA arrays. Here *n* refers to the number of calls, while *P* is the percentage of total localized calls (n = 208).

		Buoy		**Glider**	
Score	**Definition**	***n***	***P***	***n***	***P***
Absent	Calls were not pitch tracked at all because of low amplitude	41	19.7	27	13.0
Poor	Calls were not pitch tracked accurately/completely because of low amplitude or poor shape	50	24.0	34	16.3
Song	Uncertainty due to interfering species calls	8	3.8	12	5.8
Noise	Calls were not pitch tracked accurately/completely because of interfering sound	15	7.2	3	1.4
Missed	Human error (analyst chose wrong score erroneously)	2	1.0	1	0.5
Exclude	Calls were not available for pitch tracking because the platform was not monitoring	0	0.0	52	25.0
Detected	Calls were pitch tracked and scored as detected by analyst	92	44.2	79	38.0

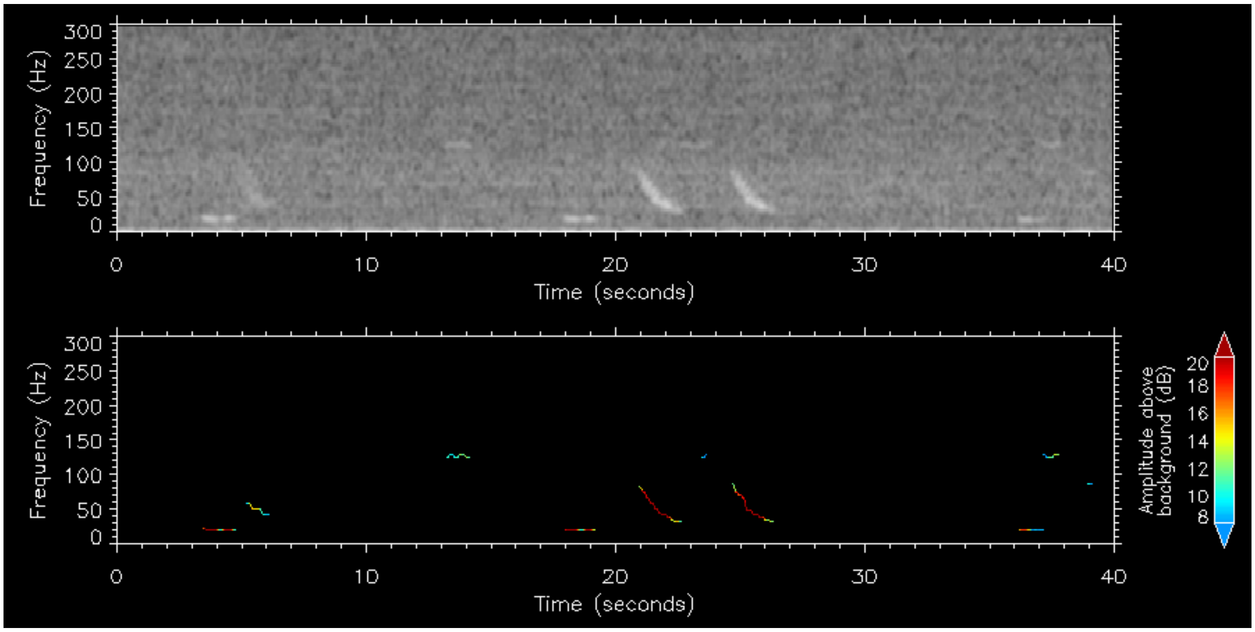


Figure 3.1. Low-frequency detection and classification system (LFDCS)
Spectrogram (top) versus pitch tracks (bottom) of sei and fin whale calls (and several other tonal sounds) generated in real time by the LFDCS.

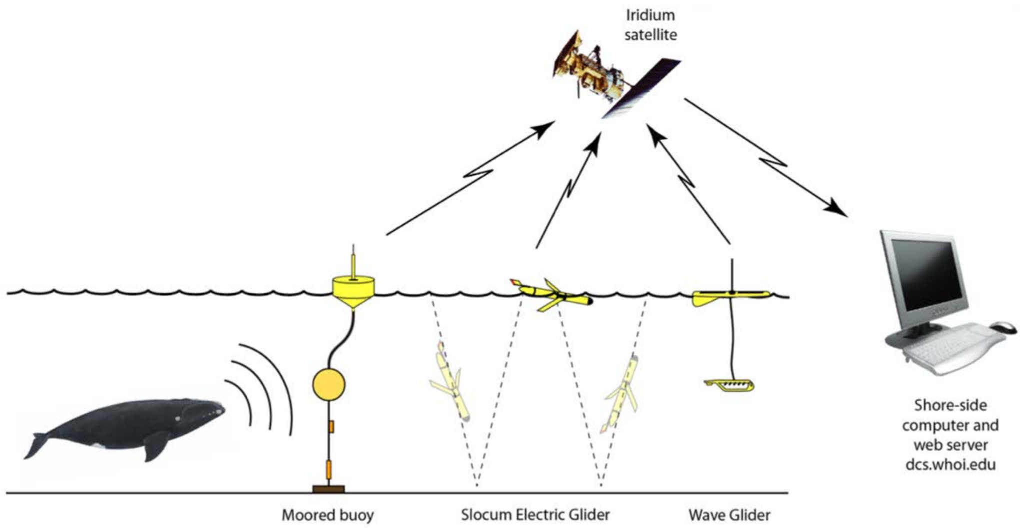


Figure 3.2. Transmission of detection information back to shore
Near real-time transmission of LFDCS pitch tracks and classification information back to shore for manual review.

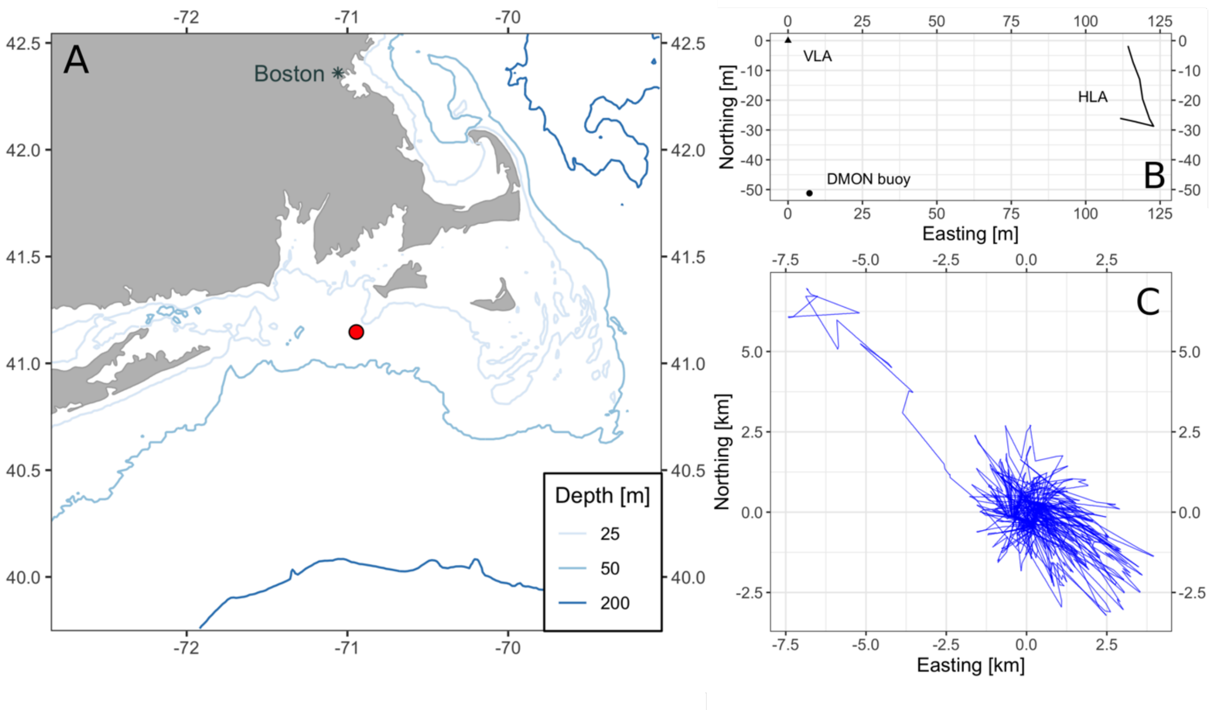


Figure 3.3. Study site and platform locations

(A) Study site (red circle) in 30m water depth 12 km southwest of Martha's Vineyard, MA, USA, (B) positions of the vertical line array (VLA), horizontal line array (HLA) and DMON-LFDCS buoy (DMON buoy) relative to the VLA, and (C) the trajectory of the glider (blue line) relative to the VLA from 28 Feb through 14 Mar.

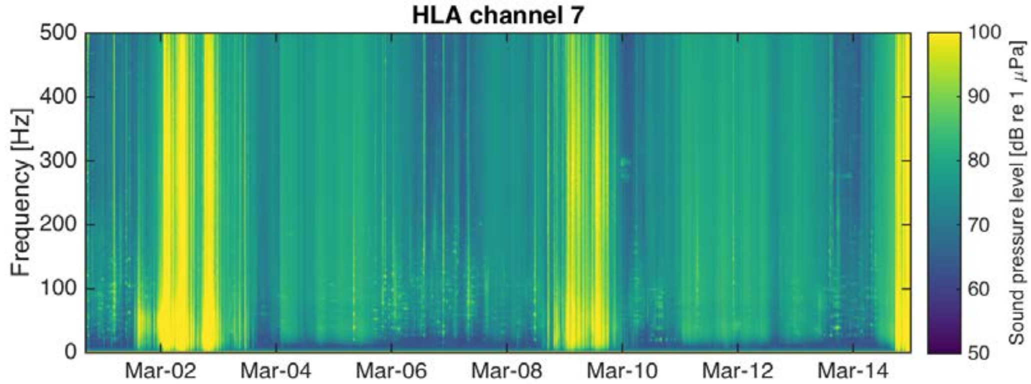


Figure 3.4. Storm-induced noise.
Acoustic energy received at the HLA (channel 7) during the study period. Yellow banding across the full frequency range indicates storm-induced noise.

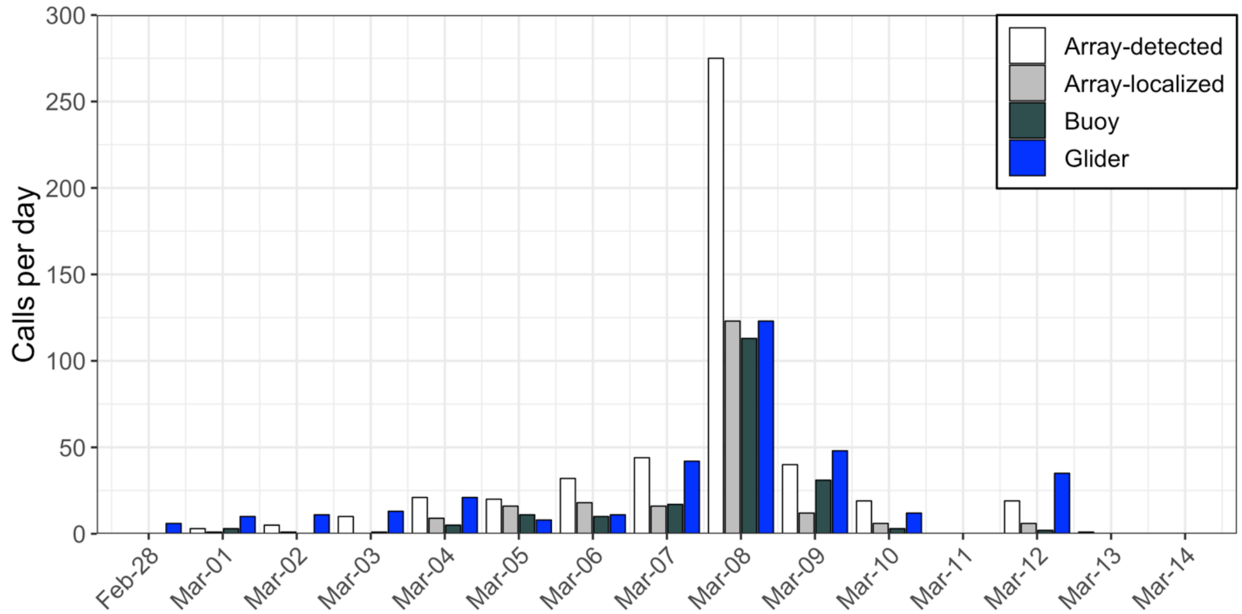


Figure 3.5. Right whale upcall detections by platform.
Daily counts of right whale upcalls detected in the array audio (white bars; n=489), the buoy pitch tracks (black bars; n = 196), and the glider pitch tracks (blue bars; n = 340), as well as numbers of calls that were successfully localized (grey bars; n = 208).

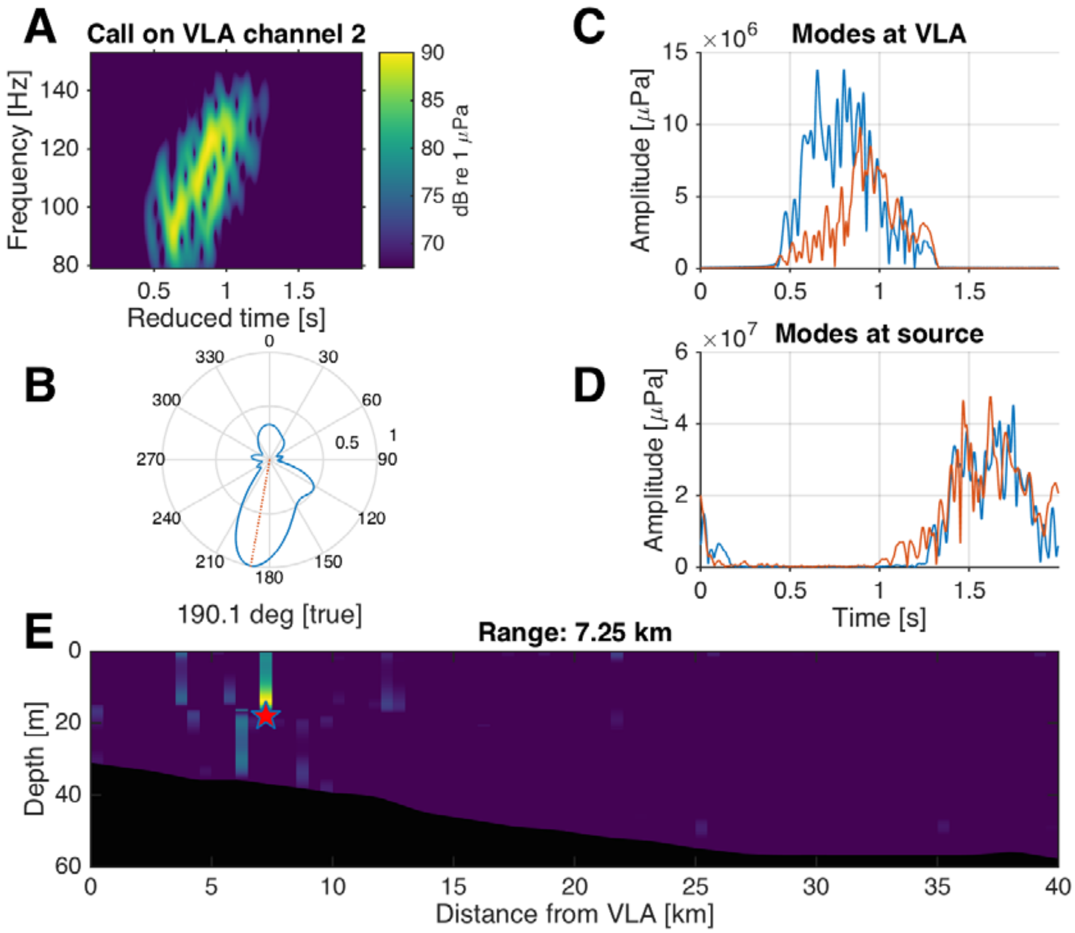


Figure 3.6. Right whale upcall localization workflow.
Example localization workflow for a single call showing (A) the call spectrogram, (B) beam pattern (blue) and arrival angle (red line), (C) received amplitudes of modes 1 (blue) and 2 (red), (D) back-propagated amplitudes of the same modes, and (E) a normalized probability map of the back-propagation results with a star indicating the most likely range and depth of the calling whale.

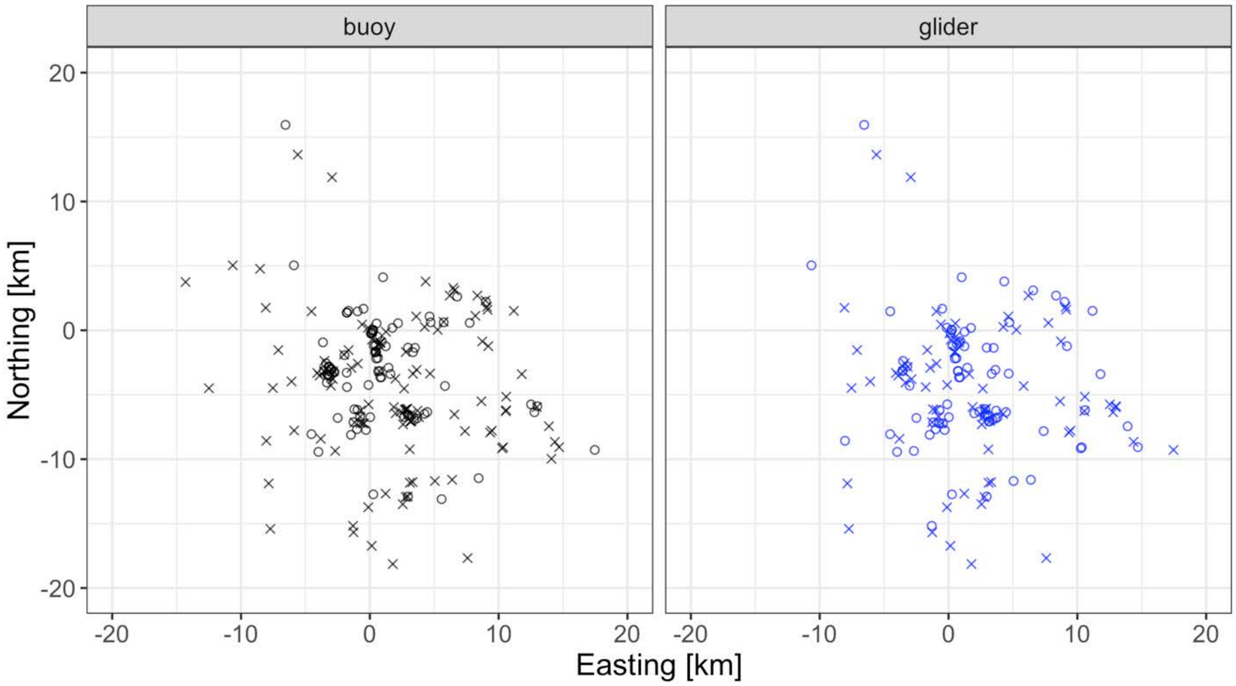


Figure 3.7. Spatial distribution of localized upcalls.
The spatial distribution of localized right whale upcalls. Open circles and crosses indicate calls detected and not detected by the buoy (black) and glider (blue), respectively.

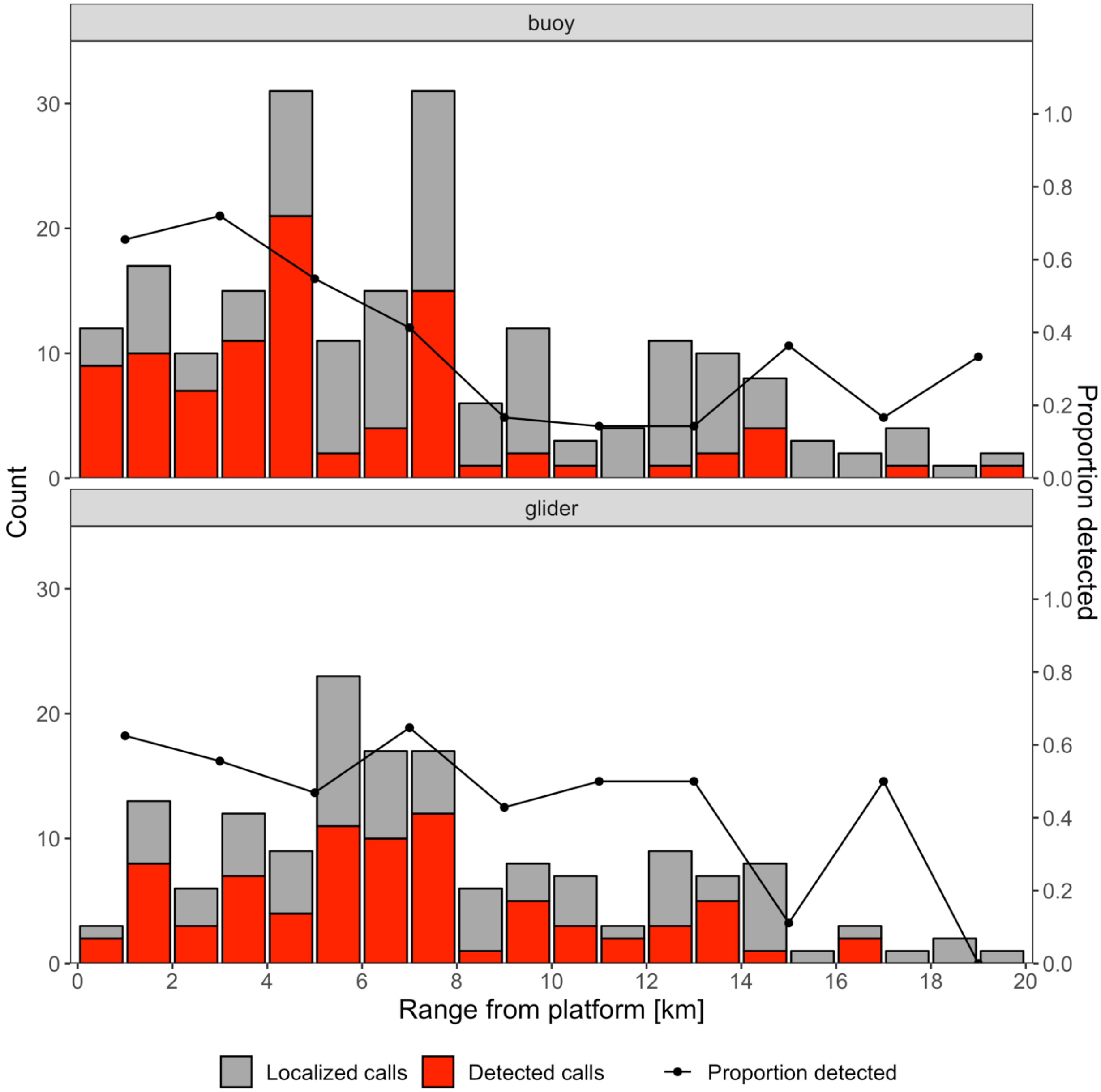


Figure 3.8. Range of localized and detected calls to the buoy and glider.
Distribution of ranges from the buoy and glider to right whale upcalls localized by the array (n = 208 for the buoy, n = 156 for the glider). Total numbers of localized calls in 1-km bins are shown in gray, localized calls detected by the DMON/LFDCS are shown in red, and the proportions of localized calls detected in 2-km bins are shown with a black line.

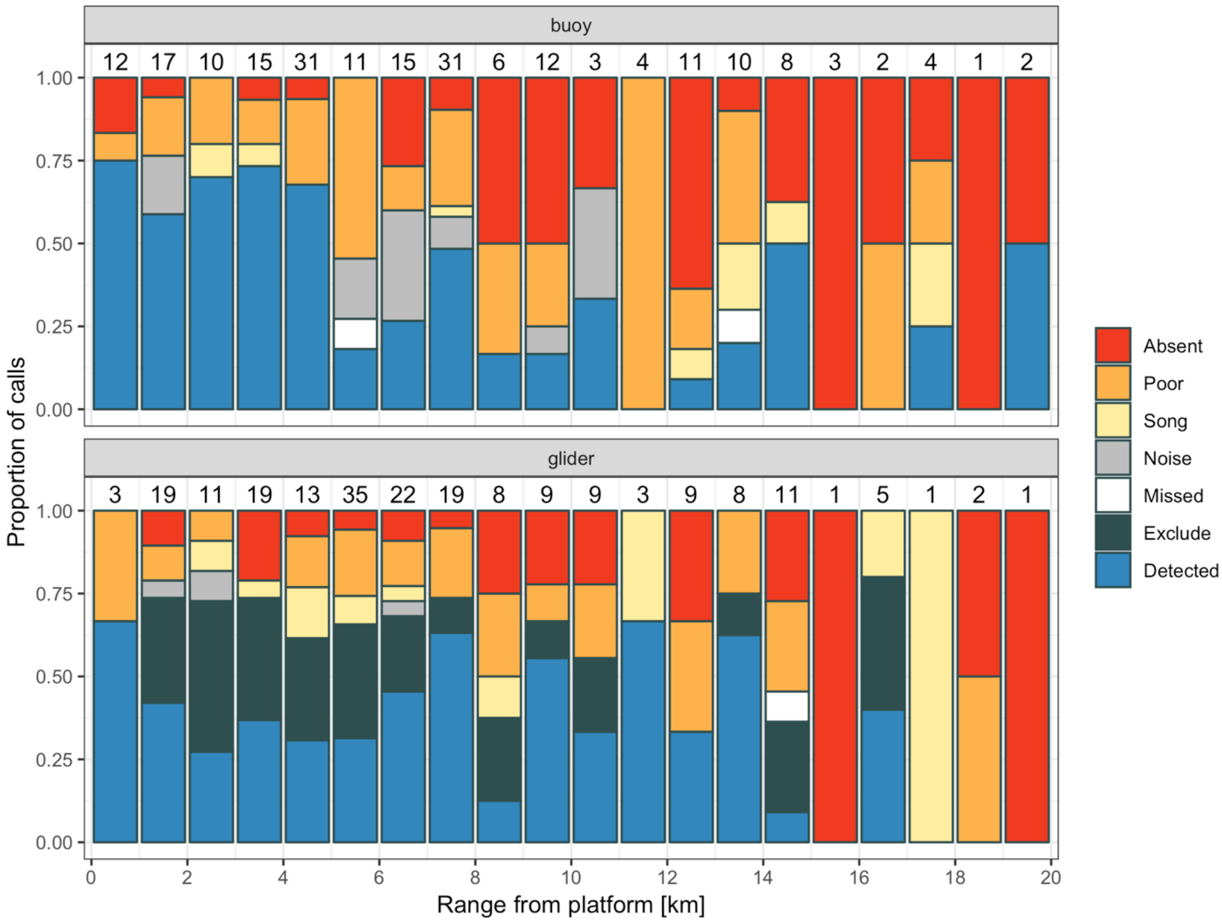


Figure 3.9. Reasons for missed calls by range.
Proportion of pitch track scores for localized calls received on each platform as a function of range (n = 208). Colors indicate the proportion of calls of a given score in 1-km range bins, while the number of calls in each bin is shown above each bar. Definitions of each category are provided in Table 3.1.

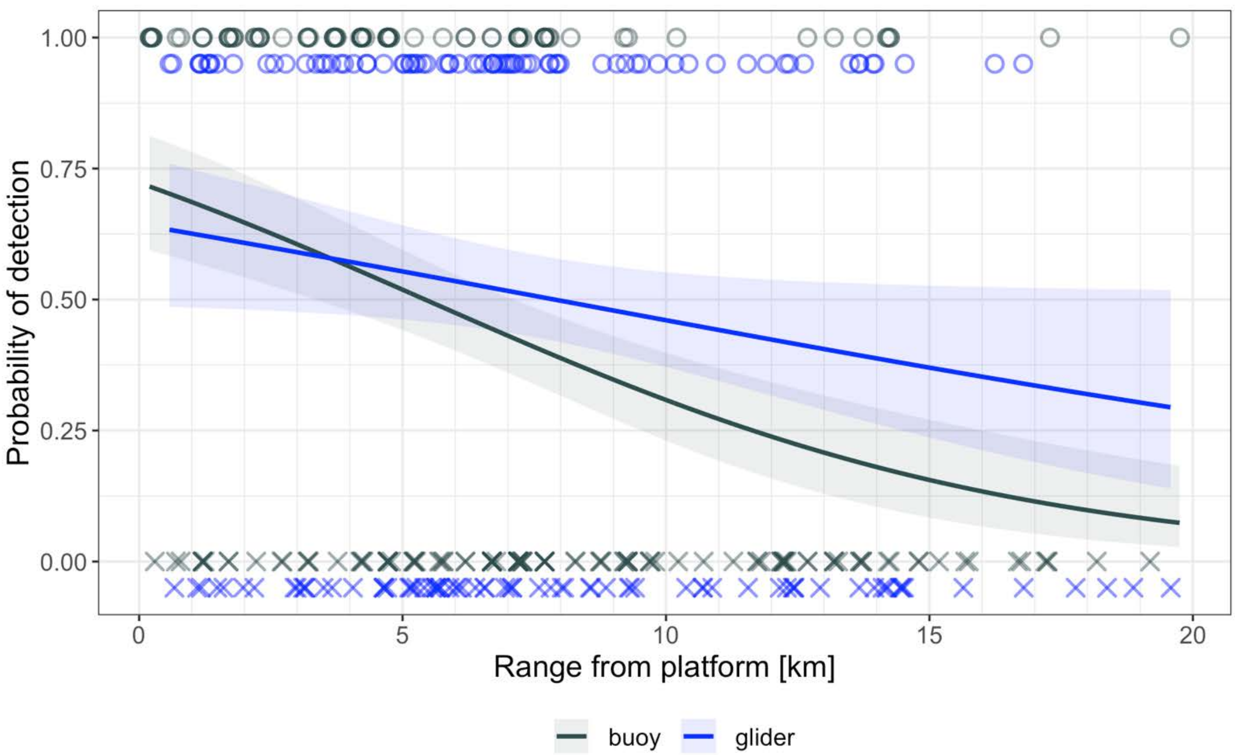


Figure 3.10. Probability of detecting localized calls.
Probability of detection of localized right whale upcalls as a function of range to the buoy (black; n = 208) and glider (blue; n = 156). The open circles and crosses indicate calls detected or undetected on either platform, respectively. The fitted regression models are shown as solid lines, while 95% confidence intervals are shown as shaded regions.

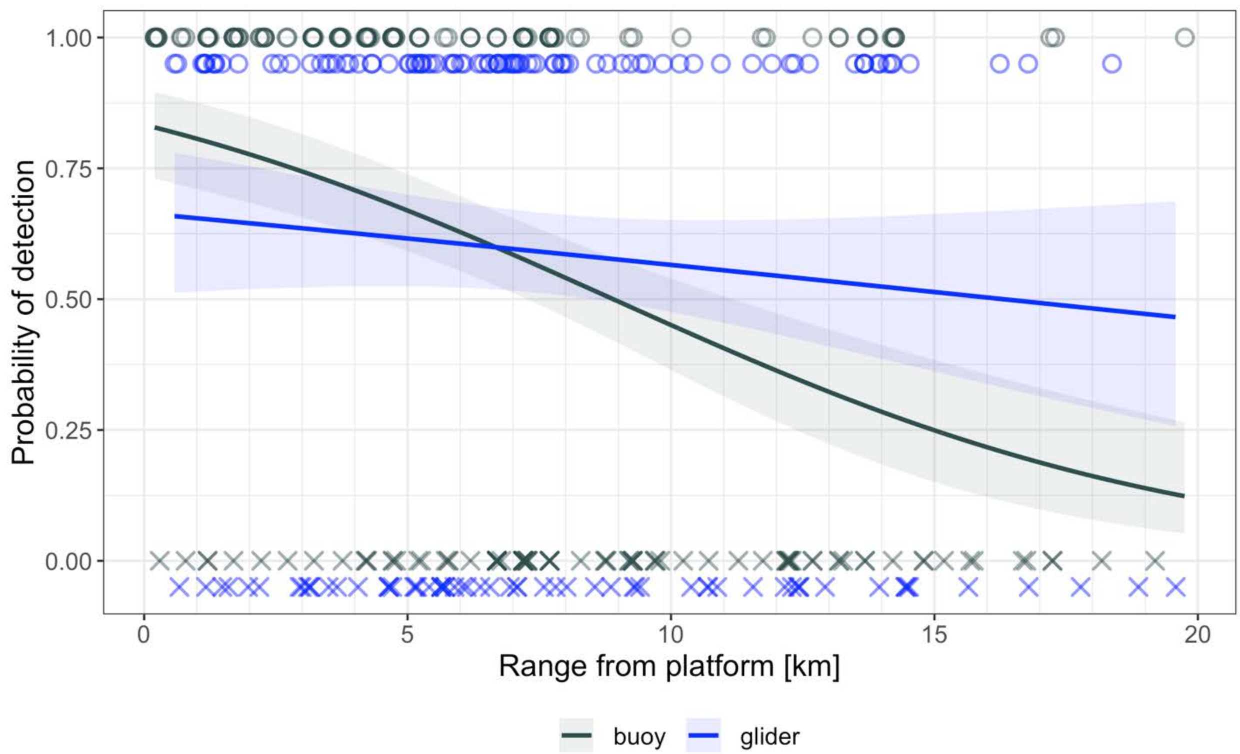


Figure 3.11. Probability of detecting localized calls, including "possible" detections.
Probability of detection of localized right whale upcalls as a function of range to the buoy (black; n = 208) and glider (blue; n = 156) when pitch tracks scored as "possible" are considered as detections. The open circles and crosses indicate calls detected or undetected on either platform, respectively. The fitted regression models are shown as solid lines, while 95% confidence intervals are shown as shaded regions.

4 References

Au, W. W. L., Pack, A. A., Lammers, M. O., Herman, L. M., Deakos, M. H., & Andrews, K. (2006). Acoustic properties of humpback whale songs. Journal of the Acoustical Society of America, 120, 1103-1110.

Baumgartner, M. F., Van Parijs, S. M., Wenzel, F. W., Tremblay, C. J., Esch, H. C., & Warde, A. M. (2008). Low frequency vocalizations attributed to sei whales (*Balaenoptera borealis*). Journal of the Acoustical Society of America, 124, 1339-1349.

Baumgartner, M. F., & Mussoline, S. E. (2011). A generalized baleen whale call detection and classification system. Journal of the Acoustical Society of America, 129, 2889-2902.

Baumgartner, M. F., Fratantoni, D. M., Hurst, T. P., Brown, M. W., Cole, T. V. N., Van Parijs, S. M., & Johnson, M. (2013). Real-time reporting of baleen whale passive acoustic detections from ocean gliders. Journal of the Acoustical Society of America, 134, 1814-1823.

Baumgartner. M. F., J. Bonnell, S. M. Van Parijs, P. J. Corkeron, C. Hotchkin, K. Ball, L. -P. Pelletier, J. Partan, D. Peters, J. Kemp, J. Pietro, K. Newhall, A. Stokes, T. V. N. Cole, E. Quintana & S. D. Kraus. Persistent near real-time passive acoustic monitoring for baleen whales from a moored buoy: system description and evaluation. Methods in Ecology and Evolution, in press.

Caswell, H., Fujiwara, M. & Brault, S. (1999). Declining survival probability threatens the North Atlantic right whale. Proceedings of the National Academy of Sciences, USA, 96, 3308-3313.

Cato, D. H. (1998). Simple methods of estimating source levels and locations of marine animal sounds. The Journal of the Acoustical Society of America, 104, 1667-1678.

Charif, R. A., Mellinger, D. K., Dunsmore, K. J., Fristrup, K. M., & Clark, C. W. (2002). Estimated source levels of fin whale *Balaenoptera physalus* vocalizations: Adjustments for surface interference. Marine Mammal Science, 18, 81-98.

Clark, C. W., Ellison, W. T., Hatch, L. T., Merrick, R. L., Van Parijs, S. M., & Wiley, D. N. (2008−2010). An ocean observing system for large-scale monitoring and mapping of noise throughout the Stellwagen Bank National Marine Sanctuary. Reports to the National Oceanographic Partnership Program, Award number: N00014-07-1-1029. Available at http://www.onr.navy.mil/reports/FY09/npclark.pdf

Clark, C. W., Brown, M. W., & P. Corkeron. (2010). Visual and acoustic surveys for North Atlantic right whales, *Eubalaena glacialis*, in Cape Cod Bay, Massachusetts, 2001–2005: Management implications. Marine Mammal Science, 26, 837-854.

Croll, D. A., Clark, C. W., Acevedo, A., Tershy, B., Flores, S., Gedamke, J., & Urban, J. (2002). Only male fin whales sing loud songs. Nature, 417, 809.

Cummings, W. C. & Holliday, D. V. (1986) Sounds and source levels from bowhead whales off Pt. Barrow, Alaska. The Journal of the Acoustical Society of America, 82, 814-821.

Davis, G. E., Baumgartner, M. F., Bonnell, J. M., Bell, J., Berchok, C. L., Bort Thornton, J., Brault, S., Buchanan, G., Charif, R. A, Cholewiak, D. M., Clark, C. W., Corkeron, P. J., Delarue, J., Dudzinski, K., Hatch, L. T., Hildebrand, J. A., Hodge, L., Klinck, H., Kraus, S. D., Martin, B. S., Mellinger, D. K., Moors-Murphy, H., Nieukirk, S., Nowacek, D. P., Parks, S. E., Read, A. J., Rice, A. N., Risch, D., Širović, A., Soldevilla, M., Stafford, K. M., Stanistreet, J. E., Summers, E., Todd, S., Warde, A., Van Parijs, S. M. (2017) Long-term passive acoustic recordings track the changing distribution of North Atlantic right whales (*Eubalaena glacialis*) from 2004 to 2014. Scientific Reports, 7, 13460-13460.

Durette-Morin, D., Davies, K. T. A, Johnson, H. D., Brown, M. W., Moors-Murphy, H., Martin, B., Taggart, C. T. (2019) Passive acoustic monitoring predicts daily variation in North Atlantic right whale presence and relative abundance in Roseway Basin, Canada. Marine Mammal Science, doi:10.1111/mms.12602

Greene, C. R. J., McLennan, M. W., Norman, R. G., McDonald, T. L, Jakubczak, R. S. & Richardson, W. J. (2004). Directional frequency and recording (DIFAR) sensors in seafloor recorders to locate calling bowhead whales during their fall migration. Journal of the Acoustical Society of America, 116, 799-813.

Hain, J. H. W., Ratnaswamy, M. J., Kenney, R. D., & Winn, H. E. (1992). The fin whale, *Balaenoptera physalus*, in waters of the northeastern United States continental shelf. Reports of the International Whaling Commission, 42, 653-669.

Harris, D., Matias, L., Thomas, L., Harwood, J. & Geissler, W. H. (2013) Applying distance sampling to fin whale calls recorded by single seismic instruments in the northeast Atlantic. The Journal of the Acoustical Society of America, 134, 3522-35.

Helble, T. A., D'Spain, G. L., Hildebrand, J. A., Campbell, G. S., Campbell, R. L. & Heaney, K. D. (2013) Site specific probability of passive acoustic detection of humpback whale calls from single fixed hydrophones. The Journal of the Acoustical Society of America, 134, 2556-70.

Johnson, M., & Hurst, T. (2007). The DMON: an open-hardware/open-software passive acoustic detector. 3rd International Workshop on the Detection and Classification of Marine Mammals using Passive Acoustics, Boston, Massachusetts, USA.

Johnson, M., Partan, J., & Hurst, T. (2013). Low complexity lossless compression of underwater sound recordings. Journal of the Acoustical Society of America, 133, 1387-1398.

Laurinolli, M., Hay, A. E., Desharnais, F. & Taggart, C. T. (2003) Localization of North Atlantic right whale sounds in the Bay of Fundy using a sonobuoy array. Marine Mammal Science, 19, 708-723.

Lin, Y. –T., Newhall, A. E. & Lynch, J. F. (2012) Low-frequency broadband sound source localization using an adaptive normal mode back-propagation approach in a shallow-water ocean. The Journal of the Acoustical Society of America, 131, 1798-1798.

Munger, L. M., Wiggins, S. M. & Hildebrand, J. A. (2011) North Pacific right whale up-call source levels and propagation distance on the southeastern Bering Sea shelf. The Journal of the Acoustical Society of America, 129, 4047-4054.

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

Morano, J. L., Salisbury, D. P., Rice, A. N., Conklin, K. L., Falk, K. L., & Clark, C. W. (2012). Seasonal and geographical patterns of fin whale song in the western North Atlantic Ocean. Journal of the Acoustical Society of America, 132, 1207-1212.

Newhall, A. E., Lin, Y-T., Lynch, J. F., Baumgartner, M. F., & Gawarkiewicz, G. G. (2012). An acoustic normal mode approach for long distance passive localization of vocalizing sei whales. Journal of the Acoustical Society of America, 131, 1814-1825.

Newhall, A. E., Lynch, J. F., Gawarkiewicz, G. G., Duda, T. F., McPhee, N. M., Bahr, F. B., Marquette, C. D., Lin, Y. –T., Jan, S., Wang, J., Chen, C. –F., Chiu, L. Y. S, Yang, Y. –J., Wei, R. –C., Emerson, C., Morton, D., Abbot, T., Abbot, P. A., Calder, B., Mayer, L. A. & Lermusiaux, P. F. J. (2010) Acoustics and oceanographic observations collected during the QPE Experiment by Research Vessels OR1, OR2 and OR3 in the East China Sea in the Summer of 2009. Woods Hole Oceanographic Institution.

Parks, S. E. & P. L. Tyack. (2005). Sound production by North Atlantic right whales (*Eubalaena glacialis*) in surface active groups. Journal of the Acoustical Society of America, 117, 3297-3306.

Parks, S. E., Johnson, M. P., Nowacek, D. P. & Tyack, P. L. (2011a) Individual right whales call louder in increased environmental noise. Biology Letters, 7, 33-35.

Parks, S. E., Searby, A., Célérier, A., Johnson, M. P., Nowacek, D. P. & Tyack, P. L. (2011b) Sound production behavior of individual North Atlantic right whales: implications for passive acoustic monitoring. Endangered Species Research, 15, 63-76.

Paul, W., & Bocconcelli, A. (1994). Drifting buoy systems using rubber stretch hose. Sea Technology, 35(11), 37-44.

Payne, R., & McVay, S. (1971). Songs of humpback whales. Science, 173, 585-597.

Porter, M. B. (1992) The KRAKEN Normal Mode Program. Naval Research Laboratory, 1-198.

Rankin, S., Barlow, J. & Oswald, J. (2008). An assessment of the accuracy and precision of localization of a stationary sound source using a 2-element towed hydrophone array. NOAA Tech Memo NOAA-TM-NMFS-SWFSC-416.

Sokal, R. R., & Rohlf, F. J. (1995). Biometry (3rd ed.), New York, NY: W. H. Freeman.

Tennessen, J. & Parks, S. E. (2016) Acoustic propagation modeling indicates vocal compensation in noise improves communication range for North Atlantic right whales. Endangered Species Research, 30, 225-237.

Thompson, P. O., Cummings, W. C., & Ha, S. J. (1986). Sounds, source levels and associated behavior of humpback whales, southeast Alaska. Journal of the Acoustical Society of America, 80, 735-740.

Van Parijs, S. M., Clark, C. W., Sousa-Lima, R. S., Parks, S. E., Rankin, S., Risch, D., & Van Opzeeland, I. C. (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.

Watkins, W. A., Tyack, P., Moore, K. E., & Bird, J. E. (1987). The 20-Hz signals of finback whales (*Balaenoptera physalus*). Journal of the Acoustical Society of America, 82, 1901-1912.

Wiggins, S. M., McDonald, M. A., Munger, L. M., Moore, S. E. & Hildebrand, J. A. (2004) Waveguide propagation allows range estimates for North Pacific right whales in the Bering sea. Canadian Acoustics, 32, 146-154.

Department of the Interior (DOI)

The Department of the Interior protects and manages the Nation's natural resources and cultural heritage; provides scientific and other information about those resources; and honors the Nation’s trust responsibilities or special commitments to American Indians, Alaska Natives, and affiliated island communities.

Bureau of Ocean Energy Management (BOEM)

The mission of the Bureau of Ocean Energy Management is to manage development of U.S. Outer Continental Shelf energy and mineral resources in an environmentally and economically responsible way.

BOEM Environmental Studies Program

The mission of the Environmental Studies Program is to provide the information needed to predict, assess, and manage impacts from offshore energy and marine mineral exploration, development, and production activities on human, marine, and coastal environments. The proposal, selection, research, review, collaboration, production, and dissemination of each of BOEM’s Environmental Studies follows the DOI Code of Scientific and Scholarly Conduct, in support of a culture of scientific and professional integrity, as set out in the DOI Departmental Manual (305 DM 3).